(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 114230675 B
(45)授权公告日 2025.02.18

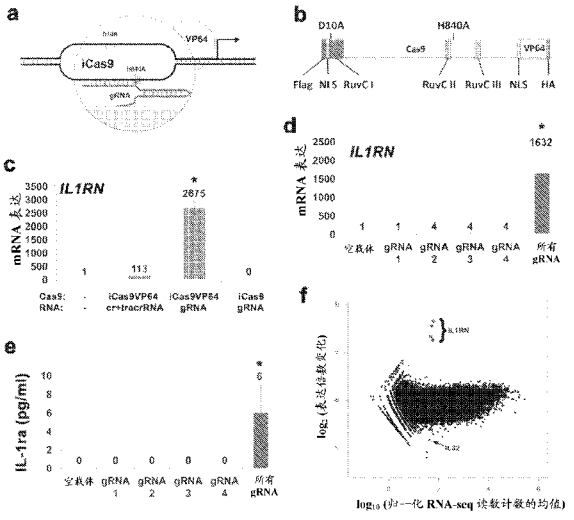
(21)申请号 202111496852.0
(22)申请日 2014.06.05
(65)同一申请的已公布的文献号
申请公布号 CN 114230675 A
(43)申请公布日 2022.03.25
(30)优先权数据
61/831481 2013.06.05 US
61/839127 2013.06.25 US
61/904911 2013.11.15 US
61/967466 2014.03.19 US
61/981575 2014.04.18 US
(62)分案原申请数据
201480044748.X 2014.06.05
(73)专利权人 杜克大学
地址 美国北卡罗来纳州
(72)发明人 C.A.格斯巴奇 I.B.希尔顿
P.佩雷兹皮内拉 A.M.卡巴迪
P.I.塔科雷 D.G.奥斯特劳特
J.B.布拉克
(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 初明明
(51)Int.Cl.
C07K 19/00(2006.01)
A61K 38/46(2006.01)
A61K 48/00(2006.01)
C12N 15/85(2006.01)
C12N 15/90(2006.01)
(56)对比文件
JP 2016521554 A,2016.07.25
Martin Jinck.A programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity.《Science》.2012,(第337期),第816-820页.
审查员 周洁
权利要求书3页 说明书98页
序列表184页 附图73页

(54)发明名称

RNA指导的基因编辑和基因调节

(57)摘要

本文公开了基于成簇有规间隔短回文重复序列(CRISPR)/CRISPR相关(Cas)9的系统相关组合物和使用所述基于CRISPR/Cas9的系统相关组合物用于改变基因表达和基因组工程改造的方法。本文还公开了使用所述组合物用于在肌肉例如骨骼肌和心肌中改变基因表达和基因组工程改造的组合物和方法。



CN 114230675 B

1.一种DNA打靶组合物,其包含:

融合蛋白,其包含第一多肽结构域和第二多肽结构域,所述第一多肽结构域包含成簇有规间隔短回文重复序列相关9(Cas9)蛋白,其中所述第二多肽结构域包含VP64结构域、KRAB结构域或p300结构域,

其中所述第二多肽结构域包含p300结构域,所述p300结构域包含SEQ ID NO: 676的氨基酸1422-2038,或者其中所述第二多肽结构域包含VP64结构域,所述VP64结构域包含SEQ ID NO: 675的氨基酸1422-1471,或者其中所述第二多肽结构域包含由包含SEQ ID NO: 466的核苷酸4222-4564的多核苷酸编码的KRAB结构域;和

至少一种指导RNA (gRNA)。

2.权利要求1的DNA打靶组合物,其中所述第二多肽结构域包含p300结构域,所述p300结构域包含SEQ ID NO: 676的氨基酸1422-2038。

3.权利要求1的DNA打靶组合物,其中所述第二多肽结构域包含VP64结构域,所述VP64结构域包含SEQ ID NO: 675的氨基酸1422-1471。

4.权利要求1的DNA打靶组合物,其中所述第二多肽结构域包含由包含SEQ ID NO: 466的核苷酸4222-4564的多核苷酸编码的KRAB结构域。

5.权利要求1的DNA打靶组合物,其中所述Cas9蛋白源自酿脓链球菌(*Streptococcus pyogenes*)、金黄色葡萄球菌(*Staphylococcus aureus*)或脑膜炎奈瑟球菌(*Neisseria meningitidis*)。

6.权利要求1的DNA打靶组合物,其中所述Cas9蛋白包含SEQ ID NO: 1的氨基酸36-1403,或者由包含SEQ ID NO: 145的核苷酸7-4122的多核苷酸编码,或者由包含SEQ ID NO: 459的核苷酸1-3168的多核苷酸编码,或者由包含SEQ ID NO: 461的核苷酸28-3270的多核苷酸编码。

7.权利要求1的DNA打靶组合物,其中所述融合蛋白包含SEQ ID NO: 675或SEQ ID NO: 676,或者由包含SEQ ID NO: 466的多核苷酸编码。

8.权利要求1的DNA打靶组合物,其中所述至少一种指导RNA靶向选自以下的基因的启动子区:*ASCL1*、*BRN2*、*MYT1L*、*NANOG*、*VEGFA*、*TERT*、*IL1B*、*IL1R2*、*IL1RN*、*HBG1*、*HBG2*和*MYOD1*;或者

其中所述至少一种指导RNA包含由SEQ ID NO: 5-40、65-144、492-515、540-563、585-625或679-750的至少一个编码的靶向结构域,或者

其中所述至少一种指导RNA靶向ASCL1,并且包含由SEQ ID NO: 5-8的至少一个编码的靶向结构域,或靶向包含SEQ ID NO: 468-515的至少一个的序列,或者

其中所述至少一种指导RNA靶向NANOG,并且包含由SEQ ID NO: 9-12的至少一个编码的靶向结构域,或者

其中所述至少一种指导RNA靶向VEGFA,并且包含由SEQ ID NO: 13-16的至少一个编码的靶向结构域,或者

其中所述至少一种指导RNA靶向TERT,并且包含由SEQ ID NO: 17-20的至少一个编码的靶向结构域,或者

其中所述至少一种指导RNA靶向IL1B,并且包含由SEQ ID NO: 21-24的至少一个编码的靶向结构域,或者

其中所述至少一种指导RNA靶向IL1R2,并且包含由SEQ ID NO: 25-28的至少一个编码的靶向结构域,或者

其中所述至少一种指导RNA靶向IL1RN,并且包含由SEQ ID NO: 29-32的至少一个编码的靶向结构域,或者

其中所述至少一种指导RNA靶向HBG1或HBG2,并且包含由SEQ ID NO: 33-36的至少一个编码的靶向结构域,或者

其中所述至少一种指导RNA靶向MYOD1,并且包含由SEQ ID NO: 37-40的至少一个编码的靶向结构域,或者

其中所述至少一种指导RNA靶向BRN2,并且靶向包含SEQ ID NO: 516-563的至少一个的序列,或者

其中所述至少一种指导RNA靶向HS2,并且靶向包含SEQ ID NO: 564-605的至少一个的序列,或者

其中所述至少一种指导RNA靶向肌养蛋白,并且包含由SEQ ID NO: 65、67、69、70、100或679-750的至少一个编码的靶向结构域,或靶向包含SEQ ID NO: 66、68、71-99、104-144或628-632的至少一个的序列。

9.一种多核苷酸,其编码权利要求1-8中任一项的DNA打靶组合物。

10.一种载体,其包含权利要求9的多核苷酸。

11.一种用于调节细胞中基因表达的组合物,所述基因选自*ASCL1*、*BRN2*、*MYT1L*、*NANOG*、*VEGFA*、*TERT*、*IL1B*、*IL1R2*、*IL1RN*、*HBG1*、*HBG2*、*MYOD1*、HS2和肌养蛋白,所述组合物包含权利要求1-8中任一项的DNA打靶组合物、编码权利要求1-8中任一项的DNA打靶组合物的多核苷酸、或包含编码权利要求1-8中任一项的DNA打靶组合物的多核苷酸的载体。

12.权利要求11的组合物,其中所述第二多肽结构域包含VP64结构域或p300结构域,并且所述基因表达增加。

13.权利要求11的组合物,其中所述第二多肽结构域包含KRAB结构域,并且所述基因表达降低。

14.一种用于诱导细胞中基因表达的组合物,所述组合物包含权利要求1-8中任一项的DNA打靶组合物、编码权利要求1-8中任一项的DNA打靶组合物的多核苷酸、或包含编码权利要求1-8中任一项的DNA打靶组合物的多核苷酸的载体,其中所述第二多肽结构域包含VP64结构域或p300结构域。

15.一种用于在受试者中调节或编辑基因的组合物,所述组合物包含权利要求1-8中任一项的DNA打靶组合物、编码权利要求1-8中任一项的DNA打靶组合物的多核苷酸、或包含编码权利要求1-8中任一项的DNA打靶组合物的多核苷酸的载体。

16.权利要求15的组合物,其中所述第二多肽结构域包含VP64结构域或p300结构域,并且所述基因表达增加。

17.权利要求15的组合物,其中所述第二多肽结构域包含KRAB结构域,并且所述基因表达降低。

18.一种细胞,其包含权利要求1-8中任一项的DNA打靶组合物、编码权利要求1-8中任一项的DNA打靶组合物的多核苷酸、包含编码权利要求1-8中任一项的DNA打靶组合物的多核苷酸的载体或权利要求11-17中任一项的组合物。

19.一种药盒,其包含权利要求1-8中任一项的DNA打靶组合物、编码权利要求1-8中任一项的DNA打靶组合物的多核苷酸、包含编码权利要求1-8中任一项的DNA打靶组合物的多核苷酸的载体、权利要求11-17中任一项的组合物或权利要求18的细胞。

20.权利要求1-8中任一项的DNA打靶组合物、权利要求9的多核苷酸、权利要求10的载体或权利要求11-17中任一项的组合物在制备用于调节细胞中至少一个内源基因的表达的药物中的用途,其中所述细胞具有选自以下的疾病或病况:骨骼肌病况、或心肌病况、或其组合。

21.权利要求1-8中任一项的DNA打靶组合物、权利要求9的多核苷酸、权利要求10的载体或权利要求11-17中任一项的组合物在制备用于激活细胞中的内源基因的药物中的用途,其中所述第二多肽结构域包含VP64结构域或p300结构域,其中所述细胞具有选自以下的疾病或病况:骨骼肌病况、或心肌病况、或其组合。

22.权利要求1-8中任一项的DNA打靶组合物、权利要求9的多核苷酸、权利要求10的载体或权利要求11-17中任一项的组合物在制备用于使细胞转分化或诱导细胞分化的药物中的用途,其中所述细胞具有选自以下的疾病或病况:骨骼肌病况、或心肌病况、或其组合。

23.权利要求1-8中任一项的DNA打靶组合物、权利要求9的多核苷酸、权利要求10的载体或权利要求11-17中任一项的组合物在制备用于在细胞中进行多重基因编辑的药物中的用途,其中所述细胞具有选自以下的疾病或病况:骨骼肌病况、或心肌病况、或其组合。

24.权利要求1-8中任一项的DNA打靶组合物、权利要求9的多核苷酸、权利要求10的载体或权利要求11-17中任一项的组合物在制备用于纠正细胞中的突变基因的药物中的用途,其中所述细胞具有选自以下的疾病或病况:骨骼肌病况、或心肌病况、或其组合。

25.权利要求1-8中任一项的DNA打靶组合物、权利要求9的多核苷酸、权利要求10的载体或权利要求11-17中任一项的组合物在制备用于在受试者肌肉中进行基因组编辑的药物中的用途,其中所述肌肉是骨骼肌或心肌,并且其中所述受试者具有选自以下的疾病或病况:骨骼肌病况、或心肌病况、或其组合。

26.权利要求1-8中任一项的DNA打靶组合物、权利要求9的多核苷酸、权利要求10的载体或权利要求11-17中任一项的组合物在制备用于治疗受试者的疾病或病况的药物中的用途,其中所述受试者具有选自以下的疾病或病况:骨骼肌病况、或心肌病况、或其组合。

RNA指导的基因编辑和基因调节

[0001] 相关申请的交叉引用

[0002] 本申请要求2013年6月5日提交的美国临时申请号61/831,481、2013年6月25日提交的美国临时申请号61/839,127、2013年11月15日提交的美国临时申请号61/904,911、2014年3月19日提交的美国临时申请号61/967,466和2014年4月18日提交的美国临时申请号61/981,575的优先权,其所有通过引用以其整体结合到本文中。

[0003] 政府利益声明

[0004] 本发明在由NIH授予的联邦资助号DP2-OD008586和R01DA036865以及美国国家科学基金会(National Science Foundation)授予的CBET-1151035的政府支持下完成。美国政府对本发明享有一定权利。

技术领域

[0005] 本公开内容涉及使用基于成簇有规间隔短回文重复序列(Clustered Regularly Interspaced Short Palindromic Repeat,CRISPR)/CRISPR相关(Cas) 9系统和病毒递送系统的基因的基因表达改变、基因组工程改造和基因组改变的领域。本公开内容还涉及肌肉(例如骨骼肌和心肌)中基因的基因组工程改造和基因组改变的领域。

发明背景

[0006] 已对合成转录因子进行了改造以控制基因表达用于哺乳动物系统中的许多不同的医学和科学应用,包括刺激组织再生、药物筛选、补偿遗传缺陷、激活沉默的肿瘤抑制基因、控制干细胞分化、进行遗传筛选和产生合成基因回路。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以识别与哺乳动物基因组正交的序列用于转基因调节。用于工程改造靶向用户定义序列的新转录因子的最通用策略基于锌指蛋白和转录-激活因子样效应物(TALE)的可编程DNA结合结构域。这两种方法包括应用这些结构域的蛋白质-DNA相互作用原理以改造具有独特DNA结合特异性的新的蛋白质。虽然这些方法对于许多应用是极其成功的,但是操作蛋白质-DNA相互作用所需要的蛋白质工程改造可能是费力的,并且需要专门技术。

[0007] 另外,这些新的蛋白质不总是有效的。其原因尚未知,但可能与外遗传修饰和染色质状态对蛋白质与基因组靶位点结合的作用有关。另外,在确保这些新的蛋白质以及其它成分递送到各细胞中的方面存在挑战。用于递送这些新的蛋白质及其多种成分的现有方法包括用不同的质粒或载体递送至细胞,由于拷贝数的差异,因此导致在各细胞中高度可变的表达水平。另外,在转染后的基因激活由于质粒DNA的稀释所致是瞬时的,且短暂的基因表达可能不足以诱导治疗作用。此外,该方法不适于不易转染的细胞类型。因此这些新的蛋白质的另一个限制是转录激活的效能。

[0008] 位点特异性核酸酶可用于在靶定的基因组基因座上引入位点特异性双链断裂。这种DNA切割刺激天然DNA修复机器,导致两种可能的修复途径之一。在不存在供体模板时,断裂可通过非同源末端连接(NHEJ) (一种导致DNA的少量插入或缺失的易错修复途径)修复。

该方法可用来有意破坏、缺失或改变靶定基因序列的读框。然而,如果供体模板连同核酸酶一起提供,则细胞机器将通过同源重组修复断裂,在DNA切割存在时,这被提高几个数量级。可采用该方法将特定变化引入DNA序列的靶位点上。工程改造的核酸酶已被用于在多种人干细胞和细胞系中的基因编辑和用于小鼠肝中的基因编辑。然而,实施这些技术的主要障碍是以有效、高效并促进成功的基因组修饰的方式体内递送至特定组织。

[0009] 在美国,遗传性遗传疾病对儿童具有毁灭性作用。这些疾病目前无法治愈,并且只能通过试图减轻症状来控制。几十年来,基因疗法领域有希望治愈这些疾病。然而有关将治疗性基因安全和有效地递送给细胞和患者的技术障碍限制了这种方法。杜兴型肌营养不良(Duchenne Muscular Dystrophy,DMD)是最普遍的遗传性单基因病,3500名男性中发生1例。DMD是肌养蛋白基因的遗传突变或自发突变的结果。肌养蛋白是负责调节肌细胞完整性和功能的蛋白质复合体的关键成分。DMD患者通常在儿童时代丧失身体上自我支持的能力,在十几岁时变得日益衰弱,在其二十岁左右死亡。对于DMD现有的实验性基因疗法策略需要重复给予瞬时基因递送载体或依赖外来遗传物质永久融合至基因组DNA中。这两种方法有着严重的安全顾虑。此外,这些策略受限于不能递送大和复杂的肌养蛋白基因序列。

发明概述

[0010] 本发明涉及包含2个异源多肽结构域的融合蛋白。第一多肽结构域包含成簇有规间隔短回文重复序列相关(Cas)蛋白,第二多肽结构域具有选自以下的活性:转录激活活性、转录阻遏活性、转录释放因子活性、组蛋白修饰活性、核酸酶活性、核酸缔合活性、甲基化酶活性和脱甲基酶活性。Cas蛋白可包含Cas9。Cas9可包含敲除Cas9的核酸酶活性的至少一个氨基酸突变。至少一个氨基酸突变可以是D10A和H840A的至少一个。Cas蛋白可包含iCas9 (SEQ ID NO: 1的氨基酸36-1403)。第二多肽结构域可具有转录激活活性。第二多肽结构域可包含至少一个VP16转录激活结构域重复序列。第二多肽结构域可包含VP16四聚体(“VP64”)或p65激活结构域。融合蛋白还可包含使第一多肽结构域与第二多肽结构域连接的接头。融合蛋白可包含iCas9-VP64。

[0011] 本发明涉及包含所述融合蛋白和至少一种指导RNA (gRNA)的DNA打靶系统。至少一个gRNA可包含靶DNA序列的12-22个碱基对互补多核苷酸序列接着前间区序列邻近基序(protospacer-adjacent motif)。至少一个gRNA可靶向基因的启动子区、基因的增强子区或基因的转录区。至少一个gRNA可靶向基因的内含子。至少一个gRNA可靶向基因的外显子。至少一个gRNA可靶向选自以下基因的启动子区:*ASCL1*、*BRN2*、*MYT1L*、*NANOG*、*VEGFA*、*TERT*、*IL1B*、*IL1R2*、*IL1RN*、*HBG1*、*HBG2*和*MYOD1*。至少一个gRNA可包含SEQ ID NO: 5-40、65-144、492-515、540-563和585-625的至少一个。

[0012] 本发明涉及与肌养蛋白基因结合的包含Cas9和至少一种指导RNA (gRNA)的DNA打靶系统。至少一种gRNA可靶向肌养蛋白基因的内含子。至少一种gRNA可靶向肌养蛋白基因的外显子。至少一种指导RNA可包含SEQ ID NO: 5-40、65-144、492-515、540-563和585-625的至少一个。DNA打靶系统可包含介于1种和10种之间的不同gRNA。

[0013] 本发明涉及编码所述融合蛋白或所述DNA打靶系统的分离多核苷酸。

[0014] 本发明涉及包含所述分离的多核苷酸的载体。

[0015] 本发明涉及包含所述分离的多核苷酸或所述载体的细胞。

[0016] 本发明涉及调节细胞中的哺乳动物基因表达的方法。所述方法包括使细胞与所述融合蛋白、所述DNA打靶系统、所述分离的多核苷酸或所述载体接触。基因表达可以是诱导的。

[0017] 本发明涉及使细胞转分化或诱导细胞分化的方法。所述方法包括使细胞与所述融合蛋白、所述DNA打靶系统、所述分离的多核苷酸或所述载体接触。细胞可以是成纤维细胞或诱导多能干细胞。成纤维细胞可被转分化成为神经元细胞或肌源细胞。可使DNA打靶系统与细胞接触,且至少一种gRNA靶向选自以下的至少一个基因的启动子区:*ASCL1*、*BRN2*、*MYOD1*和*MYT1L*。DNA打靶系统可包含靶向*ASCL1*基因的启动子区的至少一种gRNA和靶向*BRN2*基因的启动子区的至少一种gRNA。DNA打靶系统可包含介于1种和20种之间的不同的gRNA。DNA打靶系统可包含8或16种不同的gRNA。DNA打靶系统可包含dCas9-VP64。DNA打靶系统可以病毒或非病毒的方法递送到细胞中。

[0018] 本发明涉及纠正细胞中的突变基因的方法。所述方法包括将所述DNA打靶系统、所述分离的多核苷酸或所述载体给予细胞。突变基因的纠正可包括同源指导修复。所述方法还可包括将供体DNA给予细胞。突变基因可包含引起提前终止密码子和截短基因产物的移码突变。突变基因的纠正可包括核酸酶介导的非同源末端连接。突变基因的纠正可包括提前终止密码子的缺失、剪接受体位点的破坏、一个或多个外显子的缺失或剪接供体序列的破坏。一个或多个外显子的缺失可导致读框的纠正。

[0019] 本发明涉及治疗具有突变型肌养蛋白基因的有需要的受试者的方法。所述方法包括将所述DNA打靶系统、所述分离的多核苷酸或所述载体给予受试者。受试者可能患有杜兴型肌营养不良。

[0020] 本发明涉及纠正细胞的突变型肌养蛋白基因的方法。所述方法包括将所述DNA打靶系统、所述分离的多核苷酸、所述载体或所述细胞给予含有突变型肌养蛋白基因的细胞。突变型肌养蛋白基因可包含提前终止密码子、通过基因缺失的受损读框、异常剪接受体位点或异常剪接供体位点,且其中靶区为提前终止密码子、受损读框、异常剪接受体位点或异常剪接供体位点的上游或下游。突变型肌养蛋白基因的纠正可包括同源指导修复(homology-directed repair)。所述方法还包括将供体DNA给予细胞。突变型肌养蛋白基因可包含引起提前终止密码子和截短基因产物的移码突变。突变型肌养蛋白基因的纠正可包括核酸酶介导的非同源末端连接。突变型肌养蛋白基因的纠正可包括提前终止密码子的缺失、受破坏读框的纠正或通过破坏剪接受体位点或破坏剪接供体序列的剪接调节。突变型肌养蛋白基因的纠正可包括外显子45-55或外显子51的缺失。

[0021] 本发明涉及包含所述融合蛋白、所述DNA打靶系统、所述分离的多核苷酸、所述载体或所述细胞的药盒。

[0022] 本发明涉及调节细胞中的哺乳动物基因表达的方法。所述方法包括使细胞与编码DNA打靶系统的多核苷酸接触。DNA打靶系统包含所述融合蛋白和至少一种指导RNA(gRNA)。DNA打靶系统可包含介于1种和10种之间的不同gRNA。不同的gRNA可与靶基因内的不同靶区结合。靶区可被至少一个核苷酸分隔。靶区可被约15-约700个碱基对分隔。不同的gRNA各自可与至少一个不同的靶基因结合。不同的靶基因可位于同一染色体上。不同的靶基因可位于不同的染色体上。至少一个靶区可在非开放染色质区、开放染色质区、靶基因的启动子区、靶基因的增强子区、靶基因的转录区或靶基因的转录起始位点的上游区域内。至

少一个靶区可位于靶基因的转录起始位点上游约1-约1000个碱基对之间。至少一个靶区可位于靶基因的转录起始位点上游约1-约600个碱基对之间。基因表达可以是诱导的。DNA打靶系统可包含2种不同的gRNA、3种不同的gRNA、4种不同的gRNA、5种不同的gRNA、6种不同的gRNA、7种不同的gRNA、8种不同的gRNA、9种不同的gRNA或10种不同的gRNA。至少一种指导RNA可靶向选自以下基因的启动子区:*ASCL1*、*BRN2*、*MYT1L*、*NANOG*、*VEGFA*、*TERT*、*IL1B*、*IL1R2*、*IL1RN*、*HBG1*、*HBG2*和*MYOD1*。至少一种指导RNA可包含SEQ ID NO: 5-40、65-144、492-515、540-563和585-625的至少一个。至少一个靶区可在靶基因的内含子或外显子内。

[0023] 本发明涉及用于诱导细胞中哺乳动物基因表达的组合物。组合物包含所述融合蛋白和至少一种指导RNA (gRNA)。

[0024] 本发明涉及用于诱导细胞中哺乳动物基因表达的组合物。组合物包含编码所述融合蛋白和至少一种指导RNA (gRNA)的分离的多核苷酸序列。至少一种指导RNA可靶向选自以下基因的启动子区:*ASCL1*、*BRN2*、*MYT1L*、*NANOG*、*VEGFA*、*TERT*、*IL1B*、*IL1R2*、*IL1RN*、*HBG1*、*HBG2*和*MYOD1*。至少一种指导RNA可包含SEQ ID NO 5-40、65-144、492-515、540-563和585-625的至少一个。

[0025] 本发明涉及包含用于诱导细胞中哺乳动物基因表达的所述组合物的细胞。

[0026] 本发明涉及包含用于诱导细胞中哺乳动物基因表达的所述组合物或包含用于诱导细胞中哺乳动物基因表达的所述组合物的所述细胞的药盒。

[0027] 本发明涉及用于在细胞中诱导哺乳动物基因表达的药盒。药盒包含用于诱导细胞中哺乳动物基因表达的所述组合物或包含用于诱导细胞中哺乳动物基因表达的所述组合物的所述细胞。

[0028] 本发明涉及用于在受试者肌肉中进行基因组编辑的组合物。组合物包含修饰的腺伴随病毒(AAV)载体和编码位点特异性核酸酶的核苷酸序列。所述肌肉是骨骼肌或心肌。修饰的AAV载体可具有改进的心肌和骨骼肌组织嗜性。位点特异性核酸酶可包括锌指核酸酶、TAL效应核酸酶或CRISPR/Cas9系统。位点特异性核酸酶可结合肌肉细胞中的基因或基因座。基因或基因座可以是肌养蛋白基因。组合物还可包含供体DNA或转基因。

[0029] 本发明涉及包含用于在受试者肌肉中进行基因组编辑的所述组合物的药盒。

[0030] 本发明涉及在受试者肌肉中进行基因组编辑的方法。所述方法包括将用于在受试者肌肉中进行基因组编辑的所述组合物给予肌肉,其中所述肌肉是骨骼肌或心肌。基因组编辑可包括纠正突变基因或插入转基因。纠正突变基因可包括缺失、重排或置换突变基因。纠正突变基因可包括核酸酶介导的非同源末端连接或同源指导修复。

[0031] 本发明涉及治疗受试者的方法。所述方法包括将用于在受试者肌肉中进行基因组编辑的所述组合物给予受试者的肌肉,其中所述肌肉是骨骼肌或心肌。受试者可患有骨骼肌病况或遗传疾病。受试者可患有杜兴型肌营养不良。

[0032] 本发明涉及纠正受试者的突变基因的方法,所述方法包括给予用于在受试者肌肉中进行基因组编辑的所述组合物。所述肌肉是骨骼肌或心肌。可将组合物注射入受试者的骨骼肌中。可将组合物经系统注射到受试者。骨骼肌可以是胫骨前肌。

[0033] 本发明涉及用于在受试者中进行基因组编辑的修饰的慢病毒载体,所述载体包含编码所述融合蛋白的第一多核苷酸序列和编码至少一种sgRNA的第二多核苷酸序列。第一多核苷酸序列可与第一启动子有效连接。第一启动子可以是组成型启动子、诱导型启动子、

阻抑型启动子或可调节启动子。第二多核苷酸序列可编码介于1种和10种之间的不同sgRNA。第二多核苷酸序列可编码2种不同的sgRNA、3种不同的sgRNA、4种不同的sgRNA、5种不同的sgRNA、6种不同的sgRNA、7种不同的sgRNA、8种不同的sgRNA、9种不同的sgRNA或10种不同的sgRNA。编码不同sgRNA的多核苷酸序列各自可与启动子有效连接。与不同sgRNA有效连接的启动子各自可以是相同的启动子。与不同sgRNA有效连接的启动子各自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、阻抑型启动子或可调节启动子。sgRNA可与靶基因结合。sgRNA各自可与一个靶基因座内的不同靶区结合。sgRNA各自可与不同基因座内的不同靶区结合。融合蛋白可包含Cas9蛋白或iCas9-VP64蛋白。融合蛋白可包含VP64结构域、p300结构域或KRAB结构域。两个或更多个内源基因在转录上可被激活。两个或更多个内源基因可被阻遏。

[0034] 本发明涉及激活细胞中的内源基因的方法。所述方法包括使细胞与所述修饰的慢病毒载体接触。内源基因可被瞬时激活。内源基因可被稳定激活。内源基因可被瞬时阻抑。内源基因可被稳定阻抑。融合蛋白可以与sgRNA相似的水平表达。融合蛋白可以与sgRNA不同的水平表达。细胞可以是原代人细胞。

[0035] 本发明涉及在细胞中进行多重基因编辑的方法。所述方法包括使细胞与所述修饰的慢病毒载体接触。多重基因编辑可包括纠正至少一个突变基因或插入转基因。纠正突变基因可包括缺失、重排或置换至少一个突变基因。纠正至少一个突变基因可包括核酸酶介导的非同源末端连接或同源指导修复。多重基因编辑可包括缺失至少一个基因,其中基因是内源正常基因或突变基因。多重基因编辑可包括缺失至少两个基因。多重基因编辑可包括缺失介于2个和10个之间的基因。

[0036] 本发明涉及调节细胞中至少一个靶基因的基因表达的方法。所述方法包括使细胞与所述修饰的慢病毒载体接触。至少两个基因的基因表达可被调节。可调节介于2个基因和10个基因之间的基因表达。当与至少一个靶基因的正常基因表达水平相比,至少一个靶基因的基因表达水平提高或降低时,至少一个靶基因的基因表达可被调节。

附图简述

[0037] 图1显示通过iCas9-VP64的人IL1RN基因的RNA指导的激活。(a、b)通过使失活Cas9(iCas9,D10A/H840A)与VP64转录激活结构域(transactivation domain)融合产生RNA指导的转录激活因子。iCas9-VP64通过指导RNA (gRNA)与20 bp靶序列杂交识别基因组靶位点。(c)将靶向*IL1RN*启动子中的序列的4种gRNA或crRNA/tracrRNA的表达质粒与iCas9-VP64表达质粒一起共转染到HEK293T细胞中。通过qRT-PCR评价*IL1RN*表达的激活。(d) 4种gRNA表达质粒分别或组合地与iCas9-VP64一起共转染。通过qRT-PCR只在响应gRNA的组合时才观察到稳健的基因激活。(e)通过经ELISA评价培养基中的IL-1ra基因产物的分泌,来证实*IL1RN*表达的激活。只在用gRNA的组合处理的6个样品的3个中检出IL-1ra。对于(c-e),数据表示为均值±s.e.m. (n = 3个独立实验)。通过Tukey检验,用gRNA的组合处理与所有其它处理显著不同(*P ≤ 0.02)。(f)对用空表达载体处理(n = 2)或用iCas9-VP64的表达质粒和靶向*IL1RN*的4种gRNA共转染(n = 2)的样品进行RNA-seq。这些处理间,唯一的基因表达的统计显著性变化是4个*IL1RN*同种型的增加(错误发现率≤ 3 x 10-4)和*IL32*的降低(错误发现率= 0.03)。

[0038] 图2显示与细胞和基因疗法、遗传重编程和再生医学有关的RNA指导的人基因激活。将HEK293T细胞用iCas9-VP64表达质粒和4种gRNA单独或组合地转染。靶基因表达通过qRT-PCR测量,并归一化至*GAPDH* mRNA水平。数据表示为均值±s.e.m. (n = 3个独立实验)。通过Tukey检验,用gRNA的组合处理在统计上显著不同于所有其它处理(*P< 0.05)。

[0039] 图3显示iCas9-VP64的表达。转染的HEK293细胞中iCas9-VP64的表达通过针对N端Flag表位标签的蛋白质印迹证实。wt Cas9表达质粒不含表位标签。

[0040] 图4显示人靶基因的gRNA靶位点和DNA酶高敏感性的位置。各基因座的4个gRNA靶位点标为各基因之上的定制轨道(custom track),表明DNA酶高敏感性开放染色质区的DNase-seq数据显示于各基因下方。如前所述,在HEK293T细胞中进行DNase-seq以鉴定DNA酶高敏感区(Song等, *Cold Spring Harbor protocols* 2010, pdb prot5384 (2010); Song等 *Genome Res* 21, 1757-1767 (2011))。结果表明,开放染色质不是通过gRNA与iCas9-VP64组合用于基因激活的必要条件。

[0041] 图5显示通过iCas9-VP64失去核酸酶活性。将野生型Cas9或失活(D10A、H840A)iCas9-VP64表达质粒与靶向*IL1RN*启动子的4种不同的指导RNA的表达质粒一起共转染。通过Surveyor测定法,测定核酸酶活性(Guschin等, *Methods Mol Biol* 649, 247-256 (2010))。表明核酸酶活性和通过非同源末端连接的DNA修复的较低分子量条带只在用野生型Cas9处理后存在,支持核酸酶活性被iCas9-VP64灭除。

[0042] 图6显示用靶向*HBG1*和*HBG2*的gRNA处理的样品的RNA-seq。对用对照空表达载体处理(n = 3)或用iCas9-VP64的表达质粒和靶向*HBG1*的4种gRNA共转染(n = 2)的样品进行RNA-seq。这些gRNA中的3种还靶向*HBG2*。观察到*HBG1*和*HBG2*两者相对于对照的增加,但由于表达水平低而不是统计显著的。这些处理间基因表达唯一的统计显著性变化是*IL32* (错误发现率= 0.0007)和*TNFRS9* (错误发现率= 0.002)降低。

[0043] 图7显示Ascl1和γ-珠蛋白通过iCas9-VP64增量调节。将HEK293T细胞用iCas9-VP64和靶向*ASCL1*或*HBG1*启动子的4种gRNA转染。相应的Ascl1和γ-珠蛋白蛋白质产量的水平通过蛋白质印迹评价。在2个独立实验中,在HEK293T细胞中可检出低水平的这些蛋白质,而在iCas9-VP64处理后可检出表达增加。

[0044] 图8显示iCas9-VP64处理的鼠胚胎成纤维细胞中Ascl1下游靶标的激活。将小鼠胚胎成纤维细胞(MEF)用对照GFP表达质粒或iCas9-VP64表达质粒和靶向ASCL1的4种gRNA表达质粒的组合以50:50或75:25的比率转染。(a)人*ASCL1*启动子(SEQ ID NO: 3)中的gRNA靶位点在小鼠*ASCL1*启动子(SEQ ID NO: 4)中是保守的。靶位点用实线表示,转录区用短划线标明。(b)通过qRT-PCR测定,在iCas9-VP64/gRNA处理后2天,MEF中的*ASCL1*表达增加。(c-h)在神经诱导培养基中10天后,针对神经元分化的早期标志物Ascl1和Tuj1 (c-d)或更成熟的神经元分化的标志物Tuj1和MAP2 (d-f),对细胞染色。一些Tuj1阳性细胞呈现神经元形态(f-g),并发现单细胞对Tuj1和MAP2呈阳性(g)。(h)在iCas9-VP64/gRNA处理的培养物(约0.05%)中容易鉴定出Tuj1阳性细胞,但在对照中无。n = 3个独立样品和数据表示为均值±均值的标准误差。gRNA 75/25显著不同于gRNA 50/50和对照(*P < 0.01,Tukey检验)。

[0045] 图9显示(a) iCas9-VP64蛋白序列(SEQ ID NO: 1)和(b)具有U6启动子的gRNA表达盒的序列(SEQ ID NO: 2)。

[0046] 图10显示qRT-PCR的标准曲线。对于各基因,将具有最高表达水平的实验样品稀释

以绘制通过qRT-PCR测定的标准曲线以确保在适当动态范围内的有效扩增。所有扩增反应的效率在90-115%内。

[0047] 图11(a)-11(b)显示RNA指导的修复的验证。图11(a)显示在Cas9与空载体(阴性对照)或gRNA一起共转染至细胞中后2天自HEK 293T细胞收获的基因组DNA的Surveyor测定结果。图11(b)显示gRNA靶标的位置。图11(c)显示各gRNA的预期切割位点。

[0048] 图12显示如Surveyor测定法所示DMD 8036 (del48-50)细胞中RNA指导的修复。

[0049] 图13显示在整个基因座中通过PCR显示的DMD 8036 (del48-50)细胞中RNA指导的修复。野生型肌养蛋白基因的PCR产生大小为1447 bp的片段,而DMD 8036细胞系中突变基因的PCR显示缺失约817 bp。在引入基于CRISPR/Cas9的系统后的缺失条带为约630 bp。

[0050] 图14显示通过用MANDYS8 (抗肌养蛋白抗体)和GAPDH抗体(阳性对照)的蛋白质印迹所显示的DMD 8036 (del48-50)细胞中RNA指导的修复。

[0051] 图15显示说明靶向IL1RN启动子的iCas9-VP64的特异性结合的ChIP测序数据。将HEK 293T细胞用靶向IL1RN启动子的iCas9-VP64转染。

[0052] 图16显示肌养蛋白基因的CRISPR/Cas9打靶。(A)设计sgRNA序列以结合肌养蛋白基因的外显子45-55突变热区中的序列,使得基因编辑可恢复多个患者特异性突变的肌养蛋白表达。内含子的箭头表明经设计从基因组中缺失整个外显子的sgRNA靶标。外显子的箭头表明经设计产生肌养蛋白基因中的靶定移码(targeted frameshift)的sgRNA靶标。(B)在使用CR3 sgRNA通过外显子51的NHEJ DNA修复引入少量插入或缺失后读框纠正的实例。(C)经设计在具有外显子48-50缺失的患者突变中缺失外显子51并恢复肌养蛋白读框的多个sgRNA靶标的示意图。(D)经设计缺失整个外显子45-55区以克服多个DMD患者突变的多个sgRNA靶标的示意图。

[0053] 图17显示用于量化Surveyor测定结果以测量表7中的第3天基因修饰的TBE-PAGE凝胶的图像。星号标出说明核酸酶活性的预期的条带大小。

[0054] 图18显示用于量化Surveyor测定结果以测量表7中的第10天基因修饰的TBE-PAGE凝胶的图像。星号标出说明核酸酶活性的预期的条带大小。

[0055] 图19显示富集遗传修饰的DMD成肌细胞的荧光激活流式分选法。(A)使用T2A核糖体跳跃肽(ribosomal skipping peptide)序列,将表达与GFP标志物连接的人密码子优化SpCas9蛋白的质粒与携带sgRNA表达盒的一种或两种质粒一起共同电穿孔至人DMD成肌细胞中。(B)在有(底部)或没有(顶部)与SpCas9连接的GFP标志物的情况下通过T2A核糖体跳跃肽序列将所示sgRNA表达盒与表达SpCas9的各个质粒一起独立地共转染至HEK293T中。在转染后3天通过Surveyor测定法对基因修饰频率进行了评价。(C)将肌养蛋白基因中缺失外显子48-50的DMD成肌细胞用纠正这些患者细胞中的肌养蛋白读框的sgRNA处理。在电穿孔后20天对未分选(整批)或GFP+分选细胞中的基因修饰进行了评价。(D)在用指定的表达质粒电穿孔后3天DMD成肌细胞中的GFP表达。转染效率和分选细胞群由选通区表明。

[0056] 图20显示使用CRISPR/Cas9恢复肌养蛋白读框的靶定移码。(A)使用结合第一读框外终止密码子(out-of-frame stop codon)的紧接上游的sgRNA即CR3靶向外显子51的5'区。PAM:前间区序列邻近基序。(B)从用SpCas9和CR3表达盒处理的HEK293T细胞PCR扩增外显子51基因座。各个克隆的序列通过Sanger测序测定。顶部序列(粗体,外显子用红色)是天然未修饰序列。圆括号中表示各序列的克隆数。(C)产生于(B)中所示基因修饰的总基因编

辑效率和读框转换(frame conversion)的概述。(D)用SpCas9和CR3 sgRNA表达盒(图19C)处理以产生恢复肌养蛋白读框的靶定移码的人DMD成肌细胞中肌养蛋白表达的蛋白质印迹。分化6天后使用针对肌养蛋白的rod结构域的抗体探测肌养蛋白表达。

[0057] 图21显示采用多重CRISPR/Cas9基因编辑从人基因组缺失外显子51。(A)在具有外显子48-50缺失的人DMD成肌细胞中跨越外显子51基因座的终点基因组PCR。顶部箭头表明全长PCR扩增子的预期位置,2个较低的箭头表明具有因指定sgRNA组合引起的缺失的PCR扩增子的预期位置。(B)使(A)的PCR产物克隆,对各克隆测序以确定靶定基因座上存在的插入和缺失。顶行显示野生型未修饰序列,三角形表明SpCas9切割位点。在右边是显示预期缺失连接处的序列的代表性层析图。(C)用指定sgRNA处理的CRISPR/Cas9修饰的人Δ48-50 DMD成肌细胞中肌养蛋白mRNA转录物的终点RT-PCR分析。右边显示预期缺失PCR产物的代表性层析图。星号:产生于缺失产物链与未修饰链杂交的条带。(D)通过针对用GAPDH作为加载对照的肌养蛋白的蛋白质印迹,对通过CRISPR/Cas9基因组编辑拯救肌养蛋白表达进行了评价。箭头表明预期恢复的肌养蛋白条带。

[0058] 图22显示人DMD成肌细胞中通过多重CRISPR/Cas9基因编辑缺失整个外显子45-55区。(A)在HEK293T或DMD成肌细胞用指定sgRNA处理后检测内含子44和内含子55之间缺失区域的基因组DNA的终点基因组PCR。(B)通过Sanger测序以测定存在于靶定基因座上的基因组缺失的序列,来分析来自(A)中DMD成肌细胞的缺失的预期大小PCR产物的各克隆。下图是显示预期缺失连接处的序列的代表性层析图。(C)用指定sgRNA处理的CRISPR/Cas9修饰的人Δ48-50 DMD成肌细胞中肌养蛋白mRNA转录物的终点RT-PCR分析。右边显示预期缺失PCR产物的代表性层析图。(D)在将DMD成肌细胞用靶向内含子44和/或内含子55的sgRNA电穿孔后通过蛋白质印迹分析恢复的肌养蛋白表达。

[0059] 图23显示用于体内细胞移植实验的基于流式细胞术的富集基因修饰的DMD成肌细胞的证实。在有或没有CR1和CR5的sgRNA表达载体的情况下将DMD成肌细胞用Cas9处理,并通过流式细胞术分选GFP+细胞。使用位于基因座两侧的引物,通过终点PCR检测外显子51基因座上的缺失。阴性对照:仅用Cas9处理并针对GFP+细胞分选的DMD成肌细胞。

[0060] 图24显示在CRISPR/Cas9处理的人DMD成肌细胞移植到免疫缺陷型小鼠中后恢复的人肌养蛋白的体内表达。如图19所示将人Δ48-50 DMD成肌细胞用SpCas9、CR1和CR5处理以缺失外显子51并针对GFP表达分选。将这些分选细胞和未处理对照细胞注射入免疫缺陷型小鼠的后肢,并在移植后4周后评价肌纤维中的人特异性蛋白质表达。将冰冻切片按规定用抗人血影蛋白(其通过融合入小鼠肌纤维的未纠正和纠正的成肌细胞两者表达)或抗人肌养蛋白抗体染色。白色箭头表明肌纤维对人肌养蛋白呈阳性。

[0061] 图25显示探测人肌养蛋白表达的其它免疫荧光图像。左端顶部插图显示用抗人血影蛋白染色的区域的连续切片。(A-C)注射未处理人DMD成肌细胞的肌肉的切片。(D-F)通过流式细胞术富集的注射CR1/5处理的人DMD成肌细胞的肌肉的切片。白色箭头表示肌养蛋白阳性纤维。

[0062] 图26显示评价人细胞中的CRISPR/Cas9毒性和外显子51的CR1/CR5介导的缺失的脱靶作用(off-target effect)。(A)用人优化SpCas9和指定sgRNA构建体处理的HEK293T细胞中细胞毒性测定法的结果。细胞毒性基于用指定核酸酶共转染的GFP阳性细胞的存活。I-SceI是充分表征的无毒的大范围核酸酶,GZF3是已知有毒的锌指核酸酶。(B)用编码Cas9和

指定sgRNA的表达盒处理的分选hDMD细胞中脱靶位点的Surveyor分析。从在HEK293T细胞中测试的一组50个预测位点中鉴定出hDMD细胞中所测的这3个脱靶位点(图27和表4)。TGT:指定sgRNA的中靶基因座。OT:脱靶基因座。(C、D)检测(C)在用Cas9和CR1处理的HEK293T细胞或(D)在用Cas9、CR1和CR5处理的分选hDMD细胞中染色体易位的终点嵌套PCR。示意图表示针对各易位事件定制的嵌套引物对的相对位置。根据引物大小和各基因座上预测sgRNA切割位点的位置,估算各条带的预期大小。星号标出以预期大小检测的条带。通过从各末端的Sanger测序证实(C)中条带的特性(图30)。显示了HEK293T细胞中P2/P5易位的代表性层析图。

[0063] 图27显示用于量化Surveyor测定结果以测量表4中的中靶和脱靶基因修饰的TBE-PAGE凝胶的图像。星号标出表明核酸酶活性的预期的条带大小。

[0064] 图28显示检测人细胞中针对CR3和CR6/CR36由CRISPR/Cas9脱靶活性引起的染色体易位的终点嵌套PCR。在(A)按所示用Cas9和CR3处理的HEK293T或分选hDMD细胞、(B)仅用Cas9和CR36处理的HEK293T细胞、或(C)用Cas9、CR6和CR36表达盒处理的分选hDMD细胞中,采用嵌套终点PCR分析检测易位。使用各预测易位基因座的定制引物扩增易位的第二嵌套PCR反应物以使特异性最大化(参见表4)。示意图表示用于探测易位存在的嵌套引物对的相对位置。首先从自用或未用指定sgRNA处理的细胞分离的基因组DNA扩增各个可能的易位事件。使用可能产生于易位的预测PCR扩增子内的引物进行第二嵌套PCR反应。根据指定引物结合位点和各基因座上预测sgRNA切割位点估算预期大小。*表示以预期大小检测的条带,通过自各末端的Sanger测序来证实。#表示其中Sanger测序显示预测易位以外的序列的扩增子,可能为嵌套PCR期间错引发的结果。

[0065] 图29显示图28中检出的条带的Sanger测序层析图,所述条带产生于在用Cas9和CR3基因盒处理的HEK293T细胞中分别在染色体X和1上CR3和CR3-OT1之间的易位。箭头显示与在由合适的sgRNA引起的预期断裂点附近的指定染色体同源的区域。注意,因通过非同源末端连接的DNA修复的易错性质所致,测序读数在断裂点附近变成异相。

[0066] 图30显示图26C中检出的条带的Sanger测序层析图,所述条带产生于在用Cas9和CR1基因盒处理的HEK293T细胞中分别在染色体X和16上的CR1和CR1-OT1之间的易位。箭头显示与在由合适的sgRNA引起的预期断裂点附近的指定染色体同源的区域。注意,由通过非同源末端连接的DNA修复的易错性质所致,测序读数在断裂点附近变成异相。

[0067] 图31显示体内AAV注射和组织收获的概观。

[0068] 图32显示在递送AAV-SASTG-ROSA后体外和体内骨骼肌中的Rosa26 ZFN活性的Surveyor分析。箭头表示产生于Surveyor切割的预期条带。n.d.:未检出。(a)增殖的C2C12用指定量的病毒转导,在感染后4天收获。箭头表示产生于Surveyor切割的预期条带大小。(b)使C2C12在分化培养基中孵育5天,然后在24孔板中用指定量的AAV-SASTG-ROSA病毒转导。在转导后10天收集样品。(c)将指定量的AAV-SASTG-ROSA直接注射入C57BL/6J小鼠的胫骨前肌中,感染后4周收获肌肉。将收获的TA肌肉分割成8个单独的块用于基因组DNA分析,其各自在单独的泳道中显示。

[0069] 图33显示Rosa T2A opt DNA序列(SEQ ID NO: 434)和Rosa T2A opt蛋白质序列(SEQ ID NO: 435)。

[0070] 图34显示SASTG衣壳DNA序列(SEQ ID NO:436)和SASTG衣壳肽序列(SEQ ID NO:

437)。

[0071] 图35显示DZF16 ZFN靶位点序列(SEQ ID NO: 442)、DZF16-L6左侧全部氨基酸序列(SEQ ID NO: 443)和DZF16-R6右侧全部氨基酸序列(SEQ ID NO: 444)。

[0072] 图36显示E51C3 ZFN靶位点序列(SEQ ID NO: 445)、E51C3-3L左侧全部氨基酸序列(SEQ ID NO: 446)和E51C3-3R右侧全部氨基酸序列(SEQ ID NO: 447)。

[0073] 图37显示DZF15 ZFN靶位点序列(SEQ ID NO: 448)、DZF15-L6左侧全部氨基酸序列(SEQ ID NO: 449)、DZF15-R6右侧全部氨基酸序列(SEQ ID NO: 450)、DZF15-L5左侧全部氨基酸序列(SEQ ID NO: 451)、DZF15-R5右侧全部氨基酸序列(SEQ ID NO: 452)。

[0074] 图38显示E51C4 ZFN靶位点序列(SEQ ID NO: 453)、E51C4-4L左侧全部氨基酸序列(SEQ ID NO: 454)和E51C4-4R右侧全部氨基酸序列(SEQ ID NO: 455)。

[0075] 图39显示“单一载体,多重CRISPR系统”、“二元载体,多重CRISPR系统”和“单一载体,单一gRNA系统”的示意图。

[0076] 图40显示SaCas9-NLS (其中NLS用下划线表示) (SEQ ID NO: 64)和SaCas9 gRNA (SEQ ID NO: 116)的核苷酸序列。

[0077] 图41显示NmCas9 (其中NLS 1用下划线表示,NLS 2用下划线和粗体表示,HA标签用粗体表示)、来自Thomson PNAS 2013的NmCas9短发夹(SEQ ID NO: 118)和来自Church Nature Biotech 2013的NmCas9长发夹(SEQ ID NO: 119)的核苷酸序列。

[0078] 图42显示sgRNA和慢病毒Cas9表达构建体的验证。(a)将编码表达靶向AAVS1基因座的sgRNA的独特Pol III启动子的构建体或含有hU6启动子紧接多聚胸苷以终止表达的构建体(“PolyT”)转染至HEK293T细胞中。转染后2天采用终点RT-PCR以探测各指定启动子/sgRNA构建体的表达。-RT:无逆转录酶对照。(b)将HEK293T用编码AAVS1锌指核酸酶或Cas9-T2A-GFP和指定启动子/sgRNA表达盒的表达载体转染,采用Surveyor测定法评价转染后3天的基因修饰水平。(c)将HEK293T细胞用编码无sgRNA的指定Cas9-T2A-GFP构建体的慢病毒构建体转导,转导后7天,通过探测Cas9蛋白N端上的FLAG表位标签,通过蛋白质印迹评价Cas9表达。

[0079] 图43显示单一慢病毒CRISPR/Cas9表达盒的Golden Gate装配。

[0080] 图44显示多重CRISPR/Cas9系统的单一慢病毒递送。(a)将靶向不同基因组基因座的4种sgRNA克隆至表达有活性的Cas9核酸酶的慢病毒载体中。(b) HEK293T和原生人真皮成纤维细胞用表达指定sgRNA的慢病毒转导,并采用Surveyor测定法测定切割事件。转导后7天测定HEK293T。转导后10天测定人成纤维细胞。

[0081] 图45显示稳定表达dCas9-VP64的HEK293T中的瞬时基因激活。将HEK293T用慢病毒转导以稳定表达dCas9-VP64,随后用表达指定sgRNA组合的质粒转染。通过改变所递送的sgRNA数,转染后3天实现内源IL1RN (a)和HBG1 (b)基因座的可调节内源基因激活。转染后3-6天,观察到内源IL1RN (c)和HBG1 (d)的峰值水平,在第15-20天之间激活水平回到背景水平。重要的是,细胞系在第20天在第二次转染后能够起反应,虽然水平比之前观察到的低。

[0082] 图46显示使用单一慢病毒多个dCas9-VP64载体在HEK293T中稳定的基因激活。将HEK293T用慢病毒转导以稳定表达dCas9-VP64和gRNA的指定组合。通过改变所递送的sgRNA数,转导后7天实现内源IL1RN (a)和HBG1 (b)基因座的可调节内源基因激活。转导后6天观

察到内源IL1RN (c)和HBG1 (d)的峰值水平,并保持激活水平直到第21天。

[0083] 图47显示IL1RN mRNA表达水平。

[0084] 图48显示代表通过异位表达BAM神经元转录因子,成纤维细胞直接转化成神经元的示意图。

[0085] 图49显示(A) dCas9-VP64构建体的示意图。dCas9-VP64是一种与VP16转录激活结构域的四聚体融合的Cas9蛋白的催化上不活跃的形式。(B)显示RNA指导的dCas9-VP64募集至基因组靶上的机制的示意图。(C)产生具有CRISPR/Cas9转录因子的iN的实验方案的示意图。

[0086] 图50显示在用dCas9-VP64转导和用靶向*ASCL1*启动子、*ASCL1* cDNA或萤光素酶的gRNA的任一个转染的MEF中,在第3天通过(A) qRT-PCR测定的内源*ASCL1*表达或通过(B)免疫荧光检测的总*ASCL1*蛋白。星号(*)表示与4种gRNA相比,以共同递送8种gRNA的*ASCL1*表达显著增加($p<0.05$)。与通过dCas9-VP64和靶向Ascl1启动子的8种gRNA所诱导的相比,*ASCL1*的异位表达产生更多的蛋白质,但到第3天,不激活培养物中的内源基因座。

[0087] 图51显示(A)通过异位BAM因子或通过dCas9-VP64和靶向*BRN2*和*ASCL1*启动子的gRNA产生的*TUJ1*和*MAP2*阳性细胞,(B)在第11天在N3培养基中具有表达hSyn-RFP报道分子的神经元形态的细胞。

[0088] 图52显示(A)在KCl存在(底部)或不存在(顶部)时在培养基中具有对GCaMP5钙指示物呈阳性的神经元形态的细胞。(B)显示细胞在响应KCl加入时去极化的随时间变化的归一化荧光强度的迹线。

[0089] 图52显示iCas9-VP64处理的鼠胚胎成纤维细胞中使用dCas9-VP64转录因子以将成纤维细胞转化成神经元的Ascl1和Brn2 (即主调节基因)的下游靶标的激活。将小鼠胚胎成纤维细胞(MEF)用对照GFP表达质粒或iCas9-VP64表达质粒及靶向ASCL1和BRN2的8种gRNA表达质粒的组合转染。dCas9转录因子通过病毒递送。在10天后在神经诱导培养基中,针对Tuj1 (神经元分化的早期标志物)和MAP2 (更成熟的神经元分化的标志物)对细胞染色。向神经元的转化是有效的。

[0090] 图53显示用于控制哺乳动物基因调节的CRISPR/Cas9平台。A. 在由与之前是赋予靶位点特异性的可交换20bp前间区的Cas9形成复合体的恒定区组成的嵌合gRNA分子的存在下,基于Cas9的效应子结合基因组序列。B. 基于Cas9的合成转录因子通过干扰RNA聚合酶活性或通过在启动子内结合并阻断内源转录因子的结合位点阻抑靶基因的转录。C. 靶向调节元件例如增强子也可潜在地阻断多个远端基因的表达。

[0091] 图54显示使用CRISPR/dCas9-KRAB靶向HS2增强子。HS2区是远端调节>10kb下游的珠蛋白基因表达的强力增强子。设计一组单一gRNA以靶向沿增强子区的位点。

[0092] 图55显示靶向HS2增强子的单一gRNA实现珠蛋白基因的有效转录阻抑。A. 在慢病毒载体上递送dCas9和dCas9-KRAB阻抑蛋白。使单一gRNA瞬时转染用于筛选。与未接受gRNA处理的对照细胞相比,当在转染后3天通过定量RT-PCR测定时,表达dCas9-KRAB的K562实现B. γ-珠蛋白、C. ε-珠蛋白和D. β-珠蛋白基因的高达80%阻抑。E. 与β-肌动蛋白对照相比,在第3天表达dCas9或dCas9-KRAB并用Cr4或Cr8处理的细胞中的蛋白质表达显示γ-珠蛋白表达的适度阻抑。

[0093] 图56显示以不同剂量的gRNA质粒递送至A. 不用慢病毒、用B. dCas9慢病毒或C.

dCas9-KRAB慢病毒处理的细胞中的珠蛋白基因座基因的表达。增加所递送的Cr4 gRNA质粒的剂量在dCas9-KRAB处理细胞中提高阻抑,表明了dCas9-KRAB效应物和打靶gRNA两者在实现阻抑中的作用。

[0094] 图57显示用dCas9-KRAB稳定递送单一gRNA使珠蛋白基因的表达沉默。A. dCas9和dCas9-KRAB阻抑蛋白与单一gRNA一起在慢病毒载体上共表达。当在转导后7天通过定量RT-PCR测定时,与未接受慢病毒处理的对照细胞相比,表达dCas9-KRAB的K562实现B. γ-珠蛋白、C. ε-珠蛋白和D. β-珠蛋白基因的高达95%阻抑。

[0095] 图58显示分离p300 HAT“核心”用于仅通过dCas9融合的组蛋白的靶定外遗传修饰。

[0096] 图59显示酿脓链球菌(*S. pyogenes*) dCas9-VP64融合(顶部)和dCas9-p300核心融合(底部)的简图。在靶基因座上用箭头显示前间区序列邻近基序(PAM),用带阴影线的箭头显示合成指导RNA (gRNA)。

[0097] 图60A-60C显示在人293T细胞培养系中在没有任何融合的效应子结构域的情况下相对于dCas9-VP64和dCas9,使用dCas9-p300的激活功效的3个人基因座的代表性数据。

[0098] 图61A-61C显示dCas9构建体的氨基酸序列。图61A中显示了所有图61A-61C的图例。

[0099] 图62显示HAT-dCase9-p300融合蛋白未能激活基因表达。

[0100] 图63显示gRNA还与dCas9-p300核心协同起作用。

[0101] 图64显示TALEN介导来源于在肌养蛋白基因中携带不同缺失的人DMD患者的骨骼成肌细胞系中肌养蛋白的Dp427m骨骼肌同种型的5’UTR上的微小肌养蛋白(minidystrophin)的整合。将DMD患者细胞用编码在5’UTR基因座上有活性的TALEN对的构建体和携带微小肌养蛋白基因的供体模板电穿孔。(a)显示微小肌养蛋白如何整合进入5’UTR的示意图。(b)分离潮霉素抗性克隆细胞系,并使用(a)中所示引物,针对在5’UTR上的成功位点特异性整合通过PCR进行筛选。星号表明所选的用于(c)中的进一步分析的克隆。(c)使具有检出的整合事件的克隆分离的DMD成肌细胞分化6天,并评价与微小肌养蛋白C端融合的HA标签的表达。

发明详述

[0102] 如本文所述,发现某些方法和基于工程改造的CRISPR/CRISPR相关(Cas) 9的系统组合物可用于改变基因的表达、基因组工程改造和纠正或降低涉及遗传疾病的基因突变的作用。基于CRISPR/Cas9的系统包括Cas9蛋白和至少一种指导RNA,其提供系统的DNA打靶特异性。具体地说,本公开内容描述了使基于CRISPR/Cas9的系统的DNA序列打靶功能与其它活性组合的Cas9融合蛋白,因此允许改变基因表达和/或外遗传状态。该系统还可用于基因组工程改造和纠正或降低基因突变的作用。

[0103] 本公开内容还提供用于递送基于CRISPR/CRISPR相关(Cas) 9的系统和多种gRNA以靶向一个或多个内源基因的某些组合物和方法。靶向单一启动子的多种sgRNA的共转染允许协同激活,然而,由于拷贝数的差异所致,多个质粒的共转染导致各细胞中可变的表达水平。另外,在转染后的基因激活由于质粒DNA随时间推移稀释所致是瞬时的。此外,许多细胞类型不易转染,并且瞬时基因表达可能不足以引起治疗作用。为了克服这些限制,开发了

单一慢病毒系统以表达Cas9和来自独立启动子的多达4种sgRNA。公开了从单一慢病毒载体表达Cas9或dCas9融合蛋白和多达4种gRNA的平台。除自独立启动子表达的1、2、3或4种gRNA以外,慢病毒载体还表达组成型或诱导型Cas9或dCas9-VP64。该系统使得能够控制基于CRISPR/Cas9的基因调节的幅度和时机两者。此外,慢病毒平台提供有效和持续水平的基因表达,这可促进CRISPR/Cas9系统在原代细胞中的治疗应用。最后,该系统可同时用于编辑多个基因,例如几个癌基因的同时敲除。

[0104] 本公开内容还提供使用修饰的腺伴随病毒(AAV)载体将位点特异性核酸酶递送至骨骼肌和心肌的某些组合物和方法。可被工程改造的位点特异性核酸酶可用于改变基因表达、基因组工程改造、纠正或降低涉及遗传疾病的基因突变的作用或操作参与影响骨骼肌或心肌或肌肉再生的其它病况的基因。工程改造的位点特异性核酸酶可包括锌指核酸酶(ZFN)、TAL效应核酸酶(TALEN)和/或用于基因组编辑的CRISPR/Cas9系统。如本文所述,采用该独特的递送系统,在体内成功地编辑骨骼肌组织中的基因。所公开的发明提供改写人基因组用于治疗应用的方法和用于基础科学应用的目标模型类。

[0105] 基因编辑十分依赖于细胞循环和在组织与组织间变化的复杂的DNA修复途径。骨骼肌是极复杂的环境,由每个细胞超过100核的大量肌纤维组成。几十年来,基因疗法和生物制剂总的来讲因体内递送障碍而受限制。这些挑战包括体内载体的稳定性、靶向正确组织、得到足够的基因表达和有活性的基因产物,并避免可能压倒活性的毒性,这在用基因编辑工具时是常见的。在其它情况下,其它递送载体(例如质粒DNA的直接注射)在骨骼肌和心肌中起作用以表达基因,但对于获得可检测水平的基因组编辑,用这些位点特异性核酸酶无法起良好作用。

[0106] 虽然许多基因序列在AAV载体中不稳定并因此无法递送,但这些位点特异性核酸酶在AAV载体中出奇的稳定。当递送和表达这些位点特异性核酸酶时,它们在骨骼肌组织中保持活性。位点特异性核酸酶的蛋白质稳定性和活性是高度组织类型依赖性和细胞类型依赖性的。这些有活性且稳定的核酸酶能够在骨骼肌的复杂环境中修饰基因序列。本发明描述了将得力有效的并促进成功的基因组修饰的这类治疗剂的活性形式递送至骨骼肌或心肌的方法。

[0107] 本公开内容还提供某些融合外遗传效应分子,一种dCas9-p300融合蛋白,与dCas9-VP64融合相比,它提供用于合成转录调节的稳健和有效的更广泛适用的工具。与在所测的所有基因座上的dCas9-VP64融合蛋白相比,激活的靶基因至更大得多的程度。另外,p300对人基因组内的增强子具有固有的内源活性。dCas9-p300融合蛋白能够激活内源靶基因启动子和增强子区。

[0108] dCas9-p300融合蛋白可用于人组织培养细胞系以激活基因表达。该融合蛋白可用于以准确性和可预测性指导人细胞内靶基因座的外遗传状态以控制分化,调制细胞调节,并应用新发明的有效疗法。现行技术限于激活的强度和外遗传调节的程度及可持续性,可通过利用该新的融合蛋白消除的障碍。

[0109] 本节和本文公开的整个公开内容使用的章节标题仅仅用于组织目的,并无意限制。

[0110] 1.定义

[0111] 除非另有定义,否则本文所用的全部技术和科学术语具有本领域普通技术人员通

常理解的相同含义。在有冲突的情况下,以本文件(包括定义)为主。下面描述了优选的方法和材料,但是类似于本文所述的方法和材料可用于实施或测试本发明。本文提及的所有出版物、专利申请、专利和其它参考文献通过引用以其整体结合到本文中。本文公开的材料、方法和实例只是说明性的,并无意是限制性的。

[0112] 本文所用术语“包含”、“包括”、“具有”、“拥有”、“可以”、“含有”及其变体欲指不排除其它行动或结构的可能性的开放式过渡短语、术语或措词。单数形式“a”、“an”和“the”包括复数指代物,除非文中另有规定。不论是否明确陈述,本公开内容还包括“包含”本文提供的实施方案或要素的其它实施方案、“由”和“基本由”本文提供的实施方案或要素“组成”。

[0113] 对于本文数字范围的引述,明确包括其间相同精确度的各个间插数字。例如,对于6-9的范围,除6和9以外还包括数字7和8,对于范围6.0-7.0,明确包括数字6.0、6.1、6.2、6.3、6.4、6.5、6.6、6.7、6.8、6.9和7.0。

[0114] 本文可互换使用的“腺伴随病毒”或“AAV”是指属于感染人和一些其它灵长类动物物种的细小病毒科(Parvoviridae)依赖病毒属(Dependovirus)的微小病毒。目前已知AAV不引起疾病,但该病毒引起极轻微的免疫应答。

[0115] 本文所用“结合区”是指被核酸酶识别和结合的核酸酶靶区内的区域。

[0116] 本文可互换使用的“心肌”或“心脏肌肉”意指存在于壁心脏和心脏组织基础即心肌层的非自主横纹肌的类型。心肌由心肌细胞(cardiomyocyte/myocardiocyte)构成。心肌细胞显示类似于骨骼肌细胞中的条纹,但只含有唯一一个核,与多核骨骼细胞不同。

[0117] 本文所用“心肌病况”是指与心肌有关的疾病,例如心肌病、心力衰竭、心律失常和炎性心脏病。

[0118] 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)。编码序列还可包括与调节元件(包括启动子和聚腺苷酸化信号)有效连接的起始和终止信号,所述调节元件能够指导在给予其核酸的个体或哺乳动物的细胞中表达。编码序列可以是密码子优化。

[0119] 本文所用“互补序列”或“互补”意指可意味着在核酸分子的核苷酸或核苷酸类似物之间的Watson-Crick (例如A-T/U和C-G)或Hoogsteen碱基配对的核酸。“互补性”是指两个核酸序列之间共有的性质,使得当彼此反平行比对时,各位置上的核苷酸碱基都将互补。

[0120] 本文所用“纠正”、“基因组编辑”和“恢复”是指改变编码截短的蛋白质或完全不编码蛋白质的突变基因,使得获得全长功能性或部分全长功能性蛋白质表达。纠正或恢复突变基因可包括以修复机制(例如同源指导修复(HDR))置换具有突变的基因的区域或用不具有突变的基因拷贝置换整个突变基因。纠正或恢复突变基因还可包括通过在基因中产生双链断裂,然后使用非同源末端连接(NHEJ)修复,来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。在可恢复合适的读框并排除提前终止密码子的修复期间,NHEJ可加入或缺失至少一个碱基对。纠正或恢复突变基因还可包括破坏异常剪接受体位点或剪接供体序列。纠正或恢复突变基因还可包括通过两种核酸酶对相同DNA链的同时作用缺失非必需基因区段,以通过除去两个核酸酶靶位点间的DNA并修复被NHEJ断裂的DNA来恢复合适的读框。

[0121] 本文可互换使用的“供体DNA”、“供体模板”和“修复模板”是指包括目标基因的至少一部分的双链DNA片段或分子。供体DNA可编码完全功能性蛋白质或部分功能性蛋白质。

[0122] 本文可互换使用的“杜兴型肌营养不良”或“DMD”是指导致肌肉变性和最终死亡的隐性致命性X链锁病症。DMD是一种常见的遗传性单基因病3500名男性中发生1例。DMD是引起肌养蛋白基因的无义突变或移码突变的遗传突变或自发突变的结果。引起DMD的大多数肌养蛋白突变是破坏读框和引起肌养蛋白基因的提前翻译终止的外显子缺失。DMD患者通常在儿童时代丧失身体上自我支持的能力,在十几岁时变得日益衰弱,在其二十岁左右死亡。

[0123] 本文所用“肌养蛋白”是指杆状胞质蛋白质,其是肌纤维的细胞骨架与遍及细胞膜的周围胞外基质连接的蛋白质复合体的一部分。肌养蛋白提供细胞膜的营养不良蛋白聚糖复合体的结构稳定性,所述复合体负责调节肌细胞完整性和功能。本文可互换使用的肌养蛋白基因或“DMD基因”是位于基因座Xp21的2.2兆碱基。初级转录测量约2,400 kb,其成熟mRNA为约14 kb。79个外显子编码超过3500个氨基酸的蛋白质。

[0124] 本文所用“外显子51”是指肌养蛋白基因的第51个外显子。外显子51通常接近DMD患者的读框破坏性缺失,并且在临床试验中被靶中用于基于寡核苷酸的外显子跳跃。用于外显子51跳跃化合物eteplirsen的临床试验最近报告了在48周内的显著功能益处,具有与基线相比平均47%的肌养蛋白阳性纤维。外显子51中的突变通过基于NHEJ的基因组编辑,理想地适于永久纠正。

[0125] 本文可互换使用的“移码”或“移码突变”是指其中添加或缺失一个或多个核苷酸引起mRNA中密码子的读框移动的基因突变类型。读框的移动可导致蛋白质翻译上氨基酸序列的改变,例如错义突变或提前终止密码子。

[0126] 本文所用“功能性”和“完全功能性”描述具有生物活性的蛋白质。“功能基因”是指转录成mRNA的基因,所述mRNA翻译成功能性蛋白质。

[0127] 本文所用“融合蛋白”是指通过连接最初编码各自的蛋白质的两个或更多个基因所产生的嵌合蛋白。融合基因的翻译导致具有来源于原始蛋白质各自的功能性质的单一多肽。

[0128] 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。编码序列包括与调节元件(包括启动子和聚腺苷酸化信号)有效连接的起始和终止信号,所述调节元件能够在给予其核酸分子的个体的细胞中指导表达。本文所用术语“可表达形式”是指含有与编码蛋白质的编码序列有效连接的必需调节元件使得当存在于个体的细胞中时编码序列可表达的基因构建体。

[0129] 本文所用“遗传疾病”是指由基因组中的一个或多个异常区部分或完全、直接或间接引起的疾病,尤其自出生时就存在的病况。异常可以是突变、插入或缺失。异常可影响基因的编码序列或其调节序列。遗传疾病可以是但不限于DMD、血友病、囊性纤维化、亨廷顿舞蹈病、家族性高胆固醇血症(LDL受体缺陷)、肝胚细胞瘤、肝豆状核变性(Wilson’s disease)、先天性肝性卟啉病、遗传性肝代谢病症、莱-尼综合征(Lesch Nyhan syndrome)、镰状细胞性贫血、地中海贫血、着色性干皮病、Fanconi贫血、色素性视网膜炎、共济失调毛细血管扩张、布卢姆氏综合征(成视网膜细胞瘤)、成视网膜细胞瘤和泰-萨病(Tay-Sachs disease)。

[0130] 本文可互换使用的“同源指导修复”或“HDR”是指大部分在细胞周期的G2和S期,当DNA的同源段存在于胞核中时,在细胞中修复双链DNA损害的机制。HDR使用供体DNA模板指

导修复,并可用于产生基因组的特定序列变化,包括整个基因的靶定添加。如果供体模板与位点特异性核酸酶(例如基于CRISPR/Cas9的系统)一起提供,则细胞机器将通过同源重组修复断裂,在DNA切割存在时这被提高几个数量级。当同源DNA段存在时,则可发生非同源末端连接。

[0131] 本文所用"基因组编辑"是指改变基因。基因组编辑可包括纠正或恢复突变基因。基因组编辑可包括敲除基因,例如突变基因或正常基因。基因组编辑可通过改变目标基因而用于治疗疾病或增强肌肉修复。

[0132] 本文所用"相同的"或"同一性"在两个或更多个核酸或多肽序列的情况下意指序列在特定区域内具有相同残基的特定百分比。百分比可如下计算:通过最优地比对两个序列,比较特定区域内的2个序列,测定两个序列中出现的相同残基的位置数以得出匹配位置数,将匹配位置数除以特定区域中的总位置数,将结果乘以100得到序列同一性的百分比。在其中2个序列具有不同长度或比对产生一个或多个交错末端且比较的特定区域只包括单一序列的情况下,单一序列的残基包括在计算的分母中,但不在分子中。当比较DNA和RNA时,胸腺嘧啶(T)和尿嘧啶(U)可视为等同的。可手工或通过应用计算机序列算法例如BLAST或BLAST 2.0,进行同一性计算。

[0133] 本文可互换使用的"突变基因"或"突变的基因"是指进行了可检测突变的基因。突变基因经过改变,例如遗传物质的丧失、获得或交换,这影响基因的正常传递和表达。本文所用"被破坏的基因"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被破坏的基因产物被截短。

[0134] 本文所用"非同源末端连接(NHEJ)途径"是指在不需要同源模板的情况下通过直接连接断裂端修复DNA中的双链断裂的途径。通过NHEJ的不依赖模板的DNA再连接是在DNA断裂点随机引入微插入和微缺失(得失位(indel))的随机易错修复过程。该方法可被有意用来中断、缺失或改变靶定基因序列的读框。NHEJ通常使用称为微同源物(microhomologies)的短的同源DNA序列指导修复。这些微同源物常常存在于双链断裂末端上的单链突出端。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。

[0135] 本文所用"正常基因"是指未经过改变例如遗传物质的丢失、获得或交换的基因。正常基因进行正常基因传递和基因表达。

[0136] 本文所用"核酸酶介导的NHEJ"是指核酸酶(例如cas9)切断双链DNA后开始的NHEJ。

[0137] 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。单链的描述还定义了互补链的序列。因此,核酸还包括所述单链的互补链。核酸的许多变体可用于与指定核酸相同的目的。因此,核酸还包括基本相同的核酸及其互补序列。单链提供可在严格杂交条件下与靶序列杂交的探针。因此,核酸还包括在严格杂交条件下杂交的探针。

[0138] 核酸可以是单链或双链的,或可含有双链和单链序列两者的一部分。核酸可以是DNA (基因组和cDNA两者)、RNA或杂合体,其中核酸可含有脱氧核糖核苷酸和核糖核苷酸的组合,并且包括尿嘧啶、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在内的碱基的组合。可通过化学合成方法或重组方法获得核酸。

[0139] 本文所用“有效连接”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与所述启动子与它控制的基因(启动子来源于该基因)的距离大致相同。正如本领域已知,可调节该距离的变化而不丧失启动子功能。

[0140] 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质低但比非功能性蛋白质高的生物活性的蛋白质。

[0141] 本文可互换使用的“提前终止密码子”或“读框外终止密码子”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形式。

[0142] 本文所用“启动子”意指能够赋予、激活或提高细胞中核酸表达的合成或天然来源的分子。启动子可包含一个或多个特异性转录调节序列以进一步提供表达和/或改变所述序列的空间表达和/或时间表达。启动子还可包含远端增强子或阻抑蛋白元件,其可定位于距转录起始位点多达几千碱基对。启动子可来源于以下来源,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,或者在响应外部刺激例如生理应激、病原体、金属离子或诱导剂时,启动子可组成型或差异性地调节基因成分的表达。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、lac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晚期启动子和CMV IE启动子。

[0143] 本文可互换使用的“重复可变二残基”或“RVD”是指DNA识别基序内的相邻氨基酸残基对(亦称“RVD模件”),其包括TALE DNA结合结构域的33-35个氨基酸。RVD决定RVD模件的核苷酸特异性。RVD模件可以组合以产生RVD阵列。本文所用“RVD阵列长度”是指相当于被TALEN识别的TALEN靶区(即结合区)内核苷酸序列的长度的RVD模件的数目。

[0144] 本文所用“位点特异性核酸酶”是指能够特异性识别和切割DNA序列的酶。可对位点特异性核酸酶进行工程改造。工程改造的位点特异性核酸酶的实例包括锌指核酸酶(ZFN)、TAL效应核酸酶(TALEN)和基于CRISPR/Cas9的系统。

[0145] 本文所用“骨骼肌”是指横纹肌的类型,其在体细胞神经系统的控制下,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”的各成分组成,有时通称为“肌纤维”。肌细胞在称为肌生成的过程中由发育的成肌细胞(产生肌肉细胞的胚胎祖细胞的一个类型)融合形成。这些长的圆柱形多核细胞亦称肌纤维。

[0146] 本文所用“骨骼肌病况”是指与骨骼肌有关的病况,例如肌营养不良、老化、肌肉变性、伤口愈合和肌无力或萎缩。

[0147] 本文可互换使用的“间隔区”和“间隔区域”是指介于2个TALEN或ZFN的结合区之间但非其一部分的TALEN或ZFN靶区内的区域。

[0148] 本文可互换使用的“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,例如食蟹猴(cynomolgous)或猕猴、黑猩猩等)和人)。在一些实施方案中,受试者可以是人或非人。可对受试者或患者进行其它形式的治疗。

[0149] 本文所用“靶基因”是指编码已知或推定基因产物的任何核苷酸序列。靶基因可以是涉及遗传疾病的突变基因。

[0150] 本文所用“靶区”是指设计位点特异性核酸酶以与之结合并切割的靶基因的区域。

[0151] 本文所用“转录激活因子样效应物”或“TALE”是指识别并结合特定DNA序列的蛋白质结构。“TALE DNA结合结构域”是指包括一系列串联33-35个氨基酸重复序列的DNA结合结构域,亦称RVD模件,其各自特异性识别DNA的单一碱基对。RVD模件可以任何顺序排列以装配识别规定序列的阵列。

[0152] TALE DNA结合结构域的结合特异性由RVD阵列接着20个氨基酸的单一截短重复序列确定。TALE DNA结合结构域可具有12-27个RVD模件,其各自含有RVD并识别DNA的单一碱基对。已鉴定出识别4种可能的DNA核苷酸(A、T、C和G)每一个的特异性RVD。因为TALE DNA结合结构域是模块化的,识别4种不同的DNA核苷酸的重复序列可连接在一起以识别任何特定的DNA序列。然后,这些靶定的DNA结合结构域可与催化结构域结合以产生功能性酶,包括人工转录因子、甲基转移酶、整合酶、核酸酶和重组酶。

[0153] 本文可互换使用的“转录激活因子样效应核酸酶”或“TALEN”是指核酸酶的催化结构域(例如内切核酸酶*Fok*I)和可被定制DNA序列靶定的设计的TALE DNA结合结构域的工程改造融合蛋白。“TALEN单体”是指具有催化核酸酶结构域和设计的TALE DNA结合结构域的工程改造融合蛋白。可设计两个TALEN单体以靶向并切割TALEN靶区。

[0154] 本文所用“转基因”是指含有从一种生物体分离并导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。

[0155] 本文使用的有关核酸的“变体”意指(i)参考核苷酸序列的一部分或片段;(ii)参考核苷酸序列的互补序列或其部分;(iii)与参考核酸基本相同的核酸或其互补序列;或(iv)在严格条件下与参考核酸、其互补序列或与之基本相同的序列杂交的核酸。

[0156] 有关肽或多肽的“变体”是指以氨基酸的插入、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。变体还可指与具有保留至少一种生物活性的氨基酸序列的参考蛋白质基本相同的氨基酸序列的蛋白质。氨基酸的保守取代,即将氨基酸用不同的性质类似(例如亲水性、带电荷区的程度和分布)的氨基酸置换在本领域被视为通常涉及少量变化。正如本领域所了解的,在某种程度上,通过考虑氨基酸的亲水指数,可鉴定这些少量变化。Kyte等, *J. Mol. Biol.* 157:105-132 (1982)。氨基酸的亲水指数基于其疏水性和电荷的考量。本领域已知,相似亲水指数的氨基酸可被取代并仍保留蛋白质功能。一方面,取代亲水指数为±2的氨基酸。氨基酸的亲水性还可用于揭示可产生保留生物功能的蛋白质的取代。在肽的情况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。要了解,与该观察结果一致,与生物功能相容的氨基酸取代取决于通过疏水性、亲水性、电荷、大小和其它性质显示的氨基酸的相对相似性,特别是那些氨基酸的侧链。

[0157] 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RN载体。载体可以是自我复制的染色体外载体,优选是DNA质粒。例如,载体可编码包含SEQ ID NO: 1的氨基酸序列的iCas9-VP64融合蛋白或SEQ ID NO: 5-40、65-144、492-515、540-563和585-625的任一个的至少一种

gRNA核苷酸序列。或者,载体可编码Cas9和SEQ ID NO: 5-40、65-144、492-515、540-563和585-625的任一个的至少一种gRNA核苷酸序列。

[0158] 本文所用“锌指”是指识别并结合DNA序列的蛋白质结构。锌指结构域是人蛋白质组中最常见的DNA结合基序。单一锌指含有约30个氨基酸,该结构域通常通过每碱基对的单个氨基酸侧链的相互作用,经结合DNA的3个连续的碱基对而起作用。

[0159] 本文可互换使用的“锌指核酸酶”或“ZFN”是指包含与至少一种核酸酶或在完全装配时能够切割DNA的核酸酶的一部分有效连接的至少一个锌指DNA结合结构域的嵌合蛋白分子。

[0160] 除非本文另有定义,否则与本公开内容联用的科学和技术术语具有本领域普通技术人员通常理解的意义。例如,与本文所述细胞和组织培养、分子生物学、免疫学、微生物学、遗传学和蛋白质和核酸化学和杂交及其技术联用的任何术语是本领域众所周知和常用的术语。术语的含义和范围应是明确的;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词典或外来定义。另外,除非文中另有要求,否则单数形式的术语应包括复数形式,复数术语应包括单数。

[0161] 2. 用于基因组编辑的组合物

[0162] 本发明涉及用于基因组编辑、基因组改变或改变靶基因的基因表达的组合物。组合物可包括病毒载体和例如位点特异性核酸酶或CRISPR/Cas9-系统的融合蛋白与至少一种gRNA。

[0163] *a. 用于肌肉中基因组编辑的组合物*

[0164] 本发明涉及用于基因组编辑受试者骨骼肌或心肌中的靶基因的组合物。组合物包括修饰的AAV载体和编码位点特异性核酸酶的核苷酸序列。组合物将活性形式的位点特异性核酸酶递送至骨骼肌或心肌。组合物还可包含供体DNA或转基因。这些组合物可用于基因组编辑、基因组工程改造和纠正或降低突变在涉及遗传疾病和/或其它骨骼或心肌病况的基因中的作用。

[0165] 靶基因可参与细胞的分化或其中可能需要激活、阻抑或中断基因的其它过程,或可具有突变例如缺失、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,则可设计位点特异性核酸酶以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。还可以使用位点特异性核酸酶通过靶向剪接受体和供体来中断正常基因剪接以诱导提前终止密码子跳跃或恢复被中断的读框。位点特异性核酸酶可能介导或不介导脱靶改变成为基因组的蛋白质编码区。

[0166] 3. CRISPR系统

[0167] 本文可互换使用的“成簇有规间隔短回文重复序列”和“CRISPR”是指含有存在于约40%的已测序细菌和90%的已测序古细菌的基因组中的多个短的同向重复序列的基因座。CRISPR系统是微生物核酸酶系统,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介导的核酸切割特异性的非编码RNA元件的组合。将短的外来DNA区段(称为间隔区)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的3’端形成复合体,该蛋白质-RNA对通过sgRNA序列的5’端和预先确定的20 bp DNA序列(称为前间区)之间的互补碱基配

对识别其基因组靶。该复合体通过crRNA内编码的区域(即前间区)和病原体基因组内的前间区序列邻近基序(PAM)导向病原体DNA的同源基因座。同向重复序列内的非编码CRISPR阵列(CRISPR array)被转录并切割成含有各间隔区序列的短的crRNA,其将Cas核酸酶引导至靶位点(前间区)。通过简单交换表达的sgRNA的20 bp识别序列,Cas9核酸酶可被引向新的基因组靶标。使用CRISPR间隔区以类似于真核生物的RNAi的方式识别外源遗传元件并使之沉默。

[0168] 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,在4个相继步骤进行靶定DNA双链断裂以切割dsDNA。与需要多个不同的效应器起复合体作用的I型和III型效应器系统相比,II型效应器系统可在非传统情况下(例如真核细胞)起作用。II型效应器系统由长的前crRNA (pre-crRNA) (其自含有间隔区的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的间隔区的重复区杂交,因此通过内源RNase III开始dsRNA切割。该切割接着通过Cas9在各间隔区内的第二切割事件,产生保留与tracrRNA和Cas9缔合的成熟crRNA,形成Cas9:crRNA-tracrRNA复合体。

[0169] Cas9:crRNA-tracrRNA复合体解开DNA双链体,并搜寻匹配crRNA的序列以切割。靶标识别在检测靶DNA中的“前间区”序列和crRNA中的其余间隔区序列之间的互补性时发生。如果纠正前间区序列邻近基序(PAM)还存在于前间区的3’端,则Cas9介导靶DNA的切割。对于前间区打靶,所述序列必须紧接前间区序列邻近基序(PAM),一种被是DNA切割所必需的Cas9核酸酶识别的短序列。不同的II型系统具有不同的PAM要求。酿脓链球菌CRISPR系统可具有作为5’-NRG-3’的这种Cas9 (SpCas9)的PAM序列,其中R为A或G,并以该系统在人细胞中的特异性为特征。CRISPR/Cas9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组基因座的直接能力。例如,酿脓链球菌II型系统天然更喜欢使用“NGG”序列,其中“N”可为任何核苷酸,但还可接受其它PAM序列,例如工程改造系统中的“NAG” (Hsu等, *Nature Biotechnology* (2013) doi:10.1038/nbt.2647)。同样,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等, *Nature Methods* (2013) doi:10.1038/nmeth.2681)。

[0170] 4. 基于CRISPR/Cas9的系统

[0171] 显示了在人细胞中起作用的用于基因组工程改造的酿脓链球菌的II型效应器系统的一种工程改造形式。在该系统中,通过合成再造的“指导RNA” (“gRNA”,在本文也可互换用作嵌合单一指导RNA(“sgRNA”)),其是一般避免RNase III和crRNA加工的需要的crRNA-tracrRNA融合物,Cas9蛋白指向基因组靶位点(参见图53A)。本文提供基于CRISPR/Cas9的工程改造系统用于基因组编辑和治疗遗传疾病。可设计基于CRISPR/Cas9的工程改造系统以靶向任何基因,包括涉及遗传疾病、老化、组织再生或伤口愈合的基因。基于CRISPR/Cas9的系统可包括Cas9蛋白或Cas9融合蛋白和至少一种gRNA。Cas9融合蛋白可包括例如具有不同活性对Cas9是内源的结构域,例如反式激活域。

[0172] 靶基因可参与细胞的分化或其中可能需要激活基因或可具有突变例如移码突变或无义突变的其它过程。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,则可设计基于CRISPR/Cas9的系统以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的系统也可用来通过靶向剪接受体和供体中断正常基因剪接以诱导提前终止密码子跳跃或恢复被中断的读框。基于CRISPR/Cas9的系统可能介导或不介导脱靶改变成为基因组的蛋白质编码区。

[0173] *a. Cas9*

[0174] 基于CRISPR/Cas9的系统可包括Cas9蛋白或Cas9融合蛋白。Cas9蛋白是切割核酸的内切核酸酶,由CRISPR基因座编码,并参与II型CRISPR系统。Cas9蛋白可来自任何细菌或古细菌物种,例如酿脓链球菌。可使Cas9蛋白突变,使得核酸酶活性失活。没有内切核酸酶活性的来自酿脓链球菌的失活Cas9蛋白(iCas9,亦称"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。本文所用"iCas9"和"dCas9"两者是指具有氨基酸取代D10A和H840A且其核酸酶活性失活的Cas9蛋白。例如,基于CRISPR/Cas9的系统可包括SEQ ID NO: 459或461的Cas9。

[0175] *b. Cas9融合蛋白*

[0176] 基于CRISPR/Cas9的系统可包括融合蛋白。融合蛋白可包含2个异源多肽结构域,其中第一多肽结构域包含Cas蛋白,第二多肽结构域具有活性,例如转录激活活性、转录阻遏活性、转录释放因子活性、组蛋白修饰活性、核酸酶活性、核酸缔合活性、甲基化酶活性或脱甲基酶活性。融合蛋白可包括与具有以下活性的第二多肽结构域融合的上述Cas9蛋白或突变的Cas9蛋白:例如转录激活活性、转录阻遏活性、转录释放因子活性、组蛋白修饰活性、核酸酶活性、核酸缔合活性、甲基化酶活性或脱甲基酶活性。

[0177] (1) 转录激活活性

[0178] 第二多肽结构域可具有转录激活活性,即反式激活域。例如,可通过gRNA的组合将iCas9和反式激活域的融合蛋白靶向哺乳动物启动子,实现内源哺乳动物基因例如人基因的基因表达。反式激活域可包括VP16蛋白、多种VP16蛋白,例如VP48结构域或VP64结构域或NFκB激活因子活性的p65结构域。例如,融合蛋白可为iCas9-VP64。

[0179] (2) 转录阻遏活性

[0180] 第二多肽结构域可具有转录阻遏活性。第二多肽结构域可具有Kruppel相关框活性,例如KRAB结构域、ERF阻抑蛋白结构域活性、Mxi1阻抑蛋白结构域活性、SID4X阻抑蛋白结构域活性、Mad-SID阻抑蛋白结构域活性或TATA框结合蛋白活性。例如,融合蛋白可为dCas9-KRAB。

[0181] (3) 转录释放因子活性

[0182] 第二多肽结构域可具有转录释放因子活性。第二多肽结构域可具有真核释放因子1 (ERF1)活性或真核释放因子3 (ERF3)活性。

[0183] (4) 组蛋白修饰活性

[0184] 第二多肽结构域可具有组蛋白修饰活性。第二多肽结构域可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为p300或CREB结合蛋白(CBP)蛋白质或其片段。例如,融合蛋白可为dCas9-p300。

[0185] (5) 核酸酶活性

[0186] 第二多肽结构域可具有不同于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切割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,虽然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。

[0187] (6) 核酸缔合活性

[0188] 第二多肽结构域可具有核酸缔合活性,或者核酸结合蛋白-DNA-结合结构域(DBD)是含有识别双链或单链DNA的至少一个基序的独立折叠的蛋白质结构域。DBD可识别特定的DNA序列(识别序列)或对DNA具有普遍的亲和力。核酸缔合区选自螺旋-转角-螺旋区、亮氨酸拉链区、翼状螺旋区、翼状螺旋-转角-螺旋区、螺旋-环-螺旋区、免疫球蛋白折叠、B3结构域、锌指、HMG框、Wor3结构域、TAL效应子DNA结合结构域。

[0189] (7) 甲基化酶活性

[0190] 第二多肽结构域可具有甲基化酶活性,其包括将甲基转移到DNA、RNA、蛋白质、小分子、胞嘧啶或腺嘌呤上。第二多肽结构域可包括DNA甲基转移酶。

[0191] (8) 脱甲基酶活性

[0192] 第二多肽结构域可具有脱甲基酶活性。第二多肽结构域可包括从核酸、蛋白质(特别是组蛋白)和其它分子中脱去甲基(CH3-)的酶。或者,第二多肽可在使DNA脱甲基的机制中将甲基转移到羟基甲基胞嘧啶中。第二多肽可催化该反应。例如,催化该反应的第二多肽可为Tet1。

[0193] *c. gRNA*

[0194] gRNA提供基于CRISPR/Cas9的系统的打靶。gRNA是两种非编码RNA即crRNA和tracrRNA的融合物。sgRNA可通过交换20 bp前间区的序列(这通过与所需DNA靶的互补碱基配对赋予打靶特异性),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:tracrRNA双链体。可包括例如42个核苷酸crRNA和75个核苷酸tracrRNA的这种双链体,起指导Cas9切割靶核酸的作用。本文可互换使用的“靶区”、“靶序列”或“前间区”是指基于CRISPR/Cas9的系统所靶向的靶基因的区域。基于CRISPR/Cas9的系统可包括至少一种gRNA,其中gRNA靶向不同的DNA序列。靶DNA序列可以是重叠的。靶序列或前间区后面是前间区3’端的PAM序列。不同的II型系统具有不同的PAM要求。例如,酿脓链球菌II型系统使用“NGG”序列,其中“N”可为任何核苷酸。

[0195] 给予细胞的gRNA的数目可为至少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。给予细胞的gRNA的数目可介于至少1种gRNA与至少50种不同的gRNA、至少1种gRNA与至少45种不同的gRNA、至少1种gRNA与至少40种不同的gRNA、至少1种gRNA与至少35种不同的gRNA、至少1种gRNA与至少30种不同的gRNA、至少1种gRNA与至少25种不同的gRNA、至少1种gRNA与至少20种不同的gRNA、至少1种gRNA与至少16种不同的gRNA、至少1种gRNA与至少12种不同的gRNA、至少1种gRNA与至少8种不同的gRNA、至少1种gRNA与至少4种不同的gRNA、至少4种gRNA与至少50种不同的gRNA、至少4种不同的gRNA与至少45种不同的gRNA、至少4种不同的gRNA与至少40种不同的

gRNA、至少4种不同的gRNA与至少35种不同的gRNA、至少4种不同的gRNA与至少30种不同的gRNA、至少4种不同的gRNA与至少25种不同的gRNA、至少4种不同的gRNA与至少20种不同的gRNA、至少4种不同的gRNA与至少16种不同的gRNA、至少4种不同的gRNA与至少12种不同的gRNA、至少4种不同的gRNA与至少8种不同的gRNA、至少8种不同的gRNA与至少50种不同的gRNA、至少8种不同的gRNA与至少45种不同的gRNA、至少8种不同的gRNA与至少40种不同的gRNA、至少8种不同的gRNA与至少35种不同的gRNA、8种不同的gRNA与至少30种不同的gRNA、至少8种不同的gRNA与至少25种不同的gRNA、8种不同的gRNA与至少20种不同的gRNA、至少8种不同的gRNA与至少16种不同的gRNA或8种不同的gRNA与至少12种不同的gRNA之间。

[0196] gRNA可包含靶DNA序列的互补多核苷酸序列接着PAM序列。gRNA可在互补多核苷酸序列5'端包含"G"。gRNA可包含靶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接着PAM序列。PAM序列可为"NGG",其中"N"可为任何核苷酸。gRNA可靶向启动子区、增强子区或靶基因的转录区的至少一个。gRNA可包括SEQ ID NO: 5-40、65-144、492-515、540-563、585-625、462 (图40)、464 (图41)和465 (图41)的至少一个的核酸序列。

[0197] gRNA可靶向任何核酸序列。核酸序列靶可以是DNA。DNA可以是任何基因。例如,gRNA可靶向基因,例如*BRN2*、*MYT1L*、*ASCL1*、*NANOG*、*VEGFA*、*TERT*、*IL1B*、*IL1R2*、*IL1RN*、*HBG1*、*HBG2*、*MYOD1*、*OCT4*和*DMD*。

[0198] (1) 肌养蛋白

[0199] 肌养蛋白是一种杆状胞质蛋白质,其是肌纤维的细胞骨架与遍及细胞膜的周围胞外基质连接的蛋白质复合体的一部分。肌养蛋白提供细胞膜的营养不良蛋白聚糖复合体的结构稳定性。肌养蛋白基因是位于基因座Xp21的2.2兆碱基。初级转录测量约2,400 kb,其成熟mRNA为约14 kb。79个外显子编码超过3500个氨基酸的蛋白质。正常骨骼肌组织只含少量的肌养蛋白,但其存在的异常表达导致发生严重和无法治愈的症状。肌养蛋白基因中的一些突变导致在受累患者中产生有缺陷的肌养蛋白和严重的营养不良表型。肌养蛋白基因中的一些突变在受累患者中导致部分功能性肌养蛋白和更轻微得多的营养不良表型。

[0200] DMD是引起肌养蛋白基因中的无义突变或移码突变的遗传突变或自发突变的结果。对于DMD,天然存在的突变及其后果是相对充分了解的。已知rod结构域内所含的外显子45-55区中出现的符合读框的缺失可产生高度功能性肌养蛋白,而且许多携带者是无症状的或显示轻微症状。此外,理论上,可通过靶向肌养蛋白基因的这个区域的外显子来治疗超过60%的患者。已尝试通过在mRNA剪接期间跳过非必需外显子以产生内部缺失但有功能的肌养蛋白,以恢复DMD患者中的中断的肌养蛋白读框。内部肌养蛋白外显子的缺失保留合适的读框,但引起较不严重的贝克肌营养不良(Becker muscular dystrophy)。

[0201] (2) 靶向肌养蛋白的基于CRISPR/Cas9的系统

[0202] 本文公开了对肌养蛋白基因有特异性的基于CRISPR/Cas9的系统。基于CRISPR/Cas9的系统可包括Cas9和至少一种gRNA以靶向肌养蛋白基因。基于CRISPR/Cas9的系统可结合并识别靶区。可以选择紧接可能的读框外终止密码子上游的靶区,使得修复过程期间

的插入或缺失通过读框转换恢复肌养蛋白读框。靶区还可以是剪接受体位点或剪接供体位点,使得修复过程期间的插入或缺失通过剪接位点中断和外显子排除中断剪接并恢复肌养蛋白读框。靶区还可以是异常终止密码子,使得修复过程期间的插入或缺失通过消除或中断终止密码子来恢复肌养蛋白读框。

[0203] 可设计单一或多种sgRNA,通过靶向外显子45-55处的突变热点并引入外显子内的小的插入和缺失或大的一个或多个外显子的缺失,恢复肌养蛋白读框。在用Cas9和一种或多种sgRNA处理后,体外杜兴肌营养不良患者肌细胞中的肌养蛋白表达可恢复。在将遗传纠正的患者细胞移植入免疫缺陷型小鼠体内后,检出了人肌养蛋白。重要的是,CRISPR/Cas9系统独特的多重基因编辑能力使得能够通过通用编辑方法或患者特异性基因编辑方法有效地产生该突变热区的大的缺失,这可纠正高达62%的患者突变。

[0204] 基于CRISPR/Cas9的系统可使用不同序列和长度的gRNA。gRNA的实例可参见表6。基于CRISPR/Cas9的系统可靶向SEQ ID NO :65-144的核酸序列或其互补序列。gRNA可包括SEQ ID NO: 65-144的核苷酸序列或其互补序列。例如,对所公开的基于CRISPR/Cas9的系统进行工程改造以介导肌养蛋白基因的外显子51处的高度有效的基因编辑。这些基于CRISPR/Cas9的系统在DMD患者的细胞中恢复肌养蛋白表达。

[0205] (a) 外显子51和45-55

[0206] 外显子51通常与DMD中读框破坏性缺失邻接。通过外显子跳跃排除肌养蛋白转录物的外显子51可用于治疗所有DMD患者的约15%。这类DMD突变理论上适于通过基于NHEJ的基因组编辑和HDR的永久性纠正。开发了本文描述的基于CRISPR/Cas9的系统用于人肌养蛋白基因中外显子51靶定修饰。这些基于CRISPR/Cas9的系统被转染至人DMD细胞中,并介导有效的基因修饰和转换以纠正读框。蛋白质修复伴随读框修复,并在大群基于CRISPR/Cas9的系统处理细胞中检出。同样,肌养蛋白转录物的外显子45-55的排除可用于治疗所有DMD患者的约62%。

[0207] (3) AAV/CRISPR构建体

[0208] 可使用不同的构建体配置,将AAV用于递送CRISPR (参见图39)。例如,AAV可递送各个载体中的Cas9和gRNA表达盒。或者,如果使用来源于例如金黄色葡萄球菌(*Staphylococcus aureus*)或脑膜炎奈瑟球菌等菌种的小的Cas9蛋白,则在4.7 kb包装限内的单一AAV载体可结合Cas9和至多两个gRNA表达盒两者(参见图39)。

[0209] 5.基于多个CRISPR/Cas9的系统

[0210] 本公开内容涉及基于多个CRISPR/Cas9的系统(其包括基于CRISPR/CRISPR相关(Cas) 9的系统,例如Cas9或dCas9)和多种gRNA以靶向一个或多个内源基因。这个平台利用便利的Golden Gate克隆方法以将多达4个独立的sgRNA表达盒快速并入单一慢病毒载体中。各sgRNA有效地表达,并可在永生化和原代人细胞系中介导不同基因座上的多重基因编辑。已表明通过1-4种sgRNA的协同激活,稳定表达dCas9-VP64的细胞系中的瞬时转录激活是可调节的。此外,单一慢病毒载体可在永生化和原代人细胞中诱导持久长期的内源基因表达。该系统允许单一慢病毒载体的快速装配,这使在模型和原代细胞系中的多重基因编辑或激活成为可能。

[0211] 基于多个CRISPR/Cas9的系统提供转录激活和转录激活的可调节诱导的效能。除自独立启动子的1、2、3或4种sgRNA表达以外,通过Golden Gate装配容易产生的最终载体表

达组成型Cas9或dCas9-VP64。各启动子能够有效地表达指导类似水平的Cas9核酸酶活性的sgRNA。此外,表达Cas9和靶向独立基因座的4种sgRNA的单一载体的慢病毒递送导致所有4个基因座的同时多重基因编辑。使用有或没有sgRNA的Cas9的慢病毒递送,实现了在瞬时和不稳定的两种情况下2种内源基因的可调节转录激活。实现了原代人细胞中十分有效的长期基因激活。该系统因此是在人细胞中产生多重基因编辑和长期转录激活的有吸引力的有效方法。

[0212] 基于多个CRISPR/Cas9的系统允许有效的多重基因编辑以同时使多个基因失活。CRISPR/Cas9系统可通过共表达单一Cas9蛋白与两种或更多种sgRNA,使该系统特别适用于多重基因编辑或协同激活应用,来同时靶向多个不同的基因组基因座。CRISPR/Cas9系统通过简单的修饰表达的sgRNA分子,极大地加快完成分子靶向新位点的过程。单一慢病毒载体可与各种方法联合,通过化学或光遗传学调节,实现这些成分的诱导型控制,以促进在时间和空间两者上基因调节的动力学的研究。

[0213] 基于多个CRISPR/Cas9的系统可在转录上激活两个或更多个内源基因。基于多个CRISPR/Cas9的系统可在转录上阻抑两个或更多个内源基因。例如,至少两个内源基因、至少3个内源基因、至少4个内源基因、至少5个内源基因或至少10个内源基因可被基于多个CRISPR/Cas9的系统激活或阻抑。介于2个和15个之间的基因,介于2个和10个之间的基因,介于2个和5个之间的基因,介于5个和15个之间的基因或介于5个和10个之间的基因可被基于多个CRISPR/Cas9的系统激活或阻抑。

[0214] (1) 修饰的慢病毒载体

[0215] 基于多个CRISPR/Cas9的系统包括修饰的慢病毒载体。修饰的慢病毒载体包括编码融合蛋白的第一多核苷酸序列和编码至少一种sgRNA的第二多核苷酸序列。融合蛋白可以是上述基于CRISPR/Cas9的系统的融合蛋白。第一多核苷酸序列可与启动子有效连接。启动子可以是组成型启动子、诱导型启动子、阻抑型启动子或可调节启动子。

[0216] 第二多核苷酸序列编码至少1种sgRNA。例如,第二多核苷酸序列可编码至少1种sgRNA、至少2种sgRNA、至少3种sgRNA、至少4种sgRNA、至少5种sgRNA、至少6种sgRNA、至少7种sgRNA、至少8种sgRNA、至少9种sgRNA、至少10种sgRNA、至少11种sgRNA、至少12种sgRNA、至少13种sgRNA、至少14种sgRNA、至少15种sgRNA、至少16种sgRNA、至少17种sgRNA、至少18种sgRNA、至少19种sgRNA、至少20种sgRNA、至少25种sgRNA、至少30种sgRNA、至少35种sgRNA、至少40种sgRNA、至少45种sgRNA或至少50种sgRNA。第二多核苷酸序列可编码介于以下的sgRNA:1种sgRNA和50种sgRNA、1种sgRNA和45种sgRNA、1种sgRNA和40种sgRNA、1种sgRNA和35种sgRNA、1种sgRNA和30种sgRNA、1种sgRNA和25种不同的sgRNA、1种sgRNA和20种sgRNA、1种sgRNA和16种sgRNA、1种sgRNA和8种不同的sgRNA、4种不同的sgRNA和50种不同的sgRNA、4种不同的sgRNA和45种不同的sgRNA、4种不同的sgRNA和40种不同的sgRNA、4种不同的sgRNA和35种不同的sgRNA、4种不同的sgRNA和30种不同的sgRNA、4种不同的sgRNA和25种不同的sgRNA、4种不同的sgRNA和20种不同的sgRNA、4种不同的sgRNA和16种不同的sgRNA、4种不同的sgRNA和8种不同的sgRNA、8种不同的sgRNA和50种不同的sgRNA、8种不同的sgRNA和45种不同的sgRNA、8种不同的sgRNA和40种不同的sgRNA、8种不同的sgRNA和35种不同的sgRNA、8种不同的sgRNA和30种不同的sgRNA、8种不同的sgRNA和25种不同的sgRNA、8种不同的sgRNA和20种不同的sgRNA、8种不同的sgRNA和16种不同的sgRNA、16种不同的sgRNA和50

种不同的sgRNA、16种不同的sgRNA和45种不同的sgRNA、16种不同的sgRNA和40种不同的sgRNA、16种不同的sgRNA和35种不同的sgRNA、16种不同的sgRNA和30种不同的sgRNA、16种不同的sgRNA和25种不同的sgRNA或16种不同的sgRNA和20种不同的sgRNA。编码不同sgRNA的多核苷酸序列各自可与启动子有效连接。与不同的sgRNA有效连接的启动子可以是相同的启动子。与不同的sgRNA有效连接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、阻抑型启动子或可调节启动子。

[0217] 至少一种sgRNA可与靶基因或基因座结合。如果包括一种以上的sgRNA,则sgRNA各自与一个靶基因座内的不同靶区结合或sgRNA各自与不同基因座内的不同靶区结合。融合蛋白可包括Cas9蛋白或iCas9-VP64蛋白。融合蛋白可包括VP64结构域、p300结构域或KRAB结构域。

[0218] 6. 位点特异性核酸酶

[0219] 上述组合物包括编码结合并切割靶区的位点特异性核酸酶的核苷酸序列。可对位点特异性核酸酶进行工程改造。例如,工程改造的位点特异性核酸酶可以是基于CRISPR/Cas9的系统、ZFN或TALEN。位点特异性核酸酶可结合和切割骨骼肌或心肌中的细胞的基因组的基因或基因座。例如,基因或基因座可以是Rosa26基因座或肌养蛋白基因。

[0220] *a. 基于CRISPR/Cas9的系统*

[0221] 上述基于CRISPR/Cas9的系统可用于在靶定的基因组基因座处引入位点特异性双链断裂。

[0222] *b. 锌指核酸酶(ZFN)*

[0223] 位点特异性核酸酶可以是ZFN。单个锌指含有约30个氨基酸,该结构域通过每碱基对的单个氨基酸侧链的相互作用,经结合DNA的3个连续的碱基对起作用。锌指基序的模块结构允许几个结构域串联连接,允许识别并靶向多重的3个核苷酸的延长序列。这些靶定的DNA结合结构域可与核酸酶结构域(例如*FokI*)组合,以产生可在靶定的基因组基因座上引入位点特异性双链断裂的位点特异性核酸酶,称为“锌指核酸酶” (ZFN)。这种DNA切割刺激天然DNA修复机器,导致两种可能的修复途径之一,NHEJ和HDR。例如,ZFN可靶向Rosa26基因座(Perez-Pinera等,Nucleic Acids Research (2012) 40:3741-3752)或肌养蛋白基因。ZFN的实例见表1和图35-38。表1中,DNA识别螺旋用下划线表示,“Fok ELD-S”和“Fok KKR-S”是指与锌指蛋白DNA结合结构域融合的FokI核酸酶结构域。在图35-38中,靶位点中的靶DNA序列(即SEQ ID NO: 442、445、448、和453中)和ZFN氨基酸序列(即SEQ ID NO: 443、444、446、447、449-452、454和455)中的DNA识别螺旋分别用下划线表示。

[0224] 表1 鉴定的ZFN的完全氨基酸序列。

[0225]

ZFN B 左侧 Fok ELD-S 完全氨基酸序列(SEQ ID NO:438)

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGRLEPGEKPYKCPECGKSFSRKDALRGH
QRTHTGEKPYKCPECGKSFSHRTTLTNHQRTHTGEKPYKCPECGKSFSQRNALAGHQRTH
TGEKPYKCPECGKSFSHKNALQNHQRTHTGEKPYKCPECGKSFSDPGHLVRHQRTHTGEK
PYKCPECGKSFSTSGNLVRHQRTHTGAAARALVKSELEEKKSELRHKLKYVPHEYIELIE
IARNPTQDRILEMKVMEFFMKVYGYRGEHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSG
GYNLPIGQADEMERYVEENQTRDKHLNPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLT
RLNHITNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINF*

ZFN B 右侧 Fok KKR-S 完全氨基酸序列(SEQ ID NO:439)

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGRLEPGEKPYKCPECGKSFSQQRSLVGH
QRTHTGEKPYKCPECGKSFSDKKDLTRHQRTHTGEKPYKCPECGKSFSTSGHLVRHQRTH
TGEKPYKCPECGKSFSQRAHLERHQRTHTGEKPYKCPECGKSFSTSGSLVRHQRTHTGEK
PYKCPECGKSFSTSGNLVRHQRTHTGAAARALVKSELEEKKSELRHKLKYVPHEYIELIE
IARNPTQDRILEMKVMEFFMKVYGYRGEHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSG
GYNLPIGQADEMQRYVKENQTRNKHINPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLT
RLNRKTNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINF*

ZFN J 左侧 Fok KKR-S 完全氨基酸序列(SEQ ID NO:440)

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGRLEPGERPFQCRICMRNFSSKQALAVH
TRTHTGEKPFQCRICMRNFSQSTTLKRHLRTHTGEKPFQCRICMRNFSRSDHLSLHLKTH
LRGSQLVKSELEEKKSELRHKLKYVPHEYIELIEIARNPTQDRILEMKVMEFFMKVYGYR
GEHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMQRYVKENQTRNKHI
NPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNRKTNCNGAVLSVEELLIGGEMIK
AGTLTLEEVRRKFNNGEINF*

ZFN J 右侧 Fok ELD-S 完全氨基酸序列(SEQ ID NO:441)

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGRLEPGERPFQCRICMRNFSRRAHLQNH
TRTHTGEKPFQCRICMRNFSQSTTLKRHLRTHTGEKPFQCRICMRNFSDGGHLTRHLKTH
LRGSQLVKSELEEKKSELRHKLKYVPHEYIELIEIARNPTQDRILEMKVMEFFMKVYGYR
GEHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMERYVEENQTRDKHL
NPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNHITNCNGAVLSVEELLIGGEMIK
AGTLTLEEVRRKFNNGEINF*

[0226] *c. TAL效应核酸酶(TALEN)*

[0227] TALEN可用于将位点特异性双链断裂引入靶定的基因组基因座上。当2个独立的TALEN与附近的DNA序列结合,从而允许FokI二聚化并切割靶DNA时,产生位点特异性双链断裂。由于其成功有效的遗传修饰的高速率所致,TALEN具有高级基因组编辑。这种DNA切割可刺激天然DNA修复机器,导致两种可能的修复途径之一:同源指导修复(HDR)或非同源末端连接(NHEJ)途径。可设计TALEN以靶向涉及遗传疾病的任何基因。

[0228] TALEN可包括核酸酶和结合TALEN靶区中的靶基因的TALE DNA结合结构域。靶基因可具有突变例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的突变,则可设计TALEN以识别并结合提前终止密码子上游或下游的核苷酸序列。“TALEN靶区”包括2个

TALEN的结合区和出现在结合区之间的间隔区。2个TALEN结合TALEN靶区内的不同结合区，之后TALEN靶区被切割。TALEN的实例描述于国际专利申请号PCT/US2013/038536,其通过引用以其整体结合到本文中。

[0229] 7．转录激活因子

[0230] 上述组合物,包括编码激活基因的转录激活因子的核苷酸序列。可对转录激活因子进行工程改造。例如,工程改造的转录激活因子可以是基于CRISPR/Cas9的系统、锌指融合蛋白或TALE融合蛋白。

[0231] *a．基于CRISPR/Cas9的系统*

[0232] 上述基于CRISPR/Cas9的系统可用来与RNA一起激活靶基因的转录。基于CRISPR/Cas9的系统可包括上述融合蛋白,其中第二多肽结构域具有转录激活活性或组蛋白修饰活性。例如,第二多肽结构域可包括VP64或p300。

[0233] *b．锌指融合蛋白*

[0234] 转录激活因子可以是锌指融合蛋白。靶定上述DNA结合结构域的锌指可与具有转录激活活性或组蛋白修饰活性的结构域组合。例如,结构域可包括VP64或p300。

[0235] *c．TALE融合蛋白*

[0236] TALE融合蛋白可用来激活靶基因的转录。TALE融合蛋白可包括TALE DNA结合结构域和具有转录激活活性或组蛋白修饰活性的结构域。例如,结构域可包括VP64或p300。

[0237] 8．组合物

[0238] 本发明涉及用于改变细胞或受试者中的基因表达和工程改造或改变基因组DNA的组合物。组合物还可包括病毒递送系统。

[0239] *a．用于肌肉中的基因组编辑的组合物*

[0240] 本发明涉及在受试者骨骼肌或心肌中进行靶基因的基因组编辑的组合物。组合物包括修饰的AAV载体和编码位点特异性核酸酶的核苷酸序列。组合物将活性形式的位点特异性核酸酶递送至骨骼肌或心肌。组合物还可包含供体DNA或转基因。这些组合物可用于基因组编辑、基因组工程改造和纠正或降低突变在涉及遗传疾病和/或其它骨骼或心肌病况的基因中的作用。

[0241] 靶基因可参与细胞的分化或其中可能需要激活、阻抑或中断基因的其它过程,或可具有突变例如缺失、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,则可设计位点特异性核酸酶以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。还可以使用位点特异性核酸酶通过靶向剪接受体和供体来中断正常基因剪接以诱导提前终止密码子跳跃或恢复被中断的读框。位点特异性核酸酶可能介导或不介导脱靶改变成为基因组的蛋白质编码区。

[0242] *b．腺伴随病毒载体*

[0243] 上述组合物包括修饰的腺伴随病毒(AAV)载体。修饰的AAV载体可具有改进的心肌和骨骼肌组织嗜性。修饰的AAV载体将能够在哺乳动物细胞中递送和表达位点特异性核酸酶。例如,修饰的AAV载体可以是AAV-SASTG载体(Piacentino等 (2012) Human Gene Therapy 23:635-646)。修饰的AAV载体在体内可将核酸酶递送至骨骼和心肌。修饰的AAV载体可基于几种衣壳类型的一种或多种,包括AAV1、AAV2、AAV5、AAV6、AAV8和AAV9。修饰的AAV

载体可基于通过全身和局部递送有效转导骨骼肌或心肌的具有备选的嗜肌性AAV衣壳的AAV2准型,例如AAV2/1、AAV2/6、AAV2/7、AAV2/8、AAV2/9、AAV2.5和AAV/SASTG载体(Seto等,Current Gene Therapy (2012) 12:139-151)。

[0244] *c. 基于CRISPR/Cas9的系统*

[0245] 本公开内容还提供DNA打靶系统或至少一个上述基于CRISPR/Cas9的系统的组合物。这些组合物可用于基因组编辑、基因组工程改造和纠正或降低涉及遗传疾病的基因突变的作用。组合物包括基于CRISPR/Cas9的系统,其包括上述Cas9蛋白或Cas9融合蛋白。基于CRISPR/Cas9的系统还可包括至少一种上述gRNA。

[0246] *d. 基于多个CRISPR/Cas9的系统*

[0247] 本公开内容还提供上述基于多个CRISPR/Cas9的系统。这些组合物可用于基因组编辑、基因组工程改造和纠正或降低涉及遗传疾病的基因突变的作用。这些组合物可用来靶向一个以上的基因。组合物包括修饰的慢病毒载体,其包含包括上述Cas9蛋白或Cas9融合蛋白和一种以上的上述gRNA的基于CRISPR/Cas9的系统。

[0248] 9. 使用方法

[0249] 组合物的潜在应用在科学和生物工程学的多个领域中是多种多样的。所公开的组合物可用来修复引起疾病的遗传突变。所公开的组合物可用来中断基因使得基因中断导致肌肉再生或肌肉强度增加,或肌肉老化减弱。所公开的组合物可用来引入治疗性基因从骨骼肌或心肌起在全身表达,例如凝血因子或单克隆抗体。所公开的组合物可用来调节哺乳动物基因表达。所公开的组合物可用来转分化或诱导细胞分化或纠正细胞中的突变基因。提供了激活与细胞和基因疗法、遗传重编程和再生医学有关的基因的实例。RNA指导的转录激活因子可用来重编程细胞谱系规格。对于遗传重编程、转分化和/或诱导分化,激活编码细胞命运的关键调节因子的内源基因,而不是迫使这些因子过量表达,可能导致更快速有效的稳定的特异性方法。

[0250] 10. 肌肉中基因组编辑的方法

[0251] 本公开内容涉及在受试者骨骼肌或心肌中进行基因组编辑的方法。所述方法包括如上所述将组合物给予受试者的骨骼肌或心肌用于在骨骼肌或心肌中进行基因组编辑。基因组编辑可包括纠正突变基因或插入转基因。纠正突变基因可包括缺失、重排或置换突变基因。纠正突变基因可包括核酸酶介导的NHEJ或HDR。

[0252] 11. 使用基于CRISPR/Cas9的系统的方法

[0253] 基于CRISPR/Cas9的系统的潜在应用在科学和生物工程学的多个领域中是多种多样的。所公开的基于CRISPR/Cas9的系统可用来调节哺乳动物基因表达。所公开的基于CRISPR/Cas9的系统可用来转分化或诱导细胞分化或纠正细胞中的突变基因。提供了激活与细胞和基因疗法、遗传重编程和再生医学有关的基因的实例。RNA指导的转录激活因子可用来重编程细胞谱系规格。虽然在这些实验中重编程是不完全的和无效率的,但是仍有可对该方法加以改进的许多方式,包括iCas9-VP64/gRNA组合的重复转染、这些因子的稳定表达和靶向多个基因,除Ascl1用于转分化成为神经元表型以外例如Brn2和Myt1l。对于细胞的遗传重编程和转分化或诱导分化,激活编码细胞命运的关键调节因子的内源基因,而不是迫使这些因子过量表达,可能导致更快速有效的稳定的特异性方法。最后,与其它结构域的Cas9融合物,包括阻抑和外遗传修饰性结构域,可提供RNA指导转录调节因子的较大多样

性以补充用于哺乳动物细胞工程改造的其它基于RNA的工具。

[0254] *a. 激活基因表达的方法*

[0255] 本公开内容提供根据使用上述基于CRISPR/Cas9的系统通过RNA靶向启动子的转录激活因子,激活内源基因(例如哺乳动物基因)表达的机制。这根本不同于基于工程改造序列特异性DNA结合蛋白的前述方法,并可提供靶定基因调节的机会。因为gRNA表达质粒的产生只包括合成2个短的定制寡核苷酸和1个克隆步骤,所以可能快速和节约地产生许多新的基因激活因子。gRNA还可在体外转录后被直接转染至细胞中。显示了靶向单一启动子的多种gRNA,但同时靶向多个动子也将是可能的。用RNA而不是蛋白质识别基因组靶位点还可以克服靶向外遗传修饰的位点(例如甲基化DNA)的限制。

[0256] 与基于工程改造的DNA结合蛋白的现行方法形成对比,与转录激活结构域融合的Cas9可被指导RNA分子的组合靶定以诱导内源人基因的表达。这种用于靶定基因激活的简单且通用的方法克服了对工程改造的新的蛋白质的需要,并允许广泛合成基因调节。

[0257] 所述方法可包括将上述基于CRISPR/Cas9的系统、编码所述基于CRISPR/Cas9的系统的多核苷酸或载体或DNA打靶系统或至少一个基于CRISPR/Cas9的系统的组合物给予细胞或受试者。所述方法可包括给予基于CRISPR/Cas9的系统,例如给予含有转录激活结构域的Cas9融合蛋白或编码所述Cas9融合蛋白的核苷酸序列。Cas9融合蛋白可包括转录激活结构域例如VP16蛋白或转录辅激活蛋白例如p300蛋白。

[0258] (1)dCas9-VP16

[0259] Cas9融合蛋白可包括转录激活结构域例如VP16蛋白。转录激活结构域可含有至少1种VP16蛋白、至少2种VP16蛋白、至少3种VP16蛋白、至少4种VP16蛋白(即VP64激活因结构域)、至少5种VP16蛋白、至少6种VP16蛋白、至少6种VP16蛋白或至少10种VP16蛋白。Cas9蛋白可以是其中核酸酶活性失活的Cas9蛋白。例如,融合蛋白中的Cas9蛋白可以是iCas9(SEQ ID NO: 1的氨基酸36-1403),其包括D10A和H840A的氨基酸取代。Cas9融合蛋白可以是iCas9-VP64。

[0260] (2)dCas9-p300

[0261] Cas9融合蛋白可包括转录共激活结构域例如p300蛋白。p300蛋白(亦称EP300或E1A结合蛋白p300)由*EP300*基因编码,调节整个机体组织中许多基因的活性。p300蛋白在调节细胞生长和分裂、促进细胞成熟和呈特化功能(分化)和防止癌性肿瘤的生长中起作用。p300蛋白通过连接转录因子与在细胞核中执行转录的蛋白质复合体而激活转录。p300与转录因子的相互作用以下p300结构域的一个或多个的控制:核受体相互作用结构域(RID)、CREB和MYB相互作用结构域(KIX)、半胱氨酸/组氨酸区(TAZ1/CH1和TAZ2/CH3)和干扰素反应结合结构域(IBiD)。p300的最后4种结构域KIX、TAZ1、TAZ2和IBiD,各自与跨越转录因子p53的两个反式激活域9aaTAD的序列紧密结合。该蛋白质起通过染色质重建调节转录的组蛋白乙酰转移酶的作用,并且在细胞增殖和分化过程中是重要的。它通过与磷酸化CREB蛋白特异性结合介导cAMP-基因调节。

[0262] p300蛋白可激活母体抗十五褶同源物7 (Mothers against decapentaplegic homolog 7)、MAF、TSG101、过氧化物酶体增殖物激活受体α、NPAS2、PAX6、DDX5、MYBL2、母体抗十五褶同源物1、母体抗十五褶同源物2、淋巴样细胞增强子结合因子1 (Lymphoid enhancer-binding factor)、SNIP1、TRERF1、STAT3、EID1、RAR相关孤儿受体α、ELK1、HIF1A、

ING5、过氧化物酶体增殖物激活受体γ、SS18、TCF3、Zif268、雌激素受体α、GPS2、MyoD、YY1、ING4、PROX1、CITED1、HNF1A、MEF2C、MEF2D、MAML1、Twist转录因子、PTMA、IRF2、DTX1、Flap结构特异性内切核酸酶1、肌细胞特异性增强子因子2A、CDX2、BRCA1、HNRPU、STAT6、CITED2、RELA、TGS1、CEBPB、Mdm2、NCOA6、NFATC2、甲状腺激素受体α、BCL3、TFAP2A、PCNA、P53和TAL1。

[0263] 转录共激活结构域可包括人p300蛋白或其片段。转录共激活结构域可包括野生型人p300蛋白或突变型人p300蛋白或其片段。转录共激活结构域可包括人p300蛋白的核心赖氨酸-乙酰转移酶结构域,即p300 HAT核心(亦称"p300 WT核心";参见图58)。Cas9蛋白可以是其中核酸酶活性失活的Cas9蛋白。例如,融合蛋白中的Cas9蛋白可以是iCas9 (SEQ ID NO: 1的氨基酸36-1403),其包括D10A和H840A的氨基酸取代。Cas9融合蛋白可以是iCas9-p300 WT核心。

[0264] (3)gRNA

[0265] 所述方法还可包括将基于CRISPR/Cas9的系统和至少一种gRNA给予细胞或受试者,其中gRNA靶向不同的DNA序列。靶DNA序列可以是重叠的。给予细胞的gRNA的数目可为至少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。给予细胞的gRNA的数目可介于以下之间:至少1种gRNA与至少50种不同的gRNA、至少1种gRNA与至少45种不同的gRNA、至少1种gRNA与至少40种不同的gRNA、至少1种gRNA与至少35种不同的gRNA、至少1种gRNA与至少30种不同的gRNA、至少1种gRNA与至少25种不同的gRNA、至少1种gRNA与至少20种不同的gRNA、至少1种gRNA与至少16种不同的gRNA、至少1种gRNA与至少12种不同的gRNA、至少1种gRNA与至少8种不同的gRNA、至少1种gRNA与至少4种不同的gRNA、至少4种gRNA与至少50种不同的gRNA、至少4种不同的gRNA与至少45种不同的gRNA、至少4种不同的gRNA与至少40种不同的gRNA、至少4种不同的gRNA与至少35种不同的gRNA、至少4种不同的gRNA与至少30种不同的gRNA、至少4种不同的gRNA与至少25种不同的gRNA、至少4种不同的gRNA与至少20种不同的gRNA、至少4种不同的gRNA与至少16种不同的gRNA、至少4种不同的gRNA与至少12种不同的gRNA、至少4种不同的gRNA与至少8种不同的gRNA、至少8种不同的gRNA与至少50种不同的gRNA、至少8种不同的gRNA与至少45种不同的gRNA、至少8种不同的gRNA与至少40种不同的gRNA、至少8种不同的gRNA与至少35种不同的gRNA、8种不同的gRNA与至少30种不同的gRNA、至少8种不同的gRNA与至少25种不同的gRNA、8种不同的gRNA与至少20种不同的gRNA、至少8种不同的gRNA与至少16种不同的gRNA、8种不同的gRNA与至少12种不同的gRNA、至少8种不同的gRNA与至少8种不同的gRNA。

[0266] gRNA可包含靶DNA序列的互补多核苷酸序列接着是NGG。gRNA可在互补多核苷酸序列5'端包含"G"。gRNA可包含至少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个碱基对靶DNA序列的互补多核苷酸序列接着是NGG。gRNA可靶向启动子区、增强子区或靶基因的转录区的至少一个。gRNA可包括SEQ ID NO: 5-40、65-144、492-515、540-563和585-625的至少一个的核酸序列。

[0267] *b. 阻抑基因表达的方法*

[0268] 本公开内容提供使用上述基于CRISPR/Cas9的系统通过RNA,根据靶向基因组调节元件(例如远端增强子)阻抑内源基因(例如哺乳动物基因)表达的机制。Cas9融合蛋白可包括转录阻抑蛋白,例如KRAB阻抑蛋白。Cas9融合蛋白可以是dCas9-KRAB。dCas9-KRAB另外可通过将异染色质形成因子募集至靶定基因座而影响外遗传基因调节。CRISPR/dCas9-KRAB系统可用来阻抑基因的转录,但还可用来靶向通过之前通过传统阻抑方法(例如RNA干扰)达不到的基因组调节元件(图53B)。用靶定远端增强子的gRNA递送dCas9-KRAB可中断表达受该靶定增强子调节的多个基因(参见图53C)。靶定的增强子可以是基因的任何增强子,例如HS2增强子。

[0269] *a. 转分化或诱导分化的方法*

[0270] 本公开内容提供通过使用上述基于CRISPR/Cas9的系统通过RNA激活内源基因来使细胞转分化或诱导细胞分化的机制。

[0271] (1)转分化

[0272] 基于CRISPR/Cas9的系统可用来使细胞转分化。转分化,亦称谱系重编程或直接转换,是其中细胞从一种分化的细胞类型转化成另一种而无需经历中间多能状态或祖细胞类型的过程。它是一种化生类型,包括所有细胞命运转变,包括干细胞的相互转换。细胞的转分化对于疾病建模、药物研发、基因疗法和再生医学具有潜在用途。例如*BRN2*、*MYT1L*、*ASCL1*、*NANOG*和/或*MYOD1*等内源基因的激活,使用上述基于CRISPR/Cas9的系统可导致若干细胞类型例如成纤维细胞、心肌细胞、肝细胞、软骨细胞、间质祖细胞、造血干细胞或平滑肌细胞分别转分化成神经元和肌源表型。

[0273] (2)诱导分化

[0274] 基于CRISPR/Cas9的系统可用来诱导例如干细胞、心肌细胞、肝细胞、软骨细胞、间质祖细胞、造血干细胞或平滑肌细胞等的细胞分化。例如干细胞,例如胚胎干细胞或多能干细胞可被诱导分化成肌细胞或血管内皮细胞,即诱导神经元或生肌分化。

[0275] 12.基于多个CRISPR/Cas9的系统的应用

[0276] 基于多个CRISPR/Cas9的系统利用sgRNA设计的简便性和低成本,并可采用CRISPR/Cas9技术以助于利用高通量基因组研究的进展。例如,本文所述单一慢病毒载体可用于在不同的细胞系(例如本文所述原代成纤维细胞)中表达Cas9和多种sgRNA (图47)。基于多个CRISPR/Cas9的系统可与上述基于CRISPR/Cas9的系统的相同方式使用。

[0277] 除了所述转录激活和核酸酶功能性以外,该系统还可用于表达控制外遗传修饰的其它新的基于Cas9的效应器用于各种目的,包括基因组体系结构和内源基因调节途径的研究。由于内源基因调节在多种酶之间的微妙平衡,具有不同功能性的多重Cas9系统将允许检查不同调节信号间的复杂的相互作用。本文所述载体应与适体修饰的sgRNA和直向同源的Cas9相容以使使用一套sgRNA独立遗传操作成为可能。

[0278] 可使用基于多个CRISPR/Cas9的系统以激活细胞中的至少一个内源基因。所述方

法包括使细胞与修饰的慢病毒载体接触。内源基因可被瞬时激活或稳定激活。内源基因可被瞬时阻抑或稳定阻抑。融合蛋白可以与sgRNA相似的水平表达。融合蛋白可以与sgRNA相比不同的水平表达。细胞可以是原代人细胞。

[0279] 基于多个CRISPR/Cas9的系统可用于在细胞中进行多重基因编辑的方法。所述方法包括使细胞与修饰的慢病毒载体接触。多重基因编辑可包括纠正突变基因或插入转基因。纠正突变基因可包括缺失、重排或置换突变基因。纠正突变基因可包括核酸酶介导的非同源末端连接或同源指导修复。多重基因编辑可包括缺失或纠正至少一个基因,其中基因是内源正常基因或突变基因。多重基因编辑可包括缺失或纠正至少两个基因。例如,至少2个基因、至少3个基因、至少4个基因、至少5个基因、至少6个基因、至少7个基因、至少8个基因、至少9个基因或至少10个基因可缺失或纠正。

[0280] 基于多个CRISPR/Cas9的系统可用于在细胞中多重调节基因表达的方法。所述方法包括使细胞与修饰的慢病毒载体接触。所述方法可包括调节至少一个基因的基因表达水平。当与至少一个靶基因的正常基因表达水平相比至少一个靶基因的基因表达水平提高或降低时,至少一个靶基因的基因表达被调节。基因表达水平可以是RNA或蛋白质水平。

[0281] 13.纠正突变基因和治疗受试者的方法

[0282] 本公开内容还涉及纠正受试者的突变基因的方法。所述方法包括将组合物给予受试者的骨骼肌或心肌用于在上述骨骼肌或心肌中进行基因组编辑。组合物将位点特异性核酸酶递送至骨骼肌或心肌的用途可用修复模板或供体DNA恢复完全功能性或部分功能性蛋白质的表达,这可置换整个基因或含有突变的区域。位点特异性核酸酶可用来将位点特异性双链断裂引入靶定的基因组基因座中。当位点特异性核酸酶与靶DNA序列结合,因此允许切割靶DNA时,造成位点特异性双链断裂。这种DNA切割可刺激天然DNA修复机器,导致两种可能的修复途径之一:同源指导修复(HDR)或非同源末端连接(NHEJ)途径。

[0283] 本公开内容涉及在没有修复模板的情况下用位点特异性核酸酶进行基因组编辑,这可有效地纠正读框并恢复涉及遗传疾病的功能性蛋白质的表达。所公开的位点特异性核酸酶可包括采用同源指导修复或基于核酸酶介导的非同源末端连接(NHEJ)的纠正方法,其使得在可能不适于同源重组或基于选择的基因纠正的增殖受限制的原代细胞系中的有效纠正成为可能。这种策略将有活性的位点特异性核酸酶的快速稳健的装配与有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。

[0284] *a.核酸酶介导的非同源末端连接*

[0285] 从内源突变基因修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的瞬时方法形成对比,通过瞬时表达位点特异性核酸酶纠正基因组中的靶基因读框可导致各个修饰细胞及其所有子代永久性地恢复靶基因表达。

[0286] 核酸酶介导的NHEJ基因纠正可纠正突变的靶基因,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形成对比,NHEJ在细胞周期的所有阶段都有效地操作,因此可有效用于周期中的细胞和有丝分裂后细胞(例如肌纤维)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以外,这提供备选的稳健的永久性基因修复,并且在理论上将需要少至一种药物治疗。除本文所述质粒电穿孔方法以外,使用基于CRISPR/Cas9的系统的基于NHEJ的基因纠正,以及包括

大范围核酸酶和锌指核酸酶的其它工程改造的核酸酶可与用于基于细胞和基因的疗法的其它现行离体和体内平台联合。例如,通过基于mRNA的基因转移的CRISPR/Cas9型系统的递送或作为纯化的细胞可渗透蛋白质可使不含DNA的基因组编辑方法成为可能,所述方法可克服插入诱变的任何可能性。

[0287] *b. 同源指导修复*

[0288] 蛋白质表达自内源突变基因的修复可包括同源指导修复。上述方法还包括将供体模板给予细胞。供体模板可包括编码完全功能性蛋白质或部分功能性蛋白质的核苷酸序列。例如,供体模板可包括小型化肌养蛋白构建体,称为微小肌养蛋白("minidys"),一种完全功能性肌养蛋白构建体用于恢复突变型肌养蛋白基因或肌养蛋白基因的片段,其在同源指导修复后导致突变型肌养蛋白基因修复。

[0289] *c. 使用CRISPR/Cas9纠正突变基因和治疗受试者的方法*

[0290] 本公开内容还涉及用修复模板或供体DNA以基于CRISPR/Cas9的系统进行基因组编辑以恢复完全功能性或部分功能性蛋白质的表达,这可置换整个基因或含有突变的区域。基于CRISPR/Cas9的系统可用来将位点特异性双链断裂引入靶定的基因组基因座中。当基于CRISPR/Cas9的系统利用gRNA与靶DNA序列结合,因此允许切割靶DNA时,造成位点特异性双链断裂。基于CRISPR/Cas9的系统由于其成功有效的遗传修饰的高速率所致,具有高级基因组编辑的优势。这种DNA切割可刺激天然DNA修复机器,导致两种可能的修复途径之一:同源指导修复(HDR)或非同源末端连接(NHEJ)途径。例如,涉及肌养蛋白基因的基于CRISPR/Cas9的系统可包括具有SEQ ID NO: 65-115的任一个的核酸序列的gRNA。

[0291] 本公开内容涉及无需修复模板以基于CRISPR/Cas9的系统进行基因组编辑,这可有效地纠正读框并恢复涉及遗传疾病的功能性蛋白质的表达。所公开的基于CRISPR/Cas9的系统和方法可包括采用同源指导修复或基于核酸酶介导的非同源末端连接(NHEJ)的纠正方法,其使在可能不适于同源重组或基于选择的基因纠正的增殖受限制的原代细胞系中进行有效纠正成为可能。这种策略将有活性的基于CRISPR/Cas9的系统的快速稳健装配与有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引起的遗传疾病。

[0292] 本公开内容提供纠正细胞中的突变基因和治疗患有遗传疾病(例如DMD)的受试者的方法。所述方法可包括将上述基于CRISPR/Cas9的系统、编码所述基于CRISPR/Cas9的系统的多核苷酸或载体或所述基于CRISPR/Cas9的系统的组合物给予细胞或受试者。所述方法可包括给予基于CRISPR/Cas9的系统,例如给予Cas9蛋白或含有具有核酸酶活性的第二结构域的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gRNA靶向不同的DNA序列。靶DNA序列可以是重叠的。给予细胞的gRNA的数目如上所述可为至少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可包括SEQ ID NO: 65-115的至少一个的核酸序列。所述方法可包括同源指导修复或非同源末端连接。

[0293] 14. 治疗疾病的方法

[0294] 本公开内容涉及治疗有需要的受试者的方法。所述方法包括如上所述将组合物给

予受试者的组织用于在细胞或受试者基因组编辑中改变基因表达和进行工程改造或改变基因组DNA。所述方法可包括将组合物给予受试者的骨骼肌或心肌用于在上述骨骼肌或心肌中进行基因组编辑。受试者可能患有引起变性或衰弱或遗传疾病的骨骼肌或心肌病况。例如,受试者可能患有上述杜兴型肌营养不良。

[0295] *a. 杜兴型肌营养不良*

[0296] 上述方法可用于纠正肌养蛋白基因并恢复所述突变的肌养蛋白基因的完全功能性或部分功能性蛋白质表达。在一些方面和实施方案中,本公开内容提供用于降低患者的DMD的作用(例如临床症状/适应症)的方法。在一些方面和实施方案中,本公开内容提供用于治疗患者的DMD的方法。在一些方面和实施方案中,本公开内容提供用于预防患者的DMD的方法。在一些方面和实施方案中,本公开内容提供用于防止患者的DMD进一步发展的方法。

[0297] 15. 构建体和质粒

[0298] 上述组合物可包含编码本文公开的基于CRISPR/Cas9的系统的基因构建体。基因构建体(例如质粒)可包含编码基于CRISPR/Cas9的系统例如Cas9蛋白和Cas9融合蛋白和/或gRNA的至少一种的核酸。上述组合物可包含编码修饰的AAV载体的基因构建体和编码本文公开的位点特异性核酸酶的核酸序列。基因构建体(例如质粒)可包含编码位点特异性核酸酶的核酸。上述组合物可包含编码本文公开的修饰的慢病毒载体的基因构建体。基因构建体(例如质粒)可包含编码Cas9-融合蛋白和至少一种sgRNA的核酸。基因构建体可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、端粒或质粒或黏粒。

[0299] 基因构建体还可以是包括重组慢病毒、重组腺病毒和重组腺病毒相关病毒在内的重组病毒载体的基因组的一部分。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的遗传物质中的一部分。基因构建体可包含核酸编码序列的基因表达的调节元件。调节元件可以是启动子、增强子、起始密码子、终止密码子或聚腺苷酸化信号。

[0300] 核酸序列可构成可以是载体的基因构建体。载体将能够在哺乳动物的细胞中表达融合蛋白,例如Cas9-融合蛋白或位点特异性核酸酶。载体可以是重组体。载体可包含编码融合蛋白(例如Cas9-融合蛋白或位点特异性核酸酶)的异源核酸。载体可以是质粒。载体可用于用编码Cas9-融合蛋白或位点特异性核酸酶的核酸转染细胞,在其中发生Cas9-融合蛋白或位点特异性核酸酶系统表达的条件下培养并维持转化的宿主细胞。

[0301] 针对表达的稳定性和高水平,可使编码序列优化。在某些情况下,选择密码子以减少RNA的二级结构形成,例如因分子内键合形成的二级结构。

[0302] 载体可包含编码基于CRISPR/Cas9的系统或位点特异性核酸酶的异源核酸,并另可包含起始密码子(其可以是基于CRISPR/Cas9的系统或位点特异性核酸酶编码序列的上游)和终止密码子(其可以是基于CRISPR/Cas9的系统或位点特异性核酸酶编码序列的下游)。起始和终止密码子可与基于CRISPR/Cas9的系统或位点特异性核酸酶编码序列一起在读框中。载体还可包含与基于CRISPR/Cas9的系统或位点特异性核酸酶编码序列有效连接的启动子。与基于CRISPR/Cas9的系统或位点特异性核酸酶编码序列有效连接的启动子可以是来自以下的启动子:猿猴病毒40 (SV40)、小鼠乳腺瘤病毒(MMTV)启动子、人免疫缺陷病毒(HIV)启动子例如牛免疫缺陷病毒(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)启动子例如CMV即时早期启动子、EB病毒(EBV)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:人基因例如人泛素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,例如肌肉或皮肤特异性启动子。所述启动子的实例描述于美国专利申请公布号US20040175727,其内容通过引用以其整体结合到本文中。

[0303] 载体还可包含聚腺苷酸化信号,其可以是基于CRISPR/Cas9的系统或位点特异性核酸酶的下游。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载体的聚腺苷酸化信号(Invitrogen,San Diego,CA)。

[0304] 载体还可包含基于CRISPR/Cas9的系统(即Cas9蛋白或Cas9融合蛋白编码序列或sgRNA)或位点特异性核酸酶上游的增强子。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,例如来自CMV、HA、RSV或EBV的那些。多核苷酸功能增强子描述于美国专利号5,593,972、5,962,428和WO94/016737,其内容通过引用完整予以结合。载体还可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。载体还可包含调节序列,其非常适于在给予其载体的哺乳动物或人细胞中的基因表达。载体还可包含报道分子基因,例如绿色荧光蛋白(“GFP”)和/或选择标记,例如潮霉素(“Hygro”)。

[0305] 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等, Molecular Cloning and Laboratory Manual, 第2版, Cold Spring Harbor (1989),其通过引用完全予以结合。在一些实施方案中,载体可包含编码基于CRISPR/Cas9的系统的核酸序列,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,所述gRNA包含SEQ ID NO: 5-40、65-144、492-515、540-563和585-625的至少一个的核酸序列。

[0306] 16. 药物组合物

[0307] 组合物可呈药物组合物。药物组合物可包含约1 ng-约10 mg编码基于CRISPR/Cas9的系统或基于CRISPR/Cas9的系统蛋白质成分(即Cas9蛋白或Cas9融合蛋白)的DNA。药物组合物可包含约1 ng-约10 mg修饰的AAV载体的DNA和编码位点特异性核酸酶的核苷酸序列。药物组合物可包含约1 ng-约10 mg修饰的慢病毒载体的DNA。本发明的药物组合物按照待采用的给药方式配制。在其中药物组合物是注射用药物组合物的情况下,它们是灭菌、无致热原和无颗粒的。优选使用等渗制剂。一般而言,用于等渗性添加剂可包括氯化钠、葡萄糖、甘露糖醇、山梨糖醇和乳糖。在某些情况下,优选等渗溶液例如磷酸缓冲盐溶液。稳定剂包括明胶和白蛋白。在一些实施方案中,将血管收缩药加入制剂中。

[0308] 组合物还可包含药物可接受的赋形剂。药物可接受的赋形剂可为功能性分子作为载体、辅助剂、载体或稀释剂。药物可接受的赋形剂可以是转染促进剂,其可包括表面活性剂,例如免疫刺激复合体(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、囊泡例如角鲨烯和角鲨烯、透明质酸、脂质、脂质体、钙离子、病毒蛋白质、聚阴离子、聚阳离子或纳米粒或其它已知的转染促进剂。

[0309] 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,更优选聚L-谷氨酸以小于6 mg/ml的浓度存在于组合物中用于在骨骼肌或心肌中进行基因组编辑。转染促进剂还可包括表面活性剂,例如免疫刺激复合体(ISCOMS)、弗氏不完全佐剂、LPS类似物包括单磷酰脂质A、胞壁酰肽、醌类似物和囊泡例如角鲨烯和角鲨烯,而且透明质酸还可用来与基因构建体联合给予。在一些实施方案中,编码组合物的DNA载体还可包括转染促进剂例如脂质、脂质体,包括卵磷脂脂质体或本领域已知的其它脂质体,像DNA-脂质体混合物(参见例如WO9324640)、钙离子、病毒蛋白质、聚阴离子、聚阳离子或纳米粒或其它已知的转染促进剂。优选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。

[0310] 17. 递送方法

[0311] 本文提供递送用于提供基因构建体的药物制剂(优选上述组合物)的方法。组合物的递送可以是将作为在细胞中表达的核酸分子的组合物转染或电穿孔并递送至细胞表面。可采用BioRad Gene Pulser Xcell或Amaxa Nucleofector IIb装置,将核酸分子电穿孔。可使用几种不同的缓冲液,包括BioRad电穿孔溶液、Sigma磷酸缓冲盐水制品#D8537(PBS)、Invitrogen OptiMEM I (OM)或Amaxa Nucleofector溶液V (N.V.)。转染可包括转染试剂,例如Lipofectamine 2000。

[0312] 在组合物递送至组织时,载体因此进入哺乳动物细胞,转染的细胞将表达融合蛋白,例如基于CRISPR/Cas9的系统和/或位点特异性核酸酶。可将组合物给予哺乳动物以改变基因表达或再改造或改变基因组。例如,可将组合物给予哺乳动物以纠正哺乳动物的肌养蛋白基因。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,优选人、牛、猪或鸡。

[0313] *a. 基于CRISPR/Cas9的系统*

[0314] 可在有或没有体内电穿孔、脂质体介导、纳米粒促进和/或重组载体的情况下,通过DNA注射(亦称DNA接种)将编码基于CRISPR/Cas9的系统蛋白质成分(即Cas9蛋白或Cas9融合蛋白)的载体递送至哺乳动物中。重组载体可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/或重组腺伴随病毒。

[0315] 可将编码基于CRISPR/Cas9的系统蛋白质成分(即Cas9蛋白或Cas9融合蛋白)的核苷酸导入细胞以在遗传上纠正靶基因或改变基因的基因表达,例如激活或阻抑内源基因。例如,可将通过gRNA指向突变型肌养蛋白基因的编码基于CRISPR/Cas9的系统蛋白质成分(即Cas9蛋白或Cas9融合蛋白)的核苷酸导入DMD患者的成肌细胞中。或者,可将它们导入DMD患者的成纤维细胞中,且经遗传纠正的成纤维细胞用MyoD处理以诱导分化成为成肌细胞,可将其植入受试者(例如受试者的受损肌肉)中以证实经纠正的肌养蛋白是有功能的和/或治疗受试者。修饰的细胞还可以是干细胞,例如诱导多能干细胞、骨髓衍生祖细胞、骨骼肌祖细胞、DMD患者的人骨骼成肌细胞、CD133+细胞、mesoangioblast和MyoD-或Pax7-转导的细胞或其它生肌祖细胞。例如,基于CRISPR/Cas9的系统可引起诱导多能干细胞的神经元或生肌分化。

[0316] 18. 给药途径

[0317] 可通过不同的途径将组合物给予受试者,包括口服、胃肠外、舌下、经皮、直肠、跨粘膜、局部、通过吸入、通过口腔给药、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。对于兽用,按照正常的兽医实践,将组合物作为合适的可接受的制剂给药。兽医可容易地确定最合于具体动物的剂量方案和给药途径。可通过传统注射器、无针注射装置、“微弹轰击基因枪”或其它物理方法例如电穿孔(“EP”)、“流体动力方法”或超声给予组合物。

[0318] 可在有或没有体内电穿孔、脂质体介导、纳米粒促进、重组载体例如重组慢病毒、重组腺病毒和重组腺病毒相关病毒的情况下,通过几种技术包括DNA注射(亦称DNA接种)将组合物递送至哺乳动物中。可将组合物注射到骨骼肌或心肌中。例如,可将组合物注射到胫骨前肌中。

[0319] 19. 细胞类型

[0320] 这些递送方法和/或给药途径的任一种可与多种细胞类型(例如目前在研究中的基于细胞的疗法的这些细胞类型)一起使用。细胞类型可为成纤维细胞、多能干细胞、心肌细胞、肝细胞、软骨细胞、间质祖细胞、造血干细胞、平滑肌细胞或K562人类红细胞白血病细胞系。

[0321] *a. DMD*

[0322] 目前在研究中用于DMD的基于细胞的疗法的细胞类型包括永生化成肌细胞,例如野生型和DMD患者衍生系,例如Δ48-50 DMD、DMD 8036 (del48-50)、C25C14和DMD-7796细胞系、原始DMD真皮成纤维细胞、诱导多能干细胞、骨髓衍生祖细胞、骨骼肌祖细胞、DMD患者的人骨骼成肌细胞、CD133+细胞、mesoangioblasts、心肌细胞、肝细胞、软骨细胞、间质祖细胞、造血干细胞、平滑肌细胞和MyoD-或Pax7-转导的细胞或其它生肌祖细胞。人成肌细胞的永生化可用于经遗传纠正的成肌细胞的克隆衍化。可将细胞离体修饰以分离和扩增含有经遗传纠正的肌养蛋白基因且不含基因组蛋白质编码区中的其它核酸酶引导的突变的永生化DMD成肌细胞的克隆群。或者,通过非病毒或非整合性病毒基因转移,或通过含有细胞渗透性基序的纯化蛋白质和gRNA的直接递送,核酸酶的瞬时体内递送可能使具有最小或无外源DNA整合风险的高特异性原位纠正成为可能。

[0323] 20. 药盒

[0324] 本文提供药盒,其可用来编辑骨骼肌或心肌中的基因组,例如纠正突变基因。药盒包含用于在上述骨骼肌或心肌中进行基因组编辑的组合物和使用所述组合物的说明书。药盒中所包括的说明书可附在包装材料上或可作为药品说明书包括在内。虽然说明书通常为书面或印刷材料,但它们不限于此。本公开内容预期能够储存所述说明书并将其向终端用户传达的任何介质。所述介质包括包括但不限于电子存储介质(例如磁盘、磁带、盒式磁带、芯片)、光学介质(例如CD ROM)等。本文所用术语“说明书”可包括提供说明书的互联网站网址。

[0325] 用于在骨骼肌或心肌中进行基因组编辑的组合物可包括修饰的AAV载体和编码上述位点特异性核酸酶的核苷酸序列。位点特异性核酸酶可包括特异性结合并切割突变基因的ZFN、TALEN或上述基于CRISPR/Cas9的系统。可将上述位点特异性核酸酶包括在药盒中以特异性结合并靶向突变基因的特定区域。位点特异性核酸酶可能对上述突变的肌养蛋白基因有特异性。药盒还可包括上述供体DNA、gRNA或转基因。

[0326] *a. 基于CRISPR/Cas9的系统*

[0327] 本文提供可用来纠正突变基因的药盒。药盒包含用于纠正突变基因至少一种成分

和使用基于CRISPR/Cas9的系统的说明书。药盒中所包括的说明书可附在包装材料上或可作为药品说明书包括在内。虽然说明书通常为书面或印刷材料,但它们不限于此。本公开内容预期能够储存所述说明书并将其向终端用户传达的任何介质。所述介质包括包括但不限于电子存储介质(例如磁盘、磁带、盒式磁带、芯片)、光学介质(例如CD ROM)等。本文所用术语"说明书"可包括提供说明书的互联网站网址。

[0328] 至少一种成分可包括至少一种上述特异性靶向基因的基于CRISPR/Cas9的系统。药盒可包括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。药盒中可包括上述基于CRISPR/Cas9的系统以特异性结合并靶向靶基因编码区上游、内部或下游的特定靶区。例如,基于CRISPR/Cas9的系统可能对靶基因的启动子区有特异性或基于CRISPR/Cas9的系统可能对突变基因(例如上述突变的肌养蛋白基因)有特异性。药盒可包括上述供体DNA。

[0329] 21. 实施例

[0330] 对本领域技术人员极显而易见的是,本文所述公开内容的方法的其它合适的修改和改编是极适用的和值得重视的,可采用合适的等同方法实施而不偏离本公开内容的范围或本文公开的实施方案的方面。现已详细描述了本公开内容,通过参照下列实施例将更清楚地理解本公开内容,所述实施例不仅仅旨在说明本公开内容的某些方面和实施方案,不应视为限制公开内容的内容。本文引用的所有期刊参考文献、美国专利和出版物的公开内容通过引用以其整体结合到本文中。

[0331] 本发明具有多个方面,通过下列非限制性实施例说明。

[0332] 实施例1

[0333] 材料和方法

[0334] *细胞培养和转染*。HEK293T细胞通过杜克大学癌症中心机构(Duke University Cancer Center Facilities)获自美国组织保藏中心(American Tissue Collection Center,ATCC),在补充10%胎牛血清和1%青霉素/链霉素的DMEM中在37℃、5% CO_2下维持。按照生产厂说明书,将HEK293T细胞用Lipofectamine 2000 (Invitrogen)转染。转染效率按常规高于80%,在递送对照eGFP表达质粒后通过荧光显微镜术测定。按与个别gRNA表达质粒或等量的gRNA表达质粒(由等量的4种gRNA的混合物组成) 3:1的质量比转染Cas9表达质粒。

[0335] 将原生小鼠胚胎成纤维细胞(PMEF-HL,Millipore,Billerica,MA)接种(75,000个/孔)在24孔TCPS板(BD,Franklin Lakes,NJ)中,在由补充10% Premium Select FBS (Atlanta Biologicals,Lawrenceville,GA)、25 μg mL^{-1}庆大霉素(Invitrogen)、1X GlutaMAX、非必需氨基酸、丙酮酸钠和β-巯基乙醇(Invitrogen)的高葡萄糖DMEM组成的完全MEF培养基中在37℃和5% CO_2下维持。如前所述,MEF转染用单次1 μg cm^{-2}剂量的总质粒DNA进行,在用含聚(CBA-ABOL)的无血清和无抗生素的OptiMEM静电缩合后作为阳离子纳米复合物递送(Adler等, *Molecular therapy. Nucleic acids* 1, e32 (2012))。在转染后4小时,将OptiMEM用完全MEF培养基更换。在转染后48小时,对MEF进行qRT-PCR处理,或者将完全MEF培养基用含有以下的N3神经诱导培养基更换:DMEM/F-12 (Invitrogen)、1X N-2 Supplement (Invitrogen)、10 ng mL^{-1}人bFGF2 (Stemgent,Cambridge,MA)和25 μg mL^{-1}庆大霉素(Invitrogen)。使用GFP报道分子载体(pmax-GFP,3486 bp,Amaxa,Cologne,

Germany)以优化转染条件。将Cas9表达质粒按与4种gRNA表达质粒相同的混合物3:1或1:1的质量比转染。

[0336] *质粒*。编码野生型和H840A Cas9的质粒获自Addgene (质粒#39312和质粒#39316;Jinek等, *Science* 337, 816-821 (2012))。将H840A Cas9与FLAG表位标签和具有引入D10A突变的引物对的核定位序列(NLS)一起在N端符合读框地克隆至载体pcDNA3.1中。将VP64结构域、NLS和HA表位标签与Cas9 ORF符合读框地克隆在C端(图1a,图9a)。以gBlocks(Integrated DNA Technologies (IDT))订购tracrRNA和crRNA表达盒(Cong等, *Science*339, 819-823 (2013)),并克隆至具有KpnI和SacII位点的pZ供体质粒(Sigma)中。以gBlocks加以改动以包括BbsI限制位点促进新的指导RNA间隔区序列的快速克隆,订购嵌合指导RNA表达盒(Mali等, *Science*339, 823-826 (2013)) (图9b)。含有靶序列的寡核苷酸获自IDT,使之杂交,磷酸化,并使用BbsI位点克隆至合适的质粒中。靶序列见表2。

表 2 gRNA 的靶序列

[0337]

靶标	名称	序列	SEQ ID NO
ASCL1	CR1	GCTGGGTGTCCCATTGAAA	5
	CR2	CAGCCGCTCGCTGCAGCAG	6
	CR3	TGGAGAGTTTGCAAGGAGC	7
	CR4	GTTTATTCAGCCGGGAGTC	8
NANOG	CR1	CGCCAGGAGGGGTGGGTCTA	9
	CR2	CCTTGGTGAGACTGGTAGA	10
	CR3	GTCTTCAGGTTCTGTTGCT	11
	CR4	ATATTCCTGATTTAAAAGT	12
VEGFA	CR1	TTAAAAGTCGGCTGGTAGC	13
	CR2	CGGGCCGGGGGCGGGGTCC	14
	CR3	GCCCGAGCCGCGTGTGGAA	15
	CR4	CCTTCATTGCGGCGGGCTG	16
TERT	CR1	CCGACCCCTCCCGGGTCCC	17
	CR2	CAGGACCGCGCTTCCCACG	18
	CR3	TGCACCCTGGGAGCGCGAG	19
	CR4	CCGCACGCACCTGTTCCCA	20
IL1B	CR1	AAAACAGCGAGGGAGAAAC	21
	CR2	TTAACTTGATTGTGAAATC	22
	CR3	AAAACAATGCATATTTGCA	23
	CR4	AAAATCCAGTATTTTAATG	24
IL1R2	CR1	ACCCAGCACTGCAGCCTGG	25
	CR2	AACTTATGCGGCGTTTCCT	26
	CR3	TCACTTTAAAACCACCTCT	27
	CR4	GCATCTTTTTCTCTTTAAT	28
IL1RN	CR1	TGTACTCTCTGAGGTGCTC	29
	CR2	ACGCAGATAAGAACCAGTT	30
	CR3	CATCAAGTCAGCCATCAGC	31
	CR4	GAGTCACCCTCCTGGAAAC	32
HBG1/2	CR1	GCTAGGGATGAAGAATAAA	33
	CR2	TTGACCAATAGCCTTGACA	34
	CR3	TGCAAATATCTGTCTGAAA	35

[0338]

	CR4	AAATTAGCAGTATCCTCTT	36
MYOD1	CR1	CCTGGGCTCCGGGGCGTTT	37
	CR2	GGCCCCTGCGGCCACCCCG	38
	CR3	CTCCCTCCCTGCCCGGTAG	39
	CR4	AGGTTTGGAAAGGGCGTGC	40

[0339] *蛋白质印迹*。将细胞在50 mM Tris-Cl (pH 7.4)、150 mM NaCl、0.5% Triton X-100和0.1% SDS中裂解。将裂解物与加载缓冲液混合,煮沸5分钟,将等体积的蛋白质在NuPAGE® Novex 4-12%或10% Bis-Tris凝胶聚丙烯酰胺凝胶中电泳,并转移至硝酸纤维素膜上。将非特异性抗体结合用含5%脱脂奶的50 mM Tris/150 mM NaCl/0.1% Tween-20(TBS-T)封闭30分钟。将膜与在含5% BSA的TBS-T中1:1000稀释的第一抗体(HRP缀合的抗Flag (Cell Signaling,目录号2044)一起温育过夜;与在含5%牛奶的TBS-T中1:5000稀释的抗GAPDH (Cell Signaling,克隆14C10)温育30分钟;与在5% BSA中1:500稀释的抗ASCL1(Santa Cruz,克隆sc-48449)温育;或与在5%牛奶中1:500稀释的抗g-珠蛋白(Santa Cruz,克隆51-7)温育,并将膜用TBS-T洗涤30分钟。用第一抗体标记的膜与1:5000稀释的抗兔HRP缀合的抗体(Sigma-Aldrich)、抗山羊(1:3000)或抗小鼠(1:5000)温育30分钟,用TBS-T洗涤30分钟。使用Immun-Star WesternC™化学发光试剂盒(Bio-Rad)观测膜,使用ChemiDoc™ XRS+系统拍摄图像并应用ImageLab软件(Bio-Rad)处理。

[0340] *ELISA*。收集无血清培养基(OPTI-MEM),在-80℃下冷冻。按照生产商的方案,通过酶联免疫吸附测定法(ELISA)定量测定分泌到培养基中的人IL-1ra (R&D Systems,目录号DY280)。通过将重组人IL-1ra在OPTI-MEM中稀释,制备标准曲线,未经稀释地测量培养基中的IL-1ra。通过经由3 kDa MWCO滤膜离心20分钟,使样品浓缩约8倍(Amicon Ultra,目录号UFC500396)。报告值通过各样品的浓度系数校正。

[0341] 以540 nm处的波长校正,测量450 nm处的光密度。各标准品和样品以一式两份测定。对一式两份读数取平均,并通过减去平均零标准品光密度标准化。通过log转换数据生成标准曲线,并进行IL-1ra浓度与光密度的线性回归。报告值为在不同日期以对各实验取平均的技术性一式两份进行3个独立实验的均值和均值的标准误差(n = 3)。

[0342] *qRT-PCR*。使用RNeasy Plus RNA分离试剂盒(Qiagen),分离总RNA。使用SuperScript® VILO™ cDNA合成试剂盒(Invitrogen),进行cDNA合成。用应用Primer3Plus软件设计并购自IDT的表3中报告的寡核苷酸引物,用CFX96实时PCR检测系统(Bio-Rad)进行使用PerfeCTa® SYBR® Green FastMix的实时PCR。

表3 用于qRT-PCR的引物的序列

[0343]

靶标	正向引物	SEQ ID NO	反向引物	SEQ ID NO
hASCL1	GGAGCTTCTCGACTTCACCA	41	AACGCCACTGACAAGAAAGC	53
NANOG	GATTTGTGGGCCTGAAGAAA	42	CAGATCCATGGAGGAAGGAA	54
VEGFA	AAGGAGGAGGGCAGAATCAT	43	GGGTACTCCTGGAAGATGTCC	55
TERT	AAACCTTCCTCAGCTATGCCC	44	GTTTGCGACGCATGTTCCTC	56
IL1B	AGCTGATGGCCCTAAACAGA	45	AAGCCCTTGCTGTAGTGGTG	57
IL1R2	CAGGAGGACTCTGGCACCTA	46	CGGCAGGAAAGCATCTGTAT	58
IL1RN	GGAATCCATGGAGGGAAGAT	47	TGTTCTCGCTCAGGTCAGTG	59
HBG1/2	GCTGAGTGAACTGCACTGTGA	48	GAATTCTTTGCCGAAATGGA	60
MYOD1	CTCTCTGCTCCTTTGCCACA	49	GTGCTCTTCGGGTTTCAGGA	61
GAPDH	CAATGACCCCTTCATTGACC	50	TTGATTTTGGAGGGATCTCG	62
mASCL1	GGAACAAGAGCTGCTGGACT	51	GTTTTTCTGCCTCCCCATTT	63
mGAPDH	AACTTTGGCATTGTGGAAGG	52	GGATGCAGGGATGATGTTCT	64

[0344] 引物特异性通过琼脂糖凝胶电泳和解链曲线分析证实。计算在适当的动态范围内的反应效率以确保标准曲线的线性(图10)。结果表示为通过$\Delta\Delta C_T$方法归一化至GAPDH表达的目标基因mRNA表达的倍数增加。报告值为在不同日期以对各实验取平均的技术性一式两份进行的3个独立实验的均值和均值的标准误差($n = 3$)。

[0345] *RNA-Seq*。构建了RNA seq文库。简单地讲,使用SuperScript® VILO™ cDNA合成试剂盒(Invitrogen),自寡dT Dynabead® (Invitrogen)俘获的mRNA合成第一链cDNA。使用DNA聚合酶I (New England Biolabs)合成第二链cDNA。使用Agencourt AMPure XP珠粒(Beckman Coulter)纯化cDNA,并使用Nextera转座酶(Illumina;在55℃下5分钟)同时片段化并将测序引物插入双链cDNA中。使用QG缓冲液(Qiagen)终止转座反应,在AMPure XP珠粒上纯化片段化cDNA。通过6个PCR循环,产生编入索引的测序文库。

[0346] 在Illumina HiSeq 2000仪器的2条道上使用50-bp单端读数对文库测序,产生每文库2900万和7400万读数。应用Bowtie,将读数与人RefSeq转录物比对(Langmeadet等, *Genome biology*10,R25 (2009))。应用DESeq,计算差异表达的统计显著性,包括多重假设检验的校正(Anders等, *Genome biology*11, R106 (2010))。在Gene Expression Omnibus (GEO)中保存原始RNA-seq读数和与各RefSeq转录物比对的读数的数目用于公共存取,登记号目前待定。

[0347] *免疫荧光染色*。对于Tuj1和MAP2表达的检测,在N3培养基中培养第10天时,将转染的MEF用4% PFA (EMS,Hatfield,PA)在室温(RT)下固定20分钟。然后将细胞在含有0.2% Triton X-100、3% w/v BSA和10%山羊血清(Sigma-Aldrich,Saint Louis,MO)的封闭缓冲液中与兔抗Tuj1 (Covance,Princeton,New Jersey,克隆TUJ1 1-15-79,1:500)和小鼠抗MAP2 (BD,克隆Ap20,1:500)一起在室温下温育2小时,或再与小鼠抗Ascl1 (BD克隆

24B72D11.1,1:100)在4℃下温育24小时。然后将细胞用PBS洗涤3次,在封闭缓冲液中与Alexa Fluor 488山羊抗小鼠IgG和Alexa Fluor 594山羊抗兔IgG (Invitrogen,1:200)一起在室温下温育1小时,并用PBS洗涤3次。然后染色的MEF用带有ProScanII装发动机的镜台(Prior Scientific,Rockland,MA)的Nikon Eclipse TE2000-U倒置荧光显微镜扫描以产生各完全培养区的大的镶嵌图像。按照局部对比,用FIJI macro处理这些镶嵌图案以自动和均匀地对图像设置阈值,排除小的碎片,并对各孔中$Tuj1^{+}$细胞的数目计数。

[0348] 统计资料。在JMP 10 Pro中通过Tukey检验进行统计分析,α等于0.05。

[0349] 实施例2

[0350] 结果

[0351] 为了产生基于CRISPR/Cas9的转录激活系统,使Cas9的催化残基(D10A、H840A)突变以产生iCas9,并且与C端VP64酸性反式激活域遗传融合(图1a、b)。通过N端Flag表位标签的蛋白质印迹,从人胚肾(HEK) 293T细胞中的转染质粒中观察到稳健的iCas9-VP64表达(图3)。CRISPR系统通过gRNA中的20 bp序列与互补DNA靶(其后面是NGG前间区序列邻近基序(PAM)序列,其中N是任何碱基对)的碱基配对识别其靶标。靶定内源人启动子的合成转录因子的组合导致基因表达协同和稳健的激活。因此在转录起始位点500 bp内的*IL1RN*基因的启动子中鉴定出后面接NGG PAM序列的4个gRNA靶位点(图4,表2)。为了比较基于crRNA和基于gRNA的打靶策略,将4个靶位点序列引入crRNA和gRNA表达质粒中[17],并用iCas9-VP64表达质粒共转染至HEK293T细胞中。虽然在用crRNA的组合处理的样品中通过qRT-PCR观察到可观诱导的*IL1RN*表达,但用gRNA的组合实现高得多的水平(图1c)。在用gRNA和无VP64的iCas9的表达质粒处理的细胞中未观察到基因表达的改变,表明了激活结构域在调节基因表达中的关键作用(图1c)。通过在用iCas9-VP64和野生型Cas9处理的样品中进行Surveyor测定法以检测DNA修复事件,证实在iCas9-VP64系统中这些靶位点处的核酸酶活性已消除(图5)。通过单独或组合地转染4种gRNA的每一种,用gRNA的组合靶向启动子中的多个位点显示基因表达的稳健增加(图1d)。只有当gRNA组合与iCas9-VP64共转染时,才观察到高水平的*IL1RN*表达(图1d),如用其它类别的改造的转录因子观察到的一样。同样,在3个不同的实验中,只用gRNA的组合处理的6个样品中有3个观察到由*IL1RN*基因编码的IL-1受体拮抗剂(IL-1ra)蛋白的产生,而在用单一gRNA或对照质粒处理的样品中从未观察到(图1e)。为了检查由iCas9-VP64引起的基因激活的特异性,通过RNA-seq评价了用4种gRNA的组合处理的HEK293T细胞的整个基因表达(图1f)。显然,仅相对于对照表达显著增加的基因(错误发现率≤ 3×10^{-4})是从*IL1RN*基因座表达的4个同种型(图4),表明了高水平的基因激活特异性。

[0352] 为了证实该系统的普遍适用性,设计了4种gRNA以靶向与医学和生物工程学有关的8种其它基因的启动子的每一个,包括*ASCL1*、*NANOG*、*HBG1/2*、*MYOD1*、*VEGFA*、*TERT*、*IL1B*和*IL1R2* (图4,表2)。*ASCL1*和*MYOD1*的强制表达导致几个细胞类型分别转分化成神经元和肌源表型。*NANOG*是多能性的标志物,也用于遗传重编程策略。在胚胎发育期间编码γ-珠蛋白的同源物*HBG1*和*HBG2*的激活,可用作补偿镰状细胞病中的β-珠蛋白突变的治疗策略。仔细研究了*VEGFA*通过合成转录因子的增量调节作为提高组织再生和伤口愈合的策略。由*TERT*基因编码的端粒酶的强制表达,可用来使细胞系永生化。*IL1B*编码介导炎症和自身免疫的IL-1β细胞因子。可通过IL-1ra或由*IL1R2*编码的诱杀型受体的表达,阻断IL-1β信号转导。

通过qRT-PCR测定,通过iCas9-VP64和4种gRNA的表达质粒共转染至HEK293T细胞提高这些基因各自的表达(图2)。在某些情况下,单一gRNA的表达足以诱导基因表达,但在任何情况下,4种gRNA的共转染导致协同效应(图2a-d)。显然,通过DNase-seq测定的染色质可及性不是成功基因激活的预测指标(图4)。在用iCas9-VP64和靶向*HBG1*的4种gRNA (其中3种还靶向*HBG2*)转染的细胞上进行RNA-seq。这揭示了无法通过RNA-seq区分的*HBG1*和*HBG2*两者表达中的特异性和可再现的增加,虽然由于总表达水平低所致无法达到统计显著性(图6)。通过蛋白质印迹检测在用iCas9-VP64和4种gRNA处理后Ascl1和γ-珠蛋白的蛋白质表达的增加(图7),证实通过qRT-PCR观察到高较的mRNA水平(图2)。在空载体对照中可检出低的基线水平的Ascl1和γ-珠蛋白的蛋白质表达。由于通过iCas9-VP64激活基因靶的初步证据可导致基因网和细胞表型的继发改变,因此将iCas9-VP64和靶向*ASCL1*的4种gRNA的表达质粒共转染至鼠胚胎成纤维细胞(MEF)中(图8)。MEF中Ascl1的强制表达表明部分激活神经元基因网,包括下游靶标Tuj1。由于gRNA靶位点在人和小鼠*ASCL1*启动子中是保守的(图8a),因此还在用编码iCas9-VP64和4种gRNA的质粒处理的MEF中观察到*ASCL1*表达的激活(图8b)。此外,在iCas9-VP64/gRNA处理的样品中,在转染后12天通过免疫荧光染色,容易检出表达Ascl1和神经元标志物Tuj1的细胞(图8c-h)。在用对照质粒处理的细胞中未观察到Tuj1阳性细胞。

[0353] 迄今没有哺乳动物细胞中Cas9/CRISPR活性的特异性的任何全面调查。采用RNA-seq,显示靶定基因激活与可检测脱靶基因激活没有极强的特异性(图1f,图6)。对于该特异性分析选择*IL1RN*和*HBG1/2*,因为基因产物IL-1ra和γ-珠蛋白不会对HEK293T细胞中的基因表达产生继发效应。与使用单个强激活因子形成对比,利用多个弱转录激活因子的协同活性可增加特异性基因调节,因为多个相邻的脱靶位点会存在于另一个基因座上是不可能的。引人关注的是,与只用空表达质粒处理的对照样品相比,在用iCas9-VP64和靶定*IL1RN*或*HBG1/2*的gRNA处理的两种样品中,*IL32*基因适度被减量调节(错误发现率< 0.03) (图1f、图6)。因为靶定*IL1RN*和*HBG1/2*的样品两者同样受影响,所以这是与靶序列同一性有关的脱靶iCas9-VP64活性的结果是不可能的。

[0354] 为了评价iCas9-VP64与之结合基因组的特异性,使用抗HA抗体,对用iCas9-VP64和靶向IL1RN启动子的4种gRNA处理的细胞进行了ChIP测序。实验显示,iCas9靶向IL1RN启动子(图15)。此外,实验显示极高水平的特异性。iCas9只有10个潜在的脱靶结合位点(FDR < 5%)。为了进一步查询特异性,用iCas9 EGEM进行了RNA测序实验,发现相对于对照,仅IL1RN基因同种型的表达增加(FDR ≤ 3 x 10.4)。

[0355] 实施例3

[0356] 靶向肌养蛋白基因的CRISPR—方法和材料

[0357] *质粒构建体*。如前所述,使用酿脓链球菌sgRNA和人密码子优化的Cas9 (hCas9)核酸酶的表达盒(Perez-Pinera等, Nat Methods 10:973-976 (2013))。为了产生荧光报道分子系统以富集CRISPR/Cas9修饰的细胞,合成了GeneBlock (IDT)。其含有与多个克隆位点的紧接上游的T2A跳跃肽融合的Cas9编码序列3'端的一部分,随后克隆至hCas9表达载体中。然后将eGFP报道分子基因克隆至T2载体中,以供来自相同表达载体(hCas9-T2A-GFP, SEQ ID NO: 116)的Cas9和eGFP蛋白共翻译。

[0358] *细胞培养和转染*。HEK293T细胞经由Duke Cell Culture Facility获自美国组织

保藏中心(ATCC),在补充10%小牛血清和1%青霉素/链霉素的DMEM中维持。使永生化成肌细胞(Mamchaoui, K.等, *Skelet Muscle* 1, 1-11 (2011)) (一种来自野生型供体,两种来自Δ48-50 DMD患者衍生系)在补充20%小牛血清(Sigma)、50 μg/ml胎球蛋白、10 ng/ml人表皮生长因子(Sigma)、1 ng/ml人碱性成纤维细胞生长因子(Sigma)、10 μg/ml人胰岛素(Sigma)、1% GlutaMAX (Invitrogen)和1%青霉素/链霉素(Invitrogen)的骨骼肌培养基(PromoCell)中维持。原生DMD真皮成纤维细胞获自Coriell Cell存放库(GM05162A,Δ46-50),在补充10%胎牛血清、1 ng/mL人碱性成纤维细胞生长因子和1%青霉素/链霉素的DMEM中维持。所有细胞系都在37℃和5% CO_2下维持。

[0359] 按照生产商的方案,用Lipofectamine 2000 (Invitrogen)将HEK293T细胞用400 ng各表达载体在24孔板中转染。使用针对各系的优化条件,使用Gene Pulser XCell (BioRad)以PBS作为电穿孔缓冲液通过电穿孔,将永生化成肌细胞和原生成纤维细胞用5 μg各表达载体转染(图1) (Ousterout等, Mol Ther 21:1718-1726 (2013))。通过递送eGFP表达质粒(pmaxGFP,Clontech)并采用流式细胞术,测量转染效率。这些效率对于HEK293T通常≥95%,对于原生成纤维细胞和永生化成肌细胞≥70%。对于全部实验,电穿孔质粒的标示质量相当于用于各基于CRISPR/Cas9的系统的量。

[0360] *内源基因修饰的Cel-I定量测定(Surveyor测定法)*。采用Surveyor核酸酶测定法,定量测定内源靶位点处基于CRISPR/Cas9的系统诱导的损害(Guschin, D.Y.等, *Meth Mol Biol* 649, 247-256 (2010)),这可检测核酸酶介导的NHEJ的突变特性。在转染后,将细胞在37℃下孵育3或10天,使用DNeasy Blood and Tissue试剂盒(QIAGEN),提取基因组DNA。使用对各基因座有特异性的引物(参见表4),例如5'-GAGTTTGGCTCAAATTGTTACTCTT-3' (SEQ ID NO: 60)和5'-GGGAAATGGTCTAGGAGAGTAAAGT-3' (SEQ ID NO: 61),用AccuPrime High Fidelity PCR试剂盒(Invitrogen),通过35个PCR循环扩增靶基因座。

[0361] 表4 计算机预测的最前面10个脱靶位点的汇总和在用Cas9和所示sgRNA表达盒转染的HEK293T细胞中通过Surveyor测定法检测的各位点的活性。n.d.:未检出。

[0362]

SEQ ID NO.	指导	靶标	序列	PAM	评分	染色体	基因	内含子/外显子	# MMs	%得失位
67	CR3	指导	GCCTACTCAGACTGTTACTC	-	-	-	-	-	-	-
150		靶标	tCCTACTCAGACTGTTACTC	TGG	-	X	DMD	外显子	1	13.0
151		OT1	tCCTACTCAcACTGTTACTC	AGG	7.4	1	STRIP1	内含子	2	9.3
152		OT2	aCCTgCTCAcACTGTTACTC	CAG	2.5	2	ARHGAP25	内含子	3	n.d.
153		OT3	GCaTtCTCAaACTGTTACTC	AGG	2.4	13	None	无	3	n.d.
154		OT4	GgaTtCTCAcACTGTTACTC	GGG	1.3	14	PGPEP1	外显子	4	n.d.
155		OT5	aCaTACTtAtACTGTTACTC	TAG	1.3	19	MIXGA2	内含子	4	n.d.
156		OT6	tatTcCTaAGACTGTTACTC	AAG	0.9	8	LPPR1	内含子	5	n.d.
157		OT7	aaggACTaAGACTGTTACTC	GGG	0.9	9	RNF122	内含子	5	n.d.
158		OT8	GagctCTCAtACTGTTACTC	TAG	0.8	3	DNMBP	外显子	5	n.d.
159		OT9	GCaaAaTgAGACTGTTACTC	CAG	0.8	5	SLC12A2	内含子	4	n.d.
160		OT10	cCtcAtTCAGACTGTTACTC	AAG	0.8	4	KCNIP4	内含子	4	n.d.

[0363]

65	CR1	指导	GATTGGCTTTGATTTCCCTA	-	-	-	-	-	-	-
161		靶标	cATTGGCTTTGATTTCCCTA	GGG	-	X	DMD	内含子	1	8.3
162		OT1	aATTGGCATTGATTTCCCTA	GAG	7.1	16	无	无	2	0.8
163		OT2	cATTGGCTTTaATTTCCCTA	TAG	4.8	4	无	无	2	n.d.
164		OT3	GATaGGCTgTGATTTCCCTA	GAG	3.9	9	无	无	2	n.d.
165		OT4	GAaTaGCcTTGATTTCCCTA	AAG	2.4	1	无	无	3	n.d.
166		OT5	aATTtGCTTTGATTTCCCTg	AGG	1.5	1	TIMM17A	内含子	3	n.d.
167		OT6	GATgtGCTTTGATTTCCCTt	GGG	1.4	17	MYO1D	内含子	3	n.d.
168		OT7	aATTGGtTTTaATTTCCCTA	AAG	1.1	8	PIK1A	内含子	3	n.d.
169		OT8	aATTGGgTTTGATTTCCCTt	TGG	1.1	11	MS4A1	内含子	3	n.d.
170		OT9	GATgGGtTTTtATTTCCCTA	GAG	1.0	11	无	无	3	n.d.
171		OT10	GAaTGGtTTTGATTTCCCTg	GAG	1.0	11	无	无	3	n.d.
69	CR5	指导	GCAGTTGCCTAAGAACTGGT	-	-	-	-	-	-	-
172		靶标	aCAGTTGCCTAAGAACTGGT	GGG	-	X	DMD	内含子	1	14.0
173		OT1	cCAGTTGtCTAAGAACTGGg	GAG	1.5	5	NRG1	内含子	3	n.d.
174		OT2	GCAGTTGCCTgtGAACTGGT	AGG	1.4	X	无	无	2	n.d.
175		OT3	GCAGaTGCagAAGAACTGGT	GAG	1.4	19	SMIM7	内含子	3	n.d.
176		OT4	GCAGTTcCagAAGAACTGGT	GAG	0.9	11	GLB1L2	内含子	3	n.d.
177		OT5	caAcTTGCCTAtGAACTGGT	AGG	0.7	8	ASAP1	内含子	4	n.d.
178		OT6	aCAccTGCCTAAGAACTGGa	GGG	0.7	11	无	无	4	n.d.
179		OT7	tCAGgTGgCTAAGAACTGGg	TGG	0.7	14	NIN	内含子	4	n.d.
180		OT8	GaAGTTGgCcAAGAACTGGa	GAG	0.6	7	无	无	4	n.d.
181		OT9	GCtGcTGCCcAAGAACTGGc	AGG	0.6	11	AMOTL1	内含子	4	n.d.
182		OT10	tCAGcTGgCTAAGAACgGGT	AAG	0.6	7	ACTR3C	内含子	4	n.d.
70	CR6	指导	GGGGCTCCACCCTCACGAGT	-	-	-	-	-	-	-
183		靶标	aGGGCTCCACCCTCACGAGT	GGG	-	X	DMD	内含子	1	19.9
184		OT1	GcaGCTCagCCCTCACGAGT	CAG	0.8	3	无	无	4	n.d.
185		OT2	GGGGCTtCAgCaTCACGAGT	GAG	0.8	8	无	无	3	n.d.
186		OT3	GGGGCTCtcCCCTCACtAGT	GAG	0.6	8	无	无	3	n.d.

[0364]

187		OT4	GGGGaTCCACCtTCACcAGT	CAG	0.6	2	无	无	3	n.d.
188		OT5	aGGGCTggACCCTCACaAGT	AAG	0.4	16	AXIN1	内含子	4	n.d.
189		OT6	tGGtCTCCtCCCcCACGAGT	GGG	0.4	2	无	无	4	n.d.
190		OT7	aGGGCTCCcaCCcCACGAGT	GAG	0.3	5	无	无	4	n.d.
191		OT8	GaGGCTCCAtaCTCACcAGT	GAG	0.3	11	无	无	4	n.d.
192		OT9	GGaGCTgCcCCtTCACGAGT	GGG	0.3	3	无	无	4	n.d.
193		OT10	atGaCTCCACCCTCAaGAGT	AAG	0.3	8	AGPAT5	无	4	n.d.
100	CR36	指导	GCCTTCTTTATCCCCTATCG	-	-	-	-	-	-	-
194		靶标	GCCTTCTTTATCCCCTATCG	AGG	-	X	DMD	内含子	0	20.6
195		OT1	GtCTgCTgTgTCCCCTATCG	GGG	1.3	21	无	无	4	n.d.
196		OT2	cCCTTCTcTATCCCCTgTCG	TGG	1.3	8	无	无	3	n.d.
197		OT3	GCCTTCTTTATCCCCTcTCt	TGG	0.9	10	无	无	2	0.5
198		OT4	GCgcTCTTTtTCCCCTATCt	TAG	0.6	16	无	无	4	n.d.
199		OT5	GCCcTCTgTcTCCCCTgTCG	CAG	0.5	1	NFASC	无	4	n.d.
200		OT6	tCCATCTtTgTCCCCTATtG	AGG	0.5	10	无	无	4	n.d.
201		OT7	aCCtTCTCTcTCCCCTATaG	AGG	0.5	5	LOC100996485	内含子	4	n.d.
202		OT8	GttTTCTTTtTCCCCTATgG	GAG	0.5	3	无	无	4	n.d.
203		OT9	tgCTTCTTaATCCCCTATCa	AAG	0.4	7	无	无	4	n.d.
204		OT10	aCCTTCTTacTCCCCTATCc	GGG	0.4	10	ADARB2	无	4	n.d.

[0365] 所得PCR产物被随机解链,并在循环变温加热器按以下程序以各步骤间-0.3℃/s的速率重退火:95℃持续240 s,接着85℃持续60 s、75℃持续60 s、65℃持续60 s、55℃持续60 s、45℃持续60 s、35℃持续60 s和25℃持续60 s。在重退火后,将8 μL PCR产物与1 μL Surveyor核酸酶S和1 μL增强子S (Transgenomic)混合,在42℃下温育1小时。在温育后,将6 μL消化产物加载到10% TBE聚丙烯酰胺凝胶上,以200V电泳30分钟。将凝胶用溴化乙锭染色,并如前所述采用ImageLab (Bio-Rad)通过光密度测定法定量测定(Guschin等,*Meth Mol Biol* 649,247-256 (2010))。

[0366] *成肌细胞的荧光激活细胞分选*。将DMD成肌细胞用5微克各hCas9-T2A-GFP和sgRNA表达载体电穿孔,并在37℃下和5% CO_2孵育。在电穿孔后3天,对细胞进行胰蛋白消化,使用FACSvantage II分选机收集FACS分选物。收集GFP阳性细胞,使之生长用于分析。

[0367] *检测基因组缺失的基于PCR的测定法*。使用侧接各基因座的引物,通过PCR (Invitrogen AccuPrime High Fidelity PCR试剂盒),从基因组DNA扩增外显子51或外显子45-55基因座。对于外显子51,侧接引物为CelI-CR1/2-F和CelI-CR5-R,或对于外显子45-55分析为CelI-CR6-F和CelI-CR36-R (表4)。在TAE-琼脂糖凝胶中分析PCR产物,用溴化乙锭染色用于分析。

[0368] *易位的基于PCR的检测*。通过来自用仅Cas9 (对照)或Cas9与sgRNA转染的细胞的基因组DNA的两步嵌套PCR (Invitrogen AccuPrime High Fidelity PCR试剂盒用于各步骤),扩增预测的可能易位的基因座。在第一步中,使用经修饰以包括促进克隆和测序分析的限制位点的各基因座的Surveyor引物的组合,通过35个PCR循环扩增可能发生在各中靶和脱靶sgRNA靶位点的易位(表4)。使用针对每个个别预测易位定制设计的嵌套引物组,对1微升各PCR反应物进行了35轮PCR的第二轮扩增(表4)。各第二嵌套PCR引物在主要扩增子内

相同的近似区域内结合;然而,应用Primer3在线生物信息学软件优化各对以确保各易位的特异性检测。纯化相当于预测易位的预期长度且仅存在于用sgRNA处理的细胞中的PCR扩增子(QIAGEN Gel Extraction试剂盒),并通过Sanger测序分析。

[0369] *mRNA分析*。在对细胞进行胰蛋白消化前5天,通过将生长培养基用补充1%胰岛素-运铁蛋白-硒(Invitrogen #51500056)和1%青霉素/链霉素(Invitrogen #15140)的DMEM更换,使永生化成肌细胞分化成为肌纤维,并收集。按照生产商的说明书,使用RNeasy Plus Mini试剂盒(QIAGEN)从这些细胞中分离总RNA。按照生产商的说明书,使用VILO cDNA合成试剂盒(Life Technologies#11754)和1.5微克RNA在42℃下2小时,使RNA反转录为cDNA。使用通过CR1/5或CR2/5退火至外显子44和52以检测外显子51缺失的引物或通过CR6/36退火至外显子44和60以检测外显子45-55缺失的引物,用AccuPrime High Fidelity PCR试剂盒(Invitrogen)通过35个PCR循环扩增靶基因座(表4)。使PCR产物在TAE-琼脂糖凝胶上电泳,用溴化乙锭染色用于分析。使溶解的PCR条带克隆,并通过Sanger测序分析以证实预期的外显子连接处。表5列出实施例4中所用引物的序列。

[0370]

SEQ ID NO.	引物名称	引物序列	注释
205	CelI-CR1/2-F	GAGAGGTTATGTGGCTTTACCA	CR1/2 的正向 Surveyor 引物
206	CelI-CR1-R	AAAAATGCTTCCCACTTTGC	CR1 的反向 Surveyor 引物
207	CelI-CR2-R	CTCATTCTCATGCCTGGACA	CR2 的反向 Surveyor 引物
208	CelI-CR3-F	GAGTTTGGCTCAAATTGTTACTCTT	CR3 的正向 Surveyor 引物
209	CelI-CR3-R	GGGAAATGGTCTAGGAGAGTAAAGT	CR3 的反向 Surveyor 引物
210	CelI-CR4/31-F	GTTTGGCTCAAATTGTTACTCTTCA	CR4 或 CR31 的正向 Surveyor 引物
211	CelI-CR4/31-R	GTGAGAGTAATGTGTTTGCTGAGAG	CR4 或 CR31 的反向 Surveyor 引物
212	CelI-CR5-F	CGGGCTTGGACAGAACTTAC	CR5 的正向 Surveyor 引物
213	CelI-CR5-R	CTGCGTAGTGCCAAAACAAA	CR5 的反向 Surveyor 引物
214	CelI-CR6-F	TAATTTCATTGAAGAGTGGCTGAA	CR6 的正向 Surveyor 引物
215	CelI-CR6-R	AAGCCCTGTGTGGTAGTAGTCAGT	CR6 的反向 Surveyor 引物
216	CelI-CR7-F	TGAGTCATGTTGGATAACCAGTCT	CR7 的正向 Surveyor 引物
217	CelI-CR7-R	GAAGGTCAGGAACATACAATTCAA	CR7 的反向 Surveyor 引物
218	CelI-CR10/11-F	GATATGGGCATGTCAGTTTCATAG	CR10 或 CR11 的正向 Surveyor 引物
219	CelI-CR10/11-R	TGCTGTTGATTAATGGTTGATAGG	CR10 或 CR11 的反向 Surveyor 引物
220	CelI-CR12/13-F	TTTTAAATTGCCATGTTTGTGTC	CR12 或 CR13 的正向 Surveyor 引物
221	CelI-CR12/13-R	ATGAATAACCTAATGGGCAGAAAA	CR12 或 CR13 的反向 Surveyor 引物
222	CelI-CR14/15-F	TCAAGTCGCTTCATTTTGATAGAC	CR14 或 CR15 的正向 Surveyor 引物

[0371]

223	CelI-CR14/15-R	CACAACAAAACATATAGCCAAAGC	CR14 或 CR15 的反向 Surveyor 引物
224	CelI-CR16/17-F	TGCTGCTAAAATAACACAAATCAGT	CR16 或 CR17 的正向 Surveyor 引物
225	CelI-CR16/17-R	CTGTGCCTATTGTGGTTATCCTG	CR16 或 CR17 的反向 Surveyor 引物
226	CelI-CR18/19-F	ATTGATCTGCAATACATGTGGAGT	CR18 或 CR19 的正向 Surveyor 引物
227	CelI-CR18/19-R	TTTGCCTCTGCTATTACAGTATGG	CR18 或 CR19 的反向 Surveyor 引物
228	CelI-CR20/21-F	TGTAGGGTGGTTGGCTAAAATAAT	CR20 或 CR21 的正向 Surveyor 引物
229	CelI-CR20/21-R	TTTTTGCACAGTCAATAACACAAA	CR20 或 CR21 的反向 Surveyor 引物
230	CelI-CR22/23-F	GGCTGGTCTCACAATTGTACTTTA	CR22 或 CR23 的正向 Surveyor 引物
231	CelI-CR22/23-R	CATTATGGACTGAAAATCTCAGCA	CR22 或 CR23 的反向 Surveyor 引物
232	CelI-CR24/25-F	ATCATCCTAGCCATAACACAATGA	CR24 或 CR25 的正向 Surveyor 引物
233	CelI-CR24/25-R	TTCAGCTTTAACGTGATTTTCTGT	CR24 或 CR25 的反向 Surveyor 引物
234	CelI-CR26/27-F	GGATTCAGAAGCTGTTTACGAAGT	CR26 或 CR27 的正向 Surveyor 引物
235	CelI-CR26/27-R	TTTAGCTGGATTGGAAAAACAAAT	CR26 或 CR27 的反向 Surveyor 引物
236	CelI-CR28/29-F	AACTCACCCCATTGTTGGTATATT	CR28 或 CR29 的正向 Surveyor 引物
237	CelI-CR28/29-R	CCTTGTCCAAATACCGAAATACAT	CR28 或 CR29 的反向 Surveyor 引物
238	CelI-CR33-F	CACATAATTCATGAACTTGGCTTC	CR33 的正向 Surveyor 引物
239	CelI-CR33-R	TAGTAGCTGGGGAGGAAGATACAG	CR33 的反向 Surveyor 引物
240	CelI-CR34-F	TTTTTGTTTTAATTGCGACTGTGT	CR34 的正向 Surveyor 引物

[0372]

241	CelI-CR34-R	AGAAAAGGGGTTTTCTTTTGACTT	CR34 的反向 Surveyor 引物
242	CelI-CR35-F	CATTGTGACTGGATGAGAAGAAAC	CR35 的正向 Surveyor 引物
243	CelI-CR35-R	AACGGCTGTTATTAAAGTCCTCAG	CR35 的反向 Surveyor 引物
244	CelI-CR36-F	CAAGTCAGAAGTCACTTGCTTTGT	CR36 的正向 Surveyor 引物
245	CelI-CR36-R	TTTTATGTGCAGGAATCAGTCTGT	CR36 的反向 Surveyor 引物
246	Dys-E44-F	TGGCGGCGTTTTCATTAT	结合外显子 44 的正向 RT-PCR 引物
247	Dys-E52-R	TTCGATCCGTAATGATTGTTCTAGCC	结合外显子 52 的反向 RT-PCR 引物
248	Dys-E60-R	GGTCTTCCAGAGTGCTGAGG	结合外显子 60 的反向 RT-PCR 引物
249	CR3-CelI-OT1-F	TGTGTGCTTCTGTACACATCATCT	CR3 脱靶 1 的正向 Surveyor 引物
250	CR3-CelI-OT1-R	AGATTTCAACCCTCAAAAACTGAG	CR3 脱靶 1 的反向 Surveyor 引物
251	CR3-CelI-OT2-F	TAAACTCTTTCTTTTCCGCAATTC	CR3 脱靶 2 的正向 Surveyor 引物
252	CR3-CelI-OT2-R	CAAGGTGACCTGCTACCTAAAAAT	CR3 脱靶 2 的反向 Surveyor 引物
253	CR3-CelI-OT3-F	TATGACCAAGGCTATGTGTTCACT	CR3 脱靶 3 的正向 Surveyor 引物
254	CR3-CelI-OT3-R	ACAGCCTCTCTCCAGTAACATTCT	CR3 脱靶 3 的反向 Surveyor 引物
255	CR3-CelI-OT4-F	TATTCTTGCAGTGGTTTCACATTT	CR3 脱靶 4 的正向 Surveyor 引物
256	CR3-CelI-OT4-R	ATATTTTAAGCCAAGACCCAACAA	CR3 脱靶 4 的反向 Surveyor 引物
257	CR3-CelI-OT5-F	CTTTCAACTGTCTGTCTGATTGCT	CR3 脱靶 5 的正向 Surveyor 引物
258	CR3-CelI-OT5-R	AACAGCCTCTCTTCATTGTTCTCT	CR3 脱靶 5 的反向 Surveyor 引物

[0373]

259	CR3-CelI-OT6-F	CTCTGGAACTTGTCTCTGTCTTGA	CR3 脱靶 6 的正向 Surveyor 引物
260	CR3-CelI-OT6-R	CTTTCCTGCGTTCTCATGTTACTA	CR3 脱靶 6 的反向 Surveyor 引物
261	CR3-CelI-OT7-F	CCTTATATCCGTATCGCTCACTCT	CR3 脱靶 7 的正向 Surveyor 引物
262	CR3-CelI-OT7-R	CATATCTGTCTAACTTCCGCACAC	CR3 脱靶 7 的反向 Surveyor 引物
263	CR3-CelI-OT8-F	ACAGGTGTTATGTTGTCTGCATCT	CR3 脱靶 8 的正向 Surveyor 引物
264	CR3-CelI-OT8-R	ACTCCATTCCCAGATTAGTTATGC	CR3 脱靶 8 的反向 Surveyor 引物
265	CR3-CelI-OT9-F	CTGTTTTCTTTGTGAGAGTGGAGA	CR3 脱靶 9 的正向 Surveyor 引物
266	CR3-CelI-OT9-R	TGTAAGGTGGTCAAACTTGCTCTA	CR3 脱靶 9 的反向 Surveyor 引物
267	CR3-CelI-OT10-F	TTTTTCCTAGTACCCACAGATTTTT	CR3 脱靶 10 的正向 Surveyor 引物
268	CR3-CelI-OT10-R	TCCCTGATTCTCTCATTTGTGTTA	CR3 脱靶 10 的反向 Surveyor 引物
269	CR1-CelI-OT1-F	TTGGGAACATCAGAGAAAGTATGA	CR1 脱靶 1 的正向 Surveyor 引物
270	CR1-CelI-OT1-R	ACAAATTACAGTCTCCTGGGAAAG	CR1 脱靶 1 的反向 Surveyor 引物
271	CR1-CelI-OT2-F	AGTAGCTTACCTTGGCAGAGAAAA	CR1 脱靶 2 的正向 Surveyor 引物
272	CR1-CelI-OT2-R	TGACATACTGTTACCCTTTGCAGT	CR1 脱靶 2 的反向 Surveyor 引物
273	CR1-CelI-OT3-F	GAAAGGCTCAGTGAATGTTTGTT	CR1 脱靶 3 的正向 Surveyor 引物
274	CR1-CelI-OT3-R	CACTGCATCATCTCATTAAATCAA	CR1 脱靶 3 的反向 Surveyor 引物
275	CR1-CelI-OT4-F	CCCATATATTCATGATTACCCACA	CR1 脱靶 4 的正向 Surveyor 引物

[0374]

276	CR1-CelI-OT4-R	TATCAGAACGAGCACTAAAAGCAC	CR1 脱靶 4 的反向 Surveyor 引物
277	CR1-CelI-OT5-F	TTGGGAGGCTGAGGTACAAG	CR1 脱靶 5 的正向 Surveyor 引物
278	CR1-CelI-OT5-R	GAATGAAAAACAAACAGAAGGTGA	CR1 脱靶 5 的反向 Surveyor 引物
279	CR1-CelI-OT6-F	CTCCTCATCTGTACCCTTCAATCT	CR1 脱靶 6 的正向 Surveyor 引物
280	CR1-CelI-OT6-R	AGAGTGGCATCTAGTGTCAGTGAG	CR1 脱靶 6 的反向 Surveyor 引物
281	CR1-CelI-OT7-F	TACCAAAAGCTTCTCCTGTTTACC	CR1 脱靶 7 的正向 Surveyor 引物
282	CR1-CelI-OT7-R	GTAAGTTGGATGGCCTATTCTTTG	CR1 脱靶 7 的反向 Surveyor 引物
283	CR1-CelI-OT8-F	GAAGGAAATGCAAGGATACAAGAT	CR1 脱靶 8 的正向 Surveyor 引物
284	CR1-CelI-OT8-R	TGATTGAAAGAATCATTCCAGAAA	CR1 脱靶 8 的反向 Surveyor 引物
285	CR1-CelI-OT9-F	TCAGAAGGAAAATTGAAATTGGTT	CR1 脱靶 9 的正向 Surveyor 引物
286	CR1-CelI-OT9-R	CAGATGTGTTCTTCATCATTCCTC	CR1 脱靶 9 的反向 Surveyor 引物
287	CR1-CelI-OT10-F	TTCTCTTTAGGGAAAGCTCTCAAA	CR1 脱靶 10 的正向 Surveyor 引物
288	CR1-CelI-OT10-R	GGGTATAGATCATATGGAGGGAAG	CR1 脱靶 10 的反向 Surveyor 引物
289	CR5-CelI-OT1-F	AGATGATCTGCCCACCTCAG	CR5 脱靶 1 的正向 Surveyor 引物
290	CR5-CelI-OT1-R	CTTTCTTCCTCATTTAGTGGCAAT	CR5 脱靶 1 的反向 Surveyor 引物
291	CR5-CelI-OT2-F	ATGAATTGCAGATTGATGGTACTG	CR5 脱靶 2 的正向 Surveyor 引物
292	CR5-CelI-OT2-R	TCTCACCAAGAACCAAATTGTCTA	CR5 脱靶 2 的反向 Surveyor 引物

[0375]

293	CR5-Cell-OT3-F	GTAGGATACCTTGGCAACAGTCTT	CR5 脱靶 3 的正向 Surveyor 引物
294	CR5-Cell-OT3-R	TTAACGAATTGTGAGATTTGCTGT	CR5 脱靶 3 的反向 Surveyor 引物
295	CR5-Cell-OT4-F	TCAGAAAGTCAAGTAGCACACACA	CR5 脱靶 4 的正向 Surveyor 引物
296	CR5-Cell-OT4-R	AGAAGCACACACTCAGGTAAAGC	CR5 脱靶 4 的反向 Surveyor 引物
297	CR5-Cell-OT5-F	TCTTTGGGGGAATAATGACTAAAA	CR5 脱靶 5 的正向 Surveyor 引物
298	CR5-Cell-OT5-R	TTTGGCATTTATGGGAATAAAACT	CR5 脱靶 5 的反向 Surveyor 引物
299	CR5-Cell-OT6-F	ACTAATTCTGGTCAAGCCCATCA	CR5 脱靶 6 的正向 Surveyor 引物
300	CR5-Cell-OT6-R	TTAAGACATCGGATGAACAGAAAG	CR5 脱靶 6 的反向 Surveyor 引物
301	CR5-Cell-OT7-F	AGAAGCTTTCTGACATGATCTGC	CR5 脱靶 7 的正向 Surveyor 引物
302	CR5-Cell-OT7-R	TCAATTGCATTAGGACTTAGACCA	CR5 脱靶 7 的反向 Surveyor 引物
303	CR5-Cell-OT8-F	GTTAAATTACCTGTGAAGCCCTTG	CR5 脱靶 8 的正向 Surveyor 引物
304	CR5-Cell-OT8-R	CGGAAAACAGATCCACTTTATGAT	CR5 脱靶 8 的反向 Surveyor 引物
305	CR5-Cell-OT9-F	AAATCCACTGGAAACATCTTGAGT	CR5 脱靶 9 的正向 Surveyor 引物
306	CR5-Cell-OT9-R	AGTCTCTTCAGAATCATGCCCTAT	CR5 脱靶 9 的反向 Surveyor 引物
307	CR5-Cell-OT10-F	GCTTGGTGGCACATACCTGTAG	CR5 脱靶 10 的正向 Surveyor 引物
308	CR5-Cell-OT10-R	GGTAGGTAGATTTGCTTGCTTGTT	CR5 脱靶 10 的反向 Surveyor 引物
309	CR6-Cell-OT1-F	AGCTCTCAGCAGAGTAGGGATTTA	CR6 脱靶 1 的正向 Surveyor 引物

[0376]

310	CR6-CelI-OT1-R	GTGAGTCTACTGCACCCCATC	CR6 脱靶 1 的反向 Surveyor 引物
311	CR6-CelI-OT2-F	TGACACTGTGAAGTCAATTCTGTC	CR6 脱靶 2 的正向 Surveyor 引物
312	CR6-CelI-OT2-R	TCAAGAACTTGACAATGAGCAAAT	CR6 脱靶 2 的反向 Surveyor 引物
313	CR6-CelI-OT3-F	TATCCGATCCACTGTTGTGTGT	CR6 脱靶 3 的正向 Surveyor 引物
314	CR6-CelI-OT3-R	CAGGAGACCCAAAACCACTCTAC	CR6 脱靶 3 的反向 Surveyor 引物
315	CR6-CelI-OT4-F	TTGTTCTACAAATAGGGCTTCCTT	CR6 脱靶 4 的正向 Surveyor 引物
316	CR6-CelI-OT4-R	TGTTAAGTTTGGGCTTATGTTCCT	CR6 脱靶 4 的反向 Surveyor 引物
317	CR6-CelI-OT5-F	CACAAGTCTCACTGCACAAACAT	CR6 脱靶 5 的正向 Surveyor 引物
318	CR6-CelI-OT5-R	TGACCCATGATTATCTCTCTTTGA	CR6 脱靶 5 的反向 Surveyor 引物
319	CR6-CelI-OT6-F	TTCAGCTTCTGATTGGTTTTAATG	CR6 脱靶 6 的正向 Surveyor 引物
320	CR6-CelI-OT6-R	CCAATTCCTTAATTTTCCCTACAG	CR6 脱靶 6 的反向 Surveyor 引物
321	CR6-CelI-OT7-F	ATCTCAGACCAGGAGGGAGAC	CR6 脱靶 7 的正向 Surveyor 引物
322	CR6-CelI-OT7-R	CCTCAGGGTCAGTACATTTTTCAG	CR6 脱靶 7 的反向 Surveyor 引物
323	CR6-CelI-OT8-F	TTCTTAGGACATTGCTCCACATAC	CR6 脱靶 8 的正向 Surveyor 引物
324	CR6-CelI-OT8-R	GCAAACATAATGCAACTCGTAATC	CR6 脱靶 8 的反向 Surveyor 引物
325	CR6-CelI-OT9-F	GCAAGGGAGTCTGTGTCTTTG	CR6 脱靶 9 的正向 Surveyor 引物
326	CR6-CelI-OT9-R	TCATTTAAGTGGCTGTTCTGTGTT	CR6 脱靶 9 的反向 Surveyor 引物

[0377]

327	CR6-CelI-OT10-F	ACAAAACAGAGAGAAAAGGCAGAG	CR6 脱靶 10 的正向 Surveyor 引物
328	CR6-CelI-OT10-R	GTTTTGATTTCTGGTGCCTACAG	CR6 脱靶 10 的反向 Surveyor 引物
329	CR36-CelI-OT1-F	ACTGAAGCTGAAGCCCAGTC	CR36 脱靶 1 的正向 Surveyor 引物
330	CR36-CelI-OT1-R	ACATGAGCTCTCAGGTTTCTGAC	CR36 脱靶 1 的反向 Surveyor 引物
331	CR36-CelI-OT2-F	TCAAACTTAGATGGTTCCCTATGTT	CR36 脱靶 2 的正向 Surveyor 引物
332	CR36-CelI-OT2-R	GTACCCTGAAAATGTAGGGTGACT	CR36 脱靶 2 的反向 Surveyor 引物
333	CR36-CelI-OT3-F	CACTTCCCAAGTGAGGCAAT	CR36 脱靶 3 的正向 Surveyor 引物
334	CR36-CelI-OT3-R	CTATACTTGGGGCTGACTTGCTAC	CR36 脱靶 3 的反向 Surveyor 引物
335	CR36-CelI-OT4-F	TCGTATAGGTTACTTTGGCTCACA	CR36 脱靶 4 的正向 Surveyor 引物
336	CR36-CelI-OT4-R	AGGGATCTTTACTCCTCAGTGTGT	CR36 脱靶 4 的反向 Surveyor 引物
337	CR36-CelI-OT5-F	TGTAGAAGTTGGAATATCCTGCTG	CR36 脱靶 5 的正向 Surveyor 引物
338	CR36-CelI-OT5-R	GTCAACAATTTGATCTCAGGCTTC	CR36 脱靶 5 的反向 Surveyor 引物
339	CR36-CelI-OT6-F	CTCAGTACTAAAGATGGACGCTTG	CR36 脱靶 6 的正向 Surveyor 引物
340	CR36-CelI-OT6-R	AATCATTTCAGTCTTCCCAACAAT	CR36 脱靶 6 的反向 Surveyor 引物
341	CR36-CelI-OT7-F	GGGAATCACAGTAGATGTTTGTCA	CR36 脱靶 7 的正向 Surveyor 引物
342	CR36-CelI-OT7-R	AGACCAGGAGGTAAGAACATTTTG	CR36 脱靶 7 的反向 Surveyor 引物
343	CR36-CelI-OT8-F	CCACATAGAAAGAGACTTGCAGAA	CR36 脱靶 8 的正向 Surveyor 引物

[0378]

344	CR36-CelI-OT 8-R	AGAGATGCCAAAAGAACAGTCA AT	CR36 脱靶 8 的反向 Surveyor 引物
345	CR36-CelI-OT 9-F	TGTGCCTTAGGCTATGTAAACTG T	CR36 脱靶 9 的正向 Surveyor 引物
346	CR36-CelI-OT 9-R	AAACCCTTGTAACCAAAATTACC A	CR36 脱靶 9 的反向 Surveyor 引物
347	CR36-CelI-OT 10-F	TAACTGCATCAGAAGTCCTTGCT A	CR36 脱靶 10 的正向 Surveyor 引物
348	CR36-CelI-OT 10-R	GGAGACCAAGCTGCTAAAGTCA	CR36 脱靶 10 的反向 Surveyor 引物
349	CelI-CR3-F-嵌套	GTGGTGccgcggGAGTTTGGCTCAA ATTGTTACTCTT	嵌套 PCR 第一轮引物
350	CelI-CR3-R-嵌套	GTGGTGccgcggGGGAAATGGTCT AGGAGAGTAAAGT	嵌套 PCR 第一轮引物
351	CelI-CR1-F-嵌套	GTGGTGccgcggGAGAGGTTATGT GGCTTTACCA	嵌套 PCR 第一轮引物
352	CelI-CR1-R-嵌套	GTGGTGccgcggCTCATTCTCATGC CTGGACA	嵌套 PCR 第一轮引物
353	CelI-CR5-F-嵌套	GTGGTGccgcggCGGGCTTGGACA GAACTTAC	嵌套 PCR 第一轮引物
354	CelI-CR5-R-嵌套	GTGGTGccgcggCTGCGTAGTGCCA AAACAAA	嵌套 PCR 第一轮引物
355	CelI-CR6-F-嵌套	GTGGTGccgcggTAATTTCATTGAA GAGTGGCTGAA	嵌套 PCR 第一轮引物
356	CelI-CR6-R-嵌套	GTGGTGccgcggAAGCCCTGTGTG GTAGTAGTCAGT	嵌套 PCR 第一轮引物
357	CelI-CR36-F-嵌套	GTGGTGccgcggCAAGTCAGAAGT CACTTGCTTTGT	嵌套 PCR 第一轮引物
358	CelI-CR36-R-嵌套	GTGGTGccgcggTTTTATGTGCAGG AATCAGTCTGT	嵌套 PCR 第一轮引物
359	CR3-CelI-OT1-F-嵌套	GTGGTGccgcggTGTGTGCTTCTGT ACACATCATCT	嵌套 PCR 第一轮引物
360	CR3-CelI-OT1-R-嵌套	GTGGTGccgcggAGATTTCAACCCT CAAAAACTGAG	嵌套 PCR 第一轮引物
361	CR1-CelI-OT1-F-嵌套	GTGGTGccgcggTTGGGAACATCA GAGAAAGTATGA	嵌套 PCR 第一轮引物

[0379]

362	CR1-CelI-OT1-R-嵌套	GTGGTGccgcggACAAATTACAGTCTCCTGGGAAAG	嵌套 PCR 第一轮引物
363	CR36-CelI-OT3-F-嵌套	GTGGTGccgcggCACTTCCCAAGTGAGGCAAT	嵌套 PCR 第一轮引物
364	CR36-CelI-OT3-R-嵌套	GTGGTGccgcggCTATACTTGGGGCTGACTTGCTAC	嵌套 PCR 第一轮引物
365	CR3-P1/P3-F	GTGGTGccgcggTTGGCTCTTTAGCTTGTGTTTC	嵌套 PCR 第二轮引物
366	CR3-P1/P3-R	GTGGTGccgcggTGAGACTCCCAAAGGCAATC	嵌套 PCR 第二轮引物
367	CR3-P1/P4-F	GTGGTGccgcggTTGGCTCTTTAGCTTGTGTTTC	嵌套 PCR 第二轮引物
368	CR3-P1/P4-R	GTGGTGccgcggACTGAGGGGTGATCTTGGTG	嵌套 PCR 第二轮引物
369	CR3-P2/P3-F	GTGGTGccgcggGCAGAGAAAGCCAGTCGGTA	嵌套 PCR 第二轮引物
370	CR3-P2/P3-R	GTGGTGccgcggTGAGACTCCCAAAGGCAATC	嵌套 PCR 第二轮引物
371	CR3-P2/P4-F	GTGGTGccgcggGCAGAGAAAGCCAGTCGGTA	嵌套 PCR 第二轮引物
372	CR3-P2/P4-R	GTGGTGccgcggACTGAGGGGTGATCTTGGTG	嵌套 PCR 第二轮引物
373	CR1-P1/P5-F	GTGGTGccgcggCCAGAGTTCCTAGGGCAGAG	嵌套 PCR 第二轮引物
374	CR1-P1/P5-R	GTGGTGccgcggAGCTAGTCCCCACATTCCAC	嵌套 PCR 第二轮引物
375	CR1-P1/P6-F	GTGGTGccgcggCCAGAGTTCCTAGGGCAGAG	嵌套 PCR 第二轮引物
376	CR1-P1/P6-R	GTGGTGccgcggGGTGGAGGGAAACTTTAGGC	嵌套 PCR 第二轮引物
377	CR1-P2/P5-F	GTGGTGccgcggCTCATTCTCATGCCTGGACA	嵌套 PCR 第二轮引物
378	CR1-P2/P5-R	GTGGTGccgcggAGCTAGTCCCCACATTCCAC	嵌套 PCR 第二轮引物
379	CR1-P2/P6-F	GTGGTGccgcggTCTCATGCCTGGACAAGTAACT	嵌套 PCR 第二轮引物
380	CR1-P2/P6-R	GTGGTGccgcggGGTGGAGGGAAACTTTAGGC	嵌套 PCR 第二轮引物

[0380]

381	CR5-P3/P5-F	GTGGTGccgcggGGCTTGGACAGA ACTTACCG	嵌套 PCR 第二轮引物
382	CR5-P3/P5-R	GTGGTGccgcggCACCACTGTCTGC CTAAGGA	嵌套 PCR 第二轮引物
383	CR5-P4/P6-F	GTGGTGccgcggGGCTTGGACAGA ACTTACCG	嵌套 PCR 第二轮引物
384	CR5-P4/P6-R	GTGGTGccgcggGGTGGAGGGAAA CTTTAGGC	嵌套 PCR 第二轮引物
385	CR5-P3/P5-F	GTGGTGccgcggCGTAGTGCCAAA ACAAACAGT	嵌套 PCR 第二轮引物
386	CR5-P3/P5-R	GTGGTGccgcggCACCACTGTCTGC CTAAGGA	嵌套 PCR 第二轮引物
387	CR5-P4/P6-F	GTGGTGccgcggCGTAGTGCCAAA ACAAACAGT	嵌套 PCR 第二轮引物
388	CR5-P4/P6-R	GTGGTGccgcggGGTGGAGGGAAA CTTTAGGC	嵌套 PCR 第二轮引物
389	CR6-P1/P5-F	GTGGTGccgcggGCGAGGGCCTAC TTGATATG	嵌套 PCR 第二轮引物
390	CR6-P1/P5-R	GTGGTGccgcggCTTCCCAAGTGAG GCAATGC	嵌套 PCR 第二轮引物
391	CR6-P1/P6-F	GTGGTGccgcggACGTTTTGTGCTG CTGTAACA	嵌套 PCR 第二轮引物
392	CR6-P1/P6-R	GTGGTGccgcggCTGCAGGCACATT CTCTTCC	嵌套 PCR 第二轮引物
393	CR6-P2/P5-F	GTGGTGccgcggGCCCTGTGTGGTA GTAGTCA	嵌套 PCR 第二轮引物
394	CR6-P2/P5-R	GTGGTGccgcggCTTCCCAAGTGAG GCAATGC	嵌套 PCR 第二轮引物
395	CR6-P2/P6-F	GTGGTGccgcggCAGTATTAAGGG GTGGGAGCT	嵌套 PCR 第二轮引物
396	CR6-P2/P6-R	GTGGTGccgcggTCTCTTCCTCACA CAGCTGA	嵌套 PCR 第二轮引物
397	CR36-P3/P5-F	GTGGTGccgcggGGAGCTTGGAGG GAAGAGAA	嵌套 PCR 第二轮引物
398	CR36-P3/P5-R	GTGGTGccgcggCTTCCCAAGTGAG GCAATGC	嵌套 PCR 第二轮引物
399	CR36-P4/P6-F	GTGGTGccgcggATGGATGGGGAA GACACTGG	嵌套 PCR 第二轮引物
400	CR36-P4/P6-R	GTGGTGccgcggCTGCAGGCACATT CTCTTCC	嵌套 PCR 第二轮引物

[0381]

401	CR36-P3/P5-F	GTGGTGccgcggGGATGAAACAGG GCAGGAAC	嵌套 PCR 第二轮引物
402	CR36-P3/P5-R	GTGGTGccgcggTTCCCAAGTGAG GCAATGC	嵌套 PCR 第二轮引物
403	CR36-P4/P6-F	GTGGTGccgcggTTTGCAGAGCCAT GATGAGG	嵌套 PCR 第二轮引物
404	CR36-P4/P6-R	GTGGTGccgcggCGACAGCCAAAA CAGCCG	嵌套 PCR 第二轮引物

[0382] *蛋白质印迹分析*。为了评价肌养蛋白表达,通过将生长培养基用补充1%胰岛素-运铁蛋白-硒(Invitrogen)和1%抗生素/抗真菌剂(Invitrogen)的DMEM更换4-7天,例如6或7天,使永生化成肌细胞分化成肌纤维。通过诱导MyoD过量表达,并使细胞在补充1%胰岛素-运铁蛋白-硒(Invitrogen)、1%抗生素/抗真菌剂(Invitrogen)和3 μg/mL多西环素的DMEM中孵育15天,使成纤维细胞转分化成为成肌细胞。在转染HEK293T细胞后3天,评价肌养蛋白表达。对细胞进行胰蛋白消化,收集,在补充蛋白酶抑制剂混合物(Sigma)的RIPA缓冲液(Sigma)中裂解,并按照生产商的说明书(Pierce),采用二辛可宁酸测定法,定量测定总的蛋白质量。然后将样品与NuPAGE加载缓冲液(Invitrogen)和5% β-巯基乙醇混合,加热至85℃ 10分钟。在4-12% NuPAGE Bis-Tris凝胶(Invitrogen)上用MES缓冲液(Invitrogen)分离25微克蛋白质。将蛋白质在含有10-20%甲醇(例如10%甲醇)和0.01% SDS的转移缓冲液中转移到硝酸纤维素膜中达1-2小时。然后将印迹用5%牛奶-TBST在室温下封闭1小时。印迹用下列第一抗体探测:MANDYS8以检测肌养蛋白(1:100,Sigma D8168)和兔抗GAPDH (1:5000,Cell Signaling 2118S)。然后将印迹与小鼠或兔辣根过氧化物酶缀合的第二抗体(Santa Cruz)一起温育,采用ChemiDoc化学发光系统(BioRad)和Western-C ECL底物(BioRad)观测。

[0383] *移植至免疫缺陷型小鼠中*。所有动物实验按照经Duke Institutional Animal Care & Use Committee批准的方案进行。对细胞进行胰蛋白消化,收集,并在1X Hank平衡盐溶液(HBSS,Sigma)中洗涤。在临注射前,使200万个细胞沉淀,重新悬液于补充心脏毒素(Sigma #C9759)的5 μL 1X HBSS (Sigma)中。通过肌内注射,将这些细胞移植入NOD.SCID.γ (NSG)小鼠(Duke CCIF Breeding Core)的后肢胫骨前肌(TA)肌肉中。在注射后4周,使小鼠安乐死,收获TA肌肉。

[0384] *免疫荧光染色*。将收获的肌肉在30%甘油中在4℃下温育过夜后,在Optimal Cutting Temperature化合物中封片并冷冻。通过对-20℃下的包埋肌肉组织冷冻切片,获得连续的10微米切片。然后将冰冻切片在PBS中洗涤除去OCT化合物,随后在含有10%热灭活胎牛血清(用于血影蛋白检测)或5%热灭活胎牛血清(用于肌养蛋白检测)的PBS中在室温下封闭30-60分钟。将冰冻切片与下列只对人表位有特异性的第一抗体一起在4℃下温育过夜:抗血影蛋白(1:20,Leica NCL-SPEC1)或抗肌养蛋白(1:2,Leica NCL-DYS3)。在初步染色后,使用基于tyramide的免疫荧光信号扩增检测试剂盒(Life Technologies,TSA Kit #22,目录号T-20932),检测血影蛋白或肌养蛋白表达。简单地讲,将冰冻切片与1:200第二山羊抗小鼠生物素-XX (Life Technologies #B2763)在封闭缓冲液中在室温下温育1小时。然后使用链霉抗生物素-HRP缀合物(1:100,来自TSA Kit)在封闭缓冲液中使信号在室温

下扩增1小时。最后,将冰冻切片与tyramide-AlexaFluor488缀合物(1:100,TSA Kit)一起在生产商提供的扩增缓冲液中在室温下温育10分钟。然后染色的冰冻切片在ProLong AntiFade (Life Technologies #P36934)中封片,用常规荧光显微术观测。

[0385] *细胞毒性测定法*。为了定量评价潜在的sgRNA或SpCas9核酸酶相关细胞毒性,按照生产商的说明书,使用Lipofectamine 2000 (Invitrogen),将HEK293T细胞用10 ng GFP报道分子和100 ng SpCas9表达载体和100 ng sgRNA表达载体转染。通过流式细胞术,在2和5天评价GFP阳性细胞的百分比。计算存活率为从第2天到第5天GFP阳性细胞的降低,按所述归一化至用空核酸酶表达载体转染的细胞(Cornu等, Meth Mol Biol 649:237-245 (2010))。

[0386] 实施例4

[0387] 靶向肌养蛋白基因的CRISPR—结果

[0388] 设计基于CRISPR/Cas9的系统以靶向肌养蛋白基因。选择不同的gRNA以靶向基于NNNNN NNNNN NNNNN NNNNN NGG和GNNNN NNNNN NNNNN NNNNN NGG的人和小鼠肌养蛋白基因的不同区(参见表6、7和8)。

表 6

[0389]

名 称 (SEQ ID NO)	物种	基因	靶标	链	19bp 用于嵌合(5'端加入 G)	PAM
DCR1 (65)	人	DMD	内含子 50	+	attggctttgatttcccta	GGG
DCR2 (66)	人	DMD	内含子 50	-	tgtagagtaagtcagccta	TGG
DCR3 (67)	人	DMD	外显子 51-55'	+	cctactcagactgttactc	TGG
DCR4 (68)	人	DMD	外显子 51-53'	+	ttggacagaacttaccgac	TGG
DCR5 (69)	人	DMD	内含子 51	-	cagttgcctaagaactggt	GGG
DCR6 (70)	人	DMD	内含子 44	-	GGGCTCCACCCTCACGAGT	GGG
DCR7 (71)	人	DMD	内含子 55	+	TTTGCTTCGCTATAAAACG	AGG
DCR8 (72)	人	DMD	外显子 41	+	TCTGAGGATGGGGCCGCAA	TGG
DCR9 (73)	人	DMD	外显子 44	-	GATCTGTCAAATCGCCTGC	AGG
DCR10 (74)	人	DMD	外显子 45	+	CCAGGATGGCATTGGGCAG	CGG
DCR11 (75)	人	DMD	外显子 45	+	CTGAATCTGCGGTGGCAGG	AGG
DCR12	人	DMD	外显子 46	-	TTCTTTTGTTCTTCTAGCc	TGG

[0390]

名 称 (SEQ ID NO)	物种	基因	靶标	链	19bp 用于嵌合(5'端加入 G)	PAM
(76)						
DCR13 (77)	人	DMD	外显子 46	+	GAAAAGCTTGAGCAAGTCA	AGG
DCR14 (78)	人	DMD	外显子 47	+	GAAGAGTTGCCCCTGCGCC	AGG
DCR15 (79)	人	DMD	外显子 47	+	ACAAATCTCCAGTGGATAA	AGG
DCR16 (80)	人	DMD	外显子 48	-	TGTTTCTCAGGTAAAGCTC	TGG
DCR17 (81)	人	DMD	外显子 48	+	GAAGGACCATTTGACGTTa	AGG
DCR18 (82)	人	DMD	外显子 49	-	AACTGCTATTTCAGTTTCc	TGG
DCR19 (83)	人	DMD	外显子 49	+	CCAGCCACTCAGCCAGTGA	AGG
DCR20 (84)	人	DMD	外显子 50	+	gtatgcttttctgttaaag	AGG
DCR21 (85)	人	DMD	外显子 50	+	CTCCTGGACTGACCACTAT	TGG
DCR22 (86)	人	DMD	外显子 52	+	GAACAGAGGCGTCCCCAGT	TGG
DCR23 (87)	人	DMD	外显子 52	+	GAGGCTAGAACAATCATTA	CGG
DCR24	人	DMD	外显子 53	+	ACAAGAACACCTTCAGAAC	CGG

[0391]

名 称 (SEQ ID NO)	物种	基因	靶标	链	19bp 用于嵌合(5'端加入 G)	PAM
(88)						
DCR25 (89)	人	DMD	外显子 53	-	GGTTTCTGTGATTTTCTTT	TGG
DCR26 (90)	人	DMD	外显子 54	+	GGCCAAAGACCTCCGCCAG	TGG
DCR27 (91)	人	DMD	外显子 54	+	TTGGAGAAGCATTCATAAA	AGG
DCR28 (92)	人	DMD	外显子 55	-	TCGCTCACTCACCctgcaa	AGG
DCR29 (93)	人	DMD	外显子 55	+	AAAAGAGCTGATGAAACAA	TGG
DCR30 (94)	人	DMD	5'UTR/外显子 1	+	TAcACTTTTCaAAATGCTT	TGG
DCR31 (95)	人	DMD	外显子 51	+	gagatgatcatcaagcaga	AGG
DCR32 (96)	小鼠	DMD	mdx 突变	+	ctttgaaagagcaaTaaaa	TGG
DCR33 (97)	人	DMD	内含子 44	-	CACAAAAGTCAAATCGGAA	TGG
DCR34 (98)	人	DMD	内含子 44	-	ATTTCAATATAAGATTCGG	AGG
DCR35 (99)	人	DMD	内含子 55	-	CTTAAGCAATCCCGAACTC	TGG
DCR36	人	DMD	内含子 55	-	CCTTCTTTATCCCCTATCG	AGG

[0392]

名 称 (SEQ ID NO)	物种	基因	靶标	链	19bp 用于嵌合(5'端加入 G)	PAM
(100)						
DCR40 (104)	小鼠	DMD	外显子 23	-	aggccaaacctcggcttac	NNGRR
DCR41 (105)	小鼠	DMD	外显子 23	+	TTCGAAAATTTCAGgtaag	NNGRR
DCR42 (106)	小鼠	DMD	外显子 23	+	gcagaacaggagataacag	NNGRRT
DCR43 (107)	小鼠	ACVR2B	外显子 1	+	gcggccctcgcccttctct	ggggat
DCR48 (108)	人	DMD	内含子 45	-	TAGTGATCGTGGATACGAG	AGG
DCR49 (109)	人	DMD	内含子 45	-	TACAGCCCTCGGTGTATAT	TGG
DCR50 (110)	人	DMD	内含子 52	-	GGAAGGAATTAAGCCCGAA	TGG
DCR51 (111)	人	DMD	内含子 53	-	GGAACAGCTTTCGTAGTTG	AGG
DCR52 (112)	人	DMD	内含子 54	+	ATAAAGTCCAGTGTCGATC	AGG
DCR53 (113)			内含子 54	+	AAAACCAGAGCTTCGGTCA	AGG
DCR54 (114)	小鼠	Rosa26	ZFN 区	+	GAGTCTTCTGGGCAGGCTTAAAGGCTAACC	TGG
DCR55	小鼠	Rosa26	mRNA	-	TCGGGTGAGCATGTCTTTAATCTACCTCGA	TGG

[0393]

名 称 (SEQ ID NO)	物种	基因	靶标	链	19bp 用于嵌合(5'端加入 G)	PAM
(115)						
DCR49 (116)	人	DMD	Ex 51	-	gtgtcaccagagtaacagt	ctgagt
DCR50 (117)	人	DMD	Ex 51	+	tgatcatcaagcagaaggt	atgag
DCR60 (118)	小鼠	DMD	外显子 23	+	AACTTCGAAAATTTCAG gta	agccgagg
DCR61 (119)	小鼠	DMD	内含子 22	+	gaaactcatcaaatatgcgt	gttagtgt
DCR62 (120)	小鼠	DMD	内含子 22	-	tcatttacactaacacgcat	atttgatg
DCR63 (121)	小鼠	DMD	内含子 22	+	gaatgaaactcatcaaatat	gcgtgtta
DCR64 (122)	小鼠	DMD	内含子 23	-	tcatcaatatctttgaagga	ctctgggt
DCR65 (123)	小鼠	DMD	内含子 23	-	tgttttcataggaaaaatag	gcaagttg
DCR66 (124)	小鼠	DMD	内含子 23	+	aattggaaaatgtgatggga	aacagata
DCR67 (125)	人	DMD	外显子 51	+	atgatcatcaagcagaaggt	atgagaaa
DCR68 (126)	人	DMD	外显子 51	+	agatgatcatcaagcagaag	gtatgaga
DCR69	人	DMD	外显子 51	-	catttttctcataccttct	gcttgatg

[0394]

名 称 (SEQ ID NO)	物种	基因	靶标	链	19bp 用于嵌合(5'端加入 G)	PAM
(127)						
DCR70 (128)	人	DMD	外显子 51	+	tcctactcagactgttactc	tggtgaca
DCR71 (129)	人	DMD	外显子 51	-	acaggttgtgtcaccagagt	aacagtct
DCR72 (130)	人	DMD	外显子 51	-	ttatcatttttttctcatacc	ttctgctt
DCR73 (131)	人	DMD	内含子 51	-	ttgcctaagaactggtggga	aatggtct
DCR74 (132)	人	DMD	内含子 51	-	aaacagttgcctaagaactg	gtgggaaa
DCR75 (133)	人	DMD	内含子 51	+	tttcccaccagttcttaggc	aactgttt
DCR76 (134)	人	DMD	内含子 50	+	tggctttgatttccctaggg	tccagctt
DCR77 (135)	人	DMD	内含子 50	-	tagggaaatcaaagccaatg	aaacgttc
DCR78 (136)	人	DMD	内含子 50	-	gaccctagggaaatcaaagc	caatgaaa
DCR79 (137)	人	DMD	内含子 44	-	TGAGGGCTCCACCCTCACGA	GTGGGTTT
DCR80 (138)	人	DMD	内含子 44	-	AAGGATTGAGGGCTCCACCC	TCACGAGT
DCR81	人	DMD	内含子 44	-	GCTCCACCCTCACGAGTGGG	TTTGGTTC

[0395]

名 称 (SEQ ID NO)	物种	基因	靶标	链	19bp 用于嵌合(5'端加入 G)	PAM
(139)						
DCR82 (140)	人	DMD	内含子 55	-	TATCCCCTATCGAGGAAACC	ACGAGTTT
DCR83 (141)	人	DMD	内含子 55	+	GATAAAGAAGGCCTATTTCA	TAGAGTTG
DCR84 (142)	人	DMD	内含子 55	-	AGGCCTTCTTTATCCCCTAT	CGAGGAAA
DCR85 (143)	人	DMD	内含子 44	-	TGAGGGCTCCACCCTCACGA	GTGGGT
DCR86 (144)	人	DMD	内含子 55	+	GATAAAGAAGGCCTATTTCA	TAGAGT

[0396] 表7

[0397]

名称	注释	% Mod
DCR1	缺失外显子51	6.6
DCR2	缺失外显子51	10.3
DCR3	移码	13
DCR4	缺失外显子51	11.9
DCR5	缺失外显子51	12.4
DCR6	在内含子44中尽可能接近外显子44 (在患者缺失的情况下)	16.1
DCR7	在内含子55内尽可能接近外显子56 (在患者缺失的情况下)	6.8
DCR8	只可纠正外显子42-43缺失(-1/+2),(-2/+1)不可被此纠正	17.3
DCR9	跳跃外显子44 (5')	14.4
DCR10	移码	14.9
DCR11	纠正外显子45的下游	<1
DCR12	5'剪接受体/移码	<1
DCR13	纠正外显子46的下游	16.9
DCR14	移码	17.2
DCR15	纠正外显子47的下游	15.4
DCR16	移码	11.5
DCR17	纠正外显子48的下游	<1
DCR18	5'剪接受体/移码	1.8
DCR19	纠正外显子49的下游	33.7

DCR20	5’剪接受体	14.9
DCR21	纠正外显子50的下游	24.1
DCR22	移码	25.9
DCR23	纠正外显子52的下游	25.2
DCR24	移码(只可纠正+1读框)	24.8
DCR25	纠正外显子53的下游	2.6
DCR26	移码	24.5
DCR27	纠正外显子54的下游	13.4
DCR28	5’剪接受体	21.6
DCR29	纠正外显子55的下游	19.2
DCR30	在外显子1中整合微小肌养蛋白	未检测
DCR31	纠正外显子51的下游	18.9
DCR32	缺失终止密码子	未检测
DCR33	CR6的备选	1.3
DCR34	CR6的备选	13.2
DCR35	CR7的备选	22.5
DCR36	CR7的备选	26.4
DCR40	中断外显子23 5’剪接供体(纠正mdx突变)	
DCR41	中断外显子23 5’剪接供体(纠正mdx突变)	
DCR42	缺失外显子53 mdx4cv突变	
DCR43	中断肌肉生长抑制素(myostatin)受体	

表 8

[0398]

名称	Cas9		注释	使用的 Cas9
DCR49	金黄色葡萄球菌		外显子 51 中的移码	SaCas9 (来自 Zhang pX441) (NNGRRT PAM)
DCR50	金黄色葡萄球菌		中断外显子 51 的 5'端	SaCas9 (来自 Zhang pX441) (NNGRR PAM)
DCR60	脑膜炎奈瑟球菌	NNNNGANN	靶向外显子 23 的 3'剪接供体以绕过 mdx 突变	NmCas9 (NNNNGANN PAM)
DCR61	脑膜炎奈瑟球菌	NNNNGNNT	缺失外显子 23 并绕过 mdx 突变	NmCas9 (NNNNGNNT PAM)
DCR62	脑膜炎奈瑟球菌	NNNNGANN	缺失外显子 23 并绕过 mdx 突变	NmCas9 (NNNNGANN PAM)
DCR63	脑膜炎奈瑟球菌	NNNNGTTN	缺失外显子 23 并绕过 mdx 突变	NmCas9 (NNNNGTTN PAM)
DCR64	脑膜炎奈瑟球菌	NNNNGNNT	缺失外显子 23 并绕过 mdx 突变	NmCas9 (NNNNGNNT PAM)
DCR65	脑膜炎奈瑟球菌	NNNNGTTN	缺失外显子 23 并绕过 mdx 突变	NmCas9 (NNNNGTTN PAM)
DCR66	脑膜炎奈瑟球菌	NNNNGANN	缺失外显子 23 并绕过 mdx 突变	NmCas9 (NNNNGANN PAM)
DCR67	脑膜炎奈瑟球菌	NNNNGANN	靶向外显子 51 的 3'剪接供体以跳过外显子	NmCas9 (NNNNGANN PAM)

[0399]

DCR68	脑膜炎奈瑟球菌	NNNNGANN	靶向外显子 51 的 3'剪接供体以跳过外显子	NmCas9 (NNNNGANN PAM)
DCR69	脑膜炎奈瑟球菌	NNNNGANN	靶向外显子 51 的 3'剪接供体以跳过外显子	NmCas9 (NNNNGANN PAM)
DCR70	脑膜炎奈瑟球菌	NNNNGANN	外显子 51 中的移码	NmCas9 (NNNNGANN PAM)
DCR71	脑膜炎奈瑟球菌	NNNNGNNT	外显子 51 中的移码	NmCas9 (NNNNGNNT PAM)
DCR72	脑膜炎奈瑟球菌	NNNNGNNT	靶向外显子 51 的 3'剪接供体以跳过外显子	NmCas9 (NNNNGNNT PAM)
DCR73	脑膜炎奈瑟球菌	NNNNGNNT	缺失外显子 51（与 DCR5 尽可能近地结合）	NmCas9 (NNNNGNNT PAM)
DCR74	脑膜炎奈瑟球菌	NNNNGANN	缺失外显子 51（与 DCR5 尽可能近地结合）	NmCas9 (NNNNGANN PAM)
DCR75	脑膜炎奈瑟球菌	NNNNGTTN	缺失外显子 51（与 DCR5 尽可能近地结合）	NmCas9 (NNNNGTTN PAM)
DCR76	脑膜炎奈瑟球菌	NNNNGNNT	缺失外显子 51（与 DCR1/2 尽可能近地结合）	NmCas9 (NNNNGNNT PAM)
DCR77	脑膜炎奈瑟球菌	NNNNGTTN	缺失外显子 51（与 DCR1/2 尽可能近地结合）	NmCas9 (NNNNGTTN PAM)
DCR78	脑膜炎奈瑟球	NNNNGANN	缺失外显子 51（与 DCR1/2 尽可能近地结	NmCas9 (NNNNGANN PAM)

[0400]

	菌		合)	
DCR79	脑膜炎奈瑟球菌	NNNNGNNT	缺失外显子45-55以叠盖NNNNGTTN PAM，与DCR6尽可能近地结合	NmCas9 (NNNNGNNT PAM)
DCR80	脑膜炎奈瑟球菌	NNNNGANN	缺失外显子45-55，与DCR6尽可能近地结合	NmCas9 (NNNNGANN PAM)
DCR81	脑膜炎奈瑟球菌	NNNNGTTN	缺失外显子45-55，与DCR6尽可能近地结合	NmCas9 (NNNNGTTN PAM)
DCR82	脑膜炎奈瑟球菌	NNNNGNNT	缺失外显子45-55，与DCR36尽可能近地结合以叠盖NNNNGTTN PAM	NmCas9 (NNNNGNNT PAM)
DCR83	脑膜炎奈瑟球菌	NNNNGTTN	缺失外显子45-55，与DCR36尽可能近地结合	NmCas9 (NNNNGTTN PAM)
DCR84	脑膜炎奈瑟球菌	NNNNGANN	缺失外显子45-55，与DCR36尽可能近地结合	NmCas9 (NNNNGANN PAM)
DCR85	金黄色葡萄球菌	NNGRRT	缺失外显子45-55，与DCR6尽可能近地结合	SaCas9（来自Zhang pX441） (NNGRRT PAM)
DCR86	金黄色葡萄球菌	NNGRRT	缺失外显子45-55，与DCR36尽可能近地结合	SaCas9（来自Zhang pX441） (NNGRRT PAM)

[0401] 具体地说,将400 ng Cas9与400 ng空载体或靶向包括外显子51的区域的gRNA即CR1、CR2、CR3、CR4和CR5一起共转染入HEK 293T细胞中(参见图11(b))。在转染后2天收获基因组DNA,采用Surveyor测定法分析(参见图11(a)和11(c))。

[0402] 将基于CRISPR/Cas9的系统用于DMD 8036 (del48-50)细胞中以测定系统是否可修复突变型肌养蛋白基因。将5 μg Cas9与7.5 μg空载体或gRNA一起共转染至DMD 8036 (del48-50)细胞中。具体地说,使用7.5 μg CR1 (“DCR1”)、7.5 μg CR5 (“DCR5”)、15 μg CR3 (“DCR3”)或7.5 μg CR1和CR5的组合(DCR1+DCR5)。在转染后3天收获基因组DNA,采用跨越整个基因座的Surveyor测定法(图12)或PCR分析(图13)进行分析。使用侧接含有针对

CR1和CR5的基因组靶的区域的引物(正向引物:5'-gagaggttatgtggctttacca (SEQ ID NO:457),反向引物:5'-ctgcgtagtgccaaaacaaa (SEQ ID NO:458)),通过PCR扩增该基因座,得到野生型基因座的1447bp条带或缺失基因座约630bp的预期大小。在分化7天后,处理细胞的蛋白质印迹显示表达肌养蛋白(参见图14)。

[0403] 实施例5

[0404] 将CRISPR/Cas9靶向人肌养蛋白基因中的热点突变

[0405] 为了利用CRISPR/Cas9基因编辑平台以纠正广泛的肌养蛋白突变,产生靶向显子45-55间的热点突变区12种sgRNA (图16)。使用利用人密码子优化的SpCas9核酸酶和嵌合单一指导RNA(sgRNA)表达载体以指导有效的位点特异性基因编辑的酿脓链球菌系统。与实施例4用TALEN靶向外显子51相似,选择靶向符合SpCas9的5'-NRG-3' PAM要求的外显子45-55的5'和3'端的前间区。通过这些外显子内的基于NHEJ的DNA修复所产生的少量插入或缺失可产生克服各外显子周围各种肌养蛋白突变的靶定移码突变(图16A-16B)。例如,设计CR3,通过在外显子51的5'端引入少量插入或缺失以恢复下游肌养蛋白读框,来纠正外显子51周围的肌养蛋白突变或缺失(图16B)。另外,设计sgRNA以利用CRISPR/Cas9系统的多重能力,并特异缺失个别外显子或系列外显子以恢复肌养蛋白读框,与基于寡核苷酸的外显子跳跃的方法类似。为此目的,将sgRNA靶向外显子51 (图16C)或外显子45-55 (图16D)周围的内含子区。使这些sgRNA有意靶向最接近欲包括在所得转录物中的外显子下游或上游的位点,以使背景患者缺失将包括内含子sgRNA靶位点的可能性降到最小。

[0406] 实施例6

[0407] 靶向人细胞中的肌养蛋白基因的sgRNA的筛选

[0408] 对人HEK293T细胞系中的基因编辑频率进行了评价以快速测定不同的sgRNA打靶效率。将HEK293T用编码人密码子优化SpCas9和指定sgRNA的构建体转染。按规定,设计各sgRNA以修饰肌养蛋白基因。在转染后第3天或第10天,通过Surveyor测定法测定基因修饰的频率。计算在第3天和第10天所测Surveyor信号的比率以量化人细胞中各sgRNA的基因编辑频率的稳定性。如转染后3通过天Surveyor测定法测定的,29/32 (约90%)的所测sgRNA能够在预期的基因座中介导高效的基因修饰(表9,图17)。基因编辑频率对几乎所有的sgRNA都稳定(从第3天到第10天的信号变化<25%,表9,图18),表明了由每个个别sgRNA介导的基因编辑是极耐受的。一个明显的例外是CR33,其在第10天没可检测的活性,虽然活性可低于Surveyor测定法的灵敏度(估计约1%)。

表 9 测量的人细胞中 sgRNA 的活性

[0409]

靶标	sgRNA #	在第 3 天%修饰的等位基因	在第 10 天%修饰的等位基因	%变化 第 10 天/第 3 天
外显子 51 的多个缺失				
Int 50	CR1	6.6	9.3	41.8
Int 50	CR2	10.3	14.0	36.2
Ex 51	CR4	11.9	14.4	21.3
Int 51	CR5	12.4	13.3	7.8
外显子 45-55 的多个缺失				
Int 44	CR6	16.1	16.9	4.3
Int 44	CR33	1.3	<1	n.d.
Int 44	CR34	13.2	11.0	-16.6
Int 55	CR7	6.8	7.1	5.3
Int 55	CR35	22.5	20.9	-7.1
Int 55	CR36	26.4	24.7	-6.4

[0410]

靶定移码				
Ex 45	CR10	14.9	16.3	9.3
Ex 45	CR11	<1	<1	n.d.
Ex 46	CR12	<1	<1	n.d.
Ex 46	CR13	16.9	18.4	9.2
Ex 47	CR14	17.2	17.6	2.9
Ex 47	CR15	15.4	15.3	-0.9
Ex 48	CR16	11.5	10.9	-5.0
Ex 48	CR17	<1	<1	n.d.
Ex 49	CR18	1.8	2.2	20.1
Ex 49	CR19	33.7	38.4	13.9
Ex 50	CR20	14.9	13.7	-7.6
Ex 50	CR21	24.1	20.8	-13.5
Ex 51	CR3	13.0	16.7	28.0
Ex 51	CR31	18.9	16.9	-10.2
Ex 52	CR22	25.9	20.3	-21.6
Ex 52	CR23	25.2	24.0	-4.8
Ex 53	CR24	24.8	23.6	-4.6
Ex 53	CR25	2.6	2.9	9.5
Ex 54	CR26	24.5	22.0	-10.1
Ex 54	CR27	13.4	12.6	-5.9
Ex 55	CR28	21.6	19.8	-8.4
Ex 55	CR29	19.2	19.6	2.2

[0411] 实施例7

[0412] 使用基于荧光的报道分子系统富集基因编辑的细胞

[0413] 选择sgRNA以纠正DMD患者成肌细胞系的特定突变。在转染至DMD成肌细胞中后,如通过Surveyor测定法测量的,观察到未意料到的低的或未检出的基因修饰活性(图19C,整群)。采用流式细胞术,通过与SpCas9蛋白连接的2A核糖体跳跃肽选择共表达GFP的转染的细胞(图19A)。将该荧光报道分子加入SpCas9表达载体中似乎不会显著影响HEK293T细胞中的基因编辑活性(图19B)。在电穿孔后3天,低百分比的转染成肌细胞(约0.5-2%)表达荧光报道分子,尽管对照GFP表达质粒的高转染效率(通常>70%,图19D,pmaxGFP)。鉴于在容易转染的HEK293T系中高水平的CRISPR/Cas9活性,在将SpCas9-T2A-GFP和sgRNA构建体电穿孔至DMD细胞中后低效的转基因表达可解释未分选细胞中低的实测基因编辑效率。在分选GFP阳性DMD成肌细胞后,在大部分sgRNA靶基因座上,观察到可检测活性的大幅增加(图19C)。因此,所有随后的实验使用通过该荧光报道分子表达针对SpCas9表达分选的细胞。

[0414] 实施例8

[0415] 通过靶定的移码修复肌养蛋白表达

[0416] 通过NHEJ DNA修复产生的少量插入和缺失可用来产生靶定移码以纠正异常读框。设计sgRNA CR3,以通过在外显子51内引入少量插入和缺失恢复肌养蛋白读框(图16B、20A)。来自用SpCas9和CR3 sgRNA的表达质粒共转染的HEK293T细胞的基因组DNA的等位基因,通过Sanger测序评价在该基因座上通过CRISPR/Cas9产生的插入和缺失的类型(图20B)。显然,插入和缺失导致转换成所有3个读框(图20B、20C)。为了证实相关患者细胞系中的遗传纠正,将SpCas9和CR3 sgRNA的表达质粒电穿孔至具有通过在外显子51中产生移码可纠正的外显子48-50的缺失的DMD成肌细胞系中。处理的细胞经分选,通过Surveyor测定法证实具有基因修饰活性(CR3,图19C分选群),并分化成肌管以测试恢复的肌养蛋白表达。观察到肌养蛋白表达伴以可检测的核酸酶活性(图20D)。酿脓链球菌CRISPR/Cas9系统提供快速产生靶定移码以克服多种患者突变并恢复人肌养蛋白基因表达的有力方法。

[0417] 实施例9

[0418] 多重CRISPR/Cas9基因编辑介导外显子51的基因组缺失并拯救肌养蛋白表达

[0419] CRISPR/Cas9系统的多重能力提供有效产生用于靶定基因纠正的特定外显子的基因组缺失的新的方法。如图19中,将具有通过外显子51跳跃可纠正的背景缺失的DMD患者成肌细胞用侧接外显子51的sgRNA的两种组合(CR1/CR5或CR2/CR5)处理,并分选以富集基因编辑的细胞。如这些处理细胞的基因组DNA的终点PCR所检测的,当将两种sgRNA电穿孔至具有SpCas9的细胞时,预期的基因组缺失才存在(图21A)。对于两种缺失,Sanger测序证实远端染色体区段的预期连接处(图21B)。在区分出分选的成肌细胞后,只在用两种sgRNA处理的细胞中,观察到来自mRNA转录物的外显子51的缺失(图21C)。最后,在处理细胞中检测到恢复的肌养蛋白表达伴以外显子51的实测基因组水平和mRNA水平缺失(图21D)。

[0420] 实施例10

[0421] 通过多外显子大基因组缺失拯救肌养蛋白

[0422] 虽然克服患者特异性突变是CRISPR/Cas9系统的重大用途,但是开发可克服多种常见的患者缺失的单一方法将是有利的。例如,一种有前景的策略是排除整个外显子45-55区作为纠正高达62%的已知患者缺失的方法。测试基于多重CRISPR/Cas9的基因编辑以测定它是否能够在人细胞中产生外显子45-55基因座的有效缺失。在转染至HEK293T细胞中后,通过基因组DNA的PCR检出约336,000 bp的预期缺失(图22A)。同样,通过来自SpCas9/sgRNA处理的DMD患者细胞的基因组DNA的PCR检出该缺失,所述细胞带有外显子48-50的未知长度的背景缺失(图22A)。来自经处理的DMD细胞的基因组DNA的该缺失条带的Sanger测序显示内含子44和内含子55的预期连接处与sgRNA靶位点直接邻接(图22B)。在处理的DMD细胞分化后,在肌养蛋白mRNA转录物中检出外显子45-55的预期缺失,通过Sanger测序证实是外显子44和56的融合物(图22C)。通过蛋白质印迹,在含有CRISPR/Cas9诱导的来自基因组外显子45-55的缺失和所得mRNA转录物的分选细胞群中观察到恢复的蛋白质表达(图22D)。这些数据表明,多重CRISPR/Cas9编辑在超过60%的DMD患者突变中提供恢复肌养蛋白读框的单一通用方法。

[0423] 实施例11

[0424] 将纠正的成肌细胞移植到免疫缺陷型小鼠中

[0425] 对于DMD疗法的一种有前景的方法是纠正可植入患者骨骼肌组织以拯救肌养蛋白表达的自体患者肌肉祖细胞群。为了证实纠正的细胞体内表达人肌养蛋白的能力,如前所

述,移植用侧接外显子51的sgRNA CR1和CR5处理的DMD成肌细胞群,并针对GFP的表达进行分选(图19、图23)。4周后,在注射的肌肉组织的冰冻切片中,观察到通过纠正和未纠正的细胞两者表达的对人血影蛋白呈阳性的肌纤维(图24)。这些纤维的许多也对人肌养蛋白呈阳性,其表达定位于肌膜,表明了这些细胞中的功能性基因纠正(图24,图25)。在来自注射未处理DMD成肌细胞的小鼠的切片中,未观察到对人肌养蛋白呈阳性的纤维(图24、图25),表明了CRISPR/Cas9修饰的细胞是人肌养蛋白表达的来源。

[0426] 实施例12

[0427] 脱靶和细胞毒性分析

[0428] 如前所述,通过改编基于流式细胞术的GFP保留测定法,针对精选的sgRNA,评价了人细胞中CRISPR/Cas9系统的相对细胞毒性(Ousterout等, Mol Ther 21:1718-1726(2013))。在转染到人细胞中后,针对有或没有sgRNA的SpCas9共表达,观察到最小细胞毒性(图26A)。可根据sgRNA前间区序列中指定错配的预测位置偏倚和预期靶位点错配的总数,获得可公开获取的工具以评价和优选列出脱靶基因座上潜在的CRISPR/Cas9活性(Hsu等, Nat Biotechnol 31:827-826 (2013))。利用这种公众网页服务器预测用于纠正该研究中的肌养蛋白基因的sgRNA的最可能的脱靶位点(表4)。通过Surveyor测定法评价用SpCas9和CR1、CR3、CR5、CR6或CR36的各sgRNA表达盒处理的HEK293T细胞中最前面的10个潜在的脱靶位点。各自具有这10个预测的脱靶基因座之一的CR1、CR3和CR36表明基因修饰的显著水平(表4和图27)。引人关注的是,CR3脱靶序列具有与预期中靶基本相同和相似的修饰频率(在OT-1上的9.3%相对于在预期位点上的13.3%) (表4和图27)。显然,CR3-OT1是通过Surveyor测定法在分选hDMD细胞中显示显著活性水平的这3个脱靶位点中的唯一一个(图26B)。

[0429] 通过在不同染色体上在切割的靶标和脱靶基因座之间的远端再连接,在脱靶位点上的核酸酶活性可能引起非预期的染色体重排。由于使用两种或更多种核酸酶(例如在多重CRISPR/Cas9基因编辑中)引起脱靶活性的潜力增加所致,这对基于缺失的基因纠正策略提出重大顾虑。在单一和多重CRISPR/Cas9编辑策略两者中,使用高灵敏嵌套基因组PCR测定法探测潜在易位,以检测经证实的脱靶基因座上的易位(表4)。采用该测定法,在也显示高水平的脱靶活性的模式HEK293T细胞系中,在中靶和脱靶位点之间容易检出易位(图26C和图28A、28B)。PCR扩增子的Sanger测序证实各引物对的预测易位事件的特性(图29-30)。在HEK293T细胞中检出的一类易位也可在分选的hDMD成肌细胞通过嵌套PCR检出,虽然信号较弱得多,且未证实序列同一性由产物产量低所致(图26D和图28A、28C)。分别在HEK293T或用CR6或CR6/CR36处理的分选hDMD细胞中,采用该测定法未检出易位(图28),只在HEK293T细胞中在CR6-OT3处具有低水平的脱靶活性(表4)。这些结果强调选择高特异性sgRNA特别对于多重编辑应用的重要性,并且表明该方法可获益于不断的努力以改进CRISPR/Cas9系统的特异性。这些数据表明,所选的sgRNA能够纠正肌养蛋白基因而无重大毒性且仅单一强有力预测的具有可检出的活性水平的脱靶位点。

[0430] 实施例13

[0431] 论述

[0432] 基因组编辑是用于纠正遗传疾病的有力工具,CRISPR/Cas9系统的最新发展显著加快该领域的进展。表明了DMD,一种目前还没有经批准的治疗选择方法的最常见的遗传疾病得到纠正。DMD的许多基于基因和基于细胞的疗法处于临床前研究和临床试验中,而且基

因组编辑方法与许多这类方法相容。例如,基因组编辑可与DMD的患者特异性基于细胞的疗法联合。正如所示,CRISPR/Cas9系统可在人多能干细胞和其它人细胞系以及人骨骼成肌细胞中起作用。重要的是,用CRISPR/Cas9的基因编辑不会完全破坏这些细胞的生肌能力,正如在移植到免疫缺陷型小鼠中体外和体内有效的肌养蛋白表达所证实的一样。因此,这种策略应与DMD的基于细胞的疗法相容。

[0433] 另外,基因纠正的细胞的富集库显示在植入免疫缺陷型小鼠后体内人肌养蛋白的表达。如通过稳定的基因编辑频率和几种sgRNA的最低细胞毒性所观察的,CRISPR/Cas9基因编辑在人成肌细胞中没有明显的毒性作用。然而,证实了5种sgRNA间在50个预测的脱靶位点中有3个有基因编辑活性,且在中靶和脱靶位点之间可检出基因编辑活性。CRISPR/Cas9技术是用于纠正大部分肌养蛋白突变的有效和多用途的方法,可用作治疗遗传疾病的通用平台。

[0434] 另外,与本文所用的基于质粒的递送方法形成对比,sgRNA和Cas9 mRNA的直接转染可用来通过降低Cas9表达的持续时间并消除随机质粒整合的可能性,提高特异性和完全性。或者,对于该方法的体内基因组编辑和翻译,可使用通过病毒、质粒或RNA递送载体将CRISPR/Cas9系统直接递送至骨骼和/或心肌。大尺寸的酿脓链球菌Cas9基因(约4.2千碱基)对其在大小受限制的腺伴随病毒载体中的用途提出了挑战。然而,来自其它菌种(例如脑膜炎奈瑟球菌和嗜热链球菌(*S. thermophilus*)的Cas9基因短到足以将Cas9和sgRNA表达盒两者有效包装入单一AAV载体中用于体内基因编辑应用。

[0435] CRISPR/Cas9系统使有效修饰几乎90%的受测靶标成为可能,与不同基因座上该系统的稳健活性的其它报告一致。该技术的稳健性和多用途对于任意创建患者特异性基因编辑是一个重大进展。低水平的肌养蛋白,包括少至4%的野生型表达,可足以提高小鼠模型的存活、运动功能和心脏功能。CRISPR/Cas9活性的水平对于治疗益处可能是足够的。

[0436] 与CRISPR/Cas93复用的缺失外显子的用途还提供一类独特的机会和挑战。进行了从基因组缺失完整外显子以恢复肌养蛋白表达,与用在单一核酸酶作用后由基于NHEJ的DNA修复产生少量得失位恢复肌养蛋白基因的读框形成对比。经编辑的基因的蛋白质产物是可预测的,已在有天然存在的缺失的贝克肌营养不良患者中表征,与通过单一核酸酶的外显子内作用所创建的随机得失位(将导致产生来自各DNA修复事件的新的表位)形成对照。此外,对于每个成功的基因编辑事件,产生于外显子缺失的产物将导致恢复的肌养蛋白,而用外显子内的随机得失位修饰基因将只恢复导致纠正读框的编辑事件的三分之一的读框。

[0437] 所有所测的sgRNA与人细胞中的重大细胞毒性作用无关。对于所用的5种sgRNA,自50个总的受测位点中鉴定出恢复肌养蛋白表达的3个潜在脱靶位点。此外,在表达高水平的Cas9和sgRNA的HEK293T细胞中,通过高灵敏嵌套PCR测定法,可检出预期的中靶位点和这些脱靶位点之间的染色体易位。显然,在HEK293T细胞(其是一种表达极高水平的Cas9和sgRNA的永生化和非整倍的细胞系)中鉴定的脱靶活性和易位不以高水平发生,并且在某些情况下在hDMD成肌细胞中无法检出。重要的是,鉴于DMD的严重性,这种特异性水平可以是耐受的,在人细胞中缺乏明显的细胞毒性作用。

[0438] 实施例14

[0439] 开发了工程改造的AAV衣壳,称为SASTG (图34;SEQ ID NO: 436和437)用于改进

的心肌和骨骼肌组织嗜性(Piacentino等(2012) Human Gene Therapy 23:635-646)。靶向Rosa26基因座的ZFN (“Rosa26 ZFN”;图33;SEQ ID NO: 434和435)显示在小鼠细胞中是有高度活性的(Perez-Pinera等, Nucleic Acids Research (2012) 40:3741-3752)。设计编码Rosa26 ZFN蛋白的AAV-SASTG载体,随后通过UNC Viral Vector Core产生并纯化。采用Surveyor测定法(Guschin等, *Methods Mol Biol* 649, 247-256 (2010))以表明在培养的C2C12成肌细胞中递送AAV-SASTG Rosa26 ZFN后在Rosa26基因座上的NHEJ诱变,所述细胞主动循环,或通过除去血清被强制分化(未显示)。

[0440] 为了证实在AAV转移后成熟有丝分裂后骨骼肌被Rosa26 ZFN有效靶中,将编码Rosa26 ZFN的AAV-SASTG载体以1e10载体基因组(vg)或2.5e10 vg/肌肉的滴度直接注射到6周龄C57BL6/J小鼠的胫骨前肌(TA)肌肉中。在注射后4周处死小鼠,收获TA肌肉,分割成几段用于基因组DNA提取和分析(图31)。对基因组DNA进行PCR扩增,并进行Surveyor测定法以检测Rosa26靶位点上ZFN诱变的NHEJ突变特性(图32)。图32显示在递送AAV-SASTG-ROSA后Surveyor分析的体外和体内骨骼肌中的Rosa26 ZFN活性。增殖性C2C12用指定量的病毒转导,在感染后4天收获(图32(a))。使C2C12在分化培养基孵育5天,然后在24孔板中用指定量的AAV-SASTG-ROSA病毒转导(图32(b))。在转导后10天收集样品。将指定量的AAV-SASTG-ROSA直接注射到C57BL/6J小鼠的胫骨前肌,并在感染后4周收获肌肉。将收获的肌肉分割成8个单独的块用于基因组DNA分析,其各自在单独的泳道中显示(图32(c))。显然,在最高剂量(2.5e10 vg)下,在所有节段中检出高水平的基因修饰。

[0441] 实施例15

[0442] 靶向突变型肌养蛋白基因的AAV-CRISPR构建体

[0443] 设计AAV构建体以治疗性地纠正引起杜兴型肌营养不良及骨骼和心肌变性的肌养蛋白基因的突变。CRISPR/Cas9系统可使用AAV递送以通过缺失外显子51、缺失外显子45-55、中断剪接供体或受体位点或在外显子51内产生移码(Ousterout等, Molecular Therapy 2013)恢复肌养蛋白读框以恢复肌养蛋白读框和蛋白质表达。CRISPR/Cas9系统可包括具有SEQ ID NO: 64或114的序列的Cas9 (参见图40和41)。提供可与这些Cas9组合靶向其各自的PAM序列的gRNA (参见图40和41;另参见表2和3)。

[0444] 实施例16

[0445] 诱导神经元(iN)的产生

[0446] 从其它细胞谱系中产生诱导神经元(iN)在神经疾病的再生医学和研究中具有潜在应用。小鼠胚胎成纤维细胞(MEF)直接转换成功能性神经元细胞可通过递送3种神经元转录因子-*BRN2*、*ASCL1*和*MYT1L*的混合物(BAM因子,图48)而发生。其它方法可包括诱导不同神经元亚型的其它因子。这些实验需要异位递送的转录因子,并激活相应的内源基因座以保持神经元表型。将CRISPR/Cas9系统改造成多用途转录因子以激活具有通过RNA指导的机制靶向基因组中的任何启动子的哺乳动物细胞的内源基因(图49A、49B)。

[0447] 材料与方法。使用CRISPR/Cas9-转录因子激活编码*ASCL1*和*BRN2*的内源基因以指导MEF重编程成功能性诱导神经元。

[0448] 细胞培养:将MEF接种到24孔TCPS板上或聚D-赖氨酸/层粘连蛋白包被的盖玻片上。在dCas9-VP64转导和gRNA转染后(对于gRNA的序列参见表10和11),将细胞在MEF培养基(Adler等, Mol Ther Nucleic Acids 1:e32 (2012)中培养24小时,然后转移到N3神经诱

导培养基(Vierbuchen等, *Nature* 463:1035-1041 (2010))中持续实验的持续时间(图49B)。

[0449]

表 10 用于小鼠 ASCL1 的 gRNA (CR13):

寡核苷酸 (5'-3')	5'突出端		ASCL1 靶序列		SEQ ID NO
CR13-1_S:	cacc	G	CAGCCGCTCGCTGCAGCAG (SEQ ID NO: 468)		492
CR13-1_AS:	AAAC		CTGCTGCAGCGAGCGGCTG (SEQ ID NO: 469)	C	493
CR13-2_S:	cacc	G	GCTGGGTGTCCCATTGAAA (SEQ ID NO: 470)		494
CR13-2_AS:	AAAC		TTTCAATGGGACACCCAGC (SEQ ID NO: 471)	C	495
CR13-3_S:	cacc	G	GTTTATTCAGCCGGGAGTC (SEQ ID NO: 472)		496
CR13-3_AS:	AAAC		GACTCCCGGCTGAATAAAC (SEQ ID NO: 473)	C	497
CR13-4_S:	cacc	G	TGGAGAGTTTGCAAGGAGC (SEQ ID NO: 474)		498
CR13-4_AS:	AAAC		GCTCCTTGCAAACTCTCCA (SEQ ID NO: 475)	C	499
CR13-5_S:	cacc	G	CCCTCCAGACTTTCCACCT (SEQ ID NO: 476)		500
CR13-5_AS:	AAAC		AGGTGGAAAGTCTGGAGGG (SEQ ID NO: 477)	C	501

CR13-6_S:	cacc	G	AATTTTCTTCCAAGTTCTC (SEQ ID NO: 478)		502
CR13-6_AS:	AAAC		GAGAACTTGGAAGAAAATT (SEQ ID NO: 479)	C	503
CR13-7_S:	cacc	G	CTGCGGAGAGAAGAAAGGG (SEQ ID NO: 480)		504
CR13-7_AS:	AAAC		CCCTTTCTTCTCTCCGCAG (SEQ ID NO: 481)	C	505
CR13-8_S:	cacc	G	AGAGCCACCCCCTGGCTCC (SEQ ID NO: 482)		506
CR13-8_AS:	AAAC		GGAGCCAGGGGGTGGCTCT (SEQ ID NO: 483)	C	507
CR13-9_S:	cacc	G	cgaagccaaccgcggcggg (SEQ ID NO: 484)		508
CR13-9_AS:	AAAC		cccgccgcggttggcttcg (SEQ ID NO: 485)	C	509
CR13-10_S:	cacc	G	agagggaagacgatcgccc (SEQ ID NO: 486)		510
CR13-10_AS:	AAAC		gggcgatcgtcttccctct (SEQ ID NO: 487)	C	511
CR13-11_S:	cacc	G	ccccttttaactttcctccg (SEQ ID NO: 488)		512
CR13-11_AS:	AAAC		cggaggaaagttaaagggg (SEQ ID NO: 489)	C	513
CR13-12_S:	cacc	G	gcagccccgcttccttcaa (SEQ ID NO: 490)		514
CR13-12_AS:	AAAC		ttgaaggaagcggggctgc (SEQ ID NO: 491)	C	515

表 11 用于小鼠 BRN2 的 gRNA (CR16):

[0450]

寡核苷酸 (5'-3')	**5'突出端**		**BRN2 靶序列**		**SEQ ID NO**
CR16-1_S:	cacc	G	CGAGAGCGAGAGGAGGGAG (SEQ ID NO: 516)		540
CR16-1_AS:	AAAC		CTCCCTCCTCTCGCTCTCG (SEQ ID NO: 517)	C	541
CR16-2_S:	cacc	G	GAGAGAGCTTGAGAGCGCG (SEQ ID NO: 518)		542
CR16-2_AS:	AAAC		CGCGCTCTCAAGCTCTCTC (SEQ ID NO: 519)	C	543
CR16-3_S:	cacc	G	GGTGGAGGGGGCGGGGCCC (SEQ ID NO: 520)		544
CR16-3_AS:	AAAC		GGGCCCCGCCCCCTCCACC (SEQ ID NO: 521)	C	545
CR16-4_S:	cacc	G	GGTATCCACGTAAATCAAA (SEQ ID NO: 522)		546
CR16-4_AS:	AAAC		TTTGATTTACGTGGATACC (SEQ ID NO: 523)	C	547
CR16-5_S:	cacc	G	CCAATCACTGGCTCCGGTC (SEQ ID NO: 524)		548
CR16-5_AS:	AAAC		GACCGGAGCCAGTGATTGG (SEQ ID NO: 525)	C	549
CR16-6_S:	cacc	G	GGCGCCCGAGGGAAGAAGA (SEQ ID NO: 526)		550
CR16-6_AS:	AAAC		TCTTCTTCCCTCGGGCGCC (SEQ ID NO: 527)	C	551
CR16-7_S:	cacc	G	GGGTGGGGGTACCAGAGGA (SEQ ID NO: 528)		552
CR16-7_AS:	AAAC		TCCTCTGGTACCCCCACCC (SEQ ID NO: 529)	C	553
CR16-8_S:	cacc	G	CCGGGGACAGAAGAGAGGG (SEQ ID NO: 530)		554
CR16-8_AS:	AAAC		CCCTCTCTTCTGTCCCCGG (SEQ ID NO: 531)	C	555
CR16-9_S:	cacc	G	gagagagagtgggagaagc (SEQ ID NO: 532)		556
CR16-9_AS:	AAAC		gcttctcccactctctctc (SEQ ID NO: 533)	C	557
CR16-10_S:	cacc	G	aaagtaactgtcaaatgcg (SEQ ID NO: 534)		558

[0451]

CR16-10_AS:	AAAC		cgcatttgacagttacttt (SEQ ID NO: 535)	C	559
CR16-11_S:	cacc	G	ttaaccagagcgcccagtc (SEQ ID NO: 536)		560
CR16-11_AS:	AAAC		gactgggcgctctggttaa (SEQ ID NO: 537)	C	561
CR16-12_S:	cacc	G	cgtcggagctgcccgctag (SEQ ID NO: 538)		562
CR16-12_AS:	AAAC		ctagcgggcagctccgacg (SEQ ID NO: 539)	C	563

[0452] qRT-PCR和IF:在dCas9-VP64和gRNA、*ASCL1* cDNA或编码萤光素酶的阴性对照载体递送的第3天,通过在MEF中的qRT-PCR和免疫荧光,评价内源*ASCL1*的激活。通过*TUJ1*和*MAP2*共染色,且鉴定具有神经元形态和延长的细胞突的细胞,评价iN的产生。

[0453] 活细胞报道分子:在N3培养基中7-8天后,将在聚D-赖氨酸/层粘连蛋白包被的盖玻片上培养的MEF用携带hSyn-RFP和MAP2-GCamP5报道分子的病毒转导以通过钙成像和电生理学针对功能特征鉴定大部分成熟的iN (图49B)。

[0454] 结果。dCas9-VP64和靶向*ASCL1*启动子的gRNA激活MEF中的内源基因。相对于通过共同递送4种gRNA诱导激活的100倍,8种gRNA的共同递送激活内源基因400倍显著增加($p<0.05$) (图50A)。通过免疫荧光,检出MEF中核定位的*ASCL1*蛋白。与dCas9-VP64与gRNA混合物相比,异位*Asc11*表达产生更多的*ASCL1*蛋白,但是到第3天不激活内源基因座(图50A、50B)。在递送dCas9-VP64和靶向*ASCL1*和*BRN2*启动子的gRNA后,在神经发生培养基中13天之后,通过免疫荧光,鉴定出具有延长的细胞突的*TUJ1*和*MAP2*共阳性细胞(图51A第一行)。用BAM因子的异位表达,鉴定出类似数目的*TUJ1*和*MAP2*共阳性细胞(图51A第二行)。在神经发生培养基中早在第11天,在培养物中可观察到表达hSyn-RFP报道分子具有神经元形态的细胞(图51B)。表达MAP2-GCaMP5钙指示物的细胞显示用荧光显微镜检出的KCl诱导的去极化(图52A、52B)。

[0455] 小鼠胚胎成纤维细胞直接转换成具有神经元形态的*TUJ1*和*MAP2*共阳性细胞通过经由基于CRISPR/Cas9激活的转录因子的内源*BRN2*和*ASCL1*来实现。虽然dCas9-VP64比*ASCL1*的异位表达产生较少的蛋白质(图50B),但是神经元样细胞的产生相似。内源基因座的激活可诱导与用异位表达产生的在机械理论上不同的事件的重编程级联。

[0456] dCas9-VP64能够渗入异染色质,并激活稳定沉默的内源基因,唯一的一亚类“先驱(pioneer)”转录因子的特性。因此,用CRISPR/Cas9-转录因子转化细胞谱系可比转录因子的异位表达更好地克服重编程的外遗传障碍,特别是难以重编程细胞类型,例如成人细胞。这在再生医学领域可具有临床重要性,因为在细胞替代疗法常常需要使用自体源。

[0457] 实施例17

[0458] 基于多重CRISPR/Cas9的基因组工程改造—材料和方法

[0459] 质粒构建体。如上所述,使用酿脓链球菌sgRNA和人密码子优化的Cas9 (hCas9)核酸酶的表达盒。使用GeneBlocks (IDT),合成mU6 (Ohshima等, Nucleic Acids Res 9:5145-5158 (1981))、H1 (Myslinski等, Nucleic Acids Res 29:2502-2509 (2001))和7SK (Murphy等, Cell 51:81-87 (1987)) pol-III启动子的其它启动子,并克隆以替换hU6 sgRNA表达盒。将GeneBlock (IDT)克隆至Cas9编码序列的3’端以紧接Cas9之后融合T2A跳跃肽和eGFP基因以监测载体表达。然后将hCas9-T2A-GFP的编码区(SEQ ID NO: 145)转移至慢病毒表达载体中,所述载体含有驱动hCas9-T2A-GFP表达的人泛素C (hUbC)启动子以及促进紧接hUbC启动子上游的sgRNA表达盒的Golden Gate克隆的限制位点(图42A)。

[0460] 用于定制慢病毒载体装配的方案。表达高达4种选定的sgRNA和有活性的Cas9、dCas9或dCas9-VP64的定制慢病毒载体的装配在少于5天内完成。克隆方法采用Golden Gate克隆和切割其识别序列以外产生独特突出端的IIS型限制性内切酶。Golden Gate装配加快完成克隆,因为所有4种表达盒在一个步骤中连接进入最终的慢病毒载体中。除来自独立启动子的1、2、3或4种sgRNA表达以外,慢病毒载体表达有活性的Cas9、cCas9或dCas9-VP64。

[0461] 步骤1:使各含有20 bp前间区的单链寡核苷酸以产生黏性末端的所述方式退火,并连接至所需的pZ供体-启动子载体中。订购各个所需的基因组靶的2个单链寡核苷酸。为了使互补寡核苷酸退火,将8 μL有义寡+ 8 μL反义寡(两者为10mM) + 2 μL 10X连接酶缓冲液混合。使寡核苷酸解链,并在PCR机中按以下程序以步骤之间-0.3℃/秒钟速率再退火:96℃持续300秒钟,接着85℃持续20秒钟、75℃持续20秒钟、65℃持续20秒钟、55℃持续20秒钟、45℃持续20秒钟、35℃持续20秒钟和25℃持续20秒钟。为了使黏性末端磷酸化,加入1 μL 25mM ATP + 1微升T4多核苷酸激酶(NEB),在37℃下温育60分钟接着65℃温育20分钟以使酶热灭活。按照生产厂说明书,在10 μL反应体积中,使用50ng载体和1 μL退火寡核苷酸,使用在16℃下温育60分钟的T4 DNA连接酶(NEB)使各前间区连接至所需的表达载体中。按照生产商的说明书,将5微升连接物转化至XL1 blue化学感受态细菌(Agilent)中。将板变换成含有50 μg/mL卡那霉素(Sigma)的LB琼脂板,在37℃下温育过夜。按我们的经验,>90%的菌落将含有所需的连接产物。在进入步骤2之前,进行了使用M13反向标准测序引物的测序,以证实各最终的sgRNA构建体。

[0462] *步骤2:使用Golden Gate装配构建进入慢病毒目标载体的4个启动子-gRNA盒*。在完成步骤1后,有4种各自表达来自不同启动子的不同sgRNA的独立质粒。为了将4种不同的启动子-sgRNA构建体装配进入所需的目标载体中,将200 ng各sgRNA表达质粒和所需的慢病毒目标载体与1 μL T4 DNA连接酶(NEB)、1 μL BsmBI FastDigest (Fisher Scientific)和2 μL 10X T4连接酶缓冲液(NEB)在20 μL反应体积中混合。如下温育反应物:37℃持续10分钟,16℃持续15分钟,37℃持续30分钟,80℃持续5分钟。按照生产商的说明书,将5μL连接反应转化至SURE 2化学感受态细胞(Agilent)中。将板转换至含有100 μg/mL氨苄西林的LB琼脂板中,在37℃下温育过夜。任选可使用IPTG和X-gal,通过基于lacZ的蓝/白色筛选菌落;然而,按我们的经验,>90%的转化株含有合适的连接产物。由于通过反向sgRNA表达盒形成的反向重复序列,最终的构建体可能不稳定,因此我们建议将这些质粒在SURE 2细胞系中维持,并使用有义引物5'-TCGGGTTTATTACAGGGACAGCAG-3' (SEQ ID NO:464)和反义引物5'-TCTAAGGCCGAGTCTTATGAGCAG-3' (SEQ ID NO:465),用试验PCR筛选最终的质粒。这些引物跨4个启动子-gRNA区扩增。由于重复性质所致,应观察到不同的带型,其最大产物的大小为约1800 bp。

[0463] 细胞培养和转染。HEK293T细胞经由杜克大学癌症中心机构获自美国组织保藏中心(ATCC,Manassas,VA),并在补充10% FBS和1%青霉素/链霉素的DMEM中维持。原生人真皮成纤维细胞(Catalog ID:GM03348)获自Coriell Institute (Camden,New Jersey),并在补充10% FBS和1%青霉素/链霉素的DMEM中维持。将所有细胞在37℃、5% CO_2下培养。按照生产商的方案,用Lipofectamine 2000 (Life Technologies),将HEK293T细胞在24孔板中用200 ng各sgRNA表达载体(800 ng总的pDNA)转染。

[0464] 病毒产生和转导。本研究所用的所有慢病毒载体均为第二代,并采用标准病毒生产方法生产。简单地讲,每10cm培养皿接种350万个HEK293T。次日,通过磷酸钙转染方法用20μg转移载体、6μg pMD2G和10μg psPAX2转染细胞。转染后12-14小时更换培养基。在该培养基更换后24和48小时收集病毒上清液,通过0.45微米滤器,并合并。对于转导,将细胞培养基用补充4μg/mL聚凝胺(polybrene)的病毒上清液更换。12-24小时后,将病毒上清液更换为新鲜培养基。

[0465] 反转录PCR。使用miRNeasy Mini RNA分离试剂盒(Qiagen)分离RNA。使用不含DNA的试剂盒(Applied Biosystems)进行DNA酶消化。使用SuperScript VILO cDNA合成试剂盒(Invitrogen)进行cDNA合成。使用Taq DNA聚合酶(NEB)对cDNA进行PCR扩增,使所得产物在TAE琼脂糖凝胶中电泳。采用ChemiDoc XRS+系统拍摄图像,并应用ImageLab软件(Bio-Rad)处理。

[0466] 定量实时PCR。使用RNeasy Plus RNA分离试剂盒(Qiagen)分离RNA。使用SuperScript VILO cDNA合成试剂盒(Invitrogen)进行cDNA合成。用CFX96实时PCR检测系统(Bio-Rad),进行使用PerfeCTa SYBR Green FastMix (Quanta Biosciences)的实时PCR。通过琼脂糖凝胶电泳和解链曲线分析,证实引物特异性。计算适当的动态范围内的反应效率以确保标准曲线的线性。采用ΔΔCt方法,将结果表示为归一化至β-肌动蛋白表达的目标基因mRNA表达的倍数增加。报告值为2个独立实验(n = 2)的均值和S.E.M.,其中对于各实验,对技术性一式两份取平均。

[0467] 蛋白质印迹。将细胞在补充蛋白酶抑制剂混合物(Sigma)的RIPA缓冲液(Sigma)中裂解。采用BCA蛋白质测定法试剂(Thermo Scientific)和BioTek Synergy 2 Multi-Mode微量板读数器,测量蛋白质浓度。将裂解物与加载缓冲液混合,并煮沸5分钟;使25μg蛋白质在NuPage 10% Bis-Tris凝胶聚丙烯酰胺凝胶(Bio-Rad)中电泳,并转移到硝酸纤维素膜上。将非特异性抗体结合用TBST (50mM Tris,150mM NaCl和0.1% Tween-20)与5%脱脂奶在室温下封闭1小时。使膜与以下第一抗体在以下条件下温育:抗MyoD (1:250,Santa Cruz Sc-32758)在含5% BSA的TBST中在4℃下过夜;抗Myogenin (1:250,Santa Cruz Sc-12732)在5% BSA中在4℃下过夜;抗FLAG-HRP (1:1000,Cell Signaling 2044)在含5%牛奶的TBST在室温下60分钟;抗GAPDH (1:5000,Cell Signaling,克隆14C10)在含5%牛奶的TBST中在室温下30分钟。然后将膜用TBST洗涤3次共15分钟。将膜与1:5,000稀释的抗兔HRP缀合的抗体(Sigma,6154)或抗小鼠HRP缀合的抗体(Santa Cruz,SC-2005)一起温育30分钟,用TBST洗涤3次各15分钟。使用ImmunStar WesternC化学发光试剂盒(Bio-Rad)观察膜,采用ChemiDoc XRS+系统拍摄图像,并应用ImageLab软件(Bio-Rad)处理。

[0468] 内源基因修饰的Cel-I定量测定。采用可检测核酸酶介导的NHEJ的突变特性的Surveyor核酸酶测定法,定量测定内源靶位点处的CRISPR/Cas9核酸酶损害。在转染或转导后,将细胞在37℃下温育3-10天,使用DNeasy血液与组织试剂盒(Qiagen)提取基因组DNA。用AccuPrime High Fidelity PCR试剂盒(Invitrogen),通过35个PCR循环扩增靶基因座。使所得PCR产物随机解链,并在PCR机中按以下程序在步骤之间以-0.3℃/秒钟速率再退火:95℃持续240秒钟,接着85℃持续60秒钟,75℃持续60秒钟,65℃持续60秒钟,55℃持续60秒钟,45℃持续60秒钟,35℃持续60秒钟和25℃持续60秒钟。在再退火后,将8 μL PCR产物与1 μL Surveyor核酸酶S和1 μL增强子S (Transgenomic,Omaha,NE)混合,在42℃下温育1小

时。在温育后,将6 μL消化产物加载到10% TBE聚丙烯酰胺凝胶上,以200 V电泳30分钟。将凝胶用溴化乙锭染色,通过光密度测定法使用ImageLab (Bio-Rad)定量测定(Perez-Pinera等, Nucleic Acids Res 40:3741-3751 (2012))。

[0469] 统计分析。将至少2个独立实验编制为均值和均值的标准误差。应用JMP 10 Pro,用多变量ANOVA和Dunnett事后检验评价作用。

[0470] 实施例18

[0471] 用于多重CRISPR/Cas9应用的单一慢病毒载体的开发

[0472] 现行CRISPR/Cas9基因编辑系统,特别是反式激活蛋白系统的限制,是用于多重基因编辑和协同基因激活应用(尤其在难以转染细胞类型中)同时有效地递送多种sgRNA和Cas9蛋白。为了克服这种限制,我们开发了有效表达Cas9和多达4种sgRNA的单一慢病毒载体。为了使各sgRNA的表达效率最大化,该载体表达来自独立pol III启动子(人U6启动子、小鼠U6启动子、7SK和H1)的4种sgRNA。我们证实了终点RT-PCR的各启动子的sgRNA表达以检测靶向AAVS1基因座的sgRNA (图42A)。为了检测各sgRNA表达构建体的活性,我们将独立表达靶向AAVS1的sgRNA的各启动子构建体与活性Cas9表达构建体共转染至人HEK293T细胞中。显然,我们检出与充分表征的在AAVS1基因座上具有高活性的锌指核酸酶相当的各sgRNA在靶基因座上的一致和高水平的基因修饰(图42B)。此外,基于不同Cas9的构建体,包括有活性的Cas9核酸酶、死Cas9和与VP64反式激活蛋白结构域融合的死Cas9的慢病毒递送,导致HEK293T细胞中全长Cas9蛋白的表达,如通过蛋白质印迹测定的一样(图42C)。

[0473] 使用这些成分,我们开发了Golden Gate克隆方法以促进多个sgRNA表达盒快速有效地克隆至表达所需Cas9效应子的单一慢病毒载体中(图43)。在第一步中,将编码sgRNA前间区序列的寡核苷酸独立地克隆至各自具有驱动sgRNA表达的不同启动子的不同的表达载体中。在第二步中,通过Golden Gate装配,将各sgRNA表达构建体亚克隆至所选的慢病毒Cas9表达载体中。这种策略允许将多达4种sgRNA快速稳健克隆至单一慢病毒载体中用于基因编辑或激活应用。为了表达少于4种sgRNA,在未使用的启动子下游克隆PolyT终止子序列以防止从未使用的启动子转录。各载体通过2A跳跃肽共表达所选的Cas9与eGFP以使具有高感染复数的细胞的荧光激活流式细胞分选和富集成为可能。最后,含有sgRNA和Cas9表达盒的整个区域的两侧是loxP位点以介导通过Cre-lox切除的去除。

[0474] 实施例19

[0475] 用于多重基因组工程改造的单一慢病毒sgRNA/Cas9表达载体的验证

[0476] 为了证实各sgRNA的独立活性,我们装配了表达有活性的Cas9和各自靶向独立基因座的4种sgRNA的单一慢病毒载体(图44A)。作为对照载体,我们装配了表达仅一种sgRNA连同在其它3个位置处的polyT前间区的构建体。我们用表达指定sgRNA的慢病毒载体转导HEK293T和原生成纤维细胞,分别在转导后7或10天监测基因修饰频率(图44B)。在两种细胞类型中,单一慢病毒载体在所有4个基因座上介导高效的多重编辑(图44B)。引人关注的是,在成纤维细胞的4基因座中的3个中,所有4种sgRNA一起的表达导致比单一sgRNA高的修饰频率(图44B)。在按惯例是难以转染的细胞类型的成纤维细胞中,我们观察到有效的多重基因编辑。这些数据表明,单一慢病毒可有效表达4种有活性的sgRNA,该慢病毒平台可用来靶向4个不同的基因座用于多重CRISPR/Cas9基因编辑。

[0477] 实施例20

[0478] 稳定表达基于慢病毒Cas9的反式激活蛋白的细胞系中的瞬时RNA指导基因激活

[0479] 接下来,我们将兴趣放在开发通过将sgRNA转染至稳定表达Cas9的模式细胞系中使瞬时基因激活成为可能的系统。将HEK293T用不同的Cas9-T2A-GFP转导,采用流式细胞术监测GFP表达。在每2-3天正常传代后,各细胞系显示稳定的GFP表达直到转导后35天。然后将转导的HEK293T用靶向IL1RN或HBG1启动子的1-4种单独的sgRNA表达构建体转染。这些sgRNA构建体在稳定表达dCas9-VP64的细胞系中的瞬时转染导致可调节内源基因激活(图45A、45B)。在表达dCas9-VP64的细胞中sgRNA构建体瞬时转染后,基因激活在转染后约3-6天达到最大激活水平,到转染后20天下降到未检出的水平(图45CD、45D)。此外,我们能够通过第二次转染靶向各启动子的所有4种sgRNA构建体重新激活各基因,尽管激活水平显著低于第一次转染所观察到的(图45C、45D)。第二次转染后的这种活性降低可能是由于载体表达降低或未转导细胞的竞争性生长所致。尽管如此,这些数据表明慢病毒Cas9与瞬时sgRNA递送组合可用作多用途系统在Cas9稳定转导的细胞系中可调节且瞬时激活和重新激活靶基因。

[0480] 实施例21

[0481] HEK293T细胞中使用单一慢病毒sgRNA/Cas9反式激活蛋白表达载体的稳定基因激活

[0482] 慢病毒递送可通过CRISPR/Cas9反式激活使长期稳定的基因激活成为可能。为了检验这一点,使用编码dCas9-VP64和1-4种sgRNA表达盒的单一慢病毒载体转导HEK293T。与我们的瞬时转染结果类似(图45),我们能够可调节且稳健地激活内源IL1RN和HBG1基因的表达(图46A、46B)。通过用dCas9-VP64和靶向IL1RN和HBG1启动子的4种sgRNA共转染HEK293T细胞诱导的基因激活在转染后3-5天达到峰值,在转染后15-20天,基因表达回到背景水平。相比之下,dCas9-VP64和相同4种靶定IL1RN或HBG1的sgRNA的慢病毒递送诱导持续的基因激活在转导后超过20天(图46C、46D)。因此,多重dCas9-VP64反式激活蛋白的单一慢病毒递送是稳定有效地增量调节靶内源基因的有用平台。

[0483] 实施例22

[0484] dCas9-KRAB—靶向HS2增强子

[0485] HS2增强子是充分表征的激活珠蛋白基因基因座所必需的远端调节元件。递送dCas9-KRAB与靶向HS2增强子的gRNA以确定该系统是否可在K562人类红细胞白血病细胞系中阻抑γ-、ε-和β-珠蛋白表达(图54)。产生靶向沿HS2增强子的核心区(SEQ ID NO: 467)的不同位点的一组gRNA。参见表12。

表 12 HS2 gRNA 靶序列

[0486]

Cr#	前间区	完全 gRNA 序列
1	gagacacacagaaatgtaac (SEQ ID NO: 564)	gagacacacagaaatgtaacgtttTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTCC(SEQ ID NO: 585)
2	ggtggggcactgaccccgac (SEQ ID NO: 565)	ggtggggcactgaccccgacgtttTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTCC(SEQ ID NO: 586)
3	ctagagtgatgactcctatc (SEQ ID NO: 566)	ctagagtgatgactcctatcgtttTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTCC(SEQ ID NO: 587)
4	gactaaaactccacctcaaa (SEQ ID NO: 567)	gactaaaactccacctcaaagtttTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTCC(SEQ ID NO: 588)

[0487]

5	aatatgtcacattctgtctc (SEQ ID NO: 568)	aatatgtcacattctgtctcgtttTAGAGCTAGAAATAGCAAGTTAA AATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGC ACCGAGTCGGTGCTTTTTTTCC(SEQ ID NO: 589)
6	ggactatgggaggtcactaa (SEQ ID NO: 569)	ggactatgggaggtcactaagtttTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGG CACCGAGTCGGTGCTTTTTTTCC(SEQ ID NO: 590)
7	gctcatgcttggactatggg (SEQ ID NO: 570)	gctcatgcttggactatggggtttTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGG CACCGAGTCGGTGCTTTTTTTCC(SEQ ID NO: 591)
8	gttctggccaggcccctgtc (SEQ ID NO: 571)	gttctggccaggcccctgtcgtttTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGG CACCGAGTCGGTGCTTTTTTTCC(SEQ ID NO: 592)
9	agtgccccacccccgccttc (SEQ ID NO: 572)	agtgccccacccccgccttcgtttTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGG CACCGAGTCGGTGCTTTTTTTCC(SEQ ID NO: 593)
10	gtggggcactgaccccgaca (SEQ ID NO: 573)	gtggggcactgaccccgacagtttTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTG GCACCGAGTCGGTGCTTTTTTTCC(SEQ ID NO: 594)
11	aaccttctaagcaaaccttc (SEQ ID NO: 574)	aaccttctaagcaaaccttcgtttTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGG CACCGAGTCGGTGCTTTTTTTCC(SEQ ID NO: 595)
12	gttacacagaaccagaaggc (SEQ ID NO: 575)	gttacacagaaccagaaggcgtttTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTG GCACCGAGTCGGTGCTTTTTTTCC(SEQ ID NO: 596)
13	gaaggttacacagaaccaga (SEQ ID NO: 576)	gaaggttacacagaaccagagtttTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTG GCACCGAGTCGGTGCTTTTTTTCC(SEQ ID NO: 597)
14	agtcatgatgagtcatgctg (SEQ ID NO: 577)	agtcatgatgagtcatgctggtttTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGG CACCGAGTCGGTGCTTTTTTTCC(SEQ ID NO: 598)
15	gatgagtcatgctgaggctt (SEQ ID NO: 578)	gatgagtcatgctgaggcttgtttTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGG CACCGAGTCGGTGCTTTTTTTCC(SEQ ID NO: 599)
16	actctaggctgagaacatct (SEQ ID NO: 579)	actctaggctgagaacatctgtttTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGG CACCGAGTCGGTGCTTTTTTTCC(SEQ ID NO: 600)

[0488]

17	gtccccagcaggatgcttac (SEQ ID NO: 580)	gtccccagcaggatgcttacgtttTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGG CACCGAGTCGGTGCTTTTTTTCC(SEQ ID NO: 601)
18	gccctgtaagcatcctgctg (SEQ ID NO: 581)	gccctgtaagcatcctgctggtttTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGG CACCGAGTCGGTGCTTTTTTTCC(SEQ ID NO: 602)
19	cagggcagatggcaaaaaaa (SEQ ID NO: 582)	cagggcagatggcaaaaaaagtttTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTG GCACCGAGTCGGTGCTTTTTTTCC(SEQ ID NO: 603)
20	gaggtggagttttagtcagg (SEQ ID NO: 583)	gaggtggagttttagtcagggtttTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGG CACCGAGTCGGTGCTTTTTTTCC(SEQ ID NO: 604)
21	aaacggcatcataaagaaaa (SEQ ID NO: 584)	aaacggcatcataaagaaaagtttTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGG CACCGAGTCGGTGCTTTTTTTCC(SEQ ID NO: 605)

[0489] 通过CRISPR/dCas9筛选珠蛋白基因座上的单一gRNA。在用5 μg编码U6-sgRNA的表达质粒电穿孔前5-8天,用慢病毒递送dCas9和dCas9-KRAB效应物(图55A)。包括不用gRNA电穿孔的细胞(无gRNA)和用靶向不同基因座的gRNA处理的细胞(IL1RN)作为对照。在转染后3天测定时,多种gRNA实现ε-、γ-和β-珠蛋白基因的有效阻抑,达到高达80%敲除(图55B、55C、55D)。表达gRNA与dCas9或dCas9-KRAB抑制珠蛋白基因座上的基因表达。总的来讲,与仅dCas9相比,用dCas9-KRAB处理导致指定gRNA的较强阻抑,表明了KRAB结构域在募集提高阻抑的异染色质因子中的重要作用。所达到的阻抑水平依赖于只在dCas9-KRAB处理细胞中通过转染所递送的gRNA质粒的量(图56A、56B、56C)。增加Cr4 gRNA质粒的剂量直到10 μg提高在dCas9-KRAB处理细胞中沉默珠蛋白基因的水平。

[0490] 通过dCas9-KRAB使珠蛋白基因稳定沉默。通过慢病毒使dCas9/dCas9-KRAB与单一gRNA在K562中共表达(图57A)。包括不用慢病毒处理(NT)、用dCas9/dCas9-KRAB但无gRNA处理(无gRNA)和用dCas9/dCas9-KRAB和靶向不同基因座(IL1RN)的gRNA处理的细胞作为对照。从第4天到第7天选择用慢病毒处理的细胞。当在转导后7天测定时,多种gRNA实现ε-、γ-和β-珠蛋白基因的有效转录阻抑,达到高达95%敲除(图57B、57C、57D)。ε-珠蛋白的表达大多在响应靶定HS2增强子的gRNA时沉默。与用dCas9和gRNA处理相比,用dCas9-KRAB与gRNA处理导致显著更多的阻抑。

[0491] 这些观察结果表明,通过gRNA靶向HS2增强子的dCas9-KRAB实现远端珠蛋白基因的有效阻抑。这是在哺乳动物细胞中通过CRISPR/Cas9系统靶定外遗传控制的远端调节元件的第1个实例。增强子调节发育和疾病,并且本公开内容提供探测并控制增强子功能的方法,且可用来测定dCas9-KRAB对局部染色质可及性和全基因组表达的作用。

[0492] 实施例23

[0493] dCas9-p300

[0494] 设计dCas9-p300融合蛋白,并与dCas9-VP64融合蛋白进行比较(参见图59)。dCas9构建的氨基酸构建体见图61A-61C。在用Lipofectamine 2000转染试剂(Life

Technologies)转染前1天,将人胚肾(*H*uman *E*mbryonic *K*idney)组织培养系HEK293T的细胞(ATCC;CRL-11268)以1.5e5个细胞/孔的密度接种在24孔组织培养培养皿中。24小时后,将细胞用1μL Lipofectamine 2000、375ng dCas9表达构建体(分别为dCas9、dCas9VP64或dCas9p300)和125ng合并的gRNA表达质粒(4种,各为等摩尔比率)转染。表13显示gRNA信息。

表 13 gRNA 信息

[0495]

靶标位置	前间区序列 (5'-3')	基因组位置 (GRCh38 初步装配)	参考文献
IL1RN 启动子	TGTACTCTCTGAGGTGCTC (SEQ ID NO: 606)	Chr2 : 113117865-113117883	Perez-Pinera 等, Nat Methods, 2013
IL1RN 启动子	ACGCAGATAAGAACCAGTT	Chr2 : 113117714-113117732	Perez-Pinera 等, Nat

[0496]

	(SEQ ID NO: 607)		Methods, 2013
IL1RN 启动子	CATCAAGTCAGCCATCAGC (SEQ ID NO: 608)	Chr2 : 113117781-113117799	Perez-Pinera 等，Nat Methods, 2013
IL1RN 启动子	GAGTCACCCTCCTGGAAAC (SEQ ID NO: 609)	Chr2 : 113117749-113117767	Perez-Pinera 等，Nat Methods, 2013
MyoD 启动子	CCTGGGCTCCGGGGCGTTT (SEQ ID NO: 610)	Chr11 : 17719509-17719527	Perez-Pinera 等，Nat Methods, 2013
MyoD 启动子	GGCCCCTGCGGCCACCCCG (SEQ ID NO: 611)	Chr11 : 17719422-17719440	Perez-Pinera 等，Nat Methods, 2013
MyoD 启动子	CTCCCTCCCTGCCCGGTAG (SEQ ID NO: 612)	Chr11 : 17719350-17719368	Perez-Pinera 等，Nat Methods, 2013
MyoD 启动子	AGGTTTGGAAAGGGCGTGC (SEQ ID NO: 613)	Chr11 : 17719290-17719308	Perez-Pinera 等，Nat Methods, 2013
Oct4 启动子	ACTCCACTGCACTCCAGTCT (SEQ ID NO: 614)	Chr6 : 31170953-31170934	Hu 等, Nucleic Acids Res, 2014
Oct4 启动子	TCTGTGGGGGACCTGCACTG (SEQ ID NO: 615)	Chr6 : 31170885-31170866	Hu 等, Nucleic Acids Res, 2014
Oct4 启动子	GGGGCGCCAGTTGTGTCTCC (SEQ ID NO: 616)	Chr6 : 31170855-31170836	Hu 等, Nucleic Acids Res, 2014
Oct4 启动子	ACACCATTGCCACCACCATT (SEQ ID NO: 617)	Chr6 : 31170816-31170797	Hu 等, Nucleic Acids Res, 2014

[0497] 收获转染后3天的细胞,通过RT-QPCR测定mRNA表达。RT-QPCR引物序列列于表14中。

[0498] 表14 RT-QPCR引物

[0499]

引物靶标	引物序列 (5'-3')	参考文献
GAPDH 正向	CAATGACCCCTTCATTGACC (SEQ ID NO: 618)	Perez-Pinera 等, Nat Methods, 2013
GAPDH 反向	TTGATTTTGGAGGGATCTCG (SEQ ID NO: 619)	Perez-Pinera 等, Nat Methods, 2013
IL1RN 正向	GGAATCCATGGAGGGAAGAT (SEQ ID NO: 620)	Perez-Pinera 等, Nat Methods, 2013
IL1RN 反向	TGTTCTCGCTCAGGTCAGTG (SEQ ID NO: 621)	Perez-Pinera 等, Nat Methods, 2013
MyoD 正向	CTCTCTGCTCCTTTGCCACA (SEQ ID NO: 622)	Perez-Pinera 等, Nat Methods, 2013
MyoD 反向	GTGCTCTTCGGGTTTCAGGA (SEQ ID NO: 623)	Perez-Pinera 等, Nat Methods, 2013
Oct4 正向	CGAAAGAGAAAGCGAACCAGTATCGAGAAC (SEQ ID NO: 624)	Hu 等, Nucleic Acids Res, 2014
Oct4 反向	CGTTGTGCATAGTCGCTGCTTGATCGC (SEQ ID NO: 625)	Hu 等, Nucleic Acids Res, 2014

[0500] 采用$\Delta\Delta C_t$方法,将RT-QPCR归一化至*GAPDH*表达。结果表示为相对于仅用Lipofectamine而无DNA转染处理的细胞("无DNA"),目标基因表达的倍数增加(参见图60A-60C)。

[0501] 图62显示p300 HAT结构域中的残基突变引起其激活基因表达的能力丧失。图63显示多种gRNA与dCas-9-p300协同作用,如与dCas-9-VP64所示一样。

[0502] 实施例24

[0503] 图64显示来源于肌养蛋白基因中携带不同缺失的人类DMD患者的骨骼成肌细胞系中肌养蛋白的Dp427m骨骼肌同种型5'UTR的微小肌养蛋白的TALEN介导的整合。将DMD患者细胞用编码对5'UTR基因座有活性的TALEN对的构建体和携带微小肌养蛋白基因的供体模板电穿孔。图64(a)显示微小肌养蛋白如何整合至5'UTR中的示意图。图64(b)显示分离潮霉素抗性克隆细胞系,并使用图64(a)所示引物,通过PCR筛选在5'UTR上的成功位点特异性整合。星号标出选择用于图64(c)中的进一步分析的克隆。图64(c)显示使具有检出整合事件的克隆分离的DMD成肌细胞分化6天,并评价与微小肌养蛋白C端融合HA标签的表达。

[0504] 要了解,上文的描述描述和随附实施例只是说明性的,不得视为对本发明范围的限制,本发明的范围仅由随附权利要求书及其等同内容限定。

[0505] 对所公开的实施方案的各种变化和修改对于本领域人员而言将是显而易见的。可在不偏离本发明的精神和范围的情况下进行所述变化和修改,包括而不限于与本发明的化学结构、取代基、衍生物、中间体、合成法、组合物、制剂或使用方法有关的变化和修改。

[0506] 附录

hCas9-T2A-GFP DNA 序列：SEQ ID NO: 145 (SpCas9 人优化序列，*HA tag*，T2A 肽，*eGFP 序列*)

[0507]

gccaccATGGACAAGAAGTACTCCATTGGGCTCGATATCGGCACAAACAGCGTCGGCTGGGCCGTCATTACGGACGAGTA
CAAGGTGCCGAGCAAAAAATTCAAAGTTCTGGGCAATACCGATCGCCACAGCATAAAGAAGAACCTCATTGGCGCCCTCC
TGTTCGACTCCGGGGAGACGGCCGAAGCCACGCGGCTCAAAAGAACAGCACGGCGCAGATATACCCGCAGAAAGAATCGG
ATCTGCTACCTGCAGGAGATCTTTAGTAATGAGATGGCTAAGGTGGATGACTCTTTCTTCCATAGGCTGGAGGAGTCCTT
TTTGGTGGAGGAGGATAAAAAGCACGAGCGCCACCCAATCTTTGGCAATATCGTGGACGAGGTGGCGTACCATGAAAAGT
ACCCAACCATATATCATCTGAGGAAGAAGCTTGTAGACAGTACTGATAAGGCTGACTTGCGGTTGATCTATCTCGCGCTG
GCGCATATGATCAAATTTCGGGGACACTTCCTCATCGAGGGGGACCTGAACCCAGACAACAGCGATGTCGACAAACTCTT
TATCCAACTGGTTCAGACTTACAATCAGCTTTTCGAAGAGAACCCGATCAACGCATCCGGAGTTGACGCCAAAGCAATCC
TGAGCGCTAGGCTGTCCAAATCCCGGCGGCTCGAAAACCTCATCGCACAGCTCCCTGGGGAGAAGAAGAACGGCCTGTTT
GGTAATCTTATCGCCCTGTCACTCGGGCTGACCCCCAACTTTAAATCTAACTTCGACCTGGCCGAAGATGCCAAGCTTCA
ACTGAGCAAAGACACCTACGATGATGATCTCGACAATCTGCTGGCCCAGATCGGCGACCAGTACGCAGACCTTTTTTTGG
CGGCAAAGAACCTGTCAGACGCCATTCTGCTGAGTGATATTCTGCGAGTGAACACGGAGATCACCAAAGCTCCGCTGAGC
GCTAGTATGATCAAGCGCTATGATGAGCACCACCAAGACTTGACTTTGCTGAAGGCCCTTGTCAGACAGCAACTGCCTGA
GAAGTACAAGGAAATTTTCTTCGATCAGTCTAAAAATGGCTACGCCGGATACATTGACGGCGGAGCAAGCCAGGAGGAAT
TTTACAAATTTATTAAGCCCATCTTGGAAAAAATGGACGGCACCGAGGAGCTGCTGGTAAAGCTTAACAGAGAAGATCTG
TTGCGCAAACAGCGGCACTTTCGACAATGGAAGCATCCCCCACCAGATTCACCTGGGCGAACTGCACGCTATCCTCAGGCG
GCAAGAGGATTTCTACCCCTTTTTGAAAGATAACAGGGAAAAGATTGAGAAAATCCTCACATTTCGGATACCCTACTATG
TAGGCCCCCTCGCCCGGGGAAATTCCAGATTCGCGTGGATGACTCGCAAATCAGAAGAGACCATCACTCCCTGGAACTTC
GAGGAAGTCGTGGATAAGGGGGCCTCTGCCCAGTCCTTCATCGAAAGGATGACTAACTTTGATAAAAATCTGCCTAACGA
AAAGGTGCTTCCTAAACACTCTCTGCTGTACGAGTACTTCACAGTTTATAACGAGCTCACCAAGGTCAAATACGTCACAG
AAGGGATGAGAAAGCCAGCATTCCTGTCTGGAGAGCAGAAGAAAGCTATCGTGGACCTCCTCTTCAAGACGAACCGGAAA
GTTACCGTGAAACAGCTCAAAGAAGACTATTTCAAAAAGATTGAATGTTTCGACTCTGTTGAAATCAGCGGAGTGGAGGA
TCGCTTCAACGCATCCCTGGGAACGTATCACGATCTCCTGAAAATCATTAAAGACAAGGACTTCCTGGACAATGAGGAGA
ACGAGGACATTCTTGAGGACATTGTCCTCACCCTTACGTTGTTTGAAGATAGGGAGATGATTGAAGAACGCTTGAAAACT
TACGCTCATCTCTTCGACGACAAAGTCATGAAACAGCTCAAGAGGCGCCGATATACAGGATGGGGGCGGCTGTCAAGAAA
ACTGATCAATGGGATCCGAGACAAGCAGAGTGGAAAGACAATCCTGGATTTTCTTAAGTCCGATGGATTTGCCAACCGGA
ACTTCATGCAGTTGATCCATGATGACTCTCTCACCTTTAAGGAGGACATCCAGAAAGCACAAGTTTCTGGCCAGGGGGAC
AGTCTTCACGAGCACATCGCTAATCTTGCAGGTAGCCCAGCTATCAAAAAGGGAATACTGCAGACCGTTAAGGTCGTGGA
TGAACTCGTCAAAGTAATGGGAAGGCATAAGCCCGAGAATATCGTTATCGAGATGGCCCGAGAGAACCAAACTACCCAGA
AGGGACAGAAGAACAGTAGGGAAAGGATGAAGAGGATTGAAGAGGGTATAAAAGAACTGGGGTCCCAAATCCTTAAGGAA
CACCCAGTTGAAAACACCCAGCTTCAGAATGAGAAGCTCTACCTGTACTACCTGCAGAACGGCAGGGACATGTACGTGGA
TCAGGAACTGGACATCAATCGGCTCTCCGACTACGACGTGGATCATATCGTGCCCCAGTCTTTTCTCAAAGATGATTCTA
TTGATAATAAAGTGTTGACAAGATCCGATAAAAATAGAGGGAAGAGTGATAACGTCCCCTCAGAAGAAGTTGTCAAGAAA
ATGAAAAATTATTGGCGGCAGCTGCTGAACGCCAAACTGATCACACAACGGAAGTTCGATAATCTGACTAAGGCTGAACG
AGGTGGCCTGTCTGAGTTGGATAAAGCCGGCTTCATCAAAAGGCAGCTTGTTGAGACACGCCAGATCACCAAGCACGTGG
CCCAAATTCTCGATTCACGCATGAACACCAAGTACGATGAAAATGACAAACTGATTCGAGAGGTGAAAGTTATTACTCTG
AAGTCTAAGCTGGTCTCAGATTTCAGAAAGGACTTTCAGTTTTATAAGGTGAGAGAGATCAACAATTACCACCATGCGCA
TGATGCCTACCTGAATGCAGTGGTAGGCACTGCACTTATCAAAAAATATCCCAAGCTTGAATCTGAATTTGTTTACGGAG
ACTATAAAGTGTACGATGTTAGGAAAATGATCGCAAAGTCTGAGCAGGAAATAGGCAAGGCCACCGCTAAGTACTTCTTT
TACAGCAATATTATGAATTTTTTCAAGACCGAGATTACACTGGCCAATGGAGAGATTCGGAAGCGACCACTTATCGAAAC
AAACGGAGAAACAGGAGAAATCGTGTGGGACAAGGGTAGGGATTTCGCGACAGTCCGGAAGGTCCTGTCCATGCCGCAGG
TGAACATCGTTAAAAAGACCGAAGTACAGACCGGAGGCTTCTCCAAGGAAAGTATCCTCCCGAAAAGGAACAGCGACAAG
CTGATCGCACGCAAAAAAGATTGGGACCCCAAGAAATACGGCGGATTCGATTCTCCTACAGTCGCTTACAGTGTACTGGT
TGTGGCCAAAGTGGAGAAAGGGAAGTCTAAAAAACTCAAAAGCGTCAAGGAACTGCTGGGCATCACAATCATGGAGCGAT
CAAGCTTCGAAAAAAACCCCATCGACTTTCTCGAGGCGAAAGGATATAAAGAGGTCAAAAAAGACCTCATCATTAAGCTT
CCCAAGTACTCTCTCTTTGAGCTTGAAAACGGCCGGAAACGAATGCTCGCTAGTGCGGGCGAGCTGCAGAAAGGTAACGA
GCTGGCACTGCCCTCTAAATACGTTAATTTCTTGTATCTGGCCAGCCACTATGAAAAGCTCAAAGGGTCTCCCGAAGATA
ATGAGCAGAAGCAGCTGTTCGTGGAACAACACAAACACTACCTTGATGAGATCATCGAGCAAATAAGCGAATTCTCCAAA
AGAGTGATCCTCGCCGACGCTAACCTCGATAAGGTGCTTTCTGCTTACAATAAGCACAGGGATAAGCCCATCAGGGAGCA
GGCAGAAAACATTATCCACTTGTTTACTCTGACCAACTTGGGCGCGCCTGCAGCCTTCAAGTACTTCGACACCACCATAG
ACAGAAAGCGGTACACCTCTACAAAGGAGGTCCTGGACGCCACACTGATTCATCAGTCAATTACGGGGCTCTATGAAACA
AGAATCGACCTCTCTCAGCTCGGTGGAGACAGCAGGGCTGAC*CCCAAGAAGAAGAGGAAGGTGGC*TAGCG**AGGGCAGAGG**
AAGTCTTCTAACATGCGGTGACGTGGAGGAGAATCCCGGCCCTGgtacC***gtgagcaagggcgaggagctgttcaccgggg***
tggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccggcgagggcgagggcgatgcc
acctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgccgtgccctggcccaccctcgtgaccaccct
gacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaagtccgccatgcccgaag
gctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcgagggcgac
accctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagctggagtacaa
ctacaacagccacaacgtctatatcatggccgacaagcagaagaacggcatcaaggtgaacttcaagatccgccacaaca
tcgaggacggcagcgtgcagctcgccgaccactaccagcagaacacccccatcggcgacggccccgtgctgctgcccgac
aaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgt
gaccgccgccgggatcactctcggcatggacgagctgtacaagAccggTTAG

[0508] AAV/Rosa26构建体(SEQ ID NO: 456)

[0509] ggggggggggggggggggttggccactccctctctgcgcgctcgctcgctcactgaggccgggcgaccaaaggtcgcccgacgcccgggctttgcccgggcggcctcagtgagcgagcgagcgcgcagagagggagtggccaactccatcactaggggttcctagatctgaattcggtacccgttacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattga

cgtcaatgggtggagtatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgcccc
ctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttg
gcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatag
cggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaac
gggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtcta
tataagcagagctcgtttagtgaaccgtcagatcgcctggagacgccatccacgctgttttgacctccatagaaga
caccgggaccgatccagcctccggactctagaggatccggtactcgaggaactgaaaaaccagaaagttaactggt
aagtttagtctttttgtcttttatttcaggtcccggatccggtggtggtgcaaatcaaagaactgctcctcagtgg
atgttgcctttacttctaggcctgtacggaagtgttacttctgctctaaaagctgcggaattgtacccgcggcccg
ggatccaccggTGGCTAGCgtctataggcccacccccTTGGTGGAATTCGCCATGAGGTCTGACTACAAAGACCAT
GACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGATGGCCCCCAAGAAGAAGaggaagg
tgggcctcgaGCCCGGAGAAAAACCGTACAAGTGCCCTGAGTGCGGGAAATCATTCTCCGACCCTGGGGCGCTCGT
CCGGCACCAAAGGACGCATACAGGGGAAAAGCCGTATAAGTGCCCCGAGTGTGGAAAGAGCTTCTCGCAGAGAGCC
CACCTTGAACGACACCAAAGAACACACACTGGTGAGAAACCCTATAAGTGTCCAGAGTGCGGCAAATCGTTTAGCA
GATCCGATGACTTGGTGCGCCACCAGCGGACACACACGGGTGAAAAGCCCTACAAATGCCCGGAGTGTGGGAAGTC
GTTTTCAAGGTCGGATCATCTGACTACCCATCAGCGCACCCATACGGGAGCggccgcccgcgccctGGTGAAGAGC
GAGCTGGAGGAGAAGAAGTCCGAGCTGCGGCACAAGCTGAAGTACGTGCCCCACGAGTACATCGAGCTGATCGAGA
TCGCCAGGAACCCCACCCAGGACCGCATCCTGGAGATGAAGGTGATGGAGTTCTTCATGAAGGTGTACGGCTACAG
GGGAGAGCACCTGGGCGGAAGCAGAAAGCCTGACGGCGCCATCTATACAGTGGGCAGCCCCATCGATTACGGCGTG
ATCGTGGACACAAAGGCCTACAGCGGCGGCTACAATCTGCCTATCGGCCAGGCCGACGAGATGCAGAGATACGTGA
AGGAGAACCAGACCCGGAATAAGCACATCAACCCCAACGAGTGGTGGAAGGTGTACCCTAGCAGCGTGACCGAGTT
CAAGTTCCTGTTCGTGAGCGGCCACTTCAAGGGCAACTACAAGGCCCAGCTGACCAGGCTGAACCGCAAAACCAAC
TGCAATGGCGCCGTGCTGAGCGTGGAGGAGCTGCTGATCGGCGGCGAGATGATCAAAGCCGGCACCCTGACACTGG
AGGAGGTGCGGCGCAAGTTCAACAACGGCGAGATCAACTTCGAGGGCAGAGGAAGTCTTCTAACATGCGGTGACGT
GGAGGAGAATCCCGGCCCTAgatcTGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAG
GATGACGATGACAAGATGGCCCCCAAGAAGAAGaggaaggtgggcctcgagccGGGAGAGAAGCCGTACAAGTGTC
CCGAATGTGGAAAGAGCTTCTCACAGTCGGGGGACCTTCGGCGCCACCAGCGCACACATACTGGTGAAAAGCCGTA
TAAGTGTCCAGAATGTGGCAAATCATTCTCCACATCAGGGAGCCTGGTCAGGCACCAGCGAACCCACACGGGTGAG
AAGCCCTATAAGTGCCCCGAATGCGGGAAGTCCTTTTCGCAGAGAGCCCACTTGGAGAGGCACCAGAGGACCCATA
CGGGGGAGAAACCTTACAAGTGCCCTGAATGCGGAAAGTCGTTCTCGACCCATCTGGATCTCATCAGACATCAGAG
AACGCACACTGGAGAGAAACCCTACAAATGTCCCGAGTGTGGGAAGTCGTTTAGCCGAAAGGACAATCTCAAAAAC
CATCAACGGACACACACGGGTGAAAAACCATACAAATGCCCGGAGTGCGGCAAATCGTTTTCCCAACTTGCGCACT
TGCGGGCACACCAACGCACGCATACTGGAGCGGCCGCccgcgccCTGGTGAAGAGCGAGCTGGAGGAGAAGAAGTC
CGAGCTGCGGCACAAGCTGAAGTACGTGCCCCACGAGTACATCGAGCTGATCGAGATCGCCAGGAACCCCACCCAG
GACCGCATCCTGGAGATGAAGGTGATGGAGTTCTTCATGAAGGTGTACGGCTACAGGGGAGAGCACCTGGGCGGAA
GCAGAAAGCCTGACGGCGCCATCTATACAGTGGGCAGCCCCATCGATTACGGCGTGATCGTGGACACAAAGGCCTA
CAGCGGCGGCTACAATCTGCCTATCGGCCAGGCCGACGAGATGGAGAGATACGTGGAGGAGAACCAGACACGGGAT
AAGCACCTCAACCCCAACGAGTGGTGGAAGGTGTACCCTAGCAGCGTGACCGAGTTCAAGTTCCTGTTCGTGAGCG
GCCACTTCAAGGGCAACTACAAGGCCCAGCTGACCAGGCTGAACCACATCACCAACTGCAATGGCGCCGTGCTGAG

CGTGGAGGAGCTGCTGATCGGCGGCGAGATGATCAAAGCCGGCACCCTGACACTGGAGGAGGTGCGGCGCAAGTTC
AACAACGGCGAGATCAACTTCTGATTAATTAACTAATCTAGAGTcgactagagctcgctgatcagcctcgactgtg
ccttctagttgccagccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccactg
tcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctgggggggtggggtggg
gcaggacagcaaggggaggattgggaagacaatagcaggcatgctggggagagatctaggaacccctagtgatgg
agttggccactccctctctgcgcgctcgctcgctcactgaggccgcccgggcaaagcccgggcgtcgggcgacctt
tggtcgcccggcctcagtgagcgagcgagcgcgcagagagggagtggccaacccccccccccccccccctgcagcc
cagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctca
ctgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccac
agaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcg
ttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaa
acccgacaggactataaagataccaggcgtttcccccctggaagctccctcgtgcgctctcctgttccgaccctgcc
gcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcaatgctcacgctgtaggtatctc
agttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaacccccgttcagcccgaccgctgcgccttat
ccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggat
tagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaaggaca
gtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaa
ccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcc
tttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatca
aaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttt
ggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgc
ctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcga
gacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctg
caactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagttt
gcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggt
tcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcg
ttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgcc
atccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagt
tgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaac
gttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaa
ctgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaag
ggaataagggcgacacggaaatgttgaatactcatactcttccttttcaatattattgaagcatttatcagggtt
attgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggttccgcgcacatttccccg
aaaagtgccacctgacgtctaagaaaccattattatcatgacattaacctataaaaataggcgtatcacgaggccc
tttcgtctcgcgcgtttcggtgatgacggtgaaaacctctgacacatgcagctcccggagacggtcacagcttgtc
tgtaagcggatgccgggagcagacaagcccgtcagggcgcgtcagcgggtgttggcgggtgtcggggctggcttaa
ctatgcggcatcagagcagattgtactgagagtgcaccatatgcggtgtgaaataccgcacagatgcgtaaggaga
aaataccgcatcaggaaattgtaaacgttaatattttgttaaaattcgcgttaaatttttgttaaatcagctcatt
ttttaaccaataggccgaaatcggcaaaatcccttataaatcaaaagaatagaccgagatagggttgagtgttgtt

ccagtttggaacaagagtccactattaaagaacgtggactccaacgtcaaagggcgaaaaaccgtctatcagggcg
atggcccactacgtgaaccatcaccctaatcaagtttttttggggtcgaggtgccgtaaagcactaaatcggaaccc
taaagggagcccccgatttagagcttgacggggaaagccggcgaacgtggcgagaaaggaagggaagaaagcgaaa
ggagcgggcgctagggcgctggcaagtgtagcggtcacgctgcgcgtaaccaccacacccgccgcgcttaatgcgc
cgctacagggcgcgtcgcgccattcgccattcaggctacgcaactgttgggaagggcgatcggtgcgggcctcttc
gctattacgccagctggctgca

[0510] HS2增强子靶序列(SEQ ID NO:467)

[0511] Taagcttcagtttttccttagttcctgttacatttctgtgtgtctccattagtgacctcccatagtcc
aagcatgagcagttctggccaggcccctgtcggggtcagtgccccacccccgccttctggttctgtgtaaccttct
aagcaaaccttctggctcaagcacagcaatgctgagtcatgatgagtcatgctgaggcttagggtgtgtgcccaga
tgttctcagcctagagtgatgactcctatctgggtccccagcaggatgcttacagggcagatggcaaaaaaaagga
gaagctgaccacctgactaaaactccacctcaaacggcatcataaagaaaatggatgcctgagacagaatgtgaca
tattctagaatatatt

[0512] dSpCas9-KRAB序列(SEQ ID NO:466)

[0513] ATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAA
GatggcccccaagaagaagaggaaggtgggccgcggaATGGACAAGAAGTACTCCATTGGGCTCGCCATCGGCACA
AACAGCGTCGGCTGGGCCGTCATTACGGACGAGTACAAGGTGCCGAGCAAAAAATTCAAAGTTCTGGGCAATACCG
ATCGCCACAGCATAAAGAAGAACCTCATTGGCGCCCTCCTGTTCGACTCCGGGGAAACCGCCGAAGCCACGCGGCT
CAAAAGAACAGCACGGCGCAGATATACCCGCAGAAAGAATCGGATCTGCTACCtgcaGGAGATCTTTAGTAATGAG
ATGGCTAAGGTGGATGACTCTTTCTTCCATAGGCTGGAGGAGTCCTTTTTGGTGGAGGAGGATAAAAAGCACGAGC
GCCACCCAATCTTTGGCAATATCGTGGACGAGGTGGCGTACCATGAAAAGTACCCAACCATATATCATCTGAGGAA
GAAGCTTGTAGACAGTACTGATAAGGCTGACTTGCGGTTGATCTATCTCGCGCTGGCGCATATGATCAAATTTCGG
GGACACTTCCTCATCGAGGGGGACCTGAACCCAGACAACAGCGATGTCGACAAACTCTTTATCCAACTGGTTCAGA
CTTACAATCAGCTTTTCGAAGAGAACCCGATCAACGCATCCGGAGTTGACGCCAAAGCAATCCTGAGCGCTAGGCT
GTCCAAATCCCGGCGGCTCGAAAACCTCATCGCACAGCTCCCTGGGGAGAAGAAGAACGGCCTGTTTGGTAATCTT
ATCGCCCTGTCACTCGGGCTGACCCCCAACTTTAAATCTAACTTCGACCTGGCCGAAGATGCCAAGCTTCAACTGA
GCAAAGACACCTACGATGATGATCTCGACAATCTGCTGGCCCAGATCGGCGACCAGTACGCAGACCTTTTTTTGGC
GGCAAAGAACCTGTCAGACGCCATTCTGCTGAGTGATATTCTGCGAGTGAACACGGAGATCACCAAAGCTCCGCTG
AGCGCTAGTATGATCAAGCGCTATGATGAGCACCACCAAGACTTGACTTTGCTGAAGGCCCTTGTCAGACAGCAAC
TGCCTGAGAAGTACAAGGAAATTTTCTTCGATCAGTCTAAAAATGGCTACGCCGGATACATTGACGGCGGAGCAAG
CCAGGAGGAATTTTACAAATTTATTAAGCCCATCTTGGAAAAAATGGACGGCACCGAGGAGCTGCTGGTAAAGCTT
AACAGAGAAGATCTGTTGCGCAAACAGCGCACTTTCGACAATGGAAGCATCCCCCACCAGATTCACCTGGGCGAAC
TGCACGCTATCCTCAGGCGGCAAGAGGATTTCTACCCCTTTTTGAAAGATAACAGGGAAAAGATTGAGAAAATCCT
CACATTTCGGATACCCTACTATGTAGGCCCCCTCGCCCGGGGAAATTCCAGATTCGCGTGGATGACTCGCAAATCA
GAAGAGACCATCACTCCCTGGAACTTCGAGGAAGTCGTGGATAAGGGGGCCTCTGCCCAGTCCTTCATCGAAAGGA
TGACTAACTTTGATAAAAATCTGCCTAACGAAAAGGTGCTTCCTAAACACTCTCTGCTGTACGAGTACTTCACAGT
TTATAACGAGCTCACCAAGGTCAAATACGTCACAGAAGGGATGAGAAAGCCAGCATTCCTGTCTGGAGAGCAGAAG
AAAGCTATCGTGGACCTCCTCTTCAAGACGAACCGGAAAGTTACCGTGAAACAGCTCAAAGAAGACTATTTCAAAA
AGATTGAATGTTTCGACTCTGTTGAAATCAGCGGAGTGGAGGATCGCTTCAACGCATCCCTGGGAACGTATCACGA

TCTCCTGAAAATCATTAAAGACAAGGACTTCCTGGACAATGAGGAGAACGAGGACATTCTTGAGGACATTGTCCTC
ACCCTTACGTTGTTTGAAGATAGGGAGATGATTGAAGAACGCTTGAAAACTTACGCTCATCTCTTCGACGACAAAG
TCATGAAACAGCTCAAGAGGCGCCGATATACAGGATGGGGGCGGCTGTCAAGAAAACTGATCAATGGgatcCGAGA
CAAGCAGAGTGGAAAGACAATCCTGGATTTTCTTAAGTCCGATGGATTTGCCAACCGGAACTTCATGCAGTTGATC
CATGATGACTCTCTCACCTTTAAGGAGGACATCCAGAAAGCACAAGTTTCTGGCCAGGGGGACAGTCTTCACGAGC
ACATCGCTAATCTTGCAGGTAGCCCAGCTATCAAAAAGGGAATACTGCAGACCGTTAAGGTCGTGGATGAACTCGT
CAAAGTAATGGGAAGGCATAAGCCCGAGAATATCGTTATCGAGATGGCCCGAGAGAACCAAACTACCCAGAAGGGA
CAGAAGAACAGTAGGGAAAGGATGAAGAGGATTGAAGAGGGTATAAAAGAACTGGGGTCCCAAATCCTTAAGGAAC
ACCCAGTTGAAAACACCCAGCTTCAGAATGAGAAGCTCTACCTGTACTACCTGCAGAACGGCAGGGACATGTACGT
GGATCAGGAACTGGACATCAATCGGCTCTCCGACTACGACGTGGATGCCATCGTGCCCCAGTCTTTTCTCAAAGAT
GATTCTATTGATAATAAAGTGTTGACAAGATCCGATAAAAATAGAGGGAAGAGTGATAACGTCCCCTCAGAAGAAG
TTGTCAAGAAAATGAAAAATTATTGGCGGCAGCTGCTGAACGCCAAACTGATCACACAACGGAAGTTCGATAATCT
GACTAAGGCTGAACGAGGTGGCCTGTCTGAGTTGGATAAAGCCGGCTTCATCAAAAGGCAGCTTGTTGAGACACGC
CAGATCACCAAgcacGTGGCCCAAATTCTCGATTCACGCATGAACACCAAGTACGATGAAAATGACAAACTGATTC
GAGAGGTGAAAGTTATTACTCTGAAGTCTAAGCTGGTCTCAGATTTCAGAAAGGACTTTCAGTTTTATAAGGTGAG
AGAGATCAACAATTACCACCATGCGCATGATGCCTACCTGAATGCAGTGGTAGGCACTGCACTTATCAAAAAATAT
CCCAAGCTTGAATCTGAATTTGTTTACGGAGACTATAAAGTGTACGATGTTAGGAAAATGATCGCAAAGTCTGAGC
AGGAAATAGGCAAGGCCACCGCTAAGTACTTCTTTTACAGCAATATTATGAATTTTTTCAAGACCGAGATTACACT
GGCCAATGGAGAGATTCGGAAGCGACCACTTATCGAAACAAACGGAGAAACAGGAGAAATCGTGTGGGACAAGGGT
AGGGATTTCGCGACAGTCCGGAAGGTCCTGTCCATGCCGCAGGTGAACATCGTTAAAAAGACCGAAGTACAGACCG
GAGGCTTCTCCAAGGAAAGTATCCTCCCGAAAAGGAACAGCGACAAGCTGATCGCACGCAAAAAAGATTGGGACCC
CAAGAAATACGGCGGATTCGATTCTCCTACAGTCGCTTACAGTGTACTGGTTGTGGCCAAAGTGGAGAAAGGGAAG
TCTAAAAAACTCAAAAGCGTCAAGGAACTGCTGGGCATCACAATCATGGAGCGATCAAGCTTCGAAAAAAACCCCA
TCGACTTTCTCGAGGCGAAAGGATATAAAGAGGTCAAAAAAGACCTCATCATTAAGCTTCCCAAGTACTCTCTCTT
TGAGCTTGAAAACGGCCGGAAACGAATGCTCGCTAGTGCGGGCGAGCTGCAGAAAGGTAACGAGCTGGCACTGCCC
TCTAAATACGTTAATTTCTTGTATCTGGCCAGCCACTATGAAAAGCTCAAAGGGTCTCCCGAAGATAATGAGCAGA
AGCAGCTGTTCGTGGAACAACACAAACACTACCTTGATGAGATCATCGAGCAAATAAGCGAATTCTCCAAAAGAGT
GATCCTCGCCGACGCTAACCTCGATAAGGTGCTTTCTGCTTACAATAAGCACAGGGATAAGCCCATCAGGGAGCAG
GCAGAAAACATTATCCACTTGTTTACTCTGACCAACTTGGGCGCGCCTGCAGCCTTCAAGTACTTCGACACCACCA
TAGACAGAAAGCGGTACACCTCTACAAAGGAGGTCCTGGACGCCACACTGATTCATCAGTCAATTACGGGGCTCTA
TGAAACAAGAATCGACCTCTCTCAGCTCGGTGGAGACAGCAGGGCTGACCCCAAGAAGAAGAGGAAGGTGGctagC
GATGCTAAGTCACTGACTGCCTGGTCCCGGACACTGGTGACCTTCAAGGATGTGTTTGTGGACTTCACCAGGGAGG
AGTGGAAGCTGCTGGACACTGCTCAGCAGATCCTGTACAGAAATGTGATGCTGGAGAACTATAAGAACCTGGTTTC
CTTGGGTTATCAGCTTACTAAGCCAGATGTGATCCTCCGGTTGGAGAAGGGAGAAGAGCCCTGGCTGGTGGAGAGA
GAAATTCACCAAGAGACCCATCCTGATTCAGAGACTGCATTTGAAATCAAATCATCAGTTCCGAAAAAGAAACGCA
AAGttGctagCG

序列表

<110> 杜克大学
<120> RNA 指导的基因编辑和基因调节
<130> 028193-9164-W000
<140>
<141>
<150> PCT/US2014/041190
<151> 2014-06-05
<150> 61/831,481
<151> 2013-06-05
<150> 61/839,127
<151> 2013-06-25
<150> 61/904,911
<151> 2013-11-15
<150> 61/967,466
<151> 2014-03-19
<150> 61/981,575
<151> 2014-04-18
<160> 676
<170> PatentIn version 3.5
<210> 1
<211> 1483
<212> PRT
[0001]
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多肽
<400> 1

```
Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp
1               5                  10                  15
Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys Arg Lys Val
            20                  25                  30
Gly Arg Gly Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr
        35                  40                  45
Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser
    50                  55                  60
Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys
65                  70                  75                  80
Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala
                85                  90                  95
Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn
            100                 105                 110
Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val
        115                 120                 125
Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu
    130                 135                 140
```

[0002]

```
Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu
145                 150                 155                 160
Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys
                165                 170                 175
Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala
            180                 185                 190
Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp
        195                 200                 205
Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val
    210                 215                 220
Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly
225                 230                 235                 240
Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg
                245                 250                 255
Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu
            260                 265                 270
Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys
        275                 280                 285
Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp
    290                 295                 300
Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln
305                 310                 315                 320
Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu
                325                 330                 335
Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu
            340                 345                 350
Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr
        355                 360                 365
Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu
    370                 375                 380
Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly
385                 390                 395                 400
Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu
                405                 410                 415
Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp
            420                 425                 430
Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln
        435                 440                 445
Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe
    450                 455                 460
Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr
465                 470                 475                 480
Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg
                485                 490                 495
Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn
            500                 505                 510
```

[0003]

```
Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu
        515                 520                 525
Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro
    530                 535                 540
Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr
545                 550                 555                 560
Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser
                565                 570                 575
Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg
            580                 585                 590
Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu
        595                 600                 605
Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala
    610                 615                 620
Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp
625                 630                 635                 640
Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu
                645                 650                 655
Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys
            660                 665                 670
Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg
        675                 680                 685
Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly
    690                 695                 700
Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser
705                 710                 715                 720
Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser
                725                 730                 735
Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly
            740                 745                 750
Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile
        755                 760                 765
Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys
    770                 775                 780
Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg
785                 790                 795                 800
Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met
                805                 810                 815
Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys
            820                 825                 830
Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu
        835                 840                 845
Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp
    850                 855                 860
Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser
865                 870                 875                 880
```

[0004]

```
Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp
                885                 890                 895
Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys
            900                 905                 910
Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr
        915                 920                 925
Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser
    930                 935                 940
Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg
945                 950                 955                 960
Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr
                965                 970                 975
Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr
            980                 985                 990
Leu Lys Ser Lys Leu Val Ser Asp  Phe Arg Lys Asp Phe  Gln Phe Tyr
        995                 1000                1005
Lys Val  Arg Glu Ile Asn Asn  Tyr His His Ala His  Asp Ala Tyr
    1010                1015                1020
Leu Asn  Ala Val Val Gly Thr  Ala Leu Ile Lys Lys  Tyr Pro Lys
    1025                1030                1035
Leu Glu  Ser Glu Phe Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val
    1040                1045                1050
Arg Lys  Met Ile Ala Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr
    1055                1060                1065
Ala Lys  Tyr Phe Phe Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr
    1070                1075                1080
Glu Ile  Thr Leu Ala Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile
    1085                1090                1095
Glu Thr  Asn Gly Glu Thr Gly  Glu Ile Val Trp Asp  Lys Gly Arg
    1100                1105                1110
Asp Phe  Ala Thr Val Arg Lys  Val Leu Ser Met Pro  Gln Val Asn
    1115                1120                1125
Ile Val  Lys Lys Thr Glu Val  Gln Thr Gly Gly Phe  Ser Lys Glu
    1130                1135                1140
Ser Ile  Leu Pro Lys Arg Asn  Ser Asp Lys Leu Ile  Ala Arg Lys
    1145                1150                1155
Lys Asp  Trp Asp Pro Lys Lys  Tyr Gly Gly Phe Asp  Ser Pro Thr
    1160                1165                1170
Val Ala  Tyr Ser Val Leu Val  Val Ala Lys Val Glu  Lys Gly Lys
    1175                1180                1185
Ser Lys  Lys Leu Lys Ser Val  Lys Glu Leu Leu Gly  Ile Thr Ile
    1190                1195                1200
Met Glu  Arg Ser Ser Phe Glu  Lys Asn Pro Ile Asp  Phe Leu Glu
    1205                1210                1215
Ala Lys  Gly Tyr Lys Glu Val  Lys Lys Asp Leu Ile  Ile Lys Leu
    1220                1225                1230
```

Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met
1235 1240 1245
Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu
1250 1255 1260
Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu
1265 1270 1275
Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe
1280 1285 1290
Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile
1295 1300 1305
Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp
1310 1315 1320
Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg
1325 1330 1335
Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu
1340 1345 1350
Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg
1355 1360 1365
Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile
1370 1375 1380
His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser
1385 1390 1395
[0005] Gln Leu Gly Gly Asp Pro Ile Ala Gly Ser Lys Ala Ser Pro Lys
1400 1405 1410
Lys Lys Arg Lys Val Gly Arg Ala Asp Ala Leu Asp Asp Phe Asp
1415 1420 1425
Leu Asp Met Leu Gly Ser Asp Ala Leu Asp Asp Phe Asp Leu Asp
1430 1435 1440
Met Leu Gly Ser Asp Ala Leu Asp Asp Phe Asp Leu Asp Met Leu
1445 1450 1455
Gly Ser Asp Ala Leu Asp Asp Phe Asp Leu Asp Met Leu Ile Asn
1460 1465 1470
Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Ser
1475 1480

<210> 2
<211> 350
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 2

gagggcctat ttcccatgat tccttcatat ttgcatatac gatacaaggc tgttagagag 60
ataattggaa ttaatttgac tgtaaacaca aagatattag tacaaaatac gtgacgtaga 120
aagtaataat ttcttgggta gtttgcagtt ttaaaattat gttttaaaat ggactatcat 180
atgcttatcg taacttgaaa gtatttcgat ttcttggctt tatatatctt gtggaaagga 240

cgaaacaccg ggtcttcgag aagacctgtt ttagagctag aaatagcaag ttaaaataag 300
gctagtccgt tatcaacttg aaaaagtggc accgagtcgg tgctttttt 350
<210> 3
<211> 338
<212> DNA
<213> 智人
<400> 3
acccccggca gcagcccgcc ccgagcgcgc cgcctgttta ttcagccggg agtccggcac 60
gcgccaggcg cacgcactgc aacaacaaac ccagctgaat ggagagtttg caaggagcgg 120
gagaaaggaa cgggaggggg ggagaggaga ggaggagggg gagtttaggg agtgggtggg 180
aggaagaggt aagaggaggg gggggagtgg gggctgcagc cgctcgctgc agcagcgggg 240
agtgggggggc gaggcggggc cagggctgcg cgtggggctg ggtgtcccat tgaaaaggcg 300
gacgcactcc ggcagcccag cactctctca cttctggc 338
<210> 4
<211> 305
<212> DNA
<213> 小鼠
<400> 4
ccgccccgag cgcgcggcct gtttattcag ccgggagtcc ggcacgcgcc gggcgcacgc 60
actgcaacaa caaaccaggc tgaatggaga gtttgcaagg agcgggcgcg ggcaactgga 120
gggggggggg gcgagaggga gggagctgag gaggtggggg aagaggaggg gtagtggggg 180
ctgcagccgc tcgctgcagc agcggggagt ggggggcgag gcggggccag ggctgcgcgt 240
ggggctgggt gtcccattga aaaggcggcc gcaccgcagc cgcccagcag tctctcactt 300
[0006] ctggc 305
<210> 5
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 5
gctgggtgtc ccattgaaa 19
<210> 6
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 6
cagccgctcg ctgcagcag 19
<210> 7
<211> 19
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 7
tggagagttt gcaaggagc 19
<210> 8
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 8
gtttattcag ccgggagtc 19
<210> 9
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 9
[0007] cgccaggagg ggtgggtcta 20
<210> 10
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 10
ccttggtgag actggtaga 19
<210> 11
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 11
gtcttcaggt tctgttgct 19
<210> 12
<211> 19
<212> DNA
<213> 人工序列
<220>

<223> 人工序列的描述：合成
寡核苷酸
<400> 12
atattcctga tttaaaagt 19
<210> 13
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 13
ttaaaagtcg gctggtagc 19
<210> 14
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 14
cgggccgggg gcggggtcc 19

[0008]

<210> 15
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 15
gcccgagccg cgtgtggaa 19
<210> 16
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 16
ccttcattgc ggcgggctg 19
<210> 17
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成

寡核苷酸

<400> 17

ccgacccctc ccgggtccc 19

<210> 18

<211> 19

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成

寡核苷酸

<400> 18

caggaccgcg cttcccacg 19

<210> 19

<211> 19

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成

寡核苷酸

<400> 19

tgcaccctgg gagcgcgag 19

<210> 20

<211> 19

[0009]

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成

寡核苷酸

<400> 20

ccgcacgcac ctgttccca 19

<210> 21

<211> 19

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成

寡核苷酸

<400> 21

aaaacagcga gggagaaac 19

<210> 22

<211> 19

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成

寡核苷酸

<400> 22
ttaacttgat tgtgaaatc 19
<210> 23
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 23
aaaacaatgc atatttgca 19
<210> 24
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 24
aaaatccagt attttaatg 19
<210> 25
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 25
acccagcact gcagcctgg 19
<210> 26
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 26
aacttatgcg gcgtttcct 19
<210> 27
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 27

[0010]

tcactttaaa accacctct 19

<210> 28

<211> 19

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 28

gcatcttttt ctctttaat 19

<210> 29

<211> 19

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 29

tgtactctct gaggtgctc 19

<210> 30

<211> 19

<212> DNA

<213> 人工序列

[0011]

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 30

acgcagataa gaaccagtt 19

<210> 31

<211> 19

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 31

catcaagtca gccatcagc 19

<210> 32

<211> 19

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 32

gagtcaccct cctggaaac 19

<210> 33
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 33
gctagggatg aagaataaa 19
<210> 34
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 34
ttgaccaata gccttgaca 19
<210> 35
<211> 19
<212> DNA
<213> 人工序列
<220>
[0012]
<223> 人工序列的描述：合成
寡核苷酸
<400> 35
tgcaaatatc tgtctgaaa 19
<210> 36
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 36
aaattagcag tatcctctt 19
<210> 37
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 37
cctgggctcc ggggcgttt 19
<210> 38

[0013]

<210> 19
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 38
ggcccctgcg gccaccccg 19
<210> 39
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 39
ctccctccct gcccggtag 19
<210> 40
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 40
aggtttggaa agggcgtgc 19
<210> 41
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 41
ggagcttctc gacttcacca 20
<210> 42
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 42
gatttgtggg cctgaagaaa 20
<210> 43
<211> 20

[0014]

<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 43
aaggaggagg gcagaatcat 20
<210> 44
<211> 21
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 44
aaaccttcct cagctatgcc c 21
<210> 45
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 45
agctgatggc cctaaacaga 20
<210> 46
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 46
caggaggact ctggcaccta 20
<210> 47
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 47
ggaatccatg gagggaagat 20
<210> 48
<211> 21
<212> DNA

<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 48
gctgagtgaa ctgcactgtg a 21
<210> 49
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 49
ctctctgctc ctttgccaca 20
<210> 50
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 50
[0015]
caatgacccc ttcattgacc 20
<210> 51
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 51
ggaacaagag ctgctggact 20
<210> 52
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 52
aactttggca ttgtggaagg 20
<210> 53
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成
引物
<400> 53
aacgccactg acaagaaagc 20
<210> 54
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 54
cagatccatg gaggaaggaa 20
<210> 55
<211> 21
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 55
[0016] gggtactcct ggaagatgtc c 21
<210> 56
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 56
gtttgcgacg catgttcctc 20
<210> 57
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 57
aagcccttgc tgtagtggtg 20
<210> 58
<211> 20
<212> DNA
<213> 人工序列
<220>

<223> 人工序列的描述：合成
引物
<400> 58
cggcaggaaa gcatctgtat 20
<210> 59
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 59
tgttctcgct caggtcagtg 20
<210> 60
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 60
gaattctttg ccgaaatgga 20

[0017]

<210> 61
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 61
gtgctcttcg ggtttcagga 20
<210> 62
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 62
ttgattttgg agggatctcg 20
<210> 63
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成

引物

<400> 63

gtttttctgc ctccccattt 20

<210> 64

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成引物

<400> 64

ggatgcaggg atgatgttct 20

<210> 65

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 65

gattggcttt gatttcccta 20

<210> 66

<211> 23

[0018]

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 66

gtgtagagta agtcagccta tgg 23

<210> 67

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 67

gcctactcag actgttactc 20

<210> 68

<211> 23

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

[0019]

<400> 68
gttggacaga acttaccgac tgg 23
<210> 69
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 69
gcagttgcct aagaactggt 20
<210> 70
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 70
ggggctccac cctcacgagt 20
<210> 71
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 71
gtttgcttcg ctataaaacg agg 23
<210> 72
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 72
gtctgaggat ggggccgcaa tgg 23
<210> 73
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 73

[0020]

ggatctgtca aatcgcctgc agg 23

<210> 74

<211> 23

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 74

gccaggatgg cattgggcag cgg 23

<210> 75

<211> 23

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 75

gctgaatctg cggtggcagg agg 23

<210> 76

<211> 23

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 76

gttcttttgt tcttctagcc tgg 23

<210> 77

<211> 23

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 77

ggaaaagctt gagcaagtca agg 23

<210> 78

<211> 23

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 78

ggaagagttg cccctgcgcc agg 23

[0021]

<210> 79
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 79
gacaaatctc cagtggataa agg 23
<210> 80
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 80
gtgtttctca ggtaaagctc tgg 23
<210> 81
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 81
ggaaggacca tttgacgtta agg 23
<210> 82
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 82
gaactgctat ttcagtttcc tgg 23
<210> 83
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 83
gccagccact cagccagtga agg 23
<210> 84

[0022]

<210> 23
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 84
ggtatgcttt tctgttaaag agg 23
<210> 85
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 85
gctcctggac tgaccactat tgg 23
<210> 86
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 86
ggaacagagg cgtccccagt tgg 23
<210> 87
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 87
ggaggctaga acaatcatta cgg 23
<210> 88
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 88
gacaagaaca ccttcagaac cgg 23
<210> 89
<211> 23

[0023]

<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 89
gggtttctgt gattttcttt tgg 23
<210> 90
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 90
gggccaaaga cctccgccag tgg 23
<210> 91
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 91
gttggagaag cattcataaa agg 23
<210> 92
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 92
gtcgctcact caccctgcaa agg 23
<210> 93
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 93
gaaaagagct gatgaaacaa tgg 23
<210> 94
<211> 23
<212> DNA

<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 94
gtacactttt caaaatgctt tgg 23
<210> 95
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 95
ggagatgatc atcaagcaga agg 23
<210> 96
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 96
[0024]
gctttgaaag agcaataaaa tgg 23
<210> 97
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 97
gcacaaaagt caaatcggaa tgg 23
<210> 98
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 98
gatttcaata taagattcgg agg 23
<210> 99
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 99
gcttaagcaa tcccgaactc tgg 23
<210> 100
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 100
gccttcttta tcccctatcg 20
<210> 101
<400> 101
000
<210> 102
<400> 102
000
<210> 103
<400> 103
000

[0025]

<210> 104
<211> 25
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<220>
<221> 修饰的碱基
<222> (21)..(22)
<223> a, c, t, g, 未知或其它
<400> 104
gaggccaaac ctcggcttac nngrr 25
<210> 105
<211> 25
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<220>
<221> 修饰的碱基
<222> (21)..(22)

<223> a, c, t, g, 未知或其它
<400> 105
gttcgaaaat ttcaggtaag nngrr 25
<210> 106
<211> 26
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<220>
<221> 修饰的碱基
<222> (21)..(22)
<223> a, c, t, g, 未知或其它
<400> 106
ggcagaacag gagataacag nngrrt 26
<210> 107
<211> 26
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
[0026]
<400> 107
ggcggccctc gcccttctct ggggat 26
<210> 108
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 108
gtagtgatcg tggatacgag agg 23
<210> 109
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 109
gtacagccct cggtgtatat tgg 23
<210> 110
<211> 23
<212> DNA

<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 110
gggaaggaat taagcccgaa tgg 23
<210> 111
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 111
gggaacagct ttcgtagttg agg 23
<210> 112
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 112
[0027]
gataaagtcc agtgtcgatc agg 23
<210> 113
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 113
gaaaaccaga gcttcggtca agg 23
<210> 114
<211> 34
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 114
ggagtcttct gggcaggctt aaaggctaac ctgg 34
<210> 115
<211> 34
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 115
gtcgggtgag catgtcttta atctacctcg atgg 34
<210> 116
<211> 26
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 116
ggtgtcacca gagtaacagt ctgagt 26
<210> 117
<211> 25
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 117
[0028] gtgatcatca agcagaaggt atgag 25
<210> 118
<211> 29
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 118
gaacttcgaa aatttcaggt aagccgagg 29
<210> 119
<211> 29
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 119
ggaaactcat caaatatgcg tgttagtgt 29
<210> 120
<211> 29
<212> DNA
<213> 人工序列
<220>

[0029]

<223> 人工序列的描述：合成
寡核苷酸
<400> 120
gtcatttaca ctaacacgca tatttgatg 29
<210> 121
<211> 29
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 121
ggaatgaaac tcatcaaata tgcgtgtta 29
<210> 122
<211> 29
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 122
gtcatcaata tctttgaagg actctgggt 29
<210> 123
<211> 29
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 123
gtgttttcat aggaaaaata ggcaagttg 29
<210> 124
<211> 29
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 124
gaattggaaa atgtgatggg aaacagata 29
<210> 125
<211> 29
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成

寡核苷酸

<400> 125

gatgatcatc aagcagaagg tatgagaaa 29

<210> 126

<211> 29

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成

寡核苷酸

<400> 126

gagatgatca tcaagcagaa ggtatgaga 29

<210> 127

<211> 29

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成

寡核苷酸

<400> 127

gcatttttc tcataccttc tgcttgatg 29

<210> 128

<211> 29

[0030]

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成

寡核苷酸

<400> 128

gtcctactca gactgttact ctggtgaca 29

<210> 129

<211> 29

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成

寡核苷酸

<400> 129

gacaggttgt gtcaccagag taacagtct 29

<210> 130

<211> 29

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成

寡核苷酸

<400> 130
gttatcattt tttctcatac cttctgctt 29
<210> 131
<211> 29
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 131
gttgcctaag aactggtggg aaatggtct 29
<210> 132
<211> 29
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 132
gaaacagttg cctaagaact ggtgggaaa 29
<210> 133
<211> 29
<212> DNA
[0031] <213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 133
gtttcccacc agttcttagg caactgttt 29
<210> 134
<211> 29
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 134
gtggctttga tttccctagg gtccagctt 29
<210> 135
<211> 29
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 135

gtagggaaat caaagccaat gaaacgttc 29

<210> 136
<211> 29
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 136
ggaccctagg gaaatcaaag ccaatgaaa 29

<210> 137
<211> 29
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 137
gtgagggctc caccctcacg agtgggttt 29

<210> 138
<211> 29
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 138
gaaggattga gggctccacc ctcacgagt 29

<210> 139
<211> 29
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 139
ggctccaccc tcacgagtgg gtttggttc 29

<210> 140
<211> 29
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 140
gtatccccta tcgaggaaac cacgagttt 29

[0032]

<210> 141
<211> 29
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 141
ggataaagaa ggcctatttc atagagttg 29
<210> 142
<211> 29
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 142
gaggccttct ttatccccta tcgaggaaa 29
<210> 143
<211> 27
<212> DNA
<213> 人工序列
<220>
[0033]
<223> 人工序列的描述：合成
寡核苷酸
<400> 143
gtgagggctc caccctcacg agtgggt 27
<210> 144
<211> 27
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 144
ggataaagaa ggcctatttc atagagt 27
<210> 145
<211> 4932
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 145
gccaccatgg acaagaagta ctccattggg ctcgatatcg gcacaaacag cgtcggctgg 60
gccgtcatta cggacgagta caaggtgccg agcaaaaaat tcaaagttct gggcaatacc 120

[0034]

gatcgccaca gcataaagaa gaacctcatt ggcgccctcc tgttcgactc cggggagacg 180
gccgaagcca cgcggctcaa aagaacagca cggcgcagat atacccgcag aaagaatcgg 240
atctgctacc tgcaggagat ctttagtaat gagatggcta aggtggatga ctctttcttc 300
cataggctgg aggagtcctt tttggtggag gaggataaaa agcacgagcg ccacccaatc 360
tttggcaata tcgtggacga ggtggcgtac catgaaaagt acccaaccat atatcatctg 420
aggaagaagc ttgtagacag tactgataag gctgacttgc ggttgatcta tctcgcgctg 480
gcgcatatga tcaaatttcg gggacacttc ctcatcgagg gggacctgaa cccagacaac 540
agcgatgtcg acaaactctt tatccaactg gttcagactt acaatcagct tttcgaagag 600
aacccgatca acgcatccgg agttgacgcc aaagcaatcc tgagcgctag gctgtccaaa 660
tcccggcggc tcgaaaacct catcgcacag ctccctgggg agaagaagaa cggcctgttt 720
ggtaatctta tcgccctgtc actcgggctg acccccaact ttaaatctaa cttcgacctg 780
gccgaagatg ccaagcttca actgagcaaa gacacctacg atgatgatct cgacaatctg 840
ctggcccaga tcggcgacca gtacgcagac ctttttttgg cggcaaagaa cctgtcagac 900
gccattctgc tgagtgatat tctgcgagtg aacacggaga tcaccaaagc tccgctgagc 960
gctagtatga tcaagcgcta tgatgagcac caccaagact tgactttgct gaaggccctt 1020
gtcagacagc aactgcctga gaagtacaag gaaattttct tcgatcagtc taaaaatggc 1080
tacgccggat acattgacgg cggagcaagc caggaggaat tttacaaatt tattaagccc 1140
atcttggaaa aaatggacgg caccgaggag ctgctggtaa agcttaacag agaagatctg 1200
ttgcgcaaac agcgcacttt cgacaatgga agcatccccc accagattca cctgggcgaa 1260
ctgcacgcta tcctcaggcg gcaagaggat ttctaccccc ttttgaaaga taacagggaa 1320
aagattgaga aaatcctcac atttcggata ccctactatg taggcccccc cgcccgggga 1380
aattccagat tcgcgtggat gactcgcaaa tcagaagaga ccatcactcc ctggaacttc 1440
gaggaagtcg tggataaggg ggcctctgcc cagtccttca tcgaaaggat gactaacttt 1500
gataaaaatc tgcctaacga aaaggtgctt cctaaacact ctctgctgta cgagtacttc 1560
acagtttata acgagctcac caaggtcaaa tacgtcacag aagggatgag aaagccagca 1620
ttcctgtctg gagagcagaa gaaagctatc gtggacctcc tcttcaagac gaaccggaaa 1680
gttaccgtga aacagctcaa agaagactat ttcaaaaaga ttgaatgttt cgactctgtt 1740
gaaatcagcg gagtggagga tcgcttcaac gcatccctgg gaacgtatca cgatctcctg 1800
aaaatcatta aagacaagga cttcctggac aatgaggaga acgaggacat tcttgaggac 1860
attgtcctca cccttacgtt gtttgaagat agggagatga ttgaagaacg cttgaaaact 1920
tacgctcatc tcttcgacga caaagtcatg aaacagctca agaggcgccg atatacagga 1980
tgggggcggc tgtcaagaaa actgatcaat gggatccgag acaagcagag tggaaagaca 2040
atcctggatt ttcttaagtc cgatggattt gccaaccgga acttcatgca gttgatccat 2100
gatgactctc tcacctttaa ggaggacatc cagaaagcac aagtttctgg ccaggggGac 2160
agtcttcacg agcacatcgc taatcttgca ggtagcccag ctatcaaaaa gggaatactg 2220
cagaccgtta aggtcgtgga tgaactcgtc aaagtaatgg gaaggcataa gcccgagaat 2280
atcgttatcg agatggcccg agagaaccaa actacccaga agggacagaa gaacagtagg 2340
gaaaggatga agaggattga agagggtata aaagaactgg ggtcccaaat ccttaaggaa 2400
cacccagttg aaaacaccca gcttcagaat gagaagctct acctgtacta cctgcagaac 2460
ggcagggaca tgtacgtgga tcaggaactg gacatcaatc ggctctccga ctacgacgtg 2520
gatcatatcg tgccccagtc ttttctcaaa gatgattcta ttgataataa agtgttgaca 2580
agatccgata aaaatagagg gaagagtgat aacgtcccct cagaagaagt tgtcaagaaa 2640
atgaaaaatt attggcggca gctgctgaac gccaaactga tcacacaacg gaagttcgat 2700
aatctgacta aggctgaacg aggtggcctg tctgagttgg ataaagccgg cttcatcaaa 2760
aggcagcttg ttgagacacg ccagatcacc aagcacgtgg cccaaattct cgattcacgc 2820
atgaacacca agtacgatga aaatgacaaa ctgattcgag aggtgaaagt tattactctg 2880

[0035]

aagtctaagc tggtctcaga tttcagaaag gactttcagt tttataaggt gagagagatc 2940
aacaattacc accatgcgca tgatgcctac ctgaatgcag tggtaggcac tgcacttatc 3000
aaaaaatatc ccaagcttga atctgaattt gtttacggag actataaagt gtacgatgtt 3060
aggaaaatga tcgcaaagtc tgagcaggaa ataggcaagg ccaccgctaa gtacttcttt 3120
tacagcaata ttatgaattt tttcaagacc gagattacac tggccaatgg agagattcgg 3180
aagcgaccac ttatcgaaac aaacggagaa acaggagaaa tcgtgtggga caagggtagg 3240
gatttcgcga cagtccggaa ggtcctgtcc atgccgcagg tgaacatcgt taaaaagacc 3300
gaagtacaga ccggaggctt ctccaaggaa agtatcctcc cgaaaaggaa cagcgacaag 3360
ctgatcgcac gcaaaaaaga ttgggacccc aagaaatacg gcggattcga ttctcctaca 3420
gtcgcttaca gtgtactggt tgtggccaaa gtggagaaag ggaagtctaa aaaactcaaa 3480
agcgtcaagg aactgctggg catcacaatc atggagcgat caagcttcga aaaaaacccc 3540
atcgactttc tcgaggcgaa aggatataaa gaggtcaaaa aagacctcat cattaagctt 3600
cccaagtact ctctctttga gcttgaaaac ggccggaaac gaatgctcgc tagtgcgggc 3660
gagctgcaga aaggtaacga gctggcactg ccctctaaat acgttaattt cttgtatctg 3720
gccagccact atgaaaagct caaagggtct cccgaagata atgagcagaa gcagctgttc 3780
gtggaacaac acaaacacta ccttgatgag atcatcgagc aaataagcga attctccaaa 3840
agagtgatcc tcgccgacgc taacctcgat aaggtgcttt ctgcttacaa taagcacagg 3900
gataagccca tcagggagca ggcagaaaac attatccact tgtttactct gaccaacttg 3960
ggcgcgcctg cagccttcaa gtacttcgac accaccatag acagaaagcg gtacacctct 4020
acaaaggagg tcctggacgc cacactgatt catcagtcaa ttacggggct ctatgaaaca 4080
agaatcgacc tctctcagct cggtggagac agcagggctg accccaagaa gaagaggaag 4140
gtggctagcg agggcagagg aagtcttcta acatgcggtg acgtggagga gaatcccggc 4200
cctggtaccg tgagcaaggg cgaggagctg ttcaccgggg tggtgcccat cctggtcgag 4260
ctggacggcg acgtaaacgg ccacaagttc agcgtgtccg gcgagggcga gggcgatgcc 4320
acctacggca agctgaccct gaagttcatc tgcaccaccg gcaagctgcc cgtgccctgg 4380
cccaccctcg tgaccaccct gacctacggc gtgcagtgct tcagccgcta ccccgaccac 4440
atgaagcagc acgacttctt caagtccgcc atgcccgaag gctacgtcca ggagcgcacc 4500
atcttcttca aggacgacgg caactacaag acccgcgccg aggtgaagtt cgagggcgac 4560
accctggtga accgcatcga gctgaagggc atcgacttca aggaggacgg caacatcctg 4620
gggcacaagc tggagtacaa ctacaacagc cacaacgtct atatcatggc cgacaagcag 4680
aagaacggca tcaaggtgaa cttcaagatc cgccacaaca tcgaggacgg cagcgtgcag 4740
ctcgccgacc actaccagca gaacaccccc atcggcgacg gccccgtgct gctgcccgac 4800
aaccactacc tgagcaccca gtccgccctg agcaaagacc ccaacgagaa gcgcgatcac 4860
atggtcctgc tggagttcgt gaccgccgcc gggatcactc tcggcatgga cgagctgtac 4920
aagaccggtt ag 4932

<210> 146
<400> 146
000
<210> 147
<400> 147
000
<210> 148
<400> 148
000
<210> 149
<400> 149

000

<210> 150

<211> 23

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 150

tcctactcag actgttactc tgg 23

<210> 151

<211> 23

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 151

tcctactcac actgttactc agg 23

<210> 152

<211> 23

<212> DNA

<213> 人工序列

[0036]

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 152

acctgctcac actgttactc cag 23

<210> 153

<211> 23

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 153

gcattctcaa actgttactc agg 23

<210> 154

<211> 23

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 154

ggattctcac actgttactc ggg 23

[0037]

<210> 155
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 155
acatacttat actgttactc tag 23
<210> 156
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 156
tattcctaag actgttactc aag 23
<210> 157
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 157
aaggactaag actgttactc ggg 23
<210> 158
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 158
gagctctcat actgttactc tag 23
<210> 159
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 159
gcaaaatgag actgttactc cag 23
<210> 160

[0038]

<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 160
cctcattcag actgttactc aag 23
<210> 161
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 161
cattggcttt gatttcccta ggg 23
<210> 162
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 162
aattggcatt gatttcccta gag 23
<210> 163
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 163
cattggcttt aatttcccta tag 23
<210> 164
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 164
gataggctgt gatttcccta gag 23
<210> 165
<211> 23

[0039]

<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 165
gaatagcctt gatttcccta aag 23
<210> 166
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 166
aatttgcttt gatttccctg agg 23
<210> 167
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 167
gatgtgcttt gatttccctt ggg 23
<210> 168
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 168
aattggtttt aatttcccta aag 23
<210> 169
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 169
aattgggttt gatttccctt tgg 23
<210> 170
<211> 23
<212> DNA

<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 170
gatgggtttt tatttcccta gag 23
<210> 171
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 171
gaatggtttt gatttccctg gag 23
<210> 172
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 172
[0040]
acagttgcct aagaactggt ggg 23
<210> 173
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 173
ccagttgtct aagaactggg gag 23
<210> 174
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 174
gcagttgcct gtgaactggt agg 23
<210> 175
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 175
gcagatgcag aagaactggt gag 23
<210> 176
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 176
gcagttccag aagaactggt gag 23
<210> 177
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 177
[0041] caacttgcct atgaactggt agg 23
<210> 178
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 178
acacctgcct aagaactgga ggg 23
<210> 179
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 179
tcaggtggct aagaactggg tgg 23
<210> 180
<211> 23
<212> DNA
<213> 人工序列
<220>

<223> 人工序列的描述：合成
寡核苷酸
<400> 180
gaagttggcc aagaactgga gag 23
<210> 181
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 181
gctgctgccc aagaactggc agg 23
<210> 182
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 182
tcagctggct aagaacgggt aag 23

[0042]

<210> 183
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 183
agggctccac cctcacgagt ggg 23
<210> 184
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 184
gcagctcagc cctcacgagt cag 23
<210> 185
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成

寡核苷酸

<400> 185

ggggcttcag catcacgagt gag 23

<210> 186

<211> 23

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 186

ggggctctcc cctcactagt gag 23

<210> 187

<211> 23

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 187

ggggatccac cttcaccagt cag 23

<210> 188

<211> 23

[0043]

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 188

agggctggac cctcacaagt aag 23

<210> 189

<211> 23

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 189

tggtctcctc ccccacgagt ggg 23

<210> 190

<211> 23

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 190
agggctccca ccccacgagt gag 23
<210> 191
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 191
gaggctccat actcaccagt gag 23
<210> 192
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 192
ggagctgccc cttcacgagt ggg 23
<210> 193
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 193
atgactccac cctcaagagt aag 23
<210> 194
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 194
gccttcttta tcccctatcg agg 23
<210> 195
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 195

[0044]

gtctgctgtg tcccctatcg ggg 23

<210> 196

<211> 23

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 196

ccttctcta tcccctgtcg tgg 23

<210> 197

<211> 23

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 197

gccttcttta tcccctctct tgg 23

<210> 198

<211> 23

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 198

gcgctctttt tcccctatct tag 23

<210> 199

<211> 23

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 199

gccctctgtc tcccctgtcg cag 23

<210> 200

<211> 23

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 200

tccatctttg tcccctattg agg 23

[0045]

[0046]

<210> 201
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 201
accttctctc tcccctatag agg 23
<210> 202
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 202
gttttctttt tcccctatgg gag 23
<210> 203
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 203
tgcttcttaa tcccctatca aag 23
<210> 204
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 204
accttcttac tcccctatcc ggg 23
<210> 205
<211> 22
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 205
gagaggttat gtggctttac ca 22
<210> 206

[0047]

<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 206
aaaaatgctt cccactttgc 20
<210> 207
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 207
ctcattctca tgcctggaca 20
<210> 208
<211> 25
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 208
gagtttggct caaattgtta ctctt 25
<210> 209
<211> 25
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 209
gggaaatggt ctaggagagt aaagt 25
<210> 210
<211> 25
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 210
gtttggctca aattgttact cttca 25
<210> 211
<211> 25

<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 211
gtgagagtaa tgtgtttgct gagag 25
<210> 212
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 212
cgggcttgga cagaacttac 20
<210> 213
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
[0048]
<400> 213
ctgcgtagtg ccaaaacaaa 20
<210> 214
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 214
taatttcatt gaagagtggc tgaa 24
<210> 215
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 215
aagccctgtg tggtagtagt cagt 24
<210> 216
<211> 24
<212> DNA

<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 216
tgagtcatgt tggataacca gtct 24
<210> 217
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 217
gaaggtcagg aacatacaat tcaa 24
<210> 218
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 218
[0049]
gatatgggca tgtcagtttc atag 24
<210> 219
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 219
tgctgttgat taatggttga tagg 24
<210> 220
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 220
ttttaaattg ccatgtttgt gtc 23
<210> 221
<211> 24
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成
引物
<400> 221
atgaataacc taatgggcag aaaa 24
<210> 222
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 222
tcaagtcgct tcattttgat agac 24
<210> 223
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 223
[0050] cacaacaaaa catatagcca aagc 24
<210> 224
<211> 25
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 224
tgctgctaaa ataacacaaa tcagt 25
<210> 225
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 225
ctgtgcctat tgtggttatc ctg 23
<210> 226
<211> 24
<212> DNA
<213> 人工序列
<220>

<223> 人工序列的描述：合成
引物
<400> 226
attgatctgc aatacatgtg gagt 24
<210> 227
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 227
tttgcctctg ctattacagt atgg 24
<210> 228
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 228
tgtagggtgg ttggctaaaa taat 24
[0051]
<210> 229
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 229
tttttgcaca gtcaataaca caaa 24
<210> 230
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 230
ggctggtctc acaattgtac ttta 24
<210> 231
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成

引物

<400> 231

cattatggac tgaaaatctc agca 24

<210> 232

<211> 24

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成引物

<400> 232

atcatcctag ccataacaca atga 24

<210> 233

<211> 24

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成引物

<400> 233

ttcagcttta acgtgatttt ctgt 24

<210> 234

<211> 24

[0052]

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成引物

<400> 234

ggattcagaa gctgtttacg aagt 24

<210> 235

<211> 24

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成引物

<400> 235

tttagctgga ttggaaaaac aaat 24

<210> 236

<211> 24

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成引物

<400> 236
aactcacccc attgttggta tatt 24
<210> 237
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 237
ccttgtccaa ataccgaaat acat 24
<210> 238
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 238
cacataattc atgaacttgg cttc 24
<210> 239
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 239
tagtagctgg ggaggaagat acag 24
<210> 240
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 240
tttttgtttt aattgcgact gtgt 24
<210> 241
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 241

[0053]

agaaaagggg ttttcttttg actt 24

<210> 242
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 242
cattgtgact ggatgagaag aaac 24

<210> 243
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 243
aacggctgtt attaaagtcc tcag 24

<210> 244
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 244
caagtcagaa gtcacttgct ttgt 24

[0054]

<210> 245
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 245
ttttatgtgc aggaatcagt ctgt 24

<210> 246
<211> 18
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 246
tggcggcgtt ttcattat 18

<210> 247
<211> 26
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 247
ttcgatccgt aatgattgtt ctagcc 26
<210> 248
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 248
ggtcttccag agtgctgagg 20
<210> 249
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 249
tgtgtgcttc tgtacacatc atct 24
<210> 250
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 250
agatttcaac cctcaaaaac tgag 24
<210> 251
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 251
taaactcttt cttttccgca attc 24
<210> 252

[0055]

<210> 24
<211> DNA
<212> 人工序列
<213>
<220>
<223> 人工序列的描述：合成引物
<400> 252
caaggtgacc tgctacctaa aaat 24
<210> 253
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 253
tatgaccaag gctatgtgtt cact 24
<210> 254
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 254
acagcctctc tccagtaaca ttct 24
<210> 255
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 255
tattcttgca gtggtttcac attt 24
<210> 256
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 256
atattttaag ccaagaccca acaa 24
<210> 257
<211> 24

[0056]

[0057]

<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 257
ctttcaactg tctgtctgat tgct 24
<210> 258
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 258
aacagcctct cttcattgtt ctct 24
<210> 259
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 259
ctctggaact tgtctctgtc ttga 24
<210> 260
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 260
ctttcctgcg ttctcatgtt acta 24
<210> 261
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 261
ccttatatcc gtatcgctca ctct 24
<210> 262
<211> 24
<212> DNA

<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 262
catatctgtc taacttccgc acac 24
<210> 263
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 263
acaggtgtta tgttgtctgc atct 24
<210> 264
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 264
[0058]
actccattcc cagattagtt atgc 24
<210> 265
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 265
ctgttttctt tgtgagagtg gaga 24
<210> 266
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 266
tgtaaggtgg tcaaacttgc tcta 24
<210> 267
<211> 25
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物
<400> 267
tttttcctag tacccacaga ttttt 25
<210> 268
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 268
tccctgattc tctcatttgt gtta 24
<210> 269
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 269
[0059] ttgggaacat cagagaaagt atga 24
<210> 270
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 270
acaaattaca gtctcctggg aaag 24
<210> 271
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 271
agtagcttac cttggcagag aaaa 24
<210> 272
<211> 24
<212> DNA
<213> 人工序列
<220>

<223> 人工序列的描述：合成
引物
<400> 272
tgacatactg ttacccttg cagt 24
<210> 273
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 273
gaaaggctca gtgaatgttt gtt 23
<210> 274
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 274
cactgcatca tctcattaaa tcaa 24

[0060]

<210> 275
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 275
cccatatatt catgattacc caca 24
<210> 276
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 276
tatcagaacg agcactaaaa gcac 24
<210> 277
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成

引物

<400> 277

ttgggaggct gaggtacaag 20

<210> 278

<211> 24

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成引物

<400> 278

gaatgaaaaa caaacagaag gtga 24

<210> 279

<211> 24

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成引物

<400> 279

ctcctcatct gtacccttca atct 24

<210> 280

<211> 24

[0061]

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成引物

<400> 280

agagtggcat ctagtgtcag tgag 24

<210> 281

<211> 24

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成引物

<400> 281

taccaaaagc ttctcctgtt tacc 24

<210> 282

<211> 24

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成引物

<400> 282
gtaagttgga tggcctattc tttg 24
<210> 283
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 283
gaaggaaatg caaggataca agat 24
<210> 284
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 284
tgattgaaag aatcattcca gaaa 24
<210> 285
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 285
tcagaaggaa aattgaaatt ggtt 24
<210> 286
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 286
cagatgtgtt cttcatcatt cctc 24
<210> 287
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 287

[0062]

ttctctttag ggaaagctct caaa 24

<210> 288
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 288
gggtatagat catatggagg gaag 24

<210> 289
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 289
agatgatctg cccacctcag 20

<210> 290
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 290
ctttcttcct catttagtgg caat 24

[0063]

<210> 291
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 291
atgaattgca gattgatggt actg 24

<210> 292
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 292
tctcaccaag aaccaaattg tcta 24

<210> 293
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 293
gtaggatacc ttggcaacag tctt 24
<210> 294
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 294
ttaacgaatt gtgagatttg ctgt 24
<210> 295
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 295
tcagaaagtc aagtagcaca caca 24
<210> 296
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 296
agaagcacac actcaggtaa agc 23
<210> 297
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 297
tctttggggg aataatgact aaaa 24
<210> 298

[0064]

[0065]

<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 298
tttggcattt atgggaataa aact 24
<210> 299
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 299
actaattctg gtcaagccca tca 23
<210> 300
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 300
ttaagacatc ggatgaacag aaag 24
<210> 301
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 301
agaagctttc tgacatgatc tgc 23
<210> 302
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 302
tcaattgcat taggacttag acca 24
<210> 303
<211> 24

<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 303
gttaaattac ctgtgaagcc cttg 24
<210> 304
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 304
cggaaaacag atccacttta tgat 24
<210> 305
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物

[0066]

<400> 305
aaatccactg gaaacatctt gagt 24
<210> 306
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 306
agtctcttca gaatcatgcc ctat 24
<210> 307
<211> 22
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 307
gcttggtggc acatacctgt ag 22
<210> 308
<211> 24
<212> DNA

[0067]

<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 308
ggtaggtaga tttgcttgct tgtt 24
<210> 309
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 309
agctctcagc agagtaggga ttta 24
<210> 310
<211> 21
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 310
gtgagtctac tgcaccccat c 21
<210> 311
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 311
tgacactgtg aagtcaattc tgtc 24
<210> 312
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 312
tcaagaactt gacaatgagc aaat 24
<210> 313
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物
<400> 313
tatccgatcc actgttgtgt gt 22
<210> 314
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 314
caggagaccc aaaaccactc tac 23
<210> 315
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 315
[0068] ttgttctaca aatagggctt cctt 24
<210> 316
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 316
tgttaagttt gggcttatgt tcct 24
<210> 317
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 317
cacaagtctc actgcacaaa cat 23
<210> 318
<211> 24
<212> DNA
<213> 人工序列
<220>

<223> 人工序列的描述：合成
引物
<400> 318
tgacccatga ttatctctct ttga 24
<210> 319
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 319
ttcagcttct gattggtttt aatg 24
<210> 320
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 320
ccaattcctt aattttccct acag 24

[0069]

<210> 321
<211> 21
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 321
atctcagacc aggagggaga c 21
<210> 322
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 322
cctcagggtc agtacatttt tcag 24
<210> 323
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成

引物

<400> 323

ttcttaggac attgctccac atac 24

<210> 324

<211> 24

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成引物

<400> 324

gcaaacataa tgcaactcgt aatc 24

<210> 325

<211> 21

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成引物

<400> 325

gcaagggagt ctgtgtcttt g 21

<210> 326

<211> 24

[0070]

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成引物

<400> 326

tcatttaagt ggctgttctg tgtt 24

<210> 327

<211> 24

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成引物

<400> 327

acaaaacaga gagaaaaggc agag 24

<210> 328

<211> 23

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成引物

[0071]

<400> 328
gttttgattt ctggtgccta cag 23
<210> 329
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 329
actgaagctg aagcccagtc 20
<210> 330
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 330
acatgagctc tcaggtttct gac 23
<210> 331
<211> 25
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 331
tcaaacttag atggttccct atgtt 25
<210> 332
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 332
gtaccctgaa aatgtagggt gact 24
<210> 333
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 333

cacttcccaa gtgaggcaat 20

<210> 334
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 334
ctatacttgg ggctgacttg ctac 24

<210> 335
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 335
tcgtataggt tactttggct caca 24

[0072]

<210> 336
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 336
agggatcttt actcctcagt gtgt 24

<210> 337
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 337
tgtagaagtt ggaatatcct gctg 24

<210> 338
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 338
gtcaacaatt tgatctcagg cttc 24

<210> 339
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 339
ctcagtacta aagatggacg cttg 24
<210> 340
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 340
aatcatttca gtcttcccaa caat 24
<210> 341
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 341
gggaatcaca gtagatgttt gtca 24
<210> 342
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 342
agaccaggag gtaagaacat tttg 24
<210> 343
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 343
ccacatagaa agagacttgc agaa 24
<210> 344

[0073]

[0074]

<210> 24
<211> DNA
<212> 人工序列
<213>
<220>
<223> 人工序列的描述：合成引物
<400> 344
agagatgcca aaagaacagt caat 24
<210> 345
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 345
tgtgccttag gctatgtaaa ctgt 24
<210> 346
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 346
aaaccttgt aaccaaaatt acca 24
<210> 347
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 347
taactgcatc agaagtcctt gcta 24
<210> 348
<211> 22
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 348
ggagaccaag ctgctaaagt ca 22
<210> 349
<211> 37

[0075]

<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 349
gtggtgccgc gggagtttgg ctcaaattgt tactctt 37
<210> 350
<211> 37
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 350
gtggtgccgc gggggaaatg gtctaggaga gtaaagt 37
<210> 351
<211> 34
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 351
gtggtgccgc gggagaggtt atgtggcttt acca 34
<210> 352
<211> 32
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 352
gtggtgccgc ggctcattct catgcctgga ca 32
<210> 353
<211> 32
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 353
gtggtgccgc ggcgggcttg gacagaactt ac 32
<210> 354
<211> 32
<212> DNA

<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 354
gtggtgccgc ggctgcgtag tgccaaaaca aa 32
<210> 355
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 355
gtggtgccgc ggtaatttca ttgaagagtg gctgaa 36
<210> 356
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 356
[0076]
gtggtgccgc ggaagccctg tgtggtagta gtcagt 36
<210> 357
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 357
gtggtgccgc ggcaagtcag aagtcacttg ctttgt 36
<210> 358
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 358
gtggtgccgc ggttttatgt gcaggaatca gtctgt 36
<210> 359
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成
引物
<400> 359
gtggtgccgc ggtgtgtgct tctgtacaca tcatct 36
<210> 360
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 360
gtggtgccgc ggagatttca accctcaaaa actgag 36
<210> 361
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 361
[0077] gtggtgccgc ggttgggaac atcagagaaa gtatga 36
<210> 362
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 362
gtggtgccgc ggacaaatta cagtctcctg ggaaag 36
<210> 363
<211> 32
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 363
gtggtgccgc ggcacttccc aagtgaggca at 32
<210> 364
<211> 36
<212> DNA
<213> 人工序列
<220>

<223> 人工序列的描述：合成
引物
<400> 364
gtggtgccgc ggctatactt ggggctgact tgctac 36
<210> 365
<211> 34
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 365
gtggtgccgc ggttggctct ttagcttgtg tttc 34
<210> 366
<211> 32
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 366
gtggtgccgc ggtgagactc ccaaaggcaa tc 32

[0078]

<210> 367
<211> 34
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 367
gtggtgccgc ggttggctct ttagcttgtg tttc 34
<210> 368
<211> 32
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 368
gtggtgccgc ggactgaggg gtgatcttgg tg 32
<210> 369
<211> 32
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成

引物
<400> 369
gtggtgccgc gggcagagaa agccagtcgg ta 32
<210> 370
<211> 32
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 370
gtggtgccgc ggtgagactc ccaaaggcaa tc 32
<210> 371
<211> 32
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 371
gtggtgccgc gggcagagaa agccagtcgg ta 32
<210> 372
<211> 32
[0079]
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 372
gtggtgccgc ggactgaggg gtgatcttgg tg 32
<210> 373
<211> 32
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 373
gtggtgccgc ggccagagtt cctagggcag ag 32
<210> 374
<211> 32
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物

<400> 374
gtggtgccgc ggagctagtc cccacattcc ac 32
<210> 375
<211> 32
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 375
gtggtgccgc ggccagagtt cctagggcag ag 32
<210> 376
<211> 32
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 376
gtggtgccgc ggggtggagg gaaactttag gc 32
<210> 377
<211> 32
<212> DNA
[0080] <213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 377
gtggtgccgc ggctcattct catgcctgga ca 32
<210> 378
<211> 32
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 378
gtggtgccgc ggagctagtc cccacattcc ac 32
<210> 379
<211> 34
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 379

gtggtgccgc ggtctcatgc ctggacaagt aact 34

<210> 380
<211> 32
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 380
gtggtgccgc ggggtggagg gaaactttag gc 32

<210> 381
<211> 32
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 381
gtggtgccgc ggggcttgga cagaacttac cg 32

<210> 382
<211> 32
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 382
gtggtgccgc ggcaccactg tctgcctaag ga 32

<210> 383
<211> 32
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 383
gtggtgccgc ggggcttgga cagaacttac cg 32

<210> 384
<211> 32
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 384
gtggtgccgc ggggtggagg gaaactttag gc 32

[0081]

[0082]

<210> 385
<211> 33
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 385
gtggtgccgc ggcgtagtgc caaaacaaac agt 33
<210> 386
<211> 32
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 386
gtggtgccgc ggcaccactg tctgcctaag ga 32
<210> 387
<211> 33
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 387
gtggtgccgc ggcgtagtgc caaaacaaac agt 33
<210> 388
<211> 32
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 388
gtggtgccgc ggggtggagg gaaactttag gc 32
<210> 389
<211> 32
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 389
gtggtgccgc gggcgagggc ctacttgata tg 32
<210> 390

[0083]

<211> 32
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 390
gtggtgccgc ggcttcccaa gtgaggcaat gc 32
<210> 391
<211> 33
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 391
gtggtgccgc ggacgttttg tgctgctgta aca 33
<210> 392
<211> 32
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 392
gtggtgccgc ggctgcaggc acattctctt cc 32
<210> 393
<211> 32
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 393
gtggtgccgc gggccctgtg tggtagtagt ca 32
<210> 394
<211> 32
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 394
gtggtgccgc ggcttcccaa gtgaggcaat gc 32
<210> 395
<211> 33

[0084]

<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 395
gtggtgccgc ggcagtatta aggggtggga gct 33
<210> 396
<211> 32
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 396
gtggtgccgc ggtctcttcc tcacacagct ga 32
<210> 397
<211> 32
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 397
gtggtgccgc ggggagcttg gagggaagag aa 32
<210> 398
<211> 32
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 398
gtggtgccgc ggcttcccaa gtgaggcaat gc 32
<210> 399
<211> 32
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 399
gtggtgccgc ggatggatgg ggaagacact gg 32
<210> 400
<211> 32
<212> DNA

<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 400
gtggtgccgc ggctgcaggc acattctctt cc 32
<210> 401
<211> 32
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 401
gtggtgccgc ggggatgaaa cagggcagga ac 32
<210> 402
<211> 31
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 402
[0085]
gtggtgccgc ggttcccaag tgaggcaatg c 31
<210> 403
<211> 32
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 403
gtggtgccgc ggtttgcaga gccatgatga gg 32
<210> 404
<211> 30
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 404
gtggtgccgc ggcgacagcc aaaacagccg 30
<210> 405
<211> 44
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 405
aaaatatttt agctcctact cagactgtta ctctggtgac acaa 44
<210> 406
<211> 44
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 406
ttgtgtcacc agagtaacag tctgagtagg agctaaaata tttt 44
<210> 407
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 407
[0086] tagctcctac tcagactgtt actctggtga cacaac 36
<210> 408
<211> 28
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 408
tagctcctac tcagactggt gacccaac 28
<210> 409
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 409
tagctcctac tctggtgaca caac 24
<210> 410
<211> 28
<212> DNA
<213> 人工序列
<220>

<223> 人工序列的描述：合成
寡核苷酸
<400> 410
tagctcctac tcagactggt gacacaac 28
<210> 411
<211> 16
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 411
tagctcctac tcagac 16
<210> 412
<211> 27
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 412
tagctcctac tcagactgtt acacaac 27

[0087]

<210> 413
<211> 30
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 413
tagctcctac tcagactgtg gtgaggtgac 30
<210> 414
<211> 32
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 414
tagctcctac tcagactctc tggtgacaca ac 32
<210> 415
<211> 31
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成

寡核苷酸

<400> 415

tagctcctac tcagacctct ggtgacacaa c 31

<210> 416

<211> 32

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成

寡核苷酸

<400> 416

tagctcctac tcaggctgtc tggtgacaca ac 32

<210> 417

<211> 33

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成

寡核苷酸

<400> 417

tagctcctac tcagactact ctggtgacac aac 33

<210> 418

[0088]

<211> 28

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成

寡核苷酸

<400> 418

tagctcctac tcagactgtt gacacaac 28

<210> 419

<211> 13

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成

寡核苷酸

<400> 419

ctggtgacac aac 13

<210> 420

<211> 29

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成

寡核苷酸

<400> 420
tagctcctac tcagactgtt agacacaac 29
<210> 421
<211> 33
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 421
tagctcctac tcagactgct ctggtgacac aac 33
<210> 422
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 422
cagactgtta ctctggtgac 20
<210> 423
<211> 29
<212> DNA
[0089]
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 423
cagaccacct gtggtctcct actggtgac 29
<210> 424
<211> 18
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 424
gctttgattt ccctaggg 18
<210> 425
<211> 25
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 425

gctttgattt ccagttctta ggcaa 25

<210> 426

<211> 18

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成
寡核苷酸

<400> 426

gctttgattc ttaggcaa 18

<210> 427

<211> 62

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成
寡核苷酸

<220>

<221> 修饰的碱基

<222> (13)..(53)

<223> a, c, t, g, 未知或其它

<400> 427

[0090]

gctttgattt ccnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnncttaggc 60
aa 62

<210> 428

<211> 18

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成
寡核苷酸

<400> 428

tcttaaccat taccatag 18

<210> 429

<211> 32

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成
寡核苷酸

<400> 429

tcttaaccat taccatagag ttcttaggca ac 32

<210> 430

<211> 30

<212> DNA

<213> 人工序列

<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 430
tcttaaccat taccaaagtt cttaggcaac 30
<210> 431
<211> 31
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 431
tcttaaccat taccataggt tcttaggcaa c 31
<210> 432
<211> 14
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 432
gaaccaaacc cact 14

[0091]

<210> 433
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 433
gaaccaaacc cacttagggg ataa 24
<210> 434
<211> 2328
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 434
atgaggtctg actacaaaga ccatgacggt gattataaag atcatgacat cgattacaag 60
gatgacgatg acaagatggc ccccaagaag aagaggaagg tgggcctcga gcccggagaa 120
aaaccgtaca agtgccctga gtgcgggaaa tcattctccg accctggggc gctcgtccgg 180
caccaaagga cgcatacagg ggaaaagccg tataagtgcc ccgagtgtgg aaagagcttc 240
tcgcagagag cccaccttga acgacaccaa agaacacaca ctggtgagaa accctataag 300
tgtccagagt gcggcaaatc gtttagcaga tccgatgact tggtgcgcca ccagcggaca 360

[0092]

```
cacacgggtg aaaagcccta caaatgcccg gagtgtggga agtcgttttc aaggtcggat      420
catctgacta cccatcagcg cacccatacg ggagcggccg cccgcgccct ggtgaagagc      480
gagctggagg agaagaagtc cgagctgcgg cacaagctga agtacgtgcc ccacgagtac      540
atcgagctga tcgagatcgc caggaacccc acccaggacc gcatcctgga gatgaaggtg      600
atggagttct tcatgaaggt gtacggctac aggggagagc acctgggcgg aagcagaaag      660
cctgacggcg ccatctatac agtgggcagc cccatcgatt acggcgtgat cgtggacaca      720
aaggcctaca gcggcggcta caatctgcct atcggccagg ccgacgagat gcagagatac      780
gtgaaggaga accagacccg gaataagcac atcaacccca acgagtggtg gaaggtgtac      840
cctagcagcg tgaccgagtt caagttcctg ttcgtgagcg gccacttcaa gggcaactac      900
aaggcccagc tgaccaggct gaaccgcaaa accaactgca atggcgccgt gctgagcgtg      960
gaggagctgc tgatcggcgg cgagatgatc aaagccggca ccctgacact ggaggaggtg     1020
cggcgcaagt tcaacaacgg cgagatcaac ttcgagggca gaggaagtct tctaacatgc     1080
ggtgacgtgg aggagaatcc cggccctaga tctgactaca aagaccatga cggtgattat     1140
aaagatcatg acatcgatta caaggatgac gatgacaaga tggcccccaa gaagaagagg     1200
aaggtgggcc tcgagccggg agagaagccg tacaagtgtc ccgaatgtgg aaagagcttc     1260
tcacagtcgg gggaccttcg gcgccaccag cgcacacata ctggtgaaaa gccgtataag     1320
tgtccagaat gtggcaaatc attctccaca tcagggagcc tggtcaggca ccagcgaacc     1380
cacacgggtg agaagcccta taagtgcccc gaatgcggga agtcctttтc gcagagagcc     1440
cacttggaga ggcaccagag gacccatacg ggggagaaac cttacaagtg ccctgaatgc     1500
ggaaagtcgt tctcgaccca tctggatctc atcagacatc agagaacgca cactggagag     1560
aaaccctaca aatgtcccga gtgtgggaag tcgtttagcc gaaaggacaa tctcaaaaac     1620
catcaacgga cacacacggg tgaaaaacca tacaaatgcc cggagtgcgg caaatcgttt     1680
tcccaacttg cgcacttgcg ggcacaccaa cgcacgcata ctggagcggc cgcccgcgcc     1740
ctggtgaaga gcgagctgga ggagaagaag tccgagctgc ggcacaagct gaagtacgtg     1800
ccccacgagt acatcgagct gatcgagatc gccaggaacc ccacccagga ccgcatcctg     1860
gagatgaagg tgatggagtt cttcatgaag gtgtacggct acaggggaga gcacctgggc     1920
ggaagcagaa agcctgacgg cgccatctat acagtgggca gccccatcga ttacggcgtg     1980
atcgtggaca caaaggccta cagcggcggc tacaatctgc ctatcggcca ggccgacgag     2040
atggagagat acgtggagga gaaccagaca cgggataagc acctcaaccc caacgagtgg     2100
tggaaggtgt accctagcag cgtgaccgag ttcaagttcc tgttcgtgag cggccacttc     2160
aagggcaact acaaggccca gctgaccagg ctgaaccaca tcaccaactg caatggcgcc     2220
gtgctgagcg tggaggagct gctgatcggc ggcgagatga tcaaagccgg caccctgaca     2280
ctggaggagg tgcggcgcaa gttcaacaac ggcgagatca acttctga                  2328
<210> 435
<211> 775
<212> PRT
<213> 人工序列
<220>
<223> 人工序列的描述：合成
      多肽
<400> 435
Met Arg Ser Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp
1               5                  10                  15
Ile Asp Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys Arg
            20                  25                  30
Lys Val Gly Leu Glu Pro Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys
```

[0093]

```
          35                  40                  45
Gly Lys Ser Phe Ser Asp Pro Gly Ala Leu Val Arg His Gln Arg Thr
      50                  55                  60
His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe
65                  70                  75                  80
Ser Gln Arg Ala His Leu Glu Arg His Gln Arg Thr His Thr Gly Glu
                  85                  90                  95
Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Arg Ser Asp
             100                 105                 110
Asp Leu Val Arg His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys
         115                 120                 125
Cys Pro Glu Cys Gly Lys Ser Phe Ser Arg Ser Asp His Leu Thr Thr
     130                 135                 140
His Gln Arg Thr His Thr Gly Ala Ala Ala Arg Ala Leu Val Lys Ser
145                 150                 155                 160
Glu Leu Glu Glu Lys Lys Ser Glu Leu Arg His Lys Leu Lys Tyr Val
                 165                 170                 175
Pro His Glu Tyr Ile Glu Leu Ile Glu Ile Ala Arg Asn Pro Thr Gln
             180                 185                 190
Asp Arg Ile Leu Glu Met Lys Val Met Glu Phe Phe Met Lys Val Tyr
         195                 200                 205
Gly Tyr Arg Gly Glu His Leu Gly Gly Ser Arg Lys Pro Asp Gly Ala
     210                 215                 220
Ile Tyr Thr Val Gly Ser Pro Ile Asp Tyr Gly Val Ile Val Asp Thr
225                 230                 235                 240
Lys Ala Tyr Ser Gly Gly Tyr Asn Leu Pro Ile Gly Gln Ala Asp Glu
                 245                 250                 255
Met Gln Arg Tyr Val Lys Glu Asn Gln Thr Arg Asn Lys His Ile Asn
             260                 265                 270
Pro Asn Glu Trp Trp Lys Val Tyr Pro Ser Ser Val Thr Glu Phe Lys
         275                 280                 285
Phe Leu Phe Val Ser Gly His Phe Lys Gly Asn Tyr Lys Ala Gln Leu
     290                 295                 300
Thr Arg Leu Asn Arg Lys Thr Asn Cys Asn Gly Ala Val Leu Ser Val
305                 310                 315                 320
Glu Glu Leu Leu Ile Gly Gly Glu Met Ile Lys Ala Gly Thr Leu Thr
                 325                 330                 335
Leu Glu Glu Val Arg Arg Lys Phe Asn Asn Gly Glu Ile Asn Phe Glu
             340                 345                 350
Gly Arg Gly Ser Leu Leu Thr Cys Gly Asp Val Glu Glu Asn Pro Gly
         355                 360                 365
Pro Arg Ser Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp
     370                 375                 380
Ile Asp Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys Arg
385                 390                 395                 400
Lys Val Gly Leu Glu Pro Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys
```

405 410 415
Gly Lys Ser Phe Ser Gln Ser Gly Asp Leu Arg Arg His Gln Arg Thr
420 425 430
His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe
435 440 445
Ser Thr Ser Gly Ser Leu Val Arg His Gln Arg Thr His Thr Gly Glu
450 455 460
Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Gln Arg Ala
465 470 475 480
His Leu Glu Arg His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys
485 490 495
Cys Pro Glu Cys Gly Lys Ser Phe Ser Thr His Leu Asp Leu Ile Arg
500 505 510
His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys
515 520 525
Gly Lys Ser Phe Ser Arg Lys Asp Asn Leu Lys Asn His Gln Arg Thr
530 535 540
His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe
545 550 555 560
Ser Gln Leu Ala His Leu Arg Ala His Gln Arg Thr His Thr Gly Ala
565 570 575
Ala Ala Arg Ala Leu Val Lys Ser Glu Leu Glu Glu Lys Lys Ser Glu
580 585 590
[0094] Leu Arg His Lys Leu Lys Tyr Val Pro His Glu Tyr Ile Glu Leu Ile
595 600 605
Glu Ile Ala Arg Asn Pro Thr Gln Asp Arg Ile Leu Glu Met Lys Val
610 615 620
Met Glu Phe Phe Met Lys Val Tyr Gly Tyr Arg Gly Glu His Leu Gly
625 630 635 640
Gly Ser Arg Lys Pro Asp Gly Ala Ile Tyr Thr Val Gly Ser Pro Ile
645 650 655
Asp Tyr Gly Val Ile Val Asp Thr Lys Ala Tyr Ser Gly Gly Tyr Asn
660 665 670
Leu Pro Ile Gly Gln Ala Asp Glu Met Glu Arg Tyr Val Glu Glu Asn
675 680 685
Gln Thr Arg Asp Lys His Leu Asn Pro Asn Glu Trp Trp Lys Val Tyr
690 695 700
Pro Ser Ser Val Thr Glu Phe Lys Phe Leu Phe Val Ser Gly His Phe
705 710 715 720
Lys Gly Asn Tyr Lys Ala Gln Leu Thr Arg Leu Asn His Ile Thr Asn
725 730 735
Cys Asn Gly Ala Val Leu Ser Val Glu Glu Leu Leu Ile Gly Gly Glu
740 745 750
Met Ile Lys Ala Gly Thr Leu Thr Leu Glu Glu Val Arg Arg Lys Phe
755 760 765
Asn Asn Gly Glu Ile Asn Phe

770 775

<210> 436

<211> 4725

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成
多核苷酸

<400> 436

[0095]

tggccactcc ctctatgcgc actcgctcgc tcggtggggc ctggcgacca aaggtcgcca 60
gacggacgtg ctttgcacgt ccggccccac cgagcgagcg agtgcgcata gagggagtgg 120
ccaactccat cactagaggt atggcagtga cgtaacgcga agcgcgcgaa gcgagaccac 180
gcctaccagc tgcgtcagca gtcaggtgac ccttttgcga cagtttgcga caccacgtgg 240
ccgctgaggg tatatattct cgagtgagcg aaccaggagc tccattttga ccgcgaaatt 300
tgaacgagca gcagccatgc cggggttcta cgagattgtc ctgaaggtcc cgagtgacct 360
ggacgagcac ctgccgggca tttctaactc gtttgttaac tgggtggccg agaaggaatg 420
ggagctgccg ccggattctg acatggatcc gaatctgatt gagcaggcac ccctgaccgt 480
ggccgaaaag cttcagcgcg agttcctggt ggagtggcgc cgcgtgagta aggccccgga 540
ggccctcttt tttgtccagt tcgaaaaggg ggagacctac ttccacctgc acgtgctgat 600
tgagaccatc ggggtcaaat ccatggtggt cggccgctac gtgagccaga ttaaagagaa 660
gctggtgacc cgcatctacc gcggggtcga gccgcagctt ccgaactggt tcgcggtgac 720
caaaacgcga aatggcgccg ggggcgggaa caaggtggtg gacgactgct acatccccaa 780
ctacctgctc cccaagaccc agcccgagct ccagtgggcg tggactaaca tggaccagta 840
tttaagcgcc tgtttgaatc tcgcggagcg taaacggctg gtggcgcagc atctgacgca 900
cgtgtcgcag acgcaggagc agaacaaaga gaatcagaac cccaattctg acgcgccggt 960
catcaggtca aaaacctcag ccaggtacat ggagctggtc gggtggctgg tggaccgcgg 1020
gatcacgtca gaaaagcaat ggattcagga ggaccaggcc tcgtacatct ccttcaacgc 1080
cgcctccaac tcgcggtccc agatcaaggc cgcgctggac aatgcctcca agatcatgag 1140
cctgacaaag acggctccgg actacctggt gggcagcaac ccgccggagg acattaccaa 1200
aaatcggatc taccaaatcc tggagctgaa cgggtacgat ccgcagtacg cggcctccgt 1260
cttcctgggc tgggcgcaaa agaagttcgg gaagaggaac accatctggc tctttgggcc 1320
ggccacgacg ggtaaaacca acatcgcgga agccatcgcc cacgccgtgc ccttctacgg 1380
ctgcgtaaac tggaccaatg agaactttcc cttcaacgat tgcgtcgaca agatggtgat 1440
ctggtgggag gagggcaaga tgacggccaa ggtcgtggag agcgccaagg ccattctggg 1500
cggaagcaag gtgcgcgtgg accaaaagtg caagtcatcg gcccagatcg aacccactcc 1560
cgtgatcgtc acctccaaca ccaacatgtg cgccgtgatt gacgggaaca gcaccacctt 1620
cgagcatcag cagccgctgc aggaccggat gtttaaattt gaacttaccc gccgtttgga 1680
ccatgacttt gggaaggtca ccaaacagga agtaaaggac tttttccggt gggcttccga 1740
tcacgtgact gacgtggctc atgagttcta cgtcagaaag ggtggagcta agaaacgccc 1800
cgcctccaat gacgcggatg taagcgagcc aaaacggcag tgcacgtcac ttgcgcagcc 1860
gacaacgtca gacgcggaag caccggcgga ctacgcggac aggtaccaaa acaaatgttc 1920
tcgtcacgtg ggcatgaatc tgatgctttt tccctgtaaa acatgcgaga gaatgaatca 1980
aatttccaat gtctgtttta cgcatggtca aagagactgt ggggaatgct tccctggaat 2040
gtcagaatct caacccgttt ctgtcgtcaa aaagaagact tatcagaaac tgtgtccaat 2100
tcatcatatc ctgggaaggg cacccgagat tgcctgttcg gcctgcgatt tggccaatgt 2160
ggacttggat gactgtgttt ctgagcaata aatgacttaa accaggtatg gctgctgacg 2220

[0096]

gttatcttcc agattggctc gaggacaacc tttctgaagg cattcgtgag tggtgggctc 2280
tgaaacctgg agtccctcaa cccaaagcga accaacaaca ccaggacaac cgtcggggtc 2340
ttgtgcttcc gggttacaaa tacctcggac ccggtaacgg actcgacaaa ggagagccgg 2400
tcaacgaggc ggacgcggca gccctcgaac acgacaaagc ttacgaccag cagctcaagg 2460
ccggtgacaa cccgtacctc aagtacaacc acgccgacgc cgagtttcag gagcgtcttc 2520
aagaagatac gtcttttggg ggcaaccttg gcagagcagt cttccaggcc aaaaagagga 2580
tccttgagcc tcttggtctg gttgaggaag cagctaaaac ggctcctgga aagaagaggc 2640
ctgtagatca gtctcctcag gaaccggact catcatctgg tgttggcaaa tcgggcaaac 2700
agcctgccag aaaaagacta aatttcggtc agactggcga ctcagagtca gtcccagacc 2760
ctcaacctct cggagaacca ccagcagccc ccacaagttt gggatctaat acaatggctt 2820
caggcggtgg cgcaccaatg gcagacaata acgagggtgc cgatggagtg ggtaattcct 2880
caggaaattg gcattgcgat tcccaatggc tgggcgacag agtcatcacc accagcacca 2940
gaacctgggc cctgcccact tacaacaacc atctctacaa gcaaatctcc agcgcttcaa 3000
cgggagcttc aaacgacaac cactactttg gctacagcac cccttggggg tattttgact 3060
ttaacagatt ccactgccac ttctcaccac gtgactggca gcgactcatt aacaacaact 3120
ggggattccg gcccaagaaa ctcagcttca agctcttcaa catccaagtt aaagaggtca 3180
cgcagaacga tggcacgacg actattgcca ataaccttac cagcacggtt caagtgttta 3240
cggactcgga gtatcagctc ccgtacgtgc tcgggtcggc gcaccaaggc tgtctcccgc 3300
cgtttccagc ggacgtcttc atggtccctc agtatggata cctcaccctg aacaacggaa 3360
gtcaagcggt gggacgctca tccttttact gcctggagta cttcccttcg cagatgctaa 3420
ggactggaaa taacttccaa ttcagctata ccttcgagga tgtacctttt cacagcagct 3480
acgctcacag ccagagtttg gatcgcttga tgaatcctct tattgatcag tatctgtact 3540
acctgaacag aacgcaagga acaacctctg gaacaaccaa ccaatcacgg ctgcttttta 3600
gccaggctgg gcctcagtct atgtctttgc aggccagaaa ttggctacct gggccctgct 3660
accggcaaca gagactttca aagactgcta acgacaacaa caacagtaac tttccttgga 3720
cagcggccag caaatatcat ctcaatggcc gcgactcgct ggtgaatcca ggaccagcta 3780
tggccagtca caaggacgat gaagaaaaat ttttccctat gcacggcaat ctaatatttg 3840
gcaaagaagg gacaacggca agtaacgcag aattagataa tgtaatgatt acggatgaag 3900
aagagattcg taccaccaat cctgtggcaa cagagcagta tggaactgtg gcaaataact 3960
tgcagagctc aaatacagct cccacgacta gaactgtcaa tgatcagggg gccttacctg 4020
gcatggtgtg gcaagatcgt gacgtgtacc ttcaaggacc tatctgggca aagattcctc 4080
acacggatgg acactttcat ccttctcctc tgatgggagg ctttggactg aaacatccgc 4140
ctcctcaaat catgatcaaa aatactccgg taccggcaaa tcctccgacg actttcagcc 4200
cggccaagtt tgcttcattt atcactcagt actccactgg acaggtcagc gtggaaattg 4260
agtgggagct acagaaagaa aacagcaaac gttggaatcc agagattcag tacacttcca 4320
actacaacaa gtctgttaat gtggacttta ctgtagacac taatggtgtt tatagtgaac 4380
ctcgccctat tggaacccgg tatctcacac gaaacttgta atcctggtta atcaataaac 4440
cgtttaattc gtttcagttg aactttggct cttgtgcact tcttatctta tcttgtttcc 4500
atggctactg cgtagataag cagcggcctg cggcgcttgc gcttcgcggt ttacaactgc 4560
tggttaatat ttaactctcg ccatacctct agtgatggag ttggccactc cctctatgcg 4620
cactcgctcg ctcggtgggg ccggacgtgc aaagcacgtc cgtctggcga cctttggtcg 4680
ccaggcccca ccgagcgagc gagtgcgcat agagggagtg gccaa 4725

<210> 437
<211> 737
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成
多肽
<400> 437

[0097]

```
Met Ala Ala Asp Gly Tyr Leu Pro Asp Trp Leu Glu Asp Asn Leu Ser
1               5                  10                  15
Glu Gly Ile Arg Glu Trp Trp Ala Leu Lys Pro Gly Val Pro Gln Pro
            20                  25                  30
Lys Ala Asn Gln Gln His Gln Asp Asn Arg Arg Gly Leu Val Leu Pro
        35                  40                  45
Gly Tyr Lys Tyr Leu Gly Pro Gly Asn Gly Leu Asp Lys Gly Glu Pro
    50                  55                  60
Val Asn Glu Ala Asp Ala Ala Ala Leu Glu His Asp Lys Ala Tyr Asp
65                  70                  75                  80
Gln Gln Leu Lys Ala Gly Asp Asn Pro Tyr Leu Lys Tyr Asn His Ala
                85                  90                  95
Asp Ala Glu Phe Gln Glu Arg Leu Gln Glu Asp Thr Ser Phe Gly Gly
            100                 105                 110
Asn Leu Gly Arg Ala Val Phe Gln Ala Lys Lys Arg Ile Leu Glu Pro
        115                 120                 125
Leu Gly Leu Val Glu Glu Ala Ala Lys Thr Ala Pro Gly Lys Lys Arg
    130                 135                 140
Pro Val Asp Gln Ser Pro Gln Glu Pro Asp Ser Ser Ser Gly Val Gly
145                 150                 155                 160
Lys Ser Gly Lys Gln Pro Ala Arg Lys Arg Leu Asn Phe Gly Gln Thr
                165                 170                 175
Gly Asp Ser Glu Ser Val Pro Asp Pro Gln Pro Leu Gly Glu Pro Pro
            180                 185                 190
Ala Ala Pro Thr Ser Leu Gly Ser Asn Thr Met Ala Ser Gly Gly Gly
        195                 200                 205
Ala Pro Met Ala Asp Asn Asn Glu Gly Ala Asp Gly Val Gly Asn Ser
    210                 215                 220
Ser Gly Asn Trp His Cys Asp Ser Gln Trp Leu Gly Asp Arg Val Ile
225                 230                 235                 240
Thr Thr Ser Thr Arg Thr Trp Ala Leu Pro Thr Tyr Asn Asn His Leu
                245                 250                 255
Tyr Lys Gln Ile Ser Ser Ala Ser Thr Gly Ala Ser Asn Asp Asn His
            260                 265                 270
Tyr Phe Gly Tyr Ser Thr Pro Trp Gly Tyr Phe Asp Phe Asn Arg Phe
        275                 280                 285
His Cys His Phe Ser Pro Arg Asp Trp Gln Arg Leu Ile Asn Asn Asn
    290                 295                 300
Trp Gly Phe Arg Pro Lys Lys Leu Ser Phe Lys Leu Phe Asn Ile Gln
305                 310                 315                 320
Val Lys Glu Val Thr Gln Asn Asp Gly Thr Thr Thr Ile Ala Asn Asn
                325                 330                 335
```

[0098]

```
Leu Thr Ser Thr Val Gln Val Phe Thr Asp Ser Glu Tyr Gln Leu Pro
            340                 345                 350
Tyr Val Leu Gly Ser Ala His Gln Gly Cys Leu Pro Pro Phe Pro Ala
        355                 360                 365
Asp Val Phe Met Val Pro Gln Tyr Gly Tyr Leu Thr Leu Asn Asn Gly
    370                 375                 380
Ser Gln Ala Val Gly Arg Ser Ser Phe Tyr Cys Leu Glu Tyr Phe Pro
385                 390                 395                 400
Ser Gln Met Leu Arg Thr Gly Asn Asn Phe Gln Phe Ser Tyr Thr Phe
                405                 410                 415
Glu Asp Val Pro Phe His Ser Ser Tyr Ala His Ser Gln Ser Leu Asp
            420                 425                 430
Arg Leu Met Asn Pro Leu Ile Asp Gln Tyr Leu Tyr Tyr Leu Asn Arg
        435                 440                 445
Thr Gln Gly Thr Thr Ser Gly Thr Thr Asn Gln Ser Arg Leu Leu Phe
    450                 455                 460
Ser Gln Ala Gly Pro Gln Ser Met Ser Leu Gln Ala Arg Asn Trp Leu
465                 470                 475                 480
Pro Gly Pro Cys Tyr Arg Gln Gln Arg Leu Ser Lys Thr Ala Asn Asp
                485                 490                 495
Asn Asn Asn Ser Asn Phe Pro Trp Thr Ala Ala Ser Lys Tyr His Leu
            500                 505                 510
Asn Gly Arg Asp Ser Leu Val Asn Pro Gly Pro Ala Met Ala Ser His
        515                 520                 525
Lys Asp Asp Glu Glu Lys Phe Phe Pro Met His Gly Asn Leu Ile Phe
    530                 535                 540
Gly Lys Glu Gly Thr Thr Ala Ser Asn Ala Glu Leu Asp Asn Val Met
545                 550                 555                 560
Ile Thr Asp Glu Glu Glu Ile Arg Thr Thr Asn Pro Val Ala Thr Glu
                565                 570                 575
Gln Tyr Gly Thr Val Ala Asn Asn Leu Gln Ser Ser Asn Thr Ala Pro
            580                 585                 590
Thr Thr Arg Thr Val Asn Asp Gln Gly Ala Leu Pro Gly Met Val Trp
        595                 600                 605
Gln Asp Arg Asp Val Tyr Leu Gln Gly Pro Ile Trp Ala Lys Ile Pro
    610                 615                 620
His Thr Asp Gly His Phe His Pro Ser Pro Leu Met Gly Gly Phe Gly
625                 630                 635                 640
Leu Lys His Pro Pro Pro Gln Ile Met Ile Lys Asn Thr Pro Val Pro
                645                 650                 655
Ala Asn Pro Pro Thr Thr Phe Ser Pro Ala Lys Phe Ala Ser Phe Ile
            660                 665                 670
Thr Gln Tyr Ser Thr Gly Gln Val Ser Val Glu Ile Glu Trp Glu Leu
        675                 680                 685
Gln Lys Glu Asn Ser Lys Arg Trp Asn Pro Glu Ile Gln Tyr Thr Ser
    690                 695                 700
```

Asn Tyr Asn Lys Ser Val Asn Val Asp Phe Thr Val Asp Thr Asn Gly
705 710 715 720
Val Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg Asn
725 730 735
Leu

<210> 438
<211> 406
<212> PRT
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多肽
<400> 438

[0099]

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

Glu Phe Phe Met Lys Val Tyr Gly Tyr Arg Gly Glu His Leu Gly Gly
260 265 270
Ser Arg Lys Pro Asp Gly Ala Ile Tyr Thr Val Gly Ser Pro Ile Asp
275 280 285
Tyr Gly Val Ile Val Asp Thr Lys Ala Tyr Ser Gly Gly Tyr Asn Leu
290 295 300
Pro Ile Gly Gln Ala Asp Glu Met Glu Arg Tyr Val Glu Glu Asn Gln
305 310 315 320
Thr Arg Asp Lys His Leu Asn Pro Asn Glu Trp Trp Lys Val Tyr Pro
325 330 335
Ser Ser Val Thr Glu Phe Lys Phe Leu Phe Val Ser Gly His Phe Lys
340 345 350
Gly Asn Tyr Lys Ala Gln Leu Thr Arg Leu Asn His Ile Thr Asn Cys
355 360 365
Asn Gly Ala Val Leu Ser Val Glu Glu Leu Leu Ile Gly Gly Glu Met
370 375 380
Ile Lys Ala Gly Thr Leu Thr Leu Glu Glu Val Arg Arg Lys Phe Asn
385 390 395 400
Asn Gly Glu Ile Asn Phe
405

<210> 439
<211> 406
<212> PRT
[0100] <213> 人工序列
<220>
<223> 人工序列的描述：合成
多肽
<400> 439
Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp
1 5 10 15
Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys Arg Lys Val
20 25 30
Gly Arg Leu Glu Pro Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly
35 40 45
Lys Ser Phe Ser Gln Gln Arg Ser Leu Val Gly His Gln Arg Thr His
50 55 60
Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser
65 70 75 80
Asp Lys Lys Asp Leu Thr Arg His Gln Arg Thr His Thr Gly Glu Lys
85 90 95
Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Thr Ser Gly His
100 105 110
Leu Val Arg His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys
115 120 125
Pro Glu Cys Gly Lys Ser Phe Ser Gln Arg Ala His Leu Glu Arg His
130 135 140

[0101]

```
Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly
145                 150                 155                 160
Lys Ser Phe Ser Thr Ser Gly Ser Leu Val Arg His Gln Arg Thr His
                165                 170                 175
Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser
            180                 185                 190
Thr Ser Gly Asn Leu Val Arg His Gln Arg Thr His Thr Gly Ala Ala
        195                 200                 205
Ala Arg Ala Leu Val Lys Ser Glu Leu Glu Glu Lys Lys Ser Glu Leu
    210                 215                 220
Arg His Lys Leu Lys Tyr Val Pro His Glu Tyr Ile Glu Leu Ile Glu
225                 230                 235                 240
Ile Ala Arg Asn Pro Thr Gln Asp Arg Ile Leu Glu Met Lys Val Met
                245                 250                 255
Glu Phe Phe Met Lys Val Tyr Gly Tyr Arg Gly Glu His Leu Gly Gly
            260                 265                 270
Ser Arg Lys Pro Asp Gly Ala Ile Tyr Thr Val Gly Ser Pro Ile Asp
        275                 280                 285
Tyr Gly Val Ile Val Asp Thr Lys Ala Tyr Ser Gly Gly Tyr Asn Leu
    290                 295                 300
Pro Ile Gly Gln Ala Asp Glu Met Gln Arg Tyr Val Lys Glu Asn Gln
305                 310                 315                 320
Thr Arg Asn Lys His Ile Asn Pro Asn Glu Trp Trp Lys Val Tyr Pro
                325                 330                 335
Ser Ser Val Thr Glu Phe Lys Phe Leu Phe Val Ser Gly His Phe Lys
            340                 345                 350
Gly Asn Tyr Lys Ala Gln Leu Thr Arg Leu Asn Arg Lys Thr Asn Cys
        355                 360                 365
Asn Gly Ala Val Leu Ser Val Glu Glu Leu Leu Ile Gly Gly Glu Met
    370                 375                 380
Ile Lys Ala Gly Thr Leu Thr Leu Glu Glu Val Arg Arg Lys Phe Asn
385                 390                 395                 400
Asn Gly Glu Ile Asn Phe
                405
<210> 440
<211> 320
<212> PRT
<213> 人工序列
<220>
<223> 人工序列的描述：合成
      多肽
<400> 440
Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp
1               5                   10                  15
Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys Arg Lys Val
            20                  25                  30
```

[0102]

Gly Arg Leu Glu Pro Gly Glu Arg Pro Phe Gln Cys Arg Ile Cys Met
35 40 45
Arg Asn Phe Ser Ser Lys Gln Ala Leu Ala Val His Thr Arg Thr His
50 55 60
Thr Gly Glu Lys Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser
65 70 75 80
Gln Ser Thr Thr Leu Lys Arg His Leu Arg Thr His Thr Gly Glu Lys
85 90 95
Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Arg Ser Asp His
100 105 110
Leu Ser Leu His Leu Lys Thr His Leu Arg Gly Ser Gln Leu Val Lys
115 120 125
Ser Glu Leu Glu Glu Lys Lys Ser Glu Leu Arg His Lys Leu Lys Tyr
130 135 140
Val Pro His Glu Tyr Ile Glu Leu Ile Glu Ile Ala Arg Asn Pro Thr
145 150 155 160
Gln Asp Arg Ile Leu Glu Met Lys Val Met Glu Phe Phe Met Lys Val
165 170 175
Tyr Gly Tyr Arg Gly Glu His Leu Gly Gly Ser Arg Lys Pro Asp Gly
180 185 190
Ala Ile Tyr Thr Val Gly Ser Pro Ile Asp Tyr Gly Val Ile Val Asp
195 200 205
Thr Lys Ala Tyr Ser Gly Gly Tyr Asn Leu Pro Ile Gly Gln Ala Asp
210 215 220
Glu Met Gln Arg Tyr Val Lys Glu Asn Gln Thr Arg Asn Lys His Ile
225 230 235 240
Asn Pro Asn Glu Trp Trp Lys Val Tyr Pro Ser Ser Val Thr Glu Phe
245 250 255
Lys Phe Leu Phe Val Ser Gly His Phe Lys Gly Asn Tyr Lys Ala Gln
260 265 270
Leu Thr Arg Leu Asn Arg Lys Thr Asn Cys Asn Gly Ala Val Leu Ser
275 280 285
Val Glu Glu Leu Leu Ile Gly Gly Glu Met Ile Lys Ala Gly Thr Leu
290 295 300
Thr Leu Glu Glu Val Arg Arg Lys Phe Asn Asn Gly Glu Ile Asn Phe
305 310 315 320

<210> 441
<211> 320
<212> PRT
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多肽
<400> 441

Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp
1 5 10 15

[0103]

Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys Arg Lys Val
20 25 30
Gly Arg Leu Glu Pro Gly Glu Arg Pro Phe Gln Cys Arg Ile Cys Met
35 40 45
Arg Asn Phe Ser Arg Arg Ala His Leu Gln Asn His Thr Arg Thr His
50 55 60
Thr Gly Glu Lys Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser
65 70 75 80
Gln Ser Thr Thr Leu Lys Arg His Leu Arg Thr His Thr Gly Glu Lys
85 90 95
Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Asp Gly Gly His
100 105 110
Leu Thr Arg His Leu Lys Thr His Leu Arg Gly Ser Gln Leu Val Lys
115 120 125
Ser Glu Leu Glu Glu Lys Lys Ser Glu Leu Arg His Lys Leu Lys Tyr
130 135 140
Val Pro His Glu Tyr Ile Glu Leu Ile Glu Ile Ala Arg Asn Pro Thr
145 150 155 160
Gln Asp Arg Ile Leu Glu Met Lys Val Met Glu Phe Phe Met Lys Val
165 170 175
Tyr Gly Tyr Arg Gly Glu His Leu Gly Gly Ser Arg Lys Pro Asp Gly
180 185 190
Ala Ile Tyr Thr Val Gly Ser Pro Ile Asp Tyr Gly Val Ile Val Asp
195 200 205
Thr Lys Ala Tyr Ser Gly Gly Tyr Asn Leu Pro Ile Gly Gln Ala Asp
210 215 220
Glu Met Glu Arg Tyr Val Glu Glu Asn Gln Thr Arg Asp Lys His Leu
225 230 235 240
Asn Pro Asn Glu Trp Trp Lys Val Tyr Pro Ser Ser Val Thr Glu Phe
245 250 255
Lys Phe Leu Phe Val Ser Gly His Phe Lys Gly Asn Tyr Lys Ala Gln
260 265 270
Leu Thr Arg Leu Asn His Ile Thr Asn Cys Asn Gly Ala Val Leu Ser
275 280 285
Val Glu Glu Leu Leu Ile Gly Gly Glu Met Ile Lys Ala Gly Thr Leu
290 295 300
Thr Leu Glu Glu Val Arg Arg Lys Phe Asn Asn Gly Glu Ile Asn Phe
305 310 315 320

<210> 442
<211> 42
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 442

caaactagaa atgccatctt ccttgatgtt ggaggtacct gc 42

<210> 443

<211> 406

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述：合成
多肽

<400> 443

[0104]

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

290 295 300
Pro Ile Gly Gln Ala Asp Glu Met Glu Arg Tyr Val Glu Glu Asn Gln
305 310 315 320
Thr Arg Asp Lys His Leu Asn Pro Asn Glu Trp Trp Lys Val Tyr Pro
325 330 335
Ser Ser Val Thr Glu Phe Lys Phe Leu Phe Val Ser Gly His Phe Lys
340 345 350
Gly Asn Tyr Lys Ala Gln Leu Thr Arg Leu Asn His Ile Thr Asn Cys
355 360 365
Asn Gly Ala Val Leu Ser Val Glu Glu Leu Leu Ile Gly Gly Glu Met
370 375 380
Ile Lys Ala Gly Thr Leu Thr Leu Glu Glu Val Arg Arg Lys Phe Asn
385 390 395 400
Asn Gly Glu Ile Asn Phe
405

<210> 444
<211> 406
<212> PRT
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多肽
<400> 444

[0105]

Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp
1 5 10 15
Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys Arg Lys Val
20 25 30
Gly Arg Leu Glu Pro Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly
35 40 45
Lys Ser Phe Ser Gln Gln Arg Ser Leu Val Gly His Gln Arg Thr His
50 55 60
Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser
65 70 75 80
Asp Lys Lys Asp Leu Thr Arg His Gln Arg Thr His Thr Gly Glu Lys
85 90 95
Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Thr Ser Gly His
100 105 110
Leu Val Arg His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys
115 120 125
Pro Glu Cys Gly Lys Ser Phe Ser Gln Arg Ala His Leu Glu Arg His
130 135 140
Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly
145 150 155 160
Lys Ser Phe Ser Thr Ser Gly Ser Leu Val Arg His Gln Arg Thr His
165 170 175
Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser

[0106]

180 185 190
Thr Ser Gly Asn Leu Val Arg His Gln Arg Thr His Thr Gly Ala Ala
195 200 205
Ala Arg Ala Leu Val Lys Ser Glu Leu Glu Glu Lys Lys Ser Glu Leu
210 215 220
Arg His Lys Leu Lys Tyr Val Pro His Glu Tyr Ile Glu Leu Ile Glu
225 230 235 240
Ile Ala Arg Asn Pro Thr Gln Asp Arg Ile Leu Glu Met Lys Val Met
245 250 255
Glu Phe Phe Met Lys Val Tyr Gly Tyr Arg Gly Glu His Leu Gly Gly
260 265 270
Ser Arg Lys Pro Asp Gly Ala Ile Tyr Thr Val Gly Ser Pro Ile Asp
275 280 285
Tyr Gly Val Ile Val Asp Thr Lys Ala Tyr Ser Gly Gly Tyr Asn Leu
290 295 300
Pro Ile Gly Gln Ala Asp Glu Met Gln Arg Tyr Val Lys Glu Asn Gln
305 310 315 320
Thr Arg Asn Lys His Ile Asn Pro Asn Glu Trp Trp Lys Val Tyr Pro
325 330 335
Ser Ser Val Thr Glu Phe Lys Phe Leu Phe Val Ser Gly His Phe Lys
340 345 350
Gly Asn Tyr Lys Ala Gln Leu Thr Arg Leu Asn Arg Lys Thr Asn Cys
355 360 365
Asn Gly Ala Val Leu Ser Val Glu Glu Leu Leu Ile Gly Gly Glu Met
370 375 380
Ile Lys Ala Gly Thr Leu Thr Leu Glu Glu Val Arg Arg Lys Phe Asn
385 390 395 400
Asn Gly Glu Ile Asn Phe
405

<210> 445
<211> 25
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<220>
<221> 其它特征
<222> (1)..(1)
<223> 可存在或者可不存在
<220>
<221> 其它特征
<222> (25)..(25)
<223> 可存在或者可不存在
<400> 445
tatctgccca tgactggcgc aggga 25

<210> 446
<211> 320
<212> PRT
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多肽
<400> 446

[0107]

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

Thr Leu Glu Glu Val Arg Arg Lys Phe Asn Asn Gly Glu Ile Asn Phe
305 310 315 320

<210> 447
<211> 320
<212> PRT
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多肽
<400> 447

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

[0108]

Val Glu Glu Leu Leu Ile Gly Gly Glu Met Ile Lys Ala Gly Thr Leu
290 295 300
Thr Leu Glu Glu Val Arg Arg Lys Phe Asn Asn Gly Glu Ile Asn Phe
305 310 315 320
<210> 448
<211> 42
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 448
actagaaatg ccatcttcct tgatgttgga ggtacctgct ct 42
<210> 449
<211> 406
<212> PRT
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多肽
<400> 449
Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp
1 5 10 15
[0109]
Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys Arg Lys Val
20 25 30
Gly Arg Leu Glu Pro Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly
35 40 45
Lys Ser Phe Ser His Arg Thr Thr Leu Thr Asn His Gln Arg Thr His
50 55 60
Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser
65 70 75 80
Gln Arg Asn Ala Leu Ala Gly His Gln Arg Thr His Thr Gly Glu Lys
85 90 95
Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser His Lys Asn Ala
100 105 110
Leu Gln Asn His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys
115 120 125
Pro Glu Cys Gly Lys Ser Phe Ser Asp Pro Gly His Leu Val Arg His
130 135 140
Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly
145 150 155 160
Lys Ser Phe Ser Thr Ser Gly Asn Leu Val Arg His Gln Arg Thr His
165 170 175
Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser
180 185 190
Gln Ser Ser Asn Leu Val Arg His Gln Arg Thr His Thr Gly Ala Ala

[0110]

```
          195                 200                 205
Ala Arg Ala Leu Val Lys Ser Glu Leu Glu Glu Lys Lys Ser Glu Leu
      210                 215                 220
Arg His Lys Leu Lys Tyr Val Pro His Glu Tyr Ile Glu Leu Ile Glu
225                 230                 235                 240
Ile Ala Arg Asn Pro Thr Gln Asp Arg Ile Leu Glu Met Lys Val Met
                245                 250                 255
Glu Phe Phe Met Lys Val Tyr Gly Tyr Arg Gly Glu His Leu Gly Gly
            260                 265                 270
Ser Arg Lys Pro Asp Gly Ala Ile Tyr Thr Val Gly Ser Pro Ile Asp
          275                 280                 285
Tyr Gly Val Ile Val Asp Thr Lys Ala Tyr Ser Gly Gly Tyr Asn Leu
      290                 295                 300
Pro Ile Gly Gln Ala Asp Glu Met Glu Arg Tyr Val Glu Glu Asn Gln
305                 310                 315                 320
Thr Arg Asp Lys His Leu Asn Pro Asn Glu Trp Trp Lys Val Tyr Pro
                325                 330                 335
Ser Ser Val Thr Glu Phe Lys Phe Leu Phe Val Ser Gly His Phe Lys
            340                 345                 350
Gly Asn Tyr Lys Ala Gln Leu Thr Arg Leu Asn His Ile Thr Asn Cys
          355                 360                 365
Asn Gly Ala Val Leu Ser Val Glu Glu Leu Leu Ile Gly Gly Glu Met
      370                 375                 380
Ile Lys Ala Gly Thr Leu Thr Leu Glu Glu Val Arg Arg Lys Phe Asn
385                 390                 395                 400
Asn Gly Glu Ile Asn Phe
                405
<210> 450
<211> 406
<212> PRT
<213> 人工序列
<220>
<223> 人工序列的描述：合成
      多肽
<400> 450
Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp
1               5                   10                  15
Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys Arg Lys Val
            20                  25                  30
Gly Arg Leu Glu Pro Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly
          35                  40                  45
Lys Ser Phe Ser Gln Arg Asn Ala Leu Ala Gly His Gln Arg Thr His
      50                  55                  60
Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser
65                  70                  75                  80
Gln Gln Arg Ser Leu Val Gly His Gln Arg Thr His Thr Gly Glu Lys
```

[0111]

85 90 95
Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Asp Lys Lys Asp
100 105 110
Leu Thr Arg His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys
115 120 125
Pro Glu Cys Gly Lys Ser Phe Ser Thr Ser Gly His Leu Val Arg His
130 135 140
Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly
145 150 155 160
Lys Ser Phe Ser Gln Arg Ala His Leu Glu Arg His Gln Arg Thr His
165 170 175
Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser
180 185 190
Thr Ser Gly Ser Leu Val Arg His Gln Arg Thr His Thr Gly Ala Ala
195 200 205
Ala Arg Ala Leu Val Lys Ser Glu Leu Glu Glu Lys Lys Ser Glu Leu
210 215 220
Arg His Lys Leu Lys Tyr Val Pro His Glu Tyr Ile Glu Leu Ile Glu
225 230 235 240
Ile Ala Arg Asn Pro Thr Gln Asp Arg Ile Leu Glu Met Lys Val Met
245 250 255
Glu Phe Phe Met Lys Val Tyr Gly Tyr Arg Gly Glu His Leu Gly Gly
260 265 270
Ser Arg Lys Pro Asp Gly Ala Ile Tyr Thr Val Gly Ser Pro Ile Asp
275 280 285
Tyr Gly Val Ile Val Asp Thr Lys Ala Tyr Ser Gly Gly Tyr Asn Leu
290 295 300
Pro Ile Gly Gln Ala Asp Glu Met Gln Arg Tyr Val Lys Glu Asn Gln
305 310 315 320
Thr Arg Asn Lys His Ile Asn Pro Asn Glu Trp Trp Lys Val Tyr Pro
325 330 335
Ser Ser Val Thr Glu Phe Lys Phe Leu Phe Val Ser Gly His Phe Lys
340 345 350
Gly Asn Tyr Lys Ala Gln Leu Thr Arg Leu Asn Arg Lys Thr Asn Cys
355 360 365
Asn Gly Ala Val Leu Ser Val Glu Glu Leu Leu Ile Gly Gly Glu Met
370 375 380
Ile Lys Ala Gly Thr Leu Thr Leu Glu Glu Val Arg Arg Lys Phe Asn
385 390 395 400
Asn Gly Glu Ile Asn Phe
405

<210> 451
<211> 378
<212> PRT
<213> 人工序列
<220>

<223> 人工序列的描述：合成
多肽
<400> 451

[0112]

```
Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp
1               5                  10                  15
Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys Arg Lys Val
            20                  25                  30
Gly Arg Leu Glu Pro Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly
        35                  40                  45
Lys Ser Phe Ser Gln Arg Asn Ala Leu Ala Gly His Gln Arg Thr His
    50                  55                  60
Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser
65                  70                  75                  80
His Lys Asn Ala Leu Gln Asn His Gln Arg Thr His Thr Gly Glu Lys
                85                  90                  95
Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Asp Pro Gly His
            100                 105                 110
Leu Val Arg His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys
        115                 120                 125
Pro Glu Cys Gly Lys Ser Phe Ser Thr Ser Gly Asn Leu Val Arg His
    130                 135                 140
Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly
145                 150                 155                 160
Lys Ser Phe Ser Gln Ser Ser Asn Leu Val Arg His Gln Arg Thr His
                165                 170                 175
Thr Gly Ala Ala Ala Arg Ala Leu Val Lys Ser Glu Leu Glu Glu Lys
            180                 185                 190
Lys Ser Glu Leu Arg His Lys Leu Lys Tyr Val Pro His Glu Tyr Ile
        195                 200                 205
Glu Leu Ile Glu Ile Ala Arg Asn Pro Thr Gln Asp Arg Ile Leu Glu
    210                 215                 220
Met Lys Val Met Glu Phe Phe Met Lys Val Tyr Gly Tyr Arg Gly Glu
225                 230                 235                 240
His Leu Gly Gly Ser Arg Lys Pro Asp Gly Ala Ile Tyr Thr Val Gly
                245                 250                 255
Ser Pro Ile Asp Tyr Gly Val Ile Val Asp Thr Lys Ala Tyr Ser Gly
            260                 265                 270
Gly Tyr Asn Leu Pro Ile Gly Gln Ala Asp Glu Met Glu Arg Tyr Val
        275                 280                 285
Glu Glu Asn Gln Thr Arg Asp Lys His Leu Asn Pro Asn Glu Trp Trp
    290                 295                 300
Lys Val Tyr Pro Ser Ser Val Thr Glu Phe Lys Phe Leu Phe Val Ser
305                 310                 315                 320
Gly His Phe Lys Gly Asn Tyr Lys Ala Gln Leu Thr Arg Leu Asn His
                325                 330                 335
Ile Thr Asn Cys Asn Gly Ala Val Leu Ser Val Glu Glu Leu Leu Ile
```

340 345 350
Gly Gly Glu Met Ile Lys Ala Gly Thr Leu Thr Leu Glu Glu Val Arg
355 360 365
Arg Lys Phe Asn Asn Gly Glu Ile Asn Phe
370 375
<210> 452
<211> 378
<212> PRT
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多肽
<400> 452
Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp
1 5 10 15
Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys Arg Lys Val
20 25 30
Gly Arg Leu Glu Pro Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly
35 40 45
Lys Ser Phe Ser Gln Gln Arg Ser Leu Val Gly His Gln Arg Thr His
50 55 60
Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser
65 70 75 80
[0113]
Asp Lys Lys Asp Leu Thr Arg His Gln Arg Thr His Thr Gly Glu Lys
85 90 95
Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Thr Ser Gly His
100 105 110
Leu Val Arg His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys
115 120 125
Pro Glu Cys Gly Lys Ser Phe Ser Gln Arg Ala His Leu Glu Arg His
130 135 140
Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly
145 150 155 160
Lys Ser Phe Ser Thr Ser Gly Ser Leu Val Arg His Gln Arg Thr His
165 170 175
Thr Gly Ala Ala Ala Arg Ala Leu Val Lys Ser Glu Leu Glu Glu Lys
180 185 190
Lys Ser Glu Leu Arg His Lys Leu Lys Tyr Val Pro His Glu Tyr Ile
195 200 205
Glu Leu Ile Glu Ile Ala Arg Asn Pro Thr Gln Asp Arg Ile Leu Glu
210 215 220
Met Lys Val Met Glu Phe Phe Met Lys Val Tyr Gly Tyr Arg Gly Glu
225 230 235 240
His Leu Gly Gly Ser Arg Lys Pro Asp Gly Ala Ile Tyr Thr Val Gly
245 250 255
Ser Pro Ile Asp Tyr Gly Val Ile Val Asp Thr Lys Ala Tyr Ser Gly

260 265 270
Gly Tyr Asn Leu Pro Ile Gly Gln Ala Asp Glu Met Gln Arg Tyr Val
275 280 285
Lys Glu Asn Gln Thr Arg Asn Lys His Ile Asn Pro Asn Glu Trp Trp
290 295 300
Lys Val Tyr Pro Ser Ser Val Thr Glu Phe Lys Phe Leu Phe Val Ser
305 310 315 320
Gly His Phe Lys Gly Asn Tyr Lys Ala Gln Leu Thr Arg Leu Asn Arg
325 330 335
Lys Thr Asn Cys Asn Gly Ala Val Leu Ser Val Glu Glu Leu Leu Ile
340 345 350
Gly Gly Glu Met Ile Lys Ala Gly Thr Leu Thr Leu Glu Glu Val Arg
355 360 365
Arg Lys Phe Asn Asn Gly Glu Ile Asn Phe
370 375

<210> 453
<211> 26
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸

[0114]

<220>
<221> 其它特征
<222> (1)..(1)
<223> 可存在或者可不存在
<220>
<221> 其它特征
<222> (26)..(26)
<223> 可存在或者可不存在
<400> 453
tgccatcttc cttgatgttg gaggta 26

<210> 454
<211> 322
<212> PRT
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多肽
<400> 454
Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp
1 5 10 15
Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys Arg Lys Val
20 25 30
Gly Arg Leu Glu Pro Gly Glu Arg Pro Phe Gln Cys Arg Ile Cys Met
35 40 45

Arg Asn Phe Ser Ser Pro Ser Lys Leu Ala Arg His Thr Arg Thr His
50 55 60
Thr Gly Glu Lys Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser
65 70 75 80
Val Arg His Asn Leu Thr Arg His Leu Arg Thr His Thr Gly Glu Lys
85 90 95
Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Gln Arg Asn Asn
100 105 110
Leu Gly Arg His Leu Lys Thr His Thr Gly Ala Ala Ala Arg Ala Leu
115 120 125
Val Lys Ser Glu Leu Glu Glu Lys Lys Ser Glu Leu Arg His Lys Leu
130 135 140
Lys Tyr Val Pro His Glu Tyr Ile Glu Leu Ile Glu Ile Ala Arg Asn
145 150 155 160
Pro Thr Gln Asp Arg Ile Leu Glu Met Lys Val Met Glu Phe Phe Met
165 170 175
Lys Val Tyr Gly Tyr Arg Gly Glu His Leu Gly Gly Ser Arg Lys Pro
180 185 190
Asp Gly Ala Ile Tyr Thr Val Gly Ser Pro Ile Asp Tyr Gly Val Ile
195 200 205
Val Asp Thr Lys Ala Tyr Ser Gly Gly Tyr Asn Leu Pro Ile Gly Gln
210 215 220
[0115] Ala Asp Glu Met Gln Arg Tyr Val Lys Glu Asn Gln Thr Arg Asn Lys
225 230 235 240
His Ile Asn Pro Asn Glu Trp Trp Lys Val Tyr Pro Ser Ser Val Thr
245 250 255
Glu Phe Lys Phe Leu Phe Val Ser Gly His Phe Lys Gly Asn Tyr Lys
260 265 270
Ala Gln Leu Thr Arg Leu Asn Arg Lys Thr Asn Cys Asn Gly Ala Val
275 280 285
Leu Ser Val Glu Glu Leu Leu Ile Gly Gly Glu Met Ile Lys Ala Gly
290 295 300
Thr Leu Thr Leu Glu Glu Val Arg Arg Lys Phe Asn Asn Gly Glu Ile
305 310 315 320
Asn Phe

<210> 455
<211> 322
<212> PRT
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多肽
<400> 455
Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp
1 5 10 15

[0116]

Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys Arg Lys Val
20 25 30
Gly Arg Leu Glu Pro Gly Glu Arg Pro Phe Gln Cys Arg Ile Cys Met
35 40 45
Arg Asn Phe Ser Ile Pro Asn His Leu Ala Arg His Thr Arg Thr His
50 55 60
Thr Gly Glu Lys Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser
65 70 75 80
Gln Ser Ala His Leu Lys Arg His Leu Arg Thr His Thr Gly Glu Lys
85 90 95
Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser His His Asn Ser
100 105 110
Leu Thr Arg His Leu Lys Thr His Thr Gly Ala Ala Ala Arg Ala Leu
115 120 125
Val Lys Ser Glu Leu Glu Glu Lys Lys Ser Glu Leu Arg His Lys Leu
130 135 140
Lys Tyr Val Pro His Glu Tyr Ile Glu Leu Ile Glu Ile Ala Arg Asn
145 150 155 160
Pro Thr Gln Asp Arg Ile Leu Glu Met Lys Val Met Glu Phe Phe Met
165 170 175
Lys Val Tyr Gly Tyr Arg Gly Glu His Leu Gly Gly Ser Arg Lys Pro
180 185 190
Asp Gly Ala Ile Tyr Thr Val Gly Ser Pro Ile Asp Tyr Gly Val Ile
195 200 205
Val Asp Thr Lys Ala Tyr Ser Gly Gly Tyr Asn Leu Pro Ile Gly Gln
210 215 220
Ala Asp Glu Met Glu Arg Tyr Val Glu Glu Asn Gln Thr Arg Asp Lys
225 230 235 240
His Leu Asn Pro Asn Glu Trp Trp Lys Val Tyr Pro Ser Ser Val Thr
245 250 255
Glu Phe Lys Phe Leu Phe Val Ser Gly His Phe Lys Gly Asn Tyr Lys
260 265 270
Ala Gln Leu Thr Arg Leu Asn His Ile Thr Asn Cys Asn Gly Ala Val
275 280 285
Leu Ser Val Glu Glu Leu Leu Ile Gly Gly Glu Met Ile Lys Ala Gly
290 295 300
Thr Leu Thr Leu Glu Glu Val Arg Arg Lys Phe Asn Asn Gly Glu Ile
305 310 315 320
Asn Phe

<210> 456
<211> 6626
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成

多核苷酸

<400> 456

[0117]

ggggggggag ggggggggtt ggccactccc tctctgcgcg ctcgctcgct cactgaggcc 60
gggcgaccaa aggtcgcccg acgcccgggc tttgcccggg cggcctcagt gagcgagcga 120
gcgcgcagag agggagtggc caactccatc actaggggtt cctagatctg aattcggtac 180
ccgttacata acttacggta aatggcccgc ctggctgacc gcccaacgac ccccgcccat 240
tgacgtcaat aatgacgtat gttcccatag taacgccaat agggactttc cattgacgtc 300
aatgggtgga gtatttacgg taaactgccc acttggcagt acatcaagtg tatcatatgc 360
caagtacgcc ccctattgac gtcaatgacg gtaaatggcc cgcctggcat tatgcccagt 420
acatgacctt atgggacttt cctacttggc agtacatcta cgtattagtc atcgctatta 480
ccatggtgat gcggttttgg cagtacatca atgggcgtgg atagcggttt gactcacggg 540
gatttccaag tctccacccc attgacgtca atgggagttt gttttggcac caaaatcaac 600
gggactttcc aaaatgtcgt aacaactccg ccccattgac gcaaatgggc ggtaggcgtg 660
tacggtggga ggtctatata agcagagctc gtttagtgaa ccgtcagatc gcctggagac 720
gccatccacg ctgttttgac ctccatagaa gacaccggga ccgatccagc ctccggactc 780
tagaggatcc ggtactcgag gaactgaaaa accagaaagt taactggtaa gtttagtctt 840
tttgtctttt atttcaggtc ccggatccgg tggtggtgca aatcaaagaa ctgctcctca 900
gtggatgttg cctttacttc taggcctgta cggaagtgtt acttctgctc taaaagctgc 960
ggaattgtac ccgcggcccg ggatccaccg gtggctagcg tctataggcc cacccccttg 1020
gtggaattcg ccatgaggtc tgactacaaa gaccatgacg gtgattataa agatcatgac 1080
atcgattaca aggatgacga tgacaagatg gcccccaaga agaagaggaa ggtgggcctc 1140
gagcccggag aaaaaccgta caagtgccct gagtgcggga aatcattctc cgaccctggg 1200
gcgctcgtcc ggcaccaaag gacgcataca ggggaaaagc cgtataagtg ccccgagtgt 1260
ggaaagagct tctcgcagag agcccacctt gaacgacacc aaagaacaca cactggtgag 1320
aaaccctata agtgtccaga gtgcggcaaa tcgtttagca gatccgatga cttggtgcgc 1380
caccagcgga cacacacggg tgaaaagccc tacaaatgcc cggagtgtgg gaagtcgttt 1440
tcaaggtcgg atcatctgac tacccatcag cgcacccata cgggagcggc cgcccgcgcc 1500
ctggtgaaga gcgagctgga ggagaagaag tccgagctgc ggcacaagct gaagtacgtg 1560
ccccacgagt acatcgagct gatcgagatc gccaggaacc ccacccagga ccgcatcctg 1620
gagatgaagg tgatggagtt cttcatgaag gtgtacggct acaggggaga gcacctgggc 1680
ggaagcagaa agcctgacgg cgccatctat acagtgggca gccccatcga ttacggcgtg 1740
atcgtggaca caaaggccta cagcggcggc tacaatctgc ctatcggcca ggccgacgag 1800
atgcagagat acgtgaagga gaaccagacc cggaataagc acatcaaccc caacgagtgg 1860
tggaaggtgt accctagcag cgtgaccgag ttcaagttcc tgttcgtgag cggccacttc 1920
aagggcaact acaaggccca gctgaccagg ctgaaccgca aaaccaactg caatggcgcc 1980
gtgctgagcg tggaggagct gctgatcggc ggcgagatga tcaaagccgg caccctgaca 2040
ctggaggagg tgcggcgcaa gttcaacaac ggcgagatca acttcgaggg cagaggaagt 2100
cttctaacat gcggtgacgt ggaggagaat cccggcccta gatctgacta caaagaccat 2160
gacggtgatt ataaagatca tgacatcgat tacaaggatg acgatgacaa gatggccccc 2220
aagaagaaga ggaaggtggg cctcgagccg ggagagaagc cgtacaagtg tcccgaatgt 2280
ggaaagagct tctcacagtc gggggacctt cggcgccacc agcgcacaca tactggtgaa 2340
aagccgtata agtgtccaga atgtggcaaa tcattctcca catcagggag cctggtcagg 2400
caccagcgaa cccacacggg tgagaagccc tataagtgcc ccgaatgcgg gaagtccttt 2460
tcgcagagag cccacttgga gaggcaccag aggacccata cgggggagaa accttacaag 2520
tgccctgaat gcggaaagtc gttctcgacc catctggatc tcatcagaca tcagagaacg 2580
cacactggag agaaacccta caaatgtccc gagtgtggga agtcgtttag ccgaaaggac 2640

[0118]

aatctcaaaa accatcaacg gacacacacg ggtgaaaaac catacaaatg cccggagtgc 2700
ggcaaatcgt tttcccaact tgcgcacttg cgggcacacc aacgcacgca tactggagcg 2760
gccgcccgcg ccctggtgaa gagcgagctg gaggagaaga agtccgagct gcggcacaag 2820
ctgaagtacg tgccccacga gtacatcgag ctgatcgaga tcgccaggaa ccccacccag 2880
gaccgcatcc tggagatgaa ggtgatggag ttcttcatga aggtgtacgg ctacagggga 2940
gagcacctgg gcggaagcag aaagcctgac ggcgccatct atacagtggg cagccccatc 3000
gattacggcg tgatcgtgga cacaaaggcc tacagcggcg gctacaatct gcctatcggc 3060
caggccgacg agatggagag atacgtggag gagaaccaga cacgggataa gcacctcaac 3120
cccaacgagt ggtggaaggt gtaccctagc agcgtgaccg agttcaagtt cctgttcgtg 3180
agcggccact tcaagggcaa ctacaaggcc cagctgacca ggctgaacca catcaccaac 3240
tgcaatggcg ccgtgctgag cgtggaggag ctgctgatcg gcggcgagat gatcaaagcc 3300
ggcaccctga cactggagga ggtgcggcgc aagttcaaca acggcgagat caacttctga 3360
ttaattaact aatctagagt cgactagagc tcgctgatca gcctcgactg tgccttctag 3420
ttgccagcca tctgttgttt gcccctcccc cgtgccttcc ttgaccctgg aaggtgccac 3480
tcccactgtc ctttcctaat aaaatgagga aattgcatcg cattgtctga gtaggtgtca 3540
ttctattctg gggggtgggg tggggcagga cagcaagggg gaggattggg aagacaatag 3600
caggcatgct ggggagagat ctaggaaccc ctagtgatgg agttggccac tccctctctg 3660
cgcgctcgct cgctcactga ggccgcccgg gcaaagcccg ggcgtcgggc gacctttggt 3720
cgcccggcct cagtgagcga gcgagcgcgc agagagggag tggccaaccc cccccccccc 3780
ccccctgcag cccagctgca ttaatgaatc ggccaacgcg cggggagagg cggtttgcgt 3840
attgggcgct cttccgcttc ctcgctcact gactcgctgc gctcggtcgt tcggctgcgg 3900
cgagcggtat cagctcactc aaaggcggta atacggttat ccacagaatc aggggataac 3960
gcaggaaaga acatgtgagc aaaaggccag caaaaggcca ggaaccgtaa aaaggccgcg 4020
ttgctggcgt ttttccatag gctccgcccc cctgacgagc atcacaaaaa tcgacgctca 4080
agtcagaggt ggcgaaaccc gacaggacta taaagatacc aggcgtttcc ccctggaagc 4140
tccctcgtgc gctctcctgt tccgaccctg ccgcttaccg gatacctgtc cgcctttctc 4200
ccttcgggaa gcgtggcgct ttctcaatgc tcacgctgta ggtatctcag ttcggtgtag 4260
gtcgttcgct ccaagctggg ctgtgtgcac gaaccccccg ttcagcccga ccgctgcgcc 4320
ttatccggta actatcgtct tgagtccaac ccggtaagac acgacttatc gccactggca 4380
gcagccactg gtaacaggat tagcagagcg aggtatgtag gcggtgctac agagttcttg 4440
aagtggtggc ctaactacgg ctacactaga aggacagtat ttggtatctg cgctctgctg 4500
aagccagtta ccttcggaaa aagagttggt agctcttgat ccggcaaaca aaccaccgct 4560
ggtagcggtg gttttttgt ttgcaagcag cagattacgc gcagaaaaaa aggatctcaa 4620
gaagatcctt tgatcttttc tacggggtct gacgctcagt ggaacgaaaa ctcacgttaa 4680
gggattttgg tcatgagatt atcaaaaagg atcttcacct agatcctttt aaattaaaaa 4740
tgaagtttta aatcaatcta aagtatatat gagtaaactt ggtctgacag ttaccaatgc 4800
ttaatcagtg aggcacctat ctcagcgatc tgtctatttc gttcatccat agttgcctga 4860
ctccccgtcg tgtagataac tacgatacgg gagggcttac catctggccc cagtgctgca 4920
atgataccgc gagacccacg ctcaccggct ccagatttat cagcaataaa ccagccagcc 4980
ggaagggccg agcgcagaag tggtcctgca actttatccg cctccatcca gtctattaat 5040
tgttgccggg aagctagagt aagtagttcg ccagttaata gtttgcgcaa cgttgttgcc 5100
attgctacag gcatcgtggt gtcacgctcg tcgtttggta tggcttcatt cagctccggt 5160
tcccaacgat caaggcgagt tacatgatcc cccatgttgt gcaaaaaagc ggttagctcc 5220
ttcggtcctc cgatcgttgt cagaagtaag ttggccgcag tgttatcact catggttatg 5280
gcagcactgc ataattctct tactgtcatg ccatccgtaa gatgcttttc tgtgactggt 5340
gagtactcaa ccaagtcatt ctgagaatag tgtatgcggc gaccgagttg ctcttgcccg 5400

gcgtcaatac gggataatac cgcgccacat agcagaactt taaaagtgct catcattgga 5460
aaacgttctt cggggcgaaa actctcaagg atcttaccgc tgttgagatc cagttcgatg 5520
taacccactc gtgcacccaa ctgatcttca gcatctttta ctttcaccag cgtttctggg 5580
tgagcaaaaa caggaaggca aaatgccgca aaaaagggaa taagggcgac acggaaatgt 5640
tgaatactca tactcttcct ttttcaatat tattgaagca tttatcaggg ttattgtctc 5700
atgagcggat acatatttga atgtatttag aaaaataaac aaataggggt tccgcgcaca 5760
tttccccgaa aagtgccacc tgacgtctaa gaaaccatta ttatcatgac attaacctat 5820
aaaaataggc gtatcacgag gccctttcgt ctcgcgcgtt tcggtgatga cggtgaaaac 5880
ctctgacaca tgcagctccc ggagacggtc acagcttgtc tgtaagcgga tgccgggagc 5940
agacaagccc gtcagggcgc gtcagcgggt gttggcgggt gtcggggctg gcttaactat 6000
gcggcatcag agcagattgt actgagagtg caccatatgc ggtgtgaaat accgcacaga 6060
tgcgtaagga gaaaataccg catcaggaaa ttgtaaacgt taatattttg ttaaaattcg 6120
cgttaaattt ttgttaaatc agctcatttt ttaaccaata ggccgaaatc ggcaaaatcc 6180
cttataaatc aaaagaatag accgagatag ggttgagtgt tgttccagtt tggaacaaga 6240
gtccactatt aaagaacgtg gactccaacg tcaaagggcg aaaaaccgtc tatcagggcg 6300
atggcccact acgtgaacca tcaccctaat caagtttttt ggggtcgagg tgccgtaaag 6360
cactaaatcg gaaccctaaa gggagccccc gatttagagc ttgacgggga aagccggcga 6420
acgtggcgag aaaggaaggg aagaaagcga aaggagcggg cgctagggcg ctggcaagtg 6480
tagcggtcac gctgcgcgta accaccacac ccgccgcgct taatgcgccg ctacagggcg 6540
cgtcgcgcca ttcgccattc aggctacgca actgttggga agggcgatcg gtgcgggcct 6600
cttcgctatt acgccagctg gctgca 6626

[0119]

<210> 457
<211> 22
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 457
gagaggttat gtggctttac ca 22

<210> 458
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 458
ctgcgtagtg ccaaaacaaa 20

<210> 459
<211> 3252
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成多核苷酸

<400> 459

[0120]

atgaaaagga actacattct ggggctggac atcgggatta caagcgtggg gtatgggatt 60
attgactatg aaacaaggga cgtgatcgac gcaggcgtca gactgttcaa ggaggccaac 120
gtggaaaaca atgagggacg gagaagcaag aggggagcca ggcgcctgaa acgacggaga 180
aggcacagaa tccagagggt gaagaaactg ctgttcgatt acaacctgct gaccgaccat 240
tctgagctga gtggaattaa tccttatgaa gccagggtga aaggcctgag tcagaagctg 300
tcagaggaag agttttccgc agctctgctg cacctggcta agcgccgagg agtgcataac 360
gtcaatgagg tggaagagga caccggcaac gagctgtcta caaaggaaca gatctcacgc 420
aatagcaaag ctctggaaga gaagtatgtc gcagagctgc agctggaacg gctgaagaaa 480
gatggcgagg tgagagggtc aattaatagg ttcaagacaa gcgactacgt caaagaagcc 540
aagcagctgc tgaaagtgca gaaggcttac caccagctgg atcagagctt catcgatact 600
tatatcgacc tgctggagac tcggagaacc tactatgagg gaccaggaga agggagcccc 660
ttcggatgga aagacatcaa ggaatggtac gagatgctga tgggacattg cacctatttt 720
ccagaagagc tgagaagcgt caagtacgct tataacgcag atctgtacaa cgccctgaat 780
gacctgaaca acctggtcat caccagggat gaaaacgaga aactggaata ctatgagaag 840
ttccagatca tcgaaaacgt gtttaagcag aagaaaaagc ctacactgaa acagattgct 900
aaggagatcc tggtcaacga agaggacatc aagggctacc gggtgacaag cactggaaaa 960
ccagagttca ccaatctgaa agtgtatcac gatattaagg acatcacagc acggaaagaa 1020
atcattgaga acgccgaact gctggatcag attgctaaga tcctgactat ctaccagagc 1080
tccgaggaca tccaggaaga gctgactaac ctgaacagcg agctgaccca ggaagagatc 1140
gaacagatta gtaatctgaa ggggtacacc ggaacacaca acctgtccct gaaagctatc 1200
aatctgattc tggatgagct gtggcataca aacgacaatc agattgcaat ctttaaccgg 1260
ctgaagctgg tcccaaaaaa ggtggacctg agtcagcaga aagagatccc aaccacactg 1320
gtggacgatt tcattctgtc acccgtggtc aagcggagct tcatccagag catcaaagtg 1380
atcaacgcca tcatcaagaa gtacggcctg cccaatgata tcattatcga gctggctagg 1440
gagaagaaca gcaaggacgc acagaagatg atcaatgaga tgcagaaacg aaaccggcag 1500
accaatgaac gcattgaaga gattatccga actaccggga aagagaacgc aaagtacctg 1560
attgaaaaaa tcaagctgca cgatatgcag gagggaaagt gtctgtattc tctggaggcc 1620
atccccctgg aggacctgct gaacaatcca ttcaactacg aggtcgatca tattatcccc 1680
agaagcgtgt ccttcgacaa ttcctttaac aacaaggtgc tggtcaagca ggaagagaac 1740
tctaaaaagg gcaataggac tcctttccag tacctgtcta gttcagattc caagatctct 1800
tacgaaacct ttaaaaagca cattctgaat ctggccaaag gaaagggccg catcagcaag 1860
accaaaaagg agtacctgct ggaagagcgg gacatcaaca gattctccgt ccagaaggat 1920
tttattaacc ggaatctggt ggacacaaga tacgctactc gcggcctgat gaatctgctg 1980
cgatcctatt tccgggtgaa caatctggat gtgaaagtca agtccatcaa cggcgggttc 2040
acatcttttc tgaggcgcaa atggaagttt aaaaaggagc gcaacaaagg gtacaagcac 2100
catgccgaag atgctctgat tatcgcaaat gccgacttca tctttaagga gtggaaaaag 2160
ctggacaaag ccaagaaagt gatggagaac cagatgttcg aagagaagca ggccgaatct 2220
atgcccgaaa tcgagacaga acaggagtac aaggagattt tcatcactcc tcaccagatc 2280
aagcatatca aggatttcaa ggactacaag tactctcacc gggtggataa aaagcccaac 2340
agagagctga tcaatgacac cctgtatagt acaagaaaag acgataaggg gaataccctg 2400
attgtgaaca atctgaacgg actgtacgac aaagataatg acaagctgaa aaagctgatc 2460
aacaaaagtc ccgagaagct gctgatgtac caccatgatc ctcagacata tcagaaactg 2520
aagctgatta tggagcagta cggcgacgag aagaacccac tgtataagta ctatgaagag 2580
actgggaact acctgaccaa gtatagcaaa aaggataatg gccccgtgat caagaagatc 2640
aagtactatg ggaacaagct gaatgcccat ctggacatca cagacgatta ccctaacagt 2700

cgcaacaagg tggtcaagct gtcactgaag ccatacagat tcgatgtcta tctggacaac 2760
ggcgtgtata aatttgtgac tgtcaagaat ctggatgtca tcaaaaagga gaactactat 2820
gaagtgaata gcaagtgcta cgaagaggct aaaaagctga aaaagattag caaccaggca 2880
gagttcatcg cctcctttta caacaacgac ctgattaaga tcaatggcga actgtatagg 2940
gtcatcgggg tgaacaatga tctgctgaac cgcattgaag tgaatatgat tgacatcact 3000
taccgagagt atctggaaaa catgaatgat aagcgccccc ctcgaattat caaaacaatt 3060
gcctctaaga ctcagagtat caaaaagtac tcaaccgaca ttctgggaaa cctgtatgag 3120
gtgaagagca aaaagcaccc tcagattatc aaaaagggca gcggaggcaa gcgtcctgct 3180
gctactaaga aagctggtca agctaagaaa aagaaaggat cctacccata cgatgttcca 3240
gattacgctt aa 3252

<210> 460
<211> 83
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 460
gttttagtac tctggaaaca gaatctacta aaacaaggca aaatgccgtg tttatctcgt 60
caacttgttg gcgagatttt ttt 83

<210> 461
<211> 3333
<212> DNA
[0121]
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 461
atggtgccta agaagaagag aaaggtggct gccttcaaac ctaattcaat caactacatc 60
ctcggcctcg atatcggcat cgcatccgtc ggctgggcga tggtagaaat tgacgaagaa 120
gaaaacccca tccgcctgat tgatttgggc gtgcgcgtat ttgagcgtgc cgaagtaccg 180
aaaacaggcg actcccttgc catggcaagg cgtttggcgc gcagtgttcg ccgcctgacc 240
cgccgtcgcg cccaccgcct gcttcggacc cgccgcctat tgaaacgcga aggcgtatta 300
caagccgcca attttgacga aaacggcttg attaaatcct taccgaatac accatggcaa 360
cttcgcgcag ccgcattaga ccgcaaactg acgcctttag agtggtcggc agtcttgttg 420
catttaatca aacatcgcgg ctatttatcg caacggaaaa acgagggcga aactgccgat 480
aaggagcttg gcgctttgct taaaggcgta gccggcaatg cccatgcctt acagacaggc 540
gatttccgca caccggccga attggcttta aataaatttg agaaagaaag cggccatatc 600
cgcaatcagc gcagcgatta ttcgcatacg ttcagccgca aagatttaca ggcggagctg 660
attttgctgt ttgaaaaaca aaaagaattt ggcaatccgc atgtttcagg cggccttaaa 720
gaaggtattg aaaccctact gatgacgcaa cgccctgccc tgtccggcga tgccgttcaa 780
aaaatgttgg ggcattgcac cttcgaaccg gcagagccga aagccgctaa aaacacctac 840
acagccgaac gtttcatctg gctgaccaag ctgaacaacc tgcgtatttt agagcaaggc 900
agcgagcggc cattgaccga taccgaacgc gccacgctta tggacgagcc atacagaaaa 960
tccaaactga cttacgcaca agcccgtaag ctgctgggtt tagaagatac cgccttttc 1020
aaaggcttgc gctatggtaa agacaatgcc gaagcctcaa cattgatgga aatgaaggcc 1080

[0122]

taccatgcca tcagccgtgc actggaaaaa gaaggattga aagacaaaaa atccccatta 1140
aacctttctc ccgaattaca agacgaaatc ggcacggcat tctccctgtt caaaaccgat 1200
gaagacatta caggccgtct gaaagaccgt atacagcccg aaatcttaga agcgctgttg 1260
aaacacatca gcttcgataa gttcgtccaa atttccttga aagcattgcg ccgaattgtg 1320
cctctaatgg aacaaggcaa acgttacgat gaagcctgcg ccgaaatcta cggagaccat 1380
tacggcaaga agaatacgga agaaaagatt tatctgccgc cgattcccgc cgacgaaatc 1440
cgcaaccccg tcgtcttgcg cgccttatct caagcacgta aggtcattaa cggcgtggta 1500
cgccgttacg gctccccagc tcgtatccat attgaaactg caagggaagt aggtaaatcg 1560
tttaaagacc gcaaagaaat tgagaaacgc caagaagaaa accgcaaaga ccgggaaaaa 1620
gccgccgcca aattccgaga gtatttcccc aattttgtcg gagaacccaa atccaaagat 1680
attctgaaac tgcgcctgta cgagcaacaa cacggcaaat gcctgtattc gggcaaagaa 1740
atcaacttag gccgtctgaa cgaaaaaggc tatgtcgaaa tcgaccatgc cctgccgttc 1800
tcgcgcacat gggacgacag tttcaacaat aaagtactgg tattgggcag cgaaaaccaa 1860
aacaaaggca atcaaacccc ttacgaatac ttcaacggca aagacaacag ccgcgaatgg 1920
caggaattta aagcgcgtgt cgaaaccagc cgtttcccgc gcagtaaaaa acaacggatt 1980
ctgctgcaaa aattcgatga agacggcttt aaagaacgca atctgaacga cacgcgctac 2040
gtcaaccgtt tcctgtgtca atttgttgcc gaccgtatgc ggctgacagg taaaggcaag 2100
aaacgtgtct ttgcatccaa cggacaaatt accaatctgt tgcgcggctt ttggggattg 2160
cgcaaagtgc gtgcggaaaa cgaccgccat cacgccttgg acgccgtcgt cgttgcctgc 2220
tcgaccgttg ccatgcagca gaaaattacc cgttttgtac gctataaaga gatgaacgcg 2280
tttgacggta aaaccataga caaagaaaca ggagaagtgc tgcatcaaaa aacacacttc 2340
ccacaacctt gggaattttt cgcacaagaa gtcatgattc gcgtcttcgg caaaccggac 2400
ggcaaaaccg aattcgaaga agccgatacc ctagaaaaac tgcgcacgtt gcttgccgaa 2460
aaattatcat ctcgccccga agccgtacac gaatacgtta cgccactgtt tgtttcacgc 2520
gcgcccaatc ggaagatgag cgggcaaggg catatggaga ccgtcaaatc cgccaaacga 2580
ctggacgaag gcgtcagcgt gttgcgcgta ccgctgacac agttaaaact gaaagacttg 2640
gaaaaaatgg tcaatcggga gcgcgaacct aagctatacg aagcactgaa agcacggctg 2700
gaagcacata aagacgatcc tgccaaagcc tttgccgagc cgttttacaa atacgataaa 2760
gcaggcaacc gcacccaaca ggtaaaagcc gtacgcgtag agcaagtaca gaaaaccggc 2820
gtatgggtgc gcaaccataa cggtattgcc gacaacgcaa ccatggtgcg cgtagatgtg 2880
tttgagaaag gcgacaagta ttatctggta ccgatttaca gttggcaggt agcgaaaggg 2940
attttgccgg atagggctgt tgtacaagga aaagatgaag aagattggca acttattgat 3000
gatagtttca actttaaatt ctcattacac cctaatgatt tagtcgaggt tataacaaaa 3060
aaagctagaa tgtttggtta ctttgccagc tgccatcgag gcacaggtaa tatcaatata 3120
cgcattcatg atcttgatca taaaattggc aaaaatggaa tactggaagg tatcggcgtc 3180
aaaaccgccc tttcattcca aaaataccaa attgacgaac tgggcaaaga aatcagacca 3240
tgccgtctga aaaaacgccc gcctgtccgt tacccatacg atgttccaga ttacgctgca 3300
gctccagcag cgaagaaaaa gaagctggat taa 3333

<210> 462
<211> 120
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 462

gttgtagctc cctttctcat ttcggaaacg aaatgagaac cgttgctaca ataaggccgt 60
ctgaaaagat gtgccgcaac gctctgcccc ttaaagcttc tgctttaagg ggctttttt 120
<210> 463
<211> 144
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 463
gttgtagctc cctttctcat ttcgcagtgc tacaatgaaa attgtcgcac tgcgaaatga 60
gaaccgttgc tacaataagg ccgtctgaaa agatgtgccg caacgctctg ccccttaaag 120
cttctgcttt aaggggcttt tttt 144
<210> 464
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 464
tcgggtttat tacagggaca gcag 24
[0123]
<210> 465
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 465
tctaaggccg agtcttatga gcag 24
<210> 466
<211> 4564
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 466
atggactaca aagaccatga cggtgattat aaagatcatg acatcgatta caaggatgac 60
gatgacaaga tggcccccaa gaagaagagg aaggtgggcc gcggaatgga caagaagtac 120
tccattgggc tcgccatcgg cacaaacagc gtcggctggg ccgtcattac ggacgagtac 180
aaggtgccga gcaaaaaatt caaagttctg ggcaataccg atcgccacag cataaagaag 240
aacctcattg gcgccctcct gttcgactcc ggggaaaccg ccgaagccac gcggctcaaa 300
agaacagcac ggcgcagata tacccgcaga aagaatcgga tctgctacct gcaggagatc 360
tttagtaatg agatggctaa ggtggatgac tctttcttcc ataggctgga ggagtccttt 420

[0124]

ttggtggagg aggataaaaa gcacgagcgc cacccaatct ttggcaatat cgtggacgag 480
gtggcgtacc atgaaaagta cccaaccata tatcatctga ggaagaagct tgtagacagt 540
actgataagg ctgacttgcg gttgatctat ctcgcgctgg cgcatatgat caaatttcgg 600
ggacacttcc tcatcgaggg ggacctgaac ccagacaaca gcgatgtcga caaactcttt 660
atccaactgg ttcagactta caatcagctt ttcgaagaga acccgatcaa cgcatccgga 720
gttgacgcca aagcaatcct gagcgctagg ctgtccaaat cccggcggct cgaaaacctc 780
atcgcacagc tccctgggga gaagaagaac ggcctgtttg gtaatcttat cgccctgtca 840
ctcgggctga cccccaactt taaatctaac ttcgacctgg ccgaagatgc caagcttcaa 900
ctgagcaaag acacctacga tgatgatctc gacaatctgc tggcccagat cggcgaccag 960
tacgcagacc tttttttggc ggcaaagaac ctgtcagacg ccattctgct gagtgatatt 1020
ctgcgagtga acacggagat caccaaagct ccgctgagcg ctagtatgat caagcgctat 1080
gatgagcacc accaagactt gactttgctg aaggcccttg tcagacagca actgcctgag 1140
aagtacaagg aaattttctt cgatcagtct aaaaatggct acgccggata cattgacggc 1200
ggagcaagcc aggaggaatt ttacaaattt attaagccca tcttggaaaa aatggacggc 1260
accgaggagc tgctggtaaa gcttaacaga gaagatctgt tgcgcaaaca gcgcactttc 1320
gacaatggaa gcatccccca ccagattcac ctgggcgaac tgcacgctat cctcaggcgg 1380
caagaggatt tctacccctt tttgaaagat aacagggaaa agattgagaa aatcctcaca 1440
tttcggatac cctactatgt aggccccctc gcccggggaa attccagatt cgcgtggatg 1500
actcgcaaat cagaagagac catcactccc tggaacttcg aggaagtcgt ggataagggg 1560
gcctctgccc agtccttcat cgaaaggatg actaactttg ataaaaatct gcctaacgaa 1620
aaggtgcttc ctaaacactc tctgctgtac gagtacttca cagtttataa cgagctcacc 1680
aaggtcaaat acgtcacaga agggatgaga aagccagcat tcctgtctgg agagcagaag 1740
aaagctatcg tggacctcct cttcaagacg aaccggaaag ttaccgtgaa acagctcaaa 1800
gaagactatt tcaaaaagat tgaatgtttc gactctgttg aaatcagcgg agtggaggat 1860
cgcttcaacg catccctggg aacgtatcac gatctcctga aaatcattaa agacaaggac 1920
ttcctggaca atgaggagaa cgaggacatt cttgaggaca ttgtcctcac ccttacgttg 1980
tttgaagata gggagatgat tgaagaacgc ttgaaaactt acgctcatct cttcgacgac 2040
aaagtcatga aacagctcaa gaggcgccga tatacaggat gggggcggct gtcaagaaaa 2100
ctgatcaatg ggatccgaga caagcagagt ggaaagacaa tcctggattt tcttaagtcc 2160
gatggatttg ccaaccggaa cttcatgcag ttgatccatg atgactctct caccttaag 2220
gaggacatcc agaaagcaca agtttctggc caggggaca gtcttcacga gcacatcgct 2280
aatcttgcag gtagcccagc tatcaaaaag ggaatactgc agaccgttaa ggtcgtggat 2340
gaactcgtca aagtaatggg aaggcataag cccgagaata tcgttatcga gatggcccga 2400
gagaaccaaa ctacccagaa gggacagaag aacagtaggg aaaggatgaa gaggattgaa 2460
gagggtataa aagaactggg gtcccaaatc cttaaggaac acccagttga aaacacccag 2520
cttcagaatg agaagctcta cctgtactac ctgcagaacg gcagggacat gtacgtggat 2580
caggaactgg acatcaatcg gctctccgac tacgacgtgg atgccatcgt gccccagtct 2640
tttctcaaag atgattctat tgataataaa gtgttgacaa gatccgataa aaatagaggg 2700
aagagtgata acgtcccctc agaagaagtt gtcaagaaaa tgaaaaatta ttggcggcag 2760
ctgctgaacg ccaaactgat cacacaacgg aagttcgata atctgactaa ggctgaacga 2820
ggtggcctgt ctgagttgga taaagccggc ttcatcaaaa ggcagcttgt tgagacacgc 2880
cagatcacca agcacgtggc ccaaattctc gattcacgca tgaacaccaa gtacgatgaa 2940
aatgacaaac tgattcgaga ggtgaaagtt attactctga agtctaagct ggtctcagat 3000
ttcagaaagg actttcagtt ttataaggtg agagagatca acaattacca ccatgcgcat 3060
gatgcctacc tgaatgcagt ggtaggcact gcacttatca aaaaatatcc caagcttgaa 3120
tctgaatttg tttacggaga ctataaagtg tacgatgtta ggaaaatgat cgcaaagtct 3180

gagcaggaaa taggcaaggc caccgctaag tacttctttt acagcaatat tatgaatttt 3240
ttcaagaccg agattacact ggccaatgga gagattcgga agcgaccact tatcgaaaca 3300
aacggagaaa caggagaaat cgtgtgggac aagggtaggg atttcgcgac agtccggaag 3360
gtcctgtcca tgccgcaggt gaacatcgtt aaaaagaccg aagtacagac cggaggcttc 3420
tccaaggaaa gtatcctccc gaaaaggaac agcgacaagc tgatcgcacg caaaaaagat 3480
tgggacccca agaaatacgg cggattcgat tctcctacag tcgcttacag tgtactggtt 3540
gtggccaaag tggagaaagg gaagtctaaa aaactcaaaa gcgtcaagga actgctgggc 3600
atcacaatca tggagcgatc aagcttcgaa aaaaacccca tcgactttct cgaggcgaaa 3660
ggatataaag aggtcaaaaa agacctcatc attaagcttc ccaagtactc tctctttgag 3720
cttgaaaacg gccggaaacg aatgctcgct agtgcgggcg agctgcagaa aggtaacgag 3780
ctggcactgc cctctaaata cgttaatttc ttgtatctgg ccagccacta tgaaaagctc 3840
aaagggtctc ccgaagataa tgagcagaag cagctgttcg tggaacaaca caaacactac 3900
cttgatgaga tcatcgagca aataagcgaa ttctccaaaa gagtgatcct cgccgacgct 3960
aacctcgata aggtgctttc tgcttacaat aagcacaggg ataagcccat cagggagcag 4020
gcagaaaaca ttatccactt gtttactctg accaacttgg gcgcgcctgc agccttcaag 4080
tacttcgaca ccaccataga cagaaagcgg tacacctcta caaaggaggt cctggacgcc 4140
acactgattc atcagtcaat tacggggctc tatgaaacaa gaatcgacct ctctcagctc 4200
ggtggagaca gcagggctga ccccaagaag aagaggaagg tggctagcga tgctaagtca 4260
ctgactgcct ggtcccggac actggtgacc ttcaaggatg tgtttgtgga cttcaccagg 4320
gaggagtgga agctgctgga cactgctcag cagatcctgt acagaaatgt gatgctggag 4380
aactataaga acctggtttc cttgggttat cagcttacta agccagatgt gatcctccgg 4440
ttggagaagg gagaagagcc ctggctggtg gagagagaaa ttcaccaaga gacccatcct 4500
gattcagaga ctgcatttga aatcaaatca tcagttccga aaaagaaacg caaagttgct 4560

[0125]

agcg 4564

<210> 467
<211> 388
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 467

taagcttcag tttttcctta gttcctgtta catttctgtg tgtctccatt agtgacctcc 60
catagtccaa gcatgagcag ttctggccag gcccctgtcg gggtcagtgc cccacccccg 120
ccttctggtt ctgtgtaacc ttctaagcaa accttctggc tcaagcacag caatgctgag 180
tcatgatgag tcatgctgag gcttagggtg tgtgcccaga tgttctcagc ctagagtgat 240
gactcctatc tgggtcccca gcaggatgct tacagggcag atggcaaaaa aaaggagaag 300
ctgaccacct gactaaaact ccacctcaaa cggcatcata aagaaaatgg atgcctgaga 360
cagaatgtga catattctag aatatatt 388

<210> 468
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸

<400> 468
cagccgctcg ctgcagcag 19
<210> 469
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 469
ctgctgcagc gagcggctg 19
<210> 470
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 470
gctgggtgtc ccattgaaa 19
<210> 471
<211> 19
<212> DNA
[0126]
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 471
tttcaatggg acacccagc 19
<210> 472
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 472
gtttattcag ccgggagtc 19
<210> 473
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 473

gactcccggc tgaataaac 19

<210> 474

<211> 19

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 474

tggagagttt gcaaggagc 19

<210> 475

<211> 19

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 475

gctccttgca aactctcca 19

<210> 476

<211> 19

<212> DNA

<213> 人工序列

[0127]

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 476

ccctccagac tttccacct 19

<210> 477

<211> 19

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 477

aggtggaaag tctggaggg 19

<210> 478

<211> 19

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 478

aattttcttc caagttctc 19

<210> 479
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 479
gagaacttgg aagaaaatt 19
<210> 480
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 480
ctgcggagag aagaaaggg 19
<210> 481
<211> 19
<212> DNA
<213> 人工序列
<220>
[0128]
<223> 人工序列的描述：合成
寡核苷酸
<400> 481
ccctttcttc tctccgcag 19
<210> 482
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 482
agagccaccc cctggctcc 19
<210> 483
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 483
ggagccaggg ggtggctct 19
<210> 484

<210> 19
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 484
cgaagccaac cgcggcggg 19
<210> 485
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 485
cccgccgcgg ttggcttcg 19
<210> 486
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 486
agagggaaga cgatcgccc 19
<210> 487
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 487
gggcgatcgt cttccctct 19
<210> 488
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 488
ccccttaac tttcctccg 19
<210> 489
<211> 19

[0129]

<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 489
cggaggaaag ttaaagggg 19
<210> 490
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 490
gcagccccgc ttccttcaa 19
<210> 491
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
[0130]
<400> 491
ttgaaggaag cggggctgc 19
<210> 492
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 492
caccgcagcc gctcgctgca gcag 24
<210> 493
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 493
aaacctgctg cagcgagcgg ctgc 24
<210> 494
<211> 24
<212> DNA

<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 494
caccggctgg gtgtcccatt gaaa 24
<210> 495
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 495
aaactttcaa tgggacaccc agcc 24
<210> 496
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 496
[0131]
caccggttta ttcagccggg agtc 24
<210> 497
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 497
aaacgactcc cggctgaata aacc 24
<210> 498
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 498
caccgtggag agtttgcaag gagc 24
<210> 499
<211> 24
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 499
aaacgctcct tgcaaactct ccac 24
<210> 500
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 500
caccgccctc cagactttcc acct 24
<210> 501
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 501
[0132] aaacaggtgg aaagtctgga gggc 24
<210> 502
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 502
caccgaattt tcttccaagt tctc 24
<210> 503
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 503
aaacgagaac ttggaagaaa attc 24
<210> 504
<211> 24
<212> DNA
<213> 人工序列
<220>

<223> 人工序列的描述：合成
寡核苷酸
<400> 504
caccgctgcg gagagaagaa aggg 24
<210> 505
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 505
aaaccccttt cttctctccg cagc 24
<210> 506
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 506
caccgagagc cacccctgg ctcc 24
[0133]
<210> 507
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 507
aaacggagcc agggggtggc tctc 24
<210> 508
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 508
caccgcgaag ccaaccgcgg cggg 24
<210> 509
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成

寡核苷酸

<400> 509

aaaccccgcc gcggttggct tcgc 24

<210> 510

<211> 24

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 510

caccgagagg gaagacgatc gccc 24

<210> 511

<211> 24

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 511

aaacgggcga tcgtcttccc tctc 24

<210> 512

<211> 24

[0134]

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 512

caccgcccct ttaactttcc tccg 24

<210> 513

<211> 24

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 513

aaaccggagg aaagttaaag gggc 24

<210> 514

<211> 24

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 514
caccggcagc cccgcttcct tcaa 24
<210> 515
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 515
aaacttgaag gaagcggggc tgcc 24
<210> 516
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 516
cgagagcgag aggagggag 19
<210> 517
<211> 19
<212> DNA
[0135]
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 517
ctccctcctc tcgctctcg 19
<210> 518
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 518
gagagagctt gagagcgcg 19
<210> 519
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 519

cgcgctctca agctctctc 19

<210> 520
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 520

ggtggagggg gcggggccc 19

<210> 521
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 521

gggccccgcc ccctccacc 19

<210> 522
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 522

[0136]

ggtatccacg taaatcaaa 19

<210> 523
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 523

tttgatttac gtggatacc 19

<210> 524
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 524

ccaatcactg gctccggtc 19

<210> 525
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 525
gaccggagcc agtgattgg 19
<210> 526
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 526
ggcgcccgag ggaagaaga 19
<210> 527
<211> 19
<212> DNA
<213> 人工序列
<220>
[0137]
<223> 人工序列的描述：合成
寡核苷酸
<400> 527
tcttcttccc tcgggcgcc 19
<210> 528
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 528
gggtgggggt accagagga 19
<210> 529
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 529
tcctctggta cccccaccc 19
<210> 530

[0138]

<210> 19
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 530
ccggggacag aagagaggg 19
<210> 531
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 531
ccctctcttc tgtccccgg 19
<210> 532
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 532
gagagagagt gggagaagc 19
<210> 533
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 533
gcttctccca ctctctctc 19
<210> 534
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 534
aaagtaactg tcaaatgcg 19
<210> 535
<211> 19

<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 535
cgcatttgac agttacttt 19
<210> 536
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 536
ttaaccagag cgcccagtc 19
<210> 537
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
[0139]
<400> 537
gactgggcgc tctggttaa 19
<210> 538
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 538
cgtcggagct gcccgctag 19
<210> 539
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 539
ctagcgggca gctccgacg 19
<210> 540
<211> 24
<212> DNA

<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 540
caccgcgaga gcgagaggag ggag 24
<210> 541
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 541
aaacctccct cctctcgctc tcgc 24
<210> 542
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 542
[0140]
caccggagag agcttgagag cgcg 24
<210> 543
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 543
aaaccgcgct ctcaagctct ctcc 24
<210> 544
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 544
caccgggtgg agggggcggg gccc 24
<210> 545
<211> 24
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 545
aaacgggccc cgccccctcc accc 24
<210> 546
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 546
caccgggtat ccacgtaaat caaa 24
<210> 547
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 547
[0141] aaactttgat ttacgtggat accc 24
<210> 548
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 548
caccgccaat cactggctcc ggtc 24
<210> 549
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 549
aaacgaccgg agccagtgat tggc 24
<210> 550
<211> 24
<212> DNA
<213> 人工序列
<220>

<223> 人工序列的描述：合成
寡核苷酸
<400> 550
caccgggcgc ccgagggaag aaga 24
<210> 551
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 551
aaactcttct tccctcgggc gccc 24
<210> 552
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 552
caccggggtg ggggtaccag agga 24

[0142]

<210> 553
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 553
aaactcctct ggtacccccа cccc 24
<210> 554
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 554
caccgccggg gacagaagag aggg 24
<210> 555
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成

寡核苷酸

<400> 555

aaaccctct cttctgtccc cggc 24

<210> 556

<211> 24

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 556

caccggagag agagtgggag aagc 24

<210> 557

<211> 24

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 557

aaacgcttct cccactctct ctcc 24

<210> 558

<211> 24

[0143]

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 558

caccgaaagt aactgtcaaa tgcg 24

<210> 559

<211> 24

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 559

aaaccgcatt tgacagttac tttc 24

<210> 560

<211> 24

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 560
caccgttaac cagagcgccc agtc 24
<210> 561
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 561
aaacgactgg gcgctctggt taac 24
<210> 562
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 562
caccgcgtcg gagctgcccg ctag 24
<210> 563
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 563
aaacctagcg ggcagctccg acgc 24
<210> 564
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 564
gagacacaca gaaatgtaac 20
<210> 565
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 565

[0144]

[0145]

ggtggggcac tgaccccgac 20

<210> 566

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 566

ctagagtgat gactcctatc 20

<210> 567

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 567

gactaaaact ccacctcaaa 20

<210> 568

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 568

aatatgtcac attctgtctc 20

<210> 569

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 569

ggactatggg aggtcactaa 20

<210> 570

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 570

gctcatgctt ggactatggg 20

<210> 571
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 571
gttctggcca ggcccctgtc 20
<210> 572
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 572
agtgccccac ccccgccttc 20
<210> 573
<211> 20
<212> DNA
<213> 人工序列
<220>
[0146]
<223> 人工序列的描述：合成
寡核苷酸
<400> 573
gtggggcact gaccccgaca 20
<210> 574
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 574
aaccttctaa gcaaaccttc 20
<210> 575
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 575
gttacacaga accagaaggc 20
<210> 576

[0147]

<210> 20
<211> DNA
<212> 人工序列
<213>
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 576
gaaggttaca cagaaccaga 20
<210> 577
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 577
agtcatgatg agtcatgctg 20
<210> 578
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 578
gatgagtcat gctgaggctt 20
<210> 579
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 579
actctaggct gagaacatct 20
<210> 580
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 580
gtccccagca ggatgcttac 20
<210> 581
<211> 20

<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 581
gccctgtaag catcctgctg 20
<210> 582
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 582
cagggcagat ggcaaaaaaa 20
<210> 583
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
[0148]
<400> 583
gaggtggagt tttagtcagg 20
<210> 584
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 584
aaacggcatc ataaagaaaa 20
<210> 585
<211> 105
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 585
gagacacaca gaaatgtaac gttttagagc tagaaatagc aagttaaaat aaggctagtc 60
cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt tttcc 105
<210> 586
<211> 105

<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 586
ggtggggcac tgaccccgac gttttagagc tagaaatagc aagttaaaat aaggctagtc 60
cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt tttcc 105
<210> 587
<211> 105
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 587
ctagagtgat gactcctatc gttttagagc tagaaatagc aagttaaaat aaggctagtc 60
cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt tttcc 105
<210> 588
<211> 105
<212> DNA
<213> 人工序列
<220>
[0149]
<223> 人工序列的描述：合成
多核苷酸
<400> 588
gactaaaact ccacctcaaa gttttagagc tagaaatagc aagttaaaat aaggctagtc 60
cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt tttcc 105
<210> 589
<211> 105
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 589
aatatgtcac attctgtctc gttttagagc tagaaatagc aagttaaaat aaggctagtc 60
cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt tttcc 105
<210> 590
<211> 105
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 590

ggactatggg aggtcactaa gttttagagc tagaaatagc aagttaaaat aaggctagtc 60
cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt tttcc 105
<210> 591
<211> 105
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 591
gctcatgctt ggactatggg gttttagagc tagaaatagc aagttaaaat aaggctagtc 60
cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt tttcc 105
<210> 592
<211> 105
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 592
gttctggcca ggcccctgtc gttttagagc tagaaatagc aagttaaaat aaggctagtc 60
cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt tttcc 105
[0150]
<210> 593
<211> 105
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 593
agtgccccac ccccgccttc gttttagagc tagaaatagc aagttaaaat aaggctagtc 60
cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt tttcc 105
<210> 594
<211> 105
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 594
gtggggcact gaccccgaca gttttagagc tagaaatagc aagttaaaat aaggctagtc 60
cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt tttcc 105
<210> 595
<211> 105
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 595
aaccttctaa gcaaaccttc gttttagagc tagaaatagc aagttaaaat aaggctagtc 60
cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt tttcc 105
<210> 596
<211> 105
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 596
gttacacaga accagaaggc gttttagagc tagaaatagc aagttaaaat aaggctagtc 60
cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt tttcc 105
<210> 597
<211> 105
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
[0151]
<400> 597
gaaggttaca cagaaccaga gttttagagc tagaaatagc aagttaaaat aaggctagtc 60
cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt tttcc 105
<210> 598
<211> 105
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 598
agtcatgatg agtcatgctg gttttagagc tagaaatagc aagttaaaat aaggctagtc 60
cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt tttcc 105
<210> 599
<211> 105
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 599
gatgagtcat gctgaggctt gttttagagc tagaaatagc aagttaaaat aaggctagtc 60
cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt tttcc 105

<210> 600
<211> 105
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 600
actctaggct gagaacatct gttttagagc tagaaatagc aagttaaaat aaggctagtc 60
cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt tttcc 105
<210> 601
<211> 105
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 601
gtccccagca ggatgcttac gttttagagc tagaaatagc aagttaaaat aaggctagtc 60
cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt tttcc 105
<210> 602
<211> 105
<212> DNA
[0152]
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 602
gccctgtaag catcctgctg gttttagagc tagaaatagc aagttaaaat aaggctagtc 60
cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt tttcc 105
<210> 603
<211> 105
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 603
cagggcagat ggcaaaaaaa gttttagagc tagaaatagc aagttaaaat aaggctagtc 60
cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt tttcc 105
<210> 604
<211> 105
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成

多核苷酸

<400> 604

gaggtggagt tttagtcagg gttttagagc tagaaatagc aagttaaaat aaggctagtc 60

cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt tttcc 105

<210> 605

<211> 105

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成

多核苷酸

<400> 605

aaacggcatc ataaagaaaa gttttagagc tagaaatagc aagttaaaat aaggctagtc 60

cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt tttcc 105

<210> 606

<211> 19

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成

寡核苷酸

<400> 606

[0153] tgtactctct gaggtgctc 19

<210> 607

<211> 19

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成

寡核苷酸

<400> 607

acgcagataa gaaccagtt 19

<210> 608

<211> 19

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成

寡核苷酸

<400> 608

catcaagtca gccatcagc 19

<210> 609

<211> 19

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成
寡核苷酸
<400> 609
gagtcaccct cctggaaac 19
<210> 610
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 610
cctgggctcc ggggcgttt 19
<210> 611
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 611
ggcccctgcg gccacccccg 19
[0154]
<210> 612
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 612
ctccctccct gcccggtag 19
<210> 613
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 613
aggtttggaa agggcgtgc 19
<210> 614
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成

寡核苷酸

<400> 614

actccactgc actccagtct 20

<210> 615

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 615

tctgtggggg acctgcactg 20

<210> 616

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 616

ggggcgccag ttgtgtctcc 20

[0155]

<210> 617

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 617

acaccattgc caccaccatt 20

<210> 618

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成引物

<400> 618

caatgacccc ttcattgacc 20

<210> 619

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成引物

[0156]

<400> 619
ttgattttgg agggatctcg 20
<210> 620
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 620
ggaatccatg gagggaagat 20
<210> 621
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 621
tgttctcgct caggtcagtg 20
<210> 622
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 622
ctctctgctc ctttgccaca 20
<210> 623
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 623
gtgctcttcg ggtttcagga 20
<210> 624
<211> 30
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成引物
<400> 624

cgaaagagaa agcgaaccag tatcgagaac 30

<210> 625
<211> 27
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 625

cgttgtgcat agtcgctgct tgatcgc 27

<210> 626
<211> 25
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 626

gagtttggct caaattgtta ctctt 25

<210> 627
<211> 25
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 627

[0157]

gggaaatggt ctaggagagt aaagt 25

<210> 628
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 628

gattggcttt gatttcccta ggg 23

<210> 629
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 629

gcctactcag actgttactc tgg 23

[0158]

<210> 630
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 630
gcagttgcct aagaactggt ggg 23
<210> 631
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 631
ggggctccac cctcacgagt ggg 23
<210> 632
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 632
gccttcttta tcccctatcg agg 23
<210> 633
<211> 48
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<220>
<221> CDS
<222> (1)..(45)
<400> 633
gaa gga cca ttt gac gtt cag ctc cta ctc aga ctg tta ctc tgg tga 48
Glu Gly Pro Phe Asp Val Gln Leu Leu Leu Arg Leu Leu Leu Trp
1 5 10 15
<210> 634
<211> 15
<212> PRT
<213> 人工序列
<220>

<223> 人工序列的描述：合成
肽
<400> 634
Glu Gly Pro Phe Asp Val Gln Leu Leu Leu Arg Leu Leu Leu Trp
1 5 10 15
<210> 635
<211> 46
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<220>
<221> CDS
<222> (1)..(42)
<400> 635
gaa gga cca ttt gac gtt cag ctc cta ctc aga ctg tct ctg gtga 46
Glu Gly Pro Phe Asp Val Gln Leu Leu Leu Arg Leu Ser Leu
1 5 10
<210> 636
<211> 14
<212> PRT
[0159] <213> 人工序列
<220>
<223> 人工序列的描述：合成
肽
<400> 636
Glu Gly Pro Phe Asp Val Gln Leu Leu Leu Arg Leu Ser Leu
1 5 10
<210> 637
<211> 20
<212> RNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 637
gccuacucag acuguuacuc 20
<210> 638
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 638

cccaccagtt cttaggcaa 19

<210> 639

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 639

cccaccagtt cttaggcaac 20

<210> 640

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 640

tttgatttcc agttcttagg 20

<210> 641

<211> 19

<212> DNA

<213> 人工序列

[0160]

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 641

attaccatag agttcttag 19

<210> 642

<211> 19

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 642

gtggataaag gcaacaatg 19

<210> 643

<211> 16

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 643

cctcgatagg ggataa 16

<210> 644
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 644
accaaaccca cttaggggat aaa 23
<210> 645
<211> 21
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 645
ggtatcttaa ggacctccaa g 21
<210> 646
<211> 47
<212> DNA
<213> 人工序列
<220>
[0161]
<223> 人工序列的描述：合成
寡核苷酸
<400> 646
taggaagtaa attaatttga agcttgccaa ttaatccaaa tcttacc 47
<210> 647
<211> 83
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 647
tttccttttt gcaaaaccca aaatatttta gctcctactc agactgttag agtaggagga 60
gaaggagggc tgtggctctt gtg 83
<210> 648
<211> 83
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<220>
<221> 修饰的碱基

<222> (20)..(20)
<223> a, c, t, g, 未知或其它
<220>
<221> 修饰的碱基
<222> (31)..(31)
<223> a, c, t, g, 未知或其它
<220>
<221> 修饰的碱基
<222> (69)..(69)
<223> a, c, t, g, 未知或其它
<400> 648
aaaaccccca aaccctcccn ctttttcctc ntactctaac agtgtgagta ggaggccaca 60
cttagctgng gctctagtaa gtc 83
<210> 649
<211> 94
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<220>
<221> 修饰的碱基
<222> (50)..(50)
[0162]
<223> a, c, t, g, 未知或其它
<400> 649
cttctttttt ccttttgca aaaacccaaa atattttagc tcctactcan accgttttcg 60
gggggtattt tgtatttgga tgaaacactt gatg 94
<210> 650
<211> 93
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 650
tcttcttttt tccttcccgc aaaaacccaa aatattttcg ctcctactca ggacctggtc 60
gtggggtatt ttgtatttgg atgaaacact tga 93
<210> 651
<211> 104
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<220>
<221> 修饰的碱基

<222> (21)..(22)
<223> a, c, t, g, 未知或其它
<400> 651
tcaaggaaga tggcatttct nntttggaga tggcagtttc cttagtaacc acaggttgtg 60
tcaccatggc cagtaagaag gatgccacac ttgtgtgtgt cgct 104
<210> 652
<211> 86
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<220>
<221> 修饰的碱基
<222> (35)..(36)
<223> a, c, t, g, 未知或其它
<220>
<221> 修饰的碱基
<222> (42)..(44)
<223> a, c, t, g, 未知或其它
<400> 652
ttggagatag cagtttcctt ggtaagtaca ggttnntcac cnnnacagtg tgagtaggag 60
[0163] gccacactta gctgtggctc tagtaa 86
<210> 653
<211> 71
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<220>
<221> 修饰的碱基
<222> (54)..(54)
<223> a, c, t, g, 未知或其它
<220>
<221> 修饰的碱基
<222> (61)..(63)
<223> a, c, t, g, 未知或其它
<220>
<221> 修饰的碱基
<222> (69)..(69)
<223> a, c, t, g, 未知或其它
<400> 653
tttctagttt ggagatggca gtttccttag taaccacagg ttgtgtcacc aggnctcgag 60
nnntggttnt g 71
<210> 654

<211> 87
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 654
cagttgccat agtcttcact aaccgcgtca ccagagctca ggacctggtc gtggggtatt 60
ttgtatttgg atgaaacact tgatggt 87
<210> 655
<211> 92
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<220>
<221> 修饰的碱基
<222> (52)..(52)
<223> a, c, t, g, 未知或其它
<220>
<221> 修饰的碱基
<222> (59)..(59)
[0164]
<223> a, c, t, g, 未知或其它
<220>
<221> 修饰的碱基
<222> (78)..(78)
<223> a, c, t, g, 未知或其它
<400> 655
tattgtatct tggtttttct tcactgctgg ccagtttact aacaatctga antaagccna 60
ttaatccaaa tcttaccnga cgaaaaatca aa 92
<210> 656
<211> 89
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<220>
<221> 修饰的碱基
<222> (6)..(6)
<223> a, c, t, g, 未知或其它
<220>
<221> 修饰的碱基
<222> (10)..(10)
<223> a, c, t, g, 未知或其它

<220>
<221> 修饰的碱基
<222> (19)..(19)
<223> a, c, t, g, 未知或其它
<220>
<221> 修饰的碱基
<222> (30)..(30)
<223> a, c, t, g, 未知或其它
<220>
<221> 修饰的碱基
<222> (57)..(57)
<223> a, c, t, g, 未知或其它
<400> 656
atggtnttta gtcattgcng gccagttttn taacaatgtg atcaatgcca attaatncaa 60
atcttaccag aggaaaaatc aagccacag 89
<210> 657
<211> 94
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<220>
[0165]
<221> 修饰的碱基
<222> (93)..(94)
<223> a, c, t, g, 未知或其它
<400> 657
aataaaaaga aaaaaatatg ataaactgct cccagtataa aatacagagc taagacaaga 60
acgtttcatt ggctgttaaa agtcggaaat cgnn 94
<210> 658
<211> 94
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<400> 658
ttctgattta aaattccctt gaataggaag taaattaatt tgaagcttgc caattaatcc 60
aaatcttacc agaggaaaaa tcaagccaca gata 94
<210> 659
<211> 90
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物

<400> 659
caaaacaaga tgggataaaa tcttctttct gatttaaaat tcccttgaat aggaagtaaa 60
ttactagagc tgttcaaggc ctgaactcta 90
<210> 660
<211> 96
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
引物
<220>
<221> 修饰的碱基
<222> (49)..(49)
<223> a, c, t, g, 未知或其它
<220>
<221> 修饰的碱基
<222> (51)..(52)
<223> a, c, t, g, 未知或其它
<220>
<221> 修饰的碱基
<222> (54)..(54)
<223> a, c, t, g, 未知或其它
<400> 660
[0166]
ctgatttaaa attcccttga ataggaagta aattactaga gctgttcang nncngaactc 60
taggctagca tttccagccc aaaacaaaca cagttc 96
<210> 661
<211> 26
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<220>
<221> 修饰的碱基
<222> (8)..(26)
<223> a, c, t, g, 未知或其它
<400> 661
ttgtttgnnn nnnnnnnnnn nnnnnn 26
<210> 662
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<220>

<220>
<221> 修饰的碱基
<222> (6)..(24)
<223> a, c, t, g, 未知或其它
<400> 662
caccgnnnnn nnnnnnnnnn nnnn 24
<210> 663
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<220>
<221> 修饰的碱基
<222> (6)..(24)
<223> a, c, t, g, 未知或其它
<400> 663
tcccannnnn nnnnnnnnnn nnnn 24
<210> 664
<211> 24
<212> DNA
<213> 人工序列
<220>
[0167]
<223> 人工序列的描述：合成
寡核苷酸
<220>
<221> 修饰的碱基
<222> (6)..(24)
<223> a, c, t, g, 未知或其它
<400> 664
cctcgnnnnn nnnnnnnnnn nnnn 24
<210> 665
<211> 26
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<220>
<221> 修饰的碱基
<222> (5)..(23)
<223> a, c, t, g, 未知或其它
<400> 665
aaacnnnnnn nnnnnnnnnn nnncaa 26
<210> 666
<211> 24

<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<220>
<221> 修饰的碱基
<222> (5)..(23)
<223> a, c, t, g, 未知或其它
<400> 666
aaacnnnnnn nnnnnnnnnn nnnc 24
<210> 667
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<220>
<221> 修饰的碱基
<222> (5)..(23)
<223> a, c, t, g, 未知或其它
<400> 667
[0168]
aaacnnnnnn nnnnnnnnnn nnnt 24
<210> 668
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<220>
<221> 修饰的碱基
<222> (5)..(23)
<223> a, c, t, g, 未知或其它
<400> 668
aaacnnnnnn nnnnnnnnnn nnnc 24
<210> 669
<211> 26
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 669
ttgtttgtct tcgagaagac ctgttt 26

[0169]

<210> 670
<211> 26
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 670
caccgggtct tcgagaagac ctgttt 26
<210> 671
<211> 26
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 671
tcccaggtct tcgagaagac ctgttt 26
<210> 672
<211> 26
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 672
cctcgggtct tcgagaagac ctgttt 26
<210> 673
<211> 25
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 673
ggcctaccgc taaggctagc ctagt 25
<210> 674
<211> 1411
<212> PRT
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多肽
<400> 674
Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp
1 5 10 15

[0170]

```
Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys Arg Lys Val
            20                  25                  30
Gly Arg Gly Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr
        35                  40                  45
Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser
    50                  55                  60
Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys
65                  70                  75                  80
Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala
                85                  90                  95
Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn
            100                 105                 110
Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val
        115                 120                 125
Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu
    130                 135                 140
Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu
145                 150                 155                 160
Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys
                165                 170                 175
Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala
            180                 185                 190
Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp
        195                 200                 205
Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val
    210                 215                 220
Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly
225                 230                 235                 240
Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg
                245                 250                 255
Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu
            260                 265                 270
Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys
        275                 280                 285
Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp
    290                 295                 300
Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln
305                 310                 315                 320
Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu
                325                 330                 335
Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu
            340                 345                 350
Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr
        355                 360                 365
Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu
    370                 375                 380
```

[0171]

Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly
385 390 395 400
Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu
405 410 415
Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp
420 425 430
Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln
435 440 445
Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe
450 455 460
Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr
465 470 475 480
Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg
485 490 495
Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn
500 505 510
Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu
515 520 525
Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro
530 535 540
Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr
545 550 555 560
Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser
565 570 575
Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg
580 585 590
Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu
595 600 605
Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala
610 615 620
Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp
625 630 635 640
Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu
645 650 655
Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys
660 665 670
Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg
675 680 685
Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly
690 695 700
Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser
705 710 715 720
Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser
725 730 735
Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly
740 745 750

[0172]

```
Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile
        755                 760                 765
Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys
    770                 775                 780
Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg
785                 790                 795                 800
Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met
                805                 810                 815
Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys
            820                 825                 830
Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu
        835                 840                 845
Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp
    850                 855                 860
Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser
865                 870                 875                 880
Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp
                885                 890                 895
Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys
            900                 905                 910
Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr
        915                 920                 925
Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser
    930                 935                 940
Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg
945                 950                 955                 960
Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr
                965                 970                 975
Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr
            980                 985                 990
Leu Lys Ser Lys Leu Val Ser Asp  Phe Arg Lys Asp Phe  Gln Phe Tyr
        995                 1000                1005
Lys Val  Arg Glu Ile Asn Asn  Tyr His His Ala His  Asp Ala Tyr
    1010                1015                1020
Leu Asn  Ala Val Val Gly Thr  Ala Leu Ile Lys Lys  Tyr Pro Lys
    1025                1030                1035
Leu Glu  Ser Glu Phe Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val
    1040                1045                1050
Arg Lys  Met Ile Ala Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr
    1055                1060                1065
Ala Lys  Tyr Phe Phe Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr
    1070                1075                1080
Glu Ile  Thr Leu Ala Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile
    1085                1090                1095
Glu Thr  Asn Gly Glu Thr Gly  Glu Ile Val Trp Asp  Lys Gly Arg
    1100                1105                1110
```

Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn
1115 1120 1125
Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu
1130 1135 1140
Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys
1145 1150 1155
Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr
1160 1165 1170
Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys
1175 1180 1185
Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile
1190 1195 1200
Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu
1205 1210 1215
Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu
1220 1225 1230
Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met
1235 1240 1245
Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu
1250 1255 1260
Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu
1265 1270 1275

[0173]

Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe
1280 1285 1290
Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile
1295 1300 1305
Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp
1310 1315 1320
Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg
1325 1330 1335
Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu
1340 1345 1350
Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg
1355 1360 1365
Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile
1370 1375 1380
His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser
1385 1390 1395
Gln Leu Gly Gly Asp Pro Lys Lys Lys Arg Lys Val Gly
1400 1405 1410

<210> 675
<211> 1483
<212> PRT
<213> 人工序列
<220>
<223> 人工序列的描述：合成

多肽

<400> 675

Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp
1 5 10 15
Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys Arg Lys Val
20 25 30
Gly Arg Gly Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr
35 40 45
Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser
50 55 60
Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys
65 70 75 80
Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala
85 90 95
Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn
100 105 110
Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val
115 120 125
Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu
130 135 140
Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu
145 150 155 160

[0174]

Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys
165 170 175
Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala
180 185 190
Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp
195 200 205
Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val
210 215 220
Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly
225 230 235 240
Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg
245 250 255
Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu
260 265 270
Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys
275 280 285
Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp
290 295 300
Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln
305 310 315 320
Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu
325 330 335
Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu
340 345 350

[0175]

```
Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr
        355                 360                 365
Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu
    370                 375                 380
Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly
385                 390                 395                 400
Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu
                405                 410                 415
Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp
            420                 425                 430
Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln
        435                 440                 445
Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe
    450                 455                 460
Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr
465                 470                 475                 480
Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg
                485                 490                 495
Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn
            500                 505                 510
Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu
        515                 520                 525
Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro
    530                 535                 540
Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr
545                 550                 555                 560
Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser
                565                 570                 575
Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg
            580                 585                 590
Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu
        595                 600                 605
Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala
    610                 615                 620
Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp
625                 630                 635                 640
Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu
                645                 650                 655
Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys
            660                 665                 670
Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg
        675                 680                 685
Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly
    690                 695                 700
Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser
705                 710                 715                 720
```

[0176]

```
Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser
                725                 730                 735
Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly
            740                 745                 750
Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile
        755                 760                 765
Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys
    770                 775                 780
Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg
785                 790                 795                 800
Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met
                805                 810                 815
Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys
            820                 825                 830
Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu
        835                 840                 845
Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp
    850                 855                 860
Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser
865                 870                 875                 880
Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp
                885                 890                 895
Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys
            900                 905                 910
Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr
        915                 920                 925
Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser
    930                 935                 940
Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg
945                 950                 955                 960
Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr
                965                 970                 975
Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr
            980                 985                 990
Leu Lys Ser Lys Leu Val Ser Asp  Phe Arg Lys Asp Phe  Gln Phe Tyr
        995                 1000                1005
Lys Val  Arg Glu Ile Asn Asn  Tyr His His Ala His  Asp Ala Tyr
    1010                1015                1020
Leu Asn  Ala Val Val Gly Thr  Ala Leu Ile Lys Lys  Tyr Pro Lys
    1025                1030                1035
Leu Glu  Ser Glu Phe Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val
    1040                1045                1050
Arg Lys  Met Ile Ala Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr
    1055                1060                1065
Ala Lys  Tyr Phe Phe Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr
    1070                1075                1080
```

[0177]

```
Glu Ile  Thr Leu Ala Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile
    1085                1090                1095
Glu Thr  Asn Gly Glu Thr Gly  Glu Ile Val Trp Asp  Lys Gly Arg
    1100                1105                1110
Asp Phe  Ala Thr Val Arg Lys  Val Leu Ser Met Pro  Gln Val Asn
    1115                1120                1125
Ile Val  Lys Lys Thr Glu Val  Gln Thr Gly Gly Phe  Ser Lys Glu
    1130                1135                1140
Ser Ile  Leu Pro Lys Arg Asn  Ser Asp Lys Leu Ile  Ala Arg Lys
    1145                1150                1155
Lys Asp  Trp Asp Pro Lys Lys  Tyr Gly Gly Phe Asp  Ser Pro Thr
    1160                1165                1170
Val Ala  Tyr Ser Val Leu Val  Val Ala Lys Val Glu  Lys Gly Lys
    1175                1180                1185
Ser Lys  Lys Leu Lys Ser Val  Lys Glu Leu Leu Gly  Ile Thr Ile
    1190                1195                1200
Met Glu  Arg Ser Ser Phe Glu  Lys Asn Pro Ile Asp  Phe Leu Glu
    1205                1210                1215
Ala Lys  Gly Tyr Lys Glu Val  Lys Lys Asp Leu Ile  Ile Lys Leu
    1220                1225                1230
Pro Lys  Tyr Ser Leu Phe Glu  Leu Glu Asn Gly Arg  Lys Arg Met
    1235                1240                1245
Leu Ala  Ser Ala Gly Glu Leu  Gln Lys Gly Asn Glu  Leu Ala Leu
    1250                1255                1260
Pro Ser  Lys Tyr Val Asn Phe  Leu Tyr Leu Ala Ser  His Tyr Glu
    1265                1270                1275
Lys Leu  Lys Gly Ser Pro Glu  Asp Asn Glu Gln Lys  Gln Leu Phe
    1280                1285                1290
Val Glu  Gln His Lys His Tyr  Leu Asp Glu Ile Ile  Glu Gln Ile
    1295                1300                1305
Ser Glu  Phe Ser Lys Arg Val  Ile Leu Ala Asp Ala  Asn Leu Asp
    1310                1315                1320
Lys Val  Leu Ser Ala Tyr Asn  Lys His Arg Asp Lys  Pro Ile Arg
    1325                1330                1335
Glu Gln  Ala Glu Asn Ile Ile  His Leu Phe Thr Leu  Thr Asn Leu
    1340                1345                1350
Gly Ala  Pro Ala Ala Phe Lys  Tyr Phe Asp Thr Thr  Ile Asp Arg
    1355                1360                1365
Lys Arg  Tyr Thr Ser Thr Lys  Glu Val Leu Asp Ala  Thr Leu Ile
    1370                1375                1380
His Gln  Ser Ile Thr Gly Leu  Tyr Glu Thr Arg Ile  Asp Leu Ser
    1385                1390                1395
Gln Leu  Gly Gly Asp Pro Ile  Ala Gly Ser Lys Ala  Ser Pro Lys
    1400                1405                1410
Lys Lys  Arg Lys Val Gly Arg  Ala Asp Ala Leu Asp  Asp Phe Asp
    1415                1420                1425
```

Leu Asp Met Leu Gly Ser Asp Ala Leu Asp Asp Phe Asp Leu Asp
1430 1435 1440
Met Leu Gly Ser Asp Ala Leu Asp Asp Phe Asp Leu Asp Met Leu
1445 1450 1455
Gly Ser Asp Ala Leu Asp Asp Phe Asp Leu Asp Met Leu Ile Asn
1460 1465 1470
Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Ser
1475 1480

<210> 676
<211> 2048
<212> PRT
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多肽

[0178]

<400> 676
Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp
1 5 10 15
Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys Arg Lys Val
20 25 30
Gly Arg Gly Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr
35 40 45
Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser
50 55 60
Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys
65 70 75 80
Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala
85 90 95
Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn
100 105 110
Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val
115 120 125
Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu
130 135 140
Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu
145 150 155 160
Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys
165 170 175
Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala
180 185 190
Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp
195 200 205
Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val
210 215 220
Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly
225 230 235 240

[0179]

```
Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg
                245                 250                 255
Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu
            260                 265                 270
Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys
        275                 280                 285
Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp
    290                 295                 300
Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln
305                 310                 315                 320
Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu
                325                 330                 335
Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu
            340                 345                 350
Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr
        355                 360                 365
Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu
    370                 375                 380
Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly
385                 390                 395                 400
Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu
                405                 410                 415
Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp
            420                 425                 430
Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln
        435                 440                 445
Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe
    450                 455                 460
Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr
465                 470                 475                 480
Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg
                485                 490                 495
Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn
            500                 505                 510
Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu
        515                 520                 525
Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro
    530                 535                 540
Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr
545                 550                 555                 560
Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser
                565                 570                 575
Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg
            580                 585                 590
Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu
        595                 600                 605
```

[0180]

```
Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala
    610                 615                 620
Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp
625                 630                 635                 640
Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu
                645                 650                 655
Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys
            660                 665                 670
Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg
        675                 680                 685
Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly
    690                 695                 700
Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser
705                 710                 715                 720
Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser
                725                 730                 735
Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly
            740                 745                 750
Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile
        755                 760                 765
Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys
    770                 775                 780
Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg
785                 790                 795                 800
Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met
                805                 810                 815
Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys
            820                 825                 830
Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu
        835                 840                 845
Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp
    850                 855                 860
Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser
865                 870                 875                 880
Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp
                885                 890                 895
Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys
            900                 905                 910
Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr
        915                 920                 925
Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser
    930                 935                 940
Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg
945                 950                 955                 960
Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr
                965                 970                 975
```

Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr
980 985 990
Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr
995 1000 1005
Lys Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr
1010 1015 1020
Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys
1025 1030 1035
Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val
1040 1045 1050
Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr
1055 1060 1065
Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr
1070 1075 1080
Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile
1085 1090 1095
Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg
1100 1105 1110
Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn
1115 1120 1125
Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu
1130 1135 1140
[0181] Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys
1145 1150 1155
Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr
1160 1165 1170
Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys
1175 1180 1185
Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile
1190 1195 1200
Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu
1205 1210 1215
Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu
1220 1225 1230
Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met
1235 1240 1245
Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu
1250 1255 1260
Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu
1265 1270 1275
Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe
1280 1285 1290
Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile
1295 1300 1305
Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp
1310 1315 1320

[0182]

```
Lys Val  Leu Ser Ala Tyr Asn  Lys His Arg Asp Lys  Pro Ile Arg
    1325                 1330                 1335
Glu Gln  Ala Glu Asn Ile Ile  His Leu Phe Thr Leu  Thr Asn Leu
    1340                 1345                 1350
Gly Ala  Pro Ala Ala Phe Lys  Tyr Phe Asp Thr Thr  Ile Asp Arg
    1355                 1360                 1365
Lys Arg  Tyr Thr Ser Thr Lys  Glu Val Leu Asp Ala  Thr Leu Ile
    1370                 1375                 1380
His Gln  Ser Ile Thr Gly Leu  Tyr Glu Thr Arg Ile  Asp Leu Ser
    1385                 1390                 1395
Gln Leu  Gly Gly Asp Pro Ile  Ala Gly Ser Lys Ala  Ser Pro Lys
    1400                 1405                 1410
Lys Lys  Arg Lys Val Gly Arg  Ala Ile Phe Lys Pro  Glu Glu Leu
    1415                 1420                 1425
Arg Gln  Ala Leu Met Pro Thr  Leu Glu Ala Leu Tyr  Arg Gln Asp
    1430                 1435                 1440
Pro Glu  Ser Leu Pro Phe Arg  Gln Pro Val Asp Pro  Gln Leu Leu
    1445                 1450                 1455
Gly Ile  Pro Asp Tyr Phe Asp  Ile Val Lys Ser Pro  Met Asp Leu
    1460                 1465                 1470
Ser Thr  Ile Lys Arg Lys Leu  Asp Thr Gly Gln Tyr  Gln Glu Pro
    1475                 1480                 1485
Trp Gln  Tyr Val Asp Asp Ile  Trp Leu Met Phe Asn  Asn Ala Trp
    1490                 1495                 1500
Leu Tyr  Asn Arg Lys Thr Ser  Arg Val Tyr Lys Tyr  Cys Ser Lys
    1505                 1510                 1515
Leu Ser  Glu Val Phe Glu Gln  Glu Ile Asp Pro Val  Met Gln Ser
    1520                 1525                 1530
Leu Gly  Tyr Cys Cys Gly Arg  Lys Leu Glu Phe Ser  Pro Gln Thr
    1535                 1540                 1545
Leu Cys  Cys Tyr Gly Lys Gln  Leu Cys Thr Ile Pro  Arg Asp Ala
    1550                 1555                 1560
Thr Tyr  Tyr Ser Tyr Gln Asn  Arg Tyr His Phe Cys  Glu Lys Cys
    1565                 1570                 1575
Phe Asn  Glu Ile Gln Gly Glu  Ser Val Ser Leu Gly  Asp Asp Pro
    1580                 1585                 1590
Ser Gln  Pro Gln Thr Thr Ile  Asn Lys Glu Gln Phe  Ser Lys Arg
    1595                 1600                 1605
Lys Asn  Asp Thr Leu Asp Pro  Glu Leu Phe Val Glu  Cys Thr Glu
    1610                 1615                 1620
Cys Gly  Arg Lys Met His Gln  Ile Cys Val Leu His  His Glu Ile
    1625                 1630                 1635
Ile Trp  Pro Ala Gly Phe Val  Cys Asp Gly Cys Leu  Lys Lys Ser
    1640                 1645                 1650
Ala Arg  Thr Arg Lys Glu Asn  Lys Phe Ser Ala Lys  Arg Leu Pro
    1655                 1660                 1665
```

[0183]

```
Ser Thr  Arg Leu Gly Thr Phe  Leu Glu Asn Arg Val  Asn Asp Phe
    1670                 1675                 1680
Leu Arg  Arg Gln Asn His Pro  Glu Ser Gly Glu Val  Thr Val Arg
    1685                 1690                 1695
Val Val  His Ala Ser Asp Lys  Thr Val Glu Val Lys  Pro Gly Met
    1700                 1705                 1710
Lys Ala  Arg Phe Val Asp Ser  Gly Glu Met Ala Glu  Ser Phe Pro
    1715                 1720                 1725
Tyr Arg  Thr Lys Ala Leu Phe  Ala Phe Glu Glu Ile  Asp Gly Val
    1730                 1735                 1740
Asp Leu  Cys Phe Phe Gly Met  His Val Gln Glu Tyr  Gly Ser Asp
    1745                 1750                 1755
Cys Pro  Pro Pro Asn Gln Arg  Arg Val Tyr Ile Ser  Tyr Leu Asp
    1760                 1765                 1770
Ser Val  His Phe Phe Arg Pro  Lys Cys Leu Arg Thr  Ala Val Tyr
    1775                 1780                 1785
His Glu  Ile Leu Ile Gly Tyr  Leu Glu Tyr Val Lys  Lys Leu Gly
    1790                 1795                 1800
Tyr Thr  Thr Gly His Ile Trp  Ala Cys Pro Pro Ser  Glu Gly Asp
    1805                 1810                 1815
Asp Tyr  Ile Phe His Cys His  Pro Pro Asp Gln Lys  Ile Pro Lys
    1820                 1825                 1830
Pro Lys  Arg Leu Gln Glu Trp  Tyr Lys Lys Met Leu  Asp Lys Ala
    1835                 1840                 1845
Val Ser  Glu Arg Ile Val His  Asp Tyr Lys Asp Ile  Phe Lys Gln
    1850                 1855                 1860
Ala Thr  Glu Asp Arg Leu Thr  Ser Ala Lys Glu Leu  Pro Tyr Phe
    1865                 1870                 1875
Glu Gly  Asp Phe Trp Pro Asn  Val Leu Glu Glu Ser  Ile Lys Glu
    1880                 1885                 1890
Leu Glu  Gln Glu Glu Glu Glu  Arg Lys Arg Glu Glu  Asn Thr Ser
    1895                 1900                 1905
Asn Glu  Ser Thr Asp Val Thr  Lys Gly Asp Ser Lys  Asn Ala Lys
    1910                 1915                 1920
Lys Lys  Asn Asn Lys Lys Thr  Ser Lys Asn Lys Ser  Ser Leu Ser
    1925                 1930                 1935
Arg Gly  Asn Lys Lys Lys Pro  Gly Met Pro Asn Val  Ser Asn Asp
    1940                 1945                 1950
Leu Ser  Gln Lys Leu Tyr Ala  Thr Met Glu Lys His  Lys Glu Val
    1955                 1960                 1965
Phe Phe  Val Ile Arg Leu Ile  Ala Gly Pro Ala Ala  Asn Ser Leu
    1970                 1975                 1980
Pro Pro  Ile Val Asp Pro Asp  Pro Leu Ile Pro Cys  Asp Leu Met
    1985                 1990                 1995
Asp Gly  Arg Asp Ala Phe Leu  Thr Leu Ala Arg Asp  Lys His Leu
    2000                 2005                 2010
```

[0184]

Glu Phe Ser Ser Leu Arg Arg Ala Gln Trp Ser Thr Met Cys Met
2015 2020 2025
Leu Val Glu Leu His Thr Gln Ser Gln Asp Tyr Pro Tyr Asp Val
2030 2035 2040
Pro Asp Tyr Ala Ser
2045

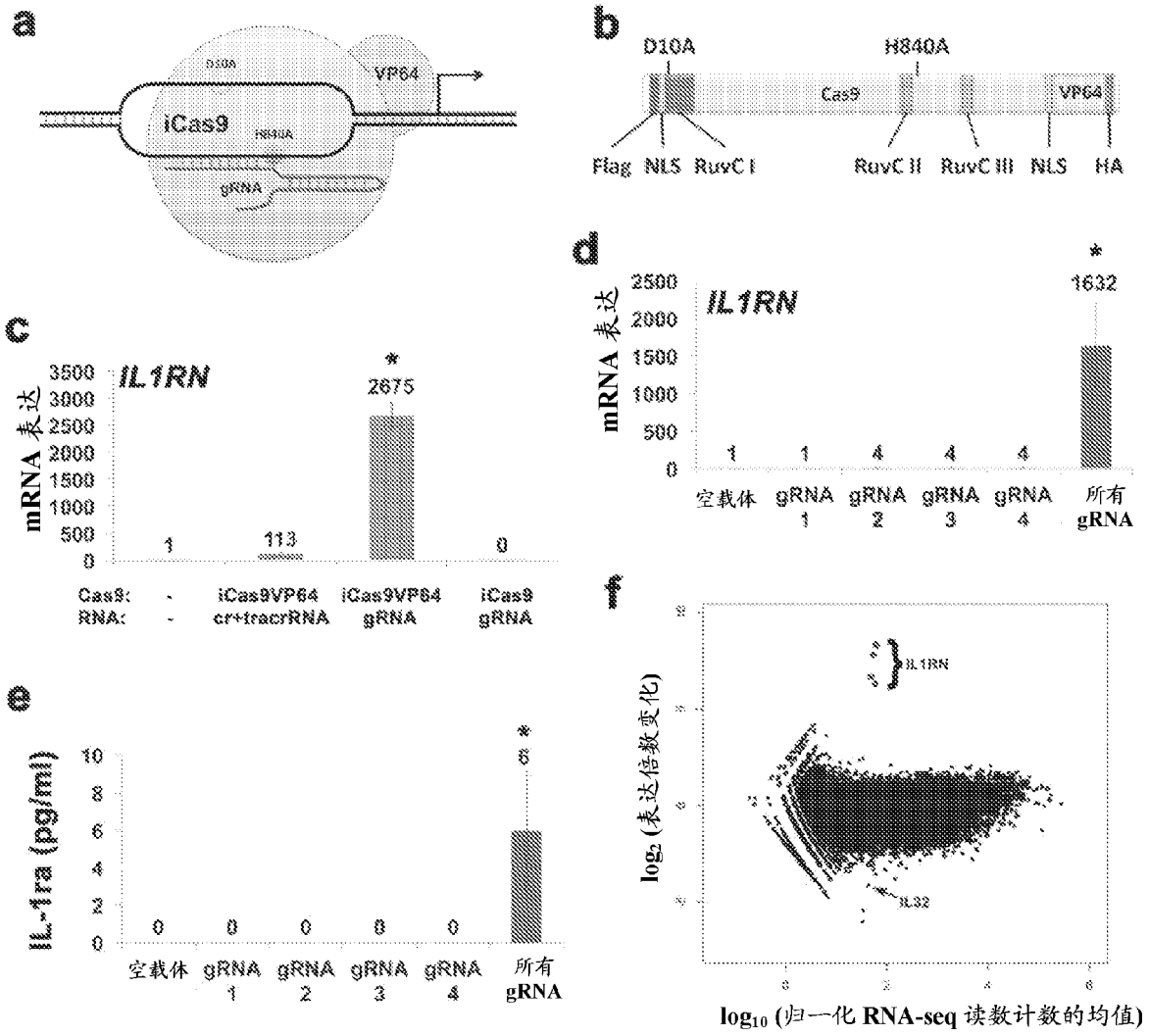

a
D10A
VP64
iCas9
H840A
gRNA
b
D10A
H840A
Cas9
VP64
Flag
NLS
RuvC I
RuvC II
RuvC III
NLS
HA
c
IL1RN
mRNA 表达
3500
3000
2500
2000
1500
1000
500
0
*
2675
1
113
0
Cas9:
RNA:
-
-
iCas9VP64
cr+tracrRNA
iCas9VP64
gRNA
iCas9
gRNA
d
IL1RN
mRNA 表达
2500
2000
1500
1000
500
0
*
1632
1
1
4
4
4
空载体
gRNA 1
gRNA 2
gRNA 3
gRNA 4
所有 gRNA
e
IL-1ra (pg/ml)
10
8
6
4
2
0
*
0
0
0
8
0
空载体
gRNA 1
gRNA 2
gRNA 3
gRNA 4
所有 gRNA
f
IL1RN
IL32
log2 (表达倍数变化)
log10 (归一化 RNA-seq 读数计数的均值)

图 1

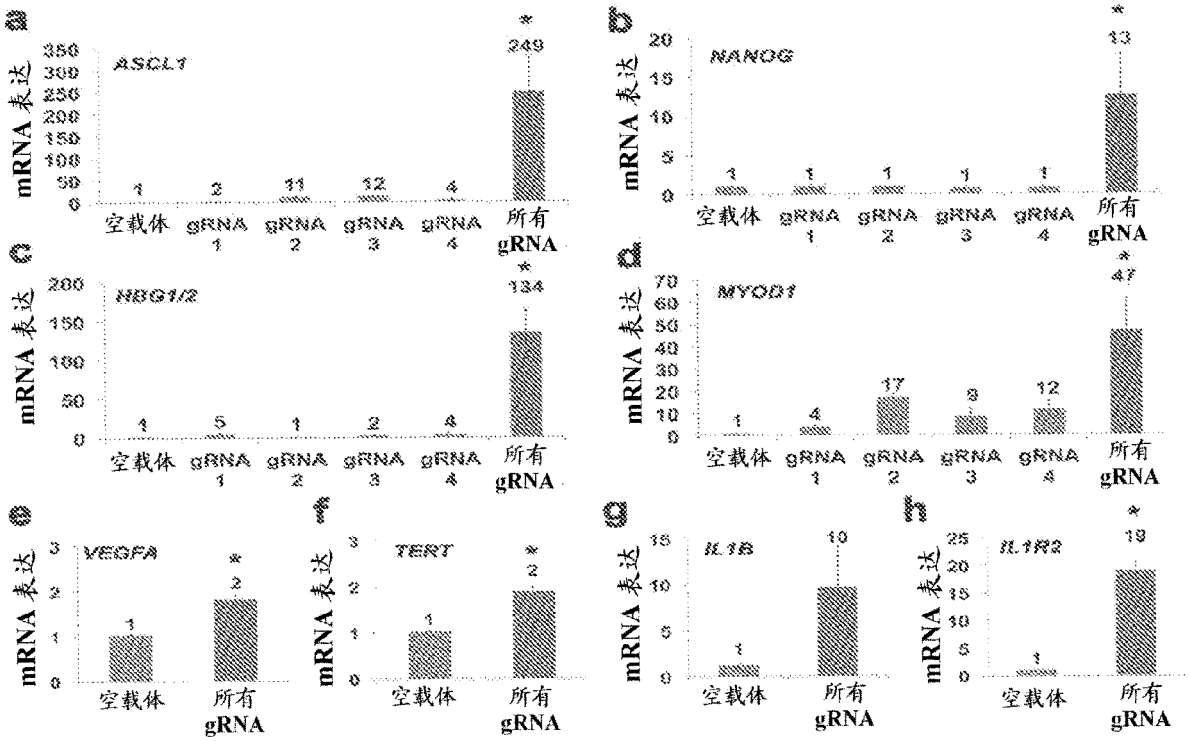

a
ASCL1
mRNA 表达
空载体
gRNA 1
gRNA 2
gRNA 3
gRNA 4
所有 gRNA
b
NANOG
c
NBG1/2
d
MYOD1
e
VEGFA
f
TERT
g
IL1B
h
IL1R2

图 2

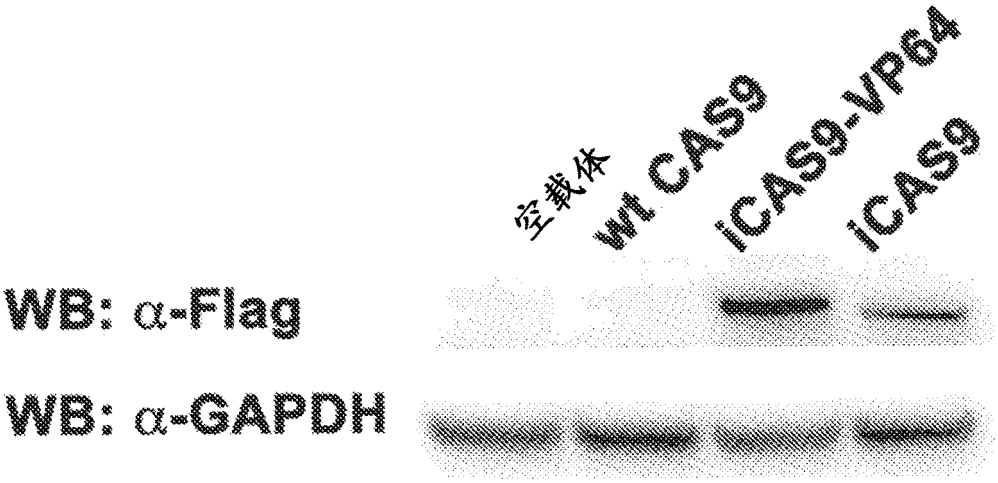

空载体
wt CAS9
iCAS9-VP64
iCAS9
WB: α-Flag
WB: α-GAPDH

图 3

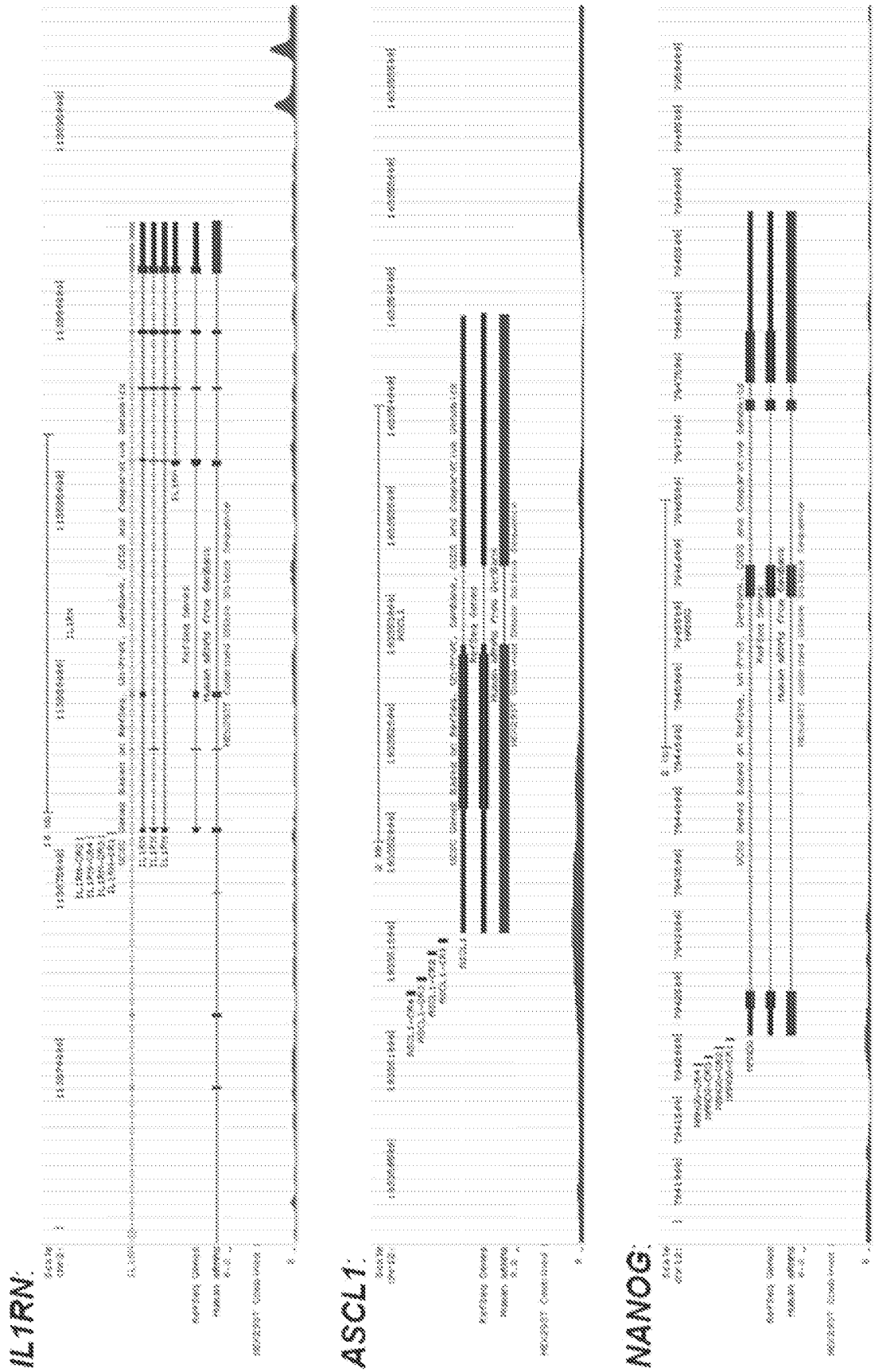


图 4

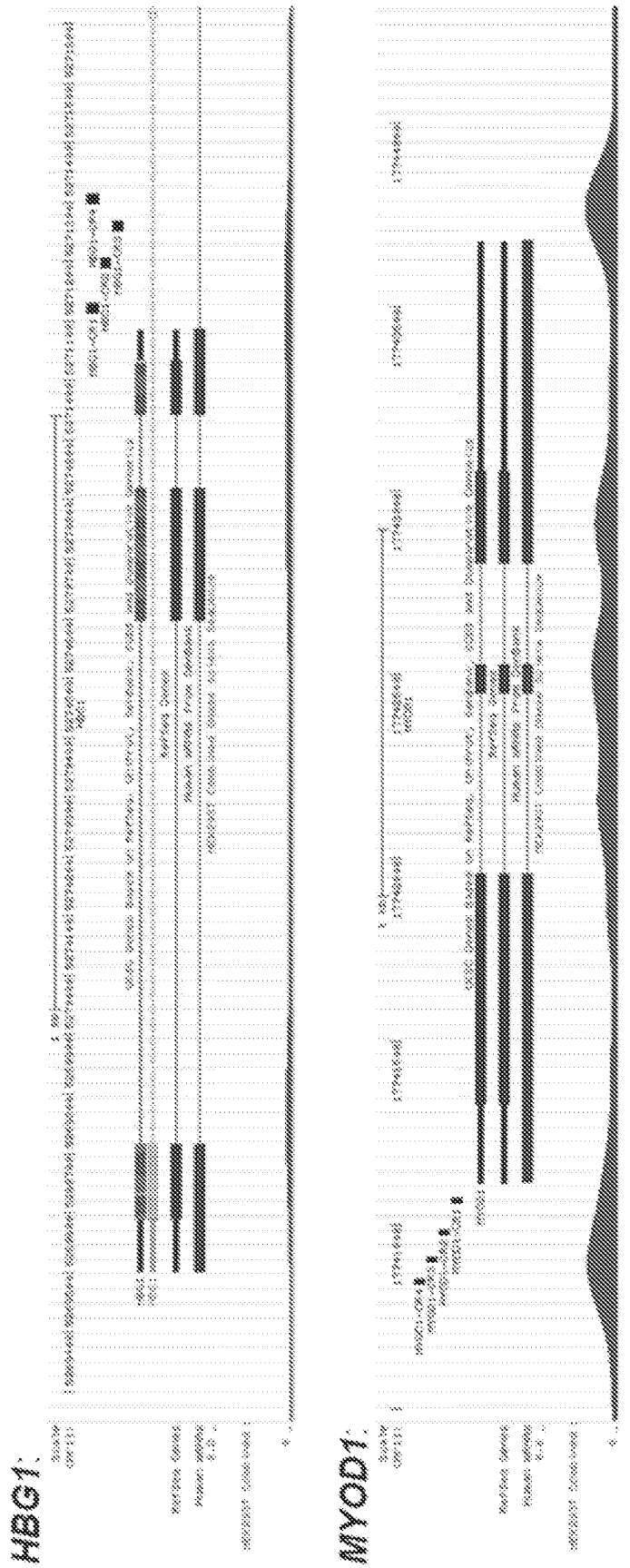


图 4(续)

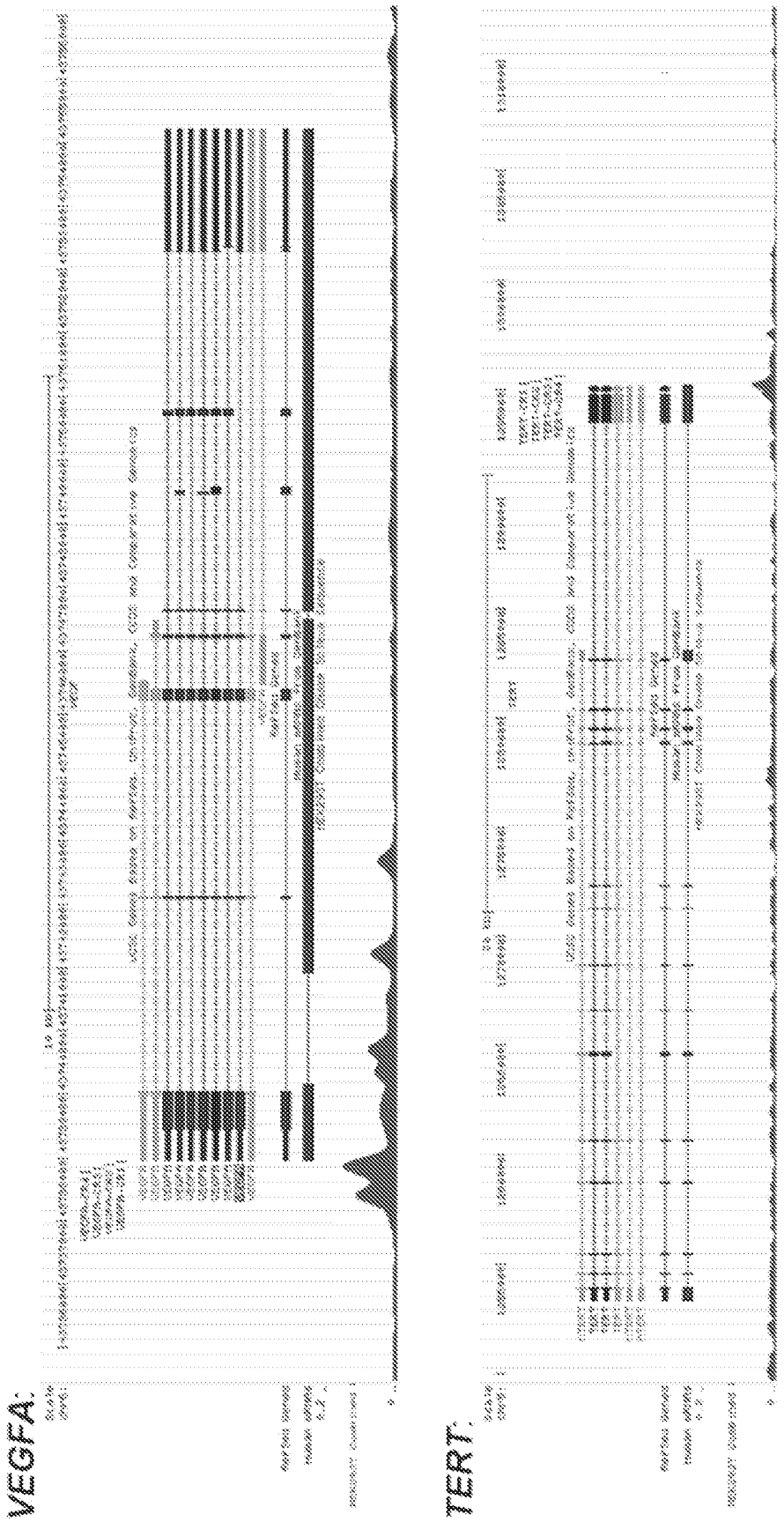


图 4(续)

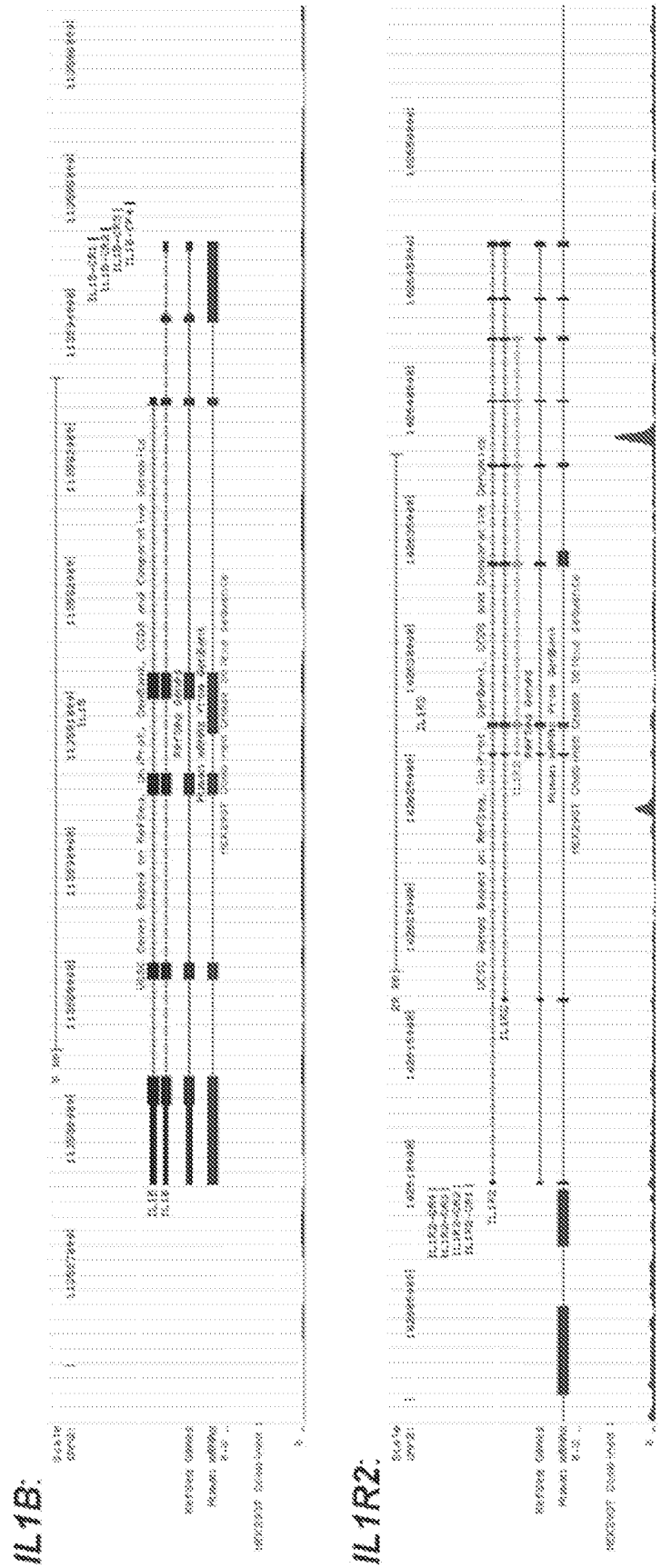


图 4(续)

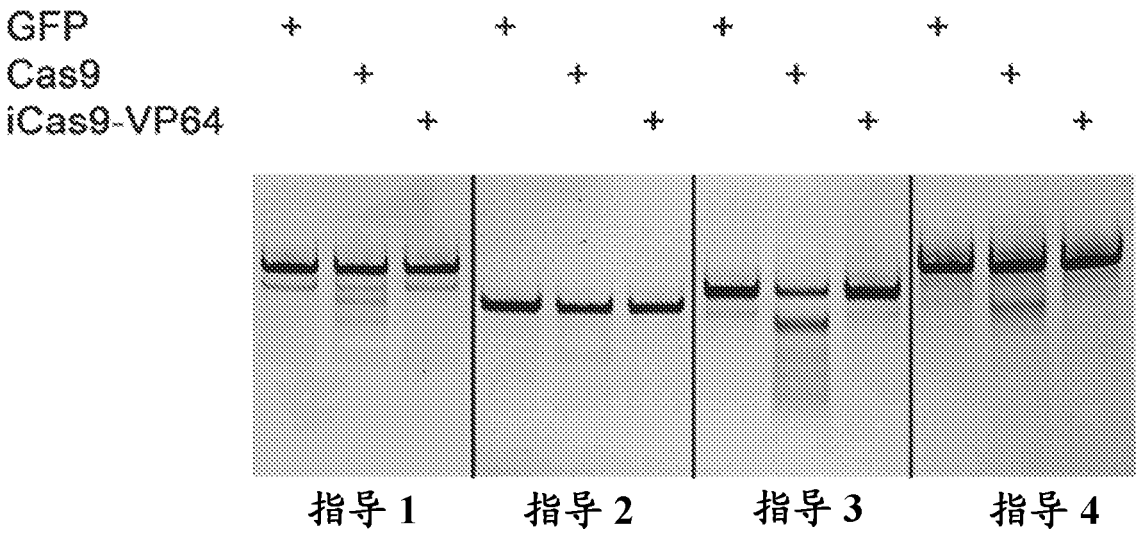

GFP
Cas9
iCas9-VP64
指导 1
指导 2
指导 3
指导 4

图 5

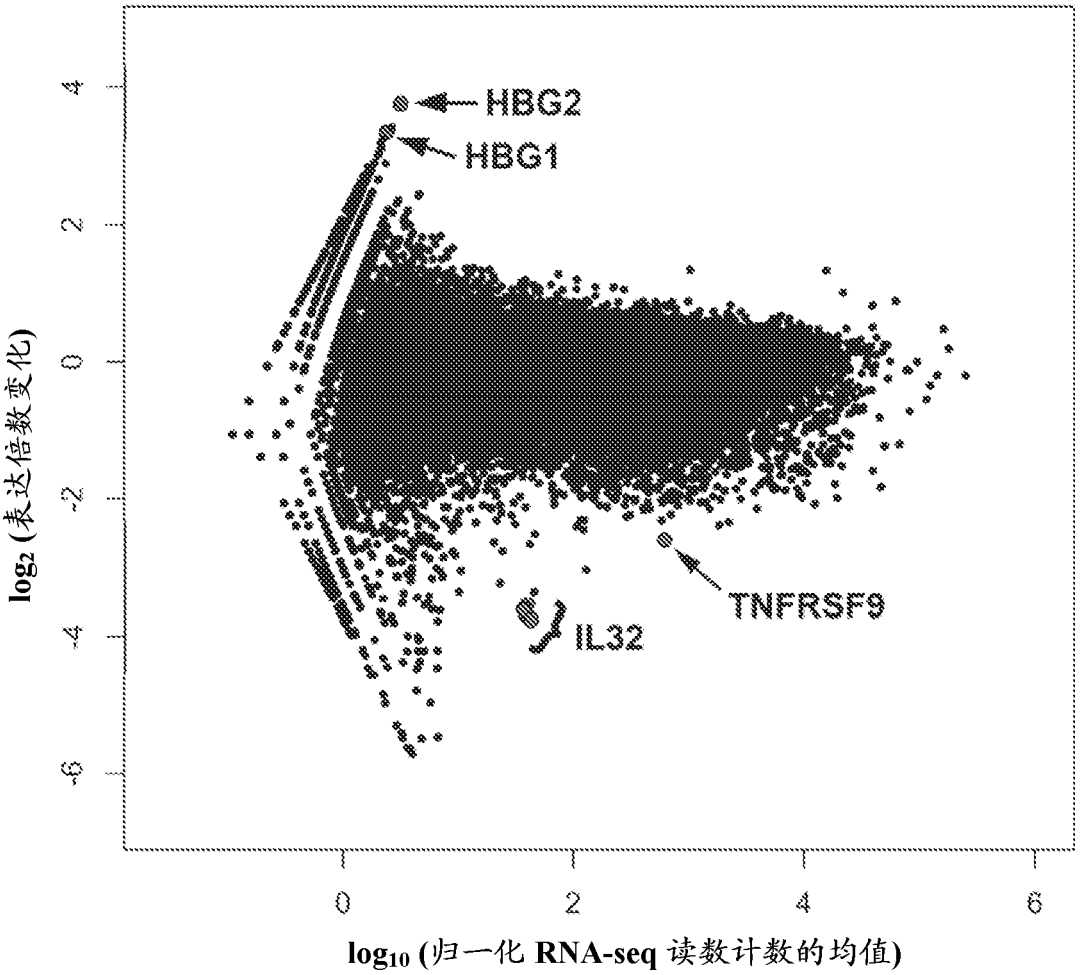

HBG2
HBG1
TNFRSF9
IL32
log2 (表达倍数变化)
log10 (归一化 RNA-seq 读数计数的均值)
4
2
0
-2
-4
-6
0
2
4
6

图 6

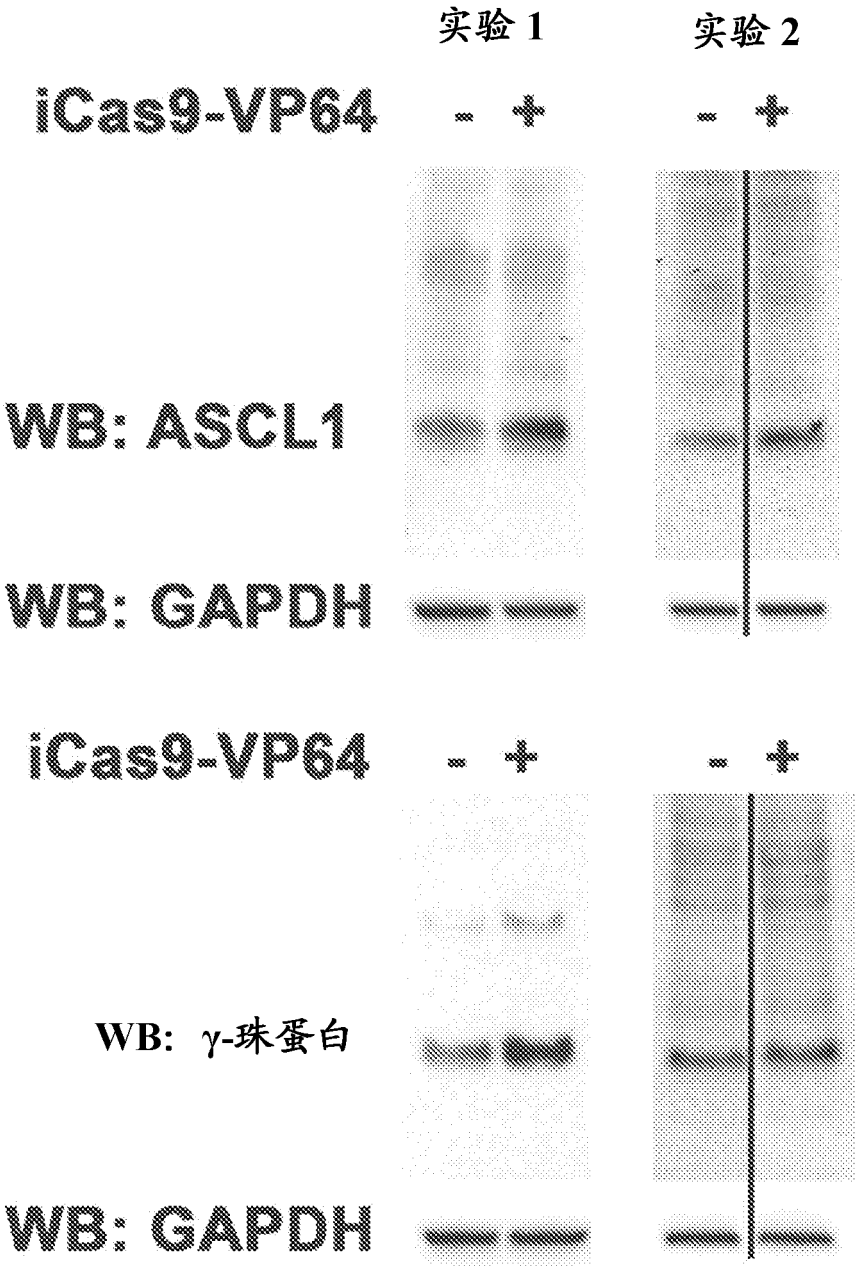

实验 1
实验 2
iCas9-VP64
- + - +
WB: ASCL1
WB: GAPDH
iCas9-VP64
- + - +
WB: γ-珠蛋白
WB: GAPDH

图 7

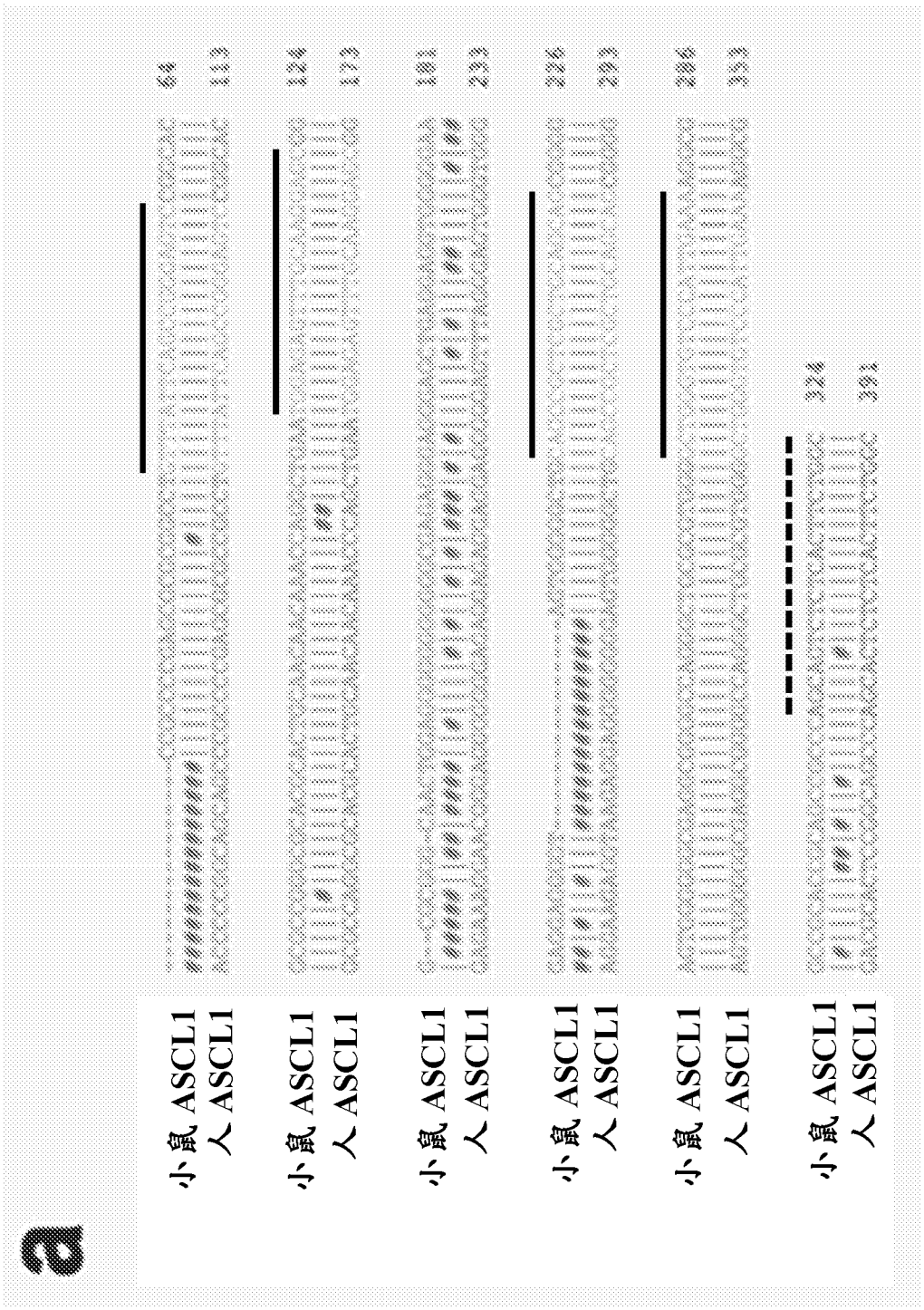

a
小鼠 ASCL1
人 ASCL1
64
113
小鼠 ASCL1
人 ASCL1
124
173
小鼠 ASCL1
人 ASCL1
181
233
小鼠 ASCL1
人 ASCL1
226
293
小鼠 ASCL1
人 ASCL1
226
353
小鼠 ASCL1
人 ASCL1
324
393

图 8

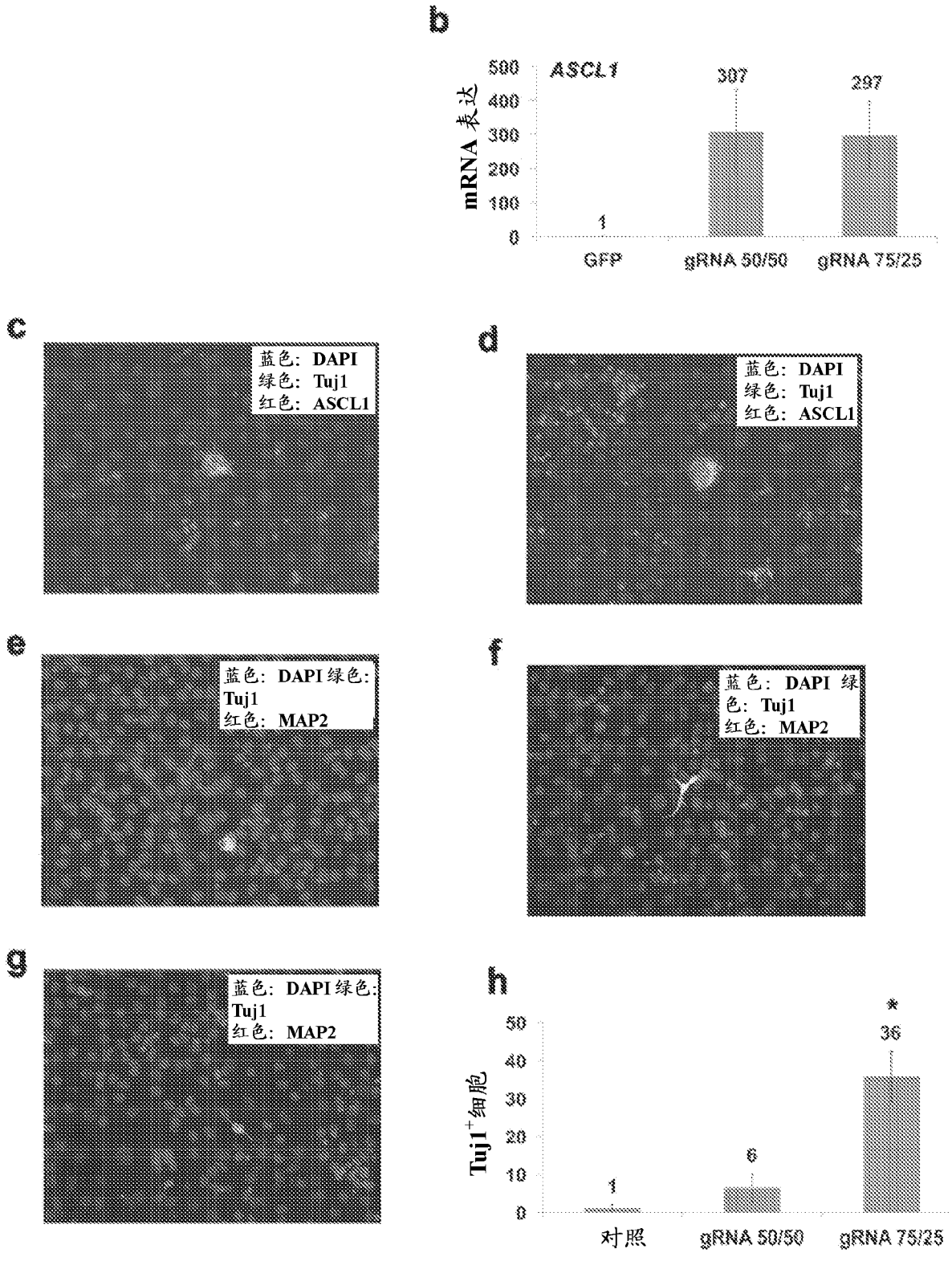

b
ASCL1
mRNA 表达
500
400
300
200
100
0
1
307
297
GFP
gRNA 50/50
gRNA 75/25
c
蓝色：DAPI
绿色：Tuj1
红色：ASCL1
d
蓝色：DAPI
绿色：Tuj1
红色：ASCL1
e
蓝色：DAPI 绿色：Tuj1
红色：MAP2
f
蓝色：DAPI 绿色：Tuj1
红色：MAP2
g
蓝色：DAPI 绿色：Tuj1
红色：MAP2
h
Tuj1+细胞
50
40
30
20
10
0
1
6
36
*
对照
gRNA 50/50
gRNA 75/25

图 8(续)

*MDYKDHDGDYKDHDIDYKDDDDK*MA**PKKKRKV**GRGMDKKYSIGLAIGTNSVGWAVITD
EYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQ
EIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVD
STDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASG
VDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQL
SKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEH
HQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEE
LLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVL
PKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDY
FKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRE
MIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFAN
RNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNE
KLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQIT
KHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAY
LNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTE
ITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESI
LPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIME
RSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSK
YVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLS
AYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITG
LYETRIDLSQLGGDPIAGSKAS**PKKKRKV**GRA***DALDDFDLDMLGSDALDDFDLDMLGS
DALDDFDLDMLGSDALDDFDLDML***INYPYDVPDYAS (SEQ ID NO: 1)

FLAG 表位表签=斜体字
核定位序列=粗体字
酿脓链球菌 **Cas9**=下划线
VP64 (4X 最小 **VP16** 结构域**)**=斜体字+粗体字
HA 表位表签=斜体字+下划线

图 9A

**GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTG
TTAGAGAGATAATTGGAATTAATTTGACTGTAAACACAAAGATATTAGTACA
AAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAA
TTATGTTTTAAAATGGACTATCATATGCTTATCGTAACTTGAAAGTATTTCGA
TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCG**GG*GTCTTC*GA*GA
AGAC*CT*GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATC
AACTTGAAAAAGTGGCACCGAGTCGGTGC**TTTTTTT** (SEQ ID NO:2)

U6 启动子=粗体字
+1 转录起始位点=下划线
在指导 **RNA** 中克隆的 **BbsI** 限制位点=斜体字+下划线
嵌合指导 **RNA** 序列=斜体字
聚 **T** 终止子序列=粗体字+下划线

图 9B

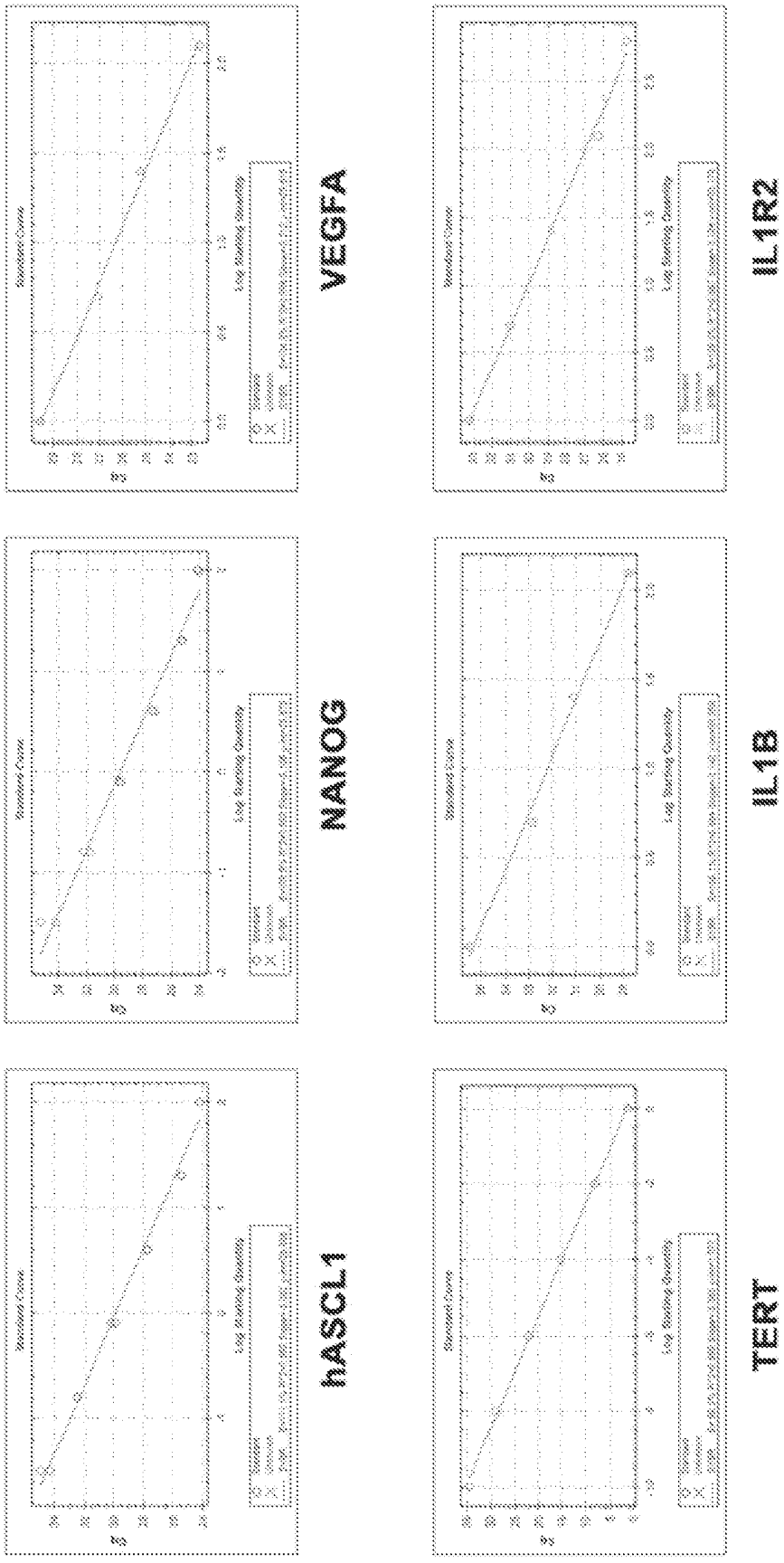


图 10

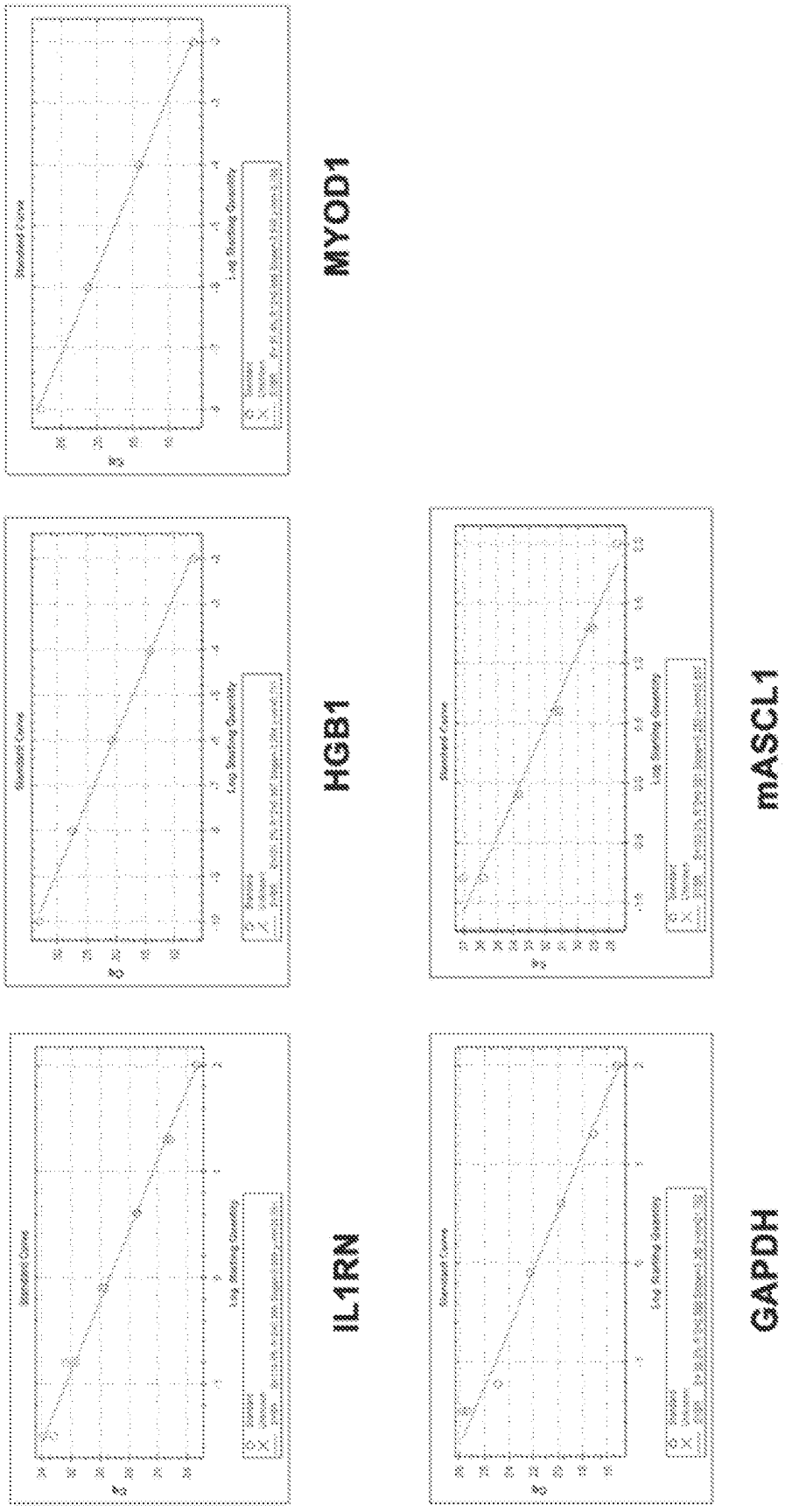
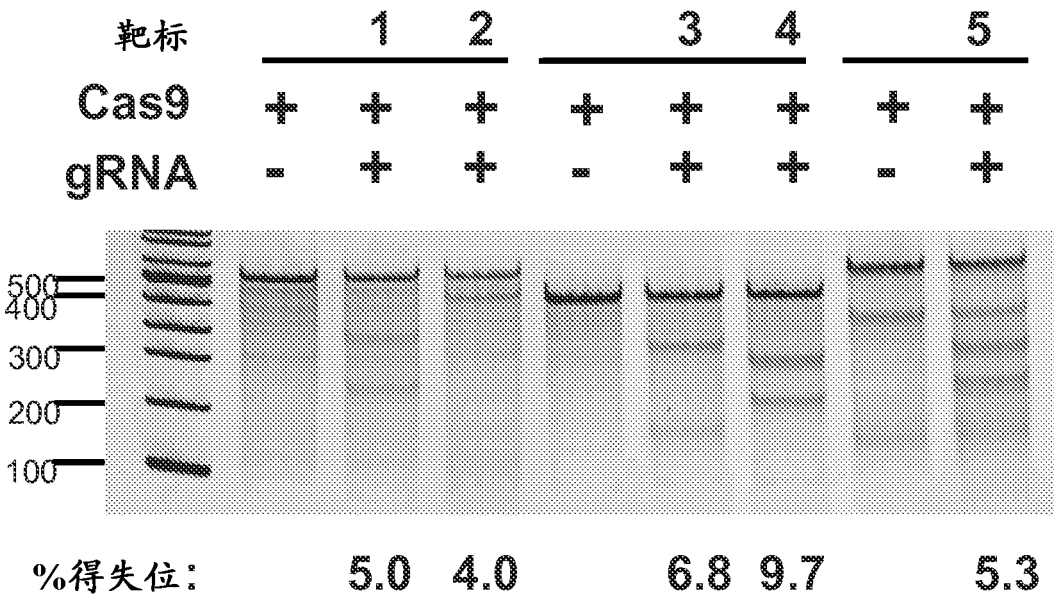
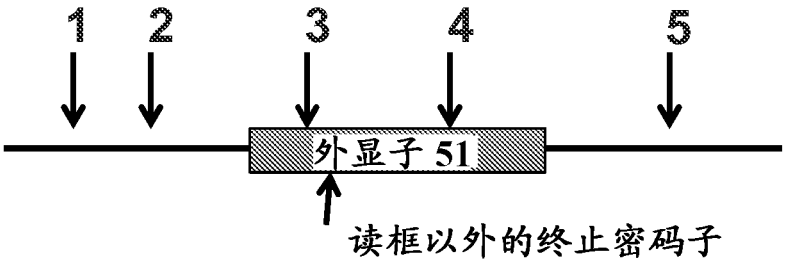


图 10(续)

a



b



c

预期切割大小

CR1=352/237

CR2=461/128

CR3=301/150

CR4=272/179

CR5=280/203

图 11

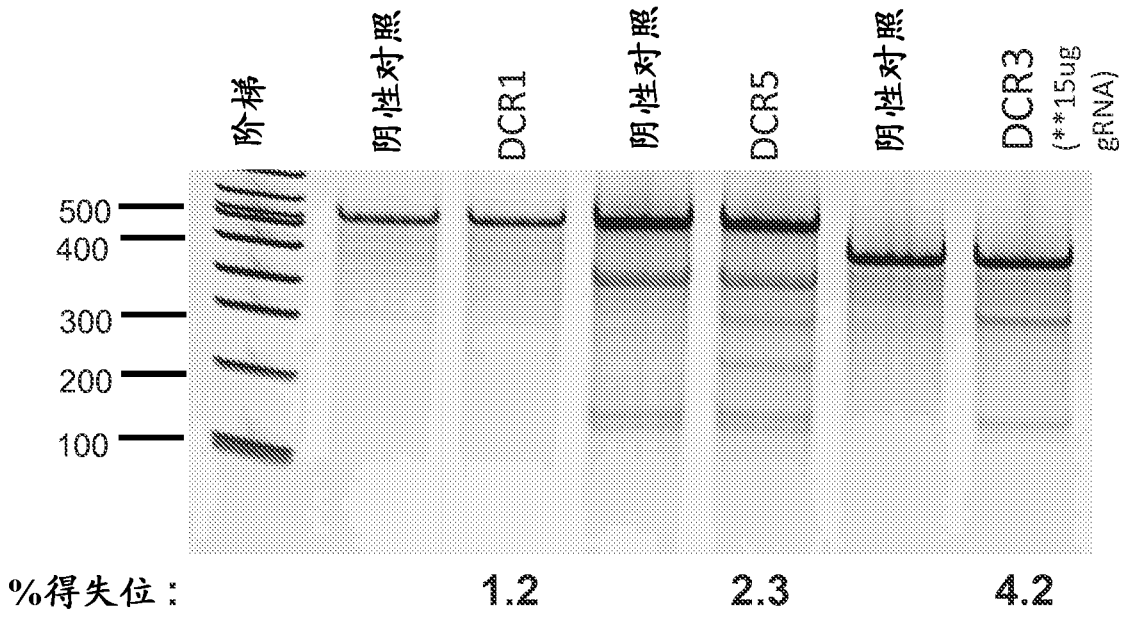

阶梯
阴性对照
DCR1
阴性对照
DCR5
阴性对照
DCR3
(**15ug gRNA)
500
400
300
200
100
%得失位：
1.2
2.3
4.2

图 12

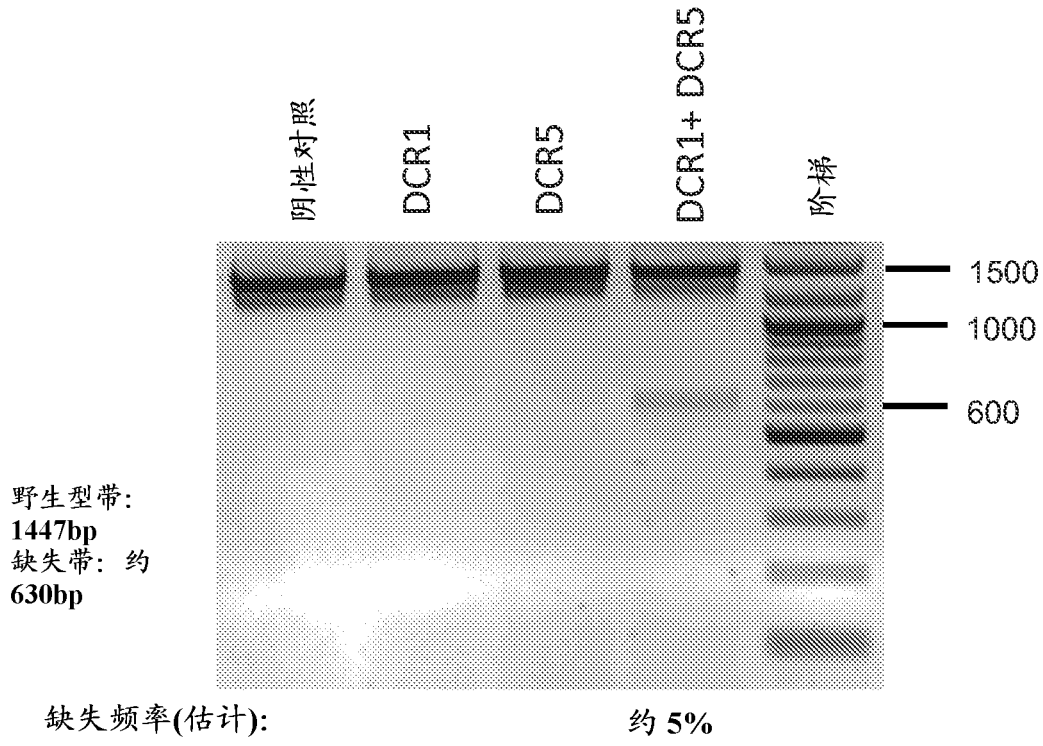

阴性对照
DCR1
DCR5
DCR1+ DCR5
阶梯
1500
1000
600
野生型带:
1447bp
缺失带：约
630bp
缺失频率(估计):
约 5%

图 13

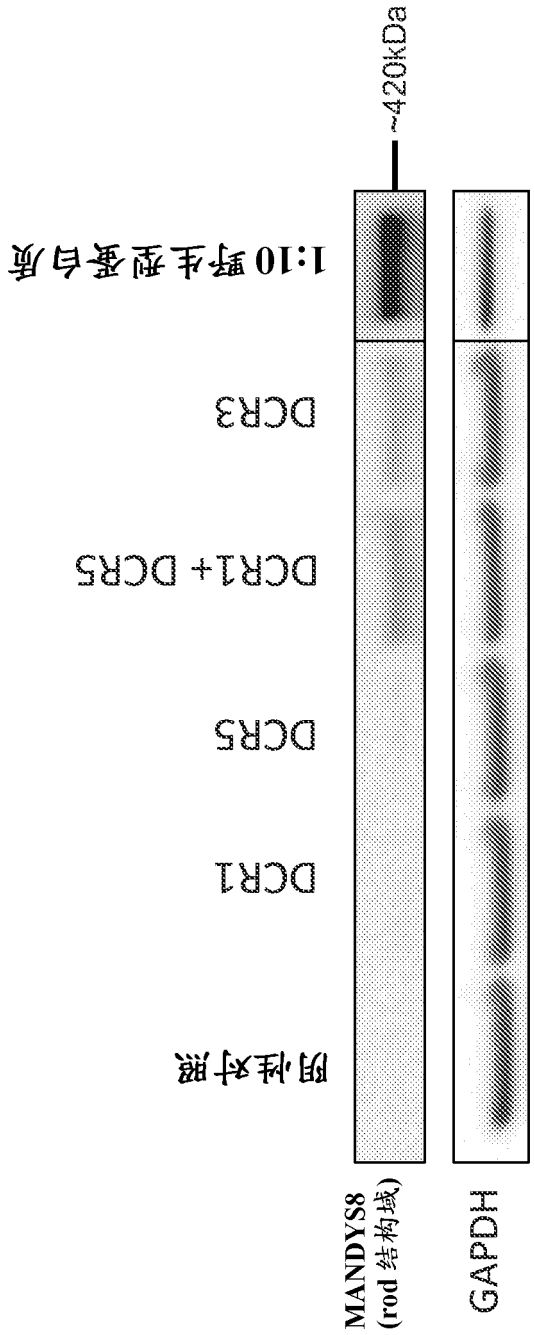

~420kDa
1:10 野生型蛋白质
DCR3
DCR1+ DCR5
DCR5
DCR1
阴性对照
MANDYS8
(rod 结构域)
GAPDH

图 14

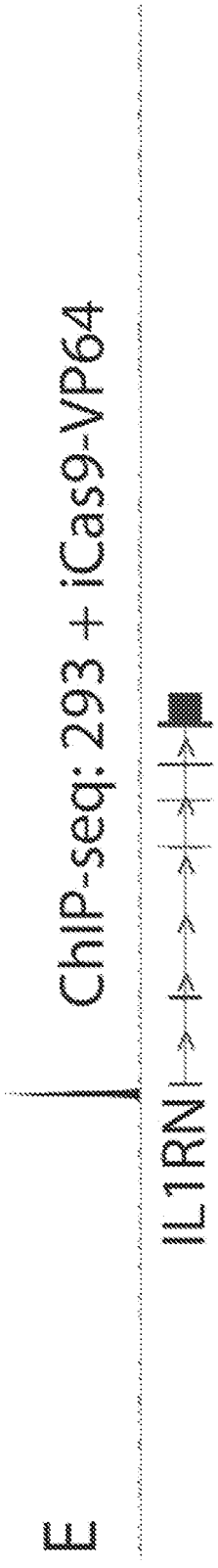


图 15

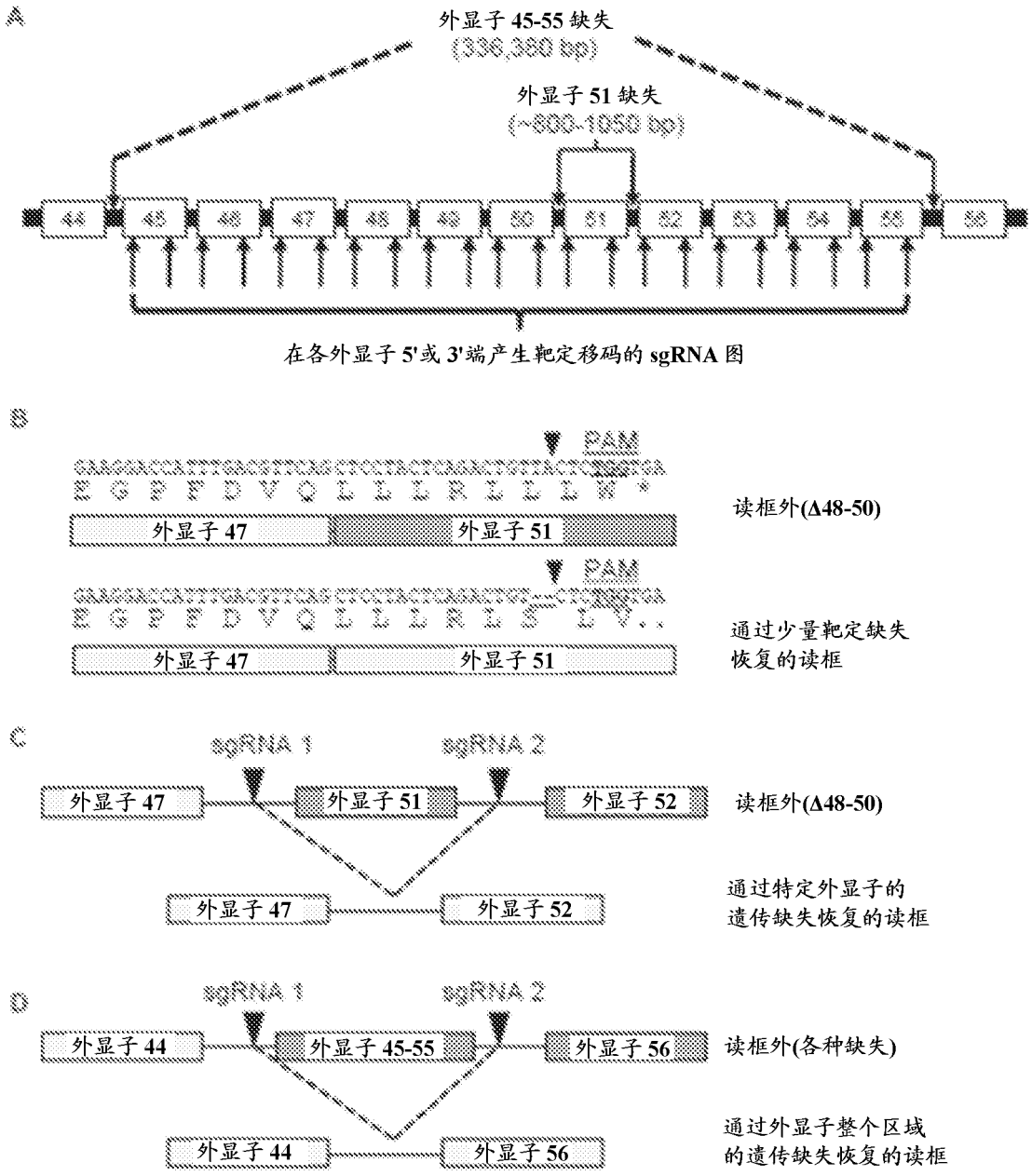

A
外显子 45-55 缺失
(336,380 bp)
外显子 51 缺失
(~800-1050 bp)
在各外显子 5'或 3'端产生靶定移码的 sgRNA 图
B
PAM
外显子 47
外显子 51
读框外(Δ48-50)
通过少量靶定缺失
恢复的读框
C
sgRNA 1
sgRNA 2
外显子 47
外显子 51
外显子 52
读框外(Δ48-50)
通过特定外显子的
遗传缺失恢复的读框
D
sgRNA 1
sgRNA 2
外显子 44
外显子 45-55
外显子 56
读框外(各种缺失)
通过外显子整个区域
的遗传缺失恢复的读框

图 16

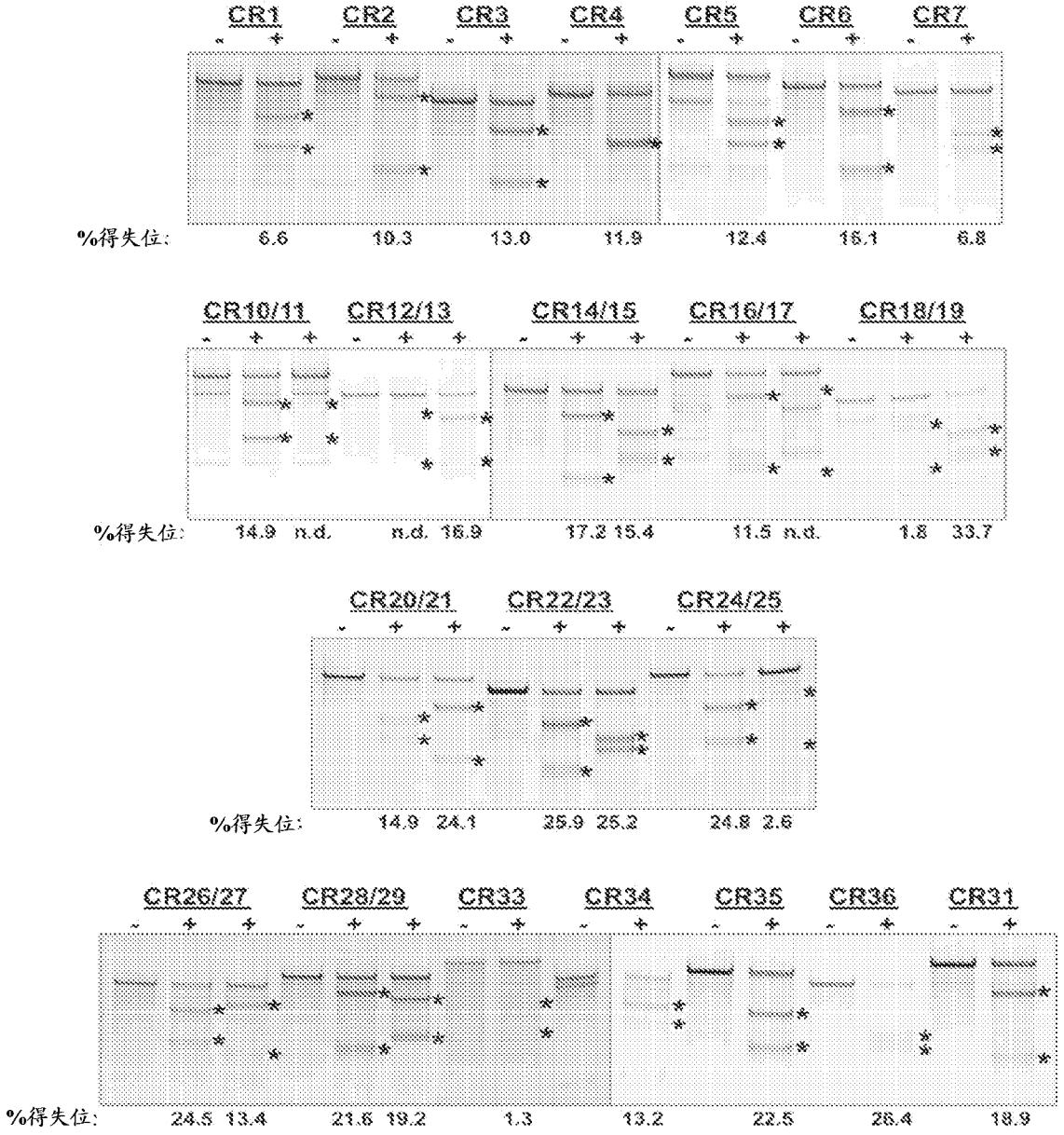

CR1
CR2
CR3
CR4
CR5
CR6
CR7
%得失位: 6.6 10.3 13.0 11.9 12.4 16.1 6.8
CR10/11
CR12/13
CR14/15
CR16/17
CR18/19
%得失位: 14.9 n.d. n.d. 16.9 17.2 15.4 11.5 n.d. 1.8 33.7
CR20/21
CR22/23
CR24/25
%得失位: 14.9 24.1 25.9 25.2 24.8 2.6
CR26/27
CR28/29
CR33
CR34
CR35
CR36
CR31
%得失位: 24.5 13.4 21.6 19.2 1.3 13.2 22.5 26.4 18.9

图 17

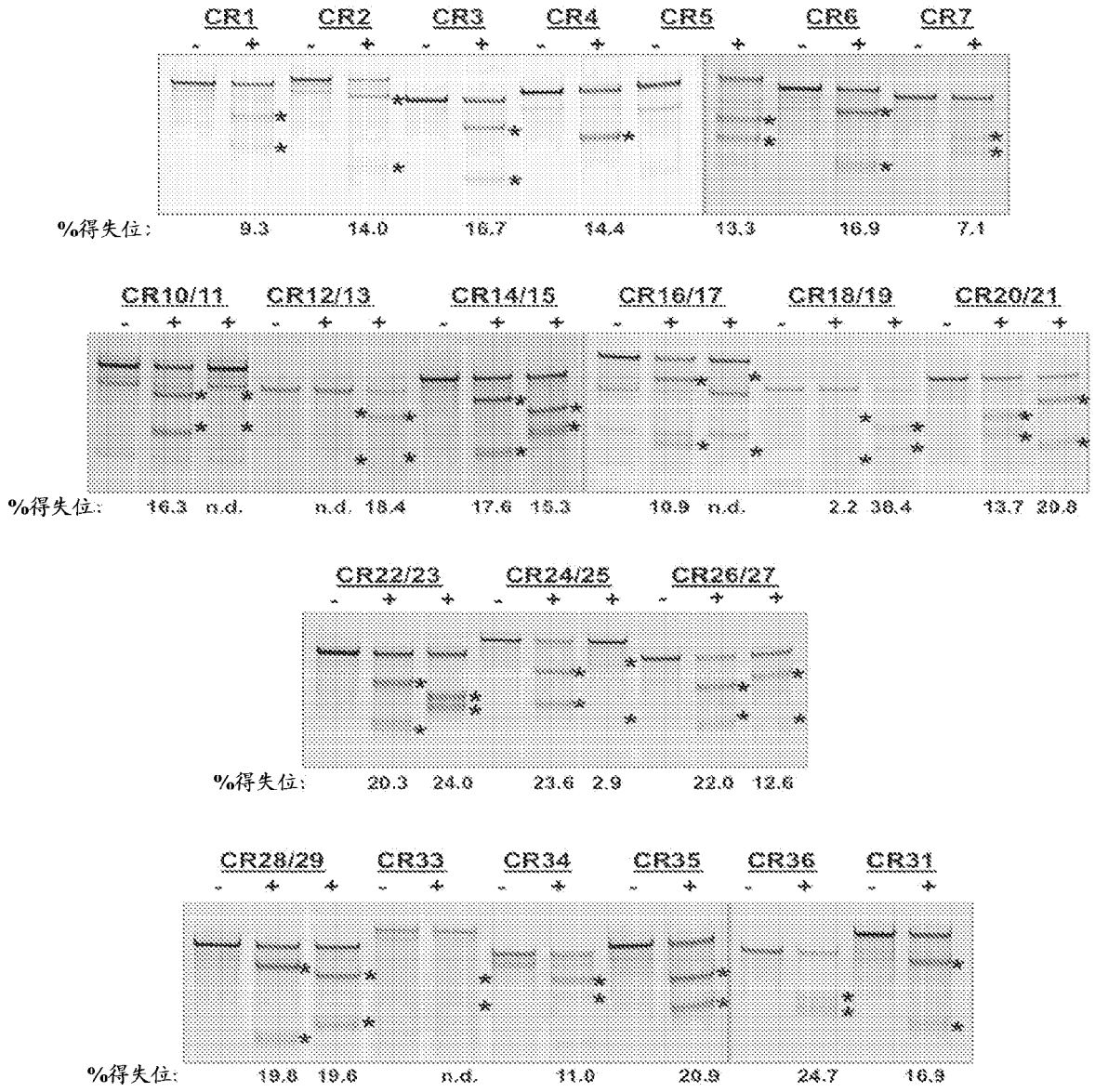

CR1
CR2
CR3
CR4
CR5
CR6
CR7
%得失位: 9.3 14.0 16.7 14.4 13.3 16.9 7.1
CR10/11
CR12/13
CR14/15
CR16/17
CR18/19
CR20/21
%得失位: 16.3 n.d. n.d. 18.4 17.6 15.3 18.9 n.d. 2.2 38.4 13.7 29.8
CR22/23
CR24/25
CR26/27
%得失位: 20.3 24.0 23.6 2.9 22.0 12.6
CR28/29
CR33
CR34
CR35
CR36
CR31
%得失位: 19.8 19.6 n.d. 11.0 20.9 24.7 16.9

图 18

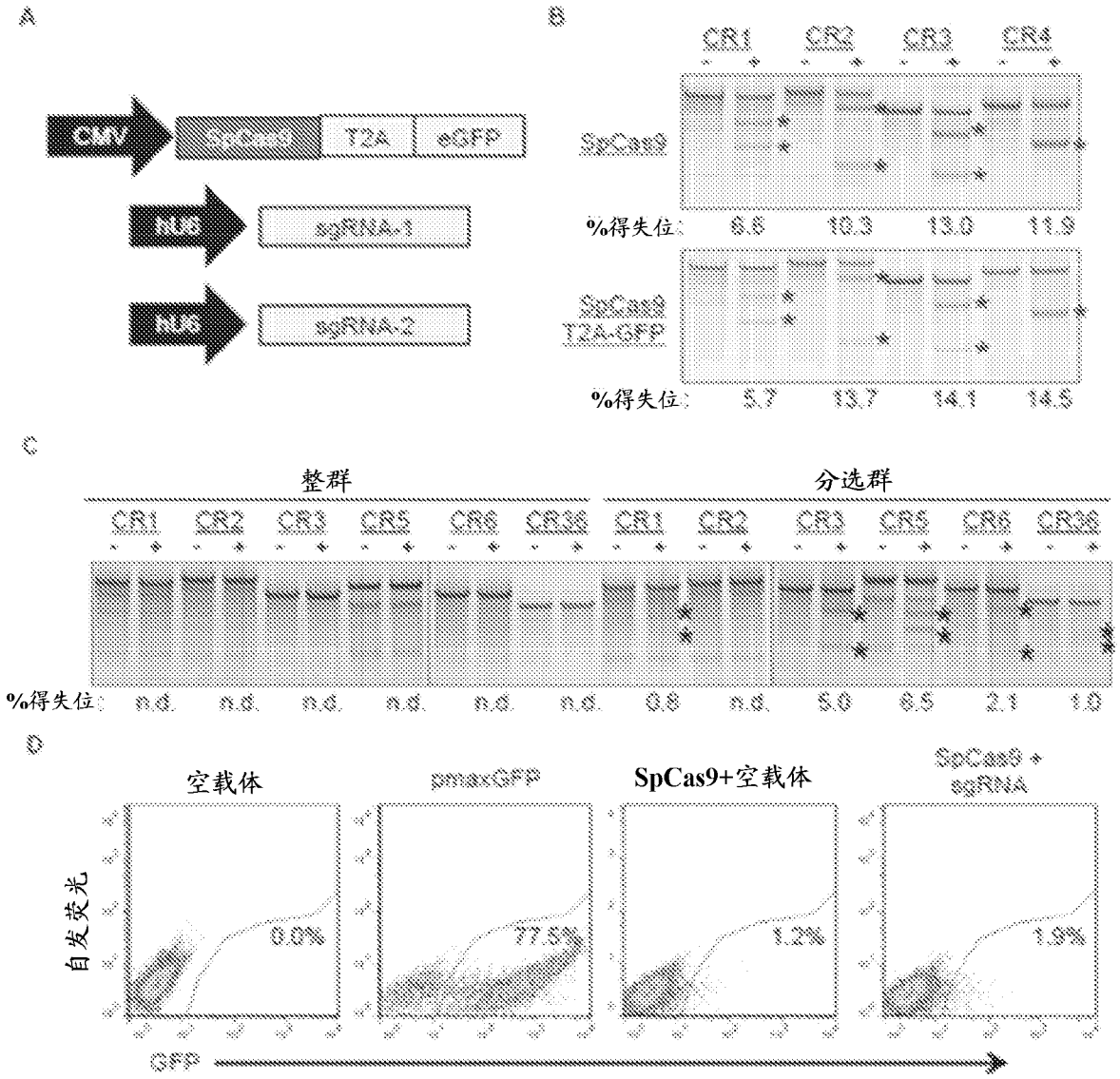

A
CMV
SpCas9
T2A
eGFP
U6
sgRNA-1
U6
sgRNA-2
B
CR1
CR2
CR3
CR4
SpCas9
%得失位:
6.6
10.3
13.0
11.9
SpCas9
T2A-GFP
%得失位:
5.7
13.7
14.1
14.5
C
整群
分选群
CR1
CR2
CR3
CR5
CR6
CR36
CR1
CR2
CR3
CR5
CR6
CR36
%得失位:
n.d.
n.d.
n.d.
n.d.
n.d.
n.d.
0.8
n.d.
5.0
8.5
2.1
1.0
D
空载体
pmaxGFP
SpCas9+空载体
SpCas9 + sgRNA
自发荧光
0.0%
77.5%
1.2%
1.9%
GFP

图 19

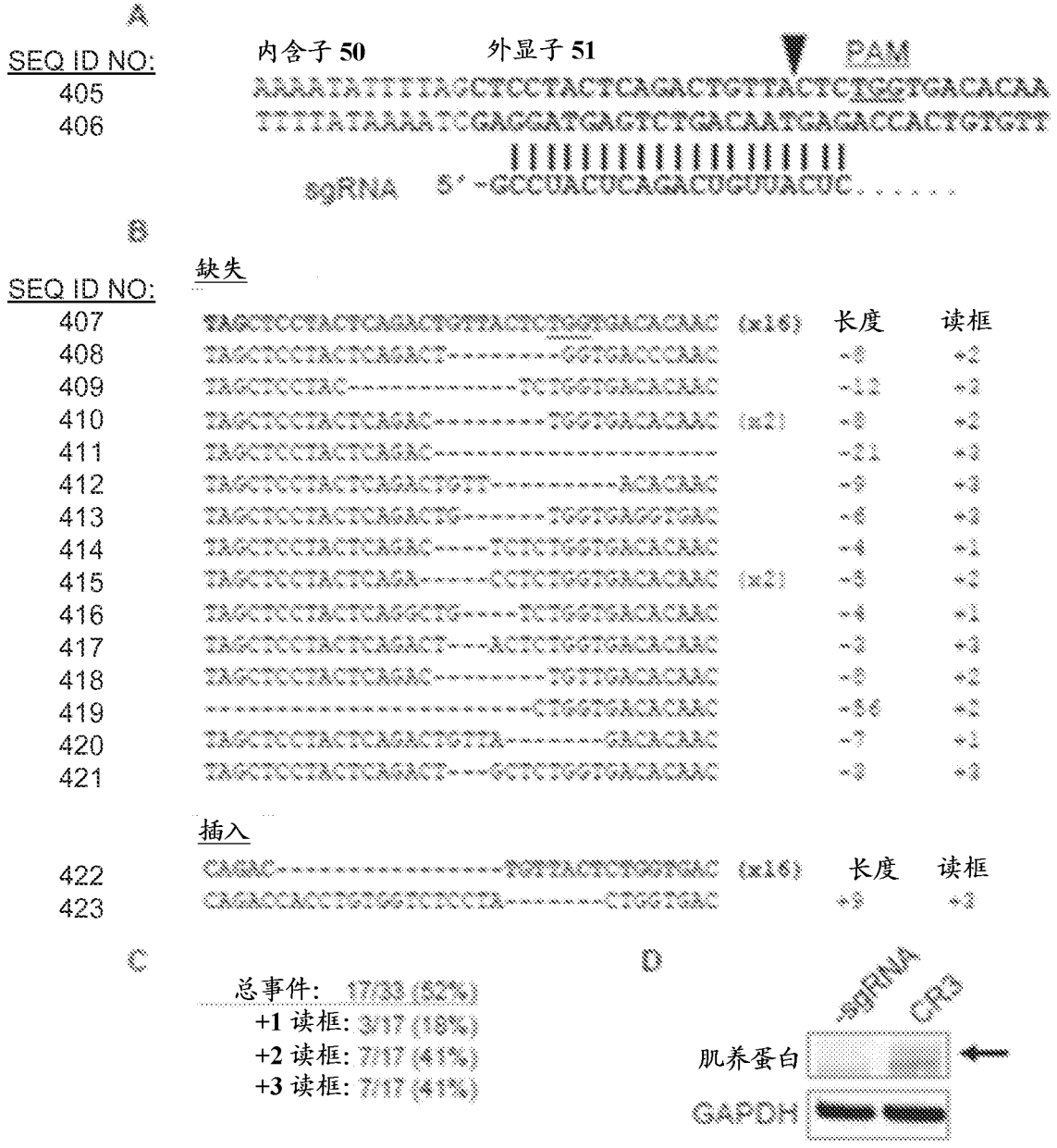

A
SEQ ID NO:
内含子 50
外显子 51
PAM
405
406
sgRNA
B
SEQ ID NO:
缺失
长度
读框
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
插入
422
423
C
总事件: 17/33 (52%)
+1 读框: 3/17 (18%)
+2 读框: 7/17 (41%)
+3 读框: 7/17 (41%)
D
sgRNA
CR3
肌养蛋白
GAPDH

图 20

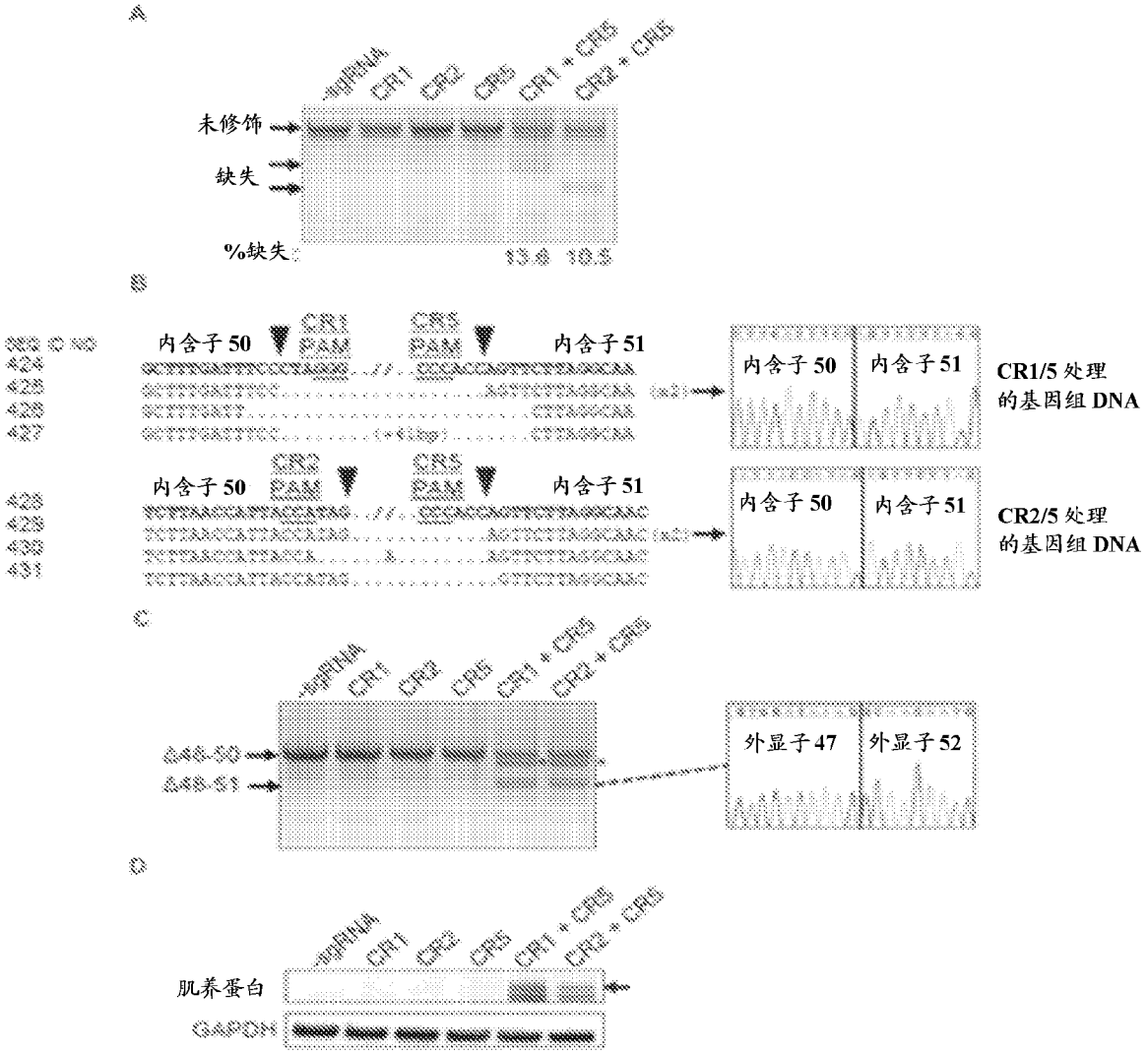

A
sgRNA
CR1
CR2
CR5
CR1 + CR5
CR2 + CR5
未修饰
缺失
%缺失
13.8
10.5
B
内含子 50
CR1
PAM
CR5
PAM
内含子 51
内含子 50
内含子 51
CR1/5 处理
的基因组 DNA
内含子 50
CR2
PAM
CR5
PAM
内含子 51
内含子 50
内含子 51
CR2/5 处理
的基因组 DNA
C
sgRNA
CR1
CR2
CR5
CR1 + CR5
CR2 + CR5
外显子 47
外显子 52
D
sgRNA
CR1
CR2
CR5
CR1 + CR5
CR2 + CR5
肌养蛋白
GAPDH

图 21

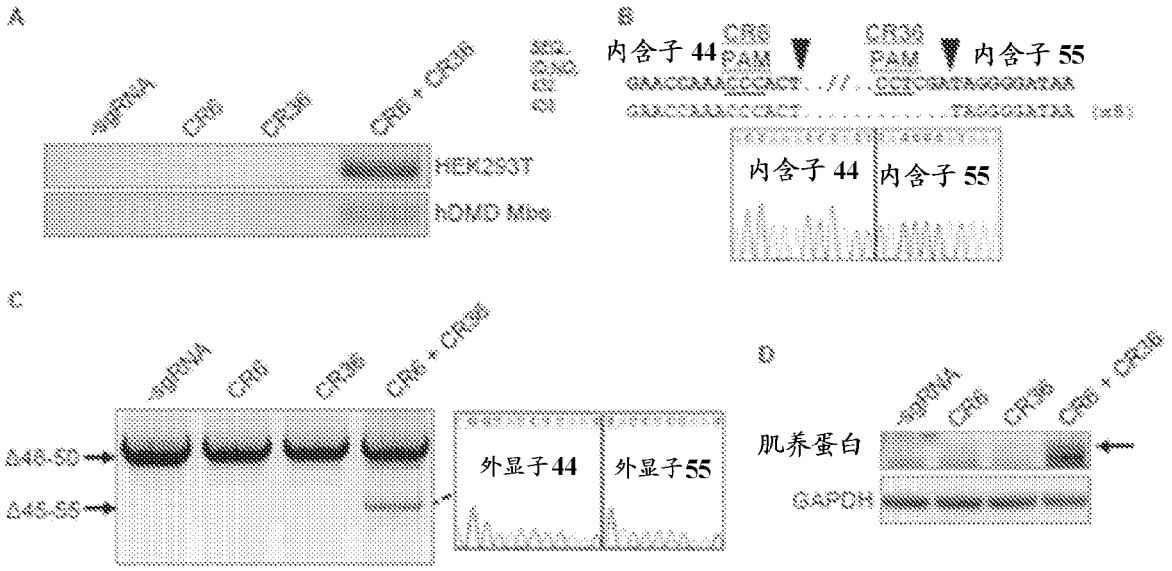

A
sgRNA
CR6
CR36
CR6 + CR36
HEK293T
NOMO Mks
B
内含子 44
内含子 55
CR6
PAM
CR36
PAM
内含子 44
内含子 55
C
sgRNA
CR6
CR36
CR6 + CR36
Δ48-SD
Δ48-SS
外显子44
外显子55
D
sgRNA
CR6
CR36
CR6 + CR36
肌养蛋白
GAPDH

图 22

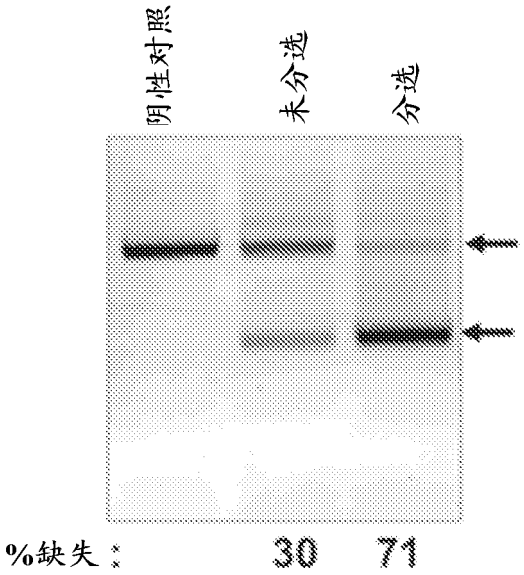

阴性对照
未分选
分选
%缺失：
30
71

图 23

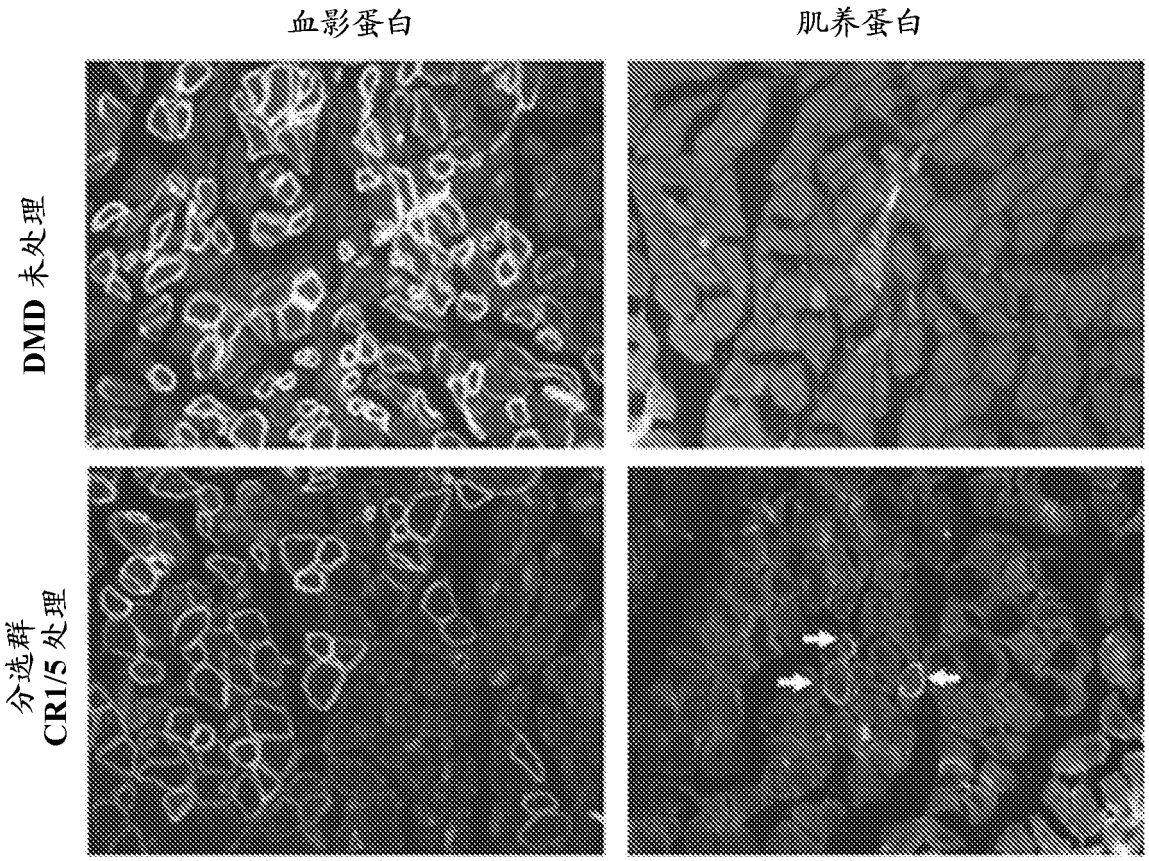


图 24

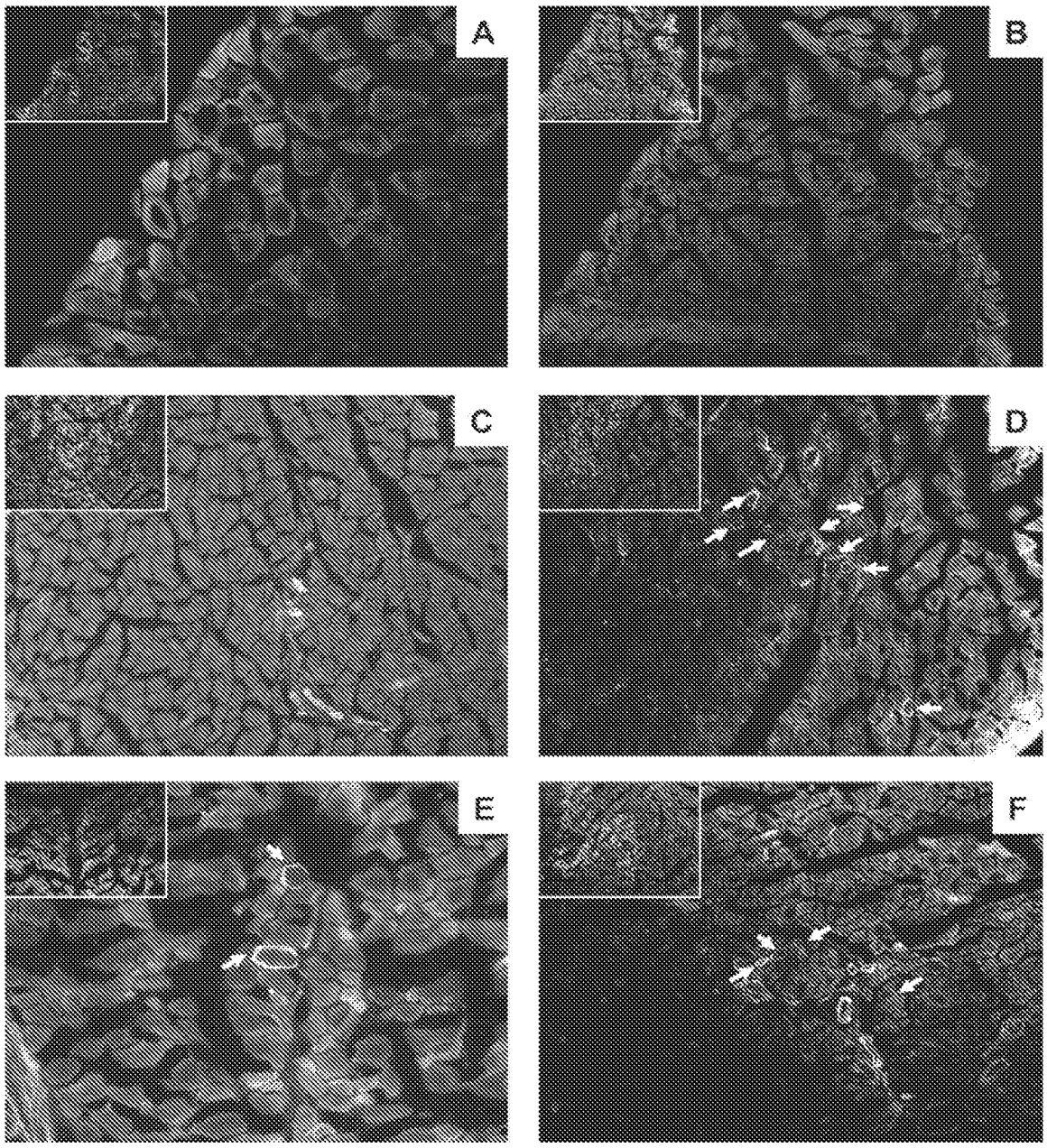


图 25

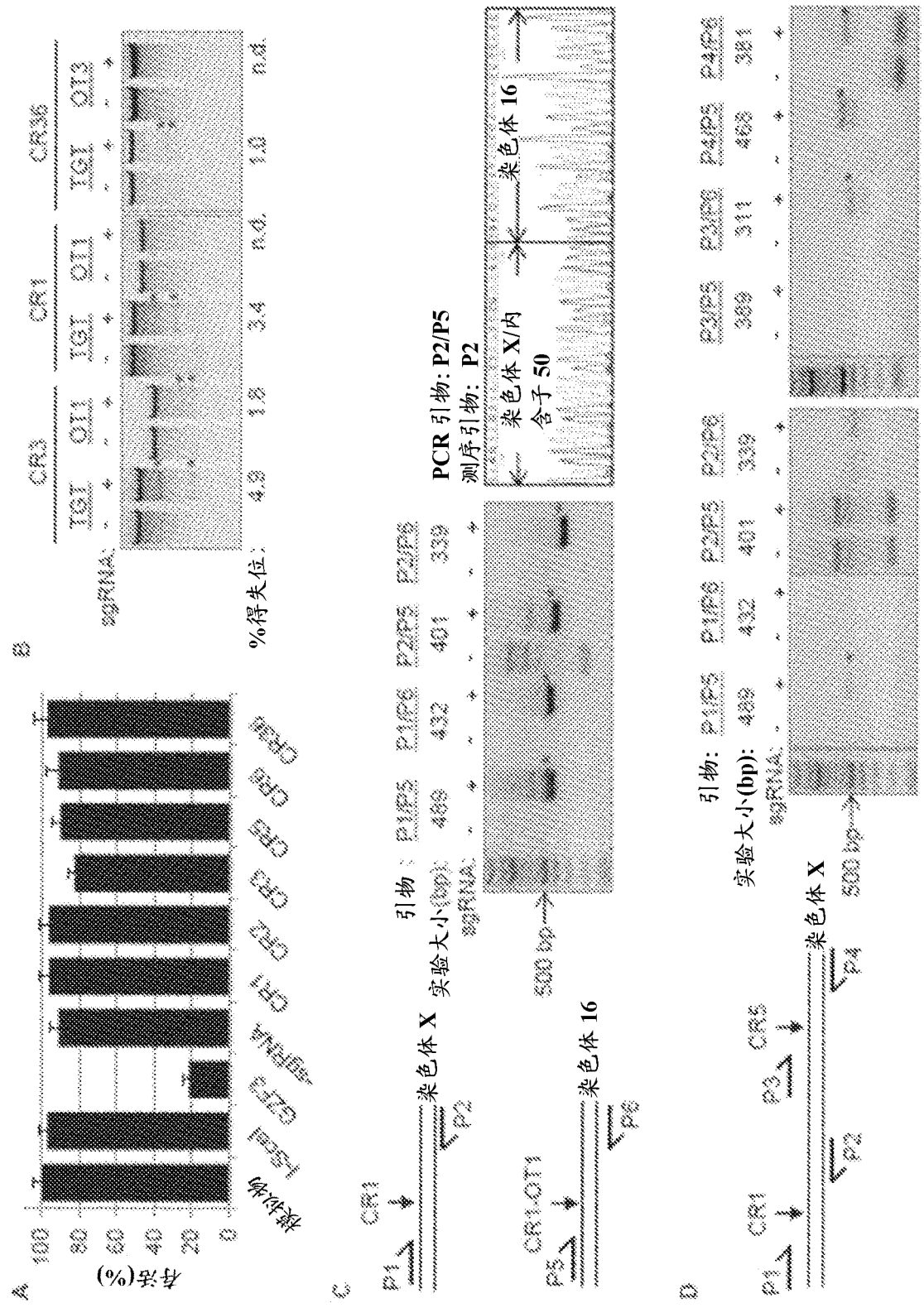

A
脱靶(%)
100
80
60
40
20
0
未处理
Cas9
Cas9 RNP
CR1
CR2
CR3
CR4
CR5
CR6
CR7
CR8
CR9
CR10
B
CR3
CR1
CR36
TGT
OT1
IGT
OT1
TGT
OT3
sgRNA:
%得失位:
4.3
1.8
3.4
n.d.
1.3
n.d.
C
CR1
P1
P2
染色体 X
引物:
P1P5
P1P6
P2P5
P2P6
实验大小(bp):
469
432
401
339
sgRNA:
500 bp
CR1-OT1
P5
P6
染色体 16
PCR 引物: P2/P5
测序引物: P2
染色体 X/内含子 50
染色体 16
D
CR1
P1
P2
CR5
P3
P4
染色体 X
引物:
P1P5
P1P6
P3P5
P2P6
P3P5
P3P6
P4P5
P4P6
实验大小(bp):
469
432
401
339
389
311
408
381
sgRNA:
500 bp

图 26

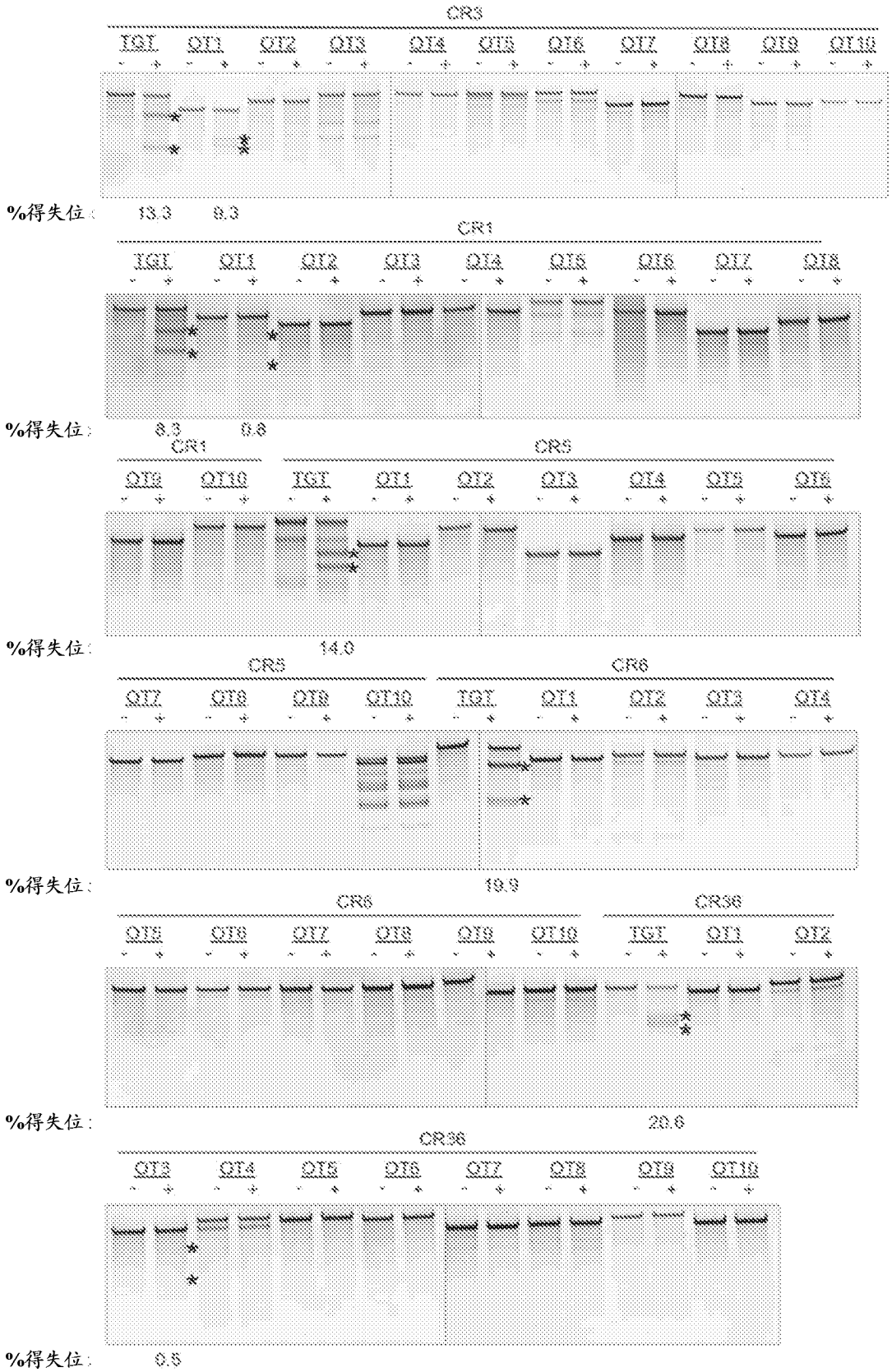

CR3
TGT
OT1
OT2
OT3
OT4
OT5
OT6
OT7
OT8
OT9
OT10
%得失位：
13.3
8.3
CR1
TGT
OT1
OT2
OT3
OT4
OT5
OT6
OT7
OT8
%得失位：
8.3
0.8
CR1
CR5
OT9
OT10
TGT
OT1
OT2
OT3
OT4
OT5
OT6
%得失位：
14.0
CR5
CR6
OT7
OT8
OT9
OT10
TGT
OT1
OT2
OT3
OT4
%得失位：
19.9
CR6
CR36
OT5
OT6
OT7
OT8
OT9
OT10
TGT
OT1
OT2
%得失位：
20.6
CR36
OT3
OT4
OT5
OT6
OT7
OT8
OT9
OT10
%得失位：
0.5

图 27

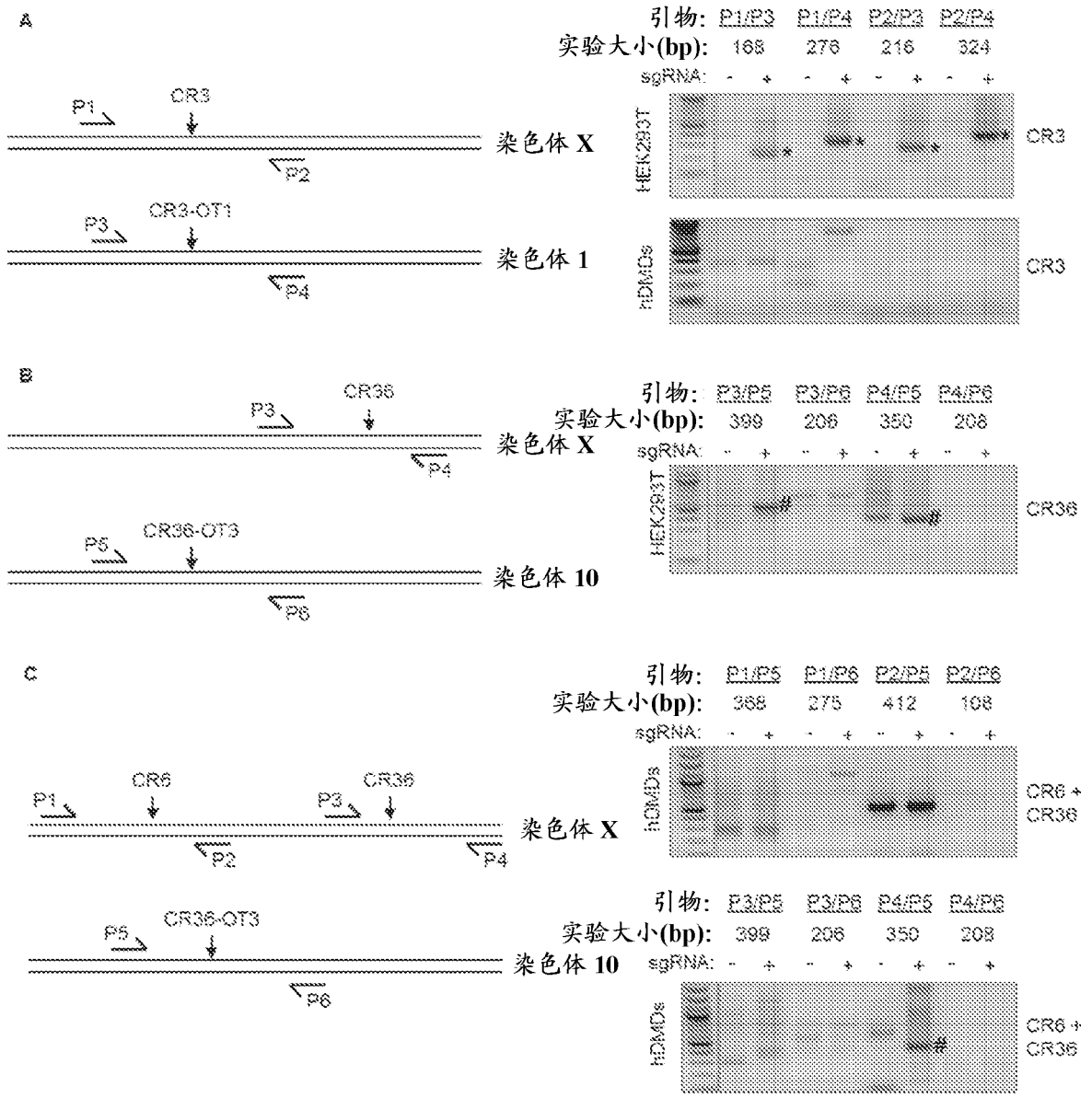

A
引物: P1/P3 P1/P4 P2/P3 P2/P4
实验大小(bp): 168 276 216 324
sgRNA: - + - + - + - +
P1
CR3
P2
染色体 X
P3
CR3-OT1
P4
染色体 1
HEK293T
hDMDs
CR3
CR3
B
引物: P3/P5 P3/P6 P4/P5 P4/P6
实验大小(bp): 399 206 350 208
sgRNA: - + - + - + - +
CR36
P3
P4
染色体 X
P5
CR36-OT3
P6
染色体 10
HEK293T
CR36
C
引物: P1/P5 P1/P6 P2/P5 P2/P6
实验大小(bp): 368 275 412 108
sgRNA: - + - + - + - +
P1
CR6
P3
CR36
P2
P4
染色体 X
hDMDs
CR6 + CR36
引物: P3/P5 P3/P6 P4/P5 P4/P6
实验大小(bp): 399 206 350 208
sgRNA: - + - + - + - +
P5
CR36-OT3
P6
染色体 10
hDMDs
CR6 + CR36

图 28

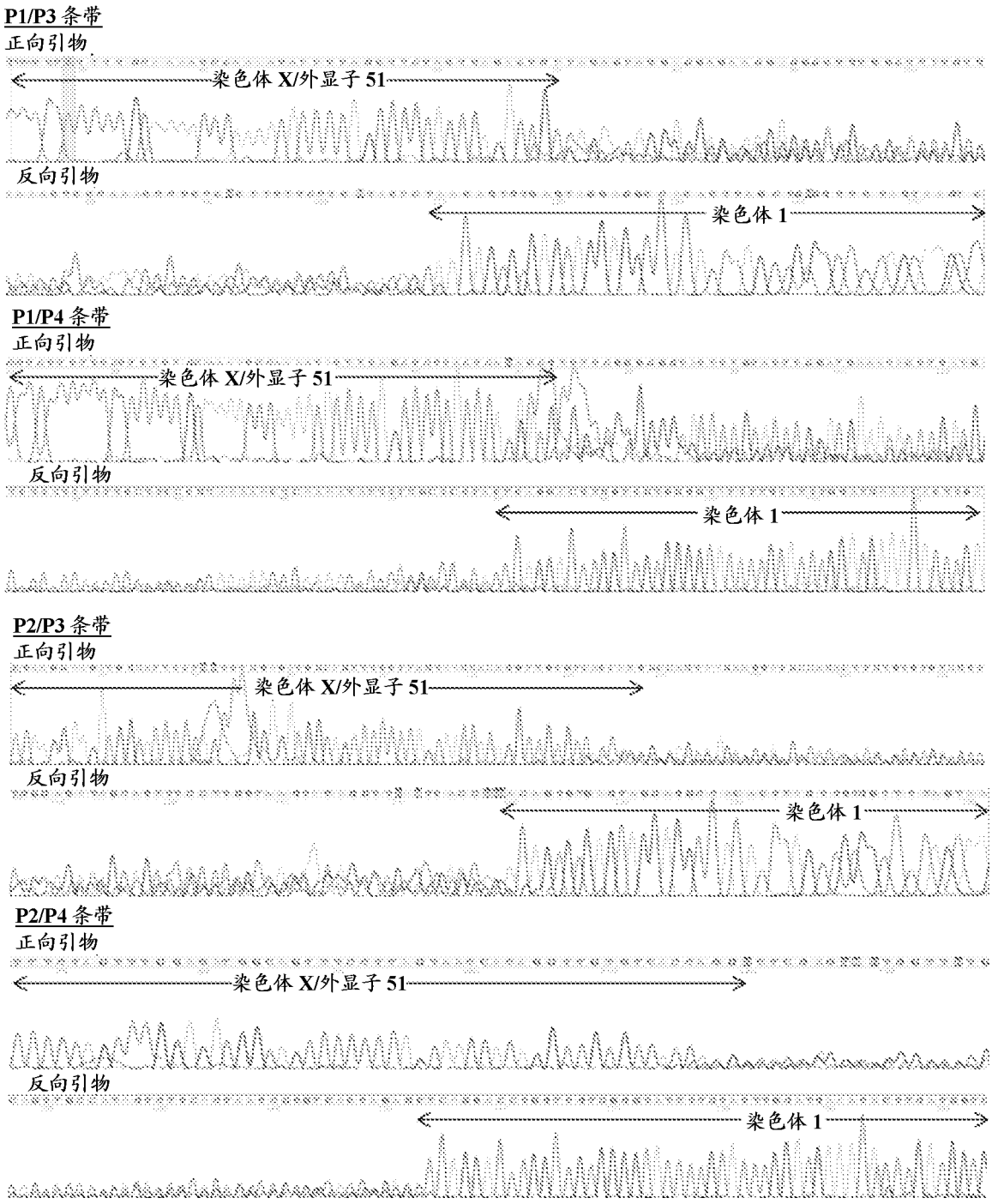

P1/P3 条带
正向引物
染色体 X/外显子 51
反向引物
染色体 1
P1/P4 条带
正向引物
染色体 X/外显子 51
反向引物
染色体 1
P2/P3 条带
正向引物
染色体 X/外显子 51
反向引物
染色体 1
P2/P4 条带
正向引物
染色体 X/外显子 51
反向引物
染色体 1

图 29

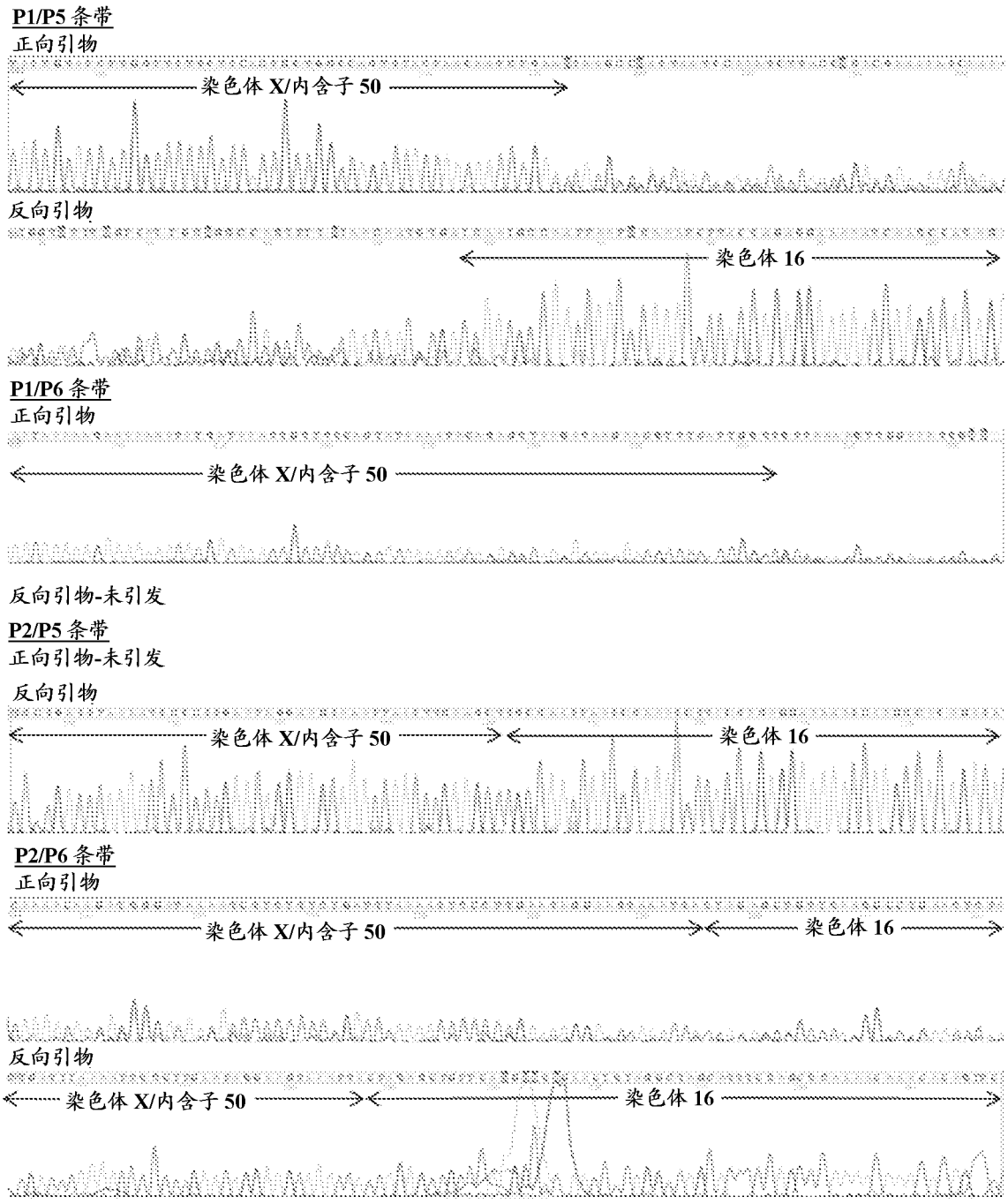

P1/P5 条带
正向引物
染色体 X/内含子 50
反向引物
染色体 16
P1/P6 条带
正向引物
染色体 X/内含子 50
反向引物-未引发
P2/P5 条带
正向引物-未引发
反向引物
染色体 X/内含子 50
染色体 16
P2/P6 条带
正向引物
染色体 X/内含子 50
染色体 16
反向引物
染色体 X/内含子 50
染色体 16

图 30

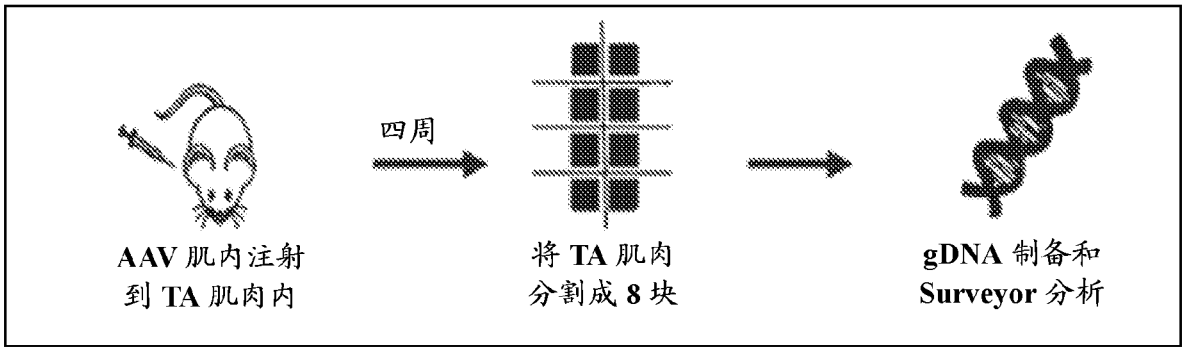


图 31

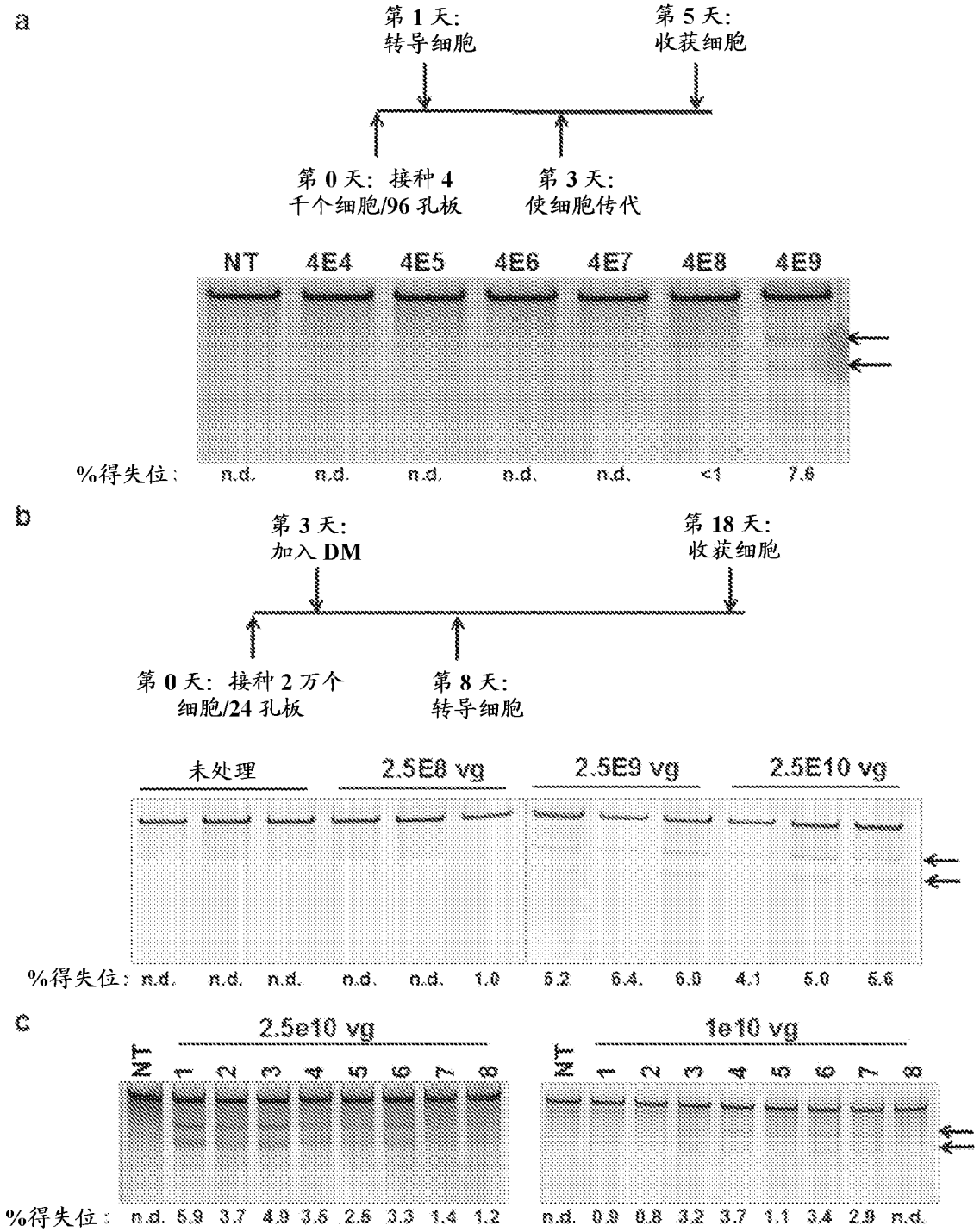

a
第 1 天:
转导细胞
第 5 天:
收获细胞
第 0 天: 接种 4
千个细胞/96 孔板
第 3 天:
使细胞传代
NT
4E4
4E5
4E6
4E7
4E8
4E9
%得失位:
n.d.
n.d.
n.d.
n.d.
n.d.
<1
7.6
b
第 3 天:
加入 DM
第 18 天:
收获细胞
第 0 天: 接种 2 万个
细胞/24 孔板
第 8 天:
转导细胞
未处理
2.5E8 vg
2.5E9 vg
2.5E10 vg
%得失位:
n.d.
n.d.
n.d.
n.d.
n.d.
1.0
6.2
6.4
6.0
4.1
5.0
5.6
c
2.5e10 vg
1e10 vg
NT
1
2
3
4
5
6
7
8
%得失位:
n.d.
5.9
3.7
4.9
3.6
2.5
3.3
1.4
1.2
n.d.
0.9
0.8
3.2
3.7
1.1
3.4
2.9
n.d.

图 32

Rosa T2A opt DNA 序列(SEQ ID NO: 434)

ATGAGGTCTGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGAT
GGCCCCCAAGAAGAAGaggaaggtgggcctcgaGCCCGGAGAAAAACCGTACAAGTGCCCTGAGTGCGGGAAATCAT
TCTCCGACCCTGGGGCGCTCGTCCGGCACCAAAGGACGCATACAGGGGAAAAGCCGTATAAGTGCCCCGAGTGTGGA
AAGAGCTTCTCGCAGAGAGCCCACCTTGAACGACACCAAAGAACACACACTGGTGAGAAACCCTATAAGTGTCCAGA
GTGCGGCAAATCGTTTAGCAGATCCGATGACTTGGTGCGCCACCAGCGGACACACACGGGTGAAAAGCCCTACAAAT
GCCCGGAGTGTGGGAAGTCGTTTTCAAGGTCGGATCATCTGACTACCCATCAGCGCACCCATACGGGAGCggcgccc
cgcgccctGGTGAAGAGCGGAGCTGGAGGAGAAGAAGTCCGAGCTGCGGCACAAGCTGAAGTACGTGCCCCACGAGTA
CATCGAGCTGATCGAGATCGCCAGGAACCCCACCCAGGACCGCATCCTGGAGATGAAGGTGATGGAGTTCTTCATGA
AGGTCTACGGCTACAGGGGCAGCCACCTGGGCGGAAGCACAAAGCCTGACGCCGCCATCTATACAGTGGGCAGCCCC
ATCGATTACGGCGTGATCGTGGACACAAAGGCCTACAGCGGCGGCTACAATCTGCCTATCGGCCAGGCCGACGAGAT
GCAGAGATACGTGGAAGGAGAACCAGACCCGGAATAAGCACATCAACCCCAACGAGTGGTGGAAGGTGTACCCTAGCA
GCGTGACCGAGTTCAAGTTCCTGTTCGTGAGCGGCCACTTCAAGGGCAACTACAAGGCCCAGCTGACCAGGCTGAAC
CGCAAAACCAACTGCAATGGCGCCGTGCTGAGCGTGGAGGAGCTGCTGATCGGCGGCGAGATGATCAAAGCCGGCAC
CCTGACACTGGAGGAGGTGCGGCGCAAGTTCAACAACGGCGAGATCAACTTCGAGGGCAGAGGAAGTCTTCTAACAT
GCGGTGACGTGGAGGAGAATCCCGGCCCTAgatcTGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATC
GATTACAAGGATGACGATGACAAGATGGCCCCCAAGAAGAAGaggaaggtgggcctcgagccGGGAGAGAAGCCGTA
CAAGTGTCCCGAATGTGGAAAGAGCTTCTCACAGTCGGGGGACCTTCGGCGCCACCAGCGCACACATACTGGTGAAA
AGCCGTATAAGTGTCCAGAATGTGGCAAATCATTCTCCACATCAGGGAGCCTGGTCAGGCACCAGCGAACCCACACG
GGTGAGAAGCCCTATAAGTGCCCCGAATGCGGGAAGTCCTTTTCGCAGAGAGCCCACTTGGAGAGGCACCAGAGGAC
CCATACGGGGGAGAAACCTTACAAGTGCCCTGAATGCGGAAAGTCGTTCTCGACCCATCTGGATCTCATCAGACATC
AGAGAACGCACACTGGAGAGAAACCCTACAAATGTCCCGAGTGTGGGAAGTCGTTTAGCCGAAAGGACAATCTCAAA
AACCATCAACGGACACACAGGGGTGAAAAACCATACAAATGCCCGGAGTGCGGCAAATCGTTTTCCCAACTTGCGCA
CTTGCGGGGCACACCAACGCACGCATACTGGAGCGGCCGCcgcgccCTGGTGAAGAGCGGAGCTGGAGGAGAAGAAGT
CCGAGCTGCGGCACAAGCTGAAGTACGTGCCCCACGAGTACATCGAGCTGATCGAGATCGCCAGGAACCCCACCCAG
GACCGCATCCTGGAGATGAAGGTGATGGAGTTCTTCATGAAGGTCTACGGCTACAGGGGCAGAGCCACCTGGGCGGAAG
CAGAAAGCCTGACGCCGCCATCTATACAGTGGGCAGCCCCATCGATTACGGCGTGATCGTGGACACAAAGGCCTACA
GCGGCGGCTACAATCTGCCTATCGGCCAGGCCGACGAGATGGAGAGATACGTGGAGGAGAACCAGACACGGGATAAG
CACCTCAACCCCAACGAGTGGTGGAAGGTGTACCCTAGCAGCGTGACCGAGTTCAAGTTCCTGTTCGTGAGCGGCCA
CTTCAAGGGCAACTACAAGGCCCAGCTGACCAGGCTGAACCACATCACCAACTGCAATGGCGCCGTGCTGAGCGTGG
AGGAGCTGCTGATCGGCGGCGAGATGATCAAAGCCGGCACCCTGACACTGGAGGAGGTGCGGCGCAAGTTCAACAAC
GGCGAGATCAACTTCTGA

Rosa T2A opt 蛋白质序列 (SEQ ID NO: 435)

MRSDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGLEPGEKPYKCPECGKSFSDPGALVPHQRTHTGEKPYKCPECG
KSFSQRARLEPHQRTHTGEKPYKCPECGKSFSRSDDLVRHQRTHTGEKPYKCPECGKSFSRSDHLTTHQRTHTGAAA
PALVKSELEEKKSELPHFLKYVPHEYIELIEIARNPTQDRILEMKVMEFFMKVYGYRGEHLGGSRKPDGAIYTVGSP
IDYGVIVDTKAYSGGYNLPIGQADEMQRYVKENQTRNKHINPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLN
HITNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINF

图 33

SASTG 衣壳序列 (SEQ ID NO: 436)

tggccactccctctatgcgcactcgctcgctcggtgggcctggcgaccaaaggtcgccagacggacgtgctttgca
cgtccggccccaccgagcgagcgagtgcgcatagagggagtggccaactccatcactagaggtatggcagtgacgta
acgcgaagcgcgcgaagcgagaccacgcctaccagctgcgtcagcagtcaggtgacccttttgcgacagtttgcgac
accacgtggccgctgagggtatatattctcgagtgagcgaaccaggagctccatttttgaccgcgaaatttgaacgag
cagcagccatgccgggggtttctacgagattgtcctgaaggtcccgagtgacctggacgagcaccctgccgggcatttct
aactcgtttgtttaactgggtggccgagaaggaatgggagctgccgccggattctgacatggatccgaatctgattga
gcaggcaccccctgaccgtggccgaaaagcttcagcgcgagtttcctggtggagtgycgccgcgtgagtaaggccccgg
aggccctctttttttgtccagttcgaaaagcggggagacctacttccacctgcacgtgcigattgagaccatcgggctc
aaatccatggtggtcggccgctacgtgagccagatttaaagagaagctggtgacccgcatctaccgcggggtcgagcc
gcagcttccgaactggttcgcggtgaccaaaacycgaaatggccgccgggggcgggaacaaggtggtggacgactgct
acatccccaactacctgctccccaagaccagccgagctccagtgggcgtggactaacatggaccagtatttaagc
gcctgtttgaatctcgcggagcgtaaacggctggtggcgcagcatctgacgcacgtgtcgcagacgcaggagcagaa
caaagagaatcagaaccccaattctgacgcgccggtcatcaggtcaaaaacctcagccaggtacatggagctggtcg
ggtggctggtggaccgcgggatcacgtcagaaaagcaatggattcaggaggaccaggcctcgtacatctccttcaac
gcgcgctccaactcgcggtcccagatcaaggccgcgctggacaatgcctccaagatcatgagcctgacaaagacggc
tccggactacctggtgggcagcaacccgccggaggacattaccaaaaatcggatctaccaaatcctggagctgaacg
ggtacgatccgcagtacgcggcctccgtcttcctgggctgggcgcaaaagaagttcgggaagaggaacaccatctgg
ctctttgggccggccacgacgggtaaaaccaacatcgcggaagccatcgcccacgccgtgccctctctacggctgcgt
aaactggaccaatgagaactttccccttcaacgattgcgtcgacaagatggtgatctggtgggaggagggcaagatga
cggccaaggtcgtggagagcgccaaggccattctgggcggaagcaaggtgcgcgtggaccaaaagtgcaagtcatcg
gcccagatcgaacccactcccgtgatcgtcacctccaacaccaacatgtgcgccgtgattgacgggaacagcaccac
cttcgagcatcagcagccgctgcaggaccggatgtttaaatttgaacttacccgccgtttggaccatgactttgggga
aggtcaccaaacaggaagtaaaggacttttccggtgggcttccgatcacgtgactgacgtggctcatgagttctac
gtcagaaagggtggagctaagaaacgcccgcctccaatgacgcggatgtaagcgagccaaaacggcagtgcacgtc
acttgcgcagccgacaacgtcagacgcggaagcaccggcggactacgcggacaggtaccaaaacaaatgttctcgtc
acgtgggcatgaatctgatgctttttccctgtaaaacatgcgagagaatgaatcaaatttccaatgtctgttttacg
catggtcaaagagaactgtggggaatgcttccctggaatgtcagaatctcaacccgtttctgtcgtcaaaaagaagac
ttatcagaaactgtgtccaattcatcatatcctgggaagggcaccgagattgcctgttcggcctgcgatttggcca
atgtggacttggatgactgtgtttctgagcaataaatgacttaaaccaggtatggctgctgacggttatcttccaga
ttggctcgaggacaaccttctgaaggcattcgtgagtggtgggctctgaaacctggagtcccctcaacccaaagcga
accaacaacaccagggacaaccgtcgggtcttgtgcttccgggttacaaatacctggaccccggtaacggactcgac
aaaggagagccggtcaacgaggcggacgcggcagccctcgaacacgacaaagcttacgaccagcagctcaaggccgg
tgacaacccgtacctcaagtacaaccacgccgacgccgagtttcaggagcgtcttcaagaagatacgtcttttgggg
gcaacttggcagagcagtcttccagggccaaaaagaggatccttgagcctcttggtctggttgaggaagcagctaaa
acggctcctggaaagaagaggcctgtagatcagtctcctcaggaaccggactcatcatctggtgttggcaaatcggg
caaacagcctgccagaaaaagactaaatttcggtcagactggcgactcagagtcagtcccagaccctcaacctctcg
gagaaccaccagcagccccccacaagtttgtggatctaatacaatggcttcaggcggtggcgcaccaatggcagacaat
aacgagggtgccgatggagtgggtaattcctcaggaaattggcattgcgattcccaatggctgggcgacagagtcat
caccaccagcaccagaacctgggccctgcccacttacaacaaccatctctacaagcaaatctccagcGCTtcaACGg
gagcttcaaacgacaaccactactttggctacagcacccttgggggtattttgactttaacagattccactgccac
ttctcaccacgtgactggcagcgactcattaacaacaactgggattccggcccaagaaactcagcttcaagctctt
caacatccaagttaaagaggtcacgcagaacgatggcacgacgactattgccaataaccttaccagcacggttcaag
tgtttacggactcggagtatcagctcccgtacgtgctcgggtcggcgcaccaaggctgtctcccgccgtttccagcg
gacgtcttcatggtccctcagtatggatacctcaccctgaacaacggaagtcaagcggtgggacgctcatcctttta
ctgcctggagtacttcccttcgcagatgctaaggactggaaataacttccaattcagctatacctttcgaggatgtac
cttttcacagcagctacgctcacagccagagtttggatcgcttgatgaatcctcttattgatcagtatctgtactac
ctgaacagaacgcaaggaacaacctctggaacaacaacaaccaatcacggctgctttttagccaggctgggcctcagtc
tatgtctttgcaggccagaaattggctacctgggccctgctaccggcaacagagactttcaaagactgctaacgaca
acaacaacagtaactttcccttggacagcggccagcaaatatcatctcaatggccgcgactcgctggtgaatccagga
ccagctatggccagtcacaaggacgatgaagaaaaatttttcccctatgcacggcaatctaatatttggcaaagaagg
gacaacggcaagtaacgcagaattagataatgtaatgattacggatgaagaagagattcgtaccaccaatcctgtgg
caacagagcagtatggaactgtggcaaataacttgcagagctcaaatacagctcccacgactagaactgtcaatgat
caggggcctacctggcatggtgtggcaagatcgtgacgtgtaccttcaaggacctatctggycaaagattcctca
cacggatggacactttcatccttctcctctgatgggaggctttggactgaaacatccgcctcctcaaatcatgatca
aaaatactccggtaccggcaaatcctccgacgactttcagcccggccaagtttgcttcatttatcactcagtactcc

图 34

actggacaggtcagcgtggaaattgagtgggagctacagaaagaaaacagcaaacgttggaatccagagattcagta
cacttccaactacaacaagtctgttaatgtggactttactgtagacactaatggtgtttatagtgaacctcgccctа
ttggaacccggtatctcacacgaaacttgtaatcctggttaatcaataaaccgtttaattcgtttcagttgaacttt
ggctcttgtgcacttcttatcttatcttgtttccatggctactgcgtagataagcagcggcctgcggcgcttgcgct
tcgcggtttacaacigctggttaatatttaactctcgccatacctctagtgatggagttggcccactcctctatgcg
cactcgctcgctcggtggggcccggacgtgcaaagcacgtccgtctggcgacctttggtcgccaggccccaccgagcg
agcgagtgcgcatagagggagtggccaa

SASTG 衣壳肽序列 (SEQ ID NO: 437)

MAADGYLPDWLEDNLSEGIREWWALKPGVPQPKANQQHQDNRRGLVLPGYKYLGPGNGLDKGEPVNEADAAALEHDK
AYDQQLKAGDNPYLKYNHADAEFQERLQEDTSFGGNLGRAVFQAKKRILEPLGLVEEPAAKTAPGKKRPVDQSPQEPD
SSSGVGKSGKQPARKRLNFGQTGDSESVPDPQPLGEPPAAPTSLGSNTMASGGGAPMADNNEGADGVGNSSGNWHCD
SQWLGDRVITTSTRTWALPTYNNHLYKQISSASTGASNDNHYFGYSTPWGYFDFNRFHCHFSPRDWQRLINNNWGFR
PKRLSFKLFNIQVKEVTQNDGTTTIANNLTSTVQVFTDSEYQLPYVLGSAHQGCLPPFPADVFMVPQYGYLTLNNGS
QAVGRSSFYCLEYFPSQMLRTGNNFQFSYTFEDVPFHSSYAHSQSLDRLMNPLIDQYLYYLNRTQGTTSGTTNQSRL
LFSQAGPQSMSLQARNWLPGPCYRQQRLSKTANDNNNSNFPWTAASKYHLNGRDSLVNPGPAMASHKDDEEKFFPMH
GNLIFGKEGTTASNAELDNVMITDEEEIRTTNPVATEQYGTVANNLQSSNTAPTTRTVNDQGALPGMVWQDRDVYLQ
GPIWAKIPHTDGHFHPSPLMGGFGLKHPPPQIMIKNTPVPANPPTTFSPAKFASFITQYSTGQVSVEIEWELQKENS
KRWNPEIQYTSNYNKSVNVDFTVDTNGVYSEPRPIGTRYLTRNL

图 34(续)

DZF16 ZFN 靶位点 (SEQ ID NO:442):

5'-CAAACTAGAAATGCCATCTTCCTTGATGTTGGACGTACCTGC-3'

3'-GTTTGATCTTTACGGTAGAAGGAACTACAACCTGCATGGACG-5'

DZF16-L6 左侧全部氨基酸序列(SEQ ID NO:443)

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGRLEPGEKPYKCPECGKSFSPKDALRGHQRTHTGEKPY
KCPECGKSFSHPTTLTNHQPTHTGEKPYKCPECGKSFSQRNALAGHQRTHTGEKPYKCPECGKSFSHKNA
LQNHQRTHTGEKPYKCPECGKSFSDPGHLVRHQPTHTGEKPYKCPECGKSFSTSGNLVRHQPTHTGAAAP
ALVKSELEEKKSELRHKLKYVPHEYIELIEIARNPTQDRILEMKVMEFFMKVYGYRGEHLGGSRKPDGAI
YTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMQRYVEENQTRNKHINPNEWWKVYPSSVTEFKFLFVSGH
FKGNYKAQLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINF*

DZF16-R6 右侧全部氨基酸序列(SEQ ID NO:444)

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGRLEPGEKPYKCPECGKSFSQQRSLVGHQRTHTGEKPY
KCPECGKSFSDKKDLTRHQRTHTGEKPYKCPECGKSFSTSGHLVRHQRTHTGEKPYECPECGKSFSQRAH
LERHQRTHTGEKPYKCPECGKSFSTSGSLVRHQRTHTGEKPYKCPECGKSFSTSGNLVRHQRTHTGAAAR
ALVKSELEEKKSELRHKLKYVPHEYIELIEIARNPTQDRILEMKVMEFFMKVYGYRGEHLGGSRKPDGAI
YTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMQRYVKENQTRNKHINPNEWWKVYPSSVTEFKFLFVSGH
FKGNYKAQLTRLNRKTNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINF*

图 35

E51C3 靶位点 (SEQ ID NO:445):

5'-(t)ATCTGCCCATGACTGGCGCAGGG(a)-3'

3'-(a)TAGACGGGTACTGACCGCGTCCC(t)-5'

E51C-3L 左侧全部氨基酸序列 (SEQ ID NO:446)

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGRLEPGEPPFQCRICMRNFSSFQALAVHTPTHTGEKPF
QCRICMFNFSQSTTLKRHLRTHTGEKPFQCRICMRNFSRSDHLSLHLKTHLPGSQLVKSELEEKKSELRH
KLKYVPHEYIELIEIARNPTQDRILEMKVMEFFMKVYGYRGEHLGGSRKPDGAIYTVGSPIDYGVIVDTK
AYSGGYNLPIGQADEMQRYVKENQTRNKHINPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNHKT
NCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINF*

E51C-3R 右侧全部氨基酸序列 (SEQ ID NO:447)

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGRLEPGEPPFQCRICMRNFSRRAHLQNHTPTHTGEKPF
QCRICMFNFSQSTTLKRHLRTHTGEKPFQCRICMRNFSDGGHLTPHLKTHLPGSQLVKSELEEKKSELRH
KLKYVPHEYIELIEIARNPTQDRILEMKVMEFFMKVYGYRGEHLGGSRKPDGAIYTVGSPIDYGVIVDTK
AYSGGYNLPIGQADEMERYVEENQTRDKHLNPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNHIT
NCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINF*

图 36

DZF15 靶位点 (SEQ ID NO:448):

5'-ACTAGAAATGCCATCTTCCTTGATGTTGGAGGTACCTGCTCT-3'
3'-TGATCTTTACGGTAGAAGGAACTACAACCTCCATGGACGAGA-5'

DZF15-L6左侧全部氨基酸序列(SEQ ID NO:449)

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGRLEPGEKPYKCPECGKSFSHRTTLTNHQRTHTGEKPY
KCPECGKSFSQKNALAGHQRTHTGEKPYKCPECGKSFSHKNALQNHQRTHTGEKPYKCPECGKSFSDPGH
LVRHQRTHTGEKPYKCPECGKSFSTSGNLVRHQRTHTGEKPYKCPECGKSFSQSSNLVRHQRTHTGAAAR
ALVKSELEEKKSELRHKLKYVPHEYIELIEIARNPTQDRILEMKVMEFFMKVYGYRGEHLGGSRKPDGAI
YTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMERYVEENQTRDKHLNPNEWWKVYPSSVTEFKFLFVSGH
FKGNYKAQLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINF*

DZF15-R6 右侧全部氨基酸序列 (SEQ ID NO:450)

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGRLEPGEKPYKCPECGKSFSQRNALAGHQRTHTGEKPY
KCPECGKSFSQQRSLVGHQRTHTGEKPYKCPECGKSFSDKKDLTRHQRTHTGEKPYKCPECGKSFSTSGH
LVRHQRTHTGEKPYKCPECGKSFSQRAHLERHQRTHTGEKPYKCPECGKSFSTSGSLVRHQRTHTGAAAR
ALVKSELEEKKSELRHKLKYVPHEYIELIEIARNPTQDRILEMKVMEFFMKVYGYRGEHLGGSRKPDGAI
YTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMERYVKENQTRNKHINPNEWWKVYPSSVTEFKFLFVSGH
FKGNYKAQLTRLNRKTNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINF*

DZF15-L5 左侧全部氨基酸序列 (SEQ ID NO:451)

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGRLEPGEKPYKCPECGKSFSQRNALAGHQRTHTGEKPY
KCPECGKSFSHKNALQNHQRTHTGEKPYKCPECGKSFSDPGHLVRHQRTHTGEKPYKCPECGKSFSTSGN
LVRHQRTHTGEKPYKCPECGKSFSQSSNLVRHQRTHTGAAARALVKSELEEKKSELRHKLKYVPHEYIEL
IEIARNPTQDRILEMKVMEFFMKVYGYRGEHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQ
ADEMERYVEENQTRDKHLNPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNHITNCNGAVLSVEEL
LIGGEMIKAGTLTLEEVRRKFNNGEINF*

DZF15-R5 右侧全部氨基酸序列 (SEQ ID NO:452)

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGRLEPGEKPYKCPECGKSFSQQRSLVGHQRTHTGEKPY
KCPECGKSFSDKKDLTRHQRTHTGEKPYKCPECGKSFSTSGHLVRHQRTHTGEKPYKCPECGKSFSQRAH
LERHQRTHTGEKPYKCPECGKSFSTSGSLVRHQRTHTGAAARALVKSELEEKKSELRHKLKYVPHEYIEL
IEIARNPTQDRILEMKVMEFFMKVYGYRGEHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQ
ADEMQRYVKENQTRNKHINPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNRKTNCNGAVLSVEEL
LIGGEMIKAGTLTLEEVRRKFNNGEINF*

图 37

E51C4 靶位点 (SEQ ID NO:453):

5′-(t)GCCATCTTCCTTGATGTTGGAGGT(a)-3′

3′-(a)CGGTAGAAGGAACTACAACCTCCA(t)-5′

E51C-4L 左侧全部氨基酸序列 (SEQ ID NO:454)

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGRLEPGERPFQCRICMRNFSSPSKLAPHTRTHTGEKPF
QCRICMRNFSVPHNLTRGLKTHTGEKPFQCRICMRNFSQPNNLGRGLKTHTGAAAPALVKSELEEKKSEL
RHKLKYVPHEYIELIEIARNPTQDRILEMKVMEFFMKVYGYRGEHLGGSRKPDGAIYTVGSPIDYGVIVD
TKAYSGGYNLPIGQADEMQRYVKENQTRNKHINPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNH
KTNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINF*

E51C-4R 右侧全部氨基酸序列 (SEQ ID NO:455)

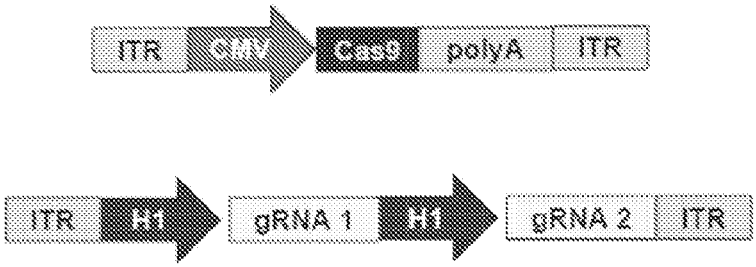
MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGRLEPGERPFQCRICMRNFSIFNHLAPHTRTHTGEKPF
QCRICMRNFSQSAHLKRHLRTHTGEKPFQCRICMRNFSHHNSLTRHLKTHTGAAAPALVKSELEEKKSEL
RHKLKYVPHEYIELIEIARNPTQDRILEMKVMEFFMKVYGYRGEHLGGSRKPDGAIYTVGSPIDYGVIVD
TKAYSGGYNLPIGQADEMERYVEENQTRDKHLNPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNH
ITNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINF*

图 38

“单一载体，多重 CRISPR 系统”



“二元载体，多重 CRISPR 系统”



“单一载体，单一 gRNA 系统”



图 39

SaCas9-NLS (NLS 用下划线表示) (SEQ ID NO: 459)

atgaaaaggaactacattctgggggctggacatcgggattacaagcgtgggtatgggattattgactatgaaacaag
gtacgtgatcgacgcaggcgtcagactgttcaaggaggccaacgtggaaaacaatgagggacggagaagcaagagg
gagccaggcgcctgaaacgacggagaaggcacagaatccagagggtgaagaaactgctgttcgattacaacctgctg
accgaccattctgagctgagtggaattaatcctttatgaagccagggtgaaaggcctgagtcagaagctgtcagagga
agagtttttccgcagctctgctgcacctggctaagcgccgaggagtgcataacgtcaatgaggtggaagaggacaccg
gcaacgagctgtctacaaaggaacagatctcacgcaatagcaaagctctggaagagaagtatgtcgcagagctgcag
ctggaacggctgaagaaagatggcgaggtgagagggtcaattaataggttcaagacaagcgactacgtcaaagaagc
caagcagctgctgaaagtgcagaaggcttaccaccagctggatcagagcttcatcgatacttatatcgacctgctgg
agactcggagaacctactatgagggaccaggagaagggagccccttcggatggaaagacatcaaggaatggtacgag
atgctgatgggacattgcacctatttccagaagagctgagaagcgtcaagtacgcttataacgcagatctgtacaa
cgccctgaatgacctgaacaacctggtcatcaccagggatgaaaacgagaaactggaatactatgagaagttccaga
tcatcgaaaacgtgtttaagcagaagaaaaagccttacactgaaacagattgctaaggagatcctggtcaacgaagag
gacatcaagggctaccgggtgacaagcactggaaaaccagagttcaccaatctgaaagtgtatcacgatattaagga
catcacagcacggaaagaaatcattgagaacgccgaactgctggatcagattgctaagatcctgactatctaccaga
gctccgaggacatccaggaagagctgactaacctgaacagcgagctgacccaggaagagatcgaaacagattagtaat
ctgaagggtacaccggaacacacaacctgtccctgaaagctatcaatctgattctggatgagctgtggcatacaaa
cgacaatcagattgcaatctttaaccggctgaagctggtcccaaaaaaggtgtgacctgagtcagcagaaagagatcc
caaccacactggtggacgatttcattctgtcaccgtggtcaagcggagcttcatccagagcatcaaagtgatcaac
gccatcatcaagaagtacggcctgcccaatgatatcattatcgagctggctagggagaagaacagcaaggacgcaca
gaagatgatcaatgagatgcagaaacgaaaccgtgcagaccaatgaacgcattgaagagattatccgaactaccggga
aagagaacgcaaagtaccgattgaaaaaatcaagctgcacgatatgcaggagggaaagtgtctgtatttcctggag
gccatcccctggaggacctgctgaacaatccattcaactacgagctcgatcatattatccccagaagcgtgtcctt
cgacaattcctttaacaacaaggtgctggtcaagcaggaagagaactctaaaaagggcaataggactcctttccagt
acctgtctagttcagattccaagatcttacgaaaccttaaaaagcacattctgaatctggccaaagtgaaagggc
cgcatcagcaagaccaaaaaggagtaccttgctggaagagcgggacatcaacagattttccgtccagaaggattttat
taaccggaatctggtggacactagatacgctactcgcggcctgatgaatctgctgcgatcctactttccgggtgaaca
atctggatgtgaaagtcaagtccatcaacggcgggttcacatctttctgaggcgcaaatggaagtttaaaaaggag
cgcaacaaagggtacaagcaccatgccgaagatgccttgattatcgcaaatgccgacttcatctttaaggagtggaa
aaagctggacaaagccaagaaagtgatggagaaccagatgttcgaagagaagcaggccgaatctatgccccgaaatcg
agacagaacaggagtacaaggaggatttcatcactcctcaccagatcaagcatatcaaggatttcaaggactacaag
tactctcaccggtgtggataaaaagcccaacagagagctgatcaatgacaccctgtatagtacaagaaaagacgataa
ggggaataccctgattgtgaacaatctgaacggactgtacgacaaagataatgacaagctgaaaaagctgatcaaca
aaagtcccgagaagctgctgatgtaccaccatgatcctcagacatatcagaaactgaagctgattatggagcagtac
ggcgacgagaagaacccactgtataagtactatgaagagactgggaactacctgaccaagtatagcaaaaaggataa
tggccccgtgatcaagaagatcaagtactatgggaacaagctgaatgcccatctggacatcacagacgattacccta
acagtcgcaacaaggtggtcaagctgtcactgaagccatacagattcgatgtctatctggacaacggcgtgtataaa
tttgtgactgtcaagaatctggatgtcatcaaaaaggagaactactatgaagtgaatagcaagtgctacgaagaggc
taaaaagctgaaaaagattagcaaccaggcagagttcatcgcctcctttacaacaacgacctgattaagatcaatg
gcgaactgtatagggtcatcgggtgaacaatgatctgctgaaccgcattgaagtgaatatgattgacatcacttac
cgagagtatctggaaaacatgaatgataagcgcccccctcgaattatcaaaacaattgcctctaagactcagagtat
caaaaagtactcaaccgacattctgggaaacctgtatgaggtgaagagcaaaaagcaccctcagattatcaaaaagg
gcagcggaggcaagcgtcctgcgctactaagaaagctggtcaagctaagaaaaagaaaaggatcctacccatacgat
gttccagattacgcttaa

SaCas9 gRNA (SEQ ID NO: 460)

[protospacer]gtttttagtactctggaaacagaatctactaaaacaaggcaaaatgccgtgtttatctcgtcaac
ttgttggcgagatttttttt

图 40

NmCas9 (NLS 1 用下划线表示,NLS 2 用下划线和粗体表示,HA 标签用粗体表示) (SEQ ID NO: 461)

atggtgcctaagaagaagaggaaggtcgtcgtgtccttcaaacctaattcaatcaactacatcctcggcctcgatatcgg
catcgcatccgtcggctgggcgatggtagaaattgacgaagaagaaaaccccatccgcctgattgatttgggcgtgc
gcgtatttgagcgtgccgaagtaccgaaaacaggcgactcccttgccatggcaacgtcgtttggcgcgcagtgttcgc
cgcctgaccccgccgtcgcgcccacccgccctgcttcggaccccgccgcctattgaaacgcgaaggcgtatttacaagccgc
caatttttgacgaaaacggcttgatttaaatcctttaccgaatacaccatggcaacttcgcgcagccgcattagaccgca
aactgacgcctttagagtggtcggcagtcttgttgcatttaatcaaacatgcggctatttatcgcaacggaaaaac
gagggcgaaactgccgataaggagcttggcgctttgcttaaaggcgtagccggcaatgcccatgccttacagacagg
cgatttccgcacaccggccgaattggcttaaataaatttgagaaagaaagcggccatatccgcaatcagcgcagcg
attattgcatacgttcagccgcaaaggattttcaggcgggagctgatttgctgtttgaaaaccaaaaagaatttgcc
aatccgcatgtttcaggcggccttaaagaaggtattgaaaccctactgatgacgcaacgccctgccctgtccggcga
tgccgttcaaaaaaatgttgggcattgcaccttcgaacccgccagagccgaaagccgctaaaaacacctacaacagccg
aacgtttcatctggctgaccaagctgaacaacctgcgtatttttagagcaaggcagcgagcggccattgaccgatacc
gaacgcgccacgcttatggacgagccatacagaaaatccaaactgacttacgcacaagcccgtaagctgctgggtttt
agaagataccgcctttttcaaaggcttgcgctatggtaaagacaatgccgaagcctcaacattgatggaaatgaagg
cttaccatgccatcagcggtgcactggaaaaagaaggatttgaaaagacaaaaaaatcccccattaaaacttttctccgcgaa
ttacaagacgaaatcggcacggcattctcccttgttcaaaaccgatgaagacattacaggccgtctgaaagaccgtat
acagcccgaaatcttagaagcgctgttgaaacacatcagcttcgataagttcgtccaaatttccttgaaagcattgc
gccgaatttgtgccttctaatggaacaaggcaaacgttacgatgaagcctgcgccgaaatctacggagaccattacggc
aagaagaatacggaagaaaagatttatctgccgccgattcccgccgacgaaatccgcaacccccgtcgtcttgcgcgc
cttatctcaagcacgtaaggtcatttaacggcgtggtacgccgttacggctccccagctcgtatccatattgaaaactg
caagggaagtaggtaaatcgtttaaagaccgcaaagaaattgagaaacgcccaagaagaaaaccgcaaagaccgggaa
aaagcgccgccaaattccgagagtatttcccgaatttgtcggagaacccaaatccaaagatattctgaaactgcgt
cctgtacgagcaacaacacggcaaatgcctgcattcgggcaaagaaatcaacttaggccgtctgaacgaaaaaggct
atgtcgaaatcgaccatgccctgccgtttctcgcgcaacatgggacgacagtttcaacaataaaqtactggtattgggc
agcgaaaaccaaaacaaaggcaatcaaaccccttacgaatacttcaacggcaaagacaaacagccgcgaatggcagga
atttaaagcgcgtgtcgaaaccagccgtttccgcgcagtaaaaaacaacggattctgctgcaaaaattcgatgaaag
acggctttaaagaacgcaatctgaacgacacgcgctacgtcaaccgtttcctgtgtcaatttgttgccgaccgtatg
cggctgacaggtaaaggcaagaaacgtgtctttgcatccaacggacaaattaccaatctgttgcgcggctttgggg
attgcgcaaagtgcgtgcggaaaacgaccgccatcacgccttggacgccgtcgtcgttgcctgctcgaccgttgcca
tgcagcagaaaattacccgtttgtacgctataaagagatgaacgcgtttgacgttaaaaccatagacaaagaaaca
ggagaagtgctgcatcaaaaaacacacttcccacaacctgggaatttttcgcacaagaagtcatgattcgcgtctt
cggcaaacccggacggcaaaccccgaattcgaagaagccgataccctagaaaaactgcgcacgttgcttgccgaaaaat
tatcatctcgccccgaagccgtacacgaatacgttacgccactgtttgtttcacgcgcgcccaatcggaagatgagc
gggcaaggtcatatggagaccgtcaaatccgccaaacgactggacgaaggcgtcagcgtgttgcgcgtaccgctgac
acagttaaaactgaaagacttggaaaaaatggtcaatcgggaqcgcgaacctaagctatacgaagcactgaaagcac
ggctggaagcacataaagacgatcctgccaaagccttttgccgagccgtttttacaaatacgatcaaagcaggcaaccgc
acccaacaggtaaaagccgtacgcgtagagcaagtacagaaaaccggcgtatggtgcgcaaccataacggtattgc
cgacaacgcaaccatggtgcgcgtagatgtgtttgagaaaggcgacaagtattatctggtaccgatttacagttggc
aggtagcgaaagggatttgccggatagggctgttgtacaaggaaaagatgaagaagattggcaacttattgatgat
agtttcaacttaaattctcattacaccctaatgatttagtcgaggttataacaaaaaaagctagaatgtttggtta
ctttgccagctgccatcgaggcacaggtaatatcaatatacgcattcatgatcttgatcataaaattggcaaaaatg
gaatactggaaggtatcggcgtcaaaacgccctttcattccaaaaatacccaaattgacgaactgggcaaagaaatc
agaccatgccgtctgaaaaaacgcccgcctgtccgt**tacccatacgatgttccagattacgctgcagctccagcagc**
ggaagaaaaagaagctggattaa

来自 Thomson PNAS 2013 的 NmCas9 短发夹(SEQ ID NO: 462)

{protospacer}GTTGTAGCTCCCTTTCTCATTTCGCAAACGAAATGACAACCGTTGCTACAATAAGGCCGTCTGA
AAAGATGTGCCGCAAGCGTCTGCCCCTTAAAGCTTCTGCTTTAAGGGGCTTTTTTT

来自 Church Nature Biotech 2013 的 NmCas9 长发夹(SEQ ID NO: 463)

[protospacer]GTTGTAGCTCCCTTTCTCATTTCGCAGTGCTACAATGAAAATTGTCGCACTGCGAAATGACAACC
GTTGCTACAATAAGGCCGTCTGAAAAGATGTGCCGCAACGCTCTGCCCCTTAAAGCTTCTGCTTTAAGGGGCTTTTT
GTT

图 41

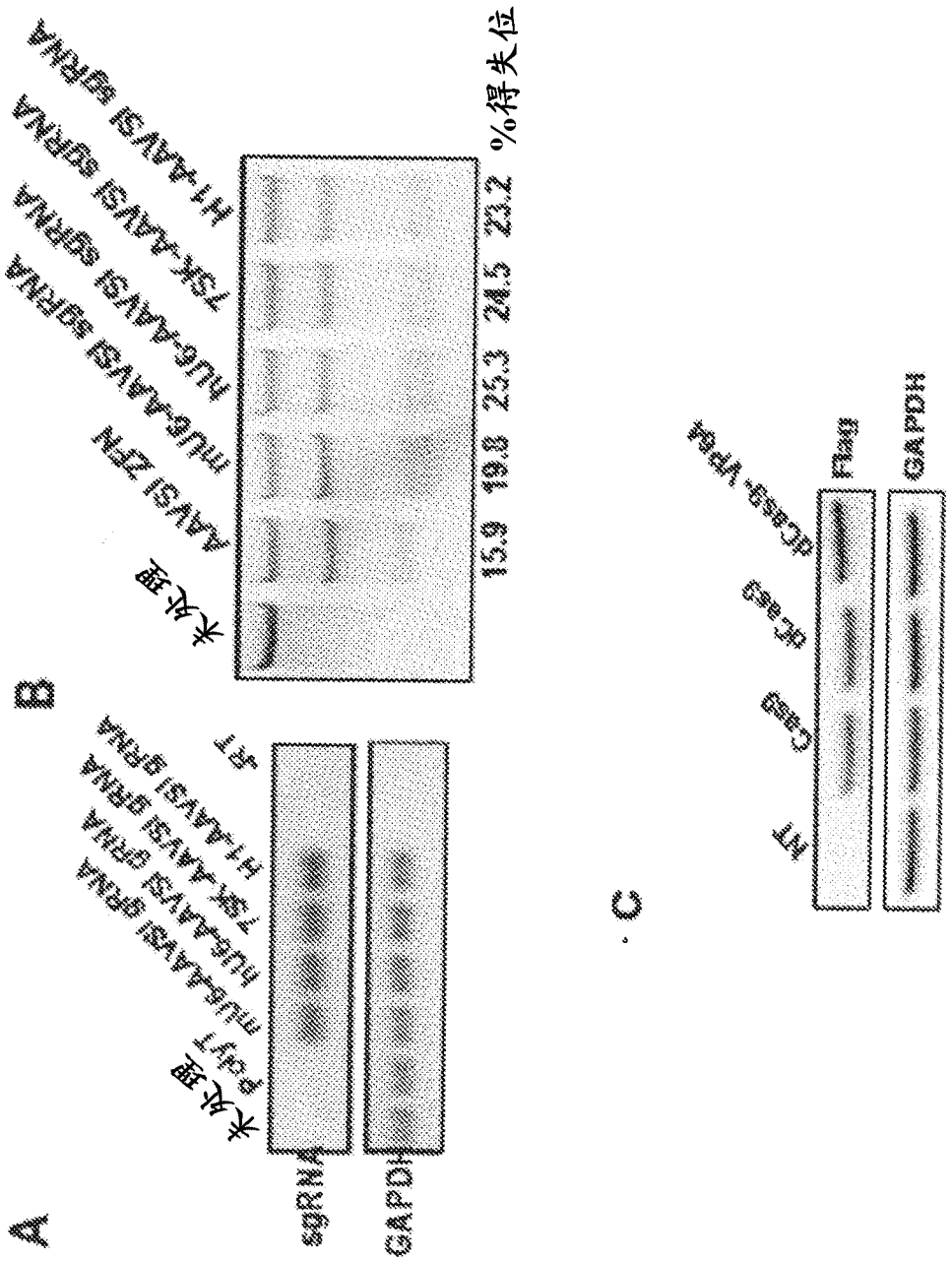

A
sgRNA
GAPDH
未处理
pCMV
mU6-AAVS1 gRNA
hU6-AAVS1 gRNA
T5K-AAVS1 gRNA
H1-AAVS1 gRNA
RT
B
未处理
AAVS1 ZFN
mU6-AAVS1 sgRNA
hU6-AAVS1 sgRNA
T5K-AAVS1 sgRNA
H1-AAVS1 sgRNA
15.9 19.8 25.3 24.5 23.2 %得失位
C
MT
Casp9
dCas9
dCas9-VP64
Flag
GAPDH

图 42

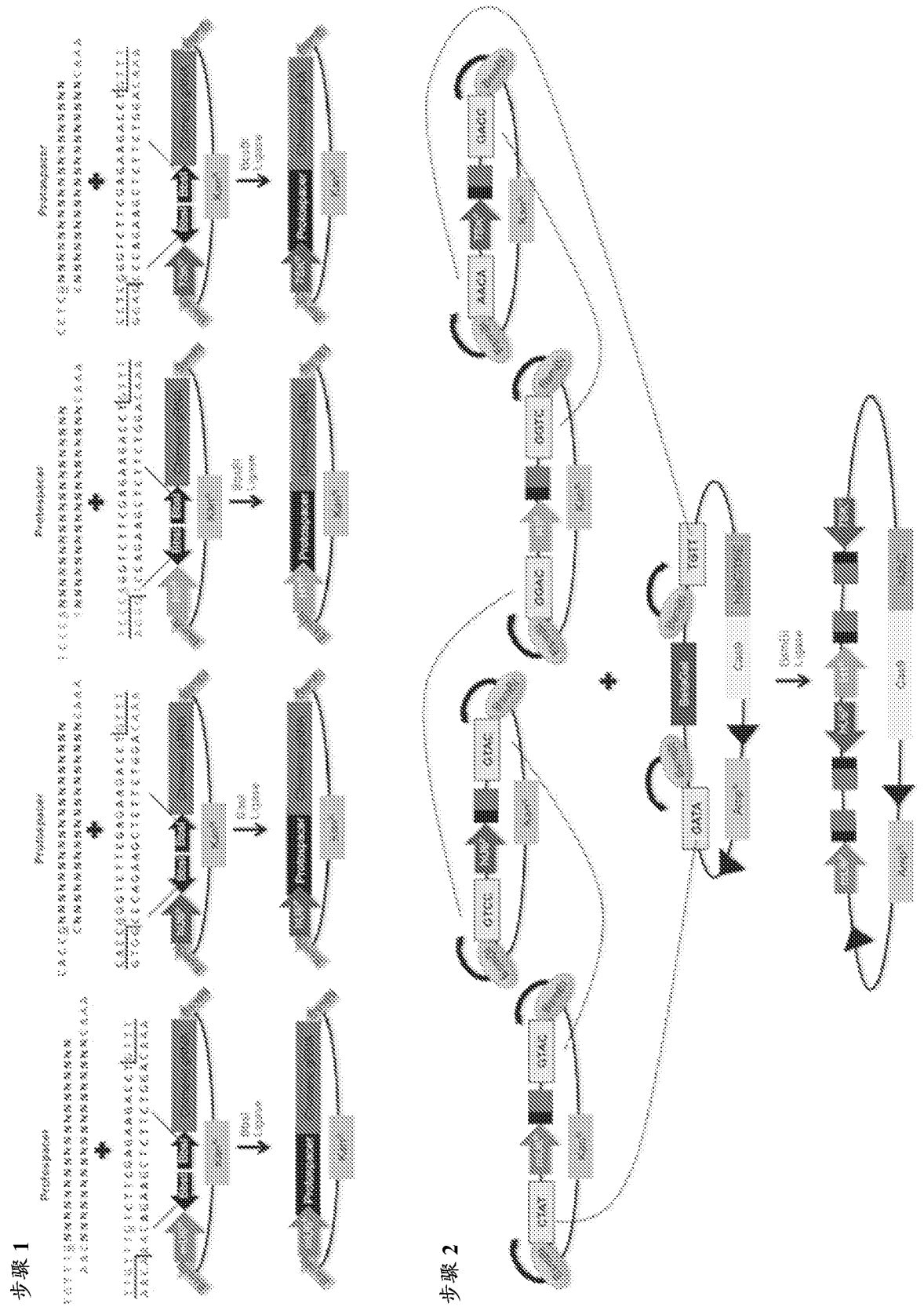


图 43

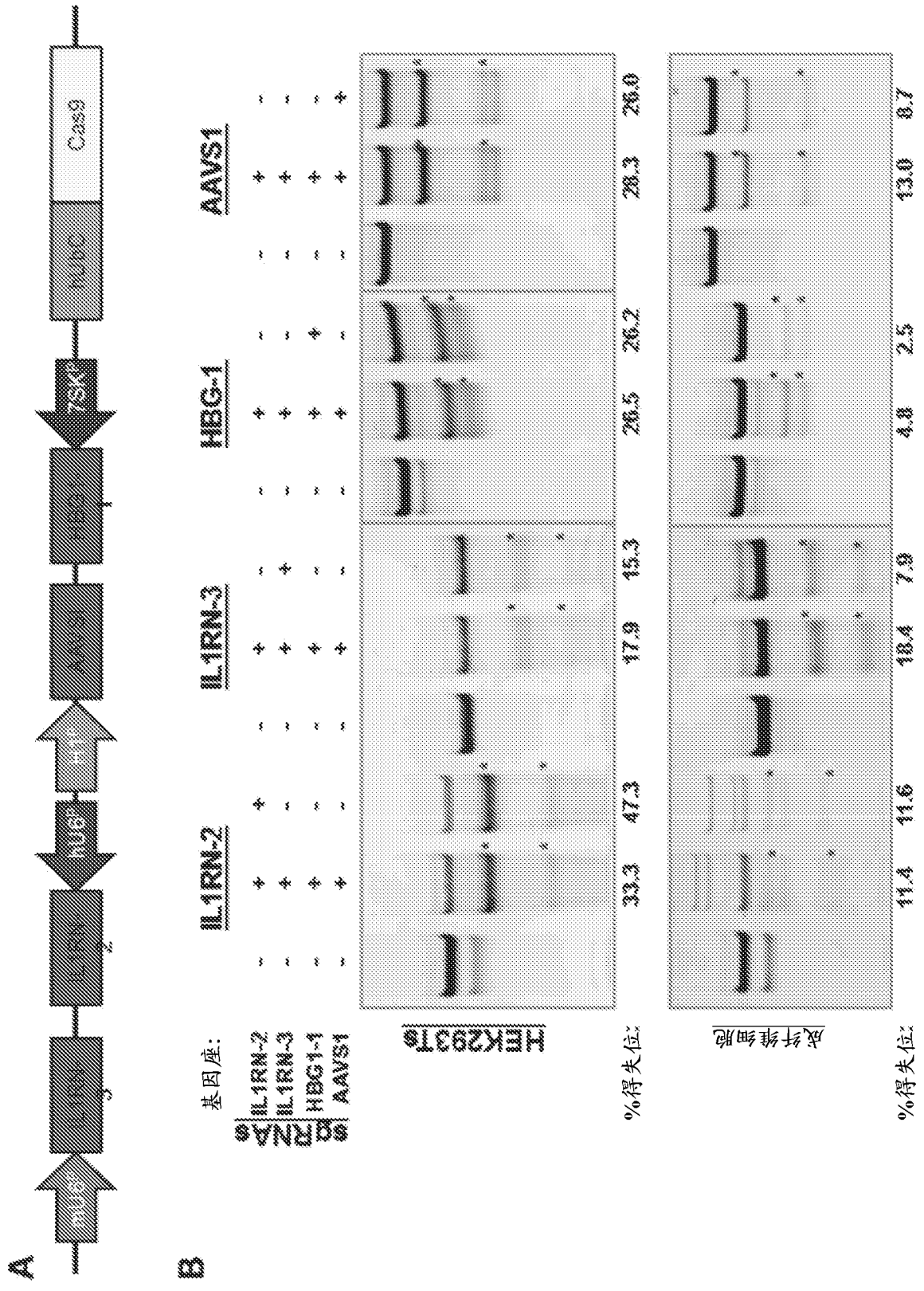

A
Cas9
7SK
B
基因座:
IL1RN-2
IL1RN-3
HBG-1
AAVS1
HEK293Ts
成纤维细胞
%得失位
33.3
47.3
17.9
15.3
26.5
26.2
28.3
26.0
11.4
11.6
18.4
7.9
4.8
2.5
13.0
8.7

图 44

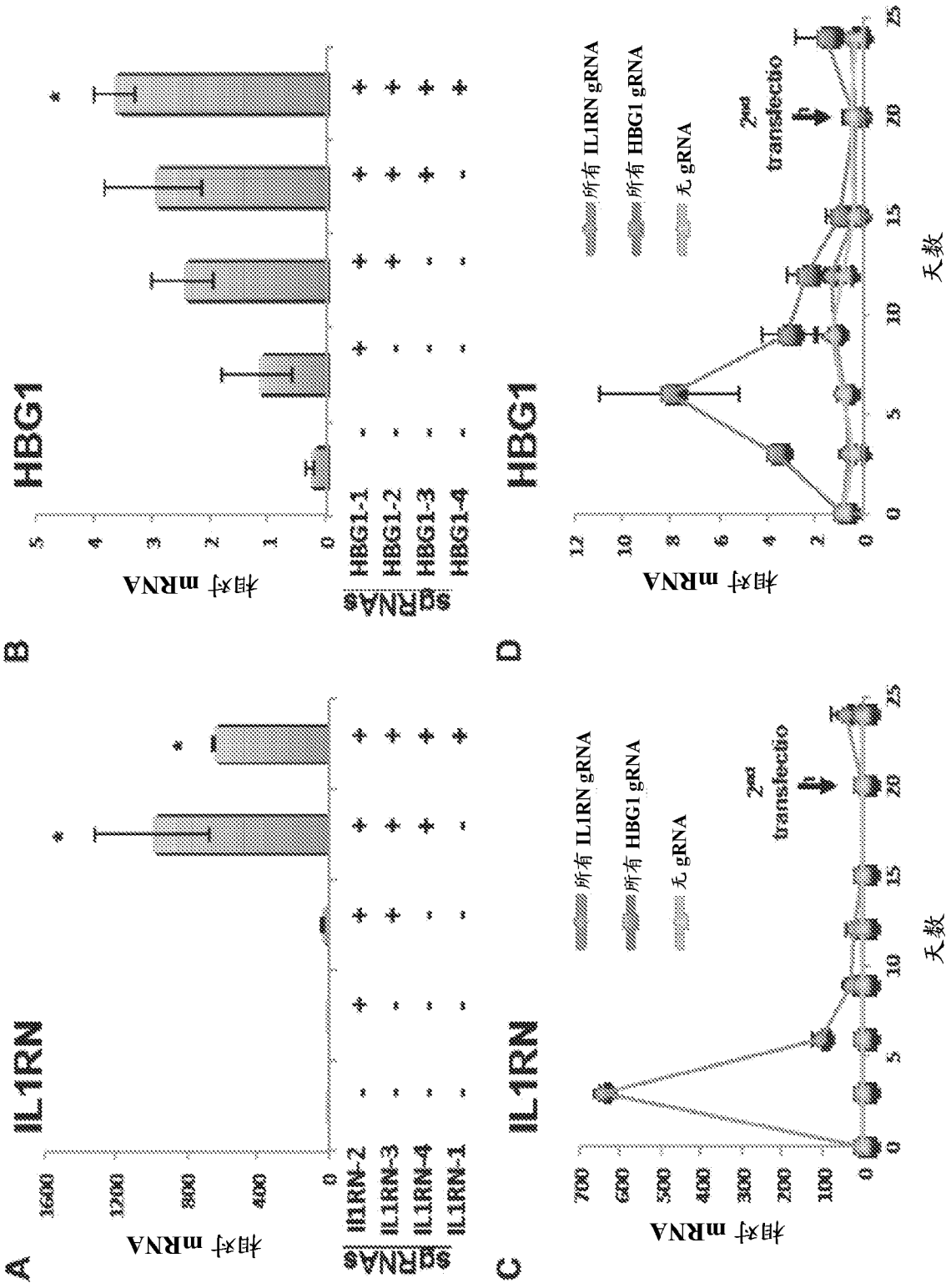

A
IL1RN
相对 mRNA
1600
1200
800
400
0
IL1RN-2
IL1RN-3
IL1RN-4
IL1RN-1
sgRNAs
B
HBG1
相对 mRNA
5
4
3
2
1
0
HBG1-1
HBG1-2
HBG1-3
HBG1-4
sgRNAs
C
IL1RN
相对 mRNA
700
600
500
400
300
200
100
0
所有 IL1RN gRNA
所有 HBG1 gRNA
无 gRNA
2nd transfectio
0
5
10
15
20
25
天数
D
HBG1
相对 mRNA
12
10
8
6
4
2
0
所有 IL1RN gRNA
所有 HBG1 gRNA
无 gRNA
2nd transfectio
0
5
10
15
20
25
天数

图 45

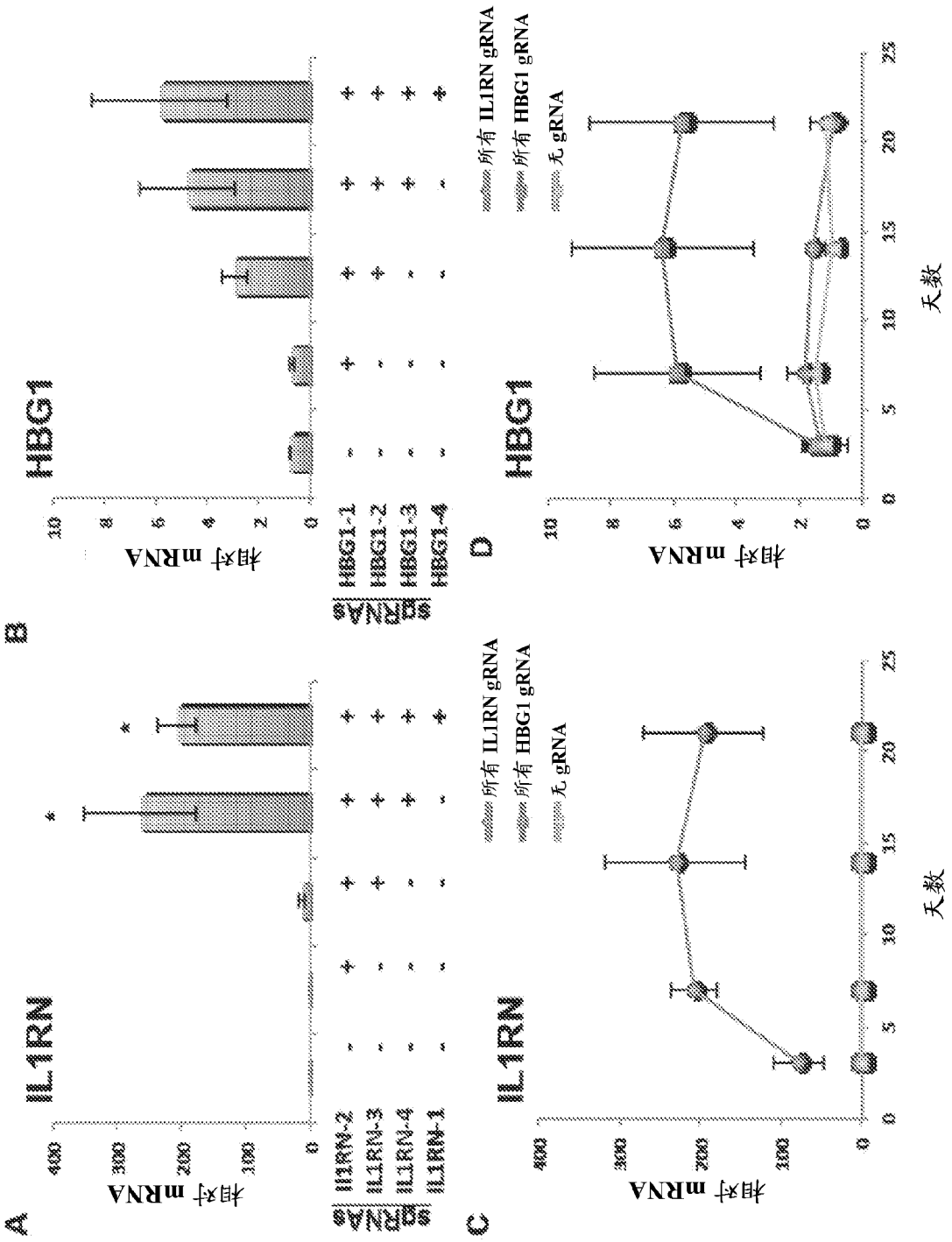

A
IL1RN
相对 mRNA
sgRNAs
IL1RN-2
IL1RN-3
IL1RN-4
IL1RN-1
B
HBG1
相对 mRNA
sgRNAs
HBG1-1
HBG1-2
HBG1-3
HBG1-4
C
IL1RN
相对 mRNA
天数
所有 IL1RN gRNA
所有 HBG1 gRNA
无 gRNA
D
HBG1
相对 mRNA
天数
所有 IL1RN gRNA
所有 HBG1 gRNA
无 gRNA

图 46

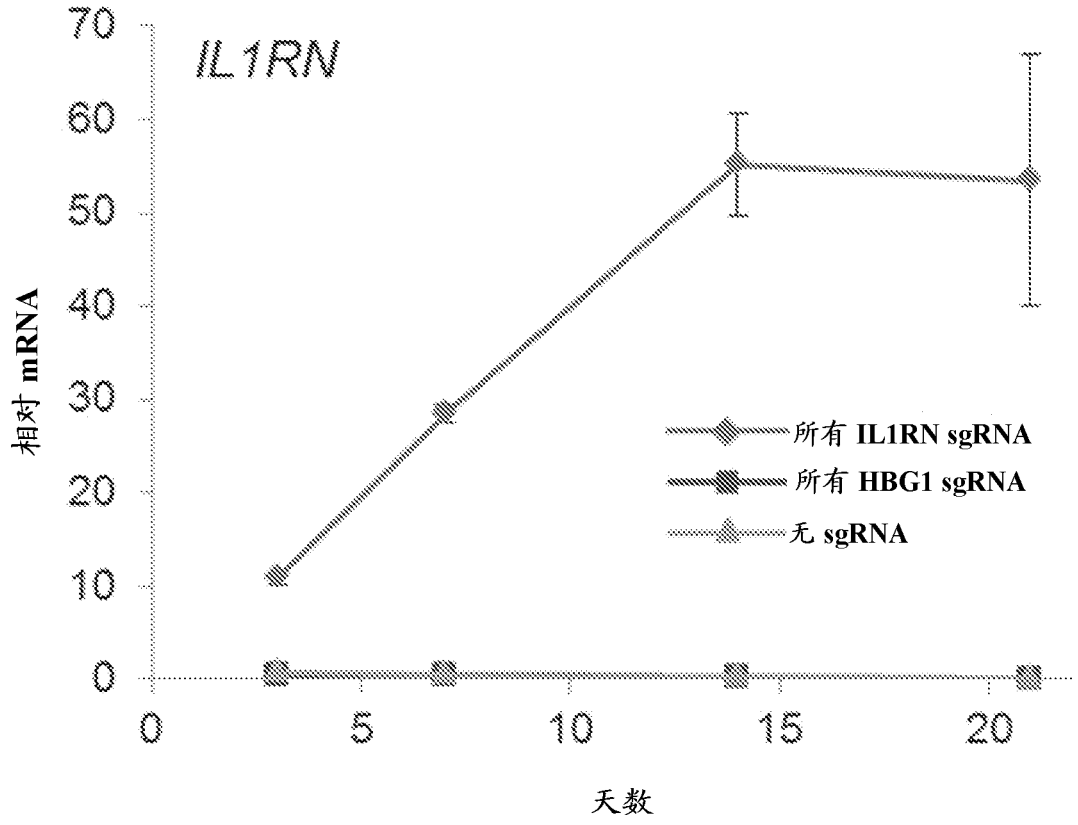

IL1RN
70
60
50
40
30
20
10
0
相对 mRNA
0
5
10
15
20
天数
所有 IL1RN sgRNA
所有 HBG1 sgRNA
无 sgRNA

图 47

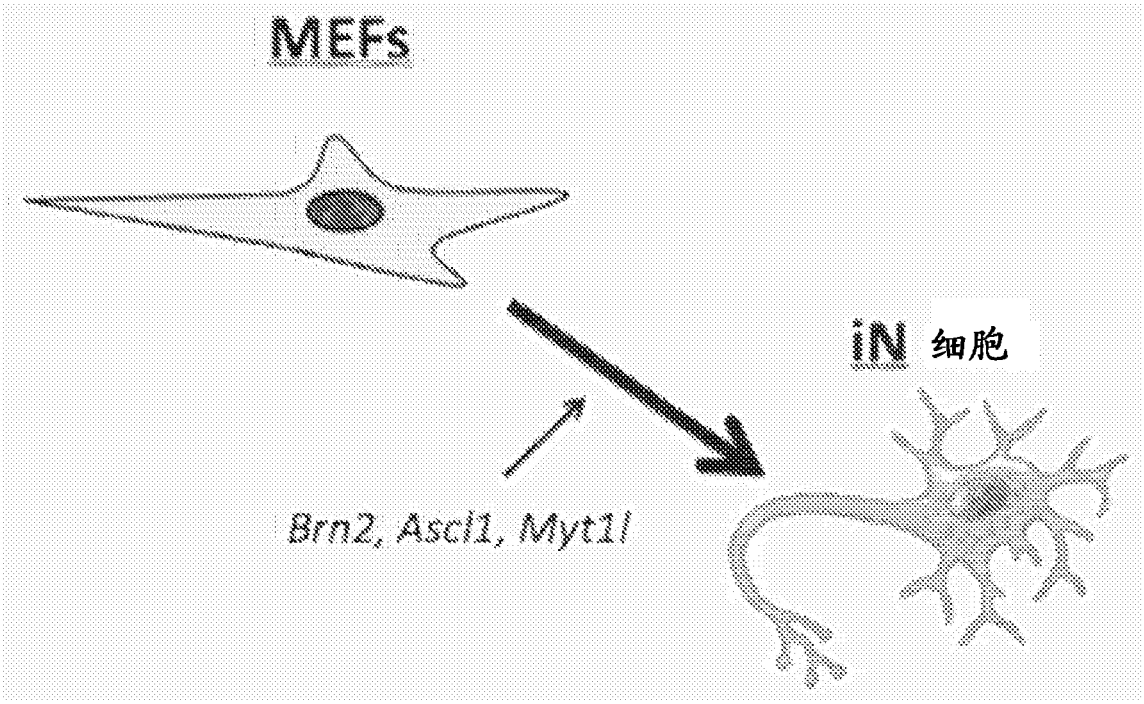

MEFs
iN 细胞
Brn2, Ascl1, Myt1l

图 48

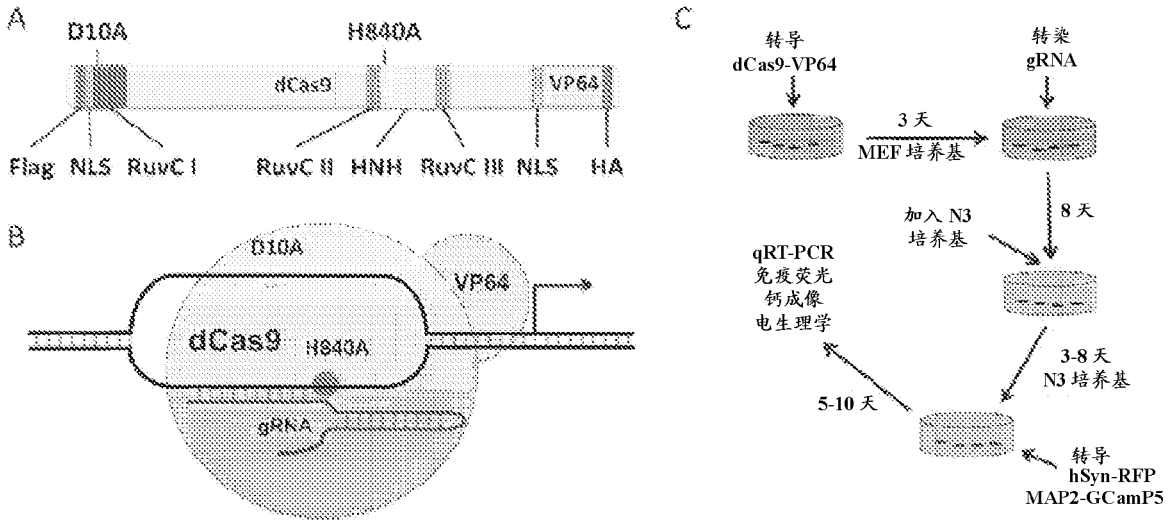

A
D10A
H840A
dCas9
VP64
Flag
NLS
RuvC I
RuvC II
HNH
RuvC III
NLS
HA
B
D10A
VP64
dCas9
H840A
gRNA
C
转导
dCas9-VP64
转染
gRNA
3 天
MEF 培养基
8 天
加入 N3
培养基
qRT-PCR
免疫荧光
钙成像
电生理学
3-8 天
N3 培养基
5-10 天
转导
hSyn-RFP
MAP2-GCamP5

图 49

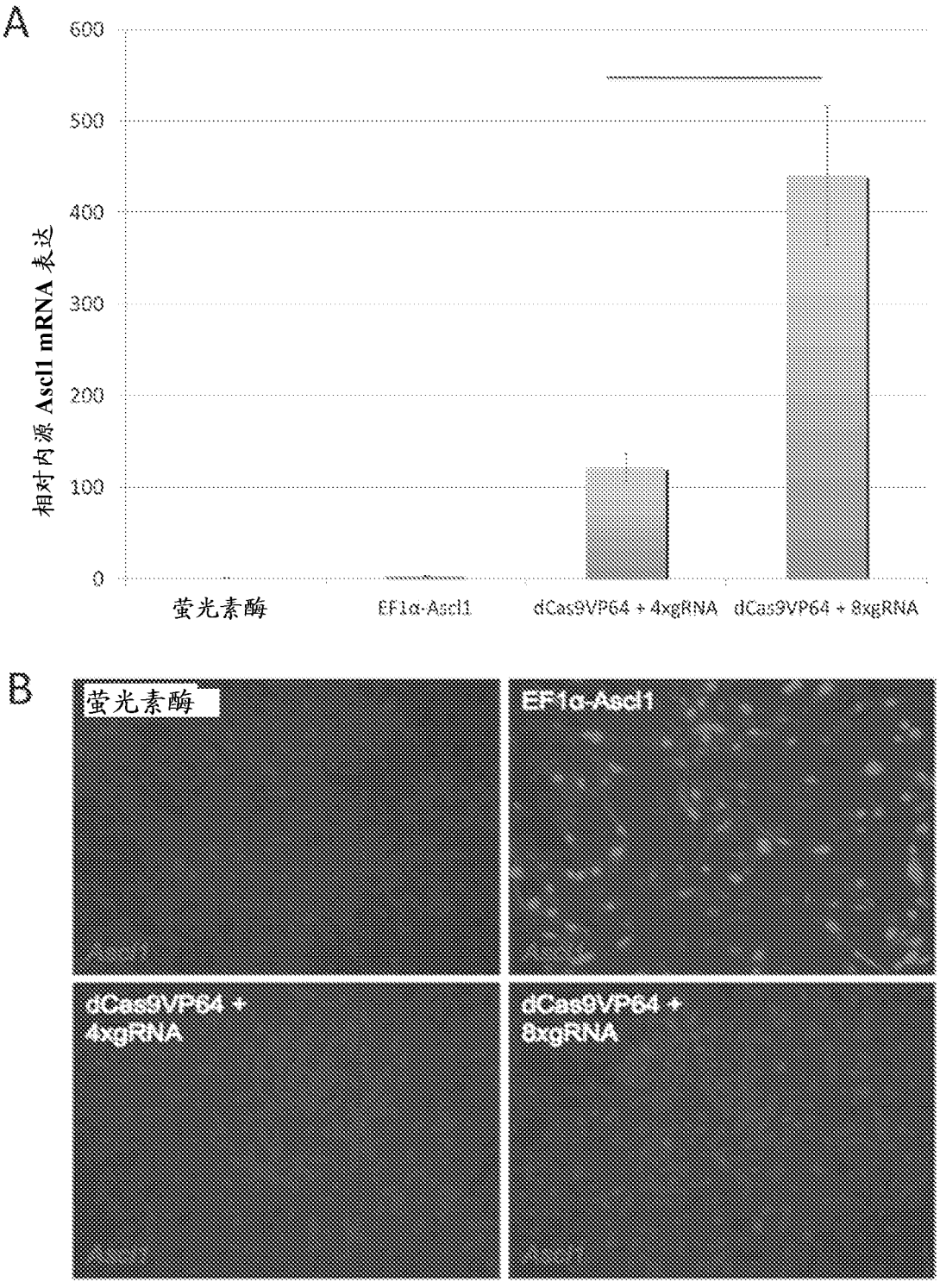

A
600
500
400
300
200
100
0
相对内源 Ascl1 mRNA 表达
萤光素酶
EF1α-Ascl1
dCas9VP64 + 4xgRNA
dCas9VP64 + 8xgRNA
B
萤光素酶
EF1α-Ascl1
dCas9VP64 + 4xgRNA
dCas9VP64 + 8xgRNA

图 50

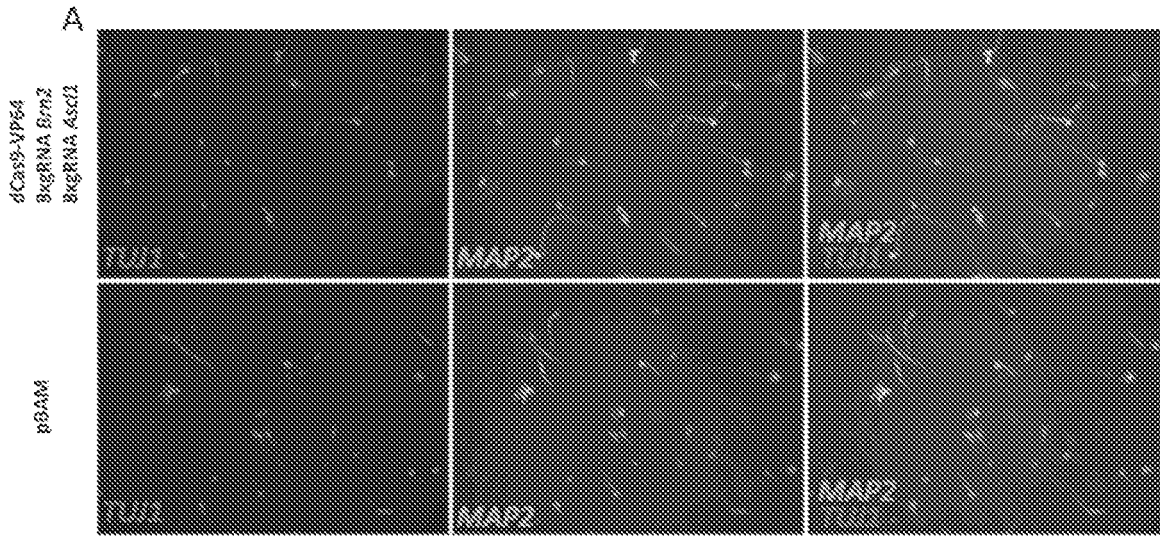


图 51

A



图 52

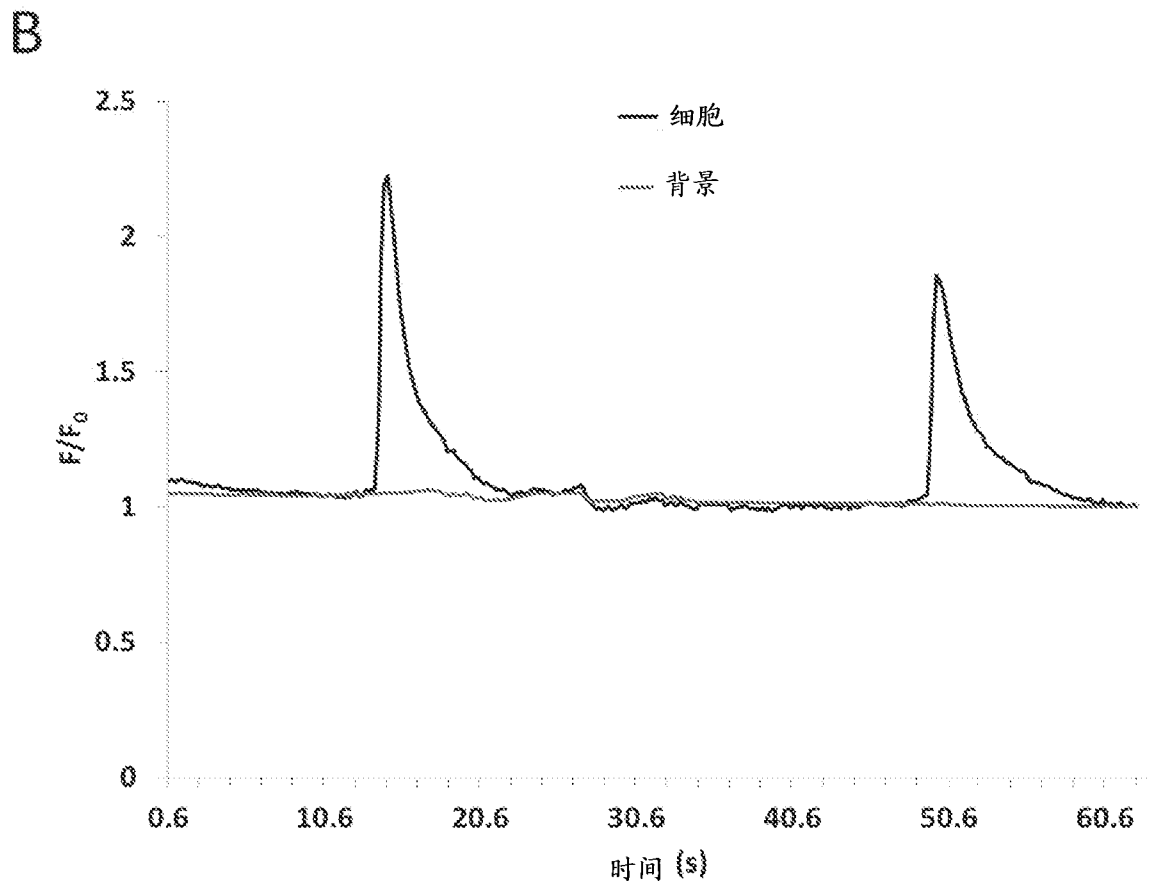


图 52(续)

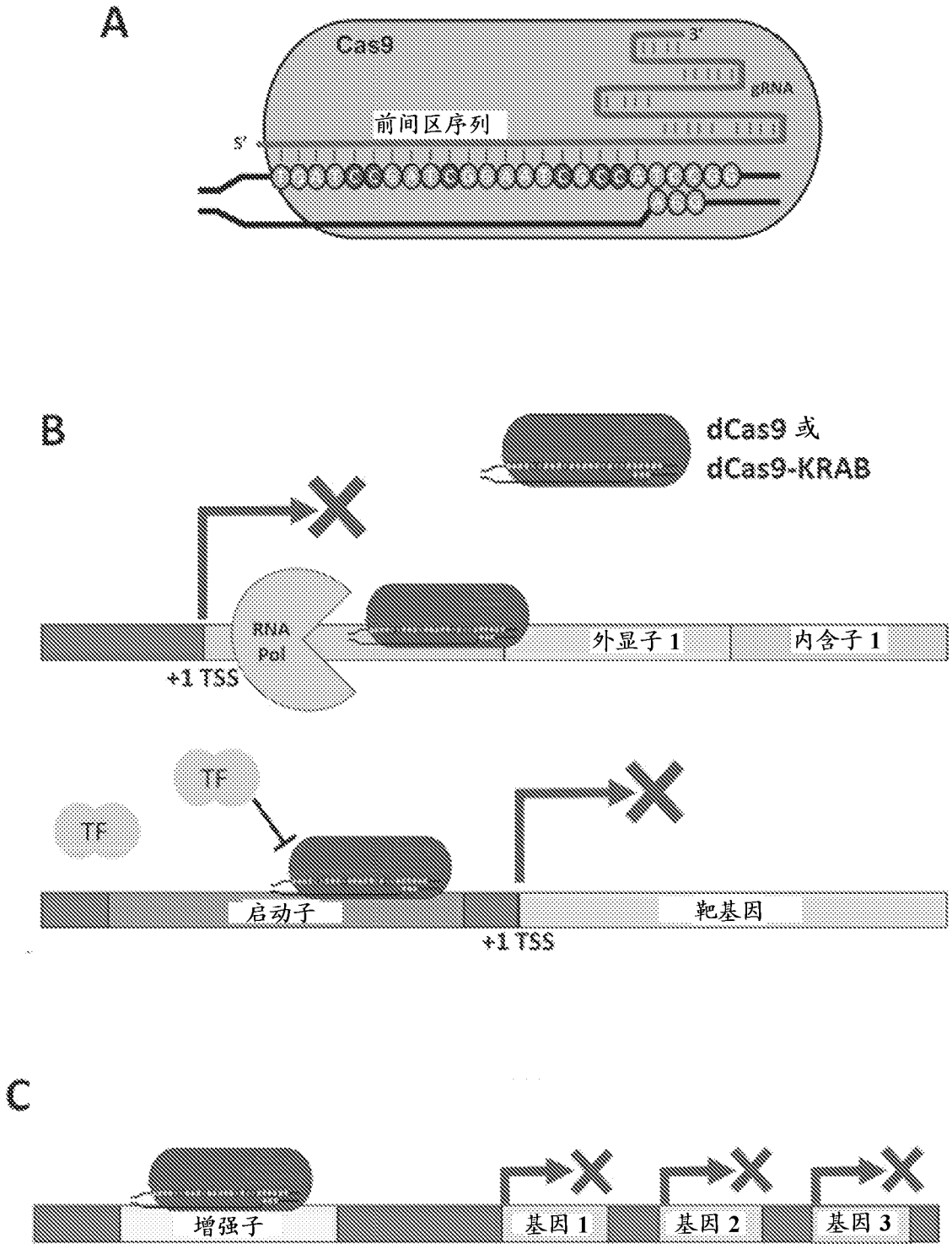

A
Cas9
3'
gRNA
前间区序列
5'
B
dCas9 或
dCas9-KRAB
RNA
Pol
+1 TSS
外显子 1
内含子 1
TF
TF
启动子
靶基因
+1 TSS
C
增强子
基因 1
基因 2
基因 3

图 53

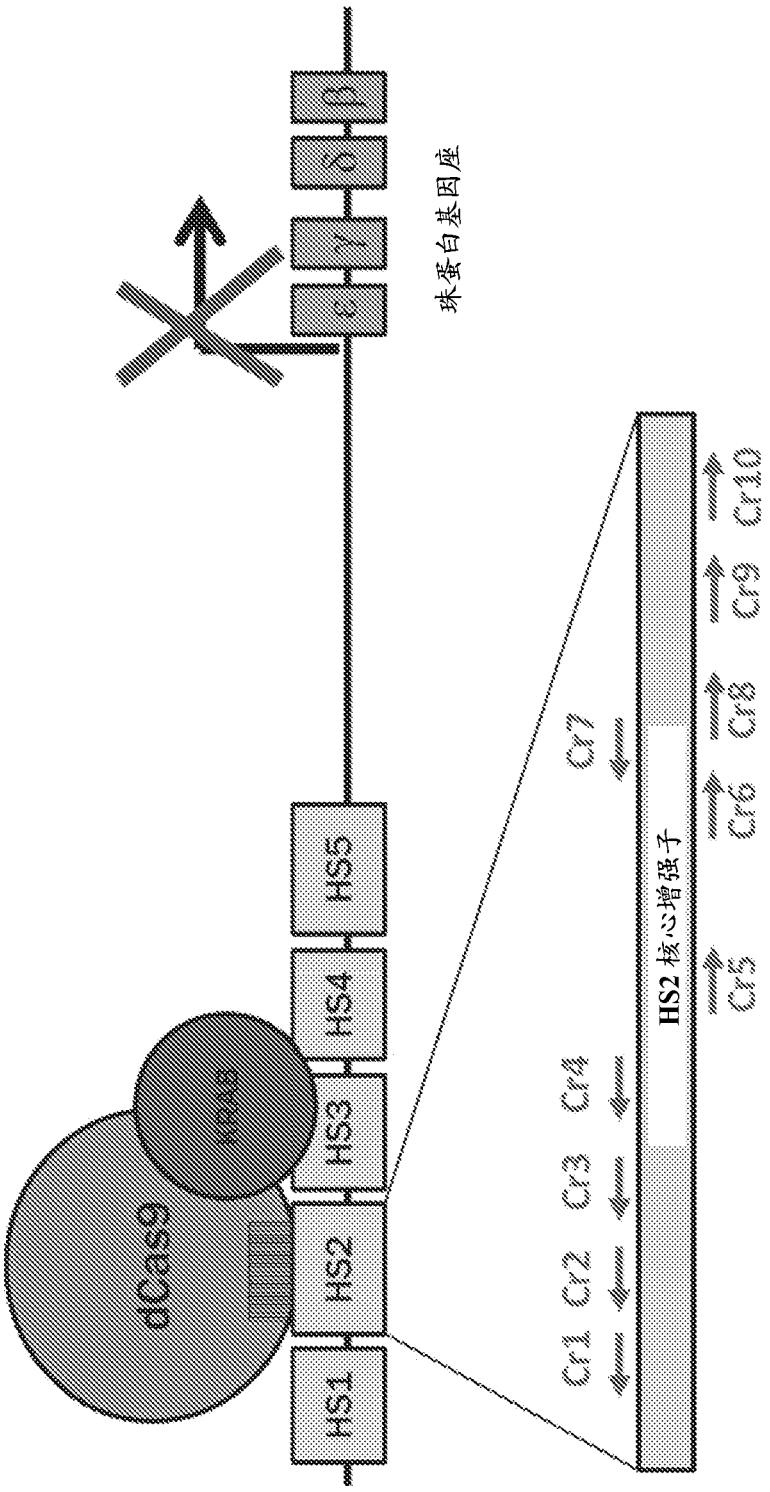

珠蛋白基因座
ε
γ
δ
β
HS5
HS4
HS3
HS2
HS1
KRAB
dCas9
Cr7
Cr6
Cr8
Cr9
Cr10
Cr5
HS2 核心增强子
Cr4
Cr3
Cr2
Cr1

图 54

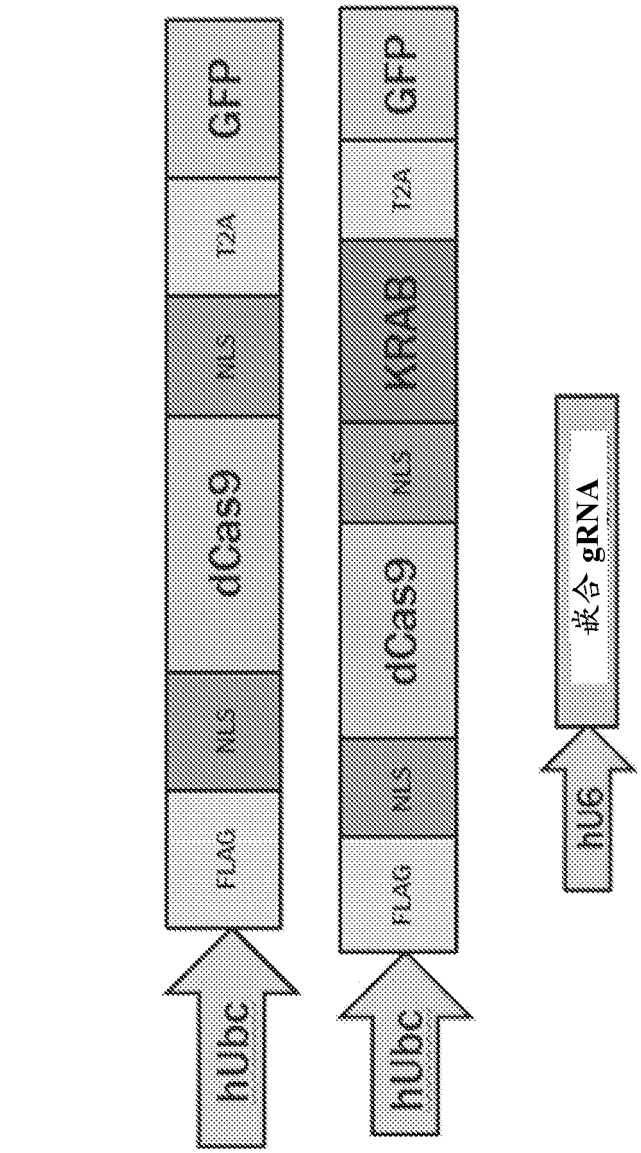

A
hUbc
FLAG
NLS
dCas9
NLS
KRAB
T2A
GFP
hUbc
FLAG
NLS
dCas9
NLS
T2A
GFP
hU6
嵌合 gRNA

图 55

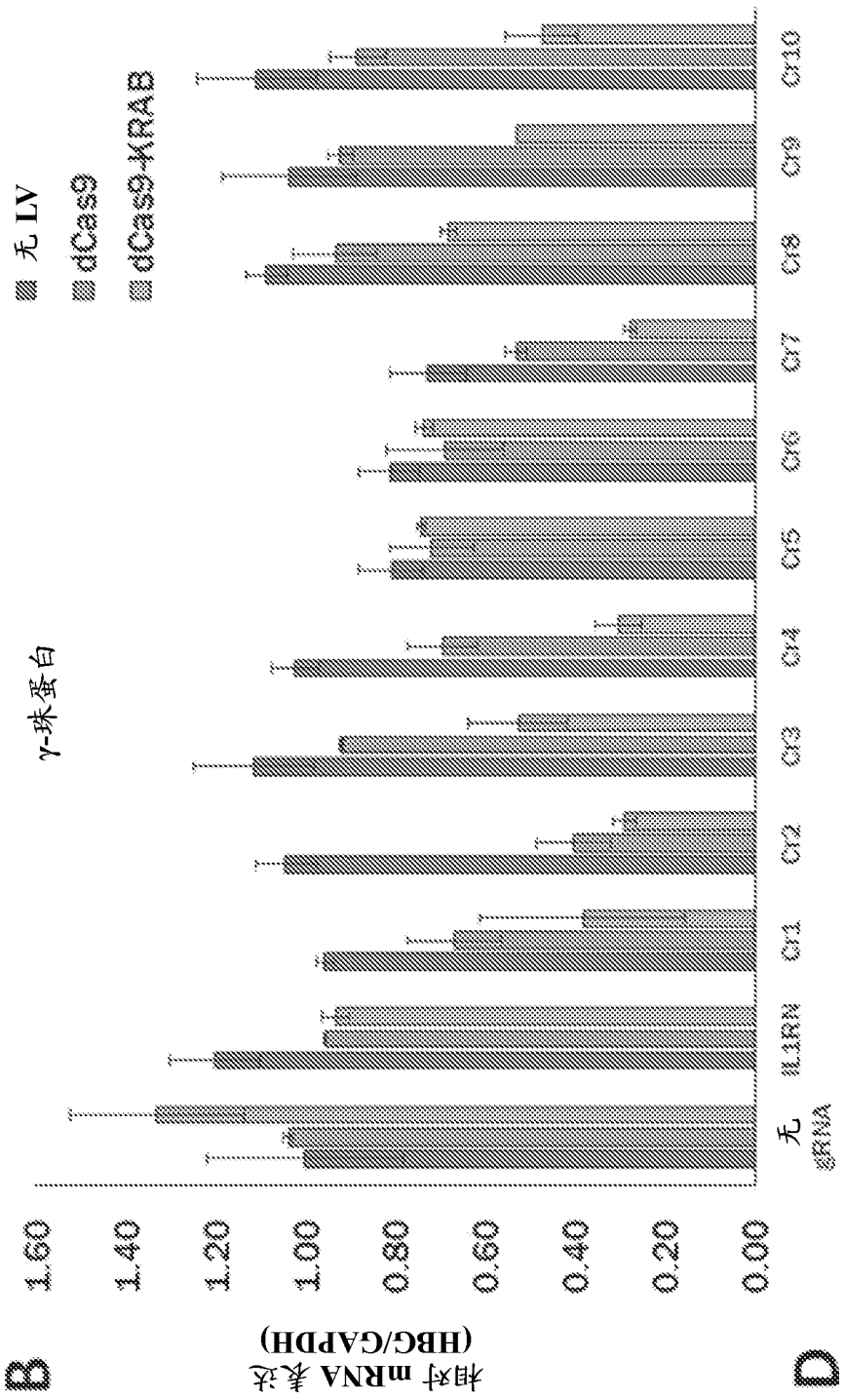

B
γ-珠蛋白
无 LV
dCas9
dCas9-KRAB
相对 mRNA 表达
(HBG/GAPDH)
1.60
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
无
gRNA
IL1RN
Cr1
Cr2
Cr3
Cr4
Cr5
Cr6
Cr7
Cr8
Cr9
Cr10
D

图 55(续)

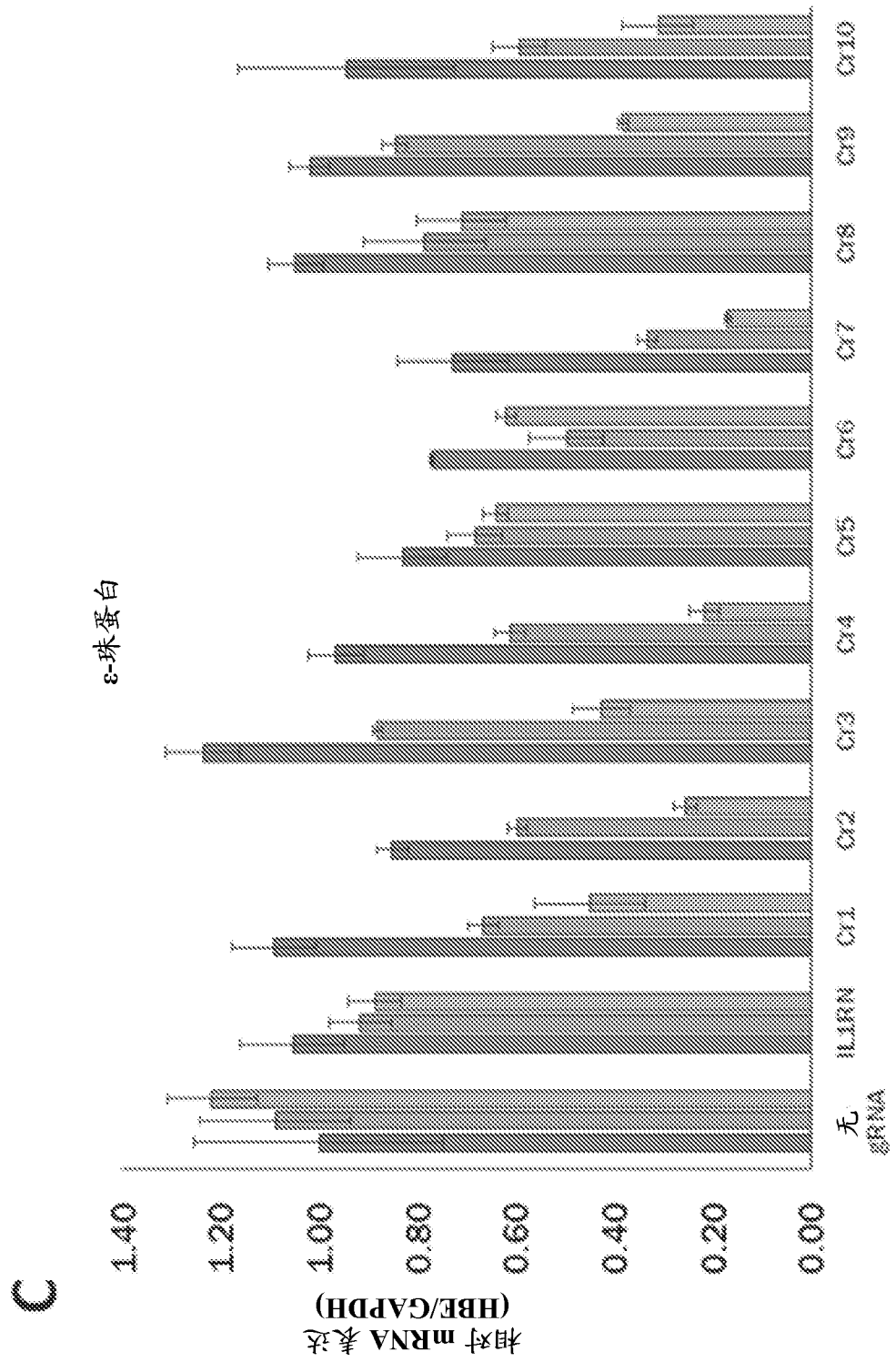

C
相对 mRNA 表达
(HBE/GAPDH)
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
ε-珠蛋白
无
gRNA
IL1RN
Cr1
Cr2
Cr3
Cr4
Cr5
Cr6
Cr7
Cr8
Cr9
Cr10

图 55(续)

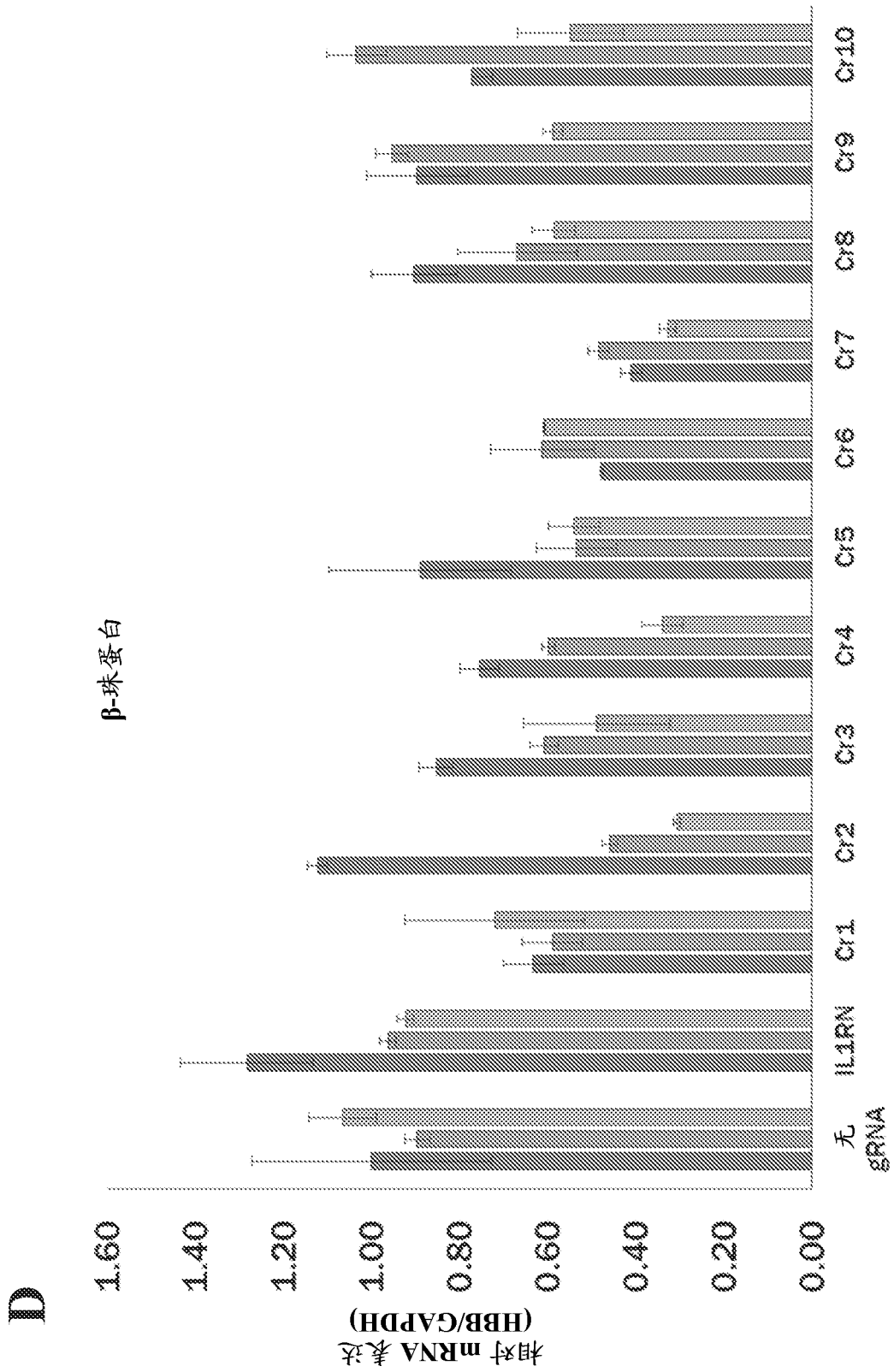

D
β-珠蛋白
相对 mRNA 表达
(HBB/GAPDH)
1.60
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
无 gRNA
IL1RN
Cr1
Cr2
Cr3
Cr4
Cr5
Cr6
Cr7
Cr8
Cr9
Cr10

图 55(续)

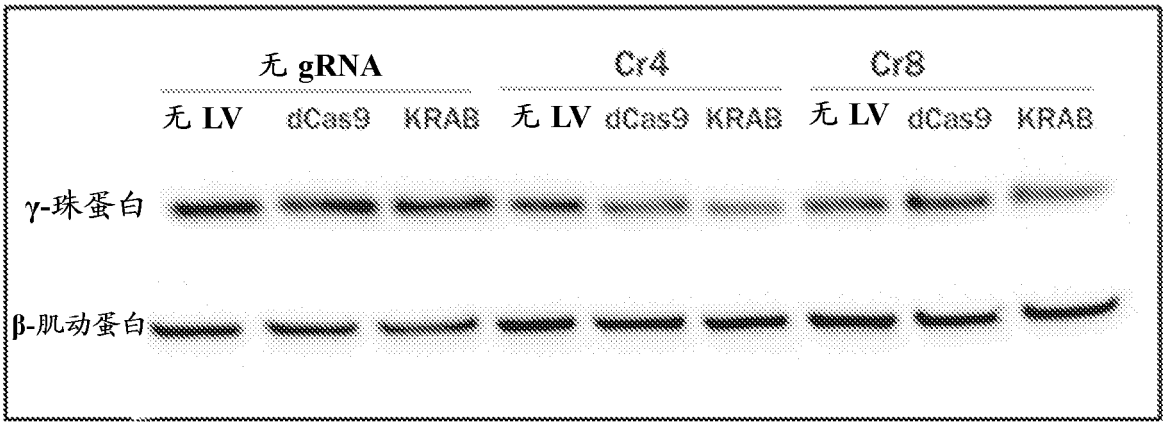

无 gRNA
Cr4
Cr8
无 LV
dCas9
KRAB
无 LV
dCas9
KRAB
无 LV
dCas9
KRAB
γ-珠蛋白
β-肌动蛋白

图 55E

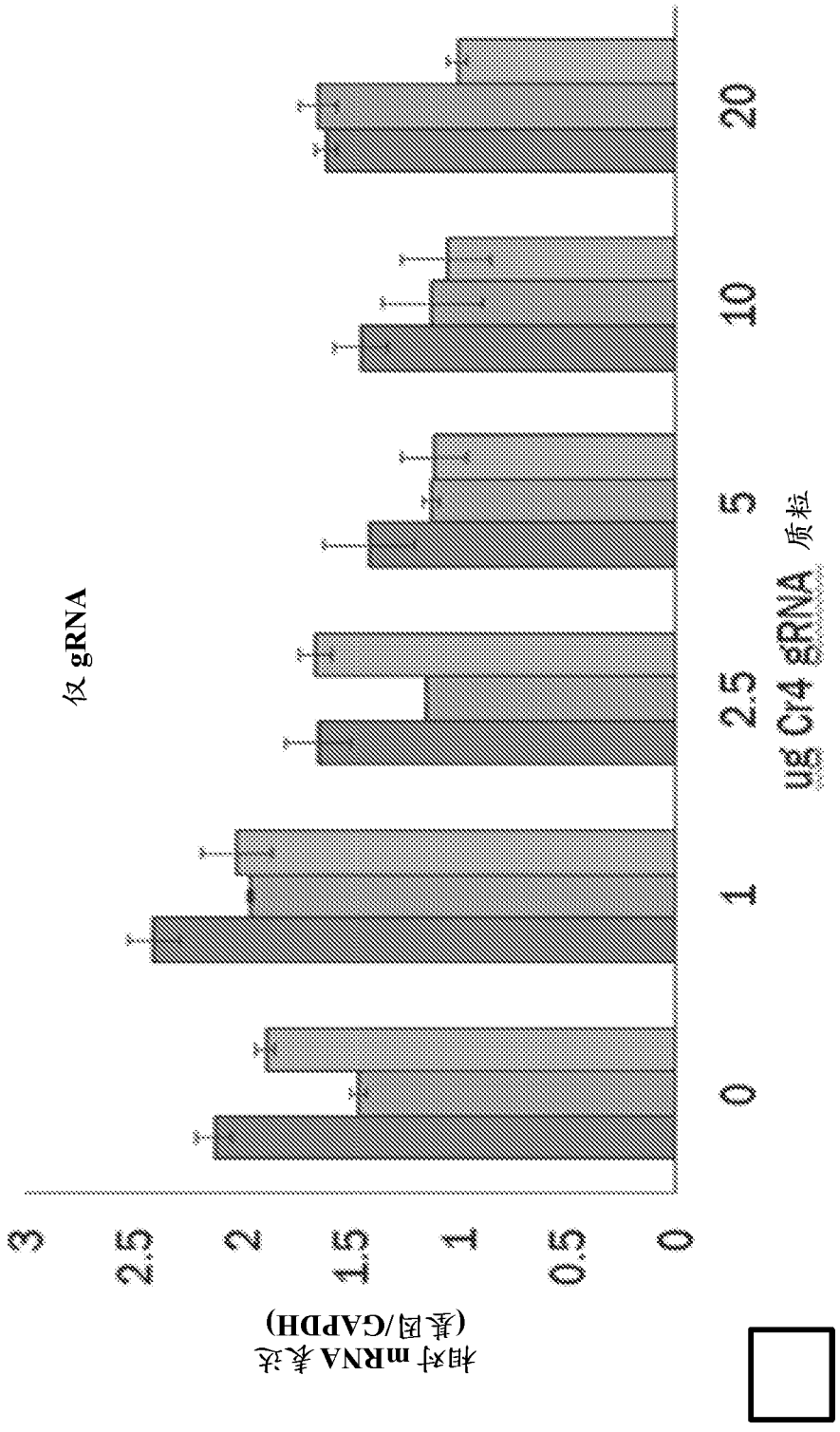

相对 mRNA 表达
(基因/GAPDH)
0
0.5
1
1.5
2
2.5
3
0
1
2.5
5
10
20
μg Cr4 eRNA 质粒
仅 gRNA

图 56A

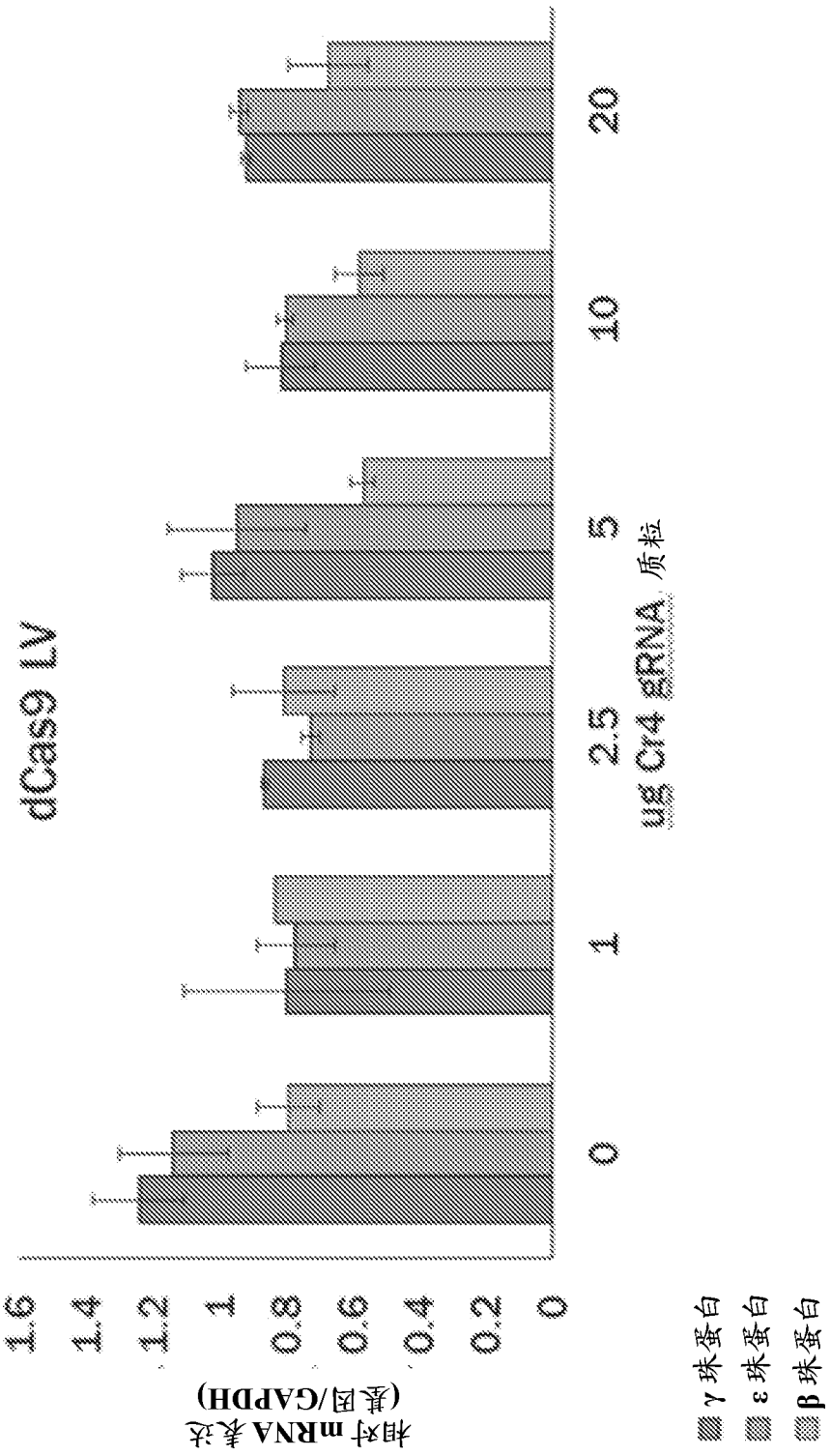

dCas9 LV
相对 mRNA 表达
(基因/GAPDH)
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
1
2.5
5
10
20
ug Cr4 gRNA 质粒
γ 珠蛋白
ε 珠蛋白
β 珠蛋白

图 56B

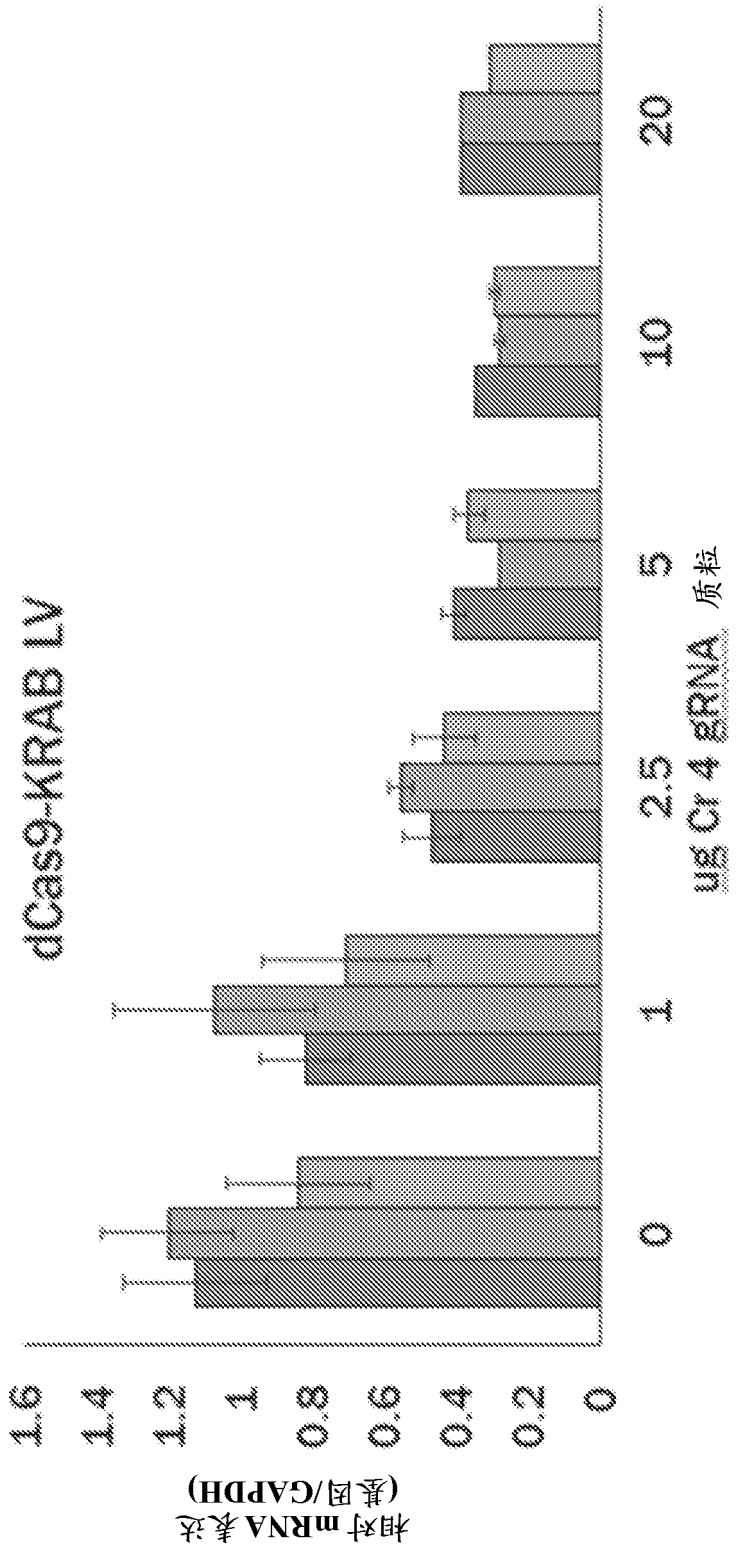
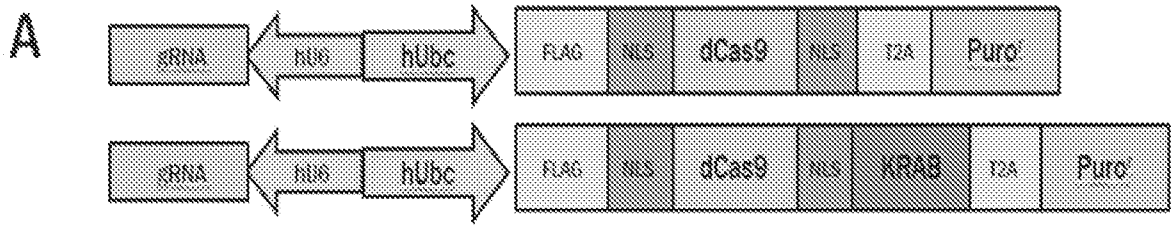

dCas9-KRAB LV
相对 mRNA 表达
(基因/GAPDH)
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
1
2.5
5
10
20
ug Cr 4 gRNA 质粒

图 56C


A
gRNA
hU6
hUbc
FLAG
NLS
dCas9
NLS
T2A
Puro
gRNA
hU6
hUbc
FLAG
NLS
dCas9
NLS
KRAB
T2A
Puro

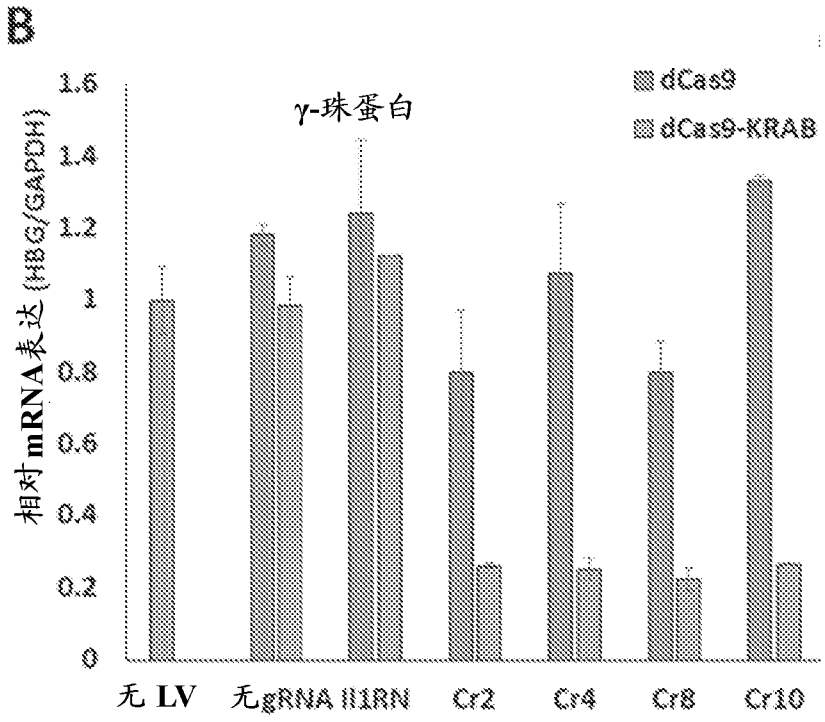

B
γ-珠蛋白
dCas9
dCas9-KRAB
相对mRNA表达 (HBG/GAPDH)
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
无 LV
无gRNA
IL1RN
Cr2
Cr4
Cr8
Cr10

图 57

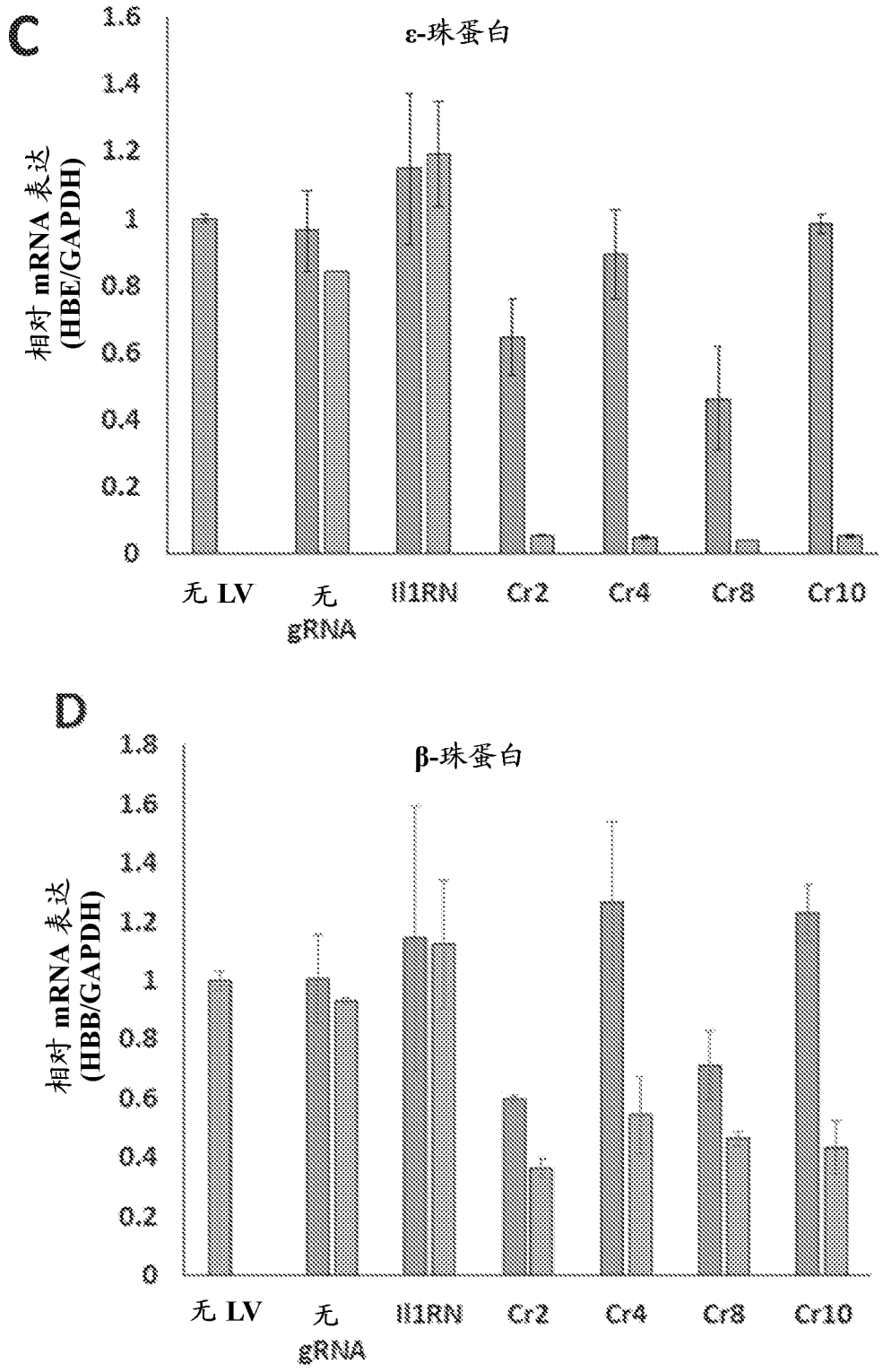

C
ε-珠蛋白
相对 mRNA 表达
(HBE/GAPDH)
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
无 LV
无 gRNA
Il1RN
Cr2
Cr4
Cr8
Cr10
D
β-珠蛋白
相对 mRNA 表达
(HBB/GAPDH)
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
无 LV
无 gRNA
Il1RN
Cr2
Cr4
Cr8
Cr10

图 57(续)

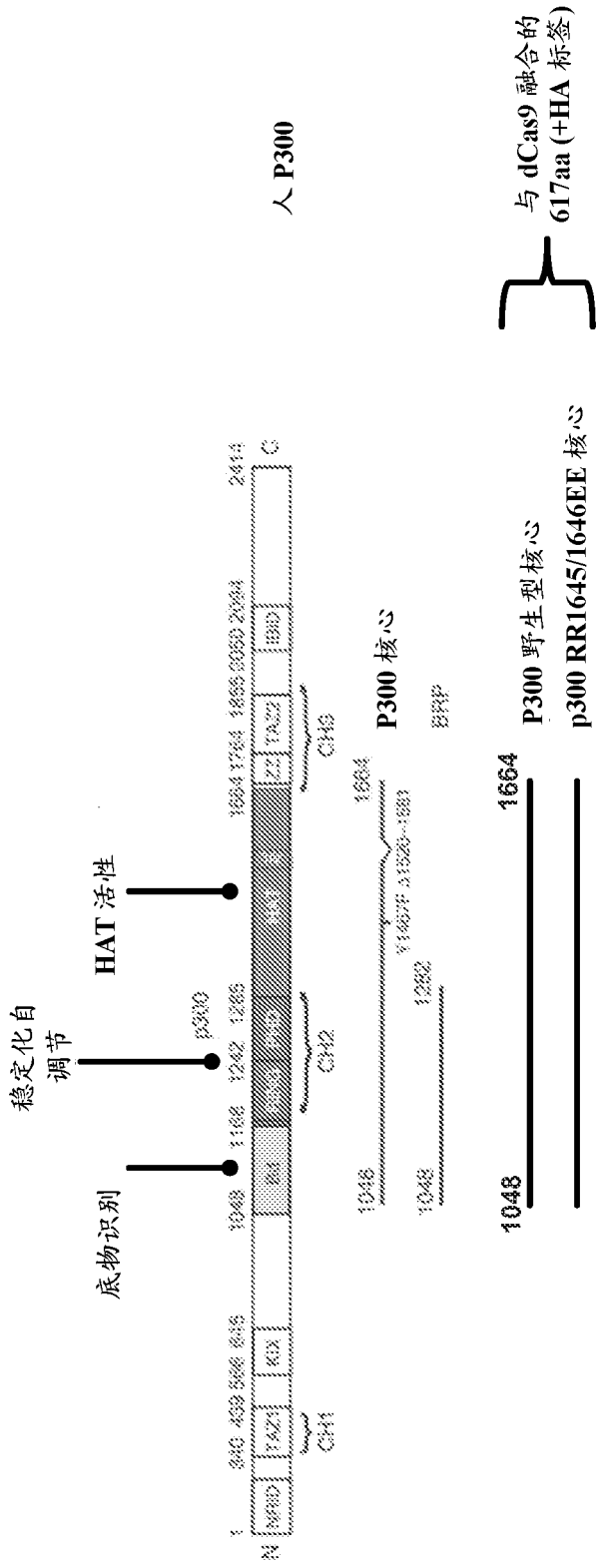

人 P300
与 dCas9 融合的
617aa (+HA 标签)
HAT 活性
稳定化自
调节
底物识别
P300 核心
P300 野生型核心
p300 RR1645/1646EE 核心
1664
1048

图 58

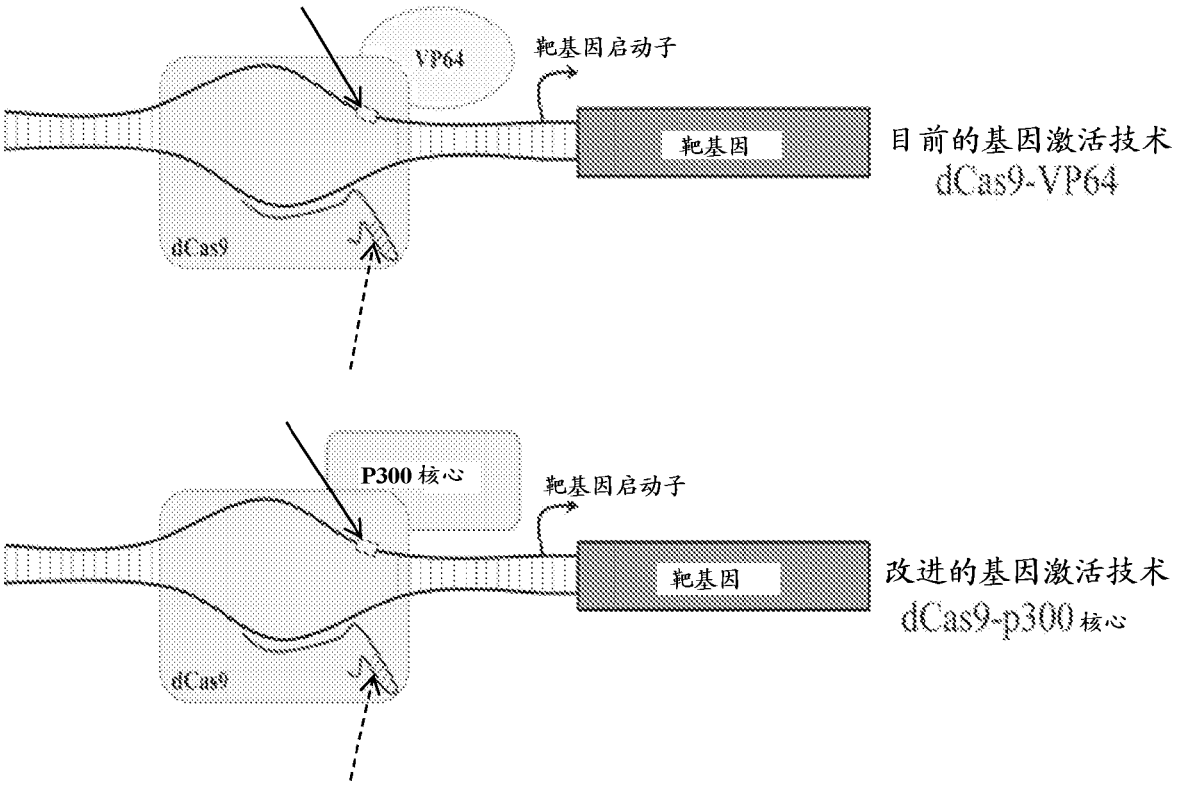

VP64
靶基因启动子
靶基因
目前的基因激活技术
dCas9-VP64
dCas9
P300 核心
靶基因启动子
靶基因
改进的基因激活技术
dCas9-p300 核心
dCas9

图 59

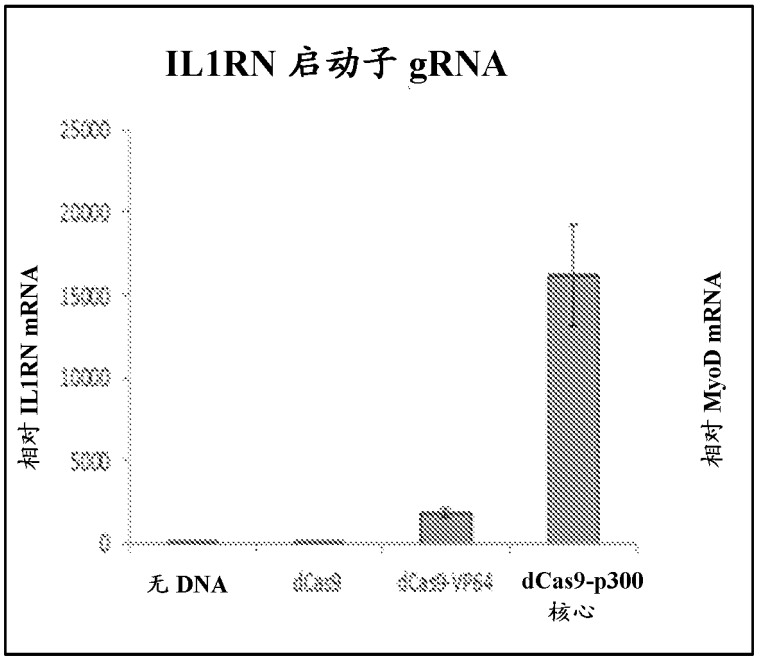

IL1RN 启动子 gRNA
相对 IL1RN mRNA
相对 MyoD mRNA
25000
20000
15000
10000
5000
0
无 DNA
dCas9
dCas9-VP64
dCas9-p300
核心

图 60A

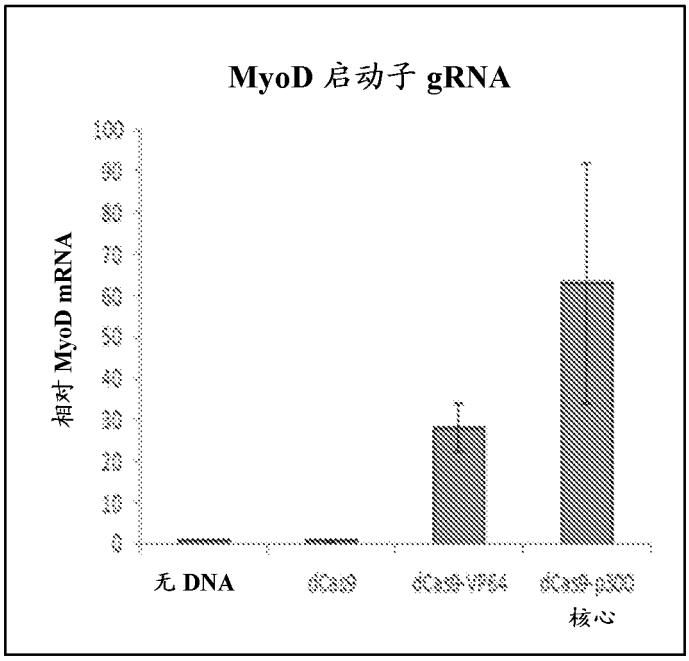

MyoD 启动子 gRNA
相对 MyoD mRNA
无 DNA
dCas9
dCas9-VP64
dCas9-p300
核心

图 60B

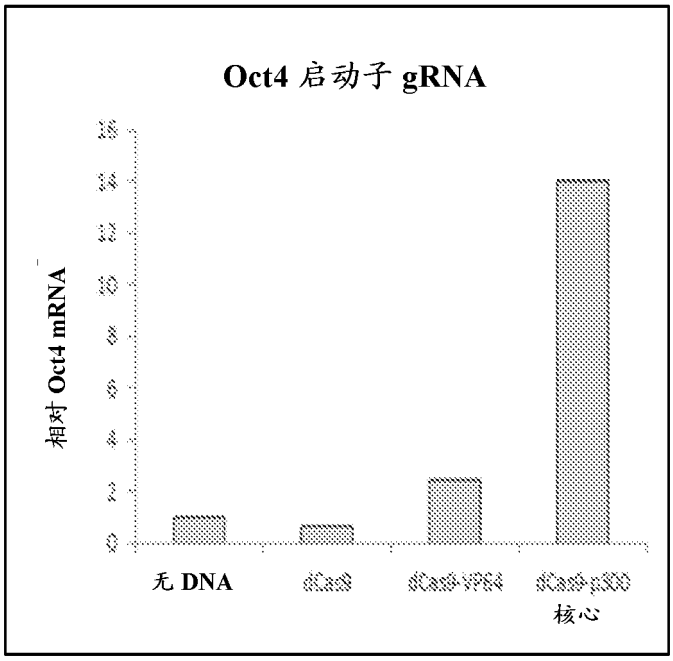

Oct4 启动子 gRNA
相对 Oct4 mRNA
无 DNA
dCas9
dCas9-VP64
dCas9-p300
核心

图 60C

i. dCas9

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGRGMDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEI
FSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDS
TDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASG
VDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ
LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRY
DEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDG
TEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILT
FRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNE
KVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLK
EDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTL
FEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVD
ELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQ
LQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRG
KSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR
QITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAH
DAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNF
FKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGF
SKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLG
ITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE
LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA
NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDA
TLIHQSITGLYETRIDLSQLGGDPKKKRKVG

氨基酸图例:

"Flag"表位
核定位序列
酿脓链球菌 **Cas9 (D10A, H840A)**
VP64 *效应子*
p300 *核心效应子*
"HA"表位

图 61A

ii. dCas9^{VRQR}

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGRQMDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEI
FSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDS
TDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASG
VDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ
LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRY
DEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDG
TEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILT
FRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNE
KVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLK
EDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTL
FEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVD
ELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQ
LQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRG
KSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR
QITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAH
DAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNF
FKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGF
SKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLG
ITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE
LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA
NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDA
TLIHQSITGLYETRIDLSQLGGDPIAGSKASPKKKRKVGRADALDDFDLDMLGSDALDDF
DLDMLGSDALDDFDLDMLGSDALDDFDLDML*INYFYDVPDYAS

图 61B

iii. dCas9$^{p300\text{核心}}$

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGRGMDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEI
FSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDS
TDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASG
VDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ
LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRY
DEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDG
TEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILT
FRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNE
KVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLK
EDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTL
FEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVD
ELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQ
LQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRG
KSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR
QITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAH
DAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNF
FKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGF
SKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLG
ITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE
LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA
NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDA
TLIHQSITGLYETRIDLSQLGGDPIAGSKASPKKKRKVGRAIFKPEELRQALMPTLEALY
RQDPESLPFRQPVDPQLLGIPDYFDIVKSPMDLSTIKRKLDTGQYQEPWQYVDDIWLMFN
NAWLYNRKTSRVYKYCSKLSEVFEQEIDPVMQSLGYCCGRKLEFSPQTLCCYGKQLCTIP
RDATYYSYQNRYHFCEKCFNEIQGESVSLGDDPSQPQTTINKEQFSKRKNDTLDPELFVE
CTECGRKMHQICVLHHEIIWPAGFVCDGCLKKSARTRKENKFSAKRLPSTRLGTFLENRV
NDFLRRQNHPESGEVTVRVVHASDKTVEVKPGMKARFVDSGEMAESFPYRTKALFAFEEI
DGVDLCFFGMHVQEYGSDCPPPNQRRVYISYLDSVHFFRPKCLRTAVYHEILIGYLEYVK
KLGYTTGHIWACPPSEGDDYIFHCHPPDQKIPKPKRLQEWYKKMLDKAVSERIVHDYKDI
FKQATEDRLTSAKELPYFEGDFWPNVLEESIKELEQEEEERKREENTSNESTDVTKGDSK
NAKKKNNKKTSKNKSSLSRGNKKKPGMPNVSNDLSQKLYATMEKHKEVFFVIRLIAGPAA
NSLPPIVDPDPLIPCDLMDGRDAFLTLARDKHLEFSSLRRAQWSTMCMLVELHTQSQDYP
YDVPDYAS

图 61C

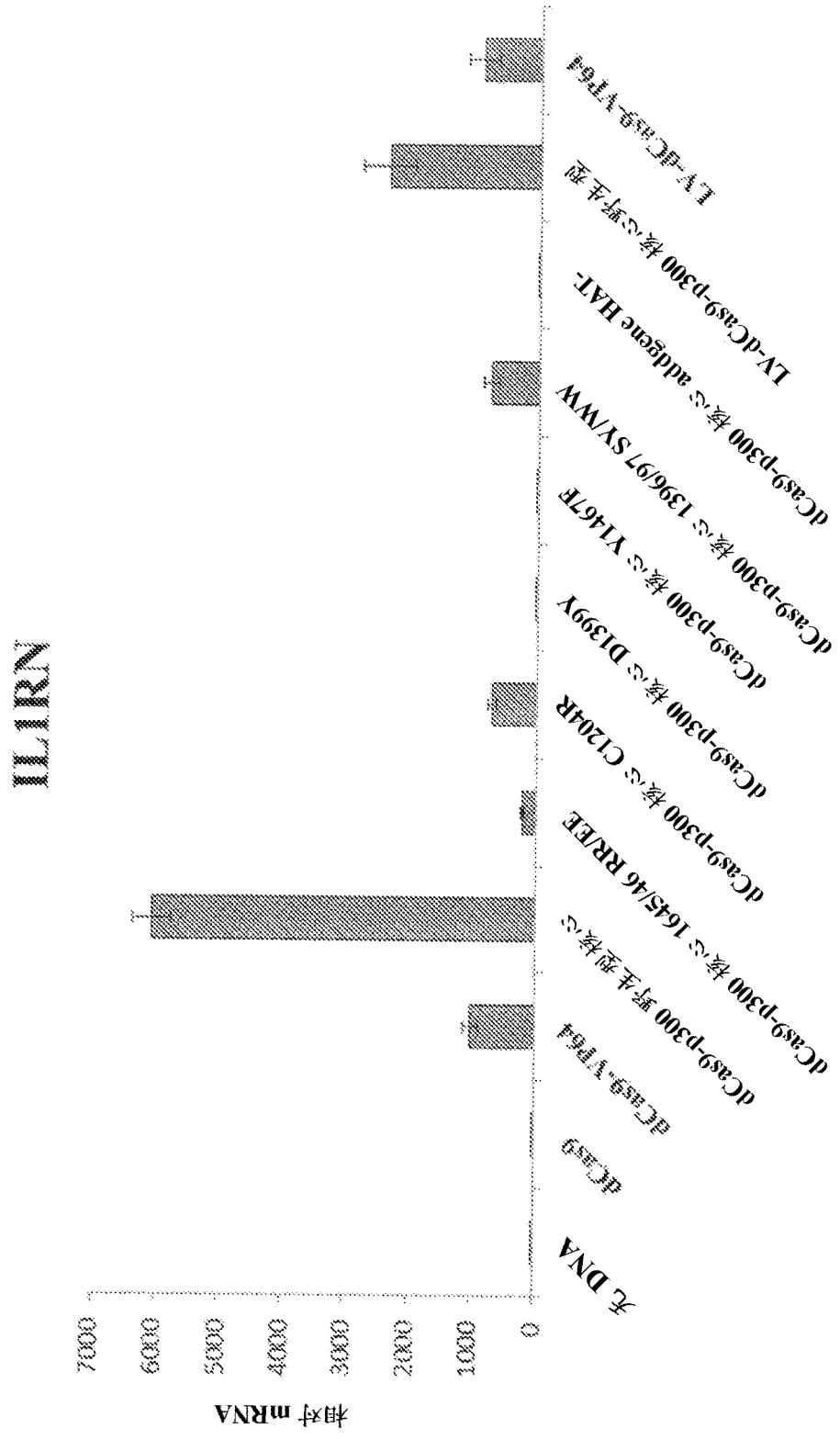

IL1RN
相对 mRNA
7000
6000
5000
4000
3000
2000
1000
0
无 DNA
dCas9
dCas9-VP64
dCas9-D300 野生型核心
dCas9-D300 核心 1645/46 RR/EE
dCas9-D300 核心 C1204R
dCas9-D300 核心 D1399Y
dCas9-D300 核心 V1467F
dCas9-D300 核心 1396/97 SV/WW
dCas9-D300 核心 addgene HAT-
LV-dCas9-D300 核心野生型
LV-dCas9-VP64

图 62

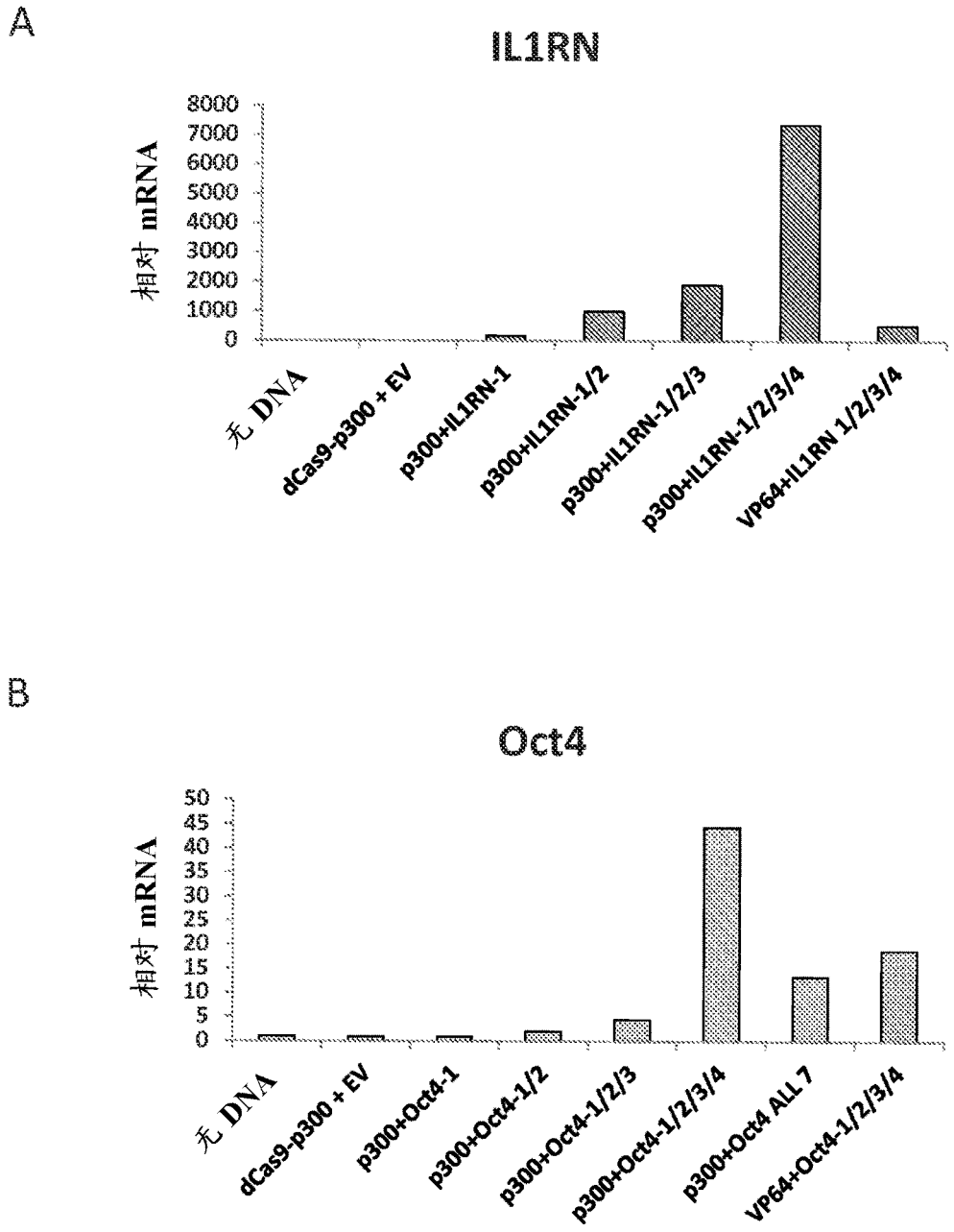

A
IL1RN
相对 mRNA
8000
7000
6000
5000
4000
3000
2000
1000
0
无 DNA
dCas9-p300 + EV
p300+IL1RN-1
p300+IL1RN-1/2
p300+IL1RN-1/2/3
p300+IL1RN-1/2/3/4
VP64+IL1RN 1/2/3/4
B
Oct4
相对 mRNA
50
45
40
35
30
25
20
15
10
5
0
无 DNA
dCas9-p300 + EV
p300+Oct4-1
p300+Oct4-1/2
p300+Oct4-1/2/3
p300+Oct4-1/2/3/4
p300+Oct4 ALL 7
VP64+Oct4-1/2/3/4

图 63

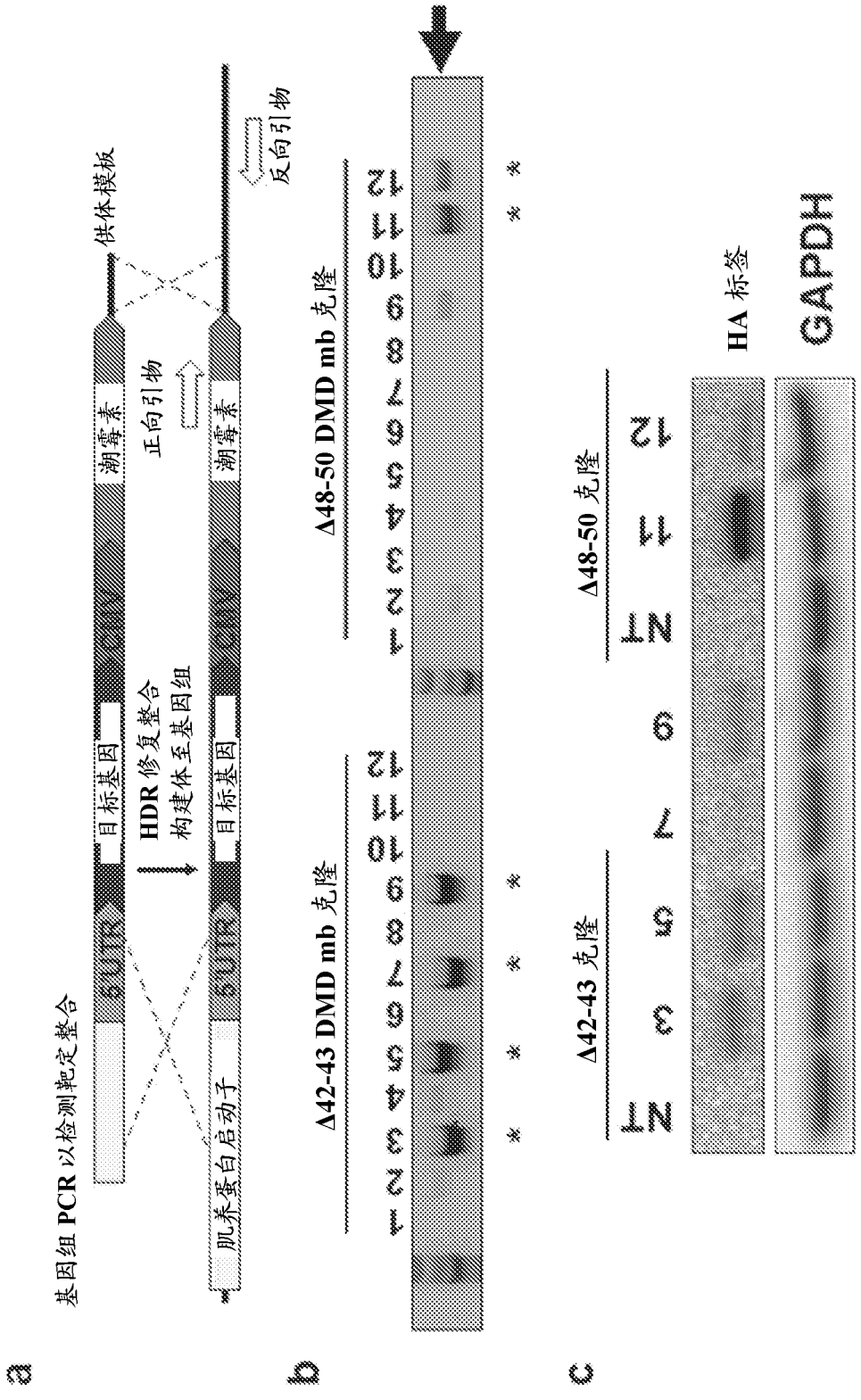

a
基因组 PCR 以检测靶定整合
5'UTR
目标基因
CDS
潮霉素
供体模板
HDR 修复整合
构建体至基因组
正向引物
反向引物
肌养蛋白启动子
b
Δ42-43 DMD mb 克隆
Δ48-50 DMD mb 克隆
c
Δ42-43 克隆
Δ48-50 克隆
HA 标签
GAPDH

图 64