(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 116194578 A
(43)申请公布日 2023.05.30

(21)申请号 202180050873.1
(22)申请日 2021.08.31
(30)优先权数据
63/072,327 2020.08.31 US
63/235,359 2021.08.20 US
(85)PCT国际申请进入国家阶段日
2023.02.17
(86)PCT国际申请的申请数据
PCT/JP2021/031973 2021.08.31
(87)PCT国际申请的公布数据
WO2022/045366 EN 2022.03.03
(71)申请人 摩大力斯医疗株式会社
地址 日本东京都
(72)发明人 山形哲也 秦园博
丽贝卡·温德姆勒
(74)专利代理机构 中原信达知识产权代理有限责任公司 11219
专利代理师 辛怡 杨青
(51)Int.Cl.
C12N 15/09(2006.01)

权利要求书2页 说明书25页
序列表56页 附图8页

(54)发明名称

通过靶向DUX4基因来治疗面肩肱型肌营养不良症(FSHD)的方法

(57)摘要

一种多核苷酸,其包含以下碱基序列:(a)编码核酸酶缺陷型CRISPR效应蛋白与转录阻遏物的融合蛋白的碱基序列,以及(b)编码向导RNA的碱基序列,所述向导RNA靶向人DUX4基因的表达调控区中SEQ ID NO:2、3、4、20、51、68、144、148、152、162、164或167所示的连续区域,所述多核苷酸预计能够用于治疗或预防面肩肱型肌营养不良症(FSHD)。

CN 116194578 A

1.一种多核苷酸,其包含以下碱基序列:

(a)编码核酸酶缺陷型CRISPR效应蛋白与转录阻遏物的融合蛋白的碱基序列,以及

(b)编码向导RNA的碱基序列,所述向导RNA靶向人DUX4基因的表达调控区中SEQ ID NO:2、3、4、20、51、68、144、148、152、162、164或167所示的连续区域。

2.根据权利要求1所述的多核苷酸,其中所述编码向导RNA的碱基序列包含SEQ ID NO:2、3、4、20、51、68、144、148、152、162、164或167所示的碱基序列,或者其中缺失、替换、插入和/或添加1至3个碱基的SEQ ID NO:2、3、4、20、51、68、144、148、152、162、164或167所示的碱基序列。

3.根据权利要求1或2所述的多核苷酸,其包含至少两个不同的所述编码向导RNA的碱基序列。

4.根据权利要求1至3中任一项所述的多核苷酸,其中所述转录阻遏物选自KRAB、MeCP2、SIN3A、HDT1、MBD2B、NIPP1和HP1A。

5.根据权利要求4所述的多核苷酸,其中所述转录阻遏物是KRAB。

6.根据权利要求1至5中任一项所述的多核苷酸,其中所述核酸酶缺陷型CRISPR效应蛋白是dCas9。

7.根据权利要求6所述的多核苷酸,其中所述dCas9来源于金黄色葡萄球菌(Staphylococcus aureus)。

8.根据权利要求1至7中任一项所述的多核苷酸,其还包含所述编码向导RNA的碱基序列的启动子序列和/或所述编码核酸酶缺陷型CRISPR效应蛋白与转录阻遏物的融合蛋白的碱基序列的启动子序列。

9.根据权利要求8所述的多核苷酸,其中所述编码向导RNA的碱基序列的启动子序列选自U6启动子、SNR6启动子、SNR52启动子、SCR1启动子、RPR1启动子、U3启动子和H1启动子。

10.根据权利要求9所述的多核苷酸,其中所述编码向导RNA的碱基序列的启动子序列是U6启动子。

11.根据权利要求8至10中任一项所述的多核苷酸,其中所述编码核酸酶缺陷型CRISPR效应蛋白与转录阻遏物的融合蛋白的碱基序列的启动子序列是广泛性启动子或神经元特异性启动子。

12.根据权利要求11所述的多核苷酸,其中所述广泛性启动子选自EFS启动子、CMV启动子和CAG启动子。

13.一种载体,其包含根据权利要求1至12中任一项所述的多核苷酸。

14.根据权利要求13所述的载体,其中所述载体是质粒载体或病毒载体。

15.根据权利要求14所述的载体,其中所述病毒载体选自腺相关病毒(AAV)载体、腺病毒载体和慢病毒载体。

16.根据权利要求15所述的载体,其中所述AAV载体选自AAV1、AAV2、AAV6、AAV7、AAV8、AAV9、Anc80、AAV_{587}MTP、AAV_{588}MTP、AAV-B1、AAVM41和AAVrh74。

17.根据权利要求16所述的载体,其中所述AAV载体是AAV9。

18.一种药物组合物,其包含根据权利要求1至12中任一项所述的多核苷酸或根据权利要求13至17中任一项所述的载体。

19.根据权利要求18所述的药物组合物,其用于治疗或预防FSHD。

20.一种治疗或预防FSHD的方法,其包括向有需要的受试者施用根据权利要求1至12中任一项所述的多核苷酸或根据权利要求13至17中任一项所述的载体。

通过靶向DUX4基因来治疗面肩肱型肌营养不良症(FSHD)的方法

技术领域

[0001] 本发明涉及通过靶向人双同源框蛋白4(DUX4)基因等来治疗面肩肱型肌营养不良症(FSHD)的方法。更具体地,本发明涉及通过使用靶向人DUX4基因的特定序列的向导RNA以及转录抑制因子与CRISPR效应蛋白的融合蛋白等抑制人DUX4基因表达来治疗或预防FSHD的方法和药剂。

背景技术

[0002] 面肩肱型肌营养不良症(FSHD)是最普遍的肌病之一,影响着所有年龄的男性和女性。

[0003] FSHD有两种类型。“FSHD1”归因于4号染色体上端粒(4q35)基因组序列(D4Z4)的缩短重复序列(10个重复序列或更少),而“FSHD2”归因于除FSHD1以外的复杂因素。在正常的D4Z4重复序列中,DNA高度甲基化。在FSHD1和FSHD2中,由于相应的基因组异常,染色质结构发生变化,伴随DNA低甲基化,并且最初不在肌肉(祖细胞)细胞中表达的基因(DUX4转录因子)被激活。尽管DUX4蛋白在发育阶段很重要,但它通常不存在于成熟细胞中,并且已知FSHD骨骼肌中的DUX4激活会导致细胞死亡。预防DUX4激活预计治疗FSHD,作为其一部分,已经尝试通过使用基因编辑技术减少DUX4 mRNA的量(非专利文献1)。

[0004] 另一方面,近年来已经开发了使用具有失活核酸酶活性的Cas9(dCas9)与转录激活结构域或转录阻遏结构域的组合的系统,其中通过使用向导RNA使蛋白质靶向基因而不切割基因的DNA序列来控制靶基因的表达(专利文献1,其以引用的方式整体并入本文)。其临床应用前景广阔(参见非专利文献2,其以引用的方式整体并入本文)。然而,存在编码dCas9、向导RNA和共转录阻遏物的复合物的序列超出最常用的病毒载体(例如,AAV)的容量的问题,所述病毒载体代表了最有前景的体内基因递送方法(参见非专利文献3,其以引用的方式整体并入本文)。

[0005] 引用列表

[0006] 专利文献

[0007] [专利文献1]WO2013/176772

[0008] 非专利文献

[0009] [非专利文献1]Mol Ther.2016年3月;24(3):527-535

[0010] [非专利文献2]Dominguez A.等人,Nat Rev Mol Cell Biol.2016年1月;17(1):5-15

[0011] [非专利文献3]Liao H.等人,Cell.2017年12月14日;171(7):1495-507

发明内容

[0012] 技术问题

[0013] 因此,本发明的一个目标是提供面肩肱型肌营养不良症(FSHD)的新颖治疗方法。

[0014] 已经通过本申请发明人的发现，即通过使用靶向人DUX4基因的特定序列的向导RNA和转录阻遏物与核酸酶缺陷型CRISPR效应蛋白的融合蛋白，可以强烈地抑制人DUX4基因(基因ID:100288687)的表达，实现了这个目标和其它目标，所述目标将在下面的详细描述过程中变得显而易见。此外，本申请发明人已经发现，使用紧凑型核酸酶缺陷型CRISPR效应蛋白和紧凑型转录阻遏物时，携带编码融合蛋白的碱基序列和编码向导RNA的碱基序列的单个AAV载体可以强烈地抑制人DUX4基因的表达。

[0015] 因此，本发明提供了：

[0016] [1]一种多核苷酸，其包含以下碱基序列：

[0017] (a)编码核酸酶缺陷型CRISPR效应蛋白与转录阻遏物的融合蛋白的碱基序列，以及

[0018] (b)编码向导RNA的碱基序列，所述向导RNA靶向人DUX4基因的表达调控区中SEQ ID NO:2、3、4、20、51、68、138、142、146、156、158或161所示的连续区域。

[0019] [2]根据[1]所述的多核苷酸，其中所述编码向导RNA的碱基序列包含SEQ ID NO:2、3、4、20、51、68、138、142、146、156、158或161所示的碱基序列，或者其中缺失、替换、插入和/或添加1至3个碱基的SEQ ID NO:2、3、4、20、51、68、138、142、146、156、158或161所示的碱基序列。

[0020] [3]根据[1]或[2]所述的多核苷酸，其包含至少两个不同的所述编码向导RNA的碱基序列。

[0021] [4]根据[1]至[3]中任一项所述的多核苷酸，其中所述转录阻遏物选自KRAB、MeCP2、SIN3A、HDT1、MBD2B、NIPP1和HP1A。

[0022] [5]根据[4]所述的多核苷酸，其中所述转录阻遏物是KRAB。

[0023] [6]根据[1]至[5]中任一项所述的多核苷酸，其中所述核酸酶缺陷型CRISPR效应蛋白是dCas9。

[0024] [7]根据[6]所述的多核苷酸，其中所述dCas9来源于金黄色葡萄球菌(Staphylococcus aureus)。

[0025] [8]根据[1]至[7]中任一项所述的多核苷酸，其还包含所述编码向导RNA的碱基序列的启动子序列和/或所述编码核酸酶缺陷型CRISPR效应蛋白与转录阻遏物的融合蛋白的碱基序列的启动子序列。

[0026] [9]根据[8]所述的多核苷酸，其中所述编码向导RNA的碱基序列的启动子序列选自U6启动子、SNR6启动子、SNR52启动子、SCR1启动子、RPR1启动子、U3启动子和H1启动子。

[0027] [10]根据[9]所述的多核苷酸，其中所述编码向导RNA的碱基序列的启动子序列是U6启动子。

[0028] [11]根据[8]至[10]中任一项所述的多核苷酸，其中所述编码核酸酶缺陷型CRISPR效应蛋白与转录阻遏物的融合蛋白的碱基序列的启动子序列是广泛性启动子或神经元特异性启动子。

[0029] [12]根据[11]所述的多核苷酸，其中所述广泛性启动子选自EFS启动子、CMV启动子和CAG启动子。

[0030] [13]一种载体，其包含根据[1]至[12]中任一项所述的多核苷酸。

[0031] [14]根据[13]所述的载体，其中所述载体是质粒载体或病毒载体。

[0032] [15]根据[14]所述的载体,其中所述病毒载体选自腺相关病毒(AAV)载体、腺病毒载体和慢病毒载体。

[0033] [16]根据[15]所述的载体,其中所述AAV载体选自AAV1、AAV2、AAV6、AAV7、AAV8、AAV9、Anc80、AAV_{587}MTP、AAV_{588}MTP、AAV-B1、AAVM41和AAVrh74。

[0034] [17]根据[16]所述的载体,其中所述AAV载体是AAV9。

[0035] [18]一种药物组合物,其包含根据[1]至[12]中任一项所述的多核苷酸或根据[13]至[17]中任一项所述的载体。

[0036] [19]根据[18]所述的药物组合物,其用于治疗或预防FSHD。

[0037] [20]一种治疗或预防FSHD的方法,其包括向有需要的受试者施用根据[1]至[12]中任一项所述的多核苷酸或根据[13]至[17]中任一项所述的载体。

[0038] 对本发明的更完整了解和许多伴随优点将容易获得,因为通过参考下面的详细描述结合附图考虑时,它们将变得更好理解。

[0039] 有益效果

[0040] 根据本发明,可以抑制人DUX4基因的表达,因此,本发明预计能够治疗FSHD。

附图说明

[0041] [图1]图1示出了所靶向的基因组区域相对于人DUX4基因的位置。

[0042] [图2]图2示出了通过使用含有由SEQ ID NO:1至76所示的靶向序列编码的crRNA的sgRNA对来源于FSHD患者的两个成淋巴细胞细胞系(LCL;GM16343 LCL和GM16414 LCL)中人DUX4基因的表达抑制作用的评估结果。水平轴示出了含有由各个靶向序列编码的crRNA的sgRNA,垂直轴示出了使用各个sgRNA时的DUX4基因表达水平与使用对照sgRNA时的DUX4基因表达水平(作为1)的比率。

[0043] [图3A]图3A示出了通过使用含有由所选择的27个靶向序列编码的crRNA的sgRNA对GM16343 LCL中人DUX4基因的表达抑制作用的评估结果。水平轴示出了含有由各个靶向序列编码的crRNA的sgRNA,垂直轴示出了使用各个sgRNA时的DUX4基因表达水平与使用对照sgRNA时的DUX4基因表达水平(作为1)的比率。

[0044] [图3B]图3B示出了通过使用含有由所选择的27个靶向序列编码的crRNA的sgRNA对GM16414 LCL中人DUX4基因的表达抑制作用的评估结果。水平轴示出了含有由各个靶向序列编码的crRNA的sgRNA,垂直轴示出了使用各个sgRNA时的DUX4基因表达水平与使用对照sgRNA时的DUX4基因表达水平(作为1)的比率。

[0045] [图4]图4示出了来自使用FSHD患者来源的LCL的验证实验中所鉴定的6个最佳sgRNA的DUX4和FSHD生物标志物TRIM43、MBD3L2和ZSCAN4的定量评估结果。水平轴示出了含有由各个靶向序列编码的crRNA的sgRNA,垂直轴示出了使用各个sgRNA时的DUX4基因表达水平与使用对照sgRNA时的DUX4基因表达水平(作为1)的比率。

[0046] [图5]图5示出了所靶向的基因组区域相对于人DUX4基因的位置。

[0047] [图6]图6示出了通过使用含有由SEQ ID NO:104至188所示的靶向序列编码的crRNA的sgRNA对来源于FSHD患者的一个成淋巴细胞细胞系(LCL;GM16343 LCL)中人DUX4基因的表达抑制作用的评估结果。水平轴示出了含有由各个靶向序列编码的crRNA的sgRNA,垂直轴示出了使用各个sgRNA时的DUX4基因表达水平与使用对照sgRNA时的DUX4基因表达

水平(作为1)的比率。

[0048] [图7]图7示出了来自使用FSHD患者来源的LCL的验证实验(N＝2)中所鉴定的6个最佳sgRNA的DUX4和FSHD生物标志物TRIM43、MBD3L2和ZSCAN4的定量评估结果。水平轴示出了含有由各个靶向序列编码的crRNA的sgRNA,垂直轴示出了使用各个sgRNA时的DUX4基因表达水平与使用对照sgRNA时的DUX4基因表达水平(作为1)的比率。

具体实施方式

[0049] 下面详细说明本发明的实施方案。

[0050] 1.多核苷酸

[0051] 本发明提供了一种包含以下碱基序列的多核苷酸(下文有时也称为“本发明的多核苷酸”):

[0052] (a)编码核酸酶缺陷型CRISPR效应蛋白与转录阻遏物的融合蛋白的碱基序列,以及

[0053] (b)编码向导RNA的碱基序列,所述向导RNA靶向人DUX4基因的表达调控区中SEQ ID NO:2、3、4、20、51、68、138、142、146、156、158或161所示的连续区域。

[0054] 将本发明的多核苷酸引入到所期望的细胞中并转录以产生核酸酶缺陷型CRISPR效应蛋白与转录阻遏物的融合蛋白,以及靶向人DUX4基因的表达调控区的特定区域的向导RNA。这些融合蛋白和向导RNA形成了复合物(在下文,所述复合物有时称为“核糖核蛋白;RNP”)并协同作用于上述特定区域,从而抑制人DUX4基因的转录。在本发明的一个实施方案中,人DUX4基因的表达可以被抑制例如不低于约40％、不低于约50％、不低于约60％、不低于约70％、不低于约75％、不低于约80％、不低于约85％、不低于约90％、不低于约95％或约100％。

[0055] (1)定义

[0056] 在本说明书中,“人双同源框蛋白4(DUX4)基因的表达调控区”意指其中可以通过RNP与该区域结合来抑制人DUX4基因表达的任何区域。即,人DUX4基因的表达调控区可以存在于人DUX4基因的任何区域如启动子区域、增强子区域、内含子和外显子中,只要RNP的结合抑制人DUX4基因的表达即可。在本说明书中,当表达调控区由特定序列表示时,表达调控区在概念上包括有义链序列和反义链序列。

[0057] 在本发明中,通过向导RNA将核酸酶缺陷型CRISPR效应蛋白与转录阻遏物的融合蛋白募集到人DUX4基因的表达调控区的特定区域中。在本说明书中,“靶向……的向导RNA”意指“将融合蛋白募集到……中的向导RNA”。

[0058] 在本说明书中,“向导RNA(也称为‘gRNA’)”是一种包含基因组特异性CRISPR-RNA(称为‘crRNA’)的RNA。crRNA是一种与靶向序列(稍后描述)的互补序列结合的RNA。当Cpf1用作CRISPR效应蛋白时,“向导RNA”是指包含由crRNA和附接至其5’末端的特定序列(例如,在FnCpf1的情况下,SEQ ID NO:80所示的RNA序列)组成的RNA的RNA。当Cas9用作CRISPR效应蛋白时,“向导RNA”是指包含crRNA和附接至其3’末端的反式激活crRNA(称为“tracrRNA”)的嵌合体RNA(称为“单向导RNA(sgRNA)”)(参见例如Zhang F.等人,Hum Mol Genet.2014年9月15日;23(R1):R40-6和Zetsche B.等人,Cell.2015年10月22日;163(3):759-71,所述文献以全文引用的方式整体并入本文)。

[0059] 在本说明书中,与crRNA在人DUX4基因的表达调控区中所结合的序列互补的序列称为“靶向序列”。即,在本说明书中,“靶向序列”是存在于人DUX4基因的表达调控区中并与PAM(原间隔子相邻基序)相邻的DNA序列。当Cpf1用作CRISPR效应蛋白时,PAM与靶向序列的5’侧相邻。当Cas9用作CRISPR效应蛋白时,PAM与靶向序列的3’侧相邻。靶向序列可以存在于人DUX4基因的表达调控区的有义链序列侧或反义链序列侧(参见例如上述Zhang F.等人,Hum Mol Genet.2014年9月15日;23(R1):R40-6和Zetsche B.等人,Cell.2015年10月22日;163(3):759-71,所述文献以引用的方式整体并入本文)。

[0060] (2)核酸酶缺陷型CRISPR效应蛋白

[0061] 在本发明中,使用核酸酶缺陷型CRISPR效应蛋白,将与其融合的转录阻遏物募集到人DUX4基因的表达调控区。本发明所用的核酸酶缺陷型CRISPR效应蛋白(下文简称为“CRISPR效应蛋白”)不受特别限制,只要它与gRNA形成复合物并且被募集到人DUX4基因的表达调控区即可。例如,可以包括核酸酶缺陷型Cas9(下文有时也称为“dCas9”)或核酸酶缺陷型Cpf1(下文有时也称为“dCpf1”)。

[0062] 上述dCas9的实例包括但不限于酿脓链球菌(Streptococcus pyogenes)来源的Cas9(SpCas9;PAM序列:NGG(N为A、G、T或C。在下文中相同))、嗜热链球菌(Streptococcus thermophilus)来源的Cas9(StCas9;PAM序列:NNAGAAW(W为A或T。在下文中相同))、脑膜炎奈瑟氏球菌(Neisseria meningitidis)来源的Cas9(NmCas9;PAM序列:NNNNGATT)或金黄色葡萄球菌来源的Cas9(SaCas9;PAM序列:NNGRRT(R为A或G。在下文中相同))等的核酸酶缺陷型变体(参见例如Nishimasu等人,Cell.2014年2月27日;156(5):935-49,Esvelt KM等人,Nat Methods.2013年11月;10(11):1116-21,Zhang Y.Mol Cell.2015年10月15日;60(2):242-55,以及Friedland AE等人,Genome Biol.2015年11月24日;16:257,所述文献以引用的方式整体并入本文)。例如,在SpCas9的情况下,可以使用其中第10位Asp残基转变为Ala残基且第840位His残基转变为Ala残基的双突变体(有时称为“dSpCas9”)(参见例如上述Nishimasu等人,Cell.2014)。或者,在SaCas9的情况下,可以使用其中第10位Asp残基转变为Ala残基且第580位Asn残基转变为Ala残基的双突变体(SEQ ID NO:81)或其中第10位Asp残基转变为Ala残基且第557位His残基转变为Ala残基的双突变体(SEQ ID NO:82)(在下文中,这些双突变体中的任一个有时称为“dSaCas9”)(参见例如上述Friedland AE等人,Genome Biol.2015,所述文献以引用的方式整体并入本文)。

[0063] 此外,在本发明的一个实施方案中,作为dCas9,还可以使用通过修饰上述dCas9的氨基酸序列的一部分而获得的变体,所述变体与gRNA形成复合物并且被募集到人DUX4基因的表达调控区。此类变体的实例包括具有部分缺失氨基酸序列的截短变体。在本发明的一个实施方案中,作为dCas9,可以使用WO2019/235627和WO2020/085441所公开的变体,所述文献以引用的方式整体并入本文。具体地,可以使用通过从作为其中第10位Asp残基转变为Ala残基且第580位Asn残基转变为Ala残基的双突变体的dSaCas9中缺失第721个至第745个氨基酸而获得的dSaCas9(SEQ ID NO:83),或其中所述缺失部分被肽接头替换的dSaCas9(例如其中所述缺失部分被GGSGGS接头(SEQ ID NO:84)替换的dSaCas9如SEQ ID NO:85所示,其中所述缺失部分被SGGGS接头(SEQ ID NO:86)替换的dSaCas9如SEQ ID NO:87所示,等等)(在下文中,这些双突变体中的任一个有时称为“dSaCas9[-25]”),或者通过从作为上述双突变体的dSaCas9中缺失第482个至第648个氨基酸而获得的dSaCas9(SEQ ID NO:88),

或其中所述缺失部分被肽接头替换的dSaCas9(其中所述缺失部分被GGSGGS接头替换的dSaCas9如SEQ ID NO:89所示)。

[0064] 上述dCpf1的实例包括但不限于新凶手弗朗西丝菌(Francisella novicida)来源的Cpf1(FnCpf1;PAM序列:NTT)、氨基酸球菌属(Acidaminococcus sp.)来源的Cpf1(AsCpf1;PAM序列:NTTT)或毛螺旋菌科(Lachnospiraceae)细菌来源的Cpf1(LbCpf1;PAM序列:NTTT)等的核酸酶缺陷型变体(参见例如Zetsche B.等人,Cell.2015年10月22日;163(3):759-71,Yamano T等人,Cell.2016年5月5日;165(4):949-62,以及Yamano T等人,Mol Cell.2017年8月17日;67(4):633-45,所述文献以引用的方式整体并入本文)。例如,在FnCpf1的情况下,可以使用其中第917位Asp残基转变为Ala残基且第1006位Glu残基转变为Ala残基的双突变体(参见例如上述Zetsche B等人,Cell.2015,所述文献以引用的方式整体并入本文)。在本发明的一个实施方案中,作为dCpf1,还可以使用通过修饰上述dCpf1的氨基酸序列的一部分而获得的变体,所述变体与gRNA形成复合物并且被募集到人DUX4基因的表达调控区。

[0065] 在本发明的一个实施方案中,dCas9用作核酸酶缺陷型CRISPR效应蛋白。在一个实施方案中,dCas9是dSaCas9,在一个具体实施方案中,dSaCas9是dSaCas9[-25]。

[0066] 包含编码CRISPR效应蛋白的碱基序列的多核苷酸可以例如以下方法进行克隆:基于其cDNA序列信息合成覆盖编码所述蛋白的所期望部分的区域的oligoDNA引物,并使用从产生所述蛋白的细胞制备的总RNA或mRNA级分作为模板,通过PCR方法扩增所述多核苷酸。此外,包含编码核酸酶缺陷型CRISPR效应蛋白的碱基序列的多核苷酸可以通过在对针对DNA切割活性非常重要的位点上利用已知的定点诱变方法将突变引入到编码所克隆的CRISPR效应蛋白的核苷酸序列中从而将氨基酸残基(例如,在SaCas9的情况下可以包括第10位Asp残基、第557位His残基和第580位Asn残基;在FnCpf1的情况下可以包括第917位Asp残基和第1006位Glu残基等等,但不限于这些)转变为其它氨基酸来获得。

[0067] 或者,包含编码核酸酶缺陷型CRISPR效应蛋白的碱基序列的多核苷酸可以基于其cDNA序列信息通过化学合成或化学合成与PCR方法或吉布森组装方法的组合来获得,也可以进一步构建为经过密码子优化以得到适合在人体内表达的密码子的碱基序列。

[0068] (3)转录阻遏物

[0069] 在本发明中,人DUX4基因表达在与核酸酶缺陷型CRISPR效应蛋白融合的转录阻遏物的作用下受到阻遏。在本说明书中,“转录阻遏物”意指能够阻遏人DUX4基因的基因转录的蛋白质或保留其功能的肽片段。本发明所用的转录阻遏物不受特别限制,只要它可以阻遏人DUX4基因的表达即可。它包括例如克鲁佩尔相关框(KRAB)、MBD2B、v-ErbA、SID(包括SID链态(SID4X))、MBD2、MBD3、DNMT家族(例如DNMT1、DNMT3A、DNMT3B)、Rb、MeCP2、ROM2、LSD1、AtHD2A、SET1、HDAC11、SETD8、EZH2、SUV39H1、PHF19、SALI、NUE、SUVR4、KYP、DIM5、HDAC8、SIRT3、SIRT6、MESOLO4、SET8、HST2、COBB、SET-TAF1B、NCOR、SIN3A、HDT1、NIPP1、HP1A、ERF阻遏结构域(ERD)及其具有转录阻遏能力的变体、其融合物等等。在本发明的一个实施方案中,KRAB用作转录阻遏物。

[0070] 包含编码转录阻遏物的碱基序列的多核苷酸可以通过化学合成或化学合成与PCR方法或吉布森组装方法的组合来构建。此外,包含编码转录阻遏物的碱基序列的多核苷酸还可以构建为密码子优化的DNA序列,以成为适合在人体内表达的密码子。

[0071] 包含编码转录阻遏物与核酸酶缺陷型CRISPR效应蛋白的融合蛋白的碱基序列的多核苷酸可以通过直接或在添加编码接头、NLS(核定位信号)的碱基序列(例如SEQ ID NO:90或SEQ ID NO:91所示的碱基序列)、标签和/或其它后将编码CRISPR效应蛋白的碱基序列连接至编码转录阻遏物的碱基序列来制备。在本发明中,转录阻遏物可以与核酸酶缺陷型CRISPR效应蛋白的N末端或C末端融合。作为接头,可以使用氨基酸数为约2至50的接头,其具体实例包括但不限于其中甘氨酸(G)和丝氨酸(S)交替连接的G-S-G-S接头等。在本发明的一个实施方案中,作为包含编码核酸酶缺陷型CRISPR效应蛋白与转录阻遏物的融合蛋白的碱基序列的多核苷酸,可以使用SEQ ID NO:92所示的碱基序列,所述碱基序列将SV40 NLS、dSaCas9(例如D10A和N580A突变体)、NLS和KRAB编码为融合蛋白。如果需要,则可以含有其它碱基序列(参见下面的“(6)其它碱基序列”)和选择标志物(例如Puro)。

[0072] (4)向导RNA

[0073] 在本发明中,核酸酶缺陷型CRISPR效应蛋白与转录阻遏物的融合蛋白可以通过向导RNA募集到人DUX4基因的表达调控区。如上述“(1)定义”所述,向导RNA包含crRNA,crRNA与靶向序列的互补序列结合。crRNA可以与靶向序列的互补序列不完全互补,只要向导RNA可以将融合蛋白募集到靶区域即可,并且可以包含其中缺失、替换、插入和/或添加至少1至3个碱基的靶向序列的碱基序列。

[0074] 例如,当dCas9用作核酸酶缺陷型CRISPR效应蛋白时,靶向序列可以使用公开gRNA设计网站(CRISPR Design Tool、CRISPR direct等)来确定。具体地,从靶基因(即,人DUX4基因)序列中,列出了PAM(例如在SaCas9的情况下为NNGRRT)与其3’侧相邻的长度为约20个核苷酸的候选靶向序列,在这些候选靶向序列中,在人类基因组中具有少量脱靶位点的候选靶向序列可以用作靶向序列。靶向序列的碱基长度为18至24个核苷酸长度,优选为20至23个核苷酸长度,更优选为21至23个核苷酸长度。作为预测脱靶位点数的初步筛检,许多生物信息学工具是已知和公开可用的,并且可以用于预测具有最低脱靶效应的靶向序列。其实例包括生物信息学工具,如Benchling(https://benchling.com)和COSMID(具有错配、插入和缺失的CRISPR脱靶位点)(可在互联网上的https://crispr.bme.gatech.edu上获得)。使用这些,可以总结出与gRNA所靶向的碱基序列的相似性。当所用的gRNA设计软件不具备检索靶基因组的脱靶位点的功能时,例如,可以通过对于候选靶向序列3’侧上的8至12个核苷酸(对所靶向的核苷酸序列具有高辨别能力的种子序列)对靶基因组进行Blast检索来检索脱靶位点。

[0075] 在本发明的一个实施方案中,在人4号染色体(Chr 4)的GRCh38/hg38中存在的区域中,“190,065,500-190,068,500”区域和“190,047,000-190,052,000”区域可能是人DUX4基因的表达调控区。因此,在本发明的一个实施方案中,在人4号染色体(Chr 4)的GRCh38/hg38中存在的“190,065,500-190,068,500”和“190,047,000-190,052,000”区域中,靶向序列可以是18至24个核苷酸的长度,优选20至23个核苷酸的长度,更优选21至23个核苷酸的长度。

[0076] 在本发明的一个实施方案中,在人4号染色体(Chr 4)的GRCh38/hg38中存在的“190,065,000-190,093,000”(D4Z4重复区域)和“190,173,000-190,176,000”(DUX4基因)区域中,靶向序列可以是18至24个核苷酸的长度,优选20至23个核苷酸的长度,更优选21至23个核苷酸的长度。

[0077] 在本发明的一个实施方案中，编码crRNA的碱基序列可以与靶向序列的碱基序列相同。例如，当SEQ ID NO:4(CCCTCCACCGGGCTGACCGGCC)所示的靶向序列作为编码crRNA的碱基序列引入到细胞中时，从所述序列转录的crRNA是CCCUCCACCGGGCUGACCGGCC(SEQ ID NO:93)并与GGCCGGTCAGCCCGGTGGAGGG(SEQ ID NO:94)结合，后者是与SEQ ID NO:4所示的碱基序列互补的序列并存在于人DUX4基因的表达调控区中。在另一个实施方案中，作为其中缺失、替换、插入和/或添加至少1至3个碱基的靶向序列的碱基序列可以用作编码crRNA的碱基序列，只要向导RNA可以将融合蛋白募集到靶区域即可。因此，在本发明的一个实施方案中，作为编码crRNA的碱基序列，可以使用SEQ ID NO:2、3、4、8、15、17、18、20、25、31、32、33、35、39、40、42、44、50、51、52、55、57、58、59、65、67、68、113、116、135、138、142、144、146、156、158、161或171所示的碱基序列，或其中缺失、替换、插入和/或添加1至3个碱基的SEQ ID NO:2、3、4、8、15、17、18、20、25、31、32、33、35、39、40、42、44、50、51、52、55、57、58、59、65、67、68、113、116、135、138、142、144、146、156、158、161或171所示的碱基序列。

[0078] 在本发明的一个优选实施方案中，作为编码crRNA的碱基序列，可以使用SEQ ID NO:2、3、4、20、51、68、138、142、146、156、158或161所示的碱基序列，或其中缺失、替换、插入和/或添加1至3个碱基的SEQ ID NO:2、3、4、20、51、68、138、142、146、156、158或161所示的碱基序列。

[0079] 当dCpf1用作核酸酶缺陷型CRISPR效应蛋白时，编码gRNA的碱基序列可以设计为编码5’末端附接有特定RNA的crRNA的DNA序列。此类附接至crRNA 5’末端的RNA和编码所述RNA的DNA序列可以由本领域普通技术人员根据所用的dCpf1适当地进行选择。例如，当使用dFnCpf1时，其中SEQ ID NO:95AATTTCTACTGTTGTAGAT附接至靶向序列5’侧的碱基序列可以用作编码gRNA的碱基序列(当转录至RNA时，加下划线部分的序列形成碱基对以形成茎-环结构)。添加至5’末端的序列可以是其中缺失、替换、插入和/或添加至少1至6个碱基的通常用于各种Cpf1蛋白的序列，只要转录后gRNA可以将融合蛋白募集到表达调控区即可。

[0080] 当dCas9用作CRISPR效应蛋白时，编码gRNA的碱基序列可以设计为其中编码已知tracrRNA的DNA序列连接至编码crRNA的DNA序列的3’末端的DNA序列。此类tracrRNA和编码所述tracrRNA的DNA序列可以由本领域普通技术人员根据所用的dCas9适当地选择。例如，当使用dSaCas9时，SEQ ID NO:96所示的碱基序列用作编码tracrRNA的DNA序列。编码tracrRNA的DNA序列可以是其中缺失、替换、插入和/或添加至少1至6个碱基的通常用于各种Cas9蛋白的编码tracrRNA的碱基序列，只要转录后gRNA可以将融合蛋白募集到表达调控区即可。

[0081] 包含如此设计的编码gRNA的碱基序列的多核苷酸可以使用已知的DNA合成方法化学合成。

[0082] 在本发明的另一个实施方案中，本发明的多核苷酸可以包含至少两个不同的编码gRNA的碱基序列。例如，多核苷酸可以包含至少两个不同的编码向导RNA的碱基序列，其中所述至少两个不同的碱基序列选自包含SEQ ID NO:2、3、4、8、15、17、18、20、25、31、32、33、35、39、40、42、44、50、51、52、55、57、58、59、65、67、68、113、116、135、138、142、144、146、156、158、161或171所示的序列的的碱基序列，优选地选自包含SEQ ID NO:2、3、4、20、51、68、138、142、146、156、158或161所示的序列的碱基序列。

[0083] (5)启动子序列

[0084] 在本发明的一个实施方案中，启动子序列可以可操作地连接至编码核酸酶缺陷型CRISPR效应蛋白与转录阻遏物的融合蛋白的碱基序列和/或编码gRNA的碱基序列中每一个的上游。可能连接的启动子不受特别限制，只要它在靶细胞中显示出启动子活性即可。可能连接至编码gRNA的碱基序列上游的启动子序列的实例包括但不限于U6启动子、SNR6启动子、SNR52启动子、SCR1启动子、RPR1启动子、U3启动子、H1启动子和tRNA启动子，它们是pol III启动子，等等。在本发明的一个实施方案中，U6启动子可以用作编码向导RNA的碱基序列的启动子序列。在本发明的一个实施方案中，当多核苷酸包含两个或更多个分别编码向导RNA的碱基序列时，单个启动子序列可以可操作地连接至所述两个或更多个碱基序列的上游。在另一个实施方案中，当多核苷酸包含两个或更多个分别编码向导RNA的碱基序列时，启动子序列可以可操作地连接至所述两个或更多个碱基序列中每一个的上游，其中可操作地连接至每一个碱基序列的启动子序列可以相同或不同。

[0085] 由于上述启动子序列可能连接至编码融合蛋白的碱基序列的上游，因此可以使用广泛性启动子或神经元特异性启动子。广泛性启动子的实例包括但不限于EF-1α启动子、EFS启动子、CMV(巨细胞病毒)启动子、hTERT启动子、SRα启动子、SV40启动子、LTR启动子、CAG启动子、RSV(劳氏肉瘤病毒)启动子等等。在本发明的一个实施方案中，EFS启动子、CMV启动子或CAG启动子可以用作广泛性启动子。神经元特异性启动子的实例包括但不限于神经元特异性烯醇酶(NSE)启动子、人神经丝轻链(NEFL)启动子。上述启动子可以具有任何修饰和/或改变，只要它在靶细胞中具有启动子活性即可。

[0086] 在本发明的一个实施方案中，U6用作编码所述向导RNA的碱基序列的启动子，而CMV启动子可以用作编码所述融合蛋白的碱基序列的启动子序列。

[0087] (6)其它碱基序列

[0088] 此外，为了改善通过编码核酸酶缺陷型CRISPR效应蛋白与转录阻遏物的融合蛋白的碱基序列的转录产生的mRNA的翻译效率，除上述那些之外，本发明的多核苷酸还可以包含诸多已知序列，如聚腺苷酸化(polyA)信号、Kozak一致序列等。例如，本发明中的聚腺苷酸化信号可以包括hGH polyA、bGH polyA、2x sNRP-1polyA(参见US7557197B2，其以引用的方式整体并入本文)等等。此外，本发明的多核苷酸可以包含编码接头序列的碱基序列、编码NLS的碱基序列和/或编码标签的碱基序列。此外，本发明的多核苷酸可以包含插入序列。插入序列的优选实例是编码IRES(内部核糖体进入位点)2A肽的序列。2A肽是来源于病毒的约20个氨基酸残基的肽序列，被细胞中存在的蛋白酶(2A肽酶)识别，并在C末端第1个残基的位置被切割。通过2A肽连接为一个单元的多个基因被转录和翻译为一个单元，然后被2A肽酶切割。2A肽酶的实例包括F2A(来源于口蹄疫病毒)、E2A(来源于马鼻炎A病毒)、T2A(来源于明脉扁刺蛾(Thosea asigna)病毒)和P2A(来源于猪捷申病毒-1)。

[0089] (7)本发明的示例实施方案

[0090] 在本发明的一个实施方案中，提供了包含以下的多核苷酸：

[0091] 编码核酸酶缺陷型CRISPR效应蛋白与转录阻遏物的融合蛋白的碱基序列，

[0092] 编码核酸酶缺陷型CRISPR效应蛋白与转录阻遏物的融合蛋白的碱基序列的启动子序列，

[0093] 一个或两个分别编码向导RNA的碱基序列，其中所述一个或两个碱基序列选自包含SEQ ID NO:2、3、4、8、15、17、18、20、25、31、32、33、35、39、40、42、44、50、51、52、55、57、58、

59、65、67、68、113、116、135、138、142、144、146、156、158、161或171所示的序列的碱基序列，优选地选自包含SEQ ID NO:2、3、4、20、51、68、138、142、146、156、158或161所示的序列的碱基序列;或包含其中缺失、替换、插入和/或添加了1至3个碱基的SEQ ID NO:2、3、4、8、15、17、18、20、25、31、32、33、35、39、40、42、44、50、51、52、55、57、58、59、65、67、68、113、116、135、138、142、144、146、156、158、161或171所示的序列的碱基序列,优选地选自包含其中缺失、替换、插入和/或添加了1至3个碱基的SEQ ID NO:2、3、4、20、51、68、138、142、146、156、158或161所示的序列的碱基序列,以及

[0094] 编码gRNA的碱基序列的启动子序列，

[0095] 其中所述核酸酶缺陷型CRISPR效应蛋白是dSaCas9或dSaCas9[-25]，

[0096] 其中所述转录阻遏物选自KRAB、MeCP2、SIN3A、HDT1、MBD2B、NIPP1和HP1A，

[0097] 其中所述编码融合蛋白的碱基序列的启动子序列选自EFS启动子、CMV启动子和CAG启动子，

[0098] 其中所述编码gRNA的碱基序列的启动子序列选自U6启动子、SNR6启动子、SNR52启动子、SCR1启动子、RPR1启动子、U3启动子和H1启动子。

[0099] 在本发明的一个实施方案中,提供了包含以下的多核苷酸：

[0100] 编码核酸酶缺陷型CRISPR效应蛋白与转录阻遏物的融合蛋白的碱基序列，

[0101] 编码核酸酶缺陷型CRISPR效应蛋白与转录阻遏物的融合蛋白的碱基序列的CMV启动子，

[0102] 一个或两个分别编码向导RNA的碱基序列,其中所述一个或两个碱基序列选自包含SEQ ID NO:2、3、4、8、15、17、18、20、25、31、32、33、35、39、40、42、44、50、51、52、55、57、58、59、65、67、68、113、116、135、138、142、144、146、156、158、161或171所示的序列的碱基序列，优选地选自包含SEQ ID NO:2、3、4、20、51、68、138、142、146、156、158或161所示的序列的碱基序列;或包含其中缺失、替换、插入和/或添加了1至3个碱基的SEQ ID NO:2、3、4、8、15、17、18、20、25、31、32、33、35、39、40、42、44、50、51、52、55、57、58、59、65、67、68、113、116、135、138、142、144、146、156、158、161或171所示的序列的碱基序列,优选地选自包含其中缺失、替换、插入和/或添加了1至3个碱基的SEQ ID NO:2、3、4、20、51、68、138、142、146、156、158或161所示的序列的碱基序列,以及

[0103] 编码向导RNA的碱基序列的U6启动子，

[0104] 其中所述核酸酶缺陷型CRISPR效应蛋白是dSaCas9，

[0105] 其中所述转录阻遏物是KRAB。

[0106] 2.载体

[0107] 本发明提供了一种包含本发明的多核苷酸的载体(下文有时称为“本发明的载体”)。本发明的载体可以是质粒载体或病毒载体。

[0108] 当本发明的载体是质粒载体时,所用的质粒载体不受特别限制,并且可以是任何质粒载体,如克隆质粒载体和表达质粒载体。所述质粒载体是通过利用已知的方法将本发明的多核苷酸插入质粒载体中而制备。

[0109] 当本发明的载体是病毒载体时,所用的病毒载体不受特别限制,其实例包括但不限于腺病毒载体、腺相关病毒(AAV)载体、慢病毒载体、逆转录病毒载体、仙台病毒载体等。在本说明书中,“病毒载体”还包括其衍生物。考虑到在基因疗法中使用,优选使用AAV载体,

因为它可以长时间表达转基因,并且它来源于非致病性病毒并具有高安全性。

[0110] 包含本发明的多核苷酸的病毒载体可以通过已知的方法制备。简而言之,制备其中插入了本发明多核苷酸的病毒表达用质粒载体,将所述载体转染到适当的宿主细胞中以允许短暂产生包含本发明多核苷酸的病毒载体,并收集病毒载体。

[0111] 在本发明的一个实施方案中,当使用AAV载体时,AAV载体的血清型不受特别限制,只要可以激活靶标中人DUX4基因的表达即可,并且可以使用AAV1、AAV2、AAV3、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAVrh.10等中的任一个(关于AAV的各种血清型,参见例如WO 2005/033321和EP2341068(A1),所述文献以引用的方式整体并入本文)。AAV变体的实例包括但不限于具有经修饰衣壳的新血清型(例如WO 2012/057363,所述文献以引用的方式整体并入本文)等。例如,在本发明的一个实施方案中,可以使用具有改善肌细胞感染性的经修饰衣壳的新血清型,如AAV_{587}MTP、AAV_{588}MTP、AAV-B1、AAVM41、AAVS1_P1和AAVS10_P1等(参见Yu等人,Gene Ther.2009年8月;16(8):953-62,Choudhury等人,Mol Ther.2016年8月;24(7):1247-57,Yang等人,Proc Natl Acad Sci U S A.2009年3月10日;106(10):3946-51,以及WO2019/207132,所述文献以引用的方式整体并入本文)。

[0112] 当制备AAV载体时,可以使用已知的方法,如(1)使用质粒的方法,(2)使用杆状病毒的方法,(3)使用单纯疱疹病毒的方法,(4)使用腺病毒的方法,或(5)使用酵母的方法(例如Appl Microbiol Biotechnol.2018;102(3):1045-1054等,所述文献以引用的方式整体并入本文)。例如,当通过使用质粒的方法制备AAV载体时,首先制备了包含在野生型AAV基因组序列两端的反向末端重复序列(ITR)和代替编码Rep蛋白和衣壳蛋白的DNA插入的本发明多核苷酸的载体质粒。另一方面,将形成病毒粒子所必需的编码Rep蛋白和衣壳蛋白的DNA插入其它质粒中。此外,将包含负责AAV增殖所必需的腺病毒辅助作用的基因(E1A、E1B、E2A、VA和E4orf6)的质粒制备为腺病毒辅助质粒。这三种质粒共转染至宿主细胞中在所述细胞中产生了重组AAV(即,AAV载体)。作为宿主细胞,优选使用能够提供负责上述辅助作用的基因的部分基因产物(蛋白质)的细胞(例如293细胞等)。当使用此类细胞时,上述腺病毒辅助质粒中不必携带编码可由宿主细胞提供的蛋白质的基因。所产生的AAV载体存在于细胞核中。因此,所期望的AAV载体是通过利用冻融破坏宿主细胞,收集病毒然后利用使用氯化铯的密度梯度超速离心法、柱法等对病毒级分进行分离和纯化来制备。

[0113] AAV载体在安全性、基因转导效率等方面具有很大优势,并用于基因疗法。然而,已知可以包装在AAV载体中的多核苷酸的大小受限制。例如,在本发明的一个实施方案中,包括包含编码dSaCas9与miniVR或microVR的融合蛋白的碱基序列、编码靶向人DUX4基因的表达调控区的gRNA的碱基序列以及作为启动子序列的EFS启动子序列或CK8启动子序列和U6启动子序列以及ITR部分的多核苷酸的碱基长度的整个长度是约4.85kb,并且它们可以包装在单个AAV载体中。

[0114] 3.药物组合物

[0115] 本发明还提供了包含本发明的多核苷酸或本发明的载体的药物组合物(下文有时称为“本发明的药物组合物”)。本发明的药物组合物可用于治疗或预防FSHD。

[0116] 本发明的药物组合物包含本发明的多核苷酸或本发明的载体作为活性成分,并且可以制备为包含此类活性成分(即,本发明的多核苷酸或本发明的载体)和通常药学上可接受的载剂的制剂。

[0117] 本发明的药物组合物经肠胃外施用，并且可以局部或全身施用。本发明的药物组合物可以通过例如但不限于静脉内施用、动脉内施用、皮下施用、腹膜内施用或肌肉内施用来施用。

[0118] 本发明的药物组合物对受试者的剂量不受特别限制，只要它是治疗和/或预防有效量即可。可以根据活性成分、剂型、受试者的年龄和体重、施用日程表、施用方法等适当地优化。

[0119] 在本发明的一个实施方案中，本发明的药物组合物不仅可以施用于受FSHD影响的受试者，而且还可以基于遗传背景分析等预防性地施用于未来可能罹患FSHD的受试者。本说明书中的术语“治疗”包括缓解疾病以及治愈疾病。此外，术语“预防”还可以包括延迟疾病发作以及预防疾病发作。本发明的药物组合物还可以称为“本发明的药剂”等。

[0120] 4.治疗或预防FSHD的方法

[0121] 本发明还提供一种治疗或预防FSHD的方法，其包括向有需要的受试者施用本发明的多核苷酸或本发明的载体(下文有时称为“本发明的方法”)。此外，本发明包括本发明的多核苷酸或本发明的载体，用于治疗或预防FSHD。此外，本发明包括本发明的多核苷酸或本发明的载体用于制造用以治疗或预防FSHD的药物组合物的用途。

[0122] 本发明的方法可以通过将上述本发明的药物组合物施用于受FSHD影响的受试者来实施，并且剂量、施用途径、受试者等与上述那些相同。

[0123] 症状的测量可以在开始使用本发明的方法治疗前和治疗后的任何时间进行，以确定受试者对治疗的反应。

[0124] 5.核糖核蛋白

[0125] 本发明提供了包含以下的核糖核蛋白(下文有时称为“本发明的RNP”)：

[0126] (c)核酸酶缺陷型CRISPR效应蛋白与转录阻遏物的融合蛋白，以及

[0127] (d)靶向人DUX4基因的表达调控区中SEQ ID NO:2、3、4、8、15、17、18、20、25、31、32、33、35、39、40、42、44、50、51、52、55、57、58、59、65、67、68、113、116、135、138、142、144、146、156、158、161或171所示的连续区域的向导RNA，优选地选自包含SEQ ID NO:2、3、4、20、51、68、138、142、146、156、158或161所示的序列的碱基序列。

[0128] 作为本发明的RNP中所包含的核酸酶缺陷型CRISPR效应蛋白、转录阻遏物和向导RNA，可以使用上述“1.多核苷酸”部分详细说明的核酸酶缺陷型CRISPR效应蛋白、转录阻遏物和向导RNA。本发明的RNP所包含的核酸酶缺陷型CRISPR效应蛋白与转录阻遏物的融合蛋白可以通过例如将编码所述融合蛋白的多核苷酸引入到细胞、细菌或其它生物体中以允许其表达或通过使用所述多核苷酸的体外翻译系统来产生。此外，本发明的RNP所包含的向导RNA可以通过例如化学合成或使用编码所述向导RNA的多核苷酸的体外转录系统来产生。将如此制备的融合蛋白和向导RNA混合以制备本发明的RNP。必要时，可以混合其它物质，如金粒子。为了将本发明的RNP直接递送到靶细胞、组织等，可以通过已知的方法将RNP包封在脂质纳米粒子(LNP)中。本发明的RNP可以通过已知的方法引入到靶细胞、组织等中。关于包封在LNP中和引入方法，可以参考例如Lee K.等人，Nat Biomed Eng.2017;1:889-901，WO 2016/153012等，所述文献以引用的方式整体并入本文。

[0129] 在本发明的一个实施方案中，本发明的RNP所包含的向导RNA靶向人4号染色体(Chr 4)的GRCh38/hg38中存在的“190,065,500-190,068,500”和“190,047,000-190,052,

000”区域中连续18至24个核苷酸的长度，优选20至23个核苷酸的长度，更优选21至23个核苷酸的长度。

[0130] 在本发明的一个实施方案中，本发明的RNP所包含的向导RNA靶向人4号染色体(Chr 4)的GRCh38/hg38中存在的“190,065,000-190,093,000”和“190,173,000-190,176,000”区域中连续18至24个核苷酸的长度，优选20至23个核苷酸的长度，更优选21至23个核苷酸的长度。

[0131] 6.其它

[0132] 本发明还提供了一种用于抑制人DUX4基因表达的组合物或药盒，其包含以下：

[0133] (e)核酸酶缺陷型CRISPR效应蛋白与转录阻遏物的融合蛋白，或编码所述融合蛋白的多核苷酸，以及

[0134] (f)向导RNA或编码所述向导RNA的多核苷酸，所述向导RNA靶向人DUX4基因的表达调控区中SEQ ID NO:2、3、4、8、15、17、18、20、25、31、32、33、35、39、40、42、44、50、51、52、55、57、58、59、65、67、68、113、116、135、138、142、144、146、156、158、161或171所示的连续区域，优选地选自包含SEQ ID NO:2、3、4、20、51、68、113、116、135、138、142、144、146、156、158、161或171所示的序列的碱基序列。

[0135] 本发明还提供了一种治疗或预防FSHD的方法，其包括施用以下(e)和(f)：

[0136] (e)核酸酶缺陷型CRISPR效应蛋白与转录阻遏物的融合蛋白，或编码所述融合蛋白的多核苷酸，以及

[0137] (f)向导RNA或编码所述向导RNA的多核苷酸，所述向导RNA靶向人DUX4的表达调控区中SEQ ID NO:2、3、4、8、15、17、18、20、25、31、32、33、35、39、40、42、44、50、51、52、55、57、58、59、65、67、68、113、116、135、138、142、144、146、156、158、161或171所示的连续区域，优选地选自包含SEQ ID NO:2、3、4、20、51、68、113、116、135、138、142、144、146、156、158、161或171所示的序列的碱基序列。

[0138] 本发明还提供了以下(e)和(f)用于制造用以治疗或预防FSHD的药物组合物的用途：

[0139] (e)核酸酶缺陷型CRISPR效应蛋白与转录阻遏物的融合蛋白，或编码所述融合蛋白的多核苷酸，以及

[0140] (f)向导RNA或编码所述向导RNA的多核苷酸，所述向导RNA靶向人DUX4基因的表达调控区中SEQ ID NO:2、3、4、8、15、17、18、20、25、31、32、33、35、39、40、42、44、50、51、52、55、57、58、59、65、67、68、113、116、135、138、142、144、146、156、158、161或171所示的连续区域，优选地选自包含SEQ ID NO:2、3、4、20、51、68、113、116、135、138、142、144、146、156、158、161或171所示的序列的碱基序列。

[0141] 作为本发明的核酸酶缺陷型CRISPR效应蛋白、转录阻遏物、向导RNA以及编码它们的多核苷酸和其中携带它们的载体，可以使用上述“1.多核苷酸”、“2.载体”和“5.核糖核蛋白”部分详细说明的那些。上述(e)和(f)的剂量、施用途径、受试者、配方等与“3.药物组合物”部分所述的那些相同。

[0142] 本发明的其它特征将在以下示例性实施方案的描述过程中变得显而易见，所述示例性实施方案是为了说明本发明而提供，而不意在对其具有限制性。

[0143] 实施例

[0144] 实施例描述了使用dCas9与转录阻遏物的融合蛋白抑制所定义的人DUX4基因的表达调控区中的基因表达，由此选择性地抑制人DUX4基因的表达。实施例还描述了赋予人DUX4基因选择性抑制而没有最低限度地影响其它基因表达的特定基因组区域的定义。本发明的抑制人DUX4基因表达的方法代表了如本文所描述和说明的FSHD的新颖治疗或预防策略。

[0145] (1)实验方法

[0146] DUX4靶向序列的选择-1

[0147] 基于人骨骼肌细胞基因组的H3K4me3、H3K27Ac模式，扫描了人DUX4基因的推定启动子区域周围约8kb的序列，以得到可以用与本文定义为靶向序列的gRNA复合的无催化活性SaCas9(D10A和N580A突变体;dSaCas9)靶向的序列。所靶向的基因组区域相对于DUX4基因的位置描绘于图1中，下面指出了其坐标：

[0148] 1.Chr4:GRCh38/hg38;190,065,500-190,068,500->～3kb(启动子A)

[0149] 2.Chr4:GRCh38/hg38;190,047,000-190,052,000->～5kb(启动子B)

[0150] 靶向序列规定为与具有序列NNGRRT的原间隔子相邻基序(PAM)相邻的21个核苷酸片段(5'-21nt靶向序列-NNGRRT-3')(表1-1至表1-3)。

[0151] [表1-1]

[0152]

SEQ ID No.		位置	链	序列	PAM
77	对照-1	N/A	N/A	ACGGAGGCTAAGCGTCGCAA	N/A
78	对照-2	N/A	N/A	CGCTTCCGCGGCCCGTTCAA	N/A
79	对照-3	N/A	N/A	GTAGGCGCGCCGCTCTCTAC	N/A
1	sgDUX4-1	190065622	-1	GGGAGGTGGAGCTGCCCCGGCT	TGGGGT
2	sgDUX4-2	190065752	-1	CTCATCCAGCAGCAGGCCGCAG	GGGAGT
3	sgDUX4-3	190065880	-1	AGCCCGGTATTCTTCCTCGCTG	AGGGGT
4	sgDUX4-4	190066039	1	CCCTCCACCGGGCTGACCGGCC	TGGGAT
5	sgDUX4-5	190066171	1	GTGGGCCGCCTACTGCGCACGC	GCGGGT
6	sgDUX4-6	190066333	-1	GGGGCCCGGTGTTTCGCGGGAC	GGGGGT
7	sgDUX4-7	190066493	-1	GCGTCCCGGTGTGCGCCGGGCC	TGGGGT
8	sgDUX4-8	190066576	1	AACGGGAGACCTAGAGGGGCGG	AAGGAT
9	sgDUX4-9	190066735	1	GGAAAAGCGGTCCTCGGCCTCC	GGGAGT
10	sgDUX4-10	190066906	-1	CGGGTGAAAACCCGACGGCAAC	CCGAGT
11	sgDUX4-11	190066994	-1	CCTGCGTGTGGCTCCTCCGTGG	CCGGGT
12	sgDUX4-12	190067146	-1	TTGCACCCTTCCCTGCATGTTT	CCGGGT
13	sgDUX4-13	190067284	1	CGCCGGGGAGGCATCTCCTCTC	TGGGGT
14	sgDUX4-14	190067420	-1	GAACTGAACCTCCGTGGGAGTC	TTGAGT
15	sgDUX4-15	190067516	1	AGAGAGCGGCTTCCCGTTCCCG	CGGGAT
16	sgDUX4-16	190067532	-1	GGCCGGCTCTCCGGACCTCTCC	AGGGAT
17	sgDUX4-17	190067677	-1	GTCGAGGCCTGGGGCCGGCCGG	CGGGGT
18	sgDUX4-18	190067696	1	GCCGGCCCCAGGCCTCGACGCC	CTGGGT
19	sgDUX4-19	190067708	1	CCTCGACGCCCTGGGTCCCTTC	CGGGGT

[0153] [表1—2]

[0154]

20	sgDUX4—20	190067753	1	GGGGGCTCACCGCCATTCATGA	AGGGGT
21	sgDUX4—21	190067757	-1	GGCAGGCAGGCTCCACCCCTTC	ATGAAT
22	sgDUX4-22	190067922	-1	GGCCATCGCGGTGAGCCCCGGC	CGGAAT
23	sgDUX4—23	190067958	—1	TTCCGCGGGGAGGGTGCTGTCC	GAGGGT
24	sgDUX4-24	190067971	-1	CTCGTCCCCGGGCTTCCGCGGG	GAGGGT
25	sgDUX4-25	190068012	-1	CTCGCTTTGGCTCGGGGTCCAA	ACGAGT
26	sgDUX4-26	190068114	1	AGGCCATCGGCATTCCGGAGCC	CAGGGT
27	sgDUX4-27	190068236	1	CGGCGAAAGCGGACCGCCGTCA	CCGGAT
28	sgDUX4-28	190068339	1	GAGAGACGGGCCTCCCGGAGTC	CAGGAT
29	sgDUX4-29	190068439	1	CCTGTGCAGCGCGGCCCCCGGC	GGGGGT
30	sgDUX4-30	190068482	1	TGGGTCGCCTTCGCCCACACCG	GCGAGT
31	sgDUX4-31	190048092	-1	TTTCAGTTTCCCTTTACTTGCA	GTGAAT
32	sgDUX4-32	190048802	-1	TCCAAAATGCAACCTGAAGCTA	CTGAGT

33	sgDUX4-33	190048960	-1	CACTTTCAATGGCGCCTTTTTC	ACGAAT
34	sgDUX4-34	190049067	1	GGAAGATTTCTCTCAAAAACGG	TTGAGT
35	sgDUX4-35	190049179	-1	GAGGAGTACGTCTTCTGCAGCC	CAGGGT
36	sgDUX4-36	190049317	-1	CAGGTCTGGGCTGAGCATTGAG	GAGGGT
37	sgDUX4-37	190049456	1	CCTCCCTCCAGGGCTATGGACC	CCGGAT
38	sgDUX4—38	190049553	1	AGCTCCCAGCATATTAGGTGGG	GCGGGT
39	sgDUX4-39	190049650	1	GTAGCCCTTTGACCAAAGGGTT	GGGAGT
40	sgDUX4-40	190049789	1	GCAGCCCTGTGACCAAAGGGCT	GGGAGT
41	sgDUX4-41	190049881	-1	CCCAGCATGCACCAGGCTCAGG	GAGAGT
42	sgDUX4-42	190050023	-1	CCCAGCATGCACCAGGCTCAGG	GAGAGT
43	sgDUX4-43	190050155	1	GAAGCTCAGTTGTGTTCACTTT	TTGGAT
44	sgDUX4-44	190050250	-1	AACTATAGCTATCTGCAATATC	ATGAAT
45	sgDUx4-45	190050823	-1	ATACGATCCAGCAATTTCACAT	CTGGGT
46	sgDUX4-46	190047610	-1	GTCAAGCAGATGCCAAGTCGT	AGGGAT
47	sgDUX4-47	190047312	-1	TCAGTAAATGTTCTGTCATTG	CAGGAT
48	sqDUX4-48	190047373	1	ATGAGCCAAGTGAAAACACCA	AGGGGT

[0155] [表1-3]

[0156]

49	s qDUX4-49	190047553	-1	GGCATCACAAATGGCAAACCA	GTGAAT
50	sgDUX4-50	190047704	1	AATGCATAAAGcCTGAAGGGA	CAGGGT
51	sgDUX4-51	190047832	1	GAACGGACAGCAAACGACAAG	CTGAGT
52	sgDUX4-52	190047916	-1	ATCGATAAATACGCTAACCAC	ATGGGT
53	sgDUX4-53	190047980	-1	GTCACCCGCATTCTTTAGAGA	GAGAAT
54	sqDUX4-54	190048279	1	GAATATTTTGAAAATTAAGTG	CAGAAT
55	sgDUX4-55	190048373	1	TGCTGGAGAAACCCTCTTTAT	GGGAGT
56	sgDUX4-56	190048520	1	TAGCATCTTAAATCTCCACCC	AGGAAT
57	s CDUX4-57	190048596	1	AATCAAAATGCACATGCTCTG	TAGAAT
58	sgDUX4-58	190048649	-1	ACTTTTAAGACGTGGAGCACT	TGGGGT
59	sqDUX4-59	190048832	1	TTTTGGAGCTTGTCCACTTAA	GTGGGT
60	s CDUX4-60	190048874	1	ACAAGACTCACGCCTCAGAAA	TGGGGT
61	sgDUX4-61	190049206	1	CGTACTCCTCTGGAAAGAACC	CTGGGT
62	sgDUX4-62	190049296	1	GGAAAAGGTAAGGTATCTTTG	TAGGGT
63	sqDUX4—63	190049365	-1	TGCCGACTCTTTGCAAGGAGA	GTGAGT
64	sgDUX4—64	190049472	—1	AGCTATAATCCCAGCATGTAC	CGGGAT
65	sgDUX4-65	190049574	-1	GCGCAGCCCTGGAAGAAGGGG	CAGAGT
66	sgDUX4—66	190049662	1	CAAAGGGTTGGGAGTGTTTAT	GAGAAT
67	sgDUX4-67	190050399	—1	GCAAGTTGAAAGAAGGTCCCT	ATGAGT
68	sgDUX4—68	190050465	—1	GTAGTTCTTAGGAGCTAGAGG	GGGAGT
69	sqDUX4—69	190050521	—1	TCATTCCACTTATGAGATACT	TAGAGT
70	sqDUX4—70	190050615	1	CATGCATGACGTACCATGTGT	CAGAAT

71	sgDUX4—71	190050709	1	TGATGGTAATTCAAACAACAC	TTGAGT
72	sgDUX4—72	190050923	1	ACATTTCCACCAACAGTGTAC	AAGGAT
73	sgDUX4—73	190051026	1	GGTATTAAGTGATATGTCATT	TGGGGT
74	sqDUX4—74	190051168	1	AAGTAGTTTGTTTTATTGTTG	CTGAAT
75	sgDUX4—75	190051280	—1	TAGGGGAAACACGTCATAACA	CTGGAT
76	sgDUX4—76	190051357	—1	CTGTAAGAAAAAATTAACTAA	CTGGAT

[0157] DUX4靶向序列的选择-2

[0158] 在之前的选择(DUX4靶向序列的选择-1)中,我们基于人骨骼肌细胞基因组的H3K4me3、H3K27Ac模式,在人DUX4基因的推定启动子区域周围大约8kb的序列中测试了靶向序列。现在我们将研究重点扩展到包括整个D4Z4重复区和DUX4编码基因的更大区域。扫描所述区域,以得到可以通过与本文定义为靶向序列的gRNA复合的无催化活性SaCas9(D10A和N580A突变体;dSaCas9)靶向的序列。所靶向的基因组区域相对于DUX4基因的位置描绘于图5中,下面指出了其坐标:

[0159] 1.Chr4:GRCh38/hg38;190,065,000-190,093,000→～28kb(D4Z4重复区)

[0160] 2.Chr4:GRCh38/hg38;chr4:190,173,000-190,176,000→～3kb(DUX4基因)

[0161] 靶向序列规定为与具有序列NNGRRT的原间隔子相邻基序(PAM)相邻的21个核苷酸片段(5'-21nt靶向序列-NNGRRT-3')(表2-1至表2-4)。

[0162] [表2-1]

[0163]

SEQ ID No.		位置	链	序列	PAM
77	对照-1	N/A	N/A	ACGGAGGCTAAGCGTCGCAA	N/A
78	对照-2	N/A	N/A	CGCTTCCGCGGCCCGTTCAA	N/A
79	对照-3	N/A	N/A	GTAGGCGCGCCGCTCTCTAC	N/A
2	sgDUX4-2	190065752	-1	CTCATCCAGCAGCAGGCCGCAG	GGGAGT
20	sgDUX4-20	190067753	1	GGGGGCTCACCGCCATTCATGA	AGGGGT
51	sgDUX4-51	190047832	1	GAACGGACAGCAAACGACAAG	CTGAGT
98	sgDUX4-82	190065019	1	CTAAGGCTTTTTCTCTCCCTCC	CAGAAT
99	sgDUX4-83	190065047	-1	TGGGAGGATTTTGCCTGTGAGT	TCGAAT
100	sgDUX4-84	190065053	-1	AGATTCTGGGAGGATTTTGCCT	GTGAGT
101	sgDUX4-85	190065061	1	GAACTCACAGGCAAAATCCTCC	CAGAAT
102	sgDUX4-86	190065066	-1	TTATGTTCTCACAAGATTCTGG	GAGGAT
103	sgDUX4-87	190065110	1	TGACTAGTTTGGCATTGCTTTT	GGGGAT
104	sgDUX4-88	190065156	-1	GATTTATAAATAATGGCATGAC	AAGGGT
105	sgDUX4-89	190065218	1	GTGAGGGGAGATGGGGAGACAT	TGGGAT
106	sgDUX4-90	190065317	1	CCAGCCAGGCCGCGCCGGCAGA	GGGGAT
107	sgDUX4-91	190065345	1	CTCCCAACCTGCCCCGGCGCGC	GGGGAT
108	sgDUX4-92	190065408	-1	TGCGGAGGCCACCGAGGAGCCT	GAGGGT
109	sgDUX4-93	190065435	-1	TCCCGGTCCTCCCGGCTTTTGC	CCGGGT
110	sgDUX4-94	190065462	-1	GGGCCCGGCAGGCCGTCGCGCT	GCGGGT
111	sgDUX4-95	190065526	1	CTCAAGCGGGGCCGCAGGGCCA	AGGGGT
112	sgDUX4-96	190065553	-1	CCACCACGGACTCCCCTGGGAC	GTGGGT
113	sgDUX4-97	190065554	1	GCTTGCGCCACCCACGTCCCAG	GGGAGT
114	sgDUX4-98	190065578	1	GAGTCCGTGGTGGGGCTGGGGC	CGGGGT
115	sgDUX4-99	190065918	1	CTGCTGGAGGAGCTTTAGGACG	CGGGGT
116	sgDUX4-100	190065929	1	GCTTTAGGACGCGGGGTTGGGA	CGGGGT
117	sgDUX4-101	190065934	1	AGGACGCGGGGTTGGGACGGGG	TCGGGT
118	sgDUX4-102	190066055	-1	CACCGGCCTGGACCTAGAAGGC	AGGAAT

[0164] [表2-2]

[0165]

119	sgDUX4—103	190066087	—1	AGAATGGCAGTTCTCCGCGGTG	TGGAGT
120	sgDUX4—104	190066110	—1	GGGATCCCCGGGATGCCCAGGA	AAGAAT
121	sgDUX4-105	190066117	1	GCCATTCTTTCCTGGGCATCCC	GGGGAT
122	sgDUX4-106	190066124	-1	CCTGGGCCGGCTCTGGGATCCC	CGGGAT
123	sgDUX4—107	190066133	—1	CTGCTGGTACCTGGGCCGGCTC	TGGGAT|
124	sgDUX4—108	190066254	-1	TGGGGATGGGGCGGTCAGGCGG	CGGGGT|
125	sgDUX4-109	190066275	-1	CCGGGGGTGGGGGGTGGGGGGT	GGGGAT

126	sgDUX4-110	190066281	-1	CGTTTTCCGGGGGTGGGGGGTG	GGGGGT
127	sgDUX4-111	190066288	-1	GACGACGCGTTTTCCGGGGGTG	GGGGGT
128	sgDUX4-112	190066295	-1	CCCAGGGGACGACGCGTTTTCC	GGGGGT
129	sgDUX4-113	190066311	1	GGAAAACGCGTCGTCCCCTGGG	CTGGGT
130	sgDUX4-114	190066759	-1	GCTGACCGTTTTCCCGGAGGGC	GGGGGT
131	sgDUX4-115	190066843	-1	CTGGGCCCCGGAACCGGGGCGA	ATGGGT
132	sgDUX4-116	190066847	-1	CTCCCTGGGCCCCGGAACCGGG	GCGAAT
133	sgDUX4-117	190066858	1	TTCGCCCCGGTTCCGGGGCCCA	GGGAGT|
134	sgDUX4-118	190066896	1	CTCCGGGACAAAAGACCGGGAC	TCGGGT
135	sgDUX4-119	190066907	1	AAGACCGGGACTCGGGTTGCCG	TCGGGT
136	sgDUX4-120	190066929	-1	GGATGTGCGGTCTGTGAACCGC	GCGGGT
137	sgDUX4-121	190066953	—1	GCCGCGTTGCAGGGCTCAGCCT	GGGGAT
138	sgDUX4-122	190067152	1	GGGCACCCGGAAACATGCAGGG	AAGGGT
139	sgDUX4-123	190067229	-1	ATTCCCGCGTGCGGCAACGTGG	GGGAGT
140	sgDUX4-124	190067239	1	ACTCCCCCACGTTGCCGCACGC	GGGAAT
141	sgDUX4-125	190067255	-1	TCCCCGGCGTGATGGCCTGACG	ATGGAT
142	sgDUX4-126	190067427	-1	GAGTGTGGAACTGAACCTCCGT	GGGAGT|
143	sgDUX4-127	190067451	-1	AAACCAGCCTGGGAGGGTGGAG	GGGAGT
144	sgDUX4-128	190067461	-1	CAGCAGGGAGAAACCAGCCTGG	GAGGGT
145	sgDUX4-129	190068022	-1	CTCGCAGGGCCTCGCTTTGGCT	CGGGGT
146	sgDUX4-130	190068065	-1	TCTCTGGTGGCGATGCCCGGGT	ACGGGT
147	sgDUX4-131	190068071	-1	AGCCGTTCTCTGGTGGCGATGC	CCGGGT
148	sgDUX4—132	190068115	—1	ACCAAATCTGGACCCTGGGCTC	CGGAAT

[0166] [表2—3]

[0167]

149	sgDUX4-133	190068133	1	GCCCAGGGTCCAGATTTGGTTT	CAGAAT
150	sgDUX4-134	190068170	1	CGCCAGCTGAGGCAGCACCGGC	GGGAAT
151	sgDUX4-135	190068252	-1	GGCTCGGAGGAGCAGGGCGGTC	TGGGAT|
152	sgDUX4-136	190068274	1	CCTGCTCCTCCGAGCCTTTGAG	AAGGAT
153	sgDUX4-137	190068332	1	CTGGCCAGAGAGACGGGCCTCC	CGGAGT|
154	sgDUX4-138	190068355	—1	CCCTTCGATTCTGAAACCAGAT	CTGAAT|
155	sgDUX4—139	190068358	1	GTCCAGGATTCAGATCTGGTTT	CAGAAT
156	sgDUX4-140	190068385	1	TCGAAGGGCCAGGCACCCGGGA	CAGGGT|
157	sgDUX4-141	190068389	-1	GCGGGCGCCCTGCCACCCTGTC	CCGGGT
158	sgDUX4—142	190068458	-1	GCGAAGGCGACCCACGAGGGAG	CAGGGT
159	sgDUX4-143	190068459	1	GCGGGGGTCACCCTGCTCCCTC	GTGGGT
160	sgDUX4—144	190068519	—1	AGCCCCAGGCGCGCAGGGCACG	TGGGGT
161	sgDUX4—145	190069348	—1	ACCGGGCCTAGACCTAGAAGGC	AGGAAT
162	sgDUX4—146	190069575	—1	GCGTTTTCCGGGGGTGGGGGGT	GGGGGT
163	sgDUX4—147	190069784	—1	CGTCCCCGGTGTGCGCCGGGCC	TGGGGT

164	sgDUX4—148	190070198	—1	CGGGTGAAGACCCGACGGCAAC	CCGAGT
165	sgDUX4—149	190070221	—1	GGATGTGGGGTCTGTGAACCGC	GCGGGT|
166	sgDUX4—150	190070530	1	GACTCCCCACGTTGCCGCACGC	GGGAAT
167	sgDUX4—151	190070946	-1	GGTGGTGGTGGTGGTGGGGGGG	GGGGGT
168	sgDUX4—152	190071909	1	CCAGCCAGGCCGCGCCGGCAGA	GGGGGT|
169	sgDUX4—153	190072645	—1	ACCGGGCCTGGACCTAGAAGGC	AGGAAT
170	sgDUX4-154	190072845	-1	TGGGGAGGGGGCGGTCAGGCGG	CGGGGT
171	sgDUX4-155	190073814	-1	GATTCCCGCGTGCGGCAACGTG	GGGAGT
172	sgDUX4-156	190173479	-1	TGGTGGTGGTGGTGGTGGGGGG	GGGGGT
173	sgDUX4—157	190175220	—1	AGAAAGGCAGTTCTCCGCGGAG	TGGAGT
174	sgDUX4—158	190175225	-1	AGGAAAGAAAGGCAGTTCTCCG	CGGAGT
175	sgDUX4-159	190175250	1	GCCTTTCTTTCCTGGGCATCCC	GGGGAT
176	sgDUX4-160	190175673	-1	GCGAGCTCCCTTGCACGTCAGC	CGGGGT
177	sgDUX4-161	190175727	1	TTGTTCTTCCGTGAAATTCTGG	CTGAAT
178	sgDUX4-162	190175732	-1	AAGGTGGGGGGAGACATTCAGC	CAGAAT

[0168] [表2-4]

[0169]

179	sgDUX4-163	190175768	1	TTCCGACGCTGTCTAGGCAAAC	CTGGAT
180	sgDUX4-164	190175774	1	CGCTGTCTAGGCAAACCTGGAT	TAGAGT|
181	sgDUX4-165	190175788	1	ACCTGGATTAGAGTTACATCTC	CTGGAT
182	sgDUX4-166	190175830	1	TATATTAAAATGCCCCCTCCCT	GTGGAT

[0170] 慢病毒转移质粒(pED316)的构建

[0171] pLentiCRISPR v2购自Genscript(https://www.genscript.com),并进行了以下修饰:将SpCas9 gRNA支架序列替换为SaCas9 gRNA支架序列;将SpCas9-FLAG替换为与克鲁佩尔相关框(KRAB)结构域融合的dSaCas9(D10A和N580A突变体)。KRAB转录抑制结构域在定位于启动子时可以通过募集抑制元件来抑制基因表达。KRAB拴系到dSaCas9的末端,下文称为dSaCas9-KRAB,并根据靶向序列的指导靶向人DUX4基因调控区(表1和表2)。所产生的骨架质粒命名为pED316。

[0172] gRNA克隆

[0173] 将三种阴性对照非靶向序列、三种阳性对照靶向序列、76种靶向序列(表1)和85种靶向序列(表2)克隆到pED316中。正向寡核苷酸和反向寡核苷酸通过Integrated DNA Technologies按以下形式合成:正向:5'CACC(G)-21碱基对靶向序列-3',反向:5'AAAC-19—21碱基对反向互补靶向序列-(C)-3',其中如果靶标不是以G开始,则添加括号中的碱基。使寡核苷酸以100μM重悬在Tris—EDTA缓冲液(pH 8.0)中。将1μl各互补寡核苷酸合并在NE Buffer 3.1(New England Biolabs)中的10μl反应物中。将反应物加热至95℃,并在热循环仪中冷却至25℃,由此退火有与克隆到pED316相容的粘性末端悬垂物的寡聚物。将退火的寡核苷酸与经BsmBI消化和凝胶纯化的慢病毒转移质粒pED316合并,并且用T4 DNA连接酶(NEB目录号:M0202S)根据制造商的方案连接。根据制造商的方案将2μl连接反应物转化至10μl NEB稳定感受态细胞(NEB目录号:C3040I)中。所得构建体通过U6启动子驱动包含由与tracrRNA(gttttagtactctggaaacagaatctactaaaacaaggcaaaatgccgtgtttatctcgtc

aacttgttggcgagatttttt;SEQ ID NO:97)融合的个别靶向序列编码的crRNA的sgRNA的表达。

[0174] 慢病毒产生

[0175] 将HEK293TA细胞以0.75×10^6个细胞/孔(对于表1所列出的靶向序列)或1×10^6个细胞/孔(对于表2所列出的靶向序列)播种在6孔细胞培养盘(VWR目录号:10062-892)中的2ml生长培养基(补充有10% FBS和2mM新鲜L-谷氨酰胺、1mM丙酮酸钠和非必需氨基酸的DMEM培养基)中,并在37℃/5% CO_2下温育24小时。次日,用1.5μg包装质粒混合物[1μg包装质粒(参见pCMV delta R8.2;addgene编号12263)和0.5μg包膜表达质粒(参见pCMV-VSV-G;addgene编号8454)]和1μg含有编码dSaCas9-KRAB的序列和所指出的sgRNA的转移质粒pED316根据制造商的方案建立TransIT-VirusGEN转染反应。在转染后48小时(对于表1所列出的靶向序列)或72小时(对于表2所列出的靶向序列),通过使培养基上清液通过0.45μm PES过滤器(VWR目录号:10218-488)来收集慢病毒。将经过纯化和等分的慢病毒储存在-80℃冰箱中直到准备使用。

[0176] FSHD患者来源的成淋巴细胞细胞系(LCL)的转导

[0177] 当使用表1所列出的靶向序列时:

[0178] 两株FSHD患者来源的B成淋巴细胞细胞系(LCL)GM16343和GM16414获自Coriell研究所。在补充有15%胎牛血清的RPMI-1640培养基中培养所述细胞。为了转导,将100,000个细胞与8μg/ml聚凝胺(Sigma目录号:TR-1003-G)混合,并向每一个孔(96孔板)加入对应于各sgRNA(表1)的200μl慢病毒上清液(参见上文)。然后将细胞和病毒混合物在1200×g、37℃下旋转1小时,然后以0.25×10^6至0.5×10^6个细胞/ml重悬在新鲜培养基中。转导后72小时,给细胞饲喂选择培养基[补充有0.5μg/ml嘌呤霉素的生长培养基(Sigma Aldrich目录号:P8833-100MG)]。每2-3天给予细胞新鲜选择培养基。细胞处于选择培养基中7-15天后,按照制造商的指导,收集细胞并用RNeasy 96试剂盒(Qiagen目录号:74182)提取RNA。

[0179] 当使用表2所列出的靶向序列时:

[0180] FSHD患者来源的B成淋巴细胞细胞系(LCL)GM16343获自Coriell研究所。在补充有15%胎牛血清的RPMI-1640培养基中培养所述细胞。为了转导,将500,000个细胞与6.66μg/ml聚凝胺(Sigma目录号:TR-1003-G)混合,并向每一个孔(24孔板)加入对应于各sgRNA(表2)的500μl慢病毒上清液(参见上文)。然后将细胞和病毒混合物在1200×g、37℃下旋转1小时,然后以0.5×10^6个细胞/ml重悬在新鲜培养基中。转导后72小时,给细胞饲喂选择培养基[补充有0.5μg/ml嘌呤霉素的生长培养基(Sigma Aldrich目录号:P8833-100MG)]。每2-3天给予细胞新鲜选择培养基。细胞处于选择培养基中14-17天后,按照制造商的指导,收集细胞并用RNeasy 96试剂盒(Qiagen目录号:74182)提取RNA。

[0181] 基因表达分析

[0182] 为了进行基因表达分析,根据大容量cDNA反转录试剂盒(Applied Biosystems;ThermoFisher目录号:4368813)方案以10μl体积由约0.5-0.8μg总RNA产生cDNA。将cDNA稀释10倍,并使用Taqman Fast Advanced Master Mix根据制造商的方案进行分析。Taqman探针(DUX4:定制设计;MBD3L2:Assay Id Hs00544743_m1;TRIM43:Assay Id Hs00299174_m1;ZSCAN4:Assay Id Hs00537549_m1;HPRT:Assay Id Hs99999909_m1 VIC_PL)获自Life Technologies。通过QuantStudio 5实时PCR系统,在Taqman Fast Advanced Master Mix方

案的指导下处理和分析基于Taqman探针的实时PCR反应。

[0183] 数据分析

[0184] 对于各个样品和对照物，通过从HPRT探针减去靶基因(DUX4、MBD3L2、TRIM43和ZSCAN4)探针的3个技术重复实验的平均Ct值来计算ΔCt值(平均Ct DUX4-平均Ct HPRT)。然后计算3个对照sgRNA的平均ΔCt。然后从平均对照ΔCt减去各个样品和对照物的ΔCt，得到各个样品和对照物的ΔΔCt。然后使用式2^-(ΔΔCt)来确定各个样品和对照物的归一化表达值。在这种情况下，各个对照物的个别表达值相对于所有三个对照样品的平均表达值进行归一化。表示对照样品的图示条柱是三个对照物在三个独立实验中的平均值。

[0185] (2)结果

[0186] 通过dSaCas9-KRAB:sgRNA抑制DUX4基因表达

[0187] 在初步sgRNA筛检实验中，产生了将dSaCas9-KRAB的表达盒和各个靶向序列的sgRNA递送到FSHD患者来源的LCL的慢病毒。选择耐受嘌呤霉素7天(对于表1所列出的靶向序列)或14天(对于表2所列出的靶向序列)的转导细胞，并使用Taqman测定法对DUX4表达进行定量。每一个样品的表达值相对于经对照sgRNA转导的细胞中的DUX4表达平均值进行归一化。

[0188] 当使用表1所列出的靶向序列时：

[0189] 如图2所示，在76个被测序列中，有27个靶向序列在两种LCL中的任一种中示出了DUX4 mRNA表达下调至少50％(图2和表3)。

[0190] [表3]

[0191] 与对照sgRNA(设为1.0)相比，用来自初步sgRNA筛检实验的所选择的27种sgRNA处理的LCL DUX4表达水平

[0192]

	GM16343	GM16414
sgDUX4-2	0.47	0.38
sgDUX4-3	0.49	0.78
sgDUX4-4	0.34	0.66
sgDUX4-8	0.48	0.78
sgDUX4-15	0.41	0.75
sgDUX4-17	0.37	0.34
sgDUX4-18	0.25	0.68
sgDUX4-20	0.53	0.42
sgDUX4-25	0.48	0.53
sgDUX4-31	0.43	1.06
sgDUX4-32	0.45	0.70
sgDUX4-33	0.50	0.68
sgDUX4-35	0.42	0.72
sgDUX4-39	0.50	0.45

	GM16343	GM16414
sgDUX4-40	0.67	0.48
sgDUX4-42	0.24	0.67
sgDUX4-44	0.48	0.98
sgDUX4-50	0.40	0.65
sgDUX4-51	0.48	0.46
sgDUX4-52	0.47	0.84
sgDUX4-55	0.44	0.89
sgDUX4-57	0.26	0.75
sgDUX4-58	0.49	0.55
sgDUX4-59	0.44	0.75
sgDUX4-65	0.41	0.81
sgDUX4-67	0.37	0.62
sgDUX4-68	0.47	0.58

[0193] 接下来，我们对由初步筛检鉴定的这27种最有效候选sgRNA进行了验证筛检，这次选择了耐受嘌呤霉素15天的转导细胞来探索更长时间处理产生更佳抑制的可能性。如图3

所示,7个sgRNA靶向序列在两种LCL中都示出了DUX4 mRNA表达下调至少50%,其中sgRNA-#2示出了约99%抑制。这一结果还表明,对于某些sgRNA,通过更长时间的处理,可以大大提高抑制效力。

[0194] DUX4的表达导致许多下游靶标的异常上调,包括生殖系中和早期发育时表达的基因。TRIM43、ZSCAN4和MBD3L2是DUX4的下游靶标,发现它们在本研究所用的FSHD患者来源的LCL培养物中也上调。为了确定dCas9-KRAB介导的DUX4抑制是否也会抑制这些DUX4靶基因,我们测量了经验证实验所鉴定的最有效sgRNA处理的样品中的TRIM43、ZSCAN4和MBD3L2的水平。正如所料,所有这些dCas9-KRAB:sgRNA都显著降低了所有三个DUX4靶标的表达,达到内源水平的约50-99%(图4),并且抑制效力与DUX4抑制效力密切相关(表4)。

[0195] [表4]

[0196] 用6种最佳sgRNA(sgDUX4-2、sgDUX4-3、sgDUX4-4、sgDUX4-20、sgDUX4-51和sgDUX4-68)处理的FSHD患者来源的LCL中DUX4、TRIM43、MBD3L2和ZSCAN4 mRNA水平的皮尔森相关性分析。

[0197]

	DUX4	TRIM43	MBD3L2	ZSCAN4
DUX4	1.000	0.984	0.961	0.965
TRIM43		1.000	0.964	0.965
MBD3L2			1.000	0.954
ZSCAN4				1.000

[0198] (表中所示的数字是DUX4、TRIM43、MBD3L2和ZSCAN4 mRNA水平中每一个的皮尔森相关性系数)

[0199] 当使用表2所列出的靶向序列时:

[0200] 除了85个新靶向序列之外,还测试了另外3个先前被证明抑制DUX4表达的靶向序列(sgDUX4-2、sgDUX4-20、sgDUX4-51)以供比较。

[0201] 根据3个筛检实验,在85个被测新靶向序列中,11个靶向序列示出了DUX4 mRNA表达平均下调至少50%(图6)。在这11个靶向序列中,有6个靶向序列在所有三个单独筛检实验中导致DUX4表达与对照非靶向序列相比<60%[第1组]。11个靶向序列中另外5个在3个单独筛检实验中的2个中导致DUX4表达<60%[第2组](图6,表5)。

[0202] [表5]

[0203] 与对照sgRNA(设为1.0)相比,用来自初步sgRNA筛检实验的11种最佳sgRNA处理的LCL的DUX4表达水平

[0204]

第1组

	表达
sgDUX4-122	0.3937
sgDUX4-126	0.2916
sgDUX4-130	0.273
sgDUX4-140	0.2856
sgDUX4-142	0.4622
sgDUX4-145	0.3729

第2组

	表达
sgDUX4-97	0.4204
sgDUX4-100	0.401
sgDUX4-119	0.3611
sgDUX4-128	0.3877
sgDUX4-155	0.4719

[0205] 接下来,我们对由初步筛检鉴定的6种最有效候选sgRNA[第1组]进行了验证筛检,这次选择了耐受嘌呤霉素17天的转导细胞来探索更长时间处理产生更佳抑制的可能性(图7)。

[0206] DUX4的表达导致许多下游靶标的异常上调,包括生殖系中和早期发育时表达的基因。TRIM43、ZSCAN4和MBD3L2是DUX4的下游靶标,发现它们在本研究所用的FSHD患者来源的LCL培养物中也上调。为了确定dCas9-KRAB介导的DUX4抑制是否也会抑制这些DUX4靶基因,我们测量了用验证实验所鉴定的最有效sgRNA处理的样品中的TRIM43、ZSCAN4和MBD3L2的水平。正如所料,所有这些dCas9-KRAB:sgRNA都显著降低了所有三个DUX4靶标的表达(图7),并且抑制效力与DUX4抑制效力密切相关(表6)。

[0207] [表6]

[0208] 用6种最佳sgRNA(sgDUX4-122、sgDUX4-126、sgDUX4-130、sgDUX4-140、sgDUX4-142、sgDUX4-145)处理的FSHD患者来源的LCL中DUX4、TRIM43、MBD3L2和ZSCAN4 mRNA水平的皮尔森相关性分析。

[0209]

	DUX4	TRIM43	MBD3L2	ZSCAN4
DUX4	1.000	0.988	0.964	0.926
TRIM43		1.000	0.983	0.885
MBD3L2			1.000	0.849
ZSCAN4				1.000

[0210] (表中所示的数字是DUX4、TRIM43、MBD3L2和ZSCAN4 mRNA水平中每一个的皮尔森相关性系数)

[0211] 工业实用性

[0212] 根据本发明,可以抑制DUX4基因在人细胞中的表达。因此,本发明预计对治疗和/或预防FSHD非常有用。

[0213] 在本文阐述数值界限或范围时,包括端点。此外,明确包括数值界限或范围内的所有值和子范围,就如同明确书写出来一样。

[0214] 如本文所用,措辞“一个”和“一种”等具有“一个或多个”的含义。

[0215] 显然,根据上述教导内容,本发明可能有许多修改和变化。因此,应当理解,在所附权利要求书的范围内,本发明可以不像本文具体描述那样实施。

[0216] 上述所有专利和其它参考文献以引用的方式全部并入本文,与详细阐述无异。

[0217] 本申请基于在美国提交的美国临时专利申请第63/072,327号(申请日期:2020年8月31日)和美国临时专利申请第63/235,359号(申请日期:2021年8月20日),二者的内容全部并入本文。

序列表

<110> 摩大力斯医疗株式会社(Modalis Therapeutics Corporation)

<120> 通过靶向DUX4基因来治疗面肩肱型肌营养不良症(FSHD)的方法

<130> 093184

<150> US63/072,327

<151> 2020-08-31

<150> US63/235,359

<151> 2021-08-20

<160> 182

<170> PatentIn第3.5版

<210> 1

<211> 22

<212> DNA

<213> 智人(Homo sapiens)

<400> 1

gggaggtgga gctgccccgg ct 22

<210> 2

<211> 22

<212> DNA

<213> 智人

<400> 2

ctcatccagc agcaggccgc ag 22

<210> 3

<211> 22

<212> DNA

<213> 智人

<400> 3

agcccggtat tcttcctcgc tg 22

<210> 4

<211> 22

<212> DNA

<213> 智人

<400> 4

ccctccaccg ggctgaccgg cc 22

<210> 5

<211> 22

<212> DNA

<213> 智人

<400> 5
gtgggccgcc tactgcgcac gc 22
<210> 6
<211> 22
<212> DNA
<213> 智人
<400> 6
ggggcccggt gtttcgcggg ac 22
<210> 7
<211> 22
<212> DNA
<213> 智人
<400> 7
gcgtcccggt gtgcgccggg cc 22
<210> 8
<211> 22
<212> DNA
<213> 智人
<400> 8
aacgggagac ctagaggggc gg 22
<210> 9
<211> 22
<212> DNA
<213> 智人
<400> 9
ggaaaagcgg tcctcggcct cc 22
<210> 10
<211> 22
<212> DNA
<213> 智人
<400> 10
cgggtgaaaa cccgacggca ac 22
<210> 11
<211> 22
<212> DNA
<213> 智人
<400> 11
cctgcgtgtg gctcctccgt gg 22
<210> 12

<211> 22
<212> DNA
<213> 智人
<400> 12
ttgcaccctt ccctgcatgt tt 22
<210> 13
<211> 22
<212> DNA
<213> 智人
<400> 13
cgccggggag gcatctcctc tc 22
<210> 14
<211> 22
<212> DNA
<213> 智人
<400> 14
gaactgaacc tccgtgggag tc 22
<210> 15
<211> 22
<212> DNA
<213> 智人
<400> 15
agagagcggc ttcccgttcc cg 22
<210> 16
<211> 22
<212> DNA
<213> 智人
<400> 16
ggccggctct ccggacctct cc 22
<210> 17
<211> 22
<212> DNA
<213> 智人
<400> 17
gtcgaggcct ggggccggcc gg 22
<210> 18
<211> 22
<212> DNA
<213> 智人

<400> 18
gccggcccca ggcctcgacg cc 22
<210> 19
<211> 22
<212> DNA
<213> 智人
<400> 19
cctcgacgcc ctgggtccct tc 22
<210> 20
<211> 22
<212> DNA
<213> 智人
<400> 20
gggggctcac cgccattcat ga 22
<210> 21
<211> 22
<212> DNA
<213> 智人
<400> 21
ggcaggcagg ctccacccct tc 22
<210> 22
<211> 22
<212> DNA
<213> 智人
<400> 22
ggccatcgcg gtgagccccg gc 22
<210> 23
<211> 22
<212> DNA
<213> 智人
<400> 23
ttccgcgggg agggtgctgt cc 22
<210> 24
<211> 22
<212> DNA
<213> 智人
<400> 24
ctcgtccccg ggcttccgcg gg 22
<210> 25

<211> 22
<212> DNA
<213> 智人
<400> 25
ctcgctttgg ctcggggtcc aa 22
<210> 26
<211> 22
<212> DNA
<213> 智人
<400> 26
aggccatcgg cattccggag cc 22
<210> 27
<211> 22
<212> DNA
<213> 智人
<400> 27
cggcgaaagc ggaccgccgt ca 22
<210> 28
<211> 22
<212> DNA
<213> 智人
<400> 28
gagagacggg cctcccggag tc 22
<210> 29
<211> 22
<212> DNA
<213> 智人
<400> 29
cctgtgcagc gcggcccccg gc 22
<210> 30
<211> 22
<212> DNA
<213> 智人
<400> 30
tgggtcgcct tcgcccacac cg 22
<210> 31
<211> 22
<212> DNA
<213> 智人

<400> 31
tttcagtttc cctttacttg ca 22
<210> 32
<211> 22
<212> DNA
<213> 智人
<400> 32
tccaaaatgc aacctgaagc ta 22
<210> 33
<211> 22
<212> DNA
<213> 智人
<400> 33
cactttcaat ggcgcctttt tc 22
<210> 34
<211> 22
<212> DNA
<213> 智人
<400> 34
ggaagatttc tctcaaaaac gg 22
<210> 35
<211> 22
<212> DNA
<213> 智人
<400> 35
gaggagtacg tcttctgcag cc 22
<210> 36
<211> 22
<212> DNA
<213> 智人
<400> 36
caggtctggg ctgagcattg ag 22
<210> 37
<211> 22
<212> DNA
<213> 智人
<400> 37
cctccctcca gggctatgga cc 22
<210> 38

<211> 22
<212> DNA
<213> 智人
<400> 38
agctcccagc atattaggtg gg 22
<210> 39
<211> 22
<212> DNA
<213> 智人
<400> 39
gtagcccttt gaccaaaggg tt 22
<210> 40
<211> 22
<212> DNA
<213> 智人
<400> 40
gcagccctgt gaccaaaggg ct 22
<210> 41
<211> 22
<212> DNA
<213> 智人
<400> 41
cccagcatgc accaggctca gg 22
<210> 42
<211> 22
<212> DNA
<213> 智人
<400> 42
cccagcatgc accaggctca gg 22
<210> 43
<211> 22
<212> DNA
<213> 智人
<400> 43
gaagctcagt tgtgttcact tt 22
<210> 44
<211> 22
<212> DNA
<213> 智人

<400> 44
aactatagct atctgcaata tc 22
<210> 45
<211> 22
<212> DNA
<213> 智人
<400> 45
atacgatcca gcaatttcac at 22
<210> 46
<211> 21
<212> DNA
<213> 智人
<400> 46
gtcaagcaga tgccaagtcg t 21
<210> 47
<211> 21
<212> DNA
<213> 智人
<400> 47
tcagtaaatg ttctgtcatt g 21
<210> 48
<211> 21
<212> DNA
<213> 智人
<400> 48
atgagccaag tgaaaacacc a 21
<210> 49
<211> 21
<212> DNA
<213> 智人
<400> 49
ggcatcacaa atggcaaacc a 21
<210> 50
<211> 21
<212> DNA
<213> 智人
<400> 50
aatgcataaa gcctgaaggg a 21
<210> 51

<211> 21
<212> DNA
<213> 智人
<400> 51
gaacggacag caaacgacaa g 21
<210> 52
<211> 21
<212> DNA
<213> 智人
<400> 52
atcgataaat acgctaacca c 21
<210> 53
<211> 21
<212> DNA
<213> 智人
<400> 53
gtcacccgca ttctttagag a 21
<210> 54
<211> 21
<212> DNA
<213> 智人
<400> 54
gaatattttg aaaattaagt g 21
<210> 55
<211> 21
<212> DNA
<213> 智人
<400> 55
tgctggagaa accctcttta t 21
<210> 56
<211> 21
<212> DNA
<213> 智人
<400> 56
tagcatctta aatctccacc c 21
<210> 57
<211> 21
<212> DNA
<213> 智人

<400> 57
aatcaaaatg cacatgctct g 21
<210> 58
<211> 21
<212> DNA
<213> 智人
<400> 58
acttttaaga cgtggagcac t 21
<210> 59
<211> 21
<212> DNA
<213> 智人
<400> 59
ttttggagct tgtccactta a 21
<210> 60
<211> 21
<212> DNA
<213> 智人
<400> 60
acaagactca cgcctcagaa a 21
<210> 61
<211> 21
<212> DNA
<213> 智人
<400> 61
cgtactcctc tggaaagaac c 21
<210> 62
<211> 21
<212> DNA
<213> 智人
<400> 62
ggaaaaggta aggtatcttt g 21
<210> 63
<211> 21
<212> DNA
<213> 智人
<400> 63
tgccgactct ttgcaaggag a 21
<210> 64

<211> 21
<212> DNA
<213> 智人
<400> 64
agctataatc ccagcatgta c 21
<210> 65
<211> 21
<212> DNA
<213> 智人
<400> 65
gcgcagccct ggaagaaggg g 21
<210> 66
<211> 21
<212> DNA
<213> 智人
<400> 66
caaagggttg ggagtgttta t 21
<210> 67
<211> 21
<212> DNA
<213> 智人
<400> 67
gcaagttgaa agaaggtccc t 21
<210> 68
<211> 21
<212> DNA
<213> 智人
<400> 68
gtagttctta ggagctagag g 21
<210> 69
<211> 21
<212> DNA
<213> 智人
<400> 69
tcattccact tatgagatac t 21
<210> 70
<211> 21
<212> DNA
<213> 智人

<400> 70
catgcatgac gtaccatgtg t 21
<210> 71
<211> 21
<212> DNA
<213> 智人
<400> 71
tgatggtaat tcaaacaaca c 21
<210> 72
<211> 21
<212> DNA
<213> 智人
<400> 72
acatttccac caacagtgta c 21
<210> 73
<211> 21
<212> DNA
<213> 智人
<400> 73
ggtattaagt gatatgtcat t 21
<210> 74
<211> 21
<212> DNA
<213> 智人
<400> 74
aagtagtttg ttttattgtt g 21
<210> 75
<211> 21
<212> DNA
<213> 智人
<400> 75
taggggaaac acgtcataac a 21
<210> 76
<211> 21
<212> DNA
<213> 智人
<400> 76
ctgtaagaaa aaattaacta a 21
<210> 77

<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 对照非靶向性靶向序列
<400> 77
acggaggcta agcgtcgcaa 20
<210> 78
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 对照非靶向性靶向序列
<400> 78
cgcttccgcg gcccgttcaa 20
<210> 79
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 对照非靶向性靶向序列
<400> 79
gtaggcgcgc cgctctctac 20
<210> 80
<211> 19
<212> RNA
<213> 新凶手弗朗西斯菌(Francisella novicid)
<220>
<221> misc_feature
<222> (1)..(19)
<223> crRNA的5'-把手
<400> 80
aauuucuacu guuguagau 19
<210> 81
<211> 1053
<212> PRT
<213> 金黄色葡萄球菌(Staphylococcus aureus)
<220>
<221> VARIANT

<222> (10)..(10)
<223> Asp残基转变为Ala残基
<220>
<221> VARIANT
<222> (580)..(580)
<223> Asn残基转变为Ala残基
<400> 81

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

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

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

Lys Tyr Tyr Gly Asn Lys Leu Asn Ala His Leu Asp Ile Thr Asp Asp
885 890 895
Tyr Pro Asn Ser Arg Asn Lys Val Val Lys Leu Ser Leu Lys Pro Tyr
900 905 910
Arg Phe Asp Val Tyr Leu Asp Asn Gly Val Tyr Lys Phe Val Thr Val
915 920 925
Lys Asn Leu Asp Val Ile Lys Lys Glu Asn Tyr Tyr Glu Val Asn Ser
930 935 940
Lys Cys Tyr Glu Glu Ala Lys Lys Leu Lys Lys Ile Ser Asn Gln Ala
945 950 955 960
Glu Phe Ile Ala Ser Phe Tyr Asn Asn Asp Leu Ile Lys Ile Asn Gly
965 970 975
Glu Leu Tyr Arg Val Ile Gly Val Asn Asn Asp Leu Leu Asn Arg Ile
980 985 990
Glu Val Asn Met Ile Asp Ile Thr Tyr Arg Glu Tyr Leu Glu Asn Met
995 1000 1005
Asn Asp Lys Arg Pro Pro Arg Ile Ile Lys Thr Ile Ala Ser Lys
1010 1015 1020
Thr Gln Ser Ile Lys Lys Tyr Ser Thr Asp Ile Leu Gly Asn Leu
1025 1030 1035
Tyr Glu Val Lys Ser Lys Lys His Pro Gln Ile Ile Lys Lys Gly
1040 1045 1050

<210> 82
<211> 1053
<212> PRT
<213> 金黄色葡萄球菌
<220>
<221> VARIANT
<222> (10)..(10)
<223> Asp残基转变为Ala残基
<220>
<221> VARIANT
<222> (557)..(557)
<223> His残基转变为Ala残基
<400> 82
Met Lys Arg Asn Tyr Ile Leu Gly Leu Ala Ile Gly Ile Thr Ser Val
1 5 10 15
Gly Tyr Gly Ile Ile Asp Tyr Glu Thr Arg Asp Val Ile Asp Ala Gly
20 25 30

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

340 345 350

Lys Ile Leu Thr Ile Tyr Gln Ser Ser Glu Asp Ile Gln Glu Glu Leu
355 360 365

Thr Asn Leu Asn Ser Glu Leu Thr Gln Glu Glu Ile Glu Gln Ile Ser
370 375 380

Asn Leu Lys Gly Tyr Thr Gly Thr His Asn Leu Ser Leu Lys Ala Ile
385 390 395 400

Asn Leu Ile Leu Asp Glu Leu Trp His Thr Asn Asp Asn Gln Ile Ala
405 410 415

Ile Phe Asn Arg Leu Lys Leu Val Pro Lys Lys Val Asp Leu Ser Gln
420 425 430

Gln Lys Glu Ile Pro Thr Thr Leu Val Asp Asp Phe Ile Leu Ser Pro
435 440 445

Val Val Lys Arg Ser Phe Ile Gln Ser Ile Lys Val Ile Asn Ala Ile
450 455 460

Ile Lys Lys Tyr Gly Leu Pro Asn Asp Ile Ile Ile Glu Leu Ala Arg
465 470 475 480

Glu Lys Asn Ser Lys Asp Ala Gln Lys Met Ile Asn Glu Met Gln Lys
485 490 495

Arg Asn Arg Gln Thr Asn Glu Arg Ile Glu Glu Ile Ile Arg Thr Thr
500 505 510

Gly Lys Glu Asn Ala Lys Tyr Leu Ile Glu Lys Ile Lys Leu His Asp
515 520 525

Met Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ala Ile Pro Leu Glu
530 535 540

Asp Leu Leu Asn Asn Pro Phe Asn Tyr Glu Val Asp Ala Ile Ile Pro
545 550 555 560

Arg Ser Val Ser Phe Asp Asn Ser Phe Asn Asn Lys Val Leu Val Lys
565 570 575

Gln Glu Glu Asn Ser Lys Lys Gly Asn Arg Thr Pro Phe Gln Tyr Leu
580 585 590

Ser Ser Ser Asp Ser Lys Ile Ser Tyr Glu Thr Phe Lys Lys His Ile
595 600 605

Leu Asn Leu Ala Lys Gly Lys Gly Arg Ile Ser Lys Thr Lys Lys Glu
610 615 620

Tyr Leu Leu Glu Glu Arg Asp Ile Asn Arg Phe Ser Val Gln Lys Asp
625 630 635 640

Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr Arg Gly Leu
645 650 655

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

965 970 975
Glu Leu Tyr Arg Val Ile Gly Val Asn Asn Asp Leu Leu Asn Arg Ile
980 985 990
Glu Val Asn Met Ile Asp Ile Thr Tyr Arg Glu Tyr Leu Glu Asn Met
995 1000 1005
Asn Asp Lys Arg Pro Pro Arg Ile Ile Lys Thr Ile Ala Ser Lys
1010 1015 1020
Thr Gln Ser Ile Lys Lys Tyr Ser Thr Asp Ile Leu Gly Asn Leu
1025 1030 1035
Tyr Glu Val Lys Ser Lys Lys His Pro Gln Ile Ile Lys Lys Gly
1040 1045 1050

<210> 83
<211> 1028
<212> PRT
<213> 人工序列
<220>
<223> 氨基酸残基(dSaCas9的第721个至第745个
氨基酸残基)缺失突变体
<220>
<221> VARIANT
<222> (10)..(10)
<223> Asp残基转变为Ala残基
<220>
<221> VARIANT
<222> (580)..(580)
<223> Asn残基转变为Ala残基
<400> 83
Met Lys Arg Asn Tyr Ile Leu Gly Leu Ala Ile Gly Ile Thr Ser Val
1 5 10 15
Gly Tyr Gly Ile Ile Asp Tyr Glu Thr Arg Asp Val Ile Asp Ala Gly
20 25 30
Val Arg Leu Phe Lys Glu Ala Asn Val Glu Asn Asn Glu Gly Arg Arg
35 40 45
Ser Lys Arg Gly Ala Arg Arg Leu Lys Arg Arg Arg Arg His Arg Ile
50 55 60
Gln Arg Val Lys Lys Leu Leu Phe Asp Tyr Asn Leu Leu Thr Asp His
65 70 75 80
Ser Glu Leu Ser Gly Ile Asn Pro Tyr Glu Ala Arg Val Lys Gly Leu
85 90 95

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

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

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

1025

<210> 84
<211> 6
<212> PRT
<213> 人工序列
<220>
<223> GGSGGS接头
<400> 84
Gly Gly Ser Gly Gly Ser
1 5
<210> 85
<211> 1034
<212> PRT
<213> 人工序列
<220>
<223> 具有GGSGGS接头的氨基酸残基(dSaCas9的第721个至第745个氨基酸残基)缺失突变体
<220>
<221> VARIANT
<222> (10)..(10)
<223> Asp残基转变为Ala残基
<220>
<221> VARIANT
<222> (580)..(580)
<223> Asn残基转变为Ala残基
<220>
<221> MISC_FEATURE
<222> (721)..(726)
<223> GGSGGS接头
<400> 85
Met Lys Arg Asn Tyr Ile Leu Gly Leu Ala Ile Gly Ile Thr Ser Val
1 5 10 15
Gly Tyr Gly Ile Ile Asp Tyr Glu Thr Arg Asp Val Ile Asp Ala Gly
20 25 30
Val Arg Leu Phe Lys Glu Ala Asn Val Glu Asn Asn Glu Gly Arg Arg
35 40 45
Ser Lys Arg Gly Ala Arg Arg Leu Lys Arg Arg Arg Arg His Arg Ile
50 55 60
Gln Arg Val Lys Lys Leu Leu Phe Asp Tyr Asn Leu Leu Thr Asp His

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

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

690 695 700
Ala Leu Ile Ile Ala Asn Ala Asp Phe Ile Phe Lys Glu Trp Lys Lys
705 710 715 720
Gly Gly Ser Gly Gly Ser Thr Glu Gln Glu Tyr Lys Glu Ile Phe Ile
725 730 735
Thr Pro His Gln Ile Lys His Ile Lys Asp Phe Lys Asp Tyr Lys Tyr
740 745 750
Ser His Arg Val Asp Lys Lys Pro Asn Arg Glu Leu Ile Asn Asp Thr
755 760 765
Leu Tyr Ser Thr Arg Lys Asp Asp Lys Gly Asn Thr Leu Ile Val Asn
770 775 780
Asn Leu Asn Gly Leu Tyr Asp Lys Asp Asn Asp Lys Leu Lys Lys Leu
785 790 795 800
Ile Asn Lys Ser Pro Glu Lys Leu Leu Met Tyr His His Asp Pro Gln
805 810 815
Thr Tyr Gln Lys Leu Lys Leu Ile Met Glu Gln Tyr Gly Asp Glu Lys
820 825 830
Asn Pro Leu Tyr Lys Tyr Tyr Glu Glu Thr Gly Asn Tyr Leu Thr Lys
835 840 845
Tyr Ser Lys Lys Asp Asn Gly Pro Val Ile Lys Lys Ile Lys Tyr Tyr
850 855 860
Gly Asn Lys Leu Asn Ala His Leu Asp Ile Thr Asp Asp Tyr Pro Asn
865 870 875 880
Ser Arg Asn Lys Val Val Lys Leu Ser Leu Lys Pro Tyr Arg Phe Asp
885 890 895
Val Tyr Leu Asp Asn Gly Val Tyr Lys Phe Val Thr Val Lys Asn Leu
900 905 910
Asp Val Ile Lys Lys Glu Asn Tyr Tyr Glu Val Asn Ser Lys Cys Tyr
915 920 925
Glu Glu Ala Lys Lys Leu Lys Lys Ile Ser Asn Gln Ala Glu Phe Ile
930 935 940
Ala Ser Phe Tyr Asn Asn Asp Leu Ile Lys Ile Asn Gly Glu Leu Tyr
945 950 955 960
Arg Val Ile Gly Val Asn Asn Asp Leu Leu Asn Arg Ile Glu Val Asn
965 970 975
Met Ile Asp Ile Thr Tyr Arg Glu Tyr Leu Glu Asn Met Asn Asp Lys
980 985 990
Arg Pro Pro Arg Ile Ile Lys Thr Ile Ala Ser Lys Thr Gln Ser Ile
995 1000 1005

Lys Lys Tyr Ser Thr Asp Ile Leu Gly Asn Leu Tyr Glu Val Lys
1010 1015 1020
Ser Lys Lys His Pro Gln Ile Ile Lys Lys Gly
1025 1030

<210> 86
<211> 5
<212> PRT
<213> 人工序列
<220>
<223> SGGGS接头
<400> 86
Ser Gly Gly Gly Ser
1 5

<210> 87
<211> 1033
<212> PRT
<213> 人工序列
<220>
<223> 具有SGGGS接头的氨基酸残基(dSaCas9的第721个至第745个氨基酸残基)缺失突变体
<220>
<221> VARIANT
<222> (10)..(10)
<223> Asp残基转变为Ala残基
<220>
<221> VARIANT
<222> (580)..(580)
<223> Asn残基转变为Ala残基
<220>
<221> MISC_FEATURE
<222> (721)..(725)
<223> SGGGS接头
<400> 87
Met Lys Arg Asn Tyr Ile Leu Gly Leu Ala Ile Gly Ile Thr Ser Val
1 5 10 15
Gly Tyr Gly Ile Ile Asp Tyr Glu Thr Arg Asp Val Ile Asp Ala Gly
20 25 30
Val Arg Leu Phe Lys Glu Ala Asn Val Glu Asn Asn Glu Gly Arg Arg
35 40 45

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

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

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

980 985 990
Pro Pro Arg Ile Ile Lys Thr Ile Ala Ser Lys Thr Gln Ser Ile Lys
995 1000 1005
Lys Tyr Ser Thr Asp Ile Leu Gly Asn Leu Tyr Glu Val Lys Ser
1010 1015 1020
Lys Lys His Pro Gln Ile Ile Lys Lys Gly
1025 1030

<210> 88
<211> 886
<212> PRT
<213> 人工序列
<220>
<223> 氨基酸残基(dSaCas9的第482个至第648个氨基酸残基)缺失突变体
<220>
<221> VARIANT
<222> (10)..(10)
<223> Asp残基转变为Ala残基
<400> 88
Met Lys Arg Asn Tyr Ile Leu Gly Leu Ala Ile Gly Ile Thr Ser Val
1 5 10 15
Gly Tyr Gly Ile Ile Asp Tyr Glu Thr Arg Asp Val Ile Asp Ala Gly
20 25 30
Val Arg Leu Phe Lys Glu Ala Asn Val Glu Asn Asn Glu Gly Arg Arg
35 40 45
Ser Lys Arg Gly Ala Arg Arg Leu Lys Arg Arg Arg Arg His Arg Ile
50 55 60
Gln Arg Val Lys Lys Leu Leu Phe Asp Tyr Asn Leu Leu Thr Asp His
65 70 75 80
Ser Glu Leu Ser Gly Ile Asn Pro Tyr Glu Ala Arg Val Lys Gly Leu
85 90 95
Ser Gln Lys Leu Ser Glu Glu Glu Phe Ser Ala Ala Leu Leu His Leu
100 105 110
Ala Lys Arg Arg Gly Val His Asn Val Asn Glu Val Glu Glu Asp Thr
115 120 125
Gly Asn Glu Leu Ser Thr Lys Glu Gln Ile Ser Arg Asn Ser Lys Ala
130 135 140
Leu Glu Glu Lys Tyr Val Ala Glu Leu Gln Leu Glu Arg Leu Lys Lys
145 150 155 160

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

465 470 475 480
Glu Thr Arg Tyr Ala Thr Arg Gly Leu Met Asn Leu Leu Arg Ser Tyr
485 490 495
Phe Arg Val Asn Asn Leu Asp Val Lys Val Lys Ser Ile Asn Gly Gly
500 505 510
Phe Thr Ser Phe Leu Arg Arg Lys Trp Lys Phe Lys Lys Glu Arg Asn
515 520 525
Lys Gly Tyr Lys His His Ala Glu Asp Ala Leu Ile Ile Ala Asn Ala
530 535 540
Asp Phe Ile Phe Lys Glu Trp Lys Lys Leu Asp Lys Ala Lys Lys Val
545 550 555 560
Met Glu Asn Gln Met Phe Glu Glu Lys Gln Ala Glu Ser Met Pro Glu
565 570 575
Ile Glu Thr Glu Gln Glu Tyr Lys Glu Ile Phe Ile Thr Pro His Gln
580 585 590
Ile Lys His Ile Lys Asp Phe Lys Asp Tyr Lys Tyr Ser His Arg Val
595 600 605
Asp Lys Lys Pro Asn Arg Glu Leu Ile Asn Asp Thr Leu Tyr Ser Thr
610 615 620
Arg Lys Asp Asp Lys Gly Asn Thr Leu Ile Val Asn Asn Leu Asn Gly
625 630 635 640
Leu Tyr Asp Lys Asp Asn Asp Lys Leu Lys Lys Leu Ile Asn Lys Ser
645 650 655
Pro Glu Lys Leu Leu Met Tyr His His Asp Pro Gln Thr Tyr Gln Lys
660 665 670
Leu Lys Leu Ile Met Glu Gln Tyr Gly Asp Glu Lys Asn Pro Leu Tyr
675 680 685
Lys Tyr Tyr Glu Glu Thr Gly Asn Tyr Leu Thr Lys Tyr Ser Lys Lys
690 695 700
Asp Asn Gly Pro Val Ile Lys Lys Ile Lys Tyr Tyr Gly Asn Lys Leu
705 710 715 720
Asn Ala His Leu Asp Ile Thr Asp Asp Tyr Pro Asn Ser Arg Asn Lys
725 730 735
Val Val Lys Leu Ser Leu Lys Pro Tyr Arg Phe Asp Val Tyr Leu Asp
740 745 750
Asn Gly Val Tyr Lys Phe Val Thr Val Lys Asn Leu Asp Val Ile Lys
755 760 765
Lys Glu Asn Tyr Tyr Glu Val Asn Ser Lys Cys Tyr Glu Glu Ala Lys
770 775 780

Lys Leu Lys Lys Ile Ser Asn Gln Ala Glu Phe Ile Ala Ser Phe Tyr
785 790 795 800
Asn Asn Asp Leu Ile Lys Ile Asn Gly Glu Leu Tyr Arg Val Ile Gly
805 810 815
Val Asn Asn Asp Leu Leu Asn Arg Ile Glu Val Asn Met Ile Asp Ile
820 825 830
Thr Tyr Arg Glu Tyr Leu Glu Asn Met Asn Asp Lys Arg Pro Pro Arg
835 840 845
Ile Ile Lys Thr Ile Ala Ser Lys Thr Gln Ser Ile Lys Lys Tyr Ser
850 855 860
Thr Asp Ile Leu Gly Asn Leu Tyr Glu Val Lys Ser Lys Lys His Pro
865 870 875 880
Gln Ile Ile Lys Lys Gly
885

<210> 89
<211> 892
<212> PRT
<213> 人工序列
<220>
<223> 具有GGSGGS接头的氨基酸残基(dSaCas9的第482个至第648个氨基酸残基)缺失突变体
<220>
<221> VARIANT
<222> (10)..(10)
<223> Asp残基转变为Ala残基
<220>
<221> MISC_FEATURE
<222> (482)..(487)
<223> GGSGGS接头
<400> 89
Met Lys Arg Asn Tyr Ile Leu Gly Leu Ala Ile Gly Ile Thr Ser Val
1 5 10 15
Gly Tyr Gly Ile Ile Asp Tyr Glu Thr Arg Asp Val Ile Asp Ala Gly
20 25 30
Val Arg Leu Phe Lys Glu Ala Asn Val Glu Asn Asn Glu Gly Arg Arg
35 40 45
Ser Lys Arg Gly Ala Arg Arg Leu Lys Arg Arg Arg Arg His Arg Ile
50 55 60
Gln Arg Val Lys Lys Leu Leu Phe Asp Tyr Asn Leu Leu Thr Asp His

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

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

690 695 700

Thr Lys Tyr Ser Lys Lys Asp Asn Gly Pro Val Ile Lys Lys Ile Lys
705 710 715 720

Tyr Tyr Gly Asn Lys Leu Asn Ala His Leu Asp Ile Thr Asp Asp Tyr
725 730 735

Pro Asn Ser Arg Asn Lys Val Val Lys Leu Ser Leu Lys Pro Tyr Arg
740 745 750

Phe Asp Val Tyr Leu Asp Asn Gly Val Tyr Lys Phe Val Thr Val Lys
755 760 765

Asn Leu Asp Val Ile Lys Lys Glu Asn Tyr Tyr Glu Val Asn Ser Lys
770 775 780

Cys Tyr Glu Glu Ala Lys Lys Leu Lys Lys Ile Ser Asn Gln Ala Glu
785 790 795 800

Phe Ile Ala Ser Phe Tyr Asn Asn Asp Leu Ile Lys Ile Asn Gly Glu
805 810 815

Leu Tyr Arg Val Ile Gly Val Asn Asn Asp Leu Leu Asn Arg Ile Glu
820 825 830

Val Asn Met Ile Asp Ile Thr Tyr Arg Glu Tyr Leu Glu Asn Met Asn
835 840 845

Asp Lys Arg Pro Pro Arg Ile Ile Lys Thr Ile Ala Ser Lys Thr Gln
850 855 860

Ser Ile Lys Lys Tyr Ser Thr Asp Ile Leu Gly Asn Leu Tyr Glu Val
865 870 875 880

Lys Ser Lys Lys His Pro Gln Ile Ile Lys Lys Gly
885 890

<210> 90
<211> 48
<212> DNA
<213> 人工序列
<220>
<223> 编码NLS的DNA序列
<400> 90
gccccaaaga agaagcggaa ggtcggtatc cacggagtcc cagcagcc 48
<210> 91
<211> 48
<212> DNA
<213> 人工序列
<220>
<223> 编码NLS的DNA序列

<400> 91
aaaaggccgg cggccacgaa aaaggccggc caggcaaaaa agaaaaag 48
<210> 92
<211> 3477
<212> DNA
<213> 人工序列
<220>
<223> 与KRAB (DNA)融合的dSaCas9
<400> 92
atggccccaa agaagaagcg gaaggtcggt atccacggag tcccagcagc caagcggaac 60
tacatcctgg gcctggccat cggcatcacc agcgtgggct acggcatcat cgactacgag 120
acacgggacg tgatcgatgc cggcgtgcgg ctgttcaaag aggccaacgt ggaaaacaac 180
gagggcaggc ggagcaagag aggcgccaga aggctgaagc ggcggaggcg gcatagaatc 240
cagagagtga agaagctgct gttcgactac aacctgctga ccgaccacag cgagctgagc 300
ggcatcaacc cctacgaggc cagagtgaag ggcctgagcc agaagctgag cgaggaagag 360
ttctctgccg ccctgctgca cctggccaag agaagaggcg tgcacaacgt gaacgaggtg 420
gaagaggaca ccggcaacga gctgtccacc aaagagcaga tcagccggaa cagcaaggcc 480
ctggaagaga aatacgtggc cgaactgcag ctggaacggc tgaagaaaga cggcgaagtg 540
cggggcagca tcaacagatt caagaccagc gactacgtga aagaagccaa acagctgctg 600
aaggtgcaga aggcctacca ccagctggac cagagcttca tcgacaccta catcgacctg 660
ctggaaaccc ggcggaccta ctatgaggga cctggcgagg gcagcccctt cggctggaag 720
gacatcaaag aatggtacga gatgctgatg ggccactgca cctacttccc cgaggaactg 780
cggagcgtga agtacgccta caacgccgac ctgtacaacg ccctgaacga cctgaacaat 840
ctcgtgatca ccagggacga gaacgagaag ctggaatatt acgagaagtt ccagatcatc 900
gagaacgtgt tcaagcagaa gaagaagccc accctgaagc agatcgccaa agaaatcctc 960
gtgaacgaag aggatattaa gggctacaga gtgaccagca ccggcaagcc cgagttcacc 1020
aacctgaagg tgtaccacga catcaaggac attaccgccc ggaaagagat tattgagaac 1080
gccgagctgc tggatcagat tgccaagatc ctgaccatct accagagcag cgaggacatc 1140
caggaagaac tgaccaatct gaactccgag ctgacccagg aagagatcga gcagatctct 1200
aatctgaagg gctataccgg cacccacaac ctgagcctga aggccatcaa cctgatcctg 1260
gacgagctgt ggcacaccaa cgacaaccag atcgctatct tcaaccggct gaagctggtg 1320
cccaagaagg tggacctgtc ccagcagaaa gagatcccca ccaccctggt ggacgacttc 1380
atcctgagcc ccgtcgtgaa gagaagcttc atccagagca tcaaagtgat caacgccatc 1440
atcaagaagt acggcctgcc caacgacatc attatcgagc tggcccgcga gaagaactcc 1500
aaggacgccc agaaaatgat caacgagatg cagaagcgga accggcagac caacgagcgg 1560
atcgaggaaa tcatccggac caccggcaaa gagaacgcca agtacctgat cgagaagatc 1620
aagctgcacg acatgcagga aggcaagtgc ctgtacagcc tggaagccat ccctctggaa 1680
gatctgctga acaacccctt caactatgag gtggaccaca tcatccccag aagcgtgtcc 1740
ttcgacaaca gcttcaacaa caaggtgctc gtgaagcagg aagaagccag caagaagggc 1800

aaccggaccc cattccagta cctgagcagc agcgacagca agatcagcta cgaaaccttc 1860
aagaagcaca tcctgaatct ggccaagggc aagggcagaa tcagcaagac caagaaagag 1920
tatctgctgg aagaacggga catcaacagg ttctccgtgc agaaagactt catcaaccgg 1980
aacctggtgg ataccagata cgccaccaga ggcctgatga acctgctgcg gagctacttc 2040
agagtgaaca acctggacgt gaaagtgaag tccatcaatg gcggcttcac cagctttctg 2100
cggcggaagt ggaagtttaa gaaagagcgg aacaaggggt acaagcacca cgccgaggac 2160
gccctgatca ttgccaacgc cgatttcatc ttcaaagagt ggaagaaact ggacaaggcc 2220
aaaaaagtga tggaaaacca gatgttcgag gaaaagcagg ccgagagcat gcccgagatc 2280
gaaaccgagc aggagtacaa agagatcttc atcacccccc accagatcaa gcacattaag 2340
gacttcaagg actacaagta cagccaccgg gtggacaaga agcctaatag agagctgatt 2400
aacgacaccc tgtactccac ccggaaggac gacaagggca acaccctgat cgtgaacaat 2460
ctgaacggcc tgtacgacaa ggacaatgac aagctgaaaa agctgatcaa caagagcccc 2520
gaaaagctgc tgatgtacca ccacgacccc cagacctacc agaaactgaa gctgattatg 2580
gaacagtacg gcgacgagaa gaatcccctg tacaagtact acgaggaaac cgggaactac 2640
ctgaccaagt actccaaaaa ggacaacggc cccgtgatca agaagattaa gtattacggc 2700
aacaaactga acgcccatct ggacatcacc gacgactacc ccaacagcag aaacaaggtc 2760
gtgaagctgt ccctgaagcc ctacagattc gacgtgtacc tggacaatgg cgtgtacaag 2820
ttcgtgaccg tgaagaatct ggatgtgatc aaaaaagaaa actactacga agtgaatagc 2880
aagtgctatg aggaagctaa gaagctgaag aagatcagca accaggccga gtttatcgcc 2940
tccttctaca acaacgatct gatcaagatc aacggcgagc tgtatagagt gatcggcgtg 3000
aacaacgacc tgctgaaccg gatcgaagtg aacatgatcg acatcaccta ccgcgagtac 3060
ctggaaaaca tgaacgacaa gaggcccccc aggatcatta agacaatcgc ctccaagacc 3120
cagagcatta agaagtacag cacagacatt ctgggcaacc tgtatgaagt gaaatctaag 3180
aagcaccctc agatcatcaa aaagggcaaa aggccggcgg ccacgaaaaa ggccggccag 3240
gcaaaaaaga aaaagggatc catggatgct aagtcactaa ctgcctggtc ccggacactg 3300
gtgaccttca aggatgtatt tgtggacttc accagggagg agtggaagct gctggacact 3360
gctcagcaga tcgtgtacag aaatgtgatg ctggagaact ataagaacct ggtttccttg 3420
ggttatcagc ttactaagcc agatgtgatc ctccggttgg agaagggaga agagccc 3477

<210> 93
<211> 22
<212> RNA
<213> 智人
<220>
<221> misc_feature
<222> (1)..(22)
<223> 与靶序列(SEQ ID NO:4)对应的crRNA
<400> 93
cccuccaccg ggcugaccgg cc 22

<210> 94

<211> 22
<212> DNA
<213> 智人
<220>
<221> misc_feature
<222> (1)..(22)
<223> 与靶序列(SEQ ID NO:4)互补的序列
<400> 94
ggccggtcag cccggtggag gg 22
<210> 95
<211> 19
<212> DNA
<213> 新凶手弗朗西斯菌
<220>
<221> misc_feature
<223> crRNA的5'-把手
<400> 95
aatttctact gttgtagat 19
<210> 96
<211> 60
<212> DNA
<213> 金黄色葡萄球菌
<220>
<221> misc_feature
<222> (1)..(83)
<223> 编码tracrRNA的序列
<400> 96
gttttagtac tctggaaaca gaatctacta aaacaaggca aaatgccgtg tttatctcgt 60
<210> 97
<211> 82
<212> DNA
<213> 金黄色葡萄球菌
<220>
<221> misc_feature
<222> (1)..(82)
<223> tracRNA
<400> 97
gttttagtac tctggaaaca gaatctacta aaacaaggca aaatgccgtg tttatctcgt 60
caacttgttg gcgagatttt tt 82

<210> 98
<211> 22
<212> DNA
<213> 智人
<400> 98
ctaaggcttt ttctctccct cc 22
<210> 99
<211> 22
<212> DNA
<213> 智人
<400> 99
tgggaggatt ttgcctgtga gt 22
<210> 100
<211> 22
<212> DNA
<213> 智人
<400> 100
agattctggg aggattttgc ct 22
<210> 101
<211> 22
<212> DNA
<213> 智人
<400> 101
gaactcacag gcaaaatcct cc 22
<210> 102
<211> 22
<212> DNA
<213> 智人
<400> 102
ttatgttctc acaagattct gg 22
<210> 103
<211> 22
<212> DNA
<213> 智人
<400> 103
tgactagttt ggcattgctt tt 22
<210> 104
<211> 22
<212> DNA

<213> 智人
<400> 104
gatttataaa taatggcatg ac 22
<210> 105
<211> 22
<212> DNA
<213> 智人
<400> 105
gtgaggggag atggggagac at 22
<210> 106
<211> 22
<212> DNA
<213> 智人
<400> 106
ccagccaggc cgcgccggca ga 22
<210> 107
<211> 22
<212> DNA
<213> 智人
<400> 107
ctcccaacct gccccggcgc gc 22
<210> 108
<211> 22
<212> DNA
<213> 智人
<400> 108
tgcggaggcc accgaggagc ct 22
<210> 109
<211> 22
<212> DNA
<213> 智人
<400> 109
tcccggtcct cccggctttt gc 22
<210> 110
<211> 22
<212> DNA
<213> 智人
<400> 110
gggcccggca ggccgtcgcg ct 22

<210> 111
<211> 22
<212> DNA
<213> 智人
<400> 111
ctcaagcggg gccgcagggc ca 22
<210> 112
<211> 22
<212> DNA
<213> 智人
<400> 112
ccaccacgga ctcccctggg ac 22
<210> 113
<211> 22
<212> DNA
<213> 智人
<400> 113
gcttgcgcca cccacgtccc ag 22
<210> 114
<211> 22
<212> DNA
<213> 智人
<400> 114
gagtccgtgg tggggctggg gc 22
<210> 115
<211> 22
<212> DNA
<213> 智人
<400> 115
ctgctggagg agctttagga cg 22
<210> 116
<211> 22
<212> DNA
<213> 智人
<400> 116
gctttaggac gcggggttgg ga 22
<210> 117
<211> 22
<212> DNA

<213> 智人
<400> 117
aggacgcggg gttgggacgg gg 22
<210> 118
<211> 22
<212> DNA
<213> 智人
<400> 118
caccggcctg gacctagaag gc 22
<210> 119
<211> 22
<212> DNA
<213> 智人
<400> 119
agaatggcag ttctccgcgg tg 22
<210> 120
<211> 22
<212> DNA
<213> 智人
<400> 120
gggatccccg ggatgcccag ga 22
<210> 121
<211> 22
<212> DNA
<213> 智人
<400> 121
gccattcttt cctgggcatc cc 22
<210> 122
<211> 22
<212> DNA
<213> 智人
<400> 122
cctgggccgg ctctgggatc cc 22
<210> 123
<211> 22
<212> DNA
<213> 智人
<400> 123
ctgctggtac ctgggccggc tc 22

<210> 124
<211> 22
<212> DNA
<213> 智人
<400> 124
tggggatggg gcggtcaggc gg 22
<210> 125
<211> 22
<212> DNA
<213> 智人
<400> 125
ccgggggtgg ggggtggggg gt 22
<210> 126
<211> 22
<212> DNA
<213> 智人
<400> 126
cgttttccgg gggtgggggg tg 22
<210> 127
<211> 22
<212> DNA
<213> 智人
<400> 127
gacgacgcgt tttccggggg tg 22
<210> 128
<211> 22
<212> DNA
<213> 智人
<400> 128
cccaggggac gacgcgtttt cc 22
<210> 129
<211> 22
<212> DNA
<213> 智人
<400> 129
ggaaaacgcg tcgtcccctg gg 22
<210> 130
<211> 22
<212> DNA

<213> 智人
<400> 130
gctgaccgtt ttcccggagg gc 22
<210> 131
<211> 22
<212> DNA
<213> 智人
<400> 131
ctgggccccg gaaccggggc ga 22
<210> 132
<211> 22
<212> DNA
<213> 智人
<400> 132
ctccctgggc cccggaaccg gg 22
<210> 133
<211> 22
<212> DNA
<213> 智人
<400> 133
ttcgccccgg ttccggggcc ca 22
<210> 134
<211> 22
<212> DNA
<213> 智人
<400> 134
ctccgggaca aaagaccggg ac 22
<210> 135
<211> 22
<212> DNA
<213> 智人
<400> 135
aagaccggga ctcgggttgc cg 22
<210> 136
<211> 22
<212> DNA
<213> 智人
<400> 136
ggatgtgcgg tctgtgaacc gc 22

<210> 137
<211> 22
<212> DNA
<213> 智人
<400> 137
gccgcgttgc agggctcagc ct 22
<210> 138
<211> 22
<212> DNA
<213> 智人
<400> 138
gggcacccgg aaacatgcag gg 22
<210> 139
<211> 22
<212> DNA
<213> 智人
<400> 139
attcccgcgt gcggcaacgt gg 22
<210> 140
<211> 22
<212> DNA
<213> 智人
<400> 140
actcccccac gttgccgcac gc 22
<210> 141
<211> 22
<212> DNA
<213> 智人
<400> 141
tccccggcgt gatggcctga cg 22
<210> 142
<211> 22
<212> DNA
<213> 智人
<400> 142
gagtgtggaa ctgaacctcc gt 22
<210> 143
<211> 22
<212> DNA

<213> 智人
<400> 143
aaaccagcct gggagggtgg ag 22
<210> 144
<211> 22
<212> DNA
<213> 智人
<400> 144
cagcagggag aaaccagcct gg 22
<210> 145
<211> 22
<212> DNA
<213> 智人
<400> 145
ctcgcagggc ctcgctttgg ct 22
<210> 146
<211> 22
<212> DNA
<213> 智人
<400> 146
tctctggtgg cgatgcccgg gt 22
<210> 147
<211> 22
<212> DNA
<213> 智人
<400> 147
agccgttctc tggtggcgat gc 22
<210> 148
<211> 22
<212> DNA
<213> 智人
<400> 148
accaaatctg gaccctgggc tc 22
<210> 149
<211> 22
<212> DNA
<213> 智人
<400> 149
gcccagggtc cagatttggt tt 22

<210> 150
<211> 22
<212> DNA
<213> 智人
<400> 150
cgccagctga ggcagcaccg gc 22
<210> 151
<211> 22
<212> DNA
<213> 智人
<400> 151
ggctcggagg agcagggcgg tc 22
<210> 152
<211> 22
<212> DNA
<213> 智人
<400> 152
cctgctcctc cgagcctttg ag 22
<210> 153
<211> 22
<212> DNA
<213> 智人
<400> 153
ctggccagag agacgggcct cc 22
<210> 154
<211> 22
<212> DNA
<213> 智人
<400> 154
cccttcgatt ctgaaaccag at 22
<210> 155
<211> 22
<212> DNA
<213> 智人
<400> 155
gtccaggatt cagatctggt tt 22
<210> 156
<211> 22
<212> DNA

<213> 智人
<400> 156
tcgaagggcc aggcacccgg ga 22
<210> 157
<211> 22
<212> DNA
<213> 智人
<400> 157
gcgggcgccc tgccaccctg tc 22
<210> 158
<211> 22
<212> DNA
<213> 智人
<400> 158
gcgaaggcga cccacgaggg ag 22
<210> 159
<211> 22
<212> DNA
<213> 智人
<400> 159
gcgggggtca ccctgctccc tc 22
<210> 160
<211> 22
<212> DNA
<213> 智人
<400> 160
agccccaggc gcgcagggca cg 22
<210> 161
<211> 22
<212> DNA
<213> 智人
<400> 161
accgggccta gacctagaag gc 22
<210> 162
<211> 22
<212> DNA
<213> 智人
<400> 162
gcgttttccg ggggtggggg gt 22

<210> 163
<211> 22
<212> DNA
<213> 智人
<400> 163
cgtccccggt gtgcgccggg cc 22
<210> 164
<211> 22
<212> DNA
<213> 智人
<400> 164
cgggtgaaga cccgacggca ac 22
<210> 165
<211> 22
<212> DNA
<213> 智人
<400> 165
ggatgtgggg tctgtgaacc gc 22
<210> 166
<211> 22
<212> DNA
<213> 智人
<400> 166
gactccccac gttgccgcac gc 22
<210> 167
<211> 22
<212> DNA
<213> 智人
<400> 167
ggtggtggtg gtggtggggg gg 22
<210> 168
<211> 22
<212> DNA
<213> 智人
<400> 168
ccagccaggc cgcgccggca ga 22
<210> 169
<211> 22
<212> DNA

<213> 智人
<400> 169
accgggcctg gacctagaag gc 22
<210> 170
<211> 22
<212> DNA
<213> 智人
<400> 170
tggggagggg gcggtcaggc gg 22
<210> 171
<211> 22
<212> DNA
<213> 智人
<400> 171
gattcccgcg tgcggcaacg tg 22
<210> 172
<211> 22
<212> DNA
<213> 智人
<400> 172
tggtggtggt ggtggtgggg gg 22
<210> 173
<211> 22
<212> DNA
<213> 智人
<400> 173
agaaaggcag ttctccgcgg ag 22
<210> 174
<211> 22
<212> DNA
<213> 智人
<400> 174
aggaaagaaa ggcagttctc cg 22
<210> 175
<211> 22
<212> DNA
<213> 智人
<400> 175
gcctttcttt cctgggcatc cc 22

<210> 176
<211> 22
<212> DNA
<213> 智人
<400> 176
gcgagctccc ttgcacgtca gc 22
<210> 177
<211> 22
<212> DNA
<213> 智人
<400> 177
ttgttcttcc gtgaaattct gg 22
<210> 178
<211> 22
<212> DNA
<213> 智人
<400> 178
aaggtggggg gagacattca gc 22
<210> 179
<211> 22
<212> DNA
<213> 智人
<400> 179
ttccgacgct gtctaggcaa ac 22
<210> 180
<211> 22
<212> DNA
<213> 智人
<400> 180
cgctgtctag gcaaacctgg at 22
<210> 181
<211> 22
<212> DNA
<213> 智人
<400> 181
acctggatta gagttacatc tc 22
<210> 182
<211> 22
<212> DNA

<213> 智人

<400> 182

tatattaaaa tgccccctcc ct 22

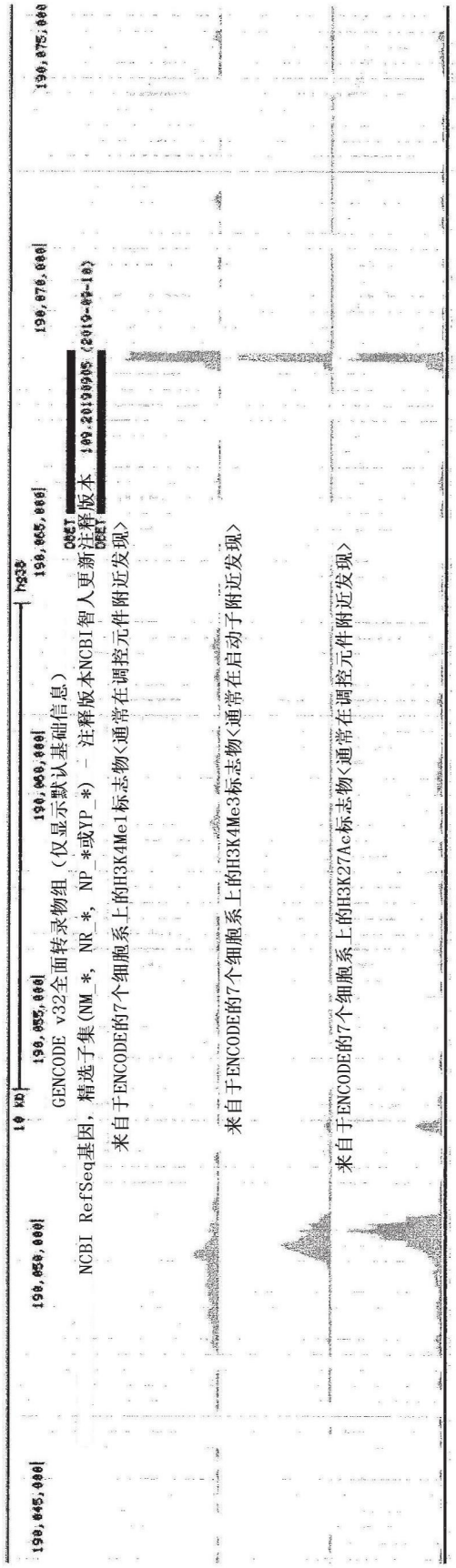

hg38
190,845,000
190,850,000
190,855,000
190,860,000
190,865,000
190,870,000
190,875,000
GENCODE v32全面转录物组（仅显示默认基础信息）
NCBI RefSeq基因，精选子集(NM_*, NR_*, NP_*或YP_*) - 注释版本NCBI智人更新注释版本 109.20190905 (2019-09-10)
来自于ENCODE的7个细胞系上的H3K4Me1标志物<通常在调控元件附近发现>
来自于ENCODE的7个细胞系上的H3K4Me3标志物<通常在启动子附近发现>
来自于ENCODE的7个细胞系上的H3K27Ac标志物<通常在调控元件附近发现>

图1

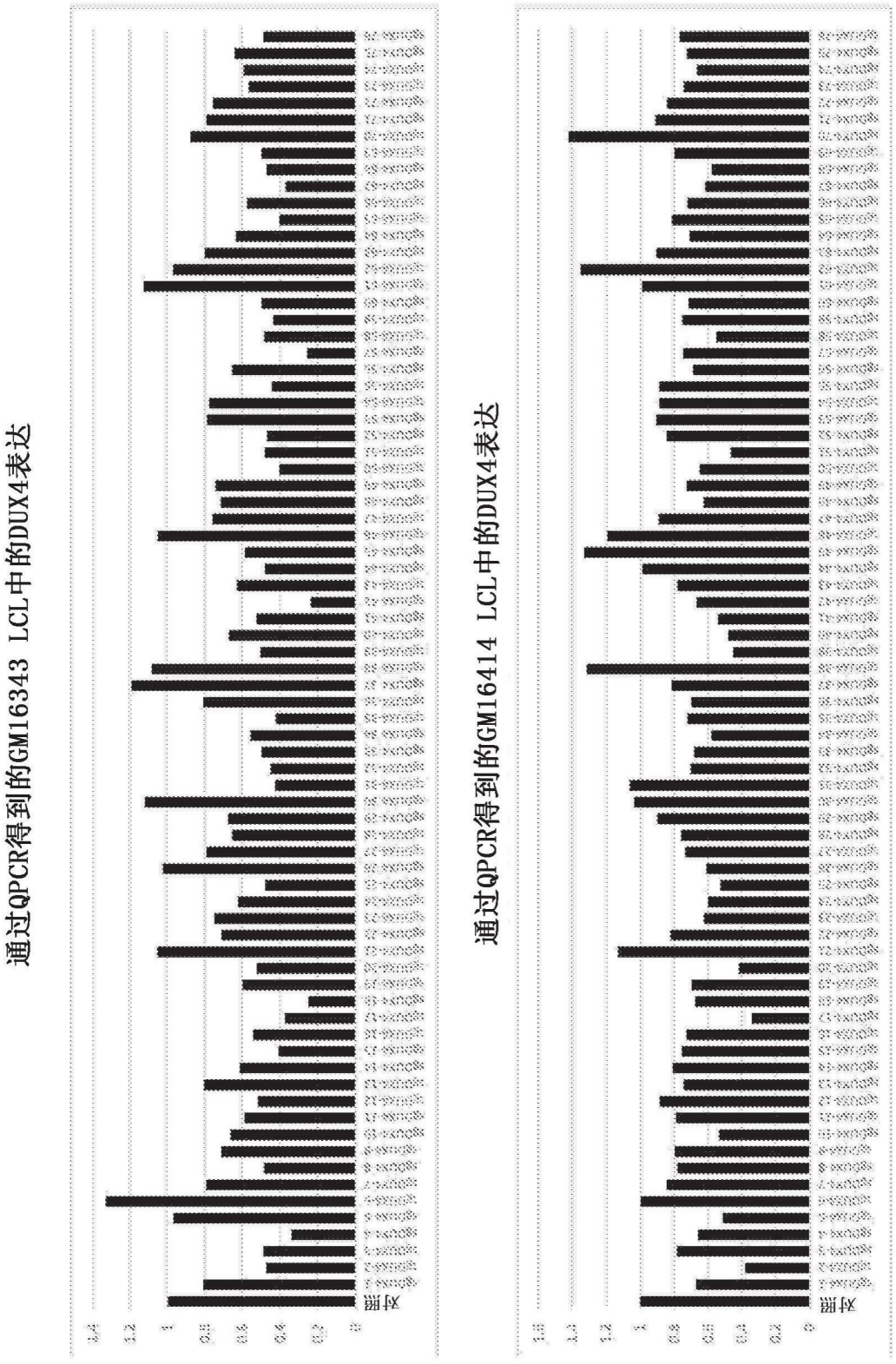

通过QPCR得到的GM16343 LCL中的DUX4表达
对照
通过QPCR得到的GM16414 LCL中的DUX4表达
对照

图2

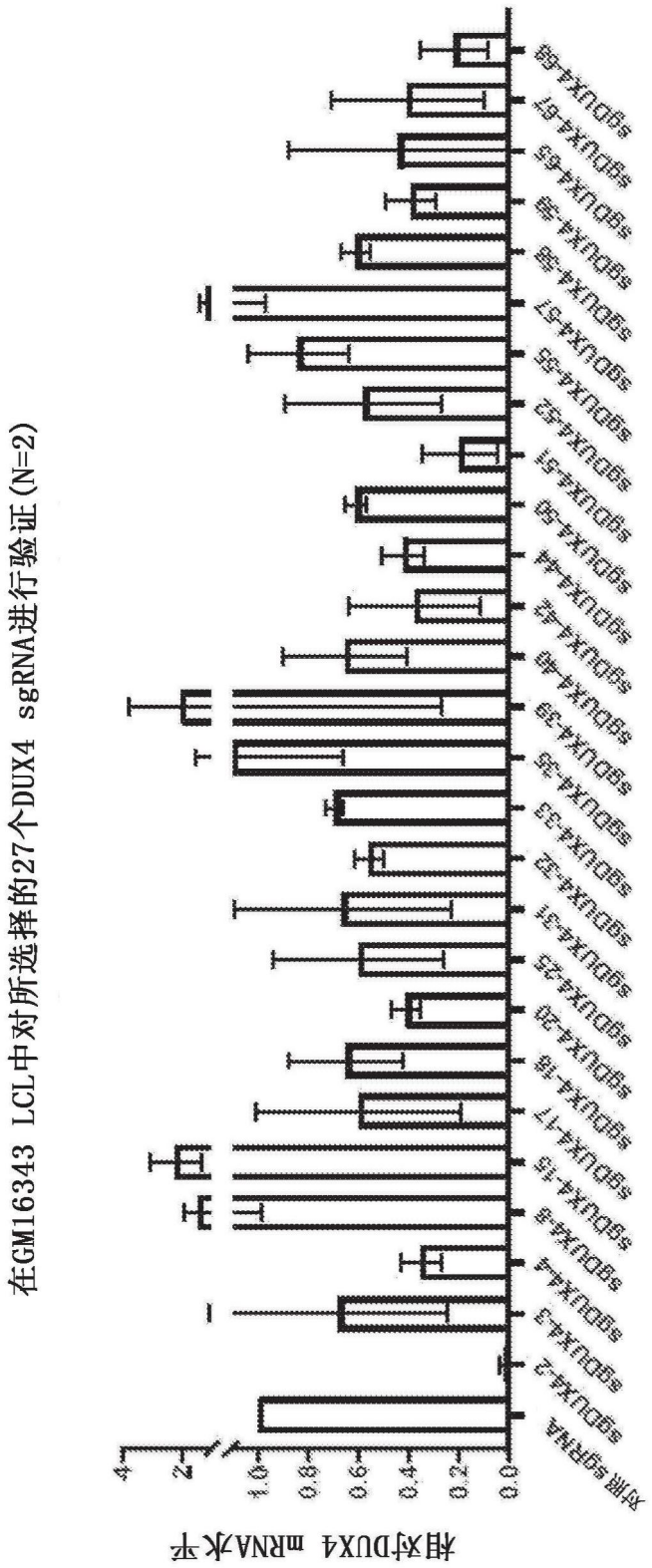

在GM16343 LCL中对所选择的27个DUX4 sgRNA进行验证(N=2)
相对DUX4 mRNA水平
4
2
1.0
0.8
0.6
0.4
0.2
0.0
对照sgRNA
sgDUX4-2
sgDUX4-3
sgDUX4-4
sgDUX4-8
sgDUX4-15
sgDUX4-17
sgDUX4-18
sgDUX4-20
sgDUX4-25
sgDUX4-31
sgDUX4-32
sgDUX4-33
sgDUX4-35
sgDUX4-39
sgDUX4-40
sgDUX4-42
sgDUX4-44
sgDUX4-50
sgDUX4-51
sgDUX4-52
sgDUX4-55
sgDUX4-57
sgDUX4-58
sgDUX4-59
sgDUX4-63
sgDUX4-67
sgDUX4-69

图3A

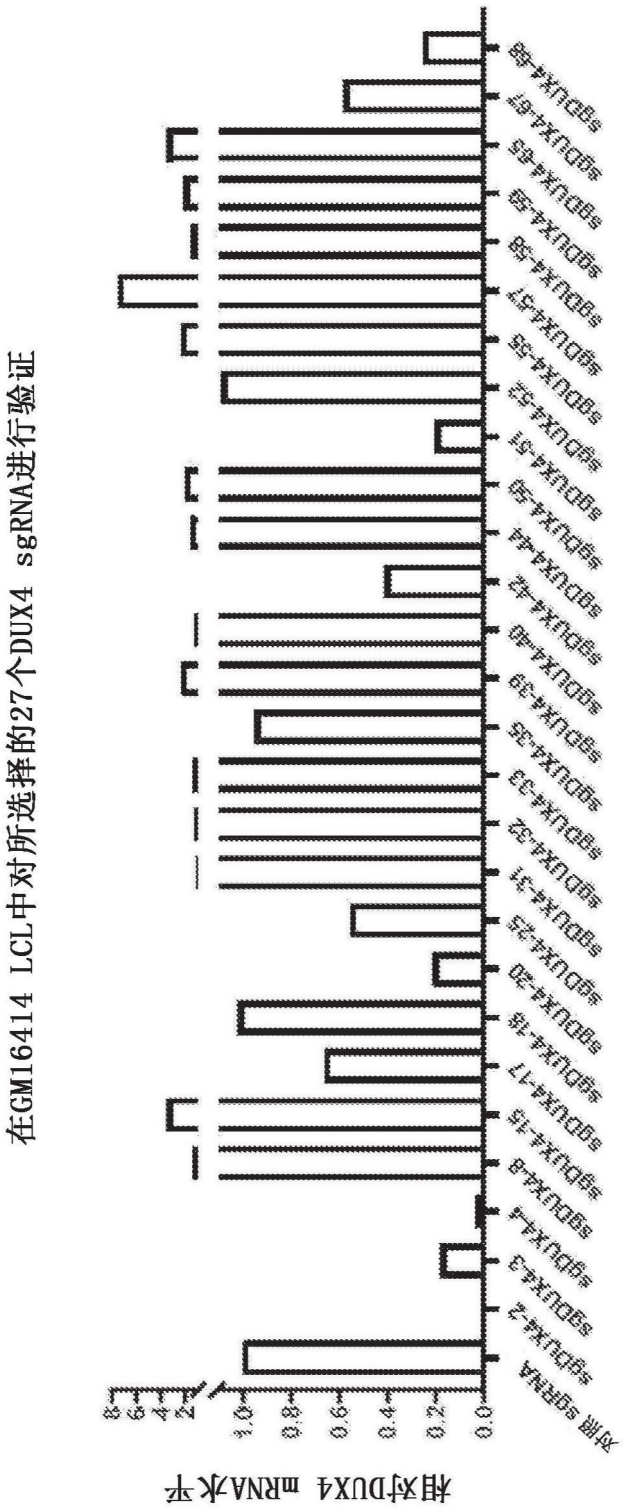

在GM16414 LCL中对所选择的27个DUX4 sgRNA进行验证
相对DUX4 mRNA水平
对照 sgRNA
sgDUX4-2
sgDUX4-3
sgDUX4-4
sgDUX4-8
sgDUX4-15
sgDUX4-17
sgDUX4-18
sgDUX4-20
sgDUX4-25
sgDUX4-31
sgDUX4-32
sgDUX4-33
sgDUX4-35
sgDUX4-39
sgDUX4-40
sgDUX4-42
sgDUX4-44
sgDUX4-50
sgDUX4-51
sgDUX4-52
sgDUX4-55
sgDUX4-57
sgDUX4-58
sgDUX4-59
sgDUX4-65
sgDUX4-67
sgDUX4-68

图3B

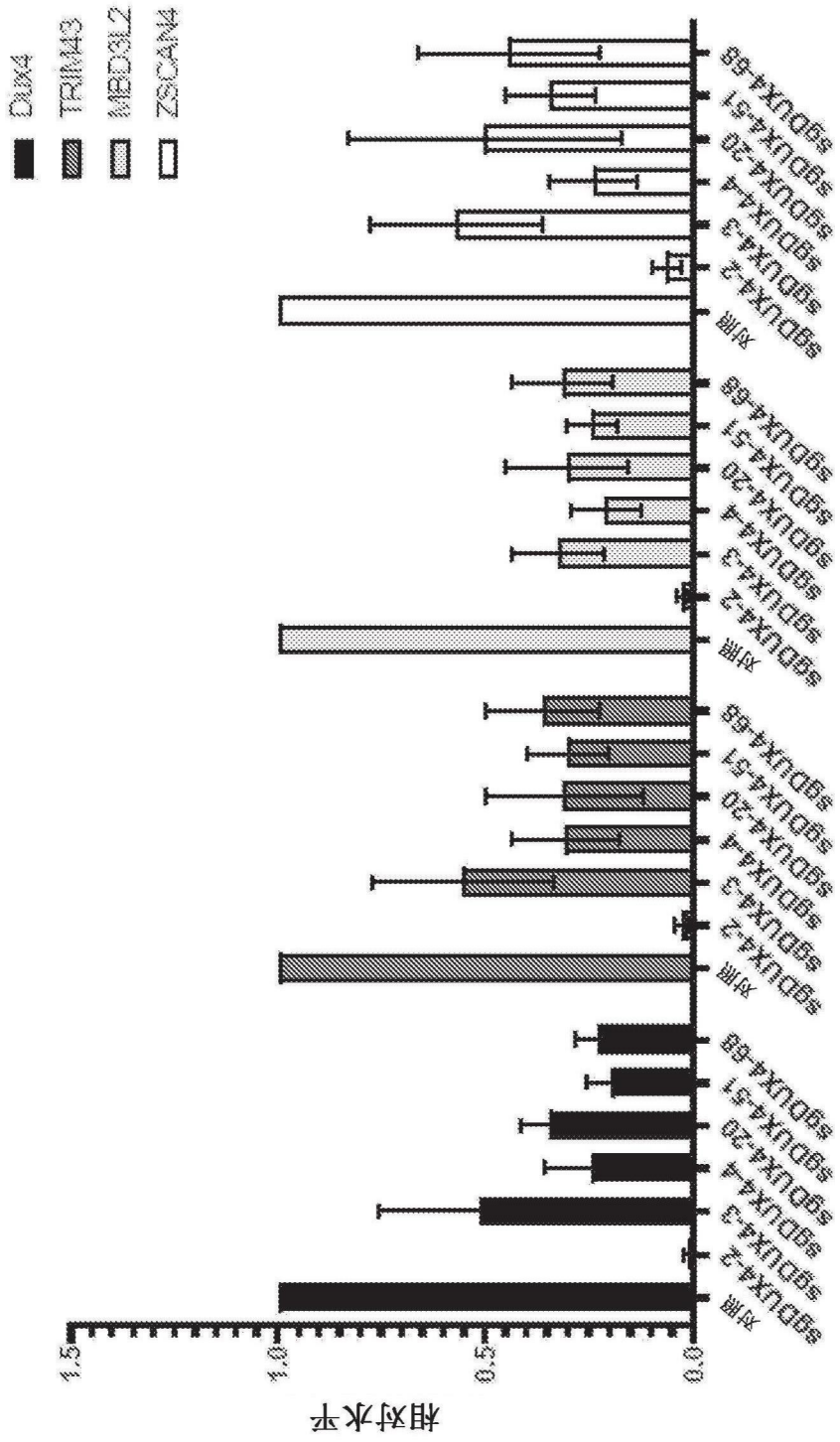

Dux4
TRIM43
MED3L2
ZSCAN4
相对水平
1.5
1.0
0.5
0.0
对照
sgDux4-2
sgDux4-3
sgDux4-4
sgDux4-20
sgDux4-51
sgDux4-68

图4

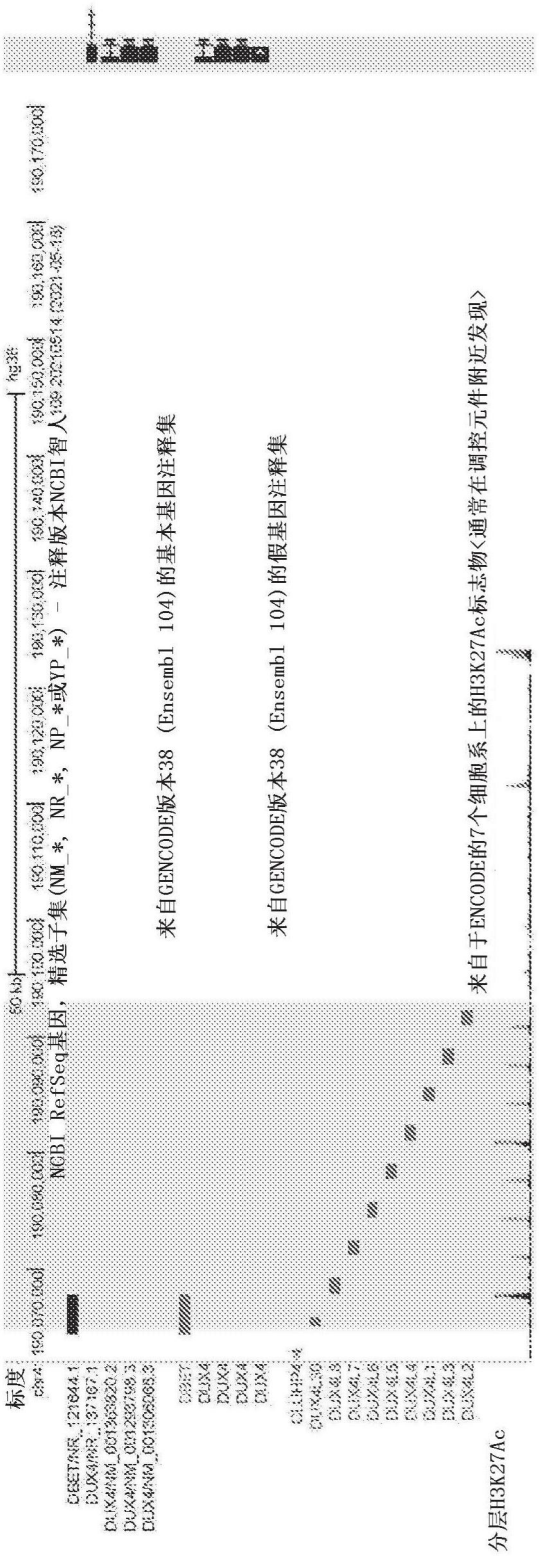

标度
chr4: 190,370,000 190,380,000 190,390,000 190,400,000 190,410,000 190,420,000 190,430,000 190,440,000 190,450,000 190,460,000 190,470,000
50 kb
hg38
NCBI RefSeq基因，精选子集(NM_*, NR_*, NP_*或YP_*) - 注释版本NCBI智人(hg38 2021-05-14 23:21:05 43)
来自GENCODE版本38 (Ensembl 104)的基本基因注释集
来自GENCODE版本38 (Ensembl 104)的假基因注释集
来自于ENCODE的7个细胞系上的H3K27Ac标志物<通常在调控元件附近发现>
D4Z4重复区
DUX4基因
分层H3K27Ac

图5

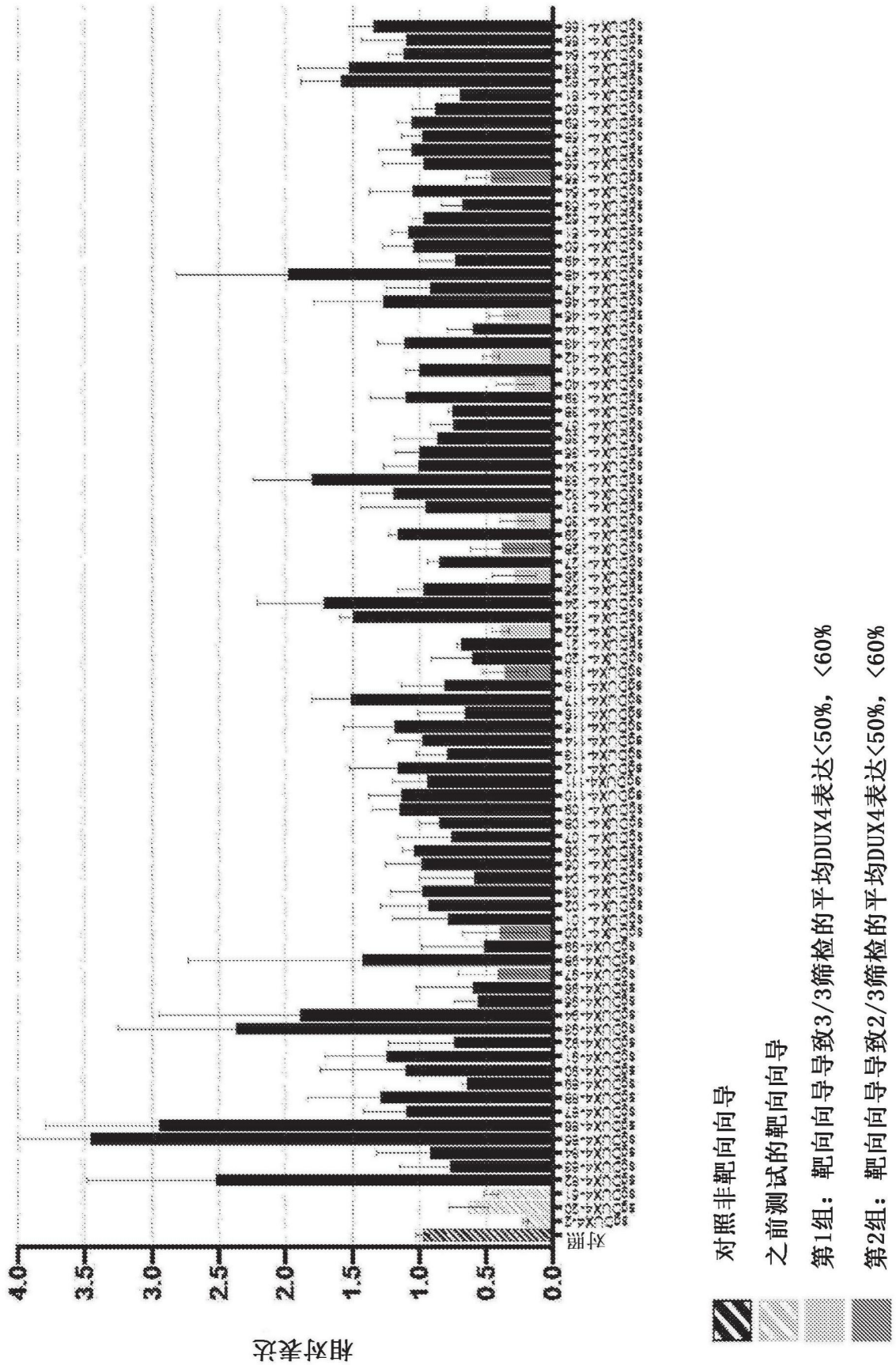

相对表达
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
对照
对照非靶向向导
之前测试的靶向向导
第1组：靶向向导导致3/3筛检的平均DUX4表达<50%，<60%
第2组：靶向向导导致2/3筛检的平均DUX4表达<50%，<60%

图6

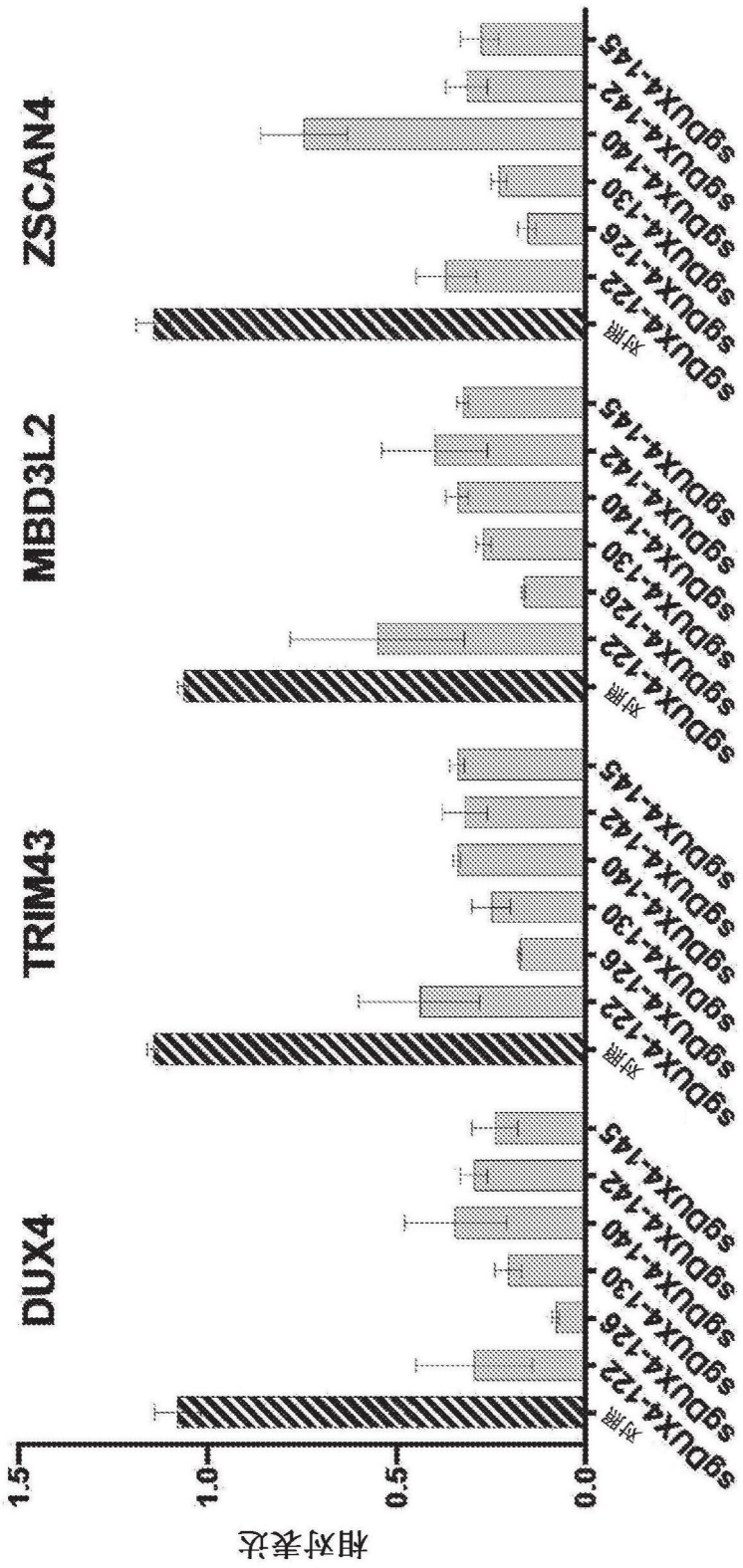

DUX4
TRIM43
MBD3L2
ZSCAN4
相对表达
1.5
1.0
0.5
0.0
对照
sgDUX4-122
sgDUX4-126
sgDUX4-130
sgDUX4-140
sgDUX4-142
sgDUX4-145

图7