(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 119836432 A
(43)申请公布日 2025.04.15

(21)申请号 202380051043.X

(22)申请日 2023.05.01

(30)优先权数据
63/337,164 2022.05.01 US
63/337,167 2022.05.01 US
63/355,083 2022.06.23 US

(85)PCT国际申请进入国家阶段日
2024.12.30

(86)PCT国际申请的申请数据
PCT/US2023/066436 2023.05.01

(87)PCT国际申请的公布数据
WO2023/215711 EN 2023.11.09

(71)申请人 恩科洛玛生物公司
地址 美国马萨诸塞州

(72)发明人 N•阿布巴克 A•弗里德兰 M•梅德 V•迈尔 F•特伦布雷 M•S•莫里森

(74)专利代理机构 北京市柳沈律师事务所 11105
专利代理师 邹宗亮

(51)Int.Cl.
C07K 14/47(2006.01)
A61K 38/00(2006.01)
A61P 3/06(2006.01)
C12N 9/10(2006.01)
C12N 9/22(2006.01)

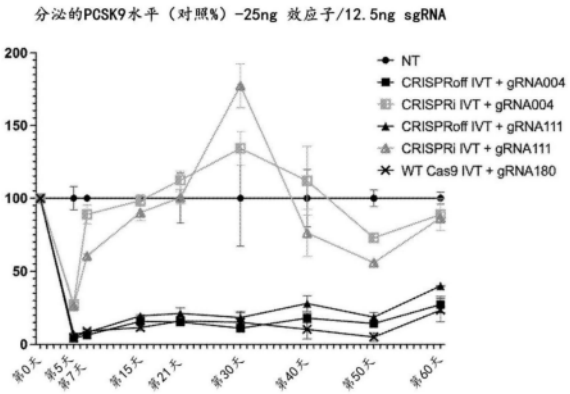
权利要求书4页 说明书188页
序列表（电子公布） 附图16页

(54)发明名称

用于PCSK9表达的表观遗传调节的组合物和方法

(57)摘要

本申请涉及包含用于PCSK9的表观遗传修饰的表观遗传编辑器的组合物和方法,以及编码该表观遗传编辑器的核酸和载体。还公开了由表观遗传编辑器表观遗传修饰的细胞。



CN 119836432 A

1.用于抑制人细胞中,任选地人肝细胞中人PCSK9基因转录的系统,所述系统包括

a)一种或多种融合蛋白,其共同包含

DNA甲基转移酶(DNMT)结构域和/或招募DNMT的结构域,任选地其中所述DNMT结构域和/或所述招募结构域包含DNMT3A结构域和/或DNMT3L结构域,并且任选地其中所招募的DNMT是DNMT3A,和

转录阻遏结构域,

每个结构域与DNA结合结构域连接,所述DNA结合结构域结合所述人PCSK9基因中的靶区域;或

b)编码所述一种或多种融合蛋白的一种或多种核酸分子。

2.根据权利要求1所述的系统,其中所述DNA结合结构域结合SEQ ID NO:1488或1489中的靶序列。

3.根据权利要求1或2所述的系统,其中所述DNA结合结构域将所述融合蛋白靶向所述PCSK9基因中选自SEQ ID NO:700-747和1036-1261的一个或多个序列。

4.根据权利要求1-3中任一项所述的系统,其中所述DNA结合结构域包括死亡CRISPR Cas(dCas)结构域、ZFP结构域或TALE结构域。

5.根据权利要求4所述的系统,其中所述DNA结合结构域包含dCas9结构域,并且所述系统还包括(i)一种或多种指导RNA,其包含SEQ ID NO:1262-1487中的任一个,或(ii)编码所述一种或多种指导RNA的核酸分子。

6.根据权利要求4或5所述的系统,其中所述dCas结构域包含dCas9序列,任选地与SEQ ID NO:12或13具有至少90%同一性的序列。

7.根据权利要求4所述的系统,其中所述ZFP结构域靶向选自SEQ ID NO:700-747的核苷酸序列。

8.根据权利要求7所述的系统,其中所述ZFP结构域包含如表1中所示的ZF001至ZF048中任一个的F1-F6氨基酸序列。

9.根据权利要求1-8中任一项所述的系统,其中所述DNMT3A结构域包含与SEQ ID NO:574或575具有至少90%同一性的序列。

10.根据权利要求1-9中任一项所述的系统,其中所述DNMT3L结构域包含与选自SEQ ID NO:578-581的序列具有至少90%同一性的序列。

11.根据权利要求1-9中任一项所述的系统,其中所述DNMT3L结构域包含与选自SEQ ID NO:582-603的序列具有至少90%同一性的序列。

12.根据权利要求1-8中任一项所述的系统,其中所述DNMT结构域包含与选自SEQ ID NO:601-603的序列具有至少90%同一性的序列。

13.根据权利要求1-12中任一项所述的系统,其中所述转录阻遏结构域包含与选自SEQ ID NO:33-570的序列具有至少90%同一性的序列。

14.根据权利要求1-12中任一项所述的系统,其中所述转录阻遏结构域包含衍生自KOX1、ZIM3、ZFP28或ZN627的KRAB结构域。

15.根据权利要求14所述的系统,其中所述KRAB结构域包含与选自SEQ ID NO:89、116、245和255的序列具有至少90%同一性的序列。

16.根据权利要求1-12中任一项所述的系统,其中所述转录阻遏结构域包含ZIM3和

KOX1 KRAB的N-和C-末端区域的融合体,并且任选地包含SEQ ID NO:571或572的氨基酸序列。

17.根据权利要求1-12中任一项所述的系统,其中所述转录阻遏结构域衍生自KAP1、MECP2、HP1a/CBX5、HP1b、CBX8、CDYL2、TOX、TOX3、TOX4、EED、EZH2、RBBP4、RCOR1或SCML2。

18.根据权利要求1-17中任一项所述的系统,其中所述系统包括

a)融合蛋白,其包含所述DNMT3A结构域、所述DNMT3L结构域、所述转录阻遏结构域和所述DNA结合结构域,

任选地,其中所述DNMT3A结构域和所述DNMT3L结构域中的一个或两个是人的,并且

任选地,其中所述DNA结合结构域是死亡CRISPR Cas结构域或ZFP结构域;或

b)编码所述融合蛋白的核酸分子。

19.根据权利要求18所述的系统,其中所述融合蛋白从N-末端到C-末端包含所述DNMT3A结构域、第一肽接头、所述DNMT3L结构域、第二肽接头、所述DNA结合结构域、第三肽接头和所述转录阻遏结构域。

20.根据权利要求19所述的系统,其中所述融合蛋白从N-末端到C-末端包含所述DNMT3A结构域、所述第一肽接头、所述DNMT3L结构域、所述第二肽接头、第一核定位信号(NLS)、所述DNA结合结构域、第二NLS、所述第三肽接头和所述转录阻遏结构域。

21.根据权利要求18所述的系统,其中所述融合蛋白从N-末端到C-末端包含第一核定位信号(NLS)、所述DNMT3A结构域、第一肽接头、所述DNMT3L结构域、第二肽接头、所述DNA结合结构域、第三肽接头、所述转录阻遏结构域和第二NLS。

22.根据权利要求18所述的系统,其中所述融合蛋白从N-末端到C-末端包含第一和第二核定位信号(NLS)、所述DNMT3A结构域、第一肽接头、所述DNMT3L结构域、第二肽接头、所述DNA结合结构域、第三肽接头、所述转录阻遏结构域以及第三和第四NLS。

23.根据权利要求18-22中任一项所述的系统,其中所述转录阻遏结构域是KRAB结构域,任选地是人KOX1、ZFP28、ZN627或ZIM3KRAB结构域。

24.根据权利要求19-23中任一项所述的系统,其中所述第二肽接头和所述第三肽接头中的一个或两个是XTEN接头,任选地选自XTEN80和XTEN16,并且进一步任选地其中所述第二肽接头是XTEN80,并且所述第三肽接头是XTEN16。

25.根据权利要求18所述的系统,其中所述融合蛋白从N-末端到C-末端包含人DNMT3A结构域、第一肽接头、人DNMT3L结构域、XTEN80肽接头、第一NLS、dSpCas9结构域、第二NLS、XTEN16肽接头和人KOX1 KRAB结构域。

26.根据权利要求25所述的系统,其中所述融合蛋白包含SEQ ID NO:658或与其至少90%相同的序列,或者SEQ ID NO:1495或与其至少90%相同的序列。

27.根据权利要求18所述的系统,其中所述融合蛋白从N末端到C末端包含人DNMT3A结构域、第一肽接头、人DNMT3L结构域、XTEN80肽接头、第一NLS、ZFP结构域、第二NLS、XTEN16接头和人KOX1 KRAB结构域。

28.根据权利要求27所述的系统,其中所述融合蛋白包含SEQ ID NO:659或与其至少90%相同的序列,或者SEQ ID NO:1496或与其至少90%相同的序列,任选地其中所述ZFP包含如表1中所示的ZF001至ZF048中任一个的F1-F6氨基酸序列。

29.根据权利要求18所述的系统,其中所述融合蛋白从N末端到C末端包含第一和第二

NLS、人DNMT3A结构域、第一肽接头、人DNMT3L结构域、XTEN80肽接头、dSpCas9结构域、XTEN16肽接头、人KOX1 KRAB结构域以及第三和第四NLS。

30.根据权利要求29所述的系统,其中所述融合蛋白包含SEQ ID NO:660或与其至少90%相同的序列。

31.根据权利要求18所述的系统,其中所述融合蛋白从N-末端到C-末端包含第一和第二NLS、人DNMT3A结构域、第一肽接头、人DNMT3L结构域、XTEN80肽接头、ZFP结构域、XTEN16肽接头、人KOX1 KRAB结构域以及第三和第四NLS,任选地其中所述融合蛋白包含SEQ ID NO:1514或与其至少90%相同的序列,或者SEQ ID NO:1523或与其至少90%相同的序列。

32.根据权利要求18所述的系统,其中所述融合蛋白从N末端到C末端包含第一和第二NLS、人DNMT3A结构域、第一肽接头、人DNMT3L结构域、XTEN80肽接头、dSpCas9结构域、XTEN16肽接头、人ZFP28 KRAB结构域以及第三和第四NLS。

33.根据权利要求32所述的系统,其中所述融合蛋白包含SEQ ID NO:661或与其至少90%相同的序列,或者SEQ ID NO:1516或与其至少90%相同的序列。

34.根据权利要求18所述的系统,其中所述融合蛋白从N末端到C末端包含第一和第二NLS、人DNMT3A结构域、第一肽接头、人DNMT3L结构域、XTEN80肽接头、ZFP结构域、XTEN16肽接头、人ZFP28 KRAB结构域以及第三和第四NLS。

35.根据权利要求18所述的系统,其中所述融合蛋白从N末端到C末端包含第一和第二NLS、人DNMT3A结构域、第一肽接头、人DNMT3L结构域、XTEN80肽接头、dSpCas9结构域、XTEN16肽接头、人ZN627 KRAB结构域以及第三和第四NLS。

36.根据权利要求35所述的系统,其中所述融合蛋白包含SEQ ID NO:662或与其至少90%相同的序列,或SEQ ID NO:1520或与其至少90%相同的序列。

37.根据权利要求18所述的系统,其中所述融合蛋白从N末端到C末端包含第一和第二NLS、人DNMT3A结构域、第一肽接头、人DNMT3L结构域、XTEN80肽接头、ZFP结构域、XTEN16肽接头、人ZN627 KRAB结构域以及第三和第四NLS。

38.根据权利要求18所述的系统,其中所述融合蛋白从N末端到C末端包含第一和第二NLS、人DNMT3A结构域、第一肽接头、人DNMT3L结构域、XTEN80肽接头、dSpCas9结构域、XTEN16肽接头、人ZIM3 KRAB结构域以及第三和第四NLS。

39.根据权利要求38所述的系统,其中所述融合蛋白包含SEQ ID NO:663或与其至少90%相同的序列,或者SEQ ID NO:1518或与其至少90%相同的序列。

40.根据权利要求18所述的系统,其中所述融合蛋白从N末端到C末端包含第一和第二NLS、人DNMT3A结构域、第一肽接头、人DNMT3L结构域、XTEN80肽接头、ZFP结构域、XTEN16肽接头、人ZIM3 KRAB结构域以及第三和第四NLS。

41.根据权利要求20-40中任一项所述的系统,其中所述NLS中的至少一个是SV40 NLS。

42.根据权利要求1-17中任一项所述的系统,其中所述系统包括:

a)第一融合蛋白,其包含第一DNA结合结构域并且包含或招募所述DNMT3A结构域,

第二融合蛋白,其包含第二DNA结合结构域并且包含或招募所述DNMT3L结构域,和

第三融合蛋白,其包含第三DNA结合结构域并且包含或招募所述转录阻遏结构域;或

b)一种或多种编码所述融合蛋白的核酸分子。

43.包含根据权利要求1-42中任一项所述的系统的人细胞,或所述细胞的后代,任选地

其中所述细胞是肝细胞。

44.药物组合物,其包含根据权利要求1-42中任一项所述的系统和药学上可接受的赋形剂,任选地其中

所述组合物包含含有所述系统的脂质纳米颗粒(LNP),和/或

所述DNA结合结构域是dCas结构域,并且所述LNP进一步包含一种或多种gRNA。

45.治疗有此需要的患者的方法,所述方法包括向所述患者施用根据权利要求1-42中任一项所述的系统或根据权利要求44所述的药物组合物,任选地静脉内施用。

46.根据权利要求45所述的方法,其中所述患者

患有心脏病,

具有升高的低密度脂蛋白胆固醇(LDL-C)或高胆固醇血症,

处于发展心肌梗塞、中风或不稳定型心绞痛的风险中,和/或

患有原发性高脂血症,任选地杂合子家族性高胆固醇血症(HeFH)或纯合子家族性高胆固醇血症(HoFH)。

47.根据权利要求1-42中任一项所述的系统或根据权利要求44所述的药物组合物,其用于治疗有此需要的患者,任选地在根据权利要求45或46所述的方法中。

48.根据权利要求1-42中任一项所述的系统在制备用于治疗有此需要的患者的药物中的用途,任选地在根据权利要求45或46所述的方法中。

用于PCSK9表达的表观遗传调节的组合物和方法

[0001] 相关申请的交叉引用

[0002] 本申请根据35U.S.C.§119(e)要求2022年5月1日提交的题为“用于PCSK9表达的表观遗传调节的组合物和方法”的美国临时申请No.63/337,164、2022年5月1日提交的题为“用于PCSK9表达的表观遗传调节的组合物和方法”的美国临时申请No.63/337,167和2022年6月23日提交的题为“用于PCSK9表达的表观遗传调节的组合物和方法”的美国临时申请No.63/355,083的权益,其各自的全部公开内容通过引用整体并入本文。

[0003] 参考电子序列表

[0004] 电子序列表(C169870034WO00-SEQ-AXW.xml;大小:1,831,551字节;和创建日期:2023年5月1日)的内容通过引用整体并入本文。

背景技术

[0005] 十多年来,基因组编辑一直被认为是一种有前途的治疗遗传疾病的治疗方法。然而,考虑到不期望的双链DNA断裂、异质性修复(包括在预期位点的大的和小的插入和缺失)以及毒性的可能性,使用传统遗传编辑器在DNA水平上进行操作仍然是有风险的。相反,靶向表观遗传修饰提供了改变基因表达而不导致双链断裂诱导的遗传毒性的潜力。

[0006] 一个有希望的表观遗传沉默候选物是前蛋白转化酶枯草杆菌蛋白酶/kexin 9型(PCSK9)基因。PCSK9是治疗心脏病的关键靶标,心脏病是全世界死亡的主要原因((Berberich et al.,Nature Rev Cardiol.(2019)16(1):9-20)。位于1号染色体上的人PCSK9基因与其食蟹猴和小鼠对应物分别具有约94%和80%的同源性。该基因在启动子区具有CpG岛,并且远离其他基因和顺式调节特征。PCSK9蛋白主要由肝脏产生。

[0007] 在人类中,PCSK9由于与LDL受体(LDLR)结合而在调节低密度脂蛋白(LDL)颗粒的循环水平中起关键作用。LDLR通过介导LDL颗粒在细胞中的内吞和降解来降低LDL颗粒的循环浓度。在不存在PCSK9的情况下,或者如果PCSK9与LDLR的相互作用被阻断,则LDLR再循环到细胞表面的速率增加,并且再循环的LDLR蛋白继续从细胞外液中去除LDL颗粒(Tombling et al.,Atherosclerosis(2021)330:52-60)。相反,当内吞的LDLR与PCSK9结合时,LDLR与其乘客LDL颗粒一起降解。临床和遗传学研究已经确定循环LDL导致动脉粥样硬化性心血管疾病(Ference et al.,Eur Heart J(2017)38:2459-72)。此外,PCSK9中的功能丧失突变与低LDL水平相关(Zhao et al.,Am J Hum Genet.(2006)79(3):514-23)。遗传或药理学降低PCSK9可减少心血管事件(Ference et al.,N Engl J Med.(2016)375(22):2144-53;Sabatine et al.,N Engl J Med.(2017)376(18):1713-22)。降低PCSK9表达有助于增加LDLR的再循环,这将导致较低的血液LDL颗粒浓度。

[0008] 鉴于PCSK9在高胆固醇血症和心血管疾病发病机制中的关键作用,需要靶向PCSK9表达的新的和改进的疗法。

[0009] 发明概述

[0010] 本公开提供了用于表观遗传修饰的系统和组合物(本文中的“表观遗传编辑器”或“表观遗传编辑系统”),以及使用其在PCSK9处产生表观遗传修饰的方法,包括在宿主细胞

和生物体中。

[0011] 在一些方面,本公开提供了用于抑制人细胞(任选地人肝细胞)中人PCSK9基因转录的系统,其包括:

[0012] a)一种或多种融合蛋白,其共同包含

[0013] DNA甲基转移酶(DNMT)结构域和/或招募DNMT的结构域,任选地其中所述DNMT结构域和/或所述招募结构域包含DNMT3A结构域和/或DNMT3L结构域,并且任选地其中所述招募的DNMT是DNMT3A,和

[0014] 转录阻遏结构域,

[0015] 每个结构域与结合人PCSK9基因中靶区域的DNA结合结构域连接;或

[0016] b)编码所述一种或多种融合蛋白的一种或多种核酸分子。

[0017] 在一些实施方案中,DNA结合结构域结合SEQ ID NO:1488或1489中的靶序列。在某些实施方案中,DNA结合结构域将融合蛋白靶向PCSK9基因中选自SEQ ID NO:700-747和1036-1261的一个或多个序列。

[0018] 在一些实施方案中,DNA结合结构域包含死亡CRISPR Cas(dCas)结构域、ZFP结构域或TALE结构域。例如,DNA结合结构域可以包含dCas9结构域,并且系统可以进一步包含(i)一种或多种指导RNA(例如,包含SEQ ID NO:1262-1487中的任一个),或(ii)编码一种或多种指导RNA的核酸分子。在某些实施方案中,dCas结构域包含dCas9序列,例如与SEQ ID NO:12或13具有至少90%同一性的序列。

[0019] 在一些实施方案中,融合蛋白包含死亡CRISPR Cas(dCas)结构域,并且系统包含本文提供的一种或多种结合PCSK9的指导RNA(gRNA)。在一些实施方案中,系统包含单个gRNA。在一些实施方案中,系统包含2个gRNA。在一些实施方案中,系统包含3个gRNA。在一些实施方案中,系统包含4个gRNA。在一些实施方案中,系统包含5个或更多个gRNA。在一些实施方案中,系统包含选自表2中提供的gRNA的gRNA。在一些实施方案中,系统包含选自表7中提供的gRNA的gRNA。在一些实施方案中,系统包含选自表8中提供的gRNA的gRNA。在一些实施方案中,系统包含选自表10中提供的gRNA的sgRNA。在一些实施方案中,系统包含选自表12中提供的gRNA的gRNA。

[0020] 在一些实施方案中,系统包含含有表10的gRNA009的gRNA靶向序列的gRNA,或结合与gRNA009相同的靶结构域序列的gRNA。在一些实施方案中,系统包含含有表10的gRNA003的gRNA靶向序列的gRNA,或结合与gRNA003相同的靶结构域序列的gRNA。在一些实施方案中,系统包含含有表10的gRNA093的gRNA靶向序列的gRNA,或结合与gRNA093相同的靶结构域序列的gRNA。在一些实施方案中,系统包含含有表10的gRNA011的gRNA靶向序列的gRNA,或结合与gRNA011相同的靶结构域序列的gRNA。在一些实施方案中,系统包含含有表10的gRNA007的gRNA靶向序列的gRNA,或结合与gRNA007相同的靶结构域序列的gRNA。在一些实施方案中,系统包含含有表10的gRNA077的gRNA靶向序列的gRNA,或结合与gRNA077相同的靶结构域序列的gRNA。在一些实施方案中,系统包含含有表10的gRNA113的gRNA靶向序列的gRNA,或结合与gRNA113相同的靶结构域序列的gRNA。在一些实施方案中,系统包含含有表10的gRNA004的gRNA靶向序列的gRNA,或结合与gRNA004相同的靶结构域序列的gRNA。在一些实施方案中,系统包含含有表10的gRNA008的gRNA靶向序列的gRNA,或结合与gRNA008相同的靶结构域序列的gRNA。在一些实施方案中,系统包含含有表10的gRNA012的gRNA靶向序

列的gRNA,或结合与gRNA012相同的靶结构域序列的gRNA。在一些实施方案中,系统包含含有表10的gRNA111的gRNA靶向序列的gRNA,或结合与gRNA111相同的靶结构域序列的gRNA。在一些实施方案中,系统包含含有表10的gRNA005的gRNA靶向序列的gRNA,或结合与gRNA005相同的靶结构域序列的gRNA。在一些实施方案中,系统包含含有表10的gRNA013的gRNA靶向序列的gRNA,或结合与gRNA013相同的靶结构域序列的gRNA。

[0021] 在一些实施方案中,系统包含融合蛋白9变体1(实施例12)和gRNA g041。在一些实施方案中,系统包含融合蛋白9变体2(实施例12)和gRNA g041。在一些实施方案中,系统包含融合蛋白10(实施例12)和gRNA g041。在一些实施方案中,系统包含融合蛋白11(实施例12)和gRNA g041。在一些实施方案中,系统包含融合蛋白12(实施例12)和gRNA g041。在一些实施方案中,系统包含融合蛋白13(实施例12)和gRNA g041。在一些实施方案中,系统包含融合蛋白14(实施例12)和gRNA g041。在一些实施方案中,系统包含融合蛋白15(实施例12)和gRNA g041。在一些实施方案中,系统包含融合蛋白9变体1(实施例12)和gRNA g049。在一些实施方案中,系统包含融合蛋白9变体2(实施例12)和gRNA g049。在一些实施方案中,系统包含融合蛋白10(实施例12)和gRNA g049。在一些实施方案中,系统包含融合蛋白11(实施例12)和gRNA g049。在一些实施方案中,系统包含融合蛋白12(实施例12)和gRNA g049。在一些实施方案中,系统包含融合蛋白13(实施例12)和gRNA g049。在一些实施方案中,系统包含融合蛋白14(实施例12)和gRNA g049。在一些实施方案中,系统包含融合蛋白15(实施例12)和gRNA g049。在一些实施方案中,系统包含融合蛋白9变体1(实施例12)以及gRNA g041和g049。在一些实施方案中,系统包含融合蛋白9变体2(实施例12)以及gRNA g041和g049。在一些实施方案中,系统包含融合蛋白10(实施例12)以及gRNA g041和g049。在一些实施方案中,系统包含融合蛋白11(实施例12)以及gRNA g041和g049。在一些实施方案中,系统包含融合蛋白12(实施例12)以及gRNA g041和g049。在一些实施方案中,系统包含融合蛋白13(实施例12)以及gRNA g041和g049。在一些实施方案中,系统包含融合蛋白14(实施例12)以及gRNA g041和g049。在一些实施方案中,系统包含融合蛋白15(实施例12)以及gRNA g041和g049。

[0022] 在一些实施方案中,DNA结合结构域包含靶向选自SEQ ID NO:700-747的核苷酸序列的ZFP结构域。在某些实施方案中,ZFP结构域依次包含如表1中所示的ZF001至ZF048中任一个的F1-F6氨基酸序列。

[0023] 在一些实施方案中,DNMT3A结构域包含与SEQ ID NO:574或575具有至少90%同一性的序列。

[0024] DNMT3L结构域可以包含例如与选自SEQ ID NO:578-581的序列具有至少90%同一性的序列。在一些实施方案中,DNMT3L结构域包含与选自SEQ ID NO:582-603的序列具有至少90%同一性的序列。在一些实施方案中,DNMT3L结构域包含与选自SEQ ID NO:601-603的序列具有至少90%同一性的序列。

[0025] 在一些实施方案中,转录阻遏结构域包含与选自SEQ ID NO:33-570的序列具有至少90%同一性的序列。在某些实施方案中,转录阻遏结构域是衍生自KOX1、ZIM3、ZFP28或ZN627的KRAB结构域。KRAB结构域可以包含例如与选自SEQ ID NO:89、116、245和255的序列具有至少90%同一性的序列。在一些实施方案中,转录阻遏结构域包含ZIM3和KOX1 KRAB的N-和C-末端区域的融合物,并且任选地包含SEQ ID NO:571或572的氨基酸序列。在某些实

施方案中,转录阻遏结构域衍生自KAP1、MECP2、HP1a/CBX5、HP1b、CBX8、CDYL2、TOX、TOX3、TOX4、EED、EZH2、RBBP4、RCOR1或SCML2。

[0026] 在一些实施方案中,所述系统包括

[0027] a)融合蛋白,其包含DNMT3A结构域、DNMT3L结构域、转录阻遏结构域和DNA结合结构域,

[0028] 任选地,其中所述DNMT3A结构域和所述DNMT3L结构域中的一个或两个是人的,并且

[0029] 任选地,其中所述DNA结合结构域是死亡CRISPR Cas结构域或ZFP结构域;或

[0030] b)编码融合蛋白的核酸分子。

[0031] 在某些实施方案中,融合蛋白从N末端到C末端包含DNMT3A结构域、第一肽接头、DNMT3L结构域、第二肽接头、DNA结合结构域、第三肽接头和转录阻遏结构域。例如,融合蛋白从N末端到C末端可以包含DNMT3A结构域、第一肽接头、DNMT3L结构域、第二肽接头、第一核定位信号(NLS)、DNA结合结构域、第二NLS、第三肽接头和转录阻遏结构域。融合蛋白从N末端到C末端可以包含第一NLS、DNMT3A结构域、第一肽接头、DNMT3L结构域、第二肽接头、DNA结合结构域、第三肽接头、转录阻遏结构域和第二NLS。融合蛋白从N末端到C末端可以包含第一和第二NLS、DNMT3A结构域、第一肽接头、DNMT3L结构域、第二肽接头、DNA结合结构域、第三肽接头、转录阻遏结构域以及第三和第四NLS。在特定的实施方案中,转录阻遏结构域是KRAB结构域,例如人KOX1、ZFP28、ZN627或ZIM3 KRAB结构域。在特定的实施方案中,第二和第三肽接头中的一个或两个是XTEN接头,其可以选自XTEN80(例如,SEQ ID NO:643)和XTEN16(例如,SEQ ID NO:638),例如,其中第二肽接头是XTEN80,并且第三肽接头是XTEN16。

[0032] 在一些实施方案中,融合蛋白从N末端到C末端可以包含人DNMT3A结构域、第一肽接头、人DNMT3L结构域、XTEN80肽接头、第一NLS、dSpCas9结构域、第二NLS、XTEN16肽接头和人KOX1 KRAB结构域。在某些实施方案中,融合蛋白包含SEQ ID NO:658或与其至少90%相同的序列。在某些实施方案中,融合蛋白包含SEQ ID NO:1495或与其至少90%相同的序列。

[0033] 在一些实施方案中,融合蛋白从N末端到C末端包含人DNMT3A结构域、第一肽接头、人DNMT3L结构域、XTEN80肽接头、第一NLS、ZFP结构域、第二NLS、XTEN16接头和人KOX1 KRAB结构域。在某些实施方案中,融合蛋白包含SEQ ID NO:659或与其至少90%相同的序列,任选地其中ZFP依次包含如表1中所示的ZF001至ZF048中任一个的F1-F6氨基酸序列。在某些实施方案中,融合蛋白包含SEQ ID NO:1496或与其至少90%相同的序列,任选地其中ZFP依次包含如表1中所示的ZF001至ZF048中任一个的F1-F6氨基酸序列。

[0034] 在一些实施方案中,融合蛋白从N末端到C末端包含第一和第二NLS、人DNMT3A结构域、第一肽接头、人DNMT3L结构域、XTEN80肽接头、dSpCas9结构域、XTEN16肽接头、人KOX1 KRAB结构域以及第三和第四NLS。在特定的实施方案中,融合蛋白可以包含SEQ ID NO:660的氨基酸序列或与其至少90%相同的序列。

[0035] 在一些实施方案中,融合蛋白从N末端到C末端包含第一和第二NLS、人DNMT3A结构域、第一肽接头、人DNMT3L结构域、XTEN80肽接头、ZFP结构域、XTEN16肽接头、人KOX1 KRAB结构域以及第三和第四NLS。

[0036] 在一些实施方案中,融合蛋白从N末端到C末端包含第一和第二NLS、人DNMT3A结构

域、第一肽接头、人DNMT3L结构域、XTEN80肽接头、dSpCas9结构域、XTEN16肽接头、人ZFP28 KRAB结构域以及第三和第四NLS。在特定的实施方案中,融合蛋白可以包含SEQ ID NO:661的氨基酸序列或与其至少90%相同的序列。

[0037] 在一些实施方案中,融合蛋白从N末端到C末端包含第一和第二NLS、人DNMT3A结构域、第一肽接头、人DNMT3L结构域、XTEN80肽接头、ZFP结构域、XTEN16肽接头、人ZFP28 KRAB结构域以及第三和第四NLS。

[0038] 在一些实施方案中,融合蛋白从N末端到C末端包含第一和第二NLS、人DNMT3A结构域、第一肽接头、人DNMT3L结构域、XTEN80肽接头、dSpCas9结构域、XTEN16肽接头、人ZN627 KRAB结构域以及第三和第四NLS。在特定的实施方案中,融合蛋白可以包含SEQ ID NO:662的氨基酸序列或与其至少90%相同的序列。

[0039] 在一些实施方案中,融合蛋白从N末端到C末端包含第一和第二NLS、人DNMT3A结构域、第一肽接头、人DNMT3L结构域、XTEN80肽接头、ZFP结构域、XTEN16肽接头、人ZN627 KRAB结构域以及第三和第四NLS。

[0040] 在一些实施方案中,融合蛋白从N末端到C末端包含第一和第二NLS、人DNMT3A结构域、第一肽接头、人DNMT3L结构域、XTEN80肽接头、dSpCas9结构域、XTEN16肽接头、人ZIM3 KRAB结构域以及第三和第四NLS。在特定的实施方案中,融合蛋白可以包含SEQ ID NO:663的氨基酸序列或与其至少90%相同的序列。

[0041] 在一些实施方案中,融合蛋白从N末端到C末端包含第一和第二NLS、人DNMT3A结构域、第一肽接头、人DNMT3L结构域、XTEN80肽接头、ZFP结构域、XTEN16肽接头、人ZIM3 KRAB结构域以及第三和第四NLS。

[0042] 在一些实施方案中,本文所述的融合蛋白中的NLS的至少一个是SV40 NLS(例如,SEQ ID NO:644)。

[0043] 在一些实施方案中,系统包括:

[0044] a)包含第一DNA结合结构域并且包含或招募DNMT3A结构域的第一融合蛋白,

[0045] 包含第二DNA结合结构域并且包含或招募DNMT3L结构域的第二融合蛋白,和

[0046] 包含第三DNA结合结构域并且包含或招募转录阻遏结构域的第三融合蛋白;或

[0047] b)一种或多种编码融合蛋白的核酸分子。

[0048] 本公开还提供了包含本文所述系统的人细胞或该细胞的后代。在一些实施方案中,所述细胞是肝细胞。

[0049] 本公开还提供了药物组合物,其包含本文所述的系统和药学上可接受的赋形剂。在一些实施方案中,组合物包含脂质纳米颗粒(LNP),其包含所述系统,和/或DNA结合结构域是dCas结构域,并且LNP还包含一种或多种gRNA。

[0050] 本公开还提供了治疗有此需要的患者的方法,其包括向患者施用本文所述的系统或药物组合物(例如,静脉内)。在一些实施方案中,患者患有心脏病;低密度脂蛋白胆固醇(LDL-C)升高或高胆固醇血症;处于发展心肌梗塞、中风或不稳定型心绞痛的风险中;和/或患有原发性高脂血症(例如,杂合子家族性高胆固醇血症(HeFH)或纯合子家族性高胆固醇血症(HoFH))。

[0051] 本公开还提供了本文所述的系统或药物组合物,其用于例如在本文所述的方法中治疗有此需要的患者。

[0052] 本公开还提供了本文所述的系统在制备用于治疗有此需要的患者的药物中的用途,例如在本文所述的方法中。

[0053] 本公开还提供了包含本文所述系统的制品和试剂盒。

[0054] 所公开的方法和组合物的其他特征、目的和优点在下面的详细描述中是显而易见的。然而,应当理解,详细描述虽然指示了所公开的方法和组合物的实施例和实施方案,但仅以说明的方式给出,而不是限制。根据详细描述,本公开范围内的各种变化和修改对于本领域技术人员将成为显而易见的。

附图说明

[0055] 图1是显示ZF蛋白和计算设计的gRNA在PCSK9基因上的预测结合位置的图。

[0056] 图2是显示用CRISPR-off(DNMT3A-3L-dCas9-KRAB)处理的细胞中第7天的相对PCSK9表达(y轴)的散点图。从gRNA靶位点到PCSK9 TSS的基因组距离显示在x轴上。

[0057] 图3是显示前40个gRNA与PCSK9基因重叠的图。

[0058] 图4A是显示在用所示gRNA处理后第7天和第28天分泌的PCSK9水平的条形图。虚线显示野生型(WT)Cas9实现的沉默。

[0059] 图4B是显示用gRNA处理后细胞中PCSK9 mRNA表达和PCSK9蛋白分泌的相关性的散点图。CRISPRi(dCas9-KRAB)代表dCas9-KRAB融合蛋白。

[0060] 图5是显示用CRISPRi(dCas9-KRAB)、CRISPR-off(DNMT3A-3L-dCas9-KRAB)和所示gRNA处理后PCSK9沉默的线图。

[0061] 图6是显示与仅用CRISPRoff系统处理的细胞相比,用CRISPRoff和辛伐他汀处理的细胞中PCSK9分泌的条形图。

[0062] 图7是显示用CRISPRoff和给定gRNA处理的Huh7肝癌细胞中PCSK9分泌减少的条形图。

[0063] 图8是显示247种靶向PCSK9的ZF蛋白的活性和毒性的散点图。相对PCSK9表达显示在x轴上,相对于pUC和脱靶对照的相应细胞计数显示在y轴上。对角线表示相对PCSK9表达与细胞计数之间的1:1相关性。

[0064] 图9是显示用ZF-off(DNMT3A-3L-ZF-KRAB)构建体处理的细胞的相对PCSK9表达(y轴)和相对于PCSK9转录起始位点(TSS)的相应靶向基因组距离(x轴)的散点图。

[0065] 图10是显示整个人PCSK9基因座侧接35.5kb和7kb的上游和下游基因组区域(67.5kb)被引入转基因小鼠并在转基因小鼠中表达的图。该转基因小鼠系在其自身(人)内源性启动子的控制下表达人PCSK9。

[0066] 图11A显示了具有变体NLS构型的融合蛋白构建体的示意图。图11B显示了具有变体KRAB结构域的另外的融合蛋白构建体的示意图。

[0067] 图12A-12B是显示在使用6.25ng RNA(图12A)或2.5ng RNA(图12B)在HeLa细胞中用具有各种NLS位置的融合蛋白构建体处理后测量的PCSK9蛋白水平的百分比的图。人和鼠DNMT3L序列分别表示为h3L和m3L。

[0068] 图13是显示具有2X NLS的构建体在沉默Hepa1-6细胞中的mPcsk9方面比CRISPR-off高效3倍的图。

[0069] 图14A是显示具有2X NLS的构建体在沉默Huh7细胞中的mPcsk9方面比CRISPR-off

更有效的图。图14B-14C是显示在第5天(图14B)和第15天(图14C)在不同剂量下,具有2X NLS的构建体在沉默Huh7细胞中的mPcsk9方面也比CRISPR-off更有效的图。

[0070] 图15是显示在Huh7细胞中,在其中dCas9被替换成锌指的CRISPR-off样型式中,2X NLS在多个ZF中提供改善的图。

[0071] 图16是显示用细菌DNMT蛋白甲基化CTLA4启动子可以诱导基因座的表观遗传沉默的一组图。

[0072] 图17是显示第30天用携带融合至dCas9的细菌DNA甲基转移酶的不同构建体处理的细胞的VIM3基因座处的甲基化谱的一组图。用M.SssI处理的样品甲基化了20%。

[0073] 图18是显示第29天通过细胞的CLTA基因座处的杂交捕获的甲基化谱比较dCas9融合物中M.SssI与鼠DNMT3A/3L的一组图。

[0074] 图19A-19D是显示当使用以CLTA-GFP作为标志物的0.5ng效应子DNA(图19A)、以GFP作为标志物的3ng效应子DNA(图19B)和以GFP作为标志物的0.5ng效应子DNA(图19C)时测试针对CRISPR-off的表观沉默活性的替代KRAB结构域的一组图。图19D显示了使用不同纳克量的效应子DNA 30天后的结果。

[0075] 发明详述

[0076] 本公开提供了用于调节PCSK9基因表达的表观遗传编辑器。通过改变PCSK9的表达,本文所述的系统、组合物和方法可用于治疗例如高胆固醇血症(例如,杂合子家族性高胆固醇血症(HeFH)、纯合子家族性高胆固醇血症(HoFH)、家族性高胆固醇血症(HF)或已确定的动脉粥样硬化性心血管疾病(ASCVD))或肾功能不全(RI)的病症。除非另有说明,“PCSK9”在本文中是指人PCSK9。人PCSK9基因序列可见于Ensembl登录号ENSG00000169174。与其他基因组工程方法相比,所述表观遗传编辑器具有几个优点,包括可逆性、易位风险降低和持久的可遗传沉默。

[0077] 在一些实施方案中,靶向用于表观遗传调节的人PCSK9基因的区域长约2kb,并且约为PCSK9 TSS的+/-1kb。在某些实施方案中,该区域具有SEQ ID NO:1488的核苷酸序列。在一些实施方案中,被靶向的PCSK9区域长约1069bp,并且约为PCSK9 TSS的+/-500bp。在某些实施方案中,被靶向的区域具有SEQ ID NO:1489的核苷酸序列。PCSK9的TSS位于基因组GRCh38的#chr1:55039548。

[0078] 在一些实施方案中,如本文所述的表观遗传编辑器可以包含一种或多种融合蛋白,其中每种融合蛋白包含DNA结合结构域,其与一个或多个用于表观遗传修饰的效应结构域连接。在某些实施方案中,其中DNA结合结构域是多核苷酸指导的DNA结合结构域,表观遗传编辑器可以进一步包含一个或多个指导多核苷酸。如本文所述的表观遗传编辑器的DNA结合结构域、效应结构域和指导多核苷酸可以以任何功能组合选自例如下文所述的那些。

[0079] 本文所述的表观遗传编辑器可以在宿主细胞中瞬时表达,或者可以整合到宿主细胞的基因组中;本公开还考虑了此类细胞及其后代。瞬时表达和整合的表观遗传编辑器或其组分都可以实现稳定的表观遗传修饰。例如,在将本文所述的表观遗传编辑器引入宿主细胞后,宿主细胞中的靶基因可以被稳定地或永久地抑制或沉默。在一些实施方案中,与不存在表观遗传编辑器时的表达水平相比,靶基因的表达降低或沉默至少1周、至少2周、至少3周、至少4周、至少5周、至少6周、至少7周、至少2个月、至少3个月、至少4个月、至少5个月、至少6个月、至少1年、至少2年或者细胞或携带细胞的受试者的整个生命期。表观遗传修饰

可以由引入表观遗传编辑器的宿主细胞的后代遗传。

[0080] 可以将本发明的表观遗传编辑器引入有此需要的患者(例如,人患者),例如引入患者的肝细胞、胆管上皮细胞(胆管细胞)、星状细胞、库普弗细胞和肝窦内皮细胞。

[0081] I.DNA结合结构域

[0082] 本文所述的表观遗传编辑器可以包含一个或多个DNA结合结构域,其将表观遗传编辑器的效应结构域引导至PCSK9基因座内或附近的靶序列。如本文所述的DNA结合结构域可以是例如多核苷酸指导的DNA结合结构域、锌指蛋白(ZFP)结构域、转录激活因子样效应子(TALE)结构域、大范围核酸酶DNA结合结构域等。DNA结合结构域的示例可以在美国专利11,162,114中找到,其通过引用整体并入本文。

[0083] 在一些实施方案中,本文所述的DNA结合结构域由其天然编码序列编码。在其他实施方案中,DNA结合结构域由已被密码子优化以在人细胞中最佳表达的核苷酸序列编码。

[0084] A.多核苷酸指导的DNA结合结构域

[0085] 在一些实施方案中,本文的DNA结合结构域可以是由指导核酸序列(例如,指导RNA序列)引导至PCSK9基因座中的靶位点的蛋白质结构域。在某些实施方案中,蛋白质结构域可以衍生自CRISPR相关核酸酶,例如I类或II类CRISPR相关核酸酶。在一些实施方案中,蛋白质结构域可以衍生自Cas核酸酶,例如II型、IIA型、IIB型、IIC型、V型或VI型Cas核酸酶。在某些实施方案中,蛋白质结构域可以衍生自II类Cas核酸酶,所述II类Cas核酸酶选自Cas1、Cas1B、Cas2、Cas3、Cas4、Cas5、Cas6、Cas7、Cas8、Cas9、Cas10、Cas14a、Cas14b、Cas14c、CasX、CasY、CasPhi、C2c4、C2c8、C2c9、C2c10、Csy1、Csy2、Csy3、Cse1、Cse2、Csc1、Csc2、Csa5、Csn2、Csm2、Csm3、Csm4、Csm5、Csm6、Cmr1、Cmr3、Cmr4、Cmr5、Cmr6、Csb1、Csb2、Csb3、Csx17、Csx14、Csx10、Csx16、CsaX、Csx3、Csx1、Csx1S、Csf1、Csf2、Cs0、Csf4及其同源物和修饰版本。“衍生自”用于表示蛋白质结构域包含亲本蛋白质的完整多肽序列,或包含其变体(例如,具有氨基酸残基缺失、插入和/或取代)。变体保留了亲本蛋白的所需功能(例如,与指导核酸序列和靶DNA形成复合物的能力)。

[0086] 在一些实施方案中,CRISPR相关蛋白结构域可以是本文所述的Cas9结构域。例如,Cas9可以指与本文所述的野生型Cas9多肽具有至少约85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%序列同一性和/或序列相似性的多肽。在一些实施方案中,所述野生型多肽是来自酿脓链球菌(Streptococcus pyogenes)的Cas9(NCBI Ref.No.NC_002737.2(SEQ ID NO:1))和/或UniProt Ref.No.Q99ZW2(SEQ ID NO:2)。在一些实施方案中,所述野生型多肽是来自金黄色葡萄球菌(Staphylococcus aureus)的Cas9(SEQ ID NO:3)。在一些实施方案中,CRISPR相关蛋白结构域是Cpf1结构域或蛋白,或与本文所述的野生型Cpf1多肽(例如,来自新杀弗朗西斯菌(Franscisella novicida)的Cpf1(UniProt Ref.No.U2UMQ6或SEQ ID NO:4))具有至少约85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%序列同一性和/或序列相似性的多肽。在某些实施方案中,CRISPR相关蛋白结构域可以是野生型蛋白的修饰形式,其包含一个或多个氨基酸残基变化,例如缺失、插入或取代;融合体或嵌合体;或其任何组合。

[0087] 已描述了各种生物体的Cas9序列和变体Cas9直系同源物的结构。可衍生本文的Cas9结构域的示例性生物体包括,但不限于,酿脓链球菌、嗜热链球菌(Streptococcus thermophilus)、链球菌属(Streptococcus sp.)、金黄色葡萄球菌、无害李斯特菌

(Listeria innocua)、格氏乳杆菌(Lactobacillus gasseri)、新杀弗朗西斯菌、产琥珀酸沃利氏菌(Wolinella succinogenes)、华德萨特菌(Sutterella wadsworthensis)、γ变形菌(Gamma proteobacterium)、脑膜炎奈瑟菌(Neisseria meningitidis)、空肠弯曲菌(Campylobacter jejuni)、多杀性巴氏杆菌(Pasteurella multocida)、产琥珀酸纤维杆菌(Fibrobacter succinogene)、深红红螺菌(Rhodospirillum rubrum)、达松维尔拟诺卡菌(Nocardiopsis dassonvillei)、始旋链霉菌(Streptomyces pristinaespiralis)、绿色产色链霉菌(Streptomyces viridochromogenes)、绿色产色链霉菌(Streptomyces viridochromogenes)、玫瑰链孢囊菌(Streptosporangium roseum)、酸热脂环酸芽孢杆菌(Alicyclobacillus acidocaldarius)、假蕈状芽孢杆菌(Bacillus pseudomycoides)、砷还原芽孢杆菌(Bacillus selenitireducens)、西伯利亚微小杆菌(Exiguobacterium sibiricum)、德氏乳杆菌(Lactobacillus delbrueckii)、唾液乳杆菌(Lactobacillus salivarius)、布氏乳杆菌(Lactobacillus buchneri)、齿垢密螺旋体(Treponema denticola)、海洋微颤菌(Microscilla marina)、伯克霍尔德氏菌(Burkholderiales bacterium)、食萘极地单胞菌(Polaromonas naphthalenivorans)、极化单胞菌属(Polaromonas sp.)、瓦氏鳄球藻(Crocosphaera watsonii)、蓝藻属(Cyanothece sp.)、铜绿微囊藻(Microcystis aeruginosa)、聚球藻属(Synechococcus sp.)、阿拉伯糖醋杆菌(Acetohalobium arabaticum)、Ammonifex degensii、热解纤维素菌(Caldicelulosiruptor becscii)、脱硫假丝酵母(Candidatus Desulforudis)、肉毒梭菌(Clostridium botulinum)、艰难梭菌(Clostridium difficile)、大芬戈尔德菌(Finegoldia magna)、嗜热盐碱厌氧菌(Natranaerobius thermophilus)、嗜热丙酸厌氧肠状菌(Pelotomaculum thermopropionium)、嗜酸性喜温硫杆菌(Acidithiobacillus caldus)、嗜酸性氧化亚铁硫杆菌(Acidithiobacillus ferrooxidans)、Allochromatium vinosum、海杆菌属(Marinobacter sp.)、亚硝化嗜盐球菌(Nitrosococcus halophilus)、瓦氏亚硝化球菌(Nitrosococcus watsoni)、浮游假交替单胞菌(Pseudoalteromonas haloplanktis)、Ktedonobacter racemifer、Methanohalobium evestigatum、多变鱼腥藻(Anabaena variabilis)、泡沫节球藻(Nodularia spumigena)、念珠藻属(Nostoc sp.)、极大节旋藻(Arthrospira maxima)、钝顶节旋藻(Arthrospira platensis)、节旋藻属(Arthrospira sp.)、螺旋藻属(Lyngbya sp.)、原型微鞘藻(Microcoleus chthonoplastes)、颤藻属(Oscillatoria sp.)、运动石袍菌(Petrotoga mobilis)、非洲栖热腔菌(Thermosipho africanus)、巴氏链球菌(Streptococcus pasteurianus)、灰色奈瑟球菌(Neisseria cinerea)、红嘴鸥弯曲杆菌(Campylobacter lari)、食清洁剂细小棒菌(Parvibaculum lavamentivorans)、白喉棒状杆菌(Corynebacterium diphtheria)和Acaryochloris marina。Cas9序列还包括来自Chylinski et al.,RNA Biol.(2013)10(5):726-37中公开的生物体和基因座的序列。

[0088] 在一些实施方案中,Cas9结构域来自酿脓链球菌(SpCas9)。在一些实施方案中,Cas9结构域来自金黄色葡萄球菌(SaCas9)。

[0089] 其他Cas结构域也被考虑用于本文的表观遗传编辑器中。这些包括例如来自CasX(Cas12E)(例如SEQ ID NO:5)、CasY(Cas12d)(例如SEQ ID NO:6)、Casφ(CasPhi)(例如SEQ ID NO:7)、Cas12f1(Cas14a)(例如SEQ ID NO:8)、Cas12f2(Cas14b)(例如SEQ ID NO:9)、

Cas12f3(Cas14c)(例如SEQ ID NO:10)和C2c8(例如SEQ ID NO:11)的Cas结构域。

[0090] 对于表观遗传编辑,核酸酶衍生的蛋白质结构域(例如,Cas9或Cpf1结构域)可以通过突变而具有降低的核酸酶活性或无核酸酶活性,使得蛋白质结构域不切割DNA或具有降低的DNA切割活性,同时保留与指导核酸序列(例如,指导RNA)和靶DNA复合的能力。例如,与野生型结构域相比,核酸酶活性可以降低至少30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%或99%。在一些实施方案中,本文所述的CRISPR相关蛋白结构域是无催化活性的("死亡")。此类结构域的示例包括例如dCas9("死亡"Cas9)、dCpf1、ddCpf1、dCasPhi、ddCas12a、dLbCpf1和dFnCpf1。例如,与野生型Cas9相比,dCas9蛋白结构域可以包含消除其核酸酶活性的一个、两个或更多个突变。已知Cas9的DNA切割结构域包括两个亚结构域:HNH核酸酶亚结构域和RuvC1亚结构域。HNH亚结构域切割与gRNA互补的链,而RuvC1亚结构域切割非互补链。这些亚结构域内的突变可以沉默Cas9的核酸酶活性。例如,突变D10A(在RuvC1中)和H840A(在HNH中)使SpCas9的核酸酶活性完全失活。类似地,SaCas9可以通过突变D10A和N580A失活。在一些实施方案中,dCas9在HNH亚结构域和/或RuvC1亚结构域中包含降低或消除核酸酶活性的至少一个突变。在一些实施方案中,dCas9仅包含RuvC1亚结构域或仅包含HNH亚结构域。应当理解,使RuvC1和/或HNH结构域失活的任何突变可以包括在本文的dCas9中,例如,RuvC1结构域和/或HNH结构域中的插入、缺失或单个或多个氨基酸取代。

[0091] 在一些实施方案中,本文的dCas9蛋白包含在对应于野生型SpCas9序列的位置D10(例如,D10A)、H840(例如,H840A)或两者的位置处的突变,如在UniProt登录号Q99ZW2(SEQ ID NO:2)提供的序列中编号的。在特定的实施方案中,dCas9包含dSpCas9(D10A和H840A)的氨基酸序列(SEQ ID NO:12)。

[0092] 在一些实施方案中,如本文所述的dCas9蛋白包含在对应于野生型SaCas9序列(例如,SEQ ID NO:3)的位置D10(例如,D10A)、N580(例如,N580A)或两者的位置处的突变。在特定的实施方案中,dCas9包含dSaCas9(D10A和N580A)的氨基酸序列(SEQ ID NO:13)。

[0093] 基于本公开和本领域知识,使Cas9失活的额外的合适突变对于本领域技术人员将是显而易见的,并且在本公开的范围内。此类突变可包括但不限于SpCas9中的D839A、N863A和/或K603R。本公开考虑降低或消除本文所述的任何Cas9的核酸酶活性的任何突变(例如,对应于本文所述的任何Cas9突变的突变)。

[0094] 与野生型Cpf1相比,dCpf1蛋白结构域可以包含降低或消除其核酸酶活性的一个、两个或更多个突变。Cpf1蛋白具有与Cas9的RuvC结构域相似的RuvC样核酸内切酶结构域,但不具有HNH核酸内切酶结构域,并且Cpf1的N末端不具有Cas9的α-螺旋识别叶。在一些实施方案中,dCpf1包含对应于如在新杀弗朗西斯菌Cpf1蛋白(FnCpf1;SEQ ID NO:4)的序列中编号的位置D917A、E1006A或D1255A的一个或多个突变。在某些实施方案中,dCpf1蛋白包含对应于D917A、E1006A、D1255A、D917A/E1006A、D917A/D1255A、E1006A/D1255A或D917A/E1006A/D1255A的突变,或本文所述的任何Cpf1氨基酸序列中的相应突变。在一些实施方案中,dCpf1包含D917A突变。在特定的实施方案中,dCpf1包含dFnCpf1的氨基酸序列(SEQ ID NO:14)。

[0095] 本文考虑的另外的核酸酶失活的CRISPR相关蛋白结构域包括来自例如dNmeCas9(例如,SEQ ID NO:15)、dCjCas9(例如,SEQ ID NO:16)、dSt1Cas9(例如,SEQ ID NO:17)、

dSt3Cas9(例如,SEQ ID NO:18)、dLbCpf1(例如,SEQ ID NO:19)、dAsCpf1(例如,SEQ ID NO:20)、denAsCpf1(例如,SEQ ID NO:21)、dHFAsCpf1(例如,SEQ ID NO:22)、dRVRAsCpf1(例如,SEQ ID NO:23)、dRRAsCpf1(例如,SEQ ID NO:24)、dCasX(例如,SEQ ID NO:25)和dCasPhi(例如,SEQ ID NO:26)的核酸酶失活的CRISPR相关蛋白结构域。

[0096] 在一些实施方案中,本文所述的Cas9结构域可以是高保真Cas9结构域,例如,包含降低Cas9结构域与DNA的糖-磷酸主链之间的静电相互作用以赋予增加的靶结合特异性的一个或多个突变。在某些实施方案中,所述高保真Cas9结构域可以是如本文所述的核酸酶失活的。

[0097] 本文所述的CRISPR相关蛋白结构域可以识别靶基因中的原间隔区相邻基序(PAM)序列。“PAM”序列通常是紧接在CRISPR相关蛋白结构域靶向的序列之后的2至6bp DNA序列。PAM序列是CRISPR蛋白结合和切割所需要的,但不是靶序列的一部分。CRISPR相关蛋白结构域可以识别天然存在的或典型的PAM序列,或者可以具有改变的PAM特异性。结合非经典PAM序列的CRISPR相关蛋白结构域已在本领域中描述。例如,结合非典型PAM序列的Cas9结构域已在Kleinstiver et al.,Nature(2015)523(7561):481-5和Kleinstiver et al.,Nat Biotechnol.(2015)33:1293-8中描述.此类Cas9结构域可以包括例如来自“VRER”SpCas9、“EQR”SpCas9、“VQR”SpCas9、“SpG Cas9”、“SpRYCas9”和“KKH”SaCas9的Cas9结构域。还考虑了这些Cas9结构域的核酸酶失活形式,例如核酸酶失活的VRER SpCas9(例如,SEQ ID NO:27)、核酸酶失活的EQR SpCas9(例如,SEQ ID NO:28)、核酸酶失活的VQR SpCas9(例如,SEQ ID NO:29)、核酸酶失活的SpG Cas9(例如,SEQ ID NO:30)、核酸酶失活的SpRY Cas9(例如,SEQ ID NO:31)和核酸酶失活的KKH SaCas9(例如,SEQ ID NO:32)。另一个示例是经工程化以识别5'-YG-3'(其中“Y”是嘧啶)的新杀弗朗西斯菌的Cas9。

[0098] 基于本公开内容,另外合适的CRISPR相关蛋白、直系同源物和变体,包括核酸酶失活变体和序列,对于本领域技术人员将是显而易见的。

[0099] 可以与本文的CRISPR相关蛋白结构域结合使用的指导RNA在下文第II节中进一步描述。

[0100] B.锌指蛋白结构域

[0101] 在一些实施方案中,本文所述的表观遗传编辑器的DNA结合结构域包含锌指蛋白(ZFP)结构域(或如本文所用的“ZF结构域”)。ZFP是具有至少一个锌指的蛋白质,并且以序列特异性方式与DNA结合。“锌指”(ZF)或“锌指基序”(ZF基序)是指包含由锌离子稳定的β-β-α(ββα)-蛋白折叠的多肽结构域。ZF结合2至4个碱基对的核苷酸,通常为3或4个碱基对(连续或非连续)。每个ZF通常包含大约30个氨基酸。ZFP结构域可以含有与其靶核酸序列串联接触的多个ZF。可以工程化ZF的串联阵列以产生结合所需核酸靶标的人工ZFP。可以通过使用包含三联体(或四联体)核苷酸序列和单个ZF氨基酸序列的数据库来合理设计ZFP,其中每个三联体或四联体核苷酸序列与结合特定三联体或四联体序列的ZF的一个或多个氨基酸序列相关。参见例如美国专利6,453,242、6,534,261和8,772,453。

[0102] ZFP广泛存在于真核细胞中,并且可以属于例如C2H2类、CCHC类、PHD类或RING类。表征这些蛋白质的一类(C2H2类)的示例性基序是-Cys-$(X)_{2-4}$-Cys-$(X)_{12}$-His-$(X)_{3-5}$-His-(SEQ ID NO:657),其中X是任何独立选择的氨基酸。在一些实施方案中,本文的ZFP结构域可包含ZF阵列,其包含各自接触三个或更多个连续核苷酸的连续C2H2-ZF。

[0103] 本文所述的表观遗传编辑器的ZFP结构域可以包括2、3、4、5、6、7、8、9、10、11、12或更多个ZF。ZFP结构域可以包括双指或三指单元的阵列,例如3、4、5、6、7、8、9或10个或更多个单元,其中每个单元结合靶序列中的亚位点。在一些实施方案中,包含至少三个ZF的ZFP结构域识别9或10个核苷酸的靶DNA序列。在一些实施方案中,包含至少四个ZF的ZFP结构域识别12至14个核苷酸的靶DNA序列。在一些实施方案中,包含至少六个ZF的ZFP结构域识别18至21个核苷酸的靶DNA序列。

[0104] 在一些实施方案中,本文所述的ZFP结构域中的ZF经由肽接头连接。肽接头的长度可以是例如2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20或更多个氨基酸。在一些实施方案中,接头包含5个或更多个氨基酸。在一些实施方案中,接头包含7-17个氨基酸。接头可以是柔性的或刚性的。

[0105] 在一些实施方案中,锌指阵列可以具有以下序列:

[0106] SRPGERPFQCRICMRNFSXXXXXXXHXXTHTGEKPFQCRICMRNFSXXXXXXXHXXTH[接头]FQCRICMRNFSXXXXXXXHXXTHTGEKPFQCRICMRNFSXXXXXXXHXXTH[接头]PFQCRICMRNFSXXXXXXXHXXTHTGEKPFQCRICMRNFSXXXXXXXHXXTHLRGS(SEQ ID NO:650),或与其至少70%、75%、80%、85%、90%、95%、96%、97%、98%或99%相同的序列,其中"XXXXXXX"代表ZF识别螺旋的氨基酸,其赋予锌指DNA结合特异性;每个X可以独立地选择。在上述序列中,斜体的"XX"可以是TR、LR或LK,并且"[接头]"表示接头序列。在一些实施方案中,接头序列是TGSQKP(SEQ ID NO:651);当ZF靶向的亚位点相邻时,可以使用该接头。在一些实施方案中,接头序列是TGGGGSQKP(SEQ ID NO:652);当锌指靶向的亚位点之间存在碱基时,可以使用该接头。所示的两个接头可以相同或不同。

[0107] 本文中的ZFP结构域可以包含两个或更多个相邻ZF的阵列,所述两个或更多个相邻ZF彼此直接相邻(例如,被短(典型的)接头序列分开),或被更长、柔性或结构化多肽序列分开。在一些实施方案中,直接相邻的指与连续的核酸序列结合,即与相邻的三核苷酸/三联体结合。在一些实施方案中,相邻的指在彼此各自的靶三联体之间交叉结合,这可有助于加强或增强靶序列的识别,并导致重叠序列的结合。在一些实施方案中,ZFP结构域内的远距离ZF可以识别(或结合)非连续核苷酸序列。

[0108] 本文示例性ZFP结构域的ZF DNA识别螺旋的氨基酸序列及其PCSK9靶序列如下表1所示,其中括号内的数字表示SEQ ID NO。

[0109] 表1.靶向PCSK9的示例性ZFP结构域的ZF序列

[0110]

ZF No.	DNA 靶标	F1	F2	F3	F4	F5	F6
ZF001	CCGGCAGAACT TGGAGTC (700)	DHSSLKR (748)	QSPHLQR (749)	YKHVLVN (750)	QMSNLDR (751)	QAETLKR (752)	RNYDLGI (753)
ZF002	GGTGGGGAGGA CTGTGCA (701)	QSTTLKR (754)	QAHGLTA (755)	DAANLRR (756)	RRDNLLR (757)	RNTHLAR (758)	EAHHLSR (759)
ZF003	GGTGGGGAGGA CTGTGCA (702)	QSTTLKR (760)	GAHGLAG (761)	DQTNLRR (762)	RQDNLQR (763)	KGDHLRR (764)	EAHHLSR (765)
ZF004	GATGGGGCTCT GGTGGCG (703)	RTDTLAR (766)	RKTALNR (767)	RNESLKV (768)	VKNTLTR (769)	RREHLVR (770)	ISHNLAR (771)
ZF005	GATGGGGCTCT GGTGGCG (704)	RRLTLDR (772)	RRDLLHN (773)	RNESLKV (774)	VGASLKR (775)	RREHLVR (776)	LEHNLGR (777)
ZF006	GATGGGGCTCT GGTGGCG (705)	RNHTLQR (778)	RREVLEN (779)	RHESLIV (780)	VGASLKR (781)	KKDHLHR (782)	LMHNLTR (783)
ZF007	GCAGGAGGATG AAAAGGG (706)	KRERLER (784)	QKGNLVV (785)	QKTHLAV (786)	QGGHLKR (787)	QRPHLTN (788)	QAATLQR (789)
ZF008	GGAGCCGCTGC TGCAACG (707)	RSQTLAQ (790)	QSTTLKR (791)	LRDSLKR (792)	LGHTLNR (793)	DPSVLTR (794)	QNSHLRR (795)
ZF009	GACTGTGCAGG AGCTGAA (708)	QRGNLAR (796)	VSNTLTR (797)	QRPHLTN (798)	QAATLQR (799)	RRQHLEL (800)	DRGNLTR (801)
ZF010	GGGGAGGACTG TGCAGGA (709)	QSAHLKR (802)	QDVSLVR (803)	RKQHLQL (804)	DRGNLTR (805)	RDDNLQR (806)	RMEHLPR (807)
ZF011	GGTGGGGAGGA CTGTGCA (710)	QSTTLKR (808)	QPHGLGA (809)	LVENLRR (810)	RVENLHR (811)	RREHLVR (812)	LSQGLAR (813)
ZF012	GGTGGGGAGGA CTGTGCA (711)	QSTTLKR (814)	QAHGLTA (815)	DGSNLAR (816)	RRDNLLR (817)	RREHLVR (818)	VDHHLRR (819)
ZF013	TGCGGTGGGGA GGACTGT (712)	RKQHLQL (820)	DRGNLTR (821)	RPDNLVR (822)	RIDKLGG (823)	EAHHLSR (824)	QGRSLRA (825)
ZF014	GTGGAAGGTGG CTGTGGT (713)	EAHHLSR (826)	QPHGLRA (827)	LKEHLTR (828)	MGHHLKR (829)	QRGNLLR (830)	RPDALPR (831)
ZF015	GCGGTGGAAGG TGGCTGT (714)	RRQHLQY (832)	LKEHLTR (833)	SKQKLQV (834)	QQTNLTR (835)	RREVLEN (836)	RPDGLAR (837)
ZF016	GCGGTGGAAGG TGGCTGT (715)	RRQHLTL (838)	DNSHLQR (839)	LPHHLQR (840)	QSNNLTR (841)	RNFILQR (842)	RKDDLKR (843)
ZF017	GCGGCGGGGGA GGACGGG (716)	RQEHLVR (844)	EGGNLMR (845)	RPDNLVR (846)	RIDKLGG (847)	RRDDLTR (848)	RKDLLHR (849)
ZF018	GCGGCGGGGGA GGACGGG (717)	RREHLVR (850)	DMGNLGR (851)	RDDNLQR (852)	RMEHLPR (853)	RKEDLAR (854)	RLDMLAR (855)
ZF019	TAGGGATGGGA GGCCGGG (718)	RREHLVR (856)	EHRGLKR (857)	RGDNLGR (858)	RSDHLSL (859)	QQAHLVR (860)	RDDNLRT (861)

[0111]

ZF No.	DNA 靶标	F1	F2	F3	F4	F5	F6
ZF020	GGCGCCGCCGG CGTGGAC (719)	DDANLRR (862)	RREVLEN (863)	LKEHLTR (864)	DPSNLRR (865)	DSSVLRR (866)	ENSKLNR (867)
ZF021	GGCGCCGCCGG CGTGGAC (720)	DRANLRR (868)	RPDALSR (869)	LKEHLTR (870)	DPSNLRR (871)	DSSVLRR (872)	ENSKLNR (873)
ZF022	GTCGTTGCAGC AGCGGCT (721)	LSQTLKR (874)	RTDTLAR (875)	QMETLKR (876)	QGGTLRR (877)	HKSSLTR (878)	DRTPLQR (879)
ZF023	GATGGGGCTCT GGTGGCG (722)	RTDTLAR (880)	RKTALNR (881)	RNESLKV (882)	VKNTLTR (883)	RREHLVR (884)	LTHNLRR (885)
ZF024	TCGGTGGGCAG CGAGGAG (723)	RLDNLDR (886)	RADNLRR (887)	SHWKLHT (888)	EKGHLNR (889)	RREVLEN (890)	RADGLQL (891)
ZF025	GGGGAGGACTG TGCAGGA (724)	QSPHLKR (892)	QSTSLQR (893)	RKQHLTL (894)	DRGNLTR (895)	RPHNLLR (896)	RREHLVR (897)
ZF026	GGGGAGGACTG TGCAGGA (725)	QQAHLVR (898)	QAETLKR (899)	RKQHLTL (900)	DRGNLTR (901)	RNTNLTR (902)	RREHLVR (903)
ZF027	GCGGCGGGGGA GGACGGG (726)	RREHLVR (904)	DPSNLQR (905)	RNTNLTR (906)	RREHLVR (907)	RTDTLAR (908)	RVDDLGR (909)
ZF028	GAGGAGTGAGC CAGGCAGT (727)	QSHSLKS (910)	ESGHLKR (911)	DLSTLRR (912)	QNEHLKV (913)	RPENLNR (914)	RRDNLNR (915)
ZF029	GGTGTGGGTGC TTGACGCC (728)	ERRGLDR (916)	DRGNLTR (917)	LSQTLNR (918)	IKHHLGR (919)	RNFILQR (920)	ERHQLVR (921)
ZF030	TCTGAGCCTGG AGGAGTGA (729)	QREHLVT (922)	RIDNLGR (923)	RMSNLVR (924)	RNESLKV (925)	RADNLGR (926)	ARNTLKG (927)
ZF031	GAGGAGTGAGC CAGGCAGT (730)	QSHSLKS (928)	EKSHLTR (929)	DSPTLRR (930)	QKVHLQV (931)	RAENLAR (932)	RRDNLLR (933)
ZF032	GAGGGCCAGGG GAGAGGTT (731)	HKSSLTR (934)	RPDNLPR (935)	QGTHLRN (936)	RTHHLIT (937)	TPSKLDR (938)	RQDNLGR (939)
ZF033	GCAGGAGGACG AGGACGGC (732)	ERAKLIR (940)	DPSNLRR (941)	RQDNLGR (942)	DQGNLGR (943)	QSAHLKR (944)	QDVSLVR (945)
ZF034	GCAGGAGGACG AGGACGGC (733)	TPSKLDR (946)	LAENLRR (947)	RQDNLGR (948)	DGGNLGR (949)	QSPHLKR (950)	QSTSLQR (951)
ZF035	GCAGGAGGACG AGGACGGC (734)	TPSKLDR (952)	DSSNLRR (953)	RQDNLGR (954)	EGGNLMR (955)	QQAHLVR (956)	QAETLKR (957)
ZF036	TGGGGAGGACT GTGCAGGA (735)	QRPHLTN (958)	QAATLQR (959)	RKECLVV (960)	QNPHLLR (961)	QSAHLKR (962)	RSDHLSL (963)
ZF037	GTTGCAGGCGG GCGCCGCC (736)	DPSVLTR (964)	EHRGLKR (965)	VPSKLKR (966)	RDDTLVR (967)	QKETLNR (968)	VRSSLRR (969)
ZF038	GTTGCAGGCGG GCGCCGCC (737)	DPSVLTR (970)	EHRGLKR (971)	KTDHLAR (972)	DKAHLVR (973)	QKETLNR (974)	VRSSLRR (975)
ZF039	GCTGTTTGGGG AGGGCGAG (738)	RADNLGR (976)	KQVTLRN (977)	KHSNLTR (978)	RREHLVR (979)	VKSSLTR (980)	VSNSLNR (981)

[0112]

ZF No.	DNA 靶标	F1	F2	F3	F4	F5	F6
ZF040	GCTGTTTGGGG AGGGCGAG (739)	RDDNLQR (982)	KNVTLTN (983)	QSAHLKR (984)	RSDHLSL (985)	VKSSLTR (986)	VSNSLNR (987)
ZF041	GCAGAGGCCGG AGGGGGTC (740)	DPSVLKR (988)	RTEHLAR (989)	QSPHLKR (990)	DQTTLRR (991)	KHSNLTR (992)	QMETLKR (993)
ZF042	TGGGGAGGGCG AGGCCGAA (741)	QNQNLAR (994)	DKSVLAR (995)	RDDNLQR (996)	KNVTLTN (997)	QSAHLKR (998)	RSDHLSL (999)
ZF043	GCGGGTGTAGG GATGGGAG (742)	KHSNLTR (1000)	RREHLTI (1001)	KKDHLHR (1002)	QTTTLKR (1003)	EEHHLTR (1004)	REDVLGR (1005)
ZF044	GCGGGTGTAGG GATGGGAG (743)	KHSNLTR (1006)	RREHLTI (1007)	KKDHLHR (1008)	QTTTLKR (1009)	EAHHLSR (1010)	RTDDLGR (1011)
ZF045	GGAGCTGGGAG CCGCTGCT (744)	LRQTLAR (1012)	VAHSLKR (1013)	DRSVLVR (1014)	QNSHLRR (1015)	VKHSLQR (1016)	QTTHLSR (1017)
ZF046	GCAAGGCGGCG GGGGAGGA (745)	QSAHLKR (1018)	QMSHLKR (1019)	RGNHLRR (1020)	LKEHLTR (1021)	RNEHLKV (1022)	QSTTLKR (1023)
ZF047	GCTGTTTGGGG AGGGCGAGG (746)	RKPHLDN (1024)	RPDVLMR (1025)	RDDNLQR (1026)	RMEHLPR (1027)	HQSSLTR (1028)	VSNSLAR (1029)
ZF048	GCTGTTTGGGG AGGGCGAGG (747)	RNIHLQT (1030)	RKDTLAR (1031)	RADNLGR (1032)	RMEHLPR (1033)	HQSSLTR (1034)	VSNSLAR (1035)

[0113] 在一些实施方案中,表观遗传编辑器的ZFP结构域结合选自SEQ ID NO:700-747中任一个的靶序列。在进一步的实施方案中,ZFP结构域依次包含如表1中所示的ZF001-ZF048中任一个的F1-F6氨基酸序列。F1-F6氨基酸序列可置于SEQ ID NO:650的ZF框架序列内,或置于本领域已知的任何其它ZF框架内。

[0114] C.TALE

[0115] 在一些实施方案中,本文所述的表观遗传编辑器的DNA结合结构域包含转录激活因子样效应子(TALE)结构域。TALE的DNA结合结构域包含约33-34个氨基酸的高度保守序列,在12和13位具有对识别特定核苷酸至关重要的重复可变二残基(RVD)。TALE可以被工程化以结合几乎任何期望的DNA序列。用于编程TALE的方法是本领域已知的。例如,此类方法描述于Carroll et al.,Genet Soc Amer.(2011)188(4):773-82;Miller et al.,Nat Biotechnol.(2007)25(7):778-85;Christian et al.,Genetics(2008)186(2):757-61;Li et al.,Nucl Acids Res.(2010)39(1):359-72;和Moscou et al.,Science(2009)326(5959):1501。

[0116] D.其他DNA结合结构域

[0117] 本文所述的表观遗传编辑器考虑了其他DNA结合结构域。在一些实施方案中,DNA结合结构域包含例如来自Natronobacterium gregoryi(NgAgo)的argonaute蛋白结构域。NgAgo是ssDNA引导的核酸内切酶,其由5'磷酸化ssDNA(gDNA)引导至其靶位点,在那里其产生双链断裂。与Cas9相反,NgAgo-gDNA系统不需要原间隔区相邻基序(PAM)。因此,使用核酸酶失活的NgAgo(dNgAgo)可以极大地扩展可以被靶向的碱基。NgAgo的表征和用途已在例如Gao et al.,Nat Biotechnol.(2016)34(7):768-73;Swarts et al.,Nature(2014)507(7491):258-61;和Swarts et al.,Nucl Acids Res.(2015)43(10):5120-9中描述。

[0118] 在一些实施方案中,DNA结合结构域包含失活的核酸酶,例如失活的大范围核酸

酶。DNA结合结构域的另外的非限制性示例包括四环素控制的阻遏物(tetR)DNA结合结构域、亮氨酸拉链、螺旋-环-螺旋(HLH)结构域、螺旋-转角-螺旋结构域、β-折叠基序、类固醇受体基序、bZIP结构域同源异型结构域和AT钩。

[0119] II.指导多核苷酸

[0120] 本文所述的包含多核苷酸指导的DNA结合结构域的表观遗传编辑器还可以包括能够与DNA结合结构域形成复合物的指导多核苷酸。指导多核苷酸可以包含RNA、DNA或两者的混合物。例如,在多核苷酸指导的DNA结合结构域是CRISPR相关蛋白结构域的情况下,指导多核苷酸可以是指导RNA(gRNA)。“指导RNA”或“gRNA”是指能够与靶序列杂交并指导CRISPR-Cas复合物与靶序列结合的核酸。使用具有可编程DNA结合蛋白(例如CRISPR相关蛋白结构域)的指导多核苷酸序列进行位点特异性DNA靶向(例如修饰基因组)的方法是本领域已知的。

[0121] 指导多核苷酸序列(例如,gRNA序列)可以包含两个部分:1)包含与靶核酸序列(“靶序列”)互补(例如与基因组靶位点中包含的核酸序列互补)的“靶向序列”的核苷酸序列;和2)结合多核苷酸引导的DNA结合结构域(例如,CRISPR-Cas蛋白结构域)的核苷酸序列。1)中的核苷酸序列可以包含与基因组核酸序列(例如基因组靶位点中包含的核酸序列)100%互补的靶向序列,并且因此可以与靶核酸序列杂交。1)中的核苷酸序列可以被称为例如crispr RNA或crRNA。2)中的核苷酸序列可以被称为指导核酸(例如tracrRNA)的支架序列或指导核酸的激活区,并且可以包含茎环结构。如上所述的部分1)和2)可以融合以形成一个单一指导(例如,单一指导RNA或sgRNA),或者可以在两个单独的核酸分子上。在一些实施方案中,指导多核苷酸包含通过接头连接的部分1)和2)。在一些实施方案中,指导多核苷酸包含通过非核酸接头(例如肽接头或化学接头)连接的部分1)和2)。

[0122] 如本文所述的指导多核苷酸的部分2(支架序列)可以是例如Jinek et al.,Science(2012)337:816-21;美国专利公开2016/0208288;或美国专利公开2016/0200779中所述。本公开还考虑了部分2)的变体。例如,gRNA支架(tracrRNA)序列的四环和茎环可以被修饰以包括RNA适体,其可以被特定的蛋白质结构域结合。在一些实施方案中,此类修饰的gRNA可用于促进与蛋白质相互作用RNA适体融合的抑制或激活结构域的招募。

[0123] 本文提供的gRNA通常包含靶向结构域和结合结构域。靶向结构域(也称为“靶向序列”)可以包含结合靶位点(例如细胞内的基因组核酸分子)的核酸序列。靶位点可以是包含PAM序列以及靶序列的双链DNA序列,其位于与PAM序列相同的链上并且直接相邻。gRNA的靶向结构域可以包含对应于靶序列的RNA序列,即,其类似于靶结构域的序列,有时具有一个或多个错配,但通常包含RNA序列而不是DNA序列。因此,gRNA的靶向结构域可以和与靶序列互补的双链靶位点的序列碱基配对(完全或部分互补),并且因此和与包含PAM序列的链互补的链碱基配对。应当理解,gRNA的靶向结构域通常不包括类似PAM序列的序列。还应理解,取决于所采用的核酸酶,PAM的位置可以是靶序列的5’或3’。例如,PAM通常位于Cas9核酸酶的靶序列的3’和Cas12a核酸酶的靶序列的5’。对于PAM的位置和gRNA与靶位点结合的机制的说明,参见例如Vanegas et al.,Fungal Biol Biotechnol.(2019)6:6的图1,其通过引用并入本文。对于RNA引导的核酸酶的gRNA靶向靶位点的机制的额外说明和描述,参见Fu etal.,Nat Biotechnol(2014)32(3):279-84和Sternberg et al.,Nature(2014)507(7490):62-7,各自通过引用并入本文。

[0124] 在一些实施方案中,靶向结构域序列包含17至30个核苷酸并且完全对应于靶序列(即,没有任何错配核苷酸)。然而,在一些实施方案中,靶向结构域序列可以包含一个或多个但通常不超过4个错配,例如1、2、3或4个错配。由于靶向结构域是gRNA(其是RNA分子)的一部分,其通常将包含核糖核苷酸,而DNA靶向结构域将包含脱氧核糖核苷酸。

[0125] 下面提供了包含22个核苷酸靶结构域和NGG PAM序列的Cas9靶位点,以及包含完全对应于靶序列(并且因此包含与包含靶序列和PAM的链互补的DNA链具有完全互补性的碱基对)的靶向结构域的gRNA的示例性说明:

[0126]

```
[                    靶结构域 (DNA)                    ][ PAM  ]
5'-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-G-G-3' (DNA)
3'-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-C-C-5' (DNA)
   | | | | | | | | | | | | | | | | | | | | | | | | | | |
5'-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-[ gRNA 支架 ]-3' (RNA)
[              靶向结构域 (RNA)                          ][  结合结构域  ]
```

[0127] 下面提供了包含22个核苷酸靶结构域和TTN PAM序列的Cas12a靶位点,以及包含完全对应于靶序列(并且因此包含与包含靶序列和PAM的链互补的DNA链具有完全互补性的碱基对)的靶向结构域的gRNA的示例性说明:

[0128]

```
                [  PAM  ][              靶结构域        ( DNA)              ]
                5'-T-T-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-3' (DNA)
                3'-A-A-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-5' (DNA)
                      | | | | | | | | | | | | | | | | | | | | | | | | |
5'-[gRNA scaffold]-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-3' (RNA)
[     结合结构域    ][                  靶向结构域 ( RNA)                  ]
```

[0129] 虽然不希望受理论束缚,但至少在一些实施方案中,据信靶向结构域的长度和与靶序列的互补性有助于gRNA/Cas9分子复合物与靶核酸相互作用的特异性。在一些实施方案中,本文提供的gRNA的靶向结构域的长度为5至50个核苷酸。在一些实施方案中,靶向结构域的长度为15至25个核苷酸。在一些实施方案中,靶向结构域的长度为18至22个核苷酸。在一些实施方案中,靶向结构域的长度为19-21个核苷酸。在一些实施方案中,靶向结构域的长度为15个核苷酸。在一些实施方案中,靶向结构域的长度为16个核苷酸。在一些实施方案中,靶向结构域的长度为17个核苷酸。在一些实施方案中,靶向结构域的长度为18个核苷酸。在一些实施方案中,靶向结构域的长度为19个核苷酸。在一些实施方案中,靶向结构域的长度为20个核苷酸。在一些实施方案中,靶向结构域的长度为21个核苷酸。在一些实施方案中,靶向结构域的长度为22个核苷酸。在一些实施方案中,靶向结构域的长度为23个核苷酸。在一些实施方案中,靶向结构域的长度为24个核苷酸。在一些实施方案中,靶向结构域的长度为25个核苷酸。在某些实施方案中,靶向结构域完全对应于本文提供的靶序列或其部分,没有错配。在一些实施方案中,本文提供的gRNA的靶向结构域相对于本文提供的靶序列包含1个错配。在一些实施方案中,靶向结构域相对于靶序列包含2个错配。在一些实施方案中,靶结构域相对于靶序列包含3个错配。

[0130] 用于设计、选择和验证gRNA的方法在本文中描述并且是本领域已知的。软件工具可用于优化对应于靶DNA序列的gRNA,例如以最小化整个基因组的总脱靶活性。例如,DNA序

列搜索算法可用于鉴定与Cas9一起使用的gRNA的crRNA中的靶序列。示例性gRNA设计工具包括Bae et al.,Bioinformatics(2014)30:1473-5中描述的那些。

[0131] 本文所述的指导多核苷酸(例如,gRNA)可以具有各种长度。在一些实施方案中,间隔区或靶向序列的长度取决于所使用的表观遗传编辑器系统的CRISPR相关蛋白组分。例如,来自不同细菌物种的Cas蛋白具有不同的最佳靶向序列长度。因此,间隔序列的长度可以包含例如5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50或超过50个核苷酸。在一些实施方案中,间隔区的长度包含10-24、11-20、11-16、18-24、19-21或20个核苷酸。在一些实施方案中,指导多核苷酸(例如gRNA)的长度为15-100(例如15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49或50)个核苷酸,并且包含与靶序列互补的至少10个(例如15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49或50)个连续核苷酸的间隔序列。在一些实施方案中,本文所述的指导多核苷酸可以被截短,例如截短1、2、3、4、5、10、15、20、25、30、40、50个或更多个核苷酸。

[0132] 在某些实施方案中,PCSK9靶序列的3'末端紧邻PAM序列(例如,典型PAM序列,如SpCas9的NGG)。指导多核苷酸的靶向序列(例如,gRNA的间隔序列)与靶序列之间的互补程度可以是至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%。在特定的实施方案中,靶向序列和靶序列可以是100%互补的。在其他实施方案中,靶向序列和靶序列可以包含例如1、2、3、4、5、6、7、8、9或10个错配。

[0133] 指导多核苷酸(例如gRNA)可以用例如化学改变和合成修饰来修饰。例如,经修饰的gRNA可以包括磷酸二酯主链键中的一个或两个非连接磷酸氧和/或一个或多个连接磷酸氧的改变或置换、核糖(例如核糖上的2'羟基)的改变、磷酸部分的改变、天然存在的核碱基的修饰或置换、核糖-磷酸主链的修饰或置换、寡核苷酸的3'末端和/或5'末端的修饰、末端磷酸基团的置换或部分、帽或接头的缀合或其任何组合。

[0134] 在一些实施方案中,gRNA的一个或多个核糖基团可以被修饰。核糖基团的化学修饰的示例包括但不限于2'-O-甲基(2'-OMe)、2'-氟(2'-F)、2'-脱氧、2'-O-(2-甲氧基乙基)(2'-MOE)、2'-NH2、2'-O-烯丙基、2'-O-乙胺、2'-O-氰基乙基、2'-O-缩醛酯或双环核苷酸,如锁核酸(LNA)、2'-(5-限制性乙基(S-cEt))、限制性MOE或2'-O,4'-C-氨基亚甲基桥连核酸(2',4'-BNANC)。2'-O-甲基修饰和/或2'-氟修饰可以增加gRNA寡核苷酸的结合亲和力和/或核酸酶稳定性。

[0135] 在一些实施方案中,gRNA的一个或多个磷酸基团可以被化学修饰。磷酸基团的化学修饰的示例包括但不限于硫代磷酸酯(PS)、膦酰基乙酸酯(PACE)、硫代膦酰基乙酸酯(硫代PACE)、酰胺、三唑、膦酸酯和磷酸三酯修饰。在一些实施方案中,本文所述的指导多核苷酸可以在5'端和/或3'端处或附近包含一个、两个、三个或更多个PS键;PS键可以是连续的或非连续的。

[0136] 在一些实施方案中,本文的gRNA包含核糖核苷酸和脱氧核糖核苷酸的混合物和/或一个或多个PS键。

[0137] 在一些实施方案中,gRNA的一个或多个核碱基可以被化学修饰。化学修饰的核碱基的示例包括但不限于2-硫尿苷、4-硫尿苷、N6-甲基腺苷、假尿苷、2,6-二氨基嘌呤、肌苷、

胸苷、5-甲基胞嘧啶、5-取代嘧啶、异鸟嘌呤、异胞嘧啶和具有卤代芳族基团的核碱基。可以在间隔区、tracr RNA区、茎环或其任何组合中进行化学修饰。

[0138] 下表2列出了用于人PCSK9的表观遗传修饰的示例性gRNA靶序列,以及人染色体1上被靶向位点的起始和终止位置的坐标(SEQ:SEQ ID NO)。该表还显示了PCSK9基因从起始坐标到TSS坐标的距离。

[0139] 表2.靶向PCSK9的gRNA的示例性靶序列

[0140]

gRNA No.	Chr. 1 链	起始	终止	gRNA 靶序列 (DNA, 5' to 3')	SEQ	TSS 距离
gRNA001	+	55039960	55039979	GGTGCTAGCCTTGCGTTCCG	1036	431
gRNA002	+	55039991	55040010	CTGGCCGAAGCACCCGAGCA	1037	462
gRNA003	-	55039529	55039548	TGCGGAAACCTTCTAGGGTG	1038	0
gRNA004	-	55039528	55039547	GCGGAAACCTTCTAGGGTGT	1039	-1
gRNA005	+	55039518	55039537	TCAAGCACCCACACCCTAGA	1040	-11
gRNA006	-	55039514	55039533	GGGTGTGGGTGCTTGACGCC	1041	-15
gRNA007	-	55039513	55039532	GGTGTGGGTGCTTGACGCCT	1042	-16
gRNA008	-	55039512	55039531	GTGTGGGTGCTTGACGCCTG	1043	-17
gRNA009	-	55039610	55039629	ACTGCCTGGCTCACTCCTCC	1044	81
gRNA010	-	55039481	55039500	TCACGCCACCAGAGCCCCAT	1045	-48
gRNA011	+	55039473	55039492	ATCGTCCGATGGGGCTCTGG	1046	-56
gRNA012	+	55039470	55039489	AGGATCGTCCGATGGGGCTC	1047	-59
gRNA013	+	55039464	55039483	TCAGATAGGATCGTCCGATG	1048	-65
gRNA014	+	55039669	55039688	GCGGCTCCCAGCTCCCAGCC	1049	140
gRNA015	-	55039678	55039697	CGGAATCCTGGCTGGGAGCT	1050	149
gRNA016	-	55039679	55039698	GCGGAATCCTGGCTGGGAGC	1051	150
gRNA017	-	55039685	55039704	GGGCGCGCGGAATCCTGGCT	1052	156
gRNA018	-	55039686	55039705	GGGGCGCGCGGAATCCTGGC	1053	157
gRNA019	+	55038610	55038629	TGAGGTCTTTGCAAACAAAG	1054	-919
gRNA020	+	55038654	55038673	CCAGCACCTAGATTCAGAGC	1055	-875
gRNA021	-	55038657	55038676	CCTGCTCTGAATCTAGGTGC	1056	-872
gRNA022	-	55038731	55038750	TCGATACTGGGAAGAAACAA	1057	-798
gRNA023	-	55038743	55038762	CTGGAAGGGCTGTCGATACT	1058	-786
gRNA024	-	55038744	55038763	TCTGGAAGGGCTGTCGATAC	1059	-785
gRNA025	-	55038757	55038776	GAGGCTTGCTCTTTCTGGAA	1060	-772
gRNA026	-	55038758	55038777	TGAGGCTTGCTCTTTCTGGA	1061	-771
gRNA027	-	55038762	55038781	GACATGAGGCTTGCTCTTTC	1062	-767
gRNA028	-	55038776	55038795	TGTACATGTGGCATGACATG	1063	-753
gRNA029	+	55038781	55038800	CATGCCACATGTACAATCTG	1064	-748
gRNA030	+	55038786	55038805	CACATGTACAATCTGAGGCC	1065	-743
gRNA031	-	55038788	55038807	CTGGCCTCAGATTGTACATG	1066	-741

[0141]

gRNA No.	Chr. 1 链	起始	终止	gRNA 靶序列 (DNA, 5' to 3')	SEQ	TSS 距离
gRNA032	-	55038807	55038826	AAAAGGGGAAAGAGAGCTCC	1067	-722
gRNA033	-	55038822	55038841	CCAGGCAGGAGGATGAAAAG	1068	-707
gRNA034	-	55038823	55038842	ACCAGGCAGGAGGATGAAAA	1069	-706
gRNA035	-	55038824	55038843	TACCAGGCAGGAGGATGAAA	1070	-705
gRNA036	+	55038830	55038849	CCTCCTGCCTGGTACACAAT	1071	-699
gRNA037	-	55038833	55038852	CCTATTGTGTACCAGGCAGG	1072	-696
gRNA038	-	55038836	55038855	ACACCTATTGTGTACCAGGC	1073	-693
gRNA039	+	55038840	55038859	GGTACACAATAGGTGTTTAC	1074	-689
gRNA040	-	55038840	55038859	GTAAACACCTATTGTGTACC	1075	-689
gRNA041	+	55038870	55038889	TCCAGTTGATTTCTTGAACA	1076	-659
gRNA042	-	55038874	55038893	ACCATGTTCAAGAAATCAAC	1077	-655
gRNA043	+	55038881	55038900	TCTTGAACATGGTGTGTAAA	1078	-648
gRNA044	+	55038906	55038925	TCTTTGCAAATTGAATCTTC	1079	-623
gRNA045	-	55038970	55038989	AGTTTGCAAAGACGTCATAT	1080	-559
gRNA046	-	55038998	55039017	GATTTATACTACAAAGATTC	1081	-531
gRNA047	+	55039075	55039094	AGTTGGTAAGGTCAGTGTGC	1082	-454
gRNA048	+	55039076	55039095	GTTGGTAAGGTCAGTGTGCA	1083	-453
gRNA049	+	55039093	55039112	GCAGGGTGCATAAAGGGCAG	1084	-436
gRNA050	+	55039097	55039116	GGTGCATAAAGGGCAGAGGC	1085	-432
gRNA051	+	55039100	55039119	GCATAAAGGGCAGAGGCCGG	1086	-429
gRNA052	+	55039137	55039156	TTTAGAAGGCTGCCAGGTTA	1087	-392
gRNA053	-	55039163	55039182	GCCCACCGAATTCTTTCCAC	1088	-366
gRNA054	+	55039167	55039186	AAAGAATTCGGTGGGCAGCG	1089	-362
gRNA055	-	55039196	55039215	CTTCTGAATCAATCCTACTG	1090	-333
gRNA056	+	55039221	55039240	CTGGTCAGCAGGAGACAAGG	1091	-308
gRNA057	+	55039202	55039221	GATTGATTCAGAAGTCTCAC	1092	-327
gRNA058	+	55039210	55039229	CAGAAGTCTCACTGGTCAGC	1093	-319
gRNA059	+	55039218	55039237	TCACTGGTCAGCAGGAGACA	1094	-311
gRNA060	+	55039228	55039247	GCAGGAGACAAGGTGGACCC	1095	-301
gRNA061	+	55039242	55039261	GGACCCAGGAAACACTGAAA	1096	-287
gRNA062	+	55039245	55039264	CCCAGGAAACACTGAAAAGG	1097	-284
gRNA063	+	55039246	55039265	CCAGGAAACACTGAAAAGGT	1098	-283
gRNA064	+	55039251	55039270	AAACACTGAAAAGGTGGGCC	1099	-278
gRNA065	+	55039281	55039300	TGGAGTCTGGCATCCCACGC	1100	-248
gRNA066	+	55039282	55039301	GGAGTCTGGCATCCCACGCA	1101	-247
gRNA067	+	55039312	55039331	CGGGAGAGGAGGAGCCCCTA	1102	-217
gRNA068	+	55039318	55039337	AGGAGGAGCCCCTAGGGCGC	1103	-211
gRNA069	-	55039329	55039348	AAGGCAGGCCGGCGCCCTAG	1104	-200
gRNA070	-	55039330	55039349	GAAGGCAGGCCGGCGCCCTA	1105	-199
gRNA071	-	55039331	55039350	GGAAGGCAGGCCGGCGCCCT	1106	-198
gRNA072	-	55039340	55039359	AACTGGGCTGGAAGGCAGGC	1107	-189
gRNA073	+	55039340	55039359	GCCTGCCTTCCAGCCCAGTT	1108	-189
gRNA074	-	55039344	55039363	TCCTAACTGGGCTGGAAGGC	1109	-185
gRNA075	+	55039346	55039365	CTTCCAGCCCAGTTAGGATT	1110	-183
gRNA076	+	55039347	55039366	TTCCAGCCCAGTTAGGATTT	1111	-182
gRNA077	-	55039348	55039367	CAAATCCTAACTGGGCTGGA	1112	-181
gRNA078	-	55039352	55039371	CTCCCAAATCCTAACTGGGC	1113	-177
gRNA079	-	55039356	55039375	AAAACTCCCAAATCCTAACT	1114	-173
gRNA080	-	55039357	55039376	AAAAACTCCCAAATCCTAAC	1115	-172
gRNA081	-	55039384	55039403	AGCGTCAGATTACGCGCAGA	1116	-145
gRNA082	-	55039385	55039404	CAGCGTCAGATTACGCGCAG	1117	-144
gRNA083	+	55039387	55039406	GCGCGTAATCTGACGCTGTT	1118	-142

[0142]

gRNA No.	Chr. 1 链	起始	终止	gRNA 靶序列 (DNA, 5' to 3')	SEQ	TSS 距离
gRNA084	+	55039388	55039407	CGCGTAATCTGACGCTGTTT	1119	-141
gRNA085	+	55039389	55039408	GCGTAATCTGACGCTGTTTG	1120	-140
gRNA086	+	55039392	55039411	TAATCTGACGCTGTTTGGGG	1121	-137
gRNA087	+	55039393	55039412	AATCTGACGCTGTTTGGGGA	1122	-136
gRNA088	+	55039398	55039417	GACGCTGTTTGGGGAGGGCG	1123	-131
gRNA089	+	55039422	55039441	CGAAACCTGATCCTCCAGTC	1124	-107
gRNA090	+	55039423	55039442	GAAACCTGATCCTCCAGTCC	1125	-106
gRNA091	+	55039424	55039443	AAACCTGATCCTCCAGTCCG	1126	-105
gRNA092	-	55039424	55039443	CGGACTGGAGGATCAGGTTT	1127	-105
gRNA093	+	55039425	55039444	AACCTGATCCTCCAGTCCGG	1128	-104
gRNA094	-	55039430	55039449	AACCCCCGGACTGGAGGATC	1129	-99
gRNA095	-	55039436	55039455	TAACGGAACCCCCGGACTGG	1130	-93
gRNA096	-	55039439	55039458	CATTAACGGAACCCCCGGAC	1131	-90
gRNA097	-	55039444	55039463	TTAAACATTAACGGAACCCC	1132	-85
gRNA098	+	55039450	55039469	CCGTTAATGTTTAATCAGAT	1133	-79
gRNA099	-	55039453	55039472	CCTATCTGATTAAACATTAA	1134	-76
gRNA100	+	55039462	55039481	AATCAGATAGGATCGTCCGA	1135	-67
gRNA101	+	55039463	55039482	ATCAGATAGGATCGTCCGAT	1136	-66
gRNA102	+	55039493	55039512	TGGCGTGATCTGCGCGCCCC	1137	-36
gRNA103	-	55039534	55039553	GTCGCTGCGGAAACCTTCTA	1138	5
gRNA104	-	55039535	55039554	CGTCGCTGCGGAAACCTTCT	1139	6
gRNA105	+	55039538	55039557	AGGTTTCCGCAGCGACGTCG	1140	9
gRNA106	+	55039547	55039566	CAGCGACGTCGAGGCGCTCA	1141	18
gRNA107	-	55039547	55039566	TGAGCGCCTCGACGTCGCTG	1142	18
gRNA108	+	55039554	55039573	GTCGAGGCGCTCATGGTTGC	1143	25
gRNA109	+	55039557	55039576	GAGGCGCTCATGGTTGCAGG	1144	28
gRNA110	+	55039558	55039577	AGGCGCTCATGGTTGCAGGC	1145	29
gRNA111	+	55039592	55039611	AGTTCAGGGTCTGAGCCTGG	1146	63
gRNA112	+	55039577	55039596	CGGGCGCCGCCGTTCAGTTC	1147	48
gRNA113	+	55039578	55039597	GGGCGCCGCCGTTCAGTTCA	1148	49
gRNA114	-	55039586	55039605	TCAGACCCTGAACTGAACGG	1149	57
gRNA115	-	55039589	55039608	GGCTCAGACCCTGAACTGAA	1150	60
gRNA116	+	55039589	55039608	TTCAGTTCAGGGTCTGAGCC	1151	60
gRNA117	+	55039603	55039622	TGAGCCTGGAGGAGTGAGCC	1152	74
gRNA118	+	55039615	55039634	AGTGAGCCAGGCAGTGAGAC	1153	86
gRNA119	+	55039620	55039639	GCCAGGCAGTGAGACTGGCT	1154	91
gRNA120	+	55039621	55039640	CCAGGCAGTGAGACTGGCTC	1155	92
gRNA121	-	55039624	55039643	CCCGAGCCAGTCTCACTGCC	1156	95
gRNA122	+	55039624	55039643	GGCAGTGAGACTGGCTCGGG	1157	95
gRNA123	+	55039625	55039644	GCAGTGAGACTGGCTCGGGC	1158	96
gRNA124	+	55039629	55039648	TGAGACTGGCTCGGGCGGGC	1159	100
gRNA125	+	55039630	55039649	GAGACTGGCTCGGGCGGGCC	1160	101
gRNA126	+	55039650	55039669	GGGACGCGTCGTTGCAGCAG	1161	121
gRNA127	-	55039651	55039670	GCTGCTGCAACGACGCGTCC	1162	122
gRNA128	-	55039690	55039709	TGAAGGGGCGCGCGGAATCC	1163	161
gRNA129	-	55039698	55039717	AGGGCGCGTGAAGGGGCGCG	1164	169
gRNA130	-	55039705	55039724	CAGGAGCAGGGCGCGTGAAG	1165	176
gRNA131	-	55039706	55039725	TCAGGAGCAGGGCGCGTGAA	1166	177
gRNA132	-	55039707	55039726	TTCAGGAGCAGGGCGCGTGA	1167	178
gRNA133	-	55039717	55039736	GGAGCTGAAGTTCAGGAGCA	1168	188
gRNA134	-	55039718	55039737	AGGAGCTGAAGTTCAGGAGC	1169	189
gRNA135	-	55039724	55039743	CTGTGCAGGAGCTGAAGTTC	1170	195

[0143]

gRNA No.	Chr. 1 链	起始	终止	gRNA 靶序列 (DNA, 5' to 3')	SEQ	TSS 距离
gRNA136	+	55039738	55039757	GCACAGTCCTCCCCACCGCA	1171	209
gRNA137	-	55039738	55039757	TGCGGTGGGGAGGACTGTGC	1172	209
gRNA138	+	55039745	55039764	CCTCCCCACCGCAAGGCTCA	1173	216
gRNA139	-	55039748	55039767	CCTTGAGCCTTGCGGTGGGG	1174	219
gRNA140	-	55039751	55039770	GCGCCTTGAGCCTTGCGGTG	1175	222
gRNA141	-	55039752	55039771	GGCGCCTTGAGCCTTGCGGT	1176	223
gRNA142	-	55039753	55039772	CGGCGCCTTGAGCCTTGCGG	1177	224
gRNA143	+	55039754	55039773	CGCAAGGCTCAAGGCGCCGC	1178	225
gRNA144	+	55039759	55039778	GGCTCAAGGCGCCGCCGGCG	1179	230
gRNA145	-	55039776	55039795	AGGCCGTGCGCGGTCCACGC	1180	247
gRNA146	+	55039778	55039797	GTGGACCGCGCACGGCCTCT	1181	249
gRNA147	-	55039786	55039805	GGAGACCTAGAGGCCGTGCG	1182	257
gRNA148	+	55039810	55039829	CAGGACAGCAACCTCTCCCC	1183	281
gRNA149	+	55039839	55039858	TGGGCACCGTCAGCTCCAGG	1184	310
gRNA150	+	55039845	55039864	CCGTCAGCTCCAGGCGGTCC	1185	316
gRNA151	-	55039848	55039867	CCAGGACCGCCTGGAGCTGA	1186	319
gRNA152	+	55039848	55039867	TCAGCTCCAGGCGGTCCTGG	1187	319
gRNA153	-	55039857	55039876	CAGCGGCCACCAGGACCGCC	1188	328
gRNA154	+	55039895	55039914	CTGCTGCTCCTGGGTCCCGC	1189	366
gRNA155	-	55039906	55039925	CACGGGCGCCCGCGGGACCC	1190	377
gRNA156	+	55039909	55039928	TCCCGCGGGCGCCCGTGCGC	1191	380
gRNA157	+	55039912	55039931	CGCGGGCGCCCGTGCGCAGG	1192	383
gRNA158	-	55039913	55039932	TCCTGCGCACGGGCGCCCGC	1193	384
gRNA159	-	55039914	55039933	CTCCTGCGCACGGGCGCCCG	1194	385
gRNA160	+	55039918	55039937	CGCCCGTGCGCAGGAGGACG	1195	389
gRNA161	+	55039922	55039941	CGTGCGCAGGAGGACGAGGA	1196	393
gRNA162	-	55039923	55039942	GTCCTCGTCCTCCTGCGCAC	1197	394
gRNA163	-	55039924	55039943	CGTCCTCGTCCTCCTGCGCA	1198	395
gRNA164	+	55039933	55039952	GGACGAGGACGGCGACTACG	1199	404
gRNA165	+	55039939	55039958	GGACGGCGACTACGAGGAGC	1200	410
gRNA166	+	55039963	55039982	GCTAGCCTTGCGTTCCGAGG	1201	434
gRNA167	+	55039967	55039986	GCCTTGCGTTCCGAGGAGGA	1202	438
gRNA168	-	55039971	55039990	GCCGTCCTCCTCGGAACGCA	1203	442
gRNA169	+	55039972	55039991	GCGTTCCGAGGAGGACGGCC	1204	443
gRNA170	-	55039980	55039999	TTCGGCCAGGCCGTCCTCCT	1205	451
gRNA171	-	55039993	55040012	CGTGCTCGGGTGCTTCGGCC	1206	464
gRNA172	-	55040006	55040025	GTGGCTGTGGTTCCGTGCTC	1207	477
gRNA173	-	55040007	55040026	GGTGGCTGTGGTTCCGTGCT	1208	478
gRNA174	-	55040019	55040038	GCAGCGGTGGAAGGTGGCTG	1209	490
gRNA175	+	55040023	55040042	CACCTTCCACCGCTGCGCCA	1210	494
gRNA176	-	55040025	55040044	CTTGGCGCAGCGGTGGAAGG	1211	496
gRNA177	-	55040028	55040047	CACCTTGGCGCAGCGGTGGA	1212	499
gRNA178	+	55040028	55040047	TCCACCGCTGCGCCAAGGTG	1213	499
gRNA179	+	55040029	55040048	CCACCGCTGCGCCAAGGTGC	1214	500
gRNA180	-	55040032	55040051	CCCGCACCTTGGCGCAGCGG	1215	503
gRNA181	+	55040070	55040089	GGGCGAACCCGCAGCCGGGA	1216	541
gRNA182	-	55040070	55040089	TCCCGGCTGCGGGTTCGCCC	1217	541
gRNA183	-	55040080	55040099	CACCGCACCGTCCCGGCTGC	1218	551
gRNA184	-	55040081	55040100	GCACCGCACCGTCCCGGCTG	1219	552
gRNA185	-	55040087	55040106	GAAACAGCACCGCACCGTCC	1220	558
gRNA186	+	55040091	55040110	GGTGCGGTGCTGTTTCCTCT	1221	562
gRNA187	+	55040092	55040111	GTGCGGTGCTGTTTCCTCTC	1222	563

[0144]

gRNA No.	Chr. 1 链	起始	终止	gRNA 靶序列 (DNA, 5' to 3')	SEQ	TSS 距离
gRNA188	-	55040109	55040128	GGGGGAAACTGAGGCCCGAG	1223	580
gRNA189	+	55040119	55040138	AGTTTCCCCCCATGTAAGAG	1224	590
gRNA190	+	55040125	55040144	CCCCCATGTAAGAGAGGAAG	1225	596
gRNA191	-	55040127	55040146	CACTTCCTCTCTTACATGGG	1226	598
gRNA192	-	55040128	55040147	CCACTTCCTCTCTTACATGG	1227	599
gRNA193	+	55040133	55040152	TAAGAGAGGAAGTGGAGTGC	1228	604
gRNA194	+	55040155	55040174	GTCGCCGAGGGCTCTTCGCT	1229	626
gRNA195	-	55040162	55040181	CGTGCCAAGCGAAGAGCCCT	1230	633
gRNA196	+	55040166	55040185	CTCTTCGCTTGGCACGATCT	1231	637
gRNA197	+	55040167	55040186	TCTTCGCTTGGCACGATCTT	1232	638
gRNA198	+	55040168	55040187	CTTCGCTTGGCACGATCTTG	1233	639
gRNA199	+	55040176	55040195	GGCACGATCTTGGGGACTGC	1234	647
gRNA200	+	55040181	55040200	GATCTTGGGGACTGCAGGCA	1235	652
gRNA201	+	55040184	55040203	CTTGGGGACTGCAGGCAAGG	1236	655
gRNA202	+	55040189	55040208	GGACTGCAGGCAAGGCGGCG	1237	660
gRNA203	+	55040190	55040209	GACTGCAGGCAAGGCGGCGG	1238	661
gRNA204	+	55040193	55040212	TGCAGGCAAGGCGGCGGGGG	1239	664
gRNA205	+	55040197	55040216	GGCAAGGCGGCGGGGGAGGA	1240	668
gRNA206	+	55040198	55040217	GCAAGGCGGCGGGGGAGGAC	1241	669
gRNA207	+	55040223	55040242	GTGGGGAGCACGGTGGAGAG	1242	694
gRNA208	+	55040224	55040243	TGGGGAGCACGGTGGAGAGC	1243	695
gRNA209	+	55040225	55040244	GGGGAGCACGGTGGAGAGCG	1244	696
gRNA210	+	55040229	55040248	AGCACGGTGGAGAGCGGGGA	1245	700
gRNA211	+	55040233	55040252	CGGTGGAGAGCGGGGACGGC	1246	704
gRNA212	+	55040277	55040296	CGTGCGGCTGCGCTATTCAG	1247	748
gRNA213	+	55040278	55040297	GTGCGGCTGCGCTATTCAGT	1248	749
gRNA214	+	55040282	55040301	GGCTGCGCTATTCAGTGGGA	1249	753
gRNA215	+	55040289	55040308	CTATTCAGTGGGAAGGTTCG	1250	760
gRNA216	+	55040290	55040309	TATTCAGTGGGAAGGTTCGC	1251	761
gRNA217	+	55040291	55040310	ATTCAGTGGGAAGGTTCGCG	1252	762
gRNA218	+	55040295	55040314	AGTGGGAAGGTTCGCGGGGT	1253	766
gRNA219	+	55040296	55040315	GTGGGAAGGTTCGCGGGGTT	1254	767
gRNA220	+	55040337	55040356	AGGGCGAGCAGAGCACTGCC	1255	808
gRNA221	+	55040382	55040401	TTTCTGCCTCGCCGCGGCAC	1256	853
gRNA222	+	55040385	55040404	CTGCCTCGCCGCGGCACAGG	1257	856
gRNA223	+	55040386	55040405	TGCCTCGCCGCGGCACAGGT	1258	857
gRNA224	-	55040391	55040410	CACCCACCTGTGCCGCGGCG	1259	862
gRNA225	-	55040396	55040415	TCCTTCACCCACCTGTGCCG	1260	867
gRNA226	+	55040405	55040424	TGGGTGAAGGAGTGAATGCC	1261	876

[0145] 在一些实施方案中,本文的gRNA不包含序列CCCGCACCUUGGCGCAGCGG(SEQ ID NO:1490)。

[0146] 本领域已知的任何tracr序列可考虑用于本文所述的gRNA。在一些实施方案中,本文描述的gRNA具有下表3所示的tracr序列,或与下面所示的tracr序列至少70%、75%、80%、85%、90%、95%、96%、97%、98%或99%相同的tracr序列(SEQ:SEQ ID NO)。

[0147] 表3.示例性TRACR序列

[0148]

SEQ	序列 (5' to 3')
653	GUUUAAGAGCUAUGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACU UGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU
654	GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUG GCACCGAGUCGGUGCUUUU
655	GUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACU UGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
656	GUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACU UGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU

[0149] 在一些实施方案中,本文的gRNA直接提供给细胞(例如,通过连同CRISPR相关蛋白结构域的RNP复合物)。在一些实施方案中,gRNA通过引入细胞中的表达载体(例如,质粒载体或病毒载体)提供给细胞,然后,其中细胞从表达载体表达gRNA。将gRNA和表达载体引入细胞的方法是本领域熟知的。

[0150] III.效应结构域

[0151] 本文所述的表观遗传编辑器包括影响靶基因的表观遗传修饰的一个或多个效应蛋白结构域(也称为"表观遗传效应结构域"或"效应结构域",如本文所用)。具有一个或多个效应结构域的表观遗传编辑器可以调节靶基因的表达而不改变其核碱基序列。在一些实施方案中,本文所述的效应结构域可以例如通过抑制转录或通过修饰或重塑染色质来提供靶基因例如PCSK9的表达的抑制或沉默。此类效应结构域在本文中也称为"抑制结构域"、"阻遏结构域"或"表观遗传阻遏结构域"。可由效应结构域介导的化学修饰的非限制性示例包括DNA或组蛋白残基的甲基化、去甲基化、乙酰化、去乙酰化、磷酸化、SUMO化和/或泛素化。

[0152] 在一些实施方案中,本文所述的表观遗传编辑器的效应结构域可以进行组蛋白尾部修饰,例如通过在组蛋白尾部上添加或去除活性标记。

[0153] 在一些实施方案中,本文所述的表观遗传编辑器的效应结构域可以包含或招募转录相关蛋白,例如转录阻遏物。转录相关蛋白可以是内源性的或外源性的。

[0154] 在一些实施方案中,本文所述的表观遗传编辑器的效应结构域可以例如包含直接或间接阻断转录因子接近含有靶序列的目的基因的蛋白质。

[0155] 效应结构域可以是保留表观遗传效应子功能的全长蛋白质或其片段("功能结构域")。能够调节(例如,抑制)基因表达的功能结构域可以衍生自较大的蛋白质。例如,可以基于阻遏蛋白的序列鉴定可以降低靶基因表达的功能结构域。调节基因表达的蛋白的氨基酸序列可以从可用的基因组浏览器获得,例如UCSD基因组浏览器或Ensembl基因组浏览器。蛋白质注释数据库如UniProt或Pfam可用于鉴定全蛋白质序列内的功能结构域。作为起点,可以测试包含由不同数据库鉴定的所有区域的最大序列的基因表达调节活性。然后可以测试各种截短物以识别最小功能单元。

[0156] 本公开还考虑了本文所述的效应结构域的变体。变体可以例如是指与本文所述的野生型效应结构域具有至少约85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%序列同一性和/或序列相似性的多肽。在特定的实施方案中,变体保留野生型效应结构域的至少约85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%的表观遗传效应子功能。

[0157] 在一些实施方案中,本文所述的效应结构域可包含两个或更多个效应结构域(例如,KOX1 KRAB和ZIM3)的融合体。效应结构域可以例如包含2、3、4、5、6、7、8、9或10个效应结

构域的融合体,例如本文所述的效应结构域。在某些实施方案中,效应结构域包含截短形式的效应结构域和第二效应结构域的融合体。在某些实施方案中,效应结构域包含两个效应结构域的截短形式的融合体(例如,两个效应结构域的N端和C端部分的融合体)。

[0158] 在一些实施方案中,本文所述的表观遗传编辑器可以包含1个效应结构域、2个效应结构域、3个效应结构域、4个效应结构域、5个效应结构域、6个效应结构域、7个效应结构域、8个效应结构域、9个效应结构域、10个效应结构域或更多个。在某些实施方案中,表观遗传编辑器包含一个或多个融合蛋白(例如,一个、两个或三个融合蛋白),每个融合蛋白具有与DNA结合结构域连接的一个或多个效应结构域(例如,一个、两个或三个效应结构域)。在一些实施方案中,效应结构域可以诱导表观遗传修饰的组合,例如转录抑制和DNA甲基化、DNA甲基化和组蛋白去乙酰化、DNA甲基化和组蛋白去甲基化、DNA甲基化和组蛋白甲基化、DNA甲基化和组蛋白磷酸化、DNA甲基化和组蛋白泛素化、DNA甲基化和组蛋白SUMO化。

[0159] 在某些实施方案中,本文所述的效应结构域(例如,DNMT3A和/或DNMT3L)由在该效应结构域的天然基因组(例如,人或鼠)中发现的核苷酸序列编码。在其他实施方案中,本文所述的效应结构域由已被密码子优化以在人细胞中最佳表达的核苷酸序列编码。

[0160] 本文所述的效应结构域可以包括例如转录阻遏物、DNA甲基转移酶和/或组蛋白修饰物,如下文进一步详述。

[0161] A.转录阻遏物

[0162] 在一些实施方案中,本文所述的表观遗传效应结构域介导靶基因表达(例如,转录)的阻遏。效应结构域可以包含例如Krüppel相关盒(KRAB)阻遏结构域、阻遏物元件沉默转录因子(REST)阻遏结构域、KRAB相关蛋白1(KAP1)结构域、MAD结构域、FKHR(横纹肌肉瘤基因中的叉头)阻遏结构域、EGR-1(早期生长反应基因产物-1)阻遏结构域、ets2阻遏物因子阻遏结构域(ERD)、MAD smSIN3相互作用结构域(SID)、毛发相关碱性螺旋-环-螺旋(bHLH)阻遏物蛋白的WRPW基序、HP1αchromo-shadow阻遏结构域、HP1β阻遏结构域或其任何组合。效应结构域可以招募一个或多个抑制靶基因表达的蛋白质结构域,例如通过支架蛋白。在一些实施方案中,效应结构域可以招募支架蛋白结构域或与招募PRMT蛋白、HDAC蛋白、SETDB1蛋白或NuRD蛋白结构域的支架蛋白结构域相互作用。

[0163] 在一些实施方案中,效应结构域包含衍生自锌指阻遏蛋白的功能结构域,例如KRAB结构域。在大约400种基于人ZFP的转录因子中发现了KRAB结构域。KRAB结构域的描述可以在例如Ecco et al.,Development(2017)144(15):2719-29和Lambert et al.,Cell(2018)172:650-65中找到。

[0164] 在某些实施方案中,效应结构域包含衍生自KOX1/ZNF10、KOX8/ZNF708、ZNF43、ZNF184、ZNF91、HPF4、HTF10或HTF34的阻遏结构域(例如,KRAB)。在一些实施方案中,效应结构域包含衍生自ZIM3、ZNF436、ZNF257、ZNF675、ZNF490、ZNF320、ZNF331、ZNF816、ZNF680、ZNF41、ZNF189、ZNF528、ZNF543、ZNF554、ZNF140、ZNF610、ZNF264、ZNF350、ZNF8、ZNF582、ZNF30、ZNF324、ZNF98、ZNF669、ZNF677、ZNF596、ZNF214、ZNF37、ZNF34、ZNF250、ZNF547、ZNF273、ZNF354、ZFP82、ZNF224、ZNF33、ZNF45、ZNF175、ZNF595、ZNF184、ZNF419、ZFP28-1、ZFP28-2、ZNF18、ZNF213、ZNF394、ZFP1、ZFP14、ZNF416、ZNF557、ZNF566、ZNF729、ZIM2、ZNF254、ZNF764、ZNF785或其任意组合的阻遏结构域(例如,KRAB)。例如,阻遏结构域可以是衍生自KOX1、ZIM3、ZFP28或ZN627的KRAB结构域。在特定的实施方案中,阻遏结构域是ZIM3

KRAB结构域。在另外的实施方案中,效应结构域衍生自人蛋白,例如人ZIM3、人KOX1、人ZFP28或人ZN627。

[0165] 可以降低或沉默靶基因表达的示例性效应结构域的序列或含有它们的蛋白质序列提供于下表4中(SEQ:SEQ ID NO)。阻遏物和转录阻遏结构域的进一步示例可以在例如PCT专利公开WO 2021/226077和Tycko et al.,Cell(2020)183(7):2020-35中找到,其各自通过引用整体并入本文。

[0166] 表4.可降低或沉默基因表达的示例性效应结构域

[0167]

蛋白	SEQ
ZIM3	33
ZNF436	34
ZNF257	35
ZNF675	36
ZNF490	37
ZNF320	38
ZNF331	39
ZNF816	40
ZNF680	41
ZNF41	42
ZNF189	43
ZNF528	44
ZNF543	45
ZNF554	46
ZNF140	47
ZNF610	48
ZNF264	49
ZNF350	50
ZNF8	51
ZNF582	52
ZNF30	53

蛋白	SEQ
ZN440_HUMAN	299
ZN583_HUMAN	300
ZN441_HUMAN	301
ZNF43_HUMAN	302
CBX5_HUMAN	303
ZN589_HUMAN	304
ZNF10_HUMAN	305
ZN563_HUMAN	306
ZN561_HUMAN	307
ZN136_HUMAN	308
ZN630_HUMAN	309
ZN527_HUMAN	310
ZN333_HUMAN	311
Z324B_HUMAN	312
ZN786_HUMAN	313
ZN709_HUMAN	314
ZN792_HUMAN	315
ZN599_HUMAN	316
ZN613_HUMAN	317
ZF69B_HUMAN	318
ZN799_HUMAN	319

[0168]

蛋白	SEQ
ZNF324	54
ZNF98	55
ZNF669	56
ZNF677	57
ZNF596	58
ZNF214	59
ZNF37A	60
ZNF34	61
ZNF250	62
ZNF547	63
ZNF273	64
ZNF354A	65
ZFP82	66
ZNF224	67
ZNF33A	68
ZNF45	69
ZNF175	70
ZNF595	71
ZNF184	72
ZNF419	73
ZFP28-1	74
ZFP28-2	75
ZNF18	76
ZNF213	77
ZNF394	78
ZFP1	79
ZFP14	80
ZNF416	81
ZNF557	82
ZNF566	83
ZNF729	84
ZIM2	85
ZNF254	86
ZNF764	87
ZNF785	88
ZNF10 (KOX1)	89
CBX5 (chromoshadow 结构域)	90

蛋白	SEQ
ZN569_HUMAN	320
ZN564_HUMAN	321
ZN546_HUMAN	322
ZFP92_HUMAN	323
YAF2_HUMAN	324
ZN723_HUMAN	325
ZNF34_HUMAN	326
ZN439_HUMAN	327
ZFP57_HUMAN	328
ZNF19_HUMAN	329
ZN404_HUMAN	330
ZN274_HUMAN	331
CBX3_HUMAN	332
ZNF30_HUMAN	333
ZN250_HUMAN	334
ZN570_HUMAN	335
ZN675_HUMAN	336
ZN695_HUMAN	337
ZN548_HUMAN	338
ZN132_HUMAN	339
ZN738_HUMAN	340
ZN420_HUMAN	341
ZN626_HUMAN	342
ZN559_HUMAN	343
ZN460_HUMAN	344
ZN268_HUMAN	345
ZN304_HUMAN	346
ZIM2_HUMAN	347
ZN605_HUMAN	348
ZN844_HUMAN	349
SUMO5_HUMAN	350
ZN101_HUMAN	351
ZN783_HUMAN	352
ZN417_HUMAN	353
ZN182_HUMAN	354
ZN823_HUMAN	355
ZN177_HUMAN	356

[0169]

蛋白	SEQ
RYBP (PRC1 的 YAF2_RYBP 组分)	91
YAF2 (PRC1 的 YAF2_RYBP 组分)	92
MGA (PRC1.6 的组分)	93
CBX1 (chromoshadow)	94
SCMH1 (SAM_1/SPM)	95
MPP8 (Chromodomain)	96
SUMO3 (Rad60-SLD)	97
HERC2 (Cyt-b5)	98
BIN1 (SH3_9)	99
PCGF2 (环指蛋白结构域)	100
TOX (HMG box)	101
FOXA1 (HNF3A C 端结构域)	102
FOXA2 (HNF3B C 端结构域)	103
IRF2BP1 (IRF-2BP1_2 N 端结构域)	104
IRF2BP2 (IRF-2BP1_2 N 端结构域)	105
IRF2BPL IRF-2BP1_2 N 端结构域	106
HOXA13 (同源异型结构域)	107
HOXB13 (同源异型结构域)	108
HOXC13 (同源异型结构域)	109
HOXA11 (同源异型结构域)	110
HOXC11 (同源异型结构域)	111
HOXC10 (同源异型结构域)	112
HOXA10 (同源异型结构域)	113
HOXB9 (同源异型结构域)	114
HOXA9 (同源异型结构域)	115
ZFP28_HUMAN	116
ZN334_HUMAN	117
ZN568_HUMAN	118
ZN37A_HUMAN	119
ZN181_HUMAN	120
ZN510_HUMAN	121
ZN862_HUMAN	122
ZN140_HUMAN	123

蛋白	SEQ
ZN197_HUMAN	357
ZN717_HUMAN	358
ZN669_HUMAN	359
ZN256_HUMAN	360
ZN251_HUMAN	361
CBX4_HUMAN	362
PCGF2_HUMAN	363
CDY2_HUMAN	364
CDYL2_HUMAN	365
HERC2_HUMAN	366
ZN562_HUMAN	367
ZN461_HUMAN	368
Z324A_HUMAN	369
ZN766_HUMAN	370
ID2_HUMAN	371
TOX_HUMAN	372
ZN274_HUMAN	373
SCMH1_HUMAN	374
ZN214_HUMAN	375
CBX7_HUMAN	376
ID1_HUMAN	377
CREM_HUMAN	378
SCX_HUMAN	379
ASCL1_HUMAN	380
ZN764_HUMAN	381
SCML2_HUMAN	382
TWST1_HUMAN	383
CREB1_HUMAN	384
TERF1_HUMAN	385
ID3_HUMAN	386
CBX8_HUMAN	387
CBX4_HUMAN	388
GSX1_HUMAN	389

[0170]

蛋白	SEQ
ZN208_HUMAN	124
ZN248_HUMAN	125
ZN571_HUMAN	126
ZN699_HUMAN	127
ZN726_HUMAN	128
ZIK1_HUMAN	129
ZNF2_HUMAN	130
Z705F_HUMAN	131
ZNF14_HUMAN	132
ZN471_HUMAN	133
ZN624_HUMAN	134
ZNF84_HUMAN	135
ZNF7_HUMAN	136
ZN891_HUMAN	137
ZN337_HUMAN	138
Z705G_HUMAN	139
ZN529_HUMAN	140
ZN729_HUMAN	141
ZN419_HUMAN	142
Z705A_HUMAN	143
ZNF45_HUMAN	144
ZN302_HUMAN	145
ZN486_HUMAN	146
ZN621_HUMAN	147
ZN688_HUMAN	148
ZN33A_HUMAN	149
ZN554_HUMAN	150
ZN878_HUMAN	151
ZN772_HUMAN	152
ZN224_HUMAN	153
ZN184_HUMAN	154
ZN544_HUMAN	155
ZNF57_HUMAN	156
ZN283_HUMAN	157
ZN549_HUMAN	158
ZN211_HUMAN	159
ZN615_HUMAN	160

蛋白	SEQ
NKX22_HUMAN	390
ATF1_HUMAN	391
TWST2_HUMAN	392
ZNF17_HUMAN	393
TOX3_HUMAN	394
TOX4_HUMAN	395
ZMYM3_HUMAN	396
I2BP1_HUMAN	397
RHXF1_HUMAN	398
SSX2_HUMAN	399
I2BPL_HUMAN	400
ZN680_HUMAN	401
CBX1_HUMAN	402
TRI68_HUMAN	403
HXA13_HUMAN	404
PHC3_HUMAN	405
TCF24_HUMAN	406
CBX3_HUMAN	407
HXB13_HUMAN	408
HEY1_HUMAN	409
PHC2_HUMAN	410
ZNF81_HUMAN	411
FIGLA_HUMAN	412
SAM11_HUMAN	413
KMT2B_HUMAN	414
HEY2_HUMAN	415
JDP2_HUMAN	416
HXC13_HUMAN	417
ASCL4_HUMAN	418
HHEX_HUMAN	419
HERC2_HUMAN	420
GSX2_HUMAN	421
BIN1_HUMAN	422
ETV7_HUMAN	423
ASCL3_HUMAN	424
PHC1_HUMAN	425
OTP_HUMAN	426

[0171]

蛋白	SEQ
ZN253_HUMAN	161
ZN226_HUMAN	162
ZN730_HUMAN	163
Z585A_HUMAN	164
ZN732_HUMAN	165
ZN681_HUMAN	166
ZN667_HUMAN	167
ZN649_HUMAN	168
ZN470_HUMAN	169
ZN484_HUMAN	170
ZN431_HUMAN	171
ZN382_HUMAN	172
ZN254_HUMAN	173
ZN124_HUMAN	174
ZN607_HUMAN	175
ZN317_HUMAN	176
ZN620_HUMAN	177
ZN141_HUMAN	178
ZN584_HUMAN	179
ZN540_HUMAN	180
ZN75D_HUMAN	181
ZN555_HUMAN	182
ZN658_HUMAN	183
ZN684_HUMAN	184
RBAK_HUMAN	185
ZN829_HUMAN	186
ZN582_HUMAN	187
ZN112_HUMAN	188
ZN716_HUMAN	189
HKR1_HUMAN	190
ZN350_HUMAN	191
ZN480_HUMAN	192
ZN416_HUMAN	193
ZNF92_HUMAN	194
ZN100_HUMAN	195
ZN736_HUMAN	196
ZNF74_HUMAN	197

蛋白	SEQ
I2BP2_HUMAN	427
VGLL2_HUMAN	428
HXA11_HUMAN	429
PDLI4_HUMAN	430
ASCL2_HUMAN	431
CDX4_HUMAN	432
ZN860_HUMAN	433
LMBL4_HUMAN	434
PDIP3_HUMAN	435
NKX25_HUMAN	436
CEBPB_HUMAN	437
ISL1_HUMAN	438
CDX2_HUMAN	439
PROP1_HUMAN	440
SIN3B_HUMAN	441
SMBT1_HUMAN	442
HXC11_HUMAN	443
HXC10_HUMAN	444
PRS6A_HUMAN	445
VSX1_HUMAN	446
NKX23_HUMAN	447
MTG16_HUMAN	448
HMX3_HUMAN	449
HMX1_HUMAN	450
KIF22_HUMAN	451
CSTF2_HUMAN	452
CEBPE_HUMAN	453
DLX2_HUMAN	454
ZMYM3_HUMAN	455
PPARG_HUMAN	456
PRIC1_HUMAN	457
UNC4_HUMAN	458
BARX2_HUMAN	459
ALX3_HUMAN	460
TCF15_HUMAN	461
TERA_HUMAN	462
VSX2_HUMAN	463

[0172]

蛋白	SEQ
CBX1_HUMAN	198
ZN443_HUMAN	199
ZN195_HUMAN	200
ZN530_HUMAN	201
ZN782_HUMAN	202
ZN791_HUMAN	203
ZN331_HUMAN	204
Z354C_HUMAN	205
ZN157_HUMAN	206
ZN727_HUMAN	207
ZN550_HUMAN	208
ZN793_HUMAN	209
ZN235_HUMAN	210
ZNF8_HUMAN	211
ZN724_HUMAN	212
ZN573_HUMAN	213
ZN577_HUMAN	214
ZN789_HUMAN	215
ZN718_HUMAN	216
ZN300_HUMAN	217
ZN383_HUMAN	218
ZN429_HUMAN	219
ZN677_HUMAN	220
ZN850_HUMAN	221
ZN454_HUMAN	222
ZN257_HUMAN	223
ZN264_HUMAN	224
ZFP82_HUMAN	225
ZFP14_HUMAN	226
ZN485_HUMAN	227
ZN737_HUMAN	228
ZNF44_HUMAN	229
ZN596_HUMAN	230
ZN565_HUMAN	231
ZN543_HUMAN	232
ZFP69_HUMAN	233
SUMO1_HUMAN	234

蛋白	SEQ
HXD12_HUMAN	464
CDX1_HUMAN	465
TCF23_HUMAN	466
ALX1_HUMAN	467
HXA10_HUMAN	468
RX_HUMAN	469
CXXC5_HUMAN	470
SCML1_HUMAN	471
NFIL3_HUMAN	472
DLX6_HUMAN	473
MTG8_HUMAN	474
CBX8_HUMAN	475
CEBPD_HUMAN	476
SEC13_HUMAN	477
FIP1_HUMAN	478
ALX4_HUMAN	479
LHX3_HUMAN	480
PRIC2_HUMAN	481
MAGI3_HUMAN	482
NELL1_HUMAN	483
PRRX1_HUMAN	484
MTG8R_HUMAN	485
RAX2_HUMAN	486
DLX3_HUMAN	487
DLX1_HUMAN	488
NKX26_HUMAN	489
NAB1_HUMAN	490
SAMD7_HUMAN	491
PITX3_HUMAN	492
WDR5_HUMAN	493
MEOX2_HUMAN	494
NAB2_HUMAN	495
DHX8_HUMAN	496
FOXA2_HUMAN	497
CBX6_HUMAN	498
EMX2_HUMAN	499
CPSF6_HUMAN	500

[0173]

蛋白	SEQ
ZNF12_HUMAN	235
ZN169_HUMAN	236
ZN433_HUMAN	237
SUMO3_HUMAN	238
ZNF98_HUMAN	239
ZN175_HUMAN	240
ZN347_HUMAN	241
ZNF25_HUMAN	242
ZN519_HUMAN	243
Z585B_HUMAN	244
ZIM3_HUMAN	245
ZN517_HUMAN	246
ZN846_HUMAN	247
ZN230_HUMAN	248
ZNF66_HUMAN	249
ZFP1_HUMAN	250
ZN713_HUMAN	251
ZN816_HUMAN	252
ZN426_HUMAN	253
ZN674_HUMAN	254
ZN627_HUMAN	255
ZNF20_HUMAN	256
Z587B_HUMAN	257
ZN316_HUMAN	258
ZN233_HUMAN	259
ZN611_HUMAN	260
ZN556_HUMAN	261
ZN234_HUMAN	262
ZN560_HUMAN	263
ZNF77_HUMAN	264
ZN682_HUMAN	265
ZN614_HUMAN	266
ZN785_HUMAN	267
ZN445_HUMAN	268
ZFP30_HUMAN	269
ZN225_HUMAN	270
ZN551_HUMAN	271

蛋白	SEQ
HXC12_HUMAN	501
KDM4B_HUMAN	502
LMBL3_HUMAN	503
PHX2A_HUMAN	504
EMX1_HUMAN	505
NC2B_HUMAN	506
DLX4_HUMAN	507
SRY_HUMAN	508
ZN777_HUMAN	509
NELL1_HUMAN	510
ZN398_HUMAN	511
GATA3_HUMAN	512
BSH_HUMAN	513
SF3B4_HUMAN	514
TEAD1_HUMAN	515
TEAD3_HUMAN	516
RGAP1_HUMAN	517
PHF1_HUMAN	518
FOXA1_HUMAN	519
GATA2_HUMAN	520
FOXO3_HUMAN	521
ZN212_HUMAN	522
IRX4_HUMAN	523
ZBED6_HUMAN	524
LHX4_HUMAN	525
SIN3A_HUMAN	526
RBBP7_HUMAN	527
NKX61_HUMAN	528
TRI68_HUMAN	529
R51A1_HUMAN	530
MB3L1_HUMAN	531
DLX5_HUMAN	532
NOTC1_HUMAN	533
TERF2_HUMAN	534
ZN282_HUMAN	535
RGS12_HUMAN	536
ZN840_HUMAN	537

[0174]

蛋白	SEQ
ZN610_HUMAN	272
ZN528_HUMAN	273
ZN284_HUMAN	274
ZN418_HUMAN	275
MPP8_HUMAN	276
ZN490_HUMAN	277
ZN805_HUMAN	278
Z780B_HUMAN	279
ZN763_HUMAN	280
ZN285_HUMAN	281
ZNF85_HUMAN	282
ZN223_HUMAN	283
ZNF90_HUMAN	284
ZN557_HUMAN	285
ZN425_HUMAN	286
ZN229_HUMAN	287
ZN606_HUMAN	288
ZN155_HUMAN	289
ZN222_HUMAN	290
ZN442_HUMAN	291
ZNF91_HUMAN	292
ZN135_HUMAN	293
ZN778_HUMAN	294
RYBP_HUMAN	295
ZN534_HUMAN	296
ZN586_HUMAN	297
ZN567_HUMAN	298

蛋白	SEQ
SPI2B_HUMAN_1	538
PAX7_HUMAN	539
NKX62_HUMAN	540
ASXL2_HUMAN	541
FOXO1_HUMAN	542
GATA3_HUMAN	543
GATA1_HUMAN	544
ZMYM5_HUMAN	545
ZN783_HUMAN	546
SPI2B_HUMAN_2	547
LRP1_HUMAN	548
MIXL1_HUMAN	549
SGT1_HUMAN	550
LMCD1_HUMAN	551
CEBPA_HUMAN	552
GATA2_HUMAN	553
SOX14_HUMAN	554
WTIP_HUMAN	555
PRP19_HUMAN	556
CBX6_HUMAN	557
NKX11_HUMAN	558
RBBP4_HUMAN	559
DMRT2_HUMAN	560
SMCA2_HUMAN	561
ZNF10_HUMAN	562
EED_HUMAN	563
RCOR1_HUMAN	564

[0175] 本公开涵盖任何上述蛋白质的功能类似物,即具有相同或基本上相同的生物学功能(例如,保留70%或更多、80%或更多、90%或更多、95%或更多或98%或更多的蛋白质的转录因子功能)的分子。例如,功能类似物可以是上面列出的蛋白质的同工型或变体,例如含有具有或不具有额外的氨基酸残基的上述蛋白质的一部分和/或含有相对于上述蛋白质的突变。在一些实施方案中,功能类似物与表4中列出的序列之一具有至少75%、80%、85%、90%、95%、98%或99%的序列同一性。还考虑了上述蛋白质的同源物、直系同源物和突变体。

[0176] 在某些实施方案中,本文所述的表观遗传编辑器包含衍生自KOX1、ZIM3、ZFP28或ZN627的KRAB结构域,和/或衍生自KAP1、MECP2、HP1a、HP1b、CBX8、CDYL2、TOX、TOX3、TOX4、EED、EZH2、RBBP4、RCOR1或SCML2的效应结构域,任选地其中亲本蛋白是人蛋白。在特定的实施方案中,本文所述的表观遗传编辑器包含衍生自KOX1、ZIM3、ZFP28和/或ZN627的结构域,任选地其中亲本蛋白是人蛋白。在某些实施方案中,表观遗传编辑器可以包含衍生自KOX1(ZNF10)(例如人KOX1)的KRAB结构域。在某些实施方案中,表观遗传编辑器可以包含衍生自

ZIM3(ZNF657或ZNF264)(例如人ZIM3)的KRAB结构域。在某些实施方案中,表观遗传编辑器可以包含衍生自ZFP28(例如人ZFP28)的KRAB结构域。在某些实施方案中,表观遗传编辑器可以包含衍生自ZN627(例如人ZN627)的KRAB结构域。在某些实施方案中,本文所述的表观遗传编辑器可以包含CDYL2(例如,人CDYL2)和/或与KOX1 KRAB结构域(例如,人KOX1 KRAB结构域)组合的TOX结构域(例如,人TOX结构域)。

[0177] 在某些实施方案中,本文所述的表观遗传效应子包含衍生自KOX1/ZNF10(SEQ ID NO:89)的阻遏结构域。例如,阻遏结构域可以包含SEQ ID NO:89的序列,或与SEQ ID NO:89具有至少75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%同一性的序列。

[0178] 在某些实施方案中,本文所述的表观遗传效应子包含衍生自KOX1/ZNF10的阻遏结构域,如下表5所示:

[0179] 表5.衍生自KOX1/ZNF10的示例性效应结构域

[0180]

蛋白	蛋白序列
KOX1/ZNF10 KRAB 1	SEQ ID NO: 565
KOX1/ZNF10 KRAB 2	SEQ ID NO: 566
KOX1/ZNF10 KRAB 3	SEQ ID NO: 567
KOX1/ZNF10 (aa 11-72)	SEQ ID NO: 568
KOX1/ZNF10 (aa 11-108)	SEQ ID NO: 569

[0181]

蛋白	蛋白序列
KOX1/ZNF10 变体	SEQ ID NO: 570
KOX1 KRAB-ZIM3 嵌合体	SEQ ID NO: 571
ZIM3-KOX1 KRAB 嵌合体	SEQ ID NO: 572

[0182] 在特定的实施方案中,阻遏结构域可以包含SEQ ID NO:565的氨基酸序列,或与SEQ ID NO:565至少75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同的序列。

[0183] 在特定的实施方案中,阻遏结构域可以包含SEQ ID NO:566的氨基酸序列,或与SEQ ID NO:566至少75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同的序列。

[0184] 在特定的实施方案中,阻遏结构域可以包含SEQ ID NO:567的氨基酸序列,或与SEQ ID NO:567至少75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同的序列。

[0185] 在特定的实施方案中,阻遏结构域可以包含SEQ ID NO:568的氨基酸序列,或与SEQ ID NO:568至少75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同的序列。

[0186] 在特定的实施方案中,阻遏结构域可以包含SEQ ID NO:569的氨基酸序列,或与SEQ ID NO:569至少75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同的序列。

[0187] 在特定的实施方案中,阻遏结构域可以包含SEQ ID NO:570的氨基酸序列,或与SEQ ID NO:570至少75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同的序列。

[0188] 在特定的实施方案中,阻遏结构域可以包含SEQ ID NO:571的氨基酸序列,或与SEQ ID NO:571至少75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同的序列。

[0189] 在特定的实施方案中,阻遏结构域可以包含SEQ ID NO:572的氨基酸序列,或与SEQ ID NO:572至少75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同的序列。

[0190] B.DNA甲基转移酶

[0191] 在一些实施方案中,本文所述的表观遗传编辑器的效应结构域通过DNA修饰如甲基化改变靶基因表达。DNA的高度甲基化区域往往比低度甲基化区域的转录活性低。DNA甲基化主要发生在CpG位点("C-磷酸-G-"或"胞嘧啶-磷酸-鸟嘌呤"位点的简写)。许多哺乳动物基因具有靠近或包括CpG岛(具有高频率CpG二核苷酸的核酸区域)的启动子区。

[0192] 本文所述的效应结构域可以是例如DNA甲基转移酶(DNMT)或其催化结构域,或者可以能够招募DNA甲基转移酶。DNMT涵盖催化甲基转移至DNA核苷酸的酶,例如催化甲基添加至基因组DNA的典型胞嘧啶-5DNMT(例如,DNMT1、DNMT3A、DNMT3B和DNMT3C)。该术语还涵盖本身不催化甲基化但招募(包括激活)催化活性DNMT的非典型家族成员;这种DNMT的非限制性示例是DNMT3L。参见例如Lyko,Nat Review(2018)19:81-92。除非另有说明,否则DNMT结构域可以指衍生自催化活性DNMT(例如,DNMT1、DNMT3A和DNMT3B)或衍生自非催化活性DNMT(例如,DNMT3L)的多肽结构域。DNMT可以通过招募抑制性调节蛋白来抑制靶基因的表达。在一些实施方案中,甲基化在CG(或CpG)二核苷酸序列处。在一些实施方案中,甲基化在CHG或CHH序列处,其中H是A、T或C中的任一个。

[0193] 在一些实施方案中,本文所述的DNMT可以是动物DNMT(例如,哺乳动物DNMT)、植物DNMT、真菌DNMT或细菌DNMT。细菌DNMT可以获自细菌物种(例如球菌属细菌、芽孢杆菌属细菌、螺旋细菌或细胞内革兰氏阳性或革兰氏阴性细菌)。在某些实施方案中,细菌物种是支原体目细菌(Mycoplasmatales bacterium)、海洋支原体(Mycoplasma marinum)或中华螺原体(Spiroplasma chinense)。在某些实施方案中,细菌物种不是穿透支原体(M.penetrans)、S.monbiae、副流感嗜血杆菌(H.parainfluenzae)、藤黄节杆菌(A.luteus)、埃及嗜血杆菌(H.aegyptius)、溶血嗜血杆菌(H.haemolyticus)、莫拉氏菌(Moraxella)、大肠杆菌(E.coli)、水生栖热菌(T.aquaticus)、新月梭菌(C.crescentus)或艰难梭菌(C.difficile)。在某些实施方案中,本文所述的表观遗传编辑器包含DNMT结构域,所述DNMT结构域包含SEQ ID NO:601或与SEQ ID NO:601至少75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同的序列。在某些实施方案中,本文所述的表观遗传编辑器包含DNMT结构域,所述DNMT结构域包含SEQ ID NO:602或与SEQ ID NO:602至少75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同的序列。在某些实施方案中,本文所述的表观遗传编辑器包含DNMT结构域,所述DNMT结构域包含SEQ ID NO:603或与SEQ ID NO:603至少75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同的序列。

[0194] 在某些实施方案中,本文所述的表观遗传编辑器中的DNMT可以包括例如DNMT1、DNMT3A、DNMT3B和/或DNMT3C。在一些实施方案中,DNMT是哺乳动物(例如,人或鼠)DNMT。在特定的实施方案中,DNMT是DNMT3A(例如,人DNMT3A)。在某些实施方案中,本文所述的表观遗传编辑器包含DNMT3A结构域,所述DNMT3A结构域包含SEQ ID NO:574或与SEQ ID NO:574至少75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同的序列。在某些实施方案中,本文所述的表观遗传编辑器包含DNMT3A结构域,所述DNMT3A结构域包含SEQ ID NO:575或与SEQ ID NO:575至少75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同的序列。在一些实施方案中,DNMT3A结构域可以具有例如在位置H739(如H739A或H739E)、R771(如R771L)和/或R836(如R836A或R836Q)或其任何组合处的突变(根据SEQ ID NO:574编号)。

[0195] 在一些实施方案中,本文所述的效应结构域可以是DNMT样结构域。如本文所用,“DNMT样结构域”是可以激活或招募其他DNMT结构域,但本身不具有甲基化活性的DNMT的调节因子。在一些实施方案中,DNMT样结构域是哺乳动物(例如,人或小鼠)DNMT样结构域。在某些实施方案中,DNMT样结构域是DNMT3L,其可以是例如人DNMT3L或小鼠DNMT3L。在某些实施方案中,本文所述的表观遗传编辑器包含DNMT3L结构域,所述DNMT3L结构域包含SEQ ID NO:578或与SEQ ID NO:578至少75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同的序列。在某些实施方案中,本文的表观遗传编辑器包含DNMT3L结构域,所述DNMT3L结构域包含SEQ ID NO:579,或与SEQ ID NO:579至少75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同的序列。在某些实施方案中,本文所述的表观遗传编辑器包含DNMT3L结构域,所述DNMT3L结构域包含SEQ ID NO:580或与SEQ ID NO:580至少75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同的序列。在某些实施方案中,本文所述的表观遗传编辑器包含DNMT3L结构域,所述DNMT3L结构域包含SEQ ID NO:581,或与SEQ ID NO:581至少75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同的序列。在一些实施方案中,DNMT3L结构域可以具有例如对应于位置D226(如D226V)、Q268(如Q268K)或两者处的突变(根据SEQ ID NO:578编号)。

[0196] 在某些实施方案中,本文的表观遗传编辑器可以包括包含DNMT和DNMT样效应结构域两者。例如,表观遗传编辑器可以包含DNMT3A-3L结构域,其中DNMT3A和DNMT3L可以共价连接。在其他实施方案中,本文所述的表观遗传编辑器可以包含效应结构域,其仅包含DNMT3A结构域(例如,人DNMT3A),或仅包含DNMT样结构域(例如,DNMT3L,其可以是人或小鼠DNMT3L)。

[0197] 下表6提供了示例性DNMT,其可以是本文所述的表观遗传效应结构域的一部分,或者可以从中衍生本文所述的表观遗传编辑器的效应结构域。

[0198] 表6.示例性DNMT序列

[0199]

蛋白名称	物种	靶标	蛋白序列
DNMT1	人	5mC	SEQ ID NO: 573
DNMT3A (h3A)	人	5mC	SEQ ID NO: 574
DNMT3A (催化结构域) (h3As)	人	5mC	SEQ ID NO: 575
DNMT3B	人	5mC	SEQ ID NO: 576
DNMT3C	小鼠	5mC	SEQ ID NO: 577
DNMT3L (h3L)	人	5mC	SEQ ID NO: 578
DNMT3L (催化结构域) (h3Ls)	人	5mC	SEQ ID NO: 579

[0200]

蛋白名称	物种	靶标	蛋白序列
DNMT3L (m3L)	小鼠	5mC	SEQ ID NO: 580
DNMT3L (催化结构域) (m3Ls)	小鼠	5mC	SEQ ID NO: 581
DNMT3L	大熊猫（*Ailuropoda melanoleuca*）	5mC	SEQ ID NO: 582
DNMT3L (催化结构域)	大熊猫	5mC	SEQ ID NO: 583
DNMT3L	眼镜猴（*Carlito syrichta*）	5mC	SEQ ID NO: 584
DNMT3L (催化结构域)	眼镜猴	5mC	SEQ ID NO: 585
DNMT3L	长爪沙鼠（*Meriones unguiculatus*）	5mC	SEQ ID NO: 586
DNMT3L (催化结构域)	长爪沙鼠	5mC	SEQ ID NO: 587
DNMT3L	北美鼠兔（*Ochotona princeps*）	5mC	SEQ ID NO: 588
DNMT3L (催化结构域)	北美鼠兔	5mC	SEQ ID NO: 589
DNMT3L	新东部灰松鼠（*Neosciurus carolinensis*）	5mC	SEQ ID NO: 590
DNMT3L (催化结构域)	新东部灰松鼠	5mC	SEQ ID NO: 591
DNMT3L	北美野牛（*Bison bison*）	5mC	SEQ ID NO: 592
DNMT3L (催化结构域)	北美野牛	5mC	SEQ ID NO: 593
DNMT3L	普氏野马（*Equus przewalskii*）	5mC	SEQ ID NO: 594
DNMT3L (催化结构域)	普氏野马	5mC	SEQ ID NO: 595
DNMT3L	卡氏小鼠（*Mus caroli*）	5mC	SEQ ID NO: 596
DNMT3L (催化结构域)	卡氏小鼠	5mC	SEQ ID NO: 597
DNMT3L	黑猩猩（*Pan troglodytes*）	5mC	SEQ ID NO: 598
DNMT3L (催化结构域)	黑猩猩	5mC	SEQ ID NO: 599
TRDMT1 (DNMT2)	人	tRNA 5mC	SEQ ID NO: 600
DNA 胞嘧啶甲基转移酶	支原体目细菌	5mC	SEQ ID NO: 601
DNA 胞嘧啶	支原体目细菌	5mC	SEQ ID NO: 602

[0201]

蛋白名称	物种	靶标	蛋白序列
甲基转移酶			
DNA (胞嘧啶-5-)-甲基转移酶	中华螺原体	5mC	SEQ ID NO: 603
M.MpeI	穿透支原体	5mC	SEQ ID NO: 604
M.SssI	单体螺原体（*Spiroplasma monobiae*）	5mC	SEQ ID NO: 605
M.HpaII	副流感嗜血杆菌	5mC (CCGG)	SEQ ID NO: 606
M.AluI	藤黄节杆菌	5mC (AGCT)	SEQ ID NO: 607
M.HaeIII	埃及嗜血杆菌（*Haemophilus aegyptius*）	5mC (GGCC)	SEQ ID NO: 608
M.HhaI	溶血嗜血杆菌（*Haemophilus haemolyticus*）	5mC (GCGC)	SEQ ID NO: 609
M.MspI	莫拉氏菌	5mC (CCGG)	SEQ ID NO: 610
Masc1	粪盘菌（*Ascobolus*）	5mC	SEQ ID NO: 611
MET1	拟南芥（*Arabidopsis*）	5mC	SEQ ID NO: 612
Masc2	粪盘菌	5mC	SEQ ID NO: 613
Dim-2	链孢霉（*Neurospora*）	5mC	SEQ ID NO: 614
dDnmt2	果蝇（*Drosophila*）	5mC	SEQ ID NO: 615
Pmt1	裂殖酵母（*S. pombe*）	5mC	SEQ ID NO: 616
DRM1	拟南芥	5mC	SEQ ID NO: 617
DRM2	拟南芥	5mC	SEQ ID NO: 618
CMT1	拟南芥	5mC	SEQ ID NO: 619
CMT2	拟南芥	5mC	SEQ ID NO: 620
CMT3	拟南芥	5mC	SEQ ID NO: 621
Rid	链孢霉	5mC	SEQ ID NO: 622
hsdM gene	细菌 (大肠杆菌, 菌株 12)	m6A	SEQ ID NO: 623
hsdS gene	细菌(大肠杆菌, 菌株 12)	m6A	SEQ ID NO: 624
M.TaqI	细菌(水生栖热菌)	m6A	SEQ ID NO: 625
M.EcoDam	大肠杆菌	m6A	SEQ ID NO: 626
M.CcrMI	新月梭菌	m6A	SEQ ID NO: 627
CamA	艰难梭菌	m6A	SEQ ID NO: 628

[0202] 本公开涵盖任何上述蛋白质的功能类似物,即具有相同或基本上相同的生物学功能(例如,保留70%或更多、80%或更多、90%或更多、95%或更多、或98%或更多的蛋白质的DNA甲基化功能或招募功能)的分子。例如,功能类似物可以是上述蛋白质的同工型或变体,例如含有具有或不具有额外的氨基酸残基的上述蛋白质的一部分和/或含有相对于上述蛋白质的突变。在一些实施方案中,功能类似物与表6中列出的序列之一具有至少75%、80%、85%、90%、95%、98%或99%的序列同一性。在一些实施方案中,本文的效应结构域

仅包含上述蛋白质的功能结构域(或其功能类似物),例如催化结构域或招募结构域。在一些实施方案中,本文的效应结构域包含选自表6的一个或多个表观遗传效应结构域,或其功能同源物、直系同源物或变体。

[0203] 如本文所用,DNMT结构域(例如,DNMT3A结构域或DNMT3L结构域)是指与亲本蛋白(例如,人或鼠DNMT3A或DNMT3L)相同的蛋白质结构域或其功能类似物(例如,具有亲本蛋白的功能片段,如催化片段或招募片段;和/或具有改善DNMT蛋白活性的突变)。

[0204] 本文的表观遗传编辑器可以在靶基因或染色体中的例如1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900或1000个或更多个CpG二核苷酸序列处实现甲基化。CpG二核苷酸序列可以位于CpG岛中的靶基因内或附近,或者可以位于不是CpG岛的区域中。CpG岛通常是指包含高频率CpG二核苷酸的核酸序列或染色体区域。例如,CpG岛可包含至少50% GC含量。CpG岛可以具有高的观察与预期CpG比率,例如至少60%的观察与预期CpG比率。如本文所用,观察与预期CpG比率由CpG数量*(序列长度)/(C数量*G数量)确定。在一些实施方案中,CpG岛具有至少60%、70%、80%、90%或更高的观察与预期CpG比率。CpG岛可以是例如至少200、250、300、350、400、450、500、550、600、650、700、750或800个核苷酸的序列或区域。在一些实施方案中,仅1个或少于2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40或50个CpG二核苷酸被表观遗传编辑器甲基化。

[0205] 在一些实施方案中,本文的表观遗传编辑器在低甲基化核酸序列(即,与标准对照相比,在5-甲基胞嘧啶核苷酸上(例如,在CpG中)可能缺乏甲基的序列)处实现甲基化。例如,相对于年轻细胞或非癌细胞,低甲基化可分别发生在衰老细胞或癌症(例如,瘤形成的早期阶段)中。

[0206] 在一些实施方案中,本文所述的表观遗传编辑器在超甲基化核酸序列处诱导甲基化。

[0207] 在一些实施方案中,甲基化可以由表观遗传编辑器在CpG二核苷酸以外的位点引入。例如,靶基因序列可以在CpA、CpT或CpC序列的C核苷酸处甲基化。在一些实施方案中,表观遗传编辑器包含DNMT3A结构域并在CpG、CpA、CpT、CpC序列或其任何组合处实现甲基化。在一些实施方案中,表观遗传编辑器包含缺乏调节亚结构域且仅维持催化结构域的DNMT3A结构域。在一些实施方案中,包含DNMT3A催化结构域的表观遗传编辑器仅在CpG序列处实现甲基化。在一些实施方案中,与包含野生型DNMT3A结构域的表观遗传编辑器相比,包含含有突变(例如R836A或R836Q突变(根据SEQ ID NO:574编号))的DNMT3A结构域的表观遗传编辑器在CpA、CpC和/或CpT序列处具有更高的甲基化活性。

[0208] C.组蛋白修饰剂

[0209] 在一些实施方案中,本文表观遗传编辑器的效应结构域介导组蛋白修饰。组蛋白修饰在基因转录中发挥结构和生化作用,例如通过形成或破坏与组蛋白结合的核小体结构并阻止基因转录。组蛋白修饰可以包括例如乙酰化、去乙酰化、甲基化、磷酸化、泛素化、SUMO化等,例如在其N-末端(“组蛋白尾”)。这些修饰维持或特异性转化染色质结构,从而控制染色体DNA上发生的如基因表达、DNA复制、DNA修复等反应。组蛋白的翻译后修饰是一种表观遗传调控机制,被认为是真核细胞遗传调控所必需的。最近的研究表明,染色质重塑因子如SWI/SNF、RSC、NURF、NRD等,其通过修饰核小体结构促进转录因子接近DNA;组蛋白乙酰

转移酶(HAT),其调节组蛋白乙酰化状态;和组蛋白去乙酰酶(HDAC),作为重要的调节因子。

[0210] 特别地,组蛋白的非结构化N-末端可以通过乙酰化、去乙酰化、甲基化、泛素化、磷酸化、SUMO化、核糖基化、瓜氨酸化、O-GlcNAc化、巴豆酰化或其任何组合进行修饰。例如,组蛋白乙酰转移酶(HAT)利用乙酰辅酶A作为辅因子并催化乙酰基向赖氨酸侧链的ε氨基的转移。这中和了赖氨酸的正电荷,削弱了组蛋白和DNA之间的相互作用,从而打开了染色体以使转录因子结合并启动转录。组蛋白乙酰转移酶对组蛋白H3的K14和K9赖氨酸的乙酰化可能与人的转录能力有关。赖氨酸乙酰化可以直接或间接产生调节转录激活的染色质修饰酶的结合位点。另一方面,组蛋白H3赖氨酸9的组蛋白甲基化可能与异染色质或转录沉默染色质相关。

[0211] 在某些实施方案中,本文所述的表观遗传编辑器的效应结构域包含组蛋白甲基转移酶结构域。效应结构域可以包含例如DOT1L结构域、SET结构域、SUV39H1结构域、G9a/EHMT2蛋白结构域、EZH1结构域、EZH2结构域、SETDB1结构域或其任何组合。在特定的实施方案中,效应结构域包含组蛋白-赖氨酸-N-甲基转移酶SETDB1结构域。

[0212] 在一些实施方案中,效应结构域包含组蛋白去乙酰酶蛋白结构域。在某些实施方案中,效应结构域包含HDAC家族蛋白结构域,例如HDAC1、HDAC3、HDAC5、HDAC7或HDAC9蛋白结构域。在特定的实施方案中,效应结构域包含核小体重塑和去乙酰酶复合物(NURD),其从组蛋白去除乙酰基。

[0213] D.其他效应结构域

[0214] 在一些实施方案中,效应结构域包含含有三部分基序的蛋白质(TRIM28、TIF1-β或KAP1)。在某些实施方案中,效应结构域包含一种或多种KAP1蛋白。本文表观遗传编辑器中的KAP1蛋白可以与表观遗传编辑器的一个或多个其他效应结构域或参与细胞环境中基因表达调节的一种或多种蛋白形成复合物。例如,KAP1可以被转录阻遏物的KRAB结构域招募。KAP1蛋白结构域可以与降低或沉默基因表达的一种或多种蛋白复合物相互作用或招募这些蛋白复合物。在一些实施方案中,KAP1与组蛋白去乙酰酶蛋白、组蛋白-赖氨酸甲基转移酶蛋白、染色质重塑蛋白和/或异染色质蛋白相互作用或招募这些蛋白。例如,KAP1蛋白结构域可以与异染色质蛋白1(HP1)蛋白、SETDB1蛋白、HDAC蛋白和/或NuRD蛋白复合物组分相互作用或招募这些蛋白。在一些实施方案中,KAP1蛋白结构域与ZFP90蛋白(例如,ZFP90的同种型2)和/或FOXP3蛋白相互作用或招募这些蛋白。示例性KAP1氨基酸序列如SEQ ID NO:629所示。

[0215] 在一些实施方案中,效应结构域包含与一个或多个DNA表观遗传标记相互作用或被一个或多个DNA表观遗传标记招募的蛋白质结构域。例如,效应结构域可以包含与靶基因中的甲基化DNA核苷酸(其可以在或可以不在靶基因的CpG岛处)相互作用的甲基CpG结合蛋白2(MECP2)蛋白。本文所述的表观遗传编辑器中的MECP2蛋白结构域可以诱导浓缩的染色质结构,从而降低或沉默靶基因的表达。在一些实施方案中,本文所述的表观遗传编辑器中的MECP2蛋白结构域可以与组蛋白去乙酰酶(例如HDAC)相互作用,从而抑制或沉默靶基因的表达。在一些实施方案中,本文所述的表观遗传编辑器中的MECP2蛋白结构域可以阻断转录因子或转录激活因子接近靶序列,从而抑制或沉默靶基因的表达。示例性MECP2氨基酸序列如SEQ ID NO:630所示。

[0216] 还考虑作为本文所述的表观遗传编辑器的效应结构域的是,例如,chromoshadow

结构域、泛素-2样Rad60 SUMO样(Rad60-SLD/SUMO)结构域、染色质组织修饰结构域(Chromo)结构域、Yaf2/RYBP C末端结合基序结构域(YAF2_RYBP)、CBX家族C末端基序结构域(CBX7_C)、锌指C3HC4型(环指)结构域(ZF-C3HC4_2)、细胞色素b5结构域(Cyt-b5)、螺旋-环-螺旋结构域(HLH)、螺旋-发夹-螺旋基序结构域(例如HHH_3)、高迁移率基团盒结构域(HMG-box)、碱性亮氨酸拉链结构域(例如bZIP_1或bZIP_2)、Myb_DNA结合结构域、同源异型结构域、具有FCS序列的MYM型锌指结构域(ZF-FCS)、干扰素调节因子2结合蛋白锌指结构域(IRF-2BP1_2)、SSX阻遏结构域(SSXRD)、B盒型锌指结构域(ZF-B_box)、CXXC锌指结构域(ZF-CXXC)、染色体缩合1结构域(RCC1)、SRC同源3结构域(SH3_9)、不育α基序结构域(SAM_1)、不育α基序结构域(SAM_2)、不育α基序/尖峰结构域(SAM_PNT)、Vestigial/Tondu家族结构域(Vg_Tdu)、LIM结构域、RNA识别基序结构域(RRM_1)、配对两亲性螺旋结构域(PAH)、蛋白酶体ATPase OB C末端结构域(Prot_ATP_ID_OB)、神经同源性2结构域(NHR2)、裂解刺激因子亚基2的铰链结构域(CSTF2_hinge)、PPAR gamma N末端区域结构域(PPARgamma_N)、CDC48 N末端结构域(CDC48_2)、WD40重复结构域(WD40)、Fip1基序结构域(Fip1)、PDZ结构域(PDZ_6)、血管性血友病因子C型结构域(VWC)、NAB保守区1结构域(NCD1)、S1 RNA结合结构域(S1)、HNF3 C末端结构域(HNF_C)、Tudor结构域(Tudor_2)、组蛋白样转录因子(CBF/NF-Y)和古细菌组蛋白结构域(CBFD_NFYB_HMF)、锌指蛋白结构域(DUF3669)、EGF样结构域(cEGF)、GATA锌指结构域(GATA)、TEA/ATTS结构域(TEA)、佛波醇酯/二酰甘油结合结构域(C1-1)、多梳状MTF2因子2结构域(Mtf2_C)、FOXO蛋白家族的转录激活结构域(FOXO-TAD)、同源异型盒KN结构域(Homeobox_KN)、BED锌指结构域(ZF-BED)、C3HC4型锌指的RING结构域(ZF-C3HC4_4)、RAD51相互作用基序结构域(RAD51_interact)、N乙基-CpG结合结构域蛋白MBD的p55结合区(MBDa)、Notch结构域、Raf样Ras结合结构域(RBD)、Spin/Ssty家族结构域(Spin-Ssty)、PHD指结构域(PHD_3)、低密度脂蛋白受体结构域A类(Ldl_recept_a)、CS结构域、DM DNA结合结构域和QLQ结构域。

[0217] 在一些实施方案中,效应结构域是包含YAF2_RYBP结构域或同源异型结构域或其任何组合的蛋白质结构域。在某些实施方案中,YAF2_RYBP结构域的同源异型结构域是PRD结构域、NKL结构域、HOXL结构域或LIM结构域。在特定的实施方案中,YAF2_RYBP结构域可以包含32个氨基酸的Yaf2/RYBP C末端结合基序结构域(32aa RYBP)。

[0218] 在一些实施方案中,效应结构域包含选自由以下组成的组的蛋白质结构域:SUMO3结构域、来自M期磷蛋白8(MPP8)的染色质结构域、来自Chromobox 1(CBX1)的chromoshadow结构域和来自Scm多梳组蛋白同源物1(SCMH1)的SAM_1/SPM结构域。

[0219] 在一些实施方案中,效应结构域包含HNF3 C末端结构域(HNF_C)。HNF_C结构域可以来自FOXA1或FOXA2。在某些实施方案中,HNF_C结构域包含EH1(engrailed同源性1)基序。

[0220] 在一些实施方案中,效应结构域可以包含干扰素调节因子2结合蛋白锌指结构域(IRF-2BP1_2)、来自DNA修复因子HERC2 E3连接酶的Cyt-b5结构域、来自桥接整合子1(BIN1)的变体SH3结构域(SH3_9)、来自转录因子TOX的HMG-box结构域或来自多梳组分PCGF2的ZF-C3HC4_2环指结构域、染色质结构域-解旋酶-DNA结合蛋白3(CHD3)结构域或ZNF783结构域。

[0221] IV.表观遗传编辑器

[0222] 本文提供了表观遗传编辑器(即,表观遗传编辑系统),其例如使用如本文所述的

一个或多个DNA结合结构域和如本文所述的一个或多个效应结构域(例如,表观遗传阻遏结构域)以任何组合将表观遗传修饰引导至感兴趣基因中的靶序列。DNA结合结构域(与如本文所述的指导多核苷酸协同,其中DNA结合结构域是多核苷酸指导的DNA结合结构域)指导效应结构域表观遗传修饰靶序列,导致基因抑制或沉默,其可以是持久的并且可跨细胞世代遗传。在一些方面,本文所述的表观遗传编辑器可以可逆地或不可逆地抑制或沉默细胞中的基因。

[0223] 在特定的实施方案中,本文所述的表观遗传编辑器包含一种或多种融合蛋白,各自包含(1)DNA结合结构域和(2)效应结构域。效应结构域可以在表观遗传编辑器包含的一种或多种融合蛋白上。例如,单个融合蛋白可以包含具有DNA结合结构域的所有效应结构域。或者,效应结构域或其子集可以在单独的融合蛋白上,每个融合蛋白具有DNA结合结构域(其可以相同或不同)。本文所述的融合蛋白可进一步包含一个或多个接头(例如,肽接头)、可检测标签、核定位信号(NLS)或其任何组合。如本文所用,"融合蛋白"是指其中两个或更多个编码序列(例如,DNA结合结构域和/或效应结构域)直接或间接共价或非共价连接的嵌合蛋白。

[0224] 在一些实施方案中,本文所述的表观遗传编辑器包含2、3、4、5、6、7、8、9、10个或更多个效应(例如,抑制/阻遏)结构域,其可以是相同的或不同的。在某些实施方案中,两个或更多个所述效应结构域协同发挥作用。效应结构域的组合可以包含DNA甲基化结构域、组蛋白去乙酰化结构域、组蛋白甲基化结构域和/或招募上述任一种的支架结构域。例如,本文所述的表观遗传编辑器可以包含一个或多个转录阻遏结构域(例如KRAB结构域如KOX1、ZIM3、ZFP28或ZN627 KRAB)与一个或多个DNA甲基化结构域(例如DNMT结构域)和/或招募结构域(例如DNMT3L结构域)的组合。这种表观遗传编辑器可以包含例如KRAB结构域、DNMT3A结构域和DNMT3L结构域。在一些实施方案中,表观遗传编辑器进一步包含额外的效应结构域(例如,KAP1、MECP2、HP1b、CBX8、CDYL2、TOX、TOX3、TOX4、EED、RBBP4、RCOR1或SCML2结构域)。在一些实施方案中,额外的效应结构域是CDYL2、TOX、TOX3、TOX4或HP1a结构域。例如,本文所述的表观遗传编辑器可以包含CDYL2和/或TOX结构域与KRAB结构域(例如,KOX1 KRAB结构域)的组合。

[0225] A.接头

[0226] 如本文所述的融合蛋白可以包含连接表观遗传编辑器的组分的一个或多个接头。接头可以是肽或非肽接头。

[0227] 在一些实施方案中,在本文提供的表观遗传编辑器中使用的一个或多个接头是肽接头,即包含肽部分的接头。肽接头可以是适用于本文所述的表观遗传编辑器融合蛋白的任何长度。在一些实施方案中,接头可包含1至200(例如,1至80)个氨基酸的肽。在一些实施方案中,接头包含长度为1至5、1至10、1至20、1至30、1至40、1至50、1至60、1至80、1至100、1至150、1至200、5至10、5至20、5至30、5至40、5至60、5至80、5至100、5至150、5至200、10至20、10至30、10至40、10至50、10至60、10至80、10至100、10至150、10至200、20至30、20至40、20至50、20至60、20至80、20至100、20至150、20至200、30至40、30至50、30至60、30至80、30至100、30至150、30至200、40至50、40至60、40至80、40至100、40至150、40至200、50至60、50至80、50至100、50至150、50至200、60至80、60至100、60至150、60至200、80至100、80至150、80至200、100至150、100至200或150至200个氨基酸。还可考虑更长或更短的接头。在一些实施方案

中,肽接头的长度为3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、25、40、45、50、55、60、65、70、75、80、85、90、95或100个氨基酸。例如,肽接头的长度可以是4、5、16、20、24、27、32、40、64、92或104个氨基酸。肽接头可以是柔性或刚性接头。在特定的实施方案中,肽接头包含SEQ ID NO:631-637和664-665中任一个的氨基酸序列或与其至少75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同的序列。

[0228] 在某些实施方案中,肽接头是XTEN接头。此类接头可以包含XTEN序列的一部分(Schellenberger et al.,Nat Biotechnol(2009)27(1):1186-90),其为仅由残基G、S、P、T、E和A组成的非结构化亲水性多肽。如本文所用,术语"XTEN"是指缺乏疏水性氨基酸残基的重组肽或多肽。XTEN接头通常是非结构化的并且包含一组有限的天然氨基酸。XTEN与蛋白质的融合改变了其流体动力学性质并降低了融合蛋白的清除和降解速率。这些XTEN融合蛋白使用重组技术产生,不需要化学修饰,并通过天然途径降解。XTEN接头的长度可以是例如5、10、16、20、26或80个氨基酸。在一些实施方案中,XTEN接头的长度为16个氨基酸。在一些实施方案中,XTEN接头的长度为80个氨基酸。在某些实施方案中,XTEN接头可以是XTEN10、XTEN16、XTEN20或XTEN80。在某些实施方案中,XTEN接头可以包含SEQ ID NO:638-643中任一个的氨基酸序列或与其至少75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同的序列。在特定的实施方案中,XTEN接头包含SEQ ID NO:638的氨基酸序列。在特定的实施方案中,XTEN接头包含SEQ ID NO:643的氨基酸序列。

[0229] 在一些实施方案中,在本文提供的表观遗传编辑器中使用的一个或多个接头是非肽接头。例如,接头可以是碳键、二硫键或碳-杂原子键。在某些实施方案中,接头是酰胺连接键的碳-氮键。在某些实施方案中,接头是环状或非环状、取代或未取代、或支链或非支链脂肪族或杂脂肪族接头。

[0230] 在一些实施方案中,在本文提供的表观遗传编辑器中使用的一个或多个接头是聚合的(例如,聚乙烯、聚乙二醇、聚酰胺、聚酯等)。接头可以包含例如氨基烷酸的单体、二聚体或聚合物;氨基烷酸(例如甘氨酸、乙酸、丙氨酸、β-丙氨酸、3-氨基丙酸、4-氨基丁酸、5-戊酸等);氨基己酸(Ahx)的单体、二聚体或聚合物;或聚乙二醇部分(PEG);或芳基或杂芳基部分。在某些实施方案中,接头可以基于碳环部分(例如,环戊烷或环己烷)或苯环。接头可以包括官能化部分以促进亲核试剂(例如硫醇、氨基)从肽附接到接头。任何亲电试剂都可以用作接头的一部分。示例性亲电试剂包括但不限于活化酯、活化酰胺、烷基卤化物、芳基卤化物、酰基卤化物和异硫氰酸酯。

[0231] 可以在表观遗传编辑器的任何两个组分之间(例如,在效应结构域(例如,阻遏结构域)和DNA结合结构域(例如,Cas9结构域)之间、在第一效应结构域和第二效应结构域之间等)采用各种接头长度和柔性。接头的范围可以从非常柔性的接头(例如富含甘氨酸/丝氨酸的接头)到更刚性的接头,以实现针对特定应用的效应结构域活性的最佳长度。在一些实施方案中,更柔性的接头是富含甘氨酸/丝氨酸的接头(富含GS的接头),其中超过45%(例如,超过48、50、55、60、70、80或90%)的残基是甘氨酸或丝氨酸残基。富含GS的接头的非限制性示例是(GGGGS)n(SEQ ID NO:664)、(G)n和W接头(SEQ ID NO:637)。在一些实施方案中,更刚性的接头呈(EAAAK)n(SEQ ID NO:665)、(SGGS)n(SEQ ID NO:631和(XP)n)的形式。在上述柔性接头和刚性接头的式中,n可以是1至30之间的任何整数。在一些实施方案中,n是1、2、3、4、5、6、7、8、9、10、11、12、13、14或15。在一些实施方案中,接头包含(GGS)n基序,其

中n是1、3或7。在一些实施方案中,接头包含(GGGGS)n基序,其中n是4(SEQ ID NO:636)。

[0232] 在一些实施方案中,本文所述的表观遗传编辑器中的接头包含核定位信号,例如,具有SEQ ID NO:644-649中任一个的氨基酸序列。在一些实施方案中,本文所述的表观遗传编辑器中的接头包含表达标签,例如可检测标签,如绿色荧光蛋白。

[0233] B.核定位信号

[0234] 本文所述的融合蛋白可以包含一个或多个核定位信号,并且在某些实施方案中,可以包含两个或更多个核定位信号。例如,融合蛋白可以包含1、2、3、4、5、6、7、8、9、10个或更多个核定位信号。如本文所用,“核定位信号”(NLS)是将蛋白质导向细胞核的氨基酸序列。在某些实施方案中,NLS可以是SV40 NLS(例如,具有SEQ ID NO:644的氨基酸序列)。融合蛋白可以在其N末端、C末端或两者处包含NLS,和/或NLS可以嵌入融合蛋白的中间(例如,在DNA结合结构域或效应结构域的N或C末端)。

[0235] 在一些实施方案中,融合蛋白可以包含两个NLS。融合蛋白可以在其N末端或C末端包含两个NLS。融合蛋白可以包含位于其N末端的一个NLS和嵌入在融合蛋白中间的一个NLS,或位于其C末端的一个NLS和嵌入在融合蛋白中间的一个NLS。融合蛋白可以包含嵌入在融合蛋白中间的两个NLS。

[0236] 在一些实施方案中,融合蛋白可以包含四个NLS。融合蛋白可以在其N末端或C末端包含至少两个(例如,两个、三个或四个)NLS。融合蛋白可以包含至少一个(例如,一个、两个、三个或四个)嵌入融合蛋白中间的NLS。在特定的实施方案中,融合蛋白可以在其N末端包含两个NLS并且在其C末端包含两个NLS。

[0237] 本文所述的NLS可以是内源性NLS序列。在某些实施方案中,本文所述的NLS包含SEQ ID NO:644-649中任一个的氨基酸序列,或与所选序列至少75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同的序列。在特定的实施方案中,NLS包含SEQ ID NO:644的氨基酸序列。额外的NLS是本领域已知的。

[0238] 在一些实施方案中,与具有在N末端不具有至少一个NLS且在C末端不具有至少一个NLS的相应融合蛋白的表观遗传编辑器相比,包含在N末端包含至少一个NLS且在C末端包含至少一个NLS的融合蛋白的表观遗传编辑器可以将表观遗传编辑器的效率提高至少5%、至少10%、至少15%、至少20%、至少30%、至少40%、至少50%、至少60%、至少70%、至少80%、至少90%、至少100%、至少200%、至少300%、至少400%、至少500%、至少600%、至少700%、至少800%、至少900%、至少1,000%、至少5,000%、至少10,000%、至少50,000%、至少100,000%或更多。

[0239] 在一些实施方案中,与具有在N端不具有两个NLS且在C端不具有两个NLS的相应融合蛋白的表观遗传编辑器相比,包含在N端包含两个NLS且在C端包含两个NLS的融合蛋白的表观遗传编辑器可以将表观遗传编辑器的效率提高至少5%、至少10%、至少15%、至少20%、至少30%、至少40%、至少50%、至少60%、至少70%、至少80%、至少90%、至少100%、至少200%、至少300%、至少400%、至少500%、至少600%、至少700%、至少800%、至少900%、至少1,000%、至少5,000%、至少10,000%、至少50,000%、至少100,000%或更多。

[0240] C.标签

[0241] 本文提供的表观遗传编辑器可以包含用于跟踪、检测和定位编辑器的一个或多个

额外的序列(“标签”)。在一些实施方案中,表观遗传编辑器包含1、2、3、4、5、6、7、8、9、10个或更多个可检测标签。每个可检测标签可以相同或不同。

[0242] 例如,表观遗传编辑器融合蛋白可以包含细胞质定位序列、输出序列(例如核输出序列)或其他定位序列,以及可用于溶解、纯化或检测融合蛋白的序列标签。本文提供的合适的蛋白质标签包括但不限于生物素羧化酶载体蛋白(BCCP)标签、myc标签、钙调蛋白标签、FLAG标签、血凝素(HA)标签、聚组氨酸标签(也称为组氨酸标签或His-标签)、麦芽糖结合蛋白(MBP)标签、nus标签、谷胱甘肽-S-转移酶(GST)标签、绿色荧光蛋白(GFP)标签、硫氧还蛋白标签、S标签、Softag(例如,Softag 1或Softag 3)、strep标签、生物素连接酶标签、FlAsH标签、V5标签和SBP标签。额外的合适序列对于本领域技术人员将是显而易见的。

[0243] D.融合蛋白构型

[0244] 本文所述的表观遗传编辑器的融合蛋白可以具有以不同构型构造的其组分。例如,DNA结合结构域可以在C末端、N末端或在两个或更多个表观遗传效应结构域或额外的结构域之间。在一些实施方案中,DNA结合结构域位于表观遗传编辑器的C末端。在一些实施方案中,DNA结合结构域位于表观遗传编辑器的N末端。在一些实施方案中,DNA结合结构域与一个或多个核定位信号连接。在一些实施方案中,DNA结合结构域在两侧侧接表观遗传效应结构域和/或额外的结构域。在一些实施方案中,其中“DBD”指示DNA结合结构域且“ED”指示效应结构域,表观遗传编辑器包含以下构型:

[0245] -N’]-[ED1]-[DBD]-[ED2]-[C’

[0246] -N’]-[ED1]-[DBD]-[ED2]-[ED3]-[C’

[0247] -N’]-[ED1]-[ED2]-[DBD]-[ED3]-[C’

[0248] 或

[0249] -N’]-[ED1]-[ED2]-DBD]-[ED3]-[ED4]-[C’。

[0250] 在一些实施方案中,表观遗传编辑器包含DNA结合结构域(DBD)、DNA甲基转移酶(DNMT)结构域和抑制或沉默靶基因表达的转录阻遏(“阻遏物”)结构域。DBD、DNMT和转录阻遏结构域可以是如本文所述的以任何组合方式的任何结构域。DBD、DNMT结构域和阻遏结构域可以是任何构型,例如,在融合蛋白的N末端、C末端或中间具有任何所述结构域。在一些实施方案中,表观遗传编辑器包含具有以下构型的融合蛋白:

[0251] N’]-[DNMT结构域]-[DBD]-[阻遏结构域]-[C’

[0252] N’]-[阻遏结构域]-[DBD]-[DNMT结构域]-[C’

[0253] N’]-[DNMT结构域]-[阻遏结构域]-[DBD]-[C’

[0254] 或

[0255] N’]-[阻遏结构域]-[DNMT结构域]-[DBD]-[C’。

[0256] 在一些实施方案中,表观遗传编辑器结构中的任一个中的连接结构“]-[”是接头,例如肽接头;可检测标签;肽键;核定位信号;和/或启动子或调节序列。在表观遗传编辑器结构中,多个连接结构“]-[”可以是相同的,或者可以各自是不同的接头、标签、NLS或肽键。在一些实施方案中,DNMT结构域可以包含表6中的任何一个结构域,或其任何组合或同源物。在特定的实施方案中,DNMT结构域包含DNMT3A或其截短型式、DNMT3L或其截短型式或两者。在特定的实施方案中,DBD是无催化活性的多核苷酸引导的DNA结合结构域(例如,dCas9)或ZFP结构域。在某些实施方案中,阻遏结构域包含表4或5中所示的结构域中的任一

个,或其任何组合或同源物。例如,阻遏结构域可以是KRAB结构域。在某些实施方案中,阻遏结构域是ZFP28、ZN627、KAP1、MeCP2、HP1b、CBX8、CDYL2、TOX、Tox3、Tox4、EED、RBBP4、RCOR1或SCML2结构域,或所述结构域中的两个的融合体(例如,ZIM3和KOX1 KRAB的N-和C-末端区域的融合体)。在特定的实施方案中,阻遏结构域是来自ZFP28、ZN627、ZIM3或KOX1的KRAB结构域。

[0257] 在一些实施方案中,表观遗传编辑器包含选自以下的构型:

[0258] N']-[DNMT3A-DNMT3L]-[DBD]-[阻遏物]-[C']

[0259] N']-[阻遏物]-[DBD]-[DNMT3A-DNMT3L]-[C']

[0260] N']-[repressor]-[DBD]-[DNMT3A]-[C']

[0261] N']-[DNMT3A]-[DBD]-[阻遏物]-[C']

[0262] N']-[阻遏物]-[DBD]-[DNMT3A]-[DNMT3L]-[C']

[0263] N']-[DNMT3A]-[DNMT3L]-[DBD]-[阻遏物]-[C']

[0264] N']-[DNMT3A]-[DBD]-[C']

[0265] N']-[DBD]-[DNMT3A]-[C']

[0266] N']-[DNMT3L]-[DBD]-[C']

[0267] N']-[DBD]-[DNMT3L]-[C']

[0268] 其中[DNMT3A-DNMT3L]指示DNMT3A和DNMT3L结构域经由肽键直接融合,并且其中连接结构]-[是如本文所述的接头、可检测标签、亲和结构域、肽键、核定位信号、启动子和/或调节序列中的任一种。DBD、阻遏物、DNMT3A和DNMT3L结构域可以是如本文所述的以任何组合方式的任何结构域。例如,DNMT3A和DNMT3L结构域可以选自表6中的那些。在特定的实施方案中,DBD是CRISPR相关蛋白结构域(例如,dCas9)或ZFP结构域;阻遏结构域是衍生自KOX1、ZIM3、ZFP28或ZN627的KRAB结构域;DNMT3A结构域是人DNMT3A结构域;并且DNMT3L结构域是人或小鼠DNMT3L结构域;本公开还考虑了这些组分的任何组合。

[0269] 在一些实施方案中,表观遗传编辑器包含选自以下的构型:

[0270] N']-[DNMT3A]-[DBD]-[SETDB1]-[C']

[0271] N']-[DNMT3A]-[DNMT3L]-[DBD]-[SETDB1]-[C']

[0272] N']-[DNMT3A-DNMT3L]-[DBD]-[SETDB1]-[C']

[0273] N']-[SETDB1]-[DBD]-[DNMT3A]-[DNMT3L]-[C']

[0274] N']-[SETDB1]-[DBD]-[DNMT3A]-[C']

[0275] 其中[DNMT3A-DNMT3L]指示DNMT3A和DNMT3L结构域经由肽键直接融合,并且其中连接结构]-[是如本文所述的接头、可检测标签、亲和结构域、肽键、核定位信号、启动子和/或调节序列中的任一种。DBD、SETDB1、DNMT3A和DNMT3L结构域可以是本文所述的以任何组合方式的任何结构域。在特定的实施方案中,DBD是CRISPR相关蛋白结构域(例如,dCas9)或ZFP结构域;SETDB1结构域衍生自人SETDB1、ZIM3、ZFP28或ZN627;DNMT3A结构域是人DNMT3A结构域;并且DNMT3L结构域是人或小鼠DNMT3L结构域;本公开还考虑这些组分的任何组合。

[0276] 本文考虑的特定构建体包括:

[0277] DNMT3A-DNMT3L-XTEN80-NLS-dCas9-NLS-XTEN16-KOX1KRAB(构型1),

[0278] DNMT3A-DNMT3L-XTEN80-NLS-ZFP结构域-NLS-XTEN16-KOX1KRAB(构型2),

[0279] NLS-DNMT3A-DNMT3L-XTEN80-dCas9-XTEN16-KOX1 KRAB-NLS(构型3),

[0280] NLS-DNMT3A-DNMT3L-XTEN80-ZFP结构域-XTEN16-KOX1KRAB-NLS(构型4),

[0281] NLS-NLS-DNMT3A-DNMT3L-XTEN80-dCas9-XTEN16-KOX1KRAB-NLS-NLS(构型5),和

[0282] NLS-NLS-DNMT3A-DNMT3L-XTEN80-ZFP结构域-XTEN16-KOX1KRAB-NLS-NLS(构型6)。

[0283] DNMT3L和DNMT3A可以衍生自人亲本蛋白、小鼠亲本蛋白或其任何组合。在某些实施方案中,DNMT3L和DNMT3A分别衍生自小鼠和人亲本蛋白(mDNMT3L和hDNMT3A)。在某些实施方案中,DNMT3L和DNMT3A均衍生自人亲本蛋白(hDNMT3L和hDNMT3A)。在一些实施方案中,dCas9是dSpCas9。在一些实施方案中,KOX1是人KOX1。还考虑了构型1-6中的任一种,其中KOX1 KRAB结构域被ZFP28、ZN627或ZIM3 KRAB结构域替换。在一些实施方案中,ZFP28、ZN627和ZIM3分别是人ZFP28、人ZN627和人ZIM3。在特定的实施方案中,融合构建体可以具有以下构型:

[0284] NLS-NLS-hDNMT3A-hDNMT3L-XTEN80-dCas9-XTEN16-KOX1KRAB-NLS-NLS(构型7),

[0285] NLS-NLS-DNMT3A-DNMT3L-XTEN80-ZFP结构域-XTEN16-KOX1KRAB-NLS-NLS(构型8),

[0286] NLS-NLS-hDNMT3A-hDNMT3L-XTEN80-dCas9-XTEN16-ZFP28KRAB-NLS-NLS(构型9),

[0287] NLS-NLS-DNMT3A-DNMT3L-XTEN80-ZFP结构域-XTEN16-ZFP28KRAB-NLS-NLS(构型10),

[0288] NLS-NLS-hDNMT3A-hDNMT3L-XTEN80-dCas9-XTEN16-ZN627KRAB-NLS-NLS(构型11),

[0289] NLS-NLS-DNMT3A-DNMT3L-XTEN80-ZFP结构域-XTEN16-ZN627KRAB-NLS-NLS(构型12),

[0290] NLS-NLS-hDNMT3A-hDNMT3L-XTEN80-dCas9-XTEN16-ZIM3KRAB-NLS-NLS(构型13),或

[0291] NLS-NLS-DNMT3A-DNMT3L-XTEN80-ZFP结构域-XTEN16-ZIM3KRAB-NLS-NLS(构型14)。

[0292] 在特定的实施方案中,本文所述的融合构建体可以具有构型1并且包含SEQ ID NO:658或与其至少75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同的序列。在以下SEQ ID NO:658中,XTEN接头带下划线,W接头带粗体、下划线和斜体,NLS序列带粗体,DNMT3A序列带斜体,DNMT3L序列带下划线和斜体,dCas9结构域带粗体和斜体,KOX1 KRAB结构域带下划线和粗体:

[0293] *MNHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCE DSITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLY EGTGRLFFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMID AKEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSN SIKQGKDQHFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSV PVIRHLFAPLKEYFACV***SSGNSNANSRGPSFSSGLVPLSLRGSH***MGPMEIYKTVSAW KRQPVRVLSLFRNIDKVLKSLGFLESGSGSGGGTLKYVEDVTNVVRRDVEKWGPFDL VYGSTQPLGSSCDRCPGWYMFQFHRILQYALPRQESQRPFFWIFMDNLLLTEDDQET TTRFLQTEAVTLQDVRGRDYQNAMRVWSNIPGLKSKHAPLTPKEEEYLQAQVRSRSK LDAPKVDLLVKNCLLPLREYFKYFSQNSLPLGGPSSGAPPPSGGSPAGSPTSTEEGT SESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSE**PKK KRKV**Y***MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALL FDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEE DKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRG HFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLA QIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKA LVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLN REDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYY VGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVL PKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLK EDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLT LTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTI LDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKG ILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGS QILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDD SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVS DFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKM IAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGR*

[0294] DFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD

[0295] SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK

[0296] DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSP

[0297] EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQA

[0298] ENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLS

[0299] QLGGDPKKKRKVSGSETPGTSESATPESTGRTLVTFKDVFVDFTREEWKLLDTAQQI

[0300] VYRNVMLENYKNLVSLGYQLTKPDVILRLEKGEEP(SEQ ID NO:658)

[0301] 在特定的实施方案中,本文所述的融合构建体可以包含以下提供的序列(SEQ ID NO:1495),或与其至少75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同的序列。在以下SEQ ID NO:1495中,XTEN接头带下划线,W接头带粗体、下划线和斜体,NLS序列带粗体,DNMT3A序列带斜体,DNMT3L序列带下划线和斜体,dCas9结构域带粗体和斜体,KOX1 KRAB结构域带下划线和粗体:

[0302] *MNHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCE DSITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLY EGTGRLFFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMID AKEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSN SIKQGKDQHFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSV PVIRHLFAPLKEYFACV**SSGNSNANSRGPSFSSGLVPLSLRGSH**MNPLEMFETVPVW RRQPVRVLSLFEDIKKELTSLGFLESGSDPGQLKHVVDVTDTVRKDVEEWGPFDLVY GATPPLGHTCDRPPSWYLFQFHRLLQYARPKPGSPRPFFWMFVDNLVLNKEDLDVAS RFLEMEPVTIPDVHGGSLQNAVRVWSNIPAIRSRHWALVSEEELSLLAQNKQSSKLA AKWPTKLVKNCFLPLREYFKYFSTELTSSL*GGPSSGAPPPSGGSPAGSPTSTEEGTS ESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSE**PKKK RKV**Y***MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEED KKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGH FLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLEN LIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQ IGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNR EDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYV GPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLP KHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKE DYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIL DFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGI LQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDS IDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGL SELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSD FRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMI AKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRD FATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDS*

[0303] *PTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKD LIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPE DNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAE NIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ LGGDPKKKRKV*SGSETPGTSESATPESTG**RTLVTFKDVFVDFTREEWKLLDTAQQIV YRNVMLENYKNLVSLGYQLTKPDVILRLEKGEEP** (SEQ ID NO: 1495)

[0304] 在特定的实施方案中,本文所述的融合构建体可以具有构型2并且包含SEQ ID NO:659或与其至少75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同的序列。在以下SEQ ID NO:659中,XTEN接头带下划线,W接头带粗体、下划线和斜体,NLS序列带粗体和下划线,DNMT3A序列带斜体,DNMT3L序列带下划线和斜体,ZFP结构域带粗体,并且KOX1 KRAB结构域带下划线和粗体。由X表示的可变氨基酸是锌指的DNA识别螺旋的氨基酸,斜体的XX可以是TR、LR或LK。

[0305] *MNHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCE DSITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLY EGTGRLFFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMID AKEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSN SIKQGKDQHFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSV PVIRHLFAPLKEYFACV***SSGNSNANSRGPSFSSGLVPLSLRGSH***MGPMEIYKTVSAW KRQPVRVLSLFRNIDKVLKSLGFLESGSGSGGGTLKYVEDVTNVVRRDVEKWGPFDL VYGSTQPLGSSCDRCPGWYMFQFHRILQYALPRQESQRPFFWIFMDNLLLTEDDQET TTRFLQTEAVTLQDVRGRDYQNAMRVWSNIPGLKSKHAPLTPKEEEYLQAQVRSRSK LDAPKVDLLVKNCLLPLREYFKYFSQNSLPLGGPSSGAPPPSGGSPAGSPTSTEEGT* SESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSE**PKK KRKVYSRPGERPFQCRICMRNFSXXXXXXXHXXTHTGEKPFQCRICMRNFSXXXXXX XHXXTH[接头]PFQCRICMRNFSXXXXXXXHXXTHTGEKPFQCRICMRNFSXXXXXX XHXXTH[接头]PFQCRICMRNFSXXXXXXXHXXTHTGEKPFQCRICMRNFSXXXXXX XHXXTHLRGSPKKKRKV**SGSETPGTSESATPESTG**RTLVTFKDVFVDFTREEWKLLD TAQQIVYRNVMLENYKNLVSLGYQLTKPDVILRLEKGEEP** (SEQ ID NO: 659)

[0306] 在某些实施方案中,SEQ ID NO:659中的六个“XXXXXXX”区依次包含表1中针对ZF001-ZF048中任一个所示的F1-F6氨基酸序列。[接头]表示接头序列。在一些实施方案中,一个或两个接头序列可以是TGSQKP(SEQ ID NO:651)。在一些实施方案中,一个或两个接头序列可以是TGGGGSQKP(SEQ ID NO:652)。在一些实施方案中,一个接头序列可以具有SEQ ID NO:651的氨基酸序列,并且另一个接头序列可以具有SEQ ID NO:652的氨基酸序列。

[0307] 在特定的实施方案中,本文所述的融合构建体可以包含以下提供的序列(SEQ ID NO:1496),或与其至少75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同的序列。在下面的SEQ ID NO:1496中,XTEN接头带下划线,W接头带粗体、下划线和斜体,NLS序列带粗体和下划线,DNMT3A序列带斜体,DNMT3L序列带下划线和斜体,ZFP结构域带粗体,KOX1 KRAB结构域带下划线和粗体。由X表示的可变氨基酸是锌指的DNA识别螺旋的氨基酸,斜体的XX可以是TR、LR或LK。

[0308] *MNHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCE DSITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLY EGTGRLFFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMID AKEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSN SIKQGKDQHFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSV PVIRHLFAPLKEYFACV***SSGNSNANSRGPSFSSGLVPLSLRGSH***MNPLEMFETVPVW RRQPVRVLSLFEDIKKELTSLGFLESGSDPGQLKHVVDVTDTVRKDVEEWGPFDLVY GATPPLGHTCDRPPSWYLFQFHRLLQYARPKPGSPRPFFWMFVDNLVLNKEDLDVAS RFLEMEPVTIPDVHGGSLQNAVRVWSNIPAIRSRHWALVSEEELSLLAQNKQSSKLA AKWPTKLVKNCFLPLREYFKYFSTELTSSL*GGPSSGAPPPSGGSPAGSPTSTEEGTS ESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSE**PKKK RKVYSRPGERPFQCRICMRNFSXXXXXXXHXXTHTGEKPFQCRICMRNFSXXXXXXX HXXTH[接头]PFQCRICMRNFSXXXXXXXHXXTHTGEKPFQCRICMRNFSXXXXXXX HXXTH[接头]PFQCRICMRNFSXXXXXXXHXXTHTGEKPFQCRICMRNFSXXXXXXX HXXTHLRGSPKKKRKV**SGSETPGTSESATPESTG**RTLVTFKDVFVDFTREEWKLLDT AQQIVYRNVMLENYKNLVSLGYQLTKPDVILRLEKGEEP** (SEQ ID NO: 1496)

[0309] 在某些实施方案中,SEQ ID NO:1496中的六个“XXXXXXX”区域依次包含表1中针对ZF001-ZF048中任一个所示的F1-F6氨基酸序列。[接头]表示接头序列。在一些实施方案中,

一个或两个接头序列可以是TGSQKP(SEQ ID NO:651)。在一些实施方案中,一个或两个接头序列可以是TGGGGSQKP(SEQ ID NO:652)。在一些实施方案中,一个接头序列可以具有SEQ ID NO:651的氨基酸序列,并且另一个接头序列可以具有SEQ ID NO:652的氨基酸序列。

[0310] 在一些实施方案中,融合蛋白可以进一步包含Dnmt3A ADD结构域,例如在上文公开的SEQ ID NO:658、659、1495或1496中公开的Dnmt3A结构域序列的下游。在一些实施方案中,ADD序列位于融合蛋白的Dnmt3A和Dnmt3L序列之间。在一些实施方案中,ADD序列位于Dnmt3A结构域的C末端。在一些实施方案中,Dnmt3A序列和ADD序列通过接头(例如本文公开的接头)分开。在一些实施方案中,ADD序列和Dnmt3L序列通过接头(例如本文公开的接头)分开。在一些实施方案中,ADD结构域包含序列:

[0311] MAAIPALDPEAEPSMDVILVGSSELSSSVSPGTGRDLIAYEVKANQRNIEDICICCGSLQVHTQHPLF EGGICAPCKDKFLDALFLYDDDGYQSYCSICCSGETLLICGNPDCTRCYCFECVDSLVGPGTSGKVHAMSNWVCYL CLPSSRSGLLQRRRKWRSQLKAFYDRESENPLEMFETVPVWRRQPVRVLSLFEDIKKELTSLGFLESGSDPGQLKH VVDVTDTVRKDVEEWGPFDLVYGATPPLGHTCDRPPSWYLFQFHRLLQYARPKPGSPRPFFWMFVDNLVLNKEDLD VASRFLEMEPVTIPDVHGGSLQNAVRVWSNIPAIRSRHWALVSEEELSLLAQNKQSSKLAAKWPTKLVKNCFLPLR EYFKYFSTELTSSL(SEQ ID NO:1497)。

[0312] 在特定的实施方案中,本文所述的融合构建体可以具有构型7并且包含SEQ ID NO:660或与其至少75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同的序列。

[0313] 在特定的实施方案中,本文所述的融合构建体可以具有构型9并且包含SEQ ID NO:661或与其至少75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同的序列。

[0314] 在特定的实施方案中,本文所述的融合构建体可以具有构型11并且包含SEQ ID NO:662或与其至少75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同的序列。

[0315] 在特定的实施方案中,本文所述的融合构建体可以具有构型13并且包含SEQ ID NO:663或与其至少75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同的序列。

[0316] 在一些实施方案中,本文所述的融合构建体(例如,构型1-14中任一个的融合构建体)在包含WPRE序列、聚腺苷酸化位点或两者的表达构建体内。在某些实施方案中,WPRE序列在3'非编码区中。在某些实施方案中,WPRE序列在聚腺苷酸化位点的上游。在特定的实施方案中,表达构建体包含融合构建体(例如,构型1-14中任一构型)和位于聚腺苷酸化位点上游的3'非编码区中的WPRE序列。

[0317] 在一些实施方案中,本文所述的融合构建体可以具有如实施例12中所示的融合蛋白1-12中任一个的序列。

[0318] 多种融合蛋白可用于实现靶基因或多个靶基因的激活或抑制。例如,包含DNA结合结构域(例如,dCas9结构域)和效应结构域的表观遗传编辑器融合蛋白可以与两个或更多个指导多核苷酸(例如,gRNA)共同递送,每个指导多核苷酸靶向不同的靶DNA序列。两个DNA结合结构域的靶位点可以相同或彼此附近,或者相隔例如约100个碱基对、约200个碱基对、约300个碱基对、约400个碱基对、约500个碱基对或约600个或更多个碱基对。另外,当靶向

双链DNA如内源基因座时,指导多核苷酸可以靶向相同或不同的链(一条或多条针对正链和/或一条或多条针对负链)。

[0319] V.靶序列

[0320] 本文的表观遗传编辑器可以针对PCSK9中的靶序列以实现PCSK9基因的表观遗传修饰。如本文所用,“靶序列”、“靶位点”或“靶区域”是存在于感兴趣的基因中的核酸序列;在一些情况下,靶序列可以在感兴趣的基因之外但在感兴趣的基因附近,其中靶序列的甲基化或阻遏物的结合抑制基因的表达。在一些实施方案中,靶序列可以是低甲基化或超甲基化的核酸序列。

[0321] 靶序列可以在靶基因的任何部分中。在一些实施方案中,靶序列是基因的非编码序列的一部分或在其附近。在一些实施方案中,靶序列是基因的外显子的一部分。在一些实施方案中,靶序列是基因的转录调节序列(如启动子或增强子)的一部分或在其附近。在一些实施方案中,靶序列与CpG岛相邻、重叠或涵盖CpG岛。在某些实施方案中,靶序列在PCSK9TSS侧翼的约3000、2900、2800、2700、2600、2500、2400、2300、2200、2100、2000、1900、1800、1700、1600、1500、1400、1300、1200、1100、1000、900、800、700、600、500、400、300、200或100个碱基对(bp)内。在某些实施方案中,靶序列在PCSK9 TSS侧翼的500bp内。在某些实施方案中,靶序列在PCSK9 TSS侧翼的1000bp内。

[0322] 在一些实施方案中,靶序列可以和与融合蛋白复合的指导多核苷酸序列(例如,gRNA)杂交,该融合蛋白包含多核苷酸指导的DNA结合结构域(例如,CRISPR蛋白,如dCas9)和效应结构域。指导多核苷酸序列可以被设计成与靶序列具有互补性,或与靶序列的相反链具有同一性。在一些实施方案中,指导多核苷酸包含与靶序列中的原间隔区序列约80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同的间隔区序列。在特定的实施方案中,指导多核苷酸包含与靶序列中的原间隔区序列100%相同的间隔区序列。

[0323] 在一些实施方案中,其中本文所述的表观遗传编辑器的DNA结合结构域是锌指阵列,靶序列可以被所述锌指阵列识别。

[0324] 在一些实施方案中,其中本文所述的表观遗传编辑器的DNA结合结构域是TALE,靶序列可以被所述TALE识别。

[0325] 本文所述的靶序列可以对靶基因的一个拷贝具有特异性,或者可以对靶基因的一个等位基因具有特异性。因此,表观遗传修饰及其表达调节可以对靶基因的一个拷贝或一个等位基因具有特异性。例如,表观遗传编辑器可以抑制含有由DNA结合结构域识别的靶序列的特定拷贝(例如,与疾病或病况相关的拷贝,或含有与疾病或病况相关的突变的拷贝)的表达。

[0326] 在一些实施方案中,靶PCSK9基因组区域可以落入以下所示的序列(chr1:55038548-55040548)内,具有或不具有末端A:

[0327] TACCTCATGGAGTCACTGTCAACCCACTGGTTGCACTGTCTTTGTGCACTGGCTCTCTGGAGTGAGGTCTTTGCAAACAAAGTGGAAAGAGCATCAACTTTGGACTCCAGCACCTAGATTCAGAGCAGGCCATTTCACTCGGAATCTGCTGTGCATCTGCAAGGGAGGATCATAAATTCGCCTTTGTTTCTTCCCAGTATCGACAGCCCTTCCAGAAAGAGCAAGCCTCATGTCATGCCACATGTACAATCTGAGGCCAGGAGCTCTCTTTCCCCTTTTCATCCTCCTGCCTGGTACACAATAGGTGTTTACTGGATGCTTGTCCAGTTGATTTCTTGAACATGGTGTGTAAAAGGAATCTTTGCAAATTGAATCTTCTGGAAAGCTGAGCTTGTGCCTACCATAGAATTCTGAATGTACCTATATGACGTCTTTGCAAACTTAAAAC

CTGAATCTTTGTAGTATAAATCCCTTGAAATGCATGTAGGCTGGACATCAAAAGCAAGCAATCTCTTCAAGGAGCA
GCTAGTTGGTAAGGTCAGTGTGCAGGGTGCATAAAGGGCAGAGGCCGGAGGGGGTCCAGGCTAAGTTTAGAAGGCT
GCCAGGTTAAGGCCAGTGGAAAGAATTCGGTGGGCAGCGAGGAGTCCACAGTAGGATTGATTCAGAAGTCTCACTG
GTCAGCAGGAGACAAGGTGGACCCAGGAAACACTGAAAAGGTGGGCCCGGCAGAACTTGGAGTCTGGCATCCCACG
CAGGGTGAGAGGCGGGAGAGGAGGAGCCCCTAGGGCGCCGGCCTGCCTTCCAGCCCAGTTAGGATTTGGGAGTTTT
TTCTTCCCTCTGCGCGTAATCTGACGCTGTTTGGGGAGGGCGAGGCCGAAACCTGATCCTCCAGTCCGGGGGTTCC
GTTAATGTTTAATCAGATAGGATCGTCCGATGGGGCTCTGGTGGCGTGATCTGCGCGCCCCAGGCGTCAAGCACCC
ACACCCTAGAAGGTTTCCGCAGCGACGTCGAGGCGCTCATGGTTGCAGGCGGGCGCCGCCGTTCAGTTCAGGGTCT
GAGCCTGGAGGAGTGAGCCAGGCAGTGAGACTGGCTCGGGCGGGCCGGGACGCGTCGTTGCAGCAGCGGCTCCCAG
CTCCCAGCCAGGATTCCGCGCGCCCCTTCACGCGCCCTGCTCCTGAACTTCAGCTCCTGCACAGTCCTCCCCCACCG
CAAGGCTCAAGGCGCCGCCGGCGTGGACCGCGCACGGCCTCTAGGTCTCCTCGCCAGGACAGCAACCTCTCCCCTG
GCCCTCATGGGCACCGTCAGCTCCAGGCGGTCCTGGTGGCCGCTGCCACTGCTGCTGCTGCTGCTGCTCCTGG
GTCCCGCGGGCGCCCGTGCGCAGGAGGACGAGGACGGCGACTACGAGGAGCTGGTGCTAGCCTTGCGTTCCGAGGA
GGACGGCCTGGCCGAAGCACCCGAGCACGGAACCACAGCCACCTTCCACCGCTGCGCCAAGGTGCGGGTGTAGGGA
TGGGAGGCCGGGGCGAACCCGCAGCCGGGACGGTGCGGTGCTGTTTCCTCTCGGGCCTCAGTTTCCCCCCATGTAA
GAGAGGAAGTGGAGTGCAGGTCGCCGAGGGCTCTTCGCTTGGCACGATCTTGGGGACTGCAGGCAAGGCGGCGGGG
GAGGACGGGTAGTGGGGAGCACGGTGGAGAGCGGGGACGGCCGGCTCTTTGGGGACTTGCTGGGGCGTGCGGCTGC
GCTATTCAGTGGGAAGGTTCGCGGGGTTGGGAGACCCGGAGGCCGAGGAAGGGCGAGCAGAGCACTGCCAGGATAT
CCTGCCCAGATTTCCCAGTTTCTGCCTCGCCGCGGCACAGGTGGGTGAAGGAGTGAATGCCTGGAACGTACTGGGA
ACTGCACCAGGCACAGAGAAAGCGGGCTTGCCATTATAGTGGGTTCCGATTTGGTTTGGAAAACATGGGCAGCGGA
GGGTGGAGGGCCTGGAGAGAAGGCCCTACCCGA(SEQ ID NO:1488)

[0328] 在一些实施方案中,靶序列可以是GRCh38 Chr1:55039228-55040296,如下所示:

[0329] GCAGGAGACAAGGTGGACCCAGGAAACACTGAAAAGGTGGGCCCGGCAGAACTTGGAGTCTGGCATCC
CACGCAGGGTGAGAGGCGGGAGAGGAGGAGCCCCTAGGGCGCCGGCCTGCCTTCCAGCCCAGTTAGGATTTGGGAG
TTTTTTCTTCCCTCTGCGCGTAATCTGACGCTGTTTGGGGAGGGCGAGGCCGAAACCTGATCCTCCAGTCCGGGGG
TTCCGTTAATGTTTAATCAGATAGGATCGTCCGATGGGGCTCTGGTGGCGTGATCTGCGCGCCCCAGGCGTCAAGC
ACCCACACCCTAGAAGGTTTCCGCAGCGACGTCGAGGCGCTCATGGTTGCAGGCGGGCGCCGCCGTTCAGTTCAGG
GTCTGAGCCTGGAGGAGTGAGCCAGGCAGTGAGACTGGCTCGGGCGGGCCGGGACGCGTCGTTGCAGCAGCGGCTC
CCAGCTCCCAGCCAGGATTCCGCGCGCCCCTTCACGCGCCCTGCTCCTGAACTTCAGCTCCTGCACAGTCCTCCCC
ACCGCAAGGCTCAAGGCGCCGCCGGCGTGGACCGCGCACGGCCTCTAGGTCTCCTCGCCAGGACAGCAACCTCTCC
CCTGGCCCTCATGGGCACCGTCAGCTCCAGGCGGTCCTGGTGGCCGCTGCCACTGCTGCTGCTGCTGCTGCTGCTC
CTGGGTCCCGCGGGCGCCCGTGCGCAGGAGGACGAGGACGGCGACTACGAGGAGCTGGTGCTAGCCTTGCGTTCCG
AGGAGGACGGCCTGGCCGAAGCACCCGAGCACGGAACCACAGCCACCTTCCACCGCTGCGCCAAGGTGCGGGTGTA
GGGATGGGAGGCCGGGGCGAACCCGCAGCCGGGACGGTGCGGTGCTGTTTCCTCTCGGGCCTCAGTTTCCCCCCAT
GTAAGAGAGGAAGTGGAGTGCAGGTCGCCGAGGGCTCTTCGCTTGGCACGATCTTGGGGACTGCAGGCAAGGCGGC
GGGGGAGGACGGGTAGTGGGGAGCACGGTGGAGAGCGGGGACGGCCGGCTCTTTGGGGACTTGCTGGGGCGTGCGG
CTGCGCTATTCAG(SEQ ID NO:1489)

[0330] VI.表观遗传修饰

[0331] 本文所述的表观遗传编辑器可以对含有靶序列的靶基因进行序列特异性表观遗

传修饰(例如,化学修饰的改变)。这种表观遗传调节可以比由于基因编辑引起的调节(例如通过产生DNA双链断裂)更安全且更容易可逆。在一些实施方案中,表观遗传调节可以减少或沉默靶基因。在一些实施方案中,修饰在靶序列的特定位点处。在一些实施方案中,修饰在靶基因的特定等位基因处。因此,表观遗传修饰可以导致含有特定等位基因的靶基因的一个拷贝而不是靶基因的另一个拷贝的表达被调节(例如,降低)。在一些实施方案中,特定等位基因与疾病、病况或病症相关。

[0332] 在一些实施方案中,表观遗传修饰减少或消除含有靶序列的靶基因的转录。在一些实施方案中,表观遗传修饰减少或消除含有由表观遗传编辑器识别的特定等位基因的靶基因拷贝的转录。在一些实施方案中,表观遗传编辑器降低由靶基因编码的蛋白质的水平或消除其表达。在一些实施方案中,表观遗传编辑器降低由含有表观遗传编辑器识别的特定等位基因的靶基因拷贝编码的蛋白质的水平或消除其表达。靶PCSK9基因可以在体外、离体或体内进行表观遗传修饰。

[0333] 本文所述的表观遗传编辑器的效应结构域可以改变(例如,放置或去除)靶基因的核苷酸处或与靶基因相关的组蛋白处的化学修饰。化学修饰可以在单个核苷酸或单个组蛋白处改变,或者可以在2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000、1500、2000、2500、3000或更多个核苷酸处改变。

[0334] 在一些实施方案中,本文所述的表观遗传编辑器的效应结构域可以改变靶基因内的CpG二核苷酸。在一些实施方案中,与基因的原始状态或未与表观遗传编辑器接触的可比细胞中的基因相比,靶序列侧翼2000、1500、1000、500或200bp内的所有CpG二核苷酸(例如,在如本文所述的改变位点中)根据本文所述的修饰类型被改变。在一些实施方案中,与基因的原始状态或未与表观遗传编辑器接触的可比细胞中的基因相比,至少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、35、40、45、50、60、70、80、90、100、150、200、250、300、350、400、450、500、550、600、650、700或更多个CpG二核苷酸被改变。在一些实施方案中,与基因的原始状态或未与表观遗传编辑器接触的可比细胞中的基因相比,至少5%、6%、7%、8%、9%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%或99%的CpG二核苷酸被改变。在一些实施方案中,与基因的原始状态或未与表观遗传编辑器接触的可比细胞中的基因相比,一个单一CpG二核苷酸被改变。

[0335] 本文所述的表观遗传编辑器的效应结构域可以改变与靶基因相关或结合的组蛋白的组蛋白修饰状态。例如,效应结构域可以在与靶基因相关的组蛋白的组蛋白尾部的一个或多个赖氨酸残基上放置修饰。在一些实施方案中,效应结构域可导致与靶基因相关的组蛋白的一个或多个组蛋白尾的去乙酰化,从而降低或沉默靶基因的表达。在一些实施方案中,组蛋白修饰状态是甲基化状态。例如,效应结构域可导致与靶基因相关的一个或多个组蛋白尾部处的H3K9、H3K27或H4K20甲基化(例如,H3K9me2、H3K9me3、H3K27me2、H3K27me3和H4K20me3甲基化中的一个或多个),从而降低或沉默靶基因的表达。

[0336] 在一些实施方案中,与染色体的原始状态或未与表观遗传编辑器接触的可比细胞中的染色体相比,与靶序列侧翼2000、1500、1000、500或200bp内的DNA核苷酸结合的组蛋白的所有组蛋白尾根据本文所述的修饰类型被改变。在一些实施方案中,与染色体的原始状

态或未与表观遗传编辑器接触的可比细胞中的染色体相比,结合的组蛋白的至少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、35、40、45、50、55、60、65、70、75、80、85、90、95、100、105、110、115、120或更多个组蛋白尾发生改变。在一些实施方案中,与染色体的原始状态或未与表观遗传编辑器接触的可比细胞中的染色体相比,结合的组蛋白的至少5%、6%、7%、8%、9%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%或99%的组蛋白尾发生改变。例如,与染色体的原始状态或未与表观遗传编辑器接触的可比细胞中的染色体相比,结合的组蛋白的一个单一组蛋白尾可以被改变。作为另一个示例,与染色体的原始状态或未与表观遗传编辑器接触的可比细胞中的染色体相比,一个单一结合的组蛋白八聚体可以被改变。

[0337] 放置在靶基因DNA核苷酸或组蛋白残基处的化学修饰可以位于或紧邻靶基因中的靶序列。在一些实施方案中,本文所述的表观遗传编辑器的效应结构域改变靶基因中的靶序列5'或3'处100-200、200-300、300-400、400-55、500-600、600-700或700-800个核苷酸的核苷酸或与其结合的组蛋白尾的化学修饰状态。在一些实施方案中,效应结构域改变靶序列侧翼的10、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000、1100、1200、1300、1400、1500、1600、1700、1800、1900或2000个核苷酸内的核苷酸或与其结合的组蛋白尾的化学修饰状态。如本文所用,"侧翼"是指特定序列(例如靶序列)的5'至5'末端和3'至3'末端的核苷酸位置。

[0338] 在一些实施方案中,效应结构域介导或诱导远离靶序列的核苷酸或与其结合的组蛋白尾的化学修饰改变。这种修饰可以在靶序列附近起始,并且随后可以扩散到靶基因中远离靶序列的一个或多个核苷酸。例如,效应结构域可以起始靶序列侧翼的10、20、30、40、50、60、70、80、90、100、200、300、400、500个核苷酸内的一个或多个核苷酸或与其结合的一个或多个组蛋白残基的化学修饰状态的改变,并且化学修饰状态的改变可以扩散到靶基因中靶序列上游或下游距靶序列至少100、200、300、400、500、600、700、800、900、1000、1100、1200、1300、1400、1500、1600、1700、1800、1900、2000、2500、3000或更多个核苷酸的一个或多个核苷酸。在某些实施方案中,化学修饰可以在靶基因中少于2、3、5、10、20、30、40、50或100个核苷酸处起始,并扩散至靶基因中至少100、200、300、400、500、600、700、800、900、1000、2000或更多个核苷酸。在一些实施方案中,化学修饰延伸至整个靶基因中的核苷酸。额外的蛋白质或转录因子,例如转录阻遏物、甲基转移酶或转录调节支架蛋白,可以参与化学修饰的扩散。或者,可以单独涉及表观遗传编辑器。

[0339] 在一些实施方案中,与对照细胞、对照组织或对照受试者(例如,在没有表观遗传编辑器的情况下)相比,本文描述的表观遗传编辑器使靶基因的表达降低至少约20%、至少约30%、至少约40%、至少约50%、至少约60%、至少约70%、至少约80%、至少约90%、至少约95%、至少约99%或更多,如通过细胞、组织或受试者中靶基因的转录测量的。在一些实施方案中,与对照细胞、对照组织或对照受试者相比,本文所述的表观遗传编辑器将靶基因拷贝的表达降低至少约20%、至少约30%、至少约40%、至少约50%、至少约60%、至少约70%、至少约80%、至少约90%、至少约95%、至少约99%或更多,如通过在细胞、组织或受试者中靶基因拷贝的转录所测量的。在某些实施方案中,靶基因的拷贝具有由表观遗传编辑器识别的特定序列或等位基因。在特定的实施方案中,表观遗传修饰的拷贝编码功能性

蛋白质,因此本文公开的表观遗传编辑器可以降低或消除蛋白质的表达和/或功能。例如,与对照细胞、对照组织或对照受试者相比,本文所述的表观遗传编辑器可以使细胞、组织或受试者中由靶基因编码的蛋白质的表达和/或功能降低至少3倍、至少4倍、至少5倍、至少6倍、至少7倍、至少8倍、至少9倍、至少10倍、至少11倍、至少12倍、至少13倍、至少14倍、至少15倍、至少20倍、至少25倍、至少30倍、至少35倍、至少40倍、至少45倍、至少50倍、至少60倍、至少70倍、至少80倍、至少90倍或至少100倍。

[0340] 靶基因表达的调节可以通过确定受靶基因表达间接或直接影响的任何参数来测定。此类参数包括例如RNA或蛋白质水平的变化;蛋白质活性的变化;产品水平的变化;下游基因表达的变化;报告基因(如,例如荧光素酶、CAT、β-半乳糖苷酶或GFP)的转录或活性的变化;信号转导的变化;磷酸化和去磷酸化的变化;受体-配体相互作用的变化;第二信使(例如cGMP、cAMP、IP3和Ca^{2+})的浓度变化;细胞生长的变化;新生血管形成的变化;和/或基因表达的任何功能效应的变化。测量可以在体外、体内和/或离体进行,并且可以通过常规方法进行,例如测量RNA或蛋白质水平、测量RNA稳定性和/或鉴定下游或报告基因表达。读出可以通过例如化学发光、荧光、比色反应、抗体结合、诱导型标志物、配体结合测定、细胞内第二信使如cGMP和三磷酸肌醇(IP3)的变化、细胞内钙水平的变化;细胞因子释放等方式。

[0341] 用于确定基因(例如表观遗传编辑器的靶标)的表达水平的方法可以包括例如通过逆转录PCR、定量RT-PCR、液滴数字PCR(ddPCR)、Northern印迹、RNA测序、DNA测序(例如,从RNA获得的互补脱氧核糖核酸(cDNA)的测序);下一代(Next-Gen)测序、纳米孔测序、焦磷酸测序或纳米串测序来确定基因的转录物水平。从基因表达的蛋白质水平可以例如通过蛋白质印迹、酶联免疫吸附测定、质谱、免疫组织化学或流式细胞术分析来测定。可以将基因表达产物水平标准化为内标,例如总信使核糖核酸(mRNA)或特定基因(例如管家基因)的表达水平。

[0342] 在一些实施方案中,可以使用报告系统检查表观遗传编辑器在调节靶基因表达中的作用。例如,表观遗传编辑器可以被设计成靶向编码报告蛋白(如荧光蛋白)的报告基因。可以通过例如流式细胞术、荧光激活细胞分选(FACS)或荧光显微镜监测报告基因在此类模型系统中的表达。在一些实施方案中,可以用含有报告基因的载体转染细胞群。可以构建载体,使得当载体转染细胞时报告基因表达。合适的报告基因包括编码荧光蛋白的基因,例如绿色、黄色、樱桃色、青色或橙色荧光蛋白。携带报告系统的细胞群可以用编码靶向报告基因的表观遗传编辑器的DNA、mRNA或载体转染。

[0343] VII.药物组合物

[0344] 在一个方面,本公开提供了药物组合物,其包含作为活性成分(或作为唯一活性成分)的一种或多种本文所述的表观遗传编辑器或其组分(例如,融合蛋白和/或指导多核苷酸),或编码所述表观遗传编辑器或其组分的核酸分子。例如,药物组合物可以包含编码本文所述的表观遗传编辑器的融合蛋白(和指导多核苷酸,如果适用)的核酸分子。在一些实施方案中,单独的药物组合物包含融合蛋白和指导多核苷酸。药物组合物还可以包含已经经历由本文提供的表观遗传编辑器介导或诱导的表观遗传修饰的细胞。

[0345] 通常,本公开的本文所述的表观遗传编辑器或其组分,或编码所述表观遗传编辑器或其组分的核酸分子适合作为与一种或多种药学上可接受的赋形剂(例如,如下所述)联

合的制剂施用。

[0346] 术语“赋形剂”在本文中用于描述除本公开的化合物以外的任何成分。赋形剂的选择将在很大程度上取决于如特定施用模式、赋形剂对溶解度和稳定性的影响以及剂型的性质等因素。如本文所用,“药学上可接受的赋形剂”包括生理上相容的任何和所有溶剂、分散介质、包衣、抗菌剂和抗真菌剂、等渗剂和吸收延迟剂等。药学上可接受的赋形剂的一些示例是水、盐水、磷酸盐缓冲盐水、葡萄糖、甘油、乙醇等,以及它们的组合。在许多情况下,优选在组合物中包括等渗剂,例如糖、多元醇如甘露醇、山梨醇或氯化钠。药学上可接受的物质的另外示例是润湿剂或少量的辅助物质,例如润湿剂或乳化剂、防腐剂或缓冲剂,其增强抗体的保质期或有效性。

[0347] 适于肠胃外施用的药物组合物的制剂通常包含与药学上可接受的载体(例如无菌水或无菌等渗盐水)组合的活性成分。此类制剂可以以适于推注施用或连续施用的形式制备、包装或销售。

[0348] VIII.递送方法

[0349] 在一些实施方案中,将表观遗传编辑器或其组分以编码表观遗传编辑器或其组分的核酸分子的形式引入靶细胞;因此,本文的药物组合物包含核酸分子。此类核酸分子可以是例如DNA、RNA或mRNA,和/或经修饰的核酸序列(例如,具有化学修饰、5’帽或一个或多个3’修饰)。在一些实施方案中,核酸分子可以例如通过转染或电穿孔作为裸DNA或RNA递送,或者可以与促进靶细胞摄取的分子(例如N-乙酰半乳糖胺)缀合。在一些实施方案中,核酸分子可以在核酸表达载体中,其可以包括表达控制序列,例如启动子、增强子、转录信号序列、转录终止序列、内含子、聚腺苷酸化信号、Kozak共有序列、内部核糖体进入位点(IRES)等。此类表达控制序列是本领域熟知的。载体还可以包含编码信号肽(例如,用于核定位、核仁定位或线粒体定位)的序列,其与编码蛋白质的序列相关联(例如,插入或融合)。

[0350] 载体的示例包括但不限于质粒载体;基于痘苗病毒、脊髓灰质炎病毒、腺病毒、腺相关病毒、SV40、单纯疱疹病毒、人免疫缺陷病毒、逆转录病毒(例如鼠白血病病毒或脾坏死病毒,衍生自逆转录病毒的载体如劳斯肉瘤病毒、哈维肉瘤病毒、禽白血病病毒、慢病毒、人免疫缺陷病毒、骨髓增生性肉瘤病毒和乳腺肿瘤病毒)的病毒载体;和其它重组载体。在某些实施方案中,载体是质粒或病毒载体。病毒颗粒或病毒样颗粒(VLP)也可用于递送编码如本文所述的表观遗传编辑器或其组分的核酸分子。例如,可以组装“空”病毒颗粒以包含任何合适的货物。病毒载体和病毒颗粒也可以被工程化以掺入靶向配体来改变靶组织特异性。

[0351] 在某些实施方案中,如本文所述的表观遗传编辑器或其组分由存在于一种或多种病毒载体中的核酸序列或任何病毒载体的合适衣壳蛋白编码。病毒载体的示例包括腺相关病毒载体(例如,衍生自AAV3、AAV3b、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAVrh8、AAV10和/或其变体);逆转录病毒载体(例如,莫洛尼鼠白血病病毒,MML-V)、腺病毒载体(例如,AD100)、慢病毒载体(例如,基于HIV和FIV的载体)和疱疹病毒载体(例如,HSV-2)。

[0352] 在一些实施方案中,递送涉及腺相关病毒(AAV)载体。当表观遗传编辑器融合蛋白的DNA结合结构域是锌指阵列时,AAV载体递送可能是特别有用的。不希望受任何理论的束缚,与较大的DNA结合结构域如Cas蛋白结构域相比,较小尺寸的锌指阵列可以允许将此类融合蛋白方便地包装在病毒载体如AAV载体中。

[0353] 任何AAV血清型,例如人AAV血清型,可用于如本文所述的AAV载体,包括但不限于AAV血清型1(AAV1)、AAV血清型2(AAV2)、AAV血清型3(AAV3)、AAV血清型4(AAV4)、AAV血清型5(AAV5)、AAV血清型6(AAV6)、AAV血清型7(AAV7)、AAV血清型8(AAV8)、AAV血清型9(AAV9)、AAV血清型10(AAV10)和AAV血清型11(AAV11)及其变体。在一些实施方案中,AAV变体与野生型AAV具有至少90%、91%、92%、93%、94%、95%、96%、97%、98%或99%氨基酸序列同一性。在某些实施方案中,可以工程化AAV变体,使得其衣壳蛋白在人中具有降低的免疫原性或增强的转导能力。在一些情况下,至少两种不同AAV血清型病毒的一个或多个区域被改组和重组以产生嵌合变体。例如,嵌合AAV可以包含与衣壳的血清型相比具有异源血清型的反向末端重复序列(ITR)。所得嵌合AAV与其亲本血清型相比可具有不同的抗原反应性或识别。在一些实施方案中,AAV的嵌合变体包括来自2、3、4、5或更多种不同AAV血清型的氨基酸序列。

[0354] 非病毒系统也被考虑用于如本文所述的递送。非病毒系统包括但不限于核酸转染方法,包括电穿孔、声穿孔、磷酸钙转染、显微注射、DNA生物弹道、脂质介导的转染、通过热休克的转染、压缩DNA介导的转染、脂质转染、阳离子试剂介导的转染和用脂质体、免疫脂质体、外泌体或阳离子表面两亲物(CFA)转染。在某些实施方案中,编码如本文所述的表观遗传编辑器融合蛋白的一种或多种mRNA可以与如本文所述的一种或多种指导多核苷酸(例如,gRNA)共同电穿孔。非病毒核酸载体的一个重要类别是纳米颗粒,其可以是有机的(例如脂质)或无机的(例如金)。例如,在本公开的某些实施方案中,有机(例如脂质和/或聚合物)纳米颗粒可适合用作递送媒介物。

[0355] 在一些实施方案中,使用脂质纳米颗粒(LNP)完成递送。LNP组合物的尺寸通常在微米或更小的数量级,并且可以包括脂质双层。在一些实施方案中,LNP是指直径小于1000nm、500nm、250nm、200nm、150nm、100nm、75nm、50nm或25nm的任何颗粒。在一些实施方案中,纳米颗粒的尺寸范围可以为1-1000nm、1-500nm、1-250nm、25-200nm、25-100nm、35-75nm或25-60nm。纳米颗粒组合物包括脂质纳米颗粒(LNP)、脂质体(例如脂质囊泡)和脂质复合物。

[0356] 如本文所述的LNP可以由阳离子、阴离子或中性脂质制成。在一些实施方案中,LNP可以包含中性脂质,例如融合性磷脂1,2-二油酰基-sn-甘油-3-磷酸乙醇胺(DOPE)或膜组分胆固醇,作为辅助脂质以增强转染活性和纳米颗粒稳定性。在一些实施方案中,LNP可以包含疏水性脂质、亲水性脂质或疏水性和亲水性脂质两者。本领域已知的任何脂质或脂质组合可用于产生LNP。脂质可以以任何摩尔比组合以产生LNP。在一些实施方案中,LNP是靶向肝脏(例如,优先地或特异性地靶向肝脏)的LNP。

[0357] 可以靶向任何类型的细胞以递送如本文所述的表观遗传编辑器或其组分。例如,细胞可以是真核的或原核的。在一些实施方案中,细胞是哺乳动物(例如人)细胞。人细胞可以包括例如肝细胞、胆管上皮细胞(胆管细胞)、星状细胞、库普弗细胞和肝窦内皮细胞。

[0358] 在一些实施方案中,将本文所述的表观遗传编辑器或其组分递送至宿主细胞用于瞬时表达,例如经由瞬时表达载体。表观遗传编辑器或其组分的瞬时表达可以导致靶基因的延长的或永久的表观遗传修饰。例如,在将表观遗传编辑器引入宿主细胞后,表观遗传修饰可以稳定至少1、2、3、4、5、6、7、8、9、10、11、或12周或更长时间;或3、4、5、6、7、8、9、10、11或12个月或更长时间。表观遗传修饰可以在宿主细胞的一次或多次有丝分裂和/或减数分

裂事件之后维持。在特定的实施方案中,表观遗传修饰在由宿主细胞产生或衍生的后代中跨代维持。

[0359] IX.表观遗传编辑器的治疗用途

[0360] 本公开还提供了用于治疗或预防受试者的病症的方法,其包括向受试者施用如本文所述的表观遗传编辑器或药物组合物。表观遗传编辑器可以实现与受试者中的疾病、病况或病症相关的靶基因中的靶多核苷酸序列的表观遗传修饰,从而调节靶基因的表达以治疗或预防疾病、病况或病症。在一些实施方案中,表观遗传编辑器将靶基因的表达降低至足以实现期望效果的程度,例如治疗相关效果,如预防或治疗疾病、病况或病症。

[0361] 在一些实施方案中,向受试者施用用于调节(例如,抑制)PCSK9表达的系统,其中所述系统包含(1)如本文所述的融合蛋白和(在相关的情况下)指导多核苷酸,或(2)编码所述融合蛋白和(在相关的情况下)指导多核苷酸的核酸分子。

[0362] “治疗(treat)”、“治疗(treating)”和“治疗(treatment)”是指减轻或消除生物病症和/或其伴随症状中的至少一种的方法。如本文所用,“减轻”疾病、病症或病况意指降低疾病、病症或病况的症状的严重性和/或发生频率。此外,本文提及的“治疗(treatment)”包括提及治愈性、姑息性和预防性治疗。在一些实施方案中,与等同的未治疗对照相比,缓解症状可涉及通过任何标准技术测量的症状减轻至少3%、5%、10%、20%、40%、50%、60%、80%、90%、95%、98%、99%、99.5%、99.9%或100%。

[0363] 在一些实施方案中,受试者可以是哺乳动物,例如人。在一些实施方案中,受试者选自非人灵长类动物,例如黑猩猩、食蟹猴或猕猴,以及其他猿和猴物种。

[0364] 在一些实施方案中,人患者患有选自高胆固醇血症(例如,家族性高胆固醇血症,如杂合性家族性高胆固醇血症(HeFH)或纯合性家族性高胆固醇血症(HoFH),或已确定的动脉粥样硬化性心血管疾病(ASCVD))或肾功能不全(RI)的病况。

[0365] 在一些实施方案中,待用本公开的表观遗传编辑器治疗的患者已经接受了针对待治疗病况(例如,高胆固醇血症(如HeFH、HoFH、HF或已确定的ASCVD)或RI)的先前治疗。在其他实施方案中,患者未接受此类先前治疗。在一些实施方案中,患者对该病况的先前治疗(例如,先前的高胆固醇血症治疗)失败。

[0366] 本公开的表观遗传编辑器可以以治疗有效量施用于患有本文所述病况的患者。如本文所用,“治疗有效量”是指所施用的治疗剂的量,其将在一定程度上缓解所治疗的病症的一种或多种症状,和/或导致医疗保健专业人员期望的临床终点。治疗的有效量可以通过其稳定患者的疾病进展和/或改善症状的能力并且优选逆转疾病进展的能力来测量。本公开的表观遗传编辑器降低或沉默PCSK9表达的能力可以通过体外测定(例如,如本文所述)以及在预测对人有功效的合适动物模型中进行评估。将选择合适的剂量方案,以便在每种特定情况下提供最佳的治疗反应,例如,作为单次推注或作为连续输注施用,并且根据每个病例的紧急情况可能调整剂量。

[0367] 本公开的表观遗传编辑器可以在没有额外治疗性治疗的情况下施用,即作为独立疗法(单一疗法)。或者,用本公开的表观遗传编辑器治疗可以包括至少一种额外的治疗性治疗(组合疗法)。在一些实施方案中,额外的治疗剂是本领域已知的用于治疗高胆固醇血症或RI的任何治疗剂。治疗剂包括但不限于他汀类、贝特类、HMG-CoA还原酶抑制剂、烟酸、胆汁酸调节剂或螯合剂、胆固醇吸收抑制剂或调节剂、CETP抑制剂、MTTP抑制剂和PPAR激动

剂。

[0368] 本公开的表观遗传编辑器或其组分(或编码表观遗传编辑器或其组分的核酸分子)可以通过本领域接受的任何方法施用,例如皮下、皮内、肿瘤内、结节内、肌肉内、静脉内、淋巴内或腹膜内。在特定的实施方案中,本公开的药物组合物静脉内施用至受试者。

[0369] X.定义

[0370] 如本文所用,术语“核酸”是指含有单链或双链形式的核苷酸(例如脱氧核糖核苷酸或核糖核苷酸)的任何寡核苷酸或多核苷酸,并且包括DNA和RNA。“核苷酸”包含糖脱氧核糖(DNA)或核糖(RNA)、碱基和磷酸基团,并通过磷酸基团连接在一起。“碱基”包括嘌呤和嘧啶,其包括天然化合物如腺嘌呤、胸腺嘧啶、鸟嘌呤、胞嘧啶、尿嘧啶、肌苷和天然类似物;以及嘌呤和嘧啶的合成衍生物,其包括但不限于放置新的反应性基团如胺、醇、硫醇、羧酸盐、卤代烷等的修饰形式。核酸可以含有已知的核苷酸类似物和/或修饰的主链残基或键,其可以是合成的、天然存在的和非天然存在的。此类核苷酸类似物、修饰残基和修饰键是本领域熟知的,并且在核酸酶存在下可以提供具有增强的细胞摄取、降低的免疫原性和/或增加的稳定性的核酸分子。

[0371] 如本文所用,“分离的”或“纯化的”核酸分子是脱离其天然环境而存在的核酸分子。例如,“分离的”或“纯化的”核酸分子(1)已与其来源的基因组DNA或细胞RNA的核酸分离;和/或(2)在自然界中不存在。在一些实施方案中,“分离的”或“纯化的”核酸分子是重组核酸分子。

[0372] 应当理解,除了本文提及的特定蛋白质和核酸分子之外,本公开还考虑使用其变体、衍生物、同源物和片段。任何给定序列的变体可以具有特定残基序列(无论是氨基酸残基还是核酸残基),其修饰方式使得所讨论的多肽或多核苷酸基本上保留其至少一种内源功能。变体序列可以通过存在于天然存在的序列中的至少一个残基(在一些实施方案中,不超过1、2、3、4、5、6、7、8、9、10、15或20个残基)的添加、缺失、取代、修饰、置换和/或变异而获得。对于本文所述的特定蛋白质(例如,本文所述的KRAB、dCas9、DNMT3A和DNMT3L蛋白质),本公开还考虑了保留其内源功能中的至少一种(例如,与所述的特定蛋白质相比,其功能的至少50%、60%、70%、80%、90%、85%、96%、97%、98%或99%)的任何蛋白质的天然存在形式或变体或同源物。

[0373] 本文提供了本公开所包括的一些示例性融合蛋白。本领域技术人员将理解,这些示例性蛋白质是非限制性示例,并且额外的蛋白质在本公开的范围内。例如,在提供包含特定结构域(例如哺乳动物DNMT3A、DNMT3L和/或KRAB结构域(例如人或小鼠DNMT3A、DNMT3L和/或KRAB结构域)的融合示例性蛋白的情况下,本领域技术人员能够确定,在一些实施方案中,本公开还包括具有相同构型但其中一个或多个哺乳动物结构域取代了来自另一哺乳动物的同源结构域(例如一个或多个小鼠结构域取代了一个或多个人结构域)的融合蛋白。例如,在提供包含小鼠DNMT3L结构域的示例性融合蛋白的情况下,也包括具有相同结构但小鼠DNMT3L取代人DNMT3L结构域的融合蛋白。

[0374] 如本文所用,本文所考虑的任何多肽或核酸序列的同源物包括与野生型氨基酸和核酸序列具有一定同源性的序列。同源序列可以包括与受试者序列至少50%、55%、65%、75%、85%、90%、91%、92%<93%、94%、95%、96%、97%、98%或99%相同的序列,例如氨基酸序列。在氨基酸或核苷酸序列的上下文中,术语“同一性百分比”是指当比对最大对应

性时两个序列中相同的残基的百分比。在一些实施方案中,出于比较目的比对的参考序列的长度为参考序列的至少30%(例如,至少40%、50%、60%、70%、80%或90%或100%)。序列同一性可以使用序列分析软件(例如,威斯康星大学生物技术中心遗传学计算机组的序列分析软件包,1710University Avenue,Madison,Wis.53705,BLAST,BESTFIT,GAP或PILEUP/PRETTYBOX程序)来测量。此类软件通过将同源度分配给各种取代、缺失和/或其它修饰来匹配相同或相似的序列。在确定同一性程度的示例性方法中,可以使用BLAST程序,其中e-3和e-100之间的概率分数指示密切相关的序列。

[0375] 两个核苷酸或多肽序列的同一性百分比通过例如BLAST®使用默认参数(可在美国国家医学图书馆的国家生物技术信息中心网站获得)来确定。在一些实施方案中,出于比较目的比对的参考序列的长度为参考序列的至少30%(例如,至少40%、50%、60%、70%、80%或90%)。

[0376] 应当理解,多肽序列中特定位置或残基的编号取决于所使用的特定蛋白质和编号方案。编号可能不同,例如在成熟蛋白的前体和成熟蛋白本身中,并且物种与物种之间序列的差异可能影响编号。本领域技术人员将能够通过本领域熟知的方法,例如通过序列比对和同源残基的测定,鉴定任何同源蛋白和相应编码核酸中的相应残基。

[0377] 术语“调节”或“改变”是指功能的量、程度或范围的变化。例如,如本文所述的表观遗传编辑器可以通过与启动子内的基序结合来调节启动子序列的活性,从而诱导、增强或抑制与启动子序列可操作地连接的基因的转录。作为其他示例,如本文所述的表观遗传编辑器可以阻断RNA聚合酶转录基因,或者可以抑制mRNA转录物的翻译。当用于提及如本文所述的表观遗传编辑器或其组分时,术语“抑制(inhibit)”、“抑制(repress)”、“抑制(suppress)”、“沉默”等是指相对于不存在表观遗传编辑器或其组分时核酸序列或蛋白质的活性而言,降低或阻止核酸序列(例如,靶基因)或蛋白质的活性(例如,转录)。该术语可包括部分或完全阻断活性,或防止或延迟活性。被抑制的活性可以比对照的活性低例如10%、20%、30%、40%、50%、60%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%,或者可以比对照的活性低例如至少1.5倍、2倍、3倍、4倍、5倍或10倍。

[0378] 术语“约”或“大约”意指在由本领域普通技术人员确定的特定值的可接受误差范围内,这将部分地取决于如何测量或确定该值,例如测量系统的限制。例如,根据惯例,“约”可以表示在给定值的一个或多于一个标准偏差内。在本申请和权利要求中描述特定值的情况下,除非另有说明,否则术语“约”应被假定为意指特定值的可接受误差范围。

[0379] 本文提供的范围应理解为该范围内的所有值的简写。例如,1至50的范围应理解为包括选自由以下组成的组的任何数字、数字的组合或子范围:1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49或50,以及前述整数之间的所有中间十进制值,例如1.1、1.2、1.3、1.4、1.5、1.6、1.7、1.8和1.9。关于子范围,特别考虑从范围的任一端点延伸的“嵌套子范围”。例如,示例性范围1至50的嵌套子范围可以在一个方向上包括1至10、1至20、1至30和1至40,或者在另一个方向上包括50至40、50至30、50至20和50至10。

[0380] 除非本文另有定义,否则结合本公开使用的科学和技术术语应具有本领域普通技术人员通常理解的含义。下面描述了示例性方法和材料,但是与本文所述的那些类似或等

同的方法和材料也可以用于本公开的实践或测试中。在冲突的情况下,以包括定义的本说明书为准。此外,除非上下文另有要求,否则单数术语应包括复数,复数术语应包括单数。贯穿本说明书和实施例,词语“具有(have)”和“包含(comprise)”或诸如“具有(has)”、“具有(having)”、“包含(comprises)”或“包含(comprising)”的变体将被理解为暗示包括所述整数或整数组,但不排除任何其他整数或整数组。除非另有说明,否则本文的元素列表的叙述包括单独或以任何组合的任何元素。本文对实施方案的叙述包括该实施方案作为单个实施方案或与本文任何其他实施方案的组合。本文提及的所有出版物、专利、专利申请和其他参考文献通过引用整体并入。通过引用并入的参考文献与说明书中包含的公开内容出现相矛盾的程度时,说明书旨在取代和/或优先于任何此类矛盾材料。尽管本文引用了许多文献,但该引用并不构成承认这些文献中的任何一个成为本领域公知常识的一部分。

[0381] 根据本公开,从属权利要求中的反向引用旨在作为一种简写,用于直接且无歧义地公开由反向引用指示的每个权利要求组合。此外,本文的标题是为了便于组织而创建的,并且不旨在以任何方式限制所要求保护的组合物和方法的范围。

[0382] 本文提供的一些蛋白质序列,例如一些融合蛋白序列,包括肽标签,例如His6标签或DYKDDDDK(SEQ ID NO:1528)标签,其可用于检测和/或纯化带标签的蛋白质,但不影响蛋白质功能。这些肽标签和另外合适的肽标签是本领域技术人员熟知的。对于本领域技术人员将显而易见的是,所公开的标签可以替代其他合适的肽标签,并且具有相同或高度相似序列但不包括此类肽标签的融合蛋白(例如,肽标签已从中切割或在没有肽标签的情况下产生)也适合于实施本公开的实施方案。

[0383] 为了更好地理解本公开,阐述了以下实施例。这些实施例仅用于说明的目的,并且不应被解释为以任何方式限制本公开的范围。

实施例

[0384] 实施例1:融合蛋白的设计与合成

[0385] 设计并构建了包含dCas9、DNMT3A、DNMT3L和KOX1KRAB(“CRISPR-off”)的融合蛋白。从N末端到C末端,蛋白质具有结构域DNMT3A-接头-DNMT3L-XTEN80-NLS-dspCas9-NLS-XTEN16-KOX1KRAB(SEQ ID NO:658和1495)。CRISPR-off质粒构建体已在Nuñez(Nuñezet al.,Cell(2021)184(9):2503-19)中描述,并且从Twist Biosciences订购。

[0386] 还构建了包含DNMT3A、3L和KOX1 KRAB的ZF融合蛋白(“ZF-off”)。构建体具有一般结构DNMT3A-接头-DNMT3L-XTEN80-NLS-ZFP结构域-NLS-XTEN16-KOX1Krab(SEQ ID NO:659和1496)。

[0387] 实施例2:用于gRNA表观遗传沉默的靶PCSK9序列的选择

[0388] 使用Benchling gRNA平台(Benchling(2021),从benchling.com检索)针对人(GRCh38)、小鼠(mm10)和猕猴(5.0)PCSK9计算设计靶向PCSK9 TSS的+/-1kb的gRNA。首先丢弃含有聚TTTT序列的gRNA。我们使用CasOFFinder进行gRNA脱靶分析(Bae et al.,Bioinformatics(2014)30(10):1473-5)。如果gRNA与每个独立物种的相应基因组构建中的多个位置匹配,则丢弃该gRNA。

[0389] 对人PCSK9 gRNA进行交叉反应性序列分析,以注释与猕猴或小鼠gRNA序列的序列错配。具体地,进行gRNA序列比对以鉴定每个核苷酸处的DNA相似性程度,包括包含至多零

个、一个或两个核苷酸错配的指导物的注释。选择最后一组226个gRNA序列用于HeLa细胞中的PCSK9初级筛选。

[0390] 实施例3:ZF靶位点的选择和用于表观遗传沉默的ZF蛋白的设计

[0391] 每个识别6bp DNA位点的两指ZFP(2F单元)文库用于设计靶向18bp DNA结合位点的更大的六指ZFP阵列。2F单元的来源是一组三指锌指蛋白,其已经使用细菌-2-杂交(B2H)选择系统选择以结合特定靶位点(Hurt etal.,PNAS(2003)100:12271-6;Maeder et al.,Mol Cell(2008)31(2):294-301)。通过产生文库中表示的6bp结合位点的所有可能的三联体组合并允许6bp靶位点之间的0或1bp来创建可靶向DNA位点的列表。为了鉴定PCSK9内的锌指靶位点,针对该列表对距TSS(人(GRCh38))+/-1kb的序列进行查询。对于每个鉴定的ZF靶位点,可以设计多种ZF蛋白。用于产生完整蛋白的六个识别螺旋的设计是通过选择双指单元并考虑许多因素进行的,例如锌指蛋白的已知结合偏好,在B2H选择系统中选择-1、2、3和6位氨基酸以结合期望的靶碱基的频率,避免在B2H中选择以结合多个不同碱基的-1、2、3和6位氨基酸,以及在可能的情况下通过匹配侧翼碱基来维持背景依赖性。完整的ZF序列衍生自天然存在的Zif268蛋白,并且选择的识别螺旋维持在其中它们在B2H中被选择的序列背景中(来自Zif268的指1-2或指2-3)。双指单元通过接头TGSQKP(SEQ ID NO:651)(其中6bp结合位点是连续的)并且通过接头TGGGGSQKP(SEQ ID NO:652)(其中1bp分隔6bp结合位点)连接。选择靶向49个不同结合位点的209个ZFP的最终集合用于HeLa细胞中的PCSK9初级筛选。

[0392] 图1显示了映射到PCSK9靶区域的gRNA和锌指蛋白的重叠。

[0393] 实施例4:HeLa细胞中的指导RNA筛选

[0394] 在HeLa细胞中进行靶向PCSK9的gRNA的初级筛选。gRNA序列作为DNA片段从Twist Biosciences订购,其中u6启动子序列在gRNA编码序列之前。

[0395] 用DNA形式的gRNA和CRISPR-off转染HeLa细胞。在含有补充有10%胎牛血清(Thermo Fisher目录号A4766)v/v、1x GlutaMAX™(Thermo Fisher目录号35050061)和1x青霉素-链霉素(Thermo Fisher目录号15140122)的DMEM(Thermo Fisher目录号11-965-092)的标准培养基中以12,000个HeLa细胞每孔(ATCC目录号CCL-2)接种6个96孔板(Sigma-Aldrich目录号M2936)。铺板后,使细胞在37℃、5%CO_2的培养箱中生长24小时。将25ng各gRNA片段和50ng CRISPR-OFF质粒(SEQ ID NO:658)重悬于DPBS缓冲液(Thermo Fisher目录号14190144)中至7.5ng/μL的浓度。另外,还将10ng的EF1a:嘌呤霉素抗性质粒添加到转染混合物中以实现85ng DNA的总有效载荷。按照制造商的说明,通过将重悬的DNA组分添加到Mirus® TransIT®-LT1转染试剂(Mirus目录号MIR2300)中来产生转染混合物。将10μL每种转染混合物一式两份加入总共六个筛选板中。使用的阳性对照是CRISPRi(dCas9-KRAB),其具有两个gRNA靶向TSS附近的位点。使用的两种CRISPRi阳性对照gRNA是gRNA004和gRNA005,如gRNA序列表中注释的。这些对照条件在初级筛选数据表中分别称为“CRISPRi-1”和“CRISPRi-2”。阴性对照是不含gRNA的CRISPR-off、具有非PCSK9基因座(CD151)靶向gRNA的CRISPR-off和空载体(pUC19;NEB目录号N3041S)。

[0396] 转染后24小时,进行嘌呤霉素抗性选择。将细胞培养基完全吸出,并将所有孔用DPBS缓冲液(Thermo Fisher目录号14190144)洗涤3次,并将200μL 1μg/μL嘌呤霉素添加到所有筛板孔中。

[0397] 转染后48小时,传代细胞。将细胞培养基完全吸出,并将所有孔用DPBS缓冲液(Thermo Fisher目录号14190144)洗涤3次。通过添加25μL胰蛋白酶-EDTA(0.25%)(Thermo Fisher目录号25200056)在37℃培养箱中孵育5分钟来酶促提取细胞。将胰蛋白酶化的细胞以1:8重悬于新鲜标准培养基中,并在更换培养基72小时后以1:4的比例重新铺板。

[0398] 为了测量分泌的PCSK9蛋白的水平,在培养基更换后24小时收获培养基,并根据制造商的建议使用Promega细胞滴度Glo™方案(目录号G7570)测定细胞板进行相对细胞计数。使用BioLegend的LEGEND MAX™人PCSK9 ELISA试剂盒(目录号443107)评估PCSK9蛋白水平。将收获的培养基铺板,并完全根据制造商的建议进行所有后续步骤。在Perkin Elmer® VICTOR®Nivo™F仪器上进行450nm处的最终平板读数。使用GraphPad Prism软件将函数拟合到标准曲线并插入未知数。PCSK9 ELISA结果通过细胞滴度Glo®测定(Promega目录号G7571)结果归一化,以校正任何孔间细胞数量变异性。

[0399] 测试了超过200个gRNA,其中40个被鉴定为靠前的序列(图2;靠前的序列指定为较黑的圆圈)。测试的gRNA的序列和功效示于表7(SEQ:SEQ ID NO)。相对PCSK9分泌("对照PCSK9%")表示经处理样品的平均PCSK9蛋白水平,其表示为相对于所有非靶向gRNA(CD151)阴性对照条件的平均值的百分比。在用多种gRNA候选治疗方法处理的细胞中观察到PCSK9的稳健沉默(阴性对照水平的30-40%)。选择具有最佳PCSK9蛋白敲除的前40个gRNA作为sgRNA进行进一步随访研究。

[0400] 表7.测试的gRNA的靶向序列

[0401]

gRNA No.或条件	SEQ	靶向序列	相对于TSS的bp	GRCh38序列起始坐标	对照PCSK9%
CD151 gRNA	N/A	N/A	N/A	N/A	100
无 gRNA	N/A	N/A	N/A	N/A	72.51
pUC	N/A	N/A	N/A	N/A	119.55
CRISPRi-1	N/A	N/A	N/A	N/A	86.89
CRISPRi-2	N/A	N/A	N/A	N/A	64.09
gRNA001	1262	GGUGCUAGCCUUGCGUUCCG	431	55039960	43.85
gRNA002	1263	CUGGCCGAAGCACCCGAGCA	462	55039991	74.87
gRNA003	1264	UGCGGAAACCUUCUAGGGUG	0	55039529	27.33

[0402]

gRNA No. 或条件	SEQ	靶向序列	相对于 TSS 的 bp	GRCh38 序列起始坐标	对照 PCSK9%
gRNA004	1265	GCGGAAACCUUCUAGGGUGU	-1	55039528	34.93
gRNA005	1266	UCAAGCACCCACACCCUAGA	-11	55039518	41.44
gRNA006	1267	GGGUGUGGGUGCUUGACGCC	-15	55039514	42.41
gRNA007	1268	GGUGUGGGUGCUUGACGCCU	-16	55039513	31.11
gRNA008	1269	GUGUGGGUGCUUGACGCCUG	-17	55039512	36.13
gRNA009	1270	ACUGCCUGGCUCACUCCUCC	81	55039610	21.34
gRNA010	1271	UCACGCCACCAGAGCCCCAU	-48	55039481	52.20
gRNA011	1272	AUCGUCCGAUGGGGCUCUGG	-56	55039473	31.00
gRNA012	1273	AGGAUCGUCCGAUGGGGCUC	-59	55039470	38.05
gRNA013	1274	UCAGAUAGGAUCGUCCGAUG	-65	55039464	41.70
gRNA014	1275	GCGGCUCCCAGCUCCCAGCC	140	55039669	85.56
gRNA015	1276	CGGAAUCCUGGCUGGGAGCU	149	55039678	43.48
gRNA016	1277	GCGGAAUCCUGGCUGGGAGC	150	55039679	75.20
gRNA017	1278	GGGCGCGCGGAAUCCUGGCU	156	55039685	53.76
gRNA018	1279	GGGGCGCGCGGAAUCCUGGC	157	55039686	56.46
gRNA019	1280	UGAGGUCUUUGCAAACAAAG	-919	55038610	76.59
gRNA020	1281	CCAGCACCUAGAUUCAGAGC	-875	55038654	60.05
gRNA021	1282	CCUGCUCUGAAUCUAGGUGC	-872	55038657	56.48
gRNA022	1283	UCGAUACUGGGAAGAAACAA	-798	55038731	67.82
gRNA023	1284	CUGGAAGGGCUGUCGAUACU	-786	55038743	59.17
gRNA024	1285	UCUGGAAGGGCUGUCGAUAC	-785	55038744	63.62
gRNA025	1286	GAGGCUUGCUCUUUCUGGAA	-772	55038757	58.91
gRNA026	1287	UGAGGCUUGCUCUUUCUGGA	-771	55038758	74.33
gRNA027	1288	GACAUGAGGCUUGCUCUUUC	-767	55038762	87.93
gRNA028	1289	UGUACAUGUGGCAUGACAUG	-753	55038776	68.47
gRNA029	1290	CAUGCCACAUGUACAAUCUG	-748	55038781	88.26
gRNA030	1291	CACAUGUACAAUCUGAGGCC	-743	55038786	79.41
gRNA031	1292	CUGGCCUCAGAUUGUACAUG	-741	55038788	106.36
gRNA032	1293	AAAAGGGGAAAGAGAGCUCC	-722	55038807	91.18
gRNA033	1294	CCAGGCAGGAGGAUGAAAAG	-707	55038822	89.34
gRNA034	1295	ACCAGGCAGGAGGAUGAAAA	-706	55038823	79.11
gRNA035	1296	UACCAGGCAGGAGGAUGAAA	-705	55038824	82.01
gRNA036	1297	CCUCCUGCCUGGUACACAAU	-699	55038830	69.28
gRNA037	1298	CCUAUUGUGUACCAGGCAGG	-696	55038833	82.57
gRNA038	1299	ACACCUAUUGUGUACCAGGC	-693	55038836	58.95
gRNA039	1300	GGUACACAAUAGGUGUUUAC	-689	55038840	81.55
gRNA040	1301	GUAAACACCUAUUGUGUACC	-689	55038840	89.49
gRNA041	1302	UCCAGUUGAUUUCUUGAACA	-659	55038870	69.65
gRNA042	1303	ACCAUGUUCAAGAAAUCAAC	-655	55038874	73.51
gRNA043	1304	UCUUGAACAUGGUGUGUAAA	-648	55038881	86.96
gRNA044	1305	UCUUUGCAAAUUGAAUCUUC	-623	55038906	80.70
gRNA045	1306	AGUUUGCAAAGACGUCAUAU	-559	55038970	85.13
gRNA046	1307	GAUUUAUACUACAAAGAUUC	-531	55038998	74.88
gRNA047	1308	AGUUGGUAAGGUCAGUGUGC	-454	55039075	66.05

[0403]

gRNA No. 或条件	SEQ	靶向序列	相对于 TSS 的 bp	GRCh38 序列起始坐标	对照 PCSK9%
gRNA048	1309	GUUGGUAAGGUCAGUGUGCA	-453	55039076	60.64
gRNA049	1310	GCAGGGUGCAUAAAGGGCAG	-436	55039093	60.72
gRNA050	1311	GGUGCAUAAAGGGCAGAGGC	-432	55039097	92.12
gRNA051	1312	GCAUAAAGGGCAGAGGCCGG	-429	55039100	53.71
gRNA052	1313	UUUAGAAGGCUGCCAGGUUA	-392	55039137	56.64
gRNA053	1314	GCCCACCGAAUUCUUUCCAC	-366	55039163	62.84
gRNA054	1315	AAAGAAUUCGGUGGGCAGCG	-362	55039167	78.78
gRNA055	1316	CUUCUGAAUCAAUCCUACUG	-333	55039196	52.04
gRNA056	1317	CUGGUCAGCAGGAGACAAGG	-328	55039200	90.01
gRNA057	1318	GAUUGAUUCAGAAGUCUCAC	-327	55039202	64.68
gRNA058	1319	CAGAAGUCUCACUGGUCAGC	-319	55039210	101.68
gRNA059	1320	UCACUGGUCAGCAGGAGACA	-311	55039218	75.12
gRNA060	1321	GCAGGAGACAAGGUGGACCC	-301	55039228	43.48
gRNA061	1322	GGACCCAGGAAACACUGAAA	-287	55039242	50.45
gRNA062	1323	CCCAGGAAACACUGAAAAGG	-284	55039245	104.49
gRNA063	1324	CCAGGAAACACUGAAAAGGU	-283	55039246	74.99
gRNA064	1325	AAACACUGAAAAGGUGGGCC	-278	55039251	78.05
gRNA065	1326	UGGAGUCUGGCAUCCCACGC	-248	55039281	86.50
gRNA066	1327	GGAGUCUGGCAUCCCACGCA	-247	55039282	64.96
gRNA067	1328	CGGGAGAGGAGGAGCCCCUA	-217	55039312	51.16
gRNA068	1329	AGGAGGAGCCCCUAGGGCGC	-211	55039318	49.85
gRNA069	1330	AAGGCAGGCCGGCGCCCUAG	-200	55039329	57.25
gRNA070	1331	GAAGGCAGGCCGGCGCCCUA	-199	55039330	71.39
gRNA071	1332	GGAAGGCAGGCCGGCGCCCU	-198	55039331	55.58
gRNA072	1333	AACUGGGCUGGAAGGCAGGC	-189	55039340	81.36
gRNA073	1334	GCCUGCCUUCCAGCCCAGUU	-189	55039340	68.77
gRNA074	1335	UCCUAACUGGGCUGGAAGGC	-185	55039344	67.81
gRNA075	1336	CUUCCAGCCCAGUUAGGAUU	-183	55039346	58.27
gRNA076	1337	UUCCAGCCCAGUUAGGAUUU	-182	55039347	53.06
gRNA077	1338	CAAAUCCUAACUGGGCUGGA	-181	55039348	31.55
gRNA078	1339	CUCCCAAAUCCUAACUGGGC	-177	55039352	52.30
gRNA079	1340	AAAACUCCCAAAUCCUAACU	-173	55039356	52.38
gRNA080	1341	AAAAACUCCCAAAUCCUAAC	-172	55039357	67.62
gRNA081	1342	AGCGUCAGAUUACGCGCAGA	-145	55039384	58.52
gRNA082	1343	CAGCGUCAGAUUACGCGCAG	-144	55039385	76.23
gRNA083	1344	GCGCGUAAUCUGACGCUGUU	-142	55039387	73.94
gRNA084	1345	CGCGUAAUCUGACGCUGUUU	-141	55039388	71.95
gRNA085	1346	GCGUAAUCUGACGCUGUUUG	-140	55039389	77.63
gRNA086	1347	UAAUCUGACGCUGUUUGGGG	-137	55039392	70.70
gRNA087	1348	AAUCUGACGCUGUUUGGGGA	-136	55039393	48.40
gRNA088	1349	GACGCUGUUUGGGGAGGGCG	-131	55039398	64.83
gRNA089	1350	CGAAACCUGAUCCUCCAGUC	-107	55039422	56.51
gRNA090	1351	GAAACCUGAUCCUCCAGUCC	-106	55039423	45.68
gRNA091	1352	AAACCUGAUCCUCCAGUCCG	-105	55039424	50.44

[0404]

gRNA No. 或条件	SEQ	靶向序列	相对于 TSS 的 bp	GRCh38 序列起始坐标	对照 PCSK9%
gRNA092	1353	CGGACUGGAGGAUCAGGUUU	-105	55039424	30.88
gRNA093	1354	AACCUGAUCCUCCAGUCCGG	-104	55039425	29.56
gRNA094	1355	AACCCCCGGACUGGAGGAUC	-99	55039430	33.66
gRNA095	1356	UAACGGAACCCCCGGACUGG	-93	55039436	35.57
gRNA096	1357	CAUUAACGGAACCCCCGGAC	-90	55039439	30.87
gRNA097	1358	UUAAACAUUAACGGAACCCC	-85	55039444	58.22
gRNA098	1359	CCGUUAAUGUUUAAUCAGAU	-79	55039450	58.99
gRNA099	1360	CCUAUCUGAUUAAACAUUAA	-76	55039453	45.83
gRNA100	1361	AAUCAGAUAGGAUCGUCCGA	-67	55039462	47.05
gRNA101	1362	AUCAGAUAGGAUCGUCCGAU	-66	55039463	57.76
gRNA102	1363	UGGCGUGAUCUGCGCGCCCC	-36	55039493	41.66
gRNA103	1364	GUCGCUGCGGAAACCUUCUA	5	55039534	52.14
gRNA104	1365	CGUCGCUGCGGAAACCUUCU	6	55039535	54.97
gRNA105	1366	AGGUUUCCGCAGCGACGUCG	9	55039538	61.69
gRNA106	1367	CAGCGACGUCGAGGCGCUCA	18	55039547	54.52
gRNA107	1368	UGAGCGCCUCGACGUCGCUG	18	55039547	35.11
gRNA108	1369	GUCGAGGCGCUCAUGGUUGC	25	55039554	25.10
gRNA109	1370	GAGGCGCUCAUGGUUGCAGG	28	55039557	43.19
gRNA110	1371	AGGCGCUCAUGGUUGCAGGC	29	55039558	29.00
gRNA111	1372	AGUUCAGGGUCUGAGCCUGG	43	55039571	39.04
gRNA112	1373	CGGGCGCCGCCGUUCAGUUC	48	55039577	45.13
gRNA113	1374	GGGCGCCGCCGUUCAGUUCA	49	55039578	32.30
gRNA114	1375	UCAGACCCUGAACUGAACGG	57	55039586	55.95
gRNA115	1376	GGCUCAGACCCUGAACUGAA	60	55039589	48.59
gRNA116	1377	UUCAGUUCAGGGUCUGAGCC	60	55039589	57.38
gRNA117	1378	UGAGCCUGGAGGAGUGAGCC	74	55039603	45.58
gRNA118	1379	AGUGAGCCAGGCAGUGAGAC	86	55039615	46.98
gRNA119	1380	GCCAGGCAGUGAGACUGGCU	91	55039620	62.91
gRNA120	1381	CCAGGCAGUGAGACUGGCUC	92	55039621	54.17
gRNA121	1382	CCCGAGCCAGUCUCACUGCC	95	55039624	48.29
gRNA122	1383	GGCAGUGAGACUGGCUCGGG	95	55039624	34.80
gRNA123	1384	GCAGUGAGACUGGCUCGGGC	96	55039625	53.96
gRNA124	1385	UGAGACUGGCUCGGGCGGGC	100	55039629	59.99
gRNA125	1386	GAGACUGGCUCGGGCGGGCC	101	55039630	34.59
gRNA126	1387	GGGACGCGUCGUUGCAGCAG	121	55039650	38.46
gRNA127	1388	GCUGCUGCAACGACGCGUCC	122	55039651	44.41
gRNA128	1389	UGAAGGGGCGCGCGGAAUCC	161	55039690	28.00
gRNA129	1390	AGGGCGCGUGAAGGGGCGCG	169	55039698	64.66
gRNA130	1391	CAGGAGCAGGGCGCGUGAAG	176	55039705	52.46
gRNA131	1392	UCAGGAGCAGGGCGCGUGAA	177	55039706	61.68
gRNA132	1393	UUCAGGAGCAGGGCGCGUGA	178	55039707	48.06
gRNA133	1394	GGAGCUGAAGUUCAGGAGCA	188	55039717	62.68
gRNA134	1395	AGGAGCUGAAGUUCAGGAGC	189	55039718	66.77
gRNA135	1396	CUGUGCAGGAGCUGAAGUUC	195	55039724	47.35

[0405]

gRNA No. 或条件	SEQ	靶向序列	相对于 TSS 的 bp	GRCh38 序列起始坐标	对照 PCSK9%
gRNA136	1397	GCACAGUCCUCCCCACCGCA	209	55039738	40.95
gRNA137	1398	UGCGGUGGGGAGGACUGUGC	209	55039738	51.32
gRNA138	1399	CCUCCCCACCGCAAGGCUCA	216	55039745	76.44
gRNA139	1400	CCUUGAGCCUUGCGGUGGGG	219	55039748	71.95
gRNA140	1401	GCGCCUUGAGCCUUGCGGUG	222	55039751	63.12
gRNA141	1402	GGCGCCUUGAGCCUUGCGGU	223	55039752	64.08
gRNA142	1403	CGGCGCCUUGAGCCUUGCGG	224	55039753	68.16
gRNA143	1404	CGCAAGGCUCAAGGCGCCGC	225	55039754	58.50
gRNA144	1405	GGCUCAAGGCGCCGCCGGCG	230	55039759	41.31
gRNA145	1406	AGGCCGUGCGCGGUCCACGC	247	55039776	46.31
gRNA146	1407	GUGGACCGCGCACGGCCUCU	249	55039778	52.25
gRNA147	1408	GGAGACCUAGAGGCCGUGCG	257	55039786	62.38
gRNA148	1409	CAGGACAGCAACCUCUCCCC	281	55039810	59.44
gRNA149	1410	UGGGCACCGUCAGCUCCAGG	310	55039839	54.56
gRNA150	1411	CCGUCAGCUCCAGGCGGUCC	316	55039845	47.26
gRNA151	1412	CCAGGACCGCCUGGAGCUGA	319	55039848	71.17
gRNA152	1413	UCAGCUCCAGGCGGUCCUGG	319	55039848	46.68
gRNA153	1414	CAGCGGCCACCAGGACCGCC	328	55039857	67.26
gRNA154	1415	CUGCUGCUCCUGGGUCCCGC	366	55039895	54.47
gRNA155	1416	CACGGGCGCCCGCGGGACCC	377	55039906	48.18
gRNA156	1417	UCCCGCGGGCGCCCGUGCGC	380	55039909	53.18
gRNA157	1418	CGCGGGCGCCCGUGCGCAGG	383	55039912	61.58
gRNA158	1419	UCCUGCGCACGGGCGCCCGC	384	55039913	67.30
gRNA159	1420	CUCCUGCGCACGGGCGCCCG	385	55039914	48.64
gRNA160	1421	CGCCCGUGCGCAGGAGGACG	389	55039918	84.11
gRNA161	1422	CGUGCGCAGGAGGACGAGGA	393	55039922	86.55
gRNA162	1423	GUCCUCGUCCUCCUGCGCAC	394	55039923	54.45
gRNA163	1424	CGUCCUCGUCCUCCUGCGCA	395	55039924	80.30
gRNA164	1425	GGACGAGGACGGCGACUACG	404	55039933	69.15
gRNA165	1426	GGACGGCGACUACGAGGAGC	410	55039939	62.54
gRNA166	1427	GCUAGCCUUGCGUUCCGAGG	434	55039963	60.30
gRNA167	1428	GCCUUGCGUUCCGAGGAGGA	438	55039967	57.71
gRNA168	1429	GCCGUCCUCCUCGGAACGCA	442	55039971	62.04
gRNA169	1430	GCGUUCCGAGGAGGACGGCC	443	55039972	79.68
gRNA170	1431	UUCGGCCAGGCCGUCCUCCU	451	55039980	40.06
gRNA171	1432	CGUGCUCGGGUGCUUCGGCC	464	55039993	59.35
gRNA172	1433	GUGGCUGUGGUUCCGUGCUC	477	55040006	45.91
gRNA173	1434	GGUGGCUGUGGUUCCGUGCU	478	55040007	70.81
gRNA174	1435	GCAGCGGUGGAAGGUGGCUG	490	55040019	72.70
gRNA175	1436	CACCUUCCACCGCUGCGCCA	494	55040023	86.51
gRNA176	1437	CUUGGCGCAGCGGUGGAAGG	496	55040025	76.28
gRNA177	1438	CACCUUGGCGCAGCGGUGGA	499	55040028	66.94
gRNA178	1439	UCCACCGCUGCGCCAAGGUG	499	55040028	48.71
gRNA179	1440	CCACCGCUGCGCCAAGGUGC	500	55040029	60.24

[0406]

gRNA No. 或条件	SEQ	靶向序列	相对于 TSS 的 bp	GRCh38 序列起始坐标	对照 PCSK9%
gRNA180	1441	CCCGCACCUUGGCGCAGCGG	503	55040032	44.64
gRNA181	1442	GGGCGAACCCGCAGCCGGGA	541	55040070	46.89
gRNA182	1443	UCCCGGCUGCGGGUUCGCCC	541	55040070	57.20
gRNA183	1444	CACCGCACCGUCCCGGCUGC	551	55040080	76.92
gRNA184	1445	GCACCGCACCGUCCCGGCUG	552	55040081	85.00
gRNA185	1446	GAAACAGCACCGCACCGUCC	558	55040087	75.31
gRNA186	1447	GGUGCGGUGCUGUUUCCUCU	562	55040091	51.49
gRNA187	1448	GUGCGGUGCUGUUUCCUCUC	563	55040092	57.43
gRNA188	1449	GGGGGAAACUGAGGCCCGAG	577	55040105	55.16
gRNA189	1450	AGUUUCCCCCCAUGUAAGAG	590	55040119	45.91
gRNA190	1451	CCCCCAUGUAAGAGAGGAAG	596	55040125	63.89
gRNA191	1452	CACUUCCUCUCUUACAUGGG	598	55040127	63.57
gRNA192	1453	CCACUUCCUCUCUUACAUGG	599	55040128	66.47
gRNA193	1454	UAAGAGAGGAAGUGGAGUGC	604	55040133	63.94
gRNA194	1455	GUCGCCGAGGGCUCUUCGCU	626	55040155	55.19
gRNA195	1456	CGUGCCAAGCGAAGAGCCCU	633	55040162	66.66
gRNA196	1457	CUCUUCGCUUGGCACGAUCU	637	55040166	57.10
gRNA197	1458	UCUUCGCUUGGCACGAUCUU	638	55040167	44.24
gRNA198	1459	CUUCGCUUGGCACGAUCUUG	639	55040168	50.40
gRNA199	1460	GGCACGAUCUUGGGGACUGC	647	55040176	73.29
gRNA200	1461	GAUCUUGGGGACUGCAGGCA	652	55040181	56.39
gRNA201	1462	CUUGGGGACUGCAGGCAAGG	655	55040184	66.89
gRNA202	1463	GGACUGCAGGCAAGGCGGCG	660	55040189	61.44
gRNA203	1464	GACUGCAGGCAAGGCGGCGG	661	55040190	49.55
gRNA204	1465	UGCAGGCAAGGCGGCGGGGG	664	55040193	41.85
gRNA205	1466	GGCAAGGCGGCGGGGGAGGA	668	55040197	95.50
gRNA206	1467	GCAAGGCGGCGGGGGAGGAC	669	55040198	120.71
gRNA207	1468	GUGGGGAGCACGGUGGAGAG	694	55040223	73.17
gRNA208	1469	UGGGGAGCACGGUGGAGAGC	695	55040224	61.04
gRNA209	1470	GGGGAGCACGGUGGAGAGCG	696	55040225	79.49
gRNA210	1471	AGCACGGUGGAGAGCGGGGA	700	55040229	42.81
gRNA211	1472	CGGUGGAGAGCGGGGACGGC	704	55040233	79.46
gRNA212	1473	CGUGCGGCUGCGCUAUUCAG	748	55040277	39.48
gRNA213	1474	GUGCGGCUGCGCUAUUCAGU	749	55040278	56.25
gRNA214	1475	GGCUGCGCUAUUCAGUGGGA	753	55040282	82.51
gRNA215	1476	CUAUUCAGUGGGAAGGUUCG	760	55040289	77.53
gRNA216	1477	UAUUCAGUGGGAAGGUUCGC	761	55040290	71.30
gRNA217	1478	AUUCAGUGGGAAGGUUCGCG	762	55040291	68.82
gRNA218	1479	AGUGGGAAGGUUCGCGGGGU	766	55040295	73.55
gRNA219	1480	GUGGGAAGGUUCGCGGGGUU	767	55040296	66.13
gRNA220	1481	AGGGCGAGCAGAGCACUGCC	808	55040337	68.85
gRNA221	1482	UUUCUGCCUCGCCGCGGCAC	853	55040382	65.59
gRNA222	1483	CUGCCUCGCCGCGGCACAGG	856	55040385	76.58
gRNA223	1484	UGCCUCGCCGCGGCACAGGU	857	55040386	75.75

[0407]

gRNA No.或条件	SEQ	靶向序列	相对于TSS的bp	GRCh38序列起始坐标	对照PCSK9%
gRNA224	1485	CACCCACCUGUGCCGCGGCG	862	55040391	81.79
gRNA225	1486	UCCUUCACCCACCUGUGCCG	867	55040396	70.86
gRNA226	1487	UGGGUGAAGGAGUGAAUGCC	876	55040405	57.69

[0408] 发现表现最佳的gRNA与PCSK9基因转录起始位点紧密对齐(图3)。

[0409] 在该初级筛选之后,所述前40个gRNA以RNA形式进行次级筛选。化学合成前40个指导物,并与体外转录的编码CRISPR-off、CRISPRi或WT Cas9构建体的mRNA共转染。在转染后7天和28天测量分泌的PCSK9水平。

[0410] 为了产生体外转录的CRISPR-off、CRISPRi和WT Cas9效应子mRNA,使用NEB®的MfeI限制性酶(目录号R3589S)线性化编码这些蛋白质的质粒构建体。根据制造商的说明书,使用1μg线性化模板通过CellScript的T7 mScript™标准mRNA生产系统(目录号C-MSC100625)建立体外转录反应。所得RNA在5'末端具有帽1结构,并且是3'聚腺苷酸化的。使用Qiagen的RNeasy®迷你试剂盒(目录号74104)纯化转录的RNA。

[0411] 从Integrated DNA Technologies获得使用标准脱盐纯化的末端修饰的sgRNA。每个指导物的5'末端的三个核苷酸和3'末端的三个核苷酸是2'-O-甲基修饰的。3'端的三个核苷间键和5'端的三个核苷间键是硫代磷酸酯核苷间键(表8;SEQ ID NO)。在表中,mX(即mA、mC、mG或mU)表示2′-O-甲基修饰的核糖核苷,rX(即rA、rC、rG或rU)表示天然核糖核苷,并且*表示硫代磷酸酯键。所有非硫代磷酸酯键的核苷间键都是磷酸酯键。

[0412] 表8.次级HeLa细胞筛选的前40个gRNA的靶向序列

[0413]

gRNA No.	SEQ	靶向序列	修饰的gRNA No.	完全修饰的RNA靶向序列
gRNA009	1270	ACUGCCUGGCUCACUCCUCC	modRNA009	mA*mC*mU*rGrCrCrUrGrGrCrUrCrArCrUrCrCrUrCrCrGrUrUrUrArArGrArGrCrUrArArGrCrUrGrGrArArArCrArGrCrArUrArGrCrArArGrUrUrUrArArArUrArArGrGrCrUrArGrUrCrCrGrUrUrArUrCrArArCrUrUrGrArArArArArGrUrGrGrCrArCrCrGrArGrUrCrGrGrUrGrCrUrUrU*mU*mU*mU

[0414]

gRNA No.	SEQ	靶向序列	修饰的 gRNA No.	完全修饰的 RNA 靶向序列
gRNA 108	1369	GUCGAGGCGCU CAUGGUUGC	modRNA108	mG*mU*mC*rGrArGrGrCrGrCrUrCr ArUrGrGrUrUrGrCrGrUrUrUrArArGr ArGrCrUrArArGrCrUrGrGrArArArCr ArGrCrArUrArGrCrArArGrUrUrUrAr ArArUrArArGrGrCrUrArGrUrCrCrGr UrUrArUrCrArArCrUrUrGrArArArAr ArGrUrGrGrCrArCrCrGrArGrUrCrGr GrUrGrCrUrUrU*mU*mU*mU
gRNA 003	1264	UGCGGAAACCU UCUAGGGUG	modRNA003	mU*mG*mC*rGrGrArArArCrCrUrUr CrUrArGrGrGrUrGrGrUrUrUrArArGr ArGrCrUrArArGrCrUrGrGrArArArCr ArGrCrArUrArGrCrArArGrUrUrUrAr ArArUrArArGrGrCrUrArGrUrCrCrGr UrUrArUrCrArArCrUrUrGrArArArAr ArGrUrGrGrCrArCrCrGrArGrUrCrGr GrUrGrCrUrUrU*mU*mU*mU
gRNA 128	1389	UGAAGGGGCGC GCGGAAUCC	modRNA128	mU*mG*mA*rArGrGrGrGrCrGrCrGr CrGrGrArArUrCrCrGrUrUrUrArArGr ArGrCrUrArArGrCrUrGrGrArArArCr ArGrCrArUrArGrCrArArGrUrUrUrAr ArArUrArArGrGrCrUrArGrUrCrCrGr UrUrArUrCrArArCrUrUrGrArArArAr ArGrUrGrGrCrArCrCrGrArGrUrCrGr GrUrGrCrUrUrU*mU*mU*mU
gRNA 110	1371	AGGCGCUCAUG GUUGCAGGC	modRNA110	mA*mG*mG*rCrGrCrUrCrArUrGrGr UrUrGrCrArGrGrCrGrUrUrUrArArGr ArGrCrUrArArGrCrUrGrGrArArArCr ArGrCrArUrArGrCrArArGrUrUrUrAr ArArUrArArGrGrCrUrArGrUrCrCrGr UrUrArUrCrArArCrUrUrGrArArArAr ArGrUrGrGrCrArCrCrGrArGrUrCrGr GrUrGrCrUrUrU*mU*mU*mU
gRNA 093	1354	AACCUGAUCCU CCAGUCCGG	modRNA093	mA*mA*mC*rCrUrGrArUrCrCrUrCr CrArGrUrCrCrGrGrGrUrUrUrArArGr ArGrCrUrArArGrCrUrGrGrArArArCr ArGrCrArUrArGrCrArArGrUrUrUrAr ArArUrArArGrGrCrUrArGrUrCrCrGr UrUrArUrCrArArCrUrUrGrArArArAr ArGrUrGrGrCrArCrCrGrArGrUrCrGr GrUrGrCrUrUrU*mU*mU*mU
gRNA 096	1357	CAUUAACGGAA CCCCCGGAC	modRNA096	mC*mA*mU*rUrArArCrGrGrArArCr CrCrCrCrGrGrArCrGrUrUrUrArArGr ArGrCrUrArArGrCrUrGrGrArArArCr ArGrCrArUrArGrCrArArGrUrUrUrAr ArArUrArArGrGrCrUrArGrUrCrCrGr UrUrArUrCrArArCrUrUrGrArArArAr ArGrUrGrGrCrArCrCrGrArGrUrCrGr GrUrGrCrUrUrU*mU*mU*mU

[0415]

gRNA No.	SEQ	靶向序列	修饰的 gRNA No.	完全修饰的 RNA 靶向序列
gRNA 092	1353	CGGACUGGAGG AUCAGGUUU	modRNA092	mC*mG*mG*rArCrUrGrGrArGrGrAr UrCrArGrGrUrUrUrGrUrUrUrArArGr ArGrCrUrArArGrCrUrGrGrArArArCr ArGrCrArUrArGrCrArArGrUrUrUrAr ArArUrArArGrGrCrUrArGrUrCrCrGr UrUrArUrCrArArCrUrUrGrArArArAr ArGrUrGrGrCrArCrCrGrArGrUrCrGr GrUrGrCrUrUrU*mU*mU*mU
gRNA 011	1272	AUCGUCCGAUG GGGCUCUGG	modRNA011	mA*mU*mC*rGrUrCrCrGrArUrGrGr GrGrCrUrCrUrGrGrGrUrUrUrArArGr ArGrCrUrArArGrCrUrGrGrArArArCr ArGrCrArUrArGrCrArArGrUrUrUrAr ArArUrArArGrGrCrUrArGrUrCrCrGr UrUrArUrCrArArCrUrUrGrArArArAr ArGrUrGrGrCrArCrCrGrArGrUrCrGr GrUrGrCrUrUrU*mU*mU*mU
gRNA 007	1268	GGUGUGGGUGC UUGACGCCU	modRNA007	mG*mG*mU*rGrUrGrGrGrUrGrCrUr UrGrArCrGrCrCrUrGrUrUrUrArArGr ArGrCrUrArArGrCrUrGrGrArArArCr ArGrCrArUrArGrCrArArGrUrUrUrAr ArArUrArArGrGrCrUrArGrUrCrCrGr UrUrArUrCrArArCrUrUrGrArArArAr ArGrUrGrGrCrArCrCrGrArGrUrCrGr GrUrGrCrUrUrU*mU*mU*mU
gRNA 077	1338	CAAAUCCUAAC UGGGCUGGA	modRNA077	mC*mA*mA*rArUrCrCrUrArArCrUr GrGrGrCrUrGrGrArGrUrUrUrArArGr ArGrCrUrArArGrCrUrGrGrArArArCr ArGrCrArUrArGrCrArArGrUrUrUrAr ArArUrArArGrGrCrUrArGrUrCrCrGr UrUrArUrCrArArCrUrUrGrArArArAr ArGrUrGrGrCrArCrCrGrArGrUrCrGr GrUrGrCrUrUrU*mU*mU*mU
gRNA 113	1374	GGGCGCCGCCG UUCAGUUCA	modRNA113	mG*mG*mG*rCrGrCrCrGrCrCrGrUr UrCrArGrUrUrCrArGrUrUrUrArArGr ArGrCrUrArArGrCrUrGrGrArArArCr ArGrCrArUrArGrCrArArGrUrUrUrAr ArArUrArArGrGrCrUrArGrUrCrCrGr UrUrArUrCrArArCrUrUrGrArArArAr ArGrUrGrGrCrArCrCrGrArGrUrCrGr GrUrGrCrUrUrU*mU*mU*mU
gRNA 094	1355	AACCCCCGGAC UGGAGGAUC	modRNA094	mA*mA*mC*rCrCrCrCrGrGrArCrUr GrGrArGrGrArUrCrGrUrUrUrArArGr ArGrCrUrArArGrCrUrGrGrArArArCr ArGrCrArUrArGrCrArArGrUrUrUrAr ArArUrArArGrGrCrUrArGrUrCrCrGr UrUrArUrCrArArCrUrUrGrArArArAr ArGrUrGrGrCrArCrCrGrArGrUrCrGr GrUrGrCrUrUrU*mU*mU*mU

[0416]

gRNA No.	SEQ	靶向序列	修饰的 gRNA No.	完全修饰的 RNA 靶向序列
gRNA 125	1386	GAGACUGGCUC GGGCGGGCC	modRNA125	mG*mA*mG*rArCrUrGrGrCrUrCrGr GrGrCrGrGrGrCrCrGrUrUrUrArArGr ArGrCrUrArArGrCrUrGrGrArArArCr ArGrCrArUrArGrCrArArGrUrUrUrAr ArArUrArArGrGrCrUrArGrUrCrCrGr UrUrArUrCrArArCrUrUrGrArArArAr ArGrUrGrGrCrArCrCrGrArGrUrCrGr GrUrGrCrUrUrU*mU*mU*mU
gRNA 122	1383	GGCAGUGAGAC UGGCUCGGG	modRNA122	mG*mG*mC*rArGrUrGrArGrArCrUr GrGrCrUrCrGrGrGrGrUrUrUrArArGr ArGrCrUrArArGrCrUrGrGrArArArCr ArGrCrArUrArGrCrArArGrUrUrUrAr ArArUrArArGrGrCrUrArGrUrCrCrGr UrUrArUrCrArArCrUrUrGrArArArAr ArGrUrGrGrCrArCrCrGrArGrUrCrGr GrUrGrCrUrUrU*mU*mU*mU
gRNA 004	1265	GCGGAAACCUU CUAGGGUGU	modRNA004	mG*mC*mG*rGrArArArCrCrUrUrCr UrArGrGrGrUrGrUrGrUrUrUrArArGr ArGrCrUrArArGrCrUrGrGrArArArCr ArGrCrArUrArGrCrArArGrUrUrUrAr ArArUrArArGrGrCrUrArGrUrCrCrGr UrUrArUrCrArArCrUrUrGrArArArAr ArGrUrGrGrCrArCrCrGrArGrUrCrGr GrUrGrCrUrUrU*mU*mU*mU
gRNA 107	1368	UGAGCGCCUCG ACGUCGCUG	modRNA107	mU*mG*mA*rGrCrGrCrCrUrCrGrAr CrGrUrCrGrCrUrGrGrUrUrUrArArGr ArGrCrUrArArGrCrUrGrGrArArArCr ArGrCrArUrArGrCrArArGrUrUrUrAr ArArUrArArGrGrCrUrArGrUrCrCrGr UrUrArUrCrArArCrUrUrGrArArArAr ArGrUrGrGrCrArCrCrGrArGrUrCrGr GrUrGrCrUrUrU*mU*mU*mU
gRNA 095	1356	UAACGGAACCC CCGGACUGG	modRNA095	mU*mA*mA*rCrGrGrArArCrCrCrCr CrGrGrArCrUrGrGrGrUrUrUrArArGr ArGrCrUrArArGrCrUrGrGrArArArCr ArGrCrArUrArGrCrArArGrUrUrUrAr ArArUrArArGrGrCrUrArGrUrCrCrGr UrUrArUrCrArArCrUrUrGrArArArAr ArGrUrGrGrCrArCrCrGrArGrUrCrGr GrUrGrCrUrUrU*mU*mU*mU
gRNA 008	1269	GUGUGGGUGCU UGACGCCUG	modRNA008	mG*mU*mG*rUrGrGrGrUrGrCrUrUr GrArCrGrCrCrUrGrGrUrUrUrArArGr ArGrCrUrArArGrCrUrGrGrArArArCr ArGrCrArUrArGrCrArArGrUrUrUrAr ArArUrArArGrGrCrUrArGrUrCrCrGr UrUrArUrCrArArCrUrUrGrArArArAr ArGrUrGrGrCrArCrCrGrArGrUrCrGr GrUrGrCrUrUrU*mU*mU*mU

[0417]

gRNA No.	SEQ	靶向序列	修饰的 gRNA No.	完全修饰的 RNA 靶向序列
gRNA 012	1273	AGGAUCGUCCG AUGGGGCUC	modRNA012	mA*mG*mG*rArUrCrGrUrCrCrGrAr UrGrGrGrGrCrUrCrGrUrUrUrArArGr ArGrCrUrArArGrCrUrGrGrArArArCr ArGrCrArUrArGrCrArArGrUrUrUrAr ArArUrArArGrGrCrUrArGrUrCrCrGr UrUrArUrCrArArCrUrUrGrArArArAr ArGrUrGrGrCrArCrCrGrArGrUrCrGr GrUrGrCrUrUrU*mU*mU*mU
gRNA 126	1387	GGGACGCGUCG UUGCAGCAG	modRNA126	mG*mG*mG*rArCrGrCrGrUrCrGrUr UrGrCrArGrCrArGrGrUrUrUrArArGr ArGrCrUrArArGrCrUrGrGrArArArCr ArGrCrArUrArGrCrArArGrUrUrUrAr ArArUrArArGrGrCrUrArGrUrCrCrGr UrUrArUrCrArArCrUrUrGrArArArAr ArGrUrGrGrCrArCrCrGrArGrUrCrGr GrUrGrCrUrUrU*mU*mU*mU
gRNA 111	1372	AGUUCAGGGUC UGAGCCUGG	modRNA111	mA*mG*mU*rUrCrArGrGrGrUrCrUr GrArGrCrCrUrGrGrGrUrUrUrArArGr ArGrCrUrArArGrCrUrGrGrArArArCr ArGrCrArUrArGrCrArArGrUrUrUrAr ArArUrArArGrGrCrUrArGrUrCrCrGr UrUrArUrCrArArCrUrUrGrArArArAr ArGrUrGrGrCrArCrCrGrArGrUrCrGr GrUrGrCrUrUrU*mU*mU*mU
gRNA 212	1473	CGUGCGGCUGC GCUAUUCAG	modRNA212	mC*mG*mU*rGrCrGrGrCrUrGrCrGr CrUrArUrUrCrArGrGrUrUrUrArArGr ArGrCrUrArArGrCrUrGrGrArArArCr ArGrCrArUrArGrCrArArGrUrUrUrAr ArArUrArArGrGrCrUrArGrUrCrCrGr UrUrArUrCrArArCrUrUrGrArArArAr ArGrUrGrGrCrArCrCrGrArGrUrCrGr GrUrGrCrUrUrU*mU*mU*mU
gRNA 170	1431	UUCGGCCAGGC CGUCCUCCU	modRNA170	mU*mU*mC*rGrGrCrCrArGrGrCrCr GrUrCrCrUrCrCrUrGrUrUrUrArArGr ArGrCrUrArArGrCrUrGrGrArArArCr ArGrCrArUrArGrCrArArGrUrUrUrAr ArArUrArArGrGrCrUrArGrUrCrCrGr UrUrArUrCrArArCrUrUrGrArArArAr ArGrUrGrGrCrArCrCrGrArGrUrCrGr GrUrGrCrUrUrU*mU*mU*mU
gRNA 136	1397	GCACAGUCCUC CCCACCGCA	modRNA136	mG*mC*mA*rCrArGrUrCrCrUrCrCr CrCrArCrCrGrCrArGrUrUrUrArArGr ArGrCrUrArArGrCrUrGrGrArArArCr ArGrCrArUrArGrCrArArGrUrUrUrAr ArArUrArArGrGrCrUrArGrUrCrCrGr UrUrArUrCrArArCrUrUrGrArArArAr ArGrUrGrGrCrArCrCrGrArGrUrCrGr GrUrGrCrUrUrU*mU*mU*mU

[0418]

gRNA No.	SEQ	靶向序列	修饰的 gRNA No.	完全修饰的 RNA 靶向序列
gRNA 144	1405	GGCUCAAGGCG CCGCCGGCG	modRNA144	mG*mG*mC*rUrCrArArGrGrCrGrCr CrGrCrCrGrGrCrGrGrUrUrUrArArGr ArGrCrUrArArGrCrUrGrGrArArArCr ArGrCrArUrArGrCrArArGrUrUrUrAr ArArUrArArGrGrCrUrArGrUrCrCrGr UrUrArUrCrArArCrUrUrGrArArArAr ArGrUrGrGrCrArCrCrGrArGrUrCrGr GrUrGrCrUrUrU*mU*mU*mU
gRNA 005	1266	UCAAGCACCCA CACCCUAGA	modRNA005	mU*mC*mA*rArGrCrArCrCrCrArCr ArCrCrCrUrArGrArGrUrUrUrArArGr ArGrCrUrArArGrCrUrGrGrArArArCr ArGrCrArUrArGrCrArArGrUrUrUrAr ArArUrArArGrGrCrUrArGrUrCrCrGr UrUrArUrCrArArCrUrUrGrArArArAr ArGrUrGrGrCrArCrCrGrArGrUrCrGr GrUrGrCrUrUrU*mU*mU*mU
gRNA 102	1363	UGGCGUGAUCU GCGCGCCCC	modRNA102	mU*mG*mG*rCrGrUrGrArUrCrUrGr CrGrCrGrCrCrCrCrGrUrUrUrArArGr ArGrCrUrArArGrCrUrGrGrArArArCr ArGrCrArUrArGrCrArArGrUrUrUrAr ArArUrArArGrGrCrUrArGrUrCrCrGr UrUrArUrCrArArCrUrUrGrArArArAr ArGrUrGrGrCrArCrCrGrArGrUrCrGr GrUrGrCrUrUrU*mU*mU*mU
gRNA 013	1274	UCAGAUAGGAU CGUCCGAUG	modRNA013	mU*mC*mA*rGrArUrArGrGrArUrCr GrUrCrCrGrArUrGrGrUrUrUrArArGr ArGrCrUrArArGrCrUrGrGrArArArCr ArGrCrArUrArGrCrArArGrUrUrUrAr ArArUrArArGrGrCrUrArGrUrCrCrGr UrUrArUrCrArArCrUrUrGrArArArAr ArGrUrGrGrCrArCrCrGrArGrUrCrGr GrUrGrCrUrUrU*mU*mU*mU
gRNA 204	1465	UGCAGGCAAGG CGGCGGGGG	modRNA204	mU*mG*mC*rArGrGrCrArArGrGrCr GrGrCrGrGrGrGrGrUrUrUrArArGr ArGrCrUrArArGrCrUrGrGrArArArCr ArGrCrArUrArGrCrArArGrUrUrUrAr ArArUrArArGrGrCrUrArGrUrCrCrGr UrUrArUrCrArArCrUrUrGrArArArAr ArGrUrGrGrCrArCrCrGrArGrUrCrGr GrUrGrCrUrUrU*mU*mU*mU
gRNA 006	1267	GGGUGUGGGUG CUUGACGCC	modRNA006	mG*mG*mG*rUrGrUrGrGrGrUrGrCr UrUrGrArCrGrCrCrGrUrUrUrArArGr ArGrCrUrArArGrCrUrGrGrArArArCr ArGrCrArUrArGrCrArArGrUrUrUrAr ArArUrArArGrGrCrUrArGrUrCrCrGr UrUrArUrCrArArCrUrUrGrArArArAr ArGrUrGrGrCrArCrCrGrArGrUrCrGr GrUrGrCrUrUrU*mU*mU*mU

[0419]

gRNA No.	SEQ	靶向序列	修饰的 gRNA No.	完全修饰的 RNA 靶向序列
gRNA 210	1471	AGCACGGUGGA GAGCGGGGA	modRNA210	mA*mG*mC*rArCrGrGrUrGrGrArGr ArGrCrGrGrGrGrArGrUrUrUrArArGr ArGrCrUrArArGrCrUrGrGrArArArCr ArGrCrArUrArGrCrArArGrUrUrUrArAr ArArUrArArGrGrCrUrArGrUrCrCrGr UrUrArUrCrArArCrUrUrGrArArArArAr ArGrUrGrGrCrArCrCrGrArGrUrCrGr GrUrGrCrUrUrU*mU*mU*mU
gRNA 109	1370	GAGGCGCUCAU GGUUGCAGG	modRNA109	mG*mA*mG*rGrCrGrCrUrCrArUrGr GrUrUrGrCrArGrGrGrUrUrUrArArGr ArGrCrUrArArGrCrUrGrGrArArArCr ArGrCrArUrArGrCrArArGrUrUrUrArAr ArArUrArArGrGrCrUrArGrUrCrCrGr UrUrArUrCrArArCrUrUrGrArArArArAr ArGrUrGrGrCrArCrCrGrArGrUrCrGr GrUrGrCrUrUrU*mU*mU*mU
gRNA 060	1321	GCAGGAGACAA GGUGGACCC	modRNA060	mG*mC*mA*rGrGrArGrArCrArArGr GrUrGrGrArCrCrCrGrUrUrUrArArGr ArGrCrUrArArGrCrUrGrGrArArArCr ArGrCrArUrArGrCrArArGrUrUrUrArAr ArArUrArArGrGrCrUrArGrUrCrCrGr UrUrArUrCrArArCrUrUrGrArArArArAr ArGrUrGrGrCrArCrCrGrArGrUrCrGr GrUrGrCrUrUrU*mU*mU*mU
gRNA 015	1276	CGGAAUCCUGG CUGGGAGCU	modRNA015	mC*mG*mG*rArArUrCrCrUrGrGrCr UrGrGrGrArGrCrUrGrUrUrUrArArGr ArGrCrUrArArGrCrUrGrGrArArArCr ArGrCrArUrArGrCrArArGrUrUrUrArAr ArArUrArArGrGrCrUrArGrUrCrCrGr UrUrArUrCrArArCrUrUrGrArArArArAr ArGrUrGrGrCrArCrCrGrArGrUrCrGr GrUrGrCrUrUrU*mU*mU*mU
gRNA 001	1262	GGUGCUAGCCU UGCGUUCCG	modRNA001	mG*mG*mU*rGrCrUrArGrCrCrUrUr GrCrGrUrUrCrCrGrGrUrUrUrArArGr ArGrCrUrArArGrCrUrGrGrArArArCr ArGrCrArUrArGrCrArArGrUrUrUrArAr ArArUrArArGrGrCrUrArGrUrCrCrGr UrUrArUrCrArArCrUrUrGrArArArArAr ArGrUrGrGrCrArCrCrGrArGrUrCrGr GrUrGrCrUrUrU*mU*mU*mU
gRNA 197	1458	UCUUCGCUUGG CACGAUCUU	modRNA197	mU*mC*mU*rUrCrGrCrUrUrGrGrCr ArCrGrArUrCrUrUrGrUrUrUrArArGr ArGrCrUrArArGrCrUrGrGrArArArCr ArGrCrArUrArGrCrArArGrUrUrUrArAr ArArUrArArGrGrCrUrArGrUrCrCrGr UrUrArUrCrArArCrUrUrGrArArArArAr ArGrUrGrGrCrArCrCrGrArGrUrCrGr GrUrGrCrUrUrU*mU*mU*mU

[0420]

gRNA No.	SEQ	靶向序列	修饰的 gRNA No.	完全修饰的 RNA 靶向序列
gRNA 127	1388	GCUGCUGCAAC GACGCGUCC	modRNA127	mG*mC*mU*rGrCrUrGrCrArArCrGr ArCrGrCrGrUrCrCrGrUrUrUrArArGr ArGrCrUrArArGrCrUrGrGrArArArCr ArGrCrArUrArGrCrArArGrUrUrUrAr ArArUrArArGrGrCrUrArGrUrCrCrGr UrUrArUrCrArArCrUrUrGrArArArAr ArGrUrGrGrCrArCrCrGrArGrUrCrGr GrUrGrCrUrUrU*mU*mU*mU
gRNA 180	1441	CCCGCACCUUG GCGCAGCGG	modRNA180	mC*mC*mC*rGrCrArCrCrUrUrGrGr CrGrCrArGrCrGrGrGrUrUrUrArArGr ArGrCrUrArArGrCrUrGrGrArArArCr ArGrCrArUrArGrCrArArGrUrUrUrAr ArArUrArArGrGrCrUrArGrUrCrCrGr UrUrArUrCrArArCrUrUrGrArArArAr ArGrUrGrGrCrArCrCrGrArGrUrCrGr GrUrGrCrUrUrU*mU*mU*mU
gRNA 112	1373	CGGGCGCCGCC GUUCAGUUC	modRNA112	mC*mG*mG*rGrCrGrCrCrGrCrCrGr UrUrCrArGrUrUrCrGrUrUrUrArArGr ArGrCrUrArArGrCrUrGrGrArArArCr ArGrCrArUrArGrCrArArGrUrUrUrAr ArArUrArArGrGrCrUrArGrUrCrCrGr UrUrArUrCrArArCrUrUrGrArArArAr ArGrUrGrGrCrArCrCrGrArGrUrCrGr GrUrGrCrUrUrU*mU*mU*mU

[0421] 使用Mirus的TransIT®-X2转染试剂(目录号MIR6003)以96孔板形式用25ng效应子和12.5ng sgRNA反向转染HeLa细胞。每周收获条件培养基长达四周,并用于使用BioLegend的LEGEND MAX™人PCSK9 ELISA试剂盒(目录号443107)测量分泌的PCSK9水平。使用Promega的CellTiter-Glo®试剂盒(目录号G7571)将ELISA数据标准化为细胞数量。虽然PCSK9沉默在CRISPRi(dCas9-KRAB)中是短暂的,并且在第28天恢复到基线,但与CRISPR-off(DNMT3A-3L-dCas9-KRAB)构建体共转染的几种sgRNA显示出稳健和持久的沉默(图4A)。用CRISPR-off测试的40个指导物中的16个显示出大于WT Cas9的沉默效率(表9)。

[0422] 表9.用修饰的gRNA和CRISPR-off处理的HeLa细胞中的相对PCSK9表达

[0423]

条件	D28, 对照% (平均)
阴性对照	
仅转染试剂	100
CRISPRi IVT + modRNA009	108.1
CRISPRi IVT + modRNA004	88.1
阳性对照	
WT Cas9 IVT + modRNA180	10.5
实验条件 (CRISPR-off 和所示修饰的 gRNA)	
modRNA009	5.7
modRNA108	12.9
modRNA003	7.5
modRNA128	16.2
modRNA110	5.7
modRNA093	11.7
modRNA096	20.7
modRNA092	12.7
modRNA011	9.6
modRNA007	11.6
modRNA077	4.2
modRNA113	4.7
modRNA094	12.0
modRNA125	67.4
modRNA122	9.5
modRNA004	1.0
modRNA107	7.7
modRNA095	30.5
modRNA008	4.7
modRNA012	6.0
modRNA126	6.1
modRNA111	4.0
modRNA212	25.7
modRNA170	14.8
modRNA136	17.2
modRNA144	35.0
modRNA005	7.9
modRNA102	25.7

[0424]

modRNA013	5.2
modRNA204	64.4
modRNA006	36.6
modRNA210	30.0
modRNA109	25.3
modRNA060	33.3
modRNA015	19.4
modRNA001	18.2
modRNA197	45.1
modRNA127	13.4
modRNA180	30.2
modRNA112	6.1

[0425] 在两周时间点,使用Zymo Research的Quick-RNA 96试剂盒(目录号R1053)提取RNA。使用Quantabio的qScript XLT一步RT-qPCR ToughMix(目录号95134-500)和TaqMan测定进行qPCR(PCSK9:Hs00545399_m1,PPIA:Hs99999904_m1)。PCSK9水平用PPIA标准化。使用δ-δCt方法进行相对定量。

[0426] 发现PCSK9分泌的抑制与第14天的mRNA沉默相关(图4B)。

[0427] 在HeLa细胞中测试modRNA004和modRNA111在60天内抑制PCSK9分泌(图5)。将WT Cas9与modRNA180共转染作为阳性对照。用25ng效应物和12.5ng gRNA处理细胞。modRNA004和modRNA111显示能介导PCSK9的持久沉默,与在HeLa细胞中通过WT Cas9经由基因编辑实现的沉默相当。

[0428] 此外,结果表明在用辛伐他汀处理的HeLa细胞中维持表观遗传沉默。已知他汀类药物治疗可经由转录机制增加PCSK9分泌。在表观遗传沉默的HeLa细胞中,他汀类药物治疗显示不增加PCSK9分泌(图6)。

[0429] 实施例5:Huh7肝癌细胞系中的指导RNA测定

[0430] Huh7肝癌细胞系适于高通量筛选和转染。在Huh7肝癌细胞系中测试来自HeLa筛选的与食蟹猴PCSK9基因具有完全同源性或单个错配的前13个指导物。用CRISPR-off构建体的表观遗传沉默显示在7天内是稳定的(图7)。

[0431] 实施例6:原代人和食蟹猴肝细胞中的指导RNA测定

[0432] BioIVT的原代人和食蟹猴HepatoPac®培养物用于测试gRNA在原代肝细胞中的功效。根据制造商的建议维持HepatoPac®培养物。简而言之,将HepatoPac®维持培养基解冻并在细胞到达后30分钟内补充。培养基更换后,使细胞在37℃、10%CO_2培养箱中适应两天。在接收后的第二天,按照制造商的说明,使用SPARK™(Precision Nanosystems)和hepato9mRNA LNP制剂试剂盒(目录号NWS0016)配制不同浓度的CRISPR-off+sgRNA、GFP-mRNA和WT CRISPR Cas9的LNP。随后,通过Quant-it™RiboGreen RNA测定试剂盒表征LNP的包封效率和总mRNA有效载荷递送。然后将LNP以指定量的总mRNA添加到培养基中以实现PCSK9沉默的临床相关水平。培养基每隔一天更换一次,持续长达四周,以评估沉默的持久性和/或遗传性。通过ELISA每七天评估PCSK9沉默以测量培养基中分泌的PCSK9水平。将使用人白蛋白ELISA(Thermo Fisher®)将PCSK9浓度控制至总肝细胞。然后将数据表示为总

PCSK9分泌占GFP-mRNA阴性对照的百分比。可以通过使用基于磁珠的抗体方法(Miltenyi Biotec)从小鼠成纤维细胞饲养层分离原代人和食蟹猴肝细胞来评估特异性。在从饲养层分离原代肝细胞后,处理细胞用于RNAseq评估和全基因组亚硫酸氢盐测序。

[0433] 选择RNA形式的前13个gRNA在原代人肝细胞(PHH)中测试。基于以下来选择靠前的gRNA:(i)在HeLa细胞中的PCSK9沉默效率和持久性,(ii)它们是否与人PCSK9基因完美比对并且与非人灵长类PCSK9基因至多一个错配。还测试gRNA的组合以确定其功效和持久性。阴性对照为只有CRISPR-off、只有gRNA片段(modRNA003)和只有CRISPRi。阳性对照是与modRNA004共转染的CRISPRi(表10;SEQ:SEQ ID NO;NHP:非人灵长类)。预测所有测试的gRNA结合人和非人灵长类PCSK9两者。

[0434] 表10.在原代人肝细胞中筛选gRNA

[0435]

gRNA No.	SEQ	gRNA 靶向序列	TSS 距离	修饰的 RNA No.	匹配至 NHP
gRNA009	1270	ACUGCCUGGCUCACUCCUCC	81	modRNA009	精确
gRNA003	1264	UGCGGAAACCUUCUAGGGUG	0	modRNA003	精确
gRNA093	1354	AACCUGAUCCUCCAGUCCGG	-104	modRNA093	精确
gRNA011	1272	AUCGUCCGAUGGGGCUCUGG	-56	modRNA011	精确
gRNA007	1268	GGUGUGGGUGCUUGACGCCU	-16	modRNA007	精确
gRNA077	1338	CAAAUCCUAACUGGGCUGGA	-181	modRNA077	1 错配
gRNA113	1374	GGGCGCCGCCGUUCAGUUCA	49	modRNA113	1 错配
gRNA004	1265	GCGGAAACCUUCUAGGGUGU	-1	modRNA004	精确
gRNA008	1269	GUGUGGGUGCUUGACGCCUG	-17	modRNA008	精确
gRNA012	1273	AGGAUCGUCCGAUGGGGCUC	-59	modRNA012	精确
gRNA111	1372	AGUUCAGGGUCUGAGCCUGG	43	modRNA111	1 错配
gRNA005	1266	UCAAGCACCCACACCCUAGA	-11	modRNA005	精确
gRNA013	1274	UCAGAUAGGAUCGUCCGAUG	-65	modRNA013	精确

[0436] 观察到稳健的PCSK9沉默。对于一些gRNA,在第7天观察到分泌的PCSK9减少>70%(取决于转染效率)。

[0437] 实施例7:HeLa细胞中的ZF测定

[0438] 从ZF文库设计针对49个PCSK9靶位点(选自GRCh38染色体1,在55038548至55040548之间)的共209个锌指蛋白(结构如SEQ ID NO:659所示)。靶位点在人基因组(GRCh38)中没有其他精确匹配。

[0439] 用DNA形式的ZF-off构建体转染HeLa细胞。在含有补充有10%胎牛血清(Thermo Fisher目录号A4766)v/v、1x GlutaMAX™(Thermo Fisher目录号35050061)和1x青霉素-链霉素(Thermo Fisher目录号15140122)的DMEM(Thermo Fisher目录号11-965-092)的标准培养基中以12,000个HeLa细胞每孔(ATCC目录号CCL-2)接种6个96孔板(Sigma-Aldrich目录号M2936)。铺板后,使细胞在37℃培养箱中于5%CO_2下生长24小时。将10ng ZF-OFF质粒重悬于DPBS缓冲液(Thermo Fisher目录号14190144)中至7.5ng/μL的浓度。此外,还将10ng的EF1a:嘌呤霉素抗性质粒和65ng的空载体(pUC19)添加到转染混合物中以实现85ng DNA的总有效载荷。通过在无血清OPTI-MEM培养基(Thermo Fisher®目录号31985062)中添加重悬的DNA并按照制造商的说明添加Mirus® TransIT®-LT1转染试剂(MIR2300)来构建转染混合物。将10μL转染混合物一式两份添加到总共六个筛选板中。阳性对照是具有高性能

gRNA(gRNA009)的CRISPR-off(SEQ ID NO:658)。阴性对照是ZF-off,具有非PCSK9基因座靶标(CLTA)和空载体(pUC19;NEB目录号N3041S)。

[0440] ZF筛选产生了活性与CRISPR相当的命中(图8)。具有高沉默效率的候选物被带到后续实验中。图9显示了以到TSS的距离划分的ZF筛选结果。总共筛选了209个ZF,其相对于阴性对照的PCSK9敲低活性如下表11所示。

[0441] 表11.ZF-off构建体活性

[0442]

ZF No.	对照%	ZF No.	对照%	ZF No.	对照%
ZF028	37.43	ZF071	66.51	ZF141	104.50
ZF001	35.90	ZF072	62.59	ZF142	71.70
ZF047	32.08	ZF073	69.50	ZF143	49.70
ZF048	15.19	ZF074	55.74	ZF144	44.01
ZF029	40.42	ZF075	51.58	ZF145	60.80
ZF030	7.74	ZF076	58.82	ZF146	42.85
ZF031	29.80	ZF077	49.52	ZF147	58.42
ZF032	40.65	ZF078	52.50	ZF148	70.50
ZF033	30.71	ZF079	66.30	ZF149	95.42
ZF034	31.95	ZF080	51.52	ZF150	74.00
ZF035	38.29	ZF081	140.85	ZF151	61.01
ZF002	33.02	ZF082	105.84	ZF152	52.34
ZF003	29.17	ZF083	72.30	ZF153	80.11
ZF004	35.39	ZF084	95.05	ZF154	79.46
ZF005	37.90	ZF085	44.18	ZF155	50.09
ZF006	37.47	ZF086	88.95	ZF156	46.99
ZF036	35.19	ZF087	44.87	ZF157	89.03
ZF037	28.76	ZF088	56.44	ZF158	99.89
ZF038	21.11	ZF089	116.41	ZF159	101.66
ZF039	9.35	ZF090	51.30	ZF160	119.21
ZF040	5.20	ZF091	57.46	ZF161	49.64
ZF041	30.50	ZF092	46.82	ZF162	68.98
ZF007	36.13	ZF093	44.03	ZF163	89.71
ZF008	34.72	ZF094	65.55	ZF164	103.15
ZF009	28.91	ZF095	56.92	ZF165	48.37
ZF010	33.58	ZF096	93.13	ZF166	81.13
ZF011	41.01	ZF097	57.08	ZF167	67.29
ZF012	19.39	ZF098	57.03	ZF168	50.88
ZF013	40.46	ZF099	59.36	ZF169	77.25
ZF014	33.06	ZF100	95.34	ZF170	108.08
ZF015	22.48	ZF101	57.13	ZF171	57.10
ZF016	34.09	ZF102	145.81	ZF172	53.73

[0443]

ZF No.	对照%	ZF No.	对照%	ZF No.	对照%
ZF017	24.71	ZF103	63.08	ZF173	62.61
ZF018	24.84	ZF104	50.06	ZF174	71.84
ZF019	14.70	ZF105	72.80	ZF175	68.12
ZF020	33.15	ZF106	75.56	ZF176	51.52
ZF021	33.87	ZF107	56.29	ZF177	48.59
ZF022	29.02	ZF108	103.79	ZF178	71.10
ZF023	30.45	ZF109	60.41	ZF179	53.08
ZF024	12.55	ZF110	53.71	ZF180	54.56
ZF042	41.13	ZF111	69.30	ZF181	78.86
ZF025	39.37	ZF112	94.36	ZF182	70.20
ZF026	28.27	ZF113	57.16	ZF183	69.07
ZF043	40.92	ZF114	51.72	ZF184	49.73
ZF044	37.27	ZF115	78.55	ZF185	69.04
ZF045	39.72	ZF116	76.55	ZF186	57.09
ZF027	38.67	ZF117	66.49	ZF187	64.90
ZF046	36.12	ZF118	60.58	ZF188	59.91
ZF049	43.53	ZF119	78.70	ZF189	59.92
ZF050	44.02	ZF120	63.82	ZF190	79.85
ZF051	46.90	ZF121	117.94	ZF191	71.95
ZF052	45.89	ZF122	82.54	ZF192	78.47
ZF053	42.71	ZF123	84.94	ZF193	74.79
ZF054	44.31	ZF124	71.86	ZF194	67.87
ZF055	44.57	ZF125	76.73	ZF195	83.86
ZF056	44.82	ZF126	119.64	ZF196	67.71
ZF057	46.52	ZF127	144.15	ZF197	64.36
ZF058	46.61	ZF128	66.74	ZF198	62.06
ZF059	47.62	ZF129	86.03	ZF199	59.08
ZF060	44.75	ZF130	103.82	ZF200	53.09
ZF061	59.99	ZF131	116.46	ZF201	57.39
ZF062	48.69	ZF132	130.63	ZF202	63.53
ZF063	50.63	ZF133	100.63	ZF203	69.94
ZF064	54.60	ZF134	42.35	ZF204	67.61
ZF065	54.25	ZF135	61.14	ZF205	58.51
ZF066	54.95	ZF136	47.14	ZF206	77.67
ZF067	69.99	ZF137	101.88	ZF207	72.94
ZF068	54.64	ZF138	99.22	ZF208	55.34
ZF069	55.99	ZF139	69.52	ZF209	64.42
ZF070	61.26	ZF140	48.41		

[0444] 与WTCas9以及CRISPR-off和gRNA003组合相比,几种ZF-off构建体在沉默PCSK9方面显示更有效。这些ZF-off构建体中ZFP结构域的靶位点和ZF序列(F1至F6)如表1所示。

[0445] 实施例8:构建体在原代人肝细胞中的全特异性筛选

[0446] 在原代人肝细胞中测试CRISPR-off和ZF-off构建体沉默PCSK9的特异性。评估特异性的读出是RNAseq、甲基化阵列和全基因组亚硫酸氢盐测序测定。表观遗传编辑后与阴性对照相比的全基因组表达和甲基化变化将被分析。

[0447] 实施例9:CpG甲基化模式

[0448] 研究了用CRISPR-off或ZF-off处理的人肝细胞(例如原代细胞或细胞系)中的CpG

甲基化模式。对亚硫酸氢盐处理的DNA进行杂交捕获测定,以研究在PCSK9 TSS周围1kb区域由CRISPR-Off或ZF-Off诱导的CpG位点的甲基化模式。

[0449] 实施例10:通过表观遗传编辑在具有野生型PCSK9的小鼠中稳定的PCSK9沉默

[0450] 测试CRISPR-off和ZF-off构建体在体内介导内源性PCSK9的表观遗传沉默的能力。使用单次IV施用配制到LNP中的mRNA(以及对于CRISPR-off沉默,gRNA)来递送构建体。在野生型小鼠中对沉默进行了为期两到六个月的测试。读出是血清PCSK9水平和血清胆固醇水平。选择每个队列的子集用于肝脏苏木精和伊红(H&E)染色RNAseq和分析。对于几种构建体,观察到稳健、稳定和可遗传的PCSK9沉默。

[0451] 实施例11:通过表观遗传编辑在表达转基因人PCSK9的小鼠中稳定的PCSK9沉默

[0452] 使用表达转基因人PCSK9的三种不同小鼠品系:hPCSK9-Tg(mPCSK9+/-)杂合小鼠、hPCSK9-Tg(mPCSK9+/+)纯合小鼠和hPCSK9-Tg(mPCSK9-/-)小鼠。使用的hPCSK9-Tg(mPCSK9-/-)小鼠系是C57BL/6J-Pcsk9-/-Tg(RP11-55M23-AbsI),其在自身启动子的控制下表达人PCSK9(图10)。参见例如Weider et al.,J Biol Chem(2016)291(32):16659-71。

[0453] 测试CRISPR-off和ZF-off构建体。经由单次IV施用配制到LNP中的mRNA/gRNA递送构建体。读出是肝脏H&E染色、测量PCSK9 mRNA水平的RNAseq和AST/ASL测量。还测试了功效,包括通过血清PCSK9蛋白水平测量的PCSK9沉默在三至四个月内的持久性。对于一些构建体,观察到血清PCSK9水平的持久和显著降低。

[0454] 持久性测试超过六至十二个月。读出是血清PCSK9水平和血清胆固醇水平。选择队列的子集用于肝脏H&E和RNAseq分析。

[0455] 实施例12:具有变体NLS构型的融合蛋白

[0456] 使用变体核定位序列(NLS)构型开发了几种改进的融合蛋白构建体,以具有显著更高的表观沉默活性。

[0457] 构建具有NLS结构域变体构型的几种构建体(图11A和11B),并在HeLa细胞的PCSK9基因座中测试(图12A-12B)。另外在Hepa1-6(图13)和HuH7(图14A-14C和图15)中的PCSK9基因座中测试构建体。示例性融合蛋白构建体氨基酸和DNA序列如下所示:

[0458]

融合蛋白1 氨基酸序列 NLS-3A-3L-dCas9-KRAB-NLS (SEQ ID NO: 1498)	MPKKKRKVPKKKRKVNHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVD RYIASEVCEDSITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPAR KGLYEGTGRLFFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDA KEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGK DQHFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLK EYFACVSSGNSNANSRGPSFSSGLVPLSLRGSHNPLEMFETVPVWRRQPVRVLSLFEDIKKE LTSLGFLESGSDPGQLKHVVDVTDTVRKDVEEWGPFDLVYGATPPLGHTCDRPPSWYLFQFH RLLQYARPKPGSPRPFFWMFVDNLVLNKEDLDVASRFLEMEPVTIPDVHGGSLQNAVRVWSN IPAIRSRHWALVSEEELSLLAQNKQSSKLAAKWPTKLVKNCFLPLREYFKYFSTELTSSLGG PSSGAPPPSGGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTST EPSEGSAPGTSTEPSEMDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKN LIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVE EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLI EGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGE KKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAK NLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSK NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGE LHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSG EQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKL INGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSI DNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKA GFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFF YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADAN LDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLI HQSITGLYETRIDLSQLGGDSGSETPGTSESATPESTGRTLVTFKDVFVDFTREEWKLLDTA QQIVYRNVMLENYKNLVSLGYQLTKPDVILRLEKGEEPPKKKRKVPKKKRKV

[0459]

融合蛋白 1 DNA 序列 (SEQ ID NO: 1499)	ATGCCAAAAAAGAAGAGAAAGGTACCGAAGAAAAAAAGAAAGGTCAACCATGATCAAGAATT CGACCCACCTAAAGTCTACCCACCTGTGCCCGCCGAAAAAAGGAAACCCATAAGGGTGCTGT CACTCTTTGATGGCATCGCCACTGGTCTCCTGGTTCTTAAGGATCTGGGAATTCAGGTCGAT CGGTACATTGCTAGCGAGGTTTGTGAGGATAGTATTACAGTGGGTATGGTGCGCCACCAGGG AAAGATCATGTATGTTGGTGACGTTAGGAGCGTCACCCAGAAACATATCCAGGAGTGGGGAC CCTTTGATTTGGTGATCGGAGGTAGTCCCTGCAATGACCTTTCCATCGTGAATCCAGCCAGG AAAGGGCTGTATGAAGGGACTGGTAGGCTCTTTTTCGAGTTTTATCGCCTGCTTCACGACGC TAGACCTAAGGAAGGTGACGATAGGCCTTTCTTTTGGCTTTTTGAGAACGTCGTGGCAATGG GAGTCTCCGACAAAAGGGACATTTCTCGCTTTCTGGAATCTAACCCCGTTATGATCGATGCC AAGGAAGTTTCTGCCGCTCACAGGGCAAGGTACTTCTGGGGCAATCTGCCCGGAATGAATCG CCCACTGGCCAGTACCGTGAATGACAAACTGGAGCTGCAGGAGTGCCTGGAGCACGGAAGAA TCGCAAAGTTTTCTAAAGTCAGGACCATTACCACTCGCAGTAACTCCATAAAACAGGGTAAG GACCAGCATTTTCCCGTCTTCATGAATGAAAAGGAAGATATTCTGTGGTGCACTGAAATGGA GAGAGTTTTCGGGTTTCCCGTGCACTATACCGATGTTTCCAACATGTCCCGCCTTGCAAGAC AAAGGCTTTTGGGCCGCTCTTGGTCTGTGCCAGTGATCCGGCACTTGTTTGCTCCCCTCAAA GAGTACTTCGCTTGCGTCAGTTCCGGAAATTCAAACGCTAACTCTCGGGGTCCATCTTTCTC CAGTGGTCTCGTGCCACTGTCTCTCCGGGGCTCTCACAATCCCCTGGAGATGTTTGAGACAG TGCCAGTCTGGCGGAGGCAGCCCGTTCGCGTTCTCTCTCTGTTCGAAGATATTAAAAAGGAA CTCACCTCCCTTGGGTTCCTGGAGAGCGGGAGCGACCCCGGACAGCTTAAGCACGTGGTCGA CGTGACTGACACCGTCCGCAAAGACGTGGAGGAATGGGGCCCCTTCGATCTGGTCTATGGGG CAACCCCTCCCCTTGGGCATACATGTGATCGGCCTCCATCCTGGTACCTGTTCCAGTTTCAC AGACTCCTGCAGTATGCCAGGCCAAAGCCAGGGAGCCCAAGGCCCTTTTTCTGGATGTTCGT CGACAACCTGGTCCTGAACAAAGAAGATCTCGACGTTGCTAGTCGCTTTCTCGAAATGGAGC CCGTGACCATTCCCGACGTGCATGGCGGTTCCCTCCAGAATGCAGTCAGGGTTTGGAGCAAT ATCCCTGCCATCAGGTCAAGGCACTGGGCACTGGTTTCAGAGGAAGAGCTGTCCCTCCTTGC CCAGAACAAGCAGTCATCCAAACTGGCAGCCAAGTGGCCAACTAAGCTGGTCAAGAACTGCT TTCTTCCCCTCAGAGAATATTTTAAGTATTTCAGTACTGAACTGACTAGCAGTCTGGGAGGG CCGAGCTCTGGCGCACCCCCACCAAGTGGAGGGTCTCCTGCCGGGTCCCCAACATCTACTGA AGAAGGCACCAGCGAATCCGCAACGCCCGAGTCAGGCCCTGGTACCTCCACAGAACCATCTG AAGGTAGTGCGCCTGGTTCCCCAGCTGGAAGCCCTACTTCCACCGAAGAAGGCACGTCAACC GAACCAAGTGAAGGATCTGCCCCTGGGACCAGCACTGAACCATCTGAGATGGACAAGAAGTA CAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACA AGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAAC CTGATCGGCGCCCTGCTGTTCGACAGCGGAGAAACAGCCGAGGCCACCCGGCTGAAGAGAAC CGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCA ACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAA GAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCA CGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCG ACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATC GAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGAC CTACAACCAGCTGTTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCC TGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAG AAGAAGAATGGCCTGTTCGGCAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAA GAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACG ACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAG AACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGC CCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGA AAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAG AACGGCTACGCCGGCTACATCGATGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAA GCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACC TGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAG CTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAA GATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACA GCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAA GTGGTGGACAAGGGCGCCAGCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAA CCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTACA ACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGC GAGCAGAAAAAAGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCA GCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGG AAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAG GACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACT GTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACA

[0460]

	AAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTG ATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGG CTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACA TCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCC GGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAA AGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCA CCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAG CTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCT GTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACC GGCTGTCCGACTACGATGTGGACGCTATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATC GATAACAAAGTGCTGACTCGGAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGA AGAGGTCGTGAAGAAGATGAAGAACTACTGGCGCCAGCTGCTGAATGCCAAGCTGATTACCC AGAGGAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCC GGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCT GGACTCCCGGATGAACACTAAGTACGACGAGAACGACAAACTGATCCGGGAAGTGAAAGTGA TCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGC GAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCT GATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACG TGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTC TACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAA GCGGCCTCTGATCGAGACAAACGGCGAAACAGGCGAGATCGTGTGGGATAAGGGCCGGGACT TTGCCACCGTGCGGAAAGTGCTGTCTATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTG CAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGACAAGCTGATCGC CAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATT CTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAG CTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGA AGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGT TCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAAC GAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCT GAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAACACTACC TGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAAT CTGGACAAGGTGCTGAGCGCCTACAACAAGCACAGAGACAAGCCTATCAGAGAGCAGGCCGA GAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTG ACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATC CACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGCGACAG CGGAAGTGAGACCCCAGGTACATCCGAATCAGCAACGCCTGAAAGCACCGGTCGGACACTGG TGACCTTCAAGGATGTATTTGTGGACTTCACCAGGGAGGAGTGGAAGCTGCTGGACACTGCT CAGCAGATCGTGTACAGAAATGTGATGCTGGAGAACTATAAGAACCTGGTTTCCTTGGGTTA TCAGCTTACTAAGCCAGATGTGATCCTCCGGTTGGAGAAGGGAGAAGAGCCCCCAAAAAAGA AGAGAAAGGTACCGAAGAAAAAAAGAAAGGTC
融合蛋白 2 氨基酸序列 3A-3L-NLS-dCas9-NLS-KRAB (SEQ ID NO: 1500)	MNHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCEDSITV GMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRLFFEF YRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSAAHRARYFWG NLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEKEDI LWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLKEYFACVSSGNSNAN SRGPSFSSGLVPLSLRGSHNPLEMFETVPVWRRQPVRVLSLFEDIKKELTSLGFLESGSDPG QLKHVVDVTDTVRKDVEEWGPFDLVYGATPPLGHTCDRPPSWYLFQFHRLLQYARPKPGSPR PFFWMFVDNLVLNKEDLDVASRFLEMEPVTIPDVHGGSLQNAVRVWSNIPAIRSRHWALVSE EELSLLAQNKQSSKLAAKWPTKLVKNCFLPLREYFKYFSTELTSSLGGPSSGAPPPSGGSPA GSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEP SEPKKKRKVMDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHER HPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD NSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFG NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAIL LSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYI DGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRR QEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIV DLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEE NEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKK

	GILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILK EHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTR SDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHH AHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNF FKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSK ESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIM ERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSK YVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSA YNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGL YETRIDLSQLGGDPKKKRKVSGSETPGTSESATPESTGRTLVTFKDVFVDFTREEWKLLDTA QQIVYRNVMLENYKNLVSLGYQLTKPDVILRLEKGEEP
融合蛋白 2 DNA 序列 (SEQ ID NO: 1501)	ATGAACCACGACCAGGAATTTGACCCTCCAAAGGTTTACCCACCTGTCCCAGCTGAGAAGAG GAAGCCCATCCGGGTGCTGTCTCTCTTTGATGGAATCGCTACAGGGCTCCTGGTGCTGAAGG ACTTGGGCATTCAGGTGGACCGCTACATTGCCTCGGAGGTGTGTGAGGACTCCATCACGGTG GGCATGGTGCGGCACCAGGGGAAGATCATGTACGTCGGGGACGTCCGCAGCGTCACACAGAA GCATATCCAGGAGTGGGGCCCATTCGATCTGGTGATTGGGGGCAGTCCCTGCAATGACCTCT CCATCGTCAACCCTGCTCGCAAGGGCCTCTACGAGGGCACTGGCCGGCTCTTCTTTGAGTTC TACCGCCTCCTGCATGATGCGCGGCCCAAGGAGGGAGATGATCGCCCCTTCTTCTGGCTCTT TGAGAATGTGGTGGCCATGGGCGTTAGTGACAAGAGGGACATCTCGCGATTTCTCGAGTCCA ACCCTGTGATGATTGATGCCAAAGAAGTGTCAGCTGCACACAGGGCCCGCTACTTCTGGGGT AACCTTCCCGGTATGAACAGGCCGTTGGCATCCACTGTGAATGATAAGCTGGAGCTGCAGGA GTGTCTGGAGCATGGCAGGATAGCCAAGTTCAGCAAAGTGAGGACCATTACTACGAGGTCAA ACTCCATAAAGCAGGGCAAAGACCAGCATTTTCCTGTCTTCATGAATGAGAAAGAGGACATC TTATGGTGCACTGAAATGGAAAGGGTATTTGGTTTCCCAGTCCACTATACTGACGTCTCCAA CATGAGCCGCTTGGCGAGGCAGAGACTGCTGGGCCGGTCATGGAGCGTGCCAGTCATCCGCC ACCTCTTCGCTCCGCTGAAGGAGTATTTTGCGTGTGTGTCTAGCGGCAATAGTAACGCTAAC AGCCGCGGGCCGAGCTTCAGCAGCGGCCTGGTGCCGTTAAGCTTGCGCGGCAGCCATAATCC CCTTGAGATGTTCGAAACCGTGCCTGTGTGGAGGAGACAGCCAGTCCGGGTGCTGTCCCTTT TTGAAGACATCAAGAAAGAGCTGACGAGTTTGGGCTTTTTGGAAAGTGGTTCTGACCCGGGA CAACTGAAGCATGTGGTTGATGTCACAGACACAGTGAGGAAGGATGTGGAGGAGTGGGGACC CTTCGATCTTGTGTACGGCGCCACACCTCCCCTGGGCCACACCTGTGACCGTCCTCCCAGCT GGTACCTGTTCCAGTTCCACCGGCTCCTGCAGTACGCACGGCCCAAGCCAGGCAGCCCCAGG CCCTTCTTCTGGATGTTCGTGGACAATCTGGTGCTGAACAAGGAAGACCTGGACGTCGCATC TCGCTTCCTGGAGATGGAGCCAGTCACCATCCCAGATGTCCACGGCGGATCCTTGCAGAATG CTGTCCGCGTGTGGAGCAACATCCCAGCCATAAGGAGCAGGCACTGGGCTCTGGTTTCGGAA GAAGAATTGTCCCTGCTGGCCCAGAACAAGCAGAGCTCGAAGCTCGCGGCCAAGTGGCCCAC CAAGCTGGTGAAGAACTGCTTTCTCCCCCTAAGAGAATATTTCAAGTATTTTTCAACAGAAC TCACTTCCTCTTTAGGAGGGCCGAGCTCTGGCGCACCCCCACCAAGTGGAGGGTCTCCTGCC GGGTCCCCAACATCTACTGAAGAAGGCACCAGCGAATCCGCAACGCCCGAGTCAGGCCCTGG TACCTCCACAGAACCATCTGAAGGTAGTGCGCCTGGTTCCCCAGCTGGAAGCCCTACTTCCA CCGAAGAAGGCACGTCAACCGAACCAAGTGAAGGATCTGCCCCTGGGACCAGCACTGAACCA TCTGAGCCAAAAAAGAAGAGAAAGGTAATGGACAAGAAGTACAGCATCGGCCTGGCCATCGG CACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCA AGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGCGCCCTGCTGTTC GACAGCGGAGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAG ACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACG ACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGG CACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCACCATCTA CCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGG CCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGAC AACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGA AAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGA GCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGC AACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGA GGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCC AGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTG CTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGAT CAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGC TGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATC GATGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGA CGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCT

[0461]

[0462]

	TCGACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGG CAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTT CCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCA GAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCCAGC GCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCT GCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTACAACGAGCTGACCAAAGTGAAAT ACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAAGCCATCGTG GACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAA GAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCC TGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAA AACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGAT CGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGC GGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAG CAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCAT GCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCG GCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAG GGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCC CGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACA GCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAA GAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAA TGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGG ACGCTATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGATAACAAAGTGCTGACTCGG AGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAA GAACTACTGGCGCCAGCTGCTGAATGCCAAGCTGATTACCCAGAGGAAGTTCGACAATCTGA CCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTG GTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAA GTACGACGAGAACGACAAACTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGG TGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCAC GCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCT GGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGA GCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTT TTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAA CGGCGAAACAGGCGAGATCGTGTGGGATAAGGGCCGGGACTTTGCCACCGTGCGGAAAGTGC TGTCTATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAA GAGTCTATCCTGCCCAAGAGGAACAGCGACAAGCTGATCGCCAGAAAGAAGGACTGGGACCC TAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAG TGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATG GAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGT GAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGA AGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAA TATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAA TGAGCAGAAACAGCTGTTTGTGGAACAGCACAAACACTACCTGGACGAGATCATCGAGCAGA TCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAGGTGCTGAGCGCC TACAACAAGCACAGAGACAAGCCTATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTAC CCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGA GGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTG TACGAGACACGGATCGACCTGTCTCAGCTGGGAGGCGACCCAAAAAAGAAGAGAAAGGTAAG CGGAAGTGAGACCCCAGGTACATCCGAATCAGCAACGCCTGAAAGCACCGGTCGGACACTGG TGACCTTCAAGGATGTATTTGTGGACTTCACCAGGGAGGAGTGGAAGCTGCTGGACACTGCT CAGCAGATCGTGTACAGAAATGTGATGCTGGAGAACTATAAGAACCTGGTTTCCTTGGGTTA TCAGCTTACTAAGCCAGATGTGATCCTCCGGTTGGAGAAGGGAGAAGAGCCC
融合蛋白 3 氨基酸序列 3A-ADD-3L-NLS-dCas9-NLS-KRAB (SEQ ID NO: 1502)	MNHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCEDSITV GMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRLFFEF YRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSAAHRARYFWG NLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEKEDI LWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLKEYFACVSSGNSNAN SRGPSFSSGLVPLSLRGSHMEVKVNRRSIEDICLCCGTLQVYTRHPLFEGGLCAPCKDKFLE SLFLYDDDGHQSYCTICCSGGTLFICESPDCTRCYCFECVDILVGPGTSERINAMACWVCFL CLPFSRSGLLQRRKRWRHQLKAFHDQEGAGPMEIYKTVSAWKRQPVRVLSLFRNIDKVLKSL GFLESGSGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPGWYMFQFHRI LQYALPRQESQRPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVRGRDYQNAMRVWSNIP GLKSKHAPLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNCLLPLREYFKYFSQNSLPLGGPSS

	GAPPPSGGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPS EGSAPGTSTEPSEPKKKRKVMDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHS IKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEES FLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRG HFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQ LPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLF LAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQI HLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPA FLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRL SRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHI ANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEE GIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLK DDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVK KTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLK SVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGEL QKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVIL ADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD ATLIHQSITGLYETRIDLSQLGGDPKKKRKVSGSETPGTSESATPESTGRTLVTFKDVFVDF TREEWKLLDTAQQIVYRNVMLENYKNLVSLGYQLTKPDVILRLEKGEEP
融合蛋白 3 DNA 序列 (SEQ ID NO: 1503)	ATGAACCATGACCAGGAATTTGACCCCCCAAAGGTTTACCCACCTGTGCCAGCTGAGAAGAG GAAGCCCATCCGCGTGCTGTCTCTCTTTGATGGGATTGCTACAGGGCTCCTGGTGCTGAAGG ACCTGGGCATCCAAGTGGACCGCTACATTGCCTCCGAGGTGTGTGAGGACTCCATCACGGTG GGCATGGTGCGGCACCAGGGAAAGATCATGTACGTCGGGGACGTCCGCAGCGTCACACAGAA GCATATCCAGGAGTGGGGCCCATTCGACCTGGTGATTGGAGGCAGTCCCTGCAATGACCTCT CCATTGTCAACCCTGCCCGCAAGGGACTTTATGAGGGTACTGGCCGCCTCTTCTTTGAGTTC TACCGCCTCCTGCATGATGCGCGGCCCAAGGAGGGAGATGATCGCCCCTTCTTCTGGCTCTT TGAGAATGTGGTGGCCATGGGCGTTAGTGACAAGAGGGACATCTCGCGATTTCTTGAGTCTA ACCCCGTGATGATTGACGCCAAAGAAGTGTCTGCTGCACACAGGGCCCGTTACTTCTGGGGT AACCTTCCTGGCATGAACAGGCCTTTGGCATCCACTGTGAATGATAAGCTGGAGCTGCAAGA GTGTCTGGAGCACGGCAGAATAGCCAAGTTCAGCAAAGTGAGGACCATTACCACCAGGTCAA ACTCTATAAAGCAGGGCAAAGACCAGCATTTCCCCGTCTTCATGAACGAGAAGGAGGACATC CTGTGGTGCACTGAAATGGAAAGGGTGTTTGGCTTCCCCGTCCACTACACAGACGTCTCCAA CATGAGCCGCTTGGCGAGGCAGAGACTGCTGGGCCGATCGTGGAGCGTGCCGGTCATCCGCC ACCTCTTCGCTCCGCTGAAGGAATATTTTGCTTGTGTGTCTAGCGGCAATAGTAACGCTAAC AGCCGCGGGCCGAGCTTCAGCAGCGGCCTGGTGCCGTTAAGCTTGCGCGGCAGCCATATGGA AGTCAAAGTGAACCGACGGAGCATTGAAGACATCTGCCTCTGCTGTGGAACTCTCCAGGTGT ACACTCGGCACCCCTTGTTTGAGGGAGGGTTATGTGCCCCATGTAAGGATAAGTTCCTGGAG TCCCTCTTCCTGTATGATGATGATGGACACCAGAGTTACTGCACCATCTGCTGTTCCGGGGG TACCCTGTTCATCTGTGAGAGCCCCGACTGTACCAGATGCTACTGTTTCGAGTGTGTGGACA TCCTGGTGGGCCCCGGGACCTCAGAGAGGATCAATGCCATGGCCTGCTGGGTTTGCTTCCTG TGCCTGCCCTTCTCACGGAGTGGACTGCTGCAGAGGCGCAAGAGGTGGCGGCACCAGCTGAA GGCCTTCCATGATCAAGAGGGAGCGGGCCCTATGGAGATATACAAGACAGTGTCTGCATGGA AGAGACAGCCAGTGCGGGTACTGAGCCTCTTCAGAAACATCGACAAGGTACTAAAGAGTTTG GGCTTCTTGGAAAGCGGTTCTGGTTCTGGGGGAGGAACGCTGAAGTACGTGGAAGATGTCAC AAATGTCGTGAGGAGAGACGTGGAGAAATGGGGCCCCTTTGACCTGGTGTACGGCTCGACGC AGCCCCTAGGCAGCTCTTGTGATCGCTGTCCCGGCTGGTACATGTTCCAGTTCCACCGGATC CTGCAGTATGCGCTGCCTCGCCAGGAGAGTCAGCGGCCCTTCTTCTGGATATTCATGGACAA TCTGCTGCTGACTGAGGATGACCAAGAGACAACTACCCGCTTCCTTCAGACAGAGGCTGTGA CCCTCCAGGATGTCCGTGGCAGAGACTACCAGAATGCTATGCGGGTGTGGAGCAACATTCCA GGGCTGAAGAGCAAGCATGCGCCCCTGACCCCAAAGGAAGAAGAGTATCTGCAAGCCCAAGT CAGAAGCAGGAGCAAGCTGGACGCCCCGAAAGTTGACCTCCTGGTGAAGAACTGCCTTCTCC CGCTGAGAGAGTACTTCAAGTATTTTTCTCAAAACTCACTTCCTCTTGGAGGGCCGAGCTCT GGCGCACCCCCACCAAGTGGAGGGTCTCCTGCCGGGTCCCCAACATCTACTGAAGAAGGCAC CAGCGAATCCGCAACGCCCGAGTCAGGCCCTGGTACCTCCACAGAACCATCTGAAGGTAGTG CGCCTGGTTCCCCAGCTGGAAGCCCTACTTCCACCGAAGAAGGCACGTCAACCGAACCAAGT GAAGGATCTGCCCCTGGGACCAGCACTGAACCATCTGAGCCAAAAAAGAAGAGAAAGGTAAT

[0463]

[0464]

GGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCA
CCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGC
ATCAAGAAGAACCTGATCGGCGCCCTGCTGTTCGACAGCGGAGAAACAGCCGAGGCCACCCG
GCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAG
AGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCC
TTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGA
GGTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCA
CCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGC
CACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTCATCCA
GCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACG
CCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAG
CTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGCAACCTGATTGCCCTGAGCCTGGGCCTGAC
CCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACA
CCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTT
CTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGA
GATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACC
TGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTC
GACCAGAGCAAGAACGGCTACGCCGGCTACATCGATGGCGGAGCCAGCCAGGAAGAGTTCTA
CAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGA
ACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATC
CACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGA
CAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGG
CCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGG
AACTTCGAGGAAGTGGTGGACAAGGGCGCCAGCGCCCAGAGCTTCATCGAGCGGATGACCAA
CTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACT
TCACCGTGTACAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCC
TTCCTGAGCGGCGAGCAGAAAAAAGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGT
GACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAA
TCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATT
ATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGCT
GACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCCACC
TGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTG
AGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCT
GAAGTCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCT
TTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATT
GCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGA
CGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAG
AGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAG
GGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCA
GAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAAC
TGGACATCAACCGGCTGTCCGACTACGATGTGGACGCTATCGTGCCTCAGAGCTTTCTGAAG
GACGACTCCATCGATAACAAAGTGCTGACTCGGAGCGACAAGAACCGGGGCAAGAGCGACAA
CGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGCCAGCTGCTGAATGCCA
AGCTGATTACCCAGAGGAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAA
CTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGT
GGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAACGACAAACTGATCCGGG
AAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTT
TACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGT
GGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACA
AGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCC
AAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGG
CGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACAGGCGAGATCGTGTGGGATA
AGGGCCGGGACTTTGCCACCGTGCGGAAAGTGCTGTCTATGCCCCAAGTGAATATCGTGAAA
AAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGA
CAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCA
CCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAG
AGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCAT
CGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTA
AGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTG
CAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCA
CTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGC
ACAAACACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTG

[0465]

	GCCGACGCTAATCTGGACAAGGTGCTGAGCGCCTACAACAAGCACAGAGACAAGCCTATCAG AGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCT TCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGAC GCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCT GGGAGGCGACCCAAAAAAGAAGAGAAAGGTAAGCGGAAGTGAGACCCCAGGTACATCCGAAT CAGCAACGCCTGAAAGCACCGGTCGGACACTGGTGACCTTCAAGGATGTATTTGTGGACTTC ACCAGGGAGGAGTGGAAGCTGCTGGACACTGCTCAGCAGATCGTGTACAGAAATGTGATGCT GGAGAACTATAAGAACCTGGTTTCCTTGGGTTATCAGCTTACTAAGCCAGATGTGATCCTCC GGTTGGAGAAGGGAGAAGAGCCC
融合蛋白 4 氨基酸序列 3A-ADD-h3L-NLS-dCas9-NLS-KRAB (SEQ ID NO: 1504)	MNHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCEDSITV GMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRLFFEF YRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSAAHRARYFWG NLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEKEDI LWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLKEYFACVSSGNSNAN SRGPSFSSGLVPLSLRGSHMEVKANQRNIEDICICCGSLQVHTQHPLFEGGICAPCKDKFLD ALFLYDDDGYQSYCSICCSGETLLICGNPDCTRCYCFECVDSLVGPGTSGKVHAMSNWVCYL CLPSSRSGLLQRRRKWRSQLKAFYDRESENPLEMFETVPVWRRQPVRVLSLFEDIKKELTSL GFLESGSDPGQLKHVVDVTDTVRKDVEEWGPFDLVYGATPPLGHTCDRPPSWYLFQFHRLLQ YARPKPGSPRPFFWMFVDNLVLNKEDLDVASRFLEMEPVTIPDVHGGSLQNAVRVWSNIPAI RSRHWALVSEEELSLLAQNKQSSKLAAKWPTKLVKNCFLPLREYFKYFSTELTSSLGGPSSG APPPSGGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSE GSAPGTSTEPSEPKKKRKVMDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSI KKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESF LVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGH FLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQL PGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFL AAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFD QSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIH LGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKII KDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLS RKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA NLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEG IKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKD DSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAK YFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKK TEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKS VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQ KGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILA DANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDA TLIHQSITGLYETRIDLSQLGGDPKKKRKVSGSETPGTSESATPESTGRTLVTFKDVFVDFT REEWKLLDTAQQIVYRNVMLENYKNLVSLGYQLTKPDVILRLEKGEEP
融合蛋白 4 DNA 序列 (SEQ ID NO: 1505)	ATGAACCATGATCAAGAATTCGACCCACCTAAAGTCTACCCACCTGTGCCCGCCGAAAAAAG GAAACCCATAAGGGTGCTGTCACTCTTTGATGGCATCGCCACTGGTCTCCTGGTTCTTAAGG ATCTGGGAATTCAGGTCGATCGGTACATTGCTAGCGAGGTTTGTGAGGATAGTATTACAGTG GGTATGGTGCGCCACCAGGGAAAGATCATGTATGTTGGTGACGTTAGGAGCGTCACCCAGAA ACATATCCAGGAGTGGGGACCCTTTGATTTGGTGATCGGAGGTAGTCCCTGCAATGACCTTT CCATCGTGAATCCAGCCAGGAAAGGGCTGTATGAAGGGACTGGTAGGCTCTTTTTCGAGTTT TATCGCCTGCTTCACGACGCTAGACCTAAGGAAGGTGACGATAGGCCTTTCTTTTGGCTTTT TGAGAACGTCGTGGCAATGGGAGTCTCCGACAAAAGGGACATTTCTCGCTTTCTGGAATCTA ACCCCGTTATGATCGATGCCAAGGAAGTTTCTGCCGCTCACAGGGCAAGGTACTTCTGGGGC AATCTGCCCGGAATGAATCGCCCACTGGCCAGTACCGTGAATGACAAACTGGAGCTGCAGGA GTGCCTGGAGCACGGAAGAATCGCAAAGTTTTCTAAAGTCAGGACCATTACCACTCGCAGTA ACTCCATAAAACAGGGTAAGGACCAGCATTTTCCCGTCTTCATGAATGAAAAGGAAGATATT CTGTGGTGCACTGAAATGGAGAGAGTTTTCGGGTTTCCCGTGCACTATACCGATGTTTCCAA CATGTCCCGCCTTGCAAGACAAAGGCTTTTGGGCCGCTCTTGGTCTGTGCCAGTGATCCGGC ACTTGTTTGCTCCCCTCAAAGAGTACTTCGCTTGCGTCAGTTCCGGAAATTCAAACGCTAAC TCTCGGGGTCCATCTTTCTCCAGTGGTCTCGTGCCACTGTCTCTCCGGGGCTCTCACATGGA AGTCAAGGCTAACCAGCGAAATATAGAAGACATCTGCATCTGCTGCGGAAGTCTCCAGGTTC

[0466]

ACACACAGCACCCTCTGTTTGAGGGAGGGATCTGCGCCCCATGTAAGGACAAGTTCCTGGAT
GCCCTCTTCCTGTACGACGATGACGGGTACCAATCCTACTGCTCCATCTGCTGCTCCGGAGA
GACGCTGCTCATCTGCGGAAACCCTGATTGCACCCGATGCTACTGCTTCGAGTGTGTGGATA
GCCTGGTCGGCCCCGGGACCTCGGGGAAGGTGCACGCCATGAGCAACTGGGTGTGCTACCTG
TGCCTGCCGTCCTCCCGAAGCGGGCTGCTGCAGCGTCGGAGGAAGTGGCGCAGCCAGCTCAA
GGCCTTCTACGACCGAGAGTCGGAGAATCCCCTGGAGATGTTTGAGACAGTGCCAGTCTGGC
GGAGGCAGCCCGTTCGCGTTCTCTCTCTGTTCGAAGATATTAAAAAGGAACTCACCTCCCTT
GGGTTCCTGGAGAGCGGGAGCGACCCCGGACAGCTTAAGCACGTGGTCGACGTGACTGACAC
CGTCCGCAAAGACGTGGAGGAATGGGGCCCCTTCGATCTGGTCTATGGGGCAACCCCTCCCC
TTGGGCATACATGTGATCGGCCTCCATCCTGGTACCTGTTCCAGTTTCACAGACTCCTGCAG
TATGCCAGGCCAAAGCCAGGGAGCCCAAGGCCCTTTTTCTGGATGTTCGTCGACAACCTGGT
CCTGAACAAAGAAGATCTCGACGTTGCTAGTCGCTTTCTCGAAATGGAGCCCGTGACCATTC
CCGACGTGCATGGCGGTTCCCTCCAGAATGCAGTCAGGGTTTGGAGCAATATCCCTGCCATC
AGGTCAAGGCACTGGGCACTGGTTTCAGAGGAAGAGCTGTCCCTCCTTGCCCAGAACAAGCA
GTCATCCAAACTGGCAGCCAAGTGGCCAACTAAGCTGGTCAAGAACTGCTTTCTTCCCCTCA
GAGAATATTTTAAGTATTTCAGTACTGAACTGACTAGCAGTCTGGGAGGGCCGAGCTCTGGC
GCACCCCCACCAAGTGGAGGGTCTCCTGCCGGGTCCCCAACATCTACTGAAGAAGGCACCAG
CGAATCCGCAACGCCCGAGTCAGGCCCTGGTACCTCCACAGAACCATCTGAAGGTAGTGCGC
CTGGTTCCCCAGCTGGAAGCCCTACTTCCACCGAAGAAGGCACGTCAACCGAACCAAGTGAA
GGATCTGCCCCTGGGACCAGCACTGAACCATCTGAGCCAAAAAAGAAGAGAAAGGTAATGGA
CAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCG
ACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATC
AAGAAGAACCTGATCGGCGCCCTGCTGTTCGACAGCGGAGAAACAGCCGAGGCCACCCGGCT
GAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGA
TCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTC
CTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCG
ACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCAC
TTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTCATCCAGCT
GGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACGCCA
AGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTG
CCCGGCGAGAAGAAGAATGGCCTGTTCGGCAACCTGATTGCCCTGAGCCTGGGCCTGACCCC
CAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCT
ACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTG
GCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGAT
CACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGA
CCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGAC
CAGAGCAAGAACGGCTACGCCGGCTACATCGATGGCGGAGCCAGCCAGGAAGAGTTCTACAA
GTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACA
GAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCAC
CTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAA
CCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCA
GGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAAC
TTCGAGGAAGTGGTGGACAAGGGCGCCAGCGCCCAGAGCTTCATCGAGCGGATGACCAACTT
CGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCA
CCGTGTACAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTC
CTGAGCGGCGAGCAGAAAAAAGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGAC
CGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCT
CCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATC
AAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGAC
CCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCCACCTGT
TCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGC
CGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAA
GTCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTA
AAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCC
AATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGA
GCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGA
ACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGC
ATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAA
CGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGG
ACATCAACCGGCTGTCCGACTACGATGTGGACGCTATCGTGCCTCAGAGCTTTCTGAAGGAC
GACTCCATCGATAACAAAGTGCTGACTCGGAGCGACAAGAACCGGGGCAAGAGCGACAACGT

		GCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGCCAGCTGCTGAATGCCAAGC TGATTACCCAGAGGAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTG GATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGC ACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAACGACAAACTGATCCGGGAAG TGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTAC AAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGG AACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGG TGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAG TACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGA GATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACAGGCGAGATCGTGTGGGATAAGG GCCGGGACTTTGCCACCGTGCGGAAAGTGCTGTCTATGCCCCAAGTGAATATCGTGAAAAAG ACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGACAA GCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCG TGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGT GTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGA CTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGT ACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAG AAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTA TGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACA AACACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCC GACGCTAATCTGGACAAGGTGCTGAGCGCCTACAACAAGCACAGAGACAAGCCTATCAGAGA GCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCA AGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCC ACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGG AGGCGACCCAAAAAAGAAGAGAAAGGTAAGCGGAAGTGAGACCCCAGGTACATCCGAATCAG CAACGCCTGAAAGCACCGGTCGGACACTGGTGACCTTCAAGGATGTATTTGTGGACTTCACC AGGGAGGAGTGGAAGCTGCTGGACACTGCTCAGCAGATCGTGTACAGAAATGTGATGCTGGA GAACTATAAGAACCTGGTTTCCTTGGGTTATCAGCTTACTAAGCCAGATGTGATCCTCCGGT TGGAGAAGGGAGAAGAGCCC
[0467]	融合蛋白 5 氨基酸序列 NLS-3A-3L-dCas9-KRAB-NLS (SEQ ID NO: 1506)	MPKKKRKVPKKKRKVNHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVD RYIASEVCEDSITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPAR KGLYEGTGRLFFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDA KEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGK DQHFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLK EYFACVSSGNSNANSRGPSFSSGLVPLSLRGSHNPLEMFETVPVWRRQPVRVLSLFEDIKKE LTSLGFLESGSDPGQLKHVVDVTDTVRKDVEEWGPFDLVYGATPPLGHTCDRPPSWYLFQFH RLLQYARPKPGSPRPFFWMFVDNLVLNKEDLDVASRFLEMEPVTIPDVHGGSLQNAVRVWSN IPAIRSRHWALVSEEELSLLAQNKQSSKLAAKWPTKLVKNCFLPLREYFKYFSTELTSSLGG PSSGAPPPSGGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTST EPSEGSAPGTSTEPSEMDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKN LIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVE EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLI EGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGE KKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAK NLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSK NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGE LHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSG EQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKL INGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSI DNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKA GFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFF YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADAN LDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLI HQSITGLYETRIDLSQLGGDSGSETPGTSESATPESTGRTLVTFKDVFVDFTREEWKLLDTA

[0468]

	QQIVYRNVMLENYKNLVSLGYQLTKPDVILRLEKGEEPSADYKDDDDKAPKKKRKVPKKKRK V
融合蛋白 5 DNA 序列 (SEQ ID NO: 1507)	ATGCCAAAAAAGAAGAGAAAGGTACCGAAGAAAAAAAGAAAGGTCAACCACGACCAGGAATT CGACCCTCCAAAGGTTTACCCACCTGTCCCAGCTGAGAAGAGGAAGCCCATCCGGGTGCTGT CTCTCTTTGATGGAATCGCTACAGGGCTCCTGGTGCTGAAGGACTTGGGCATTCAGGTGGAC CGCTACATTGCCTCGGAGGTGTGTGAGGACTCCATCACGGTGGGCATGGTGCGGCACCAGGG GAAGATCATGTACGTCGGGGACGTCCGCAGCGTCACACAGAAGCATATCCAGGAGTGGGGCC CATTCGATCTGGTGATTGGGGGCAGTCCCTGCAATGACCTCTCCATCGTCAACCCTGCTCGC AAGGGCCTCTACGAGGGCACTGGCCGGCTCTTCTTTGAGTTCTACCGCCTCCTGCATGATGC GCGGCCCAAGGAGGGAGATGATCGCCCCTTCTTCTGGCTCTTTGAGAATGTGGTGGCCATGG GCGTTAGTGACAAGAGGGACATCTCGCGATTTCTCGAGTCCAACCCTGTGATGATTGATGCC AAAGAAGTGTCAGCTGCACACAGGGCCCGCTACTTCTGGGGTAACCTTCCCGGTATGAACAG GCCGTTGGCATCCACTGTGAATGATAAGCTGGAGCTGCAGGAGTGTCTGGAGCATGGCAGGA TAGCCAAGTTCAGCAAAGTGAGGACCATTACTACGAGGTCAAACTCCATAAAGCAGGGCAAA GACCAGCATTTTCCTGTCTTCATGAATGAGAAAGAGGACATCTTATGGTGCACTGAAATGGA AAGGGTATTTGGTTTCCCAGTCCACTATACTGACGTCTCCAACATGAGCCGCTTGGCGAGGC AGAGACTGCTGGGCCGGTCATGGAGCGTGCCAGTCATCCGCCACCTCTTCGCTCCGCTGAAG GAGTATTTTGCGTGTGTGTCTAGCGGCAATAGTAACGCTAACAGCCGCGGGCCGAGCTTCAG CAGCGGCCTGGTGCCGTTAAGCTTGCGCGGCAGCCATAATCCCCTTGAGATGTTCGAAACCG TGCCTGTGTGGAGGAGACAGCCAGTCCGGGTGCTGTCCCTTTTTGAAGACATCAAGAAAGAG CTGACGAGTTTGGGCTTTTTGGAAAGTGGTTCTGACCCGGGACAACTGAAGCATGTGGTTGA TGTCACAGACACAGTGAGGAAGGATGTGGAGGAGTGGGGACCCTTCGATCTTGTGTACGGCG CCACACCTCCCCTGGGCCACACCTGTGACCGTCCTCCCAGCTGGTACCTGTTCCAGTTCCAC CGGCTCCTGCAGTACGCACGGCCCAAGCCAGGCAGCCCCAGGCCCTTCTTCTGGATGTTCGT GGACAATCTGGTGCTGAACAAGGAAGACCTGGACGTCGCATCTCGCTTCCTGGAGATGGAGC CAGTCACCATCCCAGATGTCCACGGCGGATCCTTGCAGAATGCTGTCCGCGTGTGGAGCAAC ATCCCAGCCATAAGGAGCAGGCACTGGGCTCTGGTTTCGGAAGAAGAATTGTCCCTGCTGGC CCAGAACAAGCAGAGCTCGAAGCTCGCGGCCAAGTGGCCCACCAAGCTGGTGAAGAACTGCT TTCTCCCCCTAAGAGAATATTTCAAGTATTTTTCAACAGAACTCACTTCCTCTTTAGGAGGG CCGAGCTCTGGCGCACCCCCACCAAGTGGAGGGTCTCCTGCCGGGTCCCCAACATCTACTGA AGAAGGCACCAGCGAATCCGCAACGCCCGAGTCAGGCCCTGGTACCTCCACAGAACCATCTG AAGGTAGTGCGCCTGGTTCCCCAGCTGGAAGCCCTACTTCCACCGAAGAAGGCACGTCAACC GAACCAAGTGAAGGATCTGCCCCTGGGACCAGCACTGAACCATCTGAGATGGACAAGAAGTA CAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACA AGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAAC CTGATCGGCGCCCTGCTGTTCGACAGCGGAGAAACAGCCGAGGCCACCCGGCTGAAGAGAAC CGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCA ACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAA GAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCA CGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCG ACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATC GAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGAC CTACAACCAGCTGTTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCC TGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAG AAGAAGAATGGCCTGTTCGGCAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAA GAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACG ACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAG AACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGC CCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGA AAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAG AACGGCTACGCCGGCTACATCGATGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAA GCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACC TGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAG CTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAA GATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACA GCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAA GTGGTGGACAAGGGCGCCAGCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAA CCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTACA ACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGC GAGCAGAAAAAAGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCA GCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGG AAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAG

[0469]

	GACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACT GTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACA AAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTG ATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGG CTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACA TCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCC GGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAA AGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCA CCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAG CTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCT GTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACC GGCTGTCCGACTACGATGTGGACGCTATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATC GATAACAAAGTGCTGACTCGGAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGA AGAGGTCGTGAAGAAGATGAAGAACTACTGGCGCCAGCTGCTGAATGCCAAGCTGATTACCC AGAGGAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCC GGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCT GGACTCCCGGATGAACACTAAGTACGACGAGAACGACAAACTGATCCGGGAAGTGAAAGTGA TCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGC GAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCT GATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACG TGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTC TACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAA GCGGCCTCTGATCGAGACAAACGGCGAAACAGGCGAGATCGTGTGGGATAAGGGCCGGGACT TTGCCACCGTGCGGAAAGTGCTGTCTATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTG CAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGACAAGCTGATCGC CAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATT CTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAG CTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGA AGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGT TCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAAC GAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCT GAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAACACTACC TGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAAT CTGGACAAGGTGCTGAGCGCCTACAACAAGCACAGAGACAAGCCTATCAGAGAGCAGGCCGA GAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTG ACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATC CACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGCGACAG CGGAAGTGAGACCCCAGGTACATCCGAATCAGCAACGCCTGAAAGCACCGGTCGGACACTGG TGACCTTCAAGGATGTATTTGTGGACTTCACCAGGGAGGAGTGGAAGCTGCTGGACACTGCT CAGCAGATCGTGTACAGAAATGTGATGCTGGAGAACTATAAGAACCTGGTTTCCTTGGGTTA TCAGCTTACTAAGCCAGATGTGATCCTCCGGTTGGAGAAGGGAGAAGAGCCCAGCGCTGATT ACAAAGATGATGACGATAAAGCCCCAAAAAAGAAGAGAAAGGTACCGAAGAAAAAAAGAAAG GTC
融合蛋白 6 氨基酸序列 NLS-3A-3L-dCas9-KRAB-NLS (SEQ ID NO: 1508)	MPKKKRKVPKKKRKVNHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVD RYIASEVCEDSITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPAR KGLYEGTGRLFFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDA KEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGK DQHFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLK EYFACVSSGNSNANSRGPSFSSGLVPLSLRGSHMGPMEIYKTVSAWKRQPVRVLSLFRNIDK VLKSLGFLESGSGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPGWYMF QFHRILQYALPRQESQRPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVRGRDYQNAMRV WSNIPGLKSKHAPLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNCLLPLREYFKYFSQNSLPL GGPSSGAPPPSGGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGT STEPSEGSAPGTSTEPSEMDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIK KNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHF LIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLA AKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHL GELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNF EEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL

[0470]

	SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIK DKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSR KLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIAN LAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGI KELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDD SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELD KAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYK VREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKY FFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKT EVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSV KELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQK GNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT LIHQSITGLYETRIDLSQLGGDSGSETPGTSESATPESTGRTLVTFKDVFVDFTREEWKLLD TAQQIVYRNVMLENYKNLVSLGYQLTKPDVILRLEKGEEPSADYKDDDDKAPKKKRKVPKKK RKV
融合蛋白 6 DNA 序列 (SEQ ID NO: 1509)	ATGCCAAAAAAGAAGAGAAAGGTACCGAAGAAAAAAAGAAAGGTATACAACCATGACCAGGA ATTCGACCCCCCAAAGGTTTACCCACCTGTGCCAGCTGAGAAGAGGAAGCCCATCCGCGTGC TGTCTCTCTTTGATGGGATTGCTACAGGGCTCCTGGTGCTGAAGGACCTGGGCATCCAAGTG GACCGCTACATTGCCTCCGAGGTGTGTGAGGACTCCATCACGGTGGGCATGGTGCGGCACCA GGGAAAGATCATGTACGTCGGGGACGTCCGCAGCGTCACACAGAAGCATATCCAGGAGTGGG GCCCATTCGACCTGGTGATTGGAGGCAGTCCCTGCAATGACCTCTCCATTGTCAACCCTGCC CGCAAGGGACTTTATGAGGGTACTGGCCGCCTCTTCTTTGAGTTCTACCGCCTCCTGCATGA TGCGCGGCCCAAGGAGGGAGATGATCGCCCCTTCTTCTGGCTCTTTGAGAATGTGGTGGCCA TGGGCGTTAGTGACAAGAGGGACATCTCGCGATTTCTTGAGTCTAACCCCGTGATGATTGAC GCCAAAGAAGTGTCTGCTGCACACAGGGCCCGTTACTTCTGGGGTAACCTTCCTGGCATGAA CAGGCCTTTGGCATCCACTGTGAATGATAAGCTGGAGCTGCAAGAGTGTCTGGAGCACGGCA GAATAGCCAAGTTCAGCAAAGTGAGGACCATTACCACCAGGTCAAACTCTATAAAGCAGGGC AAAGACCAGCATTTCCCCGTCTTCATGAACGAGAAGGAGGACATCCTGTGGTGCACTGAAAT GGAAAGGGTGTTTGGCTTCCCCGTCCACTACACAGACGTCTCCAACATGAGCCGCTTGGCGA GGCAGAGACTGCTGGGCCGATCGTGGAGCGTGCCGGTCATCCGCCACCTCTTCGCTCCGCTG AAGGAATATTTTGCTTGTGTGTCTAGCGGCAATAGTAACGCTAACAGCCGCGGGCCGAGCTT CAGCAGCGGCCTGGTGCCGTTAAGCTTGCGCGGCAGCCATATGGGCCCTATGGAGATATACA AGACAGTGTCTGCATGGAAGAGACAGCCAGTGCGGGTACTGAGCCTCTTCAGAAACATCGAC AAGGTACTAAAGAGTTTGGGCTTCTTGGAAAGCGGTTCTGGTTCTGGGGGAGGAACGCTGAA GTACGTGGAAGATGTCACAAATGTCGTGAGGAGAGACGTGGAGAAATGGGGCCCCTTTGACC TGGTGTACGGCTCGACGCAGCCCCTAGGCAGCTCTTGTGATCGCTGTCCCGGCTGGTACATG TTCCAGTTCCACCGGATCCTGCAGTATGCGCTGCCTCGCCAGGAGAGTCAGCGGCCCTTCTT CTGGATATTCATGGACAATCTGCTGCTGACTGAGGATGACCAAGAGACAACTACCCGCTTCC TTCAGACAGAGGCTGTGACCCTCCAGGATGTCCGTGGCAGAGACTACCAGAATGCTATGCGG GTGTGGAGCAACATTCCAGGGCTGAAGAGCAAGCATGCGCCCCTGACCCCAAAGGAAGAAGA GTATCTGCAAGCCCAAGTCAGAAGCAGGAGCAAGCTGGACGCCCCGAAAGTTGACCTCCTGG TGAAGAACTGCCTTCTCCCGCTGAGAGAGTACTTCAAGTATTTTTCTCAAAACTCACTTCCT CTTGGAGGGCCGAGCTCTGGCGCACCCCCACCAAGTGGAGGGTCTCCTGCCGGGTCCCCAAC ATCTACTGAAGAAGGCACCAGCGAATCCGCAACGCCCGAGTCAGGCCCTGGTACCTCCACAG AACCATCTGAAGGTAGTGCGCCTGGTTCCCCAGCTGGAAGCCCTACTTCCACCGAAGAAGGC ACGTCAACCGAACCAAGTGAAGGATCTGCCCCTGGGACCAGCACTGAACCATCTGAGATGGA CAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCG ACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATC AAGAAGAACCTGATCGGCGCCCTGCTGTTCGACAGCGGAGAAACAGCCGAGGCCACCCGGCT GAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGA TCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTC CTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCG ACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCAC TTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTCATCCAGCT GGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACGCCA AGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTG CCCGGCGAGAAGAAGAATGGCCTGTTCGGCAACCTGATTGCCCTGAGCCTGGGCCTGACCCC CAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCT ACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTG GCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGAT

[0471]

	CACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGA CCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGAC CAGAGCAAGAACGGCTACGCCGGCTACATCGATGGCGGAGCCAGCCAGGAAGAGTTCTACAA GTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACA GAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCAC CTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAA CCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCA GGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAAC TTCGAGGAAGTGGTGGACAAGGGCGCCAGCGCCCAGAGCTTCATCGAGCGGATGACCAACTT CGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCA CCGTGTACAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTC CTGAGCGGCGAGCAGAAAAAAGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGAC CGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCT CCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATC AAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGAC CCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCCACCTGT TCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGC CGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAA GTCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTA AAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCC AATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGA GCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGA ACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGC ATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAA CGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGG ACATCAACCGGCTGTCCGACTACGATGTGGACGCTATCGTGCCTCAGAGCTTTCTGAAGGAC GACTCCATCGATAACAAAGTGCTGACTCGGAGCGACAAGAACCGGGGCAAGAGCGACAACGT GCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGCCAGCTGCTGAATGCCAAGC TGATTACCCAGAGGAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTG GATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGC ACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAACGACAAACTGATCCGGGAAG TGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTAC AAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGG AACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGG TGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAG TACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGA GATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACAGGCGAGATCGTGTGGGATAAGG GCCGGGACTTTGCCACCGTGCGGAAAGTGCTGTCTATGCCCCAAGTGAATATCGTGAAAAAG ACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGACAA GCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCG TGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGT GTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGA CTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGT ACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAG AAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTA TGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACA AACACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCC GACGCTAATCTGGACAAGGTGCTGAGCGCCTACAACAAGCACAGAGACAAGCCTATCAGAGA GCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCA AGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCC ACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGG AGGCGACAGCGGAAGTGAGACCCCAGGTACATCCGAATCAGCAACGCCTGAAAGCACCGGTC GGACACTGGTGACCTTCAAGGATGTATTTGTGGACTTCACCAGGGAGGAGTGGAAGCTGCTG GACACTGCTCAGCAGATCGTGTACAGAAATGTGATGCTGGAGAACTATAAGAACCTGGTTTC CTTGGGTTATCAGCTTACTAAGCCAGATGTGATCCTCCGGTTGGAGAAGGGAGAAGAGCCCA GCGCTGATTACAAAGATGATGACGATAAAGCCCCAAAAAAGAAGAGAAAGGTACCGAAGAAA AAAAGAAAGGTC
融合蛋白 7 氨基酸序列 NLS-3A-3L-dCas9-ZIM-NLS	MPKKKRKVPKKKRKVNHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVD RYIASEVCEDSITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPAR KGLYEGTGRLFFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDA KEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGK DQHFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLK

	(SEQ ID NO: 1510)	EYFACVSSGNSNANSRGPSFSSGLVPLSLRGSHMGPMEIYKTVSAWKRQPVRVLSLFRNIDK VLKSLGFLESGSGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPGWYMF QFHRILQYALPRQESQRPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVRGRDYQNAMRV WSNIPGLKSKHAPLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNCLLPLREYFKYFSQNSLPL GGPSSGAPPPSGGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGT STEPSEGSAPGTSTEPSEMDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIK KNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHF LIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLA AKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHL GELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNF EEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIK DKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSR KLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIAN LAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGI KELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDD SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELD KAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYK VREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKY FFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKT EVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSV KELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQK GNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT LIHQSITGLYETRIDLSQLGGDSGSETPGTSESATPESTGMNNSQGRVTFEDVTVNFTQGEW QRLNPEQRNLYRDVMLENYSNLVSVGQGETTKPDVILRLEQGKEPWLEEEEVLGSGRAEKNG DIGGQIWKPKDVKESLSADYKDDDDKAPKKKRKVPKKKRKV
[0472]	融合蛋白 7 DNA 序列 (SEQ ID NO: 1511)	ATGCCAAAAAAGAAGAGAAAGGTACCGAAGAAAAAAAGAAAGGTATACAACCATGACCAGGA ATTCGACCCCCCAAAGGTTTACCCACCTGTGCCAGCTGAGAAGAGGAAGCCCATCCGCGTGC TGTCTCTCTTTGATGGGATTGCTACAGGGCTCCTGGTGCTGAAGGACCTGGGCATCCAAGTG GACCGCTACATTGCCTCCGAGGTGTGTGAGGACTCCATCACGGTGGGCATGGTGCGGCACCA GGGAAAGATCATGTACGTCGGGGACGTCCGCAGCGTCACACAGAAGCATATCCAGGAGTGGG GCCCATTCGACCTGGTGATTGGAGGCAGTCCCTGCAATGACCTCTCCATTGTCAACCCTGCC CGCAAGGGACTTTATGAGGGTACTGGCCGCCTCTTCTTTGAGTTCTACCGCCTCCTGCATGA TGCGCGGCCCAAGGAGGGAGATGATCGCCCCTTCTTCTGGCTCTTTGAGAATGTGGTGGCCA TGGGCGTTAGTGACAAGAGGGACATCTCGCGATTTCTTGAGTCTAACCCCGTGATGATTGAC GCCAAAGAAGTGTCTGCTGCACACAGGGCCCGTTACTTCTGGGGTAACCTTCCTGGCATGAA CAGGCCTTTGGCATCCACTGTGAATGATAAGCTGGAGCTGCAAGAGTGTCTGGAGCACGGCA GAATAGCCAAGTTCAGCAAAGTGAGGACCATTACCACCAGGTCAAACTCTATAAAGCAGGGC AAAGACCAGCATTTCCCCGTCTTCATGAACGAGAAGGAGGACATCCTGTGGTGCACTGAAAT GGAAAGGGTGTTTGGCTTCCCCGTCCACTACACAGACGTCTCCAACATGAGCCGCTTGGCGA GGCAGAGACTGCTGGGCCGATCGTGGAGCGTGCCGGTCATCCGCCACCTCTTCGCTCCGCTG AAGGAATATTTTGCTTGTGTGTCTAGCGGCAATAGTAACGCTAACAGCCGCGGGCCGAGCTT CAGCAGCGGCCTGGTGCCGTTAAGCTTGCGCGGCAGCCATATGGGCCCTATGGAGATATACA AGACAGTGTCTGCATGGAAGAGACAGCCAGTGCGGGTACTGAGCCTCTTCAGAAACATCGAC AAGGTACTAAAGAGTTTGGGCTTCTTGGAAAGCGGTTCTGGTTCTGGGGGAGGAACGCTGAA GTACGTGGAAGATGTCACAAATGTCGTGAGGAGAGACGTGGAGAAATGGGGCCCCTTTGACC TGGTGTACGGCTCGACGCAGCCCCTAGGCAGCTCTTGTGATCGCTGTCCCGGCTGGTACATG TTCCAGTTCCACCGGATCCTGCAGTATGCGCTGCCTCGCCAGGAGAGTCAGCGGCCCTTCTT CTGGATATTCATGGACAATCTGCTGCTGACTGAGGATGACCAAGAGACAACTACCCGCTTCC TTCAGACAGAGGCTGTGACCCTCCAGGATGTCCGTGGCAGAGACTACCAGAATGCTATGCGG GTGTGGAGCAACATTCCAGGGCTGAAGAGCAAGCATGCGCCCCTGACCCCAAAGGAAGAAGA GTATCTGCAAGCCCAAGTCAGAAGCAGGAGCAAGCTGGACGCCCCGAAAGTTGACCTCCTGG TGAAGAACTGCCTTCTCCCGCTGAGAGAGTACTTCAAGTATTTTTCTCAAAACTCACTTCCT CTTGGAGGGCCGAGCTCTGGCGCACCCCCACCAAGTGGAGGGTCTCCTGCCGGGTCCCCAAC ATCTACTGAAGAAGGCACCAGCGAATCCGCAACGCCCGAGTCAGGCCCTGGTACCTCCACAG AACCATCTGAAGGTAGTGCGCCTGGTTCCCCAGCTGGAAGCCCTACTTCCACCGAAGAAGGC ACGTCAACCGAACCAAGTGAAGGATCTGCCCCTGGGACCAGCACTGAACCATCTGAGATGGA CAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCG

[0473]

	ACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATC AAGAAGAACCTGATCGGCGCCCTGCTGTTCGACAGCGGAGAAACAGCCGAGGCCACCCGGCT GAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGA TCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTC CTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCG ACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCAC TTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTCATCCAGCT GGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACGCCA AGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTG CCCGGCGAGAAGAAGAATGGCCTGTTCGGCAACCTGATTGCCCTGAGCCTGGGCCTGACCCC CAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCT ACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTG GCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGAT CACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGA CCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGAC CAGAGCAAGAACGGCTACGCCGGCTACATCGATGGCGGAGCCAGCCAGGAAGAGTTCTACAA GTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACA GAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCAC CTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAA CCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCA GGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAAC TTCGAGGAAGTGGTGGACAAGGGCGCCAGCGCCCAGAGCTTCATCGAGCGGATGACCAACTT CGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCA CCGTGTACAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTC CTGAGCGGCGAGCAGAAAAAAGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGAC CGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCT CCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATC AAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGAC CCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCCACCTGT TCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGC CGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAA GTCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTA AAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCC AATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGA GCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGA ACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGC ATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAA CGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGG ACATCAACCGGCTGTCCGACTACGATGTGGACGCTATCGTGCCTCAGAGCTTTCTGAAGGAC GACTCCATCGATAACAAAGTGCTGACTCGGAGCGACAAGAACCGGGGCAAGAGCGACAACGT GCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGCCAGCTGCTGAATGCCAAGC TGATTACCCAGAGGAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTG GATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGC ACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAACGACAAACTGATCCGGGAAG TGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTAC AAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGG AACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGG TGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAG TACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGA GATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACAGGCGAGATCGTGTGGGATAAGG GCCGGGACTTTGCCACCGTGCGGAAAGTGCTGTCTATGCCCCAAGTGAATATCGTGAAAAAG ACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGACAA GCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCG TGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGT GTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGA CTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGT ACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAG AAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTA TGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACA AACACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCC GACGCTAATCTGGACAAGGTGCTGAGCGCCTACAACAAGCACAGAGACAAGCCTATCAGAGA

[0474]

	GCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCA AGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCC ACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGG AGGCGACAGCGGAAGTGAGACCCCAGGTACATCCGAATCAGCAACGCCTGAAAGCACCGGTA TGAACAATTCACAGGGGAGAGTGACATTCGAAGACGTGACCGTGAACTTCACCCAGGGAGAA TGGCAGCGCTTGAACCCAGAACAAAGGAACCTCTATCGGGACGTGATGCTGGAAAACTACTC AAATTTGGTGAGCGTTGGGCAGGGTGAGACCACTAAGCCTGACGTGATCCTGAGATTGGAAC AGGGCAAGGAGCCTTGGCTCGAGGAAGAGGAAGTCCTGGGCTCAGGGAGGGCCGAGAAAAAC GGTGATATAGGAGGCCAGATATGGAAGCCTAAGGACGTCAAGGAGAGCCTGAGCGCTGATTA CAAAGATGATGACGATAAAGCCCCAAAAAAGAAGAGAAAGGTACCGAAGAAAAAAAGAAAGG TC
融合蛋白 8 氨基酸序列 NLS-3A-3L-dCas9-ZFP-NLS (SEQ ID NO: 1512)	MPKKKRKVPKKKRKVNHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVD RYIASEVCEDSITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPAR KGLYEGTGRLFFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDA KEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGK DQHFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLK EYFACVSSGNSNANSRGPSFSSGLVPLSLRGSHMGPMEIYKTVSAWKRQPVRVLSLFRNIDK VLKSLGFLESGSGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPGWYMF QFHRILQYALPRQESQRPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVRGRDYQNAMRV WSNIPGLKSKHAPLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNCLLPLREYFKYFSQNSLPL GGPSSGAPPPSGGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGT STEPSEGSAPGTSTEPSEMDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIK KNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHF LIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLA AKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHL GELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNF EEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIK DKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSR KLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIAN LAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGI KELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDD SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELD KAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYK VREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKY FFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKT EVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSV KELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQK GNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT LIHQSITGLYETRIDLSQLGGDSGSETPGTSESATPESTGNKKLEAVGTGIEPKAMSQGLVT FGDVAVDFSQEEWEWLNPIQRNLYRKVMLENYRNLASLGLCVSKPDVISSLEQGKEPWSADY KDDDDKAPKKKRKVPKKKRKV
融合蛋白 8 DNA 序列 (SEQ ID NO: 1513)	ATGCCAAAAAAGAAGAGAAAGGTACCGAAGAAAAAAAGAAAGGTATACAACCATGACCAGGA ATTCGACCCCCCAAAGGTTTACCCACCTGTGCCAGCTGAGAAGAGGAAGCCCATCCGCGTGC TGTCTCTCTTTGATGGGATTGCTACAGGGCTCCTGGTGCTGAAGGACCTGGGCATCCAAGTG GACCGCTACATTGCCTCCGAGGTGTGTGAGGACTCCATCACGGTGGGCATGGTGCGGCACCA GGGAAAGATCATGTACGTCGGGGACGTCCGCAGCGTCACACAGAAGCATATCCAGGAGTGGG GCCCATTCGACCTGGTGATTGGAGGCAGTCCCTGCAATGACCTCTCCATTGTCAACCCTGCC CGCAAGGGACTTTATGAGGGTACTGGCCGCCTCTTCTTTGAGTTCTACCGCCTCCTGCATGA TGCGCGGCCCAAGGAGGGAGATGATCGCCCCTTCTTCTGGCTCTTTGAGAATGTGGTGGCCA TGGGCGTTAGTGACAAGAGGGACATCTCGCGATTTCTTGAGTCTAACCCCGTGATGATTGAC GCCAAAGAAGTGTCTGCTGCACACAGGGCCCGTTACTTCTGGGGTAACCTTCCTGGCATGAA CAGGCCTTTGGCATCCACTGTGAATGATAAGCTGGAGCTGCAAGAGTGTCTGGAGCACGGCA GAATAGCCAAGTTCAGCAAAGTGAGGACCATTACCACCAGGTCAAACTCTATAAAGCAGGGC AAAGACCAGCATTTCCCCGTCTTCATGAACGAGAAGGAGGACATCCTGTGGTGCACTGAAAT GGAAAGGGTGTTTGGCTTCCCCGTCCACTACACAGACGTCTCCAACATGAGCCGCTTGGCGA GGCAGAGACTGCTGGGCCGATCGTGGAGCGTGCCGGTCATCCGCCACCTCTTCGCTCCGCTG AAGGAATATTTTGCTTGTGTGTCTAGCGGCAATAGTAACGCTAACAGCCGCGGGCCGAGCTT

[0475]

CAGCAGCGGCCTGGTGCCGTTAAGCTTGCGCGGCAGCCATATGGGCCCTATGGAGATATACA
AGACAGTGTCTGCATGGAAGAGACAGCCAGTGCGGGTACTGAGCCTCTTCAGAAACATCGAC
AAGGTACTAAAGAGTTTGGGCTTCTTGGAAAGCGGTTCTGGTTCTGGGGGAGGAACGCTGAA
GTACGTGGAAGATGTCACAAATGTCGTGAGGAGAGACGTGGAGAAATGGGGCCCCTTTGACC
TGGTGTACGGCTCGACGCAGCCCCTAGGCAGCTCTTGTGATCGCTGTCCCGGCTGGTACATG
TTCCAGTTCCACCGGATCCTGCAGTATGCGCTGCCTCGCCAGGAGAGTCAGCGGCCCTTCTT
CTGGATATTCATGGACAATCTGCTGCTGACTGAGGATGACCAAGAGACAACTACCCGCTTCC
TTCAGACAGAGGCTGTGACCCTCCAGGATGTCCGTGGCAGAGACTACCAGAATGCTATGCGG
GTGTGGAGCAACATTCCAGGGCTGAAGAGCAAGCATGCGCCCCTGACCCCAAAGGAAGAAGA
GTATCTGCAAGCCCAAGTCAGAAGCAGGAGCAAGCTGGACGCCCCGAAAGTTGACCTCCTGG
TGAAGAACTGCCTTCTCCCGCTGAGAGAGTACTTCAAGTATTTTTCTCAAAACTCACTTCCT
CTTGGAGGGCCGAGCTCTGGCGCACCCCCACCAAGTGGAGGGTCTCCTGCCGGGTCCCCAAC
ATCTACTGAAGAAGGCACCAGCGAATCCGCAACGCCCGAGTCAGGCCCTGGTACCTCCACAG
AACCATCTGAAGGTAGTGCGCCTGGTTCCCCAGCTGGAAGCCCTACTTCCACCGAAGAAGGC
ACGTCAACCGAACCAAGTGAAGGATCTGCCCCTGGGACCAGCACTGAACCATCTGAGATGGA
CAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCG
ACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATC
AAGAAGAACCTGATCGGCGCCCTGCTGTTCGACAGCGGAGAAACAGCCGAGGCCACCCGGCT
GAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGA
TCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTC
CTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCG
ACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCAC
TTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTCATCCAGCT
GGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACGCCA
AGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTG
CCCGGCGAGAAGAAGAATGGCCTGTTCGGCAACCTGATTGCCCTGAGCCTGGGCCTGACCCC
CAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCT
ACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTG
GCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGAT
CACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGA
CCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGAC
CAGAGCAAGAACGGCTACGCCGGCTACATCGATGGCGGAGCCAGCCAGGAAGAGTTCTACAA
GTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACA
GAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCAC
CTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAA
CCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCA
GGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAAC
TTCGAGGAAGTGGTGGACAAGGGCGCCAGCGCCCAGAGCTTCATCGAGCGGATGACCAACTT
CGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCA
CCGTGTACAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTC
CTGAGCGGCGAGCAGAAAAAAGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGAC
CGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCT
CCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATC
AAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGAC
CCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCCACCTGT
TCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGC
CGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAA
GTCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTA
AAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCC
AATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGA
GCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGA
ACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGC
ATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAA
CGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGG
ACATCAACCGGCTGTCCGACTACGATGTGGACGCTATCGTGCCTCAGAGCTTTCTGAAGGAC
GACTCCATCGATAACAAAGTGCTGACTCGGAGCGACAAGAACCGGGGCAAGAGCGACAACGT
GCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGCCAGCTGCTGAATGCCAAGC
TGATTACCCAGAGGAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTG
GATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGC
ACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAACGACAAACTGATCCGGGAAG
TGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTAC

		AAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGG AACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGG TGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAG TACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGA GATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACAGGCGAGATCGTGTGGGATAAGG GCCGGGACTTTGCCACCGTGCGGAAAGTGCTGTCTATGCCCCAAGTGAATATCGTGAAAAAG ACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGACAA GCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCG TGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGT GTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGA CTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGT ACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAG AAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTA TGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACA AACACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCC GACGCTAATCTGGACAAGGTGCTGAGCGCCTACAACAAGCACAGAGACAAGCCTATCAGAGA GCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCA AGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCC ACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGG AGGCGACAGCGGAAGTGAGACCCCAGGTACATCCGAATCAGCAACGCCTGAAAGCACCGGTA ACAAAAAGCTTGAGGCCGTCGGAACCGGAATCGAACCAAAAGCAATGTCCCAGGGTTTGGTG ACATTTGGCGACGTGGCTGTCGATTTTTCCCAGGAAGAGTGGGAGTGGCTCAATCCTATCCA GAGGAACTTGTACCGGAAGGTGATGCTGGAGAATTATAGAAATTTGGCATCACTGGGGTTGT GCGTTAGCAAACCAGATGTTATATCTTCCCTGGAACAGGGAAAGGAGCCCTGGAGCGCTGAT TACAAAGATGATGACGATAAAGCCCCAAAAAAGAAGAGAAAGGTACCGAAGAAAAAAAGAAA GGTC
[0476]	融合蛋白 9 变体 1 氨基酸序列 NLS-NLS-3A-h3L-dCas9-KOX1KRAB-NLS-NLS (SEQ ID NO: 1514)	MGTMPKKKRKVPKKKRKVYNHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLG IQVDRYIASEVCEDSITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIV NPARKGLYEGTGRLFFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPV MIDAKEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSI KQGKDQHFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLF APLKEYFACVSSGNSNANSRGPSFSSGLVPLSLRGSHMAAIPALDPEAEPSMDVILVGSSEL SSSVSPGTGRDLIAYEVKANQRNIEDICICCGSLQVHTQHPLFEGGICAPCKDKFLDALFLY DDDGYQSYCSICCSGETLLICGNPDCTRCYCFECVDSLVGPGTSGKVHAMSNWVCYLCLPSS RSGLLQRRRKWRSQLKAFYDRESENPLEMFETVPVWRRQPVRVLSLFEDIKKELTSLGFLES GSDPGQLKHVVDVTDTVRKDVEEWGPFDLVYGATPPLGHTCDRPPSWYLFQFHRLLQYARPK PGSPRPFFWMFVDNLVLNKEDLDVASRFLEMEPVTIPDVHGGSLQNAVRVWSNIPAIRSRHW ALVSEEELSLLAQNKQSSKLAAKWPTKLVKNCFLPLREYFKYFSTELTSSLGGPSSGAPPPS GGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPG TSTEPSELEDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFD SGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGN LIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILL SDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYID GGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASA QSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEEN EDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQ SGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKG ILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKE HPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLV ETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHA HDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFF KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKE SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIME RSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKY VNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAY NKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLY ETRIDLSQLGGDSPKKKRKVGVDGSSGSETPGTSESATPESRTLVTFKDVFVDFTREEWKLL

[0477]

	DTAQQIVYRNVMLENYKNLVSLGYQLTKPDVILRLEKGEEPSADYKDDDDKAPKKKRKVPKK KRKV
融合蛋白 9 变体 1 DNA 序列 (SEQ ID NO: 1515)	ATGGGTACCATGCCAAAAAAGAAGAGAAAGGTACCGAAGAAAAAAAGAAAGGTATACAATCA CGATCAGGAGTTCGACCCCCCTAAGGTGTACCCACCAGTGCCTGCAGAGAAGAGGAAGCCAA TCCGGGTGCTGAGCCTGTTTGATGGCATCGCCACCGGCCTGCTGGTGCTGAAGGATCTGGGC ATCCAGGTGGACCGGTACATCGCCTCCGAGGTGTGCGAGGATTCTATCACCGTGGGCATGGT GCGCCACCAGGGCAAGATCATGTATGTGGGCGACGTGCGGTCCGTGACACAGAAGCACATCC AGGAGTGGGGCCCATTCGATCTGGTGATCGGCGGCAGCCCCTGTAATGACCTGTCCATCGTG AACCCTGCAAGGAAGGGACTGTACGAGGGAACCGGCCGGCTGTTCTTTGAGTTTTATAGACT GCTGCACGACGCCAGGCCTAAGGAGGGCGACGATAGACCATTCTTTTGGCTGTTCGAGAATG TGGTGGCTATGGGCGTGAGCGATAAGAGGGACATCTCCAGGTTTCTGGAGTCTAACCCCGTG ATGATCGATGCAAAGGAGGTGTCCGCCGCACACAGAGCCAGGTATTTCTGGGGCAATCTGCC AGGAATGAACAGGCCACTGGCAAGCACCGTGAATGACAAGCTGGAGCTGCAGGAGTGCCTGG AGCACGGAAGGATCGCCAAGTTTTCCAAGGTGCGCACAATCACCACACGGAGCAATTCCATC AAGCAGGGCAAGGATCAGCACTTCCCCGTGTTCATGAACGAGAAGGAGGACATCCTGTGGTG TACCGAGATGGAGAGAGTGTTCGGCTTTCCAGTGCACTACACAGACGTGTCTAACATGAGCA GGCTGGCAAGGCAGCGGCTGCTGGGCAGATCTTGGAGCGTGCCCGTGATCAGGCACCTGTTC GCCCCTCTGAAGGAGTATTTTGCCTGCGTGAGCAGCGGCAACTCCAATGCCAACAGCCGGGG CCCCTCTTTCAGCTCCGGATTGGTGCCTCTGAGCCTGAGGGGCTCCCACATGGCAGCAATCC CCGCCCTGGACCCCGAGGCCGAGCCTAGCATGGACGTGATCCTGGTGGGCTCTAGCGAGCTG TCCTCTAGCGTGTCTCCAGGAACCGGAAGGGATCTGATCGCATACGAGGTGAAGGCCAATCA GCGGAACATCGAGGACATCTGTATCTGCTGTGGCAGCCTGCAGGTGCACACACAGCACCCAC TGTTCGAGGGAGGAATCTGCGCACCCTGTAAGGATAAGTTCCTGGACGCCCTGTTTCTGTAC GACGATGACGGCTACCAGTCCTATTGCTCTATCTGCTGTTCCGGCGAGACCCTGCTGATCTG CGGCAATCCAGATTGTACAAGGTGCTATTGTTTTGAGTGCGTGGACTCTCTGGTGGGACCAG GCACCAGCGGAAAGGTGCACGCCATGTCCAACTGGGTGTGCTACCTGTGCCTGCCATCCTCT CGCAGCGGACTGCTGCAGCGGAGAAGGAAGTGGAGATCCCAGCTGAAGGCCTTCTATGATAG GGAGTCTGAGAACCCCCTGGAGATGTTTGAGACCGTGCCAGTGTGGCGCCGGCAGCCCGTGA GGGTGCTGAGCCTGTTCGAGGATATCAAGAAGGAGCTGACATCCCTGGGCTTTCTGGAGTCC GGCTCTGACCCCGGACAGCTGAAGCACGTGGTGGATGTGACCGACACAGTGCGGAAGGATGT GGAGGAGTGGGGCCCTTTCGACCTGGTGTACGGAGCAACCCCTCCACTGGGACACACATGCG ACAGACCCCCTTCTTGGTACCTGTTCCAGTTTCACCGCCTGCTGCAGTATGCAAGGCCAAAG CCAGGCAGCCCTAGACCATTCTTTTGGATGTTCGTGGATAATCTGGTGCTGAACAAGGAGGA TCTGGACGTGGCCAGCAGGTTTCTGGAGATGGAGCCAGTGACCATCCCAGACGTGCACGGCG GCTCCCTGCAGAATGCCGTGCGCGTGTGGTCTAACATCCCTGCCATCAGAAGCAGGCACTGG GCACTGGTGAGCGAGGAGGAGCTGTCCCTGCTGGCCCAGAATAAGCAGAGCAGCAAGCTGGC CGCCAAGTGGCCTACAAAGCTGGTGAAGAACTGCTTCCTGCCACTGCGGGAGTACTTCAAGT ATTTTTCCACCGAGCTGACATCTAGCCTGGGAGGACCCTCCTCTGGCGCCCCACCACCTAGC GGCGGCTCCCCTGCCGGCTCTCCAACCAGCACAGAGGAGGGCACCAGCGAGTCCGCCACACC AGAGTCTGGACCTGGCACCAGCACAGAGCCATCCGAGGGCTCTGCCCCAGGCTCTCCTGCAG GCAGCCCTACCTCCACCGAAGAGGGCACCAGCACAGAGCCTTCTGAGGGCAGCGCCCCAGGC ACCTCTACAGAGCCAAGCGAGCTCGAGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCAC CAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGG TGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGAC AGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACG GAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACA GCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCAC CCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCA CCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCC TGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAAC AGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAA CCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCA GACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGCAAC CTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGA TGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCCAGA TCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTG AGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAA GAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGC CTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGAC GGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGG CACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCG ACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAG

[0478]

	GAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCG CATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAA AGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCC CAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCC CAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACG TGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGAC CTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAA AATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGG GCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAAC GAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGA GGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGC GGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAG TCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGCA GCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCC AGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGC ATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGA GAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCC GCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAA CACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGG GCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACG CCATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGC GACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAA CTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCA AGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTG GAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTA CGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGT CCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCACGCC CACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGA AAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCG AGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTC AAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGG CGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGA GCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAG TCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAA GAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGG AAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAA AGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAA AAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGA GAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATAT GTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGA GCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCA GCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTAC AACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCT GACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGT ACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTAC GAGACACGGATCGACCTGTCTCAGCTGGGAGGCGACAGCCCCAAGAAGAAGAGAAAGGTGGG AGTCGACGGATCCAGCGGCTCCGAGACCCCAGGCACATCTGAGAGCGCCACCCCTGAGTCCC GGACCCTGGTGACCTTCAAGGATGTATTTGTGGACTTCACCAGGGAGGAGTGGAAGCTGCTG GACACTGCTCAGCAGATCGTGTACAGAAATGTGATGCTGGAGAACTATAAGAACCTGGTTTC CTTGGGTTATCAGCTTACTAAGCCAGATGTGATCCTCCGGTTGGAGAAGGGAGAAGAGCCCA GCGCTGATTACAAAGATGATGACGATAAAGCCCCAAAAAAGAAGAGAAAGGTACCGAAGAAA AAAAGAAAGGTCTGA
融合蛋白 9 变体 2 氨基酸序列 NLS-NLS-3A-h3L-dCas9-KOX1KRAB-NLS-NLS (SEQ ID NO: 1523)	MPKKKRKVPKKKRKVYNHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQV DRYIASEVCEDSITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPA RKGLYEGTGRLFFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMID AKEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQG KDQHFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPL KEYFACVSSGNSNANSRGPSFSSGLVPLSLRGSHMAAIPALDPEAEPSMDVILVGSSELSSS VSPGTGRDLIAYEVKANQRNIEDICICCGSLQVHTQHPLFEGGICAPCKDKFLDALFLYDDD GYQSYCSICCSGETLLICGNPDCTRCYCFECVDSLVGPGTSGKVHAMSNWVCYLCLPSSRSG LLQRRRKWRSQLKAFYDRESENPLEMFETVPVWRRQPVRVLSLFEDIKKELTSLGFLESGSD PGQLKHVVDVTDTVRKDVEEWGPFDLVYGATPPLGHTCDRPPSWYLFQFHRLLQYARPKPGS

		PRPFFWMFVDNLVLNKEDLDVASRFLEMEPVTIPDVHGGSLQNAVRVWSNIPAIRSRHWALV SEEELSLLAQNKQSSKLAAKWPTKLVKNCFLPLREYFKYFSTELTSSPGGPSSGAPPPSGGS PAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTST EPSELEDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGE TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDV DKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIA LSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDI LRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGA SQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDF YPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSF IERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLF KTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDI LEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGK TILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQ TVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPV ENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKN RGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA YLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTE ITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESIL PKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSS FEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNF LYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR IDLSQLGGDSPKKKRKVGVDGSSGSETPGTSESATPESRTLVTFKDVFVDFTREEWKLLDTA QQIVYRNVMLENYKNLVSLGYQLTKPDVILRLEKGEEPSADYKDDDDKAPKKKRKVPKKKRK V
[0479]	融合蛋白 9 变体 2 DNA 序列 (SEQ ID NO: 1524)	ATGGGTACCATGCCAAAAAAGAAGAGAAAGGTACCGAAGAAAAAAAGAAAGGTATACAATCA CGATCAGGAGTTCGACCCCCCTAAGGTGTACCCACCAGTGCCTGCAGAGAAGAGGAAGCCAA TCCGGGTGCTGAGCCTGTTTGATGGCATCGCCACCGGCCTGCTGGTGCTGAAGGATCTGGGC ATCCAGGTGGACCGGTACATCGCCTCCGAGGTGTGCGAGGATTCTATCACCGTGGGCATGGT GCGCCACCAGGGCAAGATCATGTATGTGGGCGACGTGCGGTCCGTGACACAGAAGCACATCC AGGAGTGGGGCCCATTCGATCTGGTGATCGGCGGCAGCCCCTGTAATGACCTGTCCATCGTG AACCCTGCAAGGAAGGGACTGTACGAGGGAACCGGCCGGCTGTTCTTTGAGTTTTATAGACT GCTGCACGACGCCAGGCCTAAGGAGGGCGACGATAGACCATTCTTTTGGCTGTTCGAGAATG TGGTGGCTATGGGCGTGAGCGATAAGAGGGACATCTCCAGGTTTCTGGAGTCTAACCCCGTG ATGATCGATGCAAAGGAGGTGTCCGCCGCACACAGAGCCAGGTATTTCTGGGGCAATCTGCC AGGAATGAACAGGCCACTGGCAAGCACCGTGAATGACAAGCTGGAGCTGCAGGAGTGCCTGG AGCACGGAAGGATCGCCAAGTTTTCCAAGGTGCGCACAATCACCACACGGAGCAATTCCATC AAGCAGGGCAAGGATCAGCACTTCCCCGTGTTCATGAACGAGAAGGAGGACATCCTGTGGTG TACCGAGATGGAGAGAGTGTTCGGCTTTCCAGTGCACTACACAGACGTGTCTAACATGAGCA GGCTGGCAAGGCAGCGGCTGCTGGGCAGATCTTGGAGCGTGCCCGTGATCAGGCACCTGTTC GCCCCTCTGAAGGAGTATTTTGCCTGCGTGAGCAGCGGCAACTCCAATGCCAACAGCCGGGG CCCCTCTTTCAGCTCCGGATTGGTGCCTCTGAGCCTGAGGGGCTCCCACATGGCAGCAATCC CCGCCCTGGACCCCGAGGCCGAGCCTAGCATGGACGTGATCCTGGTGGGCTCTAGCGAGCTG TCCTCTAGCGTGTCTCCAGGAACCGGAAGGGATCTGATCGCATACGAGGTGAAGGCCAATCA GCGGAACATCGAGGACATCTGTATCTGCTGTGGCAGCCTGCAGGTGCACACACAGCACCCAC TGTTCGAGGGAGGAATCTGCGCACCCTGTAAGGATAAGTTCCTGGACGCCCTGTTTCTGTAC GACGATGACGGCTACCAGTCCTATTGCTCTATCTGCTGTTCCGGCGAGACCCTGCTGATCTG CGGCAATCCAGATTGTACAAGGTGCTATTGTTTTGAGTGCGTGGACTCTCTGGTGGGACCAG GCACCAGCGGAAAGGTGCACGCCATGTCCAACTGGGTGTGCTACCTGTGCCTGCCATCCTCT CGCAGCGGACTGCTGCAGCGGAGAAGGAAGTGGAGATCCCAGCTGAAGGCCTTCTATGATAG GGAGTCTGAGAACCCCCTGGAGATGTTTGAGACCGTGCCAGTGTGGCGCCGGCAGCCCGTGA GGGTGCTGAGCCTGTTCGAGGATATCAAGAAGGAGCTGACATCCCTGGGCTTTCTGGAGTCC GGCTCTGACCCCGGACAGCTGAAGCACGTGGTGGATGTGACCGACACAGTGCGGAAGGATGT GGAGGAGTGGGGCCCTTTCGACCTGGTGTACGGAGCAACCCCTCCACTGGGACACACATGCG ACAGACCCCCTTCTTGGTACCTGTTCCAGTTTCACCGCCTGCTGCAGTATGCAAGGCCAAAG CCAGGCAGCCCTAGACCATTCTTTTGGATGTTCGTGGATAATCTGGTGCTGAACAAGGAGGA TCTGGACGTGGCCAGCAGGTTTCTGGAGATGGAGCCAGTGACCATCCCAGACGTGCACGGCG GCTCCCTGCAGAATGCCGTGCGCGTGTGGTCTAACATCCCTGCCATCAGAAGCAGGCACTGG GCACTGGTGAGCGAGGAGGAGCTGTCCCTGCTGGCCCAGAATAAGCAGAGCAGCAAGCTGGC

[0480]

CGCCAAGTGGCCTACAAAGCTGGTGAAGAACTGCTTCCTGCCACTGCGGGAGTACTTCAAGT
ATTTTTCCACCGAGCTGACATCTAGCCCCGGAGGACCCTCCTCTGGCGCCCCACCACCTAGC
GGCGGCTCCCCTGCCGGCTCTCCAACCAGCACAGAGGAGGGCACCAGCGAGTCCGCCACACC
AGAGTCTGGACCTGGCACCAGCACAGAGCCATCCGAGGGCTCTGCCCCAGGCTCTCCTGCAG
GCAGCCCTACCTCCACCGAAGAGGGCACCAGCACAGAGCCTTCTGAGGGCAGCGCCCCAGGC
ACCTCTACAGAGCCAAGCGAGCTCGAGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCAC
CAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGG
TGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGAC
AGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACG
GAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACA
GCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCAC
CCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCA
CCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCC
TGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAAC
AGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAA
CCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCA
GACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAATGGCCTGTTCGGCAAC
CTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGA
TGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCCAGA
TCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTG
AGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAA
GAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGC
CTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGAC
GGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGG
CACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCG
ACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAG
GAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCG
CATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAA
AGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCC
CAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCC
CAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACG
TGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGAC
CTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAA
AATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGG
GCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAAC
GAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGA
GGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGC
GGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAG
TCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGCA
GCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCC
AGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGC
ATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGA
GAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCC
GCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAA
CACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGG
GCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACG
CCATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGC
GACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAA
CTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCA
AGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTG
GAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTA
CGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGT
CCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCACGCC
CACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGA
AAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCG
AGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTC
AAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGG
CGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGA
GCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAG
TCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAA
GAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGG
AAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAA

[0481]

	AGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAA AAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGA GAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATAT GTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGA GCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCA GCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTAC AACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCT GACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGT ACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTAC GAGACACGGATCGACCTGTCTCAGCTGGGAGGCGACAGCCCCAAGAAGAAGAGAAAGGTGGG AGTCGACGGATCCAGCGGCTCCGAGACCCCAGGCACATCTGAGAGCGCCACCCCTGAGTCCC GGACCCTGGTGACCTTCAAGGATGTATTTGTGGACTTCACCAGGGAGGAGTGGAAGCTGCTG GACACTGCTCAGCAGATCGTGTACAGAAATGTGATGCTGGAGAACTATAAGAACCTGGTTTC CTTGGGTTATCAGCTTACTAAGCCAGATGTGATCCTCCGGTTGGAGAAGGGAGAAGAGCCCA GCGCTGATTACAAAGATGATGACGATAAAGCCCCAAAAAAGAAGAGAAAGGTACCGAAGAAA AAAAGAAAGGTCTGA
融合蛋白 10 氨基酸序列 NLS-NLS-3A-h3L-dCas9-ZFP-28-NLS-NLS (SEQ ID NO: 1516)	MGTMPKKKRKVPKKKRKVYNHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLG IQVDRYIASEVCEDSITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIV NPARKGLYEGTGRLFFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPV MIDAKEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSI KQGKDQHFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLF APLKEYFACVSSGNSNANSRGPSFSSGLVPLSLRGSHMAAIPALDPEAEPSMDVILVGSSEL SSSVSPGTGRDLIAYEVKANQRNIEDICICCGSLQVHTQHPLFEGGICAPCKDKFLDALFLY DDDGYQSYCSICCSGETLLICGNPDCTRCYCFECVDSLVGPGTSGKVHAMSNWVCYLCLPSS RSGLLQRRRKWRSQLKAFYDRESENPLEMFETVPVWRRQPVRVLSLFEDIKKELTSLGFLES GSDPGQLKHVVDVTDTVRKDVEEWGPFDLVYGATPPLGHTCDRPPSWYLFQFHRLLQYARPK PGSPRPFFWMFVDNLVLNKEDLDVASRFLEMEPVTIPDVHGGSLQNAVRVWSNIPAIRSRHW ALVSEEELSLLAQNKQSSKLAAKWPTKLVKNCFLPLREYFKYFSTELTSSLGGPSSGAPPPS GGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPG TSTEPSELEDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFD SGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGN LIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILL SDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYID GGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASA QSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEEN EDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQ SGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKG ILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKE HPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLV ETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHA HDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFF KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKE SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIME RSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKY VNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAY NKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLY ETRIDLSQLGGDSPKKKRKVGVDGSSGSETPGTSESATPESTGNKKLEAVGTGIEPKAMSQG LVTFGDVAVDFSQEEWEWLNPIQRNLYRKVMLENYRNLASLGLCVSKPDVISSLEQGKEPWS ADYKDDDDKAPKKKRKVPKKKRKV
融合蛋白 10 DNA 序列 (SEQ ID NO: 1517)	ATGGGTACCATGCCAAAAAAGAAGAGAAAGGTACCGAAGAAAAAAAGAAAGGTATACAATCA CGATCAGGAGTTCGACCCCCCTAAGGTGTACCCACCAGTGCCTGCAGAGAAGAGGAAGCCAA TCCGGGTGCTGAGCCTGTTTGATGGCATCGCCACCGGCCTGCTGGTGCTGAAGGATCTGGGC ATCCAGGTGGACCGGTACATCGCCTCCGAGGTGTGCGAGGATTCTATCACCGTGGGCATGGT GCGCCACCAGGGCAAGATCATGTATGTGGGCGACGTGCGGTCCGTGACACAGAAGCACATCC AGGAGTGGGGCCCATTCGATCTGGTGATCGGCGGCAGCCCCTGTAATGACCTGTCCATCGTG AACCCTGCAAGGAAGGGACTGTACGAGGGAACCGGCCGGCTGTTCTTTGAGTTTTATAGACT GCTGCACGACGCCAGGCCTAAGGAGGGCGACGATAGACCATTCTTTTGGCTGTTCGAGAATG

[0482]

TGGTGGCTATGGGCGTGAGCGATAAGAGGGACATCTCCAGGTTTCTGGAGTCTAACCCCGTG
ATGATCGATGCAAAGGAGGTGTCCGCCGCACACAGAGCCAGGTATTTCTGGGGCAATCTGCC
AGGAATGAACAGGCCACTGGCAAGCACCGTGAATGACAAGCTGGAGCTGCAGGAGTGCCTGG
AGCACGGAAGGATCGCCAAGTTTTCCAAGGTGCGCACAATCACCACACGGAGCAATTCCATC
AAGCAGGGCAAGGATCAGCACTTCCCCGTGTTCATGAACGAGAAGGAGGACATCCTGTGGTG
TACCGAGATGGAGAGAGTGTTCGGCTTTCCAGTGCACTACACAGACGTGTCTAACATGAGCA
GGCTGGCAAGGCAGCGGCTGCTGGGCAGATCTTGGAGCGTGCCCGTGATCAGGCACCTGTTC
GCCCCTCTGAAGGAGTATTTTGCCTGCGTGAGCAGCGGCAACTCCAATGCCAACAGCCGGGG
CCCCTCTTTCAGCTCCGGATTGGTGCCTCTGAGCCTGAGGGGCTCCCACATGGCAGCAATCC
CCGCCCTGGACCCCGAGGCCGAGCCTAGCATGGACGTGATCCTGGTGGGCTCTAGCGAGCTG
TCCTCTAGCGTGTCTCCAGGAACCGGAAGGGATCTGATCGCATACGAGGTGAAGGCCAATCA
GCGGAACATCGAGGACATCTGTATCTGCTGTGGCAGCCTGCAGGTGCACACACAGCACCCAC
TGTTCGAGGGAGGAATCTGCGCACCCTGTAAGGATAAGTTCCTGGACGCCCTGTTTCTGTAC
GACGATGACGGCTACCAGTCCTATTGCTCTATCTGCTGTTCCGGCGAGACCCTGCTGATCTG
CGGCAATCCAGATTGTACAAGGTGCTATTGTTTTGAGTGCGTGGACTCTCTGGTGGGACCAG
GCACCAGCGGAAAGGTGCACGCCATGTCCAACTGGGTGTGCTACCTGTGCCTGCCATCCTCT
CGCAGCGGACTGCTGCAGCGGAGAAGGAAGTGGAGATCCCAGCTGAAGGCCTTCTATGATAG
GGAGTCTGAGAACCCCCTGGAGATGTTTGAGACCGTGCCAGTGTGGCGCCGGCAGCCCGTGA
GGGTGCTGAGCCTGTTCGAGGATATCAAGAAGGAGCTGACATCCCTGGGCTTTCTGGAGTCC
GGCTCTGACCCCGGACAGCTGAAGCACGTGGTGGATGTGACCGACACAGTGCGGAAGGATGT
GGAGGAGTGGGGCCCTTTCGACCTGGTGTACGGAGCAACCCCTCCACTGGGACACACATGCG
ACAGACCCCCTTCTTGGTACCTGTTCCAGTTTCACCGCCTGCTGCAGTATGCAAGGCCAAAG
CCAGGCAGCCCTAGACCATTCTTTTGGATGTTCGTGGATAATCTGGTGCTGAACAAGGAGGA
TCTGGACGTGGCCAGCAGGTTTCTGGAGATGGAGCCAGTGACCATCCCAGACGTGCACGGCG
GCTCCCTGCAGAATGCCGTGCGCGTGTGGTCTAACATCCCTGCCATCAGAAGCAGGCACTGG
GCACTGGTGAGCGAGGAGGAGCTGTCCCTGCTGGCCCAGAATAAGCAGAGCAGCAAGCTGGC
CGCCAAGTGGCCTACAAAGCTGGTGAAGAACTGCTTCCTGCCACTGCGGGAGTACTTCAAGT
ATTTTTCCACCGAGCTGACATCTAGCCTGGGAGGACCCTCCTCTGGCGCCCCACCACCTAGC
GGCGGCTCCCCTGCCGGCTCTCCAACCAGCACAGAGGAGGGCACCAGCGAGTCCGCCACACC
AGAGTCTGGACCTGGCACCAGCACAGAGCCATCCGAGGGCTCTGCCCCGGGCTCTCCTGCAG
GCAGCCCTACCTCCACCGAAGAGGGCACCAGCACAGAGCCTTCTGAGGGCAGCGCCCCAGGC
ACCTCTACAGAGCCAAGCGAGCTCGAGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCAC
CAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGG
TGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGAC
AGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACG
GAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACA
GCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCAC
CCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCA
CCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCC
TGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAAC
AGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAA
CCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCA
GACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGCAAC
CTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGA
TGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCCAGA
TCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTG
AGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAA
GAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGC
CTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGAC
GGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGG
CACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCG
ACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAG
GAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCG
CATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAA
AGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCC
CAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCC
CAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACG
TGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGAC
CTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAA
AATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGG
GCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAAC
GAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGA

[0483]

	GGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGC GGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAG TCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGCA GCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCC AGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGC ATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGA GAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCC GCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAA CACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGG GCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACG CCATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGC GACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAA CTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCA AGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTG GAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTA CGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGT CCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCACGCC CACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGA AAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCG AGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTC AAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGG CGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGA GCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAG TCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAA GAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGG AAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAA AGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAA AAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGA GAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATAT GTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGA GCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCA GCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTAC AACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCT GACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGT ACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTAC GAGACACGGATCGACCTGTCTCAGCTGGGAGGCGACAGCCCCAAGAAGAAGAGAAAGGTGGG AGTCGACGGATCCAGCGGCTCCGAGACCCCAGGCACATCTGAGAGCGCCACCCCTGAGTCCA CCGGTAACAAAAAGCTTGAGGCCGTCGGAACCGGAATCGAACCAAAAGCAATGTCCCAGGGT TTGGTGACATTTGGCGACGTGGCTGTCGATTTTTCCCAGGAAGAGTGGGAGTGGCTCAATCC TATCCAGAGGAACTTGTACCGGAAGGTGATGCTGGAGAATTATAGAAATTTGGCATCACTGG GGTTGTGCGTTAGCAAACCAGATGTTATATCTTCCCTGGAACAGGGAAAGGAGCCCTGGAGC GCTGATTACAAAGATGATGACGATAAAGCCCCCAAGAAGAAAAGGAAGGTCCCAAAGAAAAA AAGAAAGGTGTGA
融合蛋白 11 氨基酸序列 NLS-NLS-3A-h3L-dCas9-ZIM3-NLS-NLS (SEQ ID NO: 1518)	MGTMPKKKRKVPKKKRKVYNHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLG IQVDRYIASEVCEDSITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIV NPARKGLYEGTGRLFFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPV MIDAKEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSI KQGKDQHFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLF APLKEYFACVSSGNSNANSRGPSFSSGLVPLSLRGSHMAAIPALDPEAEPSMDVILVGSSEL SSSVSPGTGRDLIAYEVKANQRNIEDICICCGSLQVHTQHPLFEGGICAPCKDKFLDALFLY DDDGYQSYCSICCSGETLLICGNPDCTRCYCFECVDSLVGPGTSGKVHAMSNWVCYLCLPSS RSGLLQRRRKWRSQLKAFYDRESENPLEMFETVPVWRRQPVRVLSLFEDIKKELTSLGFLES GSDPGQLKHVVDVTDTVRKDVEEWGPFDLVYGATPPLGHTCDRPPSWYLFQFHRLLQYARPK PGSPRPFFWMFVDNLVLNKEDLDVASRFLEMEPVTIPDVHGGSLQNAVRVWSNIPAIRSRHW ALVSEEELSLLAQNKQSSKLAAKWPTKLVKNCFLPLREYFKYFSTELTSSLGGPSSGAPPPS GGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPG TSTEPSELEDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFD SGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGN LIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILL SDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYID

	GGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASA QSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEEN EDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQ SGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKG ILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKE HPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLV ETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHA HDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFF KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKE SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIME RSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKY VNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAY NKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLY ETRIDLSQLGGDSPKKKRKVGVDGSSGSETPGTSESATPESTGMNNSQGRVTFEDVTVNFTQ GEWQRLNPEQRNLYRDVMLENYSNLVSVGQGETTKPDVILRLEQGKEPWLEEEEVLGSGRAE KNGDIGGQIWKPKDVKESLSADYKDDDDKAPKKKRKVPKKKRKV
融合蛋白 11 DNA 序列 (SEQ ID NO: 1519)	ATGGGTACCATGCCAAAAAAGAAGAGAAAGGTACCGAAGAAAAAAAGAAAGGTATACAATCA CGATCAGGAGTTCGACCCCCCTAAGGTGTACCCACCAGTGCCTGCAGAGAAGAGGAAGCCAA TCCGGGTGCTGAGCCTGTTTGATGGCATCGCCACCGGCCTGCTGGTGCTGAAGGATCTGGGC ATCCAGGTGGACCGGTACATCGCCTCCGAGGTGTGCGAGGATTCTATCACCGTGGGCATGGT GCGCCACCAGGGCAAGATCATGTATGTGGGCGACGTGCGGTCCGTGACACAGAAGCACATCC AGGAGTGGGGCCCATTCGATCTGGTGATCGGCGGCAGCCCCTGTAATGACCTGTCCATCGTG AACCCTGCAAGGAAGGGACTGTACGAGGGAACCGGCCGGCTGTTCTTTGAGTTTTATAGACT GCTGCACGACGCCAGGCCTAAGGAGGGCGACGATAGACCATTCTTTTGGCTGTTCGAGAATG TGGTGGCTATGGGCGTGAGCGATAAGAGGGACATCTCCAGGTTTCTGGAGTCTAACCCCGTG ATGATCGATGCAAAGGAGGTGTCCGCCGCACACAGAGCCAGGTATTTCTGGGGCAATCTGCC AGGAATGAACAGGCCACTGGCAAGCACCGTGAATGACAAGCTGGAGCTGCAGGAGTGCCTGG AGCACGGAAGGATCGCCAAGTTTTCCAAGGTGCGCACAATCACCACACGGAGCAATTCCATC AAGCAGGGCAAGGATCAGCACTTCCCCGTGTTCATGAACGAGAAGGAGGACATCCTGTGGTG TACCGAGATGGAGAGAGTGTTCGGCTTTCCAGTGCACTACACAGACGTGTCTAACATGAGCA GGCTGGCAAGGCAGCGGCTGCTGGGCAGATCTTGGAGCGTGCCCGTGATCAGGCACCTGTTC GCCCCTCTGAAGGAGTATTTTGCCTGCGTGAGCAGCGGCAACTCCAATGCCAACAGCCGGGG CCCCTCTTTCAGCTCCGGATTGGTGCCTCTGAGCCTGAGGGGCTCCCACATGGCAGCAATCC CCGCCCTGGACCCCGAGGCCGAGCCTAGCATGGACGTGATCCTGGTGGGCTCTAGCGAGCTG TCCTCTAGCGTGTCTCCAGGAACCGGAAGGGATCTGATCGCATACGAGGTGAAGGCCAATCA GCGGAACATCGAGGACATCTGTATCTGCTGTGGCAGCCTGCAGGTGCACACACAGCACCCAC TGTTCGAGGGAGGAATCTGCGCACCCTGTAAGGATAAGTTCCTGGACGCCCTGTTTCTGTAC GACGATGACGGCTACCAGTCCTATTGCTCTATCTGCTGTTCCGGCGAGACCCTGCTGATCTG CGGCAATCCAGATTGTACAAGGTGCTATTGTTTTGAGTGCGTGGACTCTCTGGTGGGACCAG GCACCAGCGGAAAGGTGCACGCCATGTCCAACTGGGTGTGCTACCTGTGCCTGCCATCCTCT CGCAGCGGACTGCTGCAGCGGAGAAGGAAGTGGAGATCCCAGCTGAAGGCCTTCTATGATAG GGAGTCTGAGAACCCCCTGGAGATGTTTGAGACCGTGCCAGTGTGGCGCCGGCAGCCCGTGA GGGTGCTGAGCCTGTTCGAGGATATCAAGAAGGAGCTGACATCCCTGGGCTTTCTGGAGTCC GGCTCTGACCCCGGACAGCTGAAGCACGTGGTGGATGTGACCGACACAGTGCGGAAGGATGT GGAGGAGTGGGGCCCTTTCGACCTGGTGTACGGAGCAACCCCTCCACTGGGACACACATGCG ACAGACCCCCTTCTTGGTACCTGTTCCAGTTTCACCGCCTGCTGCAGTATGCAAGGCCAAAG CCAGGCAGCCCTAGACCATTCTTTTGGATGTTCGTGGATAATCTGGTGCTGAACAAGGAGGA TCTGGACGTGGCCAGCAGGTTTCTGGAGATGGAGCCAGTGACCATCCCAGACGTGCACGGCG GCTCCCTGCAGAATGCCGTGCGCGTGTGGTCTAACATCCCTGCCATCAGAAGCAGGCACTGG GCACTGGTGAGCGAGGAGGAGCTGTCCCTGCTGGCCCAGAATAAGCAGAGCAGCAAGCTGGC CGCCAAGTGGCCTACAAAGCTGGTGAAGAACTGCTTCCTGCCACTGCGGGAGTACTTCAAGT ATTTTTCCACCGAGCTGACATCTAGCCTGGGAGGACCCTCCTCTGGCGCCCCACCACCTAGC GGCGGCTCCCCTGCCGGCTCTCCAACCAGCACAGAGGAGGGCACCAGCGAGTCCGCCACACC AGAGTCTGGACCTGGCACCAGCACAGAGCCATCCGAGGGCTCTGCCCCAGGCTCTCCTGCAG GCAGCCCTACCTCCACCGAAGAGGGCACCAGCACAGAGCCTTCTGAGGGCAGCGCCCCAGGC ACCTCTACAGAGCCAAGCGAGCTCGAGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCAC CAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGG TGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGAC AGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACG

[0484]

[0485]

GAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACA
GCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCAC
CCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCA
CCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCC
TGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAAC
AGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAA
CCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCA
GACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGCAAC
CTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGA
TGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCCAGA
TCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTG
AGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAA
GAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGC
CTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGAC
GGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGG
CACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCG
ACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAG
GAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCG
CATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAA
AGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCC
CAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCC
CAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACG
TGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGAC
CTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAA
AATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGG
GCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAAC
GAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGA
GGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGC
GGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAG
TCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGCA
GCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCC
AGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGC
ATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGA
GAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCC
GCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAA
CACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGG
GCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACG
CCATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGC
GACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAA
CTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCA
AGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTG
GAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTA
CGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGT
CCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCACGCC
CACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGA
AAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCG
AGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTC
AAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGG
CGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGA
GCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAG
TCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAA
GAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGG
AAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAA
AGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAA
AAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGA
GAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATAT
GTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGA
GCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCA
GCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTAC
AACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCT
GACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGT
ACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTAC

[0486]

	GAGACACGGATCGACCTGTCTCAGCTGGGAGGCGACAGCCCCAAGAAGAAGAGAAAGGTGGG AGTCGACGGATCCAGCGGCTCCGAGACCCCAGGCACATCTGAGAGCGCCACCCCTGAGTCCA CCGGTATGAACAATTCACAGGGGAGAGTGACATTCGAAGACGTGACCGTGAACTTCACCCAG GGAGAATGGCAGCGCTTGAACCCAGAACAAAGGAACCTCTATCGGGACGTGATGCTGGAAAA CTACTCAAATTTGGTGAGCGTTGGGCAGGGTGAGACCACTAAGCCTGACGTGATCCTGAGAT TGGAACAGGGCAAGGAGCCTTGGCTCGAGGAAGAGGAAGTCCTGGGCTCAGGGAGGGCCGAG AAAAACGGTGATATAGGAGGCCAGATATGGAAGCCTAAGGACGTCAAGGAGAGCCTGAGCGC TGATTACAAAGATGATGACGATAAAGCCCCCAAGAAGAAAAGGAAGGTCCCAAAGAAAAAAA GAAAGGTGTGA
融合蛋白 12 氨基酸序列 NLS-NLS-3A-h3L-dCas9-ZN627-NLS-NLS (SEQ ID NO: 1520)	MGTMPKKKRKVPKKKRKVYNHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLG IQVDRYIASEVCEDSITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIV NPARKGLYEGTGRLFFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPV MIDAKEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSI KQGKDQHFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLF APLKEYFACVSSGNSNANSRGPSFSSGLVPLSLRGSHMAAIPALDPEAEPSMDVILVGSSEL SSSVSPGTGRDLIAYEVKANQRNIEDICICCGSLQVHTQHPLFEGGICAPCKDKFLDALFLY DDDGYQSYCSICCSGETLLICGNPDCTRCYCFECVDSLVGPGTSGKVHAMSNWVCYLCLPSS RSGLLQRRRKWRSQLKAFYDRESENPLEMFETVPVWRRQPVRVLSLFEDIKKELTSLGFLES GSDPGQLKHVVDVTDTVRKDVEEWGPFDLVYGATPPLGHTCDRPPSWYLFQFHRLLQYARPK PGSPRPFFWMFVDNLVLNKEDLDVASRFLEMEPVTIPDVHGGSLQNAVRVWSNIPAIRSRHW ALVSEEELSLLAQNKQSSKLAAKWPTKLVKNCFLPLREYFKYFSTELTSSLGGPSSGAPPPS GGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPG TSTEPSELEDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFD SGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGN LIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILL SDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYID GGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASA QSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEEN EDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQ SGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKG ILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKE HPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLV ETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHA HDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFF KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKE SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIME RSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKY VNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAY NKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLY ETRIDLSQLGGDSPKKKRKVGVDGSSGSETPGTSESATPESTGDSVAFEDVAVNFTLEEWAL LDPSQKNLYRDVMRETFRNLASVGKQWEDQNIEDPFKIPRRNISHIPERLCESKEGGQGEES ADYKDDDDKAPKKKRKVPKKKRKV
融合蛋白 12 DNA 序列 (SEQ ID NO: 1521)	ATGGGTACCATGCCAAAAAAGAAGAGAAAGGTACCGAAGAAAAAAAGAAAGGTATACAATCA CGATCAGGAGTTCGACCCCCCTAAGGTGTACCCACCAGTGCCTGCAGAGAAGAGGAAGCCAA TCCGGGTGCTGAGCCTGTTTGATGGCATCGCCACCGGCCTGCTGGTGCTGAAGGATCTGGGC ATCCAGGTGGACCGGTACATCGCCTCCGAGGTGTGCGAGGATTCTATCACCGTGGGCATGGT GCGCCACCAGGGCAAGATCATGTATGTGGGCGACGTGCGGTCCGTGACACAGAAGCACATCC AGGAGTGGGGCCCATTCGATCTGGTGATCGGCGGCAGCCCCTGTAATGACCTGTCCATCGTG AACCCTGCAAGGAAGGGACTGTACGAGGGAACCGGCCGGCTGTTCTTTGAGTTTTATAGACT GCTGCACGACGCCAGGCCTAAGGAGGGCGACGATAGACCATTCTTTTGGCTGTTCGAGAATG TGGTGGCTATGGGCGTGAGCGATAAGAGGGACATCTCCAGGTTTCTGGAGTCTAACCCCGTG ATGATCGATGCAAAGGAGGTGTCCGCCGCACACAGAGCCAGGTATTTCTGGGGCAATCTGCC AGGAATGAACAGGCCACTGGCAAGCACCGTGAATGACAAGCTGGAGCTGCAGGAGTGCCTGG AGCACGGAAGGATCGCCAAGTTTTCCAAGGTGCGCACAATCACCACACGGAGCAATTCCATC AAGCAGGGCAAGGATCAGCACTTCCCCGTGTTCATGAACGAGAAGGAGGACATCCTGTGGTG TACCGAGATGGAGAGAGTGTTCGGCTTTCCAGTGCACTACACAGACGTGTCTAACATGAGCA GGCTGGCAAGGCAGCGGCTGCTGGGCAGATCTTGGAGCGTGCCCGTGATCAGGCACCTGTTC

[0487]

GCCCCTCTGAAGGAGTATTTTGCCTGCGTGAGCAGCGGCAACTCCAATGCCAACAGCCGGGG
CCCCTCTTTCAGCTCCGGATTGGTGCCTCTGAGCCTGAGGGGCTCCCACATGGCAGCAATCC
CCGCCCTGGACCCCGAGGCCGAGCCTAGCATGGACGTGATCCTGGTGGGCTCTAGCGAGCTG
TCCTCTAGCGTGTCTCCAGGAACCGGAAGGGATCTGATCGCATACGAGGTGAAGGCCAATCA
GCGGAACATCGAGGACATCTGTATCTGCTGTGGCAGCCTGCAGGTGCACACACAGCACCCAC
TGTTCGAGGGAGGAATCTGCGCACCCTGTAAGGATAAGTTCCTGGACGCCCTGTTTCTGTAC
GACGATGACGGCTACCAGTCCTATTGCTCTATCTGCTGTTCCGGCGAGACCCTGCTGATCTG
CGGCAATCCAGATTGTACAAGGTGCTATTGTTTTGAGTGCGTGGACTCTCTGGTGGGACCAG
GCACCAGCGGAAAGGTGCACGCCATGTCCAACTGGGTGTGCTACCTGTGCCTGCCATCCTCT
CGCAGCGGACTGCTGCAGCGGAGAAGGAAGTGGAGATCCCAGCTGAAGGCCTTCTATGATAG
GGAGTCTGAGAACCCCCTGGAGATGTTTGAGACCGTGCCAGTGTGGCGCCGGCAGCCCGTGA
GGGTGCTGAGCCTGTTCGAGGATATCAAGAAGGAGCTGACATCCCTGGGCTTTCTGGAGTCC
GGCTCTGACCCCGGACAGCTGAAGCACGTGGTGGATGTGACCGACACAGTGCGGAAGGATGT
GGAGGAGTGGGGCCCTTTCGACCTGGTGTACGGAGCAACCCCTCCACTGGGACACACATGCG
ACAGACCCCCTTCTTGGTACCTGTTCCAGTTTCACCGCCTGCTGCAGTATGCAAGGCCAAAG
CCAGGCAGCCCTAGACCATTCTTTTGGATGTTCGTGGATAATCTGGTGCTGAACAAGGAGGA
TCTGGACGTGGCCAGCAGGTTTCTGGAGATGGAGCCAGTGACCATCCCAGACGTGCACGGCG
GCTCCCTGCAGAATGCCGTGCGCGTGTGGTCTAACATCCCTGCCATCAGAAGCAGGCACTGG
GCACTGGTGAGCGAGGAGGAGCTGTCCCTGCTGGCCCAGAATAAGCAGAGCAGCAAGCTGGC
CGCCAAGTGGCCTACAAAGCTGGTGAAGAACTGCTTCCTGCCACTGCGGGAGTACTTCAAGT
ATTTTTCCACCGAGCTGACATCTAGCCTGGGAGGACCCTCCTCTGGCGCCCCACCACCTAGC
GGCGGCTCCCCTGCCGGCTCTCCAACCAGCACAGAGGAGGGCACCAGCGAGTCCGCCACACC
AGAGTCTGGACCTGGCACCAGCACAGAGCCATCCGAGGGCTCTGCCCCAGGCTCTCCTGCAG
GCAGCCCTACCTCCACCGAAGAGGGCACCAGCACAGAGCCTTCTGAGGGCAGCGCCCCAGGC
ACCTCTACAGAGCCAAGCGAGCTCGAGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCAC
CAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGG
TGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGAC
AGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACG
GAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACA
GCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCAC
CCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCA
CCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCC
TGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAAC
AGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAA
CCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCA
GACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGCAAC
CTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGA
TGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCCAGA
TCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTG
AGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAA
GAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGC
CTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGAC
GGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGG
CACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCG
ACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAG
GAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCG
CATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAA
AGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCC
CAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCC
CAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACG
TGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGAC
CTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAA
AATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGG
GCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAAC
GAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGA
GGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGC
GGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAG
TCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGCA
GCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCC
AGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGC
ATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGA
GAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCC

[0488]

	GCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAA CACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGG GCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACG CCATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGC GACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAA CTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCA AGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTG GAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTA CGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGT CCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCACGCC CACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGA AAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCG AGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTC AAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGG CGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGA GCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAG TCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAA GAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGG AAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAA AGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAA AAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGA GAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATAT GTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGA GCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCA GCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTAC AACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCT GACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGT ACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTAC GAGACACGGATCGACCTGTCTCAGCTGGGAGGCGACAGCCCCAAGAAGAAGAGAAAGGTGGG AGTCGACGGATCCAGCGGCTCCGAGACCCCAGGCACATCTGAGAGCGCCACCCCTGAGTCCA CCGGTGACTCCGTTGCTTTCGAGGACGTGGCCGTGAACTTCACACTTGAGGAATGGGCCTTG CTCGACCCAAGTCAGAAGAATCTGTACAGAGACGTGATGCGGGAGACATTCAGGAATCTCGC CAGTGTCGGAAAGCAGTGGGAAGACCAGAACATCGAAGATCCTTTCAAGATACCACGGCGCA ATATCTCCCACATTCCTGAGAGGCTGTGTGAATCTAAGGAAGGCGGACAAGGTGAGGAAAGC GCTGATTACAAAGATGATGACGATAAAGCCCCCAAGAAGAAAAGGAAGGTCCCAAAGAAAAA AAGAAAGGTGTGA
融合蛋白 13 氨基酸序列 3A-3L-NLS-dCas9-NLS-KOX1KRAB (SEQ ID NO: 1525)	MGTMNHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCEDS ITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRLF FEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSAAHRARY FWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEK EDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLKEYFACVSSGNS NANSRGPSFSSGLVPLSLRGSHMGPMEIYKTVSAWKRQPVRVLSLFRNIDKVLKSLGFLESG SGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPGWYMFQFHRILQYALP RQESQRPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVRGRDYQNAMRVWSNIPGLKSKH APLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNCLLPLREYFKYFSQNSLPLGGPSSGAPPPS GGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPG TSTEPSEPKKKRKVYMDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNL IGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEE DKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIE GDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEK KNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKN LSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGEL HAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEV VDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGE QKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKD FLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLI NGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAG SPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKEL GSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSID NKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAG FIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVRE

[0489]

	INNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFY SNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKEL LGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANL DKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDPKKKRKVSGSETPGTSESATPESTGRTLVTFKDVFVDFTREEW KLLDTAQQIVYRNVMLENYKNLVSLGYQLTKPDVILRLEKGEEP
融合蛋白 13 DNA 序列 (SEQ ID NO: 1526)	ATGGGTACCATGAACCATGACCAGGAATTTGACCCCCCAAAGGTTTACCCACCTGTGCCAGC TGAGAAGAGGAAGCCCATCCGCGTGCTGTCTCTCTTTGATGGGATTGCTACAGGGCTCCTGG TGCTGAAGGACCTGGGCATCCAAGTGGACCGCTACATTGCCTCCGAGGTGTGTGAGGACTCC ATCACGGTGGGCATGGTGCGGCACCAGGGAAAGATCATGTACGTCGGGGACGTCCGCAGCGT CACACAGAAGCATATCCAGGAGTGGGGCCCATTCGACCTGGTGATTGGAGGCAGTCCCTGCA ATGACCTCTCCATTGTCAACCCTGCCCGCAAGGGACTTTATGAGGGTACTGGCCGCCTCTTC TTTGAGTTCTACCGCCTCCTGCATGATGCGCGGCCCAAGGAGGGAGATGATCGCCCCTTCTT CTGGCTCTTTGAGAATGTGGTGGCCATGGGCGTTAGTGACAAGAGGGACATCTCGCGATTTC TTGAGTCTAACCCCGTGATGATTGACGCCAAAGAAGTGTCTGCTGCACACAGGGCCCGTTAC TTCTGGGGTAACCTTCCTGGCATGAACAGGCCTTTGGCATCCACTGTGAATGATAAGCTGGA GCTGCAAGAGTGTCTGGAGCACGGCAGAATAGCCAAGTTCAGCAAAGTGAGGACCATTACCA CCAGGTCAAACTCTATAAAGCAGGGCAAAGACCAGCATTTCCCCGTCTTCATGAACGAGAAG GAGGACATCCTGTGGTGCACTGAAATGGAAAGGGTGTTTGGCTTCCCCGTCCACTACACAGA CGTCTCCAACATGAGCCGCTTGGCGAGGCAGAGACTGCTGGGCCGATCGTGGAGCGTGCCGG TCATCCGCCACCTCTTCGCTCCGCTGAAGGAATATTTTGCTTGTGTGTCTAGCGGCAATAGT AACGCTAACAGCCGCGGGCCGAGCTTCAGCAGCGGCCTGGTGCCGTTAAGCTTGCGCGGCAG CCATATGGGCCCTATGGAGATATACAAGACAGTGTCTGCATGGAAGAGACAGCCAGTGCGGG TACTGAGCCTCTTCAGAAACATCGACAAGGTACTAAAGAGTTTGGGCTTCTTGGAAAGCGGT TCTGGTTCTGGGGGAGGAACGCTGAAGTACGTGGAAGATGTCACAAATGTCGTGAGGAGAGA CGTGGAGAAATGGGGCCCCTTTGACCTGGTGTACGGCTCGACGCAGCCCCTAGGCAGCTCTT GTGATCGCTGTCCCGGCTGGTACATGTTCCAGTTCCACCGGATCCTGCAGTATGCGCTGCCT CGCCAGGAGAGTCAGCGGCCCTTCTTCTGGATATTCATGGACAATCTGCTGCTGACTGAGGA TGACCAAGAGACAACTACCCGCTTCCTTCAGACAGAGGCTGTGACCCTCCAGGATGTCCGTG GCAGAGACTACCAGAATGCTATGCGGGTGTGGAGCAACATTCCAGGGCTGAAGAGCAAGCAT GCGCCCCTGACCCCAAAGGAAGAAGAGTATCTGCAAGCCCAAGTCAGAAGCAGGAGCAAGCT GGACGCCCCGAAAGTTGACCTCCTGGTGAAGAACTGCCTTCTCCCGCTGAGAGAGTACTTCA AGTATTTTTCTCAAAACTCACTTCCTCTTGGAGGGCCGAGCTCTGGCGCACCCCCACCAAGT GGAGGGTCTCCTGCCGGGTCCCCAACATCTACTGAAGAAGGCACCAGCGAATCCGCAACGCC CGAGTCAGGCCCTGGTACCTCCACAGAACCATCTGAAGGTAGTGCGCCTGGTTCCCCAGCTG GAAGCCCTACTTCCACCGAAGAAGGCACGTCAACCGAACCAAGTGAAGGATCTGCCCCTGGG ACCAGCACTGAACCATCTGAGCCAAAAAAGAAGAGAAAGGTATACATGGACAAGAAGTACAG CATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGG TGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTG ATCGGCGCCCTGCTGTTCGACAGCGGAGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGC CAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACG AGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAG GATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGA GAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACC TGCGGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAG GGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTA CAACCAGCTGTTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGT CTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAG AAGAATGGCCTGTTCGGCAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAG CAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACC TGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAAC CTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCC CCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAG CTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAAC GGCTACGCCGGCTACATCGATGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCC CATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGC TGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTG CACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGAT CGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCA GATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTG

[0490]

	GTGGACAAGGGCGCCAGCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCT GCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTACAACG AGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAG CAGAAAAAAGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCT GAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAG ATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGAC TTCCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTT TGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAG TGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATC AACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTT CGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCC AGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGC AGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGT GATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCC AGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTG GGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTA CCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGC TGTCCGACTACGATGTGGACGCTATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGAT AACAAAGTGCTGACTCGGAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGA GGTCGTGAAGAAGATGAAGAACTACTGGCGCCAGCTGCTGAATGCCAAGCTGATTACCCAGA GGAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGC TTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGA CTCCCGGATGAACACTAAGTACGACGAGAACGACAAACTGATCCGGGAAGTGAAAGTGATCA CCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAG ATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGAT CAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGC GGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTAC AGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCG GCCTCTGATCGAGACAAACGGCGAAACAGGCGAGATCGTGTGGGATAAGGGCCGGGACTTTG CCACCGTGCGGAAAGTGCTGTCTATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAG ACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGACAAGCTGATCGCCAG AAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTG TGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTG CTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGC CAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCG AGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAA CTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAA GGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAACACTACCTGG ACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTG GACAAGGTGCTGAGCGCCTACAACAAGCACAGAGACAAGCCTATCAGAGAGCAGGCCGAGAA TATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACA CCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCAC CAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGCGACCCAAA AAAGAAGAGAAAGGTAAGCGGAAGTGAGACCCCAGGTACATCCGAATCAGCAACGCCTGAAA GCACCGGTCGGACACTGGTGACCTTCAAGGATGTATTTGTGGACTTCACCAGGGAGGAGTGG AAGCTGCTGGACACTGCTCAGCAGATCGTGTACAGAAATGTGATGCTGGAGAACTATAAGAA CCTGGTTTCCTTGGGTTATCAGCTTACTAAGCCAGATGTGATCCTCCGGTTGGAGAAGGGAG AAGAGCCCTGA
融合蛋白 14 氨基酸序列 NLS-NLS-3A-3L-dCas9-ZIM3-NLS-NLS (SEQ ID NO: 1527)	MGTMPKKKRKVPKKKRKVYNHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLG IQVDRYIASEVCEDSITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIV NPARKGLYEGTGRLFFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPV MIDAKEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSI KQGKDQHFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLF APLKEYFACVSSGNSNANSRGPSFSSGLVPLSLRGSHMGPMEIYKTVSAWKRQPVRVLSLFR NIDKVLKSLGFLESGSGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPG WYMFQFHRILQYALPRQESQRPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVRGRDYQN AMRVWSNIPGLKSKHAPLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNCLLPLREYFKYFSQN SLPLGGPSSGAPPPSGGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTST EEGTSTEPSEGSAPGTSTEPSEMDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDR HSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKF RGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLI

[0491]

	AQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYAD LFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETIT PWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRK PAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLL KIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRI EEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGL SELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDF QFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKA TAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKK LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRV ILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEV LDATLIHQSITGLYETRIDLSQLGGDSGSETPGTSESATPESTGMNNSQGRVTFEDVTVNFT QGEWQRLNPEQRNLYRDVMLENYSNLVSVGQGETTKPDVILRLEQGKEPWLEEEEVLGSGRA EKNGDIGGQIWKPKDVKESLSADYKDDDDKAPKKKRKVPKKKRKV
融合蛋白 14 DNA 序列 (SEQ ID NO: 1528)	ATGGGTACCATGCCAAAAAAGAAGAGAAAGGTACCGAAGAAAAAAAGAAAGGTATACAACCA TGACCAGGAATTCGACCCCCCAAAGGTTTACCCACCTGTGCCAGCTGAGAAGAGGAAGCCCA TCCGCGTGCTGTCTCTCTTTGATGGGATTGCTACAGGGCTCCTGGTGCTGAAGGACCTGGGC ATCCAAGTGGACCGCTACATTGCCTCCGAGGTGTGTGAGGACTCCATCACGGTGGGCATGGT GCGGCACCAGGGAAAGATCATGTACGTCGGGGACGTCCGCAGCGTCACACAGAAGCATATCC AGGAGTGGGGCCCATTCGACCTGGTGATTGGAGGCAGTCCCTGCAATGACCTCTCCATTGTC AACCCTGCCCGCAAGGGACTTTATGAGGGTACTGGCCGCCTCTTCTTTGAGTTCTACCGCCT CCTGCATGATGCGCGGCCCAAGGAGGGAGATGATCGCCCCTTCTTCTTGGCTCTTTGAGAATG TGGTGGCCATGGGCGTTAGTGACAAGAGGGACATCTCGCGATTTCTTGAGTCTAACCCCGTG ATGATTGACGCCAAAGAAGTGTCTGCTGCACACAGGGCCCGTTACTTCTGGGGTAACCTTCC TGGCATGAACAGGCCTTTGGCATCCACTGTGAATGATAAGCTGGAGCTGCAAGAGTGTCTGG AGCACGGCAGAATAGCCAAGTTCAGCAAAGTGAGGACCATTACCACCAGGTCAAACTCTATA AAGCAGGGCAAAGACCAGCATTTCCCCGTCTTCATGAACGAGAAGGAGGACATCCTGTGGTG CACTGAAATGGAAAGGGTGTTTGGCTTCCCCGTCCACTACACAGACGTCTCCAACATGAGCC GCTTGGCGAGGCAGAGACTGCTGGGCCGATCGTGGAGCGTGCCGGTCATCCGCCACCTCTTC GCTCCGCTGAAGGAATATTTTGCTTGTGTGTCTAGCGGCAATAGTAACGCTAACAGCCGCGG GCCGAGCTTCAGCAGCGGCCTGGTGCCGTTAAGCTTGCGCGGCAGCCATATGGGCCCTATGG AGATATACAAGACAGTGTCTGCATGGAAGAGACAGCCAGTGCGGGTACTGAGCCTCTTCAGA AACATCGACAAGGTACTAAAGAGTTTGGGCTTCTTGGAAAGCGGTTCTGGTTCTGGGGGAGG AACGCTGAAGTACGTGGAAGATGTCACAAATGTCGTGAGGAGAGACGTGGAGAAATGGGGCC CCTTTGACCTGGTGTACGGCTCGACGCAGCCCCTAGGCAGCTCTTGTGATCGCTGTCCCGGC TGGTACATGTTCCAGTTCCACCGGATCCTGCAGTATGCGCTGCCTCGCCAGGAGAGTCAGCG GCCCTTCTTCTGGATATTCATGGACAATCTGCTGCTGACTGAGGATGACCAAGAGACAACTA CCCGCTTCCTTCAGACAGAGGCTGTGACCCTCCAGGATGTCCGTGGCAGAGACTACCAGAAT GCTATGCGGGTGTGGAGCAACATTCCAGGGCTGAAGAGCAAGCATGCGCCCCTGACCCCAAA GGAAGAAGAGTATCTGCAAGCCCAAGTCAGAAGCAGGAGCAAGCTGGACGCCCCGAAAGTTG ACCTCCTGGTGAAGAACTGCCTTCTCCCGCTGAGAGAGTACTTCAAGTATTTTTCTCAAAAC TCACTTCCTCTTGGAGGGCCGAGCTCTGGCGCACCCCCACCAAGTGGAGGGTCTCCTGCCGG GTCCCCAACATCTACTGAAGAAGGCACCAGCGAATCCGCAACGCCCGAGTCAGGCCCTGGTA CCTCCACAGAACCATCTGAAGGTAGTGCGCCTGGTTCCCCAGCTGGAAGCCCTACTTCCACC GAAGAAGGCACGTCAACCGAACCAAGTGAAGGATCTGCCCCTGGGACCAGCACTGAACCATC TGAGATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCG TGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGG CACAGCATCAAGAAGAACCTGATCGGCGCCCTGCTGTTCGACAGCGGAGAAACAGCCGAGGC CACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATC TGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAA GAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGT GGACGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGG ACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTC CGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTT CATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGCCAGCGGCG

[0492]

	TGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATC GCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGCAACCTGATTGCCCTGAGCCTGGG CCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCA AGGACACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGAC CTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAA CACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACC AGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATT TTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATCGATGGCGGAGCCAGCCAGGAAGA GTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGA AGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCAC CAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCT GAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCC CTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACC CCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCCAGCGCCCAGAGCTTCATCGAGCGGAT GACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACG AGTACTTCACCGTGTACAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAG CCCGCCTTCCTGAGCGGCGAGCAGAAAAAAGCCATCGTGGACCTGCTGTTCAAGACCAACCG GAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCG TGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTG AAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATAT CGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATG CCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGC AGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGA TTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCC TGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAG CACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGT GGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGG CCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATC GAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCA GCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACC AGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACGCTATCGTGCCTCAGAGCTTT CTGAAGGACGACTCCATCGATAACAAAGTGCTGACTCGGAGCGACAAGAACCGGGGCAAGAG CGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGCCAGCTGCTGA ATGCCAAGCTGATTACCCAGAGGAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTG AGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAA GCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAACGACAAACTGA TCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTC CAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGC CGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCG ACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCT ACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGC CAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACAGGCGAGATCGTGT GGGATAAGGGCCGGGACTTTGCCACCGTGCGGAAAGTGCTGTCTATGCCCCAAGTGAATATC GTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAA CAGCGACAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACA GCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAA CTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAA TCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGC TGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGC GAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGC CAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGG AACAGCACAAACACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTG ATCCTGGCCGACGCTAATCTGGACAAGGTGCTGAGCGCCTACAACAAGCACAGAGACAAGCC TATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTG CCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTG CTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTC TCAGCTGGGAGGCGACAGCGGAAGTGAGACCCCAGGTACATCCGAATCAGCAACGCCTGAAA GCACCGGTATGAACAATTCACAGGGGAGAGTGACATTCGAAGACGTGACCGTGAACTTCACC CAGGGAGAATGGCAGCGCTTGAACCCAGAACAAAGGAACCTCTATCGGGACGTGATGCTGGA AAACTACTCAAATTTGGTGAGCGTTGGGCAGGGTGAGACCACTAAGCCTGACGTGATCCTGA GATTGGAACAGGGCAAGGAGCCTTGGCTCGAGGAAGAGGAAGTCCTGGGCTCAGGGAGGGCC GAGAAAAACGGTGATATAGGAGGCCAGATATGGAAGCCTAAGGACGTCAAGGAGAGCCTGAG

[0493]

	CGCTGATTACAAAGATGATGACGATAAAGCCCCAAAAAAGAAGAGAAAGGTACCGAAGAAAA AAAGAAAGGTCTGA

[0494] 图14A中的序列可以在下面找到：

[0495]

描述	序列	SEQ ID NO:
gRNA009	ACUGCCUGGCUCACUCCUCC	1270
gRNA003	UGCGGAAACCUUCUAGGGUG	1264
gRNA093	AACCUGAUCCUCCAGUCCGG	1354
gRNA011	AUCGUCCGAUGGGGCUCUGG	1272
gRNA007	GGUGUGGGUGCUUGACGCCU	1268
gRNA077	CAAAUCCUAACUGGGCUGGA	1338
gRNA113	GGGCGCCGCCGUUCAGUUCA	1374
gRNA004	GCGGAAACCUUCUAGGGUGU	1265
gRNA008	GUGUGGGUGCUUGACGCCUG	1269
gRNA012	AGGAUCGUCCGAUGGGGCUC	1273
gRNA111	AGUUCAGGGUCUGAGCCUGG	1372
gRNA005	UCAAGCACCCACACCCUAGA	1266
gRNA013	UCAGAUAGGAUCGUCCGAUG	1274

[0496] 图15中的序列可以在下面找到：

[0497]

ZFoff-15 靶序列	GATGGGGCTCTGGTGGCG (SEQ ID NO: 676)
ZFoff-15 ZFP 氨基酸序列	SRPGERPFQCRICMRNFSRTDTLARHLRTHTGEKPFQCRICMRNFSRKTALNRHLKTHTGSQ KPFQCRICMRNFSRNESLKVHLRTHTGEKPFQCRICMRNFSVKNTLTRHLKTHTGSQKPFQC RICMRNFSRREHLVRHLRTHTGEKPFQCRICMRNFSLTHNLRRHLKTHLRGS (SEQ ID NO: 677)
ZFoff-15 全长氨基酸序列	MGTMYNHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCED SITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRL FFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSAAHRAR YFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNE KEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLKEYFACVSSGN SNANSRGPSFSSGLVPLSLRGSHMGPMEIYKTVSAWKRQPVRVLSLFRNIDKVLKSLGFLES GSGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPGWYMFQFHRILQYAL PRQESQRPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVRGRDYQNAMRVWSNIPGLKSK HAPLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNCLLPLREYFKYFSQNSLPLGGPSSGAPPP SGGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAP GTSTEPSEPKKKRKVSRPGERPFQCRICMRNFSRTDTLARHLRTHTGEKPFQCRICMRNFSR KTALNRHLKTHTGSQKPFQCRICMRNFSRNESLKVHLRTHTGEKPFQCRICMRNFSVKNTLT RHLKTHTGSQKPFQCRICMRNFSRREHLVRHLRTHTGEKPFQCRICMRNFSLTHNLRRHLKT HLRGSPKKKRKVSGSETPGTSESATPESTGRTLVTFKDVFVDFTREEWKLLDTAQQIVYRNV MLENYKNLVSLGYQLTKPDVILRLEKGEEPSADYKDDDDKA (SEQ ID NO: 678)
具有 2xNLS 的 ZFoff-15	MGTMPKKKRKVPKKKRKVYNHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLG IQVDRYIASEVCEDSITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIV NPARKGLYEGTGRLFFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPV MIDAKEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSI KQGKDQHFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLF APLKEYFACVSSGNSNANSRGPSFSSGLVPLSLRGSHMGPMEIYKTVSAWKRQPVRVLSLFR NIDKVLKSLGFLESGSGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPG WYMFQFHRILQYALPRQESQRPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVRGRDYQN AMRVWSNIPGLKSKHAPLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNCLLPLREYFKYFSQN SLPLGGPSSGAPPPSGGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTST EEGTSTEPSEGSAPGTSTEPSESRPGERPFQCRICMRNFSRTDTLARHLRTHTGEKPFQCRI CMRNFSRKTALNRHLKTHTGSQKPFQCRICMRNFSRNESLKVHLRTHTGEKPFQCRICMRNF SVKNTLTRHLKTHTGSQKPFQCRICMRNFSRREHLVRHLRTHTGEKPFQCRICMRNFSLTHN LRRHLKTHLRGSSGSETPGTSESATPESTGRTLVTFKDVFVDFTREEWKLLDTAQQIVYRNV

[0498]

	MLENYKNLVSLGYQLTKPDVILRLEKGEEPSADYKDDDDKAPKKKRKVPKKKRKV (SEQ ID NO: 679)
ZFoff-60 靶序列	GCAGAGGCCGGAGGGGGTC (SEQ ID NO: 740)
ZFoff-60 ZFP氨基酸序列	SRPGERPFQCRICMRNFSDPSVLKRHLRTHTGEKPFQCRICMRNFSRTEHLARHLKTHTGGG GSQKPFQCRICMRNFSQSPHLKRHLRTHTGEKPFQCRICMRNFSDQTTLRRHLKTHTGSQKP FQCRICMRNFSKHSNLTRHTRTHTGEKPFQCRICMRNFSQMETLKRHLRTHLRGS (SEQ ID NO: 680)
ZFoff-60全长氨基酸序列	MGTMYNHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCED SITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRL FFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSAAHRAR YFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNE KEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLKEYFACVSSGN SNANSRGPSFSSGLVPLSLRGSHMGPMEIYKTVSAWKRQPVRVLSLFRNIDKVLKSLGFLES GSGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPGWYMFQFHRILQYAL PRQESQRPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVRGRDYQNAMRVWSNIPGLKSK HAPLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNCLLPLREYFKYFSQNSLPLGGPSSGAPPP SGGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAP GTSTEPSEPKKKRKVSRPGERPFQCRICMRNFSDPSVLKRHLRTHTGEKPFQCRICMRNFSR TEHLARHLKTHTGGGGSQKPFQCRICMRNFSQSPHLKRHLRTHTGEKPFQCRICMRNFSDQT TLRRHLKTHTGSQKPFQCRICMRNFSKHSNLTRHTRTHTGEKPFQCRICMRNFSQMETLKRH LRTHLRGSPKKKRKVSGSETPGTSESATPESTGRTLVTFKDVFVDFTREEWKLLDTAQQIVY RNVMLENYKNLVSLGYQLTKPDVILRLEKGEEPSADYKDDDDKA (SEQ ID NO: 681)
具有 2xNLS 的 ZFoff-60	MGTMPKKKRKVPKKKRKVYNHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLG IQVDRYIASEVCEDSITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIV NPARKGLYEGTGRLFFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPV MIDAKEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSI KQGKDQHFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLF APLKEYFACVSSGNSNANSRGPSFSSGLVPLSLRGSHMGPMEIYKTVSAWKRQPVRVLSLFR NIDKVLKSLGFLESGSGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPG WYMFQFHRILQYALPRQESQRPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVRGRDYQN AMRVWSNIPGLKSKHAPLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNCLLPLREYFKYFSQN SLPLGGPSSGAPPPSGGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTST EEGTSTEPSEGSAPGTSTEPSESRPGERPFQCRICMRNFSDPSVLKRHLRTHTGEKPFQCRI CMRNFSRTEHLARHLKTHTGGGGSQKPFQCRICMRNFSQSPHLKRHLRTHTGEKPFQCRICM RNFSDQTTLRRHLKTHTGSQKPFQCRICMRNFSKHSNLTRHTRTHTGEKPFQCRICMRNFSQ METLKRHLRTHLRGSSGSETPGTSESATPESTGRTLVTFKDVFVDFTREEWKLLDTAQQIVY RNVMLENYKNLVSLGYQLTKPDVILRLEKGEEPSADYKDDDDKAPKKKRKVPKKKRKV (SEQ ID NO: 682)
ZFoff-137 靶序列	GGTGGGGAGGACTGTGCA (SEQ ID NO: 683)
ZFoff-137 ZFP氨基酸序列	SRPGERPFQCRICMRNFSQSTTLKRHLRTHTGEKPFQCRICMRNFSGAHGLAGHLKTHTGSQ KPFQCRICMRNFSDQTNLRRHLRTHTGEKPFQCRICMRNFSRQDNLQRHLKTHTGSQKPFQC RICMRNFSKGDHLRRHTRTHTGEKPFQCRICMRNFSEAHHLSRHLRTHLRGS (SEQ ID NO: 684)
ZFoff-137全长氨基酸序列	MGTMYNHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCED SITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRL FFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSAAHRAR YFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNE KEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLKEYFACVSSGN SNANSRGPSFSSGLVPLSLRGSHMGPMEIYKTVSAWKRQPVRVLSLFRNIDKVLKSLGFLES GSGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPGWYMFQFHRILQYAL PRQESQRPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVRGRDYQNAMRVWSNIPGLKSK HAPLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNCLLPLREYFKYFSQNSLPLGGPSSGAPPP SGGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAP GTSTEPSEPKKKRKVSRPGERPFQCRICMRNFSQSTTLKRHLRTHTGEKPFQCRICMRNFSG AHGLAGHLKTHTGSQKPFQCRICMRNFSDQTNLRRHLRTHTGEKPFQCRICMRNFSRQDNLQ RHLKTHTGSQKPFQCRICMRNFSKGDHLRRHTRTHTGEKPFQCRICMRNFSEAHHLSRHLRT HLRGSPKKKRKVSGSETPGTSESATPESTGRTLVTFKDVFVDFTREEWKLLDTAQQIVYRNV MLENYKNLVSLGYQLTKPDVILRLEKGEEPSADYKDDDDKA (SEQ ID NO: 685)

[0499]

具有 2xNLS 的 ZFoff-137	MGTMPKKKRKVPKKKRKVYNHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLG IQVDRYIASEVCEDSITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIV NPARKGLYEGTGRLFFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPV MIDAKEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSI KQGKDQHFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLF APLKEYFACVSSGNSNANSRGPSFSSGLVPLSLRGSHMGPMEIYKTVSAWKRQPVRVLSLFR NIDKVLKSLGFLESGSGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPG WYMFQFHRILQYALPRQESQRPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVRGRDYQN AMRVWSNIPGLKSKHAPLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNCLLPLREYFKYFSQN SLPLGGPSSGAPPPSGGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTST EEGTSTEPSEGSAPGTSTEPSESRPGERPFQCRICMRNFSQSTTLKRHLRTHTGEKPFQCRI CMRNFSGAHGLAGHLKTHTGSQKPFQCRICMRNFSDQTNLRRHLRTHTGEKPFQCRICMRNF SRQDNLQRHLKTHTGSQKPFQCRICMRNFSKGDHLRRHTRTHTGEKPFQCRICMRNFSEAHH LSRHLRTHLRGSSGSETPGTSESATPESTGRTLVTFKDVFVDFTREEWKLLDTAQQIVYRNV MLENYKNLVSLGYQLTKPDVILRLEKGEEPSADYKDDDDKAPKKKRKVPKKKRKV (SEQ ID NO: 686)
ZFoff-152 靶序列	GCAGGAGGACGAGGACGGC (SEQ ID NO: 687)
ZFoff-152 ZFP 氨基酸序列	SRPGERPFQCRICMRNFSTPSKLDRHTRTHTGEKPFQCRICMRNFSLAENLRRHLRTHTGSQ KPFQCRICMRNFSRQDNLGRHLRTHTGEKPFQCRICMRNFSDGGNLGRHLKTHTGGGGSQKP FQCRICMRNFSQSPHLKRHLRTHTGEKPFQCRICMRNFSQSTSLQRHLKTHLRGS (SEQ ID NO: 688)
ZFoff-152 全长氨基酸序列	MGTMYNHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCED SITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRL FFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSAAHRAR YFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNE KEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLKEYFACVSSGN SNANSRGPSFSSGLVPLSLRGSHMGPMEIYKTVSAWKRQPVRVLSLFRNIDKVLKSLGFLES GSGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPGWYMFQFHRILQYAL PRQESQRPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVRGRDYQNAMRVWSNIPGLKSK HAPLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNCLLPLREYFKYFSQNSLPLGGPSSGAPPP SGGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAP GTSTEPSEPKKKRKVSRPGERPFQCRICMRNFSTPSKLDRHTRTHTGEKPFQCRICMRNFSL AENLRRHLRTHTGSQKPFQCRICMRNFSRQDNLGRHLRTHTGEKPFQCRICMRNFSDGGNLG RHLKTHTGGGGSQKPFQCRICMRNFSQSPHLKRHLRTHTGEKPFQCRICMRNFSQSTSLQRH LKTHLRGSPKKKRKVSGSETPGTSESATPESTGRTLVTFKDVFVDFTREEWKLLDTAQQIVY RNVMLENYKNLVSLGYQLTKPDVILRLEKGEEPSADYKDDDDKA (SEQ ID NO: 689)
具有 2xNLS 的 ZFoff-152	MGTMPKKKRKVPKKKRKVYNHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLG IQVDRYIASEVCEDSITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIV NPARKGLYEGTGRLFFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPV MIDAKEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSI KQGKDQHFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLF APLKEYFACVSSGNSNANSRGPSFSSGLVPLSLRGSHMGPMEIYKTVSAWKRQPVRVLSLFR NIDKVLKSLGFLESGSGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPG WYMFQFHRILQYALPRQESQRPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVRGRDYQN AMRVWSNIPGLKSKHAPLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNCLLPLREYFKYFSQN SLPLGGPSSGAPPPSGGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTST EEGTSTEPSEGSAPGTSTEPSESRPGERPFQCRICMRNFSTPSKLDRHTRTHTGEKPFQCRI CMRNFSLAENLRRHLRTHTGSQKPFQCRICMRNFSRQDNLGRHLRTHTGEKPFQCRICMRNF SDGGNLGRHLKTHTGGGGSQKPFQCRICMRNFSQSPHLKRHLRTHTGEKPFQCRICMRNFSQ STSLQRHLKTHLRGSSGSETPGTSESATPESTGRTLVTFKDVFVDFTREEWKLLDTAQQIVY RNVMLENYKNLVSLGYQLTKPDVILRLEKGEEPSADYKDDDDKAPKKKRKVPKKKRKV (SEQ ID NO: 690)
ZFoff-153 靶序列	GCAGGAGGACGAGGACGGC (SEQ ID NO: 691)
ZFoff-153 ZFP 氨基酸序列	SRPGERPFQCRICMRNFSERAKLIRHLRTHTGEKPFQCRICMRNFSDPSNLRRHLKTHTGSQ KPFQCRICMRNFSRQDNLGRHLRTHTGEKPFQCRICMRNFSDQGNLGRHLKTHTGGGGSQKP FQCRICMRNFSQSAHLKRHLRTHTGEKPFQCRICMRNFSQDVSLVRHLKTHLRGS (SEQ ID NO: 692)
ZFoff-153 全长氨基酸序列	MGTMYNHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCED SITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRL

[0500]

	FFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSAAHRAR YFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNE KEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLKEYFACVSSGN SNANSRGPSFSSGLVPLSLRGSHMGPMEIYKTVSAWKRQPVRVLSLFRNIDKVLKSLGFLES GSGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPGWYMFQFHRILQYAL PRQESQRPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVRGRDYQNAMRVWSNIPGLKSK HAPLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNCLLPLREYFKYFSQNSLPLGGPSSGAPPP SGGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAP GTSTEPSEPKKKRKVSRPGERPFQCRICMRNFSERAKLIRHLRTHTGEKPFQCRICMRNFSD PSNLRRHLKTHTGSQKPFQCRICMRNFSRQDNLGRHLRTHTGEKPFQCRICMRNFSDQGNLG RHLKTHTGGGGSQKPFQCRICMRNFSQSAHLKRHLRTHTGEKPFQCRICMRNFSQDVSLVRH LKTHLRGSPKKKRKVSGSETPGTSESATPESTGRTLVTFKDVFVDFTREEWKLLDTAQQIVY RNVMLENYKNLVSLGYQLTKPDVILRLEKGEEPSADYKDDDDKA (SEQ ID NO:693)
具有 2xNLS 的 ZFoff-153	MGTMPKKKRKVPKKKRKVYNHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLG IQVDRYIASEVCEDSITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIV NPARKGLYEGTGRLFFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPV MIDAKEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSI KQGKDQHFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLF APLKEYFACVSSGNSNANSRGPSFSSGLVPLSLRGSHMGPMEIYKTVSAWKRQPVRVLSLFR NIDKVLKSLGFLESGSGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPG WYMFQFHRILQYALPRQESQRPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVRGRDYQN AMRVWSNIPGLKSKHAPLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNCLLPLREYFKYFSQN SLPLGGPSSGAPPPSGGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTST EEGTSTEPSEGSAPGTSTEPSESRPGERPFQCRICMRNFSERAKLIRHLRTHTGEKPFQCRI CMRNFSDPSNLRRHLKTHTGSQKPFQCRICMRNFSRQDNLGRHLRTHTGEKPFQCRICMRNF SDQGNLGRHLKTHTGGGGSQKPFQCRICMRNFSQSAHLKRHLRTHTGEKPFQCRICMRNFSQ DVSLVRHLKTHLRGSSGSETPGTSESATPESTGRTLVTFKDVFVDFTREEWKLLDTAQQIVY RNVMLENYKNLVSLGYQLTKPDVILRLEKGEEPSADYKDDDDKAPKKKRKVPKKKRKV (SEQ ID NO: 694)

[0501] 细胞培养和转染

[0502] 在含有10％FBS的DMEM中培养HeLa(ATCC-CRM-CCL-2)、Hepa1-6(PCSK9-IRES-TdTomato)、Huh7(Sekisui XenoTech,LLC)和HEK293TGriptite(CLTA-GFP)细胞。使用化学合成的指导RNA和体外转录的效应子构建体在HeLa和Huh7细胞中进行所有实验。使用Mirus的TransIT-X2转染试剂(目录号MIR6003)反向转染HeLa细胞。使用Invitrogen的MessengerMAX试剂(目录号LMRNA003)反向转染Huh7细胞。使用Biolegend的LEGEND MAX™人PCSK9 ELISA试剂盒(目录号443107)在指定时间点测量分泌的PCSK9水平。使用Promega的CellTiter-Glo试剂盒(目录号G7571)将所有ELISA数据标准化为细胞数量。

[0503] 使用Mirus的TransIT-X2转染试剂(Cat#MIR6003),将GFP敲入CLTA基因座作为框内CLTA融合物的HEK293T Griptite细胞与编码效应子构建体和人CLTA指导RNA的质粒共转染。通过FACS测量GFP表达,以GFP表达作为CLTA表达的替代。

[0504] 使用Lonza的Amaxa 4D核转染装置中的SF细胞系96孔核转染试剂盒(目录号V4SC-2096,程序代码:CM-138),将编码效应子构建体和小鼠PCSK9引导RNA的质粒共转染到Hepa1-6细胞中。在指定的时间点,对细胞进行FACS分析TdTomato表达,以TdTomato表达作为PCSK9水平的替代。

[0505] 效应子构建体和合成gRNA的体外转录

[0506] 根据制造商的说明,通过CellScript的T7 mScript™标准mRNA生产系统(目录号C-MSC100625)用1μg线性化效应子模板建立体外转录反应,以获得在5'末端具有Cap 1结构并被3'多腺苷酸化的RNA。从Integrated DNA Technologies获得末端修饰的sgRNA,其在5'和3'末端两者上具有三个具有硫代磷酸酯键的2'O-甲基修饰的核苷酸。

[0507] 甲基化谱

[0508] 使用DNAdvance DNA提取自组织试剂盒(Beckman Coulter)从96孔培养板的每个孔中提取基因组DNA。在通过高灵敏度DNA 1X试剂盒(Quant-IT)定量基因组DNA后,根据制

造商的说明,使用EZ-96DNA甲基化-Gold MagPrep试剂盒(Zymo Research)将每个基因组DNA样品亚硫酸氢盐转化。对于杂交捕获实验,使用xGen™甲基-Seq DNA文库制备试剂盒(IDT)制备DNA文库,并使用xGen™DNA文库杂交捕获试剂盒(IDT)进行杂交捕获。对于扩增子测序实验,使用xGen™甲基-Seq DNA文库制备试剂盒(IDT)制备DNA文库,并使用xGen™DNA文库杂交捕获(IDT)进行杂交捕获。使用铂Taq试剂盒(Invitrogen),将来自每个样品的亚硫酸氢盐转化的DNA用于接种对应于两个VIM扩增子中的每一个的PCR。在通过商业的服务(Azenta)测序之前,使用AMPure XP试剂盒(Beckman Coulter)清洁合并的产物,并通过Tapestation 4200(Agilent)上的D1000筛选胶带评估片段大小。

[0509] 实施例13:细菌DNA甲基转移酶

[0510] 在该实验中,筛选了一组细菌蛋白在哺乳动物细胞中的DNA甲基转移酶活性。通过使用实施例1的实验程序将这些细菌DNA甲基转移酶(表12)N-末端融合至dCas9结构域来测试表观遗传沉默活性。然后将这些构建体转染到在CTLA4的哺乳动物启动子控制下表达GFP的报告细胞系中。

[0511] 表12:细菌DNA甲基转移酶

描述	序列
M.AluI 甲基转移酶	MSKANAKYSFVDLFAGIGGFHAALAATGGVCEYAVEIDREAAAVYERNWNKPALGDITDDAN DEGVTLRGYDGPIDVLTGGFPCQPFSKSGAQHGMAETRGTLFWNIARIIEEREPTVLILENV RNLVGPRHRHEWLTIIETLRFFGYEVSGAPAIFSPHLLPAWMGGTPQVRERVFITATLVPER MRDERSTIRRATGRPLEGFPYWADSWTDFRELSRLVVIRGFQAPEREVVGDRKRYVARTDMP EGFVPASVTRPAIDETLPAWKQSHLRRNYDFFERHFAEVVAWAYRWGVYTDLFPASRRKLEW QAQDAPRLWDTVMHFRPSGIRAKRPTYLPALVAITQTSIVGPLERRLSPRETARLQGLPEWF DFGEQRAAATYKQMGNGVNVGVVRHILREHVRRDRALLKLTPAGQRIINAVLADEPDATVGA LGAAE (SEQ ID NO: 675)
M.MspI 甲基转移酶	MKPEILKLIRSKLDLTQKQASEIIEVSDKTWQQWESGKTEMHPAYYSFLQEKLKDKINFEEL SAQKTLQKKIFDKYNQNQITKNAEELAEITHIEERKDAYSSDFKFIDLFSGIGGIRQSFEVN GGKCVFSSEIDPFAKFTYYTNFGVVPFGDITKVEATTIPQHDILCAGFPCQPFSHIGKREGF EHPTQGTMFHEIVRIIETKKTPVLFLENVPGLINHDDGNTLKVIIETLEDMGYKVHHTVLDA SHFGIPQKRKRFYLVAFLNQNIHFEFPKPPMISKDIGEVLESDVTGYSISEHLQKSYLFKKD DGKPSLIDKNTTGAVKTLVSTYHKIQRLTGTFVKDGETGIRLLTTNECKAIMGFPKDFVIPV SRTQMYRQMGNSVVVPVVTKIAEQISLALKTVNQQSPQENFELELV (SEQ ID NO: 610)
M.HaeIII 甲基转移酶	MNLISLFSGAGGLDLGFQKAGFRIICANEYDKSIWKTYESNHSAKLIKGDISKISSDEFPKC DGIIGGPPCQSWSEGGSLRGIDDPRGKLFYEYIRILKQKKPIFFLAENVKGMMAQRHNKAVQ EFIQEFDNAGYDVHIILLNANDYGVAQDRKRVFYIGFRKELNINYLPPIPHLIKPTFKDVIW DLKDNPIPALDKNKTNGNKCIYPNHEYFIGSYSTIFMSRNRVRQWNEPAFTVQASGRQCQLH PQAPVMLKVSKNLNKFVEGKEHLYRRLTVRECARVQGFPDDFIFHYESLNDGYKMIGNAVPV NLAYEIAKTIKSALEICKGN (SEQ ID NO: 608)
M. HhaI 甲基转移酶	MIEIKDKQLTGLRFIDLFAGLGGFRLALESCGAECVYSNEWDKYAQEVYEMNFGEKPEGDIT QVNEKTIPDHDILCAGFPCQAFSISGKQKGFEDSRGTLFFDIARIVREKKPKVVFMENVKNF ASHDNGNTLEVVKNTMNELDYSFHAKVLNALDYGIPQKRERIYMICFRNDLNIQNFQFPKPF ELNTFVKDLLLPDSEVEHLVIDRKDLVMTNQEIEQTTPKTVRLGIVGKGGQGERIYSTRGIA ITLSAYGGGIFAKTGGYLVNGKTRKLHPRECARVMGYPDSYKVHPSTSQAYKQFGNSVVINV LQYIAYNIGSSLNFKPY(SEQ ID NO: 609)
M. SssI 甲基转移酶	MSKVENKTKKLRVFEAFAGIGAQRKALEKVRKDEYEIVGLAEWYVPAIVMYQAIHNNFHTKL EYKSVSREEMIDYLENKTLSWNSKNPVSNGYWKRKKDDELKIIYNAIKLSEKEGNIFDIRDL YKRTLKNIDLLTYSFPCQDLSQQGIQKGMKRGSGTRSGLLWEIERALDSTEKNDLPKYLLME NVGALLHKKNEEELNQWKQKLESLGYQNSIEVLNAADFGSSQARRRVFMISTLNEFVELPKG DKKPKSIKKVLNKIVSEKDILNNLLKYNLTEFKKTKSNINKASLIGYSKFNSEGYVYDPEFT GPTLTASGANSRIKIKDGSNIRKMNSDETFLYIGFDSQDGKRVNEIEFLTENQKIFVCGNSI SVEVLEAIIDKIGG (SEQ ID NO: 605)

[0512]

[0513] M.SssIDNA甲基转移酶能够有效地甲基化哺乳动物细胞中的DNA(图16),稳定的沉默长达30天。图16中的序列可以在下面找到:

[0514]

描述	序列	SEQ ID NO:
TAR087靶DNA序列	GGCGGAAGTGCTGCCTGATA	667
TAR089靶DNA序列	GCAACACCGCCTAGACCGAC	668

TAR090靶DNA序列	GGAGGTTCAGAAAGCCTAAG	669
TAR091靶DNA序列	GCAGCACTAGAGTCCCCTCA	670
TAR092靶DNA序列	GAACAACTCGTGACTGGGGT	671
TAR093靶DNA序列	GGGAGAGGAGGACAAGCGCC	672
TAR094靶DNA序列	GGACCCACCGACACCACGCC	673
TAR114靶DNA序列	CGCGGGAAAATGCAAGACGA	674

[0515] 还在第29天分析了这些细胞的甲基化谱,证实了靶基因的20%甲基化(图17-18)。

[0516] 使用实施例1的实验程序鉴定预测与M.SssI密切相关的另外三种直系同源DNA甲基转移酶并测试表观遗传沉默活性(表13)。

[0517] 表13:细菌甲基转移酶

[0518]

描述	SEQ ID NO
DNA胞嘧啶甲基转移酶 支原体目细菌	601
DNA胞嘧啶甲基转移酶 海洋支原体	602
DNA (胞嘧啶-5-)-甲基转移酶 中华螺原体	603

[0519] 预测表13的DNA甲基转移酶具有与M.SssI相似或改进的功能。在CRISPR-off的背景下测试序列,代替鼠DNMT3A/DNMT3L,并将它们的功能与M.SssIDNA甲基转移酶在HeLa TdTomato系统中沉默PCSK9基因座中的功能进行比较,以鉴定新的特征和改进的功能。

[0520] 实施例14:替代的KRAB结构域

[0521] 在该实施例中,融合蛋白用替代的KRAB结构域(表14)构建,并且当使用实施例1的实验程序测试时,与CRISPR-off相比显示出改善的活性(图19A-19D)。

[0522] 表14:替代的KRAB结构域

[0523]

描述	序列
ZFP28	MKKLEAVGTGIEPKAMSQGLVTFGDVAVDFSQEEWEWLNPIQRNLYRKVMLENYRNLASLGL CVSKPDVISSLEQGKEPWTVKRKMTRAWCPDLKAVWKIKELPLKKDFCEG (SEQ ID NO: 74)
ZN627	MDSVAFEDVAVNFTLEEWALLDPSQKNLYRDVMRETFRNLASVGKQWEDQNIEDPFKIPRRN ISHIPERLCESKEGGQGEETFSQIPDGILNKKTPGVKPCESSVCGEVGMGPSSLNRHIRDHT GREPNEYQEYGKKSYTRNQCGRALSYHRSFPVRERTHPGGKPYDCKECGETFISLVSIRRHM LTHRGGVPYKCKVCGKAFDYPSLFRIHERSHTGEKPYECKQCGKAFSCSSYIRIHERTHTGD KPYECKQCGKAFSCSKYIRIHERTHTGEKPYECKQCGKAFRCASSVRSHERTHTGEKLFECK ECGKALTCLASVRRHMIKHTGNGPYKCKVCGKAFDFPSSFRIHERTHTGEKPYDCKQCGKAF SCSSSFRKHERIHTGEKPYKCTKCGKAFSRSSYFRIHERTHTGEKPYECKQCGKAFSRSTYF RVHEKIHTGEKPYENPNPNASVVPVLS (SEQ ID NO: 666)
KAP1	MAASAAAASAAAASAASGSPGPGEGSAGGEKRSTAPSAAASASASAAASSPAGGGAEALELL EHCGVCRERLRPEREPRLLPCLHSACSACLGPAAPAAANSSGDGGAAGDGTVVDCPVCKQQC FSKDIVENYFMRDSGSKAATDAQDANQCCTSCEDNAPATSYCVECSEPLCETCVEAHQRVKY TKDHTVRSTGPAKSRDGERTVYCNVHKHEPLVLFCESCDTLTCRDCQLNAHKDHQYQFLEDA VRNQRKLLASLVKRLGDKHATLQKSTKEVRSSIRQVSDVQKRVQVDVKMAILQIMKELNKRG RVLVNDAQKVTEGQQERLERQHWTMTKIQKHQEHILRFASWALESDNNTALLLSKKLIYFQL HRALKMIVDPVEPHGEMKFQWDLNAWTKSAEAFGKIVAERPGTNSTGPAPMAPPRAPGPLSK QGSGSSQPMEVQEGYGFGSGDDPYSSAEPHVSGVKRSRSGEGEVSGLMRKVPRVSLERLDLD LTADSQPPVFKVFPGSTTEDYNLIVIERGAAAAATGQPGTAPAGTPGAPPLAGMAIVKEEET EAAIGAPPTATEGPETKPVLMALAEGPGAEGPRLASPSGSTSSGLEVVAPEGTSAPGGGPGT LDDSATICRVCQKPGDLVMCNQCEFCFHLDCHLPALQDVPGEEWSCSLCHVLPDLKEEDGSL SLDGADSTGVVAKLSPANQRKCERVLLALFCHEPCRPLHQLATDSTFSLDQPGGTLDLTLIR ARLQEKLSPPYSSPQEFAQDVGRMFKQFNKLTEDKADVQSIIGLQRFFETRMNEAFGDTKFS AVLVEPPPMSLPGAGLSSQELSGGPGDGP (SEQ ID NO: 629)

[0524]

MeCP2	MVAGMLGLREEKSEDQDLQGLKDKPLKFKKVKKDKKEEKEGKHEPVQPSAHHSAEPAEAGKA ETSEGSGSAPAVPEASASPKQRRSIIRDRGPMYDDPTLPEGWTRKLKQRKSGRSAGKYDVYL INPQGKAFRSKVELIAYFEKVGDTSLDPNDFDFTVTGRGSPSRREQKPPKKPKSPKAPGTGR GRGRPKGSGTTRPKAATSEGVQVKRVLEKSPGKLLVKMPFQTSPGGKAEGGGATTSTQVMVI KRPGRKRKAEADPQAIPKKRGRKPGSVVAAAAAEAKKKAVKESSIRSVQETVLPIKKRKTRE TVSIEVKEVVKPLLVSTLGEKSGKGLKTCKSPGRKSKESSPKGRSSSASSPPKKEHHHHHHH SESPKAPVPLLPPLPPPPPEPESSEDPTSPPEPQDLSSSVCKEEKMPRGGSLESDGCPKEPA KTQPAVATAATAAEKYKHRGEGERKDIVSSSMPRPNREEPVDSRTPVTERVS (SEQ ID NO: 630)
HP1b	MGKKQNKKKVEEVLEEEEEEYVVEKVLDRRVVKGKVEYLLKWKGFSDEDNTWEPEENLDCPD LIAEFLQSQKTAHETDKSEGGKRKADSDSEDKGEESKPKKKKEESEKPRGFARGLEPERIIG ATDSSGELMFLMKWKNSDEADLVPAKEANVKCPQVVISFYEERLTWHSYPSEDDDKKDDKN (SEQ ID NO: 94)
CBX8	GSGPPSSGGGLYRDMGAQGGRPSLIARIPVARILGDPEEESWSPSLTNLEKVVVTDVTSNFL TVTIKESNTDQGFFKEKR (SEQ ID NO: 387)
CDYL2	ASGDLYEVERIVDKRKNKKGKWEYLIRWKGYGSTEDTWEPEHHLLHCEEFIDEFNGLHMSKD KRIKSGKQSSTSKLLRDS (SEQ ID NO: 365)
TOX	KDPNEPQKPVSAYALFFRDTQAAIKGQNPNATFGEVSKIVASMWDGLGEEQKQVYKKKTEAA KKEYLKQLAAYRASLVSK (SEQ ID NO: 372)
TOX3	KDPNEPQKPVSAYALFFRDTQAAIKGQNPNATFGEVSKIVASMWDSLGEEQKQVYKRKTEAA KKEYLKALAAYRASLVSK (SEQ ID NO: 394)
TOX4	KDPNEPQKPVSAYALFFRDTQAAIKGQNPNATFGEVSKIVASMWDSLGEEQKQVYKRKTEAA KKEYLKALAAYKDNQECQ (SEQ ID NO: 1529)
EED	MSEREVSTAPAGTDMPAAKKQKLSSDENSNPDLSGDENDDAVSIESGTNTERPDTPTNTPNA PGRKSWGKGKWKSKKCKYSFKCVNSLKEDHNQPLFGVQFNWHSKEGDPLVFATVGSNRVTLY ECHSQGEIRLLQSYVDADADENFYTCAWTYDSNTSHPLLAVAGSRGIIRIINPITMQCIKHY VGHGNAINELKFHPRDPNLLLSVSKDHALRLWNIQTDTLVAIFGGVEGHRDEVLSADYDLLG EKIMSCGMDHSLKLWRINSKRMMNAIKESYDYNPNKTNRPFISQKIHFPDFSTRDIHRNYVD CVRWLGDLILSKSCENAIVCWKPGKMEDDIDKIKPSESNVTILGRFDYSQCDIWYMRFSMDF WQKMLALGNQVGKLYVWDLEVEDPHKAKCTTLTHHKCGAAIRQTSFSRDSSILIAVCDDASI WRWDRLR (SEQ ID NO: 563)
RBBP4	VWDLSKIGEEQSPEDAEDGPPELLFIHGGHTAKISDFSWNPNEPWVICSVSEDNIMQVWQMA ENIYNDEDPEGSVDPEGQ (SEQ ID NO: 1530)
RCOR1	MPAMVEKGPEVSGKRRGRNNAAASASAAAASAAASAACASPAATAASGAAASSASAAAASAA AAPNNGQNKSLAAAAPNGNSSSNSWEEGSSGSSSDEEHGGGGMRVGPQYQAVVPDFDPAKLA RRSQERDNLGMLVWSPNQNLSEAKLDEYIAIAKEKHGYNMEQALGMLFWHKHNIEKSLADLP NFTPFPDEWTVEDKVLFEQAFSFHGKTFHRIQQMLPDKSIASLVKFYYSWKKTRTKTSVMDR HARKQKREREESEDELEEANGNNPIDIEVDQNKESKKEVPPTETVPQVKKEKHSTQAKNRAK RKPPKGMFLSQEDVEAVSANATAATTVLRQLDMELVSVKRQIQNIKQTNSALKEKLDGGIEP YRLPEVIQKCNARWTTEEQLLAVQAIRKYGRDFQAISDVIGNKSVVQVKNFFVNYRRRFNID EVLQEWEAEHGKEETNGPSNQKPVKSPDNSIKMPEEEDEAPVLDVRYASAS (SEQ ID NO: 559)
SCML2	KQGFSKDPSTWSVDEVIQFMKHTDPQISGPLADLFRQHEIDGKALFLLKSDVMMKYMGLKLG PALKLCYYIEKLKEGKYS (SEQ ID NO: 382)

[0525] 实施例15:ZIM 3融合构建体

[0526] 产生了ZIM3和KOX1KRAB的新融合物。ZIM 3和KOX1KRAB都是具有广泛同源性的KRAB家族蛋白。因此,设计了代表ZIM3和KOX1KRAB之间中点的序列。这些KOX1KRAB和ZIM3构建体编码集中在蛋白质锌指结构域周围的KOX1KRAB和ZIM3的小区域。虽然KOX1KRAB和ZIM3使用的区域在其序列的前75bp内非常相似,但ZIM3在C末端也具有一个小的α-螺旋区域,KOX1KRAB中不存在。KOX1KRAB-FL序列包括相当于该额外片段的KOX1 KRAB序列,而ZIM3截短从ZIM3序列中去除该额外片段。ZIM3/KOX1KRAB嵌合体是两种蛋白质的N-和C-末端片段的融合体。ZIM3样KOX1KRAB变体都是通过以下来组装:第一,来自非人物种的ZIM3或KOX1KRAB蛋白的BLAST,以组装每个基因最接近的100个同源物(“家族”);第二,鉴定与ZIM3最相似的KOX1KRAB家族的3个成员和与KOX1KRAB最相似的ZIM3家族的3个成员;第三,合理地修饰KOX1KRAB-FL序列以类似于每组的三个(表15)。

[0527] 表15:ZIM-KOX1KRAB嵌合蛋白

描述	序列
KOX1KRAB 残基 11-72	58
KOX1KRAB-FL 残基 11-108	59
ZIM3 残基 1-100	60
KOX1KRAB-ZIM3 嵌合体 1 (KOX1KRAB 的残基 11-72, ZIM3 的残基 68-100)	61
ZIM3-KOX1KRAB 嵌合体 2 (ZIM3 的残基 1-67, KOX1KRAB 的残基 73-108)	62
截短 ZIM3 (ZIM3 的残基 1-75)	63
KOX1KRAB-FL Zim3-样(修饰的残基 11-108)	64
KOX1KRAB 样 ZIM3 1	65
KOX1KRAB 样 ZIM3 2	66
KOX1KRAB 样 ZIM3 3	67
KOX1KRAB 样 ZIM3 4	68
ZIM 样 KOX1KRAB 1	69
ZIM 样 KOX1KRAB 2	70
ZIM 样 KOX1KRAB 3	71

[0528]

[0529] 序列

[0530] 下面列出了本公开中描述的核苷酸(nt)和氨基酸(aa)序列的SEQ ID NO(SEQ)。

[0531]

SEQ	描述	序列
1	酿脓链球菌 WT Cas9 序列 (nt)	ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGC GTCGGATGGGCGGTGATCACTGATGAATATAAGGTTCCGTCTAAA AAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTATCAAAAAA AATCTTATAGGGGCTCTTTTATTTGACAGTGGAGAGACAGCGGAA GCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGG AAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGATG GCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGTCTTTT TTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGA AATATAGTAGATGAAGTTGCTTATCATGAGAAATATCCAACTATC TATCATCTGCGAAAAAAATTGGTAGATTCTACTGATAAAGCGGAT TTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGT GGTCATTTTTTGATTGAGGGAGATTTAAATCCTGATAATAGTGAT GTGGACAAACTATTTATCCAGTTGGTACAAACCTACAATCAATTA TTTGAAGAAAACCCTATTAACGCAAGTGGAGTAGATGCTAAAGCG ATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTC ATTGCTCAGCTCCCCGGTGAGAAGAAAAATGGCTTATTTGGGAAT CTCATTGCTTTGTCATTGGGTTTGACCCCTAATTTTAAATCAAAT TTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGATACT TACGATGATGATTTAGATAATTTATTGGCGCAAATTGGAGATCAA TATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGATGCTATT TTACTTTCAGATATCCTAAGAGTAAATACTGAAATAACTAAGGCT CCCCTATCAGCTTCAATGATTAAACGCTACGATGAACATCATCAA GACTTGACTCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGAA AAGTATAAAGAAATCTTTTTTGATCAATCAAAAAACGGATATGCA GGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAAATTT ATCAAACCAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTG GTGAAACTAAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTT GACAACGGCTCTATTCCCCATCAAATTCACTTGGGTGAGCTGCAT GCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTTAAAAGAC AATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTAT TATGTTGGTCCATTGGCGCGTGGCAATAGTCGTTTTGCATGGATG ACTCGGAAGTCTGAAGAAACAATTACCCCATGGAATTTTGAAGAA GTTGTCGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATG ACAAACTTTGATAAAAATCTTCCAAATGAAAAAGTACTACCAAAA CATAGTTTGCTTTATGAGTATTTTACGGTTTATAACGAATTGACA AAGGTCAAATATGTTACTGAAGGAATGCGAAAACCAGCATTTCTT TCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAACA AATCGAAAAGTAACCGTTAAGCAATTAAAAGAAGATTATTTCAAA AAAATAGAATGTTTTGATAGTGTTGAAATTTCAGGAGTTGAAGAT AGATTTAATGCTTCATTAGGTACCTACCATGATTTGCTAAAAATT ATTAAAGATAAAGATTTTTTGGATAATGAAGAAAATGAAGATATC TTAGAGGATATTGTTTTAACATTGACCTTATTTGAAGATAGGGAG ATGATTGAGGAAAGACTTAAAACATATGCTCACCTCTTTGATGAT AAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGA CGTTTGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCT GGCAAAACAATATTAGATTTTTTGAAATCAGATGGTTTTGCCAAT CGCAATTTTATGCAGCTGATCCATGATGATAGTTTGACATTTAAA GAAGACATTCAAAAAGCACAAGTGTCTGGACAAGGCGATAGTTTA CATGAACATATTGCAAATTTAGCTGGTAGCCCTGCTATTAAAAAA GGTATTTTACAGACTGTAAAAGTTGTTGATGAATTGGTCAAAGTA

[0532]

SEQ	描述	序列
		ATGGGGCGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACGT GAAAATCAGACAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGT ATGAAACGAATCGAAGAAGGTATCAAAGAATTAGGAAGTCAGATT CTTAAAGAGCATCCTGTTGAAAATACTCAATTGCAAAATGAAAAG CTCTATCTCTATTATCTCCAAAATGGAAGAGACATGTATGTGGAC CAAGAATTAGATATTAATCGTTTAAGTGATTATGATGTCGATCAC ATTGTTCCACAAAGTTTCCTTAAAGACGATTCAATAGACAATAAG GTCTTAACGCGTTCTGATAAAAATCGTGGTAAATCGGATAACGTT CCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAA CTTCTAAACGCCAAGTTAATCACTCAACGTAAGTTTGATAATTTA ACGAAAGCTGAACGTGGAGGTTTGAGTGAACTTGATAAAGCTGGT TTTATCAAACGCCAATTGGTTGAAACTCGCCAAATCACTAAGCAT GTGGCACAAATTTTGGATAGTCGCATGAATACTAAATACGATGAA AATGATAAACTTATTCGAGAGGTTAAAGTGATTACCTTAAAATCT AAATTAGTTTCTGACTTCCGAAAAGATTTCCAATTCTATAAAGTA CGTGAGATTAACAATTACCATCATGCCCATGATGCGTATCTAAAT GCCGTCGTTGGAACTGCTTTGATTAAGAAATATCCAAAACTTGAA TCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTCGTAAA ATGATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAAAA TATTTCTTTTACTCTAATATCATGAACTTCTTCAAAACAGAAATT ACACTTGCAAATGGAGAGATTCGCAAACGCCCTCTAATCGAAACT AATGGGGAAACTGGAGAAATTGTCTGGGATAAAGGGCGAGATTTT GCCACAGTGCGCAAAGTATTGTCCATGCCCCAAGTCAATATTGTC AAGAAAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAATT TTACCAAAAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGAC TGGGATCCAAAAAAATATGGTGGTTTTGATAGTCCAACGGTAGCT TATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGAAATCGAAG AAGTTAAAATCCGTTAAAGAGTTACTAGGGATCACAATTATGGAA AGAAGTTCCTTTGAAAAAAATCCGATTGACTTTTTAGAAGCTAAA GGATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACCTAAA TATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGGCT AGTGCCGGAGAATTACAAAAAGGAAATGAGCTGGCTCTGCCAAGC AAATATGTGAATTTTTTATATTTAGCTAGTCATTATGAAAAGTTG AAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTTGTGGAG CAGCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAA TTTTCTAAGCGTGTTATTTTAGCAGATGCCAATTTAGATAAAGTT CTTAGTGCATATAACAAACATAGAGACAAACCAATACGTGAACAA GCAGAAAATATTATTCATTTATTTACGTTGACGAATCTTGGAGCT CCCGCTGCTTTTAAATATTTTGATACAACAATTGATCGTAAACGA TATACGTCTACAAAAGAAGTTTTAGATGCCACTCTTATCCATCAA TCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAGCTA GGAGGTGACTGA
2	酿脓链球菌 WT Cas9 序列(aa)	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKK NLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEM AKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI YHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD VDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENL IAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDT YDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKA PLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED

[0533]

SEQ	描述	序列
		RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRE MIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMAR ENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEK LYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNK VLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDE NDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAK YFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF ATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKD WDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIME RSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLA SAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQ AENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ SITGLYETRIDLSQLGGD
3	SaCas9	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNE GRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYE ARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTK EQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEA KQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKD IKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRD ENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRV TSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQS SEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILD ELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVV KRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQ KRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEA IPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGN RTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEER DINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKS INGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKK LDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKD FKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYD KDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLY KYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNS RNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSK CYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLN RIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDIL GNLYEVKSKKHPQIIKKG
4	新杀弗朗西斯菌 WT Cpf1	MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAK DYKKAKQIIDKYHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSD DDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQE SDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFK GFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKA PEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFN NYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKT LKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIA AFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQ VFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKY LSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQN KDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKL KIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNY

[0534]

SEQ	描述	序列
		ITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYL GVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFS AKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKF IDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLT FENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKA LFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKD NPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEI NLLLKEKANDVHILSIDRGERHLAYYTLVDGKGNIIKQDTFNIIG NDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVH EIAKLVIEYNAIVVFEDLNFGFKRGRFKVEKQVYQKLEKMLIEKL NYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAG FTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFE FSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYP TKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTIL QMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDADANGAY HIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN
5	CasX	MEKRINKIRKKLSADNATKPVSRSGPMKTLLVRVMTDDLKKRLEK RRKKPEVMPQVISNNAANNLRMLLDDYTKMKEAILQVYWQEFKDD HVGLMCKFAQPASKKIDQNKLKPEMDEKGNLTTAGFACSQCGQPL FVYKLEQVSEKGKAYTNYFGRCNVAEHEKLILLAQLKPEKDSDEA VTYSLGKFGQRALDFYSIHVTKESTHPVKPLAQIAGNRYASGPVG KALSDACMGTIASFLSKYQDIIIEHQKVVKGNQKRLESLRELAGK ENLEYPSVTLPPQPHTKEGVDAYNEVIARVRMWVNLNLWQKLKLS RDDAKPLLRLKGFPSFPVVERRENEVDWWNTINEVKKLIDAKRDM GRVFWSGVTAEKRNTILEGYNYLPNENDHKKREGSLENPKKPAKR QFGDLLLYLEKKYAGDWGKVFDEAWERIDKKIAGLTSHIEREEAR NAEDAQSKAVLTDWLRAKASFVLERLKEMDEKEFYACEIQLQKWY GDLRGNPFAVEAENRVVDISGFSIGSDGHSIQYRNLLAWKYLENG KREFYLLMNYGKKGRIRFTDGTDIKKSGKWQGLLYGGGKAKVIDL TFDPDDEQLIILPLAFGTRQGREFIWNDLLSLETGLIKLANGRVI EKTIYNKKIGRDEPALFVALTFERREVVDPSNIKPVNLIGVDRGE NIPAVIALTDPEGCPLPEFKDSSGGPTDILRIGEGYKEKQRAIQA AKEVEQRRAGGYSRKFASKSRNLADDMVRNSARDLFYHAVTHDAV LVFENLSRGFGRQGKRTFMTERQYTKMEDWLTAKLAYEGLTSKTY LSKTLAQYTSKTCSNCGFTITTADYDGMLVRLKKTSDGWATTLNN KELKAEGQITYYNRYKRQTVEKELSAELDRLSEESGNNDISKWTK GRRDEALFLLKKRFSHRPVQEQFVCLDCGHEVHADEQAALNIARS WLFLNSNSTEFKSYKSGKQPFVGAWQAFYKRRLKEVWKPNA
6	CasY	MRKKLFKGYILHNKRLVYTGKAAIRSIKYPLVAPNKTALNNLSEK IIYDYEHLFGPLNVASYARNSNRYSLVDFWIDSLRAGVIWQSKST SLIDLISKLEGSKSPSEKIFEQIDFELKNKLDKEQFKDIILLNTG IRSSSNVRSLRGRFLKCFKEEFRDTEEVIACVDKWSKDLIVEGKS ILVSKQFLYWEEEFGIKIFPHFKDNHDLPKLTFFVEPSLEFSPHL PLANCLERLKKFDISRESLLGLDNNFSAFSNYFNELFNLLSRGEI KKIVTAVLAVSKSWENEPELEKRLHFLSEKAKLLGYPKLTSSWAD YRMIIGGKIKSWHSNYTEQLIKVREDLKKHQIALDKLQEDLKKVV DSSLREQIEAQREALLPLLDTMLKEKDFSDDLELYRFILSDFKSL LNGSYQRYIQTEEERKEDRDVTKKYKDLYSNLRNIPRFFGESKKE QFNKFINKSLPTIDVGLKILEDIRNALETVSVRKPPSITEEYVTK QLEKLSRKYKINAFNSNRFKQITEQVLRKYNNGELPKISEVFYRY PRESHVAIRILPVKISNPRKDISYLLDKYQISPDWKNSNPGEVVD LIEIYKLTLGWLLSCNKDFSMDFSSYDLKLFPEAASLIKNFGSCL SGYYLSKMIFNCITSEIKGMITLYTRDKFVVRYVTQMIGSNQKFP LLCLVGEKQTKNFSRNWGVLIEEKGDLGEEKNQEKCLIFKDKTDF AKAKEVEIFKNNIWRIRTSKYQIQFLNRLFKKTKEWDLMNLVLSE

[0535]

SEQ	描述	序列
		PSLVLEEEWGVSWDKDKLLPLLKKEKSCEERLYYSLPLNLVPATD YKEQSAEIEQRNTYLGLDVGEFGVAYAVVRIVRDRIELLSWGFLK DPALRKIRERVQDMKKKQVMAVFSSSSTAVARVREMAIHSLRNQI HSIALAYKAKIIYEISISNFETGGNRMAKIYRSIKVSDVYRESGA DTLVSEMIWGKKNKQMGNHISSYATSYTCCNCARTPFELVIDNDK EYEKGGDEFIFNVGDEKKVRGFLQKSLLGKTIKGKEVLKSIKEYA RPPIREVLLEGEDVEQLLKRRGNSYIYRCPFCGYKTDADIQAALN IACRGYISDNAKDAVKEGERKLDYILEVRKLWEKNGAVLRSAKFL
7	CasPhi	MADTPTLFTQFLRHHLPGQRFRKDILKQAGRILANKGEDATIAFL RGKSEESPPDFQPPVKCPIIACSRPLTEWPIYQASVAIQGYVYGQ SLAEFEASDPGCSKDGLLGWFDKTGVCTDYFSVQGLNLIFQNARK RYIGVQTKVTNRNEKRHKKLKRINAKRIAEGLPELTSDEPESALD ETGHLIDPPGLNTNIYCYQQVSPKPLALSEVNQLPTAYAGYSTSG DDPIQPMVTKDRLSISKGQPGYIPEHQRALLSQKKHRRMRGYGLK ARALLVIVRIQDDWAVIDLRSLLRNAYWRRIVQTKEPSTITKLLK LVTGDPVLDATRMVATFTYKPGIVQVRSAKCLKNKQGSKLFSERY LNETVSVTSIDLGSNNLVAVATYRLVNGNTPELLQRFTLPSHLVK DFERYKQAHDTLEDSIQKTAVASLPQGQQTEIRMWSMYGFREAQE RVCQELGLADGSIPWNVMTATSTILTDLFLARGGDPKKCMFTSEP KKKKNSKQVLYKIRDRAWAKMYRTLLSKETREAWNKALWGLKRGS PDYARLSKRKEELARRCVNYTISTAEKRAQCGRTIVALEDLNIGF FHGRGKQEPGWVGLFTRKKENRWLMQALHKAFLELAHHRGYHVIE VNPAYTSQTCPVCRHCDPDNRDQHNREAFHCIGCGFRGNADLDVA THNIAMVAITGESLKRARGSVASKTPQPLAAE
8	Cas12f1 (Cas14a)	MIKVYRYEIVKPLDLDWKEFGTILRQLQQETRFALNKATQLAWEW MGFSSDYKDNHGEYPKSKDILGYTNVHGYAYHTIKTKAYRLNSGN LSQTIKRATDRFKAYQKEILRGDMSIPSYKRDIPLDLIKENISVN RMNHGDYIASLSLLSNPAKQEMNVKRKISVIIIVRGAGKTIMDRI LSGEYQVSASQIIHDDRKNKWYLNISYDFEPQTRVLDLNKIMGID LGVAVAVYMAFQHTPARYKLEGGEIENFRRQVESRRISMLRQGKY AGGARGGHGRDKRIKPIEQLRDKIANFRDTTNHRYSRYIVDMAIK EGCGTIQMEDLTNIRDIGSRFLQNWTYYDLQQKIIYKAEEAGIKV IKIDPQYTSQRCSECGNIDSGNRIGQAIFKCRACGYEANADYNAA RNIAIPNIDKIIAESIKSGGS
9	Cas12f2 (Cas14b)	NAMIAQKTIKIKLNPTKEQIIKLNSIIEEYIKVSNFTAKKIAEIQ ESFTDSGLTQGTCSECGKEKTYRKYHLLKKDNKLFCITCYKRKYS QFTLQKVEFQNKTGLRNVAKLPKTYYTNAIRFASDTFSGFDEIIK KKQNRLNSIQNRLNFWKELLYNPSNRNEIKIKVVKYAPKTDTREH PHYYSEAEIKGRIKRLEKQLKKFKMPKYPEFTSETISLQRELYSW KNPDELKISSITDKNESMNYYGKEYLKRYIDLINSQTPQILLEKE NNSFYLCFPITKNIEMPKIDDTFEPVGIDWGITRNIAVVSILDSK TKKPKFVKFYSAGYILGKRKHYKSLRKHFGQKKRQDKINKLGTKE DRFIDSNIHKLAFLIVKEIRNHSNKPIILMENITDNREEAEKSMR QNILLHSVKSRLQNYIAYKALWNNIPTNLVKPEHTSQICNRCGHQ DRENRPKGSKLFKCVKCNYMSNADFNASINIARKFYIGEYEPFYK DNEKMKSGVNSISM
10	Cas12f3 (Cas14c)	MEVQKTVMKTLSLRILRPLYSQEIEKEIKEEEKERRKQAGGTGEL DGGFYKKLEKKHSEMFSFDRLNLLLNQLQREIAKVYNHAISELYI ATIAQGNKSNKHYISSIVYNRAYGYFYNAYIALGICSKVEANFRS NELLTQQSALPTAKSDNFPIVLHKQKGAEGEDGGFRISTEGSDLI FEIPIPFYEYNGENRKEPYKWVKKGGQKPVLKLILSTFRRQRNKG WAKDEGTDAEIRKVTEGKYQVSQIEINRGKKLGEHQKWFANFSIE QPIYERKPNRSIVGGLDVGIRSPLVCAINNSFSRYSVDSNDVFKF SKQVFAFRRRLLSKNSLKRKHGHAAHKLEPITEMTEKNDKFRKKI IERWAKEVTNFFVKNQVGIVQIEDLSTMKDREDHFFNQYLRGFWP

[0536]

SEQ	描述	序列
		YYQMQTLIENKLKEYGIEVKRVQAKYTSQLCSNPNCRYWNNYFNF EYRKVNKFPKFKCEKCNLEISADYNAARNLSTPDIEKFVAKATKG INLPEK
11	C2c8	MKVLEFKIHPTEEQVSKIDQSLAACKLLWNLSIALKEESKQRYYR KKHKFDEFSPEIWGLSYSGHYDEKEFKTLKDKEKKLLIGNPCCKI AYFKKTSNGKEYTPLNSIPIRRFMNAENIDKDAVNYLNRKKLAFY FRENTAKFIGEIETEFKKGFFKSVIKPAYDAAKKGIRGIPRFKGR RDKVETLVNGQPETIKIKSNGVIVSSKIGLLKIRGLDRLQGKAPR MAKITRKATGYYLQLTIETDDTIYKESDKCVGLDMGAVAIFTDDL GRQSEAKRYAKIQKKRLNRLQRQASRQKDNSNNQRKTYAKLARVH EKIARQRKGRNAQLAHKITSEYQSVILEDLNLKNMTAAAKPKERE DGDGYKQNGKKRKSGLNKALLDNAIGQLRTFIENKANERGRKIIR VNPKHTSQTCPNCGNIDKANRVSQSKFKCVSCGYEAHADQNAAAN ILIRGLRDEFLRAIGSLYKFPVSMIGKYPGLAGEFTPDLDANQES IGDAPIENAEHSISKQMKQEGNRTPTQPENGSQSLIFLSAPPQPC GDSHGTNNPKALPNKASKRSSKKPRGAIPENPDQLTIWDLLD
12	dSpCas9	MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKK NLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEM AKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI YHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD VDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENL IAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDT YDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKA PLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRE MIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMAR ENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEK LYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNK VLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDE NDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAK YFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF ATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKD WDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIME RSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLA SAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQ AENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ SITGLYETRIDLSQLGGD
13	dSaCas9	MKRNYILGLAIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNE GRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYE ARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTK EQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEA KQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKD IKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRD ENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRV TSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQS

[0537]

SEQ	描述	序列
		SEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILD ELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVV KRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQ KRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEA IPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEEASKKGN RTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEER DINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKS INGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKK LDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKD FKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYD KDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLY KYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNS RNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSK CYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLN RIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDIL GNLYEVKSKKHPQIIKKG
14	非活性 FnCpf1	MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAK DYKKAKQIIDKYHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSD DDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQE SDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFK GFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKA PEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFN NYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKT LKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIA AFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQ VFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKY LSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQN KDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKL KIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNY ITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYL GVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFS AKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKF IDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLT FENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKA LFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKD NPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEI NLLLKEKANDVHILSIARGERHLAYYTLVDGKGNIIKQDTFNIIG NDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVH EIAKLVIEYNAIVVFEDLNFGFKRGRFKVEKQVYQKLEKMLIEKL NYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAG FTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFE FSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYP TKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTIL QMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDADANGAY HIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN
15	dNmeCas9	MAAFKPNSINYILGLAIGIASVGWAMVEIDEEENPIRLIDLGVRV FERAEVPKTGDSLAMARRLARSVRRLTRRRAHRLLRTRRLLKREG VLQAANFDENGLIKSLPNTPWQLRAAALDRKLTPLEWSAVLLHLI KHRGYLSQRKNEGETADKELGALLKGVAGNAHALQTGDFRTPAEL ALNKFEKESGHIRNQRSDYSHTFSRKDLQAELILLFEKQKEFGNP HVSGGLKEGIETLLMTQRPALSGDAVQKMLGHCTFEPAEPKAAKN TYTAERFIWLTKLNNLRILEQGSERPLTDTERATLMDEPYRKSKL TYAQARKLLGLEDTAFFKGLRYGKDNAEASTLMEMKAYHAISRAL EKEGLKDKKSPLNLSPELQDEIGTAFSLFKTDEDITGRLKDRIQP EILEALLKHISFDKFVQISLKALRRIVPLMEQGKRYDEACAEIYG

[0538]

SEQ	描述	序列
		DHYGKKNTEEKIYLPPIPADEIRNPVVLRALSQARKVINGVVRRY GSPARIHIETAREVGKSFKDRKEIEKRQEENRKDREKAAAKFREY FPNFVGEPKSKDILKLRLYEQQHGKCLYSGKEINLGRLNEKGYVE IDAALPFSRTWDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNSR EWQEFKARVETSRFPRSKKQRILLQKFDEDGFKERNLNDTRYVNR FLCQFVADRMRLTGKGKKRVFASNGQITNLLRGFWGLRKVRAEND RHHALDAVVVACSTVAMQQKITRFVRYKEMNAFDGKTIDKETGEV LHQKTHFPQPWEFFAQEVMIRVFGKPDGKPEFEEADTLEKLRTLL AEKLSSRPEAVHEYVTPLFVSRAPNRKMSGQGHMETVKSAKRLDE GVSVLRVPLTQLKLKDLEKMVNRREREPKLYEALKARLEAHKDDPA KAFAEPFYKYDKAGNRTQQVKAVRVEQVQKTGVWVRNHNGIADNA TMVRVDVFEKGDKYYLVPIYSWQVAKGILPDRAVVQGKDEEDWQL IDDSFNFKFSLHPNDLVEVITKKARMFGYFASCHRGTGNINIRIH DLDHKIGKNGILEGIGVKTALSFQKYQIDELGKEIRPCRLKKRPP VR
16	dCjCas9	MARILAFAIGISSIGWAFSENDELKDCGVRIFTKVENPKTGESLA LPRRLARSARKRLARRKARLNHLKHLIANEFKLNYEDYQSFDESL AKAYKGSLISPYELRFRALNELLSKQDFARVILHIAKRRGYDDIK NSDDKEKGAILKAIKQNEEKLANYQSVGEYLYKEYFQKFKENSKE FTNVRNKKESYERCIAQSFLKDELKLIFKKQREFGFSFSKKFEEE VLSVAFYKRALKDFSHLVGNCSFFTDEKRAPKNSPLAFMFVALTR IINLLNNLKNTEGILYTKDDLNALLNEVLKNGTLTYKQTKKLLGL SDDYEFKGEKGTYFIEFKKYKEFIKALGEHNLSQDDLNEIAKDIT LIKDEIKLKKALAKYDLNQNQIDSLSKLEFKDHLNISFKALKLVT PLMLEGKKYDEACNELNLKVAINEDKKDFLPAFNETYYKDEVTNP VVLRAIKEYRKVLNALLKKYGKVHKINIELAREVGKNHSQRAKIE KEQNENYKAKKDAELECEKLGLKINSKNILKLRLFKEQKEFCAYS GEKIKISDLQDEKMLEIDAIYPYSRSFDDSYMNKVLVFTKQNQEK LNQTPFEAFGNDSAKWQKIEVLAKNLPTKKQKRILDKNYKDKEQK NFKDRNLNDTRYIARLVLNYTKDYLDFLPLSDDENTKLNDTQKGS KVHVEAKSGMLTSALRHTWGFSAKDRNNHLHHAIDAVIIAYANNS IVKAFSDFKKEQESNSAELYAKKISELDYKNKRKFFEPFSGFRQK VLDKIDEIFVSKPERKKPSGALHEETFRKEEEFYQSYGGKEGVLK ALELGKIRKVNGKIVKNGDMFRVDIFKHKKTNKFYAVPIYTMDFA LKVLPNKAVARSKKGEIKDWILMDENYEFCFSLYKDSLILIQTKD MQEPEFVYYNAFTSSTVSLIVSKHDNKFETLSKNQKILFKNANEK EVIAKSIGIQNLKVFEKYIVSALGEVTKAEFRQREDFKK
17	dSt1Cas9	MGSDLVLGLAIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNL VRRTNRQGRRLARRKKHRRVRLNRLFEESGLITDFTKISINLNPY QLRVKGLTDELSNEELFIALKNMVKHRGISYLDDASDDGNSSVGD YAQIVKENSKQLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHR LINVFPTSAYRSEALRILQTQQEFNPQITDEFINRYLEILTGKRK YYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEFRAA KASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAM GPAKLFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTL ETLDIEQMDRETLDKLAYVLTLNTEREGIQEALEHEFADGSFSQK QVDELVQFRKANSSIFGKGWHNFSVKLMMELIPELYETSEEQMTI LTRLGKQKTTSSSNKTKYIDEKLLTEEIYNPVVAKSVRQAIKIVN AAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKANKDEKDAAML KAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTIS IHDLINNSNQFEVDAILPLSITFDDSLANKVLVYATANQEKGQRT PYQALDSMDDAWSFRELKAFVRESKTLSNKKKEYLLTEEDISKFD VRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFT SQLRRHWGIEKTRDTYHHHAVDALIIAASSQLNLWKKQKNTLVSY SEDQLLDIETGELISDDEYKESVFKAPYQHFVDTLKSKEFEDSIL

[0539]

SEQ	描述	序列
		FSYQVDSKFNRKISDATIYATRQAKVGKDKADETYVLGKIKDIYT QDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNKQIN EKGKEVPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLGN HIDITPKDSNNKVVLQSVSPWRADVYFNKTTGKYEILGLKYADLQ FEKGTGTYKISQEKYNDIKKKEGVDSDSEFKFTLYKNDLLLVKDT ETKEQQLFRFLSRTMPKQKHYVELKPYDKQKFEGGEALIKVLGNV ANSGQCKKGLGKSNISIYKVRTDVLGNQHIIKNEGDKPKLDF
18	dSt3Cas9	MTKPYSIGLAIGTNSVGWAVITDNYKVPSKKMKVLGNTSKKYIKK NLLGVLLFDSGITAEGRRLKRTARRRYTRRRNRILYLQEIFSTEM ATLDDAFFQRLDDSFLVPDDKRDSKYPIFGNLVEEKVYHDEFPTI YHLRKYLADSTKKADLRLVYLALAHMIKYRGHFLIEGEFNSKNND IQKNFQDFLDTYNAIFESDLSLENSKQLEEIVKDKISKLEKKDRI LKLFPGEKNSGIFSEFLKLIVGNQADFRKCFNLDEKASLHFSKES YDEDLETLLGYIGDDYSDVFLKAKKLYDAILLSGFLTVTDNETEA PLSSAMIKRYNEHKEDLALLKEYIRNISLKTYNEVFKDDTKNGYA GYIDGKTNQEDFYVYLKNLLAEFEGADYFLEKIDREDFLRKQRTF DNGSIPYQIHLQEMRAILDKQAKFYPFLAKNKERIEKILTFRIPY YVGPLARGNSDFAWSIRKRNEKITPWNFEDVIDKESSAEAFINRM TSFDLYLPEEKVLPKHSLLYETFNVYNELTKVRFIAESMRDYQFL DSKQKKDIVRLYFKDKRKVTDKDIIEYLHAIYGYDGIELKGIEKQ FNSSLSTYHDLLNIINDKEFLDDSSNEAIIEEIIHTLTIFEDREM IKQRLSKFENIFDKSVLKKLSRRHYTGWGKLSAKLINGIRDEKSG NTILDYLIDDGISNRNFMQLIHDDALSFKKKIQKAQIIGDEDKGN IKEVVKSLPGSPAIKKGILQSIKIVDELVKVMGGRKPESIVVEMA RENQYTNQGKSNSQQRLKRLEKSLKELGSKILKENIPAKLSKIDN NALQNDRLYLYYLQNGKDMYTGDDLDIDRLSNYDIDHIIPQAFLK DNSIDNKVLVSSASARGKSDDFPSLEVVKKRKTFWYQLLKSKLIS QRKFDNLTKAERGGLLPEDKAGFIQRQLVETRQITKHVARLLDEK FNNKKDENNRAVRTVKIITLKSTLVSQFRKDFELYKVREINDFHH AHDAYLNAVIASALLKKYPKLEPEFVYGDYPKYNSFRERKSATEK VYFYSNIMNIFKKSISLADGRVIERPLIEVNEETGESVWNKESDL ATVRRVLSYPQVNVVKKVEEQNHGLDRGKPKGLFNANLSSKPKPN SNENLVGAKEYLDPKKYGGYAGISNSFAVLVKGTIEKGAKKKITN VLEFQGISILDRINYRKDKLNFLLEKGYKDIELIIELPKYSLFEL SDGSRRMLASILSTNNKRGEIHKGNQIFLSQKFVKLLYHAKRISN TINENHRKYVENHKKEFEELFYYILEFNENYVGAKKNGKLLNSAF QSWQNHSIDELCSSFIGPTGSERKGLFELTSRGSAADFEFLGVKI PRYRDYTPSSLLKDATLIHQSVTGLYETRIDLAKLGEG
19	dLbCpf1	MSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRAE DYKGVKKLLDRYYLSFINDVLHSIKLKNLNNYISLFRKKTRTEKE NKELENLEINLRKEIAKAFKGNEGYKSLFKKDIIETILPEFLDDK DEIALVNSFNGFTTAFTGFFDNRENMFSEEAKSTSIAFRCINENL TRYISNMDIFEKVDAIFDKHEVQEIKEKILNSDYDVEDFFEGEFF NFVLTQEGIDVYNAIIGGFVTESGEKIKGLNEYINLYNQKTKQKL PKFKPLYKQVLSDRESLSFYGEGYTSDEEVLEVFRNTLNKNSEIF SSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRD KWNAEYDDIHLKKKAVVTEKYEDDRRKSFKKIGSFSLEQLQEYAD ADLSVVEKLKEIIIQKVDEIYKVYGSSEKLFDADFVLEKSLKKND AVVAIMKDLLDSVKSFENYIKAFFGEGKETNRDESFYGDFVLAYD ILLKVDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKET DYRATILRYGSKYYLAIMDKKYAKCLQKIDKDDVNGNYEKINYKL LPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNLN DCHKLIDFFKDSISRYPKWSNAYDFNFSETEKYKDIAGFYREVEE QGYKVSFESASKKEVDKLVEEGKLYMFQIYNKDFSDKSHGTPNLH TMYFKLLFDENNHGQIRLSGGAELFMRRASLKKEELVVHPANSPI

[0540]

SEQ	描述	序列
		ANKNPDNPKKTTTLSYDVYKDKRFSEDQYELHIPIAINKCPKNIF KINTEVRVLLKHDDNPYVIGIARGERNLLYIVVVDGKGNIVEQYS LNEIINNFNGIRIKTDYHSLLDKKEKERFEARQNWTSIENIKELK AGYISQVVHKICELVEKYDAVIALEDLNSGFKNSRVKVEKQVYQK FEKMLIDKLNYMVDKKSNPCATGGALKGYQITNKFESFKSMSTQN GFIFYIPAWLTSKIDPSTGFVNLLKTKYTSIADSKKFISSFDRIM YVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKK NNVFDWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYS SFMALMSLMLQMRNSITGRTDVDFLISPVKNSDGIFYDSRNYEAQ ENAILPKNADANGAYNIARKVLWAIGQFKKAEDEKLDKVKIAISN KEWLEYAQTSVKH
20	非活性 AsCpf1	MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARND HYKELKPIIDRIYKTYADQCLQLVQLDWENLSAAIDSYRKEKTEE TRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKA ELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVF SAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENV KKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAG TEKIKGLNEVLNLAIQKNDETAHIIASLPHRFIPLFKQILSDRNT LSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSID LTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSA KEKVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAAL DQPLPTTLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPE FSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNFQMPTL ASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEK TSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSN NFIEPLEITKEIYDLNNPEKEPKKFQTAYAKKTGDQKGYREALCK WIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYH ISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYW TGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKMLNKK LKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVS HEIIKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHP ETPIIGIARGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLD NREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVV VLENLNFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEK VGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFV DPFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSF QRGLPGFMPAWDIVFEKNETQFDAKGTPFIAGKRIVPVIENHRFT GRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTM VALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPM DADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWLAYIQEL RN
21	非活性 enAsCpf1	MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARND HYKELKPIIDRIYKTYADQCLQLVQLDWENLSAAIDSYRKEKTEE TRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKA ELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYRNRKNVF SAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENV KKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAG TEKIKGLNEVLNLAIQKNDETAHIIASLPHRFIPLFKQILSDRNT LSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSID LTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSA KEKVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAAL DQPLPTTLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPE FSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNFQMPTL ARGWDVNREKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEK TSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSN

[0541]

SEQ	描述	序列
		NFIEPLEITKEIYDLNNPEKEPKKFQTAYAKKTGDQKGYREALCK WIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYH ISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYW TGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKMLNKK LKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVS HEIIKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHP ETPIIGIARGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLD NREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVV VLENLNFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEK VGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFV DPFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSF QRGLPGFMPAWDIVFEKNETQFDAKGTPFIAGKRIVPVIENHRFT GRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTM VALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPM DADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWLAYIQEL RN
22	非活性 HFAsCpf1	MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARND HYKELKPIIDRIYKTYADQCLQLVQLDWENLSAAIDSYRKEKTEE TRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKA ELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYRNRKNVF SAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENV KKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAG TEKIKGLNEVLALAIQKNDETAHIIASLPHRFIPLFKQILSDRNT LSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSID LTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSA KEKVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAAL DQPLPTTLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPE FSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNFQMPTL ARGWDVNREKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEK TSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSN NFIEPLEITKEIYDLNNPEKEPKKFQTAYAKKTGDQKGYREALCK WIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYH ISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYW TGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKMLNKK LKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVS HEIIKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHP ETPIIGIARGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLD NREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVV VLENLNFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEK VGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFV DPFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSF QRGLPGFMPAWDIVFEKNETQFDAKGTPFIAGKRIVPVIENHRFT GRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTM VALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPM DADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWLAYIQEL RN
23	非活性 RVRAsCpf1	MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARND HYKELKPIIDRIYKTYADQCLQLVQLDWENLSAAIDSYRKEKTEE TRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKA ELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVF SAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENV KKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAG TEKIKGLNEVLNLAIQKNDETAHIIASLPHRFIPLFKQILSDRNT LSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSID LTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSA

[0542]

SEQ	描述	序列
		KEKVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAAL DQPLPTTLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPE FSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNFQMPTL ARGWDVNVEKNRGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEK TSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSN NFIEPLEITKEIYDLNNPEKEPKKFQTAYAKKTGDQKGYREALCK WIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYH ISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYW TGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKMLNKK LKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVS HEIIKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHP ETPIIGIARGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLD NREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVV VLENLNFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEK VGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFV DPFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSF QRGLPGFMPAWDIVFEKNETQFDAKGTPFIAGKRIVPVIENHRFT GRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTM VALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPM DADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWLAYIQEL RN
24	非活性 RRAsCpf1	MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARND HYKELKPIIDRIYKTYADQCLQLVQLDWENLSAAIDSYRKEKTEE TRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKA ELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVF SAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENV KKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAG TEKIKGLNEVLNLAIQKNDETAHIIASLPHRFIPLFKQILSDRNT LSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSID LTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSA KEKVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAAL DQPLPTTLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPE FSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNFQMPTL ARGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEK TSEGFDKMYYDYFPDAAKMIPRCSTQLKAVTAHFQTHTTPILLSN NFIEPLEITKEIYDLNNPEKEPKKFQTAYAKKTGDQKGYREALCK WIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYH ISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYW TGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKMLNKK LKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVS HEIIKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHP ETPIIGIARGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLD NREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVV VLENLNFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEK VGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFV DPFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSF QRGLPGFMPAWDIVFEKNETQFDAKGTPFIAGKRIVPVIENHRFT GRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTM VALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPM DADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWLAYIQEL RN
25	dCasX	MEKRINKIRKKLSADNATKPVSRSGPMKTLLVRVMTDDLKKRLEK RRKKPEVMPQVISNNAANNLRMLLDDYTKMKEAILQVYWQEFKDD HVGLMCKFAQPASKKIDQNKLKPEMDEKGNLTTAGFACSQCGQPL FVYKLEQVSEKGKAYTNYFGRCNVAEHEKLILLAQLKPEKDSDEA

[0543]

SEQ	描述	序列
		VTYSLGKFGQRALDFYSIHVTKESTHPVKPLAQIAGNRYASGPVG KALSDACMGTIASFLSKYQDIIIEHQKVVKGNQKRLESLRELAGK ENLEYPSVTLPPQPHTKEGVDAYNEVIARVRMWVNLNLWQKLKLS RDDAKPLLRLKGFPSFPVVERRENEVDWWNTINEVKKLIDAKRDM GRVFWSGVTAEKRNTILEGYNYLPNENDHKKREGSLENPKKPAKR QFGDLLLYLEKKYAGDWGKVFDEAWERIDKKIAGLTSHIEREEAR NAEDAQSKAVLTDWLRAKASFVLERLKEMDEKEFYACEIQLQKWY GDLRGNPFAVEAENRVVDISGFSIGSDGHSIQYRNLLAWKYLENG KREFYLLMNYGKKGRIRFTDGTDIKKSGKWQGLLYGGGKAKVIDL TFDPDDEQLIILPLAFGTRQGREFIWNDLLSLETGLIKLANGRVI EKTIYNKKIGRDEPALFVALTFERREVVDPSNIKPVNLIGVARGE NIPAVIALTDPEGCPLPEFKDSSGGPTDILRIGEGYKEKQRAIQA AKEVEQRRAGGYSRKFASKSRNLADDMVRNSARDLFYHAVTHDAV LVFANLSRGFGRQGKRTFMTERQYTKMEDWLTAKLAYEGLTSKTY LSKTLAQYTSKTCSNCGFTITTADYDGMLVRLKKTSDGWATTLNN KELKAEGQITYYNRYKRQTVEKELSAELDRLSEESGNNDISKWTK GRRDEALFLLKKRFSHRPVQEQFVCLDCGHEVHAAEQAALNIARS WLFLNSNSTEFKSYKSGKQPFVGAWQAFYKRRLKEVWKPNA
26	dCasPhi	MPKPAVESEFSKVLKKHFPGERFRSSYMKRGGKILAAQGEEAVVA YLQGKSEEEPPNFQPPAKCHVVTKSRDFAEWPIMKASEAIQRYIY ALSTTERAACKPGKSSESHAAWFAATGVSNHGYSHVQGLNLIFDH TLGRYDGVLKKVQLRNEKARARLESINASRADEGLPEIKAEEEEV ATNETGHLLQPPGINPSFYVYQTISPQAYRPRDEIVLPPEYAGYV RDPNAPIPLGVVRNRCDIQKGCPGYIPEWQREAGTAISPKTGKAV TVPGLSPKKNKRMRRYWRSEKEKAQDALLVTVRIGTDWVVIDVRG LLRNARWRTIAPKDISLNALLDLFTGDPVIDVRRNIVTFTYTLDA CGTYARKWTLKGKQTKATLDKLTATQTVALVAIALGQTNPISAGI SRVTQENGALQCEPLDRFTLPDDLLKDISAYRIAWDRNEEELRAR SVEALPEAQQAEVRALDGVSKETARTQLCADFGLDPKRLPWDKMS SNTTFISEALLSNSVSRDQVFFTPAPKKGAKKKAPVEVMRKDRTW ARAYKPRLSVEAQKLKNEALWALKRTSPEYLKLSRRKEELCRRSI NYVIEKTRRRTQCQIVIPVIEDLNVRFFHGSGKRLPGWDNFFTAK KENRWFIQGLHKAFSDLRTHRSFYVFEVRPERTSITCPKCGHCEV GNRDGEAFQCLSCGKTCNADLDVATHNLTQVALTGKTMPKREEPR DAQGTAPARKTKKASKSKAPPAEREDQTPAQEPSQTS
27	非活性 VRER SpCas9	MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKK NLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEM AKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI YHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD VDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENL IAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDT YDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKA PLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRE MIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMAR ENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEK LYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNK VLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[0544]

SEQ	描述	序列
		TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDE NDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAK YFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF ATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKD WDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIME RSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLA SARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQ AENIIHLFTLTNLGAPAAFKYFDTTIDRKEYRSTKEVLDATLIHQ SITGLYETRIDLSQLGGD
28	非活性 EQR SpCas9	MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKK NLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEM AKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI YHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD VDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENL IAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDT YDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKA PLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRE MIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMAR ENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEK LYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNK VLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDE NDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAK YFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF ATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKD WDPKKYGGFESPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIME RSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLA SAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQ AENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQ SITGLYETRIDLSQLGGD
29	非活性 VQR SpCas9	MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKK NLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEM AKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI YHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD VDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENL IAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDT YDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKA PLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED

[0545]

SEQ	描述	序列
		RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRE MIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMAR ENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEK LYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNK VLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDE NDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAK YFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF ATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKD WDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIME RSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLA SAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQ AENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQ SITGLYETRIDLSQLGGD
30	非活性 SPG SpCas9	MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKK NLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEM AKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI YHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD VDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENL IAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDT YDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKA PLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRE MIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMAR ENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEK LYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNK VLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDE NDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAK YFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF ATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKD WDPKKYGGFLWPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIME RSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLA SAKQLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQ AENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQ SITGLYETRIDLSQLGGD
31	非活性 SpRY Cas9	MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKK NLIGALLFDSGETAERTRLKRTARRRYTRRKNRICYLQEIFSNEM AKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI YHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD VDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENL IAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDT

[0546]

SEQ	描述	序列
		YDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKA PLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRE MIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMAR ENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEK LYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNK VLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDE NDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAK YFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF ATVRKVLSMPQVNIVKKTEVQTGGFSKESIRPKRNSDKLIARKKD WDPKKYGGFLWPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIME RSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLA SAKQLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQ AENIIHLFTLTRLGAPRAFKYFDTTIDPKQYRSTKEVLDATLIHQ SITGLYETRIDLSQLGGD
32	非活性 KKH dSaCas9	MKRNYILGLAIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNE GRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYE ARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTK EQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEA KQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKD IKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRD ENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRV TSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQS SEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILD ELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVV KRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQ KRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEA IPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEEASKKGN RTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEER DINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKS INGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKK LDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKD FKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIVNNLNGLYD KDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLY KYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNS RNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSK CYEEAKKLKKISNQAEFIASFYKNDLIKINGELYRVIGVNNDLLN RIEVNMIDITYREYLENMNDKRPPHIIKTIASKTQSIKKYSTDIL GNLYEVKSKKHPQIIKKG
33	ZIM3	MNNSQGRVTFEDVTVNFTQGEWQRLNPEQRNLYRDVMLENYSNLV SVGQGETTKPDVILRLEQGKEPWLEEEEVLGSGRAEKNGDIGGQI WKPKDVKESL
34	ZNF436	MAATLLMAGSQAPVTFEDMAMYLTREEWRPLDAAQRDLYRDVMQE NYGNVVSLDFEIRSENEVNPKQEISEDVQFGTTSERPAENAEENP ESEEGFESGDRSERQW

[0547]

SEQ	描述	序列
35	ZNF257	MLENYRNLVFLGIAVSKPDLITCLEQGKEPCNMKRHEMVAKPPVMCSHIAEDLCPERDIKYFFQKVILRRYDKCEHENLQLRKGCKSVDECKVCK
36	ZNF675	MGLLTFRDVAIEFSLEEWQCLDTAQRNLYKNVILENYRNLVFLGIAVSKQDLITCLEQEKEPLTVKRHEMVNEPPVMCSHFAQEFWPEQNIKDSF
37	ZNF490	MLQMQNSEHHGQSIKTQTDSISLEDVAVNFTLEEWALLDPGQRNIYRDVMRATFKNLACIGEKWKDQDIEDEHKNQGRNLRSPMVEALCENKEDCPCGKSTSQIPDLNTNLETPTG
38	ZNF320	MALSQGLLTFRDVAIEFSQEEWKCLDPAQRTLYRDVMLENYRNLVSLDISSKCMMNTLSSTGQGNTEVIHTGTLQRQASYHIGAFCSQEIEKDIHDFVFQ
39	ZNF331	MAQGLVTFADVAIDFSQEEWACLNSAQRDLYWDVMLENYSNLVSLDLESAYENKSLPTKKNIHEIRASKRNSDRRSKSLGRNWICEGTLERPQRSRGR
40	ZNF816	MLREEATKKSKEKEPGMALPQGRLTFRDVAIEFSLEEWKCLNPAQRALYRAVMLENYRNLEFVDSSLKSMMEFSSTRHSITGEVIHTGTLQRHKSHHIGDFCFPEMKKDIHHFEFQWQ
41	ZNF680	MPGPPGSLEMGPLTFRDVAIEFSLEEWQCLDTAQRNLYRKVMFENYRNLVFLGIAVSKPHLITCLEQGKEPWNRKRQEMVAKPPVIYSHFTEDLWPEHSIKDSF
42	ZNF41	MSPPWSPALAAEGRGSSCEASVSFEDVTVDFSKEEWQHLDPAQRRLYWDVTLENYSHLLSVGYQIPKSEAAFKLEQGEGPWMLEGEAPHQSCSGEAIGKMQQQGIPGGIFFHC
43	ZNF189	MASPSPPPESKEEWDYLDPAQRSLYKDVMMENYGNLVSLDVLNRDKDEEPTVKQEIEEIEEEVEPQGVIVTRIKSEIDQDPMGRETFELVGRLDKQRGIFLWEIPRESL
44	ZNF528	MALTQGPLKFMDVAIEFSQEEWKCLDPAQRTLYRDVMLENYRNLVSLGICLPDLSVTSMLEQKRDPWTLQSEEKIANDPDGRECIKGVNTERSSKLGSN
45	ZNF543	MAASAQVSVTFEDVAVTFTQEEWGQLDAAQRTLYQEVMLETCGLLMSLGCPLFKPELIYQLDHRQELWMATKDLSQSSYPGDNTKPKTTEPTFSHLALPE
46	ZNF554	MFSQEERMAAGYLPRWSQELVTFEDVSMDFSQEEWELLEPAQKNLYREVMLENYRNVVSLEALKNQCTDVGIKEGPLSPAQTSQVTSLSSWTGYLLFQPVASSHLEQREALWIEEKGTPQASCSDWMTVLRNQDSTYKKVALQE
47	ZNF140	MSQGSVTFRDVAIDFSQEEWKWLQPAQRDLYRCVMLENYGHLVSLGLSISKPDVVSLLEQGKEPWLGKREVKRDLFSVSESSGEIKDFSPKNVIYDD
48	ZNF610	MEEAQKRKAKESGMALPQGRLTFMDVAIEFSQEEWKSLDPGQRALYRDVMLENYRNLVFLGRSCVLGSNAENKPIKNQLGLTLESHLSELQLFQAGRKIYRSNQVEKFTNHR
49	ZNF264	MAAAVLTDRAQVSVTFDDVAVTFTKEEWGQLDLAQRTLYQEVMLENCGLLVSLGCPVPKAELICHLEHGQEPWTRKEDLSQDTCPGDKGKPKTTEPTTCEPALSE
50	ZNF350	MIQAQESITLEDVAVDFTWEEWQLLGAAQKDLYRDVMLENYSNLVAVGYQASKPDALFKLEQGEQLWTIEDGIHSGACSDIWKVDHVLERLQSESLVNR
51	ZNF8	MEGVAGVMSVGPPAARLQEPVTFRDVAVDFTQEEWGQLDPTQRILYRDVMLETFGHLLSIGPELPKPEVISQLEQGTELWVAERGTTQGCHPAWEPRSESQASRKEEGLPEE
52	ZNF582	MSLGSELFRDVAIVFSQEEWQWLAPAQRDLYRDVMLETYSNLVSLGLAVSKPDVISFLEQGKEPWMVERVVSGGLCPVLESRYDTKELFPKQHVYEV

[0548]

SEQ	描述	序列
53	ZNF30	MAHKYVGLQYHGSVTFEDVAIAFSQQEWESLDSSQRGLYRDVMLE NYRNLVSMAGHSRSKPHVIALLEQWKEPEVTVRKDGRRWCTDLQL EDDTIGCKEMPTSEN
54	ZNF324	MAFEDVAVYFSQEEWGLLDTAQRALYRRVMLDNFALVASLGLSTS RPRVVIQLERGEEPWVPSGTDTTLSRTTYRRRNPGSWSLTEDRDV SG
55	ZNF98	MLENYRNLVFVGIAASKPDLITCLEQGKEPWNVKRHEMVTEPPVV YSYFAQDLWPKQGKKNYFQKVILRTYKKCGRENLQLRKYCKSMDE CKVHKECYNGLNQC
56	ZNF669	MHFRRPDPCREPLASPIQDSVAFEDVAVNFTQEEWALLDSSQKNL YREVMQETCRNLASVGSQWKDQNIEDHFEKPGKDIRNHIVQRLCE SKEDGQYGEVVSQIPNLDLNENISTGLKPCECSICGK
57	ZNF677	MALSQGLFTFKDVAIEFSQEEWECLDPAQRALYRDVMLENYRNLL SLDEDNIPPEDDISVGFTSKGLSPKENNKEELYHLVILERKESHG INNFDLKEVWENMPKFDSLW
58	ZNF596	MTFEDIIVDFTQEEWALLDTSQRKLFQDVMLENISHLVSIGKQLC KSVVLSQLEQVEKLSTQRISLLQGREVGIKHQEIPFIHHIYQKGT STISTMRS
59	ZNF214	MAVTFEDVTIIFTWEEWKFLDSSQKRLYREVMWENYTNVMSVENW NESYKSQEEKFRYLEYENFSYWQGWWNAGAQMYENQNYGETVQGT DSKDLTQQDRSQC
60	ZNF37A	MITSQGSVSFRDVTVGFTQEEWQHLDPAQRTLYRDVMLENYSHLV SVGYCIPKPEVILKLEKGEEPWILEEKFPSQSHLELINTSRNYSI MKFNEFNKG
61	ZNF34	MFEDVAVYLSREEWGRLGPAQRGLYRDVMLETYGNLVSLGVGPAG PKPGVISQLERGDEPWVLDVQGTSGKEHLRVNSPALGTRTEYKEL TSQETFGEEDPQGSEPVEACDHIS
62	ZNF250	METYGNVVSLGLPGSKPDIISQLERGEDPWVLDRKGAKKSQGLWS DYSDNLKYDHTTACTQQDSLSCPWECETKGESQNTDLSPKPLISE QTVILGKTPLGRIDQENNETKQ
63	ZNF547	MAEMNPAQGHVVFEDVAIYFSQEEWGHLDEAQRLLYRDVMLENLA LLSSLGCCHGAEDEEAPLEPGVSVGVSQVMAPKPCLSTQNTQPCE TCSSLLKDILRL
64	ZNF273	MLDNYRNLVFLGIAVSKPDLITCLEQGKEPCNMKRHAMVAKPPVV CSHFAQDLWPKQGLKDS
65	ZNF354A	MAAGQREARPQVSLTFEDVAVLFTRDEWRKLAPSQRNLYRDVMLE NYRNLVSLGLPFTKPKVISLLQQGEDPWEVEKDGSGVSSLGSKSS HKTTKSTQTQDSSFQ
66	ZFP82	MALRSVMFSDVSIDFSPEEWEYLDLEQKDLYRDVMLENYSNLVSL GCFISKPDVISSLEQGKEPWKVVRKGRRQYPDLETKYETKKLSLE NDIYEIN
67	ZNF224	MTTFKEAMTFKDVAVVFTEEELGLLDLAQRKLYRDVMLENFRNLL SVGHQAFHRDTFHFLREEKIWMMKTAIQREGNSGDKIQTEMETVS EAGTHQEW
68	ZNF33A	MFQVEQKSQESVSFKDVTVGFTQEEWQHLDPSQRALYRDVMLENY SNLVSVGYCVHKPEVIFRLQQGEEPWKQEEEFPSQSFPEVWTADH LKERSQENQSKHL
69	ZNF45	MTKSKEAVTFKDVAVVFSEEELQLLDLAQRKLYRDVMLENFRNVV SVGHQSTPDGLPQLEREEKLWMMKMATQRDNSSGAKNLKEMETLQ EVGLRYLP
70	ZNF175	MSQKPQVLGPEKQDGSCEASVSFEDVTVDFSREEWQQLDPAQRCL YRDVMLELYSHLFAVGYHIPNPEVIFRMLKEKEPRVEEAEVSHQR CQEREFGLEIPQKEISKKASFQ

[0549]

SEQ	描述	序列
71	ZNF595	MELVTFRDVAIEFSPEEWKCLDPAQQNLYRDVMLENYRNLVSLGF VISNPDLVTCLEQIKEPCNLKIHETAAKPPAICSPFSQDLSPVQG IEDSF
72	ZNF184	MSTLLQGGHNLLSSASFQESVTFKDVIVDFTQEEWKQLDPGQRDL FRDVTLENYTHLVSIGLQVSKPDVISQLEQGTEPWIMEPSIPVGT CADWETRLENSVSAPEPDISEE
73	ZNF419	MDPAQVPVAADLLTDHEEGYVTFEDVAVYFSQEEWRLLDDAQRLL YRNVMLENFTLLASLGLASSKTHEITQLESWEEPFMPAWEVVTSA IPRGCWHGAEAEEAPEQIASVG
74	ZFP28-1	MKKLEAVGTGIEPKAMSQGLVTFGDVAVDFSQEEWEWLNPIQRNL YRKVMLENYRNLASLGLCVSKPDVISSLEQGKEPWTVKRKMTRAW CPDLKAVWKIKELPLKKDFCEG
75	ZFP28-2	MSLLGEHWDYDALFETQPGLVTIKNLAVDFRQQLHPAQKNFCKNG IWENNSDLGSAGHCVAKPDLVSLLEQEKEPWMVKRELTGSLFSGQ RSVHETQELFPKQDSYAE
76	ZNF18	MLALAASQPARLEERLIRDRDLGASLLPAAPQEQWRQLDSTQKEQ YWDLILETYGKMVSGAGISHPKSDLTNSIEFGEELAGIYLHVNEK IPRPTCIGDRQENDKENLNLENH
77	ZNF213	MEGRPGETTDTCFVSGVHGPVALGDIPFYFSREEWGTLDPAQRDL FWDIKRENSRNTTLGFGLKGQSEKSLLQEMVPVVPGQTGSDVTVS WSPEEAEAWESENRPRAALGPVVGARRGRPPTRRRQFRDLA
78	ZNF394	MVAVVRALQRALDGTSSQGMVTFEDTAVSLTWEEWERLDPARRDF CRESAQKDSGSTVPPSLESRVENKELIPMQQILEEAEPQGQLQEA FQGKRPLFSKCGSTHEDRVEKQSGDP
79	ZFP1	MNKSQGSVSFTDVTVDFTQEEWEQLDPSQRILYMDVMLENYSNLL SVEVWKADDQMERDHRNPDEQARQFLILKNQTPIEERGDLFGKAL NLNTDFVSLRQVPYKYDLYEKTL
80	ZFP14	MAHGSVTFRDVAIDFSQEEWEFLDPAQRDLYRDVMWENYSNFISL GPSISKPDVITLLDEERKEPGMVVREGTRRYCPDLESRYRTNTLS PEKDIYEIYSFQWDIMER
81	ZNF416	MAAAVLRDSTSVPVTAEAKLMGFTQGCVTFEDVAIYFSQEEWGLL DEAQRLLYRDVMLENFALITALVCWHGMEDEETPEQSVSVEGVPQ VRTPEASPSTQKIQSCDMCVPFLTDILHLTDLPGQELYLTGACAV FHQDQK
82	ZNF557	MLPPTAASQREGHTEGGELVNELLKSWLKGLVTFEDVAVEFTQEE WALLDPAQRTLYRDVMLENCRNLASLGNQVDKPRLISQLEQEDKV MTEERGILSGTCPDVENPFKAKGLTPKLHVFRKEQSRNMKMER
83	ZNF566	MAQESVMFSDVSVDFSQEEWECLNDDQRDLYRDVMLENYSNLVSM GHSISKPNVISYLEQGKEPWLADRELTRGQWPVLESRCETKKLFL KKEIYEIESTQWEIMEK
84	ZNF729	MPGAPGSLEMGPLTFRDVTIEFSLEEWQCLDTVQQNLYRDVMLEN YRNLVFLGMAVFKPDLITCLKQGKEPWNMKRHEMVTKPPVMRSHF TQDLWPDQSTKDSFQEVILRTYAR
85	ZIM2	MAGSQFPDFKHLGTFLVFEELVTFEDVLVDFSPEELSSLSAAQRN LYREVMLENYRNLVSLGHQFSKPDIISRLEEEESYAMETDSRHTV ICQGE
86	ZNF254	MPGPPRSLEMGLLTFRDVAIEFSLEEWQHLDIAQQNLYRNVMLEN YRNLAFLGIAVSKPDLITCLEQGKEPWNMKRHE
87	ZNF764	MAPPLAPLPPRDPNGAGPEWREPGAVSFADVAVYFCREEWGCLRP AQRALYRDVMRETYGHLSALGIGGNKPALISWVEEEAELWGPAAQ DPE
88	ZNF785	MGPPLAPRPAHVPGEAGPRRTRESRPGAVSFADVAVYFSPEEWEC LRPAQRALYRDVMRETFGHLGALGFSVPKPAFISWVEGEVEAWSP EAQDPDGESS

[0550]

SEQ	描述	序列
89	ZNF10 (KOX1)	MDAKSLTAWSRTLVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLE NYKNLVSLGYQLTKPDVILRLEKGEEPWLVEREIHQETHPDSETA FEIKSSVSSRSIFKDKQSCDIKMEGMARNDLWYLSLEEVWKCRDQ LDKYQENPERHLRQVAFTQKKVLTQERVSESGKYGGNCLLPAQLV LREYFHKRDSHTKSLKHDLVLNGHQDSCASNSNECGQTFCQNIHL IQFARTHTGDKSYKCPDNDNSLTHGSSLGISKGIHREKPYECKEC GKFFSWRSNLTRHQLIHTGEKPYECKECGKSFSRSSHLIGHQKTH TGEEPYECKECGKSFSWFSHLVTHQRTHTGDKLYTCNQCGKSFVH SSRLIRHQRTHTGEKPYECPECGKSFRQSTHLILHQRTHVRVRPY ECNECGKSYSQRSHLVVHHRIHTGLKPFECKDCGKCFSRSSHLYS HQRTHTGEKPYECHDCGKSFSQSSALIVHQRIHTGEKPYECCQCG KAFIRKNDLIKHQRIHVGEETYKCNQCGIIFSQNSPFIVHQIAHT GEQFLTCNQCGTALVNTSNLIGYQTNHIRENAY
90	CBX5 (chromoshadow 结构域)	MGKKTKRTADSSSSEDEEEYVVEKVLDRRVVKGQVEYLLKWKGFS EEHNTWEPEKNLDCPELISEFMKKYKKMKEGENNKPREKSESNKR KSNFSNSADDIKSKKKREQSNDIARGFERGLEPEKIIGATDSCGD LMFLMKWKDTDEADLVLAKEANVKCPQIVIAFYEERLTWHAYPED AENKEKETAKS
91	RYBP (PRC1 的 YAF2_RYBP 组分)	MTMGDKKSPTRPKRQAKPAADEGFWDCSVCTFRNSAEAFKCSICD VRKGTSTRKPRINSQLVAQQVAQQYATPPPPKKEKKEKVEKQDKE KPEKDKEISPSVTKKNTNKKTKPKSDILKDPPSEANSIQSANATT KTSETNHTSRPRLKNVDRSTAQQLAVTVGNVTVIITDFKEKTRSS STSSSTVTSSAGSEQQNQSSSGSESTDKGSSRSSTPKGDMSAVND ESF
92	YAF2 (PRC1 的 YAF2_RYBP 组分)	MGDKKSPTRPKRQPKPSSDEGYWDCSVCTFRNSAEAFKCMMCDVR KGTSTRKPRPVSQLVAQQVTQQFVPPTQSKKEKKDKVEKEKSEKE TTSKKNSHKKTRPRLKNVDRSSAQHLEVTVGDLTVIITDFKEKTK SPPASSAASADQHSQSGSSSDNTERGMSRSSSPRGEASSLNGESH
93	MGA (PRC1.6 组分)	MEEKQQIILANQDGGTVAGAAPTFFVILKQPGNGKTDQGILVTNQ DACALASSVSSPVKSKGKICLPADCTVGGITVTLDNNSMWNEFYH RSTEMILTKQGRRMFPYCRYWITGLDSNLKYILVMDISPVDNHRY KWNGRWWEPSGKAEPHVLGRVFIHPESPSTGHYWMHQPVSFYKLK LTNNTLDQEGHIILHSMHRYLPRLHLVPAEKAVEVIQLNGPGVHT FTFPQTEFFAVTAYQNIQITQLKIDYNPFAKGFRDDGLNNKPQRD GKQKNSSDQEGNNISSSSGHRVRLTEGQGSEIQPGDLDPLSRGHE TSGKGLEKTSLNIKRDFLGFMDTDSALSEVPQLKQEISECLIASS FEDDSRVASPLDQNGSFNVVIKEEPLDDYDYELGECPEGVTVKQE ETDEETDVYSNSDDDPILEKQLKRHNKVDNPEADHLSSKWLPSSP SGVAKAKMFKLDTGKMPVVYLEPCAVTRSTVKISELPDNMLSTSR KDKSSMLAELEYLPTYIENSNETAFCLGKESENGLRKHSPDLRVV QKYPLLKEPQWKYPDISDSISTERILDDSKDSVGDSLSGKEDLGR KRTTMLKIATAAKVVNANQNASPNVPGKRGRPRKLKLCKAGRPPK NTGKSLISTKNTPVSPGSTFPDVKPDLEDVDGVLFVSFESKEALD IHAVDGTTEESSSLQASTTNDSGYRARISQLEKELIEDLKTLRHK QVIHPGLQEVGLKLNSVDPTMSIDLKYLGVQLPLAPATSFPFWNL TGTNPASPDAGFPFVSRTGKTNDFTKIKGWRGKFHSASASRNEGG NSESSLKNRSAFCSDKLDEYLENEGKLMETSMGFSSNAPTSPVVY QLPTKSTSYVRTLDSVLKKQSTISPSTSYSLKPHSVPPVSRKAKS QNRQATFSGRTKSSYKSILPYPVSPKQKYSHVILGDKVTKNSSGI ISENQANNFVVPTLDENIFPKQISLRQAQQQQQQQQGSRPPGLSK SQVKLMDLEDCALWEGKPRTYITEERADVSLTTLLTAQASLKTKP IHTIIRKRAPPCNNDFCRLGCVCSSLALEKRQPAHCRRPDCMFGC TCLKRKVVLVKGGSKTKHFQRKAAHRDPVFYDTLGEEAREEEEGI REEEEQLKEKKKRKKLEYTICETEPEQPVRHYPLWVKVEGEVDPE PVYIPTPSVIEPMKPLLLPQPEVLSPTVKGKLLTGIKSPRSYTPK

[0551]

SEQ	描述	序列
		PNPVIREEDKDPVYLYFESMMTCARVRVYERKKEDQRQPSSSSSP SPSFQQQTSCHSSPENHNNAKEPDSEQQPLKQLTCDLEDDSDKLQ EKSWKSSCNEGESSSTSYMHQRSPGGPTKLIEIISDCNWEEDRNK ILSILSQHINSNMPQSLKVGSFIIELASQRKSRGEKNPPVYSSRV KISMPSCQDQDDMAEKSGSETPDGPLSPGKMEDISPVQTDALDSV RERLHGGKGLPFYAGLSPAGKLVAYKRKPSSSTSGLIQVASNAKV AASRKPRTLLPSTSNSKMASSSGTATNRPGKNLKAFVPAKRPIAA RPSPGGVFTQFVMSKVGALQQKIPGVSTPQTLAGTQKFSIRPSPV MVVTPVVSSEPVQVCSPVTAAVTTTTPQVFLENTTAVTPMTAISD VETKETTYSSGATTTGVVEVSETNTSTSVTSTQSTATVNLTKTTG ITTPVASVAFPKSLVASPSTITLPVASTASTSLVVVTAAASSSMV TTPTSSLGSVPIILSGINGSPPVSQRPENAAQIPVATPQVSPNTV KRAGPRLLLIPVQQGSPTLRPVSNTQLQGHRMVLQPVRSPSGMNL FRHPNGQIVQLLPLHQLRGSNTQPNLQPVMFRNPGSVMGIRLPAP SKPSETPPSSTSSSAFSVMNPVIQAVGSSSAVNVITQAPSLLSSG ASFVSQAGTLTLRISPPEPQSFASKTGSETKITYSSGGQPVGTAS LIPLQSGSFALLQLPGQKPVPSSILQHVASLQMKRESQNPDQKDE TNSIKREQETKKVLQSEGEAVDPEANVIKQNSGAATSEETLNDSL EDRGDHLDEECLPEEGCATVKPSEHSCITGSHTDQDYKDVNEEYG ARNRKSSKEKVAVLEVRTISEKASNKTVQNLSKVQHQKLGDVKVE QQKGFDNPEENSSEFPVTFKEESKFELSGSKVMEQQSNLQPEAKE KECGDSLEKDRERWRKHLKGPLTRKCVGASQECKKEADEQLIKET KTCQENSDVFQQEQGISDLLGKSGITEDARVLKTECDSWSRISNP SAFSIVPRRAAKSSRGNGHFQGHLLLPGEQIQPKQEKKGGRSSAD FTVLDLEEDDEDDNEKTDDSIDEIVDVVSDYQSEEVDDVEKNNCV EYIEDDEEHVDIETVEELSEEINVAHLKTTAAHTQSFKQPSCTHI SADEKAAERSRKAPPIPLKLKPDYWSDKLQKEAEAFAYYRRTHTA NERRRRGEMRDLFEKLKITLGLLHSSKVSKSLILTRAFSEIQGLT DQADKLIGQKNLLTRKRNILIRKVSSLSGKTEEVVLKKLEYIYAK QQALEAQKRKKKMGSDEFDISPRISKQQEGSSASSVDLGQMFINN RRGKPLILSRKKDQATENTSPLNTPHTSANLVMTPQGQLLTLKGP LFSGPVVAVSPDLLESDLKPQVAGSAVALPENDDLFMMPRIVNVT SLATEGGLVDMGGSKYPHEVPDSKPSDHLKDTVRNEDNSLEDKGR ISSRGNRDGRVTLGPTQVFLANKDSGYPQIVDVSNMQKAQEFLPK KISGDMRGIQYKWKESESRGERVKSKDSSFHKLKMKDLKDSSIEM ELRKVTSAIEEAALDSSELLTNMEDEDDTDETLTSLLNEIAFLNQ QLNDDSVGLAELPSSMDTEFPGDARRAFISKVPPGSRATFQVEHL GTGLKELPDVQGESDSISPLLLHLEDDDFSENEKQLAEPASEPDV LKIVIDSEIKDSLLSNKKAIDGGKNTSGLPAEPESVSSPPTLHMK TGLENSNSTDTLWRPMPKLAPLGLKVANPSSDADGQSLKVMPCLA PIAAKVGSVGHKMNLTGNDQEGRESKVMPTLAPVVAKLGNSGASP SSAGK
94	CBX1 (chromoshadow)	MGKKQNKKKVEEVLEEEEEEYVVEKVLDRRVVKGKVEYLLKWKGF SDEDNTWEPEENLDCPDLIAEFLQSQKTAHETDKSEGGKRKADSD SEDKGEESKPKKKKEESEKPRGFARGLEPERIIGATDSSGELMFL MKWKNSDEADLVPAKEANVKCPQVVISFYEERLTWHSYPSEDDDK KDDKN
95	SCMH1 (SAM_1/SPM)	MLVCYSVLACEILWDLPCSIMGSPLGHFTWDKYLKETCSVPAPVH CFKQSYTPPSNEFKISMKLEAQDPRNTTSTCIATVVGLTGARLRL RLDGSDNKNDFWRLVDSAEIQPIGNCEKNGGMLQPPLGFRLNASS WPMFLLKTLNGAEMAPIRIFHKEPPSPSHNFFKMGMKLEAVDRKN PHFICPATIGEVRGSEVLVTFDGWRGAFDYWCRFDSRDIFPVGWC SLTGDNLQPPGTKVVIPKNPYPASDVNTEKPSIHSSTKTVLEHQP GQRGRKPGKKRGRTPKTLISHPISAPSKTAEPLKFPKKRGPKPGS KRKPRTLLNPPPASPTTSTPEPDTSTVPQDAATIPSSAMQAPTVC

[0552]

SEQ	描述	序列
		IYLNKNGSTGPHLDKKKVQQLPDHFGPARASVVLQQAVQACIDCA YHQKTVFSFLKQGHGGEVISAVFDREQHTLNLPAVNSITYVLRFL EKLCHNLRSDNLFGNQPFTQTHLSLTAIEYSHSHDRYLPGETFVL GNSLARSLEPHSDSMDSASNPTNLVSTSQRHRPLLSSCGLPPSTA SAVRRLCSRGVLKGSNERRDMESFWKLNRSPGSDRYLESRDASRL SGRDPSSWTVEDVMQFVREADPQLGPHADLFRKHEIDGKALLLLR SDMMMKYMGLKLGPALKLSYHIDRLKQGKF
96	MPP8 (染色质结构域)	MEQVAEGARVTAVPVSAADSTEELAEVEEGVGVVGEDNDAAARGA EAFGDSEEDGEDVFEVEKILDMKTEGGKVLYKVRWKGYTSDDDTW EPEIHLEDCKEVLLEFRKKIAENKAKAVRKDIQRLSLNNDIFEAN SDSDQQSETKEDTSPKKKKKKLRQREEKSPDDLKKKKAKAGKLKD KSKPDLESSLESLVFDLRTKKRISEAKEELKESKKPKKDEVKETK ELKKVKKGEIRDLKTKTREDPKENRKTKKEKFVESQVESESSVLN DSPFPEDDSEGLHSDSREEKQNTKSARERAGQDMGLEHGFEKPLD SAMSAEEDTDVRGRRKKKTPRKAEDTRENRKLENKNAFLEKKTVP KKQRNQDRSKSAAELEKLMPVSAQTPKGRRLSGEERGLWSTDSAE EDKETKRNESKEKYQKRHDSDKEEKGRKEPKGLKTLKEIRNAFDL FKLTPEEKNDVSENNRKREEIPLDFKTIDDHKTKENKQSLKERRN TRDETDTWAYIAAEGDQEVLDSVCQADENSDGRQQILSLGMDLQL EWMKLEDFQKHLDGKDENFAATDAIPSNVLRDAVKNGDYITVKVA LNSNEEYNLDQEDSSGMTLVMLAAAGGQDDLLRLLITKGAKVNGR QKNGTTALIHAAEKNFLTTVAILLEAGAFVNVQQSNGETALMKAC KRGNSDIVRLVIECGADCNILSKHQNSALHFAKQSNNVLVYDLLK NHLETLSRVAEETIKDYFEARLALLEPVFPIACHRLCEGPDFSTD FNYKPPQNIPEGSGILLFIFHANFLGKEVIARLCGPCSVQAVVLN DKFQLPVFLDSHFVYSFSPVAGPNKLFIRLTEAPSAKVKLLIGAY RVQLQ
97	SUMO3 (Rad60-SLD)	MSEEKPKEGVKTENDHINLKVAGQDGSVVQFKIKRHTPLSKLMKA YCERQGLSMRQIRFRFDGQPINETDTPAQLEMEDEDTIDVFQQQT GGVPESSLAGHSF
98	HERC2 (Cyt-b5)	MPSESFCLAAQARLDSKWLKTDIQLAFTRDGLCGLWNEMVKDGEI VYTGTESTQNGELPPRKDDSVEPSGTKKEDLNDKEKKDEEETPAP IYRAKSILDSWVWGKQPDVNELKECLSVLVKEQQALAVQSATTTL SALRLKQRLVILERYFIALNRTVFQENVKVKWKSSGISLPPVDKK SSRPAGKGVEGLARVGSRAALSFAFAFLRRAWRSGEDADLCSELL QESLDALRALPEASLFDESTVSSVWLEVVERATRFLRSVVTGDVH GTPATKGPGSIPLQDQHLALAILLELAVQRGTLSQMLSAILLLLQ LWDSGAQETDNERSAQGTSAPLLPLLQRFQSIICRKDAPHSEGDM HLLSGPLSPNESFLRYLTLPQDNELAIDLRQTAVVVMAHLDRLAT PCMPPLCSSPTSHKGSLQEVIGWGLIGWKYYANVIGPIQCEGLAN LGVTQIACAEKRFLILSRNGRVYTQAYNSDTLAPQLVQGLASRNI VKIAAHSDGHHYLALAATGEVYSWGCGDGGRLGHGDTVPLEEPKV ISAFSGKQAGKHVVHIACGSTYSAAITAEGELYTWGRGNYGRLGH GSSEDEAIPMLVAGLKGLKVIDVACGSGDAQTLAVTENGQVWSWG DGDYGKLGRGGSDGCKTPKLIEKLQDLDVVKVRCGSQFSIALTKD GQVYSWGKGDNQRLGHGTEEHVRYPKLLEGLQGKKVIDVAAGSTH CLALTEDSEVHSWGSNDQCQHFDTLRVTKPEPAALPGLDTKHIVG IACGPAQSFAWSSCSEWSIGLRVPFVVDICSMTFEQLDLLLRQVS EGMDGSADWPPPQEKECVAVATLNLLRLQLHAAISHQVDPEFLGL GLGSILLNSLKQTVVTLASSAGVLSTVQSAAQAVLQSGWSVLLPT AEERARALSALLPCAVSGNEVNISPGRRFMIDLLVGSLMADGGLE SALHAAITAEIQDIEAKKEAQKEKEIDEQEANASTFHRSRTPLDK DLINTGICESSGKQCLPLVQLIQQLLRNIASQTVARLKDVARRIS SCLDFEQHSRERSASLDLLLRFQRLLISKLYPGESIGQTSDISSP ELMGVGSLLKKYTALLCTHIGDILPVAASIASTSWRHFAEVAYIV

[0553]

SEQ	描述	序列
		EGDFTGVLLPELVVSIVLLLSKNAGLMQEAGAVPLLGGLLEHLDR FNHLAPGKERDDHEELAWPGIMESFFTGQNCRNNEEVTLIRKADL ENHNKDGGFWTVIDGKVYDIKDFQTQSLTGNSILAQFAGEDPVVA LEAALQFEDTRESMHAFCVGQYLEPDQEIVTIPDLGSLSSPLIDT ERNLGLLLGLHASYLAMSTPLSPVEIECAKWLQSSIFSGGLQTSQ IHYSYNEEKDEDHCSSPGGTPASKSRLCSHRRALGDHSQAFLQAI ADNNIQDHNVKDFLCQIERYCRQCHLTTPIMFPPEHPVEEVGRLL LCCLLKHEDLGHVALSLVHAGALGIEQVKHRTLPKSVVDVCRVVY QAKCSLIKTHQEQGRSYKEVCAPVIERLRFLFNELRPAVCNDLSI MSKFKLLSSLPRWRRIAQKIIRERRKKRVPKKPESTDDEEKIGNE ESDLEEACILPHSPINVDKRPIAIKSPKDKWQPLLSTVTGVHKYK WLKQNVQGLYPQSPLLSTIAEFALKEEPVDVEKMRKCLLKQLERA EVRLEGIDTILKLASKNFLLPSVQYAMFCGWQRLIPEGIDIGEPL TDCLKDVDLIPPFNRMLLEVTFGKLYAWAVQNIRNVLMDASAKFK ELGIQPVPLQTITNENPSGPSLGTIPQARFLLVMLSMLTLQHGAN NLDLLLNSGMLALTQTALRLIGPSCDNVEEDMNASAQGASATVLE ETRKETAPVQLPVSGPELAAMMKIGTRVMRGVDWKWGDQDGPPPG LGRVIGELGEDGWIRVQWDTGSTNSYRMGKEGKYDLKLAELPAAA QPSAEDSDTEDDSEAEQTERNIHPTAMMFTSTINLLQTLCLSAGV HAEIMQSEATKTLCGLLRMLVESGTTDKTSSPNRLVYREQHRSWC TLGFVRSIALTPQVCGALSSPQWITLLMKVVEGHAPFTATSLQRQ ILAVHLLQAVLPSWDKTERARDMKCLVEKLFDFLGSLLTTCSSDV PLLRESTLRRRRVRPQASLTATHSSTLAEEVVALLRTLHSLTQWN GLINKYINSQLRSITHSFVGRPSEGAQLEDYFPDSENPEVGGLMA VLAVIGGIDGRLRLGGQVMHDEFGEGTVTRITPKGKITVQFSDMR TCRVCPLNQLKPLPAVAFNVNNLPFTEPMLSVWAQLVNLAGSKLE KHKIKKSTKQAFAGQVDLDLLRCQQLKLYILKAGRALLSHQDKLR QILSQPAVQETGTVHTDDGAVVSPDLGDMSPEGPQPPMILLQQLL ASATQPSPVKAIFDKQELEAAALAVCQCLAVESTHPSSPGFEDCS SSEATTPVAVQHIRPARVKRRKQSPVPALPIVVQLMEMGFSRRNI EFALKSLTGASGNASSLPGVEALVGWLLDHSDIQVTELSDADTVS DEYSDEEVVEDVDDAAYSMSTGAVVTESQTYKKRADFLSNDDYAV YVRENIQVGMMVRCCRAYEEVCEGDVGKVIKLDRDGLHDLNVQCD WQQKGGTYWVRYIHVELIGYPPPSSSSHIKIGDKVRVKASVTTPK YKWGSVTHQSVGVVKAFSANGKDIIVDFPQQSHWTGLLSEMELVP SIHPGVTCDGCQMFPINGSRFKCRNCDDFDFCETCFKTKKHNTRH TFGRINEPGQSAVFCGRSGKQLKRCHSSQPGMLLDSWSRMVKSLN VSSSVNQASRLIDGSEPCWQSSGSQGKHWIRLEIFPDVLVHRLKM IVDPADSSYMPSLVVVSGGNSLNNLIELKTININPSDTTVPLLND CTEYHRYIEIAIKQCRSSGIDCKIHGLILLGRIRAEEEDLAAVPF LASDNEEEEDEKGNSGSLIRKKAAGLESAATIRTKVFVWGLNDKD QLGGLKGSKIKVPSFSETLSALNVVQVAGGSKSLFAVTVEGKVYA CGEATNGRLGLGISSGTVPIPRQITALSSYVVKKVAVHSGGRHAT ALTVDGKVFSWGEGDDGKLGHFSRMNCDKPRLIEALKTKRIRDIA CGSSHSAALTSSGELYTWGLGEYGRLGHGDNTTQLKPKMVKVLLG HRVIQVACGSRDAQTLALTDEGLVFSWGDGDFGKLGRGGSEGCNI PQNIERLNGQGVCQIECGAQFSLALTKSGVVWTWGKGDYFRLGHG SDVHVRKPQVVEGLRGKKIVHVAVGALHCLAVTDSGQVYAWGDND HGQQGNGTTTVNRKPTLVQGLEGQKITRVACGSSHSVAWTTVDVA TPSVHEPVLFQTARDPLGASYLGVPSDADSSAASNKISGASNSKP NRPSLAKILLSLDGNLAKQQALSHILTALQIMYARDAVVGALMPA AMIAPVECPSFSSAAPSDASAMASPMNGEECMLAVDIEDRLSPNP WQEKREIVSSEDAVTPSAVTPSAPSASARPFIPVTDDLGAASIIA ETMTKTKEDVESQNKAAGPEPQALDEFTSLLIADDTRVVVDLLKL SVCSRAGDRGRDVLSAVLSGMGTAYPQVADMLLELCVTELEDVAT

[0554]

SEQ	描述	序列
		DSQSGRLSSQPVVVESSHPYTDDTSTSGTVKIPGAEGLRVEFDRQ CSTERRHDPLTVMDGVNRIVSVRSGREWSDWSSELRIPGDELKWK FISDGSVNGWGWRFTVYPIMPAAGPKELLSDRCVLSCPSMDLVTC LLDFRLNLASNRSIVPRLAASLAACAQLSALAASHRMWALQRLRK LLTTEFGQSININRLLGENDGETRALSFTGSALAALVKGLPEALQ RQFEYEDPIVRGGKQLLHSPFFKVLVALACDLELDTLPCCAETHK WAWFRRYCMASRVAVALDKRTPLPRLFLDEVAKKIRELMADSENM DVLHESHDIFKREQDEQLVQWMNRRPDDWTLSAGGSGTIYGWGHN HRGQLGGIEGAKVKVPTPCEALATLRPVQLIGGEQTLFAVTADGK LYATGYGAGGRLGIGGTESVSTPTLLESIQHVFIKKVAVNSGGKH CLALSSEGEVYSWGEAEDGKLGHGNRSPCDRPRVIESLRGIEVVD VAAGGAHSACVTAAGDLYTWGKGRYGRLGHSDSEDQLKPKLVEAL QGHRVVDIACGSGDAQTLCLTDDDTVWSWGDGDYGKLGRGGSDGC KVPMKIDSLTGLGVVKVECGSQFSVALTKSGAVYTWGKGDYHRLG HGSDDHVRRPRQVQGLQGKKVIAIATGSLHCVCCTEDGEVYTWGD NDEGQLGDGTTNAIQRPRLVAALQGKKVNRVACGSAHTLAWSTSK PASAGKLPAQVPMEYNHLQEIPIIALRNRLLLLHHLSELFCPCIP MFDLEGSLDETGLGPSVGFDTLRGILISQGKEAAFRKVVQATMVR DRQHGPVVELNRIQVKRSRSKGGLAGPDGTKSVFGQMCAKMSSFG PDSLLLPHRVWKVKFVGESVDDCGGGYSESIAEICEELQNGLTPL LIVTPNGRDESGANRDCYLLSPAARAPVHSSMFRFLGVLLGIAIR TGSPLSLNLAEPVWKQLAGMSLTIADLSEVDKDFIPGLMYIRDNE ATSEEFEAMSLPFTVPSASGQDIQLSSKHTHITLDNRAEYVRLAI NYRLHEFDEQVAAVREGMARVVPVPLLSLFTGYELETMVCGSPDI PLHLLKSVATYKGIEPSASLIQWFWEVMESFSNTERSLFLRFVWG RTRLPRTIADFRGRDFVIQVLDKYNPPDHFLPESYTCFFLLKLPR YSCKQVLEEKLKYAIHFCKSIDTDDYARIALTGEPAADDSSDDSD NEDVDSFASDSTQDYLTGH
99	BIN1 (SH3_9)	MAEMGSKGVTAGKIASNVQKKLTRAQEKVLQKLGKADETKDEQFE QCVQNFNKQLTEGTRLQKDLRTYLASVKAMHEASKKLNECLQEVY EPDWPGRDEANKIAENNDLLWMDYHQKLVDQALLTMDTYLGQFPD IKSRIAKRGRKLVDYDSARHHYESLQTAKKKDEAKIAKPVSLLEK AAPQWCQGKLQAHLVAQTNLLRNQAEEELIKAQKVFEEMNVDLQE ELPSLWNSRVGFYVNTFQSIAGLEENFHKEMSKLNQNLNDVLVGL EKQHGSNTFTVKAQPSDNAPAKGNKSPSPPDGSPAATPEIRVNHE PEPAGGATPGATLPKSPSQLRKGPPVPPPPKHTPSKEVKQEQILS LFEDTFVPEISVTTPSQFEAPGPFSEQASLLDLDFDPLPPVTSPV KAPTPSGQSIPWDLWEPTESPAGSLPSGEPSAAEGTFAVSWPSQT AEPGPAQPAEASEVAGGTQPAAGAQEPGETAASEAASSSLPAVVV ETFPATVNGTVEGGSGAGRLDLPPGFMFKVQAQHDYTATDTDELQ LKAGDVVLVIPFQNPEEQDEGWLMGVKESDWNQHKELEKCRGVFP ENFTERVP
100	PCGF2 (环指蛋白结构域)	MHRTTRIKITELNPHLMCALCGGYFIDATTIVECLHSFCKTCIVR YLETNKYCPMCDVQVHKTRPLLSIRSDKTLQDIVYKLVPGLFKDE MKRRRDFYAAYPLTEVPNGSNEDRGEVLEQEKGALSDDEIVSLSI EFYEGARDRDEKKGPLENGDGDKEKTGVRFLRCPAAMTVMHLAKF LRNKMDVPSKYKVEVLYEDEPLKEYYTLMDIAYIYPWRRNGPLPL KYRVQPACKRLTLATVPTPSEGTNTSGASECESVSDKAPSPATLP ATSSSLPSPATPSHGSPSSHGPPATHPTSPTPPSTASGATTAANG GSLNCLQTPSSTSRGRKMTVNGAPVPPLT
101	TOX (HMG box)	MDVRFYPPPAQPAAAPDAPCLGPSPCLDPYYCNKFDGENMYMSMT EPSQDYVPASQSYPGPSLESEDFNIPPITPPSLPDHSLVHLNEVE SGYHSLCHPMNHNGLLPFHPQNMDLPEITVSNMLGQDGTLLSNSI SVMPDIRNPEGTQYSSHPQMAAMRPRGQPADIRQQPGMMPHGQLT TINQSQLSAQLGLNMGGSNVPHNSPSPPGSKSATPSPSSSVHEDE

[0555]

SEQ	描述	序列
		GDDTSKINGGEKRPASDMGKKPKTPKKKKKKDPNEPQKPVSAYAL FFRDTQAAIKGQNPNATFGEVSKIVASMWDGLGEEQKQVYKKKTE AAKKEYLKQLAAYRASLVSKSYSEPVDVKTSQPPQLINSKPSVFH GPSQAHSALYLSSHYHQQPGMNPHLTAMHPSLPRNIAPKPNNQMP VTVSIANMAVSPPPPLQISPPLHQHLNMQQHQPLTMQQPLGNQLP MQVQSALHSPTMQQGFTLQPDYQTIINPTSTAAQVVTQAMEYVRS GCRNPPPQPVDWNNDYCSSGGMQRDKALYLT
102	FOXA1 (HNF3A C-末端结构域)	MLGTVKMEGHETSDWNSYYADTQEAYSSVPVSNMNSGLGSMNSMN TYMTMNTMTTSGNMTPASFNMSYANPGLGAGLSPGAVAGMPGGSA GAMNSMTAAGVTAMGTALSPSGMGAMGAQQAASMNGLGPYAAAMN PCMSPMAYAPSNLGRSRAGGGGDAKTFKRSYPHAKPPYSYISLIT MAIQQAPSKMLTLSEIYQWIMDLFPYYRQNQQRWQNSIRHSLSFN DCFVKVARSPDKPGKGSYWTLHPDSGNMFENGCYLRRQKRFKCEK QPGAGGGGGSGSGGSGAKGGPESRKDPSGASNPSADSPLHRGVHG KTGQLEGAPAPGPAASPQTLDHSGATATGGASELKTPASSTAPPI SSGPGALASVPASHPAHGLAPHESQLHLKGDPHYSFNHPFSINNL MSSSEQQHKLDFKAYEQALQYSPYGSTLPASLPLGSASVTTRSPI EPSALEPAYYQGVYSRPVLNTS
103	FOXA2 (HNF3B C-末端结构域)	MLGAVKMEGHEPSDWSSYYAEPEGYSSVSNMNAGLGMNGMNTYMS MSAAAMGSGSGNMSAGSMNMSSYVGAGMSPSLAGMSPGAGAMAGM GGSAGAAGVAGMGPHLSPSLSPLGGQAAGAMGGLAPYANMNSMSP MYGQAGLSRARDPKTYRRSYTHAKPPYSYISLITMAIQQSPNKML TLSEIYQWIMDLFPFYRQNQQRWQNSIRHSLSFNDCFLKVPRSPD KPGKGSFWTLHPDSGNMFENGCYLRRQKRFKCEKQLALKEAAGAA GSGKKAAAGAQASQAQLGEAAGPASETPAGTESPHSSASPCQEHK RGGLGELKGTPAAALSPPEPAPSPGQQQQAAAHLLGPPHHPGLPP EAHLKPEHHYAFNHPFSINNLMSSEQQHHHSHHHHQPHKMDLKAY EQVMHYPGYGSPMPGSLAMGPVTNKTGLDASPLAADTSYYQGVYS RPIMNSS
104	IRF2BP1 (IRF-2BP1_2 N-末端结构域)	MASVQASRRQWCYLCDLPKMPWAMVWDFSEAVCRGCVNFEGADRI ELLIDAARQLKRSHVLPEGRSPGPPALKHPATKDLAAAAAQGPQL PPPQAQPQPSGTGGGVSGQDRYDRATSSGRLPLPSPALEYTLGSR LANGLGREEAVAEGARRALLGSMPGLMPPGLLAAAVSGLGSRGLT LAPGLSPARPLFGSDFEKEKQQRNADCLAELNEAMRGRAEEWHGR PKAVREQLLALSACAPFNVRFKKDHGLVGRVFAFDATARPPGYEF ELKLFTEYPCGSGNVYAGVLAVARQMFHDALREPGKALASSGFKY LEYERRHGSGEWRQLGELLTDGVRSFREPAPAEALPQQYPEPAPA ALCGPPPRAPSRNLAPTPRRRKASPEPEGEAAGKMTTEEQQQRHW VAPGGPYSAETPGVPSPIAALKNVAEALGHSPKDPGGGGGPVRAG GASPAASSTAQPPTQHRLVARNGEAEVSPTAGAEAVSGGGSGTGA TPGAPLCCTLCRERLEDTHFVQCPSVPGHKFCFPCSREFIKAQGP AGEVYCPSGDKCPLVGSSVPWAFMQGEIATILAGDIKVKKERDP
105	IRF2BP2 (IRF-2BP1_2 N-末端结构域)	MAAAVAVAAASRRQSCYLCDLPRMPWAMIWDFTEPVCRGCVNYEG ADRVEFVIETARQLKRAHGCFPEGRSPPGAAASAAAKPPPLSAKD ILLQQQQQLGHGGPEAAPRAPQALERYPLAAAAERPPRLGSDFGS SRPAASLAQPPTPQPPPVNGILVPNGFSKLEEPPELNRQSPNPRR GHAVPPTLVPLMNGSATPLPTALGLGGRAAASLAAVSGTAAASLG SAQPTDLGAHKRPASVSSSAAVEHEQREAAAKEKQPPPPAHRGPA DSLSTAAGAAELSAEGAGKSRGSGEQDWVNRPKTVRDTLLALHQH GHSGPFESKFKKEPALTAGRLLGFEANGANGSKAVARTARKRKPS PEPEGEVGPPKINGEAQPWLSTSTEGLKIPMTPTSSFVSPPPPTA SPHSNRTTPPEAAQNGQSPMAALILVADNAGGSHASKDANQVHST TRRNSNSPPSPSSMNQRRLGPREVGGQGAGNTGGLEPVHPASLPD SSLATSAPLCCTLCHERLEDTHFVQCPSVPSHKFCFPCSRQSIKQ

[0556]

SEQ	描述	序列
		QGASGEVYCPSGEKCPLVGSNVPWAFMQGEIATILAGDVKVKKERDS
106	IRF2BPL IRF-2BP1_2 N-末端结构域	MSAAQVSSSRRQSCYLCDLPRMPWAMIWDFSEPVCRGCVNYEGADRIEFVIETARQLKRAHGCFQDGRSPGPPPPVGVKTVALSAKEAAAAAAAAAAAAAQQQQQQQQQQQQQQQQQQQQQQQQQLNHVDGSSKPAVLAAPSGLERYGLSAAAAAAAAAAAAVEQRSRFEYPPPPVSLGSSSHTARLPNGLGGPNGFPKPTPEEGPPELNRQSPNSSSAAASVASRRGTHGGLVTGLPNPGGGGGPQLTVPPNLLPQTLLNGPASAAVLPPPPPHALGSRGPPTPAPPGAPGGPACLGGTPGVSATSSSASSSTSSSVAEVGVGAGGKRPGSVSSTDQERELKEKQRNAEALAELSESLRNRAEEWASKPKMVRDTLLTLAGCTPYEVRFKKDHSLLGRVFAFDAVSKPGMDYELKLFIEYPTGSGNVYSSASGVAKQMYQDCMKDFGRGLSSGFKYLEYEKKHGSGDWRLLGDLLPEAVRFFKEGVPGADMLPQPYLDASCPMLPTALVSLSRAPSAPPGTGALPPAAPSGRGAAASLRKRKASPEPPDSAEGALKLGEEQQRQQWMANQSEALKLTMSAGGFAAPGHAAGGPPPPPPPLGPHSNRTTPPESAPQNGPSPMAALMSVADTLGTAHSPKDGSSVHSTTASARRNSSSPVSPASVPGQRRLASRNGDLNLQVAPPPPSAHPGMDQVHPQNIPDSPMANSGPLCCTICHERLEDTHFVQCPSVPSHKFCFPCSRESIKAQGATGEVYCPSGEKCPLVGSNVPWAFMQGEIATILAGDVKVKKERDP
107	HOXA13 (同源异型结构域)	MTASVLLHPRWIEPTVMFLYDNGGGLVADELNKNMEGAAAAAAAAAAAAAAAGAGGGGFPHPAAAAAGGNFSVAAAAAAAAAAAAANQCRNLMAHPAPLAPGAASAYSSAPGEAPPSAAAAAAAAAAAAAAAAAAAASSSGGPGPAGPAGAEAAKQCSPCSAAAQSSSGPAALPYGYFGSGYYPCARMGPHPNAIKSCAQPASAAAAAAFADKYMDTAGPAAEEFSSRAKEFAFYHQGYAAGPYHHHQPMPGYLDMPVVPGLGGPGESRHEPLGLPMESYQPWALPNGWNGQMYCPKEQAQPPHLWKSTLPDVVSHPSDASSYRRGRKKRVPYTKVQLKELEREYATNKFITKDKRRRISATTNLSERQVTIWFQNRRVKEKKVINKLKTTS
108	HOXB13 (同源异型结构域)	MEPGNYATLDGAKDIEGLLGAGGGRNLVAHSPLTSHPAAPTLMPAVNYAPLDLPGSAEPPKQCHPCPGVPQGTSPAPVPYGYFGGGYYSCRVSRSSLKPCAQAATLAAYPAETPTAGEEYPSRPTEFAFYPGYPGTYQPMASYLDVSVVQTLGAPGEPRHDSLLPVDSYQSWALAGGWNSQMCCQGEQNPPGPFWKAAFADSSGQHPPDACAFRRGRKKRIPYSKGQLRELEREYAANKFITKDKRRKISAATSLSERQITIWFQNRRVKEKKVLAKVKNSATP
109	HOXC13 (同源异型结构域)	MTTSLLLHPRWPESLMYVYEDSAAESGIGGGGGGGGGGTGGAGGGCSGASPGKAPSMDGLGSSCPASHCRDLLPHPVLGRPPAPLGAPQGAVYTDIPAPEAARQCAPPPAPPTSSSATLGYGYPFGGSYYGCRLSHNVNLQQKPCAYHPGDKYPEPSGALPGDDLSSRAKEFAFYPSFASSYQAMPGYLDVSVVPGISGHPEPRHDALIPVEGYQHWALSNGWDSQVYCSKEQSQSAHLWKSPFPDVVPLQPEVSSYRRGRKKRVPYTKVQLKELEKEYAASKFITKEKRRRISATTNLSERQVTIWFQNRRVKEKKVVSKSKAPHLHST
110	HOXA11 (同源异型结构域)	MDFDERGPCSSNMYLPSCTYYVSGPDFSSLPSFLPQTPSSRPMTYSYSSNLPQVQPVREVTFREYAIEPATKWHPRGNLAHCYSAEELVHRDCLQAPSAAGVPGDVLAKSSANVYHHPTPAVSSNFYSTVGRNGVLPQAFDQFFETAYGTPENLASSDYPGDKSAEKGPPAATATSAAAAAAATGAPATSSSDSGGGGGCRETAAAAEEKERRRRPESSSSPESSSGHTEDKAGGSSGQRTRKKRCPYTKYQIRELEREFFFSVYINKEKRLQLSRMLNLTDRQVKIWFQNRRMKEKKINRDRLQYYSANPLL
111	HOXC11 (同源异型结构域)	MFNSVNLGNFCSPSRKERGADFGERGSCASNLYLPSCTYYMPEFSTVSSFLPQAPSRQISYPYSAQVPPVREVSYGLEPSGKWHHRNSYSSCYAAADELMHRECLPPSTVTEILMKNEGSYGGHHHPSAPHATPAGFYSSVNKNSVLPQAFDRFFDNAYCGGGDPPAEPPCSGKGEAKGE

[0557]

SEQ	描述	序列
		PEAPPASGLASRAEAGAEAEAEEENTNPSSSGSAHSVAKEPAKGA APNAPRTRKKRCPYSKFQIRELEREFFFNVYINKEKRLQLSRMLN LTDRQVKIWFQNRRMKEKKLSRDRLQYFSGNPLL
112	HOXC10 (同源异型结构域)	MTCPRNVTPNSYAEPLAAPGGGERYSRSAGMYMQSGSDFNCGVMR GCGLAPSLSKRDEGSSPSLALNTYPSYLSQLDSWGDPKAAYRLEQ PVGRPLSSCSYPPSVKEENVCCMYSAEKRAKSGPEAALYSHPLPE SCLGEHEVPVPSYYRASPSYSALDKTPHCSGANDFEAPFEQRASL NPRAEHLESPQLGGKVSFPETPKSDSQTPSPNEIKTEQSLAGPKG SPSESEKERAKAADSSPDTSDNEAKEEIKAENTTGNWLTAKSGRK KRCPYTKHQTLELEKEFLFNMYLTRERRLEISKTINLTDRQVKIW FQNRRMKLKKMNRENRIRELTSNFNFT
113	HOXA10 (同源异型结构域)	MSARKGYLLPSPNYPTTMSCSESPAANSFLVDSLISSGRGEAGGG GGGAGGGGGGGYYAHGGVYLPPAADLPYGLQSCGLFPTLGGKRNE AASPGSGGGGGGLGPGAHGYGPSPIDLWLDAPRSCRMEPPDGPPP PPQQQPPPPPQPPQPAPQATSCSFAQNIKEESSYCLYDSADKCPK VSATAAELAPFPRGPPPDGCALGTSSGVPVPGYFRLSQAYGTAKG YGSGGGGAQQLGAGPFPAQPPGRGFDLPPALASGSADAARKERAL DSPPPPTLACGSGGGSQGDEEAHASSSAAEELSPAPSESSKASPE KDSLGNSKGENAANWLTAKSGRKKRCPYTKHQTLELEKEFLFNMY LTRERRLEISRSVHLTDRQVKIWFQNRRMKLKKMNRENRIRELTA NFNFS
114	HOXB9 (同源异型结构域)	MSISGTLSSYYVDSIISHESEDAPPAKFPSGQYASSRQPGHAEHL EFPSCSFQPKAPVFGASWAPLSPHASGSLPSVYHPYIQPQGVPPA ESRYLRTWLEPAPRGEAAPGQGQAAVKAEPLLGAPGELLKQGTPE YSLETSAGREAVLSNQRPGYGDNKICEGSEDKERPDQTNPSANWL HARSSRKKRCPYTKYQTLELEKEFLFNMYLTRDRRHEVARLLNLS ERQVKIWFQNRRMKMKKMNKEQGKE
115	HOXA9 (同源异型结构域)	MATTGALGNYYVDSFLLGADAADELSVGRYAPGTLGQPPRQAATL AEHPDFSPCSFQSKATVFGASWNPVHAAGANAVPAAVYHHHHHHP YVHPQAPVAAAAPDGRYMRSWLEPTPGALSFAGLPSSRPYGIKPE PLSARRGDCPTLDTHTLSLTDYACGSPPVDREKQPSEGAFSENNA ENESGGDKPPIDPNNPAANWLHARSTRKKRCPYTKHQTLELEKEF LFNMYLTRDRRYEVARLLNLTERQVKIWFQNRRMKMKKINKDRAK DE
116	ZFP28_HUMAN	NKKLEAVGTGIEPKAMSQGLVTFGDVAVDFSQEEWEWLNPIQRNL YRKVMLENYRNLASLGLCVSKPDVISSLEQGKEPW
117	ZN334_HUMAN	KMKKFQIPVSFQDLTVNFTQEEWQQLDPAQRLLYRDVMLENYSNL VSVGYHVSKPDVIFKLEQGEEPWIVEEFSNQNYPD
118	ZN568_HUMAN	CSQESALSEEEEDTTRPLETVTFKDVAVDLTQEEWEQMKPAQRNL YRDVMLENYSNLVTVGCQVTKPDVIFKLEQEEEPW
119	ZN37A_HUMAN	ITSQGSVSFRDVTVGFTQEEWQHLDPAQRTLYRDVMLENYSHLVS VGYCIPKPEVILKLEKGEEPWILEEKFPSQSHLEL
120	ZN181_HUMAN	PQVTFNDVAIDFTHEEWGWLSSAQRDLYKDVMVQNYENLVSVAGL SVTKPYVITLLEDGKEPWMMEKKLSKGMIPDWESR
121	ZN510_HUMAN	PLRFSTLFQEQQKMNISQASVSFKDVTIEFTQEEWQQMAPVQKNL YRDVMLENYSNLVSVGYCCFKPEVIFKLEQGEEPW
122	ZN862_HUMAN	QDPSAEGLSEEVPVVFEELPVVFEDVAVYFTREEWGMLDKRQKEL YRDVMRMNYELLASLGPAAAKPDLISKLERRAAPW
123	ZN140_HUMAN	SQGSVTFRDVAIDFSQEEWKWLQPAQRDLYRCVMLENYGHLVSLG LSISKPDVVSLLEQGKEPWLGKREVKRDLFSVSES
124	ZN208_HUMAN	GSLTFRDVAIEFSLEEWQCLDTAQQNLYRNVMLENYRNLVFLGIA AFKPDLIIFLEEGKESWNMKRHEMVEESPVICSHF
125	ZN248_HUMAN	NKSQEQVSFKDVCVDFTQEEWYLLDPAQKILYRDVILENYSNLVS VGYCITKPEVIFKIEQGEEPWILEKGFPSQCHPER

[0558]

SEQ	描述	序列
126	ZN571_HUMAN	PHLLVTFRDVAIDFSQEEWECLDPAQRDLYRDVMLENYSNLISLD LESSCVTKKLSPEKEIYEMESLQWENMGKRINHHL
127	ZN699_HUMAN	EEERKTAELQKNRIQDSVVFEDVAVDFTQEEWALLDLAQRNLYRD VMLENFQNLASLGYPLHTPHLISQWEQEEDLQTVK
128	ZN726_HUMAN	GLLTFRDVAIEFSLEEWQCLDTAQKNLYRNVMLENYRNLAFLGIA VSKPDLIICLEKEKEPWNMKRDEMVDEPPGICPHF
129	ZIK1_HUMAN	RAPTQVTVSPETHMDLTKGCVTFEDIAIYFSQDEWGLLDEAQRLL YLEVMLENFALVASLGCGHGTEDEETPSDQNVSVG
130	ZNF2_HUMAN	AAVSPTTRCQESVTFEDVAVVFTDEEWSRLVPIQRDLYKEVMLEN YNSIVSLGLPVPQPDVIFQLKRGDKPWMVDLHGSE
131	Z705F_HUMAN	HSLEKVTFEDVAIDFTQEEWDMMDTSKRKLYRDVMLENISHLVSL GYQISKSYIILQLEQGKELWREGRVFLQDQNPDRE
132	ZNF14_HUMAN	DSVSFEDVAVNFTLEEWALLDSSQKKLYEDVMQETFKNLVCLGKK WEDQDIEDDHRNQGKNRRCHMVERLCESRRGSKCG
133	ZN471_HUMAN	NVEVVKVMPQDLVTFKDVAIDFSQEEWQWMNPAQKRLYRSMMLEN YQSLVSLGLCISKPYVISLLEQGREPWEMTSEMTR
134	ZN624_HUMAN	TQPDEDLHLQAEETQLVKESVTFKDVAIDFTLEEWRLMDPTQRNL HKDVMLENYRNLVSLGLAVSKPDMISHLENGKGPW
135	ZNF84_HUMAN	TMLQESFSFDDLSVDFTQKEWQLLDPSQKNLYKDVMLENYSSLVS LGYEVMKPDVIFKLEQGEEPWVGDGEIPSSDSPEV
136	ZNF7_HUMAN	EVVTFGDVAVHFSREEWQCLDPGQRALYREVMLENHSSVAGLAGF LVFKPELISRLEQGEEPWVLDLQGAEGTEAPRTSK
137	ZN891_HUMAN	RNAEEERMIAVFLTTWLQEPMTFKDVAVEFTQEEWMMLDSAQRSL YRDVMLENYRNLTSVEYQLYRLTVISPLDQEEIRN
138	ZN337_HUMAN	GPQGARRQAFLAFGDVTVDFTQKEWRLLSPAQRALYREVTLENYS HLVSLGILHSKPELIRRLEQGEVPWGEERRRRPGP
139	Z705G_HUMAN	HSLKKLTFEDVAIDFTQEEWAMMDTSKRKLYRDVMLENISHLVSL GYQISKSYIILQLEQGKELWREGRVFLQDQNPNRE
140	ZN529_HUMAN	MPEVEFPDQFFTVLTMDHELVTLRDVVINFSQEEWEYLDSAQRNL YWDVMMENYSNLLSLDLESRNETKHLSVGKDIIQN
141	ZN729_HUMAN	PGAPGSLEMGPLTFRDVTIEFSLEEWQCLDTVQQNLYRDVMLENY RNLVFLGMAVFKPDLITCLKQGKEPWNMKRHEMVT
142	ZN419_HUMAN	RDPAQVPVAADLLTDHEEGYVTFEDVAVYFSQEEWRLLDDAQRLL YRNVMLENFTLLASLGLASSKTHEITQLESWEEPF
143	Z705A_HUMAN	HSLKKVTFEDVAIDFTQEEWAMMDTSKRKLYRDVMLENISHLVSL GYQISKSYIILQLEQGKELWREGREFLQDQNPDRE
144	ZNF45_HUMAN	TKSKEAVTFKDVAVVFSEEELQLLDLAQRKLYRDVMLENFRNVVS VGHQSTPDGLPQLEREEKLWMMKMATQRDNSSGAK
145	ZN302_HUMAN	SQVTFSDVAIDFSHEEWACLDSAQRDLYKDVMVQNYENLVSVGLS VTKPYVIMLLEDGKEPWMMEKKLSKAYPFPLSHSV
146	ZN486_HUMAN	PGPLRSLEMESLQFRDVAVEFSLEEWHCLDTAQQNLYRDVMLENY RHLVFLGIIVSKPDLITCLEQGIKPLTMKRHEMIA
147	ZN621_HUMAN	LQTTWPQESVTFEDVAVYFTQNQWASLDPAQRALYGEVMLENYAN VASLVAFPFPKPALISHLERGEAPWGPDPWDTEIL
148	ZN688_HUMAN	APLLAPRPGETRPGCRKPGTVSFADVAVYFSPEEWGCLRPAQRAL YRDVMQETYGHLGALGFPGPKPALISWMEQESEAW
149	ZN33A_HUMAN	NKVEQKSQESVSFKDVTVGFTQEEWQHLDPSQRALYRDVMLENYS NLVSVGYCVHKPEVIFRLQQGEEPWKQEEEFPSQS
150	ZN554_HUMAN	CFSQEERMAAGYLPRWSQELVTFEDVSMDFSQEEWELLEPAQKNL YREVMLENYRNVVSLEALKNQCTDVGIKEGPLSPA
151	ZN878_HUMAN	DSVAFEDVAVNFTQEEWALLDPSQKNLYREVMQETLRNLTSIGKK WNNQYIEDEHQNPRRNLRRLIGERLSESKESHQHG
152	ZN772_HUMAN	MGPAQVPMNSEVIVDPIQGQVNFEDVFVYFSQEEWVLLDEAQRLL YRDVMLENFALMASLGHTSFMSHIVASLVMGSEPW

[0559]

SEQ	描述	序列
153	ZN224_HUMAN	TTFKEAMTFKDVAVVFTEEELGLLDLAQRKLYRDVMLENFRNLLS VGHQAFHRDTFHFLREEKIWMMKTAIQREGNSGDK
154	ZN184_HUMAN	DSTLLQGGHNLLSSASFQEAVTFKDVIVDFTQEEWKQLDPGQRDL FRDVTLENYTHLVSIGLQVSKPDVISQLEQGTEPW
155	ZN544_HUMAN	EARSMLVPPQASVCFEDVAMAFTQEEWEQLDLAQRTLYREVTLET WEHIVSLGLFLSKSDVISQLEQEEDLCRAEQEAPR
156	ZNF57_HUMAN	DSVVFEDVAVDFTLEEWALLDSAQRDLYRDVMLETFRNLASVDDG TQFKANGSVSLQDMYGQEKSKEQTIPNFTGNNSCA
157	ZN283_HUMAN	EESHGALISSCNSRTMTDGLVTFRDVAIDFSQEEWECLDPAQRDL YVDVMLENYSNLVSLDLESKTYETKKIFSENDIFE
158	ZN549_HUMAN	VITPQIPMVTEEFVKPSQGHVTFEDIAVYFSQEEWGLLDEAQRCL YHDVMLENFSLMASVGCLHGIEAEEAPSEQTLSAQ
159	ZN211_HUMAN	VQLRPQTRMATALRDPASGSVTFEDVAVYFSWEEWDLLDEAQKHL YFDVMLENFALTSSLGCWCGVEHEETPSEQRISGE
160	ZN615_HUMAN	MQAQESLTLEDVAVDFTWEEWQFLSPAQKDLYRDVMLENYSNLVA VGYQASKPDALSKLERGEETCTTEDEIYSRICSEI
161	ZN253_HUMAN	GPLQFRDVAIEFSLEEWHCLDTAQRNLYRDVMLENYRNLVFLGIV VSKPDLVTCLEQGKKPLTMERHEMIAKPPVMSSHF
162	ZN226_HUMAN	NMFKEAVTFKDVAVAFTEEELGLLGPAQRKLYRDVMVENFRNLLS VGHPPFKQDVSPIERNEQLWIMTTATRRQGNLGEK
163	ZN730_HUMAN	GALTFRDVAIEFSLEEWQCLDTEQQNLYRNVMLDNYRNLVFLGIA VSKPDLITCLEQEKEPWNLKTHDMVAKPPVICSHI
164	Z585A_HUMAN	SPQKSSALAPEDHGSSYEGSVSFRDVAIDFSREEWRHLDPSQRNL YRDVMLETYSHLLSVGYQVPEAEVVMLEQGKEPWA
165	ZN732_HUMAN	ELLTFRDVAIEFSPEEWKCLDPAQQNLYRDVMLENYRNLISLGVA ISNPDLVIYLEQRKEPYKVKIHETVAKHPAVCSHF
166	ZN681_HUMAN	EPLKFRDVAIEFSLEEWQCLDTIQQNLYRNVMLENYRNLVFLGIV VSKPDLITCLEQEKEPWTRKRHRMVAEPPVICSHF
167	ZN667_HUMAN	PSARGKSKSKAPITFGDLAIYFSQEEWEWLSPIQKDLYEDVMLEN YRNLVSLGLSFRRPNVITLLEKGKAPWMVEPVRRR
168	ZN649_HUMAN	TKAQESLTLEDVAVDFTWEEWQFLSPAQKDLYRDVMLENYSNLVS VGYQAGKPDALTKLEQGEPLWTLEDEIHSPAHPEI
169	ZN470_HUMAN	SQEEVEVAGIKLCKAMSLGSVTFTDVAIDFSQDEWEWLNLAQRSL YKKVMLENYRNLVSVGLCISKPDVISLLEQEKDPW
170	ZN484_HUMAN	TKSLESVSFKDVTVDFSRDEWQQLDLAQKSLYREVMLENYFNLIS VGCQVPKPEVIFSLEQEEPCMLDGEIPSQSRPDGD
171	ZN431_HUMAN	SGCPGAERNLLVYSYFEKETLTFRDVAIEFSLEEWECLNPAQQNL YMNVMLENYKNLVFLGVAVSKQDPVTCLEQEKEPW
172	ZN382_HUMAN	PLQGSVSFKDVTVDFTQEEWQQLDPAQKALYRDVMLENYCHFVSV GFHMAKPDMIRKLEQGEELWTQRIFPSYSYLEEDG
173	ZN254_HUMAN	PGPPRSLEMGLLTFRDVAIEFSLEEWQHLDIAQQNLYRNVMLENY RNLAFLGIAVSKPDLITCLEQGKEPWNMKRHEMVD
174	ZN124_HUMAN	SGHPGSWEMNSVAFEDVAVNFTQEEWALLDPSQKNLYRDVMQETF RNLASIGNKGEDQSIEDQYKNSSRNLRHIISHSGN
175	ZN607_HUMAN	SYGSITFGDVAIDFSHQEWEYLSLVQKTLYQEVMMENYDNLVSLA GHSVSKPDLITLLEQGKEPWMIVREETRGECTDLD
176	ZN317_HUMAN	DLFVCSGLEPHTPSVGSQESVTFQDVAVDFTEKEWPLLDSSQRKL YKDVMLENYSNLTSLGYQVGKPSLISHLEQEEEPR
177	ZN620_HUMAN	FQTAWRQEPVTFEDVAVYFTQNEWASLDSVQRALYREVMLENYAN VASLAFPFTTPVLVSQLEQGELPWGLDPWEPMGRE
178	ZN141_HUMAN	ELLTFRDVAIEFSPEEWKCLDPDQQNLYRDVMLENYRNLVSLGVA ISNPDLVTCLEQRKEPYNVKIHKIVARPPAMCSHF
179	ZN584_HUMAN	AGEAEAQLDPSLQGLVMFEDVTVYFSREEWGLLNVTQKGLYRDVM LENFALVSSLGLAPSRSPVFTQLEDDEQSWVPSWV

[0560]

SEQ	描述	序列
180	ZN540_HUMAN	AHALVTFRDVAIDFSQKEWECLDTTQRKLYRDVMLENYNNLVSLG YSGSKPDVITLLEQGKEPCVVARDVTGRQCPGLLS
181	ZN75D_HUMAN	KRIKHWKMASKLILPESLSLLTFEDVAVYFSEEEWQLLNPLEKTL YNDVMQDIYETVISLGLKLKNDTGNDHPISVSTSE
182	ZN555_HUMAN	DSVVFEDVAVDFTLEEWALLDSAQRDLYRDVMLETFQNLASVDDE TQFKASGSVSQQDIYGEKIPKESKIATFTRNVSWA
183	ZN658_HUMAN	NMSQASVSFQDVTVEFTREEWQHLGPVERTLYRDVMLENYSHLIS VGYCITKPKVISKLEKGEEPWSLEDEFLNQRYPGY
184	ZN684_HUMAN	ISFQESVTFQDVAVDFTAEEWQLLDCAERTLYWDVMLENYRNLIS VGCPITKTKVILKVEQGQEPWMVEGANPHESSPES
185	RBAK_HUMAN	NTLQGPVSFKDVAVDFTQEEWQQLDPDEKITYRDVMLENYSHLVS VGYDTTKPNVIIKLEQGEEPWIMGGEFPCQHSPEA
186	ZN829_HUMAN	HPEEEERMHDELLQAVSKGPVMFRDVSIDFSQEEWECLDADQMNL YKEVMLENFSNLVSVGLSNSKPAVISLLEQGKEPW
187	ZN582_HUMAN	SLGSELFRDVAIVFSQEEWQWLAPAQRDLYRDVMLETYSNLVSLG LAVSKPDVISFLEQGKEPWMVERVVSGGLCPVLES
188	ZN112_HUMAN	TKFQEMVTFKDVAVVFTEEELGLLDSVQRKLYRDVMLENFRNLLL VAHQPFKPDLISQLEREEKLLMVETETPRDGCSGR
189	ZN716_HUMAN	AKRPGPPGSREMGLLTFRDIAIEFSLAEWQCLDHAQQNLYRDVML ENYRNLVSLGIAVSKPDLITCLEQNKEPQNIKRNE
190	HKR1_HUMAN	TCMVHRQTMSCSGAGGITAFVAFRDVAVYFTQEEWRLLSPAQRTL HREVMLETYNHLVSLEIPSSKPKLIAQLERGEAPW
191	ZN350_HUMAN	IQAQESITLEDVAVDFTWEEWQLLGAAQKDLYRDVMLENYSNLVA VGYQASKPDALFKLEQGEQLWTIEDGIHSGACSDI
192	ZN480_HUMAN	AQKRRKRKAKESGMALPQGHLTFRDVAIEFSQAEWKCLDPAQRAL YKDVMLENYRNLVSLGISLPDLNINSMLEQRREPW
193	ZN416_HUMAN	DSTSVPVTAEAKLMGFTQGCVTFEDVAIYFSQEEWGLLDEAQRLL YRDVMLENFALITALVCWHGMEDEETPEQSVSVEG
194	ZNF92_HUMAN	GPLTFRDVKIEFSLEEWQCLDTAQRNLYRDVMLENYRNLVFLGIA VSKPDLITWLEQGKEPWNLKRHEMVDKTPVMCSHF
195	ZN100_HUMAN	SGCPGAERSLLVQSYFEKGPLTFRDVAIEFSLEEWQCLDSAQQGL YRKVMLENYRNLVFLAGIALTKPDLITCLEQGKEP
196	ZN736_HUMAN	GVLTFRDVAVEFSPEEWECLDSAQQRLYRDVMLENYGNLVSLGLA IFKPDLMTCLEQRKEPWKVKRQEAVAKHPAGSFHF
197	ZNF74_HUMAN	KENLEDISGWGLPEARSKESVSFKDVAVDFTQEEWGQLDSPQRAL YRDVMLENYQNLLALGPPLHKPDVISHLERGEEPW
198	CBX1_HUMAN	EESEKPRGFARGLEPERIIGATDSSGELMFLMKWKNSDEADLVPA KEANVKCPQVVISFYEERLTWHSYPSEDDDKKDDK
199	ZN443_HUMAN	ASVALEDVAVNFTREEWALLGPCQKNLYKDVMQETIRNLDCVVMK WKDQNIEDQYRYPRKNLRCRMLERFVESKDGTQCG
200	ZN195_HUMAN	TLLTFRDVAIEFSLEEWKCLDLAQQNLYRDVMLENYRNLFSVGLT VCKPGLITCLEQRKEPWNVKRQEAADGHPEMGFHH
201	ZN530_HUMAN	AAALRAPTQQVFVAFEDVAIYFSQEEWELLDEMQRLLYRDVMLEN FAVMASLGCWCGAVDEGTPSAESVSVEELSQGRTP
202	ZN782_HUMAN	NTFQASVSFQDVTVEFSQEEWQHMGPVERTLYRDVMLENYSHLVS VGYCFTKPELIFTLEQGEDPWLLEKEKGFLSRNSP
203	ZN791_HUMAN	DSVAFEDVSVSFSQEEWALLAPSQKKLYRDVMQETFKNLASIGEK WEDPNVEDQHKNQGRNLRSHTGERLCEGKEGSQCA
204	ZN331_HUMAN	AQGLVTFADVAIDFSQEEWACLNSAQRDLYWDVMLENYSNLVSLD LESAYENKSLPTEKNIHEIRASKRNSDRRSKSLGR
205	Z354C_HUMAN	AVDLLSAQEPVTFRDVAVFFSQDEWLHLDSAQRALYREVMLENYS SLVSLGIPFSMPKLIHQLQQGEDPCMVEREVPSDT
206	ZN157_HUMAN	SPQRFPALIPGEPGRSFEGSVSFEDVAVDFTRQEWHRLDPAQRTM HKDVMLETYSNLASVGLCVAKPEMIFKLERGEELW

[0561]

SEQ	描述	序列
207	ZN727_HUMAN	RVLTFRDVAVEFSPEEWECLDSAQQRLYRDVMLENYGNLFSLGLA IFKPDLITYLEQRKEPWNARRQKTVAKHPAGSLHF
208	ZN550_HUMAN	AETKDAAQMLVTFKDVAVTFTREEWRQLDLAQRTLYREVMLETCG LLVSLGHRVPKPELVHLLEHGQELWIVKRGLSHAT
209	ZN793_HUMAN	IEYQIPVSFKDVVVGFTQEEWHRLSPAQRALYRDVMLETYSNLVS VGYEGTKPDVILRLEQEEAPWIGEAACPGCHCWED
210	ZN235_HUMAN	TKFQEAVTFKDVAVAFTEEELGLLDSAQRKLYRDVMLENFRNLVS VGHQSFKPDMISQLEREEKLWMKELQTQRGKHSGD
211	ZNF8_HUMAN	DEGVAGVMSVGPPAARLQEPVTFRDVAVDFTQEEWGQLDPTQRIL YRDVMLETFGHLLSIGPELPKPEVISQLEQGTELW
212	ZN724_HUMAN	GPLTFMDVAIEFSVEEWQCLDTAQQNLYRNVMLENYRNLVFLGIA VSKPDLITCLEQGKEPWNMERHEMVAKPPGMCCYF
213	ZN573_HUMAN	HQVGLIRSYNSKTMTCFQELVTFRDVAIDFSRQEWEYLDPNQRDL YRDVMLENYRNLVSLGGHSISKPVVVDLLERGKEP
214	ZN577_HUMAN	NATIVMSVRREQGSSSGEGSLSFEDVAVGFTREEWQFLDQSQKVL YKEVMLENYINLVSIGYRGTKPDSLFKLEQGEPPG
215	ZN789_HUMAN	FPPARGKELLSFEDVAMYFTREEWGHLNWGQKDLYRDVMLENYRN MVLLGFQFPKPEMICQLENWDEQWILDLPRTGNRK
216	ZN718_HUMAN	ELLTFKDVAIEFSPEEWKCLDTSQQNLYRDVMLENYRNLVSLGVS ISNPDLVTSLEQRKEPYNLKIHETAARPPAVCSHF
217	ZN300_HUMAN	MKSQGLVSFKDVAVDFTQEEWQQLDPSQRTLYRDVMLENYSHLVS MGYPVSKPDVISKLEQGEEPWIIKGDISNWIYPDE
218	ZN383_HUMAN	AEGSVMFSDVSIDFSQEEWDCLDPVQRDLYRDVMLENYGNLVSMG LYTPKPQVISLLEQGKEPWMVGRELTRGLCSDLES
219	ZN429_HUMAN	GPLTFTDVAIEFSLEEWQCLDTAQQNLYRNVMLENYRNLVFLGIA VSKPDLITCLEKEKEPCKMKRHEMVDEPPVVCSHF
220	ZN677_HUMAN	ALSQGLFTFKDVAIEFSQEEWECLDPAQRALYRDVMLENYRNLLS LDEDNIPPEDDISVGFTSKGLSPKENNKEELYHLV
221	ZN850_HUMAN	NMEGLVMFQDLSIDFSQEEWECLDAAQKDLYRDVMMENYSSLVSL GLSIPKPDVISLLEQGKEPWMVSRDVLGGWCRDSE
222	ZN454_HUMAN	AVSHLPTMVQESVTFKDVAILFTQEEWGQLSPAQRALYRDVMLEN YSNLVSLGLLGPKPDTFSQLEKREVWMPEDTPGGF
223	ZN257_HUMAN	GPLTIRDVTVEFSLEEWHCLDTAQQNLYRDVMLENYRNLVFLGIA VSKPDLITCLEQGKEPCNMKRHEMVAKPPVMCSHI
224	ZN264_HUMAN	AAAVLTDRAQVSVTFDDVAVTFTKEEWGQLDLAQRTLYQEVMLEN CGLLVSLGCPVPKAELICHLEHGQEPWTRKEDLSQ
225	ZFP82_HUMAN	ALRSVMFSDVSIDFSPEEWEYLDLEQKDLYRDVMLENYSNLVSLG CFISKPDVISSLEQGKEPWKVVRKGRRQYPDLETK
226	ZFP14_HUMAN	AHGSVTFRDVAIDFSQEEWEFLDPAQRDLYRDVMWENYSNFISLG PSISKPDVITLLDEERKEPGMVVREGTRRYCPDLE
227	ZN485_HUMAN	APRAQIQGPLTFGDVAVAFTRIEWRHLDAAQRALYRDVMLENYGN LVSVGLLSSKPKLITQLEQGAEPWTEVREAPSGTH
228	ZN737_HUMAN	GPLQFRDVAIEFSLEEWHCLDTAQRNLYRNVMLENYRNLVFLGIV VSKPDLITCLEQGKKPLTMKKHEMVANPSVTCSHF
229	ZNF44_HUMAN	TLPRGQPEVLEWGLPKDQDSVAFEDVAVNFTHEEWALLGPSQKNL YRDVMRETIRNLNCIGMKWENQNIDDQHQNLRRNP
230	ZN596_HUMAN	PSPDSMTFEDIIVDFTQEEWALLDTSQRKLFQDVMLENISHLVSI GKQLCKSVVLSQLEQVEKLSTQRISLLQGREVGIK
231	ZN565_HUMAN	EESREIRAGQIVLKAMAQGLVTFRDVAIEFSLEEWKCLEPAQRDL YREVTLENFGHLASLGLSISKPDVVSLLEQGKEPW
232	ZN543_HUMAN	AASAQVSVTFEDVAVTFTQEEWGQLDAAQRTLYQEVMLETCGLLM SLGCPLFKPELIYQLDHRQELWMATKDLSQSSYPG
233	ZFP69_HUMAN	RESLEDEVTPGLPTAESQELLTFKDISIDFTQEEWGQLAPAHQNL YREVMLENYSNLVSVGYQLSKPSVISQLEKGEEPW

[0562]

SEQ	描述	序列
234	SUMO1_HUMAN	EGEYIKLKVIGQDSSEIHFKVKMTTHLKKLKESYCQRQGVPMNSL RFLFEGQRIADNHTPKELGMEEEDVIEVYQEQTGG
235	ZNF12_HUMAN	NKSLGPVSFKDVAVDFTQEEWQQLDPEQKITYRDVMLENYSNLVS VGYHIIKPDVISKLEQGEEPWIVEGEFLLQSYPDE
236	ZN169_HUMAN	SPGLLTTRKEALMAFRDVAVAFTQKEWKLLSSAQRTLYREVMLEN YSHLVSLGIAFSKPKLIEQLEQGDEPWREENEHLL
237	ZN433_HUMAN	MFQDSVAFEDVAVTFTQEEWALLDPSQKNLCRDVMQETFRNLASI GKKWKPQNIYVEYENLRRNLRIVGERLFESKEGHQ
238	SUMO3_HUMAN	ENDHINLKVAGQDGSVVQFKIKRHTPLSKLMKAYCERQGLSMRQI RFRFDGQPINETDTPAQLEMEDEDTIDVFQQQTGG
239	ZNF98_HUMAN	PGPLGSLEMGVLTFRDVALEFSLEEWQCLDTAQQNLYRNVMLENY RNLVFVGIAASKPDLITCLEQGKEPWNVKRHEMVT
240	ZN175_HUMAN	LSQKPQVLGPEKQDGSCEASVSFEDVTVDFSREEWQQLDPAQRCL YRDVMLELYSHLFAVGYHIPNPEVIFRMLKEKEPR
241	ZN347_HUMAN	ALTQGQVTFRDVAIEFSQEEWTCLDPAQRTLYRDVMLENYRNLAS LGISCFDLSIISMLEQGKEPFTLESQVQIAGNPDG
242	ZNF25_HUMAN	NKFQGPVTLKDVIVEFTKEEWKLLTPAQRTLYKDVMLENYSHLVS VGYHVNKPNAVFKLKQGKEPWILEVEFPHRGFPED
243	ZN519_HUMAN	ELLTFRDVAIEFSPEEWKCLDPAQQNLYRDVMLENYRNLVSLAVY SYYNQGILPEQGIQDSFKKATLGRYGSCGLENICL
244	Z585B_HUMAN	SPQKSSALAPEDHGSSYEGSVSFRDVAIDFSREEWRHLDLSQRNL YRDVMLETYSHLLSVGYQVPKPEVVMLEQGKEPWA
245	ZIM3_HUMAN	NNSQGRVTFEDVTVNFTQGEWQRLNPEQRNLYRDVMLENYSNLVS VGQGETTKPDVILRLEQGKEPWLEEEEVLGSGRAE
246	ZN517_HUMAN	AMALPMPGPQEAVVFEDVAVYFTRIEWSCLAPDQQALYRDVMLEN YGNLASLGFLVAKPALISLLEQGEEPGALILQVAE
247	ZN846_HUMAN	DSSQHLVTFEDVAVDFTQEEWTLLDQAQRDLYRDVMLENYKNLII LAGSELFKRSLMSGLEQMEELRTGVTGVLQELDLQ
248	ZN230_HUMAN	TTFKEAVTFKDVAVFFTEEELGLLDPAQRKLYQDVMLENFTNLLS VGHQPFHPFHFLREEKFWMMETATQREGNSGGKTI
249	ZNF66_HUMAN	GPLQFRDVAIEFSLEEWHCLDMAQRNLYRDVMLENYRNLVFLGIV VSKPDLITHLEQGKKPSTMQRHEMVANPSVLCSHF
250	ZFP1_HUMAN	NKSQGSVSFTDVTVDFTQEEWEQLDPSQRILYMDVMLENYSNLLS VEVWKADDQMERDHRNPDEQARQFLILKNQTPIEE
251	ZN713_HUMAN	EEEEMNDGSQMVRSQESLTFQDVAVDFTREEWDQLYPAQKNLYRD VMLENYRNLVALGYQLCKPEVIAQLELEEEWVIER
252	ZN816_HUMAN	EEATKKSKEKEPGMALPQGRLTFRDVAIEFSLEEWKCLNPAQRAL YRAVMLENYRNLEFVDSSLKSMMEFSSTRHSITGE
253	ZN426_HUMAN	EKTPAGRIVADCLTDCYQDSVTFDDVAVDFTQEEWTLLDSTQRSL YSDVMLENYKNLATVGGQIIKPSLISWLEQEESRT
254	ZN674_HUMAN	AMSQESLTFKDVFVDFTLEEWQQLDSAQKNLYRDVMLENYSHLVS VGHLVGKPDVIFRLGPGDESWMADGGTPVRTCAGE
255	ZN627_HUMAN	DSVAFEDVAVNFTLEEWALLDPSQKNLYRDVMRETFRNLASVGKQ WEDQNIEDPFKIPRRNISHIPERLCESKEGGQGEE
256	ZNF20_HUMAN	MFQDSVAFEDVAVSFTQEEWALLDPSQKNLYRDVMQETFKNLTSV GKTWKVQNIEDEYKNPRRNLSLMREKLCESKESHH
257	Z587B_HUMAN	AVVATLRLSAQGTVTFEDVAVKFTQEEWNLLSEAQRCLYRDVTLE NLALMSSLGCWCGVEDEAAPSKQSIYIQRETQVRT
258	ZN316_HUMAN	EEEEEDEDEDDLLTAGCQELVTFEDVAVYFSLEEWERLEADQRGL YQEVMQENYGILVSLGYPIPKPDLIFRLEQGEEPW
259	ZN233_HUMAN	TKFQEMVTFKDVAVVFTREELGLLDLAQRKLYQDVMLENFRNLLS VGYQPFKLDVILQLGKEDKLRMMETEIQGDGCSGH
260	ZN611_HUMAN	EEAAQKRKGKEPGMALPQGRLTFRDVAIEFSLAEWKCLNPSQRAL YREVMLENYRNLEAVDISSKCMMKEVLSTGQGNTE

[0563]

SEQ	描述	序列
261	ZN556_HUMAN	DTVVFEDVVVDFTLEEWALLNPAQRKLYRDVMLETFKHLASVDNE AQLKASGSISQQDTSGEKLSLKQKIEKFTRKNIWA
262	ZN234_HUMAN	TTFKEGLTFKDVAVVFTEEELGLLDPVQRNLYQDVMLENFRNLLS VGHHPFKHDVFLLEKEKKLDIMKTATQRKGKSADK
263	ZN560_HUMAN	SALQQEFWKIQTSNGIQMDLVTFDSVAVEFTQEEWTLLDPAQRNL YSDVMLENYKNLSSVGYQLFKPSLISWLEEEEELS
264	ZNF77_HUMAN	DCVIFEEVAVNFTPEEWALLDHAQRSLYRDVMLETCRNLASLDCY IYVRTSGSSSQRDVFGNGISNDEEIVKFTGSDSWS
265	ZN682_HUMAN	ELLTFRDVTIEFSLEEWEFLNPAQQSLYRKVMLENYRNLVSLGLT VSKPELISRLEQRQEPWNVKRHETIAKPPAMSSHY
266	ZN614_HUMAN	IKTQESLTLEDVAVEFSWEEWQLLDTAQKNLYRDVMVENYNHLVS LGYQTSKPDVLSKLAHGQEPWTTDAKIQNKNCPGI
267	ZN785_HUMAN	PAHVPGEAGPRRTRESRPGAVSFADVAVYFSPEEWECLRPAQRAL YRDVMRETFGHLGALGFSVPKPAFISWVEGEVEAW
268	ZN445_HUMAN	GCPGDQVTPTRSLTAQLQETMTFKDVEVTFSQDEWGWLDSAQRNL YRDVMLENYRNMASLVGPFTKPALISWLEAREPWG
269	ZFP30_HUMAN	ARDLVMFRDVAVDFSQEEWECLNSYQRNLYRDVILENYSNLVSLA GCSISKPDVITLLEQGKEPWMVVRDEKRRWTLDLE
270	ZN225_HUMAN	TTLKEAVTFKDVAVVFTEEELRLLDLAQRKLYREVMLENFRNLLS VGHQSLHRDTFHFLKEEKFWMMETATQREGNLGGK
271	ZN551_HUMAN	SPPSPRSSMAAVALRDSAQGMTFEDVAIYFSQEEWELLDESQRFL YCDVMLENFAHVTSLGYCHGMENEAIASEQSVSIQ
272	ZN610_HUMAN	DEEAQKRKAKESGMALPQGRLTFMDVAIEFSQEEWKSLDPGQRAL YRDVMLENYRNLVFLGICLPDLSIISMLKQRREPL
273	ZN528_HUMAN	ALTQGPLKFMDVAIEFSQEEWKCLDPAQRTLYRDVMLENYRNLVS LGICLPDLSVTSMLEQKRDPWTLQSEEKIANDPDG
274	ZN284_HUMAN	TMFKEAVTFKDVAVVFTEEELGLLDVSQRKLYRDVMLENFRNLLS VGHQLSHRDTFHFQREEKFWIMETATQREGNSGGK
275	ZN418_HUMAN	QGTVAFEDVAVNFSQEEWSLLSEVQRCLYHDVMLENWVLISSLGC WCGSEDEEAPSKKSISIQRVSQVSTPGAGVSPKKA
276	MPP8_HUMAN	AEAFGDSEEDGEDVFEVEKILDMKTEGGKVLYKVRWKGYTSDDDT WEPEIHLEDCKEVLLEFRKKIAENKAKAVRKDIQR
277	ZN490_HUMAN	VLQMQNSEHHGQSIKTQTDSISLEDVAVNFTLEEWALLDPGQRNI YRDVMRATFKNLACIGEKWKDQDIEDEHKNQGRNL
278	ZN805_HUMAN	AMALTDPAQVSVTFDDVAVTFTQEEWGQLDLAQRTLYQEVMLENC GLLVSLGCPVPRPELIYHLEHGQEPWTRKEDLSQG
279	Z780B_HUMAN	VHGSVTFRDVAIDFSQEEWECLQPDQRTLYRDVMLENYSHLISLG SSISKPDVITLLEQEKEPWIVVSKETSRWYPDLES
280	ZN763_HUMAN	DPVACEDVAVNFTQEEWALLDISQRKLYREVMLETFRNLTSIGKK WKDQNIEYEYQNPRRNFRSLIEGNVNEIKEDSHCG
281	ZN285_HUMAN	IKFQERVTFKDVAVVFTKEELALLDKAQINLYQDVMLENFRNLML VRDGIKNNILNLQAKGLSYLSQEVLHCWQIWKQRI
282	ZNF85_HUMAN	GPLTFRDVAIEFSLKEWQCLDTAQRNLYRNVMLENYRNLVFLGIT VSKPDLITCLEQGKEAWSMKRHEIMVAKPTVMCSH
283	ZN223_HUMAN	TMSKEAVTFKDVAVVFTEEELGLLDLAQRKLYRDVMLENFRNLLS VGHQPFHRDTFHFLREEKFWMMDIATQREGNSGGK
284	ZNF90_HUMAN	GPLEFRDVAIEFSLEEWHCLDTAQQNLYRDVMLENYRHLVFLGIV VTKPDLITCLEQGKKPFTVKRHEMIAKSPVMCFHF
285	ZN557_HUMAN	GHTEGGELVNELLKSWLKGLVTFEDVAVEFTQEEWALLDPAQRTL YRDVMLENCRNLASLGNQVDKPRLISQLEQEDKVM
286	ZN425_HUMAN	AEPASVTVTFDDVALYFSEQEWEILEKWQKQMYKQEMKTNYETLD SLGYAFSKPDLITWMEQGRMLLISEQGCLDKTRRT
287	ZN229_HUMAN	HSQASAISQDREEKIMSQEPLSFKDVAVVFTEEELELLDSTQRQL YQDVMQENFRNLLSVGERNPLGDKNGKDTEYIQDE

[0564]

SEQ	描述	序列
288	ZN606_HUMAN	GSLEEGRRATGLPAAQVQEPVTFKDVAVDFTQEEWGQLDLVQRTL YRDVMLETYGHLLSVGNQIAKPEVISLLEQGEEPW
289	ZN155_HUMAN	TTFKEAVTFKDVAVVFTEEELGLLDPAQRKLYRDVMLENFRNLLS VGHQPFHQDTCHFLREEKFWMMGTATQREGNSGGK
290	ZN222_HUMAN	AKLYEAVTFKDVAVIFTEEELGLLDPAQRKLYRDVMLENFRNLLS VGGKIQTEMETVPEAGTHEEFSCKQIWEQIASDLT
291	ZN442_HUMAN	RSDLFLPDSQTNEERKQYDSVAFEDVAVNFTQEEWALLGPSQKSL YRDVMWETIRNLDCIGMKWEDTNIEDQHRNPRRSL
292	ZNF91_HUMAN	PGTPGSLEMGLLTFRDVAIEFSPEEWQCLDTAQQNLYRNVMLENY RNLAFLGIALSKPDLITYLEQGKEPWNMKQHEMVD
293	ZN135_HUMAN	TPGVRVSTDPEQVTFEDVVVGFSQEEWGQLKPAQRTLYRDVMLDT FRLLVSVGHWLPKPNVISLLEQEAELWAVESRLPQ
294	ZN778_HUMAN	EQTQAAGMVAGWLINCYQDAVTFDDVAVDFTQEEWTLLDPSQRDL YRDVMLENYENLASVEWRLKTKGPALRQDRSWFRA
295	RYBP_HUMAN	PSEANSIQSANATTKTSETNHTSRPRLKNVDRSTAQQLAVTVGNV TVIITDFKEKTRSSSTSSSTVTSSAGSEQQNQSSS
296	ZN534_HUMAN	ALTQGQLSFSDVAIEFSQEEWKCLDPGQKALYRDVMLENYRNLVS LGEDNVRPEACICSGICLPDLSVTSMLEQKRDPWT
297	ZN586_HUMAN	AAAAALRAPAQSSVTFEDVAVNFSLEEWSLLNEAQRCLYRDVMLE TLTLISSLGCWHGGEDEAAPSKQSTCIHIYKDQGG
298	ZN567_HUMAN	AQGSVSFNDVTVDFTQEEWQHLDHAQKTLYMDVMLENYCHLISVG CHMTKPDVILKLERGEEPWTSFAGHTCLEENWKAE
299	ZN440_HUMAN	DPVAFKDVAVNFTQEEWALLDISQRKLYREVMLETFRNLTSLGKR WKDQNIEYEHQNPRRNFRSLIEEKVNEIKDDSHCG
300	ZN583_HUMAN	SKDLVTFGDVAVNFSQEEWEWLNPAQRNLYRKVMLENYRSLVSLG VSVSKPDVISLLEQGKEPWMVKKEGTRGPCPDWEY
301	ZN441_HUMAN	DSVAFEDVAINFTCEEWALLGPSQKSLYRDVMQETIRNLDCIGMI WQNHDIEEDQYKDLRRNLRCHMVERACEIKDNSQC
302	ZNF43_HUMAN	GPLTFMDVAIEFCLEEWQCLDIAQQNLYRNVMLENYRNLVFLGIA VSKPDLITCLEQEKEPWEPMRRHEMVAKPPVMCSH
303	CBX5_HUMAN	QSNDIARGFERGLEPEKIIGATDSCGDLMFLMKWKDTDEADLVLA KEANVKCPQIVIAFYEERLTWHAYPEDAENKEKET
304	ZN589_HUMAN	ALPAKDSAWPWEEKPRYLGPVTFEDVAVLFTEAEWKRLSLEQRNL YKEVMLENLRNLVSLAESKPEVHTCPSCPLAFGSQ
305	ZNF10_HUMAN	DAKSLTAWSRTLVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLEN YKNLVSLGYQLTKPDVILRLEKGEEPWLVEREIHQ
306	ZN563_HUMAN	DAVAFEDVAVNFTQEEWALLGPSQKNLYRYVMQETIRNLDCIRMI WEEQNTEDQYKNPRRNLRCHMVERFSESKDSSQCG
307	ZN561_HUMAN	EKTKVERMVEDYLASGYQDSVTFDDVAVDFTPEEWALLDTTEKYL YRDVMLENYMNLASVEWEIQPRTKRSSLQQGFLKN
308	ZN136_HUMAN	DSVAFEDVDVNFTQEEWALLDPSQKNLYRDVMWETMRNLASIGKK WKDQNIKDHYKHRGRNLRSHMLERLYQTKDGSQRG
309	ZN630_HUMAN	IESQEPVTFEDVAVDFTQEEWQQLNPAQKTLHRDVMLETYNHLVS VGCSGIKPDVIFKLEHGKDPWIIESELSRWIYPDR
310	ZN527_HUMAN	AVGLCKAMSQGLVTFRDVALDFSQEEWEWLKPSQKDLYRDVMLEN YRNLVWLGLSISKPNMISLLEQGKEPWMVERKMSQ
311	ZN333_HUMAN	DKVEEEAMAPGLPTACSQEPVTFADVAVVFTPEEWVFLDSTQRSL YRDVMLENYRNLASVADQLCKPNALSYLEERGEQW
312	Z324B_HUMAN	TFEDVAVYFSQEEWGLLDTAQRALYRHVMLENFTLVTSLGLSTSR PRVVIQLERGEEPWVPSGKDMTLARNTYGRLNSGS
313	ZN786_HUMAN	AEPPRLPLTFEDVAIYFSEQEWQDLEAWQKELYKHVMRSNYETLV SLDDGLPKPELISWIEHGGEPFRKWRESQKSGNII
314	ZN709_HUMAN	DSVVFEDVAVNFTQEEWALLGPSQKKLYRDVMQETFVNLASIGEN WEEKNIEDHKNQGRKLRSHMVERLCERKEGSQFGE

[0565]

SEQ	描述	序列
315	ZN792_HUMAN	AAAALRDPAQGCVTFEDVTIYFSQEEWVLLDEAQRLLYCDVMLEN FALIASLGLISFRSHIVSQLEMGKEPWVPDSVDMT
316	ZN599_HUMAN	AAPALALVSFEDVVVTFTGEEWGHLDLAQRTLYQEVMLETCRLLV SLGHPVPKPELIYLLEHGQELWTVKRGLSQSTCAG
317	ZN613_HUMAN	IKSQESLTLEDVAVEFTWEEWQLLGPAQKDLYRDVMLENYSNLVS VGYQASKPDALFKLEQGEPWTVENEIHSQICPEIK
318	ZF69B_HUMAN	GESLESRVTLGSLTAESQELLTFKDVSVDFTQEEWGQLAPAHRNL YREVMLENYGNLVSVGCQLSKPGVISQLEKGEEPW
319	ZN799_HUMAN	ASVALEDVAVNFTREEWALLGPCQKNLYKDVMQETIRNLDCVGMK WKDQNIEDQYRYPRKNLRCRMLERFVESKDGTQCG
320	ZN569_HUMAN	TESQGTVTFKDVAIDFTQEEWKRLDPAQRKLYRNVMLENYNNLIT VGYPFTKPDVIFKLEQEEEPWVMEEEVLRRHWQGE
321	ZN564_HUMAN	DSVASEDVAVNFTLEEWALLDPSQKKLYRDVMRETFRNLACVGKK WEDQSIEDWYKNQGRILRNHMEEGLSESKEYDQCG
322	ZN546_HUMAN	EETQGELTSSCGSKTMANVSLAFRDVSIDLSQEEWECLDAVQRDL YKDVMLENYSNLVSLGYTIPKPDVITLLEQEKEPW
323	ZFP92_HUMAN	AAILLTTRPKVPVSFEDVSVYFTKTEWKLLDLRQKVLYKRVMLEN YSHLVSLGFSFSKPHLISQLERGEGPWVADIPRTW
324	YAF2_HUMAN	KDKVEKEKSEKETTSKKNSHKKTRPRLKNVDRSSAQHLEVTVGDL TVIITDFKEKTKSPPASSAASADQHSQSGSSSDNT
325	ZN723_HUMAN	GPLTFTDVAIKFSLEEWQFLDTAQQNLYRDVMLENYRNLVFLGVG VSKPDLITCLEQGKEPWNMKRHKMVAKPPVVCSHF
326	ZNF34_HUMAN	RKPNPQAMAALFLSAPPQAEVTFEDVAVYLSREEWGRLGPAQRGL YRDVMLETYGNLVSLGVGPAGPKPGVISQLERGDE
327	ZN439_HUMAN	LSLSPILLYTCEMFQDPVAFKDVAVNFTQEEWALLDISQKNLYRE VMLETFWNLTSIGKKWKDQNIEYEYQNPRRNFRSV
328	ZFP57_HUMAN	AAGEPRSLLFFQKPVTFEDVAVNFTQEEWDCLDASQRVLYQDVMS ETFKNLTSVARIFLHKPELITKLEQEEEQWRETRV
329	ZNF19_HUMAN	AAMPLKAQYQEMVTFEDVAVHFTKTEWTGLSPAQRALYRSVMLEN FGNLTALGYPVPKPALISLLERGDMAWGLEAQDDP
330	ZN404_HUMAN	ARVPLTFSDVAIDFSQEEWEYLNSDQRDLYRDVMLENYTNLVSLD FNFTTESNKLSSEKRNYEVNAYHQETWKRNKTFNL
331	ZN274_HUMAN	ASRLPTAWSCEPVTFEDVTLGFTPEEWGLLDLKQKSLYREVMLEN YRNLVSVEHQLSKPDVVSQLEEAEDFWPVERGIPQ
332	CBX3_HUMAN	SKKKRDAADKPRGFARGLDPERIIGATDSSGELMFLMKWKDSDEA DLVLAKEANMKCPQIVIAFYEERLTWHSCPEDEAQ
333	ZNF30_HUMAN	AHKYVGLQYHGSVTFEDVAIAFSQQEWESLDSSQRGLYRDVMLEN YRNLVSMGHSRSKPHVIALLEQWKEPEVTVRKDGR
334	ZN250_HUMAN	AAARLLPVPAGPQPLSFQAKLTFEDVAVLLSQDEWDRLCPAQRGL YRNVMMETYGNVVSLGLPGSKPDIISQLERGEDPW
335	ZN570_HUMAN	AVGLLKAMYQELVTFRDVAVDFSQEEWDCLDSSQRHLYSNVMLEN YRILVSLGLCFSKPSVILLLEQGKAPWMVKRELTK
336	ZN675_HUMAN	GLLTFRDVAIEFSLEEWQCLDTAQRNLYKNVILENYRNLVFLGIA VSKQDLITCLEQEKEPLTVKRHEMVNEPPVMCSHF
337	ZN695_HUMAN	GLLAFRDVALEFSPEEWECLDPAQRSLYRDVMLENYRNLISLGED SFNMQFLFHSLAMSKPELIICLEARKEPWNVNTEK
338	ZN548_HUMAN	NLTEGRVVFEDVAIYFSQEEWGHLDEAQRLLYRDVMLENLALLSS LGSWHGAEDEEAPSQQGFSVGVSEVTASKPCLSSQ
339	ZN132_HUMAN	GPAQHTSWPCGSAVPTLKSMVTFEDVAVYFSQEEWELLDAAQRHL YHSVMLENLELVTSLGSWHGVEGEGAHPKQNVSVE
340	ZN738_HUMAN	SGYPGAERNLLEYSYFEKGPLTFRDVVIEFSQEEWQCLDTAQQDL YRKVMLENFRNLVFLGIDVSKPDLITCLEQGKDPW
341	ZN420_HUMAN	ARKLVMFRDVAIDFSQEEWECLDSAQRDLYRDVMLENYSNLVSLD LPSRCASKDLSPEKNTYETELSQWEMSDRLENCDL

[0566]

SEQ	描述	序列
342	ZN626_HUMAN	GPLQFRDVAIEFSLEEWHCLDTAQRNLYRNVMLENYSNLVFLGIT VSKPDLITCLEQGRKPLTMKRNEMIAKPSVMCSHF
343	ZN559_HUMAN	VAGWLTNYSQDSVTFEDVAVDFTQEEWTLLDQTQRNLYRDVMLEN YKNLVAVDWESHINTKWSAPQQNFLQGKTSSVVEM
344	ZN460_HUMAN	AAAWMAPAQESVTFEDVAVTFTQEEWGQLDVTQRALYVEVMLETC GLLVALGDSTKPETVEPIPSHLALPEEVSLQEQLA
345	ZN268_HUMAN	VLEWLFISQEQPKITKSWGPLSFMDVFVDFTWEEWQLLDPAQKCL YRSVMLENYSNLVSLGYQHTKPDIIFKLEQGEELC
346	ZN304_HUMAN	AAAVLMDRVQSCVTFEDVFVYFSREEWELLEEAQRFLYRDVMLEN FALVATLGFWCEAEHEAPSEQSVSVEGVSQVRTAE
347	ZIM2_HUMAN	AGSQFPDFKHLGTFLVFEELVTFEDVLVDFSPEELSSLSAAQRNL YREVMLENYRNLVSLGHQFSKPDIISRLEEEESYA
348	ZN605_HUMAN	IQSQISFEDVAVDFTLEEWQLLNPTQKNLYRDVMLENYSNLVFLE VWLDNPKMWLRDNQDNLKSMERGHKYDVFGKIFNS
349	ZN844_HUMAN	DLVAFEDVAVNFTQEEWSLLDPSQKNLYREVMQETLRNLASIGEK WKDQNIEDQYKNPRNNLRSLLGERVDENTEENHCG
350	SUMO5_HUMAN	KDEDIKLRVIGQDSSEIHFKVKMTTPLKKLKKSYCQRQGVPVNSL RFLFEGQRIADNHTPEELGMEEEDVIEVYQEQIGG
351	ZN101_HUMAN	DSVAFEDVAVNFTQEEWALLSPSQKNLYRDVTLETFRNLASVGIQ WKDQDIENLYQNLGIKLRSLVERLCGRKEGNEHRE
352	ZN783_HUMAN	RNFWILRLPPGSKGEAPKVPVTFDDVAVYFSELEWGKLEDWQKEL YKHVMRGNYETLVSLDYAISKPDILTRIERGEEPC
353	ZN417_HUMAN	AAAAPRRPTQQGTVTFEDVAVNFSQEEWCLLSEAQRCLYRDVMLE NLALISSLGCWCGSKDEEAPCKQRISVQRESQSRT
354	ZN182_HUMAN	SGEDSGSFYSWQKAKREQGLVTFEDVAVDFTQEEWQYLNPPQRTL YRDVMLETYSNLVFVGQQVTKPNLILKLEVEECPA
355	ZN823_HUMAN	DSVAFEDVAVNFTQEEWALLGPSQKSLYRNVMQETIRNLDCIEMK WEDQNIGDQCQNAKRNLRSHTCEIKDDSQCGETFG
356	ZN177_HUMAN	AAGWLTTWSQNSVTFQEVAVDFSQEEWALLDPAQKNLYKDVMLEN FRNLASVGYQLCRHSLISKVDQEQLKTDERGILQG
357	ZN197_HUMAN	ENPRNQLMALMLLTAQPQELVMFEEVSVCFTSEEWACLGPIQRAL YWDVMLENYGNVTSLEWETMTENEEVTSKPSSSQR
358	ZN717_HUMAN	LETYNSLVSLQELVSFEEVAVHFTWEEWQDLDDAQRTLYRDVMLE TYSSLVSLGHCITKPEMIFKLEQGAEPWIVEETPN
359	ZN669_HUMAN	RHFRRPEPCREPLASPIQDSVAFEDVAVNFTQEEWALLDSSQKNL YREVMQETCRNLASVGSQWKDQNIEDHFEKPGKDI
360	ZN256_HUMAN	AAAELTAPAQGIVTFEDVAVYFSWKEWGLLDEAQKCLYHDVMLEN LTLTTSLGGSGAGDEEAPYQQSTSPQRVSQVRIPK
361	ZN251_HUMAN	AATFQLPGHQEMPLTFQDVAVYFSQAEGRQLGPQQRALYRDVMLE NYGNVASLGFPVPKPELISQLEQGKELWVLNLLGA
362	CBX4_HUMAN	RSEAGEPPSSLQVKPETPASAAVAVAAAAAPTTTAEKPPAEAQDE PAESLSEFKPFFGNIIITDVTANCLTVTFKEYVTV
363	PCGF2_HUMAN	HRTTRIKITELNPHLMCALCGGYFIDATTIVECLHSFCKTCIVRY LETNKYCPMCDVQVHKTRPLLSIRSDKTLQDIVYK
364	CDY2_HUMAN	ASQEFEVEAIVDKRQDKNGNTQYLVRWKGYDKQDDTWEPEQHLMN CEKCVHDFNRRQTEKQKKLTWTTTSRIFSNNARRR
365	CDYL2_HUMAN	ASGDLYEVERIVDKRKNKKGKWEYLIRWKGYGSTEDTWEPEHHLL HCEEFIDEFNGLHMSKDKRIKSGKQSSTSKLLRDS
366	HERC2_HUMAN	TLIRKADLENHNKDGGFWTVIDGKVYDIKDFQTQSLTGNSILAQF AGEDPVVALEAALQFEDTRESMHAFCVGQYLEPDQ
367	ZN562_HUMAN	EKTKIGTMVEDHRSNSYQDSVTFDDVAVEFTPEEWALLDTTQKYL YRDVMLENYMNLASVDFFFCLTSEWEIQPRTKRSS
368	ZN461_HUMAN	AHELVMFRDVAIDVSQEEWECLNPAQRNLYKEVMLENYSNLVSLG LSVSKPAVISSLEQGKEPWMVVREETGRWCPGTWK

[0567]

SEQ	描述	序列
369	Z324A_HUMAN	AFEDVAVYFSQEEWGLLDTAQRALYRRVMLDNFALVASLGLSTSR PRVVIQLERGEEPWVPSGTDTTLSRTTYRRRNPGS
370	ZN766_HUMAN	AQLRRGHLTFRDVAIEFSQEEWKCLDPVQKALYRDVMLENYRNLV SLGICLPDLSIISMMKQRTEPWTVENEMKVAKNPD
371	ID2_HUMAN	SDHSLGISRSKTPVDDPMSLLYNMNDCYSKLKELVPSIPQNKKVS KMEILQHVIDYILDLQIALDSHPTIVSLHHQRPGQ
372	TOX_HUMAN	KDPNEPQKPVSAYALFFRDTQAAIKGQNPNATFGEVSKIVASMWD GLGEEQKQVYKKKTEAAKKEYLKQLAAYRASLVSK
373	ZN274_HUMAN	QEEKQEDAAICPVTVLPEEPVTFQDVAVDFSREEWGLLGPTQRTE YRDVMLETFGHLVSVGWETTLENKELAPNSDIPEE
374	SCMH1_HUMAN	DASRLSGRDPSSWTVEDVMQFVREADPQLGPHADLFRKHEIDGKA LLLLRSDMMMKYMGLKLGPALKLSYHIDRLKQGKF
375	ZN214_HUMAN	AVTFEDVTIIFTWEEWKFLDSSQKRLYREVMWENYTNVMSVENWN ESYKSQEEKFRYLEYENFSYWQGWWNAGAQMYENQ
376	CBX7_HUMAN	ELSAIGEQVFAVESIRKKRVRKGKVEYLVKWKGWPPKYSTWEPEE HILDPRLVMAYEEKEERDRASGYRKRGPKPKRLLL
377	ID1_HUMAN	GGAGARLPALLDEQQVNVLLYDMNGCYSRLKELVPTLPQNRKVSK VEILQHVIDYIRDLQLELNSESEVGTPGGRGLPVR
378	CREM_HUMAN	VVMAASPGSLHSPQQLAEEATRKRELRLMKNREAAKECRRRKKEY VKCLESRVAVLEVQNKKLIEELETLKDICSPKTDY
379	SCX_HUMAN	GGGPGGRPGREPRQRHTANARERDRTNSVNTAFTALRTLIPTEPA DRKLSKIETLRLASSYISHLGNVLLAGEACGDGQP
380	ASCL1_HUMAN	SGFGYSLPQQQPAAVARRNERERNRVKLVNLGFATLREHVPNGAA NKKMSKVETLRSAVEYIRALQQLLDEHDAVSAAFQ
381	ZN764_HUMAN	APLPPRDPNGAGPEWREPGAVSFADVAVYFCREEWGCLRPAQRAL YRDVMRETYGHLSALGIGGNKPALISWVEEEAELW
382	SCML2_HUMAN	KQGFSKDPSTWSVDEVIQFMKHTDPQISGPLADLFRQHEIDGKAL FLLKSDVMMKYMGLKLGPALKLCYYIEKLKEGKYS
383	TWST1_HUMAN	SGGGSPQSYEELQTQRVMANVRERQRTQSLNEAFAALRKIIPTLP SDKLSKIQTLKLAARYIDFLYQVLQSDELDSKMAS
384	CREB1_HUMAN	IAPGVVMASSPALPTQPAEEAARKREVRLMKNREAARECRRKKKE YVKCLENRVAVLENQNKTLIEELKALKDLYCHKSD
385	TERF1_HUMAN	SRIPVSKSQPVTPEKHRARKRQAWLWEEDKNLRSGVRKYGEGNWS KILLHYKFNNRTSVMLKDRWRTMKKLKLISSDSED
386	ID3_HUMAN	SLAIARGRGKGPAAEEPLSLLDDMNHCYSRLRELVPGVPRGTQLS QVEILQRVIDYILDLQVVLAEPAPGPPDGPHLPIQ
387	CBX8_HUMAN	GSGPPSSGGGLYRDMGAQGGRPSLIARIPVARILGDPEEESWSPS LTNLEKVVVTDVTSNFLTVTIKESNTDQGFFKEKR
388	CBX4_HUMAN	ELPAVGEHVFAVESIEKKRIRKGRVEYLVKWRGWSPKYNTWEPEE NILDPRLLIAFQNRERQEQLMGYRKRGPKPKPLVV
389	GSX1_HUMAN	VDSSSNQLPSSKRMRTAFTSTQLLELEREFASNMYLSRLRRIEIA TYLNLSEKQVKIWFQNRRVKHKKEGKGSNHRGGGG
390	NKX22_HUMAN	TPGGGGDAGKKRKRRVLFSKAQTYELERRFRQQRYLSAPEREHLA SLIRLTPTQVKIWFQNHRYKMKRARAEKGMEVTPL
391	ATF1_HUMAN	QTVVMTSPVTLTSQTTKTDDPQLKREIRLMKNREAARECRRKKKE YVKCLENRVAVLENQNKTLIEELKTLKDLYSNKSV
392	TWST2_HUMAN	KGSPSAQSFEELQSQRILANVRERQRTQSLNEAFAALRKIIPTLP SDKLSKIQTLKLAARYIDFLYQVLQSDEMDNKMTS
393	ZNF17_HUMAN	NLTEDYMVFEDVAIHFSQEEWGILNDVQRHLHSDVMLENFALLSS VGCWHGAKDEEAPSKQCVSVGVSQVTTLKPALSTQ
394	TOX3_HUMAN	KDPNEPQKPVSAYALFFRDTQAAIKGQNPNATFGEVSKIVASMWD SLGEEQKQVYKRKTEAAKKEYLKALAAYRASLVSK
395	TOX4_HUMAN	KDPNEPQKPVSAYALFFRDTQAAIKGQNPNATFGEVSKIVASMWD SLGEEQKQVYKRKTEAAKKEYLKALAAYKDNQECQ

[0568]

SEQ	描述	序列
396	ZMYM3_HUMAN	LDGSTWDFCSEDCKSKYLLWYCKAARCHACKRQGKLLETIHWRGQ IRHFCNQQCLLRFYSQQNQPNLDTQSGPESLLNSQ
397	I2BP1_HUMAN	ASVQASRRQWCYLCDLPKMPWAMVWDFSEAVCRGCVNFEGADRIE LLIDAARQLKRSHVLPEGRSPGPPALKHPATKDLA
398	RHXF1_HUMAN	MEGPQPENMQPRTRRTKFTLLQVEELESVFRHTQYPDVPTRRELA ENLGVTEDKVRVWFKNKRARCRRHQRELMLANELR
399	SSX2_HUMAN	PKIMPKKPAEEGNDSEEVPEASGPQNDGKELCPPGKPTTSEKIHE RSGPKRGEHAWTHRLRERKQLVIYEEISDPEEDDE
400	I2BPL_HUMAN	SAAQVSSSRRQSCYLCDLPRMPWAMIWDFSEPVCRGCVNYEGADR IEFVIETARQLKRAHGCFQDGRSPGPPPPVGVKTV
401	ZN680_HUMAN	PGPPGSLEMGPLTFRDVAIEFSLEEWQCLDTAQRNLYRKVMFENY RNLVFLGIAVSKPHLITCLEQGKEPWNRKRQEMVA
402	CBX1_HUMAN	NKKKVEEVLEEEEEEYVVEKVLDRRVVKGKVEYLLKWKGFSDEDN TWEPEENLDCPDLIAEFLQSQKTAHETDKSEGGKR
403	TRI68_HUMAN	LANVVEKVRLLRLHPGMGLKGDLCERHGEKLKMFCKEDVLIMCEA CSQSPEHEAHSVVPMEDVAWEYKWELHEALEHLKK
404	HXA13_HUMAN	VVSHPSDASSYRRGRKKRVPYTKVQLKELEREYATNKFITKDKRR RISATTNLSERQVTIWFQNRRVKEKKVINKLKTTS
405	PHC3_HUMAN	ENSDLLPVAQTEPSIWTVDDVWAFIHSLPGCQDIADEFRAQEIDG QALLLLKEDHLMSAMNIKLGPALKICARINSLKES
406	TCF24_HUMAN	AGPGGGSRSGSGRPAAANAARERSRVQTLRHAFLELQRTLPSVPP DTKLSKLDVLLLATTYIAHLTRSLQDDAEAPADAG
407	CBX3_HUMAN	QNGKSKKVEEAEPEEFVVEKVLDRRVVNGKVEYFLKWKGFTDADN TWEPEENLDCPELIEAFLNSQKAGKEKDGTKRKSL
408	HXB13_HUMAN	QHPPDACAFRRGRKKRIPYSKGQLRELEREYAANKFITKDKRRKI SAATSLSERQITIWFQNRRVKEKKVLAKVKNSATP
409	HEY1_HUMAN	SMSPTTSSQILARKRRRGIIEKRRRDRINNSLSELRRLVPSAFEK QGSAKLEKAEILQMTVDHLKMLHTAGGKGYFDAHA
410	PHC2_HUMAN	LVGMGHHFLPSEPTKWNVEDVYEFIRSLPGCQEIAEEFRAQEIDG QALLLLKEDHLMSAMNIKLGPALKIYARISMLKDS
411	ZNF81_HUMAN	PANEDAPQPGEHGSACEVSVSFEDVTVDFSREEWQQLDSTQRRLY QDVMLENYSHLLSVGFEVPKPEVIFKLEQGEGPWT
412	FIGLA_HUMAN	GYSSTENLQLVLERRRVANAKERERIKNLNRGFARLKALVPFLPQ SRKPSKVDILKGATEYIQVLSDLLEGAKDSKKQDP
413	SAM11_HUMAN	EEAPAPEDVTKWTVDDVCSFVGGLSGCGEYTRVFREQGIDGETLP LLTEEHLLTNMGLKLGPALKIRAQVARRLGRVFYV
414	KMT2B_HUMAN	GGTLAHTPRRSLPSHHGKKMRMARCGHCRGCLRVQDCGSCVNCLD KPKFGGPNTKKQCCVYRKCDKIEARKMERLAKKGR
415	HEY2_HUMAN	LNSPTTTSQIMARKKRRGIIEKRRRDRINNSLSELRRLVPTAFEK QGSAKLEKAEILQMTVDHLKMLQATGGKGYFDAHA
416	JDP2_HUMAN	QPVKSELDEEEERRKRRREKNKVAAARCRNKKKERTEFLQRESER LELMNAELKTQIEELKQERQQLILMLNRHRPTCIV
417	HXC13_HUMAN	LQPEVSSYRRGRKKRVPYTKVQLKELEKEYAASKFITKEKRRRIS ATTNLSERQVTIWFQNRRVKEKKVVSKSKAPHLHS
418	ASCL4_HUMAN	LPVPLDSAFEPAFLRKRNERERQRVRCVNEGYARLRDHLPRELAD KRLSKVETLRAAIDYIKHLQELLERQAWGLEGAAG
419	HHEX_HUMAN	SPFLQRPLHKRKGGQVRFSNDQTIELEKKFETQKYLSPPERKRLA KMLQLSERQVKTWFQNRRAKWRRLKQENPQSNKKE
420	HERC2_HUMAN	IAIATGSLHCVCCTEDGEVYTWGDNDEGQLGDGTTNAIQRPRLVA ALQGKKVNRVACGSAHTLAWSTSKPASAGKLPAQV
421	GSX2_HUMAN	GGSDASQVPNGKRMRTAFTSTQLLELEREFSSNMYLSRLRRIEIA TYLNLSEKQVKIWFQNRRVKHKKEGKGTQRNSHAG
422	BIN1_HUMAN	RLDLPPGFMFKVQAQHDYTATDTDELQLKAGDVVLVIPFQNPEEQ DEGWLMGVKESDWNQHKELEKCRGVFPENFTERVP

[0569]

SEQ	描述	序列
423	ETV7_HUMAN	GICKLPGRLRIQPALWSREDVLHWLRWAEQEYSLPCTAEHGFEMN GRALCILTKDDFRHRAPSSGDVLYELLQYIKTQRR
424	ASCL3_HUMAN	PNYRGCEYSYGPAFTRKRNERERQRVKCVNEGYAQLRHHLPEEYL EKRLSKVETLRAAIKYINYLQSLLYPDKAETKNNP
425	PHC1_HUMAN	LHGINPVFLSSNPSRWSVEEVYEFIASLQGCQEIAEEFRSQEIDG QALLLLKEEHLMSAMNIKLGPALKICAKINVLKET
426	OTP_HUMAN	QAGQQQGQQKQKRHRTRFTPAQLNELERSFAKTHYPDIFMREELA LRIGLTESRVQVWFQNRRAKWKKRKKTTNVFRAPG
427	I2BP2_HUMAN	AAAVAVAAASRRQSCYLCDLPRMPWAMIWDFTEPVCRGCVNYEGA DRVEFVIETARQLKRAHGCFPEGRSPPGAAASAAA
428	VGLL2_HUMAN	FSSQTPASIKEEEGSPEKERPPEAEYINSRCVLFTYFQGDISSVV DEHFSRALSQPSSYSPSCTSSKAPRSSGPWRDCSF
429	HXA11_HUMAN	DKAGGSSGQRTRKKRCPYTKYQIRELEREFFFSVYINKEKRLQLS RMLNLTDRQVKIWFQNRRMKEKKINRDRLQYYSAN
430	PDLI4_HUMAN	GAPLSGLQGLPECTRCGHGIVGTIVKARDKLYHPECFMCSDCGLN LKQRGYFFLDERLYCESHAKARVKPPEGYDVVAVY
431	ASCL2_HUMAN	RRPATAETGGGAAAVARRNERERNRVKLVNLGFQALRQHVPHGGA SKKLSKVETLRSAVEYIRALQRLLAEHDAVRNALA
432	CDX4_HUMAN	TVQVTGKTRTKEKYRVVYTDHQRLELEKEFHCNRYITIQRKSELA VNLGLSERQVKIWFQNRRAKERKMIKKKISQFENS
433	ZN860_HUMAN	EEAAQKRKEKEPGMALPQGHLTFRDVAIEFSLEEWKCLDPTQRAL YRAMMLENYRNLHSVDISSKCMMKKFSSTAQGNTE
434	LMBL4_HUMAN	DIRASQVARWTVDEVAEFVQSLLGCEEHAKCFKKEQIDGKAFLLL TQTDIVKVMKIKLGPALKIYNSILMFRHSQELPEE
435	PDIP3_HUMAN	LSPLEGTKMTVNNLHPRVTEEDIVELFCVCGALKRARLVHPGVAE VVFVKKDDAITAYKKYNNRCLDGQPMKCNLHMNGN
436	NKX25_HUMAN	DNAERPRARRRRKPRVLFSQAQVYELERRFKQQRYLSAPERDQLA SVLKLTSTQVKIWFQNRRYKCKRQRQDQTLELVGL
437	CEBPB_HUMAN	SQVKSKAKKTVDKHSDEYKIRRERNNIAVRKSRDKAKMRNLETQH KVLELTAENERLQKKVEQLSRELSTLRNLFKQLPE
438	ISL1_HUMAN	KRDYIRLYGIKCAKCSIGFSKNDFVMRARSKVYHIECFRCVACSR QLIPGDEFALREDGLFCRADHDVVERASLGAGDPL
439	CDX2_HUMAN	SLGSQVKTRTKDKYRVVYTDHQRLELEKEFHYSRYITIRRKAELA ATLGLSERQVKIWFQNRRAKERKINKKKLQQQQQQ
440	PROP1_HUMAN	QGGQRGRPHSRRRHRTTFSPVQLEQLESAFGRNQYPDIWARESLA RDTGLSEARIQVWFQNRRAKQRKQERSLLQPLAHL
441	SIN3B_HUMAN	DALTYLDQVKIRFGSDPATYNGFLEIMKEFKSQSIDTPGVIRRVS QLFHEHPDLIVGFNAFLPLGYRIDIPKNGKLNIQS
442	SMBT1_HUMAN	RLHLDSNPLKWSVADVVRFIRSTDCAPLARIFLDQEIDGQALLLL TLPTVQECMDLKLGPAIKLCHHIERIKFAFYEQFA
443	HXC11_HUMAN	AKGAAPNAPRTRKKRCPYSKFQIRELEREFFFNVYINKEKRLQLS RMLNLTDRQVKIWFQNRRMKEKKLSRDRLQYFSGN
444	HXC10_HUMAN	TTGNWLTAKSGRKKRCPYTKHQTLELEKEFLFNMYLTRERRLEIS KTINLTDRQVKIWFQNRRMKLKKMNRENRIRELTS
445	PRS6A_HUMAN	YLVSNVIELLDVDPNDQEEDGANIDLDSQRKGKCAVIKTSTRQTY FLPVIGLVDAEKLKPGDLVGVNKDSYLILETLPTE
446	VSX1_HUMAN	KASPTLGKRKKRRHRTVFTAHQLEELEKAFSEAHYPDVYAREMLA VKTELPEDRIQVWFQNRRAKWRKREKRWGGSSVMA
447	NKX23_HUMAN	EESERPKPRSRRKPRVLFSQAQVFELERRFKQQRYLSAPEREHLA SSLKLTSTQVKIWFQNRRYKCKRQRQDKSLELGAH
448	MTG16_HUMAN	VVPGSRQEEVIDHKLTEREWAEEWKHLNNLLNCIMDMVEKTRRSL TVLRRCQEADREELNHWARRYSDAEDTKKGPAPAA
449	HMX3_HUMAN	ESPEKKPACRKKKTRTVFSRSQVFQLESTFDMKRYLSSSERAGLA ASLHLTETQVKIWFQNRRNKWKRQLAAELEAANLS

[0570]

SEQ	描述	序列
450	HMX1_HUMAN	RGGVGVGGGRKKKTRTVFSRSQVFQLESTFDLKRYLSSAERAGLA ASLQLTETQVKIWFQNRRNKWKRQLAAELEAASLS
451	KIF22_HUMAN	ELLAHGRQKILDLLNEGSARDLRSLQRIGPKKAQLIVGWRELHGP FSQVEDLERVEGITGKQMESFLKANILGLAAGQRC
452	CSTF2_HUMAN	ESPYGETISPEDAPESISKAVASLPPEQMFELMKQMKLCVQNSPQ EARNMLLQNPQLAYALLQAQVVMRIVDPEIALKIL
453	CEBPE_HUMAN	AGPLHKGKKAVNKDSLEYRLRRERNNIAVRKSRDKAKRRILETQQ KVLEYMAENERLRSRVEQLTQELDTLRNLFRQIPE
454	DLX2_HUMAN	IRIVNGKPKKVRKPRTIYSSFQLAALQRRFQKTQYLALPERAELA ASLGLTQTQVKIWFQNRRSKFKKMWKSGEIPSEQH
455	ZMYM3_HUMAN	TVYQFCSPSCWTKFQRTSPEGGIHLSCHYCHSLFSGKPEVLDWQD QVFQFCCRDCCEDFKRLRGVVSQCEHCRQEKLLHE
456	PPARG_HUMAN	TMVDTEMPFWPTNFGISSVDLSVMEDHSHSFDIKPFTTVDFSSIS TPHYEDIPFTRTDPVVADYKYDLKLQEYQSAIKVE
457	PRIC1_HUMAN	GRHHAELLKPRCSACDEIIFADECTEAEGRHWHMKHFCCLECETV LGGQRYIMKDGRPFCCGCFESLYAEYCETCGEHIG
458	UNC4_HUMAN	DPDKESPGCKRRRTRTNFTGWQLEELEKAFNESHYPDVFMREALA LRLDLVESRVQVWFQNRRAKWRKKENTKKGPGRPA
459	BARX2_HUMAN	TEQPTPRQKKPRRSRTIFTELQLMGLEKKFQKQKYLSTPDRLDLA QSLGLTQLQVKTWYQNRRMKWKKMVLKGGQEAPTK
460	ALX3_HUMAN	SMELAKNKSKKRRNRTTFSTFQLEELEKVFQKTHYPDVYAREQLA LRTDLTEARVQVWFQNRRAKWRKRERYGKIQEGRN
461	TCF15_HUMAN	GGGGGAGPVVVVRQRQAANARERDRTQSVNTAFTALRTLIPTEPV DRKLSKIETVRLASSYIAHLANVLLLGDSADDGQP
462	TERA_HUMAN	IDDTVEGITGNLFEVYLKPYFLEAYRPIRKGDIFLVRGGMRAVEF KVVETDPSPYCIVAPDTVIHCEGEPIKREDEEESL
463	VSX2_HUMAN	SALNQTKKRKKRRHRTIFTSYQLEELEKAFNEAHYPDVYAREMLA MKTELPEDRIQVWFQNRRAKWRKREKCWGRSSVMA
464	HXD12_HUMAN	DGLPWGAAPGRARKKRKPYTKQQIAELENEFLVNEFINRQKRKEL SNRLNLSDQQVKIWFQNRRMKKRVVLREQALALY
465	CDX1_HUMAN	GGGGSGKTRTKDKYRVVYTDHQRLELEKEFHYSRYITIRRKSELA ANLGLTERQVKIWFQNRRAKERKVNKKKQQQQQPP
466	TCF23_HUMAN	TRAGGLALGRSEASPENAARERSRVRTLRQAFLALQAALPAVPPD TKLSKLDVLVLAASYIAHLTRTLGHELPGPAWPPF
467	ALX1_HUMAN	KCDSNVSSSKKRRHRTTFTSLQLEELEKVFQKTHYPDVYVREQLA LRTELTEARVQVWFQNRRAKWRKRERYGQIQQAKS
468	HXA10_HUMAN	NAANWLTAKSGRKKRCPYTKHQTLELEKEFLFNMYLTRERRLEIS RSVHLTDRQVKIWFQNRRMKLKKMNRENRIRELTA
469	RX_HUMAN	LSEEEQPKKKHRRNRTTFTTYQLHELERAFEKSHYPDVYSREELA GKVNLPEVRVQVWFQNRRAKWRRQEKLEVSSMKLQ
470	CXXC5_HUMAN	HMAGLAEYPMQGELASAISSGKKKRKRCGMCAPCRRRINCEQCSS CRNRKTGHQICKFRKCEELKKKPSAALEKVMLPTG
471	SCML1_HUMAN	SITKHPSTWSVEAVVLFLKQTDPLALCPLVDLFRSHEIDGKALLL LTSDVLLKHLGVKLGTAVKLCYYIDRLKQGKCFEN
472	NFIL3_HUMAN	ACRRKREFIPDEKKDAMYWEKRRKNNEAAKRSREKRRLNDLVLEN KLIALGEENATLKAELLSLKLKFGLISSTAYAQEI
473	DLX6_HUMAN	EIRFNGKGKKIRKPRTIYSSLQLQALNHRFQQTQYLALPERAELA ASLGLTQTQVKIWFQNKRSKFKKLLKQGSNPHESD
474	MTG8_HUMAN	GLHGTRQEEMIDHRLTDREWAEEWKHLDHLLNCIMDMVEKTRRSL TVLRRCQEADREELNYWIRRYSDAEDLKKGGGSSS
475	CBX8_HUMAN	ELSAVGERVFAAEALLKRRIRKGRMEYLVKWKGWSQKYSTWEPEE NILDARLLAAFEEREREMELYGPKKRGPKPKTFLL
476	CEBPD_HUMAN	AREKSAGKRGPDRGSPEYRQRRERNNIAVRKSRDKAKRRNQEMQQ KLVELSAENEKLHQRVEQLTRDLAGLRQFFKQLPS

[0571]

SEQ	描述	序列
477	SEC13_HUMAN	SGGCDNLIKLWKEEEDGQWKEEQKLEAHSDWVRDVAWAPSIGLPT STIASCSQDGRVFIWTCDDASSNTWSPKLLHKFND
478	FIP1_HUMAN	VKGVDLDAPGSINGVPLLEVDLDSFEDKPWRKPGADLSDYFNYGF NEDTWKAYCEKQKRIRMGLEVIPVTSTTNKITAED
479	ALX4_HUMAN	KADSESNKGKKRRNRTTFTSYQLEELEKVFQKTHYPDVYAREQLA MRTDLTEARVQVWFQNRRAKWRKRERFGQMQQVRT
480	LHX3_HUMAN	TAKQREAEATAKRPRTTITAKQLETLKSAYNTSPKPARHVREQLS SETGLDMRVVQVWFQNRRAKEKRLKKDAGRQRWGQ
481	PRIC2_HUMAN	GRHHAECLKPRCAACDEIIFADECTEAEGRHWHMKHFCCFECETV LGGQRYIMKEGRPYCCHCFESLYAEYCDTCAQHIG
482	MAGI3_HUMAN	IIGGDRPDEFLQVKNVLKDGPAAQDGKIAPGDVIVDINGNCVLGH THADVVQMFQLVPVNQYVNLTLCRGYPLPDDSEDP
483	NELL1_HUMAN	CCPECDTRVTSQCLDQNGHKLYRSGDNWTHSCQQCRCLEGEVDCW PLTCPNLSCEYTAILEGECCPRCVSDPCLADNITY
484	PRRX1_HUMAN	LNSEEKKKRKQRRNRTTFNSSQLQALERVFERTHYPDAFVREDLA RRVNLTEARVQVWFQNRRAKFRRNERAMLANKNAS
485	MTG8R_HUMAN	GLNGGYQDELVDHRLTEREWADEWKHLDHALNCIMEMVEKTRRSM AVLRRCQESDREELNYWKRRYNENTELRKTGTELV
486	RAX2_HUMAN	GPGEEAPKKKHRRNRTTFTTYQLHQLERAFEASHYPDVYSREELA AKVHLPEVRVQVWFQNRRAKWRRQERLESGSGAVA
487	DLX3_HUMAN	VRMVNGKPKKVRKPRTIYSSYQLAALQRRFQKAQYLALPERAELA AQLGLTQTQVKIWFQNRRSKFKKLYKNGEVPLEHS
488	DLX1_HUMAN	EVRFNGKGKKIRKPRTIYSSLQLQALNRRFQQTQYLALPERAELA ASLGLTQTQVKIWFQNKRSKFKKLMKQGGAALEGS
489	NKX26_HUMAN	GRSEQPKARQRRKPRVLFSQAQVLALERRFKQQRYLSAPEREHLA SALQLTSTQVKIWFQNRRYKCKRQRQDKSLELAGH
490	NAB1_HUMAN	LPRTLGELQLYRILQKANLLSYFDAFIQQGGDDVQQLCEAGEEEF LEIMALVGMASKPLHVRRLQKALRDWVTNPGLFNQ
491	SAMD7_HUMAN	NLSLDEDIQKWTVDDVHSFIRSLPGCSDYAQVFKDHAIDGETLPL LTEEHLRGTMGLKLGPALKIQSQVSQHVGSMFYKK
492	PITX3_HUMAN	SPEDGSLKKKQRRQRTHFTSQQLQELEATFQRNRYPDMSTREEIA VWTNLTEARVRVWFKNRRAKWRKRERSQQAELCKG
493	WDR5_HUMAN	SNLLVSASDDKTLKIWDVSSGKCLKTLKGHSNYVFCCNFNPQSNL IVSGSFDESVRIWDVKTGKCLKTLPAHSDPVSAVH
494	MEOX2_HUMAN	GNYKSEVNSKPRKERTAFTKEQIRELEAEFAHHNYLTRLRRYEIA VNLDLTERQVKVWFQNRRMKWKRVKGGQQGAAARE
495	NAB2_HUMAN	LPRTLGELQLYRVLQRANLLSYYETFIQQGGDDVQQLCEAGEEEF LEIMALVGMATKPLHVRRLQKALREWATNPGLFSQ
496	DHX8_HUMAN	PEEPTIGDIYNGKVTSIMQFGCFVQLEGLRKRWEGLVHISELRRE GRVANVADVVSKGQRVKVKVLSFTGTKTSLSMKDV
497	FOXA2_HUMAN	YAFNHPFSINNLMSSEQQHHHSHHHHQPHKMDLKAYEQVMHYPGY GSPMPGSLAMGPVTNKTGLDASPLAADTSYYQGVY
498	CBX6_HUMAN	TAAAGPAPPTAPEPAGASSEPEAGDWRPEMSPCSNVVVTDVTSNL LTVTIKEFCNPEDFEKVAAGVAGAAGGGGSIGASK
499	EMX2_HUMAN	FLLHNALARKPKRIRTAFSPSQLLRLEHAFEKNHYVVGAERKQLA HSLSLTETQVKVWFQNRRTKFKRQKLEEEGSDSQQ
500	CPSF6_HUMAN	KRIALYIGNLTWWTTDEDLTEAVHSLGVNDILEIKFFENRANGQS KGFALVGVGSEASSKKLMDLLPKRELHGQNPVVTP
501	HXC12_HUMAN	SGAPWYPINSRSRKKRKPYSKLQLAELEGEFLVNEFITRQRRREL SDRLNLSDQQVKIWFQNRRMKKKRLLLREQALSFF
502	KDM4B_HUMAN	SDNLYPESITSRDCVQLGPPSEGELVELRWTDGNLYKAKFISSVT SHIYQVEFEDGSQLTVKRGDIFTLEEELPKRVRSR
503	LMBL3_HUMAN	GIPASKVSKWSTDEVSEFIQSLPGCEEHGKVFKDEQIDGEAFLLM TQTDIVKIMSIKLGPALKIFNSILMFKAAEKNSHN

[0572]

SEQ	描述	序列
504	PHX2A_HUMAN	EPSGLHEKRKQRRIRTTFTSAQLKELERVFAETHYPDIYTREELA LKIDLTEARVQVWFQNRRAKFRKQERAASAKGAAG
505	EMX1_HUMAN	LLLHGPFARKPKRIRTAFSPSQLLRLERAFEKNHYVVGAERKQLA GSLSLSETQVKVWFQNRRTKYKRQKLEEEGPESEQ
506	NC2B_HUMAN	SSGNDDDLTIPRAAINKMIKETLPNVRVANDARELVVNCCTEFIH LISSEANEICNKSEKKTISPEHVIQALESLGFGSY
507	DLX4_HUMAN	ERRPQAPAKKLRKPRTIYSSLQLQHLNQRFQHTQYLALPERAQLA AQLGLTQTQVKIWFQNKRSKYKKLLKQNSGGQEGD
508	SRY_HUMAN	NVQDRVKRPMNAFIVWSRDQRRKMALENPRMRNSEISKQLGYQWK MLTEAEKWPFFQEAQKLQAMHREKYPNYKYRPRRK
509	ZN777_HUMAN	EITRLAVWAAVQAVERKLEAQAMRLLTLEGRTGTNEKKIADCEKT AVEFANHLESKWVVLGTLLQEYGLLQRRLENMENL
510	NELL1_HUMAN	CEKDIDECSEGIIECHNHSRCVNLPGWYHCECRSGFHDDGTYSLS GESCIDIDECALRTHTCWNDSACINLAGGFDCLCP
511	ZN398_HUMAN	AAISLWTVVAAVQAIERKVEIHSRRLLHLEGRTGTAEKKLASCEK TVTELGNQLEGKWAVLGTLLQEYGLLQRRLENLEN
512	GATA3_HUMAN	GQNRPLIKPKRRLSAARRAGTSCANCQTTTTTLWRRNANGDPVCN ACGLYYKLHNINRPLTMKKEGIQTRNRKMSSKSKK
513	BSH_HUMAN	HAELPGKHCRRRKARTVFSDSQLSGLEKRFEIQRYLSTPERVELA TALSLSETQVKTWFQNRRMKHKKQLRKSQDEPKAP
514	SF3B4_HUMAN	QDATVYVGGLDEKVSEPLLWELFLQAGPVVNTHMPKDRVTGQHQG YGFVEFLSEEDADYAIKIMNMIKLYGKPIRVNKAS
515	TEAD1_HUMAN	PIDNDAEGVWSPDIEQSFQEALAIYPPCGRRKIILSDEGKMYGRN ELIARYIKLRTGKTRTRKQVSSHIQVLARRKSRDF
516	TEAD3_HUMAN	GLDNDAEGVWSPDIEQSFQEALAIYPPCGRRKIILSDEGKMYGRN ELIARYIKLRTGKTRTRKQVSSHIQVLARKKVREY
517	RGAP1_HUMAN	DSVGTPQSNGGMRLHDFVSKTVIKPESCVPCGKRIKFGKLSLKCR DCRVVSHPECRDRCPLPCIPTLIGTPVKIGEGMLA
518	PHF1_HUMAN	SAPHSMTASSSSVSSPSPGLPRRSAPPSPLCRSLSPGTGGGVRGG VGYLSRGDPVRVLARRVRPDGSVQYLVEWGGGGIF
519	FOXA1_HUMAN	GDPHYSFNHPFSINNLMSSSEQQHKLDFKAYEQALQYSPYGSTLP ASLPLGSASVTTRSPIEPSALEPAYYQGVYSRPVL
520	GATA2_HUMAN	GQNRPLIKPKRRLSAARRAGTCCANCQTTTTTLWRRNANGDPVCN ACGLYYKLHNVNRPLTMKKEGIQTRNRKMSNKSKK
521	FOXO3_HUMAN	DSLSGSSLYSTSANLPVMGHEKFPSDLDLDMFNGSLECDMESIIR SELMDADGLDFNFDSLISTQNVVGLNVGNFTGAKQ
522	ZN212_HUMAN	TEISLWTVVAAIQAVEKKMESQAARLQSLEGRTGTAEKKLADCEK MAVEFGNQLEGKWAVLGTLLQEYGLLQRRLENVEN
523	IRX4_HUMAN	MDSGTRRKNATRETTSTLKAWLQEHRKNPYPTKGEKIMLAIITKM TLTQVSTWFANARRRLKKENKMTWPPRNKCADEKR
524	ZBED6_HUMAN	NIEKQIYLPSTRAKTSIVWHFFHVDPQYTWRAICNLCEKSVSRGK PGSHLGTSTLQRHLQARHSPHWTRANKFGVASGEE
525	LHX4_HUMAN	AKQNDDSEAGAKRPRTTITAKQLETLKNAYKNSPKPARHVREQLS SETGLDMRVVQVWFQNRRAKEKRLKKDAGRHRWGQ
526	SIN3A_HUMAN	DALSYLDQVKLQFGSQPQVYNDFLDIMKEFKSQSIDTPGVISRVS QLFKGHPDLIMGFNTFLPPGYKIEVQTNDMVNVTT
527	RBBP7_HUMAN	DDHTVCLWDINAGPKEGKIVDAKAIFTGHSAVVEDVAWHLLHESL FGSVADDQKLMIWDTRSNTTSKPSHLVDAHTAEVN
528	NKX61_HUMAN	GSILLDKDGKRKHTRPTFSGQQIFALEKTFEQTKYLAGPERARLA YSLGMTESQVKVWFQNRRTKWRKKHAAEMATAKKK
529	TRI68_HUMAN	DPTALVEAIVEEVACPICMTFLREPMSIDCGHSFCHSCLSGLWEI PGESQNWGYTCPLCRAPVQPRNLRPNWQLANVVEK
530	R51A1_HUMAN	QSLPKKVSLSSDTTRKPLEIRSPSAESKKPKWVPPAASGGSRSSS SPLVVVSVKSPNQSLRLGLSRLARVKPLHPNATST

[0573]

SEQ	描述	序列
531	MB3L1_HUMAN	AKSSQRKQRDCVNQCKSKPGLSTSIPLRMSSYTFKRPVTRITPHP GNEVRYHQWEESLEKPQQVCWQRRLQGLQAYSSAG
532	DLX5_HUMAN	VRMVNGKPKKVRKPRTIYSSFQLAALQRRFQKTQYLALPERAELA ASLGLTQTQVKIWFQNKRSKIKKIMKNGEMPPEHS
533	NOTC1_HUMAN	LQCNNHACGWDGGDCSLNFNDPWKNCTQSLQCWKYFSDGHCDSQC NSAGCLFDGFDCQRAEGQCNPLYDQYCKDHFSDGH
534	TERF2_HUMAN	ETWVEEDELFQVQAAPDEDSTTNITKKQKWTVEESEWVKAGVQKY GEGNWAAISKNYPFVNRTAVMIKDRWRTMKRLGMN
535	ZN282_HUMAN	AEISLWTVVAAIQAVERKVDAQASQLLNLEGRTGTAEKKLADCEK TAVEFGNHMESKWAVLGTLLQEYGLLQRRLENLEN
536	RGS12_HUMAN	LEKRTLFRLDLVPINRSVGLKAKPTKPVTEVLRPVVARYGLDLSG LLVRLSGEKEPLDLGAPISSLDGQRVVLEEKDPSR
537	ZN840_HUMAN	PNCLSSSMQLPHGGGRHQELVRFRDVAVVFSPEEWDHLTPEQRNL YKDVMLDNCKYLASLGNWTYKAHVMSSLKQGKEPW
538	SPI2B_HUMAN	DDYKEGDLRIMPESSESPPTEREPGGVVDGLIGKHVEYTKEDGSK RIGMVIHQVEAKPSVYFIKFDDDFHIYVYDLVKKS
539	PAX7_HUMAN	SEPDLPLKRKQRRSRTTFTAEQLEELEKAFERTHYPDIYTREELA QRTKLTEARVQVWFSNRRARWRKQAGANQLAAFNH
540	NKX62_HUMAN	AGGVLDKDGKKKHSRPTFSGQQIFALEKTFEQTKYLAGPERARLA YSLGMTESQVKVWFQNRRTKWRKRHAVEMASAKKK
541	ASXL2_HUMAN	DVMSFSVTVTTIPASQAMNPSSHGQTIPVQAFSEENSIEGTPSKC YCRLKAMIMCKGCGAFCHDDCIGPSKLCVSCLVVR
542	FOXO1_HUMAN	GGYSSVSSCNGYGRMGLLHQEKLPSDLDGMFIERLDCDMESIIRN DLMDGDTLDFNFDNVLPNQSFPHSVKTTTHSWVSG
543	GATA3_HUMAN	GGSPTGFGCKSRPKARSSTGRECVNCGATSTPLWRRDGTGHYLCN ACGLYHKMNGQNRPLIKPKRRLSAARRAGTSCANC
544	GATA1_HUMAN	GQNRPLIRPKKRLIVSKRAGTQCTNCQTTTTTLWRRNASGDPVCN ACGLYYKLHQVNRPLTMRKDGIQTRNRKASGKGKK
545	ZMYM5_HUMAN	PVALLRKQNFQPTAQQQLTKPAKITCANCKKPLQKGQTAYQRKGS AHLFCSTTCLSSFSHKRTQNTRSIICKKDASTKKA
546	ZN783_HUMAN	TEITLWTVVAAIQALEKKVDSCLTRLLTLEGRTGTAEKKLADCEK TAVEFGNQLEGKWAVLGTLLQEYGLLQRRLENVEN
547	SPI2B_HUMAN	KKQRGRPSSQPRRNIVGCRISHGWKEGDEPITQWKGTVLDQVPIN PSLYLVKYDGIDCVYGLELHRDERVLSLKILSDRV
548	LRP1_HUMAN	WTCDLDDDCGDRSDESASCAYPTCFPLTQFTCNNGRCININWRCD NDNDCGDNSDEAGCSHSCSSTQFKCNSGRCIPEHW
549	MIXL1_HUMAN	PKGAAAPSASQRRKRTSFSAEQLQLLELVFRRTRYPDIHLRERLA ALTLLPESRIQVWFQNRRAKSRRQSGKSFQPLARP
550	SGT1_HUMAN	KIKYDWYQTESQVVITLMIKNVQKNDVNVEFSEKELSALVKLPSG EDYNLKLELLHPIIPEQSTFKVLSTKIEIKLKKPE
551	LMCD1_HUMAN	DPSKEVEYVCELCKGAAPPDSPVVYSDRAGYNKQWHPTCFVCAKC SEPLVDLIYFWKDGAPWCGRHYCESLRPRCSGCDE
552	CEBPA_HUMAN	GSGAGKAKKSVDKNSNEYRVRRERNNIAVRKSRDKAKQRNVETQQ KVLELTSDNDRLRKRVEQLSRELDTLRGIFRQLPE
553	GATA2_HUMAN	GPASSFTPKQRSKARSCSEGRECVNCGATATPLWRRDGTGHYLCN ACGLYHKMNGQNRPLIKPKRRLSAARRAGTCCANC
554	SOX14_HUMAN	KPSDHIKRPMNAFMVWSRGQRRKMAQENPKMHNSEISKRLGAEWK LLSEAEKRPYIDEAKRLRAQHMKEHPDYKYRPRRK
555	WTIP_HUMAN	LYSGFQQTADKCSVCGHLIMEMILQALGKSYHPGCFRCSVCNECL DGVPFTVDVENNIYCVRDYHTVFAPKCASCARPIL
556	PRP19_HUMAN	HPSQDLVFSASPDATIRIWSVPNASCVQVVRAHESAVTGLSLHAT GDYLLSSSDDQYWAFSDIQTGRVLTKVTDETSGCS
557	CBX6_HUMAN	ELSAVGERVFAAESIIKRRIRKGRIEYLVKWKGWAIKYSTWEPEE NILDSRLIAAFEQKERERELYGPKKRGPKPKTFLL

[0574]

SEQ	描述	序列
558	NKX11_HUMAN	RTGSDSKSGKPRRARTAFTYEQLVALENKFKATRYLSVCERLNLA LSLSLTETQVKIWFQNRRTKWKKQNPGADTSAPTG
559	RBBP4_HUMAN	VWDLSKIGEEQSPEDAEDGPPELLFIHGGHTAKISDFSWNPNEPW VICSVSEDNIMQVWQMAENIYNDEDPEGSVDPEGQ
560	DMRT2_HUMAN	ERCTPAGGGAEPRKLSRTPKCARCRNHGVVSCLKGHKRFCRWRDC QCANCLLVVERQRVMAAQVALRRQQATEDKKGLSG
561	SMCA2_HUMAN	SQPGALIPGDPQAMSQPNRGPSPFSPVQLHQLRAQILAYKMLARG QPLPETLQLAVQGKRTLPGLQQQQQQQQQQQQQQQ
562	ZNF10	MDAKSLTAWSRTLVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLE NYKNLVSLGYQLTKPDVILRLEKGEEPWLVEREIHQETHPDSETA FEIKSSVSSRSIFKDKQSCDIKMEGMARNDLWYLSLEEVWKCRDQ LDKYQENPERHLRQVAFTQKKVLTQERVSESGKYGGNCLLPAQLV LREYFHKRDSHTKSLKHDLVLNGHQDSCASNSNECGQTFCQNIHL IQFARTHTGDKSYKCPDNDNSLTHGSSLGISKGIHREKPYECKEC GKFFSWRSNLTRHQLIHTGEKPYECKECGKSFSRSSHLIGHQKTH TGEEPYECKECGKSFSWFSHLVTHQRTHTGDKLYTCNQCGKSFVH SSRLIRHQRTHTGEKPYECPECGKSFRQSTHLILHQRTHVRVRPY ECNECGKSYSQRSHLVVHHRIHTGLKPFECKDCGKCFSRSSHLYS HQRTHTGEKPYECHDCGKSFSQSSALIVHQRIHTGEKPYECCQCG KAFIRKNDLIKHQRIHVGEETYKCNQCGIIFSQNSPFIVHQIAHT GEQFLTCNQCGTALVNTSNLIGYQTNHIRENAY
563	EED_HUMAN	MSEREVSTAPAGTDMPAAKKQKLSSDENSNPDLSGDENDDAVSIE SGTNTERPDTPTNTPNAPGRKSWGKGKWKSKKCKYSFKCVNSLKE DHNQPLFGVQFNWHSKEGDPLVFATVGSNRVTLYECHSQGEIRLL QSYVDADADENFYTCAWTYDSNTSHPLLAVAGSRGIIRIINPITM QCIKHYVGHGNAINELKFHPRDPNLLLSVSKDHALRLWNIQTDTL VAIFGGVEGHRDEVLSADYDLLGEKIMSCGMDHSLKLWRINSKRM MNAIKESYDYNPNKTNRPFISQKIHFPDFSTRDIHRNYVDCVRWL GDLILSKSCENAIVCWKPGKMEDDIDKIKPSESNVTILGRFDYSQ CDIWYMRFSMDFWQKMLALGNQVGKLYVWDLEVEDPHKAKCTTLT HHKCGAAIRQTSFSRDSSILIAVCDDASIWRWDRLR
564	RCOR1_HUMAN	MPAMVEKGPEVSGKRRGRNNAAASASAAAASAAASAACASPAATA ASGAAASSASAAAASAAAAPNNGQNKSLAAAAPNGNSSSNSWEEG SSGSSSDEEHGGGGMRVGPQYQAVVPDFDPAKLARRSQERDNLGM LVWSPNQNLSEAKLDEYIAIAKEKHGYNMEQALGMLFWHKHNIEK SLADLPNFTPFPDEWTVEDKVLFEQAFSFHGKTFHRIQQMLPDKS IASLVKFYYSWKKTRTKTSVMDRHARKQKREREESEDELEEANGN NPIDIEVDQNKESKKEVPPTETVPQVKKEKHSTQAKNRAKRKPPK GMFLSQEDVEAVSANATAATTVLRQLDMELVSVKRQIQNIKQTNS ALKEKLDGGIEPYRLPEVIQKCNARWTTEEQLLAVQAIRKYGRDF QAISDVIGNKSVVQVKNFFVNYRRRFNIDEVLQEWEAEHGKEETN GPSNQKPVKSPDNSIKMPEEEDEAPVLDVRYASAS
565	KOX1/ZNF10 KRAB 1	TGRTLVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLENYKNLVSL GYQLTKPDVILRLEKGEEPLEINLWITKFVKD
566	KOX1/ZNF10 KRAB 2	MYPYDVPDYASPKKKRKVGGGASMDAKSLTAW*SRTLVTFKDVFVD FTREEWKLLDTAQQIVYRNVMLENYKNLVSLGYQLTKPDVILRLE KGEEPWLV*EREIHQETHPDSETAFEIKSSV
567	KOX1/ZNF10 KRAB 3	ALSPQHSAVTQGSIIKNKEGMDAKSLTAWSRTLVTFKDVFVDFTR EEWKLLDTAQQIVYRNVMLENYKNLVSLGYQLTKPDVILRLEKGE EPWLVEREIHQETHPDSETAFEIKSSV
568	KOX1/ZNF10 (aa 11-72)	RTLVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGY QLTKPDVILRLEKGEEP
569	KOX1/ZNF10 (aa 11-108)	RTLVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGY QLTKPDVILRLEKGEEPWLVEREIHQETHPDSETAFEIKSSVSSR SIFKDKQS

[0575]

SEQ	描述	序列
570	KOX1/ZNF10 变体	RTLVTFKDVAVDFTQEEWQQLDPAQKIVYRDVMLENYSNLVSVGY QLTKPDVILRLEQKGEEPWLVEEEIHQETHPDSETAFEIKSSVSS RSIFKDKQS
571	KOX1 KRAB-ZIM3 嵌合体	RTLVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGY QLTKPDVILRLEKGEEPWLEEEEVLGSGRAEKNGDIGGQIWKPKD VKESL
572	ZIM3-KOX1 KRAB 嵌合体	MNNSQGRVTFEDVTVNFTQGEWQRLNPEQRNLYRDVMLENYSNLV SVGQGETTKPDVILRLEQGKEPWLVEREIHQETHPDSETAFEIKS SVSSRSIFKDKQS
573	人 DNMT1	MPARTAPARVPTLAVPAISLPDDVRRRLKDLERDSLTEKECVKEK LNLLHEFLQTEIKNQLCDLETKLRKEELSEEGYLAKVKSLLNKDL SLENGAHAYNREVNGRLENGNQARSEARRVGMADANSPPKPLSKP RTPRRSKSDGEAKPEPSPSPRITRKSTRQTTITSHFAKGPAKRKP QEESERAKSDESIKEEDKDQDEKRRRVTSRERVARPLPAEEPERA KSGTRTEKEEERDEKEEKRLRSQTKEPTPKQKLKEEPDREARAGV QADEDEDGDEKDEKKHRSQPKDLAAKRRPEEKEPEKVNPQISDEK DEDEKEEKRRKTTPKEPTEKKMARAKTVMNSKTHPPKCIQCGQYL DDPLKYGQHPPDAVDEPQMLTNEKLSIFDANESGFESYEALPQHK LTCFSVYCKHGHLCPIDTGLIEKNIELFFSGSAKPIYDDDPSLEG GVNGKNLGPINEWWITGFDGGEKALIGFSTSFAEYILMDPSPEYA PIFGLMQEKIYISKIVVEFLQSNSDSTYEDLINKIETTVPPSGLN LNRFTEDSLLRHAQFVVEQVESYDEAGDSDEQPIFLTPCMRDLIK LAGVTLGQRRAQARRQTIRHSTREKDRGPTKATTTKLVYQIFDTF FAEQIEKDDREDKENAFKRRRCGVCEVCQQPECGKCKACKDMVKF GGSGRSKQACQERRCPNMAMKEADDDEEVDDNIPEMPSPKKMHQG KKKKQNKNRISWVGEAVKTDGKKSYYKKVCIDAETLEVGDCVSVI PDDSSKPLYLARVTALWEDSSNGQMFHAHWFCAGTDTVLGATSDP LELFLVDECEDMQLSYIHSKVKVIYKAPSENWAMEGGMDPESLLE GDDGKTYFYQLWYDQDYARFESPPKTQPTEDNKFKFCVSCARLAE MRQKEIPRVLEQLEDLDSRVLYYSATKNGILYRVGDGVYLPPEAF TFNIKLSSPVKRPRKEPVDEDLYPEHYRKYSDYIKGSNLDAPEPY RIGRIKEIFCPKKSNGRPNETDIKIRVNKFYRPENTHKSTPASYH ADINLLYWSDEEAVVDFKAVQGRCTVEYGEDLPECVQVYSMGGPN RFYFLEAYNAKSKSFEDPPNHARSPGNKGKGKGKGKGKPKSQACE PSEPEIEIKLPKLRTLDVFSGCGGLSEGFHQAGISDTLWAIEMWD PAAQAFRLNNPGSTVFTEDCNILLKLVMAGETTNSRGQRLPQKGD VEMLCGGPPCQGFSGMNRFNSRTYSKFKNSLVVSFLSYCDYYRPR FFLLENVRNFVSFKRSMVLKLTLRCLVRMGYQCTFGVLQAGQYGV AQTRRRAIILAAAPGEKLPLFPEPLHVFAPRACQLSVVVDDKKFV SNITRLSSGPFRTITVRDTMSDLPEVRNGASALEISYNGEPQSWF QRQLRGAQYQPILRDHICKDMSALVAARMRHIPLAPGSDWRDLPN IEVRLSDGTMARKLRYTHHDRKNGRSSSGALRGVCSCVEAGKACD PAARQFNTLIPWCLPHTGNRHNHWAGLYGRLEWDGFFSTTVTNPE PMGKQGRVLHPEQHRVVSVRECARSQGFPDTYRLFGNILDKHRQV GNAVPPPLAKAIGLEIKLCMLAKARESASAKIKEEEAAKD
574	人 DNMT3A	MPAMPSSGPGDTSSSAAEREEDRKDGEEQEEPRGKEERQEPSTTA RKVGRPGRKRKHPPVESGDTPKDPAVISKSPSMAQDSGASELLPN GDLEKRSEPQPEEGSPAGGQKGGAPAEGEGAAETLPEASRAVENG CCTPKEGRGAPAEAGKEQKETNIESMKMEGSRGRLRGGLGWESSL RQRPMPRLTFQAGDPYYISKRKRDEWLARWKREAEKKAKVIAGMN AVEENQGPGESQKVEEASPPAVQQPTDPASPTVATTPEPVGSDAG DKNATKAGDDEPEYEDGRGFGIGELVWGKLRGFSWWPGRIVSWWM TGRSRAAEGTRWVMWFGDGKFSVVCVEKLMPLSSFCSAFHQATYN KQPMYRKAIYEVLQVASSRAGKLFPVCHDSDESDTAKAVEVQNKP MIEWALGGFQPSGPKGLEPPEEEKNPYKEVYTDMWVEPEAAAYAP

[0576]

SEQ	描述	序列
		PPPAKKPRKSTAEKPKVKEIIDERTRERLVYEVRQKCRNIEDICI SCGSLNVTLEHPLFVGGMCQNCKNCFLECAYQYDDDGYQSYCTIC CGGREVLMCGNNNCCRCFCVECVDLLVGPGAAQAAIKEDPWNCYM CGHKGTYGLLRRREDWPSRLQMFFANNHDQEFDPPKVYPPVPAEK RKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCEDSITVGMV RHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPAR KGLYEGTGRLFFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSD KRDISRFLESNPVMIDAKEVSAAHRARYFWGNLPGMNRPLASTVN DKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEK EDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRH LFAPLKEYFACV
575	人 DNMT3A 催化结构域	NHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQ VDRYIASEVCEDSITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPF DLVIGGSPCNDLSIVNPARKGLYEGTGRLFFEFYRLLHDARPKEG DDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSAAHRA RYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTR SNSIKQGKDQHFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMS RLARQRLLGRSWSVPVIRHLFAPLKEYFACV
576	人 DNMT3B	MKGDTRHLNGEEDAGGREDSILVNGACSDQSSDSPPILEAIRTPE IRGRRSSSRLSKREVSSLLSYTQDLTGDGDGEDGDGSDTPVMPKL FRETRTRSESPAVRTRNNNSVSSRERHRPSPRSTRGRQGRNHVDE SPVEFPATRSLRRRATASAGTPWPSPPSSYLTIDLTDDTEDTHGT PQSSSTPYARLAQDSQQGGMESPQVEADSGDGDSSEYQDGKEFGI GDLVWGKIKGFSWWPAMVVSWKATSKRQAMSGMRWVQWFGDGKFS EVSADKLVALGLFSQHFNLATFNKLVSYRKAMYHALEKARVRAGK TFPSSPGDSLEDQLKPMLEWAHGGFKPTGIEGLKPNNTQPVVNKS KVRRAGSRKLESRKYENKTRRRTADDSATSDYCPAPKRLKTNCYN NGKDRGDEDQSREQMASDVANNKSSLEDGCLSCGRKNPVSFHPLF EGGLCQTCRDRFLELFYMYDDDGYQSYCTVCCEGRELLLCSNTSC CRCFCVECLEVLVGTGTAAEAKLQEPWSCYMCLPQRCHGVLRRRK DWNVRLQAFFTSDTGLEYEAPKLYPAIPAARRRPIRVLSLFDGIA TGYLVLKELGIKVGKYVASEVCEESIAVGTVKHEGNIKYVNDVRN ITKKNIEEWGPFDLVIGGSPCNDLSNVNPARKGLYEGTGRLFFEF YHLLNYSRPKEGDDRPFFWMFENVVAMKVGDKRDISRFLECNPVM IDAIKVSAAHRARYFWGNLPGMNRPVIASKNDKLELQDCLEYNRI AKLKKVQTITTKSNSIKQGKNQLFPVVMNGKEDVLWCTELERIFG FPVHYTDVSNMGRGARQKLLGRSWSVPVIRHLFAPLKDYFACE
577	小鼠 DNMT3C	MRGGSRHLSNEEDVSGCEDCIIISGTCSDQSSDPKTVPLTQVLEA VCTVENRGCRTSSQPSKRKASSLISYVQDLTGDGDEDRDGEVGGS SGSGTPVMPQLFCETRIPSKTPAPLSWQANTSASTPWLSPASPYP IIDLTDEDVIPQSISTPSVDWSQDSHQEGMDTTQVDAESRDGGNI EYQVSADKLLLSQSCILAAFYKLVPYRESIYRTLEKARVRAGKAC PSSPGESLEDQLKPMLEWAHGGFKPTGIEGLKPNKKQPENKSRRR TTNDPAASESSPPKRLKTNSYGGKDRGEDEESREQMASDVTNNKG NLEDHCLSCGRKDPVSFHPLFEGGLCQSCRDRFLELFYMYDEDGY QSYCTVCCEGRELLLCSNTSCCRCFCVECLEVLVGAGTAEDVKLQ EPWSCYMCLPQRCHGVLRRRKDWNMRLQDFFTTDPDLEEFEPPKL YPAIPAAKRRPIRVLSLFDGIATGYLVLKELGIKVEKYIASEVCA ESIAVGTVKHEGQIKYVDDIRNITKEHIDEWGPFDLVIGGSPCND LSCVNPVRKGLFEGTGRLFFEFYRLLNYSCPEEEDDRPFFWMFEN VVAMEVGDKRDISRFLECNPVMIDAIKVSAAHRARYFWGNLPGMN RPVMASKNDKLELQDCLEFSRTAKLKKVQTITTKSNSIRQGKNQL FPVVMNGKDDVLWCTELERIFGFPEHYTDVSNMGRGARQKLLGRS WSVPVIRHLFAPLKDHFACE

[0577]

SEQ	描述	序列
578	人 DNMT3L	MAAIPALDPEAEPSMDVILVGSSELSSSVSPGTGRDLIAYEVKAN QRNIEDICICCGSLQVHTQHPLFEGGICAPCKDKFLDALFLYDDD GYQSYCSICCSGETLLICGNPDCTRCYCFECVDSLVGPGTSGKVH AMSNWVCYLCLPSSRSGLLQRRRKWRSQLKAFYDRESENPLEMFE TVPVWRRQPVRVLSLFEDIKKELTSLGFLESGSDPGQLKHVVDVT DTVRKDVEEWGPFDLVYGATPPLGHTCDRPPSWYLFQFHRLLQYA RPKPGSPRPFFWMFVDNLVLNKEDLDVASRFLEMEPVTIPDVHGG SLQNAVRVWSNIPAIRSSRHWALVSEEELSLLAQNKQSSKLAAKW PTKLVKNCFLPLREYFKYFSTELTSSL
579	人 DNMT3L 催化结构域	NPLEMFETVPVWRRQPVRVLSLFEDIKKELTSLGFLESGSDPGQL KHVVDVTDTVRKDVEEWGPFDLVYGATPPLGHTCDRPPSWYLFQF HRLLQYARPKPGSPRPFFWMFVDNLVLNKEDLDVASRFLEMEPVT IPDVHGGSLQNAVRVWSNIPAIRSRHWALVSEEELSLLAQNKQSS KLAAKWPTKLVKNCFLPLREYFKYFSTELTSSL
580	小鼠 DNMT3L	MGSRETPSSCSKTLETLDLETSDSSSPDADSPLEEQWLKSSPALK EDSVDVVLEDCKEPLSPSSPPTGREMIRYEVKVNRRSIEDICLCC GTLQVYTRHPLFEGGLCAPCKDKFLESLFLYDDDGHQSYCTICCS GGTLFICESPDCTRCYCFECVDILVGPGTSERINAMACWVCFLCL PFSRSGLLQRRKRWRHQLKAFHDQEGAGPMEIYKTVSAWKRQPVR VLSLFRNIDKVLKSLGFLESGSGSGGGTLKYVEDVTNVVRRDVEK WGPFDLVYGSTQPLGSSCDRCPGWYMFQFHRILQYALPRQESQRP FFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVRGRDYQNAMRVW SNIPGLKSKHAPLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNCLL PLREYFKYFSQNSLPL
581	小鼠 DNMT3L 催化结构域	GPMEIYKTVSAWKRQPVRVLSLFRNIDKVLKSLGFLESGSGSGGG TLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPGWYMF QFHRILQYALPRQESQRPFFWIFMDNLLLTEDDQETTTRFLQTEA VTLQDVRGRDYQNAMRVWSNIPGLKSKHAPLTPKEEEYLQAQVRS RSKLDAPKVDLLVKNCLLPLREYFKYFSQNSLPL
582	大熊猫 DNMT3L	MALSPTGTLSVETLDRSDPDPLDEGPWQATCEILLEPDAEHSTDV ILVGSSELSAPASPGPRRDLLAYEVKVNQRDIEDVCICCGSLRVH TQHPLFEGGMCAPCKDKFLDCLFLYDDDGYQSYCSICCAGETLLI CENPDCTRPSLMMKLRLFRECACLIFPSEGMLLQTVWFWKMTVVW QPGLRHLPQENPLETYKTVPVWKREPVRVLSLFGDIRRELMSLGF LESGSAPGRLKHLDDVTDVVRKDVEGWGPFDLVYGSTPPIGHACD HPPVWYLLQFHRILQYARPRPGSQQPFFWMFVDNLVLSQDDQTAA TRFLEADPVTIQDVCGRAVRNTVHVWSNIPAVRSRHSALALCEEL SLLAQDRQRTKPPAQGPAQLVKNCFLPLREYFKYFSTELTSSL
583	大熊猫 DNMT3L 催化结构域	NPLETYKTVPVWKREPVRVLSLFGDIRRELMSLGFLESGSAPGRL KHLDDVTDVVRKDVEGWGPFDLVYGSTPPIGHACDHPPVWYLLQF HRILQYARPRPGSQQPFFWMFVDNLVLSQDDQTAATRFLEADPVT IQDVCGRAVRNTVHVWSNIPAVRSRHSALALCEELSLLAQDRQRT KPPAQGPAQLVKNCFLPLREYFKYFSTELTSSL
584	眼镜猴 DNMT3L	MALSCRRTLPLESLHSSNSDLASQLDKEQWRPPCETHGIPVAAAP VLDLEAECSLDVILVGSSELSTSSSPRLGRDHIAYEVKVNQRNIE DICLCCGSFLVHTQHPLFEGGMCAPCKDKFLDTLFLYDEDGYQSY CSICCSGETLLICENPDCTRCYCFECLDTLVSPGTSEKVHAMSNW VCFLCLPFTRSGLLQRRRKWRGQLKAFYDRESESSLEMYKTVPVW KREPVRVLSLFGDIKKELMSLGFVETGSDPGRLRHLDDTTNIVRR NVEEWGPFHLLYGATPPLGHTCDRPPGWYLFQFHRLLQYARPQPG SPQPFFWMFVDNVMLTREDRAIASRFLETEPVTIPDIHGRALQNA VCVWSNIPAVRSKHSALVSEEELSLLAQDRQRAKLPTQGPTKLVK NCFLPLREYFKYFSTELTSFL
585	眼镜猴	SSLEMYKTVPVWKREPVRVLSLFGDIKKELMSLGFVETGSDPGRL RHLDDTTNIVRRNVEEWGPFHLLYGATPPLGHTCDRPPGWYLFQF

[0578]

SEQ	描述	序列
	DNMT3L 催化结构域	HRLLQYARPQPGSPQPFFWMFVDNVMLTREDRAIASRFLETEPVT IPDIHGRALQNAVCVWSNIPAVRSKHSALVSEEELSLLAQDRQRA KLPTQGPTKLVKNCFLPLREYFKYFSTELTSFL
586	长爪沙鼠 DNMT3L	MGSQETPSTRAKTPGTWNLESTDSSSPESLGHLEEQWANSSPDLK DEHSKDVEPEDSKELISSASPPSGREIIRYEISVNQRNIEDICLC CGTLQVYKQHPLFEGGICAPCKDKFLETFFLYDEDGHQSYCSICC SGGTLFICESPDCTRCYCFECVDILVGPGTSERINAMPCWVCFLC LPFTRSGLLQRRRKWRHQLKAFFDEGGASPLEMYKTVSAWKRKPM RVLSLFKNIDKELKNLGFLESGSGSEEERLKYLEDVTNVVRRDVE KWGPFDLVYGSTRPRGSSCDHCPAWYMFQFHRILQYARPPSGSEQ PFFWVFVDNLLMTEDDQITADRFLQMKAVTLQDVRGRVLQNAVRV WSNIPGVKSKHMALTEKEEQSLEAQAGTRTKLSAQKVDPLVKNCL LPLREYFKFFSQNSLPLDK
587	长爪沙鼠 DNMT3L 催化结构域	SPLEMYKTVSAWKRKPMRVLSLFKNIDKELKNLGFLESGSGSEEE RLKYLEDVTNVVRRDVEKWGPFDLVYGSTRPRGSSCDHCPAWYMF QFHRILQYARPPSGSEQPFFWVFVDNLLMTEDDQITADRFLQMKA VTLQDVRGRVLQNAVRVWSNIPGVKSKHMALTEKEEQSLEAQAGT RTKLSAQKVDPLVKNCLLPLREYFKFFSQNSLPLDK
588	北美鼠兔 DNMT3L	MALPSPETLDSLDRVPASHPDEQHWTVCDNSDPILEVEAEGSMDV ILVDDSPAPSGRDRIELEVKVNQRSIEDLCLCCGSSQVHRQHPLF QGGLCAPCKDKFLEALFLYDEDGYQSYCSICGLGDTLLVCESPDC TRGYCFACVDGLVGAGSSGHMHTVSPWVCFLCVPGSRHGLLQRRR RWRTQLKVFHEQEAAQPLEIYETVPACRRKPLRVLSLFEHIEKEL ASLGFLETGSSPGRIRHLDDVTDVVRRDVEQWGPFDLVYGSTPPL GHASPRSPGWYLFQFHRMLQYTQPTASTQRPFFWMFVDNLLLTRD DLVTATRFLEVEPATLQDVRGRVLQGAMRVWSNIPAVNSRHTELA PEAETALLAQSCRRAKASGEGLARLLKSCFLPLREYFKYFPQSPL PLRK
589	北美鼠兔 DNMT3L 催化结构域	QPLEIYETVPACRRKPLRVLSLFEHIEKELASLGFLETGSSPGRI RHLDDVTDVVRRDVEQWGPFDLVYGSTPPLGHASPRSPGWYLFQF HRMLQYTQPTASTQRPFFWMFVDNLLLTRDDLVTATRFLEVEPAT LQDVRGRVLQGAMRVWSNIPAVNSRHTELAPEAETALLAQSCRRA KASGEGLARLLKSCFLPLREYFKYFPQSPLPLRK
590	新东部灰松鼠 DNMT3L	MGGPRPAAVEESPHEIYKTVPAWKREPMRVLSLFGDIGKELTSLG FLETGSEAGRLKHLEDVTDTVRRDVEEWGPFDLVYGSTPALGHSC DRSPGWYLFQFHRLLQYARPRLGSPKPFFWMFVDNLLLTKDDQAI ASRFLEMEPVTLQDVHGRVLQNAVRVWTNVPAVKSRHSALASEEE LLLVQDGQRGRLPAQGPAALVKHCFLPLREYFKYFSQNTLPLYK
591	新东部灰松鼠 DNMT3L 催化结构域	SPHEIYKTVPAWKREPMRVLSLFGDIGKELTSLGFLETGSEAGRL KHLEDVTDTVRRDVEEWGPFDLVYGSTPALGHSCDRSPGWYLFQF HRLLQYARPRLGSPKPFFWMFVDNLLLTKDDQAIASRFLEMEPVT LQDVHGRVLQNAVRVWTNVPAVKSRHSALASEEELLLVQDGQRGR LPAQGPAALVKHCFLPLREYFKYFSQNTLPLYK
592	北美野牛 DNMT3L	MARSSPGTLNLEIMDGSDPDPALPPDREQWPPPCEILLDPEPEHS LDIILVGSSELSSPPSPGPRRDFIAYEVKVNQRDIEDVCICCGSL QLHTQHPLFEGGMCAPCKDKFLECLFLYDDDGYQSYCSICCAGET LLICENPDCTRCYCFECVDTLVGPGTSGKVHAMSNWVCFLCLPFP RSGLLQRRRKWRTWLKAFYDREAESPLVMYKTVPVWKREPIRVLS LFGDIKKELTSLGFLEDGSKPGRLKHLDDVTNIVRRDIDEWGPFD LTYGSTPTLGHTCDHPPGWYVYQFHRILQYARPLPGSPQPFFWMF VDNLVLTEEDLDVATRFLETDPVTIQDVRGRTVQNAVHVWSNIPA VKSRHSALVSQEELSLLAQDRQRVKSPVQGPATLVKNCFLPLREY FKYFSTELTSSL
593	北美野牛 DNMT3L	SPLVMYKTVPVWKREPIRVLSLFGDIKKELTSLGFLEDGSKPGRL KHLDDVTNIVRRDIDEWGPFDLTYGSTPTLGHTCDHPPGWYVYQF

[0579]

SEQ	描述	序列
	催化结构域	HRILQYARPLPGSPQPFFWMFVDNLVLTEEDLDVATRFLETDPVT IQDVRGRTVQNAVHVWSNIPAVKSRHSALVSQEELSLLAQDRQRV KSPVQGPATLVKNCFLPLREYFKYFSTELTSSL
594	普氏野马 DNMT3L	MALSSPGTLSLETLDSWDPDVAGQLDEERWQPSSEIVGRPMAAAP VLDLEEEPSMDIILVDSSELSSPPSPGPSRDMCICCGSFQVHTQH PLFEGGMCAACKDKFLSCLFLYDDDGNQSYCSICCSGETLLICEN PDCTRCYCFECVDTLVSPRTSEKVQAMSNWVCFLCLPFPRSGLLQ RRRKWRGWLKAFYDQEAVRSRSAWGRRMRSGPHLVGFLWLLVAKC PSALESPLEMYKTVPVWKREPVRVLSLFGDIKKELTTLGFLENGS DPGRLKHLDDVTNTVRRDVEEWGPFDLVYGSTPPLGHACDHPPGW YLFQFHRVLQYARPRPGSPQAFFWMFVDNLVLTEDDRAVATRFLE TDPVTIQDVCGRAVRNAVHVWSNIPAVKSRHSALFSQEESFLRAQ DRQRAKPPARGPAKLVKNCFLPLREYFKYFSTEFTSSL
595	普氏野马 DNMT3L 催化结构域	SPLEMYKTVPVWKREPVRVLSLFGDIKKELTTLGFLENGSDPGRL KHLDDVTNTVRRDVEEWGPFDLVYGSTPPLGHACDHPPGWYLFQF HRVLQYARPRPGSPQAFFWMFVDNLVLTEDDRAVATRFLETDPVT IQDVCGRAVRNAVHVWSNIPAVKSRHSALFSQEESFLRAQDRQRA KPPARGPAKLVKNCFLPLREYFKYFSTEFTSSL
596	卡氏小鼠 DNMT3L	MGSRETPSSFSKTLETLDLETSDSSSPDADSPLEEQWLKSSPALK EDNVDMVLEDCKEPLSPSSPPTGREMIRYEVKVNRRSIEDICLCC GTLQVYTQHPLFEGGICAPCKDKFLESLFLYDDDGHQSYCTICCS GGTLFICESPDCTRCYCFECVDILVGPGTSERINAMACWVCFLCL PFSRSGLLQRRKRWRHQLKAFHDQEGAGPMEIYKTVSTWKRQPVR VLSLFGNIDKVLKSLGFLESGSGSGGGTLKYVEDVTNVVRRDVEK WGPFDLVYGSTQPLGSSCDRCPGWYMFQFHRILQYALPRQESQRP FFWIFMDNLLMTEDDQETTARFLQTEAVTLQDVRGRDYQNVMRVW SNIPGLKSKHVPLTPKEEEYLQAQVRTRSKLDAQKVDLLVKNCLL PLREYFKYFS
597	卡氏小鼠 DNMT3L 催化结构域	GPMEIYKTVSTWKRQPVRVLSLFGNIDKVLKSLGFLESGSGSGGG TLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPGWYMF QFHRILQYALPRQESQRPFFWIFMDNLLMTEDDQETTARFLQTEA VTLQDVRGRDYQNVMRVWSNIPGLKSKHVPLTPKEEEYLQAQVRT RSKLDAQKVDLLVKNCLLPLREYFKYFS
598	黑猩猩 DNMT3L	MAAIPALDPEAEPSMDVILVGSSELSSSISPRTGRDLIAYEVKAN QRNIEDICICCGSLQVHTQHPLFEGGICAPCKDKSLDALFLYDDD GYQSYCSICCSGETLLICGNPDCTRCYCFECVDSLVGPGTSGKVH AMSNWVCYLCLPSSRSGLLQRRRKWRSQLKAFYDRESENPLEMFE TVPVWRRQPVRVLSLFEDIKKELTSLGFLESGSDPGQLKHVVDVT DTVRKDVEEWGPFDLVYGATPPLGHTCDRPPSWYLFQFHRLLQYA RPKPGSPRPFFWMFVDNLVLNKEDLDVASRFLEMEPVTIPDVHGG SLQNAVRVWSNIPAIRSSRHWALVSEEELSLLAQNKQSSKLAAKW PTKLVKNCFLPLREYFKYFSTELTSSL
599	黑猩猩 DNMT3L 催化结构域	NPLEMFETVPVWRRQPVRVLSLFEDIKKELTSLGFLESGSDPGQL KHVVDVTDTVRKDVEEWGPFDLVYGATPPLGHTCDRPPSWYLFQF HRLLQYARPKPGSPRPFFWMFVDNLVLNKEDLDVASRFLEMEPVT IPDVHGGSLQNAVRVWSNIPAIRSSRHWALVSEEELSLLAQNKQS SKLAAKWPTKLVKNCFLPLREYFKYFSTELTSSL
600	人 TRDMT1 (DNMT2)	MEPLRVLELYSGVGGMHHALRESCIPAQVVAAIDVNTVANEVYKY NFPHTQLLAKTIEGITLEEFDRLSFDMILMSPPCQPFTRIGRQGD MTDSRTNSFLHILDILPRLQKLPKYILLENVKGFEVSSTRDLLIQ TIENCGFQYQEFLLSPTSLGIPNSRLRYFLIAKLQSEPLPFQAPG QVLMEFPKIESVHPQKYAMDVENKIQEKNVEPNISFDGSIQCSGK DAILFKLETAEEIHRKNQQDSDLSVKMLKDFLEDDTDVNQYLLPP KSLLRYALLLDIVQPTCRRSVCFTKGYGSYIEGTGSVLQTAEDVQ

[0580]

SEQ	描述	序列
		VENIYKSLTNLSQEEQITKLLILKLRYFTPKEIANLLGFPPEFGF PEKITVKQRYRLLGNSLNVHVVAKLIKILYE
601	支原体目细菌甲基转移酶	MAEWYIPAIVSYQAIHNGFTLNKINHKIELQTMIDYLESKTLSMN SKEPVKRGFWYKKHLDEIRIVYTAVKMSEQEGNIFDVRTLFERGL SDIDLLTYSFPCQDLSQQGKQKGMGRDSQTRSGLLWEIEKALDTS KKEDLPKYLLMENVVALTHKVNAEELDEWMMKLESLGYKNDLRIL NAGDFGSSQARRRTFMISTLNEKVELPVGNKKPKSMNKILNDEPT RKDFLPALDKFDLTEYKWTKSNINKAKLINYSTFNSEAYVYDSNF TGPTLTASGANSRIKFEYNGKIRKIGAEEAYAYMGFKKSDYIKVN KLNYLNETKMIYTCGNSISVEVLRSIMTNINNNFKENK
602	海洋支原体甲基转移酶	MLFLIGTFKYVLIYITKVIRIFEAFAGIGAQRKALRNIKSNYEVS GMAEWYIPAIVSYQAIHNGFTLSRVDKKTKLTEMIKYLESKTLSM DSKEPVRTGYWFKKHKDMVRIVYSAVKLSEAEGNIFDVRTLHERK LEDIDLLTYSFPCQDLSQQGKQRGMKKDSGTRSGLLWEIEKALEA TPKDKLPKYLLMENVVALTHKTNKKDLDNWKRKLRSLGYYNDINV LNAGDFGSSQARRRAFMISTLDSKVTLPLGDKKPQAISKILNKET RSQDFMPALDEYEKTDFKRTLSNIKKCKLIDYTSFNSEAYVYDPK YTGPTLTASGANSRIKFTHQGKMRKINAEEAYRYMGFSTNDYKKV NNLNFLSETKMIYTCGNSISVEVLEEIMLKIIREDNNG
603	中华螺原体甲基转移酶	MKKIRLFEAFAGIGSQRRALKSVVGNNFEIAGLAEWYVPAIVMYQ IINNDFSKKNVLDNVPRDEVIDYLNSKCLSWDSKKPVSKNFWNRK SQDILNVIYSAVKKSEEEGNIFDVRTLHERTLESIDILTYSFPCQ DLSQQGIQKGMKKNSGTRSGLLWEIEKAIDNTPKNNLPKILLMEN VPALLNKTNELELKEWLIKLENMGYKNSIGILNAADFGSPQARRR VFMISSRNKKIELPVGKSKPGKLNDILEKNVEDKFIMTNLEKYDF SEFSLTKSNIKKCSLINYTKFNSEAYVYDPDFTGPTLTASGANSR IKIYDKGFIRRMSPLESFRYMGFDDEDYKKIDEFEFLTDTQKIFV CGNSISIEVLKAIFERIDSNE
604	穿透支原体 *M MpeI*	MNSNKDKIKVIKVFEAFAGIGSQFKALKNIARSKNWEIQHSGMVE WFVDAIVSYVAIHSKNFNPKIEQLDKDILSISNDSKMPISEYGIK KINNTIKASYLNYAKKHFNNLFDIKKVNKDNFPKNIDIFTYSFPC QDLSVQGLQKGIDKELNTRSGLLWEIERILEEIKNSFSKEEMPKY LLMENVKNLLSHKNKKNYNTWLKQLEKFGYKSKTYLLNSKNFDNC QNRERVFCLSIRDDYLEKTGFKFKELEKVKNPPKKIKDILVDSSN YKYLNLNKYETTTFRETKSNIISRSLKNYTTFNSENYVYNINGIG PTLTASGANSRIKIETQQGVRYLTPLECFKYMQFDVNDFKKVQST NLISENKMIYIAGNSIPVKILEAIFNTLEFVNNEE
605	单体螺原体 *M SssI*	MSKVENKTKKLRVFEAFAGIGAQRKALEKVRKDEYEIVGLAEWYV PAIVMYQAIHNNFHTKLEYKSVSREEMIDYLENKTLSWNSKNPVS NGYWKRKKDDELKIIYNAIKLSEKEGNIFDIRDLYKRTLKNIDLL TYSFPCQDLSQQGIQKGMKRGSGTRSGLLWEIERALDSTEKNDLP KYLLMENVGALLHKKNEEELNQWKQKLESLGYQNSIEVLNAADFG SSQARRVFMISTLNEFVELPKGDKKPKSIKKVLNKIVSEKDILN NLLKYNLTEFKKTKSNINKASLIGYSKFNSEGYVYDPEFTGPTLT ASGANSRIKIKDGSNIRKMNSDETFLYIGFDSQDGKRVNEIEFLT ENQKIFVCGNSISVEVLEAIIDKIGG
606	副流感嗜血杆菌 *M HpaII*	MKDVLDDNLLEEPAAQYSLFEPESNPNLREKFTFIDLFAGIGGFR IAMQNLGGKCIFSSEWDEQAQKTYEANFGDLPYGDITLEETKAFI PEKFDILCAGFPCQAFSIAGKRGGFEDTRGTLFFDVAEIIRRHQP KAFFLENVKGLKNHDKGRTLKTILNVLREDLGYFVPEPAIVNAKN FGVPQNRERIYIVGFHKSTGVNSFSYPEPLDKIVTFADIREEKTV PTKYYLSTQYIDTLRKHKERHESKGNGFGYEIIPDDGIANAIVVG GMGRERNLVIDHRITDFTPTTNIKGEVNREGIRKMTPREWARLQG FPDSYVIPVSDASAYKQFGNSVAVPAIQATGKKILEKLGNLYD

[0581]

SEQ	描述	序列
607	藤黄节杆菌 *M AluI*	MSKANAKYSFVDLFAGIGGFHAALAATGGVCEYAVEIDREAAAVY ERNWNKPALGDITDDANDEGVTLRGYDGPIDVLTGGFPCQPFSKS GAQHGMAETRGTLFWNIARIIEEREPTVLILENVRNLVGPRHRHE WLTIIETLRFFGYEVSGAPAIFSPHLLPAWMGGTPQVRERVFITA TLVPERMRDERIPRTETGEIDAEAIGPKPVATMNDRFPIKKGGTE LFHPGDRKSGWNLLTSGIIREGDPEPSNVDLRLTETETLWIDAWD DLESTIRRATGRPLEGFPYWADSWTDFRELSRLVVIRGFQAPERE VVGDRKRYVARTDMPEGFVPASVTRPAIDETLPAWKQSHLRRNYD FFERHFAEVVAWAYRWGVYTDLFPASRRKLEWQAQDAPRLWDTVM HFRPSGIRAKRPTYLPALVAITQTSIVGPLERRLSPRETARLQGL PEWFDFGEQRAAATYKQMGNGVNVGVVRHILREHVRRDRALLKLT PAGQRIINAVLADEPDATVGALGAAE
608	埃及嗜血杆菌 *M HaeIII*	MNLISLFSGAGGLDLGFQKAGFRIICANEYDKSIWKTYESNHSAK LIKGDISKISSDEFPKCDGIIGGPPCQSWSEGGSLRGIDDPRGKL FYEYIRILKQKKPIFFLAENVKGMMAQRHNKAVQEFIQEFDNAGY DVHIILLNANDYGVAQDRKRVFYIGFRKELNINYLPPIPHLIKPT FKDVIWDLKDNPIPALDKNKTNGNKCIYPNHEYFIGSYSTIFMSR NRVRQWNEPAFTVQASGRQCQLHPQAPVMLKVSKNLNKFVEGKEH LYRRLTVRECARVQGFPDDFIFHYESLNDGYKMIGNAVPVNLAYE IAKTIKSALEICKGN
609	溶血嗜血杆菌 *M HhaI*	MIEIKDKQLTGLRFIDLFAGLGGFRLALESCGAECVYSNEWDKYA QEVYEMNFGEKPEGDITQVNEKTIPDHDILCAGFPCQAFSISGKQ KGFEDSRGTLFFDIARIVREKKPKVVFMENVKNFASHDNGNTLEV VKNTMNELDYSFHAKVLNALDYGIPQKRERIYMICFRNDLNIQNF QFPKPFELNTFVKDLLLPDSEVEHLVIDRKDLVMTNQEIEQTTPK TVRLGIVGKGGQGERIYSTRGIAITLSAYGGGIFAKTGGYLVNGK TRKLHPRECARVMGYPDSYKVHPSTSQAYKQFGNSVVINVLQYIA YNIGSSLNFKPY
610	莫拉氏菌 *M MspI*	MKPEILKLIRSKLDLTQKQASEIIEVSDKTWQQWESGKTEMHPAY YSFLQEKLKDKINFEELSAQKTLQKKIFDKYNQNQITKNAEELAE ITHIEERKDAYSSDFKFIDLFSGIGGIRQSFEVNGGKCVFSSEID PFAKFTYYTNFGVVPFGDITKVEATTIPQHDILCAGFPCQPFSHI GKREGFEHPTQGTMFHEIVRIIETKKTPVLFLENVPGLINHDDGN TLKVIIETLEDMGYKVHHTVLDASHFGIPQKRKRFYLVAFLNQNI HFEFPKPPMISKDIGEVLESDVTGYSISEHLQKSYLFKKDDGKPS LIDKNTTGAVKTLVSTYHKIQRLTGTFVKDGETGIRLLTTNECKA IMGFPKDFVIPVSRTQMYRQMGNSVVVPVVTKIAEQISLALKTVN QQSPQENFELELV
611	粪盘菌属 Masc1	MSERRYEAGMTVALHEGSFLKIQRVYIRQYHADNRREHMLVGPLF RRTKYLKALSKKVNEVAIVHESIHVPVQDVIGVRELIITNRPFPE CRKGDEHTGRLVCRWVYNLDERAKGREYKKQRYIRRITEAEADPE YRVEDRVLRRRWFQEGYIGDEISYKEHGNGDIVDIRSESPLQVLD GWGGDLVDLENGEETSIPGPCRSASSYGRLMKPPLAQAADSNTSR KYTFGDTFCGGGGVSLGARQAGLEVKWAFDMNPNAGANYRRNFPN TDFFLAEAEQFIQLSVGISQHVDILHLSPPCQTFSRAHTIAGKND ENNEASFFAVVNLIKAVRPRLFTVEETDGIMDRQSRQFIDTALMG ITELGYSFRICVLNAIEYGVCQNRKRLIIIGAAPGEELPPFPLPT HQDFFSKDPRRDLLPAVTLDDALSTITPESTDHHLNHVWQPAEWK TPYDAHRPFKNAIRAGGGEYDIYPDGRRKFTVRELACIQGFPDEY EFVGTLTDKRRIIGNAVPPPLSAAIMSTLRQWMTEKDFERME
612	拟南芥 MET1	MVENGAKAAKRKKRPLPEIQEVEDVPRTRRPRRAAACTSFKEKSI RVCEKSATIEVKKQQIVEEEFLALRLTALETDVEDRPTRRLNDFV LFDSDGVPQPLEMLEIHDIFVSGAILPSDVCTDKEKEKGVRCTSF GRVEHWSISGYEDGSPVIWISTELADYDCRKPAASYRKVYDYFYE KARASVAVYKKLSKSSGGDPDIGLEELLAAVVRSMSSGSKYFSSG

[0582]

SEQ	描述	序列
		AAIIDFVISQGDFIYNQLAGLDETAKKHESSYVEIPVLVALREKS SKIDKPLQRERNPSNGVRIKEVSQVAESEALTSDQLVDGTDDDRR YAILLQDEENRKSMQQPRKNSSSGSASNMFYIKINEDEIANDYPL PSYYKTSEEETDELILYDASYEVQSEHLPHRMLHNWALYNSDLRF ISLELLPMKQCDDIDVNIFGSGVVTDDNGSWISLNDPDSGSQSHD PDGMCIFLSQIKEWMIEFGSDDIISISIRTDVAWYRLGKPSKLYA PWWKPVLKTARVGISILTFLRVESRVARLSFADVTKRLSGLQAND KAYISSDPLAVERYLVVHGQIILQLFAVYPDDNVKRCPFVVGLAS KLEDRHHTKWIIKKKKISLKELNLNPRAGMAPVASKRKAMQATTT RLVNRIWGEFYSNYSPEDPLQATAAENGEDEVEEEGGNGEEEVEE EGENGLTEDTVPEPVEVQKPHTPKKIRGSSGKREIKWDGESLGKT SAGEPLYQQALVGGEMVAVGGAVTLEVDDPDEMPAIYFVEYMFES TDHCKMLHGRFLQRGSMTVLGNAANERELFLTNECMTTQLKDIKG VASFEIRSRPWGHQYRKKNITADKLDWARALERKVKDLPTEYYCK SLYSPERGGFFSLPLSDIGRSSGFCTSCKIREDEEKRSTIKLNVS KTGFFINGIEYSVEDFVYVNPDSIGGLKEGSKTSFKSGRNIGLRA YVVCQLLEIVPKESRKADLGSFDVKVRRFYRPEDVSAEKAYASDI QELYFSQDTVVLPPGALEGKCEVRKKSDMPLSREYPISDHIFFCD LFFDTSKGSLKQLPANMKPKFSTIKDDTLLRKKKGKGVESEIESE IVKPVEPPKEIRLATLDIFAGCGGLSHGLKKAGVSDAKWAIEYEE PAGQAFKQNHPESTVFVDNCNVILRAIMEKGGDQDDCVSTTEANE LAAKLTEEQKSTLPLPGQVDFINGGPPCQGFSGMNRFNQSSWSKV QCEMILAFLSFADYFRPRYFLLENVRTFVSFNKGQTFQLTLASLL EMGYQVRFGILEAGAYGVSQSRKRAFIWAAAPEEVLPEWPEPMHV FGVPKLKISLSQGLHYAAVRSTALGAPFRPITVRDTIGDLPSVEN GDSRTNKEYKEVAVSWFQKEIRGNTIALTDHICKAMNELNLIRCK LIPTRPGADWHDLPKRKVTLSDGRVEEMIPFCLPNTAERHNGWKG LYGRLDWQGNFPTSVTDPQPMGKVGMCFHPEQHRILTVRECARSQ GFPDSYEFAGNINHKHRQIGNAVPPPLAFALGRKLKEALHLKKSP QHQP
613	粪盘菌属 Masc2	MELTPELSGVSTDLGGGGSIFAHWRMKEESPAPTEILDDLNVLEW EKTTRDYSKEDLRIADQLFSIEDEHQSLPFETADAEDGTPTEEEE EKELPMRTLDNFVLYDASDLELAALDLIGTELNIHAVGTVGPIYT EGEEDEQEDEDEDVSPPVRTGTQATSASVTQMTVELYIRNIVQYE FCFNDDGTVETWIQTTNAHYKLLQPAKCYTSLYRPVNDCLNVITA IITLAPESTTMSLKDLLKVMDDKAQAVSYEEVERMSEFIVQHLDQ WMETAPKKKSKLIEKSKVYIDLNNLAGIDMVSGVRPPPVRRVTGR SSAPKKRIVRNMNDAVLLHQNETTVTNWIHQLSAGMFGRALNVLG AETADVENLTCDPASAKFVVPQRRLHKRLKWETRGHIPVSEEEYK HIYQGKKYAKFFEAVRAVDESKLTIKLGDLVYVLDQDPKVTQTQF ATAGREGRKKGAEKEKIQVRFGRVLSIRQPDSNSKDAQNVFIHVQ WLVLGCDTILQEMASRRELFLTDSCDTVFADVIYGVAKLTPLGAK DIPTVEFHESMATMMGENEFFVRFKYNYQDGSFTDLKDVDAEQIG TLQPRVNTHRNPGYCSNCRIKYDNERTGDKWIYENDTEGEPRLFR SSKGWCIYAQEFVYLQPVEKQPGTTFRVGYISEINKSSVIVELLA RVDDDDKSGHISYSDPRHLYFTGTDIKVTFDKIIRKCFVFHDSGD QKAKAPLMYGTLQRDLYYYRYEKRKGKAELVPVREIRSIHEQTLN DWESRTQIERHGAVSGKKLKGLDIFAGCGGLTLGLDLSGAVDTKW DIEFAPSAANTLALNFPDAQVFNQCANVLLSRAIQSEDEGSLDIE YDLQGRVLPDLPKKGEVDFIYGGPPCQGFSGVNRYKKGNDIKNSL VATFLSYVDHYKPRFVLLENVKGLITTKLGNSKNAEGKWEGGISN GVVKFIYRTLISMNYQCRIGLVQSGEYGVPQSRPRVIFLAARMGE RLPDLPEPMHAFEVLDSQYALPHIKRYHTTQNGVAPLPRITIGEA VSDLPKFQYANPGVWPRHDPYSSAKAQPSDKTIEKFSVSKATSFV GYLLQPYHSRPQSEFQRRLRTKLVPSDEPAEKTSLLTTKLVTAHV

[0583]

SEQ	描述	序列
		TRLFNKETTQRIVCVPMWPGADHRSLPKEMRPWCLVDPNSQAEKH RFWPGLFGRLGMEDFFSTALTDVQPCGKQGKVLHPTQRRVYTVRE LARAQGFPDWFAFTDGDADSGLGGVKKWHRNIGNAVPVPLGEQIG RCIGYSVWWKDDMIAQLREDGADEDEEMIDGNDQWVEELNTQMAA DMPGLPLLVTHLLNLCVYRRLYGPNAKEFLPARVYDKKLEGGRRR LVWAML
614	链孢霉 Dim2	MDSPDRSHGGMFIDVPAETMGFQEDYLDMFASVLSQGLAKEGDYA HHQPLPAGKEECLEPIAVATTITPSPDDPQLQLQLELEQQFQTES GLNGVDPAPAPESEDEADLPDGFSDESPDDDFVVQRSKHITVDLP VSTLINPRSTFQRIDENDNLVPPPQSTPERVAVEDLLKAAKAAGK NKEDYIEFELHDFNFYVNYAYHPQEMRPIQLVATKVLHDKYYFDG VLKYGNTKHYVTGMQVLELPVGNYGASLHSVKGQIWVRSKHNAKK EIYYLLKKPAFEYQRYYQPFLWIADLGKHVVDYCTRMVERKREVT LGCFKSDFIQWASKAHGKSKAFQNWRAQHPSDDFRTSVAANIGYI WKEINGVAGAKRAAGDQLFRELMIVKPGQYFRQEVPPGPVVTEGD RTVAATIVTPYIKECFGHMILGKVLRLAGEDAEKEKEVKLAKRLK IENKNATKADTKDDMKNDTATESLPTPLRSLPVQVLEATPIESDI VSIVSSDLPPSENNPPPLTNGSVKPKAKANPKPKPSTQPLHAAHV KYLSQELVNKIKVGDVISTPRDDSSNTDTKWKPTDTDDHRWFGLV QRVHTAKTKSSGRGLNSKSFDVIWFYRPEDTPCCAMKYKWRNELF LSNHCTCQEGHHARVKGNEVLAVHPVDWFGTPESNKGEFFVRQLY ESEQRRWITLQKDHLTCYHNQPPKPPTAPYKPGDTVLATLSPSDK FSDPYEVVEYFTQGEKETAFVRLRKLLRRRKVDRQDAPANELVYT EDLVDVRAERIVGKCIMRCFRPDERVPSPYDRGGTGNMFFITHRQ DHGRCVPLDTLPPTLRQGFNPLGNLGKPKLRGMDLYCGGGNFGRG LEEGGVVEMRWANDIWDKAIHTYMANTPDPNKTNPFLGSVDDLLR LALEGKFSDNVPRPGEVDFIAAGSPCPGFSLLTQDKKVLNQVKNQ SLVASFASFVDFYRPKYGVLENVSGIVQTFVNRKQDVLSQLFCAL VGMGYQAQLILGDAWAHGAPQSRERVFLYFAAPGLPLPDPPLPSH SHYRVKNRNIGFLCNGESYVQRSFIPTAFKFVSAGEGTADLPKIG DGKPDACVRFPDHRLASGITPYIRAQYACIPTHPYGMNFIKAWNN GNGVMSKSDRDLFPSEGKTRTSDASVGWKRLNPKTLFPTVTTTSN PSDARMGPGLHWDEDRPYTVQEMRRAQGYLDEEVLVGRTTDQWKL VGNSVSRHMALAIGLKFREAWLGTLYDESAVVATATATATTAAAV GVTVPVMEEPGIGTTESSRPSRSPVHTAVDLDDSKSERSRSTTPA TVLSTSSAAGDGSANAAGLEDDDNDDMEMMEVTRKRSSPAVDEEG MRPSKVQKVEVTVASPASRRSSRQASRNPTASPSSKASKATTHEA PAPEELESDAESYSETYDKEGFDGDYHSGHEDQYSEEDEEEEYAE PETMTVNGMTIVKL
615	果蝇 dDnmt2	MVFRVLELFSGIGGMHYAFNYAQLDGQIVAALDVNTVANAVYAHN YGSNLVKTRNIQSLSVKEVTKLQANMLLMSPPCQPHTRQGLQRDT EDKRSDALTHLCGLIPECQELEYILMENVKGFESSQARNQFIESL ERSGFHWREFILTPTQFNVPNTRYRYYCIARKGADFPFAGGKIWE EMPGAIAQNQGLSQIAEIVEENVSPDFLVPDDVLTKRVLVMDIIH PAQSRSMCFTKGYTHYTEGTGSAYTPLSEDESHRIFELVKEIDTS NQDASKSEKILQQRLDLLHQVRLRYFTPREVARLMSFPENFEFPP ETTNRQKYRLLGNSINVKVVGELIKLLTIK
616	裂殖酵母 Pmt1	MLSTKRLRVLELYSGIGGMHYALNLANIPADIVCAIDINPQANEI YNLNHGKLAKHMDISTLTAKDFDAFDCKLWTMSPSCQPFTRIGNR KDILDPRSQAFLNILNVLPHVNNLPEYILIENVQGFEESKAAEEC RKVLRNCGYNLIEGILSPNQFNIPNSRSRWYGLARLNFKGEWSID DVFQFSEVAQKEGEVKRIRDYLEIERDWSSYMVLESVLNKWGHQF DIVKPDSSSCCCFTRGYTHLVQGAGSILQMSDHENTHEQFERNRM ALQLRYFTAREVARLMGFPESLEWSKSNVTEKCMYRLLGNSINVK VVSYLISLLLEPLNF

[0584]

SEQ	描述	序列
617	拟南芥 DRM1	MVMSHIFLISQIQEVEHGDSDDVNWNTDDDELAIDNFQFSPSPVH ISATSPNSIQNRISDETVASFVEMGFSTQMIARAIEETAGANMEP MMILETLFNYSASTEASSSKSKVINHFIAMGFPEEHVIKAMQEHG DEDVGEITNALLTYAEVDKLRESEDMNININDDDDDNLYSLSSDD EEDELNNSSNEDRILQALIKMGYLREDAAIAIERCGEDASMEEVV DFICAAQMARQFDEIYAEPDKKELMNNNKKRRTYTETPRKPNTDQ LISLPKEMIGFGVPNHPGLMMHRPVPIPDIARGPPFFYYENVAMT PKGVWAKISSHLYDIVPEFVDSKHFCAAARKRGYIHNLPIQNRFQ IQPPQHNTIQEAFPLTKRWWPSWDGRTKLNCLLTCIASSRLTEKI REALERYDGETPLDVQKWVMYECKKWNLVWVGKNKLAPLDADEME KLLGFPRDHTRGGGISTTDRYKSLGNSFQVDTVAYHLSVLKPLFP NGINVLSLFTGIGGGEVALHRLQIKMNVVVSVEISDANRNILRSF WEQTNQKGILREFKDVQKLDDNTIERLMDEYGGFDLVIGGSPCNN LAGGNRHHRVGLGGEHSSLFFDYCRILEAVRRKARHMRR
618	拟南芥 DRM2	MVIWNNDDDDFLEIDNFQSSPRSSPIHAMQCRVENLAGVAVTTSS LSSPTETTDLVQMGFSDEVFATLFDMGFPVEMISRAIKETGPNVE TSVIIDTISKYSSDCEAGSSKSKAIDHFLAMGFDEEKVVKAIQEH GEDNMEAIANALLSCPEAKKLPAAVEEEDGIDWSSSDDDTNYTDM LNSDDEKDPNSNENGSKIRSLVKMGFSELEASLAVERCGENVDIA ELTDFLCAAQMAREFSEFYTEHEEQKPRHNIKKRRFESKGEPRSS VDDEPIRLPNPMIGFGVPNEPGLITHRSLPELARGPPFFYYENVA LTPKGVWETISRHLFEIPPEFVDSKYFCVAARKRGYIHNLPINNR FQIQPPPKYTIHDAFPLSKRWWPEWDKRTKLNCILTCTGSAQLTN RIRVALEPYNEEPEPPKHVQRYVIDQCKKWNLVWVGKNKAAPLEP DEMESILGFPKNHTRGGGMSRTERFKSLGNSFQVDTVAYHLSVLK PIFPHGINVLSLFTGIGGGEVALHRLQIKMKLVVSVEISKVNRNI LKDFWEQTNQTGELIEFSDIQHLTNDTIEGLMEKYGGFDLVIGGS PCNNLAGGNRVSRVGLEGDQSSLFFEYCRILEVVRARMRGS
619	拟南芥 CMT1	MAARNKQKKRAEPESDLCFAGKPMSVVESTIRWPHRYQSKKTKLQ APTKKPANKGGKKEDEEIIKQAKCHFDKALVDGVLINLNDDVYVT GLPGKLKFIAKVIELFEADDGVPYCRFRWYYRPEDTLIERFSHLV QPKRVFLSNDENDNPLTCIWSKVNIAKVPLPKITSRIEQRVIPPC DYYYDMKYEVPYLNFTSADDGSDASSSLSSDSALNCFENLHKDEK FLLDLYSGCGAMSTGFCMGASISGVKLITKWSVDINKFACDSLKL NHPETEVRNEAAEDFLALLKEWKRLCEKFSLVSSTEPVESISELE DEEVEENDDIDEASTGAELEPGEFEVEKFLGIMFGDPQGTGEKTL QLMVRWKGYNSSYDTWEPYSGLGNCKEKLKEYVIDGFKSHLLPLP GTVYTVCGGPPCQGISGYNRYRNNEAPLEDQKNQQLLVFLDIIDF LKPNYVLMENVVDLLRFSKGFLARHAVASFVAMNYQTRLGMMAAG SYGLPQLRNRVFLWAAQPSEKLPPYPLPTHEVAKKFNTPKEFKDL QVGRIQMEFLKLDNALTLADAISDLPPVTNYVANDVMDYNDAAPK TEFENFISLKRSETLLPAFGGDPTRRLFDHQPLVLGDDDLERVSY IPKQKGANYRDMPGVLVHNNKAEINPRFRAKLKSGKNVVPAYAIS FIKGKSKKPFGRLWGDEIVNTVVTRAEPHNQCVIHPMQNRVLSVR ENARLQGFPDCYKLCGTIKEKYIQVGNAVAVPVGVALGYAFGMAS QGLTDDEPVIKLPFKYPECMQAKDQI
620	拟南芥 CMT2	MLSPAKCESEEAQAPLDLHSSSRSEPECLSLVLWCPNPEEAAPSS TRELIKLPDNGEMSLRRSTTLNCNSPEENGGEGRVSQRKSSRGKS QPLLMLTNGCQLRRSPRFRALHANFDNVCSVPVTKGGVSQRKFSR GKSQPLLTLTNGCQLRRSPRFRAVDGNFDSVCSVPVTGKFGSRKR KSNSALDKKESSDSEGLTFKDIAVIAKSLEMEIISECQYKNNVAE GRSRLQDPAKRKVDSDTLLYSSINSSKQSLGSNKRMRRSQRFMKG TENEGEENLGKSKGKGMSLASCSFRRSTRLSGTVETGNTETLNRR KDCGPALCGAEQVRGTERLVQISKKDHCCEAMKKCEGDGLVSSKQ ELLVFPSGCIKKTVNGCRDRTLGKPRSSGLNTDDIHTSSLKISKN

[0585]

SEQ	描述	序列
		DTSNGLTMTTALVEQDAMESLLQGKTSACGAADKGKTREMHVNST VIYLSDSDEPSSIEYLNGDNLTQVESGSALSSGGNEGIVSLDLNN PTKSTKRKGKRVTRTAVQEQNKRSICFFIGEPLSCEEAQERWRWR YELKERKSKSRGQQSEDDEDKIVANVECHYSQAKVDGHTFSLGDF AYIKGEEEETHVGQIVEFFKTTDGESYFRVQWFYRATDTIMERQA TNHDKRRLFYSTVMNDNPVDCLISKVTVLQVSPRVGLKPNSIKSD YYFDMEYCVEYSTFQTLRNPKTSENKLECCADVVPTESTESILKK KSFSGELPVLDLYSGCGGMSTGLSLGAKISGVDVVTKWAVDQNTA ACKSLKLNHPNTQVRNDAAGDFLQLLKEWDKLCKRYVFNNDQRTD TLRSVNSTKETSGSSSSSDDDSDSEEYEVEKLVDICFGDHDKTGK NGLKFKVHWKGYRSDEDTWELAEELSNCQDAIREFVTSGFKSKIL PLPGRVGVICGGPPCQGISGYNRHRNVDSPLNDERNQQIIVFMDI VEYLKPSYVLMENVVDILRMDKGSLGRYALSRLVNMRYQARLGIM TAGCYGLSQFRSRVFMWGAVPNKNLPPFPLPTHDVIVRYGLPLEF ERNVVAYAEGQPRKLEKALVLKDAISDLPHVSNDEDREKLPYESL PKTDFQRYIRSTKRDLTGSAIDNCNKRTMLLHDHRPFHINEDDYA RVCQIPKRKGANFRDLPGLIVRNNTVCRDPSMEPVILPSGKPLVP GYVFTFQQGKSKRPFARLWWDETVPTVLTVPTCHSQALLHPEQDR VLTIRESARLQGFPDYFQFCGTIKERYCQIGNAVAVSVSRALGYS LGMAFRGLARDEHLIKLPQNFSHSTYPQLQETIPH
621	拟南芥 CMT3	MAPKRKRPATKDDTTKSIPKPKKRAPKRAKTVKEEPVTVVEEGEK HVARFLDEPIPESEAKSTWPDRYKPIEVQPPKASSRKKTKDDEKV EIIRARCHYRRAIVDERQIYELNDDAYVQSGEGKDPFICKIIEMF EGANGKLYFTARWFYRPSDTVMKEFEILIKKKRVFFSEIQDTNEL GLLEKKLNILMIPLNENTKETIPATENCDFFCDMNYFLPYDTFEA IQQETMMAISESSTISSDTDIREGAAAISEIGECSQETEGHKKAT LLDLYSGCGAMSTGLCMGAQLSGLNLVTKWAVDMNAHACKSLQHN HPETNVRNMTAEDFLFLLKEWEKLCIHFSLRNSPNSEEYANLHGL NNVEDNEDVSEESENEDDGEVFTVDKIVGISFGVPKKLLKRGLYL KVRWLNYDDSHDTWEPIEGLSNCRGKIEEFVKLGYKSGILPLPGG VDVVCGGPPCQGISGHNRFRNLLDPLEDQKNKQLLVYMNIVEYLK PKFVLMENVVDMLKMAKGYLARFAVGRLLQMNYQVRNGMMAAGAY GLAQFRLRFFLWGALPSEIIPQFPLPTHDLVHRGNIVKEFQGNIV AYDEGHTVKLADKLLLKDVISDLPAVANSEKRDEITYDKDPTTPF QKFIRLRKDEASGSQSKSKSKKHVLYDHHPLNLNINDYERVCQVP KRKGANFRDFPGVIVGPGNVVKLEEGKERVKLESGKTLVPDYALT YVDGKSCKPFGRLWWDEIVPTVVTRAEPHNQVIIHPEQNRVLSIR ENARLQGFPDDYKLFGPPKQKYIQVGNAVAVPVAKALGYALGTAF QGLAVGKDPLLTLPEGFAFMKPTLPSELA
622	链孢霉 Rid	MAEQNPFVIDDEDDVIQIHDEEEVEEEVAEVIDITEDDIEPSELD RAFGSRPKEETLPSLLLRDQGFIVRPGMTVELKAPIGRFAISFVR VNSIVKVRQAHVNNVTIRGHGFTRAKEMNGMLPKQLNECCLVASI DTRDPRP
623	大肠杆菌菌株 12 hsdM	MNNNDLVAKLWKLCDNLRDGGVSYQNYVNELASLLFLKMCKETGQ EAEYLPEGYRWDDLKSRIGQEQLQFYRKMLVHLGEDDKKLVQAVF HNVSTTITEPKQITALVSNMDSLDWYNGAHGKSRDDFGDMYEGLL QKNANETKSGAGQYFTPRPLIKTIIHLLKPQPREVVQDPAAGTAG FLIEADRYVKSQTNDLDDLDGDTQDFQIHRAFIGLELVPGTRRLA LMNCLLHDIEGNLDHGGAIRLGNTLGSDGENLPKAHIVATNPPFG SAAGTNITRTFVHPTSNKQLCFMQHIIETLHPGGRAAVVVPDNVL FEGGKGTDIRRDLMDKCHLHTILRLPTGIFYAQGVKTNVLFFTKG TVANPNQDKNCTDDVWVYDLRTNMPSFGKRTPFTDEHLQPFERVY GEDPHGLSPRTEGEWSFNAEETEVADSEENKNTDQHLATSRWRKF SREWIRTAKSDSLDISWLKDKDSIDADSLPEPDVLAAEAMGELVQ ALSELDALMRELGASDEADLQRQLLEEAFGGVKE

[0586]

SEQ	描述	序列
624	大肠杆菌菌株 12 hsdS	MSAGKLPEGWVIAPVSTVTTLIRGVTYKKEQAINYLKDDYLPLIR ANNIQNGKFDTTDLVFVPKNLVKESQKISPEDIVIAMSSGSKSVV GKSAHQHLPFECSFGAFCGVLRPEKLIFSGFIAHFTKSSLYRNKI SSLSAGANINNIKPASFDLINIPIPPLAEQKIIAEKLDTLLAQVD STKARFEQIPQILKRFRQAVLGGAVNGKLTEKWRNFEPQHSVFKK LNFESILTELRNGLSSKPNESGVGHPILRISSVRAGHVDQNDIRF LECSESELNRHKLQDGDLLFTRYNGSLEFVGVCGLLKKLQHQNLL YPDKLIRARLTKDALPEYIEIFFSSPSARNAMMNCVKTTSGQKGI SGKDIKSQVVLLPPVKEQAEIVRRVEQLFAYADTIEKQVNNALAR VNNLTQSILAKAFRGELTAQWRAENPDLISGENSAAALLEKIKAE RAASGGKKASRKKS
625	水生栖热菌 M TaqI	MGLPPLLSLPSNSAPRSLGRVETPPEVVDFMVSLAEAPRGGRVLE PACAHGPFLRAFREAHGTAYRFVGVEIDPKALDLPPWAEGILADF LLWEPGEAFDLILGNPPYGIVGEASKYPIHVFKAVKDLYKKAFST WKGKYNLYGAFLEKAVRLLKPGGVLVFVVPATWLVLEDFALLREF LAREGKTSVYYLGEVFPQKKVSAVVIRFQKSGKGLSLWDTQESES GFTPILWAEYPHWEGEIIRFETEETRKLEISGMPLGDLFHIRFAA RSPEFKKHPAVRKEPGPGLVPVLTGRNLKPGWVDYEKNHSGLWMP KERAKELRDFYATPHLVVAHTKGTRVVAAWDERAYPWREEFHLLP KEGVRLDPSSLVQWLNSEAMQKHVRTLYRDFVPHLTLRMLERLPV RREYGFHTSPESARNF
626	大肠杆菌 M EcoDam	MKKNRAFLKWAGGKYPLLDDIKRHLPKGECLVEPFVGAGSVFLNT DFSRYILADINSDLISLYNIVKMRTDEYVQAARELFVPETNCAEV YYQFREEFNKSQDPFRRAVLFLYLNRYGYNGLCRYNLRGEFNVPF GRYKKPYFPEAELYHFAEKAQNAFFYCESYADSMARADDASVVYC DPPYAPLSATANFTAYHTNSFTLEQQAHLAEIAEGLVERHIPVLI SNHDTMLTREWYQRAKLHVVKVRRSISSNGGTRKKVDELLALYKP GVVSPAKK
627	新月梭菌 M CcrMI	MKFGPETIIHGDCIEQMNALPEKSVDLIFADPPYNLQLGGDLLRP DNSKVDAVDDHWDQFESFAAYDKFTREWLKAARRVLKDDGAIWVI GSYHNIFRVGVAVQDLGFWILNDIVWRKSNPMPNFKGTRFANAHE TLIWASKSQNAKRYTFNYDALKMANDEVQMRSDWTIPLCTGEERI KGADGQKAHPTQKPEALLYRVILSTTKPGDVILDPFFGVGTTGAA AKRLGRKFIGIEREAEYLEHAKARIAKVVPIAPEDLDVMGSKRAE PRVPFGTIVEAGLLSPGDTLYCSKGTHVAKVRPDGSITVGDLSGS IHKIGALVQSAPACNGWTYWHFKTDAGLAPIDVLRAQVRAGMN
628	艰难梭菌 CamA	MDDISQDNFLLSKEYENSLDVDTKKASGIYYTPKIIVDYIVKKTL KNHDIIKNPYPRILDISCGCGNFLLEVYDILYDLFEENIYELKKK YDENYWTVDNIHRHILNYCIYGADIDEKAISILKDSLTNKKVVND LDESDIKINLFCCDSLKKKWRYKFDYIVGNPPYIGHKKLEKKYKK FLLEKYSEVYKDKADLYFCFYKKIIDILKQGGIGSVITPRYFLES LSGKDLREYIKSNVNVQEIVDFLGANIFKNIGVSSCILTFDKKKT KETYIDVFKIKNEDICINKFETLEELLKSSKFEHFNINQRLLSDE WILVNKDDETFYNKIQEKCKYSLEDIAISFQGIITGCDKAFILSK DDVKLNLVDDKFLKCWIKSKNINKYIVDKSEYRLIYSNDIDNENT NKRILDEIIGLYKTKLENRRECKSGIRKWYELQWGREKLFFERKK IMYPYKSNENRFAIDYDNNFSSADVYSFFIKEEYLDKFSYEYLVG ILNSSVYDKYFKITAKKMSKNIYDYYPNKVMKIRIFRDNNYEEIE NLSKQIISILLNKSIDKGKVEKLQIKMDNLIMDSLGI
629	KAP1	MAASAAAASAAAASAASGSPGPGEGSAGGEKRSTAPSAAASASAS AAASSPAGGGAEALELLEHCGVCRERLRPEREPRLLPCLHSACSA CLGPAAPAAANSSGDGGAAGDGTVVDCPVCKQQCFSKDIVENYFM RDSGSKAATDAQDANQCCTSCEDNAPATSYCVECSEPLCETCVEA HQRVKYTKDHTVRSTGPAKSRDGERTVYCNVHKHEPLVLFCESCD TLTCRDCQLNAHKDHQYQFLEDAVRNQRKLLASLVKRLGDKHATL

[0587]

SEQ	描述	序列
		QKSTKEVRSSIRQVSDVQKRVQVDVKMAILQIMKELNKRGRVLVN DAQKVTEGQQERLERQHWTMTKIQKHQEHILRFASWALESDNNTA LLLSKKLIYFQLHRALKMIVDPVEPHGEMKFQWDLNAWTKSAEAF GKIVAERPGTNSTGPAPMAPPRAPGPLSKQGSGSSQPMEVQEGYG FGSGDDPYSSAEPHVSGVKRSRSGEGEVSGLMRKVPRVSLERLDL DLTADSQPPVFKVFPGSTTEDYNLIVIERGAAAAATGQPGTAPAG TPGAPPLAGMAIVKEEETEAAIGAPPTATEGPETKPVLMALAEGP GAEGPRLASPSGSTSSGLEVVAPEGTSAPGGGPGTLDDSATICRV CQKPGDLVMCNQCEFCFHLDCHLPALQDVPGEEWSCSLCHVLPDL KEEDGSLSLDGADSTGVVAKLSPANQRKCERVLLALFCHEPCRPL HQLATDSTFSLDQPGGTLDLTLIRARLQEKLSPPYSSPQEFAQDV GRMFKQFNKLTEDKADVQSIIGLQRFFETRMNEAFGDTKFSAVLV EPPPMSLPGAGLSSQELSGGPGDGP
630	MECP2	MVAGMLGLREEKSEDQDLQGLKDKPLKFKKVKKDKKEEKEGKHEP VQPSAHHSAEPAEAGKAETSEGSGSAPAVPEASASPKQRRSIIRD RGPMYDDPTLPEGWTRKLKQRKSGRSAGKYDVYLINPQGKAFRSK VELIAYFEKVGDTSLDPNDFDFTVTGRGSPSRREQKPPKKPKSPK APGTGRGRGRPKGSGTTRPKAATSEGVQVKRVLEKSPGKLLVKMP FQTSPGGKAEGGGATTSTQVMVIKRPGRKRKAEADPQAIPKKRGR KPGSVVAAAAAEAKKKAVKESSIRSVQETVLPIKKRKTRETVSIE VKEVVKPLLVSTLGEKSGKGLKTCKSPGRKSKESSPKGRSSSASS PPKKEHHHHHHHSESPKAPVPLLPPLPPPPPEPESSEDPTSPPEP QDLSSSVCKEEKMPRGGSLESDGCPKEPAKTQPAVATAATAAEKY KHRGEGERKDIVSSSMPRPNREEPVDSRTPVTERVS
631	接头	SGGS
632	接头	SGGSSGSETPGTSESATPESSGGS
633	接头	SGGSSGGSSGSETPGTSESATPESSGGSSGGS
634	接头	GGSGGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGS PAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESG PGSEPATSGGSGGS
635	G 接头	GSGGG
636	GX4 接头	GGGGSGGGGSGGGGSGGGGS
637	W 接头	SSGNSNANSRGPSFSSGLVPLSLRGSH
638	XTEN 接头 (XTEN16)	SGSETPGTSESATPES
639	XTEN 接头	SGGSSGGSSGSETPGTSESATPES
640	XTEN 接头	SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGS
641	XTEN 接头	SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGSSGSET PGTSESATPESSGGSSGGS
642	XTEN 接头	PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPT STEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPA TS
643	XTEN 接头 (XTEN80)	GGPSSGAPPPSGGSPAGSPTSTEEGTSESATPESGPGTSTEPSEG SAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSE
644	NLS	PKKKRKV
645	NLS	AVKRPAATKKAGQAKKKKLD
646	NLS	MSRRRKANPTKLSENAKKLAKEVEN
647	NLS	PAAKRVKLD
648	NLS	KLKIKRPVK
649	NLS	MDSLLMNRRKFLYQFKNVRWAKGRRETYLC
660	融合蛋白 (构型 7)	MGTMPKKKRKVPKKKRKVYNHDQEFDPPKVYPPVPAEKRKPIRVL SLFDGIATGLLVLKDLGIQVDRYIASEVCEDSITVGMVRHQGKIM

[0588]

SEQ	描述	序列
		YVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGT GRLFFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRF LESNPVMIDAKEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQE CLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEKEDILWCT EMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLKE YFACVSSGNSNANSRGPSFSSGLVPLSLRGSHMAAIPALDPEAEP SMDVILVGSSELSSSVSPGTGRDLIAYEVKANQRNIEDICICCGS LQVHTQHPLFEGGICAPCKDKFLDALFLYDDDGYQSYCSICCSGE TLLICGNPDCTRCYCFECVDSLVGPGTSGKVHAMSNWVCYLCLPS SRSGLLQRRRKWRSQLKAFYDRESENPLEMFETVPVWRRQPVRVL SLFEDIKKELTSLGFLESGSDPGQLKHVVDVTDTVRKDVEEWGPF DLVYGATPPLGHTCDRPPSWYLFQFHRLLQYARPKPGSPRPFFWM FVDNLVLNKEDLDVASRFLEMEPVTIPDVHGGSLQNAVRVWSNIP AIRSRHWALVSEEELSLLAQNKQSSKLAAKWPTKLVKNCFLPLRE YFKYFSTELTSSLGGPSSGAPPPSGGSPAGSPTSTEEGTSESATP ESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTE PSELEDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRH SIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIF SNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEK YPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNP DNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRR LENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQL SKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTE ITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSK NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRK QRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTF RIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSF IERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRK PAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIS GVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLF EDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIR DKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVI EMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDS IDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDK GRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRK RMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQ LFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKP IREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT LIHQSITGLYETRIDLSQLGGDSPKKKRKVGVDGSSGSETPGTSE SATPESRTLVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLENYKN LVSLGYQLTKPDVILRLEKGEEPSADYKDDDDKAPKKKRKVPKKK RKV
661	融合蛋白 (构型 9)	MGTMPKKKRKVPKKKRKVYNHDQEFDPPKVYPPVPAEKRKPIRVL SLFDGIATGLLVLKDLGIQVDRYIASEVCEDSITVGMVRHQGKIM YVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGT GRLFFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRF LESNPVMIDAKEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQE

[0589]

SEQ	描述	序列
		CLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEKEDILWCT EMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLKE YFACVSSGNSNANSRGPSFSSGLVPLSLRGSHMAAIPALDPEAEP SMDVILVGSSELSSSVSPGTGRDLIAYEVKANQRNIEDICICCGS LQVHTQHPLFEGGICAPCKDKFLDALFLYDDDGYQSYCSICCSGE TLLICGNPDCTRCYCFECVDSLVGPGTSGKVHAMSNWVCYLCLPS SRSGLLQRRRKWRSQLKAFYDRESENPLEMFETVPVWRRQPVRVL SLFEDIKKELTSLGFLESGSDPGQLKHVVDVTDTVRKDVEEWGPF DLVYGATPPLGHTCDRPPSWYLFQFHRLLQYARPKPGSPRPFFWM FVDNLVLNKEDLDVASRFLEMEPVTIPDVHGGSLQNAVRVWSNIP AIRSRHWALVSEEELSLLAQNKQSSKLAAKWPTKLVKNCFLPLRE YFKYFSTELTSSLGGPSSGAPPPSGGSPAGSPTSTEEGTSESATP ESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTE PSELEDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRH SIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIF SNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEK YPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNP DNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRR LENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQL SKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTE ITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSK NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRK QRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTF RIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSF IERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRK PAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIS GVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLF EDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIR DKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVI EMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDS IDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDK GRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRK RMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQ LFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKP IREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT LIHQSITGLYETRIDLSQLGGDSPKKKRKVGVDGSSGSETPGTSE SATPESTGNKKLEAVGTGIEPKAMSQGLVTFGDVAVDFSQEEWEW LNPIQRNLYRKVMLENYRNLASLGLCVSKPDVISSLEQGKEPWSA DYKDDDDKAPKKKRKVPKKKRKV
662	融合蛋白 (构型 11)	MGTMPKKKRKVPKKKRKVYNHDQEFDPPKVYPPVPAEKRKPIRVL SLFDGIATGLLVLKDLGIQVDRYIASEVCEDSITVGMVRHQGKIM YVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGT GRLFFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRF LESNPVMIDAKEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQE CLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEKEDILWCT EMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLKE YFACVSSGNSNANSRGPSFSSGLVPLSLRGSHMAAIPALDPEAEP

[0590]

SEQ	描述	序列
		SMDVILVGSSELSSSVSPGTGRDLIAYEVKANQRNIEDICICCGS LQVHTQHPLFEGGICAPCKDKFLDALFLYDDDGYQSYCSICCSGE TLLICGNPDCTRCYCFECVDSLVGPGTSGKVHAMSNWVCYLCLPS SRSGLLQRRRKWRSQLKAFYDRESENPLEMFETVPVWRRQPVRVL SLFEDIKKELTSLGFLESGSDPGQLKHVVDVTDTVRKDVEEWGPF DLVYGATPPLGHTCDRPPSWYLFQFHRLLQYARPKPGSPRPFFWM FVDNLVLNKEDLDVASRFLEMEPVTIPDVHGGSLQNAVRVWSNIP AIRSRHWALVSEEELSLLAQNKQSSKLAAKWPTKLVKNCFLPLRE YFKYFSTELTSSLGGPSSGAPPPSGGSPAGSPTSTEEGTSESATP ESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTE PSELEDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRH SIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIF SNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEK YPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNP DNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRR LENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQL SKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTE ITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSK NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRK QRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTF RIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSF IERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRK PAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIS GVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLF EDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIR DKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVI EMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDS IDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDK GRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRK RMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQ LFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKP IREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT LIHQSITGLYETRIDLSQLGGDSPKKKRKVGVDGSSGSETPGTSE SATPESTGDSVAFEDVAVNFTLEEWALLDPSQKNLYRDVMRETFR NLASVGKQWEDQNIEDPFKIPRRNISHIPERLCESKEGGQGEESA DYKDDDDKAPKKKRKVPKKKRKV
663	融合蛋白 (构型 13)	MGTMPKKKRKVPKKKRKVYNHDQEFDPPKVYPPVPAEKRKPIRVL SLFDGIATGLLVLKDLGIQVDRYIASEVCEDSITVGMVRHQGKIM YVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGT GRLFFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRF LESNPVMIDAKEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQE CLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEKEDILWCT EMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLKE YFACVSSGNSNANSRGPSFSSGLVPLSLRGSHMAAIPALDPEAEP SMDVILVGSSELSSSVSPGTGRDLIAYEVKANQRNIEDICICCGS LQVHTQHPLFEGGICAPCKDKFLDALFLYDDDGYQSYCSICCSGE TLLICGNPDCTRCYCFECVDSLVGPGTSGKVHAMSNWVCYLCLPS

[0591]

SEQ	描述	序列
		SRSGLLQRRRKWRSQLKAFYDRESENPLEMFETVPVWRRQPVRVL SLFEDIKKELTSLGFLESGSDPGQLKHVVDVTDTVRKDVEEWGPF DLVYGATPPLGHTCDRPPSWYLFQFHRLLQYARPKPGSPRPFFWM FVDNLVLNKEDLDVASRFLEMEPVTIPDVHGGSLQNAVRVWSNIP AIRSRHWALVSEEELSLLAQNKQSSKLAAKWPTKLVKNCFLPLRE YFKYFSTELTSSLGGPSSGAPPPSGGSPAGSPTSTEEGTSESATP ESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTE PSELEDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRH SIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIF SNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEK YPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNP DNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRR LENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQL SKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTE ITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSK NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRK QRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTF RIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSF IERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRK PAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIS GVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLF EDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIR DKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVI EMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDS IDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDK GRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRK RMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQ LFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKP IREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT LIHQSITGLYETRIDLSQLGGDSPKKKRKVGVDGSSGSETPGTSE SATPESTGMNNSQGRVTFEDVTVNFTQGEWQRLNPEQRNLYRDVM LENYSNLVSVGQGETTKPDVILRLEQGKEPWLEEEEVLGSGRAEK NGDIGGQIWKPKDVKESLSADYKDDDDKAPKKKRKVPKKKRKV
664	接头	GGGGS
665	接头	EAAAK
631	接头	SGGS

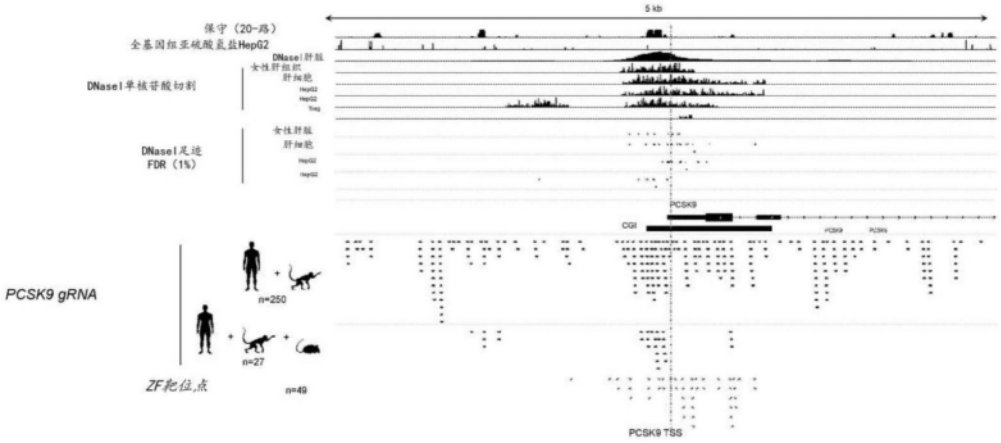

5 kb
保守（20-路）
全基因组亚硫酸氢盐HepG2
DNaseI单核苷酸切割
DNaseI足迹
FDR（1%）
PCSK9
CGI
PCSK9 gRNA
n=250
n=27
n=49
ZF靶位点
PCSK9 TSS

图1

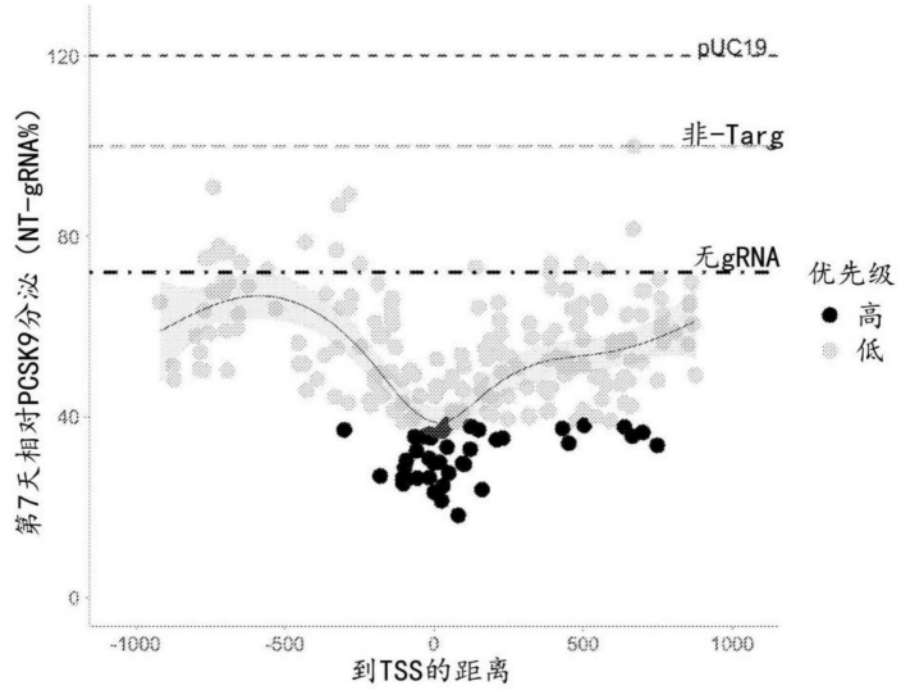

120
80
40
0
pUC19
非-Targ
无gRNA
优先级
高
低
第7天相对PCSK9分泌（NT-gRNA%）
-1000
-500
0
500
1000
到TSS的距离

图2

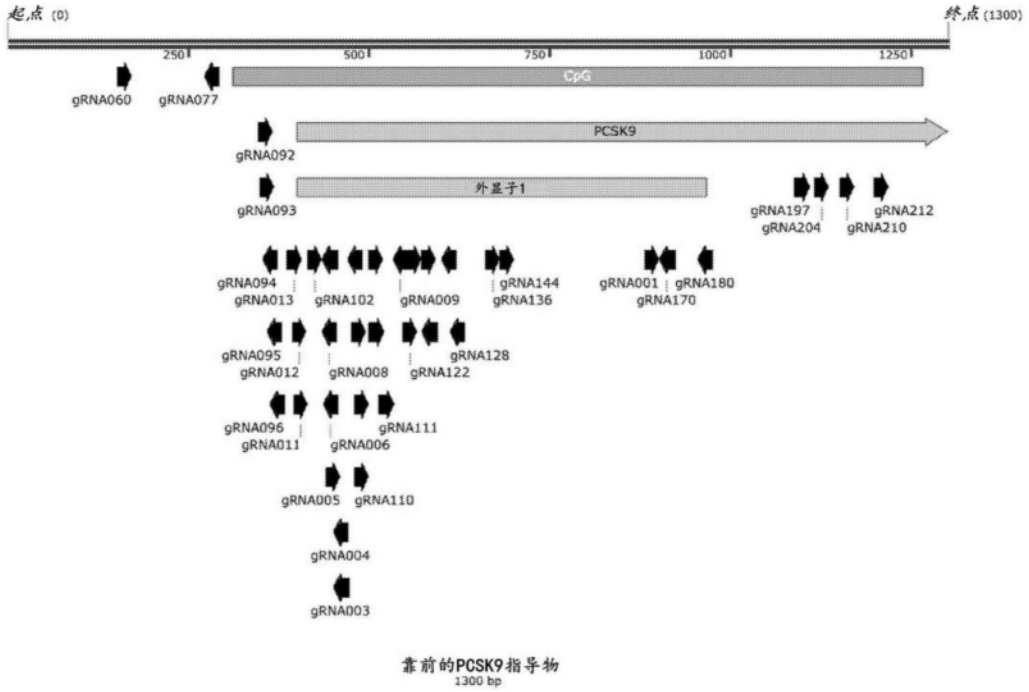

起点 (0)
终点 (1300)
250
500
750
1000
1250
CpG
PCSK9
外显子1
gRNA060
gRNA077
gRNA092
gRNA093
gRNA197
gRNA204
gRNA212
gRNA210
gRNA094
gRNA013
gRNA102
gRNA009
gRNA144
gRNA136
gRNA001
gRNA170
gRNA180
gRNA095
gRNA012
gRNA008
gRNA122
gRNA128
gRNA096
gRNA011
gRNA006
gRNA111
gRNA005
gRNA110
gRNA004
gRNA003
靠前的PCSK9指导物
1300 bp

图3

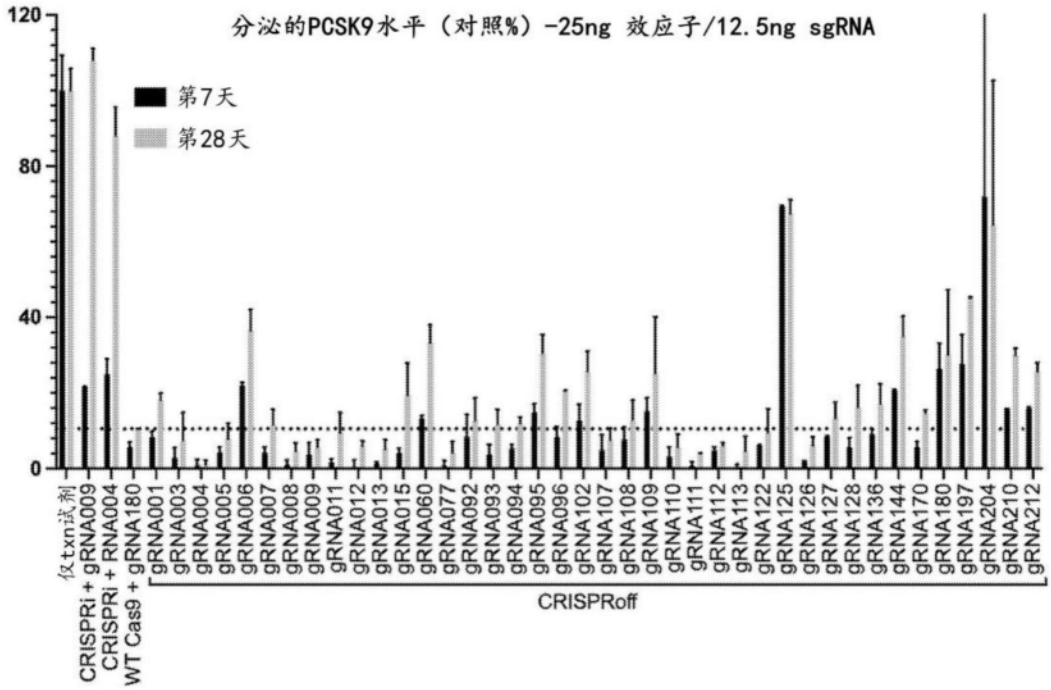

分泌的PCSK9水平（对照%）-25ng 效应子/12.5ng sgRNA
第7天
第28天
120
80
40
0
仅txn试剂
CRISPRi + gRNA009
CRISPRi + RNA004
WT Cas9 + gRNA180
gRNA001
gRNA003
gRNA004
gRNA005
gRNA006
gRNA007
gRNA008
gRNA009
gRNA011
gRNA012
gRNA013
gRNA015
gRNA060
gRNA077
gRNA092
gRNA093
gRNA094
gRNA095
gRNA096
gRNA102
gRNA107
gRNA108
gRNA109
gRNA110
gRNA111
gRNA112
gRNA113
gRNA122
gRNA125
gRNA126
gRNA127
gRNA128
gRNA136
gRNA144
gRNA170
gRNA180
gRNA197
gRNA204
gRNA210
gRNA212
CRISPRoff

图4A

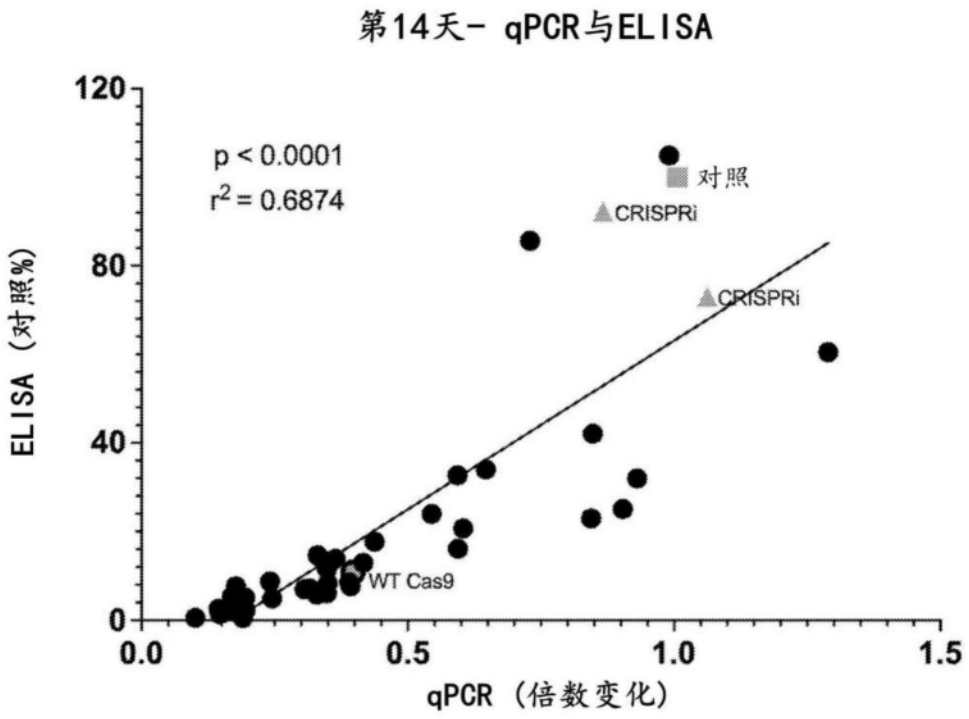

第14天- qPCR与ELISA
120
80
40
0
0.0
0.5
1.0
1.5
p < 0.0001
r2 = 0.6874
对照
CRISPRi
CRISPRi
WT Cas9
ELISA（对照%）
qPCR（倍数变化）

图4B

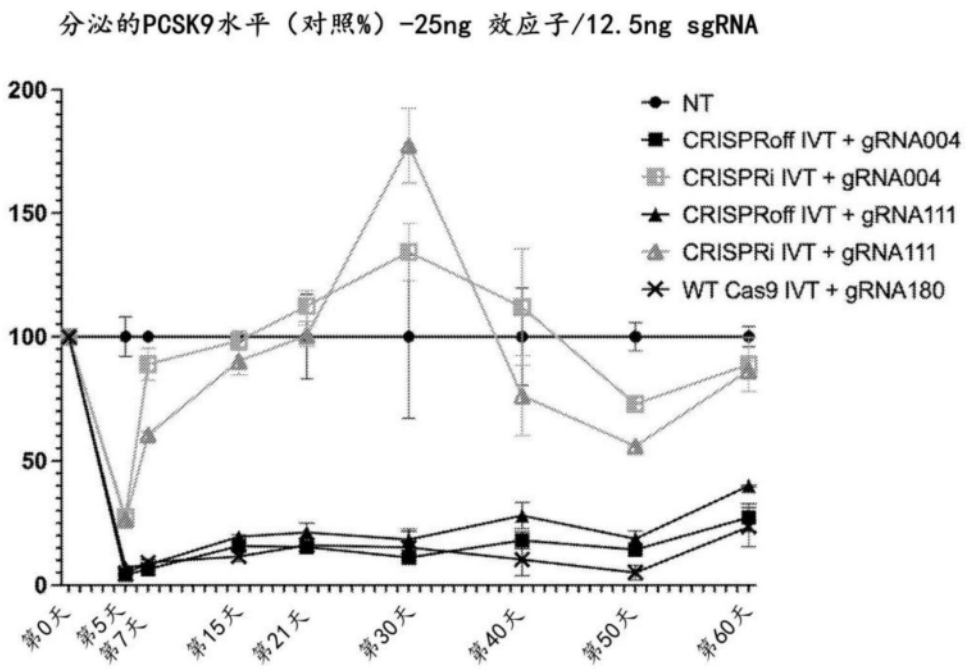

分泌的PCSK9水平（对照%）-25ng 效应子/12.5ng sgRNA
200
150
100
50
0
NT
CRISPRoff IVT + gRNA004
CRISPRi IVT + gRNA004
CRISPRoff IVT + gRNA111
CRISPRi IVT + gRNA111
WT Cas9 IVT + gRNA180
第0天
第5天
第7天
第15天
第21天
第30天
第40天
第50天
第60天

图5

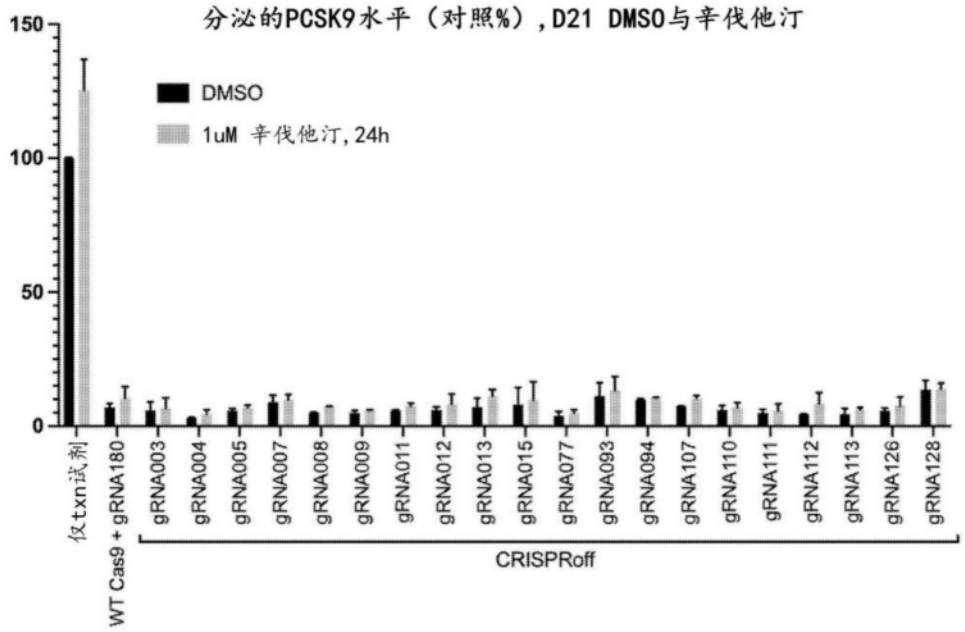

分泌的PCSK9水平（对照%），D21 DMSO与辛伐他汀
DMSO
1uM 辛伐他汀, 24h
150
100
50
0
仅txn试剂
WT Cas9 + gRNA180
gRNA003
gRNA004
gRNA005
gRNA007
gRNA008
gRNA009
gRNA011
gRNA012
gRNA013
gRNA015
gRNA077
gRNA093
gRNA094
gRNA107
gRNA110
gRNA111
gRNA112
gRNA113
gRNA126
gRNA128
CRISPRoff

图6

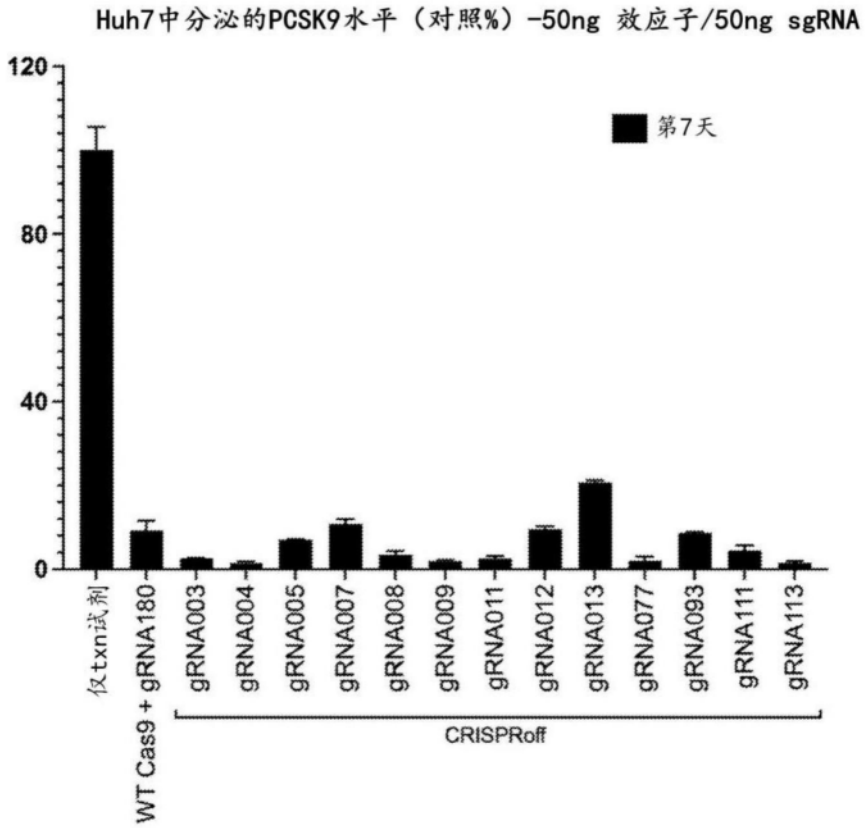

Huh7中分泌的PCSK9水平（对照%）-50ng 效应子/50ng sgRNA
第7天
120
80
40
0
仅txn试剂
WT Cas9 + gRNA180
gRNA003
gRNA004
gRNA005
gRNA007
gRNA008
gRNA009
gRNA011
gRNA012
gRNA013
gRNA077
gRNA093
gRNA111
gRNA113
CRISPRoff

图7

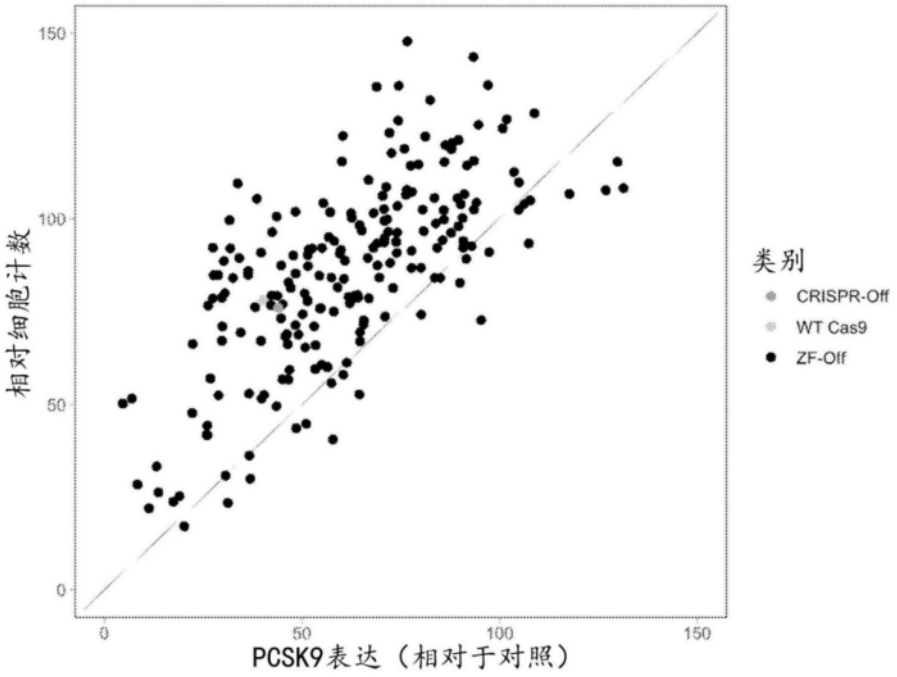

150
100
50
0
相对细胞计数
0
50
100
150
PCSK9表达（相对于对照）
类别
CRISPR-Off
WT Cas9
ZF-Off

图8

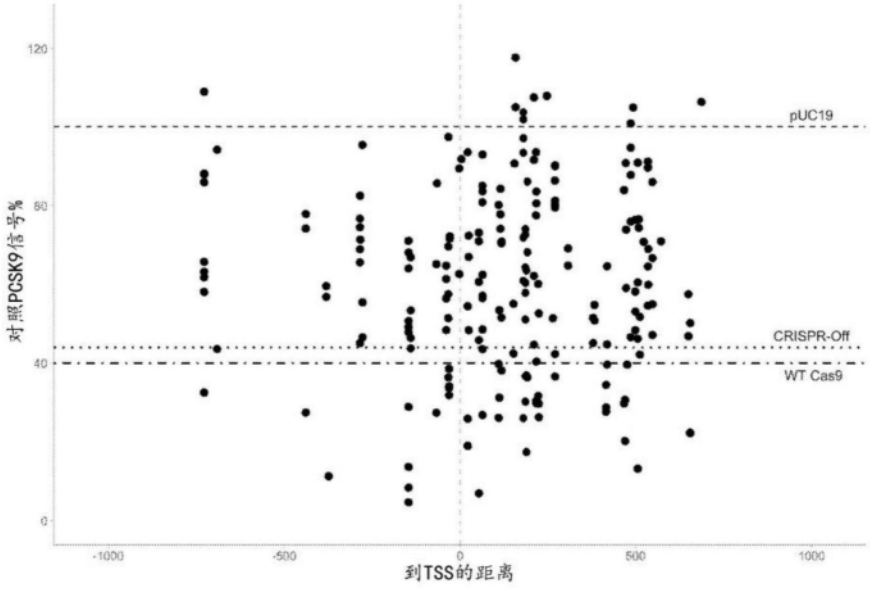

120
80
40
0
对照PCSK9信号%
pUC19
CRISPR-Off
WT Cas9
-1000
-500
0
500
1000
到TSS的距离

图9

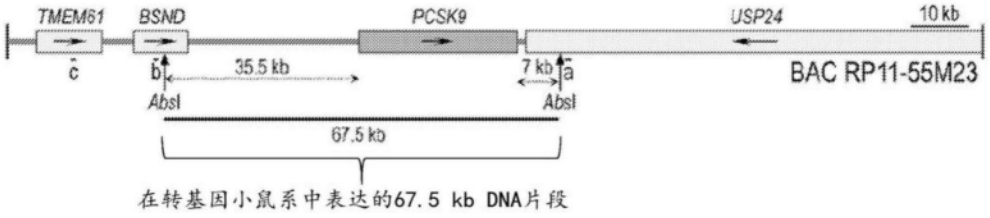


图10

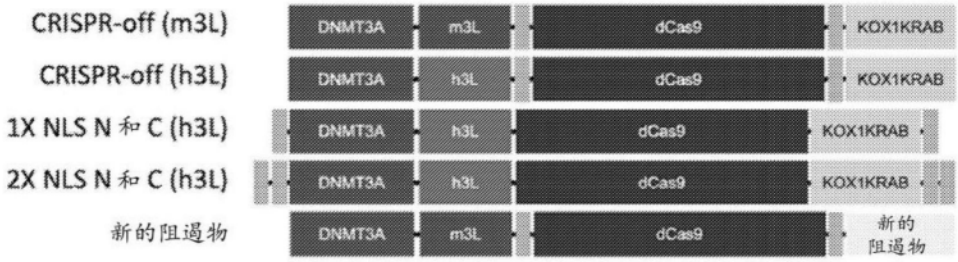


图11A

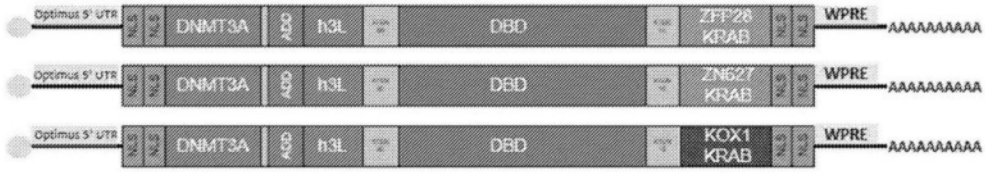


图11B

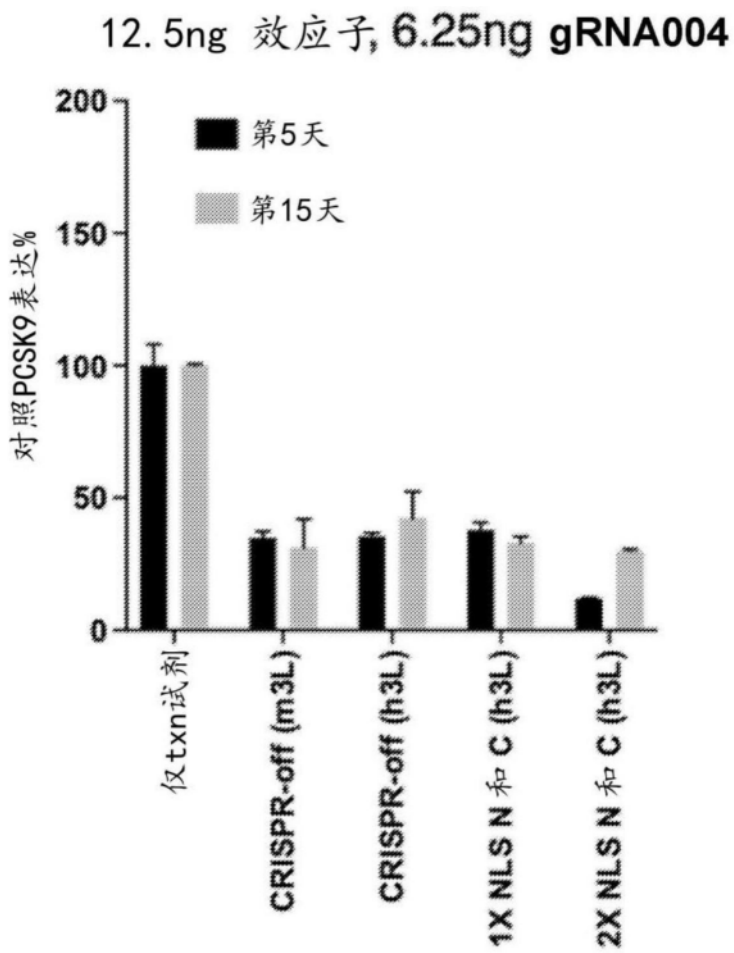

12.5ng 效应子, 6.25ng gRNA004
第5天
第15天
对照PCSK9表达%
200
150
100
50
0
仅txn试剂
CRISPR-off (m3L)
CRISPR-off (h3L)
1X NLS N 和 C (h3L)
2X NLS N 和 C (h3L)

图12A

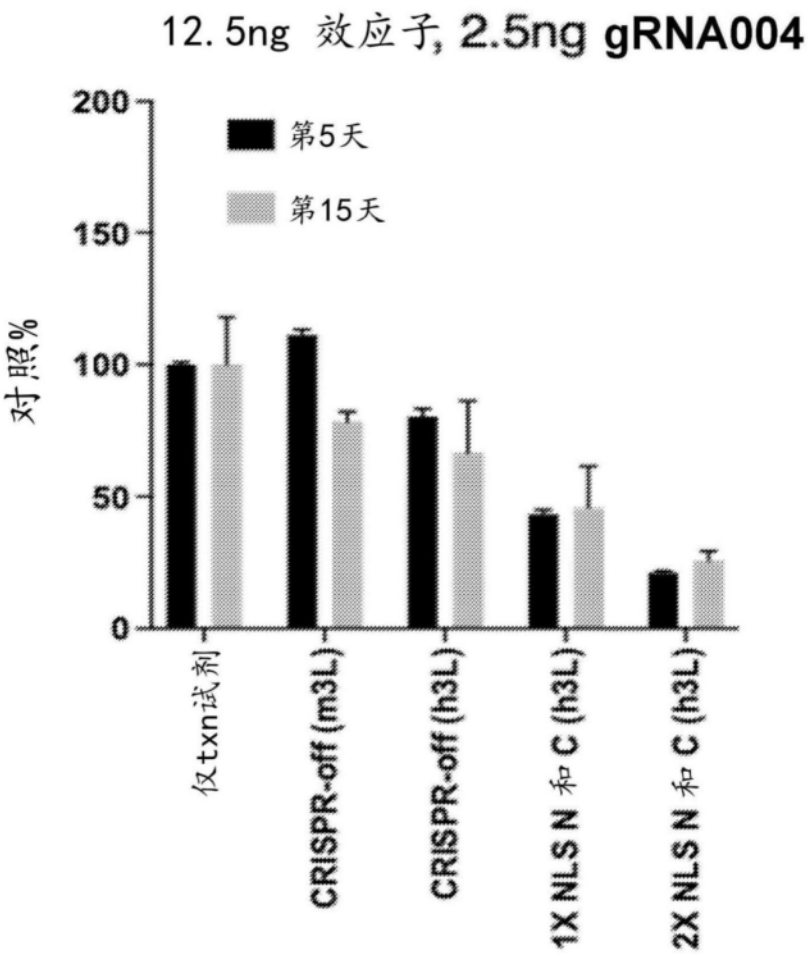

12.5ng 效应子, 2.5ng gRNA004
200
150
100
50
0
对照%
第5天
第15天
仅txn试剂
CRISPR-off (m3L)
CRISPR-off (h3L)
1X NLS N 和 C (h3L)
2X NLS N 和 C (h3L)

图12B

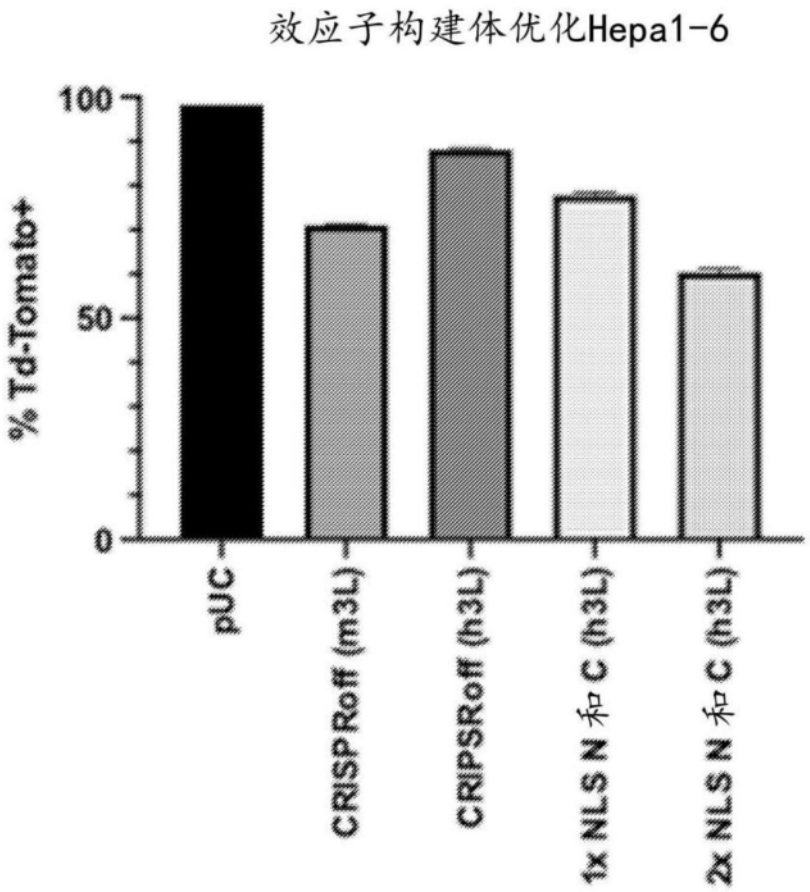

效应子构建体优化Hepa1-6
% Td-Tomato+
100
50
0
pUC
CRISPRoff (m3L)
CRIPSRoff (h3L)
1x NLS N 和 C (h3L)
2x NLS N 和 C (h3L)

图13

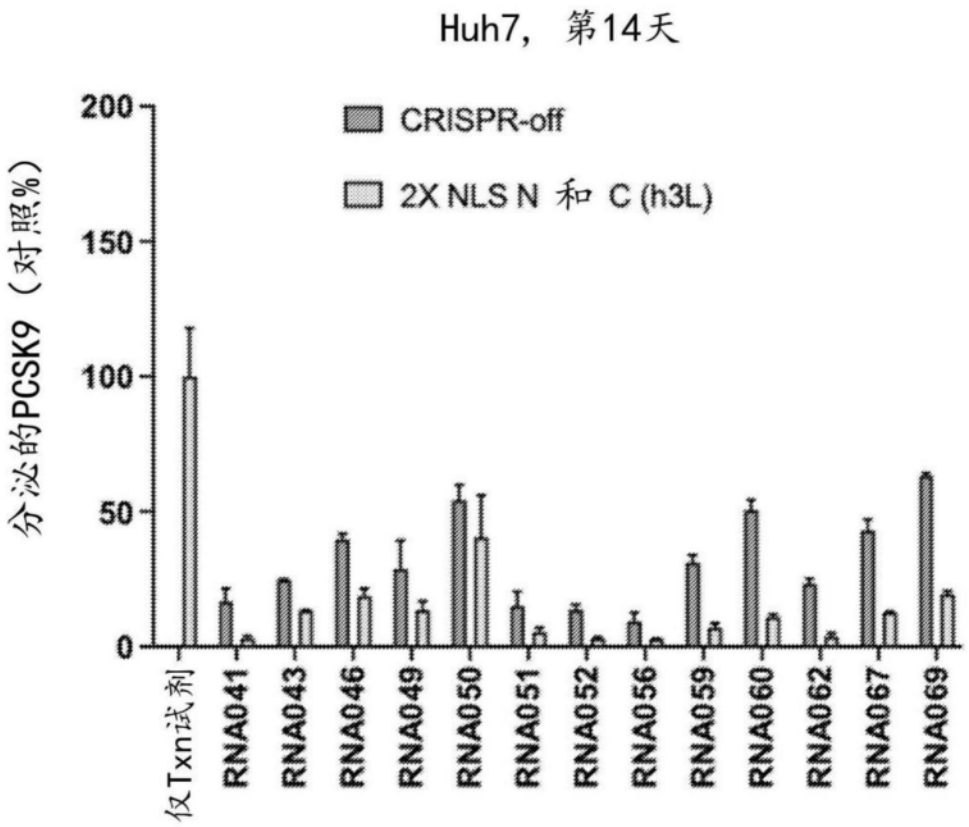

Huh7, 第14天
CRISPR-off
2X NLS N 和 C (h3L)
分泌的PCSK9（对照%）
200
150
100
50
0
仅Txn试剂
RNA041
RNA043
RNA046
RNA049
RNA050
RNA051
RNA052
RNA056
RNA059
RNA060
RNA062
RNA067
RNA069

图14A

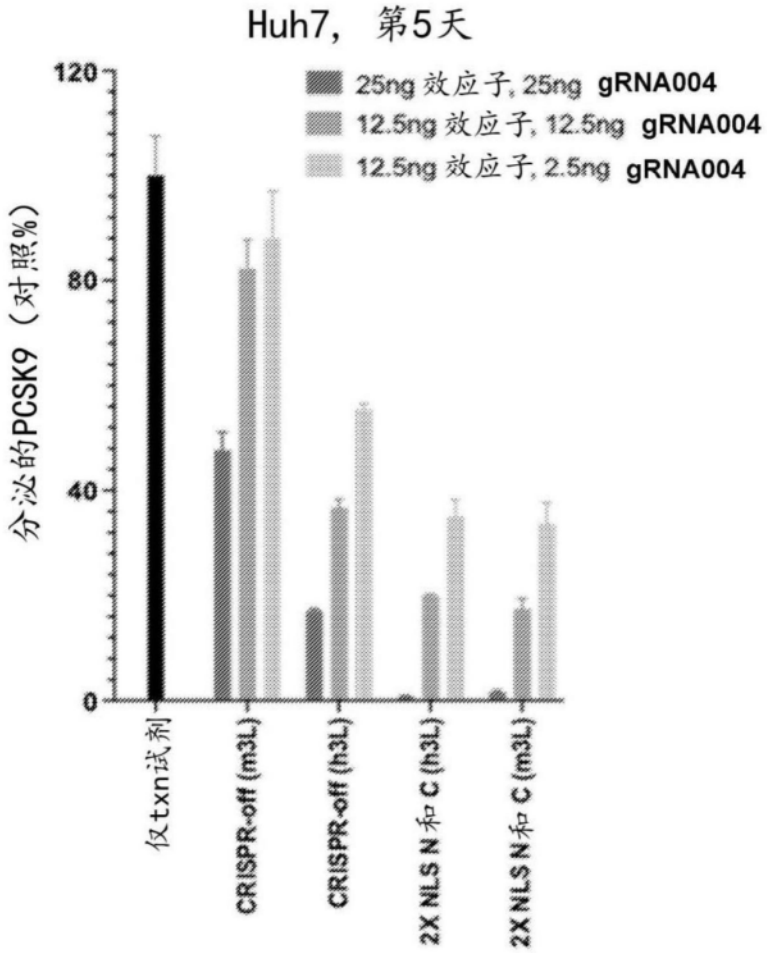

Huh7, 第5天
25ng 效应子, 25ng gRNA004
12.5ng 效应子, 12.5ng gRNA004
12.5ng 效应子, 2.5ng gRNA004
分泌的PCSK9（对照%）
120
80
40
0
仅txn试剂
CRISPR-off (m3L)
CRISPR-off (h3L)
2X NLS N和C (h3L)
2X NLS N和C (m3L)

图14B

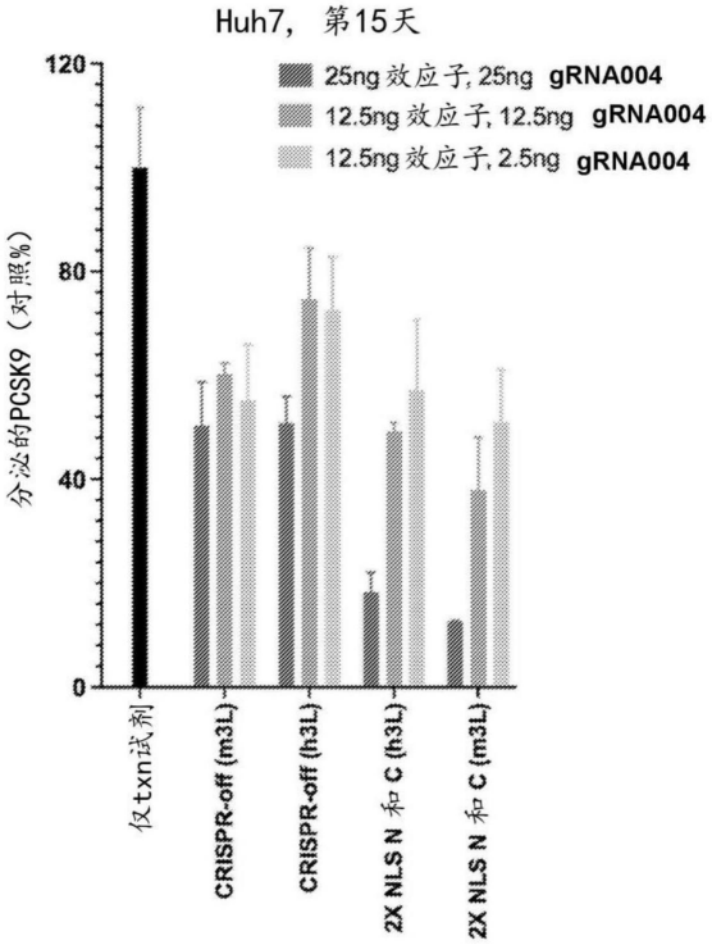

Huh7, 第15天
25ng 效应子, 25ng gRNA004
12.5ng 效应子, 12.5ng gRNA004
12.5ng 效应子, 2.5ng gRNA004
分泌的PCSK9（对照%）
120
80
40
0
仅txn试剂
CRISPR-off (m3L)
CRISPR-off (h3L)
2X NLS N 和 C (h3L)
2X NLS N 和 C (m3L)

图14C

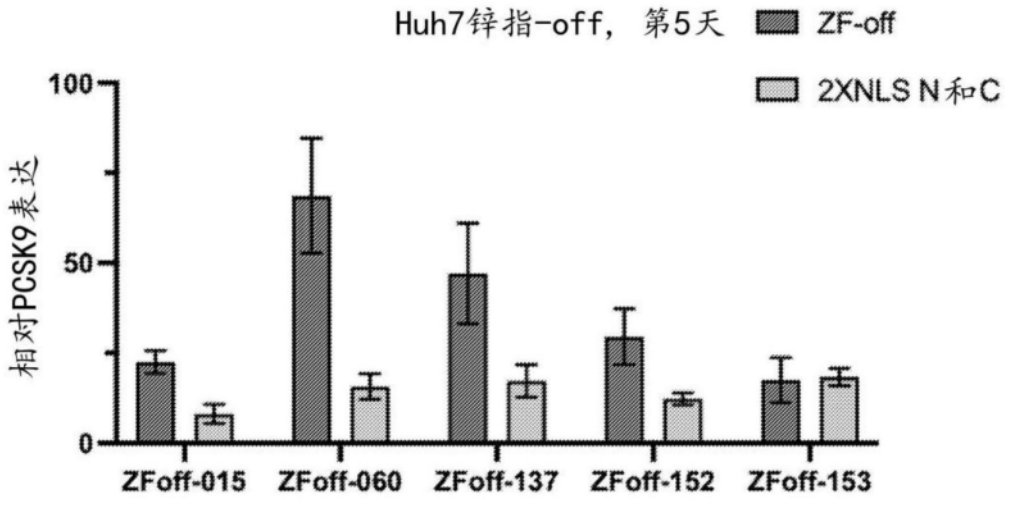

Huh7锌指-off, 第5天
ZF-off
2XNLS N和C
相对PCSK9表达
100
50
0
ZFoff-015
ZFoff-060
ZFoff-137
ZFoff-152
ZFoff-153

图15

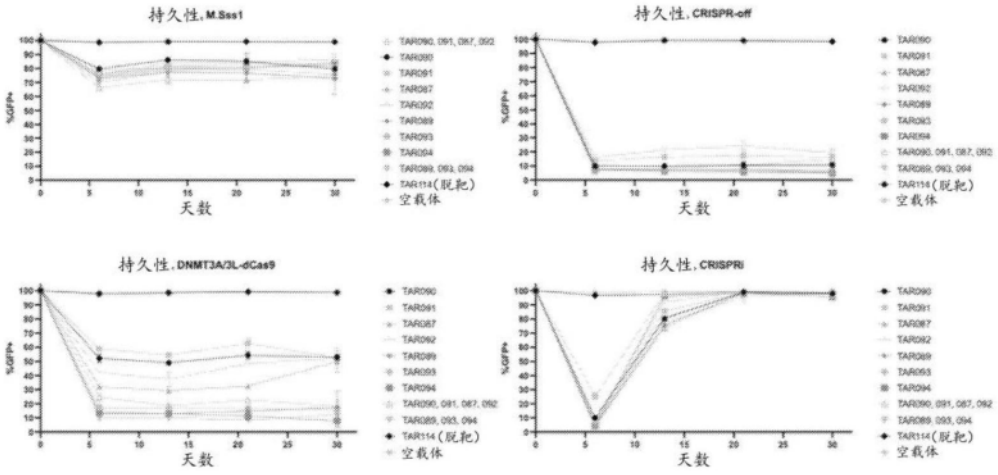

持久性, M.Sss1
持久性, CRISPR-off
持久性, DNMT3A/3L-dCas9
持久性, CRISPRi
%GFP+
天数
TAR090, 091, 087, 092
TAR090
TAR091
TAR087
TAR092
TAR089
TAR093
TAR094
TAR089, 093, 094
TAR114(脱靶)
空载体

图16

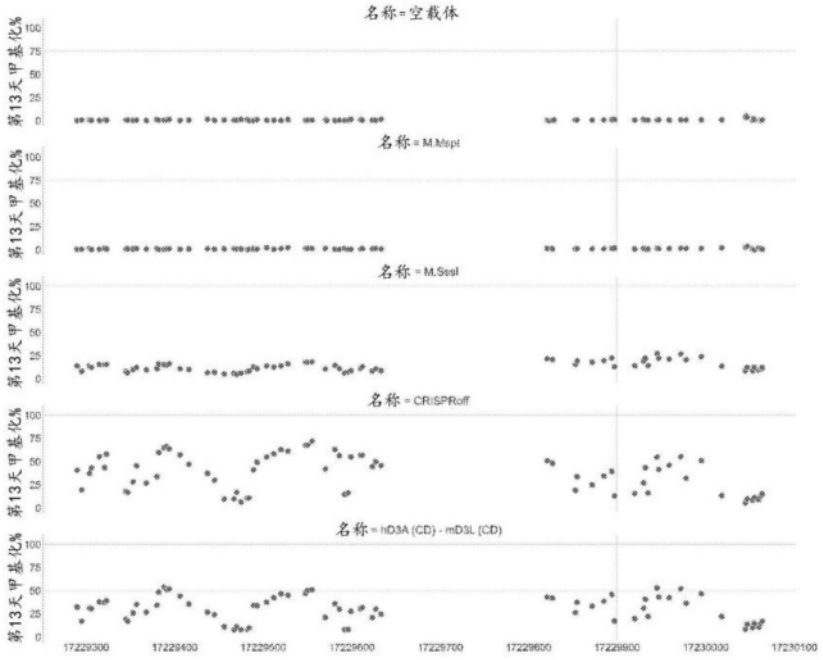

名称=空载体
名称=M.MspI
名称=M.SssI
名称=CRISPRoff
名称=hD3A (CD) - mD3L (CD)
第13天甲基化%
17229300
17229400
17229500
17229600
17229700
17229800
17229900
17230000
17230100

图17

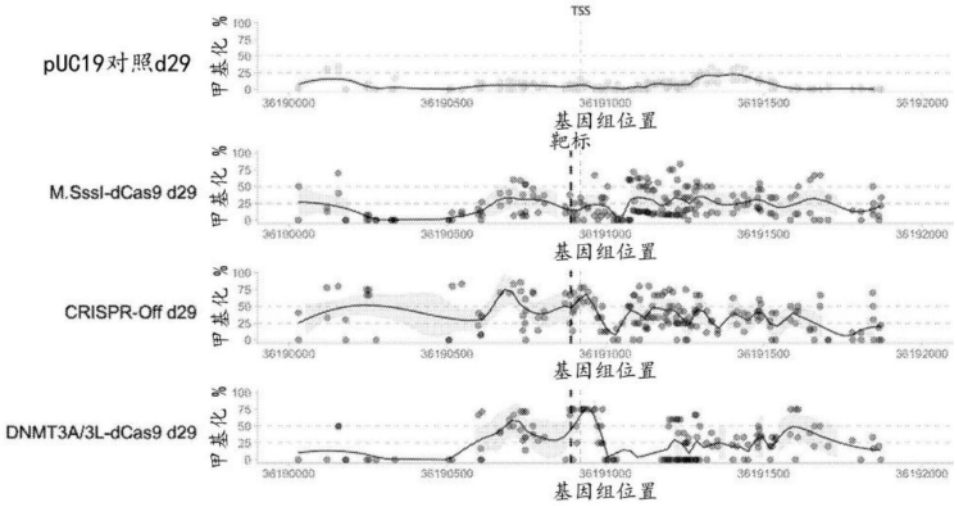

TSS
pUC19对照d29
M.SssI-dCas9 d29
CRISPR-Off d29
DNMT3A/3L-dCas9 d29
甲基化 %
基因组位置
靶标

图18

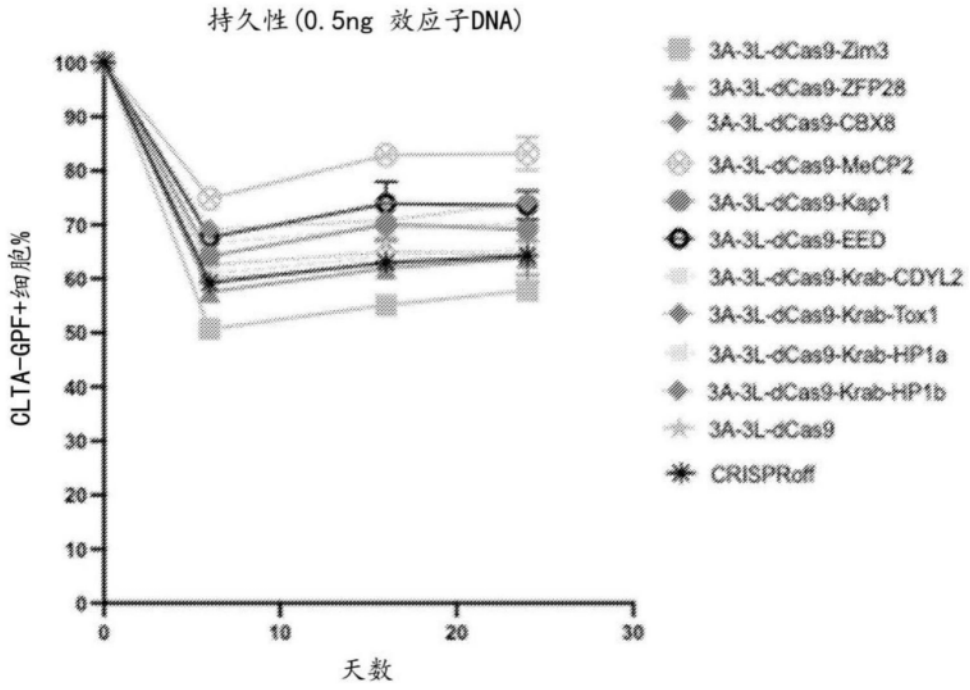

持久性(0.5ng 效应子DNA)
CLTA-GPF+细胞%
天数
3A-3L-dCas9-Zim3
3A-3L-dCas9-ZFP28
3A-3L-dCas9-C8X8
3A-3L-dCas9-MeCP2
3A-3L-dCas9-Kap1
3A-3L-dCas9-EED
3A-3L-dCas9-Krab-CDYL2
3A-3L-dCas9-Krab-Tox1
3A-3L-dCas9-Krab-HP1a
3A-3L-dCas9-Krab-HP1b
3A-3L-dCas9
CRISPRoff

图19A

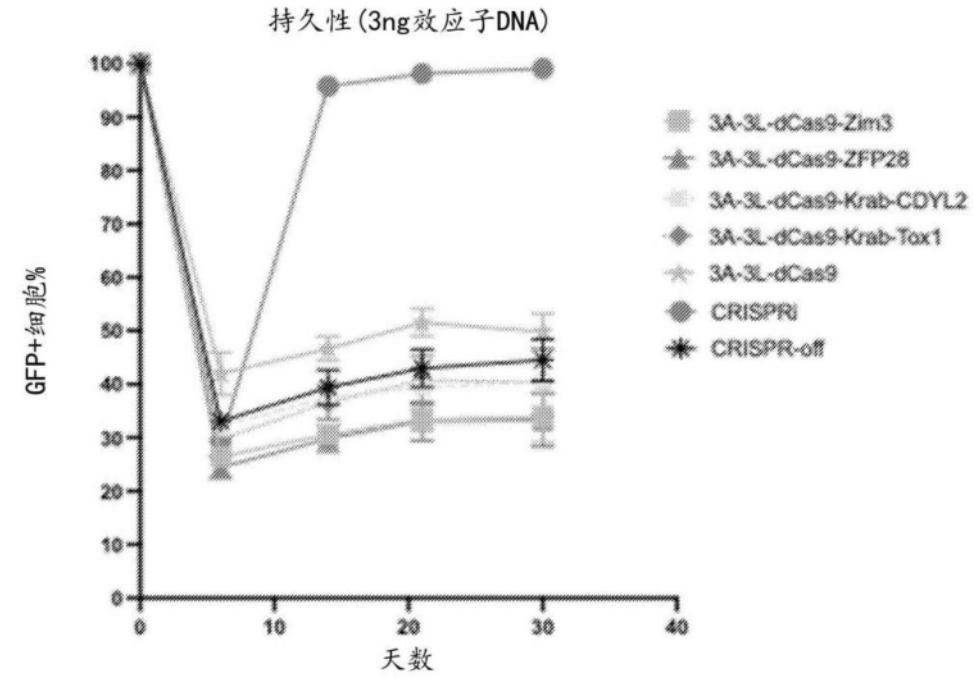

持久性(3ng效应子DNA)
GFP+细胞%
天数
3A-3L-dCas9-Zim3
3A-3L-dCas9-ZFP28
3A-3L-dCas9-Krab-CDYL2
3A-3L-dCas9-Krab-Tox1
3A-3L-dCas9
CRISPRi
CRISPR-off

图19B

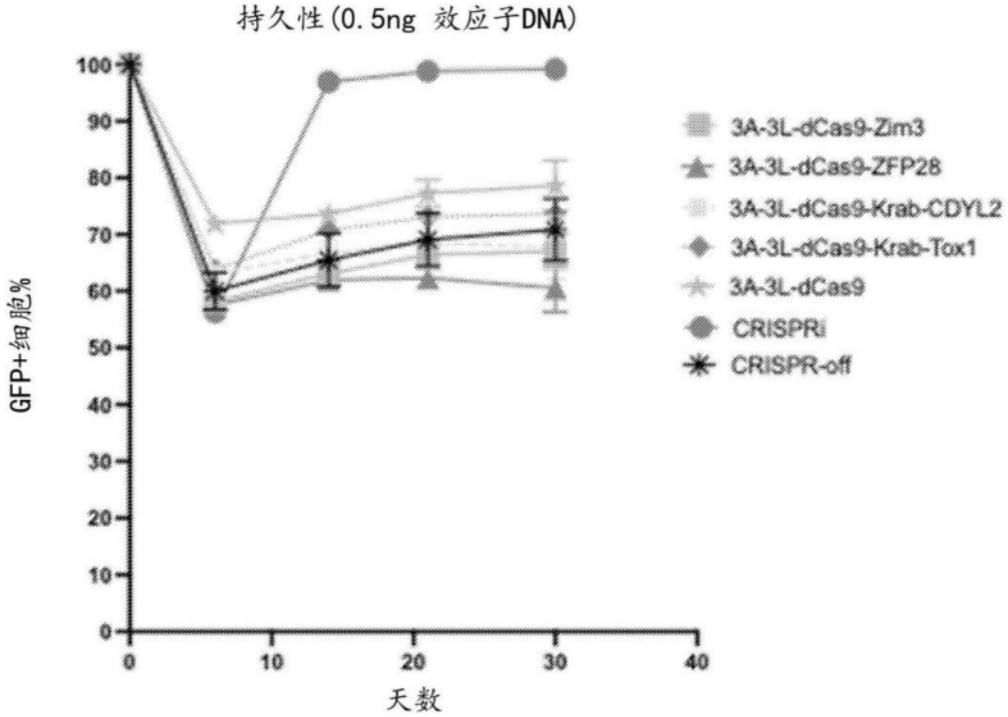

持久性(0.5ng 效应子DNA)
GFP+细胞%
天数
3A-3L-dCas9-Zim3
3A-3L-dCas9-ZFP28
3A-3L-dCas9-Krab-CDYL2
3A-3L-dCas9-Krab-Tox1
3A-3L-dCas9
CRISPRi
CRISPR-off

图19C

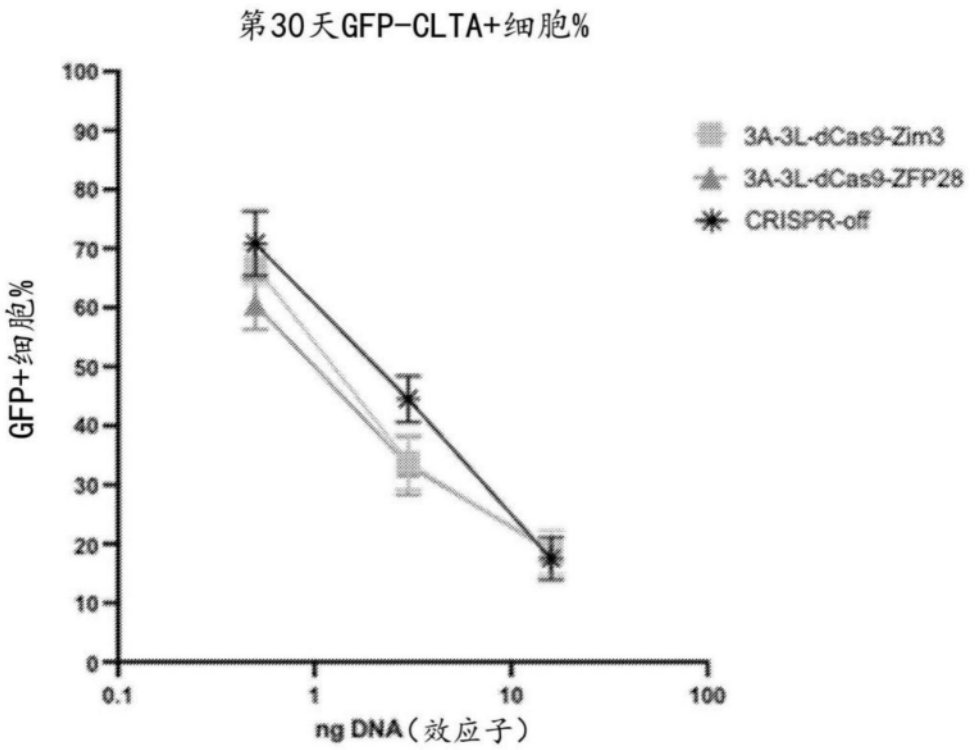

第30天GFP-CLTA+细胞%
GFP+细胞%
100
90
80
70
60
50
40
30
20
10
0
0.1
1
10
100
ng DNA（效应子）
3A-3L-dCas9-Zim3
3A-3L-dCas9-ZFP28
CRISPR-off

图19D