



(12) 发明专利申请

(10) 申请公布号 CN 105408483 A

(43) 申请公布日 2016. 03. 16

(21) 申请号 201480026276. 5

(74) 专利代理机构 北京市柳沈律师事务所
11105

(22) 申请日 2014. 03. 14

代理人 张文辉

(30) 优先权数据

61/799, 647 2013. 03. 15 US

61/838, 178 2013. 06. 21 US

61/838, 148 2013. 06. 21 US

61/921, 007 2013. 12. 26 US

(51) Int. Cl.

C12N 15/62(2006. 01)

(85) PCT国际申请进入国家阶段日

2015. 11. 09

(86) PCT国际申请的申请数据

PCT/US2014/027335 2014. 03. 14

(87) PCT国际申请的公布数据

W02014/152432 EN 2014. 09. 25

(71) 申请人 通用医疗公司

地址 美国马萨诸塞州

(72) 发明人 J. K. 乔昂格 M. 梅德

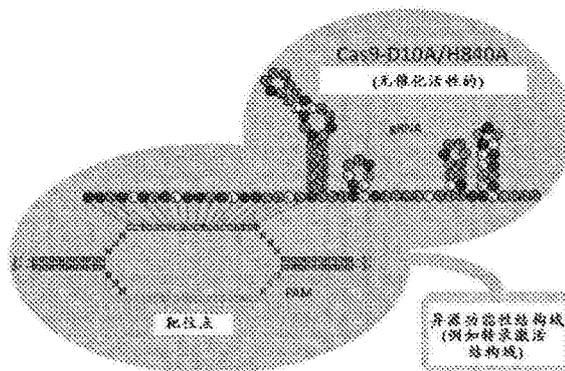
权利要求书2页 说明书25页 附图17页

(54) 发明名称

遗传和表观遗传调节蛋白至特定基因组基因座的 RNA 引导的靶向

(57) 摘要

用于将异源功能性结构域诸如转录激活物 RNA 引导靶向特定基因组基因座的方法和构建体。本发明涉及用于将基因和表观基因调控蛋白，例如转录及货物，组蛋白修饰酶，DNA 甲基化修饰物，RNA 引导靶向特定的基因座的方法和构建物。本发明至少部分基于融合蛋白的开发，包括融合到 Cas9 核酸酶的异源功能性结构域（例如，转录激活结构域），所述 Cas9 核酸酶通过突变具有其核酸酶活性失活（也称为“dCas9”）。



1. 一种融合蛋白,其包含连接于异源功能性结构域的无催化活性的 CRISPR 相关 9(dCas9) 蛋白。
2. 根据权利要求 1 所述的融合蛋白,其中所述异源功能性结构域是转录激活结构域。
3. 根据权利要求 2 所述的融合蛋白,其中所述转录激活结构域来自 VP64 或 NF- κ B p65。
4. 根据权利要求 1 所述的融合蛋白,其中所述异源功能性结构域是转录沉默子或转录抑制结构域。
5. 根据权利要求 4 所述的融合蛋白,其中所述转录抑制结构域是 Krueppel 相关盒 (KRAB) 结构域、ERF 抑制结构域 (ERD) 或 mSin3A 相互作用结构域 (SID)。
6. 根据权利要求 4 所述的融合蛋白,其中所述转录沉默子是异染色质蛋白 1 (HP1),例如 HP1 α 或 HP1 β 。
7. 根据权利要求 1 所述的融合蛋白,其中所述异源功能性结构域是修饰 DNA 的甲基化状态的酶。
8. 根据权利要求 7 所述的融合蛋白,其中修饰 DNA 的甲基化状态的酶是 DNA 甲基转移酶 (DNMT) 或 TET 蛋白。
9. 根据权利要求 8 所述的融合蛋白,其中所述 TET 蛋白是 TET1。
10. 根据权利要求 1 所述的融合蛋白,其中所述异源功能性结构域是修饰组蛋白亚单位的酶。
11. 根据权利要求 1 所述的融合蛋白,其中修饰组蛋白亚单位的酶是组蛋白乙酰转移酶 (HAT)、组蛋白脱乙酰基酶 (HDAC)、组蛋白甲基转移酶 (HMT) 或组蛋白脱甲基酶。
12. 根据权利要求 1 所述的融合蛋白,其中所述异源功能性结构域是生物系链。
13. 根据权利要求 12 所述的融合蛋白,其中所述生物系链是 MS2、Csy4 或 λ N 蛋白。
14. 根据权利要求 1 所述的融合蛋白,其中所述无催化活性的 Cas9 蛋白来自化脓性链球菌。
15. 根据权利要求 1 所述的融合蛋白,其中所述无催化活性的 Cas9 蛋白在 D10、E762、H983 或 D986 ;以及在 H840 或 N863 处包含突变。
16. 根据权利要求 15 所述的融合蛋白,其中所述突变是:
 - (i) D10A 或 D10N, 和
 - (ii) H840A、H840N 或 H840Y。
17. 根据权利要求 1 所述的融合蛋白,其中通过任选的间插接头将所述异源功能性结构域连接于无催化活性的 Cas9 蛋白的 N 末端或 C 末端,其中所述接头不干扰所述融合蛋白的活性。
18. 根据权利要求 1 所述的融合蛋白,其还任选地通过一个或多个间插接头在 N 末端、C 末端上和 / 或在所述无催化活性的 CRISPR 相关 9 (Cas9) 蛋白与所述异源功能性结构域之间中包含核定位序列和一个或多个表位标签的一者或两者。
19. 根据权利要求 18 所述的融合蛋白,其中所述表位标签是 c-myc、6His 或 FLAG。
20. 一种核酸,其编码权利要求 1-19 的任一项的融合蛋白。
21. 一种表达载体,其包含权利要求 20 的核酸。
22. 一种增加靶基因在细胞中的表达的方法,所述方法包括在所述细胞中表达权利要

求 2-3 的融合蛋白,和一个或多个被导向所述靶基因的引导 RNA。

23. 一种减少细胞中的靶基因的方法,所述方法包括在所述细胞中表达权利要求 4-6 的融合蛋白,和一个或多个被导向所述靶基因的引导 RNA。

24. 一种减少细胞中的靶基因或其启动子或增强子的 DNA 甲基化的方法,所述方法包括在所述细胞中表达权利要求 7-9 的融合蛋白,和一个或多个被导向所述相关靶基因序列的引导 RNA。

25. 一种修饰细胞中与靶基因或其启动子或增强子相关的组蛋白的方法,所述方法包括在所述细胞中表达权利要求 10-11 的融合蛋白,和一个或多个被导向所述相关靶基因序列的引导 RNA。

遗传和表观遗传调节蛋白至特定基因组基因座的 RNA 引导的靶向

[0001] 优先权的声明

[0002] 本申请要求 2013 年 3 月 15 日提交的美国专利申请系列第 61/799,647 号、2013 年 6 月 21 日提交的美国专利申请系列第 61/838,178 号、2013 年 6 月 21 日提交的美国专利申请系列第 61/838,148 号和 2013 年 12 月 26 日提交的美国专利申请系列第 61/921,007 号的权益。前述专利申请系列号的完整内容在此通过引用并入。

[0003] 联邦资助的研究或开发

[0004] 本发明是在由美国国家卫生研究院授予的基金第 DP1GM105378 号和由国防部的国防高等研究计划署 (DARPA) 授予的 W911NF-11-2-0056 下借助政府资助进行的。政府具有本发明的某些权利。

技术领域

[0005] 本发明涉及用于将遗传和表观遗传调节蛋白,例如,转录激活物、组蛋白修饰酶、DNA 甲基化修饰剂 RNA 引导靶向至特定基因组基因座的方法和组合物。

[0006] 背景

[0007] 被称为 CRISPR/Cas 系统的成簇规律间隔短回文重复 (CRISPR) 和 CRISPR 相关 (cas) 基因被不同细菌和古细菌 (archaea) 用来介导抵抗病毒和其它外来核酸的防御。这些系统使用小 RNA 来以序列特异性方式检测和沉默外来核酸。

[0008] 已描述了 3 个类型的 CRISPR/Cas 系统 (Makarova 等, Nat. Rev. Microbiol. 9, 467 (2011); Makarova 等, Biol. Direct 1, 7 (2006); Makarova 等, Biol. Direct 6, 38 (2011))。最近的工作已显示 II 型 CRISPR/Cas 系统可被工程化来通过使用与 DNA 靶位点具有互补性的单一“引导 RNA”和 Cas9 核酸酶在体外将靶向双链 DNA 断裂导向特定序列 (Jinek 等, Science 2012 ;337:816 - 821)。该可靶向的基于 Cas9 的系统也在培养的人细胞中 (Mali 等, Science. 2013 年 2 月 15 日 ;339(6121):823-6; Cong 等, Science. 2013 年 2 月 15 日 ;339(6121):819-23) 以及在斑马鱼中在体内起作用 (Hwang 和 Fu 等, Nat Biotechnol. 2013 年 3 月 ;31(3):227-9) 以诱导至内源基因内的靶向改变。

[0009] 概述

[0010] 本发明至少部分基于融合蛋白的开发,所述融合蛋白包括与其核酸酶活性已通过突变灭活的 Cas9 核酸酶 (也称为“dCas9”) 融合的异源功能性结构域 (例如,转录激活结构域)。虽然公布的研究已使用引导 RNA 将具有催化活性但失活的 Cas9 核酸酶蛋白靶向特定基因组基因座,但还没有工作改造该系统来适用于招募另外的效应子结构域。该工作还提供了导致靶基因的表达水平的升高 (而非降低) 的 RNA 引导的过程的首次例证。

[0011] 另外,本公开还提供了多重 gRNA 可用于将多种 dCas9-VP64 融合物带至单个启动子,从而导致转录的协同激活的首次例证。

[0012] 因此,在第一方面,本发明提供包含连接于异源功能性结构域 (HFD) 的无催化活性的 CRISPR 相关 9 (dCas9) 蛋白的融合蛋白,所述异源功能性结构域修饰基因表达、组蛋白

或 DNA, 例如转录激活结构域、转录阻遏物 (例如, 沉默子诸如异染色质蛋白 1 (HP1), 例如 HP1 α 或 HP1 β , 或转录抑制结构域, 例如, Krueppel 相关盒 (KRAB) 结构域、ERF 抑制结构域 (ERD) 或 mSin3A 相互作用结构域 (SID))、修饰 DNA 的甲基化状态的酶 (例如, DNA 甲基转移酶 (DNMT) 或 10-11 易位 (TET) 蛋白, 例如, 也称为 Tet 甲基胞嘧啶双加氧酶 1 的 TET1) 或修饰组蛋白亚单位的酶 (例如, 组蛋白乙酰转移酶 (HAT)、组蛋白脱乙酰酶 (HDAC) 或组蛋白脱甲基酶)。在一些实施方案中, 所述异源功能性结构域是转录激活结构域, 例如, 来自 VP64 或 NF- κ B p65 的转录激活结构域; 催化 DNA 脱甲基化的酶, 例如, TET; 或组蛋白修饰 (例如, LSD1、组蛋白甲基转移酶、HDAC 或 HAT) 或转录沉默结构域, 例如, 来自异染色质蛋白 1 (HP1), 例如, HP1 α 或 HP1 β ; 或生物系链, 例如 CRISPR/Cas 亚型 Ypest 蛋白 4 (Csy4)、MS2 或 λ N 蛋白。

[0013] 在一些实施方案中, 无催化活性的 Cas9 蛋白来自化脓性链球菌 (*S. pyogenes*)。

[0014] 在一些实施方案中, 无催化活性的 Cas9 蛋白在 D10、E762、H983 或 D986; 以及在 H840 或 N863, 例如在 D10 和 H840 处包含突变, 例如 D10A 或 D10N 和 H840A 或 H840N 或 H840Y。

[0015] 在一些实施方案中, 所述异源功能性结构域通过任选的间插接头连接于无催化活性的 Cas9 蛋白的 N 末端或 C 末端, 其中所述接头不干扰融合蛋白的活性。

[0016] 在一些实施方案中, 所述融合蛋白在 N 末端、C 末端或无催化活性的 CRISPR 相关 9 (Cas9) 蛋白与异源功能性结构域之间中包括核定位序列和一个或多个表位标签 (例如, c-myc、6His 或 FLAG 标签) 的任一者或两者, 任选地具有一个或多个间插接头。

[0017] 在其它方面, 本发明提供编码本文中描述的融合蛋白的核酸以及包含所述核酸的表达载体, 和表达所述融合蛋白的宿主细胞。

[0018] 在另外的方面, 本发明提供用于增加靶基因在细胞中的表达的方法。所述方法包括例如通过细胞与包括编码所述融合蛋白的序列的表达载体接触在所述细胞中表达如本文所述 Cas9-HFD 融合蛋白, 以及还例如通过将细胞与一个或多个包含编码一个或多个引导 RNA 的核酸序列的表达载体接触来在细胞中表达一个或多个具有针对靶基因的互补性的引导 RNA。

[0019] 除非另有定义, 否则本文所用的所有技术和 / 或科学术语具有与本发明所属领域的普通技术人员通常所理解的含义相同的含义。本文中描述了用于本发明的方法和材料; 还可使用本领域中已知的其它合适的方法和材料。所述材料、方法和实施例仅为示例性的并非旨在限制。本文中提及的所有出版物、专利申请、专利、序列、数据库条目和其它参考文献通过引用整体并入。如发生矛盾, 则以本专利说明书包括定义为准。

[0020] 根据以下详细说明和附图以及权利要求, 本发明特征和有利将是显而易见的。

[0021] 附图概述

[0022] 本专利或申请文件含有至少一个以彩色绘制的附图。具有彩色附图的本专利或专利申请公布的拷贝可在请求和支付必要的费用后由专利局提供。

[0023] 图 1A 是显示单一引导 RNA (sgRNA) 将 Cas9 核酸酶招募至特定 DNA 序列, 从而引入靶向改变的示意图。显示的引导 RNA 的序列为 GGAGCGAGCGGAGCGGUACAGUUUUAGAGCUAGAAAUAGCAAGUAAAAUAAGGCUAGUCCG (SEQ ID NO:9)。

[0024] 图 1B 是显示用于将 Cas9 核酸酶招募至特定 DNA 序列, 从而引入靶向改变的 sgRNA 的较长形式的示意图。显示的引导 RNA 的序列为 GGAGCGAGCGGAGCGGUACAGUUUUAGAGCUA

GAAAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU (SEQ ID NO:10)。

[0025] 图 1C 是显示与转录激活结构域融合并被 sgRNA 招募至特定 DNA 序列的 Cas9 蛋白, 所述 Cas9 蛋白含有 D10A 和 H840A 突变以使得蛋白质的核酸酶部分无催化活性。显示的引导 RNA 的序列为 GGAGCGAGCGGAGCGGUACAGUUUUAGAGCUAGAAAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU (SEQ ID NO:10)。

[0026] 图 1D 是描绘 dCas9-VP64 融合蛋白被嵌合 sgRNA 招募至特定基因组靶序列的示意图。

[0027] 图 1E 是举例说明 16 个 sgRNA 被靶向内源人 VEGFA 基因启动子的位置和取向的简图。小的水平箭头代表与基因组 DNA 序列互补的 gRNA 的前 20nt, 箭头从 5' 指向 3'。灰色条指示先前在人 293 细胞中确定的相对于转录起始位点 (直角箭头) 编号的 DNA 酶 I 高敏位点 (Liu 等, J Biol Chem. 2001 年 4 月 6 日 ;276(14):11323-34)。

[0028] 图 2A 是显示各种 sgRNA 对 293 细胞中的 VEGFA 蛋白表达的激活的条线图, 每种与 (灰色条块) 或不与 (黑色条块) dCas9-VP64 一起表达。计算相对于如方法中描述的脱靶 sgRNA 对照的 VEGFA 的激活倍数。以一式三份进行每一个实验, 误差条代表平均值的标准误差。星号表示如通过配对单侧 t 检验 ($p < 0.05$) 测定的被显著升高高于脱靶对照的样品。

[0029] 图 2B 是显示多重 sgRNA 表达通过 dCas9-VP64 蛋白诱导 VEGFA 蛋白表达的协同激活的条线图。显示其中将指定的 sgRNA 组合与 dCas9-VP64 共表达的 293 细胞中的 VEGFA 蛋白的激活倍数。注意, 在所有这些实验中, 每一种用于转染的单独的 sgRNA 表达质粒的量相同。如 2A 中所述计算激活倍数并以灰色条块显示。将每一个组合的由单独的 sgRNA 诱导的平均激活倍数值的计算的总和显示为黑色条块。星号表示被发现显著大于如通过方差分析 (ANOVA) ($p < 0.05$) 测定的预期总和的所有组合。

[0030] 图 3A 是举例说明被靶向内源人 NTF3 基因启动子的 6 个 sgRNA 的位置和取向的简图。水平箭头代表与基因组 DNA 序列互补的 sgRNA 的前 20nt, 箭头从 5' 指向 3'。灰线表示从 UCSC 基因组浏览器上的 ENCODE DNA 酶 I 超敏反应跟踪鉴定的潜在开放染色质的区域, 条的更粗部分表示首先转录的外显子。显示的编号是相对于转录起始位点的 (+1, 直角箭头)。

[0031] 图 3B 是显示 293 细胞中 sgRNA 引导的 dCas9-VP64 对 NTF3 基因表达的激活的条线图。显示了用指定量的 dCas9-VP64 和 NTF3 靶向 sgRNA 表达质粒共转染的 293 细胞的通过定量 RT-PCR 检测的并针对 GAPDH 对照 ($\Delta Ct \times 10^4$) 标准化的 NTF3 mRNA 的相对表达。以一式三份进行所有实验, 误差条表示平均值的标准误差。星号表示如通过配对单侧 T 检验 ($P < 0.05$) 测定比脱靶 gRNA 对照显著更大的样品。

[0032] 图 3C 是显示多重 gRNA 表达通过 dCas9-VP64 蛋白诱导 NTF3 mRNA 表达的协同激活。显示了用 dCas9-VP64 和指定的 NTF3 靶向 gRNA 表达质粒的组合共转染的 293 细胞的通过定量 RT-PCR 检测的并针对 GAPDH 对照 ($\Delta Ct \times 10^4$) 标准化的 NTF3 mRNA 的相对表达。注意, 在所有这些实验中, 每一个用于转染的单独的 gRNA 表达质粒的量相同。以一式三份进行所有实验, 误差条表示平均值的标准误差。显示了每一个组合的由单独的 gRNA 诱导的平均激活倍数值的计算的总和。

[0033] 图 4 是 sgRNA 表达载体的示例性序列。

[0034] 图 5 是 CMV-T7-Cas9D10A/H840A-3XFLAG-VP64 表达载体的示例性序列。

[0035] 图 6 是 CMV-T7-Cas9 记录的 D10A/H840A-3XFLAG-VP64 表达载体的示例性序列。

[0036] 图 7 是 Cas9-HFD 即 Cas9- 激活物的示例性序列。任选的 3xFLAG 序列加以下划线；核定位信号 PKKKRKVS (SEQ ID NO:11) 以小写显示；两个接头以粗体显示；以及 VP64 转录激活物序列 DALDDFDLMLGSDALDDFDLMLGSDALDDFDLMLGSDALDDFDLML (SEQ ID NO:12) 加框表示。

[0037] 图 8A-8B 是 (8A)dCas9-NLS-3XFLAG-HP1 α 和 (8B)dCas9-NLS-3XFLAG-HP1 β 的示例性序列。框 = 核定位信号；下划线 = 三重 flag 标签；双下划线 = HP1 α 铰链和 chromoshadow 结构域。

[0038] 图 9 为 dCas9-TET1 的示例性序列。

[0039] 图 10 是显示利用各种 dCas9-VP64 融合构建体获得的结果的条线图。所测试的那些构建体中,最优化的 dCas9-VP64 体系结构包括 N 末端 NLS (NFN) 和位于 dCas9 与 VP64 之间的另外的 NLS (N) 或 FLAG 标签 /NLS (NF)。VEGFA 基因在人 HEK293 细胞中的表达通过由 RNA 引导的 dCas9-VP64 融合物介导的转录激活来激活。将编码 dCas9-VP64 的变体的表达质粒与表达 3 个靶向 VEGFA 起始密码子的上游区域中的位点的 gRNA (在本实验中,从单一 gRNA 表达 gRNA 并且通过 Csy4 内切核糖核酸酶加工出所述 gRNA) 的质粒共转染。VEGFA 蛋白表达通过 ELISA 来测量,并且显示了两个重复的平均值,误差东躲西藏表示平均值的标准误差。

[0040] 图 11A-B 是显示具有使 Cas9 催化功能灭活的替代性取代突变的 dCas9-VP64 激活物的活性的条线图。(11A) 将表达具有对残基 D10 和 H840 的各种 Cas9 灭活取代的 dCas9-VP64 蛋白的质粒各自与单一 gRNA 或 3 个不同的靶向 VEGFA 上游区域的靶向 gRNA (分别为蓝色和红色条块) 共转染进 HEK293 细胞。(11B) 还将表达这些 dCas9-VP64 变体的质粒转染进稳定地表达单个 VEGFFA 靶向 gRNA 的 HEK293 细胞系。通过 ELISA 测定 VEGFA 蛋白水平,显示了两个重复的平均值和平均值的标准误差 (误差条)。

[0041] 详述

[0042] 本文中描述了与 Cas9 蛋白的无催化活性的形式融合 (出于使得能够在细胞和活生物体中将这功能性结构域 RNA 引导地靶向特定基因组位置的目的) 的异源功能性结构域 (例如,转录激活结构域) 的融合蛋白。

[0043] CRISPR/Cas 系统在细菌中已进化为保护细菌免受侵入质粒和病毒伤害的防御机制。来源于外来核酸的短的原型间隔序列被整合在 CRISPR 基因座中,随后被翻译和加工成短的 CRISPR RNA (crRNA)。这些与第二 tracrRNA 复合的 crRNA 随后使用它们的与侵入核酸互补的序列来引导 Cas9 介导的断裂,和随后外来核酸的破坏。在 2012 年, Doudna 和同事证明由 crRNA 与 tracrRNA 的融合物组成的单一引导 RNA (sgRNA) 可在体外介导 Cas9 核酸酶至特定 DNA 序列的招募 (图 1C;Jinek 等, Science 2012)。

[0044] 最近, sgRNA 的更长形式已被用于在人细胞和斑马鱼中引入靶向改变 (图 1B; Mali 等 Science 2013, Hwang 和 Fu 等, Nat Biotechnol. 2013 年 3 月;31(3):227-9)。Qi 等证明 Cas9 的无催化活性突变形式 (被称为 dCas9) 的 gRNA 介导的招募可导致大肠杆菌 (E. coli) 中的特定内源基因以及人细胞中的 EGFP 报告基因的抑制 (Qi 等, Cell 152, 1173 - 1183 (2013))。虽然本研究显示采用 RNA 引导的 Cas9 技术调控基因表达的潜能,

但其不测试或显示异源功能性结构域（例如 - 转录激活结构域）是否可被融合于 dCas9 而不破坏其被可编程 sgRNA 或双重 gRNA (dgRNA - 即 - 定制的 crRNA 和 tracrRNA) 招募至特定基因组位点的能力。

[0045] 如本文中所述,除了引导 Cas9 介导的核酸酶活性外,还可能使用 CRISPR 衍生的 RNA 来将与 Cas9 融合的异源功能性结构域 (Cas9-HFD) 靶向基因组中的特定位点 (图 1C)。例如,如本文中所述,可能使用单一引导 RNA (sgRNA) 来将 Cas9-HFD 例如 Cas9 转录激活物 (在下文中被称为 Cas9 激活物) 靶向特定基因的启动子,从而增加靶基因的表达。因此,可将 Cas9-HFD 定位至基因组中的位点,靶特异性由引导 RNA 的序列互补性确定。靶序列还包括 PAM 序列 (与由 RNA 指定的序列相邻的由 Cas9 蛋白指定的 2-5 个核苷酸序列)。

[0046] Cas9-HFD 通过将异源功能性结构域 (例如,例如来自 VP64 或 NF- κ B p65 的转录激活结构域) 与无催化活性的 Cas9 蛋白的 N 末端或 C 末端融合来生成。

[0047] Cas9

[0048] 许多细菌表达 Cas9 蛋白变体。来自化脓性链球菌 (*Streptococcus pyogenes*) 的 Cas9 是目前最常使用的;一些另外的 Cas9 蛋白与化脓性链球菌 Cas9 具有高水平的序列同一性,并且使用相同的引导 RNA。其它的更加多样,使用不同的 gRNA,并且同样地识别不同的 PAM 序列 (与由 RNA 指定的序列相邻的由蛋白质指定的 2-5 个核苷酸的序列)。Chylinski 等将来自一大组细菌的 Cas9 蛋白分类 (RNA Biology 10:5, 1-12; 2013), 并且许多 Cas9 蛋白列于补充图 1 和其补充表 1 中,所述图表通过引用并入本文。另外的 Cas9 蛋白描述于 Esvelt 等, Nat Methods. 2013 年 11 月; 10(11):1116-21 和 Fonfara 等, "Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems." Nucleic Acids Res. 2013 年 11 月 22 日中。[先于印刷的电子出版] doi:10.1093/nar/gkt1074。

[0049] 许多物种的 Cas9 分子可用于本文中描述的方法和组合物。虽然化脓性链球菌和嗜热链球菌 Cas9 分子是本文中的许多公开内容的主题,但同样地可使用本文中所列的其它物种的 Cas9 分子、来源于或基于所述物种的 Cas9 蛋白的 Cas9 分子。换句话说,虽然本文中的许多描述使用化脓性链球菌和嗜热链球菌 Cas9 分子,但来自其它物种的 Cas9 分子可替代它们。此类物种包括下表中所示的那些物种,所述表是基于 Chylinski 等, 2013 的补充图 1 产生的。

[0050]

替代性 Cas9 蛋白	
基因库登录号	细菌
303229466	非典型韦荣球菌(<i>Veillonella atypical</i>)ACS-134-V-Col7a
34762592	具核梭杆菌奋森亚种(<i>Fusobacterium nucleatum subsp. Vincentii</i>)
374307738	<i>Filifactor alocis</i> ATCC 35896
320528778	<i>Solobacterium moorei</i> F0204
291520705	灵巧粪球菌(<i>Coprococcus catus</i>)GD-7
42525843	齿垢密螺旋体(<i>Treponema denticola</i>)ATCC 35405
304438954	<i>Peptoniphilus duerdenii</i> ATCC BAA-1640
224543312	<i>Catenibacterium mitsuokai</i> DSM 15897
24379809	变异链球菌(<i>Streptococcus mutans</i>)UA159
15675041	酿脓链球菌(<i>Streptococcus pyogenes</i>)SF370
16801805	无害李斯特菌(<i>Listeria innocua</i> Clip)11262
116628213	嗜热链球菌(<i>Streptococcus thermophilus</i>)LMD-9
323463801	伪中间型葡萄球菌(<i>Staphylococcus pseudintermedius</i>)ED99
352684361	<i>Acidaminococcus intestini</i> RyC-MR95
302336020	<i>Olsenella uli</i> DSM 7084
366983953	<i>Oenococcus kitaharae</i> DSM 17330
310286728	两歧双歧杆菌(<i>Bifidobacterium bifidum</i>)S17
258509199	鼠李糖乳杆菌(<i>Lactobacillus rhamnosus</i>)GG
300361537	加氏乳杆菌(<i>Lactobacillus gasseri</i>)JV-V03
169823755	大芬戈尔德菌(<i>Fingoldia magna</i>)ATCC 29328
47458868	运动支原体(<i>Mycoplasma mobile</i>)163K
284931710	鸡败血性霉浆菌(<i>Mycoplasma gallisepticum</i>)菌株 F
363542550	绵羊肺炎霉浆菌(<i>Mycoplasma ovipneumoniae</i>)SC01
384393286	犬支原体(<i>Mycoplasma canis</i>)PG 14
71894592	关节液支原体(<i>Mycoplasma synoviae</i>)53
238924075	直肠真杆菌(<i>Eubacterium rectale</i>)ATCC 33656
116627542	嗜热链球菌(<i>Streptococcus thermophilus</i>)LMD-9
315149830	粪肠球菌(<i>Enterococcus faecalis</i>)TX0012
315659848	路邓葡萄球菌(<i>Staphylococcus lugdunensis</i>)M23590
160915782	细长真杆菌(<i>Eubacterium dolichum</i>)DSM 3991
336393381	棒状乳杆菌扭曲亚种(<i>Lactobacillus coryniformis subsp. Torquens</i>)
310780384	营养泥杆菌(<i>Ilyobacter polytropus</i>)DSM 2926

[0051]

替代性 Cas9 蛋白	
基因库登录号	细菌
325677756	白色瘤胃球菌(<i>Ruminococcus albus</i>)8
187736489	<i>Akkermansia muciniphila</i> ATCC BAA-835
117929158	嗜酸纤维素分解菌(<i>Acidothermus cellulolyticus</i>)11B
189440764	长双歧杆菌(<i>Bifidobacterium longum</i>)DJO10A
283456135	齿双歧杆菌(<i>Bifidobacterium dentium</i>)Bd1
38232678	白喉棒状杆菌(<i>Corynebacterium diphtheriae</i>)NCTC13129
187250660	<i>Elusimicrobium minutum</i> Pei191
319957206	<i>Nitratifractor salsuginis</i> DSM 16511
325972003	<i>Sphaerochaeta globus</i> str. Buddy
261414553	产琥珀酸丝状杆菌亚种(<i>Fibrobacter succinogenes</i> subsp. <i>Succinogenes</i>)
60683389	脆弱拟杆菌(<i>Bacteroides fragilis</i>)NCTC 9343
256819408	黄褐二氧化碳嗜纤维菌(<i>Capnocytophaga ochracea</i>)DSM 7271
90425961	沼泽红假单胞菌(<i>Rhodopseudomonas palustris</i>)BisB18
373501184	普氏菌藻(<i>Prevotella micans</i>)F0438
294674019	反刍瘤胃亚菌(<i>Prevotella ruminicola</i>)23
365959402	柱状黄杆菌(<i>Flavobacterium columnare</i>)ATCC 49512
312879015	<i>Aminomonas paucivorans</i> DSM 12260
83591793	深红螺菌(<i>Rhodospirillum rubrum</i>)ATCC 11170
294086111	<i>Candidatus Puniceispirillum marinum</i> IMCC1322
121608211	<i>Verminephrobacter eiseniae</i> EF01-2
344171927	蒲桃雷尔氏菌(<i>Ralstonia syzygii</i>)R24
159042956	<i>Dinoroseobacter shibae</i> DFL 12
288957741	固氮螺菌属(<i>Azospirillum</i>)的一个种- B510
92109262	汉氏硝化细菌(<i>Nitrobacter hamburgensis</i>)X14
148255343	慢生根瘤菌属(<i>Bradyrhizobium</i>)的一个种- BTai1
34557790	琥珀酸弧菌(<i>Wolinella succinogenes</i>)DSM 1740
218563121	<i>Campylobacter jejuni</i> subsp. <i>jejuni</i>
291276265	鼯鼠螺杆菌(<i>Helicobacter mustelae</i>)12198
229113166	蜡状芽孢杆菌(<i>Bacillus cereus</i>)Rock1-15
222109285	<i>Acidovorax ebreus</i> TPSY
189485225	未培养白蚁组 1
182624245	产气荚膜梭菌 D 菌株(<i>Clostridium perfringens</i> D str.)

[0052]

替代性 Cas9 蛋白	
基因库登录号	细菌
220930482	解纤维素梭菌(<i>Clostridium cellulolyticum</i>)H10
154250555	食清洁剂细小棒菌(<i>Parvibaculum lavamentivorans</i>)DS-1
257413184	<i>Roseburia intestinalis</i> L1-82
218767588	奈瑟氏脑膜炎球菌(<i>Neisseria meningitidis</i>)Z2491
15602992	多杀性巴氏杆菌亚种(<i>Pasteurella multocida</i> subsp. <i>Multocida</i>)
319941583	华德萨特菌(<i>Sutterella wadsworthensis</i>)3 I
254447899	γ -变形菌纲(<i>gamma proteobacterium</i>)HTCC5015
54296138	嗜肺军团菌巴黎菌株(<i>Legionella pneumophila</i> str. <i>Paris</i>)
331001027	<i>Parasutterella excrementihominis</i> YIT 11859
34557932	产琥珀酸沃廉菌(<i>Wolinella succinogenes</i>)DSM 1740
118497352	新生儿弗朗西斯菌(<i>Francisella novicida</i>)U112

[0053] 本文中描述的构建体和方法包括任何那些 Cas9 蛋白以及它们对应的引导 RNA 或相容的其它引导 RNA 的使用。已显示来自嗜热链球菌 LMD-9CRISPR1 系统的 Cas9 在 Cong 等 (Science 339, 819 (2013)) 中的人细胞中起作用。另外, Jinek 等人在体外显示来自嗜热链球菌和英诺克李斯特菌 (*L. innocua*) (但非来自奈瑟氏脑膜炎球菌或空肠弯曲菌 (*C. jejuni*) 的 Cas9 直系同源物 (其可使用不同的引导 RNA) 可被双重化脓性链球菌 gRNA 导向裂解靶质粒 DNA, 虽然效率略有降低。

[0054] 在一些实施方案中, 本系统利用来自化脓性链球菌的 Cas9 蛋白 (如在细菌中编码的或针对在哺乳动物细胞中的表达进行密码子最优化的), 其在 D10、E762、H983 或 D986 和 H840 或 N863 处含有突变, 例如 D10A/D10N 和 H840A/H840N/H840Y, 以使得蛋白的核酸酶部分催化失活; 这些位置上的取代可以是丙氨酸 (如它们在 Nishimasu 等, Cell 156, 935 - 949 (2014) 中一样) 或它们可以是其它残基, 例如谷氨酰胺、天冬酰胺、酪氨酸、丝氨酸或天冬氨酸, 例如, E762Q、H983N、H983Y、D986N、N863D、N863S 或 N863H (图 1C)。可用于本文中描述的方法和组合物的自由化来活的化脓性链球菌 Cas9 的序列如下; D10A 和 H840A 的示例性突变以粗体表示并加以下划线。

[0055]

10	20	30	40	50	60
MDKKYSIGLA	IGTNSVGVAV	ITDEYEVPSK	EPKVLGNTDR	HSIKKNLIGA	LLFDSGETAE
70	80	90	100	110	120
ATRLKPTARR	EYTRRKNRIC	YLQEIFSNEM	AKVDDSPFHR	LEESFLVEED	KKHERHPFPG
130	140	150	160	170	180
NIVDEVAYHE	KYPTIYHLAK	KLVDSTDKAD	LRLIYLALAH	MIKFRGHFLI	EGDLNFDNSD
190	200	210	220	230	240
VDKLFIQLVQ	TYNQLFEENP	INASGVDAKA	ILSARLSKSR	PLENLIAQLP	GEKKNGLFGN
250	260	270	280	290	300
LIALSLGLTP	NFKSNFDLAE	DAKLQLSKDT	YDDDLNLLA	QIGDQYADLF	LAAKNLSDAI
310	320	330	340	350	360
LLSDILKVVNT	EITKAPLSAS	MIKRYDEHMQ	DLTLKALVR	QQLPEKYKEI	FFDQSKNGYA
370	380	390	400	410	420
GYIDGGASQE	EFYKFIKPIL	EKMDGTEELL	VKLNREDLLR	KQRTFDNGSI	PHQIHLGELH
430	440	450	460	470	480
AILRBQEDFY	PFLKDNREKI	EKILTFRIFY	YVGFLARGNS	RFAMNTRKSE	ETITPWNFEE
490	500	510	520	530	540
VVDKGASAQS	FIERMTNFDK	NLPNEKVLPK	HSLLYEYFTV	YNELTKVKYV	TEGMRKPAFL
550	560	570	580	590	600
SGEQKKAIVD	LLFKTNRKVT	VKQLNEDYFK	KIECFDSVEI	SGVEDRFNAS	LCTYHDLLEI
610	620	630	640	650	660
TKDKDFLDNE	ENEDILEDIV	LTLTLFEDRE	MIEERLKYA	HLFDDKVMKQ	LKRARRYTCWG
670	680	690	700	710	720
RLSRKLINGI	RDKQEGKTI	DFLKSDFGAN	RNFMQLIHDD	SLTFKEDIQK	AQVSGGGDSL
730	740	750	760	770	780
HEHIANLAGE	PAIKKGILOT	VKVVDLVKV	MGRHKPENIV	IEMARENQTT	QKGQKNSREK
790	800	810	820	830	840
MKRIEEGIN	LGSQILKERF	VENTQLQNEK	LYLYLQNGP	EMYVDQELDI	NRLSDYDVA
850	860	870	880	890	900
IVPQSFLDD	SIDNKVLTRS	DENRKESDQV	PSEEVKRMK	NYWRQLLNAK	LITQKQFDNL
910	920	930	940	950	960
TKAERGGLE	IDKAGFIRQ	LVETROITKH	VAQILDSRMN	TKYDENDKLI	REVKVITLKS
970	980	990	1000	1010	1020
KLVSDFKDF	QFYKVRINN	YHRAHDAYLN	AVVGTALIKK	YPKLESEFVY	GDYKVYDVRK
1030	1040	1050	1060	1070	1080
MIKSEQEIG	KATAKYFFYS	NIMNFFRTEI	TLANGEIRK	FLIETNGETG	EIVWDKGRDF
1090	1100	1110	1120	1130	1140
ATVREVLSP	QVNIIVKTEV	QTGGFSKESI	LPKRNSDKLI	AKKXDWDPKK	YGGFDSPTVA

[0056]

```

1150      1160      1170      1180      1190      1200
YSVLVVAEVE NGKSKKLSV KELLGITIME RSSFENPID FLEANGYKEV KKDLIINLPE

1210      1220      1230      1240      1250      1260
YSLFELENGR KMMLASAGEL QKGNELALPS KYVNFILYLAH HYEKIKGSPF INEQKQLFVE

1270      1280      1290      1300      1310      1320
QNKHYLDEII EQISEFSKEV ILADANLDKV LSAYNRHRDK FIREQAENII NLPTLTNLGA

1330      1340      1350      1360
PRAFNYFDYI IDRKRYSSTK EVLDATLIHQ SITGLYETRI DLSQLGGD (SEQ ID NO:13)

```

[0057] 在一些实施方案中,本文中使用的 Cas9 核酸酶与化脓性链球菌 Cas9 的序列具有至少约 50% 的同一性,即与 SEQ ID NO:13 具有至少 50% 的同一性。在一些实施方案中,所述核苷酸序列与 SEQ ID NO:13 具有约 50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、99% 或 100% 的同一性。

[0058] 在一些实施方案中,本文中使用的无催化活性的 Cas9 与无催化活性的化脓性链球菌 Cas9 的序列具有至少约 50% 的同一性,即,与 SEQ ID NO:13 具有至少 50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、99% 或 100% 的同一性,其中 D10 和 H840 处的突变,例如 D10A/D10N 和 H840A/H840N/H840Y 得到保持。

[0059] 在一些实施方案中,与 SEQ ID NO:13 的任何差异在非保守区中,如通过 Chylinski 等, RNA Biology 10:5, 1-12;2013(例如,在补充图 1 及其补充表 1 中); Esvelt 等, Nat Methods. 2013 年 11 月;10(11):1116-21 和 Fonfara 等, Nucl. Acids Res. (2014) 42(4):2577-2590 中所示的序列的序列比对鉴定的。[2013 年 11 月 22 日先于印刷的电子版]doi:10.1093/nar/gkt1074,并且其中 D10 和 H840 处的突变,例如, D10A/D10N 和 H840A/H840N/H840Y 得到保持。

[0060] 为了测定两个序列的百分比同一性,为了最佳比较目的,将所述序列比对(可根据需要在第一和第二氨基酸或核酸序列的一个或两个序列中引入缺口以进行最佳比对,并且为了比较目的可忽略非同源序列)。为了比较目的而比对的参照序列的长度为至少 50% (在一些实施方案中,比对约 50%、55%、60%、65%、70%、75%、85%、90%、95% 或 100% 的参照序列的长度)。随后比较对应位置上的核苷酸或残基。当第一序列中的位置被与第二序列中的对应位置相同的核苷酸或残基占据时,则所述分子在该位置上是相同的。两个序列之间的百分比同一性是由序列共享的相同位置的数目的函数,该函数考虑了为两个序列的最佳比对而需要引入的缺口数目和每一个缺口的长度。

[0061] 序列的比较和两个序列之间的百分比同一性的测定可使用数学算法来实现。为了本申请的目的,两个氨基酸序列之间的百分比同一性使用 Needleman 和 Wunsch((1970) J. Mol. Biol. 48:444-453) 算法,使用 Blossum 62 评分矩阵,利用为 12 的缺口罚分、为 4 的缺口延伸罚分和为 5 的移码缺口罚分来测定,该算法已被整合进 GCG 软件包中的 GAP 程序。

[0062] 异源功能性结构域

[0063] 可在 Cas9 的 N 或 C 末端上融合转录激活结构域。另外,尽管本说明书举例说明转录激活结构域,但还可使用如在本领域中是已知的其它异源功能性结构域(例如,转录阻遏物(例如, KRAB、ERD、SID 和其它,例如 ets2 阻遏因子(ERF)阻遏结构域(ERD)的氨基酸 473-530、KOX1 的 KRAB 结构域的氨基酸 1-97 或 Mad mSIN3 相互作用结构域(SID)的氨基酸 1-36;参见 Beerli 等, PNAS USA 95:14628-14633(1998))或沉默子诸如异染色质蛋

白 1 (HP1, 也称为 swi6), 例如 HP1 α 或 HP1 β ; 可招募与固定的 RNA 结合序列诸如被 MS2 衣壳蛋白、内切核糖核酸酶 Csy4 或 λ N 蛋白结合的那些序列融合的长的非编码 RNA (lncRNA) 的蛋白质或肽; 修饰 DNA 的甲基化状态的酶 (例如, DNA 甲基转移酶 (DNMT) 或 TET 蛋白); 或修饰组蛋白亚单位的酶 (例如, 组蛋白乙酰转移 (HAT)、组蛋白脱乙酰酶 (HDAC)、组蛋白甲基转移酶 (例如, 用于赖氨酸或精氨酸残基的甲基化) 或组蛋白脱甲基酶 (例如, 用于赖氨酸或精氨酸残基的脱甲基化))。此类结构域的许多序列在本领域中是已知的, 例如, 催化 DNA 中的甲基化半胱氨酸的羟化的结构域。示例性蛋白包括 10-11- 易位 (TET) 1-3 家族, 将 DNA 中的 5- 甲基胞嘧啶 (5-mC) 转化成 5- 羟甲基胞嘧啶 (5-hmC) 的酶。

[0064] 人 TET1-3 的序列在本领域中是已知的并且示于下表中:

[0065]

基因库登录号		
基因	氨基酸	核酸
TET1	NP_085128.2	NM_030625.2
TET2*	NP_001120680.1(var 1) NP_060098.3(var 2)	NM_001127208.2 NM_017628.4
TET3	NP_659430.1	NM_144993.1

[0066] * 变体 (1) 代表较长的转录物并且编码较长的同种型 (a)。变体 (2) 与变体 1 在 5' UTR 和 3' UTR 中以及在编码序列中不同。所得的同种型 (b) 是较短的并且相较于同种型 a 具有不同的 C 末端。

[0067] 在一些实施方案中, 可包括催化结构域的全长序列的全部或部分, 例如包含富含半胱氨酸的延伸和由 7 个高度保守的外显子编码的 20GFeD0 结构域, 例如包含氨基酸 1580-2052 的 Tet1 催化结构域、包含氨基酸 1290-1905 的 Tet2 和包含氨基酸 966-1678 的 Tet3 的催化模块。关于举例说明所有 3 个 Tet 蛋白中的至关重要的催化残基的比对, 参见, 例如, Iyer 等, Cell Cycle. 2009 年 6 月 1 日; 8(11):1698-710. Epub 2009 年 6 月 27 日的图 1, 并且关于全长序列 (参见, 例如, seq 2c), 参见其补充材料 (可在 ftp 站点 ftp.ncbi.nih.gov/pub/aravind/DONS/supplementary_material_DONS.html 获得); 在一些实施方案中, 所述序列包括 Tet1 的氨基酸 1418-2136 或 Tet2/3 中的对应区域。

[0068] 其它催化分子可来自 Iyer 等, 2009 鉴定的蛋白质。

[0069] 在一些实施方案中, 异源性功能结构域是生物系链, 并且包含 MS2 衣壳蛋白、内切核糖核酸酶 Csy4 或 λ N 蛋白的全部或部分 (例如来自其的 DNA 结合结构域)。这些蛋白可用于将含有特定茎环结构的 RNA 分子招募至由 dCas9gRNA 靶向序列指定的场所。例如, 与 MS2 衣壳蛋白、内切核糖核酸酶 Csy4 或 λ N 融合的 dCas9 可用于招募长的非编码 RNA (lncRNA) 诸如 XIST 或 HOTAIR; 参见, 例如, Keryer-Bibens 等, Biol. Cell 100:125 - 138 (2008), 其连接于 Csy4、MS2 或 λ N 结合序列。或者, 可将 Csy4、MS2 或 λ N 蛋白结合序列连接于另一种蛋白质, 如 Keryer-Bibens 等 (同上) 中描述的, 并且可使用本文中描述的方法和组合物将所述蛋白质靶向 dCas9 结合位点。在一些实施方案中, Csy4 是无催化活性的。

[0070] 在一些实施方案中, 融合蛋白包括 dCas9 与异源功能性结构域之间的接头。可用于这些融合蛋白中 (或串联结构中的融合蛋白之间) 的接头可包括不干扰融合蛋白的功能的任何序列。在优选实施方案中, 所述接头是短的, 例如 2-20 个氨基酸, 并且通常是柔性的 (即, 包含具有高度自由的氨基酸诸如甘氨酸、丙氨酸和丝氨酸)。在一些实施方案中, 所述

接头包含一个或多个由 GGGG(S EQ ID NO:14) 或 GGGGS(S EQ ID NO:15) 组成的单位,例如, 2、3、4 或更多个 GGGG(S EQ ID NO:14) 或 GGGGS(S EQ ID NO:15) 单位的重复。还可使用其它接头序列。

[0071] 使用方法

[0072] 所述 Cas9-HFD 系统是有用的并且是修饰内源基因的表达的通用工具。用于获得该系统的当前方法需要产生针对每一个待靶向的位点的新型工程化 DNA 结合蛋白(诸如工程化锌指或转录激活因子样效应子 DNA 结合结构域)。因为这些方法需要表达专门工程化来结合每一个靶位点的大蛋白质,因此它们在它们用于多路技术的能力上受到限制。然而,Cas9-HFD 需要仅单一 Cas9-HFD 蛋白的表达,所述蛋白可通过表达多个短的 gRNA 来被靶向基因组中的多个位点。该系统因而可容易地用于同时诱导许多基因表达或将多个 Cas9-HFD 招募至单个基因、启动子或增强子。该能力将具有广泛的用途,例如用于基础生物学研究,其中其可用于研究基因功能和维持多个基因在单个途径中的表达,以及用于合成生物学,其中其将使得研究者能够在负责多个输出信号的细胞中产生回路。该技术可被执行和适用于多路复用的相对容易性使得其成为具有许多广泛应用的广泛有用的技术。

[0073] 本文所述方法包括将细胞与编码本文所述 Cas9-HFD 的核酸和编码一个或多个被导向选定的基因的引导 RNA 的核酸接触,从而调节该基因的表达。

[0074] 引导 RNA (gRNA)

[0075] 引导 RNA 一般而言出现在两个不同的系统中:系统 1,其使用一起指导 Cas9 进行裂解的单独的 crRNA 和 tracrRNA,和系统 2,其使用在单个系统中组合两个单独的引导 RNA 的嵌合 crRNA-tracrRNA 杂交体(被称为单一引导 RNA 或 sgRNA,也参见 Jinek 等,Science 2012 ;337:816 - 821)。tracrRNA 可被可变地截短,并且已显示许多长度在所述单独的系统(系统 1)和嵌合 gRNA 系统(系统 2)中都具有功能。例如,在一些实施方案中,可从其 3' 末端将 tracrRNA 截短至少 1、2、3、4、5、6、7、8、9、10、15、20、25、30、35 或 40nt。在一些实施方案中,可将 tracrRNA 分子从其 5' 末端截短至少 1、2、3、4、5、6、7、8、9、10、15、20、25、30、35 或 40nt。或者,可以从 5' 和 3' 末端截短 tracrRNA 分子,例如在 5' 末端截短至少 1、2、3、4、5、6、7、8、9、10、15 或 20nt 并且在 3' 末端截短至少 1、2、3、4、5、6、7、8、9、10、15、20、25、30、35 或 40nt。参见,例如,Jinek 等,Science 2012 ;337:816 - 821;Mali 等,Science. 2013 年 2 月 15 日 ;339(6121):823-6;Cong 等,Science. 2013 年 2 月 15 日 ;339(6121):819-23;以及 Hwang 和 Fu 等,Nat Biotechnol. 2013 年 3 月 ;31(3):227-9;Jinek 等,Elife 2, e00471(2013))。对于系统 2,一般地更长长度的嵌合 gRNA 已显示更大的中靶活性,但不同长度的 gRNA 的相对特异性目有仍未确定,从而在某些情况下可能期望使用较短的 gRNA。在一些实施方案中,gRNA 与在转录起始位点上游约 100-800bp 内,例如在转录起始位点的上游约 500bp 内,包括转录起始位点,或在转录起始位点下游约 100-800bp 内,例如约 500bp 内的区域互补。在一些实施方案中,使用编码不止一个 gRNA 的载体(例如,质粒),例如编码导向靶基因的相同区域内的不同位点的 2、3、4、5 或更多个 gRNA 的质粒。

[0076] 可使用在其 5' 末端上具有与基因组 DNA 靶位点的互补链互补的 17-20nt 的引导 RNA(例如单一 gRNA 或 tracrRNA/crRNA)将 Cas9 核酸酶导向具有例如序列 NGG 的另外的邻近的前间区序列邻近基序(PAM)的特定 17-20nt 的基因组靶。因此,本方法可包括单一引导 RNA 的使用,所述单一引导 RNA 包含与通常反式编码的 tracrRNA 融合的 crRNA,例如 Mali

等, Science 2013 年 2 月 15 日 ;339(6121):823-6 中描述的唯一 Cas9 引导 RNA, 其在 5' 末端上具有例如 25-17 个, 任选地 20 个或更少核苷酸 (nt) 的与靶序列互补的序列, 例如, 紧接前间区序列邻近基序 (PAM) 例如 NGG、NAG 或 NNGG 的 5' 的靶序列的互补链的 20、19、18 或 17nt, 优选地 17 或 18nt。在一些实施方案中, 所述单个 Cas9 引导 RNA 由如下序列组成:

[0077] (X₁₇₋₂₀)GUUUUAGAGCUAGAAAUAGCAAGUAAAAUAAGGCUAGUCCG (X_N) (SEQ ID NO :1) ;

[0078] (X₁₇₋₂₀)GUUUUAGAGCUAUGCUGAAAAGCAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUC (X_N) (SEQ ID NO :2) ;

[0079] (X₁₇₋₂₀)GUUUUAGAGCUAUGCUGUUUUGGAAACAAAACAGCAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUC (X_N) (SEQ ID NO :3) ;

[0080] (X₁₇₋₂₀)GUUUUAGAGCUAGAAAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC (X_N) (SEQ ID NO :4) ;

[0081] (X₁₇₋₂₀)GUUUUAGAGCUAGAAAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC (SEQ ID NO :5) ;

[0082] (X₁₇₋₂₀)GUUUUAGAGCUAUGCUGGAAACAGCAUAGCAAGUUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC (SEQ ID NO :6) ;或

[0083] (X₁₇₋₂₀)GUUUUAGAGCUAUGCUGGAAACAGCAUAGCAAGUUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC (SEQ ID NO :7) ;

[0084] 其中 X₁₇₋₂₀是与靶序列的 17-20 个连续核苷酸互补的核苷酸序列。先前已在文献 (Jinek 等, Science. 337(6096):816-21(2012) 和 Jinek 等, Elife. 2:e00471(2013)) 描述了编码单一引导 RNA 的 DNA。

[0085] 引导 RNA 可包含可以是不干扰核糖核酸对 Cas9 结合的任何序列的 X_N, 其中 N(在 RNA 中)可以是 0-200 例如 0-100、0-50 或 0-20。

[0086] 在一些实施方案中, 所述引导 RNA 在 3' 末端上包含一个或多个腺嘌呤 (A) 或尿嘧啶 (U) 核苷酸。在一些实施方案中, 作为用作终止 RNA PolIII 转录的终止信号的一个或多个 T 的任选的存在的结果, 所述 RNA 在分子的 3' 末端包括一个或多个 U, 例如, 1 至 8 个或更多个 U (例如, U、UU、UUU、UUUU、UUUUU、UUUUUU、UUUUUUU、UUUUUUUU)。

[0087] 虽然本文中描述的一些实例利用单一 gRNA, 但还将所述方法与双重 gRNA (例如在天然存在的系统中发现的 crRNA 和 tracrRNA) 一起使用。在该情况下, 可将单一 tracrRNA 与多个不同的使用本系统表达的 crRNA 结合使用, 例如下列序列:

[0088] (X₁₇₋₂₀)GUUUUAGAGCUA (SEQ ID NO:102) ;

[0089] (X₁₇₋₂₀)GUUUUAGAGCUAUGCUGUUUUG (SEQ ID NO:103) ;或

[0090] (X₁₇₋₂₀)GUUUUAGAGCUAUGC (SEQ ID NO:104) ;和 tracrRNA 序列。在该情况下, 将 crRNA 在本文中描述的方法和分子中用作引导 RNA, 并且从相同或不同的 DNA 分子表达 tracrRNA。在一些实施方案中, 所述方法包括将细胞与 tracrRNA 接触, 所述 tracrRNA 包含如下序列或由所述序列组成: GGAACCAUCAAACAGCAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC (SEQ ID NO:8) 或其活性部分 (活性部分是保持与 Cas9 或 dCas9 形成复合物的能力的部分)。在一些实施方案中, 可从其 3' 末端将 tracrRNA 分子截短至少 1、2、3、4、5、6、7、8、9、10、15、20、25、30、35 或 40nt。在另一个实施方案中, 可从其 5' 末端将 tracrRNA 截短至少 1、2、3、4、5、6、7、8、9、10、15、20、25、30、35 或 40nt。或者, 可从 5'

和 3' 末端将 tracrRNA 分子截短,例如,在 5' 末端截短至少 1、2、3、4、5、6、7、8、9、10、15 或 20nt 并且在 3' 末端截短至少 1、2、3、4、5、6、7、8、9、10、15、20、25、30、35 或 40nt。除了 SEQ ID NO:8 以外,示例性 tracrRNA 还包括下列序列:UAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC (SEQ ID NO:105) 或其活性部分;或 AGCAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC (SEQ ID NO:106) 或其活性部分。

[0091] 在一些实施方案中,当将 (X_{17-20}) GUUUUAGAGCUAUGCUGUUUUG (SEQ ID NO:102) 用作 crRNA 时,使用下列 tracrRNA:GGAACCAUUCAAAACAGCAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC (SEQ ID NO:8) 或其活性部分。

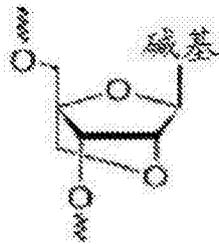
[0092] 在一些实施方案中,当将 (X_{17-20}) GUUUUAGAGCUA (SEQ ID NO:102) 用作 crRNA 时,使用下列 tracrRNA:UAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC (SEQ ID NO:105) 或其活性部分。

[0093] 在一些实施方案中,当将 (X_{17-20}) GUUUUAGAGCUAUGC (SEQ ID NO:104) 用作 crRNA 时,使用下列 tracrRNA:AGCAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC (SEQ ID NO:106) 或其活性部分。

[0094] 在一些实施方案中,所述 gRNA 被靶向与基因组的其余部分中的任何序列相异至少 3 或更多个错配的位点以使脱靶效应减小至最小。

[0095] 经修饰的 RNA 寡核苷酸诸如锁核酸 (LNA) 已被证明通过以更有利的 (稳定的) 构象锁定经修饰的寡核苷酸来提高 RNA-DNA 杂交的特异性。例如,2'-O-甲基 RNA 是其中在 2' 氧与 4' 碳之间存在另外的共价键联的经修饰的碱基,当被掺入寡核苷酸中时其可提高总体热稳定性和选择性 (式 I)。

[0096]



式 I - 锁核酸

[0097] 因此,在一些实施方案中,本文中公开的 tru-gRNA 可包含一个或多个经修饰的 RNA 寡核苷酸。例如,本文中描述的截短的引导 RNA 分子可具有与靶序列互补的引导 RNA 的一个或一些或全部区域被修饰,例如锁定的 (2'-O-4'-C 亚甲基桥)、5'-甲基胞苷、2'-O-甲基-假尿苷,或其中磷酸核糖主链已被聚酰胺链 (肽核酸) 例如合成核糖核酸替代。

[0098] 在其它实施方案中,tru-gRNA 序列的一个、一些或全部核苷酸可被修饰,例如锁定的 (2'-O-4'-C 亚甲基桥)、5'-甲基胞苷、2'-O-甲基-假尿苷,或其中磷酸核糖主链已被聚酰胺链 (肽核酸) 例如合成核糖核酸替代。

[0099] 在一些实施方案中,所述单一引导 RNA 和 / 或 crRNA 和 / 或 tracrRNA 可在 3' 末端上包含一个或多个腺嘌呤 (A) 或尿嘧啶 (U) 核苷酸。

[0100] 现有基于 Cas9 的 RGN 使用 gRNA-DNA 异源双链体形成来引导靶向目标基因组位点。然而,RNA-DNA 异源双链体可形成比它们的 DNA-DNA 对应物更加混杂的范围的结构。实

实际上,DNA-DNA 双链体对错配更加敏感,从而表明 DNA 引导的核酸酶可能不能容易地结合于脱靶序列,从而使得它们相较于 RNA 引导的核酸酶更具特异性。因此,可用于本文中描述的方法的引导 RNA 可以是杂交体,即,即其中一个或多个脱氧核糖核苷酸例如短的 DNA 寡核苷酸替代 gRNA 的全部或部分,例如 gRNA 的互补区的全部或部分。该基于 DNA 的分子可替代单一 gRNA 系统中的 gRNA 的全部或部分或可选地可替代双重 crRNA/tracrRNA 系统中的 crRNA 和 / 或 tracrRNA 的全部或部分。将 DNA 整合进互补区的此类系统应当因 DNA-DNA 双链体对错配的总体的不耐受性而相较于 RNA-DNA 双链体更容易靶向期望的基因组 DNA 序列。用于产生此类双链体的方法在本领域中是已知的,参见,例如,Barker 等,BMC Genomics. 2005 年 4 月 22 日 ;6:57 ;和 Sugimoto 等,Biochemistry. 2000 年 9 月 19 日 ;39(37):11270-81。

[0101] 另外,在使用单独的 crRNA 和 tracrRNA 的系统中,其一者或两者可以是合成的并且包含一个或多个经修饰的(例如,锁定的)核苷酸或脱氧核糖核苷酸。

[0102] 在细胞背景中,Cas9 与这些合成 gRNA 的复合物可用于提高 CRISPR/Cas9 核酸酶系统的全基因组特异性。

[0103] 所述方法可包括在细胞中表达本文中描述的 Cas9gRNA 加融合蛋白,或将所述细胞与其接触。

[0104] 表达系统

[0105] 为了使用描述的融合蛋白和引导 RNA,可能期望从编码它们的核酸表达它们。这可以以多种方式进行。例如,可将编码引导 RNA 或融合蛋白的核酸克隆入用于转化进用于复制和 / 或表达的原核或真核细胞的中间载体。中间载体通常是原核载体,例如,质粒或穿梭载体或昆虫载体,其用于贮存或操纵编码融合蛋白的核酸或用于产生融合蛋白。还可将编码引导 RNA 或融合蛋白的核酸克隆入表达载体,例如用于向植物细胞、动物细胞,优选地哺乳动物细胞或人细胞、真菌细胞、细菌细胞或原生动物细胞施用。

[0106] 为了获得表达,通常将编码引导 RNA 或融合蛋白的序列亚克隆入含有指导转录的启动子的表达载体。合适的细菌和真核启动子在本领域中是公知的,并且描述于例如 Sambrook 等,Molecular Cloning, A Laboratory Manual (2001 年第 3 版) ;Kriegler, Gene Transfer and Expression: A Laboratory Manual (1990) 和 Current Protocols in Molecular Biology (Ausubel 等,编辑,2010) 中。用于表达工程化蛋白质的细菌表达系统可在例如大肠杆菌 (*E. coli*)、芽孢杆菌属 (*Bacillus* sp.) 和沙门菌属 (*Salmonella*) (Palva 等,1983, Gene 22:229-235) 中获得。此类表达系统的试剂盒是商购可得的。用于哺乳动物细胞、酵母和昆虫细胞的真核表达系统在本领域中是公知的并且也是商购可得的。

[0107] 用于指导核酸表达的启动子取决于具体应用。例如,强组成型启动子通常用于融合蛋白的表达和纯化。相反地,当将体内施用融合蛋白以进行基因调控时,可使用组成型或诱导型启动子,这取决于融合蛋白的具体用途。另外,用于施用融合蛋白的优选启动子可以是弱启动子,诸如 HSV TK 或具有类似活性的启动子。启动子还可包括响应反式激活的元件,例如,缺氧应答元件、Gal4 应答元件、lac 阻遏应答元件和小分子控制系统诸如四环素调节的系统和 RU-486 系统(参见,例如, Gossen&Bujard, 1992, Proc. Natl. Acad. Sci. USA, 89:5547 ;Oligino 等,1998, Gene Ther. , 5:491-496 ;Wang 等,1997, Gene Ther. , 4:432-441 ;Neering 等,1996, Blood, 88:1147-55 和 Rendahl 等,1998, Nat. Biotechnol. , 16:757-761)。

[0108] 除了启动子以外,表达载体通常还含有包含核酸在宿主细胞(原核或真核的)中表达所需的所有另外的元件的转录单位或表达盒。常见表达盒从而含有可操作地连接于例如编码融合蛋白的核酸序列的启动子和例如进行转录物的高效多腺苷酸化、转录终止、核糖体结合位点或翻译终止的所需的任何信号。表达盒的另外元件可包括例如增强子和异源剪接内含子信号。

[0109] 根据融合蛋白的期望用途(例如,在植物、动物、细菌、真菌、原生动物等中表达)选择用于将遗传信息转运至细胞的特定表达载体。标准细菌表达载体包括质粒诸如基于 pBR322 的质粒、pSKF、pET23D 和商购可得的靶-融合表达系统诸如 GST 和 LacZ。优选的标签-融合蛋白是麦芽糖结合蛋白(MBP)。此类标签-融合蛋白可用于纯化工程化 TALE 重复蛋白。还可将表位标签例如 c-myc 或 FLAG 添加至重组蛋白以提供用于监测表达和用于监测细胞和亚细胞定位的方便的分离方法。

[0110] 含有来自真核病毒的调控元件的表达载体通常用于真核表达载体,例如,SV40 载体、乳头状瘤病毒载体和来源于爱泼斯坦-巴尔病毒的载体。其它示例性真核载体包括 pMSG、pAV009/A+、pMT010/A+、pMAMneo-5、杆状病毒 pDSVE 和允许在如下启动子指导下表达蛋白质的任何其它载体:SV40 早期启动子、SV40 晚期启动子、金属硫蛋白启动子、鼠乳腺肿瘤病毒启动子、劳斯肉瘤病毒启动子、多角体蛋白启动子或经显示对于在真核细胞中的表达是有效的其它启动子。

[0111] 用于表达引导 RNA 的载体可包括驱动引导 RNA 表达的 RNA Pol III 启动子,例如 H1、U6 或 7SK 启动子。这些人启动子允许在质粒转染后在哺乳动物细胞中表达 gRNA。或者, T7 启动子可用于例如体外转录,并且所述 RNA 可被体外转录和纯化。可使用适合用于短的 RNA 例如 siRNA、shRNA 或其它小的 RNA 表达的载体。

[0112] 一些表达系统具有用于选择稳定地转染的细胞系的标志物诸如胸苷激酶、潮霉素 B 磷酸物转移酶和二氢叶酸还原酶。高产表达系统也是合适的,诸如在昆虫细胞中使用杆状病毒载体,利用在多角体蛋白启动子或其它强杆状病毒启动子的指导下的融合蛋白编码序列。

[0113] 通常被包括在表达载体中的元件还包括在大肠杆菌中起作用的复制子、编码抗生素抗性以允许选择具有重组质粒的细菌的基因和允许重组序列插入的质粒的非必需区中的独特限制性位点。

[0114] 标准转染法可用于产生表达大量蛋白质的细菌、哺乳动物、酵母或昆虫细胞系,随后使用标准技术(参见,例如, Colley 等,1989, J. Biol. Chem., 264:17619-22; Guide to Protein Purification, 于 Methods in Enzymology, 第 182 卷(Deutscher, 编辑,1990)中)纯化所述蛋白质。真核和原核细胞的转化按照标准技术(参见,例如, Morrison, 1977, J. Bacteriol. 132:349-351; Clark-Curtiss&Curtiss, Methods in Enzymology 101:347-362(Wu 等,编辑,1983)来进行。

[0115] 可使用用于将外来核苷酸序列引入宿主细胞的任何已知方法。这些方法包括使用磷酸钙转染、聚凝胺、原生质体融合、电穿孔、核转染、脂质体、显微注射、裸 DNA、质粒载体、病毒载体(游离型和融合型)和用于将克隆的基因组 DNA、cDNA、合成 DNA 或其它外来遗传物质引入宿主细胞的任何其它公知的方法(参见,例如, Sambrook 等,同上)。唯一必需的是,使用的特定遗传工程方法能够成功地将至少一个基因引入能够表达选择的蛋白质的宿

主细胞。

[0116] 在一些实施方案中,所述融合蛋白包括提供给待被转运到细胞核的蛋白质的核定位结构域。几个核定位序列(NLS)是已知的,并且可使用任何适合的NLS。例如,许多NLS具有多个碱性氨基酸,称为二分基本重复(综述于 Garcia-Bustos 等,1991, *Biochim. Biophys. Acta*, 1071:83-101)中。含NLS的二分基本重复可被置于嵌合蛋白的任何部分,并且导致嵌合蛋白被定位在细胞核内。在优选实施方案中,将核定位结构域整合进最终的融合蛋白内,因为本文中描述的融合蛋白的最终功能将通常需要被定位在细胞核中的蛋白质。然而,在其中DBD结构域本身或终嵌合蛋白内的另一个功能性结构域具有固有的核转运功能的情况下,可以不必添加单独的核定位结构域。

[0117] 本发明包括所述载体和包含所述载体的细胞。

实施例

[0118] 在下列实施例中进一步描述本发明,所述实施例不限制权利要求中描述的本发明的范围。

[0119] 实施例 1. 工程化 CRISPR/Cas 激活系统:

[0120] 假设 RNA 引导的转录激活物可通过将强合成 VP64 激活结构域(Beerli 等, *Proc Natl Acad Sci USA* 95, 14628 - 14633(1998))与无催化活性的 dCas9 蛋白的羧基末端融合来产生(图 1D)。

[0121] 为了在人细胞中表达引导 RNA(gRNA),工程化载体,所述载体可表达由 U6 启动子驱动的全长嵌合 gRNA(最初由 Jinek 等(*Science*2012)描述的 crRNA 与 tracrRNA 的融合物)。如下进行 gRNA 表达质粒的构建。将成对的编码可变 20nt gRNA 靶向序列的 DNA 寡核苷酸退火在一起以产生具有 4bp 的悬突的短的双链 DNA 片段(表 1)。

[0122]

表 1. VEGFA 和 NTF3 基因靶位点和用于构建 gRNA 表达质粒的相关寡核苷酸		
gRNA	靶位点(包括 PAM)	SEQ ID NO:
V1	GTGTGCAGACGGCAGTCACTAGG	16.
V2	GAGCAGCGTCTTCGAGAGTGAGG	17.
V3	GGTGAGTGAGTGTGTGCGTGTGG	18.
V4	GTTGGAGCGGGGAGAAGGCCAGG	19.
V5	GGGTGGGGGAGTTTGCTCCTGG	20.
V6	GGCTTTGGAAAGGGGGTGGGGG	21.
V7	GGGGCGGGGTCCCGGCGGGGCGG	22.
V8	GCTCGGAGGTCGTGGCGCTGGGG	23.
V9	GACTCACCGGCCAGGGCGCTCGG	24.
V10	GGCGCAGCGGTTAGGTGGACCGG	25.
V11	GGCGCATGGCTCCGCCCCGCCGG	26.
V12	GCCACGACCTCCGAGCTACCCGG	27.
V13	GCGGCGTGAGCCCTCCCCCTTGG	28.
V14	GGAGGCGGGGTGGAGGGGGTTCGG	29.
V15	GGGTCACGCCGCGCTCCGGCGG	30.
V16	GACCCCTCCACCCCGCCTCCGG	31.
N1	GAGCGCGGAGCCATCTGGCCGGG	32.
N2	GCGCGGCGCGGAAGGGGTTAAGG	33.
N3	GCGGCGCGGCGCGGGGCCGGCGGG	34.
N4	GCCGCGCCGCCCTCCCCCGCCGG	35.
N5	GCGGTTATAACCAGCCAACCCGG	36.
N6	GTGCGCGGAGCTGTTTCGGAAGGG	37.
gRNA	顶部寡聚物	SEQ ID NO:
V1	ACACCGTGTGCAGACGGCAGTCACTG	38.
V2	ACACCGAGCAGCGTCTTCGAGAGTGG	39.
V3	ACACCGGTGAGTGAGTGTGTGCGTGG	40.
V4	ACACCGTTGGAGCGGGGAGAAGGCCG	41.
V5	ACACCGGGTGGGGGAGTTTGCTCCG	42.
V6	ACACCGGCTTTGGAAAGGGGGTGGGG	43.
V7	ACACCGGGGCGGGGTCCCGGCGGGG	44.

[0123]

V8	ACACCGCTCGGAGGTCGTGGCGCTGG	45.
V9	ACACCGACTCACC GGCCAGGGCGCTG	46.
V10	ACACCGGCGCAGCGGTTAGGTGGACG	47.
V11	ACACCGGCGCATGGCTCCGCCCCGCG	48.
V12	ACACCGCCACGACCTCCGAGCTACCG	49.
V13	ACACCGCGGCGTGAGCCCTCCCCCTG	50.
V14	ACACCGGAGGCGGGGTGGAGGGGGTG	51.
V15	ACACCGGGCTCAGCCGCGCTCCGGG	52.
V16	ACACCGACCCCTCCACCCCGCCTCG	53.
N1	ACACCGAGCGCGGAGCCATCTGGCCG	54.
N2	ACACCGCGCGGCGCGGAAGGGGTTAG	55.
N3	ACACCGCGGCGCGGCGCGGGCCGCGC	56.
N4	ACACCGCCGCGCCGCCCTCCCCGCG	57.
N5	ACACCGCGGTTATAACCAGCCAACCG	58.
N6	ACACCGTGCGCGGAGCTGTTTCGGAAG	59.
gRNA	底部寡聚物	SEQ ID NO:
V1	AAAACAGTGACTGCCGTCTGCACACG	60.
V2	AAAACCACTCTCGAAGACGCTGCTCG	61.
V3	AAAACCACGCACACACTCACTCACCG	62.
V4	AAAACGGCCTTCTCCCCGCTCCAACG	63.
V5	AAAACGGAGCAAAC TCCCCCACCCG	64.
V6	AAAACCCACCCCTTTCCAAAGCCG	65.
V7	AAAACCCCGCCGGGACCCCGCCCG	66.
V8	AAAACCAGCGCCACGACCTCCGAGCG	67.
V9	AAAACAGCGCCCTGGCCGTTGAGTCG	68.
V10	AAAACGTCCACCTAACCGCTGCGCCG	69.
V11	AAAACGCGGGGCGGAGCCATGCGCCG	70.
V12	AAAACGGTAGCTCGGAGGTCGTGGCG	71.
V13	AAAACAGGGGGAGGGCTCACGCCGCG	72.
V14	AAAACACCCCTCCACCCCGCCTCCG	73.
V15	AAAACCGGAGCGCGGCGTGAGCCCG	74.
V16	AAAACGAGGCGGGGTGGAGGGGGTTCG	75.
N1	AAAACGGCCAGATGGCTCCGCGCTCG	76.
N2	AAAAC TAACCCCTCCGCGCCGCGCG	77.
N3	AAAACGCCGGCCGCGCCGCGCCGCG	78.
N4	AAAACGCGGGGAGGGCGGCGCGCGC	79.
N5	AAAACGGTTGGCTGGTTATAACCGCG	80.
N6	AAAAC TCCGAACAGCTCCGCGCACG	81.

[0124] 将这些片段接入 BsmBI 消化的质粒 pMLM3636 以产生通过人 U6 启动子表达的编码嵌合的 ~ 102nt 单链引导 RNA 的 DNA (Mali 等, Science. 2013 年 2 月 15 日; 339(6121):823-6; Hwang 等, Nat Biotechnol. 2013 年 3 月; 31(3):227-9)。pMLM3636 质粒及其完全 DNA 序列可获自 Addgene。参见图 4。

[0125] 为了工程化 Cas9 激活物, 将 D10A、H840A 催化突变 (先前在 Jinek 等, 2012; 和 Qi 等, 2013 中描述的) 引入野生型或密码子最优化的 Cas9 序列 (图 5)。这些突变使得 Cas9 无催化活性, 从而使得其不再诱导双链断裂。在一个构建体中, 将三重 flag 标签、核定位信号和 VP64 激活结构域与失活的 Cas9 的 C 末端融合 (图 6)。该融合蛋白的表达由 CMV 启动子驱动。

[0126] 如下进行 dCas-VP64 表达质粒的构建。使用引物 (该引物将 T7 启动子位点添加至起始密码子的 5' 并且将核定位信号添加在 Cas9 编码序列的羧基末端) 通过 PCR 从质粒 pMJ841 (Addgene 质粒 #39318) 扩增编码具有灭活 D10A/H840A 突变的 Cas9 核酸酶 (dCas9) 的 DNA, 并如先前所述 (Hwang 等, Nat Biotechnol 31, 227 - 229 (2013)) 将其克隆进含有

CMV 启动子质粒以产生质粒 pMLM3629。将编码三重 FLAG 表位的寡核苷酸退火,并克隆进质粒 pMLM3629 中的 XhoI 和 PstI 位点以产生表达具有 C 末端 flag FLAG 标签的 dCas9 的质粒 pMLM3647。将编码后接合成 VP64 激活结构域的 Gly₄Ser 接头的 DNA 序列引入质粒 pMLM3647 中的 FLAG- 标记的 dCas9 的下游以产生质粒 pSL690。通过 QuikChange 定点诱变 (Agilent) 将 D10A/H840A 突变引入质粒 pJDS247,所述质粒编码已针对在人细胞中的表达进行了密码子最优化的 FLAG 标记的 Cas9 序列,以产生质粒 pMLM3668。随后将编码 Gly₄Ser 接头和 VP64 激活结构域的 DNA 序列克隆进 pMLM3668 以产生称为 pMLM3705 的密码子最优化的 dCas9-VP64 表达载体。

[0127] 如下进行细胞培养、转染和 ELISA 测定。将 Flp-In T-Rex 293 细胞维持在补充有 10% FBS、1% penstrep 和 1% Glutamax (Invitrogen) 的高级 DMEM 中。按照制造商的说明书利用脂质体 LTX (Invitrogen) 转染细胞。简言之,将 160,000 个 293 细胞接种在 24 孔板中,第二天用 250ng gRNA 质粒、250ng Cas9-VP64 质粒、30ng pmaxGFP 质粒 (Lonza)、0.5ul Plus 试剂和 1.65ul 脂质体 LTX 转染所述细胞。在转染后 40 小时收获来自感染的 293 细胞的组织培养基,使用 R&D System's Human VEGF-A ELISA 试剂盒“人 VEGF 免疫测定”来测定分泌的 VEGF-A 蛋白。

[0128] 16 个 sgRNA 被构建来靶向 293 细胞中的位于人 VEGFA 基因的转录起始位点上游、下游或其上的 3 个 DNA 酶 I 高敏位点 (HSS) 内的序列 (图 1E)。

[0129] 在测试 16 个 VEGFA 靶向 gRNA 招募新型 dCas9-VP64 融合蛋白的能力之前,首先测定这些 gRNA 的每一个的将 Cas9 核酸酶导向其在人 293 细胞中的期望的靶位点的能力。为了该目的,以 1:3 的比率转染 gRNA 和 Cas9 表达载体,因为先前的最优化实验表明使用该质粒比率在 U2OS 细胞中诱导了高水平的 Cas9 诱导的 DNA 裂解。

[0130] 除用 125ng 编码 VEGFA 靶向 gRNA 的质粒和 375ng 编码活性 Cas9 核酸酶的质粒 (pMLM3639) 转染细胞外,如上文针对 dCas9-VP16VEGFA 实验所述进行 293 细胞的转染。转染后 40 小时,按照制造商的说明书,使用 QIAamp DNA Blood Mini 试剂盒 (Qiagen) 分离基因组 DNA。使用 Phusion Hot Start II 高保真 DNA 聚合酶 (NEB),利用 3% DMSO 和下列降落 PCR 循环:98°C 的 10 个循环,10 秒;72 - 62°C, -1°C / 循环,15 秒;72°C, 30 秒,随后 98°C 的 25 个循环,10 秒;62°C, 15 秒;72°C, 30 秒进行 VEGFA 启动子中的 3 个不同靶向区的 PCR 扩增。使用引物 oFYF434 (5' -TCCAGATGGCACATTGTCAG-3' (SEQ ID NO:82)) 和 oFYF435 (5' -AGGGAGCAGGAAAGTGAGGT-3' (SEQ ID NO:83)) 扩增 -500 区域。使用引物 oFYF438 (5' -GCA CGTAACCTCACTTTCCT-3' (SEQ ID NO:84)) 和 oFYF439 (5' -CTTGCTACCTCTTTCCTTTTCT-3' (SEQ ID NO:85)) 扩增转录起始位点周围的区域。使用引物 oFYF444 (5' -AGAGAAGTCGAGGAAGA GAGAG-3' (SEQ ID NO:86)) 和 oFYF445 (5' -CAGCAGAAAGTTCATGGTTTCG-3' (SEQ ID NO:87)) 扩增 +500 区域。使用 Ampure XP 珠 (Agencourt) 纯化 PCR 产物,随后进行 T7 内切核酸酶 I 测定,并如先前所述 (Reyon 等, Nat Biotech 30, 460-465 (2012)) 在 QIAXCEL 毛细管电泳系统上进行分析。

[0131] 所有 16 个 gRNA 都能介导 Cas9 核酸酶诱导的插入缺失突变在它们各自的靶位点上的高效引入,如使用先前描述的 T7E1 基因分型测定评估的 (表 2)。因此,所有 16 个 gRNA 可与 Cas9 核酸酶复合并将其活性导向人细胞中的特定靶基因组位点。

[0132] 表 2. 由 VEGFA 靶向的 gRNA 和 Cas9 核酸酶诱导的插入缺失突变的频率

[0133]

gRNA	平均插入缺失突变频率
	(%) ± SEM
V1	18.05 ± 0.47
V2	41.48 ± 0.62
V3	33.22 ± 1.05
V4	16.97 ± 0.06
V5	7.46 ± 0.50
V6	16.99 ± 0.51
V7	1.42 ± 0.11
V8	34.07 ± 0.90
V9	24.53 ± 1.40
V10	35.65 ± 1.35
V11	4.45 ± 0.22
V12	23.95 ± 0.41
V13	9.45 ± 0.74
V14	12.17 ± 0.36
V15	14.28 ± 0.54
V16	18.82 ± 1.48

[0134] 为了测试 dCas9-VP64 蛋白是否也可被相同的 gRNA 靶向人细胞中的特定基因组位点,如下进行 VEGFA 蛋白的酶联免疫印迹测定。转染后 40 小时收获用编码 VEGFA 靶向 sgRNA 和 dCas9-VP64 的质粒转染的 Flp-In T-Rex HEK293 细胞的培养基,如先前所述 (Maeder 等, Nat Methods 10, 243 - 245 (2013)) 测量 VEGFA 蛋白表达。通过将来自其中表达 sgRNA 和 dCas9-VP64 的细胞的培养基中的 VEGFA 蛋白的浓度除以来自其中表达脱靶 sgRNA (被靶向 EGFP 报告基因中的序列) 和 dCas9-VP64 的细胞的培养基中的 VEGFA 蛋白的浓度来计算 VEGFA 表达的激活倍数。

[0135] 当与 dCas9-VP64 在人 293 细胞中共表达时,测试的 16 个 gRNA 中有 15 个诱导显著的 VEGFA 蛋白表达的增加 (图 2A)。观察到的 VEGFA 诱导的量级在 2 至 18.7- 激活倍数的范围内,平均值为 5- 倍激活。对照实验显示单独的 16 个 gRNA、单独的 dCas9-VP64 以及与被设计来结合 EGFP 受体基因序列的“脱靶”gRNA 一起的 dCas9-VP64 的每一种的表达全都不能诱导升高的 VEGFA 表达 (图 2A),这表明特定 gRNA 和 dCas9-VP64 蛋白的共表达是启动子激活所需的。因此,dCas9-VP64 被稳定地表达,并且可被 gRNA 导向激活人细胞中的特定基因组基因座的转录。在用 gRNA3 转染的细胞中观察到 VEGFA 的最大增加,所述 gRNA3 诱导蛋白质表达 18.7 倍。有趣地,3 个最好的 gRNA 以及能够诱导 3 倍或更多倍的表达的 9 个 gRNA 中的 6 个靶向 -500 区域 (转录起始位点上游离 ~ 500bp)。

[0136] 因为在一个方面,本文中描述的系统使用可变的 gRNA 来招募共同的 dCas9-VP64 激活融合物,因此可设想多重引导 RNA 在单个细胞中的表达可能使得能够多重或组合激活内源基因靶。为了测试该可能性,用 dCas9-VP64 表达质粒与各自单独地诱导从 VEGFA 启动子的表达的 4 个 gRNA (V1、V2、V3 和 V4) 的表达质粒一起转染 293 细胞。所有 4 个 gRNA 与 dCas9-VP64 的共表达诱导 VEGFA 蛋白表达的协同激活 (即,大于每一个单独的激活物的预

期的累加效应的倍数-激活)(图 2B)。另外,这 4 个激活物中的 3 个的各种组合也协同地激活 VEGFA 启动子(图 2B)。因为转录的协同激活据信由多个激活结构域至单个启动子的招募引起,因此在这些实验中,多个 gRNA/dCas9-VP64 复合物可能同时结合 VEGFA 启动子。

[0137] 这些实验表明,Cas9-HFD 例如 Cas9- 激活蛋白(具有 VP64 转录激活结构域)和具有 20nt 的与人 VEGF-A 启动子中的位点互补的序列的 sgRNA 在人 HEK293 细胞中的共表达可导致 VEGF-A 表达的上调。通过 ELISA 测定测量 VEGF-A 蛋白的增加,并且发现单独 gRNA 可与 Cas9- 激活融合蛋白一起用于使 VEGF-A 蛋白水平升高达到~18 倍(图 2A)。另外,有可能通过转录协同(通过将多个靶向相同启动子中的不同位点的 gRNA 与 Cas9- 激活融合蛋白一起引入)来实现甚至更大的激活提高(图 2B)。

[0138] 实施例 2. 工程化靶向内源人 NTF3 基因的 CRISPR/Cas 激活系统

[0139] 为了扩展本发现的一般性,我们测试 RNA 引导的激活物平台是否呆用于诱导人 NTF3 基因的表达。为了实现该目的,将 6 个 sgRNA 设计成人 NTF3 启动子中的预测的 DNA 酶 I 高敏位点(HSS),并且将表达这些 gRNA 的每一个的质粒与已针对人细胞表达进行了密码子最优化的编码 dCas9-VP64 蛋白的质粒共转染(图 3A)。

[0140] 所有 6 个测试的 gRNA 诱导显著的 NTF3 转录物水平的升高,如通过定量 RT-PCR 检测的(图 3B)。虽然不能精确地计算这 6 个 RNA 引导的激活物的倍数-激活值(因为转录物的基线水平基本上是不可检测的),但激活的 NTF3mRNA 表达的平均值在 4 倍的范围内变化。减少转染的 gRNA 和 dCas9-VP64 表达质粒的量导致更少的 NTF3 基因的激活(图 3B),这表明了明确的剂量依赖性效应。

[0141] 另外,单独地以及以单一组合和双重组合用 dCas9-VP64 和 NTF3 靶向的 gRNA 表达质粒共转染 293 细胞。通过定量 RT-PCR 检测 NTF3mRNA 的相对表达,并将所述相对表达针对 GAPDH 对照($\Delta Ct \times 10^4$)进行标准化。在所有这些实验中,用于转染的每一个单独的 gRNA 表达质粒的量是相同的。图 3B 显示该多重 gRNA 表达通过 dCas9-VP64 蛋白诱导 NTF3mRNA 表达的协同激活。

[0142] 实施例 3. 工程化 CRISPR/Cas-MS2、-Csy4 和 - λ N 融合系统 - 产生生物系链

[0143] 产生融合蛋白,其中将 MS2 衣壳蛋白、Csy4 核酸酶(优选地无催化活性的 Csy4,例如 Haurwitz 等 329(5997):1355-8(2010)中描述的 H29A 突变体)或 λ N 与灭活的 dCas9 的 N 或 C 末端融合。MS2 和 λ N 是结合特定 RNA 序列的噬菌体蛋白,从而可用作将 dCas9 蛋白系连于利用特定 MS2 或 λ N RNA 结合序列标记的异源 RNA 序列的衔接子。将 dCas9-MS2 融合物或 dCas9- λ N 融合物与在它们的 5' 或 3' 末端与 MS2 或 λ N 茎环识别序列融合的嵌合长非编码 RNA(lncRNA)共表达。嵌合 Xist 或嵌合 RepA lncRNA 将被 dCas9 融合物专门招募,并且该策略诱导靶向沉默的能力将通过测量靶基因表达来测量。所述系统将通过测试对衣壳蛋白和嵌合 RNA 的各种改变来最优化。已前已显示对 MS2 衣壳蛋白的 N55K 和 Δ FG 突变阻止蛋白质聚集并增加对茎-环 RNA 的亲合力。另外,我们将测试据报导增加对于 MS2 衣壳蛋白的亲合力的高亲和力 C-环 RNA 突变体。下文给出 MS2 和 λ N 蛋白的示例性序列; MS2 以二聚体形式起作用,从而 MS2 蛋白可包括融合单链二聚体序列。

[0144] 1. 单个 MS2 衣壳蛋白(wt、N55K 或 Δ FG)与 dCas9 的 N 末端或 C 末端的融合物的示例性序列

[0145] MS2 衣壳蛋白氨基酸序列:

[0146] MASNFTQFVLVDNNGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCVSRQSSAQNRKYTIKVEVPKVA
TQTVGGVELPVAAWRSYLNMLTPIFATNSDCELVKAMQGLLKDGNPIPSAIAANSGIY (SEQ ID NO :88)

[0147] MS2 N55K :

[0148] MASNFTQFVLVDNNGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCVSRQSSAQNRKYTIKVEVPKVA
TQTVGGVELPVAAWRSYLNMLTPIFATNSDCELVKAMQGLLKDGNPIPSAIAANSGIY (SEQ ID NO :89)

[0149] MS2 Δ FG :

[0150] MASNFTQFVLVDNNGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCVSRQSSAQNRKYTIKVEVPKGA
WRSYLNMLTPIFATNSDCELVKAMQGLLKDGNPIPSAIAANSGIY (SEQ ID NO :90)

[0151] 2. 融合的二聚 MS2 衣壳蛋白 (wt、N55K 或 Δ FG) 与 dCas9 的 N 末端或 C 末端的融合物的示例性序列

[0152] 二聚 MS2 衣壳蛋白 :

[0153] MASNFTQFVLVDNNGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCVSRQSSAQNRKYTIKVEVPKVA
TQTVGGVELPVAAWRSYLNMLTPIFATNSDCELVKAMQGLLKDGNPIPSAIAANSGLYGAMASNFTQFVLVDNNG
TGDVTVAPSNFANGVAEWISSNSRSQAYKVTCVSRQSSAQNRKYTIKVEVPKVATQTVGGVELPVAAWRSYLNML
TPIFATNSDCELVKAMQGLLKDGNPIPSAIAANSLIN (SEQ ID NO91)

[0154] MASNFTQFVLVDNNGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCVSRQSSAQNRKYTIKVEVPKVA
TQTVGGVELPVAAWRSYLNMLTPIFATNSDCELVKAMQGLLKDGNPIPSAIAANSGLYGAMASNFTQFVLVDNNG
TGDVTVAPSNFANGVAEWISSNSRSQAYKVTCVSRQSSAQNRKYTIKVEVPKVATQTVGGVELPVAAWRSYLNML
TPIFATNSDCELVKAMQGLLKDGNPIPSAIAANSLIN (SEQ ID NO :92)

[0155] 二聚 MS2 Δ FG :

[0156] MASNFTQFVLVDNNGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCVSRQSSAQNRKYTIKVEVPKGA
WRSYLNMLTPIFATNSDCELVKAMQGLLKDGNPIPSAIAANSGLYGAMASNFTQFVLVDNNGTGDVTVAPSNFA
NGVAEWISSNSRSQAYKVTCVSRQSSAQNRKYTIKVEVPKGAWRSYLNMLTPIFATNSDCELVKAMQGLLKDGN
PIPSAIAANSLIN (SEQ ID NO :93)

[0157] 3. λ N 与 dCas9 的 N 末端或 C 末端的融合物的示例性序列

[0158] λ N 氨基酸序列 :

[0159] MDAQTRRRERRAEKQAQWKAAN (SEQ ID NO :94) 或

[0160] MDAQTRRRERRAEKQAQWKAANPLLVGSAKPVNRPILSLNRKPKSRVESALNPIDLTVLAHEYHKQIES
NLQRIERKNQRTWYSKPGERGITCSGRQKIKGKSIPLI (SEQ ID NO :95)

[0161] 4. Csy4 与 dCas9 的 N 末端或 C 末端的融合物的示例性序列

[0162] 在 Haurwitz 等 329 (5997) :1355-8 (2010) 中给出了 Cys4 (例如灭活的形式) 的示例性序列。

[0163] 将构建体在也表达已在其 5' 或 3' 末端与 λ N 或 MS2 的同源茎 - 环识别序列融合的调控 RNA 例如长非编码 RNA (lncRNA) 诸如 HOTAIR、HOTTIP、XIST 或 XIST RepA 的细胞中表达。MS2 的野生型和高亲和力序列分别为 AAACAUGAGGAUUACCCAUGUCG (SEQ ID NO:96) 和 AAACAUGAGGAUCACCCAUGUCG (SEQ ID NO:97) (参见 Keryer-Bibens 等, 同上, 图 2) ; λ N 所结合的 nutL 和 nutR BoxB 分别为 GCCCUGAAGAAGGGC (SEQ ID NO:98) 和 GCCCUGAAAAAGGGC (SEQ ID NO:99)。Csy4 所结合的序列为 GTTCACTGCCGTATAGGCAG (截短的 20nt) (SEQ ID NO:100) 或 GUUCACUGCCGUUAAGGCAGCUAAGAAA (SEQ ID NO:101)。

[0164] dCas9/MS2 对表达 MS2 结合序列标记的 lncRNA 的细胞中的靶位点的结合将该 lncRNA 招募至 dCas9 结合位点；其中 lncRNA 是阻遏物，例如，XIST (dCas9 结合位点附近的基因) 被抑制。类似地，dCas9/ λ N 对表达 λ N- 结合序列标记的 lncRNA 的细胞中的靶位点的结合将该 lncRNA 招募至 dCas9 结合位点。

[0165] 实施例 4. 工程化 CRISPR/Cas-HP1 融合系统 - 序列 - 特异性沉默

[0166] 本文所述 dCas9 融合蛋白还可用于靶向沉默结构域，例如异染色质蛋白 1 (HP1, 也称为 swi6), 例如 HP1 α 或 HP1 β 。可将其中异染色质已被去除的 HP1 α 或 HP1 β 的截短的形式靶向特定基因座来诱导异染色质形成和基因沉默。与 dCas9 融合的截短的 HP1 的示例性序列示于图 8A-8B 中。可将 HP1 序列与如上所述的灭活的 dCas9 的 N 或 C 末端融合。

[0167] 实施例 5. 工程化 CRISPR/Cas-TET 融合系统 - 序列 - 特异性脱甲基化

[0168] 本文中描述的 dCas9 融合蛋白还可用于靶向修饰 DNA 的甲基化状态的酶 (例如, DNA 甲基转移酶 (DNMT) 或 TET 蛋白)。可将 TET1 的截短形式靶向特定基因座以催化 DNA 脱甲基化。与 dCas9 融合的截短的 TET1 的示例性序列示于图 9 中。可将 TET1 序列融合于如上所述的灭活的 dCas9 的 N 或 C 末端。

[0169] 实施例 6. 工程化最优化的 CRISPR/Cas-VP64 融合物

[0170] 通过改变这些融合物内的核定位信号 (NLS) 和 3xFLAG- 标签的数目和位置最优化具有 VP64 激活结构域的基于 dCas9 的转录激活物的活性 (图 10)。含有 N 末端 NLS 和位于 dCas9 与 VP64 序列之间的 NLS 的 dCas9-VP64 融合物一致地诱导较高水平的靶基因激活, 可能因激活物的增加的核定位导致 (图 10)。此外, 当将 3xFLAG 标签置于 dCas9 的 C 末端与 VP64 的 N 末端之间时, 观察到甚至更高水平的激活。3xFLAG 标签可用作人工接头, 在 dCas9 与 VP64 之间提供必要的间隔, 并且可能允许更好地折叠 VP64 结构域 (当在 dCas9 附近受约束时这可能是无法进行的) 或由招募 RNA 聚合酶 II 的转录中介体复合物更好地识别 VP64。或者, 带负电荷的 3xFLAG 标签还可用作偶然的转录激活结构域, 从而增强 VP64 结构域的效应。

[0171] 实施例 7. 最优化的无催化活性的 Cas9 蛋白 (dCas9)

[0172] 通过改变消除 dCas9 结构域中的 Cas9 的核酸酶活性的失活突变的性质来进行 dCas9-VP64 激活物的另外的最优化 (图 11A-B)。在迄今公开的研究中, 将催化残基 D10 和 H840 突变成丙氨酸 (D10A 和 H840A) 来破坏介导 DNA 的水解的活性位点网络。有人假设, 这些位置上的丙氨酸取代可能导致 dCas9 的去稳定, 从而导致亚最佳活性。因此, 测试 D10 或 H840 上结构更加保守的取代 (例如, 至天冬酰胺或酪氨酸残基 :D10N、H840N 和 H840Y) 以看它们是否可导致通过具有这些不同突变的 dCas9-VP64 融合物的更大基因激活。当将这些变异取代的 dCas9-VP64 变体与靶向内源人 VEGFA 基因的上游区域的 3 个 gRNA 共转染进 HEK293 细胞时, 除一个变体外, 对于所有这些变体观察到更多的 VEGFA 蛋白质表达 (图 11A)。然而, 当将所述 dCas9-VP64 变体仅与这些 gRNA 中的一个共转染时 (图 11A) 或当转染进表达单一 VEGFA 靶向的 gRNA 的 HEK293 来源的细胞系时 (图 11B), 该效应没有这样显著。

[0173] 其它实施方案

[0174] 应理解, 虽然已结合其详细描述说明了本发明, 但前述说明指在举例说明而不是限制本发明的范围, 本发明的范围由所述权利要求的范围来界定。其它方面、有利方面和修

改在下列权利要求的范围内。

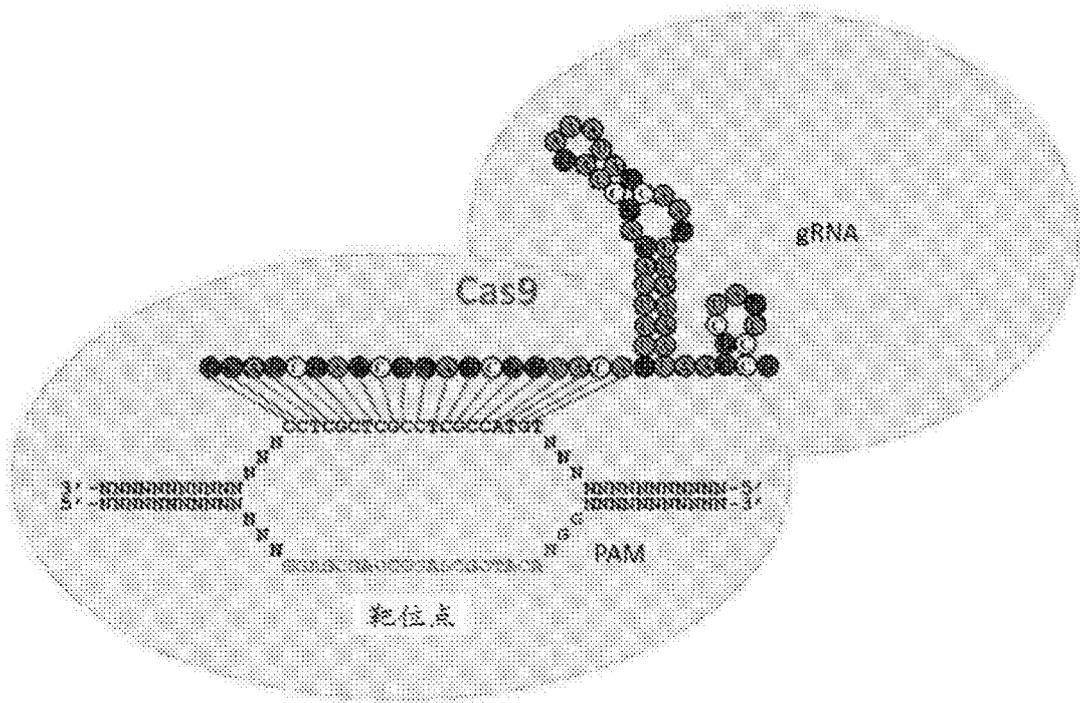


图 1A

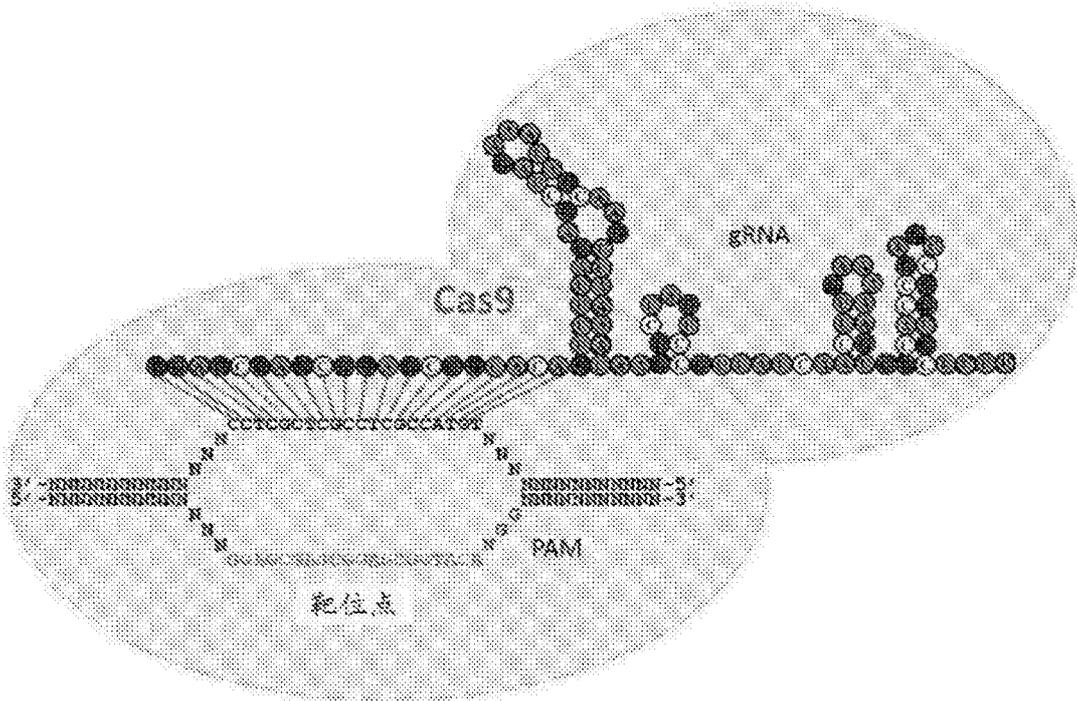


图 1B

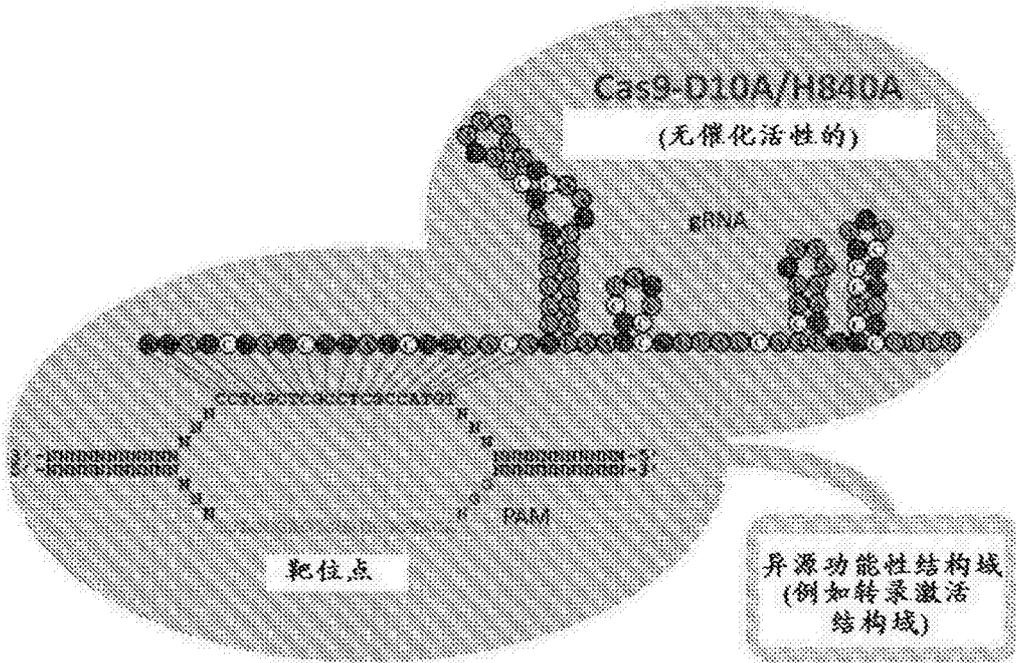


图 1C

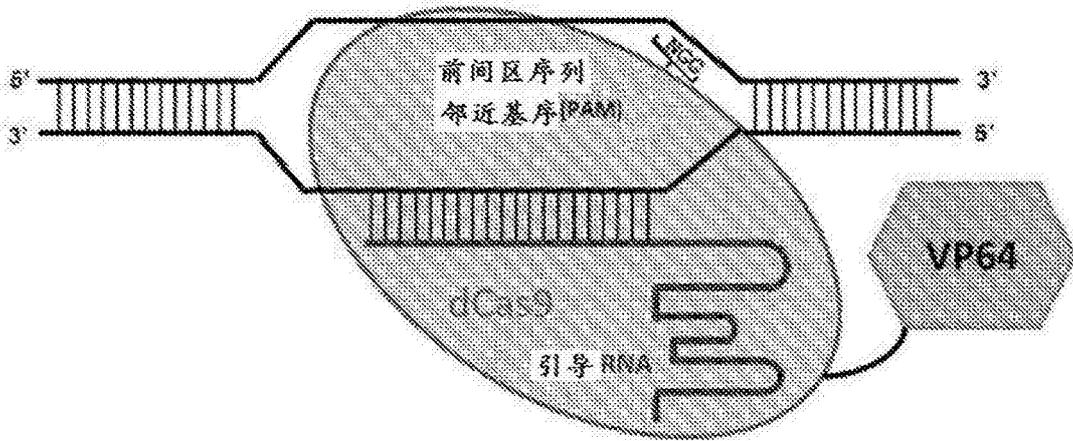


图 1D

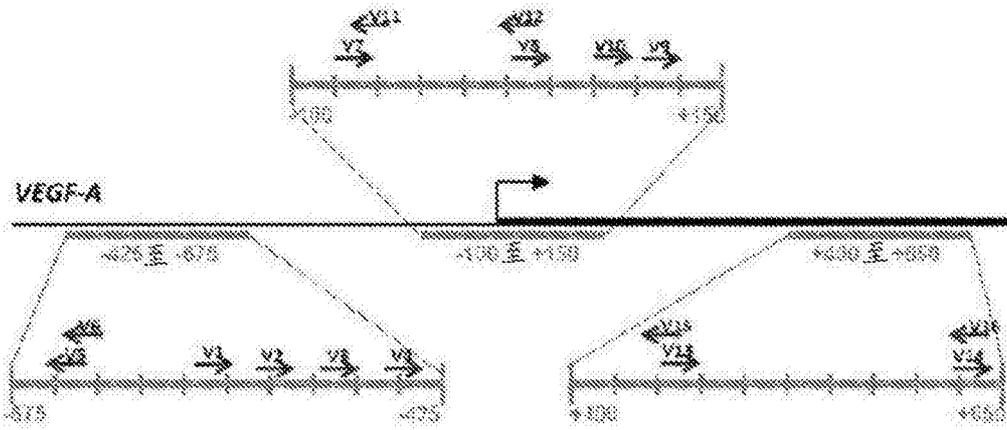


图 1E

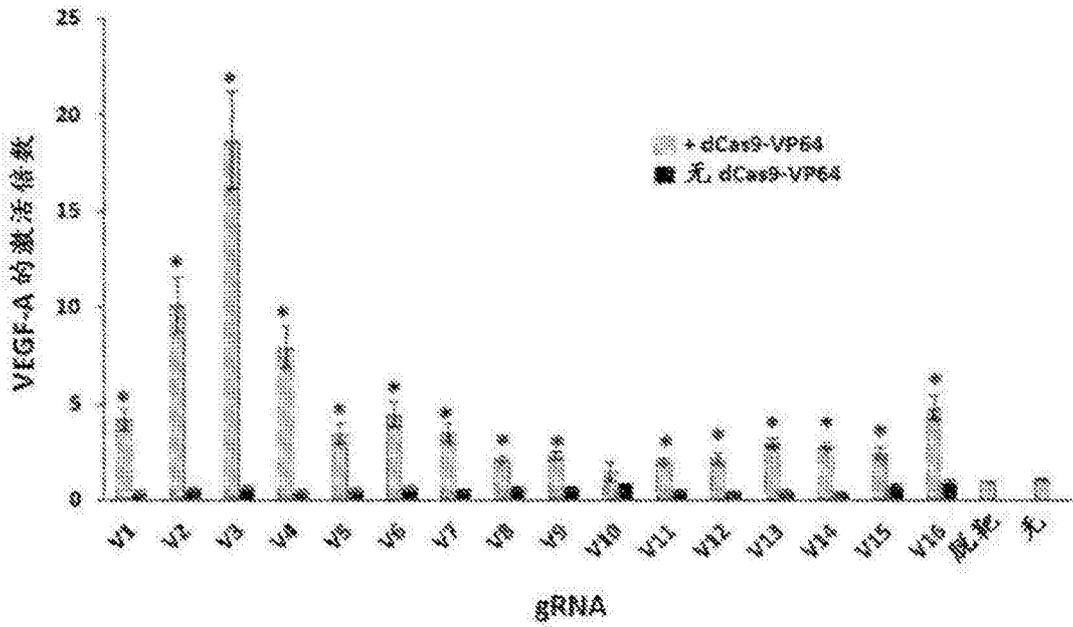


图 2A

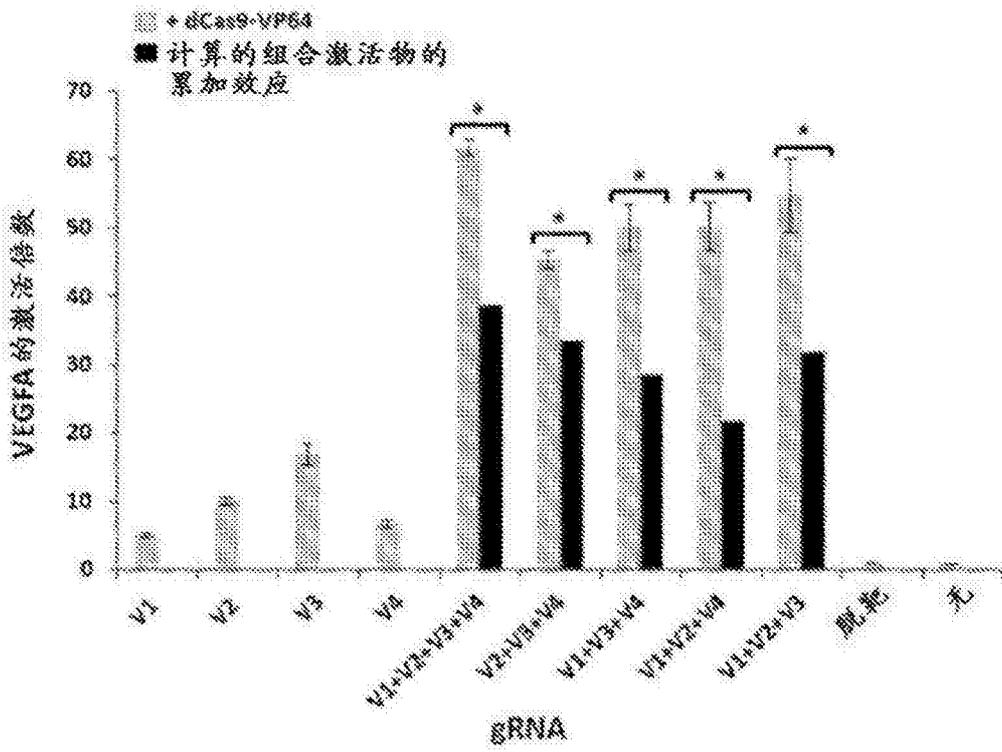


图 2B

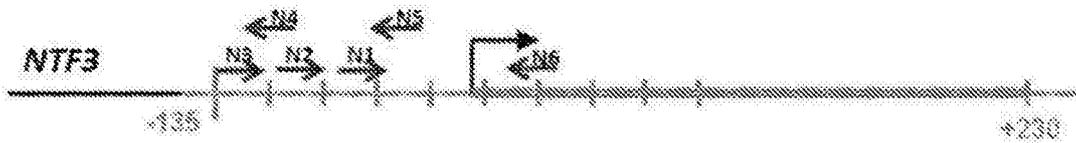


图 3A

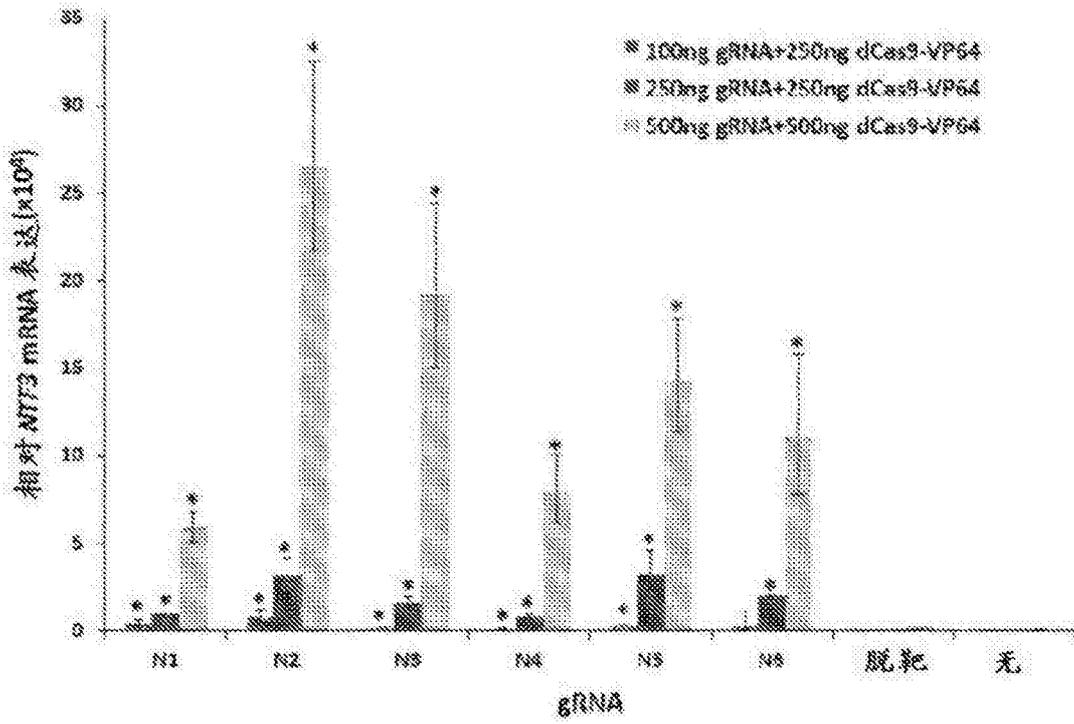


图 3B

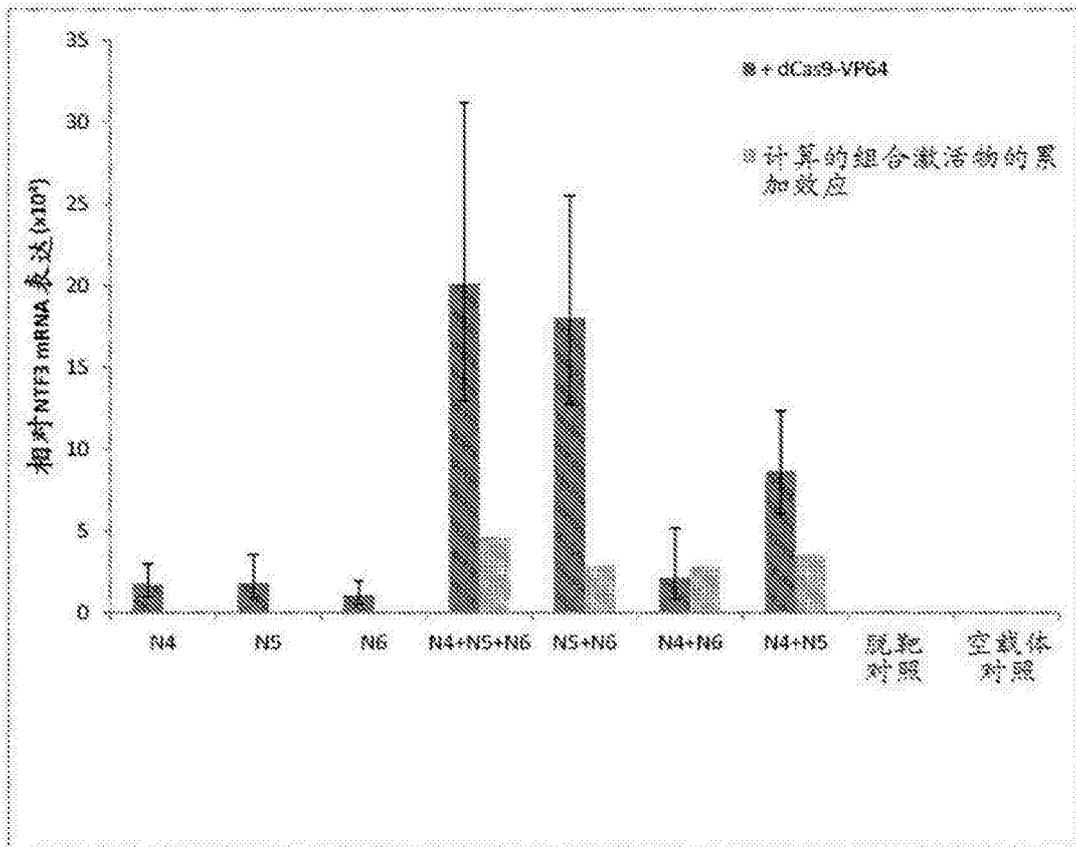


图 3C

图 4 - 引导RNA表达载体序列

```

GACGTGGCTAGCTGTACAAAAAAGCAGGCTTTAAAGGAACCAATTCAGTCCGACTGGATCCGGTACCAG
GGTCGGGCAGGAAGAGGGCCCTATTTCCCATGATTCCTTCATATTTGCATATAACGATACAAAGGCTGTTA
GAGAGATAATTAGAAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAAG
TAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTAATGTTTTAAAATGGACTATCATATGGCTTACCGT
AACTTGAAAAGTATTTGATTTCTTGGCTTTATATATCTTGTGGAAAAGGACGAARCAACNNNNNNNNNN
NNNNNNNNNNNGTTT TAGAGCTAGAAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGA AAAA
GTGGCACCAGTCCGTTGCTTTTTTAAGCTTGGGCCGCTCGAGGTACCTCTCTACATATGACATGTTGA
GCAAAAGGCCAGCAAAAAGGCCAGGAACCGTAAAAAAGGCGGCTTCCTGGCCGTTTTTCCATAGGCTCCG
CCCCCTGACGAGCATCAAAAAATCCACGCTCAAGTCAGAGGTTGGCGAAAACCCGACAGGACTATAAAA
GATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGGCTCTCTCTGTTCCGACCCCTGCCGCTTACCGGA
TACCTGTCCGCTTTCTCCCTTCGGGAAGGCTGGGCTTTCTCATAGCTCAGGCTGTAGGTATCTCAG
TTCCGCTGTAGTCTGTTCCGCTCCAAGCTGGGCTGTGTCCAGCAACCCCGCTTCAGCCCGAAGGCTCCG
CCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCC
ACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGGGGTGCACAGATTTCTTGAAGTGGTGGCCTAA
CTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAAGCCAGTTACCTTCGAAAAA
GAGTTGGTAGCTCTTGATCCGGCAAAACAAACCCGCTGGTAGCGGTGGTTTTTTTTGTTTGCAGCAG
CAGATTACGGCCAGAAAAAAGGATCTCAAGAAGATCTTTGATCTTTCTACGGGGTCTGACGCTCA
GTGGAAAGCAAACTCACGTTAAGGGATTTTTGGTCATGAGATTTATCAAAAAGGATCTTACCTAGATCC
TTTTAAATFAAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTTGGTCTGACAGTTAC
CAATGCTTAATCAGTGAAGCACCCTATCTCAGCAGTCTGTCTATTTCCGTTCCATCCATAGTTGCTGACT
CCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCTCTGGCCCCAGTGGTGCATGATGATACCGC
GAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAAACAGCCAGCCGGAAGGGCCAGCGCAGA
AGTGGTCCCTGCACTTATCCGCTCCATCCAGTCTATTAATTTGTTCCCGGGAAGCTACAGTAAAGTAG
TTCCGCAAGTAAATAGTTTGGCAACGTTGTTGCCATGCTACAGGCATCGTGGTGTACAGCTCTCTCT
TTGGTATGGCTTCATTCAGCTCCCGTTCCCAACGATCAAGGCCAGTTACATGATCCCCCATGTTCTGC
AAAAAAGCGGTTAGCTCCTTCGGTCCCTCCGATCCTTGTGAGAAAGTAAAGTTGGCCCGCAGTGTATCACT
CATGGTATGGCCAGCACTGCATAATTCTCTTACTGTCTATGCCATCCGTAAGATGCTTTTCTGTGACTG
GTGAGTACTCAACCAAGTCATTTCTGAGAATAGTGTATGCGGCGACCCAGTTGCTCTTTGCCCGGCTCA
ATACGGGATAAATACCGGCCACATAGCAGAACTTTAAAAGTCTCATCATTTGAAAAACGTTCTTCGGG
GCCAAAACCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACT
GATCTTCAGCATCTTTACTTTCCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCCAAAATGCCGC
AAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCTTTTTCAATATTATTGAAG
CATTTATCAGGGTATTGCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAG
GGTTCCGCGCACATTTCCCGAAAAGTGCCACCT (SEQ ID NO:107)

```

图 5- CMV-T7-Cas9 D10A/H840A-3XFLAG-VP64:

```

ATATGCCAAGTACGCCCTATTGACGTCAATGACGGTAAATGGCCCGCTGGCATTATGCCCAGTACATGAC
CTTATGGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGTATGCGGTTTTGGC
AGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGATTTCGAAGTCTCCACCCCATGACSTCAATGG
GAGTTTTGTTTTGGCACAAAATCAACGGGACTTTCCAAAATGTCGTAACAACCTCCGCCCATGACGCAAAATG
GGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTAGTGAACCSTCAGATCCGCTAGAG
ATCCGCGGCCGCTAATACGACTCACTATAGGGAGAGCCGCCACCATGGATAAGAAATACTCAATAGGCTTAGC
TATCGGCACAAAATAGCGTCCGATGGGCGGATCACTGATGAATATAAGSTTCCGCTTAAAAAGTTCAAGGTT
CTGGGAAATACAGACCGCCACAGTATCAAAAAAATCTTATAGGGGCTCTTTTATTTGACAGTGGAGAGACAG
CGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTATTTGTTATCTACA
GGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGTCTTTTTGGTGG
AAGAAGACAAGAAGCATGAACGTCATCTATTTTTGGAAATATAGTAGATGAAGTTGCTTATCATGAGAAATA
TCCAACATCTATCATCTGCGAAAAAATTTGGTAGATTCTACTSATAAAGCGGATTTGCGCTTAATCTATTTGG
CCTTAGCGCATATGATTAAGTTTCGTGGTCATTTTTGATTGAGGGAGATTTAAATCCTGATAATAGTGATGTG
GACAAACTATTTATCCAGTTGGTACAAACCTACAATCAATTTTGAAGAAAACCTATTAACGCAAGTGGAGT
AGATGCTAAAGCGATTCTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGCTCAGCTCCCCG
GTGAGAAGAAAAATGGCTTATTTGGGAATCTCATTGCTTTGTCATTGGGTTTGACCCCTAATTTTAAATCAAAT
TTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTAGATAATTTATGGC
GCAAATTTGGAGATCAATATGCTGATTTGTTTTGGCAGCTAAGAATTTATCAGATGCTATTTACTTTTCAGATAT
CCTAAGAGTAAATACTGAAATAACTAAGGCTCCCTATCAGCTTCAATGATTAACGCTACGATGAACATCATC
AAGACTTGACTCTTTTAAAGCTTTAGTTTCGACAACAACCTCCAGAAAAGTATAAAGAAATCTTTTTGATCAAT
CAAAAACCGGATATGCAAGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTATAAATTTATCAAACCAAT
TTTAGAAAAATGATGGTACTGAGGAATTTTGGTAAACTAAATCGTGAAGATTTGCTGCGCAAGCAAGAA
GACCTTTGACAACCGGCTTATTTCCCATCAAATCACTTTGGTGGTGGTGCATGCTATTTTGAAGAACAGAA
GACTTTTTATCCATTTTAAAGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTT
GGTCCATTGGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCATGGA
ATTTTGAAGAAGTTGTCGATAAAGGTTGCTTCAAGCTCAATCATTTATTGAACGCATGACAAACTTTGATAAAAAAT
CTTCCAAATSAAAAAGTACTACCAAAACATAGTTTGTCTTATGAGTATTTTACGGTTTATAACGAATTTGACAAA
GGTCAAATATGTTACTGAAGAAATGCGAAAAACCAGCATTCTTTCAGGTGAACAGAAAGAAAGCCATTGTTGAT
TTACTCTTCAAAAACAATCGAAAAGTAACCGTTAAGCAATTAAGAAGAAGATTATTTCAAAAAAATAGAATGTTT
TGATAGTGTGAAATTTGAGGAGTTGAAGATAGATTTAATGCTTCAATGAGTACCTACCATGATTTGCTAAAAA
TTATTAAGATAAAGATTTTTTGGATAATGAAGAAAATGAAGATATCTTAGAGGATATTGTTTAAACATTGACC
TTATTTGAAGATAGGGAGATGATTGAGSAAAGACTTAAACATATGCTCACCTCTTTGATGATAAGGTGATGA
AACAGCTTAAACSTCGCGTTATACTGGTTGGGGACGTTTGTCTCGAAAATTGATTAATGGTATTAGGGATAA
GCAATCTGGCAAAACAATATTAGATTTTTTGAATCAGATGGTTTTGCAATCGCAATTTTATGCAGCTGATCC
ATGATGATAGTTTGACATTTAAAGAAGACATTCAAAAAGCACAAGTGTCTGGACAAGGCGATAGTTTACATGA
ACATATTGCAATTTAGCTGGTAGCCCTGCTATTAATAAAGGATTTTTACAGACTGTAAAAGTTGTTGATGAAT
TGGTCAAAGTAATGGGGCGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACGTGAAAATCAGACAACCTC
AAAAGGSCAGAAAATTCGCGAGAGCGTATGAAACGAATCGAAGAAGGTATCAAAGAATTAGGAAGTCAG
ATTCTTAAAGAGCATCTGTTGAAAATACTCAATTGCAAAATGAAAAGCTCTATCTTATTATCTCCAAAATGG
AAGAGACATGTATGTGGACCAAGAATTAGATTAATCGTTTTAAGTGAATGATGTCGATGTCATTGTTCCAC
AAAGTTTCTTAAAGACGATTCAATAGACAATAAGGTCTTAAACGCGTCTGATAAAAAATCGTGGTAAATCGGA
TAACGTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAACCTCTAAACGCCAAGTTAATC
ACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTTGGAGTTTGGAGTGAACCTGATAAAGCTGGTTTTA
TCAAACGCCAATTTGGTTGAAACTCGCCAAATCACTAAGCATGTGGCACAAAATTTGGATAGTCGCATGAATAC
TAAATACGATGAAAATGATAAACTTATTCGAGAGGTTAAAGTGATTACCTTAAATCTAAATAGTTTCTGACT
TCCGAAAAGATTTCCAATCTATAAAGTACGTGAGATTAACAATTACCATCATGCCCATGATGCGTATCTAAAT
GCCGTCGTTGGAAGTCTTTGATTAAGAAATATCCAAAACCTGAATCGGAGTTTGTCTATGGTGATTATAAAGT

```

图 5- CMV-T7-Cas9 D10A/H840A-3XFLAG-VP64:

```

TTATGATGTTCTGATAAATGATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAAATATTTCTTTACT
CTAATATCATGAACCTTCTTCAAACAGAAATTACACTTGCAAATGGAGAGATTTCGAAACGCCCTCTAATCGAA
ACTAATGGGGAAACTGGAGAAATTGTCGGGATAAAGGGCCGAGATTTTCCACAGTGCAGCAAAGTATTGTCC
ATGCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAGACAGGCGGATTCTCAAGGAGTCAATTTACCAA
AAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGATCCAAAAAATATGGTGGTTTTGATAGTCC
AACGGTAGCTTATTCAGTCTAGTGGTTGCTAAGGTGGAAAAAGGGAAATCGAAGAAGTAAAAATCCGTTAA
AGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAATCCGATTGACTTTTTAGAAAGCTAAA
GGATATAAGGAAGTTAAAAAGACTTAATCATTAAACTACCTAAATATAGTCTTTTGGAGTTAGAAAACGGTC
GTAAACGGATGCTGGCTAGTGCCTGGAGAATTACAAAAAGSAAATGAGCTGGCTCTGCCAAGCAAATATGTGA
ATTTTTATATTTAGCTAGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTT
GTGGAGCAGCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTCTAAGCGTGTATTTTAGC
AGATGCCAATTTAGATAAAGTCTTAGTSCATATAACAAACATAGAGACAAACCAATACGTGAACAAAGCAGAA
AATATTATTCATTTATTTACGTTGACGAATCTTGGAGCTCCCGCTGCTTTTAAATATTTTGATACAAACAATTGAT
CGTAAACGATATACGCTACAAAAGAAGTTTTAGATGCCACTCTTATCCATCAATECCATCACTGGTCTTTATGA
AACACGCATTGATTTGAGTCAGCTAGGAGGTGACGGTTCTCCAAGAAGAAGAGGAAAGTCTCGAGCGACTA
CAAAGACCATGACGGTATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGGctgcaaggaggcgg
tggaagcgggCGCgCCGACCGCCTGGACGATTCGATCTCGACATGCTGGGTTCTGATGCCCTCGATGACTTTG
ACCTGGATATGTTGGGAAGCGACGCATTGGATGACTTTGATCTGGACATGCTCGGCTCCGATGCTCTGGACGA
TTTCGATCTCGATATGTTATAAecgggtCATCATCACCATCACCATTGAGTTTAAACCCGCTGATCAGECTCGACTG
TGCCTTCTAGTTGCCAGCCATCTGTTGTTGCCCTCCCCGTGCTTCTTACCCTGGAAAGGTGCCACTCCCA
CTGTCTTTCTAATAAAAATGAGGAAATTCATCGCATTGTCTGAGTAGGTGTCAITCTATTCTGGGGGGTGG
GGTGGGGCAGGACAGCAAGGGGGAGGATTBGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCT
ATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCGATACCCTCGACCTTAGCTAGAGCTTGGCGTAATCA
TGGTCATAGCTGTTCTGTGTGAAATTTGTTATCCGCTCACAATCCACACAACATACGAGCCGGAAGCATAAAA
GTGTAAAGCCTACGGTGCTAATGAGTGAGCTAACTACATTAATTGCGTTGCGCTCACTGCCGCTTTCCAGT
CGGTAACCTGCTGCTGCCAGCTGATTAATGAATCGGCCAACCGCGGGGAGAGGCGSTTTGCGTATGGGCG
GCTCTCCGCTTCCCTGCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
AAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACCGCAGGAAAGAACATGTGAGCAAAAAGGCCAGCA
AAAGGCCAGGAACCGTAAAAAGGCCGCTTGTGCGGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCAC
AAAAATCGACGCTCAAGTCAGAGGTGSCSAAACCCGACAGGACTATAAAGATACCAGGCGTTTTCCCCCTGGA
AGCTCCCTCGTGGCTCTCCTGTTCCGACCCTGCCGTTACCBSATACCTGTCCGCTTTCTCCCTTCGGGAAGC
GTGGCGCTTTCTCAATGCTCACGCTGTAGBTATCTCAGTTCGGTGTAGGTCGTTCCGCTCCAAGCTGGGCTGTG
GCACGAACCCCGTTCAGCCCGACCGCTGCSCTTATCCGTAACATATCGTCTTSAGTCCAACCCGTAAGA
CACGACTTATCGCCACTGGCAGCAGCCACTGSTAACAGGATTASCAGAGCGAGGATGTAGGGCGGTGCTACA
GAGTCTTSAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGC
CAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTT
TGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTACGGGGTCT
GACGCTCAGTGGAAEGAAAACCTACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTACCTAGA
TCCTTTTAAATAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAAT
GCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTGTTTCATCCATAGTTGCCTGACTCCCCGCTGCT
AGATAACTACGATACGGGAGGCTTACCATCTGGCCCAAGTGTGCAATGATACCAGCAGACCCACGCTCACC
GGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCTSCAACTTTATC
CGCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAAC
GTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTGGTTTGGTATGGCTTCATCAGCTCCGGTCCCA
ACGATCAAGGCGAGTTACATGATCCCCATGTTGTGCAAAAAAGCGGTTAGCTCTTCGGTCTCCGATCGTT
GTCAGAAGTAAGTTGGCCGAGTGTATCACTCATGGTATGSCAGCACTGCATAATTCTTACTGTCATGCC
ATCCGTAAGATGCTTTCTGTGACTGGTGAAGTACTCAACCAAGTCATCTGAGAATAGTGTATGCGGGCAGCG
AGTTGCTCTTCCCGGCGTCAATACGGGATAATACCAGCCACATAGCAGAACTTAAAAAGTGCTCATATTG
GAAAACGTTCTCGGGGCGAAAACCTCAAGGATCTTACCCTGTTGAGATCCAGTTCGATGTAACCCACTCG
TGACCCCAACTGATCTTACGATCTTTACTTTACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAAT

```

图 5 - CMV-T7-Cas9 D10A/H840A-3XFLAG-VP64:

```
GCCGCAAAAAAGGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTCAATATTATTGAA
GCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTGAATGTATTTAGAAAAATAACAAATAGGGGT
TCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTGACGGATCGGGAGATCGATCTCCCGATCCCCTAGGG
TCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGT
CGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCT
GCTTAGGGTTAGGGCTTTTGCCTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACT
AGTTATTAATAGTAATCAATTACGGGGTCATTAGTTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTAC
GGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCGCCCATTTGACGTCAATAATGACGTATGTTCCCAT
GTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGACTATTACGGTAAACTGCCCACTTGGCAGTAC
ATCAAGTGATCC (SEQ ID NO:108)
```

图 6- MV-T7-Cas9 重编码的 D10A/H840A-3XFLAG-VP64

```

ATATGCCAAGTACGCCCCCTATTGACGTCATGACGGTAAATGGCCCCCTGGCATTATGCCAGTACATGAC
CTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCCTATTACCATGGTGTGCGGTTTTGGC
AGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTCCAAGTCTCCACCCCATGACGTCATGEG
GAGTTTTGTTTTGGCACCAAATCAACGGGACTTTCCAAAATGTCGTAACAACCTCCGCCCATGACGCAAAATG
GGCGGTAGSCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAAAGTGAACCCTCAGATCCGCTAGAG
ATCCGCGGCCGCTAATACGACTCACTATAGGGAGAGCCGCCACCATGGATAAAAAAGTATTCTATTGGTTTAGC
CATCGGCACATAATTCGGTTGGATGGGCTGTCTATAACCGATGAATACAAAGTACCTTCAAAGAAATTTAAGGTG
TTGGGGAACACAGACCCTCATTCGATTAATAAAGAATCTTATCGGTGCCCTCCTATTTCGATAGTGGCGAAACGG
CAGAGGCGACTCGCCTGAAACGAACCGCTCGGAGAAAGGTATACACGTCGCAAGAACCGAATATGTTACTTAC
AAGAAATTTTTAGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTTCCACCGTTTTGGAAGAGTCTTCTTGTG
GAAGAGGACAAGAAACATGAACGGCACCCCATCTTTGGAAACATAGTAGATGAGGTGGCATATCATGAAAAG
TACCCAACGATTTATCAGCTCAGAAAAAAGCTAGTTGACTCAACTGATAAAAGCGSACCTGASSTTAATCTACTT
GGCTCTTGGCCATATGATAAAGTTCCGTGGGCACTTTCTCATTGAGGSGTATCTAAATCCGGACAACCTCGGAT
GTCGACAAACTGTTCCATCCAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGAACCCTATAAATGCAAGTG
GGGTGGATGCGAAGGCTATTCTTAGCGCCCGCCTCTAAATCCCGACGGCTAGAAAACCTGATCGCACAATT
ACCCGGAGAGAAGAAAAATGGSTTGTTCGGTAACCTTATAGCGCTCTCACTAGGCCTGACACCAAATTTAAG
TCGAACCTGACTTAGCTGAAGATGCCAAATTCAGCTTAGTAAGGACACGTACGATGACGATCTCGACAATC
TACTGGCACAATTTGGAGATCAGTATGCGGACTTATTTTGGCTGCCAAAACCTTAGCGATGCAATCCTCCTA
TCTGACATACTGAGAGTTAATACTGAGATTACCAAGCGCCGTTATCCGCTTCAATGATCAAAAGGTACGATG
AACATCACCAAGACTTGACACTTCTCAAGGCCCTAGTCCGTCAGCAACTGCCTGAGAAATAAAGGAAATATT
CTTTGATCAGTCGAAAAACGGGTACGCAAGTTATATTGACGGCGGAGCGAGTCAAGAGGAATTTACAAGTT
TATCAAACCCATATTAGAGAAGATGGATGGGACGGAAGAGTTGCTTGTAAAACCTCAATCGCGAAGATCTACT
GCGAAAAGCAGCGGACTTTGACAACCGGTAGCATTCCACATCAAATCCACTTAGGCGAATTGCATGCTATACTT
AGAAGGCAGGAGGATTTTTATCCGTTCTCAAAGACAATCGTGAAGAGATTGAGAAAATCCTAACCTTTCCGA
TACCTTACTATGTGGGACCCCTGGCCGAGGSAACTCTCGGTTGCGATGGATGACAAGAAAGTCCSAAGAA
CGATTACTCCATGGAATTTGAGGAAAGTTGTGATAAAGGTGCGTCAGCTCAATCGTTCATCGAGAGGATGAC
CACTTTGACAAGAAATTTACCGAACGAAAAAGTATTGCCTAAGCACAGTTTACTTTACGAGTATTTACAGTGT
ACAATGAACCTACGAAAGTTAAGTATGTCACTGAGGGCATGCGTAAACCCGCTTTCTAAGCGGAGAACAGA
AGAAAGCAATAGTAGATCTGTTATTCAAGACCAACCGCAAAGTGACAGTTAAGCAATTGAAAGAGGACTACTT
TAAGAAAATTTGAATGCTTCGATTCTGTGAGATCTCCGGGTAGAAAGATCGATTTAATGCGTCACTTGGTACG
TATCATGACCTCTAAAGATAATTAAGATAAAGGACTTCTGGATAACGAAGAGAAATGAAGATATCTAGAAG
ATATAGTGTGACTCTTACCCTCTTTGAAGATCGGGAAATGATTGAGSAAAGACTAAAACATACGCTCACTY
GTTTCGACGATAAGSTTATGAAACAGTTAAAGAGGGCGTCTATACGSSCTGGGGACGATTGTCGCGGAAACT
TATCAACGGGATAAGAGACAAGCAAAGTGGTAAAACCTATTCTCGATTTTCTAAAGAGCGACGGCTTCGCCAAT
AGGAACTTTATGCAGCTGATCCATGATGACTCTTAACTTCAAAGAGGATATACAAAAGGCACAGGTTTCCG
GACAAGGGGACTCATTGCAGAACATATTGCGAATCTTGTGGTTGCCCAGCCATCAAAAAGGGCATACTCCA
GACAGTCAAAGTAGTGGATGAGCTAGTTAAGSTCATGGGACGTCAAAACCGGAAAACATTGTAATCGAGAT
GGCAGCGGAAAATCAAACGACTCAGAAGGGGCAAAAAACAGTCCGAGAGCGGATGAAGAGAAATGAAAGAG
GGTATTAAGAAGACTGGGACGCCAGATCTTAAAGGAGCATCTGTGGAAAATACCCAATTGCAGAACGAGAAA
CTTTACCTCTATTACCTACAAAATGGAAGGGACATGTATGTTGATCAGGAACTGGACATAAACCGTTTTATCTGA
TTACGACGTCGATGcCATTGTACCCCAATCCTTTTTGAAGGACGATTCAATCGACAATAAAGTGCTTACACGCTC
GGATAAGAACCAGGGGAAAAGTGACAATGTTCCAAGCGAGGAAGTCTGTAAGAAAATSAAGAACTATTBSC
GGCAGCTCCTAAATGCGAAAACGATAACGCAAAGAAAGTTGATAACTTAACTAAAGCTGAGAGGGGTGGCT
TGCTGAACTTGACAAGGCCGGATTTATTAACGTCAGCTCGTGGAAAACCCGCAAATCACAAAGCATGTTGC
ACAGATACTAGATTTCCGAATGAATACGAAATACGACGAGAACGATAAGCTGATTCSGGAAAGTCAAAGTAAT
CACTTTAAAGTCAAATTTGGTGTGCGACTTCAGAAAAGGATTTTCAATTTATAAAGTTAGGGAGATAAATAACT
ACCACCATGCGCACGACGCTTATCTTAATGCCGTCTAGGGACCGCACTCATTAAAGAAATACCCGAAGCTAGA
AAGTGAGTTTGTGTATGGTGATTACAAAGTTTATGACGTCGTAAGATGATCGCGAAAAGCGAACAGGAGAT

```

图 6- MV-T7-Cas9重编码的D10A/H840A-3XFLAG-VP64

```

AGGCAAGGCTACAGCCAAATACTTCTTTTATCTAACATTATGAATTTCTTTAAGACGGAAATCACTCTGGCAA
ACGGAGAGATACGCAAACGACCTTTAATTGAAACCAATGGGGAGACAGGTGAAATCGTATGGGATAAGGGC
CGGGACTTCGCGACGGTGAGAAAAAGTTTGTCCATGCCCAAGTCAACATAGTAAAGAAAACCTGAGGTGCAG
ACCGGAGGGTTTTCAAAGGAATCGATTCTTCCAAAAAGGAATAGTGATAAGCTCATCGCTCGTAAAAAGGACT
GGGACCCGAAAAAGTACGGTGGCTTCGATAGCCCTACAGTTGCCTATTCTGTCTAGTAGTGGCAAAGTTGA
GAAGGGAAAAATCCAAGAACTGAAGTCAGTCAAAGAATTATTGGGATAACGATTATGGAGCGCTCGTCTTT
TGAAAAGAACCCCATCGACTTCTTTGAGGGCAAAGGTTACAAGGAAGTAAAAAGGATCTCATAATTAACCTA
CCAAAGTATAGTCTGTTTGAAGTTAGAAAATGGCCGAAAACGGATGTTGGCTAGCGCCGGAGAGCTTCAAAG
GGGAACGAACTCGCACTACCGTCTAAATACGTGAATTTCTGTATTTAGCGTCCATTACGAGAAGTTGAAAG
GTTACCTGAASATAACGAACAGAAGCAACTTTTTGTTGAGCAGCACAAACATTATCTCSACGAAATCATAGA
GCAAATTTCSGAATTCAGTAAGSABSTCATCTAGCTGATSECAATCTGGACAAAGTATTAAGCGCATACAAC
AAGCACAGGGATAAACCCATACGTGAGCAGGCCGAAAATATTATCCATTTGTTTACTCTTACCAACCTCGGC
CTCCAGCCGATTCAAGTATTTGACACAACGATAGATCGAAACGATACACTTCTACCAAGGAGGTGCTAGA
CGCGACTGATTACCAATCCATCACGGGATTATATGAACTCGGATAGATTGTACAGCTTGGGGGTGAC
GGATCCCCAAGAAGAAGAGGAAAAGTCTCSAGCAGTACAAAGACCATGACGGTGATTATAAAGATCATGAC
ATCGATTACAAGGATGADGATGACAAGGctgcaggagggtggaagcGGGCCGCCGACCGCTGGACGATTCG
ATCTCGACATGCTGGGTTCTGATGCCCTCGATGACTTTGACCTGGATAATGTTGGGAAGCGACGCATTTGGATGA
CTTTGACTGAGACATGCTCGGCTCCGATGCTGACGACTTTCGATCTCGATATGTTATAAcrnggCATCATCACC
ATCACATTGAGTTTAAACCCGCTGATCAGCTCAGCTGACTGTCCTTAGTTGCCAGCCATCTGTTGTTGCCCT
CCCCGTGCCCTTCTTGAACCTGGAAGGTGCCACTCCACTGTCCTTTCTAATAAAAATGAGSAAATTGCATCG
CATTGCTGAGTAGGTGTCATTCTATTCTGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGA
AGACAATAGCAGCATGCTGGGATGCGGTGGGCTCTATGGCTCTGAGGCGSAAAGAACCAGCTGGGGCT
CGATAACGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCTGTGTGAAATTTTATCCG
CTCACAATTCACACAACATACGAGCCGGAAGCATAAAGTGTAAGCCTAGGGTGCCTAATGAGTGAGTAA
CTCACATTAATTGCGTTGCGCTCACTGCCGCTTCCAGTCGGGAAACCTGTGCTGCCAGCTGCATTAATGAAT
CGGCCAACCGCGGGGAGAGGCGGTTTGCATTTGGGCGCTCTTCCGCTTCTCGCTCACTGACTCGCTGCGC
TCGGTCTTTCGGTTCGGCGAGCGGTATCACTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGG
GATAACCGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGSAAACCGTAAAAAGGCCGCGTTGCT
GGGTTTTTCCATAGGCTCCGCCCTGACGAGCATECAAAAAATGACGCTCAAGTCAGAGGTGGCGAAAC
CCGACAGGACTATAAAGATAACAGGGGTTTCCCTGGAAGCTCCCTCGTGGCTCTCTGTTCCGACCTGCC
GCTTACCGGATACCTGTCCGCTTCTCCCTCGGAAAGCGTGGCGCTTCTCAATGCTCACGCTGTAGGTATC
TCAGTTCGGTSTAGGTGCTTCCAAAGCTGGCTGTGTGACGAAACCCCGTTCAGCCGACCGCTGCGC
CTTATCCGGTAACTATGCTTGTAGTCCAACCCGGTAAGACAGACTTATCGCCACTGGCAGCAGCCACTGGT
AACAGSATTAGCAGAGCGAGGTATGTAGGGGGTGTACAGAGTCTTGAAGTGGTGGCCTAACTACGGCTAC
ACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTT
GATCCGGCAAACAACACCGCTGGTAGCGGTGGTTTTTTTGTGCAAGCAGCAGATTACGCGCAGAAAAA
AAGGATCTCAAGAAGATCCTTTGATCTTTCTACGGGGTCTGACGCTCAGTGGAAACGAAAACCTACGTTAAGG
GATTTTGGTCATGAGATTATCAAAAAGGATCTTCACTAGATCCTTTAAATTAATAAATGAATTTTAAATCAA
TCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATC
TGCTATTTCTGTTTATCCATAGTTGCCGACTCCCCGTCGTGTAGATAACTACGATACGGGAGSBBCTTACCATC
TGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACAGCCA
GCCGGAAGSBBCCGAGCGCAGAAAGTGGTCTGCAACTTTATCCGCTCCATCCAGTCTATTAATTGTTGCCGGG
AAGCTAGAGTAAGTAGTTCCGCAAGTAAAGTTTGCACAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTG
ACGCTCGTCTGTTGGTATGGCTTCACTCAGCTCCGTTCCCAACGATCAAGGCGAGTTACATGATCCCCATGT
TGTGCAAAAAGCGGTTAGCTCCTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAAGTGTATCACT
CATGGTTATGGCAGCACTGCATAATTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGGT
ACTCAACCAAGTCACTTCTGAGAATAGTGTATCGGGCAGCCAGTTGCTCTTGGCCGGCGTCAATACGGGATAA
TACCGCGCCACATAGCAGAACTTAAAAGTGTCTCATATTGGAAAACGTTCTTGGGGCGAAAACCTCTCAAGG
ATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAAGTATCTTACGATCTTTACTTTC
ACCAGCCTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAGSAAATAGGGGCGACACGGAA

```

图 6- MV-T7-Cas9重编码的D10A/H840A-3XFLAG-VP64

```

ATGTTGAATACTCATACTCTTCCTTTTCAATATTATTGAAGCATTATCAGGGTTATTGTCTCATGAGCGGATA
CATATTTGAATGTATTTAGAAAAATAACAATAGGGGTTCCGCGCACATTTCCCGAAAAAGTGCCACCTGAC
GTCGACGGATCGGGAGATCGATCTCCCGATCCCTAGGGTCCGACTCTCAGTACAATCTGCTCTGATGCCGCAT
AGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCCGCGAGCAAAATTAAGCTAC
AACAAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGAT
GTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGT
TCATAGCCCATATATGGAGTCCGCGTTACATAACTTACGGTAAATGGCCCGCTGGCTGACCGCCCAACGAC
CCCCGCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTCCATTGACGTCAATG
GGTGGACTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGATC (SEQ ID NO:109)

```

图 7- Cas9-激活蛋白

```

MDKKYSIQLAIGTNSVGNVAVITDEYKVPSSKFKVLTGNTDRRSIKKNLIGALLFDSGETAET
RLKRTAPRRYTRRNRI CYLQEIFSNEMAKVDDSFHRLEESFLVEEDKXHERHPFGNIYD
EVAYHEKYPTIYHLRKKLVDSSTKADLRILIYALAHMTKFRGHFLIEGDLNPNDSVDKLF
QLVQTYNQLFREENPINASGVDAKAIL SARLSKSRLENLIAQLPGEKKNGLFGNLI ALSLGL
TPNFKSNFDLAE DAKLQLSKDTYDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKP ILEKMDGTEELLVKLNREDLLRQRTFDNGSIPHQIHLGELHAILRQEDFYFFLK
ONEEKIEKILTFRI FYVVGPLARGNSRFAMTRKSEETITPWNFEEVVDKGSASAQSFIERMT
NFDKNLFPNEKVL PKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL LKI IKDKDFLDNEENEDILEDIV
LTLTLFRDREMIEERLKYAHLEFDOKVMKQLKRRRYTGWRLSRKLINGIRDKQSGKTI LDF
LKSDGFANENFMQLIHEDSLTFKEDIQKAQVSGQGDLSLHEHIANLAGSPA IKKGILQTVKVV
DELVKVMGRHEPENIVTEMARENQTTQKGGKNSRERMKRIEERGKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINLSDYDVAIVPQSFLKDDSIDNKVLTFRSDKNNGKSD
NVFSEEVVKMKNYWRQLLNAKLITQKRFENLTKAERGGGLSELDKAGPIKRQLVETRQITKR
VAQILDSEMNTRYDENDKLIREVKVIITLKSXLVSDFRKDFQFYKVPBINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVDVVKMI AKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
GEIRKPLIETNGETGEIVWDFGRDFATVRKVL SMPQVNI VKKTEVQTGSFSKESILFKRNS
DKLIARKKDWDPEKYGGFDSETVAYSVLVVAEVEKGRKCKLKSVKELLGITIMERSSEFEKNP
IDFLRAKGYKEVKDLI IKLFPKYSLFELENGRZRMLASAGELQKGNELALPSKYVNFY LAS
HYEELKGSPEEDNEQKQLFVEQHEHYLDEIIEQISEFSRVLILADANLDKVL SAYNKHRRDKPI
REQAENI IHLFTLTNLGAPAAFKYFDFTIDSKRYTSTKEVL DATLIBQSITGLYETRIDLSQ
LGGDGSpkkkrkvsDYKDHEGDYKDHEIDYEDDEKAAGGGGSGRADALDDFDLDMLGSDA

```

LDDFDLDMLGSDALDDFDLDMLGSDALDDFDLDML (SEQ ID NO:110)

图 8A- dCas9-NLS-3XFLAG-HP1α

```

MDKKYSIGLAIGTNSVGVAVITLEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRKKNRICYLQEIFSNEMAKVDDSSFFHRLEESFLVEEDKKHERHPIFGNIYD
SVAYHEKYPTIYHLRKKLVDSTDKADLELIYLAHAMIKFRGHFLIEGDINPINSDDVDKLF I
QLVGTYNQLFEENPINASGVDAKAILSAPLSKSRRLLENLIAQLPGEKKNGLFGNLIALSGLL
TPNFKSNFDLAEDAKLQLSKDTYDDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKRYDEHHQDLTLLKALVEQQQLPRKYKEIFFDQSKNGYAGYIDGGASQSEF
YKFIKPILEKMDGTEELLVVLNRPEDLLRQQTTFDNGSIPHQIHLGELHAILLRQEDFYFFLK
DNREKIEKILTFRI PYYVGPLARGNSRFAMWTRKSEETITPWNFEVVDKASQAQSFIERMT
NFDKNLNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMPKPAFLSGEQKKAIVDLLFKTNEK
VTYKQLKEDYFKKIECPDSEVEISGVEDRFNASLGTYHDLLKI IKDKDFLDNEENEDILEDIV
LTLTLFEDREMIERLKYAHLFDDKVMKQLKRRRYTGWGKLSKLTNGIRDKQSGKTTILDF
LKSDGFANFNFMQLIHDDSLTFKEDIQKAQVSGGGDSLEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVTEMARENQTTCKGQKNSRERMKRIEEGIKELGSQLKEHFVENTQL
QNEKLYLYYLQNGEDMYVDQELDINRLSDYDVAIVPQSFLLKDDSDENKVLTPESDKNRGKSD
NVPSEEVVKNMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIMKQLVETRQITKH
VAQILOSFMNTKYDENDKLIREVKVITLKSSELVSDFFKDFQFYKVVREINNYHEAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVVKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
GEIRKRP LIETNGETGEIVWDKGRDFATVRKVLSPQVNIIVKTEVQTTGGFSKESILPKRNS
DKLIARKNDNDPKKYGGFDSPTVAYSVLVYAKVEKSKKRLKSVKELLGITIMERSSEFEKNP
IDFLEAKGYKEVKRDLI IKLFRKYSLFELENGRERMLASAGELQKGNELALPSKYVNFLYLAS
HYEKLKGSPEDNQKQLFVEQHKRYLDEIIEQISEFSKRVILADANLEKVL SAYNKHPDKPI
EEQAENI IHLFTLTNLGAPAAFKYFDTTIERKRYTSTKEVLDATLIHQSI TGLYETRIDL SQ
LGGDGSFKKRRKVSSEYKDHDIYKDHDI DYKELDDKAAAGGGGSMKEGENNKPREKSESNKE
KSNFNSADDIKSKKREQSNDIARGFERGLEFEKLI GATDSCGDLMLFKKWKDTDEADLVL
AKEANVKCFQIVIAFYEEELTWHAYPEDAENKEKETAKS (SEQ ID NO:111)

```

图 8B- dCas9-NLS-3XFLAG-HP1 α

```

MDKKYSIGLAIGTNSVGVAVITDQYKVPSEKFKVLGNTDRHSIKKNLIGALLFDSSGETASAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSSFFHRLSEESFLVEEDKKEHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSSTDKADLRLLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLPGSLIALSLGL
TPNFKSNFDLAEDAKLQLSKDQYDODLNLQAQIGDQYALLFLAAKNLSQAILLSQILRVNT
EITKAPLSASMIKRYLEHHQDLTLLKALVRQQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLPKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLEK
LNREKIEKILTFRIPIYYVGFARGNSRFAMWTKSEETITFWNFEEVVDKASQAQSFIERMT
NFDKNLPNEKVLFKHSLLYEYFTVYNELTQVYKVTTEEMKPAFLSAGEQKKAIVDLLFKTRNK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLSTYHDLKIKDKDFLDNEENEDILEDIY
LTLTLFEDREMIEERLKYAHLFDKVMKQLRRRYTGWGRLSRKLLINGIRDKQSGKTILDF
LKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGGDSLHEBIANLAGSPAIKKGILOTVKVV
DELVKVMGRHKPFENIVIEARENQTTQKQKNSREPMKRIEKGKELSSQILKEHFVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDQDAIVPQSFLKDESIDNKVLTSSDKNRGKSD
NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETROITKH
VAQILDSRMNTKYDENDKLIPEVKVITLKSCLVSDFRKDFQFYKVFIEINNYHHAHDAYLNAV
VGTALINKYPKLESEFVYGDYKVVYDVRKMIKSEQEI GKATAKYFFYSNIMNFFTEITLAN
GEIRKRP LIETNGETGEIVWDKGRDFATVRSVLSMPQVNIKKTEVQTGGFSKESILPKRNS
DKLIANKKDWDPKKYGGFDSPTVAYSVLVAVKVEKSKKLSVKELLGITIMERSSEKNP
IDFLEARGYKEVKKDLIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFYLYAS
EYKLLKGS PEDNEOKQLFVEQKHXYLDEIIEQISEFSKRVILADANLDKVL SAYNKHRRDKPI
RQQAENI IHLFTLNLGAFAPAFKYFDFTIDRKRYTSTKEVLOATLIHQSI TGLYETRIDLSQ
LGGDGS PKKKRKYSSDYKDHGGDYKDHDIDYKDDDIKAAGGGGSAHETDKSEGGKKAQSD
SEDKGEESKPKKKKESSEKPRGEARGLEPERIIGATDSSGELMELMKWKNSDRADLVPKKA
NYKCPQVVISPYEERLTWH5YPSEODDEEDDEN. (SEQ ID NO:112)

```

框=核定位信号
下划线=三重flag标签
双下划线=HP1 α 铰链和 chromoshadow 结构域

图 9- dCas9-3XFLAG-TET1CD

```

MDKKYSIGLAIGTNSVGVAVITDSEYKVPSEKFKVLGNTORHSIKKNLIGALLFDSCGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSFDKADLNLIYLALAHMIKFRGHFLIEGDLNPNDSVDKLF
QLVQTYNQLFRENPINASGVDAKAILSARLSESRLENLIAQLPGEKKNGLFGNLIJALSLGL
TPNFKSNFDLAEADAKLQLSKDTYDQOLDNLLAQIGDQYADLFLAAKNLSDAILLSLILRVNT
EITKAPLSASMIKRYDERHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRQRTFENGSIIPHQIHLGELHAILRQQRDFYPFLK
DNEEKIEKILTFRIIPYVVGPLARGNSRFAMTRKSEETITFWNFEEVVDGASASAQSFIERMT
NFLKNLFPNEKVLPKHSLLEYEFTVYNELTKVKYVTEGMEKFAFLSGEQKFAIVDLLFKTRNK
VTYKQLKEDYFKKIECFDSEVEISGVEDRFNASLGTYSDDLKI IKDKDFLDNEENEDILEDIV
LTLTLFEDREMIEERLKYAHLEFDOKVMSQLERRRYTGWGRLSKKLLINGIEDKQSGKTILDF
LKSDGFANFNFMQLIHDDSLTFKEDIQMAQVSGQGDSLHEHIANLAGSPAIFKGLLQTVKVV
DELVKVMSRHKPENIVIEMAPENQTFQKQKNSRERMKRIEEGIKELGSOILKEHPVENTQL
QNEKLYLYYLQNGEDMYVDQELDINRLSDYDVAIVPQSFLLKDDSIDNKVLTSDKNRQKSD
NVPSEEVVKMKKNYWRQLLNAKLITQRKFNLTFAERGGLSELDKAGFIKROLVETPQITKH
VAQILDSRMTKYDENDKLIREVKVITLKSCLVDFRQDFQYKVRINNYYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVDVPRMIARSEGEIGKATAKYFFYSNIMNFFKTEITLAN
GEIFKRPLIETNGETGEIVWDKEDFAFVRKVLSPQVNIKKTEVQGGFSSKESILPKRNS
DKLIARKKDWDPEKYGGFDSPTVAYSVLVVAEVEKSKKLLKSVKELLGITIMERSSPERNP
IDFLEAKQYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFYLAS
HYEKIKGSPEDNEQQLFVEQHKHYLDEIEQISEFSKRVILADANLDKVL SAYNKHROKPI
REQASNI IHLFTLNLGAPAAFKYFDTTIDRKYRSTSTKEVLDTLIEQSITGLYETRIDLSD
LGGDGSPKKRRKSSDYKDHDGDYKDHDI DYKDDDDKAAGGGGSLPTCSCLDEVIQKDKGPFY
YTHLGAGPSVAAVREIMENRYGQKGNARIEIVVYTGKEGKSSHGCPIAKWVLRSSSDEEKV
LCLVFRQRTGHECPTAVMVVLIHWWDGIPLPMAURLYTELTENLKSYNCHPTDRCTLNENRT
CTCCGIDPPTCGASESEFGCSWMIYNGCKEGRSPSPRFRIDPSSPLHEKNLEDNLQSLATR
LAPIYKQYAPVAYONQVEYENVARECKLGSKEGEPFSGVTACLDFCAHPHRDIHNMNNGSTV
VCTLTREDNRS LGVTPODEQLHVLPLYKLSDTDFEGSMEGMEAKIKSGAIEVLAPRRYKRTC
ETQVYPRSGKKEAAMTTEVLAHKITRAVEKKEIPRIKRNKNNSTTTNNSKPSLPTLGSNTETV
QPRYKSETEPHFLLKSSDNTKTYSLMFSAPHEVKEASPGFSWSEKTASATPAPLKNQATASC
GESERSSTPHCTMPSGRLSGANAAAADGPGISQLGEVAFLPTLSAPVMEPLINSEPSITGVTE
ELTFHQENHQPSLTSFGDLASSPMEPDEQHSSEADRPFSDEPLSUDPLSPAEEKLPHIDEYW
SLGSEHIFLDANIGGVALAPANGSVLIECAARELHATTEVEHENRNHPTLFLSLVYQHKNLNK
PQHGPFELNKIFFEAKEAKNKKMEASPOKQQAANEGPEQSSEVNEINQIPSHKALTILTHQNVY
TVSPYALTHVAGPYNHVY

```

(SEQ ID NO:113)

框=核定位信号
下划线=三重flag标签
双下划线=TET1CD

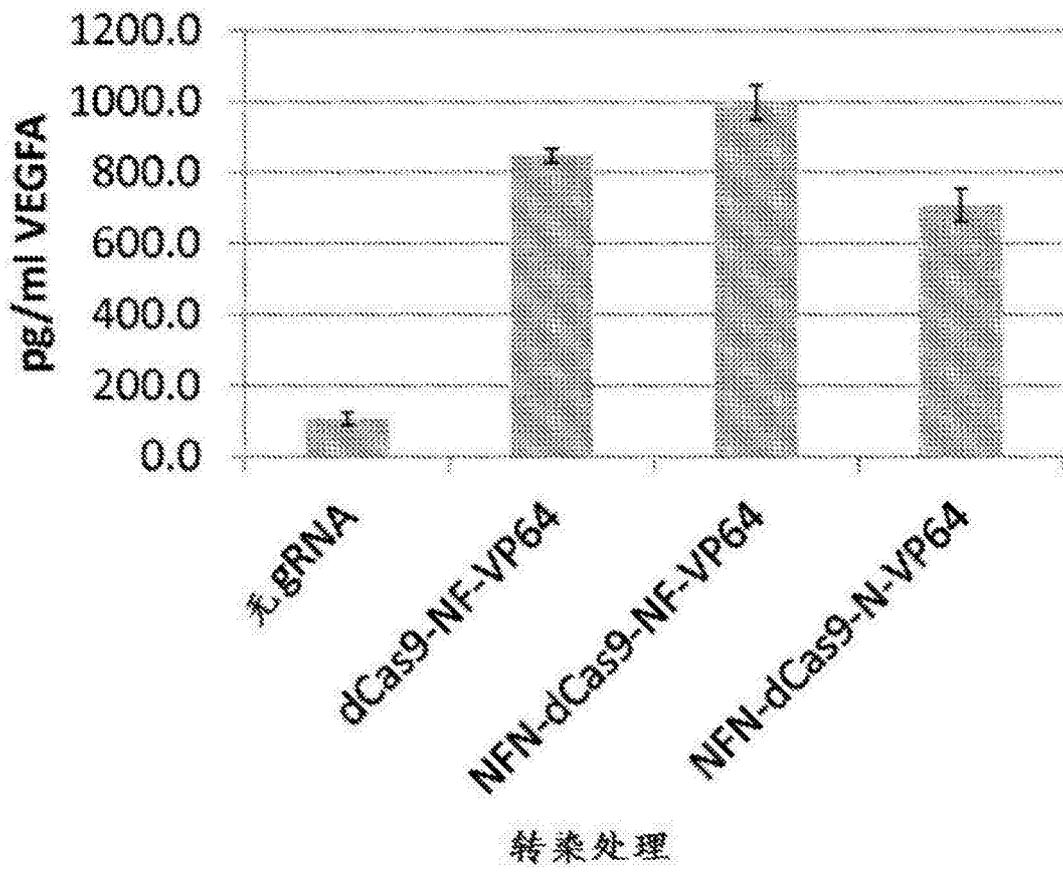


图 10

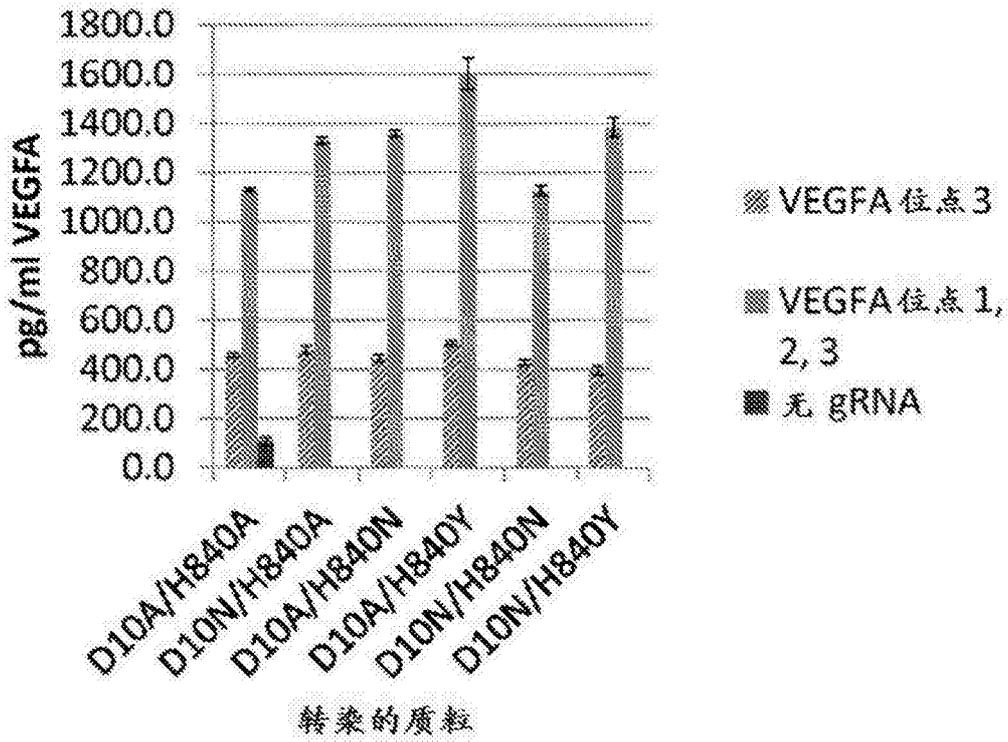


图 11A

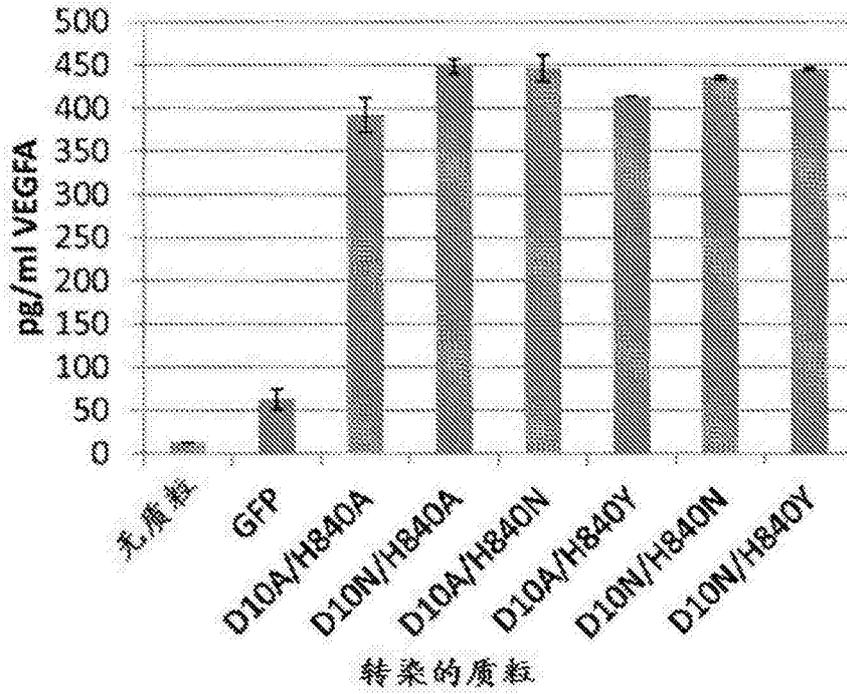


图 11B