



(12)发明专利申请

(10)申请公布号 CN 111315890 A

(43)申请公布日 2020.06.19

(21)申请号 201880071923.2

(72)发明人 章齐轩 梁锡权

(22)申请日 2018.09.07

(74)专利代理机构 中国国际贸易促进委员会专利商标事务所 11038

(30)优先权数据

62/555,862 2017.09.08 US

62/574,936 2017.10.20 US

62/626,792 2018.02.06 US

62/717,403 2018.08.10 US

代理人 傅宇昌

(51)Int.Cl.

C12N 15/63(2006.01)

(85)PCT国际申请进入国家阶段日

2020.05.07

(86)PCT国际申请的申请数据

PCT/US2018/050088 2018.09.07

(87)PCT国际申请的公布数据

W02019/051316 EN 2019.03.14

(71)申请人 生命技术公司

地址 美国加利福尼亚

权利要求书2页 说明书110页 附图59页

(54)发明名称

提高同源重组的方法和其组合物

(57)摘要

本公开涉及用于提高同源重组效率的方法、试剂盒和组合物。具体地说,本公开涉及将用于DNA编辑的核酸切割实体引入难以转染的细胞中的方法。一般来说,不使用病毒载体将所述核酸切割实体引入细胞中。本公开还涉及通过下面的组合使用:启动子捕获,短同源臂、核定位信号,和/或将一种或多种DNA结合剂(与Cas9结合的TAL效应子域或截短的向导RNA)结合到特定位点,将DNA分子直接克隆到基因组中,从而置换或重构目标基因座处的染色质,和/或增加目标基因座的进一步酶促修饰的可及性。本文提供的方法和组合物尤其可用于基因组编辑和增强其中所涉及的酶促过程。

1. 一种遗传改变细胞的方法,所述方法包含:
 - (i) 使细胞与核酸切割实体和供体DNA接触;
 - (ii) 在允许所述核酸切割实体和所述供体DNA被所述细胞吸收的条件下对所述细胞进行电穿孔;以及
 - (iii) 在非同源末端连接 (NHEJ) 抑制剂存在下培养所述细胞,从而形成基因改变的细胞;其中所述细胞是干细胞、免疫细胞、原代细胞,或悬浮生长的细胞。
2. 根据权利要求1所述的方法,其中所述核酸切割实体是锌指蛋白、转录激活因子样效应物 (TALE)、CRISPR复合物、阿格蛋白-核酸复合物、兆碱基大范围核酸酶或大范围核酸酶。
3. 根据权利要求2所述的方法,其中所述核酸切割实体包含转录激活因子样效应物 (TALE)。
4. 根据权利要求2所述的方法,其中所述核酸切割实体包含CRISPR复合物。
5. 根据权利要求1至4中任一项所述的方法,其中所述细胞是原代细胞。
6. 根据权利要求1至5中任一项所述的方法,其中所述细胞是干细胞。
7. 根据权利要求6所述的方法,其中所述干细胞是诱导型多能干细胞 (iPSC)。
8. 根据权利要求1至5中任一项所述的方法,其中所述细胞是免疫细胞。
9. 根据权利要求8所述的方法,其中所述免疫细胞是T细胞、自然杀伤 (NK) 细胞、树突状细胞、B细胞、粒细胞、单核细胞、肥大细胞或嗜中性白细胞。
10. 根据权利要求8所述的方法,其中所述免疫细胞是T细胞。
11. 根据权利要求1至10中任一项所述的方法,其中所述NHEJ抑制剂是DNA依赖性蛋白激酶 (DNA-PK) 抑制剂、DNA连接酶IV抑制剂,或其组合。
12. 根据权利要求11所述的方法,其中所述DNA-PK抑制剂是Nu7026 (2-(4-吗啉基)-4H-萘并[1,2-b]吡喃-4-酮)、Nu7441 (8-(4-二苯并噻吩基)-2-(4-吗啉基)-4H-1-苯并吡喃-4-酮)、Ku-0060648 (4-乙基-N-[4-[2-(4-吗啉基)-4-氧代-4H-1-苯并吡喃-8-基]-1-二苯并噻吩基]-1-哌嗪乙酰胺)、化合物401 (2-(4-吗啉基)-4H-嘧啶并[2,1-a]异喹啉-4-酮)、DMNB (4,5-二甲氧基-2-硝基苯甲醛)、ETP45658 (3-[1-甲基-4-(4-吗啉基)-1H-吡唑并[3,4-d]嘧啶-6-基苯酚)、LTURM 34 (8-(4-二苯并噻吩基)-2-(4-吗啉基)-4H-1,3-苯并噻吩-4-酮)、UNC2170 (3-溴-N-(3-(叔丁基氨基)丙基)苯甲酰胺)、Scr7 (2,3-二氢-6,7-二苯基-2-硫代-4(1H)-蝶啶酮,6,7-二苯基-2-硫基-二氧四氢蝶啶)、咖啡因,和/或P1 103盐酸盐 (3-[4-(4-吗啉基吡啶并[3',2':4,5]呋喃并[3,2-d]嘧啶-2-基]苯酚盐酸盐)。
13. 根据权利要求12所述的方法,其中所述DNA-PK抑制剂是Nu7026、Ku0060648和/或Nu7441。
14. 根据权利要求13所述的方法,其中所述细胞在5 μ M与60 μ M之间的Nu7026、100nM与1000nM之间的Ku0060648或0.3 μ M与2 μ M之间的Nu7441存在下培养。
15. 根据权利要求1至14中任一项所述的方法,其中使所述细胞在电穿孔之前与所述NHEJ抑制剂接触。
16. 根据权利要求1至14中任一项所述的方法,其中使所述细胞在电穿孔之后与所述NHEJ抑制剂接触。
17. 根据权利要求15所述的方法,其中步骤(ii)是使用电穿孔缓冲液来执行,并且在步

骤(ii)之前将所述NHEJ抑制剂添加到所述电穿孔缓冲液中。

18. 根据权利要求1至17中任一项所述的方法,其中步骤(iii)包含将所述细胞在细胞培养基中培养以及将所述NHEJ抑制剂添加到所述细胞培养基中。

19. 根据权利要求1至18中任一项所述的方法,其中在步骤(i)之前,将所述细胞在细胞培养基中培养,并且所述细胞培养基包含所述NHEJ抑制剂。

20. 根据权利要求1至19中任一项所述的方法,其中不将病毒载体插入所述细胞中。

21. 根据权利要求1至20中任一项所述的方法,其中所述供体DNA是聚合酶链反应(PCR)产物。

22. 根据权利要求1至21中任一项所述的方法,其中所述细胞是真核细胞。

23. 一种细胞,其根据权利要求1至22中任一项所述的方法制备,其中所述细胞不包含病毒组分。

24. 一种细胞,其包含非同源末端连接(NHEJ)抑制剂、DNA结合剂和供体DNA,其中所述细胞不包含病毒组分。

25. 一种试剂盒,其包含:

(i) 非同源末端连接(NHEJ)抑制剂;和

(ii) 电穿孔缓冲液。

26. 根据权利要求25所述的试剂盒,其进一步包含DNA结合剂。

27. 根据权利要求25或26所述的试剂盒,其进一步包含供体DNA。

28. 根据权利要求25至27中任一项所述的试剂盒,其进一步包含细胞培养基。

29. 根据权利要求25至28中任一项所述的试剂盒,其中所述DNA结合剂包含锌指蛋白、转录激活因子样效应物(TALE)、CRISPR复合物、阿格蛋白-核酸复合物,或大范围核酸酶。

30. 根据权利要求29所述的试剂盒,其中所述DNA结合剂包含转录激活因子样效应物(TALE)。

31. 根据权利要求29所述的试剂盒,其中所述DNA结合剂包含CRISPR复合物。

32. 根据权利要求25至31中任一项所述的试剂盒,其中所述NHEJ抑制剂是DNA依赖性蛋白激酶(DNA-PK)、DNA连接酶IV、DNA聚合酶1或2(PARP-1或PARP-2),或其组合。

33. 根据权利要求32所述的试剂盒,其中所述DNA-PK抑制剂是Nu7026(2-(4-吗啉基)-4H-萘并[1,2-b]吡喃-4-酮)、Nu7441(8-(4-二苯并噻吩基)-2-(4-吗啉基)-4H-1-苯并吡喃-4-酮)、Ku-0060648(4-乙基-N-[4-[2-(4-吗啉基)-4-氧代-4H-1-苯并吡喃-8-基]-1-二苯并噻吩基]-1-哌嗪乙酰胺)、化合物401(2-(4-吗啉基)-4H-嘧啶并[2,1-a]异喹啉-4-酮)、DMNB(4,5-二甲氧基-2-硝基苯甲醛)、ETP45658(3-[1-甲基-4-(4-吗啉基)-1H-吡唑并[3,4-d]嘧啶-6-基苯酚)、LTURM 34(8-(4-二苯并噻吩基)-2-(4-吗啉基)-4H-1,3-苯并噁嗪-4-酮),和/或P1 103盐酸盐(3-[4-(4-吗啉基吡啶并[3',2':4,5]呋喃并[3,2-d]嘧啶-2-基]苯酚盐酸盐)。

提高同源重组的方法和其组合物

[0001] 相关申请的交叉引用

[0002] 本申请根据35U.S.C. §119 (e) 要求2017年9月8日提交的美国临时申请第62/555,862号和2017年10月20日提交的美国临时申请第62/574,936号和 2018年2月6日提交的美国临时申请第62/626,792号和2018年8月10日提交的美国临时申请第62/717,403号的优先权,所述申请与本申请一起被共同拥有,并且其全部内容以全文引用的方式明确并入本文,如同在本文中完整阐述一样。

技术领域

[0003] 本公开大体上涉及用于提高同源重组效率和准确度的组合物和方法。本公开进一步涉及通过将启动子捕获与短同源臂组合使用而将DNA分子直接克隆到基因组中的方法;以及所克隆的DNA分子的用途,例如在各种分析、组合物、试剂盒、细胞疗法中的用途;以及用于治疗性治疗的方法。

背景技术

[0004] TALEN或CRISPR介导的基因组编辑工具的最新进展使研究人员能够有效地在哺乳动物基因组中引入双链断裂(DSB)。然后通过非同源末端连接(NHEJ)途径或同源定向修复(HDR)途径修复大部分DSB。在哺乳动物细胞中,NHEJ途径是主要的且容易出错。但是,HDR途径允许通过使用姐妹染色单体或外源DNA分子进行精确的基因组编辑。已经进行了许多尝试来提高HDR效率,但是效率仍然较低。举例来说,利用siRNA同时阻断KU70与DNA连接酶IV的基因表达使HDR效率提高4到5倍。参见Chu等人,《自然·生物技术(Nat. Biotechnol)》33:543-548 (2015)。使用Cas9切口酶和长DNA供体模板使得人类胚胎干细胞(hESC)的HDR效率达到5%。参见Rong等人,“使用CRISPR/Cas9切口酶及长DNA供体模板在人类胚胎干细胞中进行同源重组(Homologous recombination in human embryonic stem cells using CRISPR/Cas9 nickase and a long DNA donor template)”,《蛋白质细胞(Protein Cell)》5:258-260 (2014)。最近的一份报告显示,CRISPR系统和子宫内电穿孔技术的组合使用使得脑神经元中的 β -肌动蛋白基因的EGFP整合效率达到约2%。参见Uemura等人,“使用CRISPR/Cas9介导的基因敲入和子宫内电穿孔技术对脑神经元中的内源蛋白质进行荧光蛋白质标记(Fluorescent protein tagging of endogenous protein in brain neurons using CRISPR/Cas9-mediated knock-in and in utero electroporation techniques)”,《科学报告(Sci Rep.)》6: 35861 (2016)。人类POLQ和LIG4的双重损失显示随机整合被消除。但是,还观察到大量未定义的插入。参见Saito等人,“人类POLQ和LIG4的双重损失消除了随机整合(Dual loss of human POLQ and LIG4 abolishes random integration)”,《自然·通讯(Nat. Commun.)》8:16112 (2017)。使用感染复数为10⁶的腺相关病毒(AAV)系统能够以约40%效率将嵌合抗原受体(CAR)整合到TRAC基因座内。参见Eyquem等人,“利用CRISPR/Cas9使CAR靶向TRAC基因座增强了肿瘤排斥反应(Targeting a CAR to the TRAC locus with CRISPR/Cas9 enhances tumour rejection)”,《自然

(Nature) 》543:113-117 (2017)。尽管重组AAV系统被认为可以安全地治疗严重的人类疾病，但生产GMP级 AAV需要建立严格的品质控制系统。

[0005] 传统上使用长同源臂 (500bp至2kb) 将相对较大的DNA片段整合到哺乳动物基因组中，由于效率低和随机整合，因此需要构建靶向载体并筛选大量单一细胞群落。因此，这种方法通常缓慢 (约4至6个月) 且单调，妨碍了使用哺乳动物细胞表达重组蛋白。为了加速蛋白质生产过程，经常使用瞬时基因表达来排除群落筛选步骤。尽管瞬时表达导致高水平的蛋白质产生，但是转基因仅在有限的时间段内表达。因此，使用哺乳动物系统生产重组蛋白变得昂贵。为了满足未来市场对用于生物制药用途的重组蛋白的需求，需要用于快速和高效选择高生产率克隆的成本有效方法。

[0006] 本公开部分地涉及用于编辑核酸分子的组合物和方法。对于修饰基因组的有效系统和技术存在着切实的需求。还需要在不使用病毒载体的情况下 (特别是细胞难以使用传统方法转染时) 编辑核酸分子的有效方法。这些需求以及相关优点在此得到解决。

发明内容

[0007] 此处阐述的组合物和方法旨在改善基因编辑。如本文其它地方所述，已经鉴定出使基因编辑效率增加的许多组合物和方法。

[0008] 已发现使用非同源末端连接 (NHEJ) 抑制剂可增强同源重组的效率。因此，本文提供了使细胞与非同源末端连接 (NHEJ) 抑制剂接触的方法。在实施例中，NHEJ抑制剂是DNA依赖性蛋白激酶抑制剂。可以使用的其它非同源末端连接 (NHEJ) 抑制剂包括一种或多种选自以下组成的群组的化合物：(1) Nu7026、(2) Nu7441、(3) Ku-0060648、(4) DMNB、(5) ETP 45658、(6) LTURM 34、(7) UNC2170、(8) Scr7、(9) 咖啡碱 (caffeine) 和 (10) P1 103盐酸盐。

[0009] 在一个方面中，提供了一种遗传改变细胞的方法。所述方法包括 (i) 使细胞与核酸切割实体和供体DNA接触；(ii) 在允许核酸切割实体和供体DNA 被细胞吸收的条件下对细胞进行电穿孔；和 (iii) 在非同源末端连接 (NHEJ) 抑制剂存在下培养细胞，借此形成基因改变的细胞。

[0010] 在一个方面中，提供了一种遗传改变细胞的方法。所述方法包括 (i) 使细胞与核酸切割实体和供体DNA接触；(ii) 在允许核酸切割实体和供体DNA 被细胞吸收的条件下对细胞进行电穿孔；和 (iii) 在非同源末端连接 (NHEJ) 抑制剂存在下培养细胞，借此形成基因改变的细胞；其中所述细胞是干细胞、免疫细胞或原代细胞。

[0011] 在一个方面中，提供了一种遗传改变细胞的方法。所述方法包括 (i) 使细胞与核酸切割实体和供体DNA接触；(ii) 在允许核酸切割实体和供体DNA 被细胞吸收的条件下对细胞进行电穿孔；(iii) 在非同源末端连接 (NHEJ) 抑制剂存在下培养细胞，借此形成基因改变的细胞；其中细胞悬浮培养生长。

[0012] 在一个方面中，提供了一种基因改变的细胞，其通过本文所述方法发生基因改变。

[0013] 在一个方面中，提供了基因改变的细胞。基因改变的细胞包括非同源末端连接 (NHEJ) 抑制剂、核酸切割实体，和供体DNA，其中所述细胞不包含病毒组分。

[0014] 在一个方面中，提供了一种遗传改变细胞的试剂盒。试剂盒包括 (i) 非同源末端连接 (NHEJ) 抑制剂；(ii) 电穿孔缓冲液；和任选存在的 (iii) 核酸切割实体。

[0015] 一些实施例提供一种通过将启动子捕获与短同源臂组合使用而将相对较大的DNA

分子直接克隆到哺乳动物基因组中的方法。由于高效率 and 特异性,人们可以绕过克隆细胞分离步骤,使用稳定的细胞池生产重组蛋白。

[0016] 本文描述了用于在初始核酸分子中进行同源重组的方法,所述方法包含:在初始核酸分子中产生双链断裂以产生裂解的核酸分子,以及使裂解的核酸分子与供体核酸分子接触,其中初始核酸分子包含启动子和基因,其中供体核酸分子包含:(i) 长度为12bp至250bp的5'和3'末端的匹配端,(ii) 无启动子选择标记物,(iii) 报道基因,(iv) 在无启动子选择标记物的任一侧使无启动子选择标记物和报道基因或LoxP连接的自裂解肽,以及(iv) 任选地存在于无启动子选择标记物和报道基因之间的连接子。

[0017] 在一些实施例中,核酸分子中的双链断裂:(i) 距离ATG起始密码子小于或等于250bp,用于对裂解的核酸分子进行N端标记;或(ii) 距离终止密码子小于或等于250bp,用于对裂解的核酸分子进行C端标记。

[0018] 在一些实施例中,通过至少一种核酸切割实体或电穿孔诱导双链断裂。在一些实施例中,至少一种核酸切割实体包含核酸酶,其包含一种或多种锌指蛋白、一种或多种转录激活因子样效应物(TALE)、一种或多种CRISPR复合物、一种或多种阿格蛋白(argonaute)-核酸复合物,或一种或多种大范围核酸酶。在一些实施例中,使用表达载体、质粒、核糖核蛋白复合物(RNC)或mRNA来施用至少一种核酸切割实体。

[0019] 在一些实施例中,无启动子选择标记物包含蛋白质、抗生素抗性选择标记物、细胞表面标记物、细胞表面蛋白质、代谢物或其活性片段。在一些实施例中,无启动子选择标记物是蛋白质。在一些实施例中,所述蛋白质是粘着斑激酶(FAK)、血管生成素相关生长因子(AGF)受体或表皮生长因子受体(EGFR)。

[0020] 在一些实施例中,无启动子选择标记物是抗生素抗性选择标记物。在一些实施例中,抗生素抗性选择标记物是重组抗体。在一些实施例中,抗生素抗性选择标记物是人IgG抗体。

[0021] 在一些实施例中,报道基因包含荧光蛋白报道体。在一些实施例中,荧光蛋白报道体是翡翠绿荧光蛋白(EmGFP)报道体或橙色荧光蛋白(OPF)报道体。

[0022] 在一些实施例中,无启动子选择标记物:(i) 连接至报道基因的5'端,用于裂解的核酸分子进行N端标记;或(ii) 连接至报道基因的3'端,用于对裂解的核酸分子进行C端标记。

[0023] 在一些实施例中,供体核酸分子包含无启动子选择标记和报道基因之间的接头。在一些实施例中,无启动子选择标记物与报道基因之间的距离小于或等于300nt、240nt、180nt、150nt、120nt、90nt、60nt、30nt、15nt、12nt或9nt。在一些实施例中,所述距离是6个核苷酸。在一些实施例中,接头是聚甘氨酸接头(例如约2至约5个甘氨酸残基)。

[0024] 在一些实施例中,自裂解肽是自裂解2A肽。

[0025] 在一些实施例中,通过PCR扩增将匹配端添加至供体核酸分子的5'和3'末端。

[0026] 在一些实施例中,匹配端共享大于或等于95%的序列一致性。

[0027] 在一些实施例中,匹配端包含单链DNA或双链DNA。

[0028] 在一些实施例中,供体核酸分子的5'和3'末端上的匹配端具有12bp至200 bp、12bp至150bp、12bp至100bp、12bp至50bp或12bp至40bp的长度。在一些实施例中,匹配端具有35个碱基对的长度(bp)。

[0029] 在一些实施例中,初始核酸分子存在于细胞或质粒中。

[0030] 在一些实施例中,供体核酸分子包含小于或等于1kb、2kb、3kb、5kb、10kb、15kb、20kb、25kb或30kb的长度。

[0031] 在一些实施例中,供体核酸分子通过同源定向修复(HDR)整合到裂解的核酸分子中。在一些实施例中,其中HDR大于或等于10%、25%、50%、75%、90%、95%、98%、99%或100%。在一些实施例中,HDR是100%。

[0032] 在一些实施例中,供体核酸分子的整合效率大于或等于50%、75%、90%、95%、98%、99%或100%。在一些实施例中,供体核酸分子的整合效率为100%。

[0033] 在一些实施例中,所述方法进一步包含在5'末端、3'末端或5'和3'末端修饰供体核酸分子。在一些实施例中,供体核酸分子在5'和3'末端被修饰。在一些实施例中,供体核酸分子在至少一端的至少一条链中被一个或多个核酸酶抗性基团修饰。在一些实施例中,一个或多个核酸酶抗性基团包含一个或多个硫代磷酸酯基、一个或多个氨基、2'-O-甲基核苷酸、2'-脱氧-2'-氟核苷酸、2'-脱氧核苷酸、5-C-甲基核苷酸,或其组合。

[0034] 在一些实施例中,所述方法进一步包含用至少一种非同源末端连接(NHEJ)抑制剂处理供体核酸分子。在一些实施例中,至少一种NHEJ抑制剂是DNA依赖性蛋白激酶(DNA-PK)、DNA连接酶IV、DNA聚合酶1或2(PARP-1或PARP-2)或其组合。在一些实施例中,DNA-PK抑制剂是Nu7026(2-(4-吗啉基)-4H-萘并[1,2-b]吡喃-4-酮)、Nu7441(8-(4-二苯并噻吩基)-2-(4-吗啉基)-4H-1-苯并吡喃-4-酮)、Ku-0060648(4-乙基-N-[4-[2-(4-吗啉基)-4-氧代-4H-1-苯并吡喃-8-基]-1-二苯并噻吩基]-1-哌嗪乙酰胺)、化合物401(2-(4-吗啉基)-4H-嘧啶并[2,1-a]异喹啉-4-酮)、DMNB(4,5-二甲氧基-2-硝基苯甲醛)、ETP 45658(3-[1-甲基-4-(4-吗啉基)-1H-吡唑并[3,4-d]嘧啶-6-基苯酚)、LTURM 34(8-(4-二苯并噻吩基)-2-(4-吗啉基)-4H-1,3-苯并噁嗪-4-酮)、UNC2170(3-溴-N-(3-(叔丁基氨基)丙基)苯甲酰胺)、Scr7(2,3-二氢-6,7-二苯基-2-硫代-4(1H)-蝶啶酮,6,7-二苯基-2-硫基-二氧四氢蝶啶)、咖啡碱,或P1 103盐酸盐(3-[4-(4-吗啉基吡啶并[3',2':4,5]呋喃并[3,2-d]嘧啶-2-基]苯酚盐酸盐)。

[0035] 在一些实施例中,哺乳动物是人、哺乳动物实验动物、哺乳动物农场动物、哺乳动物运动动物或哺乳动物宠物。在一些实施例中,哺乳动物是人。

[0036] 在一些实施例中,细胞或质粒通过本文所述的任何同源重组方法制备。在一些实施例中,细胞是真核细胞。在一些实施例中,真核细胞是哺乳动物细胞。

[0037] 本文还描述了一种细胞疗法的方法,包含将有效量的本文所述任何细胞施用于需要其的受试者。

[0038] 在一些实施例中,细胞是T细胞,无启动子选择标记是嵌合抗原受体(CAR)。

[0039] 本文还描述了一种产生无启动子选择标记物的方法,所述方法包含激活通过本文所述的任何同源重组方法制备的细胞或质粒的启动子以产生无启动子的选择标记物。

[0040] 本文还描述了一种组合物,其包含通过用于产生本文所述的无启动子选择标记物的任何方法产生的无启动子选择标记物。

[0041] 本文还描述了一种治疗有需要的受试者的方法,所述方法包含施用有效量的通过本文所述的用于产生无启动子选择标记物的任何方法所产生的无启动子选择标记物。

[0042] 本文还描述了药物筛选分析,其包含通过本文所述的用于产生无启动子选择标记

物的任何方法产生的无启动子选择标记物。

[0043] 本文还描述了用于产生无启动子选择标记物的试剂盒,其包含在选择标记物的任一侧通过自裂解肽或LoxP与报道基因连接的无启动子选择标记物。在一些实施例中,报道基因是GFP或OFP。在一些实施例中,试剂盒进一步包含至少一种核酸切割实体。在一些实施例中,试剂盒进一步包含至少一种NHEJ 抑制剂。在一些实施例中,试剂盒进一步包含一种或多种核酸酶抗性基团。

[0044] 本文还描述了重组抗体表达盒,其包含:位于盒的5'和3'末端的匹配端,其中匹配端的长度小于或等于250bp;无启动子选择标记物;报道基因;连接无启动子选择标记物和报道基因的自裂解肽;以及任选地存在于无启动子选择标记物与报道基因之间的接头,其中无启动子选择标记物连接于报道基因的5'末端以便对裂解的核酸分子进行N端标记,或连接于报道基因的3'末端以便对裂解的核酸分子进行C端标记。

[0045] 本文还描述了用于改变细胞内存在的内源性核酸分子的组合物和方法,所述方法包含将供体核酸分子(例如供体DNA分子)引入细胞中,其中供体核酸分子可操作地连接至一个或多个细胞内靶向部分,所述细胞内靶向部分能够将供体核酸分子定位于细胞中内源核酸分子所处位置。

[0046] 在一些实施例中,细胞中内源性核酸分子所处的位置存在于细胞核、线粒体或叶绿体中。

[0047] 在一些方面中,提供了基因编辑蛋白以及相关方法,从而使细胞内核酸分子发生有效的位点特异性裂解,即使少量引入细胞中也是如此。因此,提供了即使在低浓度存在时也允许高水平的位点特异性裂解的组合物和方法。许多因素可能影响发生的细胞内核酸裂解的量。此类因素包括(1)与预定裂解的基因座接触的活性基因编辑试剂的量;(2)基因编辑试剂所表现出的裂解活性的水平;以及(3)紧邻裂解位点的供体核酸的量。更一般来说,在细胞群中的特定细胞内基因座发生的编辑的量是根据其中至少一个基因座发生裂解的细胞相对于二倍体细胞的百分比来测定。

[0048] 在一些实施例中,一个或多个细胞内靶向部分是核定位信号。在一些实施例中,核定位信号可操作地连接至供体核酸分子的5'末端。

[0049] 在一些实施例中,供体核酸分子可操作地连接至至少一个核酸切割实体。在一些实施例中,至少一种核酸切割实体包含核酸酶,所述核酸酶包含一种或多种锌指蛋白、一种或多种转录激活因子样效应物(TALE)、一种或多种 CRISPR复合物、一种或多种阿格蛋白-核酸复合物、一种或多种大范围核酸酶,或一种或多种兆碱基大范围核酸酶。

[0050] 在一些实施例中,供体DNA分子不连接至核酸切割实体。

[0051] 在一些实施例中,供体核酸分子(例如供体DNA分子)的长度为约25至约8,000个核苷酸(例如约25至约8,000个核苷酸,约25至约5,000个核苷酸,约25至约3,000个核苷酸,约25至约2,000个核苷酸,约25至约1,500个核苷酸,约30至约100个核苷酸,约30至约200个核苷酸,约50至约500个核苷酸,约50至约2,000个核苷酸,约50至约8,000个核苷酸,约75至约2,000个核苷酸,约250至约5,000个核苷酸等)。可能需要短供体核酸分子的一个实例是SNP插入或校正。例如,在这种情况下,供体核酸分子可具有两个各有15个核苷酸的同源臂和一个用于改变目标基因座的单个核苷酸。

[0052] 此外,供体核酸分子可以是单链、双链、线性或环状的。

[0053] 另外,供体核酸分子可以在至少一个末端的50个核苷酸内具有一个或多个核酸酶抗性基团。这些核酸酶抗性基团可以是硫代磷酸酯基团。此外,两个硫代磷酸酯基团可以定位于至少一个末端的50个核苷酸内。

[0054] 在一些实施例中,供体核酸分子包含正选择标记物和/或负选择标记物。此外,负选择标记物可以是单纯疱疹病毒胸苷激酶。

[0055] 在某些实施例中,供体核酸分子具有与细胞中存在的目标基因座序列互补的两个区域。此外,正选择标记物当存在时,其可以位于供体核酸分子的序列互补的两个区域之间。另外,负选择标记物当存在时,其可以不位于供体核酸分子的序列互补的两个区域之间。换句话说,负选择标记物可以位于序列互补的两个区域之外。

[0056] 在一些实施例中,可操作地连接至一个或多个细胞内靶向部分的供体核酸分子可以与结合本文阐述的其它组合物和方法使用,所述细胞内靶向部分能够使供体DNA分子定位于细胞中内源核酸分子所处的位置。因此,本文进一步提供了使细胞另外与以下一种或多种接触的方法:(1)一种或多种核酸切割实体,(2)一种或多种编码核酸切割实体的至少一种组分的核酸分子,(3)一种或多种DNA结合调节增强剂,(4)一种或多种编码DNA结合调节增强剂的至少一种组分的核酸分子,或(5)一种或多种非同源末端连接(NHEJ)抑制剂。

[0057] 另外,可以通过使用为了在目标基因座切割细胞内DNA而设计的基因编辑试剂将可操作地连接至一个或多个细胞内靶向部分的供体核酸分子引入细胞中。因此,一种或多种核酸切割实体中的至少一种可以选自由以下组成的组:(1)锌指核酸酶,(2)TAL效应物核酸酶,和(3)CRISPR复合物。类似地,本文公开了一种或多种DNA结合调节增强剂中的至少一种的用途,所述DNA结合调节增强剂选自由以下组成的组:(1)锌指核酸酶,(2)TAL效应物核酸酶,和(3)CRISPR复合物。此外,一种或多种DNA结合调节增强剂中的至少一种当使用时,可以设计成在目标基因座的50个核苷酸内结合。

[0058] 还包括在真核细胞中进行同源重组的方法,这些方法包含使细胞与以下接触:(1)供体核酸分子(例如供体DNA分子)和(2)(i)核酸切割实体,(ii)编码核酸切割实体的核酸,或(iii)核酸切割实体的至少一种组分和编码核酸切割实体的至少一种组分的核酸,其中供体核酸分子与能够将供体核酸分子定位到细胞中内源核酸分子所处位置的细胞内靶向部分结合。

[0059] 此类方法进一步包括使细胞与以下一种或多种接触:(1)一种或多种非同源末端连接(NHEJ)抑制剂,(2)一种或多种DNA结合调节增强剂,(3)一种或多种编码DNA结合调节增强剂的核酸,和(4)一种或多种DNA结合调节增强剂的至少一种组分和编码一种或多种DNA结合调节增强剂的至少一种组分的核酸。

[0060] 进一步包括包含核酸分子(例如DNA分子)的组合物,其中所述核酸分子与一个或多个细胞内靶向部分共价连接,并且其中所述核酸分子的长度为约25个核苷酸至约8,000个核苷酸(例如约25至约8,000个核苷酸,约25至约5,000个核苷酸,约25至约3,000个核苷酸,约25至约2,000个核苷酸,约25至约1,500个核苷酸,约30至约100个核苷酸,约30至约200个核苷酸,约50至约500个核苷酸,约50至约2,000个核苷酸,约50至约8,000个核苷酸,约75至约2,000个核苷酸,约250至约5,000个核苷酸等)。在某些情况下,核酸分子是供体核酸分子(例如供体DNA分子)。在一些情况下,一个或多个细胞内靶向部分是核定位信号。在另外的情况下,两个或更多个细胞内靶向部分(例如核定位信号、叶绿体靶向信号、线粒体

靶向信号等)与核酸分子共价连接。

[0061] 在一个方面中,提供了一种增加细胞中的目标基因座的可及性的方法。所述方法包括:(1)将第一DNA结合调节增强剂引入包含编码目标基因座的核酸的细胞中,其中所述第一DNA结合调节增强剂不是细胞内源的;以及(2)使第一DNA结合调节增强剂结合目标基因座的第一增强子结合序列,相对于第一DNA结合调节增强剂的缺乏,借此增加目标基因座的可及性。

[0062] 在一个方面中,提供了一种置换细胞中的目标基因座的染色质的方法。所述方法包括:(1)将第一DNA结合调节增强剂引入包含编码目标基因座的核酸的细胞中,其中所述第一DNA结合调节增强剂不是细胞内源的;以及(2)使第一DNA结合调节增强剂结合目标基因座的第一增强子结合序列,借此置换目标基因座的染色质。

[0063] 在一个方面中,提供了一种重构细胞中的目标基因座的染色质的方法。所述方法包括:(1)将第一DNA结合调节增强剂引入包含编码目标基因座的核酸的细胞中,其中所述第一DNA结合调节增强剂不是细胞内源的;以及(2)使第一DNA结合调节增强剂结合目标基因座的第一增强子结合序列,借此重构目标基因座的染色质。

[0064] 在一个方面中,提供了一种增加细胞中的目标基因座的可及性的方法。所述方法包括(1)向包括编码目标基因座的核酸的细胞中引入:(i)第一DNA结合调节增强剂,其中所述第一DNA结合调节增强剂不是细胞内源的;和(ii)第二DNA结合调节增强剂,其中所述第二DNA结合调节增强剂不是细胞内源的。(2)使第一DNA结合调节增强剂结合目标基因座的第一增强子结合序列;以及(3)使第二DNA结合调节增强剂结合目标基因座的第二增强子结合序列,相对于第一DNA结合调节增强剂或第二DNA结合调节增强剂的缺乏,借此增强目标基因座的可及性。

[0065] 在一个方面中,提供了一种置换细胞中的目标基因座的染色质的方法。所述方法包括:(1)向包含编码目标基因座的核酸的细胞中引入:(i)第一DNA结合调节增强剂,其中所述第一DNA结合调节增强剂不是细胞内源的;和(ii)第二DNA结合调节增强剂,其中第二DNA结合调节增强剂不是细胞内源的。(2)使第一DNA结合调节增强剂结合目标基因座的第一增强子结合序列;以及(3)使第二DNA结合调节增强剂结合目标基因座的第二增强子结合序列,借此置换目标基因座的染色质。

[0066] 在一个方面中,提供了一种重构细胞中的目标基因座的染色质的方法。所述方法包括:(1)向包含编码目标基因座的核酸的细胞中引入:(i)第一DNA结合调节增强剂,其中所述第一DNA结合调节增强剂不是细胞内源的;和(ii)第二DNA结合调节增强剂,其中第二DNA结合调节增强剂不是细胞内源的。(2)使第一DNA结合调节增强剂结合目标基因座的第一增强子结合序列;以及(3)使第二DNA结合调节增强剂结合目标基因座的第二增强子结合序列,借此重构目标基因座的染色质。

[0067] 在一个方面中,提供了一种增强调节蛋白或调节复合物在细胞中的目标基因座处的活性的方法。所述方法包括(1)向包含编码目标基因座的核酸的细胞中引入:(i)能够结合目标基因座的调节子结合序列的第一调节蛋白或第一调节复合物,其中所述调节子结合序列包括调节位点;和(ii)能够结合目标基因座的第一增强子结合序列的第一DNA结合调节增强剂。以及(2)使第一DNA结合调节增强剂结合第一增强子结合序列,借此增强第一调节蛋白或第一调节复合物在细胞中的目标基因座处的活性。

[0068] 在一个方面中,提供了一种调节细胞中的目标基因座的方法。所述方法包括(1)向包含编码目标基因座的核酸的细胞中引入:(i)能够结合目标基因座的调节子结合序列的第一调节蛋白或第一调节复合物,其中所述调节子结合序列包括调节位点;和(ii)能够结合目标基因座的第一增强子结合序列的第一DNA结合调节增强剂。以及(2)使第一调节蛋白或第一调节复合物调节调节位点,借此调节细胞中的目标基因座。

[0069] 在实施例中,所述方法包括引入能够结合目标基因座的第二增强子结合序列的第二DNA结合调节增强剂。

[0070] 在实施例中,第一调节蛋白或第一调节复合物不是细胞内源的。

[0071] 在实施例中,相对于第一DNA结合调节增强剂的缺乏,目标基因座处的同源重组速率增加。

[0072] 在实施例中,第二增强子结合序列通过调节子结合序列与第一增强子结合序列连接。

[0073] 在实施例中,所述方法进一步包括引入能够结合调节子结合序列的第二调节蛋白或第二调节复合物。

[0074] 在实施例中,第一调节蛋白或第二调节蛋白包括DNA结合蛋白或DNA调节酶。在实施例中,DNA结合蛋白是转录阻遏物或转录激活物。在实施例中,DNA调节酶是核酸酶、脱氨酶、甲基化酶或脱甲基酶。

[0075] 在实施例中,第一调节蛋白或第二调节蛋白包括组蛋白调节酶。在实施例中,组蛋白调节酶是脱乙酰酶或乙酰酶。

[0076] 在实施例中,第一调节蛋白是第一DNA结合核酸酶结合物。在实施例中,第二调节蛋白是第二DNA结合核酸酶结合物。在实施例中,第一DNA结合核酸酶结合物包括第一核酸酶,并且第二DNA结合核酸酶结合物包括第二核酸酶。在实施例中,第一核酸酶和第二核酸酶形成二聚体。在实施例中,第一核酸酶和第二核酸酶独立地是转录激活因子样效应物核酸酶(TALEN)。

[0077] 在实施例中,第一DNA结合核酸酶结合物包括可操作地连接至第一核酸酶(TALEN)的第一转录激活因子样(TAL)效应域。在实施例中,第一DNA结合核酸酶结合物包括可操作地连接至第一FokI核酸酶的第一TAL效应域。在实施例中,第二DNA结合核酸酶结合物包括可操作地连接至第二核酸酶(TALEN)的第二TAL效应域。在实施例中,第二DNA结合核酸酶结合物包括可操作地连接至第二FokI核酸酶的第二TAL效应域。在实施例中,第一DNA结合核酸酶结合物包括第一锌指核酸酶。在实施例中,第二DNA结合核酸酶结合物包括第一锌指核酸酶。

[0078] 在实施例中,第一调节复合物是第一核糖核蛋白复合物。在实施例中,第二调节复合物是第二核糖核蛋白复合物。在实施例中,第一核糖核蛋白复合物包括与gRNA结合的CRISPR相关蛋白9(Cas9)结构域或与向导DNA(gDNA)结合的阿格蛋白结构域。在实施例中,第二核糖核蛋白复合物包括与gRNA结合的CRISPR相关蛋白9(Cas9)结构域或与向导DNA(gDNA)结合的阿格蛋白结构域。

[0079] 在实施例中,第一调节蛋白、第一调节复合物、第二调节蛋白或第二调节复合物不是细胞内源的。在实施例中,第一调节蛋白和第二调节蛋白不是细胞内源的。在实施例中,第一调节复合物和第二调节复合物不是细胞内源的。在实施例中,第一DNA结合调节增强剂

或第二DNA结合调节增强剂不是细胞内源的。在实施例中，第一DNA结合调节增强剂和第二DNA结合调节增强剂不是细胞内源的。

[0080] 在实施例中，第一DNA结合调节增强剂是第一DNA结合蛋白或第一DNA结合核酸。在实施例中，第一DNA结合调节增强剂是第一转录激活因子样(TAL)效应蛋白或截短的第一向导RNA(gRNA)。

[0081] 在实施例中，第二DNA结合调节增强剂是第二DNA结合蛋白或第二DNA结合核酸。在实施例中，第二DNA结合调节增强剂是TAL效应蛋白或截短的gRNA。

[0082] 在实施例中，第一DNA结合调节增强剂是第一TAL效应蛋白并且第二DNA结合调节增强剂是第二TAL效应蛋白。在实施例中，第一DNA结合调节增强剂是TAL效应蛋白并且第二DNA结合调节增强剂是截短的gRNA。在实施例中，第一DNA结合调节增强剂是第一截短的gRNA并且第二DNA结合调节增强剂是截短的第二gRNA。在实施例中，第一DNA结合调节增强剂是截短的gRNA并且第二DNA结合调节增强剂是TAL效应蛋白。

[0083] 在实施例中，第一调节蛋白是第一DNA结合核酸酶结合物并且第二调节蛋白是第二DNA结合核酸酶结合物。在实施例中，第一调节蛋白是DNA结合核酸酶结合物并且第二调节复合物是核糖核蛋白复合物。在实施例中，第一调节复合物是第一核糖核蛋白复合物并且第二调节复合物是第二核糖核蛋白复合物。在实施例中，第一调节复合物是核糖核蛋白复合物并且第二调节蛋白是DNA结合核酸酶结合物。

[0084] 在实施例中，第一增强子结合序列和/或第二增强子结合序列独立地与调节子结合序列相隔少于200个核苷酸(例如约5至约180、约10至约180、约20至约180、约5至约90、约5至约70、约5至约60、约5至约50、约5至约40、约5至约30、约15至约80、约15至约60、约15至约50、约15至约40、约20至约40、约20至约40个核苷酸等)。在实施例中，第一增强子结合序列独立地与调节子结合序列相隔少于150个核苷酸。在实施例中，第一增强子结合序列和/或第二增强子结合序列与调节子结合序列相隔少于100个核苷酸。在实施例中，第一增强子结合序列和/或第二增强子结合序列独立地与调节子结合序列相隔少于50个核苷酸。在实施例中，第一增强子结合序列和/或第二增强子结合序列独立地与调节子结合序列相隔4至30个核苷酸。在实施例中，第一增强子结合序列和/或第二增强子结合序列独立地与调节子结合序列相隔7至30个核苷酸。在实施例中，第一增强子结合序列和/或第二增强子结合序列与调节子结合序列相隔4个核苷酸、7个核苷酸、12个核苷酸、20个核苷酸或30个核苷酸。

[0085] 在实施例中，第一增强子结合序列和/或第二增强子结合序列独立地与调节位点相隔10至40个核苷酸。在实施例中，第一增强子结合序列和/或第二增强子结合序列独立地与调节位点相隔33个核苷酸。

[0086] 在实施例中，第一增强子结合序列具有SEQ ID NO:26、SEQ ID NO:28、SEQ ID NO:30、SEQ ID NO:32、SEQ ID NO:34、SEQ ID NO:36、SEQ ID NO:38或SEQ ID NO:40的序列。在实施例中，第二增强子结合序列具有SEQ ID NO:27、SEQ ID NO:29、SEQ ID NO:31、SEQ ID NO:33、SEQ ID NO:35、SEQ ID NO:37、SEQ ID NO:39或SEQ ID NO:41的序列。

[0087] 在实施例中，第一DNA结合调节增强剂或第二DNA结合调节增强剂增强第一调节蛋白、第一调节复合物、第二调节蛋白或第二调节复合物在调节位点的活性。

[0088] 在一个方面中，提供了一种细胞，其包含编码目标基因座调节复合物的核酸。所述

复合物包括：(i) 目标基因座，其包括第一增强子结合序列和包含调节位点的调节子结合序列；(ii) 与调节子结合序列结合的第一调节蛋白或第一调节复合物；和(iii) 与第一增强子结合序列结合的第一DNA结合调节增强剂。

[0089] 在实施例中，目标基因座进一步包括通过调节子结合序列与第一增强子结合序列连接的第二增强子结合序列。

[0090] 在实施例中，细胞包括与第二增强子结合序列结合的第二DNA结合调节增强剂。

[0091] 在一个方面中，提供了包含编码目标基因座复合物的核酸的细胞。所述复合物包括(i) 包含第一增强子结合序列的目标基因座；和(ii) 与第一增强子结合序列结合的第一DNA结合调节增强剂，其中所述第一DNA结合调节增强剂不是细胞内源的，并且其中相对于第一DNA结合调节增强剂的缺乏，所述第一DNA结合调节增强剂能够增强目标基因座的可及性。

[0092] 在一个方面中，提供了包含编码目标基因座复合物的核酸的细胞。所述复合物包括(1) 目标基因座，所述目标基因座包括：(i) 第一增强子结合序列；和(ii) 第二增强子结合序列。(2) 与目标基因座的第一增强子结合序列结合的第一DNA结合调节增强剂，其中所述第一DNA结合调节增强剂不是细胞内源的；和(3) 与目标基因座的第二增强子结合序列结合的第二DNA结合调节增强剂，其中所述第二DNA结合调节增强剂不是细胞内源的，其中相对于第一DNA结合调节增强剂和第二DNA结合调节增强剂的缺乏，第一DNA结合调节增强剂和第二DNA结合调节增强剂能够增加目标基因座的可及性。

[0093] 在某些方面中，提供了试剂盒。本文提供的试剂盒可包括以下一种或多种：(i) 第一调节蛋白、第一调节复合物；(ii) 第一DNA结合调节增强剂；(iii) 一种或多种核酸分子；(iv) 一个或多个细胞内靶向部分；和(v) 一种或多种非同源末端连接抑制剂。

[0094] 本文还提供了基因编辑试剂，例如Cas9蛋白，以及编码此类试剂的核酸，此类试剂包含两个或更多个(例如约两个到约十二个、约三个到约十二个、约四个到约十二个、约五个到约十二个、约两个到约七个、约三个到约七个等) 核定位信号(NLS)(例如非经典、单分型和/或双分型NLS)。示例性的Cas9 蛋白是包含两个或更多个双分型核定位信号(NLS)的那些蛋白质。此外，两个或多个双分型核定位信号的全部或一些可以位于至少一个末端的二十个氨基酸之内，例如Cas9蛋白的N端和/或C端。这里的位置是指NLS中的最靠近末端的部分。因此，如果NLS的C端氨基酸后面是十个其它氨基酸，而最后一个氨基酸是蛋白质的C端，则NLS的位置距C端为十一个氨基酸。换句话说，位置计数由NLS的最后一个氨基酸决定。

[0095] 此外，基因编辑试剂(例如Cas9蛋白)可以包含氨基酸序列不同或具有相同氨基酸序列的NLS。而且，基因编辑试剂(例如Cas9蛋白)可以包含一个或多个(例如约一至约五、约一至约四个等) 亲和标签。结合基因编辑试剂使用可以包含以下氨基酸序列中的一种或多种：(A) KRTAD GSEFE SPKKK RKVE (SEQ ID NO:48)、(B) KRTAD GSEFE SPKKA RKVE (SEQ ID NO: 49)、(C) KRTAD GSEFE SPKKK AKVE (SEQ ID NO:50)、(D) KRPA TKKAG QAKKK K (SEQ ID NO:51)，以及(E) KRTAD GSEFEP AAKRV KLDE (SEQ ID NO:52)。结合基因编辑试剂使用的NLS可以包含一种或多种氨基酸序列，所述氨基酸序列属于以下式中的一种或多种的范围：(A) $KRX_{5-15}KKN_1N_2KV$ (SEQ ID NO:53)、(B) $KRX_{(5-15)}K(K/R)(K/R)_{1-2}$ (SEQ ID NO:54)、(C) $KRX_{(5-15)}K(K/R)X(K/R)_{1-2}$ (SEQ ID NO:55)，其中X是长度为5到15个氨基酸的氨基酸序列且其中 N_1 是L或A，且其中 N_2 是L、A或R。另外，可以用于本文阐述的组合物和方法中的所主张的特定Cas9

蛋白包含图41和图42中所示的氨基酸序列。

[0096] 其它目标和优势将在下面的说明中部分地阐述,且根据说明将部分地显而易见,或可以通过实施来学习。借助于所附权利要求书中具体指出的要素和组合将实现且达成所述目标和优势。

[0097] 应理解,以上一般描述和以下详细描述仅具有示例性和解释性,并且不限制权利要求书。

[0098] 并入本说明书中并且构成本说明书的一部分的附图说明了若干实施例,并且与说明书一起用于解释本文所述的原理。

附图说明

[0099] 为了更加全面地了解本文中所公开的原理和其优点,现结合附图来参考以下说明,其中:

[0100] 图1A和1B显示了通过启动子捕获和短同源臂进行的蛋白质标记。图1A 示出了N端标记。无启动子选择标记物嘌呤霉素通过自裂解2A肽与翡翠绿色荧光蛋白 (EmGFP) 报道体或橙色荧光蛋白 (OFP) 报道基因连接,随后通过 PCR在5'和3'末端添加35nt同源臂。内源启动子驱动嘌呤霉素、报道基因和内源基因的表达。TALEN或CRISPR在翻译起始位点附近诱导双链断裂 (DSB)。图1B示出了C端标记。EmGFP或OFP报道基因通过自裂解2A肽与无启动子选择标记物嘌呤霉素连接,随后在5'和3'末端添加35nt同源臂。内源启动子驱动内源基因、报道基因和嘌呤霉素的表达。TALEN或CRISPR 在翻译终止位点附近诱导DSB。终止密码子在內源基因和报道基因之间消除。在图1A和1B中,供体DNA通过同源重组插入基因组。分别使用F1/R1和F2/R2引物组,通过PCR分析5'和3'末端的接合点。

[0101] 图2A至2D显示供体形式和剂量以及同源臂长对HDR效率的影响。在图2A中,通过电穿孔将Cas9 RNP和不同量的具有35nt同源臂的供体DNA 递送到293FT细胞中。缺乏gRNA的样品用作对照。转染后48小时,通过流式细胞仪分析细胞,以测定未发生嘌呤霉素选择(-)的OFP阳性细胞的百分比。或者,在流式细胞术分析(+)之前,用嘌呤霉素处理细胞7天。在图2B 中,通过PCR扩增将各种同源臂长添加到插入盒中,然后与Cas9 RNP共转染到293FT细胞中。如图2A所述,通过流式细胞术分析细胞。在图2C和2D 中,通过电穿孔将Cas9 RNP和具有约500nt同源臂的供体质粒,或具有35nt 同源臂的单链(ss)或双链(ds) DNA供体转染到293FT或人原代T细胞中。转染后48小时,对细胞进行流式细胞术分析。

[0102] 图3A至3E显示了OFP整合到 β -肌动蛋白基因座中的克隆细胞的表征。通过电穿孔将Cas9 RNP和具有35nt同源臂的供体DNA递送入293FT细胞,然后在选择嘌呤霉素后分离克隆细胞。使用一个内部引物和一个外部引物或一对外部引物,通过接合式PCR分析克隆细胞。通过测序分析所得PCR产物。图3A和3B分别示出了具有精确HDR (1) 或具有插入缺失的HDR (2) 的N 端和C端接合点。图3A和3B中的精确HDR (1) 箭头指示基因组DNA和供体DNA或Cas9裂解位点之间的接合点。图3A和3B中以粗体显示的序列指示35nt同源臂。斜体ATG指示 β 肌动蛋白的起始密码子。图3A和3B中的具有插入缺失的HDR (2) 显示围绕接合点形成插入缺失的实例。图3C显示了克隆细胞中的合子性的表征。等位基因1在两个接合点均具有约68%的精确HDR,并且在C或N端或两端出现32%具有插入缺失的HDR。等位基因2 在约80%克隆中具有“A”插入(Δ 1ntA),在18%克隆和2%野生型(wt)中有超过2nt的缺失(Δ >2nt)。

图3D和3E显示了通过T_{al}核酸酶用OFP对 β -肌动蛋白进行N端标记。通过NEON[®]电穿孔(赛默飞世尔科技(Thermo Fisher Scientific), 目录号MPK5000)将单独TALEN mRNA或TALEN mRNA与供体DNA一起转染到HEK293FT细胞中。图3D显示了基因组编辑效率(插入缺失%)并且图3E显示了通过流式细胞术分析OFP阳性细胞(-)百分比和嘌呤霉素处理的OFP阳性细胞(+)百分比。

[0103] 图4A、4B和4C显示了A549细胞中的针对LRRK2的EmGFP的N端标记。通过电穿孔将Cas9 RNP和含有无启动子的嘌呤霉素-P2A-EmGFP片段和约35nt同源臂的供体DNA递送到细胞中。转染后48小时,对细胞进行克隆细胞分离。扩增后,将克隆细胞溶解,并且使用N端(图4A)或C端(图4B)的一个内部引物和一个外部引物、通过接合式PCR进行分析。或者,使用一对外部引物来分析两个等位基因的基因组修饰(图4C)。通过测序分析所得PCR产物。图4A和4B中以粗体表示的序列指示同源臂。底部箭头指示Cas9 裂解位点或基因组DNA与供体DNA之间的接合点。 Δ 图4C中的7nt_无HDR 指示未发生HDR,但具有7nt缺失。

[0104] 在图5A(SEQ ID NO:56-62)、5B(SEQ ID NO:63-69)和5C中,FAK 的C端用EmGFP标记。通过电穿孔将Cas9 RNP和具有短同源臂的供体DNA 转染到293FT中。嘌呤霉素选择后,对细胞进行克隆细胞分离。通过PCR扩增接合点,随后对N端接合点(图5A)或C端接合点(图5B)进行测序分析。箭头指示基因组DNA和供体DNA之间的双链断裂(DSB)或接合点。在精确HDR的情况下,还显示了短同源臂(粗体)和终止密码子(下划线)。具有Indel的HDR的实例也显示在图5A和5B中。图5C显示了两个等位基因上的基因组修饰分析。

[0105] 在图6A、6B和6C中,EGFR是用EmGFP标记的C-末端。设计了一个gRNA,以靶向终止密码子附近的EGFR基因组基因座。通过电穿孔将Cas9 RNP复合物和供体DNA输送到293FT细胞中。通过接合式PCR和测序分析克隆细胞。图6A示出了N端接合分析(SEQ ID NO:70),并且图6B示出了C端接合分析(SEQ ID NO:71)。图6C显示了每个等位基因上的基因组修饰。(c)图6C中的1ntA_noHDR是指一个“A”插入而无插入序列。

[0106] 图7A显示了DNA供体的末端修饰对HDR效率的影响,并且图7B显示了NHEJ抑制剂对HDR效率的影响。在图7A中,化学合成末端经修饰的DNA 引物且用于通过PCR扩增制备供体DNA。通过电穿孔将Cas9 RNP和供体DNA转染到原代T细胞中。转染后48小时,通过流式细胞术监测嘌呤霉素-P2A-OFP DNA片段插入 β -肌动蛋白基因座的效率。在图7B中,电穿孔后,立即将NHEJ抑制剂添加到培养基中。“F”是指正向引物;“R”是指反向引物;“PS”是指硫代磷酸酯;“NH₂”是指胺修饰,并且“ssDNA”是指单链DNA。

[0107] 图8A、8B、8C和8D显示了重组抗体在哺乳动物基因组中的克隆和表达。图8A显示了抗体表达盒,其包含无启动子的嘌呤霉素选择标记物,继之为自裂解2A肽(SEQ ID NO:5)。IgG重链(HC)和轻链(LC)的表达由CMV启动子驱动。通过PCR添加35nt同源臂。图8B(SEQ ID NO:72-76)和图8C(SEQ ID NO:77-82)分别显示了N端和C端接合分析。箭头指示双链断裂(DSB)和基因组DNA与供体DNA之间的接合点。35nt同源臂和一些额外的序列也以粗体突出显示。WP_{RE}(土拨鼠肝炎病毒转录后调控元件)和终止密码子显示在图8C中。图8D显示了通过ELISA分析所测定的产生抗体(+)或不产生抗体(-)的克隆细胞的相对百分比。

[0108] 图9.核定位信号(NLS)-供体DNA设计(SEQ ID NO:83-84)。用于连接NLS肽的结合化学试剂是琥珀酰亚氨基4-(N-马来酰亚胺甲基)环己烷-1-甲酸酯(SMCC)或CLICK-iT[®]。

[0109] 图10(SEQ ID NO:85-92)。经NLS供体DNA构建体修饰的HEK 293 株系。在左侧,

GFP基因因构成荧光团的6个核苷酸的缺失而断裂。含有6个碱基的供体的添加使GFP荧光恢复。右侧是类似的BFP基因断裂。添加具有单核苷酸多态性(SNP)的供体使BFP编码序列转换为GFP编码序列。

[0110] 图11. 相较于经NLS修饰的供体DNA, 硫代磷酸酯(PS)寡核苷酸供体DNA在添加6碱基序列以恢复GFP活性时的剂量反应。相较于**CLICK-iT®**连接的经NLS修饰的供体DNA, PS寡核苷酸供体DNA的剂量反应。

[0111] 图12. 流式细胞术对使用浓度相等的PS或NLS寡核苷酸供体DNA所编辑的细胞进行的分析。

[0112] 图13. 相较于经NLS修饰的供体DNA, PS寡核苷酸供体DNA的剂量反应, 以编辑单个碱基, 从而使表现BFP的细胞转变为表现GFP的细胞。相较于SMCC连接的经NLS修饰的供体DNA, PS寡核苷酸供体DNA的剂量反应。

[0113] 图14. 图显示了TALEN和TAL-Buddy(无核酸酶)构建体的示例性架构的示意图。

[0114] 图15. 当添加相对于TALEN结合序列以7nt间距设计的“TAL-Buddy”时, CMPK1-C目标处的插入缺失形成提高了约2倍。

[0115] 图16. “TAL-Buddy”相对于TALEN结合序列的间距长达100nt时经测试可提高TALEN裂解。

[0116] 图17. “TAL-Buddy”相对于CRISPR sgRNA结合序列的间距为20nt时使CRISPR-RNP的插入缺失形成在UFSP2-SNP目标处提高 ≥ 20 倍。

[0117] 图18. “TAL-Buddy”与sgRNA和SpCas9-HF1或eSpCas9所形成的RNP一起提高插入缺失形成。

[0118] 图19. 用于制备sgRNA和“CR-PAL”gRNA的模板的图示。

[0119] 图20. “CR-PAL”功能的图示。黑色表示具有15nt结合能力的“CR-PAL”; 灰色表示具有20nt结合能力的sgRNA。

[0120] 图21. 当“CR-PAL”与Cas9-RNP一起用于UFSP2-SNP目标时, 插入缺失的形成增加了超过60倍。

[0121] 图22. 图显示了“TAL-Buddy”的No-FokI C端片段的制备。

[0122] 图23. 图显示了测试“Buddy TAL”(293FT)。目标: CMPK1-C (SEQ ID NO:19); TALEN mRNA: 各100ng; TAL-Buddy: 7nt间距 (SEQ ID NO:18); **NEON®**: 1300/20/2。重复。

[0123] 图24. 图显示了Buddy TAL在TALEN上的测试间距。目标: CMPK1; 细胞: 293FT; **NEON®**: 1300/20/2。间距很重要, TAL不能直接紧邻TALEN。间距(0、4、7、20nt)表示TAL的18个碱基识别序列和最近的TALEN对之间的间距。

[0124] 图25. 图显示了针对提高TALEN和CRISPR效率来测试Buddy TAL。间距会影响切割效率。TAL(灰色六边形)与TALEN(深灰色箭头)相距7nt与相距20nt无差异。使用TAL和CRISPR目标(黑色圆片段), 与TAL相距20nt更好。

[0125] 图26. 图显示了用CRISPR编辑的重复TAL-buddy。293FT细胞; CRISPR 目标: USFP2; TAL-Buddy: 20nt间距; **NEON®**: 1150/20/2。重复。

[0126] 图27. 图显示针对高保真度Cas9测试TAL-Buddy-恢复低性能突变体的活性。目标: UFSP2; 细胞: 293FT; **NEON®**: 1150/20/2; TAL-Buddy: 20nt间距。HF-cas9的活性检测不到,

我们的分析表明,它的不完美超过eCas9。在没有TAL的情况下,eCas9 1.1的活性检测不到。有了TAL,可以获得wt活性水平。这很重要,因为您获得的高保真度活性仅定位在所需的目標位点(超高保真度)。

[0127] 图28.图显示了使用标准活性cas9和截短的gRNA测试CRISPR-PAL。目标:UFSP2;细胞:293FT; **NEON®**:1150/20/2;CR-PAL:15mer gRNA; CR_PAL-左间距:36nt;CR_PAL-右间距:15nt。Cas9将结合而非切割被截短的gRNA(15mer)。(Church等人,2014Kiani等人,用于基因组编辑、激活和抑制的Cas9 gRNA工程化(Cas9 gRNA engineering for genome editing, activation and repression),《自然·方法(Nat.Methods)》doi:10.1038(2015年9月7日))。使用截短的gRNA括住切割位点并打开DNA,以便标准gRNA(20mer)可以更好地切割。5%(单独)到>50%(15mer)。Cas9 v2+20mer gRNA +L/R 15mer gRNA。

[0128] 图29.图显示了Buddy TAL激活剂概念。TAL与激活结构域(例如VP64)的结合可促进活性基因表达,从而打开DNA以增强核酸酶(TALEN、Cas9等)的编辑。

[0129] 图30.此图显示了U20S中的HDR(序列验证)。供体具有HindIII位点插入;**NEON®**:1300/20/2。

[0130] 图31.此图显示了小分子/添加剂对A549细胞中的TALEN编辑的影响。目标:HTR2A-N;供体具有HindIII位点插入;**NEON®**条件:1200/20/4;培养基24小时更换;HindIII切割如图所示。NU7441(DNAPK抑制剂)和B18R(免疫反应阻遏物)。

[0131] 图32.此图显示了TALEN和TAL-Buddy相对位置的实例。然后,TALEN对在目标位点的每一侧有8个碱基的间隔。在此实例中,TAL-buddy与TALEN的间距为7nt。上链是SEQ ID NO:20,且下链是SEQ ID NO:21。

[0132] 图33.此图显示接近于UFSP2-SNP目标中的CRISPR裂解位点所设计的“TAL-Buddy”。添加100ng Lt和Rt“TAL-Buddy”mRNA以及CRISPR-RNP(1000ng的Cas9蛋白和200ng sgRNA),以便使用**NEON®**电穿孔设备(赛默飞世尔科技(Thermo Fisher Scientific),目录号MPK5000)、以1150脉冲电压、20脉冲宽度和2次脉冲数转染到约50,000个293人类胚肾细胞(293FT)中。转染后48至72小时收获细胞并且溶解。使用GENEART™基因组裂解检测试剂盒(赛默飞世尔科技,目录号A24372)分析插入缺失的形成。上链是SEQ ID NO:42,且下链是SEQ ID NO:43。

[0133] 图34.此图显示邻近于UFSP2-SNP目标中的CRISPR裂解位点所设计的“CR-PAL”。将200ng CR-PAL_Lt和CR-PAL_Rt与野生型Cas9-RNP一起培育且使用**NEON®**电穿孔设备(赛默飞世尔科技,目录号MPK5000)、以1150脉冲电压、20脉冲宽度和2次脉冲数转染到约50,000个293人类胚肾细胞(293FT)中。转染后48至72小时收获细胞并且溶解。使用“GENEART™基因组裂解检测试剂盒”(赛默飞世尔科技,目录号A24372)分析插入缺失形成。上链是SEQ ID NO:44,并且下链是SEQ ID NO:45。

[0134] 图35.此图显示了测试“Buddy TAL”(293FT)。上链是SEQ ID NO:46,并且下链是SEQ ID NO:47。

[0135] 图36是一对TAL-Buddy(在本文中也称为第一与第二DNA结合调节增强剂)结合一对TAL-FokI核酸酶融合体(在本文中也称为第一与第二DNA结合核酸酶结合物)使用的示意图。图右侧和左侧均有指示,左侧为左TAL-Buddy结合且右侧为右TAL-Buddy结合。长实线

表示细胞内核酸分子(例如染色体;在本文中也称为目标基因座)的一部分。所示核酸分子的区域A(左端和右端所示)是两个TAL-Buddy蛋白的结合位点(在本文中也称为第一和第二增强子结合序列)。区域B表示TAL-Buddy结合位点(例如第一和第二增强子结合序列)与TAL-FokI融合蛋白结合位点(在本文中也称为第一和第二结合序列)之间的距离。区域D表示两个TAL-FokI融合蛋白结合位点之间的核酸区段。区域D中的白框表示TAL-FokI融合蛋白对使核酸裂解的位点(在本文中也称为调节位点)。区域E表示核酸分子的可及性潜在得到增强的一部分。

[0136] 图37是类似于图36的示意图,例外为使用单一TAL-VP16融合体(在本文中也称为调节蛋白)替代一对TAL-FokI核酸酶融合体。未标记的圆圈表示VP16募集的转录复合物的组分。此外,由于使用了单一TAL-VP16融合体,因此存在唯一一个区域C。而且,所形成的区域B是由区域A(本文中也称为第一和第二增强子结合序列)和区域C(本文中也称为调节结合序列)之间的中间碱基对形成。

[0137] 图38显示了可用于本文阐述的各种实施例中的供体核酸分子的许多不同形式。末端的空心圆代表核酸酶抗性基团。两个圆意味着存在两个基团。黑色区域表示与另一核酸分子(例如染色体DNA)的一个或多个基因座序列同源/互补的区域。交叉阴影区域表示位于核酸片段中的序列同源/互补区域之间的核酸。此图显示了可用于不同方面中的供体核酸分子的不同变体。

[0138] 图39是基于Cas9蛋白模型酿脓链球菌Cas9的示例性Cas9形式的示意图。此1368个氨基酸蛋白由图中的实心顶线表示。标示为V1-V5的Cas9蛋白是具有核定位信号(NLS)作为组分的融合蛋白。虚线框表示单分型NLS,并且空心框表示双分型NLS。灰色框表示亲和标签(例如六组氨酸标签)。

[0139] 图40是类似于图36的示意图,其增添了TAL裂解基因座的更详细视图与供体DNA分子的示意图。左下方显示了供体DNA的线性示意图。直实线表示与目标基因座同源的区域。圆形虚线表示插入盒。“X”符号表示序列同源区域。向上和向下的虚线箭头表示供体DNA同源臂5'和3'链中的两个硫代磷酸酯键。在核酸酶消化后,这些硫代磷酸酯键经定位以便产生长度为十个核苷酸的5'突出端。左侧和右侧的空心框表示双分型NLS。线性示意图的右侧是插入盒的两个实例。上部插入盒被设计成既破坏插入位点处的功能、又表达嘌呤霉素抗性标记物。下部插入盒类似于上部插入盒,但也设计为将与组织特异性启动子可操作地连接的所关注基因插入基因座。

[0140] 图41显示了Cas9 V1的氨基酸序列(SEQ ID NO:93)。NLS和His标签如此标记。

[0141] 图42显示了Cas9 V2的氨基酸序列(SEQ ID NO:94)。NLS如此标记。

[0142] 图43显示了一系列Cas9-NLS融合蛋白的形式。“NP”是指核质蛋白NLS。

[0143] 图44A和44B显示了使用不同Cas9-NLS与两种不同细胞类型的组合所得的GCD资料。

[0144] 图45是显示共同TALE结构形式的示意图。位点1、2和3位于被认为与DNA识别和结合有关的TALE区域之外。

[0145] 图46显示TALEN蛋白的氨基酸序列(SEQ ID NO:95)。此TALEN的形式在本文中称为“TALEN V3”。N端区域包含V5抗原决定基和“G-G”接头,继之为136个氨基酸区域,随后为重复区域。136个氨基酸区域含有(1)与重复区域的个别重复序列具有一定序列同源性的一个系

列重复单元(标记为“R-3”、“R-2”、“R-1”和“R0”)和(2)“少T框”,氨基酸序列“RGA”,其可以改变以减少TALEN所结合的核酸的5'T需求量。重复区域含有三十四个氨基酸的十六个重复序列。半重复序列(标记为“R1/2”)紧邻重复区域的C末端。两个核定位信号(标记为“NLS”)的定位进一步朝向FokI核酸酶结构域之前和之后的蛋白质的C末端。

[0146] 图47显示TALEN蛋白的氨基酸序列(SEQ ID NO:96-97),但是重复区域的氨基酸序列已被去除以简化附图。另外,此图中所示的蛋白质具有三个 NLS。

[0147] 图48显示了三种不同细胞类型中的三种不同基因组基因座的基因组裂解检测数据,所述数据如以下实例8中所述产生。

[0148] 图49显示了两种不同细胞类型中的三个不同基因组基因座的基因组裂解检测 and 同源定向修复数据,所述数据如以下实例8中所述产生。

[0149] 图50显示了A549细胞中的三个不同基因组基因座的基因组裂解检测数据,所述数据如以下实例8所述产生。

[0150] 图51A显示了在用Cas9蛋白和靶向 β -肌动蛋白基因座的gRNA(β -肌动蛋白gRNA)以及含有OFP基因的供体DNA(OFP-Puro)进行电穿孔之后,对OFP一步敲入人原代T细胞进行的代表性流式细胞术分析。

[0151] 图51B显示了在用Cas9蛋白质和 β -肌动蛋白gRNA和OFP-Puro进行电穿孔之后,OFP一步敲入人原代T细胞的效率。

[0152] 图52A显示了在用Cas9蛋白和 β -肌动蛋白gRNA和OFP-Puro进行电穿孔之后,在与NHEJ抑制剂(Nu7026、Nu7441或Ku0060648)一起培育之后,对OFP一步敲入人原代T细胞进行的代表性流式细胞术分析。

[0153] 图52B显示了在用Cas9蛋白和 β -肌动蛋白gRNA和OFP Puro进行电穿孔并且与指定浓度的NHEJ抑制剂(Nu7026、Nu7441或Ku0060648)一起培育之后,OFP一步敲入人原代T细胞的效率。

[0154] 图52C显示了用指定抑制剂处理后,在电穿孔后回收的细胞百分比。

[0155] 图52D显示了在用Cas9蛋白和 β -肌动蛋白gRNA和OFP-Puro进行电穿孔且与Ku0060648一起培育之后,使用四种不同T细胞供体所得的敲入(KI)效率变化倍数。

[0156] 图53A显示了在用Cas9蛋白和 β -肌动蛋白gRNA和OFP-Puro进行电穿孔之后,在电穿孔之后在培养基中(Ku存在于培养基中)或在电穿孔期间在电穿孔缓冲液中(Ku存在于电穿孔缓冲液中)与Ku0060648一起培育之后,对OFP一步敲入人原代T细胞进行的代表性流式细胞术分析。

[0157] 图53B显示了在用Cas9蛋白和 β -肌动蛋白gRNA和OFP-Puro进行电穿孔之后,OFP一步敲入人原代T细胞的敲入效率,其中Ku0060648是在电穿孔期间以指定浓度添加至电穿孔缓冲液中。

[0158] 图53C显示了在含有指定浓度的Ku0060648的电穿孔缓冲液中进行电穿孔后所回收的原代T细胞的百分比。

[0159] 图54显示了在用Cas9蛋白和 β -肌动蛋白gRNA、Cas9蛋白和靶向T细胞受体的向导RNA(TRAC gRNA)以及OFP-Puro进行电穿孔之后,在电穿孔之后与或不与Ku0060648一起在培养基中培育之后,对OFP一步敲入人原代T细胞进行的代表性流式细胞术分析。

[0160] 图55A显示了在用Cas9蛋白、 β -肌动蛋白gRNA和OFP-Puro进行电穿孔且在电穿孔

之后与Ku0060648一起在培养基中培育之后,0FP阳性NK-92 细胞的百分比。

[0161] 图55B显示了在用Cas9蛋白、 β -肌动蛋白gRNA和0FP-嘌呤霉素进行电穿孔且在电穿孔之后与Ku0060648一起在培养基中培育之后,0FP阳性 THP-1细胞的百分比。

[0162] 图55C显示了在用Cas9蛋白、 β -肌动蛋白gRNA和0FP-Puro进行电穿孔且在电穿孔之后与Ku0060648一起在培养基中培育之后,0FP阳性杰卡特细胞(Jurkat cells)的百分比。

[0163] 图56A显示了在用Cas9蛋白、相关gRNA和供体DNA进行电穿孔且在电穿孔之后与2 μ M Nu0441或250nM Ku0060648一起在培养基中培育之后, iPSC中的各种指定目标基因的校正SNP百分比。

[0164] 图56B是图56A中的数据的图形表示。

[0165] 图57A示出了与图57B和57C的图形中的编号条柱对应的NHEJ抑制剂的组合和量。

[0166] 图57B和57C显示了在用Cas9蛋白和 β -肌动蛋白gRNA和0FP-Puro进行电穿孔之后, 0FP一步敲入人原代T细胞的敲入效率,其中NHEJ抑制剂 (如图57A所示)是在电穿孔之后添加至培养基中。

具体实施方式

[0167] 概述

[0168] 此处阐述的组合物和方法旨在改善基因编辑。作为实例,这些改善包括以下:

[0169] i.将核酸分子(例如供体DNA分子)插入细胞内核酸分子中,其中插入的核酸分子可操作地连接至存在于细胞内核酸分子中的启动子。

[0170] ii.使用非同源末端连接抑制剂促进基因编辑。

[0171] iii.使用在细胞内目标基因座处或附近结合的DNA结合分子(例如DNA 结合蛋白、DNA结合蛋白/核酸复合物),其中DNA结合蛋白促进其它DNA 结合分子对目标基因座的增强的可及性。

[0172] iv.供体DNA在细胞内递送到基因编辑基因座,以及其它DNA分子递送到细胞内的不同位置(例如含有开放阅读框的线性DNA分子可操作地连接至启动子并递送至线粒体)。

[0173] 以上改善可以个别地使用,或结合上面列出的其它方法以及另外的方法使用。

[0174] 本公开部分地涉及以下发现:针对选择标记物的启动子捕获与针对重组的短同源臂的组合使用实现了近100%的整合效率和高达100%的精确HDR。

[0175] 与使用具有0.5kb至2kb同源臂的靶向载体的传统方法不同,使用短同源臂似乎使所关注的外来DNA随机整合到基因组中的发生率最小化。最重要的是,针对选择标记物的启动子捕获的使用能够选择正确整合的物种,因为无启动子选择标记物仅在DNA分子精确插入基因组基因座时才表达。在一些实施例中,用例如硫代磷酸酯或氨基对供体DNA进行的末端修饰和/或用NHEJ 抑制剂处理进一步改善了HDR的效率。供体DNA整合的精确度具有序列依赖性。在一些基因座处,能够实现100%整合效率和100%精确HDR。

[0176] 本公开还部分地涉及用于增强与细胞内核酸区域中的细胞内核酸发生相互作用的分子或分子复合物对这些区域的可及性的组合物和方法。

[0177] 本公开另外部分地涉及用于核酸分子的细胞内定位的组合物和方法。在某些情况下,核酸分子是供体DNA分子。

[0178] 本公开还涉及上述者的各种组合,以促进诸如基因编辑、基因激活、基因抑制、DNA 甲基化等过程。

[0179] 在各方面中,包括用于增强基因编辑的组合物和方法。许多变量影响基因编辑的效率。关于同源定向修复 (HDR), 这些因素包括:

[0180] (1) (i) 供体DNA和 (ii) 位于细胞核中的位点特异性核酸酶的量,以及细胞核中的位点特异性核酸酶活性的量,

[0181] (2) 位点特异性核酸酶对目标基因座的可及程度,

[0182] (3) 与供体和核酸酶在细胞核中的存在有关的时序方面,

[0183] (4) 目标基因座裂解效率,

[0184] (5) HDR效率 (包括HDR:NHEJ比率), 以及

[0185] (6) 供体DNA结构和组成。

[0186] 预期在某些情况下,尤其是在HDR方面,可以达到接近100%的基因编辑效率。

[0187] 基因编辑试剂在细胞核中的定位: 由于人们认为影响基因编辑效率的许多因素都是基于浓度依赖性机制, 因此位点特异性核酸酶活性的量 (核酸酶活性水平与存在的核酸酶数量的组合) 越高且细胞核中的供体DNA浓度越高, 则预期越多的HDR在NHEJ中占主导地位。

[0188] 虽然细胞 (例如基于载体的系统) 中可以产生核酸分子和蛋白质, 但在许多情况下, 将基因编辑系统的组分 (例如供体DNA、位点特异性核酸酶、DNA 结合调节增强剂等) 引入细胞中。此类细胞引入可以利用例如转染和电穿孔等方法来完成。

[0189] 将基因编辑系统组分引入细胞后, 通常需要有效定位于细胞核。这样做是因为人们认为基因编辑系统组分对核酸酶的有效定位至少部分与细胞质降解 ((i) 降解活性和 (ii) 细胞质中所耗的时间量的组合) 有关。此外, 许多因素会影响核定位效率, 包括 (1) 基因编辑系统组分与一个或多个NLS的结合, (2) 所用NLS的选择, 以及 (3) 一种或多种基因编辑系统组分 (例如供体 DNA) 的化学修饰。

[0190] 在许多情况下, 本文所述方法中使用的核酸分子可以被化学修饰。化学修饰包括核酸酶抗性基团, 例如硫代磷酸酯基、氨基、2'-O-甲基核苷酸、2'-脱氧-2'-氟核苷酸、2'-脱氧核苷酸、5-C-甲基核苷酸和其组合。举例来说, 5' 和 3' 端 gRNA 分子上的三个核苷酸可以含有硫代磷酸酯键且/或可以是2'-O-甲基核苷酸。还已经发现供体DNA的胺末端修饰增强了HDR (参见例如图7A和7B)。在两种情况下都认为这至少部分是由于供体DNA分子在细胞质中的稳定所致。gRNA也被认为可以通过与Cas9蛋白结合而稳定。因此认为当gRNA与 Cas9 蛋白结合时, gRNA的胞质半衰期增加。

[0191] 数据表明并且据信, 当基因编辑系统组分就细胞质降解而言稳定并且通过细胞质迅速“穿梭”到细胞核时, 基因编辑效率提高。基因编辑系统组分快速移动通过细胞质具有另一种作用, 此作用在许多情况下将是有益的。这使得基因编辑系统组分活性在细胞核中瞬时达到高浓度, 并且获得较低的细胞质基因编辑系统组分池。因此, 一旦细胞核中高浓度的基因编辑系统组分活性耗竭, 则用于额外基因编辑活性的细胞质储备极少或没有。

[0192] 位点特异性目标基因座裂解活性: 目标基因座裂解效率由许多因素决定, 其中一些因素如上所述。这些因素包括: (1) 基因编辑系统裂解活性, (2) 目标基因座处或附近存在的介导裂解的基因编辑系统组分的量, 和 (3) 目标基因座对基因编辑系统裂解活性的可及

性,如本文所述。

[0193] 基因编辑系统裂解活性对目标基因座的可及性可随自然作用而变化,因为它对于基因组或特定细胞类型中或其间的某处来说是可及或不可及的。在那个目标基因座裂解之前诱导目标基因座转录激活可以使裂解活性更容易到达基因座。增加特定目标基因座可及性的另一种方法是通过使用DNA结合调节增强剂。

[0194] 关于位点特异性目标基因座裂解活性的一个考虑是“脱靶”效应。脱靶效应可以通过DNA结合调节增强剂、gRNA高目标基因座特异性与高保真度基因编辑试剂(例如高保真度Cas9)的单独或组合使用来最小化。

[0195] 目标基因座改变:通常执行的基因编辑的主要类型存在两种。这些基因编辑是将核酸分子插入目标基因座中以及不将核酸分子插入目标基因座中,而是改变目标基因座的核苷酸序列。此外,裂解和“修复”目标基因座时有三种可能性。目标基因座可以是(1)与裂解前的核苷酸序列相比为不变的;(2)通过缺失或添加一个或多个碱基而经修饰,而没有供体核酸插入,或(3)供体DNA插入可以在裂解位点或附近引入。这些可能性中的前两种通常是由基于NHEJ的修复机制引起的。这些可能性的第三种通常是依据基于HDR的机制。在许多情况下,特别是在需要将供体DNA插入目标基因座的情况下,第三种可能性是优选的。因此,本文提供了用于增强HDR效率和/或相对于NHEJ更有利于HDR的组合物和方法。

[0196] 已经发现许多因素引起有效的HDR并且将供体核酸分子插入裂解位点。这些因素中的一些与供体核酸分子的特征有关。这些因素之一是供体DNA同源臂的长度。在许多情况下,供体DNA分子将具有两个同源臂,其长度范围独立地为约20至约2,000个核苷酸或碱基对,这取决于供体DNA是单链还是双链。此外,双链供体DNA可在一端或两端具有3'突出物,并且这些突出物(以及5'突出物)的长度范围可为约10至约40个核苷酸。同样,供体DNA分子的一个或两个同源臂的一条或两条链可以包含位于末端或臂内其它位置的一个或多个核酸酶抗性基团(如本文其它地方所讨论)。

[0197] 有许多方法可以使HDR优于NHEJ修复。一种方法是用一种或多种NHEJ抑制剂处理即将进行基因编辑的细胞(参见图7B)。另一种是“降低”细胞内NHEJ活性。这可以通过使用例如设计成抑制一种或多种NHEJ修复途径表达的反义microRNA和/或RNAi试剂(例如DNA依赖性蛋白激酶、催化亚单元;Ku70和/或Ku80)来实现。

[0198] 定义

[0199] 如根据本公开所用,除非另外指示,否则以下术语应理解为具有以下含义:

[0200] “核酸”是指呈单链、双链或多链形式的脱氧核糖核苷酸或核糖核苷酸和其聚合物,或其补体。术语“多核苷酸”是指核苷酸的线性序列。术语“核苷酸”通常是指多核苷酸的单个单元,即单体。核苷酸可以是核糖核苷酸、脱氧核糖核苷酸或其修饰形式。本文所涵盖的多核苷酸的实例包括单链和双链DNA、单链和双链RNA(包括siRNA)以及具有单链和双链DNA和RNA混合物的杂交分子。核酸可以是线性或分支的。例如,核酸可以是核苷酸的直链,或者核酸可以是分支的,例如,使得核酸包含一个或多个核苷酸臂或分支。任选地,分支型核酸被重复地分支以形成更高级的结构,例如树枝状聚合物等。

[0201] 所述术语还包括含有已知核苷酸类似物或经修饰的骨架残基或键联的核酸,所述核酸是合成的、天然存在的和非天然存在的,具有类似于参考核酸的结合特性,并且代谢的方式类似于参考核苷酸。此类类似物的实例包括但不限于磷酸二酯衍生物,包括例如氨基

磷酸酯、二氨基磷酸酯、硫代磷酸酯 (phosphorothioate) (也称为硫代磷酸酯 (phosphothioate))、二硫代磷酸酯、膦酰基羧酸、膦酰基羧酸酯、膦酰基乙酸、膦酰基甲酸、甲基膦酸酯、硼膦酸酯或O-甲基氨基磷酸酯键联(参见Eckstein,《寡核苷酸和类似物:一种实用方法(Oligonucleotides and Analogues:A Practical Approach)》,牛津大学出版社(Oxford University Press));以及肽核酸骨架和键联。其它类似物核酸包括具有阳性骨架的那些;非离子骨架、经修饰的糖和非核糖骨架(例如二氨基磷酸酯吗啉代寡核苷酸或锁核酸(LNA)),包括以下文献中所述的那些:美国专利第 5,235,033号和第5,034,506号以及《ASC研讨会丛书580(ASC Symposium Series 580)》第6章和第7章《反义研究中的碳水化合物修饰(Carbohydrate Modifications in Antisense Research)》,Sanghui和Cook编。含有一个或多个碳环糖的核酸也包括在核酸的一个定义中。核糖-磷酸酯骨架的修饰可以出于多种原因,例如增加此类分子在生理环境中的稳定性和半衰期或作为生物芯片上的探针。可以制备天然存在的核酸和类似物的混合物;或者,可以制备不同核酸类似物的混合物,以及天然存在的核酸和类似物的混合物。在实施例1中,DNA中的核苷酸间键联是磷酸二酯、磷酸二酯衍生物或两者的组合。

[0202] 核酸可以包括非特异性序列。如本文所用,术语“非特异性序列”是指含有未被设计为与任何其它核酸序列互补或仅部分互补的一系列残基的核酸序列。举例来说,非特异性核酸序列是当与细胞或生物体接触时不起抑制性核酸作用的核酸残基序列。“抑制性核酸”是能够与目标核酸(例如可翻译成蛋白质的mRNA)结合并减少目标核酸(例如来自DNA的mRNA)的转录或减少目标核酸(例如mRNA)的翻译或改变转录物剪接(例如单链吗啉代寡核苷酸)的核酸(例如DNA、RNA、核苷酸类似物的聚合物)。

[0203] 如本文所用,术语“核酸分子”是指共价连接的核苷酸或碱基序列(例如,RNA的核糖核苷酸和DNA的脱氧核糖核苷酸,而且包括DNA/RNA杂交体,其中DNA处于单独的链中或处于相同链中),其中一个核苷酸的戊糖的3'位置是通过磷酸二酯键联与下一个核苷酸的戊糖的5'位置接合。核酸分子可以是单链或双链或部分双链的。核酸分子可在超螺旋或松弛形成时以具有钝端或粘性末端的线性或环化形式出现并且可含有“切口”。核酸分子可由完全互补的单链构成或由形成至少一个碱基错配的部分互补的单链构成。核酸分子可另外包含两个自身互补序列,其可形成双链茎区,任选地在一端被环序列隔开。包含双链茎区的核酸分子的两个区域基本上彼此互补,从而产生自身杂交。然而,茎可包括一个或多个错配、插入或缺失。如上所述,核酸分子可以包括核酸分子的化学、酶促或代谢修饰形式或其组合。化学合成的核酸分子可以指通常小于或等于150个核苷酸长(例如5至150、10至100、15至50个核苷酸长)的核酸,而酶促合成的核酸分子可涵盖较小以及较大的核酸分子,如本申请别处所述。核酸分子的酶促合成可包括使用酶(如聚合酶、连接酶、外切核酸酶、内切核酸酶等或其组合)的逐步方法。如本文所提供的术语“基因组编辑”或“基因编辑”是指涉及酶的逐步方法,所述酶例如聚合酶、连接酶、外切核酸酶、内切核酸酶等或其组合。例如,基因编辑可以包括以下方法:使核酸分子裂解,切除裂解位点或其邻近处的核苷酸,新合成新的核苷酸,并且使裂解的链接合。

[0204] 术语核酸分子还指短核酸分子,通常称为例如“引物”或“探针”。引物通常被称为用于酶促组装反应的单链起始核酸分子,而探针通常可用于检测至少部分互补的核酸分子。核酸分子具有“5'端”和“3'端”,因为核酸分子磷酸二酯键联存在于取代基单核苷酸的

戊糖环的5'碳与3'碳之间。新键联连接到5'碳的核酸分子的末端是其5'端核苷酸。新键联连接到3'碳的核酸分子的末端是其3'端核苷酸。如本文所用,末端核苷酸或碱基是处于3'端或5'端的末端位置的核苷酸。核酸分子序列即使处于较大核酸分子内部(例如,核酸分子内的序列区),也可称之为具有5'末端和3'末端。

[0205] 如本文所用,“载体”是可以用作将遗传物质转移到细胞中的媒介物的核酸分子。载体可以是质粒、病毒或噬菌体、黏粒或人工染色体,例如酵母人工染色体(YAC)、细菌人工染色体(BAC)或能够在体外或在宿主细胞内复制或被复制的其它序列,或将所需的核酸区段输送到宿主细胞内的所需位置。在实施例中,载体是指具有至少一个复制起点、多克隆位点(MCS)和一个或多个选择标记物的DNA分子。载体通常由骨架区域和至少一个插入或转基因区域或设计用于插入DNA片段或转基因的区域(例如MCS)构成。骨架区域通常含有用于在至少一个主体中繁殖的复制起点和一种或多种选择标记物。载体可以具有一个或多个限制性内切核酸酶识别位点(例如二、三、四、五、七、十个等),序列可以在所述识别位点以可测定的方式被切而不会损失载体的基本生物学功能,且可以在所述识别位点中将核酸片段剪接以便实现其复制和克隆。载体可以进一步提供引物位点(例如用于PCR)、转录和/或翻译起始和/或调控位点、重组信号、复制子、可选标记物等。显然,还可以应用插入所需核酸片段的方法而不需要使用重组、转位或限制酶(如(但不限于)尿嘧啶N糖基化酶(UDG)克隆PCR片段(美国专利第5,334,575号和第5,888,795号,两个专利均以引入的方式完全并入本文中)、T:A克隆等)将片段克隆到待根据本发明使用的克隆载体中。在实施例中,载体包含另外的特征。此类附加特征可以包括天然或合成的启动子、遗传标记物、抗生素抗性盒或选择标记物(例如毒素,例如ccdB或tse2)、抗原决定基或用于检测、操作或纯化的标签(例如V5抗原决定基、c-myc、血凝素)(HA)、FLAGTM、聚组氨酸(His)、谷胱甘肽-S-转移酶(GST)、麦芽糖结合蛋白(MBP))、支架附着区(SAR)或报道基因(例如绿色荧光蛋白(GFP)、红色荧光蛋白(RFP)、萤光素酶、 β -半乳糖苷酶等)。在实施例中,使用载体在目标宿主中分离、增殖或表达所插入的DNA片段。载体可以是例如克隆载体、表达载体、功能载体、捕捉载体、共表达载体(用于表达超过一个开放阅读框)、病毒载体或游离基因体(即能够进行染色体外复制的核酸)等。

[0206] 如本文所用,“克隆载体”包括能够用于缺失、插入、置换或组装一个或多个核酸分子的任何载体。在实施例中,克隆载体可以包含能够被另一种转基因或DNA片段去除或置换的反选择标记基因(例如ccdB或tse2)。在实施例中,克隆载体可以称为供体载体、进入载体、穿梭载体、目的地载体、目标载体、功能载体或捕捉载体。克隆载体通常包含一系列独特的限制酶裂解位点(例如II型或IIS型),用于去除、插入或置换DNA片段。或者,DNA片段可以通过TOPO[®]克隆或重组来置换或插入,例如Invitrogen/Life Technologies(Carlsbad,CA)提供的GATEWAY[®]克隆系统中采用的克隆或重组,并在本文其它地方更详细地描述。可以用于在目标宿主中表达转基因的克隆载体也可以称为表达载体。在实施例中,对克隆载体进行工程改造以获得TAL效应物结合物。

[0207] “表达载体”是为了表达转基因而设计,并且通常包含至少一个驱动转基因表达的启动子序列。如本文所用,表达是指转基因的转录或开放阅读框的转录和翻译,并且可以在无细胞环境(例如无细胞表达系统)中或在宿主细胞中发生。在实施例中,开放阅读框或基因的表达引起多肽或蛋白质的产生。表达载体通常设计成包含一个或多个调控序列,例如

控制所插入的转基因表达的增强子、启动子和终止子区域。合适的表达载体包括但不限于质粒和病毒载体。用于各种应用的载体和表达系统可获自商业供应商,例如Novagen (Madison, WI)、Clontech (Palo Alto, CA)、Stratagene (La Jolla, CA) 和Life Technologies Corp. (Carlsbad, CA)。在实施例中,对表达载体进行工程改造以表达TAL效应物融合体。

[0208] “病毒载体”通常涉及包含经修饰的病毒核酸序列的基因工程化非感染性病毒。在实施例中,病毒载体包含至少一种病毒启动子,并被设计用于插入一个或多个转基因或DNA片段。在实施例中,病毒载体与提供包装或其它功能的辅助病毒一起递送至目标宿主。在实施例中,病毒载体用于稳定地将转基因整合到宿主细胞的基因组中。病毒载体可以用于转基因的递送和/或表达。

[0209] 病毒载体可来源于噬菌体、杆状病毒、烟草花叶病毒、牛痘病毒、逆转录病毒(禽类白血肉瘤、哺乳动物C型、B型病毒、D型病毒、HTLV-BLV组、慢病毒、泡沫病毒)、腺病毒、细小病毒(如腺相关病毒)、冠状病毒、负链RNA病毒(如正粘病毒(如流感病毒)或仙台病毒)、弹状病毒(如狂犬病和水疱性口炎病毒)、副粘病毒(如麻疹和仙台病毒)、正链RNA病毒(例如小核糖核酸病毒和甲型病毒(例如塞姆利基森林病毒(Semliki Forest virus)))以及双链DNA病毒,包括腺病毒、疱疹病毒(例如1型和2型单纯疱疹病毒、爱泼斯坦-巴尔病毒(Epstein-Barr virus)、巨细胞病毒)和痘病毒(例如牛痘、鸟痘和金丝雀痘)。其它病毒包括(但不限于)诺瓦克病毒(Norwalk virus)、披膜病毒(togavirus)、黄病毒、呼肠孤病毒、乳多泡病毒、嗜肝DNA病毒和肝炎病毒。例如,用于递送基因的常见病毒载体是慢病毒载体,这是基于其相对大的包装能力、降低的免疫原性及其高效稳定地转导多种不同细胞类型的能力。此类慢病毒载体可以是“整合的”(即,能够整合到目标细胞的基因组中)或“非整合的”(即,未整合到目标细胞的基因组中)。含有来自真核病毒的调控元件的表达载体经常用于真核表达载体,例如SV40载体、乳头瘤病毒载体和源自爱泼斯坦-巴尔病毒的载体。其它示例性的真核载体包括pMSG、pAV009/A+、pMT010/A+、pMAMneo-5、杆状病毒pDSVE,以及允许蛋白质在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显示在真核细胞中有效表达的其它启动子的引导下表达的任何其它载体。

[0210] “经标记的核酸或寡核苷酸”是以共价方式、经由连接子或化学键、或以非共价方式、经由离子性、范德华力、静电或氢键与标记结合的核酸或寡核苷酸,以便可以通过检测与核酸结合的可检测标记的存在来检测核酸的存在。或者,使用高亲和力相互作用的方法可以获得相同的结果,其中一对结合搭配物中的一个结合另一个,例如生物素、链霉亲和素。在实施例中,硫代磷酸酯核酸或硫代磷酸酯聚合物骨架包括如本文所公开且本领域通常已知的可检测标记。

[0211] 如本文所用,术语“探针”或“引物”被定义为与样品发生的特异性杂交能够被检测到的一个或多个核酸片段。探针或引物可以是任何长度,此取决于它将用于的特定技术。例如,PCR引物的长度通常在10与40个核苷酸之间,而用于例如DNA印迹法的核酸探针的长度可以超过一百个核苷酸。探针可以未标记或如下所述标记,以便其对目标或样品的结合能够被检测到。探针可以由来自染色体的一个或多个特定(预选)部分的核酸来源产生,例如一个或多个克隆、分离的全染色体或染色体片段,或聚合酶链式反应(PCR)扩增产物的集

合。固定在目标元件上的核酸的长度和复杂性对于此方面来说并非关键。技术人员可以调整这些因素以提供给定杂交程序的最佳杂交和信号产生,并在不同基因或基因组位置之间提供所需的分辨率。

[0212] 探针也可以是固体表面(例如硝化纤维素、玻璃、石英、熔融石英载玻片)上所固定的经分离的核酸,如在阵列中。在一些实施例中,探针可以是例如WO 96/17958中所述的核酸阵列的成员。能够产生高密度阵列的技术也可用于此目的(参见例如Fodor《科学(Science)》251:767-773(1991);Johnston,《当代生物学(Curr.Biol.)》8:R171-R174(1998);Schummer,《生物技术(Biotechniques)》23:1087-1092(1997);Kern,《生物技术》23:120-124(1997);美国专利第5,143,854号)。

[0213] 词语“互补”或“互补性”是指多核苷酸中的核酸与第二多核苷酸中的另一核酸形成碱基对的能力。例如,序列A-G-T与序列T-C-A互补。互补性可以是部分的,其中根据碱基配对,只有一些核酸匹配,或者是完全的,其中根据碱基配对,所有核酸匹配。

[0214] 术语“分离的”当应用于核酸或蛋白质时,表示所述核酸或蛋白质基本上不含在天然状态下与其结合的其它细胞组分。它可以例如处于均质状态并且可以存在于无水或水溶液中。通常使用分析化学技术(如聚丙烯酰胺凝胶电泳或高效液相色谱)来测定纯度和均质性。作为存在于制剂中的主要物种的蛋白质基本上经纯化。

[0215] 术语“纯化”表示核酸或蛋白质在电泳凝胶中基本上产生一条色带。在一些实施例中,核酸或蛋白质为至少50%纯,任选地至少65%纯,任选地至少75%纯,任选地至少85%纯,任选地至少95%纯,并且任选地至少99%纯。

[0216] 术语“分离的”也可以指细胞或样品细胞。分离的细胞或样品细胞是单一细胞类型,其基本上不含当细胞处于其天然状态时或当细胞最初从其原生状态中去除时正常伴随细胞的许多组分。在某些实施例中,分离的细胞样品保留为了维持细胞处于期望状态所需的来自其自然状态的那些组分。在一些实施例中,分离的(例如纯化的、分开的)细胞或分离的细胞是基本上为样品中唯一的细胞类型的细胞。纯化的细胞样品可以含有至少60%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%或100%的一种细胞类型。分离的细胞样品可以通过使用细胞标记物或细胞标记物组合获得,其中任一种对于未纯化的细胞样品中的一种细胞类型来说是独特的。在一些实施例中,通过使用细胞分选仪分离细胞。在一些实施例中,使用针对细胞蛋白质的抗体来分离细胞。

[0217] 如本文所用,“野生型序列”是指可以作为模板用于后续反应或修饰的任何给定序列(例如分离的序列)。如技术人员所理解,野生型序列可以包括核酸序列(例如DNA或RNA或其组合)或氨基酸序列,或者可以由不同的化学实体构成。在一些实施例中,野生型序列可以指计算机模拟序列,其可以是这样的序列信息或可以机械设备可读和/或可编辑的形式存储在计算机可读介质中的序列数据。可以通过Web界面将野生型序列(反映给定顺序的核苷酸或氨基酸符号)输入例如客户门户。在实施例中,由客户最初提供的序列考虑到基于其的下游过程而被认为是野生型序列,而不管所述序列本身是天然序列还是修饰序列,即,相对于另一种野生序列被修饰,或完全是人工的。

[0218] 在实施例中,野生型序列也可以指实体分子,例如核酸分子(例如RNA或DNA或其组合)或由氨基酸构成的蛋白质、多肽或肽。通过化学,酶促或其它方式获得野生型序列的方法是本领域已知的。在一个实施例中,可以通过PCR扩增相应模板区域来获得实体核酸

野生型序列,或者可以基于合成寡核苷酸的组装从头合成。本文所用的野生型序列可涵盖天然存在的以及人工的(例如化学或酶促修饰的)部分或结构单元。野生型序列可以由两个或多个序列部分构成。野生型序列可以是例如编码区、开放阅读框、表达盒、效应域、重复结构域、启动子/增强子或终止子区域、非翻译区(UTR),但也可以是定义的序列基序,例如给定序列内的结合、识别或裂解位点。野生型序列可以是任何长度的DNA或RNA,并且可以是线性、环状或分支的,并且可以是单链或双链的。

[0219] 如本文所用,术语“结合物”是指原子或分子之间的结合。所述结合可以是直接或间接的。例如,本文提供的第一部分(例如核酸酶部分)和第二部分(DNA结合部分)之间的结合物可以是直接的,例如通过共价键,或间接的,例如通过非共价键(例如静电相互作用(例如离子键、氢键、卤键)、范德华相互作用(例如偶极-偶极、偶极诱导偶极、伦敦色散)、环堆积(π 效应)、疏水相互作用等)。在实施例,结合物使用结合物化学形成,包括但不限于亲核取代(例如胺和醇与酰卤、活性酯的反应)、亲电取代(例如烯胺反应)和碳-碳和碳-杂原子多重键加成(例如迈克尔反应、狄尔斯-阿尔德加成)。这些和其它有用的反应论述于例如March,《高等有机化学(ADVANCED ORGANIC CHEMISTRY)》,第3版,John Wiley&Sons,纽约,1985;Hermanson,《生物结合物技术(BIOCONJUGATE TECHNIQUES)》,Academic Press,圣地亚哥,1996;和Feeney等人,《蛋白质修饰(MODIFICATION OF PROTEINS)》;《化学系列进展(Advances in Chemistry Series)》,第198卷,美国化学协会(American Chemical Society),华盛顿特区,1982。在实施例,第一部分(例如核酸酶部分)与第二部分(肽部分)以非共价方式、通过第一部分(例如核酸酶部分)的组分与第二部分(DNA结合部分)的组分之间的非共价化学反应连接。在其它实施例,第一部分(例如多胺部分)包括一个或多个反应部分,例如如本文所述的共价反应部分(例如炔、叠氮化物、马来酰亚胺或硫醇反应部分)。在其它实施例,第一部分(例如核酸酶部分)包括具有一个或多个反应部分的接头,例如如本文所述的共价反应部分(例如炔、叠氮化物、马来酰亚胺或硫醇反应部分)。在其它实施例,第二部分(DNA结合部分)包括一个或多个反应部分,例如如本文所述的共价反应部分(例如炔、叠氮化物、马来酰亚胺或硫醇反应部分)。在其它实施例,第二部分(DNA结合部分)包括具有一个或多个反应部分的接头,例如如本文所述的共价反应部分(例如炔、叠氮化物、马来酰亚胺或硫醇反应部分)。

[0220] 如本文所用,术语“约”是指包括指定值的值范围,本领域普通技术人员将认为其与指定值合理相似。在实施例,术语“约”是指在使用本领域通常可接受的测量的标准偏差内。在实施例,约是指扩展至指定值的 $\pm 10\%$ 的范围。在实施例,约是指特定值。

[0221] 术语“多肽”、“肽”以及“蛋白质”在本文中可互换使用以指氨基酸残基的聚合物,其中聚合物可以与不由氨基酸组成的部分结合。所述术语适用于氨基酸聚合物,其中一个或多个氨基酸残基是天然存在的相应氨基酸的人工化学模拟物,以及天然存在的氨基酸聚合物和非天然存在的氨基酸聚合物。所述术语适用于大环肽、已经被非肽官能团修饰的肽、肽模拟物、聚酰胺以及大环内酰胺。“融合蛋白”是指编码作为单一部分重组表达的两个或更多个单独蛋白质序列的嵌合蛋白。

[0222] 术语“肽基”、“肽部分”、“蛋白质部分”和“肽基部分”是指单价肽或蛋白质。

[0223] 术语“氨基酸”是指天然存在的和合成的氨基酸,以及以与天然存在的氨基酸类似的方式起作用的氨基酸类似物和氨基酸模拟物。天然存在的氨基酸是那些由遗传密码编码

的氨基酸,以及后来经修饰的那些氨基酸,例如羟基脯氨酸、 γ -羧基谷氨酸以及O-磷酸化丝氨酸。氨基酸类似物是指具有与天然存在的氨基酸相同的基本化学结构的化合物,即与氢、羧基、氨基以及R基团结合的 α 碳,例如高丝氨酸、正亮氨酸、甲硫氨酸亚砷、甲硫氨酸甲基砷。此类类似物具有经修饰的R基团(例如正亮氨酸)或经修饰的肽骨架,但保留与天然存在的氨基酸相同的基本化学结构。氨基酸模拟物是指具有与氨基酸的一般化学结构不同的结构,但以与天然存在的氨基酸类似的方式起作用的化合物。术语“非天然存在的氨基酸”和“非天然氨基酸”是指自然界中未发现的氨基酸类似物、合成氨基酸以及氨基酸模拟物。

[0224] 氨基酸在本文中可以通过其公知的三字母符号或由IUPAC-IUB生物化学命名委员会推荐的单字母符号来提及。核苷酸同样可以通过它们通常接受的单字母代码来提及。

[0225] 氨基酸或核苷酸碱基“位置”由编号字表示,所述编号基于其相对于N端(或5'末端)的位置依次鉴定参考序列中的每个氨基酸(或核苷酸碱基)。由于在确定最佳比对时必须考虑的缺失、插入、截短、融合等,因此通常通过从N端简单计数确定的测试序列中的氨基酸残基数不一定与其在参考序列中的相应位置的编号相同。例如,在变异体相对于比对参考序列具有缺失的情况下,变异体中不存在对应于参考序列中缺失位点的位置的氨基酸。当所比对的参考序列中存在插入时,所述插入不对应于参考序列中所编号的氨基酸位置。在截短或融合的情况下,参考序列或比对序列中的氨基酸区段可以不对应于相应序列中的任何氨基酸。

[0226] 当在给定氨基酸或多核苷酸序列的编号的上下文中使用时,术语“编号参考”或“对应于”是指当给定氨基酸或多核苷酸与参考序列进行比较时所指定参考序列的残基的编号。

[0227] “保守修饰的变异体”适用于氨基酸序列和核酸序列两者。就特定的核酸序列而言,保守修饰的变异体是指编码一致或基本上一致氨基酸序列的那些核酸,或者其中核酸不编码氨基酸序列;基本一致的序列。由于遗传密码的简并性,因此大量功能相同的核酸编码任何给定的蛋白质。例如,密码子GCA、GCC、GCG以及GCU都编码氨基酸丙氨酸。因此,在丙氨酸由密码子指定的每个位置,密码子可以改变为所述的任何相应密码子而不改变编码的多肽。此类核酸变异是“静默变异”,它是保守修饰的变异中的一种。本文中编码多肽的每个核酸序列还描述了核酸的每种可能的静默变异。技术人员将认识到,核酸中的每个密码子(除了通常为甲硫氨酸的唯一密码子的AUG,和通常为色氨酸的唯一密码子的TGG之外)可被修饰以产生功能相同的分子。因此,编码多肽的核酸的每种静默变异隐含于关于表达产物的每个所述序列中,但不涉及实际探针序列。

[0228] 关于氨基酸序列,技术人员将认识到改变、添加或缺失所编码序列中的单个氨基酸或一小部分氨基酸的针对核酸、肽、多肽或蛋白质序列的个别取代、缺失或添加是“保守修饰的变异体”,其中改变引起氨基酸被化学上相似的氨基酸取代。提供功能上相似的氨基酸的保守取代表是本领域众所周知的。此类保守修饰的变异体除了以下之外并且不排除以下:多态性变异体、种间同源物以及等位基因。

[0229] 以下八个群组各自含有彼此为保守取代的氨基酸:1) 丙氨酸(A)、甘氨酸(G);2) 天冬氨酸(D)、谷氨酸(E);3) 天冬酰胺(N)、谷氨酰胺(Q);4) 精氨酸(R)、赖氨酸(K);5) 异亮氨酸(I)、亮氨酸(L)、甲硫氨酸(M)、缬氨酸(V);6) 苯丙氨酸(F)、酪氨酸(Y)、色氨酸(W);7) 丝氨酸(S)、苏氨酸(T);以及8) 半胱氨酸(C)、甲硫氨酸(M)。

[0230] 通过在比较窗中比较两个最佳对准的序列来确定“序列一致性百分比”，其中比较窗中的多核苷酸或多肽序列部分与参考序列(不包含添加或缺失)相比可以包含添加或缺失(即空位)以使两个序列达到最佳对准。如下计算百分比：确定两个序列中出现相同核酸碱基或氨基酸残基的位置数目以产生匹配位置的数目，将匹配位置的数目除以比较窗中的总位置数目并将结果乘以100而得到序列一致性百分比。

[0231] 在两个或更多个核酸或多肽序列的上下文中，术语“一致”或“一致性”百分比是指在比较窗或使用以下序列比较算法之一或通过人工比对和目视检查测量的指定区域上根据最大一致性进行比较和比对时，两个或更多个序列或子序列是相同的或具有特定百分比的相同氨基酸残基或核苷酸(即，例如多肽的完整多肽序列或个别结构域在指定区域上的60%一致性，任选65%、70%、75%、80%、85%、90%、95%、98%或99%一致性)。这样的序列然后被称为“基本上一致”。所述定义也指测试序列的补体。任选地，一致性存在于长度至少约50个核苷酸的区域上，或更优选存在于长度为100至500或1000或更多个核苷酸的区域上。

[0232] 序列比较时，通常一个序列充当与测试序列进行比较的参考序列。当使用序列比较算法时，将测试序列和参考序列输入计算机，必要时指定子序列坐标，并指定序列算法程序参数。可以使用默认程序参数，或者可以指定替代参数。然后序列比较算法根据程序参数计算测试序列相对于参考序列的序列一致性百分比。

[0233] 如本文所用，“比较窗”包括提及选自由以下组成的群组的连续位置数中的任一个的区段：例如全长序列或20至600、约50至约200或约100至约150个氨基酸或核苷酸，其中在两个序列最佳比对之后，可以将序列与具有相同数目个连续位置的参考序列进行比较。用于比较的序列比对方法在本领域中是公知的。用于比较的序列最佳比对可以例如根据Smith和Waterman的局部同源算法(1970)《高等应用数学(Adv. Appl. Math.)》2:482c，根据Needleman和 Wunsch的同源比对算法(1970)《分子生物学杂志(J. Mol. Biol.)》48:443，根据Pearson和Lipman的相似性检索方法(1988)《美国国家科学院院刊(Proc. Nat'l Acad. Sci. USA)》85:2444，通过这些算法的计算机实施方式(Wisconsin遗传学软件包中的GAP、BESTFIT、FASTA和TFASTA，遗传学计算机基团(Genetics Computer Group)，575 Science Dr., Madison, WI)，或通过人工比对和目视检查(参见例如Ausubel等人，《分子生物学最新方案(Current Protocols in Molecular Biology)》(1995年增刊))进行。

[0234] 适用于测定序列一致性和序列相似性的百分比的算法实例是BLAST和 BLAST 2.0 算法，其分别描述于Altschul等人(1977)《核酸研究(Nuc. Acids Res.)》25:3389-3402，和Altschul等人(1990)《分子生物学杂志(J. Mol. Biol.)》215:403-410中。进行BLAST分析的软件可以通过国家生物技术信息中心公开获得(<http://www.ncbi.nlm.nih.gov/>)。这种算法涉及首先通过鉴定查询序列中长度为W的短字来鉴定高得分序列对(HSP)，当与数据库序列中相同长度的字比对时，其匹配或满足一些正值阈值分数T。T被称为邻域字分数阈值(Altschul 等人，如上)。这些初始邻域字命中点充当开始检索以找到含有其的更长HSP 的种子。只要累积比对分数可以增加，那么字命中点就会沿着每个序列在两个方向上延伸。对于核苷酸序列，使用参数M(一对匹配残基的奖励分数；总是 >0)和N(错配残基的罚分；总是 <0)来计算累积分数。对于氨基酸序列来说，使用计分矩阵计算累积分数。当累积比对分数从其达到的最大值降低量X；累积分数因一次或多次负分残基比对的累积而变成零或低于

零;或到达任一序列的末端时,中断字命中点在各方向上的延伸。BLAST算法参数W、T以及 X 决定比对的灵敏度和速度。BLASTN程序(用于核苷酸序列)默认使用字长(W)11、期望值(E)或10、M=5、N=-4以及两条链的比较。对于氨基酸序列来说,BLASTP程序默认使用字长3和期望值(E)10和BLOSUM62计分矩阵(参见Henikoff和Henikoff,《美国国家科学院院刊》89:10915(1989))、比对(B)50、期望值(E)10、M=5、N=-4,以及两条链的比较。

[0235] BLAST算法还对两个序列之间的相似性进行统计分析(参见例如Karlin 和 Altschul,《美国国家科学院院刊》90:5873-5787(1993))。由BLAST算法提供的一种相似度度量是最小和概率(P(N)),其提供对两个核苷酸或氨基酸序列之间将偶然发生匹配的概率的指示。例如,如果测试核酸与参考核酸比较时的最小和概率小于约0.2,更优选小于约0.01,并且最优选小于约0.001,则认为核酸与参考序列相似。

[0236] 两个核酸序列或多肽基本上一致的指示是由第一核酸编码的多肽与针对由第二核酸编码的多肽产生的抗体发生免疫交叉反应,如下所述。因此,多肽通常与第二多肽基本上一致,例如,其中两个肽不同之处仅在于保守取代。两个核酸序列基本一致的另一指示是两个分子或其补体在严格条件下彼此杂交,如下所述。两个核酸序列基本上一致的又一指示是可以使用相同的引物来扩增所述序列。

[0237] “接触”是根据其清楚的普通含义使用,并且是指允许至少两个不同物种变得足够近以发生反应、相互作用或实体接触的方法。然而,应理解,所得反应产物可以直接由所加试剂之间的反应或者由来自一种或多种所加试剂的可以在反应混合物中产生的中间体来制备。在实施例,接触包括例如允许如本文所述的核糖核酸与内切核酸酶和增强子元件相互作用。

[0238] “对照”样品或值是指用作参照物(通常是已知参照物,用于与测试样品进行比较)的样品。例如,测试样品可以取自测试条件,例如在存在测试化合物(例如第一或第二DNA结合调节增强剂)的情况下,并且与来自已知条件的样品比较,例如在不存在测试化合物的情况下(阴性对照)或在存在已知化合物的情况下(阳性对照)。对照也可以表示从大量测试或结果中收集的平均值。本领域技术人员将认识到,可以设计用于评估任何数目个参数的对照。本领域技术人员将理解哪些标准对照在给定情况下最合适并且能够基于与标准对照值的比较来分析数据。标准对照对于确定数据的显著性(例如统计显著性)也是有价值的。例如,如果给定参数的值在标准对照中变化很大,则测试样品的变化不会被认为是显著的。

[0239] “标记”或“可检测部分”是可通过光谱、光化学、生物化学、免疫化学、化学或其它物理手段检测的组合物。例如,有用的标记包括³²P、荧光染料、电子致密试剂、酶(例如通常用于ELISA中的酶)、生物素、地高辛(digoxigenin)或半抗原和蛋白质或能被检测到的其它实体,例如通过将放射性标记并入与目标肽发生特异性反应的肽或抗体中。可使用本领域已知的用于使抗体与标记结合的任何适当方法,例如使用Hermanson,《生物结合物技术(Bioconjugate Techniques)》1996,Academic Press,Inc.,San Diego.中描述的方法。

[0240] “标记的蛋白质或多肽”是通过接头或化学键共价结合或通过离子、范德华力、静电或氢键非共价结合到标记上的蛋白质或多肽,使经标记的蛋白质或多肽的存在可以通过检测与经标记的蛋白质或多肽结合的标记的存在来检测。或者,使用高亲和力相互作用的方法可以获得相同的结果,其中一对结合搭配物中的一个结合另一个,例如生物素-链霉亲和素。

[0241] “生物样品”或“样品”是指获自或来源于受试者或患者的物质。生物样品包括组织切片,例如活组织检查样品和尸检样品,以及为了组织学目的而取的冷冻切片。此类样品包括体液,例如血液和血液部分或产物(例如血清、血浆、血小板、红细胞等)、痰、组织、培养的细胞(例如原代培养物、外植体以及转化的细胞)、粪便、尿液、滑液、关节组织、滑膜组织、滑膜细胞、成纤维细胞样滑膜细胞、巨噬细胞样滑膜细胞、免疫细胞、造血细胞、成纤维细胞、巨噬细胞、T细胞等。生物样品通常获自真核生物体,例如哺乳动物,如灵长类动物,例如黑猩猩或人类;牛;犬;猫;啮齿动物,例如豚鼠、大鼠、小鼠;兔;或鸟;爬行动物;或鱼。

[0242] 如本文所用,“细胞”是指执行足以保存或复制其基因组DNA的代谢或其它功能的细胞。细胞可以通过本领域熟知的方法鉴定,包括例如完整膜的存在、特定染料的染色、产生后代的能力,或者在配子的情况下,与第二配子组合产生有活力后代的能力。细胞可以包括原核细胞和真核细胞。原核细胞包括但不限于细菌。真核细胞包括但不限于酵母细胞和来源于植物和动物的细胞,例如哺乳动物、昆虫(例如夜蛾属(*Spodoptera*))以及人类细胞。

[0243] 术语“基因”是指涉及产生蛋白质的DNA片段;它包括编码区之前和之后的区域(增强子、启动子、前导序列和尾序列)以及个别编码区段(外显子)之间的中间序列(内含子)。增强子、启动子、前导序列、尾序列以及内含子包括在基因转录和翻译期间必需的调控元件。此外,“蛋白质基因产物”是从特定基因表达的蛋白质。

[0244] 如本文关于基因所用的词语“表达”或“表达的”是指所述基因的转录和/或翻译产物。DNA分子在细胞中的表达水平可以基于存在于细胞内的相应 mRNA 的量或由细胞产生的那个DNA编码的蛋白质的量来测定(Sambrook等人,1989,《分子克隆:实验室手册(Molecular Cloning:A Laboratory Manual)》, 18.1-18.88)

[0245] 转染的基因的表达可以在细胞中瞬时或稳定发生。在“瞬时表达”期间,转染的基因在细胞分裂过程中不转移至子细胞。由于其表达局限于转染的细胞,因此基因的表达随时间丧失。相反,当基因与另一个赋予转染细胞选择优势的基因共转染时,可以发生转染基因的稳定表达。此类选择优势可能是对呈递给细胞的某种毒素的抗性。

[0246] 术语“质粒”是指一种核酸分子,其编码为了基因表达而必需的基因和/或调控元件。来自质粒的基因的表达可以顺式或反式发生。如果基因以顺式表达,则基因和调控元件由相同质粒编码。反式表达是指基因和调控元件由各别质粒编码的情况。

[0247] 术语“游离基因型”是指细胞中质粒的染色体外状态。游离基因型质粒是核酸分子,其不是染色体DNA的一部分并且独立于其复制。

[0248] 术语“外源的”是指源自给定细胞或生物体外部的分子或物质(例如核酸或蛋白质)。相反,术语“内源的”是给定细胞或生物体原生的或源于给定细胞或生物体内的分子或物质。

[0249] “细胞培养物”是存在于生物体外的体外细胞群。细胞培养物可以由分离自细胞库或动物的原代细胞,或源自这些来源中的一个并且永生以进行长期体外培养的传代细胞建立。如本文所提供的细胞培养物进一步指包括合适细胞营养素并且能够在体外维持细胞的环境。所述环境可以是合适容器(例如细胞培养皿)中的液体环境、固体环境和/或半固体环境(例如琼脂、凝胶等)。可以使用细胞培养基。如本文所用,“细胞培养基”根据其在在本领域中通常接受的含义使用。细胞培养基(在本领域和本文中也称为“培养基”)包括为了支持细胞生长(例如分裂、分化、维持等)而设计的液体(例如生长因子、矿物质、维生素等)或凝

胶。在实施例中,本文提供的组合物(包括实施例)进一步包括生理学上可接受的溶液。如本文所提供的“生理学上可接受的溶液”是指任何可接受的水溶液(例如缓冲液),其中可以包含本文提供的组合物而不丧失其生物学特性。在实施例中,生理学上可接受的溶液是细胞培养基。

[0250] 术语“转染(transfection)”、“转导(transduction)”、“转染(transfecting)”或“转导(transducing)”可以互换使用并且定义为将核酸分子和/或蛋白质引入细胞的过程。可以使用基于非病毒或基于病毒的方法将核酸引入细胞。核酸分子可以是编码完整蛋白质或其功能部分的序列。典型地,核酸载体包含蛋白质表达必需的元件(例如启动子、转录起始位点等)。非病毒转染方法包括不使用病毒DNA或病毒颗粒作为递送系统将核酸分子引入细胞的任何合适方法。示例性的非病毒转染方法包括磷酸钙转染、脂质体转染、核转染、声孔效应、通过热休克转染、磁化转染以及电穿孔。对于基于病毒的方法来说,任何可用的病毒载体均可用于本文所述的方法中。病毒载体的实例包括但不限于逆转录病毒、腺病毒、慢病毒以及腺相关病毒载体。在一些方面中,按照本领域公知的标准程序使用逆转录病毒载体将核酸分子引入细胞中。术语“转染”或“转导”也指将蛋白质从外部环境引入细胞。通常,蛋白质的转导或转染依赖于能够穿过细胞膜的肽或蛋白质与所关注蛋白质的连接。参见例如Ford等人,《基因疗法(Gene Therapy)》8:1-4(2001)和Prochiantz,《自然·方法(Nat. Methods)》4:119-120(2007)。

[0251] 如本文所用,术语“特异性结合(specific binding)”或“特异性结合(specifically binds)”是指两种分子在生理条件下形成相对稳定的复合物(例如DNA结合增强剂和增强子结合序列)。

[0252] 本文提供的“核糖核蛋白复合物”、“核糖核蛋白颗粒”、“脱氧核糖核蛋白复合物”或“脱氧核糖核蛋白颗粒”是指包含核蛋白和核糖核酸或脱氧核糖核酸的复合物或颗粒。如本文所提供的“核蛋白”是指能够结合核酸(例如RNA、DNA)的蛋白质。在核蛋白结合核糖核酸的情况下,其被称作“核糖核蛋白”。在核蛋白结合脱氧核糖核酸的情况下,其被称作“脱氧核糖核蛋白”。核糖核蛋白和核糖核酸之间的相互作用或脱氧核糖核蛋白和核糖核酸之间的相互作用可以是直接的,例如通过共价键,或者间接的,例如通过非共价键(例如静电相互作用(例如离子键、氢键、卤键)、范德华相互作用(例如偶极-偶极、偶极诱导偶极、伦敦色散)、环堆积(π 效应)、疏水相互作用等)。在实施例中,核糖核蛋白包含与核糖核酸非共价结合的RNA结合基序。在实施例中,核糖核蛋白包含与脱氧核糖核酸非共价结合的RNA结合基序。例如,RNA结合基序或DNA结合基序中带正电荷的芳族氨基酸残基(例如赖氨酸残基)可与RNA或DNA的负核酸磷酸骨架形成静电相互作用,从而形成核糖核蛋白复合物或脱氧核糖核蛋白复合物(例如本文中提及的阿格蛋白复合物)。核糖核蛋白的非限制性实例包括核糖体、端粒酶、RNaseP、hnRNP、CRISPR相关蛋白9(Cas9)以及小核RNP(snRNP)。脱氧核糖核蛋白的一个实例是阿格蛋白。核糖核蛋白或脱氧核糖核蛋白可以是酶。在实施例中,核糖核蛋白或脱氧核糖核蛋白是内切核酸酶。因此,在实施例中,核糖核蛋白复合物包括内切核酸酶和核糖核酸。在实施例中,内切核酸酶是CRISPR相关蛋白9。因此,在实施例中,脱氧核糖核蛋白复合物包括内切核酸酶和脱氧核糖核酸。在实施例中,内切核酸酶是阿格蛋白核酸酶。

[0253] 如本文所提供的“向导RNA”或“gRNA”指能够结合核蛋白、从而形成核糖核蛋白复合物的核糖核苷酸序列。同样,如本文所提供的“向导DNA”或“gDNA”指能够结合核蛋白、从

而形成脱氧核糖核蛋白复合物的脱氧核糖核苷酸序列。在实施例中,向导RNA包括一个或多个RNA分子。在实施例中,向导DNA包括一个或多个DNA分子。在实施例中,gRNA包括与目标位点互补的核苷酸序列(例如调节子结合序列)。在实施例中,gDNA包括与目标位点互补的核苷酸序列(例如调节子结合序列)。互补核苷酸序列可以介导核糖核蛋白复合物或脱氧核糖核蛋白复合物与所述目标位点的结合,从而提供核糖核蛋白复合物或脱氧核糖核蛋白复合物的序列特异性。因此,在实施例中,向导RNA或向导DNA与目标核酸(例如调节子结合序列)互补。在实施例中,向导RNA结合目标核酸序列(例如调节子结合序列)。在实施例中,向导DNA结合目标核酸序列(例如调节子结合序列)。在实施例中,向导RNA与CRISPR核酸序列互补。在实施例中,向导RNA或向导DNA的补体与目标核酸(例如调节子结合序列)具有约50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%或99%的序列一致性。如本文所提供的目标核酸序列是由细胞表达的核酸序列。在实施例中,目标核酸序列是外源核酸序列。在实施例中,目标核酸序列是内源核酸序列。在实施例中,目标核酸序列(例如调节子结合序列)形成细胞基因的一部分。因此,在实施例中,向导RNA或向导DNA与细胞基因或其片段互补。在实施例中,向导RNA或向导DNA与目标核酸序列约50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%或99%互补。在实施例中,向导RNA或向导DNA与细胞基因序列约50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%或99%互补。在实施例中,向导RNA或向导DNA结合细胞基因序列。术语“目标核酸序列”是指本文提供的调节子结合序列。

[0254] 在实施例中,向导RNA或向导DNA是单链核糖核酸。在实施例中,向导RNA或向导DNA的长度为约10、20、30、40、50、60、70、80、90、100或更多个核酸残基。在实施例中,向导RNA或向导DNA的长度为约10至约30个核酸残基。在实施例中,向导RNA或向导DNA的长度为约20个核酸残基。在实施例中,向导RNA或向导DNA的长度可为至少约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、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100或更多个核酸残基或糖残基。在实施例中,向导RNA或向导DNA的长度为5至50、10至50、15至50、20至50、25至50、30至50、35至50、40至50、45至50、5至75、10至75、15至75、20至75、25至75、30至75、35至75、40至75、45至75、50至75、55至75、60至75、65至75、70至75、5至100、10至100、15至100、20至100、25至100、30至100、35至100、40至100、45至100、50至100、55至100、60至100、65至100、70至100、75至100、80至100、85至100、90至100、95至100或更多个残基。在实施例中,向导RNA或向导DNA的长度为10至15、10至20、10至30、10至40或10至50个残基。

[0255] PAM是指“原间隔子相邻基序”。这些位点通常是与Cas9结合的DNA序列相邻的2-6个碱基对DNA序列。因此,在某些情况下,可以使用除Cas9以外的DNA结合调节增强剂,而在其它情况下,可以使用单个Cas9/RNA复合物作为DNA结合调节增强剂(单独使用或与不同的DNA结合调节增强剂联合使用)。

[0256] 对于本文描述的特定蛋白质(例如Cas9、阿格蛋白)而言,所述蛋白质包括任何蛋白质的天然存在形式,或变异体或同源物,其维持蛋白质转录因子活性(例如,与原生蛋白质相比,活性在至少50%、80%、90%、95%、96%、97%、98%、99%或100%活性范围内)。

在一些实施例中,与天然存在形式相比,变异体或同源物在整个序列或序列的一部分(例如50、100、150或200个连续氨基酸部分)上具有至少90%、95%、96%、97%、98%、99%或100%氨基酸序列一致性。在其它实施例中,蛋白质是根据其NCBI序列参照物鉴定的蛋白质。在其它实施例中,蛋白质是根据其NCBI序列参照物或其功能片段或同源物鉴定的蛋白质。

[0257] 因此,本文所指的“CRISPR相关蛋白9”、“Cas9”或“Cas9蛋白”包括 Cas9内切核酸酶的任何重组或天然存在形式或其变异体或同系物,其维持 Cas9内切核酸酶活性(例如,与Cas9相比,活性在至少50%、80%、90%、95%、96%、97%、98%、99%或100%内)。在一些方面中,与天然存在的Cas9蛋白相比,所述变异体或同源物在整个序列或序列的一部分(例如50、100、150或200个连续氨基酸部分)上具有至少90%、95%、96%、97%、98%、99%或100%的氨基酸序列一致性。在实施例中,Cas9蛋白与根据UniProt参考号Q99ZW2或与其具有基本一致性的变异体或同源物鉴定的蛋白质基本上一致。Cas9是指本领域中也称为“切口酶”的蛋白质。在实施例中,Cas9结合CRISPR(成簇的有规律间隔的短回文重复序列)核酸序列。在实施例中,CRISPR核酸序列是原核核酸序列。本文适用的Cas9蛋白的实例包括但不限于cas9突变蛋白,例如如Kleinstiver,Benjamin P.等人所述的HiFi Cas9(“不具有可检测到的全基因组脱靶效应的高保真CRISPR-Cas9核酸酶(High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects.)”《自然(Nature)》(2016).PubMed PMID:26735016);Cas9蛋白质结合经修饰的PAM和直系同源Cas9蛋白,例如Prevotella和Francisella 1的CRISPR(Cpf1)。本领域中通常已知和描述的任何突变Cas9形式均可用于本文提供的方法和组合物。预期用于本文所提供的方法和组合物中的突变型 Cas9蛋白的非限制性实例描述于Slaymaker,Ian M.等人(“合理地经工程改造的具有改善的特异性的Cas9核酸酶(Rationally engineered Cas9 nucleases with improved specificity)”《科学(Science)》(2015):aad5227.PubMed PMID: 26628643)和Kleinstiver,Benjamin P.等人(“全基因组脱靶效应检测不到的高保真度CRISPR-Cas9核酸酶”《自然》(2016).PubMed PMID:26735016)。

[0258] 如本文中所提及,术语“阿格蛋白(AGO)蛋白质”、“NgAgo”或“格氏嗜盐碱杆菌(Natronobacterium gregoryi)阿格蛋白”、“格氏嗜盐碱杆菌SP2阿格蛋白”包括维持NgAgo内切核酸酶活性(例如相较于野生型NgAgo,在至少50%、80%、90%、95%、96%、97%、98%、99%或100%活性范围内的)NgAgo或其变异体或同源物的重组形式或天然存在的形式。在实施例中,与天然存在的NgAgo蛋白相比,所述变异体或同源物在整个序列或序列的一部分(例如50、100、150或200个连续氨基酸部分)上具有至少70%、75%、80%、85%、90%、95%、96%、97%、98%、99%或100%氨基酸序列一致性。在实施例中,NgAgo蛋白与美国国家生物技术信息中心(National Center for Biotechnology Information,NCBI)蛋白质标识符AFZ73749.1所鉴定的蛋白质或与其具有至少70%、75%、80%、85%、90%、95%、96%、97%、98%、99%或100%氨基酸序列一致性的变异体或同源物基本上一致。在实施例中,阿格蛋白还可以包括核酸酶域(即,脱氧核糖核酸酶或核糖核酸酶域)、额外的DNA结合域、解螺旋酶域、蛋白质-蛋白质相互作用域、二聚合域以及其它域。

[0259] 阿格蛋白还指形成结合核酸分子的复合物的蛋白质。因此,一种阿格蛋白可以结合到例如向导RNA且另一种蛋白质可以具有内切核酸酶活性。这些蛋白质都被认为是阿格

蛋白,原因是其充当与单一蛋白质(如NgAgo)执行相同功能的复合物的一部分。

[0260] 如本文中所用,术语“阿格蛋白系统”是指一组阿格蛋白和核酸,当组合时,其产生至少与阿格蛋白相关的活性(例如双链DNA的目标基因座特异性双链裂解)。

[0261] 如本文中所用,术语“阿格蛋白复合物”是指阿格蛋白和核酸(例如RNA)彼此结合而形成具有功能活性的聚集物。阿格蛋白复合物的一个实例是与对目标基因座具有特异性的向导DNA结合的格氏嗜盐碱杆菌阿格蛋白(NgAgo)蛋白质。

[0262] 在实施例中,本文提及的“阿格蛋白(AGO)蛋白质”、“NgAgo”或“格氏嗜盐碱杆菌阿格蛋白”、“格氏嗜盐碱杆菌SP2阿格蛋白”包括维持NgAgo 内切核酸酶活性(例如相较于野生型NgAgo,在至少50%、80%、90%、95%、96%、97%、98%、99%或100%活性范围内)的NgAgo或其变异体或同源物的任何重组或天然存在的形式。在实施例中,与天然存在的NgAgo蛋白相比,所述变异体或同源物在整个序列或序列的一部分(例如50、100、150或200个连续氨基酸部分)上具有至少70%、75%、80%、85%、90%、95%、96%、97%、98%、99%或100%氨基酸序列一致性。在实施例中,NgAgo蛋白与根据美国国家生物技术信息中心(NCBI)蛋白质标识符AFZ73749.1所鉴定的蛋白质或与其具有至少70%、75%、80%、85%、90%、95%、96%、97%、98%、99%或100%氨基酸序列一致性的变异体或同源物基本上一致。在实施例中,阿格蛋白还可以包括核酸酶域(即,脱氧核糖核酸酶或核糖核酸酶域)、额外的DNA结合域、解螺旋酶域、蛋白质-蛋白质相互作用域、二聚合域以及其它域。

[0263] 阿格蛋白还指形成结合核酸分子的复合物的蛋白质。因此,一种阿格蛋白可以结合到例如向导RNA且另一种蛋白质可以具有内切核酸酶活性。这些蛋白质都被认为是阿格蛋白,原因是其充当与单一蛋白质(如NgAgo)执行相同功能的复合物的一部分。

[0264] 如本文所用,术语“转录调控序列”是指核酸分子上所含的呈任何构形或几何形状的功能性核苷酸片段,其用于调控(1)一种或多种结构基因(例如二、三、四、五、七、十种等)转录成信使RNA或(2)一种或多种基因转录成非翻译RNA。转录调控序列的实例包括但不限于启动子、增强子、抑制子等。

[0265] 如本文中所用,术语“核酸靶向能力”是指分子或分子复合物根据序列特异性来识别核酸和/或与核酸结合的能力。例如,调节蛋白或调节复合物与向导DNA(gDNA)分子上的调节子结合序列或杂交区域的结合赋予核酸靶向阿格蛋白复合物的能力。

[0266] 如本文所用,本文提供的“TAL效应物”或“TAL效应蛋白”是指包括超过一个TAL重复序列并且能够以序列特异性方式结合核酸的蛋白质。在实施例中,TAL效应蛋白包括至少六个(例如至少8个、至少10个、至少12个、至少15个、至少17个、约6到约25个、约6到约35个、约8至约25个、约10至约25个、约12至约25个、约8至约22个、约10至约22个、约12至约22个、约6至约20个、约8至约20个、约10至约22个、约12至约20个、约6至约18个、约10至约18个、约12至约18个等)TAL重复序列。在实施例中,TAL效应蛋白包括18或24或17.5或23.5TAL核酸结合盒。在实施例中,TAL效应蛋白包括15.5、16.5、18.5、19.5、20.5、21.5、22.5或24.5TAL核酸结合盒。TAL效应蛋白包括至少一个与含有TAL重复序列的区域侧接的多肽区域。在实施例中,侧接区域存在于TAL重复序列的氨基端和/或羧基端。

[0267] 如本文所用,“调控序列”是指影响转录和/或翻译起始以及转录物或多肽产物的速率、稳定性和/或迁移率的核酸序列。调控序列包括但不限于启动子序列或控制元件、增强子序列、反应元件、蛋白质识别位点、诱导型元件、蛋白质结合序列、转录起始位点、终止

序列、聚腺苷酸化序列、内含子、5'和3'非翻译区(UTR)和其它可存在于编码序列内的调控序列,例如剪接位点、抑制序列元件(通常称为CNS或INS,如从某些病毒中得知)、分泌信号、核定位信号(NLS)序列、内含子、翻译偶联序列、蛋白酶裂解位点,如本文其它地方更详细所述。5'非翻译区(UTR)被转录但未翻译,并且位于转录物的起始位点和翻译起始密码子之间,并且可以包括+1个核苷酸。3'UTR可以位于翻译终止密码子和转录本末端之间。UTR可以具有特定功能,例如增加mRNA 消息稳定性或翻译衰减。3'UTR的实例包括但不限于聚腺苷酸化信号和转录终止序列。调控序列可以是通用的或宿主或组织特异性的。

[0268] 如本文所用,“启动子”是转录调控序列,当其可操作地连接至核酸片段(例如包含例如开放阅读框的转基因)时,其能够引导核酸片段的转录。启动子是位于转录起始位点上游(通常在RNA聚合酶II的起始位点附近)的核苷酸序列。启动子通常包含至少一个核心或基础基序,并且可以包括至少一个或多个控制元件或与至少一个或多个控制元件协同作用,例如上游元件(例如上游激活区(UAR))或其它调控序列或合成元件。基础基序构成了转录起始所需的转录复合物组装所需的最小序列。在实施例中,此类最小序列包括“TATA 盒”元件,其可以位于转录起始位点上游约15至约35个核苷酸之间。基础启动子还可包括“CCAAT盒”元件(通常为序列CCAAT)和/或GGGGG序列,其可位于转录起始位点上游约40至约200个核苷酸之间,通常为约60至约120个核苷酸。相邻核酸片段的转录起始于启动子区域。抑制性启动子的转录速率响应于抑制剂而降低。诱导性启动子的转录速率响应于诱导剂而升高。组成型启动子的转录速率不受特异性调控,然而其在一般代谢条件的影响下可以变化。

[0269] 表达载体中包含的启动子的选择取决于几个因素,包括但不限于效率、选择性、诱导性、所需表达水平以及细胞或组织特异性。例如,可以使用分别在特定组织、器官和细胞类型中仅赋予或主要赋予转录的组织、器官和细胞特异性启动子。在实施例中,对种子基本上具特异性的启动子(“种子优先启动子”) 可以是有用的。在实施例中,使用组成型启动子,其可以促进特定物种的大多数或所有组织中的转录。其它类别的启动子包括但不限于诱导型启动子,例如响应于外部刺激(例如化学试剂、发育刺激或环境刺激)而赋予转录的启动子。诱导型启动子可以被病原体或冷或应激(如冷、热、紫外线或高离子浓度)诱导,或可以被化学物质诱导。诱导型启动子的实例是真核金属硫蛋白启动子,其由升高的重金属含量诱导;原核lacL启动子,其响应于异丙基- β -D-硫代半乳糖吡喃糖苷(IPTG)而被诱导;以及被升高的温度诱导的真核热休克启动子。如本文所述适用的许多其它细菌和真核启动子是本领域已知的,并描述于例如 Sambrook等人,《分子克隆:实验室手册(Molecular Cloning, A Laboratory Manual)》(第2版,1989年;2001年,第3版);Kriegler,《基因转移和表达:实验室手册(Gene Transfer and Expression: A Laboratory Manual)》(1990);和 Ausubel等人,《分子生物学现代方法(Current Protocols in Molecular Biology)》。用于表达ZFP的细菌表达系统可获自例如大肠杆菌、芽孢杆菌属和沙门氏菌(Palva等人,枯草芽孢杆菌的干扰素分泌(Secretion of interferon by *Bacillus subtilis*.)《基因(Gene)》22:229-235(1983))。用于此类表达系统的试剂盒是可商购的。用于哺乳动物细胞、酵母和昆虫细胞的真核表达系统是本领域技术人员众所周知的,并且也是可商购的。

[0270] 用于原核蛋白表达的常见启动子是例如lac启动子或trc和tac启动子(IPTG诱导)、tetA启动子/操纵子(脱水四环素诱导)、PPBAD启动子(L-阿拉伯糖诱导)、r/zaPBAD启

动子(L-鼠李糖诱导)或噬菌体启动子,例如噬菌体启动子pL(对温度变化敏感)、T7、T3、SP6或T5。

[0271] 用于哺乳动物蛋白表达的常见启动子是例如巨细胞病毒(CMV)启动子、SV40启动子/增强子、牛痘病毒启动子、病毒LTR(MMTV、RSV、HIV等)、E1B启动子、组成型表达基因(肌动蛋白,GAPDH)的启动子、以组织特异性方式表达的基因(白蛋白,NSE)的启动子、诱导型基因(金属硫蛋白、类固醇激素)的启动子。

[0272] 在植物中表达核酸的许多启动子是已知的并且可以使用。此类启动子可以是组成型的、可调控的和/或组织特异性的(例如种子特异性、茎特异性、叶特异性、根特异性、果实特异性等)。可以用于植物表达的示例性启动子包括花椰菜花叶病毒35S启动子和以下基因的启动子:来自拟南芥的ACT 11和CAT 3基因、编码来自甘蓝型油菜的硬脂酰基-酰基载体蛋白去饱和酶的基因(GenBank编号X74782),以及来自玉米的编码GPC1(GenBank编号X15596)和GPC2(GenBank编号U45855)的基因。其它启动子包括烟草花叶病毒亚基因组启动子、木薯静脉花叶病毒(CVMV)启动子(在血管元元件、叶肉细胞和根尖中表现出高转录活性)、玉米的干旱诱导型启动子以及马铃薯的寒冷、干旱和高盐诱导型启动子。在美国专利第8,067,222号中发现了许多适合植物表达的其它启动子,其公开内容通过引用结合到本文中。

[0273] 在微藻(例如莱茵衣藻)叶绿体中的异源表现能够在缺乏psbA的遗传背景下使用例如psbA启动子/5'非翻译区(UTR)实现(由于psbA/D1依赖性自体衰减)或通过使强16S rRNA启动子与psbA和atpA基因的5'UTR在表达盒中融合来实现,如例如以下文献中所公开: Rasala等人,“通过启动子和5'非翻译区优化来改善异源蛋白质在莱茵衣藻叶绿体中的表达(Improved heterologous protein expression in the chloroplast of *Chlamydomonas reinhardtii* through promoter and 5'untranslated region optimization)”,《植物生物技术杂志(Plant Biotechnology Journal)》第9卷,第6期,第674-683页(2011)。用于引导TAL效应物编码核酸表达的启动子取决于特定应用。例如,强组成型启动子通常用于TAL效应物融合蛋白的表达和纯化。相反,当在体内施用TAL效应物核酸酶融合蛋白用于基因调控时,视TAL效应物核酸酶融合蛋白的特定用途和其它因素而定,可能需要使用组成型或诱导型启动子。另外,适于施用TAL效应物核酸酶融合蛋白的启动子可以是弱启动子,例如HSV胸苷激酶或具有相似活性的启动子。启动子通常还可以包括对反式激活有反应的元件,例如低氧反应元件、Gal4反应元件、lac阻遏物反应元件,以及小分子控制系统,例如tet调控系统和RU-486系统(参见例如Gossen和Bujard.通过四环素反应性启动子严格控制哺乳动物细胞中的基因表达(Tight control of gene expression in mammalian cells by tetracycline-responsive promoters).《美国国家科学院院刊》89: 5547(1992);Oligino等人,使用单纯疱疹病毒载体在大脑中实现药物诱导的转基因表达(Drug inducible transgene expression in brain using a herpes simplex virus vector).《基因疗法(Gene Ther.)》5:491-496(1998);Wang等人,利用诱导型转录调控因子对真核细胞中的基因表达实现正向和负向调控(Positive and negative regulation of gene expression in eukaryotic cells with an inducible transcriptional regulator).《基因疗法》4:432-441(1997);Neering 等人,利用重组腺病毒载体转导原始人类造血细胞(Transduction of primitive human hematopoietic cells with

recombinant adenovirus vectors).《血液(Blood)》88:1147-1155(1996);和Rendahl等人,通过两个单独的rAAV载体转导之后的基因表达的调控(Regulation of gene expression in vivo following transduction by two separate rAAV vectors).《自然·生物技术(Nat.Biotechnol.)》16:151-161 (1998)。也可以使用MNDU3启动子,并且优选在CD34+造血干细胞中具有活性。

[0274] “宿主”是指支持载体复制或载体序列所编码的蛋白质或多肽表达的细胞或生物体。宿主细胞可以是原核细胞,例如大肠杆菌,或真核细胞,例如酵母、真菌、原生动物、高等植物、昆虫或两栖动物细胞,或哺乳动物细胞,例如CHO、HeLa、293、COS-1等等,例如所培养的细胞(体外)、外植体和原代培养物(体外和离体),以及体内细胞。

[0275] 如本文中所用,短语“重组蛋白”包括参与涉及一个或多个重组位点(例如二、三、四、五、七、十、十二、十五、二十、三十、五十个等)的重组反应的切除型或整合型蛋白质、酶、辅因子或相关蛋白质,其可以是野生型蛋白质(参见Landy,《生物技术新见(Current Opinion in Biotechnology)》3:699-707 (1993))或其突变体、衍生物(例如含有重组蛋白序列或其片段的融合蛋白)、片段和变异体。重组蛋白的实例包括Cre、Int、IHF、Xis、Flp、Fis、Hin、Gin、Phi-C31、Cin、Tn3解离酶、TndX、XerC、XerD、TnpX、Hjc、SpCCE1和Par A。

[0276] 如本文所用,短语“重组位点”是指核酸分子上的识别序列,其通过重组蛋白参与整合/重组反应。重组位点是参与的核酸分子上的核酸的不连续部分或区段,其在整合或重组的初始阶段被位点特异性重组蛋白识别并结合。例如,Cre重组酶的重组位点是loxV,它是包含两个13个碱基对反向重复序列(用作重组酶结合位点)的34个碱基对序列,其侧接8个碱基对的核心序列(参见以下文献中的图1:Sauer, B.,《位点特异性重组:开发和应用(Site-specific recombination:developments and applications)》.《生物技术新见(Curr.Opin.Biotech.)》5:521-527(1994))。识别序列的其它实例包括本文所述的attB、attP、attL和attR序列,以及其突变体、片段、变异体和衍生物,它们被重组蛋白λ噬菌体整合酶和辅助蛋白整合宿主因子(IHF)、Fis和切除酶(λ噬菌体is)识别。

[0277] 如本文所用,短语“识别序列”是指蛋白质、化合物、DNA或RNA分子(例如限制性内切核酸酶、修饰甲基化酶或重组酶)识别并结合的特定序列。如本文所用,识别序列通常指重组位点。举例来说,Cre重组酶的识别序列是loxP,其为34个碱基对的序列,其包含侧接8个碱基对的核心序列的两个13个碱基对的反向重复序列(充当重组酶结合位点)(参见Sauer, B.,《生物技术新见》5:521-527(1994)的图1)。识别序列的其它实例是被重组酶λ噬菌体整合酶识别的attB、attP、attL和attR序列。attB是约25个碱基对的序列,其含有两个9个碱基对的核心型Int结合位点和7个碱基对的重叠区。attP是约240个碱基对的序列,其含有核心型Int结合位点和臂型Int结合位点,以及辅助蛋白整合宿主因子(IHF)、FIS和切除酶(λ噬菌体is)的位点。(参见Landy,《生物技术新见》3:699-707(1993))。

[0278] 在整个文档中,除非上下文另有要求,否则词语“包含(comprise)”、“包含(comprises)”和“包含(comprising)”或“含有(contain)”、“含有(contains)”或“含有(containing)”将被理解为暗示包括所述步骤或要素或一组步骤或元素,但不排除任何其它步骤或要素或一组步骤或要素。

[0279] 如本文所用,术语“同源重组”是指遗传重组的机制,其中包含相似核苷酸序列的两条DNA链交换遗传物质。细胞在减数分裂期间利用同源重组,其中同源重组不仅用于使

DNA重排以产生一组完全独特的单倍体染色体,而且用于修复损伤的DNA,尤其是用于修复双链断裂。同源重组的机制是技术人员众所周知的,并且已经描述于例如Paques和Haber (Paques F, Haber J E.;《微生物学和分子生物学评论 (Microbial.Mol.Biol.Rev.)》63: 349-404 (1999))。在各方面中,同源重组是通过分别位于所述供体DNA序列的上游(5')和下流(3')的所述第一侧接元件和所述第二侧接元件的存在来实现,所述侧接元件各自与所述目标序列内的连续DNA序列同源。

[0280] 如本文所用,术语“非同源末端连接”(NHEJ)是指通过很大程度上独立于同源性的过程将双链断裂(DSB)的两个末端连接的细胞过程。天然存在的 DSB是在DNA合成期间(此时复制叉遇到损伤的模板)以及在某些特殊的细胞过程期间(包括V(D)J重组、免疫球蛋白重链(IgH)基因座处发生的类别转换重组,和减数分裂)自发地产生。此外,细胞暴露于电离辐射(X射线和伽马射线)、紫外光、拓扑异构酶毒物或放射模拟药物会产生DSB。NHEJ(非同源末端连接)途径通过很大程度上独立于同源性的过程连接DSB的两个末端。视DSB处所产生的特定序列和化学修饰而定,NHEJ可以是精确的或突变诱发的(Lieber M.R.,通过非同源DNA末端连接途径修复双链DNA断裂的机制(The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway).《生物化学年鉴(Annu Rev Biochem)》.79:181-211)。

[0281] 如本文所用,术语“供体DNA”或“供体核酸”是指为了通过同源重组引入基因座中而设计的核酸。供体核酸具有与基因座序列同源的至少一个区域。在实施例中,供体核酸具有与基因座序列同源的两个区域。这些同源区域可以位于两个末端之一或可以位于供体核酸内部。在实施例中,期望被引入细胞中存在的核酸分子中的核酸“插入”区域将位于两个同源区域之间。

[0282] 供体核酸分子(例如供体DNA分子)可以是双链、单链或部分双链和单链的,且因此,可以在一个或两个末端具有突出端(例如两个5'突出物、两个3'突出物、一个5'和一个3'突出物、单个3'突出物或单个5'突出物)。此外,核酸分子可以是环状核酸分子的线性核酸分子(闭合的环状或带切口的核酸分子)。

[0283] 如本文所用,术语“同源重组系统”或“HR系统”是指本文阐述的系统的组件,其可用于通过同源重组来改变细胞。具体地说,锌指核酸酶、TAL效应物核酸酶、CRISPR内切核酸酶、归巢内切核酸酶和阿高编辑系统。

[0284] 如本文所用,术语“核酸切割实体”是指具有核酸切割活性(例如双链核酸切割活性)的单一分子或分子复合物。示例性的核酸切割实体包括阿高复合物、锌指蛋白、转录激活因子样效应物(TALE)、CRISPR复合物和归巢的大范围核酸酶。在实施例中,核酸切割实体将具有允许它们被核定位的活性(例如将包含核定位信号(NLS))。

[0285] 如本文所用,术语“双链断裂位点”是指核酸分子中发生双链断裂的位置。在实施例中,这将通过在两个接近的位置(例如在约3至约50个碱基对、约5至约50个碱基对、约10至约50个碱基对、约15至约50个碱基对、约20至约50个碱基对、约3至约40个碱基对、约5至约40个碱基对、约10至约40个碱基对、约15至约40个碱基对、约20至约40个碱基对范围内等)切割核酸分子来产生。通常,与包含较高GC含量的核酸区域相比,包含较高AT含量的核酸区域中的切口间距可以进一步加大。

[0286] 如本文所用,术语“匹配端”是指具有大于90%的序列一致性的核酸分子的末端。

目标基因座的DS断裂的匹配端可以是双链或单链的。供体核酸分子的匹配端通常是单链的。

[0287] 如本文所用,“同源定向修复”或“HDR”是细胞中修复DNA双链断裂 (DSB) 的机制。在一些实施例中,HDR大于或等于10%、25%、50%、75%、90%、95%、98%、99%或100%。

[0288] HDR的常见形式是“同源重组”,是指一种遗传重组机制,其中包含相似核苷酸序列的两条DNA链交换遗传物质。细胞在减数分裂期间利用同源重组,其中同源重组不仅用于使DNA重排以产生一组完全独特的单倍体染色体,而且用于修复损伤的DNA,尤其是用于修复双链断裂。同源重组的机制是技术人员众所周知的,并且已经描述于例如Paques F.,Haber J.E.,《微生物学和分子生物学评论(Microbiol.Mol.Biol.Rev.)》63:349-404(1999)。在一些实施例中,通过分别位于供体核酸分子上游(5')和下游(3')的匹配端的存在来实现同源重组,所述匹配端各自与裂解的核酸分子内的连续DNA序列同源。

[0289] 一些实施例包括为了在细胞(例如真核细胞,例如植物细胞和动物细胞,例如昆虫细胞和哺乳动物细胞,包括小鼠、大鼠、仓鼠、兔和人类细胞)中提高同源重组效率而设计的组合物和方法。在一些实施例中,同源重组效率应使得群体中大于20%的细胞将在所需目标基因座或基因座处进行过同源重组。在一些实施例中,同源重组可以在群体中10%至65%、15%至65%、20%至65%、30%至65%、35%至65%、10%至55%、20%至55%、30%至55%、35%至55%、40%至55%、10%至45%、20%至45%、30%至45%、40%至45%、30%至50%的细胞等范围内发生。

[0290] 此外,一些实施例包括用于提高细胞内同源重组效率的组合物和方法。例如,如果在一组条件下同源重组发生在10%的细胞群中,而在另一组条件下发生在40%的细胞群中,则同源重组的效率提高了300%。在一些实施例中,同源重组的效率可以增加100%至500%(例如100%至450%、100%至400%、100%至350%、100%至300%、200%至500%、200%至400%、250%至500%、250%至400%、250%至350%、300%至500%等)。

[0291] 如本文所用,“双链断裂”或“DSB”是指核酸分子中的双链断裂。在多个实施例中,DSB将通过在两个接近的位置(例如在3到50个碱基对、5到50个碱基对、10到50个碱基对、15到50个碱基对、20到50个碱基对、3到40个碱基对、5到40个碱基对、10到40个碱基对、15到40个碱基对、20到40个碱基对范围内等)切割核酸分子来产生。与包含较高GC含量的核酸区域相比,在包含较高AT含量的核酸区域中,切口的间距可进一步加大。在一些实施例中,对于核酸分子的N端标记来说,双链断裂距ATG起始密码子小于或等于250bp,或对于核酸分子的C端标记来说,距终止密码子小于或等于250 bp。

[0292] 如本文所用,“供体核酸分子”或“供体DNA”是指为了通过同源重组引入被裂解的核酸分子中而设计的核酸。供体核酸分子将具有至少一个与裂解的核酸分子序列同源的区域。在许多实施例中,供体核酸分子将具有两个与基因座序列同源的区域。这些同源区域可以位于一个或两个末端,或者可以位于供体核酸分子的内部。

[0293] 如本文所用,“整合效率”是指所关注的外来DNA区段并入初始核酸分子的频率。在一些实施例中,供体核酸分子的整合效率大于或等于50%、75%、90%、95%、98%、99%或100%。

[0294] 表1显示,在三种不同的哺乳动物细胞系中,在四个不同的基因组基因座处,发现了近100%的整合效率和高达100%的精确HDR。在某些基因座处,观察到接合点处或Cas9裂

解位点处的缺失和插入。

表 1. 外源 DNA 分子整合到人类基因组中的效率							
基因座	位置	插入/尺寸	等位基因 1 的效率		等位基因 2 的效率		功能细胞百分比
			精确的 HDR	HDR/插入缺失	精确的 HDR	无 HDR/插入缺失	
[0295]	β-肌动蛋白 SEQ ID NO: 3	N 端 嘌呤霉素-P2A- OFP/1.4 kb	68	32	0	100	80% ^a
	β-肌动蛋白 SEQ ID NO: 3	N 端 嘌呤霉素-CMV- HC-LC/4.2 kb	36	64	0	100	70% ^b
	LRRK2 SEQ ID NO: 4	N 端 嘌呤霉素-P2A- EmGFP/1.4 kb	100	0	20	80	未测定
	FAK SEQ ID NO: 1	C 端 EmGFP-P2A-嘌呤 霉素/1.4 kb	70	30	30	70	未测定
	EGFR SEQ ID NO: 2	C 端 EmGFP-P2A-嘌呤 霉素/1.4 kb	100	0	17	83	未测定

[0296] a:通过流式细胞术测定;b:通过ELISA分析测定;c:由于嵌合蛋白的表达水平低,所以不能通过流式细胞术确定。

[0297] 在一些实施例中,用硫代磷酸酯或氨基对供体DNA进行末端修饰和/或用非同源末端连接抑制剂(NHEJ)抑制剂处理可以进一步提高HDR的效率。

[0298] 如本文所用,“匹配端”是指具有大于或等于90%的序列一致性的核酸分子的末端。在一些实施例中,5'和3'末端上的匹配端具有12bp至250bp、12 bp至200bp、12bp至150bp、12bp至100bp、12bp至50bp或12bp至40 bp的长度。在一些实施例中,匹配端具有35bp的长度。在一些实施例中,匹配端将共享大于或等于90%、91%、92%、93%、94%、95%、96%、97%或99%或等于100%的序列一致性。目标基因座处的双链断裂的匹配端可以是双链或单链DNA。在一些实施例中,供体核酸分子的匹配端将是单链的。

[0299] 匹配端与核酸在目标基因座处的共有的序列一致性的量越大,则同源重组效率通常越高。当同源区域相当短时(例如50个碱基)时,特别需要高水平的序列一致性。在一些实施例中,目标基因座和匹配端之间的序列一致性的量将大于90%(例如90%至100%、90%至99%、90%至98%、95%至100%、95%至99%、95%至98%、97%至100%等)。

[0300] 如本文所用,“序列一致性百分比”意指通过在比较窗上比较两个最佳对准的核苷酸序列所测定的值,其中相较于参考序列(其不包含添加或缺失),比较窗中的核苷酸序列部分可以包含添加或缺失(即,序列比对空位)以便使两个序列达成最佳的对准。换句话说,出于定量目的,去除了序列对准空位。如下计算序列一致性百分比:测定两个序列中出现相同核酸碱基或氨基酸残基的位置数目以产生匹配位置的数目,将匹配位置的数目除以比较窗中的总位置数目并将结果乘以100,得到序列一致性百分比。

[0301] 测定序列一致性值的一种方法是使用BLAST 2.0程序组,所述程序组使用默认参数(Altschul等人,《核酸研究(Nucleic Acids Res.)》25:3389-3402 (1997))。用于进行BLAST分析的软件可公开获得,例如通过国家生物技术信息中心(National Center for Biotechnology Information)获得。

[0302] 在一些实施例中,末端在与同源重组有关的一个或多个特征上可以不同。例如,与目标基因座序列互补的末端“匹配”区域的长度可以不同。因此,一个末端可以具有四十个

具有序列互补性的核苷酸,而另一个末端可以仅具有十五个具有序列互补性的核苷酸。在一些实施例中,供体核酸分子的一个或两个末端将是部分或完全单链的。

[0303] 如本文所用,“无启动子选择标记物”是指不具有启动子的所关注的外来基因,使得其仅在插入包含启动子的基因组基因座后才表达。在一些实施例中,无启动子选择标记物是蛋白质、抗生素抗性选择标记物、细胞表面标记物、细胞表面蛋白质、代谢产物或其活性片段。在一些实施例中,无启动子选择标记物是标记(例如EmGFP或OFp)。在一个实施例中,无启动子选择标记物是嘌呤霉素、二氢叶酸还原酶或谷氨酰胺合成酶。

[0304] 无启动子选择标记物可以直接与报道基因连接,或替代地,供体核酸分子可以包含充当无启动子选择标记物和报道基因之间的“接头”的另一氨基酸序列。接头可以是多肽或本领域已知的任何其它合适的接头。在一些实施例中,接头包含大于或等于2、3、5、10、15、20、30、40、50、60、70、80或90个氨基酸。在一些实施例中,接头包含100个氨基酸。在一些实施例中,接头包含大于或等于两个选自由甘氨酸、丝氨酸、丙氨酸和苏氨酸组成的群组的氨基酸。在一些实施例中,接头是多甘氨酸接头。在一些实施例中,多甘氨酸接头包含3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个甘氨酸残基。在一个实施例中,接头是6残基的多甘氨酸。在一些实施例中,无启动子选择标记物与报道基因之间的距离小于或等于300nt、240nt、180nt、150nt、120nt、90nt、60nt、30nt、15nt、12nt或9nt。

[0305] 如本文所用,“报道基因”是指其产物容易分析的基因,并且可以作为标记物用于筛选成功修饰的细胞、研究基因表达的调控或用作标准化重组效率的对照。在一些实施例中,报道基因是可选标记物。在一些实施例中,报道基因是荧光报道分子,例如翡翠绿色荧光蛋白(EmGFP)报道体或橙色荧光蛋白(OFp)报道体。在一些实施例中,报道基因是发光报道体,例如荧光素酶(例如北美萤火虫(*P. pyralis* luciferase))。其它常用的报道基因是 β -葡萄糖醛酸苷酶和 β -半乳糖苷酶。理想情况下,报道基因应不存在于研究所用的细胞中,或容易与基因的原生形式区分开来,容易分析,线性检测范围广,并且不影响细胞的正常生理和整体健康。

[0306] 如本文所用,“自裂解肽”是指在翻译时解离为组分蛋白的肽。在一些实施例中,自裂解肽连接无启动子选择标记物和报道基因,并且在重组成初始核酸分子后,能够使无启动子选择标记物与报道基因在翻译期间解离。在一些实施例中,自裂解肽是技术人员已知的自裂解2A肽或其它自裂解肽。

[0307] 在一些实施例中,“loxP”或“X-over P1基因座”位于无启动子选择标记物的任一侧,作为自裂解肽的替代或补充。LoxP可用作Cre-lox重组策略的一部分,以促进无启动子选择标记物的复制。Cre-lox策略需要至少两种组分:1) Cre重组酶,一种催化两个loxP位点之间重组的酶;和2) loxP位点(例如特定的34碱基对序列,其由发生重组的8bp核心序列和两个侧接的13bp反向重复序列组成)或突变型lox位点。(参见例如Araki等人,PNAS 92:160-4(1995);Nagy,A.等人,《起源(Genesis)》26:99-109(2000);Araki等人,《核酸研究(Nuc Acids Res)》30(19):e103(2002);和US20100291626A1,所有文献均以引用的方式并入本文中)。示例性的loxP位点包括(但不限于)野生型lox511、lox5171、lox2272、M2、M3、M7、M11、lox71和lox66。loxP允许在以后能够去除无启动子选择标记物。这样一来,可以选择编辑过的群体,然后去除无启动子选择标记物。如果需要进一步编辑,则此允许再使用无启动子选择标记物。

[0308] 如本文所用,“非同源末端连接”(NHEJ)是指通过很大程度上独立于同源性的过程将双链断裂(DSB)的两端连接的细胞过程。天然存在的DSB是在DNA合成期间(此时复制叉遇到损伤的模板)以及在某些特殊的细胞过程期间(包括V(D)J重组、免疫球蛋白重链(IgH)基因座处发生的类别转换重组,和减数分裂)自发地产生。此外,细胞暴露于电离辐射(X射线和伽马射线)、紫外光、拓扑异构酶毒物或放射模拟药物会产生DSB。视在DSB处产生的特定序列和化学修饰而定,NHEJ可以是精确的或突变诱发的(Lieber,M.R.,《生物化学年鉴》79:181-211(2010))。

[0309] 如本文所用,“非同源末端连接抑制剂”或“NHEJ抑制剂”是指抑制非同源末端连接过程的分子。在一些实施例中,供体核酸分子用至少一种NHEJ抑制剂处理。NHEJ抑制剂的实例包括但不限于DNA依赖性蛋白激酶(DNA-PK)、DNA连接酶IV、DNA聚合酶1或2(PARP-1或PARP-2)或其组合。示例性DNA-PK抑制剂包括Nu7026(2-(4-吗啉基)-4H-萘并[1,2-b]吡喃-4-酮)、Nu7441(8-(4-二苯并噻吩基)-2-(4-吗啉基)-4H-1-苯并吡喃-4-酮)、Ku-0060648(4-乙基-N-[4-[2-(4-吗啉基)-4-氧代-4H-1-苯并吡喃-8-基]-1-二苯并噻吩基]-1-哌嗪乙酰胺)、化合物401(2-(4-吗啉基)-4H-嘧啶并[2,1a]异喹啉-4-酮)、DMNB(4,5-二甲氧基-2-硝基苯甲醛)、ETP 45658(3-[1-甲基-4-(4-吗啉基)-1H-吡唑并[3,4-d]嘧啶-6-基苯酚)、LTURM 34(8-(4-二苯并噻吩基)-2-(4-吗啉基)-4H-1,3-苯并噻嗪-4-酮)、UNC2170(3-溴-N-(3-(叔丁基氨基)丙基)苯甲酰胺)、咖啡碱、Scr7(2,3-二氢-6,7-二苯基-2-硫代-4(1H)-蝶啶酮、6,7-二苯基-2-硫基-二氧四氢蝶啶),和P1 103盐酸盐(3-[4-(4-吗啉基吡啶并[3',2':4,5]呋喃并[3,2-d]嘧啶-2-基]苯酚盐酸盐)。

[0310] 如本文所用,“目标基因座”是指核酸分子内被核酸切割实体识别和裂解的位点。当例如单一CRISPR复合物设计成使双链核酸裂解时,则目标基因座是CRISPR复合物所识别的切割位点和周围区域。当例如两种CRISPR复合物设计成将非常接近的双链核酸切割以产生双链断裂时,则被两种CRISPR复合物所识别的包括断点在内的周围区域称为目标基因座。

[0311] 如本文所用,“核酸酶抗性基团”是指一种化学基团,其可以并入核酸分子中并且能够通过酶(外切核酸酶和/或内切核酸酶)抑制含有所述基团的核酸分子降解。此类基团的实例是硫代磷酸酯基团、氨基团、2'-O-甲基核苷酸、2'-脱氧-2'-氟核苷酸、2'-脱氧核苷酸和5-C-甲基核苷酸。核酸酶抗性基团可以位于供体核酸分子中的许多位置。在一些实施例中,细胞核酸酶将消化供体核酸分子的这个部分。这些核酸酶将被核酸酶抗性基团终止或减慢,从而稳定供体核酸分子的结构。

[0312] 实施例包括包含含有一个或多个(例如一个、两个、三个、四个、五个、六个、七个等)核酸酶抗性基团的核酸分子的组合物,以及制备和使用此类供体核酸分子的方法。在许多实施例中,核酸酶抗性基团将位于供体核酸分子的一个或两个末端。相对于一个或两个末端,供体核酸分子可包含内部基团。在许多实施例中,此类供体核酸分子中的一些或全部将在细胞内加工以产生与DS断裂位点匹配的末端。

[0313] 如本文所用,术语“细胞内靶向部分”是指有助于定位到细胞内位置的化学实体(例如多肽)。细胞内靶向部分的实例包括核定位信号、叶绿体靶向信号和线粒体靶向信号。

[0314] 如本文所用,“受试者”是指人类或非人类动物(例如哺乳动物)或植物。

[0315] 如本文所用,“治疗”是指通过将有效量的无启动子选择标记物施用于受试者来减

轻受试者的疾病、病症或病状的至少一种症状。

[0316] 如本文所用,“核酸切割实体”是指具有核酸切割活性(例如双链核酸切割活性)的一种或多种分子、酶或分子复合物。在大多数实施例中,核酸切割实体组分将是蛋白质或核酸或两者的组合,但是它们可以与辅因子和/或其它分子结合。核酸切割实体通常将基于许多因素来选择,例如在目标基因座处产生DS断裂的效率、在目标基因座处或附近的合适位置处产生DS断裂的能力、在目标基因座处产生DS断裂的低潜在性、低毒性和成本问题。这些因素中的许多因素将随所用细胞和目标基因座而变化。许多核酸切割实体是本领域已知的。例如,在一些实施例中,核酸切割实体包括一种或多种锌指蛋白、转录激活因子样效应物(TALE)、CRISPR复合物(例如Cas9或CPF1)、归巢内切核酸酶或大范围核酸酶、阿格蛋白-核酸复合物,或大范围核酸酶。在一些实施例中,核酸切割实体将具有允许它们被核定位的活性(例如将包含核定位信号(NLS))。在一些实施例中,单链DNA供体可以与切口或切口组合一起发挥作用。

[0317] 如本文所用,术语“难以转染”是指将核酸、蛋白质或类似大分子插入细胞的传统方法不易处理的细胞类型。例如,细胞使用包括但不限于以下的方法的转染效率较低:脂质体介导的转染、磷酸钙转染、阳离子聚合物、FUGENE™、树枝状聚合物等。

[0318] 核酸切割实体

[0319] A. 锌指蛋白(ZFP)

[0320] 如本文所用,“锌指蛋白”(ZFP)是指包含核酸酶域和锌稳定化核酸(例如DNA)结合域的嵌合蛋白。个别DNA结合域通常被称为“指”,使得锌指蛋白或多肽具有至少一个指,更通常两个指或三个指,或甚至四个或五个指,至至少六个或更多个指。在一些实施例中,ZFP将包含三个或四个锌指。每个指通常结合DNA的二至四个碱基对。每个指可包含约30个氨基酸的锌螯合DNA结合区(参见例如美国专利公开第2012/0329067A1号,其公开内容通过引用并入本文)。

[0321] 核酸酶域的一个实例是来自IIs型限制性内切核酸酶FokI的非特异性裂解域(Kim,Y.G.等人,《美国国家科学院院刊》93:1156-60(1996)),其通常被5-7个碱基对的接头序列分隔。通常需要一对FokI裂解域以允许域发生二聚和非回文目标序列从相反的链中裂解。个别Cys₂His₂ZFN的DNA结合域通常包含3至6个个别锌指重复序列并且各自能够识别9至18个碱基对。

[0322] 与ZFP相关的一个问题是脱靶裂解的可能性,脱靶裂解可能引起供体DNA的随机整合或引起染色体重排或甚至细胞死亡,这仍然引起人们对高等生物的适用性的关注(Radecke,S.等人,《分子疗法(Mol. Ther.)》18:743-753(2010))。

[0323] B. 转录激活因子样效应物(TALE)

[0324] 如本文所用,“转录激活因子样效应物”(TALE)是指由超过一个TAL重复序列构成的蛋白质,并且能够以序列特异性方式结合核酸。TALE代表由植物病原性细菌物种(例如黄单胞菌属和青枯病菌属)在感染植物细胞后经由其III型分泌系统分泌的一类DNA结合蛋白。天然TALE尤其已经显示结合至植物启动子序列,借此调节基因表达和激活效应物特异性宿主基因,从而促进细菌繁殖(Römer,P.等人,《科学(Science)》318:645-648(2007); Boch,J.等人,《植物病理学年鉴(Annu. Rev. Phytopathol.)》48:419-436(2010); Kay,S.等人,《科学(Science)》318:648-651(2007); Kay,S.等人,《微生物学新见

(Curr.Opin.Microbiol.)》12:37-43 (2009))。

[0325] 天然TALE的特征通常为中心重复域和羧基端核定位信号序列(NLS)和转录激活域(AD)。除通常较短的羧基端重复序列(称为半重复序列)之外,中心重复域典型地由以下组成:1.5个至33.5个不同量的氨基酸重复序列,其长度通常为33-35个残基。所述重复序列大多数是一致的,但是在某些高变残基上有所不同。TALE的DNA识别特异性是由通常位于每个重复序列的位置 12和13处的高变残基(所谓的重复可变双残基(RVD),其中每个RVD靶向给定的DNA序列中的特定核苷酸)介导。因此,TAL蛋白中的重复序列的顺序倾向于与给定DNA序列中的经定义的核苷酸线性顺序相关。已鉴定出天然存在的一些TALE的潜在RVD代码,从而允许预测为了结合至给定DNA序列而必需的重复杂列顺序(Boch,J.等人,《科学》326:1509-1512 (2009);Moscou, M.J.等人,《科学》326:1501 (2009))。此外,已经表明,以新的重复序列组合产生的TAL效应物结合到根据此代码所预测的目标序列。已经表明,目标DNA 序列通常始于被TAL蛋白识别的5'胸腺嘧啶碱基。

[0326] TAL模块结构允许DNA结合域与效应分子(如核酸酶)组合。具体地说,TALE核酸酶允许开发新的基因组工程工具。

[0327] 一些实施例中使用的TALE可以产生DS断裂,或者可以具有用于产生DS 断裂的组合作用。例如,TAL-FokI核酸酶融合体可以设计成在目标基因座处或附近结合并且通过两个FokI域的结合形成双链核酸切割活性。

[0328] 在一些实施例中,TALE将包含大于或等于6个(例如大于或等于8、10、12、15或17,或6至25、6至35、8至25、10至25、12至25、8至22、10 至22、12至22、6至20、8至20、10至22、12至20、6至18、10至18、12至18个等)TAL重复序列。在一些实施例中,TALE可包含18或24或17.5 或23.5个TAL核酸结合盒。在另外的实施例中,TALE可以包含15.5、16.5、18.5、19.5、20.5、21.5、22.5或24.5个TAL核酸结合盒。TALE通常将具有至少一个多肽区域,所述区域侧接包含TAL重复序列的区域。在许多实施例中,侧接区域将存在于TAL重复序列的氨基端和羧基端。示例性的TALE阐述于美国专利公开第2013/0274129A1号,其公开内容以引入的方式并入本文中,并且可以是伯克霍尔德氏菌属(Burkholderia)、黄单胞菌属(Xanthamonas)和青枯病菌属(Ralstonia)细菌中所发现的天然存在的蛋白质的经修饰形式。

[0329] 在一些实施例中,TALE蛋白质将含有允许其转运到核的核定位信号(NLS)。

[0330] C. 基于CRISPR的系统

[0331] 术语“CRISPR”或“成簇的有规律间隔的短回文重复序列”是一种通用术语,其适用于三种类型的系统和系统亚型。一般来说,术语CRISPR是指编码CRISPR系统组分(例如编码的crRNAs)的重复区域。已鉴定出各自具有不同特征的三种类型的CRISPR系统(参见表2)。

表 2: CRISPR 系统类型概述

系统	特征	实例
I 型	多种蛋白质 (5 至 7 种典型蛋白质)、crRNA 需要 PAM。DNA 裂解是由 Cas3 催化。	表皮葡萄球菌 (<i>Staphylococcus epidermidis</i>) (IA 型)
II 型	3 至 4 种蛋白质 (一种蛋白质 (Cas9) 具有核酸酶活性)、两种 RNA 需要 PAM。目标 DNA 裂解是由 Cas9 和 RNA 组分催化。	酿脓链球菌 (<i>Streptococcus pyogenes</i>) CRISPR/Cas9、新凶手弗朗西丝氏菌 (<i>Francisella novicida</i>) U112 Cpf1
III 型	切割需要五或六种蛋白质,所需 RNA 的种数未知,但预期是 1,不需要 PAM。IIIB 型系统具有靶向 RNA 的能力。	表皮葡萄球菌 (IIIA 型); 强烈火球菌 (<i>P. furiosus</i>) (IIIB 型)。

[0332] 如本文中所用,术语“CRISPR复合物”是指CRISPR蛋白质和核酸(例如RNA)彼此结合而形成具有功能活性的聚集物。CRISPR复合物的一个实例是结合到对目标基因座具有特异性的向导RNA上的野生型Cas9(有时称为 Csn1)蛋白。

[0334] 如本文中所用,术语“CRISPR蛋白质”是指包含核酸(例如RNA)结合域核酸和效应域(例如Cas9,如酿脓链球菌Cas9或CPF1(裂解和多腺苷酸化因子1))的蛋白质。核酸结合域与具有能够和所需目标核酸(例如向导RNA)杂交的区域的第一核酸分子发生相互作用或允许与第二核酸结合,所述第二核酸具有能够与所需目标核酸(例如crRNA)杂交的区域。CRISPR蛋白质还可以包含核酸酶域(即,脱氧核糖核酸酶或核糖核酸酶域)、额外的DNA结合域、解螺旋酶域、蛋白质-蛋白质相互作用域、二聚合域以及其它域。

[0335] CRISPR蛋白还指形成结合上述第一核酸分子的复合物的蛋白质。因此,一种CRISPR蛋白质可以结合到例如向导RNA且另一种蛋白质可以具有内切核酸酶活性。这些都被认为是CRISPR蛋白质,原因是其充当与单一蛋白质(如Cas9或CPF1)执行相同功能的复合物的一部分。

[0336] 在一些实施例中,CRISPR蛋白质将含有允许其转运到核的核定位信号(NLS)。

[0337] 一些实施例中使用的CRISPR可以产生DS断裂或可以具有用于产生DS断裂的组合作用。例如,可以将突变引入CRISPR组分,以阻止CRISPR复合物造成DS断裂,但仍允许这些复合物切割DNA。已经在Cas9蛋白质中鉴定出突变,所述突变允许制备可切割DNA而不是进行双链切割的Cas9蛋白质。因此,一些实施例包括使用RuvC和/或HNH结构域中具有突变的Cas9蛋白质,所述突变将这种蛋白质的核酸酶活性限制为切割活性。

[0338] 如本文提供的术语“dCas9”是指核酸酶灭活的Cas9。在实施例中,DNA结合调节增强剂可以是与dCas9结构域结合的向导RNA。在其它实施例中,调节复合物是结合至gRNA的Cas9结构域,其中所述调节复合物进一步包括可操作地连接至Cas9结构域的VP16转录激活结构域。此类系统可用于诱导哺乳动物细胞中的例如内源基因的表达。本领域普通技术人员将立即认识到,所用DNA结合调节增强剂的类型将根据细胞类型和特定应用而变化。

[0339] 可以使用的CRISPR系统差异很大。这些系统通常具有能够形成包含蛋白质和第一核酸的复合物的功能活性,其中复合物识别第二核酸。CRISPR系统可以是I型、II型或III型系统。适合CRISPR蛋白的非限制性实例包括Cas3、Cas4、Cas5、Cas5e(或CasD)、Cas6、Cas6e、Cas6f、Cas7、Cas8a1、Cas8a2、Cas8b、Cas8c、Cas9、Cas10、Cas10d、CasF、CasG、CasH、Csy1、Csy2、Csy3、Cse1(或CasA)、Cse2(或CasB)、Cse3(或CasE)、Cse4(或CasC)、

Csc1、Csc2、Csa5、Csn2、Csm2、Csm3、Csm4、Csm5、Csm6、Cmr1、Cmr3、Cmr4、Cmr5、Cmr6、Csb1、Csb2、Csb3、Csx17、Csx14、Csx10、Csx16、CsaX、Csx3、Csz1、Csx15、Csf1、Csf2、Csf3、Csf4和Cu1966。

[0340] 在一些实施例中,CRISPR蛋白质(例如Cas9)来源于II型CRISPR系统。在特定实施例中,CRISPR系统经设计可充当寡核苷酸(例如DNA或RNA)引导的来源于Cas9蛋白质的内切核酸酶。用于本文中所阐明的这种和其它功能的Cas9蛋白质可以来源于酿脓链球菌(*Streptococcus pyogenes*)、嗜热链球菌(*Streptococcus thermophilus*)、链球菌种(*Streptococcus sp.*)、达松维尔拟诺卡放线菌(*Nocardiopsis dassonvillei*)、始旋链霉菌(*Streptomyces pristinaespiralis*)、产绿色链霉菌(*Streptomyces viridochromogenes*)、产绿色链霉菌、粉红链孢囊菌(*Streptosporangium roseum*)、酸热脂肪杆菌(*Alicyclobacillus acidocaldarius*)、假蕈状芽孢杆菌(*Bacillus pseudomycooides*)、砷还原芽孢杆菌(*Bacillus selenitireducens*)、西伯利亚微小杆菌(*Exiguobacterium sibiricum*)、戴白氏乳杆菌(*Lactobacillus delbrueckii*)、唾液乳杆菌(*Lactobacillus salivarius*)、海洋微颤菌(*Microscilla marina*)、伯克氏菌目细菌(*Burkholderiales bacterium*)、萘沃喃极单胞菌(*Polaromonas naphthalenivorans*)、极单胞菌种(*Polaromonas sp.*)、瓦氏鳄球藻(*Crocospaera watsonii*)、蓝丝菌种(*Cyanothece sp.*)、铜绿微囊藻(*Microcystis aeruginosa*)、聚球藻种(*Synechococcus sp.*)、阿拉伯糖醋杆菌(*Acetohalobium arabaticum*)、得津西氨热杆菌(*Ammonifex degensii*)、贝茨热解纤维素菌(*Caldicelulosiruptor*)、金矿菌(*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*)、嗜热丙酸氧化型佩洛菌(*Pelotomaculum thermopropionicum*)、喜温嗜酸硫杆菌(*Acidithiobacillus caldus*)、氧化亚铁嗜酸硫杆菌(*Acidithiobacillus ferrooxidans*)、酒色等着色菌(*Allochromatium vinosum*)、海杆菌种(*Marinobacter sp.*)、嗜盐片球亚硝化球菌(*Nitrosococcus halophilus*)、瓦生氏亚硝化球菌(*Nitrosococcus watsoni*)、哈河豚毒素假交替单胞菌(*Pseudoalteromonas haloplanktis*)、具总状花纤线杆菌(*Ktedonobacter racemifer*)、爱维斯特甲烷盐菌(*Methanohalobium evestigatum*)、多变念珠藻(*Anabaena variabilis*)、泡沫节球藻(*Nodularia spumigena*)、念珠藻种(*Nostoc sp.*)、极大节旋藻(*Arthrospira maxima*)、钝顶节旋藻(*Arthrospira platensis*)、节旋藻种(*Arthrospira sp.*)、螺旋藻种(*Lyngbya sp.*)、原型微鞘藻(*Microcoleus chthonoplastes*)、颤藻种(*Oscillatoria sp.*)、运动石袍菌(*Petrotoga mobilis*)、非洲高热杆菌(*Thermosiphon africanus*)、或深海藻青菌(*Acaryochloris marina*)。

[0341] D.阿格蛋白基因编辑系统

[0342] 阿格蛋白家族是内切核酸酶,其使用5'磷酸化单链核酸作为向导来裂解核酸目标。这些蛋白质,如Cas9,据信在基因表达抑制和防御外源核酸方面具有作用。

[0343] 阿格蛋白在许多方面与Cas9不同。与仅存在于原核生物中的Cas9不同,阿格蛋白在进化上是保守的并且存在于几乎所有的生物体中。已经发现一些阿格蛋白结合单链DNA并且裂解目标DNA分子。此外,阿格蛋白结合不需要特定的共同二级指导结构,并且不需要

如CRISPR系统PAM位点的序列。已经显示,可以用单链DNA向导对格氏嗜盐碱杆菌的阿格蛋白进行编程,并在哺乳动物细胞中用作基因组编辑(Gao,F.等人,《自然·生物技术(Nat. Biotechnol.)》34:768-73(2016))。

[0344] 阿格蛋白需要5'磷酸化单链向导DNA分子,其长度约为24个核苷酸。参见例如表12中SEQ ID NO:6的阿格蛋白氨基酸序列。

[0345] 将材料引入细胞

[0346] 将各种分子引入细胞中可以通过多种方法进行,包括许多标准实验室手册中所述的方法,如Davis,L.等人,《分子生物学基本方法(Basic Methods in Molecular Biology)》(1986)和Sambrook,J.等人,《分子克隆:实验室手册(Molecular Cloning:A laboratory manual)》第1卷,第2版,纽约冷泉港实验室出版社(Cold Spring Harbour Lab.Press,N.Y.) (1989)。实例包括但不限于磷酸钙转染、DEAE-聚葡萄糖介导式转染、转染、微注射、阳离子脂质介导式转染、电穿孔、转导、刮擦加载、弹道式引入、核穿孔、流体动力学冲击和感染。

[0347] 核酸切割实体和/或供体核酸分子的不同组分可以通过不同方式引入细胞。在一些实施例中,可以将单一类型的核酸切割实体分子引入细胞中,但是一些核酸切割实体分子可以在细胞内表达。一个实例是使用两个锌指-FokI融合体在细胞内核酸中产生双链断裂。在某些情况下,仅一种锌指-Fok I融合体可以被引入细胞,而另一种锌指-Fok I融合体可以在细胞内产生。

[0348] 适合的转染剂包括促进RNA、DNA和蛋白质引入细胞中的转染剂。示例性转染试剂包括但不限于TurboFect转染试剂(赛默飞世尔科技)、Pro-Ject试剂(赛默飞世尔科技)、TRANSPASSTM蛋白质转染试剂(新英格兰生物实验室(New England Biolabs))、CHARIOTTM蛋白质递送试剂(Active Motif)、PROTEOJUICETM蛋白质转染试剂(EMD Millipore)、293fectin、LIPOFECTAMINETM 2000、LIPOFECTAMINETM3000(赛默飞世尔科技)、LIPOFECTAMINETM(赛默飞世尔科技)、LIPOFECTINTM(赛默飞世尔科技)、DMRIE-C、CELLFECTINTM(赛默飞世尔科技)、OligofectamineTM(赛默飞世尔科技)、LIPOFECTACETM、FugeneTM(瑞士巴塞尔罗氏(Roche,Basel,Switzerland))、FugeneTMHD(罗氏)、TransfectamTM(Transfectam,Promega,Madison,Wis.)、Tfx-10TM(Promega)、Tfx-20TM(Promega)、Tfx-50TM(Promega)、TransfectinTM(BioRad,Hercules, Calif.)、SilentFectTM(Bio-Rad)、EffecteneTM(Qiagen,Valencia,Calif.)、DC-chol (Avanti Polar Lipids)、GENEPORTERTM(基因治疗系统(Gene Therapy Systems), San Diego,Calif.)、DHARMAFECT 1TM(Dharmacon,Lafayette,Colo.)、DHARMAFECT 2TM(Dharmacon)、DHARMAFECT 3TM(Dharmacon)、DHARMAFECT 4TM(Dharmacon)、ESCORTTMIII(Sigma,St.Louis,Mo.)和EscortTMIV(Sigma 化学公司)。

[0349] 本文所述的组合物和方法可用于高通量筛选方法中。此类方法的一个实例是反向转染。为了说明的目的,假定已经产生了gRNA分子和相应的NLS结合供体DNA分子的文库。进一步假设每个文库组合物包含(1)与细胞基因组中特定基因座具有序列同源性的gRNA分子,以及(2)具有与侧接预期基因组裂解位点同源的区域的NLS结合供体DNA分子。还假定已经产生了三百种此类文库组合物,并且这些组合物中的每一种都点加在载玻片上的单独位置。最后,在允许(1)吸收文库组合物和(2)基因编辑发生在gRNA特定目标基因座的条件下,

将表达Cas9蛋白的293FT细胞系覆盖在载玻片上。当然,此类方法的许多变化形式是可能的,包括其中所用基因编辑试剂不同(例如 TAL-FokI mRNA而非gRNA)和阵列形式不同(例如96孔板的孔而非载玻片的表面)的变化形式。

[0350] 因此,一个方面包括基因编辑试剂(例如gRNA、TAL mRNA、供体核酸分子等)的文库和用于修饰细胞中各种目标基因座的高通量方法。

[0351] 核酸定位和基因编辑效率

[0352] 一个方面还包括用于提高基因编辑效率的组合物和方法。在一些实施例中,此类组合物和方法涉及与一个或多个细胞内靶向部分连接的核酸分子,所述细胞内靶向部分将核酸分子定位在需要基因编辑的细胞内位置(例如细胞核、线粒体、叶绿体等)。一些实施例将采用细胞内靶向部分促进核酸分子在一个或多个细胞内位置处的局部浓度增加。尽管不希望受到理论的束缚,但据信,增加的基因编辑效率是由于希望进行基因编辑的位置处的供体核酸浓度增加而引起的。

[0353] 在图9中示出了一个实施例此图显示了通过两个不同的接头与单链供体 DNA分子连接的核定位信号(NLS)(细胞内靶向部分的一个实例)。这种类型的构建体可用于促进核酸递送到核。这种构建体的许多变化形式是可能的。

[0354] 如图11和13所示且如以下核定位实例中讨论,已经发现如图9所示的那些构建体能够显著提高细胞内基因编辑的效率并允许使用较少的供体核酸。具体地说,以上提到的数据证明,使用NLS修饰的供体DNA可以提高在核酸切割位点(例如用gRNA/Cas9切割的染色体基因座)进行基因组工程的效率。

[0355] 图11和13中的数据显示了接近80%的基因编辑效率。此外,发现每约 2×10^5 个细胞的NLS-供体DNA结合物为0.03皮摩尔那么少。因此,实施例包括用于细胞内基因工程的组合物和方法,其中至少75%(例如50%至75%、50%至80%、50%至85%、50%至95%、60%至95%、70%至95%、70%至90%、75%至90%等)细胞中的特定目标基因座被修饰。另外,一些实施例包括组合物和方法,其中当细胞与每 2×10^5 个细胞0.3皮摩尔或更少(例如0.001到0.3、0.005到0.3、0.01到0.3、0.05到0.3、0.001到0.2、0.005到0.2、0.001到0.15、0.001到0.1皮摩尔等)供体DNA接触时,反应混合物中至少50%的经转染细胞在目标基因座被修饰。此假定100%转染。例如,在50%转染的情况下,将获得细胞总数的约一半和经转染细胞的至少75%。

[0356] 一些实施例还涉及用于在存在需要改变的细胞内核酸分子的位置(例如核)增加供体核酸的局部浓度的组合物和方法。一些实施例包括使用细胞内靶向部分增加细胞内位置上的浓度的组合物和方法,其中局部核酸浓度的增加量是不使用细胞内靶向部分时的至少10倍(10至1,000、10至800、10至600、10至1,000、10至1,000、10至1,000、10至400、50至1,000、50至600、100至1,000、100至700倍等)。例如,可以使用例如荧光标记的核酸分子来测量核酸分子的细胞内定位的增加倍数。为了说明,NLS结合和未结合的核酸分子都可以用于在此类分析中进行比较。

[0357] 图9中所示构建体的一种变化形式是其中NLS定位于供体DNA分子的3'端(而非5'端)、两端,置于供体DNA分子的中部等。另外,一端或两端可以存在超过一个NLS。此外,核酸分子可以是:(1) DNA或RNA;(2) 单链或双链;(3) 线性的、圆形的或具有茎或发夹环的分子;和/或(4) 经过化学修饰(例如包含硫代磷酸酯键联、2'-O-甲基碱基等)。另外,如下文所述,

NLS 可以用细胞内靶向部分置换,所述细胞内靶向部分将定位引向除核之外的细胞空间(例如线粒体、叶绿体等)。因此,一些实施例包括可操作地连接至定位于需要基因编辑的细胞内位置的一个或多个细胞内靶向部分的核酸分子,以及使用此类核酸分子的方法(例如用于基因组工程)。表3列出了可以在一些实施例中使用的某些示例性细胞内靶向部分的氨基酸序列。

[0358] 表3:示例性的亚细胞/细胞器定位序列

	核定位信号	SEQ ID
	PKKKRKV	98
	AVKRPAATKKAGQAKKKKLD	99
	MSRRRKANPTKLSNAKKLAKEVEN	100
[0359]	PAAKRVKLD	101
	叶绿体靶向信号	
	LIAHPQAFPGAIAAPISYAYAVKGRKPRFQTAKGSVRI	102
	线粒体靶向信号	
	MLSLRQSIRFFKPATRTLCSRYLL	103

[0360] 另外,在使用细胞内靶向部分(例如多肽)的许多情况下,这些靶向部分可以设计成使核酸分子定位于预期存在细胞内核酸的位置(例如核、叶绿体的基质、线粒体的基质等)。换句话说,在许多情况下,可能希望将核酸定位到细胞内的特定亚空间。一些实施例包括用于将供体核酸分子定位在细胞内的位置,以及用于增强基因组工程反应在供体核酸分子所定位的细胞内的位置处的效率的组合物和方法。

[0361] 可以使用多种方法将细胞内靶向部分连接至核酸分子。实例中所述的两种方法是琥珀酰亚氨基4-(N-马来酰亚氨基甲基)环己烷-1-甲酸酯(SMCC)接头和 Click-iT®系统(赛默飞世尔科技)。无论如何,用于将细胞内靶向部分与核酸分子连接的接头通常具有某些特征,其中一些特征是(1)低细胞毒性,(2)促进细胞吸收或至少对细胞吸收的干扰水平低;以及(3)低分子量(mwt)(例如小于500mwt)。细胞内靶向部分与核酸分子的连接可以通过PCR扩增、以 NLS结合的DNA寡核苷酸为引物来进行。另外,NLS结合的DNA寡核苷酸可以用作通用引物,其中核酸部分连接至基因特定区域用于PCR扩增。此外,可以将NLS结合的DNA寡核苷酸与具有单链突出物的单链DNA供体或双链 DNA供体粘接而非共价结合细胞内靶向部分,并且然后将供体带入细胞内区室。

[0362] 结合物中的核酸分子组分的大小、类型和其它特征经常随应用而变化,其中SNP变化通常短于编码区插入。此外,与内源核酸同源的区域(当存在时)长度也将随应用而变化。结合物中的核酸分子组分(例如供体DNA)的长度可以是1至2000个或更多个(例如小于或等于1500、1000、750、500、300、250、200、150、100、75、50、40、35、30、25、20、15、10、5、4、3、2或1个)核苷酸或碱基对(取决于它们是单链还是双链)。在一些实施例中,核酸分子组分的长度为1至500、10至400、20至300、30至250、30至200或30至100个等核苷酸或碱基对。

[0363] 还包括包含基因编辑蛋白(例如Cas9蛋白、TAL蛋白等)的组合物和方法,其中所述基因编辑蛋白可操作地连接至一个或多个细胞内靶向部分,所述细胞内靶向部分能够将供体核酸分子定位于细胞中的内源核酸分子所处的位置,一个或多个细胞内靶向部分与基因编辑蛋白相关。

[0364] 可以使用的核定位信号可以具有各种结构,并且可以是例如单分型或双分型。单分型NLS通常由单簇基本残基组成。双分型NLS通常由相隔10-12个残基的两簇基本残基组

成。示例性的NLS氨基酸序列在下表4、5和6中列出。

[0365]	表 4: 基于经典的示例性 NLS 氨基酸序列	
	单分型	
	SV40	<u>PKKKRKV</u>
	Myc	<u>PAAKRVKLD</u>
	双分型	
	核质蛋白	<u>KRPAATKKAGQAKKKKL</u>
	BP-SV40	<u>KRTADGSEFESPKKKRKVE</u>
	BP-SV40A4	<u>KRTADGSEFESPKKARKVE</u>
	BP-SV40A5	<u>KRTADGSEFESPKKKAKVE</u>
	BP-SV40C	KRN ₅₋₁₅ KKN ₁ N ₂ KV

具有下划线的氨基酸序列据信与输入蛋白 α 相关, 接头序列用斜体表示。BP-SV40C 是 BP-SV40 双分型 NLS 共同序列。在某些情况下, N₁ 可以是赖氨酸或丙氨酸, 而 N₂ 可以是赖氨酸、精氨酸或丙氨酸。

[0366]	表 5: 示例性非经典/其它 NLS 氨基酸序列	
	ASPEYVNLPINGNG	SEQ ID No.
	LSPSLSPL	103
	MVQLRPRASR	104
	PPARRRRL	105
	TLSPASSPSSVSCPVIPASTDESPGSALNI	106

[0367]	表 6: 其它 NLS 序列和共同序列	
	非天然存在的 NLS	
	BP-Myc 变异体	<u>KRTADGSEFE</u> <u>PAAKRVKLDE</u>
	BP-SV40 变异体	<u>KRTADGSEFES</u> <u>PKKKRKVE</u>
	天然存在的 NLS	
	双分型 NLS1	SAA <u>KR</u> NSATVHLCPV <u>PRK</u> RS
	双分型 NLS2	AAA <u>KR</u> PADDDDNASPA <u>KRR</u> SG
	双分型 NLS3	SAA <u>KR</u> PSATVHLCDVPT <u>KKTK</u> RS
	共同序列	
	共同序列 1	KRX ₍₁₀₋₁₂₎ K(K/R)(K/R) _n
	共同序列 2	KRX ₍₁₀₋₁₂₎ K(K/R)X(K/R) _n
	共同序列 3	KRX ₍₅₋₁₅₎ K(K/R)(K/R) ₁₋₂
	共同序列 4	KRX ₍₅₋₁₅₎ K(K/R)X(K/R) ₁₋₂

X = 任何氨基酸, n = 整数, 被认为是保守的序列用下划线标出, 接头序列用斜体表示

[0368] 图39显示了可操作地连接至NLS的Cas9蛋白的一系列示意图。可以使用任何数量和类型的NLS, 并且它们在蛋白质或核酸分子中的位置将随NLS 所连接的分子的特异性和预期目的而变化。对于例如Cas9或TAL效应物等蛋白质来说, 通常需要将大量蛋白质引入细胞中, 随后使蛋白质在细胞质中保持相对短的时间段, 且使大部分蛋白质定位于核。之所以如此, 是因为据信蛋白质在细胞质中保留的时间越长, 蛋白质的降解程度就越高。还相信, 核中Cas9 的浓度越高, 切割效率就越高(当然假设所有Cas9都具有切割活性(例如与 gRNA 有关))。因此, “收集”于核中的与NLS相关的蛋白质(以及其它分子) 的数量基于: (1) 引入细胞的蛋白质数量和 (2) 蛋白质进行的速率到细胞核。

[0369] 图45是显示共同TALE结构形式的示意图。在许多情况下, TALE是通过刺激存在于细胞内特定位置(例如细胞核、线粒体、叶绿体等)的DNA而起作用。在许多情况下, 涉及DNA识别和结合的TALE蛋白区域的中断将导致DNA识别和/或结合活性降低或消除。位点1、2和3

位于被认为与DNA识别和结合有关的TALE区域之外。因此,当需要高水平的目标DNA结合时,这些是NLS定位的适合位点。

[0370] 使用图45作为参考点,NLS定位于位点1可能发生在氨基酸25左侧(N端方向)的任何位置。NLS定位于位点2和3可能发生在氨基酸814右侧(N端方向)的任何位置。这包括如下情况,其中包括天然存在的较长TALE蛋白区域,其超出图45左侧的氨基酸25和右侧的氨基酸814。此外,位点3位于效应域的C端。

[0371] 一个或多个NLS可以位于位点1、2和/或3中的一个或多个。此外,当(这些位点中的一个或多个)存在多个NLS时,它们可以是相同类型或不同类型。

[0372] 在许多情况下,NLS的位置(例如图45中的位点1、2和/或3)和类型将以引起以下的方式选定:(1)基因编辑试剂定位于细胞核的水平高和/或(2)细胞核中的功能活性水平高。这两种影响通常与核功能活性通常低于核定位的数量有关。之所以如此,是因为在许多情况下,并非所有进入细胞核的基因编辑试剂都会结合到它们具有特异性的目标基因座,而那些确实结合的基因编辑试剂可能并非总是以其被设计的方式(例如核酸裂解、转录激活等)作用于目标基因座核酸。就此而言,一个示例性的原因是,一种细胞类型中的细胞核的核酸比另一种细胞类型更易企及。此外,即使在细胞类型相同的细胞内,也存在变异,所述变异可以使得群体中的一种细胞的目标基因座比相同群体中的另一种细胞更易或不易企及。

[0373] 核定位分析和功能分析(例如基因组裂解检测)均在本文其它地方阐述。为了校正目标基因座和细胞类型的差异,比较性分析中通常将使用相同的目标基因座和细胞类型。

[0374] 此外,基因编辑效率将经常随所编辑的基因座和细胞类型而变化。这是由于多种因素引起的,包括基因编辑试剂对目标基因座的可及性以及细胞类型在同源定向修复(HDR)方面的效率。就HDR而言,具有较高HDR效率的细胞(例如293FT和U2OS细胞)展现的基因编辑率通常高于HDR效率水平较低的细胞(例如A549细胞)。

[0375] 关于NLS的位置和类型,TALE蛋白的多种形式如下表7所示。

表 7: 示例性 TALE/NLS 形式

TALE 区域				
编号	N 端 (SEQ ID NO)	重复 区域	C 端 - 效应 域 (SEQ ID NO)	效应域的 C 端 (SEQ ID NO)
1	XXKRPAATKKAGQAKKKK (116)			
2	PKKKRVD (98)			PKKKRVD (98)
3			PKKKRVD (98)	PKKKRVD (98)
4	PKKKRVD (98)		PKKKRV (98)	PKKKRVD (98)
5	PKKKRVD X ₁₋₂₀ PKKKRVD (117)			
6	XXKRPAATKKAGQAKKKK (116)			XXKRPAATKKAGQAKKKK (116)
7	一个 NLS		一个 NLS	一个 NLS
8				两个 NLS
X = 任一种氨基酸				
X ₁₋₂₀ = 1 至 20 个任一种氨基酸				

[0376] 表7中列出的示例性TALE/NLS形式在TALE蛋白内的NLS类型和NLS 位置方面有所不同。在一些情况下,TALE蛋白将含有约1至约15(例如约2 至约14、约3至约14、约4至约14、约2至约10、约2至约8、约2至约6、约3至约5、约3至约4等)个NLS。此外,当TALE蛋白中存在多个NLS 时,这些NLS可以是单部分或二部分的。

[0378] 另外,两个或更多个NLS可以位于TALE蛋白的相同区域(例如N端区域等)中。当超过一个NLS位于TALE蛋白(例如TALE蛋白的同一区域内) 时,这些NLS中的两个或多个彼此可以位于约1至约50个(例如约2至约 50、约3至约50、约5至约50、约10至约50、约15至约50、约2至约30、约5至约50、约5至约25个等)氨基酸内。在一些情况下,两个NLS可以彼此相隔两个氨基酸。此外,这些氨基酸可以属于旨在形成柔性接头(例如Gly- Gly、Gly-Ser等)的类型。

[0379] 表7列出了关于NLS的区域位置和氨基酸序列的六种特定TALE/NLS形式。进一步列出了两种更通用的TALE/NLS形式。任何数量的此类形式都是可能的。举例来说,TALE蛋白的各个区域可以独立地包含约1至约5个NLS。

[0380] 图46和47显示了NLS处于不同位置的两种不同TALEN蛋白的氨基酸序列。Cas9和TALEN蛋白的存在于分子中的NLS的数量、类型和位置都可能不同。关于图46所示的氨基酸序列,当NLS相对于重复区域位于N端时, NLS的位置距离N端比距离R-3区域更远。另外,当NLS相对于重复区域位于C端时,NLS的位置比氨基酸H R V A(图46中的氨基酸811-814) 通常更靠近C端,处于重复区域与效应域(图46所示的氨基酸序列中的FokI) 之间。因此,使用图46中的氨基酸序列进行参考,NLS可以位于三个通用位置:(1) 重复区域的N端;(2) 重复区域与效应域之间,和(3) 效应域之后。

[0381] 在一些情况下,使用图46中所示的氨基酸序列进行参考,一个或多个NLS 可以位于氨基酸768到814的区域中。举例来说,在图45中,一个或多个NLS 可以紧邻于以下氨基酸中的一个或多个之后:768、777、779、788和/或789。

[0382] 具体地说,图46尤其显示了位于重复区域N末端左侧的TAL蛋白区域(图46中的氨基酸18到153)。此TAL蛋白区域在黄单胞菌属菌种中通常是保守的,在氨基酸水平上具有超过90%一致性。此外,此区域包含四个区域(R0、R-1、R-2和R-3),这些区域与TAL重复序列具有一些序列同源性。NLS通常位于此区域之外,并且位置通常将更靠近TAL蛋白的N末端。

[0383] 另外,图46中所示的氨基酸序列在重复区域的N端仅含有例如153个氨基酸。此N端区域可以具有不同长度并且长度可以是例如约140至约400个(例如约150至约350、约150至约300、约150至约250、约150至约200、约180至约350、约185至约300、约200至约350、约200至约300个等)氨基酸。

[0384] 另外,再次使用图46中所示的氨基酸序列,氨基酸重复区域C端的区域在黄单胞菌属菌种中通常也是保守的。同样,NLS通常将位于此区域之外,并且位置通常将更靠近TAL蛋白的C末端。

[0385] 根据基因编辑分子(例如TAL蛋白、CRISPR蛋白、gRNA等)的所需细胞内水平和基因编辑活性的所需持续时间,可以将基因编辑分子作为RNA/mRNA或通过编码RNA或蛋白质基因编辑试剂的DNA引入细胞中。此外,当编码基因编辑分子的核酸位于细胞中时,编码区通常将可操作地连接至表达控制序列,例如启动子(例如组成型启动子、诱导型启动子、抑制型启动子等)。

[0386] 本文提供了各种形式的TAL蛋白(以及其它基因编辑分子)、编码这些蛋白质的核酸分子以及使用这些蛋白质修饰细胞基因组的方法。

[0387] 用于测量蛋白质和其它分子的核吸收的分析是已知的。此类分析可以基于核中的功能活性的测量结果(例如实例1中所述的GCD分析)。其它分析直接测量分子吸收并且包括基于荧光的分析。此类分析通常要求被测分子显示荧光。荧光可以天然存在于分子中,或者荧光可以由与荧光分子(例如GFP、OFP、化学标记(例如染料)等)结合而产生。

[0388] Wu等人,《生物物理学杂志(Biophysical Journal)》,96:3840-3849(2009)中提出了一种合适的分析,其中使用了双光子荧光对应显微镜来测量核输入。这些方法基于通过显微镜测量细胞质和细胞核内部多个点的平均荧光强度,然后确定比率。尽管某些基因编辑试剂可以被截留于膜和内体中,但可以将细胞质荧光水平与核酸酶荧光水平进行比较,以在一个或多个时间点确定基因编辑试剂进入细胞核的速率以及存在于细胞核中的基因编辑试剂的量。

[0389] 如Wu等人阐述的方法可以用于测量荧光标记的基因编辑试剂的基于吸收和位置的浓度。一种示例性方法是其中使用双光子荧光对应显微镜来测量Cas9蛋白的核定位。在方法的此图示中,将一系列不同的Cas9-NLS-GFP融合蛋白/gRNA复合物引入细胞系,并且对细胞进行50点荧光测量,一半在细胞质中且一半在细胞核中。然后测定每种Cas9-NLS-GFP融合蛋白/gRNA复合物的稳态细胞核对细胞质比率。本文提供允许产生细胞的组合物和方法,其中细胞内的基因编辑试剂的细胞核对细胞质比率平均为约5至约120(例如约5至约100、约15至约100、约20至约100、约25至约100、约30至约100、约35至约100、约40至约100、约50至约100、约60至约100、约70至约100、约40至约120、约50至约120等)。

[0390] 本文亦提供允许产生细胞群的组合物和方法,其中就二倍体细胞而言,群体成员的两个目标基因座中的至少一个出现至少90%裂解(例如约90%至约100%、约90%至约98%、约90%至约96%、约93%至约100%、约95%至约100%、约92%至约96%等)。在一些

情况下,当调整条件使得50,000 (+/-10%) 与约0.5至约200ng (例如约0.5至约150、约0.5至约100、约0.5至约90、约0.5至约75、约1至约200、约1.5至约200、约3至约200、约1至约50 ng、约10至约45、约12至约60ng等) Cas9/gRNA复合物在实例7所述的条件下接触时,上述裂解百分比将适用。

[0391] 采用细胞内靶向部分的组合物和方法可以通过许多方法用于改变内源核酸分子。例如,可以使用这些组合物和方法促进在内源核酸“完整”的位置处进行同源重组。此意味着内源核酸尚未被编辑基因试剂(例如CRISPR、TAL、锌指-FokI融合体等)切割。然而,在某些情况下,遗传改变位点将被切割或具有双链断裂。

[0392] 使用NHEJ抑制剂进行基因编辑的方法

[0393] 本公开部分涉及NHEJ抑制剂改善难以转染的细胞中的基因编辑(例如敲入效率)的用途。

[0394] 在一个方面中,提供了一种遗传改变细胞的方法。所述方法包括(i)使细胞与核酸切割实体和供体核酸(例如供体DNA)接触;(ii)在允许核酸切割实体和供体核酸(例如供体DNA)被细胞吸收的条件下将细胞电穿孔;(iii)在非同源末端连接(NHEJ)抑制剂存在下培养细胞,从而形成遗传改变的细胞。

[0395] 在一个方面中,提供了一种遗传改变细胞的方法。所述方法包括(i)使细胞与核酸切割实体和供体核酸(例如供体DNA)接触;(ii)在允许核酸切割实体和供体核酸(例如供体DNA)被细胞吸收的条件下将细胞电穿孔;和(iii)在非同源末端连接(NHEJ)抑制剂存在下培养细胞,从而形成遗传改变的细胞;其中所述细胞是干细胞、免疫细胞或原代细胞。

[0396] 在一个方面中,提供了一种遗传改变细胞的方法。所述方法包括(i)使细胞与核酸切割实体和供体核酸(例如供体DNA)接触;(ii)在允许核酸切割实体和供体核酸辅助物(例如供体DNA)被细胞吸收的条件下将细胞电穿孔;和(iii)在非同源末端连接(NHEJ)抑制剂存在下培养细胞,从而形成遗传改变的细胞;其中细胞悬浮培养生长。

[0397] 在一个实施例中,核酸切割实体是锌指蛋白、转录激活因子样效应物(TALE)、CRISPR复合物、阿格蛋白-核酸复合物、兆碱基大范围核酸酶或大范围核酸酶。在一个实施例中,核酸切割实体包含转录激活因子样效应物(TALE)。在一个实施例中,核酸切割实体包含CRISPR复合物。

[0398] 在一个实施例中,细胞是原代细胞。

[0399] 在一个实施例中,细胞是干细胞。在一个实施例中,干细胞是诱导型多能干细胞(iPSC)。

[0400] 在一个实施例中,细胞是免疫细胞。在一个实施例中,免疫细胞是T细胞、天然杀伤(NK)细胞、树突状细胞、B细胞、粒细胞、单核细胞、肥大细胞或嗜中性粒细胞。在一个实施例中,免疫细胞是T细胞。

[0401] 在一个实施例中,NHEJ抑制剂是DNA依赖性蛋白激酶(DNA-PK)抑制剂、DNA连接酶IV抑制剂或其组合。在一个实施例中,NHEJ抑制剂是Nu7026 (2-(4-吗啉基)-4H-萘并[1,2-b]吡喃-4-酮)、Nu7441 (8-(4-二苯并噻吩基)-2-(4-吗啉基)-4H-1-苯并吡喃-4-酮)、Ku-0060648 (4-乙基-N-[4-[2-(4-吗啉基)-4-氧代-4H-1-苯并吡喃-8-基]-1-二苯并噻吩基]-1-哌嗪乙酰胺)、化合物401 (2-(4-吗啉基)-4H-嘧啶并[2,1-a]异喹啉-4-酮)、DMNB (4,5-二甲氧基-2-硝基苯甲醛)、ETP 45658 (3-[1-甲基-4-(4-吗啉基)-1H-吡唑并[3,4-d]噻

啶-6-基苯酚)、LTURM 34 (8- (4-二苯并噻吩基)-2- (4-吗啉基)-4H-1,3-苯并噁嗪-4-酮)、UNC2170 (3-溴-N- (3- (叔丁基氨基) 丙基) 苯甲酰胺)、Scr7 (2,3-二氢-6,7-二苯基-2-硫代-4 (1H)-蝶啶酮、6,7-二苯基-2-硫基-二氧四氢蝶啶)、咖啡碱,和/或P1 103盐酸盐 (3-[4- (4-吗啉基吡啶并[3',2':4,5]呋喃并[3,2-d]嘧啶-2-基] 苯酚盐酸盐),或其任何组合。

[0402] 在一个实施例中,DNA-PK抑制剂是Nu7026、Ku0060648和/或Nu7441。

[0403] 在一个实施例中,培养基中的Nu7026浓度在1 μ M与100 μ M之间。在一个实施例中,培养基中的Nu7026浓度在1 μ M与90 μ M之间。在一个实施例中,培养基中的Nu7026浓度在1 μ M与80 μ M之间。在一个实施例中,培养基中的Nu7026浓度在1 μ M与70 μ M之间。在一个实施例中,培养基中的Nu7026浓度在1 μ M与60 μ M之间。在一个实施例中,培养基中的Nu7026浓度在1 μ M与50 μ M之间。在一个实施例中,培养基中的Nu7026浓度在1 μ M与40 μ M之间。在一个实施例中,培养基中的Nu7026浓度在1 μ M与30 μ M之间。在一个实施例中,培养基中的Nu7026浓度在1 μ M与20 μ M之间。在一个实施例中,培养基中的Nu7026浓度在10 μ M与100 μ M之间。在一个实施例中,培养基中的Nu7026浓度在20 μ M与100 μ M之间。在一个实施例中,培养基中的Nu7026浓度在30 μ M与100 μ M之间。在一个实施例中,培养基中的Nu7026浓度在40 μ M与100 μ M之间。在一个实施例中,培养基中的Nu7026浓度在50 μ M与100 μ M之间。在一个实施例中,培养基中的Nu7026浓度在5 μ M与90 μ M之间。在一个实施例中,培养基中的Nu7026浓度在5 μ M与80 μ M之间。在一个实施例中,培养基中的Nu7026浓度在5 μ M与70 μ M之间。在一个实施例中,培养基中的Nu7026浓度在5 μ M与60 μ M之间。在一个实施例中,培养基中的Nu7026浓度在10 μ M与90 μ M之间。在一个实施例中,培养基中的Nu7026浓度在10 μ M与80 μ M之间。在一个实施例中,培养基中的Nu7026浓度在10 μ M与70 μ M之间。在一个实施例中,培养基中的Nu7026浓度在10 μ M与60 μ M之间。在一个实施例中,培养基中的Nu7026浓度是约5 μ M、约10 μ M、约20 μ M、约30 μ M、约40 μ M、约50 μ M、约60 μ M、约70 μ M、约80 μ M、约90 μ M或约100 μ M。

[0404] 在一个实施例中,培养基中的Ku0060648浓度在10nM与2000nM之间。在一个实施例中,培养基中的Ku0060648浓度在10nM与1000nM之间。在一个实施例中,培养基中的Ku0060648浓度在10nM与500nM之间。在一个实施例中,培养基中的Ku0060648浓度在10nM与400nM之间。在一个实施例中,培养基中的Ku0060648浓度在10nM与300nM之间。在一个实施例中,培养基中的Ku0060648浓度在10nM与200nM之间。在一个实施例中,培养基中的Ku0060648浓度在10nM与100nM之间。在一个实施例中,培养基中的Ku0060648浓度在100nM与1000nM之间。在一个实施例中,培养基中的Ku0060648浓度在100nM与500nM之间。在一个实施例中,培养基中的Ku0060648浓度在100nM与400nM之间。在一个实施例中,培养基中的Ku0060648浓度在100nM与300nM之间。在一个实施例中,培养基中的Ku0060648浓度是约100nM、约125nM、约150nM、约200nM、约250nM、约400nM、约500nM或约1000nM。

[0405] 在一个实施例中,培养基中的Nu7441浓度在0.1 μ M与10 μ M之间。在一个实施例中,培养基中的Nu7441浓度在0.2 μ M与10 μ M之间。在一个实施例中,培养基中的Nu7441浓度在0.3 μ M与10 μ M之间。在一个实施例中,培养基中的Nu7441浓度在0.4 μ M与10 μ M之间。在一个实施例中,培养基中的Nu7441浓度在0.5 μ M与10 μ M之间。在一个实施例中,培养基中的Nu7441浓度在1 μ M与10 μ M之间。在一个实施例中,培养基中的Nu7441浓度在0.1 μ M与5 μ M之间。在一个实施例中,培养基中的Nu7441浓度在0.1 μ M与4 μ M之间。在一个实施例中,培养基中的Nu7441浓度在0.1 μ M与3 μ M之间。在一个实施例中,培养基中Nu7441的浓度在0.1 μ M与2 μ M之

间。在一个实施例中，培养基中的Nu7441浓度在0.1 μ M与1 μ M之间。在一个实施例中，培养基中的Nu7441浓度是约0.2 μ M、约0.3 μ M、约0.33 μ M、约0.4 μ M、约0.5 μ M、约0.6 μ M、约0.67 μ M、约0.7 μ M、约0.8 μ M、约0.9 μ M、约1 μ M、约1.33 μ M、约2 μ M或约3 μ M。

[0406] 在一个实施例中，培养基中的咖啡碱浓度在0.2mM与20mM之间。在一个实施例中，培养基中的咖啡碱浓度在1mM与20mM之间。在一个实施例中，培养基中的咖啡碱浓度在2mM与20mM之间。在一个实施例中，培养基中的咖啡碱浓度在5mM与20mM之间。在一个实施例中，培养基中的咖啡碱浓度在10mM与20mM之间。在一个实施例中，培养基中的咖啡碱浓度在0.2mM与10mM之间。在一个实施例中，培养基中的咖啡碱浓度在0.2mM 与8mM之间。在一个实施例中，培养基中的咖啡碱浓度在0.2mM与5mM 之间。在一个实施例中，培养基中的咖啡碱浓度为约1mM、约2mM、约3 mM、约4mM、约5mM、约6mM、约7mM、约8mM、约9mM或约10 mM。

[0407] 在一个实施例中，培养基中的SCR7浓度在0.01 μ M与10 μ M之间。在一个实施例中，培养基中的SCR7浓度在0.01 μ M与5 μ M之间。在一个实施例中，培养基中的SCR7浓度在0.01 μ M与1 μ M之间。在一个实施例中，培养基中的SCR7浓度在0.01 μ M与0.1 μ M之间。在一个实施例中，培养基中的 SCR7浓度在0.01 μ M与0.05 μ M之间。在一个实施例中，培养基中的SCR7 浓度在0.02 μ M与10 μ M之间。在一个实施例中，培养基中的SCR7浓度在 0.05 μ M与10 μ M之间。在一个实施例中，培养基中的SCR7浓度在0.1 μ M 与10 μ M之间。在一个实施例中，培养基中的SCR7浓度在1 μ M与10 μ M之间。在一个实施例中，培养基中的SCR7浓度在2 μ M与10 μ M之间。在一个实施例中，培养基中的SCR7浓度在5 μ M与10 μ M之间。在一个实施例中，培养基中的SCR7浓度是约0.05 μ M、约0.1 μ M、约0.5 μ M、约1 μ M、约2 μ M或约5 μ M。

[0408] 在一个实施例中，培养基中的UNC2170浓度在1 μ M与60 μ M之间。在一个实施例中，培养基中的UNC2170浓度在1 μ M与50 μ M之间。在一个实施例中，培养基中的UNC2170浓度在1 μ M与40 μ M之间。在一个实施例中，培养基中的UNC2170浓度在1 μ M与30 μ M之间。在一个实施例中，培养基中的UNC2170浓度在1 μ M与20 μ M之间。在一个实施例中，培养基中的UNC2170浓度在1 μ M与10 μ M之间。在一个实施例中，培养基中的UNC2170 浓度在10 μ M与60 μ M之间。在一个实施例中，培养基中的UNC2170浓度在 20 μ M与60 μ M之间。在一个实施例中，培养基中的UNC2170浓度在30 μ M 与60 μ M之间。在一个实施例中，培养基中的UNC2170浓度在40 μ M与60 μ M之间。在一个实施例中，培养基中的UNC2170浓度是约5 μ M、约10 μ M、约20 μ M、约30 μ M、约40 μ M、约50 μ M或约60 μ M。

[0409] 在一个实施例中，培养基中的LTURM34浓度在1nM与100nM之间。在一个实施例中，培养基中的LTURM34浓度在1nM与80nM之间。在一个实施例中，培养基中的LTURM34浓度在1nM与60nM之间。在一个实施例中，培养基中的LTURM34浓度在1nM与50nM之间。在一个实施例中，培养基中的LTURM34浓度在1nM与40nM之间。在一个实施例中，培养基中的LTURM34浓度在1nM与30nM之间。在一个实施例中，培养基中的 LTURM34浓度在1nM与20nM之间。在一个实施例中，培养基中的LTURM34 浓度在1nM与10nM之间。在一个实施例中，培养基中的LTURM34浓度在 10nM与100nM之间。在一个实施例中，培养基中的LTURM34浓度在20nM 与100nM之间。在一个实施例中，培养基中的LTURM34浓度在30nM与100 nM之间。在一个实施例中，培养基中的LTURM34浓度在40nM与100nM 之间。在一个实施例中，培养基中的LTURM34浓度在50nM与100nM之间。在一个实施例中，培养基中的LTURM34浓度在60nM与100nM之间。在一个实施例中，培养基中的LTURM34浓度在70nM与100nM之间。在一个实施例中，培养基中的

LTURM34浓度在80nM与100nM之间。在一个实施例中,培养基中的LTURM34浓度是约10nM、约20nM、约30nM、约40nM、约50nM、约60nM、约70nM、约80nM、约90nM或约100nM。

[0410] 在实施例中,将细胞在电穿孔缓冲液中电穿孔,并且在电穿孔之前将NHEJ抑制剂添加到电穿孔缓冲液中。在此类实施例中,通常以比添加到培养基中的浓度高10至100倍的浓度添加NHEJ抑制剂。在实施例中,电穿孔缓冲液包含NHEJ抑制剂,其浓度比培养基中的浓度高约50倍。在实施例中,将包含NHEJ抑制剂、细胞和其它组分的电穿孔缓冲液在电穿孔后直接添加到细胞培养基中。

[0411] 在实施例中,使细胞在电穿孔之前与NHEJ抑制剂接触。在实施例中,使细胞在电穿孔之后与NHEJ抑制剂接触。

[0412] 在实施例中,使用电穿孔缓冲液进行电穿孔步骤,并且在电穿孔之前将NHEJ抑制剂添加到电穿孔缓冲液中。

[0413] 在实施例中,培养步骤包括在细胞培养基中培养细胞,以及将NHEJ抑制剂添加到细胞培养基中。在实施例中,在使细胞与核酸切割实体和供体DNA接触之前,在细胞培养基中培养细胞,并且所述细胞培养基包含NHEJ抑制剂。

[0414] 在实施例中,病毒载体不插入细胞中。即,核酸切割实体、gRNA和/或供体DNA中无一者通过病毒载体提供给细胞。

[0415] 在实施例中,供体核酸是聚合酶链反应(PCR)产物。在一些实施例中,供体核酸是单链的。在一些实施例中,供体核酸是双链或部分双链的。

[0416] 在实施例中,真核细胞是肝细胞。

[0417] 在一个方面中,提供了基因改变的细胞。遗传改变的细胞包括非同源末端连接(NHEJ)抑制剂、核酸切割实体和供体核酸(例如供体DNA)。在实施例中,细胞不包含病毒组分。

[0418] 在一个方面中,提供了一种基因改变的细胞,其通过本文所述方法发生基因改变。

[0419] 在一个方面中,提供了一种遗传改变细胞的试剂盒。试剂盒包括(i)非同源末端连接(NHEJ)抑制剂;(ii)电穿孔缓冲液;和任选存在的(iii)核酸切割实体。

[0420] 在实施例中,试剂盒包括供体核酸,例如供体DNA。

[0421] 在实施例中,试剂盒包括细胞培养基。

[0422] 在实施例中,DNA结合剂包含锌指蛋白、转录激活因子样效应物(TALE)、CRISPR复合物、阿格蛋白-核酸复合物或大范围核酸酶。在实施例中,DNA结合剂是转录激活因子样效应物(TALE)。在实施例中,DNA结合剂是CRISPR复合物。

[0423] 在实施例中,NHEJ抑制剂是DNA依赖性蛋白激酶(DNA-PK)、DNA连接酶IV或其组合。

[0424] 在实施例中,DNA-PK抑制剂是Nu7026 (2-(4-吗啉基)-4H-萘并[1,2-b]吡喃-4-酮)、Nu7441 (8-(4-二苯并噻吩基)-2-(4-吗啉基)-4H-1-苯并吡喃-4-酮)、Ku-0060648 (4-乙基-N-[4-[2-(4-吗啉基)-4-氧代-4H-1-苯并吡喃-8-基]-1-二苯并噻吩基]-1-哌嗪乙酰胺)、化合物401 (2-(4-吗啉基)-4H-嘧啶并[2,1-a]异喹啉-4-酮)、DMNB (4,5-二甲氧基-2-硝基苯甲醛)、ETP 45658 (3-[1-甲基-4-(4-吗啉基)-1H-吡唑并[3,4-d]嘧啶-6-基苯酚)、LTURM 34 (8-(4-二苯并噻吩基)-2-(4-吗啉基)-4H-1,3-苯并噁嗪-4-酮)、UNC2170 (3-溴-N-(3-(叔丁基氨基)丙基)苯甲酰胺)、Scr7 (2,3-二氢-6,7-二苯基-2-硫代-4(1H)-蝶啶酮,6,7-二苯基-2-硫基-二氧四氢蝶啶)、咖啡碱,和/或P1 103盐酸盐 (3-[4-(4-吗啉基吡

啉并[3',2':4,5]呋喃并[3,2-d]嘧啶-2-基]苯酚盐酸盐),或其任何组合。

[0425] 方法

[0426] 本文所提供的方法和组合物尤其适用于调节目标基因座(例如基因、基因组区域或转录调控序列(例如启动子、增强子)),包括染色质(与DNA结合的组蛋白)、DNA、与DNA结合的蛋白质,或其组合。如本文所用,术语“目标基因座”是指细胞基因组内的区域。目标基因座包括结合蛋白质或核酸的一个或多个结合序列,其结合引起目标基因座发生结构和或化学上的修饰。使用本文所提供的方法和组合物,目标基因座可以通过将一种或多种DNA结合剂(例如第一或第二DNA结合调节增强剂)结合到形成目标基因座的一部分的特定位置而发生结构或化学上的修饰。所述DNA结合剂(例如第一或第二DNA结合调节增强剂)的结合可以引起例如目标基因座处的染色质置换或重构,且/或其可以增强其它内源或外源调节剂对目标基因座进一步修饰的可及性。举例来说,通过增强DNA在基因座的裂解位点和周围序列的可及性,本文所提供的方法适用于提高核酸酶(TALEN、Cas9)在基因组基因座的效率和特异性。因此,本文提供的方法和组合物尤其适用于基因组编辑和增强其中涉及的酶促过程。

[0427] 因此,在一个方面中,提供了一种增强细胞中的目标基因座可及性的方法。所述方法包括:(1)将第一DNA结合调节增强剂引入包含编码目标基因座的核酸的细胞中,其中所述第一DNA结合调节增强剂不是细胞内源的;以及(2)使第一DNA结合调节增强剂结合目标基因座的第一增强子结合序列,相对于第一DNA结合调节增强剂的缺乏,借此增强目标基因座的可及性。

[0428] 可以在本文提供的增强子结合序列的上游或下游增强目标基因座的可及性。因此,相对于DNA结合调节增强剂的缺乏,DNA结合调节增强剂结合到增强子结合位点后,位于增强子结合序列5'和3'的染色质可以更易企及。

[0429] 在实施例中,目标基因座包括与目标基因座的多个增强子结合序列(例如2、4、6、8、10个增强子结合序列)结合的多种DNA结合调节增强剂。多个增强子结合序列中的每一个可以通过长度为20-60个核苷酸的序列彼此分开。在实施例中,目标基因座包括第一、第二、第三、第四、第五和第六增强子结合序列,其中第一增强子结合序列通过第二增强子结合序列与第三增强子结合序列连接,第三增强子结合序列通过第四增强子结合序列与第五增强子结合序列连接,并且第四增强子结合序列通过第五增强子结合序列与第六增强子结合序列连接。第一和第二增强子结合序列、第二和第三增强子结合序列、第三和第四增强子结合序列、第四和第五增强子结合序列以及第五和第六增强子结合序列可以各自相隔20-50个核苷酸。在实施例中,第一和第二增强子结合序列、第二和第三增强子结合序列、第三和第四增强子结合序列、第四和第五增强子结合序列以及第五和第六增强子结合序列各自相隔50个核苷酸。

[0430] 在另一个方面中,提供了一种置换细胞中的目标基因座的染色质的方法。所述方法包括:(1)将第一DNA结合调节增强剂引入包含编码目标基因座的核酸的细胞中,其中所述第一DNA结合调节增强剂不是细胞内源的;以及(2)使第一DNA结合调节增强剂结合目标基因座的第一增强子结合序列,借此置换目标基因座的染色质。

[0431] 在另一个方面中,提供了一种重构细胞中的目标基因座的染色质的方法。所述方法包括:(1)将第一DNA结合调节增强剂引入包含编码目标基因座的核酸的细胞中,其中所

述第一DNA结合调节增强剂不是细胞内源的;以及(2)使第一DNA结合调节增强剂结合目标基因座的第一增强子结合序列,借此重构目标基因座的染色质。

[0432] 如上所述,本文提供的方法和组合物可以包括一种或多种DNA结合剂(例如第一或第二DNA结合调节增强剂)的结合以完成目标基因座的调节。因此,在另一个方面中,提供了一种增强细胞中的目标基因座的可达性的方法。所述方法包括(1)向包括编码目标基因座的核酸的细胞中引入:(i)第一DNA结合调节增强剂,其中所述第一DNA结合调节增强剂不是细胞内源的;和(ii)第二DNA结合调节增强剂,其中所述第二DNA结合调节增强剂不是细胞内源的。(2)使第一DNA结合调节增强剂结合目标基因座的第一增强子结合序列;以及(3)使第二DNA结合调节增强剂结合目标基因座的第二增强子结合序列,相对于第一DNA结合调节增强剂或第二DNA结合调节增强剂的缺乏,借此增强目标基因座的可达性。如本文所提供,增强(增加)目标基因座的可达性是指目标基因座的结构调节,其引起调节蛋白或复合物(例如酶(例如核酸酶))在目标基因座上的功能活性增强。从染色质清除目标基因座和/或重构目标基因座处的DNA以使调节蛋白更好地结合和/或增强活性。因此,术语增强(增加)目标基因座的可达性包括调节目标基因座的结构以使调节蛋白的活性增加,其中所述活性包括例如酶活性、DNA结合活性、转录活性。

[0433] 在一个方面中,提供了一种置换细胞中的目标基因座的染色质的方法。所述方法包括:(1)向包含编码目标基因座的核酸的细胞中引入:(i)第一DNA结合调节增强剂,其中第一DNA结合调节增强剂不是细胞内源的;和(ii)第二DNA结合调节增强剂,其中第二DNA结合调节增强剂不是细胞内源的。(2)使第一DNA结合调节增强剂结合目标基因座的第一增强子结合序列;以及(3)使第二DNA结合调节增强剂结合目标基因座的第二增强子结合序列,借此置换目标基因座的染色质。

[0434] 在一个方面中,提供了一种重构细胞中的目标基因座的染色质的方法。所述方法包括:(1)向包含编码目标基因座的核酸的细胞中引入:(i)第一DNA结合调节增强剂,其中第一DNA结合调节增强剂不是细胞内源的;和(ii)第二DNA结合调节增强剂,其中第二DNA结合调节增强剂不是细胞内源的。(2)使第一DNA结合调节增强剂结合目标基因座的第一增强子结合序列;以及(3)使第二DNA结合调节增强剂结合目标基因座的第二增强子结合序列,借此重构目标基因座的染色质。

[0435] 如上所述,本文提供的方法和组合物可以增强目标基因座的可达性并且借此可以允许募集目标基因座的调节活性。因此,提供了一种调节细胞中的目标基因座的方法。所述方法包括(1)向包含编码目标基因座的核酸的细胞中引入:(i)能够结合目标基因座的调节子结合序列的第一调节蛋白或第一调节复合物,其中所述调节子结合序列包括调节位点;和(ii)能够结合目标基因座的第一增强子结合序列的第一DNA结合调节增强剂。以及(2)使第一调节蛋白或第一调节复合物调节调节位点,借此调节细胞中的目标基因座。

[0436] 由于一种或多种DNA结合剂(例如第一或第二DNA结合调节增强剂)结合到目标基因座,因此目标基因座变得更易企及,从而允许增强调节蛋白或调节复合物在目标基因座处的效率和/或特异性。例如,使用本文提供的方法和组合物可以通过例如同源重组来增强基因编辑反应的效率。在实施例,由于一种或多种DNA结合剂(例如第一或第二DNA结合调节增强剂)的存在,因此核酸酶在目标基因座处的核酸酶活性得以增强。

[0437] 因此,在一个方面中,提供了一种增强调节蛋白或调节复合物在细胞中的目标基

因座处的活性的方法。所述方法包括(1)向包含编码目标基因座的核酸的细胞中引入:(i)能够结合目标基因座的调节子结合序列的第一调节蛋白或第一调节复合物,其中调节子结合序列包括调节位点;和(ii)能够结合目标基因座的第一增强子结合序列的第一DNA结合调节增强剂。以及(2)使第一DNA结合调节增强剂结合第一增强子结合序列,借此增强第一调节蛋白或第一调节复合物对细胞中的目标基因座的活性。

[0438] 1.DNA结合调节增强剂

[0439] 如本文提供的“DNA结合调节增强剂”是能够结合细胞中的目标基因座的相应序列(增强子结合序列)并且借此在化学或结构上调节目标基因座的药剂。与目标基因座结合后,本文提供的DNA结合调节增强剂(包括其实例)可调节基因座处的染色质。DNA结合调节增强剂在结合后,可以将增强子结合序列上游(5')或下游(3')的致密堆积异染色质区转化成堆积不太致密的常染色质区。可以通过将组蛋白从它们在目标位点上所结合的DNA上解离(染色质置换)来实现转化。或者,组蛋白可以在目标基因座处的染色质内重排(染色质重构)。一旦改变了目标基因座上的染色质结构,DNA就变得更易企及以便随后修饰目标基因座。通过一种或多种DNA结合调节增强剂(例如第一或第二DNA结合调节增强剂)的结合可以实现此效果。因此,在实施例中,本文阐述的方法包括引入能够结合目标基因座的第二增强子结合序列的第二DNA结合调节增强剂。

[0440] 对于本文提供的方法来说,可以多种方式将增强剂和调节蛋白或复合物引入细胞。可以通过转染编码增强剂和调节蛋白或复合物的核酸(载体)来引入增强剂和调节蛋白或复合物。或者,可以通过转染编码增强剂和调节蛋白或复合物的mRNA来引入增强剂和调节蛋白或复合物。增强剂和调节蛋白或复合物可以通过直接转染实际药剂、调节蛋白或调节复合物来进一步引入。本领域普通技术人员将立即认识到药剂、调节蛋白或复合物在细胞中的半衰期(药剂在细胞中具有活性和/或表达的时间)是由其递送到细胞的实体形式决定。不受任何特定科学理论束缚,相较于增强剂和作为实际蛋白质或复合物转染的调节蛋白或复合物,编码增强剂、调节蛋白或复合物的核酸的递送将使得细胞中表达/存在的增强剂、调节蛋白或复合物延长。

[0441] 在实施例中,引入第一DNA结合调节增强剂包括引入编码第一DNA结合调节增强剂的载体。在实施例中,引入第一DNA结合调节增强剂包括引入编码第一DNA结合调节增强剂的mRNA。在实施例中,引入第一DNA结合调节增强剂包括引入第一DNA结合蛋白或第一DNA结合核酸。

[0442] 在实施例中,引入第二DNA结合调节增强剂包括引入编码第二DNA结合调节增强剂的载体。在实施例中,引入第二DNA结合调节增强剂包括引入编码第二DNA结合调节增强剂的mRNA。在实施例中,引入第二DNA结合调节增强剂包括引入第二DNA结合蛋白或第二DNA结合核酸。

[0443] 在实施例中,引入第一调节蛋白包括引入编码第一调节蛋白的载体。在实施例中,引入第一调节蛋白包括引入编码第一调节蛋白的mRNA。在实施例中,引入第一调节蛋白包括引入第一调节蛋白。在实施例中,引入第一调节复合物包括引入编码第一调节复合物的载体。在实施例中,引入第一调节复合物包括引入编码第一调节复合物的mRNA。在实施例中,引入第一调节复合物包括引入第一调节复合物。

[0444] 在实施例中,引入第二调节蛋白包括引入编码第二调节蛋白的载体。在实施例中,

引入第二调节蛋白包括引入编码第二调节蛋白的mRNA。在实施例中,引入第二调节蛋白包括引入第二调节蛋白。在实施例中,引入第二调节复合物包括引入编码第二调节复合物的载体。在实施例中,引入第二调节复合物包括引入编码第二调节复合物的mRNA。在实施例中,引入第二调节复合物包括引入第二调节复合物。

[0445] 可用于本文提供的方法和组合物的示例性DNA结合调节增强剂包括 DNA结合蛋白或DNA结合核酸。第一DNA结合调节增强剂和第二DNA结合调节增强剂可以是相同的或在化学上不同。在实施例中,第一DNA结合调节增强剂不是细胞内源的。在实施例中,第二DNA结合调节增强剂不是细胞内源的。在实施例中,第一DNA结合调节增强剂是第一DNA结合蛋白或第一DNA结合核酸。在实施例中,第一DNA结合调节增强剂是第一转录激活因子样 (TAL) 效应蛋白或截短的第一向导RNA (gRNA)。在实施例中,第一 DNA结合调节增强剂是第一锌指DNA结合蛋白。在实施例中,第二DNA结合调节增强剂是第二DNA结合蛋白或第二DNA结合核酸。在实施例中,第二DNA结合调节增强剂是TAL效应蛋白或截短的gRNA。

[0446] “截短的gRNA”或“截短的向导RNA”是对应于野生型向导RNA的核糖核酸,但与所述野生型向导RNA相比包含更少的核苷酸。如本文所提供,截短的gRNA可以与Cas9蛋白结合。因此,如本文所提供的截短的向导RNA 可以是与Cas9蛋白结合并且能够结合调节子结合序列的RNA。与截短的 gRNA结合的Cas9蛋白无法裂解调节子结合序列。因此,在实施例中,DNA结合调节增强剂是与Cas9蛋白结合的截短的gRNA。在实施例中,结合至截短的gRNA的Cas9蛋白是酿脓链球菌Cas9蛋白。如本文提供的酿脓链球菌 Cas9蛋白是来源于细菌酿脓链球菌的Cass9蛋白。

[0447] 本文提供的截短的gRNA的长度可以小于16个核苷酸。在实施例中,截短的gRNA的长度不超过15个核苷酸。在实施例中,截短的gRNA的长度为 10至15个核苷酸。在实施例中,截短的gRNA的长度为11至15个核苷酸。在实施例中,截短的gRNA的长度为12至15个核苷酸。在实施例中,截短的 gRNA的长度为13至15个核苷酸。在实施例中,截短的gRNA的长度为10至14个核苷酸。在实施例中,截短的gRNA的长度为10至13个核苷酸。在实施例中,截短的gRNA的长度为10至12个核苷酸。在实施例中,截短的 gRNA的长度为16个核苷酸。在实施例中,截短的gRNA的长度小于15个核苷酸。在实施例中,截短的gRNA的长度为15个核苷酸。在实施例中,截短的gRNA的长度小于14个核苷酸。在实施例中,截短的gRNA的长度为14个核苷酸。在实施例中,截短的gRNA的长度小于13个核苷酸。在实施例中,截短的gRNA的长度为13个核苷酸。在实施例中,截短的gRNA的长度小于 12个核苷酸。在实施例中,截短的gRNA的长度为12个核苷酸。在实施例中,截短的gRNA的长度小于11个核苷酸。在实施例中,截短的gRNA的长度为11个核苷酸。在实施例中,截短的gRNA的长度小于10个核苷酸。在实施例中,截短的gRNA的长度为10个核苷酸。在实施例中,截短的gRNA的长度小于9个核苷酸。在实施例中,截短的gRNA的长度为9个核苷酸。在实施例中,截短的gRNA的长度小于8个核苷酸。在实施例中,截短的gRNA的长度为8个核苷酸。在实施例中,截短的gRNA的长度小于7个核苷酸。在实施例中,截短的gRNA的长度为7个核苷酸。在实施例中,截短的gRNA的长度小于6个核苷酸。在实施例中,截短的gRNA的长度为6个核苷酸。在实施例中,截短的gRNA的长度小于5个核苷酸。在实施例中,截短的gRNA的长度为5个核苷酸。在实施例中,截短的gRNA的长度小于4个核苷酸。在实施例中,截短的gRNA的长度为4个核苷酸。

[0448] 2. 增强子结合序列

[0449] 如本文提供的“增强子结合序列”是形成目标基因座的一部分并被DNA 结合调节增强剂结合的核酸序列。在实施例中,增强子结合序列是TAL核酸结合盒。如本文所用,“TAL核酸结合盒”(也称为“TAL盒”)是指编码多肽的核酸,所述多肽允许包括所述多肽的蛋白质结合单个碱基对(例如A、T、C或G)。在实施例中,蛋白质将包含超过一种由TAL核酸结合盒编码的多肽。所编码的多聚体的个别氨基酸序列被称为“TAL重复序列”。在实施例中,TAL重复序列的长度将在二十八个与四十个氨基酸之间并且(对于存在的氨基酸来说)与以下三十四个氨基酸序列将共有至少60%(例如至少约65%、至少约 70%、至少约75%、至少约80%、约60%至约95%、约65%至约95%、约70%至约95%、约75%至约95%、约80%至约95%、约85%至约95%、约60%至约90%、约60%至约85%、约65%至约90%、约70%至约90%、约75%至约 90%等)一致性:LTPDQVVAIA SXXGKGQALE TVQRLLPVLC QAHG (SEQ ID NO:118)。

[0450] 在实施例中,上述序列中位置十二和十三处的两个X代表TAL核酸结合盒中也识别核酸分子中的特定碱基的氨基酸。

[0451] 在实施例中,存在于一连串重复序列的羧基端的最后TAL重复序列通常是部分TAL重复序列,其中羧基端可以缺失(例如大致为此最后TAL重复序列的氨基端15到20个氨基酸)。

[0452] 在实施例中,增强子结合序列是能够与向导RNA结合序列或向导DNA 结合序列结合(杂交)的核酸序列。在实施例中,第一增强子结合序列具有SEQ ID NO:26、SEQ ID NO:28、SEQ ID NO:30、SEQ ID NO:32、SEQ ID NO:34、SEQ ID NO:36、SEQ ID NO:38或SEQ ID NO:40的序列。在实施例中,第二增强子结合序列具有SEQ ID NO:27、SEQ ID NO:29、SEQ ID NO:31、SEQ ID NO:33、SEQ ID NO:35、SEQ ID NO:37、SEQ ID NO:39或SEQ ID NO: 41的序列。

[0453] 3. 调节蛋白和调节复合物

[0454] 本文提供的调节蛋白和调节复合物可以是细胞内源的,或可以不是细胞内源的。如本文提供的术语“调节蛋白”和“调节复合物”分别是指能够在结构和/或化学上改变目标基因座的分子(例如蛋白质或蛋白质结合物)或分子复合物(例如核糖核蛋白复合物)。目标基因座结构或化学组成的变化可以包括整个目标基因座或其部分的变化。调节蛋白的实例包括但不限于双链核酸酶、切口酶、转录激活因子、转录阻遏物、核酸甲基化酶、核酸脱甲基酶、拓扑异构酶、旋转螺、连接酶、甲基转移酶、转座酶、糖基化酶、整合酶、激酶、磷酸酶、硫脲酶、聚合酶、荧光活性和重组酶。本文提供的调节复合物的非限制性实例包括核糖核蛋白复合物和脱氧核糖核蛋白复合物。

[0455] 在实施例中,第一调节蛋白或第二调节蛋白包括DNA结合蛋白或DNA 调节酶。DNA结合蛋白可以是转录阻遏物或转录激活因子。在实施例中,DNA 调节酶是核酸酶、脱氨酶、甲基化酶或脱甲基酶。在实施例中,第一调节蛋白或第二调节蛋白包括组蛋白调节酶。在实施例中,组蛋白调节酶是脱乙酰酶或乙酰酶。

[0456] 在实施例中,第一调节蛋白或第二调节蛋白包括可操作地连接至第一 DNA修饰域的第一DNA结合域。在实施例中,第一DNA结合域是TAL效应域,并且第一DNA修饰域是转录激活域或转录阻遏域。在实施例中,第一 DNA修饰域是VP16域。在实施例中,第一DNA修饰域是VP64域。在实施例中,第一DNA修饰域是VP16域、VP32域或VP64转录激活域或KRAB转录阻遏

域。

[0457] 在实施例中,第一调节蛋白是第一DNA结合核酸酶结合物。在实施例中,第二调节蛋白是第二DNA结合核酸酶结合物。如本文所用,“DNA结合核酸酶结合物”是指具有核酸切割活性(例如双链核酸切割活性)的一种或多种分子、酶或分子复合物。在大多数实施例中,DNA结合核酸酶结合物组分将是蛋白质或核酸或两者的组合,但是它们可以与辅因子和/或其它分子结合。DNA结合核酸酶结合物通常将基于多种因素进行选择,例如目标基因座处的DS断裂产生效率、在目标基因座处或附近的合适位置处产生DS断裂产生的能力、在非所需基因座处产生DS断裂的潜在性低、低毒性和成本问题。这些因素中的许多因素将随所用细胞和目标基因座而变化。许多DNA结合核酸酶结合物是本领域已知的。例如,在一些实施例中,DNA结合核酸酶结合物包括一种或多种锌指蛋白、转录激活因子样效应物(TALE)、CRISPR复合物(例如Cas9或CPF1)、归巢内切核酸酶或兆碱基大范围核酸酶、阿格蛋白-核酸复合物,或大范围核酸酶。在一些实施例中,DNA结合核酸酶结合物将具有允许它们被核定位的活性(例如将包含核定位信号(NLS))。在一些实施例中,单链DNA供体可以与切口或切口组合一起发挥作用。

[0458] 在实施例中,DNA结合核酸酶结合物是TAL效应物融合体。如本文所提供,“TAL效应物融合体”是指TAL效应物连接至自然界中与其并非天然结合的另一种多肽或蛋白质(例如阿格蛋白)。在实施例中,TAL效应物融合体中的非TAL组分将赋予融合蛋白功能活性(例如酶促活性)。在实施例中,TAL效应物融合体可具有结合活性或可具有直接或间接触发核酸修饰的活性,例如核酸酶活性。

[0459] 在实施例中,第一DNA结合核酸酶结合物包括第一核酸酶,并且第二DNA结合核酸酶结合物包括第二核酸酶。在实施例中,第一核酸酶和第二核酸酶形成二聚体。在实施例中,第一核酸酶和第二核酸酶独立地是转录激活因子样效应物核酸酶(TALEN)。在实施例中,第一核酸酶和第二核酸酶独立地是FokI核酸酶裂解域突变体KKR Sharkey。在实施例中,第一核酸酶和第二核酸酶独立地是FokI核酸酶裂解域突变体ELD Sharkey。

[0460] 在实施例中,第一DNA结合核酸酶结合物包括可操作地连接至第一核酸酶(TALEN)的第一转录激活因子样(TAL)效应域(例如TAL蛋白的DNA结合部分)。在实施例中,第一DNA结合核酸酶结合物包括可操作地连接至第一FokI核酸酶的第一TAL效应域。在实施例中,第二DNA结合核酸酶结合物包括可操作地连接至第二核酸酶(TALEN)的第二TAL效应域。在实施例中,第二DNA结合核酸酶结合物包括可操作地连接至第二FokI核酸酶的第二TAL效应域。在实施例中,第一DNA结合核酸酶结合物包括第一锌指核酸酶。在实施例中,第二DNA结合核酸酶结合物包括第一锌指核酸酶。

[0461] 如本文所用,术语“锌指核酸酶”是指一种蛋白质,其包含具有锌稳定化核酸(例如DNA)结合域的多肽。个别DNA结合域通常被称为“指”,使得锌指蛋白或多肽具有至少一个指,更通常为两个指或三个指,或甚至四个或五个指,至至少六个或更多个指。在某些方面中,锌指核酸酶将包含三个或四个锌指。每个指通常结合DNA的二至四个碱基对。每个指通常包含约30个氨基酸的锌螯合DNA结合区(参见例如美国专利公开第2012/0329067A1号,其公开内容通过引用并入本文)。

[0462] 形成本文提供的结合物的一部分的核酸酶蛋白的一个实例是来自IIS型限制性内切核酸酶FokI的非特异性裂解域(Kim,Y.G.等人,《美国国家科学院院刊》93:1156-60

(1996)),其通常被5-7个碱基对的接头序列分开。通常需要一对FokI裂解域以允许域发生二聚和非回文目标序列从相反的链裂解。个别 Cys₂His₂ ZFN的DNA结合域通常包含3至6个个别锌指重复序列并且各自能够识别9至18个碱基对。

[0463] 如本文所用,“转录激活因子样效应物”(TALE)是指由超过一个TAL重复序列构成的蛋白质,并能够以序列特异性方式结合核酸。TALE代表由植物病原性细菌物种(例如黄单胞菌属和青枯病菌属)在感染植物细胞后经由其III型分泌系统分泌的一类DNA结合蛋白。天然TALE尤其已经显示结合至植物启动子序列,借此调节基因表达和激活效应物特异性宿主基因,从而促进细菌繁殖(Römer,P.等人,《科学(Science)》318:645-648(2007);Boch,J.等人,《植物病理学年鉴(Annu.Rev.Phytopathol.)》48:419-436(2010);Kay,S.等人,《科学(Science)》318:648-651(2007);Kay,S.等人,《微生物学新见(Curr.Opin.Microbiol.)》12:37-43(2009))。

[0464] 天然TALE的特征通常为中心重复域和羧基端核定位信号序列(NLS)和转录激活域(AD)。除通常较短的羧基端重复序列(称为半重复序列)之外,中心重复域典型地由以下组成:1.5个至33.5个不同量的氨基酸重复序列,其长度通常为33-35个残基。所述重复序列大多数是一致的,但是在某些高变残基上有所不同。TALE的DNA识别特异性是由通常位于每个重复序列的位置 12和13的高变残基(所谓的重复可变双残基(RVD),其中每个RVD靶向给定DNA序列中的特定核苷酸)介导。因此,TAL蛋白中的重复序列顺序倾向于与给定DNA序列中的经定义的核苷酸线性顺序相关。已鉴定出天然存在的一些TALE的潜在RVD代码,从而允许预测为了结合至给定DNA序列而必需的重复杂列顺序(Boch,J.等人,《科学》326:1509-1512(2009);Moscou,M.J.等人,《科学》326:1501(2009))。此外,已经表明,以新的重复序列组合产生的TAL效应物结合到根据此代码所预测的目标序列。已经表明,目标DNA序列通常始于被TAL蛋白识别的5'胸腺嘧啶碱基。

[0465] TAL模块结构允许DNA结合域与效应分子(如核酸酶)组合。具体地说,TALE核酸酶允许开发新的基因组工程工具。一些实施例中使用的TALE可以产生DS断裂,或者可以具有用于产生DS断裂的组合作用。例如,TAL-FokI 核酸酶融合体可以设计成在目标基因座处或附近结合并且通过两个FokI域的结合形成双链核酸切割活性。

[0466] 在一些实施例中,TALE将包含大于或等于6个(例如大于或等于8、10、12、15或17,或6至25、6至35、8至25、10至25、12至25、8至22、10至22、12至22、6至20、8至20、10至22、12至20、6至18、10至18、12至18个等)TAL重复序列。在一些实施例中,TALE可包含18或24或17.5或23.5个TAL核酸结合盒。在另外的实施例中,TALE可以包含15.5、16.5、18.5、19.5、20.5、21.5、22.5或24.5个TAL核酸结合盒。TALE通常将具有至少一个多肽区域,所述区域侧接包含TAL重复序列的区域。在许多实施例中,侧接区域将存在于TAL重复序列的氨基端和羧基端。示例性的TALE阐述于美国专利公开第2013/0274129A1号,其公开内容以引入的方式并入本文中,并且可以是伯克霍尔德氏菌属(Burkholderia)、黄单胞菌属(Xanthomonas)和青枯病菌属(Ralstonia)细菌中所发现的天然存在的蛋白质的经修饰形式。在一些实施例中,TALE蛋白质将含有使其转运到核的核定位信号(NLS)。

[0467] 对于本文提供的方法和组合物来说,相对于DNA结合调节增强剂的缺乏,调节蛋白或调节复合物的核酸靶向能力增强。在实施例中,相对于DNA结合调节增强剂的缺乏,目标基因座处的同源重组速率增加。

[0468] 在实施例中,第一调节复合物是第一核糖核蛋白复合物。在实施例中,第二调节复合物是第二核糖核蛋白复合物。在实施例中,第一核糖核蛋白复合物包括与gRNA结合的CRISPR相关蛋白9 (Cas9) 结构域或与向导DNA (gDNA) 结合的阿格蛋白结构域。在实施例中,第二核糖核蛋白复合物包括与gRNA结合的CRISPR相关蛋白9 (Cas9) 结构域或与向导DNA (gDNA) 结合的阿格蛋白结构域。

[0469] 在实施例中,第一DNA结合调节增强剂是第一TAL效应蛋白并且第二DNA结合调节增强剂是第二TAL效应蛋白。在实施例中,第一DNA结合调节增强剂是TAL效应蛋白并且第二DNA结合调节增强剂是截短的gRNA。在实施例中,第一DNA结合调节增强剂是截短的第一gRNA并且第二DNA结合调节增强剂是截短的第二gRNA。在实施例中,第一DNA结合调节增强剂是截短的gRNA并且第二DNA结合调节增强剂是TAL效应蛋白。

[0470] 在实施例中,第一调节蛋白是第一DNA结合核酸酶结合物并且第二调节蛋白是第二DNA结合核酸酶结合物。在实施例中,第一调节蛋白是DNA结合核酸酶结合物并且第二调节复合物是核糖核蛋白复合物。在实施例中,第一调节复合物是第一核糖核蛋白复合物并且第二调节复合物是第二核糖核蛋白复合物。在实施例中,第一调节复合物是核糖核蛋白复合物并且第二调节蛋白是DNA结合核酸酶结合物。

[0471] 本文提供的药剂对于表达它们的细胞可以是内源的或非内源的。因此,在实施例中,第一调节蛋白或第一调节复合物不是细胞内源的。在实施例中,第一调节蛋白、第一调节复合物、第二调节蛋白或第二调节复合物不是细胞内源的。在实施例中,第一调节蛋白和第二调节蛋白不是细胞内源的。在实施例中,第一调节复合物和第二调节复合物不是细胞内源的。在实施例中,第一DNA结合调节增强剂或第二DNA结合调节增强剂不是细胞内源的。在实施例中,第一DNA结合调节增强剂和第二DNA结合调节增强剂不是细胞内源的。

[0472] 申请人惊奇地发现,第一和/或第二增强子结合位点相对于调节子结合序列的距离影响DNA结合调节增强剂对调节蛋白或调节复合物的活性的影响。第一增强子结合位点与调节子结合序列之间的距离是使第一DNA结合调节增强剂的大部分3'核苷酸与调节子结合序列的大部分5'核苷酸连接的核苷酸的数目。类似地,第二增强子结合位点与调节子结合序列之间的距离是使调节子结合序列的大部分3'核苷酸和第一DNA结合调节增强剂的大部分5'核苷酸连接的核苷酸的数目。调节子结合序列可以被蛋白质(例如DNA结合蛋白)或核酸(例如gRNA或gDNA)结合。调节子结合序列中所包括的调节位点是调节子结合序列中的核苷酸的位置,其被调节蛋白或调节复合物识别,并且对应于其与调节子结合序列的其余部分的键被水解的核苷酸。

[0473] 在实施例中,第一增强子结合序列或第二增强子结合序列与调节子结合序列相隔少于200个核苷酸。在实施例中,第一增强子结合序列或第二增强子结合序列与调节子结合序列相隔少于150个核苷酸。在实施例中,第一增强子结合序列或第二增强子结合序列与调节子结合序列相隔少于100个核苷酸。在实施例中,第一增强子结合序列或第二增强子结合序列与调节子结合序列相隔少于50个核苷酸。在实施例中,第一增强子结合序列或第二增强子结合序列与调节子结合序列相隔4至30个核苷酸。在实施例中,第一增强子结合序列或第二增强子结合序列与调节子结合序列相隔7至30个核苷酸。在实施例中,第一增强子结合序列或第二增强子结合序列与调节子结合序列相隔4个核苷酸。在实施例中,第一增强子结合序列或第二增强子结合序列与调节子结合序列相隔7个核苷酸。在实施例中,第一增强子

结合序列或第二增强子结合序列与调节子结合序列相隔12个核苷酸。在实施例中,第一增强子结合序列或第二增强子结合序列与调节子结合序列相隔20个核苷酸。在实施例中,第一增强子结合序列或第二增强子结合序列与调节子结合序列相隔30个核苷酸。

[0474] 在实施例中,第一增强子结合序列或第二增强子结合序列与调节位点相隔 10至40个核苷酸。在实施例中,第一增强子结合序列或第二增强子结合序列与调节位点相隔33个核苷酸。

[0475] 在实施例中,第一增强子结合序列与调节子结合序列相隔30个核苷酸,并且第二增强子结合序列与调节子结合序列相隔19个核苷酸。在其它实施例中,第一增强子结合序列和第二增强子结合序列的长度独立地为18个核苷酸。在另一个其它实施例中,调节子结合序列包括第一结合序列和第二结合序列,其中第一结合序列和第二结合序列的长度独立地为18个核苷酸并且相隔16 个核苷酸序列。

[0476] 在实施例中,第一DNA结合调节增强剂或第二DNA结合调节增强剂增强第一调节蛋白、第一调节复合物、第二调节蛋白或第二调节复合物在调节位点的活性。

[0477] 本文提供了多种形式,用于增强细胞中存在的其它组分(例如供体DNA 分子、调节蛋白、调节复合物等)对目标基因座的可及性。调节增强剂(其结合目标基因座中的特定DNA序列(增强子结合序列))的结合增强了可及性。DNA结合调节增强剂可以是截短的gRNA或TAL效应域。在实施例中,两种 DNA结合调节增强剂(例如第一和第二DNA结合调节增强剂)结合目标基因座。在两种DNA结合调节增强剂(例如两个TAL效应域或两个截短的gRNA) 结合目标基因座的情况下,它们可以侧接包括例如核酸酶裂解位点的调节序列。相对于DNA结合调节增强剂的缺乏,通过DNA结合调节增强剂对它们相应的增强子结合序列的结合,目标基因座的调节序列可以更易企及。

[0478] 一个方面尤其提供了一种目标基因座,其包括两个各自与其相应结合序列(增强子结合序列)结合的TAL效应域,其中所述增强子结合序列侧接具有调节位点(例如核酸酶裂解位点)的调节子结合序列。在增强子结合序列侧接调节子结合序列的情况下,第一增强子结合序列通过调节子结合序列与第二增强子结合序列连接。因此,在5'至3'方向上,目标基因座可编码与调节子结合序列连接的第一增强子结合序列,所述调节子结合序列与第二增强子结合序列连接。两个TAL效应域对其相应结合序列(增强子结合序列)的结合允许尤其在调节子结合序列被两种TALEN结合物结合和/或修饰的目标基因座的可及性。两个增强子结合序列的每一个与调节子结合序列可以相隔例如7个核苷酸。在两个增强子结合序列中的每一个与调节子结合序列相隔7个核苷酸的情况下,第一个增强子结合序列的大部分3'核苷酸(即最后一个核苷酸)通过7 个连续核苷酸的序列与调节子结合序列的大部分5'核苷酸(即第一核苷酸)连接。类似地,第二增强子结合序列的大部分5'核苷酸(即,第一核苷酸)通过 7个连续核苷酸的序列与调节子结合序列的大部分3'核苷酸(即最后一个核苷酸)连接。在两种调节蛋白或调节复合物或其组合结合调节子结合序列的情况下,它们可以通过分别结合独立的结合序列(第一结合序列和第二结合序列) 来结合。第一结合序列可以包括在调节子结合序列的5'部分中,而第二结合序列可以形成调节子结合序列的3'部分的一部分。因此,在5'至3'方向上,调节子结合序列可包括通过至少一个核苷酸连接至第二结合序列的第一结合序列。在实施例中,调节子结合序列的大部分5'核苷酸(即,第一核苷酸)是第一结合位点的大部分5'核苷酸,并且调节子结合序列的大部分3'核苷酸(即最后一

个核苷酸)是第一个结合位点的大部分3'核苷酸。

[0479] 此外,两个增强子结合序列中的每一个可以与裂解位点(调节位点)相隔 33个核苷酸。在两个增强子结合序列中的每一个与调节位点相隔33个核苷酸的情况下,第一增强子结合序列的大部分3'核苷酸通过33个连续核苷酸的序列与调节位点的5'核苷酸连接。类似地,第二增强子结合序列的大部分5'核苷酸通过33个连续核苷酸的序列连接至调节位点的3'核苷酸。

[0480] 因此,在一个实施例中,目标基因座包括与第一TAL效应蛋白结合的第一增强子结合序列;与第二TAL效应蛋白结合的第二增强子结合序列;第一 DNA结合核酸酶结合物,其由可操作地连接至第一TALEN的第一TAL效应域组成,其中所述第一结合物在第一结合位点结合至调节子结合序列;以及第二DNA结合核酸酶结合物,其由可操作地连接至第二TALEN的第二TAL效应域组成,其中所述第二结合物在第二结合位点结合至调节子结合序列。在另一个实施例中,第一增强子结合序列与调节子结合序列相隔7个核苷酸,并且第二增强子结合序列与调节子结合序列相隔7个核苷酸。在另一个实施例中,第一增强子结合序列与调节子结合序列的第一结合序列相隔7个核苷酸,并且第二增强子结合序列与调节子结合序列的第二结合序列相隔7个核苷酸。在另一个实施例中,第一增强子结合序列与调节位点相隔33个核苷酸,并且第二增强子结合序列与调节位点相隔33个核苷酸。在另一个实施例中,第一增强子结合序列与调节子结合序列相隔12个核苷酸,并且第二增强子结合序列与调节子结合序列相隔12个核苷酸。在另一个实施例中,第一增强子结合序列与调节子结合序列相隔4个核苷酸,并且第二增强子结合序列与调节子结合序列相隔4个核苷酸。

[0481] 在一个实施例中,目标基因座包括与第一TAL效应蛋白结合的第一增强子结合序列;与第二TAL效应蛋白结合的第二增强子结合序列;第一DNA结合核酸酶结合物,其由可操作地连接至第一TALEN的第一TAL效应域组成,其中所述第一结合物在第一结合位点结合至调节子结合序列;以及第二DNA结合核酸酶结合物,其由可操作地连接至第二TALEN的第二TAL效应域组成,其中所述第二结合物在第二结合位点结合至调节子结合序列。在另一个实施例中,第一增强子结合序列与调节子结合序列相隔30个核苷酸,并且第二增强子结合序列与调节子结合序列相隔19个核苷酸。在另一个实施例中,第一增强子结合序列与调节子结合序列的第一结合序列相隔30个核苷酸,并且第二增强子结合序列与调节子结合序列的第二结合序列相隔19个核苷酸。在另一个实施例中,第一增强子结合序列的长度为18个核苷酸,并且第二增强子结合序列的长度为18个核苷酸。在另一个实施例中,调节子结合序列的第一结合序列与调节子结合序列的第二结合序列相隔16个核苷酸。

[0482] 在一个实施例中,目标基因座包括与第一TAL效应蛋白结合的第一增强子结合序列;与第二TAL效应蛋白结合的第二增强子结合序列;以及核糖核蛋白复合物,其由与向导RNA结合的Cas9结构域组成,其中所述核糖核蛋白复合物与调节子结合序列结合。在另一个实施例中,第一增强子结合序列与调节子结合序列相隔7个核苷酸,并且第二增强子结合序列与调节子结合序列相隔7个核苷酸。在另一个实施例中,第一增强子结合序列与调节子结合序列相隔20个核苷酸,并且第二增强子结合序列与调节子结合序列相隔20个核苷酸。

[0483] 在一个实施例中,目标基因座包括与第一TAL效应蛋白结合的第一增强子结合序列;与第二TAL效应蛋白结合的第二增强子结合序列;以及DNA结合结合物,其由可操作地连接至转录激活域的TAL效应域组成,其中所述DNA结合结合物与调节子结合序列结合。在另

一个实施例中,第一增强子结合序列与调节子结合序列相隔30个核苷酸,并且第二增强子结合序列与调节子结合序列相隔30个核苷酸。在另一个实施例中,第一增强子结合序列的长度为18个核苷酸,并且第二增强子结合序列的长度为18个核苷酸。在另一个实施例中,调节子结合序列的长度为18个核苷酸。

[0484] 在另一个实施例中,目标基因座包括与结合至Cas9蛋白的第一截短的向导RNA结合的第一增强子结合序列;与结合至Cas9蛋白的第二截短的向导RNA结合的第二增强子结合序列;以及核糖核蛋白复合物,其由与向导RNA结合的Cas9结构域组成,其中所述核糖核蛋白复合物与调节子结合序列结合。在另一个实施例中,第一增强子结合序列与调节子结合序列相隔30个核苷酸,并且第二增强子结合序列与调节子结合序列相隔15个核苷酸。

[0485] 在一个实施例中,调节子结合序列的长度为52个核苷酸。在另一个实施例中,第一结合序列的长度为18个核苷酸。在另一个实施例中,第二结合序列的长度为18个核苷酸。

[0486] 在一个实施例中,第一DNA结合调节增强剂是第一TAL效应蛋白,并且第二DNA结合调节增强剂是第二TAL效应蛋白。

[0487] 在一个实施例中,第一DNA结合调节增强剂是第一TAL效应蛋白,并且第二DNA结合调节增强剂是截短的gRNA。在另一个实施例中,截短的gRNA与Cas9蛋白结合。

[0488] 在一个实施例中,第一DNA结合调节增强剂是截短的第一gRNA,并且第二DNA结合调节增强剂是截短的第二gRNA。在另一个实施例中,截短的第一gRNA与第一Cas9蛋白结合,并且截短的第二gRNA与第二Cas9蛋白结合。

[0489] 在一个实施例中,第一DNA结合调节增强剂是第一TAL效应蛋白;第二DNA结合调节增强剂是第二TAL效应蛋白;第一调节蛋白是第一DNA结合核酸酶结合物,其由可操作地连接至第一TALEN的第一TAL效应域组成;并且第二调节蛋白是第二DNA结合核酸酶结合物,其由可操作地连接至第二TALEN的第二TAL效应域组成。

[0490] 在一个实施例中,第一DNA结合调节增强剂是第一TAL效应蛋白;第二DNA结合调节增强剂是第二TAL效应蛋白;并且调节蛋白复合物是与向导RNA结合的Cas9结构域。

[0491] 在一个实施例中,第一DNA结合调节增强剂是与Cas9蛋白结合的截短的第一gRNA;第二DNA结合调节增强剂是与Cas9蛋白结合的截短的第二gRNA;并且调节蛋白复合物是与向导RNA结合的Cas9结构域。

[0492] 在一个实施例中,第一DNA结合调节增强剂是与Cas9蛋白结合的截短的第一gRNA;第二DNA结合调节增强剂是与Cas9蛋白结合的截短的第二gRNA;第一调节蛋白是第一DNA结合核酸酶结合物,其由可操作地连接至第一TALEN的第一TAL效应域组成;并且第二调节蛋白是第二DNA结合核酸酶结合物,其由可操作地连接至第二TALEN的第二TAL效应域组成。

[0493] B. 用于细胞内变化的核酸分子

[0494] 供体核酸分子(例如供体DNA分子)通常将包含至少一个与目标基因座处或附近的核酸对应的同源区域和为了修饰目标基因座而设计的惰性区域。为了同源重组而设计的供体核酸分子通常将具有呈以下顺序的至少三个区域:(1)与目标基因座或附近的核酸对应的第一同源区域;(2)插入区域;和(3)对应于目标基因座处或附近的核酸的第二同源区域(参见图38)。此外,供体核酸分子可以是单链(SS)或双链(DS),并且它们的一个或两个末端可为钝化末端,或者可以在一个或两个末端上具有突出物。突出物当存在时,可以是5'、3'或3'和5'。另外,突出物的长度可以变化。供体核酸分子通常也将包含“插入”区域,所述区

域可以是约一个核苷酸至约数千个核苷酸。

[0495] 当供体核酸分子的一个或两个末端与设计要引入的双链断裂的末端“匹配”时,同源重组的效率会提高。此外,供体核酸分子进入细胞之后(以及进入细胞之前),可暴露于核酸酶(例如内切核酸酶、内切核酸酶等)。为了限制内切核酸酶在改变供体核酸分子方面的作用,可以存在一个或多个核酸酶抗性基团。

[0496] 预定修饰的细胞内核酸分子可以是任何细胞内核酸分子,包括染色体、核质粒、叶绿体基因组和线粒体基因组。此外,预定修饰的细胞内核酸分子可以位于细胞中的任何位置。

[0497] 图38显示了可以在本文阐述的方法中使用的供体核酸分子的许多变异体。末端的空心圆代表核酸酶抗性基团。此类基团可以位于供体核酸分子中的许多位置。供体核酸分子编号6显示位于核酸酶抗性基团之后的较低链的3'末端区域。在某些情况下,细胞核酸酶将消化供体核酸分子的这一部分。这些核酸酶将被核酸酶抗性基团终止或减慢,借此稳定了较低链的3'区域末端的结构。

[0498] 包含含有一个或多个(例如一个、两个、三个、四个、五个、六个、七个等)核酸酶抗性基团的核酸分子的组合物可以用于实施本文所述的方法。在许多情况下,核酸酶抗性基团将位于供体核酸分子的一个末端或两个末端。相对于一个或两个末端,供体核酸分子可包含内部基团。在许多情况下,一些或所有这样的供体核酸分子将在细胞内加工以产生与双链断裂位点匹配端。

[0499] 同源区域可以具有不同的长度,并且可以与目标基因座处的核酸具有不同数量的序列一致性。通常,同源重组效率随着同源区域的长度和序列一致性的增加而增加。所用同源区域的长度通常由以下因素决定:例如大核酸分子的脆性、转染效率,以及容易产生含有同源区域的核酸分子。

[0500] 同源区域的总长度可以是约20个碱基至约10,000个碱基(例如约20个碱基至约100个碱基、约30个碱基至约100个碱基、约40个碱基至约100个碱基、约50个碱基至约8,000个碱基、约50个碱基至约7,000个碱基、约50个碱基至约6,000个碱基、约50个碱基至约5,000个碱基、约50个碱基至约3,000个碱基、约50个碱基至约2,000个碱基、约50个碱基至约1,000个碱基、约50个碱基至约800个碱基、约50个碱基至约600个碱基、约50个碱基至约500个碱基、约50个碱基至约400个碱基、约50个碱基至约300个碱基、约50个碱基至约200个碱基、约100个碱基至约8,000个碱基、约100个碱基至约2,000个碱基、约100个碱基至约1,000个碱基、约100个碱基至约700个碱基、约100个碱基至约600个碱基、约100个碱基至约400个碱基、约100个碱基至约300个碱基、约150个碱基至约1,000个碱基、约150个碱基至约500个碱基、约150个碱基至约400个碱基、约200个碱基至约1,000个碱基、约200个碱基至约600个碱基、约200个碱基至约400个碱基、约200个碱基至约300个碱基、约250个碱基至约2,000个碱基、约250个碱基至约1,000个碱基、约350个碱基至约2,000个碱基、约350个碱基至约1,000个碱基等)。

[0501] 在某些情况下,可能需要使用长度小于200个碱基的序列同源区域。当供体核酸分子包含小的插入片段(例如小于约300个碱基)和/或当供体核酸分子具有与双链断裂位点匹配的一个或两个突出端时,通常会是这样情况。

[0502] 突出端可以具有各种长度,并且在相同供体核酸分子的每个末端可以具有不同的

长度。在许多情况下,这些突出物将形成序列同源区域。图38例如显示了具有30个核苷酸单链突出物的一系列供体核酸分子。这些供体核酸分子显示为单链和双链。图38中的供体核酸分子编号1是具有30个核苷酸序列的单链分子,其具有预定双股断裂位点、30个核苷酸插入和位于每一端的两个核酸酶抗性基团。

[0503] 同源区域与目标基因座核酸共享序列一致性的量越大,通常则同源重组效率越高。当同源区域相当短时(例如50个碱基)时,特别需要高水平的序列一致性。通常,目标基因座与同源区域之间的序列一致性的量将大于90%(例如约90%至约100%、约90%至约99%、约90%至约98%、约95%至约100%、约95%至约99%、约95%至约98%、约97%至约100%等)。

[0504] 供体核酸分子的插入区可以具有多种长度,这取决于其预期的应用。在许多情况下,供体核酸分子的长度是约1至约4,000个碱基(例如约1至3,000、约1至2,000、约1至1,500、约1至1,000、约2至1,000、约3至1,000、约5至1,000、约10至1,000、约10至400、约10至50、约15至65、约2至15个碱基等)。

[0505] 本文还提供用于将少量碱基(例如约1至约10、约1至约6、约1至约5、约1至约2、约2至约10、约2至约6、约3至约8个等)引入细胞内核酸的组合物和方法)。为了说明的目的,可以制备长度为五十一个碱基对的供体核酸分子。此供体核酸分子可具有长度为25个碱基对的两个同源区域,其中插入区域为单个碱基对。当目标基因座周围的核酸基本上与同源区域匹配而没有中间碱基对时,同源重组将在目标基因座处引入单个碱基对。诸如此类的同源重组反应可用于例如破坏蛋白质编码阅读框,从而在细胞内核酸中引入移码。因此提供了将一种或少量碱基引入细胞内核酸分子的组合物和方法。

[0506] 一个方面包括用于改变细胞内核酸分子中的短核苷酸序列的组合物和方法。一个此实例是单核苷酸位置的改变,一个实例是单核苷酸多态性(SNP)的校正或改变。为了说明的目的,使用SNP改变,可以设计具有长度为25个碱基对的两个同源区域的供体核酸分子。这些同源区域之间定位有单个碱基对,所述碱基对基本上是细胞内核酸分子中的相应碱基对的“错配”。因此,同源重组可以通过将碱基对改为一个被认为是野生型的碱基对或另一个碱基(例如不同的SNP)而用于改变SNP。可以通过随后对目标基因座进行测序来鉴定已经正确进行了同源重组的细胞。

[0507] 一个方面包括改变基因组(包括SNP改变)以用于治疗应用的组合物和方法。为了说明的目的,以下列出了由SNP改变引起的两种遗传病。

[0508] 与镰状细胞性贫血相关的最常见SNP是rs334,其引起单个密码子从GAG变为GTG的改变。这种变化导致谷氨酸残基被缬氨酸残基置换。本文阐述的组合物和方法适合于将此SNP从GTG改变为GAG,特别是SNP rs334的个别纯合子。这些原因之一涉及将核酸分子引入细胞可以诱导毒性相关效应。此外,这些效应的等级是其随核酸引入细胞中的量而增加。如下实例所示,基因组插入效率使得需要将相对少量的供体DNA引入细胞中(参见例如图11和13中的供体DNA-NLS结合物数据)。

[0509] 用于改变患者中SNP rs334的一种示例性离体工作流程将包括从患者中去除骨髓组织、改变SNP rs334,随后将编辑细胞再引回至患者体内。

[0510] 与囊性纤维化相关的最常见基因组改变之一是基于囊性纤维化跨膜传导调节因子(CFTR)中的三碱基对缺失(SNP rs199826652),其引起位置508的氨基酸苯丙氨酸缺失。

[0511] 用于改变患者中SNP rs199826652的体内工作流程包括在发生三碱基对插入以校正SNP rs199826652的条件下,将供体DNA分子递送至患者的气管细胞。

[0512] 有效基因编辑所需的低剂量供体核酸也可用于全身递送。之所以如此,是因为低剂量与降低的毒性相关。当使用经修饰的核酸分子(例如具有硫代磷酸酯键的核酸分子)时,低供体DNA分子水平尤其重要。

[0513] 供体核酸分子可以与细胞外靶向部分以及细胞内靶向部分结合。“细胞外靶向部分”是将供体核酸分子引导至一种或多种细胞类型的分子。此类部分包括细胞表面受体配体和抗体。假单胞菌的结构域II已显示参与跨细胞膜的转运。(Jinno等人,《生物化学杂志(J.Biol.Chem.)》263:13203-13207(1988))。因此,一种用于将核酸分子递送至生物体中的亚细胞位置的示例性系统可能涉及以下成分:(1)供体DNA分子、(2)核结合信号(NLS),和(3)融合蛋白,其包含结合至细胞表面受体和假单胞菌外毒素的结构域II的抗体,其中NLS和融合蛋白与供体DNA分子共价结合。这种类型的供体DNA分子允许全身递送供体DNA分子,其中供体DNA分子将被递送至包含细胞表面受体的细胞内的亚细胞位置。

[0514] 在上述两个实例的每一个中,为了使患者获得可观的收益,仅需改变一个等位基因的一个副本。但是,在许多细胞中,SNP的两个副本都会被改变。因此,一个方面包括治疗由纯合和杂合基因组分所致的罹病。

[0515] 供体核酸分子也可以设计成引入功能性编码区染色体开放阅读框。一个实例是将开放阅读框末端的终止密码子去除。此类终止密码子可以被去除,因为它们不存在于野生型开放阅读框中(即,代表“野生型”的改变),或者它们可以天然地存在于开放阅读框的末端。终止密码子也可以引入编码区域。当试图破坏开放阅读框架时,这尤其有用。

[0516] 此外,可以引入标签编码区,使得蛋白质表达产生经标记的蛋白质。可以将此类标签引入细胞内核酸中,使得标签存在于蛋白质的氨基端、羧基端或内部的一个或多个处。标签的实例包括抗原决定基标签(例如His标签、麦芽糖结合蛋白(MBP)标签、纤维素结合域(CBD)标签和谷胱甘肽S-转移酶(GST)标签等)和酶标签(例如辣根过氧化物酶(HRP)标签和碱性磷酸酶(AP)标签等)。

[0517] 因此,一个方面包括用于产生非天然存在的蛋白质而无需克隆编码非天然存在的蛋白质的核酸分子的组合物和方法。这些方法部分地基于将多肽编码区引入细胞内核酸分子中产生融合蛋白的位置,所述融合蛋白由经修饰的细胞内核酸分子编码,随后表达被编码的融合蛋白且从细胞中分离出融合蛋白。

[0518] C. 细胞

[0519] 本文提供的细胞(包括其实例)包括能够增强细胞中的基因组基因座的可及性的复合物。所提供的复合物可通过包含增强子蛋白来增强调节蛋白或复合物在基因组(目标)基因座处的活性,所述增强子蛋白可增加调节蛋白对基因座的可及性。例如,在本文所提供的DNA结合调节增强剂结合到基因组基因座(目标基因座)后,使基因座更易被核酸酶或其它酶活性企及,从而增强了所述核酸酶或其它酶活性的效率和效力。

[0520] 在一个方面中,提供了一种细胞,其包含编码目标基因座调节复合物的核酸。所述复合物包括:(i)目标基因座,其包括第一增强子结合序列和包含调节位点的调节子结合序列;(ii)与调节子结合序列结合的第一调节蛋白或第一调节复合物;以及(iii)与第一增强子结合序列结合的第一DNA结合调节增强剂。

[0521] 在实施例中,目标基因座进一步包括通过调节子结合序列与第一增强子结合序列连接的第二增强子结合序列。

[0522] 在实施例中,细胞包括与第二增强子结合序列结合的第二DNA结合调节增强剂。

[0523] 在一个方面中,提供了包含编码目标基因座复合物的核酸的细胞。所述复合物包括 (i) 包含第一增强子结合序列的目标基因座;以及 (ii) 与第一增强子结合序列结合的第一DNA结合调节增强剂,其中所述第一DNA结合调节增强剂不是细胞内源的,并且其中相对于第一DNA结合调节增强剂的缺乏,所述第一DNA结合调节增强剂能够增强目标基因座的可及性。

[0524] 在一个方面中,提供了包含编码目标基因座复合物的核酸的细胞。所述复合物包括 (1) 目标基因座,其包括: (i) 第一增强子结合序列;和 (ii) 第二增强子结合序列。(2) 与目标基因座的第一增强子结合序列结合的第一DNA结合调节增强剂,其中所述第一DNA结合调节增强剂不是细胞内源的;以及 (3) 与目标基因座的第二增强子结合序列结合的第二DNA结合调节增强剂,其中所述第二DNA结合调节增强剂不是细胞内源的,其中相对于第一DNA结合调节增强剂和第二DNA结合调节增强剂的缺乏,第一DNA结合调节增强剂和第二DNA结合调节增强剂能够增强目标基因座的可及性。

[0525] 本文所述的组合物和方法可用于产生可用于许多目的的细胞系。例如,可以改变单个基因座或多个基因座。可以产生的细胞系的一个实例是用于产生人源化抗体的CHO细胞系。为了产生此类细胞系,在设计成插入CHO细胞基因组的条件下,将编码人源化抗体序列的供体核酸分子引入CHO细胞系。通常,还将可选标记物引入基因组以允许选择经修饰的细胞。当然,可以使用任何合适的细胞系和基本上任何期望的编码序列。因此,一方面包括用于产生细胞的组合物和方法,所述细胞可用于生物生产基因产物(例如蛋白质)。

[0526] 本文所述的组合物和方法也可以用于产生原代细胞或癌细胞的统一池。沿着这些思路,高效的基因编辑可以直接或在最少的选择后改变可用于“下游”应用的细胞。

[0527] 一种示例性工作流程涉及辛伐他汀(simvastatin)前体(莫那可林J(monacolin J))的合成。辛伐他汀前体能够在化学上利用多步骤工艺制备,所述工艺涉及洛伐他汀(lovastatin)(一种通过土曲霉(*Aspergillus terreus*)产生的真菌聚酮化合物)。在产黄青霉菌(*Penicillium chrysogenum*)中发现的洛伐他汀水解酶已得到鉴定并且表征。此水解酶高效地使洛伐他汀水解为莫纳可林J,但对于辛伐他汀的活性则检测不到(参见Huang等人,通过工程改造土曲霉的工业菌株来单步生产辛伐他汀前体莫纳可林J(Single-step production of the simvastatin precursor monacolin J by engineering of an industrial strain of *Aspergillus terreus*),《代谢工程(Metabolic Engineering)》42:109-114(2017))。

[0528] 在此工作流程中,通过将产黄青霉菌洛伐他汀水解酶稳定地引入到土曲霉基因组中,开发出了土曲霉生产细胞系。由于洛伐他汀是土曲霉产生的天然聚酮化合物产品,因此工程化细胞然后将洛伐他汀在细胞内转化为莫纳可林J。因此,一种工作流程是使用本文所述方法对土曲霉细胞进行工程改造,其中有足够百分比的细胞群(例如超过60%)表达洛伐他汀水解酶,使得所述细胞群可直接用于莫纳可林J的生产。一种替代的工作流是在使用前选择工程化土曲霉细胞或针对工程化土曲霉细胞进行选择。

[0529] 与上述相似的工作流程可用于产生供筛选分析中使用的细胞(例如原代哺乳动物

细胞、永生化哺乳动物细胞等)。一个实例是修饰原代肝细胞,然后根据与药物相关的肝毒性用于筛选。

[0530] D. 试剂盒

[0531] 一个方面包括试剂盒,其部分地用于组装和/或储存核酸分子以及用于编辑细胞基因组。作为这些试剂盒的一部分,提供了组装核酸分子和制备反应混合物的材料和说明书用于试剂盒组分的储存和使用。

[0532] 试剂盒可以含有下列组分中的一种或多种:

[0533] 1. 一种或多种DNA结合调节增强剂(例如TAL效应蛋白或与Cas9蛋白结合的截短的gRNA),

[0534] 2. 一种或多种调节蛋白(例如DNA结合核酸酶结合物,其包含与核酸酶连接的TAL效应域),

[0535] 3. 一种或多种调节复合物(例如一个或多个与gRNA结合的Cas9结构域、与向导DNA结合的阿格蛋白结构域等),以及

[0536] 4. 如何使用试剂盒组分的说明书。

[0537] 试剂盒试剂可以提供于任何适合容器中。试剂盒可以提供例如一种或多种反应或储存缓冲液。试剂可以能用于特定反应的形式提供,或以使用之前需要添加一种或多种其它组分的形式(例如浓缩物或冻干形式)提供。缓冲液可以是任何缓冲液,包括(但不限于)碳酸钠缓冲液、碳酸氢钠缓冲液、硼酸盐缓冲液、Tris缓冲液、MOPS缓冲液、HEPES缓冲液以及其组合。在一些实施例中,缓冲液呈碱性。在一些实施例中,缓冲液具有约7到约10的pH。

[0538] 实例

[0539] 提供下述实例,以阐释某些公开的实施例,并且决不视为限制本公开的范围。实例并非想表示以下实验是所进行的全部或唯一实验。尽力确保所用数字(例如数量、温度等)的准确性,但应考虑一些实验误差和偏差。除非另有说明,否则份数是重量份,分子量是重均分子量,温度是摄氏度,并且压力是大气压或接近大气压。

[0540] 实例1:启动子插入

[0541] 材料

[0542] GENEART™PLATINUM™Cas9核酸酶、GENEART™CRISPR gRNA设计工具、GENEART™精密gRNA合成试剂盒、293FT细胞、杜尔贝科氏修改的伊格尔培养基(Dulbecco's Modified Eagle Medium, DMEM)培养基、胎牛血清(FBS)、TRYPLE™表现酶、2% E-Gel®EX琼脂糖凝胶、TranscriptAid T7高产量转录试剂盒、MEGACLEAR™转录纯化试剂盒、ZERO BLUNT® TOPO®PCR克隆试剂盒、PURELINK®Pro Quick96质粒纯化试剂盒、PURELINK™PCR纯化试剂盒、QUBIT®RNA BR分析试剂盒、NEON®转染系统10μL试剂盒、GIBCO®OPTIMIZER™CTS™T细胞扩增SFM、重组人类IL-2(介白素2) CTS™、DYNABEADS™MYONE™抗生蛋白链菌素C1、DYNABEADS™人类T扩增因子 CD3/CD28、DYNABEADS™UNTOUCHED™人类T细胞试剂盒、IgG(总)人类 ELISA试剂盒、多克隆β-肌动蛋白抗体、多克隆表皮生长因子(EGFR)抗体,和Phusion Flash高保真度PCR预混液得自赛默飞世尔科技。Ficoll-Paque PLUS 购自GE Healthcare Life Sciences。NU 7026是从Tocris Bioscience订购的。本研究中使用的DNA寡核苷酸和供体DNA的序列列于表12中。

[0543] gRNA设计

[0544] gRNA合成使用的DNA寡核苷酸和引物是利用GENEART™CRISPR gRNA设计工具设计。然后使用GENEART™精密gRNA合成试剂盒合成gRNA。gRNA的浓度是利用QUBIT®RNA BR分析试剂盒来测定。

[0545] 通过不对称PCR产生长单链DNA

[0546] 首先用正向引物和生物素化的反向引物扩增供体DNA模板。将所得PCR产物(20ng)添加到Phusion Flash高保真度PCR预混液中,所述预混液含有0.2μM正向引物和0.01μM生物素化反向引物,总体积为50μl。配置总共24个反应且使用以下PCR程序:98℃历时30秒,一个循环;然后为98℃历时5秒、55℃历时10秒和72℃历时45秒,总共24个循环。最终延伸物在72℃下培育3分钟。为了去除双链DNA模板,将PCR产物与300μl DYNABEADS™MYONE™抗生蛋白链菌素C1合并并且在室温下、在轻轻旋转下培育20分钟。用磁铁去除磁珠且使用4根柱对上清液进行PURELINK™PCR纯化,然后使用快速蒸发进行浓缩。得到约5μg单链DNA。

[0547] 基因组裂解和检测分析

[0548] 利用GENEART®基因组裂解检测试剂盒(赛默飞世尔科技,目录号A24372),根据制造商说明书测定基因组裂解效率。表12中描述了用于PCR扩增每个基因组基因座的引物序列。转染后48至72小时分析细胞。裂解效率是基于琼脂糖凝胶色带相对强度的计算,所述强度是使用运作ALPHAVIEW® 3.4.0.0版的ALPHAIMAGER®凝胶记录系统(ProteinSimple(美国加州圣何塞(San Jose,CA,USA)))来定量。

[0549] 人类原代T细胞的分离

[0550] 利用Ficoll-Paque PLUS密度梯度,根据制造商说明书,从周边血液中分离出人类周边血液单核细胞(PBMC)。然后使用DYNABEADS™UNTOUCHED™人类T细胞试剂盒分离出人类原代T细胞且使用补充有200IU/mL IL-2的OPTIMIZER™CTS™T-Cell扩增SFM进行扩增。人类T细胞的激活和扩增是使用DYNABEADS™人类T扩增因子CD3/CD28试剂盒进行。在激活的第3天,收获T细胞用于转染。

[0551] 细胞转染

[0552] 使293FT或A549细胞在补充有10%FBS的DMEM培养基中维持。转染当天,将细胞从培养瓶中拆离并计数。每次电穿孔时,将1.5μg Cas9蛋白和360ng gRNA添加到再悬浮缓冲液R中直到7μl的最终体积,但添加的Cas9蛋白加gRNA的总体积小于1μl。混合后,将样品在室温下培育5至10分钟,以形成Cas9 RNP复合物。同时,含有 1×10^6 个细胞的等分试样用不含Ca²⁺和Mg²⁺的DPBS洗涤一次并且将细胞集结粒再悬浮于50μl再悬浮缓冲液R中。然后将细胞悬浮液的5μl等分试样与7μl Cas9 RNP混合,随后添加1μl指定量的供体DNA。将10μl含有Cas9 RNP和供体的细胞悬浮液施加于NEON®电穿孔设备(赛默飞世尔科技,目录号MPK5000)上,相应地,所述电穿孔设备的电压设定为1150V,脉冲宽度设定为20ms且脉冲数设定为2。将电穿孔的细胞转移至含有0.5ml培养基的48孔板中。不含gRNA或供体DNA的样品充当对照。转染后第48小时,利用流式细胞术分析细胞。或者,基因组基因座用相应引物进行PCR扩增。使用GENEART®基因组裂解检测分析对所得PCR片段进行分析。对编辑过的细胞进一步进行有限稀释,随后进行克隆细胞分离。通过N端和C端接头的PCR扩增和测序来表

征克隆细胞。使用 VECTOR NTI ADVANCE® 11.5 软件 (赛默飞世尔科技) 分析测序数据。

[0553] 转染原代T细胞时,每个NEON®电穿孔设备使用 1×10^5 个细胞,相应地,所述设备的电压设定为1700V,脉冲宽度设定为20ms且脉冲数设定为1。为了评估化学修饰对HDR效率的影响,在化学合成过程中,在寡核苷酸的特定位置添加硫代磷酸酯或胺修饰的核苷酸。然后使用所得经修饰的寡核苷酸扩增供体DNA。用NU 7026抑制剂处理细胞时,如上所述转染细胞,然后添加到含有30 μ M Nu 7026的细胞培养基中。转染后第48小时分析细胞。

[0554] 蛋白质标记策略

[0555] 蛋白质标记使研究人员可以目测蛋白质的亚细胞定位并且研究其功能。标记内源细胞蛋白的策略描绘于图1中。无启动子的嘌呤霉素选择标记物通过自裂解的2A肽与报道基因相连。嘌呤霉素基因位于融合蛋白的5'末端(用于N端标记)或3'末端(用于C端标记)。通过PCR扩增将35nt同源臂添加到供体DNA的5'和3'末端。嘌呤霉素的表达由内源启动子驱动,而报道基因与内源基因同框融合。TALEN或CRISPR设计成靶向ATG起始密码子附近的基因组基因座以进行N端标记或靶向终止密码子附近的基因组基因座用于C端标记。然后将所得TALEN或CRISPR以及供体DNA通过脂质介导的转染或电穿孔递送到细胞中。转染后48小时,用嘌呤霉素处理细胞7天,然后通过荧光显微镜目测或通过接合式PCR和测序进行分析。

[0556] N端蛋白质标记实例

[0557] 为了评估标记内源蛋白的策略,我们将OFP基因与 β -肌动蛋白的N端融合。 β -肌动蛋白是真核生物中最丰裕的蛋白质之一,因此很容易使用荧光显微镜进行监测。设计并且合成一种gRNA,其靶向 β -肌动蛋白的靠近ATG起始密码子的基因组基因座(表12),然后与Cas9核酸酶复合而形成RNP。通过PCR扩增将相关的35nt同源臂添加至序列经验证的无启动子嘌呤霉素-P2A-OFP DNA片段。使用PureLink™PCR纯化试剂盒纯化所得供体PCR片段,然后使用快速蒸发浓缩至约1 μ g/ μ l的最终浓度。为了检查供体剂量对HDR效率的影响,我们将Cas9 RNP的量保持恒定,并改变了供体DNA的量。通过电穿孔将Cas9 RNP和供体DNA转染到293FT中。转染后第48小时,通过荧光显微镜分析细胞。当用单独Cas9 RNP或Cas9蛋白和供体DNA转染细胞时,未检测到OFP阳性细胞,而当用Cas9 RNP和供体DNA转染细胞时,未观察到OFP阳性细胞。通过流式细胞术分析测定OFP阳性细胞百分比。在没有选择的情况下,当供体DNA的量从25ng增加到500ng时,OFP阳性细胞的百分比从约5%增加到20%(图2A)。供体DNA的最佳量是每个反应约500 ng。另一方面,用1 μ g/ml嘌呤霉素处理转染的细胞7天后,约80%的细胞是OFP阳性的。在不同量的供体DNA之间,OFP阳性细胞的百分比没有显著差异(图2A)。接下来,我们研究了同源臂长对HDR效率的影响。通过PCR扩增将各种长度的同源臂添加至无启动子的嘌呤霉素-P2A-OFP DNA片段。如图2B所示,当同源臂长度从12nt增加到80nt时,OFP阳性细胞的百分比增加,然后稳定在35nt附近。

[0558] 传统上,质粒供体用于将大的DNA分子并入基因组中。为了比较,我们构建了包含大约500nt同源臂的供体质粒。此外,我们通过不对称PCR制备了具有35nt同源臂的长单链DNA供体。通过电穿孔将Cas9 RNP和各种形式的供体DNA递送到293FT细胞或人原代T细胞中。转染后第48小时,我们使用流式细胞仪分析了OFP阳性细胞的百分比。如图1C和1D中所描绘,在293FT和原代T细胞中,使用具有35nt同源臂的单链(ss)或双链DNA(ds)片段的OFP阳性细胞百分比显著高于使用具有长同源臂的供体质粒的OFP阳性细胞百分比。使用ssDNA供体的效率高于使用dsDNA供体的效率,但它们在293FT中的效率相似。

[0559] 为了检查整合位点的一致性,对经Cas9 RNP和供体DNA转染的细胞进行嘌呤霉素选择、有限稀释和克隆细胞分离。随机挑选总共48个群落用于接合式PCR分析。在48个群落中,只有一个不能生长和产生PCR产物。当使用一种外部引物和一种内部引物时,所有其它47个群落均会产生N端和C端接合的PCR产物。当使用尺寸约为420bp的一对外部引物时,也观察到了PCR产物,其对应于没有插入的基因组DNA片段。未观察到含有插入的大PCR产物的原因是因为没有插入的较小DNA片段被优先扩增。PCR产物的测序分析证实,大约82%的N端接合在基因组DNA和供体DNA之间的接合处显示出精确的HDR(图3A(1))。其它18%的克隆细胞也包含插入,但在接合区有突变(图3A(2))。大多数突变是缺失和插入。部分或全长同源臂的双重复序列有时插入接合点。在C端接合点,大约78%的克隆细胞具有精确的HDR(图3B(1)),而其它22%的细胞在接合点形成有插入缺失(图3B(2))。C端接合点很少缺失相对较大的供体DNA片段(最大165nt)。总体而言,所有克隆细胞均在一个等位基因的正确基因组基因座包含一个供体DNA拷贝,其中68%的细胞在N端和C端都具有精确的HDR,而32%的细胞在N端或C端或两端都具有不完善的HDR。其它等位基因不包含任何插入。取而代之的是,大约80%的克隆在Cas9裂解位点有一个“A”插入,而20%的克隆则具有超过2nt的缺失。在第二个等位基因上仅检测到一个野生型克隆(图3C)。如蛋白质印记分析所证实,大多数克隆表达野生型 β -肌动蛋白和 β -肌动蛋白的OFP融合体。

[0560] TALEN(TAL效应物核酸酶)是在哺乳动物基因组中引入双链断裂的另一种方法。设计并且合成了三对TALEN mRNA,其靶向 β -肌动蛋白ATG密码子附近的区域。使用1150伏、20毫秒(ms)和两个脉冲,经由NEON®电穿孔设备将单独的TALEN mRNA或TALEN mRNA与供体DNA转染到HEK293FT细胞中。转染后第48小时,使细胞溶解以测量基因组编辑效率(图3D)或通过流式细胞术分析(图3E)以测定OFP阳性细胞(-)的百分比。或者,在流式细胞术分析(+)之前,用嘌呤霉素处理细胞7天(图3E)。如图3D中所描绘,虽然T1和T3目标产生约60%和35%插入缺失频率,但在无嘌呤霉素选择的情况下,OFP阳性细胞的百分比极低。然而,在嘌呤霉素选择后,对于所有三种不同的目标来说,OFP阳性的百分比上升至约60%(图3E)。

[0561] 除了 β -肌动蛋白之外,我们还评估了不同细胞系中的不同蛋白质。LRRK2蛋白与帕金森氏病(Parkinson's disease)有关,分子量约为280kd。设计靶向起始密码子附近的LRRK2基因组基因座的gRNA。通过PCR扩增,将约35nt的同源臂添加至序列经验证的无启动子嘌呤霉素-P2A-EmGFP DNA片段。使用1050伏、30毫秒和2个脉冲,经由NEON®电穿孔设备将Cas9 RNP和供体DNA共递送到A549细胞中。由于LRRK2是丰度相对较低的蛋白质,因此我们无法检测到细胞内的EmGFP信号。一些商业抗体也无法通过蛋白质印迹法在全细胞溶解物中检测出内源性野生型LRRK2蛋白。为了检查整合效率,在转染后第48小时用0.75 μ g/ml嘌呤霉素处理细胞7天,然后进行有限稀释和克隆细胞分离。使用一个内部引物和一个外部引物或一对外部引物,通过PCR分析接合点。通过测序分析所得PCR产物以测定整合精度。令人惊讶的是,所有86个群落都含有插入片段的至少一份拷贝物。对于所有群落来说,N端和C端都具有精确的HDR,基因组DNA和供体DNA之间的接合正确(图4A和4B)。分离基因组DNA后,我们能够检测到杂合子的两种PCR产物和纯合子的一种大PCR产物。基于测序分析,约20%群落在两个等位基因中均具有供体DNA的精确整合,而其余80%群落中的第二等位基因不包含任何插入片段,而是仅具有7nt缺失(图4C)。这些结果表明,可以实现100%整合效率和100%精确HDR。

[0562] C端蛋白质标记的实例

[0563] C端蛋白标记的启动子捕获策略与N端蛋白标记的启动子捕获策略稍有不同之处在于,用于C端标记的无启动子选择标记物位于报道基因之后,而用于N端标记的无启动子选择标记物位于报道基因之前(图1)。作为实例,我们将EmGFP标签与粘着斑激酶(FAK)的C端融合。设计并合成靶向终止密码子附近的FAK基因组基因座的gRNA(表12)。通过PCR将短同源臂添加到序列经验证的EmGFP-2A-嘌呤霉素盒中。经由**NEON®**电穿孔设备将Cas9 RNP和供体DNA递送到293FT细胞中。转染后第48小时,用0.75μg/ml嘌呤霉素选择细胞7天,然后进行有限稀释和克隆细胞分离。通过PCR和测序分析接合点。如图5A和5B所描绘,约95%和85%的克隆分别在N端或C端具有正确接合点。其它克隆也包含插入盒,但是在接合点或Cas9裂解位点形成有插入缺失。再次,我们观察到部分或全长同源臂的重复序列被插入基因组。总体而言,所有检查的克隆均包含至少一个供体DNA拷贝,其中约70%的克隆在N端和C端均具有精确的HDR,而其它30%的克隆在一个等位基因中含有不精确的HDR。约30%克隆细胞的两个等位基因中均整合有供体。约70%细胞在第二个等位基因处没有插入片段,但在Cas9裂解位点的接合点形成有插入缺失。在第二个等位基因中仅检测到一个野生型克隆(图5C)。

[0564] 除了FAK之外,我们还检查了其它蛋白质,例如表皮生长因子受体(EGFR)。EGFR有几种同工型。在这项研究中,我们将EmGFP与EGFR同工型1的C端融合。gRNA设计成使终止密码子附近的EGFR基因组基因座裂解。通过PCR将短同源臂添加至插入盒。经由电穿孔将Cas9 RNP和供体DNA送入293FT细胞中。嘌呤霉素选择后,对细胞进行克隆分离。令人惊讶的是,所有19个群落在一个等位基因上都具有一个插入盒,N端和C端的接合100%正确。约17%群落发生了双等位基因整合,而83%群落在第二个等位基因上不含插入,而仅在Cas9裂解位点处具有“A”插入(图6)。通过蛋白质印记法检测EmGFP对EGFR的基因组修饰。

[0565] DNA供体的末端修饰和NHEJ抑制剂对HDR效率的影响

[0566] 线性ds-DNA或ss-DNA供体能够在体内被外切核酸酶降解。供体DNA的末端修饰可能能够防止它们降解。为了检验这种假设,化学合成了在5'末端具有不同修饰的DNA引物(表12)。然后使用经修饰的DNA引物,通过PCR扩增来制备含有无启动子嘌呤霉素-P2A-OFP片段的供体DNA。所得PCR产物使用PURELINK™PCR纯化试剂盒加以纯化,随后使用快速蒸发进行浓缩。通过电穿孔将靶向β-肌动蛋白基因组基因座的Cas9 RNP与各种形式的供体DNA共递送到原代T细胞中。转染后第48小时,通过流式细胞术分析测定OFP阳性细胞的百分比。如图7A所述,与未修饰的供体DNA相比,经硫代磷酸酯修饰的DNA供体使HDR效率提高了约2倍。有趣的是,胺修饰的供体也提高了HDR效率,尤其是当修饰发生在修饰反义链5'末端的反向引物上时。使用两端均经过胺修饰的供体DNA,OFP阳性细胞的百分比增加了约4倍。ssDNA供体的末端修饰也提高了HDR效率。但是,使用胺修饰的dsDNA供体的效率比使用经修饰的ssDNA供体的效率高约2倍。

[0567] 已知NHEJ修复途径的破坏可提高HDR效率。在这里,我们检查了那些NHEJ抑制剂如何影响相对大的DNA分子整合到人类原代T细胞中。经由电穿孔将Cas9 RNP和供体DNA送入原代T细胞后,我们立即将细胞转移到含有30μM Nu7026的培养基中。转染后第48小时,我们通过流式细胞仪分析了细胞。如图7B所示,用Nu7026处理细胞使OFP阳性细胞的百分比对于未修饰的供体DNA而言增加约5倍,而对于胺修饰的供体DNA而言增加2倍。其它DNAPK抑制

剂(包括Nu7441和Ku-0060648)获得了相似的结果。

[0568] 潜在应用

[0569] 使用上述方法,我们可以容易地以接近100%的整合效率将一大片DNA 整合到哺乳动物基因组中,从而使研究人员可以将感兴趣的外来DNA直接克隆到哺乳动物基因组中并表达蛋白质以用于治疗应用。

[0570] 表达盒实例

[0571] 作为实例,我们制备约4.2kb的人类IgG表达盒,其含有无启动子选择标记物、细胞巨大病毒(CMV) 启动子、IgG重链、IgG轻链和WPRE (土拨鼠肝炎病毒转录后调节元件)。CMV启动子驱动IgG重链和轻链的表达,其通过 2A自裂解肽连接(图8A)。通过PCR将35nt短同源臂添加至表达盒中,随后进行PCR柱纯化。如上所述,将表达盒插入293FT细胞中的 β -肌动蛋白基因座。嘌呤霉素选择7天后,我们使用ELISA分析测量了稳定细胞池中的IgG 产生的滴度。作为对照,将含有IgG重链和轻链表达盒的质粒DNA瞬时共转染到细胞中。转染后第5天收获培养基。工程化细胞池中的IgG的表达水平为约0.5克/升,而瞬时质粒表达系统中的IgG水平为约0.3克/升。

[0572] 为了表征稳定池中的每个克隆细胞,我们进行了有限稀释和克隆细胞分离。通过PCR和测序分析整合的接合点。如图8B和8C中所描绘,约88%的克隆细胞在N端接合点具有精确整合,而12%的克隆细胞在接合点插有一些额外的序列。另一方面,约41%的克隆细胞在C端具有正确的接合,而59%的克隆细胞在接合点具有小的突变。例如,我们观察到碱基取代,在WPRE聚腺苷酸尾部区域存在一个或几个核苷酸插入。小突变发生在终止密码子之后,可能不会影响IgG的表达。为了证实这一点,我们检查了每个克隆细胞的IgG 滴度。如图8D中所示,约70%的克隆细胞能够产生抗体。

[0573] 在这项研究中,内源蛋白被标记。嵌合蛋白的表达水平取决于细胞内内源蛋白的丰度。对于丰裕的蛋白质(例如 β -肌动蛋白)来说,使用常规的宽视野荧光显微镜容易检测到嵌合融合蛋白。但是,对于丰度低的蛋白质(例如 LRRK2)来说,常规的宽视野荧光显微镜不足以进行检测。使用高分辨率荧光技术,例如荧光共振能量转移(FRET)和连续波超声可切换荧光(CW-USF),可以使活细胞内部的荧光分子可视化,并具有改善的时空分辨率(Sekar等人,“活细胞蛋白质定位的荧光共振能量转移(FRET)显微镜成像(Fluorescence resonance energy transfer(FRET) microscopy imaging of live cell protein localizations)”,《细胞生物学杂志(J.Cell Biol.)》160:629-33(2003)和Cheng 等人,“经由新型造影剂、以高信噪比和高灵敏度在厘米深的组织体模中的高分辨率超声波可切换荧光成像(High-resolution ultrasound-switchable fluorescence imaging in centimeter-deep tissue phantoms with high signal-to-noise ratio and high sensitivity via novel contrast agents)”PLoS One.11:e0165963 (2016))。虽然尚未完全理解为什么一种等位基因中的嵌合蛋白的表达水平显著低于另一种野生型等位基因,但是当将转基因插入基因组时,某些转录或翻译调控元件可能会被破坏。

[0574] 实例2:通过将核定位信号附着到供体DNA来提高哺乳动物细胞中的基于同源性的编辑速率。

[0575] 假定将供体DNA(单链或双链、线性或环状)递送至细胞核将增加供体 DNA在编辑发生处附近的局部浓度,且因此使修复偏向于此供体DNA而非 NHEJ。

[0576] Zanta, M.A. 等人, 《美国国家科学院院刊 (Proc. Natl. Acad. Sci. (USA))》96:91-96 (1999) 证明与DNA区段结合的NLS可以增加DNA区段向核的递送。因此有理由认为, 可以使用类似的方法来增强供体ssDNA向核的递送, 并且供体DNA在核内的增加可能会增加供体DNA在“切割位点”的整合频率。

[0577] 对于NLS来说, 使用进化的SV40 NLS (BP-SV40, KRTADGSEFESPKKKRKVEGG) (SEQ ID NO:13)。Hodel, M.R. 等人, 《生物化学杂志》276 (2):1317-1325 (2001) 报道此序列有效地定位于细胞核。使用琥珀酰亚氨基4-(N-马来酰亚氨基甲基) 环己烷-1-甲酸酯 (SMCC) 或 **Click-iT®** 化学物质, 以使NLS肽与ssDNA供体序列结合。所得NLS-寡核苷酸结合物通过HPLC纯化。NLS-寡核苷酸的质量通过MALDI-TOF测定。制备如图9所示的两种构建体。如图10所示, 这些供体DNA允许通过荧光进行筛选。

[0578] 第1部分: 使用NLS结合的寡核苷酸供体将缺失6个碱基的GFP转化为功能性GFP。

[0579] 经由**Click-iT®** 化学物质使NLS肽BP-SV40 (SEQ ID NO.:13) 的羧基末端与寡核苷酸的5'末端结合:

[0580] 5'CGGGGTAGCGGCTGAAGCACTGCACGCCGTAGGTCAGGGTGGTCA CGAG

[0581] GGTGGGCCAGGGCACGGGCAGCTTGCCGGTGGTGCAGATGAACTTC AG-3' (SEQ ID NO.:14)。

所得NLS-寡核苷酸结合物通过HPLC纯化。NLS-寡核苷酸的质量通过MALDI-TOF测定。

[0582] 在转染之前的当天, 以每孔 1×10^5 个的细胞密度将破碎的 EmGFPGRIPITITE™293细胞系接种于24孔板上。转染当天, 将0.5μg Cas9 mRNA和150ng靶向断裂的EmGFP基因的gRNA (GCACGCCGTAGGTGGTCACGAGG) (SEQ ID NO.:15) 添加到无菌试管中的25μl **OPTI-MEM®** 中。将NLS-寡核苷酸结合物溶于水中, 并且将不同数量的NLS-寡核苷酸添加到装有Cas9和gRNA的试管中。硫代磷酸酯修饰 (PS) 的寡核苷酸用作对照, 其在寡核苷酸的5'末端具有两个硫代磷酸酯且在3'末端具有两个硫代磷酸酯。在单独的试管中, 将1.5μl的 LIPOFECTAMINE™ MESSENGERMAX™ 添加到25μl **OPTI-MEM®** 培养基中。然后将稀释的 LIPOFECTAMINE™ MESSENGERMAX™ 转移到含有Cas9、gRNA和指定量的NLS-寡核苷酸或PS-寡核苷酸的试管中。在室温下培育5分钟后, 将混合物添加到含有0.5ml生长培养基的24孔中。转染后第48小时, 通过流式细胞仪分析细胞, 以测定EmGFP阳性细胞的百分比。

[0583] 如图11所示, NLS-供体使得细胞系的编辑明显更高。多达52%的细胞在 0.1皮摩尔NLS-供体的最优剂量下呈GFP阳性, 相比之下, 标准PS-供体需要 3皮摩尔才能达到30倍更多材料所达到的最高36%的最优编辑。图12展现 NLS-供体在0.03皮摩尔的相同低剂量下, 向GFP+细胞的转化高得多。总之, 发现编辑的转化更高, 如NLS-供体在低得多的剂量下、根据GFP阳性细胞所测量。

[0584] 第2部分: 通过使用NLS结合的寡核苷酸供体改变单个碱基而将BFP转化为功能GFP。

[0585] 通过SMCC化学物质使NLS肽BP-SV40 (SEQ ID NO.:13) 的羧基末端与寡核苷酸的5'末端结合:

[0586] 5'-GCTGCCCCGTGCCCTGGCCACCCTCGTGACCACCCTGACCT

[0587] ACGGCGTGCACTGCTTCAGCCGCTACCCCGACCACATGA-3'

[0588] (SEQ ID NO.:16)。所得NLS-寡核苷酸结合物通过HPLC纯化。NLS-寡核苷酸的质量通过MALDI-TOF测定。

[0589] 转染前一天,将eBFP 293FT稳定细胞系以每孔 1×10^5 的细胞密度接种在24孔板上。转染当天,将 $0.5 \mu\text{g}$ Cas9 mRNA和 150 ng 靶向eBFP基因的gRNA (CTCGTGACCACCCTGACCCACGG) (SEQ ID NO:17)添加到无菌试管中的 $25 \mu\text{l}$ Opti-MEM®中。将NLS-寡核苷酸溶解在水中,并且将不同数量的NLS-寡核苷酸添加到装有Cas9和gRNA的试管中。未修饰的寡核苷酸用作对照。在单独的试管中,将 $1.5 \mu\text{l}$ 的LIPOFECTAMINE™MESSENGERMAX™添加到 $25 \mu\text{l}$ Opti-MEM®培养基中。然后将稀释的LIPOFECTAMINE™MESSENGERMAX™转移到含有Cas9、gRNA和指定量的NLS-寡核苷酸或未修饰的寡核苷酸的试管中。在室温下培育5分钟后,将混合物添加到含有 0.5 ml 生长培养基的24孔中。转染后第48小时,通过流式细胞术分析细胞以测定GFP阳性细胞的百分比。

[0590] 如图13所示,NLS-供体再次使得细胞系的编辑明显更高。在 0.3 皮摩尔的最优剂量下,多达76%的细胞从BFP转化为GFP阳性,相比之下,在 10 皮摩尔的对照PS寡核苷酸存在下,则为58.5%。再次发现NLS-供体在30倍更低剂量下,存在更高的编辑。随着剂量的降低,用NLS-寡核苷酸可以维持较高的编辑水平,其中在 0.01 皮摩尔下,21%细胞被编辑,相比之下,在对照PS寡核苷酸存在下,在 0.03 皮摩尔下,6%细胞被编辑。

[0591] 本文描述的方法在细胞工程、细胞疗法和生物生产等方面具有广泛的应用。与瞬时质粒表达不同,相对较大的表达盒能够直接插入基因组中的特定基因座用于生物生产。可以使用具有所需相对强度的内源启动子靶向安全港区域。重复区域也可能潜在地成为目标,以合并有效载荷的多个拷贝以获得更高的表达水平。独立于选择标记物,强启动子可用于驱动所关注的外来基因的表达。由于高整合效率和特异性,因此稳定的细胞池可直接用于蛋白质生产,而无需分离克隆细胞,从而节省了时间和成本。在一些实施例中,这种方法用于在ExpiCHO细胞中产生重组抗体。

[0592] 实例3. 目标dsDNA断裂邻近的DNA结合可促进染色质的置换和/或DNA展开,并且促进设计核酸酶的进入改善。

[0593] “TAL-Buddy”由18个重复的TAL结合子组成。“TAL-Buddy”设计成紧邻设计核酸酶结合区,每一侧一个(左=Lt,右=Rt,TALEN对和TAL-Buddy结合序列列于表12)。“TAL-Buddy”如下制备:使用BsaI、经由Golden Gate 组装反应来组装含有T7启动子和转录/翻译起始元件的N端片段与TAL的氨基端片段、六TAL RVD三聚体,以及含有C端域、核定位信号和终止密码子的C端片段(图14中所示)。SEQ ID NO:35中列出了“TAL-Buddy”(CMPK1-TALEN2_7nt_TAL-Buddy_Lt)核苷酸序列的实例。CMPK1-C目标的相邻基因组序列示于SEQ ID NO:36;并且TALEN和TAL Buddy的相对位置示于SEQ ID NO:20和SEQ ID NO:21。此实例的进一步描述提供于图14-18、22和32-36。

[0594] 全长“TAL-Buddy”是使用引物对TD1-F2和TD8-R2 (SEQ ID NO:22-23)、通过扩增来富集,并且进一步用作模板以便使用mMESSAGE mACHINE™T7 ULTRA转录试剂盒(赛默飞世尔科技)制备mRNA。添加 0 、 25 、 50 或 100 ng 的Lt和Rt“TAL-Buddy”mRNA以及 100 ng 的TALEN mRNA对,以便使用NEON®电穿孔设备(赛默飞世尔科技)、以 1300 脉冲电压、 20 脉冲宽度和 2 次脉冲数转染到 $\sim 50,000$ 个293人类胚肾细胞(293FT)中。转染后48至72小时收获细胞并且溶解。使用GENEART™基因组裂解检测试剂盒(赛默飞世尔科技,目录号A24372)分析插入缺失的形成。(图15)

[0595] 可以用于组装TAL的一类方法是Golden Gate方法。

[0596] 在Golden Gate中,组装和克隆是基于具有“粘性”末端的核酸区段的产生,所述核酸区段是通过用一种或多种II型限制性内切核酸酶裂解而产生的,通常随后将组装好的核酸分子引入合适的宿主细胞中。使用II型限制性内切核酸酶是因为它们识别不对称序列并且在距识别位点的限定距离处使这些序列裂解。此外,DNA分子的末端能够设计为侧接II型限制位点,使得片段的消化除去酶识别位点并产生互补的突出物。此类末端能够无缝接合,从而形成缺少原始位点或疤痕的接合点。

[0597] 此外,II型限制性内切核酸酶可以并且已经用于产生TAL效应物的重复区域。II型限制性内切核酸酶也可用于将合适的末端蛋白质编码核酸连接到 TAL效应物重复区的侧面,并将TAL效应物编码区连接到其它核酸分子(例如载体,其中TAL效应物编码核酸可操作地连接到启动子)。II型限制性核酸内切酶TAL效应物组装方法阐述于例如Morbiter等人,“通过模块化克隆来组装定制的TALE型DNA结合域(Assembly of custom TALE-type DNA binding domains by modular cloning)”,《核酸研究(Nucleic Acids Res.)》39:5790-9(2011)。

[0598] 结果:当“TAL-Buddy”设计成相对于TALEN结合序列间隔7nt(即,相对于TALEN裂解位点为33nt)时,CMPK1-C目标处的插入缺失形成增加约2倍(图15)。

[0599] 实例4.“TAL-Buddies”设计成相对于TALEN结合序列具有不同的间距(表12)并且使用实例3中所述的相同方法在293FT细胞中进行测试。

[0600] 结果:当相对于TALEN结合序列间隔7-30nt时,“TAL-Buddy”起作用(图16)。当TAL-buddies相对于TALEN距离4到30nt时,TALEN切割出现最佳的增强。将TAL-buddies紧邻着TALEN或距离其大于50nt,将不会增强TALEN切割(图16和图17)。

[0601] 实例5.紧邻着靶向UFSP2-SNP位点的CRISPR sgRNA的“TAL-Buddy”设计成相对于CRISPR sgRNA结合序列的间距为7nt或20nt。

[0602] UFSP2-SNP目标的基因组序列列于SEQ ID NO:25和SEQ ID NO:43。

[0603] 结果:当“TAL-Buddy”设计成相对于性能不良的CRISPR sgRNA结合序列的间距为7nt或20nt(即,相对于CRISPR裂解位点分别为23nt和37nt)时,插入缺失的形成增加10至20倍。结果示于图17中。

[0604] 实例6.为了最小化野生型SpCas9的脱靶效应,测试突变形式

[0605] 通过增强DNA目标基因座的可及性,可以增强以下蛋白质的活性:性能不良的cas9蛋白(例如Kleinstiver,Benjamin P.等人所述的HiFi Cas9(“全基因组脱靶效应检测不到的高保真度CRISPR-Cas9核酸酶(High-fidelity CRISPR- Cas9 nucleases with no detectable genome-wide off-target effects)”,《自然》(2016).PubMed PMID: 26735016;结合经修饰的PAM的Cas9蛋白和其它直系同源Cas9蛋白,例如来自普雷沃氏菌属(Prevotella)和弗朗西斯氏菌属(Francisella)的CRISPR 1(Cpf1)。本领域中通常已知和描述的任何突变Cas9形式均可用于本文提供的方法和组合物。预期用于本文所提供的方法和组合物中的突变型Cas9蛋白的非限制性实例描述于Slaymaker,Ian M.等人(“合理地经工程改造的具有改善的特异性的Cas9核酸酶(Rationally engineered Cas9 nucleases with improved specificity)”《科学(Science)》(2015):aad5227.PubMed PMID: 26628643)和Kleinstiver,Benjamin P.等人(“全基因组脱靶效应检测不到的高保真度

CRISPR-Cas9核酸酶”《自然》(2016).PubMed PMID:26735016),所述文献以全文引用的方式并入用于所有目的。这两种突变形式的在靶裂解效率也受到损害。对相对于sgRNA结合序列的间距为20nt的“TAL-Buddy”与RNP的组合进行测试,所述RNP用sgRNA和eSpCas9或SpCas9-HF1形成。添加100ng的Lt和Rt“TAL-Buddy”mRNA以及CRISPR-RNP(1000ng的SpCas9-HF1或eSpCas9蛋白和200ng sgRNA),以便利用NEON®电穿孔设备(赛默飞世尔科技)、以1150脉冲电压、20脉冲宽度和2次脉冲数转染到约50,000个293人类胚肾细胞(293FT)中。转染后48至72小时收获细胞并且溶解。使用GENEART™基因组裂解检测试剂盒(赛默飞世尔科技,目录号 A24372)分析插入缺失的形成。

[0606] 结果:当添加相对于sgRNA结合序列的间距为20nt的“TAL-Buddy”时,使用sgRNA和SpCas9-HF1或eSpCas9形成的CRISPR-RNP分别获得5倍和14倍的插入缺失形成(图18)。

[0607] 实例7.长度为15nt的截短gRNA(“CR-PAL”)已显示dsDNA结合活性,但当存在野生型Cas9时则没有裂解活性。

[0608] 用于制备sgRNA和“CR-PAL”的模板的架构图示于图19中。CR-PAL的功能图示于图20中。通过体外转录设计并制备了CRISPR裂解位点附近的15聚体gRNA(“CR-PAL”)。全长sgRNA结合序列的基因组DNA序列和相对位置在SEQ ID NO:44和SEQ ID NO:45中列出。图19和图20。

[0609] 结果:对于左(Lt)和右(Rt)CR_PAL(图21和图34)来说,插入缺失的形成均增加了超过60倍。

[0610] 实例8:Cas9 NLS变异体

[0611] 对Cas9 v2(BPsv40标签/核质蛋白)、IDT(目录号1074181)与Cas9 v1(-/3xsv40)在A549细胞中针对两个目标(HPRT和PRKCG)进行比较,进行4倍连续稀释以确定蛋白质浓度如何影响功能执行。HPRT被认为容易修饰目标,而PRKCG则更难以修饰。

[0612] 使用1μg Cas9蛋白(来自各种来源)和250ng gRNA(HPRT或PRKCG)形成RNP复合物。培育10分钟之后,通过在适当体积的OPTI-MEM™中稀释初始浓度来制备RNP复合物的4倍连续稀释液。根据手册将每种连续稀释液与LIPOFECTAMINE™CRISPRMAX™混合且然后添加到约50,000个293FT细胞中。使经转染的细胞生长3天,并通过基因组裂解检测分析来测量编辑效率。

[0613] 还测试了spy Cas9骨架变异体的多种不同形式(参见图43,数据未示出),其中N或C端添加有各种NLS或亲和标签。

[0614] 在图44所示数据呈现的三种形式中,Cas9 v2在稀释范围内具有显著最高的活性。

[0615] 实例9:TALEN裂解和同源定向修复效率

[0616] 下表8中阐述了下文针对目标所示的TALEN设计,其用于产生图48-50中所列的数据。下表9-11阐述了用于产生图48-50的数据。

[0617]	表 8	SID ID NO
--------	-----	--------------

[0618]	HTR2A_3目标:		
	TALEN-F结合		
	TGTTTTGCTG	ACTTCAAAAA	119
	ACAAAACGAC	TGAAGTTTTT	120
	GACGTACGTT	CTCGACTCGG	
	TALEN-R结合		
	EFEMP1_4目标:		
	TALEN-F结合		
	AGTTAGGAAA	AGGGCTTTCA	121
	TCAATCCTTT	TCCCGAAAGT	123
	CLRN1_2目标:		
	TALEN-F结合		
	GCCCTGAGGC	ATTGACGAGC	124
	CGGGACTCCG	TAACTGCTCG	125
		TALEN-R结合	

[0619] 对于96孔培养板中所生长的各50,000个细胞来说,100ng正向和100ng 反向TALEN mRNA和/或10皮摩尔供体单链寡核苷酸在中部含有6个核苷酸 HindIII识别位点且在5'末端和3'末端含有35个核苷酸同源臂。5'和3'末端的两个远端核苷酸均具有硫代磷酸酯键,以防止核酸酶降解。

[0620] 转染当天,预备的细胞如下制备:(1)计算所需的细胞总数(各50,000个细胞);(2)拆分细胞且对细胞数进行计数;(3)以1,000rpm将所需数目个细胞离心5分钟;(4)细胞集结粒用DPBS洗涤一次,然后以1,000rpm离心5分钟;(5)将细胞集结粒再悬浮于每50,000个细胞5μl的NEON®再悬浮缓冲液R(赛默飞世尔科技,NEON®转染系统,100μL试剂盒,目录号MPK10096);以及(6)将100ng正向TALEN引物、100ng反向TALEN引物、10pmol供体单链寡核苷酸和5μl R缓冲液添加到各5μl含有细胞的R缓冲液中。

[0621] 使用10μl NEON®移液管进行电穿孔(赛默飞世尔科技,目录号 MPK5000)。电穿孔条件如下:对于293FT细胞而言,1300(脉冲电压)、20(脉冲宽度)、2(脉冲数);对于U2OS细胞而言,1400(脉冲电压)、20(脉冲宽度)、2(脉冲数);对于A549细胞而言,1150(脉冲电压)、30(脉冲宽度)、2(脉冲数)。

[0622] 然后将电穿孔的细胞转移到96孔培养板中的100μl预热生长培养基中。转染后48-72小时收获细胞且使用GENEART®基因组裂解检测试剂盒(赛默飞世尔科技,目录号A24372)分析裂解效率,且利用HindIII消化来测定HDR 效率。

表 9: 图 48 GCD 数据

基因	细胞类型	Ave	SD	Ave	SD
		V1		V3	
HTR2A-N_3	U2OS	62.675	0.304	58.900	3.620
HTR2A-N_3	293FT	57.895	2.482	66.830	0.608
HTR2A-N_3	A549	45.780	1.513	57.675	1.435
EFEMP1-N_4	U2OS	24.830	2.885	58.665	2.001
EFEMP1-N_4	293FT	15.630	1.669	61.725	3.486
EFEMP1-N_4	A549	28.01	1.400	54.420	1.824
CLRN1-SNP_2	U2OS	1.465	0.477	36.535	3.048
CLRN1-SNP_2	293FT	6.905	2.368	46.070	1.499
CLRN1-SNP_2	A549	6.470	0.622	34.615	0.615

表 10: 图 49 HDR 数据

基因	细胞类型	Ave	SD	Ave	SD
		V1		V3	
HTR2A-N_3	293FT	16.570	0.3818	22.000	2.164
	U2OS	24.540	3.691	22.000	1.365
EFEMP1-N_4	293FT	2.635	0.417	33.755	0.615
	U2OS	6.775	0.912	31.040	0.990
CLRN1-SNP_2	293FT	1.970	0.976	22.400	2.461
	U2OS	0.250	0.0990	15.415	0.728

表 11: 图 50 HDR 数据 (A549 细胞)

基因	Ave		SD	
	V1	V3	V1	V3
HTR2A-N_3	4.485	7.925	0.557	0.304
EFEMP1-N_4	0.855	11.785	0.926	1.874
CLRN1-SNP_2	0	6.975	0	0.167

[0625] 例10: 使用Cas9一步敲入原代T细胞

[0626] 使用FICOLL-PAQUE™PLUS培养基和INVITROGEN™ DYNABEADS™UNTOUCHED™人类T细胞试剂盒,从健康供体的 LeukoPak血液产物中分离出人原代T细胞。然后在含有2%人血清的GIBCO™ OPTIMIZER™CTS™T细胞扩增培养基中培养细胞,并且用Gibco DYNABEADS人T扩增因子CD3/CD28激活。这项研究中的所有实验均在激活后的第三天进行。

[0627] 激活后的第三天,用存在于电穿孔缓冲液中的以下各物转染细胞:1μg TRUECUT™Cas9蛋白v2(赛默飞世尔科技)、240ng靶向β-肌动蛋白基因的第一外显子的向导RNA,和300ng或1μg供体DNA(橙色荧光蛋白(OFP) 基因与嘌呤霉素基因融合的1.4kb PCR产物OFP-Puro)。使用NEON®转染系统进行电穿孔。所用程序是:程序16(1400V,20ms,2个脉冲);程序23(1500V,10ms,3个脉冲);和程序24(1600V,10ms,3个脉冲)。

[0628] 转染后第二天,通过FACS分析对细胞的侧向散射区(SSC-A)和OFP进行分析。SSC-A是基于尺寸来识别细胞。

[0629] 如图51A中所示,使用1μg供体DNA的敲入效率为7.71%、5.46%或6.99%,此取决于所用NEON®程序。敲入效率取决于供体DNA浓度(图51B)。程序16在测试条件下产生了最高的敲入效率。

[0630] 实例11:在NHEJ抑制剂存在下,使用Cas9一步敲入原代T细胞

[0631] 如同实例10分离并激活人原代T细胞。激活后第三天,细胞用存在于电穿孔缓冲液中的以下各物转染:1μg TRUECUT™Cas9蛋白v2、240ng靶向β-肌动蛋白基因的第一外显子的

向导RNA (β -肌动蛋白gRNA) 和1 μ g OFP- Puro。使用**NEON®**转染系统、程序16进行电穿孔。电穿孔后,在含有NHEJ 抑制剂(Nu7026、Nu7441或Ku0060648) 的培养基中培养细胞两天,然后通过FACS分析对SSC-A和OFP进行分析。

[0632] 如图52A和52B所示,用NHEJ抑制剂培养细胞以剂量依赖性方式提高敲入效率。细胞的存活率不受影响(图52C)。不管T细胞供体如何,敲入效率均增加(图52D)。

[0633] 实例12:在含有NHEJ抑制剂的电穿孔缓冲液存在下,使用Cas9一步敲入原代T细胞

[0634] 如同实例10分离且激活人原代T细胞并且如同实例11进行转染,例外之处为电穿孔之前,将NHEJ抑制剂(Ku0060648)添加到电穿孔缓冲液中(为培养基中所用浓度的50倍)。电穿孔后,将细胞、各组分和电穿孔缓冲液直接添加到细胞培养基中,并将细胞培养两天,然后通过FACS分析对SSC-A和 OFP进行分析。

[0635] 观察到敲入效率出现剂量依赖性增加(图53A和53B)。细胞的存活率不受影响(图53C)。

[0636] 实例13:在NHEJ抑制剂存在下,使用两种Cas9/gRNA复合物一步敲入原代T细胞

[0637] 如同实例10分离且激活人原代T细胞。激活之后的第三天,用存在于电穿孔缓冲液中的1 μ g TRUECUT™Cas9蛋白v2与250ng β -肌动蛋白gRNA、1 μ g TRUECUT™Cas9蛋白v2与250ng靶向T细胞受体的向导RNA (TRAC gRNA) 以及1 μ g OFP-Puro转染细胞。使用**NEON®**转染系统、程序24进行电穿孔。电穿孔后,将细胞在含有NHEJ抑制剂(Ku0060648) 的培养基中培养两天,然后通过FACS分析对SSC-A和OFP进行分析。

[0638] 如图54所示,用Ku0060648培育细胞使得敲入效率提高(Q3),而不会影响敲除效率(Q3和Q4)。

[0639] 实例14:在NHEJ抑制剂存在下使用Cas9进行的一步敲入使多种细胞系中的敲入效率增加

[0640] 用存在于电穿孔缓冲液中的1 μ g TRUECUT™Cas9蛋白v2和250ng β -肌动蛋白gRNA以及0.25 μ g、0.5 μ g或1 μ g OFP-Puro转染多种细胞类型(NK- 92、THP-1、Jurkat)。使用**NEON®**转染系统进行电穿孔。电穿孔后,将细胞在含有NHEJ抑制剂(Ku0060648) 的培养基中培养两天,然后通过FACS分析对SSC-A和OFP进行分析。

[0641] 转染后将每种细胞类型与Ku0060648一起培育使每种细胞系中的敲入效率提高。NK-92的结果提供于图55A中;THP-1的结果提供于图55B中;且 Jurkat的结果提供于图55C中。

[0642] 实例15:在NHEJ抑制剂存在下,使用Cas9一步敲入iPSC

[0643] 使用以下文献中所述的方案,用针对指定目标的向导RNA、校正指定SNP 的供体DNA (长度约100bp的寡核苷酸) 和1 μ g TRUECUT™Cas9蛋白v2 转染诱导型多能干细胞(iPSCs;Gibco):“通过电穿孔进行CRISPR-Cas9基因组编辑以便研究StemFlex™培养基中所培养的人类多能干细胞(CRISPR-Cas9 Genome Editing for Research of Human Pluripotent Stem Cells Cultured in StemFlex™Medium via Electroporation)”,赛默飞世尔科技,公开号 MAN0016956 Rev. A.0,2017年4月20日(获得于assets.thermofisher.com/TFS- Assets/LSG/manuals/MAN0016956_CRISPR_Cas9_GenomeEditing_StemFlex_via_Electropor_UB.pdf)。目标是LRRK2、NAV1.5、TNNT2和SNCA基因。

[0644] 如图56A和56B所示,对敲入效率的影响是抑制剂和目标特有的。

[0645] 实例16:在多种NHEJ抑制剂存在下,使用Cas9一步敲入T细胞

[0646] 如实例10中所述制备T细胞,并且如实例12中所述,通过电穿孔用Cas9、gRNA和供体DNA转染。电穿孔后,将细胞在包含Nu7026、Nu7441和/或 Ku0060648中的一种或多种的培养基中培养,如图57A所示。来自两种不同的T细胞供体的数据提供于图57B和57C。

[0647] 氨基酸和核苷酸序列的描述

[0648] 表12提供了本文引用的某些序列的清单。

[0649] 表12:本文所提及的各种核苷酸和氨基酸序列

[0650]

描述	序列	SEQ ID NO
FAK	CTCGATGTCATTGACCAAGCAAGACTGAAAATGGTGAGCAAGGGCGAGGAGCT GTTACCGGGGTGGTGCCCATCTGGTCGAGCTGGACGGCGACGTAAACGGCCA CAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGA CCCTGAAGTTCATCTGCACCACCGCAAGCTGCCCCGTGCCCTGGCCACCTCG TGACCACCTTCACCTACGGCGTGCACTGCTTCGCCCGCTACCCCGACCACATGA AGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCA CCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCG AGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAG GACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACAACAGCCACAAGGT	1

[0651]

描述	序列	SEQ ID NO
	CTATATACCGCCGACAAGCAGAAGAACGGCATCAAGGTGAAC TTCAAGACCCG CCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACA CCCCCATCGGCGACGGCCCCGTGCTGCTGCCCCGACAACCACTACCTGAGCACCC AGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCTGCTG GAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGGA AGCGGAGCTACTAACTTCAGCCTGCTGAAGCAGGCTGGAGACGTGGAGGAGAA CCCTGGACCTATGACCGAGTACAAGCCACAGTGCGGCTGGCCACCAGGGACG ATGTGCCTAGAGCTGTGCGGACACTGGCCGCTGCCTTCGCCGATTACCCTGCCAC CAGACACACCGTGGACCCCGACAGACACATCGAGAGAGTGACCGAGCTGCAGG AACTGTTTCTGACCAGAGTGGGCCTGGACATCGGCAAAAGTGTGGGTGGCCGATG ATGGCGCCGCTGTGGCTGTGTGGACAACCCCTGAGTCTGTGGAAGCCGGCGCTG TGTTCGCCGAGATCGGACCTAGAATGGCCGAGCTGAGCGGCTCTAGACTGGCTG CCCAGCAGCAGATGGAAGGCCTGCTGGCCCCCACAGACCTAAAGAGCCTGCC TGGTTTCTGGCCACCGTGGGCGTGTACCTGACCACCAGGGCAAGGGACTGGG ATCTGCTGTGGTGCTGCCTGGCGTGGAAGCTGCTGAAAGGGCTGGCGTGCCCGC CTTCCTGGAACAAGCGCCCCAGAAACCTGCCCTTCTACGAGAGACTGGGCTT CACCGTGACCGCCGACGTGGAAGTGCCTGAGGGCCCTAGAACCTGGTGCATGA CCAGAAAGCCTGGCGCCCTTGGGCAGACGAGACCACACTGAGCCTCCCC	
EGFR	GGTTCGCGCCACAAAGCAGTGAATTTATTGGAGCATGGGTGAGCAAGGGCGAGG AGCTGTTACCGGGGTGGTGCCATCCTGGTCGAGCTGGACGGCGACGTAAACG GCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAG CTGACCTGAAGTTTCTATCTGCACACCGGCAAGCTGCCCCGTGCCCTGGCCACC CTCGTGACCACCTTACCTACGGCGTGCAGTGCTTCGCCCCGTACCCCGACCAC ATGAAGCAGCAGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAG CGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAA GTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCA AGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACCTACAACAGCCAC AAGGTCTATATACCGCCGACAAGCAGAAGAACGGCATCAAGGTGAAC TTCAAG ACCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCA GAACACCCCATCGGCGACGGCCCCGTGCTGCTGCCCCGACAACCACTACCTGAG CACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCC TGCTGGAGTTCGTGACCGCCGCGGGATCACTCTCGGCATGGACGAGCTGTACA AGGGAAGCGGAGCTACTAACTTCAGCCTGCTGAAGCAGGCTGGAGACGTGGAG GAGAACCCTGGACCTATGACCGAGTACAAGCCACAGTGCGGCTGGCCACCAG GGACGATGTGCCTAGAGCTGTGCGGACACTGGCCGCTGCCTTCGCCGATTACCC TGCCACCAGACACACCGTGGACCCCGACAGACACATCGAGAGAGTGACCGAGC TGCAGGAACTGTTTCTGACCAGAGTGGGCCTGGACATCGGCAAAAGTGTGGGTG GCCGATGATGGCGCCGCTGTGGCTGTGTGGACAACCCCTGAGTCTGTGGAAGCC GGCCTGTGTTTCGCCGAGATCGGACCTAGAATGGCCGAGCTGAGCGGCTCTAGA CTGGCTGCCAGCAGCAGATGGAAGGCCTGCTGGCCCCCACAGACCTAAAGA GCCTGCCTGGTTTCTGGCCACCGTGGGCGTGTACCTGACCACCAGGGCAAGGG ACTGGGATCTGCTGTGGTGCTGCCTGGCGTGGAAGCTGCTGAAAGGGCTGGCGT GCCCGCCTTCCTGGAACAAGCGCCCCAGAAACCTGCCCTTCTACGAGAGACT GGGCTTACCGTGACCGCCGACGTGGAAGTGCCTGAGGGCCCTAGAACCTGGT GCATGACCAGAAAGCCTGGCGCCTACCACGGAGGATAGTATGAGCCCTAAAAAT CCAG	2
β肌动蛋白	CACAGCGCGCCCGGCTATTCTCGCAGCTCACCATGACCGAGTACAAGCCACAG TGCGGCTGGCCACCAGGGACGATGTGCCTAGAGCTGTGCGGACACTGGCCGCTG	3

[0652]

描述	序列	SEQ ID NO
	CCTTCGCCGATTACCTGCCACCAGACACACCGTGGACCCCGACAGACACATCG AGAGAGTGACCGAGCTGCAGGAACTGTTTCTGACCAGAGTGGGCCTGGACATC GGCAAAGTGTGGGTGGCCGATGATGGCGCCGCTGTGGCTGTGTGGACAACCCCT GAGTCTGTGGAAGCCGGCGCTGTGTTCCGCCGAGATCGGACCTAGAATGGCCGAG CTGAGCGGCTCTAGACTGGCTGCCAGCAGCAGATGGAAGGCCTGCTGGCCCCC CACAGACCTAAAGAGCCTGCCTGGTTTCTGGCCACCGTGGGCGTGTACCTGAC CACCAGGGCAAGGGACTGGGATCTGCTGTGGTGCTGCCTGGCGTGGAAGCTGC TGAAAGGGCTGGCGTGCCCGCCTTCCTGGAAACAAGCGCCCCCAGAAACCTGC CCTTCTACGAGAGACTGGGCTTACCGTGACCGCCGACGTGGAAGTGCCTGAGG GCCCTAGAACCTGGTGCATGACCAGAAAGCCTGGCGCCGGAAGCGGAGCTACT AACTTCAGCCTGCTGAAGCAGGCTGGAGACGTGGAGGAGAACCCTGGACCTAA CCTGAGCAAAAACGTGAGCGTGAGCGTGTATATGAAGGGGAACGTCAACAATC ATGAGTTTGAGTACGACGGGGAAGGTGGTGGTGATCCTTATACAGGTAAATATTC CATGAAGATGACGCTACGTGGTCAAAATTCCTACCCTTTCTATGATATCATT CCACGGCATTTAGTATGGTTTCCGCGTATTTACAAAATACCTGAGGGAATTGTT GACTATTTTAAGGATTTCGCTTCCCGACGCATTCCAGTGGAACAGACGAATTGTGT TTGAAGATGGTGGAGTACTAAACATGAGCAGTGATATCACATATAAAGATAATGT TCTGCATGGTGACGTAAGGCTGAGGGAGTGAACCTCCCGCCGAATGGGCCAGT GATGAAGAATGAAATTGTGATGGAGGAACCGACTGAAGAAACATTTACTCCAAA AAACGGGGTTCTGTGGCTTTTGTCCCAAAGCGTACTTACTTAAAGATGGTTCC TATTACTATGGAATATGACAACATTTACAGATCCAAGAAATCTGGCCAGGCAC CTCCTGGGTATCATTGTGTAAGCATCGTCTCGTCAAGACCAATGTGGGACATGG ATTTAAGACGGTTGAGCAGACTGAATATGCCACTGCTCATGTCAGTGATCTTCCC AAATTCGAAGCTGATGATGATATCGCCGCGCTCGTCGTCGACAACGG	
LRRK2	GAGGGCGGCGGGTTGGAAGCAGGTGCCACC <u>ATG</u> ACCGAGTACAAGCCCACAGT GCGGCTGGCCACCAGGGACGATGTGCCTAGAGCTGTGCGGACACTGGCCGCTG CCTTCGCCGATTACCTGCCACCAGACACACCGTGGACCCCGACAGACACATCG AGAGAGTGACCGAGCTGCAGGAACTGTTTCTGACCAGAGTGGGCCTGGACATC GGCAAAGTGTGGGTGGCCGATGATGGCGCCGCTGTGGCTGTGTGGACAACCCCT GAGTCTGTGGAAGCCGGCGCTGTGTTCCGCCGAGATCGGACCTAGAATGGCCGAG CTGAGCGGCTCTAGACTGGCTGCCAGCAGCAGATGGAAGGCCTGCTGGCCCCC CACAGACCTAAAGAGCCTGCCTGGTTTCTGGCCACCGTGGGCGTGTACCTGAC CACCAGGGCAAGGGACTGGGATCTGCTGTGGTGCTGCCTGGCGTGGAAGCTGC TGAAAGGGCTGGCGTGCCCGCCTTCCTGGAAACAAGCGCCCCCAGAAACCTGC CCTTCTACGAGAGACTGGGCTTACCGTGACCGCCGACGTGGAAGTGCCTGAGG GCCCTAGAACCTGGTGCATGACCAGAAAGCCTGGCGCCGGAAGCGGAGCTACT AACTTCAGCCTGCTGAAGCAGGCTGGAGACGTGGAGGAGAACCCTGGACCTGT GAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCATCCTGGTCGAGCTGG ACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGAT GCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCC GTGCCCTGGCCACCCCTCGTGACCACCTTACCTACGGCGTGCACTGCTTCGCC CGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAA GGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACC CGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAA GGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACA ACTACAACAGCCACAAGGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCA AGGTGAACCTCAAGACCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCC GACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGAC	4

描述	序列	SEQ ID NO
	AACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGCTAGTGGCAGCTGTCAGGGGTGCGAAGAGGAC	
人 IgG 表达盒	CGACCTCTTTTGTGCCCTGATATAGTTTCGCCATGACCGAGTACAAGCCCACAGTGCGGCTGGCCACCAGGGACGATGTGCCTAGAGCTGTGCGGACACTGGCCGCTGCCCTCGCCGATTACCTTGCCACCAGACACACCGTGGACCCGACAGACACATCGAGAGAGTGACCGAGCTGCAGGAACTGTTTCTGACCAGAGTGGGCCTGGACATCGGCAAAGTGTGGGTGGCCGATGATGGCGCCGCTGTGGCTGTGTGGACAACCCCTGAGTCTGTGGAAGCCGGCGCTGTGTTCCGCCGAGATCGGACCTAGAATGGCCGAGCTGAGCGGCTCTAGACTGGCTGCCCAGCAGCAGATGGAAGGCCTGCTGGCCCCCACAGACCTAAAGAGCCTGCCTGGTTTCTGGCCACCGTGGGCGTGTACCTGACCACAGGGCAAGGGACTGGGATCTGCTGTGGTGTGCTGCTGGCGTGGAAGCTGCTGAAAGGGCTGGCGTGCCCGCCTTCCTGGAACAAGCGCCCCAGAAAACCTGCCCTTCTACGAGAGACTGGGCTTACCGTGACCGCCGACGTGGAAGTGCCTGAGGGCCCTAGAACCTGGTGCATGACCAGAAAAGCCTGGCGCCTGAGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCTATTAGTTCATAGCCCATATATGGAGTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCGCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTACGGTAAACTGCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGA CTCACGGGGATTCCAAGTCTCCACCCTATTGACGTCAATGGGAGTTTGT TTTGGCACAAAATCAACGGGACTTTCCAA AATGTCGTAACAACCTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCTGGAGACGCCATCCACGCTGTTTGTGACCTCCATAGAAGACACCGGGACCGATCCAGCCTCCGGACTCTAGAGGATCGAACCCTTGCCACCATGGGTTGGAGCCTCATCTTGCTCTTCCTTGTCGCTGTTGCTACGCGTGTCCTGTCCCAGGTACA ACTGCAGCAGCCTGGGGCTGAGCTGGTGAAGCCTGGGGCCTCAGTGAAGATGTCCTGCAAGGCTTCTGGCTACACATTTACCAGTTACAATATGCACTGGGTAAAACAGACACCTGGTCGGGGCCTGGAATGGATTGGAGCTATTTATCCCGGAAATGGTGATACTTCCTACAATCAGAAGTTCAAAGGCAAGGCCACATTGACTGCAGACAAATCCTCCAGCACAGCCTACATGCAGCTCAGCAGCCTGACATCTGAGGACTCTGCGGTCTATTACTGTGCAAGATCGACTTACTACGGCGGTGACTGGTACTTCAATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCTGCAGCTAGCACCAAGGGCCCATCGGTCTTCCCCCTGGCACCCCTCTCCAAGAGCACCTCTGGGGGCACCGCGGCCCTGGGCTGCCTGGTCAAGGACTACTTCCCCGAACCGGTGACGGTGTCTGTGGA ACTCAGGCGCCCTGACCAGCGGGTGTCACACCTTCCCGGCTGTCCTACAGTCCTCAGGACTCTACTCCCTCAGCAGCGTGGTGACCGTGCCCTCCAGCAGCTTGGGCACCCAGACCTACATCTGCAACGTGAATCACAAGCCCAGCAACACCAAGGTGGACAAGAAAGCAGAGCCCAATCTTGTGACAAA ACTCACACATGCCACCGTGCCAGCACCTGAACTCCTGGGGGGACCGTCAGTCTTCTCTTCCCCCAAAAACCAAGGACACCCCTCATGATCTCCCGGACCCCTGAGGTCAATGCGTGGTGGTGGACGTGAGCCACGAAGACCCTGAGGTCAAGTTCAACTGGTACGTGGACGGCGTGGAGGTGCATAATGCCAAGACAAAGCCA CGGGAGGAGCAGTACAACAGCACGTACCGTGTGGTCAGCGTCTCACCCTCTGC CACCAGGACTGGCTGAATGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAAGCCCTCCAGCCCCATCGAGAAAACCATCTCCAAGGCCAAAGGGCAGCCCCGAG	5

[0653]

[0654]

描述	序列	SEQ ID NO
	AACCACAGGTGTACACCCTGCCCCATCCCGGATGAGCTGACCAAGAACCAGG TCAGCCTGACCTGCCTGGTCAAAGGCTTCTATCCCAGCGACATCGCCGTGGAGT GGGAGAGCAATGGGCAGCCGAGAACAACTACAAGACCACGCCTCCCGTGCTG GACTCCGACGGCTCCTTCTTCTCTACAGCAAGCTCACCCTGGACAAGAGCAGG TGGCAGCAGGGGAACGTCTTCTCATGCTCCCGTGATGCATGAGGCTCTGCACAAC CACTACACGCAGAAGAGCCTCTCCCTGTCTCCGGGTAAACGTAAACGAAGAGGC AGCGGGGCTACTAACTTCAGCCTGCTGAAGCAGGCTGGAGACGTGGAGGAGAA CCCTGGACCTATGGATTTTCAGGTGCAGATTATCAGCTTCCTGCTAATCAGTGCTT CAGTCATAATGTCCAGAGGACAAATTGTTCTCTCCAGTCTCCAGCAATCCTGTC TGCATCTCCAGGGGAGAAGGTCACAATGACTTGACAGGGCCAGCTCAAGTGTAAAG TTACATCCACTGGTTCAGCAGAAGCCAGGATCCTCCCCAAACCCTGGATTTAT GCCACATCCAACCTGGCTTCTGGAGTCCCTGTTTCGCTTCAGTGGCAGTGGGTCT GGGACTTCTTACTCTCTCACAATCAGCAGAGTGGAGGCTGAAGATGCTGCCACT TATTACTGCCAGCAGTGGACTAGTAACCCACCCACGTTCCGAGGGGGGACCAAG CTGGAAATCAAACGTACGGTGGCTGCACCATCTGTCTTCATCTTCCCGCCATCTG ATGAGCAGTTGAAATCTGGAACCTGCCTCTGTTGTGTGCCTGCTGAATAACTCTA TCCCAGAGAGGCCAAAGTACAGTGGAAGGTGGATAACGCCCTCCAATCGGGTAA CTCCCAGGAGAGTGTACAGAGCAGGACAGCAAGGACAGCACCTACAGCCTCA GCAGCACCTGACGCTGAGCAAAGCAGACTACGAGAAACACAAAGTCTACGCC TGCGAAGTCACCCATCAGGGCCTGAGCTCGCCCGTCACAAAGAGCTTCAACAG GGGAGAGTGTGAAAGGGTTCGATCCCTACCGGTTAGTAATGAGTTTGATATCTC GACAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACT ATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTAAATGCCTTTGTATCATGCTA TTGCTTCCCGTATGGCTTTCATTTCTCCTCCTTGATATAAATCCTGGTTGCTGTCTC TTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGCACTGTGT TTGCTGACGCAACCCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTC CGGGACTTTCGCTTTCCTCTCCCTATTGCCACGGCGGAACATCGCCGCCTGC CTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATCCGTGGTG TTGTCGGGGAAGCTGACGCTCTTCCATGGCTGCTCGCCTGTGTTGCCACCTGG ATTCTGCGCGGGACGTCCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACC TTCCTTCCCGCGGCCTGCTGCCGGCTCTGCGGCCTCTCCGCGTCTTCGCTTCG CCCTCAGACGAGTCGGATCTCCCTTTGGGCCGCTCCCCGCTGGGATGACGATA TCGCTGCGCTCGTTGTCGACAACGG	
格氏嗜盐碱杆菌阿格蛋白氨基酸序列	MTVIDLDSTT TADELTSGHT YDISVTLTG YDNTDEQHPR MSLAFEQDNG ERRYITLWKN TTPKDVFTYD YATGSTYIFT NIDYEVKDG YENLTATYQTT VENATAQEVG TTDEDEFAG GEPLDHHLLD ALNETPDDAE TESDSGHVMT SFASRDQLPE WTLHTYTLTA TDGAKTDTEY ARRTLAYTVR QELYTDHDA PVATDGLMLL TPEPLGETPL DLDCGVRVEA DETRTLDTYTT AKDRLLAREL VEEGLKRSLW DDYLVRGIDE VLSKEPVLTC DEFDLHERYD LSVEVGHSGR AYLHINFRHR FVPKLTADI DDDNIYPGLR VKTTYRPRRG HIVWGLRDEC ATDSLNTLGN QSVVAYHRNN QTPINTDLLD AIEAADRRV ETRRQGHGDD AVSFPQELLA VEPNTHQIKQ FASDGFHQQA RSKTRLSASR CSEKAQAF RLDPVRLNGS TVEFSSEFFT GNNEQQLRL YENGESVLTF RDGARGAHPD ETFSKGIVNP PESFEVAVVL PEQQADTCKA QWDTMADLLN QAGAPPTRSE TVQYDAFSSP ESISLNVAGA IDPSEVDAAF VVLPPDQEGF ADLASPTETY DELKKALANM GIYSQMAYFD RFRDAKIFYT RNVALGLLAA AGGVAFTTEH AMPGDADMFI GIDVRSYPE DGASGQINIA ATATAVYKDG TILGHSSTRP QLGEKLQSTD VRDIMKNAIL GYQQVTGES THIVIHRRDGF MNEDLDPATE	6

[0655]

描述	序列	SEQ ID NO
	FLNEQGVEYD IVEIRKQPQT RLLAVSDVQY DTPVKSI AAI NQNEPRATVA TFGAPEYLAT RDGGGLPRPI QIERVAGETD IETLTRQVYL LSQSHIQVHN STARLPITTA YADQASTHAT KGYLVQTGAF ESNVGFL	
核定位信号 (NLS)	PKKKRKV	7
NLS	AVKRPAATKKAGQAKKKKLD	8
NLS	MSRRRKANPTKLSENAKKLAKEVEN	9
NLS	PAAKRVKLD	10
叶绿体靶向信号	LIAHPQAFPGAIAAPISYAYAVKGRKPRFQTAKGSVRI	11
线粒体靶向信号	MLSLRQSIRFFKPATRTLCSRYLL	12
BP-SV40, NLES 肽	KRTADGSEFESPKKKRKVEGG	13
寡核苷酸实例	CGGGGTAGCGGCTGAAGCACTGCACGCCGTAGGTCAGGGTGG TCACGAGGGTGGGCCAGGGCACGGGCAGCTTGCCGGTGGTGC AGATGAACTTCAG	14
断裂的 EmGFP	GCACGCCGTAGGTGGTCACGAGG	15
寡核苷酸实例	GCTGCCCCGTGCCCTGGCCCCACCCTCGTGACCACCCTGACCTAC GGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGA	16
eBFP	CTCGTGACCACCCTGACCCACGG	17
CMPK1-TALEN2_7nt_TAL-Buddy_Lt	CTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTA ATACGACTCACTATAGGGAGTCCCAAGCTGGCTAGCGTTTAAA CTTCTGCGGCCGCGCCACCATGGGAAAACCTATTCTTAATCCT CTGCTGGGCCTGGATTCTACCGGAGGCGTGGACCTGAGAACA CTGGGATATTCTCAGCAGCAGCAGGAGAAGATCAAGCCCAAG GTGAGATCTACAGTGGCCCAGCACCACGAAGCCCTGGTGGGA CACGGATTACACACGCCACATTGTGGCCCTGTCTCAGCACC CTGCCGCCCTGGGAACAGTGGCCGTGAAATATCAGGATATGAT TGCCGCCCTGCCTGAGGCCACACACGAAGCCATTGTGGGAGT GGGAAAACGAGGCGCTGGAGCCAGAGCCCTGGAAGCCCTGCT GACAGTGGCCGAGAACTGAGAGGACCTCCTCTGCAGCTGGA TACAGGACAGCTGCTGAAGATTGCCAAAAGGGGCGGAGTGAC CGCGGTGGAAGCCGTGCACGCCTGGAGAAATGCCCTGACAGG AGCCCCTCTGAACCTGACCCCCGAACAGGTGGTGGCCATTGCC AGCCACGACGGCGGCAAGCAGGCCCTGGAAACCGTGCAGAG ACTGCTGCCCCTGCTGTGCCAGGCCCATGGCCTGACACCTGAA CAGGTGGTGGCTATCGCCTCTAATATCGGAGGAAAACAGGCTC TGGAACAGTGCAGCGGCTGCTGCCTGTGCTGTGTCAGGCTC ACGGCTTGACTCCAGAACAGGTGGTGGCTATTGCTTCCAATAT TGGGGGAAACAGGCCCTGGAACTGTGCAGCGCCTGCTGCC AGTGCTGTGCCAGGCTCACGGA CTGACCCCCGAACAGGTGGT GGCCATTGCCAGCAACATCGGCGGCAAGCAGGCCCTGGAAAC	18

[0656]

描述	序列	SEQ ID NO
	CGTGCAGAGACTGCTGCCCCGTGCTGTGCCAGGCCCATGGCCTG ACACCTGAACAGGTGGTGGCTATCGCCTCTAATATCGGAGGAA AACAGGCTCTGGAAACAGTGCAGCGGCTGCTGCCTGTGCTGT GTCAGGCTCACGGCTTGACTCCAGAACAGGTGGTGGCTATTGC TTCCAATATTGGGGGGAAACAGGCCCTGGAACTGTGCAGCG CCTGCTGCCAGTGTGTGCCAGGCTCACGGGCTGACCCCCGA ACAGGTGGTGGCCATTGCCAGCCACGACGGCGGCAAGCAGGC CCTGGAAACCGTGCAGAGACTGCTGCCCCGTGCTGTGCCAGGC CCATGGCCTGACACCTGAACAGGTGGTGGCTATCGCCTCTCAC GACGGAGGAAAACAGGCTCTGGAAACAGTGCAGCGGCTGCT GCCTGTGCTGTGTGTCAGGCTCACGGCTTGACTCCAGAACAGGT GGTGGCTATTGCTTCCAACGGCGGGGGGAAACAGGCCCTGGA AACTGTGCAGCGCCTGCTGCCAGTGTGTGCCAGGCTCACGG CCTCACTCCCGAACAGGTGGTGGCCATTGCCAGCAACAACGG CGGCAAGCAGGCCCTGGAAACCGTGCAGAGACTGCTGCCCCGT GCTGTGCCAGGCCCATGGCCTGACACCTGAACAGGTGGTGGC TATCGCCTCTAACGGCGGAGGAAAACAGGCTCTGGAAACAGT GCAGCGGCTGCTGCCTGTGCTGTGTGTCAGGCTCACGGCTTGACT CCAGAACAGGTGGTGGCTATTGCTTCCAATATTGGGGGGAAAC AGGCCCTGGAACTGTGCAGCGCCTGCTGCCAGTGTGTGCC AGGCTCACGGACTGACCCCCGAACAGGTGGTGGCCATTGCCA GCAACATCGGCGGCAAGCAGGCCCTGGAAACCGTGCAGAGAC TGCTGCCCCGTGCTGTGCCAGGCCCATGGCCTGACACCTGAACA GGTGGTGGCTATCGCCTCTAATATCGGAGGAAAACAAGCACTC GAGACAGTGCAGCGGCTGCTGCCTGTGCTGTGTGTCAGGCTCAC GGCTTGACTCCAGAACAGGTGGTGGCTATTGCTTCCAACAACG GGGGGAAACAGGCCCTGGAACTGTGCAGCGCCTGCTGCCAG TGCTGTGCCAGGCTCACGGCCTGACCCCCGAACAGGTGGTGG CCATTGCCAGCAACAACGGCGGCAAGCAGGCCCTGGAAACCG TGCAGAGACTGCTGCCCCGTGCTGTGCCAGGCCCATGGCCTGAC ACCTGAACAGGTGGTGGCTATCGCCTCTAATATCGGAGGAAAA CAGGCTCTGGAAACAGTGCAGCGGCTGCTGCCTGTGCTGTGT CAGGCTCACGGCTTGACTCCACAGCAGGTGCTGGCAATTGCTA GCAATATCGGCGGACGGCCCGCCCTGGAGAGCATTGTGGCCCCA GCTGTCTAGACCTGATCCTGCCCTGGCCGCCCTGACAAATGAT CACCTGGTGGCCCTGGCCTGTCTGGGAGGCAGACCTGCCCTG GATGCCGTGAAAAAAGGACTGCCTCACGCCCCTGCCCTGATCA AGAGAACAAATAGAAGAATCCCCGAGCGGACCTCTCACAGAG TGGCCGGATCCCCTAAGAAAAAGCGGAAGGTGGGATCCT TGAA AGCTTCTCGAGTCTAGAGGGCCCGTTTAAACCCGCTGATCAGC CTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCCT CCCCCGTGCC	
CMPK1-C 目标	AACTCAAGTGATCTGCCCCGCTCGACCTCCCAAAGTGCTGGGA TTACAGATGTGAGCCACCGCGCCCCGCCAAATTTGATTATTTTT AATAAGAACTTAGCTGTATGGTATTTTAAACAGTACCTGCTTTTA AAATTATTATCATCTTT TTCCCTTTACAGGTTTTTGATGAAGTGT GCAGATTTTTGACAAGGAAGGCTAATTCTAAACCTGAAGGCAT CCTTGAAATCATGCTTGAAATATTGCTTTGATAGCTGCTATCATGA	19

[0657]

描述	序列	SEQ ID NO
	CCCCTTTTAAAGGCAATTCTAATCTTTCATAACTACATCTCAATT AGTGGCTGGAAAGTACATGGTAAAACAAAGTAAATTTTTTAT GTTCTTTTTTTTGGTCACAGGAGTAGACAGTGAATTCAGGTTTA ACTTCACCTTAGTTATGGTGCTCACCAAACGAAGGGTATCAGC TATTTTTTTTAAATTCAAAAAGAATATCCCTTTTATAGTTTGTG CCTTCTGTGAGCAAACTTTTATAGTACGCGTATATATCCCTCTA GTAATCACAACATTTTAGGATT	
TD1-F2, 5'引物, 用于 TALEN/TAL-Buddy 全长富集	CTGGCTAACTAGAGAACCCACTGCTTACTG	22
TD8-R2, 3'引物, 用于 TALEN/TAL-Buddy 全长富集	GGCACGGGGGAGGGGCAAACAACAGATGGC	23
CMPK1-TALEN2_F, 位于 CMPK1-C 目标处的正向 TALEN	TGTGCAGATTTTGTGACAA	24
CMPK1-TALEN2_R, 位于 CMPK1-C 目标处的反向 TALEN	TCAAGGATGCCTTCAGGT	25
CMPK1-TALEN2_7nt_TAL-Buddy_Lt, 间隔 7 nt 的左 TAL-Buddy	CAAAAACCTGTAAAGGAA	26
CMPK1-TALEN2_7nt_TAL-Buddy_Rt, 间隔 7nt 的右 TAL-Buddy	CTTGAATATTGCTTTGAT	27
CMPK1-TALEN2_0nt_TAL-Buddy_Lt, 间隔 0 nt 的左 TAL-Buddy	ACTTCATCAAAAACCTGT	28
CMPK1-TALEN2_0nt_TAL-Buddy_Rt, 间隔 0 nt 的右 TAL-Buddy	AATCATGCTTGAATATTG	29
CMPK1-TALEN2_4nt_TAL-Buddy_Lt, 间隔 4 nt 的左 TAL-Buddy	CATCAAAAACCTGTAAAG	30
CMPK1-TALEN2_4nt_TAL-Buddy_Rt, 间隔 4 nt 的右 TAL-Buddy	ATGCTTGAATATTGCTTT	31
CMPK1-TALEN2_12nt_TAL-Buddy_Lt, 间隔 12 nt 的左 TAL-Buddy	ACCTGTAAAGGAAAAAGA	32
CMPK1-TALEN2_12nt_TAL-Buddy_Rt, 间隔 12 nt 的右 TAL-Buddy	ATATTGCTTTGATAGCTG	33
CMPK1-TALEN2_20nt_TAL-Buddy_Lt, 间隔 20 nt 的左 TAL-Buddy	AGGAAAAAGATGATAATA	34
CMPK1-TALEN2_20nt_TAL-Buddy_Rt, 间隔 20 nt 的右 TAL-Buddy	TTGATAGCTGCTATCATG	35
CMPK1-TALEN2_30nt_TAL-Buddy_Lt, 间隔 30 nt 的左 TAL-Buddy	TGATAATAATTTTAAAAG	36
CMPK1-TALEN2_30nt_TAL-Buddy_Rt, 间隔 30 nt 的右 TAL-Buddy	CTATCATGACCCCTTTT	37
CMPK1-TALEN2_50nt_TAL-Buddy_Lt, 间隔 50 nt 的左 TAL-Buddy	GGTACTGTAAATACCA	38
CMPK1-TALEN2_50nt_TAL-Buddy_Rt, 间隔 50 nt 的右 TAL-Buddy	GGCAATTCTAATCTTTCA	39
CMPK1-TALEN2_100nt_TAL-Buddy_Lt, 间隔 100 nt 的左 Left TAL-Buddy	TTGGCGGGGCGCGGTGGC	40

描述	序列	SEQ ID NO
[0658]	CMPK1-TALEN2_100nt_TAL-Buddy_Rt, 间隔 100 nt 的右 TAL-Buddy ATGGTAAACAAAGTAAA	41

[0659] 本文提供的名称、标题和子标题不应解释为限制本公开的各个方面。因此,以下所定义的术语通过整体参考说明书而被更充分地定义。

[0660] 除非另外定义,否则本文所用的技术和科学术语具有与本领域普通技术人员通常所理解相同的含义。参见例如Singleton等人,《微生物学和分子生物学词典 (DICTIONARY OF MICROBIOLOGY AND MOLECULAR BIOLOGY)》第二版,J.Wiley&Sons (纽约州纽约,1994); Sambrook等人,《分子克隆:实验室手册 (MOLECULAR CLONING, A LABORATORY MANUAL)》,冷泉港出版社 (纽约州冷泉港,1989)。可以使用与本文所述相似或等效的任何方法、装置以及材料。本文提供定义以促进理解本文中经常使用的某些术语,并且不意味着限制本发明的范围。

[0661] 在本申请中,除非另外说明,否则“或”的使用意指“和/或”。在多个从属权利要求的上下文中,“或”的使用仅替代地指超过一个在先独立或从属权利要求。更应注意,如本说明书和所附权利要求书中所使用,单数形式“一个 (a)”、“一个 (an)”和“所述 (the)”和任何词语的任何单数使用,包括复数形式,除非明确和确切地限于一个指示物。如本文所用,术语“包括”和其语法变型旨在是非限制性的,使得清单中所列项不排除可以取代或添加到所列项中的其它相似项。

[0662] 如本文所述,任何浓度范围、百分比范围、比率范围或整数范围应理解为包括所述范围内的任何整数值,以及在适当时包括其分数 (例如整数的十分之一和百分之一),除非另有说明。

[0663] 单位、前缀和符号是以国际单位制 (SI) 接受的形式表示。数值范围包括限定范围的数值。考虑到有效数字和与测量相关的误差,测量值应理解为近似值。

[0664] 前述书面说明书被认为足以使本领域技术人员能够实施实施例。前述描述和实例详述了某些实施例并描述了发明人设想的最佳模式。然而,应当理解,无论前述内容在文本中的详细程度如何,实施例可以多种方式实施,并且应所述根据所附权利要求书和其任何等效物来解释。本说明书和示例性实施例不应视为是限制性的。

[0665] 出于本说明书和所附权利要求书的目的,除非另外指出,否则表示数量、百分比或比例的所有数值和本说明书和权利要求书中使用的其它数值在所有情况下应理解为被术语“约”修饰,直至它们还未被如此修饰的程度。因此,除非相反指出,否则说明书和所附权利要求书中阐释的数值参数是近似值,其可根据寻求获得的期望特性而改变。最低限度地,并且不试图将等同原则的应用限于权利要求书的范围,至少应根据所报告的有效数字的位数并且通过应用一般舍入技术来解释每个数值参数。

[0666] 当术语如“小于或等于”或“大于或等于”位于数值或范围的清单之前时,术语修饰清单中所提供的所有值或范围。在一些实施例中,数值四舍五入到最接近的整数或有效数字。

[0667] 实施例

[0668] 实施例P-1. 一种用于在初始核酸分子中进行同源重组的方法,包含:(a) 在所述

初始核酸分子中产生双链断裂,以产生裂解的核酸分子;以及(b)使所述裂解的核酸分子与供体核酸分子接触,其中所述初始核酸分子包含启动子和基因,其中所述供体核酸分子包含:(i)长度为12bp至250bp的5'和3'末端上的匹配端、(ii)无启动子选择标记物、(iii)报道基因、(iv)使所述无启动子选择标记物与位于所述无启动子选择标记物任一侧的所述报道基因或 LoxP连接的自裂解肽,以及(iv)任选地存在于所述无启动子选择标记物与所述报道基因之间的接头。

[0669] 实施例P-2.根据实施例P-1的方法,其中所述核酸分子中的双链断裂:(i)相对于ATG起始密码子,小于或等于250bp,用于对所述裂解的核酸分子进行N端标记;或(ii)相对于终止密码子,小于或等于250bp,用于对所述裂解的核酸分子进行C端标记。

[0670] 实施例P-3.根据实施例P-1的方法,其中所述双链断裂是通过至少一种核酸切割实体或电穿孔来诱导。

[0671] 实施例P-4.根据实施例P-3的方法,其中所述至少一种核酸切割实体包含核酸酶,包含一种或多种锌指蛋白、一种或多种转录激活因子样效应物 (TALE)、一种或多种CRISPR复合物、一种或多种阿格蛋白-核酸复合物,或一种或多种大范围核酸酶。

[0672] 实施例P-5.根据实施例P-3的方法,其中所述至少一种核酸切割实体是使用表达载体、质粒、核糖核蛋白复合物(RNC)或mRNA施用。

[0673] 实施例P-6.根据实施例P-1的方法,其中所述无启动子选择标记物包含蛋白质、抗生素抗性选择标记物、细胞表面标记物、细胞表面蛋白、代谢物或其活性片段。

[0674] 实施例P-7.根据实施例P-6的方法,其中所述无启动子选择标记物是蛋白质。

[0675] 实施例P-8.根据实施例P-7的方法,其中所述蛋白质是粘着斑激酶 (FAK)、血管生成素相关生长因子 (AGF) 受体,或表皮生长因子受体 (EGFR)。

[0676] 实施例P-9.根据实施例P-6的方法,其中所述无启动子选择标记物是抗生素抗性选择标记物。

[0677] 实施例P-10.根据实施例P-9的方法,其中所述抗生素抗性选择标记物是重组抗体。

[0678] 实施例P-11.根据实施例P-9的方法,其中所述抗生素抗性选择标记物是人类IgG抗体。

[0679] 实施例P-12.根据实施例P-1的方法,其中所述报道基因包含荧光蛋白报道体。

[0680] 实施例P-13.根据实施例P-12的方法,其中所述荧光蛋白报道体是翡翠绿色荧光蛋白 (EmGFP) 报道体或橙色荧光蛋白 (OFp) 报道体。

[0681] 实施例P-14.根据实施例P-1的方法,其中所述无启动子选择标记物:(i) 连接至报道基因的5'末端,以便对所述裂解的核酸分子进行N端标记;或(ii) 连接至报道基因的3'末端,以便对所述裂解的核酸分子进行C端标记。

[0682] 实施例P-15.根据实施例P-1的方法,其中所述供体核酸分子包含介于所述无启动子选择标记物与所述报道基因之间的接头。

[0683] 实施例P-16.根据实施例P-15所述的方法,其中所述无启动子选择标记物与所述报道基因之间的距离小于或等于300nt、240nt、180nt、150nt、120 nt、90nt、60nt、30nt、15nt、12nt或9nt。

[0684] 实施例P-17.根据实施例P-16的方法,其中所述距离是6nt。

- [0685] 实施例P-18.根据实施例P-15的方法,其中所述接头是聚甘氨酸接头。
- [0686] 实施例P-19.根据实施例P-1的方法,其中所述自裂解肽是自裂解2A肽。
- [0687] 实施例P-20.根据实施例P-1的方法,其中通过PCR扩增将所述匹配端添加到所述供体核酸分子的5'和3'末端。
- [0688] 实施例P-21.根据实施例P-1的方法,其中所述匹配端共享大于或等于 95%的序列一致性。
- [0689] 实施例P-22.根据实施例P-1的方法,其中所述匹配端包含单链DNA或双链DNA。
- [0690] 实施例P-23.根据实施例P-1的方法,其中所述供体核酸分子的5'和3'末端的匹配端具有12bp至200bp、12bp至150bp、12bp至100bp、12bp至 50bp或12bp至40bp的长度。
- [0691] 实施例P-24.根据实施例P-23的方法,其中所述匹配端具有35bp的长度。
- [0692] 实施例P-25.根据实施例P-1的方法,其中所述初始核酸分子存在于细胞或质粒中。
- [0693] 实施例P-26.根据实施例P-1的方法,其中所述供体核酸分子包含小于或等于1kb、2kb、3kb、5kb、10kb、15kb、20kb、25kb或30kb的长度。
- [0694] 实施例P-27.根据实施例P-1的方法,其中所述供体核酸分子通过同源定向修复(HDR)整合于所述裂解的核酸分子中。
- [0695] 实施例P-28.根据实施例P-27的方法,其中所述HDR大于或等于10%、25%、50%、75%、90%、95%、98%、99%或100%。
- [0696] 实施例P-29.根据实施例P-1的方法,其中所述供体核酸分子的整合效率大于或等于50%、75%、90%、95%、98%、99%或100%。
- [0697] 实施例P-30.根据实施例P-1的方法,进一步包含在5'末端、3'末端或5'与3'末端修饰所述供体核酸分子。
- [0698] 实施例P-31.根据实施例P-30的方法,其中所述供体核酸分子在5'与3'末端经修饰。
- [0699] 实施例P-32.根据实施例P-30的方法,其中所述供体核酸分子经修饰而在至少一个末端的至少一条链中具有一个或多个核酸酶抗性基团。
- [0700] 实施例P-33.根据实施例P-32的方法,其中所述一个或多个核酸酶抗性基团包含一个或多个硫代磷酸酯基团、一个或多个氨基、2'-O-甲基核苷酸、2'-脱氧-2'-氟核苷酸、2'-脱氧核苷酸、5-C-甲基核苷酸或其组合。
- [0701] 实施例P-34.根据实施例P-1的方法,进一步包含用至少一种非同源末端连接(NHEJ)抑制剂处理所述供体核酸分子。
- [0702] 实施例P-35.根据实施例P-34的方法,其中所述至少一种NHEJ抑制剂是DNA依赖性蛋白激酶(DNA-PK)、DNA连接酶IV、DNA聚合酶1或2(PARP-1或PARP-2),或其组合。
- [0703] 实施例P-36.根据实施例P-35的方法,其中所述DNA-PK抑制剂是 Nu7026 (2-(4-吗啉基)-4H-萘并[1,2-b]吡喃-4-酮)、Nu7441 (8-(4-二苯并噻吩基)-2-(4-吗啉基)-4H-1-苯并吡喃-4-酮)、Ku-0060648 (4-乙基-N-[4-[2-(4-吗啉基)-4-氧代-4H-1-苯并吡喃-8-基]-1-二苯并噻吩基]-1-哌嗪乙酰胺)、化合物401 (2-(4-吗啉基)-4H-嘧啶并[2,1-a]异喹啉-4-酮)、DMNB (4,5-二甲氧基-2-硝基苯甲醛)、ETP 45658 (3-[1-甲基-4-(4-吗啉基)-1H-吡唑并[3,4-d]嘧啶-6-基苯酚)、LTURM 34 (8-(4-二苯并噻吩基)-2-(4-吗啉基)-4H-1,

3-苯并噁嗪-4-酮),或P1 103盐酸盐(3-[4-(4-吗啉基吡啶并[3',2':4,5]呋喃并[3,2-d]嘧啶-2-基]苯酚盐酸盐)。

[0704] 实施例P-37.根据实施例P-1的方法,其中所述哺乳动物是人类、哺乳动物实验动物、哺乳动物农畜、哺乳动物运动动物或哺乳动物宠物。

[0705] 实施例P-38.根据实施例P-37的方法,其中所述哺乳动物是人类。

[0706] 实施例P-39.一种利用实施例P-1的方法制备的细胞或质粒。

[0707] 实施例P-40.根据实施例P-39的细胞,其中所述细胞是真核细胞。

[0708] 实施例P-41.根据实施例P-40的细胞,其中所述真核细胞是哺乳动物细胞。

[0709] 实施例P-42.一种细胞治疗方法,包含向有需要的受试者施用有效量的根据实施例P-41的细胞。

[0710] 实施例P-43.根据实施例P-42的方法,其中所述细胞是T细胞且所述无启动子选择标记物是嵌合抗原受体(CAR)。

[0711] 实施例P-44.一种用于产生无启动子选择标记物的方法,包含激活通过实施例P-1的方法所制备的细胞或质粒的启动子以产生所述无启动子选择标记物。

[0712] 实施例P-45.一种组合物,其包含通过实施例P-44的方法所产生的无启动子选择标记物。

[0713] 实施例P-46.一种对有需要的受试者进行治疗性治疗的方法,包含施用有效量的通过实施例P-44的方法所产生的无启动子选择标记物。

[0714] 实施例P-47.一种药物筛选分析,包含通过实施例P-44的方法所产生的无启动子选择标记物。

[0715] 实施例P-48.一种用于产生无启动子选择标记物的试剂盒,其包含无启动子选择标记物,所述无启动子选择标记物通过自裂解肽或LoxP而在选择标记物的任一侧连接至报道基因。

[0716] 实施例P-49.根据实施例P-48的试剂盒,其中所述报道基因是GFP或 OFP。

[0717] 实施例P-50.根据实施例P-48的试剂盒,其进一步包含至少一种核酸切割实体。

[0718] 实施例P-51.根据实施例P-48的试剂盒,其进一步包含至少一种NHEJ 抑制剂。

[0719] 实施例P-52.根据实施例P-48的试剂盒,其进一步包含一个或多个核酸酶抗性基团。

[0720] 实施例P-53.一种重组抗体表达盒,其包含:(i)所述表达盒的5'和3'末端的匹配端,其中所述匹配端的长度小于或等于250bp;(ii)无启动子选择标记物;(iii)报道基因;(iv)使所述无启动子选择标记物与所述报道基因连接的自裂解肽;和(v)任选地介于所述无启动子选择标记物与所述报道基因之间的接头,其中所述无启动子选择标记物连接于所述报道基因的5'末端以便对裂解的核酸分子进行N端标记,或连接于所述报道基因的3'末端以便对裂解的核酸分子进行C端标记。

[0721] 实施例P-54.一种增强细胞中的目标基因座的可及性的方法,所述方法包含:(1)将第一DNA结合调节增强剂引入包含编码目标基因座的核酸的细胞中,其中所述第一DNA结合调节增强剂不是所述细胞内源的;和(2)使所述第一DNA结合调节增强剂结合所述目标基因座的第一增强子结合序列,相对于所述第一DNA结合调节增强剂的缺乏,借此增强所述目标基因座的可及性。

[0722] 实施例P-55.根据实施例P-54的方法,其中所述引入第一DNA结合调节增强剂包含引入编码所述第一DNA结合调节增强剂的载体。

[0723] 实施例P-56.根据实施例P-54的方法,其中所述引入第一DNA结合调节增强剂包含引入编码所述第一DNA结合调节增强剂的mRNA。

[0724] 实施例P-57.根据实施例P-54的方法,其中所述引入第一DNA结合调节增强剂包含引入第一DNA结合蛋白或第一DNA结合核酸。

[0725] 实施例P-58.根据实施例P-54的方法,其中相对于所述第一DNA结合调节增强剂的缺乏,在所述目标基因座发生同源重组的比率增加。

[0726] 实施例P-59.根据实施例P-54的方法,其中所述第一DNA结合调节增强剂是第一DNA结合蛋白或第一DNA结合核酸。

[0727] 实施例P-60.根据实施例P-54的方法,其中所述第一DNA结合调节增强剂是第一转录激活因子样(TAL)效应蛋白或截短的第一向导RNA(gRNA)。

[0728] 实施例P-61.根据实施例P-54的方法,其中所述第一增强子结合序列具有SEQ ID NO:26、SEQ ID NO:28、SEQ ID NO:30、SEQ ID NO:32、SEQ ID NO:34、SEQ ID NO:36、SEQ ID NO:38或SEQ ID NO:40的序列。

[0729] 实施例P-62.一种置换细胞中的目标基因座的染色质的方法,所述方法包含:(1)将第一DNA结合调节增强剂引入包含编码目标基因座的核酸的所述细胞中,其中所述第一DNA结合调节增强剂不是所述细胞内源的;和(2)使所述第一DNA结合调节增强剂结合所述目标基因座的第一增强子结合序列,借此置换所述目标基因座的染色质。

[0730] 实施例P-63.根据实施例P-62的方法,其中所述引入第一DNA结合调节增强剂包含引入编码所述第一DNA结合调节增强剂的载体。

[0731] 实施例P-64.根据实施例P-62的方法,其中所述引入第一DNA结合调节增强剂包含引入编码所述第一DNA结合调节增强剂的mRNA。

[0732] 实施例P-65.根据实施例P-62的方法,其中所述引入第一DNA结合调节增强剂包含引入第一DNA结合蛋白或第一DNA结合核酸。

[0733] 实施例P-66.根据实施例P-62的方法,其中相对于所述第一DNA结合调节增强剂的缺乏,在所述目标基因座处发生同源重组的比率增加。

[0734] 实施例P-67.根据实施例P-62的方法,其中所述第一DNA结合调节增强剂是第一DNA结合蛋白或第一DNA结合核酸。

[0735] 实施例P-68.根据实施例P-62的方法,其中所述第一DNA结合调节增强剂是第一转录激活因子样(TAL)效应蛋白或截短的第一向导RNA(gRNA)。

[0736] 实施例P-69.一种重构细胞中的目标基因座的染色质的方法,所述方法包含:(1)将第一DNA结合调节增强剂引入包含编码目标基因座的核酸的细胞,其中所述第一DNA结合调节增强剂不是所述细胞内源的;和(2)使所述第一DNA结合调节增强剂结合所述目标基因座的第一增强子结合序列,借此重构所述目标基因座的染色质。

[0737] 实施例P-70.根据实施例P-69的方法,其中所述引入第一DNA结合调节增强剂包含引入编码所述第一DNA结合调节增强剂的载体。

[0738] 实施例P-71.根据实施例P-69的方法,其中所述引入第一DNA结合调节增强剂包含引入编码所述第一DNA结合调节增强剂的mRNA。

[0739] 实施例P-72.根据实施例P-69的方法,其中所述引入第一DNA结合调节增强剂包含引入第一DNA结合蛋白或第一DNA结合核酸。

[0740] 实施例P-73.根据实施例P-69的方法,其中相对于所述第一DNA结合调节增强剂的缺乏,在所述目标基因座处发生同源重组的比率增加。

[0741] 实施例P-74.根据实施例P-69的方法,其中所述第一DNA结合调节增强剂是第一DNA结合蛋白或第一DNA结合核酸。

[0742] 实施例P-75.根据实施例P-69的方法,其中所述第一DNA结合调节增强剂是第一转录激活因子样(TAL)效应蛋白或截短的第一向导RNA(gRNA)。

[0743] 实施例P-76.一种增强细胞中的目标基因座的可达性的方法,所述方法包含:(1)向包含编码目标基因座的核酸的细胞中引入:(i)第一DNA结合调节增强剂,其中所述第一DNA结合调节增强剂不是所述细胞内源的;和(ii)第二DNA结合调节增强剂,其中所述第二DNA结合调节增强剂不是所述细胞内源的;(2)使所述第一DNA结合调节增强剂结合所述目标基因座的第一增强子结合序列;和(3)使所述第二DNA结合调节增强剂结合所述目标基因座的第二增强子结合序列,相对于所述第一DNA结合调节增强剂或所述第二DNA结合调节增强剂的缺乏,借此增强所述目标基因座的可达性。

[0744] 实施例P-77.根据实施例P-76的方法,其中所述引入第一DNA结合调节增强剂包含引入编码所述第一DNA结合调节增强剂的载体。

[0745] 实施例P-78.根据实施例P-76的方法,其中所述引入第一DNA结合调节增强剂包含引入编码所述第一DNA结合调节增强剂的mRNA。

[0746] 实施例P-79.根据实施例P-76的方法,其中所述引入第一DNA结合调节增强剂包含引入第一DNA结合蛋白或第一DNA结合核酸。

[0747] 实施例P-80.根据实施例P-76的方法,其中所述引入第二DNA结合调节增强剂包含引入编码所述第二DNA结合调节增强剂的载体。

[0748] 实施例P-81.根据实施例P-76的方法,其中所述引入第二DNA结合调节增强剂包含引入编码所述第二DNA结合调节增强剂的mRNA。

[0749] 实施例P-82.根据实施例P-76的方法,其中所述引入第二DNA结合调节增强剂包含引入第二DNA结合蛋白或第二DNA结合核酸。

[0750] 实施例P-83.根据实施例P-76的方法,其中相对于所述第一DNA结合调节增强剂的缺乏,在所述目标基因座处发生同源重组的比率增加。

[0751] 实施例P-84.根据实施例P-76的方法,其中所述第一DNA结合调节增强剂是第一DNA结合蛋白或第一DNA结合核酸。

[0752] 实施例P-85.根据实施例P-76的方法,其中所述第一DNA结合调节增强剂是第一转录激活因子样(TAL)效应蛋白或截短的第一向导RNA(gRNA)。

[0753] 实施例P-86.根据实施例P-76的方法,其中所述第二DNA结合调节增强剂是第二DNA结合蛋白或第二DNA结合核酸。

[0754] 实施例P-87.根据实施例P-76的方法,其中所述第二DNA结合调节增强剂是TAL效应蛋白或截短的gRNA。

[0755] 实施例P-88.根据实施例P-76的方法,其中所述第一DNA结合调节增强剂是第一TAL效应蛋白且所述第二DNA结合调节增强剂是第二TAL效应蛋白。

[0756] 实施例P-89.根据实施例P-76的方法,其中所述第一DNA结合调节增强剂是TAL效应蛋白且所述第二DNA结合调节增强剂是截短的gRNA。

[0757] 实施例P-90.根据实施例P-76的方法,其中所述第一DNA结合调节增强剂是截短的第一gRNA且所述第二DNA结合调节增强剂是截短的第二gRNA。

[0758] 实施例P-91.根据实施例P-76的方法,其中所述第一DNA结合调节增强剂是截短的gRNA且所述第二DNA结合调节增强剂是TAL效应蛋白。

[0759] 实施例P-92.根据实施例P-76的方法,其中所述第一增强子结合序列具有SEQ ID NO:26、SEQ ID NO:28、SEQ ID NO:30、SEQ ID NO:32、SEQ ID NO:34、SEQ ID NO:36、SEQ ID NO:38或SEQ ID NO:40的序列。

[0760] 实施例P-93.根据实施例P-76的方法,其中所述第二增强子结合序列具有SEQ ID NO:27、SEQ ID NO:29、SEQ ID NO:31、SEQ ID NO:33、SEQ ID NO:35、SEQ ID NO:37、SEQ ID NO:39或SEQ ID NO:41的序列。

[0761] 实施例P-94.一种置换细胞中的目标基因座的染色质的方法,所述方法包含:(1)向包含编码目标基因座的核酸的细胞中引入:(i)第一DNA结合调节增强剂,其中所述第一DNA结合调节增强剂不是所述细胞内源的;和(ii)第二DNA结合调节增强剂,其中所述第二DNA结合调节增强剂不是所述细胞内源的;(2)使所述第一DNA结合调节增强剂结合所述目标基因座的第一增强子结合序列;和(3)使所述第二DNA结合调节增强剂结合所述目标基因座的第二增强子结合序列,借此置换所述目标基因座的染色质。

[0762] 实施例P-95.根据实施例P-94的方法,其中所述引入第一DNA结合调节增强剂包含引入编码所述第一DNA结合调节增强剂的载体。

[0763] 实施例P-96.根据实施例P-94的方法,其中所述引入第一DNA结合调节增强剂包含引入编码所述第一DNA结合调节增强剂的mRNA。

[0764] 实施例P-97.根据实施例P-94的方法,其中所述引入第一DNA结合调节增强剂包含引入第一DNA结合蛋白或第一DNA结合核酸。

[0765] 实施例P-98.根据实施例P-94的方法,其中所述引入第二DNA结合调节增强剂包含引入编码所述第二DNA结合调节增强剂的载体。

[0766] 实施例P-99.根据实施例P-94的方法,其中所述引入第二DNA结合调节增强剂包含引入编码所述第二DNA结合调节增强剂的mRNA。

[0767] 实施例P-100.根据实施例P-94的方法,其中所述引入第二DNA结合调节增强剂包含引入第二DNA结合蛋白或第二DNA结合核酸。

[0768] 实施例P-101.根据实施例P-94的方法,其中相对于所述第一DNA结合调节增强剂的缺乏,在所述目标基因座发生同源重组的比率增加。

[0769] 实施例P-102.根据实施例P-94的方法,其中所述第一DNA结合调节增强剂是第一DNA结合蛋白或第一DNA结合核酸。

[0770] 实施例P-103.根据实施例P-94的方法,其中所述第一DNA结合调节增强剂是第一转录激活因子样(TAL)效应蛋白或截短的第一向导RNA(gRNA)。

[0771] 实施例P-104.根据实施例P-94的方法,其中所述第二DNA结合调节增强剂是第二DNA结合蛋白或第二DNA结合核酸。

[0772] 实施例P-105.根据实施例P-94的方法,其中所述第二DNA结合调节增强剂是TAL效

应蛋白或截短的gRNA。

[0773] 实施例106.根据实施例P-94的方法,其中所述第一DNA结合调节增强剂是第一TAL效应蛋白且所述第二DNA结合调节增强剂是第二TAL效应蛋白。

[0774] 实施例P-107.根据实施例P-94的方法,其中所述第一DNA结合调节增强剂是TAL效应蛋白且所述第二DNA结合调节增强剂是截短的gRNA。

[0775] 实施例P-108.根据实施例P-94的方法,其中所述第一DNA结合调节增强剂是截短的第一gRNA且所述第二DNA结合调节增强剂是截短的第二 gRNA。

[0776] 实施例P-109.根据实施例P-94的方法,其中所述第一DNA结合调节增强剂是截短的gRNA且所述第二DNA结合调节增强剂是TAL效应蛋白。

[0777] 实施例P-110.根据实施例P-94的方法,其中所述第一增强子结合序列具有SEQ ID NO:26、SEQ ID NO:28、SEQ ID NO:30、SEQ ID NO:32、SEQ ID NO:34、SEQ ID NO:36、SEQ ID NO:38或SEQ ID NO:40的序列。

[0778] 实施例P-111.根据实施例P-94的方法,其中所述第二增强子结合序列具有SEQ ID NO:27、SEQ ID NO:29、SEQ ID NO:31、SEQ ID NO:33、SEQ ID NO:35、SEQ ID NO:37、SEQ ID NO:39或SEQ ID NO:41的序列。

[0779] 实施例P-112.一种重构细胞中的目标基因座的染色质的方法,所述方法包含:(1)向包含编码目标基因座的核酸的细胞中引入:(i) 第一DNA结合调节增强剂,其中所述第一DNA结合调节增强剂不是所述细胞内源的;和(ii) 第二DNA结合调节增强剂,其中所述第二DNA结合调节增强剂不是所述细胞内源的;(2)使所述第一DNA结合调节增强剂结合所述目标基因座的第一增强子结合序列;和(3)使所述第二DNA结合调节增强剂结合所述目标基因座的第二增强子结合序列,借此置换所述目标基因座的染色质。

[0780] 实施例P-113.根据实施例P-112的方法,其中所述引入第一DNA结合调节增强剂包含引入编码所述第一DNA结合调节增强剂的载体。

[0781] 实施例P-114.根据实施例P-112的方法,其中所述引入第一DNA结合调节增强剂包含引入编码所述第一DNA结合调节增强剂的mRNA。

[0782] 实施例P-115.根据实施例P-112的方法,其中所述引入第一DNA结合调节增强剂包含引入第一DNA结合蛋白或第一DNA结合核酸。

[0783] 实施例P-116.根据实施例P-112的方法,其中所述引入第二DNA结合调节增强剂包含引入编码所述第二DNA结合调节增强剂的载体。

[0784] 实施例P-117.根据实施例P-112的方法,其中所述引入第二DNA结合调节增强剂包含引入编码所述第二DNA结合调节增强剂的mRNA。

[0785] 实施例P-118.根据实施例P-112的方法,其中所述引入第二DNA结合调节增强剂包含引入第二DNA结合蛋白或第二DNA结合核酸。

[0786] 实施例P-119.根据实施例P-112的方法,其中相对于所述第一DNA结合调节增强剂的缺乏,在所述目标基因座处发生同源重组的比率增加。

[0787] 实施例P-120.根据实施例P-112的方法,其中所述第一DNA结合调节增强剂是第一DNA结合蛋白或第一DNA结合核酸。

[0788] 实施例P-121.根据实施例P-112的方法,其中所述第一DNA结合调节增强剂是第一转录激活因子样(TAL)效应蛋白或截短的第一向导RNA(gRNA)。

[0789] 实施例P-122.根据实施例P-112的方法,其中所述第二DNA结合调节增强剂是第二DNA结合蛋白或第二DNA结合核酸。

[0790] 实施例P-123.根据实施例P-112的方法,其中所述第二DNA结合调节增强剂是TAL效应蛋白或截短的gRNA。

[0791] 实施例P-124.根据实施例P-112的方法,其中所述第一DNA结合调节增强剂是第一TAL效应蛋白且所述第二DNA结合调节增强剂是第二TAL效应蛋白。

[0792] 实施例P-125.根据实施例P-112的方法,其中所述第一DNA结合调节增强剂是TAL效应蛋白且所述第二DNA结合调节增强剂是截短的gRNA。

[0793] 实施例P-126.根据实施例P-112的方法,其中所述第一DNA结合调节增强剂是截短的第一gRNA且所述第二DNA结合调节增强剂是截短的第二gRNA。

[0794] 实施例P-127.根据实施例P-112的方法,其中所述第一DNA结合调节增强剂是截短的gRNA且所述第二DNA结合调节增强剂是TAL效应蛋白。

[0795] 实施例P-128.一种增强调节蛋白或调节复合物在细胞中的目标基因座处的活性的方法,所述方法包含:(1)向包含编码目标基因座的核酸的细胞中引入:(i)能够结合所述目标基因座的调节子结合序列的第一调节蛋白或第一调节复合物,其中所述调节子结合序列包含调节位点;和(ii)能够结合所述目标基因座的第一增强子结合序列的第一DNA结合调节增强剂;以及(2)使所述第一DNA结合调节增强剂结合所述第一增强子结合序列,借此增强所述第一调节蛋白或所述第一调节复合物在细胞中的目标基因座处的活性。

[0796] 实施例P-129.根据实施例P-128的方法,进一步包含引入能够结合所述目标基因座的第二增强子结合序列的第二DNA结合调节增强剂。

[0797] 实施例P-130.根据实施例P-128的方法,其中所述引入第一DNA结合调节增强剂包含引入编码所述第一DNA结合调节增强剂的载体。

[0798] 实施例P-131.根据实施例P-128的方法,其中所述引入第一DNA结合调节增强剂包含引入编码所述第一DNA结合调节增强剂的mRNA。

[0799] 实施例P-132.根据实施例P-128的方法,其中所述引入第一DNA结合调节增强剂包含引入第一DNA结合蛋白或第一DNA结合核酸。

[0800] 实施例P-133.根据实施例P-129的方法,其中所述引入第二DNA结合调节增强剂包含引入编码所述第二DNA结合调节增强剂的载体。

[0801] 实施例P-134.根据实施例P-129的方法,其中所述引入第二DNA结合调节增强剂包含引入编码所述第二DNA结合调节增强剂的mRNA。

[0802] 实施例P-135.根据实施例P-129的方法,其中所述引入第二DNA结合调节增强剂包含引入第二DNA结合蛋白或第二DNA结合核酸。

[0803] 实施例P-136.根据实施例P-128的方法,其中所述第一调节蛋白或所述第一调节复合物不是所述细胞内源的。

[0804] 实施例P-137.根据实施例P-128的方法,其中相对于所述第一DNA结合调节增强剂的缺乏,在所述目标基因座处发生同源重组的比率增加。

[0805] 实施例P-138.根据实施例P-129的方法,其中所述第二增强子结合序列通过所述调节子结合序列连接至所述第一增强子结合序列。

[0806] 实施例P-139.根据实施例P-128的方法,进一步包含引入能够结合所述调节子结

合序列的第二调节蛋白或第二调节复合物。

[0807] 实施例P-140.根据实施例P-128的方法,其中所述引入第一调节蛋白包含引入编码所述第一调节蛋白的载体。

[0808] 实施例P-141.根据实施例P-128的方法,其中所述引入第一调节蛋白包含引入编码所述第一调节蛋白的mRNA。

[0809] 实施例P-142.根据实施例P-128的方法,其中所述引入第一调节蛋白包含引入第一调节蛋白。

[0810] 实施例P-143.根据实施例P-128的方法,其中所述引入第一调节复合物包含引入编码所述第一调节复合物的载体。

[0811] 实施例P-144.根据实施例P-128的方法,其中所述引入第一调节复合物包含引入编码所述第一调节复合物的mRNA。

[0812] 实施例P-145.根据实施例P-128的方法,其中所述引入第一调节复合物包含引入第一调节复合物。

[0813] 实施例P-146.根据实施例P-139的方法,其中所述引入第二调节蛋白包含引入编码所述第二调节蛋白的载体。

[0814] 实施例P-147.根据实施例P-139的方法,其中所述引入第二调节蛋白包含引入编码所述第二调节蛋白的mRNA。

[0815] 实施例P-148.根据实施例P-139的方法,其中所述引入第二调节蛋白包含引入第二调节蛋白。

[0816] 实施例P-149.根据实施例P-139的方法,其中所述引入第二调节复合物包含引入编码所述第二调节复合物的载体。

[0817] 实施例P-150.根据实施例P-139的方法,其中所述引入第二调节复合物包含引入编码所述第二调节复合物的mRNA。

[0818] 实施例P-151.根据实施例P-139的方法,其中所述引入第二调节复合物包含引入第二调节复合物。

[0819] 实施例P-152.根据实施例P-139的方法,其中所述第一调节蛋白或所述第二调节蛋白包含DNA结合蛋白或DNA调节酶。

[0820] 实施例P-153.根据实施例P-152的方法,其中所述DNA结合蛋白是转录抑制因子或转录激活因子。

[0821] 实施例P-154.根据实施例P-152的方法,其中所述DNA调节酶是核酸酶、脱氨酶、甲基化酶或脱甲基酶。

[0822] 实施例P-155.根据实施例P-128的方法,其中所述第一调节蛋白或所述第二调节蛋白包含组蛋白调节酶。

[0823] 实施例P-156.根据实施例P-155的方法,其中所述组蛋白调节酶是脱乙酰酶或乙酰酶。

[0824] 实施例P-157.根据实施例P-128的方法,其中所述第一调节蛋白是第一 DNA结合蛋白核酸酶结合物。

[0825] 实施例P-158.根据实施例P-139的方法,其中所述第二调节蛋白是第二 DNA结合蛋白核酸酶结合物。

[0826] 实施例P-159.根据实施例P-158的方法,其中所述第一DNA结合蛋白核酸酶结合物包含第一核酸酶且所述第二DNA结合蛋白核酸酶结合物包含第二核酸酶。

[0827] 实施例P-160.根据实施例P-159的方法,其中所述第一核酸酶与所述第二核酸酶形成二聚体。

[0828] 实施例P-161.根据实施例P-159的方法,其中所述第一核酸酶和所述第二核酸酶独立地是转录激活因子样效应物核酸酶 (TALEN)。

[0829] 实施例P-162.根据实施例P-159的方法,其中所述第一DNA结合蛋白核酸酶结合物包含可操作地连接到第一核酸酶 (TALEN) 的第一转录激活因子样 (TAL) 效应域。

[0830] 实施例P-163.根据实施例P-159的方法,其中所述第一DNA结合蛋白核酸酶结合物包含可操作地连接到第一FokI核酸酶的第一TAL效应域。

[0831] 实施例P-164.根据实施例P-159的方法,其中所述第二DNA结合蛋白核酸酶结合物包含可操作地连接到第二核酸酶 (TALEN) 的第二TAL效应域。

[0832] 实施例P-165.根据实施例P-159的方法,其中所述第二DNA结合蛋白核酸酶结合物包含可操作地连接到第二FokI核酸酶的第二TAL效应域。

[0833] 实施例P-166.根据实施例P-159的方法,其中所述第一DNA结合蛋白核酸酶结合物包含第一锌指核酸酶。

[0834] 实施例P-167.根据实施例P-159的方法,其中所述第二DNA结合蛋白核酸酶结合物包含第一锌指核酸酶。

[0835] 实施例P-168.根据实施例P-128的方法,其中所述第一调节复合物是第一核糖核蛋白复合物。

[0836] 实施例P-169.根据实施例P-139的方法,其中所述第二调节复合物是第二核糖核蛋白复合物。

[0837] 实施例P-170.根据实施例P-168的方法,其中所述第一核糖核蛋白复合物包含结合至gRNA的CRISPR相关蛋白9 (Cas9) 结构域或结合至向导DNA (gDNA) 的阿格蛋白结构域。

[0838] 实施例P-171.根据实施例P-169的方法,其中所述第二核糖核蛋白复合物包含结合至gRNA的CRISPR相关蛋白9 (Cas9) 结构域或结合至向导DNA (gDNA) 的阿格蛋白结构域。

[0839] 实施例P-172.根据实施例P-139的方法,其中所述第一调节蛋白、所述第一调节复合物、所述第二调节蛋白或所述第二调节复合物不是所述细胞内源的。

[0840] 实施例P-173.根据实施例P-139的方法,其中所述第一调节蛋白和所述第二调节蛋白不是所述细胞内源的。

[0841] 实施例P-174.根据实施例P-139的方法,其中所述第一调节复合物和所述第二调节复合物不是所述细胞内源的。

[0842] 实施例P-175.根据实施例P-168的方法,其中所述第一DNA结合调节增强剂或所述第二DNA结合调节增强剂不是所述细胞内源的。

[0843] 实施例P-176.根据实施例P-129的方法,其中所述第一DNA结合调节增强剂和所述第二DNA结合调节增强剂不是所述细胞内源的。

[0844] 实施例P-177.根据实施例P-128的方法,其中所述第一DNA结合调节增强剂是第一DNA结合蛋白或第一DNA结合核酸。

[0845] 实施例P-178.根据实施例P-128的方法,其中所述第一DNA结合调节增强剂是第一

转录激活因子样 (TAL) 效应蛋白或截短的第一向导RNA (gRNA)。

[0846] 实施例P-179.根据实施例P-139的方法,其中所述第二DNA结合调节增强剂是第二DNA结合蛋白或第二DNA结合核酸。

[0847] 实施例P-180.根据实施例P-129的方法,其中所述第二DNA结合调节增强剂是TAL效应蛋白或截短的gRNA。

[0848] 实施例P-181.根据实施例P-129的方法,其中所述第一DNA结合调节增强剂是第一TAL效应蛋白且所述第二DNA结合调节增强剂是第二TAL效应蛋白。

[0849] 实施例P-182.根据实施例P-129的方法,其中所述第一DNA结合调节增强剂是TAL效应蛋白且所述第二DNA结合调节增强剂是截短的gRNA。

[0850] 实施例P-183.根据实施例P-129的方法,其中所述第一DNA结合调节增强剂是截短的第一gRNA且所述第二DNA结合调节增强剂是截短的第二 gRNA。

[0851] 实施例P-184.根据实施例P-129的方法,其中所述第一DNA结合调节增强剂是截短的gRNA且所述第二DNA结合调节增强剂是TAL效应蛋白。

[0852] 实施例P-185.根据实施例P-139的方法,其中所述第一调节蛋白是第一 DNA结合核酸酶结合物且所述第二调节蛋白是第二DNA结合核酸酶结合物。

[0853] 实施例P-186.根据实施例P-139的方法,其中所述第一调节蛋白是DNA 结合核酸酶结合物且所述第二调节复合物是核糖核蛋白复合物。

[0854] 实施例P-187.根据实施例P-139的方法,其中所述第一调节复合物是第一核糖核蛋白复合物且所述第二调节复合物是第二核糖核蛋白复合物。

[0855] 实施例P-188.根据实施例P-139的方法,其中所述第一调节复合物是核糖核蛋白复合物且所述第二调节蛋白是DNA结合核酸酶结合物。

[0856] 实施例P-189.根据实施例P-129的方法,其中所述第一增强子结合序列和/或第二增强子结合序列与所述调节子结合序列相隔少于200个核苷酸、少于150个核苷酸、少于100个核苷酸或少于50个核苷酸。

[0857] 实施例P-190.根据实施例P-129的方法,其中所述第一增强子结合序列和/或第二增强子结合序列与所述调节子结合序列相隔4至30个核苷酸或7至 30个核苷酸。

[0858] 实施例P-191.根据实施例P-129的方法,其中所述第一增强子结合序列和/或第二增强子结合序列与所述调节子结合序列相隔4个核苷酸、7个核苷酸、12个核苷酸、20个核苷酸或30个核苷酸。

[0859] 实施例P-192.根据实施例P-129的方法,其中所述第一增强子结合序列和/或第二增强子结合序列与所述调节子结合序列相隔少于200个核苷酸、少于150个核苷酸、少于100个核苷酸,或少于50个核苷酸。

[0860] 实施例P-193.根据实施例P-129的方法,其中所述第一增强子结合序列和/或所述第二增强子结合序列与所述调节位点相隔10至40个核苷酸。

[0861] 实施例P-194.根据实施例P-129的方法,其中所述第一增强子结合序列和/或所述第二增强子结合序列与所述调节位点相隔33个核苷酸。

[0862] 实施例P-195.根据实施例P-139的方法,其中所述第一DNA结合调节增强剂或所述第二DNA结合调节增强剂增强所述第一调节蛋白、所述第一调节复合物、所述第二调节蛋白或所述第二调节复合物在所述调节位点处的活性。

[0863] 实施例P-196.一种调节细胞中的目标基因座的方法,所述方法包含:(1) 向包含编码目标基因座的核酸的细胞中引入:(i) 能够结合所述目标基因座的调节子结合序列的第一调节蛋白或第一调节复合物,其中所述调节子结合序列包含调节位点;和(ii) 能够结合所述目标基因座的第一增强子结合序列的第一DNA结合调节增强剂;以及(2) 使所述第一调节蛋白或所述第一调节复合物调节所述调节位点,借此调节细胞中的所述目标基因座。

[0864] 实施例P-197.根据实施例P-196的方法,进一步包含引入能够结合所述目标基因座的第二增强子结合序列的第二DNA结合调节增强剂。

[0865] 实施例P-198.根据实施例P-196的方法,其中所述引入第一DNA结合调节增强剂包含向细胞中引入:(1) 编码所述第一DNA结合调节增强剂的载体;(2) 编码所述第一DNA结合调节增强剂的mRNA;或(3) 第一DNA结合调节增强。

[0866] 实施例P-199.根据实施例P-197的方法,其中所述引入第二DNA结合调节增强剂包含向细胞中引入:(1) 编码所述第一DNA结合调节增强剂的载体;(2) 编码所述第一DNA结合调节增强剂的mRNA;或(3) 第一DNA结合调节增强。

[0867] 实施例P-200.根据实施例P-199的方法,其中所述引入第二DNA结合调节增强剂包含引入编码所述第二DNA结合调节增强剂的mRNA。

[0868] 实施例P-201.根据实施例P-197的方法,其中所述引入第二DNA结合调节增强剂包含引入第二DNA结合蛋白或第二DNA结合核酸。

[0869] 实施例P-202.根据实施例P-196的方法,其中所述第一调节蛋白或所述第一调节复合物不是所述细胞内源的。

[0870] 实施例P-203.根据实施例P-196的方法,其中相对于所述第一DNA结合调节增强剂的缺乏,在所述目标基因座处发生同源重组的比率增加。

[0871] 实施例P-204.根据实施例P-197的方法,其中所述第二增强子结合序列通过所述调节子结合序列连接至所述第一增强子结合序列。

[0872] 实施例P-205.根据实施例P-196的方法,进一步包含引入能够结合所述调节子结合序列的第二调节蛋白或第二调节复合物。

[0873] 实施例P-206.根据实施例P-196的方法,其中所述引入第一调节蛋白包含引入编码所述第一调节蛋白的载体。

[0874] 实施例P-207.根据实施例P-196的方法,其中所述引入第一调节蛋白包含引入编码所述第一调节蛋白的mRNA。

[0875] 实施例P-208.根据实施例P-196的方法,其中所述引入第一调节蛋白包含引入第一调节蛋白。

[0876] 实施例P-209.根据实施例P-196的方法,其中所述引入第一调节复合物包含引入编码所述第一调节复合物的载体。

[0877] 实施例P-210.根据实施例P-196的方法,其中所述引入第一调节复合物包含引入编码所述第一调节复合物的mRNA。

[0878] 实施例P-211.根据实施例P-196的方法,其中所述引入第一调节复合物包含引入第一调节复合物。

[0879] 实施例P-212.根据实施例P-205的方法,其中所述引入第二调节蛋白包含引入编码所述第二调节蛋白的载体。

[0880] 实施例P-213.根据实施例P-205的方法,其中所述引入第二调节蛋白包含引入编码所述第二调节蛋白的mRNA。

[0881] 实施例P-214.根据实施例P-205的方法,其中所述引入第二调节蛋白包含引入第二调节蛋白。

[0882] 实施例P-215.根据实施例P-205的方法,其中所述引入第二调节复合物包含引入编码所述第二调节复合物的载体。

[0883] 实施例P-216.根据实施例P-205的方法,其中所述引入第二调节复合物包含引入编码所述第二调节复合物的mRNA。

[0884] 实施例P-217.根据实施例P-205的方法,其中所述引入第二调节复合物包含引入第二调节复合物。

[0885] 实施例P-218.根据实施例P-205的方法,其中所述第一调节蛋白或所述第二调节蛋白包含DNA结合蛋白或DNA调节酶。

[0886] 实施例P-219.根据实施例P-218的方法,其中所述DNA结合蛋白是转录抑制因子或转录激活因子。

[0887] 实施例P-220.根据实施例P-218的方法,其中所述DNA调节酶是核酸酶、脱氨酶、甲基化酶或脱甲基酶。

[0888] 实施例P-221.根据实施例P-205的方法,其中所述第一调节蛋白或所述第二调节蛋白包含组蛋白调节酶。

[0889] 实施例P-222.根据实施例P-221的方法,其中所述组蛋白调节酶是脱乙酰酶或乙酰酶。

[0890] 实施例P-223.根据实施例P-196的方法,其中所述第一调节蛋白是第一 DNA结合蛋白核酸酶结合物。

[0891] 实施例P-224.根据实施例P-205的方法,其中所述第二调节蛋白是第二DNA结合蛋白核酸酶结合物。

[0892] 实施例P-225.根据实施例P-224的方法,其中所述第一DNA结合蛋白核酸酶结合物包含第一核酸酶且所述第二DNA结合蛋白核酸酶结合物包含第二核酸酶。

[0893] 实施例P-226.根据实施例P-225的方法,其中所述第一核酸酶与所述第二核酸酶形成二聚体。

[0894] 实施例P-227.根据实施例P-225的方法,其中所述第一核酸酶和所述第二核酸酶独立地是转录激活因子样效应物核酸酶 (TALEN)。

[0895] 实施例P-228.根据实施例P-225的方法,其中所述第一DNA结合蛋白核酸酶结合物包含可操作地连接到第一核酸酶 (TALEN) 的第一转录激活因子样 (TAL) 效应域。

[0896] 实施例P-229.根据实施例P-228的方法,其中所述第一DNA结合蛋白核酸酶结合物包含可操作地连接到第一FokI核酸酶的第一TAL效应域。

[0897] 实施例P-230.根据实施例P-227的方法,其中所述第二DNA结合蛋白核酸酶结合物包含可操作地连接到第二核酸酶 (TALEN) 的第二TAL效应域。

[0898] 实施例P-231.根据实施例P-230的方法,其中所述第二DNA结合蛋白核酸酶结合物包含可操作地连接到第二FokI核酸酶的第二TAL效应域。

[0899] 实施例P-232.根据实施例P-196的方法,其中所述第一DNA结合蛋白核酸酶结合物

包含第一锌指核酸酶。

[0900] 实施例P-233.根据实施例P-205的方法,其中所述第二DNA结合蛋白核酸酶结合物包含第一锌指核酸酶。

[0901] 实施例P-234.根据实施例P-196的方法,其中所述第一调节复合物是第一核糖核蛋白复合物。

[0902] 实施例P-235.根据实施例P-197的方法,其中所述第二调节复合物是第二核糖核蛋白复合物。

[0903] 实施例P-236.根据实施例P-234的方法,其中所述第一核糖核蛋白复合物包含结合至gRNA的CRISPR相关蛋白9 (Cas9) 结构域或结合至向导DNA (gDNA) 的阿格蛋白结构域。

[0904] 实施例P-237.根据实施例P-235的方法,其中所述第二核糖核蛋白复合物包含结合至gRNA的CRISPR相关蛋白9 (Cas9) 结构域或结合至向导DNA (gDNA) 的阿格蛋白结构域。

[0905] 实施例P-238.根据实施例P-205的方法,其中所述第一调节蛋白、所述第一调节复合物、所述第二调节蛋白或所述第二调节复合物不是所述细胞内源的。

[0906] 实施例P-239.根据实施例P-205的方法,其中所述第一调节蛋白和所述第二调节蛋白不是所述细胞内源的。

[0907] 实施例P-240.根据实施例P-205的方法,其中所述第一调节复合物和所述第二调节复合物不是所述细胞内源的。

[0908] 实施例P-241.根据实施例P-197的方法,其中所述第一DNA结合调节增强剂或所述第二DNA结合调节增强剂不是所述细胞内源的。

[0909] 实施例P-242.根据实施例P-197的方法,其中所述第一DNA结合调节增强剂和所述第二DNA结合调节增强剂不是所述细胞内源的。

[0910] 实施例P-243.根据实施例P-196的方法,其中所述第一DNA结合调节增强剂是第一DNA结合蛋白或第一DNA结合核酸。

[0911] 实施例P-244.根据实施例P-196的方法,其中所述第一DNA结合调节增强剂是第一转录激活因子样 (TAL) 效应蛋白或截短的第一向导RNA (gRNA) 。

[0912] 实施例P-245.根据实施例P-197的方法,其中所述第二DNA结合调节增强剂是第二DNA结合蛋白或第二DNA结合核酸。

[0913] 实施例P-246.根据实施例P-197的方法,其中所述第二DNA结合调节增强剂是TAL效应蛋白或截短的gRNA。

[0914] 实施例P-247.根据实施例P-197的方法,其中所述第一DNA结合调节增强剂是第一TAL效应蛋白且所述第二DNA结合调节增强剂是第二TAL效应蛋白。

[0915] 实施例P-248.根据实施例P-197的方法,其中所述第一DNA结合调节增强剂是TAL效应蛋白且所述第二DNA结合调节增强剂是截短的gRNA。

[0916] 实施例P-249.根据实施例P-197的方法,其中所述第一DNA结合调节增强剂是截短的第一gRNA且所述第二DNA结合调节增强剂是截短的第二 gRNA。

[0917] 实施例P-250.根据实施例P-197的方法,其中所述第一DNA结合调节增强剂是截短的gRNA且所述第二DNA结合调节增强剂是TAL效应蛋白。

[0918] 实施例P-251.根据实施例P-205的方法,其中所述第一调节蛋白是第一 DNA结合核酸酶结合物且所述第二调节蛋白是第二DNA结合核酸酶结合物。

[0919] 实施例P-252.根据实施例P-205的方法,其中所述第一调节蛋白是DNA 结合核酸酶结合物且所述第二调节复合物是核糖核蛋白复合物。

[0920] 实施例P-253.根据实施例P-252的方法,其中所述第一调节复合物是第一核糖核蛋白复合物且所述第二调节复合物是第二核糖核蛋白复合物。

[0921] 实施例P-254.根据实施例P-205的方法,其中所述第一调节复合物是核糖核蛋白复合物且所述第二调节蛋白是DNA结合核酸酶结合物。

[0922] 实施例P-255.根据实施例P-196的方法,其中所述第一增强子结合序列与所述调节子结合序列相隔少于200个核苷酸、少于150个核苷酸、少于100 个核苷酸,或少于50个核苷酸。

[0923] 实施例P-256.根据实施例P-196的方法,其中所述第一增强子结合序列与所述调节子结合序列相隔4至30个核苷酸或7至30个核苷酸。

[0924] 实施例P-257.根据实施例P-196的方法,其中所述第一增强子结合序列与所述调节子结合序列相隔4个核苷酸、7个核苷酸、12个核苷酸、20个核苷酸或30个核苷酸。

[0925] 实施例P-258.根据实施例P-197的方法,其中所述第二增强子结合序列与所述调节子结合序列相隔少于200个核苷酸、少于150个核苷酸、少于100 个核苷酸或少于50个核苷酸。

[0926] 实施例P-259.根据实施例P-197的方法,其中所述第二增强子结合序列与所述调节子结合序列相隔4至30个核苷酸或7至30个核苷酸。

[0927] 实施例P-260.根据实施例P-197的方法,其中所述第二增强子结合序列与所述调节子结合序列相隔4个核苷酸、7个核苷酸、12个核苷酸、20个核苷酸或30个核苷酸。

[0928] 实施例P-261.根据实施例P-197的方法,其中所述第一增强子结合序列或所述第二增强子结合序列与所述调节位点相隔10至40个核苷酸。

[0929] 实施例P-262.根据实施例P-197的方法,其中所述第一增强子结合序列或所述第二增强子结合序列与所述调节位点相隔33个核苷酸。

[0930] 实施例P-263.根据实施例P-197的方法,其中所述第一DNA结合调节增强剂或所述第二DNA结合调节增强剂增强所述第一调节蛋白、所述第一调节复合物、所述第二调节蛋白或所述第二调节复合物在所述调节位点处的活性。

[0931] 实施例P-264.一种包含编码目标基因座调节复合物的核酸的细胞,所述复合物包含:(i) 包含第一增强子结合序列和含有调节位点的调节子结合序列的目标基因座;(ii) 结合至所述调节子结合序列的第一调节蛋白或第一调节复合物;和(iii) 结合至所述第一增强子结合序列的第一DNA结合调节增强剂。

[0932] 实施例P-265.根据实施例P-264的细胞,其中所述目标基因座进一步包含通过所述调节子结合序列与所述第一增强子结合序列连接的第二增强子结合序列。

[0933] 实施例P-266.根据实施例P-264的细胞,包含结合至所述第二增强子结合序列的第二DNA结合调节增强剂。

[0934] 实施例P-267.一种包含编码目标基因座复合物的核酸的细胞,所述复合物包含:(i) 包含第一增强子结合序列的目标基因座;和(ii) 结合至所述第一增强子结合序列的第一DNA结合调节增强剂,其中所述第一DNA结合调节增强剂不是所述细胞内源的且其中相对于所述第一DNA结合调节增强剂的缺乏,所述第一DNA结合调节增强剂能够增强所述目标基

因座的可及性。

[0935] 实施例P-268.一种包含编码目标基因座复合物的核酸的细胞,所述复合物包含:
(1) 目标基因座,所述目标基因座包含:(i) 第一增强子结合序列;和(ii) 第二增强子结合序列;
(2) 结合至所述目标基因座的所述第一增强子结合序列的第一DNA结合调节增强剂,其中所述第一DNA结合调节增强剂不是所述细胞内源的;以及
(3) 结合至所述目标基因座的所述第二增强子结合序列的第二DNA结合调节增强剂,其中所述第二DNA结合调节增强剂不是所述细胞内源的,其中相对于所述第一DNA结合调节增强剂和所述第二DNA结合调节增强剂的缺乏,所述第一DNA结合调节增强剂和所述第二DNA结合调节增强剂能够增强所述目标基因座的可及性。

[0936] 实施例P-269.一种试剂盒,其包含:(i) 第一调节蛋白或第一调节复合物;以及(ii) 第一DNA结合调节增强剂。

[0937] 实施例P-270.一种用于改变存在于细胞内的内源核酸分子的方法,所述方法包含将供体DNA分子引入所述细胞内,其中所述供体DNA分子可操作地连接到一个或多个细胞内靶向部分,所述细胞内靶向部分能够使所述供体DNA分子定位到所述细胞中的所述内源核酸分子所处的位置。

[0938] 实施例P-271.根据实施例P-270的方法,其中所述细胞中的所述内源核酸分子所处的位置位于细胞核、线粒体或叶绿体中。

[0939] 实施例P-272.根据实施例P-270的方法,其中所述一个或多个细胞内目标部分是核定位信号。

[0940] 实施例P-273.根据实施例P-270的方法,其中所述供体DNA分子的长度为约25至约8,000个核苷酸。

[0941] 实施例P-274.根据实施例P-270的方法,其中所述供体DNA分子是单链的。

[0942] 实施例P-275.根据实施例P-270的方法,其中所述供体DNA分子在至少一个末端的50个核苷酸内具有一个或多个核酸酶抗性基团。

[0943] 实施例P-276.根据实施例P-275的方法,其中所述核酸酶抗性基团是硫代磷酸酯基团。

[0944] 实施例P-277.根据实施例P-276的方法,其中存在两个硫代磷酸酯基团,所述硫代磷酸酯基团位于至少一个末端的50个核苷酸内。

[0945] 实施例P-278.根据实施例P-270的方法,其中所述供体DNA分子含有正向可选标记物和负向可选标记物。

[0946] 实施例P-279.根据实施例P-278的方法,其中所述负向可选标记物是单纯疱疹病毒胸苷激酶。

[0947] 实施例P-280.根据实施例P-270的方法,其中所述供体DNA分子具有序列与存在于所述细胞中的目标基因座互补的两个区域。

[0948] 实施例P-281.根据实施例P-278的方法,其中所述负向可选标记物定位于供体DNA分子的两个具有序列互补性的区域之间。

[0949] 实施例P-282.根据实施例P-278的方法,其中所述负向可选标记物不定位于所述供体DNA分子的两个具有序列互补性的区域之间。

[0950] 实施例P-283.根据实施例P-270的方法,其中使所述细胞与以下中的一种或多种

接触：(1) 一种或多种核酸切割实体；(2) 编码核酸切割实体中的至少一种组分的一种或多种核酸分子；(3) 一种或多种DNA结合调节增强剂；(4) 编码DNA结合调节增强剂中的至少一种组分的一种或多种核酸分子；或 (5) 一种或更多种非同源末端连接 (NHEJ) 抑制剂。

[0951] 实施例P-284. 根据实施例P-283的方法, 其中所述一种或更多种非同源末端连接 (NHEJ) 抑制剂是DNA依赖性蛋白激酶抑制剂。

[0952] 实施例P-285. 根据实施例P-284的方法, 其中所述一种或更多种非同源末端连接 (NHEJ) 抑制剂中的至少一种选自由以下组成的组：(1) Nu7026、(2) Nu7441、(3) Ku-0060648、(4) DMNB、(5) ETP 45658、(6) LTURM 34, 和 (7) P1 103盐酸盐。

[0953] 实施例P-286. 根据实施例P-283的方法, 其中所述一种或多种核酸切割实体中的至少一种选自以下组成的组：(1) 锌指核酸酶、(2) TAL效应核酸酶和 (3) CRISPR复合物。

[0954] 实施例P-287. 根据实施例P-283的方法, 其中所述一种或更多种DNA结合调节增强剂中的至少一种选自以下组成的组：(1) 锌指核酸酶、(2) TAL效应核酸酶和 (3) CRISPR复合物。

[0955] 实施例P-288. 根据实施例P-270的方法, 其中所述一种或更多种DNA结合调节增强剂中的至少一种设计成在所述目标基因座的50个核苷酸内结合。

[0956] 实施例P-289. 一种用于在真核细胞中进行同源重组的方法, 所述方法包含使所述细胞与以下接触：(1) 供体DNA分子和 (2) (i) 核酸切割实体、(ii) 编码核酸切割实体的核酸或 (iii) 核酸切割实体的至少一种组分和编码核酸切割实体的至少一种组分的核酸, 其中所述供体DNA分子结合至细胞内靶向部分, 所述细胞内靶向部分能够使所述供体DNA分子定位至所述细胞中的所述内源核酸分子所处的位置。

[0957] 实施例P-290. 根据实施例P-289的方法, 进一步使所述细胞与以下中的一种或多种接触：(1) 一种或更多种非同源末端连接 (NHEJ) 抑制剂、(2) 一种或更多种DNA结合调节增强剂、(3) 一种或多种编码DNA结合调节增强剂的核酸, 以及 (4) 一种或更多种DNA结合调节增强剂的至少一种组分和编码一种或更多种DNA结合调节增强剂的至少一种组分的核酸。

[0958] 实施例P-291. 一种包含DNA分子的组合物, 其中所述DNA分子共价连接到一个或多个细胞内靶向部分并且其中所述DNA分子的长度为约25个核苷酸到约8,000个核苷酸。

[0959] 实施例P-292. 根据实施例P-291的组合物, 其中所述DNA分子是供体 DNA分子。

[0960] 实施例P-293. 根据实施例P-291的组合物, 其中所述一个或更多个细胞内靶向部分是核定位信号。

[0961] 实施例P-294. 根据实施例P-291的组合物, 其中两个或更多个细胞内靶向部分共价连接到DNA分子。

[0962] 实施例P-295. 根据实施例P-291的组合物, 其中所述一个或更多个细胞内靶向部分选自以下组成的组：(1) 核定位信号、(2) 叶绿体靶向信号, 和 (3) 线粒体靶向信号。

[0963] 实施例P-296. 一种Cas9蛋白, 其包含两个或更多个双分型核定位信号。

[0964] 实施例P-297. 根据实施例P-296的Cas9蛋白, 其中所述两个或更多个双分型核定位信号定位于至少一个末端的二十个氨基酸内。

[0965] 实施例P-298. 根据实施例P-296的Cas9蛋白, 其中所述两个或更多个双分型核定位信号个别地定位于所述蛋白质的N端和C端的二十个氨基酸内。

[0966] 实施例P-299. 根据实施例P-296的Cas9蛋白, 其中所述两个或更多个双分型核定

位信号包含不同的氨基酸序列。

[0967] 实施例P-300.根据实施例P-296的Cas9蛋白,其进一步包含至少一种单分型核定位信号。

[0968] 实施例P-301.根据实施例P-296的Cas9蛋白,其进一步包含亲和标签。

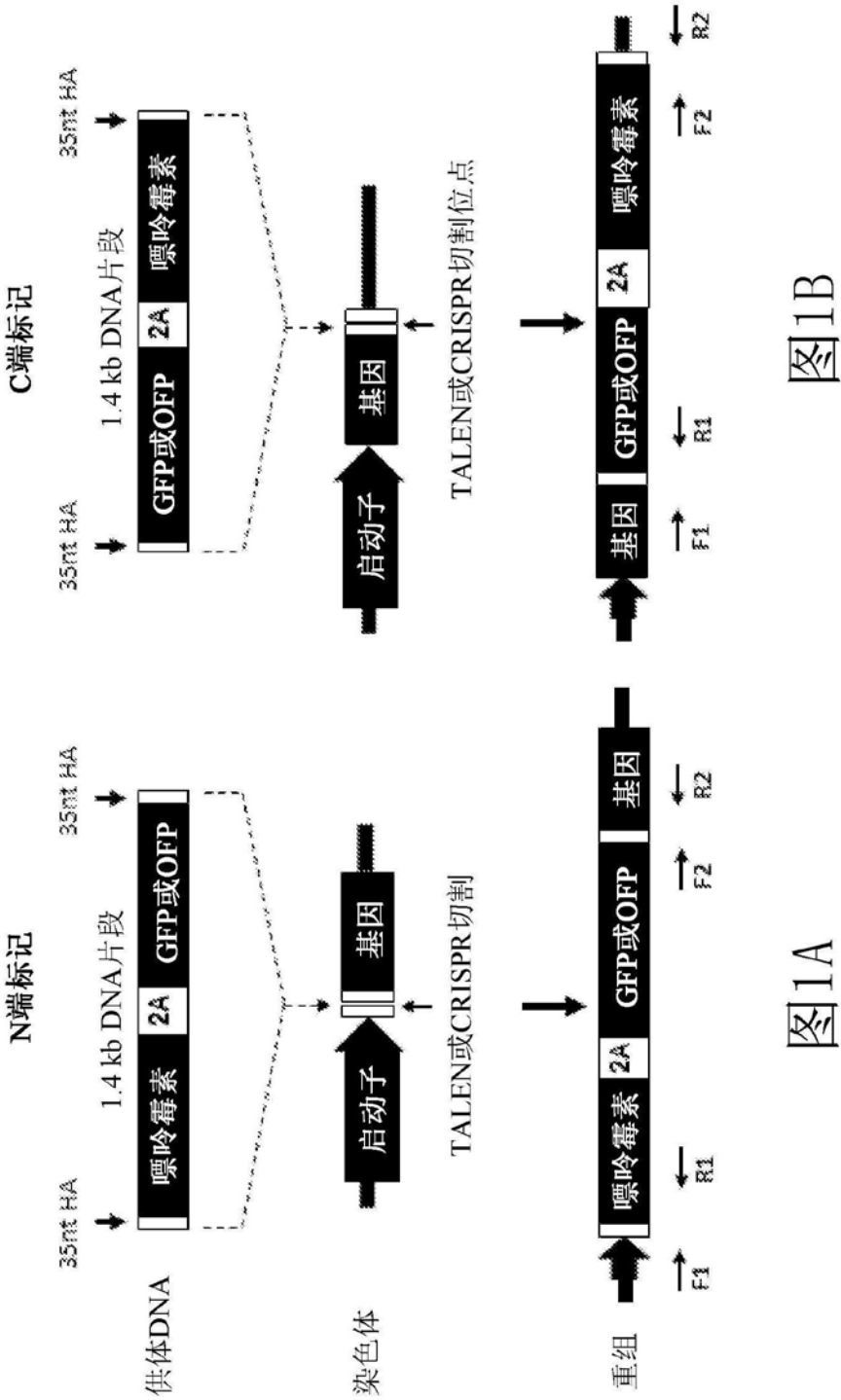
[0969] 实施例P-302.根据实施例P-296的Cas9蛋白,其中所述核定位信号中的至少一种具有选自以下组成的组的氨基酸序列:(1) KRTADGSEFESPCKKKRKVE、(2) KRTADGSEFESPCKARKVE、(3) KRTADGSEFESPCKKAKVE、(4) KRPAATKKAGQAKKKK和(5) KRTADGSEFEPAAKRVKLDE。

[0970] 实施例P-303.根据实施例P-296的Cas9蛋白,其中所述核定位信号中的至少一种具有选自以下组成的组的氨基酸序列:(1) KRX5-15 KKN1N2KV、(2) KRX (5-15) K (K/R) (K/R) 1-2,和(3) KRX (5-15) K (K/R) X (K/R) 1-2,其中X 是长度为5至15个氨基酸的氨基酸序列且其中N1是L或A,且其中N2是 L、A或R。

[0971] 实施例P-304.根据实施例P-296的Cas9蛋白,其包含图42中所示的氨基酸序列。

[0972] 参考文献

[0973] Liang等人,“通过改进gRNA、Cas9核酸酶和供体DNA的设计和递送来增强CRISPR/Cas9介导的精确基因组编辑(Enhanced CRISPR/Cas9-mediated precise genome editing by improved design and delivery of gRNA,Cas9 nuclease, and donor DNA)”,《生物技术杂志(J.Biotechnol)》241:136-146(2017)。



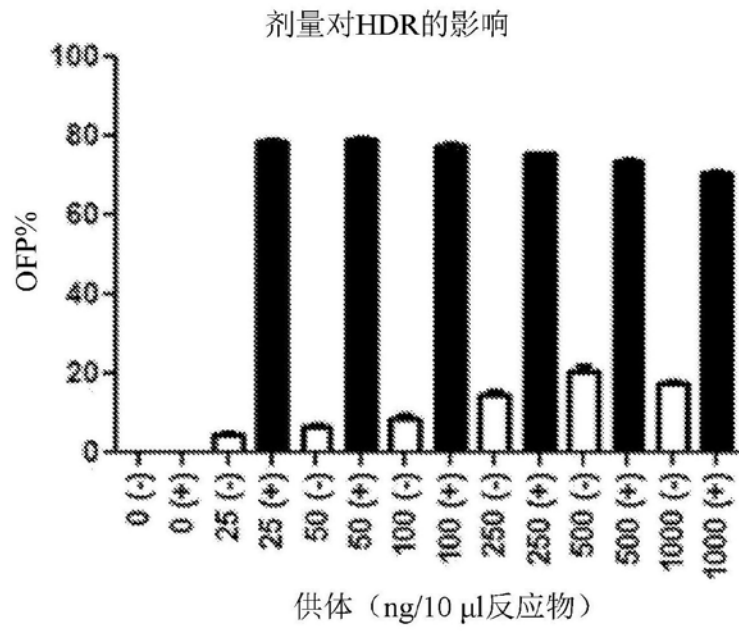


图2A

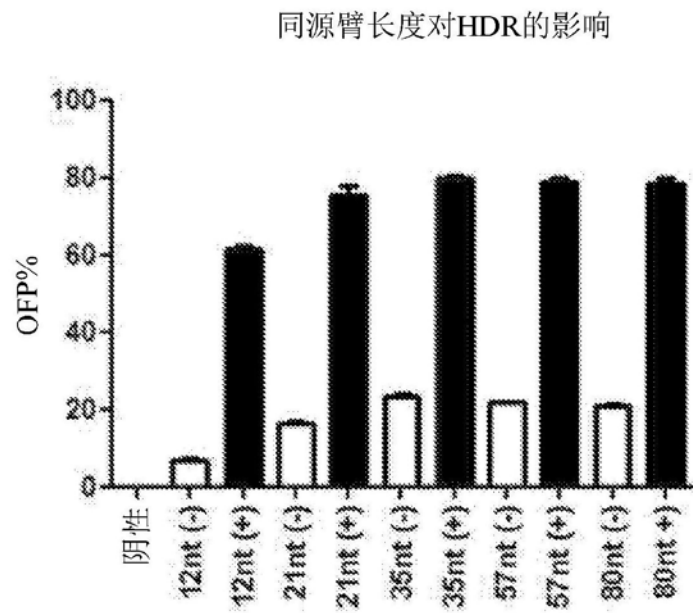


图2B

供体形式对HDR的影响

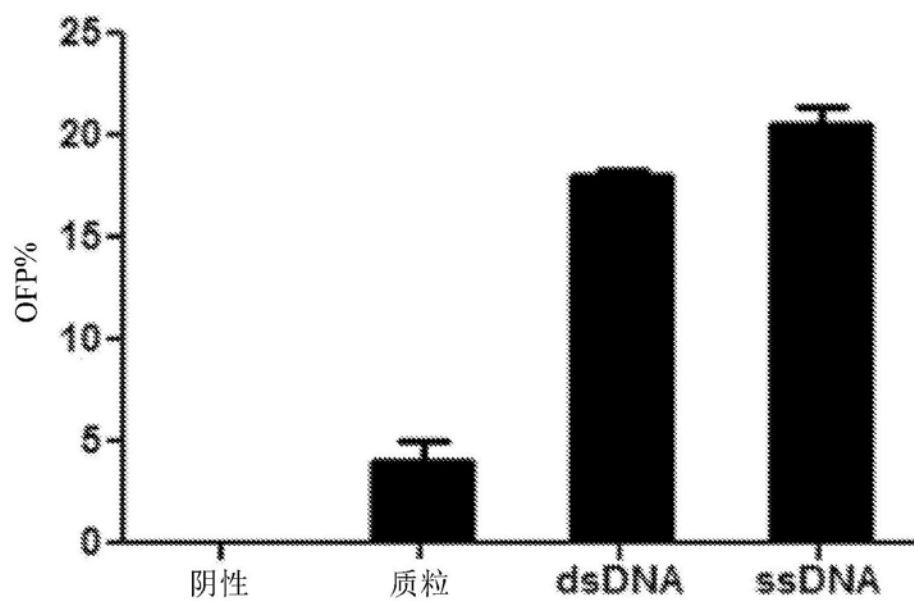


图2C

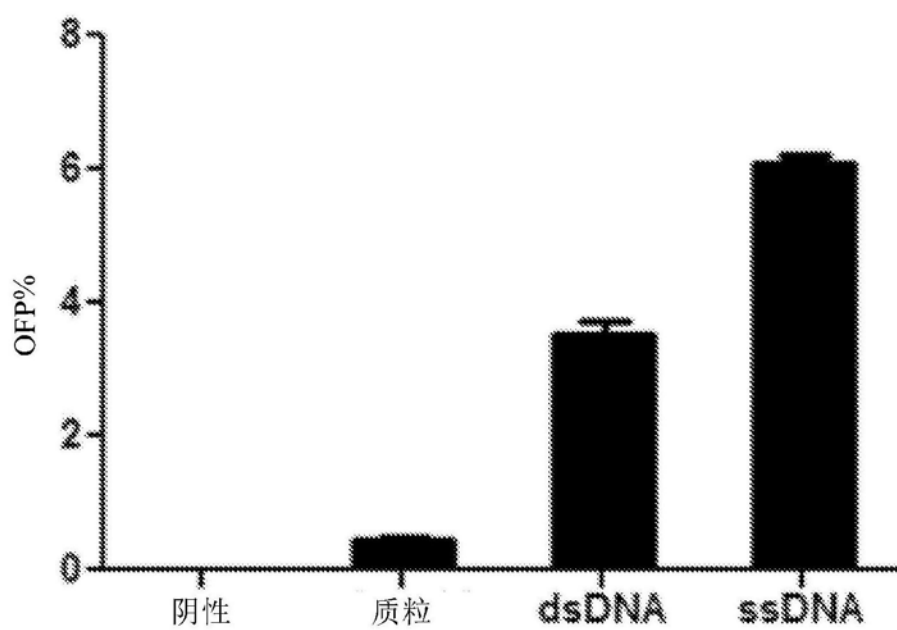


图2D

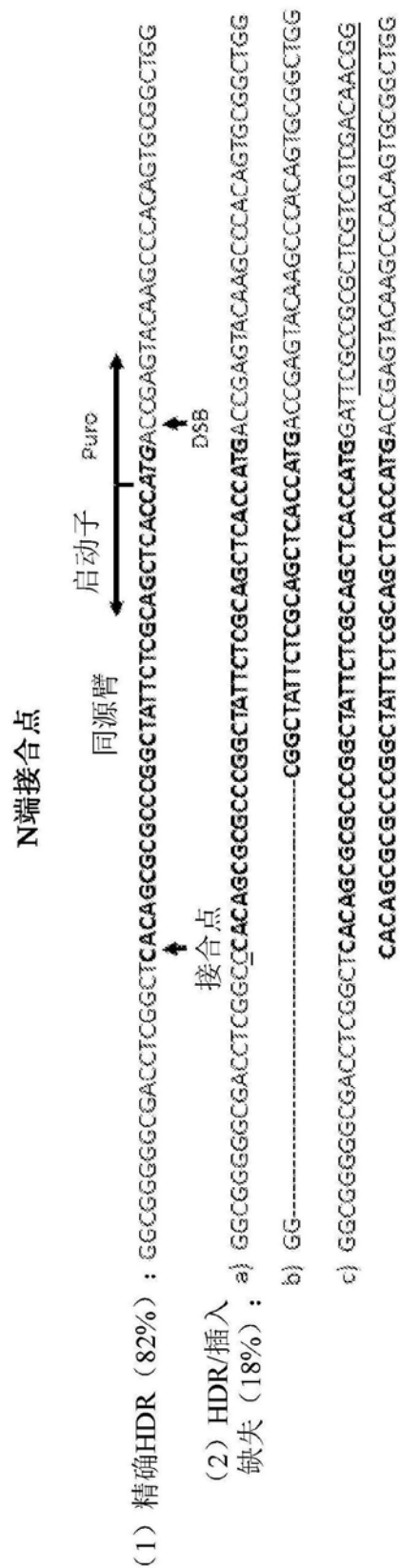


图3A

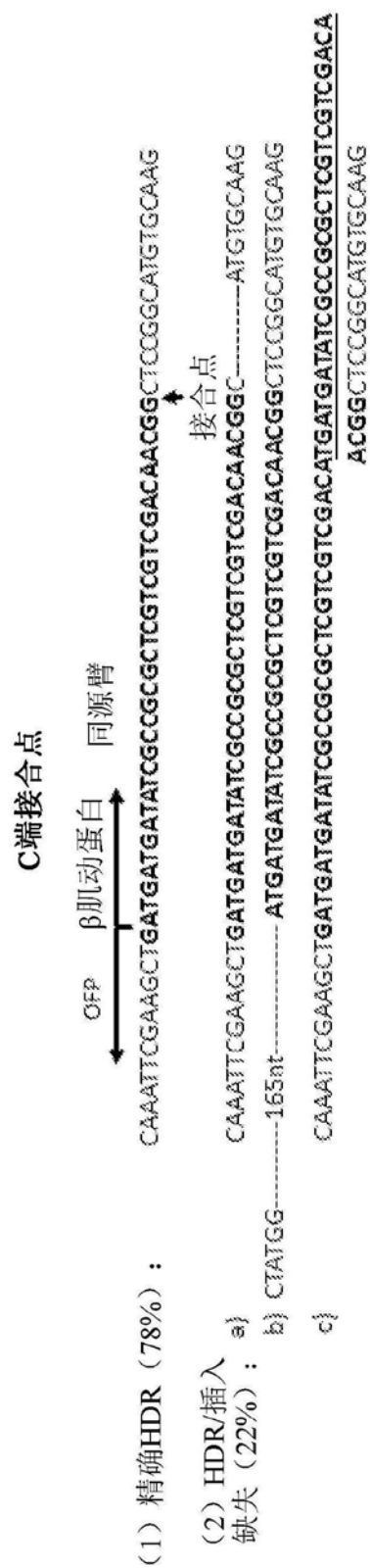


图3B

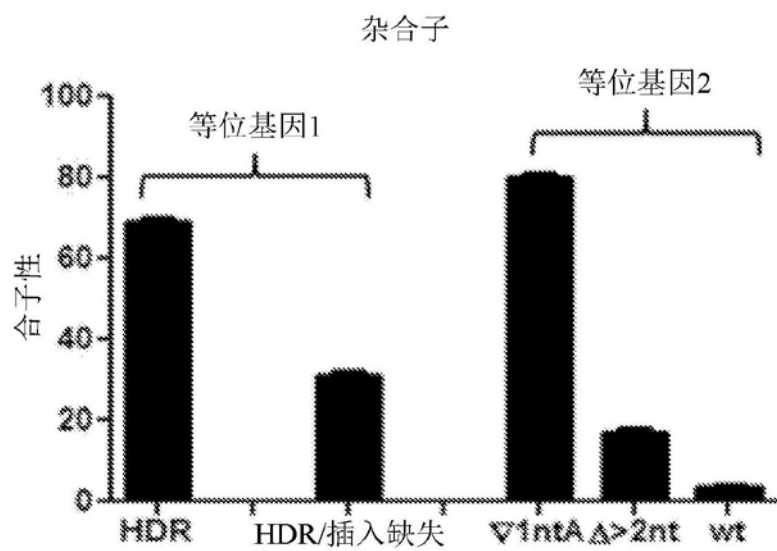


图3C

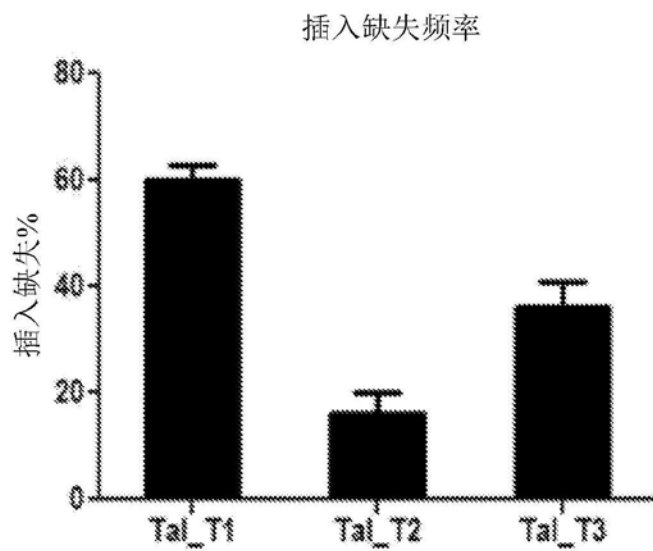


图3D

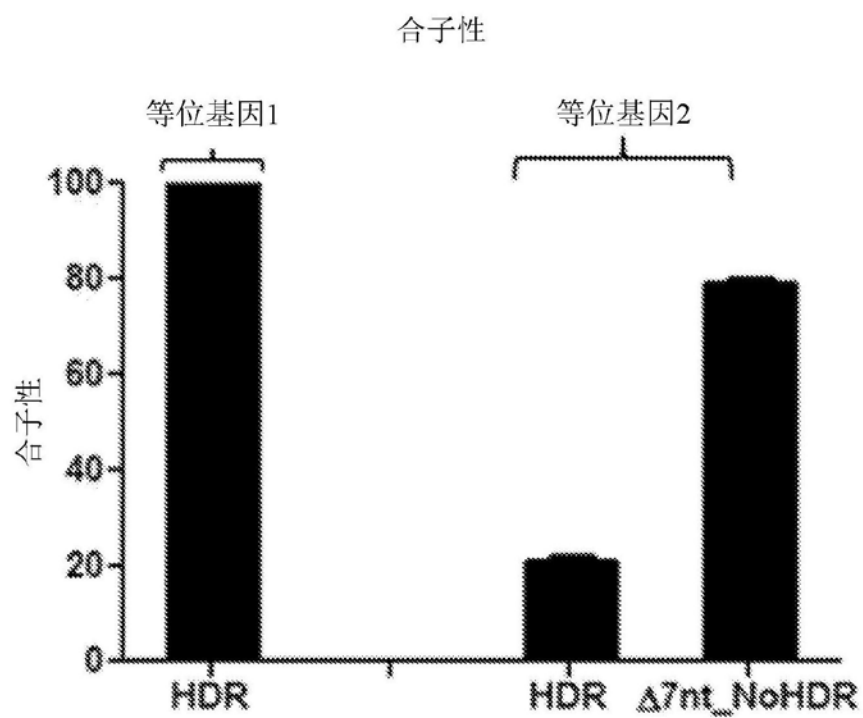


图4C

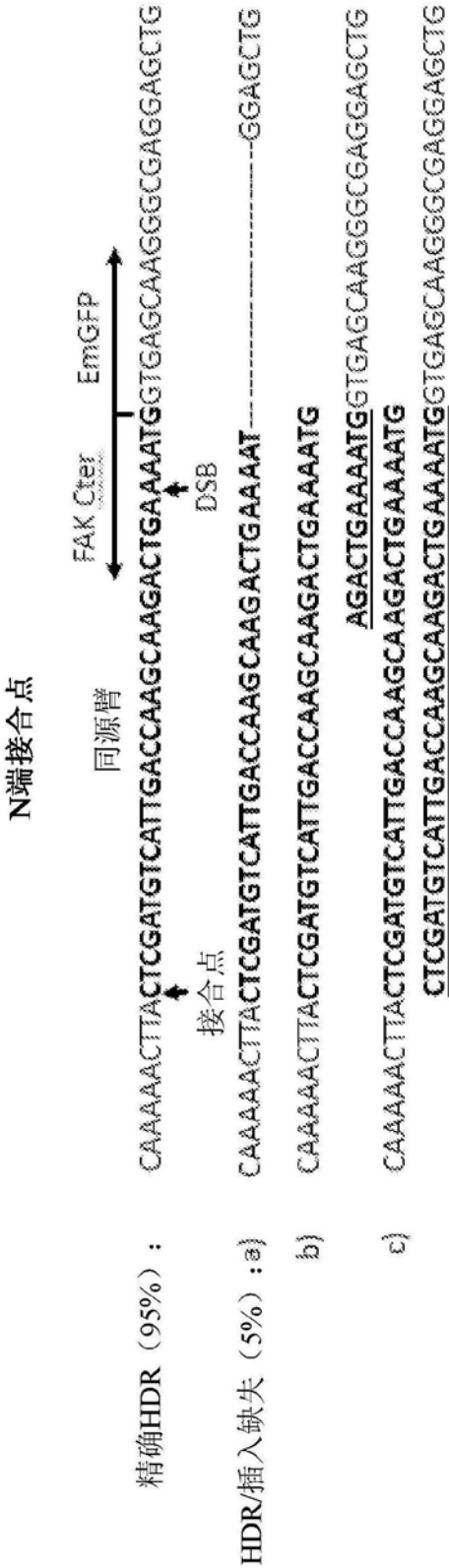


图5A

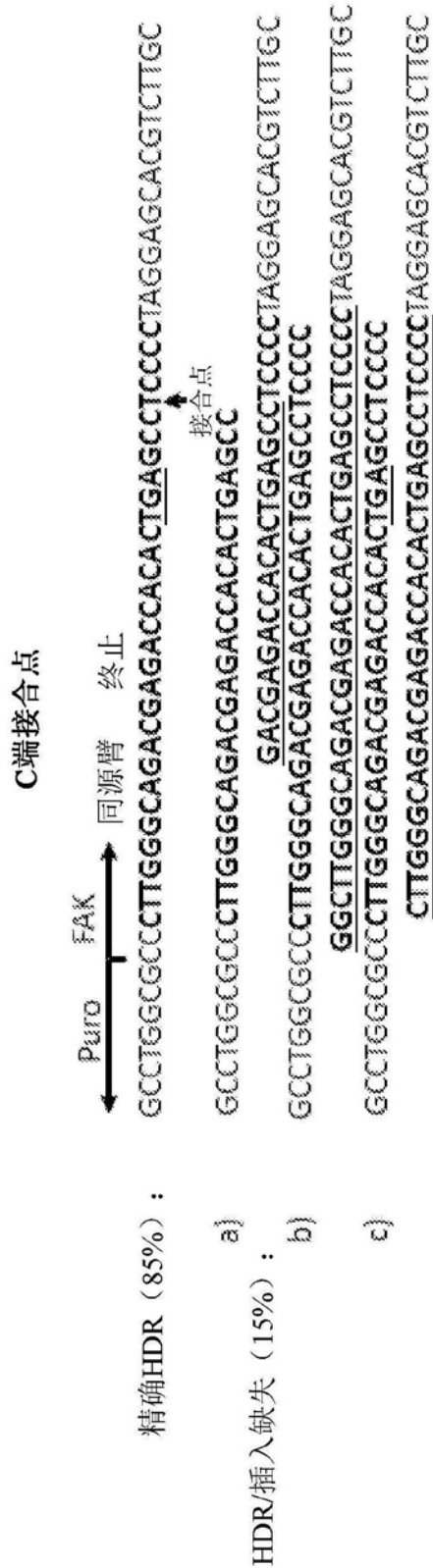


图5B

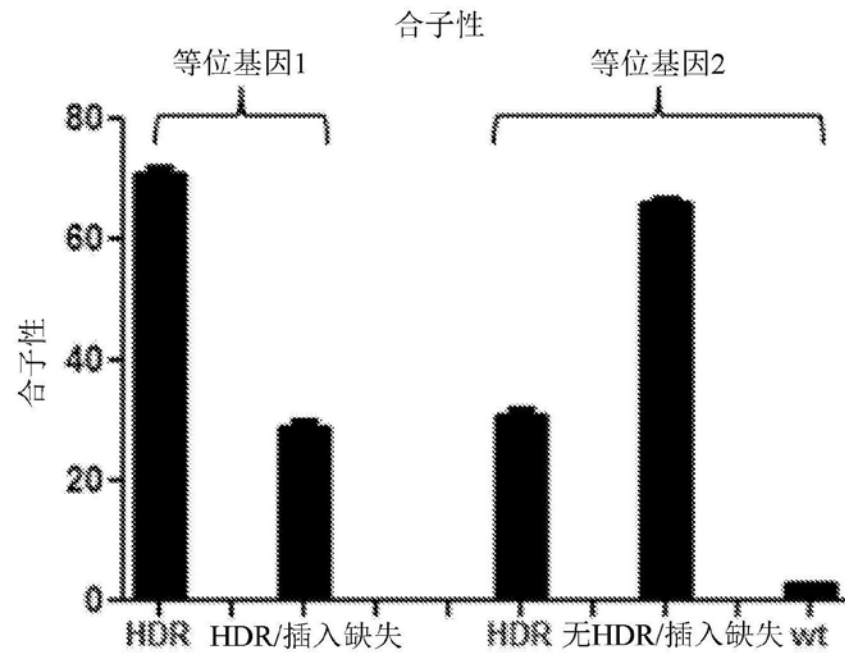


图5C

对克隆细胞的N端接合点分析（精确HDR：19/19）



图6A

对克隆细胞的C端接合点分析（精确HDR：19/19）



图6B

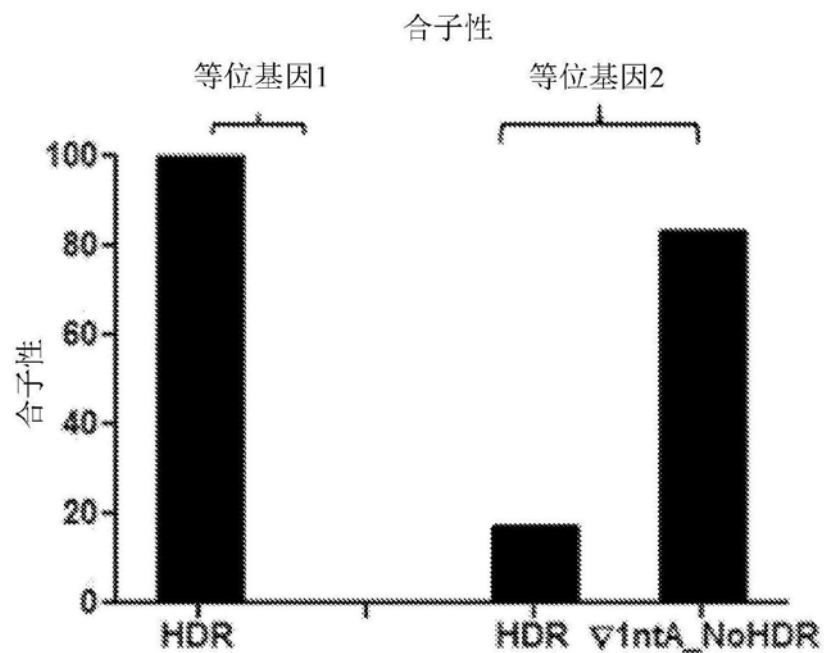


图6C

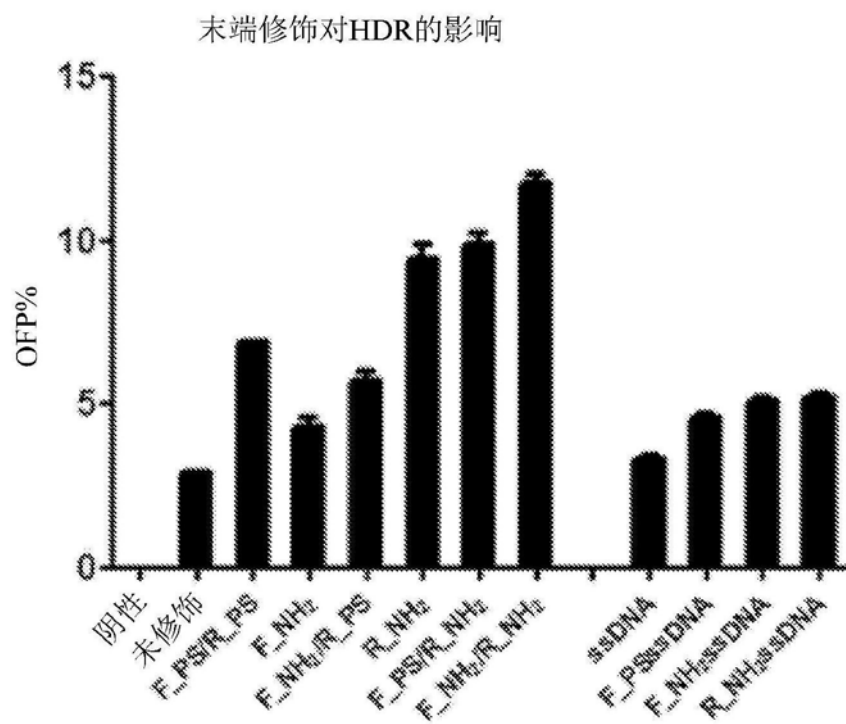


图7A

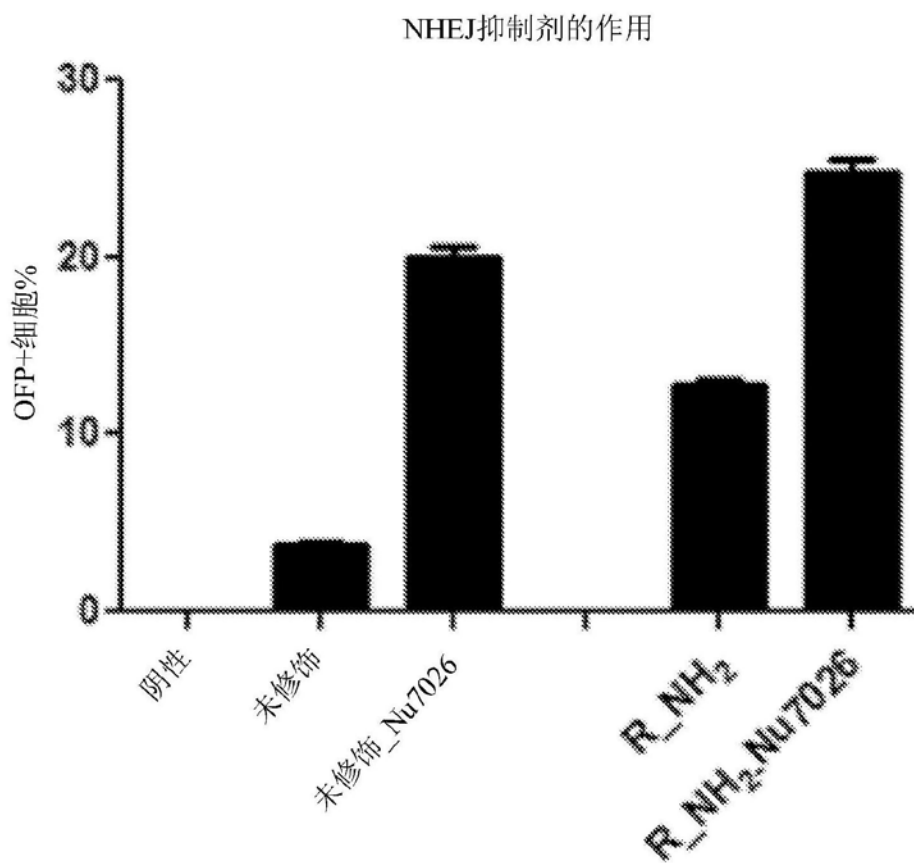


图7B

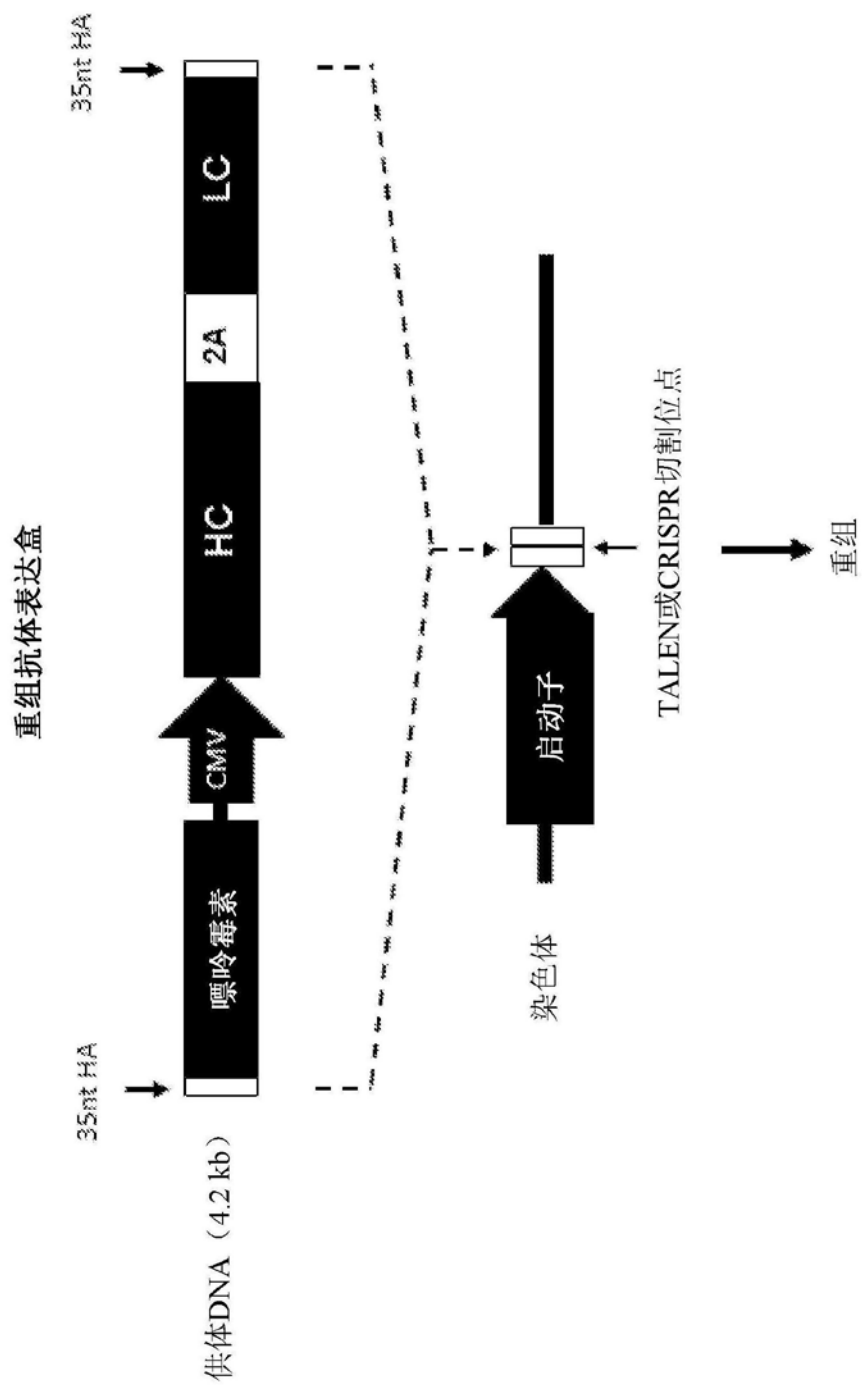


图8A



图8B

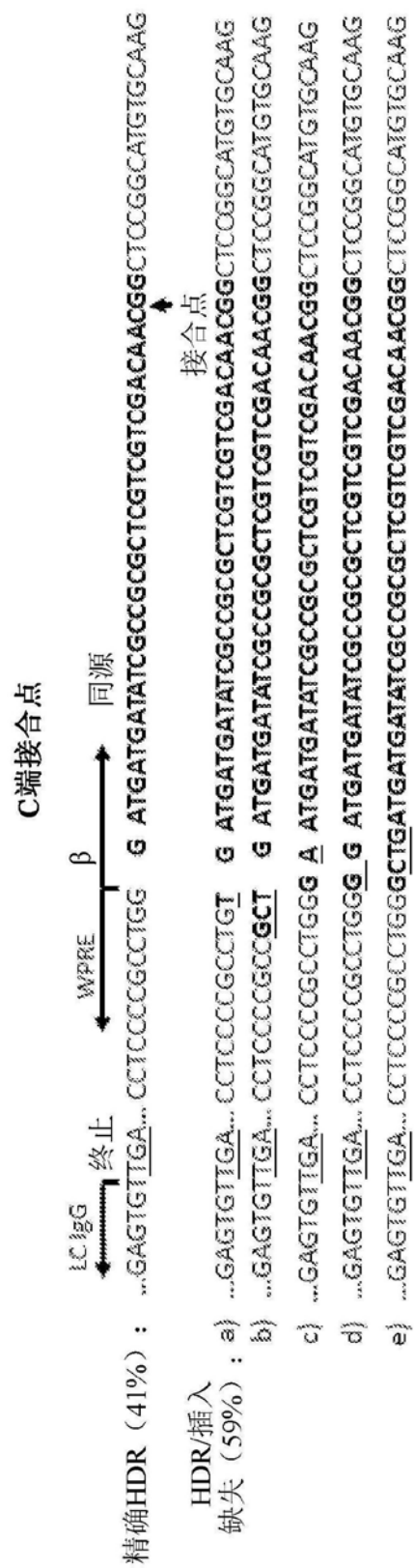


图8C

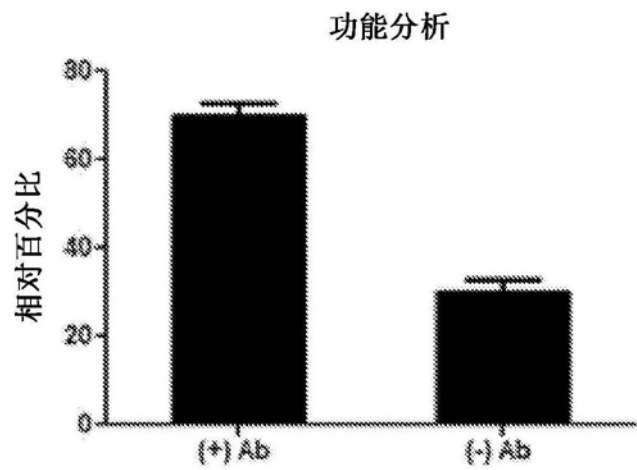


图8D



项目名称	描述
1 - BFP	肽的N端-KRTADGSEFESPKKKRKVEGG- 接头-5'- GETGCCCCGTGCCCTGGCCDACCCTCGTGACCACCCCTGACCTACGGCGTGCAAGTGCTTCAGCCGCTACCCCGACACATGA 3' "Cys-SMCC-C6" 接头: 0.5-1 OD (19-38 μg), 具有>85% HPLC纯度
2-6碱基	肽的N端-KRTADGSEFESPKKKRKVEGG- 接头-5'- CGGGGTAGCGGGCGAAGCACTGCACGCCGTAGGTGAAGGTGSTCACGAGGGTGGGCCAGGGSCACGGGCAGCTTGCCGGTG GTGCAGATGAACCTTCAG 3' "Click化学接头": 0.5-1 OD (20-38 μg), 具有>85% HPLC纯度

图9

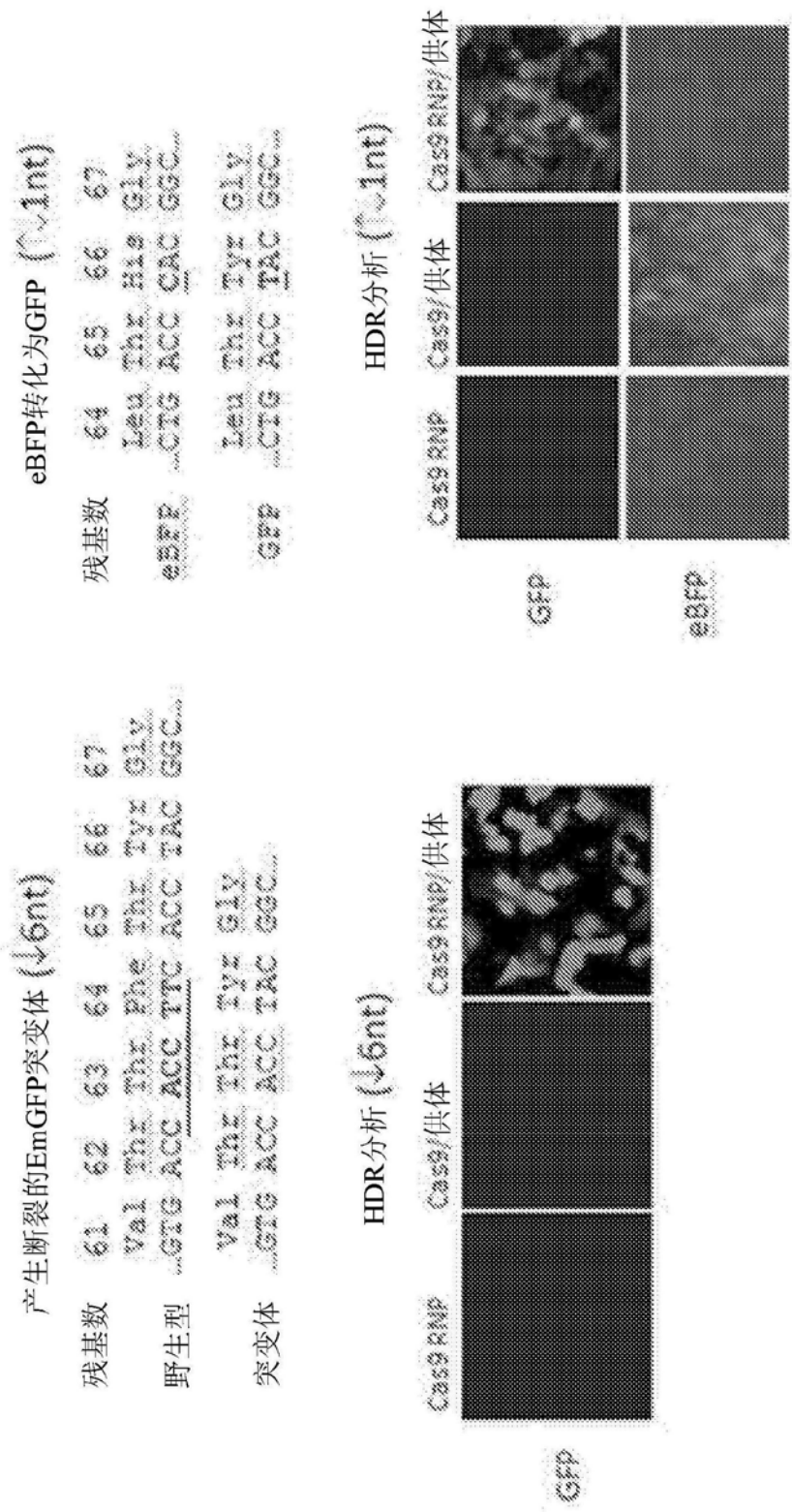


图10

		经硫代磷酸酯修饰的供体（皮摩尔/反应物）					NLS修饰的供体（皮摩尔/反应物）				
		0.03	0.1	0.3	1	3	10	0.01	0.03	0.1	0.3
平均值	0	3.24	9.58	18.81	31.50	35.80	36.00	29.10	46.10	51.90	56.25
标准差	0	0.37	0.59	0.44	0.71	1.13	1.13	2.26	1.56	0.14	1.20

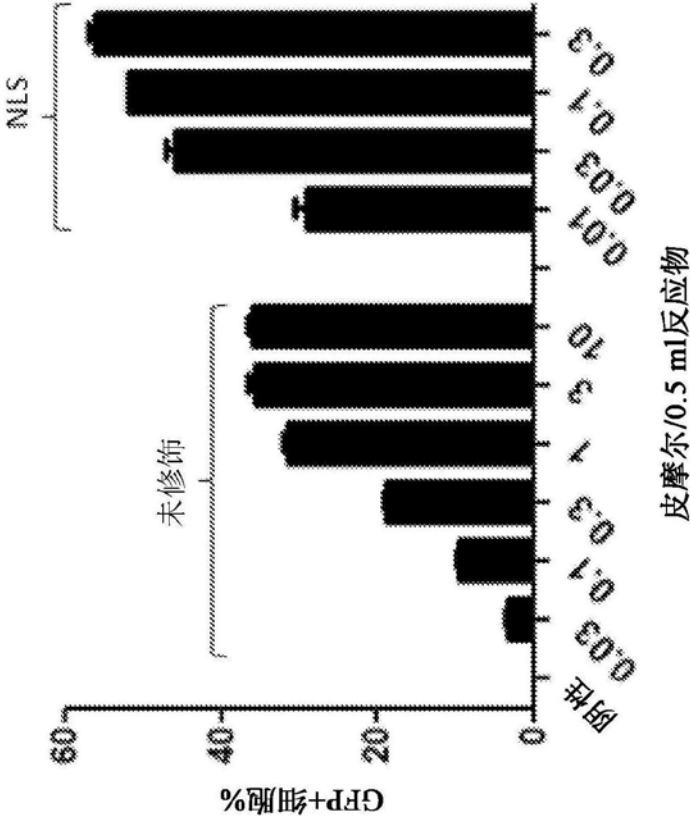


图11

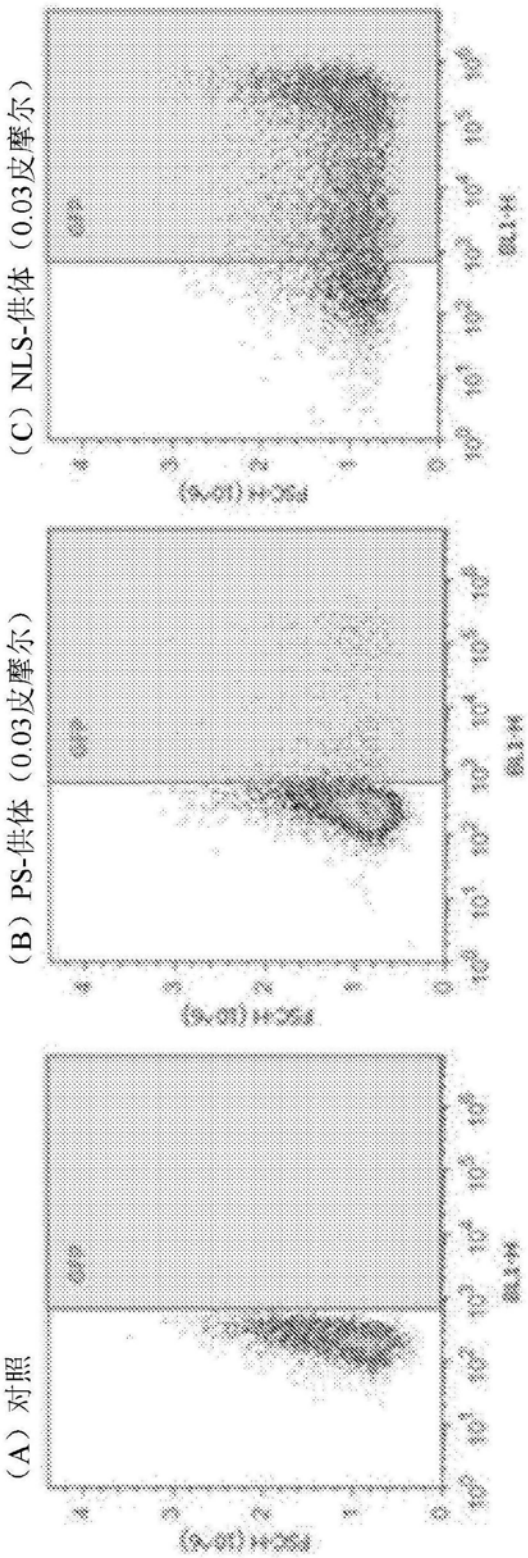


图12

未修饰的供体 (皮摩尔/反应物)				NLS修饰的供体 (皮摩尔/反应物)			
阴性	0.03	0.1	0.3	1	3	10	0.03
平均值	0	6.15	8.3	15.1	33	52.05	21.15
标准差	0	0.21	0.28	0.57	2.83	0.78	0.21
							0.03
							33.1
							43.8
							1.20
							75.85
							0.78

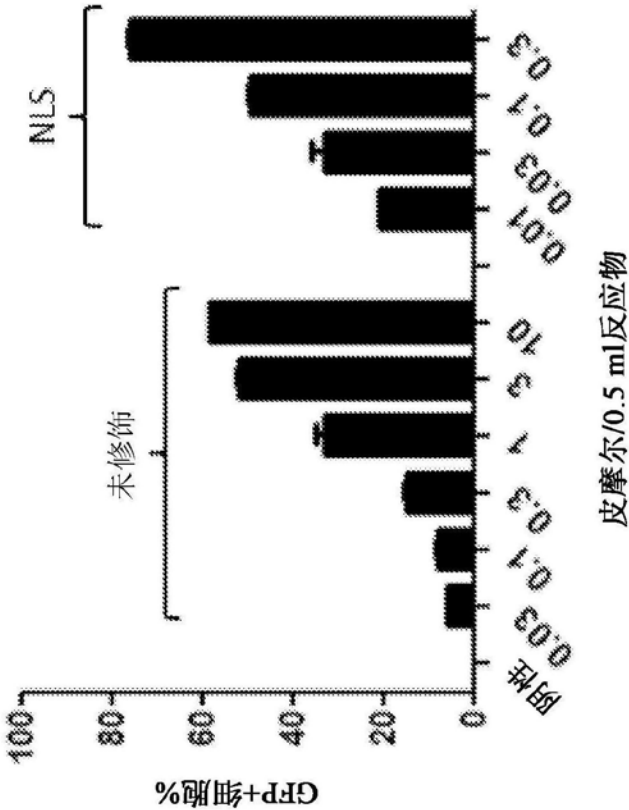


图13

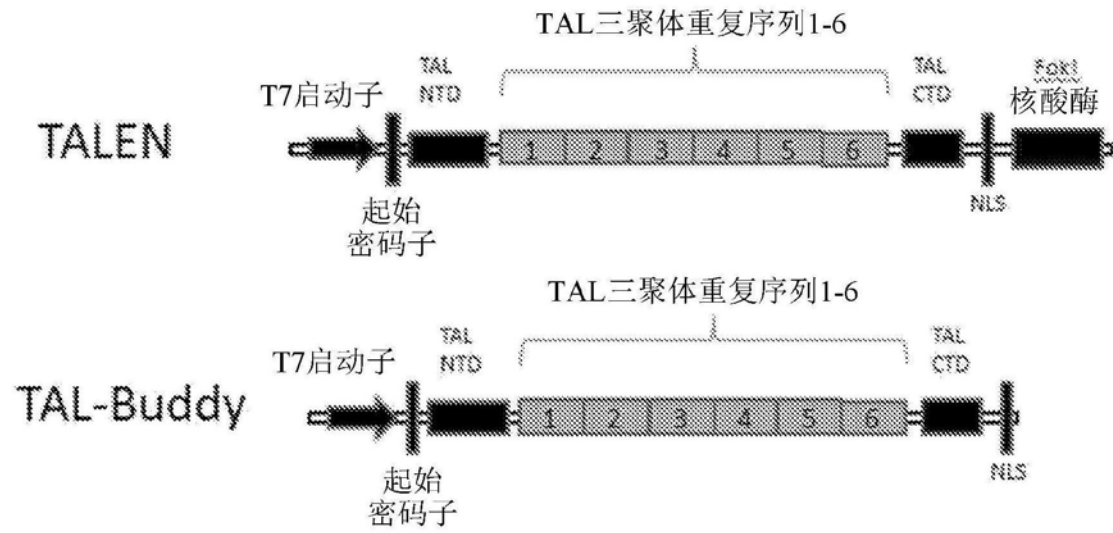


图14

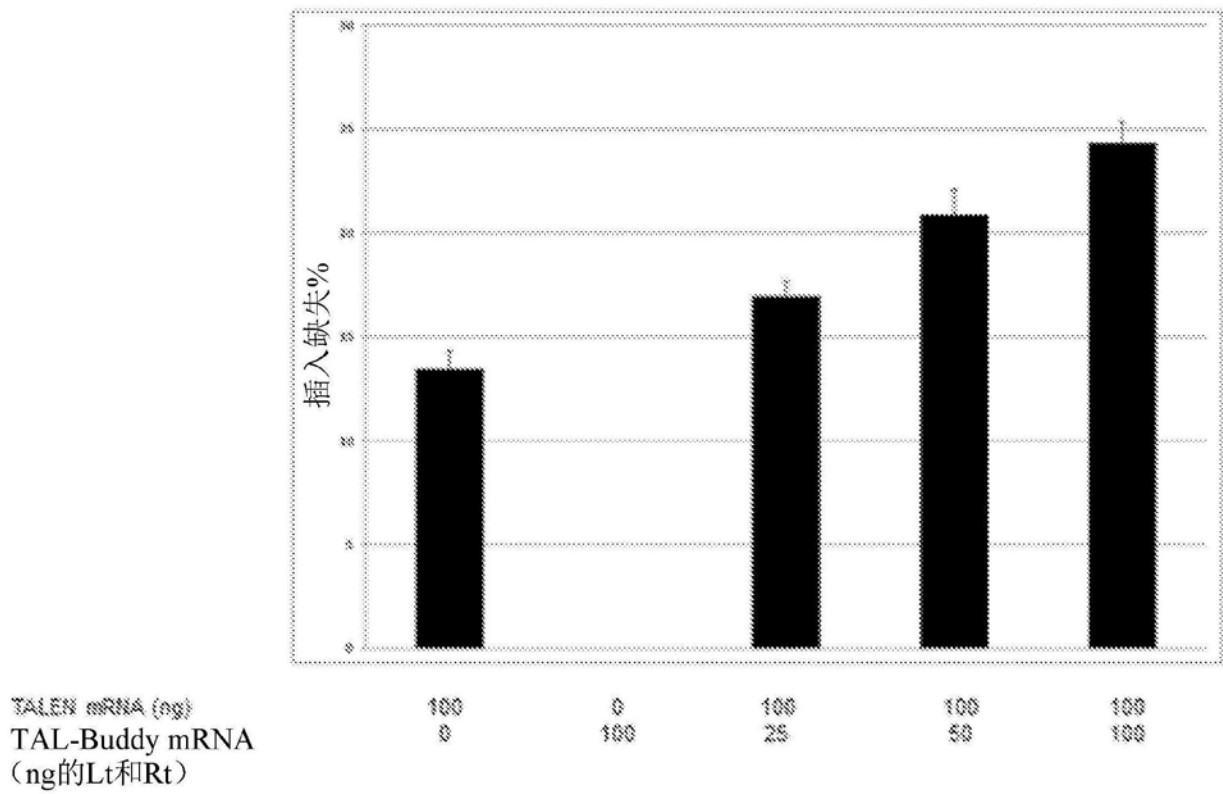


图15

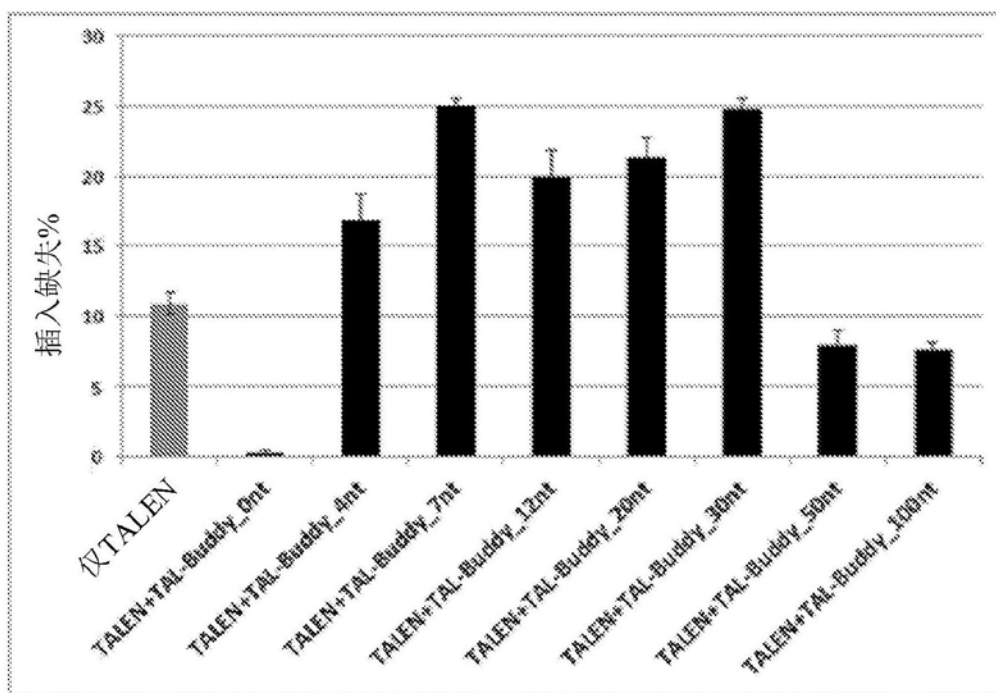


图16

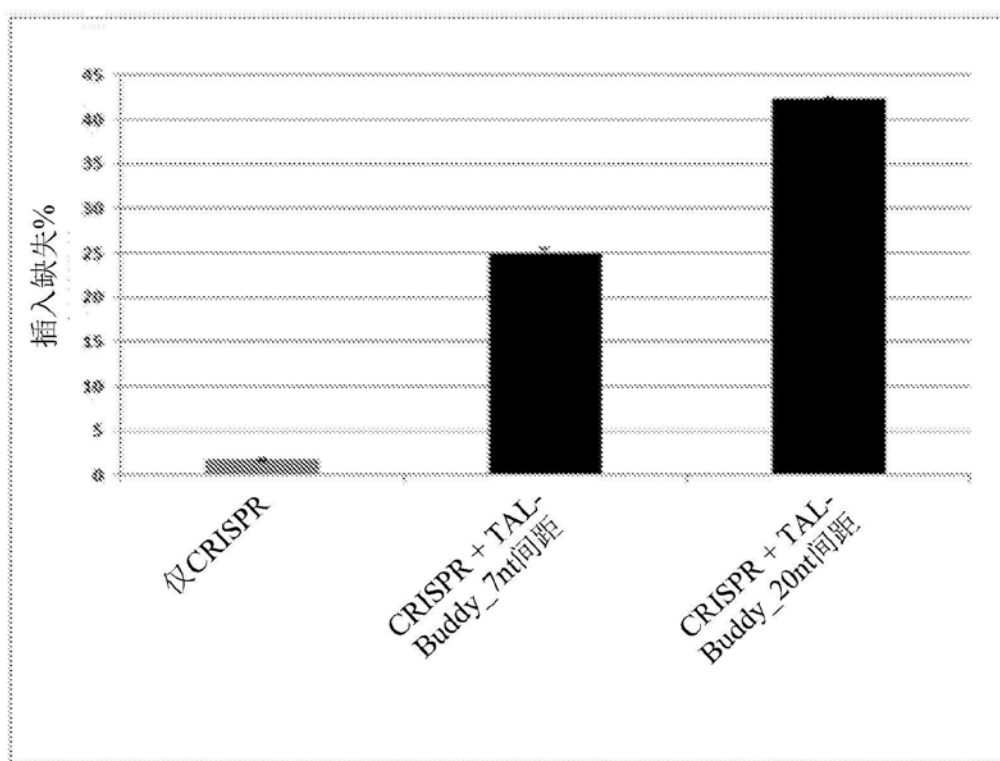


图17

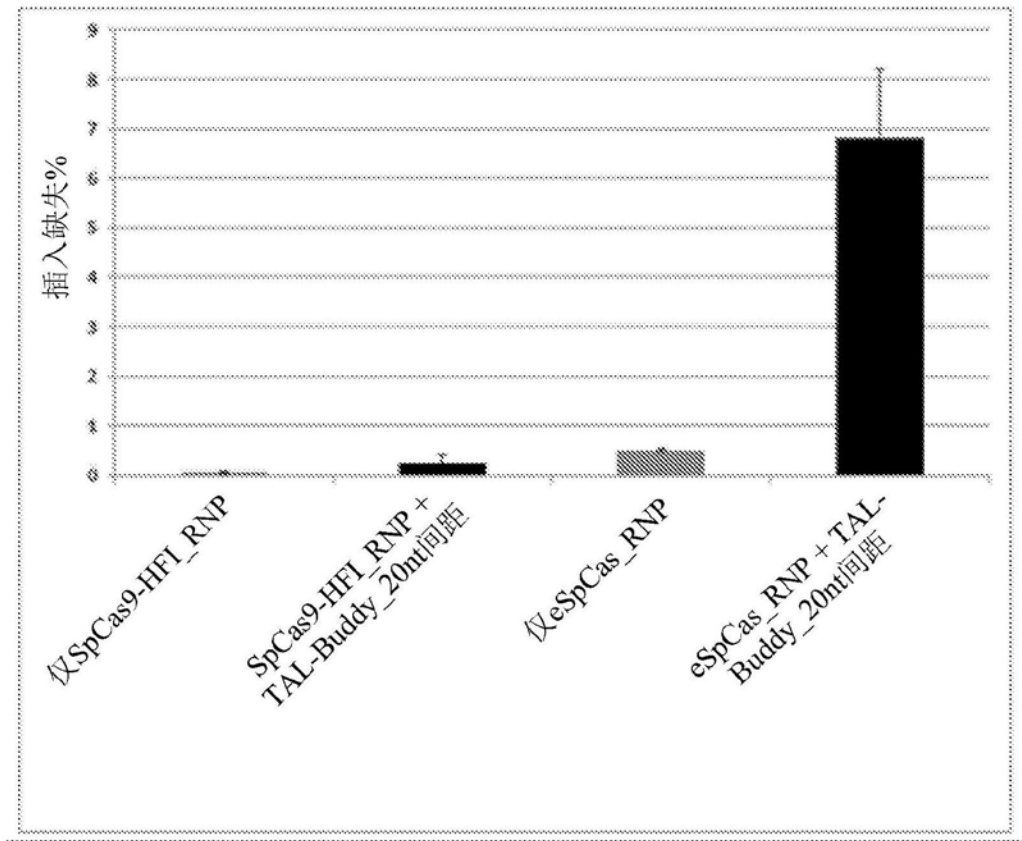


图18

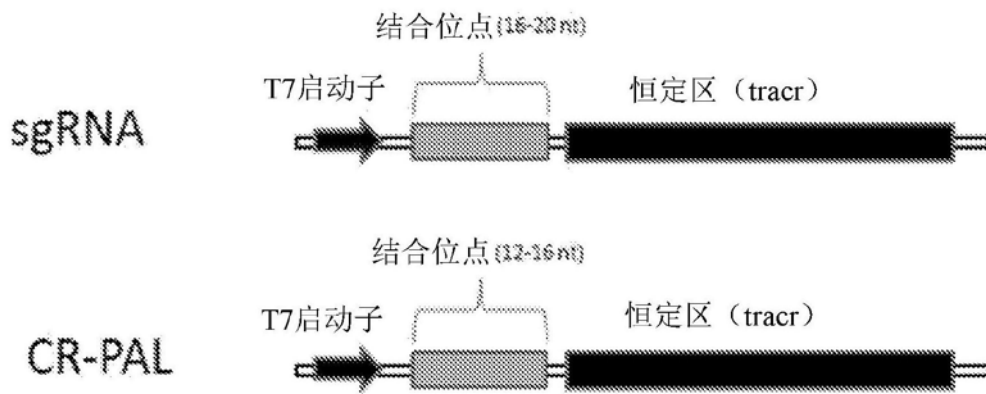


图19

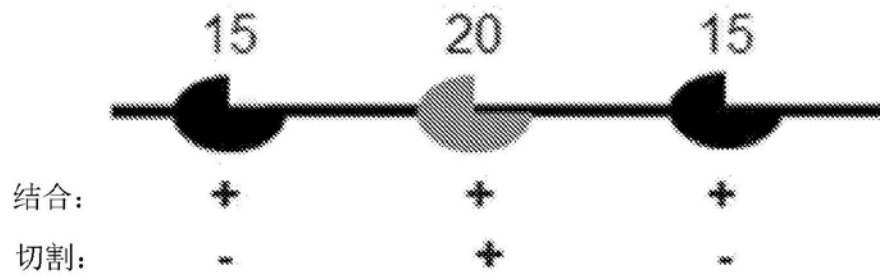


图20

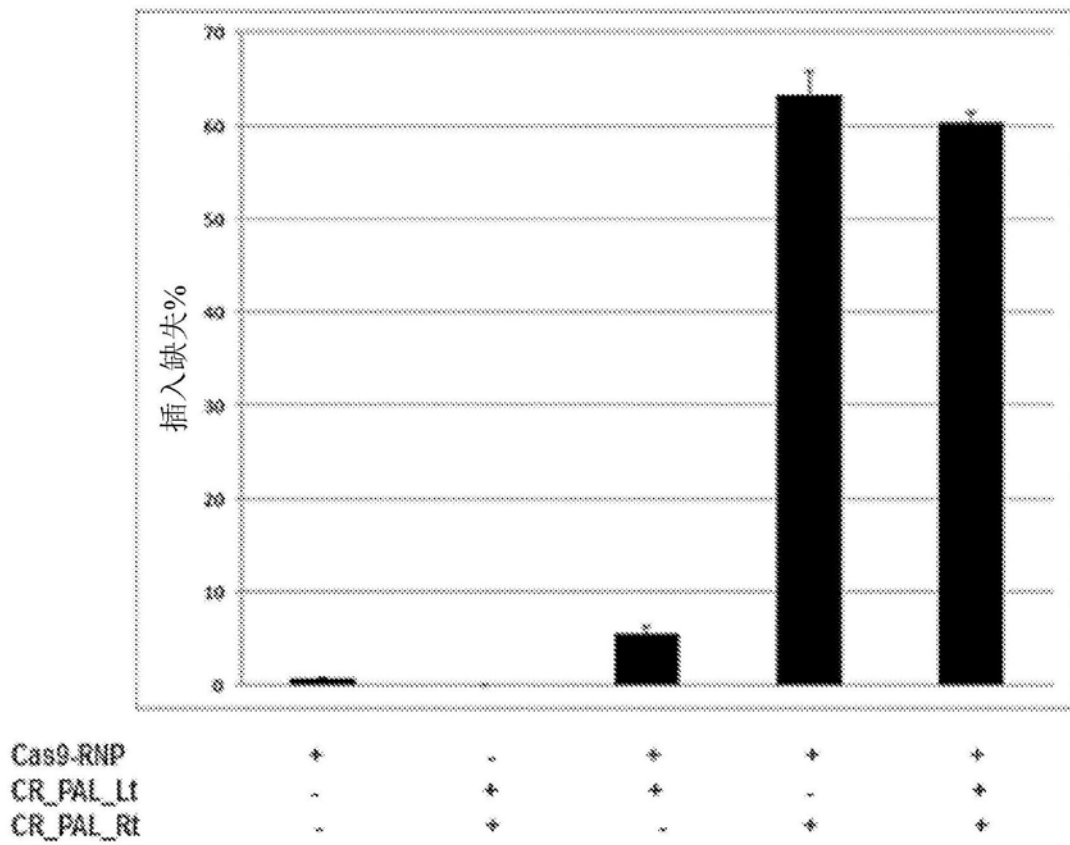


图21

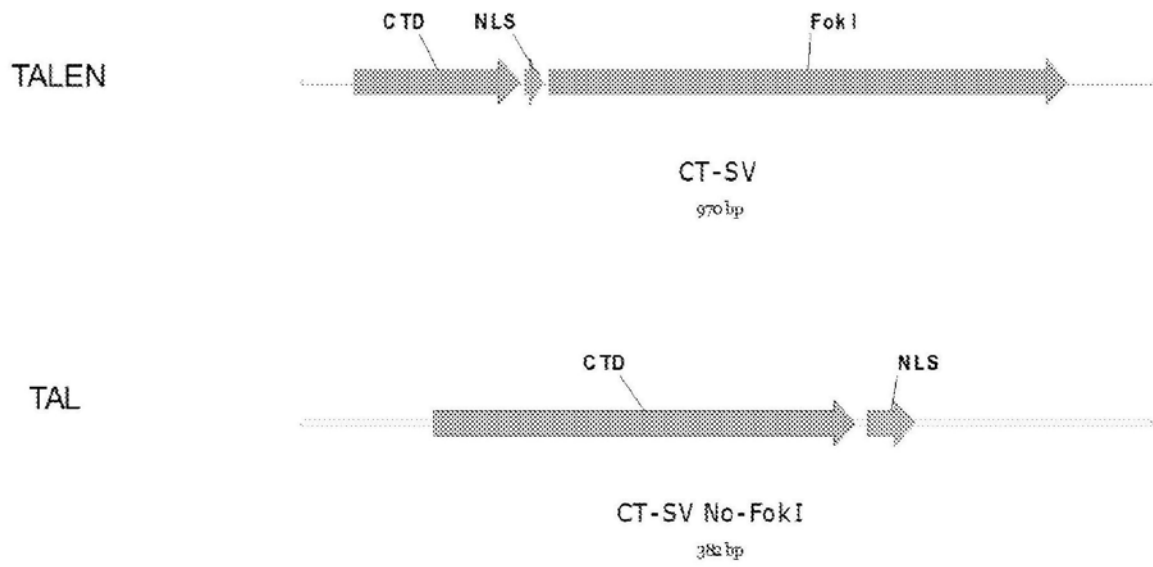


图22

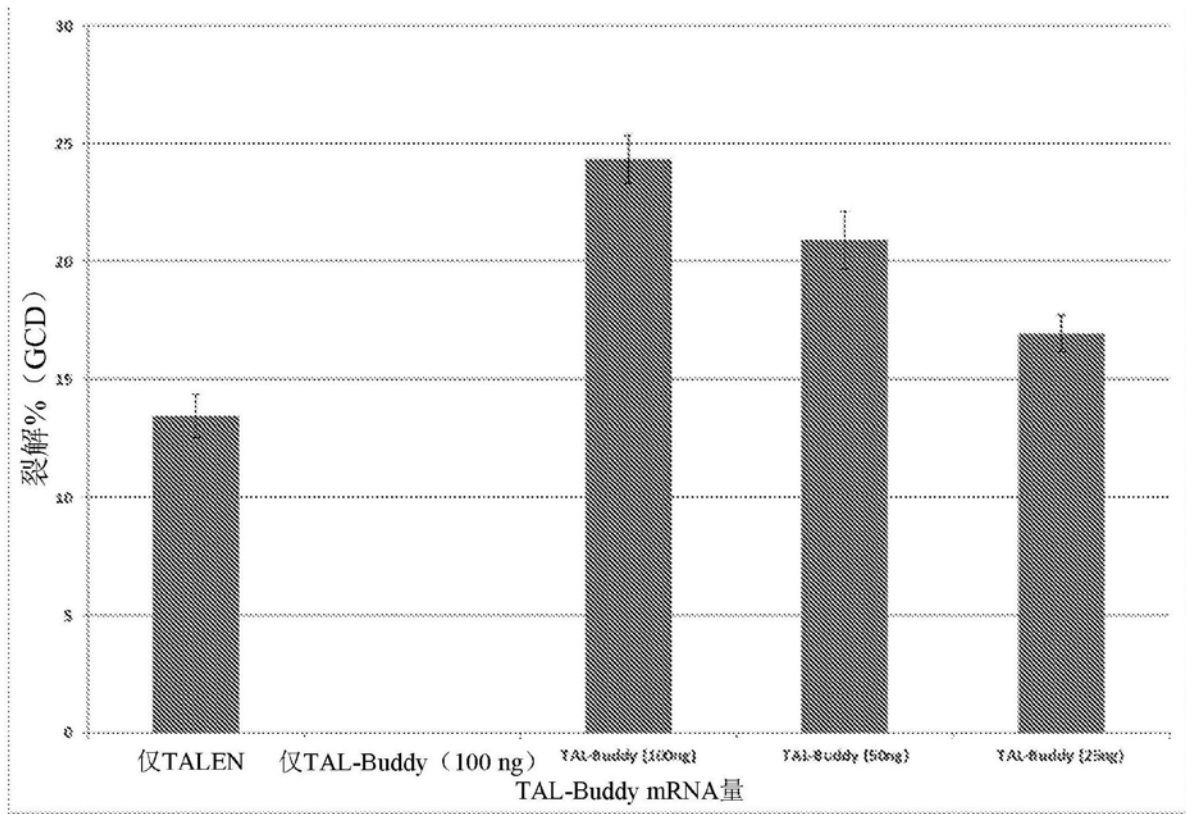


图23

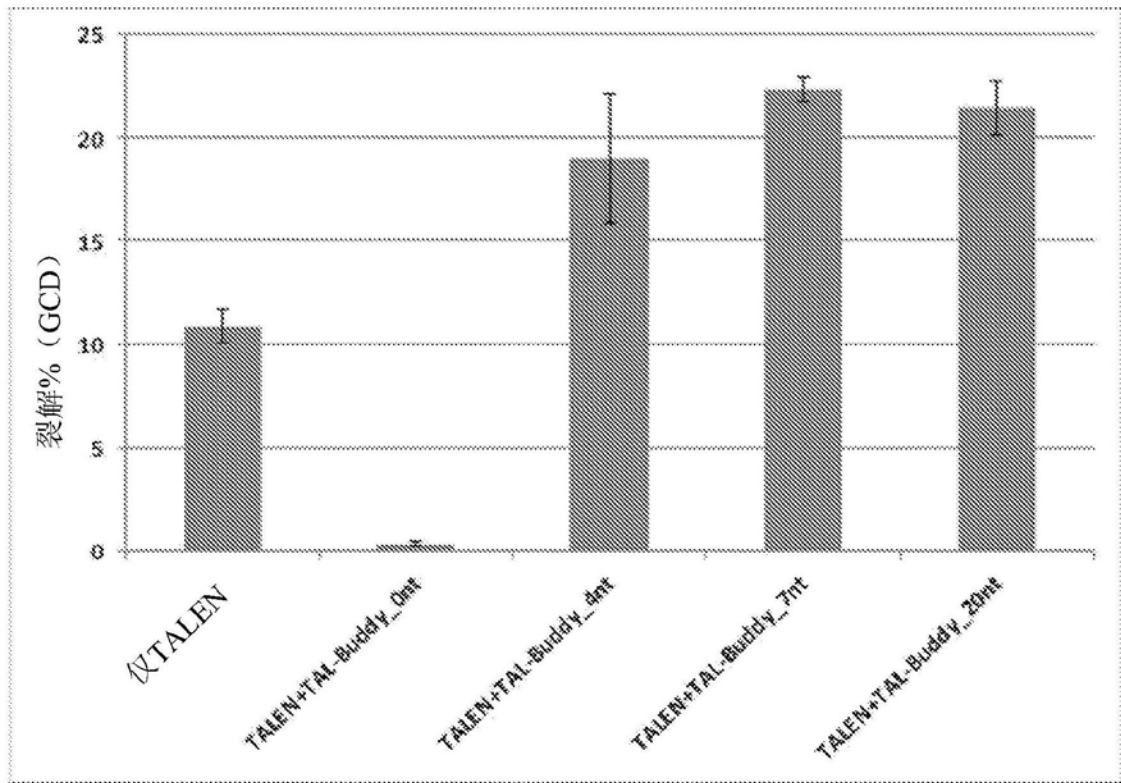


图24

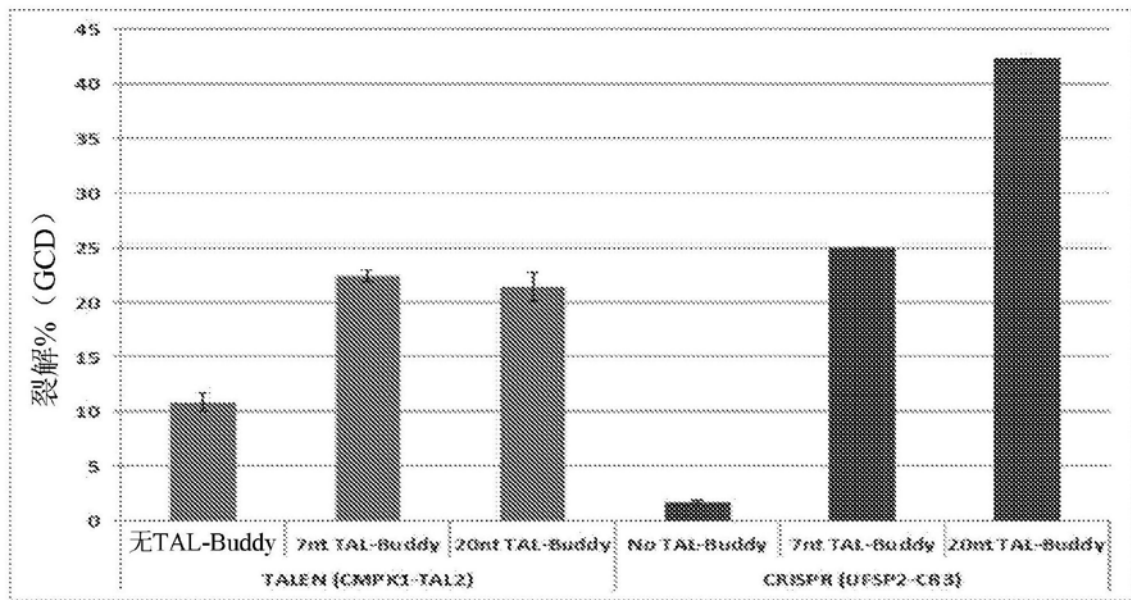


图25

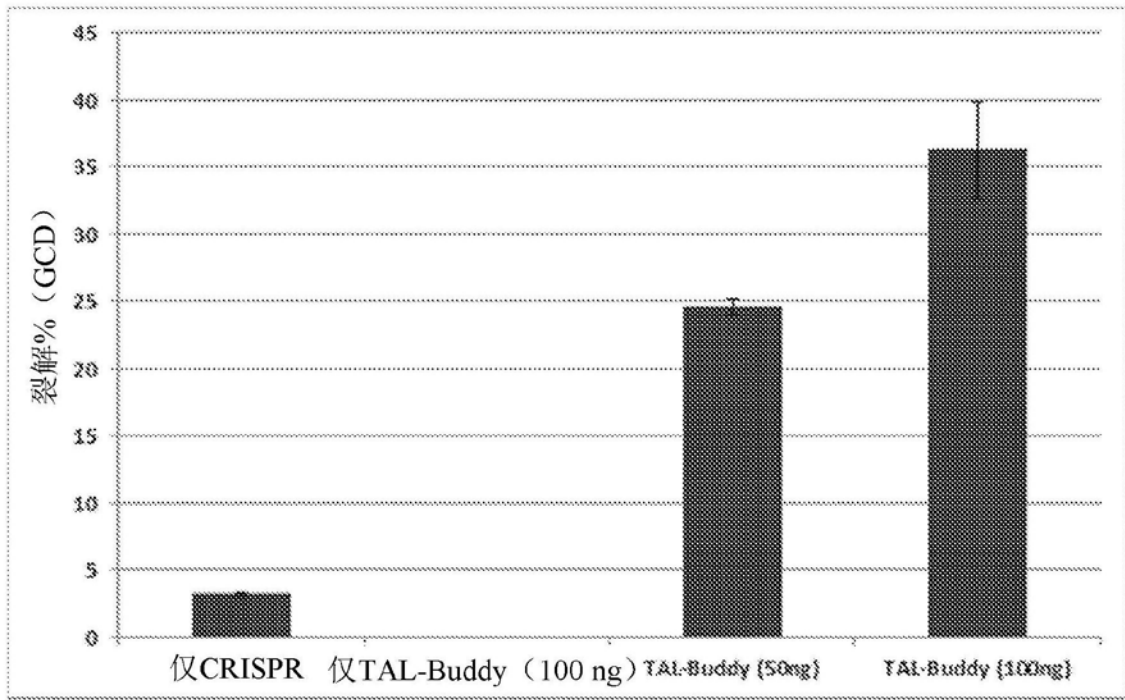


图26

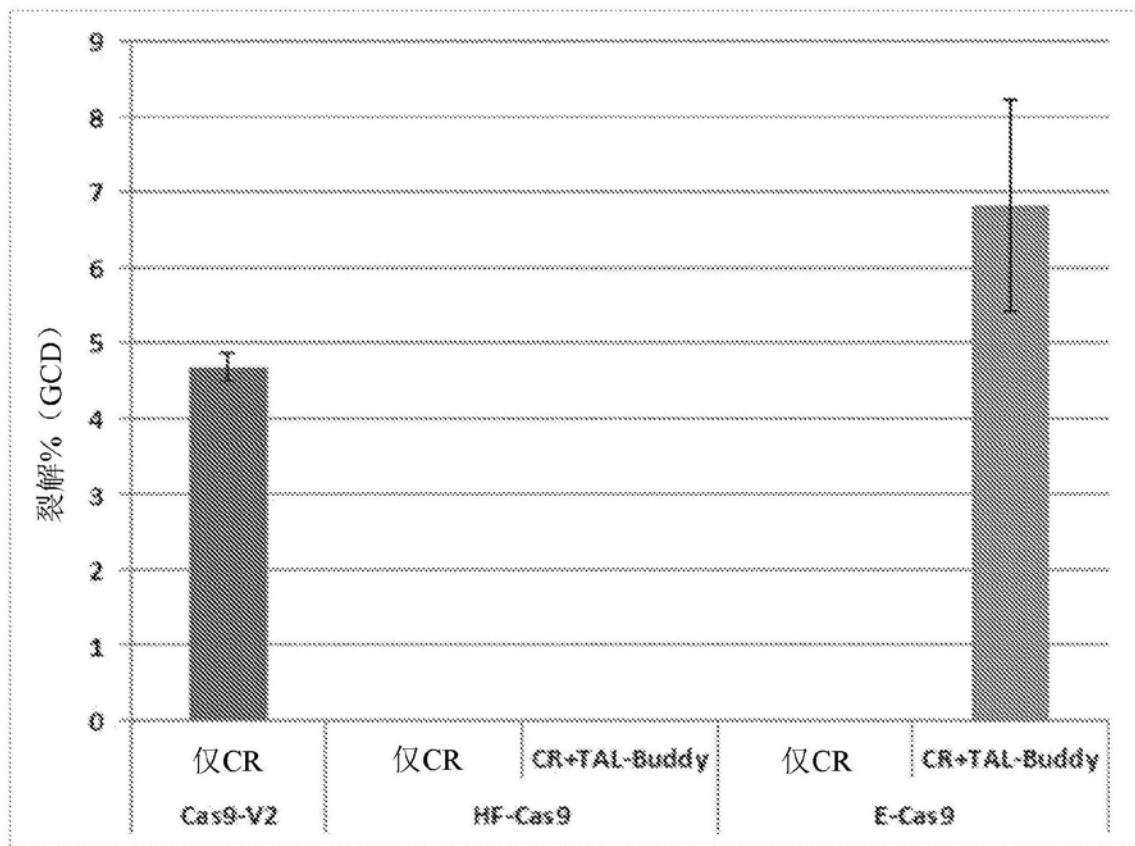


图27

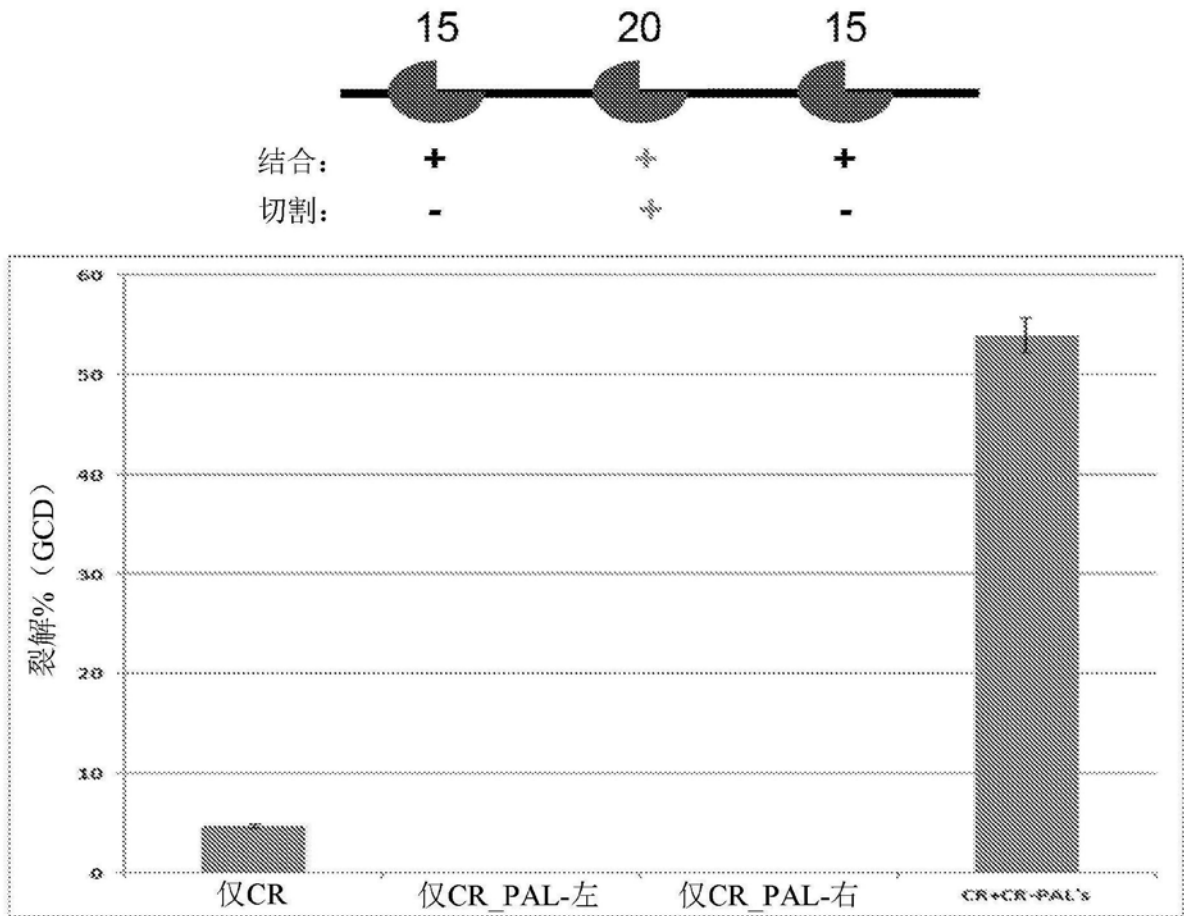


图28

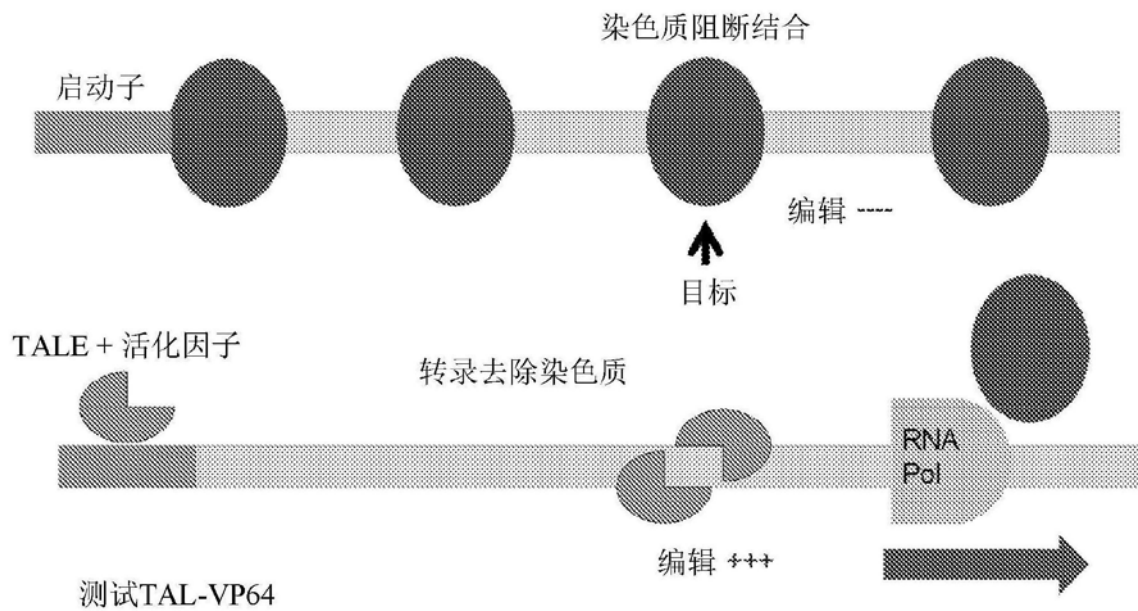


图29

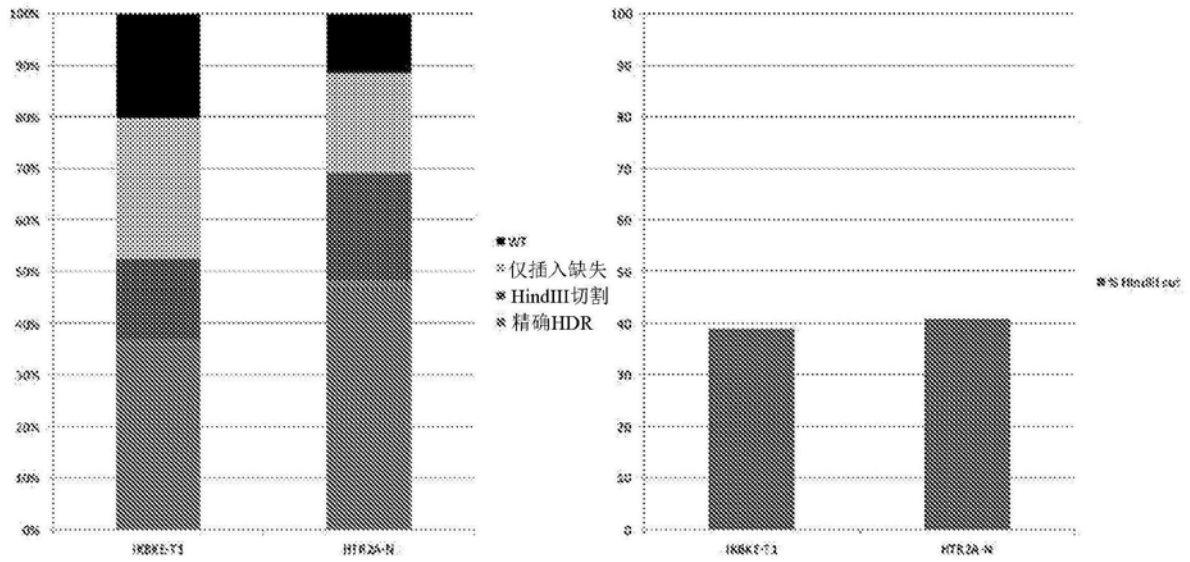


图30

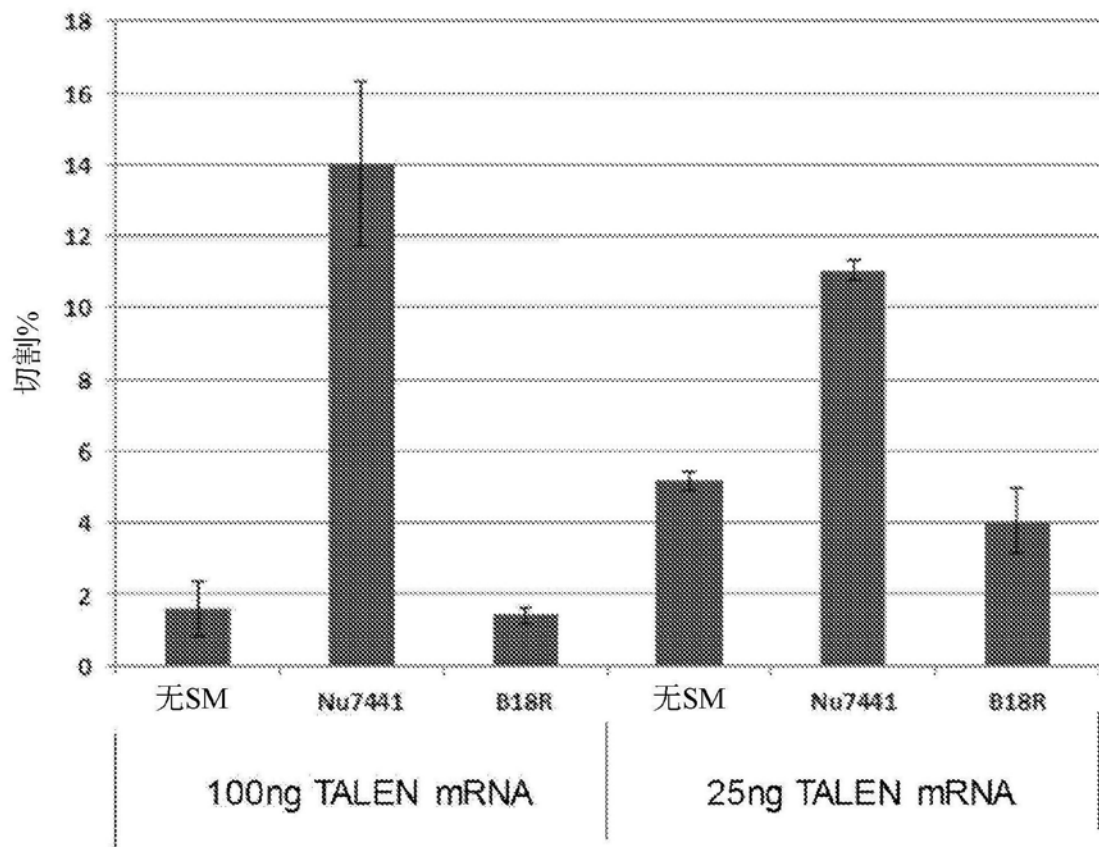


图31



图32



图33

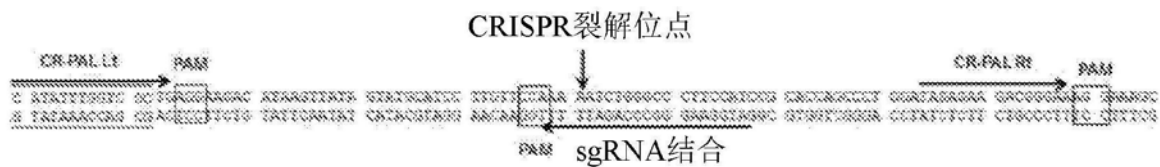


图34

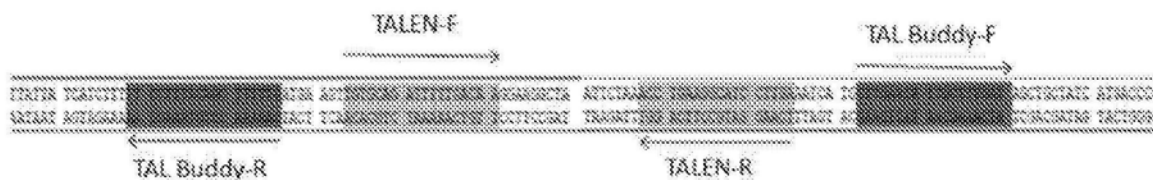


图35

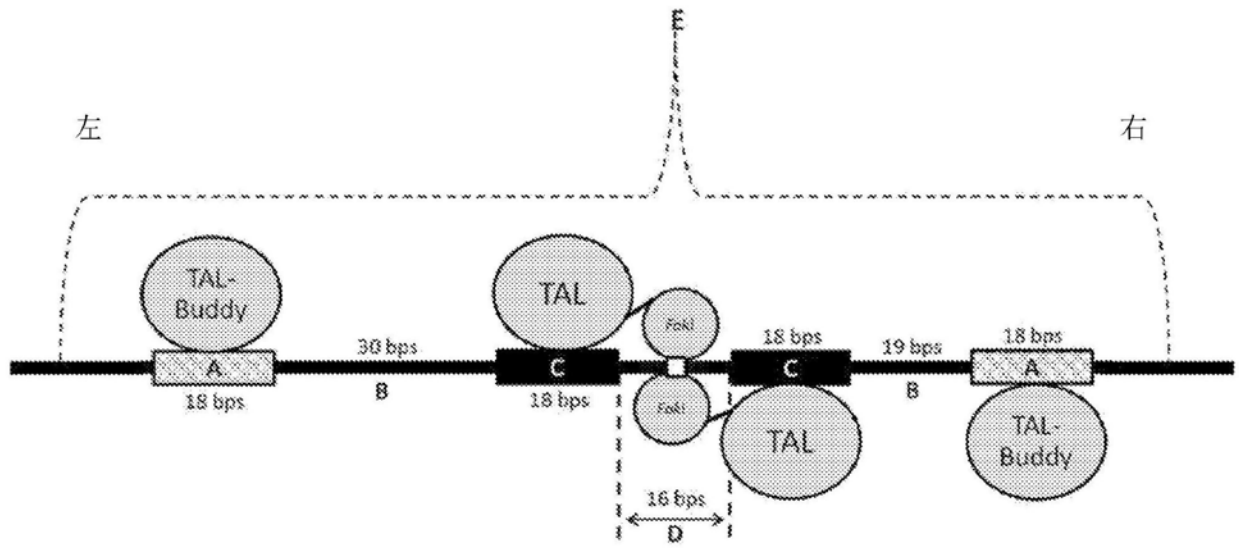


图36

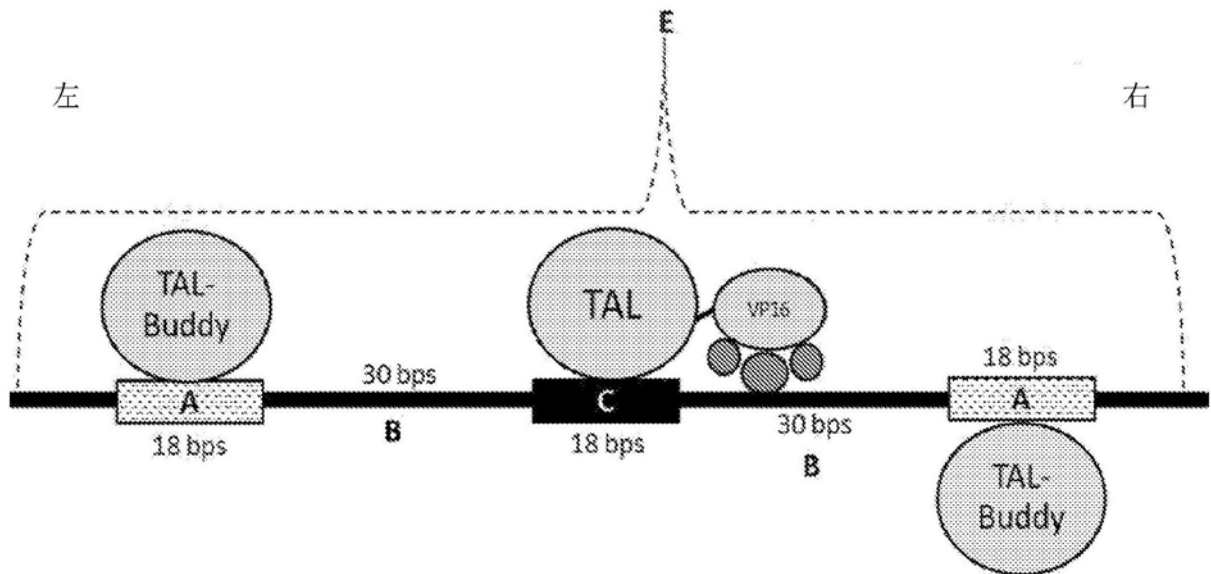


图37

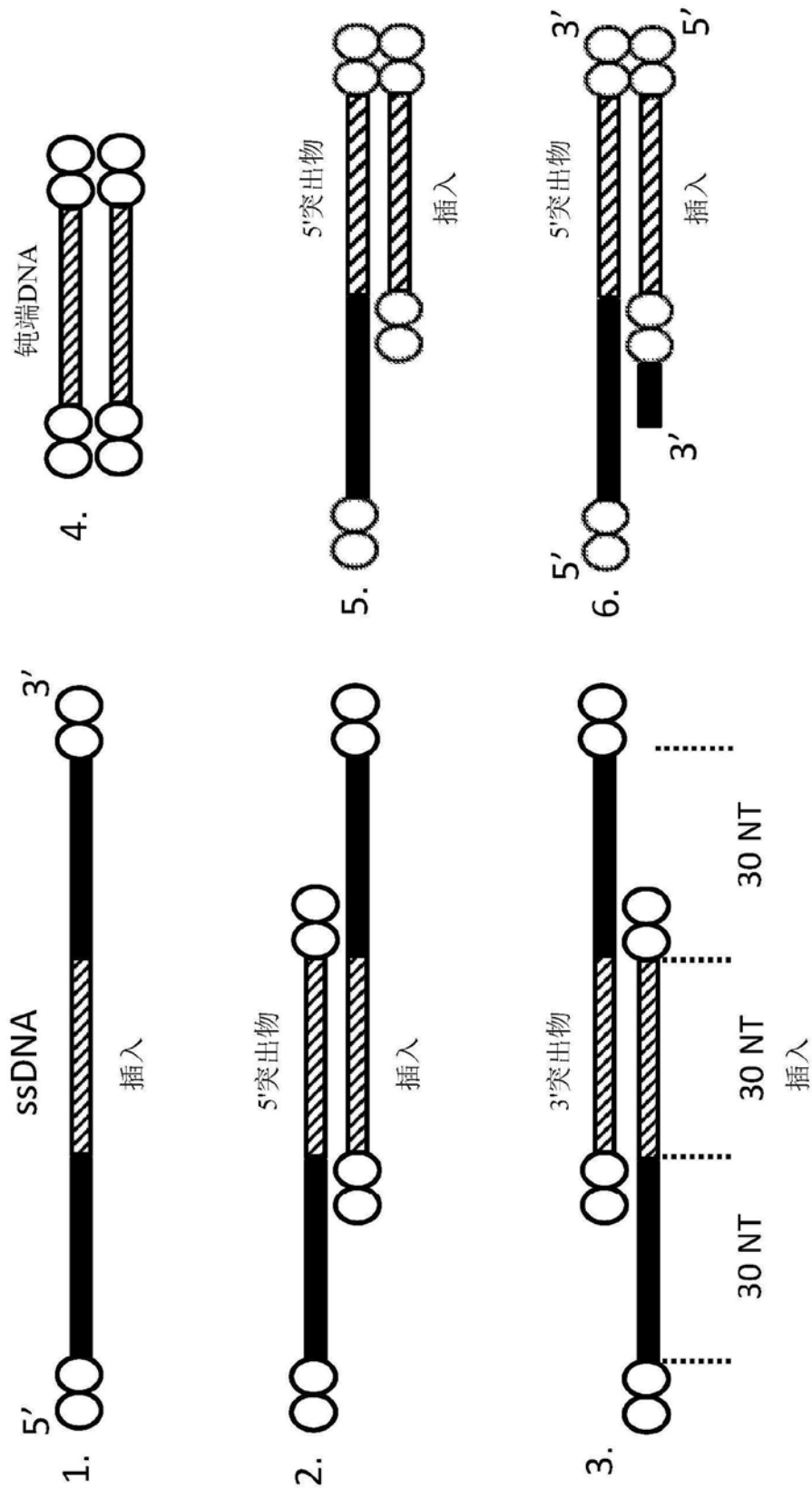


图38

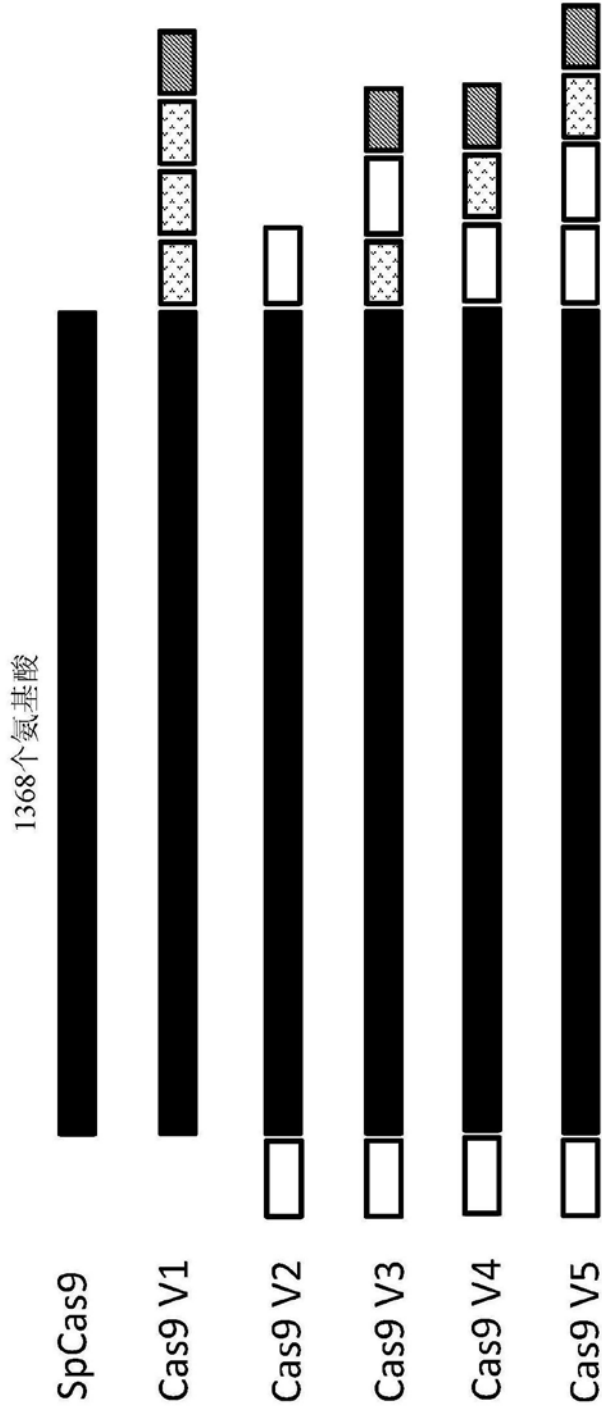


图39

MDKKYSIGLD IGTSNVGWAV ITDEYKVPSK KFKVLGNTDR HSIKKNLIGA LLFDSGETAE ATRLKRTARR
 RYTRRKNRIC YLQEIFSNEM AKVDDSFHR LEESFLVEED KKHERHPIFG NIVDEVAYHE KYPTIYHLRK
 KLVDSTDKAD LRLIYLALAH MIKFRGHFLI EGDLPDNDSD VDKLFIQLVQ TYNQLFEENP INASGVDAKA
 ILSARLSKSR RLENLIAQLP GEKKNGLFGN LIALSLGLTP NFKSNFDLAE DAKLQLSKDT YDDDLNLLA
 QIGDQYADLF LAAKNLSDAI LLSDILRVNT EITKAPLSAS MIKRYDEHHQ DLTLKALVR QQLPKEYKEI
 FFDQSKNGYA GYIDGGASQE EFYKFIKPIL EKMDGTEELL VKLNREDLLR KQRTFDNGSI PHQIHLGELH
 AILRRQEDFY PFLKDNREKI EKILTFRIPY YVGPLARGNS RFAWMTRKSE ETITPWNFEE VVDKGASAQs
 FIERMTNFDK NLPNEKVLPK HSLLYEYFTV YNELTKVKYV TEGMRKPAFL SGEQKKAIVD LLFKTNRKVT
 VKQLKEDYFK KIECFDSVEI SGVEDRFNAS LGTYHDLKI IKDKDFLDNE ENEDILEDIV LTLTLFEDRE
 MIEERLKTya HLFDDKVMKQ LKRRRYTGWG RLSRKLINGI RDKQSGKTIL DFLKSDGFAN RNFMQLIHDD
 SLTFKEDIQK AQVSGQDLSL HEHIANLAGS PAIKKGILQT VKVDELVKV MGRHKPENIV IEMARENQTT
 QKGQKNSRER MKRIEEGIKE LGSQILKEHP VENTQLQNEK LYLYLQNGR DMVVDQELDI NRLSDYDVDH
 IVPQSFLKDD SIDNKVLTRS DKNRGKSDNV PSEEVVKMK NYWRQLLNAK LITQRKFDNL TKAERGGLSE
 LDKAGFIKRQ LVETRQITKH VAQILD SRMN TKYDENDKLI REVKVITLKS KLVSDFRKDF QFYKVVREINN
 YHHAHDAYLN AVVGTALIKK YPKLESEFVY GDYKVYDVRK MIAKSEQEIG KATAKYFFYS NIMNFFKTEI
 TLANGEIRKR PLIETNGETG EIVWDKGRDF ATVRKVL SMP KGKSKKLKSV KELLGITIME RSSFEKNPID FLEAKGYKEV
 ARKKDWDPPK YGGFDSPTVA YSVLVVAKVE KRM LASAGEL QKGNELALPS KYVNFLYLAS HYEKLKGSPE DNEQQLFVE
 KKDIIKLPK YSLFELENGR KRMLASAGEL QKGNELALPS KYVNFLYLAS HYEKLKGSPE DNEQQLFVE
 QHKHYLDEII EQISEFSKRV ILADANLDKV ISAYNKHHRDK PIREQAENII HLFTLTNLGA PAAFKYFDTT
 IDRKRYTSTK EVLDATLIHQ SITGLYETRI DLSQLGGDRS RA **DPKKRRKV** **NLS2**
 SRGSGSA **HHH HHH** **NLS3** **NLS4**
 His标签

图41

NLS1

1 MKRTADGSEF **ESPKKKRKVE** DKKYSIGLDI GTNSVGWAVI TDEYKVP SKK FKVLGNTDRH SIKKNLIGAL
 71 LFDSGETAEA TRLKRTARRR YTRRKNRICY LQEIFSNEMA KVDDSEFFHRL EESFLVEEDK KHERHPIFGN
 141 IVDEVAYHEK YPTIYHLRKK LVDSTDKADL RLIYLALAHM IKFRGHFLIE GDLPDNDSDV DKLFIQLVQT
 211 YNQLFEENPI NASGVDAKAI LSARLSKRR IGDQYADLFL AAKNLSDAIL LSDILRVNTE IALSGLTPN FKSDFDLAED
 281 AKQLSKD TY DDDLDNLLAQ IDQYADLFL AAKNLSDAIL LSDILRVNTE ITKAPLSASM IKRYDEHHQD
 351 LTLLKALVRQ QLPEKYKEIF FDQSKNGYAG YIDGGASQEE FYKFIKPILE KMDGTEELLV KLNREDLLRK
 421 QRTFDNGSIP HQIHLGELHA ILRRQEDFYP FLKDNREKIE KILTFRIPYY VGPLARGNSR FAWMTRKSEE
 491 TITPWNFEV VDKGASAQSF IERMTNFDKN LPNEKVLPHK SLLYEYFTVY NELTKVKYVT EGMKRPAPLS
 561 GEQKKAIVDL LFKTNRKVTV KQLKEDYFKK IECFDSVEIS GVEDRFNASL GTYHDLKII KDKDFLDNEE
 631 NEDILEDIVL TLTLFEDREM IEERLKTYAH LFDDKVMKQL KRRRYTGWR LSRKLINGIR DKQSGKTILD
 701 FLKSDGFANR NFMQLIHDDS LTFKEDIQKA QVSGQGDSLH EHIANLAGSP AIKKGILQTV KVVDELVKVM
 771 GRHKPENIVI EMARENQTTQ KGQKNSRERM KRIEEGIKEL GSQILKEHPV ENTQLQNEKL YLYYLQNGRD
 841 MYVDQELDIN RLSYDQVDHI VPQSFLKDDS IDNKVLTRSD KNRGKSDNVP SEEVVKMKMN YWRQLLNAKL
 911 ITQRKFDNLT KAERGGLSEL DKAGFIKRQL VETRQITKHV AQILDSRMNT KYDENDKLIR EVKVITLKS
 981 LVSDFRKDFQ FYKVREINNY HHAHDAYLNA VVGITALIKKY PKLESEFVYG DYKVYDVRKM IAKSEQEIGK
 1051 ATAKYFFYSN IMNFFKTEIT LANGEIRKRP LIETNGETGE IVWDKGRDEFA TVRKVLSMPQ VNIVKKTEVQ
 1121 TGGFSKESIL PKRNSDKLIA RKKDWDPKKY GGFDSPTVAY SVLVVAKVEK GKSKKLKSVK ELLGITIMER
 1191 SSFEKNPIDF LEAKGYKEVK KDLIKLPKY SLFELENGRK RMLASAGELQ KGNEIALPSK YVNFYLLASH
 1261 YEKLKGPED NEQQLFVEQ HKHYLDEIIE QISEFSKRVI LADANLDKVL SAYNKHDKP IREQAENIIH
 1331 LFTLTNLGAP AAFKYFDTTI DRKRYSTKE VLDATLIHQ S ITGLYETRID LSQLG **DKRP AATKAGQAK**
 1401 **KKK**

NLS2

图42

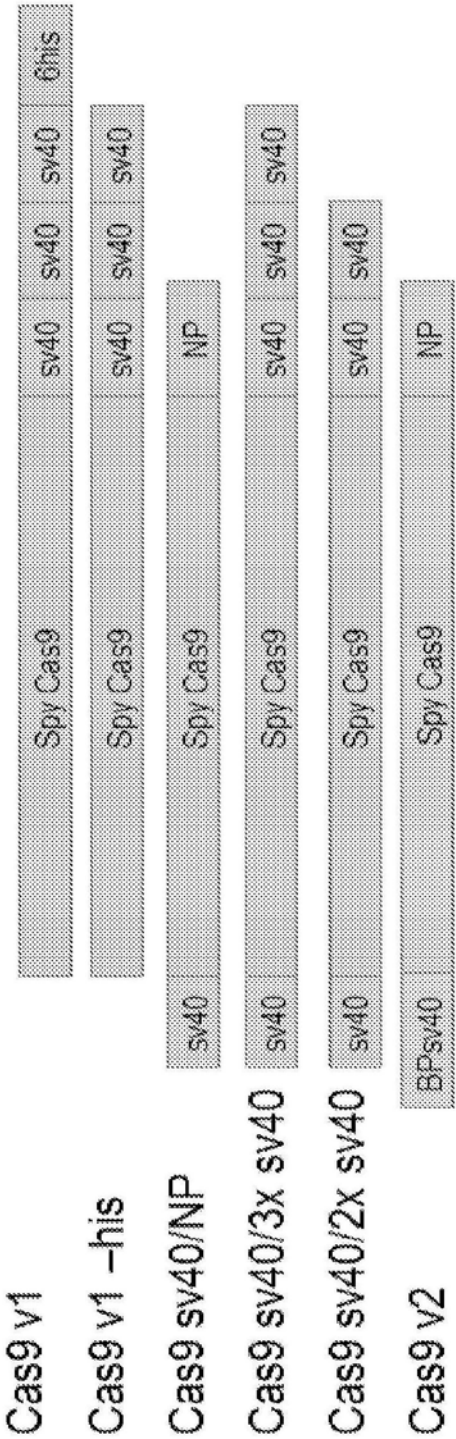


图43

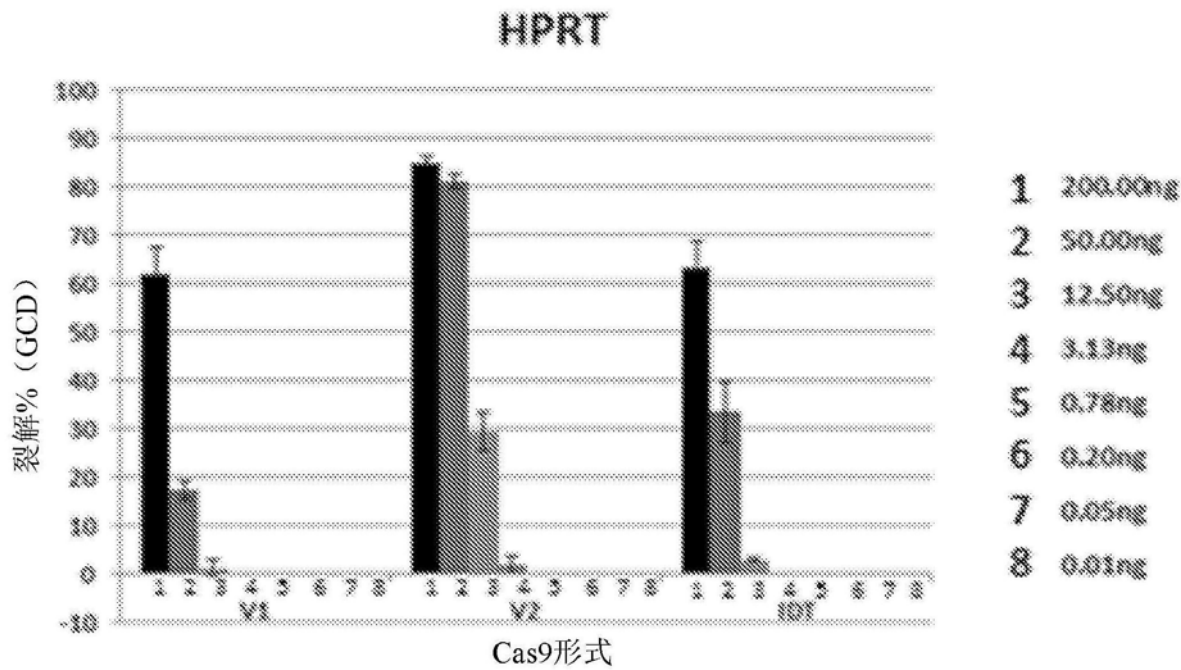


图44A

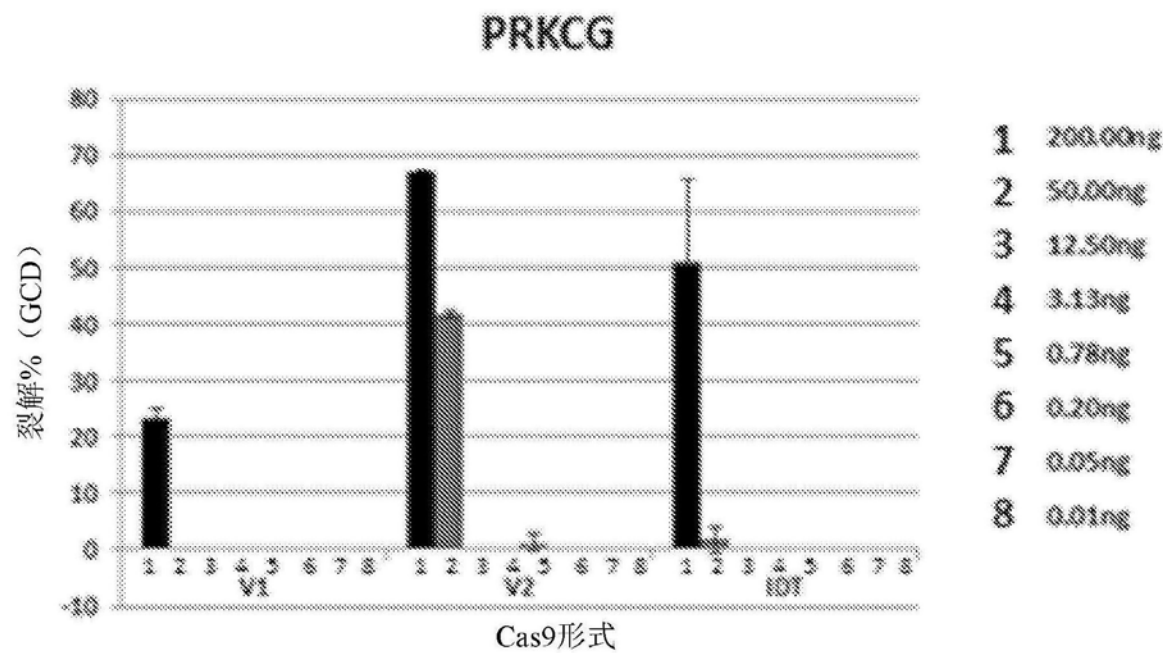


图44B

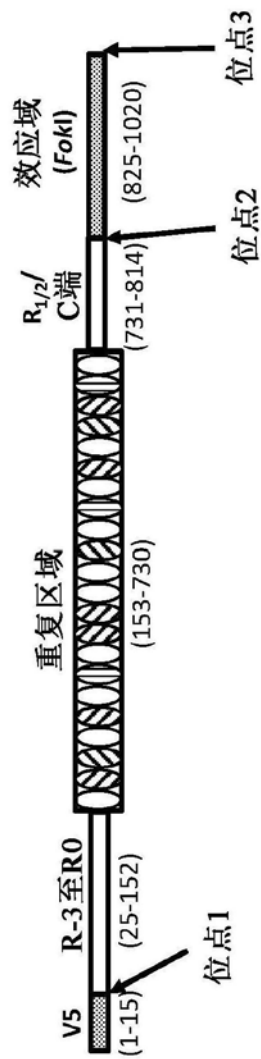


图45

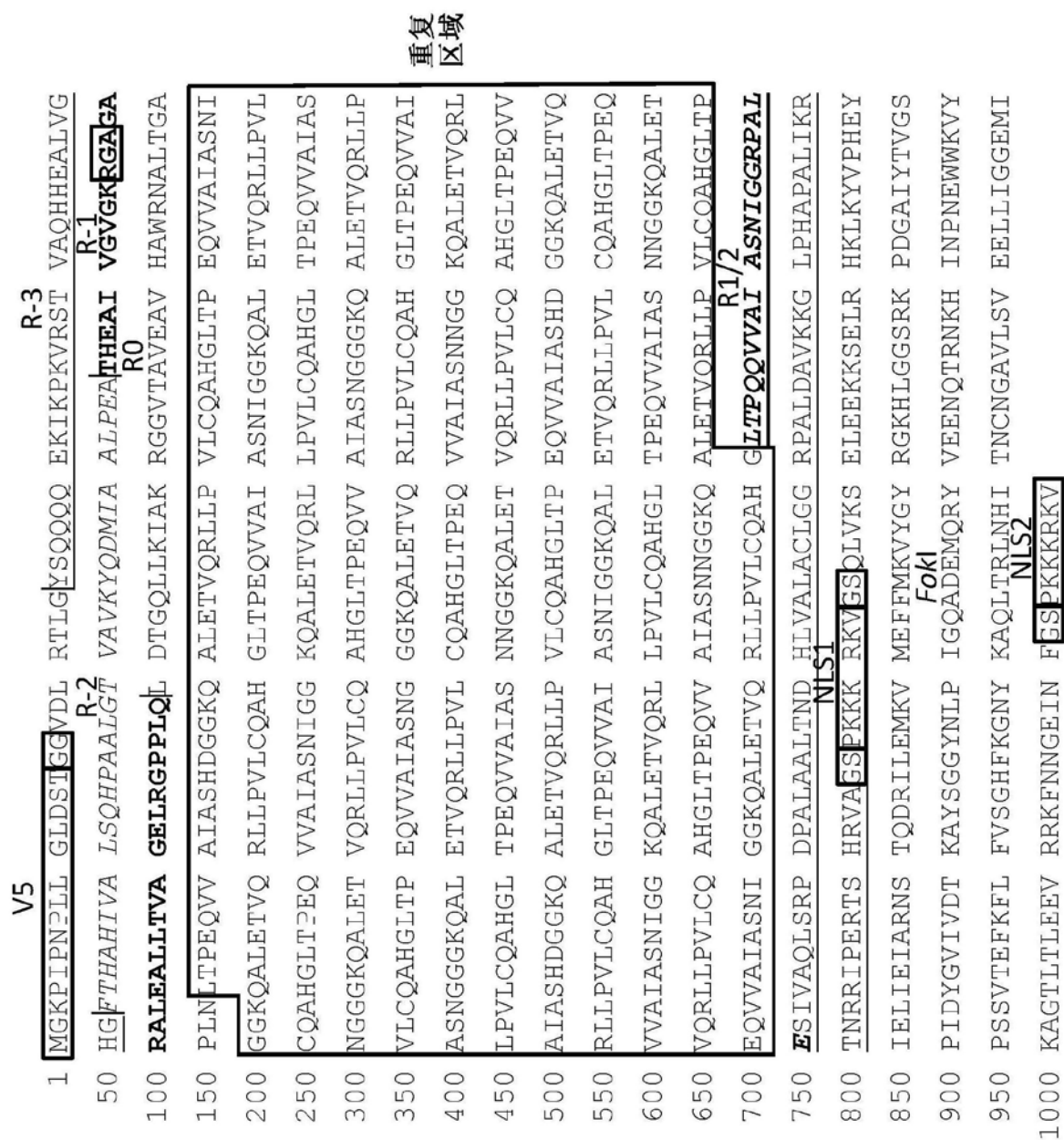


图46

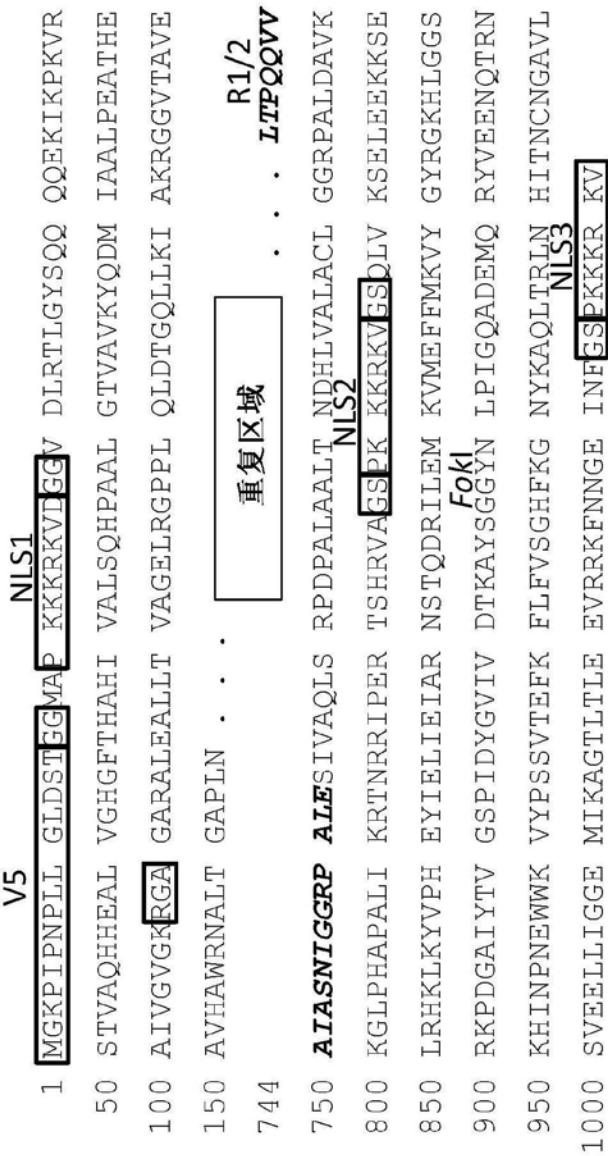


图47

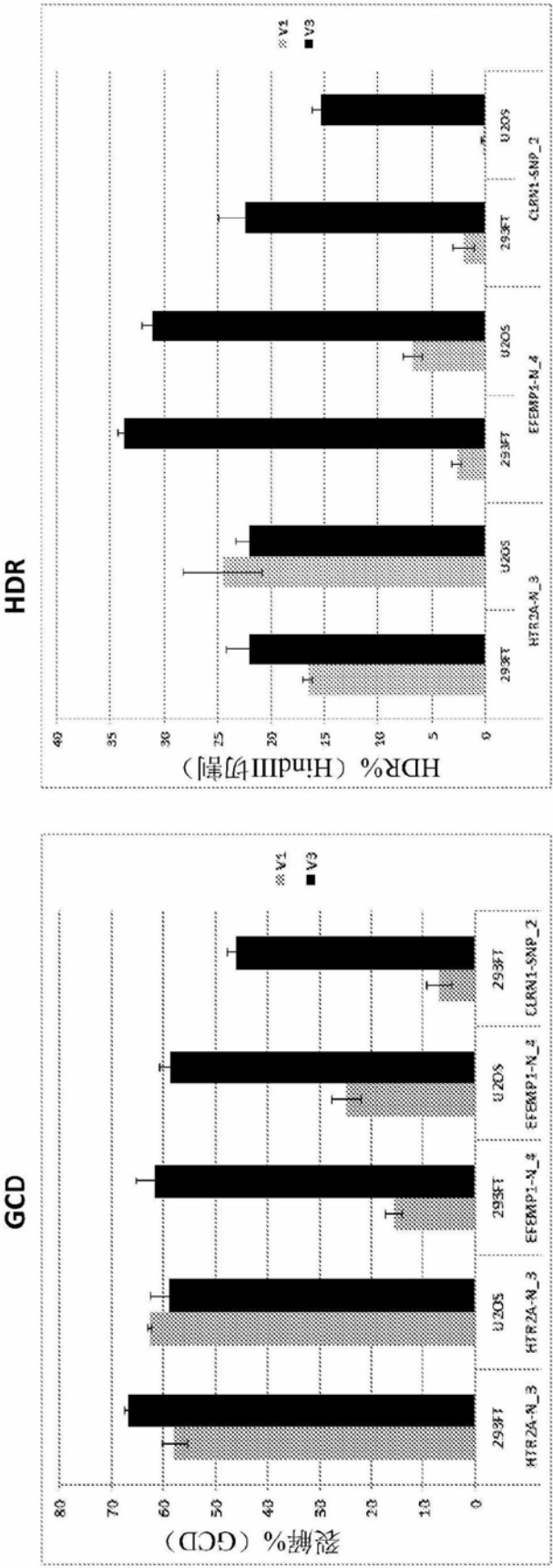


图49

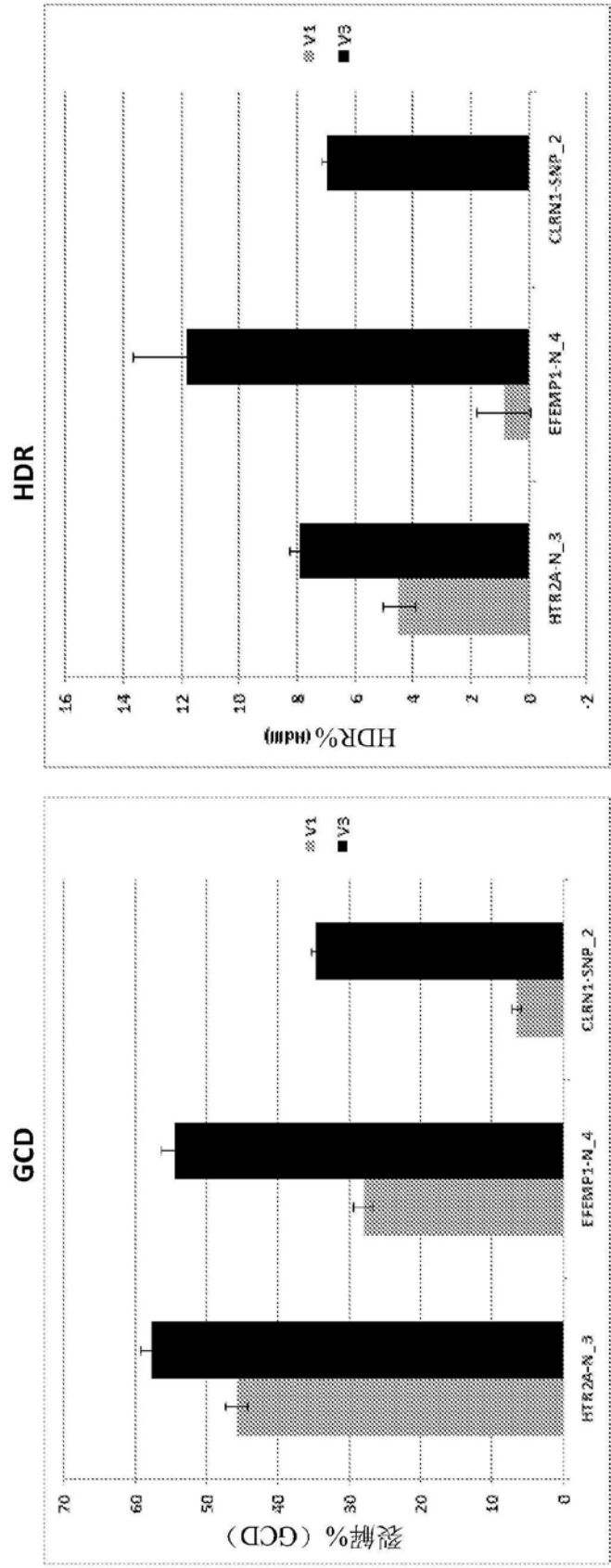


图50

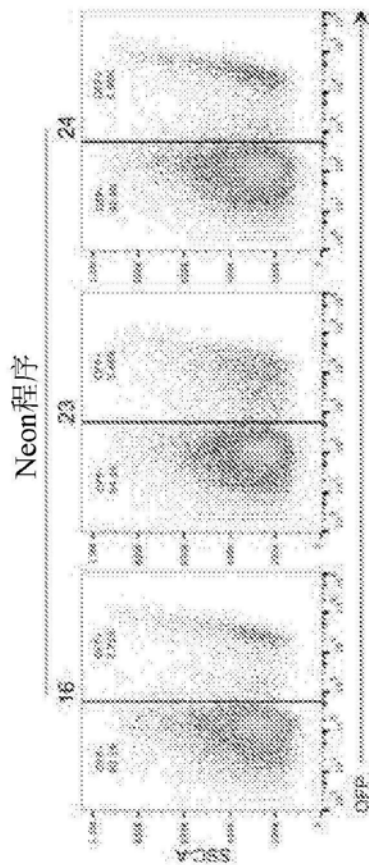


图51A

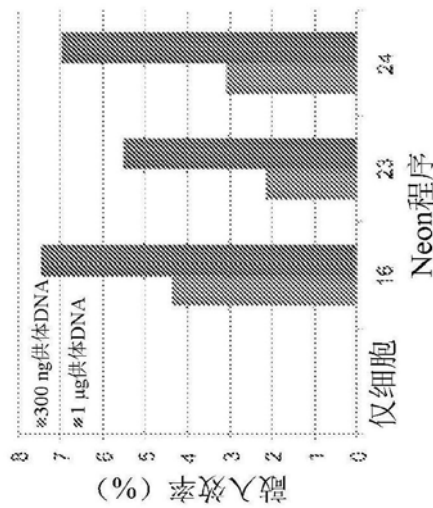


图51B

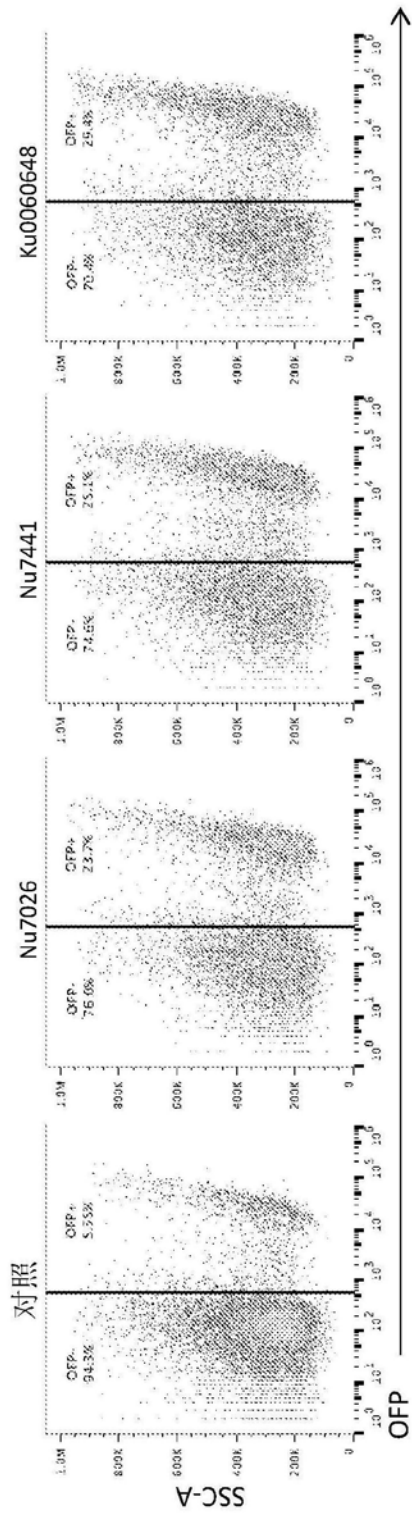


图52A

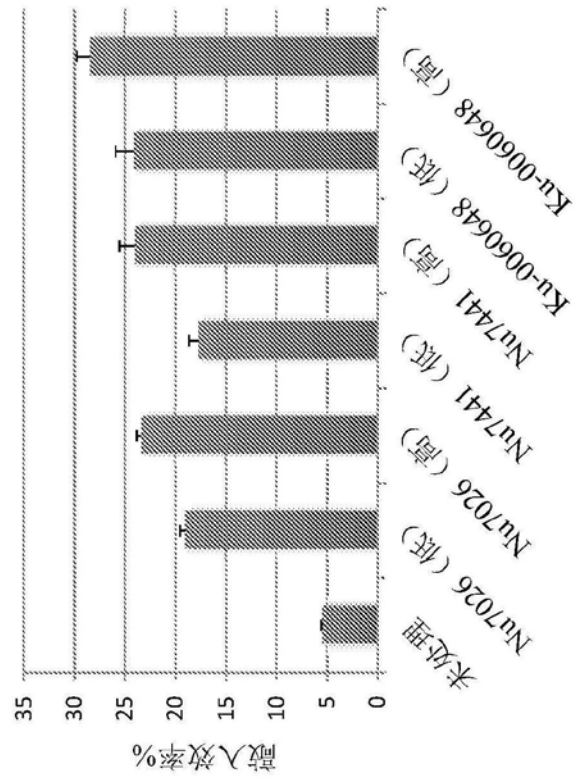


图52B

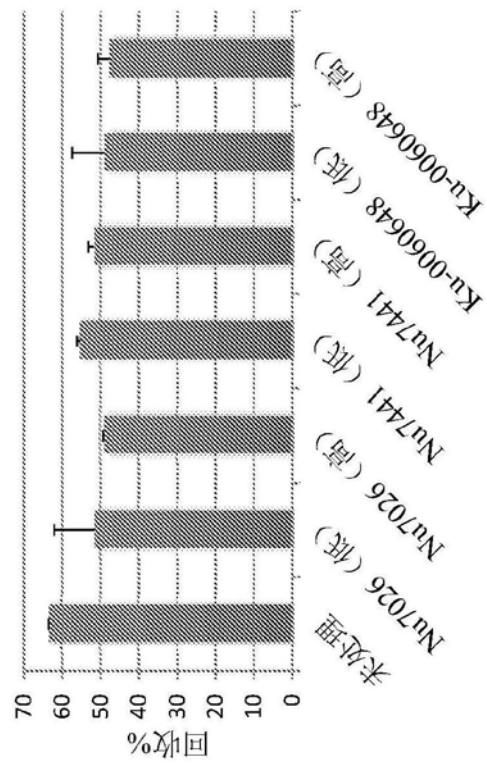


图52C

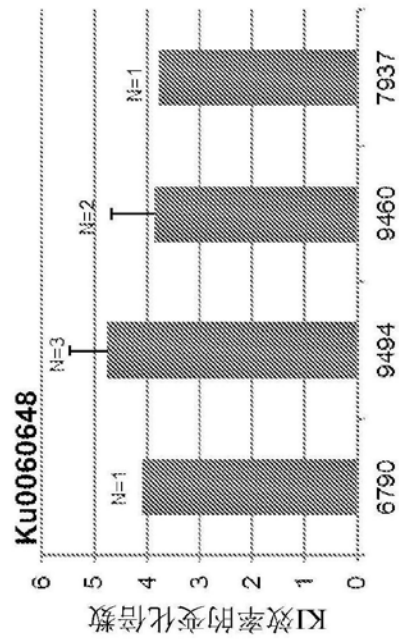


图52D

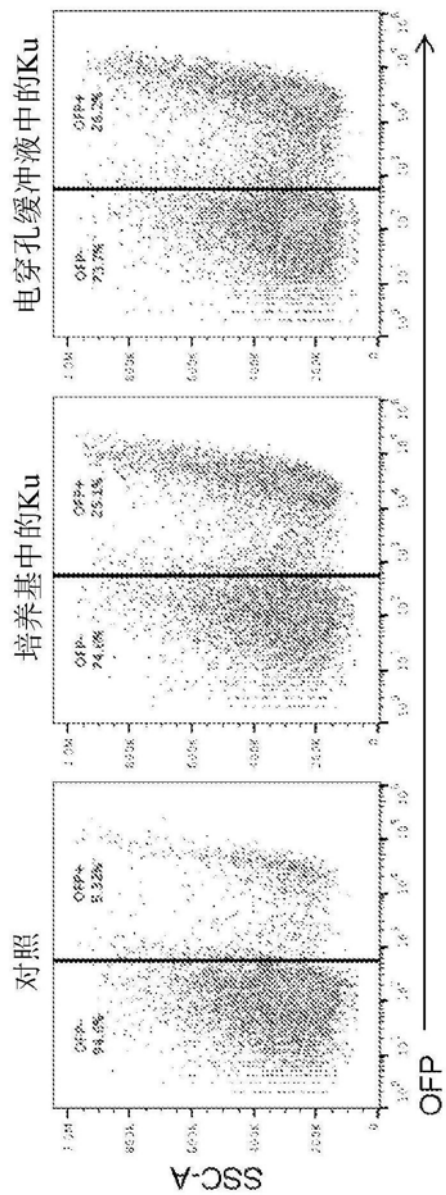


图53A

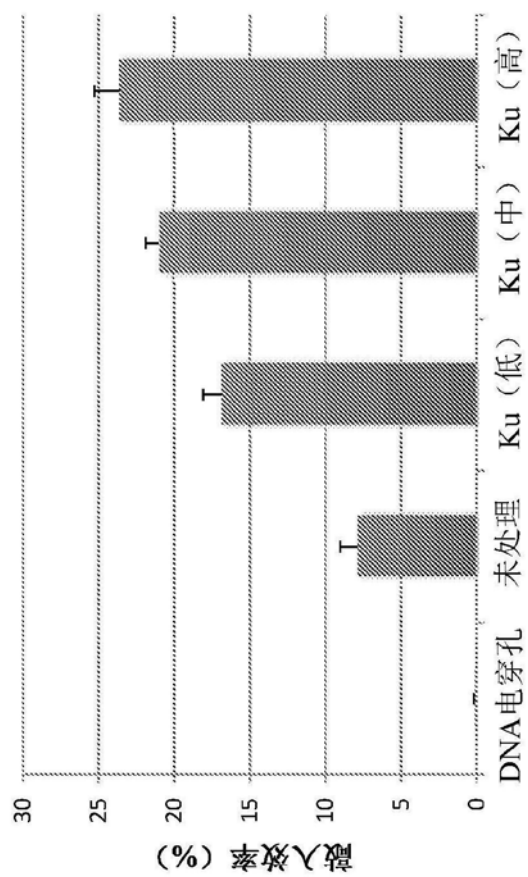


图53B

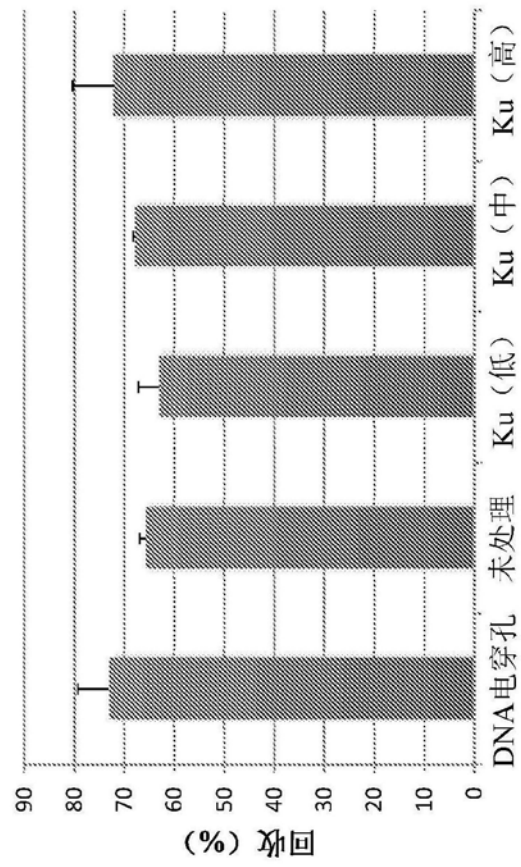


图53C

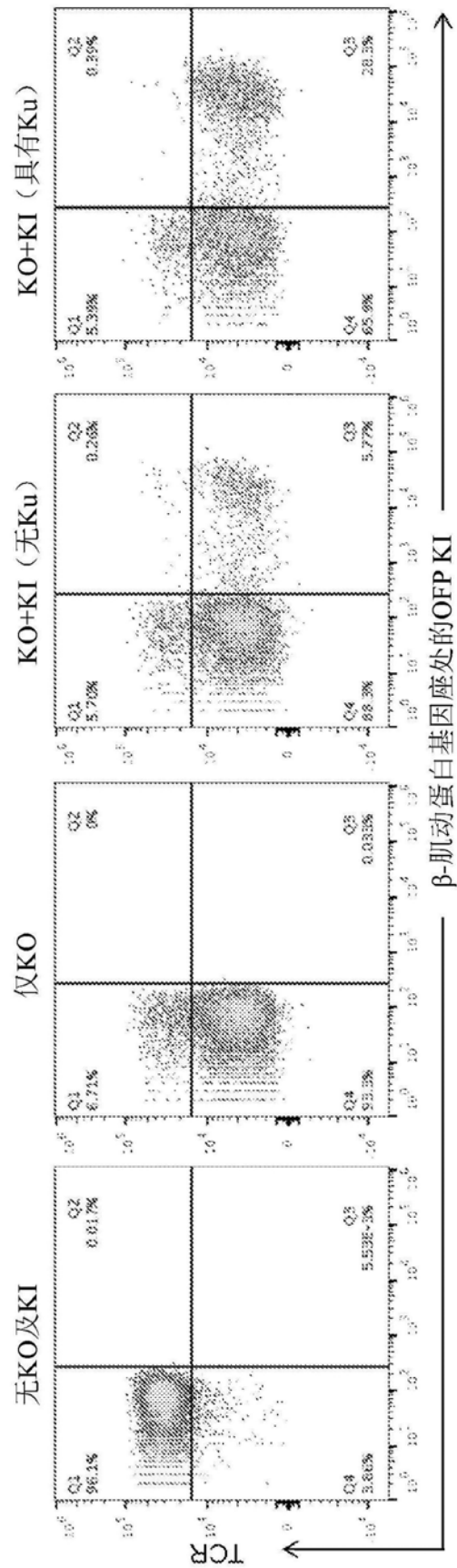


图54

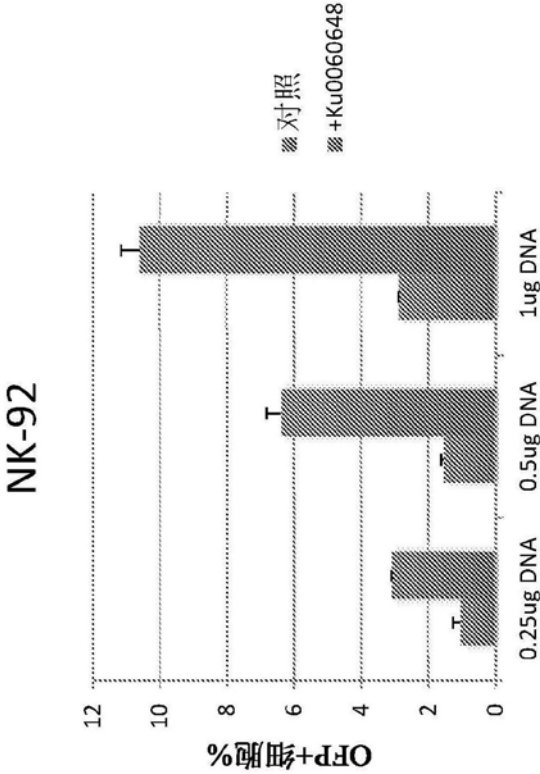


图55A

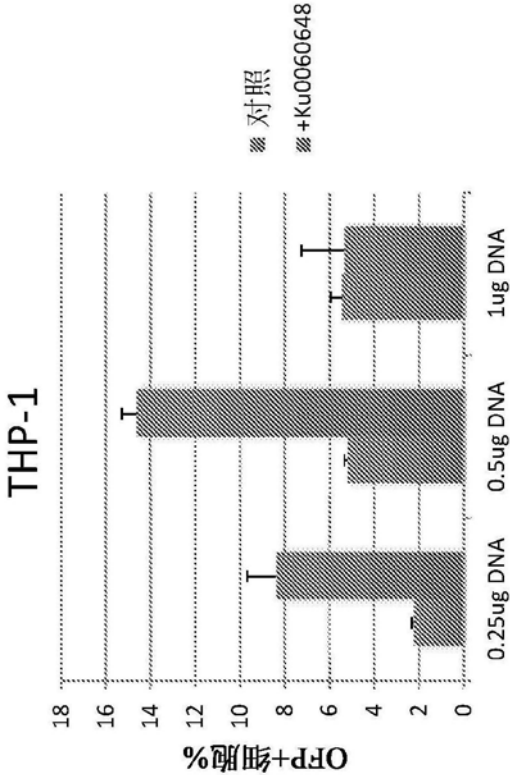


图55B

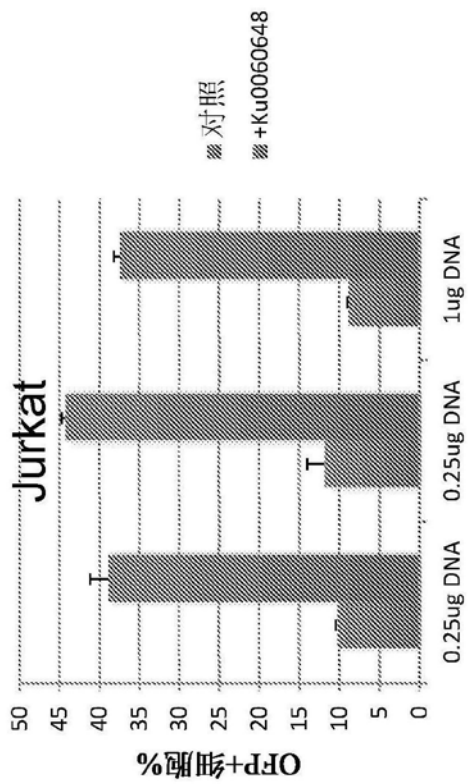


图55C

具有TrueCm Cas9和E1v2的GIBCO iPSC的经校正的SNPE百分比				
目标	无增强子	具有增强子2	具有增强子3	
目标1	29%	37%	41%	
目标2	41%	55%	62%	
目标3	32%	48%	41%	
目标4	21%	27%	NGS不起作用	
目标5	21%	15%	6%	

图56A

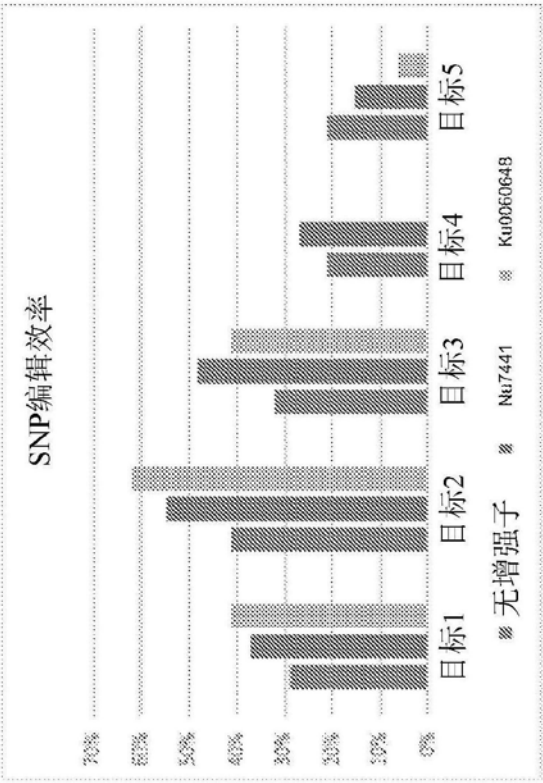


图56B

- 1. Nu7026
- 2. Nu7441
- 3. Ku0060648
- 4. Nu7026 + Nu7441
- 5. Nu7026 + Ku0060648
- 6. Nu7441 + Ku0060648
- 7. Nu7026 + Nu7441 + Ku0060648

图57A

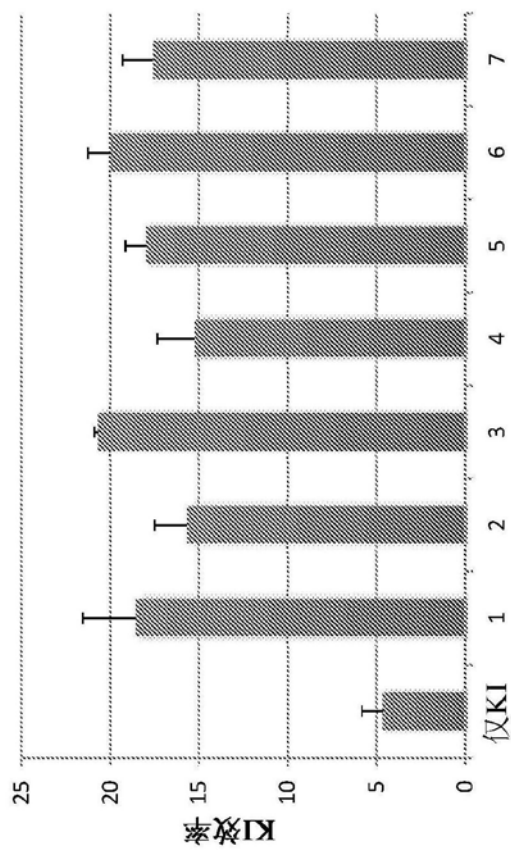


图57B

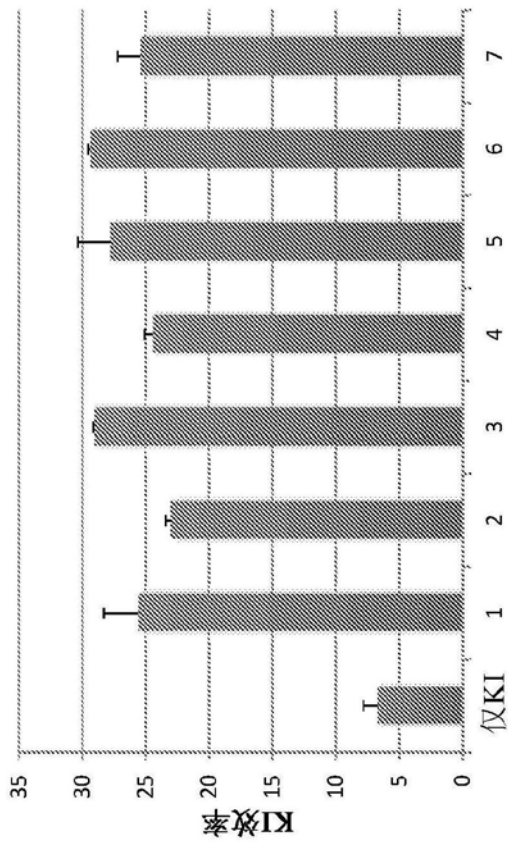


图57C