(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 120092083 A
(43)申请公布日 2025.06.03

(21)申请号 202380073912.9
(22)申请日 2023.10.19
(30)优先权数据
10-2022-0134734 2022.10.19 KR
(85)PCT国际申请进入国家阶段日
2025.04.18
(86)PCT国际申请的申请数据
PCT/KR2023/016230 2023.10.19
(87)PCT国际申请的公布数据
WO2024/085674 KO 2024.04.25
(71)申请人 财团法人峨山社会福祉财团
地址 韩国
申请人 蔚山大学校产学协力团
成均馆大学校产学协力团
(72)发明人 金容涉 权志妍 金大植
(74)专利代理机构 北京银龙知识产权代理有限公司 11243
专利代理师 许静 宋海花
(51)Int.Cl.
C12N 9/22(2006.01)
C12N 9/78(2006.01)
C12N 15/62(2006.01)
C12N 15/10(2006.01)
C12N 15/113(2006.01)

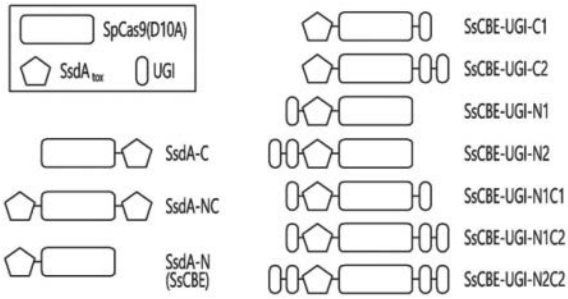
权利要求书2页 说明书21页
序列表（电子公布） 附图12页

(54)发明名称

包含Cas蛋白和细菌毒素的融合蛋白及其用途

(57)摘要

本发明涉及一种包含Cas蛋白和细菌毒素的融合蛋白及其用途，根据一方案的融合蛋白、其多肽、以及包含该融合蛋白的CRISPR-Cas系统，可以进行有效的碱基编辑。此外，由于多肽的尺寸小，具有易于通过载体递送的效果。另外，通过其诱导插入缺失时，插入缺失效率显著增加，对于插入缺失诱导时，插入缺失的效率显著提升，形成插入缺失的核苷酸的大小也增加，从而可以有效地用于基因的敲除。



CN 120092083 A

1.一种融合蛋白,其包含Cas蛋白和细菌毒素,其中,所述细菌毒素为单链DNA脱氨酶毒素A。

2.根据权利要求1所述的融合蛋白,其中,所述Cas蛋白为Cas9蛋白或Cas12蛋白。

3.根据权利要求1所述的融合蛋白,其中,所述单链DNA脱氨酶毒素A为胞苷脱氨酶。

4.根据权利要求1所述的融合蛋白,其中,所述单链DNA脱氨酶毒素A为失活的单链DNA脱氨酶毒素A。

5.根据权利要求1所述的融合蛋白,其中,所述单链DNA脱氨酶毒素A包含SEQ ID NO:1的氨基酸序列。

6.根据权利要求1所述的融合蛋白,其中,所述单链DNA脱氨酶毒素A包含与SEQ ID NO:1的氨基酸序列具有至少85%以上的序列同一性的序列。

7.根据权利要求4所述的融合蛋白,其中,所述失活的单链DNA脱氨酶毒素A在单链DNA脱氨酶毒素A的催化活性位点发生了氨基酸突变。

8.根据权利要求4所述的融合蛋白,其中,所述失活的单链DNA脱氨酶毒素A在SEQ ID NO:1的氨基酸序列中具有G302D、E349A、或其对应的氨基酸突变。

9.根据权利要求4所述的融合蛋白,其中,所述失活的单链DNA脱氨酶毒素A具有SEQ ID NO:17的氨基酸序列。

10.根据权利要求4所述的融合蛋白,其中,所述失活的单链DNA脱氨酶毒素A包含与SEQ ID NO:17的氨基酸序列具有至少85%以上的序列同一性的序列。

11.根据权利要求1所述的融合蛋白,其中,所述细菌毒素结合于Cas蛋白的C端、N端、或C端和N端两端。

12.根据权利要求1所述的融合蛋白,其中,所述融合蛋白还包含DNA糖基化酶抑制剂。

13.根据权利要求12所述的融合蛋白,其中,所述DNA糖基化酶抑制剂为胸腺嘧啶糖基化酶抑制剂、尿嘧啶糖基化酶抑制剂、氧鸟嘌呤糖基化酶抑制剂、或烷基鸟嘌呤DNA糖基化酶抑制剂。

14.根据权利要求1所述的融合蛋白,其中,所述融合蛋白具有距靶序列5`端1bp至20bp范围的编辑窗口。

15.一种多核苷酸,其编码权利要求1至14中的任一项所述的融合蛋白。

16.一种载体,其包含权利要求15所述的多核苷酸。

17.一种CRISPR-Cas系统,其包含融合蛋白或编码所述融合蛋白的多核苷酸、以及向导多核苷酸,其中,所述融合蛋白包含Cas蛋白和细菌毒素,所述细菌毒素为单链DNA脱氨酶毒素A。

18.根据权利要求17所述的CRISPR-Cas系统,其中,所述向导多核苷酸包含CRISPR RNA和反式激活RNA,所述向导多核苷酸为双链向导RNA或单链向导RNA。

19.根据权利要求17所述的CRISPR-Cas系统,其中,所述系统在靶核酸分子的核苷酸序列中形成至少一个以上的核苷酸的缺失、插入、置换或插入缺失。

20.根据权利要求19所述的CRISPR-Cas系统,其中,所述系统在靶核酸分子的核苷酸序列中形成1bp至60bp的核苷酸的缺失、1bp至60bp的核苷酸的插入、或1bp至60bp的核苷酸的插入缺失。

21.根据权利要求17所述的CRISPR-Cas系统,其中,所述系统具有距靶序列5`端1bp至

20bp范围的编辑窗口。

22.一种编辑核酸的方法,其包含将核酸分子与CRISPR-Cas系统接触的步骤,其中,

所述编辑在所述核酸分子的核苷酸序列中形成至少一个以上的核苷酸序列的缺失、插入、置换或插入缺失,

所述CRISPR-Cas系统包含融合蛋白或编码所述融合蛋白的多核苷酸、以及向导多核苷酸,

所述融合蛋白包含Cas蛋白和细菌毒素,所述细菌毒素为单链DNA脱氨酶毒素A。

23.根据权利要求22所述的编辑核酸的方法,其中,所述编辑为所述核酸分子的核苷酸序列中1bp至60bp的核苷酸的缺失、1bp至60bp的核苷酸的插入、或1bp至60bp的核苷酸的插入缺失。

包含Cas蛋白和细菌毒素的融合蛋白及其用途

技术领域

[0001] 本发明涉及一种包含Cas蛋白和细菌毒素的融合蛋白及其用途。

背景技术

[0002] 基因组编辑(Genome editing)是一种自由编辑生物体遗传信息的技术。通过生命科学领域的进步和基因组测序技术的发展,我们能够广泛地了解各种遗传信息。例如,已经确保了有关动植物的繁殖、疾病和生长、引发各种人类遗传疾病的基因突变、以及对用于生物燃料生产等基因的了解,但要达到直接应用其改善生物体、治疗人类疾病的水平,还需要进一步的技术进步。

[0003] 整体编辑技术可以通过改变包含人类在内的动物、植物、微生物的遗传信息来极大地扩大其应用范围。基因剪刀是为精确切割所需的遗传信息而设计和制造的分子工具,在基因组编辑技术中发挥着关键作用。与将基因测序领域推向新高度的下一代测序技术(Next generation sequencin)一样,基因剪刀正在成为扩大基因信息应用速度范围,创造新产业领域的核心技术。

[0004] 然而,APOBEC或AID蛋白是通常用于使用Cas9进行胞嘧啶碱基编辑的蛋白,其尺寸相对较大,存在难以制作成用于细胞治疗剂的载体的问题。因此,需要生产一种易于作为载体生产且具有优异的碱基编辑效率的CRISPR-Cas系统。

发明内容

[0005] 技术课题

[0006] 在一方案,提供一种融合蛋白,其包含Cas蛋白(CRISPR-associated protein)和细菌毒素(toxin)。

[0007] 在另一方案,提供一种多核苷酸(polynucleotide),其编码所述融合蛋白。

[0008] 又另一方案,提供一种载体,其包含所述多核苷酸。

[0009] 又另一方案,提供一种CRISPR-Cas系统,其包含所述融合蛋白或编码该融合蛋白的多核苷酸、以及向导多核苷酸。

[0010] 又另一方案,提供一种编辑核酸的方法,其包含将核酸(nucleic acid)分子与所述CRISPR-Cas系统接触的步骤。

[0011] 技术方案

[0012] 在一方案,提供一种融合蛋白,其包含Cas蛋白(CRISPR-associated protein)和细菌毒素(toxin)。

[0013] 在本说明书中,术语“Cas蛋白(CRISPR-associated protein)”可以是CRISPR结合核酸内切酶。所述Cas蛋白可以切割特定靶多核苷酸序列的全部或部分。

[0014] 所述Cas蛋白可以是2类Cas(Class 2CRISPR associated system)蛋白。所述2类Cas蛋白可以包括在II型、V型、或VI型系统中。

[0015] 所述II型系统可以具有cas1、cas2、以及cas9基因。II型系统可进一步分为三个亚

型,即,II-A、II-B、以及II-C亚型。II-A亚型可以包含额外的基因csn2。具有II-A亚型系统的生物体可以有嗜热链球菌(Streptococcus thermophilus)。II-B亚型缺失csn2,但可以具有cas4。具有II-B亚型系统的生物体可以有嗜肺军团菌(Legionella pneumophila)。II-C亚型是细菌中最常见的II型系统,可以仅具有Cas1、Cas2、以及Cas9的三种蛋白。具有II-C亚型系统的生物体可以有乳酸奈瑟球菌(Neisseria lactamica)。

[0016] 所述V型系统可以具有cas12基因以及cas1和cas2基因。Cas12基因可以编码Cas12蛋白,其具有与Cas9各个区域同源的RuvC样核酸酶结构域,但缺乏存在于Cas9蛋白的HNH核酸酶结构域。

[0017] 所述VI型系统可以具有cas13基因以及cas1和cas2基因。

[0018] 在所述II型系统中,Cas9的RuvC样核酸酶(RNase H折叠)结构域和HNH(McrA样)核酸酶结构域分别可以切割靶核酸中的一条链。II型系统的Cas9切割活性还可能需要crRNA的tracrRNA杂交,以便形成促进由Cas9引起的crRNA和靶向结合的双链体。

[0019] 在所述V型系统中,Cas12的RuvC样核酸酶结构域可以以交错结构切割靶核酸的两条链,产生5`突出端(overhang)。这些5`突出端可能有助于通过非同源末端连接方法插入DNA。V型系统的Cas12切割活性也不需要crRNA与tracrRNA的杂交来形成双链体,V型系统的crRNA可以使用具有形成内部双链体的茎环结构的单个crRNA。所述V型系统可以在靶序列的位点诱导单链或双链的断裂。链断裂可以是利用5`突出端的交错切割。

[0020] 所述Cas蛋白可以包含Cas9或Cas12。

[0021] 所述Cas12蛋白可以指来源于源自多种细菌物种的蛋白质。所述Cas蛋白可以是源自链球菌属(Streptococcus)、弯曲杆菌属(Campylobacter)、硝化杆菌属(Nitratifractor)、葡萄球菌属(Staphylococcus)、微小杆菌属(Parvibaculum)、罗氏菌属(Roseburia)、奈瑟菌属(Neisseria)、葡萄糖醋杆菌属(Gluconacetobacter)、固氮螺菌属(Azospirillum)、球毛菌属(Sphaerochaeta)、乳杆菌属(Lactobacillus)、真杆菌属(Eubacterium)、棒状杆菌属(Corynebacter)、肉杆菌属(Carnobacterium)、红杆菌属(Rhodobacter)、李斯特菌属(Listeria)、帕卢迪杆菌属(Paludibacter)、梭菌属(Clostridium)、毛螺菌科(Lachnospiraceae)、梭菌属(Clostridiaridium)、纤毛菌属(Leptotrichia)、弗朗西斯菌属(Francisella)、军团菌属(Legionella)、脂环酸芽孢杆菌属(Alicyclobacillus)、嗜甲烷菌属(Methanomethyophilus)、卟啉单胞菌属(Porphyromonas)、普雷沃氏菌属(Prevotella)、拟杆菌门(Bacteroidetes)、螺杆球菌属(Helcococcus)、钩端螺旋体属(Letospira)、脱硫弧菌属(Desulfovibrio)、脱磺钠菌属(Desulfonatronum)、奥匹图菌属(Opitutaceae)、结核杆菌属(Tuberibacillus)、芽孢杆菌属(Bacillus)、短杆菌属(Brevibacilus)、甲基杆菌属(Methylobacterium)、或氨基酸球菌属(Acidaminococcus)。更具体地,所述Cas12蛋白可以是源自选自下组中的一种细菌物种:土拉弗朗西斯菌1、土拉弗朗西斯菌诺维奇亚种、口腔普雷沃氏菌、毛螺菌科细菌MC20171、丁酸弧菌属的解蛋白丁酸弧菌、佩列格林细菌属的细菌GW2011_GWA2_33_10、朴古菌属的细菌GW2011_GWC2_44_17、斯氏密螺旋体种SCADC、氨基酸球菌种BV3L6、毛螺菌科细菌MA2020、热甲烷等离子体坎迪达图斯、优雅真杆菌、博氏莫拉菌237、伊那达钩端螺旋体、毛螺菌科细菌ND2006、犬齿龈卟啉单胞菌3、狄氏普雷沃氏菌、以及猕猴卟啉单胞菌。

[0022] 此外,所述Cas12蛋白可以是选自Cas12a、mgCas12a、Cas12b、Cas12c、Cas12d、

Cas12e、Cas12f、Cas12g、Cas12h、Cas12i、以及Cas12j中的一种。所述Cas12蛋白可以包含Cas12蛋白的修饰。所述Cas12蛋白具有核酸酶活性时,Cas12蛋白可以被修饰为具有降低的核酸酶活性,例如,与野生型酶相比,至少70%、至少80%、至少90%、至少95%、至少97%、或100%的核酸酶失活。

[0023] 所述Cas9可以是肺炎链球菌(S.pneumoniae)、化脓性链球菌(S.pyogenes)、嗜热链球菌(S.thermophilus)、或空肠弯曲杆菌(C.jejuni)Cas9,可以包含源自这些生物体的突变Cas9。该酶可以是Cas9同源物或直系同源物(ortholog)。在一具体实例中,CRISPR酶可以针对在真核细胞中的表达进行密码子优化。在一具体实例中,CRISPR酶可以在靶序列的位点诱导单链或双链的断裂。所述Cas9蛋白可以包含Cas9蛋白的修饰。所述Cas9蛋白具有核酸酶活性时,Cas9蛋白可以被修饰为具有降低的核酸酶活性,例如,与野生型酶相比,至少70%、至少80%、至少90%、至少95%、至少97%、或100%的核酸酶失活。根据一具体实例,所述Cas9可以是Cas9 D10A。

[0024] 在一具体实例中,所述Cas蛋白可以是Cas9蛋白或Cas12蛋白。

[0025] 在本发明的一实施例中,至少一种核定位信号(nuclear localization signal;NLS)可以附着至编码Cas蛋白的核酸序列。在一具体实例中,可以附着至少一个以上的C端或N端核定位信号。可以编码Cas蛋白、其直系同源物或同系物,其包含一个以上的核定位序列(NLS),例如,约1、2、3、4、5、6、7、8、9、10个以上,或更多个核定位序列。在本文所述的Cas蛋白复合物的优选实例中,密码子优化的Cas蛋白可以包含附着至蛋白质C端的核定位序列。在特定实例中,其他定位标签可以与Cas蛋白融合,例如,将Cas定位到细胞内的特定位置,如细胞器,例如:线粒体、质体、叶绿体、囊泡、高尔基体、(核或细胞)膜、核糖体、核仁、内质网(ER)、细胞骨架、液泡、中心体、核小体、颗粒、中心粒等(不限于此)。

[0026] 所述Cas蛋白和细菌毒素可以通过接头(linker)融合。所述接头可以位于Cas蛋白的C端、N端、或C端和N端两端,通过所述接头细菌毒素可以与Cas蛋白结合。所述合适的接头基序和接头配制包含文献【Chen et al.,Fusion protein linkers:property,design and functionality.Adv Drug Deliv Rev.2013;65(10):1357-69】中所述的内容,其全部内容可以通过引用并入本说明书。

[0027] 在一具体实例中,所述细菌毒素可以是单链DNA脱氨酶毒素A(SsdA(single-stranded DNAdeaminase toxin A))。

[0028] 所述SsdA可以源自假单胞菌属(Pseudomonas SP.)菌株。所述假单胞菌属菌株可以是丁香假单胞菌(Pseudomonas syringae)、凝结假单胞菌(Pseudomonas congelans)、沙门氏假单胞菌(Pseudomonas savastanoi)、绿色假单胞菌(Pseudomonas viridiflava)、冠状假单胞菌(Pseudomonas coronafaciens)、荧光假单胞菌(Pseudomonas fluorescens)、假单胞菌属MPC6、假单胞菌属GL-R-26、或假单胞菌属GL-RE-26。

[0029] 所述SsdA的N端可以具有PAAR结构域,SsdA的C端可以具有DYW脱氨酶(deaminase)结构域。所述SsdA在具有常见的氨基酸基序(motif)HxE和CxxC基序这一点上,与现有碱基编辑中使用的脱氨酶相同,但在还具有SGW基序这一点上,与现有碱基编辑中使用的脱氨酶存在差异。所述SsdA是一种脱氨酶,其在结构和进化上与现有碱基编辑技术中使用的其他脱氨酶不同,被归类为DYW样脱氨酶(DYW-kike deaminase)。SsdA和现有碱基编辑技术中使用的其他脱氨酶的系统发育树与组成结构域的差异,如图1所示。

[0030] 根据一具体实例,所述SsdA可以包含SEQ ID NO:1的氨基酸序列。

[0031] 所述SsdA可以包含毒素结构域(domain)。所述SsdA的毒素结构域是具有脱氨酶活性的部分,可以具有100至200个氨基酸的长度,例如,120至180个氨基酸的长度、120至170个氨基酸的长度、130至160个氨基酸的长度、或140至160个氨基酸的长度。具体地,所述SsdA的毒素结构域可以包含SEQ ID NO:2的氨基酸序列。所述毒素结构域的碱基序列,示于表1。

[0032] 表1

[0033]

	氨基酸序列	SEQ ID NO
毒素结构域(domain)	KVSNIAESEAALGRASQARADLPQSKELKV KTVSSNDKKTLSGWGNKKPEGYERISAEQV KAKSEEIGHEVKSHPYDRDYKGQYFSSHAE KQMSIASPNHPLGVSKPMCTDCQGYFSQLA KYSKVEQTVADPKAIRIFKTDGSVETIMRSE	SEQ ID NO: 2

[0034] 所述SsdA的SEQ ID NO:2的氨基酸序列(毒素结构域)可以包含催化活性位点(catalytic active site)。所述催化活性位点(catalytic active site)可以包含HxE基序、CxxC基序、或SGW基序。SGW基序是只有SsdA酶才具有的附加基序,与现有的APOBEC、AID等脱氨酶不同。具体地,HxE基序可以包含SEQ ID NO:1(包含PAAR结构域的蛋白质(PAAR domaomn-containing protein))的第301至303位氨基酸,HxE基序可以包含第347至349位氨基酸,SGW基序可以包含第301至303位氨基酸。

[0035] 在一具体实例中,所述SsdA与SEQ ID NO:1的氨基酸序列(包含PAAR结构域的蛋白质)可以具有至少80%以上、85%以上、90%以上、91%以上、92%以上、93%以上、94%以上、95%以上、96%以上、97%以上、98%以上、或99%以上的序列同一性(sequence identity)。

[0036] 在一具体实例中,所述融合蛋白可以是Cas蛋白与DYW脱氨酶蛋白结合而成的。

[0037] 在一具体实例中,所述SsdA可以是失活的SsdA。

[0038] 所述失活的SsdA可以是在活化的SsdA的催化活性位点上发生氨基酸突变。所述失活的SsdA可以具有比SsdA更低的细胞毒性。

[0039] 根据一具体实例,所述失活的SsdA可以在SEQ ID NO:1(包含PAAR结构域的蛋白质)的氨基酸序列中G302、E349位点具有氨基酸突变。所述氨基酸突变是指野生型(wild type)蛋白质被G302、E349位点具有的氨基酸以外的其他氨基酸置换。所述其他氨基酸可以是选自精氨酸(R)、组氨酸(H)、赖氨酸(K)、天冬氨酸(D)、谷氨酸(E)、丝氨酸(S)、苏氨酸(T)、天冬酰胺(N)、谷氨酰胺(Q)、半胱氨酸(C)、硒代半胱氨酸(U)、甘氨酸(G)、脯氨酸(P)、丙氨酸(A)、缬氨酸(V)、异亮氨酸(I)、亮氨酸(L)、甲硫氨酸(M)、苯丙氨酸(F)、酪氨酸(Y)、色氨酸(W)以及所述氨基酸的所有变体组成的组中的任意一种,不包括野生型蛋白质在突变位点具有的氨基酸。具体地,所述失活的SsdA可以在SEQ ID NO:1的氨基酸序列中具有G302D、E349A或其相应的氨基酸突变。更具体地,在所述SEQ ID NO:1(包含PAAR结构域的蛋白质)的氨基酸序列中具有G302D突变的失活的SsdA可以是SEQ ID NO:17。

[0040] 在一具体实例中,所述失活的SsdA与SEQ ID NO:17的氨基酸序列可以具有至少80%以上、85%以上、90%以上、91%以上、92%以上、93%以上、94%以上、95%以上、96%以上、97%以上、98%以上、或99%以上的序列同一性(sequence identity)。

[0041] 所述SsdA可以诱导单链DNA脱氨基,所述SsdA可以是胞苷脱氨酶(cytidine deaminase)。

[0042] 在本说明书中,术语“胞苷脱氨酶(cytidine deaminase)”是指具有除去胞嘧啶(cytosine)、胞苷(cytidine)、或脱氧胞苷(deoxycytidine)的氨基(-NH2)基团的活性的酶。在本说明书中,所述胞苷(cytidine)脱氨酶作为包含胞嘧啶(cytosine)脱氨酶的概念来使用。在本说明书中,所述胞苷(cytidine)脱氨酶可以与所述胞嘧啶(cytosine)脱氨酶混合使用。

[0043] 所述胞苷脱氨酶(cytidine deaminase)是指所有具有将存在于核苷酸中的碱基胞嘧啶(例如,存在于DNA或RNA中的胞嘧啶)转化为尿嘧啶(C-to-U转化或C-to-U编辑(C-to-U conversion or C-to-U editing))的活性的酶,将位于靶位点序列(靶核酸序列)的PAM序列所在的链上的胞嘧啶转化为尿嘧啶。

[0044] 所述细菌毒素可以是结合至Cas蛋白的末端。例如,所述细菌毒素可以结合至Cas蛋白的C端、N端、或C端和N端两端。

[0045] 在一具体实例中,所述融合蛋白还可以包含DNA糖基化酶抑制剂(DNA glycosylase inhibitor)。

[0046] 所述DNA糖基化酶抑制剂可以是胸腺嘧啶糖基化酶抑制剂(thymine glycosylase inhibitor)、尿嘧啶糖基化酶抑制剂(uracil glycosylase inhibitor)、氧鸟嘌呤糖基化酶抑制剂(oxoguanine glycosylase inhibitor)、或烷基鸟嘌呤DNA糖基化酶抑制剂(alkylguanine glycosylase inhibitor)。

[0047] 所述尿嘧啶DNA糖基化酶抑制剂可以是源自枯草芽孢杆菌(Bacillus subtilis)噬菌体、PBS1的尿嘧啶DNA糖基化酶抑制剂,源自枯草芽孢杆菌噬菌体或PBS2的尿嘧啶DNA糖基化酶抑制剂,但不限于此。

[0048] 在另一方案,提供一种多核苷酸(polynucleotide),其编码所述融合蛋白。

[0049] 又另一方案,提供一种载体,其包含所述多核苷酸。

[0050] 在本说明书中,术语“载体”可以指可以运输与其连接的另一核酸的核酸分子。载体可以是单链、双链、或部分双链的核酸分子;包含一个以上的自由端的核酸分子、没有自由端的核酸分子(例如,环状的);包含DNA、RNA或两者兼有的核酸分子;以及包含本领域已知的其他各种多核苷酸。载体的一种类型有“质粒”,其可以指环状双链DNA环,其中可以插入额外的DNA片段,例如通过标准分子克隆技术。另一种类型的载体是病毒载体,其中病毒衍生的DNA或RNA序列可以存在于载体中以包装成病毒(例如,逆转录病毒、复制缺陷型逆转录病毒、腺病毒、复制缺陷型腺病毒、以及腺相关病毒)。重组表达载体可以包含适合于在宿主细胞中表达核酸的形式的本发明的核酸,这可以意味着重组表达载体包含一个以上的调控元件,一个以上的调控元件可以基于待用于表达的宿主细胞进行选择,可以可操作地连接至待表达的核酸序列。在重组表达载体中,“可操作地连接”可以指感兴趣的核苷酸序列以允许所述核苷酸序列的表达的方式连接至调控元件(例如,在体外转录/翻译系统中或在宿主细胞中,如果载体被引入宿主细胞)。

[0051] 在本说明书中,术语“调控元件”可以包含启动子、增强子、内部核糖体进入位点(internal ribosomal enter site;IRES)、以及其他表达控制元件(例如,转录终止信号,例如,聚腺苷酸化信号和聚U序列)。调控元件在例如文献【Goeddel,GENE EXPRESSION TECHNOLOGY:METHODS IN ENZYMOLOGY185,Academic Press San Diego,Calif.(1990)】中有所描述。调控元件可以包括指导核苷酸序列在许多类型的宿主细胞中组成性表达的元件和指导核苷酸序列仅在特定宿主细胞中表达的元件(例如,组织特异性调控序列)。在一具体实例中,载体可以包含一个以上的Pol III启动子(例如,1、2、3、4、5、或更多的Pol III启动子)、一个以上的Pol II启动子(例如,1、2、2、3、4、5、或更多的POLII启动子)、一个以上的Pol I启动子(例如,1、2、3,、4、5、或更多的pol I启动子)、或其组合。pol III启动子的实例可以包括但不限于U6和H1启动子。pol II启动子的实例可以包括但不限于逆转录病毒劳斯肉瘤病毒(Rous sarcoma virus;RSV)LTR启动子(任选地具有RSV增强子)、巨细胞病毒(CMV)启动子(任选地具有CMV增强子)、SV40启动子、二氢叶酸还原酶启动子、β-肌动蛋白启动子、磷酸甘油激酶(phosphoglycerol kinase;PGK)启动子、以及EF1α启动子。例如,载体可以包括慢病毒和腺相关病毒(AAV1、AAV2、AAV3、AAV4、AAV5、AAV6、AAV7、AAV8、或AAV9),并且还可以选择载体的类型来靶向特定类型的细胞。

[0052] 此外,所述载体系统内的多个核酸分子可以位于相同或不同的载体上。

[0053] 在一具体实例中,载体,例如,质粒或病毒载体,如通过肌内注射递送至感兴趣的组织,而在其他情况下,递送可以通过静脉内、经皮、鼻腔内、口腔、粘膜、或其他方法进行。这种递送可以通过单剂量或多剂量中任一种进行。本领域技术人员将理解,在本说明书中递送的实际剂量可以根据多种因素而有很大变化,例如,载体选择、靶细胞、生物体、或组织、待治疗受试者的一般状况、寻求的转化/修饰的程度、给药途径、给药方法、寻求的转化/修饰类型等。给药剂量还可以包含,例如,载体(水、盐水、乙醇、甘油、乳糖、蔗糖、磷酸钙、明胶、葡聚糖、琼脂、果胶、花生油、芝麻油等)、稀释剂、药学上可接受的载体(例如,磷酸盐缓冲盐水)、药学上可接受的赋形剂、和/或本领域已知的其他化合物。给药剂量还可以包含一种以上的药学上可接受的盐,例如,无机酸盐,如盐酸盐、氢溴酸盐、磷酸盐、硫酸盐等;以及有机酸盐,例如,乙酸盐、丙酸盐、丙二酸盐、苯甲酸盐等。进一步地,本说明书中还可以给出辅助组分,例如,润湿剂或乳化剂、pH缓冲组分、凝胶或胶凝物质、调味剂、着色剂、微球、聚合物、悬浮剂等。进一步地,还可以存在一种以上的其他常规药物成分,例如,防腐剂、保水剂、悬浮剂、表面活性剂、抗氧化剂、填充剂、螯合剂、包衣剂、化学稳定剂等。适合的示例性成分包括微晶纤维素、羧甲基纤维素钠、聚山梨酯80、苯乙醇、三氯叔丁醇、山梨酸钾、山梨酸、二氧化硫、没食子酸丙酯、对羟基苯甲酸酯、乙基香草醛、甘油、苯酚、对氯苯酚、明胶、白蛋白及其组合。例如,可以通过AAV实现疾病治疗的递送。用于将AAV体内递送至人类的治疗有效剂量可以是约20ml至约50ml的盐水溶液,每ml溶液含有约1Y1010至约1Y10100的AAV。可以调整给药剂量以平衡治疗益处和任何副作用。

[0054] 又另一方案,提供一种CRISPR-Cas系统,其包含:包含Cas蛋白和细菌毒素的融合蛋白或编码所述融合蛋白的多核苷酸;以及向导多核苷酸。

[0055] 所述融合蛋白和编码该融合蛋白的多核苷酸,如上所述。

[0056] 所述向导多核苷酸可以包含靶向序列和/或激活序列。

[0057] 在本说明书中,术语“靶向序列”可以指包含与靶核酸中的序列互补的DNA、或DNA

和RNA的混合物的多核苷酸。在特定实施方案中,靶向序列还可以包含其他核酸、或核酸类似物、或其组合物。在特定实施方案中,靶向序列可以仅由DNA组成,因为这种构建体在宿主细胞降解的可能性较小。在一些实施方案中,此类配置可以增加靶向序列识别特异性和/或减少脱靶结合/杂交的发生。所述靶向序列可以包含向导序列或间隔序列。所述靶向序列的结构域的长度可以是至少12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、或30个核苷酸。

[0058] 在本说明书中,术语"激活序列"可以指包含RNA或DNA、或DNA和RNA的混合物的多核苷酸的一部分,其能够与Cas蛋白相互作用、或缔合、或结合。在一具体实例中,激活区域还可以包含其他核酸、或核酸类似物、或其组合物。在一具体实例中,激活序列可以邻近或连接至靶序列。在一具体实例中,激活区域可以位于靶向区域的下游。在一具体实例中,激活区域可以位于靶向区域的上游。所述激活序列可以包含直接重复序列、CRISPR RNA (CRISPR RNA:crRNA)和/或反式激活RNA(trans-activating RNA:tracrRNA)。

[0059] 在一具体实例中,向导多核苷酸、向导RNA、成熟crRNA、以及未成熟crRNA可以包含直接重复序列和向导序列或间隔序列,或由其组成。在一具体实例中,向导RNA或成熟crRNA可以包含与向导序列或间隔序列连接的直接重复序列,或由其组成。在一具体实例中,直接重复序列可以位于向导序列或间隔序列的上游(即5`)。

[0060] 在一具体实例中,所述向导多核苷酸可以包含crRNA和tracrRNA。

[0061] 在一具体实例中,所述向导多核苷酸可以是双链向导RNA(dual guide RNA)或单链向导RNA(single-chain guide RNA:sgRNA)。

[0062] 所述系统可以形成靶核酸(nucleic acid)分子的核苷酸(nucleotide)序列中至少一个以上的核苷酸的缺失(deletion)、插入(insertion)、置换(substitution)、或插入缺失(insertion and deletion;indel)。

[0063] 所述核酸可以是RNA或DNA。

[0064] 根据一具体实例,所述系统可以在靶核酸分子的核苷酸序列中形成1bp至60bp的核苷酸的缺失,例如,1bp至55bp、1bp至50bp、1bp至45bp、1bp至40bp、1bp至35bp、1bp至30bp、1bp至25bp、1bp至20bp、1bp至15bp、1bp至10bp、1bp至5bp、5bp至60bp、5bp至55bp、5bp至50bp、5bp至45bp、5bp至40bp、5bp至35bp、5bp至30bp、5bp至25bp、5bp至20bp、5bp至15bp、5bp至10bp、10bp至60bp、10bp至55bp、10bp至50bp、10bp至45bp、10bp至40bp、10bp至35bp、10bp至30bp、10bp至25bp、10bp至20bp、10bp至15bp、15bp至60bp、15bp至55bp、15bp至50bp、15bp至45bp、15bp至40bp、15bp至35bpbp、15bp至30bp、15bp至25bp、15bp至20bp、20bp至60bp、20bp至55bp、20bp至50bp、20bp至45bp、20bp至40bp、20bp至35bp、20bp至30bp、20bp至25bp、25bp至60bp、25bp至55bp、25bp至50bp、25bp至45bp、25bp至40bp、25bp至35bp、25bp至30bp、30bp至60bp、30bp至55bp、30bp至50bp、30bp至45bp、30bp至40bp、30bp至35bp、35bp至60bp、35bp至55bp、35bp至50bp、35bp至45bp、35bp至40bp、40bp至60bp、40bp至55bp、40bp至50bp、40bp至45bp、45bp至60bp、45bp至55bp、45bp至50bp、50bp至60bp、50bp至55bp、或55bp至60bp。

[0065] 根据一具体实例,所述系统可以在靶核酸分子的核苷酸序列中形成1bp至60bp的核苷酸的插入,例如,1bp至55bp、1bp至50bp、1bp至45bp、1bp至40bp、1bp至35bp、1bp至30bp、1bp至25bp、1bp至20bp、1bp至15bp、1bp至10bp、1bp至5bp、5bp至60bp、5bp至55bp、5bp

至50bp、5bp至45bp、5bp至40bp、5bp至35bp、5bp至30bp、5bp至25bp、5bp至20bp、5bp至15bp、5bp至10bp、10bp至60bp、10bp至55bp、10bp至50bp、10bp至45bp、10bp至40bp、10bp至35bp、10bp至30bp、10bp至25bp、10bp至20bp、10bp至15bp、15bp至60bp、15bp至55bp、15bp至50bp、15bp至45bp、15bp至40bp、15bp至35bp、15bp至30bp、15bp至25bp、15bp至20bp、20bp至60bp、20bp至55bp、20bp至50bp、20bp至45bp、20bp至40bp、20bp至35bp、20bp至30bp、20bp至25bp、25bp至60bp、25bp至55bp、25bp至50bp、25bp至45bp、25bp至40bp、25bp至35bp、25bp至30bp、30bp至60bp、30bp至55bp、30bp至50bp、30bp至45bp、30bp至40bp、30bp至35bp、35bp至60bp、35bp至55bp、35bp至50bp、35bp至45bp、35bp至40bp、40bp至60bp、40bp至55bp、40bp至50bp、40bp至45bp、45bp至60bp、45bp至55bp、45bp至50bp、50bp至60bp、50bp至55bp、或55bp至60bp。

[0066] 根据一具体实例,所述系统可以在靶核酸分子的核苷酸序列中形成1bp至60bp的核苷酸的插入缺失,例如,1bp至55bp、1bp至50bp、1bp至45bp、1bp至40bp、1bp至35bp、1bp至30bp、1bp至25bp、1bp至20bp、1bp至15bp、1bp至10bp、1bp至5bp、5bp至60bp、5bp至55bp、5bp至50bp、5bp至45bp、5bp至40bp、5bp至35bp、5bp至30bp、5bp至25bp、5bp至20bp、5bp至15bp、5bp至10bp、10bp至60bp、10bp至55bp、10bp至50bp、10bp至45bp、10bp至40bp、10bp至35bp、10bp至30bp、10bp至25bp、10bp至20bp、10bp至15bp、15bp至60bp、15bp至55bp、15bp至50bp、15bp至45bp、15bp至40bp、15bp至35bp、15bp至30bp、15bp至25bp、15bp至20bp、20bp至60bp、20bp至55bp、20bp至50bp、20bp至45bp、20bp至40bp、20bp至35bp、20bp至30bp、20bp至25bp、25bp至60bp、25bp至55bp、25bp至50bp、25bp至45bp、25bp至40bp、25bp至35bp、25bp至30bp、30bp至60bp、30bp至55bp、30bp至50bp、30bp至45bp、30bp至40bp、30bp至35bp、35bp至60bp、35bp至55bp、35bp至50bp、35bp至45bp、35bp至40bp、40bp至60bp、40bp至55bp、40bp至50bp、40bp至45bp、45bp至60bp、45bp至55bp、45bp至50bp、50bp至60bp、50bp至55bp、或55bp至60bp。

[0067] 根据一具体实例,所述系统形成核苷酸的插入缺失的效率可以是5%至50%,例如,5%至45%、5%至40%、5%至35%、5%至30%、5%至25%、5%至20%、5%至15%、5%至10%、10%至10%、50%、10%至45%、10%至40%、10%至35%、10%至30%、10%至25%、10%至20%、10%至15%、15%至50%、15%至45%、15%至40%、15%至35%、15%至30%、15%至25%、15%至20%、20%至50%、20%至45%、20%至40%、20%至35%、20%至30%、20%至25%、25%至50%、25%至45%、25%至40%、25%至35%、25%至30%、30%至50%、30%至45%、30%至40%、30%至35%、35%至50%、35%至45%、35%至40%、40%至50%、40%至45%、或45%到50%。

[0068] 根据一具体实例,所述系统形成核苷酸的置换的效率可以是1%至20%,例如,1%至18%、1%至16%、1%至14%、1%至12%、1%至10%、1%至8%、1%至6%、1%至4%、1%至2%、2%至20%、2%至18%、2%至16%、2%至14%、2%至12%、2%至10%、2%至8%、2%至6%、2%至4%、4%至20%、4%至18%、4%至16%、4%至14%、4%至12%、4%至10%、4%至8%、4%至6%、6%至20%、6%至18%、6%至16%、6%至14%、6%至12%、6%至10%、6%至8%、8%至20%、8%至18%、8%至16%、8%至14%、8%至12%、8%至10%、10%至20%、10%至18%、10%至16%、10%至14%、10%至12%、12%至20%、12%至18%、12%至16%、12%至14%、14%至20%、14%至18%、14%至16%、16%至20%、16%至18%、

或18%至20%。

[0069] 所述系统可以形成靶核酸(nucleic acid)分子的核苷酸(nucleotide)序列中至少4个核苷酸的编辑窗口(editing window)。根据一具体实例,所述系统可以具有至少50个核苷酸的编辑窗口,例如,至少49个核苷酸、至少48个核苷酸、至少47个核苷酸、至少46个核苷酸、至少45个核苷酸、至少44个核苷酸、至少43个核苷酸、至少42个核苷酸、至少41个核苷酸、至少40个核苷酸、至少39个核苷酸、至少38个核苷酸、至少37个核苷酸、至少36个核苷酸、至少35个核苷酸、至少34个核苷酸、至少33个核苷酸、至少32个核苷酸、至少31个核苷酸、至少30个核苷酸、至少29个核苷酸、至少28个核苷酸、至少27个核苷酸、至少26个核苷酸、至少25个核苷酸、至少24个核苷酸、至少23个核苷酸、至少22个核苷酸、至少21个核苷酸、至少20个核苷酸、至少19个核苷酸、至少18个核苷酸、至少17个核苷酸、至少16个核苷酸、至少15个核苷酸、至少14个核苷酸、至少13个核苷酸、至少12个核苷酸、至少11个核苷酸、至少10个核苷酸、至少9个核苷酸、至少8个核苷酸、至少7个核苷酸、至少6个核苷酸、至少5个核苷酸、或至少4个核苷酸。

[0070] 根据一具体实例,所述系统可以具有距gRNA靶碱基序列5`端1bp至20bp、1bp至19bp、1bp至18bp、1bp至17bp、1bp至16bp、1bp至15bp、1bp至14bp、1bp至13bp、1bp至12bp、1bp至11bp、1bp至10bp、1bp至9bp、1bp至8bp、2bp至20bp、2bp至19bp、2bp至18bp、2bp至17bp、2bp至16bp、2bp至15bp、2bp至14bp、2bp至13bp、2bp至12bp、2bp至11bp、2bp至10bp、2bp至9bp、2bp至8bp、3bp至20bp、3bp至19bp、3bp至18bp、3bp至17bp、3bp至16bp、3bp至15bp、3bp至14bp、3bp至13bp、3bp至12bp、3bp至11bp、3bp至10bp、3bp至9bp、3bp至8bp、4bp至20bp、4bp至19bp,4bp至18bp、4bp至17bp、4bp至16bp、4bp至15bp、4bp至14bp、4bp至13bp、4bp至12bp、4bp至11bp、4bp至10bp、4bp至9bp、或4bp至8bp的编辑窗口(editing window)。

[0071] 根据一具体实例,所述核苷酸编辑窗口(Editing window)可以显示距gRNA靶碱基序列5`端位于第1位的胞嘧啶(C_1)至位于第20位的胞嘧啶(C_{20})、C_2至C_{20}、C_3至C_{20}、C_4至C_{20}、C_1至C_{19}、C_2至C_{19}、C_3至C_{19}、C_4至C_{19}、C_1至C_{18}、C_2至C_{18}、C_3至C_{18}、C_4至C_{18}、C_1至C_{17}、C_2至C_{17}、C_3至C_{17}、C_4至C_{17}、C_1至C_{16}、C_2至C_{16}、C_3至C_{16}、C_4至C_{16}、C_1至C_{15}、C_2至C_{15}、C_3至C_{15}、C_4至C_{15}、C_1至C_{14}、C_2至C_{14}、C_3至C_{14}、C_4至C_{14}、C_1至C_{13}、C_2至C_{13}、C_3至C_{13}、C_4至C_{13}、C_1至C_{12}、C_2至C_{12}、C_3至C_{12}、C_4至C_{12}、C_1至C_{11}、C_2至C_{11}、C_3至C_{11}、C_4至C_{11}、C_1至C_{10}、C_2至C_{10}、C_3至C_{10}、C_4至C_{10}、C_1至C_9、C_2至C_9、C_3至C_9、C_4至C_9、C_1至C_8、C_2至C_8、C_3至C_8、或C_4至C_8范围的碱基编辑。

[0072] 又另一方案,提供了一种编辑核酸的方法,其包含将核酸分子与CRISPR-Cas系统接触的步骤。

[0073] 所述核酸、以及CRISPR-Cas系统,如上所述。

[0074] 在一具体实例中,所述编辑可以形成所述核酸分子的核苷酸序列中至少一个以上的核苷酸序列的缺失、插入、置换、或插入缺失。

[0075] 根据一具体实例,修饰所述核酸的方法中修饰可以是核酸分子的核苷酸序列中形成1bp至60bp的核苷酸的缺失,例如,1bp至55bp、1bp至50bp、1bp至45bp、1bp至40bp、1bp至35bp、1bp至30bp、1bp至25bp、1bp至20bp、1bp至15bp、1bp至10bp、1bp至5bp、5bp至60bp、5bp至55bp、5bp至50bp、5bp至45bp、5bp至40bp、5bp至35bp、5bp至30bp、5bp至25bp、5bp至20bp、5bp至15bp、5bp至10bp、10bp至60bp、10bp至55bp、10bp至50bp、10bp至45bp、10bp至40bp、10bp至35bp、10bp至30bp、10bp至25bp、10bp至20bp、10bp至15bp、15bp至60bp、15bp至55bp、

15bp至50bp、15bp至45bp、15bp至40bp、15bp至35bp、15bp至30bp、15bp至25bp、15bp至20bp、20bp至60bp、20bp至55bp、20bp至50bp、20bp至45bp、20bp至40bp、20bp至35bp、20bp至30bp、20bp至25bp、25bp至60bp、25bp至55bp、25bp至50bp、25bp至45bp、25bp至40bp、25bp至35bp、25bp至30bp、30bp至60bp、30bp至55bp、30bp至50bp、30bp至45bp、30bp至40bp、30bp至35bp、35bp至60bp、35bp至55bp、35bp至50bp、35bp至45bp、35bp至40bp、40bp至60bp、40bp至55bp、40bp至50bp、40bp至45bp、45bp至60bp、45bp至55bp、45bp至50bp、50bp至60bp、50bp至55bp、或55bp至60bp。

[0076] 根据一具体实例,修饰所述核酸的方法中修饰可以是核酸分子的核苷酸序列中形成1bp至60bp的核苷酸的插入,例如,1bp至55bp、1bp至50bp、1bp至45bp、1bp至40bp、1bp至35bp、1bp至30bp、1bp至25bp、1bp至20bp、1bp至15bp、1bp至10bp、1bp至5bp、5bp至60bp、5bp至55bp、5bp至50bp、5bp至45bp、5bp至40bp、5bp至35bp、5bp至30bp、5bp至25bp、5bp至20bp、5bp至15bp、5bp至10bp、10bp至60bp、10bp至55bp、10bp至50bp、10bp至45bp、10bp至40bp、10bp至35bp、10bp至30bp、10bp至25bp、10bp至20bp、10bp至15bp、15bp至60bp、15bp至55bp、15bp至50bp、15bp至45bp、15bp至40bp、15bp至35bp、15bp至30bp、15bp至25bp、15bp至20bp、20bp至60bp、20bp至55bp、20bp至50bp、20bp至45bp、20bp至40bp、20bp至35bp、20bp至30bp、20bp至25bp、25bp至60bp、25bp至55bp、25bp至50bp、25bp至45bp、25bp至40bp、25bp至35bp、25bp至30bp、30bp至60bp、30bp至55bp、30bp至50bp、30bp至45bp、30bp至40bp、30bp至35bp、35bp至60bp、35bp至55bp、35bp至50bp、35bp至45bp、35bp至40bp、40bp至60bp、40bp至55bp、40bp至50bp、40bp至45bp、45bp至60bp、45bp至55bp、45bp至50bp、50bp至60bp、50bp至55bp、或55bp至60bp。

[0077] 根据一具体实例,修饰所述核酸的方法中修饰可以是核酸分子的核苷酸序列中形成1bp至60bp的核苷酸的插入缺失,例如1bp至55bp、1bp至50bp、1bp至45bp、1bp至40bp、1bp至35bp、1bp至30bp、1bp至25bp、1bp至20bp、1bp至15bp、1bp至10bp、1bp至5bp、5bp至60bp、5bp至55bp、5bp至50bp、5bp至45bp、5bp至40bp、5bp至35bp、5bp至30bp、5bp至25bp、5bp至20bp、5bp至15bp、5bp至10bp、10bp至60bp、10bp至55bp、10bp至50bp、10bp至45bp、10bp至40bp、10bp至35bp、10bp至30bp、10bp至25bp、10bp至20bp、10bp至15bp、15bp至60bp、15bp至55bp、15bp至50bp、15bp至45bp、15bp至40bp、15bp至35bp、15bp至30bp、15bp至25bp、15bp至20bp、20bp至60bp、20bp至55bp、20bp至50bp、20bp至45bp、20bp至40bp、20bp至35bp、20bp至30bp、20bp至25bp、25bp至60bp、25bp至55bp、25bp至50bp、25bp至45bp、25bp至40bp、25bp至35bp、25bp至30bp、30bp至60bp、30bp至55bp、30bp至50bp、30bp至45bp、30bp至40bp、30bp至35bp、35bp至60bp、35bp至55bp、35bp至50bp、35bp至45bp、35bp至40bp、40bp至60bp、40bp至55bp、40bp至50bp、40bp至45bp、45bp至60bp、45bp至55bp、45bp至50bp、50bp至60bp、50bp至55bp、或55bp至60bp。

[0078] 根据一具体实例,修饰所述核酸的方法形成核苷酸序列的插入缺失的效率可以是5%至50%,例如,5%至45%、5%至40%、5%至35%、5%至30%、5%至25%、5%至20%、5%至15%、5%至10%、10%至50%、10%至45%、10%至40%、10%至35%、10%至30%、10%至25%、10%至20%、10%至15%、15%至50%、15%至45%、15%至40%、15%至35%、15%至30%、15%至25%、15%至20%、20%至50%、20%至45%、20%至40%、20%至35%、20%至30%、20%至25%、25%至50%、25%至45%、25%至40%、25%至35%、25%至30%、

30％至50％、30％至45％、30％至40％、30％至35％、35％至50％、35％至45％、35％至40％、40％至50％、40％至45％、或45％至50％。

[0079] 根据一具体实例,修饰所述核酸的方法形成核苷酸序列的置换的效率可以是1％至20％,例如,1％至18％、1％至16％、1％至14％、1％至12％、1％至10％、1％至8％、1％至6％、1％至4％、1％至2％、2％至20％、2％至18％、2％至16％、2％至14％、2％至12％、2％至10％、2％至8％、2％至6％、2％至4％、4％至20％、4％至18％、4％至16％、4％至14％、4％至12％、4％至10％、4％至8％、4％至6％、6％至20％、6％至18％、6％至16％、6％至14％、6％至12％、6％至10％、6％至8％、8％至20％、8％至18％、8％至16％、8％至14％、8％至12％、8％至10％、10％至20％、10％至18％、10％至16％、10％至14％、10％至12％、12％至20％、12％至18％、12％至16％、12％至14％、14％至20％、14％至18％、14％至16％、16％至20％、16％至18％、或18％至20％。

[0080] 修饰所述核酸的方法在靶核酸(nucleic acid)分子的核苷酸(nucleotide)序列中至少可以形成4个核苷酸的编辑窗口(editing window)。根据一具体实例,修饰所述核酸的方法至少具有50个核苷酸的编辑窗口,例如,至少49个核苷酸、至少48个核苷酸、至少47个核苷酸、至少46个核苷酸、至少45个核苷酸、至少44个核苷酸、至少43个核苷酸、至少42个核苷酸、至少41个核苷酸、至少40个核苷酸、至少39个核苷酸、至少38个核苷酸、至少37个核苷酸、至少36个核苷酸、至少35个核苷酸、至少34个核苷酸、至少33个核苷酸、至少32个核苷酸、至少31个核苷酸、至少30个核苷酸、至少29个核苷酸、至少28个核苷酸、至少27个核苷酸核苷酸、至少26个核苷酸、至少25个核苷酸、至少24个核苷酸、至少23个核苷酸、至少22个核苷酸、至少21个核苷酸、至少20个核苷酸、至少19个核苷酸、至少18个核苷酸、至少17个核苷酸、至少16个核苷酸、至少15个核苷酸、至少14个核苷酸、至少13个核苷酸、至少12个核苷酸、至少11个核苷酸、至少10个核苷酸、至少9个核苷酸、至少8个核苷酸、至少7个核苷酸、至少6个核苷酸、或至少5个核苷酸。

[0081] 根据一具体实例,修饰所述核酸的方法可以具有距gRNA靶碱基序列5`端1bp至20bp、1bp至19bp、1bp至18bp、1bp至17bp、1bp至16bp、1bp至15bp、1bp至14bp、1bp至13bp、1bp至12bp、1bp至11bp、1bp至10bp、1bp至9bp、1bp至8bp、2bp至20bp、2bp至19bp、2bp至18bp、2bp至17bp、2bp至16bp、2bp至15bp,2bp至14bp、2bp至13bp、2bp至12bp、2bp至11bp、2bp至10bp、2bp至9bp、2bp至8bp、3bp至20bp、3bp至19bp、3bp至18bp、3bp至17bp、3bp至16bp、3bp至15bp、3bp至14bp、3bp至13bp、3bp至12bp、3bp至11bp、3bp至10bp、3bp至9bp、3bp至8bp、4bp至20bp、4bp至19bp、4bp至18bp、4bp至17bp、4bp至16bp、4bp至15bp、4bp至14bp、4bp至13bp、4bp至12bp、4bp至11bp、4bp至10bp、4bp至9bp、或4bp至8bp的编辑窗口(editing window)。

[0082] 根据一具体实例,修饰所述核算的方法的编辑窗口(Editing window)可以显示距gRNA靶碱基序列5`端位于第一位的胞嘧啶(C_1)至位于第20位的胞嘧啶(C_{20})、C_2至C_{20}、C_3至C_{20}、C_4至C_{20}、C_1至C_{19}、C_2至C_{19}、C_3至C_{19}、C_4至C_{19}、C_1至C_{18}、C_2至C_{18}、C_3至C_{18}、C_4至C_{18}、C_1至C_{17}、C_2至C_{17}、C_3至C_{17}、C_4至C_{17}、C_1至C_{16}、C_2至C_{16}、C_3至C_{16}、C_4至C_{16}、C_1至C_{15}、C_2至C_{15}、C_3至C_{15}、C_4至C_{15}、C_1至C_{14}、C_2至C_{14}、C_3至C_{14}、C_4至C_{14}、C_1至C_{13}、C_2至C_{13}、C_3至C_{13}、C_4至C_{13}、C_1至C_{12}、C_2至C_{12}、C_3至C_{12}、C_4至C_{12}、C_1至C_{11}、C_2至C_{11}、C_3至C_{11}、C_4至C_{11}、C_1至C_{10}、C_2至C_{10}、C_3至C_{10}、C_4至C_{10}、C_1至C_9、C_2至C_9、C_3至C_9、C_4至C_9、C_1至C_8、C_2至C_8、C_3至C_8、或C_4至C_8范围的碱基编辑。

[0083] 发明效果

[0084] 根据一方案,融合蛋白、其多肽、以及包含其的CRISPR-Cas系统,可以进行有效的碱基编辑。此外,由于多肽的尺寸小,易于与Cas蛋白结合,当用作细胞治疗剂时,具有促进通过载体递送的效果。

[0085] 另外,当通过此诱导插入缺失时,插入缺失效率和形成插入缺失的核苷酸的大小也增加,从而可以有效地用于基因的敲除(knock-out)。

附图说明

[0086] 图1是显示SsdA与现有碱基编辑技术中使用的其他脱氨酶的结构/进化差异的示意图。

[0087] 图2是显示用SsdA处理具有SEQ ID NO:4的碱基序列的单链DNA使其脱氨基,并用UDG和NaOH处理切割单链DNA过程的图,以及显示确认SsdA脱氨基与否的蛋白质免疫印迹结果的照片。

[0088] 图3是显示包含Cas蛋白、SsdA和向导多核苷酸的CRISPR-Cas系统在靶DNA的靶位点的碱基编辑效率的图:图3a是显示包含Cas蛋白、SsdA和向导多核苷酸的CRISPR-Cas系统在RNF2 DNA的靶位点的碱基编辑效率的图,图3b是显示包含Cas蛋白、SsdA和向导多核苷酸的CRISPR-Cas系统在HEK2 DNA的靶位点的碱基编辑效率的图。

[0089] 图4是分别显示Cas9(D10A)-SsdA融合蛋白、Cas9(D10A)-SsdA-UGI融合蛋白、dCas9-SsdA融合蛋白、以及dCas9-SsdA-UGI融合蛋白的细胞毒性的图。

[0090] 图5是显示包含Cas9和SsdA融合蛋白的CRISPR-Cas系统的碱基编辑效率和插入缺失形成效率的图:

[0091] 图5a是显示包含Cas9(D10A)-SsdA融合蛋白、Cas9(D10A)-SsdA-UGI融合蛋白、dCas9-SsdA融合蛋白、或dCas9-SsdA-UGI融合蛋白的每个CRISPR-Cas系统在HEK2、HEK3、HEK4、以及RNF2中的碱基编辑效率和插入缺失形成效率的图,图5b是显示包含cjCas9(D8A)-SsdA融合蛋白、cjCas9(D8A)-SsdA-UGI融合蛋白、或cjCas9(L58Y/D900K)(D8A)-SsdA-UGI融合蛋白的每个CRISPR-Cas系统在EPAS1_e2、EPAS1_e5、HIF_e8、HIF_e9、以及TFPi中的碱基编辑效率的图,图5c是显示在尿嘧啶-DNA糖基化酶(UDG)表达抑制细胞系中,包含cjCas9(D8A)-SsdA融合蛋白、cjCas9(D8A)-SsdA-UGI融合蛋白、或cjCas9(L58Y/D900K)(D8A)-SsdA-UGI融合蛋白的每个CRISPR-Cas系统在EPAS1_e2、EPAS1_e5、HIF_e8、HIF_e9、以及TFPi中的碱基编辑效率的图。

[0092] 图6是显示用于制备SsdA和尿嘧啶糖基化酶抑制剂(UGI)结合至Cas蛋白的C端、N端、以及N端和C端两端的融合蛋白的质粒的示意图。

[0093] 图7是显示根据SsdA和Cas蛋白的结合位点确认的碱基编辑效率的图:

[0094] 图7a是显示包含Cas9(D10A)-SsdA融合蛋白(SsdA-C)、SsdA-Cas9(D10A)融合蛋白(SsdA-N、SsCBE)、或SsdA-Cas9(D10A)-SsdA(SsdA-NC)融合蛋白的每个CRISPR-Cas系统在HEK2、HEK3、HEK4、RNF2、FANCF、TYRO3、CCR5、或EMX1中的胞嘧啶碱基编辑效率的图,图7b是显示包含SsdA-Cas9(D10A)融合蛋白(SsdA-N、SsCBE)的C端、N端、以及C端和N端两端结合1个或2个UGI的融合蛋白的每个CRISPR-Cas系统在HEK2、HEK3、HEK4、RNF2、FANCF、TYRO3、CCR5、或EMX1中的胞嘧啶碱基编辑效率的图,图7c是显示为了准确比较碱基编辑效率,在N

端结合两个碱基编辑效率明显高的UGI的UGI-UGI-SsdA-Cas9(D10A)融合蛋白(SsCBE-UGI-N2)或在N端、C端两端各结合两个UGI的UGI-UGI-SsdA-Cas9(D10A)-UGI-UGI融合蛋白(SsCBE-UGI-N2C2),在20个靶位点(HEK2-1、HEK2-2、HEK2-3、HEK2-4、HEK3-1、HEK3-2、HEK3-3、HEK3-6、HEK3-7、HEK3-8、HEK4-1、HEK4-2、HEK4-3、HEK4-4、HEK4-5、HEK4-6、HEK4-7、HEK4-8、RNF2-3、RNF2-4)中的胞嘧啶碱基编辑效率的图,图7d是针对图7c所示的20个靶位点的碱基编辑,对碱基编辑技术的特征即编辑范围(编辑窗口)进行分析的图。

[0095] 图8是显示UGI-UGI-SsdA-Cas9(D10A)融合蛋白(SsCBE-UGI-N2)、BE3和BE4max在靶位点(HEK2、HEK3、HEK2-2、HEK3-8、HEK4-2、或HEK4-7位点)的置换频率(substitution frequency)(%)的图。

[0096] 图9是显示包含Cas9(D10A)-SsdA(G54D)-UGI融合蛋白、Cas9(D10A)-SsdA(G54D)融合蛋白、dCas9-SsdA(G54D)-UGI融合蛋白、或dCas9-SsdA(G54D)融合蛋白的每个CRISPR-Cas系统在HEK2、HEK3、以及HEK4中的碱基编辑效率和插入缺失形成效率的图表。

具体实施方式

[0097] 通过以下实施例进行更详细地描述。但是,这些实施例仅用于示例性说明,本发明的范围不局限于这些实施例。

[0098] 实施例1质粒克隆(plasmid cloning)

[0099] 将编码具有SEQ ID NO:1的氨基酸序列的His6-SsdA与具有SEQ ID NO:3的氨基酸序列的SsdAI的gBlock双链DNA片段(Integrated DNA Technologies)和细菌表达载体pET-28b(+)DNA(Novagen)分别用XbaI和XhoI限制酶(New England Biolabs)在37℃下处理3小时。然后,使用琼脂糖凝胶提取(Geneall)来纯化线性化的pET-28b载体,并使用快速连接酶(New England Biolabs)将其连接到所述gBlock双链DNA片段。

[0100] 然后,使用pCMV质粒DNA,通过PCR扩增获得Cas9和UGI的编码序列(coding sequence),使用Gibson Assembly Master Mix(New England Biolabs)在gBlock中扩增SsdA序列,并亚克隆到pCMV质粒中。

[0101] SEQ ID NO:1的含PAAR结构域的蛋白质氨基酸序列和SEQ ID NO:3的SsdAI氨基酸序列示于表2。

[0102] 表2

[0103]

	氨基酸序列	SEQ ID NO
SsdA 蛋白(含 PAAR 结构域	MSAAARVNDPIEHTGSLTGLLAGLAIGAIGAALV VGTGGLAAVAIVGASAATGAGVGQLIGSLSCCNH QTGQIVSGSSNVYINGEPAARAHADQAKCDEHTS RPQVIAQGSSNVYINGHPAARVGDRTACDAKIVV	SEQ ID NO: 1

[0104]

的蛋白质）	GSSNVFIGGGTETTDPINPEVPELLERSILLVGLAS AVVLASPVIVIAGLVGGIAGGTVGSMGGAQLFGE GTDGQKLMAFGGALLGGGLGAKGGKWFDTRY DIKVQGVGSNLGNLKITPKGAAKVSNIAESEAAL GRASQARADLPQSKELKVKTVSSNDKKTLSGWG NKKPEGYERISAEQVKAKSEEIGHEVKSHPYDRD YKGQYFSSHAEKQMSIASPNHPLGVSKPMCTDC QGYFSQLAKYSKVEQTVADPKAIRIFKTDGSVETI MRSE	
SsdA 蛋白	MNNKSKVLIEKLLLEVAKSPEGELILPLRKLLWN TITEDETAAKKKAILTALDVMCVRQGVNFWIKKF GDNEPLNYILNIALETAEGKFDESKALGLRDEFY VSIVEDQEYEVEEYPAMFVGHAAANTIARAVDDF QFEPYDHRVDRDLDPEGFESSYLVASAFAGGLSE DGDPKLRRAFWEWYLSIAVPQVV	SEQ ID NO：3

[0105] 实施例2SsdA的纯化(purification)

[0106] 使用大肠杆菌BL21纯化实施例1中克隆的SsdA蛋白。

[0107] 更具体地,使用0.5mM IPTG将pET-28b-His6-SsdA-SsdAI施用于大肠杆菌BL21,并使用Ni-NTA琼脂糖珠(Qiagen)纯化His6-SsdA和SsdAI复合蛋白。然后为了从SsdAI中分离His6-SsdA,用变性缓冲液(8M尿素、50mM Tris-HCl pH7.5、500mM NaCl和1mM DTT)使His6-SsdA和SsdAI复合蛋白变性后,在4℃下孵育16小时。将包含变性蛋白质复合物的悬浮缓冲液与Ni-NTA琼脂糖珠(Qiagen)混合并加载到重力流柱(gravity-flow column)上,以去除未结合的SsdAI。随后,用浓度递减的尿素(6M、4M、2M、1M和0M),处理变性缓冲液以事先为了SsdA的重折叠。用含有300mM咪唑的洗脱缓冲液洗脱与Ni-NTA琼脂糖珠结合的重折叠蛋白后,并将洗脱的蛋白质对20mM Tris-HCl pH7.5、200mM NaCl、1mM DTT、以及40％(w/v)甘油进行透析,并使用Amicon Ultra-15离心过滤装置(Millipore)进行浓缩。然后通过SDS-PAGE分析His6-SsdA蛋白的浓度。

[0108] 实施例3细胞培养和转染(transfection)

[0109] 将HEK293T细胞(ATCC CRL-11268)保存在包含10％胎牛血清(fetal bovine serum:FBS)和1％青霉素/链霉素(Welgene)的杜氏改良伊格尔培养基(Dulbecco's modified Eagle's medium(DMEM))中,然后将所述HEK293T细胞以每孔密度$6x10^4$个细胞接种到经TC处理的48孔板(Corning Life Sciences)中。接种后24小时,在细胞汇合度约60％下,使用500ng质粒(250ng Cas9-SsdA表达质粒和250ng gRNA表达质粒)和1.5uL Lipofectamine 2000(Thermo Fisher Scientific)进行转染。然后,将转染的细胞在37℃下孵育3天,并使用裂解缓冲液(pH7.5的10mM Tris-HCl,0.05％ SDS,100mg/mL蛋白酶K;QIAGEN)直接裂解细胞,并制备基因组DNA。将所述细胞裂解物在56℃下孵育30分钟,并进一步在99℃下孵育15分钟以灭活蛋白酶K。

[0110] 实施例4靶向深度测序(deep sequencing)和数据分析

[0111] 通过PCR扩增了靶位点(总共2或3次),并通过Illumina MiniSeq或iSeq100测序系统进行测序。

[0112] 更具体地说,将3mL细胞裂解物或1mL分离的基因组DNA应用于第一次PCR后,将1mL第一次PCR产物用于第二次PCR。使用1mL第二次PCR产物将Illumina TruSeq HT双标签接头序列附着至标签PCR引物对。在2%琼脂糖凝胶上确认PCR扩增子的大小,并使用Illumina MiniSeq或iSeq100测序系统对扩增子进行测序。使用MAUND(https://github.com/ibscge/maund)进行靶向深度测序分析,所有结果均通过Cas-Analyzer(http://www.rgenome.net/cas-analyzer/)进行确认。

[0113] 实施例5确认SsdA的单链DNA中胞嘧啶的脱氨基

[0114] 为了确认SsdA是否将单链DNA的胞嘧啶碱基转化为尿嘧啶,用SsdA处理含有FAM的单链DNA。

[0115] 更具体地,将具有SEQ ID NO:4的碱基序列(5`-Aaaaaaaaaaaaaaaagcgaaaaaaaaaaaaaaaaaa-3`)的单链DNA用1至200nM的SsdA在37℃下处理1小时后,用尿嘧啶DNA糖基化酶(UDG)在37℃条件下处理30分钟,除去变为尿嘧啶的DNA碱基,从而生成脱碱基位点(abasic site)。然后,用100mM NaOH处理并在95℃条件下孵育2分钟以切割脱碱基位点。并且通过蛋白质免疫印迹确认其切割,结果如图2所示。

[0116] 图2是显示用SsdA处理具有SEQ ID NO:4的碱基序列的单链DNA使其脱氨基,用UDG和NaOH处理并切割单链DNA过程的图,以及显示确认SsdA脱氨基与否的蛋白质免疫印迹结果的照片。

[0117] 如图2所示,确认只有用SsdA、UDG和NaOH处理单链DNA时才会发生DNA切割。这意味着,SsdA可以通过脱氨基有效地将单链DNA的胞嘧啶转化为尿嘧啶。

[0118] 实施例6确认包含Cas蛋白、SsdA和向导多核苷酸的CRISPR-Cas系统的碱基编辑效率

[0119] 为了确认包含Cas蛋白、SsdA和向导多核苷酸的CRISPR-Cas系统的碱基编辑效率,用包含SsdA、dCas9和gRNA的CRISPR-Cas系统处理靶DNA(RNF2和HEK2)确定胞嘧啶的脱氨基形成与否。所述靶DNA(RNF2和HEK2)的碱基序列示于表3。

[0120] 表3

[0121]

靶DNA	靶位点(target site)	SEQ ID NO
RNF2	GTC3ATC6TTAGTC12ATTACCTGAGG	SEQ ID NO:5
HEK2	GAAC4AC6AAAGC11ATAGACTGCGGG	SEQ ID NO:6

[0122] 更具体地,分别用100nM Cas9、300nM sgRNA、以及40nM SsdA处理RNF2 DNA和HEK2 DNA后,在37℃下孵育8小时,以诱导靶DNA的靶位点胞嘧啶转化为尿嘧啶。然后,8小时后,通过RNase和Protease K处理去除sgRNA、Cas9、以及SsdA。并且利用Qiagen DNA提取试剂盒(DNAextraction kit)纯化DNA(DNA purification)。随后,使用包含靶位点的引物(primer)进行PCR,并通过深度测序(Deep sequencing)测量碱基编辑效率,结果如图3所示。

[0123] 图3是显示包含Cas蛋白、SsdA和向导多核苷酸的CRISPR-Cas系统在靶DNA的靶位点的碱基编辑效率的图:

[0124] 图3a是显示包含Cas蛋白、SsdA和向导多核苷酸的CRISPR-Cas系统在RNF2 DNA的

靶位点的碱基编辑效率的图,图3b是显示包含Cas蛋白、SsdA和向导多核苷酸的CRISPR-Cas系统在HEK2 DNA的靶位点的碱基编辑效率的图。

[0125] 如图3a所示,当使用靶向RNF2的gRNA时,确认胞嘧啶转化为尿嘧啶的效率在C_3处约为7%,在C_{12}处约为11%。此外,如图3b所示,当使用靶向HEK2的gRNA时,确认在C_4、C_6、以及C_{11}处胞嘧啶转化为尿嘧啶效率显示约为18%至19%。

[0126] 这意味着,所述CRISPR-Cas系统只能对DNA双链的一条链进行基因编辑,因此,考虑到最大碱基编辑效率为50%,这一结果显示了相当高水平的碱基编辑效率。

[0127] 此外,进一步地,为了确认包含所述Cas蛋白、SsdA和向导多核苷酸的CRISPR-Cas系统引起的碱基转化是否是胞嘧啶转化为尿嘧啶,用尿嘧啶特异性切除酶尿嘧啶特异性切除试剂(Uracil-Specific Excision Reagent(USER))处理脱氨基DNA,并通过PCR扩增靶位点。结果确认胞嘧啶转化为尿嘧啶的读取数消失。

[0128] 这些结果意味着,所述CRISPR-Cas系统可以有效地将胞嘧啶转化为尿嘧啶。

[0129] 实施例7确认包含Cas蛋白和SsdA的融合蛋白的细胞毒性

[0130] 确认包含Cas蛋白、SsdA和UGI的融合蛋白的细胞毒性,其结果示于图4。

[0131] 为了确认CRISPR-Cas系统的SsdA在真核细胞中是否有毒性,构建表达与Cas9(D10A)结合的蛋白质的质粒。然后,将HEK293细胞以6×10^4细胞/孔的浓度分装至48孔板(well plate)后,将所述质粒转染(transfection)至HEK293细胞中。转染48小时后,将活细胞用胰蛋白酶(trypsinization)消化并利用血细胞计数器(hemacytometer)计数HEK293细胞的数量。

[0132] 图4是显示Cas9(D10A)-SsdA融合蛋白、Cas9(D10A)-SsdA-UGI融合蛋白、dCas9-SsdA融合蛋白、以及dCas9-SsdA-UGI融合蛋白各自的细胞毒性的图。

[0133] 如图4所示,确认包含Cas9、SsdA和UGI的融合蛋白几乎没有细胞毒性。

[0134] 实施例8确认包含Cas蛋白和SsdA的融合蛋白的CRISPR-Cas系统的碱基编辑效率和插入缺失形成效率

[0135] 为了确认CRISPR-Cas系统的SsdA在真核细胞中是否有效引起胞嘧啶脱氨基,构建了表达与Cas9结合的蛋白质的质粒,并转染(transfection)至HEK293细胞中。然后,通过PCR扩增转染细胞的靶位点(HEK2、HEK3、HEK4、RNF2、EPAS1_e2、EPAS1_e5、HIF_e8、HIF_e9、或TFPi位点)的碱基序列,并利用下一代测序法(NGS)来分析突变的引入与否和插入缺失形成与否,结果如图5所示。

[0136] 使用的Cas9是spCas9(Cas9(D10A))或cjCas9(D8A)。尿嘧啶-DNA糖基化酶(Uracil-DNAglycosylase(UDG))敲低细胞系(UNG KD)是通过表达敲低UDG的shRNA(5`-GTCTACAGACATAGAGGATTT-3:SEQ ID NO:7)的质粒和基于HIV的包装质粒(含有基因:Gag/Pol、Rev、VSV-G)与Opti MEM(Invitrogen)中的Lipofectamine3000(Invitrogen)试剂混合而转染HEK293细胞生成的。

[0137] 图5是显示包含Cas9和SsdA融合蛋白的CRISPR-Cas系统的碱基编辑效率和插入缺失形成效率的图:

[0138] 图5a是显示包含Cas9(D10A)-SsdA融合蛋白、Cas9(D10A)-SsdA-UGI融合蛋白、dCas9-SsdA融合蛋白、或dCas9-SsdA-UGI融合蛋白的每个CRISPR-Cas系统在HEK2、HEK3、HEK4、以及RNF2中的碱基编辑效率和插入缺失形成效率的图,图5b是显示包含cjCas9

(D8A)-SsdA融合蛋白、cjCas9(D8A)-SsdA-UGI融合蛋白、或cjCas9(L58Y/D900K)(D8A)-SsdA-UGI融合蛋白的每个CRISPR-Cas系统在EPAS1_e2、EPAS1_e5、HIF_e8、HIF_e9、以及TFPi中的碱基编辑效率的图,图5c是显示在尿嘧啶-DNA糖基化酶(UDG)敲低细胞系中,包含cjCas9(D8A)-SsdA融合蛋白、cjCas9(D8A)-SsdA-UGI融合蛋白、或cjCas9(L58Y/D900K)(D8A)-SsdA-UGI融合蛋白的每个CRISPR-Cas系统在EPAS1_e2、EPAS1_e5、HIF_e8、HIF_e9、以及TFPi中的碱基编辑效率的图。

[0139] 如图5a所示,Cas9(D10A)-SsdA-UGI的情况下,确认在HEK2、HEK3、HEK4、以及RNF2位点中均发生4%至11%的胞嘧啶碱基编辑,并且在在靶位点形成碱基序列的插入缺失。此外,在dCas9-SsdA-UGI的情况下,确认在1%至3.5%水平发生胞嘧啶碱基编辑,并且未形成插入缺失。

[0140] 如图5b所示,cjCas9(D8A)-SsdA-UGI和cjCas9(L58Y/D900K)(D8A)-SsdA-UGI的情况,均引起约3%的碱基编辑,并与未结合UGI的融合蛋白相比,显示出提高的碱基编辑效率。

[0141] 如图5c所示,在UDG敲低细胞系中,cjCas9(D8A)-SsdA-UGI和cjCas9(L58Y/D900K)(D8A)-SsdA-UGI的情况下,均显示出约15%的碱基编辑效率,与未结合UGI的cjCas9(D8A)-SsdA融合蛋白相比,显示出显著提高的碱基编辑效率。

[0142] 表4是显示由Cas9(D10A)-Ssda-UGI和HEK2靶gRNA产生的碱基编辑的序列的分析结果。

[0143] 表4

[0144]

序列	读数	SEQ ID NO
GGAAACTGGAACACAAAGCATAGACTG CGGGGCGGGCCA	3,912 (WT)	SEQ ID NO: 8
GGAAACTGGAACATAAAGCATAGACTG CGGGGCGGGCCA	237	SEQ ID NO: 9
GGAAACTGGAATATAAAGCATAGACTGC GGGGCGGGCCA	103	SEQ ID NO: 10
GGAAACTGGAATACAAAGCATAGACTG CGGGGCGGGCCA	61	SEQ ID NO: 11
GGAAACTGGAACACAAAGTATAGACTG CGGGGCGGGCCA	29	SEQ ID NO: 12

[0145] 实施例9确认根据SsdA和Cas蛋白的结合位点的碱基编辑效率

[0146] 尿嘧啶-DNA糖基化酶(UDG)是一种众所周知的修复细胞内DNA中产生的胞嘧啶脱氨基的蛋白质。为了准确确认SsdA确实导致细胞内DNA中胞嘧啶脱氨基,首先制作了敲除UDG的HEK293细胞系(HEK293 UDG-KO)。

[0147] 为了确认SsdA与Cas蛋白的C端、N端、以及N端和C端两端的融合蛋白的碱基编辑效率和插入缺失形成效率,通过改变SsdA的位置来制备融合蛋白,并由此确认碱基编辑效率和插入缺失形成效率。

[0148] 更具体地,构建如图6所示的pCMV质粒,并将其与根据实施例3的每个靶位点的

gRNA一起引入HEK293 UDG-KO细胞中。然后,通过PCR扩增转染细胞的靶位点(HEK2、HEK3、HEK4、RNF2、FANCF、TYRO3、CCR5、或EMX1位点)的碱基序列,并利用下一代测序法(NGS)分析突变的引入和插入缺失形成与否,其结果示于图7。

[0149] 图6是显示用于制备SsdA和尿嘧啶糖基化酶抑制剂(UGI)结合至Cas蛋白的C端、N端、以及N端和C端两端的融合蛋白的质粒的示意图。

[0150] 图7是显示根据SsdA和Cas蛋白的结合位点确认的碱基编辑效率的图:

[0151] 图7a是显示包含Cas9(D10A)-SsdA融合蛋白(SsdA-C)、SsdA-Cas9(D10A)融合蛋白(SsdA-N、SsCBE)、或SsdA-Cas9(D10A)-SsdA(SsdA-NC)融合蛋白的每个CRISPR-Cas系统在HEK2、HEK3、HEK4、RNF2、FANCF、TYRO3、CCR5、或EMX1中的胞嘧啶碱基编辑效率的图,图7b是显示包含SsdA-Cas9(D10A)融合蛋白(SsdA-N、SsCBE)的C端、N端、以及C端和N端两端结合1个或2个UGI的融合蛋白的每个CRISPR-Cas系统在HEK2、HEK3、HEK4、RNF2、FANCF、TYRO3、CCR5、或EMX1中的胞嘧啶碱基编辑效率的图,图7c是显示为了准确比较碱基编辑效率,在N端结合两个碱基编辑效率明显高的UGI的UGI-UGI-SsdA-Cas9(D10A)融合蛋白(SsCBE-UGI-N2)或在N端、C端两端各结合两个UGI的UGI-UGI-SsdA-Cas9(D10A)-UGI-UGI融合蛋白(SsCBE-UGI-N2C2),分别在20个靶位点(HEK2-1、HEK2-2、HEK2-3、HEK2-4、HEK3-1、HEK3-2、HEK3-3、HEK3-6、HEK3-7、HEK3-8、HEK4-1、HEK4-2、HEK4-3、HEK4-4、HEK4-5、HEK4-6、HEK4-7、HEK4-8、RNF2-3、RNF2-4)中的胞嘧啶碱基编辑效率的图,图7d是针对图7c所示的20个靶位点的碱基编辑,对碱基编辑技术的特征即编辑范围(编辑窗口)来分析的图。

[0152] 所述靶基因的靶位点的碱基序列示于表5。

[0153] 表5

[0154]

靶 DNA	靶位点 (taregt site)	SEQ ID NO
HEK2	GAACACAAAGCATAGACTGC	18
HEK2-1	CCAGCCCGCTGGCCCTGTAA	19
HEK2-2	GCTGGCCCTGTAAAGGAAAC	20
HEK2-3	GTTTCCTTTACAGGGCCAGC	21
HEK2-4	GCACTTGTTTGCAGCTATTC	22

[0155]

HEK3	GGCCCAGACTGAGCACGTGA	23
HEK3-1	CTGCTTCTCCAGCCCTGGCC	24
HEK3-2	CCCTGGCCTGGGTCAATCCT	25
HEK3-3	GACTGAGCACGTGATGGCAG	26
HEK3-6	CTTCCTCCAGAGGGCGTCGC	27
HEK3-7	CAGGACAGCTTTTCCTAGAC	28
HEK3-8	CAGCTCCTGCACCGGGATAC	29
HEK4	GGCACTGCGGCTGGAGGTGG	30
HEK4-1	GGGGCACCGCGGCGCCCCGG	31
HEK4-2	GCGGCGCCCCGGTGGCACTG	32
HEK4-3	CGCCCCGGTGGCACTGCGGC	33
HEK4-4	TCCCTTCCTTCCACCCAGCC	34
HEK4-5	CCCTGCCTGTCATCCTGCTT	35
HEK4-6	GCAGTGCCACCGGGGCGCCG	36
HEK4-7	CTCCAGCCGCAGTGCCACCG	37
HEK4-8	ACCTCCAGCCGCAGTGCCAC	38
RNF2	GTCATCTTAGTCATTACCTG	39
RNF2-3	TACACGTCTCATATGCCCCT	40
RNF2-4	TCAACCATTAAGCAAAACAT	41
EMX1	GTCACCTCCAATGACTAGGG	42
FANCF	GGAATCCCTTCTGCAGCACC	43
TYRO3	GGCCACACTAGCGTTGCTGC	44
CCR5	TGACATCAATTATTATACAT	45

[0156] 如图7a所示,在SsdA结合至Cas的C端的SsdA-Cas9(D10A)(SsdA-C)的情况下,确认与SsdA结合至N端或N端和C端两端的情况相比,胞嘧啶碱基编辑效率略有降低。另一方面,在SsdA结合至Cas的N端的情况(SsdA-N、SsCBE)下,确认碱基编辑效率提升。

[0157] 如图7b所示,在碱基编辑效率高的SsdA-Cas9(D10A)(SsdA-N,SsCBE)中进一步结合UGI时,与不存在UGI时相比,碱基编辑效率提高,在N端结合两个UGI的UGI-UGI-SsdA-Cas9(D10A)融合蛋白(SsCBE-UGI-N2)或在N端、C端两端各结合两个UGI的UGI-UGI-SsdA-Cas9(D10A)-UGI-UGI融合蛋白(SsCBE-UGI-N2C2)的情况下,确认碱基编辑效率显著提高。

[0158] 如图7c所示,在N端结合两个碱基编辑效率最高的UGI的UGI-UGI-SsdA-Cas9(D10A)融合蛋白(SsCBE-UGI-N2);或在N端、C端两端各结合两个UGI的UGI-UGI-SsdA-Cas9(D10A)-UGI-UGI融合蛋白(SsCBE-UGI-N2C2);在20个靶位点比较的结果显示,在N端结合两个UGI的UGI-UGI-SsdA-Cas9(D10A)融合蛋白(SsCBE-UGI-N2);确认显示出略高的碱基编辑效率。

[0159] 如图7d所示,在N端结合两个碱基编辑效率最高的UGI的UGI-UGI-SsdA-Cas9(D10A)融合蛋白(SsCBE-UGI-N2);或其在N端、C端两端各结合两个UGI的UGI-UGI-SsdA-Cas9(D10A)-UGI-UGI融合蛋白(SsCBE-UGI-N2C2);在20个靶位点比较编辑窗口的结果显

示,在距gRNA靶碱基序列5`端4bp至8bp之间的位点,确认显示出高的碱基编辑效率。

[0160] 实施例10确认在野生型HEK293细胞中利用UGI-UGI-SsdA-Cas9(D10A)融合蛋白(SsCBE-UGI-N2)的碱基编辑能力

[0161] 确认已确认到具有最高的碱基编辑能力的UGI-UGI-SsdA-Cas9(D10A)融合蛋白(SsCBE-UGI-N2)是否在野生型HEK293细胞中也显示碱基编辑效率。

[0162] 更具体地,根据实施例3,将UGI-UGI-SsdA-Cas9(D10A)融合蛋白(SsCBE-UGI-N2);和现有的BE3或BE4max与各靶位点的gRNA一起引入HEK293细胞中。然后,通过PCR扩增转染细胞中的靶位点(HEK2、HEK3、HEK2-2、HEK3-8、HEK4-2、或HEK4-7位点)的碱基序列,并利用下一代测序法(NGS)分析突变的引入和插入缺失形成与否,其结果示于图8。

[0163] 图8是显示UGI-UGI-SsdA-Cas9(D10A)融合蛋白(SsCBE-UGI-N2)、BE3和BE4max在靶位点(HEK2、HEK3、HEK2-2、HEK3-8、HEK4-2、或HEK4-7位点)处的置换频率(%)的图。

[0164] 如图8所示,UGI-UGI-SsdA-Cas9(D10A)融合蛋白(SsCBE-UGI-N2)成功在六个靶位点(HEK2、HEK3、HEK2-2、HEK3-8、HEK4-2、或HEK4-7位点)显示出碱基编辑能力。

[0165] 实施例11确认利用SsdA突变体的CRISPR-Cas系统的细胞毒性和碱基编辑能力

[0166] 为了确认SsdA突变体的细胞毒性和碱基编辑能力,制作失活的SsdA,并确认其效果。

[0167] 更具体地,将G54D突变引入SsdA的催化活性位点后,所述突变的SsdA蛋白与Cas9结合,并确认其细胞内功能。其结果示于图9。SsdA的催化活性位点的G54D突变是SEQ ID NO:1的氨基酸序列(包含PAAR结构域的蛋白质)中的G302D突变。具体地,所述SsdA催化活性位点的G54D突变可以具有SEQ ID NO:17的序列。所述SsdA的催化活性位点G54D突变的氨基酸序列示于表6。

[0168] 表6

[0169]

	氨基酸序列	SEQ ID NO
SsdA 的催化活性位点的 G54D 突变序列	KVSNIAESEAALGRASQARADLPQSKELKV KTVSSNDKKTLS**D**WGNKKPEGYERISAEQ VKAKSEEIGHEVKSHPYDRDYKGQYFSSHA EKQMSIASPNHPLGVSKPMCTDCQGYFSQL AKYSKVEQTVADPKAIRIFKTDGSVETIMRS E	17

[0170] 图9是显示包含Cas9(D10A)-SsdA(G54D)-UGI融合蛋白、Cas9(D10A)-SsdA(G54D)融合蛋白、dCas9-SsdA(G54D)-UGI融合蛋白、或dCas9-SsdA(G54D)融合蛋白的每个CRISPR-Cas系统在HEK2、HEK3、以及HEK4中的碱基编辑效率和插入缺失形成效率的图表。

[0171] 如图9所示,已确认SsdA的G54D突变与Cas9(D10A)结合时,无论是否存在UGI,都会以约18%至34%的非常高的效率发生插入缺失。这意味着,无论UGI功能如何,SsdA(G54D)都可以在靶位点以更高效率引起插入缺失。

[0172] 此外,当使用SsdA(G54D)时,确认不仅细胞内毒性显著降低,而且缺失的碱基序列的大小显著增加。

[0173] 表7是显示由SsdA(G54D)-Cas9(D10A)产生的碱基编辑序列的分析结果。

[0174] 如表7所示,对SsdA(G54D)-Cas9(D10A)产生的插入缺失序列进行分析的结果可以确认,与通常形成1bp缺失的Cas9不同,在SsdA(G54D)-Cas9(D10A)的情况下,缺失的碱基序列的大小显著增加。这些结果意味着,包含SsdA(G54D)的CRISPR-Cas系统与仅使用Cas9的系统相比,可以提高基因敲除效率。

[0175] 表7

[0176]

序列	读数	SEQ ID NO
AATTTTCCAGCCCGCTGGCCCTGTAAA GGAAACTGGAACACAAAGCATAGACT GCGGGGCGGGCCAGCCTGAATA	1,010 (WT)	SEQ ID NO: 13
AATTTTCCAGCCTGAATA	503	SEQ ID NO: 14
AATTTTCCAGCCCGCTGGCCCTGTAAA GGAAACTGGAACACAAAGCGGGGCGG GCCAGCCTGAATA	354	SEQ ID NO: 15
AATTTTCCAGCCCGCTGGCCCTGTAAA GGAAACTGGAACACAAAGCATA	349	SEQ ID NO: 16

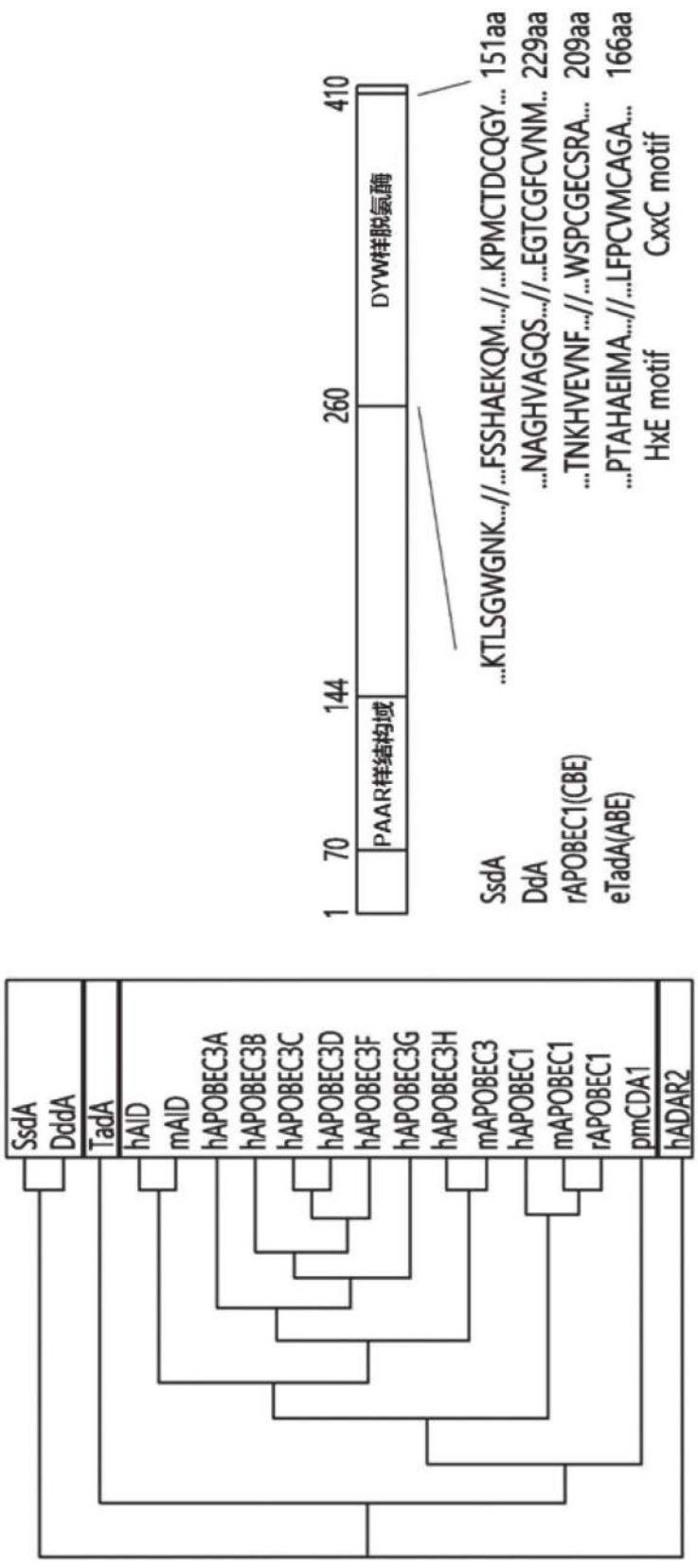

1
70
144
260
410
PAAR样结构域
DYW样脱氨酶
SsdA ...KTLSGWGNK...//...FSSHAEKQM...//...KPMCTDCQGY... 151aa
DdA ...NAGHVAGQS...//...EGTCGFCVNM... 229aa
rAPOBEC1(CBE) ...TNKHVEVNF...//...WSPCGECSRA... 209aa
eTadA(ABE) ...PTAHAEIMA...//...LFPCVMCAGA... 166aa
HxE motif CxxC motif
SsdA
DddA
TadA
hAID
mAID
hAPOBEC3A
hAPOBEC3B
hAPOBEC3C
hAPOBEC3D
hAPOBEC3F
hAPOBEC3G
hAPOBEC3H
mAPOBEC3
hAPOBEC1
mAPOBEC1
rAPOBEC1
pmCDA1
hADAR2

图1

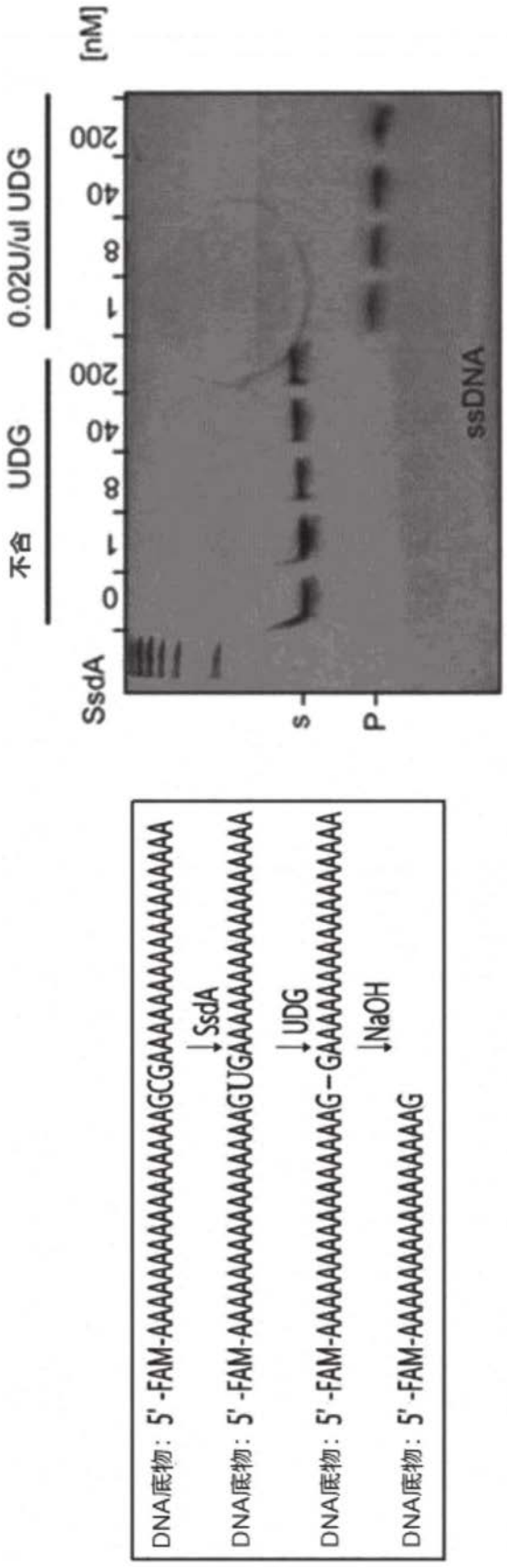

[nM]
0.02U/ul UDG
不含 UDG
200
40
8
1
0
SsdA
S
P
ssDNA
DNA底物：5'-FAM-AAAAAAAAAAAAAAAAGCGAAAAAAAAAAAAAAA
↓SsdA
DNA底物：5'-FAM-AAAAAAAAAAAAAAAGUGAAAAAAAAAAAAAAA
↓UDG
DNA底物：5'-FAM-AAAAAAAAAAAAAAAG−GAAAAAAAAAAAAAAA
↓NaOH
DNA底物：5'-FAM-AAAAAAAAAAAAAAAG

图2

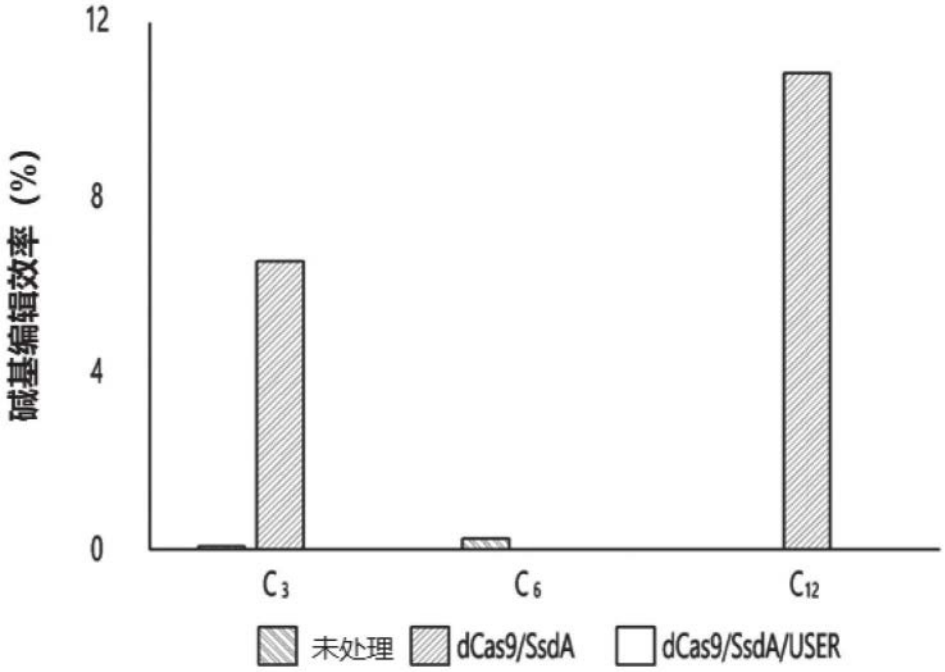

碱基编辑效率（%）
12
8
4
0
C3
C6
C12
未处理
dCas9/SsdA
dCas9/SsdA/USER

图3a

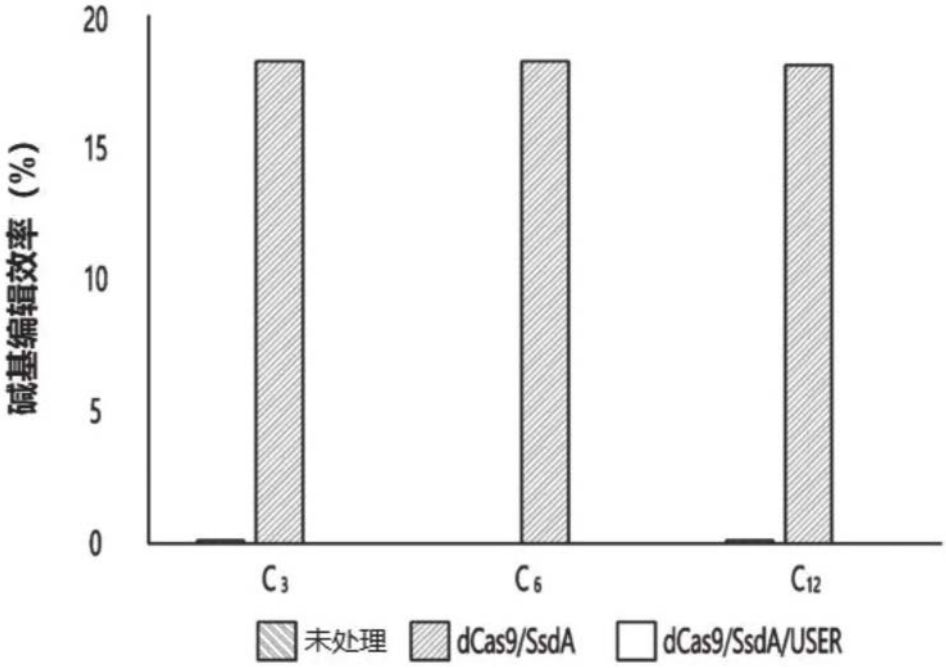

碱基编辑效率（%）
20
15
10
5
0
C3
C6
C12
未处理
dCas9/SsdA
dCas9/SsdA/USER

图3b

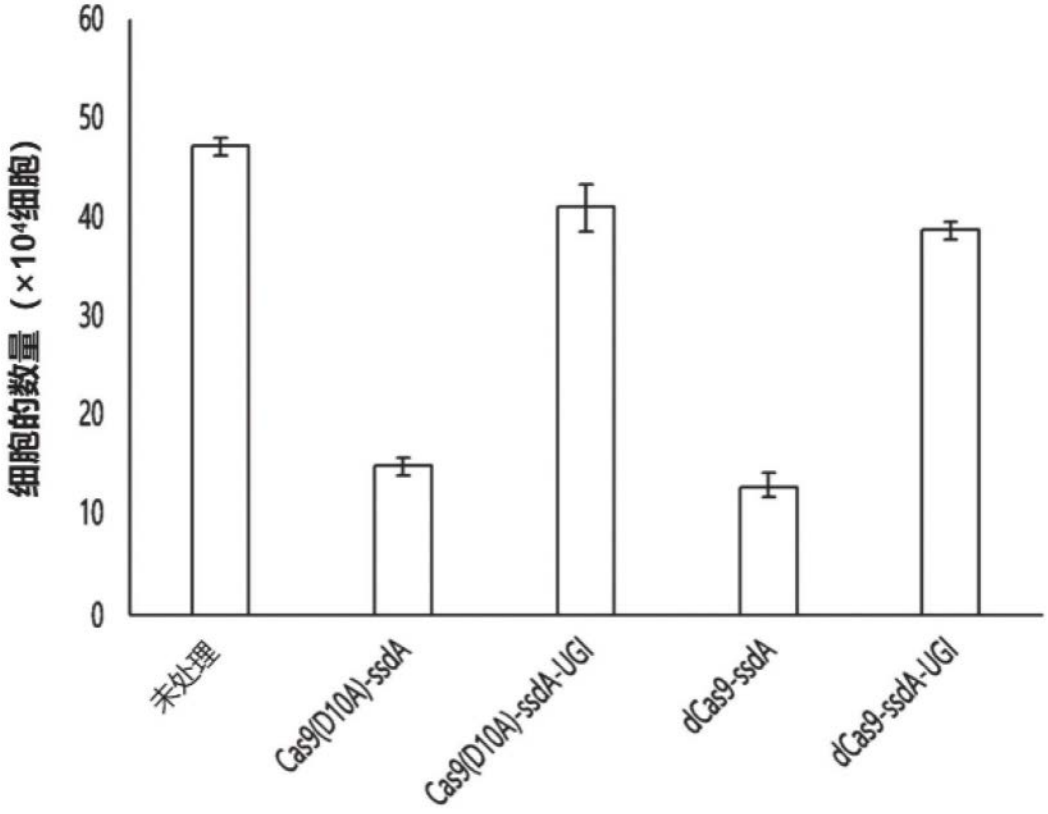

细胞的数量（×10⁴细胞）
60
50
40
30
20
10
0
未处理
Cas9(D10A)-ssdA
Cas9(D10A)-ssdA-UGI
dCas9-ssdA
dCas9-ssdA-UGI

图4

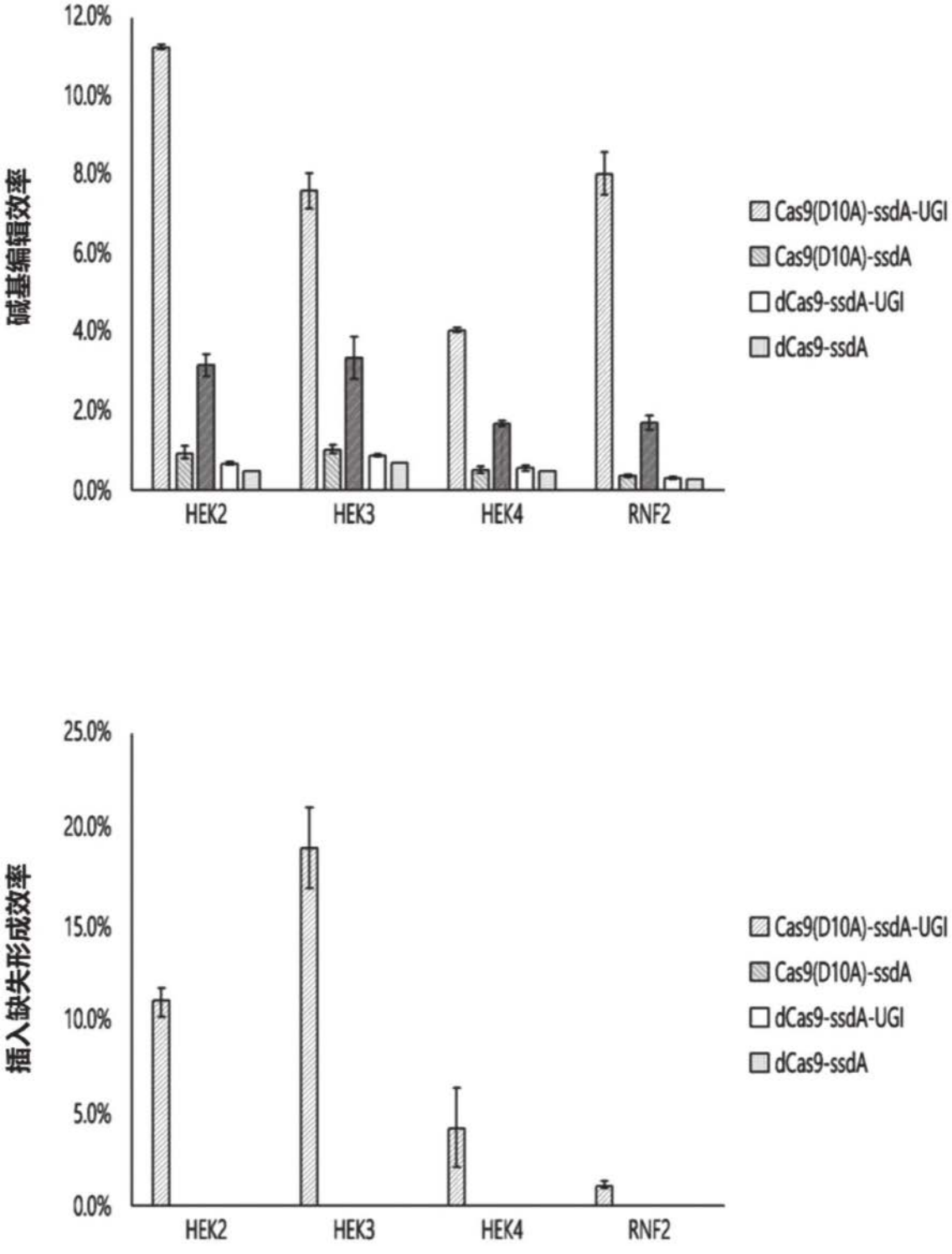

碱基编辑效率
12.0%
10.0%
8.0%
6.0%
4.0%
2.0%
0.0%
HEK2
HEK3
HEK4
RNF2
Cas9(D10A)-ssdA-UGI
Cas9(D10A)-ssdA
dCas9-ssdA-UGI
dCas9-ssdA
插入缺失形成效率
25.0%
20.0%
15.0%
10.0%
5.0%
0.0%
HEK2
HEK3
HEK4
RNF2
Cas9(D10A)-ssdA-UGI
Cas9(D10A)-ssdA
dCas9-ssdA-UGI
dCas9-ssdA

图5a

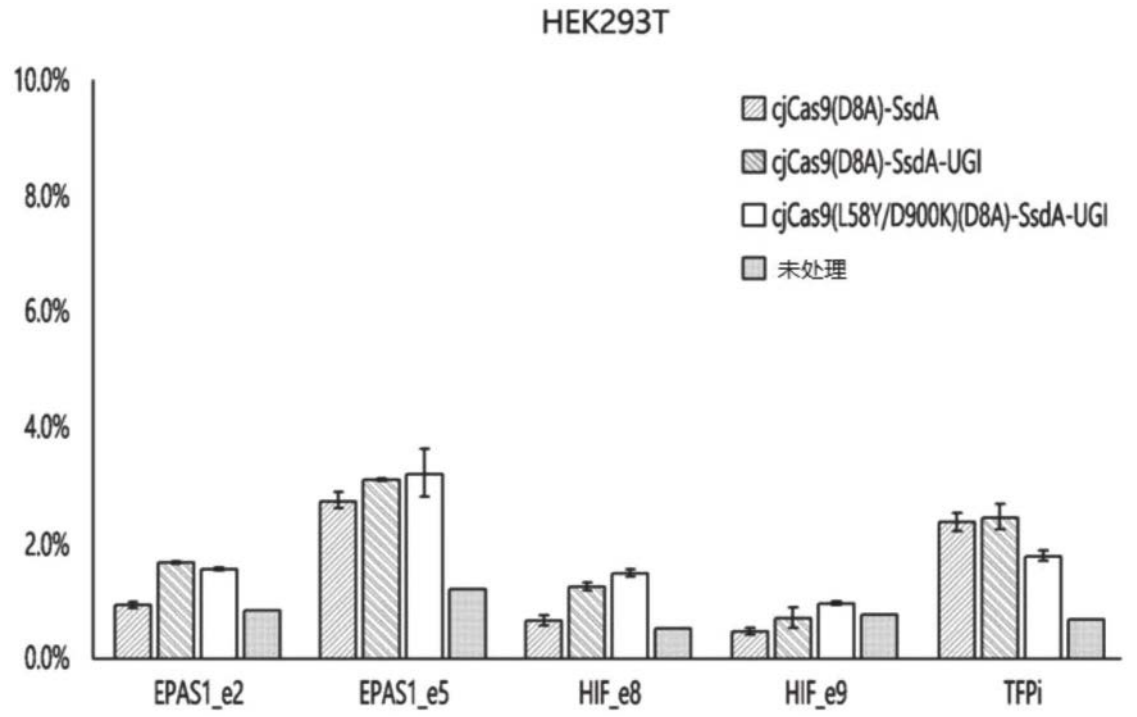

HEK293T
10.0%
8.0%
6.0%
4.0%
2.0%
0.0%
cjCas9(D8A)-SsdA
cjCas9(D8A)-SsdA-UGI
cjCas9(L58Y/D900K)(D8A)-SsdA-UGI
未处理
EPAS1_e2
EPAS1_e5
HIF_e8
HIF_e9
TFPi

图5b

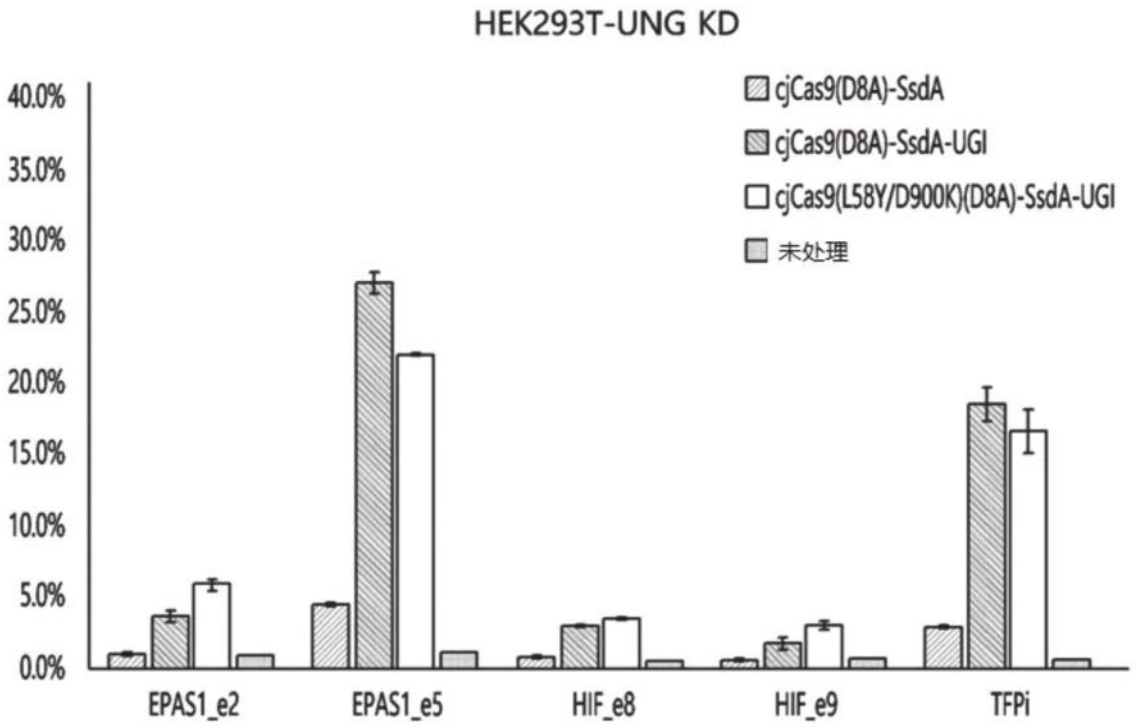

HEK293T-UNG KD
40.0%
35.0%
30.0%
25.0%
20.0%
15.0%
10.0%
5.0%
0.0%
cjCas9(D8A)-SsdA
cjCas9(D8A)-SsdA-UGI
cjCas9(L58Y/D900K)(D8A)-SsdA-UGI
未处理
EPAS1_e2
EPAS1_e5
HIF_e8
HIF_e9
TFPi

图5c

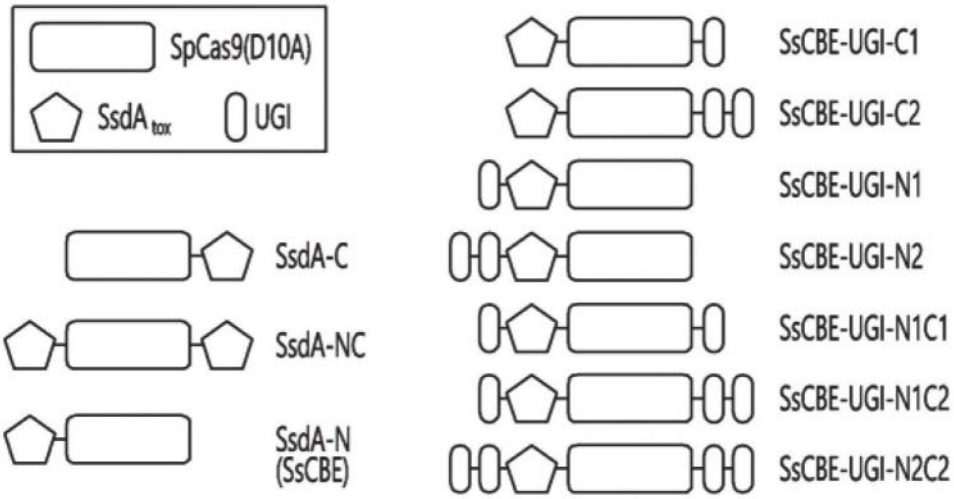

SpCas9(D10A)
SsdA tox
UGI
SsdA-C
SsdA-NC
SsdA-N
(SsCBE)
SsCBE-UGI-C1
SsCBE-UGI-C2
SsCBE-UGI-N1
SsCBE-UGI-N2
SsCBE-UGI-N1C1
SsCBE-UGI-N1C2
SsCBE-UGI-N2C2

图6

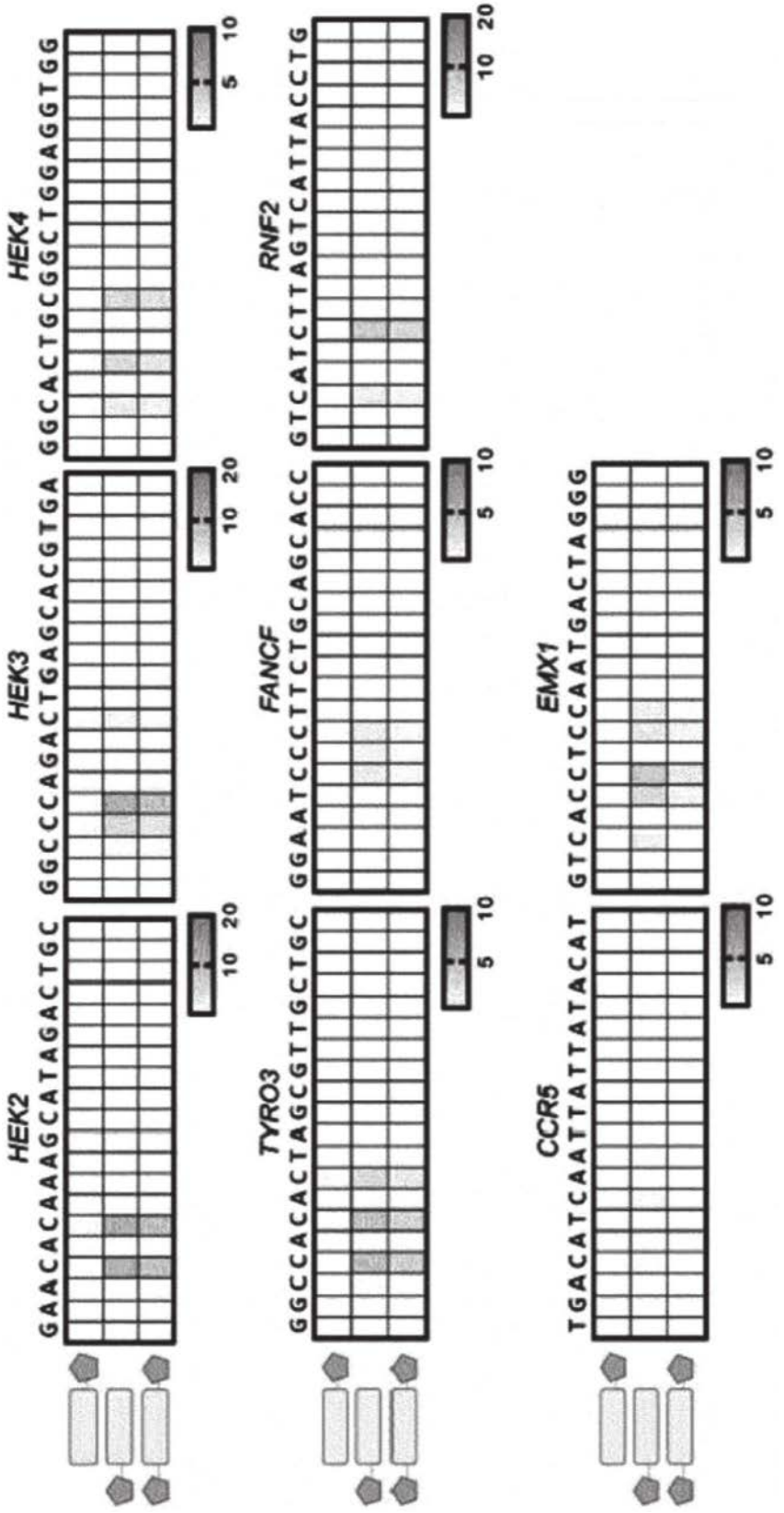

HEK4
GGCACTGCGGCTGGAGGTGG
HEK3
GGCCCAGACTGAGCACGTGA
HEK2
GAACACAAAGCATAGACTGC
RNF2
GTCATCTTAGTCATTACCTG
FANCF
GGAATCCCTTCTGCAGCACC
TYRO3
GGCCACACTAGCGTTGCTGC
EMX1
GTCACCTCCAATGACTAGGG
CCR5
TGACATCAATTATTATACAT
5 10
10 20

图7a

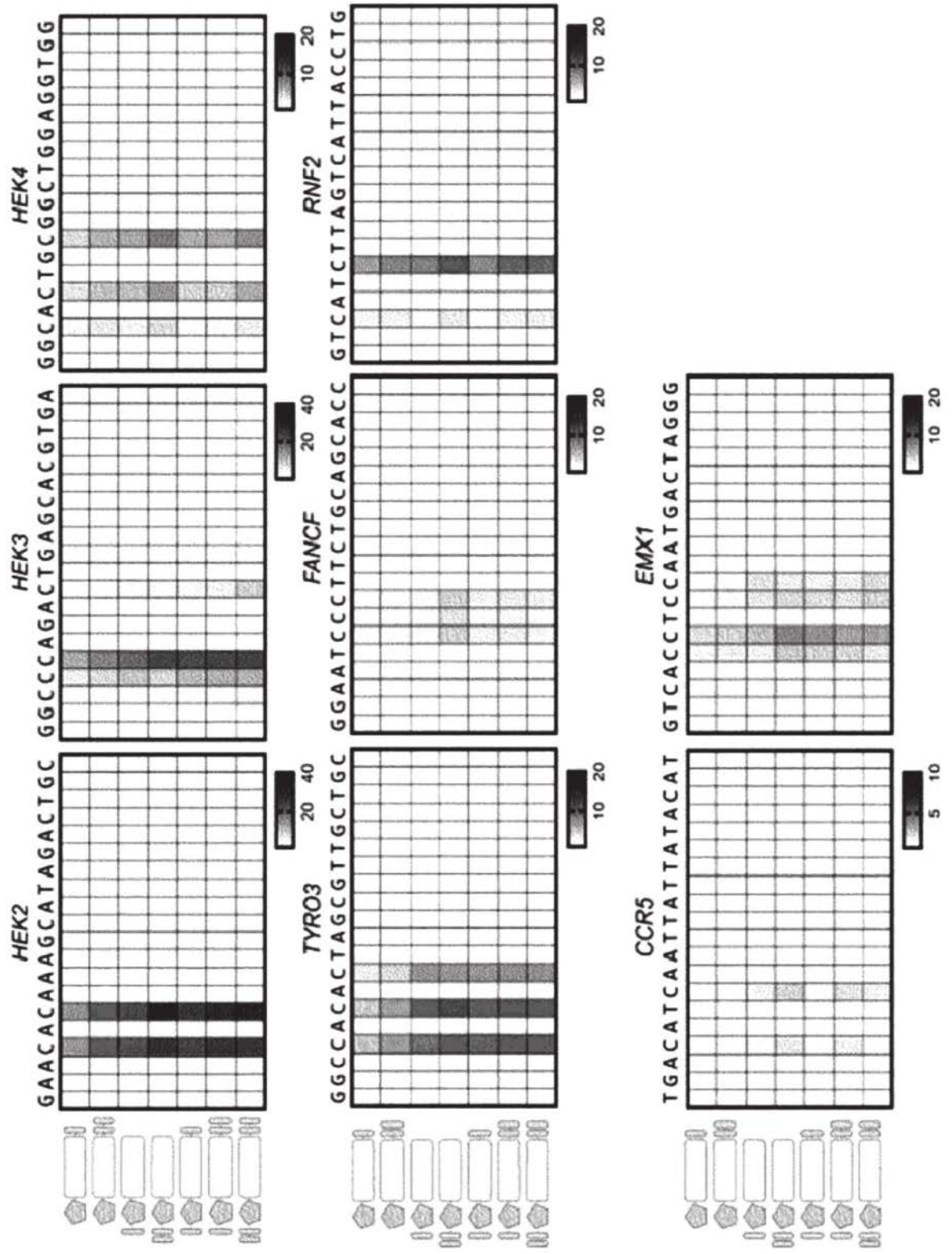

HEK4
GGCACTGCGGCTGGAGGTGG
HEK3
GGCCCAGACTGAGCACGTGA
HEK2
GAACACAAAGCATAGACTGC
RNF2
GTCATCTTAGTCATTACCTG
FANCF
GGAATCCCTTCTGCAGCACC
TYRO3
GGCCACACTAGCGTTGCTGC
EMX1
GTCACCTCCAATGACTAGGG
CCR5
TGACATCAATTATTATACAT

图7b

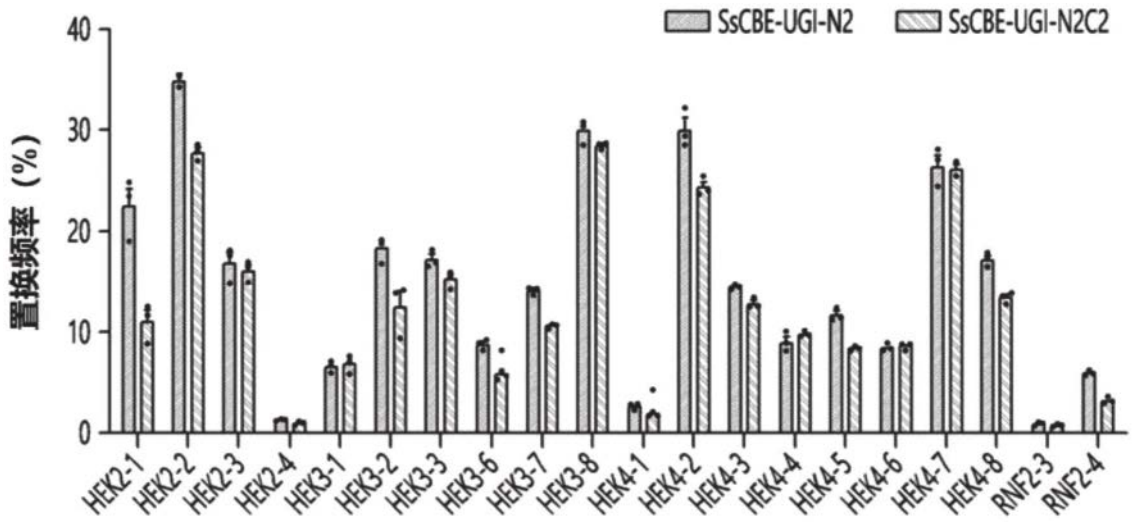

SsCBE-UGI-N2
SsCBE-UGI-N2C2
置换频率（%）
40
30
20
10
0
HEK2-1
HEK2-2
HEK2-3
HEK2-4
HEK3-1
HEK3-2
HEK3-3
HEK3-6
HEK3-7
HEK3-8
HEK4-1
HEK4-2
HEK4-3
HEK4-4
HEK4-5
HEK4-6
HEK4-7
HEK4-8
RNF2-3
RNF2-4

图7c

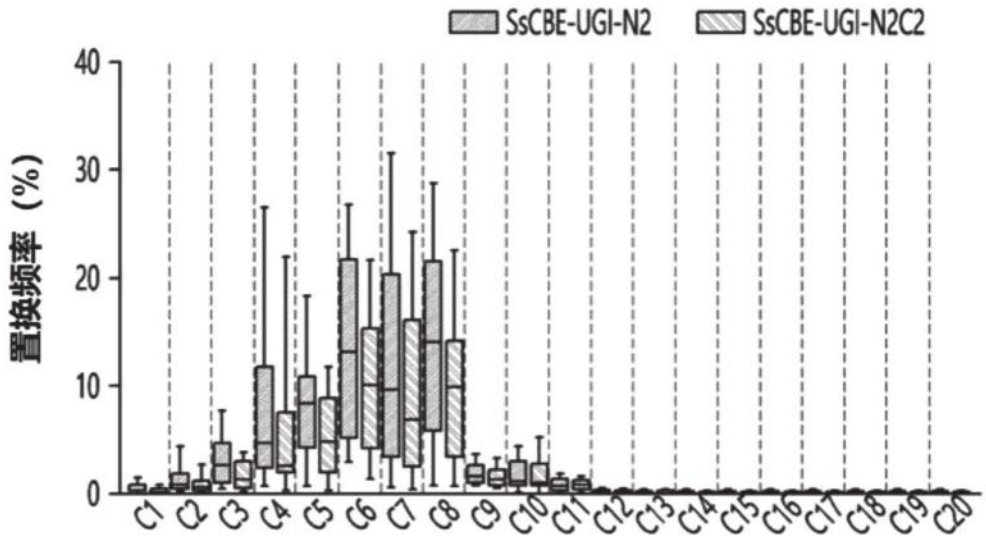

SsCBE-UGI-N2
SsCBE-UGI-N2C2
置换频率（%）
40
30
20
10
0
C1
C2
C3
C4
C5
C6
C7
C8
C9
C10
C11
C12
C13
C14
C15
C16
C17
C18
C19
C20

图7d

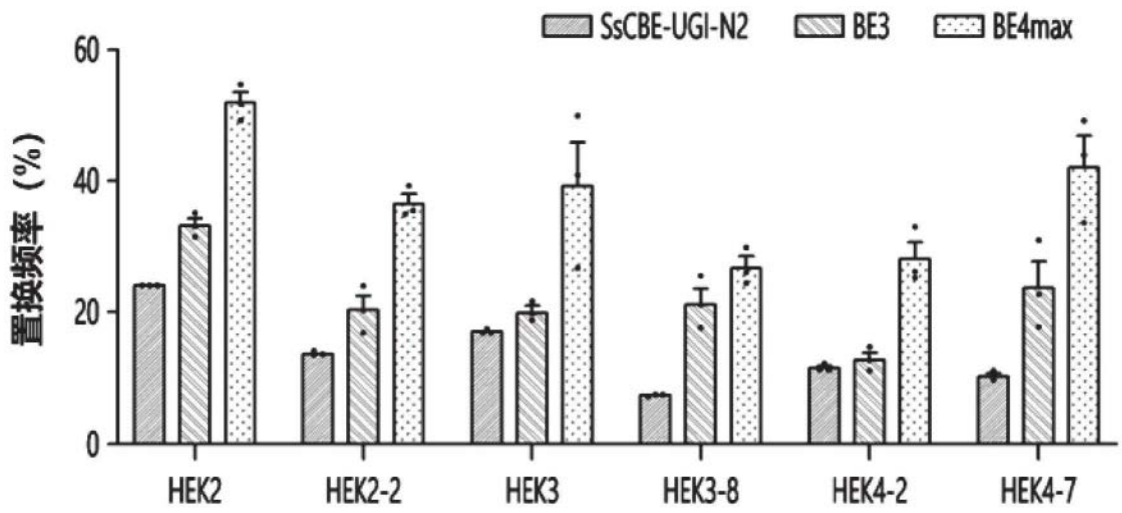

SsCBE-UGI-N2
BE3
BE4max
置换频率（%）
60
40
20
0
HEK2
HEK2-2
HEK3
HEK3-8
HEK4-2
HEK4-7

图8

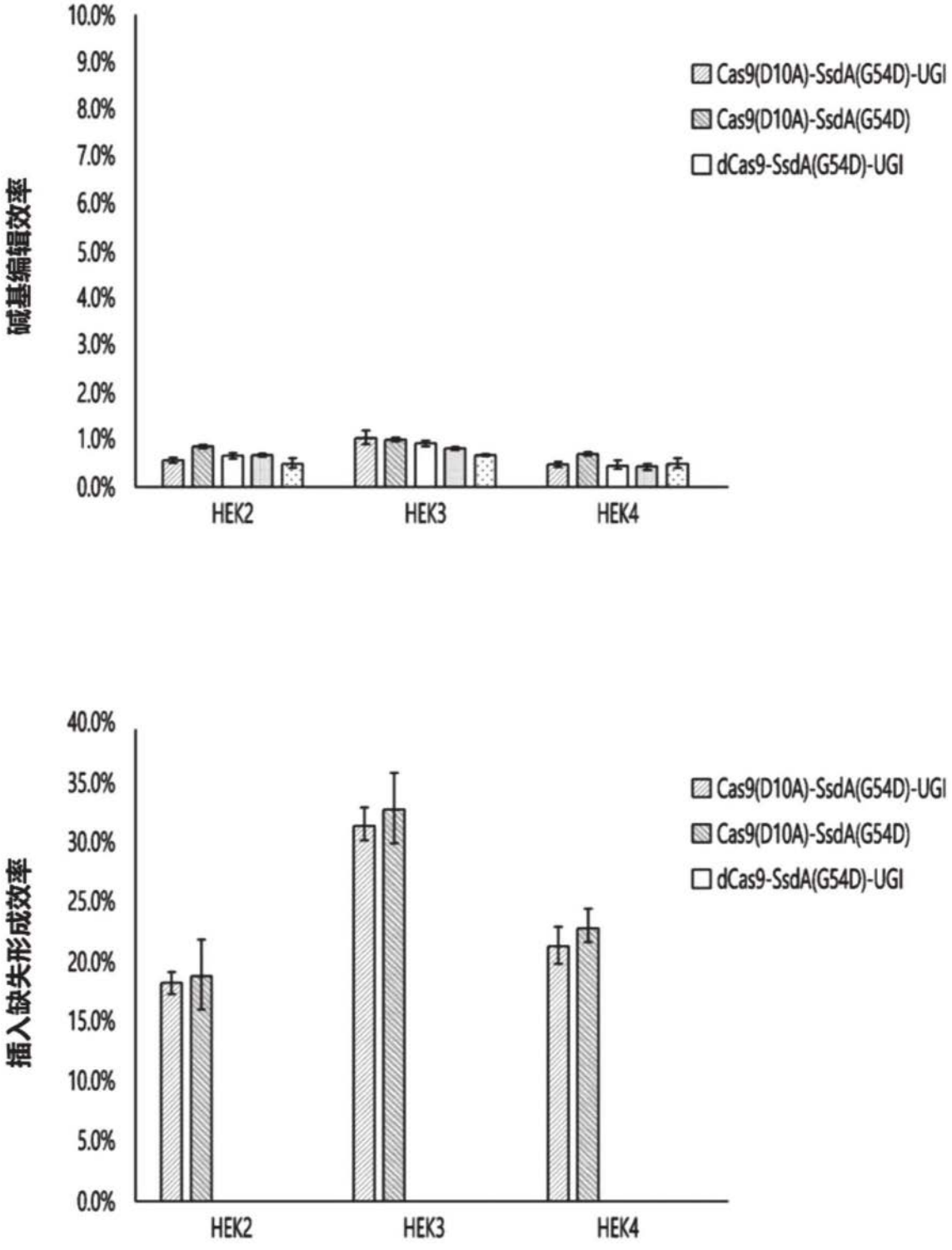

碱基编辑效率
10.0%
9.0%
8.0%
7.0%
6.0%
5.0%
4.0%
3.0%
2.0%
1.0%
0.0%
HEK2
HEK3
HEK4
Cas9(D10A)-SsdA(G54D)-UGI
Cas9(D10A)-SsdA(G54D)
dCas9-SsdA(G54D)-UGI
插入缺失形成效率
40.0%
35.0%
30.0%
25.0%
20.0%
15.0%
10.0%
5.0%
0.0%
HEK2
HEK3
HEK4
Cas9(D10A)-SsdA(G54D)-UGI
Cas9(D10A)-SsdA(G54D)
dCas9-SsdA(G54D)-UGI

图9