(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 114929287 B
(45)授权公告日 2025.06.03

(21)申请号 202080050214.3
(22)申请日 2020.05.08
(65)同一申请的已公布的文献号
申请公布号 CN 114929287 A
(43)申请公布日 2022.08.19
(30)优先权数据
62/846,422 2019.05.10 US
62/927,585 2019.10.29 US
(85)PCT国际申请进入国家阶段日
2022.01.10
(86)PCT国际申请的申请数据
PCT/US2020/032226 2020.05.08
(87)PCT国际申请的公布数据
WO2020/231863 EN 2020.11.19
(73)专利权人 比姆医疗股份有限公司
地址 美国马萨诸塞州
(72)发明人 E·斯梅卡洛瓦 M·帕克
F·葛瑞格尔 L·巴雷拉
G·恰拉梅拉
(74)专利代理机构 北京戈程知识产权代理有限公司 11314
专利代理师 程伟
(51)Int.Cl.
A61K 48/00(2006.01)
C12N 15/113(2006.01)
C12N 15/85(2006.01)
C12N 15/90(2006.01)
C12N 15/10(2006.01)
C12N 9/22(2006.01)
(56)对比文件
US 2018236103 A1,2018.08.23
US 2019093099 A1,2019.03.28
WO 2019023680 A1,2019.01.31 (续)
审查员 周凯燕

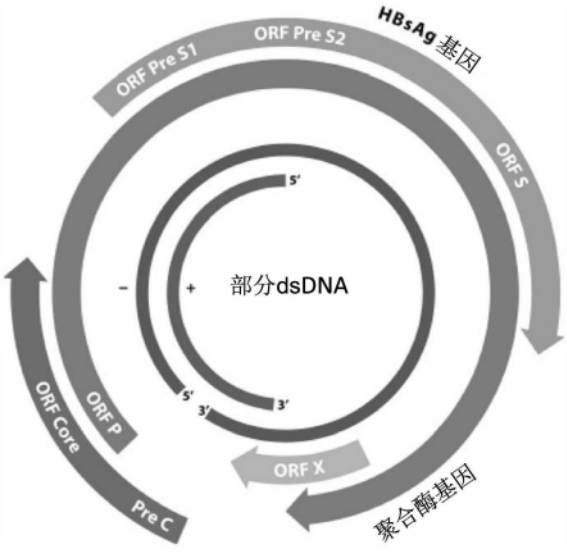
权利要求书8页 说明书403页
序列表（电子公布） 附图37页

(54)发明名称

用于治疗乙型肝炎的组合物和方法

(57)摘要

本发明的特征在于用于将突变引入乙型肝炎(HBV)基因组中的组合物和方法。



CN 114929287 B

[转续页]

[接上页]

(56)对比文件

D Kostiushev等.“A novel CRISPR/Cas9-based approach to transient activation of intracellular host restriction factors results in strong suppression of hepatitis B virus and degradation of cccDNA”.《Journal of Viral Hepatitis》.2018,第25卷(第S2期),第16页右栏第2段-第17页左栏第1段.

D Kostiushev等.“A novel CRISPR/Cas9-based approach to transient activation of intracellular host restriction factors results in strong suppression of hepatitis B virus and degradation of cccDNA”.《Journal of Viral Hepatitis》.2018,第25卷(第S2期),第16页右栏第2段-第17页左栏第1段.

1.一种碱基编辑器和一种或多种向导RNA在制备用于编辑乙型肝炎病毒(HBV)基因组的核碱基的药物的用途,所述碱基编辑器包含Cas9结构域和腺苷脱氨酶结构域,所述Cas9结构域是无核酸酶活性的或是切口酶,

其特征在于,所述腺苷脱氨酶结构域:

i)是TadA*7.10腺苷脱氨酶,其由MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO:29)的氨基酸序列或其仅缺失N-端甲硫氨酸的片段所组成;或

ii)是TadA*8.8腺苷脱氨酶,其由MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO:29)的氨基酸序列或其仅缺失N-端甲硫氨酸的片段所组成,且参考SEQ ID NO:29所示的氨基酸序列进行由Y123H、Y147R和Q154R所组成的一组氨基酸改变;其中,所述一种或多种向导RNA介导所述碱基编辑器与靶标位点的结合,以产生HBV基因组的所述核碱基的改变,其中,所述核碱基在选自由编码如SEQ ID NO:45或46所示的氨基酸序列的pol基因、编码如SEQ ID NO:48、50或51所示的氨基酸序列的X基因、编码如SEQ ID NO:52所示的氨基酸序列的S基因和编码如SEQ ID NO:49所示的氨基酸序列的核心区基因所组成群组的HBV基因中,从而在所述基因中引入抑制HBV复制或中断HBV共价闭合环状DNA活性的错义突变,

其中:

a)当所述腺苷脱氨酶结构域是TadA*7.10腺苷脱氨酶结构域时,所述一种或多种向导RNA包含间隔序列,所述间隔序列选自以下任一核苷酸序列:UAUGGAUGAUGUGGUAUUGG、AGAAGAUGAGGCAUAGCAGC或CACCACGAGUCUAGACUCUG;以及

b)当所述腺苷脱氨酶结构域是TadA*8.8腺苷脱氨酶结构域时,所述一种或多种向导RNA包含间隔序列,所述间隔序列选自以下任一核苷酸序列:UAUGGAUGAUGUGGUAUUGG、GAGAAGUCCACCACGAGUCU、UGGACUUCUCUCAAUUUUCU、GAAGAACUCCCUCGCCUCGC、UUUGCUGACGCAACCCCCAC、AAAGAAUUUGGAGCUACUGU、AAAUUGAGAGAAGUCCACCA、AACCCCCACUGGCUGGGGCU、AAUGUCAACGACCGACCUUG、AGAAAGGCCUUGUAAGUUGG、AGAAGAUGAGGCAUAGCAGC、AGAGAGGUGCGCCCCGUGGU、CACCACGAGUCUAGACUCUG、GAAGAUGAGGCAUAGCAGCA、GACGCAACCCCCACUGGCUG、GAGGUGAAGCGAAGUGCACA、GCAAACACUUGGCACAGACC、GGGGCGCACCUCUCUUUACG、GUGAAAAAGUUGCAUGGUGC、UGACGCAACCCCCACUGGCU或UGCCCAAGGUCUUACAUAAG。

2.根据权利要求1所述的用途,其特征在于,所述结合是在真核细胞中。

3.根据权利要求2所述的用途,其特征在于,所述细胞是体内的或离体的。

4.根据权利要求1所述的用途,其特征在于,所述错义突变是在所述pol基因中,且所述错义突变导致由所述pol基因编码的HBV聚合酶蛋白中的E24G、D380G、F381P、D540G或D689G。

5.根据权利要求1所述的用途,其特征在于,所述错义突变是在所述核心区基因中,且所述错义突变导致由所述核心区基因编码的HBV核心区蛋白中的T160A。

6.根据权利要求1所述的用途,其特征在于,所述错义突变是在所述X基因中,且所述错

义突变导致由所述X基因编码的HBV X蛋白中的H86R、W120R、E122K、E121K或L141P。

7.根据权利要求1所述的用途,其特征在于,所述错义突变是在所述S基因中,且所述错义突变导致由所述S基因编码的HBV S蛋白中的T37A、S38P或M197V。

8.根据权利要求1所述的用途,其特征在于,所述Cas9是酿脓链球菌Cas9、金黄色葡萄球菌Cas9、嗜热链球菌1 Cas9或犬链球菌Cas9。

9.根据权利要求1所述的用途,其特征在于,所述Cas9具有前间隔序列邻近基序(PAM),所述PAM对于选自5'-NGG-3'、5'-NAG-3'、5'-NGA-3'、5'-NAA-3'、5'-NNAGGA-3'或 5'-NNACCA-3'的核酸序列具有特异性。

10.根据权利要求1所述的用途,其特征在于,所述Cas9具有前间隔序列邻近基序(PAM),所述PAM对于选自5'-NNNRRT-3'、5'-NGA-3'、5'-NGCG-3'、5'-NGN-3'、5'-NGCN-3'、5'-NGTN-3'或5'-NAA-3'的核酸序列具有特异性。

11.根据权利要求1所述的用途,其特征在于,所述一种或多种向导RNA包含CRISPR RNA (crRNA)和反式编码小RNA (tracrRNA),其中所述crRNA包含与HBV核酸序列互补的核酸序列。

12.根据权利要求1所述的用途,其特征在于,所述碱基编辑器与单向导RNA (sgRNA)复合,所述sgRNA包含与HBV核酸序列互补的核酸序列。

13.根据权利要求1所述的用途,其包含使所述HBV基因组与向导RNA接触,所述向导RNA介导所述碱基编辑器与两种、三种或四种HBV核酸序列的结合。

14.根据权利要求1所述的用途,其特征在于,所述间隔序列由UAUGGAUGAUGUGGUAUUGG的核苷酸序列所组成。

15.根据权利要求14所述的用途,其特征在于,所述错义突变导致由所述S基因编码的HBV S蛋白中的D540G。

16.一种融合蛋白或编码所述融合蛋白的多核苷酸以及一种或多种向导RNA在制备治疗受试者的乙型肝炎病毒(HBV)感染的药物的用途,所述融合蛋白包含Cas9结构域和腺苷脱氨酶结构域,所述Cas9结构域是无核酸酶活性的或是切口酶,

其特征在于,所述腺苷脱氨酶结构域:

i)是TadA*7.10腺苷脱氨酶,其由MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO:29)的氨基酸序列或其仅缺失N-端甲硫氨酸的片段所组成;或

ii)是TadA*8.8腺苷脱氨酶,其由MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO:29)的氨基酸序列或其仅缺失N-端甲硫氨酸的片段所组成,且参考SEQ ID NO:29所示的氨基酸序列进行由Y123H、Y147R和Q154R所组成的一组氨基酸改变;

其中,所述一种或多种向导RNA介导所述腺苷脱氨酶结构域与靶标位点的结合,以在编码选自由如SEQ ID NO:45或46所示的氨基酸序列的pol、如SEQ ID NO:48、50或51所示的氨基酸序列的X、如SEQ ID NO:52所示的氨基酸序列的S和如SEQ ID NO:49所示的氨基酸序列的核心区所组成群组的HBV多肽的多核苷酸中产生A·T到 G·C的改变,从而在编码HBV多肽

的多核苷酸中引入抑制HBV复制或中断HBV共价闭合环状DNA活性的错义突变,以及其中:

a)当所述腺苷脱氨酶结构域是TadA*7.10腺苷脱氨酶结构域时,所述一种或多种向导RNA包含间隔序列,所述间隔序列选自以下任一核苷酸序列:UAUGGAUGAUGUGGUAUUGG、AGAAGAUGAGGCAUAGCAGC或CACCACGAGUCUAGACUCUG;以及

b)当所述腺苷脱氨酶结构域是TadA*8.8腺苷脱氨酶结构域时,所述一种或多种向导RNA包含间隔序列,所述间隔序列选自以下任一核苷酸序列:UAUGGAUGAUGUGGUAUUGG、GAGAAGUCCACCACGAGUCU、UGGACUUCUCUCAAUUUUCU、GAAGAACUCCCUCGCCUCGC、UUUGCUGACGCAACCCCCAC、AAAGAAUUUGGAGCUACUGU、AAAUUGAGAGAAGUCCACCA、AACCCCCACUGGCUGGGGCU、AAUGUCAACGACCGACCUUG、AGAAAGGCCUUGUAAGUUGG、AGAAGAUGAGGCAUAGCAGC、AGAGAGGUGCGCCCCGUGGU、CACCACGAGUCUAGACUCUG、GAAGAUGAGGCAUAGCAGCA、GACGCAACCCCCACUGGCUG、GAGGUGAAGCGAAGUGCACA、GCAAACACUUGGCACAGACC、GGGGCGCACCUCUCUUUACG、GUGAAAAAGUUGCAUGGUGC、UGACGCAACCCCCACUGGCU或UGCCCAAGGUCUUACAUAAG。

17.一种多核苷酸在制备治疗受试者的乙型肝炎病毒(HBV)感染的药物的用途,其特征在于,所述多核苷酸编码的融合蛋白,其包含Cas9结构域和腺苷脱氨酶结构域,所述Cas9结构域是无核酸酶活性的或是切口酶,

其中,所述腺苷脱氨酶结构域:

i)是TadA*7.10腺苷脱氨酶,其由MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO:29)的氨基酸序列或其仅缺失N-端甲硫氨酸的片段所组成;或

ii)是TadA*8.8腺苷脱氨酶,其由MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO:29)的氨基酸序列或其仅缺失N-端甲硫氨酸的片段所组成,且参考SEQ ID NO:29所示的氨基酸序列进行由Y123H、Y147R和Q154R所组成的一组氨基酸改变;

其中,所述多核苷酸包含一种或多种向导RNA介导所述腺苷脱氨酶结构域与靶标位点的结合,以在编码选自由如SEQ ID NO:45或46所示的氨基酸序列的pol、如SEQ ID NO:48、50或51所示的氨基酸序列的X、如SEQ ID NO:52所示的氨基酸序列的S和如SEQ ID NO:49所示的氨基酸序列的核心区所组成群组的HBV多肽的多核苷酸中产生A·T到 G·C的改变,从而在编码HBV多肽的多核苷酸中引入抑制HBV复制或中断HBV共价闭合环状DNA活性的错义突变,以及其中:

a)当所述腺苷脱氨酶结构域是TadA*7.10腺苷脱氨酶结构域时,所述一种或多种向导RNA包含间隔序列,所述间隔序列选自以下任一核苷酸序列:UAUGGAUGAUGUGGUAUUGG、AGAAGAUGAGGCAUAGCAGC或CACCACGAGUCUAGACUCUG;以及

b)当所述腺苷脱氨酶结构域是TadA*8.8腺苷脱氨酶结构域时,所述一种或多种向导RNA包含间隔序列,所述间隔序列选自以下任一核苷酸序列:UAUGGAUGAUGUGGUAUUGG、GAGAAGUCCACCACGAGUCU、UGGACUUCUCUCAAUUUUCU、GAAGAACUCCCUCGCCUCGC、UUUGCUGACGCAACCCCCAC、AAAGAAUUUGGAGCUACUGU、AAAUUGAGAGAAGUCCACCA、

AACCCCCACUGGCUGGGGCU、AAUGUCAACGACCGACCUUG、AGAAAGGCCUUGUAAGUUGG、AGAAGAUGAGGCAUAGCAGC、AGAGAGGUGCGCCCCGUGGU、CACCACGAGUCUAGACUCUG、GAAGAUGAGGCAUAGCAGCA、GACGCAACCCCCACUGGCUG、GAGGUGAAGCGAAGUGCACA、GCAAACACUUGGCACAGACC、GGGGCGCACCUCUCUUUACG、GUGAAAAAGUUGCAUGGUGC、UGACGCAACCCCCACUGGCU或UGCCCAAGGUCUUACAUAAG。

18.根据权利要求16或权利要求17所述的用途,其特征在于,所述受试者是哺乳动物。

19.根据权利要求16或权利要求17所述的用途,其特征在于,所述药物给药至所述受试者的细胞。

20.根据权利要求19所述的用途,其特征在于,所述细胞是肝细胞。

21.根据权利要求16或权利要求17所述的用途,其特征在于,所述Cas9是酿脓链球菌Cas9、金黄色葡萄球菌Cas9、嗜热链球菌1 Cas9或犬链球菌Cas9。

22.根据权利要求21所述的用途,其特征在于,所述Cas9具有前间隔序列邻近基序(PAM),所述PAM对于选自5'-NGG-3'、5'-NAG-3'、5'-NGA-3'、5'-NAA-3'、5'-NNAGGA-3'或5'-NNACCA-3'的核酸序列具有特异性。

23.根据权利要求16或权利要求17所述的用途,其特征在于,所述Cas9具有前间隔序列邻近基序(PAM),所述PAM对于选自5'-NNNRRT-3'、5'-NGA-3'、5'-NGCG-3'、5'-NGN-3'、5'-NGCN-3'、5'-NGTN-3'或5'-NAA-3'的核酸序列具有特异性。

24.根据权利要求16或权利要求17所述的用途,其特征在于,所述一种或多种向导RNA包含CRISPR RNA (crRNA)和反式编码小RNA (tracrRNA),其中所述crRNA包含与HBV核酸序列互补的核酸序列。

25.根据权利要求16或权利要求17所述的用途,其特征在于,所述融合蛋白与单向导RNA (sgRNA)复合,所述sgRNA包含与HBV核酸序列互补的核酸序列。

26.根据权利要求16或权利要求17所述的用途,其特征在于,所述向导RNA其介导所述腺苷脱氨酶结构域与两种、三种或四种HBV核酸序列的结合。

27.根据权利要求16或权利要求17所述的用途,其特征在于,所述间隔序列由UAUGGAUGAUGUGGUAUUGG的核苷酸序列所组成。

28.根据权利要求27所述的用途,其特征在于,所述错义突变导致由S基因编码的HBV S蛋白中的D540G。

29.根据权利要求16或权利要求17所述的用途,其特征在于,所述错义突变是在pol基因中,且所述pol基因中的所述错义突变导致由所述pol基因编码的HBV聚合酶中的E24G、D380G、F381P、D540G或D689G。

30.根据权利要求16或权利要求17所述的用途,其特征在于,所述错义突变是在核心区基因中,且所述核心区基因中的所述错义突变导致由所述核心区基因编码的HBV核心区蛋白中的T160A。

31.根据权利要求16或权利要求17所述的用途,其特征在于,所述错义突变是在X基因中,且所述错义突变导致由所述X基因编码的HBV X蛋白中的H86R、W120R、E122K、E121K或L141P。

32.根据权利要求16或权利要求17所述的用途,其特征在于,所述错义突变是在S基因中,且所述S基因中的所述错义突变导致由所述S基因编码的HBV S蛋白中的T37A、S38P或

M197V。

33.根据权利要求1、16或17所述的用途,其特征在于,所述Cas9由具有D1135V、G1218R、R1335Q和T1337R的氨基酸替换的下列氨基酸序列或其仅缺失N-端甲硫氨酸的变体所组成:MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD。

34.一种组合物,其包含与向导RNA结合的碱基编辑器,其特征在于,所述向导RNA包含与选自由编码如SEQ ID NO:45或46所示的氨基酸序列的pol基因、编码如SEQ ID NO:48、50或51所示的氨基酸序列的X基因、编码如SEQ ID NO:52所示的氨基酸序列的S基因和编码如SEQ ID NO:49所示的氨基酸序列的核心区基因所组成群组的HBV基因互补的核酸序列以介导所述碱基编辑器与靶标位点的结合,以引入抑制HBV复制或中断HBV共价闭合环状DNA活性的错义突变,其中,所述碱基编辑器包含Cas9结构域和腺苷脱氨酶结构域,所述Cas9结构域是无核酸酶活性的或是切口酶,

其中,所述腺苷脱氨酶结构域:

i)是TadA*7.10腺苷脱氨酶,其由MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO:29)的氨基酸序列或其仅缺失N-端甲硫氨酸的片段所组成;或

ii)是TadA*8.8腺苷脱氨酶,其由MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO:29)的氨基酸序列或其仅缺失N-端甲硫氨酸的片段所组成,且参考SEQ ID NO:29所示的氨基酸序列进行由Y123H、Y147R和Q154R所组成的一组氨基酸改变;

其中:

a)当所述腺苷脱氨酶结构域是TadA*7.10腺苷脱氨酶结构域时,所述向导RNA包含间隔序列,所述间隔序列选自以下任一核苷酸序列:UAUGGAUGAUGUGGUAUUGG、AGAAGAUGAGGCAUAGCAGC或CACCACGAGUCUAGACUCUG;以及

b)当所述腺苷脱氨酶结构域是TadA*8.8腺苷脱氨酶结构域时,所述向导RNA包含间隔序列,所述间隔序列选自以下任一核苷酸序列:UAUGGAUGAUGUGGUAUUGG、GAGAAGUCCACCACGAGUCU、UGGACUUCUCUCAAUUUUCU、GAAGAACUCCCUCGCCUCGC、UUUGCUGACGCAACCCCCAC、AAAGAAUUUGGAGCUACUGU、AAAUUGAGAGAAGUCCACCA、AACCCCCACUGGCUGGGGCU、AAUGUCAACGACCGACCUUG、AGAAAGGCCUUGUAAGUUGG、AGAAGAUGAGGCAUAGCAGC、AGAGAGGUGCGCCCCGUGGU、CACCACGAGUCUAGACUCUG、GAAGAUGAGGCAUAGCAGCA、GACGCAACCCCCACUGGCUG、GAGGUGAAGCGAAGUGCACA、GCAAACACUUGGCACAGACC、GGGGCGCACCUCUCUUUACG、GUGAAAAAGUUGCAUGGUGC、UGACGCAACCCCCACUGGCU或UGCCCAAGGUCUUACAUAAG。

35.根据权利要求34所述的组合物,其特征在于,所述间隔序列由下列UAUGGAUGAUGUGGUAUUGG的核苷酸序列所组成。

36.根据权利要求35所述的组合物,其特征在于,所述错义突变导致由所述S基因编码的HBV S蛋白中的D540G。

37.根据权利要求34所述的组合物,其特征在于,进一步包含脂质,所述脂质是阳离子脂质。

38.根据权利要求34所述的组合物,其特征在于,进一步包含药学上可接受的赋形剂。

39.一种用于治疗HBV感染的药物组合物,其包含

位于药学上可接受的赋形剂中的碱基编辑器或编码所述碱基编辑器的核酸以及一种或多种向导RNA (gRNA),所述gRNA包含与选自由编码如SEQ ID NO:45或46所示的氨基酸序列的pol基因、编码如SEQ ID NO:48、50或51所示的氨基酸序列的X基因、编码如SEQ ID NO:52所示的氨基酸序列的S基因和编码如SEQ ID NO:49所示的氨基酸序列的核心区基因所组成群组的HBV基因互补的核酸序列以介导所述碱基编辑器与靶标位点的结合,以引入抑制HBV复制或中断共价闭合环状DNA活性的错义突变,其中,所述碱基编辑器包含Cas9结构域和腺苷脱氨酶结构域,所述Cas9结构域是无核酸酶活性的或是切口酶,

其中,所述腺苷脱氨酶结构域:

i)是TadA*7.10腺苷脱氨酶,其由MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO:29)的氨基酸序列或其仅缺失N-端甲硫氨酸的片段所组成;或

ii)是TadA*8.8腺苷脱氨酶,其由MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO:29)的氨基酸序列或其仅缺失N-端甲硫氨酸的片段所组成,且参考SEQ ID NO:29所示的氨基酸序列进行由Y123H、Y147R和Q154R所组成的一组氨基酸改变;

其中:

a)当所述腺苷脱氨酶结构域是TadA*7.10腺苷脱氨酶结构域时,所述一种或多种向导

RNA包含间隔序列,所述间隔序列选自以下任一核苷酸序列:UAUGGAUGAUGUGGUAUUGG、AGAAGAUGAGGCAUAGCAGC或CACCACGAGUCUAGACUCUG;以及

b)当所述腺苷脱氨酶结构域是TadA*8.8腺苷脱氨酶结构域时,所述一种或多种向导RNA包含间隔序列,所述间隔序列选自以下任一核苷酸序列:UAUGGAUGAUGUGGUAUUGG、GAGAAGUCCACCACGAGUCU、UGGACUUCUCUCAAUUUUCU、GAAGAACUCCCUCGCCUCGC、UUUGCUGACGCAACCCCCAC、AAAGAAUUUGGAGCUACUGU、AAAUUGAGAGAAGUCCACCA、AACCCCCACUGGCUGGGGCU、AAUGUCAACGACCGACCUUG、AGAAAGGCCUUGUAAGUUGG、AGAAGAUGAGGCAUAGCAGC、AGAGAGGUGCGCCCCGUGGU、CACCACGAGUCUAGACUCUG、GAAGAUGAGGCAUAGCAGCA、GACGCAACCCCCACUGGCUG、GAGGUGAAGCGAAGUGCACA、GCAAACACUUGGCACAGACC、GGGGCGCACCUCUCUUUACG、GUGAAAAAGUUGCAUGGUGC、UGACGCAACCCCCACUGGCU或UGCCCAAGGUCUUACAUAAG。

40.根据权利要求39所述的药物组合物,其特征在于,所述Cas9由具有D1135V、G1218R、R1335Q和T1337R的氨基酸替换的下列氨基酸序列或其仅缺失N-端甲硫氨酸的变体所组成:MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD。

41.根据权利要求39或40所述的药物组合物,其特征在于,所述一种或多种gRNA和所述碱基编辑器配制在一起或单独配制。

42.根据权利要求39或40所述的药物组合物,其特征在于,所述间隔序列由UAUGGAUGAUGUGGUAUUGG的核苷酸序列所组成。

43.根据权利要求42所述的药物组合物,其特征在于,所述错义突变导致由所述S基因编码的HBV S蛋白中的D540G。

44.根据权利要求39或40所述的药物组合物,其特征在于,进一步包含适用于在哺乳动物细胞中表达的载体,所述载体包含编码所述碱基编辑器的多核苷酸。

45.根据权利要求44所述的药物组合物,其特征在于,所述载体是病毒载体。

46.根据权利要求45所述的药物组合物,其特征在于,所述病毒载体是逆转录病毒载体、腺病毒载体、疱疹病毒载体或腺相关病毒载体。

47.根据权利要求1、16或17所述的用途或根据权利要求39或40所述的药物组合物,其特征在于,所述一种或多种向导RNA由下列核苷酸序列所组成:

5'-usasus GGATGATGTGGTATTGG GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUsususu-3',其中,a、u是2'-O-甲基修饰的核苷酸,s是硫代磷酸酯。

48.根据权利要求47所述的用途或药物组合物,其特征在于,所述错义突变导致由所述S基因编码的HBV S蛋白中的D540G。

49.根据权利要求39或40所述的药物组合物,其特征在于,进一步包含脂质。

50.根据权利要求39或40所述的药物组合物,其特征在于,编码所述碱基编辑器的所述核酸是mRNA。

51.一种根据权利要求38所述的组合物在制备治疗HBV感染的药物的用途。

52.一种根据权利要求39至46中任一项所述的药物组合物在制备治疗HBV感染的药物的用途。

用于治疗乙型肝炎的组合物和方法

[0001] 相关申请的交叉引用

[0002] 本国际PCT申请主张2019年5月10日提交的美国临时专利申请号62/846,422和2019年10月29日提交的美国临时专利申请号62/927,585的优先权和权益,其各自的内容通过引用以其整体并入本文。

技术领域

[0003] 本公开关于将突变引入乙型肝炎(HBV)基因组中的组合物和方法。

背景技术

[0004] 乙型肝炎是由乙型肝炎病毒(HBV)引起的严重的肝脏感染。HBV是一种小型DNA嗜肝病毒,它通过RNA中间体复制并且可以通过整合到宿主基因组中而在被感染细胞中长期存在。全世界范围内有大约2.57亿的HBV慢性感染者,包括美国中8.5万到220万人之间。慢性HBV感染表现为慢性肝炎、肝硬化和/或肝细胞癌。20％到30％的患有慢性HBV感染的成年人发展为肝细胞癌或肝硬化。HBV感染每年导致600,000到1,000,000人死亡。

[0005] HBV感染的现有治疗方法具有严重的限制性。抗病毒药物,例如,替诺福韦(一种核苷酸逆转录酶抑制剂)可以减少病毒复制但不能治愈HBV感染患者。这些抗病毒疗法每个月花费患者多达500到1500美元。由于HBV造成的肝损伤程度,在一些情况下,移植变得必要。处理器官移植的固有风险之外,成本也令人望而却步。因此,亟需用于治疗HBV感染的改善的方法。

发明内容

[0006] 如下文所述,本发明的特征在于通过将改变引入乙型肝炎病毒(HBV)基因组内来治疗HBV感染的组合物和方法。在特定实施方案中,本发明提供一种碱基编辑器系统(例如,包含可编程DNA结合蛋白、核碱基编辑器和gRNA的融合蛋白),其用于修饰HBV基因组以引入改变诸如提前终止密码子或将改变引入HBV的编码序列中,或者将核碱基的脱氨作用引入HBV共价闭合环状DNA(cccDNA)中。

[0007] 本文提供了用于编辑乙型肝炎(HBV基因组的方法和组合物以及相关治疗及其用途。一方面,本文提供了一种编辑乙型肝炎病毒(HBV)基因组的核碱基的方法,其中该方法包括使HBV基因组与一种或多种向导RNA和碱基编辑器接触,所述碱基编辑器包含多核苷酸可编程DNA结合结构域和腺苷脱氨酶或胞苷脱氨酶结构域,其中所述向导RNA将所述碱基编辑器靶向以使得HBV基因组的核碱基发生改变。在一些实施方案中,HBV基因组的核碱基位于编码HBV蛋白的多核苷酸中。在一些实施方案中,接触是在真核细胞、哺乳动物细胞或人细胞中。在一些实施方案中,接触在细胞中以体内或离体方式进行。在一些实施方案中,胞苷脱氨酶将HBV基因组中的靶标C转化为U。在一些实施方案中,胞苷脱氨酶将编码HBV蛋白的多核苷酸中的靶标C·G转化为T·A。在一些实施方案中,腺苷脱氨酶将编码HBV蛋白的多核苷酸中的靶标A·T转化为G·C。

[0008] 在上述方法的一些实施方案中,在编码HBV蛋白的多核苷酸中的HBV基因组中的核碱基的改变导致提前终止密码子。在一些实施方案中,核碱基的改变导致HBV X蛋白中的R87*或W120*终止。在一些实施方案中,核碱基的改变导致HBV S蛋白中的W35*或W36*终止。在一些实施方案中,HBV多核苷酸的改变是错义突变。在一些实施方案中,错义突变位于HBV pol基因中。在一些实施方案中,错义突变导致由HBV pol基因编码的HBV聚合酶蛋白中的E24G、L25F、P26F、R27C、V48A、V48I、S382F、V378I、V378A、V379I、V379A、L377F、D380G、D380N、F381P、R376G、A422T、F423P、A432V、M433V、P434S、D540G、A688V、D689G、A717T、E718K、P713S、P713L或L719P。在一些实施方案中,错义突变位于HBV核心区基因中。在一些实施方案中,错义突变导致由HBV核心区基因编码的HBV核心区蛋白中的T160A、T160A、P161F、S162L、C183R或*184Q。在一些实施方案中,错义突变位于HBV X基因中。在一些实施方案中,错义突变导致由HBV X基因编码的HBV X蛋白中的H86R、W120R、E122K、E121K或L141P。在某些实施方案中,错义突变导致由HBV S基因编码的HBV S蛋白中的S38F、L39F、W35R、W36R、T37I、T37A、R78Q、S34L、F80P或D33G。

[0009] 在上述方法的一些实施方案中,本文提供的多核苷酸可编程DNA结合结构域是选自以下的Cas9:酿脓链球菌(Streptococcus pyogenes)Cas9(SpCas9)、金黄色葡萄球菌(Staphylococcus aureus)Cas9(SaCas9)、嗜热链球菌(Streptococcus thermophilus)1Cas9(St1Cas9)、犬链球菌(Steptococcus canis)Cas9(ScCas9)、或其变体。在一些实施方案中,Cas9具有前间隔序列邻近基序(PAM),所述PAM对于选自5'-NGG-3'、5'-NAG-3'、5'-NGA-3'、5'-NAA-3'、5'-NNAGGA-3'或5'-NNACCA-3'的核酸序列具有特异性。在一些实施方案中,多核苷酸可编程DNA结合结构域包含对于改变的前间隔序列邻近基序(PAM)具有特异性的经修饰的Cas9。在一些实施方案中,改变的PAM选自5'-NNNRRT-3'、NGA-3'、5'-NGCG-3'、5'-NGN-3'、NGCN-3'、5'-NGTN-3'或5'-NAA-3'。在一些实施方案中,多核苷酸可编程DNA结合结构域是无核酸活性的或切口酶变体。在一些实施方案中,无核酸活性的或切口酶变体是核酸酶灭活的Cas9(dCas9),其包含氨基酸置换D10A或其相应的氨基酸置换。

[0010] 在上述方法的一些实施方案中,腺苷脱氨酶结构域能够将脱氧核糖核酸(DNA)中的腺嘌呤脱氨。在一些实施方案中,腺苷脱氨酶是TadA脱氨酶。在一些实施方案中,TadA脱氨酶是TadA*7.10、TadA*8.1、TadA*8.2、TadA*8.3、TadA*8.4、TadA*8.5、TadA*8.6、TadA*8.7、TadA*8.8、TadA*8.9、TadA*8.10、TadA*8.11、TadA*8.12、TadA*8.13、TadA*8.14、TadA*8.15、TadA*8.16、TadA*8.17、TadA*8.18、TadA*8.19、TadA*8.20、TadA*8.21、TadA*8.22、TadA*8.23或TadA*8.24。在一些实施方案中,胞苷脱氨酶能够将DNA中的胞苷脱氨基。在一些实施方案中,胞苷脱氨酶是APOBEC或其衍生物。在一些实施方案中,碱基编辑器进一步包含尿嘧啶糖苷酶抑制剂(UGI)。在一些实施方案中,碱基编辑器不包含尿嘧啶糖苷酶抑制剂(UGI)。

[0011] 在上述方法的一些实施方案中,一种或多种用于编辑HBV基因组中的核碱基的向导RNA包含CRISPR RNA(crRNA)和反式编码小RNA(tracrRNA),其中所述crRNA包含与HBV核酸序列互补的核酸序列。在一些实施方案中,碱基编辑器与单向导RNA(sgRNA)复合,所述sgRNA包含与HBV核酸序列互补的核酸序列。一些实施方案中,HBV蛋白是HBV S、聚合酶(pol)、核心区或X蛋白。

[0012] 在一些方面,上述用于编辑乙型肝炎病毒(HBV)基因组的方法包括编辑一个或多

个核碱基。在一些实施方案中,该方法包含两种或更多种向导RNA,所述向导RNA靶向两种或更多种HBV核酸序列。在一些实施方案中,向导RNA包含选自下列中的一者或多者的从5'至3'的序列,或者其1、2、3、4或5个核苷酸5'截短片段:UCAAUCCCAACAAGGACACC、GGGAACAAGAUCUACAGCAU、AAGCCCAGGAUGAUGGGAUG、CUGCCAACUGGAUCCUGCGC、GACACAUCCAGCGAUAACCA、GCUGCCAACUGGAUCCUGCG、UAUGGAUGAUGUGGUAUUGG、CCAUGCCCCAAAGCCACCCA、AAGCCACCCAAGGCACAGCU、GAGAAGUCCACCACGAGUCU、CUUCUCUCAAUUUUCUAGGG、GACGACGAGGCAGGUCCCCU、CCCAACAAGGACACCUGGCC、UGCCAACUGGAUCCUGCGCG、AGGAGUUCCGCAGUAUGGAU、CCGCAGUAUGGAUCGGCAGA、CCUCUGCCGAUCCAUACUGC、CGCCCACCGAAUGUUGCCCA、GACUUCUCUCAAUUUUCUAG、GUUCCGCAGUAUGGAUCGGC、UACUAACAUUGAGGUUCCCG、UCCGCAGUAUGGAUCGGCAG、UCCUCUGCCGAUCCAUACUG、GUAGCUCCAAAUUCUUUAUA或AAUCCACACUCCGAAAGACA。

[0013] 另一方面,提供了一种治疗受试者的乙型肝炎病毒(HBV)感染的方法,其中该方法包括向有此需要的受试者给药融合蛋白或编码该融合蛋白的多核苷酸,该融合蛋白包含多核苷酸可编程DNA结合结构域和碱基编辑器结构域,该碱基编辑器结构域是腺苷脱氨酶或胞苷脱氨酶结构域;以及一种或多种向导多核苷酸,所述向导多核苷酸将碱基编辑器结构域靶向以在编码HBV多肽的核酸序列中产生A·T到G·C、C·G到T·A、或C·G到U·A的改变。

[0014] 另一方面,提供了一种治疗受试者的乙型肝炎病毒(HBV)感染的方法,其中该方法包括向有此需要的受试者给药编码多核苷酸可编程DNA结合结构域的多核苷酸和碱基编辑器结构域,所述碱基编辑器结构域是腺苷脱氨酶或胞苷脱氨酶结构域;以及一种或多种向导多核苷酸,所述向导多核苷酸将碱基编辑器结构域靶向以在编码HBV多肽的核酸序列中产生A·T到G·C、C·G到T·A、或C·G到U·A的改变。

[0015] 在上述治疗方法的一些实施方案中,受试者是哺乳动物或人。在一些实施方案中,该方法包括将述融合蛋白、编码所述融合蛋白的多核苷酸、或编码多核苷酸可编程DNA结合结构域的一种或多种多核苷酸和碱基编辑器结构域、以及所述一种或多种向导多核苷酸给药至受试者的细胞。在一些实施方案中,细胞是肝细胞。在一些实施方案中,多核苷酸可编程DNA结合结构域是选自以下的Cas9:酿脓链球菌Cas9(SpCas9)、金黄色葡萄球菌Cas9(SaCas9)、嗜热链球菌1Cas9(St1Cas9)、犬链球菌Cas9(ScCas9)、或其变体。在一些实施方案中,Cas9具有前间隔序列邻近基序(PAM),所述PAM对于选自5'-NGG-3'、5'-NAG-3'、5'-NGA-3'、5'-NAA-3'、5'-NNAGGA-3'或5'-NNACCA-3'的核酸序列具有特异性。在一些实施方案中,多核苷酸可编程DNA结合结构域包含对于改变的前间隔序列邻近基序(PAM)具有特异性的经修饰的Cas9。在一些实施方案中,改变的PAM的核酸序列选自5'-NNNRRT-3'、NGA-3'、5'-NGCG-3'、5'-NGN-3'、NGCN-3'、5'-NGTN-3'或5'-NAA-3'。在一些实施方案中,多核苷酸可编程DNA结合结构域是无核酸活性的或切口酶变体。在一些实施方案中,无核酸活性的或切口酶变体是核酸酶灭活的Cas9(dCas9),其包含氨基酸置换D10A或其相应的氨基酸置换。在一些实施方案中,腺苷脱氨酶结构域能够将脱氧核糖核酸(DNA)中的腺嘌呤脱氨基。在一些实施方案中,腺苷脱氨酶是TadA脱氨酶。在一些实施方案中,TadA脱氨酶是TadA*7.10、TadA*8.1、TadA*8.2、TadA*8.3、TadA*8.4、TadA*8.5、TadA*8.6、TadA*8.7、TadA*8.8、TadA*8.9、TadA*8.10、TadA*8.11、TadA*8.12、TadA*8.13、TadA*8.14、TadA*8.15、TadA*8.16、

TadA*8.17、TadA*8.18、TadA*8.19、TadA*8.20、TadA*8.21、TadA*8.22、TadA*8.23或TadA*8.24。在一些实施方案中,胞苷脱氨酶能够将DNA中的胞苷脱氨基。在一些实施方案中,胞苷脱氨酶是APOBEC或其衍生物。在一些实施方案中,碱基编辑器进一步包含一种或多种尿嘧啶糖苷酶抑制剂(UGI)。在一些实施方案中,碱基编辑器不包含尿嘧啶糖苷酶抑制剂(UGI)。在一些实施方案中,一种或多种向导RNA包含CRISPR RNA(crRNA)和反式编码小RNA(tracrRNA),其中crRNA包含与HBV核酸序列互补的核酸序列。在一些实施方案中,碱基编辑器与单向导RNA(sgRNA)复合,所述sgRNA包含与HBV核酸序列互补的核酸序列。在一些实施方案中,sgRNA包含核酸序列,所述核酸序列包含与所述HBV核酸序列互补的至少10个连续核苷酸。在一些实施方案中,sgRNA包含核酸序列,所述核酸序列包含与HBV核酸序列互补的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或40个连续核苷酸。

[0016] 在一些实施方案中,上述方法包括编辑一个或多个核碱基。在一些实施方案中,上述方法包含两种或更多种向导RNA,所述向导RNA靶向两种或更多种HBV核酸序列。在一些实施方案中,上述方法包含两种或更多种向导RNA,所述向导RNA靶向三种、四种或五种HBV核酸序列。在上述方法的一些实施方案中,HBV核酸序列编码选自HBV聚合酶、HBV核心区蛋白、HBV S蛋白、HBV X蛋白或其组合的一种或多种HBV蛋白。在上述方法的一些实施方案中,一种或多种向导RNA包含选自下列中的一者或多者的从5'至3'的序列,或者其1、2、3、4或5个核苷酸5'截短片段:UCAAUCCCAACAAGGACACC、GGGAACAAGAUCUACAGCAU、AAGCCCAGGAUGAUGGGAUG、CUGCCAACUGGAUCCUGCGC、GACACAUCCAGCGAUAACCA、GCUGCCAACUGGAUCCUGCG、UAUGGAUGAUGUGGUAUUGG、CCAUGCCCCAAAGCCACCCA、AAGCCACCCAAGGCACAGCU、GAGAAGUCCACCACGAGUCU、CUUCUCUCAAUUUUCUAGGG、GACGACGAGGCAGGUCCCCU、CCCAACAAGGACACCUGGCC、UGCCAACUGGAUCCUGCGCG、AGGAGUUCCGCAGUAUGGAU、CCGCAGUAUGGAUCGGCAGA、CCUCUGCCGAUCCAUACUGC、CGCCCACCGAAUGUUGCCCA、GACUUCUCUCAAUUUUCUAG、GUUCCGCAGUAUGGAUCGGC、UACUAACAUUGAGGUUCCCG、UCCGCAGUAUGGAUCGGCAG、UCCUCUGCCGAUCCAUACUG、GUAGCUCCAAAUUCUUUAUA或AAUCCACACUCCGAAAGACA。在一些实施方案中,编码HBV蛋白的多核苷酸的改变是提前终止密码子。在一些实施方案中,核酸序列的改变导致由该核酸编码的HBV X蛋白中的R87*或W120*。在一些实施方案中,核酸序列的改变导致由该核酸编码的HBV S蛋白中的W35*或W36*。在一些实施方案中,编码HBV蛋白的多核苷酸的改变是错义突变。在一些实施方案中,错义突变位于HBV pol基因中。在一些实施方案中,HBV pol基因中的错义突变导致由HBV pol基因编码的HBV聚合酶中的E24G、L25F、P26F、R27C、V48A、V48I、S382F、V378I、V378A、V379I、V379A、L377F、D380G、D380N、F381P、R376G、A422T、F423P、A432V、M433V、P434S、D540G、A688V、D689G、A717T、E718K、P713S、P713L或L719P。在一些实施方案中,错义突变位于HBV核心区基因中。在一些实施方案中,HBV核心区基因中的错义突变导致由HBV核心区基因编码的HBV核心区蛋白中的T160A、T160A、P161F、S162L、C183R或*184Q。在一些实施方案中,错义突变位于HBV X基因中。在一些实施方案中,错义突变导致由HBV X基因编码的HBV X蛋白中的H86R、W120R、E122K、E121K或L141P。在一些实施方案中,错义突变位于HBV S基因中。在一些实施方案中,错义突变导致由HBV S基因编码的HBV S蛋白中的S38F、L39F、W35R、W36R、T37I、T37A、R78Q、S34L、F80P或D33G。在一些实施方案中,碱基编辑

器是BE4或BE4的变体,其中APOBEC-1被替换为APOBEC-3A的序列,且/或Cas9被替换为包含V1134、R1217、Q1334和R1336的Cas9变体(称为SpCas9-VRQR)。

[0017] 一方面,提供了例如用于治疗HBV感染的组合物。一方面,提供了一种组合物,其中该组合物包含与向导RNA结合的碱基编辑器,其中所述向导RNA包含与HBV基因互补的核酸序列。在一种实施方案中,碱基编辑器是腺苷碱基编辑器或胞苷脱氨酶。在一种实施方案中,腺苷脱氨酶结构域能够将脱氧核糖核酸(DNA)中的腺嘌呤脱氨基。在一种实施方案中,腺苷脱氨酶是TadA脱氨酶。在一种实施方案中,TadA脱氨酶是TadA*7.10、TadA*8.1、TadA*8.2、TadA*8.3、TadA*8.4、TadA*8.5、TadA*8.6、TadA*8.7、TadA*8.8、TadA*8.9、TadA*8.10、TadA*8.11、TadA*8.12、TadA*8.13、TadA*8.14、TadA*8.15、TadA*8.16、TadA*8.17、TadA*8.18、TadA*8.19、TadA*8.20、TadA*8.21、TadA*8.22、TadA*8.23或TadA*8.24。在一种实施方案中,胞苷脱氨酶能够将DNA中的胞苷脱氨基。在一种实施方案中,胞苷脱氨酶是APOBEC或其衍生物。在一种实施方案中,碱基编辑器进一步包含一种或多种尿嘧啶糖苷酶抑制剂(UGI)。在一种实施方案中,碱基编辑器不包含尿嘧啶糖苷酶抑制剂(UGI)。在一种实施方案中,碱基编辑器(i)包含Cas9切口酶;

[0018] (ii)包含无核酸酶活性的Cas9;

[0019] (iii)不包含UGI结构域;

[0020] (iv)包含APOBEC-1或APOBEC-3A胞苷脱氨酶;

[0021] (v)包含与下述序列为至少80%、85%、90%、95%、96%、97%、98%、99%或100%相同的氨基酸序列:

[0022] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDN

EQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFD
TTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSKRTADGSEFESPKKKRKVE;或者

[0023] (vi)包含与下述序列为至少80％、85％、90％、95％、96％、97％、98％、99％或100％相同的氨基酸序列：

[0024] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIE
KFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIM
TEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPP
HILWATGLKSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLG
NTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQL
VQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKL
QLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVR
QQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHL
GELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFI
ERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDY
FKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL
HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQIL
KEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDK
LIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAK
SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEK
NPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDN
EQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFD
TTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSKRTADGSEFESPKKKRKVE。

[0025] 在一种实施方案中,该组合物的向导RNA包含与编码HBV聚合酶、HBV核心区蛋白、HBV S蛋白或HBV X蛋白的HBV基因互补的核酸序列。在一种实施方案中,向导RNA包含与编码HBV X蛋白的HBV基因完美互补的15、16、17、18、19、20、21、22、23、24、25、26、27、28、29或30个连续核苷酸。在一种实施方案中,向导RNA包含与编码HBV S蛋白的HBV基因完美互补的15、16、17、18、19、20、21、22、23、24、25、26、27、28、29或30个连续核苷酸。在一种实施方案中,向导RNA包含与编码HBV聚合酶的HBV基因完美互补的15、16、17、18、19、20、21、22、23、24、25、26、27、28、29或30个连续核苷酸。在一种实施方案中,向导RNA包含与编码HBV核心区蛋白的HBV基因完美互补的15、16、17、18、19、20、21、22、23、24、25、26、27、28、29或30个连续核苷酸。在一种实施方案中,向导RNA包含来自选自由以下所组成的组的核酸的5’末端1、2、3、4或5个核酸的截短:从5’至3’,UCAAUCCCAACAAGGACACC;GGGAACAAGAUCUACAGCAU;AAGCCCAGGAUGAUGGGAUG;CUGCCAACUGGAUCCUGCGC;GACACAUCCAGCGAUAACCA;GCUGCCAACUGGAUCCUGCG;UAUGGAUGAUGUGGUAUUGG;CCAUGCCCCAAAGCCACCCA;AAGCCACCCAAGGCACAGCU;GAGAAGUCCACCACGAGUCU;CUUCUCUCAAUUUUCUAGGG;

GACGACGAGGCAGGUCCCCU;CCCAACAAGGACACCUGGCC;UGCCAACUGGAUCCUGCGCG;AGGAGUUCCGCAGUAUGGAU;CCGCAGUAUGGAUCGGCAGA;CCUCUGCCGAUCCAUACUGC;CGCCCACCGAAUGUUGCCCA;GACUUCUCUCAAUUUUCUAG;GUUCCGCAGUAUGGAUCGGC;UACUAACAUUGAGGUUCCCG;UCCGCAGUAUGGAUCGGCAG;UCCUCUGCCGAUCCAUACUG;GUAGCUCCAAAUUCUUUAUA和AAUCCACACUCCGAAAGACA。在一种实施方案中,向导RNA包含选自由以下所组成的组的核酸:从5'至3',UCAAUCCCAACAAGGACACC;GGGAACAAGAUCUACAGCAU;AAGCCCAGGAUGAUGGGAUG;CUGCCAACUGGAUCCUGCGC;GACACAUCCAGCGAUAACCA;GCUGCCAACUGGAUCCUGCG;UAUGGAUGAUGUGGUAUUGG;CCAUGCCCCAAAGCCACCCA;AAGCCACCCAAGGCACAGCU;GAGAAGUCCACCACGAGUCU;CUUCUCUCAAUUUUCUAGGG;GACGACGAGGCAGGUCCCCU;CCCAACAAGGACACCUGGCC;UGCCAACUGGAUCCUGCGCG;AGGAGUUCCGCAGUAUGGAU;CCGCAGUAUGGAUCGGCAGA;CCUCUGCCGAUCCAUACUGC;CGCCCACCGAAUGUUGCCCA;GACUUCUCUCAAUUUUCUAG;GUUCCGCAGUAUGGAUCGGC;UACUAACAUUGAGGUUCCCG;UCCGCAGUAUGGAUCGGCAG;UCCUCUGCCGAUCCAUACUG;GUAGCUCCAAAUUCUUUAUA和AAUCCACACUCCGAAAGACA。在一种实施方案中,上述组合物进一步包含脂质。在一种实施方案中,脂质是阳离子脂质。在一种实施方案中,组合物进一步包含药学可接受的赋形剂。

[0026] 另一方面,提供了一种药物组合物,其中该药物组合物包含位于药学上可接受的赋形剂中的碱基编辑器或编码所述碱基编辑器的核酸以及一种或多种RNA(gRNA),所述gRNA包含与HBV基因互补的核酸序列。在一种实施方案中,碱基编辑器(i)包含Cas9切口酶;

[0027] (ii)包含无核酸酶活性的Cas9;

[0028] (iii)不包含UGI结构域;

[0029] (iv)包含APOBEC-1或APOBEC-3A胞苷脱氨酶;

[0030] (v)包含与下述序列为至少80%、85%、90%、95%、96%、97%、98%、99%或100%相同的氨基酸序列:

[0031] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS

EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDK
LIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAK
SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEK
NPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDN
EQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFD
TTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSKRTADGSEFESPKKKRKVE;或者

[0032] (vi)包含与下述序列为至少80％、85％、90％、95％、96％、97％、98％、99％或100％相同的氨基酸序列：

[0033] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIE
KFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIM
TEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPP
HILWATGLKSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLG
NTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQL
VQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKL
QLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVR
QQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHL
GELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFI
ERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDY
FKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL
HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQIL
KEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDK
LIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAK
SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEK
NPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDN
EQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFD
TTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSKRTADGSEFESPKKKRKVE。

[0034] 在一种实施方案中,碱基编辑器包含Cas9或包含V1134、R1217、Q1334和R1336的Cas9变体(SpCas9-VRQR)。在一种实施方案中,gRNA和碱基编辑器配制在一起或单独配制。在一种实施方案中,向导RNA包含选自以下一者或多者的从5’至3’的序列,或其1、2、3、4或5个核苷酸5’截短片段：

[0035] UCAAUCCCAACAAGGACACC;GGGAACAAGAUCUACAGCAU;AAGCCCAGGAUGAUGGGAUG;
CUGCCAACUGGAUCCUGCGC;GACACAUCCAGCGAUAACCA;GCUGCCAACUGGAUCCUGCG;
UAUGGAUGAUGUGGUAUUGG;CCAUGCCCCAAAGCCACCCA;AAGCCACCCAAGGCACAGCU;
GAGAAGUCCACCACGAGUCU;CUUCUCUCAAUUUUCUAGGG;GACGACGAGGCAGGUCCCCU;

CCCAACAAGGACACCUGGCC;UGCCAACUGGAUCCUGCGCG;AGGAGUUCCGCAGUAUGGAU;CCGCAGUAUGGAUCGGCAGA;CCUCUGCCGAUCCAUACUGC;CGCCCACCGAAUGUUGCCCA;GACUUCUCUCAAUUUUCUAG;GUUCCGCAGUAUGGAUCGGC;UACUAACAUUGAGGUUCCCG;UCCGCAGUAUGGAUCGGCAG;UCCUCUGCCGAUCCAUACUG;GUAGCUCCAAAUUCUUUAUA或AAUCCACACUCCGAAAGACA。在药物组合物的一种实施方案中,进一步包含适用于在哺乳动物细胞中表达的载体,其中载体包含编码碱基编辑器的多核苷酸。在一种实施方案中,载体是病毒载体。在一种实施方案中,病毒载体是逆转录病毒载体、腺病毒载体、慢病毒载体、疱疹病毒载体或腺相关病毒载体(AAV)。在一种实施方案中,药物组合物进一步包含适用于在哺乳动物细胞中表达的核糖核粒子。

[0036] 另一方面,提供了一种治疗HBV感染的方法,其中该方法包括向有此需要的受试者给药如上述组合物或药物组合物。

[0037] 另一方面,提供了一种HBV基因组,其包含选自由以下组成的组的改变:

[0038] 提前终止密码子,其将R87STOP或W120STOP引入X基因内;

[0039] 提前终止密码子,其将W35STOP或W36STOP引入S基因内;

[0040] 所述HBV pol基因中的错义突变,其将E24G、L25F、P26F、R27C、V48A、V48I、S382F、V378I、V378A、V379I、V379A、L377F、D380G、D380N、F381P、R376G、A422T、F423P、A432V、M433V、P434S、D540G、A688V、D689G、A717T、E718K、P713S、P713L或L719P引入HBV聚合酶中;

[0041] 所述HBV核心区基因中的错义突变,其将T160A、T160A、P161F、S162L、C183R或STOP184Q引入HBV核心区多肽中;

[0042] 所述X基因中的错义突变,其将H86R、W120R、E122K、E121K或L141P引入HBV X多肽中;和

[0043] 所述S基因中的错义突变,其将S38F、L39F、W35R、W36R、T37I、T37A、R78Q、S34L、F80P或D33G引入HBV S多肽中。

[0044] 在一些实施方案中,HBV基因组包含上述改变中的两种或更多种。

[0045] 在上述方法或上述HBV基因组的实施方案中,HBV是基因型C或基因型D。

[0046] 另一方面,提供了上述任何方面和实施方案中所述的组合物用于治疗受试者的HBV感染的用途。

[0047] 另一方面,提供了上述任何方面和实施方案中所述的药物组合物用于治疗受试者的HBV感染的用途。

[0048] 在上述用途的一种实施方案中,受试者是哺乳动物。在上述用途的一种实施方案中,受试者是人。

[0049] 在上述方法或药物组合物的一种实施方案中,一种或多种向导RNA如表26中所列。

[0050] 另一方面,提供了一种向导RNA(gRNA),其包含与HBV基因互补的核酸序列。在一种实施方案中,向导RNA包含与编码HBV聚合酶、HBV核心区蛋白、HBV S蛋白或HBV X蛋白的HBV基因互补的核酸序列。在一种实施方案中,向导RNA包含与编码HBV X蛋白的HBV基因完美互补的15、16、17、18、19、20、21、22、23、24、25、26、27、28、29或30个连续核苷酸。在一种实施方案中,向导RNA包含与编码HBV S蛋白的HBV基因完美互补的15、16、17、18、19、20、21、22、23、24、25、26、27、28、29或30个连续核苷酸。在一种实施方案中,向导RNA包含与编码HBV聚合酶的HBV基因完美互补的15、16、17、18、19、20、21、22、23、24、25、26、27、28、29或30个连续核苷

酸。在一种实施方案中,向导RNA包含与编码HBV核心区蛋白的HBV基因完美互补的15、16、17、18、19、20、21、22、23、24、25、26、27、28、29或30个连续核苷酸。在一种实施方案中,向导RNA包含来自选自由以下所组成的组的核酸的5'末端1、2、3、4或5个核酸的截短:从5'至3',UCAAUCCCAACAAGGACACC;GGGAACAAGAUCUACAGCAU;AAGCCCAGGAUGAUGGGAUG;CUGCCAACUGGAUCCUGCGC;GACACAUCCAGCGAUAACCA;GCUGCCAACUGGAUCCUGCG;UAUGGAUGAUGUGGUAUUGG;CCAUGCCCCAAAGCCACCCA;AAGCCACCCAAGGCACAGCU;GAGAAGUCCACCACGAGUCU;CUUCUCUCAAUUUUCUAGGG;GACGACGAGGCAGGUCCCCU;CCCAACAAGGACACCUGGCC;UGCCAACUGGAUCCUGCGCG;AGGAGUUCCGCAGUAUGGAU;CCGCAGUAUGGAUCGGCAGA;CCUCUGCCGAUCCAUACUGC;CGCCCACCGAAUGUUGCCCA;GACUUCUCUCAAUUUUCUAG;GUUCCGCAGUAUGGAUCGGC;UACUAACAUUGAGGUUCCCG;UCCGCAGUAUGGAUCGGCAG;UCCUCUGCCGAUCCAUACUG;GUAGCUCCAAAUUCUUUAUA和AAUCCACACUCCGAAAGACA。在一种实施方案中,向导RNA包含选自由以下所组成的组的核酸:从5'至3',UCAAUCCCAACAAGGACACC;GGGAACAAGAUCUACAGCAU;AAGCCCAGGAUGAUGGGAUG;CUGCCAACUGGAUCCUGCGC;GACACAUCCAGCGAUAACCA;GCUGCCAACUGGAUCCUGCG;UAUGGAUGAUGUGGUAUUGG;CCAUGCCCCAAAGCCACCCA;AAGCCACCCAAGGCACAGCU;GAGAAGUCCACCACGAGUCU;CUUCUCUCAAUUUUCUAGGG;GACGACGAGGCAGGUCCCCU;CCCAACAAGGACACCUGGCC;UGCCAACUGGAUCCUGCGCG;AGGAGUUCCGCAGUAUGGAU;CCGCAGUAUGGAUCGGCAGA;CCUCUGCCGAUCCAUACUGC;CGCCCACCGAAUGUUGCCCA;GACUUCUCUCAAUUUUCUAG;GUUCCGCAGUAUGGAUCGGC;UACUAACAUUGAGGUUCCCG;UCCGCAGUAUGGAUCGGCAG;UCCUCUGCCGAUCCAUACUG;GUAGCUCCAAAUUCUUUAUA和AAUCCACACUCCGAAAGACA。

[0051] 另一方面,提供了一种药物组合物,其中该药物组合物包含:(i)编码碱基编辑器的核酸;和(ii)上述任何方面和实施方案中所述的向导RNA。在一种实施方案中,药物组合物进一步包含脂质。在药物组合物的一种实施方案中,编码碱基编辑器的核酸是mRNA。

[0052] 本发明定义的组合物和物品是关于下文所提供的实施例而分离或另外制造的。从具体实施方式和权利要求书可明了本发明的其它特征和优点。

[0053] 定义

[0054] 以下定义补充了本领域中的定义且针对本申请,并且不归咎于任何相关或不相关的情况,例如,任何共同拥有的专利或申请。尽管与本文中揭示者类似或等效的方法和材料可用于测试本公开的实践中,但优选的方法和材料在本文中描述。因此,本文所使用的技术仅用于描述特定实施方案的目的,而非试图限制。

[0055] 除非另做定义,否则本文中使用的所有科技术语均具有本发明所属领域技术人员所一般理解的意义。下述参考文献对技术人员提供本发明中使用的多个术语的一般性定义:Singleton等人所著《微生物学和分子生物学词典(第二版)》(Singleton et al., Dictionary of Microbiology and Molecular Biology(2nd ed.1994));《剑桥科技词典》(The Cambridge Dictionary of Science and Technology(Walker ed.,1988));Rieger等人编撰的《遗传性术语表(第五版)》(The Glossary of Genetics,5th Ed.,R.Rieger et al.(eds.),Springer Verlag(1991));以及Hale和Marham所著《哈珀·柯林斯生物学词典》(Hale&Marham,The Harper Collins Dictionary of Biology(1991)).如本文中所用,除

非明确排除,否则下述术语具有归于其下方的意义。

[0056] 在本申请中,除非具体地另做指定,否则单数的使用包括复数。必须注意,如本说明书中所用,除非上下文中明确排除,否则单数形式“一”和“该/所述”包括复数个对象。在本申请中,除非另做指定,否则“或”的使用意为“和/或”。此外,术语“包括”以及其他形式(诸如主动形式和被动形式)不是限制性的。

[0057] 如本说明书和权利要求书中所用,词语“包含”(以及其他形式的包含,诸如复数形式和单数形式)、“具有”(以及其他形式的具有,诸如复数形式和单数形式)、“包括”(以及其他形式的包括,诸如复数形式和单数形式)或“含有”(以及其他形式的含有,诸如单数形式和复数形式)是包含性的或开放性的,并且不排除另外的、未引用的元件或方法步骤。设想本说明书中讨论的任何实施方案可以关于本公开的任何方法或组合物来实现,反之亦然。此外,本公开的组合物可用来实现本公开的方法。

[0058] 术语“约”或“大约”意为处于特定数值的可接受的误差范围内,如由本领域技术人员所确定的,其将部分地取决于该数值如何测量或测定,即,测量系统的限制。例如,根据本领域的实践,“约”可以意为处于1个或超过1个标准偏差内。另选地,“约”可以意为给定值的至多20%、至多10%、至多5%或至多1%的范围。另选地,特别是关于生物系统或过程,该术语可以意为处于数量级内,例如处于数值的5倍以内或2倍以内。如果在申请和权利要求中描述特定数值,除非另做指定,否则应假设术语“约”意为处于该特定数值的可接受的误差范围内。

[0059] 说明书中对“一些实施方案”、“实施方案”、“一个实施方案”或“其他实施方案”的引用意味着,与这些实施方案关联描述的特定特征、结构或特点包括在本公开的至少一些实施方案中,但不必包括在全部实施方案中。

[0060] “腺苷脱氨酶”意为能够催化腺嘌呤或腺苷的水解性脱氨反应的多肽或其片段。在一些实施方案中,脱氨酶或脱氨酶结构域是催化腺苷到肌苷的水解性脱氨反应或腺苷到脱氧肌苷的脱氧反应的腺苷脱氨酶。在一些实施方案中,腺苷脱氨酶催化脱氧核糖核酸(DNA)中的腺嘌呤或腺苷的水解性脱氨反应。本文提供的腺苷脱氨酶(例如,经工程化的腺苷脱氨酶、进化的腺苷脱氨酶)可以来自诸如细菌的任何生物体。

[0061] 在一些实施方案中,脱氨酶或脱氨酶结构域是来自生物体,诸如人、黑猩猩、大猩猩、猴、牛、狗、大鼠或小鼠的天然出现脱氨酶的变体。在一些实施方案中,脱氨酶或脱氨酶结构域不出现在自然界中。例如,在一些实施方案中,脱氨酶或脱氨酶结构域列与天然出现的脱氨酶为至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相同。在一些实施方案中,腺苷脱氨酶来自细菌,诸如大肠杆菌(Escherichia coli)、金黄色葡萄球菌(Staphylococcus aureus)、鼠伤寒沙门氏菌(Salmonella typhimurium)、腐败希瓦氏菌(Shewanella putrefaciens)、流感嗜血杆菌(Haemophilus influenzae)或新月柄杆菌(Caulobacter crescentus)。在一些实施方案中,腺苷脱氨酶是TadA脱氨酶。在一些实施方案中,TadA脱氨酶是大肠杆菌TadA(ecTadA)脱氨酶或其片段。

[0062] 例如,脱氨酶结构域在国际PCT申请号PCT/2017/045381(WO 2018/027078)和PCT/US2016/058344(WO 2017/070632)中描述,其各自通过引用而整体并入本文。也参见,Komor,A.C.,et al.,“Programmable editing of a target base in genomic DNA

without double-stranded DNA cleavage"Nature 533,420-424(2016);Gaudelli,N.M.,et al.,"Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage"Nature 551,464-471(2017);Komor,A.C.,et al.,"Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:Abase editors with higher efficiency and product purity"Science Advances 3:eaao4774(2017))和Rees,H.A.,et al.,"Base editing:precision chemistry on the genome and transcriptome of living cells."Nat Rev Genet.2018Dec;19(12):770-788.doi:10.1038/s41576-018-0059-1,其整体内容通过引用并入本文。

[0063] 野生型TadA(wt)腺苷脱氨酶具有以下序列(也称为TadA参考序列):

[0064] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD。

[0065] 在一些实施方案中,腺苷脱氨酶包含以下序列中的改变:

[0066] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD(也称为TadA*7.10)。

[0067] 在一些实施方案中,TadA*7.10包含至少一个改变。在一些实施方案中,TadA*7.10包含位于氨基酸82和/或166处的改变。在特定实施方案中,上文述参考的序列的变体包含以下改变中的一种或多种:Y147T、Y147R、Q154S、Y123H、V82S、T166R和/或Q154R。本文中,改变Y123H称为改变H123H,在TadA*7.10中被修复为Y123H TadA(wt)。在其他实施方案中,TadA*7.10序列的变体包含选自由以下的组的改变的组合:Y147T+Q154R;Y147T+Q154S;Y147R+Q154S;V82S+Q154S;V82S+Y147R;V82S+Q154R;V82S+Y123H;I76Y+V82S;V82S+Y123H+Y147T;V82S+Y123H+Y147R;V82S+Y123H+Q154R;Y147R+Q154R+Y123H;Y147R+Q154R+I76Y;Y147R+Q154R+T166R;Y123H+Y147R+Q154R+I76Y;V82S+Y123H+Y147R+Q154R;和I76Y+V82S+Y123H+Y147R+Q154R。

[0068] 在其他实施方案中,本发明提供包括相对于TadA*7.10的缺失的腺苷脱氨酶变体,例如TadA*8,该腺苷脱氨酶变体包含C端的缺失,该缺失起始于残基149、150、151、152、153、154、155、156或157,或另一TadA中的对应突变。在其他实施方案中,腺苷脱氨酶变体是TadA(例如,TadA*8)单体,其包含以下改变中的一种或多种:相对于TadA*7.10(TadA参考序列),Y147T、Y147R、Q154S、Y123H、V82S、T166R和/或Q154R,或另一TadA中的相应突变。在其他实施方案中,腺苷脱氨酶变体是TadA(例如,TadA*8)单体,其包含以下改变:相对于TadA*7.10(TadA参考序列),Y147T+Q154R、Y147T+Q154S、Y147R+Q154S、V82S+Q154S、V82S+Y147R、V82S+Q154R、V82S+Y123H、I76Y+V82S、V82S+Y123H+Y147T、V82S+Y123H+Y147R、V82S+Y123H+Q154R、Y147R+Q154R+Y123H、Y147R+Q154R+I76Y、Y147R+Q154R+T166R、Y123H+Y147R+Q154R+I76Y、V82S+Y123H+Y147R+Q154R、和I76Y+V82S+Y123H+Y147R+Q154R,或另一TadA中的相应突变。

[0069] 在又其他实施方案中,腺苷脱氨酶变体是同源二聚体,其包含两个腺苷脱氨酶结构域,每个腺苷脱氨酶结构域具有以下突变中的一种或多种:相对于TadA*7.10(TadA参考序列),Y147T、Y147R、Q154S、Y123H、V82S、T166R和/或Q154R,或另一TadA中的相应突变。在

其他实施方案中,腺苷脱氨酶变体是同源二聚体,其包含两个腺苷脱氨酶结构域(例如,TadA*8),每个腺苷脱氨酶结构域具有选自以下的组的改变的组合:相对于TadA*7.10(TadA参考序列),Y147T+Q154R、Y147T+Q154S、Y147R+Q154S、V82S+Q154S、V82S+Y147R、V82S+Q154R、V82S+Y123H、I76Y+V82S、V82S+Y123H+Y147T、V82S+Y123H+Y147R、V82S+Y123H+Q154R、Y147R+Q154R+Y123H、Y147R+Q154R+I76Y、Y147R+Q154R+T166R、Y123H+Y147R+Q154R+I76Y、V82S+Y123H+Y147R+Q154R、和I76Y+V82S+Y123H+Y147R+Q154R,或另一TadA中的相应突变。在其他实施方案中,腺苷脱氨酶变体是异源二聚体,其包含野生型TadA腺苷脱氨酶结构域和腺苷脱氨酶变体结构域(例如,TadA*8),所述腺苷脱氨酶变体结构域包含以下突变中的一种或多种:相对于TadA*7.10(TadA参考序列),Y147T、Y147R、Q154S、Y123H、V82S、T166R和/或Q154R,或另一TadA中的相应突变。在其他实施方案中,腺苷脱氨酶变体是异源二聚体,其包含野生型TadA腺苷脱氨酶结构域和腺苷脱氨酶变体结构域(例如,TadA*8),所述腺苷脱氨酶变体结构域包含选自以下的组的改变的组合:相对于TadA*7.10(TadA参考序列),Y147T+Q154R、Y147T+Q154S、Y147R+Q154S、V82S+Q154S、V82S+Y147R、V82S+Q154R、V82S+Y123H、I76Y+V82S、V82S+Y123H+Y147T、V82S+Y123H+Y147R、V82S+Y123H+Q154R、Y147R+Q154R+Y123H、Y147R+Q154R+I76Y、Y147R+Q154R+T166R、Y123H+Y147R+Q154R+I76Y、V82S+Y123H+Y147R+Q154R、和I76Y+V82S+Y123H+Y147R+Q154R,或另一TadA中的相应突变。

[0070] 在其他实施方案中,腺苷脱氨酶变体是异源二聚体,其包含TadA*7.10结构域和腺苷脱氨酶变体结构域(例如,TadA*8),所述腺苷脱氨酶变体结构域包含以下突变中的一种或多种:相对于TadA*7.10(TadA参考序列),Y147T、Y147R、Q154S、Y123H、V82S、T166R和/或Q154R,或另一TadA中的相应突变。在其他实施方案中,腺苷脱氨酶变体是异源二聚体,其包含TadA*7.10结构域和腺苷脱氨酶变体结构域(例如,TadA*8),所述腺苷脱氨酶变体结构域包含以下改变的组合:相对于TadA*7.10(TadA参考序列),Y147T+Q154R、Y147T+Q154S、Y147R+Q154S、V82S+Q154S、V82S+Y147R、V82S+Q154R、V82S+Y123H、I76Y+V82S、V82S+Y123H+Y147T、V82S+Y123H+Y147R、V82S+Y123H+Q154R、Y147R+Q154R+Y123H、Y147R+Q154R+I76Y、Y147R+Q154R+T166R、Y123H+Y147R+Q154R+I76Y、V82S+Y123H+Y147R+Q154R、或I76Y+V82S+Y123H+Y147R+Q154R,或另一TadA中的相应突变。在一个实施方案中,腺苷脱氨酶是TadA*8,其包含以下序列或其具有腺苷脱氨酶活性的片段或基本上由其组成:

[0071] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD。

[0072] 在一些实施方案中,TadA*8是经截短的。在一些实施方案中,相对于全长度TadA*8,经截短的TadA*8失去了1、2、3、4、5,6、7、8、9、10、11、12、13、14、15、6、17、18、19或20个N端氨基酸残基。在一些实施方案中,相对于全长度TadA*8,经截短的TadA*8失去了1、2、3、4、5,6、7、8、9、10、11、12、13、14、15、6、17、18、19或20个C端氨基酸残基。在一些实施方案中,腺苷脱氨酶变体是全长度TadA*8。在特定实施方案中,腺苷脱氨酶异源二聚体包含TadA*8结构域和选自以下的腺苷脱氨酶结构域:

[0073] 金黄色葡萄球菌(Staphylococcus aureus(S.aureus))TadA:

[0074] MGSHMTNDIYFMTLAIEEAKKAAQLGEVPIGAIITKDDEVIARAHNLRETLQQPTAHAEHIAIERAAKVLGSWRLEGCTLYVTLEPCVMCAGTIVMSRIPRVVYGADDPKGGCSGSLMNLLQQSNFNHRAIVDKGVLKEACSTLLT

TFFKNLRANKKSTN

[0075] 枯草芽孢杆菌(Bacillus subtilis(B.subtilis))TadA:

[0076] MTQDELYMKEAIKEAKKAEEKGEVPIGAVLVINGEIIARAHNLRETEQRSIAHAEMLVIDEACKALGTW RLEGATLYVTLEPCPMCAGAVVLSRVEKVVFGAFDPKGGCSGTLMNLLQEERFNHQAEVVSGVLEEECGGMLSAFFR ELRKKKKAARKNLSE

[0077] 鼠伤寒沙门氏菌(Salmonella typhimurium(S.typhimurium))TadA:

[0078] MPPAFITGVTSLSDVELDHEYWMRHALTLAKRAWDEREVPVGAVLVHNHRVIGEGWNRPIGRHDPTAHA EIMALRQGGLVLQNYRLLDTTLYVTLEPCVMCAGAMVHSRIGRVVFGARDAKTGAAGSLIDVLHHPGMNHRVEIIEG VLRDECATLLSDFFRMRRQEIKALKKADRAEGAGPAV

[0079] 腐败希瓦氏菌(Shewanella putrefaciens(S.putrefaciens))TadA:

[0080] MDEYWMQVAMQMAEKAEAAGEVPVGAVLVKDGQQIATGYNLSISQHDPTAHAEILCLRSAGKKLENYRL LDATLYITLEPCAMCAGAMVHSRIARVVYGARDEKTGAAGTVVNLLQHPAFNHQVEVTSGVLAEACSAQLSRFFKRR RDEKKALKLAQRAQQGIE

[0081] 流感嗜血杆菌(Haemophilus influenzae)F3031(H.influenzae)TadA:

[0082] MDAAKVRSEFDEKMMRYALELADKAEALGEIPVGAVLVDDARNIIGEGWNLSIVQSDPTAHAEIIA LRNGAKNIQNYRLLNSTLYVTLEPCTMCAGAILHSRIKRLVFGASDYKTGAIGSRFHFFDDYKMNHTLEITSGVLAE ECSQKLSTFFQKRREEKKIEKALLKSLSDK

[0083] 新月柄杆菌(Caulobacter crescentus(C.crescentus))TadA:

[0084] MRTDESEDQDHRMMRLALDAARAAAEAGETPVGAVILDPSTGEVIATAGNGPIAAHDPTAHAEIAAMRA AAAKLGNYRLTDLTLVVTLEPCAMCAGAISHARIGRVVFGADDPKGGAVVHGPKFFAQPTCHWRPEVTGGVLADESA DLLRGFFRARRKAKI

[0085] 硫还原地杆菌(Geobacter sulfurreducens(G.sulfurreducens))TadA:

[0086] MSSLKKTPIRDDAYWMGKAIREAAKAAARDEVPIGAVIVRDGAVIGRGHNLREGSNDPSAHAEMIAIR QAARRSANWRLTGATLYVTLEPCLMCMGAIILARLERVVFGCYDPKGGAAGSLYDLSADPRLNHQVRLSPGVCQEE CGTMLSDFFRDLRRRKKAKATPALF IDERKVPPEPTadA*7.10

[0087] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGL VMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAAL LCYFFRMPRQVFNAQKKAQSSTD。

[0088] “腺苷脱氨酶碱基编辑器8(ABE8)多肽”或“ABE8”意为一种如本文所定义的碱基编辑器,其包含腺苷脱氨酶变体,该变体包含位于以下参考序列的氨基酸位置82和/或166处的改变:MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC YFFRMPRQVFNAQKKAQSSTD

[0089] 在一些实施方案中,ABE8包含进一步的改变,如本文所述,相对于参考序列。

[0090] “腺苷脱氨酶碱基编辑器8(ABE8)多核苷酸”意为编码ABE8的多核苷酸。

[0091] 本文中,“给药”是指向患者或受试者提供本文所述的一种或多种组合物。举例而言且并非限制,可通过静脉内(i.v.)注射、皮下(s.c.)注射、皮内(i.d.)注射、腹膜内(i.p.)注射或肌肉内(i.m.)注射来执行组合物给药(例如,注射)。可采用一种或多种此类途径。例如,可通过单次快速注射或通过随时间渐进灌注进行肠胃外给药。另选地,或同时,

可通过口服途径给药。

[0092] “剂”意为任何小分子化学化合物、抗体、核酸分子、或肽、或其片段。

[0093] “改变”意为基因或多肽的表达水平或活性的变化(增加或减少),如通过该领域已知的方法如本文所述的那些方法所检测。如本文中所用,改变包括表达水平变化至少10%、变化25%、变化40%,以及表达水平变化50%或更高。

[0094] “缓解”意为减少、阻抑、衰减、缩减、迟滞或稳定化疾病的发展或进展。

[0095] “类似物”意为不相同但具有类似的功能或结构特征的分子。例如,多肽类似物保留相对应的天然多肽的生物学活性,但具有某些令该类似物功能相对于天然多肽增强的生化修饰。此类生化修饰将会增加类似物的蛋白酶抗性、膜渗透性或半衰期,而不改变例如配体结合。类似物可以包括非天然氨基酸。

[0096] “碱基编辑器(BE)”或“核碱基编辑器(NBE)意为结合多核苷酸并且具有核碱基修饰活性的试剂。在多个实施方案中,碱基编辑器包含核碱基修饰多肽(例如,脱氨酶)和多核苷酸可编程核苷酸结合结构域,其与向导多核苷酸(例如,向导RNA)协力。在多个实施方案中,该试剂是生物分子复合物,其包含具有碱基编辑活性的蛋白结构域,即能够修饰核酸分子(例如,DNA)中的碱基(例如,A、T、C、G或U)的结构域。在一些实施方案中,多核苷酸可编程DNA结合结构域融合或连接至脱氨酶结构域。在一个实施方案中,该试剂是融合蛋白,其包含一个或多个具有碱基编辑活性的结构域。在另一实施方案中,具有碱基编辑活性的蛋白结构域连接至向导RNA(例如,经由向导RNA上的RNA结合基序和融合至脱氨酶的RNA结合结构域)。在一些实施方案中,具有碱基编辑活性的结构域能够将核酸分子内的碱基脱氨基。在一些实施方案中,碱基编辑器能够将DNA分子内的一个或多个碱基脱氨基。在一些实施方案中,碱基编辑器能够将DNA内的胞嘧啶(C)或腺苷(A)脱氨基。在一些实施方案中,碱基编辑器能够将DNA内的胞嘧啶(C)和腺苷(A)脱氨基。在一些实施方案中,碱基编辑器是胞苷编辑器(CBE)。在一些实施方案中,碱基编辑器是腺苷碱基编辑器(ABE)。在一些实施方案中,碱基编辑器是腺苷碱基编辑器(ABE)和胞苷编辑器(CBE)。在一些实施方案中,碱基编辑器是融合至腺苷脱氨酶的无核酸酶活性Cas9(dCas9)。在一些实施方案中,Cas9是环状高突变(permutant)的Cas9(例如,spCas9或saCas9)。环状高突变的Cas9是本领域中已知的,并且在例如Oakes et al.,Cell 176,254-267,2019中有所描述。在一些实施方案中,碱基编辑器融合至碱基切除修复(BER)的抑制剂,例如,UGI结构域或dISN结构域。在一些实施方案中,融合蛋白包含融合至脱氨酶的Cas9切口酶和碱基切除修复的抑制剂诸如UGI结构域或dISN结构域。在其他实施方案中,碱基编辑器是脱碱基的碱基编辑器。

[0097] 在一些实施方案中,腺苷脱氨酶是从TadA进化的。在一些实施方案中,多核苷酸可编程DNA结合结构域是CRISPR相关(例如,Cas或Cpf1)酶。在一些实施方案中,碱基编辑器是融合至脱氨酶结构域的催化死亡Cas9(dCas9)。在一些实施方案中,碱基编辑器是融合至脱氨酶结构域的Cas9切口酶(nCas9)。在一些实施方案中,碱基编辑器融合至碱基切除修复(BER)的抑制剂。在一些实施方案中,碱基切除修复的抑制剂是尿嘧啶DNA糖苷酶抑制剂(UGI)。在一些实施方案中,碱基切除修复的抑制剂是肌苷碱基切除修复抑制剂。碱基编辑器的细节在国际PCT申请号PCT/2017/045381(WO2018/027078)和PCT/US2016/058344(WO2017/070632)中描述,其各自通过引用而整体并入本文。也参见,Komor,A.C.,et al.,“Programmable editing of a target base in genomic DNA without double-stranded

DNA cleavage"Nature 533,420-424(2016);Gaudelli,N.M.,et al.,"Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage"Nature 551,464-471(2017);Komor,A.C.,et al.,"Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity"Science Advances3:eaao4774(2017))和Rees,H.A.,et al.,"Base editing:precision chemistry on the genome and transcriptome of living cells."Nat Rev Genet.2018Dec;19(12):770-788.doi:10.1038/s41576-018-0059-1,其整体内容通过引用并入本文。

[0098] 在一些实施方案中,通过将腺苷脱氨酶变体(例如,TadA*8)克隆到包括环状高突变体(permutant)Cas9(例如,spCAS9)和二分体核定位序列的支架中而产生碱基编辑器(例如,ABE8)。环状高突变的Cas9是本领域中已知的,并且在例如Oakes et al.,Cell 176,254-267,2019中有所描述。示例性的环状高突变序列详述于下,其中粗体序列指示来源于Cas9的序列,斜体序列表示连接子序列,而下划线序列表示二分体核定位序列。

[0099] CP5(具有MSP"NGC=具有规则Cas9突变如NGG的Pam变体"、PID=蛋白质相互作用结构域和"D10A"切口酶):

[0100] **EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKK**
TEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPTVAYSVLVVAKVEKGKSKKLKSVKELLGIT
IMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAKFLQKGNELALPSKYVNFLY
LASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAE
NIIHLFTLTNLGAPRAFKYFDTTIARKEYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD*GGSGGSGGS*
GGSGGSGGSGGM**DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA**
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYH
EKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEEN
PINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKD
TYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALV
RQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGS
IPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE
EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIV
DLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDI
VLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELD
INRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDN
LTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKD
FQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEGADKRTADGSE
FESPKKKRKV*

[0101] 在一些实施方案中,ABE8选自来自上文表8的碱基编辑器。在一些实施方案中,ABE8含有从TadA进化的腺苷脱氨酶变体。在一些实施方案中,ABE8的腺苷脱氨酶变体是如

上文表8中所述的TadA*8变体。在一些实施方案中,腺苷脱氨酶变体是TadA*7.10变体(例如,TadA*8),其包含选自由以下所组成的组的一种或多种改变:Y147T、Y147R、Q154S、Y123H、V82S、T166R和/或Q154R。在多个实施方案中,ABE8包含TadA*7.10变体(例如,TadA*8),其具有选自以下的组的改变组合:Y147T+Q154R;Y147T+Q154S;Y147R+Q154S;V82S+Q154S;V82S+Y147R;V82S+Q154R;V82S+Y123H、I76Y+V82S;V82S+Y123H+Y147T;V82S+Y123H+Y147R;V82S+Y123H+Q154R;Y147R+Q154R+Y123H;Y147R+Q154R+I76Y;Y147R+Q154R+T166R;Y123H+Y147R+Q154R+I76Y;V82S+Y123H+Y147R+Q154R;以及I76Y+V82S+Y123H+Y147R+Q154R。

[0102] 在一些实施方案中,ABE8是单体构建体。在一些实施方案中,ABE8是异源二聚体构建体。一些实施方式中,ABE8碱基编辑器包含序列:MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD

[0103] 举例而言,如本文所述的碱基编辑组合物、系统和方法将要使用的腺苷碱基编辑器ABE具有如下文提供的核酸序列(8877个碱基对)(Addgene,Watertown,MA.;Gaudelli NM,et al.,Nature.2017Nov 23;551(7681):464-471.doi:10.1038/nature24644;Koblan LW,et al.,Nat Biotechnol.2018Oct;36(9):843-846.doi:10.1038/nbt.4172.)。也包括与ABE核酸序列具有至少95%或更高同一性的多核苷酸序列。

[0104] ATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTAGAGATCCGCGGCCGCTAATACGACTCACTATAGGGAGAGCCGCCACCATGAAACGGACAGCCGACGGAAGCGAGTTCGAGTCACCAAAGAAGAAGCGGAAAGTCTCTGAAGTCGAGTTTAGCCACGAGTATTGGATGAGGCACGCACTGACCCTGGCAAAGCGAGCATGGGATGAAAGAGAAGTCCCCGTGGGCGCCGTGCTGGTGCACAACAATAGAGTGATCGGAGAGGGATGGAACAGGCCAATCGGCCGCCACGACCCTACCGCACACGCAGAGATCATGGCACTGAGGCAGGGAGGCCTGGTCATGCAGAATTACCGCCTGATCGATGCCACCCTGTATGTGACACTGGAGCCATGCGTGATGTGCGCAGGAGCAATGATCCACAGCAGGATCGGAAGAGTGGTGTTCGGAGCACGGGACGCCAAGACCGGCGCAGCAGGCTCCCTGATGGATGTGCTGCACCACCCCGGCATGAACCACCGGGTGGAGATCACAGAGGGAATCCTGGCAGACGAGTGCGCCGCCCTGCTGAGCGATTTCTTTAGAATGCGGAGACAGGAGATCAAGGCCCAGAAGAAGGCACAGAGCTCCACCGACTCTGGAGGATCTAGCGGAGGATCCTCTGGAAGCGAGACACCAGGCACAAGCGAGTCCGCCACACCAGAGAGCTCCGGCGGCTCCTCCGGAGGATCCTCTGAGGTGGAGTTTTCCCACGAGTACTGGATGAGACATGCCCTGACCCTGGCCAAGAGGGCACGCGATGAGAGGGAGGTGCCTGTGGGAGCCGTGCTGGTGCTGAACAATAGAGTGATCGGCGAGGGCTGGAACAGAGCCATCGGCCTGCACGACCCAACAGCCCATGCCGAAATTATGGCCCTGAGACAGGGCGGCCTGGTCATGCAGAACTACAGACTGATTGACGCCACCCTGTACGTGACATTCGAGCCTTGCGTGATGTGCGCCGGCGCCATGATCCACTCTAGGATCGGCCGCGTGGTGTTTGGCGTGAGGAACGCAAAAACCGGCGCCGCAGGCTCCCTGATGGACGTGCTGCACTACCCCGGCATGAATCACCGCGTCGAAATTACCGAGGGAATCCTGGCAGATGAATGTGCCGCCCTGCTGTGCTATTTCTTTCGGATGCCTAGACAGGTGTTCAATGCTCAGAAGAAGGCCCAGAGCTCCACCGACTCCGGAGGATCTAGCGGAGGCTCCTCTGGCTCTGAGACACCTGGCACAAGCGAGAGCGCAACACCTGAAAGCAGCGGGGGCAGCAGCGGGGGGTCAGACAAGAAGTACAGCATCGGCCTGGCCATCGGC

ACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACAC
CGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGC
TGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAG
ATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCG
GCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGA
AACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGC
CACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTA
CAACCAGCTGTTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCA
AGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCC
CTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGA
CACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGA
ACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCT
ATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAA
GTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGT
TCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGAC
CTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCT
GCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCC
CCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACC
CCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAA
GAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCA
AAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTG
CTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTC
CGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGG
ACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGAC
AGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCG
GAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCC
TGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAA
GAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCC
CGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCG
AGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAG
CGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAA
CGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGT
CCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGA
AGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCA
GCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAAC
TGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGAC
TCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCT
GGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCT
ACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTAC

AAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTA
CAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGA
CAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCC
CAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAG
CGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATT
CTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACC
ATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGA
CCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCG
AACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAG
CTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCAT
CGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACA
AGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCT
GCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCT
GATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGTGACTCTGGCGGCTCAA
AAAGAACCGCCGACGGCAGCGAATTCGAGCCCAAGAAGAAGAGGAAAGTCTAACCGGTCATCATCACCATCACCATT
GAGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCT
TCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAG
GTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTG
GGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCGATACCGTCGACCTCTAGCTAGAGC
TTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGG
AAGCATAAAGTGTAAAGCCTAGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTT
TCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGG
CGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAA
GGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCA
GGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGC
TCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTC
TCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCT
CACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCC
GACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGC
CACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCT
ACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGA
TCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATC
TCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACACTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCA
TGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATAT
GAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATC
CATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGA
TACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGT
GGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAA
TAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCT

CCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCG
ATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCAT
GCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGA
GTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAA
CGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAA
CTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGG
GAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTAT
TGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAA
AGTGCCACCTGACGTCGACGGATCGGGAGATCGATCTCCCGATCCCCTAGGGTCGACTCTCAGTACAATCTGCTCTG
ATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAA
GCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGA
TGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTC
ATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGC
CCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTA
TTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATC

[0105] 举例而言,如本文所述的碱基编辑组合物、系统和方法中使用的胞苷碱基编辑器(CBE)具有如下文提供的以下核酸序列(8877个碱基对)(Addgene,Watertown,MA.;Komor AC,et al.,2017,Sci Adv.,30;3(8):eaao4774.doi:10.1126/sciadv.aao4774)。也包括与BE4核酸序列具有至少95%或更高同一性的多核苷酸序列。

[0106] 1 ATATGCCAAG TACGCCCCCT ATTGACGTCA ATGACGGTAA ATGGCCCGCC TGGCATTATG

[0107] 61 CCCAGTACAT GACCTTATGG GACTTTCCTA CTTGGCAGTA CATCTACGTA TTAGTCATCG

[0108] 121 CTATTACCAT GGTGATGCGG TTTTGGCAGT ACATCAATGG GCGTGGATAG CGGTTTGACT

[0109] 181 CACGGGGATT TCCAAGTCTC CACCCCATTG ACGTCAATGG GAGTTTGTTT TGGCACCAAA

[0110] 241 ATCAACGGGA CTTTCCAAAA TGTCGTAACA ACTCCGCCCC ATTGACGCAA ATGGGCGGTA

[0111] 301 GGCGTGTACG GTGGGAGGTC TATATAAGCA GAGCTGGTTT AGTGAACCGT CAGATCCGCT

[0112] 361 AGAGATCCGC GGCCGCTAAT ACGACTCACT ATAGGGAGAG CCGCCACCAT GAGCTCAGAG

[0113] 421 ACTGGCCCAG TGGCTGTGGA CCCCACATTG AGACGGCGGA TCGAGCCCCA TGAGTTTGAG

[0114] 481 GTATTCTTCG ATCCGAGAGA GCTCCGCAAG GAGACCTGCC TGCTTTACGA AATTAATTGG

[0115] 541 GGGGGCCGGC ACTCCATTTG GCGACATACA TCACAGAACA CTAACAAGCA CGTCGAAGTC

[0116] 601 AACTTCATCG AGAAGTTCAC GACAGAAAGA TATTTCTGTC CGAACACAAG GTGCAGCATT

[0117] 661 ACCTGGTTTC TCAGCTGGAG CCCATGCGGC GAATGTAGTA GGGCCATCAC TGAATTCCTG

[0118] 721 TCAAGGTATC CCCACGTCAC TCTGTTTATT TACATCGCAA GGCTGTACCA CCACGCTGAC

[0119] 781 CCCCGCAATC GACAAGGCCT GCGGGATTTG ATCTCTTCAG GTGTGACTAT CCAAATTATG

[0120] 841 ACTGAGCAGG AGTCAGGATA CTGCTGGAGA AACTTTGTGA ATTATAGCCC GAGTAATGAA

[0121] 901 GCCCACTGGC CTAGGTATCC CCATCTGTGG GTACGACTGT ACGTTCTTGA ACTGTACTGC

[0122] 961 ATCATACTGG GCCTGCCTCC TTGTCTCAAC ATTCTGAGAA GGAAGCAGCC ACAGCTGACA

[0123] 1021 TTCTTTACCA TCGCTCTTCA GTCTTGTCAT TACCAGCGAC TGCCCCCACA CATTCTCTGG

[0124] 1081 GCCACCGGGT TGAAATCTGG TGGTTCTTCT GGTGGTTCTA GCGGCAGCGA

GACTCCCGGG

[0125] 1141 ACCTCAGAGT CCGCCACACC CGAAAGTTCT GGTGGTTCTT CTGGTGGTTC TGATAAAAAG

[0126] 1201 TATTCTATTG GTTTAGCCAT CGGCACTAAT TCCGTTGGAT GGGCTGTCAT AACCGATGAA

[0127] 1261 TACAAAGTAC CTTCAAAGAA ATTTAAGGTG TTGGGGAACA CAGACCGTCA TTCGATTAAA

[0128] 1321 AAGAATCTTA TCGGTGCCCT CCTATTCGAT AGTGGCGAAA CGGCAGAGGC GACTCGCCTG

[0129] 1381 AAACGAACCG CTCGGAGAAG GTATACACGT CGCAAGAACC GAATATGTTA CTTACAAGAA

[0130] 1441 ATTTTTAGCA ATGAGATGGC CAAAGTTGAC GATTCTTTCT TTCACCGTTT GGAAGAGTCC

[0131] 1501 TTCCTTGTCG AAGAGGACAA GAAACATGAA CGGCACCCCA TCTTTGGAAA CATAGTAGAT

[0132] 1561 GAGGTGGCAT ATCATGAAAA GTACCCAACG ATTTATCACC TCAGAAAAAA GCTAGTTGAC

[0133] 1621 TCAACTGATA AAGCGGACCT GAGGTTAATC TACTTGGCTC TTGCCCATAT GATAAAGTTC

[0134] 1681 CGTGGGCACT TTCTCATTGA GGGTGATCTA AATCCGGACA ACTCGGATGT CGACAAACTG

[0135] 1741 TTCATCCAGT TAGTACAAAC CTATAATCAG TTGTTTGAAG AGAACCCTAT AAATGCAAGT

[0136] 1801 GGCGTGGATG CGAAGGCTAT TCTTAGCGCC CGCCTCTCTA AATCCCGACG GCTAGAAAAC

[0137] 1861 CTGATCGCAC AATTACCCGG AGAGAAGAAA AATGGGTTGT TCGGTAACCT TATAGCGCTC

[0138] 1921 TCACTAGGCC TGACACCAAA TTTTAAGTCG AACTTCGACT TAGCTGAAGA TGCCAAATTG

[0139] 1981 CAGCTTAGTA AGGACACGTA CGATGACGAT CTCGACAATC TACTGGCACA AATTGGAGAT

[0140] 2041 CAGTATGCGG ACTTATTTTT GGCTGCCAAA AACCTTAGCG ATGCAATCCT CCTATCTGAC

[0141] 2101 ATACTGAGAG TTAATACTGA GATTACCAAG GCGCCGTTAT CCGCTTCAAT GATCAAAAGG

[0142] 2161 TACGATGAAC ATCACCAAGA CTTGACACTT CTCAAGGCCC TAGTCCGTCA GCAACTGCCT

[0143] 2221 GAGAAATATA AGGAAATATT CTTTGATCAG TCGAAAAACG GGTACGCAGG TTATATTGAC

[0144] 2281 GGCGGAGCGA GTCAAGAGGA ATTCTACAAG TTTATCAAAC CCATATTAGA GAAGATGGAT

[0145] 2341 GGGACGGAAG AGTTGCTTGT AAAACTCAAT CGCGAAGATC TACTGCGAAA GCAGCGGACT

[0146] 2401 TTCGACAACG GTAGCATTCC ACATCAAATC CACTTAGGCG AATTGCATGC TATACTTAGA

[0147] 2461 AGGCAGGAGG ATTTTTATCC GTTCCTCAAA GACAATCGTG AAAAGATTGA GAAAATCCTA

[0148] 2521 ACCTTTCGCA TACCTTACTA TGTGGGACCC CTGGCCCGAG GGAACTCTCG GTTCGCATGG

[0149] 2581 ATGACAAGAA AGTCCGAAGA AACGATTACT CCATGGAATT TTGAGGAAGT TGTCGATAAA

[0150] 2641 GGTGCGTCAG CTCAATCGTT CATCGAGAGG ATGACCAACT TTGACAAGAA TTTACCGAAC

[0151] 2701 GAAAAAGTAT TGCCTAAGCA CAGTTTACTT TACGAGTATT TCACAGTGTA CAATGAACTC

[0152] 2761 ACGAAAGTTA AGTATGTCAC TGAGGGCATG CGTAAACCCG CCTTTCTAAG CGGAGAACAG

[0153] 2821 AAGAAAGCAA TAGTAGATCT GTTATTCAAG ACCAACCGCA AAGTGACAGT TAAGCAATTG

[0154] 2881 AAAGAGGACT ACTTTAAGAA AATTGAATGC TTCGATTCTG TCGAGATCTC CGGGGTAGAA

[0155] 2941 GATCGATTTA ATGCGTCACT TGGTACGTAT CATGACCTCC TAAAGATAAT TAAAGATAAG

[0156] 3001 GACTTCCTGG ATAACGAAGA GAATGAAGAT ATCTTAGAAG ATATAGTGTT GACTCTTACC

[0157] 3061 CTCTTTGAAG ATCGGGAAAT GATTGAGGAA AGACTAAAAA CATACGCTCA CCTGTTCGAC

[0158] 3121 GATAAGGTTA TGAAACAGTT AAAGAGGCGT CGCTATACGG GCTGGGGACG ATTGTCGCGG

[0159] 3181 AAACTTATCA ACGGGATAAG AGACAAGCAA AGTGGTAAAA CTATTCTCGA TTTTCTAAAG

[0160] 3241 AGCGACGGCT TCGCCAATAG GAACTTTATG CAGCTGATCC ATGATGACTC TTTAACCTTC

[0161] 3301 AAAGAGGATA TACAAAAGGC ACAGGTTTCC GGACAAGGGG ACTCATTGCA CGAACATATT

[0162] 3361 GCGAATCTTG CTGGTTCGCC AGCCATCAAA AAGGGCATAC TCCAGACAGT CAAAGTAGTG

[0163] 3421 GATGAGCTAG TTAAGGTCAT GGGACGTCAC AAACCGGAAA ACATTGTAAT

CGAGATGGCA

[0164] 3481 CGCGAAAATC AAACGACTCA GAAGGGGCAA AAAAACAGTC GAGAGCGGAT GAAGAGAATA

[0165] 3541 GAAGAGGGTA TTAAAGAACT GGGCAGCCAG ATCTTAAAGG AGCATCCTGT GGAAAATACC

[0166] 3601 CAATTGCAGA ACGAGAAACT TTACCTCTAT TACCTACAAA ATGGAAGGGA CATGTATGTT

[0167] 3661 GATCAGGAAC TGGACATAAA CCGTTTATCT GATTACGACG TCGATCACAT TGTACCCCAA

[0168] 3721 TCCTTTTTGA AGGACGATTC AATCGACAAT AAAGTGCTTA CACGCTCGGA TAAGAACCGA

[0169] 3781 GGGAAAAGTG ACAATGTTCC AAGCGAGGAA GTCGTAAAGA AAATGAAGAA CTATTGGCGG

[0170] 3841 CAGCTCCTAA ATGCGAAACT GATAACGCAA AGAAAGTTCG ATAACTTAAC TAAAGCTGAG

[0171] 3901 AGGGGTGGCT TGTCTGAACT TGACAAGGCC GGATTTATTA AACGTCAGCT CGTGGAAACC

[0172] 3961 CGCCAAATCA CAAAGCATGT TGCACAGATA CTAGATTCCC GAATGAATAC GAAATACGAC

[0173] 4021 GAGAACGATA AGCTGATTCG GGAAGTCAAA GTAATCACTT TAAAGTCAAA ATTGGTGTCG

[0174] 4081 GACTTCAGAA AGGATTTTCA ATTCTATAAA GTTAGGGAGA TAAATAACTA CCACCATGCG

[0175] 4141 CACGACGCTT ATCTTAATGC CGTCGTAGGG ACCGCACTCA TTAAGAAATA CCCGAAGCTA

[0176] 4201 GAAAGTGAGT TTGTGTATGG TGATTACAAA GTTTATGACG TCCGTAAGAT GATCGCGAAA

[0177] 4261 AGCGAACAGG AGATAGGCAA GGCTACAGCC AAATACTTCT TTTATTCTAA CATTATGAAT

[0178] 4321 TTCTTTAAGA CGGAAATCAC TCTGGCAAAC GGAGAGATAC GCAAACGACC TTTAATTGAA

[0179] 4381 ACCAATGGGG AGACAGGTGA AATCGTATGG GATAAGGGCC GGGACTTCGC GACGGTGAGA

[0180] 4441 AAAGTTTTGT CCATGCCCCA AGTCAACATA GTAAAGAAAA CTGAGGTGCA GACCGGAGGG

[0181] 4501 TTTTCAAAGG AATCGATTCT TCCAAAAAGG AATAGTGATA AGCTCATCGC TCGTAAAAAG

[0182] 4561 GACTGGGACC CGAAAAAGTA CGGTGGCTTC GATAGCCCTA CAGTTGCCTA TTCTGTCCTA

[0183] 4621 GTAGTGGCAA AAGTTGAGAA GGGAAAATCC AAGAAACTGA AGTCAGTCAA AGAATTATTG

[0184] 4681 GGGATAACGA TTATGGAGCG CTCGTCTTTT GAAAAGAACC CCATCGACTT CCTTGAGGCG

[0185] 4741 AAAGGTTACA AGGAAGTAAA AAAGGATCTC ATAATTAAAC TACCAAAGTA TAGTCTGTTT

[0186] 4801 GAGTTAGAAA ATGGCCGAAA ACGGATGTTG GCTAGCGCCG GAGAGCTTCA AAAGGGGAAC

[0187] 4861 GAACTCGCAC TACCGTCTAA ATACGTGAAT TTCCTGTATT TAGCGTCCCA TTACGAGAAG

[0188] 4921 TTGAAAGGTT CACCTGAAGA TAACGAACAG AAGCAACTTT TTGTTGAGCA GCACAAACAT

[0189] 4981 TATCTCGACG AAATCATAGA GCAAATTTCG GAATTCAGTA AGAGAGTCAT CCTAGCTGAT

[0190] 5041 GCCAATCTGG ACAAAGTATT AAGCGCATAC AACAAGCACA GGGATAAACC CATACGTGAG

[0191] 5101 CAGGCGGAAA ATATTATCCA TTTGTTTACT CTTACCAACC TCGGCGCTCC AGCCGCATTC

[0192] 5161 AAGTATTTTG ACACAACGAT AGATCGCAAA CGATACACTT CTACCAAGGA GGTGCTAGAC

[0193] 5221 GCGACACTGA TTCACCAATC CATCACGGGA TTATATGAAA CTCGGATAGA TTTGTCACAG

[0194] 5281 CTTGGGGGTG ACTCTGGTGG TTCTGGAGGA TCTGGTGGTT CTACTAATCT GTCAGATATT

[0195] 5341 ATTGAAAAGG AGACCGGTAA GCAACTGGTT ATCCAGGAAT CCATCCTCAT GCTCCCAGAG

[0196] 5401 GAGGTGGAAG AAGTCATTGG GAACAAGCCG GAAAGCGATA TACTCGTGCA CACCGCCTAC

[0197] 5461 GACGAGAGCA CCGACGAGAA TGTCATGCTT CTGACTAGCG ACGCCCCTGA ATACAAGCCT

[0198] 5521 TGGGCTCTGG TCATACAGGA TAGCAACGGT GAGAACAAGA TTAAGATGCT CTCTGGTGGT

[0199] 5581 TCTGGAGGAT CTGGTGGTTC TACTAATCTG TCAGATATTA TTGAAAAGGA GACCGGTAAG

[0200] 5641 CAACTGGTTA TCCAGGAATC CATCCTCATG CTCCCAGAGG AGGTGGAAGA AGTCATTGGG

[0201] 5701 AACAAGCCGG AAAGCGATAT ACTCGTGCAC ACCGCCTACG ACGAGAGCAC CGACGAGAAT

[0202] 5761 GTCATGCTTC TGACTAGCGA CGCCCCTGAA TACAAGCCTT GGGCTCTGGT

CATACAGGAT

[0203] 5821 AGCAACGGTG AGAACAAGAT TAAGATGCTC TCTGGTGGTT CTCCCAAGAA GAAGAGGAAA

[0204] 5881 GTCTAACCGG TCATCATCAC CATCACCATT GAGTTTAAAC CCGCTGATCA GCCTCGACTG

[0205] 5941 TGCCTTCTAG TTGCCAGCCA TCTGTTGTTT GCCCCTCCCC CGTGCCTTCC TTGACCCTGG

[0206] 6001 AAGGTGCCAC TCCCACTGTC CTTTCCTAAT AAAATGAGGA AATTGCATCG CATTGTCTGA

[0207] 6061 GTAGGTGTCA TTCTATTCTG GGGGGTGGGG TGGGGCAGGA CAGCAAGGGG GAGGATTGGG

[0208] 6121 AAGACAATAG CAGGCATGCT GGGGATGCGG TGGGCTCTAT GGCTTCTGAG GCGGAAAGAA

[0209] 6181 CCAGCTGGGG CTCGATACCG TCGACCTCTA GCTAGAGCTT GGCGTAATCA TGGTCATAGC

[0210] 6241 TGTTTCCTGT GTGAAATTGT TATCCGCTCA CAATTCCACA CAACATACGA GCCGGAAGCA

[0211] 6301 TAAAGTGTAA AGCCTAGGGT GCCTAATGAG TGAGCTAACT CACATTAATT GCGTTGCGCT

[0212] 6361 CACTGCCCGC TTTCCAGTCG GGAAACCTGT CGTGCCAGCT GCATTAATGA ATCGGCCAAC

[0213] 6421 GCGCGGGGAG AGGCGGTTTG CGTATTGGGC GCTCTTCCGC TTCCTCGCTC ACTGACTCGC

[0214] 6481 TGCGCTCGGT CGTTCGGCTG CGGCGAGCGG TATCAGCTCA CTCAAAGGCG GTAATACGGT

[0215] 6541 TATCCACAGA ATCAGGGGAT AACGCAGGAA AGAACATGTG AGCAAAAGGC CAGCAAAAGG

[0216] 6601 CCAGGAACCG TAAAAAGGCC GCGTTGCTGG CGTTTTTCCA TAGGCTCCGC CCCCCTGACG

[0217] 6661 AGCATCACAA AAATCGACGC TCAAGTCAGA GGTGGCGAAA CCCGACAGGA CTATAAAGAT

[0218] 6721 ACCAGGCGTT TCCCCCTGGA AGCTCCCTCG TGCGCTCTCC TGTTCCGACC CTGCCGCTTA

[0219] 6781 CCGGATACCT GTCCGCCTTT CTCCCTTCGG GAAGCGTGGC GCTTTCTCAT AGCTCACGCT

[0220] 6841 GTAGGTATCT CAGTTCGGTG TAGGTCGTTC GCTCCAAGCT GGGCTGTGTG CACGAACCCC

[0221] 6901 CCGTTCAGCC CGACCGCTGC GCCTTATCCG GTAACTATCG TCTTGAGTCC AACCCGGTAA

[0222] 6961 GACACGACTT ATCGCCACTG GCAGCAGCCA CTGGTAACAG GATTAGCAGA GCGAGGTATG

[0223] 7021 TAGGCGGTGC TACAGAGTTC TTGAAGTGGT GGCCTAACTA CGGCTACACT AGAAGAACAG

[0224] 7081 TATTTGGTAT CTGCGCTCTG CTGAAGCCAG TTACCTTCGG AAAAAGAGTT GGTAGCTCTT

[0225] 7141 GATCCGGCAA ACAAACCACC GCTGGTAGCG GTGGTTTTTT TGTTTGCAAG CAGCAGATTA

[0226] 7201 CGCGCAGAAA AAAAGGATCT CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCTC

[0227] 7261 AGTGGAACGA AAACTCACGT TAAGGGATTT TGGTCATGAG ATTATCAAAA AGGATCTTCA

[0228] 7321 CCTAGATCCT TTTAAATTAA AAATGAAGTT TTAAATCAAT CTAAAGTATA TATGAGTAAA

[0229] 7381 CTTGGTCTGA CAGTTACCAA TGCTTAATCA GTGAGGCACC TATCTCAGCG ATCTGTCTAT

[0230] 7441 TTCGTTCATC CATAGTTGCC TGACTCCCCG TCGTGTAGAT AACTACGATA CGGGAGGGCT

[0231] 7501 TACCATCTGG CCCCAGTGCT GCAATGATAC CGCGAGACCC ACGCTCACCG GCTCCAGATT

[0232] 7561 TATCAGCAAT AAACCAGCCA GCCGGAAGGG CCGAGCGCAG AAGTGGTCCT GCAACTTTAT

[0233] 7621 CCGCCTCCAT CCAGTCTATT AATTGTTGCC GGGAAGCTAG AGTAAGTAGT TCGCCAGTTA

[0234] 7681 ATAGTTTGCG CAACGTTGTT GCCATTGCTA CAGGCATCGT GGTGTCACGC TCGTCGTTTG

[0235] 7741 GTATGGCTTC ATTCAGCTCC GGTTCCCAAC GATCAAGGCG AGTTACATGA TCCCCCATGT

[0236] 7801 TGTGCAAAAA AGCGGTTAGC TCCTTCGGTC CTCCGATCGT TGTCAGAAGT AAGTTGGCCG

[0237] 7861 CAGTGTTATC ACTCATGGTT ATGGCAGCAC TGCATAATTC TCTTACTGTC ATGCCATCCG

[0238] 7921 TAAGATGCTT TTCTGTGACT GGTGAGTACT CAACCAAGTC ATTCTGAGAA TAGTGTATGC

[0239] 7981 GGCGACCGAG TTGCTCTTGC CCGGCGTCAA TACGGGATAA TACCGCGCCA CATAGCAGAA

[0240] 8041 CTTTAAAAGT GCTCATCATT GGAAAACGTT CTTCGGGGCG AAAACTCTCA AGGATCTTAC

[0241] 8101 CGCTGTTGAG ATCCAGTTCG ATGTAACCCA CTCGTGCACC CAACTGATCT

TCAGCATCTT

[0242] 8161 TTACTTTCAC CAGCGTTTCT GGGTGAGCAA AAACAGGAAG GCAAAATGCC GCAAAAAAGG

[0243] 8221 GAATAAGGGC GACACGGAAA TGTTGAATAC TCATACTCTT CCTTTTTCAA TATTATTGAA

[0244] 8281 GCATTTATCA GGGTTATTGT CTCATGAGCG GATACATATT TGAATGTATT TAGAAAAATA

[0245] 8341 AACAAATAGG GGTTCCGCGC ACATTTCCCC GAAAAGTGCC ACCTGACGTC GACGGATCGG

[0246] 8401 GAGATCGATC TCCCGATCCC CTAGGGTCGA CTCTCAGTAC AATCTGCTCT GATGCCGCAT

[0247] 8461 AGTTAAGCCA GTATCTGCTC CCTGCTTGTG TGTTGGAGGT CGCTGAGTAG TGCGCGAGCA

[0248] 8521 AAATTTAAGC TACAACAAGG CAAGGCTTGA CCGACAATTG CATGAAGAAT CTGCTTAGGG

[0249] 8581 TTAGGCGTTT TGCGCTGCTT CGCGATGTAC GGGCCAGATA TACGCGTTGA CATTGATTAT

[0250] 8641 TGACTAGTTA TTAATAGTAA TCAATTACGG GGTCATTAGT TCATAGCCCA TATATGGAGT

[0251] 8701 TCCGCGTTAC ATAACTTACG GTAAATGGCC CGCCTGGCTG ACCGCCCAAC GACCCCCGCC

[0252] 8761 CATTGACGTC AATAATGACG TATGTTCCCA TAGTAACGCC AATAGGGACT TTCCATTGAC

[0253] 8821 GTCAATGGGT GGAGTATTTA CGGTAAACTG CCCACTTGGC AGTACATCAA GTGTATC

[0254] 在一些实施例中,胞苷碱基编辑器是BE4,其具有选自以下一项的核酸序列:

[0255] 原始BE4核酸序列:

[0256] ATGagctcagagactggcccagtggctgtggaccccacattgagacggcggatcgagccccatgagtt tgaggtattcttcgatccgagagagctccgcaaggagacctgcctgctttacgaaattaattgggggggccggcac tccatttggcgacatacatcacagaacactaacaagcacgtcgaagtcaacttcatcgagaagttcacgacagaaa gatatttctgtccgaacacaaggtgcagcattacctggtttctcagctggagccgcgaatgtagtagggccatcac tgaattcctgtcaaggtatccccacgtcactctgtttatttacatcgcaaggctgtaccaccacgctgacccccgc aatcgacaaggcctgcgggatttgatctcttcaggtgtgactatccaaattatgactgagcaggagtcaggatact gctggagaaactttgtgaattatagcccgagtaatgaagcccactggcctaggtatccccatctgtgggtacgact gtacgttcttgaactgtactgcatcatactgggcctgcctccttgtctcaacattctgagaaggaagcagccacag ctgacattctttaccatcgctcttcagtcttgtcattaccagcgactgcccccacacattctctgggccaccgggt tgaaatctggtggttcttctggtggttctagcggcagcgagactcccgggacctcagagtccgccacacccgaaag ttctggtggttcttctggtggttctgataaaaagtattctattggtttagccatcggcactaattccgttggatgg gctgtcataaccgatgaatacaaagtaccttcaaagaaatttaaggtgttggggaacacagaccgtcattcgatta aaaagaatcttatcggtgccctcctattcgatagtggcgaaacggcagaggcgactcgcctgaaacgaaccgctcg

gagaaggtatacacgtcgcaagaaccgaatatgttacttacaagaaatttttagcaatgagatggccaaagttgac
gattctttctttcaccgtttggaagagtccttccttgtcgaagaggacaagaaacatgaacggcaccccatctttg
gaaacatagtagatgaggtggcatatcatgaaaagtacccaacgatttatcacctcagaaaaaagctagttgactc
aactgataaagcggacctgaggttaatctacttggctcttgcccatatgataaagttccgtgggcactttctcatt
gagggtgatctaaatccggacaactcggatgtcgacaaactgttcatccagttagtacaaacctataatcagttgt
ttgaagagaaccctataaatgcaagtggcgtggatgcgaaggctattcttagcgcccgcctctctaaatcccgacg
gctagaaaacctgatcgcacaattacccggagagaagaaaaatgggttgttcggtaaccttatagcgctctcacta
ggcctgacaccaaattttaagtcgaacttcgacttagctgaagatgccaaattgcagcttagtaaggacacgtacg
atgacgatctcgacaatctactggcacaaattggagatcagtatgcggacttatttttggctgccaaaaaccttag
cgatgcaatcctcctatctgacatactgagagttaatactgagattaccaaggcgccgttatccgcttcaatgatc
aaaaggtacgatgaacatcaccaagacttgacacttctcaaggccctagtccgtcagcaactgcctgagaaatata
aggaaatattctttgatcagtcgaaaaacgggtacgcaggttatattgacggcggagcgagtcaagaggaattcta
caagtttatcaaacccatattagagaagatggatgggacggaagagttgcttgtaaaactcaatcgcgaagatcta
ctgcgaaagcagcggactttcgacaacggtagcattccacatcaaatccacttaggcgaattgcatgctatactta
gaaggcaggaggatttttatccgttcctcaaagacaatcgtgaaaagattgagaaaatcctaacctttcgcatacc
ttactatgtgggacccctggcccgagggaactctcggttcgcatggatgacaagaaagtccgaagaaacgattact
ccatggaatttgaggaagttgtcgataaaggtgcgtcagctcaatcgttcatcgagaggatgaccaactttgaca
agaatttaccgaacgaaaaagtattgcctaagcacagtttactttacgagtatttcacagtgtacaatgaactcac
gaaagttaagtatgtcactgagggcatgcgtaaacccgcctttctaagcggagaacagaagaaagcaatagtagat
ctgttattcaagaccaaccgcaaagtgacagttaagcaattgaaagaggactactttaagaaaattgaatgcttcg
attctgtcgagatctccggggtagaagatcgatttaatgcgtcacttggtacgtatcatgacctcctaaagataat
taaagataaggacttcctggataacgaagagaatgaagatatcttagaagatatagtgttgactcttaccctcttt
gaagatcgggaaatgattgaggaaagactaaaaacatacgctcacctgttcgacgataaggttatgaaacagttaa
agaggcgtcgctatacgggctggggacgattgtcgcggaaacttatcaacgggataagagacaagcaaagtggtaa
aactattctcgattttctaaagagcgacggcttcgccaataggaactttatgcagctgatccatgatgactcttta
accttcaaagaggatatacaaaaggcacaggtttccggacaaggggactcattgcacgaacatattgcgaatcttg
ctggttcgccagccatcaaaaagggcatactccagacagtcaaagtagtggatgagctagttaaggtcatgggacg
tcacaaaccggaaaacattgtaatcgagatggcacgcgaaaatcaaacgactcagaaggggcaaaaaaacagtcga
gagcggatgaagagaatagaagagggtattaaagaactgggcagccagatcttaaaggagcatcctgtggaaaata
cccaattgcagaacgagaaactttacctctattacctacaaaatggaagggacatgtatgttgatcaggaactgga
cataaaccgtttatctgattacgacgtcgatcacattgtaccccaatccttttgaaggacgattcaatcgacaat
aaagtgcttacacgctcggataagaaccgagggaaaagtgacaatgttccaagcgaggaagtcgtaaagaaaatga
agaactattggcggcagctcctaaatgcgaaactgataacgcaaagaaagttcgataacttaactaaagctgagag
gggtggcttgtctgaacttgacaaggccggatttattaaacgtcagctcgtggaaacccgccaaatcacaaagcat
gttgcacagatactagattcccgaatgaatacgaaatacgacgagaacgataagctgattcgggaagtcaaagtaa
tcactttaaagtcaaaattggtgtcggacttcagaaaggattttcaattctataaagttagggagataaataacta
ccaccatgcgcacgacgcttatcttaatgccgtcgtagggaccgcactcattaagaaatacccgaagctagaaagt
gagtttgtgtatggtgattacaaagtttatgacgtccgtaagatgatcgcgaaaagcgaacaggagataggcaagg
ctacagccaaatacttcttttattctaacattatgaatttctttaagacggaaatcactctggcaaacggagagat

acgcaaacgacctttaattgaaaccaatggggagacaggtgaaatcgtatgggataagggccgggacttcgcgacg
gtgagaaaagttttgtccatgccccaagtcaacatagtaaagaaaactgaggtgcagaccggagggttttcaaagg
aatcgattcttccaaaaaggaatagtgataagctcatcgctcgtaaaaaggactgggacccgaaaaagtacggtgg
cttcgatagccctacagttgcctattctgtcctagtagtggcaaaagttgagaagggaaaatccaagaaactgaag
tcagtcaaagaattattggggataacgattatggagcgctcgtcttttgaaaagaaccccatcgacttccttgagg
cgaaaggttacaaggaagtaaaaaaggatctcataattaaactaccaaagtatagtctgtttgagttagaaaatgg
ccgaaaacggatgttggctagcgccggagagcttcaaaaggggaacgaactcgcactaccgtctaaatacgtgaat
ttcctgtatttagcgtcccattacgagaagttgaaaggttcacctgaagataacgaacagaagcaacttttttgttg
agcagcacaaacattatctcgacgaaatcatagagcaaatttcggaattcagtaagagagtcatcctagctgatgc
caatctggacaaagtattaagcgcatacaacaagcacagggataaacccatacgtgagcaggcggaaaatattatc
catttgtttactcttaccaacctcggcgctccagccgcattcaagtattttgacacaacgatagatcgcaaacgat
acacttctaccaaggaggtgctagacgcgacactgattcaccaatccatcacgggattatatgaaactcggataga
tttgtcacagcttgggggtgactctggtggttctggaggatctggtggttctactaatctgtcagatattattgaa
aaggagaccggtaagcaactggttatccaggaatccatcctcatgctcccagaggaggtggaagaagtcattggga
acaagccggaaagcgatatactcgtgcacaccgcctacgacgagagcaccgacgagaatgtcatgcttctgactag
cgacgcccctgaatacaagccttgggctctggtcatacaggatagcaacggtgagaacaagattaagatgctctct
ggtggttctggaggatctggtggttctactaatctgtcagatattattgaaaaggagaccggtaagcaactggtta
tccaggaatccatcctcatgctcccagaggaggtggaagaagtcattgggaacaagccggaaagcgatatactcgt
gcacaccgcctacgacgagagcaccgacgagaatgtcatgcttctgactagcgacgcccctgaatacaagccttgg
gctctggtcatacaggatagcaacggtgagaacaagattaagatgctctctggtggttctAAAAGGACGGCGGACG
GATCAGAGTTCGAGAGTCCGAAAAAAAAACGAAAGGTCGAAtaa

[0257] BE4密码子优化1核酸序列：

[0258] ATGTCATCCGAAACCGGGCCAGTGGCCGTAGACCCAACACTCAGGAGGCGGATAGAACCCCATGAGTT
TGAAGTGTTCTTCGACCCCAGAGAGCTGCGCAAAGAGACTTGCCTCCTGTATGAAATAAATTGGGGGGGTCGCCAT
TCAATTTGGAGGCACACTAGCCAGAATACTAACAAACACGTGGAGGTAAATTTTATCGAGAAGTTTACCACCGAAA
GATACTTTTGCCCCAATACACGGTGTTCAATTACCTGGTTTCTGTCATGGAGTCCATGTGGAGAATGTAGTAGAGC
GATAACTGAGTTCCTGTCTCGATATCCTCACGTCACGTTGTTTATATACATCGCTCGGCTTTATCACCATGCGGAC
CCGCGGAACAGGCAAGGTCTTCGGGACCTCATATCCTCTGGGGTGACCATCCAGATAATGACGGAGCAAGAGAGCG
GATACTGCTGGCGAAACTTTGTTAACTACAGCCCAAGCAATGAGGCACACTGGCCTAGATATCCGCATCTCTGGGT
TCGACTGTATGTCCTTGAACTGTACTGCATAATTCTGGGACTTCCGCCATGCTTGAACATTCTGCGGCGGAAACAA
CCACAGCTGACCTTTTTCACGATTGCTCTCCAAAGTTGTCACTACCAGCGATTGCCACCCCACATCTTGTGGGCTA
CTGGACTCAAGTCTGGAGGAAGTTCAGGCGGAAGCAGCGGGTCTGAAACGCCCGGAACCTCAGAGAGCGCAACGCC
CGAAAGCTCTGGAGGGTCAAGTGGTGGTAGTGATAAGAAATACTCCATCGGCCTCGCCATCGGTACGAATTCTGTC
GGTTGGGCCGTTATCACCGATGAGTACAAGGTCCCTTCTAAGAAATTCAAGGTTTTGGGCAATACAGACCGCCATT
CTATAAAAAAAAACCTGATCGGCGCCCTTTTGTTTGACAGTGGTGAGACTGCTGAAGCGACTCGCCTGAAGCGAAC
TGCCAGGAGGCGGTATACGAGGCGAAAAAACCGAATTTGTTACCTCCAGGAGATTTTCTCAAATGAAATGGCCAAG
GTAGATGATAGTTTTTTTCACCGCTTGGAAGAAAGTTTTCTCGTTGAGGAGGACAAAAAGCACGAGAGGCACCCAA
TCTTTGGCAACATAGTCGATGAGGTCGCATACCATGAGAAATATCCTACGATCTATCATCTCCGCAAGAAGCTGGT
CGATAGCACGGATAAAGCTGACCTCCGGCTGATCTACCTTGCTCTTGCTCACATGATTAAATTCAGGGGCCATTTC

CTGATAGAAGGAGACCTCAATCCCGACAATTCTGATGTCGACAAACTGTTTATTCAGCTCGTTCAGACCTATAATC
AACTCTTTGAGGAGAACCCCATCAATGCTTCAGGGGTGGACGCAAAGGCCATTTTGTCCGCGCGCTTGAGTAAATC
ACGACGCCTCGAGAATTTGATAGCTCAACTGCCGGGTGAGAAGAAAAACGGGTTGTTTGGGAATCTCATAGCGTTG
AGTTTGGGACTTACGCCAAACTTTAAGTCTAACTTTGATTTGGCCGAAGATGCCAAATTGCAGCTGTCCAAAGATA
CCTATGATGACGACTTGGATAACCTTCTTGCGCAGATTGGTGACCAATACGCGGATCTGTTTCTTGCCGCAAAAAA
TCTGTCCGACGCCATACTCTTGTCCGATATACTGCGCGTCAATACTGAGATAACTAAGGCTCCCCTCAGCGCGTCC
ATGATTAAAAGATACGATGAGCACCACCAAGATCTCACTCTGTTGAAAGCCCTGGTTCGCCAGCAGCTTCCAGAGA
AGTATAAGGAGATATTTTTCGACCAATCTAAAAACGGCTATGCGGGTTACATTGACGGTGGCGCCTCTCAAGAAGA
ATTCTACAAGTTTATAAAGCCGATACTTGAGAAAATGGACGGTACAGAGGAATTGTTGGTTAAGCTCAATCGCGAG
GACTTGTTGAGAAAGCAGCGCACATTTGACAATGGTAGTATTCCACACCAGATTCATCTGGGCGAGTTGCATGCCA
TTCTTAGAAGACAAGAAGATTTTTATCCGTTTCTGAAAGATAACAGAGAAAAGATTGAAAAGATACTTACCTTTCG
CATACCGTATTATGTAGGTCCCCTGGCTAGAGGGAACAGTCGCTTCGCTTGGATGACTCGAAAATCAGAAGAAACA
ATAACCCCCTGGAATTTTGAAGAAGTGGTAGATAAAGGTGCGAGTGCCCAATCTTTTATTGAGCGGATGACAAATT
TTGACAAGAATCTGCCTAACGAAAAGGTGCTTCCCAAGCATTCCCTTTTGTATGAATACTTTACAGTATATAATGA
ACTGACTAAAGTGAAGTACGTTACCGAGGGGATGCGAAAGCCAGCTTTTCTCAGTGGCGAGCAGAAAAAAGCAATA
GTTGACCTGCTGTTCAAGACGAATAGGAAGGTTACCGTCAAACAGCTCAAAGAAGATTACTTTAAAAAGATCGAAT
GTTTTGATTCAGTTGAGATAAGCGGAGTAGAGGATAGATTTAACGCAAGTCTTGGAACTTATCATGACCTTTTGAA
GATCATCAAGGATAAAGATTTTTTGGACAACGAGGAGAATGAAGATATCCTGGAAGATATAGTACTTACCTTGACG
CTTTTTGAAGATCGAGAGATGATCGAGGAGCGACTTAAGACGTACGCACATCTCTTTGACGATAAGGTTATGAAAC
AATTGAAACGCCGGCGGTATACTGGCTGGGGCAGGCTTTCTCGAAAGCTGATTAATGGTATCCGCGATAAGCAGTC
TGGAAAGACAATCCTTGACTTTCTGAAAAGTGATGGATTTGCAAATAGAAACTTTATGCAGCTTATACATGATGAC
TCTTTGACGTTCAAGGAAGACATCCAGAAGGCACAGGTATCCGGCCAAGGGGATAGCCTCCATGAACACATAGCCA
ACCTGGCCGGCTCACCAGCTATTAAAAAGGGAATATTGCAAACCGTTAAGGTTGTTGACGAACTCGTTAAGGTTAT
GGGCCGACACAAACCAGAGAATATCGTGATTGAGATGGCTAGGGAGAATCAGACCACTCAAAAAGGTCAGAAAAAT
TCTCGCGAAAGGATGAAGCGAATTGAAGAGGGAATCAAAGAACTTGGCTCTCAAATTTTGAAAGAGCACCCGGTAG
AAAACACTCAGCTGCAGAATGAAAAGCTGTATCTGTATTATCTGCAGAATGGTCGAGATATGTACGTTGATCAGGA
GCTGGATATCAATAGGCTCAGTGACTACGATGTCGACCACATCGTTCCTCAATCTTTCCTGAAAGATGACTCTATC
GACAACAAAGTGTTGACGCGATCAGATAAGAACCGGGGAAAATCCGACAATGTACCCTCAGAAGAAGTTGTCAAGA
AGATGAAAAACTATTGGAGACAATTGCTGAACGCCAAGCTCATAACACAACGCAAGTTCGATAACTTGACGAAAGC
CGAAAGAGGTGGGTTGTCAGAATTGGACAAAGCTGGCTTTATTAAGCGCCAATTGGTGGAGACCCGGCAGATTACG
AAACACGTAGCACAAATTTTGGATTCACGAATGAATACCAAATACGACGAAAACGACAAATTGATACGCGAGGTGA
AAGTGATTACGCTTAAGAGTAAGTTGGTTTCCGATTTCAGGAAGGATTTTCAGTTTTACAAAGTAAGAGAAATAAA
CAACTACCACCACGCCCATGATGCTTACCTCAACGCGGTAGTTGGCACAGCTCTTATCAAAAAATATCCAAAGCTG
GAAAGCGAGTTCGTTTACGGTGACTATAAAGTATACGACGTTCGGAAGATGATAGCCAAATCAGAGCAGGAAATTG
GGAAGGCAACCGCAAAATACTTCTTCTATTCAAACATCATGAACTTCTTTAAGACGGAGATTACGCTCGCGAACGG
CGAAATACGCAAGAGGCCCCTCATAGAGACTAACGGCGAAACCGGGGAGATCGTATGGGACAAAGGACGGGACTTT
GCGACCGTTAGAAAAGTACTTTCAATGCCACAAGTGAATATTGTTAAAAAGACAGAAGTACAAACAGGGGGGTTCA
GTAAGGAATCCATTTTGCCCAAGCGGAACAGTGATAAATTGATAGCAAGGAAAAAAGATTGGGACCCTAAGAAGTA
CGGTGGTTTCGACTCTCCTACCGTTGCATATTCAGTCCTTGTAGTTGCGAAAGTGGAAAAGGGGAAAAGTAAGAAG

CTTAAGAGTGTTAAAGAGCTTCTGGGCATAACCATAATGGAACGGTCTAGCTTCGAGAAAAATCCAATTGACTTTC
TCGAGGCTAAAGGTTACAAGGAGGTAAAAAAGGACCTGATAATTAAACTCCCAAAGTACAGTCTCTTCGAGTTGGA
GAATGGGAGGAAGAGAATGTTGGCATCTGCAGGGGAGCTCCAAAAGGGGAACGAGCTGGCTCTGCCTTCAAAATAC
GTGAACTTTCTGTACCTGGCCAGCCACTACGAGAAACTCAAGGGTTCTCCTGAGGATAACGAGCAGAAACAGCTGT
TTGTAGAGCAGCACAAGCATTACCTGGACGAGATAATTGAGCAAATTAGTGAGTTCTCAAAAAGAGTAATCCTTGC
AGACGCGAATCTGGATAAAGTTCTTTCCGCCTATAATAAGCACCGGGACAAGCCTATACGAGAACAAGCCGAGAAC
ATCATTCACCTCTTTACCCTTACTAATCTGGGCGCGCCGGCCGCCTTCAAATACTTCGACACCACGATAGACAGGA
AAAGGTATACGAGTACCAAAGAAGTACTTGACGCCACTCTCATCCACCAGTCTATAACAGGGTTGTACGAAACGAG
GATAGATTTGTCCCAGCTCGGCGGCGACTCAGGAGGGTCAGGCGGCTCCGGTGGATCAACGAATCTTTCCGACATA
ATCGAGAAAGAAACCGGCAAACAGTTGGTGATCCAAGAATCAATCCTGATGCTGCCTGAAGAAGTAGAAGAGGTGA
TTGGCAACAAACCTGAGTCTGACATTCTTGTCCACACCGCGTATGACGAGAGCACGGACGAGAACGTTATGCTTCT
CACTAGCGACGCCCCTGAGTATAAACCATGGGCGCTGGTCATCCAAGATTCCAATGGGGAAAACAAGATTAAGATG
CTTAGTGGTGGGTCTGGAGGGAGCGGTGGGTCCACGAACCTCAGCGACATTATTGAAAAAGAGACTGGTAAACAAC
TTGTAATACAAGAGTCTATTCTGATGTTGCCTGAAGAGGTGGAGGAGGTGATTGGGAACAAACCGGAGTCTGATAT
ACTTGTTCATACCGCCTATGACGAATCTACTGATGAGAATGTGATGCTTTTaACGTCAGACGCTCCCGAGTACAAA
CCCTGGGCTCTGGTGATTCAGGACAGCAATGGTGAGAATAAGATTAAAATGTTGAGTGGGGGCTCAAAGCGCACGG
CTGACGGTAGCGAATTTGAGAGCCCCAAAAAAAAACGAAAGGTCGAAtaa

[0259] BE4密码子优化2核酸序列：

[0260] ATGAGCAGCGAGACAGGCCCTGTGGCTGTGGATCCTACACTGCGGAGAAGAATCGAGCCCCACGAGTT
CGAGGTGTTCTTCGACCCCAGAGAGCTGCGGAAAGAGACATGCCTGCTGTACGAGATCAACTGGGGCGGCAGACAC
TCTATCTGGCGGCACACAAGCCAGAACACCAACAAGCACGTGGAAGTGAACTTTATCGAGAAGTTTACGACCGAGC
GGTACTTCTGCCCCAACACCAGATGCAGCATCACCTGGTTTCTGAGCTGGTCCCCTTGCGGCGAGTGCAGCAGAGC
CATCACCGAGTTTCTGTCCAGATATCCCCACGTGACCCTGTTCATCTATATCGCCCGGCTGTACCACCACGCCGAT
CCTAGAAATAGACAGGGACTGCGCGACCTGATCAGCAGCGGAGTGACCATCCAGATCATGACCGAGCAAGAGAGCG
GCTACTGCTGGCGGAACTTCGTGAACTACAGCCCCAGCAACGAAGCCCACTGGCCTAGATATCCTCACCTGTGGGT
CCGACTGTACGTGCTGGAACTGTACTGCATCATCCTGGGCCTGCCTCCATGCCTGAACATCCTGAGAAGAAAGCAG
CCTCAGCTGACCTTCTTCACAATCGCCCTGCAGAGCTGCCACTACCAGAGACTGCCTCCACACATCCTGTGGGCCA
CCGGACTTAAGAGCGGAGGATCTAGCGGCGGCTCTAGCGGATCTGAGACACCTGGCACAAGCGAGTCTGCCACACC
TGAGAGTAGCGGCGGATCTTCTGGCGGCTCCGACAAGAAGTACTCTATCGGACTGGCCATCGGCACCAACTCTGTT
GGATGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACA
GCATCAAGAAGAATCTGATCGGCGCCCTGCTGTTCGACTCTGGCGAAACAGCCGAAGCCACCAGACTGAAGAGAAC
CGCCAGGCGGAGATACACCCGGCGGAAGAACCGGATCTGCTACCTGCAAGAGATCTTCAGCAACGAGATGGCCAAG
GTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGACAAGAAGCACGAGCGGCACCCCA
TCTTCGGCAACATCGTGGATGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGT
GGACAGCACCGACAAGGCCGACCTGAGACTGATCTACCTGGCTCTGGCCCACATGATCAAGTTCCGGGGCCACTTT
CTGATCGAGGGCGATCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACC
AGCTGTTCGAGGAAAACCCCATCAACGCCTCTGGCGTGGACGCCAAGGCTATCCTGTCTGCCAGACTGAGCAAGAG
CAGAAGGCTGGAAAACCTGATCGCCCAGCTGCCTGGCGAGAAGAAGAATGGCCTGTTCGGCAACCTGATTGCCCTG
AGCCTGGGACTGACCCCTAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACA

CCTACGACGACGACCTGGACAATCTGCTGGCCCAGATCGGCGATCAGTACGCCGACTTGTTTCTGGCCGCCAAGAA
CCTGTCCGACGCCATCCTGCTGAGCGATATCCTGAGAGTGAACACCGAGATCACAAAGGCCCCTCTGAGCGCCTCT
ATGATCAAGAGATACGACGAGCACCACCAGGATCTGACCCTGCTGAAGGCCCTCGTTAGACAGCAGCTGCCAGAGA
AGTACAAAGAGATTTTCTTCGATCAGTCCAAGAACGGCTACGCCGGCTACATTGATGGCGGAGCCAGCCAAGAGGA
ATTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTGGTCAAGCTGAACAGAGAG
GACCTGCTGCGGAAGCAGCGGACCTTCGACAATGGCTCTATCCCTCACCAGATCCACCTGGGAGAGCTGCACGCCA
TTCTGCGGAGACAAGAGGACTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCAG
GATCCCCTACTACGTGGGACCACTGGCCAGAGGCAATAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACC
ATCACACCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCCAGCGCTCAGTCCTTCATCGAGCGGATGACCAACT
TCGATAAGAACCTGCCTAACGAGAAGGTGCTGCCCAAGCACTCCCTGCTGTATGAGTACTTCACCGTGTACAACGA
GCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTTCTGAGCGGCGAGCAGAAAAAGGCCATT
GTGGATCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGT
GCTTCGACAGCGTGGAAATCAGCGGCGTGGAAGATCGGTTCAATGCCAGCCTGGGCACATACCACGACCTGCTGAA
AATTATCAAGGACAAGGACTTCCTGGACAACGAAGAGAACGAGGACATTCTCGAGGACATCGTGCTGACCCTGACA
CTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACATACGCCCACCTGTTCGACGACAAAGTGATGAAGC
AACTGAAGCGGAGGCGGTACACAGGCTGGGGCAGACTGTCTCGGAAGCTGATCAACGGCATCCGGGATAAGCAGTC
CGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGAC
AGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAAGGCGATTCTCTGCACGAGCACATTGCCA
ACCTGGCCGGATCTCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTTGTGAAAGTGAT
GGGCAGACACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACACAGAAGGGCCAGAAGAAC
AGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGG
AAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGACGGGATATGTACGTGGACCAAGA
GCTGGACATCAACCGGCTGAGCGACTACGATGTGGACCATATCGTGCCCCAGAGCTTTCTGAAGGACGACTCCATC
GATAACAAGGTCCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGATAACGTGCCCTCCGAAGAGGTGGTCAAGA
AGATGAAGAACTACTGGCGACAGCTGCTGAACGCCAAGCTGATTACCCAGCGGAAGTTCGATAACCTGACCAAGGC
CGAGAGAGGCGGCCTGAGCGAACTTGATAAGGCCGGCTTCATTAAGCGGCAGCTGGTGGAAACCCGGCAGATCACC
AAACACGTGGCACAGATTCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTGA
AAGTCATCACCCTGAAGTCTAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTCTACAAAGTGCGGGAAATCAA
CAACTACCATCACGCCCACGACGCCTACCTGAATGCCGTTGTTGGAACAGCCCTGATCAAGAAGTATCCCAAGCTG
GAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAACAAGAGATCG
GCAAGGCTACCGCCAAGTACTTTTTCTACAGCAACATCATGAACTTTTTCAAGACAGAGATCACCCTGGCCAACGG
CGAGATCCGGAAAAGACCCCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCAGAGATTTT
GCCACAGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAGAAAACCGAGGTGCAGACAGGCGGCTTCA
GCAAAGAGTCTATCCTGCCTAAGCGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTA
CGGCGGCTTCGATAGCCCTACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAAAAG
CTCAAGAGCGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTTGAGAAGAACCCGATCGACTTTC
TGGAAGCCAAGGGCTACAAAGAAGTCAAGAAGGACCTCATCATCAAGCTCCCCAAGTACAGCCTGTTCGAGCTGGA
AAATGGCCGGAAGCGGATGCTGGCCTCAGCAGGCGAACTGCAGAAAGGCAATGAACTGGCCCTGCCTAGCAAATAC
GTCAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCAGCCCCGAGGACAATGAGCAAAAGCAGCTGT

TTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGC
CGACGCTAACCTGGATAAGGTGCTGTCTGCCTATAACAAGCACCGGGACAAGCCTATCAGAGAGCAGGCCGAGAAT
ATCATCCACCTGTTTACCCTGACCAACCTGGGAGCCCCTGCCGCCTTCAAGTACTTCGACACCACCATCGACCGGA
AGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACACTGATCCACCAGTCTATCACCGGCCTGTACGAAACCCG
GATCGACCTGTCTCAGCTCGGCGGCGATTCTGGTGGTTCTGGCGGAAGTGGCGGATCCACCAATCTGAGCGACATC
ATCGAAAAAGAGACAGGCAAGCAGCTCGTGATCCAAGAATCCATCCTGATGCTGCCTGAAGAGGTTGAGGAAGTGA
TCGGCAACAAGCCTGAGTCCGACATCCTGGTGCACACCGCCTACGATGAGAGCACCGATGAGAACGTCATGCTGCT
GACAAGCGACGCCCCTGAGTACAAGCCTTGGGCTCTCGTGATTCAGGACAGCAATGGGGAGAACAAGATCAAGATG
CTGAGCGGAGGTAGCGGAGGCAGTGGCGGAAGCACAAACCTGTCTGATATCATTGAAAAAGAAACCGGGAAGCAAC
TGGTCATTCAAGAGTCCATTCTCATGCTCCCGGAAGAAGTCGAGGAAGTCATTGGAAACAAACCCGAGAGCGATAT
TCTGGTCCACACAGCCTATGACGAGTCTACAGACGAAAACGTGATGCTCCTGACCTCTGACGCTCCCGAGTATAAG
CCCTGGGCACTTGTTATCCAGGACTCTAACGGGGAAAACAAAATCAAAATGTTGTCCGGCGGCAGCAAGCGGACAG
CCGATGGATCTGAGTTCGAGAGCCCCAAGAAGAAACGGAAGGTgGAGtaa

[0261] “碱基编辑活性”意为发挥作用以化学地改变多核苷酸内的碱基。在一个实施方案中,将第一碱基转化为第二碱基。在一个实施方案中,碱基编辑活性是胞苷脱氨酶活性,例如,将靶标C·G转化为T·A。在另一实施方案中,碱基编辑活性是腺苷或腺嘌呤脱氨酶活性,例如,将A·T转化为G·C。在另一实施方案中,碱基编辑活性是胞苷脱氨酶活性,例如,将靶标C·G转化为T·A,以及腺苷或腺嘌呤脱氨酶活性,例如,将A·T转化为G·C。

[0262] 术语“碱基编辑器系统”是指用于编辑靶标核苷酸序列的核碱基的系统。在多种实施方案中,碱基编辑器(BE)系统包含(1)多核苷酸可编程核苷酸结合结构域、脱氨酶结构域、和用于将靶标核苷酸序列中的核碱基脱氨的胞苷脱氨酶结构域;和(2)与该多核苷酸可编程核苷酸结合结构域协同作用的一种或多种向导多核苷酸(例如,向导RNA)。在多种实施方案中,碱基编辑器(BE)系统包含选自腺苷脱氨酶或胞苷脱氨酶的核碱基编辑器结构域以及具有核酸序列特异性结合活性的结构域。在一些实施方案中,碱基编辑器系统包含(1)包含多核苷酸可编程DNA结合结构域和用于将靶标核苷酸序列中的一个或多个核碱基脱氨的脱氨酶结构域的碱基编辑器(BE);和(2)与该多核苷酸可编程DNA结合结构域协同作用的一种或多种向导RNA。在一些实施方案中,多核苷酸可编程核苷酸结合结构域是多核苷酸可编程DNA结合结构域。在一些实施方案中,碱基编辑器是胞苷编辑器(CBE)。在一些实施方案中,碱基编辑器是腺嘌呤或腺苷碱基编辑器(ABE)。在一些实施方案中,碱基编辑器是腺嘌呤或腺苷碱基编辑器(ABE)或者胞苷编辑器(CBE)。

[0263] 术语“Cas9”或“Cas9结构域”是指包含Cas9蛋白的RNA引导的核酸酶或其片段(例如,一种蛋白质,其包含Cas9的活性、失活或部分活性的DNA裂解结构域和/或Cas9的gRMA结合结构域)。Cas9核酸酶有时也称为Casn1核酸酶或CRISPR(成簇规则间隔短回文重复序列)相关核酸酶。示例性Cas9是酿脓链球菌(Streptococcus pyogenes)Cas9(spCas9),其氨基酸序列提供于下:

MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETAEATRLKRTARR
RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRK
KLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQIYNQLFEENPINASRVDAKA
ILSARLSKSRRLENLIAQLPGEKRNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLA
QIGDQYADLFLAAKNLSDAILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQS
FIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVT
VKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRG
[0264] MIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGHSLHEQIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQTTQ
KGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHI
VPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL
DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNY
HHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA
RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVK
KDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ
HKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTI
DRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD（单下划线:HNH结构域;双下划线:RuvC结构域)

[0265] 术语“保守氨基酸置换”或“保守突变”是指一种氨基酸被另一种具有共同特性的氨基酸替换。一种定义个体氨基酸之间的共同特性的功能性途径是分析同源生物体的相应蛋白质之间的氨基酸变化的标准化频率(Schulz,G.E.and Schirmer,R.H.,Principles of Protein Structure,Springer-Verlag,New York(1979))。根据此类分析,可定义多组氨基酸,其中一个组内的氨基酸优先彼此交换,并因此在对于总体蛋白质结构的影响中与彼此最为类似(Schulz,G.E.and Schirmer,R.H.,如前)。保守突变的非限制性示例包括以下氨基酸的氨基酸置换:例如,氨基酸置换精氨酸且反之亦然,使得正电荷可得以保持;谷氨酸置换天冬氨酸且反之亦然,使得负电荷可得以保持;丝胺酸置换苏氨酸,使得游离-OH可得以保持;以及谷氨酰胺置换天冬酰胺,使得游离-NH_2可得以保持。

[0266] 如本文中可互换使用的术语“编码序列”或“蛋白质编码序列”是指编码蛋白质的多核苷酸链段。该区域或序列的起始密码子在5’末端附近合,并且终止密码子在3’末端附近。可与本文所述碱基编辑器合用的终止密码子包括下列:

[0267]

谷氨酰胺	**C**AG → **T**AG 终止密码子
	CAA → **T**AA
精氨酸	CGA → **T**GA
色氨酸	TG**G** → TG**A**
	T**G**G → T**A**G
	T**GG** → T**AA**

[0268] 编码序列也可以称为开放阅读框。

[0269] “胞苷脱氨酶”意为能够催化将氨基转化为羰基的脱氨反应的多肽或其片段。在一个实施方案中,胞苷脱氨酶将胞嘧啶转化为尿嘧啶或将5-甲基胞嘧啶转化为胸腺嘧啶。来源于海七鳃鳗(Petromyzon marinus)的PmCDA1(海七鳃鳗胞嘧啶脱氨酶1,“PmCDA1”)、来源于哺乳动物(例如,人、猪、牛、马、猴等)的AID(激活诱导胞苷脱氨酶;AICDA)和APOBEC是示例性的胞苷脱氨酶。

[0270] 如本文中所用,术语“脱氨酶”或“脱氨酶结构域”是指催化脱氨反应的蛋白质或酶。在一些实施方案中,脱氨酶或脱氨酶结构域是分别催化胞苷或脱氧胞苷到尿苷或脱氧尿苷的水解脱氨反应的胞苷脱氨酶。在一些实施方案中,脱氨酶或脱氨酶结构域是催化胞嘧啶到尿嘧啶的水解脱氨反应的胞嘧啶脱氨酶。在一些实施方案中,脱氨酶是腺苷脱氨酶,其催化腺嘌呤至次黄嘌呤的水解性脱氨反应。在一些实施方案中,脱氨酶是腺苷脱氨酶,其催化腺苷或腺嘌呤(A)至肌苷(I)的水解性脱氨反应。在一些实施方案中,脱氨酶或脱氨酶结构域是分别催化腺苷或脱氧肌苷到肌苷或脱氧肌苷的脱氧反应的腺苷脱氨酶。在一些实施方案中,腺苷脱氨酶催化脱氧核糖核酸(DNA)中的腺苷的水解性脱氨反应。本文提供的腺苷脱氨酶(例如,经工程化的腺苷脱氨酶、进化的腺苷脱氨酶)可以来自任何生物体诸如细菌。在一些实施方案中,腺苷脱氨酶来自细菌,诸如大肠杆菌(Escherichia coli)、金黄色葡萄球菌(Staphylococcus aureus)、鼠伤寒沙门氏菌(Salmonella typhimurium)、腐败希瓦氏菌(Shewanella putrefaciens)、流感嗜血杆菌(Haemophilus influenzae)或新月柄杆菌(Caulobacter crescentus)。在一些实施方案中,腺苷脱氨酶是TadA脱氨酶。在一些实施方案中,脱氨酶或脱氨酶结构域是来自生物体诸如人、黑猩猩、大猩猩、猴、牛、狗、大鼠或小鼠的天然出现脱氨酶的变体。在一些实施方案中,脱氨酶或脱氨酶结构域不出现在自然界中。例如,在一些实施方案中,脱氨酶或脱氨酶结构域与天然出现的脱氨酶具有至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%、至少99.1%、至少99.2%、至少99.3%、至少99.4%、至少99.5%、至少99.6%、至少99.7%、至少99.8%或至少99.9%的同一性。

[0271] “检测”指的是证实待检测的被分析物的存在、不存在或量。在一种实施方案中,检测了多核苷酸或多肽中的序列改变。在另一实施方案中,检测了插入缺失(indel)的存在。

[0272] “可检测的标记”意为一种组合物,当连接至感兴趣的分子时,该组合物使得后者可经由光谱学、光化学、生物化学、免疫化学或化学手段检测。例如,有用的标记包括放射性同位素、磁性微珠、金属微珠、胶体离子、荧光染料、高电子密度试剂、酶(例如,一般用于酶

联免疫吸收试验(ELISA)、生物素、地高辛或半抗原。

[0273] “疾病”意为损害或干扰细胞、组织或器官的正常功能的任何病症或病变。疾病的示例包括HBV感染,以及相关疾病或疾患,包括肝硬化、肝细胞癌(HCC)和任何与HBV感染有关或HBV感染导致的其他疾病。

[0274] “有效量”意为相对于未治疗患者缓解疾病症状所需的化合物的量。用来实践本发明的用于治疗性处理疾病的活性化合物的有效量可变,取决于给药模式以及受试者的年龄、体重和一般健康情况。基本上,主治医生或兽医将决定适宜的量和用药方案。这一量被称为“有效”量。在一种实施方案中,有效量是本发明的碱基编辑器的足以向细胞内的HBV基因组中引入改变的量。在一种实施方案中,有效量是实现治疗效果(例如,削弱或控制HBV感染)所需的碱基编辑器的量。此类治疗效果不必足以改变受试者、组织或器官中所有细胞中的HBV基因组,而仅改变受试者、组织或器官中存在的大约1%、5%、10%、25%、50%、75%或更多的细胞中的HBV基因组。在一些实施方案中,有效量足以缓解HBV的一种或多种症候。

[0275] 在一些实施方案中,本文所提供的融合蛋白(例如,包含nCas9结构域和脱氨酶结构域(例如,腺苷脱氨酶、胞苷脱氨酶)的核碱基编辑器)的有效量,是指足以诱导对于通过本文所述的核碱基编辑器特异性地结合并编辑的靶标位点进行编辑的量。如本领域技术人员将会知悉的,药剂例如融合蛋白的有效量可以基于多种因素而变,例如,基于所希望的生物学响应,例如,基于待编辑的具体基因组或靶标为点,基于所靶向的细胞或组装,和/或基于所使用的药剂。

[0276] 在一些实施方案中,本文所提供的融合蛋白(例如,包含nCas9结构域和脱氨酶结构域的融合蛋白)的有效量,是指融合蛋白的足以诱导对于通过该融合蛋白特异性地结合并编辑的靶标位点进行编辑的量。如本领域技术人员将会知悉的,药剂例如融合蛋白的有效量可以基于多种因素而变,例如,基于所希望的生物学响应,例如,基于待编辑的具体基因组或靶标为点,基于所靶向的细胞或组装,和/或基于所使用的药剂。

[0277] “片段”意为多肽或核酸分子的一部分。这一部分优选地含有参考核酸分子或多肽总长度的至少10%、20%、30%、40%、50%、60%、70%、80%或90%。片段可包含10、20、30、40、50、60、70、80、90、或100、200、300、400、500、600、700、800、900、或1000个核苷酸或氨基酸。

[0278] “向导RNA”或“gRNA”意为多核苷酸,其对于靶标序列为特异性的并且可与多核苷酸可编程核苷酸结合结构域蛋白(例如,Cas9或Cpf1)形成复合物。在一种实施方案中,向导多核苷酸是向导RNA(gRNA)。gRNA可作为两种或更多种RNA的复合物存在,或作为单RNA分子存在。作为单RNA分子存在的gRNA可以称为单向导RNA(sgRNA),但“gRNA”用来可互换地指代作为单分子存在或作为两个或更多个分子的复合物存在的向导RNA。典型地,作为单RNA物质存在的gRNA包含两个结构域:(1)与靶标核酸共享同源性的结构域(例如,并且引导Cas9复合物结合至靶标);和(2)结合Cas9蛋白的结构域。在一些实施方案中,结构域(2)对应于作为tracrRNA而为人所知的序列,并且包括茎环结构。例如,在一些实施方案中,结构域(2)与Jinek et al.,Science 337:816-821(2012)中提供的tracrRNA相同或同源,该文献的整体内容通过引用并入本文。gRNA(例如,包括结构域(2)的那些)可以见于名称为“Switchable Cas9 Nucleases and Uses Thereof”的美国专利公开第20160208288号和名称为“Delivery System For Functional Nucleases”的美国专利第9,737,604号,其各自

的整体内容通过引用而以其整体并入本文。在一些实施方案中,gRNA包含两个或更多个结构域(1)和(2),并且可以称为“扩展的gRNA”。扩展的gRNA将会在两个或更多个不同的区域结合两个或更多个Cas9蛋白并且结合靶标核酸,如本文所述。gRNA包含与靶标位点互补的核苷酸序列,该核苷酸序列介导核酸酶/RNA复合物与该靶标位点的结合,提供核酸酶:RNA复合物的序列特异性。

[0279] “HBV聚合酶蛋白”意为与野生型HBV聚合酶S蛋白氨基酸序列具有至少约95%同一性的多肽或其片段,其在乙型肝炎病毒感染中发挥作用。在一种实施方案中,HBV聚合酶由HBV A、B、C、D、E、F、G或H基因型编码。在一种实施方案中,HBV聚合酶氨基酸序列以UniPr登录号Q8B5R0-1提供,其再现如下:

[0280]

10	20	30	40	50
MPLSYQHFRR	LLLLDDEAGP	LEEELPRLAD	EGLNRRVAED	LNLGNLNVSI
60	70	80	90	100
PWTHKVGNFT	GLYSSTVPVF	NPHWKTPSFP	NIHLHQDIIK	KCEQFVGPLT
110	120	130	140	150
VNEKRRLQLI	MPARFYPKVT	KYLPLDKGIK	PYYPEHLVNH	YFQTRHYLHT
160	170	180	190	
LWKAGILYKR	ETTHSASFCG	SPYSWEQDLQ	HGAESFHQQS	

[0281] HBV聚合酶中的突变包括:E24G、L25F、P26F、R27C、V48A、V48I、S382F、V378I、V378A、V379I、V379A、L377F、D380G、D380N、F381P、R376G、A422T、F423P、A432V、M433V、P434S、D540G、A688V、D689G、A717T、E718K、P713S、P713L或L719P。

[0282] 其他示例性HBV DNA聚合酶包括,例如,NCBI登录号AAB59972.1,其具有下述序列。

[0283] MPLSYQHFRKLLLLDDEAGPLEEELPRLADEGLNRRVAEDLNLGNLNVSIPWTHKVGNFTGLYSSTVPV
FNPHWKTPSFPNIHLHQDIIKKCEQFVGPLTVNEKRRLQLIMPARFYPKVTKYLPLDKGIKPYYPEHLVNHYFQTRH
YLHTLWKAGILYKRETTHSASFCGSPYSWEQDLQHGAESFHQQSSGILSRPPVGSSLQSKHSKSRLGLQSQQGHLAR
RQQGRSWSIRAGFHPTARRPFGVEPSGSGHTTNFASKSASCLHQSPDRKAAYPAVSTFEKHSSSGHAVEFHNLSPNS
ARSQSERPVFPCWWLQFRSSKPCSDYCLSLIVNLLEDWGPCAEHGEHHIRIPRTPSRVTGGVFLVDKNPHNTAESRL
VVDFSQFSRGNYRVSWPKFAVPNLQSLTNLLSSNLSWLSLDVSAAFYHLPLHPAAMPHLLVGSSGLSRYVARLSSNS
RILNHQHGTMPNLHDYCSRNLYVSLLLLYQTFGRKLHLYSHPIILGFRKIPMGVGLSPFLLAQFTSAICSVVRRAFP
HCLAFSYMDDVVLGAKSVQHLESLFTAVTNFLLSLGIHLNPNKTKRWGYSLNFMGYVIGSYGSLPQEHIIQKIKECF
RKLPINRPIDWKVCQRIVGLLGFAAPFTQCGYPALMPLYACIQSKQAFTFSPTYKAFLCKQYLNLYPVARQRPGLCQ
VFADATPTGWGLVMGHQRVRGTFSAPLPIHTAELLAACFARSRSGANIIGTDNSVVLSRKYTSYPWLLGCAANWILR
GTSFVYVPSALNPADDPSRGRLGLSRPLLRLPFRPTTGRTSLYADSPSVPSHLPDRVHFASPLHVAWRPP

[0284] “HBV聚合酶基因”意为编码HBV聚合酶的多核苷酸。

[0285] “乙型肝炎表面抗原(HBsAg)多肽”意为与NCBI登录号AAB59969.1具有至少约85%同一性的抗原蛋白或其片段,其在HBV病毒感染中发挥作用。示例性HBsAg氨基酸序列提供于下:

[0286] MENITSGFLGPLLVLQAGFFLLTRILTIPQSLDSWWTSLNFLGGTTVCLGQNSQSPTSNHSPTSCPPTC
PGYRWMCLRRFIIFLFILLLCLIFLLVLLDYQGMLPVCPLIPGSSTTSTGPCRTCMTTAQGTSMYPSCCCTKPSDGN

CTCIPIPSSWAFGKFLWEWASARFSWLSLLVPFVQWFVGLSPTVWLSVIWMMWYWGPSLYSILSPFLPLLPIFFCLW VYI

[0287] “HBsAg多核苷酸”意为编码HBsAg蛋白的多核苷酸。

[0288] “HBV X蛋白”意为与NCBI登录号AAB59970.1具有至少约85％同一性的多肽或其片段,其在HBV病毒感染中发挥作用。示例性氨基酸序列提供于下:

[0289] 1 maarlccqld pardvlclrp vgaescgrpf sgslgtlssp spsavptdhg ahlslrglpv

[0290] 61 cafssagpca lrftsarrme ttvnahrmlp kvlhkrtlgl samsttdlea yfkdclfkdw

[0291] 121 eelgeeirlk vfvlggcrhk lvcapapcnf ftsa

[0292] “核心区抗原前体”意为与NCBI登录号AAB59971.1具有至少约85％同一性的多肽或其片段,其在HBV病毒感染中发挥作用。

[0293] “HBV核心区蛋白”意为与野生型HBV核心区蛋白氨基酸序列具有至少约95％同一性的多肽或其片段。在一种实施方案中,HBV核心区蛋白在乙型肝炎病毒感染中发挥作用。在一种实施方案中,HBV核心区蛋白由HBV A、B、C、D、E、F、G或H基因型编码。在一种实施方案中,HBV核心区蛋白氨基酸序列以NCBI GenBank登录号AXG50928.1提供,提供如下:

[0294] 1 mdidpykefg asvellsflp sdffpsirdl ldtasalyre alespehcsp hhtalrqail

[0295] 61 cwgelmnlat wvgsnledpa srelvvsyvn vnmglkirql lwfhiscltf gretvleylv

[0296] 121 sfgvwirtpp ayrppnapil stlpettvvr rrgrsprrrt psprrrrsqs prrrrsqsre

[0297] 181 sqc

[0298] “HBV X蛋白”意为编码HBV X蛋白的多核苷酸。

[0299] “HBV X蛋白(基因型B)”意为与野生型HBV基因型B X蛋白氨基酸序列具有至少约95％同一性的多肽或其片段。在一种实施方案中,HBV X蛋白在乙型肝炎病毒感染中发挥作用。在一种实施方案中,HBV基因型B X蛋白氨基酸序列以NCBI GenBank登录号BAQ95575.1提供,提供如下:

[0300] 1 maarlccqld pardvlclrp vgaesrgrpl pgplgalppa sppvvpsdhg ahlslrglpv

[0301] 61 cafssxgpca lrftsarrme ttvnahrnlp kvlhkrtlgl samsttdlea yfkdcvfxew

[0302] 121 eelgeexrlk vfvlggcrhk lvcspapcnf ftsa

[0303] “HBV X蛋白(基因型C)”意为与野生型HBV基因型C X蛋白氨基酸序列具有至少约95％同一性的多肽或其片段。在一种实施方案中,HBV X蛋白在乙型肝炎病毒感染中发挥作用。在一种实施方案中,HBV基因型C X蛋白氨基酸序列以NCBI GenBank登录号BAQ95563.1提供,提供如下:

[0304] 1 maarvccqld pardvlclrp vgaesrgrpv sgpfgplpsp sssavpadyg ahlslrglpv

[0305] 61 cafssagpca lrftsarrme ttvnahqvlp kllhkrtlgl samsttdlea yfkdclfkdw

[0306] 121 eelgeeirlk vfvlggcrhk lvcspapcnf ftsa

[0307] “HBV S蛋白”意为与野生型HBV S蛋白氨基酸序列具有至少约95％同一性的多肽或其片段。在一种实施方案中,HBV S蛋白在乙型肝炎病毒感染中发挥作用。在一种实施方案中,HBV S蛋白由HBV A、B、C、D、E、F、G或H基因型编码。在一种实施方案中,HBV S蛋白氨基酸序列以NCBI GenBank登录号ABV02793.1提供,提供如下:

[0308] 1 menttsgflg pllvlqagff lltrnltipq sldswwtsln flggaptcpg qnsqsptsnh

[0309] 61 sptscppicp gyrwmclrrf iiflfilllc lifllvlldy qgmlpvcpll pgtsttstgp

[0310] 121 cktctipaqg tsmfpsccct kpsdgnctci pipsswafar flwewasvrf swlsllvpfv

[0311] 181 qwfvglsptv wlsviwmmwy wgpslynils pflpllpiff clwvyi

[0312] 乙型肝炎病毒亚型ayw的全基因组,是包括编码HBV聚合酶、HBsAg蛋白、HBV X蛋白和核心区抗原前体的多核苷酸的全基因组,以GenBank登录号U95551.1提供,其再现如下:

[0313] 1 aattccacaa cctttcacca aactctgcaa gatcccagag tgagaggcct gtatttccct

[0314] 61 gctggtggct ccagttcagg agcagtaaac cctgttccga ctactgcctc tcccttatcg

[0315] 121 tcaatcttct cgaggattgg ggaccctgcg ctgaacatgg agaacatcac atcaggattc

[0316] 181 ctaggacccc ttctcgtgtt acaggcgggg tttttcttgt tgacaagaat cctcacaata

[0317] 241 ccgcagagtc tagactcgtg gtggacttct ctcaattttc tagggggaac taccgtgtgt

[0318] 301 cttggccaaa attcgcagtc cccaacctcc aatcactcac caacctcctg tcctccaact

[0319] 361 tgtcctggtt atcgctggat gtgtctgcgg cgtttttatca tcttcctctt catcctgctg

[0320] 421 ctatgcctca tcttcttgtt ggttcttctg gactatcaag gtatgttgcc cgtttgtcct

[0321] 481 ctaattccag gatcctcaac caccagcacg ggaccatgcc gaacctgcat gactactgct

[0322] 541 caaggaacct ctatgtatcc ctcctgttgc tgtaccaaac cttcggacgg aaattgcacc

[0323] 601 tgtattccca tcccatcatc ctgggctttc ggaaaattcc tatgggagtg ggcctcagcc

[0324] 661 cgtttctcct ggctcagttt actagtgcca tttgttcagt ggttcgtagg gctttccccc

[0325] 721 actgtttggc tttcagttat atggatgatg tggtattggg ggccaagtct gtacagcatc

[0326] 781 ttgagtccct ttttaccgct gttaccaatt ttcttttgtc tttgggtata catttaaacc

[0327] 841 ctaacaaaac aaagagatgg ggttactctc tgaattttat gggttatgtc attggaagtt

[0328] 901 atgggtcctt gccacaagaa cacatcatac aaaaaatcaa agaatgtttt agaaaacttc

[0329] 961 ctattaacag gcctattgat tggaaagtat gtcaacgaat tgtgggtctt ttgggttttg

[0330] 1021 ctgccccatt tacacaatgt ggttatcctg cgttaatgcc cttgtatgca tgtattcaat

[0331] 1081 ctaagcaggc tttcactttc tcgccaactt acaaggcctt tctgtgtaaa caatacctga

[0332] 1141 acctttaccc cgttgcccgg caacggccag gtctgtgcca agtgtttgct gacgcaaccc

[0333] 1201 ccactggctg gggcttggtc atgggccatc agcgcgtgcg tggaaccttt tcggctcctc

[0334] 1261 tgccgatcca tactgcggaa ctcctagccg cttgttttgc tcgcagcagg tctggagcaa

[0335] 1321 acattatcgg gactgataac tctgttgtcc tctcccgcaa atatacatcg tatccatggc

[0336] 1381 tgctaggctg tgctgccaac tggatcctgc gcgggacgtc ctttgtttac gtcccgtcgg

[0337] 1441 cgctgaatcc tgcggacgac ccttctcggg gtcgcttggg actctctcgt ccccttctcc

[0338] 1501 gtctgccgtt ccgaccgacc acggggcgca cctctcttta cgcggactcc ccgtctgtgc

[0339] 1561 cttctcatct gccggaccgt gtgcacttcg cttcacctct gcacgtcgca tggagaccac

[0340] 1621 cgtgaacgcc caccgaatgt tgcccaaggt cttacataag aggactcttg gactctctgc

[0341] 1681 aatgtcaacg accgaccttg aggcatactt caaagactgt ttgtttaaag actgggagga

[0342] 1741 gttgggggag gagattagat taaaggtctt tgtactagga ggctgtaggc ataaattggt

[0343] 1801 ctgcgcacca gcaccatgca acttttttcac ctctgcctaa tcatctcttg ttcatgtcct

[0344] 1861 actgttcaag cctccaagct gtgccttggg tggctttggg gcatggacat cgacccttat

[0345] 1921 aaagaatttg gagctactgt ggagttactc tcgtttttgc cttctgactt ctttccttca

[0346] 1981 gtacgagatc ttctagatac cgcctcagct ctgtatcggg aagccttaga gtctcctgag

[0347] 2041 cattgttcac ctcaccatac tgcactcagg caagcaattc tttgctgggg ggaactaatg

[0348] 2101 actctagcta cctgggtggg tgttaatttg gaagatccag catctagaga cctagtagtc

[0349] 2161 agttatgtca acactaatat gggcctaaag ttcaggcaac tcttgtggtt tcacatttct

[0350] 2221 tgtctcactt ttggaagaga aaccgttata gagtatttgg tgtctttcgg agtgtggatt

[0351] 2281 cgcactcctc cagcttatag accaccaaat gcccctatcc tatcaacact tccggaaact

[0352] 2341 actgttgtta gacgacgagg caggtcccct agaagaagaa ctccctcgcc tcgcagacga

[0353] 2401 aggtctcaat cgccgcgtcg cagaagatct caatctcggg aacctcaatg ttagtattcc

[0354] 2461 ttggactcat aaggtgggga actttactgg tctttattct tctactgtac ctgtctttaa

[0355] 2521 tcctcattgg aaaacaccat cttttcctaa tatacattta caccaagaca ttatcaaaaa

[0356] 2581 atgtgaacag tttgtaggcc cacttacagt taatgagaaa agaagattgc aattgattat

[0357] 2641 gcctgctagg ttttatccaa aggttaccaa atatttacca ttggataagg gtattaaacc

[0358] 2701 ttattatcca gaacatctag ttaatcatta cttccaaact agacactatt

tacacactct

[0359] 2761 atggaaggcg ggtatattat ataagagaga aacaacacat agcgcctcat tttgtgggtc

[0360] 2821 accatattct tgggaacaag atctacagca tggggcagaa tctttccacc agcaatcctc

[0361] 2881 tgggattctt tcccgaccac cagttggatc cagccttcag agcaaacaca gcaaatccag

[0362] 2941 attgggactt caatcccaac aaggacacct ggccagacgc caacaaggta ggagctggag

[0363] 3001 cattcgggct gggtttcacc ccaccgcacg gaggcctttt ggggtggagc cctcaggctc

[0364] 3061 agggcatact acaaactttg ccagcaaatc cgcctcctgc ctccaccaat cgccagacag

[0365] 3121 gaaggcagcc taccccgctg tctccacctt tgagaaacac tcatcctcag gccatgcagt

[0366] 3181 gg

[0367] “异源二聚体”意为一种融合蛋白,其包含两个结构域,诸如野生型TadA结构域和TadA结构域的变体(例如,TadA*8),或者两个变体TadA结构域(例如,TadA*7.10和TadA*8,或者两个TadA*8结构域)。

[0368] “杂交”意为互补核碱基之间的氢键键合,其可以是沃森-克里克(Watson-Crick)、胡斯坦(Hoogsteen)和反向胡斯坦氢键键合。例如,腺嘌呤和胸腺嘧啶是通过形成氢键而配对的互补核碱基。

[0369] 术语“分离的”、“纯化的”或“生物学纯的”指的是物质没有不同程度的在其天然状态下发现的正常伴生组分。“分离”表示与原始来源或周围物质分隔的程度。“纯化”表示高于“分离”的分隔程度。“纯化的”或“生物学纯的”蛋白质最够不含其它材料,使得任何杂质均不在材料上影响该蛋白质的生物学性质或不造成其它负面后果。换言之,如果本发明的核酸或肽基本上在通过重组DNA技术生产时不含细胞物质、病毒物质或培养基,或者当化学合成时不含化学前体或其它化学品,则该核酸或肽是纯化的。典型使用分析化学技术例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。术语“纯化的”可表示核酸或蛋白质在电泳凝胶中给出实质上的一条谱带。对于可进行修饰如磷酸化或糖基化的蛋白质,不同的修饰可给出不同的分离的蛋白,它们可独立纯化。

[0370] “分离的多核苷酸”意为一种核酸(例如,DNA),其不含在本发明的核酸分子所源自的有机体的天然出现的基因组中位于该基因侧翼的基因。该术语因此包括,举例而言,重组DNA,其被并入载体内、并入自主复制质粒或病毒内、或并入原核生物或真核生物的基因组DNA内,或作为独立于其它序列的单独分子而存在(举例而言,通过PCR或限制性内切酶消化制备的cDNA或基因组或cDNA片段)。此外,该术语包括从DNA分子转录的RNA分子,以及作为编码附加多肽序列的杂交基因的一部分的重组DNA。

[0371] “增加”意为至少10%、25%、50%、75%或100%的正向改变。

[0372] 术语“碱基修复的抑制剂”、“碱基修复抑制剂”、“IBR”或其语法等效物是指一种蛋

白质,其能够抑制核酸修复酶(例如,碱基切除修复酶)的活性。在一些实施方案中,IBR是肌苷碱基切除修复的抑制剂。示例性的碱基修复抑制剂包括APE1、Endo III、Endo IV、Endo V、Endo VIII、Fpg、hOGG1、hNEIL1、T7 Endo1、T4PDG、UDG、hSMUG1和hAAG的抑制剂。在一些实施方案中,碱基修复抑制剂是Endo V或hAAG的抑制剂。在一些实施方案中,IBR是Endo V或hAAG的抑制剂。在一些实施方案中,IBR是无催化活性的EndoV或无催化活性的hAAG。在一些实施方案中,碱基修复抑制剂是无催化活性的EndoV或无催化活性的hAAG。在一些实施方案中,碱基修复抑制剂是尿嘧啶糖苷酶抑制剂(UGI)。UGI是指一种蛋白质,其能够抑制尿嘧啶-DNA糖苷酶碱基切除修复酶。在一些实施方案中,UGI结构域包含野生型UGI或野生型UGI的片段。在一些实施方案中,本文所提供的UGI蛋白包括UGI的片段和与UGI或UGI片段同源的蛋白质。在一些实施方案中,碱基修复抑制剂是肌苷碱基切除修复的抑制剂。在一些实施方案中,碱基修复抑制剂是“无催化活性的肌苷特异性核酸酶”或“死亡的肌苷特异性核酸酶”。不欲受缚于任何特定理论,无催化活性的肌苷糖苷酶(例如,烷基腺嘌呤糖苷酶(AAG))可以结合肌苷,但不能创建无碱基位点或移除该肌苷,从而在空间上阻断DNA损坏/修复机制对新形成的肌苷部分的影响。在一些实施方案中,无催化活性的肌苷特异性核酸酶可以能够结合核酸内的肌苷但不裂解该核酸。非限制性的示例性无催化活性的肌苷特异性核酸酶包括例如来自人的无催化活性的烷基腺苷糖苷酶(AAG核酸酶),和例如来自大肠杆菌的无催化活性的核酸内切酶V(EndoV核酸酶)。在一些实施方案中,无催化活性的AAG核酸酶包含E125Q突变或另一AAG核酸酶中的相应突变。

[0373] “内含肽”是蛋白质的片段,在一个名为蛋白质剪接的过程中,其能够切除自身并且将剩余片段(外显肽)通过肽键接合。内含肽也称为“蛋白质内含子”。内含肽切除自身并且将蛋白质的剩余部分接合的过程在本文中称为“蛋白质剪接”或“内含肽介导的蛋白质剪接”。在一些实施方案中,前体蛋白质(在内含肽介导的蛋白质剪接之前的含有内含肽的蛋白质)的内含肽来自两个基因。本文中,此类内含肽称为分裂内含肽(例如,分裂内含肽-N和分裂内含肽-C)。例如,在蓝细菌中,DnaE(DNA聚合酶III的催化性亚基)由两个独立的基因(dnaE-n和dnaE-c)编码。由dnaE-n基因编码的内含肽在本文中可以称为“内含肽-N”。由dnaE-c基因编码的内含肽在本文中可以称为“内含肽-C”。

[0374] 也可使用其他内含肽系统。例如,基于dnaE内含肽的合成内含肽,Cfa-N(例如,分裂内含肽-N)和Cfa-C(例如,分裂内含肽-C)内含肽对,已经有所描述(例如,在Stevens et al.,J Am Chem Soc.2016Feb.24;138(7):2162-5中,该文献通过引用并入本文)。可根据本公开使用的内含肽对的非限制性示例包括:Cfa DnaE内含肽、Ssp GyrB内含肽、Ssp DnaX内含肽、Ter DnaE3内含肽、Ter ThyX内含肽、Rma DnaB内含肽和Cne Prp8内含肽(例如,如美国专利第8,394,604号中所述,该专利通过引用并入本文)。

[0375] 提供了内含肽的示例性核苷酸和氨基酸序列。

[0376] DnaE内含肽-N DNA:

[0377] TGCCTGTCATACGAAACCGAGATACTGACAGTAGAATATGGCCTTCTGCCAATCGGGAAGATTGTGGAGAAACGGATAGAATGCACAGTTTACTCTGTCGATAACAATGGTAACATTTATACTCAGCCAGTTGCCCAGTGGCACGACCGGGGAGAGCAGGAAGTATTCGAATACTGTCTGGAGGATGGAAGTCTCATTAGGGCCACTAAGGACCACAAATTTATGACAGTCGATGGCCAGATGCTGCCTATAGACGAAATCTTTGAGCGAGAGTTGGACCTCATGCGAGTTGACAACCTTCCTAAT

[0378] DnaE内含肽-N蛋白：

[0379] CLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYCLEDGSLIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNL PN

[0380] DnaE内含肽-C DNA：

[0381] ATGATCAAGATAGCTACAAGGAAGTATCTTGGCAAACAAAACGTTTATGATATTGGAGTCGAAAGAGATCACAACTTTGCTCTGAAGAACGGATTCATAG CTTCTAAT

[0382] 内含肽-C:MIKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN

[0383] Cfa-N DNA：

[0384] TGCCTGTCTTATGATACCGAGATACTTACCGTTGAATATGGCTTCTTGCCTATTGGAAAGATTGTCGAAGAGAGAATTGAATGCACAGTATATACTGTAGACAAGAATGGTTTCGTTTACACACAGCCCATTGCTCAATGGCACAATCGCGGCGAACAAGAAGTATTTGAGTACTGTCTCGAGGATGGAAGCATCATACGAGCAACTAAAGATCATAAATTCATGACCACTGACGGGCAGATGTTGCCAATAGATGAGATATTCGAGCGGGGCTTGGATCTCAAACAAGTGGATGGATTGCCA

[0385] Cfa-N蛋白：

[0386] CLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCLEDGSIIRATKDHKFMTTDGQMLPIDEIFERGLDLKQVDGLP

[0387] Cfa-C DNA：

[0388] ATGAAGAGGACTGCCGATGGATCAGAGTTTGAATCTCCCAAGAAGAAGAGGAAAGTAAAGATAATATCTCGAAAAAGTCTTGGTACCCAAAATGTCTATGATATTGGAGTGGAGAAAGATCACAACTTCCTTCTCAAGAACGGTCTCGTAGCCAGCAAC

[0389] Cfa-C蛋白：

[0390] MKRTADGSEFESPKKKRKVKIISRKSLGTQNVYDIGVEKDHNFLLKNGLVASN

[0391] 内含肽-N和内含肽-C可以分别融合至分裂Cas9的N端部分和分裂Cas9的C端部分，以将分裂Cas9的N端部分和分裂Cas9的C端部分接合。例如，在一些实施方案中，内含肽-N融合至分裂Cas9的N端部分的C端，即，以形成N--[分裂Cas9的N端部分]-[内含肽-N]--C的结构。在一些实施方案中，内含肽-C融合至分裂Cas9的C端部分的N端，即，以形成N-[内含肽-C]--[分裂Cas9的C端部分]-C的结构。用于将内含肽所融合的蛋白质(例如，分裂Cas9)接合的内含肽介导的蛋白质剪接的机制是本领域中已知的，例如，如Shah et al.,Chem Sci.2014;5(1):446-461中所述，该文献通过引用并入本文。用于设计和使用内含肽的方法是本领域中已知的，并且在例如PCT专利公开第WO2014004336号、PCT专利公开第WO2017132580号、美国专利公开第20150344549号和美国专利公开第20180127780号中描述，该些专利各自通过引用而以其整体并入本文。

[0392] “分离的多肽”意为已经与其天然伴随组分分离的本发明的多肽。典型地，以重量计，当多肽的至少60％不含其天然关联的蛋白质和天然出现的有机分子时，则该多肽是分离的。以重量计，优选该制剂含有至少75％、更优选至少90％、最优选99％的本发明的多肽。可通过例如从天然来源提取、编码多肽的重组核酸的表达、或化学合成蛋白质来获得本发明的分离的多肽。可通过任何适宜的方法例如柱层析、聚丙烯酰胺凝胶电泳或HPLC分析来测量纯度。

[0393] 如本文中所用，术语“连接子”可以指共价连接子(例如，共价键)、非共价连接子、

化学基团或分子,其连接两个分子或部分,例如,蛋白复合网或核糖核酸复合物的两个组分,或者融合蛋白的两个结构域,例如,多核苷酸可编程DNA结合结构域(例如,dCas9)和脱氨酶结构域(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)。连接子可以接合碱基编辑器系统的不同组分或者多个组分的不同部分。例如,在一些实施方案中,连接子可以接合多核苷酸可编程核苷酸接合结构域的向导多核苷酸接合结构域与脱氨酶的催化结构域。在一些实施方案中,连接子可以接合CRISPR多肽与脱氨酶。在一些实施方案中,连接子可以接合Cas9与脱氨酶。在一些实施方案中,连接子可以接合dCas9与脱氨酶。在一些实施方案中,连接子可以接合nCas9与脱氨酶。在一些实施方案中,连接子可以接合向导多核苷酸与脱氨酶。在一些实施方案中,连接子可以接合碱基编辑器系统的脱氨组分与多核苷酸可编程核苷酸结合组分。在一些实施方案中,连接子可以结合碱基编辑器的脱氨组分的RNA结合部分与多核苷酸可编程核苷酸结合组分。在一些实施方案中,连接子可以结合碱基编辑器的脱氨组分的RNA结合部分与多核苷酸可编程核苷酸结合组分的RNA结合部分。连接子可以定位在两个基团、分子或其他部分之间或者侧翼具有两个基团、分子或其他部分,并且经由共价键或非共价相互作用与每个基团、分子或其他部分连结,从而将两者连结。在一些实施方案中,连接子可以是有机分子、基团、聚合物或化学部分。在一些实施方案中,连接子可以是多核苷酸。在一些实施方案中,连接子可以是DNA连接子。在一些实施方案中,连接子可以是RNA连接子。在一些实施方案中,连接子可以包含能够结合至配体的适体。在一些实施方案中,配体可以是碳水化合物、肽、蛋白质或核酸。在一些实施方案中,连接子可以包含可以衍生自核糖开关的适体。适体自其衍生的核糖开关可以选自茶碱核糖开关、硫胺素焦磷酸(TPP)核糖开关、腺苷钴胺素(AdoCbl)核糖开关、S-腺苷甲硫氨酸(SAM)核糖开关、SAH核糖开关、黄素单核苷酸(FMN)核糖开关、四氢叶酸核糖开关、赖氨酸核糖开关、甘氨酸核糖开关、嘌呤核糖开关、GlmS核糖开关或前-去氮鸟苷1(pre-queosine 1(PreQ1))核糖开关。在一些实施方案中,连接子可以包含键合之多肽或蛋白质结构域诸如多肽配体的适体。在一些实施方案中,多肽配体可以是K同源(KH)结构域、MS2外壳蛋白结构域、PP7外壳蛋白结构域、SfMu外壳蛋白结构域、不育α基序、端粒酶Ku结合基序和Ku蛋白、端粒酶Sm7结合基序和Sm7蛋白、或RNA识别基序。在一些实施方案中,多肽配体可以是碱基编辑器系统组分的一部分。例如,核碱基编辑组分可以包含脱氨酶结构域和RNA识别基序。

[0394] 在一些实施方案中,连接子可以是氨基酸或多个氨基酸(例如,肽或蛋白质)。一些实施方式中,连接子可以是约5至100个氨基酸的长度,例如,约5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、20至30、30至40、40至50、50至60、60至70、70至80、80至90或90至100个氨基酸的长度。在一些实施方案中,连接子可以是约100至150、150至200、200至250、250至300、300至350、350至400、400至450或450至500个氨基酸的长度。也可以设想更长或更短的连接子。

[0395] 在一些实施方案中,连接子接合RNA可编程核酸酶的gRNA接合结构域(包括Cas9核酸酶结构域)与核酸编辑蛋白(例如,胞苷或腺苷脱氨酶)的催化结构域。在一些实施方案中,连接子接合dCas9与核酸编辑蛋白。例如,连接子可以定位在两个基团、分子或其他部分之间或者侧翼具有两个基团、分子或其他部分,并且经由共价键与每个基团、分子或其他部分连结,从而将两者连结。在一些实施方案中,连接子是氨基酸或多个氨基酸(例如,肽或蛋白质)。在一些实施方案中,连接子是有机分子、基团、聚合物或化学部分。在一些实施方案

中,连接子为5-200个氨基酸的长度,例如,5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、35、45、50、55、60、60、65、70、70、75、80、85、90、90、95、100、101、102、103、104、105、110、120、130、140、150、160、175、180、190或200个氨基酸的长度。

[0396] 在一些实施方案中,碱基编辑器的多个结构域经由连接子融合,该连接子包含SGGSSGSETPGTSESATPESSGGS、SGGSSGGSSGSETPGTSESATPESSGGSSGGS或GGSGGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSGGS的氨基酸序列。在一些实施方案中,核碱基编辑器的多个结构域经由连接子融合,该连接子包含氨基酸序列SGSETPGTSESATPES,其也可称为XTEN连接子。在一些实施方案中,连接子为24个氨基酸的长度。在一些实施方案中,连接子包含氨基酸序列SGGSSGGSSGSETPGTSESATPES。在一些实施方案中,连接子为40个氨基酸的长度。在一些实施方案中,连接子包含氨基酸序列SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGS。在一些实施方案中,连接子为64个氨基酸的长度。在一些实施方案中,连接子包含氨基酸序列SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGSSGSETPGTSESATPESSGGSSGGS。在一些实施方案中,连接子为92个氨基酸的长度。在一些实施方案中,连接子包含氨基酸序列PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATS。

[0397] “标记物”意为任意蛋白质或多核苷酸,其在表达水平或活性上有与疾病或病变相关的改变。

[0398] 如本文中所用,术语“突变”是指序列(例如,核酸或氨基酸序列)内的残基被置换为另一残基,或序列内一个或多个残基的缺失或插入。典型地,本文中通过下述来描述突变:鉴定原始残基,然后鉴定该残基在序列内的位置和新置换的残基的身份。用于进行本文所提供的氨基酸置换(突变)的方法是本领域中众所周知的,并且提供在例如《分子克隆:实验室手册(第四版)》(Green and Sambrook,Molecular Cloning:A Laboratory Manual(4th ed.,Cold Spring Harbor Laboratory Press,Cold Spring Harbor,N.Y.(2012)))中。在一些实施方案中,本文公开的碱基编辑器可以在核酸(例如,受试者基因组内的核酸)中有效地产生“预期突变”诸如点突变,而不产生显著数量的非预期突变诸如非预期点突变。在一些实施方案中,预期突变是通过结合至向导多核苷酸(例如,gRNA)的特异性碱基编辑器(例如,胞苷碱基编辑器或腺苷碱基编辑器)产生的突变,该特异性碱基编辑器经过特异性地涉及以产生预期突变。

[0399] 通常,在序列(例如,本文所述的氨基酸序列)中进行或鉴定的突变是相对于参考(或野生型)序列即不含有该突变的序列进行编号的。本领域技术人员将会容易地理解如何相对于参考序列确定氨基酸序列和核酸序列中突变的位置。

[0400] 术语“非保守突变”包括不同组之间的氨基酸置换,例如,赖氨酸置换色氨酸,或苯丙氨酸置换丝氨酸等。在这种情况下,优选该非保守氨基酸置换不干扰或抑制功能性变体的生物学活性。非保守氨基酸置换可以增强功能性变体的生物学活性,使得该功能性变体的生物学活性比野生型蛋白质增加。

[0401] 在一些实施方案中,本文公开的碱基编辑器可以在核酸(例如,受试者基因组内的核酸)中有效地产生“预期突变”诸如点突变,而不产生显著数量的非预期突变诸如非预期点突变。在一些实施方案中,预期突变是通过结合至向导多核苷酸(例如,gRNA)的特异性碱

基编辑器(例如,胞苷碱基编辑器或腺苷碱基编辑器)产生的突变,该特异性碱基编辑器经过特异性地涉及以产生预期突变。

[0402] 通常,在序列(例如,本文所述的氨基酸序列)中进行或鉴定的突变是相对于参考(或野生型)序列即不含有该突变的序列进行编号的。本领域技术人员将会容易地理解如何相对于参考序列确定氨基酸序列和核酸序列中突变的位置。

[0403] 术语"非保守突变"包括不同组之间的氨基酸置换,例如,赖氨酸置换色氨酸,或苯丙氨酸置换丝氨酸等。在这种情况下,优选该非保守氨基酸置换不干扰或抑制功能性变体的生物学活性。非保守氨基酸置换可以增强功能性变体的生物学活性,使得该功能性变体的生物学活性比野生型蛋白质增加。

[0404] 术语"核定位序列"、"核定位信号"或"NLS"是指促使蛋白质进入细胞核内的氨基酸序列。核定位序列是本领域中已知的,并且在例如Plank等人在2000年11月23日提交的国际PCT申请PCT/EP2000/011690并且在2001年5月31日公布为WO/2001/038547中有所描述,该专利通过引用其对于示例性核定位序列的公开内容而并入本文。在其他实施方案中,NLS是经优化的NLS,例如,Koblan etal.,Nature Biotech.2018doi:10.1038/nbt.4172中所述。在一些实施方案中,NLS包含氨基酸序列KRTADGSEFESPKKKRKV、KRPAATKKAGQAKKKK、KKTELQTTNAENKTKKL、KRGINDRNFWRGENGRKTR、RKSGKIAAIVVKRPRK、PKKKRKV或MDSLLMNRRKFLYQFKNVRWAKGRRETYLC。

[0405] 本文中,术语"核碱基"、"含氮碱基"或"碱基"可互换地使用,是指含有氮的生物化合物,其形成核苷,核苷继而成为核苷酸的组分。核碱基形成碱基对并且一个接一个堆叠的能力直接导致长链螺旋结构,诸如核糖核酸(RNA)和脱氧核糖核酸(DNA)。五种核碱基,即,腺嘌呤(A)、胞嘧啶(C)、鸟嘌呤(G)、胸腺嘧啶(T)和尿嘧啶(U),被称为基本碱基或标准碱基。腺嘌呤和鸟嘌呤衍生自嘌呤,而胞嘧啶、尿嘧啶和胸腺嘧啶衍生自嘧啶。DNA和RNA也可以含有经修饰的其他(非基本)碱基。非限制性的示例性经修饰的核碱基可以包括次黄嘌呤、黄嘌呤、7-甲基鸟嘌呤、5,6-二氢尿嘧啶、5-甲基胞嘧啶(m5C)和5-羟甲基胞嘧啶。次黄嘌呤和黄嘌呤可以通过诱变剂的存在而产生,两者都是通过脱氨反应(将氨基替换为羰基)产生的。次黄嘌呤可以从腺嘌呤修饰而得。黄嘌呤可以从鸟嘌呤修饰而得。尿嘧啶可以从胞嘧啶的脱氨反应获得。"核苷"由核碱基和五碳糖(或核糖或脱氧核糖)组成。核苷的示例包括腺苷、鸟苷、尿苷、胞苷、5-甲基尿苷(m5U)、脱氧腺苷、脱氧鸟苷、胸苷、脱氧尿苷和脱氧胞苷。具有经修饰的核碱基的核苷的示例包括肌苷(I)、黄苷(X)、7-甲基鸟苷(m7G)、二氢尿苷(D)、5-甲基胞苷(m5C)和假尿苷(Ψ)。"核苷酸"由核碱基、五碳糖(或核糖或脱氧核糖)和至少一个磷酸酯基团组成。

[0406] 如本文中所用,术语"核酸"和"核酸分子"是指一种化合物,其包含核碱基和酸性部分,例如,核苷、核苷酸或核苷酸的聚合物。典型地,包含三个或更多个核苷酸的聚合性核酸(例如,核酸分子)是线性分子,其中相邻的核苷酸经由磷酸二酯链接而彼此连接。在一些实施方案中,"核酸"是指个体核酸残基(例如,核苷酸和/或核苷)。在一些实施方案中,"核酸"是指包含三个或更多个个体核苷酸残基的寡核苷酸链。如本文中所用,术语"寡核苷酸"和"多核苷酸"可以互换地使用,是指核苷酸的聚合物(例如,一串至少三个核苷酸)。在一些实施方案中,"核酸"涵盖RNA以及单链和/或双链DNA。核酸可以天然地出现在例如基因组、转录物、mRNA、tRNA、rRNA、siRNA、snRNA、质粒、粘粒、染色体、染色单体或其他天然出现的核

酸分子的情境中。另一方面,核酸分子可以是非天然出现的分子,例如,重组DNA或RNA、人工染色体、经工程化的基因组、或其片段,或者合成DNA、RNA、DNA/RNA杂交物,或者包括非天然出现的核苷酸或核苷。此外,术语“核酸”、“DNA”、“RNA”和/或类似术语包括核酸类似物,例如,具有除磷酸二酯以外的主链的类似物。核酸可以从天然来源纯化,使用重组表达系统生产并任选地进行纯化,化学地合成等。如果合适,例如,在化学合成的分子的情况下,核酸可以包含核苷类似物诸如具有经化学修饰的碱基或糖以及主链修饰的类似物。除非另做说明,否则核酸序列以5’至3’方向呈现。在一些实施方案中,核酸是或包含天然核苷(例如,腺苷、胸苷、鸟苷、胞苷、尿苷、脱氧腺苷、脱氧胸苷、脱氧尿苷和脱氧胞苷);核苷类似物(例如,2-氨基腺苷、2-硫代胸苷、肌苷、吡咯并嘧啶、3-甲基腺苷、5-甲基胞苷、2-氨基腺苷、C5-溴尿苷、C5-氟尿苷、C5-碘尿苷、C5-丙炔基尿苷、C5-丙炔基胞苷、C5-甲基胞苷、2-氨基腺苷、7-脱氮腺苷、7-脱氮鸟苷、8-氧代腺苷、8-氧代鸟苷、O(6)-甲基鸟嘌呤和2-硫代胞苷);经化学修饰的碱基;经生物学修饰的碱基(例如,经甲基化的碱基);经插入的碱基;经修饰的糖(2’修饰,例如,氟核糖、核糖、2’-脱氧核糖、阿拉伯糖和己糖);和/或经修饰的磷酸酯基团(例如,硫代磷酸酯和5’-N-亚磷酰胺链接)。

[0407] 术语“核酸可编程DNA结合蛋白”或“napDNAbp”可以与“多核苷酸可编程核苷酸结合结构域”互换地使用,是指与核酸缔合的蛋白质(例如,DNA或RNA),诸如将napDNAbp引导至具体核酸序列的向导核酸或向导多核苷酸(例如,gRNA)。在一些实施方案中,多核苷酸可编程核苷酸结合结构域是多核苷酸可编程DNA结合结构域。在一些实施方案中,多核苷酸可编程核苷酸结合结构域是多核苷酸可编程RNA结合结构域。在一些实施方案中,多核苷酸可编程核苷酸结合结构域是Cas9蛋白。Cas9蛋白可以与将该Cas9蛋白引导至具体DNA序列的向导RNA缔合,该DNA序列与向导RNA互补。在一些实施方案中,napDNAbp是Cas9结构域,例如,核酸酶活性Cas9、Cas9切口酶(nCas9)或无核酸酶活性的Cas9(dCas9)。核酸可编程DNA结合蛋白的非限制性示例包括Cas9(例如,dCas9和nCas9)、Cas12a/Cpf1、Cas12b/C2c1、Cas12c/C2c3、Cas12d/CasY、Cas12e/CasX、Cas12g、Cas12h和Cas12i。Cas酶的非限制性示例包括Cas1、Cas1B、Cas2、Cas3、Cas4、Cas5、Cas5d、Cas5t、Cas5h、Cas5a、Cas6、Cas7、Cas8、Cas8a、Cas8b、Cas8c、Cas9(也称为Csn1或Csx12)、Cas10、Cas10d、Cas12a/Cpf1、Cas12b/C2c1、Cas12c/C2c3、Cas12d/CasY、Cas12e/CasX、Cas12g、Cas12h、Cas12i、Csy1、Csy2、Csy3、Csy4、Cse1、Cse2、Cse3、Cse4、Cse5e、Csc1、Csc2、Csa5、Csn1、Csn2、Csm1、Csm2、Csm3、Csm4、Csm5、Csm6、Cmr1、Cmr3、Cmr4、Cmr5、Cmr6、Csb1、Csb2、Csb3、Csx17、Csx14、Csx10、Csx16、CsaX、Csx3、Csx1、Csx1S、Csx11、Csf1、Csf2、Cs0、Csf4、Csd1、Csd2、Cst1、Cst2、Csh1、Csh2、Csa1、Csa2、Csa3、Csa4、Csa5、II型Cas效应蛋白、V型Cas效应蛋白、VI型Cas效应蛋白、CARF、DinG、其同源物、或其经修饰或工程化的版本。其他核酸可编程DNA结合蛋白也处于本公开的范畴内,但它们可能不在本公开中具体地列举。参见,例如,Makarova et al.“Classification and Nomenclature of CRISPR-Cas Systems:Where from Here?”CRISPR J.2018Oct;1:325-336.doi:10.1089/crispr.2018.0033;Yan etal.,“Functionally diverse type V CRISPR-Cas systems”Science.2019Jan 4;363(6422):88-91.doi:10.1126/science.aav7271,其各自的整体内容通过引用并入本文。

[0408] 如本文中所用,术语“核碱基编辑结构域”或“核碱基编辑蛋白”是指一种蛋白质或酶,其可以催化RNA或DNA中的核碱基修饰,诸如胞嘧啶(或胞苷)到尿嘧啶(或尿苷)或胸腺

嘧啶(或胸苷)以及腺嘌呤(或腺苷)到次黄嘌呤(或肌苷),以及非模板化的核苷酸加成和插入。在一些实施方案中,核碱基编辑结构域是脱氨酶结构域(例如,腺嘌呤脱氨酶或腺苷脱氨酶;或者胞苷脱氨酶或胞嘧啶脱氨酶)。在一些实施方案中,核碱基编辑结构域是超过一个脱氨酶结构域(例如,腺嘌呤脱氨酶或腺苷脱氨酶;以及胞苷或胞嘧啶脱氨酶)。在一些实施方案中,核碱基编辑结构域可以是天然出现的核碱基编辑结构域。在一些实施方案中,核碱基编辑结构域可以是从天然出现的核碱基编辑结构域工程化或进化的核碱基编辑结构域。核碱基编辑结构域可以来自任何生物体,诸如细菌、人、黑猩猩、大猩猩、猴、牛、狗、大鼠或小鼠。

[0409] 如本文中所用,如“获得一剂”中的“获得”包括合成、购买或以其它方法取得该剂。

[0410] 如本文中所用,“患者”或“受试者”是指哺乳动物受试者或个体,其被诊断患有疾病或疾患、处于患有或发展出疾病或疾患的风险下或易患或发展出疾病或疾患。在一些实施方案中,术语“患者”是指哺乳动物受试者,其具有高于平均水平的发展出疾病或疾患的可能性。示例性患者可以是人、非人灵长动物、猫、狗、猪、牛、猫、马、骆驼、骆马、山羊、绵羊、啮齿动物(例如,小鼠、兔、大鼠或豚鼠)和其他哺乳动物,其可能受益于本文所公开的疗法。示例性人类患者可以是男性和/或女性。

[0411] 本文中,“有此需要的患者”或“有此需要的受试者”是指患者,其被诊断患有疾病或疾患、处于患有疾病或疾患的风险下、预先查明患有疾病或疾患或易患疾病或疾患。

[0412] 本文中,术语“蛋白质”、“肽”、“多肽”及其语法等效物可互换地使用,并且是指通过肽(酰胺)键连接在一起的氨基酸残基的聚合物。这些术语是指任何大小、结构或功能的蛋白质、肽或多肽。典型地,蛋白质、肽或多肽的长度将会为至少三个氨基酸。蛋白质、肽或多肽可以指单个蛋白质或蛋白质的集合。可例如通过加入化学整体诸如碳水化合物基团、羟基、磷酸根、法尼基、异法尼基、脂肪酸基团、用于缀合、功能化或其他修饰等的链接基或其它修饰而修饰蛋白质、肽或多肽中的一个或多个氨基酸。蛋白质、肽或多肽也可以为单分子或可以为多分子复合物。蛋白质、肽或多肽可以仅为天然出现的蛋白质或肽的片段。蛋白质、肽或多肽可以为天然出现的、重组的或合成的,或其任意组合。如本文所用,术语“融合蛋白”是指一种杂交多肽,其包含来自至少两种不同蛋白质的蛋白结构域。一个蛋白质可以定位在融合蛋白的氨基端(N端)部分或定位在羧基端(C端)蛋白,从而分别形成氨基端融合蛋白或羧基端融合蛋白。蛋白质可以包含不同的结构域,例如,核酸结合结构域(例如,Cas9的引导该蛋白质结合至靶标位点的gRNA结合结构域)和核酸编辑蛋白的核酸裂解结构域或催化结构域。在一些实施方案中,蛋白质包含蛋白部分(例如,构成核酸结合结构域的氨基酸序列)和有机化合物(例如,可充当核酸裂解剂的化合物)。在一些实施方案中,蛋白质与核酸例如RNA或DNA复合或缔合。本文所提供的任何蛋白质均可通过本领域已知的任何方法生产。例如,本文所提供的蛋白质可以经由重组蛋白质表达和纯化生产,该方法尤其适用于包含肽连接子的融合蛋白。用于重组蛋白表达和纯化的方法是周知的,并且包括《分子克隆:实验室手册(第四版)》(Green and Sambrook,Molecular Cloning:A Laboratory Manual(4th ed.,Cold Spring Harbor Laboratory Press,Cold Spring Harbor,N.Y.(2012)))中描述的那些,该文献的整体内容通过引用并入本文。

[0413] 本文所公开的多肽和蛋白质(包括其功能性部分和功能性变体)可以包含合成氨基酸来代替一个或多个天然出现的氨基酸。此类合成氨基酸是本领域中已知的,并且包括

例如氨基环己烷甲酸、正亮氨酸、α-氨基正癸酸、高丝氨酸、S-乙酰基氨基甲基-半胱氨酸、反式-3-和反式-4-羟基脯氨酸、4-氨基苯丙氨酸、4-硝基苯丙氨酸、4-氯苯丙氨酸、4-羧基苯丙氨酸、β-苯基丝氨酸、β-羟基苯丙氨酸、苯基甘氨酸、α-萘基丙氨酸、环己基丙氨酸、环己基甘氨酸、吲哚啉-2-甲酸、1,2,3,4-四氢异喹啉-3-甲酸、氨基阿洛糖酸、氨基阿洛糖酸单酰胺、N'-苄基-N'-甲基-赖氨酸、N',N'-二苄基-赖氨酸、6-羟基赖氨酸、鸟氨酸、α-氨基环戊烷甲酸、α-氨基环己烷甲酸、α-氨基环庚烷甲酸、α-(2-氨基-2-降莰烷)-甲酸、α,γ-二氨基丁酸、α,β-二氨基丙酸、高苯丙氨酸和α-叔丁基甘氨酸。多肽和蛋白质可以与多肽构建体的一个或多个氨基酸的转录后修饰缔合。转录后修饰的非限制性示例包括磷酸化、酰化(包括乙酰化和甲酰化)、糖基化(包括N-联糖基化和O-联糖基化)、酰胺化、羟基化、烷基化(包括甲基化和乙基化)、泛素化、吡咯烷酮甲酸的加成、二硫桥的形成、硫酸化、豆蔻酰化、棕榈酰化、异戊二烯化、法尼基化、香叶基化、糖基磷脂酰肌醇化、脂化和碘化。

[0414] 如本文所用,术语"重组"在蛋白质或核酸的语境中是指不出现在自然界中而是作为人工工程化产物的蛋白质或核酸。例如,在一些实施方案中,重组蛋白质或核酸包含氨基酸或核苷酸序列,与任何天然出现的序列相比,该氨基酸或核苷酸序列包含至少一个、至少两个、至少三个、至少四个、至少五个、至少六个、至少七个突变。

[0415] "降低"意为至少10%、25%、50%、75%或100%的负向改变。

[0416] "参考"意为标准或对照条件。在一种实施方案中,将存在于用本文所述的碱基编辑器系统处理的细胞中的病毒负载与存在于未经处理的对照细胞中的HBV感染水平进行比较,该对照用作参考。在另一种实施方案中,将存在于与本文所述的碱基编辑器系统接触的细胞中的HBV基因组序列与存在于未经处理的对照细胞中的HBV基因组序列进行比较。

[0417] "参考序列"是用作序列比对基础的定义序列。参考序列可以是特定序列的一部分或整体;例如,全长度cDNA或基因序列的链段、或完整的cDNA或基因序列。对于多肽,参考多肽序列的长度将通常为至少约16个氨基酸、至少约20个氨基酸、至少约25个氨基酸、约35个氨基酸、约50个氨基酸或约100个氨基酸。对于核酸,参考核酸序列的长度将通常为至少约50个核苷酸、至少约60个核苷酸、至少约75个核苷酸、约100个核苷酸或约300个核苷酸或与之接近或之间的任何整数。在一些实施方案中,参考序列是感兴趣的蛋白质的野生型序列。在其他实施方案中,参考序列是编码野生型蛋白质的多核苷酸序列。

[0418] 术语"RNA可编程核酸酶"和"RNA引导的核酸酶"与并非裂解靶标的一种或多种RNA合用(例如,与之结合或缔合)。在一些实施方案中,当与RNA复合时,RNA可编程核酸酶可以称为核酸酶:RNA复合物。典型地,所结合的RNA称为向导RNA(gRNA)。在一些实施方案中,RNA可编程核酸酶是(CRISPR相关系统)Cas9核酸内切酶,例如,来自酿脓链球菌(Streptococcus pyogenes)的Cas9(Csn1)(参见,例如,"Complete genome sequence of an M1 strain of Streptococcus pyogenes."Ferretti J.J.,McShan W.M.,Ajdic D.J.,Savic D.J.,Savic G.,Lyon K.,Primeaux C,Sezate S.,Suvorov A.N.,Kenton S.,Lai H.S.,Lin S.P.,Qian Y.,Jia H.G.,Najar F.Z.,Ren Q.,Zhu H.,Song L.,White J.,Yuan X.,Clifton S.W.,Roe B.A.,McLaughlin R.E.,Proc.Natl.Acad.Sci.U.S.A.98:4658-4663(2001);"CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III."Deltcheva E.,Chylinski K.,Sharma CM.,Gonzales K.,Chao Y.,Pirzada Z.A.,Eckert M.R.,Vogel J.,Charpentier E.,Nature 471:602-607(2011))。

[0419] “特异性地结合”意为核酸分子、多肽或其复合物(例如,核酸可编程DNA结合结构域和向导核酸)、化合物或分子识别并且结合本发明的多肽和/或核酸分子,但它们基本上不识别并且结合样品例如生物样品中的其他分子。

[0420] 可用于本发明方法的核酸分子包括编码本发明的多肽或其片段的任何核酸分子。此类核酸分子无须与内源性核酸序列100%一致,但典型将会展现基本一致性。具有与内源性序列“基本一致性”的多核苷酸典型能与双链核酸分子的至少一条链杂交。可用于本发明方法的核酸分子包括编码本发明的多肽或其片段的任何核酸分子。此类核酸分子无须与内源性核酸序列100%一致,但典型将会展现基本一致性。具有与内源性序列“基本一致性”的多核苷酸典型能与双链核酸分子的至少一条链杂交。“杂交”意为在互补多核苷酸序列(如,本文中揭示的基因)或其部分在多种严格条件下配对以形成双链分子。(见,例如,Wahl,G.M.and S.L.Berger(1987)Methods Enzymol.152:399;Kimmel,A.R.(1987)Methods Enzymol.152:507)。

[0421] 举例而言,严格的盐浓度一般为少于约750mM NaCl和75mM柠檬酸三钠,优选少于约500mM NaCl和50mM柠檬酸三钠,且更优选少于约250mM NaCl和25mM柠檬酸三钠。低严格度杂交可在不存在有机溶剂如甲酰胺的情况下获得,而高严格度杂交可在存在至少约35%甲酰胺且更优选至少约50%甲酰胺的情况下获得。严格的温度条件一般将包括至少约30℃,更优选至少约37℃,且最优选至少约42℃的温度。可变的附加因素,如杂交时间、洗涤剂如十二烷基硫酸钠(SDS)的浓度、以及包含或不饱和载体DNA,是该领域技术人员所周知的。通过按需要组合这些多种条件来实施各种水平的严格度。在优选的实施方案中,杂交将出现在30℃的750mM NaCl、75mM柠檬酸三钠和1% SDS中。在更优选的实施方案中,杂交将出现在37℃的500mM NaCl、50mM柠檬酸三钠、1% SDS、35%甲酰胺和100μg/ml变性鲑鱼精DNA(ssDNA)中。在最优选的实施方案中,杂交将出现在42℃的250mM NaCl、25mM柠檬酸三钠、1% SDS、50%甲酰胺和200μg/ml ssDNA中。对这些条件的有用改变对于该领域技术人员是显而易见的。

[0422] 对于大多数应用,杂交后的洗涤步骤的严格度也将改变。可通过盐浓度和温度来界定洗涤严格度条件。如上,可通过降低盐浓度或增加温度来增加洗涤严格度。举例而言,洗涤步骤的严格的盐浓度为少于约30mM NaCl和3mM柠檬酸三钠,或少于约15mM NaCl和1.5mM柠檬酸三钠。用于洗涤步骤的严格的温度条件一般将包括至少约25℃、至少约42℃、或至少约68℃的温度。在一种实施方案中,洗涤步骤将出现在25℃的30mM NaCl、3mM柠檬酸三钠和0.1% SDS中。在另一种实施方案中,洗涤步骤将出现在42℃的15mM NaCl、1.5mM柠檬酸三钠和0.1% SDS中。在更优选的实施方案中,洗涤步骤将出现在68℃的15mM NaCl、1.5mM柠檬酸三钠和0.1% SDS中。对这些条件的额外改变对于该领域技术人员是显而易见的。杂交技术是该领域技术人员所周知的,且揭示在例如Benton and Davis(Science 196:180,1977);Grunstein and Hogness(Proc.Natl.Acad.Sci.,USA 72:3961,1975);《分子生物学现代技术》(Ausubel et al.(Current Protocols in Molecular Biology,Wiley Interscience,New York,2001));《分子克隆技术指南》(Berger and Kimmel(Guide to Molecular Cloning Techniques,1987,Academic Press,New York));以及《分子克隆:实验室手册》(Sambrook et al.,Molecular Cloning:ALaboratory Manual,Cold Spring Harbor Laboratory Press,New York)中。

[0423] “分裂”意为分为两个或更多个片段。

[0424] “分裂Cas9蛋白”或“分裂Cas9”是指一种Cas9蛋白,其作为由两个独立的核苷酸序列编码的N端片段和C端片段而提供。对应于Cas9蛋白的N端部分和C端部分的多肽可以被分裂以形成“重建”Cas9蛋白。在特定实施方案中,Cas9蛋白被分为处于该蛋白质的无序区内的两个片段,例如,如Nishimasu et al.,Cell,Volume 156,Issue5,pp.935-949,2014中所述或如Jiang et al.(2016)Science351:867-871.PDB file:5F9R中所述,其各自通过引用并入本文。在一些实施方案中,在位于SpCas9的介于大约氨基酸A292至G364、F445至K483或E565至T637之间的区域内的任何C、T、A或S处,或在任何其他Cas9、Cas9变体(例如,nCas9、dCas9)或其他napDNAbp的相应位置处,将蛋白质分为两个片段。在一些实施方案中,在SpCas9T310、T313、A456、S469或C574处,将蛋白质分为两个片段。在一些实施方案中,将蛋白质分为两个片段的过程称为“剪接”该蛋白质。

[0425] 在其他实施方案中,Cas9蛋白的N端部分包含酿脓链球菌Cas9野生型(SpCas9)(NCBI参考序列:NC_002737.2,Uniprot参考序列:Q99ZW2)的氨基酸1至573或1至637S,并且Cas9蛋白的C端部分包含SpCas9野生型的氨基酸574至1368或638至1368部分。

[0426] 分裂Cas9的C端部分可以与分裂Cas9的N端部分接合以形成完全的Cas9蛋白。在一些实施方案中,Cas9蛋白的C端部分始于Cas9蛋白的N端部分结束之处。因此,在一些实施方案中,分裂Cas9的C端部分包含spCas9的氨基酸(551至651)至1368部分。“(551至651)至1368”意为始于氨基酸551至651(含)之间的氨基酸处并且终于氨基酸1368处。例如,分裂Cas9的C端部分可包含spCas9的氨基酸551至1368、552至1368、553至1368、554至1368、555至1368、556至1368、557至1368、558至1368、559至1368、560至1368、561至1368、562至1368、563至1368、564至1368、565至1368、566至1368、567至1368、568至1368、569至1368、570至1368、571至1368、572至1368、573至1368、574至1368、575至1368、576至1368、577至1368、578至1368、579至1368、580至1368、581至1368、582至1368、583至1368、584至1368、585至1368、586至1368、587至1368、588至1368、589至1368、590至1368、591至1368、592至1368、593至1368、594至1368、595至1368、596至1368、597至1368、598至1368、599至1368、600至1368、601至1368、602至1368、603至1368、604至1368、605至1368、606至1368、607至1368、608至1368、609至1368、610至1368、611至1368、612至1368、613至1368、614至1368、615至1368、616至1368、617至1368、618至1368、619至1368、620至1368、621至1368、622至1368、623至1368、624至1368、625至1368、626至1368、627至1368、628至1368、629至1368、630至1368、631至1368、632至1368、633至1368、634至1368、635至1368、636至1368、637至1368、638至1368、639至1368、640至1368、641至1368、642至1368、643至1368、644至1368、645至1368、646至1368、647至1368、648至1368、649至1368、650至1368或651至1368中任一者的部分。在一些实施方案中,分裂Cas9蛋白的C端部分包含spCas9的氨基酸574至1368或638至1368部分。

[0427] “受试者”意为哺乳动物,包括但不限于人类和非人哺乳动物,诸如非人灵长动物(猴)、牛、马、犬、羊或猫科动物。在一些实施方案中,本文所述的受试者被HBV感染或具有发展出HBV的潜力。

[0428] “基本相同”意为多肽或核酸分子展现与参考氨基酸序列(例如,任何一种本文所述的氨基酸序列)或核酸序列(例如,任何一种本文所述的核酸序列)的至少50%一致性。在

一种实施方案中,此序列在氨基酸水平或核酸水平上与用于比较的序列的一致性为至少60%、80%、85%、90%、95%或甚至99%。

[0429] 通常使用序列分析软件(例如,威斯康星大学生物技术中心的遗传学计算机公司(Genetics Computer Group,University of Wisconsin Biotechnology Center,1710University Avenue,Madison,Wis.53705)的序列分析软件包BLAST、BESTFIT、GAP、或PILEUP/PRETTYBOX程序)测量序列一致性。该软件通过设定多种替换、删除、及/或其它修饰的同源性程度来匹配一致或相似的序列。保守替换典型包括下述各组的组内替换:甘氨酸、丙氨酸;缬氨酸、异亮氨酸、亮氨酸;天冬氨酸、谷氨酸、天冬酰胺、谷氨酰胺;丝氨酸、苏氨酸;赖氨酸、精氨酸;以及苯丙氨酸、酪氨酸。在例示性的测定一致性程度的途径中,可使用BLAST程序,其中介于e^{-3}与e^{-100}之间的可能性得分指示密切相关的序列。使用了例如具有以下参数的COBALT:

[0430] a)比对参数:空位罚分为-11、-1,并且末端空位罚分为-5、-1,

[0431] b)CDD参数:使用RPS BLAST开启;Blast E-值0.003;找到保存的列并重新计算(Find Conserved columns and Recompute)开启,和

[0432] c)查询聚类参数:使用查询簇开启;字长4;最大簇间距离0.8;

[0433] 自体风格为Regular。

[0434] 使用了例如具有以下参数的EMBOSS Needle:

[0435] a)基质:BLOSUM62;

[0436] b)空位开头(GAP OPEN):10;

[0437] c)空位延长(GAP EXTEND):0.5;

[0438] d)输出格式:对;

[0439] e)末端空位罚分:错;

[0440] f)末端空位开头:10;和

[0441] g)末端空位延长:0.5。

[0442] 术语“靶标位点”是指核酸分子内的序列,其被脱氨酶或包含脱氨酶(例如,胞苷或腺嘌呤脱氨酶)的融合蛋白(本文所公开的融合蛋白或碱基编辑器)脱氨基。

[0443] 因为RNA可编程核酸酶(例如,Cas9)使用RNA:DNA杂交作用来靶向DNA裂解位点,所以这些蛋白质原则上能够被靶向至向导RNA所特异化的任何序列。使用RNA可编程核酸酶诸如Cas9进行位点特异性裂解(例如,以修饰基因组)的方法是本领域中已知的(参见,例如,Cong,L.et al.,Multiplex genome engineering using CRISPR/Cas systems.Science 339,819-823(2013);Mali,P.et al.,RNA-guided human genome engineering via Cas9.Science 339,823-826(2013);Hwang,W.Y.et al.,Efficient genome editing in zebrafish using a CRISPR-Cas system.Nature biotechnology31,227-229(2013);Jinek,M.et al.,RNA-programmed genome editing in human cells.eLife 2,e00471(2013);Dicarlo,J.E.et al.,Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems.Nucleic acids research(2013);Jiang,W.et al.,RNA-guided editing of bacterial genomes using CRISPR-Cas systems.Nature biotechnology 31,233-239(2013);其各自的整体内容通过引用并入本文)。

[0444] 如本文中所用,“治疗”等指的是减轻或缓解与其相关的病变和/或症状或者获得

所希望的药理学和/或生理学效果。应知晓,尽管未排除,但治疗病症或症状并不需要完全消除与该病症或症状相关的病症、症候或症状。在一些实施方案中,该效果是治疗性的,即,而不限于,部分地或完全地削弱、减少、废止、减弱、减轻、降低疾病和/或可归因于疾病的不良症状的强度或治愈该疾病和/或可归因于疾病的不良症状。在一些实施方案中,该效果是预防性的,即,保护或防止疾病或病症的发作或复发的效果。就此而言,本发明公开的方法包括给药治疗有效量的如本文所述的组合物。在一种实施方案中,本发明提供对HBV感染的治疗。

[0445] “尿嘧啶糖苷酶抑制剂”或“UGI”意为抑制尿嘧啶切除修复系统的药剂。在一种实施方案中,该药剂是蛋白质或其片段,其结合宿主尿嘧啶-DNA糖苷酶并且防止尿嘧啶残基被从DNA中移除。在一种实施方案中,UGI是指一种蛋白质、其片段或结构域,其能够抑制尿嘧啶-DNA糖苷酶碱基切除修复酶。在一些实施方案中,UGI结构域包含野生型UGI或其经修饰的版本。在一些实施方案中,UGI结构域包含下文详述的示例性氨基酸序列的片段。在一些实施方案中,UGI片段包含氨基酸序列,所述氨基酸序列包含下文提供的示例性UGI序列的至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%。在一些实施方案中,UGI包含与示例性GI氨基酸序列或其片段同源的氨基酸序列,如下文详述。在一些实施方案中,UGI或其部分与野生型UGI或UGI序列或其部分为至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%、至少99.5%、至少99.9%或100%相同,如下文详述。示例性UGI包含以下氨基酸序列:

[0446] >splP14739IUNGI_BPPB2尿嘧啶-DNA糖苷酶抑制剂

[0447] MTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLT SD APE YKPW ALVIQDS NGENKIKML.

[0448] 本文中提供的范围理解为该范围内所有值的略写。例如,1至50的范围理解为包括来自由1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49或50所组成的组的任意数字、数字组合或子范围。

[0449] 本文中,对任何变量定义中一系列化学基团的描述包括该变量作为任何单一基团或作为所列基团的组合的定义。对本文中变量或方面的实施方案的描述包括该实施方案作为任何单一输送方案或作为与任何其它实施方案或其部分的组合。

[0450] 本文中提供的任何组合物或方法可与本文中提供的一种或多种任何其它组合物和方法组合。

[0451] 本文的说明和实施例详细地例示性说明本公开的实施方案。应理解,本公开不限于本文所述的特定实施方案并因此可变。本领域技术人员应了解,存在大量关于本公开的改变和修改,而这些改变和修改还该在本公开范畴内。

[0452] 所有术语预期被理解为本领域人员将会理解的意思。除非另做定义,否则本文中使用的所有科技术语均具有与本公开所属领域技术人员所一般理解的相同的意义。

[0453] 除非另做指定,否则本文所公开的一些实施方案的实践采用免疫学、生物化学、化学、分子生物学、微生物学、细胞生物学、基因组和重组DNA的传统技术,这些技术属于本领域的技术。参见,例如,Sambrook和Green的《分子克隆:实验室手册(第四版)》(Molecular

Cloning:A Laboratory Manual,4th Edition(2012));F.M.Ausubel等人编撰的《分子生物学实验指南》的系列(the series Current Protocols in Molecular Biology);《酶学方法》的系列(the series Methods In Enzymology(Academic Press,Inc.));M.J.MacPherson\B.D.Hames和G.R.Taylor编撰的《PCR 2:实用方法》(PCR 2:A Practical Approach(1995));Harlow和Lane编撰的《抗体,实验室手册》((1988)Antibodies,A Laboratory Manual)以及R.I.Freshney编撰的《动物细胞培养:基本技术和专业应用手册》(Culture of Animal Cells:A Manual of Basic Technique and Specialized Applications,6th Edition(2010))。

[0454] 尽管本公开的多个特征可以在单实施方案上下文中描述,但这些特征也可以单独提供或以任何合适的组合方式提供。相反,尽管为了清楚起见,本公开可以在本文的单独实施方案上下文中描述,但本公开也可以在单实施方案中实现。本文使用的章节标题仅用于组织性目的,而不视为限制所描述的主题。

[0455] 本公开的特征在所附权利要求书中具体阐述。参考以下阐述例示性实施方案的具体实施方式(其中利用了本公开的原理),并且参照如下文所述的附图,可以获得对本公开的特征和优点的更佳理解。

附图说明

[0456] 图1是例示性说明,其显示用于乙型肝炎表面抗原(HBsAg)基因、聚合酶基因、蛋白X基因和核心区基因的部分双链的和重叠开放阅读框(ORF)。HBsAg基因包含ORF PreS1、ORF PreS2和ORF S,其分别编码大、中和小表面蛋白。ORF核心区和Pre C编码衣壳蛋白。

[0457] 图2是例示性说明,其描绘HBV生命周期。术语“ER”表示内质网。术语“HBsAg”表示乙型肝炎表面抗原。“HBx转录激活因子”是聚合酶II和III启动子的HBV特异性转录激活因子。

[0458] 图3A提供了用于将终止密码子引入病毒基因内和用于产生无碱基位点以治疗慢性HBV的碱基编辑策略的总结。

[0459] 图3B提供了适合用于引入终止密码子和用于经由碱基编辑产生无碱基的向导RNA删选策略的总结。

[0460] 图3C是示意图,例示性说明用于在cccDNA中产生无碱基位点的保守gRNA设计。

[0461] 图3D是HBV cccDNA的示意图,显示被预期产生在HBV基因组中出现率小于0.05%的氨基酸的16个向导RNA(以三角形描绘)的相对位置。

[0462] 图3E是图表,显示了通过gRNA候选物产生的碱基编辑的最高百分比。

[0463] 图3F是总结关于gRNA候选物信息的图表。

[0464] 图4A和4B描绘碱基编辑器。图4A描绘了具有APOBEC胞苷脱氨酶结构域、Cas9结构域和两个尿嘧啶糖基化酶抑制剂(UGI)结构域的碱基编辑器。图4B提供BE4的示意图。

[0465] 图5是例示性说明,其显示本公开的向导RNA在何处映射至HBV基因组。每个三角形代表独特的向导RNA。

[0466] 图6是示意性的例示性说明,其总结了对于靶向HBV基因的向导RNA分子的筛选和从该筛选中观察到的结果的子集。“PAM”是指前间隔序列邻近基序。“Pol”是指HBV聚合酶基因;“S”是指HBV表面蛋白;“X”是指HBV蛋白X基因;而“核心区”是指HBV核心区蛋白。

MSPbeam52、50、…、等等是指向导RNA,在本申请中,它们也称为M52、M50、…、等等。该筛选鉴定了12种gRNA,其表现出大于20％的上靶碱基编辑。

[0467] 图7包含将BE4和A3ABE4碱基编辑器进行比较的图表。这些图表显示针对与各碱基编辑器一起使用的不同向导RNA观察到的编辑百分比。在本申请中,MSPbeam39、40、…、等等也称为M39、M40、…、等等。

[0468] 图8是图表,其例示性说明在用HBV慢病毒的载体转导并且用编码碱基编辑器的核酸构建体转染的HepG2-NTCP细胞系中观察到的功能性碱基编辑。所测试的核酸构建体是DNA分子、野生型RNA分子、或包含在N1残基处修饰的假尿苷(PsU)的RNA分子。“NTCP”是指牛胆酸钠共同转运多肽。在本申请中,MSPbeam39、40、…、等等也称为M39、M40、…、等等。

[0469] 图9是图表,其例示性说明在用HBV慢病毒的载体转导并且用编码碱基编辑器的核酸构建体转染的HepG2-NTCP细胞系中观察到的功能性碱基编辑。所测试的核酸构建体是DNA格式转染(两种质粒,一种编码碱基编辑器,一种编码gRNA)或RNA格式(编码碱基编辑器的PsU修饰的内部mRNA,其中RNA在N1残基处被修饰;以及合成的gRNA)。当使用碱基编辑器和RNA格式的效果更佳的Stop/功能改变(“FC”)gRNA时,在HepG2-NTCP lenti HBV细胞共观察到高达80％的编辑。“NTCP”是指牛胆酸钠共同转运多肽。在本申请中,MSPbeam39、40、…、等等也称为M39、M40、…、等等。

[0470] 图10是图表,其例示性说明在用HBV慢病毒的载体转导并且用编码BE4、BE4-VRQR或ABE碱基编辑器的三种核酸构建体之一转染的HepG2-NTCP细胞系中观察到的功能性碱基编辑。在本申请中,MSPbeam39、40、…、等等也称为M39、M40、…、等等。

[0471] 图11是例示性说明,其描绘映射至HBV基因组的保守区的向导RNA。

[0472] 图12A是示意图,例示性说明长期原代肝细胞共培养物。图12B提供了肝细胞单层或肝细胞共培养物的实验时间线。图12C显示经转导的来自在共培养系统中使用的供体(RSE、TVR)的原代肝细胞的图像。

[0473] 图13A至13F表征HBV感染的原代人肝细胞(PHH)系统。图13A是时间线,显示从铺板到研究终点的13天感染和治疗时间表。图13B是图表,其显示在不治疗HBV感染的PHH细胞、用干扰素治疗或用替诺福韦治疗后,PHH培养物中存在的细胞外HBV DNA的量。作为阴性对照,将PHH细胞暴露于没有聚乙二醇的HBV病毒。图13C是图表,其显示暴露于针对图13B和13C的图例中所述实验条件下的PHH培养物中存在的HBV表面抗原(HBsAg)的量。图13D是图表,其显示暴露于针对图13B和13C的图例中所述实验条件下的PHH培养物中存在的细胞内HBV DNA的量。图13E是图表,其显示暴露于针对图13B和13C的图例中所述实验条件下的PHH培养物中存在的总HBV RNA的量。图13F是图表,其显示暴露于针对图13B和13C的图例中所述实验条件下的PHH培养物中存在的前基因组RNA(pgRNA)的量。

[0474] 图14是图表,其显示用BE4和gRNA进行转染导致了HBV感染的PHH中HBV标记物水平的降低。使用了向导RNA 52和190,其分别将BE4碱基编辑器靶向至HBV基因组的Pol和X基因区。测试了具有和不具有尿苷糖基化酶抑制剂(UGI)结构域的BE4。

[0475] 图15是图表,其显示对在HBV感染的PHH细胞中进行的筛选中的功能性向导RNA的鉴定,其中作为cccDNA替代指标的HBsAg说的下降是功能性gRNA的标志。向导RNA引入的终止密码子标注为终止-39等。向导RNA引入的在保守氨基酸处的改变表示为保守-4等。选择进行进一步分析的gRNA(Stop-191、保守-12)以方框指明。

[0476] 图16A和16B示例性说明了作用于HBV上的碱基编辑作用的机理方面。图16A是图表,其显示HBV感染的PHH细胞中的HBsAg水平,该细胞用编码下列的mRNA转染:具有UGI结构域的BE4碱基编辑器(BE4)、没有UGI结构域的BE4碱基编辑器(BE4_noUGI)、Cas9、没有UGI结构域的催化上死亡(即,没有切口酶活性)的BE4(dBE4_noUGI)、或死亡的Cas9(dCas9)。仅使用编码碱基编辑器的mRNA转染细胞,或者除了该mRNA外也用gRNA191或gRNA12转染细胞。图16B是图表,其显示在如针对图16A所述那样转染的HBV感染的PHH细胞中的细胞外HBV DNA水平。

[0477] 图17A和17B将在HepG2-NTCP Lenti-HBV和HBV感染的PHH中的碱基编辑进行比较。图17A是图表,其显示在用BE4和UGI转染与用BE4而不用UGI转染的HepG2-NTCP Lenti-HBV中观察到的编辑效率。图17B是图表,其显示在用BE4和UGI转染与用BE4而不用UGI转染的HBV感染的PH中观察到的编辑效率。

[0478] 图18是将使用gRNA190在HBV-Lenti-HepG2与HBV感染的PHH中的碱基编辑、插入缺失率和颠换率(即,C到A或G)进行比较的图。

[0479] 图19显示示意性时间线,涉及使用HBV病毒感染的原代肝细胞共培养物(PHH)作为临床相关系统来评估本文所述碱基编辑试剂的抗病毒活性。在一些实施方案中,将用HBV感染的PHH共培养物用于本文所述的实验中以测定并评估碱基编辑器的抗病毒效力。简而言之,通过两次脂质转染历经两周进程将碱基编辑试剂(碱基编辑器mRNA和合成gRNA)转染至PHH共培养物中。第一次转染在注射HBV后3天时进行,以确保cccDNA在病毒转染的时间点完全形成。在所述实验的整个进程中监控细胞外参数(HBsAg、HBeAg和HBV DNA),并且在所述实验结束时监控细胞内参数(HBV DNA、病毒RNA和编辑)。HbsAg是指HBV的表面蛋白抗原。其检测表明个体中的HBV感染。HBeAg是指乙型肝炎e抗原,一种当该病毒被主动复制是在被感染血液中循环的HBV蛋白抗原。HBeAg的存在表明,该个体被感染并且能够将病毒传播给其他个体。

[0480] 图20显示柱状图,其呈现采用HBV感染的原代肝细胞共培养物(PHH)和gRNA12进行的14天实验的结果,gRNA12靶向HBV聚合酶与S基因序列的交叉点中的多核苷酸序列。使用抗病毒药物恩替卡韦(entecavir)作为对照,评估碱基编辑器(BE4和BE3-noUGI的效力)。如观察到的,BE4-noUGI碱基编辑器和gRNA12导致所测试的全部4种病毒标记物参数的下降,即,HBV DNA、HBsAg、HBeAg和pgRNA标记物产生的量或水平的下降。此外,与恩替卡韦相比,BE4-noUGI碱基编辑器和gRNA12在降低所测试的全部4种HBV参数方面显示了总体上杰出的性能。因此证明,与HBV抗病毒药物恩替卡韦相比,本文所述的碱基编辑方法在降低所测试的病毒(HBV)参数方面更为有效。

[0481] 图21显示柱状图,其呈现采用多种gRNA(多重gRNA)与BE4协同作用的结果。所评估的HBV参数包括pgRNA、HBsAg、HBeAg和HBV总DNA。结果指示了特定HBV参数的gRNA特异性降低,其中gRNA19表现出比所测试的其他gRNA改善的HBV抑制活性。此外,使用多重gRNA时,特别是使用gRNA19+gRNA190的组合时,观察到了HBV抑制的可测量的改善,其中使用gRNA190、gRNA12、gRNA40和gRNA52的组合时观察到了最优HBV抑制活性。

[0482] 图22显示柱状图,其呈现使用HBV感染的PHH培养系统和BE4碱基编辑器的碱基编辑结果。对从HBV感染的PHH培养物中纯化的总DNA并且对富集cccDNA的相同样品执行NGS测序。结果表明在富集cccDNA的样品中,碱基编辑显著增加(%碱基编辑),因此表明BE4碱基

编辑器和gRNA成功地对HBV cccDNA进行了碱基编辑。发现从HBV-PHH中纯化的总基因组DNA中的碱基编辑减少,表示经编辑的cccDNA失去了传播进入可复制型病毒颗粒内的能力。

[0483] 图23显示柱状图,其呈现采用多种gRNA(多重gRNA)与BE4和noUGI(BE4_noUGI)时的结果,如下文实施例10中所述。所评估的HBV参数包括HBsAg、HBeAg、pgRNA和HBV总DNA。结果指示HBV参数的显著的gRNA特异性抑制,其中gRNA12和gRNA19表现出增加的抑制活性。此外,发现gRNA19与BE_noUGI的HBV抑制活性与所测试的其他gRNA的组合同样有效。

[0484] 图24显示柱状图,其呈现使用HBV感染的PHH培养系统和BE4_noUGI碱基编辑器的碱基编辑结果。对从HBV感染的PHH培养物中纯化的总DNA并且对富集cccDNA的相同样品执行NGS测序。结果表明在富集cccDNA的样品中,碱基编辑显著增加(%碱基编辑),因此表明BE4_noUGI和gRNA成功地对HBV cccDNA进行了碱基编辑。发现从HBV-PHH中纯化的总基因组DNA中的稳健碱基编辑活性,表示经编辑的cccDNA失去了传播进入可复制型病毒颗粒内的能力。

[0485] 图25A至25D显示图表和柱状图,其与在长期(例如,25天)实验中使用不具有切口酶活性的碱基编辑器dBE4_noUGI(H840A)和HBV感染的PHH系统来评估碱基编辑器对于HBV病毒参数HBsAg(图25A)、细胞外HBV DNA(图25B)、HBeAg(图25C)和白蛋白(细胞活力/代谢速率)(图25D)的效果相关。该实验的结果显示,dBE4_noUGI(D10A_H840A)和gRNA12降低了HBV感染的PHH中的病毒参数。此外,尽管干扰素和碱基编辑组分(dBE4_noUGI+gRNA12)两者均降低HBV病毒参数,但发现干扰素治疗比使用本文所述的碱基编辑器和碱基编辑系统时的毒性更大。碱基编辑器dBE4_noUGI(H840A)包含氨基酸序列MSSETGPVAVDPTLRRRIEPHEFE
VFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAIT
EFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLY
VLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGGSSGGSSGSETPGTSESATPESSG
GSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRY
TRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKA
DLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLI
AQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS
DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPIL
EKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLAR
GNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGM
RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEE
NEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGF
ANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMAR
ENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIV
PQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR
QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTA
LIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIV
WDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE
KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELA
LPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIRE

QAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSKRTADGSEF ESPKKKRKVE

[0486] 图26A至26C呈现图表和柱状图,其显示长期(例如,25天)实验的结果,该实验涉及用基因型D和C的HBV感染的PHH培养物来评估本文所述的碱基编辑器(例如,dBE4_no UGI)和BE系统(例如,dBE4_no UGI+gRNA,例如gRNA12)对HBV的削弱或抑制,该评估通过评估HBV参数即HBsAg(图26A)、HBeAg(图26B)和细胞外HBV DNA(图26C)进行。实验表明,基因型C的HBV感染的细胞侵略性更强,因为该细胞在实验结束时的病毒负载更高。此外,用dBE4_no UGI和gRNA12转染HBV感染的PHH培养物导致病毒参数与关于基因型D的HBV和基因型C的HBV两者的对照相比有所减少。

[0487] 图27A和27B呈现柱状图,其表明用腺嘌呤碱基编辑器ABE7.10和HBV特异性gRNA(例如,gRNA94,其靶向HBV聚合酶活性位点)转染HBV感染的PHH培养物的结果。如图所示,相对于对照(未治疗的PHH和仅用ABE7.10治疗的PHH),ABE7.10+gRNA94在所测定的PHH培养物中显示显著的gRNA特异性HBV抑制和HBV标记物HBsAg、HBeAg、pgRNA和HBV总DNA的减少(图27A)。此外,ABE7.10+gRNA94在HBV感染的PHH中导致稳健的HBV cccDNA编辑(图27B)。在总HBV基因组DNA中观察到的碱基编辑的缺失表明,被编辑的HBV cccDNA不能传播到复制竞争性病毒颗粒中。

具体实施方式

[0488] 本发明的特征在于用于编辑HBV基因组的组合物和方法。例如,在一些实施方案中,本文设想的组合物可包括碱基编辑器和靶向HBV基因中特定核苷酸的向导核酸。在一些实施方案中,该编辑将提前终止密码子引入病毒蛋白之一的编码序列中。在另一实施方案中,该编辑将一种或多种功能性置换引入一种或多种病毒蛋白的编码序列中。

[0489] HBV基因组包含3.2kb的部分双链DNA和编码七种蛋白质的开放阅读框(ORF)。参照图1,开放阅读框(ORF)P编码病毒聚合酶。ORF C/PreC编码衣壳蛋白。ORF PreS1、ORF PreS2和ORF S分别编码大(L)、中(M)和小(S)表面蛋白。ORF X编码分泌型X蛋白。

[0490] 部分双链HBV基因组被宿主抑制转化为共价闭合环状DNA(cccDNA)。cccDNA被宿主RNA聚合酶转录以产生包括前基因组RNA(pgRNA)在内的病毒mRNA。pgRNA被HBV聚合酶逆转录为基因组HBV DNA,其能够被转化为cccDNA,封装在病毒体中,或整合到宿主细胞基因组中(图2)。cccDNA是HBV生命周期中的一种关键组分,是对慢性HBV干扰负责的一种稳定的分子。对HBV基因组的编辑可以中断cccDNA的形成,从而降低病毒的致病性。

[0491] 存在十种不同的HBV基因型(A-J)。“基因型”的特征在于与任何其他基因组的<92%序列一致性,而亚基因型的特征在于<96至92%序列一致性。基因型D的HBV在美国最为流行。HBV基因型D的研究模型可用,包括病毒原液(例如,基因型D,亚基因型ayw(Imquest))和小鼠模型(例如,人源化小鼠模型(Phoenixbio))。因此,在一些实施方案中,提供靶向HBV ORF进行编辑的方法和组合物。这些组合物可包含具有Cas9或其他核酸可编程DNA结合蛋白结构域和腺苷或胞嘧啶脱氨酶结构域的核碱基编辑器在一些实施方案中,碱基编辑器将一个或多个改变引入HBV ORF中。在一些实施方案中,该改变导致HBV蛋白的保守部分中的突变。在特定实施方案中,该改变引入一个或多个终止密码子。在整篇说明书中,导致蛋白质的提前终止的终止密码子的引入以氨基酸符号、氨基酸位置和术语STOP表示(例如,

R87STOP表示编码位于氨基酸位置87处的精氨酸被替换为终止密码子的密码子)。有利地,本发明的方法不会引入HBV基因组中的双链断裂。

[0492] 本发明提供使用碱基编辑来治疗慢性HBV的策略(图3A)。本文描述了筛选方法,该方法用于鉴定将终止密码子或功能性突变引入HBV基因中的向导RNA,或用于鉴定在HBV基因组的超保守区域内产生无碱基位点的gRNA(图3B)。使用本文所述的方法和组合物将终止密码子引入病毒基因中可在不产生双链断裂的情况下实施,从而消除或降低在将HBV整合到宿主基因组内之后切割宿主遗传物质的风险。此外,该组合物采用作为天然HBV抗病毒限制因子的脱氨酶。例如,用干扰素α或淋巴毒素β受体(LTBR)诱导APOBEC胞苷脱氨酶促使无碱基位点形成和cccDNA降解(图3A)。此外,使用不具有尿嘧啶糖基化酶抑制剂结构域的碱基编辑器可将细胞尿嘧啶糖基化酶靶向至cccDNA并促使其降解。

[0493] 提供的另一种筛选方式鉴定可用来在cccDNA中产生无碱基位点的保守gRNA。参照图3C,鉴定了7种向导RNA,当使用慢病毒来引入碱基编辑器和gRNA(Lenti-HBV)时,这些向导RNA具有大于20%的编辑效率。靶向保守区域的gRNA显示在图3D中。几种gRNA具有至少45%的编辑效率(图3E和3F)。

[0494] 在一些方面,提供了使用包含胞苷脱氨酶或腺苷脱氨酶结构域的碱基编辑器来编辑HBV cccDNA的方法和组合物。在一种实施方案中,碱基编辑器包含APOBEC胞苷脱氨酶结构域、Cas9结构域,并且任选地包含一个或多个尿嘧啶糖基化酶抑制剂(UGI)结构域(图4A、4B)。

[0495] 核碱基编辑器

[0496] 本文公开了一种用于编辑、修饰或改变多核苷酸的靶标核苷酸序列的碱基编辑器或核碱基编辑器。本文描述了核碱基编辑器或碱基编辑器,其包含多核苷酸可编程核苷酸结合结构域和核碱基编辑结构域(例如,腺苷脱氨酶、胞苷脱氨酶)。当与向导多核苷酸(例如,gRNA)结合时,多核苷酸可编程核苷酸结合结构域可以特异性地结合至靶标多核苷酸序列(即,经由所结合的向导核酸的碱基与靶标多核苷酸序列的碱基之间的互补碱基配对),并因此将碱基编辑器定位在希望进行编辑的靶标核酸序列处。在一些实施方案中,靶标多核苷酸序列包含单链DNA或双链DNA。在一些实施方案中,靶标多核苷酸序列包含RNA。在一些实施方案中,靶标多核苷酸序列包含DNA-RNA杂交物。

[0497] 多核苷酸可编程核苷酸结合结构域

[0498] 应知悉,多核苷酸可编程核苷酸结合结构域也可以包括结合RNA的核酸可编程蛋白质。例如,多核苷酸可编程核苷酸结合结构域可以与核酸缔合,该核酸将该多核苷酸可编程核苷酸结合结构域引导至RNA。其他核酸可编程DNA结合蛋白也处于本公开的范畴内,但它们不在本公开中具体地列举。

[0499] 碱基编辑器的多核苷酸可编程核苷酸结合结构域自身可以包含一个或多个结构域。例如,多核苷酸可编程核苷酸结合结构域可以包含一个或多个核酸酶结构域。在一些实施方案中,多核苷酸可编程核苷酸结合结构域的核酸酶结构域可以包含核酸内切酶或核酸外切酶。本文中,术语“核酸外切酶”是指能够从游离末端消化核酸(例如,RNA或DNA)的蛋白质或多肽,而术语“核酸内切酶”是指能够催化(例如,裂解)核酸(例如,DNA或RNA)中的内部区域的蛋白质或多肽。在一些实施方案中,核酸内切酶可以裂解双链核酸的一条链。在一些实施方案中,核酸内切酶可以裂解双链核酸分子的两条链。在一些实施方案中,多核苷酸可

编程核苷酸结合结构域可以是脱氧核糖核酸酶。在一些实施方案中,多核苷酸可编程核苷酸结合结构域可以是核糖核酸酶。

[0500] 在一些实施方案中,多核苷酸可编程核苷酸结合结构域的核酸酶结构域可以切割靶标多核苷酸的零、一或两条链。在一些情况下,多核苷酸可编程核苷酸结合结构域可以包含切口酶结构域。本文中,术语“切口酶”是指包含核酸酶结构域的多核苷酸可编程核苷酸结合结构域,其能够仅裂解双螺旋核酸分子(例如,DNA)中的两条链中的一条链。在一些实施方案中,切口酶可以通过经一个或多个突变引入活性多核苷酸可编程核苷酸结合结构域中而衍生自完全催化活性(例如,天然)形式的多核苷酸可编程核苷酸结合结构域。例如,如果多核苷酸可编程核苷酸结合结构域包含衍生自Cas9的切口酶结构域,则衍生自Cas9的切口酶结构域可以包括D10A突变和位于位置840处的组氨酸。在这种情况下,残基H840保持催化活性并且因此可以裂解核酸双螺旋的一条链。在另一实施例中,衍生自Cas9的切口酶结构域可以包含H840A突变,而位于位置10处的氨基酸保持为D。在一些实施方案中,切口酶可以通过去除切口酶活性所不需要的全部或部分核酸酶结构域而衍生自完全催化活性(例如,天然)形式的多核苷酸可编程核苷酸结合结构域。例如,如果多核苷酸可编程核苷酸结合结构域包含衍生自Cas9的切口酶结构域,则衍生自Cas9的切口酶结构域可以包含RuvC结构域或HNH结构域的全部或部分缺失。

[0501] 示例性的催化活性Cas9的氨基酸序列如下:

[0502] MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD。

[0503] 因此,包含具有切口酶结构域的多核苷酸可编程核苷酸结合结构域的碱基编辑器能够在具体的(例如,通过所结合的向导核酸的互补序列确定的)多核苷酸靶标序列处产生单链DNA断裂(切口)。在一些实施方案中,被包含切口酶结构域(例如,衍生自Cas9的切口酶

结构域)的碱基编辑器裂解的核酸双螺旋靶标多核苷酸序列的链是没有通过该碱基编辑器编辑的链(即,被碱基编辑器裂解的链与包含待编辑甲基的链相对)。在其他实施方案中,包含切口酶结构域(例如,衍生自Cas9的切口酶结构域)的碱基编辑器可以裂解被靶向以进行编辑的DNA分子的链。在这种情况下,非靶向链不被裂解。

[0504] 本文中还提供包含碱基编辑器,其包含催化上死亡(即,不能够裂解靶标多核苷酸序列)的多核苷酸可编程核苷酸结合结构域。本文中,术语"催化上死亡的"和"核酸酶死亡的"可互换使用,并且是指一种多核苷酸可编程核苷酸结合结构域,其具有一个或多个突变和/或缺失,导致其失去了裂解核酸的链的能力。在一些实施方案中,作为一个或多个核酸酶结构域中特异性点突变的结果,催化上死亡的多核苷酸可编程核苷酸结合结构域碱基编辑器可能缺乏核酸酶活性。例如,在碱基编辑器包含Cas9结构域的情况下,该Cas9可以包含D10A突变和H840A突变两者。此类突变令两个核酸酶结构域失去活性,从而导致核酸酶活性的丧失。在其他实施方案中,催化上死亡的多核苷酸可编程核苷酸结合结构域可以包含全部或部分的催化结构域(例如RuvC1结构域和/或HNH结构域)的一个或多个缺失。在进一步的实施方案中,催化上死亡的多核苷酸可编程核苷酸结合结构域包含点突变(例如,D10A或H840A)以及全部或部分的核酸酶结构域的缺失。

[0505] 本文还设想了能够从多核苷酸可编程核苷酸结合结构域的先前功能性版本产生催化上死亡的多核苷酸可编程核苷酸结合结构域的突变。例如,在催化上死亡的Cas9("dCas9")的情况下,提供了具有除D10A和H840A外的突变的变体,其得到无核酸酶活性的Cas9。举例而言,此类突变包括位于D10和H840处的其他氨基酸置换,或Cas9的核酸酶结构域内的其他置换(例如,HNH核酸酶亚结构域和/或RuvC1亚结构域中的置换)。基于本公开和本领域的知识,其他合适的无核酸酶活性的dCas9结构域对于本领域技术人员是显而易见的,并且处于本公开的范畴内。此类其他示例性的合适的无核酸酶活性的Cas9结构域包括但不限于,D10A/H840A、D10A/D839A/H840A和D10A/D839A/H840A/N863A突变结构域(参见例如,Prashant et al.,CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering.Nature Biotechnology.2013;31(9):833-838,其整体内容通过引用并入本文)。

[0506] 可并入碱基编辑器中的多核苷酸可编程核苷酸结合结构域的非限制性示例包括源自CRISPR蛋白的结构域、限制性核酸酶、大范围核酸酶、TAL核酸酶(TALEN)和锌指核酸酶(ZFN)。在一些情况下,碱基编辑器包含具有天然或修饰的蛋白质或其部分的多核苷酸可编程核苷酸结合结构域,在CRISPR(即,成簇规律间隔短回文重复序列)介导的核酸修饰期间,其能够经由所结合的向导核酸结合至核酸序列。本文中,此类蛋白质称为"CRISPR蛋白"。据此,本文公开了包含具有全部或部分的CRISPR蛋白的多核苷酸可编程核苷酸结合结构域的碱基编辑器(即,包含全部或部分的CRISPR蛋白作为结构域的碱基编辑器,该结构域也称为碱基编辑器的"衍生自CRISPR蛋白的结构域")。与CRISPR蛋白的野生型或天然版本相比,被并入碱基编辑器中的衍生自CRISPR蛋白的结构域可以经修饰。例如,如下文所述,衍生自CRISPR蛋白的结构域可以包含性对于CRISPR蛋白的野生型或天然版本的一个或多个突变、插入、缺失、重排和/或重组。

[0507] CRISPR是适应性免疫系统,其提供针对可动遗传因子(病毒、转座因子和接合质粒)的保护。CRISPR簇含有间隔序列、与祖先可动因子互补的序列和侵入靶标的核酸。

CRISPR簇被转录并加工为CRISPR RNA(crRNA)。在II型CRISPR系统中,正确加工pre-crRNA需要反式编码的小RNA(tracrRNA)、内源核糖核酸酶3(rnc)和Cas9蛋白。tracrRNA在核糖核酸酶3辅助的pre-crRNA加工中充当向导。随后,Cas9/crRNA/tracrRNA核酸内切溶解地裂解与间隔序列互补的线性或环状dsDNA靶标。不与crRNA互补的靶标链首先被核酸内切溶解地切割,然后被3'-5'核酸外切溶解地修剪。本质上中,DNA结合和裂解通常需要蛋白质和两种RNA。但是,单向导RNA("sgRNA"或简称为"gNRA")可以工程化改造,以将crRNA和tracrRNA两者的各个方面合并到单个RNA物种中。参见例如,Jinek M.,Chylinski K.,Fonfara I.,Hauer M.,Doudna J.A.,Charpentier E.Science 337:816-821(2012),其整体内容通过引用并入本文。Cas9识别CRISPR重复序列中的短基序(PAM或前间隔序列邻近基序)以帮助区分自我与非自我。

[0508] 在一些实施方案中,本文所述的方法可以利用经工程化的Cas蛋白。向导RNA(gRNA)是短合成RNA,由Cas结合所必需的支架序列和使用者定义的约20个核苷酸的定义待修饰的基因组(或多核苷酸,例如DNA或RNA)靶标的间隔序列构成。因此,技术人员可以通过改变gRNA中存在的靶标序列来改变Cas蛋白的基因组或多核苷酸靶标。Cas蛋白的特异性部分地由gRNA靶向序列对于基因组多核苷酸靶标序列与基因组剩余部分的特异性如何而决定。

[0509] 在一些实施方案中,gRNA支架序列如下:GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUUUU。

[0510] 在一种实施方案中,RNA支架包含茎环。在一种实施方案中,RNA支架包含核酸序列:

[0511] GUUUUUGUACUCUCAAGAUUUAAGUAACUGUACAACGAAACUUACACAGUUACUUAAAUCUUGCAGAAGCUACAAAGAUAAGGCUUCAUGCCGAAAUCAACACCCUGUCAUUUUAUGGCAGGGUG。在一种实施方案中,RNA支架包含核酸序列:

[0512] GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU。

[0513] 在一种实施方案中,酿脓链球菌(S.pyrogenes)sgRNA支架多核苷酸序列如下:

[0514] GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC。

[0515] 在一种实施方案中,金黄色葡萄球菌(S.aureus)sgRNA支架多核苷酸序列如下:

[0516] GUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCUACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGA。

[0517] 在一种实施方案中,BhCas12b sgRNA支架具有下述多核苷酸序列:GUUCUGTCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUGCAGGGUGUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUACGAGGCAUUAGCAC。

[0518] 在一种实施方案中,BvCas12b sgRNA支架具有下述多核苷酸序列:GACCUAUAGGGUCAAUGAAUCUGUGCGUGUGCCAUAAGUAAUUAAAAAUUACCCACCACACAGGAGCACCUGAAAACAGGUGCUUGGCAC。

[0519] 在一些实施方案中,被并入碱基编辑器中的衍生自CRISPR蛋白的结构域是核酸内切酶(例如,脱氧核糖核酸酶或核糖核酸酶),当与所结合的向导核酸协同作用时,该核酸内切酶能够结合靶标多核苷酸。在一些实施方案中,被并入碱基编辑器中的衍生自CRISPR蛋

白的结构域是切口酶,当与所结合的向导核酸协同作用时,该切口酶能够结合靶标多核苷酸。在一些实施方案中,被并入碱基编辑器中的衍生自CRISPR蛋白的结构域是催化上死亡的结构域,当与所结合的向导核酸协同作用时,该催化上死亡的结构域能够结合靶标多核苷酸。在一些实施方案中,本衍生自碱基编辑器的CRISPR蛋白结合的靶标多核苷酸是DNA。在一些实施方案中,本衍生自碱基编辑器的CRISPR蛋白结合的靶标多核苷酸是RNA。

[0520] 本文中可使用的Cas蛋白包括1类和2类。Cas蛋白的非限制性示例包括Cas1、Cas1B、Cas2、Cas3、Cas4、Cas5、Cas5d、Cas5t、Cas5h、Cas5a、Cas6、Cas7、Cas8、Cas9(也称为Csn1或Csx12)、Cas10、Csy1、Csy2、Csy3、Csy4、Cse1、Cse2、Cse3、Cse4、Cse5e、Csc1、Csc2、Csa5、Csn1、Csn2、Csm1、Csm2、Csm3、Csm4、Csm5、Csm6、Cmr1、Cmr3、Cmr4、Cmr5、Cmr6、Csb1、Csb2、Csb3、Csx17、Csx14、Csx10、Csx16、CsaX、Csx3、Csx1、Csx1S、Csf1、Csf2、Cs0、Csf4、Csd1、Csd2、Cst1、Cst2、Csh1、Csh2、Csa1、Csa2、Csa3、Csa4、Csa5、Cas12a/Cpf1、Cas12b/C2c1、Cas12c/C2c3、Cas12d/CasY、Cas12e/CasX、Cas12g、Cas12h和Cas12i、CARF、DinG、其同源物、或其经修饰的版本。未经修饰的CRISPR酶可能具有DNA裂解活性,诸如Cas9,其具有两个功能性核酸内切酶结构域:RuvC和HNH。CRISPR酶可以引导一条或两条链在靶标序列处(诸如靶标序列内和/或靶标序列的补体内)的裂解。例如,CRISPR酶可以引导从靶标序列的第一个或最后一个核苷酸起约1、2、3、4、5、6、7、8、9、10、15、20、25、50、100、200、500个或更多个碱基对内的一条或两条链。

[0521] 可以使用一种载体,该载体编码CRISPR酶,该酶被关于野生型酶突变以使得经突变的CRISPR酶缺乏裂解含有靶标序列的靶标多核苷酸的一条或两条链的能力。Cas9可以指与野生型示例性Cas9多肽(例如,来自酿脓链球菌的Cas9)具有至少或至少约50%、60%、70%、80%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%序列同一性和/或序列同源性的多肽。Cas9可以指与野生型示例性Cas9多肽(例如,来自酿脓链球菌)具有至多或至多约50%、60%、70%、80%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%序列同一性和/或序列同源性的多肽。Cas9可以指野生型或经修饰形式的Cas9蛋白,其可以包含氨基酸变化诸如缺失、插入、置换、变异、突变、融合、嵌合或其任意组合。

[0522] 在一些实施方案中,碱基编辑器的衍生自CRISPR蛋白的结构域可以包括来自以下的Cas9的全部或部分:溃疡棒状杆菌(Corynebacteriumulcerans)(NCBI Refs:NC_015683.1,NC_017317.1);白喉棒状杆菌(Corynebacteriumdiphtheria)(NCBI Refs:NC_016782.1,NC_016786.1);螺原体(Spiroplasma syrphidicola)(NCBI Ref:NC_021284.1);中间普雷沃菌(Prevotella intermedia)(NCBI Ref:NC_017861.1);中国台湾螺原体(Spiroplasma taiwanense,China)(NCBI Ref:NC_021846.1);海豚链球菌(Streptococcus iniae)(NCBI Ref:NC_021314.1);拟杆菌(Belliella baltica)(NCBI Ref:NC_018010.1);扭曲冷弯曲菌(Psychroflexus torquis)(NCBI Ref:NC_018721.1);嗜热链球菌(Streptococcus thermophilus)(NCBI Ref:YP_820832.1);无害利斯特菌(Listeria innocua)(NCBI Ref:NP_472073.1);空肠弯曲杆菌(Campylobacter jejuni)(NCBI Ref:YP_002344900.1);脑膜炎奈瑟氏菌(Neisseria meningitidis)(NCBI Ref:YP_002342100.1);酿脓链球菌(Streptococcus pyogenes)或金黄色葡萄球菌(Staphylococcus aureus)。

[0523] 核碱基编辑器的Cas9结构域

[0524] Cas9核酸酶序列和结构是本领域技术人员所周知的(参见例如,“Complete genome sequence of an M1 strain of Streptococcus pyogenes.”Ferretti et al., J.J.,McShan W.M.,Ajdic D.J.,Savic D.J.,Savic G.,Lyon K.,Primeaux C,Sezate S., Suvorov A.N.,Kenton S.,Lai H.S.,Lin S.P.,Qian Y.,Jia H.G.,Najar F.Z.,Ren Q., Zhu H.,Song L.,White J.,Yuan X.,Clifton S.W.,Roe B.A.,McLaughlin R.E., Proc.Natl.Acad.Sci.U.S.A.98:4658-4663(2001);“CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.”Deltcheva E.,Chylinski K.,Sharma C.M.,Gonzales K.,Chao Y.,Pirzada Z.A.,Eckert M.R.,Vogel J.,Charpentier E., Nature 471:602-607(2011);和“A programmabledual-RNA-guided DNA endonuclease in adaptive bacterial immunity.”Jinek M.,Chylinski K.,Fonfara I.,Hauer M.,Doudna J.A.,Charpentier E.Science 337:816-821(2012),其各自的整体内容通过引用并入本文)。Cas9直接同源物已经在多种物种中描述,包括但不限于酿脓链球菌和嗜热链球菌。基于本公开,其他合适的Cas9核酸酶和序列对于本领域技术人员将是显而易见的,并且此类Cas9核酸酶和序列包括来自生物体的Cas9序列和以下文献中公开的基因座:Chylinski, Rhun,and Charpentier,“The tracrRNA and Cas9 families of type IICRISPR-Cas immunity systems”(2013)RNA Biology 10:5,726-737;其整体内容通过引用并入本文。

[0525] 在一些方面,核酸可编程DNA结合蛋白(napDNAbp)是Cas9结构域。本文中提供了非限制性的示例性Cas9结构域。Cas9结构域可以是核酸酶活性Cas9结构域、无核酸酶活性的Cas9结构域或Cas9切口酶。在一些实施方案中,Cas9结构域是核酸酶活性结构域。例如,Cas9结构域可以是切割双螺旋核酸的两条链(例如,双螺旋DNA分子的两条链)的Cas9结构域。在一些实施方案中,Cas9结构域包含本文所详述的任何一个氨基酸序列。在一些实施方案中,Cas9结构域包含氨基酸序列,所述氨基酸序列与本文详述的任何一个氨基酸序列为至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相同。在一些实施方案中,Cas9结构域包含氨基酸序列,与本文详述的任何一个氨基酸序列相比,该氨基酸序列具有1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、21、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50或更多个突变。在一些实施方案中,Cas9结构域包含氨基酸序列,与本文详述的任何一个氨基酸序列相比,该氨基酸序列具有至少10个、至少15个、至少20个、至少30个、至少40个、至少50个、至少60个、至少70个、至少80个、至少90个、至少100个、至少150个、至少200个、至少250个、至少300个、至少350个、至少400个、至少500个、至少600个、至少700个、至少800个、至少900个、至少1000个、至少1100个或至少1200个相同的连续氨基酸残基。

[0526] 在一些实施方案中,提供了包含Cas9的片段的蛋白质。例如,在一些实施方案中,蛋白质包含一个或两个Cas9结构域:(1)Cas9的gRNA结合结构域;或(2)Cas9的DNA切割结构域。在一些实施方案中,包含Cas9或其片段的蛋白质称为“Cas9变体”。Cas9变体享有与Cas9或其片段的同源性。例如,Cas9变体与野生型Cas9至少约70%相同,至少约80%相同,至少约90%相同,至少约95%相同,至少约96%相同,至少约97%相同,至少约98%相同,至少约99%相同,至少约99.5%相同,或至少约99.9%相同。在一些实施方案中,与野生型Cas9相

比,Cas9变体可以具有1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、21、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个或更多氨基酸变化。在一些实施方案中,Cas9变体包含Cas9的片段(例如,gRNA结合结构域或DNA切割结构域),使得该片段与野生型Cas9的相应片段至少约70%相同,至少约80%相同,至少约90%相同,至少约95%相同,至少约96%相同,至少约97%相同,至少约98%相同,至少约99%相同,至少约99.5%相同,或至少约99.9%相同。在一些实施方案中,该片段是相应野生型Cas9的氨基酸长度的至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%。在一些实施方案中,该片段为至少100个氨基酸的长度。一些实施方式中,该片段为至少100、150、200、250、300、350、400、450、500、550、600、650、700、750、800、850、900、950、1000、1050、1100、1150、1200、1250个或至少1300个氨基酸的长度。

[0527] 在一些实施方案中,如本文所提供的Cas9融合蛋白包含Cas9蛋白的全长度氨基酸序列,例如,本文所提供的Cas9序列之一。然而,在其他实施方案中,如本文所提供的融合蛋白不包含全长度Cas9序列,而仅包含其一个或多个片段。合适的Cas9结构域和Cas9片段的示例性氨基酸序列在本文中提供,并且Cas9结构域和片段的其他合适的序列对于本领域技术人员而言将会是显而易见的。

[0528] Cas9蛋白可以与将该Cas9蛋白引导至具体DNA序列的向导RNA缔合,该DNA序列与向导RNA互补。在一些实施方案中,多核苷酸可编程核苷酸结合结构域是Cas9结构域,例如,核酸酶活性Cas9、Cas9切口酶(nCas9)或无核酸酶活性的Cas9(dCas9)。核酸可编程DNA结合蛋白的示例包括但不限于,Cas9(例如,dCas9和nCas9)、CasX、CasY、Cpf1、Cas12b/C2C1和Cas12c/C2C3。

[0529] 在一些实施方案中,野生型Cas9对应于来自酿脓链球菌(Streptococcus pyogenes)的Cas9(NCBI参考序列:NC_017053.1,核苷酸和氨基酸序列如下)。

[0530] ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGATCACTGATGATTATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTATCAAAAAAAATCTTATAGGGGCTCTTTTATTTGGCAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGTCTTTTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAAGTTGCTTATCATGAGAAATATCCAACTATCTATCATCTGCGAAAAAAATTGGCAGATTCTACTGATAAAGCGGATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAGATTTAAATCCTGATAATAGTGATGTGGACAAACTATTTATCCAGTTGGTACAAATCTACAATCAATTATTTGAAGAAAACCCTATTAACGCAAGTAGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGCTCAGCTCCCCGGTGAGAAGAGAAATGGCTTGTTTGGGAATCTCATTGCTTTGTCATTGGGATTGACCCCTAATTTTAAATCAAATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTAGATAATTTATTGGCGCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGATGCTATTTTACTTTCAGATATCCTAAGAGTAAATAGTGAAATAACTAAGGCTCCCCTATCAGCTTCAATGATTAAGCGCTACGATGAACATCATCAAGACTTGACTCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTATAAAGAAATCTTTTTTGATCAATCAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAAATTTATCAAACCAATTTTAGAAAAAA

TGGATGGTACTGAGGAATTATTGGTGAAACTAAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACAACGGC
TCTATTCCCCATCAAATTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTTAAA
AGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTCCATTGGCGCGTGGCAATA
GTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCATGGAATTTTGAAGAAGTTGTCGATAAAGGT
GCTTCAGCTCAATCATTTATTGAACGCATGACAAACTTTGATAAAAATCTTCCAAATGAAAAAGTACTACCAAAACA
TAGTTTGCTTTATGAGTATTTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTGAGGGAATGCGAAAAC
CAGCATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAACAAATCGAAAAGTAACCGTTAAG
CAATTAAAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGAAATTTCAGGAGTTGAAGATAGATTTAA
TGCTTCATTAGGCGCCTACCATGATTTGCTAAAAATTATTAAAGATAAAGATTTTTTGGATAATGAAGAAAATGAAG
ATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAAGATAGGGGGATGATTGAGGAAAGACTTAAAACATAT
GCTCACCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTTTGTCTCGAAA
ATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTAGATTTTTTGAAATCAGATGGTTTTGCCAATC
GCAATTTTATGCAGCTGATCCATGATGATAGTTTGACATTTAAAGAAGATATTCAAAAAGCACAGGTGTCTGGACAA
GGCCATAGTTTACATGAACAGATTGCTAACTTAGCTGGCAGTCCTGCTATTAAAAAAGGTATTTTACAGACTGTAAA
AATTGTTGATGAACTGGTCAAAGTAATGGGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACGTGAAAATCAGA
CAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAATCGAAGAAGGTATCAAAGAATTAGGAAGTCAG
ATTCTTAAAGAGCATCCTGTTGAAAATACTCAATTGCAAAATGAAAAGCTCTATCTCTATTATCTACAAAATGGAAG
AGACATGTATGTGGACCAAGAATTAGATATTAATCGTTTAAGTGATTATGATGTCGATCACATTGTTCCACAAAGTT
TCATTAAAGACGATTCAATAGACAATAAGGTACTAACGCGTTCTGATAAAAATCGTGGTAAATCGGATAACGTTCCA
AGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAACTTCTAAACGCCAAGTTAATCACTCAACGTAAGTT
TGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTGAACTTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTG
AAACTCGCCAAATCACTAAGCATGTGGCACAAATTTTGGATAGTCGCATGAATACTAAATACGATGAAAATGATAAA
CTTATTCGAGAGGTTAAAGTGATTACCTTAAAATCTAAATTAGTTTCTGACTTCCGAAAAGATTTCCAATTCTATAA
AGTACGTGAGATTAACAATTACCATCATGCCCATGATGCGTATCTAAATGCCGTCGTTGGAACTGCTTTGATTAAGA
AATATCCAAAACTTGAATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTCGTAAAATGATTGCTAAGTCT
GAGCAAGAAATAGGCAAAGCAACCGCAAAATATTTCTTTTACTCTAATATCATGAACTTCTTCAAAACAGAAATTAC
ACTTGCAAATGGAGAGATTCGCAAACGCCCTCTAATCGAAACTAATGGGGAAACTGGAGAAATTGTCTGGGATAAAG
GGCGAGATTTTGCCACAGTGCGCAAAGTATTGTCCATGCCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAGACA
GGCGGATTCTCCAAGGAGTCAATTTTACCAAAAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGATCC
AAAAAAATATGGTGGTTTTGATAGTCCAACGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGAAAT
CGAAGAAGTTAAAATCCGTTAAAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAAATCCGATT
GACTTTTTAGAAGCTAAAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACCTAAATATAGTCTTTTTGA
GTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAATTACAAAAAGGAAATGAGCTGGCTCTGCCAAGCA
AATATGTGAATTTTTTATATTTAGCTAGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAA
TTGTTTGTGGAGCAGCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCGTGTTATTTT
AGCAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAGACAAACCAATACGTGAACAAGCAGAAA
ATATTATTCATTTATTTACGTTGACGAATCTTGGAGCTCCCGCTGCTTTTAAATATTTTGATACAACAATTGATCGT
AAACGATATACGTCTACAAAAGAAGTTTTAGATGCCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACACG
CATTGATTTGAGTCAGCTAGGAGGTGACTGA

MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETAEATRLKRTARR
RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRK
KLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQIYNQLFEENPINASRVDAKA
ILSARLSKSRRLENLIAQLPGEKRNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLA
[0531] QIGDQYADLFLAAKNLSDAILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQS
FIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVT
VKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRG
MIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGHSLHEQIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQTTQ
KGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHI
VPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL
DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNY
[0532] HHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA
RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVK
KDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ
HKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTI
DRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0533] (单下划线:HNH结构域;双下划线:RuvC结构域)

[0534] 在一些实施方案中,野生型Cas9对应于或包含以下核苷酸和/或氨基酸序列:

[0535] ATGGATAAAAAGTATTCTATTGGTTTAGACATCGGCACTAATTCCGTTGGATGGGCTGTCATAACCGAT
GAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACCGTCATTCGATTAAAAAGAATCTTATCGG
TGCCCTCCTATTCGATAGTGGCGAAACGGCAGAGGCGACTCGCCTGAAACGAACCGCTCGGAGAAGGTATACACGTC
GCAAGAACCGAATATGTTACTTACAAGAAATTTTTAGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTTCACCGT
TTGGAAGAGTCCTTCCTTGTCGAAGAGGACAAGAAACATGAACGGCACCCCATCTTTGGAAACATAGTAGATGAGGT
GGCATATCATGAAAAGTACCCAACGATTTATCACCTCAGAAAAAAGCTAGTTGACTCAACTGATAAAGCGGACCTGA
GGTTAATCTACTTGGCTCTTGCCCATATGATAAAGTTCCGTGGGCACTTTCTCATTGAGGGTGATCTAAATCCGGAC
AACTCGGATGTCGACAAACTGTTCATCCAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGAACCCTATAAATGC
AAGTGGCGTGGATGCGAAGGCTATTCTTAGCGCCCGCCTCTCTAAATCCCGACGGCTAGAAAACCTGATCGCACAAT
TACCCGGAGAGAAGAAAAATGGGTTGTTCGGTAACCTTATAGCGCTCTCACTAGGCCTGACACCAAATTTTAAGTCG
AACTTCGACTTAGCTGAAGATGCCAAATTGCAGCTTAGTAAGGACACGTACGATGACGATCTCGACAATCTACTGGC
ACAAATTGGAGATCAGTATGCGGACTTATTTTTGGCTGCCAAAAACCTTAGCGATGCAATCCTCCTATCTGACATAC
TGAGAGTTAATACTGAGATTACCAAGGCGCCGTTATCCGCTTCAATGATCAAAAGGTACGATGAACATCACCAAGAC
TTGACACTTCTCAAGGCCCTAGTCCGTCAGCAACTGCCTGAGAAATATAAGGAAATATTCTTTGATCAGTCGAAAAA
CGGGTACGCAGGTTATATTGACGGCGGAGCGAGTCAAGAGGAATTCTACAAGTTTATCAAACCCATATTAGAGAAGA

TGGATGGGACGGAAGAGTTGCTTGTAAAACTCAATCGCGAAGATCTACTGCGAAAGCAGCGGACTTTCGACAACGGT
AGCATTCCACATCAAATCCACTTAGGCGAATTGCATGCTATACTTAGAAGGCAGGAGGATTTTTATCCGTTCCTCAA
AGACAATCGTGAAAAGATTGAGAAAATCCTAACCTTTCGCATACCTTACTATGTGGGACCCCTGGCCCGAGGGAACT
CTCGGTTCGCATGGATGACAAGAAAGTCCGAAGAAACGATTACTCCATGGAATTTTGAGGAAGTTGTCGATAAAGGT
GCGTCAGCTCAATCGTTCATCGAGAGGATGACCAACTTTGACAAGAATTTACCGAACGAAAAAGTATTGCCTAAGCA
CAGTTTACTTTACGAGTATTTCACAGTGTACAATGAACTCACGAAAGTTAAGTATGTCACTGAGGGCATGCGTAAAC
CCGCCTTTCTAAGCGGAGAACAGAAGAAAGCAATAGTAGATCTGTTATTCAAGACCAACCGCAAAGTGACAGTTAAG
CAATTGAAAGAGGACTACTTTAAGAAAATTGAATGCTTCGATTCTGTCGAGATCTCCGGGGTAGAAGATCGATTTAA
TGCGTCACTTGGTACGTATCATGACCTCCTAAAGATAATTAAAGATAAGGACTTCCTGGATAACGAAGAGAATGAAG
ATATCTTAGAAGATATAGTGTTGACTCTTACCCTCTTTGAAGATCGGGAAATGATTGAGGAAAGACTAAAAACATAC
GCTCACCTGTTCGACGATAAGGTTATGAAACAGTTAAAGAGGCGTCGCTATACGGGCTGGGGACGATTGTCGCGGAA
ACTTATCAACGGGATAAGAGACAAGCAAAGTGGTAAAACTATTCTCGATTTTCTAAAGAGCGACGGCTTCGCCAATA
GGAACTTTATGCAGCTGATCCATGATGACTCTTTAACCTTCAAAGAGGATATACAAAAGGCACAGGTTTCCGGACAA
GGGGACTCATTGCACGAACATATTGCGAATCTTGCTGGTTCGCCAGCCATCAAAAAGGGCATACTCCAGACAGTCAA
AGTAGTGGATGAGCTAGTTAAGGTCATGGGACGTCACAAACCGGAAAACATTGTAATCGAGATGGCACGCGAAAATC
AAACGACTCAGAAGGGGCAAAAAAACAGTCGAGAGCGGATGAAGAGAATAGAAGAGGGTATTAAAGAACTGGGCAGC
CAGATCTTAAAGGAGCATCCTGTGGAAAATACCCAATTGCAGAACGAGAAACTTTACCTCTATTACCTACAAAATGG
AAGGGACATGTATGTTGATCAGGAACTGGACATAAACCGTTTATCTGATTACGACGTCGATCACATTGTACCCCAAT
CCTTTTTGAAGGACGATTCAATCGACAATAAAGTGCTTACACGCTCGGATAAGAACCGAGGGAAAAGTGACAATGTT
CCAAGCGAGGAAGTCGTAAAGAAAATGAAGAACTATTGGCGGCAGCTCCTAAATGCGAAACTGATAACGCAAAGAAA
GTTCGATAACTTAACTAAAGCTGAGAGGGGTGGCTTGTCTGAACTTGACAAGGCCGGATTTATTAAACGTCAGCTCG
TGGAAACCCGCCAAATCACAAAGCATGTTGCACAGATACTAGATTCCCGAATGAATACGAAATACGACGAGAACGAT
AAGCTGATTCGGGAAGTCAAAGTAATCACTTTAAAGTCAAAATTGGTGTCGGACTTCAGAAAGGATTTTCAATTCTA
TAAAGTTAGGGAGATAAATAACTACCACCATGCGCACGACGCTTATCTTAATGCCGTCGTAGGGACCGCACTCATTA
AGAAATACCCGAAGCTAGAAAGTGAGTTTGTGTATGGTGATTACAAAGTTTATGACGTCCGTAAGATGATCGCGAAA
AGCGAACAGGAGATAGGCAAGGCTACAGCCAAATACTTCTTTTATTCTAACATTATGAATTTCTTTAAGACGGAAAT
CACTCTGGCAAACGGAGAGATACGCAAACGACCTTTAATTGAAACCAATGGGGAGACAGGTGAAATCGTATGGGATA
AGGGCCGGGACTTCGCGACGGTGAGAAAAGTTTTGTCCATGCCCCAAGTCAACATAGTAAAGAAAACTGAGGTGCAG
ACCGGAGGGTTTTCAAAGGAATCGATTCTTCCAAAAAGGAATAGTGATAAGCTCATCGCTCGTAAAAAGGACTGGGA
CCCGAAAAAGTACGGTGGCTTCGATAGCCCTACAGTTGCCTATTCTGTCCTAGTAGTGGCAAAAGTTGAGAAGGGAA
AATCCAAGAAACTGAAGTCAGTCAAAGAATTATTGGGGATAACGATTATGGAGCGCTCGTCTTTTGAAAAGAACCCC
ATCGACTTCCTTGAGGCGAAAGGTTACAAGGAAGTAAAAAAGGATCTCATAATTAAACTACCAAAGTATAGTCTGTT
TGAGTTAGAAAATGGCCGAAAACGGATGTTGGCTAGCGCCGGAGAGCTTCAAAAGGGGAACGAACTCGCACTACCGT
CTAAATACGTGAATTTCCTGTATTTAGCGTCCCATTACGAGAAGTTGAAAGGTTCACCTGAAGATAACGAACAGAAG
CAACTTTTTGTTGAGCAGCACAAACATTATCTCGACGAAATCATAGAGCAAATTTCGGAATTCAGTAAGAGAGTCAT
CCTAGCTGATGCCAATCTGGACAAAGTATTAAGCGCATACAACAAGCACAGGGATAAACCCATACGTGAGCAGGCGG
AAAATATTATCCATTTGTTTACTCTTACCAACCTCGGCGCTCCAGCCGCATTCAAGTATTTTGACACAACGATAGAT
CGCAAACGATACACTTCTACCAAGGAGGTGCTAGACGCGACACTGATTCACCAATCCATCACGGGATTATATGAAAC
TCGGATAGATTTGTCACAGCTTGGGGGTGACGGATCCCCCAAGAAGAAGAGGAAAGTCTCGAGCGACTACAAAGACC

ATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGGCTGCAGGA

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR
RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRK
KLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKA
ILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLA
QIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQS
FIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVT
VKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRE
MIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTT
QKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDH
IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE
LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLI
ARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV
KKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE
QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTT
IDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0536]

[0537] (单下划线:HNH结构域;双下划线:RuvC结构域)。

[0538] 在一些实施方案中,野生型Cas9对应于来自酿脓链球菌(Streptococcuspyogenes)的Cas9(NCBI参考序列:NC_002737.2(核苷酸序列如下);和Uniprot参考序列:Q99ZW2(氨基酸序列如下)):

[0539] ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGATCACTGAT
GAATATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTATCAAAAAAAATCTTATAGG
GGCTCTTTTATTTGACAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTC
GGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGA
CTTGAAGAGTCTTTTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAAGT
TGCTTATCATGAGAAATATCCAACTATCTATCATCTGCGAAAAAAATTGGTAGATTCTACTGATAAAGCGGATTTGC
GCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAGATTTAAATCCTGAT
AATAGTGATGTGGACAAACTATTTATCCAGTTGGTACAAACCTACAATCAATTATTTGAAGAAAACCCTATTAACGC
AAGTGGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGCTCAGC
TCCCCGGTGAGAAGAAAAATGGCTTATTTGGGAATCTCATTGCTTTGTCATTGGGTTTGACCCCTAATTTTAAATCA
AATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTAGATAATTTATTGGC
GCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGATGCTATTTTACTTTCAGATATCC

TAAGAGTAAATACTGAAATAACTAAGGCTCCCCTATCAGCTTCAATGATTAAACGCTACGATGAACATCATCAAGAC
TTGACTCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTATAAAGAAATCTTTTTTGATCAATCAAAAAA
CGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAAATTTATCAAACCAATTTTAGAAAAAA
TGGATGGTACTGAGGAATTATTGGTGAAACTAAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACAACGGC
TCTATTCCCCATCAAATTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTTAAA
AGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTCCATTGGCGCGTGGCAATA
GTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCATGGAATTTTGAAGAAGTTGTCGATAAAGGT
GCTTCAGCTCAATCATTTATTGAACGCATGACAAACTTTGATAAAAATCTTCCAAATGAAAAAGTACTACCAAAACA
TAGTTTGCTTTATGAGTATTTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTGAAGGAATGCGAAAAC
CAGCATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAACAAATCGAAAAGTAACCGTTAAG
CAATTAAAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGAAATTTCAGGAGTTGAAGATAGATTTAA
TGCTTCATTAGGTACCTACCATGATTTGCTAAAAATTATTAAAGATAAAGATTTTTTGGATAATGAAGAAAATGAAG
ATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAAGATAGGGAGATGATTGAGGAAAGACTTAAAACATAT
GCTCACCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTTTGTCTCGAAA
ATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTAGATTTTTTGAAATCAGATGGTTTTGCCAATC
GCAATTTTATGCAGCTGATCCATGATGATAGTTTGACATTTAAAGAAGACATTCAAAAAGCACAAGTGTCTGGACAA
GGCGATAGTTTACATGAACATATTGCAAATTTAGCTGGTAGCCCTGCTATTAAAAAAGGTATTTTACAGACTGTAAA
AGTTGTTGATGAATTGGTCAAAGTAATGGGGCGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACGTGAAAATC
AGACAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAATCGAAGAAGGTATCAAAGAATTAGGAAGT
CAGATTCTTAAAGAGCATCCTGTTGAAAATACTCAATTGCAAAATGAAAAGCTCTATCTCTATTATCTCCAAAATGG
AAGAGACATGTATGTGGACCAAGAATTAGATATTAATCGTTTAAGTGATTATGATGTCGATCACATTGTTCCACAAA
GTTTCCTTAAAGACGATTCAATAGACAATAAGGTCTTAACGCGTTCTGATAAAAATCGTGGTAAATCGGATAACGTT
CCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAACTTCTAAACGCCAAGTTAATCACTCAACGTAA
GTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTGAACTTGATAAAGCTGGTTTTATCAAACGCCAATTGG
TTGAAACTCGCCAAATCACTAAGCATGTGGCACAAATTTTGGATAGTCGCATGAATACTAAATACGATGAAAATGAT
AAACTTATTCGAGAGGTTAAAGTGATTACCTTAAAATCTAAATTAGTTTCTGACTTCCGAAAAGATTTCCAATTCTA
TAAAGTACGTGAGATTAACAATTACCATCATGCCCATGATGCGTATCTAAATGCCGTCGTTGGAACTGCTTTGATTA
AGAAATATCCAAAACTTGAATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTCGTAAAATGATTGCTAAG
TCTGAGCAAGAAATAGGCAAAGCAACCGCAAAATATTTCTTTTACTCTAATATCATGAACTTCTTCAAAACAGAAAT
TACACTTGCAAATGGAGAGATTCGCAAACGCCCTCTAATCGAAACTAATGGGGAAACTGGAGAAATTGTCTGGGATA
AAGGGCGAGATTTTGCCACAGTGCGCAAAGTATTGTCCATGCCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAG
ACAGGCGGATTCTCCAAGGAGTCAATTTTACCAAAAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGA
TCCAAAAAAATATGGTGGTTTTGATAGTCCAACGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGA
AATCGAAGAAGTTAAAATCCGTTAAAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAAATCCG
ATTGACTTTTTAGAAGCTAAAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACCTAAATATAGTCTTTT
TGAGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAATTACAAAAAGGAAATGAGCTGGCTCTGCCAA
GCAAATATGTGAATTTTTTATATTTAGCTAGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAA
CAATTGTTTGTGGAGCAGCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCGTGTTAT
TTTAGCAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAGACAAACCAATACGTGAACAAGCAG

AAAATATTATTCATTTATTTACGTTGACGAATCTTGGAGCTCCCGCTGCTTTTAAATATTTTGATACAACAATTGAT
CGTAAACGATATACGTCTACAAAAGAAGTTTTAGATGCCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAAC
ACGCATTGATTTGAGTCAGCTAGGAGGTGACTGA

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR
RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRK
KLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKA
ILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLA
QIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQS
FIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVT
VKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRE
[0540] MIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTT
QKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDH
IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE
LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLI
ARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV
KKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE
QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTT
IDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD(单下划线:HNH结构域;双下划线:
RuvC结构域)

[0541] 在一些实施方案中,Cas9是指来自以下的Cas9:溃疡棒状杆菌(Corynebacteriumulcerans)(NCBI Refs:NC_015683.1,NC_017317.1);白喉棒状杆菌(Corynebacteriumdiphtheria)(NCBI Refs:NC_016782.1,NC_016786.1);螺原体(Spiroplasma syrphidicola)(NCBI Ref:NC_021284.1);中间普雷沃菌(Prevotella intermedia)(NCBI Ref:NC_017861.1);中国台湾螺原体(Spiroplasma taiwanense, China)(NCBI Ref:NC_021846.1);海豚链球菌(Streptococcus iniae)(NCBI Ref:NC_021314.1);拟杆菌(Belliella baltica)(NCBI Ref:NC_018010.1);扭曲冷弯曲菌(Psychroflexus torquisI)(NCBI Ref:NC_018721.1);嗜热链球菌(Streptococcus thermophilus)(NCBI Ref:YP_820832.1);无害利斯特菌(Listeria innocua)(NCBI Ref:NP_472073.1);空肠弯曲杆菌(Campylobacter jejuni)(NCBI Ref:YP_002344900.1);脑膜炎奈瑟氏菌(Neisseria meningitidis)(NCBI Ref:YP_002342100.1)或者是指来自任何其他生物体的Cas9。

[0542] 应知悉,其他Cas9蛋白(例如,核酸酶死亡的Cas9(dCas9)、Cas9切口酶(nCas9)或

核酸酶活性的Cas9),包括其变体和同源物,处于本公开的范畴内。示例性Cas9蛋白包括而不限于下文提供的那些。在一些实施方案中,Cas9蛋白是核酸酶死亡的Cas9(dCas9)。在一些实施方案中,Cas9蛋白是Cas9切口酶(nCas9)。在一些实施方案中,Cas9蛋白是核酸酶活性的Cas9。

[0543] 在一些实施方案中,Cas9结构域是无核酸酶活性的Cas9结构域(dCas9)。例如,dCas9结构域可以结合至双螺旋核酸分子(例如,经由gRNA分子)而不裂解该双螺旋核酸分子的任何链。在一些实施方案中,无核酸酶活性的dCas9结构域包含本文所详述的氨基酸序列的D10X突变和H840X突变,或本文所提供的任何氨基酸序列的相应突变,其中X为任何氨基酸变化。在一些实施方案中,无核酸酶活性的dCas9结构域包含本文所详述的氨基酸序列的D10A突变和H840A突变,或本文所提供的任何氨基酸序列的相应突变。作为一个示例,无核酸酶活性的Cas9结构域包含克隆载体pPlatTET-gRNA2(登录号BAV54124)中详述的氨基酸序列。

[0544] 示例性的无催化活性的Cas9(dCas9)的氨基酸序列如下:MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0545] (参见例如,Qi et al.,“Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression.”Cell.2013;152(5):1173-83,其整体内容通过引用并入本文)。

[0546] 在一些实施方案中,Cas9核酸酶具有无活性的(例如,灭活的)DNA裂解结构域,换言之,Cas9是切口酶,称为“nCas9”蛋白(对于“切口酶”Cas9)。经核酸酶灭活的Cas9蛋白可以互换地称为“dCas9”蛋白(对于核酸酶“死亡的”Cas9)或无催化活性的Cas9。用于产生具有无活性的DNA裂解结构域的Cas9蛋白(或其片段)的方法是已知的(参见例如,Jinek et al.,Science.337:816-821(2012);Qi et al.,“Repurposing CRISPR as an RNA-Guided

Platform for Sequence-Specific Control of Gene Expression”(2013)Cell.28;152(5):1173-83,其各自的整体内容通过引用并入本文)。例如,Cas9的DNA裂解结构域已知包括两个亚结构域,HNH核酸酶亚结构域和RuvC1亚结构域。HNH亚结构域裂解与gRNA互补的链,而RuvC1亚结构域裂解非互补链。这些亚结构域中的突变可以静默Cas9的核酸酶活性。例如,突变D10A和H840A将酿脓链球菌Cas9的核酸酶活性完全灭活(Jinek et al.,Science.337:816-821(2012);Qi et al.,Cell.28;152(5):1173-83(2013))。

[0547] 在一些实施方案中,dCas9结构域包含氨基酸序列,所述氨基酸序列与本文提供的任何一个dCas9结构域为至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相同。在一些实施方案中,Cas9结构域包含氨基酸序列,与本文详述的任何一个氨基酸序列相比,该氨基酸序列具有1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、21、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50或更多个突变。在一些实施方案中,Cas9结构域包含氨基酸序列,与本文详述的任何一个氨基酸序列相比,该氨基酸序列具有至少10个、至少15个、至少20个、至少30个、至少40个、至少50个、至少60个、至少70个、至少80个、至少90个、至少100个、至少150个、至少200个、至少250个、至少300个、至少350个、至少400个、至少500个、至少600个、至少700个、至少800个、至少900个、至少1000个、至少1100个或至少1200个相同的连续氨基酸残基。

[0548] 在一些实施方案中,dCas9部分地或完整地对应于或包含具有一个或多个突变的Cas9氨基酸序列,该突变将Cas9核酸酶活性灭活。例如,在一些实施方案中,dCas9结构域包含D10A和H840A突变或另一Cas9中的相应突变。

[0549] 在一些实施方案中,dCas9包含dCas9的氨基酸序列(D10A和H840A):

[0550] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRK

KLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKA
ILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLA
QIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQS
FIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVT
VKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRE
MIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
[0551] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTT
QKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDA
IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE
LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLI
ARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV
KKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE
QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTT
IDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD（单下划线:HNH结构域;双下划线:RuvC结构域）。

[0552] 在一些实施方案中,Cas9结构域包含D10A突变,而位于位置840处的残基保持为上文提供的氨基酸序列中的组氨酸,或者位于本文提供的任何氨基酸序列中的相应位置处。

[0553] 在其他实施方案中,提供了具有除D10A和H840A之外的突变的dCas9变体,该突变例如导致经核酸酶灭活的Cas9(dCas9)。举例而言,此类突变包括位于D10和H840处的其他氨基酸置换,或Cas9的核酸酶结构域内的其他置换(例如,HNH核酸酶亚结构域和/或RuvC1亚结构域中的置换)。在一些实施方案中,提供了dCas9的变体或同源物,该变体或同源物至少约70%相同,至少约80%相同,至少约70%相同,至少约80%相同,至少约90%相同,至少约95%相同,至少约98%相同,至少约99%相同,至少约99.5%相同,或至少约99.9%相同。在一些实施方案中,提供了dCas9的变体,该变体的氨基酸序列更短或更长,短了或长了约5个氨基酸、约10个氨基酸、约15个氨基酸、约20个氨基酸、约25个氨基酸、约30个氨基酸、约40个氨基酸、约50个氨基酸、约75个氨基酸、约100个氨基酸或更多。

[0554] 在一些实施方案中,Cas9结构域是Cas9切口酶。Cas9切口酶可以是能够仅裂解双螺旋核酸分子(例如,双螺旋DNA分子)的一条链的Cas9蛋白。在一些实施方案中,Cas9切口酶裂解双螺旋核酸分子的靶标链,意味着Cas9切口酶裂解与结合至Cas9的gRNA(例如,sgRNA)碱基配对(互补)的链。在一些实施方案中,Cas9切口酶包含D10A突变并且具有位于位置840处的组氨酸。在一些实施方案中,Cas9切口酶裂解双螺旋核酸分子的非靶标、未经碱基编辑的链,意味着Cas9切口酶裂解与结合至Cas9的gRNA(例如,sgRNA)不进行碱基配对(互补)的链。在一些实施方案中,Cas9切口酶包含H840A突变并且具有位于位置10处的天冬氨酸残基,或相应突变。在一些实施方案中,Cas9切口酶包含氨基酸序列,所述氨基酸序列

与本文提供的任何一个Cas9切口酶为至少60％、至少65％、至少70％、至少75％、至少80％、至少85％、至少90％、至少95％、至少96％、至少97％、至少98％、至少99％或至少99.5％相同。基于本公开和本领域的知识,其他合适的Cas9切口酶对于本领域技术人员将是显而易见的,并且处于本公开的范畴内。

[0555] 示例性的催化Cas9切口酶(nCas9)的氨基酸序列如下:

[0556] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKP
IREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0557] 在一些实施方案中,Cas9是指来自古生菌(例如,纳古菌)的Cas9,其构建单细胞原核微生物的结构域和界。在一些实施方案中,可编程核苷酸结合蛋白可以是CasX或CasY蛋白,它们已经在例如Burstein etal.,"New CRISPR-Cas systems from uncultivated microbes."Cell Res.2017Feb 21.doi:10.1038/cr.2017.21中有所描述,该文献的整体内容通过引用并入本文。使用基因组解析的宏基因组学,鉴定了大量CRISPR-Cas系统,包括在古生菌生命结构域中首次报道的Cas9。这种趋异Cas9蛋白是作为活性CRISPR-Cas系统的一部分在很少研究的纳古菌中被发现的。在细菌中,发现了两种先前未知的系统,CRISPR-CasX和CRISPR-CasY,它们是迄今为止所发现的最紧凑的系统之一。在一些实施方案中,在本文所述的碱基编辑器系统中,Cas9被替换为CasX或CasX的变体。在一些实施方案中,在本文所述的碱基编辑器系统中,Cas9被替换为CasY或CasY的变体。应知悉,其他RNA引导的DNA结合蛋白可以用作核酸可编程DNA结合蛋白(napDNAbp),并且处于本公开的范畴内。

[0558] 在一些实施方案中,本文所提供的任何融合蛋白的核酸可编程DNA结合蛋白(napDNAbp)可以是CasX或CasY蛋白。在一些实施方案中,napDNAbp是CasX蛋白。在一些实施方案中,napDNAbp是CasY蛋白。在一些实施方案中,napDNAbp包含氨基酸序列,所述氨基酸序列与天然出现的CasX或CasY蛋白为至少85％、至少90％、至少91％、至少92％、至少93％、至少94％、至少95％、至少96％、至少97％、至少98％、至少99％或至少99.5％相同。在一些

实施方案中,可编程核苷酸结合蛋白是天然出现的CasX或CasY蛋白。在一些实施方案中,可编程核苷酸结合蛋白包含氨基酸序列,所述氨基酸序列与本文描述的任何CasX或CasY蛋白为至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相同。应知悉,来自其他细菌物种的CasX和CasY也可以根据本公开使用。

[0559] 示例性CasX((uniprot.org/uniprot/F0NN87;

[0560] uniprot.org/uniprot/F0NH53)tr|F0NN87|F0NN87_SULIHCRISPR-相关Casx蛋白OS=冰岛硫化叶菌(菌株HVE10/4)GN=SiH_0402PE=4SV=1)氨基酸序列如下:

[0561] MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAERRGKAKKKKGEEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKECEEVSAPSFVKPEFYEFGRSPGMVERTRRVKLEVEPHYLIIAAAGWVLTRLGKAKVSEGDYVGVNVFTPTRGILYSLIQNVNGIVPGIKPETAFGLWIARKVVSSVTNPNVSVVRIYTISDAVGQNPTTINGGFSIDLTKLLEKRYLLSERLEAIARNALSISSNMRERYIVLANYIYEYLTG SKRLEDLLYFANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG。

[0562] 示例性CasX(>tr|F0NH53|F0NH53_SULIR CRISPR相关蛋白,Casx OS=冰岛硫化叶菌(Sulfolobus islandicus)(菌株REY15A)GN=SiRe_0771PE=4SV=1)氨基酸序列如下:

[0563] MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAERRGKAKKKKGEEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKECEEVSAPSFVKPEFYKFGRSPGMVERTRRVKLEVEPHYLIMAAAGWVLTRLGKAKVSEGDYVGVNVFTPTRGILYSLIQNVNGIVPGIKPETAFGLWIARKVVSSVTNPNVSVVSIYTISDAVGQNPTTINGGFSIDLTKLLEKRDLLSERLEAIARNALSISSNMRERYIVLANYIYEYLTGSKRLEDLLYFANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG。

[0564] δ变形菌CasX

[0565] MEKRINKIRKKLSADNATKPVSRSGPMKTLLVRVMTDDLKKRLEKRRKKPEVMPQVISNNAANNLRMLLDDYTKMKEAILQVYWQEFKDDHVGLMCKFAQPASKKIDQNKLKPEMDEKGNLTTAGFACSQCGQPLFVYKLEQVSEKGKAYTNYFGRCNVAEHEKLILLAQLKPVKDSDEAVTYSLGKFGQRALDFYSIHVTKESTHPVKPLAQIAGNRYASGPVGKALSDACMGTIASFLSKYQDIIIEHQKVVKGNQKRLESLRELAGKENLEYPSVTLPPQPHTKEGVDfAYNEVIARVRMWVNLNLWQKLKLSRDDAKPLLRLKGFPSFPVVERRENEVDWWNTINEVKKLIDAKRDMGRVFWSGVTAEKRNTILEGYNYLPNENDHKKREGSLENPKKPAKRQFGDLLLYLEKKYAGDWGKVFDEAWERIDKKIAGLTSHIEREEARNAEDAQSKAVLTDWLRAKASFVLERLKEMDEKEFYACEIQLQKWYGDLRGNPFAVEAENRVVDISGFSIGSDGHSIQYRNLLAWKYLENGKREFYLLMNYGKKGRIRFTDGTDIKKSGKWQGLLYGGGKAKVIDLTFDPDDEQLIILPLAFGTRQGREFIWNDLLSLETGLIKLANGRVIEKTIYNKKIGRDEPALFVALTFERREVVDPSNIKPVNLIGVARGENIPAVIALTDPEGCPLPEFKDSSGGPTDILRIGEGYKEKQRAIQAAKEVEQRRAGGYSRKFASKSRNLADDMVRNSARDLFYHAVTHDAVLVFANLSRGFGRQGKRTFMTERQYTKMEDWLTAKLAYEGLTSKTYLSKTLAQYTSKTCSNCGFTITYADMDVMLVRLKKTSDGWATTLNNKELKAEYQITYYNRYKRQTVEKELSAELDRLSEESGNNDISKWTKGRRDEALFLLKKRFSHRPVQEQFVCLDCGHEVHAAEQAALNIARSWLFLNSNSTEFKSYKSGKQPFVGAWQAFYKRRLKEVWKPNA

[0566] 示例性CasY((ncbi.nlm.nih.gov/protein/APG80656.1)>APG80656.1 CRISPR-相关蛋白CasY[未经培养的俭菌(Parcubacteria)菌群细菌])氨基酸序列如下:

[0567] MSKRHPRISGVKGYRLHAQRLEYTGKSGAMRTIKYPLYSSPSGGRTVPREIVSAINDDYVGLYGLSNFDDLYNAEKRNEEKVYSVLDFWYDCVQYGAVFSYTAPGLLKNVAEVRGGSYELTKTLKGSHLYDELQIDKVIKFLNKKEISRANGSLDKLKKDIIDCFKAEYRERHKDQCNKLADDIKNAKKDAGASLGERQKKLFRDFFGISEQSENDKPSF

TNPLNLTCCLLPFDTVNNNRNRGEVLFNKLKEYAQKLDKNEGSLEMWEYIGIGNSGTAFSNFLGEGFLGRLRENKI
TELKKAMMDITDAWRGQEQEEELEKRLRILAALTIKLREPKFDNHWGGYRSDINGKLSSWLQNYINQTVKIKEDLK
GHKKDLKKAKEMINRFGESDTKEEAVVSSLLESIEKIVPDDSADDEKPDIPAIAIYRRFLSDGRLTLNRFVQREDV
QEALIKERLEAEKKKKPKKRKKKSDAEDEKETIDFKELFPHLAKPLKLVPNFYGDSKRELYKKYKNAAIYTDALWK
AVEKIYKSAFSSSLKNSFFDTDFDKDFFIKRLQKIFSVYRRFNTDKWKPIVKNSFAPYCDIVSLAENEVLYKPKQS
RSRKSAAIDKNRVRLPSTENIAKAGIALARELSVAGFDWKDLLKKEEHEEYIDLIELHKTALALLLAVTETQLDIS
ALDFVENGTVKDFMKTRDGNLVLEGRFLEMFSQSIVFSELRGLAGLMSRKEFITRSAIQTMNGKQAELLYIPHEFQ
SAKITTPKEMSRAFLDLAPAEFATSLEPESLSEKSLLKLKQMRYYPHYFGYELTRTGQGIDGGVAENALRLEKSPV
KKREIKCKQYKTLGRGQNKIVLYVRSSYYQTQFLEWFLHRPKNVQTDVAVSGSFLIDEKKVKTRWNYDALTVALEP
VSGSERVFVSQPFTIFPEKSAEEEGQRYLGIDIGEYGIAYTALEITGDSAKILDQNFISDPQLKTLREEVKGLKLD
QRRGTFAMPSTKIARIRESLVHSLRNRIHHLALKHKAKIVYELEVSRFEEGKQKIKKVYATLKKADVYSEIDADKN
LQTTVWGKLAVASEISASYTSQFCGACKKLWRAEMQVDETITTQELIGTVRVIKGGTLIDAIKDFMRPPIFDENDT
PFPKYRDFCDKHHISKKMRGNSCLFICPFCRANADADIQASQTIALLRYVKEEKKVEDYFERFRKLKNIKVLGQMK
KI。

[0568] 在一些实施方案中,核酸可编程DNA结合蛋白(napDNAbp)是微生物CRISPR-Cas系统的单个效应子。微生物CRISPR-Cas系统的单个效应子包括而不限于,Cas9、Cpf1、Cas12b/C2c1和Cas12c/C2c3。典型地,微生物CRISPR-Cas系统被分为1类系统和2类系统。1类系统具有多亚基效应子复合物,而2类系统具有单蛋白效应子。例如,Cas9和Cpf1是2类效应子。除了Cas9和Cpf1之外,三种截然不同的2类CRISPR-Cas系统(Cas12b/C2c1和Cas12c/C2c3)已经由Shmakov etal.,“Discovery and Functional Characterization of Diverse Class 2CRISPR Cas Systems”,Mol.Cell,2015Nov.5;60(3):385-397描述,其整体内容通过引用并入本文。两种系统Cas12b/C2c1和Cas12c/C2c3的效应子含有与Cpf1相关的Ruv样核酸内切酶结构域。第三种系统含有具有两个称为HEPN RNase的结构域的效应子。不同于通过Cas12b/C2c1产生CRISPR RNA,成熟CRISPR RNA的生产为tracrRNA-独立的。Cas12b/C2c1依赖CRISPR RNA和tracrRNA两者进行DNA裂解。

[0569] 嗜酸耐热性脂环酸芽孢杆菌(Alicyclobaccillus acidoterrastris)Cas12b/C2c1(AacC2c1)的晶体结构已经作为具有嵌合单分子向导RNA(sgRNA)的复合物而有所报道。参见例如,Liu etal.,“C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism”,Mol.Cell,2017Jan.19;65(2):310-322,其整体内容通过引用并入本文。晶体结构已经作为结合至靶标DNA的嗜酸耐热性脂环酸芽孢杆菌(Alicyclobaccillus acidoterrastris)C2c1三元配合物有所报道。参见例如,Yang et al.,“PAM-dependent Target DNA Recognition and Cleavage by C2C1 CRISPR-Cas endonuclease”,Cell,2016Dec.15;167(7):1814-1828,其整体内容通过引用并入本文。AacC2c1的带有靶标和非靶标DNA链的具有催化活性的构象,已经被捕获并独立地定位在单个RuvC催化带中,且Cas12b/C2c1介导的裂解导致靶标DNA的交错的七核苷酸断裂。Cas12b/C2c1三元配合物与先前鉴定的Cas9和Cpf1配对物之间的结构比较表明了CRISPR-Cas9系统所使用的机制的多样性。

[0570] 在一些实施方案中,本文所提供的任何融合蛋白的核酸可编程DNA结合蛋白(napDNAbp)可以是Cas12b/C2c1或Cas12c/C2c3蛋白。在一些实施方案中,napDNAbp是

Cas12b/C2c1蛋白。在一些实施方案中,napDNAbp是Cas12c/C2c3蛋白。在一些实施方案中,napDNAbp包含氨基酸序列,所述氨基酸序列与天然出现的Cas12b/C2c1或Cas12c/C2c3蛋白为至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相同。在一些实施方案中,napDNAbp是天然出现的Cas12b/C2c1或Cas12c/C2c3蛋白。在一些实施方案中,napDNAbp包含氨基酸序列,所述氨基酸序列与本文所提供的任何一种napDNAbp序列为至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相同。应知悉,来自其他细菌物种的Cas12b/C2c1或Cas12c/C2c3也可以根据本公开使用。

[0571] Cas12b/C2c1((uniprot.org/uniprot/T0D7A2#2)sp|T0D7A2|C2C1_ALIAG CRISPR相关核酸内切酶C2c1 OS=嗜酸耐热性脂环酸芽孢杆菌(Alicyclobacillus acido-terrestris)(菌株ATCC 49025/DSM 3922/CIP 106132/NCIMB 13137/GD3B)GN=c2c1 PE=1SV=1)氨基酸序列如下:

[0572] MAVKSIKVKLRLDDMPEIRAGLWKLHKEVNAGVRYYTEWLSLLRQENLYRRSPNGDGEQECDKTAEEC
KAELLERLRARQVENGHRGPAGSDDELLQLARQLYELLVPQAIGAKGDAQQIARKFLSPLADKDAVGGLGIAKAGN
KPRWVRMREAGEPGWEEEKEKAETRKSADRTADVLRALADFGLKPLMRVYTDSEMSSVEWKPLRKGQAVRTWDRDM
FQQAIERMMSWESWNQRVGQEYAKLVEQKNRFEQKNFVGQEHLVHLVNQLQQDMKEASPGLESKEQTAHYVTGRAL
RGSDKVFEKWGKLAPDAPFDLYDAEIKNVQRRNTRRFGSHDLFAKLAEPEYQALWREDASFLTRYAVYNSILRKLN
HAKMFATFTLPDATAHPIWTRFDKLGGNLHQYTFLFNEFGERRHAIRFHKLLKVENGVAREVDDVTVPISMSEQLD
NLLPRDPNEPIALYFRDYGAEQHFTGEFGGAKIQCRRDQLAHMHRRRGARDVYLNVSVRVQSQSEARGERRPPYAA
VFRLVGDNHRAFVHFDKLSDYLAEHPDDGKLGSEGLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNSKGRVPF
FFPIKGNDNLVAVHERSQLLKLPGETESKDLRAIREERQRTLRQLRTQLAYLRLLVRCGSEDVGRRERSWAKLIEQ
PVDAANHMTPDWREAFENELQKLKSLHGICSDKEWMDAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYAKDV
VGGNSIEQIEYLERQYKFLKSWSFFGKVSGQVIRAEKGSRFAITLREHIDHAKEDRLKKLADRIIMEALGYVYALD
ERGKGKWVAKYPPCQLILLEELSEYQFNNDRPPSENNQLMQWSHRGVFQELINQAQVHDLLVGTMYAAFSSRFDAR
TGAPGIRCRRVPARCTQEHNPEPFPWWLNKFVVEHTLDACPLRADDLIPTGEGEIFVSPFSAEEGDFHQIHADLNA
AQNLQQRLWSDFDISQIRLRCDWGEVDGELVLIPRLTGKRTADSYSNKVFYTNTGVTYYERERGKKRRKVFAQEKL
SEEEAELLVEADEAREKSVVLMRDPSGIINRGNWTRQKEFWSMV NQRIEGYLVKQIRSRVPLQDSACENTGDI

[0573] BhCas12b(外村尚芽孢杆菌(Bacillus hisashii))NCBI参考序列:WP_095142515

[0574] MAPKKKRKVGIHGVPAAATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYEHHEQD
PKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVEKKGEANQLSNKFLYPLVDPNSQ
SGKGTASSGRKPRWYNLKIAGDPSWEEEKKKWEEDKKKDPLAKILGKLAEYGLIPLFIPYTDSNEPIVKEIKWMEKS
RNQSVRRLDKDMFIQALERFLSWESWNLKVKEEYEKVEKEYKTLEERIKEDIQALKALEQYEKERQEQLLRDTLNTN
EYRLSKRGLRGWREIIQKWLKMDENEPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYATF
CEIDKKKKDAKQQATFTLADPINHPLWVRFEERSGSNLNKYRILTEQLHTEKLKKKLTVQLDRLIYPTESGGWEEKG
KVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGTLGGARVQFDRDHLRRYPHKVESGNVGRIYFNMTVN
IEPTESPVSKSLKIHRDDFPKVVNFKPKELTEWIKDSKGKKLKSGIESLEIGLRVMSIDLGQRQAAAASIFEVVDQK
PDIEGKLFFPIKGTELYAVHRASFNIKLPGETLVKSREVLRKAREDNLKLMNQKLNFLRNVLHFQQFEDITEREKRV
TKWISRQENSDVPLVYQDELIQIRELMYKPYKDWVAFLKQLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDE

IDRTRKFLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHLNALKEDRLKKMANTIIMHALGYCYDVRKKKWQAKNPACQ
IILFEDLSNYNPYEERSRFENSKLMKWSRREIPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKL
QDNRFFKNLQREGRLTLDKIAVLKEGDLYPDKGGEKFISLSKDRKCVTTHADINAAQNLQKRFWTRTHGFYKVYCKA
YQVDGQTVYIPESKDQKQKIIEEFGEGYFILKDGVYEWVNAGKLKIKKGSSKQSSSELVDSDILKDSFDLASELKGE
KLMLYRDPSGNVFPSDKWMAAGVFFGKLERILISKLTNQYSISTIEDDSSKQSMKRPAATKKAGQAKKKK

[0575] 在一些实施方案中,Cas12b是BvCas12B,其是BhCas12b的变体并且包含以下相对于BhCas12B的变化:S893R、K846R和E837G。BvCas12b(芽孢杆菌属种V3-13)NCBI参考序列:WP_101661451.1

[0576] MAIRSIKLKMKTNSGTDSIYLRKALWRTHQLINEGIAYYMNLLTLYRQEAIGDKTKEAYQAELINIIRN
QQRNNGSSEEHGSDQEILALLRQLYELIIPSSIGESGDANQLGNKFLYPLVDPNSQSGKGTSNAGRKPRWKRLKEEG
NPDWELEKKKDEERKAKDPTVKIFDNLNKYGLLPLFPLFTNIQKDIEWLPLGKRQSVRKWDKDMFIQAIERLLSWES
WNRRVADEYKQLKEKTESYYKEHLTGGEEWIEKIRKFEKERNMELEKNAFAPNDGYFITSRQIRGWDRVYEKWSKLP
ESASPEELWKVVAEQQNKMSEGFGDPKVFSFLANRENRDIWRGHSERIYHIAAYNGLQKKLSRTKEQATFTLPDAIE
HPLWIRYESPGGTNLNLFKLEEKQKKNYYVTLSKIIWPSEEKWIEKENIEIPLAPSIQFNRQIKLKQHVKGKQEISF
SDYSSRISLDGVLGGSRIQFNRKYIKNHKELLGEGDIGPVFFNLVVDVAPLQETRNGRLQSPIGKALKVISSDFSKV
IDYKPKELMDWMNTGSASNSFGVASLLEGMRVMSIDMGQRTSASVSIFEVVKELPKDQEQKLFYSINDTELFAIHKR
SFLLNLPGEVVTKNNKQQRQERRKKRQFVRSQIRMLANVLRLETKKTPDERKKAIHKLMEIVQSYDSWTASQKEVWE
KELNLLTNMAAFNDEIWKESLVELHHRIEPYVGQIVSKWRKGLSEGRKNLAGISMWNIDELEDTRRLLISWSKRSRT
PGEANRIETDEPFGSSLLQHIQNVKDDRLKQMANLIIMTALGFKYDKEEKDRYKRWKETYPACQIILFENLNRYLFN
LDRSRRENSRLMKWAHRSIPRTVSMQGEMFGLQVGDVRSEYSSRFHAKTGAPGIRCHALTEEDLKAGSNTLKRLIED
GFINESELAYLKKGDIIPSQGGELFVTLSKRYKKDSDNNELTVIHADINAAQNLQKRFWQQNSEVYRVPCQLARMGE
DKLYIPKSQTETIKKYFGKGSFVKNNTEQEVYKWEKSEKMKIKTDTTFDLQDLDGFEDISKTIELAQEQQKKYLTMF
RDPSGYFFNNETWRPQKEYWSIVNNIIKSCLKKKILSNKVEL

[0577] Cas9核酸酶具有两个功能性核酸内切酶结构域:RuvC和HNH。当靶向结合核酸酶结构域的那些位置以裂解靶标DNA的相对链时,Cas9经历构型变化。Cas9介导的DNA裂解的最终结果是靶标DNA内(PAM序列上游大约3至4个核苷酸)的双链断裂(DSB)。然后,通过两种常规修复途径中的一种修复所得DSB:(1)高效但易错的非同源末端接合(NHEJ)途径;或(2)低效但高保真同源导向的修复(HDR)途径。

[0578] 非同源末端接合(NHEJ)和/或同源导向的修复(HDR)的“效率”可以通过任何传统方法计算。例如,在一些情况下,效率可以表达为成功的HDR的百分比。例如,检验员核酸酶测定可用来产生裂解结果,并且产物与底物的比率可以用来计算该百分比。例如,可以使用检验员核酸酶,该核酸酶直接裂解含有新并入的限制序列的作为成功HDR的结果的DNA。更多被裂解的底物表明更高百分比的HDR(更高的HDR效率)。作为说明性示例,可以使用以下方程计算HDR的分数(百分比):[(裂解产物)/(底物加上裂解产物)](例如,(b+c)/(a+b+c),其中“a”是DNA底物的条带强度,而“b”和“c”是裂解产物)。

[0579] 在一些情况下,效率可以表达为成功的NHEJ的百分比。例如,T7核酸内切酶I测定可用来产生裂解结果,并且产物与底物的比率可以用来计算该百分比NHEJ。T7核酸内切酶I裂解误配的异源双螺旋DNA,其来自野生型与突变DNA链的杂交(NHEJ在原始段落的位点处产生小的随机插入或缺失(插入缺失(indel)))。更多裂解表明更高百分比的NHEJ(更高的

NHEJ效率)。作为说明性示例,可使用以下方程计算NHEJ的分数(百分比):(1-(1-(b+c)/(a+b+c))$^{1/2}$)×100,其中"a"是DNA底物的条带强度,而"b"和"c"是裂解产物(Ran et.al.,Cell.2013Sep.12;154(6):1380-9;和Ran et al.,Nat Protoc.2013Nov.;8(11):2281-2308)。

[0580] NHEJ修复途径是最具活性的修复机制,并且它经常在DSB位点处造成小核苷酸插入或缺失(indel)。NHEJ介导的DSB修复的随机性具有重要的实践意义,因为表达Cas9和gRNA的细胞群或向导多核苷酸可以导致多样化的突变。在大多数情况下,NHEJ在靶标DNA中给出了小indel,其导致氨基酸缺失、插入或框移突变,这些突变导致所靶向基因的开放阅读框(ORF)内的提取终止密码子。理想的最终结果是所靶向的基因内的失能性突变。

[0581] 尽管NHEJ介导的DSB修复常常中断基因的开放阅读框,当同源导向的修复(HDR)可以用来产生从单个核苷酸变化到大插入如添加荧光团或标签范围内的特异性核苷酸变化。

[0582] 为了利用HDR进行基因编辑,可以使用gRNA以及Cas9或Cas9切口酶将含有所希望序列的DNA修复模板递送至感兴趣的细胞类型中。修复模板可以含有所希望的编辑以及位于紧邻靶标的上游或下游的额外同源序列(称为左侧同源臂和右侧同源臂)。每个同源臂的长度可能取决于所引入的变化尺寸,越大的插入需要越长的同源臂。修复模板可以是单链寡核苷酸、双链寡核苷酸或双链DNA质粒。即便在表达Cas9、gRNA和外源息服务模板的细胞中,HDR的效率通常也较低(<10%的经修饰的等位基因)。HDR的效率可以通过将细胞同步化而提升,因为HDR在细胞周期的S期和G2期发生。NHEJ中牵涉的化学或遗传学地抑制基因也可能增加HDR频率。

[0583] 在一些实施方案中,Cas9是经修饰的Cas9。贯穿存在部分同源性的基因组范围内,给定的gRNA靶向序列可能具有额外位点。这些位点称为脱靶,并且在设计gRNA时需要考虑这些位点。除了优化gRNA设计之外,CRISPR特异性也可应通过对Cas9的修饰而得以增加。Cas9通过两个核酸酶结构域RuvC和HNH的组合活性而产生双链断裂(DSB)。Cas9切口酶,SpCas9的D10A突变体,保留一个核酸酶结构域并且产生DNA切口而非DSB。该切口酶系统也可以与HDR介导的基因编辑组合以进行特异性基因编辑。

[0584] 在一些情况下,Cas9是变体Cas9蛋白。变体Cas9多肽具有当与野生型Cas9蛋白的氨基酸序列比较时相异一个氨基酸(例如,具有缺失、插入、置换、融合)的氨基酸序列。在一些实例中,变体Cas9多肽具有降低Cas9多肽的核酸酶活性的氨基酸变化(例如,缺失、插入或置换)。例如,在一些实例中,变体Cas9多肽具有相应野生型Cas9蛋白的低于50%、低于40%、低于30%、低于20%、低于10%、低于5%或低于1%的核酸酶活性。在一些情况下,变体Cas9蛋白没有实质性的核酸酶活性。当受试Cas9蛋白是没有实质性核酸酶活性的变体Cas9蛋白时,它可以称为"dCas9"。

[0585] 在一些情况下,变体Cas9蛋白具有降低的核酸酶活性。例如,变体Cas9蛋白表现出野生型Cas9蛋白(例如,野生型Cas9蛋白)的低于约20%、低于约15%、低于约10%、低于约5%、低于约1%或低于约0.1%的核酸内切酶活性。

[0586] 在一些情况下,变体Cas9蛋白可以裂解向导靶标序列的互补链,但裂解双链向导靶标序列的非互补链的能力降低。例如,变体Cas9蛋白可以具有降低RuvC结构域的功能的突变(氨基酸置换)。作为非限制性示例,在一些实施方案中,变体Cas9蛋白具有D10A(在氨基酸位置10处的天冬氨酸变为丙氨酸),并且因此可以裂解双链向导靶标序列的互补链但

裂解双链向导靶标序列的非互补链的能力降低(因此,当该变体Cas9蛋白裂解双链靶标核酸时,导致单链断裂(SSB)而非双链断裂(DSB))(参见例如,Jinek et al.,Science.2012Aug.17;337(6096):816-21)。

[0587] 在一些情况下,变体Cas9蛋白可以裂解双链向导靶标序列的非互补链,但裂解该向导靶标序列的互补链的能力降低。例如,变体Cas9蛋白可以具有降低HNH结构域(RuvC/HNH/RuvC结构域基序)的功能的突变(氨基酸置换)。作为非限制性示例,在一些实施方案中,变体Cas9蛋白具有H840A(在氨基酸位置840处的组氨酸变为丙氨酸)突变,并且因此可以裂解该向导靶标序列的非互补链但裂解该向导靶标序列的互补链的能力降低(因此,当该变体Cas9蛋白裂解靶标序列时,导致SSB而非DSB)。此类Cas9蛋白裂解向导靶标序列(例如,单链向导靶标序列)的能力降低,但保留结合向导靶标序列(例如,单链向导靶标序列)的能力。

[0588] 在一些情况下,变体Cas9蛋白裂解双链靶标DNA的互补链和非互补链两者的能力降低。作为非限制性示例,在一些情况下,变体Cas9蛋白荷有D10A和H840A两种突变,使得该多肽裂解双链靶标DNA的互补链和非互补链的能力降低。此类Cas9蛋白裂解靶标DNA(例如,单链靶标DNA)的能力降低,但保留结合靶标DNA(例如,单链靶标DNA)的能力。

[0589] 作为另一种非限制性示例,在一些情况下,变体Cas9蛋白荷有W476A和W1126A突变,使得该多肽裂解靶标DNA的能力降低。此类Cas9蛋白裂解靶标DNA(例如,单链靶标DNA)的能力降低,但保留结合靶标DNA(例如,单链靶标DNA)的能力。

[0590] 作为另一种非限制性示例,在一些情况下,变体Cas9蛋白荷有P475A、W476A、N477A、D1125A、W1126A和D1127A突变,使得该多肽裂解靶标DNA的能力降低。此类Cas9蛋白裂解靶标DNA(例如,单链靶标DNA)的能力降低,但保留结合靶标DNA(例如,单链靶标DNA)的能力。

[0591] 作为另一种非限制性示例,在一些情况下,变体Cas9蛋白荷有H840A、W476A和W1126A突变,使得该多肽裂解靶标DNA的能力降低。此类Cas9蛋白裂解靶标DNA(例如,单链靶标DNA)的能力降低,但保留结合靶标DNA(例如,单链靶标DNA)的能力。作为另一种非限制性示例,在一些情况下,变体Cas9蛋白荷有H840A、D10A、W476A和W1126A突变,使得该多肽裂解靶标DNA的能力降低。此类Cas9蛋白裂解靶标DNA(例如,单链靶标DNA)的能力降低,但保留结合靶标DNA(例如,单链靶标DNA)的能力。在一些实施方案中,变体Cas9具有位于Cas9 HNH结构域中位置840处的经修复的His残基(A840H)。

[0592] 作为另一种非限制性示例,在一些情况下,变体Cas9蛋白荷有H840A、P475A、W476A、N477A、D1125A、W1126A和D1127A突变,使得该多肽裂解靶标DNA的能力降低。此类Cas9蛋白裂解靶标DNA(例如,单链靶标DNA)的能力降低,但保留结合靶标DNA(例如,单链靶标DNA)的能力。作为另一种非限制性示例,在一些情况下,变体Cas9蛋白荷有D10A、H840A、P475A、W476A、N477A、D1125A、W1126A和D1127A突变,使得该多肽裂解靶标DNA的能力降低。此类Cas9蛋白裂解靶标DNA(例如,单链靶标DNA)的能力降低,但保留结合靶标DNA(例如,单链靶标DNA)的能力。在一些情况下,当变体Cas9蛋白荷有W476A和W1126A突变时或当变体Cas9蛋白荷有P475A、W476A、N477A、D1125A、W1126A和D1127A突变时,变体Cas9蛋白不与PAM序列有效地结合。因此,在一些此类情况下,当在结合方法中使用此类变体Cas9蛋白时,该方法不需要PAM序列。换言之,在一些情况下,当在结合方法中使用此类变体Cas9蛋白时,该

方法可以包括向导RNA,但该方法可以在不存在PAM序列的情况下执行(因此,由向导RNA的靶向链段提供结合的特异性)。可以突变其他残基以实现上述效果(即,将一个或另一股核酸酶部分灭活)。作为非限制性示例,可以改变(即,置换)残基D10、G12、G17、E762、H840、N854、N863、H982、H983、A984、D986和/或A987。而且,除丙氨酸置换之外的突变是合适的。

[0593] 在一些实施方案中,具有降低的催化活性的变体Cas9蛋白(例如,当Cas9蛋白具有D10、G12、G17、E762、H840、N854、N863、H982、H983、A984、D986和/或A987突变例如D10A、G12A、G17A、E762A、H840A、N854A、N863A、H982A、H983A、A984A和/或D986A时),变体Cas9蛋白仍可以位点特异性的方式结合至靶标DNA(因为它仍然被向导RNA引导至靶标DNA序列),只要它保留与向导RNA相互作用的能力即可。

[0594] 在一些实施方案中,变体Cas蛋白可以是spCas9、spCas9-VRQR、spCas9-VRER、xCas9(sp)、saCas9、saCas9-KKH、spCas9-MQKSER、spCas9-LRKIQK或spCas9-LRVSQL。

[0595] 在一些实施方案中,使用了包括氨基酸置换D1135M、S1136Q、G1218K、E1219F、A1322R、D1332A、R1335E和T1337R(SpCas9-MQKFRAER)并且具有针对改变的PAM 5'-NGC-3'的特异性的经修饰的SpCas9。

[0596] 作为酿脓链球菌Cas9的替代品,可以包括来自Cpf1家族的RNA引导的核酸内切酶,其在哺乳动物细胞中展示裂解活性。来自普雷沃菌属和弗朗西斯氏菌属1的CRISPR(CRISPR/Cpf1)是CRISPR/Cas9系统的DNA编辑技术同源物。Cpf1是II类CRISPR/Cas系统的RNA引导的核酸内切酶。这一获得性免疫机制见于普雷沃菌属和弗朗西斯氏菌属细菌中。Cpf1基因与CRISPR基因座缔合,编码使用向导RNA来发现并裂解病毒DNA的核酸内切酶。Cpf1是比Cas9更小且更简单的核酸内切酶,克服了CRISPR/Cas9系统的一些限制。不同于Cas9核酸酶,Cpf1介导的DNA裂解结果是具有短3'核苷酸的双链断裂。Cpf1的交错裂解模式可以打开定向基因转移的可能性,类似于传统的限制酶克隆,这可能增加基因编辑的效率。与上述Cas9变体和直系同源物类似,Cpf1也可以扩充能被CRISPR靶向至富AT区域或富AT基因组的位点的数目,该区域或基因组缺乏SpCas9所青睐的NGG PAM位点。Cpf1基因座含有混合的α/β结构域、后面跟随一个螺旋区域的RuvC-I、RuvC-II和锌指样结构域。Cpf1蛋白具有类似于Cas9的RuvC结构域的RuvC样核酸内切酶结构域。此外,Cpf1不具有HNH核酸内切酶结构域,并且Cpf1的N端不具有Cas9的α-螺旋识别叶。Cpf1 CRISPR-Cas结构域架构显示,Cpf1在功能上是独特的,归类为2类V型CRISPR系统。Cpf1基因座编码Cas1、Cas2和Cas4蛋白,相较于II型系统,更类似于I型和III型系统。功能性Cpf1不需要反向激活CRISPR RNA(tracrRNA),因此,仅需要CRISPR(crRNA)。这有益于基因组编辑,因为Cpf1不仅比Cas9小,而且它具有较小的sgRNA分子(核苷酸数大约为Cas9的一半)。与被Cas9靶向的富G的PAM不同,Cpf1-crRNA复合物通过鉴定前间隔序列邻近基序5'-YTN-3'来裂解靶标DNA或RNA。在鉴定PAM之后,Cpf1引入4或5个核苷酸突出的粘性末端样DNA双链断裂。

[0597] 本公开的一些方面提供包含充当核酸可编程DAN结合蛋白的结构域的融合蛋白,其可以被用来将蛋白质诸如碱基编辑器引导至特定核酸(例如,DNA或RNA)序列。在特定实施方案中,融合蛋白包含核酸可编程DNA结合蛋白结构域或脱氨酶结构域。DNA结合蛋白包括而不限于,Cas9(例如,dCas9和nCas9)、Cas12a/Cpf1、Cas12b/C2c1、Cas12c/C2c3、Cas12d/CasY、Cas12e/CasX、Cas12g、Cas12h和Cas12i。具有不同于Cas9的PAM特异性的可编程多核苷酸结合蛋白的一个示例是来自普雷沃菌属和弗朗西丝氏菌属1(Cpf1)的成簇规则

间隔短回文重复序列。类似于Cas9,Cpf1也是2类CRISPR效应子。已经显示,Cpf1介导强劲的DNA干扰,其特征与Cas9截然不同。Cpf1是单个RNA引导的核酸内切酶缺乏性tracrRNA,并且其利用富T的前间隔序列邻近基序(TTN、TTTN或YTN)。此外,Cpf1经由交错的DNA双链断裂来裂解DNA。16种Cpf1家族蛋白质中,来自氨基酸球菌属(Acidaminococcus)和毛螺菌科(Lachnospiraceae)的两种酶显示在人类细胞中具备有效的基因组编辑活性。Cpf1蛋白是本领域中已知的,并且先前已经描述在例如Yamano et al.,“Crystal structure of Cpf1 in complex with guide RNA and target DNA.”Cell(165)2016,p.949-962中,其整体内容通过引用并入本文。

[0598] 也可用于本发明组合物和方法中的是无核酸酶活性的Cpf1(dCpf1)变体,这些变体可以用作向导核苷酸序列可编程多核苷酸结合蛋白结构域。Cpf1蛋白具有类似于Cas9的RuvC结构域的RuvC样核酸内切酶结构域但不具有HNH核酸内切酶结构域,并且Cpf1的N端不具有Cas9的α-螺旋识别叶。Zetsche et al.,Cell,163,759-771,2015(其通过引用并入本文)中显示,Cpf1的RuvC样结构域的主要责任是裂解两条DNA链,并且RuvC样结构域的灭活将Cpf1核酸酶活性灭活。例如,对应于新凶手弗朗西丝氏菌(Francisella novicida)Cpf1中的D917A、E1006A或D1255A的突变将Cpf1核酸酶活性灭活。在一些实施方案中,本公开的dCpf1包含对应于D917A、E1006A、D1255A、D917A/E1006A、D917A/D1255A、E1006A/D1255A或D917A/E1006A/D1255A的突变。应理解,任何将Cpf1的RuvC结构域灭活的突变,例如,置换突变、缺失或插入,均可根据本公开使用。

[0599] 在一些实施方案中,本文所提供的任何融合蛋白的核酸可编程核苷酸结合蛋白可以是Cpf1蛋白。在一些实施方案中,Cpf1蛋白是Cpf1切口酶(nCpf1)。在一些实施方案中,Cpf1蛋白是无核酸酶活性的Cpf1(dCpf1)。在一些实施方案中,Cpf1、nCpf1或dCpf1包含氨基酸序列,所述氨基酸序列与本文公开的Cpf1序列为至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相同。在一些实施方案中,dCpf1包含氨基酸序列,所述氨基酸序列与本文公开的Cpf1序列为至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相同,并且包含对应于D917A、E1006A、D1255A、D917A/E1006A、D917A/D1255A、E1006A/D1255A或D917A/E1006A/D1255A的突变。应知悉,来自其他细菌物种的Cpf1也可以根据本公开使用。

[0600] 野生型新凶手弗朗西丝氏菌(Francisella novicida)Cpf1的氨基酸序列如下。D917、E1006和D1255加粗并加下划线。

[0601] MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSSVC
ISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESDLIL
WLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPK
FLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITK
FNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVT
TMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEY
ITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFD
EIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEH
FYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYL
GVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKN
GSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENIS
ESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKK
ITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKAND
VHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEM
[0602] KEGYLSQVVHEIAKLVIEYNAIVVF**E**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGG
VLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLD
KGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGEC
IKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**D**ANGAY
HIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN.

[0603] 新凶手弗朗西丝氏菌(Francisella novicida)Cpf1 D917A的氨基酸序列如下。(A917、E1006和D1255加粗并加下划线)。

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSSVC
ISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESDLIL
WLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPK
FLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITK
FNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVT
TMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEY
ITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFD
EIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEH
FYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYL

[0604] GVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKN
GSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENIS
ESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKK
ITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKAND
VHILSI**A**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEM
KEGYLSQVVHEIAKLVIEYNAIVVF**E**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGG
VLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLD
KGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGEC
IKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**D**ANGAY
HIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN.

[0605] 新凶手弗朗西丝氏菌(Francisella novicida)Cpf1 E1006A的氨基酸序列如下。(D917、A1006和D1255加粗并加下划线)。

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSSVC

[0606] ISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESDLIL
WLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPK

FLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITK
FNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVT
TMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEY
ITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFD
EIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEH
FYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYL
GVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKN
GSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENIS
ESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKK
ITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKAND
VHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEM
KEGYLSQVVHEIAKLVIEYNAIVVF**A**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGG
VLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLD
KGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGEC
IKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**D**ANGAY
HIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN.

[0607]

[0608] 新凶手弗朗西丝氏菌(Francisella novicida)Cpf1 D1255A的氨基酸序列如下。(D917、E1006和A1255突变位置加粗并加下划线)。

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSSVC
ISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESDLIL
WLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPK
FLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITK
FNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVT
TMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEY
ITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFD
EIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEH
FYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYL
GVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKN
GSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENIS
ESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKK
ITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKAND
VHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEM
KEGYLSQVVHEIAKLVIEYNAIVVF**E**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGG
VLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLD
KGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGEC

[0609]

[0610] IKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**A**ANGAY
HIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0611] 新凶手弗朗西丝氏菌(Francisella novicida)Cpf1 D917A/E1006A的氨基酸序列如下。(A917、A1006和D1255加粗并加下划线)。

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSSVC
ISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESDLIL
WLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPK
FLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITK
FNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVT
TMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEY
ITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFD
EIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEH
FYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYL

[0612] GVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKN
GSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENIS
ESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKK
ITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKAND
VHILSI**A**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEM
KEGYLSQVVHEIAKLVIEYNAIVVF**A**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGG
VLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLD
KGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGEC
IKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**D**ANGAY
HIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN.

[0613] 新凶手弗朗西丝氏菌(Francisella novicida)Cpf1D917A/D1255A的氨基酸序列如下。(A917、E1006和A1255加粗并加下划线)。

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSSVC
ISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESDLIL
WLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPK
FLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITK

[0614] FNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVT
TMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEY
ITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFD
EIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEH
FYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYL

GVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKN
GSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENIS
ESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKK
ITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKAND
VHILSI**A**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEM
[0615] KEGYLSQVVHEIAKLVIEYNAIVVF**E**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGG
VLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLD
KGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGEC
IKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**A**ANGAY
HIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN.

[0616] 新凶手弗朗西丝氏菌(Francisella novicida)Cpf1E1006A/D1255A的氨基酸序列如下。(D917、A1006和A1255加粗并加下划线)。

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSSVC
ISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESDLIL
WLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPK
FLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITK
FNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVT
TMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEY
ITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFD
EIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEH
FYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYL
[0617] GVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKN
GSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENIS
ESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKK
ITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKAND
VHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEM
KEGYLSQVVHEIAKLVIEYNAIVVF**A**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGG
VLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLD
KGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGEC
IKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**A**ANGAY
HIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN.

[0618] 新凶手弗朗西丝氏菌(Francisella novicida)Cpf1D917A/E1006A/D1255A的氨基酸序列如下。(A917、A1006和A1255加粗并加下划线)。

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSSVC
ISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESDLIL
WLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPK
FLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITK
FNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVT
TMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEY
ITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFD
EIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEH
FYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYL
[0619] GVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKN
GSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENIS
ESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKK
ITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKAND
VHILSI**A**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEM
KEGYLSQVVHEIAKLVIEYNAIVVF**A**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGG
VLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLD
KGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGEC
IKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**A**ANGAY
HIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN.

[0620] 在一些实施方案中,存在于融合蛋白中的Cas9结构域之一可以替换为不需要PAM序列的向导核苷酸序列可编程DNA结合蛋白结构域。

[0621] 在一些实施方案中,Cas结构域是来自金黄色葡萄球菌(Staphylococcus aureus)的Cas9结构域(SaCas9)。在一些实施方案中,SaCas9结构域是核酸酶活性SaCas9、无核酸酶活性的SaCas9(SaCas9d)或SaCas9切口酶(SaCas9n)。在一些实施方案中,SaCas9结构域包含N579A突变,或本文所提供的任何氨基酸序列中的相应突变。

[0622] 在一些实施方案中,SaCas9结构域、SaCas9d结构域或SaCas9n结构域可以结合至具有非经典PAM的核酸序列。在一些实施方案中,SaCas9结构域、SaCas9d结构域或SaCas9n结构域可以结合至具有NNGRRT或NNGRRT PAM的核酸序列。在一些实施方案中,SaCas9结构域包含E781X、N967X和R1014X突变中的一种或多种,或本文所提供的任何氨基酸序列的相应突变,其中X是任何氨基酸。在一些实施方案中,SaCas9结构域包含E781K、N967K和R1014H突变中的一种或多种,或本文所提供的任何氨基酸序列的相应突变。在一些实施方案中,SaCas9结构域包含E781K、N967K或R1014H突变,或本文所提供的任何氨基酸序列的相应突变。

[0623] 示例性SaCas9的氨基酸序列如下:

[0624] MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKKL
LFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISR
NSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRT
YYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEK
FQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQ
IAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNR
LKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKM
INEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIP
RSVSFDNSFNNKVLVKQEE**N**SKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEER
DINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKH
HAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYK
YSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKL
KLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLK
PYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYR
VIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQII
KKG。

[0625] 该序列中,加下划线并以粗体显示的残基N579,可以经突变(例如,突变为A579)以得到SaCas9切口酶。

[0626] 示例性SaCas9n的氨基酸序列如下:

[0627] KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKKLL
FDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRN
SKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTY
YEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKF
QIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQI
AKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRL
KLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMI
NEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPR
SVSFDNSFNNKVLVKQEE**A**SKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERD
INRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHH
AEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKY
SHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLK
LIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKP
YRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRV
IGVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIK
KG。

[0628]

[0629] 该序列中,可以从N579突变以得到SaCas9切口酶的残基A579加下划线并以粗体显示。

[0630] 示例性SaKKH Cas9的氨基酸序列如下：

[0631] KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKKLL
FDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRN
SKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTY
YEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKF
QIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQI
AKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRL
KLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMI
NEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPR
SVSFDNSFNNKVLVKQEE**A**SKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERD
INRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHH
AEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKY
SHRVDKKPNR*A*LINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLK
LIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKP
YRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFY*K*NDLIKINGELYRV
IGVNNDLLNRIEVNMIDITYREYLENMNDKRPP*H*IIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIK
KG。

[0632] 上述可以从N579突变以得到SaCas9切口酶的残基A579加下划线并以粗体显示。上述可以从E781、N967和R1014突变以得到SaKKH Cas9的残基K781、K967和H1014加下划线并以斜体显示。

[0633] 高保真度Cas9结构域

[0634] 本公开的一些方面提供高保真度Cas9结构域。在一些实施方案中，高保真度Cas9结构域是经工程化的Cas9结构域，相对于相应的野生型Cas9结构域，该高保真度Cas9结构域包含一种或多种减少Cas9结构域与DNA的糖-磷酸酯主链之间的静电相互作用的突变。与DNA的糖-磷酸酯主链的静电相互作用减少的高保真度Cas9结构域可能具有较低的脱靶效应。在一些实施方案中，Cas9结构域(例如，野生型Cas9结构域)包含一种或多种突变，该突变减少Cas9结构域与DNA的糖-磷酸酯主链之间的缔合。在一些实施方案中，Cas9结构域包含一种或多种突变，该突变将Cas9结构域与DNA的糖-磷酸酯主链之间的缔合减少至少1％、至少2％、至少3％、至少4％、至少5％、至少10％、至少15％、至少20％、至少25％、至少30％、至少35％、至少40％、至少45％、至少50％、至少55％、至少60％、至少65％或至少70％。

[0635] 在一些实施方案中，本文所提供的任何Cas9融合蛋白包含N497X、R661X、Q695X和/或Q926X突变中的一种或多种，或本文所提供的任何氨基酸序列中的相应突变，其中X是任何氨基酸。在一些实施方案中，本文所提供的任何Cas9融合蛋白包含N497A、R661A、Q695A和/或Q926A突变中的一种或多种，或本文所提供的任何氨基酸序列中的相应突变。在一些实施方案中，Cas9结构域包含D10A突变，或本文所提供的任何氨基酸序列中的相应突变。具有高保真度的Cas9结构域是本领域中已知的并且对于技术人员将会是显而易见的。例如，具有高保真度的Cas9结构域已经在以下文献中有所描述：Kleinstiver,B.P.,et al.“High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-

target effects."Nature529,490-495(2016);和Slaymaker,I.M.,et al."Rationally engineered Cas9nucleases with improved specificity."Science 351,84-88(2015);各自的整体内容通过引用并入本文。

[0636] 在一些实施方案中,经修饰的Cas9是高保真度Cas9酶。在一些实施方案中,高保真度Cas9酶是SpCas9(K855A)、eSpCas9(1.1)、SpCas9-HF1或超高精度Cas9变体(HypaCas9)。经修饰的Cas9eSpCas9(1.1)含有谷氨酸置换,该置换弱化了HNH/RuvC凹槽与非靶标DNA链之间的相互作用,防止链分离和在脱靶位点切割。类似地,SpCas9-HF1通过丙氨酸置换降低了脱靶编辑,该丙氨酸置换中断了Cas9与DNA磷酸酯主链的相互作用。HypaCas9含有位于REC3结构域中的突变(SpCas9 N692A/M694A/Q695A/H698A),该突变增加Cas9校对和靶标辨别力。全部三种高保真度酶都产生比野生型Cas9少的脱靶编辑。

[0637] 示例性的高保真度Cas9提供于下。

[0638] 相对于Cas9的高保真度Cas9结构域突变以黑体并加下划线显示。

[0639] MDKKYSIGL**A**IGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR
RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRK
KLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKA
ILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLA
QIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQS
FIERMT**A**FDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVT
VKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRE
MIEERLKTYAHLFDDKVMKQLKRRRYTGWG**A**LSRKLINGIRDKQSGKTILDFLKSDGFANRNFM**A**LIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTT
QKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDH
IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE
LDKAGFIKRQLVETR**A**ITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLI
ARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV
KKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE
QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTT
IDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD。

[0640] 向导多核苷酸

[0641] 在一种实施方案中,向导多核苷酸是向导RNA。RNA/Cas复合物可以辅助将Cas蛋白"引导"至靶标DNA。Cas9/crRNA/tracrRNA核酸内切溶解地裂解与间隔序列互补的线性或环状dsDNA靶标。不与crRNA互补的靶标链首先被核酸内切溶解地切割,然后被3'-5'核酸外切溶解地修剪。本质上中,DNA结合和裂解通常需要蛋白质和两种RNA。但是,单向导RNA

(“sgRNA”或简称为“gNRA”)可以工程化改造,以将crRNA和tracrRNA两者的各个方面合并到单个RNA物种中。参见例如,Jinek M.,et al.,Science 337:816-821(2012),其整体内容通过引用并入本文。Cas9识别CRISPR重复序列中的短基序(PAM或前间隔序列邻近基序)以帮助区分自我与非自我。Cas9核酸酶序列和结构是本领域技术人员已知的(参见例如,“Complete genome sequence of an M1 strain of Streptococcus pyogenes.”Ferretti,J.J.et al.,Natl.Acad.Sci.U.S.A.98:4658-4663(2001);“CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.”Deltcheva E.et al.,Nature 471:602-607(2011);和“Programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.”Jinek M.et al,Science 337:816-821(2012),其各自的整体内容通过引用并入本文)。Cas9直接同源物已经在多种物种中描述,包括但不限于酿脓链球菌和嗜热链球菌。基于本公开,其他合适的Cas9核酸酶和序列对于本领域技术人员可以是显而易见的,并且此类Cas9核酸酶和序列包括来自生物体的Cas9序列和以下文献中公开的基因座:Chylinski,Rhun,and Charpentier,“The tracrRNA and Cas9 families of type IICRISPR-Cas immunity systems”(2013)RNA Biology 10:5,726-737;其整体内容通过引用并入本文。在一些实施方案中,Cas9核酸酶具有无活性的(例如,灭活的)DNA裂解结构域,换言之,Cas9是切口酶。

[0642] 在一些实施方案中,向导多核苷酸是至少一个单向导RNA(“sgRNA”或“gRNA”)。在一些实施方案中,向导多核苷酸是至少一个tracrRNA。在一些实施方案中,向导多核苷酸不需要PAM序列将多核苷酸可编程DNA结合结构域(例如,Cas9或Cpf1)引导至靶标核苷酸序列。

[0643] 本文所公开的碱基编辑器的多核苷酸可编程核苷酸结合结构域(例如,衍生自CRISPR的结构域)可以通过与向导多核苷酸缔合而识别靶标多核苷酸序列。向导多核苷酸(例如,gRNA)典型是单链的,并且可以编程以位点特异性地结合(即,经由互补碱基配对)至多核苷酸的靶标序列,从而将与该向导核酸协同作用的碱基编辑器导向至该靶标序列。向导多核苷酸可以是DNA。向导多核苷酸可以是RNA。在一些情况下,向导多核苷酸包含天然核苷酸(例如,腺苷)。在一些情况下,向导多核苷酸包含非天然(非自然)核苷酸(例如,肽核酸或核苷酸类似物)。在一些情况下,向导核酸序列的靶向区域可以是至少15、16、17、18、19、20、21、22、23、24、25、26、27、28、29或30个核苷酸的长度。向导核酸的靶向区域可以是10至30个核苷酸之间的长度,或15至25个核苷酸之间的长度,或15至20个核苷酸之间的长度。

[0644] 在一些实施方案中,向导多核苷酸包含两个或更多个个体多核苷酸,这些多核苷的可以经由例如互补性碱基配对(例如,双向导多核苷酸)而彼此相互作用。例如,向导多核苷酸可以包含CRISPR RNA(crRNA)和反向激活CRISPR RNA(tracrRNA)。例如,向导多核苷酸可以包含一个或多个反向激活CRISPR RNA(tracrRNA)。

[0645] 在II型CRISPR系统中,通过CRISPR蛋白(例如,Cas9)靶向核酸典型需要在包含识别靶标序列的序列的第一RNA分子(crRNA)与包含重复序列的第二RNA分子(trRNA)之间进行互补性碱基配对,这形成了稳定化向导RNA-CRISPR蛋白复合物的支架区域。此类双向导RNA系统可以用作向导多核苷酸来将本文所公开的碱基编辑器导向至靶标多核苷酸序列。

[0646] 在一些实施方案中,本文所提供的碱基编辑器利用单向导多核苷酸(例如,gRNA)。在一些实施方案中,本文所提供的碱基编辑器利用双向导多核苷酸(例如,双gRNA)。在一些

实施方案中,本文所提供的碱基编辑器利用一个或多个向导多核苷酸(例如,多gRNA)。在一些实施方案中,将单向导多核苷酸用于不同的本文所述碱基编辑器中。例如,单向导多核苷酸可用于胞苷碱基编辑器和腺苷碱基编辑器。

[0647] 在其他实施方案中,向导多核苷酸可以在单分子(即,单分子向导核酸)中包含核酸的多核苷酸靶向部分和核酸的支架部分两者。例如,单分子向导多核苷酸可以是单向导RNA(sgRNA或gRNA)。本文中,术语向导多核苷酸设想了能够与碱基编辑器相互作用并将其导向至靶标多核苷酸序列的任何单、双或多分子核酸。

[0648] 典型地,向导多核苷酸(例如,crRNA/trRNA复合物或gRNA)包含包括能够识别并结合至靶标多核苷酸序列的序列的"多核苷酸靶向链段"和将碱基编辑器的多核苷酸可编程核苷酸结合结构域内的向导多核苷酸"蛋白质结合链段"。在一些实施方案中,向导多核苷酸的多核苷酸靶向链段识别并结合至DNA多核苷酸,从而促进DNA内碱基的编辑。在其他情况下,向导多核苷酸的多核苷酸靶向链段识别并结合至RNA多核苷酸,从而促进RNA内碱基的编辑。本文中,"链段"是指分子的节段或区域,例如,向导多核苷酸中的核苷酸的连续延伸。链段也可以指复合物的区域/节段,使得链段可以包含超过一个分子的区域。例如,若向导多核苷酸包含多个核酸分子,则蛋白结合链段可以包括多个独立分子的全部或部分,这些独立分子例如沿着互补性区域杂交。在一些实施方案中,包含两个独立分子的靶向DNA的RNA的蛋白结合链段可以包含:(i)长度为100个碱基对的第一RNA分子的碱基对40至75;和(ii)长度为50个碱基对的第一RNA分子的碱基对10至25。除非在特定语境中具体地另做定义,否则"链段"的定义不限于具体数目的总碱基对,不限于来自给定RNA分子的任何特定数目的碱基对,不限于复合物内的特定数目的独立分子,并且可以包括任何总长度的RNA分子的区域并且可以包括与其他分子具有互补性的区域。

[0649] 向导RNA或向导多核苷酸可以包含两个或更多个RNA,例如,CRISPR RNA(crRNA)和反向激活crRNA(tracrRNA)。向导RNA或向导多核苷酸有时可以包含单链RNA,或通过融合crRNA与tracrRNA的部分(例如,功能性部分)形成的单向导RNA(sgRNA)。向导RNA或向导多核苷酸也可以是包含crRNA和tracrRNA的双RNA。此外,crRNA可以与靶标DNA杂交。

[0650] 如上所述,向导RNA或向导多核苷酸可以是表达产物。例如,编码向导RNA的DNA可以是包含编码向导RNA的序列的载体。可以通过用分离的向导RNA或包含编码向导RNA的序列和启动子的质粒DNA转染细胞而将向导RNA或向导多核苷酸转移到该细胞内。也可以通过其他途径诸如使用病毒介导的基因递送将向导RNA或向导多核苷酸转移到细胞内。

[0651] 向导RNA或向导多核苷酸可以是经分离的。例如,可以将向导RNA以经分离的RNA的形式转移到细胞或生物体内。向导RNA可以使用本领域中已知的体外转录系统通过体外转录制备。可以将向导RNA以经分离的RNA的形式而不是以包含向导RNA的编码序列的质粒的形式转移到细胞内。

[0652] 向导RNA或向导多核苷酸可以包含三个区域:位于5'末端的可以与染色体序列中的靶标位点互补的第一区域、可以形成茎环结构的第二内部区域、和可以为单链的第三3'区域。每个向导RNA的第一区域也可以是不同的,使得每个向导RNA将融合蛋白引导至特定的靶标位点。再者,在全部向导RNA中,每个向导RNA的第二区域和第三区域可以是相同的。

[0653] 向导RNA或向导多核苷酸的第一区域可以与位于染色体序列中的靶标位点处的序列互补,适当向导RNA的第一区域可以与靶标位点进行碱基配对。在一些情况下,向导RNA的

第一区域可以包含或包含约10个核苷酸至25个核苷酸(即,10个核苷酸至核苷酸;或约10个核苷酸至约25个核苷酸;或10个核苷酸至约25个核苷酸;或约10个核苷酸至25个核苷酸)或更多。例如,在向导RNA的第一区域与染色体序列的靶标位点之间进行碱基配对的区域可以是或可以是约10、11、12、13、14、15、16、17、18、19、20、22、23、24、25或更多个核苷酸的长度。有时,向导RNA的第一区域可以是或可以是约19、20或21个核苷酸的长度。

[0654] 向导RNA或向导多核苷酸也可以包含形成二级结构的第二区域。例如,由向导RNA形成的二级结构可以包含茎(或发夹)和环。环和茎的长度可变。例如,环的长度范围可以是或可以是约3至10个核苷酸,并且茎的长度范围可以是或可以是约6至20个碱基对。茎可以包含一个或多个1至10个或约10个核苷酸的凸起部。第二区域的总体长度可以在或可以在约16至60个核苷酸的长度范围内。例如,环可以是或可以是约4个核苷酸的长度,并且茎可以是或可以是约12个碱基对。

[0655] 向导RNA或向导多核苷酸也可以包含位于3'末端的第三区域,该第三区域可以基本上是单链的。例如,第三区域有时不与感兴趣的细胞中的任何染色体序列互补,而有时不与向导RNA的剩余部分互补。再者,第三区域的长度可变。第三区域可以是超过或是超过约4个核苷酸的长度。例如,第三区域的总体长度可以在或可以在约5至60个核苷酸的长度范围内。

[0656] 向导RNA或向导多核苷酸可以靶向基因靶标的任何外显子或内含子。在一些情况下,向导可以靶向基因的外显子1或2;在其他情况下,向导可以靶向基因的外显子3或4。组合物可以包含多个全部靶向相同外显子的RNA,或者在一些情况下,包含多个可以靶向不同外显子的RNA。基因的外显子和内含子可以被靶向。

[0657] 向导RNA或向导多核苷酸可以靶向(约)20个核苷酸的核酸序列。靶标核酸可以少于(约)20个核苷酸。靶标核酸可以是至少(约)5、10、15、16、17、18、19、20、21、22、23、24、25、30个核苷酸或1至100个核苷酸之间任一处的长度。靶标核酸可以是至多(约)5、10、15、16、17、18、19、20、21、22、23、24、25、30、40、50个核苷酸或1至100个核苷酸之间任一处的长度。靶标核酸序列可以是紧邻PAM的5'第一个核苷酸的(约)20个碱基。向导RNA可以靶向核酸序列。靶标核酸可以是至少(约)1至10、1至20、1至30、1至40、1至50、1至60、1至70、1至80、1至90或1至100个核苷酸。

[0658] 向导多核苷酸,例如,向导RNA,可以指能与细胞基因组中的另一核酸(例如,靶标核酸或前间隔序列)杂交的核酸。向导多核苷酸可以是RNA。向导多核苷酸可以是DNA。向导多核苷酸可以被编程或设计为位点特异性地结合至核酸序列。向导多核苷酸可以包含一个多核苷酸链并且可以称为单向导多核苷酸。向导多核苷酸可以包含两个多核苷酸链并且可以称为双向导多核苷酸。向导RNA可以作为RNA分子被引入细胞内。例如,RNA分子可以在体外被转录和/或化学合成。RNA可以从合成的DNA分子(例如,gBlocks®基因片段)转录。然后,向导RNA可以作为RNA分子被引入细胞中。也可以将向导RNA以非RNA核酸分子(例如,DNA分子)的形式引入细胞内。例如,可以将编码向导RNA的DNA可操作地连接至启动子控制序列,以在感兴趣的细胞内进行向导RNA的表达。可以将RNA编码序列可操作地连接至被RNA聚合酶III(Pol III)识别的启动子序列。可用于表达向导RNA的质粒载体包括但不限于,px330载体和px333载体。在一些情况下,质粒载体(例如,px333载体)可以包含至少两个编码向导RNA的DNA序列。

[0659] 用于选择、设计和验证向导多核苷酸(例如,向导RNA和靶向序列)的方法在本文中描述并且是本领域技术人员已知的。例如,为了最小化核碱基编辑器系统中脱氨酶结构域(例如,AID结构域)的潜在底物混杂性的影响,可以最小化可能无意中被脱氨反应所靶向的残基(例如,可能潜在地驻留在靶标核酸基因座内的ssDNA上的脱靶C残基)的数目。此外,可以使用软件工具来优化对应于靶标核酸序列的gRNA,例如,以最小化跨基因座的总脱靶活性。例如,对于使用酿脓链球菌Cas9的每种可能的靶向结构域选择,可以跨基因组鉴定全部脱靶序列(先前选定的PAM,例如,NAG或NGG),该序列含有多达某一数目(例如,1、2、3、4、5、6、7、8、9或10个)的误配碱基对。可以鉴定与靶标位点互补的gRNA的第一区域,并且可以根据预测的总脱靶得分将全部第一区域(例如,crRNA)排序;排序靠前的靶向结构域代表那些可能具有最大上靶活性和最小脱靶活性的结构域。可以使用本领域已知和/或如本文详述的方法对候选靶向gRNA进行功能性评估。

[0660] 作为非限制性示例,用于与Cas9合用的向导RNA的crRNA中的靶标DNA杂交序列可以使用DNA序列检索算法鉴定,可以使用自定义gRNA设计软件基于公共工具cas-offinder完成gRNA设计,如Bae S.,Park J.,&Kim J.-S.Cas-OFFinder:A fast and versatile algorithm that searches for potential off-target sites of Cas9RNA-guided endonucleases.Bioinformatics 30,1473-1475(2014)中所述。这一软件在计算向导的全基因组脱靶倾向性之后对向导进行评分。典型地,对于17到24长度的向导范围,考虑从完美匹配到存在7个误配的匹配范围。一旦通过计算确定了脱靶位点,就计算每个向导的聚集得分并使用web界面将其汇总到表格输出中。除了鉴定邻近PAM序列的潜在靶标位点之外,该软件还鉴定全部PAM邻近序列,这些序列与所选择的靶标位点相异1、2、3或超过3个核苷酸。可以使用公共可获得的总结例如RepeatMasker程序,获得靶标核酸序列(例如,靶标基因)的基因组DNA序列和重复元件。RepeatMasker检索输入DNA序列的重复元件和低复杂度区域。输出是给定查询序列中存在的重复序列的详细注释。

[0661] 鉴定之后,可以将向导RNA(例如,crRNA)的第一区域基于以下各项排序为几个层次:它们与靶标位点的距离、它们的正交性和用于与相关PAM序列密切匹配的5'核苷酸(例如,基于对含有相关PAM(例如,对于酿脓链球菌是NGG PAM,对于金黄色葡萄球菌是NNGRRT或NNGRRV PAM)的人类基因组中的密切匹配的鉴定的5'G)的存在。如本文所用,正交性是指人类基因组中含有最低数目的与靶标序列的误配的序列的数目。例如,"高水平的正交性"或"良好正交性"可以指20聚体靶向结构域,该结构域除了预期靶标之外没有人类基因组中的相同序列,也不存在任何含有靶标序列中一个或两个误配的序列。可以选择具有良好正交性的靶向结构域以最小化脱靶DNA裂解。

[0662] 在一些实施方案中,可以使用报告系统来检测碱基编辑活性和测试候选向导多核苷酸。在一些实施方案中,报告系统可以包含基于报告基因的测定,其中碱基编辑活性导致该报告基因的表达。例如,报告系统可以包括包含经灭活的启动密码子(例如,在模板链上的3'-TAC-5'到3'-CAC-5'的突变)的报告基因。当对靶标C成功脱氨基后,相应mRNA将被转录为5'-AUG-3'而不是5'-GUG-3',使得报告基因能够翻译。合适的报告基因对于本领域技术人员将是显而易见的。报告基因的非限制性示例包括编码绿色荧光蛋白(GFP)、红色荧光蛋白(RFP)、荧光素酶、分泌的碱性磷酸酯酶(SEAP)的基因,或其表达可检测并且对于本领域技术人员显而易见的任何其他基因。报告系统可用来测试多种不同的gRNA,例如,以便确

定各脱氨酶将以靶标DNA序列的哪个残基为靶标。也可以测试靶向非模板链的sgRNA,以便评估特定碱基编辑蛋白(例如,Cas9脱氨酶融合蛋白)的脱靶效应。在一些实施方案中,可以设计此类gRNA,使得突变的起始密码子将不与gRNA进行碱基配对。向导多核苷酸可以包含标准核糖核苷酸、经修饰的核糖核苷酸(例如,假尿苷)、核糖核苷酸异构体和/或核糖核苷酸类似物。在一些实施方案中,向导多核苷酸可以包含至少一个可检测的标记物。可检测的表及其可以是荧光团(例如,FAM、TMR、Cy3、Cy5、德克萨斯红、俄勒冈绿、Alexa Fluors、Halo标签或合适的荧光染料)、检测标签(例如,生物素、地高辛等)、量子点或金粒子。

[0663] 向导多核苷酸可以互相合成、酶促合成或经其组合合成。例如,向导RNA可以使用基于亚磷酰胺的标准固相合成方法合成。作为另一种选择,向导RNA可以通过将编码向导RNA的DNA可操作地连接至被噬菌体RNA聚合酶识别的启动子控制序列而体外合成。合适的是具体启动子序列的示例包括T7、T3、SP6启动子序列或其变种。在其中向导RNA包含两个独立分子(例如,crRNA和tracrRNA)的实施方案中,crRNA可以化学地合成,而tracrRNA可以酶促地合成。

[0664] 在一些实施方案中,碱基编辑器系统可以包含多个向导多核苷酸,例如,gRNA。例如,gRNA可以靶向碱基编辑器系统中的一个或多个靶标基因座(例如,至少1个gRNA、至少2个gRNA、至少5个gRNA、至少10个gRNA、至少20个gRNA、至少30g RNA、至少50个gRNA)。多个gRNA序列可以随机排列,并且优选通过直接重复序列分隔。

[0665] 编码向导RNA(gRNA)或向导多核苷酸的DNA序列也可以是载体的一部分。再者,载体可以包含额外的表达控制序列(例如,增强子序列、Kozak序列、聚腺苷酸化序列、转录终止序列等)、可选择的标志物序列(例如,GFP或抗生素耐药基因诸如嘌呤霉素)、复制起点等。编码向导RNA的DNA分子可以是线性的。编码向导RNA(gRNA)或向导多核苷酸的DNA分子也可以是环状的。

[0666] 在一些实施方案中,碱基编辑器系统的一个或多个组分可以由DNA序列编码。此类DNA序列可以被一起或单独地引入表达系统(例如,细胞)中。例如,编码多核苷酸可编程核苷酸结合结构域和向导RNA的DNA序列可以被引入细胞中,每个DNA序列可以是独立分子(例如,一个含有多核苷酸可编程核苷酸结合结构域编码序列的载体和含有向导RNA编码序列的第二载体)的一部分,或者两者都可以是同一分子(例如,一个含有用于多核苷酸可编程核苷酸结合结构域和向导RNA两者的编码(和调节)序列的载体)的一部分。

[0667] 向导多核苷酸可以包含一种或多种修饰以向核酸提供新的或增强的特征。向导多核苷酸可以包含核酸亲和性标签。向导多核苷酸可以包含合成核苷酸、合成核苷酸类似物、核苷酸衍生物和/或经修饰的核苷酸。

[0668] 在一些情况下,gRNA或向导多核苷酸可以包含修饰。修饰可以在gRNA或向导多核苷酸的任意位置处做出。可以对单gRNA或向导多核苷酸做出超过一种修饰。gRNA或向导多核苷酸可以在修饰后经历质量控制。在一些情况下,质量控制可以包括PAGE、HPLC、MS或其任何组合。

[0669] gRNA或向导多核苷酸的修饰可以是置换、插入、缺失、化学修饰、物理修饰、稳定化、纯化或其任何组合。

[0670] gRNA或向导多核苷酸也可以通过以下进行修饰:5'腺苷酸化、5'鸟苷-三磷酸酯封端、5'N7-甲基鸟苷-三磷酸酯封端、5'三磷酸酯封端、3'磷酸酯、3'硫代磷酸酯、5'粒氨酸

酯、5’硫代磷酸酯、Cis-Syn胸腺嘧啶二聚体、三聚体、C12间隔序列、C3间隔序列、C6间隔序列、dSpacer、PC间隔序列、rSpacer、间隔序列18、间隔序列9、3’-3’修饰、5’-5’修饰、无碱基、吖啶、偶氮苯、生物素、生物素BB、生物素TEG、胆甾醇基TEG、脱硫生物素TEG、DNP TEG、DNP-X、DOTA、dT-生物素、双生物素、PC生物素、补骨脂素C2、补骨脂素C6、TINA、3’DABCYL、黑洞淬灭剂1、黑洞淬灭剂2、DABCYL SE、dT-DABCYL、IRDye QC-1、QSY-21、QSY-35、QSY-7、QSY-9、羧基连接子、硫醇连接子、2’-脱氧核糖核苷类似物嘌呤、2’-脱氧核糖核苷类似物嘧啶、核糖核苷类似物、2’-O-甲基核糖核苷类似物、糖修饰的类似物、摆动/通用碱基、荧光染料标记物、2’-氟RNA、2’-O-甲基RNA、甲基磷酸酯、磷酸二酯DNA、磷酸二酯RNA、硫代磷酸酯DNA、硫代磷酸酯RNA、UNA、假尿苷-5’-三磷酸酯、5’-甲基胞苷-5’-三磷酸酯或其任何组合。

[0671] 在一些情况下,修饰是永久的。在其他情况下,修饰是暂时的。在一些情况下,对gRNA或向导多核苷酸做出多种修饰。gRNA或向导多核苷酸修饰可以改变核苷酸的生理化学性质,诸如它们的构象、极性、疏水性、化学反应性、碱基配对相互作用或其任何组合。

[0672] PAM序列可以是本领域中已知的任何PAM序列。合适的PAM序列包括但不限于,NGG、NGA、NGC、NGN、NGT、NGCG、NGAG、NGAN、NGNG、NGCN、NGCG、NGTN、NNGRRT、NNNRRT、NNGRR(N)、TTTV、TYCV、TYCV、TATV、NNNNGATT、NNAGAAW或NAAAAC。Y是嘧啶;N是任何核苷酸碱基;W是A或T。

[0673] 修饰也可以是硫代磷酸酯替代物。在一些情况下,天然磷酸二酯键可能易受细胞核酸酶造成的快速降解;并且,使用硫代磷酸酯(PS)键替代物对核苷酸间链接进行修饰可能对于通过细胞降解进行的水解而言更为稳定。修饰可以增加gRNA或向导多核苷酸中的稳定性。修饰也可以增强生物学活性。在一些情况下,经硫代磷酸酯增强的RNA gRNA可以抑制RNase A、RNase T1、牛血清核酸酶或其任何组合。这些性质可以运行将PS-RNA gRNA用于在体内或体外暴露于核酸酶的可能性很高的应用中。例如,硫代磷酸酯(PS)键可以被引入到位于gRNA的5’-或‘’-末端的最末端3-5个核苷酸之间,这可以抑制核酸外切酶降解。在一些情况下,硫代磷酸酯键可以添加到整个gRNA中的任何位置处以减少被核酸内切酶攻击。

[0674] 前间隔序列邻近基序

[0675] 术语“前间隔序列邻近基序(protospacer adjacent motif,PAM)”或PAM样基序是指紧跟在被CRISPR细菌适应性免疫系统中的Cas9核酸酶所靶向的DNA序列会后的2-6个碱基对DNA序列。在一些实施方案中,PAM可以是5’PAM(即,定位在前间隔序列的5’末端的上游)。在其他实施方案中,PAM可以是5’PAM(即,定位在前间隔序列的3’末端的下游)。

[0676] PAM序列对于靶标结合而言是必不可少的,但确切的序列取决于Cas蛋白的类型。

[0677] 本文所提供的碱基编辑器可以包含衍生自CRISPR蛋白的结构域,其能够结合含有经典的或非经典的前间隔序列邻近基序(PAM)序列的核苷酸序列。PAM位点是接近多核苷酸序列的核苷酸序列。本公开的一些方面提供碱基编辑器,这些碱基编辑器包含具有不同PAM特异性的CRISPR蛋白质的全部或部分。例如,典型地,Cas9蛋白诸如来自酿脓链球菌的Cas9(spCas9)需要经典的NGG PAM序列来结合特定的核酸区域,其中“NGG”中的“N”是腺嘌呤(A)、胸腺嘧啶(T)、鸟嘌呤(G)或胞嘧啶(C),并且G是鸟苷。PAM可以是CRISPR蛋白特异性的,并且在包含不同的衍生自CRISPR蛋白的结构域的不同碱基编辑器之间可以是不同的。PAM可以位于靶标序列的5’或3’。PAM可以在靶标序列的上游或下游。PAM可以是1、2、3、4、5、6、7、8、9、10个或更多个核苷酸的长度。通常,PAM是2至6个核苷酸之间的长度。几种PAM变体在

下表1中描述。

[0678] 表1.Cas9蛋白和相应PAM序列

[0679]

变体	PAM
spCas9	NGG
spCas9-VRQR	NGA
spCas9-VRER	NGCG
spCas9-MQKFRAER	NGC
xCas9 (sp)	NGN

[0680]

变体	PAM
saCas9	NNGRRT
saCas9-KKH	NNNRRT
spCas9-MQKSER	NGCG
spCas9-MQKSER	NGCN
spCas9-LRKIQK	NGTN
spCas9-LRVSQK	NGTN
spCas9-LRVSQL	NGTN
SpyMacCas9	NAA
Cpf1	5’ (TTTV)

[0681] 在一些实施方案中,PAM是NGC。在一些实施方案中,NGC PAM被Cas9变体识别。在一些实施方案中,NGC PAM变体包括选自下列的一种或多种氨基酸置换:D1135M、S1136Q、G1218K、E1219F、A1322R、D1332A、R1335E和T1337R(统称“MQKFRAER”)。

[0682] 在一些实施方案中,PAM是NGT。在一些实施方案中,NGT PAM被Cas9变体识别。在一些实施方案中,通过在一个或多个残基1335、1337、1135、1136、1218和/或1219处的靶向突变来产生NGT PAM变体。在一些实施方案中,通过在一个或多个残基1219、1335、1337、1218处的靶向突变来产生NGT PAM变体。在一些实施方案中,通过在一个或多个残基1135、1136、1218、1219和1335处的靶向突变来产生NGT PAM变体。在一些实施方案中,NGT PAM变体选自下表2和表3中提供的靶向突变的集合。

[0683] 表2:位于残基1219、1335、1337、1218处的NGT PAM变体突变

[0684]

变体	E1219V	R1335Q	T1337	G1218
1	F	V	T	
2	F	V	R	

[0685]

变体	E1219V	R1335Q	T1337	G1218
3	F	V	Q	
4	F	V	L	
5	F	V	T	R
6	F	V	R	R
7	F	V	Q	R
8	F	V	L	R
9	L	L	T	
10	L	L	R	
11	L	L	Q	
12	L	L	L	
13	F	I	T	
14	F	I	R	
15	F	I	Q	
16	F	I	L	
17	F	G	C	
18	H	L	N	
19	F	G	C	A
20	H	L	N	V
21	L	A	W	
22	L	A	F	
23	L	A	Y	
24	I	A	W	
25	I	A	F	
26	I	A	Y	

[0686] 表3:位于残基1135、1136、1218、1219和1335处的NGT PAM变体突变

[0687]

变体	D1135L	S1136R	G1218S	E1219V	R1335Q
27	G				

变体	D1135L	S1136R	G1218S	E1219V	R1335Q
28	V				
29	I				
30		A			
31		W			
32		H			
33		K			
34			K		
35			R		
36			Q		
37			T		
38			N		
39				I	
40				A	
41				N	
42				Q	
43				G	
44				L	
45				S	
46				T	
47					L
48					I
49					V
50					N
51					S
52					T
53					F
54					Y
55	N1286Q	I1331F			

[0688]

[0689] 在一些实施方案中,NGT PAM变体选自表2和表3中的变体5、7、28、31或36。在一些

实施方案中,变体具有改善的NGT PAM识别。

[0690] 在一些实施方案中,NGT PAM变体具有位于残基1219、1335、1337和/或1218处的突变。在一些实施方案中,从下表4中提供的变体中选择具有用于改善的识别的突变的NGT PAM变体。

[0691] 表4:位于残基1219、1335、1337和1218处的NGT PAM变体突变

[0692]

变体	E1219V	R1335Q	T1337	G1218
1	F	V	T	
2	F	V	R	
3	F	V	Q	
4	F	V	L	
5	F	V	T	R
6	F	V	R	R
7	F	V	Q	R
8	F	V	L	R

[0693] 在一些实施方案中,从下表5中提供的变体中选择NGT PAM。

[0694] 表5.NGT PAM变体

[0695]

	NGTN变体	D1135	S1136	G1218	E1219	A1322R	R1335	T1337
变体1	LRKIQK	L	R	K	I	-	Q	K
变体2	LRSVQK	L	R	S	V	-	Q	K
变体3	LRSVQL	L	R	S	V	-	Q	L
变体4	LRKIRQK	L	R	K	I	R	Q	K
变体5	LRSVRQK	L	R	S	V	R	Q	K
变体6	LRSVRQL	L	R	S	V	R	Q	L

[0696] 在一些实施方案中,NGTN变体是变体1。在一些实施方案中,NGTN变体是变体2。在一些实施方案中,NGTN变体是变体3。在一些实施方案中,NGTN变体是变体4。在一些实施方案中,NGTN变体是变体5。在一些实施方案中,NGTN变体是变体6。

[0697] 在一些实施方案中,Cas9结构域是来自酿脓链球菌(Streptococcus pyogenes)的Cas9结构域(SpCas9)。在一些实施方案中,SpCas9结构域是核酸酶活性SpCas9、无核酸酶活性的SpCas9(SpCas9d)或SpCas9切口酶(SpCas9n)。在一些实施方案中,SaCas9包含N579A突变,或本文所提供的任何氨基酸序列中的相应突变,其中X是除D以外的任何氨基酸。在一些实施方案中,SpCas9包含D9A突变,或本文所提供的任何氨基酸序列中的相应突变。在一些实施方案中,SpCas9结构域、SpCas9d结构域或SpCas9n结构域可以结合至具有非经典PAM的核酸序列。在一些实施方案中,SpCas9结构域、SpCas9d结构域或SpCas9n结构域可以结合至具有NGG、NGA或NGCG PAM的核酸序列。

[0698] 在一些实施方案中,SpCas9结构域包含D1135X、R1335X和T1337X突变中的一种或多种,或本文所提供的任何氨基酸序列的相应突变,其中X是任何氨基酸。在一些实施方案中,SpCas9结构域包含D1135E、R1335Q和T1337R突变中的一种或多种,或本文所提供的任何氨基酸序列的相应突变。在一些实施方案中,SpCas9结构域包含D1135E、R1335Q和T1337R突变,或本文所提供的任何氨基酸序列的相应突变。在一些实施方案中,SpCas9结构域包含

D1135X、R1335X和T1337X突变中的一种或多种,或本文所提供的任何氨基酸序列的相应突变,其中X是任何氨基酸。在一些实施方案中,SpCas9结构域包含D1135V、R1335Q和T1337R突变中的一种或多种,或本文所提供的任何氨基酸序列的相应突变。在一些实施方案中,SpCas9结构域包含D1135V、R1335Q和T1337R突变,或本文所提供的任何氨基酸序列的相应突变。在一些实施方案中,SpCas9结构域包含D1135X、G1218X、R1335X和T1337X突变中的一种或多种,或本文所提供的任何氨基酸序列的相应突变,其中X是任何氨基酸。在一些实施方案中,SpCas9结构域包含D1135V、G1218R、R1335Q和T1337R突变中的一种或多种,或本文所提供的任何氨基酸序列的相应突变。在一些实施方案中,SpCas9结构域包含D1135V、G1218R、R1335Q和T1337R突变,或本文所提供的任何氨基酸序列的相应突变。

[0699] 在一些实施方案中,本文所提供的任何融合蛋白的Cas9结构域包含氨基酸序列,所述氨基酸序列与本文所述的Cas9多肽为至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相同。在一些实施方案中,本文所提供的任何融合蛋白的Cas9结构域包含本文所述的任何Cas9多肽的氨基酸序列。在一些实施方案中,本文所提供的任何融合蛋白的Cas9结构域由本文所述的任何Cas9多肽的氨基酸序列组成。

[0700] 在一些实施例中,被本文所公开的碱基编辑器的衍生自CRISPR蛋白的结构域识别的PAM可在独立于编码碱基编辑器的插入物(例如,AAV插入物)的寡核苷酸上提供至细胞。在此类实施方案中,提供位于独立的寡核苷酸上的PAM可以允许靶标序列的裂解,该靶标序列不能以其他方式裂解,因为同一多核苷酸上不存在作为靶标序列的邻近PAM。

[0701] 在一种实施方案中,酿脓链球菌Cas9(SpCas9)可以用作CRISPR核酸内切酶以进行基因组工程化。但是,可以使用其他的。在一些实施方案中,不同的核酸内切酶可以用来靶向某些基因组靶标。在一些实施方案中,可以使用具有非NGG PAM序列的衍生自SpCas9的合成变体。此外,已经鉴定了来自多个物种的其他Cas9直系同源物,并且这些“非SpCas9”可以结合也能用于本公开的多种PAM序列。例如,相对大尺寸的SpCas9(大约4kb编码序列)可能导致质粒携带不能在细胞中有效表达的SpCas9 cDNA。相反,金黄色葡萄球菌Cas9(SaCas9)的编码序列比SpCas9短大约1千碱基,可能使得它在细胞中被有效地表达。类似于SpCas9,SaCas9核酸内切酶能够在体外修饰哺乳动物细胞中的靶标基因,并且在小鼠体内修饰细胞中的靶标基因。在一些实施方案中,Cas蛋白可以靶向不同的PAM序列。在一些实施方案中,靶标基因可以邻近Cas9 PAM,例如,5’-NGG。在其他实施方案中,其他Cas9直系同源物可能具有不同的PAM要求。例如,其他PAM诸如嗜热链球菌(S.thermophilus)的PAM(对于CRISPR1是5’-NNAGAA,而对于CRISPR3是5’-NGGNG)和脑膜炎奈瑟氏菌(Neisseria meningiditis)的PAM(5’-NNNNGATT)也可以被发现邻近靶标基因。

[0702] 在一些实施方案中,对于酿脓链球菌系统,靶标基因序列可能位于5’-NGG PAM之前(即,位于其5),并且20-nt向导RNA序列可以与相对链碱基配对以介导邻近PAM的Cas9裂解。在一些实施方案中,邻近切割可以在PAM上游的(约)3个碱基对处。在一些实施方案中,邻近切割可以在PAM上游的(约)10个碱基对处。在一些实施方案中,邻近切割可以在PAM上游的(约)0至20个碱基对处。例如,邻近切割可以在PAM上游的1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29或30个碱基对旁边。邻近切割也可以在PAM下游1至30个碱基对处。能够结合PAM序列的示例性SpCas9蛋白的序列如下:

[0703] 示例性PAM结合SpCas9的氨基酸序列如下：

[0704] MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA
RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVD
STDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSR
RLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNL
SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNEL
TKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMG
RHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAE
RGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGE
IRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYG
GFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELEN
GRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD
ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI
DLSQLGGD。

[0705] 示例性PAM结合SpCas9n的氨基酸序列如下：

[0706] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA
RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVD
STDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSR
RLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNL
SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNEL
TKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMG
RHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAE
RGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGE
IRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYG
GFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELEN
GRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD
ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI

DLSQLGGD。

[0707] 示例性PAM结合SpEQR Cas9的氨基酸序列如下：

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR
RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKK
LVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAI
LSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQ
IGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIF
FDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHA
ILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSF
IERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTV
KQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREM
[0708] IEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDS
LTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQ
KGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHI
VPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL
DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNY
HHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA
RKKDWDPKKYGGF**E**SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVK
KDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ
HKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTI
DRK**Q**Y**R**STKEVLDATLIHQSITGLYETRIDLSQLGGD。在该序列中,可能从D1135、R1335和T1337突变以得到SpEQR Cas9的残基E1135、Q1335和R1337加下划线并以粗体显示。

[0709] 示例性PAM结合SpVQR Cas9的氨基酸序列如下：

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR
RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRK
KLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKA
ILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLA
QIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQS
FIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVT
VKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRE
[0710] MIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTT
QKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDH
IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE
LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLI
ARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV
KKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE
QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTT
IDRKQYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD。在该序列中,可能从D1135、R1335和T1337突变以得到SpVQR Cas9的残基V1135、Q1335和R1337加下划线并以粗体显示。

[0711] 示例性PAM结合SpVRER Cas9的氨基酸序列如下:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR
RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRK
KLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKA
ILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLA
[0712] QIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQS
FIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVT
VKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRE
MIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTT
QKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDH
IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE
LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN

[0713] YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLI
ARKKDWDPKKYGGF**V**SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV
KKDLIIKLPKYSLFELENGRKRMLASA**R**ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE
QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTT
IDRK**E**Y**R**STKEVLDATLIHQSITGLYETRIDLSQLGGD。在上述序列中,可能从D1134、G1217、R1335和T1337突变以得到SpVRER Cas9的残基V1135、R1218、Q1335和R1337加下划线并以粗体显示。

[0714] 在一些实施方案中,Cas9结构域是重组Cas9结构域。在一些实施方案中,重组Cas9结构域是SpyMacCas9结构域。在一些实施方案中,SpyMacCas9结构域是核酸酶活性SpyMacCas9、无核酸酶活性的SpyMacCas9(SpyMacCas9d)或SpyMacCas9切口酶(SpyMacCas9n)。在一些实施方案中,SaCas9结构域、SaCas9d结构域或SaCas9n结构域可以结合至具有非经典PAM的核酸序列。在一些实施方案中,SpyMacCas9结构域、SpCas9d结构域或SpCas9n结构域可以结合至具有NAA PAM序列的核酸序列。

[0715] 示例性SpyMacCas9

[0716] MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETAEATRLKRTA
RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLAD
STDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQIYNQLFEENPINASRVDAKAILSARLSKSR
RLENLIAQLPGEKRNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNL
SDAILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNEL
TKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDRGMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGHSLHEQIANLAGSPAIKKGILQTVKIVDELVKVMG
HKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELD
INRLSDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAER
GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNY
HHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI
RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEIQTVGQNGGLFDDNPKSPLEVTPSKLVPLKKELN
PKKYGGYQKPTTAYPVLLITDTKQLIPISVMNKKQFEQNPVKFLRDRGYQQVGKNDFIKLPKYTLVDIGDGIKRLW
ASSKEIHKGNQLVVSKKSQILLYHAHHLDSDLSNDYLQNHNQQFDVLFNEIISFSKKCKLGKEHIQKIENVYSNKK
NSASIEELAESFIKLLGFTQLGATSPFNFLGVKLNQKQYKGKKDYILPCTEGTLIRQSITGLYETRVDLSKIGED。

[0717] 在一些情况下,变体Cas9蛋白荷有H840A、P475A、W476A、N477A、D1125A、W1126A和

D1218A突变,使得该多肽裂解靶标DNA或RNA的能力降低。此类Cas9蛋白裂解靶标DNA(例如,单链靶标DNA)的能力降低,但保留结合靶标DNA(例如,单链靶标DNA)的能力。作为另一种非限制性示例,在一些情况下,变体Cas9蛋白荷有D10A、H840A、P475A、W476A、N477A、D1125A、W1126A和D1218A突变,使得该多肽裂解靶标DNA的能力降低。此类Cas9蛋白裂解靶标DNA(例如,单链靶标DNA)的能力降低,但保留结合靶标DNA(例如,单链靶标DNA)的能力。在一些情况下,当变体Cas9蛋白荷有W476A和W1126A突变时或当变体Cas9蛋白荷有P475A、W476A、N477A、D1125A、W1126A和D1218A突变时,变体Cas9蛋白不与PAM序列有效地结合。因此,在一些此类情况下,当在结合方法中使用此类变体Cas9蛋白时,该方法不需要PAM序列。换言之,在一些情况下,当在接恶化方法中使用此类变体Cas9蛋白时,该方法可以包括向导RNA,但该方法可以在不存在PAM序列的情况下执行(因此,由向导RNA的靶向链段提供结合的特异性)。可以突变其他残基以实现上述效果(即,将一个或另一股核酸酶部分灭活)。作为非限制性示例,可以改变(即,置换)残基D10、G12、G17、E762、H840、N854、N863、H982、H983、A984、D986和/或A987。而且,除丙氨酸置换之外的突变是合适的。

[0718] 在一些实施方案中,碱基编辑器的衍生自CRISPR蛋白的结构域可以包含具有经典PAM序列(NGG)的Cas9蛋白的全部或一部分。在其他实施方案中,碱基编辑器的衍生自Cas9的结构域可以采用非经典PAM序列。此类序列已经在本领域中有所描述并且对于技术人员将会是显而易见的。例如,结合非经典PAM序列的Cas9结构域已经在以下文献中有所描述:Kleinstiver,B.P.,et al.,"Engineered CRISPR-Cas9 nucleases with altered PAM specificities"Nature 523,481-485(2015);和Kleinstiver,B.P.,et al.,"Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition"Nature Biotechnology 33,1293-1298(2015);各自的整体内容通过引用并入本文。

[0719] 具有降低的排他性的Cas9结构域

[0720] 典型地,Cas9蛋白诸如来自酿脓链球菌的Cas9(spCas9)需要经典的NGG PAM序列来结合特定的核酸区域,其中"NGG"中的"N"是腺苷(A)、胸苷(T)或胞苷(C),并且G是鸟苷。这可能限制了编辑基因组内所希望的碱基的能力。在一些实施方案中,本文所提供的碱基编辑融合蛋白可能需要被放置在精确位置处,例如,在包含位于PAM上游的靶标碱基的区域。参见例如,Komor,A.C.,et al.,"Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage"Nature533,420-424(2016),其整体内容通过引用并入本文。据此,在一些实施方案中,本文所提供的任何融合蛋白可以含有能够结合不含经典(例如,NGG)PAM序列的核苷酸序列的Cas9结构域。结合至非经典PAM序列的Cas9结构域已经在本领域中有所描述并且对于技术人员将会是显而易见的。例如,结合非经典PAM序列的Cas9结构域已经在以下文献中有所描述:Kleinstiver,B.P.,et al.,"Engineered CRISPR-Cas9 nucleases with altered PAM specificities"Nature523,481-485(2015);Kleinstiver,B.P.,et al.,"Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition"NatureBiotechnology 33,1293-1298(2015);Nishimasu,H.,et al.,"Engineered CRISPR-Cas9 nuclease with expanded targeting space"Science.2018Sep21;361(6408):1259-1262;Chatterjee,P.,et al.,Minimal PAM specificity of a highly

similar SpCas9 ortholog"Sci Adv.2018Oct 24;4(10):eaau0766.doi:10.1126/sciadv.aau0766,各自的整体内容通过引用并入本文。

[0721] 包含Cas9结构域以及胞苷脱氨酶和/或腺苷脱氨酶的融合蛋白

[0722] 本公开的一些方面提供包含Cas9结构域或其他核酸可编程DNA结合蛋白以及一个或多个腺苷脱氨酶、胞苷脱氨酶和/或DNA糖苷酶结构域的融合蛋白。应知悉,Cas9结构域可以是本文所提供的任何Cas9结构域或Cas9蛋白(例如,dCas9或nCas9)。在一些实施方案中,本文所提供的任何Cas9结构域或Cas9蛋白(例如,dCas9或nCas9)可以与本文所提供的任何胞苷脱氨酶和腺苷脱氨酶融合。本文所公开的碱基编辑器的结构域可以任何次序排列。

[0723] 例如且不限于,在一些实施方案中,融合蛋白包含结构:

[0724] NH_2-[胞苷脱氨酶]-[Cas9结构域]-[腺苷脱氨酶]-COOH;

[0725] NH_2-[腺苷脱氨酶]-[Cas9结构域]-[胞苷脱氨酶]-COOH;

[0726] NH_2-[腺苷脱氨酶]-[胞苷脱氨酶]-[Cas9结构域]-COOH;

[0727] NH_2-[胞苷脱氨酶]-[腺苷脱氨酶]-[Cas9结构域]-COOH;

[0728] NH_2-[Cas9结构域]-[腺苷脱氨酶]-[胞苷脱氨酶]-COOH;或

[0729] NH_2-[Cas9结构域]-[胞苷脱氨酶]-[腺苷脱氨酶]-COOH。

[0730] 在一些实施方案中,融合蛋白的腺苷脱氨酶包含TadA*8和胞苷脱氨酶。在一些实施方案中,TadA*8是TadA*8.1、TadA*8.2、TadA*8.3、TadA*8.4、TadA*8.5、TadA*8.6、TadA*8.7、TadA*8.8、TadA*8.9、TadA*8.10、TadA*8.11、TadA*8.12、TadA*8.13、TadA*8.14、TadA*8.15、TadA*8.16、TadA*8.17、TadA*8.18、TadA*8.19、TadA*8.20、TadA*8.21、TadA*8.22、TadA*8.23或TadA*8.24。

[0731] 示例性融合蛋白结构包括下列:

[0732] NH_2-[腺苷脱氨酶]-[Cas9]-[胞苷脱氨酶]-COOH;

[0733] NH_2-[胞苷脱氨酶]-[Cas9]-[腺苷脱氨酶]-COOH;

[0734] NH_2-[TadA*8]-[Cas9]-[胞苷脱氨酶]-COOH;或

[0735] NH_2-[胞苷脱氨酶]-[Cas9]-[TadA*8]-COOH。

[0736] 在一些实施方案中,包含胞苷脱氨酶、无碱基编辑器、以及腺苷脱氨酶和napDNAbp(例如,Cas9结构域)的融合蛋白不包括连接子序列。在一些实施方案中,连接子存在于胞苷脱氨酶和腺苷脱氨酶结构域与napDNAbp之间。在一些实施方案中,上述通用结构中使用的"-"表示存在任选的连接子。在一些实施方案中,胞苷脱氨酶和腺苷脱氨酶与napDNAbp经由本文所提供的任何连接子融合。例如,在一些实施方案中,胞苷脱氨酶和/或腺苷脱氨酶与napDNAbp经由下面题为"连接子"的章节中所提供的任何连接子融合。

[0737] 在一些实施方案中,具有胞苷脱氨酶、腺苷脱氨酶、以及Cas9或Cas12结构域的例示性Cas9或Cas12融合蛋白的通用结构包含下列结构中的任何一种,其中NLS是核定位序列(例如,本文所提供的任何NLS),NH_2是融合蛋白的N端,而COOH是融合蛋白的C端.

[0738] NH_2-NLS-[胞苷脱氨酶]-[Cas9结构域]-[腺苷脱氨酶]-COOH;

[0739] NH_2-NLS-[腺苷脱氨酶]-[Cas9结构域]-[胞苷脱氨酶]-COOH;

[0740] NH_2-NLS-[腺苷脱氨酶]-[胞苷脱氨酶]-[Cas9结构域]-COOH;

[0741] NH_2-NLS-[胞苷脱氨酶]-[腺苷脱氨酶]-[Cas9结构域]-COOH;

[0742] NH_2-NLS-[Cas9结构域]-[腺苷脱氨酶]-[胞苷脱氨酶]-COOH;

[0743] NH_2-NLS-[Cas9结构域]-[胞苷脱氨酶]-[腺苷脱氨酶]-COOH;

[0744] NH_2-[胞苷脱氨酶]-[Cas9结构域]-[腺苷脱氨酶]-NLS-COOH;

[0745] NH_2-[腺苷脱氨酶]-[Cas9结构域]-[胞苷脱氨酶]-NLS-COOH;

[0746] NH_2-[腺苷脱氨酶]-[胞苷脱氨酶]-[Cas9结构域]-NLS-COOH;

[0747] NH_2-[胞苷脱氨酶]-[腺苷脱氨酶]-[Cas9结构域]-NLS-COOH;

[0748] NH_2-[Cas9结构域]-[腺苷脱氨酶]-[胞苷脱氨酶]-NLS-COOH;或

[0749] NH_2-[Cas9结构域]-[胞苷脱氨酶]-[腺苷脱氨酶]-NLS-COOH。

[0750] 在一些实施方案中,NLS存在于连接子中,或者NLS侧翼具有连接子,例如,本文所述。在一些实施方案中,NLS的N端或C端是二分NLS。二分NLS包含两个碱性氨基酸簇,这两个簇被相对较短的间隔序列分隔(因此,二分即2个部分,而单分NLS则不然)。核质蛋白的NLS,KR[PAATKKAGQA]KKKK,是普遍存在的二分信号的原型:两个碱性氨基酸簇,通过约10个氨基酸的间隔序列分隔。示例性二分NLS的序列如下:PKKKRKVEGADKRTADGSEFESPKKKRKV。

[0751] 在一些实施方案中,包含胞苷脱氨酶、腺苷脱氨酶、Cas9结构域和NLS的融合蛋白不包括连接子序列。在一些实施方案中,在结构域或蛋白(例如,胞苷脱氨酶、腺苷脱氨酶、Cas9结构域或NLS)中的一者或多者之间存在连接子。

[0752] 应知悉,本公开的融合蛋白可以包含一个或多个额外特征。例如,在一些实施方案中,融合蛋白可以包含抑制剂、细胞质定位序列、输出序列诸如核输出序列、或其他定位序列,以及可用于融合蛋白的溶解、纯化或检测的序列标签。本文所提供的合适的蛋白标签包括但不限于,生物素羧化酶载体蛋白(BCCP)标签、myc标签、钙调蛋白标签、FLAG标签、血细胞凝集素(HA)标签、聚组氨酸标签(也称为组氨酸标签或His标签)、麦芽糖结合蛋白(MBP)标签、nus标签、谷胱甘肽-S-转移酶(GST)标签、绿色荧光蛋白(GFP)标签、硫氧还蛋白标签、S标签、Softag(例如,Softag 1、Softag 3)、链霉素标签、生物素连接酶标签、FlAsH标签、V5标签和SBP标签。其他合适的序列对于本领域技术人员将是显而易见的。在一些实施方案中,融合蛋白包含一个或多个His标签。

[0753] 示例性而非限制性融合蛋白的细节在国际PCT申请号PCT/2017/044935和PCT/US2016/016288中描述,其各自通过引用而整体并入本文。

[0754] 包含核定位序列(NLS)的融合蛋白

[0755] 在一些实施方案中,本文所提供的融合蛋白进一步包含一个或多个(例如,2、3、4、5个)核靶向序列,例如,核定位序列(NLS)。在一种实施方案中,使用二分NLS。在一些实施方案中,NLS包含促进将包含NLS的蛋白质并入细胞核(例如,通过核转运)中的氨基酸序列。在一些实施方案中,本文所提供的任何融合蛋白进一步包含核定位序列(NLS)。在一些实施方案中,NLS融合至融合蛋白的N端。在一些实施方案中,NLS融合至融合蛋白的C端。在一些实施方案中,NLS融合至Cas9结构域的N端。在一些实施方案中,NLS融合至Cas9结构域或dCas9结构域的C端。在一些实施方案中,NLS融合至脱氨酶的N端。在一些实施方案中,NLS融合至脱氨酶的C端。在一些实施方案中,NLS经由一个或多个连接子融合至融合蛋白。在一些实施方案中,NLS融合至融合蛋白而没有连接子。在一些实施方案中,NLS包含本文所提供或引用的任何一种NLS序列的氨基酸序列。额外的核定位序列是本领域中已知的并且对于技术人员将会是显而易见的。例如,NLS序列在Plank等人的PCT/EP2000/011690中有所描述,其内容通过引用其对于示例性核定位序列的公开而并入本文。在一些实施方案中,NLS包含氨基

酸序列PKKKRKVEGADKRTADGSEFESPKKKRKV、KRTADGSEFESPKKKRKV、KRPAATKKAGQAKKKK、KKTELQTTNAENKTKKL、KRGINDRNFWRGENGRKTR、RKSGKIAAIVVKRPRKPKKKRKV或MDSLLMNRRKFLYQFKNVRWAKGRRETYLC。

[0756] 在一些实施方案中,NLS存在于连接子中,或者NLS侧翼具有连接子,例如,本文所述的连接子。在一些实施方案中,NLS的N端或C端是二分NLS。二分NLS包含两个碱性氨基酸簇,这两个簇被相对较短的间隔序列分隔(因此,二分即2个部分,而单分NLS则不然)。核质蛋白的NLS,KR[PAATKKAGQA]KKKK,是普遍存在的二分信号的原型:两个碱性氨基酸簇,通过约10个氨基酸的间隔序列分隔。示例性二分NLS的序列如下:

[0757] PKKKRKVEGADKRTADGSEFESPKKKRKV

[0758] 在一些实施方案中,融合蛋白不包含连接子序列。在一些实施方案中,存在位于一个或多个结构域或蛋白质之间的连接子序列。在一些实施方案中,具有腺苷脱氨酶或胞苷脱氨酶以及Cas9结构域的例示性Cas9融合蛋白的通用结构包含下列结构中的任何一种,其中NLS是核定位序列(例如,本文所提供的任何NLS),NH_2是融合蛋白的N端,而COOH是融合蛋白的C端:

[0759] NH_2-NLS-[腺苷脱氨酶]-[Cas9结构域]-COOH;

[0760] NH_2-NLS-[Cas9结构域]-[腺苷脱氨酶]-COOH;

[0761] NH_2-[腺苷脱氨酶]-[Cas9结构域]-NLS-COOH;

[0762] NH_2-[Cas9结构域]-[腺苷脱氨酶]-NLS-COOH;

[0763] NH_2-NLS-[胞苷脱氨酶]-[Cas9结构域]-COOH;

[0764] NH_2-NLS-[Cas9结构域]-[胞苷脱氨酶]-COOH;

[0765] NH_2-[胞苷脱氨酶]-[Cas9结构域]-NLS-COOH;或

[0766] NH_2-[Cas9结构域]-[胞苷脱氨酶]-NLS-COOH。

[0767] 应知悉,本公开的融合蛋白可以包含一个或多个额外特征。例如,在一些实施方案中,融合蛋白可以包含抑制剂、细胞质定位序列、输出序列诸如核输出序列、或其他定位序列,以及可用于融合蛋白的溶解、纯化或检测的序列标签。本文所提供的合适的蛋白标签包括但不限于,生物素羧化酶载体蛋白(BCCP)标签、myc标签、钙调蛋白标签、FLAG标签、血细胞凝集素(HA)标签、聚组氨酸标签(也称为组氨酸标签或His标签)、麦芽糖结合蛋白(MBP)标签、nus标签、谷胱甘肽-S-转移酶(GST)标签、绿色荧光蛋白(GFP)标签、硫氧还蛋白标签、S标签、Softag(例如,Softag 1、Softag 3)、链霉素标签、生物素连接酶标签、FlAsH标签、V5标签和SBP标签。其他合适的序列对于本领域技术人员将是显而易见的。在一些实施方案中,融合蛋白包含一个或多个His标签。

[0768] 可使用编码包含一个或多个核定位序列(NLS)的CRISPR酶的载体。例如,可存在(约)1、2、3、4、5、6、7、8、9、10个被使用的NLS。CRISPR酶可以包含位于氨基端处或附近的NLS,大约或超过大约1、2、3、4、5、6、7、8、9、10个位于羧基端处或附近的NLS,或这些的任何组合(例如,一个或多个位于氨基端处的NLS和一个或多个位于羧基端处的NLS)。当存在超过一个NLS时,每一个NLS可以独立于其他NLS而选择,使得单个NLS可以存在于超过一个拷贝中及/或与一个或多个其他NLS组合存在于一个或多个拷贝中。

[0769] 该方法中使用的CRISPR酶可以包含大约6个NLS。当最接近NLS的氨基酸处于沿着多肽链从N端或C端起约50个氨基酸内,例如,1、2、3、4、5、10、15、20、25、30、40或50个氨基酸

内时,则将NLS视为位于N端或C端附近。

[0770] 具有内部插入物的融合蛋白

[0771] 本文提供融合蛋白,该融合蛋白包含与核酸可编程核酸集合蛋白(例如,napDNAbp)融合的异源多肽。异源多肽可以是未见于天然或野生型napDNAbp多肽序列中的多肽。异源多肽可以在napDNAbp的C端、napDNAbp的N-端与napDNAbp融合,或者可以插入到napDNAbp的内部位置处。在一些实施方案中,异源多肽插入到napDNAbp的内部位置处。

[0772] 在一些实施方案中,异源多肽是脱氨酶或其功能性片段。例如,融合蛋白可包含位于Cas9或Cas12(例如,Cas12b/C2c1)多肽的N端片段和C端片段的两侧。融合蛋白中的脱氨酶可以是腺苷脱氨酶。在一些实施方案中,腺苷脱氨酶是TadA(例如,TadA7.10或TadA*8)。在一些实施方案中,TadA是TadA*8。本文所述的TadA序列(例如,TadA7.10或TadA*8)是用于上述融合蛋白的合适脱氨酶。

[0773] 脱氨酶可以是环状高突变脱氨酶。例如,脱氨酶可以是环状高突变脱氨酶。在一些实施方案中,脱氨酶是环状高突变TadA,在如TadA参考序列中编号的氨基酸残基116处被环状高突变。在一些实施方案中,脱氨酶是环状高突变TadA,在如TadA参考序列中编号的氨基酸残基136处被环状高突变。在一些实施方案中,脱氨酶是环状高突变TadA,在如TadA参考序列中编号的氨基酸残基65处被环状高突变。

[0774] 融合蛋白可以包含超过一种脱氨酶。融合蛋白可包含,例如,1、2、3、4、5或更多种脱氨酶。在一些实施方案中,融合蛋白包含一种脱氨酶。在一些实施方案中,融合蛋白包含两种脱氨酶。融合蛋白中的两种或更多种脱氨酶可以是腺苷脱氨酶、胞苷含脱氨酶或其组合。两种或更多种脱氨酶可以是同源二聚体。两种或更多种脱氨酶可以是异源二聚体。两种或更多种脱氨酶可以串联插入napDNAbp中。在一些实施方案中,两种或更多种脱氨酶可以不串联插入napDNAbp中。

[0775] 在一些实施方案中,融合蛋白中的napDNAbp是Cas9多肽或其片段。Cas9多肽可以是变体Cas9多肽。在一些实施方案中,Cas9多肽是Cas9切口酶(nCas9)多肽或其片段。在一些实施方案中,Cas9多肽是核酸酶死亡的Cas9(dCas9)多肽或其片段。融合蛋白中的Cas9多肽可以是全长Cas9多肽。在一些情况下,融合蛋白中的Cas9多肽可以不是全长Cas9多肽。Cas9多肽可以在例如N端或C端相对于天然出现的Cas9蛋白截短。Cas9多肽可以是环状高突变的Cas9蛋白。Cas9多肽可以是Cas9多肽的片段、部分或结构域,其仍能够结合靶标多核苷酸和向导核酸序列。

[0776] 在一些实施方案中,Cas9多肽是酿脓链球菌Cas9(SpCas9)、金黄色葡萄球菌Cas9(SaCas9)、嗜热链球菌1Cas9(St1Cas9),或其片段或变体。

[0777] 融合蛋白的Cas9可以包含氨基酸序列,所述氨基酸序列与天然出现的Cas9多肽为至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相同。

[0778] 融合蛋白的Cas9可包含氨基酸序列,所述氨基酸序列与下文详述的Cas9氨基酸序列(下文称为“Cas9参考序列”)为至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相同:

[0779] MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

[0780] RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRK
KLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKA
ILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLA
QIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQS
FIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVT
VKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRE
MIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTT
QKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDH
IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE
LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLI
ARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV
KKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE
QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTT
IDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD（单下划线:HNH结构域;双下划线:RuvC结构域）。

[0781] 包含侧翼具有Cas9多肽的N端和C端的异源催化结构域的融合蛋白也可用于进行本文所述方法中的碱基编辑。包含Cas9和一种或多种脱氨酶结构域（例如，腺苷脱氨酶）或包含侧翼具有Cas9序列的腺苷脱氨酶结构域的融合蛋白也可用于对靶标序列进行高度特异性的有效碱基编辑。在一种实施方案中，嵌合Cas9融合蛋白含有插入Cas9多肽内的异源催化结构域（例如，腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶）。在一些实施方案中，融合蛋白包含腺苷脱氨酶结构域和插入到Cas9内的胞苷脱氨酶结构域。在一些实施方案中，腺苷脱氨酶融合在Cas9内，并且胞苷脱氨酶融合至C端。在一些实施方案中，腺苷脱氨酶融合在Cas9内，并且胞苷脱氨酶融合至N端。在一些实施方案中，胞苷脱氨酶融合在Cas9内，并且腺苷脱氨酶融合至C端。在一些实施方案中，胞苷脱氨酶融合在Cas9内，并且腺苷脱氨酶融合至N端。

[0782] 具有腺苷脱氨酶和胞苷脱氨酶以及Cas9的融合蛋白的示例性结构提供如下：

[0783] NH_2-[Cas9（腺苷脱氨酶）]-[胞苷脱氨酶]-COOH；

[0784] NH_2-[胞苷脱氨酶]-[Cas9（腺苷脱氨酶）]-COOH；

[0785] NH_2-[Cas9（胞苷脱氨酶）]-[腺苷脱氨酶]-COOH；或

[0786] NH_2-[腺苷脱氨酶]-[Cas9（胞苷脱氨酶）]-COOH。

[0787] 在一些实施方案中，上述通用结构中使用的“-”表示存在任选的连接子。

[0788] 在多种实施方案中，催化机构与具有DNA修饰活性（例如，脱氨酶活性），诸如腺苷

脱氨酶活性。在一些实施方案中,腺苷脱氨酶是TadA(例如,TadA7.10)。在一些实施方案中,TadA是TadA*8。在一些实施方案中,TadA*8融合在Cas9内,并且胞苷脱氨酶融合至C端。在一些实施方案中,TadA*8融合在Cas9内,并且胞苷脱氨酶融合至N端。在一些实施方案中,胞苷脱氨酶融合在Cas9内,并且TadA*8融合至C端。在一些实施方案中,胞苷脱氨酶融合在Cas9内,并且TadA*8融合至N端。具有TadA*8和胞苷脱氨酶以及Cas9的融合蛋白的示例性结构提供如下:

[0789] NH_2-[Cas9(TadA*8)]-[胞苷脱氨酶]-COOH;

[0790] NH_2-[胞苷脱氨酶]-[Cas9(TadA*8)]-COOH;

[0791] NH_2-[Cas9(胞苷脱氨酶)]-[TadA*8]-COOH;或

[0792] NH_2-[TadA*8]-[Cas9(胞苷脱氨酶)]-COOH。

[0793] 在一些实施方案中,上述通用结构中使用的“-”表示存在任选的连接子。

[0794] 可以将异源多肽(例如,脱氨酶)插入napDNAbp(例如,Cas9或Cas12(例如,Cas12b/C2c1))的合适位置处,例如,使得napDNAbp保留其结合靶标多核苷酸和向导核酸的能力。可以将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入napDNAbp中而不破坏该脱氨酶的功能(例如,碱基编辑活性)或该napDNAbp的功能(例如,结合至靶标核酸和向导核酸的能力)。可以将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入napDNAbp的例如无序区域内或包含高温因子或B-因子的区域中,如通过晶体学研究所示。蛋白质的不太有序、无序或非结构化区域,例如暴露于溶剂的区域和环,可以用于插入而不破坏结构或功能。可以将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入napDNAbp的柔性环区域中或暴露于溶剂的区域中。在一些实施方案中,将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入Cas9或Cas12b/C2c1多肽的柔性环中。

[0795] 在一些实施方案中,通过对Cas9多肽晶体结构的B-因子分子来确定脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)的插入位置。在一些实施方案中,将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入Cas9肽的包含高于平均值的B因子(例如,与总蛋白或包含无序区域的蛋白结构域相比,更高的B因子)的区域中。B-因子或温度因子可以表明原子从其平均位置的波动(例如,作为温度依赖性原子振动或晶格中的静态无序的结果)。主链原子的高B-因子(例如,高于平均B-因子)可以指示具有相对高的局部迁移率的区域的标志。此类区域可用于插入脱氨酶而不破坏结构和功能。可以将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入某一位置处,该位置具有含Cα原子的残基且B-因子比总蛋白的平均B因子高50%、60%、70%、80%、90%、100%、110%、120%、130%、140%、150%、160%、170%、180%、190%、200%或超过200%。可以将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入某一位置处,该位置具有含Cα原子的残基且B-因子比包含该残基的Cas9蛋白结构域的平均B因子高50%、60%、70%、80%、90%、100%、110%、120%、130%、140%、150%、160%、170%、180%、190%、200%或超过200%。包含高于平均值的B因子的Cas9多肽位置可以包括,例如,残基768、792、1052、1015、1022、1026、1029、1067、1040、1054、1068、1246、1247和1248,如上述Cas9参考序列中编号。包含高于平均值的B因子的Cas9多肽区域可以包括,例如,残基792至872、792至906和2至791,如上述Cas9参考序列中编号。

[0796] 可以将异源多肽(例如,脱氨酶)插入napDNAbp中的选自由以下所组成的组的氨基酸残基处:768、791、792、1015、1016、1022、1023、1026、1029、1040、1052、1054、1067、1068、1069、1246、1247和1248,如上述Cas9参考序列中编号,或者另一Cas9多肽的相应氨基酸残基处。在一些实施方案中,将异源多肽插入如上述Cas9参考序列中编号的氨基酸位置768至769、791至792、792至793、1015至1016、1022至1023、1026至1027、1029至1030、1040至1041、1052至1053、1054至1055、1067至1068、1068至1069、1247至1248或1248至1249之间或其相应的氨基酸位置处。在一些实施方案中,将异源多肽插入如上述Cas9参考序列中编号的氨基酸位置769至770、792至793、793至794、1016至1017、1023至1024、1027至1028、1030至1031、1041至1042、1053至1054、1055至1056、1068至1069、1069至1070、1248至1249或1249至1250之间或其相应的氨基酸位置处。在一些实施方案中,异源多肽代替选自由下列组成的组的氨基酸残基:768、791、792、1015、1016、1022、1023、1026、1029、1040、1052、1054、1067、1068、1069、1246、1247和1248,如上述Cas9参考序列中编号,或者另一Cas9多肽的相应氨基酸残基处。应理解,关于插入位置而对上述Cas9参考序列的引用仅用于例示性说明目的。本文所讨论的插入不限于上述Cas9参考序列的Cas9多肽序列,而是包括在变体Cas9多肽(例如,Cas9切口酶(nCas9)、核酸酶死亡的Cas9(dCas9)、缺乏核酸酶结构域的Cas9变体、截短的Cas9、或缺乏部分或完整的HNH结构域的Cas9结构域)中相应位置处的插入。

[0797] 可以将异源多肽(例如,脱氨酶)插入napDNAbp中的选自由以下所组成的组的氨基酸残基处:768、792、1022、1026、1040、1068和1247,如上述Cas9参考序列中编号,或者另一Cas9多肽的相应氨基酸残基处。在一些实施方案中,将异源多肽插入如上述Cas9参考序列中编号的氨基酸位置768至769、792至793、1022至1023、1026至1027、1029至1030、1040至1041、1068至1069或1247至1248之间或其相应的氨基酸位置处。在一些实施方案中,将异源多肽插入如上述Cas9参考序列中编号的氨基酸位置769至770、793至794、1023至1024、1027至1028、1030至1031、1041至1042、1069至1070或1248至1249之间或其相应的氨基酸位置处。在一些实施方案中,异源多肽代替选自由下列组成的组的氨基酸残基:768、792、1022、1026、1040、1068和1247,如上述Cas9参考序列中编号,或者另一Cas9多肽的相应氨基酸残基处。

[0798] 异源多肽(例如,脱氨酶)可以在如本文所述的氨基酸残基处或另一Cas9多肽的相应氨基酸残基处插入napDNAbp内。在一种实施方案中,可以将异源多肽(例如,脱氨酶)插入napDNAbp中的选自由以下所组成的组的氨基酸残基处:1002、1003、1025、1052至1056、1242至1247、1061至1077、943至947、686至691、569至578、530至539和1060至1077,如上述Cas9参考序列中编号,或者另一Cas9多肽的相应氨基酸残基处。可将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入在残基的N端或C端或替换该残基。在一些实施方案中,将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入在残基的C端。

[0799] 在一些实施方案中,将腺苷脱氨酶(例如,TadA)插入选自由下列组成的组的氨基酸残基处:1015、1022、1029、1040、1068、1247、1054、1026、768、1067、1248、1052和1246,如上述Cas9参考序列中编号,或者另一Cas9多肽的相应氨基酸残基处。在一些实施方案中,将腺苷脱氨酶(例如,TadA)插入在残基792至872、792至906或2至791的位置处,如上述Cas9参考序列中编号,或者另一Cas9多肽的对应氨基酸残基处。在一些实施方案中,将腺苷脱氨酶

插入选自由下列所组成的组的氨基酸的N端:1015、1022、1029、1040、1068、1247、1054、1026、768、1067、1248、1052和1246,如上述Cas9参考序列中编号,或者另一Cas9多肽的相应氨基酸残基处。在一些实施方案中,将腺苷脱氨酶插入选自由下列所组成的组的氨基酸的C端:1015、1022、1029、1040、1068、1247、1054、1026、768、1067、1248、1052和1246,如上述Cas9参考序列中编号,或者另一Cas9多肽的相应氨基酸残基处。在一些实施方案中,将腺苷脱氨酶插入以替换选自由下列所组成的组的氨基酸:1015、1022、1029、1040、1068、1247、1054、1026、768、1067、1248、1052和1246,如上述Cas9参考序列中编号,或者另一Cas9多肽的相应氨基酸残基处。

[0800] 在一些实施方案中,将CBE(例如,APOBEC1)插入选自由下列组成的组的氨基酸残基处:1016、1023、1029、1040、1069和1247,如上述Cas9参考序列中编号,或者另一Cas9多肽的相应氨基酸残基处。在一些实施方案中,将ABE插入选自由下列所组成的组的氨基酸的N端:1016、1023、1029、1040、1069和1247,如上述Cas9参考序列中编号,或者另一Cas9多肽的相应氨基酸残基处。在一些实施方案中,将ABE插入选自由下列所组成的组的氨基酸的C端:1016、1023、1029、1040、1069和1247,如上述Cas9参考序列中编号,或者另一Cas9多肽的相应氨基酸残基处。在一些实施方案中,将ABE插入以替换选自由下列所组成的组的氨基酸:1016、1023、1029、1040、1069和1247,如上述Cas9参考序列中编号,或者另一Cas9多肽的相应氨基酸残基处。

[0801] 在一些实施方案中,将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入氨基酸残基768处,如上述Cas9参考序列中编号,或者另一Cas9多肽的对应氨基酸残基处。在一些实施方案中,将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入氨基酸残基768的N端处,如上述Cas9参考序列中编号,或者另一Cas9多肽的对应氨基酸残基处。在一些实施方案中,将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入氨基酸残基768的C端处,如上述Cas9参考序列中编号,或者另一Cas9多肽的对应氨基酸残基处。在一些实施方案中,将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入以替换氨基酸残基768,如上述Cas9参考序列中编号,或者另一Cas9多肽的对应氨基酸残基处。

[0802] 在一些实施方案中,将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入氨基酸残基791处或插入氨基酸残基792处,如上述Cas9参考序列中编号,或者另一Cas9多肽的对应氨基酸残基处。在一些实施方案中,将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入氨基酸残基791的N端处或插入氨基酸792的N端处,如上述Cas9参考序列中编号,或者另一Cas9多肽的对应氨基酸残基处。在一些实施方案中,将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入氨基酸791的C端处或插入氨基酸792的N端处,如上述Cas9参考序列中编号,或者另一Cas9多肽的对应氨基酸残基处。在一些实施方案中,将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入以替换氨基酸791或插入以替换氨基酸792,如上述Cas9参考序列中编号,或者另一Cas9多肽的对应氨基酸残基处。

[0803] 在一些实施方案中,将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入氨基酸残基1016处,如上述Cas9参考序列中编号,或者另一Cas9多肽的对应氨基酸残基处。在一些实施方案中,将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷

脱氨酶和胞苷脱氨酶)插入氨基酸残基1016的N端处,如上述Cas9参考序列中编号,或者另一Cas9多肽的对应氨基酸残基处。在一些实施方案中,将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入氨基酸残基1016的C端处,如上述Cas9参考序列中编号,或者另一Cas9多肽的对应氨基酸残基处。在一些实施方案中,将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入以替换氨基酸残基1016,如上述Cas9参考序列中编号,或者另一Cas9多肽的对应氨基酸残基处。

[0804] 在一些实施方案中,将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入氨基酸残基1022处或插入氨基酸残基1023处,如上述Cas9参考序列中编号,或者另一Cas9多肽的对应氨基酸残基处。在一些实施方案中,将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入氨基酸残基1022的N端处或插入氨基酸残基1023的N端处,如上述Cas9参考序列中编号,或者另一Cas9多肽的对应氨基酸残基处。在一些实施方案中,将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入氨基酸残基1022的C端处或插入氨基酸残基1023的C端处,如上述Cas9参考序列中编号,或者另一Cas9多肽的对应氨基酸残基处。在一些实施方案中,将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入以替换氨基酸残基1022或插入以替换氨基酸残基1023,如上述Cas9参考序列中编号,或者另一Cas9多肽的对应氨基酸残基处。

[0805] 在一些实施方案中,将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入氨基酸残基1026处或插入氨基酸残基1029处,如上述Cas9参考序列中编号,或者另一Cas9多肽的对应氨基酸残基处。在一些实施方案中,将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入氨基酸残基1026的N端处或插入氨基酸残基1029的N端处,如上述Cas9参考序列中编号,或者另一Cas9多肽的对应氨基酸残基处。在一些实施方案中,将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入氨基酸残基1026的C端处或插入氨基酸残基1029的C端处,如上述Cas9参考序列中编号,或者另一Cas9多肽的对应氨基酸残基处。在一些实施方案中,将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入以替换氨基酸残基1026或插入以替换氨基酸残基1029,如上述Cas9参考序列中编号,或者另一Cas9多肽的对应氨基酸残基处。

[0806] 在一些实施方案中,将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入氨基酸残基1040处,如上述Cas9参考序列中编号,或者另一Cas9多肽的对应氨基酸残基处。在一些实施方案中,将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入氨基酸残基1040的N端处,如上述Cas9参考序列中编号,或者另一Cas9多肽的对应氨基酸残基处。在一些实施方案中,将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入氨基酸残基1040的C端处,如上述Cas9参考序列中编号,或者另一Cas9多肽的对应氨基酸残基处。在一些实施方案中,将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入以替换氨基酸残基1040,如上述Cas9参考序列中编号,或者另一Cas9多肽的对应氨基酸残基处。

[0807] 在一些实施方案中,将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入氨基酸残基1052处或插入氨基酸残基1054处,如上述Cas9参考序列中编

号,或者另一Cas9多肽的对应氨基酸残基处。在一些实施方案中,将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入氨基酸残基1052的N端处或插入氨基酸残基1054的N端处,如上述Cas9参考序列中编号,或者另一Cas9多肽的对应氨基酸残基处。在一些实施方案中,将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入氨基酸残基1052的C端处或插入氨基酸残基1054的C端处,如上述Cas9参考序列中编号,或者另一Cas9多肽的对应氨基酸残基处。在一些实施方案中,将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入以替换氨基酸残基1052或插入以替换氨基酸残基1054,如上述Cas9参考序列中编号,或者另一Cas9多肽的对应氨基酸残基处。

[0808] 在一些实施方案中,将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入氨基酸残基1067处、或插入氨基酸残基1068处、或插入氨基酸残基1069处,如上述Cas9参考序列中编号,或者另一Cas9多肽的对应氨基酸残基处。在一些实施方案中,将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入氨基酸残基1067的N端处或插入氨基酸残基1068的N端处或插入氨基酸残基1069的N端处,如上述Cas9参考序列中编号,或者另一Cas9多肽的对应氨基酸残基处。在一些实施方案中,将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入氨基酸残基1067的C端处或插入氨基酸残基1068的C端处或插入氨基酸残基1069的C端处,如上述Cas9参考序列中编号,或者另一Cas9多肽的对应氨基酸残基处。在一些实施方案中,将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入以替换氨基酸残基1067、或插入以替换氨基酸残基1068、或插入以替换氨基酸残基1069,如上述Cas9参考序列中编号,或者另一Cas9多肽的对应氨基酸残基处。

[0809] 在一些实施方案中,将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入氨基酸残基1246处、或插入氨基酸残基1247处、或插入氨基酸残基1248处,如上述Cas9参考序列中编号,或者另一Cas9多肽的对应氨基酸残基处。在一些实施方案中,将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入氨基酸残基1246的N端处或插入氨基酸残基1247的N端处或插入氨基酸残基1248的N端处,如上述Cas9参考序列中编号,或者另一Cas9多肽的对应氨基酸残基处。在一些实施方案中,将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入氨基酸残基1246的C端处或插入氨基酸残基1247的C端处或插入氨基酸残基1248的C端处,如上述Cas9参考序列中编号,或者另一Cas9多肽的对应氨基酸残基处。在一些实施方案中,将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入以替换氨基酸残基1246、或插入以替换氨基酸残基1247、或插入以替换氨基酸残基1248,如上述Cas9参考序列中编号,或者另一Cas9多肽的对应氨基酸残基处。

[0810] 在一些实施方案中,将异源多肽(例如,脱氨酶)插入Cas9多肽的柔性环中。柔性环部分可以选自下列所组成的组:如上述Cas9参考序列中编号的530至537、569至570、686至691、943至947、1002至1025、1052至1077、1232至1247或1298至1300,或另一Cas9肽的相应氨基酸残基。柔性环部分可以选自由下列所组成的组:1至529、538至568、580至685、692至942、948至1001、1026至1051、1078至1231或1248至1297,如上述Cas9参考序列中编号,或者另一Cas9多肽的相应氨基酸残基处。

[0811] 异源多肽(例如,腺苷脱氨酶)可以被插入对应于以下氨基酸残基的Cas9多肽区域中:如上述Cas9参考序列中编号的1017至1069、1242至1247、1052至1056、1060至1077、1002至1003、943至947、530至537、568至579、686至691、1242至1247、1298至1300、1066至1077、1052至1056或1060至1077,或另一Cas9多肽的相应氨基酸残基。

[0812] 可以将异源多肽(例如,腺苷脱氨酶)插入Cas9多肽的被删除区域的位置处。被删除区域可以对应于Cas9多肽的N端或C端。在一些实施方案中,被删除区域对应于如上述Cas9参考序列中编号的残基792至872,或另一Cas9多肽的相应氨基酸残基。在一些实施方案中,被删除区域对应于如上述Cas9参考序列中编号的残基792至906,或另一Cas9多肽的相应氨基酸残基。在一些实施方案中,被删除区域对应于如上述Cas9参考序列中编号的残基2至791,或另一Cas9多肽的相应氨基酸残基。在一些实施方案中,被删除区域对应于如上述Cas9参考序列中编号的残基1017至1069,或其相应氨基酸残基。

[0813] 示例性的内部融合碱基编辑器提供在下表5A中:

[0814] 表5A:Cas9蛋白中的插入基因座

[0815]

BE ID	修饰	其他ID
IBE001	Cas9 TadA ins 1015	ISLAY01
IBE002	Cas9 TadA ins 1022	ISLAY02
IBE003	Cas9 TadA ins 1029	ISLAY03
IBE004	Cas9 TadA ins 1040	ISLAY04
IBE005	Cas9 TadA ins 1068	ISLAY05
IBE006	Cas9 TadA ins 1247	ISLAY06
IBE007	Cas9 TadA ins 1054	ISLAY07
IBE008	Cas9 TadA ins 1026	ISLAY08
IBE009	Cas9 TadA ins 768	ISLAY09
IBE020	delta HNH TadA 792	ISLAY20
IBE021	N端融合单TadA螺旋截短165-末端	ISLAY21
IBE029	TadA-环状高突变体116 ins 1067	ISLAY29
IBE031	TadA-环状高突变体136 ins 1248	ISLAY31
IBE032	TadA-环状高突变体136ins 1052	ISLAY32
IBE035	δ792-872TadA ins	ISLAY35
IBE036	δ792-906TadA ins	ISLAY36
IBE043	TadA-环状高突变体65 ins1246	ISLAY43
IBE044	TadA ins C-端截短2 791	ISLAY44

[0816] 可以将异源多肽(例如,脱氨酶)插入Cas9多肽的结构性或功能性结构域内。可以将异源多肽(例如,脱氨酶)插入Cas9多肽的两个结构性或功能性结构域之间。可以将异源多肽(例如,脱氨酶)插入Cas9多肽的结构性或功能性结构域的位置处,例如,在从Cas9多肽删除该结构域之后。Cas9多肽的结构性或功能性结构域可以包括,例如,RuvC I、RuvC II、RuvC III、Rec1、Rec2、PI或HNH。

[0817] 在一些实施方案中,Cas9多肽缺少选自以下所组成的组的一个或多个结构域:RuvC I、RuvC II、RuvC III、Rec1、Rec2、PI或HNH结构域。在一些实施方案中,Cas9多肽缺少

核酸酶结构域。在一些实施方案中,Cas9多肽缺少HNH结构域。在一些实施方案中,Cas9多肽缺少HNH结构域的一部分,使得该Cas9多肽的HNH活性下降或消失。在一些实施方案中,Cas9多肽包含核酸酶结构域的缺失,并且插入脱氨酶以代替该核酸酶结构域。在一些实施方案中,HNH结构域被删除,并且脱氨酶被插入到其位置处。在一些实施方案中,一个或多个RuvC结构域被删除,并且脱氨酶被插入到其位置处。

[0818] 包含异源多肽的融合蛋白侧翼可以具有napDNAbp的N端和C端片段。在一些实施方案中,融合蛋白包含侧翼具有napDNAbp的N端和C端片段的脱氨酶。该N端片段或C端片段可以结合靶标多核苷酸序列。该N端片段的C端或该C端片段的N端可以包含Cas9多肽的柔性环的一部分。该N端片段的C端或该C端片段的N端可以包含Cas9多肽的α-螺旋结构的一部分。该C端片段的N端可以包含DNA结合结构域。该C端片段的N端可以包含RuvC结构域。该C端片段的N端可以包含HNH结构域。在一些实施方案中,该N端片段和C端片段都不包含HNH结构域。

[0819] 在一些实施方案中,该N端Cas9片段的C端包含一个氨基酸,当融合蛋白将靶标核碱基脱氨基时,该氨基酸接近该靶标核碱基。在一些实施方案中,该C端Cas9片段的N端包含一个氨基酸,当融合蛋白将靶标核碱基脱氨基时,该氨基酸接近该靶标核碱基。不同脱氨酶的插入位置可以不同,以便在靶标核碱基与N端Cas9片段的C端或C端Cas9片段的N端之间具有近距离。例如,ABE的插入位置可以位于选自由以下组成的组的氨基酸残基处:1015、1022、1029、1040、1068、1247、1054、1026、768、1067、1248、1052和1246,如上述Cas9参考序列中编号,或者另一Cas9多肽的相应氨基酸残基处。

[0820] 融合蛋白的N端Cas9片段(即,位于融合蛋白中脱氨酶侧翼的N端Cas9片段)可以包含Cas9多肽的N端。融合蛋白的N端Cas9片段可以包含至少约下列的长度:100、200、300、400、500、600、700、800、900、1000、1100、1200或1300个氨基酸。融合蛋白的N端Cas9片段可以包含对应于以下氨基酸残基的序列:如上述Cas9参考序列中编号的1至56、1至95、1至200、1至300、1至400、1至500、1至600、1至700、1至718、1至765、1至780、1至906、1至918或1至1100,或另一Cas9多肽的相应氨基酸残基。融合蛋白的N端Cas9片段可以包含序列,该序列包含与以下氨基酸序列至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%序列同一性:如上述Cas9参考序列中编号的1至56、1至95、1至200、1至300、1至400、1至500、1至600、1至700、1至718、1至765、1至780、1至906、1至918或1至1100,或另一Cas9多肽的相应氨基酸残基。

[0821] 融合蛋白的C端Cas9片段(即,位于融合蛋白中脱氨酶侧翼的C端Cas9片段)可以包含Cas9多肽的C端。融合蛋白的C端Cas9片段可以包含至少约下列的长度:100、200、300、400、500、600、700、800、900、1000、1100、1200或1300个氨基酸。融合蛋白的C端Cas9片段可以包含对应于以下氨基酸残基的序列:1099至1368、918至1368、906至1368、780至1368、765至1368、718至1368、94至1368或56至1368,如上述Cas9参考序列中编号,或者另一Cas9多肽的相应氨基酸残基处。融合蛋白的N端Cas9片段可以包含序列,该序列包含与以下氨基酸序列至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%序列同一性:1099至1368、918至1368、906至1368、780至1368、765至1368、718至1368、94至1368或56至1368,如上述Cas9参考序列中编号,或者另一Cas9多肽的相应氨基酸残基处。

[0822] 融合蛋白的N端Cas9片段和C端Cas9片段整体看来可能并不对应于全长度天然出现的Cas9多肽序列,例如,如上述Cas9参考序列中详述的。

[0823] 本文所述的融合蛋白可以实现靶向脱氨基,并且在非靶标位点(例如,脱靶位点)处的脱氨基作用减少,诸如全基因组假脱氨基作用减少。本文所述的融合蛋白可以实现靶向脱氨基,并且在非靶标位点的旁观者脱氨基作用减少。与例如包含融合至Cas9多肽的N端或C端的脱氨酶的末端融合蛋白相比,不希望的脱氨基作用或脱靶脱氨基作用可以减少至少30%、至少40%、至少50%、至少60%、至少70%、至少80%、至少90%、至少95%或至少99%。与例如包含融合至Cas9多肽的N端或C端的脱氨酶的末端融合蛋白相比,不希望的脱氨基作用或脱靶脱氨基作用可以减少1倍、至少2倍、至少3倍、至少4倍、至少5倍、至少10倍、至少15倍、至少20倍、至少30倍、至少40倍、至少50倍、至少60倍、至少70倍、至少80倍、至少90倍或至少100倍。

[0824] 在一些实施方案中,融合蛋白的脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)将R环范围内的不超过两个核碱基脱氨基。在一些实施方案中,融合蛋白的脱氨酶将R环范围内的不超过三个核碱基脱氨基。在一些实施方案中,融合蛋白的脱氨酶将R环范围内的不超过2、3、4、5、6、7、8、9或10个核碱基脱氨基。R环是一种三链核酸结构,其包括DNA:RNA杂交物、DNA:DNA或RNA:RNA互补结构亦即与单链DNA缔合。如本文所用,R环可在靶标多核苷酸与CRISPR复合物或碱基编辑复合物接触时形成,其中向导多核苷酸(例如,向导RNA)的一部分与靶标多核苷酸(例如,靶标DNA)的一部分杂交并替换该靶标多核苷酸的一部分。在一些实施方案中,R环包含间隔序列与靶标DNA互补序列的经杂交区域。R环区域可以为约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或50个核碱基对的长度。在一些实施方案中,R环区域为约20个核碱基对的长度。应理解,如本文所用,R环区域不限于与向导多核苷酸杂交的靶标DNA链。例如,R环区域内靶标核碱基的编辑可以是对包含与向导RNA的互补链的DNA链的编辑,或者可以是对作为与向导RNA的互补链相反的链的DNA链的编辑。在一些实施方案中,R环区域中的编辑包含编辑靶标DNA序列中向导RNA的非互补链(前间隔序列链)上的核碱基。

[0825] 本文所述的融合蛋白可以在不同于经典碱基编辑的编辑窗内实现靶标脱氨基。在一些实施方案中,靶标核碱基位于靶标多核苷酸序列中PAM序列的上游约1至约20个碱基处。在一些实施方案中,靶标核碱基位于靶标多核苷酸序列中PAM序列的上游约2至约12个碱基处。在一些实施方案中,靶标核碱基位于与PAM序列相距或该PAM序列上游约1至9个碱基对、约2至10个碱基对、约3至11个碱基对、约4至12个碱基对、约5至13个碱基对、约6至14个碱基对、约7至15个碱基对、约8至16个碱基对、约9至17个碱基对、约10至18个碱基对、约11至19个碱基对、约12至20个碱基对、约1至7个碱基对、约2至8个碱基对、约3至9个碱基对、约4至10个碱基对、约5至11个碱基对、约6至12个碱基对、约7至13个碱基对、约8至14个碱基对、约9至15个碱基对、约10至16个碱基对、约11至17个碱基对、约12至18个碱基对、约13至19个碱基对、约14至20个碱基对、约1至5个碱基对、约2至6个碱基对、约3至7个碱基对、约4至8个碱基对、约5至9个碱基对、约6至10个碱基对、约7至11个碱基对、约8至12个碱基对、约9至13个碱基对、约10至14个碱基对、约11至15个碱基对、约12至16个碱基对、约13至17个碱基对、约14至18个碱基对、约15至19个碱基对、约16至20个碱基对、约1至3个碱基对、约2至4

个碱基对、约3至5个碱基对、约4至6个碱基对、约5至7个碱基对、约6至8个碱基对、约7至9个碱基对、约8至10个碱基对、约9至11个碱基对、约10至12个碱基对、约11至13个碱基对、约12至14个碱基对、约13至15个碱基对、约14至16个碱基对、约15至17个碱基对、约16至18个碱基对、约17至19个碱基对、约18至20个碱基对处。在一些实施方案中,靶标核碱基位于与PAM序列相距或该PAM序列上游约1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20或更多个碱基对处。在一些实施方案中,靶标核碱基位于PAM序列上游约1、2、3、4、5、6、7、8或9个碱基对处。在一些实施方案中,靶标核碱基位于PAM序列上游约2、3、4或6个碱基对处。

[0826] 融合蛋白可以包含超过一种异源多肽。例如,融合蛋白可以额外地包含一种或多种UGI结构域和/或一种或多种核定位信号。该两种或更多种异源结构域可以串联插入。可以将两种或更多种异源结构域插入多个位置处,使得它们在NapDNAbp中不是串联的。

[0827] 融合蛋白可包含位于脱氨酶与napDNAbp多肽之间的连接子。连接子可以是肽或非肽连接子。例如,连接子可以是XTEN、(GGGS)n、(GGGGS)n、(G)n、(EAAAK)n、(GGS)n、SGSETPGTSESATPES。在一些实施方案中,融合蛋白包含位于N端Cas9片段与脱氨酶之间的连接子。在一些实施方案中,融合蛋白包含位于C端Cas9片段与脱氨酶之间的连接子。在一些实施方案中,napDNAbp的N端和C端片段通过连接子连结至脱氨酶。在一些实施方案中,N端和C端片段不通过连接子连结至脱氨酶结构域。在一些实施方案中,融合蛋白包含位于N端Cas9片段与脱氨酶之间的连接子,但不包含位于C端Cas9片段与脱氨酶之间的连接子。在一些实施方案中,融合蛋白包含位于C端Cas9片段与脱氨酶之间的连接子,但不包含位于N端Cas9片段与脱氨酶之间的连接子。

[0828] 在一些实施方案中,融合蛋白中的napDNAbp是Cas12多肽(例如,Cas12b/C2c1)或其片段。Cas12多肽可以是变体Cas12多肽。在其他实施方案中,Cas12多肽的N端或C端片段包含核酸可编程DNA结合结构域或RuvC结构域。在其他实施方案中,融合蛋含有位于Cas12多肽与催化结构域之间的连接子。在其他实施方案中,连接子的氨基酸序列是GGSGGS或GSSGSETPGTSESATPESSG。在其他实施方案中,连接子是刚性连接子。在上述方面的其他实施方案中,连接子由GGAGGCTCTGGAGGAAGC或GGCTCTTCTGGATCTGAAACACCTGGCACAAGCGAGAGCGCCACCCCTGAGAGCTCTGGC编码。

[0829] 包含侧翼具有Cas12多肽的N端和C端的异源催化结构域的融合蛋白也可用于进行本文所述方法中的碱基编辑。包含Cas12和一种或多种脱氨酶结构域(例如,腺苷脱氨酶)或包含侧翼具有Cas12序列的腺苷脱氨酶结构域的融合蛋白也可用于对靶标序列进行高度特异性的有效碱基编辑。在一种实施方案中,嵌合Cas12融合蛋白含有插入Cas12多肽内的异源催化结构域(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)。在一些实施方案中,融合蛋白包含腺苷脱氨酶结构域和插入到Cas12内的胞苷脱氨酶结构域。在一些实施方案中,腺苷脱氨酶融合在Cas12内,并且胞苷脱氨酶融合至C端。在一些实施方案中,腺苷脱氨酶融合在Cas12内,并且胞苷脱氨酶融合至N端。在一些实施方案中,胞苷脱氨酶融合在Cas12内,并且腺苷脱氨酶融合至C端。在一些实施方案中,胞苷脱氨酶融合在Cas12内,并且腺苷脱氨酶融合至N端。具有腺苷脱氨酶和胞苷脱氨酶以及Cas12的融合蛋白的示例性结构提供如下:

[0830] NH_2-[Cas12(腺苷脱氨酶)]-[胞苷脱氨酶]-COOH;

[0831] NH_2-[胞苷脱氨酶]-[Cas12(腺苷脱氨酶)]-COOH;

[0832] NH_2-[Cas12(胞苷脱氨酶)]-[腺苷脱氨酶]-COOH;或

[0833] NH_2-[腺苷脱氨酶]-[Cas12(胞苷脱氨酶)]-COOH。

[0834] 在一些实施方案中,上述通用结构中使用的“-”表示存在任选的连接子。

[0835] 在多种实施方案中,催化机构与具有DNA修饰活性(例如,脱氨酶活性),诸如腺苷脱氨酶活性。在一些实施方案中,腺苷脱氨酶是TadA(例如,TadA7.10)。在一些实施方案中,TadA是TadA*8。在一些实施方案中,TadA*8融合在Cas12内,并且胞苷脱氨酶融合至C端。在一些实施方案中,TadA*8融合在Cas12内,并且胞苷脱氨酶融合至N端。在一些实施方案中,胞苷脱氨酶融合在Cas12内,并且TadA*8融合至C端。在一些实施方案中,胞苷脱氨酶融合在Cas12内,并且TadA*8融合至N端。具有TadA*8和胞苷脱氨酶以及Cas12的融合蛋白的示例性结构提供如下:

[0836] N-[Cas12(TadA*8)]-[胞苷脱氨酶]-C;

[0837] N-[胞苷脱氨酶]-[Cas12(TadA*8)]-C;

[0838] N-[Cas12(胞苷脱氨酶)]-[TadA*8]-C;或

[0839] N-[TadA*8]-[Cas12(胞苷脱氨酶)]-C。

[0840] 在一些实施方案中,上述通用结构中使用的“-”表示存在任选的连接子。

[0841] 在其他实施方案中,融合蛋白含有一个或多个催化结构域。在其他实施方案中,一个或多个催化结构域中的至少一个插入到Cas12多肽内或融合在Cas12 N端或C端。在其他实施方案中,一个或多个催化结构域中的至少一个插入到Cas12多肽的环、α螺旋区、非结构化部分或溶剂可及部分。在其他实施方案中,Cas12多肽是Cas12a、Cas12b、Cas12c、Cas12d、Cas12e、Cas12g、Cas12h或Cas12i。在其他实施方案中,Cas12多肽与外村尚芽孢杆菌Cas12b、热噬淀粉芽孢杆菌Cas12b、芽孢杆菌属菌种V3-13 Cas12b或嗜酸性脂环酸芽孢杆菌Cas12b具有至少约85%的氨基酸序列同一性。在其他实施方案中,Cas12多肽与外村尚芽孢杆菌Cas12b、热噬淀粉芽孢杆菌Cas12b、芽孢杆菌属菌种V3-13 Cas12b或嗜酸性脂环酸芽孢杆菌Cas12b具有至少约90%的氨基酸序列同一性。在其他实施方案中,Cas12多肽与外村尚芽孢杆菌Cas12b、热噬淀粉芽孢杆菌Cas12b、芽孢杆菌属菌种V3-13 Cas12b或嗜酸性脂环酸芽孢杆菌Cas12b具有至少约95%的氨基酸序列同一性。在其他实施方案中,Cas12多肽含有外村尚芽孢杆菌Cas12b、热噬淀粉芽孢杆菌Cas12b、芽孢杆菌属菌种V3-13 Cas12b或嗜酸性脂环酸芽孢杆菌Cas12b的片段或基本上尤其组成。

[0842] 在其他实施方案中,将催化结构域插入BhCas12b的氨基酸位置153至154、255至256、306至307、980至981、1019至1020、534至535、604至605或344至345之间或Cas12a、Cas12c、Cas12d、Cas12e、Cas12g、Cas12h或Cas12i的相应的氨基酸残基处。在其他实施方案中,催化结构域插入在BhCas12b的氨基酸P153和S154之间。在其他实施方案中,催化结构域插入在BhCas12b的氨基酸K255和E256之间。在其他实施方案中,催化结构域插入在BhCas12b的氨基酸D980和G981之间。在其他实施方案中,催化结构域插入在BhCas12b的氨基酸K1019和L1020之间。在其他实施方案中,催化结构域插入在BhCas12b的氨基酸F534和P535之间。在其他实施方案中,催化结构域插入在BhCas12b的氨基酸K604和G605之间。在其他实施方案中,催化结构域插入在BhCas12b的氨基酸H344和F345之间。在其他实施方案中,催化结构域插入在BvCas12b的位置147和148,248和249,299和300,991和992、或者1031和1032之间,或Cas12a、Cas12c、Cas12d、Cas12e、Cas12g、Cas12h或Cas12i的对应氨基酸残基

处。在其他实施方案中,催化结构域插入在BvCas12b的氨基酸P147和D148之间。在其他实施方案中,催化结构域插入在BvCas12b的氨基酸G248和G249之间。在其他实施方案中,催化结构域插入在BvCas12b的氨基酸P299和E300之间。在其他实施方案中,催化结构域插入在BvCas12b的氨基酸G991和E992之间。在其他实施方案中,催化结构域插入在BvCas12b的氨基酸K1031和M1032之间。在其他实施方案中,催化结构域插入在AaCas12b的位置157和158,258和259,310和311,1008和1009、或者1044和1045之间,或Cas12a、Cas12c、Cas12d、Cas12e、Cas12g、Cas12h或Cas12i的对应氨基酸残基处。在其他实施方案中,催化结构域插入在AaCas12b的氨基酸P157和G158之间。在其他实施方案中,催化结构域插入在AaCas12b的氨基酸V258和G259之间。在其他实施方案中,催化结构域插入在AaCas12b的氨基酸D310和P311之间。在其他实施方案中,催化结构域插入在AaCas12b的氨基酸G1008和E1009之间。在其他实施方案中,催化结构域插入在AaCas12b的氨基酸G1044和K1045之间。

[0843] 在其他实施方案中,融合蛋白含有核定位信号(例如,二分体核定位信号)。在其他实施方案中,核定位信号的氨基酸序列是MAPKKKRKVGIHGVPAA。在上述方面的其他实施方案中,核定位信号由以下序列编码:

[0844] ATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCC。在其他实施方案中,Cas12b多肽含有突变,该突变静默RuvC结构域的催化活性。在其他实施方案中,Cas12b多肽含有D574A、D829A和/或D952A突变。在其他实施方案中,融合蛋白进一步含有标签(例如,流感病毒血凝素标签)。

[0845] 在一些实施方案中,融合蛋白包含napDNAbp结构域(例如,衍生自Cas12的结构域)和内部融合的核碱基编辑结构域(例如,脱氨酶结构域(例如,腺苷脱氨酶够用)的全部或一部分)的融合蛋白。在一些实施方案中,napDNAbp是Cas12b。在一些实施方案中,碱基编辑器包含BhCas12b结构域和插入在下表6中提供的基因座处的内部融合的TadA*8结构域。

[0846] 表6:Cas12b蛋白中的插入基因座

[0847]

BhCas12b	插入位点	在 aa 之间插入
位置 1	153	PS
位置 2	255	KE
位置 3	306	DE
位置 4	980	DG
位置 5	1019	KL
位置 6	534	FP
位置 7	604	KG
位置 8	344	HF
BvCas12b	**插入位点**	**在 aa 之间插入**
位置 1	147	PD
位置 2	248	GG
位置 3	299	PE
位置 4	991	GE
位置 5	1031	KM
AaCas12b	**插入位点**	**在 aa 之间插入**
位置 1	157	PG
位置 2	258	VG
位置 3	310	DP
位置 4	1008	GE
位置 5	1044	GK

[0848] 作为非限制性示例,可以将腺苷脱氨酶(例如,ABE8.13)插入到BhCas12b内以产生融合蛋白(例如,ABE8.13-BhCas12b),该融合蛋白有效地编辑核酸序列。

[0849] 在一些实施方案中,本文所述的碱基编辑系统包含具有插入到Cas9中的TadA的ABE。提供了相关的具有插入到Cas9中的TadA的ABE的序列。

[0850] 101 Cas9 TadAins 1015

[0851] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE

NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVL
NNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGA
AGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDYDVRKMIAKSEQEIGKATAKYFF
YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0852] 102 Cas9 TadAins 1022

[0853] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVP
VGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGV
RNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDAKSEQEIGKATAKYFF
YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT

LIHQSITGLYETRIDLSQLGGD

[0854] 103 Cas9 TadAins 1029

[0855] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRA
RDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRI
GRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGKATAKYFF
YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0856] 103 Cas9 TadAins 1040

[0857] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSGSSGSETPGTSESATPESSGSEVEFSHEYW
MRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCV

MCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSS
TDNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0858] 105 Cas9 TadAins 1068

[0859] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGEGS
SGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEI
MALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGIL
ADECAALLCYFFRMPRQVFNAQKKAQSSTDTGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0860] 106 Cas9 TadAins 1247

[0861] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE

MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLV
LNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTG
AAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0862] 107 Cas9 TadAins 1054

[0863] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGSSGSETPGTSESATP
ESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYR
LIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRM
PRQVFNAQKKAQSSTDGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0864] 108 Cas9 TadAins 1026

[0865] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK

PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDE
REVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRV
VFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDQEIGKATAKYFF
YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0866] 109 Cas9 TadAins 768

[0867] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLH
DPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHR
VEITEGILADECAALLCYFFRMPRTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRD
MYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFD
NLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITL
ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK
KYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFEL
ENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILA
DANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI
DLSQLGGD

[0868] 110.1 Cas9 TadAins 1250

[0869] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAV
LVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAK
TGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPREDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILA
DANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI
DLSQLGGD

[0870] 110.2 Cas9 TadAins 1250

[0871] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPGSSGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPV

GAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVR
NAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPREDNEQKQLFVEQHKHYLDEIIEQISEFSKRV
ILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYE
TRIDLSQLGGD

[0872] 110.3 Cas9 TadAins 1250

[0873] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPGSSGSSGSETPGTSESATPESGSSSGSEVEFSHEYWMRHALTLAKRARDERE
VPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVF
GVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPREDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITG
LYETRIDLSQLGGD

[0874] 110.4 Cas9 TadAins 1250

[0875] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF

IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPGSSGSSGSETPGTSESATPESGSSSGSEVEFSHEYWMRHALTLAKRARDERE
VPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVF
GVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMRREDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITG
LYETRIDLSQLGGD

[0876] 110.5 Cas9 TadAins 1249

[0877] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSGSSGSSGSETPGTSESATPESGSSSGSEVEFSHEYWMRHALTLAKRARDEREV
PVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFG
VRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMRRPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITG
LYETRIDLSQLGGD

[0878] 110.5 Cas9 TadAinsδ59-66 1250

[0879] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT

EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPGSSGSSGSETPGTSESATPESGSSGSEVEFSHEYWMRHALTLAKRARDEREV
PVGAVLVLNNRVIGEGWNRAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGA
AGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDEDNEQKQLFVEQHKHYLDEIIEQ
ISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH
QSITGLYETRIDLSQLGGD

[0880] 110.6 Cas9 TadAins 1251

[0881] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPEGSSGSSGSETPGTSESATPESGSSSGSEVEFSHEYWMRHALTLAKRARDER
EVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVV
FGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMRRDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITG
LYETRIDLSQLGGD

[0882] 110.7 Cas9 TadAins 1252

[0883] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST

DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPEDGSSGSSGSETPGTSESATPESGSSSGSEVEFSHEYWMRHALTLAKRARDE
REVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRV
VFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMRRNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITG
LYETRIDLSQLGGD

[0884] 110.8Cas9 TadAinsδ59-66C-截短1250

[0885] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAV
LVLNNRVIGEGWNRAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLM
DVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKV

LSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGG
D

[0886] 111.1 Cas9 TadAins 997

[0887] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDAT
LYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVF
NAQKKAQSSTDGSSGSETPGTSESATPESSGIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIM
NFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLI
ARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIK
LPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQIS
EFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQS
ITGLYETRIDLSQLGGD

[0888] 111.2 Cas9 TadAins 997

[0889] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDAT

LYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVF
NAQKKAQSSTDGSSGSSGSETPGTSESATPESSGGSSIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYF
FYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVK
KDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDE
IIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDA
TLIHQSITGLYETRIDLSQLGGD

[0890] 112δHNH TadA

[0891] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRA
IGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPG
MNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMN
TKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYD
VRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNI
VKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMER
SSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGS
PEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK
YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0892] 113 N端单TadA螺旋截短165-末端

[0893] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
YFFRMPRSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTD
RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYN
QLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDT
YDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILR
RQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKN
LPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSV
EISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRR

YTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNE
KLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQL
LNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLV
SDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS
NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSD
KLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDL
IIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIE
QISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLI
HQSITGLYETRIDLSQLGGD

[0894] 114 N端单TadA螺旋截短165-末端δ59-65

[0895] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRTAHAEIMALRQGGLVMQNYRLI
DATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPR
SGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKN
LIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIV
DEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN
LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ
SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYP
FLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVL
PKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRL
SRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQ
TVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYL
QNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLIT
QRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDF
QFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFK
TEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKK
DWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKY
SLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSK
RVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGL
YETRIDLSQLGGD

[0896] 115.1 Cas9 TadAins 1004

[0897] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP

LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRA
IGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPG
MNHRVEITEGILADECAALLCYFFRMPRQLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP
KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVIL
ADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGD

[0898] 115.2 Cas9 TadAins 1005

[0899] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNR
AIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCYFFRMPRQESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP
KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVIL
ADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGD

[0900] 115.3 Cas9 TadAins 1006

[0901] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR

RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLEGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWN
RAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHY
PGMNHRVEITEGILADECAALLCYFFRMPRQSEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP
KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVIL
ADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGD

[0902] 115.4 Cas9 TadAins 1007

[0903] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGW
NRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLH
YPGMNHRVEITEGILADECAALLCYFFRMPRQEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP
KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE

LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVIL
ADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGD

[0904] 116.1 Cas9 TadAins C-端截短2 792

[0905] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREV
PVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFG
VRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQSQILKEHPVENTQLQNEKLYLYYLQNGR
DMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYK
VREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP
KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVIL
ADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGD

[0906] 116.2 Cas9 TadAins C-端截短2 791

[0907] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVP
VGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGV
RNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQGSQILKEHPVENTQLQNEKLYLYYLQNGR

DMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYK
VREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP
KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVIL
ADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGD

[0908] 116.3 Cas9 TadAins C-端截短2 790

[0909] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKEGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPV
GAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVR
NAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQLGSQILKEHPVENTQLQNEKLYLYYLQNGR
DMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYK
VREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP
KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVIL
ADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGD

[0910] 117 Cas9δ1017-1069

[0911] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD

NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYSSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLH
DPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHR
VEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKE
SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAK
GYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQH
KHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTST
KEVLDATLIHQSITGLYETRIDLSQLGGD

[0912] 118 Cas9 TadA-CP116ins 1067

[0913] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNMNHR
VEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAK
RARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHS
RIGRVVFGVRNAKTGAAGSLMDVLHYPGGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0914] 119 Cas9 TadAins 701

[0915] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIG
EGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMD
VLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDLTFKEDIQKAQVSGQGDSLHEHIANLAGS
PAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ
NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWR
QLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSK
LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFF
YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0916] 120 Cas9 TadACP136ins 1248

[0917] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTS
ESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0918] 121 Cas9 TadACP136ins 1052

[0919] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAMNHRVEITEGILADECA
ALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAV
LVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAK
TGAAGSLMDVLHYPGNGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0920] 122 Cas9 TadACP136ins 1041

[0921] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSMNHRVEITEGILADECAALLCYFFRMPRQV
FNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWN
RAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHY

PGNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0922] 123 Cas9 TadACP139ins 1299

[0923] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRMNH
RVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLA
KRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIH
SRIGRVVFGVRNAKTGAAGSLMDVLHYPGDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0924] 124 Cas9δ792-872 TadAins

[0925] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRA

IGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPG
MNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHA
HDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRP
LIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT
VAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLA
SAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLS
AYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0926] 125 Cas9δ792-906 TadAins

[0927] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRA
IGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPG
MNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT
KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDV
RKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIV
KKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS
SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSP
EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKY
FDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0928] 126 TadA CP65ins 1003

[0929] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE

MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVL
HYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWM
RHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFF
YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0930] 127 TadA CP65ins 1016

[0931] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNA
KTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESS
GSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPYDVRKMIAKSEQEIGKATAKYFF
YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0932] 128 TadA CP65ins 1022

[0933] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK

PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMITAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRV
VFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSE
SATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPAKSEQEIGKATAKYFF
YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0934] 129 TadA CP65ins 1029

[0935] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEITAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMI
HSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGS
ETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPGKATAKYFF
YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0936] 130 TadA CP65ins 1041

[0937] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSTAHAEIMALRQGGLVMQNYRLIDATLYVTF
EPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKK
AQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLH
DPNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0938] 131 TadA CP65ins 1054

[0939] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANTAHAEIMALRQGGLVM
QNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCY
FFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNN
RVIGEGWNRAIGLHDPGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN

SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0940] 132 TadA CP65ins 1246

[0941] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRN
AKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPES
SGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[0942] 在一些实施方案中,产生了腺苷脱氨酶碱基编辑器,以将TadA或其变体在经鉴定位置处插入到Cas9多肽中。

[0943] 示例性但非限制性的融合蛋白在国际PCT申请号PCT/US2020/016285以及美国专利临时申请号62/852,228和62/852,224中有所描述,其内容通过引用而以其整体并入本文。

[0944] 核碱基编辑结构域

[0945] 本文描述了包含融合蛋白的碱基编辑器,该融合蛋白包括多核苷酸可编程核苷酸结合结构域和核碱基编辑结构域(例如,脱氨酶结构域)。可将碱基编辑器编程,以通过与能够识别靶标序列的向导多核苷酸相互作用来编辑靶标多核苷酸序列中的一个或多个碱基。一旦靶标序列被识别,碱基编辑器就被锚定在待发生编辑之处的多核苷酸上,然后,脱氨酶结构域组分可以编辑靶标碱基。

[0946] 在一些实施方案中,核碱基编辑结构域包括脱氨酶结构域。如本文中具体描述的,脱氨酶结构域包括胞苷脱氨酶或腺苷脱氨酶。在一些实施方案中,“胞嘧啶脱氨酶”和“胞苷

脱氨酶”可以互换地使用。在一些实施方案中,“腺嘌呤脱氨酶”和“腺苷脱氨酶”可以互换地使用。核碱基编辑蛋白的细节在国际PCT申请号PCT/2017/045381(WO2018/027078)和PCT/US2016/058344(WO2017/070632)中描述,其各自通过引用而整体并入本文。也参见,Komor,A.C.,et al.,“Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage”Nature 533,420-424(2016);Gaudelli,N.M.,et al.,“Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage”Nature 551,464-471(2017)和Komor,A.C.,et al.,“Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity”Science Advances 3:eaao4774(2017)),其整体内容通过引用并入本文。

[0947] A到G的编辑

[0948] 在一些实施方案中,本文所述的碱基编辑器可以包含脱氨酶结构域,该脱氨酶结构域包括腺苷脱氨酶。碱基编辑器的此类腺苷脱氨酶结构域可以促进通过将腺嘌呤(A)脱氨基以形成肌苷(I)而A核碱基编辑为鸟嘌呤(G)核碱基,肌苷展现了G的碱基配对特性。腺苷脱氨酶能够将脱氧核糖核酸(DNA)中的脱氧腺苷残基的腺嘌呤脱氨基(即,移除氨基团)。

[0949] 在一些实施方案中,本文所提供的核碱基编辑器可以通过将一个或多个蛋白结构域融合在一起从而产生融合蛋白而制备。在某些实施方案中,本文所提供的融合蛋白包含一种或多种特征,该特征改善融合蛋白的碱基编辑活性(例如,效率、选择性和特异性)。例如,本文所提供的融合蛋白可以包含具有削弱的核酸酶活性的Cas9结构域。在一些实施方案中,本文所提供的融合蛋白可以具有不具有核酸酶活性的Cas9结构域(dCas9)或者被称为Cas9切口酶(nCas9)的切割双螺旋DNA分子一条链的Cas9结构域。不受缚于任何特定理论,催化残基(例如,H840)的存在保持了Cas9裂解含有与所靶向的A相对的T的未经编辑(例如,未经脱氨)的链。Cas9的催化残基的突变(例如,D10到A10)防止含有所靶向的A残基的经编辑的链的裂解。此类Cas9变体能够基于gRNA定义的靶标序列而在具体位置处生成单链DNA断裂(切口),导致未经编辑的链的修复,最终造成未经编辑的链上T到C的变化。在一些实施方案中,A到G碱基编辑器进一步包含肌苷碱基切除修复的抑制剂,例如,尿嘧啶糖苷酶抑制剂(UGI)结构域或无催化活性的肌苷特异性核酸酶。不受缚于任何特定理论,UGI结构域或无催化活性的肌苷特异性核酸酶可以抑制或防止经脱氨基的腺苷残基(例如,肌苷)的碱基切除修复,这可以改善碱基编辑器的活性或效率。

[0950] 包含腺苷脱氨酶的碱基编辑器可以作用于任何多核苷酸,包括DNA、RNA和DNA-RNA杂交物。在某些实施方案中,包含腺苷脱氨酶的碱基编辑器可以将包含RNA的多核苷酸的靶标A脱氨基。例如,碱基编辑器可以包含能够将RNA多核苷酸和/或DNA-RNA杂交多核苷酸的靶标A脱氨基的腺苷脱氨酶结构域。在一种实施方案中,被并入碱基编辑器中的腺苷脱氨酶包含作用于RNA的腺苷脱氨酶(ADAR,例如,ADAR1或ADAR2)的全部或一部分。在另一实施方案中,被并入碱基编辑器中的腺苷脱氨酶包含作用于tRNA的腺苷脱氨酶(ADAT)的全部或一部分。包含腺苷脱氨酶结构域的碱基编辑器也可以能够将DNA多核苷酸的A核碱基脱氨基。在一种实施方案中,碱基编辑器的腺苷脱氨酶包含ADAT的全部或一部分,其包含一个或多个允许ADAT将DNA中的靶标A脱氨基的突变。例如,碱基编辑器可以包含来自大肠杆菌的ADAT(EcTadA)的全部或一部分,其包含以下突变中的一个或多个:D108N、A106V、D147Y、

E155V、L84F、H123Y、I156F或另一腺苷脱氨酶中的相应突变。

[0951] 腺苷脱氨酶可以源自任何合适的生物体(例如,大肠杆菌)。在一些实施方案中,腺苷脱氨酶是天然出现的腺苷脱氨酶,其包括对应于本文所提供的任何突变(例如,ecTadA中的突变)的一个或多个突变。任何同源蛋白质中的相应残基可以通过例如序列比对和同源残基的确定而鉴定。可以据此产生任何天然出现的腺苷脱氨酶(例如,具有与ecTadA的同源性)的对应于本文所述任何突变(例如,ecTadA中鉴定的任何突变)的突变。

[0952] 腺苷脱氨酶

[0953] 在一些实施方案中,本文所述的融合蛋白可以包含脱氨酶结构域,该脱氨酶结构域包括腺苷脱氨酶。碱基编辑器的此类腺苷脱氨酶结构域可以促进通过将腺嘌呤(A)脱氨基以形成肌苷(I)而A核碱基编辑为鸟嘌呤(G)核碱基,肌苷展现了G的碱基配对特性。腺苷脱氨酶能够将脱氧核糖核酸(DNA)中的脱氧腺苷残基的腺嘌呤脱氨基(即,移除氨基团)。

[0954] 在一些实施方案中,本文所提供的腺苷脱氨酶能够将腺嘌呤脱氨基。在一些实施方案中,本文所提供的腺苷脱氨酶能够将DNA的脱氧腺苷残基中的腺嘌呤脱氨基。在一些实施方案中,腺苷脱氨酶是天然出现的腺苷脱氨酶,其包括对应于本文所提供的任何突变(例如,ecTadA中的突变)的一个或多个突变。本领域技术人员将能够鉴定任何同源蛋白质中的相应残基,例如,通过序列比对和同源残基的确定。据此,本领域技术人员将能够产生任何天然出现的腺苷脱氨酶(例如,具有与ecTadA的同源性)的对应于本文所述任何突变(例如,ecTadA中鉴定的任何突变)的突变。在一些实施方案中,腺苷脱氨酶来自原核生物。在一些实施方案中,腺苷脱氨酶来自细菌。在一些实施方案中,腺苷脱氨酶来自大肠杆菌、金黄色葡萄球菌、伤寒沙门氏菌、腐败希瓦氏菌、流感嗜血杆菌、新月柄杆菌或枯草芽孢杆菌。在一些实施方案中,腺苷脱氨酶来自大肠杆菌。

[0955] 本公开提供具有增加的效率(>50-60%)和特异性的腺苷脱氨酶变体。特别地,本文所述的腺苷脱氨酶变体更有可能编辑多核苷酸内的所希望的碱基,并且更不可能编辑不预期改变的碱基(即,“旁观者”)。

[0956] 在特定实施方案中,TadA是PCT/US2017/045381(WO 2018/027078)中描述的任何一种TadA,该专利通过引用以其整体并入本文。

[0957] 在一些实施方案中,本公开的核碱基编辑器是腺苷脱氨酶变体,其包含以下序列中的改变:

[0958] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD(也称为TadA*7.10)。

[0959] 在特定实施方案中,融合蛋白包含单个(例如,作为单体提供)TadA*8变体。在一些实施方案中,TadA*8与Cas9切口酶连接。在一些实施方案中,本公开的融合蛋白包含,作为异源二聚体,野生型TadA(TadA(wt))与TadA*8变体连接。在一些实施方案中,本公开的融合蛋白包含,作为异源二聚体,TadA*7.10与TadA*8变体连接。在一些实施方案中,碱基编辑器是包含TadA*8变体单体的ABE8。在一些实施方案中,碱基编辑器是ABE8,其包含TadA*8变体与TadA(wt)的异源二聚体。在一些实施方案中,碱基编辑器是ABE8,其包含TadA*8变体与TadA*7.10的异源二聚体。在一些实施方案中,碱基编辑器是ABE8,其包含TadA*8变体的异源二聚体。在一些实施方案中,TadA*8选自表8。在一些实施方案中,ABE8选自表8、9或10。相

关序列如下：

[0960] 野生型TadA(TadA(wt))或“TadA参考序列”

[0961] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD

[0962] TadA*7.10：

[0963] MSEVEFSHEYW MRHALTLAKR ARDEREVPVGAVLVLNNRVI GEGWNRAIGL HDPTAHAEIM

[0964] ALRQGGLVMQ NYRLIDATLY VTFEPCVMCA GAMIHSRIGR VVFGVRNAKT GAAGSLMDVL

[0965] HYPGMNHRVE ITEGILADEC AALLCYFFRM PRQVFNAQKK AQSSTD

[0966] 在一些实施方案中,腺苷脱氨酶包含氨基酸序列,所述氨基酸序列与下文提供的任何腺苷脱氨酶中详述的任何一个氨基酸序列为至少60％、至少65％、至少70％、至少75％、至少80％、至少85％、至少90％、至少95％、至少96％、至少97％、至少98％、至少99％或至少99.5％相同。应知悉,本文所提供的腺苷脱氨酶可以包括一个或多个突变(例如,本文所提供的任何突变)。本公开提供任何具有某一百分比的同一性加上任何本文所述突变或其组合的脱氨酶结构域。在一些实施方案中,腺苷脱氨酶包含氨基酸序列,与参考序列或本文所提供的任何腺苷脱氨酶相比,该氨基酸序列具有1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、21、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50或更多个突变。在一些实施方案中,腺苷脱氨酶包含氨基酸序列,与本领域中已知或本文所述的任何一种氨基酸序列相比,该氨基酸序列具有至少5、至少10、至少15、至少20、至少25、至少30、至少35、至少40、至少45、至少50、至少60、至少70、至少80、至少90、至少100、至少110、至少120、至少130、至少140、至少150、至少160或至少170个相同的连续氨基酸残基。

[0967] 在一些实施方案中,TadA脱氨酶是全长大肠杆菌TadA脱氨酶。例如,在某些实施方案中,腺苷脱氨酶包含氨基酸序列：

[0968] MRRAFITGVFFLSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD。

[0969] 但应知悉,可用于本申请的其他腺苷脱氨酶对于技术人员是显而易见的并且处于本公开范畴内。例如,腺苷脱氨酶可以是作用于tRNA(ADAT)的腺苷脱氨酶同源物。非限制性地,示例性ADAT同源物的氨基酸序列包括以下：

[0970] 金黄色葡萄球菌(Staphylococcus aureus)TadA：

[0971] MGSHMTNDIYFMTLAIEEAKKAAQLGEVPIGAIITKDDEVIARAHNLRETLQQPTAHAEHIAIERAAKVLGSWRLEGCTLYVTLEPCVMCAGTIVMSRIPRVVYGADDPKGGCSGSLMNLLQQSNFNHRAIVDKGVLKEACSTLLTTFFKNLRANKKSTN

[0972] 枯草芽孢杆菌(Bacillus subtilis)TadA：

[0973] MTQDELYMKEAIKEAKKAEEKGEVPIGAVLVINGEIIARAHNLRETEQRSIAHAEMLVIDEACKALGTWRLEGATLYVTLEPCPMCAGAVVLSRVEKVVFGAFDPKGGCSGTLMNLLQEERFNHQAEVVSGVLEEECGGMLSAFFRELRKKKKAARKNLSE

[0974] 鼠伤寒沙门氏菌(Salmonella typhimurium(S.typhimurium))TadA：

[0975] MPPAFITGVTSLSDVELDHEYWMRHALTLAKRAWDEREVPVGAVLVHNHRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVLQNYRLLDTTLYVTLEPCVMCAGAMVHSRIGRVVFGARDAKTGAAGSLIDVLHHPGMNHRVEIIEGVLRDECATLLSDFFRMRRQEIKALKKADRAEGAGPAV

[0976] 腐败希瓦氏菌(Shewanella putrefaciens(S.putrefaciens))TadA:

[0977] MDEYWMQVAMQMAEKAEAAGEVPVGAVLVKDGQQIATGYNLSISQHDPTAHAEILCLRSAGKKLENYRLLDATLYITLEPCAMCAGAMVHSRIARVVYGARDEKTGAAGTVVNLLQHPAFNHQVEVTSGVLAEACSAQLSRFFKRRRDEKKALKLAQRAQQGIE

[0978] 流感嗜血杆菌(Haemophilus influenzae)F3031(H.influenzae)TadA:

[0979] MDAAKVRSEFDEKMMRYALELADKAEALGEIPVGAVLVDDARNIIGEGWNLSIVQSDPT A HAEIIALRNGAKNIQNYRLLNSTLYVTLEPCTMCAGAILHSRIKRLVFGASDYKTGAIGSRFHFFDDYKMNHTLEITSGVLAEECSQKLSTFFQKRREEKKIEKALLKSLSDK

[0980] 新月柄杆菌(Caulobacter crescentus(C.crescentus))TadA:

[0981] MRTDESEDQDHRMMRLALDAARAAAEAGETPVGAVILDPSTGEVIATAGNGPIAAHDPTAHAEIAAMRAAAKLGNYRLTDLTLVVTLEPCAMCAGAISHARIGRVVFGADDPKGGAVVHGPKFFAQPTCHWRPEVTGGVLADESADLLRGFFRARRKAKI

[0982] 硫还原地杆菌(Geobacter sulfurreducens(G.sulfurreducens))TadA:

[0983] MSSLKKTPIRDDAYWMGKAIREAAKAAARDEVPIGAVIVRDGAVIGRGHNLREGSNDPSAHAEMIAIRQAARRSANWRLTGATLYVTLEPCLMCMGAIILARLERVVFGCYDPKGGAAGSLYDLSADPRLNHQVRLSPGVCQEECGTMLSDFFRDLRRRKKAKATPALFIDERKVPPEP

[0984] 大肠杆菌TadA(ecTadA)的一种实施方案包括以下:

[0985] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

[0986] 在一些实施方案中,腺苷脱氨酶来自原核生物。在一些实施方案中,腺苷脱氨酶来自细菌。在一些实施方案中,腺苷脱氨酶来自大肠杆菌、金黄色葡萄球菌、伤寒沙门氏菌、腐败希瓦氏菌、流感嗜血杆菌、新月柄杆菌或枯草芽孢杆菌。在一些实施方案中,腺苷脱氨酶来自大肠杆菌。

[0987] 在一种实施方案中,本公开的融合蛋白包含与TadA*7.10连接的野生型TadA,TadA*7.10与Cas9切口酶连接。在特定实施方案中,融合蛋白包含单个TadA*7.10结构域(例如,作为单体提供)。在其他实施方案中,ABE7.10编辑器包含TadA*7.10和TadA(wt),它们能够形成异源二聚体。

[0988] 在一些实施方案中,腺苷脱氨酶包含氨基酸序列,所述氨基酸序列与下文提供的任何腺苷脱氨酶中详述的任何一个氨基酸序列为至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相同。应知悉,本文所提供的腺苷脱氨酶可以包括一个或多个突变(例如,本文所提供的任何突变)。本公开提供任何具有某一百分比的同一性加上任何本文所述突变或其组合的脱氨酶结构域。在一些实施方案中,腺苷脱氨酶包含氨基酸序列,与参考序列或本文所提供的任何腺苷脱氨酶相比,该氨基酸序列具有1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、21、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、

39、40、41、42、43、44、45、46、47、48、49、50或更多个突变。在一些实施方案中,腺苷脱氨酶包含氨基酸序列,与本领域中已知或本文所述的任何一种氨基酸序列相比,该氨基酸序列具有至少5、至少10、至少15、至少20、至少25、至少30、至少35、至少40、至少45、至少50、至少60、至少70、至少80、至少90、至少100、至少110、至少120、至少130、至少140、至少150、至少160或至少170个相同的连续氨基酸残基。

[0989] 应知悉,本文所提供的任何突变(例如,基于TadA参考序列)可以被引入其他腺苷脱氨酶中,诸如大肠杆菌TadA(ecTadA)、金黄色葡萄球菌TadA(saTadA)或其他腺苷脱氨酶(例如,细菌腺苷脱氨酶)。对于技术人员将会是显而易见的是,可以类似地比对其他脱氨酶以鉴定可能如本文所提供的那样突变的同源氨基酸残基。因此,可以在具有同源氨基酸残基的其他腺苷脱氨酶(例如,ecTadA)中做出在TadA参考序列中鉴定的任何突变。也应知悉,本文所提供的任何突变均可以独立地或以任何组合形式在TadA参考序列或另一腺苷脱氨酶中做出。

[0990] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的D108X突变或另一腺苷脱氨酶(例如,ecTadA)中的相应突变,其中X表示除野生型腺苷脱氨酶中的相应氨基酸以外的任何氨基酸。在一些实施方案汇总,腺苷脱氨酶包含D108G、D108N、D108V、D108A或D108Y突变,或者另一腺苷脱氨酶中的相应突变。

[0991] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的A106X突变或另一腺苷脱氨酶(例如,ecTadA)中的相应突变,其中X表示除野生型腺苷脱氨酶中的相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的A106V突变,或另一腺苷脱氨酶(例如,野生型TadA或ecTadA)中的相应突变。

[0992] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的E155X突变或另一腺苷脱氨酶(例如,ecTadA)中的相应突变,其中X的存在表示除野生型腺苷脱氨酶中的相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的E155D、E155G或E155V突变,或另一腺苷脱氨酶(例如,ecTadA)中的相应突变。

[0993] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的D147X突变或另一腺苷脱氨酶(例如,ecTadA)中的相应突变,其中X的存在表示除野生型腺苷脱氨酶中的相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的D147Y突变,或另一腺苷脱氨酶(例如,ecTadA)中的相应突变。

[0994] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的A106X、E155X或D147X突变或另一腺苷脱氨酶(例如,ecTadA)中的相应突变,其中X表示除野生型腺苷脱氨酶中的相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含E155D、E155G或E155V突变。在一些实施方案中,腺苷脱氨酶包含D147Y。

[0995] 例如,腺苷脱氨酶可以含有包含TadA参考序列中的D108N、A106V、155V和/或D147Y突变,或另一腺苷脱氨酶(例如,野生型TadA或ecTadA)中的相应突变。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的以下突变组(突变组通过“;”分隔),或另一腺苷脱氨酶(例如,或ecTadA)中的相应突变:D108N和A106V;D108N和E155V;D108N和D147Y;A106V和E155V;A106V和D147Y;E155V和D147Y;D108N、A106V和E155V;D108N、A106V和D147Y;D108N、E155V和D147Y;A106V、E155V和D147Y;以及D108N、A106V、E155V和D147Y。但应知悉,本文所提供的相应突变的任何组合均可以在腺苷脱氨酶(例如,ecTadA)中做出。

[0996] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H8X、T17X、L18X、W23X、L34X、W45X、R51X、A56X、E59X、E85X、M94X、I95X、V102X、F104X、A106X、R107X、D108X、K110X、M118X、N127X、A138X、F149X、M151X、R153X、Q154X、I156X和/或K157X突变中的一个或多个,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个相应突变,其中X的存在表示除野生型腺苷脱氨酶中的相应氨基酸外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H8Y、T17S、L18E、W23L、L34S、W45L、R51H、A56E、或A56S、E59G、E85K、或E85G、M94L、I95L、V102A、F104L、A106V、R107C、或R107H、或R107P、D108G、或D108N、或D108V、或D108A、或D108Y、K110I、M118K、N127S、A138V、F149Y、M151V、R153C、Q154L、I156D和/或K157R突变中的一个或多个,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个相应突变。

[0997] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H8X、D108X和/或N127X突变中的一个或多个,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个相应突变,其中X表示任何氨基酸的存在。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H8Y、D108N和/或N127S突变中的一个或多个,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个相应突变。

[0998] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H8X、R26X、M61X、L68X、M70X、A106X、D108X、A109X、N127X、D147X、R152X、Q154X、E155X、K161X、Q163X和/或T166X突变中的一个或多个,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个相应突变,其中X表示除野生型腺苷脱氨酶中的相应氨基酸外的任何氨基酸的存在。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H8Y、R26W、M61I、L68Q、M70V、A106T、D108N、A109T、N127S、D147Y、R152C、Q154H或Q154R、E155G或E155V或E155D、K161Q、Q163H和/或T166P突变中的一个或多个,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个相应突变。

[0999] 在一些实施方案中,腺苷脱氨酶包含选自由TadA参考序列中的H8X、D108X、N127X、D147X、R152X和Q154X所组成的组中的一个、两个、三个、四个、五个或六个突变,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个相应突变,其中X表示除野生型腺苷脱氨酶中的相应氨基酸外的任何氨基酸的存在。在一些实施方案中,腺苷脱氨酶包含选自由TadA参考序列中的H8X、M61X、M70X、D108X、N127X、Q154X、E155X和Q163X突变所组成的组中的一个、两个、三个、四个、五个、六个、七个或八个突变,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个相应突变,其中X表示除野生型腺苷脱氨酶中的相应氨基酸外的任何氨基酸的存在。在一些实施方案中,腺苷脱氨酶包含选自由TadA参考序列中的H8X、D108X、N127X、E155X和T166X所组成的组中的一个、两个、三个、四个或五个突变,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个相应突变,其中X表示除野生型腺苷脱氨酶中的相应氨基酸外的任何氨基酸的存在。

[1000] 在一些实施方案中,腺苷脱氨酶包含选自由的H8X、A106X、D108X、另一腺苷脱氨酶中的一个或多个突变所组成的组中的一个、两个、三个、四个、五个或六个突变,其中X表示除野生型腺苷脱氨酶中的相应氨基酸外的任何氨基酸的存在。在一些实施方案中,腺苷脱氨酶包含选自由H8X、R26X、L68X、D108X、N127X、D147X和E155X突变所组成的组中的一个、两个、三个、四个、五个、六个、七个或八个突变,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个相应突变,其中X表示除野生型腺苷脱氨酶中的相应氨基酸外的任何氨基酸的存在。在一些实施方案中,腺苷脱氨酶包含选自由TadA参考序列中的H8X、D108X、A109X、N127X和

E155X所组成的组中的一个、两个、三个、四个或五个突变,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个相应突变,其中X表示除野生型腺苷脱氨酶中的相应氨基酸外的任何氨基酸的存在。

[1001] 在一些实施方案中,腺苷脱氨酶包含选自由TadA参考序列中的H8Y、D108N、N127S、D147Y、R152C和Q154H所组成的组中的一个、两个、三个、四个、五个或六个突变,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个相应突变。在一些实施方案中,腺苷脱氨酶包含选自由TadA参考序列中的H8Y、M61I、M70V、D108N、N127S、Q154R、E155G和Q163H所组成的组中的一个、两个、三个、四个、五个、六个、七个或八个突变,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个相应突变。在一些实施方案中,腺苷脱氨酶包含选自由TadA参考序列中的H8Y、D108N、N127S、E155V和T166P所组成的组中的一个、两个、三个、四个或五个突变,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个相应突变。在一些实施方案中,腺苷脱氨酶包含选自由TadA参考序列中的H8Y、A106T、D108N、N127S、E155D和K161Q所组成的组中的一个、两个、三个、四个、五个或六个突变,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个相应突变。在一些实施方案中,腺苷脱氨酶包含选自由TadA参考序列中的H8Y、R26W、L68Q、D108N、N127S、D147Y和E155V所组成的组中的一个、两个、三个、四个、五个、六个、七个或八个突变,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个相应突变。在一些实施方案中,腺苷脱氨酶包含选自由TadA参考序列中的H8Y、D108N、A109T、N127S和E155G所组成的组中的一个、两个、三个、四个或五个突变,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个相应突变。

[1002] 本文所提供的任何突变和任何其他突变(例如,基于ecTadA氨基酸序列)可以被引入任何其他腺苷脱氨酶中。本文所提供的任何突变均可以独立地或以任何组合形式在TadA参考序列或另一腺苷脱氨酶(例如,ecTadA)中做出。

[1003] A到G的核碱基编辑蛋白质的细节在国际PCT申请号PCT/2017/045381(WO2018/027078)和Gaudelli,N.M.,et al.,"Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage"Nature,551,464-471(2017)中有所描述,其整体内容通过引用并入本文。

[1004] 在一些实施方案中,腺苷脱氨酶包含另一腺苷脱氨酶(例如,ecTadA)中的一个或多个相应突变。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的D108N、D108G或D108V突变,或另一腺苷脱氨酶(例如,ecTadA)中的相应突变。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的A106V和D108N突变,或另一腺苷脱氨酶(例如,ecTadA)中的相应突变。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的R107C和D108N突变,或另一腺苷脱氨酶(例如,ecTadA)中的相应突变。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H8Y、D108N、N127S、D147Y和Q154H突变,或另一腺苷脱氨酶(例如,ecTadA)中的相应突变。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H8Y、D108N、N127S、D147Y和E155V突变,或另一腺苷脱氨酶(例如,ecTadA)中的相应突变。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的D108N、D147Y和E155V突变,或另一腺苷脱氨酶(例如,ecTadA)中的相应突变。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H8Y、D108N和N127S突变,或另一腺苷脱氨酶(例如,ecTadA)中的相应突变。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的A106V、D108N、D147Y和E155V突变,或另一腺苷脱氨酶(例如,ecTadA)中

的相应突变。

[1005] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的S2X、H8X、I49X、L84X、H123X、N127X、I156X和/或K160X突变中的一个或多个,或另一腺苷脱氨酶中的一个或多个相应突变,其中X的存在表示除野生型腺苷脱氨酶中的相应氨基酸外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的S2A、H8Y、I49F、L84F、H123Y、N127S、I156F和/或K160S突变中的一个或多个,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个相应突变。

[1006] 在一些实施方案中,腺苷脱氨酶包含L84X突变腺苷脱氨酶,其中X表示除野生型腺苷脱氨酶中的相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的L84F突变,或另一腺苷脱氨酶(例如,ecTadA)中的相应突变。

[1007] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H123X突变或另一腺苷脱氨酶(例如,ecTadA)中的相应突变,其中X表示除野生型腺苷脱氨酶中的相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H123Y突变,或另一腺苷脱氨酶(例如,ecTadA)中的相应突变。

[1008] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的I156X突变或另一腺苷脱氨酶(例如,ecTadA)中的相应突变,其中X表示除野生型腺苷脱氨酶中的相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的I156F突变,或另一腺苷脱氨酶(例如,ecTadA)中的相应突变。

[1009] 在一些实施方案中,腺苷脱氨酶包含选自由TadA参考序列中的L84X、A106X、D108X、H123X、D147X、E155X和I156X突变所组成的组中的一个、两个、三个、四个、五个、六个或七个突变,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个相应突变,其中X表示除野生型腺苷脱氨酶中的相应氨基酸外的任何氨基酸的存在。在一些实施方案中,腺苷脱氨酶包含选自由TadA参考序列中的S2X、I49X、A106X、D108X、D147X和E155X所组成的组中的一个、两个、三个、四个、五个或六个突变,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个相应突变,其中X表示除野生型腺苷脱氨酶中的相应氨基酸外的任何氨基酸的存在。在一些实施方案中,腺苷脱氨酶包含选自由TadA参考序列中的H8X、A106X、D108X、N127X和K160X所组成的组中的一个、两个、三个、四个或五个突变,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个相应突变,其中X表示除野生型腺苷脱氨酶中的相应氨基酸外的任何氨基酸的存在。

[1010] 在一些实施方案中,腺苷脱氨酶包含选自由TadA参考序列中的L84F、A106V、D108N、H123Y、D147Y、E155V和I156F所组成的组中的一个、两个、三个、四个、五个、六个或七个突变,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个相应突变。在一些实施方案中,腺苷脱氨酶包含选自由TadA参考序列中的S2A、I49F、A106V、D108N、D147Y和E155V所组成的组中的一个、两个、三个、四个、五个或六个突变。

[1011] 在一些实施方案中,腺苷脱氨酶包含选自由TadA参考序列中的H8Y、A106T、D108N、N127S和K160S所组成的组中的一个、两个、三个、四个或五个突变,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个相应突变。

[1012] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的E25X、R26X、R107X、A142X和/或A143X突变中的一个或多个,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个相应

突变,其中X的存在表示除野生型腺苷脱氨酶中的相应氨基酸外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的E25M、E25D、E25A、E25R、E25V、E25S、E25Y、R26G、R26N、R26Q、R26C、R26L、R26K、R107P、R107K、R107A、R107N、R107W、R107H、R107S、A142N、A142D、A142G、A143D、A143G、A143E、A143L、A143W、A143M、A143S、A143Q和/或A143R突变中的一个或多个,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个相应突变。在一些实施方案中,腺苷脱氨酶包含对应于TadA参考序列中的本文所述突变中的一个或多个,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个相应突变。

[1013] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的E25X突变或另一腺苷脱氨酶(例如,ecTadA)中的相应突变,其中X表示除野生型腺苷脱氨酶中的相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的E25M、E25D、E25A、E25R、E25V、E25S或E25Y突变,或另一腺苷脱氨酶(例如,ecTadA)中的相应突变。

[1014] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的R26X突变或另一腺苷脱氨酶(例如,ecTadA)中的相应突变,其中X表示除野生型腺苷脱氨酶中的相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的R26G、R26N、R26Q、R26C、R26L或R26K突变,或另一腺苷脱氨酶(例如,ecTadA)中的相应突变。

[1015] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的R107X突变或另一腺苷脱氨酶(例如,ecTadA)中的相应突变,其中X表示除野生型腺苷脱氨酶中的相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的R107P、R107K、R107A、R107N、R107W、R107H或R107S突变,或另一腺苷脱氨酶(例如,ecTadA)中的相应突变。

[1016] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的A142X突变或另一腺苷脱氨酶(例如,ecTadA)中的相应突变,其中X表示除野生型腺苷脱氨酶中的相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的A142N、A142D、A142G突变,或另一腺苷脱氨酶(例如,ecTadA)中的相应突变。

[1017] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的A143X突变或另一腺苷脱氨酶(例如,ecTadA)中的相应突变,其中X表示除野生型腺苷脱氨酶中的相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的A143D、A143G、A143E、A143L、A143W、A143M、A143S、A143Q和/或A143R突变,或另一腺苷脱氨酶(例如,ecTadA)中的相应突变。

[1018] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H36X、N37X、P48X、I49X、R51X、M70X、N72X、D77X、E134X、S S146X、Q154X、K157X和/或K161X突变中的一个或多个,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个相应突变,其中X的存在表示除野生型腺苷脱氨酶中的相应氨基酸外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H36L、N37T、N37S、P48T、P48L、I49V、R51H、R51L、M70L、N72S、D77G、E134G、S146R、S146C、Q154H、K157N和/或K161T突变中的一个或多个,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个相应突变。

[1019] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H36X突变或另一腺苷脱氨酶(例如,ecTadA)中的相应突变,其中X表示除野生型腺苷脱氨酶中的相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H36L突变,或另一腺苷脱氨酶(例如,ecTadA)中的相应突变。

[1020] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的N37X突变或另一腺苷脱氨酶(例如,ecTadA)中的相应突变,其中X表示除野生型腺苷脱氨酶中的相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的N37T或N37S突变,或另一腺苷脱氨酶(例如,ecTadA)中的相应突变。

[1021] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的P48X突变或另一腺苷脱氨酶(例如,ecTadA)中的相应突变,其中X表示除野生型腺苷脱氨酶中的相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的P48T或P48L突变,或另一腺苷脱氨酶(例如,ecTadA)中的相应突变。

[1022] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的R51X突变或另一腺苷脱氨酶中的相应突变,其中X表示除野生型腺苷脱氨酶中的相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的R51H或R51L突变,或另一腺苷脱氨酶(例如,ecTadA)中的相应突变。

[1023] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的S146X突变或另一腺苷脱氨酶(例如,ecTadA)中的相应突变,其中X表示除野生型腺苷脱氨酶中的相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的S146R或S146C突变,或另一腺苷脱氨酶(例如,ecTadA)中的相应突变。

[1024] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的K157X突变或另一腺苷脱氨酶(例如,ecTadA)中的相应突变,其中X表示除野生型腺苷脱氨酶中的相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的K157N突变,或另一腺苷脱氨酶(例如,ecTadA)中的相应突变。

[1025] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的P48X突变或另一腺苷脱氨酶(例如,ecTadA)中的相应突变,其中X表示除野生型腺苷脱氨酶中的相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的P48S、P48T或P48A突变,或另一腺苷脱氨酶(例如,ecTadA)中的相应突变。

[1026] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的A142X突变或另一腺苷脱氨酶(例如,ecTadA)中的相应突变,其中X表示除野生型腺苷脱氨酶中的相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的A142N突变,或另一腺苷脱氨酶(例如,ecTadA)中的相应突变。

[1027] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的W23X突变或另一腺苷脱氨酶(例如,ecTadA)中的相应突变,其中X表示除野生型腺苷脱氨酶中的相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的W23R或W23L突变,或另一腺苷脱氨酶(例如,ecTadA)中的相应突变。

[1028] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的R152X突变或另一腺苷脱氨酶(例如,ecTadA)中的相应突变,其中X表示除野生型腺苷脱氨酶中的相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的R152P或R52H突变,或另一腺苷脱氨酶(例如,ecTadA)中的相应突变。

[1029] 在一种实施方案中,腺苷脱氨酶可以包含突变H36L、R51L、L84F、A106V、D108N、H123Y、S146C、D147Y、E155V、I156F和K157N。在一些实施方案中,腺苷脱氨酶包含相对于TadA参考序列的突变的下列组合,其中组合的每个突变通过“_”分隔并且突变的每个组合

位于括号内：

[1030] (A106V_D108N)、

[1031] (R107C_D108N)、

[1032] (H8Y_D108N_N127S_D147Y_Q154H)、

[1033] (H8Y_D108N_N127S_D147Y_E155V)、

[1034] (D108N_D147Y_E155V)、

[1035] (H8Y_D108N_N127S)、

[1036] (H8Y_D108N_N127S_D147Y_Q154H)、

[1037] (A106V_D108N_D147Y_E155V)、

[1038] (D108Q_D147Y_E155V)、

[1039] (D108M_D147Y_E155V)、

[1040] (D108L_D147Y_E155V)、

[1041] (D108K_D147Y_E155V)、

[1042] (D108I_D147Y_E155V)、

[1043] (D108F_D147Y_E155V)、

[1044] (A106V_D108N_D147Y)、

[1045] (A106V_D108M_D147Y_E155V)、

[1046] (E59A_A106V_D108N_D147Y_E155V)、

[1047] (E59A催化上死亡的_A106V_D108N_D147Y_E155V)、

[1048] (L84F_A106V_D108N_H123Y_D147Y_E155V_I156Y)、

[1049] (L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)、

[1050] (D103A_D104N)、

[1051] (G22P_D103A_D104N)、

[1052] (D103A_D104N_S138A)、

[1053] (R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F)、

[1054] (E25G_R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F)、

[1055] (E25D_R26G_L84F_A106V_R107K_D108N_H123Y_A142N_A143G_D147Y_E155V_I156F)、

[1056] (R26Q_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F)、

[1057] (E25M_R26G_L84F_A106V_R107P_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F)、

[1058] (R26C_L84F_A106V_R107H_D108N_H123Y_A142N_D147Y_E155V_I156F)、

[1059] (L84F_A106V_D108N_H123Y_A142N_A143L_D147Y_E155V_I156F)、

[1060] (R26G_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F)、

[1061] (E25A_R26G_L84F_A106V_R107N_D108N_H123Y_A142N_A143E_D147Y_E155V_I156F)、

[1062] (R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F)、

[1063] (A106V_D108N_A142N_D147Y_E155V)、

[1064] (R26G_A106V_D108N_A142N_D147Y_E155V)、

[1065] (E25D_R26G_A106V_R107K_D108N_A142N_A143G_D147Y_E155V)、

[1066] (R26G_A106V_D108N_R107H_A142N_A143D_D147Y_E155V)、

[1067] (E25D_R26G_A106V_D108N_A142N_D147Y_E155V)、

[1068] (A106V_R107K_D108N_A142N_D147Y_E155V)、

[1069] (A106V_D108N_A142N_A143G_D147Y_E155V)、

[1070] (A106V_D108N_A142N_A143L_D147Y_E155V)、

[1071] (H36L_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N)、

[1072] (N37T_P48T_M70L_L84F_A106V_D108N_H123Y_D147Y_I49V_E155V_I156F)、

[1073] (N37S_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K161T)、

[1074] (H36L_L84F_A106V_D108N_H123Y_D147Y_Q154H_E155V_I156F)、

[1075] (N72S_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F)、

[1076] (H36L_P48L_L84F_A106V_D108N_H123Y_E134G_D147Y_E155V_I156F)、

[1077] (H36L_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K157N)、

[1078] (H36L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F)、

[1079] (L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F_K161T)、

[1080] (N37S_R51H_D77G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)、

[1081] (R51L_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K157N)、

[1082] (D24G_Q71R_L84F_H96L_A106V_D108N_H123Y_D147Y_E155V_I156F_K160E)、

[1083] (H36L_G67V_L84F_A106V_D108N_H123Y_S146T_D147Y_E155V_I156F)、

[1084] (Q71L_L84F_A106V_D108N_H123Y_L137M_A143E_D147Y_E155V_I156F)、

[1085] (E25G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_Q159L)、

[1086] (L84F_A91T_F104I_A106V_D108N_H123Y_D147Y_E155V_I156F)、

[1087] (N72D_L84F_A106V_D108N_H123Y_G125A_D147Y_E155V_I156F)、

[1088] (P48S_L84F_S97C_A106V_D108N_H123Y_D147Y_E155V_I156F)、

[1089] (W23G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)、

[1090] (D24G_P48L_Q71R_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_Q159L)、

[1091] (L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F)、

[1092] (H36L_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F_K157N)，

[1093] (N37S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_K161T)、(L84F_A106V_D108N_D147Y_E155V_I156F)、

[1094] (R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K161T)、

[1095] (L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K161T)、

[1096] (L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K160E_K161T)、

[1097] (L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K160E)、

[1098] (R74Q_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)、

[1099] (R74A_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)、

[1100] (L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)、

[1101] (R74Q_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)、

[1102] (L84F_R98Q_A106V_D108N_H123Y_D147Y_E155V_I156F)、

[1103] (L84F_A106V_D108N_H123Y_R129Q_D147Y_E155V_I156F)、

[1104] (P48S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F)、(P48S_A142N)、

[1105] (P48T_I49V_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_L157N)、

[1106] (P48T_I49V_A142N)、

[1107] (H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N)、

[1108] (H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S146C_A142N_D147Y_E155V_I156F (H36L_P48T_I49V_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N)、

[1109] (H36L_P48T_I49V_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F_K157N)、

[1110] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N)、

[1111] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F_K157N)、

[1112] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_A142N_D147Y_E155V_I156F_K157N)、

[1113] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N)、

[1114] (W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N)、

[1115] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F_K161T)、

[1116] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152H_E155V_I156F_K157N)、

[1117] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F_K157N)、

[1118] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F_K157N)、

[1119] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142A_S146C_D147Y_E155V_I156F_K157N)、

[1120] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142A_S146C_D147Y_R152P_E155V_I156F_K157N)、

[1121] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F_K161T)、

[1122] (W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F_K157N)、

[1123] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_R152P_E155V_

I156F_K157N)。

[1124] 在某些实施方案中,本文所提供的融合蛋白包含一种或多种特征,该特征改善融合蛋白的碱基编辑活性。例如,本文所提供的任何融合蛋白可以包含具有削弱的核酸酶活性的Cas9结构域。在一些实施方案中,本文所提供的任何融合蛋白可以具有不具有核酸酶活性的Cas9结构域(dCas9)或者被称为Cas9切口酶(nCas9)的切割双螺旋DNA分子一条链的Cas9结构域。

[1125] 在一些实施方案中,腺苷脱氨酶是TadA*7.10。在一些实施方案中,TadA*7.10包含至少一个改变。在特定实施方案中,TadA*7.10包含以下改变中的一种或多种:Y147T、Y147R、Q154S、Y123H、V82S、T166R和Q154R。本文中,改变Y123H也称为H123H(TadA*7.10中的改变H123Y修复为Y123H(wt))。在其他实施方案中,TadA*7.10包含选自以下组的改变的组合:Y147T+Q154R;Y147T+Q154S;Y147R+Q154S;V82S+Q154S;V82S+Y147R;V82S+Q154R;V82S+Y123H;I76Y+V82S;V82S+Y123H+Y147T;V82S+Y123H+Y147R;V82S+Y123H+Q154R;Y147R+Q154R+Y123H;Y147R+Q154R+I76Y;Y147R+Q154R+T166R;Y123H+Y147R+Q154R+I76Y;V82S+Y123H+Y147R+Q154R;and I76Y+V82S+Y123H+Y147R+Q154R。在特定实施方案中,相对于TadA*7.10(TadA参考序列),腺苷脱氨酶变体包含C端的缺失,该缺失起始于残基149、150、151、152、153、154、155、156和157,或另一TadA中的对应突变。

[1126] 在其他实施方案中,本公开的碱基编辑器是包含腺苷脱氨酶变体(例如,TadA*8)的单体,该腺苷脱氨酶变体包含一种或多种以下改变:相对于TadA*7.10(TadA参考序列),Y147T、Y147R、Q154S、Y123H、V82S、T166R和/或Q154R,或另一TadA中的相应突变。在其他实施方案中,腺苷脱氨酶变体(TadA*8)是单体,其包含选自以下组的改变的组合:相对于TadA*7.10(TadA参考序列),Y147T+Q154R、Y147T+Q154S、Y147R+Q154S、V82S+Q154S、V82S+Y147R、V82S+Q154R、V82S+Y123H、I76Y+V82S、V82S+Y123H+Y147T、V82S+Y123H+Y147R、V82S+Y123H+Q154R、Y147R+Q154R+Y123H、Y147R+Q154R+I76Y、Y147R+Q154R+T166R、Y123H+Y147R+Q154R+I76Y、V82S+Y123H+Y147R+Q154R、和I76Y+V82S+Y123H+Y147R+Q154R,或另一TadA中的相应突变。在其他实施方案中,碱基编辑器是异源二聚体,其包含野生型腺苷脱氨酶和腺苷脱氨酶变体(例如,TadA*8),所述腺苷脱氨酶变体结构域包含以下突变中的一种或多种:相对于TadA*7.10(TadA参考序列),Y147T、Y147R、Q154S、Y123H、V82S、T166R和/或Q154R,或另一TadA中的相应突变。在其他实施方案中,碱基编辑器是异源二聚体,其包含TadA*7.10结构域和腺苷脱氨酶变体结构域(例如,TadA*8),所述腺苷脱氨酶变体结构域包含选自以下组的改变的组合:相对于TadA*7.10(TadA参考序列),Y147T+Q154R、Y147T+Q154S、Y147R+Q154S、V82S+Q154S、V82S+Y147R、V82S+Q154R、V82S+Y123H、I76Y+V82S、V82S+Y123H+Y147T、V82S+Y123H+Y147R、V82S+Y123H+Q154R、Y147R+Q154R+Y123H、Y147R+Q154R+I76Y、Y147R+Q154R+T166R、Y123H+Y147R+Q154R+I76Y、V82S+Y123H+Y147R+Q154R、和I76Y+V82S+Y123H+Y147R+Q154R,或另一TadA中的相应突变。

[1127] 在一个实施方案中,腺苷脱氨酶是TadA*8,其包含以下序列或其具有腺苷脱氨酶活性的片段或基本上由其组成:

[1128] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD

[1129] 在一些实施方案中,TadA*8是经截短的。在一些实施方案中,相对于全长度TadA*8,经截短的TadA*8失去了1、2、3、4、5,6、7、8、9、10、11、12、13、14、15、6、17、18、19或20个N端氨基酸残基。在一些实施方案中,相对于全长度TadA*8,经截短的TadA*8失去了1、2、3、4、5,6、7、8、9、10、11、12、13、14、15、6、17、18、19或20个C端氨基酸残基。在一些实施方案中,腺苷脱氨酶变体是全长度TadA*8。

[1130] 在一些实施方案中,TadA*8是TadA*8.1、TadA*8.2、TadA*8.3、TadA*8.4、TadA*8.5、TadA*8.6、TadA*8.7、TadA*8.8、TadA*8.9、TadA*8.10、TadA*8.11、TadA*8.12、TadA*8.13、TadA*8.14、TadA*8.15、TadA*8.16、TadA*8.17、TadA*8.18、TadA*8.19、TadA*8.20、TadA*8.21、TadA*8.22、TadA*8.23或TadA*8.24。

[1131] 在一种实施方案中,本公开的融合蛋白包含与本文所述的腺苷脱氨酶变体(例如,TadA*8)连接的野生型TadA,该腺苷脱氨酶变体与Cas9切口酶连接。在特定实施方案中,融合蛋白包含单个TadA*8结构域(例如,作为单体提供)。在其他实施方案中,碱基编辑器包含TadA*8和TadA(wt),它们能够形成异源二聚体。示例性序列如下:TadA(wt)或“TadA参考序列”:

[1132] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD

[1133] TadA*7.10:

[1134] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

[1135] TadA*8:

[1136] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD。

[1137] 在一些实施方案中,腺苷脱氨酶包含氨基酸序列,所述氨基酸序列与下文提供的任何腺苷脱氨酶中详述的任何一个氨基酸序列为至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相同。应知悉,本文所提供的腺苷脱氨酶可以包括一个或多个突变(例如,本文所提供的任何突变)。本公开提供任何具有某一百分比的同一性加上任何本文所述突变或其组合的脱氨酶结构域。在一些实施方案中,腺苷脱氨酶包含氨基酸序列,与参考序列或本文所提供的任何腺苷脱氨酶相比,该氨基酸序列具有1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、21、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50或更多个突变。在一些实施方案中,腺苷脱氨酶包含氨基酸序列,与本领域中已知或本文所述的任何一种氨基酸序列相比,该氨基酸序列具有至少5、至少10、至少15、至少20、至少25、至少30、至少35、至少40、至少45、至少50、至少60、至少70、至少80、至少90、至少100、至少110、至少120、至少130、至少140、至少150、至少160或至少170个相同的连续氨基酸残基。

[1138] 在特定实施方案中,TadA*8包含位于以粗体显示的以下任何位置处的一个或多个

突变。在其它实施方案中,TadA*8包含位于以下划线显示的任何位置处的一个或多个突变:

MSEVEFSHEY WMRHALTLAK RARDEREVPV GAVLVLNNRV IGEGWNRAIG 50
LHDPTAHAEI MALRQGGLVM QNYRLIDATL YVTFEPCVMC AGAMIHSRIG 100
RVVFGVRNAK TGAAGSLMDV LHYPGMNHRV EITEGILADE CAALLCYFFR 150
MPRQVFNAQK KAQSSTD

[1139] 例如,相对于TadA*7.10(TadA参考序列),TadA*8包含位于氨基酸位置82和/或166处的改变(例如,V82S、T166R),该改变单独存在或与以下Y147T、Y147R、Q154S、Y123H和/或Q154R中的任何一个或多个组合,或另一TadA中的相应突变。在特定实施方案中,改变的组合选自下列的组:相对于TadA*7.10(TadA参考序列),Y147T+Q154R、Y147T+Q154S、Y147R+Q154S、V82S+Q154S、V82S+Y147R、V82S+Q154R、V82S+Y123H、I76Y+V82S、V82S+Y123H+Y147T、V82S+Y123H+Y147R、V82S+Y123H+Q154R、Y147R+Q154R+Y123H、Y147R+Q154R+I76Y、Y147R+Q154R+T166R、Y123H+Y147R+Q154R+I76Y、V82S+Y123H+Y147R+Q154R、和I76Y+V82S+Y123H+Y147R+Q154R,或另一TadA中的相应突变。

[1140] 在一些实施方案中,腺苷脱氨酶是TadA*8,其包含以下序列或其具有腺苷脱氨酶活性的片段或基本上由其组成:

[1141] MSEVEFSHEY WMRHALTLAK RARDEREVPVGAVLVLNNRV IGEGWNRAIG LHDPTAHAEI

[1142] MALRQGGLVM QNYRLIDATL YVTFEPCVMC AGAMIHSRIG

[1143] RVVFGVRNAK TGAAGSLMDV LHYPGMNHRV EITEGILADE CAALLCTFFR

[1144] MPRQVFNAQK KAQSSTD

[1145] 在一些实施方案中,TadA*8是经截短的。在一些实施方案中,相对于全长度TadA*8,经截短的TadA*8失去了1、2、3、4、5,6、7、8、9、10、11、12、13、14、15、6、17、18、19或20个N端氨基酸残基。在一些实施方案中,相对于全长度TadA*8,经截短的TadA*8失去了1、2、3、4、5,6、7、8、9、10、11、12、13、14、15、6、17、18、19或20个C端氨基酸残基。在一些实施方案中,腺苷脱氨酶变体是全长度TadA*8。

[1146] 在一种实施方案中,本公开的融合蛋白包含与本文所述的腺苷脱氨酶变体(例如,TadA*8)连接的野生型TadA,该腺苷脱氨酶变体与Cas9切口酶连接。在特定实施方案中,融合蛋白包含单个TadA*8结构域(例如,作为单体提供)。在其他实施方案中,碱基编辑器包含TadA*8和TadA(wt),它们能够形成异源二聚体。

[1147] C到T的编辑

[1148] 在一些实施方案中,本文所公开的碱基编辑器包含融合蛋白,该融合蛋白包含能够将多核苷酸的靶标胞苷(C)碱基脱氨基以产生尿苷(U)的胞苷脱氨酶,尿苷具有胸腺嘧啶的碱基配对特性。在一些实施方案中,例如,若多核苷酸为双链(例如,DNA),则可将尿苷碱基取代为胸苷碱基(例如,通过细胞修复结构)以引起C:G到T:A的转换。在其他实施方案中,通过碱基编辑器将核酸中的C脱氨基为U不能伴随U到T的置换。

[1149] 将多核苷酸中的靶标C脱氨基以导致U,是一种能够通过本文所述的碱基编辑器执行的碱基编辑类型的非限制性示例。在另一实施例中,包含胞苷脱氨酶结构域的碱基编辑器可以接到胞嘧啶(C)碱基到鸟苷(G)碱基的转化。例如,通过碱基编辑器的胞苷脱氨酶将胞苷脱氨基产生的多核苷酸的U可以通过碱基切除修复机制(例如,通过尿嘧啶DNA糖苷酶

(UDG)结构域)从多核苷酸切除,产生无碱基位点。然后,通可以过例如跨损伤聚合酶将与无碱基位点相对的核碱基置换(例如,通过碱基修复机构)为另一碱基诸如C。尽管将与无碱基位点相对的核碱基替换为C是典型的,但也可能发生其他置换(例如,A、G或T)。

[1150] 据此,在一些实施方案中,本文所述的碱基编辑器包含能够将多核苷酸中的靶标C脱氨基为U的脱氨反应结构域(例如,胞苷脱氨酶结构域)。再者,如下文所述,碱基编辑器可以包含额外结构域,该额外结构域促进脱氨反应所得的U转化为(在一些实施方案中)T或G。例如,包含胞苷脱氨酶结构域的碱基编辑器可以进一步包含尿嘧啶糖苷酶抑制剂(UGI)结构域以介导通过T置换U,完成C到T的碱基编辑事件。在另一实施例中,碱基编辑器可以将跨损伤聚合酶并入以改善C到G的碱基编辑的效率,因为跨损伤聚合酶可以促进与无碱基位点相对的C的并入(即,造成G并入到无碱基位点处,完成C到G的碱基编辑事件)。

[1151] 包含胞苷脱氨酶作为结构域的碱基编辑器可以将任何多核苷酸中的靶标C脱氨基,该多核苷酸包括DNA、RNA和DNA-RNA杂交物。典型地,胞苷脱氨酶催化定位在多核苷酸的单链部分情境中的C核碱基。在一些实施方案中,包含靶标C的整体多核苷酸可以是单链的。例如,被并入碱基编辑器中的胞苷脱氨酶可以将单链RNA多核苷酸中的靶标C脱氨基。在其他实施方案中,包含胞苷脱氨酶结构域的碱基编辑器可以作用于双链多核苷酸,但靶标C可以定位在多核苷酸的一部分中,该部分在脱氨基反应发生时处于单链状态。例如,在其中NAGPB结构域包含Cas9结构域的实施方案中,在Cas9-gRNA-靶标DNA复合物的形成过程中,可以留下几个未配对的核苷酸,导致Cas9“R-环复合物”的形成。这些未配对的核苷酸可以形成单链DNA泡,其充当单链特异性核苷酸脱氨酶(例如,胞苷脱氨酶)的底物。

[1152] 在一些实施方案中,碱基编辑器的胞苷脱氨酶可以包含载脂蛋白BmRNA编辑复合物(APOBEC)家族脱氨酶的全部或一部分。APOBEC是进化上保守的胞苷脱氨酶家族。这一家族的成员是C至U编辑酶。APOBEC样蛋白的N端结构域是催化结构域,而C端结构域是假催化结构域。更详细地,催化结构域是锌依赖性胞苷脱氨酶结构域,并且对于胞苷脱氨反应而言至关重要。APOBEC家族成员包括APOBEC1、APOBEC2、APOBEC3A、APOBEC3B、APOBEC3C、APOBEC3D(现在称为“APOBEC3E”)、APOBEC3F、APOBEC3G、APOBEC3H、APOBEC4和激活诱导的(胞苷)脱氨酶。在一些实施方案中,被并入碱基编辑器的脱氨酶包含APOBEC1脱氨酶的全部或一部分。在一些实施方案中,被并入碱基编辑器的脱氨酶包含APOBEC2脱氨酶的全部或一部分。在一些实施方案中,被并入碱基编辑器的脱氨酶包含APOBEC3脱氨酶的全部或一部分。在一些实施方案中,被并入碱基编辑器的脱氨酶包含APOBEC3A脱氨酶的全部或一部分。在一些实施方案中,被并入碱基编辑器的脱氨酶包含APOBEC3B脱氨酶的全部或一部分。在一些实施方案中,被并入碱基编辑器的脱氨酶包含APOBEC3C脱氨酶的全部或一部分。在一些实施方案中,被并入碱基编辑器的脱氨酶包含APOBEC3D脱氨酶的全部或一部分。在一些实施方案中,被并入碱基编辑器的脱氨酶包含APOBEC3E脱氨酶的全部或一部分。在一些实施方案中,被并入碱基编辑器的脱氨酶包含APOBEC3F脱氨酶的全部或一部分。在一些实施方案中,被并入碱基编辑器的脱氨酶包含APOBEC3G脱氨酶的全部或一部分。在一些实施方案中,被并入碱基编辑器的脱氨酶包含APOBEC3H脱氨酶的全部或一部分。在一些实施方案中,被并入碱基编辑器的脱氨酶包含APOBEC4脱氨酶的全部或一部分。在一些实施方案中,被并入碱基编辑器的脱氨酶包含激活诱导的脱氨酶(AID)的全部或一部分。在一些实施方案中,被并入碱基编辑器的脱氨酶包含胞苷脱氨酶1(CDA1)的全部或一部分。应知悉,碱基

编辑器可以包含来自任何合适的生物体(例如,人或大鼠)的脱氨酶。在一些实施方案中,碱基编辑器的脱氨酶来自人、黑猩猩、大猩猩、猴、牛、狗、大鼠或小鼠。在一些实施方案中,碱基编辑器的脱氨酶结构域源自大鼠(例如,大鼠APOBEC1)。在一些实施方案中,碱基编辑器的脱氨酶结构域是人APOBEC1。在一些实施方案中,碱基编辑器的脱氨酶结构域是pmCDA1。

[1153] PmCDA1的氨基酸和核酸序列如下文所示。

[1154] >tr|A5H718|A5H718_PETMA胞嘧啶脱氨酶OS=海七鳃鳗OX=7757PE=2SV=1氨基酸序列:

[1155] MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQSGTERGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLKIWACKLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKRRSELSIMIQVKILHTTKSPAV核酸序列:>EF094822.1海七鳃鳗分离物PmCDA.21胞嘧啶脱氨酶mRNA,完全cds:

[1156] TGACACGACACAGCCGTGTATATGAGGAAGGGTAGCTGGATGGGGGGGGGGGGAATACGTTCAGAGAGGACATTAGCGAGCGTCTTGTTGGTGGCCTTGAGTCTAGACACCTGCAGACATGACCGACGCTGAGTACGTGAGAATCCATGAGAAGTTGGACATCTACACGTTTAAGAAACAGTTTTTCAACAACAAAAAATCCGTGTCGCATAGATGCTACGTTCTCTTTGAATTAAAACGACGGGGTGAACGTAGAGCGTGTTTTTGGGGCTATGCTGTGAATAAACCACAGAGCGGGACAGAACGTGGAATTCACGCCGAAATCTTTAGCATTAGAAAAGTCGAAGAATACCTGCGCGACAACCCCGGACAATTCACGATAAATTGGTACTCATCCTGGAGTCCTTGTGCAGATTGCGCTGAAAAGATCTTAGAATGGTATAACCAGGAGCTGCGGGGGAACGGCCACACTTTGAAAATCTGGGCTTGCAAACTCTATTACGAGAAAAATGCGAGGAATCAAATTGGGCTGTGGAACCTCAGAGATAACGGGGTTGGGTTGAATGTAATGGTAAGTGAACACTACCAATGTTGCAGGAAAATATTCATCCAATCGTCGCACAATCAATTGAATGAGAATAGATGGCTTGAGAAGACTTTGAAGCGAGCTGAAAAACGACGGAGCGAGTTGTCCATTATGATTCAGGTAAAAATACTCCACACCACTAAGAGTCCTGCTGTTTAAGAGGCTATGCGGATGGTTTTC

[1157] 人激活诱导型胞苷脱氨酶(AID)的编码序列(CDS)的氨基酸和核酸序列如下所示。

[1158] >tr|Q6QJ80|Q6QJ80_HUMAN激活诱导的胞苷脱氨酶OS=智人OX=9606GN=AICDAPE=2SV=1氨基酸序列:

[1159] MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKAPV

[1160] 人激活诱导型胞苷脱氨酶(AID)的编码序列(CDS)的氨基酸和核酸序列如下所示。

[1161] >tr|Q6QJ80|Q6QJ80_HUMAN激活诱导的胞苷脱氨酶OS=智人OX=9606GN=AICDAPE=2SV=1氨基酸序列:

[1162] MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKAPV

[1163] 核酸序列:>NG_011588.1:5001-15681智人激活诱导的胞苷脱氨酶(AICDA),染色体12上的RefSeqGene(LRG_17):

[1164] AGAGAACCATCATTAATTGAAGTGAGATTTTTCTGGCCTGAGACTTGCAGGGAGGCAAGAAGACACTCTGGACACCACTATGGACAGGTAAAGAGGCAGTCTTCTCGTGGGTGATTGCACTGGCCTTCCTCTCAGAGCAAATCTGAGTAATGAGACTGGTAGCTATCCCTTTCTCTCATGTAACTGTCTGACTGATAAGATCAGCTTGATCAATATGCATATATATTTTTTGATCTGTCTCCTTTTCTTCTATTCAGATCTTATACGCTGTCAGCCCAATTCTTTCTGTTTCAGACTTCTCTTGATTTCCCTCTTTTTCATGTGGCAAAAGAAGTAGTGCGTACAATGTACTGATTCGTCCTGAGATTTGTACCATG

GTTGAAACTAATTTATGGTAATAATATTAACATAGCAAATCTTTAGAGACTCAAATCATGAAAAGGTAATAGCAGTA
CTGTACTAAAAACGGTAGTGCTAATTTTCGTAATAATTTTGTAAATATTCAACAGTAAAACAACTTGAAGACACACT
TTCCTAGGGAGGCGTTACTGAAATAATTTAGCTATAGTAAGAAAATTTGTAATTTTAGAAATGCCAAGCATTCTAAA
TTAATTGCTTGAAAGTCACTATGATTGTGTCCATTATAAGGAGACAAATTCATTCAAGCAAGTTATTTAATGTTAAA
GGCCCAATTGTTAGGCAGTTAATGGCACTTTTACTATTAACTAATCTTTCCATTTGTTCAGACGTAGCTTAACTTAC
CTCTTAGGTGTGAATTTGGTTAAGGTCCTCATAATGTCTTTATGTGCAGTTTTTGATAGGTTATTGTCATAGAACTT
ATTCTATTCCTACATTTATGATTACTATGGATGTATGAGAATAACACCTAATCCTTATACTTTACCTCAATTTAACT
CCTTTATAAAGAACTTACATTACAGAATAAAGATTTTTTAAAAATATATTTTTTTGTAGAGACAGGGTCTTAGCCCA
GCCGAGGCTGGTCTCTAAGTCCTGGCCCAAGCGATCCTCCTGCCTGGGCCTCCTAAAGTGCTGGAATTATAGACATG
AGCCATCACATCCAATATACAGAATAAAGATTTTTAATGGAGGATTTAATGTTCTTCAGAAAATTTTCTTGAGGTCA
GACAATGTCAAATGTCTCCTCAGTTTACACTGAGATTTTGAAAACAAGTCTGAGCTATAGGTCCTTGTGAAGGGTCC
ATTGGAAATACTTGTTCAAAGTAAAATGGAAAGCAAAGGTAAAATCAGCAGTTGAAATTCAGAGAAAGACAGAAAAG
GAGAAAAGATGAAATTCAACAGGACAGAAGGGAAATATATTATCATTAAGGAGGACAGTATCTGTAGAGCTCATTAG
TGATGGCAAAATGACTTGGTCAGGATTATTTTTAACCCGCTTGTTTCTGGTTTGCACGGCTGGGGATGCAGCTAGGG
TTCTGCCTCAGGGAGCACAGCTGTCCAGAGCAGCTGTCAGCCTGCAAGCCTGAAACACTCCCTCGGTAAAGTCCTTC
CTACTCAGGACAGAAATGACGAGAACAGGGAGCTGGAAACAGGCCCCTAACCAGAGAAGGGAAGTAATGGATCAACA
AAGTTAACTAGCAGGTCAGGATCACGCAATTCATTTCACTCTGACTGGTAACATGTGACAGAAACAGTGTAGGCTTA
TTGTATTTTCATGTAGAGTAGGACCCAAAAATCCACCCAAAGTCCTTTATCTATGCCACATCCTTCTTATCTATACT
TCCAGGACACTTTTTCTTCCTTATGATAAGGCTCTCTCTCTCTCCACACACACACACACACACACACACACACACAC
ACACACACACACAAACACACACCCCGCCAACCAAGGTGCATGTAAAAAGATGTAGATTCCTCTGCCTTTCTCATCTA
CACAGCCCAGGAGGGTAAGTTAATATAAGAGGGATTTATTGGTAAGAGATGATGCTTAATCTGTTTAACACTGGGCC
TCAAAGAGAGAATTTCTTTTCTTCTGTACTTATTAAGCACCTATTATGTGTTGAGCTTATATATACAAAGGGTTATT
ATATGCTAATATAGTAATAGTAATGGTGGTTGGTACTATGGTAATTACCATAAAAATTATTATCCTTTTAAAATAAA
GCTAATTATTATTGGATCTTTTTTAGTATTCATTTTATGTTTTTTATGTTTTTGATTTTTTAAAAGACAATCTCACC
CTGTTACCCAGGCTGGAGTGCAGTGGTGCAATCATAGCTTTCTGCAGTCTTGAACTCCTGGGCTCAAGCAATCCTCC
TGCCTTGGCCTCCCAAAGTGTTGGGATACAGTCATGAGCCACTGCATCTGGCCTAGGATCCATTTAGATTAAAATAT
GCATTTTAAATTTTAAAATAATATGGCTAATTTTTACCTTATGTAATGTGTATACTGGCAATAAATCTAGTTTGCTG
CCTAAAGTTTAAAGTGCTTTCCAGTAAGCTTCATGTACGTGAGGGGAGACATTTAAAGTGAAACAGACAGCCAGGTG
TGGTGGCTCACGCCTGTAATCCCAGCACTCTGGGAGGCTGAGGTGGGTGGATCGCTTGAGCCCTGGAGTTCAAGACC
AGCCTGAGCAACATGGCAAAACGCTGTTTCTATAACAAAAATTAGCCGGGCATGGTGGCATGTGCCTGTGGTCCCAG
CTACTAGGGGGCTGAGGCAGGAGAATCGTTGGAGCCCAGGAGGTCAAGGCTGCACTGAGCAGTGCTTGCGCCACTGC
ACTCCAGCCTGGGTGACAGGACCAGACCTTGCCTCAAAAAAATAAGAAGAAAAATTAAAAATAAATGGAAACAACTA
CAAAGAGCTGTTGTCCTAGATGAGCTACTTAGTTAGGCTGATATTTTGGTATTTAACTTTTAAAGTCAGGGTCTGTC
ACCTGCACTACATTATTAAAATATCAATTCTCAATGTATATCCACACAAAGACTGGTACGTGAATGTTCATAGTACC
TTTATTCACAAAACCCCAAAGTAGAGACTATCCAAATATCCATCAACAAGTGAACAAATAAACAAAATGTGCTATAT
CCATGCAATGGAATACCACCCTGCAGTACAAAGAAGCTACTTGGGGATGAATCCCAAAGTCATGACGCTAAATGAAA
GAGTCAGACATGAAGGAGGAGATAATGTATGCCATACGAAATTCTAGAAAATGAAAGTAACTTATAGTTACAGAAAG
CAAATCAGGGCAGGCATAGAGGCTCACACCTGTAATCCCAGCACTTTGAGAGGCCACGTGGGAAGATTGCTAGAACT
CAGGAGTTCAAGACCAGCCTGGGCAACACAGTGAAACTCCATTCTCCACAAAAATGGGAAAAAAAGAAAGCAAATCA

GTGGTTGTCCTGTGGGGAGGGGAAGGACTGCAAAGAGGGAAGAAGCTCTGGTGGGGTGAGGGTGGTGATTCAGGTTC
TGTATCCTGACTGTGGTAGCAGTTTGGGGTGTTTACATCCAAAAATATTCGTAGAATTATGCATCTTAAATGGGTGG
AGTTTACTGTATGTAAATTATACCTCAATGTAAGAAAAAATAATGTGTAAGAAAACTTTCAATTCTCTTGCCAGCAA
ACGTTATTCAAATTCCTGAGCCCTTTACTTCGCAAATTCTCTGCACTTCTGCCCCGTACCATTAGGTGACAGCACTA
GCTCCACAAATTGGATAAATGCATTTCTGGAAAAGACTAGGGACAAAATCCAGGCATCACTTGTGCTTTCATATCAA
CCATGCTGTACAGCTTGTGTTGCTGTCTGCAGCTGCAATGGGGACTCTTGATTTCTTTAAGGAAACTTGGGTTACCA
GAGTATTTCCACAAATGCTATTCAAATTAGTGCTTATGATATGCAAGACACTGTGCTAGGAGCCAGAAAACAAAGAG
GAGGAGAAATCAGTCATTATGTGGGAACAACATAGCAAGATATTTAGATCATTTTGACTAGTTAAAAAAGCAGCAGA
GTACAAAATCACACATGCAATCAGTATAATCCAAATCATGTAAATATGTGCCTGTAGAAAGACTAGAGGAATAAACA
CAAGAATCTTAACAGTCATTGTCATTAGACACTAAGTCTAATTATTATTATTAGACACTATGATATTTGAGATTTAA
AAAATCTTTAATATTTTAAAATTTAGAGCTCTTCTATTTTTCCATAGTATTCAAGTTTGACAATGATCAAGTATTAC
TCTTTCTTTTTTTTTTTTTTTTTTTTTTTTTGAGATGGAGTTTTGGTCTTGTTGCCCATGCTGGAGTGGAATGGCAT
GACCATAGCTCACTGCAACCTCCACCTCCTGGGTTCAAGCAAAGCTGTCGCCTCAGCCTCCCGGGTAGATGGGATTA
CAGGCGCCCACCACCACACTCGGCTAATGTTTGTATTTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGT
CTCAAACTCCTGACCTCAGAGGATCCACCTGCCTCAGCCTCCCAAAGTGCTGGGATTACAGATGTAGGCCACTGCGC
CCGGCCAAGTATTGCTCTTATACATTAAAAAACAGGTGTGAGCCACTGCGCCCAGCCAGGTATTGCTCTTATACATT
AAAAAATAGGCCGGTGCAGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAAGCCAAGGCGGGCAGAACACCCGAGG
TCAGGAGTCCAAGGCCAGCCTGGCCAAGATGGTGAAACCCCGTCTCTATTAAAAATACAAACATTACCTGGGCATGA
TGGTGGGCGCCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGGATCCGCGGAGCCTGGCAGATCTGCCTGAGC
CTGGGAGGTTGAGGCTACAGTAAGCCAAGATCATGCCAGTATACTTCAGCCTGGGCGACAAAGTGAGACCGTAACAA
AAAAAAAAAAATTTAAAAAAAGAAATTTAGATCAAGATCCAACTGTAAAAAGTGGCCTAAACACCACATTAAAGAGT
TTGGAGTTTATTCTGCAGGCAGAAGAGAACCATCAGGGGGTCTTCAGCATGGGAATGGCATGGTGCACCTGGTTTTT
GTGAGATCATGGTGGTGACAGTGTGGGGAATGTTATTTTGGAGGGACTGGAGGCAGACAGACCGGTTAAAAGGCCAG
CACAACAGATAAGGAGGAAGAAGATGAGGGCTTGGACCGAAGCAGAGAAGAGCAAACAGGGAAGGTACAAATTCAAG
AAATATTGGGGGGTTTGAATCAACACATTTAGATGATTAATTAAATATGAGGACTGAGGAATAAGAAATGAGTCAAG
GATGGTTCCAGGCTGCTAGGCTGCTTACCTGAGGTGGCAAAGTCGGGAGGAGTGGCAGTTTAGGACAGGGGGCAGTT
GAGGAATATTGTTTTGATCATTTTGAGTTTGAGGTACAAGTTGGACACTTAGGTAAAGACTGGAGGGGAAATCTGAA
TATACAATTATGGGACTGAGGAACAAGTTTATTTTATTTTTTGTTTCGTTTTCTTGTTGAAGAACAAATTTAATTGT
AATCCCAAGTCATCAGCATCTAGAAGACAGTGGCAGGAGGTGACTGTCTTGTGGGTAAGGGTTTGGGGTCCTTGATG
AGTATCTCTCAATTGGCCTTAAATATAAGCAGGAAAAGGAGTTTATGATGGATTCCAGGCTCAGCAGGGCTCAGGAG
GGCTCAGGCAGCCAGCAGAGGAAGTCAGAGCATCTTCTTTGGTTTAGCCCAAGTAATGACTTCCTTAAAAAGCTGAA
GGAAAATCCAGAGTGACCAGATTATAAACTGTACTCTTGCATTTTCTCTCCCTCCTCTCACCCACAGCCTCTTGATG
AACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCCGCTGGGCTAAGGGTCGGCGTGAGACCTACCTGTGCTACGT
AGTGAAGAGGCGTGACAGTGCTACATCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGGTATCAATTAAAGTCG
GCTTTGCAAGCAGTTTAATGGTCAACTGTGAGTGCTTTTAGAGCCACCTGCTGATGGTATTACTTCCATCCTTTTTT
GGCATTTGTGTCTCTATCACATTCCTCAAATCCTTTTTTTTATTTCTTTTTCCATGTCCATGCACCCATATTAGACA
TGGCCCAAAATATGTGATTTAATTCCTCCCCAGTAATGCTGGGCACCCTAATACCACTCCTTCCTTCAGTGCCAAGA
ACAACTGCTCCCAAACTGTTTACCAGCTTTCCTCAGCATCTGAATTGCCTTTGAGATTAATTAAGCTAAAAGCATTT
TTATATGGGAGAATATTATCAGCTTGTCCAAGCAAAAATTTTAAATGTGAAAAACAAATTGTGTCTTAAGCATTTTT

GAAAATTAAGGAAGAAGAATTTGGGAAAAAATTAACGGTGGCTCAATTCTGTCTTCCAAATGATTTCTTTTCCCTCC
TACTCACATGGGTCGTAGGCCAGTGAATACATTCAACATGGTGATCCCCAGAAAACTCAGAGAAGCCTCGGCTGATG
ATTAATTAAATTGATCTTTCGGCTACCCGAGAGAATTACATTTCCAAGAGACTTCTTCACCAAAATCCAGATGGGTT
TACATAAACTTCTGCCCACGGGTATCTCCTCTCTCCTAACACGCTGTGACGTCTGGGCTTGGTGGAATCTCAGGGAA
GCATCCGTGGGGTGGAAGGTCATCGTCTGGCTCGTTGTTTGATGGTTATATTACCATGCAATTTTCTTTGCCTACAT
TTGTATTGAATACATCCCAATCTCCTTCCTATTCGGTGACATGACACATTCTATTTCAGAAGGCTTTGATTTTATCA
AGCACTTTCATTTACTTCTCATGGCAGTGCCTATTACTTCTCTTACAATACCCATCTGTCTGCTTTACCAAAATCTA
TTTCCCCTTTTCAGATCCTCCCAAATGGTCCTCATAAACTGTCCTGCCTCCACCTAGTGGTCCAGGTATATTTCCAC
AATGTTACATCAACAGGCACTTCTAGCCATTTTCCTTCTCAAAAGGTGCAAAAAGCAACTTCATAAACACAAATTAA
ATCTTCGGTGAGGTAGTGTGATGCTGCTTCCTCCCAACTCAGCGCACTTCGTCTTCCTCATTCCACAAAAACCCATA
GCCTTCCTTCACTCTGCAGGACTAGTGCTGCCAAGGGTTCAGCTCTACCTACTGGTGTGCTCTTTTGAGCAAGTTGC
TTAGCCTCTCTGTAACACAAGGACAATAGCTGCAAGCATCCCCAAAGATCATTGCAGGAGACAATGACTAAGGCTAC
CAGAGCCGCAATAAAAGTCAGTGAATTTTAGCGTGGTCCTCTCTGTCTCTCCAGAACGGCTGCCACGTGGAATTGCT
CTTCCTCCGCTACATCTCGGACTGGGACCTAGACCCTGGCCGCTGCTACCGCGTCACCTGGTTCACCTCCTGGAGCC
CCTGCTACGACTGTGCCCGACATGTGGCCGACTTTCTGCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCG
CGCCTCTACTTCTGTGAGGACCGCAAGGCTGAGCCCGAGGGGCTGCGGCGGCTGCACCGCGCCGGGGTGCAAATAGC
CATCATGACCTTCAAAGGTGCGAAAGGGCCTTCCGCGCAGGCGCAGTGCAGCAGCCCGCATTCGGGATTGCGATGCG
GAATGAATGAGTTAGTGGGGAAGCTCGAGGGGAAGAAGTGGGCGGGGATTCTGGTTCACCTCTGGAGCCGAAATTAA
AGATTAGAAGCAGAGAAAAGAGTGAATGGCTCAGAGACAAGGCCCCGAGGAAATGAGAAAATGGGGCCAGGGTTGCT
TCTTTCCCCTCGATTTGGAACCTGAACTGTCTTCTACCCCCATATCCCCGCCTTTTTTTCCTTTTTTTTTTTTTGAA
GATTATTTTTACTGCTGGAATACTTTTGTAGAAAACCACGAAAGAACTTTCAAAGCCTGGGAAGGGCTGCATGAAAA
TTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCCTTTTGGTAAGGGGCTTCCTCGCTTTTTAAATTTTCTTTCTTT
CTCTACAGTCTTTTTTGGAGTTTCGTATATTTCTTATATTTTCTTATTGTTCAATCACTCTCAGTTTTCATCTGATG
AAAACTTTATTTCTCCTCCACATCAGCTTTTTCTTCTGCTGTTTCACCATTCAGAGCCCTCTGCTAAGGTTCCTTTT
CCCTCCCTTTTCTTTCTTTTGTTGTTTCACATCTTTAAATTTCTGTCTCTCCCCAGGGTTGCGTTTCCTTCCTGGTC
AGAATTCTTTTCTCCTTTTTTTTTTTTTTTTTTTTTTTTTTAAACAAACAAACAAAAAACCCAAAAAAACTCTTTC
CCAATTTACTTTCTTCCAACATGTTACAAAGCCATCCACTCAGTTTAGAAGACTCTCCGGCCCCACCGACCCCCAAC
CTCGTTTTGAAGCCATTCACTCAATTTGCTTCTCTCTTTCTCTACAGCCCCTGTATGAGGTTGATGACTTACGAGAC
GCATTTCGTACTTTGGGACTTTGATAGCAACTTCCAGGAATGTCACACACGATGAAATATCTCTGCTGAAGACAGTG
GATAAAAAACAGTCCTTCAAGTCTTCTCTGTTTTTATTCTTCAACTCTCACTTTCTTAGAGTTTACAGAAAAAATAT
TTATATACGACTCTTTAAAAAGATCTATGTCTTGAAAATAGAGAAGGAACACAGGTCTGGCCAGGGACGTGCTGCAA
TTGGTGCAGTTTTGAATGCAACATTGTCCCCTACTGGGAATAACAGAACTGCAGGACCTGGGAGCATCCTAAAGTGT
CAACGTTTTTCTATGACTTTTAGGTAGGATGAGAGCAGAAGGTAGATCCTAAAAAGCATGGTGAGAGGATCAAATGT
TTTTATATCAACATCCTTTATTATTTGATTCATTTGAGTTAACAGTGGTGTTAGTGATAGATTTTTCTATTCTTTTC
CCTTGACGTTTACTTTCAAGTAACACAAACTCTTCCATCAGGCCATGATCTATAGGACCTCCTAATGAGAGTATCTG
GGTGATTGTGACCCCAAACCATCTCTCCAAAGCATTAATATCCAATCATGCGCTGTATGTTTTAATCAGCAGAAGCA
TGTTTTTATGTTTGTACAAAAGAAGATTGTTATGGGTGGGGATGGAGGTATAGACCATGCATGGTCACCTTCAAGCT
ACTTTAATAAAGGATCTTAAAATGGGCAGGAGGACTGTGAACAAGACACCCTAATAATGGGTTGATGTCTGAAGTAG
CAAATCTTCTGGAAACGCAAACTCTTTTAAGGAAGTCCCTAATTTAGAAACACCCACAAACTTCACATATCATAATT

AGCAAACAATTGGAAGGAAGTTGCTTGAATGTTGGGGAGAGGAAAATCTATTGGCTCTCGTGGGTCTCTTCATCTCA
GAAATGCCAATCAGGTCAAGGTTTGCTACATTTTGTATGTGTGTGATGCTTCTCCCAAAGGTATATTAACTATATAA
GAGAGTTGTGACAAAACAGAATGATAAAGCTGCGAACCGTGGCACACGCTCATAGTTCTAGCTGCTTGGGAGGTTGA
GGAGGGAGGATGGCTTGAACACAGGTGTTCAAGGCCAGCCTGGGCAACATAACAAGATCCTGTCTCTCAAAAAAAAA
AAAAAAAAAAAGAAAGAGAGAGGGCCGGGCGTGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGCCGG
GCGGATCACCTGTGGTCAGGAGTTTGAGACCAGCCTGGCCAACATGGCAAAACCCCGTCTGTACTCAAAATGCAAAA
ATTAGCCAGGCGTGGTAGCAGGCACCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCAG
GAGGTGGAGGTTGCAGTAAGCTGAGATCGTGCCGTTGCACTCCAGCCTGGGCGACAAGAGCAAGACTCTGTCTCAGA
AAAAAAAAAAAAAAAGAGAGAGAGAGAGAAAGAGAACAATATTTGGGAGAGAAGGATGGGGAAGCATTGCAAGGAAA
TTGTGCTTTATCCAACAAAATGTAAGGAGCCAATAAGGGATCCCTATTTGTCTCTTTTGGTGTCTATTTGTCCCTAA
CAACTGTCTTTGACAGTGAGAAAAATATTCAGAATAACCATATCCCTGTGCCGTTATTACCTAGCAACCCTTGCAAT
GAAGATGAGCAGATCCACAGGAAAACTTGAATGCACAACTGTCTTATTTTAATCTTATTGTACATAAGTTTGTAAAA
GAGTTAAAAATTGTTACTTCATGTATTCATTTATATTTTATATTATTTTGCGTCTAATGATTTTTTATTAACATGAT
TTCCTTTTCTGATATATTGAAATGGAGTCTCAAAGCTTCATAAATTTATAACTTTAGAAATGATTCTAATAACAACG
TATGTAATTGTAACATTGCAGTAATGGTGCTACGAAGCCATTTCTCTTGATTTTTAGTAAACTTTTATGACAGCAAA
TTTGCTTCTGGCTCACTTTCAATCAGTTAAATAAATGATAAATAATTTTGGAAGCTGTGAAGATAAAATACCAAATA
AAATAATATAAAAGTGATTTATATGAAGTTAAAATAAAAAATCAGTATGATGGAATAAACTTG

[1165] 根据本公开的多个方面,可以融合至Cas19的其他示例性脱氨酶提供于下。在一些实施方案中,脱氨酶是激活诱导的脱氨酶(AID)。应理解,在一些实施方案中,可使用各序列的活性结构域,例如,没有定位信号(例如,核定位序列、没有核输出信号、细胞质定位信号)的结构域。

[1166] 人AID:

[1167] MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELL
FLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRK
AEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRIL
LPLYEVDDLRDAFRTLGL (下划线:核定位序列;双下划线:核输出信号)

[1168] 小鼠AID:

[1169] MDSLLMKQKKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSCSLDFGHLRNKSGCHVELL
FLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVAEFLRWNPNLSLRIFTARLYFCEDRK
AEPEGLRRLHRAGVQIGIMTFKDYFYCWNTFVENRERTFKAWEGLHENSVRLTRQLRRIL
LPLYEVDDLRDAFRMLGF (下划线:核定位序列;双下划线:核输出信号)

[1170] 犬AID:

[1171] MDSLLMKQRKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLDFGHLRNKSGCHVELL
FLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFAARLYFCEDRK
AEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENREKTFKAWEGLHENSVRLSRQLRRIL
LPLYEVDDLRDAFRTLGL (下划线:核定位序列;双下划线:核输出信号)牛AID:

MDSLLKKQRQFLYQFKNVRWAKGRHETYLCYVVKRRDSPTSFSLDFGHLRNKAGCHVELL

[1172] FLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFTARLYFCDKER

KAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRI

LLPLYEVDDLRDAFRTLGL (下划线:核定位序列;双下划线:核输出信号)大鼠AID:

MAVGSKPKAALVGPHWERERIWCFLCSTGLGTQQTGQTSRWLRPAATQDPVSPPRSLLMK

[1173] QRKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLDFGYLRNKSGCHVELLFLRYISD

WDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLTGWGALPAGLMSPA

RPSDYFYCWNTFVENHERTFKAWEGLHENSVRLSRRLRRILLPLYEVDDLRDAFRTLGL

(下划线:核定位序列;双下划线:核输出信号)

[1174] clAID家犬亚种(Canis lupus familiaris):

[1175] MDSLLMKQRKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLDFGHLRNKSGCHVELLFLRYISDWD LDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFAARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFY CWNTFVENREKTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL

[1176] btAID黄牛(Bos Taurus):

[1177] MDSLLKKQRQFLYQFKNVRWAKGRHETYLCYVVKRRDSPTSFSLDFGHLRNKAGCHVELLFLRYISDWD LDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFTARLYFCDKERKAEPEGLRRLHRAGVQIAIMTFKDYF YCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL

[1178] mAID小家鼠(Mus musculus):

[1179] MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELLFLRYISDWD LDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFY CWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL

[1180] rAPOBEC-1褐家鼠(Rattus norvegicus):

[1181] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEK FTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTE QESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHIL WATGLK

[1182] maAPOBEC-1金黄仓鼠(Mesocricetus auratus):

[1183] MSSETGPVVVDPTLRRRIEPHEFDAFFDQGELRKETCLLYEIRWGGRHNIWRHTGQNTSRHVEINFIEK FTSERYFYPSTRCSIVWFLSWSPCGECSKAITEFLSGHPNVTLFIYAARLYHHTDQRNRQGLRDLISRGVTIRIMTE QEYCYCWRNFVNYPPSNEVYWPRYPNLWMRLYALELYCIHLGLPPCLKIKRRHQYPLTFFRLNLQSCHYQRIPPHIL WATGFI

[1184] ppAPOBEC-1猩猩(Pongo pygmaeus):

[1185] MTSEKGPSTGDPTLRRRIESWEFDVFYDPRELRKETCLLYEIKWGMSRKIWRSSGKNTTNHVEVNFIKK FTSERRFHSSISCSITWFLSWSPCWECSQAIREFLSQHPGVTLVIYVARLFWHMDQRNRQGLRDLVNSGVTIQIMRA SEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKISRRWQNHLAFFRLHLQNCHYQTIPPHIL LATGLIHPSVTWR

[1186] ocAPOBEC1穴兔(Oryctolagus cuniculus)：

[1187] MASEKGPSNKDYTLRRRIEPWEFEVFFDPQELRKEACLLYEIKWGASSKTWRSSGKNTTNHVEVNFLEKLTSEGRLGPSTCCSITWFLSWSPCWECSMAIREFLSQHPGVTLIIFVARLFQHMDRRNRQGLKDLVTSGVTVRVMSVSEYCYCWENFVNYPPGKAAQWPRYPPRWMLMYALELYCIILGLPPCLKISRRHQKQLTFFSLTPQYCHYKMIPPYILLATGLLQPSVPWR

[1188] mdAPOBEC-1灰短尾负鼠(Monodelphis domestica)：

[1189] MNSKTGPSVGDATLRRRIKPWEFVAFFNPQELRKETCLLYEIKWGNQNIWRHSNQNTSQHAEINFMEKFTAERHFNSSVRCSITWFLSWSPCWECSKAIRKFLDHYPNVTLAIFISRLYWHMDQQHRQGLKELVHSGVTIQIMSYSEYHYCWRNFVDYPQGEEDYWPKYPYLWIMLYVLELHCIILGLPPCLKISGSHSNQLALFSLDLQDCHYQKIPYNVLVATGLVQPFVTWR

[1190] ppAPOBEC-2猩猩(Pongo pygmaeus)：

[1191] MAQKEEAAAATEAASQNGEDLENLDDPEKLKELIELPPFEIVTGERLPANFFKFQFRNVEYSSGRNKTFLCYVVEAQGKGGQVQASRGYLEDEHAAAHAEEAFFNTILPAFDPALRYNVTWYVSSSPCAACADRIIKTLSKTKNLRLLILVGRLFMWEELEIQDALKKLKEAGCKLRIMKPQDFEYVWQNFVEQEEGESKAFQPWEDIQENFLYYEEKLADILK

[1192] btAPOBEC-2黄牛(Bos Taurus)：

[1193] MAQKEEAAAAAEPASQNGEEVENLEDPEKLKELIELPPFEIVTGERLPAHYFKFQFRNVEYSSGRNKTFLCYVVEAQSKGGQVQASRGYLEDEHATNHAEEAFFNSIMPTFDPALRYMVTWYVSSSPCAACADRIVKTLNKTKNLRLLILVGRLFMWEEPEIQAALRKLKEAGCRLRIMKPQDFEYIWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADILK

[1194] mAPOBEC-3-(1)小家鼠(Mus musculus)：

[1195] MQPQRLGPRAGMGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLHHGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQIVRFLATHHNLSLDIFSSRLYNVQDPETQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWKRLLTNFRYQDSKLQEILRPCYISVPSSSSSTLSNICLTKGLPETRFWVEGRRMDPLSEEEFYSQFYNQRVKHLCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKGKQHAEILFLDKIRSMELSQVTITCYLTWSPCPNCAWQLAAFKRDRPDLILHIYTSRLYFHWKRPFQKGLCSLWQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLRRIKESWGLQDLVNDFGNLQLGPPMS

[1196] 小鼠APOBEC-3-(2)：

MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLHHGVFKNKDNI*HAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFEC*AEQIVRFLATHHNLSLDIFSSRLYNVQDPETQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWK

[1197] RLLTNFRYQDSKLQEILRPCYIPVPSSSSSTLSNICLTKGLPETRFCVEGRRMDPLSEEEFYSQFYNQRVKHLCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKGKQ*HAEILFLDKIRSMELSQVTITCYLTWSPCPNC*AWQLAAFKRDRPDLILHIYTSRLYFHWKRPFQKGLCSLWQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLRRIKESWGLQDLVNDFGNLQLGPPMS(斜体:核酸编辑结构域)

[1198] 大鼠APOBEC-3：

MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNRLRYAIDRKDTFLCYEVTRKDCDSPVSL
HHGVFKNKDNI*HAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFEC*AEQVLRFLATH
HNLSLDIFSSRLYNIRDPENQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPW

[1199] KKLLTNFRYQDSKLQEILRPCYIPVPSSSSSTLSNICLTKGLPETRFCVERRRVHLLSEE
EFYSQFYNQRVKHLCYYHGVKPYLCYQLEQFNGQAPLKGCLLSEKGKQ*HAEILFLDKIRS*
*MELSQVIITCYLTWSPCPNC*AWQLAAFKRDRPDLILHIYTSRLYFHWKRPFQKGLCSLWQ
SGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLHRIKESWGLQDLVNDF
GNLQLGPPMS(斜体:核酸编辑结构域)

[1200] hAPOBEC-3A智人(Homo sapiens):

[1201] MEASPASGPRHLMDPHIFTSNFNNGIGRHKTYLCYEVERLDNGTSVKMDQHRGFLHNQAKNLLCGFYGR
HAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGEVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDA
GAQVSIMTYDEFKHCWDTFVDHQGCPFQPWDGLDEHSQALSGRLRAILQNQGN

[1202] hAPOBEC-3F智人(Homo sapiens):

[1203] MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPRLDAKIFRGQVYSQPEHHAEMC
FLSWFCGNQLPAYKCFQITWFVSWTPCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERDYRRALCRLSQAGARVKI
MDDEEFAYCWENFVYSEGQPFMPWYKFDDNYAFLHRTLKEILRNPMEAMYPHIFYFHFKNLRKAYGRNESWLCFTME
VVKHHSPVSWKRGVFRNQVDPETHCHAERCFLSWFCDDILSPNTNYEVTWYTSWSPCPECAGEVAEFLARHSNVNLT
IFTARLYYFWDTDYQEGLRSLSQEGASVEIMGYKDFKYCWENFVYNDDEPFKPWKGLKYNFLFLDSKLQEILE

[1204] 恒河猴(Rhesus macaque)APOBEC-3G:
<u>MVEPMDPRTFVSNFNNRPILSGLNTVWLCCEVKTKDPSGPPLDAKIFQGKVYSKAKYHPE
MRFLRWFHKWRQLHHDQEYKVTWYVSWSPCTRCANSVATFLAKDPKVTLTIFVARLYYFW</u>

[1205] KPDYQQALRILCQKRGGPHATMKIMNYNEFQDCWNKFVDGRGKPFKPRNNLPKHYTLLQA
TLGELLRHLMDPGTFTSNFNNKPWVSGQHETYLCYKVERLHNDTWVPLNQHRGFLRNQAP
NIHGFPKGRHAELCFLDLIPFWKLDGQQYRVTCFTSWSPCFSCAQEMAKFISNNEHVSLC
IFAARIYDDQGRYQEGLRALHRDGAKIAMMNYSEFEYCWDTFVDRQGRPFQPWDGLDEHS
QALSGRLRAI(斜体:核酸编辑结构域;下划线:细胞质定位信号)

[1206] 黑猩猩APOBEC-3G:
<u>MKPHFRNPVERMYQDTFSDNFYNRPILSHRNTVWLCYEVKTKGPSRPPLDAKIFRGQVYS</u>
KLKY*HPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKC*TRDVATFLAEDPKVTLTIFV
ARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEPWNNLPK

[1207] YYILLHIMLGEILRHSMDPPTFTSNFNNELWVRGRHETYLCYEVERLHNDTWVLLNQRRG
FLCNQAPHKHGFLEGR*HAELCFLDVIPFWKLDLHQDYRVTCFTSWSPCFSC*AQEMAKFIS
NNKHVSLCIFAARIYDDQGRCQEGLRTLAKAGAKISIMTYSEFKHCWDTFVDHQGCPFQP
WDGLEEHSQALSGRLRAILQNQGN

[1208] (斜体:核酸编辑结构域;下划线:细胞质定位信号)

[1209] 绿猴APOBEC-3G:

MNPQIRNMVEQMEPDIFVYYFNNRPILSGRNTVWLCYEVKTKDPSGPPLDANIFQGKLYP

EAKD*HPEMKFLHWFRKWRQLHRDQEYEVTWYVSWSPCTRC*ANSVATFLAEDPKVTLTIFV

ARLYYFWKPDYQQALRILCQERGGPHATMKIMNYNEFQHCWNEFVDGQGKPFKPRKNLPK

[1210] HYTLLHATLGELLRHVMDPGTFTSNFNNKPWVSGQRETYLCYKVERSHNDTWVLLNQHRG

FLRNQAPDRHGFPKGR*HAELCFLDLIPFWKLDDQQYRVTCFTSWSPCFSC*AQKMAKFISN

NKHVSLCIFAARIYDDQGRCQEGLRTLHRDGAKIAVMNYSEFEYCWDTFVDRQGRPFQPW

DGLDEHSQALSGRLRAI

[1211] (斜体:核酸编辑结构域;下划线:细胞质定位信号)

[1212] 人APOBEC-3G:

MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQVYS

ELKY*HPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKC*TRDMATFLAEDPKVTLTIFV

ARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEPWNNLPK

[1213] YYILLHIMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRG

FLCNQAPHKHGFLEGR*HAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSC*AQEMAKFIS

KNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVDHQGCPFQP

WDGLDEHSQDLSGRLRAILQNQEN

[1214] (斜体:核酸编辑结构域;下划线:细胞质定位信号)

[1215] 人APOBEC-3F:

MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPRLDAKIFRGQVYS

QPEH*HAEMCFLSWFCGNQLPAYKCFQITWFVSWTPCPDC*VAKLAEFLAEHPNVTLTISAA

RLYYYWERDYRRALCRLSQAGARVKIMDDEEFAYCWENFVYSEGQPFMPWYKFDDNYAFL

[1216] HRTLKEILRNPMEAMYPHIFYFHFKNLRKAYGRNESWLCFTMEVVKHHSPVSWKRGVFRN

QVDPETH*CHAERCFLSWFCDDILSPNTNYEVTWYTSWSPCPEC*AGEVAEFLARHSNVNLT

IFTARLYYFWDTDYQEGLRSLSQEGASVEIMGYKDFKYCWENFVYNDDEPFKPWKGLKYN

FLFLDSKLQEILE

[1217] (斜体:核酸编辑结构域)

[1218] 人APOBEC-3B:

MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGQVY

[1219] FKPQY*HAEMCFLSWFCGNQLPAYKCFQITWFVSWTPCPDC*VAKLAEFLSEHPNVTLTISA

ARLYYYWERDYRRALCRLSQAGARVTIMDYEEFAYCWENFVYNEGQQFMPWYKFDENYAF

LHRTLKEILRYLMDPDTFTFNFNNDPLVLRRRQTYLCYEVERLDNGTWVLMDQHMGFLCN

EAKNLLCGFY*GRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGC*AGEVRAFLQEN

[1220] THVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVYRQGCPFQPWD GLEEHSQALSGRLRAILQNQGN

[1221] (斜体:核酸编辑结构域)

[1222] 大鼠APOBEC-3B:

[1223] MQPQGLGPNAGMGPVCLGCSHRRPYSPIRNPLKKLYQQTFYFHFKNVRYAWGRKNNFLCYEVNGMDCAL PVPLRQGVFRKQGHIHAELCFIYWFHDKVLRVLSPMEEFKVTWYMSWSPCSKCAEQVARFLAAHRNLSLAIFSSRLY YYLRNPNYQQKLCRLIQEGVHVAAMDLPEFKKCWNKFVDNDGQPFRPWMRLRINFSFYDCKLQEIFSRMNLLREDVF YLQFNNSHRVKPVQNRYYRRKSYLCYQLERANGQEPLKGYLLYKKGEQHVEILFLEKMRSMELSQVRITCYLTWSPC PNCARQLAAFKKDHPDLILRIYTSRLYFWRKKFQKGLCTLWRSGIHVDVMDLPQFADCWTNFVNPQRPFRPWNELEK NSWRIQRRLRRIKESWGL

[1224] 牛APOBEC-3B:

[1225] DGWEVAFRSGTVLKAGVLGVSMTEGWAGSGHPGQGACVWTPGTRNTMNLLREVLFKQQFGNQPRVPAPY YRRKTYLCYQLKQRNDLTLDRGCFRNKKQRHAERFIDKINSLDLNPSQSYKIICYITWSPCPNCANELVNFITRNNH LKLEIFASRLYFHWIKSFKMGLQDLQNAGISVAVMTHTEFEDCWEQFVDNQSRPFQPWDKLEQYSASIRRRLQRILT API

[1226] 黑猩猩APOBEC-3B:

[1227] MNPQIRNPMEWMYQRTFYYNFENEPILYGRSYTWLCYEVKIRRGHSNLLWDTGVFRGQMYSQPEHHAEM CFLSWFCGNQLSAYKCFQITWFVSWTPCPDCVAKLAKFLAEHPNVTLTISAARLYYYWERDYRRALCRLSQAGARVK IMDDEEFAYCWENFVYNEGQPFMPWYKFDDNYAFLHRTLKEIIRHLMDPDTFTFNFNNDPLVLRRHQTYLCYEVERL DNGTWVLMDQHMGFLCNEAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGQVRAFLQEN THVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVYRQGCPFQPWDGLEEHSQALSGRLRAIL QVRASSLCMVPHRPPPPPQSPGPCLPLCSEPPLGSLLPTGRPAPSLPFLLTASFSFPPPASLPPLPSLSLSPGHLPV PSFHSLTSCSIQPPCSSRIRETEGWASVSKEGRDLG

[1228] 人APOBEC-3C:

[1229] MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFRNQVD SETH*CHAERCFLSWFCDDILSPNTKYQVTWYTSWSPCPDC*AGEVAEFLARHSNVNLTIFT

[1230] ARLYYFQYPCYQEGLRSLSQEGVAVEIMDYEDFKYCWENFVYNDNEPFKPWKGLKTNFRL LKRRLRESLQ

[1231] (斜体:核酸编辑结构域)

[1232] 大猩猩APOBEC-3C

[1233] MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFRNQVD SETH*CHAERCFLSWECDDILSPNTNYQVTWYTSWSPCPEC*AGEVAEFLARHSNVNLTIFT ARLYYFQDTDYQEGLRSLSQEGVAVKIMDYKDFKYCWENFVYNDDEPFKPWKGLKYNFRF LKRRLQEILE

[1234] (斜体:核酸编辑结构域)

[1235] 人APOBEC-3A:

[1236] MEASPASGPRHLMDPHIFTSNFNNGIGRHKTYLCYEVERLDNGTSVKMDQHRGFLHNQAK NLLCGFYGR*HAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGC*AGEVRAFLQENTHV RLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFKHCWDTFVDHQGCPFQPWDGLD EHSQALSGRLRAILQNQGN

[1237] (斜体:核酸编辑结构域)

[1238] 恒河猴(Rhesus macaque)APOBEC-3A:

[1239] MDGSPASRPRHLMDPNTFTFNFNNDLSVRGRHQTYLCYEVERLDNGTWVPMDERRGFLCN KAKNVPCGDYGC*HVELRFLCEVPSWQLDPAQTYRVTWFISWSPC*FRRGCAGQVRVFLQEN KHVRLRIFAARIYDYDPLYQEALRTLRDAGAQVSIMTYEEFKHCWDTFVDRQGRPFQPWD GLDEHSQALSGRLRAILQNQGN

[1240] (斜体:核酸编辑结构域)

[1241] 牛APOBEC-3A:

[1242] MDEYTFTENFNNQGWPSKTYLCYEMERLDGDATIPLDEYKGFVRNKGLDQPEKPC*HAELY FLGKIHSWNLDRNQHYRLTCFISWSPC*YDCAQKLTTFLKENHHISLHILASRIYTHNRFG CHQSGLCELQAAGARITIMTFEDFKHCWETFVDHKGKPFQPWEGLNVKSQALCTELQAIL KTQQN

[1243] (斜体:核酸编辑结构域)

[1244] 人APOBEC-3H:

[1245] MALLTAETFRLQFNNKRRLRRPYYPRKALLCYQLTPQNGSTPTRGYFENKKKC*HAEICFI NEIKSMGLDETQCYQVTCYLTWSPCSSC*AWELVDFIKAHDHLNLGIFASRLYYHWCKPQQ

[1246] KGLRLLCGSQVPVEVMGFPKFADCWENFVDHEKPLSFNPYKMLEELDKNSRAIKRRLERI KIPGVRAQGRYMDILCDAEV

[1247] (斜体:核酸编辑结构域)

[1248] 恒河猴(Rhesus macaque)APOBEC-3H:

[1249] MALLTAKTFSLQFNNKRRVNKPYYPRKALLCYQLTPQNGSTPTRGHLKNKKKDHAEIRFINKIKSMGLD ETQCYQVTCYLTWSPCPSCAGELVDFIKAHRHLNLRIFASRLYYHWRPNYQEGLLLLCGSQVPVEVMGLPEFTDCWE NFVDHKEPPSFNPSEKLEELDKNSQAIKRRLERIKSRSVDVLENGLRSLQLGPVTPSSSIRNSR

[1250] 人APOBEC-3D:

MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGPVL
PKRQSNHRQEVYFRFEN*HAEMCFLSWFCGNRLPANRRFQITWFVSWNPC*LPCVVKVTKFL
AEHPNVTLTISAARLYYYRDRDWRWVLLRLHKAGARVKIMDYEDFAYCWENFVCNEGQPF

[1251] MPWYKFDDNYASLHRTLKEILRNPMEAMYPHIFYFHFKNLLKACGRNESWLCFTMEVTKH
HSAVFRKRGVFRNQVDPETHC*HAERCFLSWFCDDILSPNTNYEVTWYTSWSPCPEC*AGEV
AEFLARHSNVNLTIFTARLCYFWDTDYQEGLCSLSQEGASVKIMGYKDFVSCWKNFVYSD
DEPFKPWKGLQTNFRLLKRRLREILQ

[1252] (斜体:核酸编辑结构域)

[1253] 人APOBEC-1:

[1254] MTSEKGPSTGDPTLRRRIEPWEFDVFYDPRELRKEACLLYEIKWGMSRKIWRSSGKNTTNHVEVNFIKK
FTSERDFHPSMSCSITWFLSWSPCWECSQAIREFLSRHPGVTLVIYVARLFWHMDQQNRQGLRDLVNSGVTIQIMRA
SEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKISRRWQNHLTFFRLHLQNCHYQTIPPHIL
LATGLIHPSVAWR

[1255] 小鼠APOBEC-1:

[1256] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSVWRHTSQNTSNHVEVNFLEK
FTTERYFRPNTRCSITWFLSWSPCGECSRAITEFLSRHPYVTLFIYIARLYHHTDQRNRQGLRDLISSGVTIQIMTE
QEYCYCWRNFVNYPPSNEAYWPRYPHLWVKLYVLELYCIILGLPPCLKILRRKQPQLTFFTITLQTCHYQRIPPHLL
WATGLK

[1257] 大鼠APOBEC-1:

[1258] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEK
FTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTE
QESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHIL
WATGLK

[1259] 人APOBEC-2:

[1260] MAQKEEAAVATEAASQNGEDLENLDDPEKLKELIELPPFEIVTGERLPANFFKFQFRNVEYSSGRNKTF
LCYVVEAQGKGGQVQASRGYLEDEHAAAHAEEAFFNTILPAFDPALRYNVTWYVSSSPCAACADRIIKTLSKTKNLR
LLILVGRLFMWEEPEIQAALKKLKEAGCKLRIMKPQDFEYVWQNFVEQEEGESKAFQPWEDIQENFLYYEEKLADIL
K

[1261] 小鼠APOBEC-2:

[1262] MAQKEEAAEAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFRNVEYSSGRNKTF
LCYVVEVQSKGGQAQATQGYLEDEHAGAHAEEAFFNTILPAFDPALKYNVTWYVSSSPCAACADRILKTLSKTKNLR
LLILVSRLFMWEEPEVQAALKKLKEAGCKLRIMKPQDFEYIWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADIL
K

[1263] 大鼠APOBEC-2:

[1264] MAQKEEAAEAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFRNVEYSSGRNKTF
LCYVVEAQSKGGQVQATQGYLEDEHAGAHAEEAFFNTILPAFDPALKYNVTWYVSSSPCAACADRILKTLSKTKNLR
LLILVSRLFMWEEPEVQAALKKLKEAGCKLRIMKPQDFEYLWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADIL

K

[1265] 牛APOBEC-2:

[1266] MAQKEEAAAAAEPASQNGEEVENLEDPEKLKELIELPPFEIVTGERLPAHYFKFQFRNVEYSSGRNKTF LCYVVEAQSKGGQVQASRGYLEDEHATNHAEEAFFNSIMPTFDPALRYMVTWYVSSSPCAACADRIVKTLNKTKNLR LLILVGRLFMWEEPEIQAALRKLKEAGCRLRIMKPQDFEYIWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADIL K

[1267] 海七鳃鳗(Petromyzon marinus)CDA1(pmCDA1):

[1268] MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQSGTERGIHAEI FSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLKIWACKLYYEKNARNQIGLWNLRDNGV GLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKRRSELSFMIQVKILHTTKSPAV

[1269] 人APOBEC3G D316R D317R:

[1270] MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQVYSELKYHPEMR FFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPKVTLTIFVARLYYFWDPDYQEALRSLCQKRDGPR ATMKFNYDEFQHCWSKFVYSQRELFEPWNNLPKYYILLHFMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYLCYEV ERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISK KHVSLCIFTARIYRRQGRCQEGLRTLAEAGAKISFTYSEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQ NQEN

[1271] 人APOBEC3G链A:

[1272] MDPPTFTFNFNNEPWWGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVI PFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISFTYSEFK HCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQ

[1273] 人APOBEC3G链A D120R D121R:

[1274] MDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDV IPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIYRRQGRCQEGLRTLAEAGAKISFMTYSE FKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQ

[1275] hAPOBEC-4智人(Homo sapiens):

[1276] MEPIYEEYLANHGTIVKPYYWLSFSLDCSNCPYHIRTGEEARVSLTEFCQIFGFPYGTTFPQTKHLTFY ELKTSSGSLVQKGHASSCTGNYIHPESMLFEMNGYLDSAIYNNDSIRHIILYSNNSPCNEANHCCISKMYNFLITYP GITLSIYFSQLYHTEMDFPASAWNREALRSLASLWPRVVLSPISGGIWHSVLHSFISGVSGSHVFQPILTGRALADR HNAYEINAITGVKPYFTDVLLQTKRNPNTKAQEALESYPLNNAFPGQFFQMPSGQLQPNLPPDLRAPVVFVLVPLRD LPPMHMGQNPNKPRNIVRHLNMPQMSFQETKDLGRLPTGRSVEIVEITEQFASSKEADEKKKKKGKK

[1277] mAPOBEC-4小家鼠(Mus musculus):

[1278] MDSLLMKQKKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSCSLDFGHLRNKSGCHVELLFLRYISDWD LDPGRCYRVTWFTSWSPCYDCARHVAEFLRWNPNLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIGIMTFKDYFY CWNTFVENRERTFKAWEGLHENSVRLTRQLRRILLPLYEVDDLRDAFRMLGF

[1279] rAPOBEC-4褐家鼠(Rattus norvegicus):

[1280] MEPLYEEYLTHSGTIVKPYYWLSVSLNCTNCPYHIRTGEEARVPYTEFHQTFGFPWSTYPQTKHLTFYE LRSSSGNLIQKGLASNCTGSHTHPESMLFERDGYLDSLIFHDSNIRHIILYSNNSPCDEANHCCISKMYNFLMNYPE VTLSVFFSQLYHTENQFPTSAWNREALRGLASLWPQVTLSAISGGIWQSILETFVSGISEGLTAVRPFTAGRTLTDR

YNAYEINCITEVKPYFTDALHSWQKENQDQKVWAASENQPLHNTTPAQWQPDMSQDCRTPAVFMLVPYRDLPPIHVN PSPQKPRTVVRHLNTLQLSASKVKALRKSPSGRPVKKEEARKGSTRSQEANETNKSKWKKQTLFIKSNICHLLEREQ KKIGILSSWSV

[1281] mfAPOBEC-4食蟹猴(Macaca fascicularis)：

[1282] MEPTYEEYLANHGTIVKPYYWLSFSLDCSNCPYHIRTGEEARVSLTEFCQIFGFPYGTTYPQTKHLTFY ELKTSSGSLVQKGHASSCTGNYIHPESMLFEMNGYLDSAIYNNDSIRHIILYCNNSPCNEANHCCISKVYNFLITYP GITLSIYFSQLYHTEMDFPASAWNREALRSLASLWPRVVLSPISGGIWHSVLHSFVSGVSGSHVFQPILTGRALTDR YNAYEINAITGVKPFFTDVLLHTKRNPNTKAQMALESYPLNNAFPGQSFQMTSGIPPDLRAPVVFVLLPLRDLPPMH MGQDPNKPRNIIRHLNMPQMSFQETKDLERLPTRRSVETVEITERFASSKQAEEKTKKKKGKK

[1283] pmCDA-1海七鳃鳗(Petromyzon marinus)：

[1284] MAGYECVRVSEKLDFDTFEFQFENLHYATERHRTYVIFDVKPQSAGGRSRRLWGYIINNPNVCHAELIL MSMIDRHLESNPGVYAMTWYMSWSPCANCSSKLNPWLKNLLEEQGHTLTMHFSRIYDRDREGDHRGLRGLKHVSNSF RMGVVGRAEVKECLAEYVEASRRTLTWLDTTESMAAKMRRKLFCILVRCAGMRESGIPLHLFTLQTPLLSGRVVWWR V

[1285] pmCDA-2海七鳃鳗(Petromyzon marinus)：

[1286] MELREVVDCALASCVRHEPLSRVAFLRCFAAPSQKPRGTVILFYVEGAGRGVTGGHAVNYNKQGTSIHA EVLLLSAVRAALLRRRRCEDGEEATRGCTLHCYSTYSPCRDCVEYIQEFGASTGVRVVIHCCRLYELDVNRRRSEAE GVLRSLSRLGRDFRLMGPRDAIALLLGGRLANTADGESGASGNAWVTETNVVEPLVDMTGFGDEDLHAQVQRNKQIR EAYANYASAVSLMLGELHVDPDKFPFLAEFLAQTSVEPSGTPRETRGRPRGASSRGPEIGRQRPADFERALGAYGLF LHPRIVSREADREEIKRDLIVVMRKHNYQGP

[1287] pmCDA-5海七鳃鳗(Petromyzon marinus)：

[1288] MAGDENVRVSEKLDFDTFEFQFENLHYATERHRTYVIFDVKPQSAGGRSRRLWGYIINNPNVCHAELIL MSMIDRHLESNPGVYAMTWYMSWSPCANCSSKLNPWLKNLLEEQGHTLMMHFSRIYDRDREGDHRGLRGLKHVSNSF RMGVVGRAEVKECLAEYVEASRRTLTWLDTTESMAAKMRRKLFCILVRCAGMRESGMPLHLFT

[1289] yCD酿酒酵母(Saccharomyces cerevisiae)：

[1290] MVTGGMASKWDQKGMDIAYEEAALGYKEGGVPIGGCLINNKDGSVLGRGHNMRFQKGSATLHGEISTLE NCGRLEGKVYKDTTLYTTLSPCDMCTGAIIMYGIPRCVVGENVNFKSKGEKYLQTRGHEVVVVDDERCKKIMKQFID ERPQDWFEDIGE

[1291] rAPOBEC-1(δ177-186)：

[1292] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEK FTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTE QESGYCWRNFVNYSPSNEAHWPRYPHLWVRGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK

[1293] rAPOBEC-1(δ202-213)：

[1294] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEK FTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTE QESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQHYQRLPPHILWATGLK

[1295] 小鼠APOBEC-3：

MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HGVFKNKDNI*HAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFEC*AEQIVRFLATHH
NLSLDIFSSRLYNVQDPETQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWK

[1296] RLLTNFRYQDSKLQEILRPCYIPVPSSSSSTLSNICLTKGLPETRFCVEGRRMDPLSEEE
FYSQFYNQRVKHLCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKGKQ*HAEILFLDKIRSM*
*ELSQVTITCYLTWSPCPNC*AWQLAAFKRDRPDLILHIYTSRLYFHWKRPFQKGLCSLWQS
GILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLRRIKESWGLQDLVNDFG
NLQLGPPMS (斜体:核酸编辑结构域)

[1297] 本公开的一些方面基于以下认知:例如通过在脱氨酶结构域中做出点突变来调整本文所述任何融合蛋白的脱氨酶结构域催化活性将会影响该融合蛋白(例如,碱基编辑器)的持续合成能力。例如,降低但不消除碱基编辑融合蛋白内的脱氨酶结构域的催化活性的突变可能使得脱氨酶结构域将更不可能催化与靶标残基相邻的残基的脱氨基,从而缩窄脱氨基窗。缩窄脱氨基窗的能力可防止不希望的与特异性靶标残基相邻的残基的脱氨基,这种能力可能降低或阻止脱靶效应。

[1298] 例如,在一些实施方案中,并入碱基编辑器中的APOBEC脱氨酶可包含选自由下列组成的组的一种或多种突变:rAPOBEC1的H121X、H122X、R126X、R126X、R118X、W90X、W90X和R132X或另一APOBEC脱氨酶中的一种或多种相对应的突变,其中X是任意氨基酸。在一些实施方案中,并入碱基编辑器中的APOBEC脱氨酶可包含选自由下列组成的组的一种或多种突变:rAPOBEC1的H121R、H122R、R126A、R126E、R118A、W90A、W90Y和R132E或另一APOBEC脱氨酶中的一种或多种相对应的突变。

[1299] 在一些实施方案中,并入碱基编辑器中的APOBEC脱氨酶可包含选自由下列组成的组的一种或多种突变:hAPOBEC3G的D316X、D317X、R320X、R320X、R313X、W285X、W285X、R326X或另一APOBEC脱氨酶中的一种或多种相对应的突变,其中X是任意氨基酸。在一些实施方案中,本文提供的任何融合蛋白包含APOBEC脱氨酶,该脱氨酶包含一个或多个选自由下列组成的组的一种或多种突变:hAPOBEC3G的D316R、D317R、R320A、R320E、R313A、W285A、W285Y、R326E或另一APOBEC脱氨酶的一种或多种相对应的突变。

[1300] 在一些实施方案中,并入碱基编辑器中的APOBEC脱氨酶包含rAPOBEC1的H121R和H122R突变或另一APOBEC脱氨酶的一种或多种相对应的突变。在一些实施方案中,并入碱基编辑器中的APOBEC脱氨酶包含具有下列的APOBEC脱氨酶:rAPOBEC1的R126A突变或另一APOBEC脱氨酶的一种或多种相对应的突变。在一些实施方案中,并入碱基编辑器中的APOBEC脱氨酶包含具有下列的APOBEC脱氨酶:rAPOBEC1的R126E突变或另一APOBEC脱氨酶的一种或多种相对应的突变。在一些实施方案中,并入碱基编辑器中的APOBEC脱氨酶包含具有下列的APOBEC脱氨酶:rAPOBEC1的R118A突变或另一APOBEC脱氨酶的一种或多种相对应的突变。在一些实施方案中,并入碱基编辑器中的APOBEC脱氨酶包含具有下列的APOBEC脱氨酶:rAPOBEC1的W90A突变或另一APOBEC脱氨酶的一种或多种相对应的突变。在一些实施方案中,并入碱基编辑器中的APOBEC脱氨酶包含具有下列的APOBEC脱氨酶:rAPOBEC1的W90Y突变或另一APOBEC脱氨酶的一种或多种相对应的突变。在一些实施方案中,并入碱基

编辑器中的APOBEC脱氨酶包含具有下列的APOBEC脱氨酶:rAPOBEC1的R132E突变或另一APOBEC脱氨酶的一种或多种相对应的突变。在一些实施方案中,并入碱基编辑器中的APOBEC脱氨酶包含具有下列的APOBEC脱氨酶:rAPOBEC1的W90Y和R126E突变或另一APOBEC脱氨酶的一种或多种相对应的突变。在一些实施方案中,并入碱基编辑器中的APOBEC脱氨酶包含具有下列的APOBEC脱氨酶:rAPOBEC1的R126E和R132E突变或另一APOBEC脱氨酶的一种或多种相对应的突变。在一些实施方案中,并入碱基编辑器中的APOBEC脱氨酶包含具有下列的APOBEC脱氨酶:rAPOBEC1的W90Y和R132E突变或另一APOBEC脱氨酶的一种或多种相对应的突变。在一些实施方案中,并入碱基编辑器中的APOBEC脱氨酶包含具有下列的APOBEC脱氨酶:rAPOBEC1的W90Y、R126E和R132E突变或另一APOBEC脱氨酶的一种或多种相对应的突变。

[1301] 在一些实施方案中,并入碱基编辑器中的APOBEC脱氨酶包含具有下列的APOBEC脱氨酶:hAPOBEC3G的D316R和D317R突变或另一APOBEC脱氨酶的一种或多种相对应的突变。在一些实施方案中,本文提供的任何融合蛋白包含APOBEC脱氨酶,该脱氨酶包含hAPOBEC3G的R320A突变或另一APOBEC脱氨酶的一种或多种相对应的突变。在一些实施方案中,并入碱基编辑器中的APOBEC脱氨酶包含具有下列的APOBEC脱氨酶:hAPOBEC3G的R320E突变或另一APOBEC脱氨酶的一种或多种相对应的突变。在一些实施方案中,并入碱基编辑器中的APOBEC脱氨酶包含具有下列的APOBEC脱氨酶:hAPOBEC3G的R313A突变或另一APOBEC脱氨酶的一种或多种相对应的突变。在一些实施方案中,并入碱基编辑器中的APOBEC脱氨酶包含具有下列的APOBEC脱氨酶:hAPOBEC3G的W285A突变或另一APOBEC脱氨酶的一种或多种相对应的突变。在一些实施方案中,并入碱基编辑器中的APOBEC脱氨酶包含具有下列的APOBEC脱氨酶:hAPOBEC3G的W285Y突变或另一APOBEC脱氨酶的一种或多种相对应的突变。在一些实施方案中,并入碱基编辑器中的APOBEC脱氨酶包含具有下列的APOBEC脱氨酶:hAPOBEC3G的R326E突变或另一APOBEC脱氨酶的一种或多种相对应的突变。在一些实施方案中,并入碱基编辑器中的APOBEC脱氨酶包含具有下列的APOBEC脱氨酶:hAPOBEC3G的W285Y和R320E突变或另一APOBEC脱氨酶的一种或多种相对应的突变。在一些实施方案中,并入碱基编辑器中的APOBEC脱氨酶包含具有下列的APOBEC脱氨酶:hAPOBEC3G的R320E和R326E突变或另一APOBEC脱氨酶的一种或多种相对应的突变。在一些实施方案中,并入碱基编辑器中的APOBEC脱氨酶包含具有下列的APOBEC脱氨酶:hAPOBEC3G的W285Y和R326E突变或另一APOBEC脱氨酶的一种或多种相对应的突变。在一些实施方案中,并入碱基编辑器中的APOBEC脱氨酶包含具有下列的APOBEC脱氨酶:hAPOBEC3G的W285Y、R320E和R326E突变或另一APOBEC脱氨酶的一种或多种相对应的突变。

[1302] 多种经修饰的胞苷脱氨酶是商业上可获得的,包括但不限于SaBE3、SaKKH-BE3、VQR-BE3、EQR-BE3、VRER-BE3、YE1-BE3、EE-BE3、YE2-BE3和YEE-BE3,它们可以从Addgene获得(质粒85169、85170、85171、85172、85173、85174、85175、85176、85177)。在一些实施方案中,被并入碱基编辑器的脱氨酶包含APOBEC1脱氨酶的全部或一部分。

[1303] C到T核碱基编辑蛋白的细节在国际PCT申请号PCT/US2016/058344(WO2017/070632)和Komor,A.C.,et al.,"Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage"Nature 533,420-424(2016)中有所描述,其整体内容通过引用并入本文。

[1304] 胞苷脱氨酶

[1305] 本文提供的融合蛋白包含一个或过个胞苷脱氨酶。在一些实施方案中,本文所提供的胞苷脱氨酶能够将胞嘧啶或5-甲基胞嘧啶脱氨基化为尿嘧啶或胸腺嘧啶。在一些实施方案中,本文所体用的胞苷脱氨酶能够将DNA中的胞嘧啶脱氨基。胞苷脱氨酶可以源自任何合适的生物体。在一些实施方案中,胞苷脱氨酶是天然出现的胞苷脱氨酶,其包括对应于本文所提供的任何突变的一个或多个突变。本领域技术人员将能够鉴定任何同源蛋白质中的相应残基,例如,通过序列比对和同源残基的确定。据此,本领域技术人员将能够产生任何天然出现的胞苷脱氨酶中的对应于本文所述任何突变的突变。在一些实施方案中,胞苷脱氨酶来自原核生物。在一些实施方案中,胞苷脱氨酶来自细菌。在一些实施方案中,胞苷脱氨酶来自哺乳动物(例如,人)。

[1306] 在一些实施方案中,胞苷脱氨酶包含氨基酸序列,所述氨基酸序列与本文详述的任何一个胞苷脱氨酶氨基酸序列为至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相同。应知悉,本文所提供的胞苷脱氨酶可以包括一个或多个突变(例如,本文所提供的任何突变)。本公开提供任何具有某一百分比的同一性加上任何本文所述突变或其组合的脱氨酶结构域。在一些实施方案中,胞苷脱氨酶包含氨基酸序列,与参考序列或本文所提供的任何胞苷脱氨酶相比,该氨基酸序列具有1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、21、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50或更多个突变。在一些实施方案中,胞苷脱氨酶包含氨基酸序列,与本领域中已知或本文所述的任何一种氨基酸序列相比,该氨基酸序列具有至少5、至少10、至少15、至少20、至少25、至少30、至少35、至少40、至少45、至少50、至少60、至少70、至少80、至少90、至少100、至少110、至少120、至少130、至少140、至少150、至少160或至少170个相同的连续氨基酸残基。

[1307] 本发明的融合蛋白包含两个或更多个核酸编辑结构域。在一些实施方案中,核酸编辑结构域可以催化C到U的碱基变化。在一些实施方案中,核酸编辑结构域是脱氨酶结构域。在一些实施方案中,脱氨酶是胞苷脱氨酶或腺苷脱氨酶。在一些实施方案中,脱氨酶是载脂蛋白BmRNA编辑复合物(APOBEC)家族脱氨酶。在一些实施方案中,脱氨酶是APOBEC1脱氨酶。在一些实施方案中,脱氨酶是APOBEC2脱氨酶。在一些实施方案中,脱氨酶是APOBEC3脱氨酶。在一些实施方案中,脱氨酶是APOBEC3A脱氨酶。在一些实施方案中,脱氨酶是APOBEC3B脱氨酶。在一些实施方案中,脱氨酶是APOBEC3C脱氨酶。在一些实施方案中,脱氨酶是APOBEC3D脱氨酶。在一些实施方案中,脱氨酶是APOBEC3E脱氨酶。在一些实施方案中,脱氨酶是APOBEC3F脱氨酶。在一些实施方案中,脱氨酶是APOBEC3G脱氨酶。在一些实施方案中,脱氨酶是APOBEC3H脱氨酶。在一些实施方案中,脱氨酶是APOBEC4脱氨酶。在一些实施方案中,脱氨酶是激活诱导的脱氨酶(AID)。在一些实施方案中,脱氨酶是脊椎动物脱氨酶。在一些实施方案中,脱氨酶是无脊椎动物脱氨酶。在一些实施方案中,脱氨酶是人、黑猩猩、大猩猩、猴、牛、狗、大鼠或小鼠脱氨酶。在一些实施方案中,脱氨酶是人脱氨酶。在一些实施方案中,脱氨酶是大鼠脱氨酶,例如,rAPOBEC1。在一些实施方案中,脱氨酶是海七鳃鳗胞苷脱氨酶1(pmCDA1)。在一些实施方案中,脱氨酶是人APOBEC3G。在一些实施方案中,脱氨酶是人APOBEC3G的片段。在一些实施方案中,脱氨酶是包含D3167R和D317R突变的人APOBEC3G变

体。在一些实施方案中,脱氨酶是人APOBEC3G的检定并且包含对应于D3167R和D317R突变的突变。在一些实施方案中,核酸编辑结构域与本文所述任何脱氨酶的脱氨酶结构域为至少80%、至少85%、至少90%、至少92%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相同。

[1308] 在某些实施方案中,本文所提供的融合蛋白包含一种或多种特征,该特征改善融合蛋白的碱基编辑活性。例如,本文所提供的任何融合蛋白可以包含具有削弱的核酸酶活性的Cas9结构域。在一些实施方案中,本文所提供的任何融合蛋白可以具有不具有核酸酶活性的Cas9结构域(dCas9)或者被称为Cas9切口酶(nCas9)的切割双螺旋DNA分子一条链的Cas9结构域。

[1309] 具有向导RNA的Cas9复合物

[1310] 本公开的一些方面提供复合物,其包含本文提供的任何融合蛋白和结合至融合蛋白的Cas9结构域(例如,dCas9、核酸酶活性Cas9或Cas9切口酶)的向导RNA。在一些实施方案中,向导核酸(例如,向导RNA)为15至100个核苷酸长,并且包含与靶标序列互补的至少10个连续核苷酸。在一些实施方案中,向导RNA为15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49或50个核苷酸长。在一些实施方案中,向导RNA包含与靶标序列互补的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或40个连续核苷酸的序列。在一些实施方案中,靶标序列为DNA序列。在一些实施方案中,靶标序列为细菌、酵母、真菌、昆虫、植物或动物基因组中的序列。在一些实施方案中,靶标序列为人类基因组中的序列。在一些实施方案中,靶标序列的3'末端紧邻经典PAM序列(NGG)。在一些实施方案中,靶标序列的3'紧邻非经典PAM序列(例如,表1中所列的序列或5'-NAA-3')。在一些实施方案中,向导核酸(例如,向导RNA)与感兴趣的基因(例如,HBV基因组中的基因)中的序列互补。

[1311] 本公开的一些方面提供使用本文所提供的融合蛋白或复合物的方法。例如,本公开的一些方面提供包括以下步骤的方法:使DNA分子与本文提供的任何具有至少一种向导RNA的融合蛋白接触,其中该向导RNA为约15至100个核苷酸长并且包含与靶标序列互补的至少10个连续核苷酸。在一些实施方案中,靶标序列的3'末端紧邻AGC、GAG、TTT、GTG或CAA序列。在一些实施方案中,靶标序列的3'末端紧邻NGA、NGCG、NGN、NNGRRT、NNNRRT、NGCG、NGCN、NGTN、NGTN、NGTN或5'(TTTV)序列。

[1312] 应理解,各序列中具体位置或残基的编号取决于所使用的特定蛋白质和编号方案。编号在例如成熟蛋白的前体和成熟蛋白本身中可以是不同的,并且来自不同物种的序列之间的差异可能影响编号。本领域技术人员将能够通过本领域周知的方法鉴定任何同源蛋白质和相应编码核酸中的各残基,例如,通过序列比对和同源残基的确定。

[1313] 对于本领域技术人员将会是显而易见的是,为了使本文公开的任何融合蛋白靶向靶标位点,例如,包含待编辑的突变的位点,将融合蛋白与向导RNA共表达通常是必要的。如本文其余处更详细解释的,向导RNA通常包含允许Cas9结合的tracrRNA框架以及向Cas9:核酸编辑酶/结构域融合蛋白提供序列特异性的向导序列。作为另一种选择,向导RNA和tracrRNA可以作为两个核酸分子而独立地提供。在一些实施方案中,向导RNA包含一结构,其中该向导序列包含与靶标序列互补的序列。向导序列通常为20个核苷酸长。基于本公开,适用于使Cas9:核酸编辑酶/结构域融合蛋白靶向特定基因组靶标位点的向导RNA的序列对

于本领域技术人员将会是显而易见的。此类合适的向导RNA序列通常包含与位于待编辑的靶标核苷酸上游或下游50个核苷酸内的核酸序列互补的向导序列。本文提供了一些适用于使所提供的任何融合蛋白靶向特定靶标序列的示例性向导RNA序列。

[1314] 额外结构域

[1315] 本文所述的碱基编辑器可以包括帮助促进多核苷酸的核碱基的核碱基编辑、修饰或改变的任何结构域。在一些实施方案中,碱基编辑器包含多核苷酸可编程核苷酸结合结构域(例如,Cas9)、核碱基编辑结构域(例如,脱氨酶结构域)和一个或多个额外结构域。在一些情况下,额外结构域可以促进碱基编辑器的酶促或催化功能、碱基编辑器的结合功能,或者可以是将会干扰所希望的碱基编辑结果的细胞机构(例如,酶)的抑制剂。在一些实施方案中,碱基编辑器可以包含核酸酶、切口酶、重组酶、脱氨酶、甲基转移酶、甲基化酶、乙酰化酶、乙酰转移酶、转录激活因子或转录抑制因子结构域。

[1316] 在一些实施方案中,碱基编辑器可以包含尿嘧啶糖苷酶抑制剂(UGI)结构域。UGI结构域可以,例如,通过抑制通过C的脱氨反应形成的U转化回C核碱基,改善包含包含胞苷脱氨酶结构域的碱基编辑器的效率。在一些情况下,应答U:G异源双螺旋DNA的存在的细胞DNA修复可能是细胞内核碱基编辑效率下降的原因。在这种情况下,尿嘧啶DNA糖苷酶(UDG)可以催化U从细胞内的DNA中移除,这可能启动碱基切除修复(BER),主要导致U:G对逆转为C:G对。在这种情况下,BER可以在包含一个或多个结构域的碱基编辑器中被抑制,该结构域结合单链,阻断被编辑的碱基,抑制UGI,抑制BER,保护被编辑的碱基和/或促使未经编辑的链的修复。因此,本公开设想了包含UGI结构域的碱基编辑器融合蛋白。

[1317] 在一些实施方案中,碱基编辑器包含双链断裂(DSB)结合蛋白的全部或一部分作为结构域。例如,DSB结合蛋白可以包括细菌噬菌体Mu的Gam蛋白,其结合至DSB的末端并且可以包含DSB免于降解。参见,Komor,A.C.,et al.,“Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity”Science Advances 3:eaao4774(2017),其整体内容通过引用并入本文。

[1318] 此外,在一些实施方案中,Gam蛋白可以融合至碱基编辑器的N端。在一些实施方案中,Gam蛋白可以融合至碱基编辑器的C端。细菌噬菌体Mu的Gam蛋白可以结合至双链断裂(DSB)的末端并且保护它们免于降解。在一些实施方案中,使用Gam来结合DSB的自由末端可以减少碱基编辑过程中的插入缺失形成。在一些实施方案中,将174个残基的Gam蛋白融合至碱基编辑器的N端。参见,Komor,A.C.,et al.,“Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity”Science Advances 3:eaao4774(2017)。在一些实施方案中,一个或多个突变可以改变碱基编辑器结构域相对于野生型结构域的长度。例如,在至少一个结构域中至少一个氨基酸的缺失可以缩短碱基编辑器的长度。在另一种情况下,一个或多个突变不改变结构域相对于野生型结构域的长度。例如,任何结构域中的一种或多种置换不改变碱基编辑器的长度。

[1319] 在一些实施方案中,碱基编辑器可以包含核酸聚合酶(NAP)的全部或一部分作为结构域。例如,碱基编辑器可以包含真核生物NAP的全部或一部分。在一些实施方案中,被并入碱基编辑器中的NAP或其部分是DNA聚合酶。在一些实施方案中,被并入碱基编辑器中的

NAP或其部分具有跨损伤聚合酶活性。在一些情况下,被并入碱基编辑器中的NAP或其部分是跨损伤DNA聚合酶。在一些实施方案中,被并入碱基编辑器中的NAP或其部分是Rev7、Rev1复合物、聚合酶ι、聚合酶κ或聚合酶η。在一些实施方案中,被并入碱基编辑器中的NAP或其部分是真核生物聚合酶α、β、γ、δ、ε、ζ、γ、η、ι、κ、λ、μ或ν组分。在一些实施方案中,被并入碱基编辑器中的NAP或其部分包含氨基酸序列,该氨基酸序列与核酸聚合酶(例如,跨损伤DNA聚合酶)为至少75%、80%、85%、90%、95%、96%、97%、98%、99%或至少99.5%相同。

[1320] 碱基编辑器系统

[1321] 本文所提供的碱基编辑器用包括以下步骤:(a)将受试者的多核苷酸(例如,双链DNA或RNA、单链DNA或RNA)的靶标核苷酸序列与包含腺苷脱氨酶结构域或胞苷脱氨酶结构域的碱基编辑器系统和至少一种向导多核酸(例如,gRNA)接触,其中前述结构域融合至多核苷酸结合结构域,从而形成能够在如本文所述的核酸分子内的一个或多个碱基处诱导改变的核碱基编辑器,其中该靶标核苷酸序列包含所靶向的碱基对;(b)诱导靶标区域的链分离;(c)将靶标区域的单链中的靶标核碱基对的第一核碱基转化为第二核碱基;以及(d)切割靶标区域的不超过一条链,其中与第一核碱基互补的第三核碱基被替换为与第二核碱基互补的第四核碱基。应知悉,在一些实施方案中,省略了步骤(b)。在一些实施方案中,被靶向的核碱基对是一个或多个基因中的多个核碱基对。在一些实施方案中,本文所提供的碱基编辑器系统能够对一个或多个基因中的多个核碱基对进行多元编辑。在一些实施方案中,多个核碱基对定位在同一基因中。在一些实施方案中,多个核碱基对定位在一个或多个基因中,其中至少一个基因定位在不同的基因座中。

[1322] 在一些实施方案中,被切割的单链(被切口的链)与向导核酸杂交。在一些实施方案中,被切割的单链与包含第一核碱基的链相对。在一些实施方案中,碱基编辑器包含Cas9结构域。在一些实施方案中,第一碱基是腺嘌呤,并且第二碱基不是G、C、A或T。在一些实施方案中,第二碱基是肌苷。

[1323] 本文所提供的碱基编辑系统提供新的基因组编辑途径,该系统使用含有催化缺陷性酿脓链球菌Cas9、胞苷脱氨酶和碱基切除修复抑制剂的融合蛋白来诱导DNA中的可编程单核苷酸(C→T或A→G)变化而不产生双链DNA断裂、不需要供体DNA模板并且不诱导过量的随机插入和缺失。

[1324] 本文提供用于使用碱基编辑器系统来编辑核碱基的系统、组合物和方法。在一些实施方案中,碱基编辑器系统包含:包含多核苷酸可编程核苷酸结合结构域和用于编辑核碱基的核碱基编辑结构域(例如,脱氨酶结构域)的碱基编辑器(BE);和与该多核苷酸可编程核苷酸结合结构域协同作用的向导多核苷酸(例如,向导RNA)。在一些实施方案中,碱基编辑器系统包含:包含多核苷酸可编程核苷酸结合结构域和用于编辑核碱基的核碱基编辑结构域(例如,脱氨酶结构域)的碱基编辑器(BE);和与该多核苷酸可编程核苷酸结合结构域协同作用的向导多核苷酸(例如,向导RNA)。在一些实施方案中,多核苷酸可编程核苷酸结合结构域是多核苷酸可编程DNA结合结构域。在一些实施方案中,多核苷酸可编程核苷酸结合结构域是多核苷酸可编程RNA结合结构域。在一些实施方案中,脱氨酶结构域可以是胞嘧啶脱氨酶或胞苷脱氨酶、腺嘌呤脱氨酶或腺苷脱氨酶。在一些实施方案中,“胞嘧啶脱氨酶”和“胞苷脱氨酶”可以互换地使用。在一些实施方案中,“腺嘌呤脱氨酶”和“腺苷脱氨酶”可以互换地使用。在一些情况下,脱氨酶结构域可以是胞嘧啶脱氨酶或胞苷脱氨酶。在一些

情况下,脱氨酶结构域可以是腺嘌呤脱氨酶或腺苷脱氨酶。

[1325] 核碱基编辑蛋白的细节在国际PCT申请号PCT/2017/045381(WO2018/027078)和PCT/US2016/058344(WO2017/070632)中描述,其各自通过引用而整体并入本文。也参见,Komor,A.C.,et al.,"Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage"Nature 533,420-424(2016);Gaudelli,N.M.,et al.,"Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage"Nature 551,464-471(2017)和Komor,A.C.,et al.,"Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity"Science Advances 3:eaao4774(2017)),其整体内容通过引用并入本文。

[1326] 在一些实施方案中,单向导多核苷酸可用将脱氨酶靶向靶标核酸序列。在一些实施方案中,一对向导多核苷酸可用将不同的脱氨酶靶向靶标核酸序列。碱基编辑器系统的核碱基组分和多核苷酸可编程核苷酸结合组分可以彼此共价或非共价地缔合。例如,在一些实施方案中,脱氨酶结构域可以通过多核苷酸可编程核苷酸结合结构域被靶向至靶标核苷酸序列。在一些实施方案中,多核苷酸可编程核苷酸结合结构域可以融合或连接至脱氨酶结构域。在一些实施方案中,多核苷酸可编程核苷酸结合结构域可以通过与脱氨酶结构域非共价地相互作用或缔合而将脱氨酶结构域靶向至核苷酸序列。例如,在一些实施方案中,核碱基编辑组分(例如,脱氨酶组分)可以包含额外的异源部分或结构域,该异源部分或结构域能够与作为多核苷酸可编程核苷酸结合结构域的一部分的额外异源部分或结构域相互作用、缔合或形成复合物。在一些实施方案中,额外异源部分可以能够与多肽结合、相互作用、缔合或形成复合物。在一些实施方案中,额外异源部分可以能够与多核苷酸结合、相互作用、缔合或形成复合物。在一些实施方案中,额外异源部分可以能够结合至向导多核苷酸。在一些实施方案中,额外异源部分可以能够结合至多肽连接子。在一些实施方案中,额外异源部分可以能够结合至多核苷酸连接子。额外异源部分可以是蛋白结构域。在一些实施方案中,额外异源部分可以是K同源(KH)结构域、MS2外壳蛋白结构域、PP7外壳蛋白结构域、SfMu外壳蛋白结构域、不育α基序、端粒酶Ku结合基序和Ku蛋白、端粒酶Sm7结合基序和Sm7蛋白、或RNA识别基序。

[1327] 碱基编辑器系统可以进一步包含向导多核苷酸组分。应知悉,碱基编辑器系统的组分可以彼此经由共价键彼此缔合、非共价相互作用、或其缔合及相互作用的任何组合。在一些实施方案中,脱氨酶结构域可以通过向导多核苷酸被靶向至靶标核苷酸序列。例如,在一些实施方案中,碱基编辑器系统的核碱基编辑组分(例如,脱氨酶组分)可以包含额外异源部分或结构域(例如,多核苷酸结合结构域诸如RNA或DNA结合蛋白),该额外异源部分或结构域能够与向导多核苷酸的部分或链段(例如,多核苷酸基序)相互作用、缔合或形成复合物。在一些实施方案中,额外异源部分或结构域(例如,多核苷酸结合结构域诸如RNA或DNA结合蛋白)可以融合或连接至脱氨酶结构域。在一些实施方案中,额外异源部分可以能够与多肽结合、相互作用、缔合或形成复合物。在一些实施方案中,额外异源部分可以能够与多核苷酸结合、相互作用、缔合或形成复合物。在一些实施方案中,额外异源部分可以能够结合至向导多核苷酸。在一些实施方案中,额外异源部分可以能够结合至多肽连接子。在一些实施方案中,额外异源部分可以能够结合至多核苷酸连接子。额外异源部分可以是蛋

白结构域。在一些实施方案中,额外异源部分可以是K同源(KH)结构域、MS2外壳蛋白结构域、PP7外壳蛋白结构域、SfMu外壳蛋白结构域、不育α基序、端粒酶Ku结合基序和Ku蛋白、端粒酶Sm7结合基序和Sm7蛋白、或RNA识别基序。

[1328] 在一些实施方案中,碱基编辑器系统可以进一步包含碱基切除修复(BER)组分的抑制剂。应知悉,碱基编辑器系统的组分可以彼此经由共价键彼此缔合、非共价相互作用、或其缔合及相互作用的任何组合。BER组分的抑制剂可以包含碱基切除修复抑制剂。在一些实施方案中,碱基切除修复的抑制剂可以是尿嘧啶DNA糖苷酶抑制剂(UGI)。在一些实施方案中,碱基切除修复的抑制剂可以是肌苷碱基切除修复抑制剂。在一些实施方案中,碱基切除修复的抑制剂可以通过多核苷酸可编程核苷酸结合结构域被靶向至靶标核苷酸序列。在一些实施方案中,多核苷酸可编程核苷酸结合结构域可以融合或连接至碱基切除修复的抑制剂。在一些实施方案中,多核苷酸可编程核苷酸结合结构域可以融合或连接至脱氨酶结构域和碱基切除修复的抑制剂。在一些实施方案中,多核苷酸可编程核苷酸结合结构域可以通过与碱基切除修复的抑制剂非共价地相互作用或缔合而将碱基切除修复的抑制剂靶向至核苷酸序列。例如,在一些实施方案中,碱基切除修复组分的抑制剂可以包含额外的异源部分或结构域,该异源部分或结构域能够与作为多核苷酸可编程核苷酸结合结构域的一部分的额外异源部分或结构域相互作用、缔合或形成复合物。在一些实施方案中,碱基切除修复的抑制剂可以通过向导多核苷酸被靶向至靶标核苷酸序列。例如,在一些实施方案中,碱基切除修复的抑制剂可以包含额外异源部分或结构域(例如,多核苷酸结合结构域诸如RNA或DNA结合蛋白),该额外异源部分或结构域能够与向导多核苷酸的部分或链段(例如,多核苷酸基序)相互作用、缔合或形成复合物。在一些实施方案中,向导多核苷酸的额外异源部分或结构域(例如,多核苷酸结合结构域诸如RNA或DNA结合蛋白)可以融合或连接至碱基切除修复的抑制剂。在一些实施方案中,额外异源部分可以能够与多核苷酸结合、相互作用、缔合或形成复合物。在一些实施方案中,额外异源部分可以能够结合至向导多核苷酸。在一些实施方案中,额外异源部分可以能够结合至多肽连接子。在一些实施方案中,额外异源部分可以能够结合至多核苷酸连接子。额外异源部分可以是蛋白结构域。在一些实施方案中,额外异源部分可以是K同源(KH)结构域、MS2外壳蛋白结构域、PP7外壳蛋白结构域、SfMu外壳蛋白结构域、不育α基序、端粒酶Ku结合基序和Ku蛋白、端粒酶Sm7结合基序和Sm7蛋白、或RNA识别基序。

[1329] 在一些实施方案中,碱基编辑器抑制被编辑链的碱基切除修复。在一些实施方案中,碱基编辑器保护或结合未经编辑的链。在一些实施方案中,碱基编辑器包含UGI活性。在一些实施方案中,碱基编辑器包含无催化活性的肌苷特异性核酸酶。在一些实施方案中,碱基编辑器包含切口酶活性。在一些实施方案中,预期的碱基对编辑位于PAM位点的上游。在一些实施方案中,预期的碱基对编辑位于PAM位点的上游1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个核苷酸处。在一些实施方案中,预期的碱基对编辑位于PAM位点的下游。在一些实施方案中,预期的被编辑碱基对位于PAM位点的下游1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个核苷酸处。

[1330] 在一些实施方案中,该方法不需要经典(例如,NGG)PAM位点。在一些实施方案中,核碱基编辑器包含连接子或间隔序列。在一些实施方案中,连接子或间隔序列为1至25个氨基酸的长度。在一些实施方案中,连接子或间隔序列为5至20个氨基酸的长度。在一些实施

方案中,连接子或间隔序列为10、11、12、13、14、15、16、17、18、19或20个氨基酸的长度。

[1331] 在一些实施方案中,靶标区域包含靶标窗,其中靶标窗包含靶标核碱基对。在一些实施方案中,靶标窗包含1至10个核苷酸。在一些实施方案中,靶标窗为1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个核苷酸的长度。在一些实施方案中,预期的碱基对编辑位于靶标窗内。在一些实施方案中,靶标窗包含预期的碱基对编辑。在一些实施方案中,使用本文所提供的任何碱基编辑器执行该方法。在一些实施方案中,靶标窗是脱氨基窗。

[1332] 在一些实施方案中,非限制性的示例性胞苷碱基编辑器(CBE)包括BE1(APOBEC1-XTEN-dCas9)、BE2(APOBEC1-XTEN-dCas9-UGI)、BE3(APOBEC1-XTEN-dCas9(A840H)-UGI)、BE3-Gam、saBE3、saBE4-Gam、BE4、BE4-Gam、saBE4或saB4E-Gam。BE4将APOBEC1-Cas9n(D10A)连接子延长至32个氨基酸并将Cas9n-UGI连接子延长至9个氨基酸,并且将UGI的第二个拷贝附加到构建体的C端并将另一个9个氨基酸的连接子附加到单碱基编辑器构建体内。碱基编辑器saBE3和saBE4的酿脓链球菌Cas9n(D10A)被较小的金黄色葡萄球菌Cas9n(D10A)替代。BE3-Gam、saBE3-Gam、BE4-Gam和saBE4-Gam具有经由16个氨基酸的XTEN连接子与BE3、saBE3、BE4和saBE4的N端融合的174个残基的Gam蛋白。

[1333] 在一些实施方案中,腺苷碱基编辑器(ABE)可以将DNA中的腺嘌呤脱氨基。在一些实施方案中,ABE是通过将BE3的APOBEC1组分替换为天然或经工程化的大肠杆菌TadA、人ADAR2、小鼠ADA或人ADAT2而产生的。在一些实施方案中,ABE包含进化的TadA变体。在一些实施方案中,ABE是ABE 1.2(TadA*-XTEN-nCas9-NLS)。在一些实施方案中,TadA*包含A106V和D108N突变。

[1334] 在一些实施方案中,ABE是第二代ABE。在一些实施方案中,ABE是ABE2.1,其包含TadA*(TadA*2.1)中的额外突变D147Y和E155V。在一些实施方案中,ABE是ABE2.2,其是融合至无催化活性版本的人烷基腺嘌呤DNA糖苷酶(具有E125Q突变的AAG)的ABE2.1。在一些实施方案中,ABE是ABE2.3,其是融合至无催化活性版本的大肠杆菌Endo V(具有D35A突变而无活性)的ABE2.1。在一些实施方案中,ABE是ABE2.6,其具有两倍于ABE2.1中的连接子的连接子(32个氨基酸,$(SGGS)_2$-XTEN-$(SGGS)_2$)。在一些实施方案中,ABE是ABE2.7,其是以额外的野生型TadA单体系结的ABE2.1。在一些实施方案中,ABE是ABE2.8,其是以额外的TadA*2.1单体系结的ABE2.1。在一些实施方案中,ABE是ABE2.9,其是进化的TadA(TadA*2.1)直接融合至ABE2.1的N端的融合物。在一些实施方案中,ABE是ABE2.10,其是野生型TadA直接融合至ABE2.1的N端的融合物。在一些实施方案中,ABE是ABE2.11,其是在TadA*单体的N端处具有无活性E59A突变的ABE2.9。在一些实施方案中,ABE是ABE2.12,其是在内部TadA*单体处具有无活性E59A突变的ABE2.9。

[1335] 在一些实施方案中,ABE是第三代ABE。在一些实施方案中,ABE是ABE3.1,其是具有三个额外TadA突变(L84F、H123Y和I157F)的ABE2.3。

[1336] 在一些实施方案中,ABE是第四代ABE。在一些实施方案中,ABE是ABE4.3,其是具有额外TadA突变A142N(TadA*4.3)的ABE3.1。

[1337] 在一些实施方案中,ABE是第五代ABE。在一些实施方案中,ABE是ABE5.1,其通过将来自存活克隆体的一组共有突变(H36L、R51L、S146C和K157N)导入ABE3.1中而产生的。在一些实施方案中,ABE是ABE5.3,其具有含有野生型大肠杆菌TadA与内部进化的TadA*融合的

异二聚构建体。在一些实施方案中,ABE是ABE5.2、ABE5.4、ABE5.5、ABE5.6、ABE5.7、ABE5.8、ABE5.9、ABE5.10、ABE5.11、ABE5.12、ABE5.13或ABE5.14,如下表7中所示。在一些实施方案中,ABE是第六代ABE。在一些实施方案中,ABE是ABE6.1、ABE6.2、ABE6.3、ABE6.4、ABE6.5或ABE6.6,如下表7中所示。在一些实施方案中,ABE是第七代ABE。在一些实施方案中,ABE是ABE7.1、ABE7.2、ABE7.3、ABE7.4、ABE7.5、ABE7.6、ABE7.7、ABE7.8、ABE 7.9或ABE7.10,如下表7中所示。

[1338] 表7.ABE的基因型

[1339]

	23	26	36	37	48	49	51	72	84	87	106	108	123	125	142	146	147	152	155	156	157	161
ABE0.1	W	R	H	N	P		R	N	L	S	A	D	H	G	A	S	D	R	E	I	K	K
ABE0.2	W	R	H	N	P		R	N	L	S	A	D	H	G	A	S	D	R	E	I	K	K
ABE1.1	W	R	H	N	P		R	N	L	S	A	N	H	G	A	S	D	R	E	I	K	K
ABE1.2	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	D	R	E	I	K	K
ABE2.1	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.2	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.3	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.4	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.5	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.6	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.7	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.8	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.9	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.10	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.11	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.12	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE3.1	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.2	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.3	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.4	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.5	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.6	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.7	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.8	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE4.1	W	R	H	N	P		R	N	L	S	V	N	H	G	N	S	Y	R	V	I	K	K
ABE4.2	W	G	H	N	P		R	N	L	S	V	N	H	G	N	S	Y	R	V	I	K	K
ABE4.3	W	R	H	N	P		R	N	F	S	V	N	Y	G	N	S	Y	R	V	F	K	K
ABE5.1	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.2	W	R	H	S	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	T
ABE5.3	W	R	L	N	P		L	N	I	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.4	W	R	H	S	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	T
ABE5.5	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.6	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.7	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.8	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K

[1340]

	23	26	36	37	48	49	51	72	84	87	106	108	123	125	142	146	147	152	155	156	157	161
ABE5.9	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.10	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.11	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.12	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.13	W	R	H	N	P		L	D	F	S	V	N	Y	A	A	S	Y	R	V	F	K	K
ABE5.14	W	R	H	N	S		L	N	F	C	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE6.1	W	R	H	N	S		L	N	F	S	V	N	Y	G	N	S	Y	R	V	F	K	K
ABE6.2	W	R	H	N	T	V	L	N	F	S	V	N	Y	G	N	S	Y	R	V	F	N	K
ABE6.3	W	R	L	N	S		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE6.4	W	R	L	N	S		L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
ABE6.5	W	R	L	N	T	V	L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE6.6	W	R	L	N	T	V	L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
ABE7.1	W	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE7.2	W	R	L	N	A		L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
ABE7.3	L	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE7.4	R	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE7.5	W	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	H	V	F	N	K
ABE7.6	W	R	L	N	A		L	N	I	S	V	N	Y	G	A	C	Y	P	V	F	N	K
ABE7.7	L	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	P	V	F	N	K
ABE7.8	L	R	L	N	A		L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
ABE7.9	L	R	L	N	A		L	N	F	S	V	N	Y	G	N	C	Y	P	V	F	N	K
ABE7.10	R	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	P	V	F	N	K

[1341] 在一些实施方案中,碱基编辑器是第八代ABE(ABE8)。在一些实施方案中,ABE8含有TadA*8变体。在一些实施方案中,ABE8具有含有TadA*8变体的单体构建体(“ABE8.x-m”)。在一些实施方案中,ABE8是ABE8.1-m,其具有含有具有Y147T突变的TadA*7.10的单体构建体(TadA*8.1)。在一些实施方案中,ABE8是ABE8.2-m,其具有含有具有Y147R突变的TadA*7.10(TadA*8.2)的单体构建体。在一些实施方案中,ABE8是ABE8.3-m,其具有含有具有Q154S突变的TadA*7.10(TadA*8.3)的单体构建体。在一些实施方案中,ABE8是ABE8.4-m,其具有含有具有Y123H突变的TadA*7.10(TadA*8.4)的单体构建体。在一些实施方案中,ABE8是ABE8.5-m,其具有含有具有V82S突变的TadA*7.10(TadA*8.5)的单体构建体。在一些实施方案中,ABE8是ABE8.6-m,其具有含有具有T166R突变的TadA*7.10(TadA*8.6)的单体构建体。在一些实施方案中,ABE8是ABE8.7-m,其具有含有具有Q154R突变的TadA*7.10(TadA*8.7)的单体构建体。在一些实施方案中,ABE8是ABE8.8-m,其具有含有具有Y147R、Q154R和Y123H突变的TadA*7.10(TadA*8.8)的单体构建体。在一些实施方案中,ABE8是ABE8.9-m,其具有含有具有Y147R、Q154R和I76Y突变的TadA*7.10(TadA*8.9)的单体构建体。在一些实施方案中,ABE8是ABE8.10-m,其具有含有具有Y147R、Q154R和T166R突变的TadA*7.10(TadA*8.10)的单体构建体。在一些实施方案中,ABE8是ABE8.11-m,其具有含有具有Y147T和Q154R突变的TadA*7.10(TadA*8.11)的单体构建体。在一些实施方案中,ABE8是ABE8.12-m,其具有含有具有Y147T和Q154S突变的TadA*7.10(TadA*8.12)的单体构建体。在一些实施方案

中,ABE8是ABE8.13-m,其具有含有具有Y123H(Y123H从H123Y逆转)、Y147R、Q154R和I76Y突变的TadA*7.10(TadA*8.13)的单体构建体。在一些实施方案中,ABE8是ABE8.14-m,其具有含有具有I76Y和V82S突变的TadA*7.10(TadA*8.14)的单体构建体。在一些实施方案中,ABE8是ABE8.15-m,其具有含有具有V82S和Y147R突变的TadA*7.10(TadA*8.15)的单体构建体。在一些实施方案中,ABE8是ABE8.16-m,其具有含有具有V82S、Y123H(Y123H从H123Y逆转)和Y147R突变的TadA*7.10(TadA*8.16)的单体构建体。在一些实施方案中,ABE8是ABE8.17-m,其具有含有具有V82S和Q154R突变的TadA*7.10(TadA*8.17)的单体构建体。在一些实施方案中,ABE8是ABE8.18-m,其具有含有具有V82S、Y123H(Y123H从H123Y逆转)和Q154R突变的TadA*7.10(TadA*8.18)的单体构建体。在一些实施方案中,ABE8是ABE8.19-m,其具有含有具有V82S、Y123H(Y123H从H123Y逆转)、Y147R和Q154R突变的TadA*7.10(TadA*8.19)的单体构建体。在一些实施方案中,ABE8是ABE8.20-m,其具有含有具有I76Y、V82S、Y123H(Y123H从H123Y逆转)、Y147R和Q154R突变的TadA*7.10(TadA*8.20)的单体构建体。在一些实施方案中,ABE8是ABE8.21-m,其具有含有具有Y147R和Q154S突变的TadA*7.10(TadA*8.21)的单体构建体。在一些实施方案中,ABE8是ABE8.22-m,其具有含有具有V82S和Q154S突变的TadA*7.10(TadA*8.22)的单体构建体。在一些实施方案中,ABE8是ABE8.23-m,其具有含有具有V82S和Y123H(Y123H从H123Y逆转)突变的TadA*7.10(TadA*8.23)的单体构建体。在一些实施方案中,ABE8是ABE8.24-m,其具有含有具有V82S、Y123H(Y123H从H123Y逆转)和Y147T突变的TadA*7.10(TadA*8.24)的单体构建体。

[1342] 在一些实施方案中,ABE8具有含有野生型大肠杆菌与TadA*8变体融合的异二聚构建体(“ABE8.x-d”)。在一些实施方案中,ABE8是ABE8.1-d,其具有含有野生型大肠杆菌TadA与具有Y147T突变的TadA*7.10(TadA*8.1)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.2-d,其具有含有野生型大肠杆菌TadA与具有Y147R突变的TadA*7.10(TadA*8.2)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.3-d,其具有含有野生型大肠杆菌TadA与具有Q154S突变的TadA*7.10(TadA*8.3)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.4-d,其具有含有野生型大肠杆菌TadA与具有Y123H突变的TadA*7.10(TadA*8.4)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.5-d,其具有含有野生型大肠杆菌TadA与具有V82S突变的TadA*7.10(TadA*8.5)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.6-d,其具有含有野生型大肠杆菌TadA与具有T166R突变的TadA*7.10(TadA*8.6)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.7-d,其具有含有野生型大肠杆菌TadA与具有Q154R突变的TadA*7.10(TadA*8.7)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.8-d,其具有含有野生型大肠杆菌TadA与具有Y147R、Q154R和Y123H突变的TadA*7.10(TadA*8.8)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.9-d,其具有含有野生型大肠杆菌TadA与具有Y147R、Q154R和I76Y突变的TadA*7.10(TadA*8.9)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.10-d,其具有含有野生型大肠杆菌TadA与具有Y147R、Q154R和T166R突变的TadA*7.10(TadA*8.10)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.11-d,其具有含有野生型大肠杆菌TadA与具有Y147T和Q154R突变的TadA*7.10(TadA*8.11)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.12-d,其具有含有野生型大肠杆菌TadA与具有Y147T和Q154S突变的TadA*7.10(TadA*8.12)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.13-d,其具有含有野生型大肠杆菌TadA与具有Y123H

(Y123H从H123Y逆转)、Y147R、Q154R和I76Y突变的TadA*7.10(TadA*8.13)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.14-d,其具有含有野生型大肠杆菌TadA与具有I76Y和V82S突变的TadA*7.10(TadA*8.14)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.15-d,其具有含有野生型大肠杆菌TadA与具有V82S和Y147R突变的TadA*7.10(TadA*8.15)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.16-d,其具有含有野生型大肠杆菌TadA与具有V82S、Y123H(Y123H从H123Y逆转)和Y147R突变的TadA*7.10(TadA*8.16)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.17-d,其具有含有野生型大肠杆菌TadA与具有V82S和Q154R突变的TadA*7.10(TadA*8.17)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.18-d,其具有含有野生型大肠杆菌TadA与具有V82S、Y123H(Y123H从H123Y逆转)和Q154R突变的TadA*7.10(TadA*8.18)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.19-d,其具有含有野生型大肠杆菌TadA与具有V82S、Y123H(Y123H从H123Y逆转)、Y147R和Q154R突变的TadA*7.10(TadA*8.19)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.20-d,其具有含有野生型大肠杆菌TadA与具有I76Y、V82S、Y123H(Y123H从H123Y逆转)、Y147R和Q154R突变的TadA*7.10(TadA*8.20)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.21-d,其具有含有野生型大肠杆菌TadA与具有Y147R和Q154S突变的TadA*7.10(TadA*8.21)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.22-d,其具有含有野生型大肠杆菌TadA与具有V82S和Q154S突变的TadA*7.10(TadA*8.22)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.23-d,其具有含有野生型大肠杆菌TadA与具有V82S和Y123H(Y123H从H123Y逆转)突变的TadA*7.10(TadA*8.23)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.24-d,其具有含有野生型大肠杆菌TadA与具有V82S、Y123H(Y123H从H123Y逆转)和Y147T突变的TadA*7.10(TadA*8.24)融合的异二聚构建体。

[1343] 在一些实施方案中,ABE8具有含有TadA*7.10与TadA*8变体融合的异二聚构建体("ABE8.x-7")。在一些实施方案中,ABE8是ABE8.1-7,其具有含有TadA*7.10与具有Y147T突变的TadA*7.10(TadA*8.1)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.2-7,其具有含有TadA*7.10与具有Y147R突变的TadA*7.10(TadA*8.2)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.3-7,其具有含有TadA*7.10与具有Q154S突变的TadA*7.10(TadA*8.3)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.4-7,其具有含有TadA*7.10与具有Y123H突变的TadA*7.10(TadA*8.4)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.5-7,其具有含有TadA*7.10与具有V82S突变的TadA*7.10(TadA*8.5)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.6-7,其具有含有TadA*7.10与具有T166R突变的TadA*7.10(TadA*8.6)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.7-7,其具有含有TadA*7.10与具有Q154R突变的TadA*7.10(TadA*8.7)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.8-7,其具有含有TadA*7.10与具有Y147R、Q154R和Y123H突变的TadA*7.10(TadA*8.8)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.9-7,其具有含有TadA*7.10与具有Y147R、Q154R和I76Y突变的TadA*7.10(TadA*8.9)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.10-7,其具有含有TadA*7.10与具有Y147R、Q154R和T166R突变的TadA*7.10(TadA*8.10)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.11-7,其具有含有TadA*7.10与具有Y147T和Q154R突变的TadA*7.10(TadA*8.11)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.12-7,其具有含有TadA*7.10与具有

Y147T和Q154S突变的TadA*7.10(TadA*8.12)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.13-7,其具有含有TadA*7.10与具有Y123H(Y123H从H123Y逆转)、Y147R、Q154R和I76Y突变的TadA*7.10(TadA*8.13)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.14-7,其具有含有TadA*7.10与具有I76Y和V82S突变的TadA*7.10(TadA*8.14)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.15-7,其具有含有TadA*7.10与具有V82S和Y147R突变的TadA*7.10(TadA*8.15)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.16-7,其具有含有TadA*7.10与具有V82S、Y123H(Y123H从H123Y逆转)和Y147R突变的TadA*7.10(TadA*8.16)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.17-7,其具有含有TadA*7.10与具有V82S和Q154R突变的TadA*7.10(TadA*8.17)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.18-7,其具有含有TadA*7.10与具有V82S、Y123H(Y123H从H123Y逆转)和Q154R突变的TadA*7.10(TadA*8.18)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.19-7,其具有含有TadA*7.10与具有V82S、Y123H(Y123H从H123Y逆转)、Y147R和Q154R突变的TadA*7.10(TadA*8.19)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.20-7,其具有含有TadA*7.10与具有I76Y、V82S、Y123H(Y123H从H123Y逆转)、Y147R和Q154R突变的TadA*7.10(TadA*8.20)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.21-7,其具有含有TadA*7.10与具有Y147R和Q154S突变的TadA*7.10(TadA*8.21)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.22-7,其具有含有TadA*7.10与具有V82S和Q154S突变的TadA*7.10(TadA*8.22)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.23-7,其具有含有TadA*7.10与具有V82S和Y123H(Y123H从H123Y逆转)突变的TadA*7.10(TadA*8.23)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.24-7,其具有含有TadA*7.10与具有V82S、Y123H(Y123H从H123Y逆转)和Y147T突变的TadA*7.10(TadA*8.24)融合的异二聚构建体。

[1344] 在一些实施方案中,ABE是ABE8.1-m、ABE8.2-m、ABE8.3-m、ABE8.4-m、ABE8.5-m、ABE8.6-m、ABE8.7-m、ABE8.8-m、ABE8.9-m、ABE8.10-m、ABE8.11-m、ABE8.12-m、ABE8.13-m、ABE8.14-m、ABE8.15-m、ABE8.16-m、ABE8.17-m、ABE8.18-m、ABE8.19-m、ABE8.20-m、ABE8.21-m、ABE8.22-m、ABE8.23-m、ABE8.24-m、ABE8.1-d、ABE8.2-d、ABE8.3-d、ABE8.4-d、ABE8.5-d、ABE8.6-d、ABE8.7-d、ABE8.8-d、ABE8.9-d、ABE8.10-d、ABE8.11-d、ABE8.12-d、ABE8.13-d、ABE8.14-d、ABE8.15-d、ABE8.16-d、ABE8.17-d、ABE8.18-d、ABE8.19-d、ABE8.20-d、ABE8.21-d、ABE8.22-d、ABE8.23-d或ABE8.24-d,如下表8中所示。

[1345] 表8:腺苷脱氨酶碱基编辑器8变体

[1346]

ABE8	腺苷脱氨酶	腺苷脱氨酶说明
ABE8.1-m	TadA*8.1	单体_TadA*7.10 + Y147T
ABE8.2-m	TadA*8.2	单体_TadA*7.10 + Y147R
ABE8.3-m	TadA*8.3	单体_TadA*7.10 + Q154S
ABE8.4-m	TadA*8.4	单体_TadA*7.10 + Y123H
ABE8.5-m	TadA*8.5	单体_TadA*7.10 + V82S
ABE8.6-m	TadA*8.6	单体_TadA*7.10 + T166R
ABE8.7-m	TadA*8.7	单体_TadA*7.10 + Q154R
ABE8.8-m	TadA*8.8	单体_TadA*7.10 + Y147R_Q154R_Y123H
ABE8.9-m	TadA*8.9	单体_TadA*7.10 + Y147R_Q154R_I76Y
ABE8.10-m	TadA*8.10	单体_TadA*7.10 + Y147R_Q154R_T166R
ABE8.11-m	TadA*8.11	单体_TadA*7.10 + Y147T_Q154R
ABE8.12-m	TadA*8.12	单体_TadA*7.10 + Y147T_Q154S
ABE8.13-m	TadA*8.13	单体_TadA*7.10 + Y123H_Y147R_Q154R_I76Y
ABE8.14-m	TadA*8.14	单体_TadA*7.10 + I76Y_V82S
ABE8.15-m	TadA*8.15	单体_TadA*7.10 + V82S_Y147R
ABE8.16-m	TadA*8.16	单体_TadA*7.10 + V82S_Y123H_Y147R
ABE8.17-m	TadA*8.17	单体_TadA*7.10 + V82S_Q154R
ABE8.18-m	TadA*8.18	单体_TadA*7.10 + V82S_Y123H_Q154R
ABE8.19-m	TadA*8.19	单体_TadA*7.10 + V82S_Y123H_Y147R_Q154R
ABE8.20-m	TadA*8.20	单体_TadA*7.10 + I76Y_V82S_Y123H_Y147R_ Q154R
ABE8.21-m	TadA*8.21	单体_TadA*7.10 + Y147R_Q154S
ABE8.22-m	TadA*8.22	单体_TadA*7.10 + V82S_Q154S
ABE8.23-m	TadA*8.23	单体_TadA*7.10 + V82S_Y123H
ABE8.24-m	TadA*8.24	单体_TadA*7.10 + V82S_Y123H_Y147T
ABE8.1-d	TadA*8.1	异源二聚体_(WT) + (TadA*7.10 + Y147T)
ABE8.2-d	TadA*8.2	异源二聚体_(WT) + (TadA*7.10 + Y147R)
ABE8.3-d	TadA*8.3	异源二聚体_(WT) + (TadA*7.10 + Q154S)
ABE8.4-d	TadA*8.4	异源二聚体_(WT) + (TadA*7.10 + Y123H)
ABE8.5-d	TadA*8.5	异源二聚体_(WT) + (TadA*7.10 + V82S)
ABE8.6-d	TadA*8.6	异源二聚体_(WT) + (TadA*7.10 + T166R)
ABE8.7-d	TadA*8.7	异源二聚体_(WT) + (TadA*7.10 + Q154R)

[1347]

ABE8.8-d	TadA*8.8	异源二聚体_(WT) + (TadA*7.10 + Y147R_Q154R_ Y123H)
ABE8.9-d	TadA*8.9	异源二聚体_(WT)+(TadA*7.10 + Y147R_Q154R_I76Y)
ABE8.10-d	TadA*8.10	异源二聚体_(WT) + (TadA*7.10 + Y147R_Q154R_T166R)
ABE8.11-d	TadA*8.11	异源二聚体_(WT) + (TadA*7.10 + Y147T_Q154R)
ABE8.12-d	TadA*8.12	异源二聚体_(WT) + (TadA*7.10 + Y147T_Q154S)
ABE8.13-d	TadA*8.13	异源二聚体_(WT) + (TadA*7.10 + Y123H_Y147T_Q154R_I76Y)
ABE8.14-d	TadA*8.14	异源二聚体_(WT) + (TadA*7.10 + I76Y_V82S)
ABE8.15-d	TadA*8.15	异源二聚体_(WT) + (TadA*7.10 + V82S_ Y147R)
ABE8.16-d	TadA*8.16	异源二聚体_(WT) + (TadA*7.10 + V82S_Y123H_Y147R)
ABE8.17-d	TadA*8.17	异源二聚体_(WT) + (TadA*7.10 + V82S_Q154R)
ABE8.18-d	TadA*8.18	异源二聚体_(WT) + (TadA*7.10 + V82S_Y123H_Q154R)
ABE8.19-d	TadA*8.19	异源二聚体_(WT) + (TadA*7.10 + V82S_Y123H_Y147R_Q154R)
ABE8.20-d	TadA*8.20	异源二聚体_(WT) + (TadA*7.10 + I76Y_V82S_Y123H_Y147R_ Q154R)
ABE8.21-d	TadA*8.21	异源二聚体_(WT) + (TadA*7.10 + Y147R_Q154S)
ABE8.22-d	TadA*8.22	异源二聚体_(WT) + (TadA*7.10 + V82S_Q154S)
ABE8.23-d	TadA*8.23	异源二聚体_(WT) + (TadA*7.10 + V82S_Y123H)
ABE8.24-d	TadA*8.24	异源二聚体_(WT) + (TadA*7.10 + V82S_Y123H_Y147T)

[1348] 在一些实施方案中,通过将腺苷脱氨酶变体(例如,TadA*8)克隆到包括环状完全突变的Cas9(例如,CP5或CP6)和二分体核定位序列的支架中而产生碱基编辑器(例如,ABE8)。在一些实施方案中,碱基编辑器(例如,ABE7.9、ABE7.10或ABE8)是NGC PAM CP5变体(酿脓链球菌Cas9或spVRQR Cas9)。在一些实施方案中,碱基编辑器(例如,ABE7.9、ABE7.10或ABE8)是AGA PAM CP5变体(酿脓链球菌Cas9或spVRQR Cas9)。在一些实施方案中,碱基编辑器(例如,ABE7.9、ABE7.10或ABE8)是NGC PAM CP6变体(酿脓链球菌Cas9或spVRQR Cas9)。在一些实施方案中,碱基编辑器(例如,ABE7.9、ABE7.10或ABE8)是AGA PAM CP6变体(酿脓链球菌Cas9或spVRQR Cas9)。

[1349] 在一些实施方案中,ABE具有下表9中所示的基因型。

[1350] 表9.ABE的基因型

[1351]

	23	26	36	37	48	49	51	72	84	87	105	108	123	125	142	145	147	152	155	156	157	161
ABE7.9	L	R	L	N	A		L	N	F	S	V	N	Y	G	N	C	Y	P	V	F	N	K
ABE7.10	R	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	P	V	F	N	K

[1352] 如下表10中所示,描述了40种ABE8的基因型。指明了ABE的进化的大肠杆菌TadA部分中的残基位置。当不同于ABE7.10突变时,显示了ABE8中的突变性变化。在一些实施方案中,ABE具有下表10中所示的ABE之一的基因型。

[1353] 表10.进化的TadA中的残基身份

[1354]

	23	36	48	51	76	82	84	106	108	123	146	147	152	154	155	156	157	166
ABE7.10	**R**	**L**	**A**	**L**	**I**	**V**	**F**	**V**	**N**	**Y**	**C**	**Y**	**P**	**Q**	**V**	**F**	**N**	**T**
ABE8.1-m												T						
ABE8.2-m												R						
ABE8.3-m														S				
ABE8.4-m										H								
ABE8.5-m						S												
ABE8.6-m																		R
ABE8.7-m														R				
ABE8.8-m										H		R		R				
ABE8.9-m					Y							R		R				
ABE8.10-m												R		R				R
ABE8.11-m												T		R				
ABE8.12-m												T		S				
ABE8.13-m					Y					H		R		R				
ABE8.14-m					Y	S												
ABE8.15-m						S						R						
ABE8.16-m						S				H		R						
ABE8.17-m						S								R				
ABE8.18-m						S				H				R				
ABE8.19-m						S				H		R		R				
ABE8.20-m					Y	S				H		R		R				
ABE8.21-m												R		S				
ABE8.22-m						S								S				
ABE8.23-m						S				H								
ABE8.24-m						S				H		T						
ABE8.1-d												T						
ABE8.2-d												R						
ABE8.3-d														S				
ABE8.4-d										H								
ABE8.5-d						S												

[1355]

ABE8.6-d																		R
ABE8.7-d														R				
ABE8.8-d										H		R		R				
ABE8.9-d					Y							R		R				
ABE8.10-d												R		R				R
ABE8.11-d												T		R				
ABE8.12-d												T		S				
ABE8.13-d					Y					H		R		R				
ABE8.14-d					Y	S												
ABE8.15-d						S						R						
ABE8.16-d						S				H		R						
ABE8.17-d						S								R				
ABE8.18-d						S				H				R				
ABE8.19-d						S				H		R		R				
ABE8.20-d					Y	S				H		R		R				
ABE8.21-d												R		S				
ABE8.22-d						S								S				
ABE8.23-d						S				H								
ABE8.24-d						S				H		T						

[1356] 在一些实施方案中,碱基编辑器是ABE8.1,其包含以下序列或其具有腺苷脱氨酶活性的片段或基本上由其组成:

[1357] ABE8.1_Y147T_CP5_NGC PAM_单体

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRL
IDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCTFFR
MPRQVFNAQKKAQSSTD***SGGSSGGSSGSETPGTSESATPESSGGSSGGS***EIGKATAKYFFYSNIMNFFKTEITL
ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDW
DPKKYGGFMQPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAKFLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQIS
EFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYRSTKEVLDATLI
[1358] **HQSITGLYETRIDLSQLGGD*GGSGGSGGSGGSGGSGGSGGM*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKV**
LGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVE
EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVD
KLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQ
DLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRT
FDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNF
EEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLL

[1359] FKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLF EDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQ KNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLK DDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLV ETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTA LIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEGADKRTADGSEFESPKKKRKV*

[1360] 在上述序列中,正常文字表示腺苷脱氨酶序列,粗体序列指示衍生自Cas9的序列,斜体序列表示连接子序列,而下划线序列表示二分体核定位序列。

[1361] 在一些实施方案中,碱基编辑器是ABE8.1,其包含以下序列或其具有腺苷脱氨酶活性的片段或基本上由其组成:

[1362] pNMG-B335 ABE8.1_Y147T_CP5_NGC PAM_单体:

[1363] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRL IDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCTFFR MPRQVFNAQKKAQSSTDSGGSSG*GSSGSETPGTSESATPESSGGSSGGS*EIGKATAKYFFYSNIMNFFKTEITLA NGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD PKKYGGFMQPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKY SLFELENGRKRMLASAKFLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISE FSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYRSTKEVLDATLIH QSITGLYETRIDLSQLGGD*GGSGGSGGSGGSGGSGGSGGM*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVL GNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEE DKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDK LFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFD LAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQD LTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLF KTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFE DREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDS LTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQK NSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKD DSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVE TRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAL IKKYPKLESEFVYGDYKVYDVRKMIAKSEQEGADKRTADGSEFESPKKKRKV*

[1364] 在上述序列中,正常文字表示腺苷脱氨酶序列,粗体序列指示衍生自Cas9的序列,斜体序列表示连接子序列,而下划线序列表示二分体核定位序列。

[1365] 在一些实施方案中,碱基编辑器是ABE8.14,其包含以下序列或其具有腺苷脱氨酶活性的片段或基本上由其组成:

[1366] 具有NGC PAM CP5的pNMG-357_ABE8.14

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRL
IDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFR
MRRQEIKAQKKAQSSTDGGSSGGS*SGSETPGTSESATPESSGGSSGGS*MSEVEFSHEYWMRHALTLAKRARDERE
VPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRV
VFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD*SGGSSGGSSG
*SETPGTSESATPESSGGSSGGS*EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRD
FATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPTVAYSVLVVAKVEKGK
SKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAKFLQKGNELA
LPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKP
IREQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD*GGSGGSG
*GSGGSGGSGGSGGM*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEA
TRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI
YHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAK
AILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIG
DQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN
GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYP
FLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNE
KVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVE
ISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKR
RRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL
AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPV
ENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEV
VKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIA
KSEQ*EGADKRTADGSEFESPKKKRKV**

[1367]

[1368] 在上述序列中,正常文字表示腺苷脱氨酶序列,粗体序列指示衍生自Cas9的序列,斜体序列表示连接子序列,而下划线序列表示二分体核定位序列。

[1369] 在一些实施方案中,碱基编辑器是ABE8.8-m,其包含以下序列或其具有腺苷脱氨酶活性的片段或基本上由其组成:

[1370] ABE8.8-m

[1371] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRL
IDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFR
MPRRVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGS*DKKYSIGLAIGTNSVGWAVITDEYKV
PSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRL
EESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLT
PNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIK
RYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED
LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEE
TITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQK
KAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDI
VLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF
MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMAREN
QTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHI
VPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAG
FIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLI
ETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSP
TVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRK
RMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA
NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[1372] 在上述序列中,正常文字表示腺苷脱氨酶序列,粗体序列指示衍生自Cas9的序列,斜体序列表示连接子序列,下划线序列表示二分体核定位序列,而双下划线序列指示突变。

[1373] 在一些实施方案中,碱基编辑器是ABE8.8-d,其包含以下序列或其具有腺苷脱氨酶活性的片段或基本上由其组成:

[1374] ABE8.8-d

[1375] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRL
IDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFR
MRRQEIKAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGS*SEVEFSHEYWMRHALTLAKRARDERE

[1376] VPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRV
VFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSG*
*SETPGTSESATPESSGGSSGGS*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF
DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVA
YHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPI
NASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDL
DNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKE
IFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAIL
RRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTN
FDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKK
IECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDD
KVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDS
LHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGS
QILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKS
DNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT
KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKV
YDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKEL
LGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLY
LASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIH
LFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPK
KKRKV*

[1377] 在上述序列中,正常文字表示腺苷脱氨酶序列,粗体序列指示衍生自Cas9的序列,斜体序列表示连接子序列,下划线序列表示二分体核定位序列,而双下划线序列指示突变。

[1378] 在一些实施方案中,碱基编辑器是ABE8.13-m,其包含以下序列或其具有腺苷脱氨酶活性的片段或基本上由其组成:

[1379] ABE8.13-m

[1380] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRL
YDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFR
MPRRVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGS*DKKYSIGLAIGTNSVGWAVITDEYKV
PSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRL
EESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLT
PNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIK

RYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED
LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEE
TITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQK
KAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDI
VLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF
MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMAREN
QTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHI
[1381] VPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAG
FIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLI
ETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSP
TVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRK
RMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA
NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[1382] 在上述序列中,正常文字表示腺苷脱氨酶序列,粗体序列指示衍生自Cas9的序列,斜体序列表示连接子序列,下划线序列表示二分体核定位序列,而双下划线序列指示突变。

[1383] 在一些实施方案中,碱基编辑器是ABE8.13-d,其包含以下序列或其具有腺苷脱氨酶活性的片段或基本上由其组成:

[1384] ABE8.13-d

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRL
IDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFR
MRRQEIKAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGS*SEVEFSHEYWMRHALTLAKRARDERE
VPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRV
VFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSG*
*SETPGTSESATPESSGGSSGGS*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF
DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVA
[1385] YHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPI
NASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDL
DNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKE
IFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAIL
RRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTN
FDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKK
IECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDD

KVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDS
LHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGS
QILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKS
DNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT
KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKV
[1386] YDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKEL
LGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLY
LASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIH
LFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPK
KKRKV*

[1387] 在上述序列中,正常文字表示腺苷脱氨酶序列,粗体序列指示衍生自Cas9的序列,斜体序列表示连接子序列,下划线序列表示二分体核定位序列,而双下划线序列指示突变。

[1388] 在一些实施方案中,碱基编辑器是ABE8.17-m,其包含以下序列或其具有腺苷脱氨酶活性的片段或基本上由其组成:

[1389] ABE8.17-m

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRL
IDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFR
MPRRVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGS*DKKYSIGLAIGTNSVGWAVITDEYKV
PSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRL
EESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLT
PNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIK
RYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED
LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEE
[1390] TITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQK
KAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDI
VLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF
MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMAREN
QTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHI
VPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAG
FIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLI
ETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSP
TVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRK
RMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA
[1391] NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[1392] 在上述序列中,正常文字表示腺苷脱氨酶序列,粗体序列指示衍生自Cas9的序列,

斜体序列表示连接子序列,下划线序列表示二分体核定位序列,而双下划线序列指示突变。

[1393] 在一些实施方案中,碱基编辑器是ABE8.17-d,其包含以下序列或其具有腺苷脱氨酶活性的片段或基本上由其组成:

[1394] ABE8.17-d

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRL
IDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFR
MRRQEIKAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGS*SEVEFSHEYWMRHALTLAKRARDERE
VPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYSTFEPCVMCAGAMIHSRIGRV
VFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSG*
*SETPGTSESATPESSGGSSGGS*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF
DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVA
YHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPI
NASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDL
DNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKE
IFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAIL
RRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTN
FDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKK
IECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDD
KVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDS
LHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGS
QILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKS
DNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT
KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKV
YDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKEL
LGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLY
LASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIH
LFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD*EGADKRTADGSEFESPK*
*KKRKV**

[1396] 在上述序列中,正常文字表示腺苷脱氨酶序列,粗体序列指示衍生自Cas9的序列,斜体序列表示连接子序列,下划线序列表示二分体核定位序列,而双下划线序列指示突变。

[1397] 在一些实施方案中,碱基编辑器是ABE8.20-m,其包含以下序列或其具有腺苷脱氨酶活性的片段或基本上由其组成:

[1398] ABE8.20-m

[1399]

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRL
YDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFR
MPRRVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGS*DKKYSIGLAIGTNSVGWAVITDEYKV
PSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRL
EESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLT
PNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIK
RYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED
LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEE
TITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQK
KAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDI
VLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF
MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMAREN
QTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHI
VPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAG
FIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLI
ETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSP
TVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRK
RMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA
NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[1400] 在上述序列中,正常文字表示腺苷脱氨酶序列,粗体序列指示衍生自Cas9的序列,斜体序列表示连接子序列,下划线序列表示二分体核定位序列,而双下划线序列指示突变。

[1401] 在一些实施方案中,碱基编辑器是ABE8.20-d,其包含以下序列或其具有腺苷脱氨酶活性的片段或基本上由其组成:

[1402] ABE8.20-d

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRL
IDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFR
MRRQEIKAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGS*SEVEFSHEYWMRHALTLAKRARDERE
VPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYSTFEPCVMCAGAMIHSRIGRV
VFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSG*
*SETPGTSESATPESSGGSSGGS*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF
DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVA
YHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPI
NASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDL
DNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKE
IFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAIL
RRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTN
[1403] FDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKK
IECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDD
KVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDS
LHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGS
QILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKS
DNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT
KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKV
YDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKEL
LGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLY
LASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIH
LFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPK
KKRKV*

[1404] 在上述序列中,正常文字表示腺苷脱氨酶序列,粗体序列指示衍生自Cas9的序列,斜体序列表示连接子序列,下划线序列表示二分体核定位序列,而双下划线序列指示突变。

[1405] 在一些实施方案中,本发明的ABE8选自以下序列:

[1406] 01.单ABE8.1_bpNLS+Y147T

[1407] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
TFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL

KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD
SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[1408] 02.单ABE8.1_bpNLS+Y147R

[1409] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
RFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD
SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[1410] 03.单ABE8.1_bpNLS+Q154S

[1411] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
YFFRMPRSVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS

DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD
SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[1412] 04.单ABE8.1_bpNLS+Y123H

[1413] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLC
YFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD
SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[1414] 05.单ABE8.1_bpNLS+V82S

[1415] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
YFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD
SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[1416] 06.单ABE8.1_bpNLS+T166R

[1417] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
YFFRMPRQVFNAQKKAQSSRDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD
SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG

GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[1418] 07.单ABE8.1_bpNLS+Q154R

[1419] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
YFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD
SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[1420] 08.单ABE8.1_bpNLS+Y147R_Q154R_Y123H

[1421] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLC
RFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD

SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[1422] 09.单ABE8.1_bpNLS+Y147R_Q154R_I76Y

[1423] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
RFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD
SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[1424] 10.单ABE8.1_bpNLS+Y147R_Q154R_T166R

[1425] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
RFFRMPRRVFNAQKKAQSSRDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT

LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD
SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[1426] 11.单ABE8.1_bpNLS+Y147T_Q154R

[1427] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
TFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD
SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[1428] 12.单ABE8.1_bpNLS+Y147T_Q154S

[1429] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC

TFFRMPRSVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD
SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[1430] 13.单ABE8.1_bpNLS+H123Y123H_Y147R_Q154R_I76Y

[1431] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLC
RFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD
SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL

FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[1432] 14.单ABE8.1_bpNLS+V82S+Q154R

[1433] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
YFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD
SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[1434] 在一些实施方案中,碱基编辑器基因不包含一个结构域,该结构域包含尿嘧啶糖苷酶抑制剂(UGI)的全部或一部分。在一些实施方案中,碱基编辑器基因不包含一个结构域,该结构域包含尿嘧啶结合蛋白(UBP)(诸如尿嘧啶DNA糖苷酶(UDG))的全部或一部分。在一些实施方案中,碱基编辑器基因不包含一个结构域,该结构域包含核酸聚合酶的全部或一部分。在一些实施方案中,被并入碱基编辑器中的核酸聚合酶或其部分是跨损伤DNA聚合酶。

[1435] 在一些实施方案中,碱基编辑器的结构域可以包含多个结构域。例如,包含衍生自Cas9的多核苷酸可编程核苷酸结合结构域的碱基编辑器可以包含对应于野生型或天然Cas9的REC叶和NUC叶的REC叶和NUC叶。在另一实施例中,碱基编辑器可以包含RuvCI结构域、BH结构域、REC1结构域、REC2结构域、RuvCII结构域、L1结构域、HNH结构域、L2结构域、RuvCIII结构域、WED结构域、TOPO结构域或CTD结构域中的一个或多个。在一些实施方案中,碱基编辑器的一个或多个结构域包含相对于包含该结构域的野生型版本多肽的突变(例如,置换、插入、缺失)。例如,多核苷酸可编程DNA结合结构域的HNH结构域可以包含H840A置换。在另一实施例中,多核苷酸可编程DNA结合结构域的RuvCI结构域可以包含D10A置换。

[1436] 本文所公开的碱基编辑器的不同结构域(例如,邻近结构域)可以使用或不是使用

一个或多个连接子结构域(例如,XTEN连接子结构域)彼此联接。在一些实施方案中,连接子结构域可以是键(例如,共价键)、化学基团或连接连个分子或部分(例如,融合蛋白的两个结构域,诸如举例而言,第一结构域(例如,衍生自Cas9的结构域)和第二结构域(例如,腺苷脱氨酶结构或胞苷脱氨酶结构域))的分子。在一些实施方案中,连接子是共价键(例如,碳-碳键、二硫键、碳-杂原子键等)。在某些实施方案中,连接子是酰胺链结的碳-氮键。在某些实施方案中,连接子是环状或非环状的、经取代或未经取代的、分支的或未分支的脂肪族或杂脂肪族连接子。在某些实施方案中,连接子是聚合性的(例如,聚乙烯、聚乙二醇、聚酰胺、聚酯等)。在某些实施方案中,连接子包含氨基烷酸的单体、二聚体或聚合物。在一些实施方案中,连接子包含氨基烷酸(例如,甘氨酸、乙酸、丙氨酸、β-丙氨酸、3-氨基丙酸、4-氨基丁酸、5-戊酸等)。在一些实施方案中,连接子包含氨基己酸(Ahx)的单体、二聚体或聚合物。在某些实施方案中,连接子是基于碳环部分(例如,环戊烷、环己烷)的。在其他实施方案中,连接子包含聚乙二醇部分(PEG)。在某些实施方案中,连接子包含芳基或杂芳基部分。在某些实施方案中,连接子是基于苯环的。连接子可以包括经功能化的部分以促进来自肽的亲核体(例如,巯基、氨基)附接至连接子。任何亲电体均可用作连接子的一部分。示例性的亲电体包括但不限于,活化的酯、活化的酰胺、Michael受体、烷基卤化物、芳基卤化物、酰卤和异硫氰酸酯。在一些实施方案中,连接子接合RNA可编程核酸酶的gRNA接合结构域(包括Cas9核酸酶结构域)与核酸编辑蛋白的催化结构域。在一些实施方案中,连接子接合dCas9与第二结构域(例如,UGI、胞苷脱氨酶等)。

[1437] 典型地,连接子可以定位在两个基团、分子或其他部分之间或者侧翼具有两个基团、分子或其他部分,并且经由共价键与每个基团、分子或其他部分连结,从而将两者连结。在一些实施方案中,连接子是氨基酸或多个氨基酸(例如,肽或蛋白质)。在一些实施方案中,连接子是有机分子、基团、聚合物或化学部分。在一些实施方案中,连接子为2至100个氨基酸的长度,例如,2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、30至35、35至40、40至45、45至50、50至60、60至70、70至80、80至90、90至100、100至150或150至200个氨基酸的长度。在一些实施方案中,连接子为约3至约104(例如,5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、55、60、65、70、75、80、85、90、95或100)个氨基酸的长度。也设想了更长或更短的连接子。在一些实施方案中,连接子结构域包含氨基酸序列SGSETPGTSESATPES,其也可以称为XTEN连接子。可以采用任何用于连接融合蛋白结构域的方法(例如,从非常柔性的(SGGS)n、(GGGS)n、(GGGGS)n和(G)n形式的连接子到更为刚性的(EAAAK)n、(GGS)n、SGSETPGTSESATPES形式的连接子(参见例如,Guilinger JP,Thompson DB,Liu DR.Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification.Nat.Biotechnol.2014;32(6):577-82;其整体内容通过引用并入本文)范围内,或(XP)n基序),以便实现针对核碱基编辑器活性的最优长度。在一些实施方案中,n是1、2、3、4、5、6、7、8、9、10、11、12、13、14或15。在一些实施方案中,连接子包含(GGS)n基序,其中n是1、3或7。在一些实施方案中,本文所提供的融合蛋白的Cas9结构域经由包含氨基酸序列SGSETPGTSESATPES的连接子融合。在一些实施方案中,连接子包含多个脯氨酸残基,并且为5至21、5至14、5至9、5至7个氨基酸的长度,例如,PAPAP、PAPAPA、PAPAPAP、PAPAPAPA、P(AP)4、P(AP)7、P(AP)10(参见例

如,Tan J,Zhang F,Karcher D,Bock R.Engineering of high-precision base editors for site-specific single nucleotide replacement.Nat Commun.2019Jan25;10(1):439;其整体内容通过引用并入本文)。此类富含脯氨酸的连接子也称为"刚性"连接子。

[1438] 连接子

[1439] 在某些实施方案中,连接子可以用来连接本发明的任何多肽或多肽结构域。连接子可以简单到是共价键,或者它可以是长度为很多原子的聚合性连接子。在某些实施方案中,连接子是多肽或是基于多个氨基酸的。在其他实施方案中,连接子不是肽样的。在某些实施方案中,连接子是共价键(例如,碳-碳键、二硫键、碳-杂原子键等)。在某些实施方案中,连接子是酰胺链结的碳-氮键。在某些实施方案中,连接子是环状或非环状的、经取代或未经取代的、分支的或未分支的脂肪族或杂脂肪族连接子。在某些实施方案中,连接子是聚合性的(例如,聚乙烯、聚乙二醇、聚酰胺、聚酯等)。在某些实施方案中,连接子包含氨基烷酸的单体、二聚体或聚合物。在某些实施方案中,连接子包含氨基烷酸(例如,甘氨酸、乙酸、丙氨酸、β-丙氨酸、3-氨基丙酸、4-氨基丁酸、5-戊酸等)。在某些实施方案中,连接子包含氨基己酸(Ahx)的单体、二聚体或聚合物。在某些实施方案中,连接子是基于碳环部分(例如,环戊烷、环己烷)的。在其他实施方案中,连接子包含聚乙二醇部分(PEG)。在其他实施方案中,连接子包含氨基酸。在某些实施方案中,连接子包含肽。在某些实施方案中,连接子包含芳基或杂芳基部分。在某些实施方案中,连接子是基于苯环的。连接子可以包括经功能化的部分以促进来自肽的亲核体(例如,巯基、氨基)附接至连接子。任何亲电体均可用作连接子的一部分。示例性的亲电体包括但不限于,活化的酯、活化的酰胺、Michael受体、烷基卤化物、芳基卤化物、酰卤和异硫氰酸酯。

[1440] 在一些实施方案中,连接子是氨基酸或多个氨基酸(例如,肽或蛋白质)。在一些实施方案中,连接子是键(例如,共价键)、有机分子、基团、聚合物或化学部分。在一些实施方案中,连接子为约3至约104(例如,5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、55、60、65、70、75、80、85、90、95或100)个氨基酸的长度。

[1441] 在一些实施方案中,胞苷脱氨酶和腺苷脱氨酶与napDNAbp经由长度为4、16、32或104个氨基酸的连接子融合。在一些实施方案中,连接子为约3至约104个氨基酸的长度。在一些实施方案中,本文所提供的任何融合蛋白包含经由连接子融合至彼此的胞苷脱氨酶、腺苷脱氨酶和Cas9结构域。可以在脱氨酶结构域(例如,经工程化的ecTadA)与Cas9结构域之间采用各种连接子长度和柔软度(例如,从非常柔性的$(GGGS)_n$、$(GGGGS)_n$和$(G)_n$形式的连接子到更为刚性的$(EAAAK)_n$、$(SGGS)_n$、SGSETPGTSESATPES形式的连接子(参见例如,Guilinger JP,Thompson DB,Liu DR.Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification.Nat.Biotechnol.2014;32(6):577-82;整体内容通过引用并入本文)范围内和$(XP)_n$),以便实现针对核碱基编辑器的活性的最优长度。在一些实施方案中,n是1、2、3、4、5、6、7、8、9、10、11、12、13、14或15。在一些实施方案中,连接子包含$(GGS)_n$基序,其中n是1、3或7。在一些实施方案中,本文提供的任何融合蛋白的胞苷芳与腺苷脱氨酶与Cas9结构域经由(例如,XTEN连接子)融合,该连接子包含氨基酸序列SGSETPGTSESATPES。

[1442] 在一些实施方案中,靶标区域包含靶标窗,其中靶标窗包含靶标核碱基对。在一些

实施方案中,靶标窗包含1至10个核苷酸。在一些实施方案中,靶标窗为1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个核苷酸的长度。在一些实施方案中,预期的碱基对编辑位于靶标窗内。在一些实施方案中,靶标窗包含预期的碱基对编辑。在一些实施方案中,使用本文所提供的任何碱基编辑器执行该方法。在一些实施方案中,靶标窗是脱氨基窗。

[1443] 此外,在一些情况下,Gam蛋白可以融合至碱基编辑器的N端。在一些情况下,Gam蛋白可以融合至碱基编辑器的C端。细菌噬菌体Mu的Gam蛋白可以结合至双链断裂(DSB)的末端并且保护它们免于降解。在一些实施方案中,使用Gam来结合DSB的自由末端可以减少碱基编辑过程中的插入缺失形成。在一些实施方案中,将174个残基的Gam蛋白融合至碱基编辑器的N端。参见,Komor,A.C.,et al.,"Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity"Science Advances 3:eaao4774(2017)。在一些情况下,一个或多个突变可以改变碱基编辑器结构域相对于野生型结构域的长度。例如,在至少一个结构域中至少一个氨基酸的缺失可以缩短碱基编辑器的长度。在另一种情况下,一个或多个突变不改变结构域相对于野生型结构域的长度。例如,任何结构域中的一种或多种置换不改变碱基编辑器的长度。

[1444] 在一些实施方案中,本文所提供的碱基编辑融合蛋白需要被定位在精确位置处,例如,靶标碱基被放置在定义区域(例如,"脱氨基窗")内。在一些情况下,靶标可以位于4个碱基的区域内。在一些情况下,此类定义的靶标区域可以位于PAM的上游大约15个碱基处。参见,Komor,A.C.,etal.,"Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage"Nature 533,420-424(2016);Gaudelli,N.M.,et al.,"Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage"Nature 551,464-471(2017)和Komor,A.C.,et al.,"Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity"Science Advances3:eaao4774(2017)),其整体内容通过引用并入本文。

[1445] 定义的靶标区域可以是脱氨基窗。脱氨基窗可以是定义的区域,在该区域中,碱基编辑器作用于靶标核苷酸并将该靶标核苷酸脱氨基。在一些实施方案中,脱氨基窗位于2、3、4、5、6、7、8、9或10个碱基的区域内。在一些实施方案中,脱氨基窗位于PAM的上游5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24或25个碱基处。

[1446] 本公开的碱基编辑器可以包含任何促进靶标多核苷酸序列的编辑的结构域、特征和氨基酸序列。例如,在一些实施方案中,碱基编辑器包含核定位序列(NLS)。在一些实施方案中,碱基编辑器的NLS定位在脱氨酶结构域与多核苷酸可编程核苷酸结合结构域之间。在一些实施方案中,碱基编辑器的NLS定位在多核苷酸可编程核苷酸结合结构域的C端。

[1447] 可存在于本所公开的碱基编辑器中的其他示例性特征是定位序列,诸如细胞质定位序列、输出序列诸如核输出序列、或其他定位序列,以及可用于融合蛋白的溶解、纯化或检测的序列标签。本文所提供的合适的蛋白标签包括但不限于,生物素羧化酶载体蛋白(BCCP)标签、myc标签、钙调蛋白标签、FLAG标签、血细胞凝集素(HA)标签、聚组氨酸标签(也称为组氨酸标签或His标签)、麦芽糖结合蛋白(MBP)标签、nus标签、谷胱甘肽-S-转移酶

(GST)标签、绿色荧光蛋白(GFP)标签、硫氧还蛋白标签、S标签、Softag(例如,Softag 1、Softag 3)、链霉素标签、生物素连接酶标签、FlAsH标签、V5标签和SBP标签。其他合适的序列对于本领域技术人员将是显而易见的。在一些实施方案中,融合蛋白包含一个或多个His标签。

[1448] 可以包括在融合蛋白中的蛋白结构域的非限制性示例包括脱氨酶结构域(例如,胞苷脱氨酶、腺苷脱氨酶)、尿嘧啶糖苷酶抑制剂(UGI)结构域、表位标签和报告基因序列。

[1449] 表位标签的非限制性示例包括组氨酸(His)标签、V5标签、FLAG标签、流感病毒血凝素(HA)标签、Myc标签、VSV-G标签和硫氧还蛋白(Trx)标签。报告基因的示例包括但不限于,谷胱甘肽-5-转移酶(GST)、辣根过氧化物酶(HRP)、氯霉素乙酰转移酶(CAT)、β-半乳糖苷酶、β-葡糖醛酸糖苷酶、荧光素酶、绿色荧光蛋白(GFP)、HcRed、DsRed、青色荧光蛋白(CFP)、黄色荧光蛋白(YFP)和包括蓝色荧光蛋白(BFP)在内的自体荧光蛋白。额外蛋白质序列可以包括结合DNA分子或结合其他细胞分子的氨基酸序列,包括但不限于,麦芽糖结合蛋白(MBP)、S标签、Lex A DNA结合结构域(DBD)融合物、GAL4 DNA结合结构域融合物和单纯疱疹病毒(HSV)BP16蛋白融合物。

[1450] Base编辑器效率

[1451] CRISPR-Cas9核酸酶已经广泛用于介导靶向基因组编辑。在大多数基因组编辑应用中,Cas9与向导多核苷酸(例如,单向导RNA(sgRNA))形成复合物并且诱导被该sgRNA序列针对的靶标位点处的双链DNA断裂(DSB)。细胞主要通过非同源末端接合(NHEJ)修复途径来应答这一DSB,导致可能造成中断该基因的移码突变的随机插入或缺失(indel)。在存在与位于DSB侧翼的序列具有高度同源性的供体DNA目标的情况下,可以通过被称为同源引导修复(HDR)的替代途径实现基因改变。不幸的是,在大多数非微扰条件下,HDR是低效的,取决于细胞状态和细胞类型,并且更高频率的插入缺失占主导地位。本文所提供的碱基编辑系统提供新的途径来提供基因组编辑而不产生双链DNA断裂,不需要供体DNA目标,并且不诱导过量的随机插入和缺失。

[1452] 本文提供的碱基编辑器能够修饰特定核苷酸碱基而不产生显著的插入缺失部分。如本文所用,术语“插入缺失(indel)”是指核酸内核苷酸碱基的插入或缺失。此类插入或缺失可以导致基因的编码区域内的移码突变。在一些实施方案中,希望产生碱基编辑器,其有效地修饰(例如,突变或脱氨基)核酸内的特定核苷酸而不在靶标核苷酸序列中产生大量的插入或缺失(即,插入缺失)。在某些实施方案中,本文提供的任何碱基编辑器都能够产生相比于插入缺失更符合预期的修饰(例如,点突变或脱氨基)部分。

[1453] 在一些实施方案中,本文提供的任何碱基编辑器系统导致在靶标多核苷酸序列中低于50%、低于40%、低于30%、低于20%、低于19%、低于18%、低于17%、低于16%、低于15%、低于14%、低于13%、低于12%、低于11%、低于10%、低于9%、低于8%、低于7%、低于6%、低于5%、低于4%、低于3%、低于2%、低于1%、低于0.9%、低于0.8%、低于0.7%、低于0.6%、低于0.5%、低于0.4%、低于0.3%、低于0.2%、低于0.1%、低于0.09%、低于0.08%、低于0.07%、低于0.06%、低于0.05%、低于0.04%、低于0.03%、低于0.02%或低于0.01%的插入缺失形成。

[1454] 在一些实施方案中,包含本文所述的一种ABE8碱基编辑器变体的任何碱基编辑器系统导致在靶标多核苷酸序列中低于50%、低于40%、低于30%、低于20%、低于19%、低于

18%、低于17%、低于16%、低于15%、低于14%、低于13%、低于12%、低于11%、低于10%、低于9%、低于8%、低于7%、低于6%、低于5%、低于4%、低于3%、低于2%、低于1%、低于0.9%、低于0.8%、低于0.7%、低于0.6%、低于0.5%、低于0.4%、低于0.3%、低于0.2%、低于0.1%、低于0.09%、低于0.08%、低于0.07%、低于0.06%、低于0.05%、低于0.04%、低于0.03%、低于0.02%或低于0.01%的插入缺失形成。

[1455] 在一些实施方案中,包含本文所述的一种ABE8碱基编辑器变体的任何碱基编辑器系统导致在靶标多核苷酸序列中低于0.8%的插入缺失形成。在一些实施方案中,包含本文所述的一种ABE8碱基编辑器变体的任何碱基编辑器系统导致在靶标多核苷酸序列中至多0.8%的插入缺失形成。在一些实施方案中,包含本文所述的一种ABE8碱基编辑器变体的任何碱基编辑器系统导致在靶标多核苷酸序列中低于0.3%的插入缺失形成。在一些实施方案中,与包含一种ABE7碱基编辑器的碱基编辑器系统相比,包含本文所述的一种ABE8碱基编辑器变体的任何碱基编辑器系统导致在靶标多核苷酸序列中更低的插入缺失形成。在一些实施方案中,与包含ABE7.10的碱基编辑器系统相比,包含本文所述的一种ABE8碱基编辑器变体的任何碱基编辑器系统导致在靶标多核苷酸序列中更低的插入缺失形成。

[1456] 在一些实施方案中,与包含一种ABE7碱基编辑器的碱基编辑器系统相比,包含本文所述的一种ABE8碱基编辑器变体的任何碱基编辑器系统具有降低的插入缺失频率。在一些实施方案中,与包含一种ABE7碱基编辑器的碱基编辑器系统相比,包含本文所述的一种ABE8碱基编辑器变体的任何碱基编辑器系统具有降低至少0.01%、至少1%、至少2%、至少3%、至少4%、至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%或至少95%的插入缺失频率。在一些实施方案中,与包含ABE7.10的碱基编辑器系统相比,包含本文所述的一种ABE8碱基编辑器变体的碱基编辑器系统具有降低至少0.01%、至少1%、至少2%、至少3%、至少4%、至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%或至少95%的插入缺失频率。

[1457] 本公开提供腺苷脱氨酶变体(例如,ABE8变体),其具有增加的效率和特异性。特别地,本文所述的腺苷脱氨酶变体更有可能编辑多核苷酸内的所希望的碱基,并且更不可能编辑不预期改变的碱基(例如,“旁观者”)。

[1458] 在一些实施方案中,包含本文所述的一种ABE8碱基编辑器变体的任何碱基编辑系统具有减少的旁观者编辑或突变。在一些实施方案中,非预期编辑或突变是旁观者突变或旁观者编辑,例如,对靶标核苷酸序列的靶标窗中的非预期或非靶标位置中靶标碱基(例如,A或C)的碱基编辑。在一些实施方案中,与包含ABE7碱基编辑器例如ABE7.10的碱基编辑器系统相比,包含本文所述的一种ABE8碱基编辑器变体的任何碱基编辑系统具有减少的旁观者编辑或突变。在一些实施方案中,与包含ABE7碱基编辑器例如ABE7.10的碱基编辑器系统相比,包含本文所述的一种ABE8碱基编辑器变体的任何碱基编辑系统的旁观者编辑或突变减少至少1%、至少2%、至少3%、至少4%、至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%或至少99%。在一些实施方

案中,与包含ABE7碱基编辑器例如ABE7.10的碱基编辑器系统相比,包含本文所述的一种ABE8碱基编辑器变体的任何碱基编辑系统的旁观者编辑或突变减少至少1.1倍、至少1.2倍、至少1.3倍、至少1.4倍、至少1.5倍、至少1.6倍、至少1.7倍、至少1.8倍、至少1.9倍、至少2.0倍、至少2.1倍、至少2.2倍、至少2.3倍、至少2.4倍、至少2.5倍、至少2.6倍、至少2.7倍、至少2.8倍、至少2.9倍或至少3.0倍。

[1459] 在一些实施方案中,包含本文所述的一种ABE8碱基编辑器变体的任何碱基编辑系统具有减少的假编辑。在一些实施方案中,非预期编辑或突变是假突变或假编辑,例如,对基因组的非预期或非靶标区域中的靶标碱基(例如,A或C)的非特异性编辑或不依赖于向导的编辑。在一些实施方案中,与包含ABE7碱基编辑器例如ABE7.10的碱基编辑器系统相比,包含本文所述的一种ABE8碱基编辑器变体的任何碱基编辑系统具有减少的假编辑。在一些实施方案中,与包含ABE7碱基编辑器例如ABE7.10的碱基编辑器系统相比,包含本文所述的一种ABE8碱基编辑器变体的任何碱基编辑系统的假编辑减少至少1%、至少2%、至少3%、至少4%、至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%或至少99%。在一些实施方案中,与包含ABE7碱基编辑器例如ABE7.10的碱基编辑器系统相比,包含本文所述的一种ABE8碱基编辑器变体的任何碱基编辑系统的假编辑减少至少1.1倍、至少1.2倍、至少1.3倍、至少1.4倍、至少1.5倍、至少1.6倍、至少1.7倍、至少1.8倍、至少1.9倍、至少2.0倍、至少2.1倍、至少2.2倍、至少2.3倍、至少2.4倍、至少2.5倍、至少2.6倍、至少2.7倍、至少2.8倍、至少2.9倍或至少3.0倍。

[1460] 本公开的一些方面是基于以下认知的:本文提供的任何碱基编辑器都能够有效地在核酸(例如,受试者基因组中的核酸)中产生预期突变诸如点突变,而不产生显著数量的非预期突变,诸如非预期点突变(即,旁观者的突变)。在一些实施方案中,本文提供的任何碱基编辑器都能够产生至少0.01%的预期突变(即,至少0.01%的碱基编辑效率)。在一些实施方案中,本文提供的任何碱基编辑器都能够产生至少0.01%、1%、2%、3%、4%、5%、10%、15%、20%、25%、30%、40%、45%、50%、60%、70%、80%、90%、95%或99%的预期突变。

[1461] 在一些实施方案中,本文所述的任何ABE8碱基编辑器变体具有至少0.01%、至少1%、至少2%、至少3%、至少4%、至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%或至少99%的碱基编辑效率。在一些实施方案中,可以通过计算细胞群中被编辑的核碱基的百分比来测量碱基编辑效率。在一些实施方案中,如通过细胞群中被编辑的核碱基所测量的,本文所述的任何ABE8碱基编辑器变体具有至少0.01%、至少1%、至少2%、至少3%、至少4%、至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%或至少99%的碱基编辑效率。

[1462] 在一些实施方案中,本文所述的任何ABE8碱基编辑器变体具有比ABE7碱基编辑器高的碱基编辑效率。在一些实施方案中,与ABE7碱基编辑器例如ABE7.10相比,本文所述的任何ABE8碱基编辑器具有高出至少1%、至少2%、至少3%、至少4%、至少5%、至少10%、至

少15％、至少20％、至少25％、至少30％、至少35％、至少40％、至少45％、至少50％、至少55％、至少60％、至少65％、至少70％、至少75％、至少80％、至少85％、至少90％、至少95％、至少99％、至少100％、至少105％、至少110％、至少115％、至少120％、至少125％、至少130％、至少135％、至少140％、至少145％、至少150％、至少155％、至少160％、至少165％、至少170％、至少175％、至少180％、至少185％、至少190％、至少195％、至少200％、至少210％、至少220％、至少230％、至少240％、至少250％、至少260％、至少270％、至少280％、至少290％、至少300％、至少310％、至少320％、至少330％、至少340％、至少350％、至少360％、至少370％、至少380％、至少390％、至少400％、至少450％或至少500％的碱基编辑效率。

[1463] 在一些实施方案中,与ABE7碱基编辑器例如ABE7.10相比,本文所述的任何ABE8碱基编辑器具有高出至少1.1倍、至少1.2倍、至少1.3倍、至少1.4倍、至少1.5倍、至少1.6倍、至少1.7倍、至少1.8倍、至少1.9倍、至少2.0倍、至少2.1倍、至少2.2倍、至少2.3倍、至少2.4倍、至少2.5倍、至少2.6倍、至少2.7倍、至少2.8倍、至少2.9倍、至少3.0倍、至少3.1倍、至少3.2、至少3.3倍、至少3.4倍、至少3.5倍、至少3.6倍、至少3.7倍、至少3.8倍、至少3.9倍、至少4.0倍、至少4.1倍、至少4.2倍、至少4.3倍、至少4.4倍、至少4.5倍、至少4.6倍、至少4.7倍、至少4.8倍、至少4.9倍或至少5.0倍的碱基编辑效率。

[1464] 在一些实施方案中,本文所述的任何ABE8碱基编辑器变体具有至少0.01％、至少1％、至少2％、至少3％、至少4％、至少5％、至少10％、至少15％、至少20％、至少25％、至少30％、至少35％、至少40％、至少45％、至少50％、至少55％、至少60％、至少65％、至少70％、至少75％、至少80％、至少85％、至少90％、至少95％或至少99％的上靶碱基编辑效率。在一些实施方案中,如通过细胞群中被编辑的靶标核碱基所测量的,本文所述的任何ABE8碱基编辑器变体具有至少0.01％、至少1％、至少2％、至少3％、至少4％、至少5％、至少10％、至少15％、至少20％、至少25％、至少30％、至少35％、至少40％、至少45％、至少50％、至少55％、至少60％、至少65％、至少70％、至少75％、至少80％、至少85％、至少90％、至少95％或至少99％的上靶碱基编辑效率。

[1465] 在一些实施方案中,本文所述的任何ABE8碱基编辑器变体具有比ABE7碱基编辑器高的上靶碱基编辑效率。在一些实施方案中,与ABE7碱基编辑器例如ABE7.10相比,本文所述的任何ABE8碱基编辑器具有高出至少1％、至少2％、至少3％、至少4％、至少5％、至少10％、至少15％、至少20％、至少25％、至少30％、至少35％、至少40％、至少45％、至少50％、至少55％、至少60％、至少65％、至少70％、至少75％、至少80％、至少85％、至少90％、至少95％、至少99％、至少100％、至少105％、至少110％、至少115％、至少120％、至少125％、至少130％、至少135％、至少140％、至少145％、至少150％、至少155％、至少160％、至少165％、至少170％、至少175％、至少180％、至少185％、至少190％、至少195％、至少200％、至少210％、至少220％、至少230％、至少240％、至少250％、至少260％、至少270％、至少280％、至少290％、至少300％、至少310％、至少320％、至少330％、至少340％、至少350％、至少360％、至少370％、至少380％、至少390％、至少400％、至少450％或至少500％的上靶碱基编辑效率。

[1466] 在一些实施方案中,与ABE7碱基编辑器例如ABE7.10相比,本文所述的任何ABE8碱基编辑器具有高出至少1.1倍、至少1.2倍、至少1.3倍、至少1.4倍、至少1.5倍、至少1.6倍、

至少1.7倍、至少1.8倍、至少1.9倍、至少2.0倍、至少2.1倍、至少2.2倍、至少2.3倍、至少2.4倍、至少2.5倍、至少2.6倍、至少2.7倍、至少2.8倍、至少2.9倍、至少3.0倍、至少3.1倍、至少3.2倍、至少3.3倍、至少3.4倍、至少3.5倍、至少3.6倍、至少3.7倍、至少3.8倍、至少3.9倍、至少4.0倍、至少4.1倍、至少4.2倍、至少4.3倍、至少4.4倍、至少4.5倍、至少4.6倍、至少4.7倍、至少4.8倍、至少4.9倍或至少5.0倍的上靶碱基编辑效率。

[1467] 可以经由质粒、载体、LNP复合物或mRNA将本文所述的ABE8碱基编辑器变体递送至宿主细胞。在一些实施方案中,将本文所述的任何ABE8碱基编辑器变体作为mRNA递送至宿主细胞。在一些实施方案中,如通过被编辑的核碱基测量的,经由基于核酸的递送系统(例如,mRNA)递送的ABE8碱基编辑器具有至少1%、至少2%、至少3%、至少4%、至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%或至少99%的上靶编辑效率。在一些实施方案中,与通过质粒或载体系统递送的ABE8碱基编辑器相比,通过mRNA系统递送的ABE8碱基编辑器具有更高的碱基编辑效率。在一些实施方案中,与通过质粒或载体系统递送时相比,本文所述的任何ABE8碱基编辑器变体在通过mRNA系统递送时具有高出至少1%、至少2%、至少3%、至少4%、至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少99%、至少100%、至少105%、至少110%、至少115%、至少120%、至少125%、至少130%、至少135%、至少140%、至少145%、至少150%、至少155%、至少160%、至少165%、至少170%、至少175%、至少180%、至少185%、至少190%、至少195%、至少200%、至少210%、至少220%、至少230%、至少240%、至少250%、至少260%、至少270%、至少280%、至少290%、至少300%、至少310%、至少320%、至少330%、至少340%、至少350%、至少360%、至少370%、至少380%、至少390%、至少400%、至少450%或至少500%的上靶编辑效率。在一些实施方案中,与通过质粒或载体系统递送时相比,本文所述的任何ABE8碱基编辑器变体在通过mRNA系统递送时具有高出至少1.1倍、至少1.2倍、至少1.3倍、至少1.4倍、至少1.5倍、至少1.6倍、至少1.7倍、至少1.8倍、至少1.9倍、至少2.0倍、至少2.1倍、至少2.2倍、至少2.3倍、至少2.4倍、至少2.5倍、至少2.6倍、至少2.7倍、至少2.8倍、至少2.9倍、至少3.0倍、至少3.1倍、至少3.2倍、至少3.3倍、至少3.4倍、至少3.5倍、至少3.6倍、至少3.7倍、至少3.8倍、至少3.9倍、至少4.0倍、至少4.1倍、至少4.2倍、至少4.3倍、至少4.4倍、至少4.5倍、至少4.6倍、至少4.7倍、至少4.8倍、至少4.9倍或至少5.0倍的上靶编辑效率。

[1468] 在一些实施方案中,包含本文所述的一种ABE8碱基编辑器变体的任何碱基编辑器系统导致在靶标多核苷酸序列中低于50%、低于40%、低于30%、低于20%、低于19%、低于18%、低于17%、低于16%、低于15%、低于14%、低于13%、低于12%、低于11%、低于10%、低于9%、低于8%、低于7%、低于6%、低于5%、低于4%、低于3%、低于2%、低于1%、低于0.9%、低于0.8%、低于0.7%、低于0.6%、低于0.5%、低于0.4%、低于0.3%、低于0.2%、低于0.1%、低于0.09%、低于0.08%、低于0.07%、低于0.06%、低于0.05%、低于0.04%、低于0.03%、低于0.02%或低于0.01%的脱靶编辑。

[1469] 在一些实施方案中,与通过质粒或载体系统递送时相比,本文所述的任何ABE8碱基编辑器变体在通过mRNA系统递送时具有较低的引导脱靶编辑效率。在一些实施方案中,

与通过质粒或载体系统递送时相比,本文所述的任何ABE8碱基编辑器变体在通过mRNA系统递送时具有低出至少1%、至少2%、至少3%、至少4%、至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%或至少99%的引导脱靶编辑效率。在一些实施方案中,与通过质粒或载体系统递送时相比,本文所述的任何ABE8碱基编辑器变体在通过mRNA系统递送时具有低出至少1.1倍、至少1.2倍、至少1.3倍、至少1.4倍、至少1.5倍、至少1.6倍、至少1.7倍、至少1.8倍、至少1.9倍、至少2.0倍、至少2.1倍、至少2.2倍、至少2.3倍、至少2.4倍、至少2.5倍、至少2.6倍、至少2.7倍、至少2.8倍、至少2.9倍或至少3.0倍的引导脱靶编辑效率。在一些实施方案中,与通过质粒或载体系统递送时相比,本文所述的任何ABE8碱基编辑器变体在通过mRNA系统递送时具有降低至少约2.2倍的引导脱靶编辑效率。

[1470] 在一些实施方案中,与通过质粒或载体系统递送时相比,本文所述的任何ABE8碱基编辑器变体在通过mRNA系统递送时具有较低的不依赖于向导的脱靶编辑效率。在一些实施方案中,与通过质粒或载体系统递送时相比,本文所述的任何ABE8碱基编辑器变体在通过mRNA系统递送时具有低出至少1%、至少2%、至少3%、至少4%、至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%或至少99%的不依赖于向导的脱靶编辑效率。在一些实施方案中,与通过质粒或载体系统递送时相比,本文所述的任何ABE8碱基编辑器变体在通过mRNA系统递送时具有低出至少1.1倍、至少1.2倍、至少1.3倍、至少1.4倍、至少1.5倍、至少1.6倍、至少1.7倍、至少1.8倍、至少1.9倍、至少2.0倍、至少2.1倍、至少2.2倍、至少2.3倍、至少2.4倍、至少2.5倍、至少2.6倍、至少2.7倍、至少2.8倍、至少2.9倍、至少3.0倍、至少5.0倍、至少10.0倍、至少20.0倍、至少50.0倍、至少70.0倍、至少100.0倍、至少120.0倍、至少130.0倍或至少150.0倍的不依赖于向导的脱靶编辑效率。在一些实施方案中,与通过质粒或载体系统递送时相比,本文所述的ABE8碱基编辑器变体在通过mRNA系统递送时具有降低134.0倍的不依赖于向导的脱靶编辑效率(例如,假RNA脱氨基)。在一些实施方案中,本文所述的ABE8碱基编辑器变体不增加不依赖于向导的跨基因组突变率。

[1471] 本公开的一些方面是基于以下认知的:本文提供的任何碱基编辑器都能够有效地在核酸(例如,受试者基因组中的核酸)中产生预期突变诸如点突变,而不产生显著数量的非预期突变,诸如非预期点突变。在一些实施方案中,本文提供的任何碱基编辑器都能够产生至少0.01%的预期突变(例如,伪脱靶编辑或旁观者编辑)。在一些实施方案中,预期突变是由结合至gRNA的特异性碱基编辑器产生的,该碱基编辑器被特异性地设计为改变或校正靶标基因中的突变。本公开的一些方面是基于以下认知的:本文提供的任何碱基编辑器都能够有效地在核酸(例如,受试者基因组中的核酸)中产生预期突变,而不产生显著数量的非预期突变。在一些实施方案中,预期突变是由结合至gRNA的特异性碱基编辑器产生的,该碱基编辑器被特异性地设计为改变或校正预期突变。在一些实施方案中,预期突变是在基因的编码区内产生终止密码子(例如,提前终止密码子)的突变。在一些实施方案中,预期突变是消除终止密码子的突变。在一些实施方案中,预期突变是改变基因剪接的突变。在一些实施方案中,预期突变是改变基因的调节序列(例如,基因启动子或基因抑制因子)的突变。

在一些实施方案中,本文提供的碱基编辑器能够产生大于1:1的预期点突变与插入缺失的比率。在一些实施方案中,本文提供的碱基编辑器能够产生至少1.5:1、至少2:1、至少2.5:1、至少3:1、至少3.5:1、至少4:1、至少4.5:1、至少5:1、至少5.5:1、至少6:1、至少6.5:1、至少7:1、至少7.5:1、至少8:1、至少8.5:1、至少9:1、至少10:1、至少11:1、至少12:1、至少13:1、至少14:1、至少15:1、至少20:1、至少25:1、至少30:1、至少40:1、至少50:1、至少100:1、至少200:1、至少300:1、至少400:1、至少500:1、至少600:1、至少700:1、至少800:1、至少900:1、或至少1000:1或更高的预期点突变与插入缺失的比率。

[1472] 可以使用任何合适的方法测定预期突变和插入缺失的数目,例如,如国际PCT申请号PCT/2017/045381(WO2018/027078)和PCT/US2016/058344(WO2017/070632);Komor,A.C.,et al.,"Programmable editing of atarget base in genomic DNA without double-stranded DNA cleavage"Nature 533,420-424(2016);Gaudelli,N.M.,et al.,"Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage"Nature 551,464-471(2017);以及Komor,A.C.,et al.,"Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity"Science Advances 3:eaao4774(2017)中所述,其整体内容通过引用并入本文。

[1473] 在一些实施方案中,为了计算插入缺失频率,对测序读取进行扫描,以获得与位于插入缺失可能发生的窗口两侧的两个10-bp序列的精确匹配。如果没有定位到精确匹配,则将该读取从分析中排除。如果这一插入缺失窗的长度与参考序列确切匹配,则将该读取归类为不含插入缺失。如果插入缺失窗比参考序列长或短两个或更多个碱基,则将该测序读取分别归类为插入或缺失。在一些实施方案中,本文提供的碱基编辑器可以限制插入缺失在核酸区域中的形成。在一些实施方案中,该区域位于碱基编辑器所靶向的核苷酸处,或者是位于碱基编辑器所靶向的核苷酸相距2、3、4、5、6、7、8、9或10个核苷酸内的区域。

[1474] 在靶标核苷酸区域处形成的插入缺失的数目可能取决于核酸(例如,细胞基因组内的核酸)被暴露于碱基编辑器的时间量。在一些实施方案中,在将靶标核苷酸序列(例如,细胞基因组内的核酸)暴露于碱基编辑器至少1小时、至少2小时、至少6小时、至少12小时、至少24小时、至少36小时、至少48小时、至少3天、至少4天、至少5天、至少7天、至少10天或至少14天之后,测定插入缺失的数目或比例。应知悉,本文所述的碱基编辑器的特征可以应用于本文提供的任何融合蛋白或使用该融合蛋白的方法。

[1475] 在一些实施方案中,本文提供的碱基编辑器能够限制插入缺失在核酸区域中的形成。在一些实施方案中,该区域位于碱基编辑器所靶向的核苷酸处,或者是位于碱基编辑器所靶向的核苷酸相距2、3、4、5、6、7、8、9或10个核苷酸内的区域。在一些实施方案中,本文提供的任何碱基编辑器能够将核酸区域处的插入缺失的形成限制为低于1%、低于1.5%、低于2%、低于2.5%、低于3%、低于3.5%、低于4%、低于4.5%、低于5%、低于6%、低于7%、低于8%、低于9%、低于10%、低于12%、低于15%或低于20%。在核苷酸区域处形成的插入缺失的数目可能取决于核酸(例如,细胞基因组内的核酸)被暴露于碱基编辑器的时间量。在一些实施方案中,在将核酸(例如,细胞基因组内的核酸)暴露于碱基编辑器至少1小时、至少2小时、至少6小时、至少12小时、至少24小时、至少36小时、至少48小时、至少3天、至少4天、至少5天、至少7天、至少10天或至少14天之后,测定插入缺失的任何数目或比例。

[1476] 多元编辑

[1477] 在一些实施方案中,本文所提供的碱基编辑器系统能够对一个或多个基因中的多个核碱基对进行多元编辑。在一些实施方案中,多个核碱基对定位在同一基因中。在一些实施方案中,多个核碱基对定位在一个或多个基因中,其中至少一个基因定位在不同的基因座中。在一些实施方案中,多元编辑可以包含一个或多个向导多核苷酸。在一些实施方案中,多元编辑可以包含一个或多个碱基编辑器系统。在一些实施方案中,多元编辑可以包含一个或多个碱基编辑器系统和单向导多核苷酸。在一些实施方案中,多元编辑可以包含一个或多个碱基编辑器系统和多个向导多核苷酸。在一些实施方案中,多元编辑可以包含一个或多个向导多核苷酸和单个碱基编辑器系统。在一些实施方案中,多元编辑可以包含至少一个向导多核苷酸,该向导多核苷酸不需要PAM序列来靶向结合至靶标多核苷酸序列。在一些实施方案中,多元编辑可以包含至少一个向导多核苷酸,该向导多核苷酸需要PAM序列来靶向结合至靶标多核苷酸序列。在一些实施方案中,多元编辑可以包含至少一个不需要PAM序列来靶向结合至靶标多核苷酸序列的向导多核苷酸与至少一个需要PAM序列来靶向结合至靶标多核苷酸序列的向导多核苷酸的混合物。应知悉,使用本文所述任何碱基编辑器的多元编辑的特征可以应用至使用本文所提供的任何碱基编辑器的方法的组合。也应知悉,使用本文所述任何碱基编辑器的多元编辑可以包含多个核碱基对的依序编辑。

[1478] 在一些实施方案中,多个核碱基对位于一个或多个基因中。在一些实施方案中,多个核碱基对位于同一基因中。在一些实施方案中,一个或多个基因中的至少一个基因定位在不同基因座中。

[1479] 在一些实施方案中,该编辑是编辑至少一个蛋白质编码区域内的多个核碱基。在一些实施方案中,该编辑是编辑至少一个蛋白质非编码区域内的多个核碱基。在一些实施方案中,该编辑是编辑至少一个蛋白质编码区域和至少一个蛋白质非编码区域内的多个核碱基。

[1480] 在一些实施方案中,该编辑与一个或多个向导多核苷酸协同作用。在一些实施方案中,碱基编辑器系统可以包含一个或多个碱基编辑器系统。在一些实施方案中,碱基编辑器系统可以包含一个或多个与单向导多核苷酸协同作用的碱基编辑器系统。在一些实施方案中,碱基编辑器系统可以包含一个或多个与多个向导多核苷酸协同作用的碱基编辑器系统。在一些实施方案中,该编辑与一个或多个向导多核苷酸和单个碱基编辑器系统协同作用。在一些实施方案中,该编辑与至少一个向导多核苷酸协同作用,该向导多核苷酸不需要PAM序列来靶向结合至靶标多核苷酸序列。在一些实施方案中,该编辑与至少一个向导多核苷酸协同作用,该向导多核苷酸需要PAM序列来靶向结合至靶标多核苷酸序列。在一些实施方案中,该编辑与至少一个不需要PAM序列来靶向结合至靶标多核苷酸序列的向导多核苷酸与至少一个需要PAM序列来靶向结合至靶标多核苷酸序列的向导多核苷酸的混合物协同作用。应知悉,使用本文所述任何碱基编辑器的多元编辑的特征可以应用至使用本文所提供的任何碱基编辑器的方法的组合。也应知悉,该编辑可以包含多个核碱基对的依序编辑。

[1481] 使用碱基编辑器的方法

[1482] 编辑HBV基因组为治疗和基础研究提供了新策略。

[1483] 本公开提供用于治疗被诊断患有HBV感染的受试者的方法。例如,在一些实施方案中,提供了包括下述的方法:向患有此类疾病(例如,HBV感染所致的疾病)的受试者给药有

效量的核碱基编辑器(例如,腺苷脱氨酶碱基编辑器或胞苷脱氨酶碱基编辑器),该核碱基编辑器改变HBV基因组中的核碱基。

[1484] 在某些方面,提供了用于治疗HBV感染的方法,该HBV感染可通过对HBV基因中的至少一种进行脱氨酶介导的基因编辑来治疗。

[1485] 应理解,各序列(例如,分别为疾病相关基因或其编码的蛋白质的多核苷酸或氨基酸序列)中具体位置或残基的编号取决于所使用的特定蛋白质和编号方案。编号在例如成熟蛋白的前体和成熟蛋白本身中可以是不同的,并且来自不同物种的序列之间的差异可能影响编号。本领域技术人员将能够通过本领域周知的方法鉴定任何同源蛋白质和相应编码核酸中的各残基,例如,通过序列比对和同源残基的确定。

[1486] 本文提供了使用碱基编辑器或碱基编辑器系统对HBV基因组的靶标核苷酸序列中的核碱基进行编辑的方法。在一些实施方案中,碱基编辑器(例如,包含腺苷脱氨酶和Cas9结构域)的活性导致HBV基因组中的核碱基的改变。在一些实施方案中,靶标DNA序列改变以包括G→A点突变,并且其中A碱基的脱氨基导致了降低或消除HBV功能的序列。在一些实施方案中,靶标DNA序列改变以包括T→C点突变,并且其中突变的C碱基的脱氨基导致了降低或消除HBV功能的序列。

[1487] 在一些实施方案中,靶标DNA序列编码蛋白质(例如,HBV蛋白诸如聚合酶或表面抗原),并且该改变导致由野生型密码子编码的氨基酸产生变化。在一些实施方案中,A核苷酸的脱氨基作用导致由野生型密码子编码的氨基酸中的变化。在一些实施方案中,野生型A的脱氨基作用导致编码突变氨基酸的密码子。在一些实施方案中,野生型C的脱氨基作用导致由野生型密码子编码的氨基酸产生变化。在一些实施方案中,野生型C的脱氨基作用导致编码突变氨基酸的密码子。在一些实施方案中,受试者患有或已经被诊断患有HBV感染。

[1488] 在一些实施方案中,本文所提供的腺苷脱氨酶能够将DNA的脱氧腺苷残基的腺嘌呤脱氨基。本公开的其他方面提供融合蛋白,该融合蛋白包含腺苷脱氨酶(例如,如本文所述将DNA中的脱氧腺苷脱氨基的腺苷脱氨酶)和能够结合至特异性核苷酸序列的结构域(例如,Cas9或Cpf1蛋白)。例如,腺苷可以被转化为肌苷残基,而肌苷残基通常与胞嘧啶残基配对。除其他之外,此类融合蛋白可用于核酸序列的靶向编辑。此类融合蛋白可用于HBV DNA的体外靶向编辑;用于引入靶向突变,例如,用于细胞中的HBV基因组的离体改变;以及用于引入体外靶向突变,例如,感染the alteration of the HBV的受试者体内细胞中的HBV基因组的改变。本公开提供利用脱氨酶和核碱基编辑器的脱氨酶、融合蛋白、核酸、载体、细胞、组合物、方法、试剂盒、系统等。

[1489] 产生预期突变

[1490] 本文提供的核碱基编辑蛋白可以被验证用于基于基因编辑的体外人类疗法,例如,通过将降低或消除感染或其症候的改变引入HBV基因组中。技术人员将会理解,本文提供的核碱基编辑蛋白,例如,包含多核苷酸可编程核苷酸结合结构域(例如,Cas9)和核碱基编辑结构域(例如,腺苷脱氨酶结构域或胞苷脱氨酶结构域)的融合蛋白,可以用来将A改变为G或将C改变为T。在第一种情况下,A脱氨基化为I改变了HBV基因组序列,而在后一种情况下,与HBV基因组中的T进行碱基配对的A的脱氨基作用以及随后的一轮复制引入了该突变。

[1491] 在一些实施方案中,本文公开提供碱基编辑器,其可以在受试者基因组中的核酸(例如,受试者基因组内的核酸)中有效地产生“预期突变”,而不产生显著数量的非预期突

变诸如非预期点突变。在一些实施方案中,预期突变是通过结合至向导多核苷酸(例如,gRNA)的特异性碱基编辑器(例如,胞苷碱基编辑器或腺苷碱基编辑器)产生的突变,该特异性碱基编辑器经过特异性地涉及以产生预期突变。在一些实施方案中,预期突变是腺嘌呤(A)至鸟嘌呤(G)的点突变。在一些实施方案中,预期突变是胞嘧啶(C)至胸腺嘧啶(T)的点突变。在一些实施方案中,预期突变是基因的编码区或非编码区内的腺嘌呤(A)至鸟嘌呤(G)的点突变。在一些实施方案中,预期突变是基因的编码区或非编码区内的胞嘧啶(C)至胸腺嘧啶(T)的点突变。在一些实施方案中,预期突变是点突变,该点突变在基因的编码区内产生终止密码子,例如,提前终止密码子。在一些实施方案中,预期突变是消除终止密码子的突变。

[1492] 在一些实施方案中,本文提供的碱基编辑器能够产生大于1:1的预期突变与非预期突变的比率(例如,预期点突变:非预期点突变)。在一些实施方案中,本文提供的碱基编辑器能够产生至少1.5:1、至少2:1、至少2.5:1、至少3:1、至少3.5:1、至少4:1、至少4.5:1、至少5:1、至少5.5:1、至少6:1、至少6.5:1、至少7:1、至少7.5:1、至少8:1、至少10:1、至少12:1、至少15:1、至少20:1、至少25:1、至少30:1、至少40:1、至少50:1、至少100:1、至少150:1、至少200:1、至少250:1、至少500:1、或至少1000:1或更高的预期突变与非预期突变的比率(例如,预期点突变:非预期点突变)。

[1493] 碱基编辑器效率的细节在国际PCT申请号PCT/2017/045381(WO2018/027078)和PCT/US2016/058344(WO2017/070632)中描述,其各自通过引用而整体并入本文。也参见,Komor,A.C.,et al.,"Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage"Nature 533,420-424(2016);Gaudelli,N.M.,et al.,"Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage"Nature 551,464-471(2017)和Komor,A.C.,et al.,"Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity"Science Advances 3:eaao4774(2017)),其整体内容通过引用并入本文。

[1494] 在一些实施方案中,编辑一个或多个基因中的多个核碱基对导致了至少一个预期突变的形成。在一些实施方案中,至少一个预期突变的形成导致了HBV基因组的改变。应知悉,本文所述的碱基编辑器的多元编辑的特征可以应用至使用本文所提供的碱基编辑器的方法的组合。

[1495] HBV基因的修饰

[1496] 在一些实施方案中,HBV基因的精确修饰降低病毒的毒力并且/或降低病毒在体外传播的能力。修饰可以是提前终止密码子或者是修复或以其他方式削弱HBV蛋白活性的突变。在一些实施方案中,被靶向以进行修饰的HBV基因是前核心区基因的pol、X、S、pre-S1、pre-S2或核心区,或其组合。

[1497] 融合蛋白在宿主细胞内的表达

[1498] 本公开的包含腺苷脱氨酶变体的融合蛋白可以使用技术人员已知的常规方法在包括但不限于,细菌、酵母、真菌、昆虫、植物和动物细胞在内的几乎任何感兴趣的宿主细胞内表达。例如,编码本公开的腺苷脱氨酶的DNA可以通过基于cDNA序列为CDS的上游或下游设计合适的引物而被克隆。可以将克隆的DNA直接地、或在当需要时用限制酶消化后、或在

添加合适的连接子和/或核定位信号之后,用编码碱基编辑系统的一个或多个额外组分的DNA连接。碱基编辑系统在宿主细胞中被翻译以形成复合物。

[1499] 编码本文所述的蛋白结构域的DNA可以通过化学合成该DNA而获得,或通过利用PCR方法和Gibson组装方法来联接合成的部分重叠oligoDNA短链而获得,以构建编码其全长度的DNA。通过化学合成或PCR方法或Gibson组装方法来构建全长度DNA的优势在于,可以根据该DNA所引入至其中的宿主而以CDS全长度设计待使用的密码子。在异源DNA的表达中,预计蛋白质表达水平通过将其DNA序列转化为以高频率用于宿主生物体中的码码字而得以增加。作为密码子在待用宿主中的使用频率的数据,例如,可使用Kazusa DNA研究所主页中公开的遗传密码使用频率数据库(http://www.kazusa.or.jp/codon/index.html),或者可以指显示密码子在每种宿主中使用频率的文献。参照所获得的数据和待引入的DNA序列,可以将在用于DNA序列的密码子中显示在该宿主中的低使用频率的密码子转化为编码相同氨基酸但显示高使用频率的密码子。

[1500] 例如,可以通过在合适的表达载体中将DNA连接至启动子的下游,产生含有编码核酸序列识别模块和/或核酸碱基转化酶的表达载体。

[1501] 作为表达载体,使用了源自大肠杆菌的质粒(例如,pBR322、pBR325、pUC12、pUC13);源自枯草芽孢杆菌的质粒(例如,pUB110、pTP5、pC194);源自酵母的质粒(例如,pSH19、pSH15);昆虫细胞表达质粒(例如,pFast-Bac);动物细胞表达质粒(例如,pA1-11、pXT1、pRc/CMV、pRc/RSV、pcDNAI/Neo);细菌噬菌体,诸如λ噬菌体等;昆虫病毒载体,诸如杆状病毒等(例如,BmNPV、AcNPV);动物病毒载体,诸如逆转录病毒、牛痘病毒、腺病毒等。

[1502] 在一些实施方案中,可以使用任何适用于待用于基因表达的宿主的任何启动子。在使用DSB的传统方法中,由于宿主细胞的存活率有时由于毒性而显著下降,希望通过使用诱导型启动子来启动诱导而增加细胞的数目。但是,由于也可以通过表达本公开的核酸修饰酶复合物而获得足够的细胞增殖,因此也可以无限制地使用构造型启动子。

[1503] 例如,当宿主为动物细胞时,使用了SRα启动子、SV40启动子、LTR启动子、CMV(巨细胞病毒)启动子、RSV(鲁斯氏肉瘤病毒)启动子、MoMuLV(莫洛尼鼠白血病病毒)LTR、HSV-TK(单纯疱疹病毒胸苷激酶)启动子等。其中,CMV启动子、SRα启动子等为优选。

[1504] 当宿主为大肠杆菌时,trp启动子、lac启动子、recA启动子、λP_L启动子、lpp启动子、T7启动子等为优选。

[1505] 当宿主为芽孢杆菌属时,SPO1启动子、SPO2启动子、penP启动子等为优选。

[1506] 当宿主为酵母时,Gal1/10启动子、PHO5启动子、PGK启动子、GAP启动子、ADH启动子等为优选。

[1507] 当宿主为昆虫细胞时,多角体蛋白启动子、P10启动子等为优选。

[1508] 当宿主为植物细胞时,CaMV35S启动子、CaMV19S启动子、NOS启动子等为优选。

[1509] 作为表达载体,除了上述那些之外,可使用根据需要而含有增强子、剪接信号、终止子、polyA加入信号、选择标记物(诸如耐药基因、营养缺陷型互补基因等)、复制起点等的表达载体。

[1510] 编码本文所述蛋白结构域的RNA可以通过下述制备,例如,使用编码DNA(该DNA编码上述核酸序列识别模块和/或核碱基转化酶)的载体作为模板,在本身已知的体外转录系统)中转录为mRNA。

[1511] 通过将含有编码核酸识别模块和/或核酸碱基转化酶的表达载体引入宿主细胞中,并且培养该宿主细胞,可以细胞内地表达本公开的融合蛋白。

[1512] 作为宿主,使用了埃希氏杆菌属、芽孢杆菌属、酵母、昆虫细胞、昆虫、动物细胞等。

[1513] 作为埃希氏杆菌属,使用了大肠杆菌K12.cndot.DH1(Proc.Natl.Acad.Sci.USA,60,160(1968))、大肠杆菌JM103(Nucleic Acids Research,9,309(1981))、大肠杆菌JA221(Journal of Molecular Biology,120,517(1978))、大肠杆菌HB101(Journal of Molecular Biology,41,459(1969))、大肠杆菌C600(Genetics,39,440(1954))等。

[1514] 作为芽孢杆菌属,使用了枯草芽孢杆菌M1114(Gene,24,255(1983))、枯草芽孢杆菌207-21(Journal of Biochemistry,95,87(1984))等。

[1515] 作为酵母,使用了酿酒酵母AH22、AH22R$^-$、NA87-11A、DKD-5D,20B-12,粟酒裂殖酵母NCYC1913、NCYC2036,毕赤酵母KM71等。

[1516] 作为昆虫细胞,当病毒是AcNPV时,使用了卷心菜粘虫幼虫衍生的建立系的细胞(草地贪夜蛾(Spodoptera frugiperda)细胞;Sf细胞)、衍生自粉纹夜蛾(Trichoplusia ni)中肠的MG1细胞、衍生自粉纹夜蛾卵的High Five^{TM}细胞、衍生自甘蓝夜蛾(Mamestra brassicae)的细胞、衍生自甲虫(Estigmena acrea)的细胞等。当病毒为BmNPV时,使用家蚕(Bombyx mori)衍生的建立系细胞(家蚕N细胞;BmN细胞)等作为昆虫细胞。作为Sf细胞,例如,使用了Sf9细胞(ATCC CRL1711)、Sf21细胞(上述全部,In Vivo,13,213-217(1977))等。

[1517] 作为昆虫,例如,使用了家蚕幼虫、果蝇(Drosophila)、蟋蟀等(Nature,315,592(1985))。

[1518] 作为动物细胞,使用了细胞系诸如猴COS-7细胞、猴Vero细胞、中国仓鼠卵巢(CHO)细胞、dhfr基因缺陷CHO细胞、小鼠L细胞、小鼠AtT-20细胞、小鼠骨髓瘤细胞、大鼠GH3细胞、人FL细胞等,多能干细胞诸如人或其他哺乳动物的iPS细胞等,以及从各种组织制备的原代培养细胞。此外,也可以使用斑马鱼胚胎、爪蟾(Xenopus)卵细胞等。

[1519] 作为植物细胞,使用了从各种植物(例如,谷物诸如稻、小麦、玉米等,产品作物诸如西红柿、黄瓜、茄子等,花园植物诸如康乃馨、洋桔梗等,实验植物诸如烟草、拟南芥(arabidopsis thaliana)等)制备的悬浮培养的细胞、愈伤组织、原生质体、叶段、根段等。

[1520] 上述宿主细胞全部可以是单倍体(一倍体)或多倍体(例如,二倍体、三倍体、四倍体等)。在传统突变引入方法中,原则上仅将突变引入一个同源染色体中以产生异质基因类型。因此,除非出现显性突变,否则不会表达所希望的表型,并且纯合性不方便地耗时耗力。相比之下,根据本公开,由于可以将突变引入基因组中同源染色体上的任何等位基因中,即便在隐性突变的情况下也可以在一代中表达所希望的表型,这由于可以解决传统方法的问题而极其有用。

[1521] 表达载体可以根据宿主的种类通过已知方法(例如,溶菌酶法、感受态法、PEG法、$CaCl_2$共沉淀法、电穿孔法、显微注射法、粒子枪法、脂质转染法、农杆菌属(Agrobacterium)法等)引入。

[1522] 可以根据例如Proc.Natl.Acad.Sci.USA,69,2110(1972),Gene,17,107(1982)等中描述的方法转化大肠杆菌。

[1523] 可以根据例如Molecular&General Genetics,168,111(1979)等中描述的方法将芽孢杆菌属引入载体中。

[1524] 可以根据例如Methods in Enzymology,194,182-187(1991),Proc.Natl.Acad.Sci.USA,75,1929(1978)等中描述的方法将酵母引入载体中。

[1525] 可以根据例如Bio/Technology,6,47-55(1988)等中描述的方法将昆虫细胞和昆虫引入载体中。

[1526] 可以根据例如Cell Engineering additional volume 8,New Cell Engineering Experiment Protocol,263-267(1995)(秀润社(Shujunsha)出版)和Virology,52,456(1973)中描述的方法将动物细胞引入载体中。

[1527] 使用载体引入的细胞可以是根据已知方法根据宿主的种类培养的。

[1528] 例如,当培养大肠杆菌或芽孢杆菌属时,优选液体培养基作为待用于培养的培养基。培养基优选含有转化株生长所必需的碳源、氮源、无机物质等。碳源的示例包括葡萄糖、糊精、可溶淀粉、蔗糖等;氮源的示例包括无机或有机物质,诸如铵盐、硝酸盐、玉米浆、蛋白胨、酪蛋白、肉膏、豆粕、马铃薯提取物等;并且无机物质的示例包括氯化钙、磷酸二氢钠、氯化镁等。培养基可以含有酵母提取物、微生物、生长促进因子等。培养基的pH优选为约5至约8。

[1529] 作为用于培养大肠杆菌的培养基,例如,优选为含有葡萄糖、酪蛋白氨基酸的M9培养基(Journal of Experiments in Molecular Genetics,431-433,Cold Spring Harbor Laboratory,New York 1972)。若必要,例如,可以将诸如3β-吲哚基丙烯酸的试剂添加到培养基中以确保启动子的有效功能。大肠杆菌通常在约15℃至约43℃培养。若必要,可以执行曝气和搅拌。

[1530] 芽孢杆菌属通常在约30℃至约40℃培养。若必要,可以执行曝气和搅拌。

[1531] 用于培养酵母的培养基的示例包括Burkholder基础培养基(Proc.Natl.Acad.Sci.USA,77,4505(1980))、含有0.5%酪蛋白氨基酸的SD培养基(Proc.Natl.Acad.Sci.USA,81,5330(1984))等。培养基的pH优选为约5至约8。培养通常在约20℃至约35℃执行。若必要,可以执行曝气和搅拌。

[1532] 作为用于培养昆虫细胞或昆虫的培养基,例如,使用了适宜地含有诸如灭活的10%牛血清等添加剂的Grace昆虫培养基(Nature,195,788(1962))等。培养基的pH优选为约6.2至约6.4。培养通常在约27℃执行。若必要,可以执行曝气和搅拌。

[1533] 作为用于培养动物细胞的培养基,使用了含有约5%至约20%胎牛血清的最小必需培养基(MEM)(Science,122,501(1952))、杜尔贝科改进伊格尔培养基(DMEM)(Virology,8,396(1959))、RPMI 1640培养基(The Journal of the American Medical Association,199,519(1967))、199培养基(Proceeding of the Society for the Biological Medicine,73,1(1950))等。培养基的pH优选为约6至约8。培养通常在约30℃至约40℃执行。若必要,可以执行曝气和搅拌。

[1534] 作为用于培养植物细胞的培养基,例如,使用了MS培养基、LS培养基、B5培养基等。培养基的pH优选为约5至约8。培养通常在约20℃至约30℃执行。若必要,可以执行曝气和搅拌。

[1535] 当使用诸如动物细胞、昆虫细胞、植物细胞等高等真核细胞作为宿主细胞时,在诱导型启动子(例如,金属硫蛋白启动子(通过重金属离子诱导)、热休克蛋白启动子(通过热休克诱导)、Tet-ON/Tet-OFF系统启动子(通过添加或移除四环素或其衍生物诱导)、类固醇

应答型启动子(通过类固醇激素或其衍生物诱导)等)的调节下将编码本公开的碱基编辑系统(例如,包含腺苷脱氨酶变体)的DNA引入宿主细胞内,在适宜的阶段将诱导物质添加到培养基中(或自培养基中移除)以诱导核酸修饰酶复合物的表达,执行给定时间段的培养以完成碱基编辑并且将突变引入引入靶标基因中,可以实现碱基编辑系统的瞬时表达。

[1536] 原核细胞诸如大肠杆菌等可以利用诱导型启动子。诱导型启动子的示例包括但不限于,lac启动子(通过IPTG诱导)、cspA启动子(通过冷休克诱导)、araBAD启动子(通过阿拉伯糖诱导)等。

[1537] 作为另一种选择,当使用诸如动物细胞、昆虫细胞、植物细胞等高等真核细胞作为宿主细胞时,也可利用上述诱导型启动子作为载体移除机制。换言之,载体封固有在宿主细胞中发挥作用的复制起点和编码复制所必需的蛋白质的核酸(例如,对于动物细胞,SV40和大T抗原、oriP和EBNA-1等),并且通过上述诱导型启动子调节编码该蛋白质的核酸的表达。因此,尽管载体在存在诱导物质的情况下可以自主复制,但当诱导物质被移除时无法进行自主复制,并且载体随着细胞分裂自然地脱落(不可能通过将四环素和强力霉素添加到Tet-OFF系统载体中实现自主复制)。

[1538] 递送系统

[1539] 本文公开的碱基编辑器可以被编码在病毒载体中含有的核酸上。病毒载体可包括慢病毒、腺病毒、逆转录病毒和腺相关病毒(AAV)。可基于应用而选择病毒载体。例如,AAV由于它们的轻微免疫原性而常常用于体内基因递送。腺病毒因为其诱导强烈的免疫原性应答而常常用作疫苗。病毒载体的封装能力可能限制可以被封装到载体中的碱基编辑器的尺寸。例如,AAV的封装能力为约4.5Kb,包括两个145个碱基的末端反向重复序列(ITR)。

[1540] AAV是小的、单链DNA依赖性病毒,属于细小病毒家族。4.7kb野生型(wt)AAV基因组由分别编码四种复制蛋白和三种衣壳蛋白的两个基因组成,并且两侧各具有145-bp的末端反向重复序列(ITR)。病毒体由三种衣壳蛋白(Vp1、Vp2和Vp3)构成,以1:1:10的比率产生自同一个开读框但来自差异剪接位点(Vp1)和替代性翻译起始位点(分别为Vp2和Vp3)。Vp3是病毒体中丰度最高的亚基,并且参与细胞表面处的定义病毒向性的受体复制。已经在Vp1的独特N端中鉴定了在病毒感染性中起作用的磷脂酶结构域。

[1541] 类似于wt AAV,重组AAV(rAAV)利用顺式作用145-bp ITR与病毒转基因盒两侧相接,提供高达4.5kb进行外来DNA的封装。感染会后,rAAV可以表达本发明的融合蛋白,并且通过附加基因地存在于环状头至尾串联体中而持续存在,不被整合到宿主基因组中。尽管存在使用这一系统的体外和体内rAAV成功的大量示例,但当基因编码序列的长度在尺寸上大于或等于wt AAV基因组时,有限的封装能力已经限制了AAV介导的基因递送的用途。

[1542] AAV载体的低封装能力使得大量超过这一尺寸的基因的递送和/或大的生理学调节元件的使用颇具挑战性。这些挑战可以例如通过将待递送的蛋白质分为两个或更多个片段而得以解决,其中该N端片段融合至分裂内含肽-N,并且C端片段融合至分裂内含肽-C。然后将这些片段封装到两个或更多个AAV载体中。如本文所用,“内含肽”是指自剪接蛋白内含子(例如,肽),其连接位于两侧的N端和C端外显肽(例如,待接合的片段)。某些内含肽用于接合异源蛋白片段的用途在例如Wood et al.,J.Biol.Chem.289(21);14512-9(2014)中有所描述。例如,当融合至独立的蛋白片段时,内含肽IntN和IntC彼此识别,将其自身剪出并同步连接位于两侧的它们所融合的蛋白片段的N端和C端外显肽,从而从两个蛋白片段重构

全长度蛋白质。其他合适的内含肽对于本领域技术人员将是显而易见的。

[1543] 本发明融合蛋白的片段的长度可变。在一些实施方案中,蛋白片段的长度为2个氨基酸至约1000个氨基酸的范围。在一些实施方案中,蛋白片段的长度为约5个氨基酸至约500个氨基酸的范围。在一些实施方案中,蛋白片段的长度为约20个氨基酸至约200个氨基酸的范围。在一些实施方案中,蛋白片段的长度为约10个氨基酸至约100个氨基酸的范围。其他长度的合适的蛋白片段对于本领域技术人员将会是显而易见的。

[1544] 在一些实施方案中,将核酸酶(例如,Cas9)的一部分或片段融合至内含肽。核酸酶可以融合至内含肽的N端或C端。在一些实施方案中,融合蛋白的一部分或片段融合至内含肽并且融合至AAV衣壳蛋白。内含肽、核酸酶和衣壳蛋白可以任何排列(例如,核酸酶-内含肽-衣壳、内含肽-核酸酶-衣壳、衣壳-内含肽-核酸酶等)融合在一起。在一些实施方案中,内含肽的N端融合至融合蛋白的C端,并且内含肽的C端融合至AAV衣壳蛋白的N端。

[1545] 在一种实施方案中,通过将大的转基因盒剪接为独立的两半(5'和3'末端,或头部和尾部)而产生双AAV载体,其中该盒的每一半都被封装到单个AAV载体(尺寸<5kb)中。然后,在通过两个双AAV载体对同一细胞进行共转染后,遵循下列实现全长度转基因表达盒的重新组装:(1)5'基因组与3'基因组(双AAV重叠载体)之间的同源重组(HR);(2)5'基因组和3'基因组(双AAV反式剪接载体)的ITR介导的尾至头连环化;或(3)这两种机制的组合(双AAV杂交载体)。在体内使用双AAV载体导致全长度蛋白质的表达。双AAV载体平台的使用为尺寸>4.7kb的转基因提供了有效且可行的基因转移策略。

[1546] 所公开的用于设计碱基编辑器的策略可用于生成能够被封装在病毒载体内的碱基编辑器。基于RNA或DNA病毒的系统的用于递送碱基编辑器的用途利用了高度进化的过程的优势,这些过程用于将病毒靶向至培养物中或宿主中的特定细胞以及将病毒有效载荷转运到细胞核或宿主细胞基因组。病毒载体可以直接给药至培养物中的细胞、患者(体内),或者它们可以用来在体外处理细胞,并且经修饰的细胞可以任选地被给药至患者(离体)。传统的基于病毒的系统可包括用于基因转移的逆转录病毒、慢病毒、腺病毒、腺相关病毒和单纯疱疹病毒。使用逆转录病毒、慢病毒和腺相关病毒基因转移方法整合到宿主基因组中是可能的,经常导致感兴趣的转基因的长期表达。此外,已经在很多不同的细胞类型和靶标组织中观察到了高转导效率。

[1547] 可以通过并入外来包膜蛋白改变逆转录病毒的向性,拓展靶标细胞的潜在靶标群。慢病毒载体是逆转录病毒载体,其能够转导或感染非分裂细胞并且通常产生高病毒滴度。逆转录病毒基因转移系统的选择将会因此取决于靶标组织。逆转录病毒载体由顺式作用长末端重复序列构成,其封装能力高达6-10kb的外来序列。最小顺式作用LTR对于载体的复制和封装而言是足够的,其随后被用来将治疗性基因整合到靶标细胞内,以提供永久性转基因表达。广泛使用的逆转录病毒载体包括那些基于以下的载体:鼠白血病病毒(MuLV)、长臂猿白血病病毒(GaLV)、猿猴免疫缺陷病毒(SIV)、人免疫缺陷病毒(HIV)及其组合(参见例如,Buchscher et al.,J.Virol.66:2731-2739(1992);Johann et al.,J.Virol.66:1635-1640(1992);Sommnerfelt et al.,Virol.176:58-59(1990);Wilson et al.,J.Virol.63:2374-2378(1989);Miller et al.,J.Virol.65:2220-2224(1991);PCT/US94/05700)。

[1548] 逆转录病毒载体,尤其是慢病毒载体,可能需要小于给定长度的多核苷酸序列,以

有效地整合到靶标细胞中。例如,与尺寸较小的逆转录病毒载体相比,长度大于9kb的逆转录病毒载体可能导致低病毒滴度。在一些方面,本公开的碱基编辑器的尺寸是足够的,从而能够进行有效封装并经由逆转录病毒载体被递送到靶标细胞内。在一些情况下,即便当与向导核酸和/或可靶向的核酸酶系统的其他组分一起被表达时,碱基编辑器的尺寸也允许进行有效封装和递送。

[1549] 在优选瞬时表达的应用中,可以使用基于腺病毒的系统。基于腺病毒的系统能够在很多细胞类型中表现出非常高的转导效率并且不需要细胞分裂。使用此类载体,已经获得了高滴度和表达水平。这一载体可以在相对简单的系统中大量生产。腺相关病毒("AAV")载体也可以用来使用靶标核酸转染细胞,例如在体外生产核酸和肽,以及进行体内和离体的基因治疗过程(参见例如,West et al.,Virology 160:38-47(1987);美国专利号4,797,368;WO 93/24641;Kotin,Human Gene Therapy 5:793-801(1994);Muzyczka,J.Clin.Invest.94:1351(1994))。重组AAV载体的构建在大量出版物中有所描述,包括美国专利号5,173,414;Tratschin et al.,Mol.Cell.Biol.5:3251-3260(1985);Tratschin,et al.,Mol.Cell.Biol.4:2072-2081(1984);Hermonat&Muzyczka,PNAS 81:6466-6470(1984);和Samulski et al.,J.Virol.63:03822-3828(1989)。

[1550] 本文所述的碱基编辑器因此可以使用病毒载体递送。碱基编辑器系统的一个或多个组分可以被编码在一种或多种病毒载体上。例如,碱基编辑器和向导核酸可以被编码在单个病毒载体上。在其他情况下,碱基编辑器和向导核酸被编码在不同的病毒载体上。在每种情况下,碱基编辑器和向导核酸可以各自可操作地连接至启动子和终止子。

[1551] 被编码在病毒载体上的组分的组合可以通过所选病毒载体的载荷尺寸约束条件确定。

[1552] 碱基编辑器的非病毒递送

[1553] 也可使用碱基编辑器的非病毒递送途径。非病毒核酸载体的一个重要类别是纳米粒子,其可以是有机的或无机的。纳米粒子是本领域周知的。任何合适的纳米粒子设计可以用来递送基因组编辑系统组分或编码此类组分的核酸。例如,在本公开的某些实施方案中,有机(例如,脂质和/或聚合物)纳米粒子可能适合用作递送媒介。用于纳米粒子制剂和/或基因转移的示例性脂质显示在(下)表11中。

[1554] 表11

基因转移中使用的脂质

脂质	缩写	特征
1,2-二油酰基-sn-甘油-3-磷脂酰胆碱	DOPC	辅助
1,2-二油酰基-sn-甘油-3-磷脂酰乙醇胺	DOPE	辅助
胆固醇		辅助
N-[1-(2,3-二油酰氧基)丙基]N,N,N-三甲基氯化铵	DOTMA	阳离子
1,2-二油酰氧基-3-三甲基铵-丙烷	DOTAP	阳离子
双十八烷基酰胺基氨基乙酰精胺	DOGS	阳离子
N-(3-氨基丙基)-N,N-二甲基-2,3-双(十二烷氧基)-1-丙铵溴化物	GAP-DLRIE	阳离子
鲸蜡基三甲基溴化铵	CTAB	阳离子
[1555] 鸟氨酸 6-月桂酰己酯	LHON	阳离子
1-(2,3-二油酰氧基丙基)-2,4,6-三甲基吡啶鎓	2Oc	阳离子
2,3-二油酰氧基-N-[2-(精胺甲酰胺基)-乙基]-N,N-二甲基-1-丙铵三氟乙酸盐	DOSPA	阳离子
1,2-二油烯基-3-三甲基铵-丙烷	DOPA	阳离子
N-(2-羟基乙基)-N,N-二甲基-2,3-双(十四烷氧基)-1-丙铵溴化物	MDRIE	阳离子
(二肉豆蔻氧基丙基)二甲基羟乙基溴化铵	DMRI	阳离子
3β-[N-(N',N'-二甲基氨基乙烷)-氨基甲酰基]胆固醇	DC-Chol	阳离子
双-胍-二乙三胺-胆固醇(Bis-guanidium-tren-cholesterol)	BGTC	阳离子
1,3-二油酰氧基-2-(6-羧基-精胺酰基)-丙基酰胺	DOSPER	阳离子

[1556]

基因转移中使用的脂质		
脂质	缩写	特征
二甲基十八烷基溴化铵	DDAB	阳离子
双十八烷基酰胺基甘氨酰亚精胺	DSL	阳离子
外消旋-[(2,3-双十八烷氧基丙基)(2-羟基乙基)]二甲基氯化铵	CLIP-1	阳离子
外消旋-[2(2,3-双十六烷氧基丙氧基甲氧基)乙基]三甲基溴化铵	CLIP-6	阳离子
乙基二肉豆蔻酰磷脂酰胆碱	EDMPC	阳离子
1,2-二硬脂酰氧基-N,N-二甲基-3-氨基丙烷	DSDMA	阳离子
1,2-二肉豆蔻酰基-三甲基铵丙烷	DMTAP	阳离子
0,0'-二肉豆蔻基-N-赖氨酰基天冬氨酸酯	DMKE	阳离子
1,2-二硬脂酰基-sn-甘油-3-乙基磷酰胆碱	DSEPC	阳离子
N-棕榈酰-D-赤型-鞘氨苷-氨基甲酰基-精胺	CCS	阳离子
N-叔丁基-N0-十四烷基-3-十四烷基氨基丙脒	diC14-amidine	阳离子
十八烯酰氧基[乙基-2-十七烯基-3-羟基乙基]咪唑鎓氯化物	DOTIM	阳离子
N1-胆固醇氧羰基-3,7-二氮杂壬烷-1,9-二胺	CDAN	阳离子
2-(3-[双(3-氨基-丙基)-氨基]丙基氨基)-N双十四烷基氨基甲酰基甲基-乙酰胺	RPR209120	阳离子
1,2-二亚油酰氧基-3-二甲基氨基丙烷	DLinDMA	阳离子
2,2-二亚油酰基-4-二甲基氨基乙基-[1,3]-二氧杂环戊烷基	DLin-KC2-DMA	阳离子
二亚油酰基-甲基-4-二甲基氨基丁酸酯	DLin-MC3-DMA	阳离子

[1557] 表12列出了用于基因转移和/或纳米粒子制剂的示例性聚合物。

[1558] 表12

[1559]

用于基因转移的聚合物	
聚合物	缩写
聚乙二醇	PEG
聚乙烯亚胺	PEI
二硫代双(琥珀酰亚胺基丙酸酯)	DSP
3,3'-二硫代双丙亚氨酸二甲酯	DTBP
聚乙烯亚胺双氨基甲酸酯	PEIC
聚(L-赖氨酸)	PLL
组氨酸修饰的 PLL	
聚(N-乙烯基吡咯烷酮)	PVP
聚丙烯亚胺	PPI
聚(酰胺基胺)	PAMAM
聚(酰胺基乙烯亚胺)	SS-PAEI
三乙四胺	TETA
聚(β-氨基酯)	
聚(4-羟基-L-脯氨酸酯)	PHP
聚(烯丙基胺)	
聚(α-[4-氨基丁基]-L-乙醇酸)	PAGA
聚(D,L-乳酸-共-乙醇酸)	PLGA
聚(N-乙基-4-乙烯基吡啶鎓溴化物)	
聚磷腈	PPZ
聚磷酸酯	PPE
聚(氨基磷酸酯)	PPA
聚(N-(2-羟基丙基)甲基丙烯酰胺)	pHPMA
聚(甲基丙烯酸 2-(二甲基氨基)乙酯)	pDMAEMA
聚(丙烯磷酸 2-氨基乙酯)	PPE-EA
壳聚糖	
半乳糖基化壳聚糖	
N-脱乙酰化壳聚糖	

用于基因转移的聚合物	
聚合物	缩写
组蛋白	
胶原	
葡聚糖-精胺	D-SPM

[1560]

[1561] 表13总结了用于编码本发明的融合蛋白的多核苷酸的递送方法。

[1562] 表13

递送	载体/模式	递送至非分裂细胞内	表达的持续时间	基因组整合	所递送的分子类型
物理	(例如,电穿孔、粒子枪、磷酸钙转染	是	瞬时	否	核酸和蛋白质
病毒	逆转录病毒	否	稳定	是	RNA
	慢病毒	是	稳定	是/否具有修饰	RNA
	腺病毒	是	瞬时	否	DNA
	腺相关病毒(AAV)	是	稳定	否	DNA
	牛痘病毒	是	非常短暂	否	DNA
	单纯疱疹病毒	是	稳定	否	DNA
非病毒	阳离子脂质体	是	瞬时	取决于递送的是什么	核酸和蛋白质
	聚合物纳米粒子	是	瞬时	取决于递送的是什么	核酸和蛋白质
生物非病毒	减毒细菌	是	瞬时	否	核酸
递送媒介	工程化的细菌噬菌体	是	瞬时	否	核酸
	哺乳动物病毒样粒子	是	瞬时	否	核酸
	生物脂质体:红细胞血影和外	是	瞬时	否	核酸

[1563]

递送	载体/模式	递送至非分裂细胞内	表达的持续时间	基因组整合	所递送的分子类型
	泌体				

[1564]

[1565] 在另一方面,基因组编辑系统组分或编码此类组分的核酸(例如,核酸结合蛋白诸如,举例而言,Cas9或其变体)以及靶向感兴趣的基因组核酸序列的gRNA的递送,可以通过

递送核糖核蛋白(RNP)至细胞而得以实现。RNP包含处于与靶向gRNA的复合物中的核酸结合蛋白,例如,Cas9。可以使用已知方法将RNP递送至细胞,所述方法诸如电穿孔、核转染或阳离子脂质介导的方法,例如,如Zuris,J.A.et al.,2015,Nat.Biotechnology,33(1):73-80中报道的。就用于CRISPR碱基编辑系统而言,尤其是针对难以转染的细胞诸如原代细胞,RNP具有优势。此外,RNP也可以缓解在细胞内进行蛋白质表达时可能出现的困难,尤其是当不能良好地表达可能用于CRISPR质粒中的真核启动子(例如,CMV或EF1A)时。有利的是,RNP的使用并不需要将外来DNA递送到细胞内。此外,因为包含核酸结合蛋白和gRNA复合物的RNP随着时间推移而降解,RNP的使用具有限制脱靶效应的潜力。以类似于基于质粒的技术所采用的方式,RNP可以用来递送结合蛋白(例如,Cas9变体)并且引导同源诱导修复(HDR)。

[1566] 用来驱动编码核酸分子表达的碱基编辑器的启动子可以包括AAV ITR。这对于消除对于可能占据载体内空间的额外启动子元件的需求而言可能是有利的。释放的额外空间可以用来驱动额外元件诸如向导核酸或可选标记物的表达。ITR活性相对较弱,因此它可以用来降低由于所选核酸酶的过度表达带来的潜在毒性。

[1567] 任何合适的启动子可以用来驱动碱基编辑器和(若适宜)向导核酸的表达。对于泛素化表达,可使用的启动子包括CMV、CAG、CBh、PGK、SV40、铁蛋白重链或轻链等。对于脑或其他CNS细胞表达,合适的启动子包括:用于全部神经元的突触蛋白I、用于兴奋性神经元的CaMKIIα、用于GABA神经元的GAD67或GAD65或VGAT等。对于肝细胞表达,合适的启动子包括白蛋白启动子。对于非细胞表达,合适的启动子可以包括SP-B。对于内皮细胞,合适的启动子可以包括ICAM。对于造血细胞,合适的启动子可以包括IFNβ或CD45。对于成骨细胞,合适的启动子可以包括OG-2。

[1568] 在一些情况下,本公开的碱基编辑器的尺寸足够小,以致于允许独立的启动子驱动碱基编辑器和同一核酸分子中的兼容向导核酸的表达。例如,载体或病毒载体可以包含可操作地连接至编码碱基编辑器的核酸的第一启动子以及可操作地连接至向导核酸的第二启动子。

[1569] 用来驱动向导核酸的表达的启动子可以包括:Pol III启动子,诸如U6或H1。Pol II启动子和内含子盒用来表达gRNA腺相关病毒(AAV)。

[1570] 本文所述的具有或不具有一种或多种核酸的碱基编辑器可以使用腺相关病毒(AAV)、慢病毒、腺病毒或其他质粒或病毒载体类型递送,尤其是使用来自例如美国专利号8,454,972(关于腺病毒的制剂、剂量)、美国专利号8,404,658(关于AAV的制剂、剂量)和美国专利号5,846,946(关于DNA质粒的制剂、剂量)以及来自牵涉慢病毒、AAV和腺病毒的临床试验与关于该临床试验的出版物的制剂和剂量递送。例如,对于AAV,给药途径、制剂和剂量可以如美国专利号8,454,972和牵涉AAV的临床试验中所示。例如,对于腺病毒,给药途径、制剂和剂量可以如美国专利号8,404,658和牵涉腺病毒的临床试验中所示。例如,对于质粒递送,给药途径、制剂和剂量可以如美国专利号5,846,946和牵涉质粒的临床研究中所示。剂量可以基于或外推至平均体重为70kg的个体(例如,成年男性),并且可以针对患者、受试者、不同体重和物种的哺乳动物进行调节。给药频率处于医疗或兽医从业者(例如,医师、兽医)的能力范围内,取决于包括以下的因素:患者或受试者的年龄、性别、一般健康情况、其他病症以及待解决的特定病症或症状。可以将病毒载体注射到感兴趣的组织内。对于细胞类型特异性的碱基编辑,碱基编辑器和任选的向导核酸的表达可以由细胞类型特异性启动

子驱动。

[1571] 对于体内基因递送,AAV可能比其他病毒载体有利。在一些情况下,AAV导致低毒性,这可能是由于纯化方法不需要对细胞离子进行超离心,而超离心可能激活免疫应答。在一些情况下,因为AAV不整合到宿主基因组中,所以导致造成插入性诱变的可能性较低。

[1572] AAV具有4.5或4.75Kb的封装限制。这意味着所公开的碱基编辑器以及启动子和转录终止子可以装入单个病毒载体中。大于4.5或4.75Kb的构建体可能导致显著减少的病毒产生。例如,SpCas9很大,该基因本身超过4.1Kb,这使得它难以封装到AAV中。因此,本公开的实施方案包括利用所公开的长度比传统碱基编辑器短的碱基编辑器。在一些实施例中,碱基编辑器小于4kb。所公开的碱基编辑器可以小于4.5kb、4.4kb、4.3kb、4.2kb、4.1kb、4kb、3.9kb、3.8kb、3.7kb、3.6kb、3.5kb、3.4kb、3.3kb、3.2kb、3.1kb、3kb、2.9kb、2.8kb、2.7kb、2.6kb、2.5kb、2kb或1.5kb。在一些情况下,所公开的碱基编辑器的长度为4.5kb或更短。

[1573] AAV可以是AAV1、AAV2、AAV5或其任何组合。可以关于待靶向的细胞而选择AAV的类型,例如,可以选择AAV血清型1、2、5或杂交衣壳AAV1、AAV2、AAV5或其任何组合来靶向脑或神经元细胞;并且可以选择AAV4来靶向心脏组织。AAV8可用于递送至肝脏。关于这些细胞的某些AAV血清型的列表可以在Grimm,D.et al,J.Virol.82:5887-5911(2008)中找到。

[1574] 慢病毒是复合逆转录病毒,其具有感染并且在有丝分裂细胞和有丝分裂后细胞中表达其基因的能力。最常见的已知慢病毒是人免疫缺陷病毒(HIV),其使用其他病毒的包膜糖蛋白来靶向大范围的细胞类型。

[1575] 慢病毒可以如下制备。在克隆pCasES10(其含有慢病毒转移质粒主链)后,在转染前一天,将低代(p=5)的HEK293FT接种到T-75烧瓶中的具有10%胎牛血清且不含抗生素的DMEM中,达到50%汇合。20小时后,将培养基变为OptiMEM(无血清)培养基,并在4小时后进行转染。用10μg的慢病毒转移质粒(pCasES10)和下列封装质粒转染细胞:5μg的pMD2.G(VSV-g假型)和7.5μg的psPAX2(gag/pol/rev/tat)。转染可以在4mL OptiMEM中用阳离子脂质递送剂(50μl Lipofectamine 2000和100ul Plus试剂)进行。6小时后,将培养基变为具有10%胎牛血清的无抗生素DMEM。这些方法在细胞培养期间使用血清,但优选无血清的方法。

[1576] 慢病毒可以如下纯化。在48小时后收获病毒上清液。首先清除上清液中的碎屑,并且通过0.45μm低蛋白结合(PVDF)过滤器过滤。然后将它们在超离心机中以24,000rpm旋转2小时。将病毒球丸重新悬浮在50μl的DMEM中,在4℃过夜。然后将它们分为等量小样并立即在-80℃冷冻。

[1577] 在另一实施方案中,还考虑了基于马感染性贫血病毒(EIAV)的最小非灵长动物慢病毒载体。在另一实施例中,考虑经由视网膜下注射递送RETINOSTAT®,一种表达抑制血管生成的蛋白质(即,内皮抑素和血管抑素)的基于马感染性贫血病毒的慢病毒基因治疗载体。在另一实施方案中,考虑使用自灭活的慢病毒载体。

[1578] 该系统的任何RNA,例如,向导RNA或编码碱基编辑器的mRNA,可以RNA的形式递送。可以使用体外转录产生编码碱基编辑器的mRNA。例如,可以使用含有下列元件的PCR盒合成核酸酶mRNA:T7启动子、任选的kozak序列(GCCACC)、核酸酶序列和3'UTR诸如来自β珠蛋白-polyA尾部的3'UTR。该盒可以用于通过T7聚合酶进行的转录。向导多核苷酸(例如,gRNA)也

可以使用体外转录从含有T7启动子、其后的序列“GG”和向导多核苷酸序列的盒转录。

[1579] 为了增强表达并降低可能的毒性,可以例如使用假-U或5-甲基-C将编码碱基编辑器的序列和/或向导核酸修饰为包括以或多个经修饰的核苷酸。

[1580] 在一些实施方案中,本公开提出修饰细胞或生物体的方法。细胞可以是原核细胞或真核细胞。细胞可以是哺乳动物细胞。哺乳动物细胞可以是非人灵长动物、牛、猪、啮齿动物或小鼠细胞。通过本公开的碱基编辑器、组合物和方法引入细胞内的修饰可以使得该细胞或该细胞的后裔被改变,以改善生物产物诸如抗体、淀粉、醇或其他所希望的细胞产出的生产。通过本公开的方法引入细胞内的修饰可使得该细胞或该细胞的后裔包括改变,该改变使得所生产的生物产物发生变化。

[1581] 该系统可以包括一种或多种不同的载体。在一方面,对碱基编辑器进行密码子优化以用于在所希望的细胞类型中表达,以真核细胞为优先,优选哺乳动物细胞或人细胞。

[1582] 通常,密码子优化是指,通过将原生序列的至少一个密码子(例如,约或超过约1、2、3、4、5、10、15、20、25、50或更多个密码子)替换为更频繁或最频繁地用于该宿主细胞的基因中但同时维持原生氨基酸序列的密码子,修饰核酸序列以增强在感兴趣的宿主细胞中的表达的过程。多个物种针对特定氨基酸的某些密码子表现出特定的偏倚。密码子偏倚(生物体之间的密码子选择差异)经常与信使RNA(mRNA)的翻译效率相关联,而信使RNA(mRNA)的翻译效率据信取决于被翻译的密码子的性质和特定转移RNA(tRNA)分子的可用性等。细胞中占据主导地位的所选tRNA通常反映了肽合成中最频繁使用的密码子。据此,可基于密码子优化裁剪基因以在给定生物体中实现最优基因表达。密码子选择表可以例如在www.kazusa.orjp/codon/(2002年7月9日)获得的“密码子选择数据库”中轻易地获得,并且这些表可以通过多种方式调整。参见,Nakamura,Y.,et al.″Codon usage tabulated from the international DNA sequence databases:status for the year 2000″Nucl.Acids Res.28:292(2000)。用于针对在特定宿主细胞中表达而对特定序列进行密码子优化的计算机算法也是可获得的,诸如Gene Forge(Aptagen;Jacobus,Pa.)。在一些实施方案中,编码经工程化的核酸酶的序列中的一个或多个密码子(例如,1、2、3、4、5、10、15、20、25、50或更多个或全部密码子)对应于针对特定氨基酸最频繁使用的密码子。

[1583] 通常使用封装细胞来形成能够感染宿主细胞的病毒粒子。此类细胞包括封装腺病毒的293细胞,以及封装逆转录病毒的ψ2细胞或PA317细胞。用于基因疗法中的病毒载体一般通过生产将核酸载体封装到病毒粒子中的细胞系而产生。载体通常含有封装和后续整合到宿主中所需的最小病毒序列,其他病毒序列则替换为用于待表达的多核苷酸的表达盒。缺失的病毒功能通常由封装细胞系以反式提供。例如,用于基因疗法的AAV载体通常仅具备封装和整合到宿主基因组中所需的来自AAV基因组的ITR序列。病毒DNA可以封装在细胞系中,该细胞系含有编码其他AAV基因的名为rep和cap的辅助质粒但缺少ITR序列。细胞系也可以被作为辅助者的腺病毒感染。辅助病毒可以促进AAV载体的复制以及AAV基因从辅助质粒表达。在一些情况下,辅助质粒由于缺少ITR序列而不以显著的量被封装。可以通过例如热处理减少腺病毒的污染,因为与AAV相比,腺病毒对于热处理更加敏感。

[1584] 内含肽

[1585] 在一些实施方案中,将核酸酶(例如,Cas9)的一部分或片段融合至内含肽。核酸酶可以融合至内含肽的N端或C端。在一些实施方案中,融合蛋白的一部分或片段融合至内含

肽并且融合至AAV衣壳蛋白。内含肽、核酸酶和衣壳蛋白可以任何排列(例如,核酸酶-内含肽-衣壳、内含肽-核酸酶-衣壳、衣壳-内含肽-核酸酶等)融合在一起。在一些实施方案中,内含肽的N端融合至融合蛋白的C端,并且内含肽的C端融合至AAV衣壳蛋白的N端。

[1586] 内含肽(干预蛋白)是见于多种不同生物体中的自加工结构域,其完成被称为蛋白质剪接的过程。蛋白质剪接是一种多步骤生化反应,由肽键的裂解和形成构成。尽管蛋白质剪接的内源性底物是在含有内含肽的生物体中发现的蛋白质,但内含肽也可以用来化学地操纵几乎任何多肽主链。

[1587] 在蛋白质剪接中,内含肽通过裂解两个肽键而从前体多肽切除自身,从而经由形成新的肽键而连接位于两侧的外显肽(外部蛋白质)。这一重排发生在翻译后(或可能与翻译同时发生)。内含肽介导的蛋白质剪接自发地发生,仅需要内含肽结构域的折叠。

[1588] 大约5%的内含肽是分裂内含肽,其被转录并翻译为两个独立的多肽,即N-内含肽和C-内含肽,各自融合至一个外显肽。翻译之后,内含肽片段自发地且非共价地组装为经典内含肽结构以完成U反式蛋白质剪接。蛋白质剪接的机制包括一系列酰基转移反应,这些反应导致位于内含肽-外显肽接合处的两个肽键断裂以及N-外显肽与C外显肽之间的新肽键形成。这一过程通过激活接合N-外显肽与内含肽的N端的肽键启动。几乎全部内含肽均具有位于其N端的半胱氨酸或丝氨酸,该氨基酸供给C端N-外显肽残基的羰基碳。保守的苏氨酸和组氨酸(称为TXXH基序)以及常见的天冬氨酸一起促进了这种N到O/S酰基的转移,导致线性(硫代)酯中间体的形成。之后,这一中间体通过第一个C-外显肽残基(+1)的亲核攻击而历经反式(硫代)酯化,该残基是半胱氨酸、丝氨酸或苏氨酸。所得的支化(硫代)酯中间体通过内含肽的高度保守的C端天冬酰胺的独特的转化:环化得以拆分。这一过程为组氨酸(见于高度保守的HNF基序中)和倒数第二的组氨酸所促进,并且也可能牵涉该天冬氨酸。这一琥珀酰亚胺形成反应将内含肽从反应性复合物中切除,并留下通过非肽链接附接的外显肽。这一结构以不依赖于内含肽的模式迅速重排为稳定的肽键。

[1589] 在一些实施方案中,碱基编辑器(例如,ABE、CBE)的N端片段融合至分裂内含肽-N,并且C端片段融合至分裂内含肽-C。这些片段随后被封装到两个或更多个AAV载体中。某些内含肽用于接合异源蛋白片段的用途在例如Wood et al.,J.Biol.Chem.289(21);14512-9(2014)中有所描述。例如,当融合至独立的蛋白片段时,内含肽IntN和IntC彼此识别,将其自身剪出并同步连接位于两侧的它们所融合的蛋白片段的N端和C端外显肽,从而从两个蛋白片段重构全长度蛋白质。其他合适的内含肽对于本领域技术人员将是显而易见的。

[1590] 在一些实施方案中,ABE在所选的SpCas9区域内的Ala、Ser、Thr或Cys残基处分裂为N端片段和C端片段。这些区域对应于通过Cas9晶体结构分析检定的环区域。在氨基酸位置S303、T310、T313、S355、A456、S460、A463、T466、S469、T472、T474、C574、S577、A589和S590处,每个片段的N端融合至内含肽-N,并且每个片段的C端融合至内含肽-C,这些位置在下述序列中以粗体大写字母表示。

1 mdkkysigld igtnsvgwav itdeykvpsk kfkvlgntdr hsikknliga llfdsgetae
61 atrlkrtarr rytrrknric ylqeifsnem akvddsffhr leesflveed kkherhpifg
121 nivdevayhe kyptiyhlrk klvdstdkad lrliylalah mikfrghfli egdlnpdnsd
181 vdklfiqlvq tynqlfeenp inasgvdaka ilsarlsksr rlenliaqlp gekknglfgn
241 lialslgltp nfksnfdlae daklqlskdt ydddldnlla qigdqyadlf laaknlsdai
301 ll**S**dilrvn**T** ei**T**kaplsas mikrydehhq dltllkalvr qqlpekykei ffdq**S**kngya
361 gyidggasqe efykfikpil ekmdgteell vklnredllr kqrtfdngsi phqihlgelh
421 ailrrqedfy pflkdnreki ekiltfripy yvgpl**A**rgn**S** rf**A**wm**T**rk**S**e e**T**i**T**pwnfee
481 vvdkgasaqs fiermtnfdk nlpnekvlpk hsllyeyftv yneltkvkyv tegmrkpafl
541 sgeqkkaivd llfktnrkvt vkqlkedyfk kie**C**fd**S**vei sgvedrfn**AS** lgtyhdllki
601 ikdkdfldne enedilediv ltltlfedre mieerlktya hlfddkvmkq lkrrrytgwg
661 rlsrklingi rdkqsgktil dflksdgfan rnfmqlihdd sltfkediqk aqvsgqgdsl
721 hehianlags paikkgilqt vkvvdelvkv mgrhkpeniv iemarenqtt qkgqknsrer
781 mkrieegike lgsqilkehp ventqlqnek lylyylqngr dmyvdqeldi nrlsdydvdh
841 ivpqsflkdd sidnkvltrs dknrgksdnv pseevvkkmk nywrqllnak litqrkfdnl
901 tkaergglse ldkagfikrq lvetrqitkh vaqildsrmn tkydendkli revkvitlks
961 klvsdfrkdf qfykvreinn yhhahdayln avvgtalikk ypklesefvy gdykvydvrk
1021 miakseqeig katakyffys nimnffktei tlangeirkr plietngetg eivwdkgrdf
1081 atvrkvlsmp qvnivkktev qtggfskesi lpkrnsdkli arkkdwdpkk yggfdsptva
1141 ysvlvvakve kgkskklksv kellgitime rssfeknpid fleakgykev kkdliiklpk
1201 yslfelengr krmlasagel qkgnelalps kyvnflylas hyeklkgspe dneqkqlfve
1261 qhkhyldeii eqisefskrv iladanldkv lsaynkhrdk pireqaenii hlftltnlga
[1591]

1321 paafkyfdtt idrkrytstk evldatlihq sitglyetri dlsqlggd
[1592]

[1593] 药物组合物

[1594] 本公开的其他方面涉及药物组合物,其包含本文所述的任何碱基编辑器、融合蛋白或融合蛋白-向导多核苷酸复合物。如本文中所用,术语“药用组合物”可以指配制为用于药学用途的组合物。在一些实施方案中,药物组合物进一步包含药学可接受的载剂。在一些实施方案中,药物组合物包含额外试剂(例如,用于特异性递送、增加半衰期或其他治疗性化合物)。

[1595] 如本文中所用,术语“药学上可接受的载剂”意为药学上可接受的材料、组合物或媒介物,诸如液体或固体填充物、稀释剂、赋形剂、制造助剂(例如,润滑剂、滑石、硬脂酸镁、硬脂酸钙、硬脂酸锌或硬脂酸)或溶剂胶囊化材料,其牵涉入将化合物从身体的一个位点(例如,递送位点)携带或运输至身体的另一位点(例如,器官、组织或胜肽的部分)。就与制剂的其他成分的相容性而言,药学上可接受的载剂是“可接受的”并且对受试者的组织无害(例如,生理学上相容的、无菌的、生理pH等)。

[1596] 可用作药学上可接受的载剂的材料的一些非限制性示例包括:(1)糖类,诸如乳糖、葡萄糖和蔗糖;(2)淀粉类,诸如玉米淀粉和马铃薯淀粉;(3)纤维素及其衍生物,诸如羧

甲基纤维素钠、甲基纤维素、乙基纤维素、微晶纤维素和醋酸纤维素;(4)黄蓍粉;(5)麦芽;(6)明胶;(7)润滑剂,诸如硬脂酸镁、十二烷基硫酸钠和滑石;(8)赋形剂,诸如可可脂和栓蜡;(9)油类,诸如花生油、棉籽油、葵花籽油、芝麻油、橄榄油、玉米油和大豆油;(10)二醇类,诸如丙二醇;(11)多元醇类,诸如甘油、山梨醇、甘露醇和聚乙二醇(PEG);(12)酯类,诸如油酸乙酯和月桂酸乙酯;(13)琼脂;(14)缓冲剂,诸如氢氧化镁和氢氧化铝;(15)海藻酸;(16)无热原水;(17)等张盐水;(18)Ringer溶液;(19)乙醇;(20)pH缓冲溶液;(21)聚酯类、聚碳酸酯类和/或聚酐类;(22)填充剂,诸如多肽和氨基酸;(23)血清醇类,诸如乙醇;和(23)其他用于药物制剂中的无毒相容物质。润湿剂、着色剂、释放剂、包衣剂、甜味剂、风味剂、芳香剂、防腐剂和抗氧化剂也可以存在于制剂中。术语诸如“赋形剂”、“载剂”、“药学上可接受的载剂”、“媒介物”等在本文中可互换地使用。

[1597] 药物组合物可以包含一种或多种pH缓冲化合物以将制剂的pH维持在预定水平,该水平反映了生理pH,诸如约5.0至约8.0的范围内。水性液体制剂中使用的pH缓冲化合物可以是氨基酸或氨基酸的混合物,诸如组氨酸或者氨基酸的混合物诸如组氨酸和甘氨酸。作为另一种选择,pH缓冲化合物优选为将制剂的pH维持在预定水平(诸如约5.0至约8.0的范围内)并且不螯合钙离子的试剂。此类pH缓冲化合物的说明性示例包括但不限于,咪唑和醋酸根离子。pH缓冲化合物可能以任何适于将制剂的pH维持在预定水平的量存在。

[1598] 药物组合物也可以含有一种或多种渗透调节剂的化合物,即,将制剂的渗透性质(例如,紧张性、渗透性和/或渗透压)调节至接纳者个体的血流和血细胞可以接受的水平。渗透调节剂可以是不螯合钙离子的试剂。渗透调节剂可以是本领域技术人员已知或可获得的调节制剂的渗透性质的任何化合物。本领域技术人员可以经验性地确定给定渗透调节剂用于本发明制剂中的适用性。合适类型的渗透调节剂的说明性示例包括但不限于:盐类,诸如氯化钠和醋酸钠;糖类,诸如蔗糖、葡萄糖和甘露醇;氨基酸,诸如甘氨酸;以及一种或多种这些试剂的混合物和/或一种或多种这些类型的试剂的混合物。渗透调节剂可能以足以调节制剂的渗透性质的任何浓度存在。

[1599] 在一些实施方案中,将药物组合物配制为用于递送至受试者以例如进行基因编辑。给药本文所述药物组合物的合适途径包括而不限于:外用、皮下、透皮、皮内、病灶内、关节内、腹膜内、膀胱内、跨粘膜、齿龈、皮内、耳蜗内、经鼓膜、器官内、硬膜外、鞘内、肌肉内、静脉内、血管内、骨内、眼周、肿瘤内、脑内和脑室内给药。

[1600] 在一些实施方案中,本文所述的药物组合物局部给药。在一些实施方案中,通过注射、通过导管、通过栓剂、或通过植入物将本文所述的药物组合物给药至受试者,该植入物是多孔的、无孔的或凝胶状材料,包括膜(诸如硅橡胶膜)或纤维。

[1601] 在其他实施方案中,将本文所述的药物组合物以控制释放系统的方式递送。在一个实施方案中,可以使用泵(参见,例如,Langer,1990,Science249:1527-1533;Sefton,1989,CRC Crit.Ref.Biomed.Eng.14:201;Buchwald et al.,1980,Surgery 88:507;Saudek et al,1989,N.Engl.J.Med.321:574)。在另一实施方案中,可以使用聚合材料。(参见,例如,Medical Applications of Controlled Release(Langer和Wise编,CRC Press,Boca Raton,Fla.,1974);Controlled Drug Bioavailability,Drug Product Design and Performance(Smolen和Ball编,Wiley,New York,1984);Ranger and Peppas,1983,Macromol.Sci.Rev.Macromol.Chem.23:61。也参见,Levy et al.,1985,Science 228:190;

During et al.,1989,Ann.Neurol.25:351;Howard et ah,1989,J.Neurosurg.71:105。)例如,在上文的Langer文献中讨论了其他控制释放系统。

[1602] 在一些实施方案中,根据常规过程将药物组合物配制为适用于静脉内或皮下给药至受试者(例如,人)的组合物。在一些实施方案中,用于通过注射给药的药物组合物是无菌等渗溶液,用作增溶剂和局部麻醉剂(诸如利多卡因)以缓解注射部位的疼痛。通常,这些成分以单位剂型单独地或混合在一起,例如,作为干燥冻干粉或无水浓缩物,提供在标注活性剂数量的密封容器诸如安瓿或小袋中。若药物待通过输液给药,则可使用含有无菌医药级水或盐水的输液瓶分配。若药物组合物通过注射给药,则可提供无菌注射用水或盐水的安瓿,使得可以在给药前混合这些成分。

[1603] 用于全身性给药的药物组合物可以是液体,例如,无菌盐水、乳酸盐林格溶液或汉克溶液。此外,药物组合物可以是固体形式并且在重新溶解或悬浮后立即使用。也预期冻干形式。药物组合物可以包含在脂质颗粒或媒介物诸如脂质体或微晶体内,这也适用于肠胃外给药。颗粒可以是任何适合的结构,诸如单层的或多层的,只要组合物包含在其中即可。化合物可以包埋在“稳定化的质粒-脂质颗粒”(SPLP)中,该颗粒含有融合脂质二油酰基磷脂酰乙醇胺(DOPE)、低含量(5-10mol%)的阳离子脂质,并且通过聚乙二醇(PEG)涂层予以稳定化(Zhang Y.P.et al.,Gene Ther.1999,6:1438-47)。对于此类颗粒和媒介物,尤其优选带正电的脂质诸如N-[1-(2,3-二油酰氧基)丙基]-N,N,N-三甲基-甲基硫酸铵或“DOTAP”。此类脂质颗粒的制备是众所周知的。见,例如,美国专利号4,880,635、4,906,477、4,911,928、4,917,951、4,920,016和4,921,757,其各自通过引用并入本文。

[1604] 例如,本文所述的药物组合物可以作为单位剂量给药或封装。当用于本公开的药物组合物时,术语“单位剂量”是指适合作为单一剂量用于受试者的物理上离散的单位,每个单位包含预定数量的活性材料,该预定数量计算为产生所希望的治疗效果,且该活性材料与需要的稀释剂即载体或媒介物联合。

[1605] 再者,药物组合物可以提供为包含以下向的药物试剂盒:(a)含有冻干形式的本发明化合物的容器,和(b)含有药学上可结束的稀释剂(例如,用于重构或稀释冻干的本发明化合物的无菌稀释剂)。与此类容器任选地关联的可以是由政府机构规定的调节药物或生物产品的制造、使用或规格形式的注意事项,该注意事项反映了该机构批准的用于人类给药的制造、使用或规格。

[1606] 在其他方面,包括含有上述可用于治疗疾病的物质的制品。在一些实施方案中,该制品包含容器和标签。合适的容器包括,例如,瓶、小瓶、注射器和试管。容器可由各种材料诸如玻璃或塑料制成。在一些实施方案中,该容器容纳本文所述的对于治疗疾病有效的组合物并且可以具有无菌接口。例如,容器可以是静脉注射溶液袋或具有可被皮下注射针刺穿的塞子的小瓶。组合物中的活性剂是本发明的化合物。在一些实施方案中,位于容器上或与容器关联的标签指示该组合物用于治疗所选择的疾病。该制品可以进一步包含第二容器,该第二容器包含药学上可接受的缓冲剂,诸如磷酸盐缓冲盐水、Ringer溶液或葡萄糖溶液。其可进一步包括从商业和使用者角度所需的其他材料,包括其他缓冲剂、稀释剂、过滤器、针头、注射器和具有使用说明书的包装插页。

[1607] 在一些实施方案中,本文所述的任何融合蛋白、gRNA和/或复合物作为药物组合物的一部分提供。在一些实施方案中,药物组合物包含本文提供的任何融合蛋白。在一些实施

方案中,药物组合物包含本文提供的任何复合物。在一些实施方案中,药物合物包含和碳核蛋白复合物,该复合物包含与gRNA和阳离子脂质形成复合物的RNA引导的核酸酶(例如Cas9)。在一些实施方案中,药物组合物包含gRNA、核酸可编程DNA结合蛋白、阳离子脂质和药学可接受的赋形剂。药物组合物可以任选地包含一种或多种额外的治疗活性物质。

[1608] 核碱基编辑器用于靶向HBV ORFS中的核苷酸的用途

[1609] 如本文所述,评估了核碱基编辑器靶向HBV ORFS中的核苷酸的适用性。在一种实施方案中,用编码本文所述的核碱基编辑器的一种或多种核酸分子与少量编码报告基因(例如,GFP)的载体一起转染、转导或以其他方式修饰感兴趣的单细胞(例如,感染HBV的动物细胞)。这些细胞可以是永生化的人细胞系,诸如293T、K562或U20S。作为另一种选择,可以使用原代人细胞。此类细胞可能与最终的细胞靶标相关。

[1610] 可以使用病毒载体执行递送。在一种实施方案中,可以使用脂质转染(诸如Lipofectamine或Fugene)或通过电穿孔执行转染)。转染之后,可通过荧光显微镜检查或通过流式细胞术测定GFP的表达,以证实持续且高水平的转染。这些初步转染可包含不同的核碱基编辑器以确定编辑器的哪些组合给出最大活性。

[1611] 如本文所述,例如,通过将HBV的基因组或ORF测序以检测靶标序列中的改变,来评估核碱基编辑器的活性。对于Sanger测序,将纯化的PCR扩增子克隆到质粒主链中,转化,小规模制备并使用单引物测序。也可以使用下一代测序技术执行测序。当使用下一代测序时,扩增子可以是具有位于不对称位置的预期切口位点的300-500bp。在PCR之后,可以将下一代测序适配器和条码(例如,Illumina多路复用适配器和索引)添加到扩增子的末端,例如,以在高通量测序中(例如在Illumina MiSeq上)使用。

[1612] 可以选择在初始测试中诱导最高水平的靶标特异性改变的融合蛋白进行进一步评估。

[1613] 在特定实施方案中,使用核碱基编辑器来靶向感兴趣的多核苷酸。在一种实施方案中,将本发明的核碱基编辑器与向导RNA协同递送至细胞(例如,肝脏),该向导RNA用于靶向HBV基因组内的核酸序列,从而改变HBV基因。

[1614] 在一些实施方案中,碱基编辑器为向导RNA所靶向以将一个或多个错义突变或提前终止密码子引入HBV基因组的聚合酶(pol)基因中。在一些实施方案中,通过碱基编辑器引入pol基因中的一种或多种突变编码E24G、L25F、P26F、R27C、V48A、V48I、S382F、V378I、V378A、V379I、V379A、L377F、D380G、D380N、F381P、R376G、A422T、F423P、A432V、M433V、P434S、D540G、A688V、D689G、A717T、E718K、P713S、P713L和L719P。在一些实施方案中,通过碱基编辑器引入pol基因中的一种或多种突变选自由下列组成的组:E24G、L25F、P26F、R27C、V48A、V48I、S382F、V378I、V378A、V379I、V379A、L377F、D380G、D380N、F381P、R376G、A422T、F423P、A432V、M433V、P434S、D540G、A688V、D689G、A717T、E718K、P713S、P713L和L719P。

[1615] 在一些实施方案中,如本文所述的碱基编辑器为向导RNA所靶向以将一个或多个错义突变或提前终止密码子引入HBV基因组的核心区基因中。在一些实施方案中,被引入核心区基因中的一种或多种突变包含T160A、T160A、P161F、S162L、C183R和/或*(STOP)184Q。在一些实施方案中,被引入核心区基因中的一种或多种突变选自由以下组成的组:T160A、T160A、P161F、S162L、C183R和*(STOP)184Q。

[1616] 在一些实施方案中,如本文所述的碱基编辑器为向导RNA所靶向以将一个或多个错义突变或提前终止密码子引入HBV基因组的X基因中。在一些实施方案中,被引入X基因中的一种或多种突变包含H86Y、R87*(STOP)、H86R、W120R、W120STOP、E122K、E121K和/或L141P。在一些实施方案中,被引入X基因中的一种或多种突变选自由以下组成的组:H86Y、R87*(STOP)、H86R、W120R、W120STOP、E122K、E121K和L141P。

[1617] 在一些实施方案中,如本文所述的碱基编辑器为向导RNA所靶向以将一个或多个错义突变或提前终止密码子引入HBV基因组的S基因中。在一些实施方案中,该一种或多种突变包含S38F、L39F、W35STOP、W35R、W36STOP、W36R、T37I、T37A、R78Q、S34L、F80P和/或D33G。在一些实施方案中,被引入S基因中的一种或多种突变选自由以下组成的组S38F、L39F、W35*(STOP)、W35R、W36*(STOP)、W36R、T37I、T37A、R78Q、S34L、F80P和D33G。

[1618] 试剂盒

[1619] 本公开的多个方法提供包含碱基编辑器系统的试剂盒。在一种实施方案中,试剂盒包含核酸构建体,该核酸构建体包含编码核碱基编辑器融合蛋白的核苷酸序列。融合蛋白包含脱氨酶(例如,胞苷脱氨酶或腺嘌呤脱氨酶)和核酸可编程DNA结合蛋白(napDNAbp)。在一些实施方案中,试剂盒包含至少一种能够靶向HBV基因的向导RNA。在一些实施方案中,试剂盒包含核酸构建体,该核酸构建体包含编码至少一种向导RNA的核苷酸序列。

[1620] 在一些实施方案中,试剂盒提供关于使用该试剂盒来编辑HBV基因组中的一种或多种基因的使用说明书。使用说明书通常将包括关于该试剂盒用于编辑核酸尤其是病毒核酸的用途的信息。在其他实施方案中,使用说明书包括以下中的至少一项:注意事项;警告;临床研究;和/或参考文献。使用说明书可直接印在容器(当存在时)上,或作为用于该容器的标签,或作为独立的页、册、卡或折叠式印刷品提供在给容器内或与该容器一起提供。在其他实施方案中,试剂盒可以包含关于合适的操作参数的标签或独立插页(包装插页)形式的使用说明书。在又一实施方案中,试剂盒可以包含一个或多个带有适宜的正对照和负对照或对照样品的容器,该对照或对照样品待作为标准品用于检测、校准或归一化。

[1621] 在某些实施方案中,试剂盒可用于治疗患有HBV感染的受试者。

[1622] 治疗HBV感染的方法

[1623] 本发明提供治疗HBV感染或其症候的方法,该方法包括向受试者(例如,哺乳动物,诸如人)给药治疗有效量的药物组合物,该药物组合物包含编码本文所述碱基编辑器系统(例如,碱基编辑器和gRNA)的多核苷酸。在一些实施方案中,碱基编辑器是融合蛋白,其包含多核苷酸可编程DNA结合结构域和腺苷脱氨酶结构域或胞苷脱氨酶结构域。用该碱基编辑器和一种或多种向导多核苷酸转导受试者的细胞,该向导多核苷酸靶向该碱基编辑器以实现编码HBV多肽的核酸序列的A·T到G·C改变(如果用腺苷脱氨酶结构域转导细胞)或C·G到U·A改变(如果用胞苷脱氨酶结构域转导细胞)。

[1624] 在一些实施方案中,治疗慢性乙型肝炎包括多种方法的联合。例如,可以向感染HBV的受试者给药治疗有效量的上述药物组合,该药物组合靶向并修饰HBV cccDNA。例如,不含UGI结构域的BE4碱基编辑器可与靶向HBV基因组区域的gRNA合用。不受限于理论,忽略抑制剂使得C->U脱氨基作用对尿嘧啶糖基化酶敏感,该酶损坏HBV cccDNA,从而使其不稳定。该治疗可以与降低或抑制HBsAg表达(包括来自整合的HBV DNA的表达)的治疗联合使用,后者诸如使用本文所述的碱基编辑器和向导RNA靶向HBV基因组的S或pol基因。治疗也

可以包括刺激免疫系统。在一些实施方案中,这三种不同的治疗目标中的每一个可以使用本文所述的碱基编辑试剂和技术实施。

[1625] 本文的方法包括对该受试者(包括被证实需要此治疗的受试者)给药有效量的本文所述的组合物。证实受试者需要此治疗可为专业医护人员对受试者的判断,且可以是主观的(例如,观点)或客观的(例如,通过测试或诊断方法可测)。

[1626] 通常,本发明的治疗方法包括给药治疗有效量的药物组合物,该药物组合物包含例如编码碱基编辑器和gRNA的载体,该gRNA靶向有此需要的受试者(例如,人类)的HBV基因。此治疗将被适当地给予苦于、患有、易患HBV感染或处于HBV感染风险下的受试者,尤其是人类。本文的组合物也可用于治疗可能涉及HBV感染的任何其他疾患。

[1627] 在一种实施方案中,本发明提供监测治疗进展的方法。该方法包括以下步骤:确定苦于或易患与HBV感染相关的疾患或其症候的受试者体内的诊断性标记物(标记物)(例如,病毒负载)或诊断性措施(例如,扫描、测定)的水平,其中该受试者已经被给药足以治疗该疾病或其症候的治疗量的本文的组合物。可以将该方法中确定的标记物水平与健康的正常对照或其他被折磨的患者中的已知标记物水平进行比较,以建立该受试者的疾病状态。在优选实施方案中,在晚于确定第一水平的时间点确定该受试者体内的第二水平,并且比较两个水平以监测病程或疗效。在某些优选实施方案中,在开始根据本发明进行治疗之前确定受试者体内标记物的治疗前水平;然后可以将这一标记物的治疗前水平与治疗开始之后该受试者体内的标记物水平进行比较,以确定治疗的效力。

[1628] 除非明确指定,否则本发明的实践采用传统的分子生物学(包括重组技术)、微生物学、细胞生物学、生物化学和免疫学技术,这些技术处于该领域技术人员的知识范围内。此类技术在文献中完整地诠释,如,《分子克隆:实验室手册(第二版)》(Molecular Cloning:A Laboratory Manual,second edition,(Sambrook,1989));《寡核苷酸合成》(Oligonucleotide Synthesis(Gait,1984));《动物细胞培养》(Animal Cell Culture(Freshney,1987));《实验免疫学手册》的《酶学方法》(Methods in Enzymology Handbook of Experimental Immunology(Weir,1996));《用于哺乳动物细胞的转基因载体》(Gene Transfer Vectors for Mammalian Cells(Miller and Calos,1987));和《分子生物学现代方法》(Current Protocols in Molecular Biology"(Ausubel,1987));《PCR:聚合酶链反应》(PCR:The Polymerase Chain Reaction,(Mullis,1994));《免疫学现代方法》(Current Protocols in Immunology(Coligan,1991))。这些技术可用来生产本发明的多核苷酸和多肽,且因此可在作成并实践本发明中考虑这些技术。尤其可用于特定实施方式的技术将在下文中讨论。

[1629] 提出下述实施例以向该领域技术人员提供对于如何制作并使用本发明的测定、筛选和治疗方法的完全公开和说明,但并不试图限制发明人所认定的发明范畴。

[1630] 实施例

[1631] 实施例1:对HBV基因组的靶向碱基编辑导致病毒基因中的提前终止密码子或功能性置换。

[1632] 评估了碱基编辑器(图4A、4B)编辑HBV基因组的能力。向导RNA被设计为将碱基编辑器靶向以将终止密码子或功能性置换引入HBV基因的保守区域处(图5)。使用可在Benchling Software上获得的CRISPR设计工具来分析HBV基因型D亚基因型ayw以鉴定潜在

的向导RNA,并且针对每种向导RNA确定跨全部HBV基因型的序列保守水平。选择97中向导RNA进行测试,这些向导RNA会将碱基编辑器靶向以基于编辑窗口中的胞嘧啶位置将终止密码子或功能性改变引入HBV基因中(表14)。这些向导RNA与NGG、NAG和NNNRRT(包括NNGRRT)PAM序列相邻或接近。

[1633]

表 14：向导 RNA

向导	gRNA	%终止编辑	向导链	PAM	基因	全部基因型保守%
AATACCGCAGAGTCTAGACT	M1	0.47	1	NNNRRT	S	37.15943898
TCCACCACGAGTCTAGACTC	M2	15.09	-1	NNNRRT	S	96.35660164
CACATCCAGCGATAACCAGG	M3	21.26	-1	NNNRRT	S	54.84443011
AAAGCCCAGGATGATGGGAT	M4	1.85	-1	NNGRRT	S	42.35047558
CGAACCACTGAACAAATGGC	M5	6.5	-1	NNNRRT	S	92.95502176
ATCCATATAACTGAAAGCCA	M6	28.28	-1	NNNRRT	S	74.33499919
ATACCCAAAGACAAAAGAAA	M7	3.03	-1	NNNRRT	S	75.57633403
ATGCAACTTTTTCACCTCTG	M8	0.01	1	NNNRRT	核心区	94.08350798
TCAAGCCTCCAAGCTGTGCC	M9	1.67	1	NNGRRT	核心区	97.93648235
CCCAGCAAAGAATTGCTTGC	M10	0.1	-1	NNGRRT	核心区	9.285829437
CCACCCAGGTAGCTAGAGTC	M11	4.04	-1	NNNRRT	核心区	21.32838949
AATCCACACTCCGAAAGACA	M12	8.95	-1	NNNRRT	核心区	14.78316943
GTCTCAATCGCCGCGTCGCA	M13	9.66	1	NNNRRT	核心区	68.25729486
AAGATCTCAATCTCGGGAAC	M14	13.47	1	NNNRRT	核心区	5.271642754
TTTTCCAATGAGGATTAAAG	M15	56.89	-1	NNNRRT	Pol	3.111397711
TTTACACCAAGACATTATCA	M16	7.15	1	NNNRRT	Pol	10.23698211
GAACAGTTTGTAGGCCCACT	M17	0.17	1	NNNRRT	Pol	3.256488796

[1634]

向导	gRNA	%终止编辑	向导链	PAM	基因	全部基因型保守%
ATTGCAATTGATTATGCCTG	M18	0.73	1	NNNRRT	Pol	3.320973722
CGCCTTCCATAGAGTGTGTA	M19	ND	-1	NNNRRT	Pol	15.25068515
CCCAAGAATATGGTGACCCA	M20	6.59	-1	NNNRRT	Pol	53.32903434
ACAAGATCTACAGCATGGGG	M21	13.57	1	NNGRRT	Pol	2.579397066
CAGCCTTCAGAGCAAACACA	M22	18.65	1	NNNRRT	Pol	1.338062228
TCAGAGCAAACACAGCAAAT	M23	0.55	1	NNNRRT	Pol	1.35418346
CCCCAATCCTCGAGAAGATT	M24	0.31	-1	NNNRRT	Pol	18.49105272
CCAGGACAAGTTGGAGGACA	M25	3.1	-1	NNNRRT	Pol	28.43785265
GCTGTACCAAACCTTCGGAC	M26	12.17	1	NNNRRT	Pol	13.92874416
ACCCCATCTCTTTGTTTTGT	M27	0.15	-1	NNGRRT	Pol	4.981460584
CCACAAGAACACATCATACA	M28	0.29	1	NNNRRT	Pol	1.837820409
ACTTTCCAATCAATAGGCCT	M29	4.1	-1	NNNRRT	Pol	2.063517653
GTCAACGAATTGTGGGTCTT	M30	0.02	1	NNGRRT	Pol	3.981944221
ACACAATGTGGTTATCCTGC	M31	0.17	1	NNNRRT	Pol	17.58826374
CGGCAACGGCCAGGTCTGTG	M32	1.28	1	NNNRRT	Pol	20.13541835
GTGGTCTCCATGCGACGTGC	M33	34.2	-1	NNNRRT	Pol	68.4829921
TACCGCAGAGTCTAGACTCG	M34	6.7	1	NGG	S	37.06271159
CGCAGAGTCTAGACTCGTGG	M35	0.53	1	NGG	S	37.12719652

[1635]

向导	gRNA	%终止编辑	向导链	PAM	基因	全部基因型保守%
CACCACGAGTCTAGACTCTG	M36	3.14	-1	NGG	S	94.29308399
GAAAGCCCAGGATGATGGGA	M37	16.59	-1	NGG	S	40.59326133
CCACCCAAGGCACAGCTTGG	M38	0.31	-1	NGG	核心区	94.53490247
AAGCCACCCAAGGCACAGCT	M39	27.31	-1	NGG	核心区	94.61550862
CCATGCCCCAAAGCCACCCA	M40	30.21	-1	NGG	核心区	64.59777527
TCAGGCAAGCAATTCTTTGC	M41	4.01	1	NGG	核心区	9.592132839
CAGGCAAGCAATTCTTTGCT	M42	6.22	1	NGG	核心区	9.576011607
AGGCAAGCAATTCTTTGCTG	M43	0.91	1	NGG	核心区	9.592132839
GGCAAGCAATTCTTTGCTGG	M44	0.32	1	NGG	核心区	8.189585684
GCAAGCAATTCTTTGCTGGG	M45	0.02	1	NGG	核心区	8.221828148
TCCAAGGAATACTAACATTG	M46	50.79	-1	NGG	Pol	2.950185394
TTCCAATGAGGATTAAAGAC	M47	45.95	-1	NGG	Pol	3.337094954
TGCAATTGATTATGCCTGCT	M48	1.16	1	NGG	Pol	8.866677414
TGGGAACAAGATCTACAGCA	M49	30.73	1	NGG	Pol	22.31178462
GGGAACAAGATCTACAGCAT	M50	51.93	1	NGG	Pol	22.27954216
GGAACAAGATCTACAGCATG	M51	40.48	1	NGG	Pol	3.320973722
TCAATCCCAACAAGGACACC	M52	58.84	1	NGG	Pol	15.29904885
GACGCCAACAAGGTAGGAGC	M53	36.32	1	NGG	Pol	15.91165565

[1636]

向导	gRNA	%终止编辑	向导链	PAM	基因	全部基因型保守%
TGCTCCAGCTCCTACCTTGT	M54	45.57	-1	NGG	Pol	16.16959536
CCACCAATCGCCAGACAGGA	M55	27.87	1	NGG	Pol	0.709334193
AGCCACCAGCAGGGAAATAC	M56	41.17	-1	NGG	Pol	16.73383847
ACCAGGACAAGTTGGAGGAC	M57	18.71	-1	NGG	Pol	16.21795905
CGATAACCAGGACAAGTTGG	M58	44.43	-1	NGG	Pol	18.2814767
CCCATCTCTTTGTTTTGTTA	M59	0.45	-1	NGG	Pol	4.933096889
CCCCATCTCTTTGTTTTGTT	M60	2.03	-1	NGG	Pol	5.223279059
TCAACGAATTGTGGGTCTTT	M61	5.2	1	NGG	Pol	23.48863453
CAACGAATTGTGGGTCTTTT	M62	1.6	1	NGG	Pol	24.81057553
GGTCTCCATGCGACGTGCAG	M63	27.5	-1	NGG	Pol	68.88602289
ACAGGCGGGGTTTTTCTTGT	M64	ND	1	NGA	S	86.8128325
GCAGAGTCTAGACTCGTGGT	M65	0.02	1	NGA	S	37.19168144
GAGAAGTCCACCACGAGTCT	M66	26.51	-1	NGA	S	97.16266323
GACACATCCAGCGATAACCA	M67	37.7	-1	NGA	S	54.97339997
AAGCCCAGGATGATGGATG	M68	45.01	-1	NGA	S	42.38271804
CCACTCCCATAGGAATTTTC	M69	1.1	-1	NGA	S	22.198936
CTGAGCCAGGAGAAACGGGC	M70	30.01	-1	NGA	S	22.10220861
TACCACATCATCCATATAAC	M71	7.13	-1	NGA	S	57.55279703

[1637]

向导	gRNA	%终止编辑	向导链	PAM	基因	全部基因型保守%
AGCCACCCAAGGCACAGCTT	M72	7.39	-1	NGA	核心区	94.32532645
CCAGCAAAGAATTGCTTGCC	M73	2.07	-1	NGA	核心区	9.576011607
GACGACGAGGCAGGTCCCCT	M74	20.83	1	NGA	核心区	52.34563921
GACGAGGCAGGTCCCCTAGA	M75	0.63	1	NGA	核心区	50.66903111
GGTCTCAATCGCCGCGTCGC	M76	23.22	1	NGA	核心区	69.57923585
CTCAATCGCCGCGTCGCAGA	M77	ND	1	NGA	核心区	91.76205062
AGTCCAAGGAATACTAACAT	M78	28.77	-1	NGA	Pol	32.01676608
TGTTTTCCAATGAGGATTAA	M79	22.75	-1	NGA	Pol	3.320973722
TTTCCACCAGCAATCCTCTG	M80	7.42	1	NGA	Pol	16.18571659
CCCAACAAGGACACCTGGCC	M81	2.41	1	NGA	Pol	15.17007899
ACGCCAACAAGGTAGGAGCT	M82	25.55	1	NGA	Pol	16.0889892
CCTCCACCAATCGCCAGACA	M83	8.63	1	NGA	Pol	0.693212961
CCAATCCTCGAGAAGATTGA	M84	0.12	-1	NGA	Pol	21.66693535
TCCCCAATCCTCGAGAAGAT	M85	24.44	-1	NGA	Pol	22.10220861
AGGGTCCCCAATCCTCGAGA	M86	45.43	-1	NGA	Pol	19.70014509
CTTCTCTCAATTTTCTAGGG	M87	21.49	1	NGA	Pol	82.52458488
CCAGGACAAGTTGGAGGACA	M88	15	-1	NGA	Pol	27.63179107
GATAACCAGGACAAGTTGGA	M89	50.24	-1	NGA	Pol	18.2814767

[1638]

向导	gRNA	%终止编辑	向导链	PAM	基因	全部基因型保守%
CCACCAGCACGGGACCATGC	M90	15.43	1	NGA	Pol	7.818797356
GCTGTACCAACCTTCGGAC	M91	29.59	1	NGA	Pol	14.23504756
CTCCATGCGACGTGCAGAGG	M92	26.43	-1	NGA	Pol	68.99887151
GTGGTCTCCATGCGACGTGC	M93	33.81	-1	NGA	Pol	68.4829921
TATGGATGATGTGGTATTGG	M94	30.62	1	NGG	Pol	67.22553603
TGCCAACTGGATCCTGCGCG	M189	0.23	1	NGA	X	72.48105755
GCTGCCAACTGGATCCTGCG	M190	36.53	1	NGG	X	76.28566823
CTGCCAACTGGATCCTGCGC	M191	41.43	1	NGG	X	72.49717878
CGCCCACCGAATGTTGCCCA	E95	45.29	1	NGG	X	
CTCCTCCCAGTCTTTAAACA	E96	0.01	1	NNNRRT	X	

[1639] 基因效率如下评估。对于质粒转染,将编码向导RNA的载体和编码碱基编辑器的载体瞬时转染到先前使用包含HBV基因组的慢病毒载体转导的HEK293T细胞中。所测试的碱基

编辑器包括:来自下列中的ABE、BE4(图4B)或BE4变体:

[1640] A3A-BE4表示BE4的一种变体,其中APOBEC-1被替换为APOBEC-3A的序列;APOBEC-3A(A3A)由例如Gehrke et al.,Nature Biotechnology volume 36,pages 977-982(2018)描述;

[1641] A3A-BE4-VRQR表示BE4的一种变体,其中APOBEC-1被替换为APOBEC-3A的序列,且Cas9被替换为包含V1134、R1217、Q1334和R1336的Cas9变体(称为SpCas9-VRQR);并且

[1642] BE4-VRQR表示BE4变体,其中Cas9被替换为包含V1134、R1217、Q1334和R1336的Cas9变体(称为SpCas9-VRQR)。

[1643] 与特异性向导合用的碱基编辑器的汇总提供在表13中。

[1644] 使用针对Hek293或HepG2-NTCP细胞优化的高效、低毒DNA转染试剂(Mirus TransIT293或Lipofectamine 2000)以3:1比率使用250ng的gRNA质粒和750ng的碱基编辑器质粒进行转染。对于mRNA转染,用转录的碱基编辑器mRNA和合成gRNA以3:1比率使用lipofectamin Messengermax对HepG2-NTCP细胞进行体外转染。在四天(对于质粒转染)和两天(对于RNA转染)后,用0.05% SDS、25μg/ml蛋白酶K和10mM Tris pH 8.0的简单裂解缓冲液萃取基因组DNA,然后在85℃加热孵育。基因组位点经PCR扩增并在MiSeq上测序。大约60%的所测试向导RNA将终止密码子引入HBV基因组中。在所测试的97种向导RNA中,12种跨全部HBV基因组具有大于20%的上靶编辑和大于15%的保守(图6、表15(M81和M189包括在该表中但用作阴性对照))。

[1645]

表 15：HBV 靶标汇总

疾病	基因符号	gRNA	编辑器	前间隔序列	PAM
HBV	S	M68	BE4-VRQR 或 A3A-BE4-VRQR	AAGCCCAGGATGATGGGATG	NGA
HBV	S	M67	BE4-VRQR 或 A3A-BE4-VRQR	GACACATCCAGCGATAACCA	NGA
HBV	S	M66	BE4-VRQR 或 A3A-BE4-VRQR	GAGAAGTCCACCACGAGTCT	NGA
HBV	Pol	M87	BE4-VRQR 或 A3A-BE4-VRQR	CTTCTCTCAATTTTCTAGGG	NGA
HBV	核心区	M74	BE4-VRQR 或 A3A-BE4-VRQR	GACGACGAGGCAGGTCCCCT	NGA
HBV	Pol	M81	BE4-VRQR 或 A3A-BE4-VRQR	CCCAACAAGGACACCTGGCC	NGA
HBV	X	M189	BE4-VRQR 或 A3A-BE4-VRQR	TGCCAACTGGATCCTGCGCG	NGA
HBV	Pol	M52	BE4 或 A3A-BE4	TCAATCCCAACAAGGACACC	NGG
HBV	Pol	M50	BE4 或 A3A-BE4	GGGAACAAGATCTACAGCAT	NGG
HBV	X	M191	BE4 或 A3A-BE4	CTGCCAACTGGATCCTGCGC	NGG
HBV	X	M190	BE4 或 A3A-BE4	GCTGCCAACTGGATCCTGCG	NGG
HBV	Pol	M94	ABE	TATGGATGATGTGGTATTGG	NGG
HBV	前核心区	M40	BE4 或 A3A-BE4	CCATGCCCCAAAGCCACCCA	NGG
HBV	前核心区	M39	BE4 或 A3A-BE4	AAGCCACCCAAGGCACAGCT	NGG

[1646] 这12种向导RNA中,M94将功能性改变突变引入HBV聚合酶的催化结构域中(D540G置换)。已知在该点处的突变抑制聚合酶的引发和延伸活性。

[1647] 实施例2:不同的碱基编辑器将改变引入HBV基因组中

[1648] 比较了不同碱基编辑器(即,BE4、BE4-VRQR、A3A-BE4-VRQR和ABE(TadA7.10))的效率和特异性。碱基编辑效率是指具有经编辑序列的测序读段数除以所分析的测序读段总数。碱基编辑特异性是指预期的碱基编辑相对于旁观者碱基编辑的比率。如上所述,将碱基

编辑器和向导RNA引入细胞内。具体地,使用BE4和A3ABE4碱基编辑器测试了12种具有至少20％上靶碱基编辑的向导RNA。BE4包含APOBEC胞苷脱氨酶、Cas9结构域和两个尿嘧啶DNA糖基化酶抑制剂(UGI)结构域,而A3A-BE4包含APOBEC3A胞苷脱氨酶、Cas9结构域和两个UGI结构域。参照图7,与包含APOBEC3A胞苷脱氨酶(A3ABE4)的碱基编辑器相比,BE4碱基编辑剂对于每种向导RNA都具有更佳的效率并且对于大多数向导RNA都具有更佳的特异性。

[1649] 将编码BE4碱基编辑器的DNA和RNA构建体与向导RNA一起引入HepG2-NTCP-lenti-HBV细胞系内。具体地,测试了DNA构建体、野生型RNA构建体和在N1部分处包含假尿嘧啶修饰的RNA构建体。测试了两种不同量(400ng和800ng)的经修饰的RNA构建体。每种构建体均产生碱基编辑,但包含假尿嘧啶的RNA构建体似乎表现出比DNA构建体或野生型RNA构建体更大的碱基编辑效率(图8和9)。

[1650] 使用BE4、BE4-VRQR和ABE碱基编辑器进一步测试向导RNA。将在N1部分处包含假尿嘧啶修饰并且编码碱基编辑器的RNA构建体与向导RNA一起引入HepG2-NTCP-lenti-HBV细胞系内。参照图10,每种碱基编辑器均表现出大于40％的功能性编辑。表16和17提供了BE4和ABE碱基编辑系统的潜在碱基编辑结果。

[1651]

表 16：针对被选择用于引入功能性变化或终止密码子的向导 RNA 的 BE4 碱基编辑的表征

gRNA的序列	PAM	向导Id	通过BE4碱基编辑器引入的突变	预期氨基酸改变的频率(跨经测序的HBV蛋白，根据Hbvdb.ibcp.fr的乙型肝炎病毒)	碱基编辑窗口内的胞苷数
AAGCCACCCAA GGCACAGCT	NGG	M39			4
CCATGCCCCAA AGCCACCCA	NGG	M40			3
GGGAACAAGAT CTACAGCAT	NGG	M50	Pol:Q177Stop	Pol:177*=0.00%，Pol:177Q=98.74%	1
TCAATCCCAAC AAGGACACC	NGG	M52	Pol:S216F，Pol:Q217Stop	Pol:216F=0.01%，Pol:216S=22.91%，Pol:217*=0.00%，Pol:217Q=56.79%	3
GCTGCCAACTG GATCCTGCG	NGG	M190	X:Q8Stop，Pol:A757V	Pol:757A=99.91%，Pol:757V=0.00%，X:8*=0.00%，X:8Q=99.25%	2
CTGCCAACTGG ATCCTGCGC	NGG	M191	X:Q8Stop，Pol:A757V	Pol:757A=99.91%，Pol:757V=0.00%，X:8*=0.00%，X:8Q=99.25%	3
GACGACGAGGC AGGTCCCCT	NGA	M74	核心区:R152Stop	核心区:152*=0.00%，核心区:152R=99.59%	1
CTTCTCTCAAT TTTCTAGGG	NGA	M87	S:S38F，S:L39F，Pol:S382F，Pol:Q383Stop	Pol:382F=0.05%，Pol:382S=99.86%，Pol:383*=0.00%，Pol:383Q=99.93%，S:38F=0.01%，S:38S=99.93%，S:39F=0.04%，S:39L=99.84%	3

[1652]

GAGAAGTCCAC CACGAGTCT	NGA	M66	S:W36Stop，Pol:D380N	Pol:380D=99.88% ， Pol:380N=0.01% ， S:36*=0.00%，S:36W=99.55%	1
GACACATCCAG CGATAACCA	NGA	M67	S:W74Stop，S:M75I，Pol:D418N，Pol:V419M	Pol:418D=99.93% ， Pol:418N=0.01% ， Pol:419M=0.05% ， Pol:419V=99.85% ， S:74*=0.00%，S:74W=99.15%，S:75I=0.07%，S:75M=99.79%	2
AAGCCCAGGAT GATGGGATG	NGA	M68	S:W156Stop，S:A157T，Pol:G500N	Pol:500G=99.91% ， Pol:500N=0.00% ， S:156*=0.00% ， S:156W=99.65% ， S:157A=99.83%，S:157T=0.02%	3
CCCAACAAGGA CACCTGGCC	NGA	M81	Pol:Q218Stop	Pol:218*=0.00%，Pol:218Q=99.00%	1
TGCCAACTGGA TCCTGCGCG	NGA	M189	X:Q8Stop	X:8*=0.00%，X:8Q=99.25%	2
TATGGATGATG TGGTATTGG	NGG	M94			0

[1653] 表17:针对被选择用于引入功能性变化或终止密码子的向导RNA的ABE碱基编辑的表征

[1654]

向导Id	通过BE4碱基编辑器引入的突变	等位基因频率	碱基编辑窗口内的胞苷编号
M39			1
M40			0
M50	Pol:E176G，Pol:Q177R	Pol:176E=99.94%，Pol:176G=0.02%，Pol:177Q=98.74%，Pol:177R=0.05%	4
M52			1
M190	X:Q8R，Pol:N758G	Pol:758G=0.00%，Pol:758N=99.78%，X:8Q=99.25%，X:8R=0.07%	2
M191	X:Q8R，Pol:N758G	Pol:758G=0.00%，Pol:758N=99.78%，X:8Q=99.25%，X:8R=0.07%	2
M74	Pol:D16G，Pol:E17G	Pol:16D=69.15%，Pol:16G=0.12%，Pol:17E=94.58%，Pol:17G=0.59%	2
M87			0
M66	S:S38P，Pol:F381P	Pol:381F=99.93%，Pol:381P=0.00%，S:38P=0.04%，S:38S=99.93%	2
M67	S:M75T，S:C76R，Pol:V419A	Pol:419A=0.06%，Pol:419V=99.85%，S:75M=99.79%，S:75T=0.06%，S:76C=96.88%，S:76R=0.04%	2
M68	S:W156R，Pol:L499P	Pol:499L=95.19%，Pol:499P=0.01%，S:156R=0.02%，S:156W=99.65%	1
M81	Pol:Q217R，Pol:Q218R	Pol:217Q=56.79%，Pol:217R=0.72%，	4

[1655]

		Pol:218Q=99.00%，Pol:218R=0.01%	
M189	X:Q8R，Pol:N758G	Pol:758G=0.00%，Pol:758N=99.78%，X:8Q=99.25%，X:8R=0.07%	2
M94	S:M197V，Pol:D540G	Pol:540D=99.94%，Pol:540G=0.01%，S:197M=99.59%，S:197V=0.01%	1

[1656] 实施例3:使用碱基编辑器来靶向HBV基因组的保守区域。

[1657] 评估了碱基编辑器编辑器HBV基因组的保守区域的能力(图11)。表18鉴定了126种RNA,这些RNA靶向在>80％的HBV A和D基因型中保守(“超保守”)的区域并且被设计为最小化脱靶效应(即,靶向人类基因组区域)。预期这些向导RNA基于在HBV基因组中的高度保守而引入功能性改变。不受限于理论,使用碱基编辑器(包括不含UGI的碱基编辑器)将脱氨酶靶向至HBV cccDNA,具有促使形成无嘧啶位点和cccDNA损伤和/或降解的潜力。表18提供向导RNA的序列的列表,该向导RNA将碱基编辑器靶向HBV基因组的保守区域。这些向导RNA所靶向的序列在HBV基因型A、B、C和D之间具有大于50％的保守度并且具有至少一个与人类基因组hg38中最相似序列的错配,从而最小化脱靶效应。

[1658] 表18:保守的gRNA列表

[1659]

向导序列	向导id	PAM
AAAAACCCCGCCTGTAACAC	新型	NAG
AAAACGCCGCAGACACATCC	新型	NGC
AAAGGCCTTGTAAGTTGGCG	新型	NGA
AAAGGCCTTGTAAGTTGGCGA	新型	NNNRRT
AACCACTGAACAAATGGCAC	新型	NAG
AACCCCGCCTGTAACACGAG	新型	NAG
AACGCCGCAGACACATCCAG	新型	NGA
AAGAAGTCAGAAGGCAAAAA	新型	NGA
AAGCCACCCAAGGCACAGCT	MSPbeam39	NGG
AAGCCCTACGAACCACTGAA	新型	NAA
AAGCCTCCAAGCTGTGCCTT	新型	NGG
AAGGCACAGCTTGGAGGCTT	新型	NAA
AAGGCACAGCTTGGAGGCTTG	EMSbeam1	NNNRRT
AAGGCCTTGTAAGTTGGCGA	新型	NAA
AATCGCCGCGTCGCAGAAGAT	新型	NNNRRT
AATGTCAACGACCGACCTTG	新型	NGG
ACAGGCGGGGTTTTTCTTGT	MSPbeam64	NGA
ACCAATTTATGCCTACAGCCT	EMSbeam2	NNNRRT
ACGAACCACTGAACAAATGGC	MSPbeam5	NNNRRT
ACGGGGCGCACCTCTCTTTA	新型	NGC
ACTCCCTCGCCTCGCAGACG	新型	NAG
ACTGAACAAATGGCACTAGT	新型	NAA
ACTTCTCTCAATTTTCTAGG	新型	NGG
ACTTTCTCGCCAACTTACAA	新型	NGC
AGAAAGGCCTTGTAAGTTGG	新型	NGA
AGAAGAACTCCCTCGCCTCG	新型	NAG
AGACAAAAGAAAATTGGTAA	新型	NAG
AGCCACCCAAGGCACAGCTT	MSPbeam72	NGA

[1660]

向导序列	向导id	PAM
AGCCCTACGAACCACTGAAC	新型	NAA
AGCTGTGCCTTGGGTGGCTT	新型	NGG
AGGAGGCTGTAGGCATAAAT	EMSbeam3	NGG
AGGAGTTCCGCAGTATGGAT	EMSbeam4	NGG
AGGCAGGTCCCCTAGAAGAA	新型	NAA
AGGCCTTGTAAGTTGGCGAG	新型	NAA
AGGCGAGGGAGTTCTTCTTC	新型	NAG
AGTCCAAGAGTCCTCTTATG	新型	NAA
AGTTCCGCAGTATGGATCGG	新型	NAG
AGTTCTTCTTCTAGGGGACC	新型	NGC
ATCCATACTGCGGAACTCCT	新型	NGC
ATCTTCTGCGACGCGGCGAT	新型	NGA
ATGTCAACGACCGACCTTGA	新型	NGC
ATTTGTTCAGTGGTTCGTAG	新型	NGC
CAAATGGCACTAGTAAACTG	新型	NGC
CAAGCCTCCAAGCTGTGCCT	EMSbeam5	NGG
CAAGGCACAGCTTGGAGGCT	EMSbeam6	NGA
CACCACGAGTCTAGACTCTG	MSPbeam36	NGG
CACCCAAGGCACAGCTTGGA	新型	NGC
CACTGAACAAATGGCACTAG	新型	NAA
CATACTGCGGAACTCCTAGC	新型	NGC
CATGCAACTTTTTCACCTCTG	MSPbeam8	NNNRRT
CATGCCCCAAAGCCACCCAA	新型	NGC
CCACCCAAGGCACAGCTTGG	MSPbeam38	NGG
CCATGCAACTTTTTCACCTC	新型	NGC
CCATGCCCCAAAGCCACCCA	MSPbeam40	NGG
CCCAAAGCCACCCAAGGCAC	新型	NGC
CCCCAAAGCCACCCAAGGCA	新型	NAG
CCCGTCTGTGCCTTCTCATC	新型	NGC

[1661]

向导序列	向导id	PAM
CCGCAGTATGGATCGGCAGA	EMSbeam7	NGA
CCTACGAACCACTGAACAAA	新型	NGG
CCTCCAAGCTGTGCCTTGGG	新型	NGG
CCTCTGCCGATCCATACTGC	EMSbeam8	NGA
CCTGCTGGTGGCTCCAGTTC	新型	NGG
CGCCGCGTCGCAGAAGATCT	新型	NAA
CGCGTAAAGAGAGGTGCGCCC	新型	NNNRRT
CGTGCAGAGGTGAAGCGAAG	新型	NGC
CTACGAACCACTGAACAAAT	新型	NGC
CTAGACTCGTGGTGGACTTCT	EMSbeam9	NNNRRT
CTCAATCGCCGCGTCGCAGA	MSPbeam77	NGA
CTCCAAGCTGTGCCTTGGGT	新型	NGC
CTCTGCCGATCCATACTGCG	新型	NAA
CTGTGCCTTGGGTGGCTTTG	新型	NGG
CTGTTCAAGCCTCCAAGCTG	新型	NGC
CTTCTCTCAATTTTCTAGGG	MSPbeam87	NGA
CTTCTGCGACGCGGCGATTG	新型	NGA
GAAAAACCCCGCCTGTAACA	新型	NGA
GAAAGCCCTACGAACCACTGA	新型	NNNRRT
GAACTCCCTCGCCTCGCAGAC	EMSbeam10	NNNRRT
GAACTGGAGCCACCAGCAGG	新型	NAA
GAAGAACTCCCTCGCCTCGC	EMSbeam11	NGA
GACAAACGGGCAACATACCTT	新型	NNNRRT
GACTCGTGGTGGACTTCTCT	新型	NAA
GACTTCTCTCAATTTTCTAG	EMSbeam12	NGG
GAGAAGTCCACCACGAGTCT	MSPbeam66	NGA
GAGGACAAACGGGCAACATAC	新型	NNNRRT
GAGGCAGGTCCCCTAGAAGA	新型	NGA
GATCCATACTGCGGAACTCC	新型	NAG

[1662]

向导序列	向导id	PAM
GATTGAGATCTTCTGCGACGC	新型	NNNRRT
GCAACTTTTTCACCTCTGCC	新型	NAA
GCACAGCTTGGAGGCTTGAA	新型	NAG
GCCACCCAAGGCACAGCTTG	新型	NAG
GCGTCAGCAAACACTTGGCA	新型	NAG
GCTGCTATGCCTCATCTTCTT	EMSbeam14	NNNRRT
GCTGTGCCTTGGGTGGCTTT	新型	NGG
GGAAAGCCCTACGAACCACT	新型	NAA
GGACTTCTCTCAATTTTCTA	EMSbeam15	NGG
GGAGTGCGAATCCACACTCC	新型	NAA
GGAGTTCCGCAGTATGGATC	新型	NGC
GGCACTAGTAAACTGAGCCA	新型	NGA
GGCCTTGTAAGTTGGCGAGA	新型	NAG
GGCGGGGTTTTTCTTGTTGAC	新型	NNGRRT
GGCGGGGTTTTTCTTGTTGAC	新型	NNNRRT
GGCGTTCACGGTGGTCTCCA	新型	NGC
GGGAAAGCCCTACGAACCAC	新型	NGA
GGGACTGCGAATTTTGGCCA	新型	NGA
GGGGCGCACCTCTCTTTACG	新型	NGG
GGGTTGCGTCAGCAAACACT	新型	NGG
GGTCACCATATTCTTGGGAA	新型	NAA
GGTGGAGCCCTCAGGCTCAG	新型	NGC
GGTTGCGTCAGCAAACACTT	新型	NGC
GTCACCATATTCTTGGGAAC	新型	NAG
GTCCACCACGAGTCTAGACTC	EMSbeam16/MSPbeam2	NNNRRT
GTCCATGCCCCAAAGCCACC	新型	NAA
GTCCCGTCGGCGCTGAATCC	新型	NGC
GTCGCAGAAGATCTCAATCT	新型	NGG

[1663]

向导序列	向导id	PAM
GTCTGTGCCTTCTCATCTGC	新型	NGG
GTGGACTTCTCTCAATTTTC	新型	NAG
GTTCCGCAGTATGGATCGGC	EMSbeam17	NGA
GTTCCGCAGTATGGATCGGC	新型	NNAGGA
GTTTACTAGTGCCATTTGTT	新型	NAG
GTTTACTAGTGCCATTTGTTC	新型	NNNRRT
TAAAACGCCGCAGACACATC	新型	NAG
TAAAACGCCGCAGACACATCC	EMSbeam18	NNNRRT
TAGGCAGAGGTGAAAAAGTTG	新型	NNNRRT
TCACCATATTCTTGGGAACA	新型	NGA
TCAGTTTACTAGTGCCATTTG	新型	NNNRRT
TCCATGCCCCAAAGCCACCC	新型	NAG
TCCCCCTAGAAAATTGAGAG	新型	NAG
TCCGCAGTATGGATCGGCAG	EMSbeam19	NGG
TCCTCTGCCGATCCATACTG	EMSbeam20	NGG
TCGCAGAAGATCTCAATCTC	新型	NGG
TCTCAATCGCCGCGTCGCAG	新型	NAG
TCTTCTGCGACGCGGCGATT	新型	NAG
TCTTGTTCCCAAGAATATGG	新型	NGA
TGAGATCTTCTGCGACGCGG	新型	NGA
TGAGCCTGAGGGCTCCACCC	新型	NAA
TGAGGCATAGCAGCAGGATG	新型	NAG
TGGACTTCTCTCAATTTTCT	EMSbeam21	NGG
TGGCCAAAATTCGCAGTCCC	新型	NAA
TGGCTTTGGGGCATGGACAT	新型	NGA
TGTGCACTTCGCTTCACCTC	新型	NGC
TGTGCCTTGGGTGGCTTTGG	新型	NGC
TTAGGCAGAGGTGAAAAAGT	新型	NGC
TTCAAGCCTCCAAGCTGTGCC	EMSbeam22/MS	NNGRRT

[1664]

向导序列	向导id	PAM
	Pbeam9	
TTCAAGCCTCCAAGCTGTGCC	EMSbeam22	NNNRRT
TTCCCGAGATTGAGATCTTC	新型	NGC
TTCCGCAGTATGGATCGGCA	新型	NAG
TTCGCTTCACCTCTGCACGT	新型	NGC
TTTGCTGACGCAACCCCCAC	EMSbeam23_whb	NGG

[1665] 在表18A中描述所测试的向导核酸(也参见图11)。

[1666] 表18A:向导核酸

[1667]

gRNA	序列	PAM	CAS	所靶向的区域
E1	AAGGCACAGCTTGGAGGCTTG	NNNRRT	SaCas9	核心区
E2	ACCAATTTATGCCTACAGCCT	NNNRRT	SaCas9	X
E3	AGGAGGCTGTAGGCATAAAT	NGG	SpCas9	X
E4	AGGAGTTCC GCAGTATGGAT	NGG	SpCas9	Pol
E5	CAAGCCTCCAAGCTGTGCCT	NGG	SpCas9	核心区
E6	CAAGGCACAGCTTGGAGGCT	NGA	SpCas9	核心区
E7	CCGCAGTATGGATCGGCAGA	NGA	SpCas9	Pol
E8	CCTCTGCCGATCCATACTGC	NGA	SpCas9	Pol
E9	CTAGACTCGTGGTGGACTTCT	NNNRRT	SaCas9	Pol，S
E10	GAACTCCCTCGCCTCGCAGAC	NNNRRT	SaCas9	Pol，核心区
E11	GAAGAACTCCCTCGCCTCGC	NGA	SpCas9	Pol，核心区
E12	GACTTCTCTCAATTTTCTAG	NGG	SpCas9	Pol，S
E13 (M66)	GAGAAGTCCACCACGAGTCT	NGA	SpCas9	Pol，S
E14	GCTGCTATGCCTCATCTTCTT	NNNRRT	SaCas9	Pol，S
E15	GGACTTCTCTCAATTTTCTA	NGG	SpCas9	Pol，S
E16	GTCCACCACGAGTCTAGACTC	NNNRRT	SaCas9	Pol，S
E17	GTTCCGCAGTATGGATCGGC	NGA	SpCas9	Pol
E18	TAAAACGCCGCAGACACATCC	NNNRRT	SaCas9	Pol，S
E19	TCCGCAGTATGGATCGGCAG	NGG	SpCas9	Pol
E20	TCCTCTGCCGATCCATACTG	NGG	SpCas9	Pol
E21	TGGACTTCTCTCAATTTTCT	NGG	SpCas9	Pol，S
E22	TTCAAGCCTCCAAGCTGTGCC	NNGRRT	SaCas9	核心区
E22	TTCAAGCCTCCAAGCTGTGCC	NNNRRT	SaCas9	核心区
E23	TTTGCTGACGCAACCCCCAC	NGG	SpCas9	Pol
E24	TACTAACATTGAGGTTCCCG	NGA	SpCas9	Pol，核心区

[1668] 在用向导RNA和碱基编辑器(pBtx448(BE4)、pBtx543(VRQR-BE4)或pBTx546(sa-Cas9))转染前一天,将50,000个HBV感染的细胞(pBtx693:用编码X和核心区HBV基因的HBV序列转导的细胞系;和pBtx536:用用编码Pol和S HBV基因的HBV序列转导的细胞系)铺板。用250ng的向导RNA和750ng的编码碱基编辑器的构建体转染细胞。在转染后1天和3天更换培养基。从收获的细胞中分离DNA,并通过PCR使用10μM引物扩增经编辑的区域。表中的每种细胞鉴定了扩增反应中使用的引物(例如,1058/ES86)、向导RNA(例如,E1)和碱基编辑器

(例如,546)。PCR反应的条件如下:初始步骤:98℃,30秒;然后是30个循环:98℃10秒,60℃20秒和72℃20秒;然后是最终步骤:72℃,5分钟。

[1669] 后续对扩增产物进行测序以鉴定向导RNA所靶向的哪些核苷酸被编辑(即,编辑的特异性)以及编辑的效率。表19是针对表20中所列的向导RNA中每一种的碱基编辑效率和特异性的汇总。

[1670] 表19:碱基编辑效率和特异性

[1671]

gRNA	所编辑的碱基位置(粗体显示的行) 在每个位置处的编辑%(下方的行)							平均最高的编辑%
E1	**5**	**7**	**10**					10.37
	0.24	17.23	4.1					
	0.31	3.5	0.63					
E2	**2**	**3**	**13**	**16**				24.18
	0	0	23.17	4.32				
	2.49	4.12	33.23	5.29				
	0.62	1.14	16.15	2.5				
E3	**7**							1.31
	2.62							
	0							
E4		**8**	**9**	**11**				54.23
		59	55.25	1.55				
		47.73	45.42	0.45				
		55.97	51.17	1.73				
E5	**1**	**5**	**6**	**8**	**9**			22.89
	1.35	2.17	10.25	11.75	9.78			
	9.53	7.19	29.63	35.37	28.98			
	2.64	11.46	28.78	32.25	28.56			
E6		**6**	**8**					36.04
		39.98	35.13					
		28.63	39.09					
		21.92	33.91					
E7		**1**	**2**	**4**	**14**			51.29
		46.08	34.59	8.9	28.72			
		56.62	44.53	10.5	37.31			
		51.17	41.05	9.7	33.93			
E8		**4**	**7**	**8**	**12**			44.72
		44.54	13.45	13.85	13.88			
		42.28	8.85	10.45	13.28			
		47.35	14.53	17.45	17.51			
E9		**1**	**6**	**8**				4.65

[1672]

		3.64	2.63	5.87				
		2.24	1.47	4.73				
		1.64	1.29	3.35				
E10	**4**	**6**	**7**	**8**	**10**	**12**	**13**	38.72
	1.01	29.15	19.25	18.69	42.17	2.12	9.14	
	1.25	24.05	15.7	15.43	31.85	2.71	8.28	
	1.36	29.67	19.88	20.05	42.15	2.06	8.01	
E11		**7**	**9**	**10**	**11**	**13**		11.79
		9.26	13.19	5.05	3.85	4.67		
		6.07	10.83	4.07	2.53	3.4		
		8.3	11.35	4.63	3.67	4.58		
E12		**3**	**6**	**8**	**10**			43.91
		0.33	32.61	36.17	8.22			
		0.13	9.03	9.33	2.49			
		0.17	11.97	86.22	2.82			
E13 (M66)		**8**	**9**	**11**	**12**			19.47
		20.53	8.47	2.31	0.75			
		26.03	9.21	1.99	0.54			
		11.84	4.79	1.52	0.42			
E14		**5**	**10**	**11**	**13**	**16**		7.54
		0.01	0.33	2.24	5.16	4.92		
		0	0.67	3.17	7.04	6.61		
		0.01	1.3	4.84	10.41	10.3		
E15		**4**	**7**	**9**	**11**			5
		1.6	4.3	3.76	2.14			
		1.79	4.57	3.52	2.2			
		2.73	6.13	5.33	3.61			
E16		**3**	**4**	**6**	**7**	**9**	**14**	28.76
		19.79	5.81	7.58	10.3	26.74	44.32	
		20.55	7.4	7.12	7.99	25.89	38.01	
		1.99	0.91	1.05	1.19	2.63	3.95	
E17		**4**	**5**	**7**	**17**			
		57.01	55.99	13.37	1.97			57.71

[1673]

		55.63	51.62	12.52	4.73			
		60.48	57.62	17.93	3.82			
E18		**6**	**8**	**9**	**11**			
		1.68	0.52	1.52	0.34			10.22
		16.99	5.23	15.96	4.49			
		11.99	3.59	10.91	3.6			
E19		**2**	**3**	**5**				
		53.21	49.12	24.46				44.63
		38.64	35.38	18.57				
		42.03	38.94	23.8				
E20		**2**	**3**	**5**	**8**	**9**	**13**	
		5.05	5.88	52.37	19.33	19.03	16.47	50.74
		4.83	5.64	50.42	18.55	21.86	21.38	
		5.73	7.13	49.43	13.98	16.3	21.61	
E21		**5**	**8**	**10**	**12**			15.40
		6.93	8.98	5.85	2.7			
		16.16	20.1	13.3	5.95			
		13.32	17.13	12.13	5.67			
E22		**3**	**7**	**8**	**10**	**11**		0.01
		0	0	0	0	0		
		0.02	0	0.01	0.01	0.02		
		0.02	0.01	0	0.02	0.01		
E23		**5**	**9**					6.35
		3.61	1.83					
		8.62	3.59					
		6.83	3.31					
E24		**3**	**7**	**17**				47.51
		37.98	54.1	6.99				
		34.43	49.5	6.77				
		24.76	38.92	7.49				
E95		**3**	**4**	**5**	**7**	**8**		45.29
		3.44	11.92	14.63	16.61	16.01		
		7.71	33.79	46.19	56.45	52.91		
		15.11	48.31	62.82	63.65	66.96		

[1674]

E96		3	4	6	7	8		0.01
		0	0	0	0	0		
		0.02	0.01	0	0.01	0		
		0	0	0	0	0		

[1675] 表19和20鉴定氨基酸中的置换和它们的频率,这将是使用被设计成靶向HBV基因组的超保守区域的向导RNA进行的BE4和ABE介导的编辑的结果。

[1676]

表 20：对使用被选择用于靶向 HBV 基因组的保守区域的特定向导 RNA 进行的 BE4 碱基编辑所导致的氨基酸置换的计算机分析

向导序列	PAM	向导Id	通过BE4碱基编辑器引入的突变	氨基酸置换频率	每个窗口中的C数
GTTCCGCAGTATGGATCGGC	NGA	E17	Pol:A717T，Pol:E718K	Pol:717A=99.79%，Pol:717T=0.02%，Pol:718E=99.81%，Pol:718K=0.05%	3
AGGAGTTCCGCAGTATGGAT	NGG	E4	Pol:E718K	Pol:718E=99.81%，Pol:718K=0.05%	1
CCGCAGTATGGATCGGCAGA	NGA	E7	Pol:A717T	Pol:717A=99.79%，Pol:717T=0.02%	1
TCCTCTGCCGATCCATACTG	NGG	E20	Pol:P713S	Pol:713P=99.78%，Pol:713S=0.01%	2
TACTAACATTGAGGTTCCCG	NGA	E24_剪接(不是保守的)	核心区:C183Y，Pol:V48I	核心区:183C=98.81%，核心区:183Y=0.02%，Pol:48I=0.01%，Pol:48V=99.47%	1
CGCCCACCGAATGTTGCCCA	NGG	E95(X-终止，不是保守的)	X:H86Y，X:R87Stop	X:86H=81.50%，X:86Y=0.08%，X:87*=0.00%，X:87R=22.91%	4
CCTCTGCCGATCCATACTGC	NGA	E8	Pol:P713L	Pol:713L=0.06%，Pol:713P=99.78%	3
TCCGCAGTATGGATCGGCAG	NGG	E19	Pol:A717T	Pol:717A=99.79%，Pol:717T=0.02%	1
GACTTCTCTCAATTTTCTAG	NGG	E12	S:S38F，S:L39F，Pol:S382F	Pol:382F=0.05%,Pol:382S=99.86%,S:38F=0.01%，S:38S=99.93%，S:39F=0.04%，S:39L=99.84%	2

[1677]

GAACTCCCTCGCCTCGCAGAC	NNNRRT	E10	核心区:T160I，核心区 :P161F，核心区:S162L，Pol:L25F，Pol:P26F，Pol:R27C	核心区:160I=0.02%，核心区:160T=99.81%，核心区 :161F=0.00%，核心区 :161P=99.82%，核心区 :162L=0.03%，核心区 :162S=99.85%，Pol:25F=0.01%,Pol:25L=99.84%,Pol:26F=0.00%,Pol:26P=99.88%,Pol:27C=0.02%,Pol:27R=98.51%	5
CAAGGCACAGCTTGGAGGCT	NGA	E6			2
GTCCACCACGAGTCTAGACTC	NNNRRT	E16/2	S:W35Stop，S:W36Stop，Pol:V378I，Pol:V379I，Pol:D380N	Pol:378I=0.00%，Pol:378V=99.80%，Pol:379I=0.00%，Pol:379V=99.86%，Pol:380D=99.88%,Pol:380N=0.01%,S:35*=0.00%,S:35W=99.77%，S:36*=0.00%，S:36W=99.55%	5
ACCAATTTATGCCTACAGCCT	NNNRRT	E2			1
CAAGCCTCCAAGCTGTGCCT	NGG	E5			3
GAGAAGTCCACCACGAGTCT	NGA	M66/E13	S:W36Stop，Pol:D380N	Pol:380D=99.88%,Pol:380N=0.01%,S:36*=0.00%,S:36W=99.55%	1
TGGACTTCTCTCAATTTTCT	NGG	E21	S:T37I，S:S38F	S:37I=0.02%，S:37T=99.87%，S:38F=0.01%，S:38S=99.93%	2
GAAGAACTCCCTCGCCTCGC	NGA	E11	核心区:T160I，Pol:L25F	核心区:160I=0.02%，核心区:160T=99.81%，Pol:25F=0.01%，Pol:25L=99.84%	1
AAGGCACAGCTTGGAGGCTTG	NNNRRT	E1			3

[1678]

TAAAACGCCGCAGACAC ATCC	NNNRRT	E18	S:R78Q , S:R79H , Pol:A422T	Pol:422A=99.91%,Pol:422T=0.01%,S:78Q=0.03%, S:78R=99.91%, S:79H=0.36%, S:79R=99.38%	3
GCTGCTATGCCTCATCT TCTT	NNNRRT	E14	Pol:A432V, Pol:P434S	Pol:432A=99.51% , Pol:432V=0.04% , Pol:434P=99.89%, Pol:434S=0.06%	2
TTTGCTGACGCAACCCC CAC	NGG	E23	Pol:A688V	Pol:688A=99.75%, Pol:688V=0.01%	1
GGACTTCTCTCAATTTT CTA	NGG	E15	S:T37I, S:S38F	S:37I=0.02% , S:37T=99.87% , S:38F=0.01% , S:38S=99.93%	2
CTAGACTCGTGGTGGAC TTCT	NNNRRT	E9	S:S34L, Pol:L377F	Pol:377F=0.01%,Pol:377L=99.91%,S:34L=0.41%, S:34S=99.40%	2
AGGAGGCTGTAGGCATA AAT	NGG	E3			1
TTCAAGCCTCCAAGCTG TGCC	NNGRRT	E22/9			4
ACTCCTCCCAGTCTTTA AACA	NNNRRT	E96	X:W120Stop, X:E121K, X:E122K	X:120*=0.00%, X:120W=99.79%, X:121E=99.82%, X:121K=0.04%, X:122E=99.13%, X:122K=0.05%	5

[1679]

表 21：对使用被选择用于靶向 HBV 基因组的保守区域的特定向导 RNA 进行的 ABE 碱基编辑所导致的氨基酸置换的计算机分析

向导ID	通过ABE碱基编辑器引入的突变	等位基因频率	最小等位基因频率	每个窗口中的A数
E17			100	1
E4	Pol:L719P	Pol:719L=99.80%，Pol:719P=0.08%	0.082362631	1
E7			100	2
E20			100	0
E24_剪接(不是保守的)	核心区:C183R，核心区:Stop184Q，Pol:V48A	核心区:183C=98.81%，核心区:183R=0.07%，Pol:48A=0.22%，Pol:48V=99.47%	0.073421439	3
E95(X-终止，不是保守的)	X:H86R	X:86H=81.50%，X:86R=11.10%	11.09879032	1
E8			100	0
E19			100	1
E12			100	0
E10	核心区:T160A	核心区:160A=0.04%，核心	0.03671072	1

[1680]

		区:160T=99.81%		
E6			100	1
E16/2	S:W35R，S:W36R，Pol:V378A，Pol:V379A	Pol:378A=0.09%，Pol:378V=99.80%，Pol:379A=0.07%，Pol:379V=99.86%，S:35R=0.07%，S:35W=99.77%，S:36R=0.10%，S:36W=99.55%	0.070596541	2
E2	X:L141P	X:141L=99.53%，X:141P=0.01%	0.010080645	3
E5			100	0
M66/E13	S:S38P，Pol:F381P	Pol:381F=99.93%，Pol:381P=0.00%，S:38P=0.04%，S:38S=99.93%	0	2
E21	S:T37A，Pol:D380G	Pol:380D=99.88%，Pol:380G=0.04%，S:37A=0.06%，S:37T=99.87%	0.03529827	1
E11	核心区:T160A，Pol:E24G	核心区:160A=0.04%，核心区:160T=99.81%，Pol:24E=99.75%，Pol:24G=0.11%	0.03671072	2

[1681]

E1			100	2
E18	S:F80P，Pol:F423P	Pol:423F=99.84%，Pol:423P=0.00%，S:80F=99.26%，S:80P=0.00%	0	3
E14	Pol:M433V	Pol:433M=99.88%，Pol:433V=0.02%	0.02353218	1
E23_whb	Pol:D689G	Pol:689D=99.87%，Pol:689G=0.02%	0.02353218	1
E15			100	0
E9	S:D33G，Pol:R376G	Pol:376G=0.02%，Pol:376R=99.85%，S:33D=99.75%，S:33G=0.08%	0.02353218	2
E3			100	1
E22/9			100	2
E96	X:W120R	X:120R=0.13%，X:120W=99.79%	0.131048387	1
			100	2

[1682] 实施例4:设想额外的gRNA靶向HBV

[1683] 表22提供了向导RNA所靶向的序列的列表,该向导RNA被设计为使用ABE碱基编辑

器将错义突变引入HBV基因组中。该向导RNA所靶向的序列在HBV基因型A和D之间至少50％保守。对应用gRNA和ABE时将会导致的氨基酸置换进行计算机分析，并且我们进一步选择那些将会产生在0.05％的已知测序的HBV基因中出现的氨基酸置换的gRNA。那将意味着使用所选择的gRNA进行碱基编辑将会导致HBV蛋白功能失调。

[1684] 表22:ABE gRNA功能列表

[1685]

向导序列	向导 id	PAM
AAAAACCCCGCCTGTAACAC	新型	NAG
AAAAAGTTGCATGGTGCTGG	新型	NGC
AAAACAAGCGGCTAGGAGTTC	新型	NNNRRT
AAAACGCCGCAGACACATCC	新型	NGC
AAACAAGCGGCTAGGAGTTC	新型	NGC
AAAGAATTTGGAGCTACTGT	新型	NGA
AAAGCCAAACAGTGGGGGAA	新型	NGC
AAATTGAGAGAAGTCCACCA	新型	NGA
AACAAATGGCACTAGTAAAC	新型	NGA
AACAAGAAAAACCCCGCCTG	新型	NAA
AACAGTGGGGGAAAGCCCTA	新型	NGA
AACATCACATCAGGATTCCT	新型	NGG
AACATGGAGAACATCACATC	新型	NGG
AACCACTGAACAAATGGCAC	新型	NAG
AACCCCCACTGGCTGGGGCT	新型	NGG
AAGAAGATGAGGCATAGCAG	新型	NAG
AAGAAGTCAGAAGGCAAAAA	新型	NGA
AAGAAGTCAGAAGGCAAAAAC	新型	NNGRRT
AAGAAGTCAGAAGGCAAAAAC	新型	NNNRRT
AAGAATCCTCACAATACCGC	新型	NGA
AAGAATTTGGAGCTACTGTG	新型	NAG
AAGGAAAGAAGTCAGAAGGC	新型	NAA
AATGTCAACGACCGACCTTG	新型	NGG
AATTGAGAGAAGTCCACCAC	新型	NAG
ACAAATGGCACTAGTAAACT	新型	NAG
ACAAGAATCCTCACAATACCG	新型	NNGRRT
ACAAGAATCCTCACAATACCG	新型	NNNRRT
ACACATCCAGCGATAACCAGG	M3	NNNRRT

[1686]

向导序列	向导 id	PAM
ACAGGAGGTTGGTGAGTGAT	新型	NGG
ACAGGAGGTTGGTGAGTGATT	新型	NNNRRT
ACAGTGGGGGAAAGCCCTAC	新型	NNACCA
ACATGGAGAACATCACATCA	新型	NGA
ACATTTAAACCCTAACAAAA	新型	NAA
ACCAATTTATGCCTACAGCCT	E2	NNNRRT
ACGAACCACTGAACAAATGGC	M5	NNNRRT
ACGCAACCCCCACTGGCTGG	新型	NGC
ACTGAACAAATGGCACTAGT	新型	NAA
AGAAAGGCCTTGTAAGTTGG	新型	NGA
AGAACATCACATCAGGATTC	新型	NNAGGA
AGAAGAACTCCCTCGCCTCG	新型	NAG
AGAAGATGAGGCATAGCAGC	新型	NGG
AGAAGGGGACGAGAGAGTCC	新型	NAA
AGACAAAAGAAAATTGGTAA	新型	NAG
AGACAAAAGAAAATTGGTAAC	新型	NNNRRT
AGACACATCCAGCGATAACC	新型	NGG
AGACGGAGAAGGGGACGAGA	新型	NAG
AGAGAAGTCCACCACGAGTC	新型	NAG
AGAGAGGTGCGCCCCGTGGT	新型	NGG
AGATGAGAAGGCACAGACGG	新型	NGA
AGCGCCGACGGGACGTAAAC	新型	NAA
AGCGCCGACGGGACGTAAAC	新型	NNAGGA
AGCTCCAAATTCTTTATAAGG	新型	NNNRRT
AGGAAAGAAGTCAGAAGGCA	新型	NAA
AGGAGGTTGGTGAGTGATTG	新型	NAG
AGGCAAAAACGAGAGTAACTC	新型	NNNRRT
AGGCATAGCAGCAGGATGAA	新型	NAG
AGGGTTTAAATGTATACCCA	新型	NAG

[1687]

向导序列	向导 id	PAM
AGGTATGTTGCCCGTTTGTCC	新型	NNNRRT
AGTAACTCCACAGTAGCTCC	新型	NAA
AGTCCAAGGAATACTAACAT	M78	NGA
AGTCCCCAACCTCCAATCAC	新型	NNACCA
AGTGATTGGAGGTTGGGGACT	新型	NNGRRT
AGTGATTGGAGGTTGGGGACT	新型	NNNRRT
AGTGCGAATCCACACTCCGA	新型	NAG
AGTGTGGATTCGCACTCCTC	新型	NAG
ATAAAGAATTTGGAGCTACTG	新型	NNGRRT
ATAAAGAATTTGGAGCTACTG	新型	NNNRRT
ATAAATTGGTCTGCGCACCA	新型	NNACCA
ATATGGATGATGTGGTATTG	新型	NGG
ATCCATACTGCGGAACTCCT	新型	NGC
ATGAGGCATAGCAGCAGGAT	新型	NAA
ATGATGTGGTATTGGGGGCC	新型	NAG
ATGGATGATGTGGTATTGGG	新型	NGC
ATGGGATGGGAATACAGGTG	新型	NAA
ATGTCAACGACCGACCTTGA	新型	NGC
ATTCAGCGCCGACGGGACGT	新型	NAA
ATTGTGAGGATTCTTGTCAA	新型	NAA
ATTTAAACCCTAACAAAACAA	新型	NNNRRT
CAAAAACGAGAGTAACTCCA	新型	NAG
CAAAAGAAAATTGGTAACAG	新型	NGG
CAAAATTCGCAGTCCCCAACC	新型	NNNRRT
CAAATGGCACTAGTAAACTG	新型	NGC
CAAGAATCCTCACAATACCG	新型	NAG
CAATACCGCAGAGTCTAGACT	M1	NNNRRT
CACCACGAGTCTAGACTCTG	M36	NGG
CACTGAACAAATGGCACTAG	新型	NAA

[1688]

向导序列	向导 id	PAM
CAGACACATCCAGCGATAAC	新型	NAG
CAGACGGAGAAGGGGACGAG	新型	NGA
CAGGAGGTTGGTGAGTGATT	新型	NGA
CATAAATTGGTCTGCGCACC	新型	NGC
CATACTGCGGAACTCCTAGC	新型	NGC
CATTTAAACCCTAACAAAAC	新型	NAA
CCCAACCTCCAATCACTCAC	新型	NAA
CCCGAGATTGAGATCTTCTG	新型	NGA
CCCTAGAAGAAGAACTCCCT	新型	NGC
CCTACGAACCACTGAACAAA	新型	NGG
CCTAGAAAATTGAGAGAAGT	新型	NNACCA
CCTCACAATACCGCAGAGTC	新型	NAG
CCTGAACTGGAGCCACCAGC	新型	NGG
CGAATTTTGGCCAAGACACA	新型	NGG
CGAGCAAAACAAGCGGCTAG	新型	NAG
CGCAGAAGATCTCAATCTCG	新型	NGA
CGCAGACCAATTTATGCCTA	新型	NAG
CGCAGAGTCTAGACTCGTGG	M35	NGG
CGCCGACGGGACGTAAACAA	新型	NGG
CGCGTAAAGAGAGGTGCGCCC	新型	NNNRRT
CGTGCAGAGGTGAAGCGAAG	新型	NGC
CTACGAACCACTGAACAAAT	新型	NGC
CTAGACTCGTGGTGGACTTCT	E9	NNNRRT
CTCAATCGCCGCGTCGCAGA	M77	NGA
CTCACAATACCGCAGAGTCT	新型	NGA
CTCACCAACCTCCTGTCCTC	新型	NAA
CTGAAAGCCAAACAGTGGGG	新型	NAA
CTGAACTGGAGCCACCAGCAG	新型	NNNRRT
CTGACGCAACCCCCACTGGC	新型	NGG

[1689]

向导序列	向导 id	PAM
CTGCGAGCAAAACAAGCGGCT	新型	NNGRRT
CTGCGAGCAAAACAAGCGGCT	新型	NNNRRT
CTGTGCCAAGTGTTTGCTGA	新型	NGC
CTTGTAAGTTGGCGAGAAAG	新型	NGA
CTTGTCAACAAGAAAAACCC	新型	NGC
CTTGTTGACAAGAATCCTCA	新型	NAA
GAAAAACCCCGCCTGTAACA	新型	NGA
GAAAATTGAGAGAAGTCCACC	新型	NNGRRT
GAAAATTGAGAGAAGTCCACC	新型	NNNRRT
GAAAATTGGTAACAGCGGTA	新型	NAA
GAAAGCCAAACAGTGGGGGA	新型	NAG
GAAAGCCCTACGAACCACTGA	新型	NNNRRT
GAACATCACATCAGGATTCC	新型	NAG
GAACATGGAGAACATCACAT	新型	NAG
GAACTCCCTCGCCTCGCAGAC	E10	NNNRRT
GAAGAACTCCCTCGCCTCGC	E11	NGA
GAAGAAGAACTCCCTCGCCT	新型	NGC
GAAGATGAGGCATAGCAGCA	新型	NGA
GAAGGAAAGAAGTCAGAAGG	新型	NAA
GACAAAAGAAAATTGGTAAC	新型	NGC
GACAAACGGGCAACATACCT	新型	NGA
GACAAACGGGCAACATACCTT	新型	NNNRRT
GACAAGAATCCTCACAATAC	新型	NGC
GACACATCCAGCGATAACCA	M67	NGA
GACGCAACCCCCACTGGCTG	新型	NGG
GACGTAAACAAAGGACGTCC	新型	NGC
GAGAAGTCCACCACGAGTCT	M66	NGA
GAGAGTAACTCCACAGTAGCT	新型	NNNRRT
GAGGACAAACGGGCAACATAC	新型	NNNRRT

[1690]

向导序列	向导 id	PAM
GAGGACAGGAGGTTGGTGAG	新型	NGA
GAGGCATAGCAGCAGGATGA	新型	NGA
GAGGCATAGCAGCAGGATGA	新型	NNAGGA
GAGGTGAAAAAGTTGCATGG	新型	NGC
GAGGTGAAAAAGTTGCATGGT	新型	NNNRRT
GAGGTGAAGCGAAGTGCACA	新型	NGG
GAGTAACTCCACAGTAGCTC	新型	NAA
GAGTCCAAGGAATACTAACAT	新型	NNNRRT
GATCCATACTGCGGAACTCC	新型	NAG
GATGAGAAGGCACAGACGGG	新型	NAG
GATGAGGCATAGCAGCAGGA	新型	NGA
GATGATGTGGTATTGGGGGC	新型	NAA
GATTCAGCGCCGACGGGACG	新型	NAA
GATTGGAGGTTGGGGACTGC	新型	NAA
GCAAACACTTGGCACAGACC	新型	NGG
GCAACTTTTTCACCTCTGCC	新型	NAA
GCAGAAGATCTCAATCTCGG	新型	NAA
GCAGACACATCCAGCGATAA	新型	NNAGGA
GCAGACCAATTTATGCCTAC	新型	NGC
GCAGAGTCTAGACTCGTGGT	M65	NGA
GCATAAATTGGTCTGCGCAC	新型	NAG
GCCCAAGGTCTTACATAAGA	新型	NGA
GCCGACGGGACGTAAACAAA	新型	NGA
GCCGCAGACACATCCAGCGA	新型	NAA
GCCGCAGACACATCCAGCGA	新型	NNACCA
GCGAATCCACACTCCGAAAG	新型	NNACCA
GCGAGCAAAACAAGCGGCTA	新型	NGA
GCGCCGACGGGACGTAAACA	新型	NAG
GCTGCTATGCCTCATCTTCTT	E14	NNNRRT

[1691]

向导序列	向导 id	PAM
GGAAAGAAGTCAGAAGGCAA	新型	NAA
GGAAAGCCCTACGAACCACT	新型	NAA
GGCACTAGTAAACTGAGCCA	新型	NGA
GGCAGACGGAGAAGGGGACGA	新型	NNGRRT
GGCAGACGGAGAAGGGGACGA	新型	NNNRRT
GGCATAGCAGCAGGATGAAG	新型	NGG
GGCGTTCACGGTGGTCTCCA	新型	NGC
GGGAAAGCCCTACGAACCAC	新型	NGA
GGGACTGCGAATTTTGGCCA	新型	NGA
GGGGACTGCGAATTTTGGCC	新型	NAG
GGGGCATTTGGTGGTCTATA	新型	NGC
GGGGCGCACCTCTCTTTACG	新型	NGG
GGTATACATTTAAACCCTAA	新型	NAA
GGTGAGTGATTGGAGGTTGG	新型	NGA
GTAAACAAAGGACGTCCCGC	新型	NNAGGA
GTAAACAAAGGACGTCCCGCG	新型	NNGRRT
GTAAACAAAGGACGTCCCGCG	新型	NNNRRT
GTAGGCATAAATTGGTCTGC	新型	NNACCA
GTATACATTTAAACCCTAAC	新型	NAA
GTCCAAGGAATACTAACATT	新型	NAG
GTCCATGCCCCAAAGCCACC	新型	NAA
GTCGCAGAAGATCTCAATCT	新型	NGG
GTCTAGACTCTGCGGTATTG	新型	NGA
GTCTAGACTCTGCGGTATTG	新型	NNAGGA
GTCTAGACTCTGCGGTATTGT	新型	NNGRRT
GTCTAGACTCTGCGGTATTGT	新型	NNNRRT
GTGAAAAAGTTGCATGGTGC	新型	NGG
GTGAGGATTCTTGTCAACAA	新型	NAA
GTGCCAAGTGTTTGCTGACG	新型	NAA

[1692]

向导序列	向导 id	PAM
GTGCGAATCCACACTCCGAA	新型	NGA
GTGGACTTCTCTCAATTTTC	新型	NAG
GTGTGGATTCGCACTCCTCC	新型	NGC
GTTAATCATTACTTCCAAAC	新型	NAG
GTTACCAATTTTCTTTTGTCT	新型	NNGRRT
GTTACCAATTTTCTTTTGTCT	新型	NNNRRT
GTTAGTATTCCTTGGACTCA	新型	NAA
GTTGGTGAGTGATTGGAGGT	新型	NGG
GTTTACGTCCCGTCGGCGCT	新型	NAA
GTTTACTAGTGCCATTTGTT	新型	NAG
GTTTACTAGTGCCATTTGTTC	新型	NNNRRT
TAAAACGCCGCAGACACATC	新型	NAG
TAAAACGCCGCAGACACATCC	E18	NNNRRT
TAAACAAAGGACGTCCCGCG	新型	NAG
TAAAGAATTTGGAGCTACTG	新型	NGG
TACTAGTGCCATTTGTTCAG	新型	NGG
TAGCAGCAGGATGAAGAGGAA	新型	NNNRRT
TATACATTTAAACCCTAACA	新型	NAA
TATATGGATGATGTGGTATT	新型	NGG
TATGGATGATGTGGTATTGG	M94	NGG
TCACCATATTCTTGGGAACA	新型	NGA
TCAGTTTACTAGTGCCATTTG	新型	NNNRRT
TCCATGCCCCAAAGCCACCC	新型	NAG
TCCTGAACTGGAGCCACCAG	新型	NAG
TCGCAGAAGATCTCAATCTC	新型	NGG
TCTAGACTCTGCGGTATTGT	新型	NAG
TCTCAATCGCCGCGTCGCAG	新型	NAG
TGAAAGCCAAACAGTGGGGG	新型	NAA
TGACGCAACCCCCACTGGCT	新型	NGG

[1693]

向导序列	向导 id	PAM
TGAGGCATAGCAGCAGGATG	新型	NAG
TGATTGGAGGTTGGGGACTG	新型	NGA
TGCCCAAGGTCTTACATAAG	新型	NGG
TGGACTTCTCTCAATTTTCT	E21	NGG
TGGATGATGTGGTATTGGGGG	新型	NNNRRT
TGGCCAAAATTCGCAGTCCC	新型	NAA
TGGGGACTGCGAATTTTGGC	新型	NAA
TGGTGAGTGATTGGAGGTTG	新型	NGG
TGTAAGTTGGCGAGAAAGTG	新型	NAA
TGTAGGCATAAATTGGTCTG	新型	NGC
TGTGAGGATTCTTGTCAACA	新型	NGA
TGTGCACTTCGCTTCACCTC	新型	NGC
TGTGGAGTTACTCTCGTTTT	新型	NGC
TGTTAGTATTCCTTGGACTCA	新型	NNNRRT
TGTTTACGTCCCGTCGGCGC	新型	NGA
TTAAACCCTAACAAAACAAA	新型	NAG
TTCCCCCACTGTTTGGCTTT	新型	NAG
TTCCCGAGATTGAGATCTTC	新型	NGC
TTGCTGACGCAACCCCCACT	新型	NGC
TTGGAGGACAGGAGGTTGGTG	新型	NNNRRT
TTGGTGAGTGATTGGAGGTT	新型	NGG
TTGTAAGTTGGCGAGAAAGT	新型	NAA
TTGTGAGGATTCTTGTCAAC	新型	NAG
TTTGCTGACGCAACCCCCAC	E23_whb	NGG
TTTGTTTACGTCCCGTCGGCG	新型	NNGRRT
TTTGTTTACGTCCCGTCGGCG	新型	NNNRRT

[1694] 表23提供了向导RNA的列表,该向导RNA使用BE4碱基编辑器将错义突变引入HBV基因组中。该向导RNA的序列在HBV基因型A和D之间至少50％保守。对应用gRNA和BE4时将会导致的氨基酸置换进行计算机分析,并且我们进一步选择那些将会产生在0.05％的已知测序

的HBV基因中出现的氨基酸置换的gRNA。那将意味着使用所选择的gRNA进行碱基编辑将会导致HBV蛋白功能失调。

[1695] 表23:BE4 gRNA功能列表

[1696]

向导序列	*向导 id*	*PAM*
AAAAACCCCGCCTGTAACAC	新型	NAG
AAAACAAGCGGCTAGGAGTTC	新型	NNNRRT
AAAACGCCGCAGACACATCC	新型	NGC
AAACAAGCGGCTAGGAGTTC	新型	NGC
AAAGCCAAACAGTGGGGGAA	新型	NGC
AACCACTGAACAAATGGCAC	新型	NAG
AACCCCCACTGGCTGGGGCT	新型	NGG
AACCCCGCCTGTAACACGAG	新型	NAG
AACGCCGCAGACACATCCAG	新型	NGA
AAGAAGTCAGAAGGCAAAAA	新型	NGA
AAGAAGTCAGAAGGCAAAAAC	新型	NNGRRT
AAGAAGTCAGAAGGCAAAAAC	新型	NNNRRT
AAGAATCCTCACAATACCGC	新型	NGA
AAGCCCCAGCCAGTGGGGGT	新型	NGC
AAGCCCTACGAACCACTGAA	新型	NAA
AAGGGGACGAGAGAGTCCCA	新型	NGC
AAGTCAGAAGGCAAAAACGA	新型	NAG
AATCGCCGCGTCGCAGAAGAT	新型	NNNRRT
AATTCGCAGTCCCCAACCTC	新型	NAA
ACAAGAATCCTCACAATACCG	新型	NNGRRT
ACAAGAATCCTCACAATACCG	新型	NNNRRT
ACAAGCGGCTAGGAGTTCCG	新型	NAG
ACACATCCAGCGATAACCAGG	M3	NNNRRT
ACAGGCGGGGTTTTTCTTGT	M64	NGA
ACATCCAGCGATAACCAGGA	新型	NAA
ACGAACCACTGAACAAATGGC	M5	NNNRRT
ACGCAACCCCCACTGGCTGG	新型	NGC
ACGGGGCGCACCTCTCTTTA	新型	NGC

[1697]

向导序列	向导 id	PAM
ACTCCCTCGCCTCGCAGACG	新型	NAG
ACTGAACAAATGGCACTAGT	新型	NAA
ACTTCTCTCAATTTTCTAGG	新型	NGG
ACTTTCTCGCCAACTTACAA	新型	NGC
AGAAGAACTCCCTCGCCTCG	新型	NAG
AGAAGTCAGAAGGCAAAAAC	新型	NAG
AGAATCCTCACAATACCGCA	新型	NAG
AGACACATCCAGCGATAACC	新型	NGG
AGCCCTACGAACCACTGAAC	新型	NAA
AGCGCCGACGGGACGTAAAC	新型	NAA
AGCGCCGACGGGACGTAAAC	新型	NNAGGA
AGCGGCTAGGAGTTCCGCAGT	新型	NNGRRT
AGCGGCTAGGAGTTCCGCAGT	新型	NNNRRT
AGCTCCAAATTCTTTATAAGG	新型	NNNRRT
AGGAGTTCCGCAGTATGGAT	E4	NGG
AGGCATAGCAGCAGGATGAA	新型	NAG
AGGGCTTTCCCCCACTGTTT	新型	NGC
AGTAACTCCACAGTAGCTCC	新型	NAA
AGTAGCTCCAAATTCTTTAT	新型	NAG
AGTCCAAGAGTCCTCTTATG	新型	NAA
AGTCCAAGGAATACTAACAT	M78	NGA
AGTCCCCAACCTCCAATCAC	新型	NNACCA
AGTTCCGCAGTATGGATCGG	新型	NAG
ATCCATACTGCGGAACTCCT	新型	NGC
ATGAGGCATAGCAGCAGGAT	新型	NAA
ATTCCTTGGACTCATAAGGT	新型	NGG
ATTTAAACCCTAACAAAACAA	新型	NNNRRT
ATTTGTTCAGTGGTTCGTAG	新型	NGC
CAAAATTCGCAGTCCCCAACC	新型	NNNRRT

[1698]

向导序列	向导 id	PAM
CAAGAATCCTCACAATACCG	新型	NAG
CAATACCGCAGAGTCTAGACT	M1	NNNRRT
CAATGCTCAGGAGACTCTAA	新型	NGC
CACCACGAGTCTAGACTCTG	M36	NGG
CACTGAACAAATGGCACTAG	新型	NAA
CAGACACATCCAGCGATAAC	新型	NAG
CAGCCAGTGGGGGTTGCGTC	新型	NGC
CAGCGCCGACGGGACGTAAA	新型	NAA
CAGTAGCTCCAAATTCTTTA	新型	NAA
CAGTAGCTCCAAATTCTTTAT	新型	NNGRRT
CAGTAGCTCCAAATTCTTTAT	新型	NNNRRT
CATAGCAGCAGGATGAAGAG	新型	NAA
CCAGCCAGTGGGGGTTGCGT	新型	NAG
CCGCAGTATGGATCGGCAGA	E7	NGA
CCGTGTGTCTTGGCCAAAATT	新型	NNNRRT
CCTACGAACCACTGAACAAA	新型	NGG
CCTCACAATACCGCAGAGTC	新型	NAG
CCTTGGACTCATAAGGTGGG	新型	NAA
CGAGCAAAACAAGCGGCTAG	新型	NAG
CGCAGACCAATTTATGCCTA	新型	NAG
CGCCGCGTCGCAGAAGATCT	新型	NAA
CGCGTAAAGAGAGGTGCGCCC	新型	NNNRRT
CGGCTAGGAGTTCCGCAGTA	新型	NGG
CGTCAGCAAACACTTGGCAC	新型	NGA
CGTGTGTCTTGGCCAAAATT	新型	NGC
CTACGAACCACTGAACAAAT	新型	NGC
CTAGACTCGTGGTGGACTTCT	E9	NNNRRT
CTAGCCGCTTGTTTTGCTCG	新型	NAG
CTCACCAACCTCCTGTCCTC	新型	NAA

[1699]

向导序列	向导 id	PAM
CTCTCGTCCCCTTCTCCGTC	新型	NGC
CTGAAAGCCAAACAGTGGGG	新型	NAA
CTGACGCAACCCCCACTGGC	新型	NGG
CTGCGAGCAAAACAAGCGGC	新型	NAG
CTGCGAGCAAAACAAGCGGCT	新型	NNGRRT
CTGCGAGCAAAACAAGCGGCT	新型	NNNRRT
CTGTGCCAAGTGTTTGCTGA	新型	NGC
CTTCTCTCAATTTTCTAGGG	M87	NGA
CTTCTGCGACGCGGCGATTG	新型	NGA
GAAAAACCCCGCCTGTAACA	新型	NGA
GAAAGCCAAACAGTGGGGGA	新型	NAG
GAAAGCCCTACGAACCACTGA	新型	NNNRRT
GAACTCCCTCGCCTCGCAGAC	E10	NNNRRT
GAACTCCTAGCCGCTTGTTT	新型	NGC
GAAGAACTCCCTCGCCTCGC	E11	NGA
GAAGTCAGAAGGCAAAAACG	新型	NGA
GACAAACGGGCAACATACCT	新型	NGA
GACAAACGGGCAACATACCTT	新型	NNNRRT
GACACATCCAGCGATAACCA	M67	NGA
GACGCAACCCCCACTGGCTG	新型	NGG
GACTTCTCTCAATTTTCTAG	E12	NGG
GAGAAGTCCACCACGAGTCT	M66	NGA
GAGCCTGAGGGCTCCACCCC	新型	NAA
GAGGACAAACGGGCAACATAC	新型	NNNRRT
GAGGACAGGAGGTTGGTGAG	新型	NGA
GAGGCATAGCAGCAGGATGA	新型	NGA
GAGGCATAGCAGCAGGATGA	新型	NNAGGA
GAGTCCAAGGAATACTAACAT	新型	NNNRRT
GATCCATACTGCGGAACTCC	新型	NAG

[1700]

向导序列	向导 id	PAM
GATGAGGCATAGCAGCAGGA	新型	NGA
GCAAACACTTGGCACAGACC	新型	NGG
GCAGACACATCCAGCGATAA	新型	NNAGGA
GCAGACCAATTTATGCCTAC	新型	NGC
GCAGAGTCTAGACTCGTGGT	M65	NGA
GCATAGCAGCAGGATGAAGA	新型	NGA
GCCGCAGACACATCCAGCGA	新型	NAA
GCCGCAGACACATCCAGCGA	新型	NNACCA
GCGAATCCACACTCCGAAAG	新型	NNACCA
GCGAGCAAAACAAGCGGCTA	新型	NGA
GCGCCGACGGGACGTAAACA	新型	NAG
GCGTCAGCAAACACTTGGCA	新型	NAG
GCTCAGGAGACTCTAAGGCTT	新型	NNNRRT
GCTCCTCTGCCGATCCATAC	新型	NGC
GCTGCGAGCAAAACAAGCGG	新型	NNAGGA
GCTGCTATGCCTCATCTTCTT	E14	NNNRRT
GGACTTCTCTCAATTTTCTA	E15	NGG
GGAGTTCCGCAGTATGGATC	新型	NGC
GGCACTAGTAAACTGAGCCA	新型	NGA
GGCATAGCAGCAGGATGAAG	新型	NGG
GGCGGGGTTTTTCTTGTTGAC	新型	NNGRRT
GGCGGGGTTTTTCTTGTTGAC	新型	NNNRRT
GGGACTGCGAATTTTGGCCA	新型	NGA
GGGGACTGCGAATTTTGGCC	新型	NAG
GGGGCATTTGGTGGTCTATA	新型	NGC
GGGGCGCACCTCTCTTTACG	新型	NGG
GGGTTGCGTCAGCAAACACT	新型	NGG
GGTATACATTTAAACCCTAA	新型	NAA
GGTTGCGTCAGCAAACACTT	新型	NGC

[1701]

向导序列	向导 id	PAM
GTAGCTCCAAATTCTTTATA	新型	NGG
GTATACATTTAAACCCTAAC	新型	NAA
GTCCAAGAGTCCTCTTATGT	新型	NAG
GTCCAAGGAATACTAACATT	新型	NAG
GTCCACCACGAGTCTAGACTC	E6/M2	NNNRRT
GTCCATGCCCCAAAGCCACC	新型	NAA
GTCCCGTCGGCGCTGAATCC	新型	NGC
GTCCTTTGTTTACGTCCCGT	新型	NGG
GTCGCAGAAGATCTCAATCT	新型	NGG
GTCTGTGCCTTCTCATCTGC	新型	NGG
GTGCCAAGTGTTTGCTGACG	新型	NAA
GTGGACTTCTCTCAATTTTC	新型	NAG
GTTAATCATTACTTCCAAAC	新型	NAG
GTTACCAATTTTCTTTTGTCT	新型	NNGRRT
GTTACCAATTTTCTTTTGTCT	新型	NNNRRT
GTTATCGCTGGATGTGTCTG	新型	NGG
GTTCCGCAGTATGGATCGGC	E17	NGA
GTTCCGCAGTATGGATCGGC	新型	NNAGGA
GTTTACTAGTGCCATTTGTT	新型	NAG
GTTTACTAGTGCCATTTGTTC	新型	NNNRRT
TAAAACGCCGCAGACACATC	新型	NAG
TAAAACGCCGCAGACACATCC	E18	NNNRRT
TACCGCAGAGTCTAGACTCG	M34	NGG
TAGCAGCAGGATGAAGAGGAA	新型	NNNRRT
TAGCCGCTTGTTTTGCTCGC	新型	NGC
TAGCCGCTTGTTTTGCTCGCA	新型	NNNRRT
TAGGCAGAGGTGAAAAAGTTG	新型	NNNRRT
TAGGGCTTTCCCCCACTGTT	新型	NGG
TAGTATTCCTTGGACTCATA	新型	NGG

[1702]

向导序列	向导 id	PAM
TATACATTTAAACCCTAACA	新型	NAA
TATGTTGCCCGTTTGTCCTC	新型	NAA
TATTCCTTGGACTCATAAGG	新型	NGG
TCAGTTTACTAGTGCCATTTG	新型	NNNRRT
TCCACCACGAGTCTAGACTC	新型	NGC
TCCCCCTAGAAAATTGAGAG	新型	NAG
TCCGCAGTATGGATCGGCAG	E19	NGG
TCCTAGCCGCTTGTTTTGCT	新型	NGC
TCCTCTGCCGATCCATACTG	E20	NGG
TCCTCTTCATCCTGCTGCTA	新型	NGC
TCGCAGAAGATCTCAATCTC	新型	NGG
TCTCAATCGCCGCGTCGCAG	新型	NAG
TCTTCTGCGACGCGGCGATT	新型	NAG
TCTTGTTCCCAAGAATATGG	新型	NGA
TGAAAGCCAAACAGTGGGGG	新型	NAA
TGACGCAACCCCCACTGGCT	新型	NGG
TGAGCCTGAGGGCTCCACCC	新型	NAA
TGAGGCATAGCAGCAGGATG	新型	NAG
TGCCCAAGGTCTTACATAAG	新型	NGG
TGCCCGTTTGTCCTCTAATT	新型	NNAGGA
TGCCCGTTTGTCCTCTAATTC	新型	NNGRRT
TGCCCGTTTGTCCTCTAATTC	新型	NNNRRT
TGCGAGCAAAACAAGCGGCT	新型	NGG
TGCTGGTGGCTCCAGTTCAGG	新型	NNNRRT
TGGACTTCTCTCAATTTTCT	E21	NGG
TGGCCAAAATTCGCAGTCCC	新型	NAA
TGGGGACTGCGAATTTTGGC	新型	NAA
TGGTTATCGCTGGATGTGTC	新型	NGC
TGTGCACTTCGCTTCACCTC	新型	NGC

[1703]

向导序列	*向导 id*	*PAM*
TGTGTCTTGGCCAAAATTCG	新型	NAG
TTAAACCCTAACAAAACAAA	新型	NAG
TTACTCTCGTTTTTGCCTTC	新型	NGA
TTAGGCAGAGGTGAAAAAGT	新型	NGC
TTATCGCTGGATGTGTCTGC	新型	NGC
TTCCCCCACTGTTTGGCTTT	新型	NAG
TTCCCGAGATTGAGATCTTC	新型	NGC
TTCCGCAGTATGGATCGGCA	新型	NAG
TTCGCTTCACCTCTGCACGT	新型	NGC
TTGCTGACGCAACCCCCACT	新型	NGC
TTGGAGGACAGGAGGTTGGTG	新型	NNNRRT
TTTGCTGACGCAACCCCCAC	E23	NGG

[1704] 表24提供了一系列被向导RNA靶向的序列,该向导RNA将提前终止(STOP)密码子引入HBV基因中。

[1705] 表24:STOP gRNA列表

[1706]

向导序列	向导 id	PAM
AACAAGATCTACAGCATGGGG	M21	NNGRRT
AACCCCATCTCTTTGTTTTGT	M27	NNGRRT
AAGCCACCCAAGGCACAGCT	M39(前核心区)	NGG
AAGCCCAGGATGATGGGATG	M68	NGA
ACACATCCAGCGATAACCAGG	M3	NNNRRT
ACAGGCGGGGTTTTTCTTGT	M64	NGA
ACCAGGACAAGTTGGAGGAC	M57	NGG
ACCAGGACAAGTTGGAGGACA	M25	NNNRRT
ACGAACCACTGAACAAATGGC	M5	NNNRRT
ACGCCAACAAGGTAGGAGCT	M82	NGA
AGCCACCAGCAGGGAAATAC	M56	NGG
AGCCACCCAAGGCACAGCTT	M72	NGA
AGGCAAGCAATTCTTTGCTG	M43	NGG
AGGGTCCCCAATCCTCGAGA	M86	NGA
AGTCCAAGGAATACTAACAT	M78	NGA
ATTTACACCAAGACATTATCA	M16	NNNRRT
CAACGAATTGTGGGTCTTTT	M62	NGG
CAATACCGCAGAGTCTAGACT	M1	NNNRRT
CACCACGAGTCTAGACTCTG	M36	NGG
CAGGCAAGCAATTCTTTGCT	M42	NGG
CATCCATATAACTGAAAGCCA	M6	NNNRRT
CATGCAACTTTTTCACCTCTG	M8	NNNRRT
CCAATCCTCGAGAAGATTGA	M84	NGA
CCACCAATCGCCAGACAGGA	M55	NGG
CCACCAGCACGGGACCATGC	M90	NGA
CCACCCAAGGCACAGCTTGG	M38	NGG
CCACTCCCATAGGAATTTTC	M69	NGA
CCAGCAAAGAATTGCTTGCC	M73	NGA

[1707]

向导序列	向导 id	PAM
CCAGCCTTCAGAGCAAACACA	M22	NNNRRT
CCAGGACAAGTTGGAGGACA	M88	NGA
CCATGCCCCAAAGCCACCCA	M40(前核心区)	NGG
CCCAACAAGGACACCTGGCC	M81	NGA
CCCACCCAGGTAGCTAGAGTC	M11	NNNRRT
CCCATCTCTTTGTTTTGTTA	M59	NGG
CCCCAGCAAAGAATTGCTTGC	M10	NNGRRT
CCCCATCTCTTTGTTTTGTT	M60	NGG
CCGCCTTCCATAGAGTGTGTA	M19	NNNRRT
CCGGCAACGGCCAGGTCTGTG	M32	NNNRRT
CCTCCACCAATCGCCAGACA	M83	NGA
CGATAACCAGGACAAGTTGG	M58	NGG
CGCAGAGTCTAGACTCGTGG	M35	NGG
CTCAATCGCCGCGTCGCAGA	M77	NGA
CTCCATGCGACGTGCAGAGG	M92	NGA
CTGAGCCAGGAGAAACGGGC	M70	NGA
CTGCCAACTGGATCCTGCGC	M191	NGG
CTTCTCTCAATTTTCTAGGG	M87	NGA
GAAAGCCCAGGATGATGGGA	M37	NGG
GAAAGCCCAGGATGATGGGAT	M4	NNGRRT
GAAGATCTCAATCTCGGGAAC	M14	NNNRRT
GAATCCACACTCCGAAAGACA	M12	NNNRRT
GACACATCCAGCGATAACCA	M67	NGA
GACGACGAGGCAGGTCCCCT	M74	NGA
GACGAGGCAGGTCCCCTAGA	M75	NGA
GACGCCAACAAGGTAGGAGC	M53	NGG
GAGAAGTCCACCACGAGTCT	M66	NGA
GATAACCAGGACAAGTTGGA	M89	NGA
GATTGCAATTGATTATGCCTG	M18	NNNRRT

[1708]

向导序列	向导 id	PAM
GCAAGCAATTCTTTGCTGGG	M45	NGG
GCAGAGTCTAGACTCGTGGT	M65	NGA
GCCACAAGAACACATCATACA	M28	NNNRRT
GCTGCCAACTGGATCCTGCG	M190	NGG
GCTGTACCAAACCTTCGGAC	M91	NGA
GGAACAAGATCTACAGCATG	M51	NGG
GGCAAGCAATTCTTTGCTGG	M44	NGG
GGGAACAAGATCTACAGCAT	M50	NGG
GGTCTCAATCGCCGCGTCGC	M76	NGA
GGTCTCAATCGCCGCGTCGCA	M13	NNNRRT
GGTCTCCATGCGACGTGCAG	M63	NGG
GGTGGTCTCCATGCGACGTGC	M33	NNNRRT
GTCCACCACGAGTCTAGACTC	E16/M2	NNNRRT
GTGGTCTCCATGCGACGTGC	M93	NGA
GTTTTCCAATGAGGATTAAAG	M15	NNNRRT
TACACAATGTGGTTATCCTGC	M31	NNNRRT
TACCACATCATCCATATAAC	M71	NGA
TACCGCAGAGTCTAGACTCG	M34	NGG
TACTTTCCAATCAATAGGCCT	M29	NNNRRT
TATACCCAAAGACAAAAGAAA	M7	NNNRRT
TCAACGAATTGTGGGTCTTT	M61	NGG
TCAATCCCAACAAGGACACC	M52	NGG
TCAGGCAAGCAATTCTTTGC	M41	NGG
TCCAAGGAATACTAACATTG	M46	NGG
TCCCAAGAATATGGTGACCCA	M20	NNNRRT
TCCCCAATCCTCGAGAAGAT	M85	NGA
TCCCCAATCCTCGAGAAGATT	M24	NNNRRT
TGAACAGTTTGTAGGCCCACT	M17	NNNRRT
TGCAATTGATTATGCCTGCT	M48	NGG

[1709]

向导序列	向导 id	PAM
TGCCAACTGGATCCTGCGCG	M189	NGA
TGCTCCAGCTCCTACCTTGT	M54	NGG
TGCTGTACCAAACCTTCGGAC	M26	NNNRRT
TGGGAACAAGATCTACAGCA	M49	NGG
TGTCAACGAATTGTGGGTCTT	M30	NNGRRT
TGTTTTCCAATGAGGATTAA	M79	NGA
TTCAAGCCTCCAAGCTGTGCC	E22/M9	NNGRRT
TTCAGAGCAAACACAGCAAAT	M23	NNNRRT
TTCCAATGAGGATTAAAGAC	M47	NGG
TTTCCACCAGCAATCCTCTG	M80	NGA
AAAGCCAAACAGTGGGGGAA	新型	NGC
AAAGCCCAGGATGATGGGAT	新型	NGG
AACAAGATCTACAGCATGGGG	新型	NNNRRT
AACCACTGAACAAATGGCAC	新型	NAG
AACCAGGACAAGTTGGAGGA	新型	NAG
AACCCCATCTCTTTGTTTTGT	新型	NNNRRT
AACTCTGCAAGATCCCAGAG	新型	NGA
AAGCCCCAGCCAGTGGGGGT	新型	NGC
AATGTATACCCAAAGACAAAA	新型	NNNRRT
AATTCGCAGTCCCCAACCTC	新型	NAA
ACATCATCCATATAACTGAA	新型	NGC
ACATCCAGCGATAACCAGGA	新型	NAA
ACCCAAAGACAAAAGAAAAT	新型	NGG
ACCCAGGTAGCTAGAGTCAT	新型	NAG
ACCCCATCTCTTTGTTTTGT	新型	NAG
ACCTCAATGTTAGTATTCCT	新型	NGG
ACGACGAGGCAGGTCCCCTA	新型	NAA
ACTCCCATAGGAATTTTCCG	新型	NAA
ACTCCTCCCAGTCTTTAAAC	新型	NAA

[1710]

向导序列	向导 id	PAM
ACTCCTCCCAGTCTTTAAACA	E96	NNNRRT
ACTCTGCAAGATCCCAGAGT	新型	NAG
AGACGACGAGGCAGGTCCCC	新型	NAG
AGATTGCAATTGATTATGCC	新型	NGC
AGCCAGGAGAAACGGGCTGA	新型	NGC
AGCCCAGGATGATGGGATGG	新型	NAA
AGCCTTCAGAGCAAACACAG	新型	NAA
AGCGATAACCAGGACAAGTT	新型	NNAGGA
AGGTCTCAATCGCCGCGTCG	新型	NAG
ATACTACAAACTTTGCCAGC	新型	NAA
ATCCACACTCCGAAAGACAC	新型	NAA
ATCCAGTTGGCAGCACAGCC	新型	NAG
ATGGGGCAGAATCTTTCCAC	新型	NAG
ATGGGGCAGAATCTTTCCACC	新型	NNNRRT
ATTTGTTCAGTGGTTCGTAG	新型	NGC
ATTTTGGCCAAGACACACGG	新型	NAG
CAAAATTCGCAGTCCCCAACC	新型	NNNRRT
CAACCACCAGCACGGGACCA	新型	NGC
CAATCCCAACAAGGACACCT	新型	NGC
CAATCGCCAGACAGGAAGGC	新型	NGC
CACCAAACTCTGCAAGATCC	新型	NAG
CACCAATCGCCAGACAGGAA	新型	NGC
CACCAGCACGGGACCATGCC	新型	NAA
CACCAGTTGGATCCAGCCTT	新型	NAG
CACTCCCATAGGAATTTTCC	新型	NAA
CAGACGAAGGTCTCAATCGC	新型	NGC
CAGGATCCAGTTGGCAGCAC	新型	NGC
CATACTACAAACTTTGCCAG	新型	NAA
CATCCAGCGATAACCAGGAC	新型	NAG

[1711]

向导序列	向导 id	PAM
CATCCATATAACTGAAAGCC	新型	NAA
CCAATACCACATCATCCATA	新型	NAA
CCACAAGAACACATCATACA	新型	NAA
CCCCAGCAAAGAATTGCTTGC	新型	NNNRRT
CCCCCCAGCAAAGAATTGCT	新型	NGC
CCCGGCAACGGCCAGGTCTG	新型	NGC
CCGCCTTCCATAGAGTGTGT	新型	NAA
CCTCAATGTTAGTATTCCTT	新型	NGA
CCTCCCAGTCTTTAAACAAA	新型	NAG
CCTTCCATAGAGTGTGTAAA	新型	NAG
CGACGAGGCAGGTCCCCTAG	新型	NAG
CGCCAACAAGGTAGGAGCTG	新型	NAG
CGCCCACCGAATGTTGCCCA	E95	NGG
CGGACGACCCTTCTCGGGGT	新型	NGC
CGTCTGGCCAGGTGTCCTTGT	新型	NNGRRT
CGTCTGGCCAGGTGTCCTTGT	新型	NNNRRT
CGTTCCGACCGACCACGGGG	新型	NGC
CTACGAACCACTGAACAAAT	新型	NGC
CTCAATCTCGGGAACCTCAAT	新型	NNNRRT
CTCCACCAATCGCCAGACAG	新型	NAA
CTCCACCCCAAAAGGCCTCCG	新型	NNNRRT
CTCCCATAGGAATTTTCCGA	新型	NAG
CTCTGCAAGATCCCAGAGTG	新型	NGA
CTGAAAGCCAAACAGTGGGG	新型	NAA
CTGCAAGATCCCAGAGTGAG	新型	NGG
CTGGCCAGGTGTCCTTGTTG	新型	NGA
CTGTACCAAACCTTCGGACG	新型	NAA
CTGTGCCAAGTGTTTGCTGA	新型	NGC
GAAAGCCAAACAGTGGGGGA	新型	NAG

[1712]

向导序列	向导 id	PAM
GAAAGCCCAGGATGATGGGAT	新型	NNNRRT
GAACAAGATCTACAGCATGG	新型	NGC
GAGCCACCAGCAGGGAAATA	新型	NAG
GAGCCAGGAGAAACGGGCTG	新型	NGG
GAGTCCAAGGAATACTAACAT	新型	NNNRRT
GATCTCAATCTCGGGAACCT	新型	NAA
GCAGGATCCAGTTGGCAGCA	新型	NAG
GCCAACAAGGTAGGAGCTGG	新型	NGC
GCCACAAGAACACATCATAC	新型	NAA
GCCACCAGCAGGGAAATACA	新型	NGC
GCCAGGTGTCCTTGTTGGGAT	新型	NNNRRT
GCCGTTCCGACCGACCACGG	新型	NGC
GCCTTCAGAGCAAACACAGC	新型	NAA
GCGAATCCACACTCCGAAAG	新型	NNACCA
GCGATAACCAGGACAAGTTG	新型	NAG
GCTCCAGCTCCTACCTTGTT	新型	NGC
GGCATACTACAAACTTTGCCA	新型	NNNRRT
GGCTCCAGTTCAGGAGCAGT	新型	NAA
GGGCAGAATCTTTCCACCAG	新型	NAA
GGGCCATCAGCGCGTGCGTG	新型	NAA
GTACGAGATCTTCTAGATAC	新型	NGC
GTATACCCAAAGACAAAAGA	新型	NAA
GTCCAAGGAATACTAACATT	新型	NAG
GTCTCAATCGCCGCGTCGCA	新型	NAA
GTCTGGCCAGGTGTCCTTGT	新型	NGG
GTGAAACCACAAGAGTTGCC	新型	NGA
GTGCCAAGTGTTTGCTGACG	新型	NAA
GTGCTGCCAACTGGATCCTG	新型	NGC
GTGTTTTCCAATGAGGATTA	新型	NAG

[1713]

向导序列	向导 id	PAM
TAACCAGGACAAGTTGGAGG	新型	NNAGGA
TAAGCAGGCTTTCACTTTCT	新型	NGC
TACACCAAGACATTATCAAA	新型	NAA
TCACCAAACTCTGCAAGATCC	新型	NNGRRT
TCACCAAACTCTGCAAGATCC	新型	NNNRRT
TCAGGCTCAGGGCATACTAC	新型	NAA
TCATCCATATAACTGAAAGC	新型	NAA
TCCAATCAATAGGCCTGTTA	新型	NNAGGA
TCCACCAATCGCCAGACAGG	新型	NAG
TCCACCACGAGTCTAGACTC	新型	NGC
TCCCAACAAGGACACCTGGC	新型	NAG
TCCCATAGGAATTTTCCGAA	新型	NGC
TCCCGACCACCAGTTGGATC	新型	NAG
TCGCAGACGAAGGTCTCAAT	新型	NGC
TCTCAATCGCCGCGTCGCAG	新型	NAG
TCTGCAAGATCCCAGAGTGA	新型	NAG
TCTGGCCAGGTGTCCTTGTT	新型	NGG
TGAAACCACAAGAGTTGCCT	新型	NAA
TGAAAGCCAAACAGTGGGGG	新型	NAA
TGAGCCAGGAGAAACGGGCT	新型	NAG
TGCCACAAGAACACATCATA	新型	NAA
TGCCGAACCTGCATGACTAC	新型	NGC
TGGCCAAAATTCGCAGTCCC	新型	NAA
TGGCTCCAGTTCAGGAGCAG	新型	NAA
TGGGGCAGAATCTTTCCACC	新型	NGC
TGGTCTCCATGCGACGTGCA	新型	NAG
TGTACCAAACCTTCGGACGG	新型	NAA
TGTATACCCAAAGACAAAAG	新型	NAA
TGTCAACGAATTGTGGGTCTT	新型	NNNRRT

[1714]

向导序列	向导 id	PAM
TTACACCAAGACATTATCAA	新型	NAA
TTCTCTCAATTTTCTAGGGG	新型	NAA
TTGCAATTGATTATGCCTGC	新型	NAG
TTTACACAATGTGGTTATCC	新型	NGC
TTTACACCAAGACATTATCA	新型	NAA
TTTCCAATCAATAGGCCTGT	新型	NAA
TTTCCAATGAGGATTAAAGA	新型	NAG

[1715] 表25是一系列被向导RNA靶向的HBV基因组中的序列。

[1716] 表25:用于靶向HBV基因组的gRNA

[1717]

向导序列	向导 id	PAM
AAAAAATCAAAGAATGTTTT	新型	NGA
AAAAAATGTGAACAGTTTGT	新型	NGG
AAAAACCCCGCCTGTAACAC	新型	NAG
AAAAAGTTGCATGGTGCTGG	新型	NGC
AAAAATCAAAGAATGTTTTA	新型	NAA
AAAAATGTGAACAGTTTGTA	新型	NGC
AAAACAAGCGGCTAGGAGTTC	新型	NNNRRT
AAAACATTCTTTGATTTTTTG	新型	NNNRRT
AAAACCCCGCCTGTAACACG	新型	NGA
AAAACGCCGCAGACACATCC	新型	NGC
AAAAGATGGTGTTTTCCAAT	新型	NAG
AAAAGGTTCCACGCACGCGC	新型	NGA
AAAAGTGAGACAAGAAATGT	新型	NAA
AAAATCAAAGAATGTTTTAG	新型	NAA
AAAATTGGTAACAGCGGTAA	新型	NAA
AAACAAAGGACGTCCCGCGC	新型	NGG
AAACAAGCGGCTAGGAGTTC	新型	NGC
AAACACAGCAAATCCAGATT	新型	NGG
AAACACTCATCCTCAGGCCA	新型	NGC
AAACCCAGCCCGAATGCTCC	新型	NGC
AAACCCCGCCTGTAACACGA	新型	NAA
AAACGAGAGTAACTCCACAG	新型	NAG
AAACGGGCAACATACCTTGA	新型	NAG
AAACTACTGTTGTTAGACGA	新型	NGA
AAACTGAGCCAGGAGAAACG	新型	NGC
AAACTGTTCACATTTTTTGA	新型	NAA
AAACTTCCTATTAACAGGCCT	新型	NNNRRT
AAAGAATCCCAGAGGATTGC	新型	NGG

[1718]

向导序列	向导 id	PAM
AAAGAATTTGGAGCTACTGT	新型	NGA
AAAGACTGGGAGGAGTTGGG	新型	NGA
AAAGACTGGGAGGAGTTGGG	新型	NNAGGA
AAAGAGATGGGGTTACTCTC	新型	NGA
AAAGATGGTGTTTTCCAATG	新型	NGG
AAAGATTCTGCCCCATGCTG	新型	NAG
AAAGCCAAACAGTGGGGGAA	新型	NGC
AAAGCCCAGGATGATGGGAT	新型	NGG
AAAGGCCTTGTAAGTTGGCG	新型	NGA
AAAGGCCTTGTAAGTTGGCGA	新型	NNNRRT
AAAGGTGGAGACAGCGGGGT	新型	NGG
AAAGTATGTCAACGAATTGT	新型	NGG
AAAGTGAGACAAGAAATGTG	新型	NAA
AAAGTGAGACAAGAAATGTG	新型	NNACCA
AAAGTTGCATGGTGCTGGTG	新型	NGC
AAAGTTTGTAGTATGCCCTG	新型	NGC
AAATACAGGCCTCTCACTCT	新型	NGG
AAATATTTACCATTGGATAA	新型	NGG
AAATCAAAGAATGTTTTAGA	新型	NAA
AAATGGGGCAGCAAAACCCA	新型	NAA
AAATGTATACCCAAAGACAA	新型	NAG
AAATGTATATTAGGAAAAGA	新型	NGG
AAATGTGAAACCACAAGAGT	新型	NGC
AAATTAACACCCACCCAGGT	新型	NGC
AAATTGAGAGAAGTCCACCA	新型	NGA
AAATTGGTAACAGCGGTAAA	新型	NAG
AACAAACAGTCTTTGAAGTA	新型	NGC
AACAAAGGACGTCCCGCGCA	新型	NGA
AACAAATGGCACTAGTAAAC	新型	NGA

[1719]

向导序列	向导 id	PAM
AACAAGAAAAACCCCGCCTG	新型	NAA
AACAAGAAGATGAGGCATAG	新型	NAG
AACAAGAGATGATTAGGCAG	新型	NGG
AACAAGATCTACAGCATGGGG	M21	NNGRRT
AACAAGATCTACAGCATGGGG	新型	NNNRRT
AACAAGGACACCTGGCCAGA	新型	NGC
AACAATGCTCAGGAGACTCT	新型	NAG
AACACAGCAAATCCAGATTG	新型	NGA
AACACATCATACAAAAAATC	新型	NAA
AACACATCATACAAAAAATCA	新型	NNGRRT
AACACATCATACAAAAAATCA	新型	NNNRRT
AACACCCACCCAGGTAGCTA	新型	NAG
AACACGAGAAGGGGTCCTAG	新型	NAA
AACAGAGTTATCAGTCCCGA	新型	NAA
AACAGCGGTAAAAAGGGACT	新型	NAA
AACAGTAGGACATGAACAAG	新型	NGA
AACAGTAGGACATGAACAAGA	新型	NNNRRT
AACAGTCTTTGAAGTATGCCT	新型	NNNRRT
AACAGTGGGGGAAAGCCCTA	新型	NGA
AACATACCTTGATAGTCCAG	新型	NAG
AACATCACATCAGGATTCCT	新型	NGG
AACATGGAGAACATCACATC	新型	NGG
AACATTGAGGTTCCCGAGAT	新型	NGA
AACCACTGAACAAATGGCAC	新型	NAG
AACCAGGACAAGTTGGAGGA	新型	NAG
AACCCATAAAATTCAGAGAG	新型	NAA
AACCCCATCTCTTTGTTTTGT	M27	NNGRRT
AACCCCATCTCTTTGTTTTGT	新型	NNNRRT
AACCCCCACTGGCTGGGGCT	新型	NGG

[1720]

向导序列	向导 id	PAM
AACCCCGCCTGTAACACGAG	新型	NAG
AACCCCGCCTGTAACACGAGA	新型	NNGRRT
AACCCCGCCTGTAACACGAGA	新型	NNNRRT
AACCCTAACAAAACAAAGAGA	新型	NNGRRT
AACCCTAACAAAACAAAGAGA	新型	NNNRRT
AACCTAGCAGGCATAATCAAT	新型	NNNRRT
AACCTGCATGACTACTGCTC	新型	NAG
AACCTTTCACCAAACTCTGC	新型	NAG
AACCTTTGGATAAAACCTAG	新型	NAG
AACCTTTTCGGCTCCTCTGC	新型	NGA
AACGAGAGTAACTCCACAGT	新型	NGC
AACGCAGGATAACCACATTGT	新型	NNNRRT
AACGCCCACCGAATGTTGCC	新型	NAA
AACGCCGCAGACACATCCAG	新型	NGA
AACGGTTTCTCTTCCAAAAG	新型	NGA
AACTAATGACTCTAGCTACC	新型	NGG
AACTACCGTGTGTCTTGGCC	新型	NAA
AACTACTGTTGTTAGACGAC	新型	NAG
AACTAGATGTTCTGGATAAT	新型	NAG
AACTCCCTCGCCTCGCAGAC	新型	NAA
AACTCCTCCCAGTCTTTAAA	新型	NAA
AACTCTGCAAGATCCCAGAG	新型	NGA
AACTCTGTTGTCCTCTCCCG	新型	NAA
AACTGAAAGCCAAACAGTGG	新型	NGG
AACTGGAGCCACCAGCAGGG	新型	NAA
AACTTCCAATGACATAACCCA	新型	NNNRRT
AACTTGTCCTGGTTATCGCT	新型	NGA
AAGAAAATTGGTAACAGCGG	新型	NAA
AAGAACCAACAAGAAGATGA	新型	NGC

[1721]

向导序列	向导 id	PAM
AAGAAGATGAGGCATAGCAG	新型	NAG
AAGAAGATTGCAATTGATTA	新型	NGC
AAGAAGTCAGAAGGCAAAAA	新型	NGA
AAGAAGTCAGAAGGCAAAAAC	新型	NNGRRT
AAGAAGTCAGAAGGCAAAAAC	新型	NNNRRT
AAGAATCCTCACAATACCGC	新型	NGA
AAGAATTGCTTGCCTGAGTG	新型	NAG
AAGAATTTGGAGCTACTGTG	新型	NAG
AAGACATTATCAAAAAATGT	新型	NAA
AAGACATTATCAAAAAATGTG	新型	NNNRRT
AAGACTGGGAGGAGTTGGGG	新型	NAG
AAGACTGTTTGTTTAAAGAC	新型	NGG
AAGAGAAACCGTTATAGAGTA	新型	NNNRRT
AAGAGAGAAACAACACATAG	新型	NGC
AAGAGATGATTAGGCAGAGG	新型	NGA
AAGAGATGGGGTTACTCTCT	新型	NAA
AAGAGGACTCTTGGACTCTC	新型	NGC
AAGATCTACAGCATGGGGCA	新型	NAA
AAGATCTCGTACTGAAGGAA	新型	NGA
AAGATGGTGTTTTCCAATGA	新型	NGA
AAGATTCTGCCCCATGCTGT	新型	NGA
AAGATTGACGATAAGGGAGA	新型	NGC
AAGCAATTCTTTGCTGGGGG	新型	NAA
AAGCCACCCAAGGCACAGCT	M39	NGG
AAGCCCAGGATGATGGGATG	M68	NGA
AAGCCCCAGCCAGTGGGGGT	新型	NGC
AAGCCCTACGAACCACTGAA	新型	NAA
AAGCCTCCAAGCTGTGCCTT	新型	NGG
AAGCGAAGTGCACACGGTCC	新型	NGC

[1722]

向导序列	向导 id	PAM
AAGCGAAGTGCACACGGTCCG	新型	NNNRRT
AAGGAAAGAAGTCAGAAGGC	新型	NAA
AAGGACACCTGGCCAGACGC	新型	NAA
AAGGCACAGCTTGGAGGCTT	新型	NAA
AAGGCACAGCTTGGAGGCTTG	E1	NNNRRT
AAGGCCTCCGTGCGGTGGGG	新型	NGA
AAGGCCTTGTAAGTTGGCGA	新型	NAA
AAGGCGGGTATATTATATAA	新型	NAG
AAGGCTTCCCGATACAGAGC	新型	NGA
AAGGGACTCAAGATGCTGTA	新型	NAG
AAGGGGACGAGAGAGTCCCA	新型	NGC
AAGGGGTCCTAGGAATCCTGA	新型	NNNRRT
AAGGGTCGATGTCCATGCCC	新型	NAA
AAGGGTCGTCCGCAGGATTC	新型	NGC
AAGGTAGGAGCTGGAGCATT	新型	NGG
AAGGTCGGTCGTTGACATTG	新型	NAG
AAGGTGGAGACAGCGGGGTA	新型	NGC
AAGGTTACCAAATATTTACCA	新型	NNGRRT
AAGGTTACCAAATATTTACCA	新型	NNNRRT
AAGGTTTGGTACAGCAACAG	新型	NAG
AAGGTTTGGTACAGCAACAGG	新型	NNGRRT
AAGGTTTGGTACAGCAACAGG	新型	NNNRRT
AAGTCAGAAGGCAAAAACGA	新型	NAG
AAGTGAAAGCCTGCTTAGAT	新型	NGA
AAGTTATGGGTCCTTGCCAC	新型	NAG
AAGTTGGAGGACAGGAGGTTG	新型	NNGRRT
AAGTTGGAGGACAGGAGGTTG	新型	NNNRRT
AAGTTTTCTAAAACATTCTT	新型	NGA
AATACAGGCCTCTCACTCTG	新型	NGA

[1723]

向导序列	向导 id	PAM
AATACTAACATTGAGGTTCC	新型	NGA
AATAGTGTCTAGTTTGGAAG	新型	NAA
AATAGTGTCTAGTTTGGAAGT	新型	NNNRRT
AATATGGTGACCCACAAAAT	新型	NAG
AATATTTGGTAACCTTTGGA	新型	NAA
AATCAATAGGCCTGTTAATA	新型	NGA
AATCCCAGAGGATTGCTGGT	新型	NGA
AATCCGCCTCCTGCCTCCAC	新型	NAA
AATCCTCTGGGATTCTTTCC	新型	NGA
AATCCTCTGGGATTCTTTCC	新型	NNACCA
AATCCTGCGGACGACCCTTCT	新型	NNGRRT
AATCCTGCGGACGACCCTTCT	新型	NNNRRT
AATCGCCGCGTCGCAGAAGAT	新型	NNNRRT
AATCTCGGGAACCTCAATGT	新型	NAG
AATCTTCTTTTCTCATTAACT	新型	NNNRRT
AATGACTCTAGCTACCTGGG	新型	NGG
AATGATTAACTAGATGTTCT	新型	NGA
AATGATTAACTAGATGTTCTG	新型	NNNRRT
AATGGCACTAGTAAACTGAG	新型	NNAGGA
AATGGGGCAGCAAAACCCAA	新型	NAG
AATGTATACCCAAAGACAAA	新型	NGA
AATGTATACCCAAAGACAAAA	新型	NNNRRT
AATGTCAACGACCGACCTTG	新型	NGG
AATGTTTGCTCCAGACCTGC	新型	NGC
AATTCGCAGTCCCCAACCTC	新型	NAA
AATTCGTTGACATACTTTCCA	新型	NNNRRT
AATTCTTTGCTGGGGGGAAC	新型	NAA
AATTGAGAGAAGTCCACCAC	新型	NAG
AATTGGTAACAGCGGTAAAA	新型	NGG

[1724]

向导序列	向导 id	PAM
AATTTATGCCTACAGCCTCC	新型	NAG
AATTTGGAAGATCCAGCATC	新型	NAG
AATTTTATGGGTTATGTCAT	新型	NGG
AATTTTATGGGTTATGTCATT	新型	NNNRRT
AATTTTCCGAAAGCCCAGGA	新型	NGA
ACAAACTTTGCCAGCAAATC	新型	NGC
ACAAAGAGATGGGGTTACTCT	新型	NNGRRT
ACAAAGAGATGGGGTTACTCT	新型	NNNRRT
ACAAATGGCACTAGTAAACT	新型	NAG
ACAACAGTAGTTTCCGGAAGT	新型	NNNRRT
ACAACCTTTCACCAAACTCTG	新型	NNNRRT
ACAAGAAATGTGAAACCACA	新型	NGA
ACAAGAACACATCATACAAA	新型	NAA
ACAAGAAGATGAGGCATAGC	新型	NGC
ACAAGAATCCTCACAATACCG	新型	NNGRRT
ACAAGAATCCTCACAATACCG	新型	NNNRRT
ACAAGAGTTGCCTGAACTTT	新型	NGG
ACAAGATCTACAGCATGGGG	新型	NAG
ACAAGCGGCTAGGAGTTCCG	新型	NAG
ACAAGGCCTTTCTGTGTAAA	新型	NAA
ACAAGGGCATTAACGCAGGA	新型	NAA
ACAAGGGCATTAACGCAGGA	新型	NNACCA
ACAATGCTCAGGAGACTCTA	新型	NGG
ACAATGTGGTTATCCTGCGT	新型	NAA
ACACACGGTAGTTCCCCCTA	新型	NAA
ACACATAGCGCCTCATTTTG	新型	NGG
ACACATCATACAAAAAATCA	新型	NAG
ACACATCCAGCGATAACCAGG	M3	NNNRRT
ACACCTGGCCAGACGCCAAC	新型	NAG

[1725]

向导序列	向导 id	PAM
ACACGGTAGTTCCCCCTAGA	新型	NAA
ACACGGTCCGGCAGATGAGA	新型	NGG
ACACTATTTACACACTCTAT	新型	NGA
ACACTCATCCTCAGGCCATG	新型	NAG
ACAGAAAGGCCTTGTAAGTT	新型	NGC
ACAGACGGGGAGTCCGCGTA	新型	NAG
ACAGAGCTGAGGCGGTATCT	新型	NGA
ACAGAGCTGAGGCGGTATCTA	新型	NNNRRT
ACAGCAAATCCAGATTGGGAC	新型	NNNRRT
ACAGCAACAGGAGGGATACA	新型	NAG
ACAGCAACAGGAGGGATACAT	新型	NNNRRT
ACAGCGGTAAAAAGGGACTC	新型	NAG
ACAGCTTGGAGGCTTGAACA	新型	NNAGGA
ACAGGAGGTTGGTGAGTGAT	新型	NGG
ACAGGAGGTTGGTGAGTGATT	新型	NNNRRT
ACAGGCGGGGTTTTTCTTGT	M64	NGA
ACAGGTACAGTAGAAGAATA	新型	NAG
ACAGGTGCAATTTCCGTCCG	新型	NAG
ACAGTAGTTTCCGGAAGTGT	新型	NGA
ACAGTCTTTGAAGTATGCCT	新型	NAA
ACAGTGGGGGAAAGCCCTAC	新型	NAA
ACAGTGGGGGAAAGCCCTAC	新型	NNACCA
ACAGTTTGTAGGCCCACTTA	新型	NAG
ACATAACCCATAAAATTCAG	新型	NGA
ACATAACTGACTACTAGGTCT	新型	NNNRRT
ACATACCTTGATAGTCCAGA	新型	NGA
ACATCACATCAGGATTCCTA	新型	NGA
ACATCATACAAAAAATCAAA	新型	NAA
ACATCATCCATATAACTGAA	新型	NGC

[1726]

向导序列	向导 id	PAM
ACATCCAGCGATAACCAGGA	新型	NAA
ACATCGTATCCATGGCTGCT	新型	NGG
ACATGAACAAGAGATGATTA	新型	NGC
ACATGGAGAACATCACATCA	新型	NGA
ACATTATCAAAAAATGTGAA	新型	NAG
ACATTCTTTGATTTTTTGTA	新型	NGA
ACATTGAGGTTCCCGAGATT	新型	NAG
ACATTGTGTAAATGGGGCAG	新型	NAA
ACATTTAAACCCTAACAAAA	新型	NAA
ACATTTCTTGTCTCACTTTT	新型	NGA
ACCAAACTCTGCAAGATCCC	新型	NGA
ACCAAATATTTACCATTGGA	新型	NAA
ACCAAATATTTACCATTGGAT	新型	NNGRRT
ACCAAATATTTACCATTGGAT	新型	NNNRRT
ACCAAATGCCCCTATCCTAT	新型	NAA
ACCAACAAGAAGATGAGGCA	新型	NAG
ACCAATTTATGCCTACAGCCT	E2	NNNRRT
ACCAATTTTCTTTTGTCTTT	新型	NGG
ACCACATCATCCATATAACT	新型	NAA
ACCACATTGTGTAAATGGGG	新型	NAG
ACCAGGACAAGTTGGAGGAC	M57	NGG
ACCAGGACAAGTTGGAGGACA	M25	NNNRRT
ACCAGTAAAGTTCCCCACCTT	新型	NNGRRT
ACCAGTAAAGTTCCCCACCTT	新型	NNNRRT
ACCAGTTGGATCCAGCCTTC	新型	NGA
ACCCAAAGACAAAAGAAAAT	新型	NGG
ACCCAGGTAGCTAGAGTCAT	新型	NAG
ACCCATAACTTCCAATGACA	新型	NAA
ACCCCATCTCTTTGTTTTGT	新型	NAG

[1727]

向导序列	向导 id	PAM
ACCCCGAGAAGGGTCGTCCG	新型	NAG
ACCCCGCCTGTAACACGAGA	新型	NGG
ACCCCGCTGTCTCCACCTTT	新型	NAG
ACCCCTTCTCGTGTTACAGG	新型	NGG
ACCCTAACAAAACAAAGAGA	新型	NGG
ACCCTTATCCAATGGTAAATA	新型	NNNRRT
ACCCTTCTCGGGGTCGCTTG	新型	NGA
ACCGACCTTGAGGCATACTT	新型	NAA
ACCGCCTCAGCTCTGTATCG	新型	NGA
ACCTAGCAGGCATAATCAAT	新型	NGC
ACCTCAATGTTAGTATTCCT	新型	NGG
ACCTCACCATACTGCACTCA	新型	NGC
ACCTCCTGTCCTCCAACTTGT	新型	NNNRRT
ACCTGCATGACTACTGCTCA	新型	NGG
ACCTGGCCGTTGCCGGGCAA	新型	NGG
ACCTGGGTGGGTGTTAATTT	新型	NGA
ACCTGGGTGGGTGTTAATTTG	新型	NNNRRT
ACCTGTCTTTAATCCTCATT	新型	NGA
ACCTTATTATCCAGAACATC	新型	NAG
ACCTTCGTCTGCGAGGCGAG	新型	NGA
ACCTTGGGCAACATTCGGTG	新型	NGC
ACCTTTACCCCGTTGCCCGG	新型	NAA
ACCTTTCACCAAACTCTGCA	新型	NGA
ACCTTTGGATAAAACCTAGC	新型	NGG
ACGAACCACTGAACAAATGGC	M5	NNNRRT
ACGACGAGGCAGGTCCCCTA	新型	NAA
ACGAGAAGGGGTCCTAGGAAT	新型	NNNRRT
ACGAGGCAGGTCCCCTAGAA	新型	NAA
ACGATGTATATTTGCGGGAG	新型	NGG

[1728]

向导序列	向导 id	PAM
ACGCAACCCCCACTGGCTGG	新型	NGC
ACGCACGCGCTGATGGCCCA	新型	NGA
ACGCACGCGCTGATGGCCCA	新型	NNACCA
ACGCCAACAAGGTAGGAGCT	M82	NGA
ACGCCCACCGAATGTTGCCC	新型	NAG
ACGCGCTGATGGCCCATGAC	新型	NAA
ACGGAGGCCTTTTGGGGTGG	新型	NGC
ACGGCAGACGGAGAAGGGGA	新型	NGA
ACGGGCTGAGGCCCACTCCC	新型	NNAGGA
ACGGGGAGTCCGCGTAAAGA	新型	NAG
ACGGGGCGCACCTCTCTTTA	新型	NGC
ACGGTGGTCTCCATGCGACG	新型	NGC
ACGGTTTCTCTTCCAAAAGT	新型	NAG
ACGTCGCATGGAGACCACCG	新型	NGA
ACTAATATGGGCCTAAAGTT	新型	NAG
ACTAATGACTCTAGCTACCT	新型	NGG
ACTACCGTGTGTCTTGGCCA	新型	NAA
ACTACTAGGTCTCTAGATGC	新型	NGG
ACTACTGTTGTTAGACGACG	新型	NGG
ACTAGATGTTCTGGATAATA	新型	NGG
ACTAGGAGGCTGTAGGCATAA	新型	NNNRRT
ACTAGTAAACTGAGCCAGGA	新型	NAA
ACTATTTACACACTCTATGG	新型	NAG
ACTCATAAGGTGGGGAACTTT	新型	NNNRRT
ACTCCCATAGGAATTTTCCG	新型	NAA
ACTCCCTCGCCTCGCAGACG	新型	NAG
ACTCCTCCAGCTTATAGACCA	新型	NNNRRT
ACTCCTCCCAGTCTTTAAAC	新型	NAA
ACTCCTCCCAGTCTTTAAACA	E96	NNNRRT

[1729]

向导序列	向导 id	PAM
ACTCTAAGGCTTCCCGATAC	新型	NGA
ACTCTAGCTACCTGGGTGGGT	新型	NNNRRT
ACTCTGCAAGATCCCAGAGT	新型	NAG
ACTCTGTTGTCCTCTCCCGC	新型	NAA
ACTGAAAGCCAAACAGTGGG	新型	NGA
ACTGAACAAATGGCACTAGT	新型	NAA
ACTGAAGGAAAGAAGTCAGA	新型	NGG
ACTGGCTGGGGCTTGGTCAT	新型	NGG
ACTGGGAGGAGTTGGGGGAG	新型	NAG
ACTGTTGTTAGACGACGAGG	新型	NAG
ACTGTTTGTTTAAAGACTGG	新型	NAG
ACTGTTTGTTTAAAGACTGGG	新型	NNGRRT
ACTGTTTGTTTAAAGACTGGG	新型	NNNRRT
ACTTACAAGGCCTTTCTGTG	新型	NAA
ACTTACAGTTAATGAGAAAA	新型	NAA
ACTTCAAAGACTGTTTGTTT	新型	NAA
ACTTCCAATGACATAACCCA	新型	NAA
ACTTCTCTCAATTTTCTAGG	新型	NGG
ACTTTCTCGCCAACTTACAA	新型	NGC
AGAAAATTGGTAACAGCGGT	新型	NAA
AGAAACCGTTATAGAGTATT	新型	NGG
AGAAAGGCCTTGTAAGTTGG	新型	NGA
AGAACATCACATCAGGATTC	新型	NNAGGA
AGAAGAACCAACAAGAAGAT	新型	NAG
AGAAGAACTCCCTCGCCTCG	新型	NAG
AGAAGATCTCGTACTGAAGG	新型	NAA
AGAAGATGAGGCATAGCAGC	新型	NGG
AGAAGATTGACGATAAGGGA	新型	NAG
AGAAGGGGACGAGAGAGTCC	新型	NAA

[1730]

向导序列	向导 id	PAM
AGAAGGGGTCCTAGGAATCC	新型	NGA
AGAAGTCAGAAGGCAAAAAC	新型	NAG
AGAATCCTCACAATACCGCA	新型	NAG
AGACAAAAGAAAATTGGTAA	新型	NAG
AGACAAAAGAAAATTGGTAAC	新型	NNNRRT
AGACAAGAAATGTGAAACCA	新型	NAA
AGACAAGAAATGTGAAACCAC	新型	NNGRRT
AGACAAGAAATGTGAAACCAC	新型	NNNRRT
AGACACACGGTAGTTCCCCC	新型	NAG
AGACACATCCAGCGATAACC	新型	NGG
AGACACCAAATACTCTATAA	新型	NGG
AGACAGGTACAGTAGAAGAA	新型	NAA
AGACCACCGTGAACGCCCAC	新型	NGA
AGACCTGCTGCGAGCAAAAC	新型	NAG
AGACCTTCGTCTGCGAGGCG	新型	NGG
AGACCTTCGTCTGCGAGGCGA	新型	NNGRRT
AGACCTTCGTCTGCGAGGCGA	新型	NNNRRT
AGACCTTGGGCAACATTCGG	新型	NGG
AGACGACGAGGCAGGTCCCC	新型	NAG
AGACGGAGAAGGGGACGAGA	新型	NAG
AGACGGGGAGTCCGCGTAAA	新型	NAG
AGACGGGGAGTCCGCGTAAAG	新型	NNNRRT
AGACTGGGAGGAGTTGGGGG	新型	NGG
AGACTGGGAGGAGTTGGGGGA	新型	NNNRRT
AGACTGTTTGTTTAAAGACT	新型	NGG
AGAGAAGTCCACCACGAGTC	新型	NAG
AGAGAGGTGCGCCCCGTGGT	新型	NGG
AGAGAGTCCCAAGCGACCCC	新型	NAG
AGAGATGATTAGGCAGAGGT	新型	NAA

[1731]

向导序列	向导 id	PAM
AGAGCAAACACAGCAAATCC	新型	NGA
AGAGCTGAGGCGGTATCTAG	新型	NAG
AGAGGAAGATGATAAAACGC	新型	NGC
AGAGGCCTGTATTTCCCTGC	新型	NGG
AGAGTATTTGGTGTCTTTCG	新型	NAG
AGAGTCCCAAGCGACCCCGA	新型	NAA
AGAGTCCCAAGCGACCCCGAG	新型	NNGRRT
AGAGTCCCAAGCGACCCCGAG	新型	NNNRRT
AGAGTTTGGTGAAAGGTTGT	新型	NGA
AGATCCAGCATCTAGAGACC	新型	NAG
AGATCCAGCATCTAGAGACCT	新型	NNNRRT
AGATCTCGTACTGAAGGAAA	新型	NAA
AGATCTTCTAGATACCGCCT	新型	NAG
AGATCTTGTTCCCAAGAATA	新型	NGG
AGATGAGAAGGCACAGACGG	新型	NGA
AGATGATTAGGCAGAGGTGA	新型	NAA
AGATGCTGGATCTTCCAAAT	新型	NAA
AGATGCTGTACAGACTTGGCC	新型	NNNRRT
AGATTAAAGGTCTTTGTACT	新型	NGG
AGATTGAATACATGCATACA	新型	NGG
AGATTGCAATTGATTATGCC	新型	NGC
AGCAAAACCCAAAAGACCCA	新型	NAA
AGCAAATCCGCCTCCTGCCT	新型	NNACCA
AGCAACAGGAGGGATACATA	新型	NAG
AGCAATTCTTTGCTGGGGGGA	新型	NNNRRT
AGCAGTAGTCATGCAGGTTC	新型	NGC
AGCCACCAGCAGGGAAATAC	M56	NGG
AGCCACCCAAGGCACAGCTT	M72	NGA
AGCCAGGAGAAACGGGCTGA	新型	NGC

[1732]

向导序列	向导 id	PAM
AGCCCAGGATGATGGGATGG	新型	NAA
AGCCCTACGAACCACTGAAC	新型	NAA
AGCCGAAAAGGTTCCACGCA	新型	NGC
AGCCTGAGGGCTCCACCCCA	新型	NAA
AGCCTTCAGAGCAAACACAG	新型	NAA
AGCGATAACCAGGACAAGTT	新型	NGA
AGCGATAACCAGGACAAGTT	新型	NNAGGA
AGCGCAGGGTCCCCAATCCT	新型	NGA
AGCGCCGACGGGACGTAAAC	新型	NAA
AGCGCCGACGGGACGTAAAC	新型	NNAGGA
AGCGGCTAGGAGTTCCGCAGT	新型	NNGRRT
AGCGGCTAGGAGTTCCGCAGT	新型	NNNRRT
AGCTCCAAATTCTTTATAAGG	新型	NNNRRT
AGCTCTGTATCGGGAAGCCT	新型	NAG
AGCTGTGCCTTGGGTGGCTT	新型	NGG
AGCTTGGAGGCTTGAACAGT	新型	NGG
AGGAAAGAAGTCAGAAGGCA	新型	NAA
AGGAAGATGATAAAACGCCG	新型	NAG
AGGAAGTTTTCTAAAACATTC	新型	NNNRRT
AGGAATTTTCCGAAAGCCCA	新型	NGA
AGGAATTTTCCGAAAGCCCAG	新型	NNNRRT
AGGACAAGTTGGAGGACAGG	新型	NGG
AGGACCCCTTCTCGTGTTAC	新型	NGG
AGGACGTCCCGCGCAGGATC	新型	NAG
AGGAGCAGTAAACCCTGTTC	新型	NGA
AGGAGCTGGAGCATTCGGGC	新型	NGG
AGGAGGCGGATTTGCTGGCA	新型	NAG
AGGAGGCTGTAGGCATAAAT	E3	NGG
AGGAGGTTGGTGAGTGATTG	新型	NAG

[1733]

向导序列	向导 id	PAM
AGGAGTGCGAATCCACACTC	新型	NGA
AGGAGTTCCGCAGTATGGAT	E4	NGG
AGGAGTTGGGGGAGGAGATT	新型	NGA
AGGATCCTCAACCACCAGCA	新型	NGG
AGGATCCTGGAATTAGAGGA	新型	NAA
AGGATGAAGAGGAAGATGAT	新型	NAA
AGGATGATGGGATGGGAATA	新型	NAG
AGGATTAAAGACAGGTACAG	新型	NAG
AGGATTCTTGTCAACAAGAA	新型	NAA
AGGATTGGGGACCCTGCGCT	新型	NAA
AGGCAAAAACGAGAGTAACTC	新型	NNNRRT
AGGCAAGCAATTCTTTGCTG	M43	NGG
AGGCAGGAGGCGGATTTGCT	新型	NGC
AGGCAGGTCCCCTAGAAGAA	新型	NAA
AGGCATAGCAGCAGGATGAA	新型	NAG
AGGCCCACTTACAGTTAATG	新型	NGA
AGGCCTCCGTGCGGTGGGGT	新型	NAA
AGGCCTTGTAAGTTGGCGAG	新型	NAA
AGGCGAGGGAGTTCTTCTTC	新型	NAG
AGGCGGGTATATTATATAAG	新型	NGA
AGGCTGCCTTCCTGTCTGGCG	新型	NNNRRT
AGGCTTCCCGATACAGAGCT	新型	NAG
AGGCTTGAACAGTAGGACAT	新型	NAA
AGGGACTCAAGATGCTGTAC	新型	NGA
AGGGAGAGGCAGTAGTCGGAA	新型	NNGRRT
AGGGAGAGGCAGTAGTCGGAA	新型	NNNRRT
AGGGATACATAGAGGTTCCT	新型	NGA
AGGGCTTTCCCCCACTGTTT	新型	NGC
AGGGGACCTGCCTCGTCGTC	新型	NAA

[1734]

向导序列	向导 id	PAM
AGGGGCATTTGGTGGTCTAT	新型	NAG
AGGGTCCCCAATCCTCGAGA	M86	NGA
AGGGTCGATGTCCATGCCCC	新型	NAA
AGGGTTTAAATGTATACCCA	新型	NAG
AGGGTTTACTGCTCCTGAAC	新型	NGG
AGGTACAGTAGAAGAATAAAG	新型	NNNRRT
AGGTAGGAGCTGGAGCATTC	新型	NGG
AGGTATGTTGCCCGTTTGTCC	新型	NNNRRT
AGGTATTGTTTACACAGAAA	新型	NGC
AGGTCGGTCGTTGACATTGC	新型	NGA
AGGTCGGTCGTTGACATTGCA	新型	NNGRRT
AGGTCGGTCGTTGACATTGCA	新型	NNNRRT
AGGTCTCAATCGCCGCGTCG	新型	NAG
AGGTCTGGAGCAAACATTAT	新型	NGG
AGGTGCGCCCCGTGGTCGGT	新型	NGG
AGGTGTCCTTGTTGGGATTG	新型	NAG
AGGTTCAGGTATTGTTTACA	新型	NAG
AGGTTCCACGCACGCGCTGA	新型	NGG
AGGTTGGGGACTGCGAATTT	新型	NGG
AGGTTTGGTACAGCAACAGG	新型	NGG
AGGTTTTATCCAAAGGTTAC	新型	NAA
AGTAAAGTTCCCCACCTTAT	新型	NAG
AGTAACTCCACAGTAGCTCC	新型	NAA
AGTAATGATTAACTAGATGTT	新型	NNGRRT
AGTAATGATTAACTAGATGTT	新型	NNNRRT
AGTAGAAGAATAAAGACCAG	新型	NAA
AGTAGCTCCAAATTCTTTAT	新型	NAG
AGTAGGACATGAACAAGAGA	新型	NGA
AGTAGTTTCCGGAAGTGTTG	新型	NNAGGA

[1735]

向导序列	向导 id	PAM
AGTAGTTTCCGGAAGTGTTGA	新型	NNGRRT
AGTAGTTTCCGGAAGTGTTGA	新型	NNNRRT
AGTCATTAGTTCCCCCCAGC	新型	NAA
AGTCATTAGTTCCCCCCAGCA	新型	NNGRRT
AGTCATTAGTTCCCCCCAGCA	新型	NNNRRT
AGTCCAAGAGTCCTCTTATG	新型	NAA
AGTCCAAGGAATACTAACAT	M78	NGA
AGTCCAGAAGAACCAACAAG	新型	NAG
AGTCCCAAGCGACCCCGAGA	新型	NGG
AGTCCCCAACCTCCAATCAC	新型	NNACCA
AGTCCCGATAATGTTTGCTC	新型	NAG
AGTCCGCGTAAAGAGAGGTG	新型	NGC
AGTCCTCTTATGTAAGACCT	新型	NGG
AGTCTTTAAACAAACAGTCTT	新型	NNNRRT
AGTCTTTGAAGTATGCCTCA	新型	NGG
AGTGAAAGCCTGCTTAGATT	新型	NAA
AGTGATTGGAGGTTGGGGAC	新型	NGC
AGTGATTGGAGGTTGGGGACT	新型	NNGRRT
AGTGATTGGAGGTTGGGGACT	新型	NNNRRT
AGTGCACACGGTCCGGCAGA	新型	NGA
AGTGCGAATCCACACTCCGA	新型	NAG
AGTGTCTAGTTTGGAAGTAA	新型	NGA
AGTGTGGATTCGCACTCCTC	新型	NAG
AGTGTTGATAGGATAGGGGCA	新型	NNNRRT
AGTTAATGAGAAAAGAAGAT	新型	NGC
AGTTATGGGTCCTTGCCACA	新型	NGA
AGTTATGTCAACACTAATAT	新型	NGG
AGTTCCCCACCTTATGAGTC	新型	NAA
AGTTCCCCACCTTATGAGTC	新型	NNAGGA

[1736]

向导序列	向导 id	PAM
AGTTCCCCCCAGCAAAGAAT	新型	NGC
AGTTCCCCCTAGAAAATTGA	新型	NAG
AGTTCCCCCTAGAAAATTGAG	新型	NNNRRT
AGTTCCGCAGTATGGATCGG	新型	NAG
AGTTCTTCTTCTAGGGGACC	新型	NGC
AGTTGCATGGTGCTGGTGCG	新型	NAG
AGTTGGCAGCACAGCCTAGC	新型	NGC
AGTTTCCGGAAGTGTTGATA	新型	NGA
ATAAAGAATTTGGAGCTACTG	新型	NNGRRT
ATAAAGAATTTGGAGCTACTG	新型	NNNRRT
ATAAATTGGTCTGCGCACCA	新型	NNACCA
ATAACCACATTGTGTAAATG	新型	NGG
ATAACTCTGTTGTCCTCTCC	新型	NGC
ATAACTCTGTTGTCCTCTCCC	新型	NNNRRT
ATAACTGAAAGCCAAACAGT	新型	NGG
ATAACTGACTACTAGGTCTC	新型	NAG
ATAAGAGAGAAACAACACAT	新型	NGC
ATAAGGGAGAGGCAGTAGTC	新型	NGA
ATAATATACCCGCCTTCCAT	新型	NGA
ATACAGGTGCAATTTCCGTC	新型	NGA
ATACAGGTGCAATTTCCGTCC	新型	NNNRRT
ATACATAGAGGTTCCTTGAG	新型	NAG
ATACATAGAGGTTCCTTGAGC	新型	NNNRRT
ATACATGCATACAAGGGCAT	新型	NAA
ATACCGCCTCAGCTCTGTAT	新型	NGG
ATACGATGTATATTTGCGGG	新型	NGA
ATACGATGTATATTTGCGGG	新型	NNAGGA
ATACTAACATTGAGGTTCCC	新型	NAG
ATACTACAAACTTTGCCAGC	新型	NAA

[1737]

向导序列	向导 id	PAM
ATAGAGTATTTGGTGTCTTT	新型	NGG
ATAGCAGCAGGATGAAGAGG	新型	NAG
ATAGCGCCTCATTTTGTGGG	新型	NNACCA
ATAGGAATTTTCCGAAAGCC	新型	NAG
ATAGTCCAGAAGAACCAACA	新型	NGA
ATAGTCCAGAAGAACCAACAA	新型	NNNRRT
ATATAATATACCCGCCTTCCA	新型	NNGRRT
ATATAATATACCCGCCTTCCA	新型	NNNRRT
ATATACATCGTATCCATGGC	新型	NGC
ATATGGATGATGTGGTATTG	新型	NGG
ATATGGGCCTAAAGTTCAGG	新型	NAA
ATATGGTGACCCACAAAATG	新型	NGG
ATATTTGCGGGAGAGGACAA	新型	NAG
ATATTTGGTAACCTTTGGAT	新型	NAA
ATCAATAGGCCTGTTAATAG	新型	NAA
ATCCAAAGGTTACCAAATAT	新型	NNACCA
ATCCAACTGGTGGTCGGGAA	新型	NGA
ATCCAATGGTAAATATTTGG	新型	NAA
ATCCACACTCCGAAAGACAC	新型	NAA
ATCCAGTTGGCAGCACAGCC	新型	NAG
ATCCATACTGCGGAACTCCT	新型	NGC
ATCCCAGAGGATTGCTGGTG	新型	NAA
ATCCCAGAGGATTGCTGGTGG	新型	NNNRRT
ATCCCATCATCCTGGGCTTT	新型	NGG
ATCCCTCCTGTTGCTGTACC	新型	NAA
ATCCTCGAGAAGATTGACGA	新型	NAA
ATCCTGCGGACGACCCTTCT	新型	NGG
ATCGACCCTTATAAAGAATT	新型	NGG
ATCTAGAAGATCTCGTACTG	新型	NAG

[1738]

向导序列	向导 id	PAM
ATCTAGTTAATCATTACTTC	新型	NAA
ATCTCTTTGTTTTGTTAGGGT	新型	NNNRRT
ATCTTCCAAATTAACACCCAC	新型	NNNRRT
ATCTTCTGCGACGCGGCGAT	新型	NGA
ATCTTCTTGTTGGTTCTTCT	新型	NGA
ATCTTGCAGAGTTTGGTGAA	新型	NGG
ATCTTTCCACCAGCAATCCTC	新型	NNGRRT
ATCTTTCCACCAGCAATCCTC	新型	NNNRRT
ATGAACAAGAGATGATTAGG	新型	NAG
ATGAACAAGAGATGATTAGGC	新型	NNNRRT
ATGACATAACCCATAAAATT	新型	NAG
ATGACCAAGCCCCAGCCAGT	新型	NGG
ATGACTCTAGCTACCTGGGT	新型	NGG
ATGAGGATTAAAGACAGGTA	新型	NAG
ATGAGGCATAGCAGCAGGAT	新型	NAA
ATGAGTGTTTCTCAAAGGTG	新型	NAG
ATGATGGGATGGGAATACAGG	新型	NNNRRT
ATGATGTGGTATTGGGGGCC	新型	NAG
ATGATGTGTTCTTGTGGCAA	新型	NGA
ATGATTAGGCAGAGGTGAAA	新型	NAG
ATGCATACAAGGGCATTAAC	新型	NNAGGA
ATGCATACAAGGGCATTAACG	新型	NNGRRT
ATGCATACAAGGGCATTAACG	新型	NNNRRT
ATGCATGTATTCAATCTAAG	新型	NAG
ATGCCTACAGCCTCCTAGTA	新型	NAA
ATGCCTGCTAGGTTTTATCC	新型	NAA
ATGCGACGTGCAGAGGTGAAG	新型	NNNRRT
ATGCTGTAGATCTTGTTCCC	新型	NAG
ATGGACATCGACCCTTATAA	新型	NGA

[1739]

向导序列	向导 id	PAM
ATGGATACGATGTATATTTG	新型	NGG
ATGGATCGGCAGAGGAGCCG	新型	NAA
ATGGATCGGCAGAGGAGCCGA	新型	NNNRRT
ATGGATGATGTGGTATTGGG	新型	NGC
ATGGCACTAGTAAACTGAGC	新型	NAG
ATGGCCCATGACCAAGCCCC	新型	NGC
ATGGCCCATGACCAAGCCCCA	新型	NNNRRT
ATGGGATGGGAATACAGGTG	新型	NAA
ATGGGCCATCAGCGCGTGCG	新型	NGG
ATGGGGCAGAATCTTTCCAC	新型	NAG
ATGGGGCAGAATCTTTCCACC	新型	NNNRRT
ATGGGGCAGCAAAACCCAAA	新型	NGA
ATGGTAAATATTTGGTAACCT	新型	NNGRRT
ATGGTAAATATTTGGTAACCT	新型	NNNRRT
ATGGTCCCGTGCTGGTGGTT	新型	NAG
ATGGTGAGGTGAACAATGCT	新型	NAG
ATGTATACCCAAAGACAAAA	新型	NAA
ATGTCAACACTAATATGGGCC	新型	NNNRRT
ATGTCAACGACCGACCTTGA	新型	NGC
ATGTGATGTTCTCCATGTTC	新型	NGC
ATGTGGTTATCCTGCGTTAA	新型	NGC
ATGTTCTCCATGTTCAGCGC	新型	NGG
ATGTTCTGGATAATAAGGTT	新型	NAA
ATGTTGCCCAAGGTCTTACA	新型	NAA
ATTAAAGACAGGTACAGTAG	新型	NAG
ATTAAAGGTCTTTGTACTAG	新型	NAG
ATTAACACCCACCCAGGTAGC	新型	NNGRRT
ATTAACACCCACCCAGGTAGC	新型	NNNRRT
ATTAACAGGCCTATTGATTG	新型	NAA

[1740]

向导序列	向导 id	PAM
ATTAACTGTAAGTGGGCCTA	新型	NAA
ATTCAGCGCCGACGGGACGT	新型	NAA
ATTCCAGGATCCTCAACCAC	新型	NAG
ATTCCCATCCCATCATCCTG	新型	NGC
ATTCCTATGGGAGTGGGCCT	新型	NAG
ATTCCTTGGACTCATAAGGT	新型	NGG
ATTCGCACTCCTCCAGCTTA	新型	NAG
ATTCTTGGGAACAAGATCTA	新型	NAG
ATTCTTTCCCGACCACCAGT	新型	NGG
ATTGAATACATGCATACAAG	新型	NGC
ATTGAGACCTTCGTCTGCGA	新型	NGC
ATTGAGATCTTCTGCGACGC	新型	NGC
ATTGGGACTTCAATCCCAAC	新型	NAG
ATTGGGGGCCAAGTCTGTAC	新型	NGC
ATTGGTAACAGCGGTAAAAA	新型	NGG
ATTGTGAGGATTCTTGTCAA	新型	NAA
ATTGTGGGTCTTTTGGGTTT	新型	NGC
ATTGTGTAAATGGGGCAGCA	新型	NAA
ATTTAAACCCTAACAAAACA	新型	NAG
ATTTAAACCCTAACAAAACAA	新型	NNNRRT
ATTTACACCAAGACATTATC	新型	NAA
ATTTACACCAAGACATTATCA	M16	NNNRRT
ATTTCTTGTCTCACTTTTGG	新型	NAG
ATTTGCGGGAGAGGACAACA	新型	NAG
ATTTGCTGTGTTTGCTCTGA	新型	NGG
ATTTGGAAGATCCAGCATCT	新型	NGA
ATTTGGTGGTCTATAAGCTG	新型	NAG
ATTTGGTGGTCTATAAGCTGG	新型	NNGRRT
ATTTGGTGGTCTATAAGCTGG	新型	NNNRRT

[1741]

向导序列	向导 id	PAM
ATTTGGTGTCTTTCGGAGTG	新型	NGG
ATTTGTTCAGTGGTTCGTAG	新型	NGC
ATTTTATGGGTTATGTCATT	新型	NGA
ATTTTGGCCAAGACACACGG	新型	NAG
ATTTTTTGATAATGTCTTGGT	新型	NNNRRT
CAAAAAATCAAAGAATGTTT	新型	NAG
CAAAAAATGTGAACAGTTTG	新型	NAG
CAAAAACGAGAGTAACTCCA	新型	NAG
CAAAAGAAAATTGGTAACAG	新型	NGG
CAAAAGGCCTCCGTGCGGTG	新型	NGG
CAAAAGTGAGACAAGAAATG	新型	NGA
CAAAATTCGCAGTCCCCAACC	新型	NNNRRT
CAAACACAGCAAATCCAGAT	新型	NGG
CAAACACTTGGCACAGACCT	新型	NGC
CAAAGAATTGCTTGCCTGAG	新型	NGC
CAAAGACAAAAGAAAATTGG	新型	NAA
CAAAGGACGTCCCGCGCAGGA	新型	NNNRRT
CAAAGGTGGAGACAGCGGGG	新型	NAG
CAAAGTTTGTAGTATGCCCT	新型	NAG
CAAATATACATCGTATCCAT	新型	NGC
CAAATATTTACCATTGGATA	新型	NGG
CAAATGGCACTAGTAAACTG	新型	NGC
CAAATTAACACCCACCCAGG	新型	NAG
CAACACATAGCGCCTCATTTT	新型	NNGRRT
CAACACATAGCGCCTCATTTT	新型	NNNRRT
CAACATACCTTGATAGTCCA	新型	NAA
CAACATTCGGTGGGCGTTCA	新型	NGG
CAACATTCGGTGGGCGTTCAC	新型	NNNRRT
CAACCACCAGCACGGGACCA	新型	NGC

[1742]

向导序列	向导 id	PAM
CAACCTTTCACCAAACTCTG	新型	NAA
CAACGAATTGTGGGTCTTTT	M62	NGG
CAACTTGTCCTGGTTATCGC	新型	NGG
CAAGAAATGTGAAACCACAA	新型	NAG
CAAGAAGATGAGGCATAGCA	新型	NNAGGA
CAAGAAGATGAGGCATAGCAG	新型	NNGRRT
CAAGAAGATGAGGCATAGCAG	新型	NNNRRT
CAAGAATATGGTGACCCACA	新型	NAA
CAAGAATCCTCACAATACCG	新型	NAG
CAAGACATTATCAAAAAATG	新型	NGA
CAAGAGTTGCCTGAACTTTA	新型	NGC
CAAGATCTACAGCATGGGGC	新型	NGA
CAAGCAATTCTTTGCTGGGG	新型	NGA
CAAGCCTCCAAGCTGTGCCT	E5	NGG
CAAGGACCCATAACTTCCAA	新型	NGA
CAAGGCACAGCTTGGAGGCT	E6	NGA
CAAGTTGGAGGACAGGAGGT	新型	NGG
CAATACCGCAGAGTCTAGACT	M1	NNNRRT
CAATCAATAGGCCTGTTAAT	新型	NGG
CAATCAATAGGCCTGTTAATA	新型	NNNRRT
CAATCCCAACAAGGACACCT	新型	NGC
CAATCGCCAGACAGGAAGGC	新型	NGC
CAATCTGGATTTGCTGTGTT	新型	NGC
CAATGCTCAGGAGACTCTAA	新型	NGC
CAATGTCAACGACCGACCTT	新型	NAG
CAATTCGTTGACATACTTTC	新型	NAA
CACAACCTTTCACCAAACTC	新型	NGC
CACAAGAACACATCATACAA	新型	NAA
CACAAGAGTTGCCTGAACTT	新型	NAG

[1743]

向导序列	向导 id	PAM
CACACGGTAGTTCCCCCTAG	新型	NAA
CACACGGTCCGGCAGATGAG	新型	NAG
CACAGAAAGGCCTTGTAAGT	新型	NGG
CACAGACCTGGCCGTTGCCG	新型	NGC
CACAGACGGGGAGTCCGCGT	新型	NAA
CACATAGCGCCTCATTTTGT	新型	NGG
CACATCATACAAAAAATCAA	新型	NGA
CACATCATCCATATAACTGA	新型	NAG
CACATTTCTTGTCTCACTTT	新型	NGG
CACATTTTTTGATAATGTCT	新型	NGG
CACCAAACTCTGCAAGATCC	新型	NAG
CACCAATCGCCAGACAGGAA	新型	NGC
CACCACGAGTCTAGACTCTG	M36	NGG
CACCAGCACGGGACCATGCC	新型	NAA
CACCAGTTGGATCCAGCCTT	新型	NAG
CACCATACTGCACTCAGGCA	新型	NGC
CACCCAAGGCACAGCTTGGA	新型	NGC
CACCCCAAAAGGCCTCCGTG	新型	NGG
CACCGCACGGAGGCCTTTTG	新型	NGG
CACCTCACCATACTGCACTC	新型	NGG
CACCTCTGCACGTCGCATGG	新型	NGA
CACCTCTGCACGTCGCATGG	新型	NNACCA
CACCTGGCCAGACGCCAACA	新型	NGG
CACGGAGGCCTTTTGGGGTG	新型	NAG
CACGGTCCGGCAGATGAGAA	新型	NGC
CACTAGTAAACTGAGCCAGG	新型	NGA
CACTATTTACACACTCTATG	新型	NAA
CACTCAGGCAAGCAATTCTT	新型	NGC
CACTCCCATAGGAATTTTCC	新型	NAA

[1744]

向导序列	向导 id	PAM
CACTGAACAAATGGCACTAG	新型	NAA
CACTGGCTGGGGCTTGGTCA	新型	NGG
CACTGTTTGGCTTTCAGTTAT	新型	NNGRRT
CACTGTTTGGCTTTCAGTTAT	新型	NNNRRT
CACTTACAGTTAATGAGAAA	新型	NGA
CACTTACAGTTAATGAGAAAA	新型	NNNRRT
CACTTTCTCGCCAACTTACA	新型	NGG
CAGAAGAACCAACAAGAAGA	新型	NGA
CAGACACATCCAGCGATAAC	新型	NAG
CAGACCTGCTGCGAGCAAAA	新型	NAA
CAGACCTGGCCGTTGCCGGG	新型	NAA
CAGACGAAGGTCTCAATCGC	新型	NGC
CAGACGGAGAAGGGGACGAG	新型	NGA
CAGACGGGGAGTCCGCGTAA	新型	NGA
CAGAGCAAACACAGCAAATC	新型	NAG
CAGAGCTGAGGCGGTATCTA	新型	NAA
CAGAGTTTGGTGAAAGGTTG	新型	NGG
CAGATGAGAAGGCACAGACG	新型	NGG
CAGCAAAGAATTGCTTGCCT	新型	NAG
CAGCAACAGGAGGGATACAT	新型	NGA
CAGCACAGCCTAGCAGCCAT	新型	NGA
CAGCAGGATGAAGAGGAAGA	新型	NGA
CAGCATGGGGCAGAATCTTT	新型	NNACCA
CAGCCAGTGGGGGTTGCGTC	新型	NGC
CAGCCTAGCAGCCATGGATA	新型	NGA
CAGCCTCCTAGTACAAAGACC	新型	NNNRRT
CAGCGATAACCAGGACAAGT	新型	NGG
CAGCGCCGACGGGACGTAAA	新型	NAA
CAGCGGTAAAAAGGGACTCA	新型	NGA

[1745]

向导序列	向导 id	PAM
CAGCTCTGTATCGGGAAGCCT	新型	NNGRRT
CAGCTCTGTATCGGGAAGCCT	新型	NNNRRT
CAGCTTGGAGGCTTGAACAG	新型	NAG
CAGGACAAGTTGGAGGACAG	新型	NAG
CAGGAGACTCTAAGGCTTCC	新型	NGA
CAGGAGGCGGATTTGCTGGC	新型	NAA
CAGGAGGTTGGTGAGTGATT	新型	NGA
CAGGATCCAGTTGGCAGCAC	新型	NGC
CAGGATGAAGAGGAAGATGA	新型	NAA
CAGGCAAGCAATTCTTTGCT	M42	NGG
CAGGGTCCCCAATCCTCGAG	新型	NAG
CAGGTACAGTAGAAGAATAA	新型	NGA
CAGGTACAGTAGAAGAATAA	新型	NNACCA
CAGGTATTGTTTACACAGAA	新型	NGG
CAGGTGCAATTTCCGTCCGA	新型	NGG
CAGGTGTCCTTGTTGGGATT	新型	NAA
CAGTAAAGTTCCCCACCTTA	新型	NGA
CAGTAGAAGAATAAAGACCAG	新型	NNNRRT
CAGTAGCTCCAAATTCTTTA	新型	NAA
CAGTAGCTCCAAATTCTTTAT	新型	NNGRRT
CAGTAGCTCCAAATTCTTTAT	新型	NNNRRT
CAGTATGGTGAGGTGAACAA	新型	NGC
CAGTCTTTGAAGTATGCCTC	新型	NAG
CAGTTAATGAGAAAAGAAGAT	新型	NNNRRT
CAGTTATGTCAACACTAATA	新型	NGG
CAGTTGGATCCAGCCTTCAG	新型	NGC
CAGTTGGCAGCACAGCCTAG	新型	NAG
CAGTTTGTAGGCCCACTTACA	新型	NNNRRT
CATAAATTGGTCTGCGCACC	新型	NGC

[1746]

向导序列	向导 id	PAM
CATAACCCATAAAATTCAGA	新型	NAG
CATAAGGTGGGGAACTTTAC	新型	NGG
CATACAAGGGCATTAACGCA	新型	NGA
CATACCTTGATAGTCCAGAA	新型	NAA
CATACCTTGATAGTCCAGAA	新型	NNACCA
CATACTACAAACTTTGCCAG	新型	NAA
CATACTGCGGAACTCCTAGC	新型	NGC
CATAGAGGTTCCTTGAGCAG	新型	NAG
CATAGCAGCAGGATGAAGAG	新型	NAA
CATAGGAATTTTCCGAAAGC	新型	NNAGGA
CATAGGAATTTTCCGAAAGCC	新型	NNGRRT
CATAGGAATTTTCCGAAAGCC	新型	NNNRRT
CATATTAGTGTTGACATAAC	新型	NGA
CATCATCCTGGGCTTTCGGA	新型	NAA
CATCCAGCGATAACCAGGAC	新型	NAG
CATCCATATAACTGAAAGCC	新型	NAA
CATCCATATAACTGAAAGCCA	M6	NNNRRT
CATCGTATCCATGGCTGCTA	新型	NGC
CATCTTCCTCTTCATCCTGC	新型	NGC
CATCTTCTTGTTGGTTCTTC	新型	NGG
CATCTTGAGTCCCTTTTTAC	新型	NGC
CATGACCAAGCCCCAGCCAG	新型	NGG
CATGACCAAGCCCCAGCCAGT	新型	NNGRRT
CATGACCAAGCCCCAGCCAGT	新型	NNNRRT
CATGCAACTTTTTCACCTCTG	M8	NNNRRT
CATGCATACAAGGGCATTAA	新型	NGC
CATGCCCCAAAGCCACCCAA	新型	NGC
CATGCGACGTGCAGAGGTGA	新型	NGC
CATGCTGTAGATCTTGTTCC	新型	NAA

[1747]

向导序列	向导 id	PAM
CATGCTGTAGATCTTGTTCCC	新型	NNGRRT
CATGCTGTAGATCTTGTTCCC	新型	NNNRRT
CATGGACATCGACCCTTATA	新型	NAG
CATGGTCCCGTGCTGGTGGT	新型	NGA
CATGGTCCCGTGCTGGTGGT	新型	NNAGGA
CATGGTCCCGTGCTGGTGGTT	新型	NNGRRT
CATGGTCCCGTGCTGGTGGTT	新型	NNNRRT
CATGGTGCTGGTGCGCAGAC	新型	NAA
CATGTTCAGCGCAGGGTCCC	新型	NAA
CATTAGTTCCCCCCAGCAAA	新型	NAA
CATTCGGTGGGCGTTCACGG	新型	NGG
CATTGAGGTTCCCGAGATTG	新型	NGA
CATTGGAAAACACCATCTTTT	新型	NNNRRT
CATTGGAAGTTATGGGTCCT	新型	NGC
CATTGTGTAAATGGGGCAGC	新型	NAA
CATTGTTCACCTCACCATAC	新型	NGC
CATTTAAACCCTAACAAAAC	新型	NAA
CATTTACACCAAGACATTAT	新型	NAA
CATTTCTTGTCTCACTTTTG	新型	NAA
CATTTGGTGGTCTATAAGCT	新型	NGA
CATTTGGTGGTCTATAAGCT	新型	NNAGGA
CATTTGTTCAGTGGTTCGTA	新型	NGG
CCAAAAGACCCACAATTCGT	新型	NGA
CCAAAAGGCCTCCGTGCGGT	新型	NGG
CCAAACCTTCGGACGGAAAT	新型	NGC
CCAAACTCTGCAAGATCCCA	新型	NAG
CCAAATATTTACCATTGGAT	新型	NAG
CCAACAAGAAGATGAGGCAT	新型	NGC
CCAAGAATATGGTGACCCAC	新型	NAA

[1748]

向导序列	向导 id	PAM
CCAAGTCTGTACAGCATCTT	新型	NAG
CCAATACCACATCATCCATA	新型	NAA
CCAATCAATAGGCCTGTTAA	新型	NAG
CCAATCCTCGAGAAGATTGA	M84	NGA
CCAATCGCCAGACAGGAAGG	新型	NAG
CCAATGAGGATTAAAGACAGG	新型	NNNRRT
CCACAAGAACACATCATACA	新型	NAA
CCACAATTCGTTGACATACTT	新型	NNNRRT
CCACATCATCCATATAACTG	新型	NAA
CCACATTGTGTAAATGGGGC	新型	NGC
CCACCAATCGCCAGACAGGA	M55	NGG
CCACCAGCACGGGACCATGC	M90	NGA
CCACCCAAGGCACAGCTTGG	M38	NGG
CCACCGCACGGAGGCCTTTT	新型	NGG
CCACTCCCATAGGAATTTTC	M69	NGA
CCACTGCATGGCCTGAGGAT	新型	NAG
CCACTTACAGTTAATGAGAA	新型	NAG
CCAGACGCCAACAAGGTAGG	新型	NGC
CCAGAGGATTGCTGGTGGAA	新型	NGA
CCAGATTGGGACTTCAATCC	新型	NAA
CCAGCAAAGAATTGCTTGCC	M73	NGA
CCAGCATCTAGAGACCTAGTA	新型	NNNRRT
CCAGCCAGTGGGGGTTGCGT	新型	NAG
CCAGCCTTCAGAGCAAACAC	新型	NGC
CCAGCCTTCAGAGCAAACACA	M22	NNNRRT
CCAGCTTATAGACCACCAAA	新型	NGC
CCAGGACAAGTTGGAGGACA	M88	NGA
CCAGGATGATGGGATGGGAAT	新型	NNNRRT
CCAGGTCTGTGCCAAGTGTT	新型	NGC

[1749]

向导序列	向导 id	PAM
CCAGGTGTCCTTGTTGGGAT	新型	NGA
CCAGTGGGGGTTGCGTCAGC	新型	NAA
CCAGTTGGATCCAGCCTTCA	新型	NAG
CCATACTGCACTCAGGCAAG	新型	NAA
CCATAGAGTGTGTAAATAGTG	新型	NNNRRT
CCATATAACTGAAAGCCAAA	新型	NAG
CCATCATCCTGGGCTTTCGG	新型	NAA
CCATGCAACTTTTTCACCTC	新型	NGC
CCATGCCCCAAAGCCACCCA	M40	NGG
CCATGCGACGTGCAGAGGTG	新型	NAG
CCATGGATACGATGTATATT	新型	NGC
CCATGGCTGCTAGGCTGTGC	新型	NGC
CCATTTGTTCAGTGGTTCGT	新型	NGG
CCCAAAAGGCCTCCGTGCGG	新型	NGG
CCCAAAGCCACCCAAGGCAC	新型	NGC
CCCAACAAGGACACCTGGCC	M81	NGA
CCCAACCTCCAATCACTCAC	新型	NAA
CCCAACTCCTCCCAGTCTTT	新型	NAA
CCCAAGAATATGGTGACCCA	新型	NAA
CCCACCCAGGTAGCTAGAGTC	M11	NNNRRT
CCCACCGCACGGAGGCCTTT	新型	NGG
CCCACTTACAGTTAATGAGA	新型	NAA
CCCAGAGGATTGCTGGTGGA	新型	NAG
CCCATCATCCTGGGCTTTCG	新型	NAA
CCCATCTCTTTGTTTTGTTA	M59	NGG
CCCCAAAAGGCCTCCGTGCGG	新型	NNGRRT
CCCCAAAAGGCCTCCGTGCGG	新型	NNNRRT
CCCCAAAGCCACCCAAGGCA	新型	NAG
CCCCAACTCCTCCCAGTCTT	新型	NAA

[1750]

向导序列	向导 id	PAM
CCCCACCGCACGGAGGCCTTT	新型	NNGRRT
CCCCACCGCACGGAGGCCTTT	新型	NNNRRT
CCCCACCTTATGAGTCCAAG	新型	NAA
CCCCAGCAAAGAATTGCTTGC	M10	NNGRRT
CCCCAGCAAAGAATTGCTTGC	新型	NNNRRT
CCCCATCTCTTTGTTTTGTT	M60	NGG
CCCCCCAGCAAAGAATTGCT	新型	NGC
CCCCGAGAAGGGTCGTCCGC	新型	NGG
CCCCGCCTGTAACACGAGAA	新型	NGG
CCCCGCTGTCTCCACCTTTG	新型	NGA
CCCCGTTGCCCGGCAACGGC	新型	NAG
CCCCTTCTCGTGTTACAGGC	新型	NGG
CCCGACCACCAGTTGGATCC	新型	NGC
CCCGAGAAGGGTCGTCCGCA	新型	NGA
CCCGAGATTGAGATCTTCTG	新型	NGA
CCCGCCTGTAACACGAGAAG	新型	NGG
CCCGCCTTCCATAGAGTGTG	新型	NAA
CCCGCGCAGGATCCAGTTGG	新型	NAG
CCCGCTGTCTCCACCTTTGA	新型	NAA
CCCGGCAACGGCCAGGTCTG	新型	NGC
CCCGTCGGCGCTGAATCCTG	新型	NGG
CCCGTCTGTGCCTTCTCATC	新型	NGC
CCCGTGGTCGGTCGGAACGG	新型	NAG
CCCGTTGCCCGGCAACGGCC	新型	NGG
CCCGTTTGTCCTCTAATTCC	新型	NGG
CCCTAACAAAACAAAGAGAT	新型	NGG
CCCTAGAAGAAGAACTCCCT	新型	NGC
CCCTATCCTATCAACACTTC	新型	NGG
CCCTGCTGGTGGCTCCAGTT	新型	NAG

[1751]

向导序列	向导 id	PAM
CCCTTATCGTCAATCTTCTC	新型	NAG
CCCTTCTCCGTCTGCCGTTC	新型	NGA
CCCTTCTCGTGTTACAGGCG	新型	NGG
CCGAAAAGGTTCCACGCACG	新型	NGC
CCGAAAAGGTTCCACGCACGC	新型	NNNRRT
CCGAAGGTTTGGTACAGCAA	新型	NAG
CCGACCTTGAGGCATACTTC	新型	NAA
CCGCAGGATTCAGCGCCGAC	新型	NGG
CCGCAGTATGGATCGGCAGA	E7	NGA
CCGCCTCAGCTCTGTATCGG	新型	NAA
CCGCCTTCCATAGAGTGTGT	新型	NAA
CCGCCTTCCATAGAGTGTGTA	M19	NNNRRT
CCGCGCAGGATCCAGTTGGC	新型	NGC
CCGCTGTCTCCACCTTTGAG	新型	NAA
CCGCTTGTTTTGCTCGCAGC	新型	NGG
CCGGAAGTGTTGATAGGATA	新型	NGG
CCGGCAACGGCCAGGTCTGTG	M32	NNNRRT
CCGTCGGCGCTGAATCCTGC	新型	NGA
CCGTGCGGTGGGGTGAAACC	新型	NAG
CCGTGGTCGGTCGGAACGGC	新型	NGA
CCGTGTGTCTTGGCCAAAATT	新型	NNNRRT
CCGTTGCCGGGCAACGGGGT	新型	NAA
CCGTTTCTCCTGGCTCAGTTT	新型	NNNRRT
CCGTTTGTCCTCTAATTCCA	新型	NGA
CCTAACAAAACAAAGAGATG	新型	NGG
CCTAATATACATTTACACCA	新型	NGA
CCTACAGCCTCCTAGTACAA	新型	NGA
CCTACGAACCACTGAACAAA	新型	NGG
CCTAGAAAATTGAGAGAAGT	新型	NNACCA

[1752]

向导序列	向导 id	PAM
CCTATCCTATCAACACTTCC	新型	NGA
CCTCAATGTTAGTATTCCTT	新型	NGA
CCTCACAATACCGCAGAGTC	新型	NAG
CCTCCAAGCTGTGCCTTGGG	新型	NGG
CCTCCACCAATCGCCAGACA	M83	NGA
CCTCCAGCTTATAGACCACC	新型	NAA
CCTCCCAGTCTTTAAACAAA	新型	NAG
CCTCCTAGTACAAAGACCTT	新型	NAA
CCTCGAGAAGATTGACGATA	新型	NGG
CCTCGCCTCGCAGACGAAGGT	新型	NNNRRT
CCTCTATGTATCCCTCCTGT	新型	NGC
CCTCTCACTCTGGGATCTTG	新型	NAG
CCTCTGCCGATCCATACTGC	E8	NGA
CCTCTGGGATTCTTTCCCGA	新型	NNACCA
CCTGAACTGGAGCCACCAGC	新型	NGG
CCTGAACTTTAGGCCCATAT	新型	NAG
CCTGAGGATGAGTGTTTCTC	新型	NAA
CCTGAGGGCTCCACCCCAAA	新型	NGG
CCTGCATGACTACTGCTCAA	新型	NGA
CCTGCCTCGTCGTCTAACAA	新型	NAG
CCTGCCTCGTCGTCTAACAAC	新型	NNNRRT
CCTGCGGACGACCCTTCTCG	新型	NGG
CCTGCGTTAATGCCCTTGTA	新型	NGC
CCTGCTGCGAGCAAAACAAG	新型	NGG
CCTGCTGGTGGCTCCAGTTC	新型	NGG
CCTGGAATTAGAGGACAAAC	新型	NGG
CCTGGCCAGACGCCAACAAG	新型	NNAGGA
CCTGGCCGTTGCCGGGCAAC	新型	NGG
CCTGGGTGGGTGTTAATTTG	新型	NAA

[1753]

向导序列	向导 id	PAM
CCTGTCTGGCGATTGGTGGA	新型	NGC
CCTGTCTTTAATCCTCATTG	新型	NAA
CCTGTTAATAGGAAGTTTTC	新型	NAA
CCTTATCGTCAATCTTCTCG	新型	NGG
CCTTATTATCCAGAACATCTA	新型	NNNRRT
CCTTCAGTACGAGATCTTCT	新型	NGA
CCTTCCATAGAGTGTGTAAA	新型	NAG
CCTTCCTGTCTGGCGATTGG	新型	NGG
CCTTCGTCTGCGAGGCGAGG	新型	NAG
CCTTGATAGTCCAGAAGAAC	新型	NAA
CCTTGGACTCATAAGGTGGG	新型	NAA
CCTTGGGTGGCTTTGGGGCA	新型	NGG
CCTTGTTGGGATTGAAGTCC	新型	NAA
CCTTTGGATAAAACCTAGCA	新型	NGC
CCTTTTGGGGTGGAGCCCTC	新型	NGG
CGAAGGTTTGGTACAGCAAC	新型	NGG
CGAAGTGCACACGGTCCGGC	新型	NGA
CGAATTTTGGCCAAGACACA	新型	NGG
CGAATTTTGGCCAAGACACAC	新型	NNNRRT
CGACCCCGAGAAGGGTCGTC	新型	NGC
CGACCCTTATAAAGAATTTG	新型	NAG
CGACCCTTCTCGGGGTCGCT	新型	NGG
CGACCTTGAGGCATACTTCA	新型	NAG
CGACGAGGCAGGTCCCCTAG	新型	NAG
CGACGTGCAGAGGTGAAGCG	新型	NAG
CGAGAAAGTGAAAGCCTGCT	新型	NAG
CGAGAAGATTGACGATAAGG	新型	NAG
CGAGCAAAACAAGCGGCTAG	新型	NAG
CGAGGCAGGTCCCCTAGAAG	新型	NAG

[1754]

向导序列	向导 id	PAM
CGAGGGAGTTCTTCTTCTAG	新型	NGG
CGATAACCAGGACAAGTTGG	M58	NGG
CGATGTATATTTGCGGGAGA	新型	NGA
CGATTGAGACCTTCGTCTGC	新型	NAG
CGATTGGTGGAGGCAGGAGG	新型	NGG
CGCAACCCCCACTGGCTGGGG	新型	NNNRRT
CGCACGGAGGCCTTTTGGGG	新型	NGG
CGCAGAAGATCTCAATCTCG	新型	NGA
CGCAGACCAATTTATGCCTA	新型	NAG
CGCAGAGTCTAGACTCGTGG	M35	NGG
CGCAGGATAACCACATTGTG	新型	NAA
CGCAGGATTCAGCGCCGACG	新型	NGA
CGCAGGGTCCCCAATCCTCG	新型	NGA
CGCAGGGTCCCCAATCCTCGA	新型	NNNRRT
CGCAGTATGGATCGGCAGAG	新型	NAG
CGCATGGAGACCACCGTGAA	新型	NGC
CGCCAACAAGGTAGGAGCTG	新型	NAG
CGCCCACCGAATGTTGCCCA	E95	NGG
CGCCCCGTGGTCGGTCGGAA	新型	NGG
CGCCGACGGGACGTAAACAA	新型	NGG
CGCCGCGTCGCAGAAGATCT	新型	NAA
CGCCTCAGCTCTGTATCGGG	新型	NAG
CGCCTCCTGCCTCCACCAAT	新型	NGC
CGCCTCGCAGACGAAGGTCT	新型	NAA
CGCGCTGATGGCCCATGACC	新型	NAG
CGCGTAAAGAGAGGTGCGCCC	新型	NNNRRT
CGCGTGCGTGGAACCTTTTC	新型	NGC
CGGAAAATTCCTATGGGAGT	新型	NGG
CGGAAACTACTGTTGTTAGA	新型	NGA

[1755]

向导序列	向导 id	PAM
CGGAACGGCAGACGGAGAAG	新型	NGG
CGGAAGTGTTGATAGGATAG	新型	NGG
CGGACGACCCTTCTCGGGGT	新型	NGC
CGGCAGACGGAGAAGGGGAC	新型	NAG
CGGCGATTGAGACCTTCGTC	新型	NGC
CGGCTAGGAGTTCCGCAGTA	新型	NGG
CGGGAAGCCTTAGAGTCTCC	新型	NGA
CGGGCAACGGGGTAAAGGTT	新型	NAG
CGGGCTGAGGCCCACTCCCA	新型	NAG
CGGGCTGAGGCCCACTCCCAT	新型	NNGRRT
CGGGCTGAGGCCCACTCCCAT	新型	NNNRRT
CGGGCTGGGTTTCACCCCAC	新型	NGC
CGGGGAGTCCGCGTAAAGAG	新型	NGG
CGGGGTTTTTCTTGTTGACA	新型	NGA
CGGGTATATTATATAAGAGA	新型	NAA
CGGTAAAAAGGGACTCAAGA	新型	NGC
CGGTCGGAACGGCAGACGGA	新型	NAA
CGGTGGGGTGAAACCCAGCC	新型	NGA
CGGTTTCTCTTCCAAAAGTG	新型	NGA
CGTAAACAAAGGACGTCCCG	新型	NGC
CGTACTGAAGGAAAGAAGTC	新型	NGA
CGTCAATCTTCTCGAGGATT	新型	NGG
CGTCAGCAAACACTTGGCAC	新型	NGA
CGTCCCGCGCAGGATCCAGT	新型	NGG
CGTCCGAAGGTTTGGTACAG	新型	NAA
CGTCGCATGGAGACCACCGT	新型	NAA
CGTCTAACAACAGTAGTTTC	新型	NGG
CGTCTAACAACAGTAGTTTCC	新型	NNNRRT
CGTCTGGCCAGGTGTCCTTGT	新型	NNGRRT

[1756]

向导序列	向导 id	PAM
CGTCTGGCCAGGTGTCCTTGT	新型	NNNRRT
CGTGAACGCCCACCGAATGT	新型	NGC
CGTGCAGAGGTGAAGCGAAG	新型	NGC
CGTGCGGTGGGGTGAAACCC	新型	NGC
CGTGCTGGTGGTTGAGGATCC	新型	NNGRRT
CGTGCTGGTGGTTGAGGATCC	新型	NNNRRT
CGTGTGTCTTGGCCAAAATT	新型	NGC
CGTTCACGGTGGTCTCCATG	新型	NGA
CGTTCCGACCGACCACGGGG	新型	NGC
CGTTGACATTGCAGAGAGTC	新型	NAA
CGTTGACATTGCAGAGAGTCC	新型	NNGRRT
CGTTGACATTGCAGAGAGTCC	新型	NNNRRT
CGTTGCCGGGCAACGGGGTA	新型	NAG
CTAACAACAGTAGTTTCCGG	新型	NAG
CTAATATGGGCCTAAAGTTC	新型	NGG
CTAATGACTCTAGCTACCTGG	新型	NNGRRT
CTAATGACTCTAGCTACCTGG	新型	NNNRRT
CTACCTTGTTGGCGTCTGGC	新型	NAG
CTACGAACCACTGAACAAAT	新型	NGC
CTACTAGGTCTCTAGATGCT	新型	NGA
CTACTGTTGTTAGACGACGA	新型	NGC
CTACTGTTGTTAGACGACGAG	新型	NNNRRT
CTAGAAGATCTCGTACTGAA	新型	NGA
CTAGACTCGTGGTGGACTTCT	E9	NNNRRT
CTAGACTCTGCGGTATTGTG	新型	NGG
CTAGAGTCATTAGTTCCCCC	新型	NAG
CTAGCAGGCATAATCAATTG	新型	NAA
CTAGCCGCTTGTTTTGCTCG	新型	NAG
CTAGCTACCTGGGTGGGTGT	新型	NAA

[1757]

向导序列	向导 id	PAM
CTAGGTTTTATCCAAAGGTTA	新型	NNNRRT
CTAGTAAACTGAGCCAGGAG	新型	NAA
CTAGTAGTCAGTTATGTCAAC	新型	NNNRRT
CTATAACGGTTTCTCTTCCA	新型	NAA
CTATCCTATCAACACTTCCG	新型	NAA
CTATGTGTTGTTTCTCTCTTA	新型	NNNRRT
CTATTAACAGGCCTATTGAT	新型	NGG
CTATTGATTGGAAAGTATGT	新型	NAA
CTATTTACACACTCTATGGA	新型	NGG
CTCAAGATGCTGTACAGACT	新型	NGG
CTCAATCGCCGCGTCGCAGA	M77	NGA
CTCAATCTCGGGAACCTCAAT	新型	NNNRRT
CTCACAATACCGCAGAGTCT	新型	NGA
CTCACCAACCTCCTGTCCTC	新型	NAA
CTCACCATACTGCACTCAGG	新型	NAA
CTCAGGCTCAGGGCATACTA	新型	NAA
CTCATCCTCAGGCCATGCAG	新型	NGG
CTCCAACTTGTCCTGGTTAT	新型	NGC
CTCCAAGCTGTGCCTTGGGT	新型	NGC
CTCCACCAATCGCCAGACAG	新型	NAA
CTCCACCCCAAAAGGCCTCCG	新型	NNNRRT
CTCCAGACCTGCTGCGAGCA	新型	NAA
CTCCATGCGACGTGCAGAGG	M92	NGA
CTCCATGTTCAGCGCAGGGTC	新型	NNNRRT
CTCCCATAGGAATTTTCCGA	新型	NAG
CTCCCTCGCCTCGCAGACGA	新型	NGG
CTCCTACCTTGTTGGCGTCT	新型	NGC
CTCCTGCCTCCACCAATCGC	新型	NAG
CTCCTGGCTCAGTTTACTAG	新型	NGC

[1758]

向导序列	向导 id	PAM
CTCGAGAAGATTGACGATAA	新型	NGG
CTCGAGGATTGGGGACCCTG	新型	NGC
CTCTAAGGCTTCCCGATACA	新型	NAG
CTCTAGATGCTGGATCTTCC	新型	NAA
CTCTATAACGGTTTCTCTTC	新型	NAA
CTCTATAACGGTTTCTCTTCC	新型	NNNRRT
CTCTCACTCTGGGATCTTGC	新型	NGA
CTCTCGTCCCCTTCTCCGTC	新型	NGC
CTCTCTGCAATGTCAACGAC	新型	NGA
CTCTGAAGGCTGGATCCAAC	新型	NGG
CTCTGAAGGCTGGATCCAACT	新型	NNNRRT
CTCTGCAAGATCCCAGAGTG	新型	NGA
CTCTGCCGATCCATACTGCG	新型	NAA
CTCTGGGATCTTGCAGAGTT	新型	NGG
CTCTGGGATTCTTTCCCGACC	新型	NNNRRT
CTCTGTATCGGGAAGCCTTA	新型	NAG
CTCTTATGTAAGACCTTGGG	新型	NAA
CTCTTCCAAAAGTGAGACAA	新型	NAA
CTCTTTGTTTTGTTAGGGTT	新型	NAA
CTGAAAGCCAAACAGTGGGG	新型	NAA
CTGAACTGGAGCCACCAGCA	新型	NGG
CTGAACTGGAGCCACCAGCAG	新型	NNNRRT
CTGAAGGAAAGAAGTCAGAA	新型	NGC
CTGACGCAACCCCCACTGGC	新型	NGG
CTGACTACTAGGTCTCTAGA	新型	NGC
CTGACTTCTTTCCTTCAGTA	新型	NGA
CTGAGCCAGGAGAAACGGGC	M70	NGA
CTGAGGATGAGTGTTTCTCA	新型	NAG
CTGAGGGCTCCACCCCAAAA	新型	NGC

[1759]

向导序列	向导 id	PAM
CTGAGTGCAGTATGGTGAGG	新型	NGA
CTGCAAGATCCCAGAGTGAG	新型	NGG
CTGCATGACTACTGCTCAAG	新型	NAA
CTGCCAACTGGATCCTGCGC	M191	NGG
CTGCCGGACCGTGTGCACTT	新型	NGC
CTGCCGTTCCGACCGACCAC	新型	NGG
CTGCCTCCACCAATCGCCAG	新型	NNAGGA
CTGCCTTCCTGTCTGGCGAT	新型	NGG
CTGCGAATTTTGGCCAAGACA	新型	NNNRRT
CTGCGAGCAAAACAAGCGGC	新型	NAG
CTGCGAGCAAAACAAGCGGCT	新型	NNGRRT
CTGCGAGCAAAACAAGCGGCT	新型	NNNRRT
CTGCTGCGAGCAAAACAAGC	新型	NGC
CTGCTGGTGGCTCCAGTTCA	新型	NGA
CTGGAATTAGAGGACAAACG	新型	NGC
CTGGATCCAACTGGTGGTCG	新型	NGA
CTGGCCAGACGCCAACAAGG	新型	NAG
CTGGCCAGGTGTCCTTGTTG	新型	NGA
CTGGCCGTTGCCGGGCAACG	新型	NGG
CTGGCGATTGGTGGAGGCAG	新型	NAG
CTGGCTGGGGCTTGGTCATG	新型	NGC
CTGGGAGGAGTTGGGGGAGG	新型	NGA
CTGGGGGGAACTAATGACTC	新型	NAG
CTGGGTGGGTGTTAATTTGG	新型	NAG
CTGGGTTTCACCCCACCGCA	新型	NGG
CTGGTGCGCAGACCAATTTA	新型	NGC
CTGTAACACGAGAAGGGGTC	新型	NNAGGA
CTGTACAGACTTGGCCCCCA	新型	NNACCA
CTGTACCAAACCTTCGGACG	新型	NAA

[1760]

向导序列	向导 id	PAM
CTGTATTTCCCTGCTGGTGGC	新型	NNNRRT
CTGTCTGGCGATTGGTGGAG	新型	NNAGGA
CTGTCTTTAATCCTCATTGG	新型	NAA
CTGTGCCAAGTGTTTGCTGA	新型	NGC
CTGTGCCTTGGGTGGCTTTG	新型	NGG
CTGTGCTGCCAACTGGATCC	新型	NGC
CTGTGTTTGCTCTGAAGGCT	新型	NGA
CTGTTAATAGGAAGTTTTCT	新型	NAA
CTGTTCAAGCCTCCAAGCTG	新型	NGC
CTGTTGCTGTACCAAACCTT	新型	NGG
CTGTTGTTAGACGACGAGGC	新型	NGG
CTGTTTGGCTTTCAGTTATA	新型	NGG
CTGTTTGTTTAAAGACTGGG	新型	NGG
CTTACAAGGCCTTTCTGTGT	新型	NAA
CTTACAAGGCCTTTCTGTGTA	新型	NNNRRT
CTTACAGTTAATGAGAAAAG	新型	NAG
CTTATCCAATGGTAAATATT	新型	NGG
CTTATCGTCAATCTTCTCGA	新型	NGA
CTTATGAGTCCAAGGAATAC	新型	NAA
CTTCAAAGACTGTTTGTTTA	新型	NAG
CTTCACCTCTGCACGTCGCA	新型	NGG
CTTCCAAATTAACACCCACC	新型	NAG
CTTCCAATGACATAACCCAT	新型	NAA
CTTCCGGAAACTACTGTTGT	新型	NAG
CTTCCTATTAACAGGCCTAT	新型	NGA
CTTCCTGTCTGGCGATTGGT	新型	NGA
CTTCTCTCAATTTTCTAGGG	M87	NGA
CTTCTGCGACGCGGCGATTG	新型	NGA
CTTCTTTTCTCATTAACTGT	新型	NAG

[1761]

向导序列	向导 id	PAM
CTTGAACAGTAGGACATGAA	新型	NAA
CTTGCCTGAGTGCAGTATGG	新型	NGA
CTTGGCACAGACCTGGCCGT	新型	NGC
CTTGGGTGGCTTTGGGGCAT	新型	NGA
CTTGGTGTAAATGTATATTA	新型	NGA
CTTGTAAGTTGGCGAGAAAG	新型	NGA
CTTGTCAACAAGAAAAACCC	新型	NGC
CTTGTCTCACTTTTGGAAGA	新型	NAA
CTTGTTCATGTCCTACTGTT	新型	NAA
CTTGTTGACAAGAATCCTCA	新型	NAA
CTTTAAACAAACAGTCTTTG	新型	NAG
CTTTAGGCCCATATTAGTGT	新型	NGA
CTTTCCACCAGCAATCCTCT	新型	NGG
CTTTCGGAAAATTCCTATGG	新型	NAG
CTTTCTGTGTAAACAATACC	新型	NGA
CTTTGAGAAACACTCATCCT	新型	NAG
CTTTGGATAAAACCTAGCAGG	新型	NNNRRT
CTTTGTACTAGGAGGCTGTA	新型	NGC
CTTTGTTTACGTCCCGTCGG	新型	NGC
CTTTTCTCATTAACTGTAAG	新型	NGG
CTTTTGGAAGAGAAACCGTTA	新型	NNGRRT
CTTTTGGAAGAGAAACCGTTA	新型	NNNRRT
CTTTTGGGGTGGAGCCCTCA	新型	NGC
GAAAAACCCCGCCTGTAACA	新型	NGA
GAAAAGATGGTGTTTTCCAA	新型	NGA
GAAAAGATGGTGTTTTCCAA	新型	NNAGGA
GAAAAGATGGTGTTTTCCAAT	新型	NNGRRT
GAAAAGATGGTGTTTTCCAAT	新型	NNNRRT
GAAAATTGAGAGAAGTCCACC	新型	NNGRRT

[1762]

向导序列	向导 id	PAM
GAAAATTGAGAGAAGTCCACC	新型	NNNRRT
GAAAATTGGTAACAGCGGTA	新型	NAA
GAAACACTCATCCTCAGGCCA	新型	NNNRRT
GAAACCCAGCCCGAATGCTC	新型	NAG
GAAAGACACCAAATACTCTA	新型	NAA
GAAAGACACCAAATACTCTAT	新型	NNNRRT
GAAAGCCAAACAGTGGGGGA	新型	NAG
GAAAGCCCAGGATGATGGGA	M37	NGG
GAAAGCCCAGGATGATGGGAT	M4	NNGRRT
GAAAGCCCAGGATGATGGGAT	新型	NNNRRT
GAAAGCCCTACGAACCACTGA	新型	NNNRRT
GAAAGGCCTTGTAAGTTGGC	新型	NAG
GAAAGTATGTCAACGAATTG	新型	NGG
GAAAGTGAAAGCCTGCTTAGA	新型	NNGRRT
GAAAGTGAAAGCCTGCTTAGA	新型	NNNRRT
GAAATACAGGCCTCTCACTC	新型	NGG
GAACAAGAGATGATTAGGCA	新型	NAG
GAACAAGATCTACAGCATGG	新型	NGC
GAACAATGCTCAGGAGACTC	新型	NAA
GAACACATCATACAAAAAAT	新型	NAA
GAACAGGGTTTACTGCTCCT	新型	NAA
GAACAGTAGGACATGAACAA	新型	NAG
GAACATCACATCAGGATTCC	新型	NAG
GAACATGGAGAACATCACAT	新型	NAG
GAACCTGCATGACTACTGCT	新型	NAA
GAACCTGCATGACTACTGCT	新型	NNAGGA
GAACCTTTACCCCGTTGCCC	新型	NGC
GAACGCCCACCGAATGTTGCC	新型	NNNRRT
GAACTACCGTGTGTCTTGGC	新型	NAA

[1763]

向导序列	向导 id	PAM
GAACTCCCTCGCCTCGCAGA	新型	NGA
GAACTCCCTCGCCTCGCAGAC	E10	NNNRRT
GAACTCCTAGCCGCTTGTTT	新型	NGC
GAACTGGAGCCACCAGCAGG	新型	NAA
GAAGAACCAACAAGAAGATG	新型	NGG
GAAGAACTCCCTCGCCTCGC	E11	NGA
GAAGAAGAACTCCCTCGCCT	新型	NGC
GAAGATCTCAATCTCGGGAAC	M14	NNNRRT
GAAGATCTCGTACTGAAGGA	新型	NAG
GAAGATCTCGTACTGAAGGAA	新型	NNNRRT
GAAGATGAGGCATAGCAGCA	新型	NGA
GAAGATTGACGATAAGGGAG	新型	NGG
GAAGATTGACGATAAGGGAGA	新型	NNNRRT
GAAGCGAAGTGCACACGGTC	新型	NGG
GAAGGAAAGAAGTCAGAAGG	新型	NAA
GAAGGCACAGACGGGGAGTC	新型	NGC
GAAGGCGGGTATATTATATA	新型	NGA
GAAGGGGACGAGAGAGTCCC	新型	NAG
GAAGGGTCGTCCGCAGGATT	新型	NAG
GAAGGTCTCAATCGCCGCGT	新型	NGC
GAAGGTTTGGTACAGCAACA	新型	NGA
GAAGTCAGAAGGCAAAAACG	新型	NGA
GAAGTTATGGGTCCTTGCCA	新型	NAA
GAATACTAACATTGAGGTTCC	新型	NNNRRT
GAATATGGTGACCCACAAAA	新型	NGA
GAATCCACACTCCGAAAGACA	M12	NNNRRT
GAATCCCAGAGGATTGCTGG	新型	NGG
GAATTGCTTGCCTGAGTGCAG	新型	NNNRRT
GACAAAAGAAAATTGGTAAC	新型	NGC

[1764]

向导序列	向导 id	PAM
GACAAACGGGCAACATACCT	新型	NGA
GACAAACGGGCAACATACCTT	新型	NNNRRT
GACAACAGAGTTATCAGTCC	新型	NGA
GACAACAGAGTTATCAGTCCC	新型	NNNRRT
GACAAGAAATGTGAAACCAC	新型	NAG
GACAAGAATCCTCACAATAC	新型	NGC
GACACACGGTAGTTCCCCCT	新型	NGA
GACACACGGTAGTTCCCCCTA	新型	NNNRRT
GACACATCCAGCGATAACCA	M67	NGA
GACACCTGGCCAGACGCCAA	新型	NAA
GACACTATTTACACACTCTA	新型	NGG
GACAGGAAGGCAGCCTACCC	新型	NGC
GACAGGTACAGTAGAAGAAT	新型	NAA
GACATAACCCATAAAATTCA	新型	NAG
GACATACTTTCCAATCAATA	新型	NGC
GACATGAACAAGAGATGATT	新型	NGG
GACCAAGCCCCAGCCAGTGG	新型	NGG
GACCACCGTGAACGCCCACC	新型	NAA
GACCCCGAGAAGGGTCGTCC	新型	NNAGGA
GACCCCGAGAAGGGTCGTCCG	新型	NNGRRT
GACCCCGAGAAGGGTCGTCCG	新型	NNNRRT
GACCCCTTCTCGTGTTACAGG	新型	NNGRRT
GACCCCTTCTCGTGTTACAGG	新型	NNNRRT
GACCCTGCGCTGAACATGGA	新型	NAA
GACCCTTATAAAGAATTTGG	新型	NGC
GACCCTTCTCGGGGTCGCTT	新型	NGG
GACCTAGTAGTCAGTTATGT	新型	NAA
GACCTGCTGCGAGCAAAACA	新型	NGC
GACCTGGCCGTTGCCGGGCAA	新型	NNGRRT

[1765]

向导序列	向导 id	PAM
GACCTGGCCGTTGCCGGGCAA	新型	NNNRRT
GACCTTCGTCTGCGAGGCGA	新型	NGG
GACCTTGAGGCATACTTCAA	新型	NGA
GACCTTGGGCAACATTCGGT	新型	NGG
GACGACGAGGCAGGTCCCCT	M74	NGA
GACGAGGCAGGTCCCCTAGA	M75	NGA
GACGCAACCCCCACTGGCTG	新型	NGG
GACGCCAACAAGGTAGGAGC	M53	NGG
GACGGGGAGTCCGCGTAAAG	新型	NGA
GACGTAAACAAAGGACGTCC	新型	NGC
GACTACTGCCTCTCCCTTATC	新型	NNNRRT
GACTCGTGGTGGACTTCTCT	新型	NAA
GACTCTAAGGCTTCCCGATA	新型	NAG
GACTGGGAGGAGTTGGGGGA	新型	NGA
GACTGTTTGTTTAAAGACTG	新型	NGA
GACTGTTTGTTTAAAGACTG	新型	NNAGGA
GACTTCTCTCAATTTTCTAG	E12	NGG
GACTTCTTTCCTTCAGTACG	新型	NGA
GAGAAAGTGAAAGCCTGCTT	新型	NGA
GAGAAGATTGACGATAAGGG	新型	NGA
GAGAAGTCCACCACGAGTCT	M66	NGA
GAGACCTTCGTCTGCGAGGC	新型	NAG
GAGAGAGTCCCAAGCGACCC	新型	NGA
GAGAGGCAGTAGTCGGAACA	新型	NGG
GAGAGTAACTCCACAGTAGCT	新型	NNNRRT
GAGAGTCCCAAGCGACCCCG	新型	NGA
GAGATGATTAGGCAGAGGTG	新型	NAA
GAGATGATTAGGCAGAGGTGA	新型	NNNRRT
GAGATTGAGATCTTCTGCGA	新型	NGC

[1766]

向导序列	向导 id	PAM
GAGCAGTAGTCATGCAGGTT	新型	NGG
GAGCCACCAGCAGGGAAATA	新型	NAG
GAGCCAGGAGAAACGGGCTG	新型	NGG
GAGCCTGAGGGCTCCACCCC	新型	NAA
GAGCTGAGGCGGTATCTAGA	新型	NGA
GAGGACAAACGGGCAACATAC	新型	NNNRRT
GAGGACAACAGAGTTATCAGT	新型	NNNRRT
GAGGACAGGAGGTTGGTGAG	新型	NGA
GAGGACTCTTGGACTCTCTG	新型	NAA
GAGGAGCCGAAAAGGTTCCA	新型	NGC
GAGGAGTTGGGGGAGGAGAT	新型	NAG
GAGGATTCTTGTCAACAAGA	新型	NAA
GAGGATTGGGGACCCTGCGC	新型	NGA
GAGGCAGGAGGCGGATTTGC	新型	NGG
GAGGCAGGTCCCCTAGAAGA	新型	NGA
GAGGCATAGCAGCAGGATGA	新型	NGA
GAGGCATAGCAGCAGGATGA	新型	NNAGGA
GAGGCTTGAACAGTAGGACA	新型	NGA
GAGGGAGTTCTTCTTCTAGG	新型	NGA
GAGGTGAAAAAGTTGCATGG	新型	NGC
GAGGTGAAAAAGTTGCATGGT	新型	NNNRRT
GAGGTGAAGCGAAGTGCACA	新型	NGG
GAGTAACTCCACAGTAGCTC	新型	NAA
GAGTCATTAGTTCCCCCCAG	新型	NAA
GAGTCCAAGGAATACTAACAT	新型	NNNRRT
GAGTCCCAAGCGACCCCGAG	新型	NAG
GAGTCCCTTTTTACCGCTGTT	新型	NNNRRT
GAGTGCGAATCCACACTCCG	新型	NAA
GAGTTTGGTGAAAGGTTGTG	新型	NAA

[1767]

向导序列	向导 id	PAM
GATAACCACATTGTGTAAAT	新型	NGG
GATAACCAGGACAAGTTGGA	M89	NGA
GATAAGGGAGAGGCAGTAGT	新型	NGG
GATAATGTTTGCTCCAGACC	新型	NGC
GATACGATGTATATTTGCGG	新型	NAG
GATAGGATAGGGGCATTTGG	新型	NGG
GATAGTCCAGAAGAACCAAC	新型	NAG
GATCCAACTGGTGGTCGGGA	新型	NAG
GATCCATACTGCGGAACTCC	新型	NAG
GATCCTCAACCACCAGCACG	新型	NGA
GATCCTCAACCACCAGCACG	新型	NNACCA
GATCGGCAGAGGAGCCGAAA	新型	NGG
GATCTCAATCTCGGGAACCT	新型	NAA
GATCTCGTACTGAAGGAAAG	新型	NAG
GATCTTCTAGATACCGCCTC	新型	NGC
GATCTTGCAGAGTTTGGTGA	新型	NAG
GATGAGAAGGCACAGACGGG	新型	NAG
GATGAGGCATAGCAGCAGGA	新型	NGA
GATGAGTGTTTCTCAAAGGT	新型	NGA
GATGATGTGGTATTGGGGGC	新型	NAA
GATGATTAGGCAGAGGTGAA	新型	NAA
GATGGCCCATGACCAAGCCC	新型	NAG
GATGTGATGTTCTCCATGTT	新型	NAG
GATGTTCTCCATGTTCAGCG	新型	NAG
GATTAAAGACAGGTACAGTA	新型	NAA
GATTAAAGACAGGTACAGTAG	新型	NNGRRT
GATTAAAGACAGGTACAGTAG	新型	NNNRRT
GATTAAAGGTCTTTGTACTA	新型	NGA
GATTAACTAGATGTTCTGGA	新型	NAA

[1768]

向导序列	向导 id	PAM
GATTCAGCGCCGACGGGACG	新型	NAA
GATTGAATACATGCATACAA	新型	NGG
GATTGACGATAAGGGAGAGG	新型	NAG
GATTGACGATAAGGGAGAGGC	新型	NNNRRT
GATTGAGACCTTCGTCTGCG	新型	NGG
GATTGAGATCTTCTGCGACG	新型	NGG
GATTGAGATCTTCTGCGACGC	新型	NNNRRT
GATTGCAATTGATTATGCCTG	M18	NNNRRT
GATTGCTGGTGGAAAGATTC	新型	NGC
GATTGGAGGTTGGGGACTGC	新型	NAA
GATTGGGACTTCAATCCCAA	新型	NAA
GATTGGGACTTCAATCCCAA	新型	NNAGGA
GATTGGTGGAGGCAGGAGGC	新型	NGA
GATTTGCTGTGTTTGCTCTG	新型	NAG
GCAAACACTTGGCACAGACC	新型	NGG
GCAAACATTATCGGGACTGA	新型	NAA
GCAAAGAATTGCTTGCCTGAG	新型	NNNRRT
GCAAAGTTTGTAGTATGCCC	新型	NGA
GCAAATATACATCGTATCCA	新型	NGG
GCAAATCCAGATTGGGACTT	新型	NAA
GCAAATCCGCCTCCTGCCTCC	新型	NNNRRT
GCAACAGGAGGGATACATAG	新型	NGG
GCAACATACCTTGATAGTCC	新型	NGA
GCAACGGCCAGGTCTGTGCC	新型	NAG
GCAACTTTTTCACCTCTGCC	新型	NAA
GCAAGCAATTCTTTGCTGGG	M45	NGG
GCAATGTCAACGACCGACCT	新型	NGA
GCAATTTCCGTCCGAAGGTT	新型	NGG
GCACACGGTCCGGCAGATGA	新型	NAA

[1769]

向导序列	向导 id	PAM
GCACAGACCTGGCCGTTGCC	新型	NGG
GCACAGACGGGGAGTCCGCG	新型	NAA
GCACAGCCTAGCAGCCATGGA	新型	NNNRRT
GCACAGCTTGGAGGCTTGAA	新型	NAG
GCACGGAGGCCTTTTGGGGT	新型	NGA
GCACGGGACCATGCCGAACC	新型	NGC
GCACTAGTAAACTGAGCCAG	新型	NAG
GCACTCCTCCAGCTTATAGA	新型	NNACCA
GCAGAAGATCTCAATCTCGG	新型	NAA
GCAGACACATCCAGCGATAA	新型	NNAGGA
GCAGACCAATTTATGCCTAC	新型	NGC
GCAGACGGAGAAGGGGACGA	新型	NAG
GCAGAGGTGAAAAAGTTGCA	新型	NGG
GCAGAGGTGAAGCGAAGTGCA	新型	NNNRRT
GCAGAGTCTAGACTCGTGGT	M65	NGA
GCAGATGAGAAGGCACAGAC	新型	NGG
GCAGATGAGAAGGCACAGACG	新型	NNGRRT
GCAGATGAGAAGGCACAGACG	新型	NNNRRT
GCAGCACAGCCTAGCAGCCA	新型	NGG
GCAGGAGGCGGATTTGCTGG	新型	NAA
GCAGGATAACCACATTGTGT	新型	NAA
GCAGGATCCAGTTGGCAGCA	新型	NAG
GCAGGCTTTCACTTTCTCGC	新型	NAA
GCAGGGTCCCCAATCCTCGA	新型	NAA
GCAGGTTCGGCATGGTCCCG	新型	NGC
GCAGGTTCGGCATGGTCCCGT	新型	NNNRRT
GCAGTAGTCATGCAGGTTCGG	新型	NNNRRT
GCAGTATGGATCGGCAGAGG	新型	NGC
GCATAAATTGGTCTGCGCAC	新型	NAG

[1770]

向导序列	向导 id	PAM
GCATACAAGGGCATTAACGC	新型	NGG
GCATAGCAGCAGGATGAAGA	新型	NGA
GCATAGCAGCAGGATGAAGAG	新型	NNNRRT
GCATCTAGAGACCTAGTAGT	新型	NAG
GCATGGACATCGACCCTTAT	新型	NAA
GCATGGACATCGACCCTTATA	新型	NNGRRT
GCATGGACATCGACCCTTATA	新型	NNNRRT
GCATGTATTCAATCTAAGCA	新型	NGC
GCATTTGGTGGTCTATAAGC	新型	NGG
GCCAACAAGGTAGGAGCTGG	新型	NGC
GCCAAGTCTGTACAGCATCT	新型	NGA
GCCACAAGAACACATCATAC	新型	NAA
GCCACAAGAACACATCATACA	M28	NNNRRT
GCCACCAGCAGGGAAATACA	新型	NGC
GCCACCCAAGGCACAGCTTG	新型	NAG
GCCAGACGCCAACAAGGTAG	新型	NAG
GCCAGGTGTCCTTGTTGGGAT	新型	NNNRRT
GCCAGTGGGGGTTGCGTCAG	新型	NAA
GCCATTTGTTCAGTGGTTCG	新型	NAG
GCCCAAGGTCTTACATAAGA	新型	NGA
GCCCACTTACAGTTAATGAG	新型	NAA
GCCCATGACCAAGCCCCAGC	新型	NAG
GCCCCGTGGTCGGTCGGAAC	新型	NGC
GCCCGTTTGTCCTCTAATTC	新型	NAG
GCCGACGGGACGTAAACAAA	新型	NGA
GCCGCAGACACATCCAGCGA	新型	NAA
GCCGCAGACACATCCAGCGA	新型	NNACCA
GCCGCTTGTTTTGCTCGCAG	新型	NAG
GCCGGGCAACGGGGTAAAGGT	新型	NNNRRT

[1771]

向导序列	向导 id	PAM
GCCGTTCCGACCGACCACGG	新型	NGC
GCCGTTGCCGGGCAACGGGG	新型	NAA
GCCGTTGCCGGGCAACGGGGT	新型	NNNRRT
GCCTAAAGTTCAGGCAACTCT	新型	NNNRRT
GCCTACAAACTGTTCACATTT	新型	NNNRRT
GCCTACAGCCTCCTAGTACA	新型	NAG
GCCTCAGCCCGTTTCTCCTGG	新型	NNNRRT
GCCTCAGCTCTGTATCGGGA	新型	NGC
GCCTCCACCAATCGCCAGAC	新型	NGG
GCCTCGTCGTCTAACAACAG	新型	NAG
GCCTCTCACTCTGGGATCTTG	新型	NNGRRT
GCCTCTCACTCTGGGATCTTG	新型	NNNRRT
GCCTGAGGATGAGTGTTTCT	新型	NAA
GCCTGAGGATGAGTGTTTCTC	新型	NNNRRT
GCCTGAGGGCTCCACCCCAA	新型	NAG
GCCTGCTAGGTTTTATCCAA	新型	NGG
GCCTGCTTAGATTGAATACA	新型	NGC
GCCTGTATTTCCCTGCTGGT	新型	NGC
GCCTTCAGAGCAAACACAGC	新型	NAA
GCCTTCTGACTTCTTTCCTT	新型	NAG
GCCTTTTGGGGTGGAGCCCT	新型	NAG
GCGAAGTGCACACGGTCCGG	新型	NAG
GCGAATCCACACTCCGAAAG	新型	NNACCA
GCGACGTGCAGAGGTGAAGC	新型	NAA
GCGAGCAAAACAAGCGGCTA	新型	NGA
GCGAGGGAGTTCTTCTTCTA	新型	NGG
GCGATAACCAGGACAAGTTG	新型	NAG
GCGATTGAGACCTTCGTCTG	新型	NGA
GCGCAGGGTCCCCAATCCTC	新型	NAG

[1772]

向导序列	向导 id	PAM
GCGCCGACGGGACGTAAACA	新型	NAG
GCGCGTGCGTGGAACCTTTT	新型	NGG
GCGCTGATGGCCCATGACCA	新型	NGC
GCGGGAGAGGACAACAGAGTT	新型	NNNRRT
GCGGGGTTTTTCTTGTTGAC	新型	NAG
GCGGGTATATTATATAAGAG	新型	NGA
GCGTCAGCAAACACTTGGCA	新型	NAG
GCTAGGCTGTGCTGCCAACT	新型	NGA
GCTATGCCTCATCTTCTTGT	新型	NGG
GCTCAGGAGACTCTAAGGCTT	新型	NNNRRT
GCTCCAGACCTGCTGCGAGC	新型	NAA
GCTCCAGCTCCTACCTTGTT	新型	NGC
GCTCCTACCTTGTTGGCGTC	新型	NGG
GCTCCTCTGCCGATCCATAC	新型	NGC
GCTCCTGAACTGGAGCCACC	新型	NGC
GCTCGCAGCAGGTCTGGAGC	新型	NAA
GCTCTGTATCGGGAAGCCTT	新型	NGA
GCTGAGGCCCACTCCCATAG	新型	NAA
GCTGCCAACTGGATCCTGCG	M190	NGG
GCTGCCCCATTTACACAATG	新型	NGG
GCTGCGAGCAAAACAAGCGG	新型	NNAGGA
GCTGCTAGGCTGTGCTGCCAA	新型	NNGRRT
GCTGCTAGGCTGTGCTGCCAA	新型	NNNRRT
GCTGCTATGCCTCATCTTCTT	E14	NNNRRT
GCTGGATCCAACTGGTGGTC	新型	NGG
GCTGGTGGCTCCAGTTCAGG	新型	NGC
GCTGGTGGTTGAGGATCCTG	新型	NAA
GCTGTACCAAACCTTCGGAC	M91	NGA
GCTGTAGATCTTGTTCCCAA	新型	NAA

[1773]

向导序列	向导 id	PAM
GCTGTAGGCATAAATTGGTC	新型	NGC
GCTGTGCCTTGGGTGGCTTT	新型	NGG
GCTGTGTTTGCTCTGAAGGC	新型	NGG
GCTTCCCGATACAGAGCTGA	新型	NGC
GCTTGCCTGAGTGCAGTATGG	新型	NNNRRT
GCTTGGAGGCTTGAACAGTA	新型	NGA
GCTTGGTCATGGGCCATCAG	新型	NGC
GCTTTCCCCCACTGTTTGGCT	新型	NNNRRT
GCTTTCGGAAAATTCCTATG	新型	NGA
GGAAAATTCCTATGGGAGTG	新型	NGC
GGAAAGAAGTCAGAAGGCAA	新型	NAA
GGAAAGATTCTGCCCCATGCT	新型	NNNRRT
GGAAAGCCCTACGAACCACT	新型	NAA
GGAAATACAGGCCTCTCACTC	新型	NNGRRT
GGAAATACAGGCCTCTCACTC	新型	NNNRRT
GGAACAAGATCTACAGCATG	M51	NGG
GGAACAGGGTTTACTGCTCC	新型	NGA
GGAACGGCAGACGGAGAAGG	新型	NGA
GGAACTAATGACTCTAGCTAC	新型	NNGRRT
GGAACTAATGACTCTAGCTAC	新型	NNNRRT
GGAACTACCGTGTGTCTTGGC	新型	NNNRRT
GGAAGATCCAGCATCTAGAGA	新型	NNNRRT
GGAAGATGATAAAACGCCGC	新型	NGA
GGAAGCCTTAGAGTCTCCTG	新型	NGC
GGAAGGCGGGTATATTATAT	新型	NAG
GGAAGTGTTGATAGGATAGG	新型	NGC
GGAATTAGAGGACAAACGGG	新型	NAA
GGACAAGTTGGAGGACAGGAG	新型	NNNRRT
GGACACCTGGCCAGACGCCAA	新型	NNNRRT

[1774]

向导序列	向导 id	PAM
GGACATGAACAAGAGATGAT	新型	NAG
GGACCCCTTCTCGTGTTACA	新型	NGC
GGACCCTGCGCTGAACATGG	新型	NGA
GGACCTGCCTCGTCGTCTAA	新型	NAA
GGACCTGCCTCGTCGTCTAAC	新型	NNNRRT
GGACTTCTCTCAATTTTCTA	E15	NGG
GGAGACCACCGTGAACGCCCA	新型	NNGRRT
GGAGACCACCGTGAACGCCCA	新型	NNNRRT
GGAGAGGACAACAGAGTTAT	新型	NAG
GGAGAGGCAGTAGTCGGAAC	新型	NGG
GGAGCAAACATTATCGGGAC	新型	NGA
GGAGCTGGAGCATTCGGGCT	新型	NGG
GGAGGTTGGTGAGTGATTGG	新型	NGG
GGAGTCCGCGTAAAGAGAGG	新型	NGC
GGAGTGCGAATCCACACTCC	新型	NAA
GGAGTTCCGCAGTATGGATC	新型	NGC
GGATAAAACCTAGCAGGCATA	新型	NNNRRT
GGATAACCACATTGTGTAAA	新型	NGG
GGATACATAGAGGTTCCTTG	新型	NGC
GGATACGATGTATATTTGCG	新型	NGA
GGATCCAACTGGTGGTCGGG	新型	NAA
GGATCCAACTGGTGGTCGGGA	新型	NNGRRT
GGATCCAACTGGTGGTCGGGA	新型	NNNRRT
GGATCCTCAACCACCAGCAC	新型	NGG
GGATCCTGGAATTAGAGGAC	新型	NAA
GGATCGGCAGAGGAGCCGAA	新型	NAG
GGATCTTGCAGAGTTTGGTG	新型	NAA
GGATGAAGAGGAAGATGATA	新型	NAA
GGATGAGTGTTTCTCAAAGG	新型	NGG

[1775]

向导序列	向导 id	PAM
GGATGATGGGATGGGAATAC	新型	NGG
GGATTAAAGACAGGTACAGT	新型	NGA
GGATTCTTTCCCGACCACCAG	新型	NNGRRT
GGATTCTTTCCCGACCACCAG	新型	NNNRRT
GGATTTGCTGGCAAAGTTTG	新型	NAG
GGATTTGCTGTGTTTGCTCT	新型	NAA
GGCAACATACCTTGATAGTC	新型	NAG
GGCAACGGCCAGGTCTGTGC	新型	NAA
GGCAAGCAATTCTTTGCTGG	M44	NGG
GGCACAGACCTGGCCGTTGC	新型	NGG
GGCACTAGTAAACTGAGCCA	新型	NGA
GGCAGACGGAGAAGGGGACG	新型	NGA
GGCAGACGGAGAAGGGGACGA	新型	NNGRRT
GGCAGACGGAGAAGGGGACGA	新型	NNNRRT
GGCAGATGAGAAGGCACAGA	新型	NGG
GGCAGCAAAACCCAAAAGACC	新型	NNNRRT
GGCAGGAGGCGGATTTGCTGG	新型	NNNRRT
GGCATACTACAAACTTTGCC	新型	NGC
GGCATACTACAAACTTTGCCA	新型	NNNRRT
GGCATAGCAGCAGGATGAAG	新型	NGG
GGCATGGACATCGACCCTTA	新型	NAA
GGCCAGACGCCAACAAGGTA	新型	NGA
GGCCCACTTACAGTTAATGA	新型	NAA
GGCCCATATTAGTGTTGACA	新型	NAA
GGCCTCAGCCCGTTTCTCCT	新型	NGC
GGCCTCCGTGCGGTGGGGTG	新型	NAA
GGCCTCTCACTCTGGGATCT	新型	NGC
GGCCTGTATTTCCCTGCTGG	新型	NGG
GGCCTTGTAAGTTGGCGAGA	新型	NAG

[1776]

向导序列	向导 id	PAM
GGCGAGAAAGTGAAAGCCTGC	新型	NNNRRT
GGCGAGGGAGTTCTTCTTCT	新型	NGG
GGCGATTGGTGGAGGCAGGA	新型	NGC
GGCGATTGGTGGAGGCAGGAG	新型	NNGRRT
GGCGATTGGTGGAGGCAGGAG	新型	NNNRRT
GGCGGATTTGCTGGCAAAGTT	新型	NNNRRT
GGCGGGGTTTTTCTTGTTGA	新型	NAA
GGCGGGGTTTTTCTTGTTGAC	新型	NNGRRT
GGCGGGGTTTTTCTTGTTGAC	新型	NNNRRT
GGCGGGTATATTATATAAGA	新型	NAG
GGCGTTCACGGTGGTCTCCA	新型	NGC
GGCTAGGAGTTCCGCAGTAT	新型	NGA
GGCTCCAGTTCAGGAGCAGT	新型	NAA
GGCTGAGGCCCACTCCCATA	新型	NGA
GGCTGGATCCAACTGGTGGT	新型	NGG
GGCTTCCCGATACAGAGCTG	新型	NGG
GGCTTCCCGATACAGAGCTGA	新型	NNNRRT
GGCTTTCAGTTATATGGATGA	新型	NNNRRT
GGCTTTCGGAAAATTCCTAT	新型	NGG
GGGAAAGAATCCCAGAGGAT	新型	NGC
GGGAAAGAATCCCAGAGGATT	新型	NNNRRT
GGGAAAGCCCTACGAACCAC	新型	NGA
GGGAACAAGATCTACAGCAT	M50	NGG
GGGAAGCCTTAGAGTCTCCT	新型	NAG
GGGACCATGCCGAACCTGCA	新型	NGA
GGGACCCTGCGCTGAACATG	新型	NAG
GGGACTGCGAATTTTGGCCA	新型	NGA
GGGAGAGGCAGTAGTCGGAA	新型	NAG
GGGAGGAGTTGGGGGAGGAGA	新型	NNNRRT

[1777]

向导序列	向导 id	PAM
GGGATACATAGAGGTTCCTT	新型	NAG
GGGATACATAGAGGTTCCTTG	新型	NNNRRT
GGGATCTTGCAGAGTTTGGT	新型	NAA
GGGATCTTGCAGAGTTTGGTG	新型	NNNRRT
GGGCAACATTCGGTGGGCGTT	新型	NNNRRT
GGGCAACGGGGTAAAGGTTC	新型	NGG
GGGCAGAATCTTTCCACCAG	新型	NAA
GGGCATACTACAAACTTTGC	新型	NAG
GGGCCAAGTCTGTACAGCATC	新型	NNGRRT
GGGCCAAGTCTGTACAGCATC	新型	NNNRRT
GGGCCATCAGCGCGTGCGTG	新型	NAA
GGGCCTCAGCCCGTTTCTCC	新型	NGG
GGGCGCACCTCTCTTTACGC	新型	NGA
GGGCTGAGGCCCACTCCCAT	新型	NGG
GGGCTTGGTCATGGGCCATC	新型	NGC
GGGCTTTCGGAAAATTCCTA	新型	NGG
GGGCTTTCGGAAAATTCCTAT	新型	NNGRRT
GGGCTTTCGGAAAATTCCTAT	新型	NNNRRT
GGGGAACTACCGTGTGTCTT	新型	NGC
GGGGACCCTGCGCTGAACAT	新型	NGA
GGGGACGAGAGAGTCCCAAG	新型	NGA
GGGGACTGCGAATTTTGGCC	新型	NAG
GGGGCATTTGGTGGTCTATA	新型	NGC
GGGGCGCACCTCTCTTTACG	新型	NGG
GGGGCTTGGTCATGGGCCAT	新型	NAG
GGGGGAACTACCGTGTGTCT	新型	NGG
GGGGGAGGAGATTAGATTAA	新型	NGG
GGGGTGAAACCCAGCCCGAA	新型	NGC
GGGGTGGAGCCCTCAGGCTC	新型	NGG

[1778]

向导序列	向导 id	PAM
GGGGTTTTTCTTGTTGACAA	新型	NAA
GGGTAGGCTGCCTTCCTGTC	新型	NGG
GGGTAGGCTGCCTTCCTGTCT	新型	NNNRRT
GGGTATATTATATAAGAGAG	新型	NAA
GGGTATTAAACCTTATTATC	新型	NAG
GGGTCACCATATTCTTGGGAA	新型	NNNRRT
GGGTCCTAGGAATCCTGATG	新型	NGA
GGGTCGATGTCCATGCCCCA	新型	NAG
GGGTCGTCCGCAGGATTCAG	新型	NGC
GGGTGGAGCCCTCAGGCTCA	新型	NGG
GGGTGTTAATTTGGAAGATC	新型	NAG
GGGTTACTCTCTGAATTTTA	新型	NGG
GGGTTGCGTCAGCAAACACT	新型	NGG
GGGTTTAAATGTATACCCAA	新型	NGA
GGGTTTACTGCTCCTGAACT	新型	NGA
GGGTTTCACCCCACCGCACG	新型	NAG
GGTAAATATTTGGTAACCTT	新型	NGG
GGTAACCTTTGGATAAAACC	新型	NAG
GGTAGGAGCTGGAGCATTCG	新型	NGC
GGTAGGCTGCCTTCCTGTCT	新型	NGC
GGTAGTTCCCCCTAGAAAAT	新型	NGA
GGTATACATTTAAACCCTAA	新型	NAA
GGTATTAAACCTTATTATCC	新型	NGA
GGTATTGTGAGGATTCTTGT	新型	NAA
GGTCACCATATTCTTGGGAA	新型	NAA
GGTCATGGGCCATCAGCGCG	新型	NGC
GGTCCCGTGCTGGTGGTTGA	新型	NGA
GGTCGATGTCCATGCCCCAA	新型	NGC
GGTCGGAACGGCAGACGGAG	新型	NAG

[1779]

向导序列	向导 id	PAM
GGTCGGGAAAGAATCCCAGA	新型	NGA
GGTCGGTCGGAACGGCAGAC	新型	NGA
GGTCGGTCGTTGACATTGCA	新型	NAG
GGTCTCAATCGCCGCGTCGC	M76	NGA
GGTCTCAATCGCCGCGTCGCA	M13	NNNRRT
GGTCTCCATGCGACGTGCAG	M63	NGG
GGTCTCTAGATGCTGGATCTT	新型	NNNRRT
GGTCTGGAGCAAACATTATC	新型	NGG
GGTCTGTGCCAAGTGTTTGC	新型	NGA
GGTGAGGTGAACAATGCTCA	新型	NGA
GGTGAGTGATTGGAGGTTGG	新型	NGA
GGTGCAATTTCCGTCCGAAGG	新型	NNNRRT
GGTGCGCCCCGTGGTCGGTC	新型	NGA
GGTGGAAAGATTCTGCCCCA	新型	NGC
GGTGGAGCCCTCAGGCTCAG	新型	NGC
GGTGGGGTGAAACCCAGCCC	新型	NAA
GGTGGTCGGGAAAGAATCCC	新型	NGA
GGTGGTCGGGAAAGAATCCC	新型	NNAGGA
GGTGGTCGGGAAAGAATCCCA	新型	NNGRRT
GGTGGTCGGGAAAGAATCCCA	新型	NNNRRT
GGTGGTCTCCATGCGACGTG	新型	NAG
GGTGGTCTCCATGCGACGTGC	M33	NNNRRT
GGTGTAAATGTATATTAGGA	新型	NAA
GGTGTTAATTTGGAAGATCC	新型	NGC
GGTGTTTTCCAATGAGGATT	新型	NAA
GGTTACCAAATATTTACCAT	新型	NGG
GGTTACTCTCTGAATTTTAT	新型	NGG
GGTTATGTCATTGGAAGTTA	新型	NGG
GGTTCAGGTATTGTTTACAC	新型	NGA

[1780]

向导序列	向导 id	PAM
GGTTCCACGCACGCGCTGAT	新型	NGC
GGTTCCTTGAGCAGTAGTCA	新型	NGC
GGTTCCTTGAGCAGTAGTCAT	新型	NNNRRT
GGTTCGGCATGGTCCCGTGC	新型	NGG
GGTTCGGCATGGTCCCGTGCT	新型	NNNRRT
GGTTGAGGATCCTGGAATTA	新型	NAG
GGTTGCGTCAGCAAACACTT	新型	NGC
GGTTGGGGACTGCGAATTTT	新型	NGC
GGTTTAATACCCTTATCCAA	新型	NGG
GGTTTACTGCTCCTGAACTG	新型	NAG
GGTTTCACCCCACCGCACGG	新型	NGG
GGTTTGGTACAGCAACAGGA	新型	NGG
GGTTTTATCCAAAGGTTACC	新型	NAA
GTAAACAAAGGACGTCCCGC	新型	NNAGGA
GTAAACAAAGGACGTCCCGCG	新型	NNGRRT
GTAAACAAAGGACGTCCCGCG	新型	NNNRRT
GTAAACCCTGTTCCGACTAC	新型	NGC
GTAAACTGAGCCAGGAGAAA	新型	NGG
GTAAAGAGAGGTGCGCCCCG	新型	NGG
GTAAATAGTGTCTAGTTTGG	新型	NAG
GTAAATAGTGTCTAGTTTGGA	新型	NNNRRT
GTAAATATTTGGTAACCTTT	新型	NGA
GTAAATGGGGCAGCAAAACC	新型	NAA
GTAACACGAGAAGGGGTCCT	新型	NGG
GTAACCTTTGGATAAAACCT	新型	NGC
GTAAGACCTTGGGCAACATT	新型	NGG
GTAAGTTGGCGAGAAAGTGA	新型	NAG
GTACGAGATCTTCTAGATAC	新型	NGC
GTACTAGGAGGCTGTAGGCA	新型	NAA

[1781]

向导序列	向导 id	PAM
GTACTGAAGGAAAGAAGTCA	新型	NAA
GTAGAAGAATAAAGACCAGT	新型	NAA
GTAGCTCCAAATTCTTTATA	新型	NGG
GTAGGAGCTGGAGCATTCGGG	新型	NNGRRT
GTAGGAGCTGGAGCATTCGGG	新型	NNNRRT
GTAGGCATAAATTGGTCTGC	新型	NNACCA
GTAGGCCCACTTACAGTTAA	新型	NGA
GTAGTATGCCCTGAGCCTGA	新型	NGG
GTAGTCAGTTATGTCAACAC	新型	NAA
GTAGTCATGCAGGTTCGGCA	新型	NGG
GTAGTCGGAACAGGGTTTAC	新型	NGC
GTAGTTCCCCCTAGAAAATT	新型	NAG
GTAGTTTCCGGAAGTGTTGA	新型	NAG
GTATACATTTAAACCCTAAC	新型	NAA
GTATACCCAAAGACAAAAGA	新型	NAA
GTATCTAGAAGATCTCGTAC	新型	NGA
GTATGATGTGTTCTTGTGGC	新型	NAG
GTATGCATGTATTCAATCTA	新型	NGC
GTATGGATCGGCAGAGGAGC	新型	NGA
GTATTAAACCTTATTATCCA	新型	NAA
GTATTCCCATCCCATCATCC	新型	NGG
GTATTTGGTGTCTTTCGGAGT	新型	NNGRRT
GTATTTGGTGTCTTTCGGAGT	新型	NNNRRT
GTCAACACTAATATGGGCCT	新型	NAA
GTCAATCTTCTCGAGGATTG	新型	NGG
GTCACCATATTCTTGGGAAC	新型	NAG
GTCATTAGTTCCCCCCAGCA	新型	NAG
GTCCAAGAGTCCTCTTATGT	新型	NAG
GTCCAAGGAATACTAACATT	新型	NAG

[1782]

向导序列	向导 id	PAM
GTCCACCACGAGTCTAGACTC	E16/M2	NNNRRT
GTCCAGAAGAACCAACAAGA	新型	NGA
GTCCATGCCCCAAAGCCACC	新型	NAA
GTCCCAAGCGACCCCGAGAA	新型	NGG
GTCCCGATAATGTTTGCTCC	新型	NGA
GTCCCGCGCAGGATCCAGTT	新型	NGC
GTCCCGTCGGCGCTGAATCC	新型	NGC
GTCCGGCAGATGAGAAGGCA	新型	NAG
GTCCTACTGTTCAAGCCTCC	新型	NAG
GTCCTCTTATGTAAGACCTT	新型	NGG
GTCCTTTGTTTACGTCCCGT	新型	NGG
GTCGCAGAAGATCTCAATCT	新型	NGG
GTCGGAACGGCAGACGGAGA	新型	NGG
GTCGGTCGGAACGGCAGACG	新型	NAG
GTCGGTCGTTGACATTGCAG	新型	NGA
GTCTAACAACAGTAGTTTCC	新型	NGA
GTCTAGACTCTGCGGTATTG	新型	NGA
GTCTAGACTCTGCGGTATTG	新型	NNAGGA
GTCTAGACTCTGCGGTATTGT	新型	NNGRRT
GTCTAGACTCTGCGGTATTGT	新型	NNNRRT
GTCTATAAGCTGGAGGAGTG	新型	NGA
GTCTCAATCGCCGCGTCGCA	新型	NAA
GTCTGCGCACCAGCACCATG	新型	NAA
GTCTGGAGCAAACATTATCG	新型	NGA
GTCTGGCCAGGTGTCCTTGT	新型	NGG
GTCTGGCGATTGGTGGAGGC	新型	NGG
GTCTGTGCCTTCTCATCTGC	新型	NGG
GTCTTACATAAGAGGACTCT	新型	NGG
GTCTTGGTGTAAATGTATAT	新型	NAG

[1783]

向导序列	向导 id	PAM
GTCTTTAAACAAACAGTCTT	新型	NGA
GTCTTTGAAGTATGCCTCAAG	新型	NNNRRT
GTCTTTGTACTAGGAGGCTG	新型	NAG
GTGAAAAAGTTGCATGGTGC	新型	NGG
GTGAAACCACAAGAGTTGCC	新型	NGA
GTGAGAGGCCTGTATTTCCC	新型	NGC
GTGAGAGGCCTGTATTTCCCT	新型	NNNRRT
GTGAGGATTCTTGTCAACAA	新型	NAA
GTGAGGTGAACAATGCTCAG	新型	NAG
GTGATGTTCTCCATGTTCAG	新型	NGC
GTGCACACGGTCCGGCAGAT	新型	NAG
GTGCAGTATGGTGAGGTGAA	新型	NAA
GTGCCAAGTGTTTGCTGACG	新型	NAA
GTGCGAATCCACACTCCGAA	新型	NGA
GTGCGCCCCGTGGTCGGTCG	新型	NAA
GTGCTGCCAACTGGATCCTG	新型	NGC
GTGCTGGTGGTTGAGGATCC	新型	NGG
GTGGACTTCTCTCAATTTTC	新型	NAG
GTGGAGACAGCGGGGTAGGC	新型	NGC
GTGGAGGCAGGAGGCGGATT	新型	NGC
GTGGGTCACCATATTCTTGG	新型	NAA
GTGGGTCTTTTGGGTTTTGC	新型	NGC
GTGGTCGGGAAAGAATCCCA	新型	NAG
GTGGTCTCCATGCGACGTGC	M93	NGA
GTGGTTGAGGATCCTGGAAT	新型	NAG
GTGTAAATAGTGTCTAGTTT	新型	NGA
GTGTAAATGTATATTAGGAA	新型	NAG
GTGTGGATTCGCACTCCTCC	新型	NGC
GTGTTGACATAACTGACTAC	新型	NAG

[1784]

向导序列	向导 id	PAM
GTGTTTCTCAAAGGTGGAGA	新型	NAG
GTGTTTTCCAATGAGGATTA	新型	NAG
GTTAATCATTACTTCCAAAC	新型	NAG
GTTACCAAATATTTACCATT	新型	NGA
GTTACCAATTTTCTTTTGTCT	新型	NNGRRT
GTTACCAATTTTCTTTTGTCT	新型	NNNRRT
GTTAGTATTCCTTGGACTCA	新型	NAA
GTTATCGCTGGATGTGTCTG	新型	NGG
GTTATGGGTCCTTGCCACAA	新型	NAA
GTTATGTCAACACTAATATG	新型	NGC
GTTATGTCATTGGAAGTTAT	新型	NGG
GTTCACATTTTTTGATAATGT	新型	NNNRRT
GTTCAGGTATTGTTTACACA	新型	NAA
GTTCCCCACCTTATGAGTCC	新型	NAG
GTTCCCCACCTTATGAGTCCA	新型	NNGRRT
GTTCCCCACCTTATGAGTCCA	新型	NNNRRT
GTTCCCCCTAGAAAATTGAG	新型	NGA
GTTCCGCAGTATGGATCGGC	E17	NGA
GTTCCGCAGTATGGATCGGC	新型	NNAGGA
GTTCTTGTGGCAAGGACCCA	新型	NAA
GTTGACATTGCAGAGAGTCC	新型	NAG
GTTGAGGATCCTGGAATTAG	新型	NGG
GTTGATAGGATAGGGGCATT	新型	NGG
GTTGATAGGATAGGGGCATTT	新型	NNNRRT
GTTGCATGGTGCTGGTGCGC	新型	NGA
GTTGCATGGTGCTGGTGCGC	新型	NNACCA
GTTGCCCAAGGTCTTACATA	新型	NGA
GTTGCCCAAGGTCTTACATA	新型	NNAGGA
GTTGCCGGGCAACGGGGTAA	新型	NGG

[1785]

向导序列	向导 id	PAM
GTTGGAGGACAGGAGGTTGG	新型	NGA
GTTGGATCCAGCCTTCAGAG	新型	NAA
GTTGGGGGAGGAGATTAGAT	新型	NAA
GTTGGGGGAGGAGATTAGATT	新型	NNNRRT
GTTGGTGAGTGATTGGAGGT	新型	NGG
GTTGGTTCTTCTGGACTATC	新型	NAG
GTTTAAAGACTGGGAGGAGT	新型	NGG
GTTTAATACCCTTATCCAATG	新型	NNNRRT
GTTTACACAGAAAGGCCTTG	新型	NAA
GTTTACGTCCCGTCGGCGCT	新型	NAA
GTTTACTAGTGCCATTTGTT	新型	NAG
GTTTACTAGTGCCATTTGTTC	新型	NNNRRT
GTTTACTGCTCCTGAACTGG	新型	NGC
GTTTCACCCCACCGCACGGA	新型	NGC
GTTTCTCAAAGGTGGAGACAG	新型	NNGRRT
GTTTCTCAAAGGTGGAGACAG	新型	NNNRRT
GTTTGCTCCAGACCTGCTGC	新型	NAG
GTTTGGAAGTAATGATTAAC	新型	NAG
GTTTGGTACAGCAACAGGAG	新型	NGA
GTTTGTTTAAAGACTGGGAG	新型	NAG
GTTTTCCAATGAGGATTAAAG	M15	NNNRRT
GTTTTGCTCGCAGCAGGTCT	新型	NGA
GTTTTGCTGCCCCATTTACA	新型	NAA
TAAAACCTAGCAGGCATAAT	新型	NAA
TAAAACGCCGCAGACACATC	新型	NAG
TAAAACGCCGCAGACACATCC	E18	NNNRRT
TAAACAAAGGACGTCCCGCG	新型	NAG
TAAACCCTAACAAAACAAAG	新型	NGA
TAAACCTTATTATCCAGAACA	新型	NNNRRT

[1786]

向导序列	向导 id	PAM
TAAACTGAGCCAGGAGAAAC	新型	NGG
TAAAGAATTTGGAGCTACTG	新型	NGG
TAAAGACAGGTACAGTAGAA	新型	NAA
TAAAGACTGGGAGGAGTTGG	新型	NGG
TAAAGAGAGGTGCGCCCCGTG	新型	NNNRRT
TAAAGGTCTTTGTACTAGGA	新型	NGC
TAAAGTTCAGGCAACTCTTG	新型	NGG
TAAATGGGGCAGCAAAACCC	新型	NAA
TAAATGTATACCCAAAGACA	新型	NAA
TAACACCCACCCAGGTAGCT	新型	NGA
TAACACGAGAAGGGGTCCTA	新型	NGA
TAACAGCGGTAAAAAGGGACT	新型	NNNRRT
TAACAGGCCTATTGATTGGA	新型	NAG
TAACATTGAGGTTCCCGAGAT	新型	NNNRRT
TAACCACATTGTGTAAATGG	新型	NGC
TAACCAGGACAAGTTGGAGG	新型	NNAGGA
TAACTAGATGTTCTGGATAA	新型	NAA
TAACTGAAAGCCAAACAGTG	新型	NGG
TAACTGACTACTAGGTCTCT	新型	NGA
TAAGAGGACTCTTGGACTCTC	新型	NNNRRT
TAAGCAGGCTTTCACTTTCT	新型	NGC
TAAGGGAGAGGCAGTAGTCG	新型	NAA
TAAGGTTTAATACCCTTATC	新型	NAA
TAAGGTTTAATACCCTTATCC	新型	NNNRRT
TAAGTTGGCGAGAAAGTGAA	新型	NGC
TAATAAGGTTTAATACCCTTA	新型	NNNRRT
TAATACCCTTATCCAATGGT	新型	NAA
TAATATACCCGCCTTCCATA	新型	NAG
TAATATGGGCCTAAAGTTCA	新型	NGC

[1787]

向导序列	向导 id	PAM
TAATCTCCTCCCCCAACTCCT	新型	NNNRRT
TAATGAGAAAAGAAGATTGCA	新型	NNNRRT
TAATGATTAACTAGATGTTC	新型	NGG
TAATGCCCTTGTATGCATGTA	新型	NNNRRT
TACAAACTGTTCACATTTTT	新型	NGA
TACAAACTGTTCACATTTTTT	新型	NNNRRT
TACACAATGTGGTTATCCTGC	M31	NNNRRT
TACACCAAGACATTATCAAA	新型	NAA
TACAGAGCTGAGGCGGTATC	新型	NAG
TACAGGTGCAATTTCCGTCC	新型	NAA
TACAGTAGAAGAATAAAGAC	新型	NAG
TACATCGTATCCATGGCTGC	新型	NAG
TACCAATTTTCTTTTGTCTT	新型	NGG
TACCACATCATCCATATAAC	M71	NGA
TACCATTGGATAAGGGTATT	新型	NAA
TACCCCGCTGTCTCCACCTT	新型	NGA
TACCCCGTTGCCCGGCAACGG	新型	NNNRRT
TACCCGCCTTCCATAGAGTGT	新型	NNNRRT
TACCGCAGAGTCTAGACTCG	M34	NGG
TACCGCCTCAGCTCTGTATC	新型	NGG
TACCTGAACCTTTACCCCGT	新型	NGC
TACCTGGGTGGGTGTTAATT	新型	NGG
TACCTGTCTTTAATCCTCAT	新型	NGG
TACCTTGTTGGCGTCTGGCC	新型	NGG
TACGATGTATATTTGCGGGA	新型	NAG
TACTAACATTGAGGTTCCCG	E24_splice	NGA
TACTAGGAGGCTGTAGGCAT	新型	NAA
TACTAGTGCCATTTGTTCAG	新型	NGG
TACTGAAGGAAAGAAGTCAG	新型	NAG

[1788]

向导序列	向导 id	PAM
TACTGCCTCTCCCTTATCGT	新型	NAA
TACTTCAAAGACTGTTTGTT	新型	NAA
TACTTTCCAATCAATAGGCCT	M29	NNNRRT
TAGAAAACTTCCTATTAACA	新型	NGC
TAGAAGAATAAAGACCAGTA	新型	NAG
TAGAAGATCTCGTACTGAAG	新型	NAA
TAGACTCTGCGGTATTGTGA	新型	NGA
TAGAGTATTTGGTGTCTTTC	新型	NGA
TAGAGTCATTAGTTCCCCCC	新型	NGC
TAGAGTGTGTAAATAGTGTC	新型	NAG
TAGATGTTCTGGATAATAAGG	新型	NNNRRT
TAGATTAAAGGTCTTTGTAC	新型	NAG
TAGATTGAATACATGCATAC	新型	NAG
TAGCAGCAGGATGAAGAGGA	新型	NGA
TAGCAGCAGGATGAAGAGGAA	新型	NNNRRT
TAGCCGCTTGTTTTGCTCGC	新型	NGC
TAGCCGCTTGTTTTGCTCGCA	新型	NNNRRT
TAGCTCCAAATTCTTTATAA	新型	NGG
TAGGAAAAGATGGTGTTTTC	新型	NAA
TAGGAATTTTCCGAAAGCCC	新型	NGG
TAGGACCCCTTCTCGTGTTA	新型	NAG
TAGGCAGAGGTGAAAAAGTTG	新型	NNNRRT
TAGGCCCACTTACAGTTAAT	新型	NAG
TAGGCTGCCTTCCTGTCTGG	新型	NGA
TAGGGCTTTCCCCCACTGTT	新型	NGG
TAGGGGCATTTGGTGGTCTA	新型	NAA
TAGGGTTTAAATGTATACCC	新型	NAA
TAGTACAAAGACCTTTAATC	新型	NAA
TAGTATGCCCTGAGCCTGAG	新型	NGC

[1789]

向导序列	向导 id	PAM
TAGTATTCCTTGGACTCATA	新型	NGG
TAGTCCAGAAGAACCAACAA	新型	NAA
TAGTGTTGACATAACTGACTA	新型	NNNRRT
TAGTTCCCCCTAGAAAATTG	新型	NGA
TAGTTTCCGGAAGTGTTGAT	新型	NGG
TAGTTTGGAAGTAATGATTAA	新型	NNNRRT
TATAACGGTTTCTCTTCCAA	新型	NAG
TATAACTGAAAGCCAAACAG	新型	NGG
TATAAGAGAGAAACAACACA	新型	NAG
TATAATATACCCGCCTTCCA	新型	NAG
TATACATTTAAACCCTAACA	新型	NAA
TATACCCAAAGACAAAAGAAA	M7	NNNRRT
TATAGAGTATTTGGTGTCTTT	新型	NNGRRT
TATAGAGTATTTGGTGTCTTT	新型	NNNRRT
TATATGGATGATGTGGTATT	新型	NGG
TATATTATATAAGAGAGAAA	新型	NAA
TATATTTGCGGGAGAGGACAA	新型	NNGRRT
TATATTTGCGGGAGAGGACAA	新型	NNNRRT
TATCATCTTCCTCTTCATCC	新型	NGC
TATCCATGGCTGCTAGGCTG	新型	NGC
TATCCCTCCTGTTGCTGTAC	新型	NAA
TATCCTATCAACACTTCCGG	新型	NAA
TATCTAGAAGATCTCGTACT	新型	NAA
TATCTAGAAGATCTCGTACT	新型	NNAGGA
TATGATGTGTTCTTGTGGCA	新型	NGG
TATGCCTCAAGGTCGGTCGT	新型	NGA
TATGCCTGCTAGGTTTTATC	新型	NAA
TATGCCTGCTAGGTTTTATCC	新型	NNNRRT
TATGGATCGGCAGAGGAGCC	新型	NAA

[1790]

向导序列	向导 id	PAM
TATGGATGATGTGGTATTGG	M94	NGG
TATGGTGACCCACAAAATGA	新型	NGC
TATGGTGAGGTGAACAATGC	新型	NNAGGA
TATGTAAGACCTTGGGCAACA	新型	NNNRRT
TATGTATCCCTCCTGTTGCT	新型	NNACCA
TATGTTGCCCGTTTGTCCTC	新型	NAA
TATTAACAGGCCTATTGATT	新型	NGA
TATTAACAGGCCTATTGATTG	新型	NNNRRT
TATTAGGAAAAGATGGTGTTT	新型	NNNRRT
TATTATCCAGAACATCTAGT	新型	NAA
TATTCCCATCCCATCATCCT	新型	NGG
TATTCCTTGGACTCATAAGG	新型	NGG
TATTCTTCTACTGTACCTGTC	新型	NNNRRT
TATTGATTGGAAAGTATGTCA	新型	NNGRRT
TATTGATTGGAAAGTATGTCA	新型	NNNRRT
TATTGGGGGCCAAGTCTGTA	新型	NAG
TATTTACACACTCTATGGAA	新型	NGC
TATTTACACACTCTATGGAAG	新型	NNGRRT
TATTTACACACTCTATGGAAG	新型	NNNRRT
TATTTCCCTGCTGGTGGCTC	新型	NAG
TATTTGCGGGAGAGGACAAC	新型	NGA
TATTTGGTAACCTTTGGATA	新型	NAA
TCAACACTAATATGGGCCTA	新型	NAG
TCAACGAATTGTGGGTCTTT	M61	NGG
TCAAGATGCTGTACAGACTT	新型	NGC
TCAAGGTCGGTCGTTGACAT	新型	NGC
TCAATAGGCCTGTTAATAGG	新型	NAG
TCAATCCCAACAAGGACACC	M52	NGG
TCAATCTTCTCGAGGATTGG	新型	NGA

[1791]

向导序列	向导 id	PAM
TCACCAAACTCTGCAAGATCC	新型	NNGRRT
TCACCAAACTCTGCAAGATCC	新型	NNNRRT
TCACCATACTGCACTCAGGC	新型	NAG
TCACCATACTGCACTCAGGCA	新型	NNNRRT
TCACCATATTCTTGGGAACA	新型	NGA
TCACCTCACCATACTGCACT	新型	NAG
TCACCTCTGCACGTCGCATG	新型	NAG
TCACTCTGGGATCTTGCAGAG	新型	NNNRRT
TCACTTTCTCGCCAACTTAC	新型	NAG
TCAGAAGGCAAAAACGAGAG	新型	NAA
TCAGCCCGTTTCTCCTGGCT	新型	NAG
TCAGGCAAGCAATTCTTTGC	M41	NGG
TCAGGCTCAGGGCATACTAC	新型	NAA
TCAGGGCATACTACAAACTT	新型	NGC
TCAGGTATTGTTTACACAGA	新型	NAG
TCAGTTTACTAGTGCCATTTG	新型	NNNRRT
TCATCCATATAACTGAAAGC	新型	NAA
TCATTAGTTCCCCCCAGCAA	新型	NGA
TCCAAATTCTTTATAAGGGT	新型	NGA
TCCAACTGGTGGTCGGGAAA	新型	NAA
TCCAACTTGTCCTGGTTATCG	新型	NNGRRT
TCCAACTTGTCCTGGTTATCG	新型	NNNRRT
TCCAAGAGTCCTCTTATGTA	新型	NGA
TCCAAGGAATACTAACATTG	M46	NGG
TCCAATCAATAGGCCTGTTA	新型	NNAGGA
TCCACACTCCGAAAGACACC	新型	NAA
TCCACCAATCGCCAGACAGG	新型	NAG
TCCACCACGAGTCTAGACTC	新型	NGC
TCCACCCCAAAAGGCCTCCG	新型	NGC

[1792]

向导序列	向导 id	PAM
TCCAGCATCTAGAGACCTAG	新型	NAG
TCCAGCCTTCAGAGCAAACA	新型	NAG
TCCAGTTGGCAGCACAGCCT	新型	NGC
TCCATGCCCCAAAGCCACCC	新型	NAG
TCCATGCGACGTGCAGAGGT	新型	NAA
TCCCAACAAGGACACCTGGC	新型	NAG
TCCCAAGAATATGGTGACCCA	M20	NNNRRT
TCCCAGAGGATTGCTGGTGG	新型	NAA
TCCCATAGGAATTTTCCGAA	新型	NGC
TCCCATCATCCTGGGCTTTC	新型	NGA
TCCCATCATCCTGGGCTTTCG	新型	NNNRRT
TCCCCAATCCTCGAGAAGAT	M85	NGA
TCCCCAATCCTCGAGAAGATT	M24	NNNRRT
TCCCCACCTTATGAGTCCAA	新型	NGA
TCCCCCTAGAAAATTGAGAG	新型	NAG
TCCCGACCACCAGTTGGATC	新型	NAG
TCCCTGCTGGTGGCTCCAGT	新型	NNAGGA
TCCCTTATCGTCAATCTTCT	新型	NGA
TCCCTTATCGTCAATCTTCT	新型	NNAGGA
TCCCTTATCGTCAATCTTCTC	新型	NNGRRT
TCCCTTATCGTCAATCTTCTC	新型	NNNRRT
TCCCTTTTTACCGCTGTTAC	新型	NAA
TCCGAAAGCCCAGGATGATG	新型	NGA
TCCGAAGGTTTGGTACAGCA	新型	NNAGGA
TCCGCAGGATTCAGCGCCGA	新型	NGG
TCCGCAGTATGGATCGGCAG	E19	NGG
TCCGGAAGTGTTGATAGGAT	新型	NGG
TCCGGCAGATGAGAAGGCAC	新型	NGA
TCCGTCCGAAGGTTTGGTAC	新型	NGC

[1793]

向导序列	向导 id	PAM
TCCTAATATACATTTACACC	新型	NAG
TCCTACCTTGTTGGCGTCTGG	新型	NNNRRT
TCCTACTGTTCAAGCCTCCA	新型	NGC
TCCTAGCCGCTTGTTTTGCT	新型	NGC
TCCTAGTACAAAGACCTTTAA	新型	NNNRRT
TCCTCCAGCTTATAGACCAC	新型	NAA
TCCTCGAGAAGATTGACGAT	新型	NAG
TCCTCTAATTCCAGGATCCT	新型	NAA
TCCTCTAATTCCAGGATCCT	新型	NNACCA
TCCTCTGCCGATCCATACTG	E20	NGG
TCCTCTTATGTAAGACCTTG	新型	NGC
TCCTCTTCATCCTGCTGCTA	新型	NGC
TCCTGAACTGGAGCCACCAG	新型	NAG
TCCTGCCTCCACCAATCGCC	新型	NGA
TCCTGCGGACGACCCTTCTC	新型	NGG
TCCTGGAATTAGAGGACAAA	新型	NGG
TCCTGTCCTCCAACTTGTCC	新型	NGG
TCCTGTCTGGCGATTGGTGG	新型	NGG
TCCTTCAGTACGAGATCTTC	新型	NAG
TCCTTGAGCAGTAGTCATGC	新型	NGG
TCCTTGGACTCATAAGGTGG	新型	NGA
TCCTTTGTTTACGTCCCGTC	新型	NGC
TCGACCCTTATAAAGAATTT	新型	NGA
TCGAGAAGATTGACGATAAG	新型	NGA
TCGCAGAAGATCTCAATCTC	新型	NGG
TCGCAGACGAAGGTCTCAAT	新型	NGC
TCGGAAAATTCCTATGGGAG	新型	NGG
TCGGAACGGCAGACGGAGAA	新型	NGG
TCGGCATGGTCCCGTGCTGG	新型	NGG

[1794]

向导序列	向导 id	PAM
TCGGCGCTGAATCCTGCGGA	新型	NGA
TCGGTCGGAACGGCAGACGG	新型	NGA
TCGGTCGTTGACATTGCAGA	新型	NAG
TCGTACTGAAGGAAAGAAGT	新型	NAG
TCGTCAATCTTCTCGAGGAT	新型	NGG
TCGTCCGCAGGATTCAGCGC	新型	NGA
TCGTTGACATACTTTCCAAT	新型	NAA
TCTAACAACAGTAGTTTCCG	新型	NAA
TCTAAGGCTTCCCGATACAG	新型	NGC
TCTAATTCCAGGATCCTCAA	新型	NNACCA
TCTAGAAGATCTCGTACTGA	新型	NGG
TCTAGACTCTGCGGTATTGT	新型	NAG
TCTAGTTAATCATTACTTCC	新型	NAA
TCTAGTTTGGAAGTAATGAT	新型	NAA
TCTATAACGGTTTCTCTTCC	新型	NAA
TCTATAAGCTGGAGGAGTGC	新型	NAA
TCTCAAAGGTGGAGACAGCG	新型	NGG
TCTCAATCGCCGCGTCGCAG	新型	NAG
TCTCACTCTGGGATCTTGCA	新型	NAG
TCTCCGTCTGCCGTTCCGAC	新型	NGA
TCTCCGTCTGCCGTTCCGAC	新型	NNACCA
TCTCCTCCCCCAACTCCTCC	新型	NAG
TCTCCTGAGCATTGTTCACC	新型	NNACCA
TCTCTAGATGCTGGATCTTC	新型	NAA
TCTCTCTTATATAATATACC	新型	NGC
TCTCTTCCAAAAGTGAGACA	新型	NGA
TCTGACTTCTTTCCTTCAGTA	新型	NNNRRT
TCTGCAAGATCCCAGAGTGA	新型	NAG
TCTGCCGTTCCGACCGACCA	新型	NGG

[1795]

向导序列	向导 id	PAM
TCTGGACTATCAAGGTATGT	新型	NGC
TCTGGAGCAAACATTATCGGG	新型	NNNRRT
TCTGGCCAGGTGTCCTTGTT	新型	NGG
TCTGGCGATTGGTGGAGGCA	新型	NGA
TCTGTGCCTTCTCATCTGCC	新型	NGA
TCTTACATAAGAGGACTCTT	新型	NGA
TCTTCCAAAAGTGAGACAAG	新型	NAA
TCTTCTACTGTACCTGTCTT	新型	NAA
TCTTCTGCGACGCGGCGATT	新型	NAG
TCTTCTTTTCTCATTAACTG	新型	NAA
TCTTGGACTCTCTGCAATGT	新型	NAA
TCTTGGTGTAAATGTATATT	新型	NGG
TCTTGTCTCACTTTTGGAAG	新型	NGA
TCTTGTTCCCAAGAATATGG	新型	NGA
TCTTTAAACAAACAGTCTTT	新型	NAA
TCTTTAATCCTCATTGGAAA	新型	NNACCA
TCTTTCCACCAGCAATCCTC	新型	NGG
TCTTTGCTGGGGGGAACTAA	新型	NGA
TCTTTGTACTAGGAGGCTGT	新型	NGG
TCTTTGTTTTGTTAGGGTTT	新型	NAA
TCTTTTCCTAATATACATTT	新型	NNACCA
TGAAACCACAAGAGTTGCCT	新型	NAA
TGAAAGCCAAACAGTGGGGG	新型	NAA
TGAACAAGAGATGATTAGGC	新型	NGA
TGAACAGTAGGACATGAACA	新型	NGA
TGAACAGTTTGTAGGCCCACT	M17	NNNRRT
TGAACATGGAGAACATCACA	新型	NNAGGA
TGAACATGGAGAACATCACAT	新型	NNGRRT
TGAACATGGAGAACATCACAT	新型	NNNRRT

[1796]

向导序列	向导 id	PAM
TGAACCTTTACCCCGTTGCC	新型	NGG
TGAACTGGAGCCACCAGCAG	新型	NGA
TGAAGAGGAAGATGATAAAA	新型	NGC
TGAAGGCTGGATCCAACTGG	新型	NGG
TGACATAACCCATAAAATTC	新型	NGA
TGACATAACCCATAAAATTCA	新型	NNGRRT
TGACATAACCCATAAAATTCA	新型	NNNRRT
TGACATACTTTCCAATCAAT	新型	NGG
TGACATTGCAGAGAGTCCAA	新型	NAG
TGACCAAGCCCCAGCCAGTG	新型	NGG
TGACGATAAGGGAGAGGCAG	新型	NAG
TGACGCAACCCCCACTGGCT	新型	NGG
TGACTACTAGGTCTCTAGATG	新型	NNGRRT
TGACTACTAGGTCTCTAGATG	新型	NNNRRT
TGACTTCTTTCCTTCAGTAC	新型	NAG
TGAGAAAAGAAGATTGCAAT	新型	NGA
TGAGACCTTCGTCTGCGAGG	新型	NGA
TGAGATCTTCTGCGACGCGG	新型	NGA
TGAGCCAGGAGAAACGGGCT	新型	NAG
TGAGCCTGAGGGCTCCACCC	新型	NAA
TGAGGATGAGTGTTTCTCAA	新型	NGG
TGAGGATTCTTGTCAACAAG	新型	NAA
TGAGGCATAGCAGCAGGATG	新型	NAG
TGAGGTGAACAATGCTCAGG	新型	NGA
TGAGTCCCTTTTTACCGCTG	新型	NNACCA
TGAGTGCAGTATGGTGAGGT	新型	NAA
TGAGTGCAGTATGGTGAGGTG	新型	NNNRRT
TGAGTGTTTCTCAAAGGTGG	新型	NGA
TGATAGTCCAGAAGAACCAA	新型	NAA

[1797]

向导序列	向导 id	PAM
TGATGGGATGGGAATACAGG	新型	NGC
TGATGTTCTCCATGTTCAGCG	新型	NNGRRT
TGATGTTCTCCATGTTCAGCG	新型	NNNRRT
TGATTGGAAAGTATGTCAAC	新型	NAA
TGATTGGAGGTTGGGGACTG	新型	NGA
TGCAAGATCCCAGAGTGAGA	新型	NGC
TGCAATTGATTATGCCTGCT	M48	NGG
TGCACACGGTCCGGCAGATG	新型	NGA
TGCATACAAGGGCATTAACG	新型	NAG
TGCATGTATTCAATCTAAGC	新型	NGG
TGCCAACTGGATCCTGCGCG	M189	NGA
TGCCACAAGAACACATCATA	新型	NAA
TGCCCAAGGTCTTACATAAG	新型	NGG
TGCCCGTTTGTCCTCTAATT	新型	NNAGGA
TGCCCGTTTGTCCTCTAATTC	新型	NNGRRT
TGCCCGTTTGTCCTCTAATTC	新型	NNNRRT
TGCCCTTGTATGCATGTATT	新型	NAA
TGCCGAACCTGCATGACTAC	新型	NGC
TGCCGTTCCGACCGACCACG	新型	NGG
TGCCTACAGCCTCCTAGTAC	新型	NAA
TGCCTCCACCAATCGCCAGA	新型	NAG
TGCCTGAGTGCAGTATGGTG	新型	NGG
TGCCTGCTAGGTTTTATCCA	新型	NAG
TGCGACGTGCAGAGGTGAAG	新型	NGA
TGCGAGCAAAACAAGCGGCT	新型	NGG
TGCGGTGGGGTGAAACCCAGC	新型	NNGRRT
TGCGGTGGGGTGAAACCCAGC	新型	NNNRRT
TGCTAGGCTGTGCTGCCAAC	新型	NGG
TGCTAGGTTTTATCCAAAGG	新型	NNACCA

[1798]

向导序列	向导 id	PAM
TGCTCCAGACCTGCTGCGAG	新型	NAA
TGCTCCAGCTCCTACCTTGT	M54	NGG
TGCTCCTGAACTGGAGCCAC	新型	NAG
TGCTCGCAGCAGGTCTGGAG	新型	NAA
TGCTGGCAAAGTTTGTAGTA	新型	NGC
TGCTGGTGGCTCCAGTTCAG	新型	NAG
TGCTGGTGGCTCCAGTTCAGG	新型	NNNRRT
TGCTGGTGGTTGAGGATCCT	新型	NGA
TGCTGTACAGACTTGGCCCC	新型	NAA
TGCTGTACCAAACCTTCGGA	新型	NGG
TGCTGTACCAAACCTTCGGAC	M26	NNNRRT
TGCTGTAGATCTTGTTCCCA	新型	NGA
TGGAAAACACCATCTTTTCC	新型	NAA
TGGAAAGTATGTCAACGAATT	新型	NNGRRT
TGGAAAGTATGTCAACGAATT	新型	NNNRRT
TGGAACCTTTTCGGCTCCTC	新型	NGC
TGGAACCTTTTCGGCTCCTCT	新型	NNNRRT
TGGAAGGCGGGTATATTATA	新型	NAA
TGGACATCGACCCTTATAAA	新型	NAA
TGGACTCTCTGCAATGTCAA	新型	NGA
TGGACTTCTCTCAATTTTCT	E21	NGG
TGGATAAAACCTAGCAGGCA	新型	NAA
TGGATACGATGTATATTTGC	新型	NGG
TGGATCCAACTGGTGGTCGG	新型	NAA
TGGATCGGCAGAGGAGCCGA	新型	NAA
TGGATGATGTGGTATTGGGGG	新型	NNNRRT
TGGATTTGCTGTGTTTGCTC	新型	NGA
TGGCAAGGACCCATAACTTC	新型	NAA
TGGCACTAGTAAACTGAGCC	新型	NGG

[1799]

向导序列	向导 id	PAM
TGGCAGCACAGCCTAGCAGCC	新型	NNGRRT
TGGCAGCACAGCCTAGCAGCC	新型	NNNRRT
TGGCCAAAATTCGCAGTCCC	新型	NAA
TGGCCAGACGCCAACAAGGT	新型	NGG
TGGCGATTGGTGGAGGCAGG	新型	NGG
TGGCTCCAGTTCAGGAGCAG	新型	NAA
TGGCTGCTAGGCTGTGCTGC	新型	NAA
TGGCTTTGGGGCATGGACAT	新型	NGA
TGGGAACAAGATCTACAGCA	M49	NGG
TGGGACTTCAATCCCAACAA	新型	NGA
TGGGATCTTGCAGAGTTTGG	新型	NGA
TGGGATTCTTTCCCGACCAC	新型	NAG
TGGGATTGAAGTCCCAATCT	新型	NGA
TGGGCCATCAGCGCGTGCGT	新型	NGA
TGGGGACCCTGCGCTGAACA	新型	NGG
TGGGGACTGCGAATTTTGGC	新型	NAA
TGGGGCAGAATCTTTCCACC	新型	NGC
TGGGGGAGGAGATTAGATTA	新型	NAG
TGGGGGGAACTAATGACTCT	新型	NGC
TGGGGTGGAGCCCTCAGGCT	新型	NAG
TGGGGTTACTCTCTGAATTTT	新型	NNGRRT
TGGGGTTACTCTCTGAATTTT	新型	NNNRRT
TGGGTGGGTGTTAATTTGGA	新型	NGA
TGGGTTATGTCATTGGAAGTT	新型	NNGRRT
TGGGTTATGTCATTGGAAGTT	新型	NNNRRT
TGGGTTTCACCCCACCGCAC	新型	NGA
TGGGTTTTGCTGCCCCATTTA	新型	NNNRRT
TGGTCCCGTGCTGGTGGTTG	新型	NGG
TGGTCGGGAAAGAATCCCAG	新型	NGG

[1800]

向导序列	向导 id	PAM
TGGTCGGTCGGAACGGCAGA	新型	NGG
TGGTCTATAAGCTGGAGGAG	新型	NGC
TGGTCTATAAGCTGGAGGAGT	新型	NNGRRT
TGGTCTATAAGCTGGAGGAGT	新型	NNNRRT
TGGTCTCCATGCGACGTGCA	新型	NAG
TGGTCTGCGCACCAGCACCA	新型	NGC
TGGTGACCCACAAAATGAGG	新型	NGC
TGGTGAGGTGAACAATGCTC	新型	NGG
TGGTGAGTGATTGGAGGTTG	新型	NGG
TGGTGGCTCCAGTTCAGGAG	新型	NAG
TGGTGGTCGGGAAAGAATCC	新型	NAG
TGGTGGTCTATAAGCTGGAG	新型	NAG
TGGTGTAAATGTATATTAGG	新型	NAA
TGGTGTAAATGTATATTAGGA	新型	NNNRRT
TGGTGTTTTCCAATGAGGAT	新型	NAA
TGGTTATCGCTGGATGTGTC	新型	NGC
TGGTTGAGGATCCTGGAATT	新型	NGA
TGGTTGAGGATCCTGGAATT	新型	NNAGGA
TGTAAATAGTGTCTAGTTTG	新型	NAA
TGTAAATGTATATTAGGAAA	新型	NGA
TGTAAATGTATATTAGGAAAA	新型	NNNRRT
TGTAACACGAGAAGGGGTCC	新型	NAG
TGTAACACGAGAAGGGGTCCT	新型	NNGRRT
TGTAACACGAGAAGGGGTCCT	新型	NNNRRT
TGTAAGTTGGCGAGAAAGTG	新型	NAA
TGTACCAAACCTTCGGACGG	新型	NAA
TGTAGATCTTGTTCCCAAGAA	新型	NNNRRT
TGTAGGCATAAATTGGTCTG	新型	NGC
TGTAGTATGCCCTGAGCCTG	新型	NGG

[1801]

向导序列	向导 id	PAM
TGTATACCCAAAGACAAAAG	新型	NAA
TGTATATTTGCGGGAGAGGA	新型	NAA
TGTATGATGTGTTCTTGTGG	新型	NAA
TGTATGATGTGTTCTTGTGG	新型	NNAGGA
TGTATGCATGTATTCAATCT	新型	NAG
TGTCAACACTAATATGGGCC	新型	NAA
TGTCAACGAATTGTGGGTCTT	M30	NNGRRT
TGTCAACGAATTGTGGGTCTT	新型	NNNRRT
TGTCCTACTGTTCAAGCCTC	新型	NAA
TGTCCTTGTTGGGATTGAAGT	新型	NNNRRT
TGTCTGGCGATTGGTGGAGG	新型	NAG
TGTCTTGGTGTAAATGTATA	新型	NNAGGA
TGTCTTTAATCCTCATTGGA	新型	NAA
TGTCTTTCGGAGTGTGGATT	新型	NGC
TGTGAGGATTCTTGTCAACA	新型	NGA
TGTGCACTTCGCTTCACCTC	新型	NGC
TGTGCCTTGGGTGGCTTTGG	新型	NGC
TGTGGAGTTACTCTCGTTTT	新型	NGC
TGTGGGTCACCATATTCTTG	新型	NGA
TGTGTAAATAGTGTCTAGTT	新型	NGG
TGTGTAAATAGTGTCTAGTTT	新型	NNNRRT
TGTGTCTTGGCCAAAATTCG	新型	NAG
TGTGTTGTTTCTCTCTTATA	新型	NAA
TGTTAATAGGAAGTTTTCTA	新型	NAA
TGTTAGTATTCCTTGGACTCA	新型	NNNRRT
TGTTCATGTCCTACTGTTCA	新型	NGC
TGTTCTCCATGTTCAGCGCA	新型	NGG
TGTTGACATAACTGACTACT	新型	NGG
TGTTGCCCAAGGTCTTACAT	新型	NAG

[1802]

向导序列	向导 id	PAM
TGTTGCTGTACCAAACCTTC	新型	NGA
TGTTGGGATTGAAGTCCCAAT	新型	NNGRRT
TGTTGGGATTGAAGTCCCAAT	新型	NNNRRT
TGTTGGTTCTTCTGGACTAT	新型	NAA
TGTTTACGTCCCGTCGGCGC	新型	NGA
TGTTTCTCAAAGGTGGAGAC	新型	NGC
TGTTTGCTCCAGACCTGCTG	新型	NGA
TGTTTGGCTTTCAGTTATAT	新型	NGA
TGTTTGGCTTTCAGTTATATG	新型	NNNRRT
TGTTTGTTTAAAGACTGGGA	新型	NGA
TGTTTTAGAAAACTTCCTAT	新型	NAA
TGTTTTCCAATGAGGATTAA	M79	NGA
TGTTTTGCTCGCAGCAGGTC	新型	NGG
TTAAACCCTAACAAAACAAA	新型	NAG
TTAAAGACAGGTACAGTAGA	新型	NGA
TTAAAGACTGGGAGGAGTTG	新型	NGG
TTAAAGGTCTTTGTACTAGG	新型	NGG
TTAAATGTATACCCAAAGAC	新型	NAA
TTAACACCCACCCAGGTAGC	新型	NAG
TTAACAGGCCTATTGATTGG	新型	NAA
TTAACTAGATGTTCTGGATAA	新型	NNNRRT
TTAACTGTAAGTGGGCCTAC	新型	NAA
TTAATACCCTTATCCAATGG	新型	NAA
TTAATCATTACTTCCAAACT	新型	NGA
TTAATGAGAAAAGAAGATTG	新型	NAA
TTACACACTCTATGGAAGGC	新型	NGG
TTACACCAAGACATTATCAA	新型	NAA
TTACAGTTAATGAGAAAAGA	新型	NGA
TTACCATTGGATAAGGGTAT	新型	NAA

[1803]

向导序列	向导 id	PAM
TTACCCCGTTGCCCGGCAAC	新型	NGC
TTACGCGGACTCCCCGTCTG	新型	NGC
TTACTCTCGTTTTTGCCTTC	新型	NGA
TTACTGCTCCTGAACTGGAG	新型	NNACCA
TTAGAAAACTTCCTATTAAC	新型	NGG
TTAGATTAAAGGTCTTTGTA	新型	NNAGGA
TTAGATTGAATACATGCATA	新型	NAA
TTAGGCAGAGGTGAAAAAGT	新型	NGC
TTAGGGTTTAAATGTATACC	新型	NAA
TTAGTATTCCTTGGACTCAT	新型	NAG
TTATATGGATGATGTGGTAT	新型	NGG
TTATCAGTCCCGATAATGTT	新型	NGC
TTATCGCTGGATGTGTCTGC	新型	NGC
TTCAAAGACTGTTTGTTTAA	新型	NGA
TTCAAGCCTCCAAGCTGTGCC	E22/M9	NNGRRT
TTCAAGCCTCCAAGCTGTGCC	E22	NNNRRT
TTCACCTCTGCACGTCGCAT	新型	NGA
TTCACTTTCTCGCCAACTTA	新型	NAA
TTCAGAGCAAACACAGCAAAT	M23	NNNRRT
TTCAGGTATTGTTTACACAG	新型	NAA
TTCCAAATTAACACCCACCC	新型	NGG
TTCCAATGACATAACCCATA	新型	NAA
TTCCAATGAGGATTAAAGAC	M47	NGG
TTCCAGGATCCTCAACCACC	新型	NGC
TTCCCCACCTTATGAGTCCA	新型	NGG
TTCCCCCACTGTTTGGCTTT	新型	NAG
TTCCCCCTAGAAAATTGAGA	新型	NAA
TTCCCGAGATTGAGATCTTC	新型	NGC
TTCCCGATACAGAGCTGAGG	新型	NGG

[1804]

向导序列	向导 id	PAM
TTCCGAAAGCCCAGGATGAT	新型	NGG
TTCCGCAGTATGGATCGGCA	新型	NAG
TTCCGGAAACTACTGTTGTT	新型	NGA
TTCCGGAAGTGTTGATAGGA	新型	NAG
TTCCGTCCGAAGGTTTGGTA	新型	NAG
TTCCTAATATACATTTACAC	新型	NAA
TTCCTATGGGAGTGGGCCTC	新型	NGC
TTCCTGTCTGGCGATTGGTG	新型	NAG
TTCCTTGAGCAGTAGTCATG	新型	NAG
TTCCTTGGACTCATAAGGTG	新型	NGG
TTCGCACTCCTCCAGCTTAT	新型	NGA
TTCGCACTCCTCCAGCTTAT	新型	NNACCA
TTCGCTTCACCTCTGCACGT	新型	NGC
TTCTCAAAGGTGGAGACAGC	新型	NGG
TTCTCATCTGCCGGACCGTG	新型	NGC
TTCTCGAGGATTGGGGACCC	新型	NGC
TTCTCTCAATTTTCTAGGGG	新型	NAA
TTCTCTTCCAAAAGTGAGAC	新型	NAG
TTCTCTTCCAAAAGTGAGACA	新型	NNNRRT
TTCTTGGGAACAAGATCTAC	新型	NGC
TTCTTGTCTCACTTTTGGAA	新型	NAG
TTCTTTCCCGACCACCAGTT	新型	NGA
TTGAACAGTAGGACATGAAC	新型	NAG
TTGAACAGTAGGACATGAACA	新型	NNNRRT
TTGAAGTCCCAATCTGGATT	新型	NGC
TTGACATACTTTCCAATCAA	新型	NAG
TTGACATTGCAGAGAGTCCA	新型	NGA
TTGAGAAACACTCATCCTCA	新型	NGC
TTGAGGATCCTGGAATTAGA	新型	NGA

[1805]

向导序列	向导 id	PAM
TTGATTGGAAAGTATGTCAA	新型	NGA
TTGCAATCTTCTTTTCTCAT	新型	NAA
TTGCAATTGATTATGCCTGC	新型	NAG
TTGCATGGTGCTGGTGCGCAG	新型	NNNRRT
TTGCCAGCAAATCCGCCTCC	新型	NGC
TTGCCCAAGGTCTTACATAA	新型	NAG
TTGCCTGAACTTTAGGCCCAT	新型	NNNRRT
TTGCCTGAGTGCAGTATGGT	新型	NAG
TTGCTCTGAAGGCTGGATCCA	新型	NNNRRT
TTGCTGACGCAACCCCCACT	新型	NGC
TTGCTGTGTTTGCTCTGAAGG	新型	NNGRRT
TTGCTGTGTTTGCTCTGAAGG	新型	NNNRRT
TTGCTTGCCTGAGTGCAGTA	新型	NGG
TTGGAAGAGAAACCGTTATA	新型	NAG
TTGGAAGATCCAGCATCTAG	新型	NGA
TTGGAGGACAGGAGGTTGGT	新型	NAG
TTGGAGGACAGGAGGTTGGTG	新型	NNNRRT
TTGGATCCAGCCTTCAGAGC	新型	NAA
TTGGCGAGAAAGTGAAAGCC	新型	NGC
TTGGCTTTCAGTTATATGGA	新型	NGA
TTGGGACTTCAATCCCAACA	新型	NGG
TTGGGATTGAAGTCCCAATC	新型	NGG
TTGGGGGAGGAGATTAGATT	新型	NAA
TTGGGTATACATTTAAACCC	新型	NAA
TTGGTAACAGCGGTAAAAAG	新型	NGA
TTGGTGAGTGATTGGAGGTT	新型	NGG
TTGGTGGTCTATAAGCTGGA	新型	NGA
TTGGTGTAAATGTATATTAG	新型	NAA
TTGGTTCTTCTGGACTATCA	新型	NGG

[1806]

向导序列	向导 id	PAM
TTGTAAGTTGGCGAGAAAGT	新型	NAA
TTGTACTAGGAGGCTGTAGGC	新型	NNNRRT
TTGTAGTATGCCCTGAGCCT	新型	NAG
TTGTATGCATGTATTCAATC	新型	NAA
TTGTCTCACTTTTGGAAGAG	新型	NAA
TTGTCTTTGGGTATACATTT	新型	NAA
TTGTGAGGATTCTTGTCAAC	新型	NAG
TTGTGGCAAGGACCCATAACT	新型	NNNRRT
TTGTGGGTCACCATATTCTT	新型	NGG
TTGTTCATGTCCTACTGTTC	新型	NAG
TTGTTGGTTCTTCTGGACTAT	新型	NNNRRT
TTGTTTACACAGAAAGGCCTT	新型	NNNRRT
TTTAAACCCTAACAAAACAA	新型	NGA
TTTAAAGACTGGGAGGAGTT	新型	NGG
TTTAAATGTATACCCAAAGA	新型	NAA
TTTAATCTAATCTCCTCCCC	新型	NAA
TTTACACAATGTGGTTATCC	新型	NGC
TTTACACACTCTATGGAAGG	新型	NGG
TTTACACAGAAAGGCCTTGT	新型	NAG
TTTACACCAAGACATTATCA	新型	NAA
TTTACCCCGTTGCCCGGCAA	新型	NGG
TTTAGAAAACTTCCTATTAA	新型	NAG
TTTATAAGGGTCGATGTCCA	新型	NGC
TTTATGGGTTATGTCATTGG	新型	NAG
TTTCAGTTATATGGATGATG	新型	NGG
TTTCCAATCAATAGGCCTGT	新型	NAA
TTTCCAATGAGGATTAAAGA	新型	NAG
TTTCCACCAGCAATCCTCTG	M80	NGA
TTTCCGAAAGCCCAGGATGA	新型	NGG

[1807]

向导序列	向导 id	PAM
TTTCCTTCAGTACGAGATCTT	新型	NNNRRT
TTTCTCAAAGGTGGAGACAG	新型	NGG
TTTCTCATTAACTGTAAGTG	新型	NGC
TTTCTCCTGGCTCAGTTTAC	新型	NAG
TTTCTCTTCCAAAAGTGAGA	新型	NAA
TTTCTGTGTAAACAATACCT	新型	NAA
TTTCTTGTCTCACTTTTGGA	新型	NGA
TTTCTTGTTGACAAGAATCCT	新型	NNNRRT
TTTGAAGTATGCCTCAAGGT	新型	NGG
TTTGAGAAACACTCATCCTC	新型	NGG
TTTGCCTTCTGACTTCTTTCC	新型	NNNRRT
TTTGCTCCAGACCTGCTGCG	新型	NGC
TTTGCTCGCAGCAGGTCTGG	新型	NGC
TTTGCTCTGAAGGCTGGATC	新型	NAA
TTTGCTGACGCAACCCCCAC	E23_whb	NGG
TTTGCTGCCCCATTTACACAA	新型	NNNRRT
TTTGCTGTGTTTGCTCTGAA	新型	NGC
TTTGGAAGAGAAACCGTTAT	新型	NGA
TTTGGAAGATCCAGCATCTA	新型	NAG
TTTGGAAGTAATGATTAACT	新型	NGA
TTTGGTGGTCTATAAGCTGG	新型	NGG
TTTGGTGTCTTTCGGAGTGT	新型	NGA
TTTGTAGGCCCACTTACAGT	新型	NAA
TTTGTAGTATGCCCTGAGCC	新型	NGA
TTTGTATGATGTGTTCTTGT	新型	NGC
TTTGTCTTTGGGTATACATT	新型	NAA
TTTGTGGGTCACCATATTCT	新型	NGG
TTTGTTTACGTCCCGTCGGCG	新型	NNGRRT
TTTGTTTACGTCCCGTCGGCG	新型	NNNRRT

[1808]

向导序列	向导 id	PAM
TTTTATGGGTTATGTCATTG	新型	NAA
TTTTCCGAAAGCCCAGGATGA	新型	NNGRRT
TTTTCCGAAAGCCCAGGATGA	新型	NNNRRT
TTTTCTCATTAACTGTAAGT	新型	NGG
TTTTGATAATGTCTTGGTGT	新型	NAA
TTTTGCTCGCAGCAGGTCTG	新型	NAG
TTTTGGAAGAGAAACCGTTA	新型	NAG
TTTTGTATGATGTGTTCTTG	新型	NGG
TTTTTGATAATGTCTTGGTG	新型	NAA

[1809] 实施例5:HBV感染的原代人肝细胞(PHH)系统的建立和验证

[1810] 通过在自组装原代肝细胞共培养系统中持续HBV感染30天,建立了包含HBV感染的原代人肝细胞(PHHs)的系统(图12A)。为了生成该共培养系统,将原代人肝细胞(PHH)(BioIVT)以35万个细胞每孔铺板到I型胶原蛋白包被的24孔板(Corning,354408)中,并在37℃、5% CO_2下孵育以生成贴壁细胞单层。铺板后4小时,用CP培养基(BioIVT)洗涤铺板的肝细胞以去除任何未贴壁的细胞。将3T3-J2小鼠胚胎成纤维细胞(Kerafast,从Howard Green(Harvard),Boston分销)以每孔95%肝细胞:5%成纤维细胞的比率接种,并在37℃、5% CO_2下再培养12小时以形成共培养物。每2天更换培养基(500μL每孔)以用于持续维持。图12C中显示了从两个不同人肝脏分离的来自两位不同人类肝细胞供体RSE和TVR的经转导的原代肝细胞的图像。它们由BioIVT(Maryland,US)分离、冷冻保存并分销用于研究目的。

[1811] 在第二天,在肝细胞共培养物形成后,将慢病毒以MOI 500滴加至每孔的培养基中。将共培养物转导16小时,然后将培养基更换为新鲜的CP培养基(BioIVT)。每2天更换培养基(500μL每孔)以用于持续维持。在历经7天后,在肝细胞共培养物中出现蛋白质表达。在转导后第7天,使用与gRNA和碱基编辑器mRNA配制在一起的基于脂质转染的试剂转染共培养物。在转染后48小时,将细胞裂解并收获gDNA。图12B提供肝细胞单层或肝细胞共培养物中的体外转导时间表的时间线,其显示了代表性的时间点。重要的是,在感染期间为使用聚乙二醇(PEG)治疗的PHH培养物倾向于具有非常低的百分比的HBV感染的PHH细胞。

[1812] 为了确定该系统是否概括了原代肝细胞的生理环境,用干扰素(其抑制复制并减少cccDNA)或替诺福韦(其抑制复制但不减少cccDNA水平)治疗感染和未感染HBV的经培养的PHH细胞,并测量HBV标记物(图13A)。如预期的,通过HBV标记物判定为感染的细胞对治疗产生应答。参照图13B,到感染后第7天,细胞外HBV DNA(黄金标准HBV复制标记物)与未经治疗的HBV感染的对照细胞相比显著减少。在经治疗的PHH培养物中检测到的细胞外HBV DNA的量类似于在阴性对照细胞培养物中观察到的量。此外,经治疗的细胞和对照细胞在感染后5至13天之间显示接近基线水平的细胞外HBV DNA,但在未经治疗的HBV感染的细胞培养物中,细胞外HBV DNA的浓度在整个观察期内(即感染后5至13天)增加。

[1813] 与HBV感染相关的其他标记物包括HBV表面抗原(HBsAg)、细胞内HBV DNA、总HBV RNA和前基因组RNA(pgRNA)。用感染时或替诺福韦治疗的培养物具有相对于阴性对照降低的水平的HBV表面抗原(HBsAg)、细胞内HBV DNA、总HBV RNA和前基因组RNA(pgRNA)(分别为图13C、13D、13E和13F)。这些结果显示,HBV标记物如预期的对治疗产生应答。它还确定PHH系统是生理相关的并且可用来评估HBV复制和cccDNA活性。

[1814] 实施例6:用BE4和gRNA进行转染导致了HBV感染的PHH中HBV标记物水平的降低

[1815] 使用本发明的碱基编辑器和向导RNA对PHH细胞中的HBV进行碱基编辑的疗效使用上述系统评估。详而言之,用编码BE4碱基编辑器(具有或不具有UGI结构域)和gRNA的构建体转染细胞。选择M52和M190向导RNA。这些gRNA分别将碱基编辑器引导至HBV基因组的pol和X基因区域并促进碱基变化,在这些基因中导致提前终止密码子。参照图14,用BE4(不具有UGI结构域)和HBV靶向gRNA治疗的细胞一贯地导致全部所测试的HBV标记物的显著减少,类似于使用干扰素治疗时的观察结果。不具有UGI结构域的BE4碱基编辑器比具有UGI结构域的BE4表现更好。不受限于理论,忽略UGI结构域使得C->U脱氨基作用对尿嘧啶糖基化酶敏感,该酶损坏HBV cccDNA,从而促进其降解。

[1816] 实施例7:基于HBV-PHH中的HBsAg水平(cccDNA的替代性标记物)的筛选鉴定功能性gRNA

[1817] 使用实施例5中所述的系统鉴定效果较佳的功能性向导RNA。将候选的向导RNA与BE4碱基编辑器一起引入HBV感染的PHH中,并确定HBsAg的水平。在用功能性向导RNA治疗的细胞中观察到的HBsAg水平低于在未经治疗的细胞中观察到的水平(图15),并且有几个达到了在干扰素治疗的细胞中观察到的水平。在用靶向前核心区、pol和X基因的功能性向导RNA治疗的细胞中的HBsAg水平显示在图15中。当与HBV靶向gRNA联合使用时,不具有UGI结构域的BE4碱基编辑器表现更好,证据为其HBsAg水平低于具有UGI结构域的BE4碱基编辑器。基于该筛选的结果,选择最佳的gRNA(靶向X基因的gRNA191和靶向pol基因的gRNA12)。在用gRNA191和BE4_noUGI治疗的细胞中的HBsAg(cccDNA的替代性标记物)水平与使用干扰素治疗时所得的HBsAg水平相当。

[1818] 实施例8:作用于HBV的碱基编辑的机制方面

[1819] 用各自单独使用或与gRNA 191或12联合使用的编码BE4碱基编辑器、缺乏UGI结构域的BE4碱基编辑器、缺乏切口酶活性(即,“死”)并且还缺乏UGI结构域的BE4碱基编辑器、Cas9和死Cas9的mRNA转染HBV感染的PHH细胞。gRNA 191将碱基编辑器靶向以将提前终止密码子引入X基因中,而gRNA 12将碱基编辑器靶向至HBV Pol/S基因的保守区域。

[1820] 作为对于用碱基编辑器和gRNA进行转染的应答,HBsAg和HBV细胞外DNA(HBV的标记物)显著减少(图16A和16B)。将HBsAg水平归一化至未经治疗的HBV感染的细胞(作为HBV参考)。核酸酶活性对于Cas9活性而言是必要的,但对于碱基编辑活性而言则不需要。有趣的是,仅引入Cas9 mRNA而不使用向导RNA也降低了病毒参数。不受限于理论,mRNA可诱导细胞免疫应答,这有助于HBV抑制。尽管HBV可以游离形式地存在,但它也可以整合到宿主细胞基因组中。缺乏UGI结构域的BE4碱基编辑器与gRNA一起给药显示与Cas9与gRNA一起给药相当的活性。如果通过尿嘧啶糖基化酶的活性而在经转换的C→U上形成无碱基位点,则缺乏切口酶活性的碱基编辑器(例如,死BE4)可能是有利的,因为它削弱产生基因组双链断裂的潜力(参见图3A)。不受限于理论,为了治疗慢性乙型肝炎,包括以下的复杂途径可能是有利

的:1)靶向cccDNA;2)抑制HBsAg合成(包括从整合的HBV DNA合成);和3)刺激免疫系统。这些特征可以在用于碱基编辑的试剂中发现。

[1821] 实施例9:在HBV-lenti-HepG2和HBV感染的PHH中的碱基编辑的比较

[1822] 对在HEPG2-NTCP Lenti-HBV和HBV感染的PHH细胞中的碱基编辑效率进行比较。在HBV感染的PHH细胞中观察到了与HEPG2-NTCP Lenti-HBV细胞相比显著降低的编辑率(图17A和17B)。此外,针对两种不同的细胞类型,观察到了不同模式的gRNA疗效。当使用gRNA190时,HEPG2-NTCP Lenti-HBV细胞显示比PHH-HBV细胞高的碱基编辑效率(图18)。有趣的是,在PHH-HBV细胞中没有观察到插入缺失或颠换(图18)。如预期的,在用gRNA190和BE4_noUGI编辑的Lenti-HBV细胞中观察到了显著数量的插入缺失和颠换。使用BE4_noUGI时的PHH-HBV中没有插入缺失或颠换,指示经编辑的cccDNA与被整合的HBV DNA的不同命运。

[1823] 实施例10.使用HBV病毒感染的原代肝细胞共培养物(PHH)作为临床相关系统来评估碱基编辑试剂的抗病毒活性

[1824] 执行实验以确定使用本文所述的HBV感染的原代人肝细胞(PHH)培养物的碱基编辑途径是否在降低几个不同的病毒参数方面与已知HBV抗病毒药物(恩替卡韦)和/或其他对照相比更为有效。

[1825] 实验采用使用HBV感染的PHH培养物系统的途径并且用如图19中所示的碱基编辑器系统的组分(碱基编辑器)转染PHH。用HBV转染原代肝细胞(PHH)共培养物,以便测试碱基编辑器的抗病毒疗效。通过两次脂质转染历经两周进程转染碱基编辑试剂或组分(碱基编辑器mRNA和合成gRNA)。第一次转染在感染后3天执行,以允许在转染的时间点完全形成病毒的共价闭合环状DNA(cccDNA)。在实验的整个进程中监控细胞外参数(HBsAg、HBeAg和HBV DNA),并且在实验结束时监控细胞内参数(HBV DNA、病毒RNA和编辑)。简而言之,在第0天,用HBV(MOI=500)转染PHH。在HBV感染后第3天,用碱基编辑组分(例如,BE和gRNA)或对照转染细胞。在第10天,对PHH进行另一轮转染。在实验结束时,例如,在第17天或更久,例如,在第25天,将PHH裂解并且评估细胞外HBsAg、HBeAg和HBV DNA的量。还评估了细胞内HBV DNA、病毒RNA的量,以及由碱基编辑器系统的组分进行碱基编辑的效率,并与对照进行比较。转染细节:RNA格式,每孔800ng总RNA,根据供应商的方案,在24孔板(600ng碱基编辑器+200ng gRNA)中通过脂质转染(lipofectamin messengerMax,Thermofisher Scientific)进行转染。

[1826] 为了确定使用本文所述的碱基编辑试剂时靶向了HBV生命周期的那个步骤,评估了四(4)个不同的病毒参数(HBsAg、HBeAg、HBV DNA和从cccDNA生产的前基因组RNA(pgRNA))。从cccDNA生成的前基因组RNA(pgRNA)在病毒基因组扩增和复制中发挥重要作用。pgRNA和cccDNA的肝表达水平指示被感染细胞中的病毒持续存在和复制活性。

[1827] 在14天PHH共培养物实验中,用碱基编辑器BE4或BE4_noUGI和gRNA(即,gRNA12,其靶向HBV聚合酶和S基因的交点)转染HBV感染的PHH(图19),该实验包括常用的HBV抗病毒药物恩替卡韦作为比较剂。恩替卡韦将病毒聚合酶灭活,从而抑制复制,但不干扰cccDNA活性。基于其在本领域的用途,恩替卡韦治疗减少了HBV DNA,然而,它不影响HBsAg和pgRNA的表达。该结果指示,所建立的系统是临床相关的,用于评估HBV抗病毒试剂的疗效。来自该实验的发现表明,与恩替卡韦相比,用BE4_noUGI和gRNA12治疗HBV感染的PHH导致了所评估的

全部4种HBV标记物参数即HBV DNA、HBsAg、HBeAg和pgRNA的减少,因此指示使用本文所述的组分进行碱基编辑抑制了HBV复制并中断了cccDNA活性(图20)。特别地,碱基编辑器BE4_noUGI+gRNA提供了与使用碱基编辑器BE4_UGI和使用恩替卡韦相比更大的全部HBV参数降幅(图20)。

[1828] 实施例11.与已知HBV抗病毒药物(恩替卡韦)相比,碱基编辑在抑制HBV及降低HBV病毒参数方面有效

[1829] 为了评估不同gRNA的组合(多重gRNA)在HBV感染的PHH培养物中是否将会导致与单个gRNA相比有所改善的HBV抑制,执行了使用至多4种靶向HBV cccDNA的不同区域的gRNA的组合进行的实验。该实验针对BE4和BE4_noUGI两种碱基编辑器进行,并在实验结束时针对4个病毒参数(HBsAg、HBeAg、HBV中DNA、pgRNA)对结果进行评估。与先前的结果相一致,用单个gRNA和BE4对PHH进行转染导致了特定病毒参数的小幅但明显的降低;抑制水平取决于gRNA。对于所评估的全部参数,BE+gRNA19的效果略优于所测试的其他gRNA。使用多重gRNA(gRNA19+gRNA190)和BE时检测到了HBV抑制的小幅改善,而BE加上四种gRNA(190+12+40+52)在抑制HBV方面相对最佳(图21、22、23和24)。

[1830] 表26呈现本文所述实验(例如,图21至27)中使用的gRNA和相应的核苷酸序列。在表26中,a、c、g、u:2'-O-甲基残基;s:硫代磷酸酯;A、C、G、U:RNA核碱基残基。

[1831]

表 26

向导	gRNA 鉴定者	序列
CTGCCAACTGGATCCTGCGC	M191 (gRNA191)	5’ csusgs CCAACUGGAUCCUGCGC GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUsususu-3’
GCTGCCAACTGGATCCTGCG	M190 (gRNA190)	5’ gscsus GCCAACUGGAUCCUGCG GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUsususu-3’
GAAAGCCCAGGATGATGGGA	M37 (gRNA37)	5’ gsasas AGCCCAGGAUGAUGGGA GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUsususu-3’
TCCGCAGTATGGATCGGCAG	E19 (gRNA19)	5’ uscscs GCAGUAUGGAUCGGCAG GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUsususu-3’
GACTTCTCTCAATTTTCTAG	E12 (gRNA12)	5’ gsascs UUCUCUCAAUUUUCUAG GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUsususu-3’
TCAATCCCAACAAGGACACC	M52 (gRNA52)	5’ uscsas AUCCCAACAAGGACACC GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUsususu-3’
TCCTCTGCCGATCCATACTG	E20 (gRNA20)	5’ uscscs UCUGCCGAUCCAUACUG GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUsususu-3’
CCATGCCCCAAAGCCACCCA	M40 (gRNA40)	5’ cscsas TGCCCCAAAGCCACCCA GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUsususu-3’
TATGGATGATGTGGTATTGG	M94 (gRNA94)	5’ usasus GGATGATGTGGTATTGG GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUsususu-3’

[1832] 图22显示从HBV感染的PHH纯化的总DNA和通过ExoI/ExoIII消化来富集cccDNA的相同样品的下一代测序(NGS)结果。图22显示在不同条件(单个gRNA或gRNA的组合)下使用

BE4在特定HBV靶标位点处的碱基编辑结果。对于富集HBV cccDNA的样品,观察到了显著增加的碱基编辑(增加的%编辑),这指示成功的cccDNA编辑。发现从HBV-PHH中纯化的总DNA(基因组DNA加上病毒DNA)中的%编辑下降,指示经编辑的cccDNA失去了传播进入可复制型病毒颗粒内的能力。

[1833] 与BE4(即BE4-UGI)相比,用碱基编辑器BE4_noUGI和不同的gRNA进行转染导致了对四个病毒参数更大程度的抑制,该结果与先前实验相一致(图23)。发现,在实验中,与4种gRNA(gRNA12、gRNA19、gRNA191和gRNA40)相比,gRNA19在抑制HBV方面表现最好。用BE4_noUGI和gRNA19进行转染导致了稳健的抗病毒应答并且导致了所测试的全部病毒参数(HBsAg、HBeAg、pgRNA和HBV中DNA)的减少。使用BE4_noUGI与多重gRNA时没有检测到抗病毒疗效的改善。与使用BE4_noUGI与gRNA的组合时相比,用BE4_noUGI和gRNA19进行转染导致了病毒参数的相同水平的减少(图23)。

[1834] 还对从HBV感染的PHH纯化的总DNA并且对通过ExoI/ExoIII消化来富集cccDNA的相同样品执行NGS测序。图24显示在不同条件(单个gRNA或gRNA的组合)下使用BE4_noUGI对特定HBV靶标位点的碱基编辑。对于富集HBV cccDNA的样品,观察到了显著增加的碱基编辑(增加的%编辑),这指示成功的cccDNA编辑。发现从HBV-PHH中纯化的总DNA(基因组DNA加上病毒DNA)中的%编辑下降,指示经编辑的cccDNA失去了传播进入可复制型病毒颗粒内的能力(图24)。

[1835] 实施例12没有切口酶活性的碱基编辑器在减少感染的HBV参数方面有效

[1836] 碱基编辑器BE4_noUGI具备切口酶活性,可在所靶向的DNA位点中造成双链断裂(dsbreak)(Komor AC et al.,2016).cccDNA中的双链断裂可进一步导致HBV DNA被整合到人类基因组中,而这是应该避免的。为了最小化在碱基编辑过程期间产生dsbreak的可能性,生成并评估了不具有切口酶活性而仅具备脱氨酶活性的突变BE4_noUGI(H840A),即dBE4_noUGI。在长期(25天)实验中,使用HBV感染的PHH评估dBE4_noUGI与gRNA12(Pol/S)协同作用的碱基编辑活性。尽管单独使用碱基编辑器(无gRNA)进行转染不导致所评估的病毒参数的显著减少,但用dBE4_noUGI和gRNA12进行转染导致了稳健的HBV抑制,如通过测量病毒参数HBsAg、HBV DNA和HBeAg的水平所确定(图25A至25C)。

[1837] 此外,在长期实验中,使用HBV感染的PHH,就抑制HBV方面将干扰素(其已知用于治疗HBV)的抗病毒活性与碱基编辑器dBE4_noUGI和gRNA的碱基编辑活性机芯比较。尽管干扰素可以有效地治疗一些患者的HBV,但该药剂也已知是有毒的并且可能造成大量不良副作用。实验结果表明,干扰素和使用dBE4_noUGI+gRNA12进行的碱基编辑两者均将HBV病毒参数降低至相当的水平(图25A至25C)。通过在实验结束时测试分泌到细胞培养基中的白蛋白的水平,也可以评价细胞活力和代谢活性。干扰素治疗导致了分泌白蛋白的显著减少,这指示干扰素治疗的毒性。与干扰素治疗相比,通过dBE4_noUGI和gRNA(例如,gRNA12)进行的碱基编辑造成了分泌白蛋白水平的最小限度减少,从而指示缺乏与使用碱基编辑途径相关的毒性(图25D)。

[1838] 实施例13.碱基编辑减少了HBV感染的PHH中的基因型D和C的HBV的病毒参数(长期实验)

[1839] 执行实验以确定碱基编辑途径是否针对不同基因型的HBV导致类似的抗病毒疗效。因此,在平行实验中,在HBV感染的PHH共培养物中评估dBE4_noUGI和gRNA12(靶向HBV聚

合酶和S基因的交点(P/S)),在该平行实验中,用基因型D(Gen.D)的HBV或基因型C(Gen.C)的HBV感染PHH。用基因型C的HBV感染PHH(无治疗条件)在实验结束时导致了更高的病毒负载。尽管在实验中评估了更高水平的病毒参数,但与对照相比,用碱基编辑器dBE4_no UGI和gRNA12进行转染时,HBV Gen.D和HBV Gen.C两者均显示了稳健的HBV抑制(图26A至26C)。结果支持本文所述的碱基编辑和碱基编辑器系统用于治疗由不同基因型的HBV引起的感染的用途。

[1840] 实施例14用ABE7.10和靶向HBV聚合酶活性位点的HBV特异性gRNA进行转染减少了HBV感染的PHH中的HBV标记物

[1841] 腺嘌呤碱基编辑器(ABE)具备稳健的碱基编辑活性,同时显示最低限度的脱靶效应。(Gaudelli,N.M.,et al.,"Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage"Nature 551,464-471(2017)。尽管已经报道了内源性胞苷脱氨酶将cccDNA脱氨基,但腺嘌呤脱氨酶并未表现出此类机制或活性。为了确定使用ABE进行HBV编辑是否抑制HBV感染的原代肝细胞中的病毒复制,在牵涉使用HBV感染的PHH共培养系统的实验中,评估了与gRNA协同作用的如本文所述的含有腺嘌呤脱氨酶的碱基编辑器ABE7.10。因此,用将ABE7.10 mRNA和gRNA94转染HBV感染的PHH培养物,gRNA94被设计成将静默突变引入HBV聚合酶活性位点中,如上文的实施例10和图19中所述。对实验结果的分析显示,与单独使用ABE7.10 mRNA时不造成病毒参数降低相比,联合使用碱基编辑器ABE7.10 mRNA和gRNA94(靶向HBV聚合酶活性位点)时检测到了病毒参数HBsAg、HBeAg、pgRNA和HBV总DNA的稳健下降(图27A)。对从HBV感染的PHH纯化的总DNA并且对通过ExoI/ExoIII消化来富集cccDNA的相同样品执行下一代测序(NGS)。在cccDNA富集的样品中检测到了通过ABE7.10 mRNA和gRNA94进行的大约50%的碱基编辑,指示成功的cccDNA编辑。从HBV-PHH中纯化的总DNA中不存在碱基编辑,指示经编辑的cccDNA失去了传播进入可复制型病毒颗粒内的能力(图27B)。这些数据联合表明,ABE用作有效的抗病毒试剂用于降低HBV病毒载量。

[1842] 其它实施方案

[1843] 从前述说明书可知,可对本文所述的发明作出改变和修改以令其适应各种用途和条件。此类实施方案也处于所附权利要求书的范畴内。

[1844] 本文中,对任何变量定义中一系列元件的描述包括该变量作为任何单一元件或作为所列元件的组合(或亚组合)的定义。对本文中实施方案的描述包括该实施方案作为任何单一实施方案或作为与任何其它实施方案或其部分的组合。

[1845] 本说明书中提及的全部专利和出版物通过引用并入本文,其并入程度与各独立的专利和出版物具体且独立地指明以待通过引用并入相同。

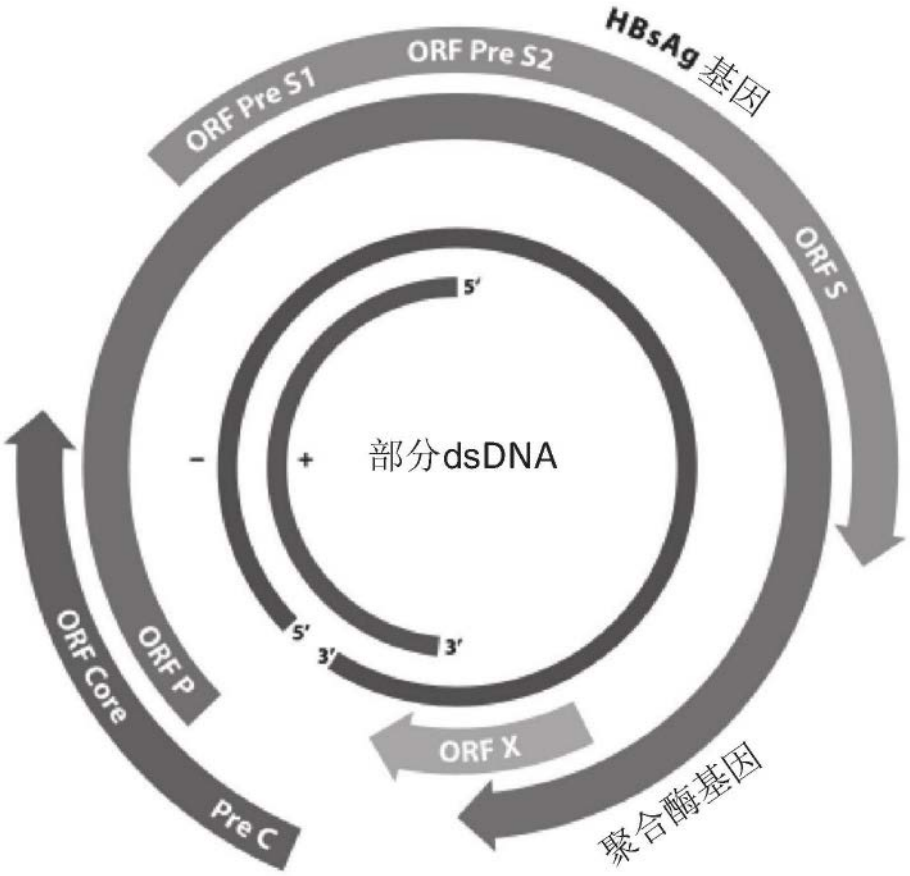

HBsAg 基因
ORF Pre S1
ORF Pre S2
ORF S
5′
部分dsDNA
-
+
5′
3′
3′
ORF P
ORF Core
Pre C
ORF X
聚合酶基因

图1

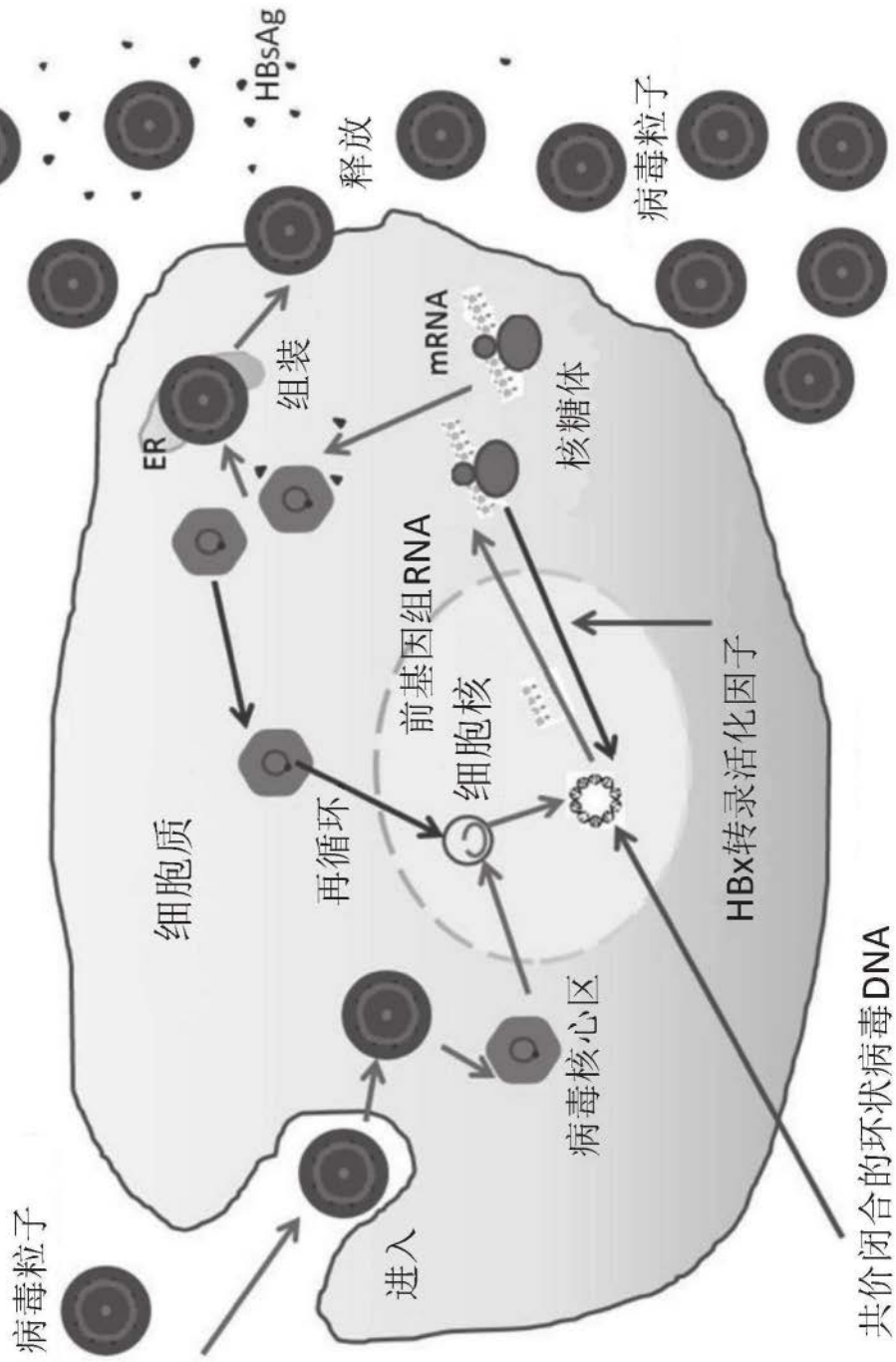

HBsAg
释放
病毒粒子
ER
组装
mRNA
核糖体
细胞质
再循环
前基因组RNA
细胞核
HBx转录活化因子
病毒核心区
进入
病毒粒子
共价闭合的环状病毒DNA

图2

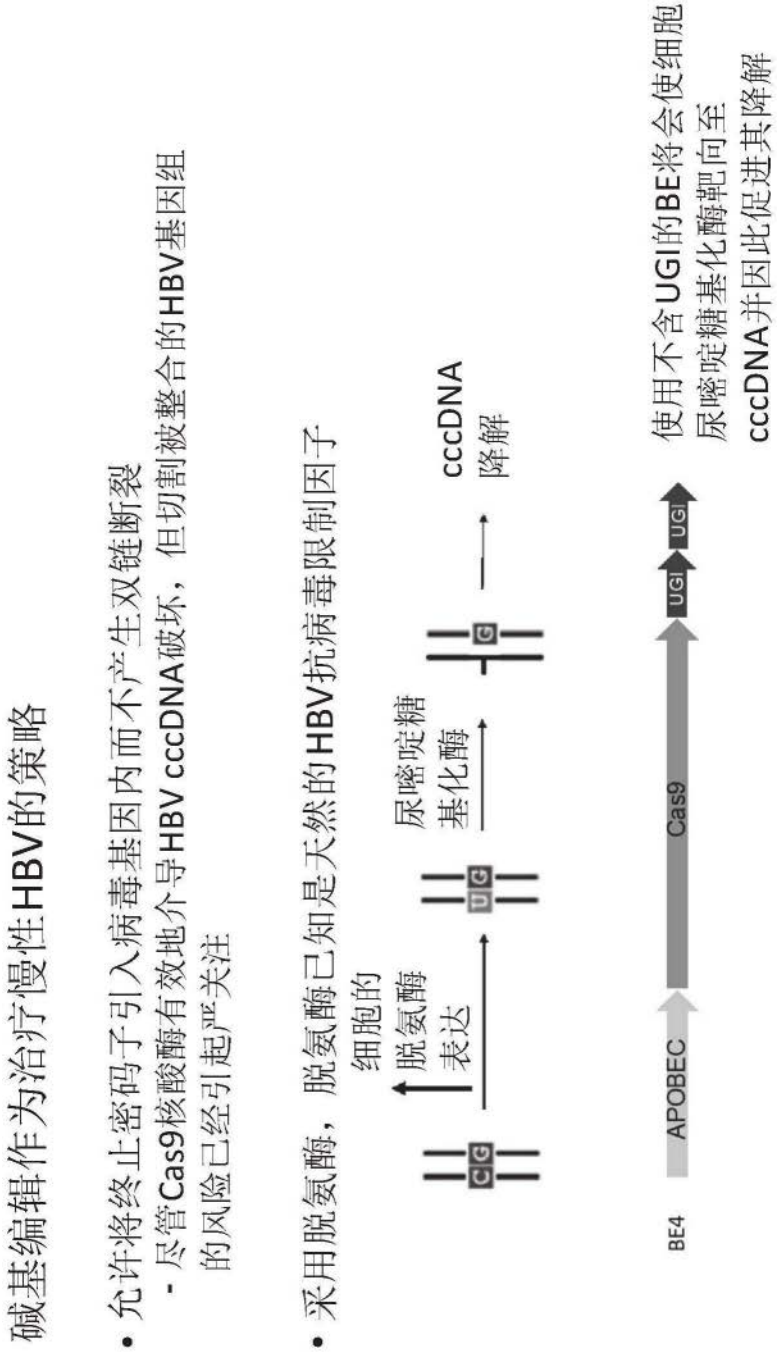

碱基编辑作为治疗慢性HBV的策略
• 允许将终止密码子引入病毒基因内而不产生双链断裂
- 尽管Cas9核酸酶有效地介导HBV cccDNA破坏，但切割被整合的HBV基因组的风险已经引起严关注
• 采用脱氨酶，脱氨酶已知是天然的HBV抗病毒限制因子
细胞的脱氨酶表达
尿嘧啶糖基化酶
cccDNA降解
BE4
APOBEC
Cas9
UGI
UGI
使用不含UGI的BE将会使细胞尿嘧啶糖基化酶靶向至cccDNA并因此促进其降解

图3A

向导RNA筛选策略

以基因型D亚基因型ayw为重点鉴定保守的HBV区

- 在美国最为常见的
- 细胞和动物模型的存在(Fabien Zoulim lab (INSERM))。人源化小鼠模型

	标准	
	筛选 1	筛选 2
gRNA设计	∘ 将终止密码子或功能性突变引入HBV基因中的gRNA	∘ 超保守gRNA ∘ 主要目的是在HBV基因组中产生无碱基位点 ∘ 将会避免已知的抗性突变体
跨HBV基因型的gRNA保守	没有初始过滤	在全部基因型中均>50%
碱基编辑器 (BE4)	具有或不具有UGI	
筛选策略	Hed293-Lenti-HBV_DNA/质粒格式的编辑效率	

图3B

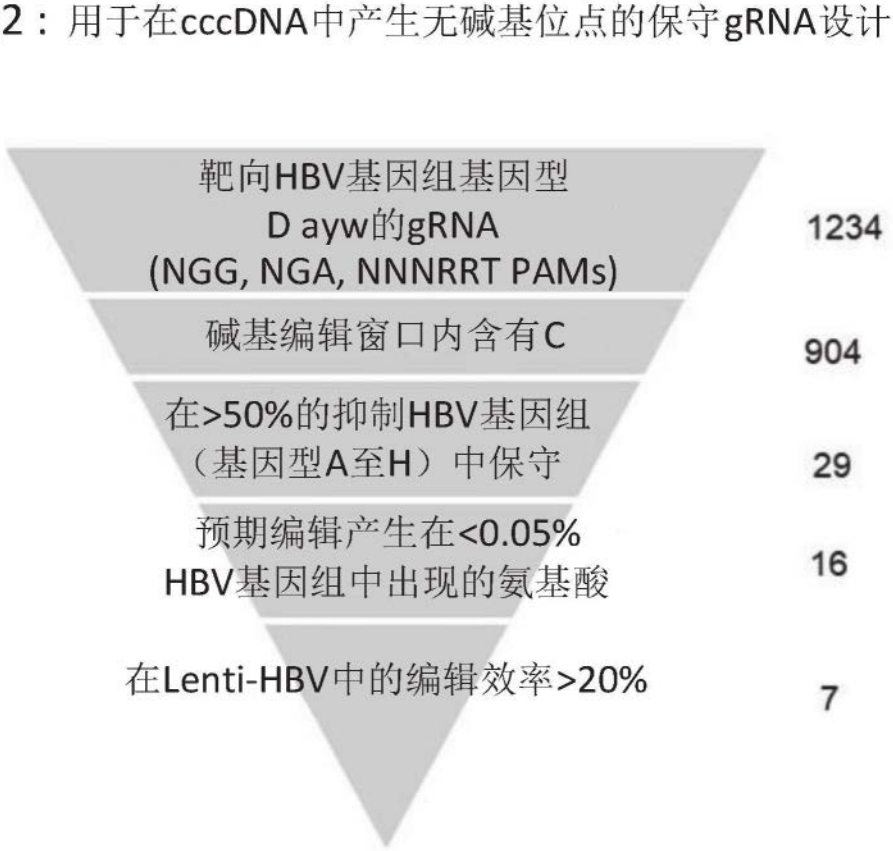

筛选2：用于在cccDNA中产生无碱基位点的保守gRNA设计
靶向HBV基因组基因型
D ayw的gRNA
(NGG, NGA, NNNRRT PAMs)
1234
碱基编辑窗口内含有C
904
在>50%的抑制HBV基因组
（基因型A至H）中保守
29
预期编辑产生在<0.05%
HBV基因组中出现的氨基酸
16
在Lenti-HBV中的编辑效率>20%
7

图3C

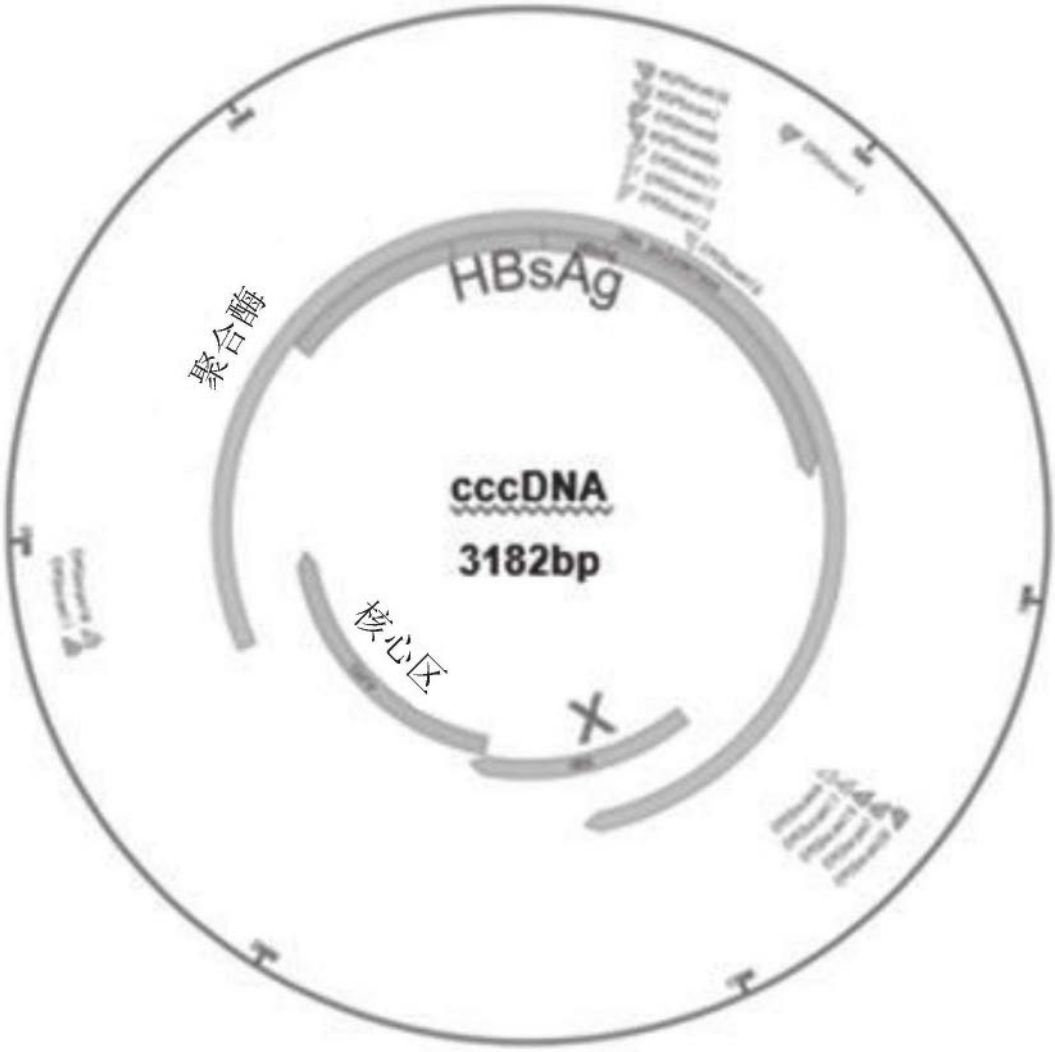


图3D

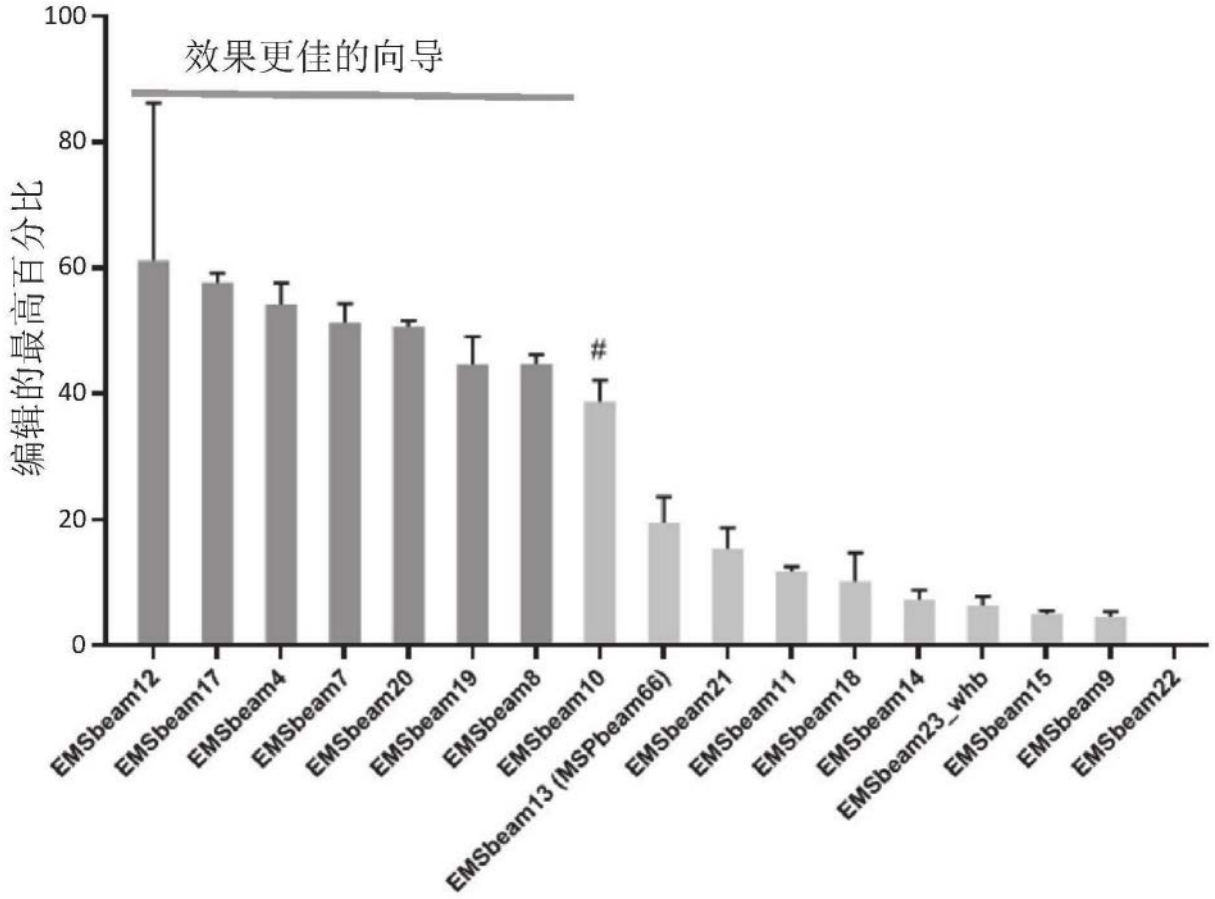


图3E

gRNA ID	编辑的最高百分比	导引链	PAM	基因	在全部基因型中的保守百分比
Cons-12	61.0	1	NGG	S/Pol	96
Cons-17	58.0	-1	NGA	Pol	93
Cons-19	45.0	-1	NGG	Pol	92
Cons-20	51.0	1	NGG	Pol	92
Cons-4	54.0	-1	NGG	Pol	95
Cons-7	51.0	-1	NGA	Pol	92
Cons-8	45.0	1	NGA	Pol	92

图3F



图4A

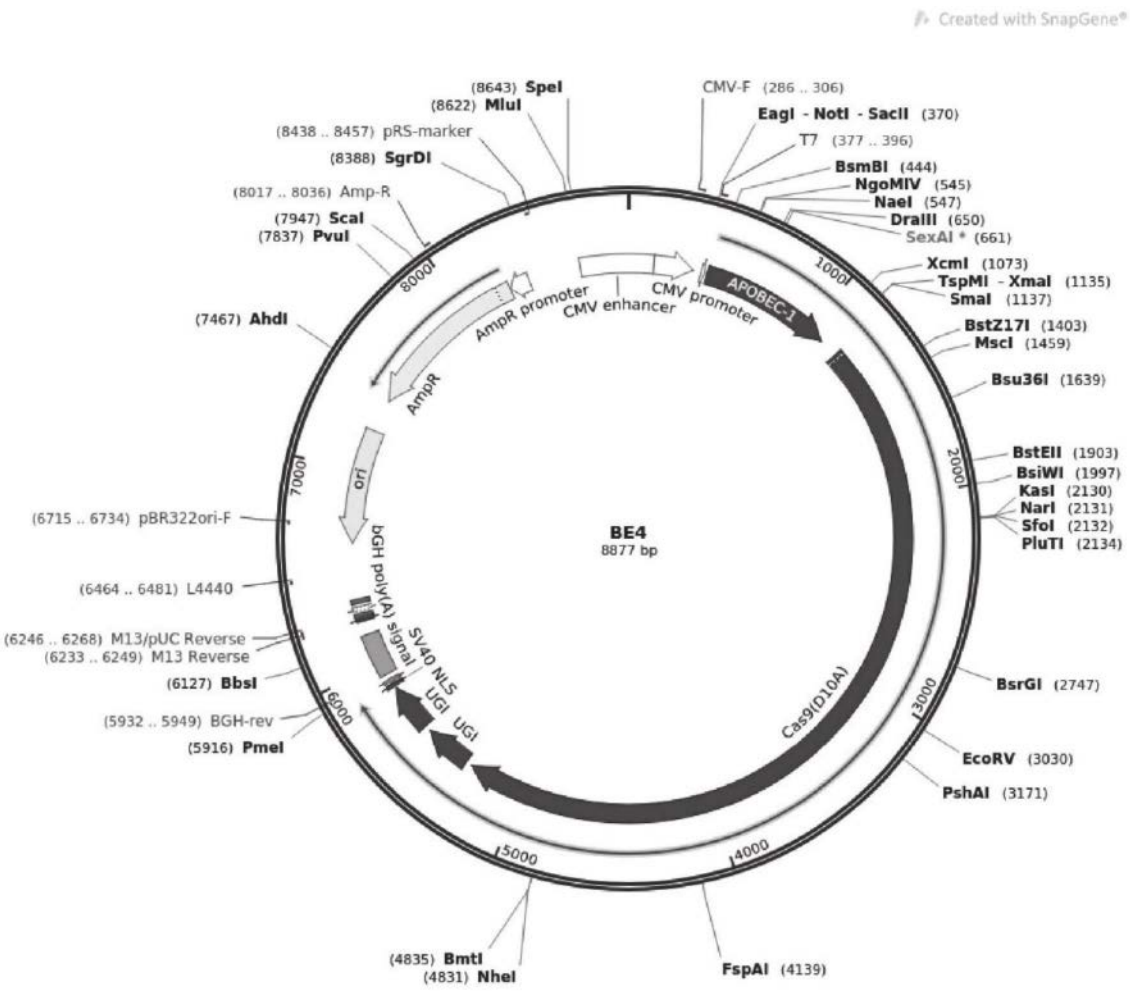

Created with SnapGene®
BE4
8877 bp
(8643) SpeI
(8622) MluI
(8438 .. 8457) pRS-marker
(8388) SgrDI
(8017 .. 8036) Amp-R
(7947) ScaI
(7837) PvuI
(7467) AhdI
(6715 .. 6734) pBR322ori-F
(6464 .. 6481) L4440
(6246 .. 6268) M13/pUC Reverse
(6233 .. 6249) M13 Reverse
(6127) BbsI
(5932 .. 5949) BGH-rev
(5916) PmeI
(4835) BmtI
(4831) NheI
FspAI (4139)
PshAI (3171)
EcoRV (3030)
BsrGI (2747)
BstEII (1903)
BsiWI (1997)
KasI (2130)
NarI (2131)
SfoI (2132)
PluTI (2134)
Bsu36I (1639)
BstZ17I (1403)
MscI (1459)
XcmI (1073)
TspMI - XmaI (1135)
SmaI (1137)
BsmBI (444)
NgoMIV (545)
NaeI (547)
DraIII (650)
SexAI * (661)
CMV-F (286 .. 306)
EagI - NotI - SacII (370)
T7 (377 .. 396)
AmpR promoter
CMV enhancer
CMV promoter
APOBEC-1
AmpR
ori
BGH poly(A) signal
SV40 NLS
UGI
UGI
Cas9(D10A)
1000
2000
3000
4000
5000
6000
7000
8000

图4B

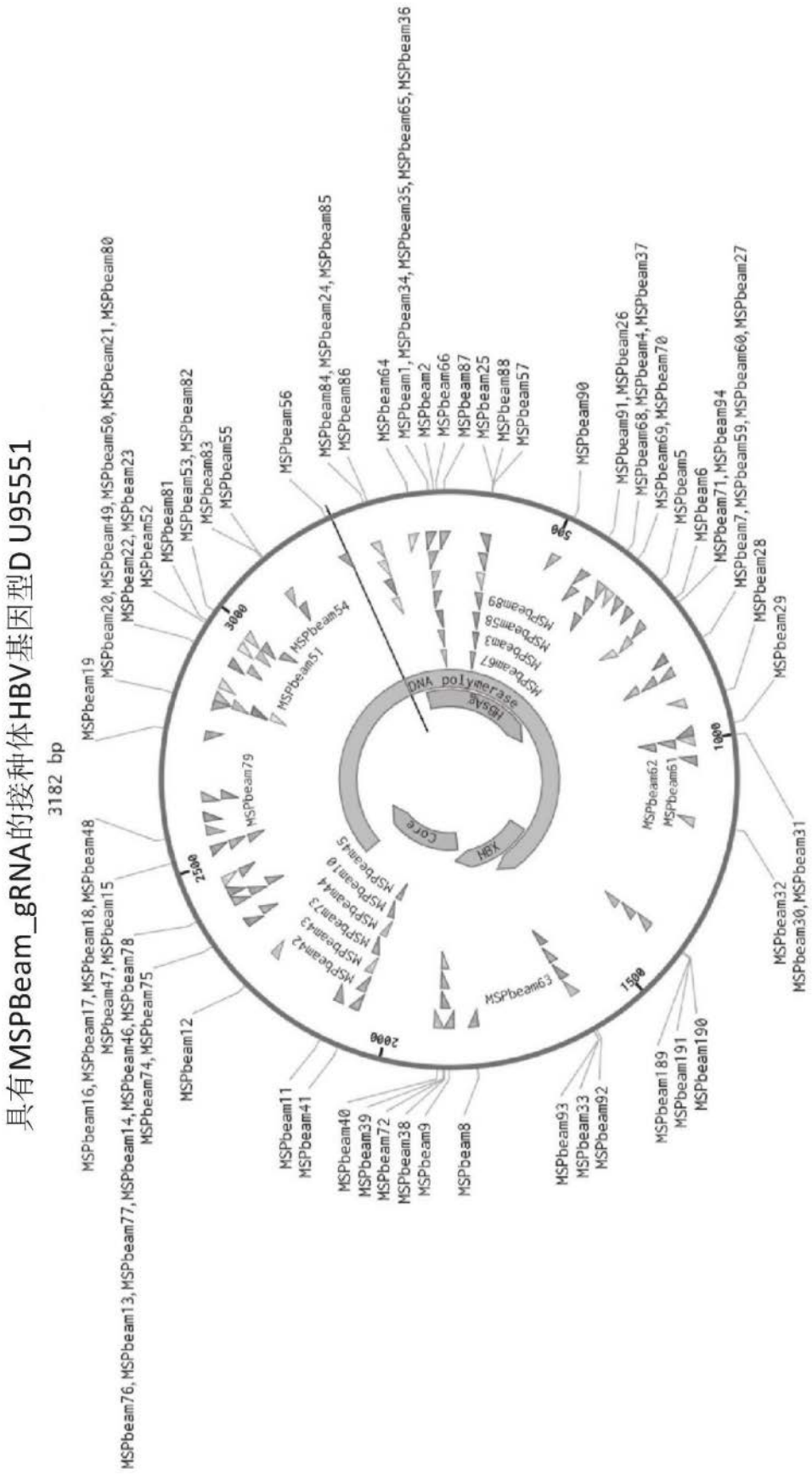

具有MSPBeam_gRNA的接种体HBV基因型D U95551
3182 bp
MSPbeam19
MSPbeam20, MSPbeam49, MSPbeam50, MSPbeam21, MSPbeam80
MSPbeam22, MSPbeam23
MSPbeam52
MSPbeam81
MSPbeam53, MSPbeam82
MSPbeam83
MSPbeam55
MSPbeam56
MSPbeam84, MSPbeam24, MSPbeam85
MSPbeam86
MSPbeam64
MSPbeam1, MSPbeam34, MSPbeam35, MSPbeam65, MSPbeam36
MSPbeam2
MSPbeam66
MSPbeam87
MSPbeam25
MSPbeam88
MSPbeam57
MSPbeam90
MSPbeam91, MSPbeam26
MSPbeam68, MSPbeam4, MSPbeam37
MSPbeam69, MSPbeam70
MSPbeam5
MSPbeam6
MSPbeam71, MSPbeam94
MSPbeam7, MSPbeam59, MSPbeam60, MSPbeam27
MSPbeam28
MSPbeam29
MSPbeam32
MSPbeam30, MSPbeam31
MSPbeam62
MSPbeam61
MSPbeam63
MSPbeam93
MSPbeam33
MSPbeam92
MSPbeam189
MSPbeam191
MSPbeam190
MSPbeam11
MSPbeam41
MSPbeam40
MSPbeam39
MSPbeam72
MSPbeam38
MSPbeam9
MSPbeam8
MSPbeam12
MSPbeam16, MSPbeam17, MSPbeam18, MSPbeam48
MSPbeam47, MSPbeam15
MSPbeam76, MSPbeam13, MSPbeam77, MSPbeam14, MSPbeam46, MSPbeam78
MSPbeam74, MSPbeam75
MSPbeam79
MSPbeam54
MSPbeam51
MSPbeam42
MSPbeam43
MSPbeam73
MSPbeam44
MSPbeam10
MSPbeam45
MSPbeam67
MSPbeam3
MSPbeam58
MSPbeam89
HBSAg
DNA polymerase
Core
HBX
500
1000
1500
2000
2500
3000

图5

12种效果更佳的HBV gRNA/所靶向的全部4种HBV基因
STOP突变
名称 停止编辑百分比 PAM 基因 在全部基因型中的保守百分比
MSPbeam52 59 NGG Pol 15
MSPbeam50 52 NGG Pol 22
MSPbeam68 45 NGA S 42
MSPbeam191 41 NGG X 72
MSPbeam67 38 NGA S 55
MSPbeam190 37 NGG X 76
MSPbeam40 30 NGG Core 65
MSPbeam39 27 NGG Core 95
MSPbeam66 27 NGA S 97
MSPbeam87 21 NGA Pol 83
MSPbeam74 21 NGA Core 52
FC
名称 停止编辑百分比 PAM 基因 在全部基因型中的保守百分比
MSPbeam94 31 NGG Pol 67
*所靶向的患者相对于全局患者群体的百分比
通过Hek293-Lenti HBV中的质粒转染筛选了总计97种引入终止密码子或造成功能性改变(FC)的gRNA
>20%上靶碱基编辑
跨全部HBV基因型的>15%保守
CONFIDENTIAL

图6

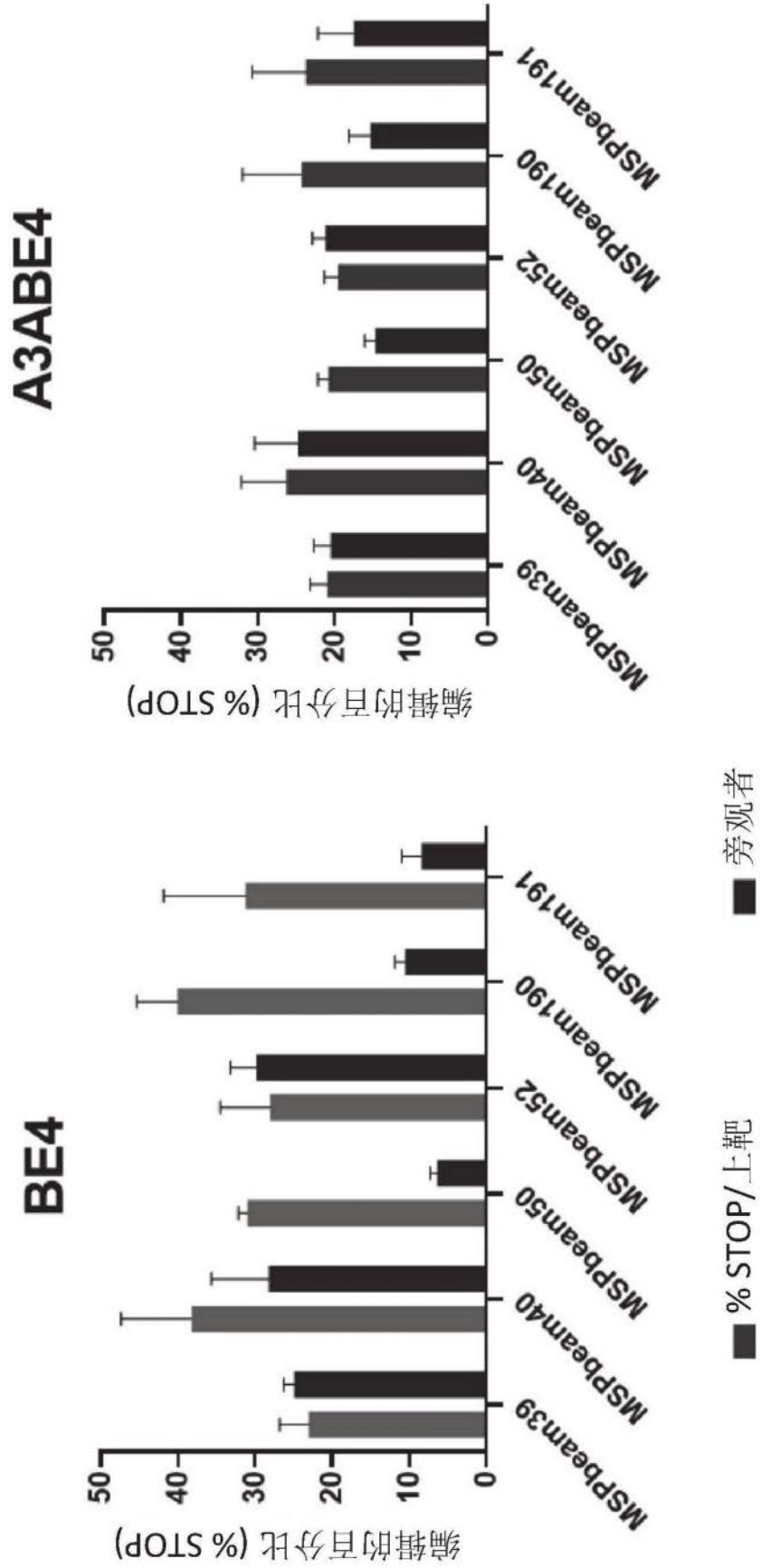

A3ABE4
编辑的百分比 (% STOP)
0
10
20
30
40
50
MSPbeam39
MSPbeam40
MSPbeam50
MSPbeam52
MSPbeam190
MSPbeam191
BE4
编辑的百分比 (% STOP)
0
10
20
30
40
50
MSPbeam39
MSPbeam40
MSPbeam50
MSPbeam52
MSPbeam190
MSPbeam191
% STOP/上靶
旁观者

图7

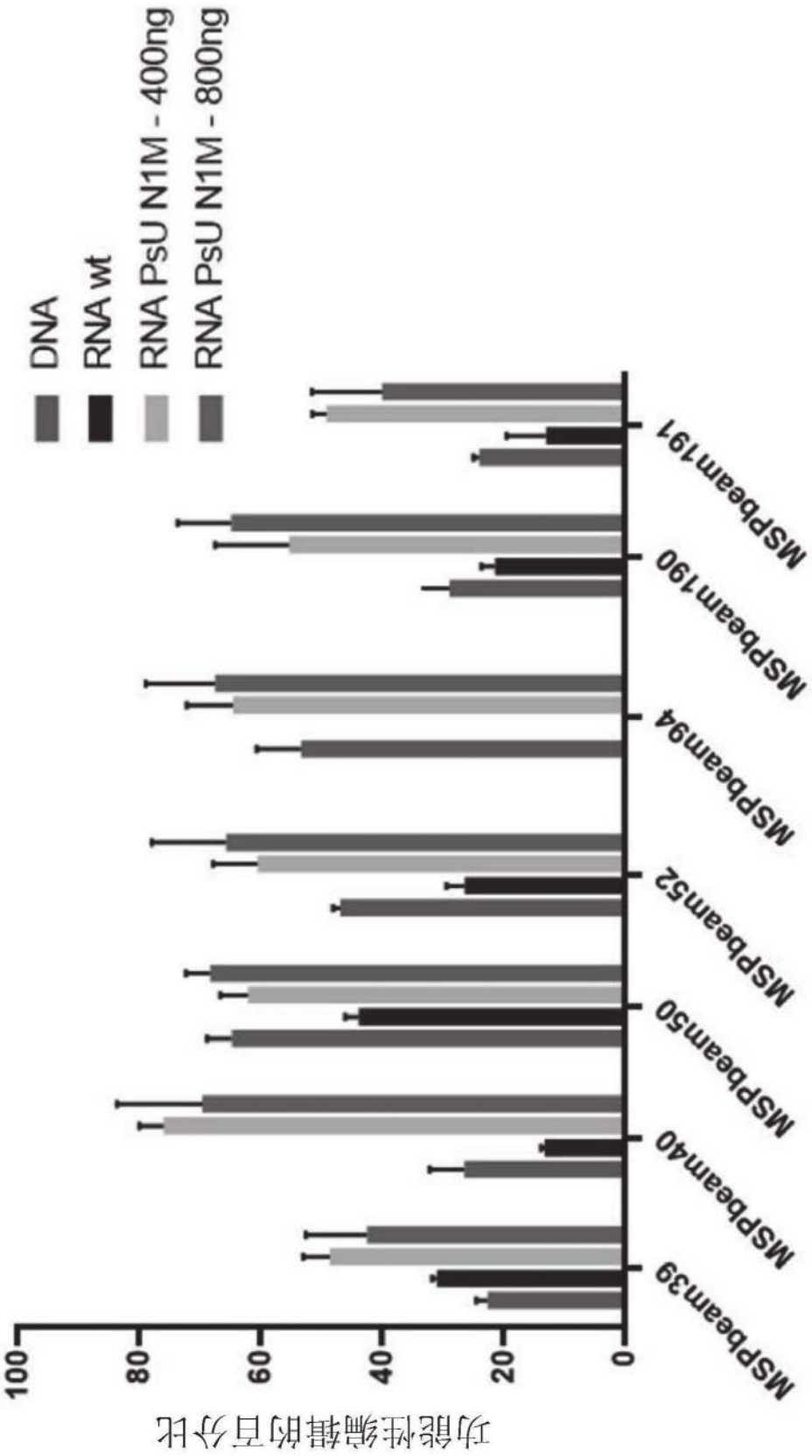

DNA
RNA wt
RNA PsU N1M - 400ng
RNA PsU N1M - 800ng
MSPbeam191
MSPbeam190
MSPbeam94
MSPbeam52
MSPbeam50
MSPbeam40
MSPbeam39
0
20
40
60
80
100
功能性编辑的百分比

图8

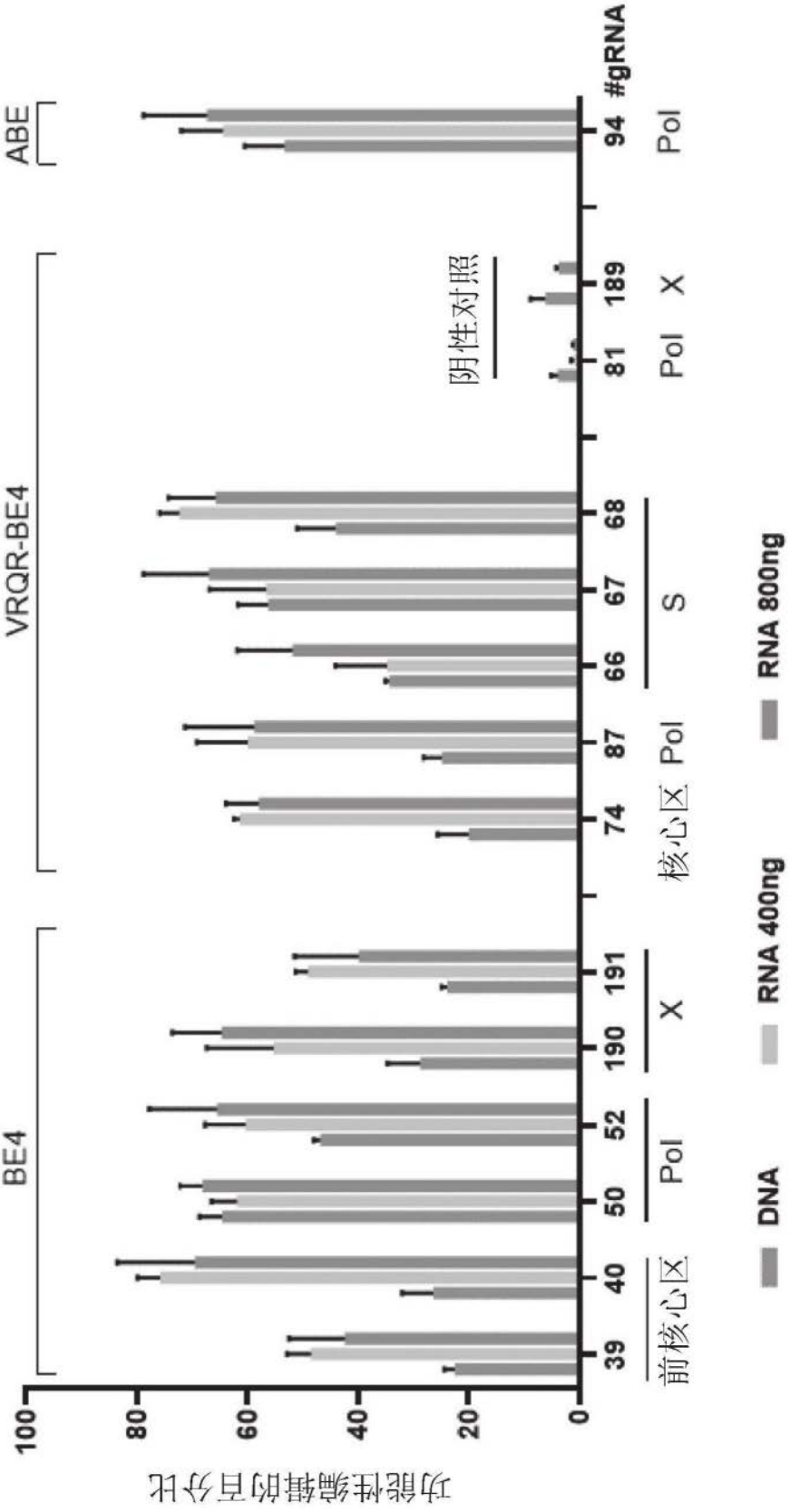

功能性编辑的百分比
100
80
60
40
20
0
BE4
ABE
VRQR-BE4
前核心区
Pol
X
核心区
Pol
S
Pol
X
Pol
阴性对照
39
40
50
52
190
191
74
87
66
67
68
81
189
94
#gRNA
DNA
RNA 400ng
RNA 800ng

图9

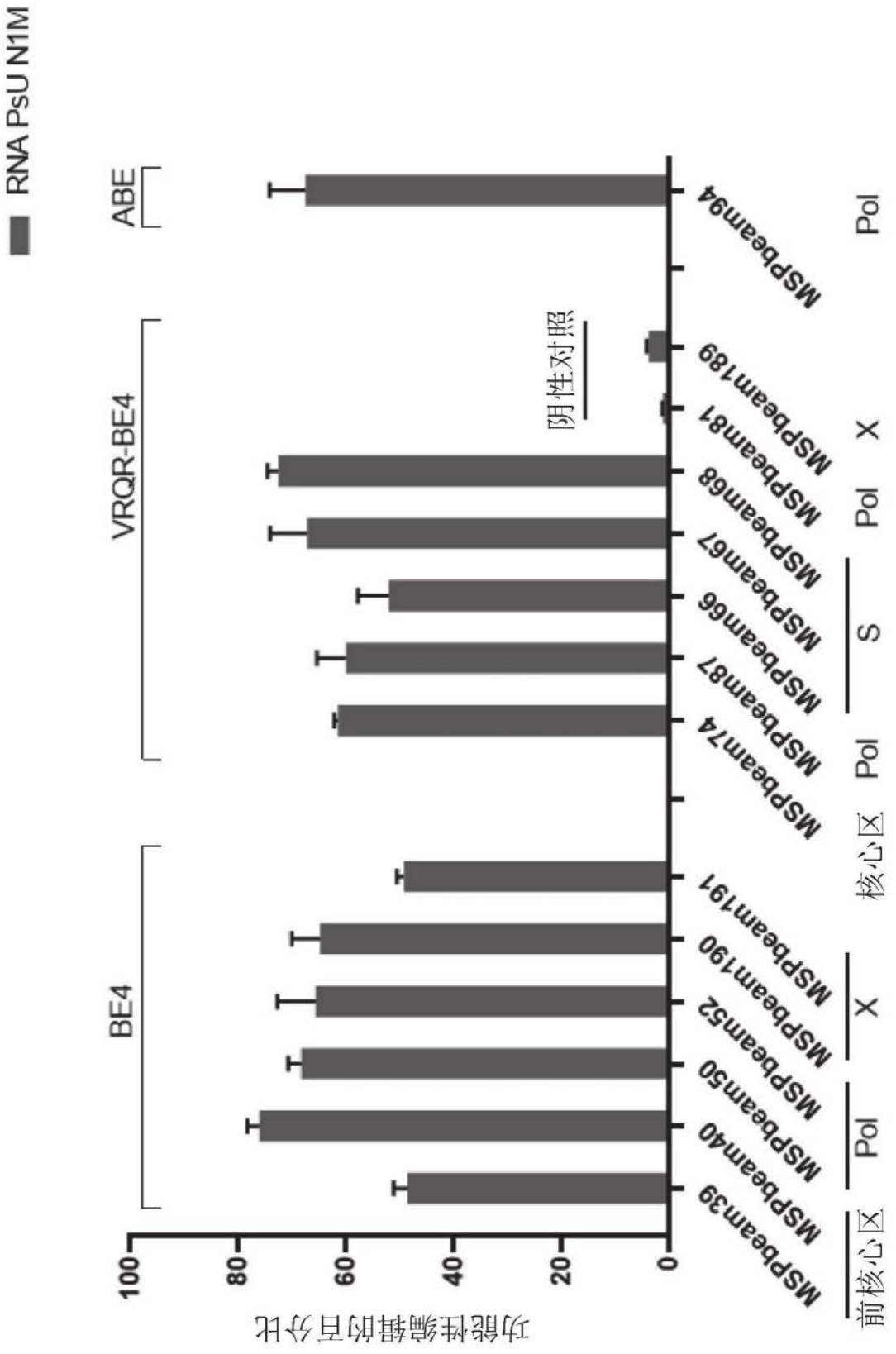

RNA PsU N1M
ABE
VRQR-BE4
BE4
阴性对照
功能性编辑的百分比
0
20
40
60
80
100
MSPbeam39
MSPbeam40
MSPbeam50
MSPbeam52
MSPbeam190
MSPbeam191
MSPbeam74
MSPbeam87
MSPbeam66
MSPbeam67
MSPbeam68
MSPbeam81
MSPbeam189
MSPbeam94
前核心区
Pol
X
核心区
Pol
S
Pol
X
Pol

图10

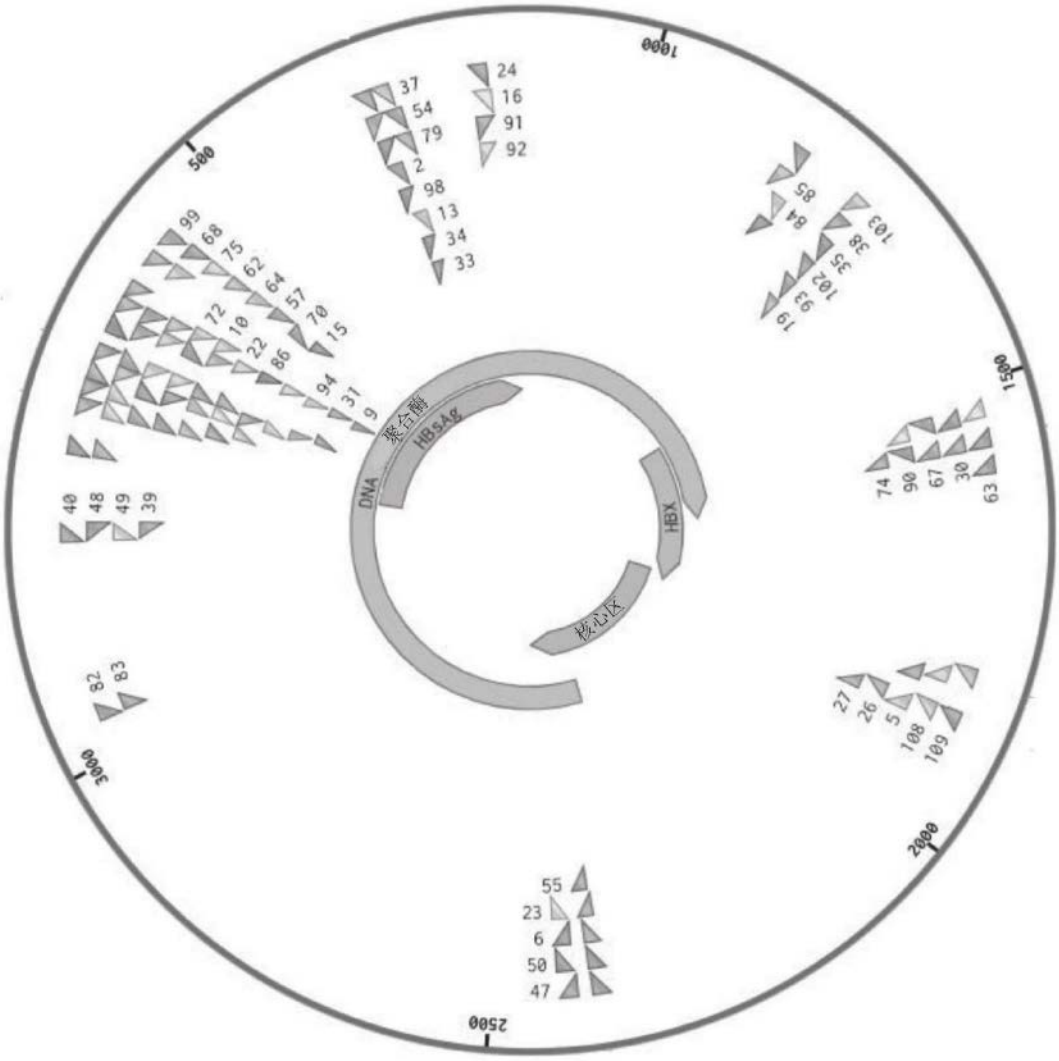

500
1000
1500
2000
2500
3000
DNA
聚合酶
HBSAg
HBX
核心区
37
54
79
2
98
13
34
33
24
16
91
92
99
68
75
62
64
57
70
15
72
10
22
86
94
31
9
40
48
49
39
82
83
85
84
103
38
35
102
93
19
74
90
67
30
63
27
26
5
108
109
55
23
6
50
47

图11

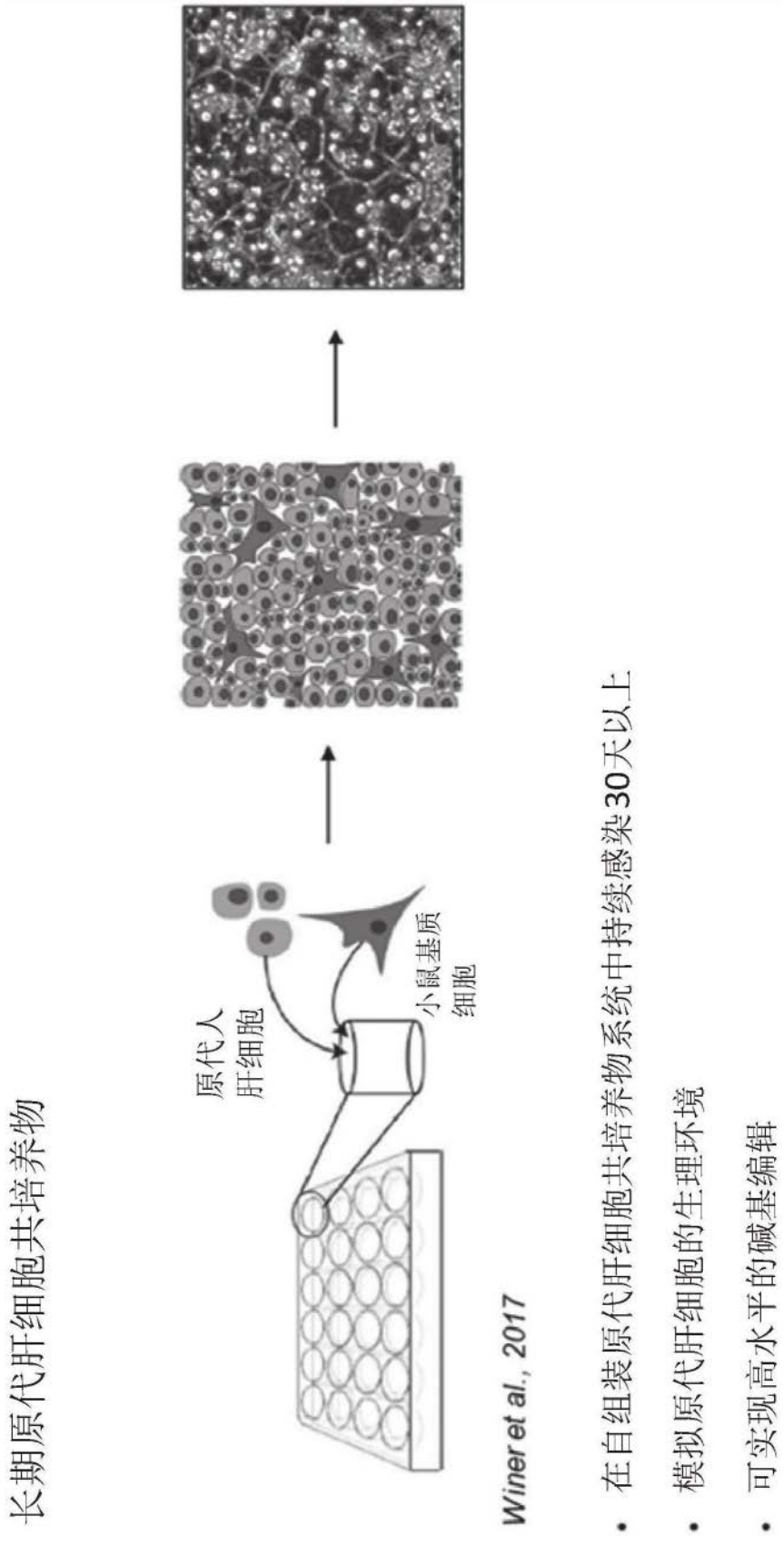

长期原代肝细胞共培养物
原代人
肝细胞
小鼠基质
细胞
Winer et al., 2017
• 在自组装原代肝细胞共培养物系统中持续感染30天以上
• 模拟原代肝细胞的生理环境
• 可实现高水平的碱基编辑

图12A

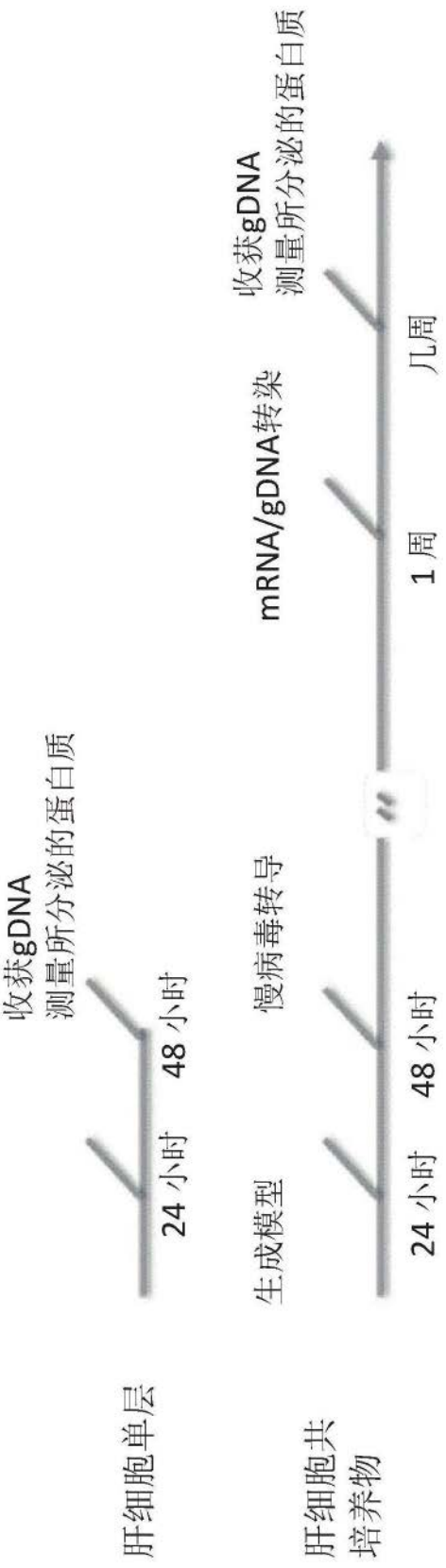

肝细胞单层
24 小时
生成模型
48 小时
收获gDNA
测量所分泌的蛋白质
肝细胞共
培养物
24 小时
慢病毒转导
48 小时
1 周
mRNA/gDNA转染
几周
收获gDNA
测量所分泌的蛋白质

图12B

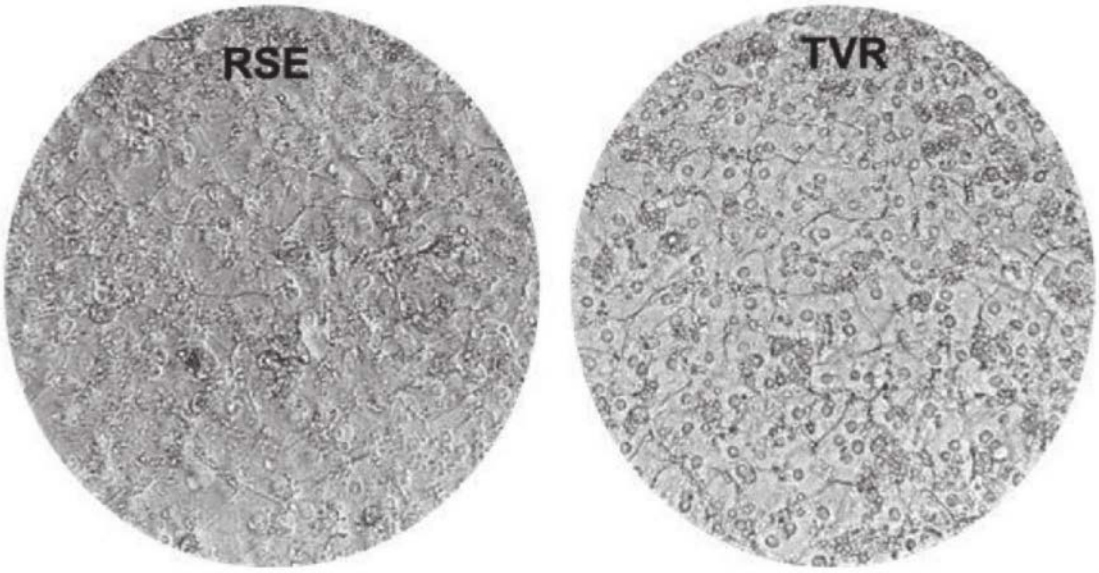


图12C

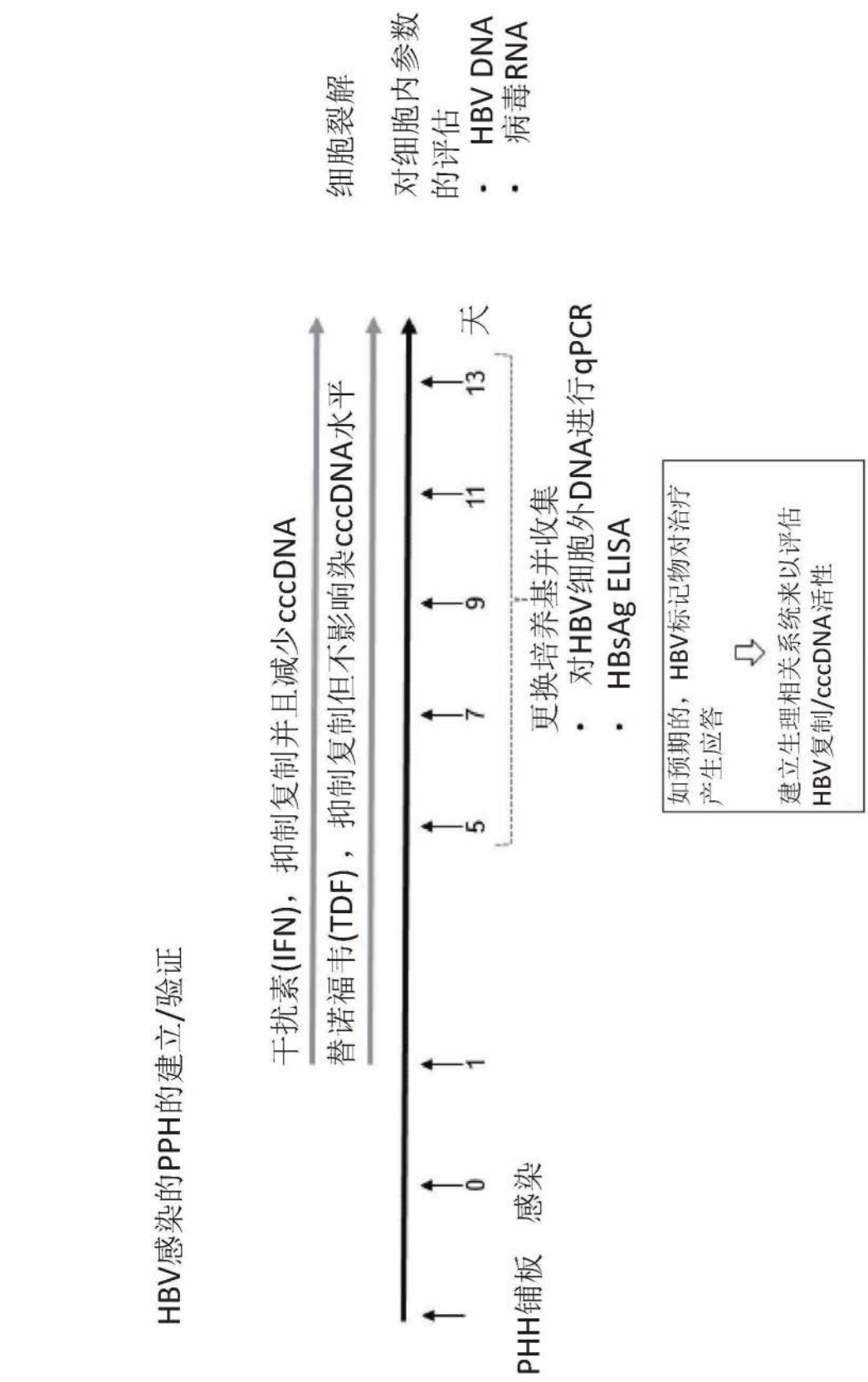

HBV感染的PPH的建立/验证
干扰素(IFN)，抑制复制并且减少cccDNA
替诺福韦(TDF)，抑制复制但不影响染cccDNA水平
细胞裂解
对细胞内参数的评估
• HBV DNA
• 病毒RNA
PHH铺板
感染
0
1
5
7
9
11
13
天
更换培养基并收集
• 对HBV细胞外DNA进行qPCR
• HBsAg ELISA
如预期的，HBV标记物对治疗产生应答
建立生理相关系统来以评估HBV复制/cccDNA活性

图13A

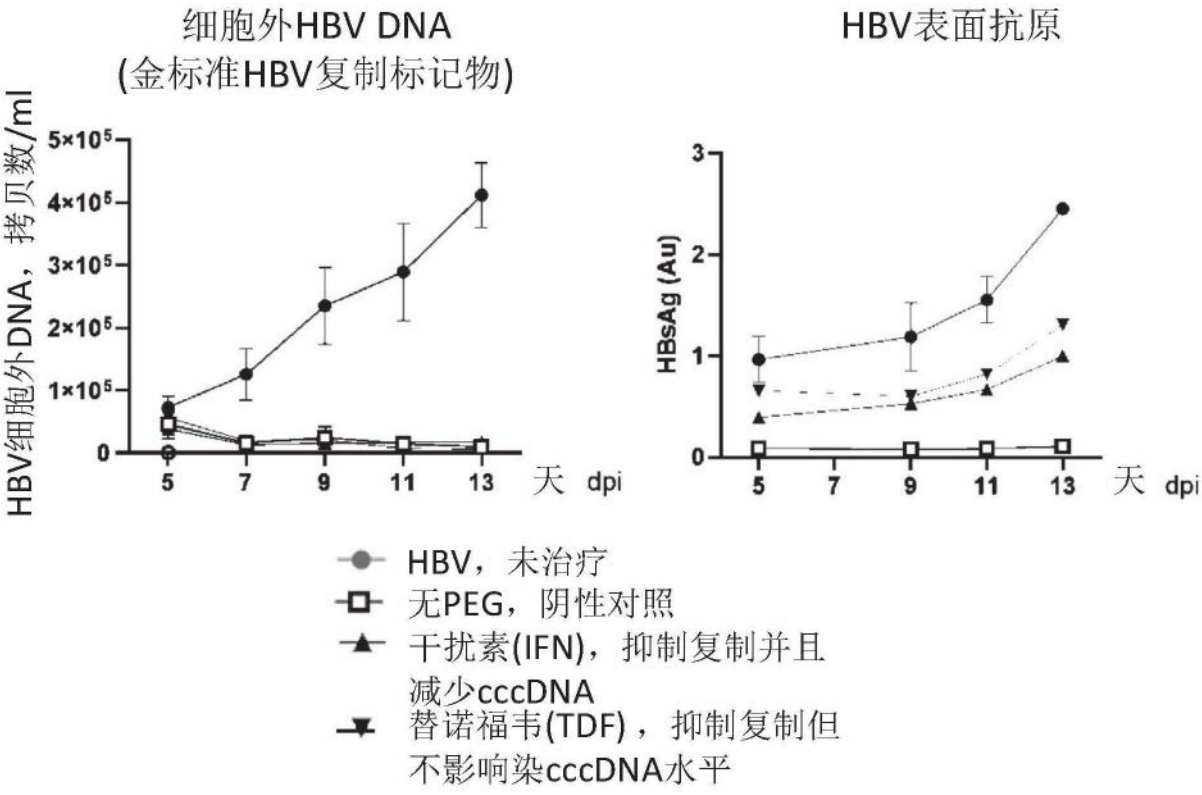


图 13B 图 13C

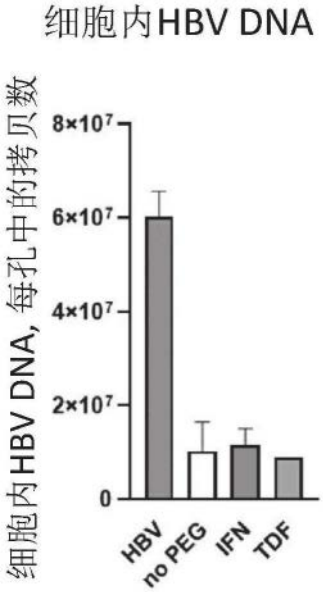


图13D

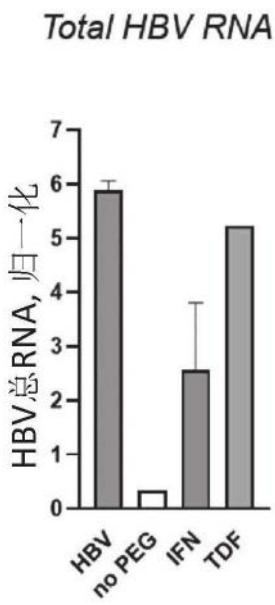

Total HBV RNA
HBV总RNA, 归一化
0
1
2
3
4
5
6
7
HBV
no PEG
IFN
TDF

图13E

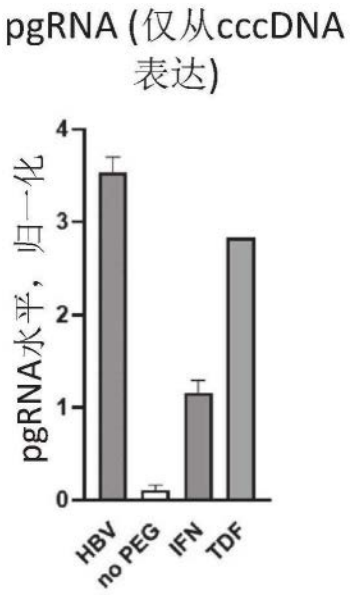

pgRNA (仅从cccDNA
表达)
pgRNA水平，归一化
0
1
2
3
4
HBV
no PEG
IFN
TDF

图13F

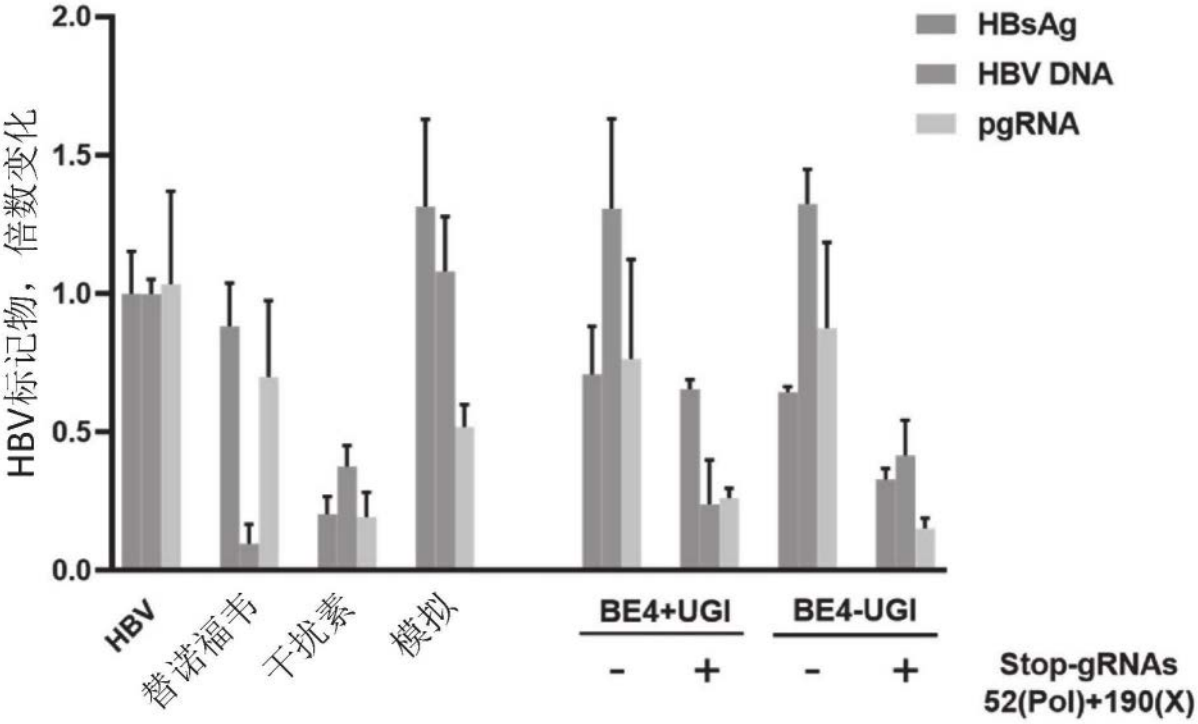

HBsAg
HBV DNA
pgRNA
HBV标记物，倍数变化
2.0
1.5
1.0
0.5
0.0
HBV
替诺福韦
干扰素
模拟
BE4+UGI
BE4-UGI
-
+
-
+
Stop-gRNAs
52(Pol)+190(X)

图14

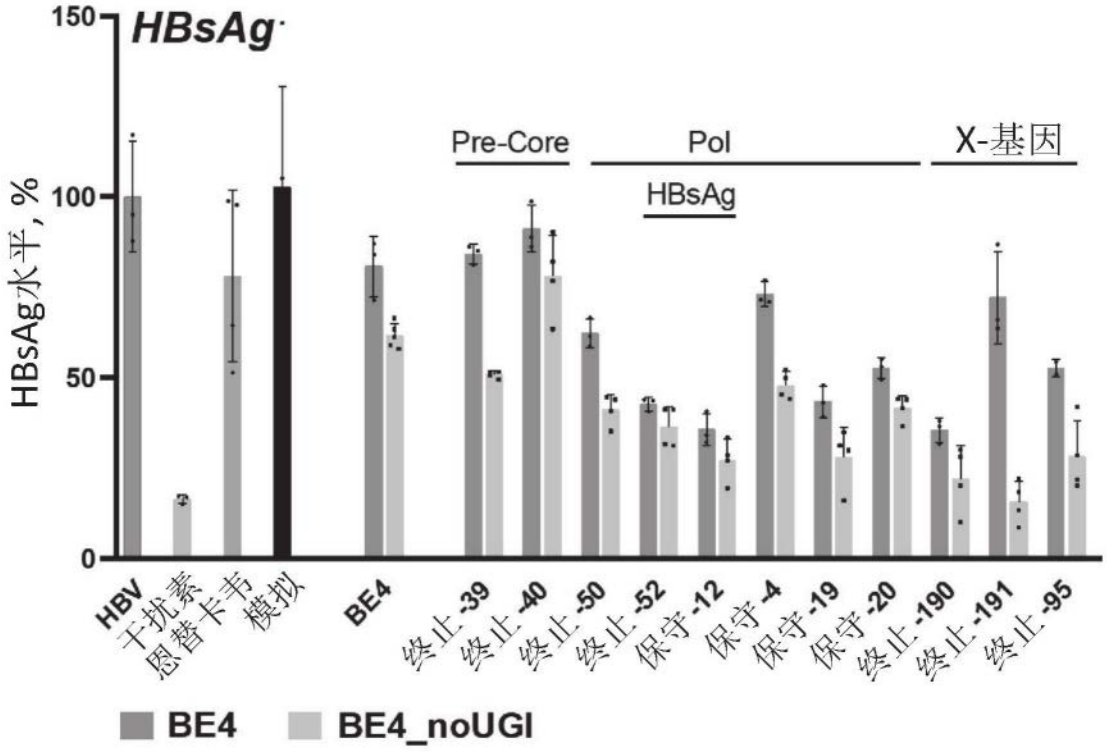

HBsAg
HBsAg水平，%
150
100
50
0
Pre-Core
Pol
X-基因
HBsAg
HBV
干扰素
恩替卡韦
模拟
BE4
终止-39
终止-40
终止-50
终止-52
保守-12
保守-4
保守-19
保守-20
终止-190
终止-191
终止-95
BE4
BE4_noUGI

图15

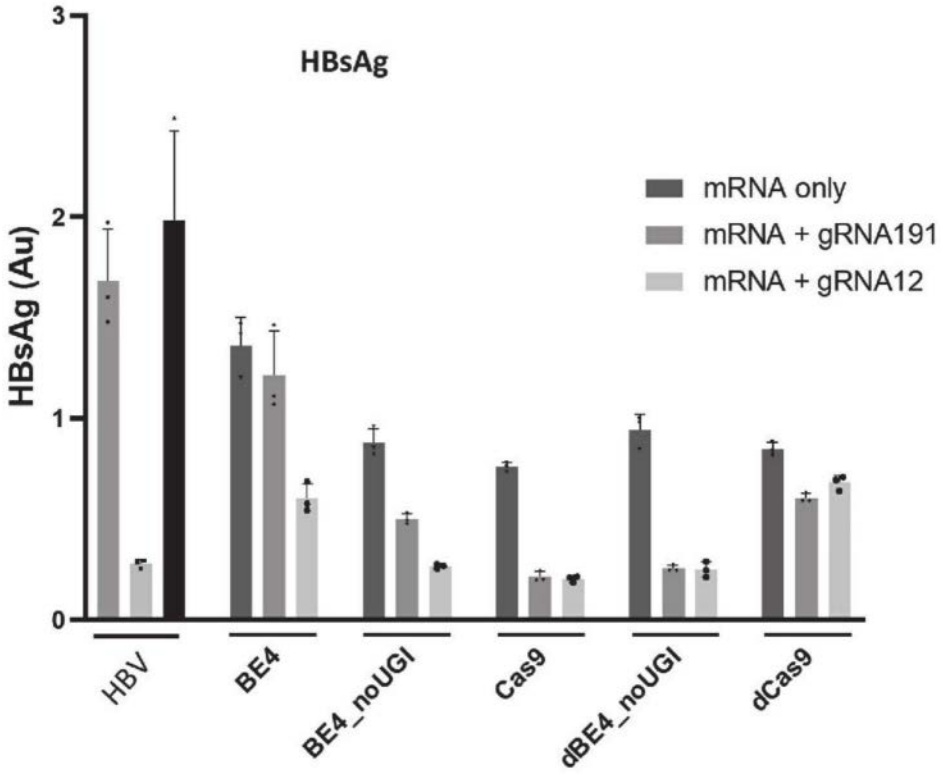

HBsAg
HBsAg (Au)
3
2
1
0
mRNA only
mRNA + gRNA191
mRNA + gRNA12
HBV
BE4
BE4_noUGI
Cas9
dBE4_noUGI
dCas9

图16A

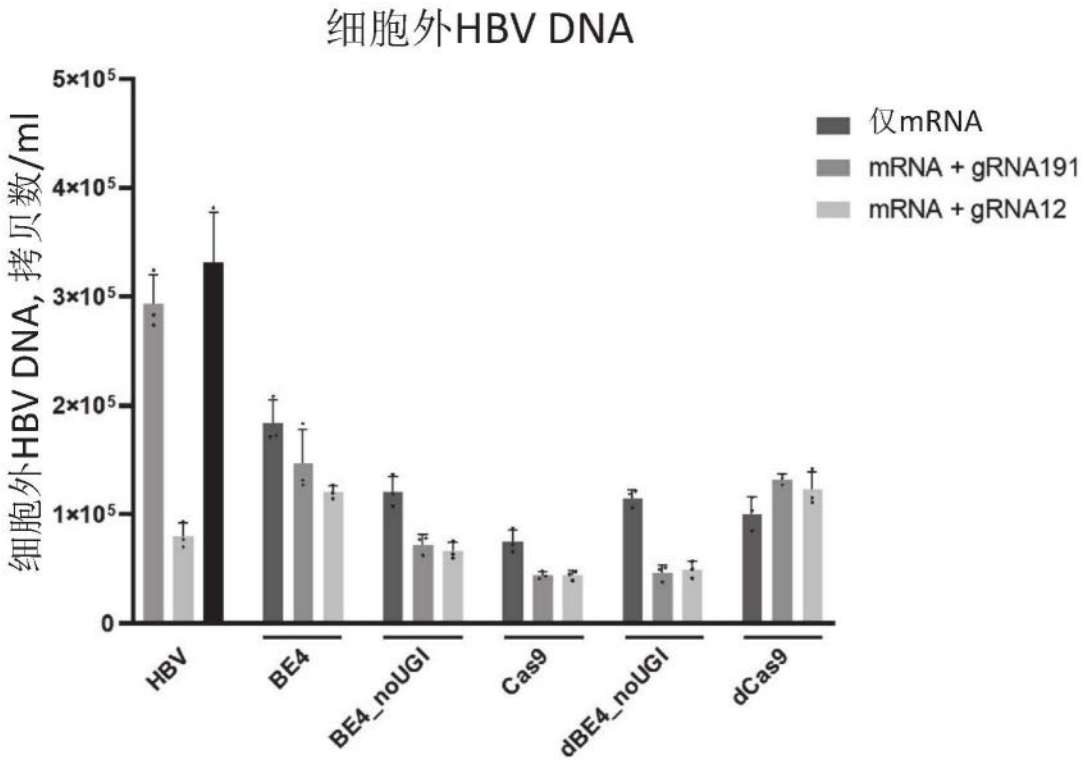

细胞外HBV DNA
细胞外HBV DNA, 拷贝数/ml
5×10⁵
4×10⁵
3×10⁵
2×10⁵
1×10⁵
0
仅mRNA
mRNA + gRNA191
mRNA + gRNA12
HBV
BE4
BE4_noUGI
Cas9
dBE4_noUGI
dCas9

图16B

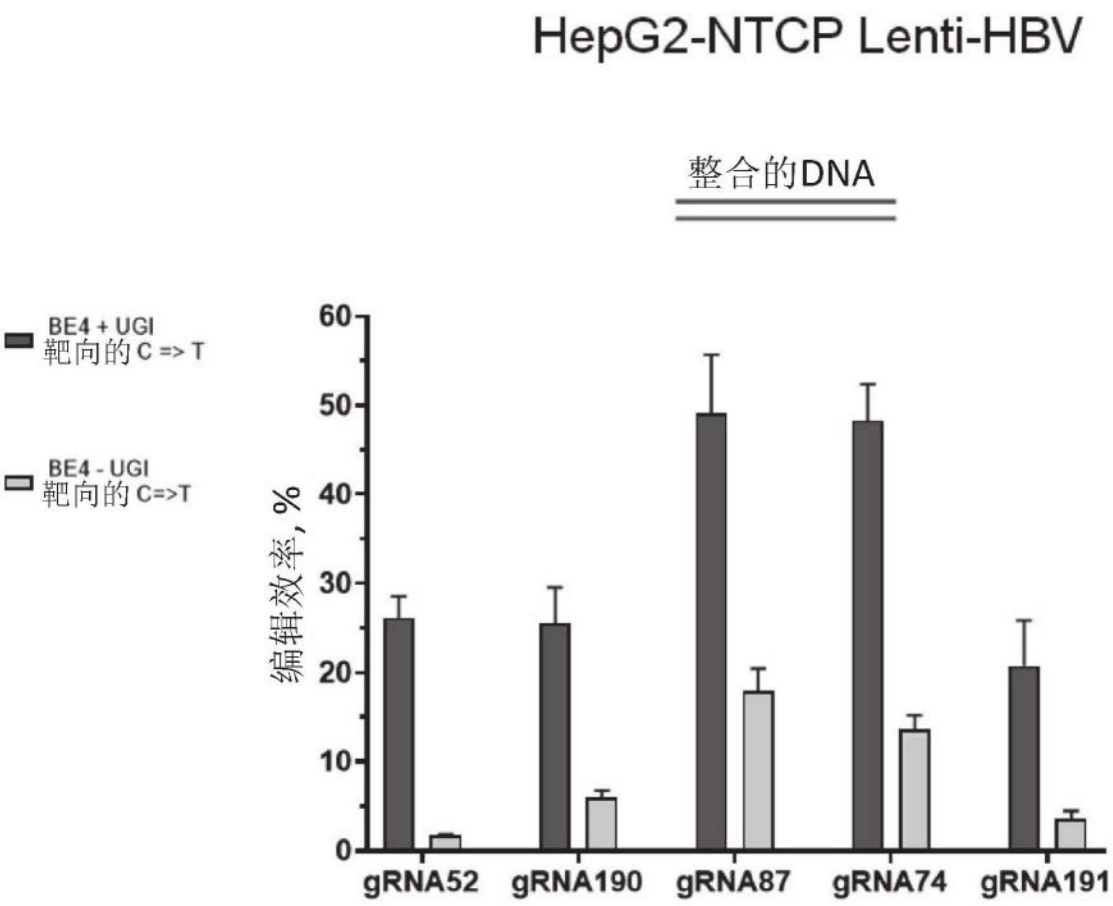

HepG2-NTCP Lenti-HBV
整合的DNA
BE4 + UGI
靶向的C => T
BE4 - UGI
靶向的C=>T
编辑效率，%
60
50
40
30
20
10
0
gRNA52
gRNA190
gRNA87
gRNA74
gRNA191

图17A

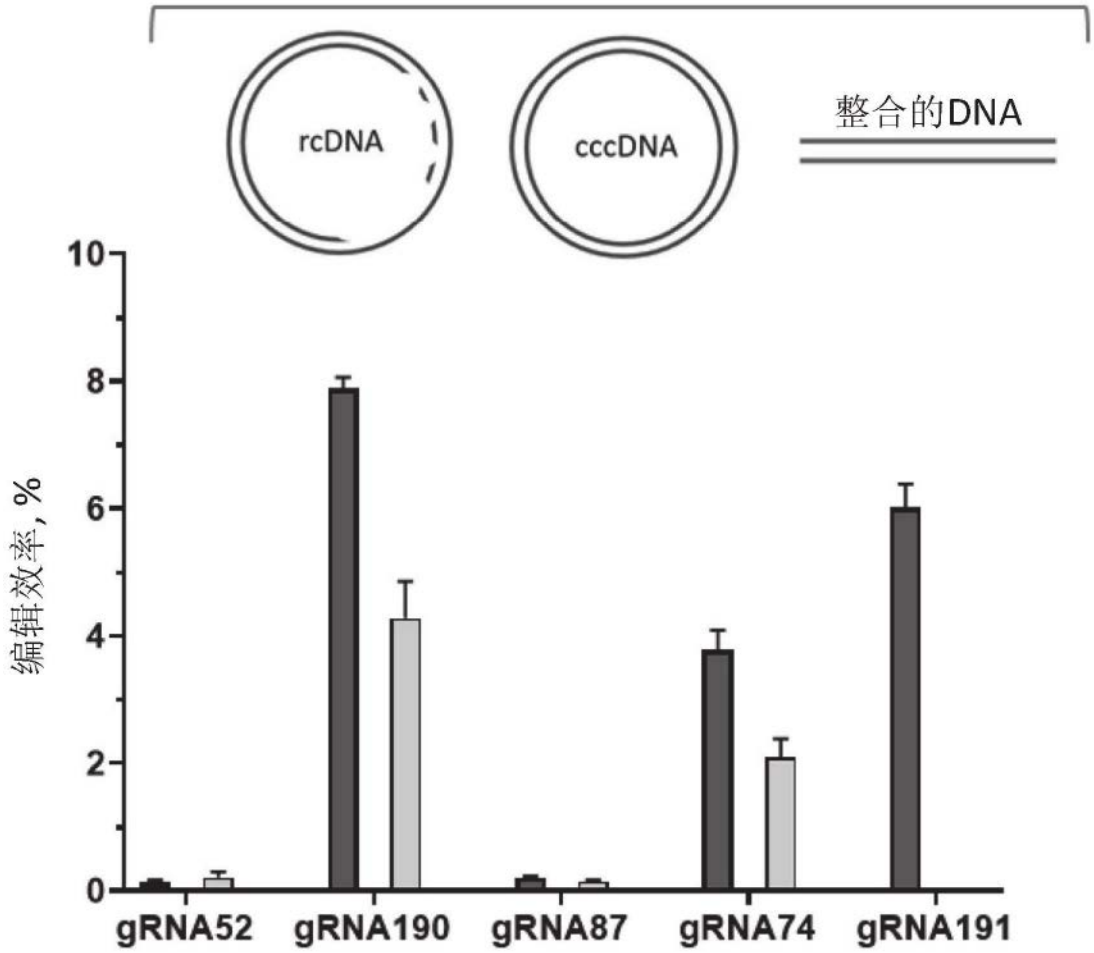

rcDNA
cccDNA
整合的DNA
编辑效率，%
10
8
6
4
2
0
gRNA52
gRNA190
gRNA87
gRNA74
gRNA191

图17B

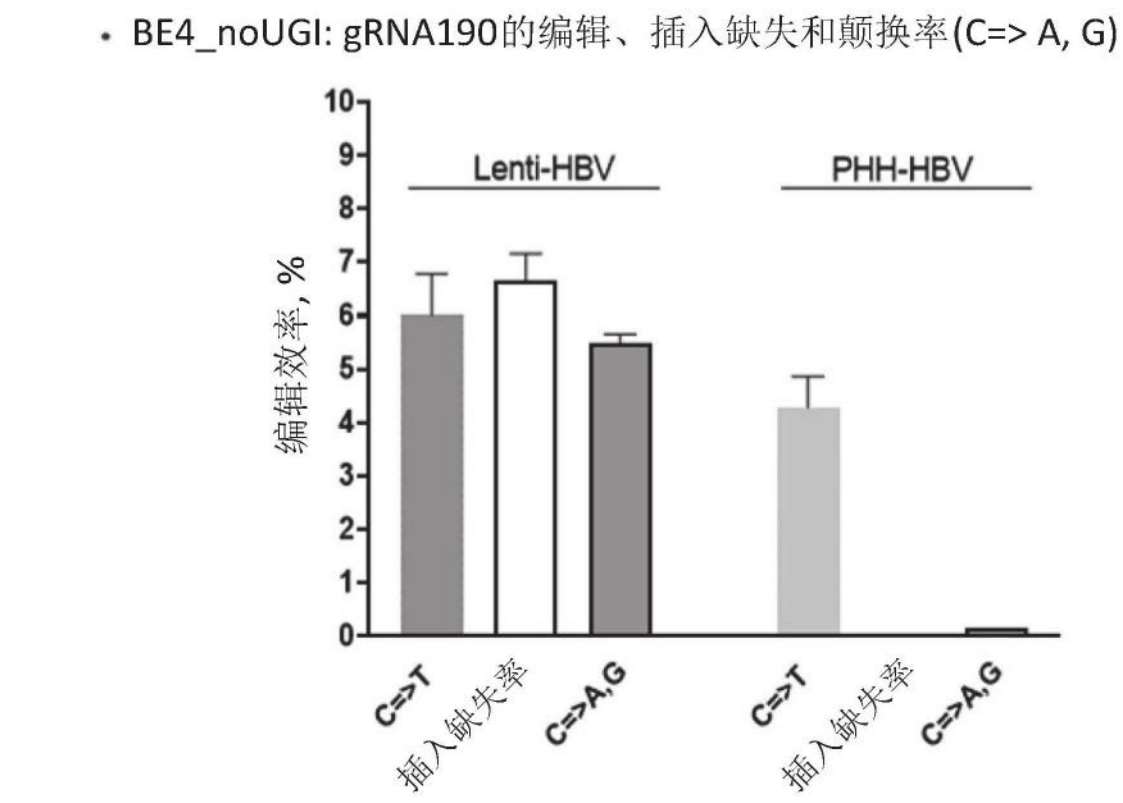

• BE4_noUGI: gRNA190的编辑、插入缺失和颠换率(C=> A, G)
编辑效率，%
10
9
8
7
6
5
4
3
2
1
0
Lenti-HBV
PHH-HBV
C=>T
插入缺失率
C=>A,G
C=>T
插入缺失率
C=>A,G

图18

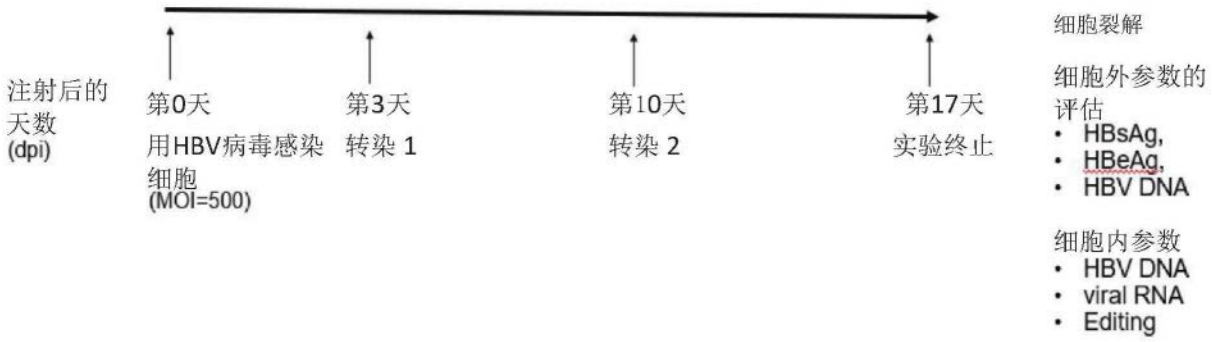

注射后的天数 (dpi)
第0天
用HBV病毒感染细胞 (MOI=500)
第3天
转染 1
第10天
转染 2
第17天
实验终止
细胞裂解
细胞外参数的评估
• HBsAg,
• HBeAg,
• HBV DNA
细胞内参数
• HBV DNA
• viral RNA
• Editing

图19

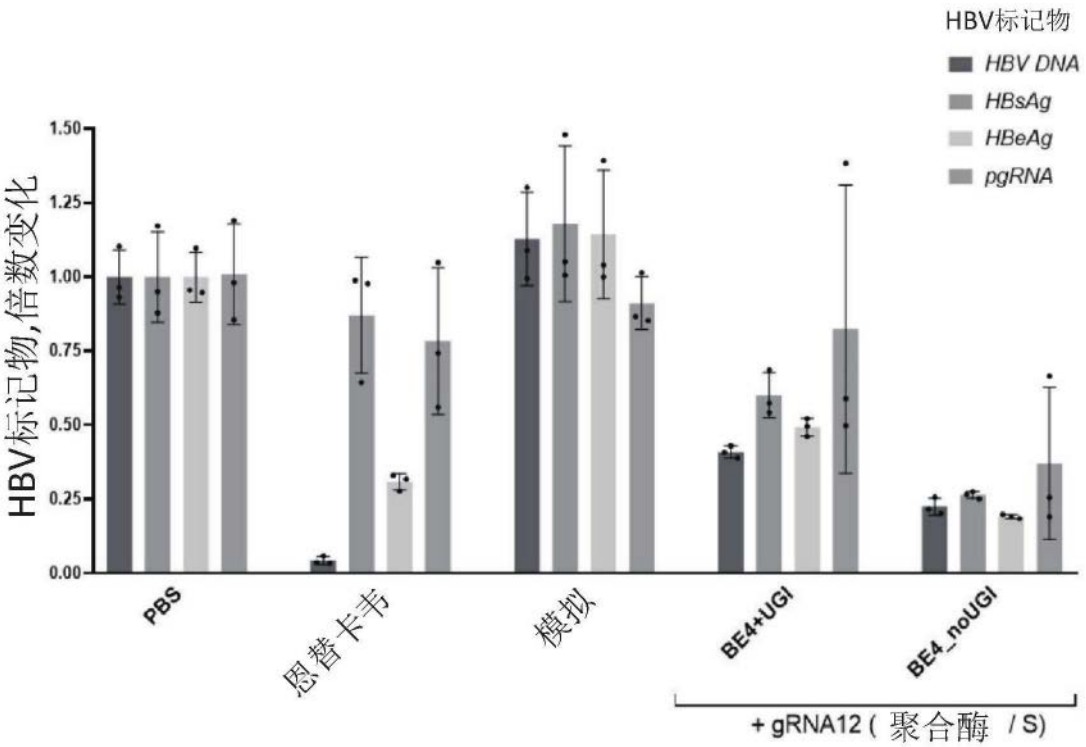

HBV标记物
HBV DNA
HBsAg
HBeAg
pgRNA
HBV标记物,倍数变化
1.50
1.25
1.00
0.75
0.50
0.25
0.00
PBS
恩替卡韦
模拟
BE4+UGI
BE4_noUGI
+ gRNA12 (聚合酶 / S)

图20

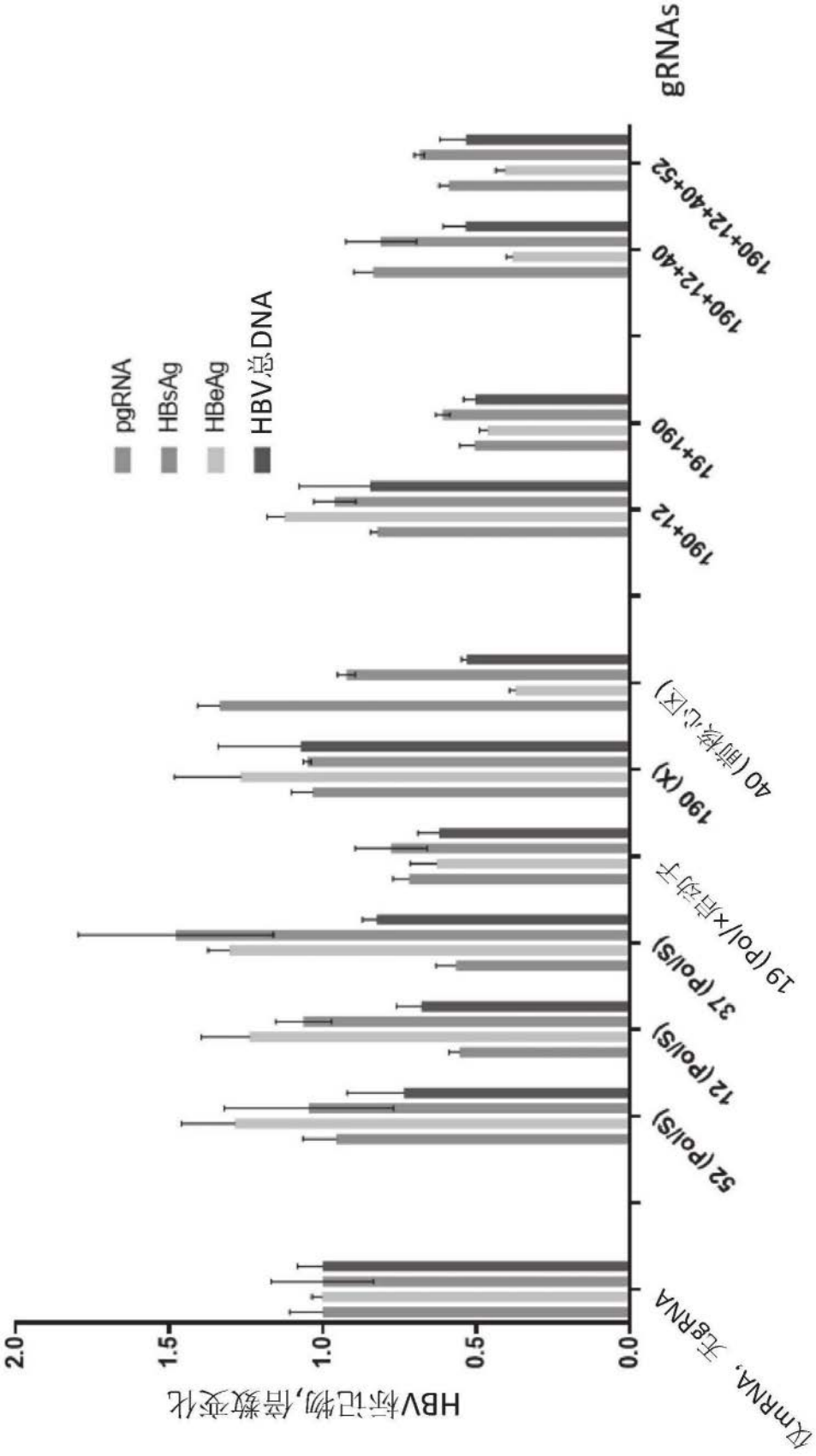

gRNAs
pgRNA
HBsAg
HBeAg
HBV总DNA
190+12+40+52
190+12+40
19+190
190+12
40 (前核心区)
190 (X)
19 (Pol/×后边子
37 (Pol/S)
12 (Pol/S)
52 (Pol/S)
总RNA，无囊RNA
2.0
1.5
1.0
0.5
0.0
HBV标记物，倍数变化

图21

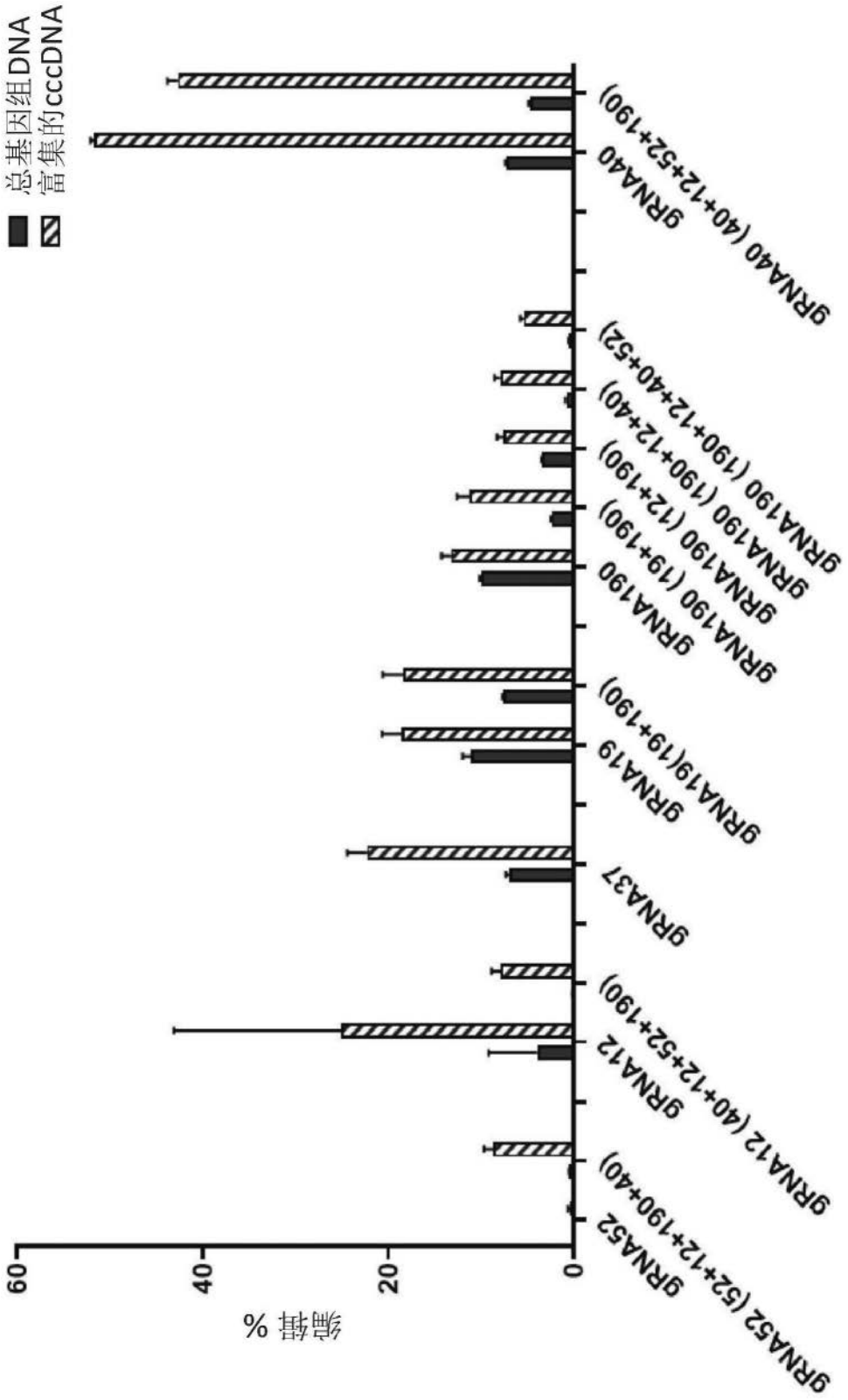

总基因组DNA
富集的cccDNA
编辑 %
0
20
40
60
gRNA52
gRNA52 (52+12+190+40)
gRNA12
gRNA12 (40+12+52+190)
gRNA37
gRNA19
gRNA19(19+190)
gRNA190
gRNA190 (19+190)
gRNA190 (12+190)
gRNA190 (190+12+40)
gRNA40
gRNA40 (40+12+52+190)

图22

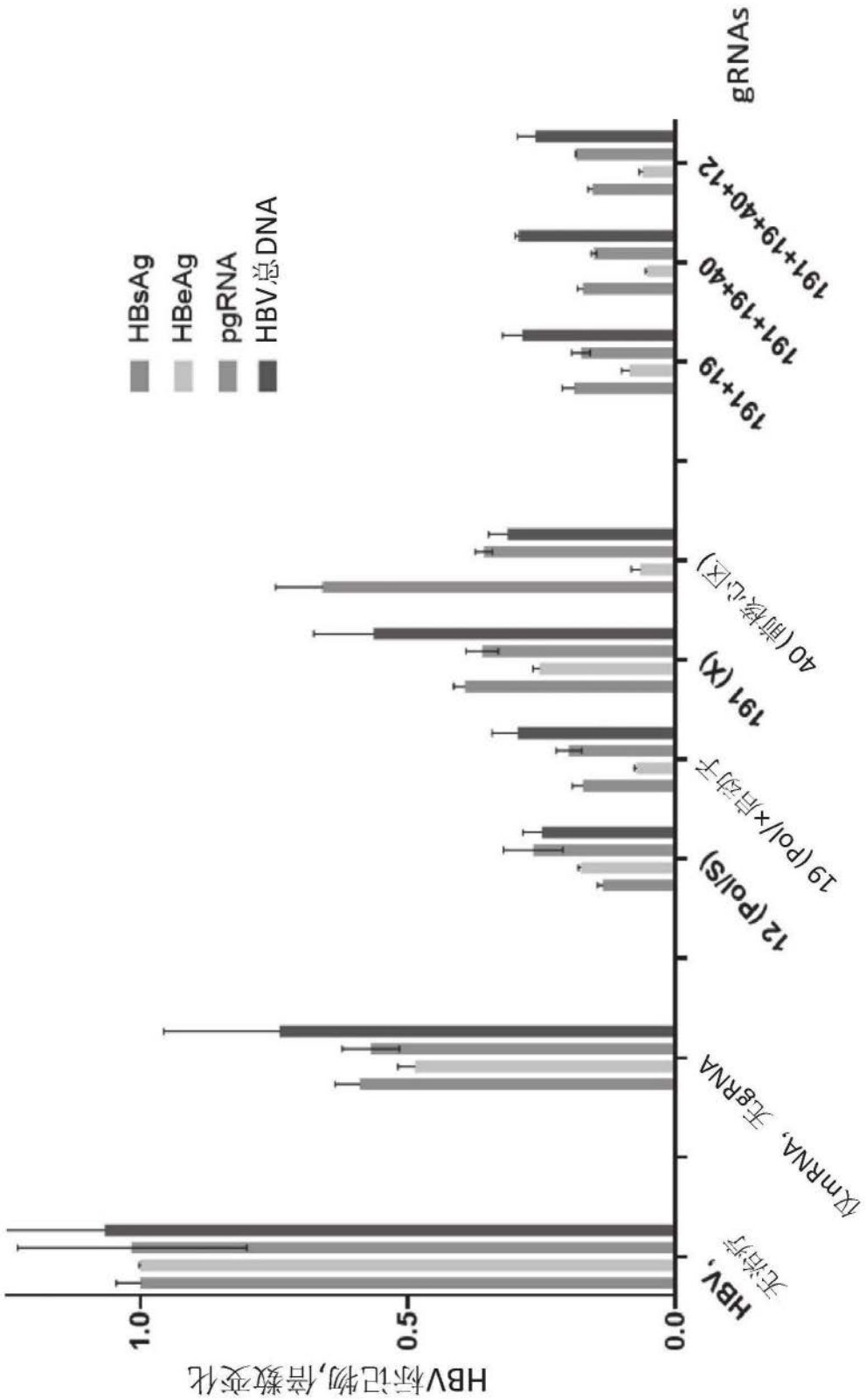

gRNAs
HBsAg
HBeAg
pgRNA
HBV总DNA
HBV标记物,倍数变化
1.0
0.5
0.0
HBV,
无治疗
cccDNA, 无cRNA
12 (Pol/s)
19 (Pol/×启动子
191 (X)
40 (前核心区)
191+19
191+19+40
191+19+40+12

图23

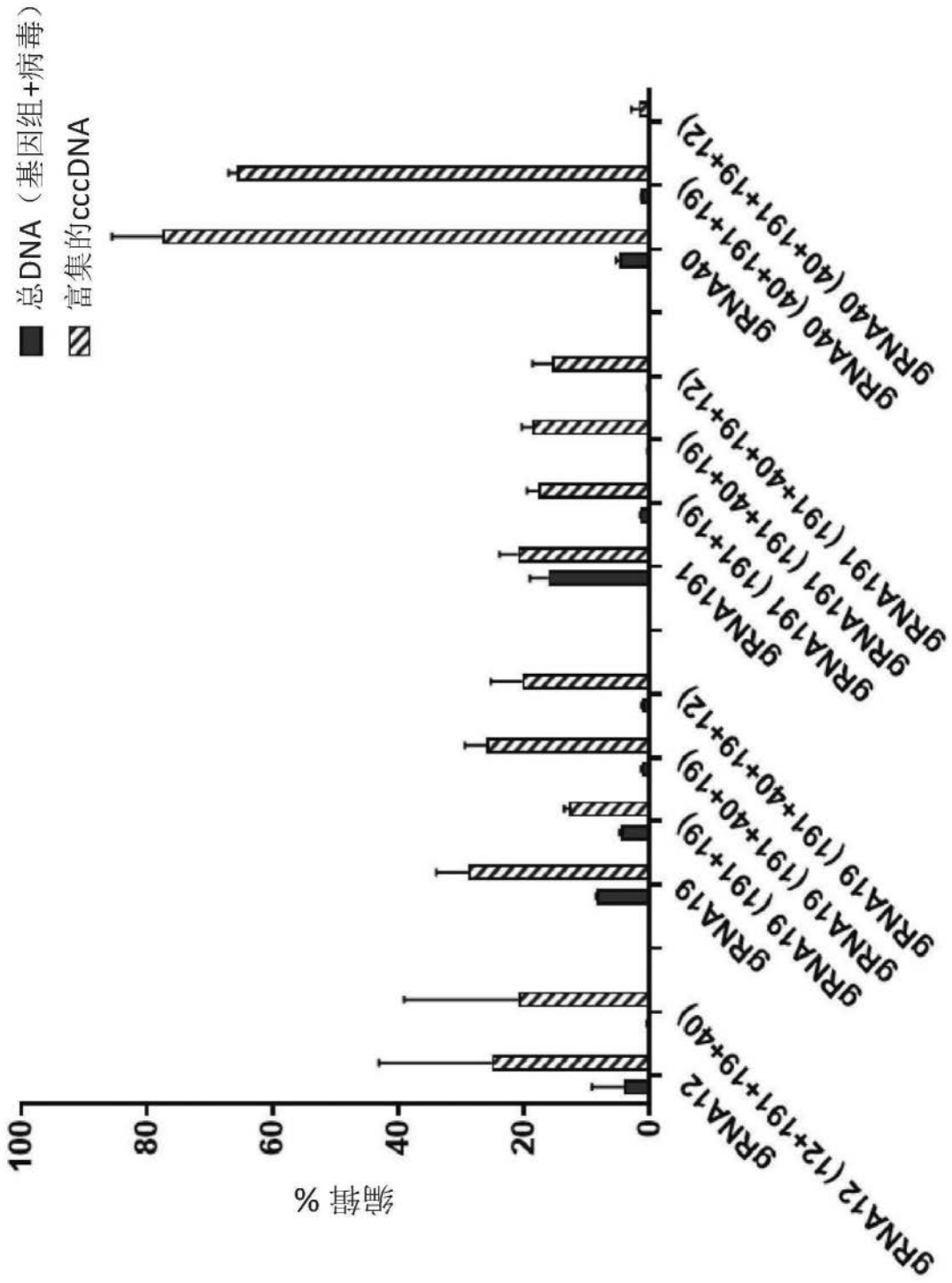

总DNA（基因组+病毒）
富集的cccDNA
编辑 %
100
80
60
40
20
0
gRNA12
gRNA12 (12+191+19+40)
gRNA19
gRNA19 (191+19)
gRNA19 (191+40+19)
gRNA19 (191+40+19+12)
gRNA191
gRNA191 (191+19)
gRNA191 (191+40+19)
gRNA191 (191+40+19+12)
gRNA40
gRNA40 (40+191+19)
gRNA40 (40+191+19+12)

图24

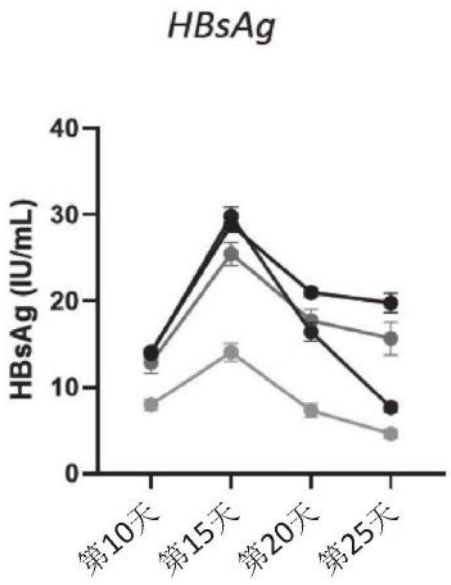

HBsAg
HBsAg (IU/mL)
40
30
20
10
0
第10天
第15天
第20天
第25天

图25A

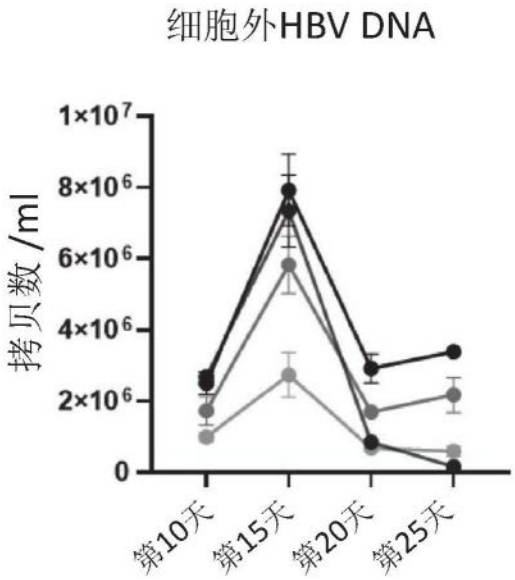

细胞外HBV DNA
拷贝数 /ml
1×10^7
8×10^6
6×10^6
4×10^6
2×10^6
0
第10天
第15天
第20天
第25天

图25B

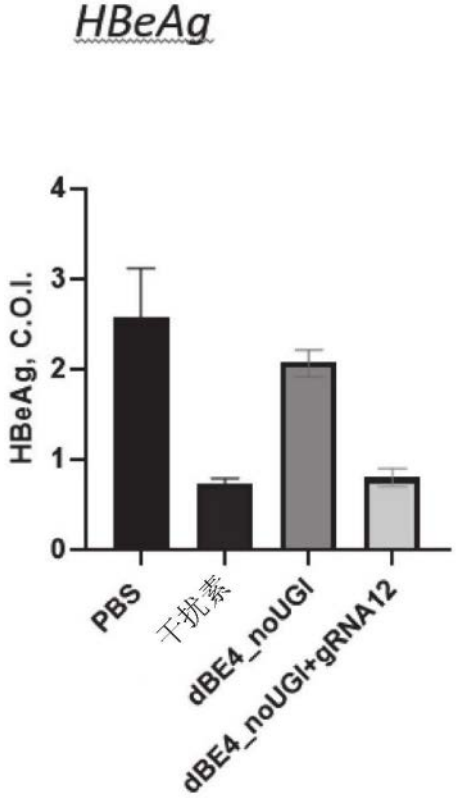


图25C

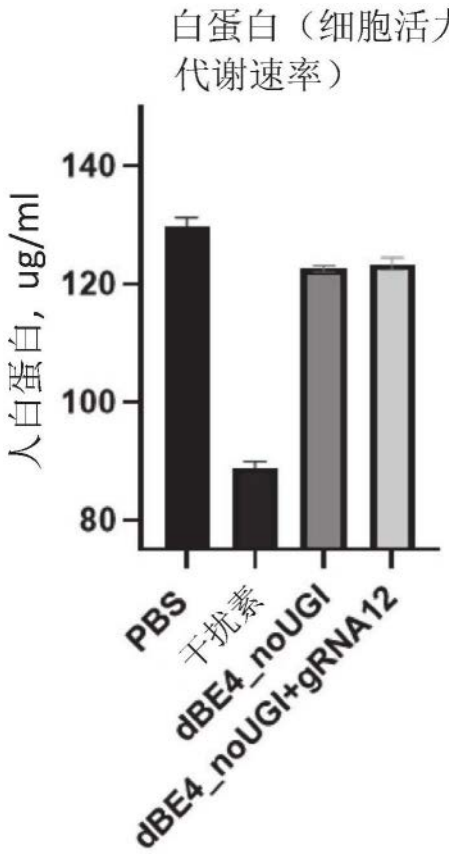

白蛋白（细胞活力/
代谢速率）
人白蛋白，ug/ml
140
120
100
80
PBS
干扰素
dBE4_noUGI
dBE4_noUGI+gRNA12

图25D

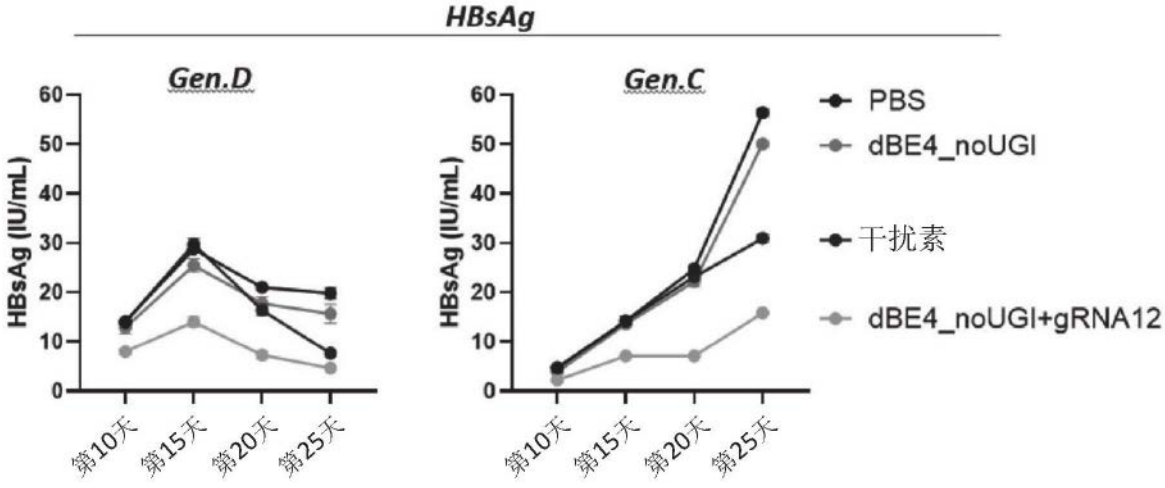

HBsAg
Gen.D
Gen.C
HBsAg (IU/mL)
60
50
40
30
20
10
0
第10天
第15天
第20天
第25天
PBS
dBE4_noUGI
干扰素
dBE4_noUGI+gRNA12

图26A

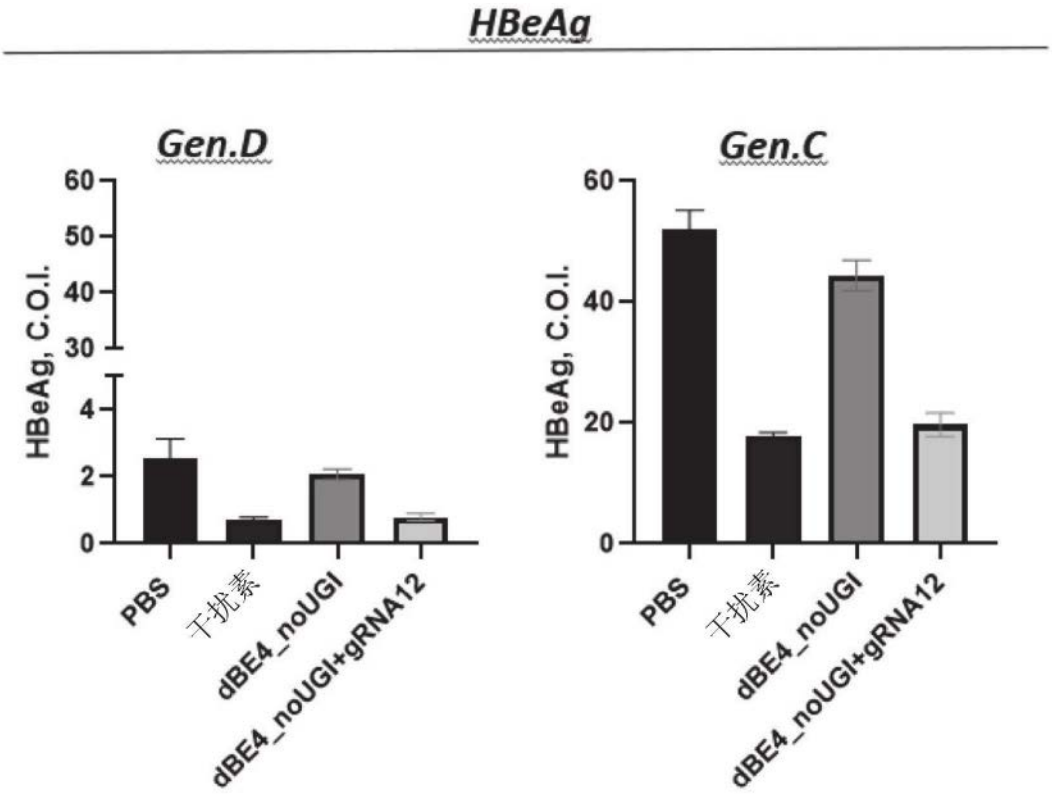

HBeAg
Gen.D
Gen.C
HBeAg, C.O.I.
60
50
40
30
4
2
0
20
PBS
干扰素
dBE4_noUGI
dBE4_noUGI+gRNA12

图26B

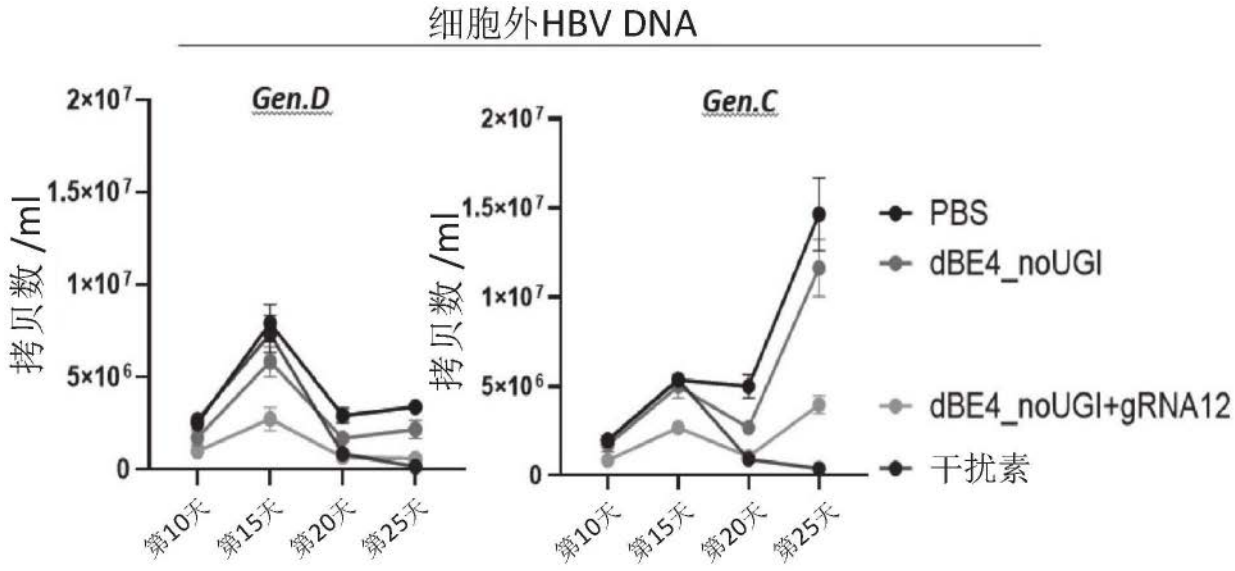

细胞外HBV DNA
Gen.D
Gen.C
拷贝数 /ml
2×10^7
1.5×10^7
1×10^7
5×10^6
0
第10天
第15天
第20天
第25天
PBS
dBE4_noUGI
dBE4_noUGI+gRNA12
干扰素

图26C

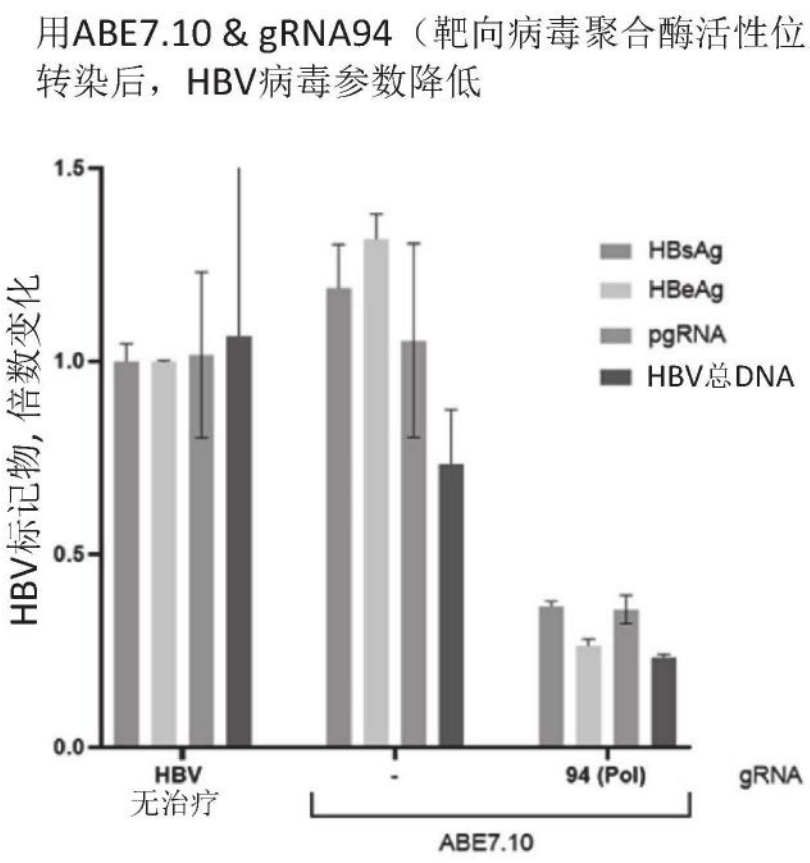

用ABE7.10 & gRNA94（靶向病毒聚合酶活性位点）
转染后，HBV病毒参数降低
HBV标记物，倍数变化
1.5
1.0
0.5
0.0
HBsAg
HBeAg
pgRNA
HBV总DNA
HBV
无治疗
-
94 (Pol)
gRNA
ABE7.10

图27A

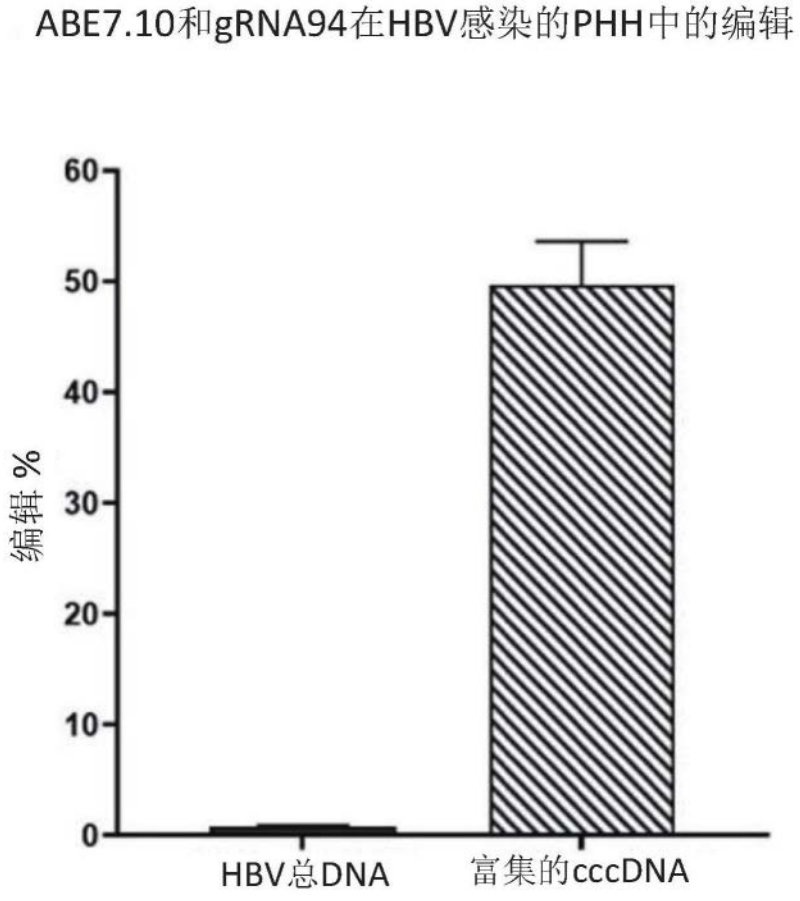

ABE7.10和gRNA94在HBV感染的PHH中的编辑
编辑 %
60
50
40
30
20
10
0
HBV总DNA
富集的cccDNA

图27B