(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 114026237 B
(45)授权公告日 2025.03.25

(21)申请号 202080028267.5
(22)申请日 2020.02.13
(65)同一申请的已公布的文献号
申请公布号 CN 114026237 A
(43)申请公布日 2022.02.08
(30)优先权数据
62/805,271 2019.02.13 US
62/852,228 2019.05.23 US
62/852,224 2019.05.23 US
62/876,354 2019.07.19 US
62/912,992 2019.10.09 US
62/931,722 2019.11.06 US
62/941,569 2019.11.27 US
62/966,526 2020.01.27 US
(85)PCT国际申请进入国家阶段日
2021.10.12
(86)PCT国际申请的申请数据
PCT/US2020/018124 2020.02.13
(87)PCT国际申请的公布数据
WO2020/168088 EN 2020.08.20
(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 N·戈代尔利 M·帕克
I·斯雷梅克 Y·于 B·蔡澈
Y·阿拉泰恩 F·葛瑞格尔
G·伦格 D·A·玻恩 S-J·李
(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.
C12N 15/55(2006.01)
C12N 15/113(2010.01)
C12N 15/10(2006.01)
C12N 9/78(2006.01)
C12N 9/22(2006.01)
C12N 5/10(2006.01)
A61K 48/00(2006.01)
A61K 38/50(2006.01)
A61K 38/46(2006.01)
A61P 3/08(2006.01)
(56)对比文件
CN 109328231 A,2019.02.12
WO 2018027078 A1,2018.02.08
Ribeiro LF等.Protein Engineering Strategies to Expand CRISPR-Cas9 Applications.《Int J Genomics》.2018,第2018卷doi: 10.1155/2018/1652567,特别是第7-8页.
Ribeiro LF等.Protein Engineering Strategies to Expand CRISPR-Cas9 Applications.《Int J Genomics》.2018,第2018卷同上.
审查员 廖日滔

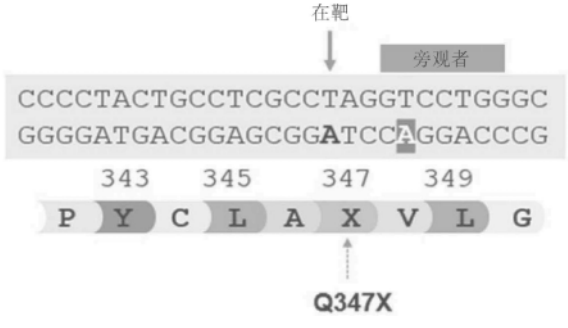
权利要求书12页 说明书267页
序列表（电子公布） 附图23页

(54)发明名称

用于治疗1a型糖原贮积病的组成物和方法

(57)摘要

本发明提供了包含具有提高效率的新型腺苷碱基编辑器(例如ABE8)的组成物和使用包含腺苷脱氨酶变体的碱基编辑器来改变与1a型糖原贮积病(GSD1a)相关的突变的方法。



CN 114026237 B

1.一种在体外编辑包含与1a型糖原贮积病(GSD1a)相关的单核苷酸多态性(SNP)的葡萄糖-6-磷酸酶(G6PC)多核苷酸的方法,所述方法包含将所述G6PC多核苷酸与碱基编辑器在具有一种或多种向导RNA的复合物中接触,其中,一种或多种所述向导RNA引导所述碱基编辑器以实现与GSD1a相关的所述SNP的A·T到G·C改变,其中,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结构域,其中,所述腺苷脱氨酶结构域由以下氨基酸序列MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEI MALRQGGLV MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDV LHYPGMNHRVEITEGILADECAAL LCYFFRMPRQVFNAQKKAQSSTD(SEQ ID NO:3)或其仅缺少N端甲硫氨酸的片段所组成,所述腺苷脱氨酶结构域具有一或多个改变,其中:

a)所述一或多个改变由以下所组成:

选自参考SEQ ID NO:3的V82S、Y147T、Y147R、Q154R和Q154S的改变,或选自以下参考SEQ ID NO:3的改变的组合:

V82S和Y147R,

V82S和Q154R,

V82S和Q154S,

V82S和Y123H,以及

V82S、Y123H和Y147T;与GSD1a相关的所述SNP导致具有Q347X突变的G6PC多肽的表达,其中X代表终止密码子;以及一种或多种所述向导RNA包含由核酸序列GACCUAGGCGAGGCAGUAGG组成的间隔区;或

b)所述一或多个改变由以下所组成:

选自参考SEQ ID NO:3的V82S、Y147T、Y147R和Q154S的改变,或选自以下参考SEQ ID NO:3的改变的组合:

Y147T和Q154S,

Y147R和Q154S,

V82G、Y147T和Q154S,以及

R20A、K21A、Y147T、和Q154S;

与GSD1a相关的所述SNP导致具有R83C突变的G6PC多肽的表达;以及一种或多种所述向导RNA包含由核酸序列CAGUAUGGACACUGUCCAAA或AGUAUGGACACUGUCCAAA组成的间隔区。

2.根据权利要求1所述的方法,其中,所述接触是在细胞中。

3.根据权利要求1所述的方法,其中,所述多核苷酸可编程DNA结合结构域是化脓性链球菌Cas9(SpCas9)。

4.根据权利要求3所述的方法,其中,所述多核苷酸可编程DNA结合结构域是对非G原型间隔区相邻基序(PAM)具有特异性的SpCas9。

5.根据权利要求4所述的方法,其中,所述SpCas9对核酸序列5'-NGA-3'具有特异性,其中,N是A、T、C或G。

6.根据权利要求4或5所述的方法,其中,所述SpCas9对由核酸序列5'-AGA-3'或5'-GGA-3'组成的原型间隔区相邻基序(PAM)具有特异性。

7.根据权利要求4所述的方法,其中,所述SpCas9对核酸序列NGA具有特异性,其中,N是A、T、C或G。

8.根据权利要求1所述的方法,其中,所述多核苷酸可编程DNA结合结构域是金黄色葡萄球菌Cas9(SaCas9)。

9.根据权利要求8所述的方法,其中,所述SaCas9对核酸序列5'-NNGRRT-3'具有原型间隔区相邻基序(PAM)的特异性,其中,N是A、T、C或G,且R是A或G。

10.根据权利要求9所述的方法,其中,所述SaCas9对由核酸序列5'-GAGAAT-3'组成的原型间隔区相邻基序(PAM)具有特异性。

11.根据权利要求8所述的方法,其中,所述SaCas9对由核苷酸序列NNGRRT组成的原型间隔区相邻基序(PAM)具有特异性,其中,N是A、T、C或G,且R是A或G。

12.根据权利要求1所述的方法,其中,所述多核苷酸可编程DNA结合结构域是核酸酶无活性的。

13.根据权利要求12所述的方法,其中,所述多核苷酸可编程DNA结合结构域是切口酶。

14.根据权利要求13所述的方法,其中,所述切口酶包含以下氨基酸序列:

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYT
RRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKA
DLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENL
IAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAIL
LSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVG
PLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY
VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKD
FLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILD
FLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPE
NIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL
SDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLS
ELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAH
DAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRP
LIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSP
TVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRM
LASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL
GGD。

15.根据权利要求1所述的方法,其中,所述腺苷脱氨酶结构域能够使脱氧核糖核酸(DNA)中的腺苷脱氨基。

16.根据权利要求1所述的方法,其中,所述碱基编辑器包含单个腺苷脱氨酶结构域。

17.根据权利要求1所述的方法,其中,所述碱基编辑器进一步包含第二腺苷脱氨酶结构域,其中,所述第二腺苷脱氨酶结构域是野生型腺苷脱氨酶结构域。

18.根据权利要求1所述的方法,其中,所述腺苷脱氨酶结构域由以下序列组成:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEI MALRQGGLVMQ
NYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDV LHYPGMNHRVEITEGILADECAALLC

TFFRMPRQVFNAQKKAQSSTD。

19.一种细胞,包含:

碱基编辑器或编码所述碱基编辑器的多核苷酸,其中,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结构域,其中,所述腺苷脱氨酶结构域由以下氨基酸序列

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTA HAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGS LMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD(SEQ ID NO:3)或其仅缺少N端甲硫氨酸的片段所组成,所述腺苷脱氨酶结构域具有一或多个改变,以及一或多种引导所述碱基编辑器以实现与GSD1a相关的SNP的A·T到G·C改变的向导RNA,其中:a)所述一或多个改变由以下所组成:

选自参考SEQ ID NO:3的V82S、Y147T、Y147R、Q154R和Q154S的改变,或选自以下参考SEQ ID NO:3的改变的组合:

V82S和Y147R,

V82S和Q154R,

V82S和Q154S,

V82S和Y123H,以及

V82S、Y123H和Y147T;与GSD1a相关的所述SNP导致具有Q347X突变的葡萄糖-6-磷酸酶(G6PC)多肽的表达,其中X代表终止密码子;以及一种或多种所述向导RNA包含由核酸序列GACCUAGGCGAGGCAGUAGG组成的间隔区;或

b)所述一或多个改变由以下所组成:

选自参考SEQ ID NO:3的V82S、Y147T、Y147R和Q154S的改变,或选自以下参考SEQ ID NO:3的改变的组合:

Y147T和Q154S,

Y147R和Q154S,

V82G、Y147T和Q154S,以及

R20A、K21A、Y147T、和Q154S;

与GSD1a相关的所述SNP导致具有R83C突变的G6PC多肽的表达;以及一种或多种所述向导RNA包含由核酸序列CAGUAUGGACACUGUCCAAA或AGUAUGGACACUGUCCAAA组成的间隔区。

20.根据权利要求19所述的细胞,其中,所述细胞是肝细胞或肝细胞前体。

21.根据权利要求19所述的细胞,其中,所述细胞表达G6PC多肽。

22.根据权利要求19所述的细胞,其中,所述细胞来自患有GSD1a的个体。

23.根据权利要求19所述的细胞,其中,所述细胞是哺乳动物细胞。

24.根据权利要求19所述的细胞,其中,所述细胞是人类细胞。

25.根据权利要求19所述的细胞,其中,与GSD1a相关的所述SNP的A·T至G·C改变导致在所述G6PC的氨基酸序列第347位处的终止密码替换为谷氨酰胺或将所述G6PC的氨基酸序列第83位处的半胱氨酸替换为精氨酸。

26.根据权利要求19所述的细胞,其中,所述多核苷酸可编程DNA结合结构域是化脓性链球菌Cas9(SpCas9)。

27.根据权利要求26所述的细胞,其中,所述多核苷酸可编程DNA结合结构域是对非G原

型间隔区相邻基序(PAM)具有特异性的SpCas9。

28.根据权利要求27所述的细胞,其中,所述SpCas9对由核酸序列5'-NGA-3'组成的原型间隔区相邻基序(PAM)具有特异性,其中,N是A、T、C或G。

29.根据权利要求27或28所述的细胞,其中,所述SpCas9对由核酸序列5'-AGA-3'或5'-GGA-3'组成的原型间隔区相邻基序(PAM)具有特异性。

30.根据权利要求29所述的细胞,其中,所述SpCas9对由核酸序列NGA组成的原型间隔区相邻基序(PAM)具有特异性,其中,N是A、T、C或G。

31.根据权利要求19所述的细胞,其中,所述多核苷酸可编程DNA结合结构域是金黄色葡萄球菌Cas9(SaCas9)。

32.根据权利要求31所述的细胞,其中,所述SaCas9对由核酸序列5'-NNGRRT-3'组成的原型间隔区相邻基序(PAM)具有特异性,其中,N是A、T、C或G,且R是A或G。

33.根据权利要求32所述的细胞,其中,所述SaCas9对由核酸序列5'-GAGAAT-3'组成的原型间隔区相邻基序(PAM)具有特异性。

34.根据权利要求31所述的细胞,其中,所述SaCas9对由核酸序列NNGRRT组成的原型间隔区相邻基序(PAM)具有特异性,其中,N是A、T、C或G,且R是A或G。

35.根据权利要求19所述的细胞,其中,所述多核苷酸可编程DNA结合结构域是核酸酶无活性的。

36.根据权利要求19所述的细胞,其中,所述多核苷酸可编程DNA结合结构域是切口酶。

37.根据权利要求36所述的细胞,其中,所述切口酶包含以下氨基酸序列:

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD。

38.根据权利要求19所述的细胞,其中,所述腺苷脱氨酶结构域能够使脱氧核糖核酸

(DNA)中的腺苷脱氨基。

39.根据权利要求19所述的细胞,其中,所述碱基编辑器包含单个腺苷脱氨酶结构域。

40.根据权利要求19所述的细胞,其中,所述碱基编辑器进一步包含第二腺苷脱氨酶结构域,其中,所述第二腺苷脱氨酶结构域是野生型腺苷脱氨酶结构域。

41.根据权利要求19所述的细胞,其中,所述腺苷脱氨酶结构域由以下序列组成:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEI MALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDV LHYPGMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD。

42.根据权利要求19所述的细胞,其中,所述一种或多种向导RNA的支架由以下序列所组成:

GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU。

43.根据权利要求19所述的细胞,其中,所述一种或多种向导RNA的支架由以下序列所组成:

GUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCUACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUU。

44.一种碱基编辑器或编码所述碱基编辑器的多核苷酸在制造用于治疗个体的1a型糖原贮积病(GSD1a)的药物中的用途,其中,所述治疗包含向所述个体施用:

所述碱基编辑器或编码所述碱基编辑器的多核苷酸,其中,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结构域,其中,所述腺苷脱氨酶结构域由以下氨基酸序列MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEI MALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDV LHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD(SEQ ID NO:3)或其仅缺少N端甲硫氨酸的片段所组成,所述腺苷脱氨酶结构域具有一或多个改变,以及一或多种引导所述碱基编辑器以实现与GSD1a相关的SNP的A•T到G•C改变的向导RNA,其中:

a)所述一或多个改变由以下所组成:

选自参考SEQ ID NO:3的V82S、Y147T、Y147R、Q154R和Q154S的改变,或选自以下参考SEQ ID NO:3的改变的组合:

V82S和Y147R,

V82S和Q154R,

V82S和Q154S,

V82S和Y123H,以及

V82S、Y123H和Y147T;与GSD1a相关的所述SNP导致具有Q347X突变的葡萄糖-6-磷酸酶(G6PC)多肽的表达,其中X代表终止密码子;以及一种或多种所述向导RNA包含由核酸序列GACCUAGGCGAGGCAGUAGG组成的间隔区;或

b)所述一或多个改变由以下所组成:

选自参考SEQ ID NO:3的V82S、Y147T、Y147R和Q154S的改变,或选自以下参考SEQ ID NO:3的改变的组合:

Y147T和Q154S,

Y147R和Q154S,

V82G、Y147T和Q154S,以及

R20A、K21A、Y147T、和Q154S;

与GSD1a相关的所述SNP导致具有R83C突变的G6PC多肽的表达;以及一种或多种所述向导RNA包含由核酸序列CAGUAUGGACACUGUCCAAA或AGUAUGGACACUGUCCAAA组成的间隔区。

45.根据权利要求44所述的用途,其中,所述个体是哺乳动物。

46.根据权利要求44所述的用途,其中,所述GSD1a的治疗包含将所述碱基编辑器或编码所述碱基编辑器的多核苷酸和所述一种或多种向导RNA递送至所述个体的细胞。

47.根据权利要求46所述的用途,其中,所述细胞是肝细胞或肝细胞前体。

48.根据权利要求44所述的用途,其中,所述多核苷酸可编程DNA结合结构域是化脓性链球菌Cas9(SpCas9)。

49.根据权利要求48所述的用途,其中,所述多核苷酸可编程DNA结合结构域是对非G原型间隔区相邻基序(PAM)具有特异性的SpCas9。

50.根据权利要求49所述的用途,其中,所述SpCas9对由核酸序列5'-NGA-3'组成的原型间隔区相邻基序(PAM)具有特异性,其中,N是A、T、C或G。

51.根据权利要求49所述的用途,其中,所述SpCas9对由核酸序列5'-AGA-3'或5'-GGA-3'组成的原型间隔区相邻基序(PAM)具有特异性。

52.根据权利要求49所述的用途,其中,所述SpCas9对由核酸序列NGA组成的原型间隔区相邻基序(PAM)具有特异性,其中,N是A、T、C或G。

53.根据权利要求44所述的用途,其中,所述多核苷酸可编程DNA结合结构域是金黄色葡萄球菌Cas9(SaCas9)。

54.根据权利要求53所述的用途,其中,所述SaCas9对由核酸序列5'-NNGRRT-3'组成的原型间隔区相邻基序(PAM)具有特异性,其中,N是A、T、C或G,且R是A或G。

55.根据权利要求53或54所述的用途,其中,所述SaCas9对由核酸序列5'-GAGAAT-3'组成的原型间隔区相邻基序(PAM)具有特异性。

56.根据权利要求53所述的用途,其中,所述SaCas9对由核酸序列NNGRRT组成的原型间隔区相邻基序(PAM)具有特异性,其中,N是A、T、C或G,且R是A或G。

57.根据权利要求44所述的用途,其中,所述多核苷酸可编程DNA结合结构域是核酸酶无活性的。

58.根据权利要求44所述的用途,其中,所述多核苷酸可编程DNA结合结构域是切口酶。

59.根据权利要求58所述的用途,其中,所述切口酶包含以下氨基酸序列:

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKD

FLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILD FLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPE NIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL SDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLS ELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAH DAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRP LIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSP TVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRM LASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL GGD。

60.根据权利要求44所述的用途,其中,所述腺苷脱氨酶结构域能够使脱氧核糖核酸(DNA)中的腺苷脱氨基。

61.根据权利要求44所述的用途,其中,所述碱基编辑器包含单个腺苷脱氨酶结构域。

62.根据权利要求44所述的用途,其中,所述碱基编辑器进一步包含第二腺苷脱氨酶结构域,其中,所述第二腺苷脱氨酶结构域是野生型腺苷脱氨酶结构域。

63.一种在体外产生肝细胞的方法,所述方法包含:

(a)引入包含与1a型糖原贮积病(GSD1a)相关的单核苷酸多态性(SNP)的诱导多能干细胞或肝细胞祖细胞,以及

碱基编辑器或编码所述碱基编辑器的多核苷酸,其中,所述碱基编辑器包含多核苷酸可编程核苷酸结合结构域和腺苷脱氨酶结构域,其中,所述腺苷脱氨酶结构域由以下氨基酸序列MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEI MALRQGGLV MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDV LHYPGMNHRVEITEGILADECAAL LCYFFRMPRQVFNAQKKAQSSTD(SEQ ID NO:3)或其仅缺少N端甲硫氨酸的片段所组成,所述腺苷脱氨酶结构域具有一或多个改变,以及一种或多种引导所述碱基编辑器以实现与GSD1a相关的所述SNP的A·T到G·C改变的向导RNA,其中:

i)所述一或多个改变由以下所组成:

选自参考SEQ ID NO:3的V82S、Y147T、Y147R、Q154R和Q154S的改变,或选自以下参考SEQ ID NO:3的改变的组合:

V82S和Y147R,

V82S和Q154R,

V82S和Q154S,

V82S和Y123H,以及

V82S、Y123H和Y147T;与GSD1a相关的所述SNP导致具有Q347X突变的葡萄糖-6-磷酸酶(G6PC)多肽的表达,其中X代表终止密码子;以及一种或多种所述向导RNA包含由核酸序列GACCUAGGCGAGGCAGUAGG组成的间隔区;或

ii)所述一或多个改变由以下所组成:

选自参考SEQ ID NO:3的V82S、Y147T、Y147R和Q154S的改变,或选自以下参考SEQ ID NO:3的改变的组合:

Y147T和Q154S,

Y147R和Q154S,

V82G、Y147T和Q154S,以及

R20A、K21A、Y147T、和Q154S;

与GSD1a相关的所述SNP导致具有R83C突变的G6PC多肽的表达;以及一种或多种所述向导RNA包含由核酸序列CAGUAUGGACACUGUCCAAA或AGUAUGGACACUGUCCAAA组成的间隔区;和(b)将所述诱导多能干细胞或肝细胞祖细胞分化为肝细胞。

64.根据权利要求63所述的方法,其中,所述肝细胞祖细胞获自患有GSD1a的个体。

65.根据权利要求63所述的方法,其中,所述肝细胞是哺乳动物细胞。

66.根据权利要求63所述的方法,其中,步骤(a)的所述诱导多能干细胞表达具有Q347X突变的G6PC多肽,其中X代表终止密码子。

67.根据权利要求63所述的方法,其中,步骤(a)的所述诱导多能干细胞表达具有R83C突变的G6PC多肽。

68.根据权利要求63所述的方法,其中,所述腺苷脱氨酶结构域能够使脱氧核糖核酸(DNA)中的腺苷脱氨基。

69.根据权利要求63所述的方法,其中,所述碱基编辑器包含单个腺苷脱氨酶结构域。

70.根据权利要求63所述的方法,其中,所述碱基编辑器进一步包含第二腺苷脱氨酶结构域,其中,所述第二腺苷脱氨酶结构域是野生型腺苷脱氨酶结构域。

71.一种用于治疗1a型糖原贮积病(GSD1a)的医药组成物,其包含:

有效量的碱基编辑器,其中,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结构域,其中,所述腺苷脱氨酶结构域由以下氨基酸序列

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD(SEQ ID NO:3)或其仅缺少N端甲硫氨酸的片段所组成,所述腺苷脱氨酶结构域具有一或多个改变,以及一或多种引导所述碱基编辑器以实现与GSD1a相关的SNP的A•T到G•C改变的向导RNA,其中:

a)所述一或多个改变由以下所组成:

选自参考SEQ ID NO:3的V82S、Y147T、Y147R、Q154R和Q154S的改变,或选自以下参考SEQ ID NO:3的改变的组合:

V82S和Y147R,

V82S和Q154R,

V82S和Q154S,

V82S和Y123H,以及

V82S、Y123H和Y147T;与GSD1a相关的所述SNP导致具有Q347X突变的葡萄糖-6-磷酸酶(G6PC)多肽的表达,其中X代表终止密码子;以及一种或多种所述向导RNA包含由核酸序列GACCUAGGCGAGGCAGUAGG组成的间隔区;或

b)所述一或多个改变由以下所组成:

选自参考SEQ ID NO:3的V82S、Y147T、Y147R和Q154S的改变,或选自以下参考SEQ ID NO:3的改变的组合:

Y147T和Q154S,

Y147R和Q154S,

V82G、Y147T和Q154S,以及

R20A、K21A、Y147T、和Q154S;

与GSD1a相关的所述SNP导致具有R83C突变的G6PC多肽的表达;以及一种或多种所述向导RNA包含由核酸序列CAGUAUGGACACUGUCCAAA或AGUAUGGACACUGUCCAAA组成的间隔区。

72.一种用于治疗1a型糖原贮积病(GSD1a)的医药组成物,其包含有效量的权利要求19至43中任一项所述的细胞。

73.根据权利要求71或72所述的医药组成物,还包含药学上可接受的赋形剂。

74.一种用于治疗1a型糖原贮积病(GSD1a)的试剂盒,所述试剂盒包含碱基编辑器,其中,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结构域,其中,所述腺苷脱氨酶结构域由以下氨基酸序列

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD(SEQ ID NO:3)或其仅缺少N端甲硫氨酸的片段所组成,所述腺苷脱氨酶结构域具有一或多个改变,以及一种或多种引导所述碱基编辑器以实现与GSD1a相关的SNP的A•T到G•C改变的向导RNA,其中:

a)所述一或多个改变由以下所组成:

选自参考SEQ ID NO:3的V82S、Y147T、Y147R、Q154R和Q154S的改变,或选自以下参考SEQ ID NO:3的改变的组合:

V82S和Y147R,

V82S和Q154R,

V82S和Q154S,

V82S和Y123H,以及

V82S、Y123H和Y147T;其中,与GSD1a相关的所述SNP导致具有Q347X突变的葡萄糖-6-磷酸酶(G6PC)多肽的表达,其中X代表终止密码子;以及一种或多种所述向导RNA包含由核酸序列GACCUAGGCGAGGCAGUAGG组成的间隔区;或

b)所述一或多个改变由以下所组成:

选自参考SEQ ID NO:3的V82S、Y147T、Y147R和Q154S的改变,或选自以下参考SEQ ID NO:3的改变的组合:

Y147T和Q154S,

Y147R和Q154S,

V82G、Y147T和Q154S,以及

R20A、K21A、Y147T、和Q154S;

与GSD1a相关的所述SNP导致具有R83C突变的G6PC多肽的表达;以及一种或多种所述向导RNA包含由核酸序列CAGUAUGGACACUGUCCAAA或AGUAUGGACACUGUCCAAA组成的间隔区。

75.一种用于治疗1a型糖原贮积病(GSD1a)的试剂盒,所述试剂盒包含根据权利要求19至43中任一项所述的细胞。

76.一种碱基编辑器系统,其包含在具有一种或多种向导RNA的复合物中的碱基编辑

器,其中,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结构域,其中,所述腺苷脱氨酶结构域由以下氨基酸序列

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD(SEQ ID NO:3)或其仅缺少N端甲硫氨酸的片段所组成,所述腺苷脱氨酶结构域具有一或多个改变,其中一种或多种所述向导RNA引导所述碱基编辑器以实现与GSD1a相关的SNP的A•T到G•C改变的,以及其中:

a)所述一或多个改变由选自以下参考SEQ ID NO:3的改变或改变的组合所组成:

V82S和Y147R,

V82S和Q154R,

V82S和Q154S,

V82S和Y123H,以及

V82S、Y123H和Y147T;与GSD1a相关的所述SNP导致具有Q347X突变的葡萄糖-6-磷酸酶(G6PC)多肽的表达,其中X代表终止密码子,以及一种或多种所述向导RNA包含由核酸序列GACCUAGGCGAGGCAGUAGG组成的间隔区;或

b)所述一或多个改变由以下所组成:

选自参考SEQ ID NO:3的V82S、Y147T、Y147R和Q154S的改变,或选自以下参考SEQ ID NO:3的改变的组合:

Y147T和Q154S,

Y147R和Q154S,

V82G、Y147T和Q154S,以及

R20A、K21A、Y147T、和Q154S;

与GSD1a相关的所述SNP导致具有R83C突变的G6PC多肽的表达;以及一种或多种所述向导RNA包含由核酸序列CAGUAUGGACACUGUCCAAA或AGUAUGGACACUGUCCAAA组成的间隔区。

77.根据权利要求76所述的碱基编辑器系统,其中,所述碱基编辑器进一步包含第二腺苷脱氨酶结构域,其中,所述第二腺苷脱氨酶结构域是野生型腺苷脱氨酶结构域。

78.根据权利要求76所述的碱基编辑器系统,其中,所述多核苷酸可编程DNA结合结构域是金黄色葡萄球菌Cas9(SaCas9)、嗜热链球菌1Cas9(St1Cas9)、或化脓性链球菌Cas9(SpCas9)。

79.根据权利要求78所述的碱基编辑器系统,其中,所述多核苷酸可编程DNA结合结构域的原型间隔区相邻基序是对非G原型间隔区相邻基序(PAM)具有特异性的SpCas9。

80.根据权利要求78所述的碱基编辑器系统,其中,所述多核苷酸可编程DNA结合结构域是核酸酶无活性Cas9。

81.根据权利要求78所述的碱基编辑器系统,其中,所述多核苷酸可编程DNA结合结构域是Cas9切口酶。

82.一种碱基编辑器系统,包含一种或多种向导RNA和融合蛋白,所述融合蛋白包含多核苷酸可编程DNA结合结构域和腺苷脱氨酶结构域,所述多核苷酸可编程DNA结合结构域由以下序列所组成:

EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTE

VQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFE
KNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAKFLQKGNELALPSKYVNFLYLASHYEKLKGSPED
NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAFKYF
DTTIARKEYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGSGGSGGSGGSGGSGGSGGMDKKYSIGLAIGTNS
VGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMA
KVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGH
FLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIA
LSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSA
SMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNR
EDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEE
TITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL
TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHD
DSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQK
NSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDS
IDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI
TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK
LESEFVYGDYKVYDVRKMIAKSEQEGADKRTADGSEFESPKKKRKV,并且所述腺苷脱氨酶结构域由以下氨基酸序列

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQN
YRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYF
FRMPRQVFNAQKKAQSSTD(SEQ ID NO:3)或其仅缺少N端甲硫氨酸的片段所组成,所述腺苷脱氨酶结构域具有一或多个改变,其中一种或多种所述向导RNA引导所述碱基编辑器以实现与GSD1a相关的SNP的A·T到G·C改变的,以及其中:

a)所述一或多个改变由以下所组成:

选自参考SEQ ID NO:3的V82S、Y147T、Y147R、Q154R和Q154S的改变,或选自以下参考SEQ ID NO:3的改变的组合:

V82S和Y147R,

V82S和Q154R,

V82S和Q154S,

V82S和Y123H,以及

V82S、Y123H和Y147T;与GSD1a相关的所述SNP导致具有Q347X突变的葡萄糖-6-磷酸酶(G6PC)多肽的表达,其中X代表终止密码子;以及一种或多种所述向导RNA包含由核酸序列GACCUAGGCGAGGCAGUAGG组成的间隔区;或

b)所述一或多个改变由以下所组成:

选自参考SEQ ID NO:3的V82S、Y147T、Y147R和Q154S的改变,或选自以下参考SEQ ID NO:3的改变的组合:

Y147T和Q154S,

Y147R和Q154S,

V82G、Y147T和Q154S,以及

R20A、K21A、Y147T、和Q154S;

与GSD1a相关的所述SNP导致具有R83C突变的G6PC多肽的表达;以及一种或多种所述向导RNA包含由核酸序列CAGUAUGGACACUGUCCAAA或AGUAUGGACACUGUCCAAA组成的间隔区。

83.一种细胞,包含根据权利要求76至82中任一项所述的碱基编辑器系统。

84.根据权利要求83所述的细胞,其中,所述细胞是哺乳动物细胞。

85.根据权利要求83所述的细胞,其中,所述细胞是体内的或体外的。

用于治疗1a型糖原贮积病的组成物和方法

[0001] 相关申请的交叉引用

[0002] 本申请是一项国际PCT(Internatinoal PCT)申请,要求于2019年2月13日提交的美国临时申请第62/805,271号;于2019年5月23日提交的第62/852,228号;于2019年5月23日提交的第62/852,224号;于2019年7月19日提交的第62/876,354号;于2019年10月9日提交的第62/912,992号;于2019年11月6日提交的第62/931,722号;于2019年11月27日提交的第62/941,569号;和于2020年1月27日提交的第62/966,526号的优先权和权益,其全部内容通过引用整体并入本文。

[0003] 引用加入

[0004] 本说明书中提及的所有出版物、专利和专利申请均通过引用并入本文,其程度如同个别的出版物、专利或专利申请被具体地和分别地指示为通过引用并入一样。除非另有说明,本说明书中提及的出版物、专利和专利申请通过引用整体并入本文。

技术领域

[0005] 本公开提供包含具有提高效率的新型腺苷碱基编辑器(例如,ABE8)的组合物以及使用包含腺苷脱氨酶变体的碱基编辑器改变与1a型糖原贮积病(GSD1a)相关的突变的方法。

背景技术

[0006] 对于大多数已知的遗传疾病,需要校正目标基因座中的点突变,而不是随机破坏基因,以研究或解决疾病的根本原因。当前的基因组编辑技术利用成簇的规则间隔短回文重复(CRISPR)系统在目标基因座引入双链DNA断裂,作为基因校正的第一步。为回应双链DNA断裂,细胞DNA的修复过程主要通过非同源末端连接在DNA切割位点,导致随机插入缺失(indels)。尽管大多数遗传疾病都是由点突变引起的,但当前的点突变校正方法效率低下,并且通常会在目标基因座上诱发大量随机插入和缺失(indels),这是由细胞对dsDNA断裂的反应引起的。因此,需要一种改进的基因组编辑形式,它更有效,并且非预期的产物少得多,例如随机插入缺失(indels)或易位。

[0007] 糖原贮积病1型(也称为GSD1或冯吉尔克病(Von Gierke Disease))是一种遗传性疾病,导致糖原分解和糖异生缺乏,组织中糖原和脂质积累,引起危及生命的低血糖和乳酸酸中毒,并导致潜在的中枢神经系统损伤和长期肝肾并发症,如脂肪变性、肝腺瘤和肝细胞癌。

[0008] GSD1有两种类型,1a型(GSD1a)和1b型(GSD1b),它们是由不同的基因突变引起的。GSD1a是由葡萄糖6-磷酸酶(G6PC)基因突变引起的,影响约80%的GSD1患者。在美国,大约每100,000名新生儿中就有一个患有GSD1a,大约22%的患者携带隐性突变Q347*,37%的患者携带隐性突变R83C。

[0009] 没有批准用于GSD1a的药物疗法。尽管肝移植是治愈性的,但尚无批准的疗法,目前的治疗方案几乎只涉及持续的玉米淀粉喂食。如果长期不治疗,患者会出现严重的乳酸

性酸中毒,可进展为肾功能衰竭,并在婴儿期或儿童期死亡。GSD1a是一个严重未满足医疗需求的领域。因此,需要有用于治疗患有GSD1a的患者的新组合物和方法。

[0010] 引用加入

[0011] 本说明书中提及的所有出版物、专利和专利申请均通过引用并入本文,其程度如同个别的出版物、专利或专利申请被具体地和分别地指示为通过引用并入一样。除非另有说明,本说明书中提及的出版物、专利和专利申请通过引用整体并入本文。

发明内容

[0012] 本发明的特征在于,使用可编程核碱基编辑器,以精确校正致病氨基酸的组成物和方法。特别地,本发明的组成物和方法可用于治疗1a型糖原贮积病(GSD1a)。因此,本发明提供了治疗GSD1a的组成物和方法,使用腺苷(A)碱基编辑器(ABE)(例如ABE8)精确校正内源性G6PC基因中的单核苷酸多态性,以校正有害突变(例如Q347X、R83C)。

[0013] 一方面,本发明提供了一种编辑包含与糖原贮积病1a型(GSD1a)相关的单核苷酸多态性(SNP)的G6PC多核苷酸的方法,所述方法包含将G6PC多核苷酸与腺苷脱氨酶碱基编辑器8(ABE8)在具有一种或多种向导多核苷酸的复合物中接触,其中所述ABE8包含多核苷酸可编程DNA结合结构域和腺苷脱氨酶结构域,并且其中一种或多种所述向导多核苷酸靶向所述碱基编辑器,以实现与GSD1a相关的所述SNP的A·T至G·C的改变。另一方面,本发明提供一种包含所述腺苷脱氨酶碱基编辑器8(ABE8)的细胞,或编码所述碱基编辑器的多核苷酸,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,以及一种或多种靶向所述碱基编辑器的向导多核苷酸,以实现与GSD1a相关的所述SNP的A·T到G·C的改变。另一方面,本发明提供了一种治疗个体的GSD1a的方法,包括向所述个体施用:一种腺苷脱氨酶碱基编辑器8(ABE8)或包含编码所述碱基编辑器的多核苷酸,其中所述腺苷脱氨酶碱基编辑器8(ABE8)包含多核苷酸可编程DNA结合结构域和腺苷脱氨酶结构域,以及一种或多种靶向所述腺苷脱氨酶碱基编辑器8(ABE8)的向导多核苷酸,以实现与GSD1a相关的所述SNP的A·T至G·C的改变。另一方面,本发明提供了一种产生肝细胞或其祖细胞的方法,所述方法包括:a)引入包含与GSD1a相关的所述SNP、腺苷脱氨酶碱基编辑器8(ABE8)、或编码所述腺苷脱氨酶碱基编辑器8(ABE8)的多核苷酸或诱导多能干细胞或肝细胞祖细胞,其中所述碱基编辑器包含多核苷酸可编程核苷酸结合结构域和腺苷脱氨酶结构域,以及一种或多种向导多核苷酸,其中一种或多种所述向导多核苷酸靶向所述碱基编辑器,以实现与GSD1a相关的所述SNP的A·T到G·C的改变;b)将所述诱导多能干细胞或肝细胞祖细胞分化为肝细胞。

[0014] 一方面,本发明提供了一种编辑包含与1a型糖原贮积病(GSD1a)相关的单核苷酸多态性(SNP)的葡萄糖-6-磷酸酶(G6PC)多核苷酸的方法,所述方法包含将G6PC多核苷酸与腺苷脱氨酶碱基编辑器8(ABE)在具有一种或多种向导多核苷酸的复合物中接触,其中所述腺苷脱氨酶碱基编辑器8(ABE8)包含插入Cas9或Cas12多肽内的腺苷脱氨酶变体结构域,其中一种或多种所述向导多核苷酸靶向所述碱基编辑器,以实现与GSD1a相关的所述SNP的A·T到G·C的改变。另一方面,本发明提供了一种治疗个体1a型糖原贮积病(GSD1a)的方法,所述方法包括向所述个体施用:一种腺苷脱氨酶碱基编辑器8(ABE8)或编码所述碱基编辑器8(ABE8)的多核苷酸,其中所述腺苷脱氨酶碱基编辑器8(ABE8)包含插入Cas9或Cas12多肽内的腺苷脱氨酶变体,以及一种或多种靶向腺苷脱氨酶碱基编辑器8(ABE8)的向导多

核苷酸,以实现与GSD1a相关的所述SNP的A·T至G·C的改变,从而治疗所述个体的GSD1a。又一方面,本发明提供了一种用于治疗个体1a型糖原贮积病(GSD1a)的方法,所述方法包括向所述个体施用:一种包含插入Cas9或Cas12多肽内的腺苷脱氨酶变体的融合蛋白或编码其融合蛋白的多核苷酸,和一种或多种向导多核苷酸靶向所述融合蛋白,以实现与GSD1a相关的单核苷酸多态性(SNP)的A·T到G·C改变,从而治疗所述个体的GSD1a。

[0015] 一方面,本发明提供了一种用于治疗1a型糖原贮积病(GSD1a)的医药组成物,其包含有效量的腺苷脱氨酶碱基编辑器8(ABE8),其中所述腺苷脱氨酶碱基编辑器8(ABE8)包含多核苷酸可编程DNA结合结构域和腺苷脱氨酶变体结构域。在一些实施例中,所述医药组成物包含一种或多种向导多核苷酸,其能够靶向所述腺苷脱氨酶碱基编辑器8(ABE8),以实现与GSD1a相关的所述SNP的A·T至G·C的改变。另一方面,本发明提供用于治疗1a型糖原贮积病(GSD1a)的医药组成物,其包含有效量的本文提供的任何细胞。在一些实施例中,所述医药组成物包含药学上可接受的赋形剂。

[0016] 另一方面,本发明提供一种用于治疗1a型糖原贮积病(GSD1a)的试剂盒,所述试剂盒包含腺苷脱氨酶碱基编辑器8(ABE8),其中所述腺苷脱氨酶碱基编辑器8(ABE8)包含多核苷酸可编程DNA结合结构域和腺苷脱氨酶结构域,以及一种或多种向导多核苷酸,其能够靶向所述腺苷脱氨酶碱基编辑器8(ABE8),以实现与GSD1a相关的所述SNP的A·T到G·C的改变。又一方面,本发明提供用于治疗1a型糖原贮积病(GSD1a)的试剂盒,所述试剂盒包含本文提供的任何细胞。

[0017] 在一些实施例中,所述接触是在细胞、真核细胞、哺乳动物细胞或人类细胞中进行。在一些实施例中,所述细胞在体内。在一些实施例中,所述细胞是离体的。在一些实施例中,所述细胞是肝细胞、肝细胞前体或iPSc衍生的肝细胞。在一些实施例中,所述细胞表达G6PC多肽。在一些实施例中,所述细胞或肝细胞祖细胞来自患有GSD1a的个体。在一些实施例中,所述个体是哺乳动物或人。在一些实施例中,所述肝细胞或肝细胞祖细胞是哺乳动物细胞或人类细胞。在一些实施例中,将所述腺苷脱氨酶碱基编辑器8(ABE8)或编码所述腺苷脱氨酶碱基编辑器8(ABE8)的多核苷酸,和一种或多种向导多核苷酸递送至个体的细胞。

[0018] 在上述方面或本文描述的本发明任何其他方面的各种实施例中,与GSD1a相关的SNP位于葡萄糖-6-磷酸酶(G6PC)基因中。在一个实施例中,与1a型糖原贮积病(GSD1a)相关的SNP处的A·T到G·C的改变将谷氨酰胺(Q)改变为非谷氨酰胺(X)氨基酸。在一个实施例中,与1a型糖原贮积病(GSD1a)相关的所述SNP的A·T到G·C的改变将G6PC多肽中的精氨酸(R)改变为非精氨酸(X)。在一个实施例中,与GSD1a相关的所述SNP导致在位置347处具有非谷氨酰胺氨基酸(X)或在位置83处具有非精氨酸氨基酸(X)的G6PC多肽的表达。在一个实施例中,所述碱基编辑器的校正将位置347处的非谷氨酰胺氨基酸(X)替换为谷氨酰胺。在另一个实施例中,所述碱基编辑器的校正将位置83处的非精氨酸氨基酸(X)替换为精氨酸。在一个实施例中,与GSD1a相关的SNP处的A·T至G·C改变导致在位置347的氨基酸处过早终止或在位置83处编码半胱氨酸的G6PC多肽的表达。在一些实施例中,所述SNP处的改变是Q347X和/或R83C中的一个或多个。

[0019] 在本文描述的本发明的上述方面或任何其他方面的各种实施例中,所述腺苷脱氨酶变体被插入到Cas9或Cas12多肽的柔性环、α螺旋区域、非结构化部分或溶剂可及部分内。在一些实施例中,所述腺苷脱氨酶变体侧接所述Cas9或Cas12多肽的N-末端片段和C-末端

片段。在一些实施例中,所述融合蛋白或腺苷脱氨酶碱基编辑器8(ABE8)包含NH2-[所述Cas9或Cas12多肽的N-末端片段]-[腺苷脱氨酶变体]-[所述Cas9或Cas12多肽的C-末端片段]-COOH的结构,其中每个"]-["是任选的接头。在一个实施例中,所述N-末端片段的C-末端或C-末端片段的N-末端包含所述Cas9或Cas12多肽的柔性环的一部分。在一个实施例中,所述柔性环包含靠近目标核碱基的氨基酸。在一些实施例中,所述一种或多种向导多核苷酸引导所述融合蛋白或腺苷脱氨酶碱基编辑器8(ABE8)以实现目标核碱基的脱氨基作用。在一些实施例中,所述SNP目标核碱基的脱氨基作用以非野生型核碱基替换所述目标核碱基,并且其中所述目标核碱基的脱氨基作用改善了GSD1a的症状。在一个实施例中,所述目标核碱基与目标多核苷酸序列中的PAM序列相距1-20个核碱基。在一个实施例中,所述目标核碱基是所述PAM序列上游的2-12个核碱基。

[0020] 在一个实施例中,所述Cas9或Cas12多肽的N-末端片段或C-末端片段与所述目标多核苷酸序列结合。在一个实施例中,所述N-末端片段或C-末端片段包含RuvC结构域;所述N-末端片段或C-末端片段包含HNH结构域;所述N-末端片段和C-末端片段均不包含HNH结构域;或所述N-末端片段和C-末端片段均不包含RuvC结构域。在一个实施例中,所述Cas9或Cas12多肽在一个或多个结构域中包含部分或完全缺失,并且其中所述脱氨酶插入在所述Cas9或Cas12多肽的部分或完全缺失处。在一个实施例中,所述缺失位于RuvC结构域内;所述缺失位于HNH结构域内;或所述缺失将RuvC结构域和C-末端结构域桥接、L-I结构域和HNH结构域桥接、或RuvC结构域和L-I结构域桥接。

[0021] 在各种实施例中,所述多核苷酸可编程DNA结合结构域是Cas9多肽。在一些实施例中,所述融合蛋白或腺苷脱氨酶碱基编辑器8(ABE8)包含插入所述Cas9多肽中的腺苷脱氨酶变体结构域。在一些实施例中,所述Cas9多肽是化脓性链球菌Cas9(SpCas9)、金黄色葡萄球菌Cas9(SaCas9)、嗜热链球菌1Cas9(St1Cas9)或其变体。在一些实施例中,所述Cas9多肽具有以下氨基酸序列(Cas9参考序列):

[0022] MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSG
ETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNP

DNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG
LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNL
SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSK
NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHL
GELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW
NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRK
PAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHD
LLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRY
TGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ
[0023] GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKN
SRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDY
DVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK
FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVI
TLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVY
DVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDK
GRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDS
PTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLII
KLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQ
LFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT
NLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD（单下划线:HNH结构域;双下划线:RuvC结构域;(Cas9参考序列)),或其相应区域。

[0024] 在一些实施例中,所述Cas9多肽包含在所述Cas9多肽参考序列中编号为1017-1069的氨基酸或其相应氨基酸的缺失;所述Cas9多肽包含在所述Cas9多肽参考序列中编号为792-872的氨基酸或其相应氨基酸的缺失;或所述Cas9多肽包含在所述Cas9多肽参考序列中编号为792-906的氨基酸或其相应氨基酸的缺失。在一些实施例中,所述腺苷脱氨酶变体被插入到所述Cas9多肽的柔性环内。在一些实施例中,所述柔性环包含选自由以下组成的组的区域:在所述Cas9参考序列中位置编号为530-537、569-579、686-691、768-793、943-947、1002-1040、1052-1077、1232-1248和1298-1300的氨基酸残基的位置或与其相应的氨基酸位置。在一些实施例中,脱氨酶被插入在所述Cas9参考序列中编号为768-769、791-792、792-793、1015-1016、1022-1023、1026-1027、1029-1030、1040-1041、1052-1053、1054-1055、1067-1068、1068-1069、1247-1248或1248-1249之间的氨基酸位置或与其相应的氨基酸位置。在一些实施例中,脱氨酶被插入在所述Cas9参考序列中编号为768-769、792-793、1022-1023、1026-1027、1040-1041、1068-1069或1247-1248之间的氨基酸位置或与其相应的氨基酸位置。在一些实施例中,脱氨酶被插入在所述Cas9参考序列中编号为1016-1017、

1023-1024、1029-1030、1040-1041、1069-1070或1247-1248之间的氨基酸位置或与其相应的氨基酸位置。在一些实施例中,在,所述腺苷脱氨酶变体被插入到表10A中鉴定的基因座处的Cas9多肽内。在一些实施例中,所述N-末端片段包含所述Cas9参考序列的1-529、538-568、580-685、692-942、948-1001、1026-1051、1078-1231和/或1248-1297之间的氨基酸残基或与其相应的残基。在一些实施例中,所述C-末端片段包含所述Cas9参考序列的1301-1368、1248-1297、1078-1231、1026-1051、948-1001、692-942、580-685和/或538-568之间的氨基酸残基或与其相应的残基。

[0025] 在一些实施例中,所述Cas9多肽是切口酶,或其中所述Cas9多肽是核酸酶无活性的。在一些实施例中,所述Cas9多肽是经修饰的SpCas9,并且对改变的PAM具有特异性或对非G PAM具有特异性。在一些实施例中,所述经修饰的SpCas9多肽包括了D1135M、S1136Q、G1218K、E1219F、A1322R、D1332A、R1335E和T1337R(SpCas9-MQKFRAER)的氨基酸置换,并且对改变的PAM 5'-NGC-3'具有特异性。

[0026] 在各种实施例中,所述多核苷酸可编程DNA结合结构域是经修饰的化脓性链球菌Cas9(SpCas9)或其变体。在本文描述的本发明的上述方面或任何其他方面的各种实施例中,所述多核苷酸可编程DNA结合结构域包含所述经修饰的SpCas9,其具有改变的原型间隔区相邻基序(PAM)特异性或对非G PAM的特异性。在一个实施例中,所述经修饰的SpCas9对核酸序列5'-NGA-3'具有特异性。在一个实施例中,所述经修饰的SpCas9对核酸序列5'-AGA-3'或5'-GGA-3'具有特异性。在一个实施例中,所述经修饰的SpCas9对NGA PAM变体具有特异性。

[0027] 在各种实施例中,所述多核苷酸可编程DNA结合结构域是金黄色葡萄球菌Cas9(SaCas9)或其变体。在一个实施例中,所述SaCas9对核酸序列5'-NNGRRT-3'具有特异性。在一个实施例中,所述SaCas9对核酸序列5'-GAGAAT-3'具有特异性。在一个实施例中,所述SaCas9对NNGRRT PAM变体具有特异性。

[0028] 在各种实施例中,所述多核苷酸可编程DNA结合结构域是Cas12多肽。在一个实施例中,所述腺苷脱氨酶变体被插入到所述Cas12多肽中。在一个实施例中,所述Cas12多肽是Cas12a、Cas12b、Cas12c、Cas12d、Cas12e、Cas12g、Cas12h或Cas12i。在一个实施例中,所述腺苷脱氨酶变体被插入到以下氨基酸位置之间:a)BhCas12b的153-154、255-256、306-307、980-981、1019-1020、534-535、604-605或344-345之间,或Cas12a、Cas12c、Cas12d、Cas12e、Cas12g、Cas12h或Cas12i的相应氨基酸残基;b)BvCas12b的147和148、248和249、299和300、991和992,或1031和1032之间,或Cas12a、Cas12c、Cas12d、Cas12e、Cas12i、Cas12h的相应氨基酸残基;或c)AaCas12b的157和158、258和259、310和311、1008和1009、或1044和1045之间,或Cas12a、Cas12c、Cas12d、Cas12e、Cas12i、Cas12h的相应氨基酸残基。在一个实施例中,所述腺苷脱氨酶变体被插入到表10B中鉴定的基因座处的所述Cas12多肽内。在一个实施例中,所述Cas12多肽是Cas12b。在一个实施例中,所述Cas12多肽包含BhCas12b结构域、BvCas12b结构域或AACas12b结构域。

[0029] 在各种实施例中,所述多核苷酸可编程DNA结合结构域是核酸酶无活性变体。在其他实施例中,所述多核苷酸可编程DNA结合结构域是切口酶变体。在一个实施例中,所述切口酶变体包含D10A氨基酸置换或其相应的氨基酸置换。在一些实施例中,所述腺苷脱氨酶结构域能够使脱氧核糖核酸(DNA)中的腺苷脱氨基。在一些实施例中,所述腺苷脱氨酶结构

域是包含腺苷脱氨酶变体的单体。在一些实施例中,所述腺苷脱氨酶结构域是包含野生型腺苷脱氨酶结构域和腺苷脱氨酶变体的异源二聚体。

[0030] 在一些实施例中,所述腺苷脱氨酶变体包含以下氨基酸序列:MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD;其中所述氨基酸序列包含至少一个改变。在一些实施例中,相对于上述序列,所述腺苷脱氨酶变体在氨基酸位置82和/或166处包含改变。在一些实施例中,相对于上述序列,所述至少一种改变包含:V82S、Y147T、Y147R、Q154S、Y123H和/或Q154R。在一些实施例中,所述至少一种改变包含选自由以下组成的组的改变的组合: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。在一些实施例中,相对于上述序列,所述至少一种改变是Y147T+Q154S。

[0031] 在一些实施例中,所述腺苷脱氨酶变体包含在选自由149、150、151、152、153、154、155、156和157组成的组中的残基开始的C-末端的缺失。在一些实施例中,所述腺苷脱氨酶变体是包含TadA*8腺苷脱氨酶变体结构域的腺苷脱氨酶单体。在一些实施例中,所述腺苷脱氨酶变体是包含野生型腺苷脱氨酶结构域和TadA*8腺苷脱氨酶变体结构域的腺苷脱氨酶异源二聚体。在一些实施例中,所述腺苷脱氨酶变体是包含TadA结构域和TadA*8腺苷脱氨酶变体结构域的腺苷脱氨酶异源二聚体。

[0032] 在一些实施例中,所述向导多核苷酸包含选自下组的核酸序列:

[0033] a)GACCUAGGCGAGGCAGUAGG;

[0034] b)CCAGUAUGGACACUGUCCAAA;

[0035] c)CAGUAUGGACACUGUCCAAA;和

[0036] d)AGUAUGGACACUGUCCAAAG。

[0037] 在一些实施例中,所述一种或多种向导RNA包含CRISPR RNA(crRNA)和反式编码的小RNA(tracrRNA),其中所述crRNA包含与G6PC核酸序列互补的核酸序列,所述G6PC核酸序列包含与GSD1a相关的所述SNP。在一些实施例中,所述腺苷脱氨酶碱基编辑器8(ABE8)与包含与G6PC核酸序列互补的核酸序列的单向导RNA(sgRNA)复合,所述G6PC核酸序列包含与GSD1a相关的所述SNP。

[0038] 在一些实施例中,所述腺苷脱氨酶是TadA脱氨酶。在一个实施例中,所述TadA脱氨酶是TadA*8变体。在一些实施例中,所述TadA*8变体选自由以下组成的组: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和TadA*8.24。在一些实施例中,所述腺苷脱氨酶碱基编辑器8(ABE8)选自由以下组成的组: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。

[0039] 在一些实施例中,所述腺苷脱氨酶碱基编辑器8(ABE8)包含或基本上由以下具有腺苷脱氨酶活性的序列或其片段组成:

[0040] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD。

[0041] 在一些实施例中,所述gRNA包含具有以下序列的支架:

[0042] GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU。

[0043] 在一些实施例中,所述gRNA包含具有以下序列的支架:

[0044] GUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCUACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUU。

[0045] 在一个方面,本文提供了一种碱基编辑器,其包含与一种或多种向导多核苷酸复合的腺苷脱氨酶碱基编辑器8(ABE8),其中所述腺苷脱氨酶碱基编辑器8(ABE8)包含多核苷酸可编程DNA结合结构域和腺苷脱氨酶结构域,并且其中一种或多种所述向导多核苷酸靶向所述碱基编辑器,以实现与GSD1a相关的所述SNP的A·T到G·C改变。在一些实施例中,所述腺苷脱氨酶变体包含V82S和/或T166R的改变。在一些实施例中,所述腺苷脱氨酶变体还包含一个或多个以下改变:Y147T、Y147R、Q154S、Y123H和Q154R。在一些实施例中,所述碱基编辑器结构域包含腺苷脱氨酶异源二聚体,所述腺苷脱氨酶异源二聚体包含野生型腺苷脱氨酶结构域和腺苷脱氨酶变体。在一些实施例中,所述腺苷脱氨酶变体是截短的TadA8,所述截短的TadA8相对于全长TadA8,缺失了1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个N-末端氨基酸残基。在一些实施例中,所述腺苷脱氨酶变体是截短的TadA8,所述截短的TadA8相对于全长TadA8,缺失了1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个C-末端氨基酸残基。在一些实施例中,所述多核苷酸可编程DNA结合结构域是经修饰的金黄色葡萄球菌Cas9(SaCas9)、嗜热链球菌1Cas9(St1Cas9)、经修饰的化脓性链球菌Cas9(SpCas9)或其变体。在一些实施例中,所述多核苷酸可编程DNA结合结构域是SpCas9的变体,其具有对改变的原型间隔区相邻基序(PAM)的特异性或对非G PAM的特异性。在一些实施例中,所述多核苷酸可编程DNA结合结构域是核酸酶无活性Cas9。在一些实施例中,所述多核苷酸可编程DNA结合结构域是Cas9切口酶。

[0046] 一方面,本文提供了一种碱基编辑器系统,其包含一种或多种向导RNA和融合蛋白,所述融合蛋白包含多核苷酸可编程DNA结合结构域,所述多核苷酸可编程DNA结合结构域包含以下序列:

[0047] EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAKFLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGSGGSGGSGGSGGSGGSGGMDKKYSIGLAI

GTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS
NEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIK
FRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFG
NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKA
PLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLV
KLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTR
KSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGE
QKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDI
VLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQ
LIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQ
KGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFL
KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVE
TRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIK
KYPKLESEFVYGDYKVYDVRKMIAKSEQEGADKRTADGSEFESPKKKRKV*,其中粗体序列表示源自Cas9的序列,斜体序列表示接头序列,下划线序列表示双分核定位序列,并且至少一个碱基编辑器结构域包含腺苷脱氨酶变体,所述腺苷脱氨酶变体包含在氨基酸位置82和/或166处的改变:

[0048] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAAL
LCYFFRMPRQVFNAQKKAQSSTD,并且其中一种或多种所述向导多核苷酸靶向所述碱基编辑器以实现与GSD1a相关的所述SNP的A·T到G·C改变。

[0049] 在一个方面,提供了一种包含以上描述的任一碱基编辑器系统的细胞。在一些实施例中,所述细胞是人细胞或哺乳动物细胞。在一些实施例中,所述细胞是离体的、体内的或体外的。

[0050] 此处的描述和示例详细说明了本公开的实施例。应当理解,本公开不限于本文描述的特定实施例并且因此可以有所不同。本领域技术人员将认识到,本公开内容有许多变化和修改,这些变化和修改都包含在其范围内。

[0051] 除非另外指明,否则本文公开的一些实施例的实践采用免疫学、生物化学、化学、分子生物学、微生物学、细胞生物学、基因组学和重组DNA的常规技术,这些技术在本领域技术范围内。参见例如Sambrook和Green,Molecular Cloning:A Laboratory Manual,4th Edition(2012);Current Protocols in Molecular Biology系列(F.M.Ausubel等人编辑);Methods In Enzymology系列(Academic Press,Inc.),PCR 2:A Practical Approach(M.J.MacPherson、B.D.Hames和G.R.Taylor编辑(1995)),Harlow和Lane编辑(1988)Antibodies,A Laboratory Manual,and Culture of Animal Cells:A Manual of Basic Technique and Specialized Applications,6th Edition(R.I.Freshney编辑(2010))。

[0052] 尽管可以在单个实施例的上下文中描述本公开的各种特征,但是也可以单独地或以任何合适的组合来提供这些特征。相反,尽管为了清楚起见可以在单独的实施例的上下文中描述本公开,但是本公开也可以在单个实施例中实施。此处使用的章节标题仅用于组织目的,不应被解释为限制所描述的主题。

[0053] 本公开的特征在所附权利要求中具体阐述。通过参考以下说明性实施例的详细描述,其中利用了本公开的原理,并鉴于如下所述的附图,将获得对本发明的特征和优点更好的理解。

[0054] 定义

[0055] 以下定义补充了本领域中的定义,并且针对当前申请,并且不归咎于任何相关或不相关的案例,例如,任何共同拥有的专利或申请。尽管与本文所述的那些相似或等效的任何方法和材料可用于测试本公开的实践中,但本文描述了首选的材料和方法。因此,本文中使用的术语仅用于描述特定实施例的目的,并不旨在进行限制。

[0056] 除非另有定义,本文使用的所有技术和科学术语均具有本发明所属领域的技术人员通常理解的含义。以下参考文献为技术人员提供了本发明中使用的许多术语的一般定义:Singleton等人,Dictionary of Microbiology and Molecular Biology(第2版,1994);The Cambridge Dictionary of Science and Technology(Walker编辑,1988);The Glossary of Genetics,5th Ed.,R.Rieger等人(编辑),Springer Verlag(1991);和Hale&Marham,The Harper Collins Dictionary of Biology(1991)。

[0057] 在本申请中,除非另有特别说明,否则单数的使用包括复数。必须注意,在说明书中使用的单数形式“一个(a)”、“一个(an)”和“所述(the)”包括复数参考,除非上下文另有明确规定。在本申请中,除非另有说明,否则“或”的使用是指“和/或”,并且被理解为包括在内的。此外,术语“包括(including)”以及诸如“包括(include)”、“包括(includes)”和“包括(included)”等其他形式的使用不是限制性的。

[0058] 如在本说明书和权利要求中使用的,词语“包括(comprising)”(以及任何形式的包括,例如“包含(comprise)”和“包含(comprises)”)、“具有(having)”(以及任何形式的具有,例如“具有(have)”和“具有(has)”)、“包括(including)”(以及任何形式的包括,例如“包括(includes)”和“包括(include)”)或“包含(containing)”(以及任何形式的包含,例如“包含(contains)”和“包含(contain)”)是包含性或开放式的,并且不排除额外的、未提及的要素或方法步骤。预期本说明书中讨论的任何实施例可关于本公开的任何方法或组合实施,反之亦然。此外,本公开的组成物可用于实现本公开的方法。

[0059] 术语“约(about)”或“约(approximately)”是指在由本领域普通技术人员确定的特定值的可接受误差范围内,这将部分取决于如何测量或确定该值,即测量系统的局限。例如,根据本领域的实践,“约”可以表示在1个标准偏差以内或超过1个标准偏差。或者,“约”可以表示给定值的最多20%、最多10%、最多5%或最多1%的范围。或者,特别是对于生物系统或过程,所述术语可表示在一个数量级内,例如在值的5倍或2倍内。在申请和权利要求中描述特定值的情况下,除非另有说明,术语“约”的含义应假定在特定值的可接受误差范围内。

[0060] 此处提供的范围应理解为所述范围内所有值的简写。例如,1至50的范围被理解为包括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中的任何数字、数字组合或子范围。

[0061] 说明书中对“一些实施例(some embodiments)”、“一个实施例(an embodiment)”、“一个实施例(one embodiment)”或“其他实施例(other embodiments)”的引用是指结合实

施例描述的特定特征、结构或特性包括在至少一些实施例中,但不一定是本公开的所有实施例。

[0062] “腺苷脱氨酶”是指能够催化腺嘌呤或腺苷水解脱氨基的多肽或其片段。在一些实施例中,所述脱氨酶或脱氨酶结构域催化腺苷水解脱氨基为肌苷,或脱氧腺苷水解脱氨基为脱氧肌苷的腺苷脱氨酶。在一些实施例中,所述腺苷脱氨酶催化脱氧核糖核酸(DNA)中腺嘌呤或腺苷的水解脱氨基。本文提供的腺苷脱氨酶(例如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,例如细菌。

[0063] 在一些实施例中,所述腺苷脱氨酶是TadA脱氨酶。在一些实施例中,所述TadA脱氨酶是TadA变体。在一些实施例中,所述TadA变体是TadA*8。在一些实施例中,所述脱氨酶或脱氨酶结构域是来自生物体例如人、黑猩猩、大猩猩、猴、牛、狗、大鼠或小鼠的天然存在的脱氨酶的变体。在一些实施例中,所述脱氨酶或脱氨酶结构域在自然界中不存在。例如,在一些实施例中,所述脱氨酶或脱氨结构酶域与天然存在的脱氨酶至少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%相同。例如,国际PCT申请号PCT/2017/045381(WO 2018/027078)和PCT/US2016/058344(WO 2017/070632)中描述了脱氨酶结构域,它们各自通过引用整体并入本文。另请见Komor,A.C.等人,“Programmable editing of atarget base in genomic DNA without double-stranded DNA cleavage”Nature 533,420-424(2016);Gaudelli,N.M.等人,“Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage”Nature 551,464-471(2017);Komor,AC等人,“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)和Rees,HA等人,“Base editing:precision chemistry on the genome and transcriptome of living cells.”Nat Rev Genet2018Dec;19(12):770-788.doi:10.1038/s41576-018-0059-1,其全部内容通过引用并入本文。

[0064] 野生型TadA(wt)腺苷脱氨酶具有以下序列(也称为TadA参考序列):

[0065] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD

[0066] 在一些实施例中,所述腺苷脱氨酶包含以下序列的改变:

[0067] MSEVEFSHEY WMRHALTLAK RARDEREVPVGAVLVLNNRV IGEGWNRAIG LHDPTAHAEI MALRQGGLVM QNYRLIDATL YVTFEPCVMC AGAMIHSRIG RVVFGVRNAK TGAAGSLMDV LHYPGMNHRV EITEGILADE CAALLCYFFR MPRQVFNAQK KAQSSTD

[0068] (也称为TadA*7.10)。

[0069] 在一些实施例中,TadA*7.10包括至少一种改变。在一些实施例中,TadA*7.10包含氨基酸82和/或166处的改变。在特定实施例中,上述序列的变体包含一个或多个以下改变:Y147T、Y147R、Q154S、Y123H、V82S、T166R和/或Q154R。改变Y123H在本文中也称为H123H(TadA*7.10中H123Y的改变恢复到Y123H(wt))。在其他实施例中,TadA*7.10序列的变体包

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;和I76Y+V82S+Y123H+Y147R+Q154R。

[0070] 在其他实施例中,本发明提供了包含缺失的腺苷脱氨酶变体例如TadA*8,相对于TadA*7.10、TadA参考序列或另一个TadA中的相应突变,其包含从残基149、150、151、152、153、154、155、156或157处开始的C末端的缺失。在其他实施例中,腺苷脱氨酶变体是TadA(例如TadA*8)单体,相对于TadA*7.10、TadA参考序列或另一个TadA中的相应突变,其包含一个或多个以下改变:Y147T、Y147R、Q154S、Y123H、V82S、T166R和/或Q154R。在其他实施例中,相对于TadA*7.10、TadA参考序列或另一个TadA中的相应突变,腺苷脱氨酶变体是包含选自以下组的改变的组合的单体: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。

[0071] 在其他实施例中,所述腺苷脱氨酶变体是包含两个腺苷脱氨酶结构域(例如TadA*8)的同型二聚体,相对于TadA*7.10、TadA参考序列或另一个TadA中的相应突变,各具有一种或多种以下改变:Y147T、Y147R、Q154S、Y123H、V82S、T166R和/或Q154R。在其他实施例中,所述腺苷脱氨酶变体是包含两个腺苷脱氨酶结构域(例如TadA*8)的同型二聚体,相对于TadA*7.10、TadA参考序列或另一个TadA中的相应突变,各具有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;和I76Y+V82S+Y123H+Y147R+Q154R。

[0072] 在其他实施例中,所述腺苷脱氨酶变体是包含野生型TadA腺苷脱氨酶结构域和腺苷脱氨酶变体结构域(例如TadA*8)的异源二聚体,相对于TadA*7.10、TadA参考序列或另一个TadA中的相应突变,其包含一种或多种以下改变:Y147T、Y147R、Q154S、Y123H、V82S、T166R和/或Q154R。在其他实施例中,所述腺苷脱氨酶变体是包含野生型TadA腺苷脱氨酶结构域和腺苷脱氨酶变体结构域(例如TadA*8)的异源二聚体,相对于TadA*7.10、TadA参考序列或另一个TadA中的相应突变,其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;和I76Y+V82S+Y123H+Y147R+Q154R。

[0073] 在其他实施例中,所述腺苷脱氨酶变体是包含TadA*7.10结构域和腺苷脱氨酶变体结构域(例如TadA*8)的异源二聚体,相对于TadA*7.10、TadA参考序列或另一个TadA中的相应突变,其包含一个或多个以下改变Y147T、Y147R、Q154S、Y123H、V82S、T166R和/或Q154R。在其他实施例中,所述腺苷脱氨酶变体是包含TadA*7.10结构域和腺苷脱氨酶变体结构域(例如TadA*8)的异源二聚体,相对于TadA*7.10、TadA参考序列或另一个TadA中的相应突变,其包含以下改变的组合: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;或I76Y+V82S+Y123H+Y147R+Q154R。

[0074] 在一个实施例中,所述腺苷脱氨酶是TadA*8,其包含或基本上由以下具有腺苷脱氨酶活性的序列或其片段组成:

[0075] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD.

[0076] 在一些实施例中,所述TadA*8被截短。在一些实施例中,相对于全长TadA*8,所述截短的TadA*8缺失了1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个N-末端氨基酸残基。在一些实施例中,相对于全长TadA*8,所述截短的TadA*8缺失了1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个C-末端氨基酸残基。在一些实施例中,所述腺苷脱氨酶变体是全长TadA*8。

[0077] 在特定实施例中,腺苷脱氨酶异源二聚体包含选自以下之一的TadA*8结构域和腺苷脱氨酶结构域:

[0078] 金黄色葡萄球菌(S.aureus)TadA:

[0079] MGSHMTNDIYFMTLAIEEAKKAAQLGEVPIGAIITKDDEVIARAHNLRETLQQPTAHAEHIAIERAAKVLGSWRLEGCTLYVTLEPCVMCAGTIVMSRIPRVVYGADDPKGGCSGSLMNLLQQSNFNHRAIVDKGVLKEACSTLLTTFFKNLRANKKSTN

[0080] 枯草芽孢杆菌(B.subtilis)TadA:

[0081] MTQDELYMKEAIKEAKKAEEKGEVPIGAVLVINGEIIARAHNLRETEQRSIAHAEMLVIDEACKALGTWRLEGATLYVTLEPCPMCAGAVVLSRVEKVVFGAFDPKGGCSGTLMNLLQEERFNHQAEVVSGVLEEECGGMLSAFFRELRKKKKAARKNLSE

[0082] 鼠伤寒沙门氏杆菌(S.typhimurium)TadA:

[0083] MPPAFITGVTSLSDVELDHEYWMRHALTLAKRAWDEREVPVGAVLVHNHRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVLQNYRLLDTTLYVTLEPCVMCAGAMVHSRIGRVVFGARDAKTGAAGSLIDVLHHPGMNHRVEIIEGVLRDECATLLSDFFRMRRQEIKALKKADRAEGAGPAV

[0084] 腐败希瓦氏菌(S.putrefaciens)TadA:

[0085] MDEYWMQVAMQMAEKAEAAGEVPVGAVLVKDGQQIATGYNLSISQHDPTAHAEILCLRSAGKKLENYRLLDATLYITLEPCAMCAGAMVHSRIARVVYGARDEKTGAAGTVVNLLQHPAFNHQVEVTSGVLAEACSAQLSRFFKRRRDEKKALKLAQRAQQGIE

[0086] 流感嗜血杆菌F3031(H.influenzae)TadA:

[0087] MDAAKVRSEFDEKMMRYALELADKAEALGEIPVGAVLVDDARNIIGEGWNLSIVQSDPTAHAEIIALRNGAKNIQNYRLLNSTLYVTLEPCTMCAGAILH

[0088] SRIKRLVFGASDYKTGAIGSRFHFFDDYKMNHTLEITSGVLAEECSQKLSTFFQKRREEKKIEKALLKSLSDK

[0089] 新月柄杆菌(C.crescentus)TadA:

[0090] MRTDESEDQDHRMMRLALDAARAAAEAGETPVGAVILDPSTGEVIATAGNGPIAAHDPTAHAEIAAMRAAAAKLGNYRLTDLTLVVTLEPCAMCAGAISHARIGRVVFGADDPKGGAVVHGPKFFAQPTCHWRPEVTGGVLADESA

DLLRGFFRARRKAKI

[0091] 硫还原地杆菌(G.sulfurreducens)TadA:

[0092] MSSLKKTPIRDDAYWMGKAIREAAKAAARDEVPIGAVIVRDGAVIGRGHNLREGSNDPSAHAEMIAIRQAARRSANWRLTGATLYVTLEPCLMCMGAIILARLERVVFGCYDPKGGAAGSLYDLSADPRLNHQVRLSPGVCQEECGTMLSDFFRDLRRRKKAKATPALFIDERKVPPEP

[0093] TadA*7.10

[0094] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

[0095] “腺苷脱氨酶基地编辑器8(ABE8)多肽”或“ABE8”是指一种碱基编辑器,如本文所定义的,其包含腺苷脱氨酶的变体,所述腺苷脱氨酶变体包含下面参考序列中氨基酸位置82和/或166处的改变:MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD一些实施例中,如本文所述的,相对于参考序列,ABE8还包含进一步改变。

[0096] “腺苷脱氨酶碱基编辑器8(ABE8)多核苷酸”是指编码ABE8的多核苷酸。

[0097] “给药”在本文中是指向患者或个体提供本文所述的一种或多种组成物。例如但不限于,组成物给药,例如注射,可以通过静脉内(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.)进行注射。可以采用一种或多种所述的途径。肠胃外给药可以是,例如,通过弹丸式注射或随时间逐渐灌注。或者,或同时,可以通过口服途径给药。

[0098] “药剂”是指任何小分子化合物、抗体、核酸分子或多肽,或其片段。

[0099] “改变”是指基因或多肽的结构、表达水平或活性的变化(例如增加或减少),如通过标准本领域已知方法例如本文所述的那些方法检测的。如本文所用,改变包括多核苷酸或多肽序列的变化或表达水平的变化,例如25%变化、40%变化、50%变化或更大。

[0100] “改善”是指减少、抑制、减弱、减弱、阻止或稳定疾病的发展或进展。

[0101] “类似物”是指不相同但具有类似功能或结构特征的分子。例如,多核苷酸或多肽类似物保留了相应的天然存在的多核苷酸或多肽的生物活性,同时相对于天然存在的多核苷酸或多肽,具有增强所述类似物功能的某些修饰。这种修饰可以增加类似物对DNA的亲和力、效率、特异性、蛋白酶或核酸酶抗性、膜渗透性和/或半衰期,而不改变例如配体结合。类似物可包括非天然核苷酸或氨基酸。

[0102] “碱基编辑器(BE)”或“核碱基编辑器(NBE)”是指结合多核苷酸并具有核碱基修饰活性的物质。在各种实施例中,所述碱基编辑器包含核碱基修饰多肽(例如脱氨酶)和与向导多核苷酸(例如向导RNA)结合的核酸可编程核苷酸结合结构域。在各种实施例中,所述物质是包含具有碱基编辑活性的蛋白质结构域的生物分子复合物,即能够修饰核酸分子(例如A、T、C、G或U)内的碱基(例如脱氧核糖核酸)。在一些实施例中,多核苷酸可编程DNA结合结构域与脱氨酶结构域融合或连接。在一个实施例中,所述物质是包含具有碱基编辑活性的结构域的融合蛋白。在另一个实施例中,具有碱基编辑活性的蛋白质结构域与向导RNA连接(例如通过向导RNA上的RNA结合基序和与脱氨酶融合的RNA结合结构域)。在一些实施例

中,所述具有碱基编辑活性的结构域能够使核酸分子内的碱基脱氨基。在一些实施例中,所述碱基编辑器能够使DNA分子内的一个或多个碱基脱氨基。在一些实施例中,所述碱基编辑器能够使DNA内的腺苷(A)脱氨基。在一些实施例中,所述碱基编辑器是腺苷碱基编辑器(ABE)。

[0103] 在一些实施例中,碱基编辑器是通过将腺苷脱氨酶变体(例如TadA*8)克隆到包含环状置换Cas9(例如spCAS9或saCAS9)和双分核定位序列支架来产生的(例如ABE8)。环状排列Cas9s是本领域已知并且被描述于例如Oakes等人,Cell 176,254-267,2019。示例性环状排列如下,其中粗体序列表示源自Cas9的序列,斜体序列表示接头序列,下划线序列表示双分核定位序列。

[0104] CP5(带有MSP“NGG=带有突变常规Cas9如NGG的Pam变体”PID=蛋白质相互作用结构域和“D10A”切口酶):

[0105] EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELE
NGRKRMLASAKFLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD
EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAFKYF
DTTIARKEYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD*GGSGGSGGSGGSGGSGGSG*
*GM*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF
GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFG
NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDA
ILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGY
AGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGEL
HAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE
EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF
LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLK
IIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGW
GRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDS
LHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRE
RMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDN
LTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLK
SKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVR
KMIAKSEQEGADKRTADGSEFESPKKKRKV*

[0106] 在一些实施例中,所述ABE8选自下表7或9的碱基编辑器。在一些实施例中,所述ABE8含有从TadA进化而来的腺苷脱氨酶变体。在一些实施例中,所述ABE8的腺苷脱氨酶变体是如下表7或9中所述的TadA*8变体。在一些实施例中,所述腺苷脱氨酶变体是包含一种或多种选自Y147T、Y147R、Q154S、Y123H、V82S、T166R和/或Q154R的改变的TadA*7.10变体(例如TadA*8)。在各种实施例中,所述ABE8包含TadA*7.10变体(例如TadA*8),以及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;和I76Y+V82S+Y123H+Y147R+Q154R。在一些实施例中,ABE8

是单体构建体。在一些实施例中,ABE8是异源二聚体构建体。在一些实施例中,所述腺苷脱氨酶碱基编辑器8(ABE8)包含以下序列:MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD.

[0107] 在一些实施例中,所述多核苷酸可编程DNA结合结构域是CRISPR相关(例如Cas或Cpf1)酶。在一些实施例中,所述碱基编辑器是与脱氨酶结构域融合的催化死亡的Cas9(dCas9)。在一些实施例中,所述碱基编辑器是与脱氨酶结构域融合的Cas9切口酶(nCas9)。所述碱基编辑器的细节在国际PCT申请号PCT/2017/045381(WO 2018/027078)和PCT/US2016/058344(WO 2017/070632)中进行了描述,它们各自通过引用整体并入本文。另见Komor,A.C.等人,"Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage"Nature 533,420-424(2016);Gaudelli,N.M.等人,"Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage"Nature 551,464-471(2017);Komor,A.C.等人,"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)和Rees,H.A.等人,"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,其全部内容通过引用并入本文。

[0108] 举例来说,本文描述的碱基编辑组成物、系统和方法中使用的腺嘌呤碱基编辑器(ABE)具有如下核酸序列(8877个碱基对),(Addgene,Watertown,MA.;Gaudelli NM等人,Nature.2017Nov23;551(7681):464-471.doi:10.1038/nature24644;Koblan LW等人,Nat Biotechnol.20180ct;36(9):843-846.doi:8/10 4172)。还包括与ABE核酸序列具有至少95%或更高同一性的多核苷酸序列。

[0109] ATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTAGAGATCCGCGGCCGCTAATACGACTCACTATAGGGAGAGCCGCCACCATGAAACGGACAGCCGACGGAAGCGAGTTCGAGTCACCAAAGAAGAAGCGGAAAGTCTCTGAAGTCGAGTTTAGCCACGAGTATTGGATGAGGCACGCACTGACCCTGGCAAAGCGAGCATGGGATGAAAGAGAAGTCCCCGTGGGCGCCGTGCTGGTGCACAACAATAGAGTGATCGGAGAGGGATGGAACAGGCCAATCGGCCGCCACGACCCTACCGCACACGCAGAGATCATGGCACTGAGGCAGGGAGGCCTGGTCATGCAGAATTACCGCCTGATCGATGCCACCCTGTATGTGACACTGGAGCCATGCGTGATGTGCGCAGGAGCAATGATCCACAGCAGGATCGGAAGAGTGGTGTTCGGAGCACGGGACGCCAAGACCGGCGCAGCAGGCTCCCTGATGGATGTGCTGCACCACCCCGGCATGAACCACCGGGTGGAGATCACAGAGGGAATCCTGGCAGACGAGTGCGCCGCCCTGCTGAGCGATTTCTTTAGAATGCGGAGACAGGAGATCAAGGCCCAGAAGAAGGCACAGAGCTCCACCGACTCTGGAGGATCTAGCGGAGGATCCTCTGGAAGCGAGACACCAGGCACAAGCGAGTCCGCCACACCAGAGAGCTCCGGCGGCTCCTCCGGAGGATCCTCTGAGGTGGAGTTTTCCCACGAGTACTGGATGAGACATGCCCTGACCCTGGCCAAGAGGGCACGCGATGAGAGGGAGGTGCCTGTGGGAGCCGTGCTGGTGCTGAACAATAGAGTGATCGGCGAGGGCTGGAACAGAGCCATCGGCCTGCACGACCCAACAGCC

CATGCCGAAATTATGGCCCTGAGACAGGGCGGCCTGGTCATGCAGAACTACAGACTGATTGACGCCACCCTGTACGT
GACATTCGAGCCTTGCGTGATGTGCGCCGGCGCCATGATCCACTCTAGGATCGGCCGCGTGGTGTTTGGCGTGAGGA
ACGCAAAAACCGGCGCCGCAGGCTCCCTGATGGACGTGCTGCACTACCCCGGCATGAATCACCGCGTCGAAATTACC
GAGGGAATCCTGGCAGATGAATGTGCCGCCCTGCTGTGCTATTTCTTTCGGATGCCTAGACAGGTGTTCAATGCTCA
GAAGAAGGCCCAGAGCTCCACCGACTCCGGAGGATCTAGCGGAGGCTCCTCTGGCTCTGAGACACCTGGCACAAGCG
AGAGCGCAACACCTGAAAGCAGCGGGGGCAGCAGCGGGGGGTCAGACAAGAAGTACAGCATCGGCCTGGCCATCGGC
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
TCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATC
TCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACACTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCA

TGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATAT
GAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATC
CATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGA
TACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGT
GGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAA
TAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCT
CCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCG
ATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCAT
GCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGA
GTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAA
CGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAA
CTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGG
GAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTAT
TGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAA
AGTGCCACCTGACGTCGACGGATCGGGAGATCGATCTCCCGATCCCCTAGGGTCGACTCTCAGTACAATCTGCTCTG
ATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAA
GCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGA
TGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTC
ATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGC
CCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTA
TTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATC

[0110] “碱基编辑活性”是指用于化学改变多核苷酸内的碱基。在一个实施例中,第一个碱基被转化为第二个碱基。在一个实施例中,碱基编辑活性是胞苷脱氨酶活性,例如将目标C·G转化为T·A。在另一个实施例中,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。在另一个实施例中,碱基编辑活性是胞苷脱氨酶活性,例如将目标C·G转化为T·A和腺苷或腺嘌呤脱氨酶活性,例如将A·T转化为G·C。在一些实施例中,碱基编辑活性通过编辑效率来评估。碱基编辑效率可以通过任何合适的方式来测量,例如,通过桑格(Sanger)测序或下一代测序。在一些实施例中,碱基编辑效率通过受碱基编辑器影响的核碱基转化的总测序读数百分比来测量,例如,转化为G.C碱基对的目标A.T碱基对的总测序读数百分比。在一些实施例中,当在细胞群中进行碱基编辑时,碱基编辑效率通过受碱基编辑器影响的核碱基转化的总细胞百分比来测量。

[0111] 术语“碱基编辑器系统”是指用于编辑目标核苷酸序列的核碱基的系统。在各种实施例中,所述碱基编辑器系统包含(1)多核苷酸可编程核苷酸结合结构域(例如Cas9);(2)用于使所述核碱基脱氨基的脱氨酶结构域(例如腺苷脱氨酶);(3)一种或多种向导多核苷酸(例如向导RNA)。在一些实施例中,所述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。在一些实施例中,所述碱基编辑器是腺嘌呤或腺苷碱基编辑器(ABE)。在一些实施例中,所述碱基编辑器系统是ABE8。

[0112] 在一些实施例中,碱基编辑器系统可以包括不止一个碱基编辑组分。例如,碱基编辑器系统可包括不止一种脱氨酶。在一些实施例中,碱基编辑器系统可包括一种或多种腺

苷脱氨酶。在一些实施例中,可利用单向导多核苷酸将不同的脱氨酶靶向目标核酸序列。在一些实施例中,单对向导多核苷酸可用于将不同脱氨酶靶向目标核酸序列。

[0113] 所述碱基编辑器系统的脱氨酶结构域和多核苷酸可编程核苷酸结合组分可以共价或非共价地缔合,或通过其缔合和相互作用的任何组合。例如,在一些实施例中,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靶向目标核苷酸序列。在一些实施例中,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。在一些实施例中,所述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相互作用或缔合,将脱氨酶结构域靶向目标核苷酸序列。例如,在一些实施例中,所述脱氨酶结构域可包含额外异源部分或结构域,其能够与作为多核苷酸可编程核苷酸结合结构域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。在一些实施例中,所述额外异源部分可能能够与多肽结合、产生相互作用、缔合或形成复合物。在一些实施例中,所述额外异源部分可能能够与多核苷酸结合、相互作用、缔合或形成复合物。在一些实施例中,所述额外异源部分可能能够结合向导多核苷酸。在一些实施例中,所述额外异源部分可能能够结合多肽接头。在一些实施例中,所述额外异源部分可能能够结合多核苷酸接头。所述额外的异源部分可以是蛋白质结构域。在一些实施例中,所述额外的异源部分可以是K同源(KH)结构域、MS2外壳蛋白结构域、PP7外壳蛋白结构域、SfMu Com外壳蛋白结构域、无菌α基序(sterile alpha motif)、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。

[0114] 碱基编辑器系统可进一步包含向导多核苷酸组分。应当理解,所述碱基编辑器系统的组分可以共价或非共价地缔合,或通过其缔合和相互作用的组合。在一些实施例中,脱氨酶结构域可以通过向导多核苷酸靶向目标核苷酸序列。例如,在一些实施例中,所述脱氨酶结构域可包含额外的异源部分或结构域(例如多核苷酸结合结构域,如RNA或DNA结合蛋白),其能够与向导多核苷酸的一部分或区段(例如多核苷酸基序)相互作用、缔合或形成复合物。在一些实施例中,所述额外的异源部分或结构域(例如多核苷酸结合结构域,如RNA或DNA结合蛋白)可以与脱氨酶结构域融合或连接。在一些实施例中,所述额外的异源部分可能能够与多肽结合、相互作用、缔合或形成复合物。在一些实施例中,所述额外的异源部分可能能够与多核苷酸结合、相互作用、缔合或形成复合物。在一些实施例中,所述额外的异源部分可能能够结合向导多核苷酸。在一些实施例中,所述额外的异源部分能够结合多肽接头。在一些实施例中,所述额外异源部分能够结合多核苷酸接头。所述额外异源部分可以是蛋白质结构域。在一些实施例中,所述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 Com外壳蛋白结构域、无菌α基序(sterile alpha motif)、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。

[0115] 在一些实施例中,碱基编辑器系统可进一步包含碱基切除修复(BER)组分的抑制剂。应当理解,碱基编辑器系统的组分可以通过共价、非共价相互作用或其缔合和相互作用的任何组合彼此缔合。BER组分的抑制剂可以包括BER抑制剂。在一些实施例中,所述BER抑制剂可以是尿嘧啶DNA糖基化酶抑制剂(UGI)。在一些实施例中,BER抑制剂可以是肌苷BER抑制剂。在一些实施例中,所述BER抑制剂可以通过多核苷酸可编程核苷酸结合结构域靶向目标核苷酸序列。在一些实施例中,所述多核苷酸可编程核苷酸结合结构域可以与BER抑制剂融合或连接。在一些实施例中,所述多核苷酸可编程核苷酸结合结构域可以与脱氨酶结构域和BER抑制剂融合或连接。在一些实施例中,所述多核苷酸可编程核苷酸结合结构域可

以通过与BER抑制剂非共价相互作用或缔合,来将BER抑制剂靶向目标核苷酸序列。例如,在一些实施例中,所述BER组分的抑制剂可包含额外的异源部分或结构域,其能够与作为多核苷酸可编程核苷酸结合结构域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。

[0116] 在一些实施例中,所述BER抑制剂可以通过向导多核苷酸靶向目标核苷酸序列。例如,在一些实施例中,所述BER抑制剂可包含额外异源部分或结构域(例如多核苷酸结合结构域,如RNA或DNA结合蛋白),其能够与向导多核苷酸的一部分或区段(例如多核苷酸基序)相互作用、缔合或形成复合物。在一些实施例中,所述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,如RNA或DNA结合蛋白)可以与BER抑制剂融合或连接。在一些实施例中,所述额外异源部分可能能够与多核苷酸结合、相互作用、缔合或形成复合物。在一些实施例中,所述额外异源部分可能能够结合向导多核苷酸。在一些实施例中,所述额外异源部分能够结合多肽接头。在一些实施例中,所述额外异源部分能够结合多核苷酸接头。所述额外异源部分可以是蛋白质结构域。在一些实施例中,所述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 Com外壳蛋白结构域、无菌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。

[0117] 术语“Cas9”或“Cas9结构域”是指包含Cas9蛋白或其片段(例如包含Cas9的活性、无活性或部分活性DNA切割结构域,和/或gRNA结合结构域的一种蛋白质)的RNA向导核酸酶。Cas9核酸酶有时也称为Casnl核酸酶或CRISPR(成簇的规则间隔短回文重复序列)相关核酸酶。CRISPR是一种适应性免疫系统,可提供针对移动遗传组分(病毒、转座组分和接合质粒)的保护。CRISPR簇包含间隔区、与前述移动组分互补的序列和目标入侵核酸。CRISPR簇被转录并加工成CRISPR RNA(crRNA)。在II型CRISPR系统中,正确处理pre-crRNA需要转编码的小RNA(tracrRNA)、内源性核糖核酸酶3(rnc)和Cas9蛋白。tracrRNA作为核糖核酸酶3辅助处理pre-crRNA的指引。随后,Cas9/crRNA/tracrRNA核酸内切切割与间隔区互补的线性或环状dsDNA目标。与crRNA不互补的目标链首先通过核酸内切方式切割,然后通过核酸外切方式修剪3′-5’。在自然界中,DNA结合和切割通常需要蛋白质和两种RNA。然而,可以对单向导RNA(“sgRNA”,或简称为“gRNA”)进行工程改造,以便将crRNA和tracrRNA的各个方面整合到单一RNA种类中。见例如Jinek M.,Chylinski K.,Fonfara I.,Hauer M.,Doudna J.A.,Charpentier E.Science 337:816-821(2012),其全部内容通过引用并入本文。Cas9识别CRISPR重复序列(PAM或原型间隔区相邻基序)中的短基序,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众所周知的(见例如“Complete genome sequence of an M1 strain of Streptococcus pyogenes.”Ferretti等人,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.Science337:816-821(2012),其全部内容通过引用并入本文)。Cas9直向同源物已在各种物种中得到描述,包括但不限于化脓链球菌和嗜热链球菌。基于本公开内容,其他合适的Cas9核酸酶和序列对于本领域技术人员来说将是显而易见的,并且此类Cas9核酸酶和序列包括来自以下公开的生物体和基因座的Cas9序列:Chylinski、Rhun和Charpentier,"The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems"(2013)RNA Biology 10:5,726-737;其全部内容通过引用并入本文。

[0118] 一个Cas9示例是化脓性链球菌Cas9(spCas9),其氨基酸序列提供如下:

[0119]
MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFG
NIVDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD
VDKLFIQLVQIYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLFGN
LIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE
VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL
SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDRGMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGHSL
HEQIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQTTQKGQKNSRERM
KRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHI
VPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT
KAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSK
LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKM
IAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFA
TVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAY
SVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKY
SLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ
HKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAP
AAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0120]

[0121] (单下划线:HNH结构域;双下划线:RuvC结构域)

[0122] 核酸酶无活性的Cas9蛋白可互换地称为"dCas9"蛋白(对于核酸酶-"死的"Cas9)或无催化活性的Cas9。用于产生具有无活性DNA切割结构域的Cas9蛋白(或其片段)的方法

是已知的(见例如Jinek等人,Science.337:816-821(2012);Qi等人,“Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression”(2013)Cell.28;152(5):1173-83,其全部内容通过引用并入本文)。

[0123] 例如,已知Cas9的DNA切割结构域包括两个子结构域,HNH核酸酶子结构域和RuvC1子结构域。HNH子结构域切割与gRNA互补的链,而RuvC1子结构域切割非互补链。这些子结构域内的突变可以使Cas9的核酸酶活性沉默。例如,突变D10A和H840A使化脓性链球菌Cas9的核酸酶活性完全失活(Jinek等人,Science.337:816-821(2012);Qi等人,Cell.28;152(5):1173-83(2013))。在一些实施例中,Cas9核酸酶具有无活性(例如失活的)DNA切割结构域,即,Cas9是切口酶,称为“nCas9”蛋白(对于“切口酶”Cas9)。在一些实施例中,提供了包含Cas9片段的蛋白质。例如,在一些实施例中,蛋白质包含两个Cas9结构域之一:(1)Cas9的gRNA结合结构域;或(2)Cas9的DNA切割结构域。在一些实施例中,包含Cas9或其片段的蛋白质被称为“Cas9变体”。Cas9变体与Cas9或其片段具有同源性。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、在至少约98%相同、至少约99%相同、至少约99.5%相同或至少约99.9%相同。在一些实施例中,Cas9变体与野生型Cas9相比,可具有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、50或更多的氨基酸变化。在一些实施例中,Cas9变体包含Cas9的片段(例如gRNA结合结构域或DNA切割结构域),使得所述片段与野生型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。在一些实施例中,所述片段与相应野生型Cas9的氨基酸长度至少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%相同、至少96%、至少97%、至少98%、至少99%或至少99.5%相同。

[0124] 在一些实施例中,所述片段的长度为至少100个氨基酸。在一些实施例中,所述片段是至少100、150、200、250、300、350、400、450、500、550、600、650、700、750、800、850、900、1050、1100、1150、1200、1250或至少1300个氨基酸的长度。

[0125] 在一些实施例中,野生型Cas9对应于来自化脓性链球菌的Cas9(NCBI参考序列:NC_017053.1,核苷酸和氨基酸序列如下)。

[0126] ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGATCACTGATGATTATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTATCAAAAAAAATCTTATAGGGGCTCTTTTATTTGGCAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGTCTTTTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAAGTTGCTTATCATGAGAAATATCCAACTATCTATCATCTGCGAAAAAAATTGGCAGATTCTACTGATAAAGCGGATTTGCGCTTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAGATTTAAATCCTGATAATAGTGATGTGGACAAACTATTTATCCAGTTGGTACAAATCTACAATCAATTATTTGAAGAAAACCCTATTAACGCAAGTAGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGCTCAGCTCCCCGGTGAGAAGAGAAATGGCTTGTTTGGGAATCTCATTGCTTTGTCATTGGGATTGACCCCTAATTTTAAATCA

AATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTAGATAATTTATTGGC
GCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGATGCTATTTTACTTTCAGATATCC
TAAGAGTAAATAGTGAAATAACTAAGGCTCCCCTATCAGCTTCAATGATTAAGCGCTACGATGAACATCATCAAGAC
TTGACTCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTATAAAGAAATCTTTTTTGATCAATCAAAAAA
CGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAAATTTATCAAACCAATTTTAGAAAAAA
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

MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFG
NIVDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD
VDKLFIQLVQIYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLFGN
LIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE
VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL
SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDRGMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
[0127] RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGHSL
HEQIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQTTQKGQKNSRERM
KRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHI
VPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT
KAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSK
LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKM
IAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFA
TVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAY
SVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKY
SLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ
HKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAP
AAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0128] (单下划线:HNH结构域;双下划线:RuvC结构域)

[0129] 在一些实施例中,野生型Cas9对应于或包含以下核苷酸和/或氨基酸序列:

[0130] 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

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFG
NIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD
VDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGN
LIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE
VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL
SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
[0131] RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL
HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER
MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDH
IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK
MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF
ATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVA
YSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE
QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA
PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0132] (单下划线:HNH结构域;双下划线:RuvC结构域)

[0133] 在一些实施例中,野生型Cas9对应于来自化脓性链球菌的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)。

[0134] ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGATCACTGAT
GAATATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTATCAAAAAAAATCTTATAGG
GGCTCTTTTATTTGACAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTC
GGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGA
CTTGAAGAGTCTTTTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAAGT
TGCTTATCATGAGAAATATCCAACTATCTATCATCTGCGAAAAAAAATTGGTAGATTCTACTGATAAAGCGGATTTGC
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

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFG
NIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD
VDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGN
LIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE
VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL
SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
[0135] RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL
HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER
MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDH
IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK
MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF
ATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVA
YSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE
QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA
PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO:1.

单下划线:HNH结构域;双下划线:RuvC结构域)

[0136] 在一些实施例中,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);栖蚜蝇螺原体(Spiroplasma syrphidicola)(NCBI Ref:NC_021284.1);中间普氏菌(Prevotella intermedia)(NCBI Ref:NC_017861.1);中国台湾螺原体(Spiroplasma taiwanense,China)(NCBI Ref:NC_021846.1);链球菌(Streptococcus iniae)(NCBI Ref:NC_021314.1);波罗的海贝尔氏菌(Belliella baltica)(NCBI Ref:NC_018010.1);旋螺冷弯曲菌I(Psychroflexus TorquisI)(NCBI Ref:NC_018721.1);嗜热链球菌(Streptococcus thermophilus)(NCBI Ref:YP_820832.1)、无害李斯特菌(Listeria innocua)(NCBI Ref:NP_472073.1)、空肠弯曲菌(Campylobacter jejuni)(NCBI Ref:YP_002344900.1)或脑膜炎奈瑟菌(Neisseria meningitidis)(NCBI Ref.:YP_002342100.1)或来自任何其他生物体的Cas9。

[0137] 在一些实施例中,Cas9来自脑膜炎奈瑟菌(Nme)。在一些实施例中,Cas9是Nme1、Nme2或Nme3。在一些实施例中,Nme1、Nme2或Nme3的PAM相互作用结构域分别是N_4GAT、N_4CC和N_4CAAA(见例如Edraki,A.等人,A Compact,High-Accuracy Cas9 with a Dinucleotide PAM for In Vivo Genome Editing,Molecular Cell(2018))。一个脑膜炎奈瑟菌Cas9蛋白示例,Nme1Cas9,(NCBI参考文献:WP_002235162.1;II型CRISPR RNA引导的核酸内切酶Cas9)具有以下氨基酸序列:

[0138]

```
   1 maafkpnpin yilgldigia svgwamveid edenpiclid lgvrvferae vpktgdslam
  61 arrlarsvrr ltrrrahrll rarrllkreg vlqaadfden glikslpntp wqlraaaldr
 121 kltplewsav llhlikhrgy lsqrkneget adkelgallk gvadnahalq tgdfrtpael
 181 alnkfekesg hirnqrgdys htfsrkdlqa elillfekqk efgnphvsgg lkegietllm
 241 tqrpalsgda vqkmlghctf epaepkaakn tytaerfiwl tklnnlrile qgserpltdt
 301 eratlmdepy rkskltyaqa rkllgledta ffkglrygkd naeastlmem kayhaisral
 361 ekeglkdkks plnlspelqd eigtafslfk tdeditgrlk driqpeilea llkhisfdkf
 421 vqislkalrr ivplmeqgkr ydeacaeiyg dhygkkntee kiylppipad eirnpvvlra
 481 lsqarkving vvrrygspar ihietarevg ksfkdrkeie krqeenrkdr ekaaakfrey
 541 fpnfvgepks kdilklrlye qqhgkclysg keinlgrlne kgyveidhal pfsrtwddsf
 601 nnkvlvlgse nqnkgnqtpy eyfngkdnsr ewqefkarve tsrfprskkq rillqkfded
 661 gfkernlndt ryvnrflcqf vadrmrltgk gkkrvfasng qitnllrgfw glrkvraend
 721 rhhaldavvv acstvamqqk itrfvrykem nafdgktidk etgevlhqkt hfpqpweffa
 781 qevmirvfgk pdgkpefeea dtpeklrtll aeklssrpea vheyvtplfv srapnrkmsg
 841 qghmetvksa krldegvsvl rvpltqlklk dlekmvnrer epklyealka rleahkddpa
 901 kafaepfyky dkagnrtqqv kavrveqvqk tgvwvrnhng iadnatmvrv dvfekgdkyy
 961 lvpiyswqva kgilpdravv qgkdeedwql iddsfnfkfs lhpndlvevi tkkarmfgyf
1021 aschrgtgni nirihdldhk igkngilegi gvktalsfqk yqidelgkei rpcrlkkrpp
1081 vr
```

[0139]

[0140] 另一个脑膜炎奈瑟菌Cas9蛋白示例,Nme2Cas9,(NCBI参考文献:WP_002230835;II

型CRISPR RNA引导的核酸内切酶Cas9)具有以下氨基酸序列：

[0141]
```
   1 maafkpnpin yilgldigia svgwamveid eeenpirlid lgvrvferae vpktgdslam
  61 arrlarsvrr ltrrrahrll rarrllkreg vlqaadfden glikslpntp wqlraaaldr
 121 kltplewsav llhlikhrgy lsqrkneget adkelgallk gvannahalq tgdfrtpael
 181 alnkfekesg hirnqrgdys htfsrkdlqa elillfekqk efgnphvsgg lkegietllm
 241 tqrpalsgda vqkmlghctf epaepkaakn tytaerfiwl tklnnlrile qgserpltdt
 301 eratlmdepy rkskltyaqa rkllgledta ffkglrygkd naeastlmem kayhaisral
 361 ekeglkdkks plnlsselqd eigtafslfk tdeditgrlk drvqpeilea llkhisfdkf
 421 vqislkalrr ivplmeqgkr ydeacaeiyg dhygkkntee kiylppipad eirnpvvlra
 481 lsqarkving vvrrygspar ihietarevg ksfkdrkeie krqeenrkdr ekaaakfrey
 541 fpnfvgepks kdilklrlye qqhgkclysg keinlvrlne kgyveidhal pfsrtwddsf
 601 nnkvlvlgse nqnkgnqtpy eyfngkdnsr ewqefkarve tsrfprskkq rillqkfded
 661 gfkecnlndt ryvnrflcqf vadhilltgk gkrrvfasng qitnllrgfw glrkvraend
 721 rhhaldavvv acstvamqqk itrfvrykem nafdgktidk etgkvlhqkt hfpqpweffa
 781 qevmirvfgk pdgkpefeea dtpeklrtll aeklssrpea vheyvtplfv srapnrkmsg
 841 ahkdtlrsak rfvkhnekis vkrvwlteik ladlenmvny kngreielye alkarleayg
 901 gnakqafdpk dnpfykkggq lvkavrvekt qesgvllnkk naytiadngd mvrvdvfckv
 961 dkkgknqyfi vpiyawqvae nilpdidckg yriddsytfc fslhkydlia fqkdekskve
1021 fayyincdss ngrfylawhd kgskeqqfri stqnlvliqk yqvnelgkei rpcrlkkrpp
1081 vr
```

[0142] 在一些实施例中,dCas9对应于,或部分或全部包含具有一个或多个使Cas9核酸酶活性失活的突变的Cas9氨基酸序列。例如,在一些实施例中,dCas9结构域包含D10A和H840A突变或另一个Cas9中的相应突变。在一些实施例中,dCas9包含dCas9(D10A和H840A)的氨基酸序列：

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFG
NIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD
VDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGN
LIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE
VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL
SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
[0143] RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL
HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER
MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDA
IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK
MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF
ATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVA
YSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE
QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA
PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0144] (单下划线:HNH结构域;双下划线:RuvC结构域)。

[0145] 在一些实施例中,Cas9结构域包含D10A突变,而840位的残基在上面提供的氨基酸序列中或在本文提供的任何氨基酸序列中的相应位置处仍为组氨酸。

[0146] 在其他实施例中,提供了具有除D10A和H840A之外的突变的dCas9变体,其例如导致核酸酶失活的Cas9(dCas9)。例如,此类突变包括在D10和H840处的其他氨基酸置换,或Cas9核酸酶结构域内的其他置换(例如HNH核酸酶子结构域和/或RuvC1子结构域中的置换)。在一些实施例中,提供了dCas9的变体或同源物,其至少约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。在一些实施例中,提供了dCas9的变体,其具有以下较短或较长的氨基酸序列: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。

[0147] 在一些实施例中,本文提供的Cas9融合蛋白包含Cas9蛋白的全长氨基酸序列,例如本文提供的Cas9序列之一。然而,在其他实施例中,本文提供的融合蛋白不包含全长Cas9序列,而仅包含其一个或多个片段。本文提供了合适的Cas9结构域和Cas9片段的示例氨基酸序列,并且额外的Cas9结构域和片段的合适序列对本领域技术人员来说是显而易见的。

[0148] 应当理解,额外的Cas9蛋白(例如核酸酶死亡Cas9(dCas9)、Cas9切口酶(nCas9)或核酸酶活性Cas9),包括其变体和同系物,在本公开的范围内。示例Cas9蛋白包括但不限于以下提供的那些。在一些实施例中,Cas9蛋白是核酸酶死亡的Cas9(dCas9)。在一些实施例中,Cas9蛋白是Cas9切口酶(nCas9)。在一些实施例中,Cas9蛋白是核酸酶活性Cas9。

[0149] 示例性无催化活性的Cas9(dCas9):

[0150] DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR
RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTD
KADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLEN
LIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAIL
LSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPL
ARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDN
EENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSD
GFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEM
ARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDA
IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFI
KRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG
TALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE
IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK
VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE
LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0151] 示例性催化Cas9切口酶(nCas9):

[0152] DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR
RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTD
KADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLEN
LIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAIL
LSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPL
ARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDN
EENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSD
GFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEM

ARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDH
IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFI
KRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG
TALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE
IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK
VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE
LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0153] 示例性催化活性Cas9:

[0154] DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDS
TDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRR
LENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY
KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIP
YYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELT
KVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKII
KDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGK
TILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGR
HKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELD
INRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAER
GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNY
HHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI
RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGG
FDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG
RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA
NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRID
LSQLGGD.

[0155] 在一些实施例中,Cas9是指来自古细菌(例如纳米古细菌)的Cas9,其构成单细胞原核微生物的域和界。在一些实施例中,Cas9是指CasX或CasY,其已在例如Burstein等人的“New CRISPR-Cas system from uncultivated microbes”Cell Res.2017Feb 21.doi:10.1038/cr.2017.21中描述,其全部内容通过引用并入本文。使用基因组解析的宏基因组学,确定了许多CRISPR-Cas系统,包括在古菌域首次报道的Cas9。作为活性CRISPR-Cas系统的一部分,这种各有差异的Cas9蛋白在很少被研究的纳米古细菌中发现。在细菌中,发现了两个以前未知的系统,CRISPR-CasX和CRISPR-CasY,它们是迄今为止发现的最紧凑的系统之一。在一些实施例中,Cas9是指CasX,或CasX的变体。在一些实施例中,Cas9是指CasY或CasY的变体。应当理解,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),并且在本公开的范围内。

[0156] 在特定实施例中,可用于本发明的方法中的napDNAbps包括本领域已知的,并且例

如由Oakes等人于Cell 176,254-267,2019描述的环状排列。示例性环状排列如下,粗体序列表示源自Cas9的序列,斜体序列表示接头序列,下划线序列表示双分核定位序列。

[0157] CP5(带有MSP“NGG=带有突变常规Cas9如NGG的Pam变体”PID=蛋白质相互作用结构域和“D10A”切口酶)。

[0158]

EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELE
NGRKRMLASAKFLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD
EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAFKYF
DTTIARKEYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD*GGSGGSGGSGGSGGSGGSG*
***GM*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA**
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF
GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFG
NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDA
ILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGY
AGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGEL
HAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE
EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF
LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLK
IIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGW
GRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDS
LHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRE
RMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDN
LTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLK
SKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVR
KMIAKSEQEGADKRTADGSEFESPKKKRKV*

[0160] 可并入碱基编辑器的多核苷酸可编程核苷酸结合结构域的非限制性实例包括CRISPR蛋白衍生结构域、限制性核酸酶、大范围核酸酶、TAL核酸酶(TALEN)和锌指核酸酶(ZFN)。

[0161] 在一些实施例中,本文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。在一些实施例中,所述napDNAbp是CasX蛋白。在一些实

施例中,所述napDNAbp是CasY蛋白。在一些实施例中,所述napDNAbp包含的氨基酸序列与天然存在的CasX或CasY蛋白至少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。在一些实施例中,所述napDNAbp是天然存在的CasX或CasY蛋白。在一些实施例中,所述napDNAbp包含的氨基酸序列与本文所述的任何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。应当理解,根据本公开也可以使用来自其他细菌种类的Cas12b/C2c1、CasX和CasY。

[0162] Cas12b/C2c1(uniprot.org/uniprot/T0D7A2#2)

[0163] sp|T0D7A2|C2C1_ALIAG CRISPR相关内切核酸酶C2c1 OS=脂环酸杆菌(Alicyclobacillus acido-terrestris)(菌株ATCC 49025/DSM 3922/CIP 106132/NCIMB 13137/GD3B)GN=c2c1PE=1SV=1

[0164] MAVKSIKVKLRLDDMPEIRAGLWKLHKEVNAGVRYYTEWLSLLRQENLYRRSPNGDGEQECDKTAEEC
KAELLERLRARQVENGHRGPAGSDDELLQLARQLYELLVPQAIGAKGDAQQIARKFLSPLADKDAVGGLGIAKAGN
KPRWVRMREAGEPGWEEEKEKAETRKSADRTADVLRALADFGLKPLMRVYTDSEMSSVEWKPLRKGQAVRTWDRDM
FQQAIERMMSWESWNQRVGQEYAKLVEQKNRFEQKNFVGQEHLVHLVNQLQQDMKEASPGLESKEQTAHYVTGRAL
RGSDKVFEKWGKLAPDAPFDLYDAEIKNVQRRNTRRFGSHDLFAKLAEPEYQALWREDASFLTRYAVYNSILRKLN
HAKMFATFTLPDATAHPIWTRFDKLGGNLHQYTFLFNEFGERRHAIRFHKLLKVENGVAREVDDVTVPISMSEQLD
NLLPRDPNEPIALYFRDYGAEQHFTGEFGGAKIQCRRDQLAHMHRRRGARDVYLNVSVRVQSQSEARGERRPPYAA
VFRLVGDNHRAFVHFDKLSDYLAEHPDDGKLGSEGLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNSKGRVPF
FFPIKGNDNLVAVHERSQLLKLPGETESKDLRAIREERQRTLRQLRTQLAYLRLLVRCGSEDVGRRERSWAKLIEQ
PVDAANHMTPDWREAFENELQKLKSLHGICSDKEWMDAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYAKDV
VGGNSIEQIEYLERQYKFLKSWSFFGKVSGQVIRAEKGSRFAITLREHIDHAKEDRLKKLADRIIMEALGYVYALD
ERGKGKWVAKYPPCQLILLEELSEYQFNNDRPPSENNQLMQWSHRGVFQELINQAQVHDLLVGTMYAAFSSRFDAR
TGAPGIRCRRVPARCTQEHNPEPFPWWLNKFVVEHTLDACPLRADDLIPTGEGEIFVSPFSAEEGDFHQIHADLNA
AQNLQQRLWSDFDISQIRLRCDWGEVDGELVLIPRLTGKRTADSYSNKVFYTNTGVTYYERERGKKRRKVFAQEKL
SEEEAELLVEADEAREKSVVLMRDPSGIINRGNWTRQKEFWSMV NQRIEGYLVKQIRSRVPLQDSACENTGDI

[0165] CasX

[0166] (uniprot.org/uniprot/F0NN87;uniprot.org/uniprot/F0NH53)

[0167] >tr|F0NN87|F0NN87_SULIH CRISPR相关Casx蛋白OS=冰岛硫化叶菌(Sulfolobus islandicus)(菌株HVE10/4)GN=SiH_0402PE=4SV=1

[0168] MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAERRGKAKKKKGE
EGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKECEEVSAPSFVKPEFYEFGRSPGMVERTRR
VKLEVEPHYLIIAAAGWVLTRLGKAKVSEGDYVGVNVFTPTRGILYSLIQNVNGIVPGIKPETAFGLWIARKVVSSV
TNPNVSVVRIYTISDAVGQNPTTINGGFSIDLTKLLEKRYLLSERLEAIARNALSISSNMRERYIVLANYIYEYLTG

[0169] SKRLEDLLYFANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG

[0170] >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

[0171] MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAERRGKAKKKKGE
EGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKECEEVSAPSFVKPEFYKFGRSPGMVERTRR

VKLEVEPHYLIMAAAGWVLTRLGKAKVSEGDYVGVNVFTPTRGILYSLIQNVNGIVPGIKPETAFGLWIARKVVSSV
TNPNVSVVSIYTISDAVGQNPTTINGGFSIDLTKLLEKRDLLSERLEAIARNALSISSNMRERYIVLANYIYEYLTG
SKRLEDLLYFANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG

[0172] δ变形杆菌(Deltaproteobacteria)CasX

[0173] MEKRINKIRKKLSADNATKPVSRSGPMKTLLVRVMTDDLKKRLEKRRKKPEVMPQVISNNAANNLRMLL
DDYTKMKEAILQVYWQEFKDDHVGLMCKFAQPASKKIDQNKLKPEMDEKGNLTTAGFACSQCGQPLFVYKLEQVSEK
GKAYTNYFGRCNVAEHEKLILLAQLKPVKDSDEAVTYSLGKFGQRALDFYSIHVTKESTHPVKPLAQIAGNRYASGP
VGKALSDACMGTIASFLSKYQDIIIEHQKVVKGNQKRLESLRELAGKENLEYPSVTLPPQPHTKEGVDAYNEVIARV
RMWVNLNLWQKLKLSRDDAKPLLRLKGFPSFPVVERRENEVDWWNTINEVKKLIDAKRDMGRVFWSGVTAEKRNTIL
EGYNYLPNENDHKKREGSLENPKKPAKRQFGDLLLYLEKKYAGDWGKVFDEAWERIDKKIAGLTSHIEREEARNAED
AQSKAVLTDWLRAKASFVLERLKEMDEKEFYACEIQLQKWYGDLRGNPFAVEAENRVVDISGFSIGSDGHSIQYRNL
LAWKYLENGKREFYLLMNYGKKGRIRFTDGTDIKKSGKWQGLLYGGGKAKVIDLTFDPDDEQLIILPLAFGTRQGRE
FIWNDLLSLETGLIKLANGRVIEKTIYNKKIGRDEPALFVALTFERREVVDPSNIKPVNLIGVARGENIPAVIALTD
PEGCPLPEFKDSSGGPTDILRIGEGYKEKQRAIQAAKEVEQRRAGGYSRKFASKSRNLADDMVRNSARDLFYHAVTH
DAVLVFANLSRGFGRQGKRTFMTERQYTKMEDWLTAKLAYEGLTSKTYLSKTLAQYTSKTCSNCGFTITYADMDVML
VRLKKTSDGWATTLNNKELKAEYQITYYNRYKRQTVEKELSAELDRLSEESGNNDISKWTKGRRDEALFLLKKRFSH
RPVQEQFVCLDCGHEVHAAEQAALNIARSWLFLNSNSTEFKSYKSGKQPFVGAWQAFYKRRLKEVWKPNA

[0174] CasY(ncbi.nlm.nih.gov/protein/APG80656.1)

[0175] >APG80656.1 CRISPR相关蛋白CasY[未培养的俭菌菌群细菌(Parcubacteria group bacterium)]

[0176] MSKRHPRISGVKGYRLHAQRLEYTGKSGAMRTIKYPLYSSPSGGRTVPREIVSAINDDYVGLYGLSNFD
DLYNAEKRNEEKVYSVLDFWYDCVQYGAVFSYTAPGLLKNVAEVRGGSYELTKTLKGSHLYDELQIDKVIKFLNKKE
ISRANGSLDKLKKDIIDCFKAEYRERHKDQCNKLADDIKNAKKDAGASLGERQKKLFRDFFGISEQSENDKPSFTNP
LNLTCCLLPFDTVNNNRNRGEVLFNKLKEYAQKLDKNEGSLEMWEYIGIGNSGTAFSNFLGEGFLGRLRENKITELK
KAMMDITDAWRGQEQEEELEKRLRILAALTIKLREPKFDNHWGGYRSDINGKLSSWLQNYINQTVKIKEDLKGHKKD
LKKAKEMINRFGESDTKEEAVVSSLLESIEKIVPDDSADDEKPDIPAIAIYRRFLSDGRLTLNRFVQREDVQEALIK
ERLEAEKKKKPKKRKKKSDAEDEKETIDFKELFPHLAKPLKLVPNFYGDSKRELYKKYKNAAIYTDALWKAVEKIYK
SAFSSSLKNSFFDTDFDKDFFIKRLQKIFSVYRRFNTDKWKPIVKNSFAPYCDIVSLAENEVLYKPKQSRSRKSAAI
DKNRVRLPSTENIAKAGIALARELSVAGFDWKDLLKKEEHEEYIDLIELHKTALALLLAVTETQLDISALDFVENGT
VKDFMKTRDGNLVLEGRFLEMFSQSIVFSELRGLAGLMSRKEFITRSAIQTMNGKQAELLYIPHEFQSAKITTPKEM
SRAFLDLAPAEFATSLEPESLSEKSLLKLKQMRYYPHYFGYELTRTGQGIDGGVAENALRLEKSPVKKREIKCKQYK
TLGRGQNKIVLYVRSSYYQTQFLEWFLHRPKNVQTDVAVSGSFLIDEKKVKTRWNYDALTVALEPVSGSERVFVSQP
FTIFPEKSAEEEGQRYLGIDIGEYGIAYTALEITGDSAKILDQNFISDPQLKTLREEVKGLKLDQRRGTFAMPSTKI
ARIRESLVHSLRNRIHHLALKHKAKIVYELEVSRFEEGKQKIKKVYATLKKADVYSEIDADKNLQTTVWGKLAVASE
ISASYTSQFCGACKKLWRAEMQVDETITTQELIGTVRVIKGGTLIDAIKDFMRPPIFDENDTPFPKYRDFCDKHHIS
KKMRGNSCLFICPFCRANADADIQASQTIALLRYVKEEKKVEDYFERFRKLKNIKVLGQMKKI

[0177] 术语“Cas12”或“Cas12结构域”是指包含Cas12蛋白或其片段的RNA向导核酸酶(例如包含Cas12的活性、无活性或部分活性的DNA切割结构域的蛋白质,和/或Cas12的gRNA结合结构域)。Cas12属于2类V型CRISPR/Cas系统。Cas12核酸酶有时也称为CRISPR(成簇的规

则间隔短回文重复序列)相关核酸酶。下面提供了示例性久氏芽孢杆菌(Bacillus hisashii)Cas 12b(BhCas12b)Cas 12结构域的序列:

[0178] MAPKKKRKVGIHGVPAAATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYEHHEQDPKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVEKKGEANQLSNKFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEEKKKWEEDKKKDPLAKILGKLAEYGLIPLFIPYTDSNEPIVKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWESWNLKVKEEYEKVEKEYKTLEERIKEDIQALKALEQYEKERQEQLLRDTLNTNEYRLSKRGLRGWREIIQKWLKMDENEPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYATFCEIDKKKKDAKQQATFTLADPINHPLWVRFEERSGSNLNKYRILTEQLHTEKLKKKLTVQLDRLIYPTESGGWEEKGKVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGTLGGARVQFDRDHLRRYPHKVESGNVGRIYFNMTVNIEPTESPVSKSLKIHRDDFPKVVNFKPKELTEWIKDSKGKKLKSGIESLEIGLRVMSIDLGQRQAAAASIFEVVDQKPDIEGKLFFPIKGTELYAVHRASFNIKLPGETLVKSREVLRKAREDNLKLMNQKLNFLRNVLHFQQFEDITEREKRVTKWISRQENSDVPLVYQDELIQIRELMYKPYKDWVAFLKQLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDEIDRTRKFLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHLNALKEDRLKKMANTIIMHALGYCYDVRKKKWQAKNPACQIILFEDLSNYNPYEERSRFENSKLMKWSRREIPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKLQDNRFFKNLQREGRLTLDKIAVLKEGDLYPDKGGEKFISLSKDRKCVTTHADINAAQNLQKRFWTRTHGFYKVYCKAYQVDGQTVYIPESKDQKQKIIEEFGEGYFILKDGVYEWVNAGKLKIKKGSSKQSSSELVDSDILKDSFDLASELKGEKLMLYRDPSGNVFPSDKWMAAGVFFGKLERILISKLTNQYSISTIEDDSSKQSMKRPAATKKAGQAKKKK.

[0179] 与BhCas12b氨基酸序列具有至少85％或更高同一性的氨基酸序列也可用于本发明的方法。

[0180] 术语“保守氨基酸置换”或“保守突变”是指一个氨基酸被具有共同特性的另一个氨基酸置换。定义单个氨基酸之间共同特性的一种功能方法是分析同源生物的相应蛋白质之间氨基酸变化的标准化频率(Schulz,G.E.和Schirmer,R.H.,Principles of Protein Structure,Springer-Verlag,New York(1979))。根据这样的分析,可以定义氨基酸组,其中组内的氨基酸优先相互交换,因此在它们对整体蛋白质结构的影响方面彼此最相似(Schulz,G.E.和Schirmer,R.H.,supra)。保守突变的非限制性实例包括氨基酸的氨基酸置换,例如精氨酸的赖氨酸和反之亦然,从而可以保持正电荷;谷氨酸对天冬氨酸,反之亦然,以保持负电荷;苏氨酸的丝氨酸,这样可以保持一个游离的-OH;和谷氨酰胺对天冬酰胺,这样可以保持游离的-NH_2。

[0181] 如本文可互换使用的术语“编码序列”或“蛋白质编码序列”是指编码蛋白质的多核苷酸片段。所述区域或序列在靠近5’端的地方有一个起始密码子,在靠近3’端的地方有一个终止密码子。编码序列也可称为开放阅读框。

[0182] 如本文所用,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。在一些实施例中,所述脱氨酶是腺苷脱氨酶,其催化腺嘌呤水解脱氨为次黄嘌呤。在一些实施例中,所述脱氨酶是腺苷脱氨酶,其催化腺苷或腺嘌呤(A)水解脱氨为肌苷(I)。在一些实施例中,所述脱氨酶或脱氨酶结构域分别催化腺苷或脱氧腺苷水解脱氨基为肌苷或脱氧肌苷的腺苷脱氨酶。在一些实施例中,所述腺苷脱氨酶催化脱氧核糖核酸(DNA)中腺苷的水解脱氨基。本文提供的腺苷脱氨酶(例如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,例如细菌。在一些实施例中,所述腺苷脱氨酶来自细菌,例如大肠杆菌

(Escherichia coli)、金黄色葡萄球菌(Staphylococcus aureus)、鼠伤寒沙门氏杆菌(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)。

[0183] 在一些实施例中,所述腺苷脱氨酶是TadA脱氨酶。在一些实施例中,所述TadA脱氨酶是TadA变体。在一些实施例中,所述TadA变体是TadA*8。在一些实施例中,所述脱氨酶或脱氨酶结构域是来自生物体例如人、黑猩猩、大猩猩、猴、牛、狗、大鼠或小鼠的天然存在的脱氨酶的变体。在一些实施例中,所述脱氨酶或脱氨酶结构域在自然界中不存在。例如,在一些实施例中,脱氨酶或脱氨酶结构域与天然存在的脱氨酶为至少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%相同。例如,国际PCT申请号PCT/2017/045381(WO 2018/027078)和PCT/US2016/058344(WO 2017/070632)中描述了脱氨酶结构域,它们各自通过引用整体并入本文。另见Komor,A.C.等人,"Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage"Nature 533,420-424(2016);Gaudelli,N.M.等人,"Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage"Nature 551,464-471(2017);Komor,A.C.等人,"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)),和Rees,H.A.,等人,"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,其全部内容通过引用并入本文。

[0184] "检测"是指鉴定待检测分析物的存在、不存在或量。在一个实施例中,检测多核苷酸或多肽中的序列改变。在另一个实施例中,检测插入缺失的存在。

[0185] "可检测标记"是指一种组成物,当与感兴趣的分子连接时,通过光谱、光化学、生物化学、免疫化学或化学手段使后者可检测。例如,有用的标记包括放射性同位素、磁珠、金属珠、胶体颗粒、荧光染料、电子致密试剂、酶(例如通常用于ELISA)、生物素、地高辛或半抗原。

[0186] "疾病"是指损害或干扰细胞、组织或器官的正常功能的任何病症或病症。疾病的一个例子包括糖原贮积病1型(也称为GSD1或冯吉尔克病)。在一些实施例中,GSD1是1a型(GSD1a)。

[0187] "有效量"是指相对于未治疗的患者改善疾病症状所需的量。用于实施本发明以治疗疾病的活性化合物的有效量取决于给药方式、个体的年龄、体重和一般健康状况。最终,主治医师或兽医将决定适当的量和剂量方案。这种量被称为"有效"量。在一个实施例中,有效量是足以在细胞(例如体外或体内细胞)中引入感兴趣的基因(例如G6PC)改变的本发明碱基编辑器的量。在一个实施例中,有效量是实现治疗效果(例如减少或控制GSD1a或其症状或病症)所需的碱基编辑器的量。这种治疗效果不需要足以改变个体、组织或器官的所有细胞中的G6PC,而只需改变存在于个体、组织或器官中约1%、5%、10%、25%、50%、75%或更多细胞中的G6PC。在一个实施例中,有效量足以改善GSD1a的一种或多种症状。

[0188] “片段”是指多肽或核酸分子的一部分。该部分包含参考核酸分子或多肽全长的至少约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。

[0189] “葡萄糖-6-磷酸酶(G6PC)多肽”是指与NCBI登录号AAA16222.1具有至少约95％氨基酸序列同一性的多肽或其片段。在特定实施例中,本发明提供了一种编辑包含与1a型糖原贮积病(GSD1a)相关的单核苷酸多态性(SNP)的G6PC多核苷酸的方法。在一个实施例中,与GSD1a相关的所述SNP处的A•T到G•C的改变将所述G6PC多肽中的谷氨酰胺(Q)改变为非谷氨酰胺(X)氨基酸。在另一个实施例中,与GSD1a相关的所述SNP处的A•T到G•C改变将所述G6PC多肽中的精氨酸(R)改变为非精氨酸(X)。在一个实施例中,与GSD1a相关的所述SNP导致在位置347处具有非谷氨酰胺(X)氨基酸或在位置83处具有非精氨酸(X)氨基酸的所述G6PC多肽的表达。在一个实施例中,所述碱基编辑器校正用非谷氨酰胺氨基酸(X)替换位置347处的谷氨酰胺。在另一个实施例中,所述碱基编辑器的校正用非精氨酸氨基酸(X)替换位置83处的精氨酸。

[0190] 在特定实施例中,G6PC包含一个或多个相对于以下参考序列的改变。在特定实施例中,与GSD1a相关的G6PC包含一个或多个选自Q347X和R83C的突变。下面提供了来自智人的示例性G6PC氨基酸序列:

```
1       MEEGMNVLHD FGIQSTHYLQ VNYQDSQDWF ILVSVIADLR NAFYVLFPIW FHLQEAVGIK
61       LLWVAVIGDW LNLVFKWILF GQRPYWWVLD TDYYSNTSVP LIKQFPVTCE TGPGSPSGHA
121 MGTAGVYYVM VTSTLSIFQG KIKPTYRFRC LNVILWLGFW AVQLNVCLSR IYLAAHFPHQ
181 VVAGVLSGIA VAETFSHIHS IYNASLKKYF LITFFLFSFA IGFYLLLKGL GVDLLWTLEK
241 AQRWCEQPEW VHIDTTPFAS LLKNLGTLFG LGLALNSSMY RESCKGKLSK WLPFRLSSIV
301 ASLVLLHVFD SLKPPSQVEL VFYVLSFCKS AVVPLASVSV IPYCLAQVLG QPHKKSL
```

[0191]

[0192] “葡萄糖-6-磷酸酶多核苷酸”是指编码G6PC多肽的多核苷酸。下面提供了来自智人的示例性G6PC核苷酸序列(GenBank:U01120.1):

1 ATAGCAGAGC AATCACCACC AAGCCTGGAA TAACTGCAAG GGCTCTGCTG ACATCTTCCT

61 GAGGTGCCAA GGAAATGAGG ATGGAGGAAG GAATGAATGT TCTCCATGAC TTTGGGATCC

121 AGTCAACACA TTACCTCCAG GTGAATTACC AAGACTCCCA GGACTGGTTC ATCTTGGTGT

181 CCGTGATCGC AGACCTCAGG AATGCCTTCT ACGTCCTCTT CCCCATCTGG TTCCATCTTC

241 AGGAAGCTGT GGGCATTAAA CTCCTTTGGG TAGCTGTGAT TGGAGACTGG CTCAACCTCG

301 TCTTTAAGTG GTAAGAACCA TATAGAGAGG AGATCAGCAA GAAAAGAGGC TGGCATTCGC

361 TCTCGCAATG TCTGTCCATC AGAAGTTGCT TTCCCCAGGC TATTCAGGAA GCCACGGGCT

421 ACTCATGCTT CCAACCCCTC TCTCTGACTT TGGATCATCT ACATAAAGGG GGAAGACAGA

481 AAAAATCCTA CCAGTGAGTT GAAAATACAG GAAAGCCTAT TTCATATGGG TTAAAGGGTA

541 GGACAGTTGA ATTTCGTGAA AAGTCTGAGT TATATAGGCT TTGAGCAAAG AGTTTTATTA

601 GTATGAAGCA GAAGAGGTAA CATAAAGAAA GATGTATGGG GCCAGGCATG GTGGCTCACA

661 CCTGTAATCC CAGCACTTTG GGAGGCCGAG GTGGGCGAAT CACTCCTGGG TGAACTCAGG

721 AGTTCAAGAC CAGCCTGGGC AACATGGCGA AACTCCATCT CTACAAAAAC ATTACGAAAA

[0193] 781 TTAGCTGGGC GTGTTGGTGC TGTAGTCCCA GCTACTCAGG AGGCTGAGGT GAGAGGCGGA

841 GGAGGTTGCA GTGAGTCAAG ATCATGCCAC TGCACTCCAG CCTGGGCAAC AGAGTAAGAC

901 CCTGTCTCAA AAAAAAAAAA AAGATAGATG ATGTATGCTG TATGAAAAAA GGAAACACAC

961 AGATGATTCA ACAGCCTGTT TTGTGGGGTA ATGAAAAGTC ACCCTGGGAA CTGGGCTCCA

1021 GCCCTCGTTC TGCCACCCAC CAACTACATG TCCTTGGCAA GTCATATCAA TTATCTGAGT

1081 TTCTGTTTTA TAATCTACAA ATAGGTTATC TCTGGCAGCT TAATAATAAT CAGGGTTAAC

1141 ATTTATTAAA CAGTGTGTGC CAGTCCATGT GCTATGTGCT TTTCTGTGAG GTAGTTACTG

1201 CTATTTACAG AAACAGTAGA TGCAGAGACC AAGGTGCTGA GTTAAATGAT TAGGCCAACA

1261 AGGTTAGTAC ATGCCGAGCC AGGATGGAAG CCCAGGTAGG CAGGCTGGCT TCCGCGGCAA

1321 TGCTCTTATG AACTATGTTA CGTCCAGTGC TGATAAACTG ACTCTCTGGG GAGCAGGGGA

1381 AAGCCCTGAG TTTAGCATTT GCCAATTTCT ATCACGTAAA CATTCCCATT CTGGCCACTT

1441 TCTTTCTTTC TTTCTTTTGT TTGTTTGTTT GAGATGGAGT CTCGCACTGT TGCCTGGCTG

1501 GAGTGCAATG GTGCAATCTC AGCTCACTGC AACCTCTGCC TCTCCGGTTC AAGTGATTCT

1561 CCTGCCTCAG CCTCCCAAGT AGCTGGGATT ACAGGTGCCC GCCACCATGC CCAGCTAATT

[0194]

1621 TTTTTTGTAT TTTTAGTAGA GACATGGTTT CACTATGTTG ACTAGGCTGG TCTCGAACTC

1681 CTGACCTCAT GATCTGCCTG CCTTGGCCTC CCTAAGTGCT AGGATTACAG GCGTGAGCCA

1741 CTACACCCAG CCGCATGATT CTAAAAAATA AAAAGATGAA GTGTTATTCC AAACATCTGA

1801 TCTCCATTGA AGAACCATGC AATCTCTCTG GGTTGATAGA GGCCAGAGTT AGTGGCTCTC

1861 CCTGATTTCG GTGAGAAATC ACTATTCCAC CATCACGGGA TAAAAGGCAT CCTGACTGGC

1921 GGTTGACACC TATTTCCACA GTGAAAGATA TATCTAGTAC TTTTAAAGGG GAAGTGGTTT

1981 GTCTGAGATA CTCTGTTTCA AAGTAGAGAG GATACAGAAC AAGCATCTGA AGCTATATAC

2041 ATCCTTACAG AGAGCAATTC TGATGGAAAT GCAGGCCATG TTTCCCTGGG GGGGGCTCGT

2101 CCTAGGGGCT GGAGTGCATT CTCTGATGTC AGAGGAAATG CAAGATTCCC TGAGGCCTGA

2161 GGGAACCCAT GGTATATGCA AGTCCAAGTT TCAAACTGTA GTTCCATATG CATTCTTCCA

2221 GGACAAATAC TTCTTGAGGT TAAAAAAAAA AAGTCACATA GCTGCCATTT TATGGATTTC

2281 AGGATTTTTT TTTTTTTTTT TTTGAGATGG AGTCTTGCTC TGTCACCCAG CCTGTAGTGC

2341 AGTGGCATAA TCTCGGCTCA CGGCAACCTC CGCCTCCCAG GTTCAAGCGA TTCTCTTGCC

2401 TTAGCCTCCC GAGTAGCTGG GATTACAGTC ACGCACCACC ACATCTGGCT AATTCTTTAT

2461 ATTTTTTGGT AGAAACGGTG TTTCACCATG TTGGCCAGGC TGGTCTCAAA CTCCTGACCT

2521 CATGTGATCT GCCTGCCTTG GCCTCCCAAA GTGCTGAGAT TACAGGTGTG AGCCACCGCG

2581 CCTGCCTGGA GTTCAGAATC TTGGGCTTCA TTATTTGTGT TTAAATAGAT CATACAGTCA

2641 GGCACGGTGG CTCATGCCTG TAATCCCAGC ACTTTGGGAG GCTGAGGTGG GAGGATTGCC

2701 TGAGTTCAGG AGATGGAGAC CAGCCTGGGC AACATGGTGA AACCCCGTCT CTACTAAAAA

2761 TACAAAAACT AGCTGGATGT GGTGGCACAC ACCTGTAGTC CCAGCTATTC AGGAGGCTGA

2821 GGTGGGAGGA TCCCAGGAGG TAGAGGTCAC AATGAGCCGA GATTGCGCCA CTGCACTCCA

2881 GGCTGGGTTA CTGAGCCAGA TCCTGTCTCA AAAAAAAAAA AGATAATACA TTCAAACAGT

2941 TCAAAATGCA AAAGTTACAT ACATAAGGAA GTGTCATGAA ATATCTCCCT CTCACACTTC

3001 TCCCCAGCCA CCCAGTTCTC CCTTCTAGAG GCAACATGTG AAATCCTTCT CAGGCTACAC

3061 TCTTCTTGAA GGTGTAGGCT TTGGGCAAAA GCATTCATTC AGTAACCCCA GAAACTTGTT

3121 CTGTTTTTCC ATAGGATTCT CTTTGGACAG CGTCCATACT GGTGGGTTTT GGATACTGAC

3181 TACTACAGCA ACACTTCCGT GCCCCTGATA AAGCAGTTCC CTGTAACCTG TGAGACTGGA

3241 CCAGGTAAGC GTCCCAGCCC CTGCAGACAG AAGCTGAGTG GACCTCGTTT ACCTGTTATG

3301 GATGAAACTG ACCTTGAGGG GACATGAGGA GAGCCATTCC TTTGTACTTT TGTCATGCTC

3361 TTCAATTGGC ACAAATTAAT TCACTTCTGC AATACTTTCC TGAATAGCAC AGTAGTATTG

3421 GAAATCTGCC TATTACAGAA CCTGGATGGA GTCCAGAGAG GCACGGGCAT CCATGGGCAA

3481 AGGGCTCGTG AGAGTCACCG CCCTGCAGCG CTGTGTCCTG AGAAAGGAGG GGGCAGAAGC

3541 CTGAGCTTCT GGGGGTCCTT CCCAATGGCC TGGCCCACTG GATGTGCCCT CCTGAGCTGA

3601 CCGTCCAATC CCTTGCCCTC TCTGTGCCTA CGTTTTATTA GTTACAGCCA GATGGTTACT

3661 GTCAAATCAA ATGATAGATT TCATTTTCAG TATGTAATAG GAAGCCCCTC CCTCACCCTA

3721 AAGTCTCAGC TGCCCTCTAA GACTAGTACT CTCTAAGGTA CTAGTATCCC TTCCTCAGAG

[0195]

3781 ACCCTTTCCC TGACCCCAAA ACTAGGGAAG GTCCCTTAGT TATTTGCTCT CACAGACCAC

3841 GCATTTACCT CAGAGCATAT TCACTCATTC AGCTGTTACT TACCAAGCAC CTACTGGGAG

3901 CTATACACTG TTCTATGTGC TAGGGATACC TCTGTCAGTG AACAACACAG ACACAAAGAT

3961 CCCTGCCCTT GTGGAGCTGA AATCTGAATA GAGGAGGTGA AATATACAAA AATTATAATA

4021 AATAAGTAAA CTAGGCCAGT TGTGGTTGCT CATGCCTGTA ATCCCAGCAC TTTGGGAAGC

4081 CAAGGTAGGT AGATCACCTG AGGTCAGGAG TTCAAAACCA GCCTGGCCAA CATTGCAAAA

4141 TCCTGTCTTT ACTAAAAATG GAAAAATTGG TCAGGCGTGA TGGCACACGC CTGTAGTCTC

4201 AGCTACCTGG GAGGCTGAGG CAGGAGAATC GCTTGAACCT GGGAGGCAGA GGTTGCAGTG

4261 AACCGAGATC GGACCACTGC ACTCCAGCCT GAATGACAGA ACGAGACTCT GTCTCAAAAA

4321 AAAAGTAAAC TATTAATATG TAGGATAGGC CAGGCACGGT GGCTCACCCT GTAATCCCAG

4381 CACTTTGGGA GGCTGAGGCG GGTGGATCAC CTGAGGTGAG GAGTTCAAGA CCAGCCTGGC

4441 CAACATGGCA AAACCCTGTC TCTACTAAAA ATACAAAAAT TAGCTGGGTG TCCTGGTGCA

4501 TGCCTGTAAT CTGAGCTACT CAGGAGGCTA AGGCAGGAGA ATCGCTTGAA CCTGGGAGGT

4561 GGTGAGCCAA GATTGCGCCA TTGCACTCCA GCCTGGGCGA CAAAATGAGA CACCATCTGA

4621 AAAAAAAAAA AAAATATATA TATATATACA CACACACACA CACACACACA CACACACACA

4681 TATAATACTA GAAAATGATT GTTTATAGGC AAAAAAAAAA AAAAAGAAGA AGAAGAAGAA

4741 AAGGAAAGGA GAAGGAAAGA AGGACCAAAC ATCTTTTGTA GAAATATGTT TGCTTTCATC

4801 ATAACAGCTT GTTATCAAGG ATGAATTTCT CCCTGAAATT AATGGAGGCA CAGACTGGAA

4861 AGTTTAAAGT GGCTTTAAGA GGTTATTTTA TTTAGTCCTC TGTCTTAATA GAAGCAAATT

4921 ATTATCTCTG CTCCTTAGGT AGAGTAGCTA AGGCTCAGAA AGTAGGCCGG GCGCGGTGGC

4981 TCACGCCTGT AATCCTAGCA CTTTGGGAGG CCAACGCAGG TGGATCACCT GAGGTCAGGA

5041 GTTTGAGACC AGCCTGGCCA ACATGGTGAA ACCTCGTCAC TAATAAAAAA ATACAAAAAC

5101 TTAGCCAGGC ATGGTGGCGG GCGCCTGTAA TCCCAGCTAC CCAGGAGGCT GCGGCAGGAG

5161 AATCACTTCA ACCCGGGAGG CAGAGGTTGC AGTGAGCTGA AATCACACCA CTGCACTCCA

5221 GCCTTGGTGA CAGAGAAAGA TTCTGTCAGG AAAAAAAAAA AAAAGTTTAA ATGAATTACC

5281 CAAGGTATAT AATTGTTAGT GTTAGAAGGA AGAAGAAGGG AGGGAGGAAG GAAGGGAGAA

5341 AGAAAGGGAA GGAGGAAGGG AGGGAGGGAA GAAAGCCTTT ATTTATCTAT GGGGTTCCCT

5401 GGAAAGCAGG CTGAAATGGA GATTCACGTG CAGGAGTTTA GATACTCTGG GGAACTATAC

5461 TTGTAGAAGG GAAGGAACAG GAACAGGGCA GAAGGAGAGG TCCGGTTGTG ATTCTGCCTC

5521 ATCCAACCCC ACAGCGAGCT CTGAAGCTGG GGATGGCTCC TCAGAGTTGG TCCAAGTTGG

5581 GACAAGGGAA TCAGACCCTG GGGAGAGCGT AACCTTGATC AAGGCGACTC TCTTTAGCCC

5641 AGGGCAATGC CAGGAGAAGG CTGAGAGCAG AAAGCCATCT ACCATCACAC TCTCAACAGC

5701 TACGAAATAA GTCCTGCAGT TCAGGAGGGA GGTCTGGGCG GCACATCTCA GGACCCTCTA

5761 TCTCTCAGGG TAGAGGAATT AAGAATGGGA TGGGAACCAG ACGGGCCATG GTGGCTCACA

5821 CCTATAATCC CAACACTTTG GGAGGCCAAG GGTAGGAGGA TTGCTTGAGC CCAAGAGTTC

5881 AAAACCAGCC TGGGCAAAAA CAATCAAACA AACAAACAAA ACACATTTAA AAAATTTGCT

[0196]

5941 GTGTGTGGTG GTGTGCACCT GTGGTCCCAG CTACTCAGGG GGCTGAGGTG GGAGGATTGC

6001 TTGAGTCCAG GAGGTCGAGG CTGCAGTGAG CTATGATCAT GGCACTGCAT TGCAGCCTAG

6061 GAGACAAAGC AAGACACTGT CTCTAAAAAA ACAAAAAACA AACAAATAAA AAAACGGAAC

6121 CGGTTGCAAG CAGGGTTAAA TAGCGTGGTC AGAGTAGGAC TCACTGAGAA TATGAGATCT

6181 GAGTCAAGTC TTCAAGGATG TGAGGAAGTA AGTTTCTGGC AGAAGAGCTG TGAAGGGCTG

6241 TCTGGCCAGA GAAGATTGCA ATGCAAAAGC CCTGAGGTGG GAACGTGTTT GGTGTGTTTA

6301 AAGGAAAGCA ATGAGGCCAG TGTAGCCAGA ACAGAGTGTG CAAGGAGAGA AGGAACAGAA

6361 GATGTGGAGG GCAGATCAGT TTGTAATTGT ACGCCCAGTA TGCTGATTCT TTGTGTAATC

6421 TCCAGACTGT ATTAAACTGC AAGAGCAGGG CCCCTCTCTG GCTTTGCTCA TCATTGTATT

6481 CCCAGAGCCT TGCACAATGC TTGGTGCATA GGAGATGGAA ATTTGTTAAA TAAATGAATT

6541 ATGGATAACG AATGGATGGT AAGATGGGTG GATGGATGGG GGGTGAACGG ATGGATGGGG

6601 GGTGAATGGA TGGATGAATG GGTAGATGGG TGGATAGGGG GATGGCTGGG TGGCTGGGTA

6661 GATGATGCAC TGTCTCCCAG ATGAGGACCT TTTCACCTTT ACTCCATTCT CTTTCCTGCC

6721 CTTTAGGGAG CCCCTCTGGC CATGCCATGG GCACAGCAGG TGTATACTAC GTGATGGTCA

6781 CATCTACTCT TTCCATCTTT CAGGGAAAGA TAAAGCCGAC CTACAGATTT CGGTAAGAAC

6841 TCACCACTGG GGTGTAGGTG GTGGAGGGCA GGAGGCAGCT CTCTCTGTAG CTGACACACC

6901 ACGTATTCTT CCTCACATCC CCCTAGCCCG CTCCCACACC TGGGCAGCCG CTGATTAAGA

6961 GTTGTGGCAC TTTGGATAGG GATAAACCTC AGAGTCAGGG AATGTTTGGG CTGAAAGGGA

7021 TCCAGTAGTG CAATCCGTTG TTTTACAGAT AAGGAAACAA AGCCCAACAC CATGAAGGGA

7081 CTTATAAAAA TAAGGTAGTG AAGTAGCAGC AGGGCTTAAA TAAAAACCCA TGTCTGTACC

7141 AACCACAGAG TCACCCATCC AGGTTAAAAT AACCAGAGAA ACAGAAGATA TTCCTACTAC

7201 AGAGAATTCC GGGTGTGCAG CCACAGTGCA AATCCTTTTT ATTTTTATTT TTGAGATGCA

7261 GTCTCGCTCT GTCATCCAGG CTGAAGTGCA GTGGCACGAT CATGTCTCGC TGCAACCTCT

7321 GCCTCCCAGG CTCAAGCGAT CCTCCCACCT CAGCCATCTG AGTAGCTGGG ACCACAGGCC

7381 ACACACCACA CCCAGCTAAT TTCTCGTATC TTTTTGTAGA GACAGAGTTC TGCTATGTTG

7441 CCCAGGCTCA GGCTGGTCTT GATCTCAAGC AATTGGCTTG CCTCAGCCTC CTAAAATATT

7501 GGGATTACAG GCATGAGCCA CCGCGCCAGC CATGCAAATC CTTAATTATC AAACAGATAA

7561 AATAGGGAAG TTAAAATTCA TATACACAAG GGTTAACCAC TTGCCACAGG CATTTTTTTT

7621 TTTTTTTTGA GACGGAATCT CGCTCTGTTG CCCAGGCTGG AGTGCAGTGG CGCCATCTCG

7681 CCTCACTGCA ACCTCCGCTT CCTGGGTTCA AGCTATTCTT CTGCCTCAGC CTACCGAGTA

7741 GCTGGGACTA CAGGCACGTG CCACCACACC TGGCTAATTT TTTTATTTTT AGTAGAGATG

7801 GGGTTTCACC ATATTGGCCA GGCTGGTCTT GAACTCCTGA CCTAGTGATC CATCCGCCTC

7861 AGCCTCCCAA AGTGCTGGGA TTGCAGGCAT GAGCCACCGC GCCTGGCCTT TTTTTTTTTT

7921 TTTTGAGACG GAGTTTTGCT CTTGTTGCCC AGGCTAGAGT GCAGTGGCGC AGTCTCGGCT

7981 CACTGTAACC TCCACCTCCT GAGTTCAAGC AATTCTCCTG CCTCAGCCTC TCAAATAGCT

8041 GGGATTACAG GCGTGAGCCA CCCCACCTGG CTAATTTTGT AATTTTTTTT TTAGTAGAGA

[0197]

8101 TGGGGTTTCA CCTGTTGATC AGGCTGGTCT CAAACTCCTG ACCTCAAGTG ATCCACCCAC

8161 CTCGGCCTCC CAAAGTGCTG GGATTACAAG CATAAGCCAC CGTGCCTGGT CAATTTTGAT

8221 CTTTTTTAAA GAGACAGGGG TCTTGCTATG TTGCCCAGAC TAGTCTTGAA CTCCTGGCCT

8281 CAAGTGATCC TCTCACCTCG GCCTCCCAAA GTATTGGGAT TACAGGTCTG AGCCGCTGCA

8341 CCCAGCCCCC AACAGGCATC TTTGGACTTT TGAGTACTGG CTTTAATTTA CAAAAATTCC

8401 ACTGAGAGCA CCTAAGTTTG CCAGGCTCCA ACATTTCTGC AGGGGCTGTT TTCTTTGCTG

8461 AAGGATCTGC ACCTGTGTTC TGTTATGGTT GCCTCTTCTG TTGCAGGTGC TTGAATGTCA

8521 TTTTGTGGTT GGGATTCTGG GCTGTGCAGC TGAATGTCTG TCTGTCACGA ATCTACCTTG

8581 CTGCTCATTT TCCTCATCAA GTTGTTGCTG GAGTCCTGTC AGGTATGGGC TGATCTGACT

8641 CCCTTCCTTC TCCCCCAAAC CCCATTCCGT TTCTCTCCCT AATCAGGACA AAATCCCAGC

8701 ATTCCAGCCA CATCCTGTGT GTAATCAGTA CTGTTAGCAT TTCTGTGGGT TGAAAGTCAA

8761 GAATGAGCAA CTTGAAATGA TTAATTTCTA TAAGAGTGCC CAGATCTATA GAATGAATTG

8821 TGTAGAAGTT ACCATACATC AAATTAACGC ACCAAATTGA ATTAGCTTGA AATCTCAGAG

8881 CTTTTTACAA TCTTTATTTC TTACTGGTCT TCAACAGGCC CTAATTTACT TTTCAGGGAA

8941 TCTGCCAAAT TTAACAAATT AACACGATGT CCTAGGAAAG CTGTTCATTT AAATACATTC

9001 ATTTGCAAAC CTAATAGATA ACTGCAGTTG ATCTCTTTTA TAGGTTCAGA GTTTTGAATA

9061 TGTTTTTTTT TGTTTTTTTT TTTTGAGATG GAGTCTCGCT CTGTGACCCA GGCTAGAGTG

9121 CAGTGGTGCG ATCTCGGCTC ACTGCAAGCT CCACCTCCTG GGTTCACGCC ATTCTCCTGC

9181 CTCAGCCTCT CCGAGTAGCT GGGACTACAG GCGCCCGCCA CCATGCCCGG CTAATTTTTT

9241 GTATTTTTAG CAGAGACGGG GTTTCACCGT GGTCTTGATC TCCTGACCTC GTGATCCGCC

9301 CGCCTCGGCC TCCCAAAGCG CTGGGATTAC AAGGGTGAGC CACCGCACCC TGCCTGAATA

9361 TGTGTTTTCT TAGATCCAAT TAACAAGGGT AAGACAAGAT TTAAGTTAAG CATAAGAAAG

9421 ATTTTGTGGG AGGCACTGGA ATATAAGACC TTAACAAAAC TGTGGAATTT CTCCCCTGGA

9481 GATTTGTAAG AACGGAACAT AGCAGCATTC AAAGAAGAAT GTTGAGAACA AGGGAGATAA

9541 TGGTTTCATG GTAATCACAA AAGTAACACA GCATTTAGTA CTGGGTTCCA TGTTTGAGGA

9601 AGAACCTGGA AGCCATATCA CATGAAAAAC CTGGGAATGT TTAGGTTAGA GAGAATAACT

9661 GTGTTCAAAT GTGTGACAGA GGGACTAGAT TCATCACTTA CTAACTCCTG CAGAAAGAAC

9721 TGAGAAAAAT AGACAGTATT AGAGGGGGAC CAGTTTCACA CAGACAAGGA AGAACTATTC

9781 AGCAATCAAT TCCGTTCAAA GATAAAATGG ACTGTTATAG TGGGGGTGAG CTCCCTACCT

9841 CTGAGGGTAT TTCAAGTAGA GATAGGAGGA CCTCCTGGTA GGAAATTTGC ATACGGTGGG

9901 AGATTGTACG TGATATGGCA CCTCCATCTG AAAGAGTCTA TATTGAGGGC AGGCTGGAGT

9961 CACACATGGG AATAAGCCAG GCGACCCTCC CATCTGCCAT CTGTGATTTA ATTCCACAGT

10021 CGCAGAACGG ATGGCATGTC ACCCACTCCT CCAAACCCAC CTCTAGCAAA GGTCCCAAAT

10081 CCTTCCTATC TCTCACAGTC ATGCTTTCTT CCACTCAGGC ATTGCTGTTA CAGAAACTTT

10141 CAGCCACATC CACAGCATCT ATAATGCCAG CCTCAAGAAA TATTTTCTCA TTACCTTCTT

10201 CCTGTTCAGC TTCGCCATCG GATTTTATCT GCTGCTCAAG GGACTGGGTG TAGACCTCCT

[0198]

10261 GTGGACTCTG GAGAAAGCCC AGAGGTGGTG CGAGCAGCCA GAATGGGTCC ACATTGACAC

10321 CACACCCTTT GCCAGCCTCC TCAAGAACCT GGGCACGCTC TTTGGCCTGG GGCTGGCTCT

10381 CAACTCCAGC ATGTACAGGG AGAGCTGCAA GGGGAAACTC AGCAAGTGGC TCCCATTCCG

10441 CCTCAGCTCT ATTGTAGCCT CCCTCGTCCT CCTGCACGTC TTTGACTCCT TGAAACCCCC

10501 ATCCCAAGTC GAGCTGGTCT TCTACGTCTT GTCCTTCTGC AAGAGTGCGG TAGTGCCCCT

10561 GGCATCCGTC AGTGTCATCC CCTACTGCCT CGCCCAGGTC CTGGGCCAGC CGCACAAGAA

10621 GTCGTTGTAA GAGATGTGGA GTCTTCGGTG TTTAAAGTCA ACAACCATGC CAGGGATTGA

10681 GGAGGACTAC TATTTGAAGC AATGGGCACT GGTATTTGGA GCAAGTGACA TGCCATCCAT

10741 TCTGCCGTCG TGGAATTAAA TCACGGATGG CAGATTGGAG GGTCGCCTGG CTTATTCCCA

10801 TGTGTGACTC CAGCCTGCCC TCAGCACAGA CTCTTTCAGA TGGAGGTGCC ATATCACGTA

10861 CACCATATGC AAGTTTCCCG CCAGGAGGTC CTCCTCTCTC TACTTGAATA CTCTCACAAG

10921 TAGGGAGCTC ACTCCCACTG GAACAGCCCA TTTTATCTTT GAATGGTCTT CTGCCAGCCC

10981 ATTTTGAGGC CAGAGGTGCT GTCAGCTCAG GTGGTCCTCT TTTACAATCC TAATCATATT

11041 GGGTAATGTT TTTGAAAAGC TAATGAAGCT ATTGAGAAAG ACCTGTTGCT AGAAGTTGGG

11101 TTGTTCTGGA TTTTCCCCTG AAGACTTACT TATTCTTCCG TCACATATAC AAAAGCAAGA

11161 CTTCCAGGTA GGGCCAGCTC ACAAGCCCAG GCTGGAGATC CTAACTGAGA ATTTTCTACC

11221 TGTGTTCATT CTTACCGAGA AAAGGAGAAA GGAGCTCTGA ATCTGATAGG AAAAGAAGGC

11281 TGCCTAAGGA GGAGTTTTTA GTATGTGGCG TATCATGCAA GTGCTATGCC AAGCCATGTC

11341 TAAATGGCTT TAATTATATA GTAATGCACT CTCAGTAATG GGGGACCAGC TTAAGTATAA

11401 TTAATAGATG GTTAGTGGGG TAATTCTGCT TCTAGTATTT TTTTTACTGT GCATACATGT

11461 TCATCGTATT TCCTTGGATT TCTGAATGGC TGCAGTGACC CAGATATTGC ACTAGGTCAA

11521 AACATTCAGG TATAGCTGAC ATCTCCTCTA TCACATTACA TCATCCTCCT TATAAGCCCA

11581 GCTCTGCTTT TTCCAGATTC TTCCACTGGC TCCACATCCA CCCCACTGGA TCTTCAGAAG

11641 GCTAGAGGGC GACTCTGGTG GTGCTTTTGT ATGTTTCAAT TAGGCTCTGA AATCTTGGGC

11701 AAAATGACAA GGGGAGGGCC AGGATTCCTC TCTCAGGTCA CTCCAGTGTT ACTTTTAATT

11761 CCTAGAGGGT AAATATGACT CCTTTCTCTA TCCCAAGCCA ACCAAGAGCA CATTCTTAAA

11821 GGAAAAGTCA ACATCTTCTC TCTTTTTTTT TTTTTTTGAG ACAGGGTCTC ACTATGTTGC

11881 CCAGGCTGCT CTTGAATTCC TGGGCTCAAG CAGTCCTCCC ACCCTACCAC AGCGTCCCGC

11941 GTAGCTGGGA CTACAGGTGC AAGCCACTAT GTCCAGCTAG CCAACTCCTC CTTGCCTGCT

12001 TTTCTTTTTT TTTCTTTTTT TGAGACGGCG CACCTATCAC CCAGGCTGGA GTGGAGTGGC

12061 ACGATCTTGG CTCACTGCAA CCTCTTCCTC CTGGTTCAAG CGATTCTCAT GTCTCAGCCT

12121 CCTCAGTAGC TAGGACTACC GGCGTGCACC ACCATGCCAG GCTAATTTTT ATATTTTTAG

12181 AATTTTAGAA GAGATGGGAT TTCATCATGT TGGCCAGGCT GGTCTCGAAC TCCTGACCTC

12241 AAGTGATCCA CCTGCCTTGG CCTCCCAAGG TGCTAGGATT ACAGGCATGA GCCACCGCAC

12301 CGGGCCCTCC TTGCCTGTTT TTCAATCTCA TCTGATATGC AGAGTATTTC TGCCCCACCC

12361 ACCTACCCCC CAAAAAAAGC TGAAGCCTAT TTATTTGAAA GTCCTTGTTT TTGCTACTAA

12421 TTATATAGTA TACCATACAT TATCATTCAA AACAACCATC CTGCTCATAA CATCTTTGAA

12481 AAGAAAAATA TATATGTGCA GTATTTTATT AAAGCAACAT TTTATTTAAG AATAAAGTCT

12541 TGTTAATTAC TATATTTTAG ATGCAATGTG ATCTGAAGTT TCTAATTCTG GCCCAACTAA

12601 ATTTCTAGCT CTGTTTCCCT AAACAAATAA TTTGGTTTCT CTGTGCCTGC ATTTTCCCTT

12661 TGGAGAAGAA AAGTGCTCTC TCTTGAGTTG ACCGAGAGTC CCATTAGGGA TAGGGAGACT

12721 TAAATGCATC CACAGGGGCA CAGGCAGAGT TGAGCACATA AACGGAGGCC CAAAATCAGC

12781 ATAGAACCAG AAAGATTCAG AGTTGGCCAA GAATGAACAT TGGCTACCAG ACCACAAGTC

12841 AGCATGAGTT GCTCTATGGC ATCAAATTGC AACTTGAGAG TAGATGGGCA GGGTCACTAT

12901 CAAATTAAGC AATCAGGGCA CACAAGTTGC AGTAACACAA CAAGACTAGG CCAGCTCTGG

12961 AATCCAGTAA CTCAGTGTCA GCAAGGTTTT GGGTTATAGT TCAAGAAAGT CTAAACAGAG

[0199] 13021 CCAGTCACAG CACCAAGGAA TGCTCAAGGG AGCTATTGCA GGTTTCTCTG CTAAGAGATT

13081 TATTTCATCC TGGGTGCAGG GTTCGACCTC CAAAGGCCTC AAATCATCAC CGTATCAATG

13141 GATTTCCTGA GGGTAAGCTC CGCTATTTCA CACCTGAACT CCGGAGTCTG TATATTCAGG

13201 GAAGATTGCA TTCTCCTACT GGATTTGGGC TCTCAGAGGG CGTTGTGGGA ACCAGGCCCC

13261 TCACAGAATC AAATGGTCCC AACCAGGGAG AAAGAAAATA GTCTTTTTTT TTTTTTAAT

13321 AGAGATGGGG GTCTCACTAT GCTGCCCAGG CTGGTCTTGA ACTCCTGGGT TCAAGTGATC

13381 CTCCTGCCTC AGCCTCCCAA AGTGCTGGGA TTACAGTGTG AGCCACTGCG CTTGGCCAGA

13441 AATGGTTTTG ATCTGTCTGA ACTGAACCCT ACTGCTTAGG CATAGCCCCA TCCTTGATAA

13501 TCTATTTGCT CCCAAGGACC AAGTCCAAGA TCCTTACAAG AAAGGTCTGC CAGAAAGTAA

13561 ATACTGCCCC CACTCCCTGA AGTTTATGAG GTTGATAAGA AAACATAACA GATAAAGTTT

13621 ATTGAGTGCT AACTTTA

[0200] “向导RNA”或“gRNA”是指可以特异于目标序列并且可以与多核苷酸可编程核苷酸结合结构域蛋白(例如Cas9或Cpf1)形成复合物的多核苷酸。在一个实施例中,向导多核苷酸是向导RNA(gRNA)。gRNA可以作为两个或多个RNA的复合物存在,也可以作为单个RNA分子存在。以单个RNA分子形式存在的gRNA可称为单向导RNA(sgRNA),但“gRNA”可互换使用,指以单个分子或两个或多个分子的复合体形式存在的向导RNA。通常,作为单一RNA种类存在的gRNA包含两个结构域:(1)与目标核酸具有同源性的结构域(例如引导Cas9复合物与目标的结合);(2)结合Cas9蛋白的结构域。在一些实施例中,结构域(2)对应于称为tracrRNA的序列,并且包含茎环结构。例如,在一些实施例中,结构域(2)与Jinek等人,Science337:816-821(2012)中提供的tracrRNA相同或同源,该文献的全部内容通过引用并入本文。gRNA的其他示例(例如包括结构域2的那些)可以在美国临时专利申请号U.S.S.N.61/874,682,于2013年9月6日提交,题为“Switchable Cas9 Nucleases and Uses Thereof”,以及美国临时专利申请号U.S.S.N.61/874,746,提交于2013年9月6日,题为“Delivery System For Functional Nucleases”中找到,其各自的全部内容通过引用整体并入本文。在一些实施例中,gRNA包含结构域(1)和(2)中的两个或更多个,并且可以被称为“扩展的gRNA”。如本文所述,延伸的gRNA将结合两个或更多个Cas9蛋白并在两个或更多个不同区域结合目标核酸。gRNA包含与靶位点互补的核苷酸序列,其介导核酸酶/RNA复合物与所述靶位点的结合,提

供核酸酶:RNA复合物的序列特异性。如本领域技术人员将理解的,RNA多核苷酸序列,例如gRNA序列,包括核碱基尿嘧啶(U),一种嘧啶衍生物,而不是包含在DNA多核苷酸序列中的核碱基胸腺嘧啶(T)。在RNA中,尿嘧啶与腺嘌呤碱基配对并在DNA转录过程中取代胸腺嘧啶。

[0201] “异源二聚体”是指包含两个结构域的融合蛋白,例如野生型TadA结构域和TadA结构域的变体(例如TadA*8)或两个变体的TadA结构域(例如TadA*7.10和TadA*8或两个TadA*8结构域)。

[0202] “杂交”是指互补核碱基之间的氢键,其可以是瓦生克立克(Watson-Crick)、胡斯坦(Hoogsteen)或反向胡斯坦氢键。例如,腺嘌呤和胸腺嘧啶是互补的核碱基,它们通过形成氢键配对。

[0203] 术语“碱基修复抑制剂(inhibitor of base repair)”或“IBR”是指能够抑制核酸修复酶例如碱基切除修复(BER)酶的活性的蛋白质。在一些实施例中,IBR是肌苷碱基切除修复的抑制剂。碱基修复的示例性抑制剂包括APE1、Endo III、Endo IV、Endo V、Endo VIII、Fpg、hOGG1、hNEIL1、T7 Endol、T4PDG、UDG、hSMUG1和hAAG的抑制剂。在一些实施例中,IBR是Endo V或hAAG的抑制剂。在一些实施例中,IBR是无催化活性的EndoV或无催化活性的hAAG。在一些实施例中,碱基修复抑制剂是Endo V或hAAG的抑制剂。在一些实施例中,碱基修复抑制剂是无催化活性的EndoV或无催化活性的hAAG。

[0204] 在一些实施例中,碱基修复抑制剂是尿嘧啶糖基化酶抑制剂(UGI)。UGI是指能够抑制尿嘧啶-DNA糖基化酶碱基切除修复酶的蛋白质。在一些实施例中,UGI结构域包含野生型UGI或野生型UGI的片段。在一些实施例中,本文提供的UGI蛋白质包括UGI的片段和与UGI或UGI片段同源的蛋白质。在一些实施例中,碱基修复抑制剂是肌苷碱基切除修复抑制剂。在一些实施例中,碱基修复抑制剂是“无催化活性的肌苷特异性核酸酶”或“死肌苷特异性核酸酶”。不希望受任何特定理论的束缚,无催化活性的肌苷糖基化酶(例如烷基腺嘌呤糖基化酶(AAG))可以结合肌苷,但不能产生脱碱基位点或去除肌苷,从而在空间上阻断新形成的肌苷部分免受DNA损伤/修复机制。在一些实施例中,无催化活性的肌苷特异性核酸酶能够结合核酸中的肌苷但不切割核酸。非限制性、示例性无催化活性肌苷特异性核酸酶包括无催化活性烷基腺苷糖基化酶(AAG核酸酶),例如来自人,和无催化活性内切核酸酶V(EndoV核酸酶),例如来自大肠杆菌。在一些实施例中,无催化活性的AAG核酸酶包含E125Q突变或另一种AAG核酸酶中的相应突变。

[0205] “增加”是指至少10%、25%、50%、75%或100%的正变化。

[0206] “内含肽(intein)”是一种蛋白质片段,它能够自我切除并在称为蛋白质剪接的过程中用肽键连接剩余的片段(外显肽)。内含肽也称为“蛋白质内含子”。内含肽自身切除并连接蛋白质剩余部分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”。在一些实施例中,前体蛋白的内含肽(在内含肽介导的蛋白质剪接之前含有内含肽的蛋白质)来自两个基因。这种内含肽在本文中被称为分裂内含肽(例如分裂内含肽-N和分裂内含肽-C)。例如,在蓝藻中,DNA聚合酶III的催化亚基a的DnaE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称为“内含肽-N”。由dnaE-c基因编码的内含肽在本文中可称为“内含肽-C”。

[0207] 也可以使用其他内含肽系统。例如,基于dnaE内含肽、Cfa-N(例如分裂的内含肽-N)和Cfa-C(例如分裂的内含肽-C)内含肽对的合成内含肽已被描述(例如在Stevens等人,J

Am Chem Soc.2016Feb.24;138(7):2162-5,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实例包括: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(例如描述于美国专利号8,394,604中,该专利通过引用并入本文)。

[0208] 以下提供了内含肽的示例性核苷酸和氨基酸序列。

[0209] DnaE内含肽-N DNA:

[0210] TGCCTGTCATACGAAACCGAGATACTGACAGTAGAATATGGCCTTCTGCCAATCGGGAAGATTGTGGAGAAACGGATAGAATGCACAGTTTACTCTGTCGATAACAATGGTAACATTTATACTCAGCCAGTTGCCCAGTGGCACGACCGGGGAGAGCAGGAAGTATTCGAATACTGTCTGGAGGATGGAAGTCTCATTAGGGCCACTAAGGACCACAAATTTATGACAGTCGATGGCCAGATGCTGCCTATAGACGAAATCTTTGAGCGAGAGTTGGACCTCATGCGAGTTGACAACCTTCCTAAT

[0211] DnaE内含肽-N蛋白质:

[0212] CLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYCLEDGSLIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPN

[0213] DnaE内含肽-C DNA:

[0214] ATGATCAAGATAGCTACAAGGAAGTATCTTGGCAAACAAAACGTTTATGATATTGGAGTCGAAAGAGATCACAACTTTGCTCTGAAGAACGGATTCATAGCTTCTAAT

[0215] 内含肽-C:MIKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN

[0216] Cfa-N DNA:

[0217] TGCCTGTCTTATGATACCGAGATACTTACCGTTGAATATGGCTTCTTGCCTATTGGAAAGATTGTCGAAGAGAGAATTGAATGCACAGTATATACTGTAGACAAGAATGGTTTCGTTTACACACAGCCCATTGCTCAATGGCACAATCGCGGCGAACAAGAAGTATTTGAGTACTGTCTCGAGGATGGAAGCATCATACGAGCAACTAAAGATCATAAATTCATGACCACTGACGGGCAGATGTTGCCAATAGATGAGATATTCGAGCGGGGCTTGGATCTCAAACAAGTGGATGGATTGCCA

[0218] Cfa-N蛋白质:

[0219] CLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCLEDGSIIRATKDHKFMTTDGQMLPIDEIFERGLDLKQVDGLP

[0220] Cfa-C DNA:

[0221] ATGAAGAGGACTGCCGATGGATCAGAGTTTGAATCTCCCAAGAAGAAGAGGAAAGTAAAGATAATATCTCGAAAAAGTCTTGGTACCCAAAATGTCTATGATATTGGAGTGGAGAAAGATCACAACTTCCTTCTCAAGAACGGTCTCGTAGCCAGCAAC

[0222] Cfa-C蛋白质:

[0223] MKRTADGSEFESPKKKRKVKIISRKSLGTQNVYDIGVEKDHNFLLKNGLVASN

[0224] 内含肽-N和内含肽-C可以分别融合到分裂Cas9的N-末端部分和分裂Cas9的C-末端部分,用于连接分裂Cas9的N-末端部分和分裂Cas9的C-末端部分。例如,在一些实施例中,内含肽-N融合到分裂Cas9的N端部分的C端,即形成N--[分裂Cas9的N-末端部分]-[内含肽-N]--C的结构。在一些实施例中,内含肽-C融合到分裂Cas9的C-末端部分的N-末端,即形成N-[内含肽-C]--[分裂Cas9的C-末端部分]-C。内含肽介导的蛋白质剪接机制用于连接内含肽融合到的蛋白质(例如分裂Cas9)是本领域已知的,例如,如Shah等人描述,Chem

Sci.2014;5(1):446-461,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已知的并且被例如WO2014004336、WO2017132580、US20150344549和US20180127780描述,它们各自通过引用整体并入本文。

[0225] 术语“分离的”、“纯化的”或“生物纯的”是指材料在不同程度上不含在其天然状态下通常伴随的组分。“隔离”表示与原始来源或周围环境的分离程度。“纯化”表示高于隔离度的分离度。“纯化的”或“生物学纯的”蛋白质充分不含其他材料,使得任何杂质不会实质性地影响蛋白质的生物学特性或引起其他不利后果。也就是说,如果本发明的核酸或肽在通过重组DNA技术生产时基本上不含细胞材料、病毒材料或培养基,或者在化学合成时基本上不含化学前体或其他化学品,则该核酸或肽被纯化。纯度和均匀性通常使用分析化学技术确定,例如聚丙烯酰胺凝胶电泳或高效液相色谱。术语“纯化的”可以表示核酸或蛋白质在电泳凝胶中产生基本上一条带。对于可以进行修饰(例如磷酸化或糖基化)的蛋白质,不同的修饰可能会产生不同的分离蛋白质,这些蛋白质可以单独纯化。

[0226] “分离的多核苷酸”是指不含基因的核酸(例如DNA),在本发明的核酸分子所源自的生物体的天然存在的基因组中,该基因位于该基因的侧翼。因此,该术语包括,例如,整合到载体中的重组DNA;进入自主复制的质粒或病毒;或进入原核生物或真核生物的基因组DNA;或作为独立于其他序列的独立分子(例如通过PCR或限制性内切核酸酶消化产生的cDNA或基因组或cDNA片段)存在。此外,该术语包括从DNA分子转录的RNA分子,以及作为编码额外多肽序列的杂合基因的一部分的重组DNA。

[0227] “分离的多肽”是指已与天然伴随的组分分离的本发明的多肽。通常,当多肽至少60%(重量)不含蛋白质和天然存在的有机分子时,多肽就被分离出来。优选地,该制剂重量是本发明多肽的至少75%,更优选至少90%,而最优选至少99%。本发明的分离的多肽可以通过以下方式获得,例如,通过从天然来源中提取,通过编码这种多肽的重组核酸的表达;或通过化学合成蛋白质。纯度可以通过任何合适的方法测量,例如柱色谱、聚丙烯酰胺凝胶电泳或通过HPLC分析。

[0228] 如本文所用,术语“接头”可指共价接头(例如共价键)、非共价接头、化学基团或连接两个分子或部分(例如蛋白质复合物的两个组分)的分子或核糖核复合体,或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)。接头可以连接碱基编辑器系统的不同组分或组分的不同部分。例如,在一些实施例中,接头可以连接多核苷酸可编程核苷酸结合结构域的向导多核苷酸结合结构域和脱氨酶的催化结构域。在一些实施例中,接头可以连接CRISPR多肽和脱氨酶。在一些实施例中,接头可以连接Cas9和脱氨酶。在一些实施例中,接头可以连接dCas9和脱氨酶。在一些实施例中,接头可以连接nCas9和脱氨酶。在一些实施例中,接头可以连接向导多核苷酸和脱氨酶。在一些实施例中,接头可以连接碱基编辑器系统的脱氨基组分和多核苷酸可编程核苷酸结合组分。在一些实施例中,接头可以连接碱基编辑器系统的脱氨基组分和多核苷酸可编程核苷酸结合组分的RNA结合部分。在一些实施例中,接头可以连接碱基编辑器系统的脱氨基组分的RNA结合部分和多核苷酸可编程核苷酸结合组分的RNA结合部分。接头可以位于两个基团、分子或其他部分之间或两侧,并通过共价键或非共价相互作用连接到每一个,从而连接这两者。在一些实施例中,接头可以是有机分子、基团、聚合物或化学部分。在一些实施例中,接头可以是多核苷酸。在一些实施例中,接头可以是DNA接头。在一些实施例中,接

头可以是RNA接头。在一些实施例中,接头可包含能够结合配体的适体。在一些实施例中,配体可以是碳水化合物、肽、蛋白质或核酸。在一些实施例中,接头可包含可衍生自核糖开关的适体。衍生适体的核糖开关可以选自茶碱核糖开关、焦磷酸硫胺素(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核糖开关或前奎氨酸(PreQ1)核糖开关。在一些实施例中,接头可包含与多肽或蛋白质结构域例如多肽配体结合的适体。在一些实施例中,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 Com外壳蛋白结构域、无菌α基序(sterile alpha motif)、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。在一些实施例中,多肽配体可以是碱基编辑器系统组分的一部分。例如,核碱基编辑组分可包含脱氨酶结构域和RNA识别基序。

[0229] 在一些实施例中,接头可以是一个氨基酸或多个氨基酸(例如肽或蛋白质)。在一些实施例中,接头的长度可为约5-100个氨基酸,例如约5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、20-30、30-40、40-50、50-60、60-70、70-80、80-90,或90-100个氨基酸长度。在一些实施例中,接头可为约100-150、150-200、200-250、250-300、300-350、350-400、400-450,或450-500个氨基酸长度。也可以考虑更长或更短的接头。

[0230] 在一些实施例中,接头连接RNA可编程核酸酶的gRNA结合结构域,包括Cas9核酸酶结构域和核酸编辑蛋白(例如腺苷脱氨酶)的催化结构域。在一些实施例中,接头连接dCas9和核酸编辑蛋白。例如,接头位于两个基团、分子或其他部分之间或两侧,并通过共价键与每一个连接,从而将两者连接起来。在一些实施例中,接头是一个氨基酸或多个氨基酸(例如肽或蛋白质)。在一些实施例中,接头是有机分子、基团、聚合物或化学部分。在一些实施例中,接头的长度为5-200个氨基酸,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。也考虑更长或更短的接头。

[0231] 在一些实施例中,核碱基编辑器的结构域通过包含以下氨基酸序列的接头融合:SGGSSGSETPGTSESATPESSGGS、SGGSSGGSSGSETPGTSESATPESSGGSSGGS、或

[0232] GGSGGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGG SGGS。在一些实施例中,核碱基编辑器的结构域通过包含氨基酸序列SGSETPGTSESATPES的接头融合,它也可以称为XTEN接头。在一些实施例中,接头包含氨基酸序列SGGS。在一些实施例中,接头包含$(SGGS)_n$、$(GGGS)_n$、$(GGGGS)_n$、$(G)_n$、$(EAAAK)_n$、$(GGS)_n$、SGSETPGTSESATPES、或$(XP)_n$基序、或任何这些的组合,其中n独立地是1和30之间的整数,并且其中X是任何氨基酸。在一些实施例中,n是1、2、3、4、5、6、7、8、9、10、11、12、13、14或15。

[0233] 在一些实施例中,接头的长度为24个氨基酸。在一些实施例中,接头包含氨基酸序列SGGSSGGSSGSETPGTSESATPES。在一些实施例中,接头的长度为40个氨基酸。在一些实施例中,接头包含氨基酸序列SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGS。在一些实施例中,接头的长度为64个氨基酸。在一些实施例中,接头包含氨基酸序列SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGSSGSETPGTSESATPESSGGSSGGS。在一些实施例中,接头的长度为92个氨基酸。在一些实施例中,接头包含氨基酸序列

[0234] PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATS。

[0235] “标记”是指在与疾病或病症相关的表达水平或活性方面具有改变的任何蛋白质或多核苷酸。

[0236] 如本文所用,术语“突变”是指序列(例如核酸或氨基酸序列)内的残基被另一残基取代,或序列内一个或多个残基的缺失或插入。突变在本文中通常通过鉴定原始残基随后是该残基在序列中的位置,以及通过新取代残基的身份来描述。用于进行本文提供的氨基酸置换(突变)的各种方法是本领域公知的,并且由例如Green和Sambrook,Molecular Cloning:A Laboratory Manual(4th ed.,Cold Spring Harbor Laboratory Press,Cold Spring Harbor,N.Y.(2012))提供。在一些实施例中,当前公开的碱基编辑器可以有效地在核酸(例如个体基因组内的核酸)中产生“预期突变”,例如点突变,而不会产生大量非预期突变,例如意外的点突变。在一些实施例中,预期突变是由与向导多核苷酸(例如gRNA)结合的特定碱基编辑器(例如腺苷碱基编辑器)产生的突变,该向导多核苷酸(例如gRNA)专门设计用于产生预期突变。

[0237] 通常,在序列(例如本文所述的氨基酸序列)中产生或鉴定的突变是相对于参考(或野生型)序列,即不包含突变的序列,进行编号。本领域技术人员将容易理解如何确定氨基酸和核酸序列中相对于参考序列的突变位置。

[0238] 术语“非保守性突变”涉及不同组之间的氨基酸置换,例如,色氨酸为赖氨酸,或丝氨酸为苯丙氨酸等。在这种情况下,非保守性氨基酸置换最好是不干扰、或抑制功能变体的生物活性。非保守性氨基酸置换可增强功能变体的生物活性,从而与野生型蛋白质相比,功能变体的生物活性增加。

[0239] 术语“核定位序列”、“核定位信号”或“NLS”是指促进蛋白质输入细胞核的氨基酸序列。核定位序列是本领域已知的,并且描述于例如Plank等人的国际PCT申请:PCT/EP2000/011690,提交于2000年11月23日,2001年5月31日作为WO/2001/038547公开,其内容通过引用并入本文以公开示例性核定位序列。在其他实施例中,NLS是优化的NLS,例如由Koblan等人于Nature Biotech.2018doi:10.1038/nbt.4172所描述。在一些实施例中,NLS包含选自以下的氨基酸序列:KRTADGSEFESPKKKRKV、KRPAATKKAGQAKKKK、KKTELQTTNAENKTKKL、KRGINDRNFWRGENGRKTR、RKSGKIAAIVVKRPRK、PKKKRKV或MDSLLMNRRKFLYQFKNVRWAKGRRETYLC。

[0240] 如本文所用,术语“核酸”和“核酸分子”是指包含核碱基和酸性部分的化合物,例如核苷、核苷酸或核苷酸的聚合物。通常,聚合核酸,例如包含三个或更多个核苷酸的核酸分子是线性分子,其中相邻核苷酸通过磷酸二酯键相互连接。在一些实施例中,“核酸”是指单个核酸残基(例如核苷酸和/或核苷)。在一些实施例中,“核酸”是指包含三个或更多个单独核苷酸残基的寡核苷酸链。如本文所用,术语“寡核苷酸”和“多核苷酸”可互换使用以指核苷酸的聚合物(例如至少三个核苷酸的链)。在一些实施例中,“核酸”包括RNA以及单链和/或双链DNA。核酸可以是天然存在的,例如在基因组、转录物、mRNA、tRNA、rRNA、siRNA、snRNA、质粒、粘粒、染色体、染色单体或其他天然存在的核酸分子的背景下。另一方面,核酸分子可以是非天然存在的分子,例如重组DNA或RNA、人工染色体、工程基因组或其片段,或合成的DNA、RNA、DNA/RNA杂交体、或包括非天然存在的核苷酸或核苷。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骨架之外的其他骨架的类似物。核酸可以从天然来源纯化、使用重组表达系统产生和任选地纯化、化学合成等。在合适的情况下,例如在化学合成分子的情况下,核酸可以包含核苷类似物,例如具有化学修饰碱基的类似物或糖和骨架修饰。除非另有说明,否则核酸序列以5’到3’方向呈现。在一些实施例中,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、脱氧鸟苷和脱氧胞苷);核苷类似物(例如2-氨基腺苷、2-硫胸苷、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)-甲基鸟苷和2-硫代胞苷);化学经修饰的碱基;生物修饰碱基(例如甲基化碱基);插入的碱基;经修饰的糖(2’-例如氟核糖、核糖、2’-脱氧核糖、阿拉伯糖和己糖);和/或经修饰的磷酸基团(例如硫代磷酸酯和5’-N-亚磷酰胺键)。

[0241] 术语“核酸可编程的DNA结合蛋白”或“napDNAbp”可以与“多核苷酸可编程的核苷酸结合结构域”互换使用,以指与核酸(例如DNA或RNA)相关的蛋白质,例如向导核酸或向导多核苷酸(例如gRNA),将napDNAbp引导至特定的核酸序列。在一些实施例中,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。在一些实施例中,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。在一些实施例中,多核苷酸可编程核苷酸结合结构域是Cas9蛋白。Cas9蛋白可以与向导RNA缔合,该向导RNA将Cas9蛋白引导至与向导RNA互补的特定DNA序列。在一些实施例中,napDNAbp是Cas9结构域,例如核酸酶活性Cas9、Cas9切口酶(nCas9)或核酸酶无活性Cas9(dCas9)。核酸可编程DNA结合蛋白的非限制性实例包括:Cas9(例如dCas9和nCas9)、Cas12a/Cpf1、Cas12b/C2c1、Cas12c/C2c3、Cas12d/CasY、Cas12e/CasX、Cas12g、Cas12h和Cas12i。Cas酶的非限制性实例包括:Cas1、Cas1B、Cas2、Cas3、Cas4、Cas5、Cas5d、Cas5t、Cas5h、Cas5a、Cas6、Cas7、Cas8、Cas8a、Cas8b、Cas8c、Cas9(也被称为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、其同源物或其修饰或工程化版本。其他核酸可编程的DNA结合蛋白也在本公开的范围内,尽管它们可能未在本公开中具体列出。见例如Makarova等人,“Classification and Nomenclature of CRISPR-Cas Systems:Where from Here?”CRISPR J.2018Oct;1:325-336.doi:10.1089/crispr.2018.0033;Yan等人,“Functionally diverse type V CRISPR-Cas systems”Science.2019Jan 4;363(6422):88-91.doi:10.1126/science.aav7271,其各自的全部内容通过引用并入本文。

[0242] 术语“核碱基”、“含氮碱基”或“碱基”在本文中可互换使用,是指形成核苷的含氮生物化合物,核苷又是核苷酸的组分。核碱基形成碱基对并相互堆叠的能力直接导致长链螺旋结构,例如核糖核酸(RNA)和脱氧核糖核酸(DNA)。五种核碱基——腺嘌呤(A)、胞嘧啶(C)、鸟嘌呤(G)、胸腺嘧啶(T)和尿嘧啶(U)——被称为基或典型。腺嘌呤和鸟嘌呤来源于嘌呤,胞嘧啶、尿嘧啶和胸腺嘧啶来源于嘧啶。DNA和RNA还可以包含其他(非主要)经修饰的碱

基。非限制性示例性修饰核碱基可包括次黄嘌呤、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-氢甲基胞嘧啶。次黄嘌呤和黄嘌呤可以通过诱变剂的存在产生,它们都是通过脱氨基(用羰基替换胺基团)产生的。次黄嘌呤可以从腺嘌呤修饰而来。黄嘌呤可由鸟嘌呤修饰。尿嘧啶可由胞嘧啶脱氨基产生。“核苷”由核碱基和五碳糖(核糖或脱氧核糖)组成。核苷的实例包括腺苷、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实例包括肌苷(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(核糖或脱氧核糖)和至少一个磷酸基团组成。

[0243] 如本文所用,术语“核碱基编辑结构域”或“核碱基编辑蛋白”是指可催化RNA或DNA中的核碱基经修饰的蛋白质或酶,例如胞嘧啶(或胞苷)为尿嘧啶(或尿苷)或胸腺嘧啶(或胸苷)和腺嘌呤(或腺苷)到次黄嘌呤(或肌苷)脱氨基,以及非模板化的核苷酸添加和插入。在一些实施例中,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)。在一些实施例中,核碱基编辑结构域可以是天然存在的核碱基编辑结构域。在一些实施例中,核碱基编辑结构域可以是来自天然存在的核碱基编辑结构域的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,例如细菌、人、黑猩猩、大猩猩、猴、牛、狗、大鼠或小鼠。

[0244] 如本文所用,“获得”如同在“获得物质”中,包括合成、购买或以其他方式获得该物质。

[0245] 如本文所用,“患者”或“个体”是指哺乳动物个体或个体,其被诊断患有、有风险患有或患上、或怀疑患有或患上疾病或病症。在一些实施例中,术语“患者”是指哺乳动物个体,其具有高于平均患上疾病或病症的可能性。示例性患者可以是人类、非人类灵长类动物、猫、狗、猪、牛、猫、马、骆驼、美洲驼、山羊、绵羊、啮齿动物(例如小鼠、兔、大鼠或豚鼠)和其他可以受益于本文公开的疗法的哺乳动物。示例性的人类患者可以是男性和/或女性。

[0246] “有需要的患者”或“有需要的个体”在本文中是指被诊断患有或怀疑患有疾病或病症的患者,例如但不限于糖原贮积病1型(GSD1或冯吉尔克病)。

[0247] 术语“致病突变”、“致病变异”、“疾病外壳突变”、“致病变异”、“有害突变”或“易感突变”是指增加个体对某种疾病或病症的易感性的遗传改变或突变。在一些实施例中,致病性突变包含至少一种野生型氨基酸被基因编码的蛋白质中的至少一种致病性氨基酸取代。

[0248] 术语“蛋白质”、“肽”、“多肽”及其语法等价物在本文中可互换使用,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。通常,蛋白质、肽或多肽的长度至少为三个氨基酸。蛋白质、肽或多肽可以指单个蛋白质或蛋白质的集合。蛋白质、肽或多肽中的一个或多个氨基酸可以被修饰,例如,通过添加化学实体例如碳水化合物基团、羟基、磷酸基团、法呢基、异法呢基、脂肪酸基团、用于缀合、功能化或其他经修饰的接头等。蛋白质、肽或多肽也可以是单个分子或可以是多分子复合物。蛋白质、肽或多肽可以只是天然存在的蛋白质或肽的片段。蛋白质、肽或多肽可以是天然存在的、重组的或合成的,或其任何组合。如本文所用,术语“融合蛋白”是指包含来自至少两种不同蛋白质的蛋白质结构域的杂合多肽。一种蛋白质可以位于融合蛋白的氨基末端(N-末端)部分或羧基末端(C-末端)蛋白,从而分别形成氨基末端融合蛋白或羧基末端融合蛋白。蛋白质可以包含不同的结构域,例如,核酸结合结构域(例如引导蛋白质与

靶位点结合的Cas9的gRNA结合结构域)和核酸切割结构域,或核酸编辑蛋白的催化结构域。在一些实施例中,蛋白质包含蛋白质部分,例如构成核酸结合结构域的氨基酸序列,和有机化合物,例如可以充当核酸切割剂的化合物。在一些实施例中,蛋白质与核酸例如RNA或DNA形成复合物或缔合。本文提供的任何蛋白质可以通过本领域已知的任何方法产生。例如,本文提供的蛋白质可以通过重组蛋白质表达和纯化产生,这尤其适用于包含肽接头的融合蛋白质。重组蛋白表达和纯化的方法是众所周知的,包括Green和Sambrook于Molecular Cloning:ALaboratory Manual(4th ed.,Cold Spring Harbor Laboratory Press,Cold Spring Harbor,N.Y.(2012))描述的那些,其全部内容通过引用并入本文。

[0249] 本文公开的多肽和蛋白质(包括其功能部分和功能变体)可包含合成氨基酸代替一种或多种天然存在的氨基酸。此类合成氨基酸是本领域已知的,包括例如氨基环己烷羧酸、正亮氨酸、α-氨基正癸酸、高丝氨酸、S-乙酰氨基甲基-半胱氨酸、反式-3-和反式-4-羟脯氨酸、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-苯丝氨酸、β-羟基苯丙氨酸、苯甘氨酸、α-萘丙氨酸、环己基丙氨酸、环己基甘氨酸、二氢吲哚-2-羧酸、1,2,3,4-四氢异喹啉-3-羧酸、氨基丙二酸、氨基丙二酸单酰胺、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-氨基环戊烷羧酸、α-氨基环己烷羧酸、α-氨基环庚烷羧酸、α-(2-氨基-2-降冰片烷)-羧酸、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结构的一个或多个氨基酸的翻译后修饰相关。翻译后经修饰的非限制性实例包括磷酸化、酰化(包括乙酰化和甲酰化)、糖基化(包括N-连接和O-连接)、酰胺化、羟基化、烷基化(包括甲基化和乙基化)、泛素化、添加吡咯烷酮羧酸、形成二硫键、硫酸化、肉豆蔻酰化、棕榈酰化、异戊二烯化、法呢基化、香叶基化、糖基化、脂酰化和碘化。

[0250] 如本文在蛋白质或核酸的背景下使用的术语"重组"是指在自然界中不存在但为人类工程产物的蛋白质或核酸。例如,在一些实施例中,重组蛋白或核酸分子包含氨基酸或核苷酸序列,该序列与任何自然存在的序列相比,包含至少1个、至少2个、至少3个、至少4个、至少5个、至少6个或至少7个突变。

[0251] "减少"是指至少10%、25%、50%、75%或100%的负变化。

[0252] "参照"是指标准或对照条件。在一个实施例中,参照是野生型或健康细胞。在其他实施例中且不限于,参照是未处理的细胞,其未经受测试条件,或经受安慰剂或生理盐水、培养基、缓冲液和/或不含目标多核苷酸的对照载体。

[0253] "参考序列"是用作序列比较基础的定义序列。参考序列可以是特定序列的子集或全部;例如,全长cDNA或基因序列的片段,或完整的cDNA或基因序列。对于多肽,参考多肽序列的长度通常为至少约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个核苷酸或其附近或之间的任何整数。在一些实施例中,参考序列是目的蛋白质的野生型序列。在其他实施例中,参考序列是编码野生型蛋白质的多核苷酸序列。

[0254] 术语"RNA可编程核酸酶"和"RNA引导的核酸酶"与一种或多种不是切割目标的RNA一起使用(例如结合或缔合)。在一些实施例中,当与RNA形成复合物时,RNA可编程核酸酶可被称为核酸酶:RNA复合物。通常,结合的RNA被称为向导RNA(gRNA)。gRNA可以作为两个或多

个RNA的复合物存在,也可以作为单个RNA分子存在。以单个RNA分子形式存在的gRNA可称为单向导RNA(sgRNA),尽管“gRNA”可互换使用,指以单个分子或两个或多个分子的复合体形式存在的向导RNA。通常,作为单一RNA种类存在的gRNA包含两个结构域:(1)与目标核酸具有同源性的结构域(例如引导Cas9复合物与目标的结合);(2)结合Cas9蛋白的结构域。在一些实施例中,结构域(2)对应于称为tracrRNA的序列,并且包含茎环结构。例如,在一些实施例中,结构域(2)与Jinek等人于Science337:816-821(2012)中提供的tracrRNA相同或同源,其全部内容通过引用并入本文。gRNA的其他示例(例如包括结构域2的那些)可以在美国临时专利申请号U.S.S.N.61/874,682,于2013年9月6日提交,题为“Switchable Cas9 Nucleases and Uses Thereof”中找到,以及美国临时专利申请号U.S.S.N.61/874,746,提交于2013年9月6日,名称为“Delivery System For Functional Nucleases”,其各自的全部内容通过引用整体并入本文。在一些实施例中,gRNA包含域(1)和(2)中的两个或更多个,并且可以被称为“扩展的gRNA”。例如,如本文所述,延伸的gRNA将,例如,结合两个或更多个Cas9蛋白并在两个或更多个不同区域结合目标核酸。gRNA包含与靶位点互补的核苷酸序列,其介导核酸酶/RNA复合物与所述靶位点的结合,提供核酸酶:RNA复合物的序列特异性。

[0255] 在一些实施例中,RNA可编程核酸酶是(CRISPR相关系统)Cas9核酸内切酶,例如来自化脓性链球菌的Cas9(Casnl)(见,例如"Complete genome sequence of an Ml 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).

[0256] 由于RNA可编程核酸酶(例如Cas9)使用RNA:DNA杂交来靶向DNA切割位点,因此这些蛋白质原则上能够靶向向导RNA指定的任何序列。使用RNA可编程核酸酶(例如Cas9)进行位点特异性切割(例如修饰基因组)的方法是本领域已知的(参见例如Cong,L.等人,Multiplex genome engineering using CRISPR/Cas systems.Science339,819-823(2013);Mali,P.等人,RNA-guided human genome engineering via Cas9.Science 339,823-826(2013);Hwang,W.Y.等人,Efficient genome editing in zebrafish using a CRISPR-Cas system.Nature biotechnology 31,227-229(2013);Jinek,M.等人,RNA-programmed genome editing in human cells.eLife 2,e00471(2013);Dicarlo,J.E.等人,Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems.Nucleic acids research(2013);Jiang,W.等人,RNA-guided editing of bacterial genomes using CRISPR-Cas systems.Nature biotechnology 31,233-239(2013);其全部内容通过引用并入本文)。

[0257] 术语“单核苷酸多态性(SNP)”是发生在基因组特定位置的单个核苷酸的变异,其中每个变异在群体中都存在一定程度(例如>1%)。例如,在人类基因组的特定碱基位置,C核苷酸可以出现在大多数个体中,但在少数个体中,该位置被A占据。这意味着在该特定位置存在SNP,并且两个可能的核苷酸变异,C或A,被称为该位置的等位基因。SNP是疾病易感

性差异的基础。疾病的严重程度和我们身体对治疗的反应方式也是遗传变异的表现。SNP可以落入基因的编码区、基因的非编码区或基因间区(基因之间的区域)。在一些实施例中,由于遗传密码的简并性,编码序列内的SNP不一定改变所产生蛋白质的氨基酸序列。编码区的SNP有两种类型:同义SNP和非同义SNP。同义SNP不影响蛋白质序列,非同义SNP改变蛋白质的氨基酸序列。非同义SNP有两种类型:错义和无义。不在蛋白质编码区的SNP仍然可以影响基因剪接、转录因子结合、信使RNA降解或非编码RNA的序列。受此类SNP影响的基因表达称为eSNP(表达SNP),可以位于基因的上游或下游。单核苷酸变异(SNV)是单个核苷酸的变异,没有任何频率限制,可以在体细胞中出现。体细胞单核苷酸变异也可称为单核苷酸改变。

[0258] “特异性结合”是指核酸分子、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,其识别并结合本发明的多肽和/或核酸分子,但基本上不识别和结合样品中的其他分子,例如生物样品。

[0259] 可用于本发明方法的核酸分子包括编码本发明多肽或其片段的任何核酸分子。这样的核酸分子不需要与内源核酸序列100%相同,但通常会表现出基本的同一性。与内源序列具有“基本同一性”的多核苷酸通常能够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多肽或其片段的任何核酸分子。这样的核酸分子不需要与内源核酸序列100%相同,但通常会表现出基本的同一性。与内源序列具有“基本同一性”的多核苷酸通常能够与双链核酸分子的至少一条链杂交。“杂交”是指在各种严格条件下,在互补多核苷酸序列(例如本文所述的基因)或其部分之间配对以形成双链分子。(见例如Wahl,G.M.和S.L.Berger(1987)Methods Enzymol.152:399;Kimmel,A.R.(1987)Methods Enzymol.152:507).

[0260] 例如,严格盐浓度通常低于约750mM氯化钠和75mM柠檬酸三钠,优选低于约500mM氯化钠和50mM柠檬酸三钠,更优选低于约250mM氯化钠和25mM柠檬酸三钠。可以在没有有机溶剂例如甲酰胺的情况下,获得低严格杂交,而可以在至少约35%甲酰胺,更优选至少约50%甲酰胺存在下,获得高严格杂交。严格的温度条件通常包括至少约30℃,更优选至少约37℃,最优选至少约42℃的温度。不同的附加参数,例如杂交时间、去污剂的浓度如十二烷基硫酸钠(SDS),以及载体DNA的包含或排除,是本领域技术人员众所周知的。通过根据需要组合这些不同的条件来实现不同程度的严格性。在一个实施例中,杂交将在30℃下,在750mM氯化钠、75mM柠檬酸三钠和1% SDS中进行。在另一个实施例中,杂交将在37℃下,在500mM氯化钠、50mM柠檬酸三钠、1% SDS、35%甲酰胺和100μg/ml变性鲑鱼精子DNA(ssDNA)中进行。在另一个实施例中,杂交将在42℃下,在250mM氯化钠、25mM柠檬酸三钠、1% SDS、50%甲酰胺和200μg/ml ssDNA中进行。这些条件的有用变化对本领域技术人员来说是显而易见的。

[0261] 对于大多数应用,杂交后的洗涤步骤在严格性方面也会有所不同。洗涤严格条件可以通过盐浓度和温度来定义。如上所述,可以通过降低盐浓度或提高温度来增加洗涤严格性。例如,洗涤步骤的严格盐浓度优选小于约30mM氯化钠和3mM柠檬酸三钠,最优选小于约15mM氯化钠和1.5mM柠檬酸三钠。洗涤步骤的严格温度条件通常包括至少约25℃,更优选至少约42℃,甚至更优选至少约68℃的温度。在一个实施例中,洗涤步骤将在25℃下,在30mM氯化钠、3mM柠檬酸三钠和0.1% SDS中进行。在更优选的实施例中,洗涤步骤将在42℃下,在15mM氯化钠、1.5mM柠檬酸三钠和0.1% SDS中进行。在更优选的实施例中,洗涤步骤

将在68℃下在15mM NaCl、1.5mM柠檬酸三钠和0.1% SDS中进行。这些条件的其他变化对于本领域技术人员来说将是显而易见的。杂交技术是本领域技术人员众所周知的,并且描述于例如Benton和Davis(Science 196:180,1977);Grunstein和Hogness(Proc.Natl.Acad.Sci.,USA 72:3961,1975);Ausubel等人(Current Protocols in Molecular Biology,Wiley Interscience,New York,2001);Berger和Kimmel(Guide to Molecular Cloning Techniques,1987,Academic Press,New York);和Sambrook等人,Molecular Cloning:A Laboratory Manual,Cold Spring Harbor Laboratory Press,New York.

[0262] “分裂”是指分成两个或更多个片段。

[0263] “分裂Cas9蛋白”或“分裂Cas9”是指作为由两个单独的核苷酸序列编码的N-末端片段和C-末端片段提供的Cas9蛋白。对应于Cas9蛋白的N-末端部分和C-末端部分的多肽可以被剪接以形成“重建的”Cas9蛋白。在特定实施例中,Cas9蛋白质在蛋白质的无序区域内被分成两个片段,例如,如Nishimasu等人于Cell,Volume156,Issue 5,pp.935-949,2014所述,或描述于Jiang等人(2016)Science 351:867-871.PDB file:5F9R,各自都通过引用并入本文。在一些实施例中,该蛋白质在SpCas9区域内约氨基酸A292-G364、F445-K483或E565-T637之间的任何C、T、A或S处分成两个片段,或在任何其他Cas9、Cas9变体(例如nCas9、dCas9)或其他napDNAbp。在一些实施例中,蛋白质在SpCas9 T310、T313、A456、S469或C574处被分成两个片段。在一些实施例中,将蛋白质分成两个片段的过程称为“分裂”蛋白质。

[0264] 在其他实施例中,Cas9蛋白的N-末端部分包含化脓性链球菌Cas9野生型(SpCas9)的氨基酸1-573或1-637(NCBI参考序列:NC_002737.2,Uniprot参考序列:Q99ZW2),而Cas9蛋白的C-末端部分包含SpCas9野生型的氨基酸574-1368或638-1368,或其相应位置的一部分。

[0265] 分裂的Cas9的C-末端部分可以与分裂的Cas9的N-末端部分连接以形成完整的Cas9蛋白。在一些实施例中,Cas9蛋白的C-末端部分从Cas9蛋白的N-末端部分结束的地方开始。因此,在一些实施例中,分裂的Cas9的C-末端部分包含spCas9的氨基酸(551-651)-1368的一部分。“(551-651)-1368”是指起始于氨基酸551-651(含)之间的氨基酸并终止于氨基酸1368。例如,分裂Cas9的C-末端部分可以包含任何一个以下spCas9氨基酸的一部分:551-1368、552-1368、553-1368、554-1368、555-1368、556-1368、557-1368、558-1368、559-1368、560-1368、561-1368、562-1368、563-1368、564-1368、565-1368、566-1368、567-1368、568-1368、569-1368、570-1368、571-1368、572-1368、573-1368、574-1368、575-1368、576-1368、577-1368、578-1368、579-1368、580-1368、581-1368、582-1368、583-1368、584-1368、585-1368、586-1368、587-1368、588-1368、589-1368、590-1368、591-1368、592-1368、593-1368、594-1368、595-1368、596-1368、597-1368、598-1368、599-1368、600-1368、601-1368、602-1368、603-1368、604-1368、605-1368、606-1368、607-1368、608-1368、609-1368、610-1368、611-1368、612-1368、613-1368、614-1368、615-1368、616-1368、617-1368、618-1368、619-1368、620-1368、621-1368、622-1368、623-1368、624-1368、625-1368、626-1368、627-1368、628-1368、629-1368、630-1368、631-1368、632-1368、633-1368、634-1368、635-1368、636-1368、637-1368、638-1368、639-1368、640-1368、641-1368、642-1368、643-1368、644-

1368、645-1368、646-1368、647-1368、648-1368、649-1368、650-1368、或651-1368。在一些实施例中,分裂Cas9蛋白的C-末端部分包含SpCas9的氨基酸574-1368或638-1368的一部分。

[0266] “个体”是指哺乳动物,包括但不限于人类或非人类哺乳动物,例如牛、马、犬、绵羊或猫。个体包括家畜、饲养以生产劳动力和提供商品如食物的驯养动物,包括但不限于牛、山羊、鸡、马、猪、兔和绵羊。

[0267] “基本相同”是指多肽或核酸分子与参考氨基酸序列(例如本文所述的任何一种氨基酸序列)或核酸序列(任何一种本文所述的核酸序列)。在一个实施例中,这样的序列与用于比较的序列在氨基酸水平或核酸,有至少60%、80%或85%、90%、95%或甚至99%相同。

[0268] 序列同一性通常使用序列分析软件(例如威斯康星大学生物技术中心遗传学计算机组的序列分析软件包,1710University Avenue,Madison,Wis.53705,BLAST,BESTFIT,COBALT,EMBOSS Needle,GAP,或PILEUP/PRETTYBOX程序)。此类软件通过指定各种替换、缺失和/或其他经修饰的同源性程度来匹配相同或相似的序列。保守替换通常包括以下组内的替换: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;和苯丙氨酸、酪氨酸。在确定同一性程度的示例性方法中,可以使用BLAST程序,其中e^{-3}和e^{-100}之间的概率分数表示密切相关的序列。

[0269] 例如,COBALT与以下参数一起使用:

[0270] a)对齐参数(alignment parameters):Gap penalties-11,-1和End-Gap penalties-5,-1,

[0271] b)CDD参数:Use RPS BLAST on;Blast E-value 0.003;Find Conserved columns and Recompute on,和

[0272] c)Query Clustering参数:Use query clusters on;Word Size4;Max cluster distance 0.8;Alphabet Regular。

[0273] 例如,使用EMBOSS Needle时具有以下参数:

[0274] a)矩陣(Matrix):BLOSUM62;

[0275] b)空位开放(GAP OPEN):10;

[0276] c)空位扩展(GAP EXTEND):0.5;

[0277] d)输出格式(OUTPUT FORMAT):对(pair);

[0278] e)結束空位罚分(END GAP PENALTY):错误(false);

[0279] f)結束空位开放(END GAP OPEN):10;和

[0280] g)結束空位扩展(END GAP EXTEND):0.5。

[0281] 术语“靶位点”是指被核碱基编辑器经修饰的核酸分子内的序列。在一个实施例中,靶位点被脱氨酶或包含脱氨酶(例如腺嘌呤脱氨酶)的融合蛋白脱氨基。

[0282] 如本文所用,术语“治疗(treat)”、“治疗(treating)”、“治疗(treatment)”等是指减少或改善与其相关的病症和/或症状或获得期望的药理学和/或生理学效果。应当理解,尽管不排除,治疗病症或病况并不要求完全消除病症、病况或与其相关的症状。在一些实施例中,该作用是治疗性的,即但不限于,该作用部分地或完全地减少、减弱、消除、减退、减轻、降低疾病的强度,或治愈该疾病和/或可归因于该疾病的不利症状。在一些实施例中,该作用是预防性的,即该作用保护或防止疾病或病症的发生或复发。为此,目前公开的方法包

括施用治疗有效量的如本文所述的组成物。

[0283] “尿嘧啶糖基化酶抑制剂”或“UGI”是指抑制尿嘧啶切除修复系统的物质。在一个实施例中,所述物质是结合宿主尿嘧啶-DNA糖基化酶并防止从DNA中去除尿嘧啶残基的蛋白质或其片段。在一个实施例中,UGI是能够抑制尿嘧啶-DNA糖基化酶碱基切除修复酶的蛋白质、其片段或结构域。在一些实施例中,UGI结构域包含野生型UGI或其修饰版。在一些实施例中,UGI结构域包含下面阐述的示例性氨基酸序列的片段。在一些实施例中,UGI片段包含的氨基酸序列包含下面提供的示例性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%、至少至少95%、至少96%、至少97%、至少98%、至少99%或100%。在一些实施例中,UGI包含与下述示例性UGI氨基酸序列或其片段同源的氨基酸序列。在一些实施例中,UGI或其一部分与下述野生型UGI或UGI序列或其一部分至少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。示例性的UGI包含如下氨基酸序列:

[0284] >splP14739IUNGI_BPPB2尿嘧啶-DNA糖基化酶抑制剂

[0285] MTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML。

[0286] 术语“载体”是指将核酸序列引入细胞中从而产生转化细胞的手段。载体包括质粒、转座子、噬菌体、病毒、脂质体和附加体。“表达载体”是包含待在受体细胞中表达的核苷酸序列的核酸序列。表达载体可以包括额外的核酸序列以促进和/或促进引入序列的表达,例如起始、终止、增强子、启动子和分泌序列。

[0287] 本文提供的任何组成物或方法可以与本文提供的任何其他组成物和方法中的一种或多种组合。

[0288] DNA编辑已成为通过在基因水平上校正致病突变来改变疾病状态的可行手段。直到最近,所有DNA编辑平台的功能都是通过在特定基因组位点诱导DNA双链断裂(DSB)并依靠内源性DNA修复途径以半随机方式确定产物结果,从而产生复杂的基因产物群体。虽然可以通过同源定向修复(HDR)途径实现精确的、用户定义的修复结果,但许多挑战阻碍了在治疗相关细胞类型中使用HDR进行高效修复。在实践中,该途径相对于竞争性、易出错的非同源末端连接途径而言效率低下。此外,HDR被严格限制在细胞周期的G1和S期,阻止了有丝分裂后细胞中DSB的精确修复。因此,已经证明很难或不可能以用户定义的、可编程的方式在这些群体中高效地改变基因组序列。

附图说明

[0289] 本公开的特征在所附权利要求中具体阐述。通过参考以下阐述说明性实施例的详细描述,其中利用了本公开的原理,以及附图,将获得对本发明的特征和优点的更好理解:

[0290] 图1描绘了G6PC核苷酸目标序列和相应的氨基酸序列,指示旁观者(bystander)和目标A>G碱基以校正GSD1a Q347X突变。

[0291] 图2描绘了精确的碱基校正和旁观者编辑。图2A描绘了目标核碱基和旁观者核碱基的位置。图2B描绘了使用ABE8变体的HEK293T细胞中GSD1a G6PC Q347X突变的精确在靶和旁观者校正百分比。

[0292] 图3A和3B描绘了用于校正HEK293T细胞中GSD1a G6PC Q347X突变的编辑器优化。

图3A描绘了G6PC核苷酸目标序列和指示旁观者和在靶A>G碱基上的相应氨基酸序列以及用于校正GSD1a Q347X突变的GGA PAM序列。图3B是描绘使用ABE8单体和异源二聚体变体校正GSD1a G6PC Q347X突变的百分比的图。

[0293] 图4是描绘在HEK293T细胞中使用ABE8双突变变体校正GSD1aG6PC Q347X突变的百分比的图,比较旁观者(A2)和在靶(A6)A>G碱基。

[0294] 图5是描绘在源自患者的B淋巴细胞中使用ABE8变体精确校正GSD1a Q347X突变的百分比的图。

[0295] 图6A和6B描绘了复合杂合(Q347X、G222R)患者iPS衍生肝细胞中GSD1a G6PC Q347X突变的精确校正。图6A描绘了G6PC核苷酸目标序列、相应的氨基酸序列和GGA PAM序列,指示旁观者和在靶A>G碱基以校正GSD1a Q347X突变。图6B是描绘使用ABE8变体比较在靶与旁观者校正的GSD1a Q347X突变的A>G碱基编辑效率的图。

[0296] 图7A和7B描绘了用于校正患者iPS衍生肝细胞中的GSD1a Q347X突变的编辑器优化。图7A显示了GSD1a的NGA PAM序列和对应的目标序列,指示旁观者和在靶A>G碱基。图7B是描绘使用ABE8变体的GSD1a Q347X突变的碱基编辑效率的图。

[0297] 图8A和8B提供了原代肝细胞共培养系统中GSD1a Q347X突变的体外转导时间表。图8A提供了显示代表性时间点的肝细胞单层或肝细胞共培养物中体外转导时间表的时间线。图8B显示了来自用于GSD1aQ347X突变的共培养系统中的供体的转导的原代肝细胞的图像。

[0298] 图9显示了用含有GSD1a Q347X突变的慢病毒载体以30、100和300慢病毒的多重感染(MOI)转导的原代肝细胞共培养细胞中第6天(D6)的GFP表达(GFP,亮视野,合并)的图像。

[0299] 图10A、10B和10C描绘了慢病毒转导的原代肝细胞共培养系统中GSD1a Q347X突变的校正。图10A显示了原代肝细胞共培养细胞(供体RSE)中GFP表达的图像,所述细胞以MOI为500的含有GSD1aQ347X突变的慢病毒载体转导。图10B是描绘转导的原代肝细胞共培养物中GSD1a Q347X突变和插入缺失的在靶校正的A>G碱基编辑效率的图。虚线代表治疗益处的A>G碱基编辑效率。图10C是描绘GSD1a Q347X突变在转导的原代肝细胞共培养物中的A>G碱基编辑效率的图,培养基含有或不含聚乙二醇8000(PEG8K),并用胶原酶III型、IV型和透明质酸处理或保持未经处理。

[0300] 图11描绘了G6PC核苷酸目标序列和相应的氨基酸序列,指示旁观者和在靶A>G碱基以校正GSD1a R83C突变。

[0301] 图12A和12B描绘了HEK293T细胞中GSD1a G6PC R83C突变的精确校正。图12A描绘了G6PC核苷酸目标序列和相应的氨基酸序列,指示旁观者、同义和在靶A>G碱基,用于校正GSD1a R83C突变。图12B是描绘使用ABE8变体的GSD1a R83C突变的A>G碱基编辑效率的图,比较在靶与旁观者校正。

[0302] 图13A和13B描绘了在HEK293T慢病毒模型细胞中通过质粒转染对G6PC R83C突变进行的碱基编辑。图13A显示GSD1a gRNA#820的GAGAAT PAM序列和对应的目标序列以及GSD1a gRNA#1121的AGA PAM序列和对应的目标序列,指示旁观者和目标序列的在靶A>G碱基。图13B是描绘使用具有gRNA1121或gRNA820的ABE碱基编辑器对GSD1aR83C突变的在靶和旁观者校正百分比的图。

[0303] 图14是描绘使用saABE8变体的GSD1a R83C突变的A>G碱基编辑效率的图。

[0304] 图15是描绘使用saABE8双突变变体的GSD1a R83C突变的A>G碱基编辑效率的图。

[0305] 图16是描绘在HEK293T细胞中使用ABE8变体对GSD1a R83C突变的在靶、旁观者和同义旁观者校正的A>G碱基编辑效率的图。

[0306] 图17A和17B描绘了从GSD1a转基因小鼠模型分离的原代小鼠肝细胞中GSD1a Q347X突变的校正。图17A显示了从ASC转基因小鼠模型huG6PC、R83C(V166L)中分离的原代小鼠肝细胞的图像。图17B是描绘位置A12G、A10G、A6G和插入缺失的碱基编辑效率的图,用于校正从使用ABE8变体的GSD1a转基因小鼠模型分离的原代小鼠肝细胞中的GSD1a R83C突变。

[0307] 图18是描绘使用TadA-SaCas9 ABE编辑器结合如图所示的不同长度的向导RNA在靶(12A)和脱靶(6A)位点处的A>G碱基编辑水平的图。在HEK293T细胞中获得数据。还提供了靶位点和其他编辑详细信息。

[0308] 图19是描绘使用ABE8s(TadA*8变体-SaCas9)结合20nt和21nt向导RNA在靶(12A)和脱靶(6A)位点的A>G碱基编辑(百分比编辑)水平的图。在HEK293T细胞中获得数据。

[0309] 图20是描绘使用ABE碱基编辑器(TadA变体-SaCas9)结合20nt或21nt向导RNA在靶(12A)和脱靶(6A)位点的A>G碱基编辑水平(R83C的校正百分比)的图。在HEK293T细胞中获得数据。

[0310] 图21是描绘使用ABE碱基编辑器(TadA变体-SaCas9)结合20nt或21nt向导RNA在靶(12A)和脱靶(6A)位点处的A>G(%)碱基编辑水平的图。在GSD1a R83C的原代人肝细胞慢病毒模型中获得数据。

[0311] 图22是描绘使用ABE碱基编辑器(TadA变体-SaCas9)结合20nt或21nt向导RNA在靶(12A)和脱靶(6A)位点处的A>G碱基编辑水平(R83C的校正百分比)的图。在GSD1a R83C的原代人肝细胞慢病毒模型中获得数据。

[0312] 图23是描绘杂合转基因GSD1a R83C小鼠中在靶位点和脱靶位点的A>G(%)精确碱基编辑水平的图。

[0313] 图24是描述用于访问NRNN PAM的所有可能PAM的Cas9变体的表。仅列出了需要识别其PAM中三个或更少定义核苷酸的Cas9变体。非G PAM变体包括SpCas9-NRRH、SpCas9-NRTH和SpCas9-NRCH。

具体实施方式

[0314] 本发明提供了包含具有提高的效率的新型腺苷碱基编辑器(例如ABE8)的组成物和使用包含腺苷脱氨酶变体的碱基编辑器来改变与1a型糖原贮积病(GSD1a)相关的突变的方法。

[0315] 本发明至少部分基于以下发现:以腺苷脱氨酶变体(即ABE8)为特征的碱基编辑器可精确校正内源性葡萄糖6-磷酸酶(G6PC)基因(例如R83C、Q347X)中的单核苷酸多态性。

[0316] GSD1a突变R83C和Q347X是胞苷到胸苷(C→T)的转换突变,导致C·G到T·A碱基对的替换。这些替换可以恢复为带有腺苷碱基编辑器(ABE)的野生型、非致病基因组序列,ABE可催化A·T到G·C的替换。通过扩展,GSD1a引起的突变是使用ABE恢复为野生型序列的潜在目标,而不会像使用基因治疗那样导致G6PC基因过度表达的风险。因此,A·T到G·C DNA碱基编辑精确地校正了G6PC基因中一个或多个最普遍的由GSD1a引起的突变。

[0317] 核碱基编辑器

[0318] 本文公开了用于编辑、修饰或改变多核苷酸的目标核苷酸序列的碱基编辑器或核碱基编辑器。本文描述了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)的核碱基编辑器或碱基编辑器。多核苷酸可编程核苷酸结合结构域,当与结合的向导多核苷酸(例如gRNA)结合时,可以特异性结合目标多核苷酸序列(即通过结合的向导核酸的碱基和目标多核苷酸的碱基之间的互补碱基配对序列),从而将碱基编辑器靶向需要编辑的目标核酸序列。在一些实施例中,所述目标多核苷酸序列包含单链DNA或双链DNA。在一些实施例中,所述目标多核苷酸序列包含RNA。在一些实施例中,所述目标多核苷酸序列包含DNA-RNA杂交体。

[0319] 多核苷酸可编程核苷酸结合结构域

[0320] 应当理解,多核苷酸可编程核苷酸结合结构域还可以包括结合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将多核苷酸可编程核苷酸结合结构域引导至RNA的核酸缔合。其他核酸可编程的DNA结合蛋白也在本公开的范围内,尽管它们没有在本公开中具体列出。

[0321] 碱基编辑器的多核苷酸可编程核苷酸结合结构域本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包含一个或多个核酸酶结构域。在一些实施例中,所述多核苷酸可编程核苷酸结合结构域的核酸酶结构域可包含核酸内切酶或核酸外切酶。在本文中,术语“外切核酸酶”是指能够从游离末端分解核酸(例如RNA或DNA)的蛋白质或多肽,而术语“内切核酸酶”是指能够催化(例如切割)核酸(例如DNA或RNA)中内部区域的蛋白质或多肽。在一些实施例中,所述核酸内切酶可以切割双链核酸的单链。在一些实施例中,所述内切核酸酶可以切割双链核酸分子的两条链。在一些实施例中,所述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。在一些实施例中,所述多核苷酸可编程核苷酸结合结构域可以是核糖核酸酶。

[0322] 在一些实施例中,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目标多核苷酸的零、一条或两条链。在一些实施例中,所述多核苷酸可编程核苷酸结合结构域可包含切口酶结构域。在本文中,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,该核酸酶结构域能够仅切割双链核酸分子(例如DNA)中的两条链中的一条链。在一些实施例中,切口酶通过将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,可以衍生自完全催化活性(例如天然的)形式的多核苷酸可编程核苷酸结合结构域。例如,在多核苷酸可编程核苷酸结合结构域包含源自Cas9的切口酶结构域的情况下,源自Cas9的切口酶结构域可在位置840处包含D10A突变和组氨酸。在此类实施例中,所述残基H840保留催化活性并由此可切割核酸双链体的单链。在另一个实施例中,源自Cas9的切口酶结构域可包含H840A突变,而位置10处的氨基酸残基仍为D。在一些实施例中,切口酶通过去除切口酶活性不需要的全部或部分核酸酶结构域,可以衍生自完全催化活性(例如天然的)形式的多核苷酸可编程核苷酸结合结构域。例如,在多核苷酸可编程核苷酸结合结构域包含源自Cas9的切口酶结构域的情况下,所述源自Cas9的切口酶结构域可包含RuvC结构域或HNH结构域的全部或部分的缺失。

[0323] 示例性催化活性Cas9的氨基酸序列如下:

[0324] MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA

RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVD
STDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSR
RLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNL
SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNEL
TKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMG
RHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAE
RGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGE
IRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYG
GFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELEN
GRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD
ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI
DLSQLGGD.

[0325] 包含了包含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因此能够在特定多核苷酸目标序列(例如由结合的向导核酸的互补序列确定)处产生单链DNA断裂(切口)。在一些实施例中,被包含切口酶结构域(例如Cas9衍生的切口酶结构域)的碱基编辑器切割的核酸双链体目标多核苷酸序列的链是未被碱基编辑器编辑的链(即被碱基编辑器切割的链与包含要编辑的碱基的链相反)。在其他实施例中,包含切口酶结构域(例如Cas9衍生的切口酶结构域)的碱基编辑器可以切割被靶向进行编辑的DNA分子的链。在此类实施例中,非目标链未被切割。

[0326] 本文还提供了碱基编辑器,其包含催化死亡(即不能切割目标多核苷酸序列)的多核苷酸可编程核苷酸结合结构域。在本文中,术语“催化死亡”和“核酸酶死亡”可互换使用,指具有一个或多个突变和/或缺失导致其无法切割核酸链的多核苷酸可编程核苷酸结合结构域。在一些实施例中,催化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由于一个或多个核酸酶结构域中的特定点突变而缺乏核酸酶活性。例如,在碱基编辑器包含Cas9结构域的情况下,Cas9可以包含D10A突变和H840A突变。此类突变使两个核酸酶结构域失活,从而导致核酸酶活性丧失。在其他实施例中,催化死亡的多核苷酸可编程核苷酸结合结构域可包含催化结构域(例如RuvC1和/或HNH结构域)的全部或部分的一个或多个缺失。在进一步的实施例中,催化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)以及全部或部分核酸酶结构域的缺失。

[0327] 本文还考虑了能够从多核苷酸可编程核苷酸结合结构域的先前功能版本产生催化死亡的多核苷酸可编程核苷酸结合结构域的突变。例如,在催化死亡的Cas9(“dCas9”)的情况下,提供了具有除D10A和H840A以外的突变的变体,这导致核酸酶失活的Cas9。例如,此类突变包括在D10和H840处的其他氨基酸置换,或Cas9核酸酶结构域内的其他置换(例如

HNH核酸酶子结构域和/或RuvC1子结构域中的置换)。基于本公开内容和本领域的知识,其他合适的核酸酶无活性的dCas9结构域对于本领域技术人员来说是显而易见的,并且在本公开内容的范围内。这种额外的示例性合适的核酸酶无活性的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等人,CAS9transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering.Nature Biotechnology.2013;31(9):833-838,其全部内容通过引用并入本文)。

[0328] 可并入碱基编辑器的多核苷酸可编程核苷酸结合结构域的非限制性实例包括CRISPR蛋白衍生结构域、限制性核酸酶、大范围核酸酶、TAL核酸酶(TALEN)和锌指核酸酶(ZFN)。在一些实施例中,碱基编辑器包含多核苷酸可编程核苷酸结合结构域,其包含天然或经修饰的蛋白质或其部分,通过结合的向导核酸,其能够在CRISPR(即,成簇的规则间隔短回文重复序列)-介导的核酸修饰期间结合核酸序列。这种蛋白质在本文中称为“CRISPR蛋白质”。因此,本文公开了包含多核苷酸可编程核苷酸结合结构域的碱基编辑器,所述多核苷酸可编程核苷酸结合结构域包含CRISPR蛋白的全部或部分(即包含CRISPR蛋白的全部或部分作为结构域的碱基编辑器,也称为所述碱基编辑器的“CRISPR蛋白衍生结构域”)。与CRISPR蛋白的野生型或天然版本相比,可以对掺入碱基编辑器的CRISPR蛋白衍生结构域进行修饰。例如,如下所述,CRISPR蛋白衍生结构域可包含一个或多个相对于CRISPR蛋白的野生型或天然形式的突变、插入、缺失、重排和/或重组。

[0329] CRISPR是一种适应性免疫系统,可针对移动遗传组件(病毒、转座组件和接合质粒)提供保护。CRISPR簇包含间隔区、与先行移动组件互补的序列和目标入侵核酸。CRISPR簇被转录并加工成CRISPR RNA(crRNA)。在II型CRISPR系统中,pre-crRNA的正确加工需要转编码的小RNA(tracrRNA)、内源性核糖核酸酶3(rnc)和Cas9蛋白。tracrRNA用作核糖核酸酶3辅助加工前crRNA的指引。随后,Cas9/crRNA/tracrRNA核酸内切切割与间隔区互补的线性或环状dsDNA目标。与crRNA不互补的目标链首先被核酸内切切割,然后3′-5′核酸外切修剪。在自然界中,DNA结合和切割通常需要蛋白质和两种RNA。然而,可以对单向导RNA(“sgRNA”或简称为“gRNA”)进行工程改造,以便将crRNA和tracrRNA的各个方面整合到单个RNA种类中。参见例如Jinek M.,Chylinski K.,Fonfara I.,Hauer M.,Doudna J.A.,Charpentier E.Science 337:816-821(2012),其全部内容通过引用并入本文。Cas9识别CRISPR重复序列(PAM或原型间隔区相邻基序)中的短基序,以帮助区分自我与非自我。

[0330] 在一些实施例中,本文所述的方法可以利用工程化的Cas蛋白。向导RNA(gRNA)是一种短的合成RNA,由Cas结合所需的支架序列和用户定义的~20个核苷酸间隔区组成,该间隔区定义了要修改的基因组目标。因此,技术人员可以改变Cas蛋白的基因组目标特异性,部分取决于gRNA靶向序列与基因组其余部分相比,对基因组目标的特异性。

[0331] 在一些实施例中,gRNA支架序列如下:GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUUUU.

[0332] 在一些实施例中,掺入碱基编辑器的CRISPR蛋白衍生结构域是当与结合的向导核酸结合时能够结合目标多核苷酸的核酸内切酶(例如脱氧核糖核酸酶或核糖核酸酶)。在一些实施例中,掺入碱基编辑器的CRISPR蛋白衍生结构域是当与结合的向导核酸结合时能够结合目标多核苷酸的切口酶。在一些实施例中,掺入碱基编辑器的CRISPR蛋白衍生结构域

是当与结合的向导核酸结合时能够结合目标多核苷酸的催化死亡结构域。在一些实施例中,由碱基编辑器的CRISPR蛋白衍生结构域结合的目标多核苷酸是DNA。在一些实施例中,由碱基编辑器的CRISPR蛋白衍生结构域结合的目标多核苷酸是RNA。

[0333] 可用于本文的Cas蛋白包括第1类和第2类。Cas蛋白的非限制性实例包括:Cas1、Cas1B、Cas2、Cas3、Cas4、Cas5、Cas5d、Cas5t、Cas5h、Cas5a、Cas6、Cas7、Cas8、Cas9(also known as Csn1 or 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、其同系物,或其修改版本。未经经修饰的CRISPR酶可以具有DNA切割活性,例如Cas9,它具有两个功能性核酸内切酶结构域:RuvC和HNH。CRISPR酶可以指导切割目标序列处的一条或两条链,例如在目标序列内和/或在目标序列的互补序列内。例如,CRISPR酶可以指导切割目标序列从第一个或最后一个核苷酸开始的约1、2、3、4、5、6、7、8、9、10、15、20、25、50、100、200、500,或更多个碱基对。

[0334] 可以使用编码CRISPR酶的载体,所述CRISPR酶相对于相应的野生型酶突变,使得突变的CRISPR酶缺乏切割含有目标序列的目标多核苷酸的一条或两条链的能力。Cas9可指具有与野生型示例性Cas9多肽(例如来自化脓性链球菌的Cas9)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可以指具有与野生型示例性Cas9多肽(例如来自化脓性链球菌)最多或最多约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可指Cas9蛋白的野生型或修饰形式,其可包含如缺失、插入、取代、变体、突变、融合、嵌合体或其任何组合的氨基酸变化。

[0335] 在一些实施例中,碱基编辑器的CRISPR蛋白衍生结构域可包括来自以下的Cas9的全部或部分:溃疡棒状杆菌(NCBI Refs:NC_015683.1,NC_017317.1);白喉棒状杆菌(NCBI Refs:NC_016782.1,NC_016786.1);栖蚜蝇螺原体(NCBI Ref:NC_021284.1);中间普氏菌(NCBI Ref:NC_017861.1);中国台湾螺原体(NCBI Ref:NC_021846.1);链球菌(NCBI Ref:NC_021314.1);波罗的海贝尔氏菌(NCBI Ref:NC_018010.1);旋螺冷弯曲菌(NCBI Ref:NC_018721.1);嗜热链球菌(NCBI Ref:YP_820832.1);无害李斯特菌(NCBI Ref:NP_472073.1);空肠弯曲菌(NCBI Ref:YP_002344900.1);脑膜炎奈瑟菌(NCBI Ref:YP_002342100.1)、化脓性链球菌或金黄色葡萄球菌。

[0336] 核碱基编辑器的Cas9结构域

[0337] Cas9核酸酶序列和结构是本领域技术人员众所周知的(参见例如“Complete genome sequence of an Ml strain of Streptococcus pyogenes.”Ferretti等人,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"Aprogrammabledual-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),其全部内容通过引用并入本文)。Cas9直向同源物已在各种物种中得到描述,包括但不限于化脓链球菌和嗜热链球菌。基于本公开内容,其他合适的Cas9核酸酶和序列对于本领域技术人员来说将是显而易见的,并且此类Cas9核酸酶和序列包括来自Chylinski、Rhun和Charpentier,"The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems"(2013)RNA Biology 10:5,726-737;中公开的生物体和基因座的Cas9序列,其全部内容通过引用并入本文。

[0338] 在一些实施例中,核酸可编程DNA结合蛋白(napDNAbp)是Cas9结构域。本文提供了非限制性的示例性Cas9结构域。所述Cas9结构域可以是核酸酶活性Cas9结构域、核酸酶无活性Cas9结构域(dCas9)或Cas9切口酶(nCas9)。在一些实施例中,所述Cas9结构域是核酸酶活性结构域。例如,所述Cas9结构域可以是切割双链核酸的两条链(例如双链DNA分子的两条链)的Cas9结构域。在一些实施例中,所述Cas9结构域包含如本文所述的任一氨基酸序列。在一些实施例中,所述Cas9结构域包含与本文所述的任一氨基酸序列至少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%相同的氨基酸序列。在一些实施例中,所述Cas9结构域包含与本文列出的任一氨基酸序列相比,具有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、50或更多个突变的氨基酸序列。在一些实施例中,所述Cas9结构域包含与本文所述的任一氨基酸序列相比,具有至少10、至少15、至少20、至少30、至少40、至少50、至少60、至少70、至少80、至少90、至少100、至少150、至少200、至少250、至少300、至少350、至少400、至少500、至少600、至少700、至少800、至少900、至少1000、至少1100或至少1200个相同的连续氨基酸残基。

[0339] 在一些实施例中,提供了包含Cas9片段的蛋白质。例如,在一些实施例中,蛋白质包含两个Cas9结构域之一:(1)Cas9的gRNA结合结构域;或(2)Cas9的DNA切割结构域。在一些实施例中,包含Cas9或其片段的蛋白质被称为"Cas9变体"。Cas9变体与Cas9或其片段具有同源性。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。在一些实施例中,Cas9变体与野生型Cas9相比,可具有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、50或更多的氨基酸变化。在一些实施例中,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。在一些实施例中,该片段为对应野生型Cas9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%相同、至

少96％、至少97％、至少98％、至少99％或至少99.5％。在一些实施例中，该片段的长度为至少100个氨基酸。在一些实施例中，该片段是至少100、150、200、250、300、350、400、450、500、550、600、650、700、750、800、850、900、950、1000、1050、1100、1150、1200、1250或至少1300个氨基酸的长度。

[0340] 在一些实施例中，本文提供的Cas9融合蛋白包含Cas9蛋白的全长氨基酸序列，例如本文提供的Cas9序列之一。然而，在其他实施例中，本文提供的融合蛋白不包含全长Cas9序列，而仅包含其一个或多个片段。本文提供了合适的Cas9结构域和Cas9片段的示例性氨基酸序列，并且Cas9结构域和片段的其他合适的序列对本领域技术人员来说是显而易见的。

[0341] Cas9蛋白可以与向导RNA缔合，该向导RNA将Cas9蛋白引导至与向导RNA互补的特定DNA序列。在一些实施例中，所述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)或核酸酶无活性Cas9(dCas9)。核酸可编程DNA结合蛋白的例子包括但不限于Cas9(例如dCas9和nCas9)、CasX、CasY、Cpf1、Cas12b/C2c1和Cas12c/C2c3。在一些实施例中，野生型Cas9对应于来自化脓性链球菌的Cas9(NCBI参考序列:NC_017053.1，核苷酸和氨基酸序列如下)。

[0342] ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGATCACTGAT
GATTATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTATCAAAAAAAATCTTATAGG
GGCTCTTTTATTTGGCAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTC
GGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGA
CTTGAAGAGTCTTTTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAAGT
TGCTTATCATGAGAAATATCCAACTATCTATCATCTGCGAAAAAAATTGGCAGATTCTACTGATAAAGCGGATTTGC
GCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAGATTTAAATCCTGAT
AATAGTGATGTGGACAAACTATTTATCCAGTTGGTACAAATCTACAATCAATTATTTGAAGAAAACCCTATTAACGC
AAGTAGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGCTCAGC
TCCCCGGTGAGAAGAGAAATGGCTTGTTTGGGAATCTCATTGCTTTGTCATTGGGATTGACCCCTAATTTTAAATCA
AATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTAGATAATTTATTGGC
GCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGATGCTATTTTACTTTCAGATATCC
TAAGAGTAAATAGTGAAATAACTAAGGCTCCCCTATCAGCTTCAATGATTAAGCGCTACGATGAACATCATCAAGAC
TTGACTCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTATAAAGAAATCTTTTTTGATCAATCAAAAAA
CGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAAATTTATCAAACCAATTTTAGAAAAAA
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

MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFG
[0343] NIVDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD
VDKLFIQLVQIYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLFGN
LIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[0344] LLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE
VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL
SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDRGMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGHSL
HEQIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQTTQKGQKNSRERM
KRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHI
VPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT
KAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSK
LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKM
IAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFA
TVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAY
SVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKY
SLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ
HKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAP
AAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0345] (单下划线:HNH结构域;双下划线:RuvC结构域)

[0346] 在一些实施例中,野生型Cas9对应于或包含以下核苷酸和/或氨基酸序列:

[0347] 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

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFG
NIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD
VDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGN
LIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE
VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL
SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
[0348] RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL
HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER
MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDH
IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK
MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF
ATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVA
YSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE
QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA
PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0349] (单下划线:HNH结构域;双下划线:RuvC结构域)。

[0350] 在一些实施例中,野生型Cas9对应于来自化脓性链球菌的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):

[0351] 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

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFG
NIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD
VDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGN
LIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE
VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL
SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
[0352] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL
HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER
MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDH
IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK
MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF
ATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVA
YSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE
QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA
[0353] PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (单下划线:HNH

结构域;双下划线:RuvC结构域)

[0354] 在一些实施例中,Cas9是指来自以下的Cas9:溃疡棒状杆菌(NCBI Refs:NC_015683.1、NC_017317.1);白喉棒状杆菌(NCBI Refs:NC_016782.1,NC_016786.1);栖蚜蝇螺原体(NCBI Ref:NC_021284.1);中间普氏菌(NCBI Ref:NC_017861.1);中国台湾螺原体(NCBI Ref:NC_021846.1);链球菌(NCBI Ref:NC_021314.1);波罗的海贝尔氏菌(NCBI Ref:NC_018010.1);旋螺冷弯曲菌I(NCBI Ref:NC_018721.1);嗜热链球菌(NCBI Ref:YP_820832.1)、无害李斯特菌(NCBI Ref:NP_472073.1)、空肠弯曲菌(NCBI Ref:YP_002344900.1)或脑膜炎奈瑟菌(NCBI Ref.:YP_002342100.1)或来自任何其他生物体的Cas9。

[0355] 应当理解,额外的Cas9蛋白(例如核酸酶死亡Cas9(dCas9)、Cas9切口酶(nCas9)或核酸酶活性Cas9),包括其变体和同系物,在本公开的范围内。示例性的Cas9蛋白包括但不限于以下提供的那些。在一些实施例中,Cas9蛋白是核酸酶死亡的Cas9(dCas9)。在一些实施例中,Cas9蛋白是Cas9切口酶(nCas9)。在一些实施例中,Cas9蛋白是核酸酶活性的Cas9。

[0356] 在一些实施例中,Cas9结构域是核酸酶无活性的Cas9结构域(dCas9)。例如,dCas9结构域可以结合双链核酸分子(例如通过gRNA分子)而不切割双链核酸分子的任何一条链。在一些实施例中,核酸酶失活的dCas9结构域包含本文所列氨基酸序列的D10X突变和H840X突变,或本文提供的任何氨基酸序列中的相应突变,其中X是任何氨基酸变化。在一些实施例中,核酸酶无活性的dCas9结构域包含本文所述氨基酸序列的D10A突变和H840A突变,或本文提供的任何氨基酸序列中的相应突变。作为一个例子,核酸酶无活性的Cas9结构域包含在克隆载体pPlatTET-gRNA2(登录号BAV54124)中列出的氨基酸序列。

[0357] 示例性无催化活性的Cas9(dCas9)的氨基酸序列如下:

[0358] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI

DLSQLGGD(见例如Qi等人,“Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression.”Cell.2013;152(5):1173-83,其全部内容通过引用并入本文)。

[0359] 基于本公开内容和本领域知识,其他合适的核酸酶失活dCas9结构域对于本领域技术人员来说将是显而易见的,并且在本公开内容的范围内。这种额外的示例性合适的核酸酶无活性的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(见例如Prashant等人,CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering.Nature Biotechnology.2013;31(9):833-838,其全部内容通过引用并入本文)。

[0360] 在一些实施例中,Cas9核酸酶具有无活性(例如失活的)DNA切割结构域,即,Cas9是切口酶,称为“nCas9”蛋白(对于“切口酶”Cas9)。核酸酶无活性的Cas9蛋白可互换地称为“dCas9”蛋白(对于核酸酶-“死的”Cas9)或无催化活性的Cas9。用于产生具有无活性DNA切割结构域的Cas9蛋白(或其片段)的方法是已知的(参见例如Jinek等人,Science.337:816-821(2012);Qi等人,“Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression”(2013)Cell.28;152(5):1173-83,其全部内容通过引用并入本文)。例如,已知Cas9的DNA切割结构域包括两个子结构域,HNH核酸酶子结构域和RuvC1子结构域。HNH子结构域切割与gRNA互补的链,而RuvC1子结构域切割非互补链。这些子结构域内的突变可以使Cas9的核酸酶活性沉默。例如,突变D10A和H840A使化脓性链球菌Cas9的核酸酶活性完全失活(Jinek等人,Science.337:816-821(2012);Qi等人,Cell.28;152(5):1173-83(2013))。

[0361] 在一些实施例中,dCas9结构域包含与本文提供的任何一个dCas9结构域相比,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%相同的氨基酸序列。在一些实施例中,Cas9结构域包含与本文所列氨基酸序列中的任一者相比,具有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、50或更多个突变的氨基酸序列。在一些实施例中,Cas9结构域包含与本文所述的任一氨基酸序列相比,具有至少10、至少15、至少20、至少30、至少40、至少50、至少60、至少70、至少80、至少90、至少100、至少150、至少200、至少250、至少300、至少350、至少400、至少500、至少600、至少700、至少800、至少900、至少1000、至少1100或至少1200个相同的连续氨基酸残基。

[0362] 在一些实施例中,dCas9对应于,或部分或全部包含具有一个或多个使Cas9核酸酶活性失活的突变的Cas9氨基酸序列。例如,在一些实施例中,dCas9结构域包含D10A和H840A突变或另一个Cas9中的相应突变。

[0363] 在一些实施例中,dCas9包含dCas9(D10A和H840A)的氨基酸序列:

[0364] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFG
NIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD
VDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGN
LIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE
VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL
SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL
HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER
MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDA
IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK
MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF
ATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVA
YSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE
QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA
PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (单下划线:HNH结构域;双下划线:RuvC结构域)。

[0365] 在一些实施例中,Cas9结构域包含D10A突变,而位置840处的残基在上文提供的氨基酸序列中或在本文提供的任何氨基酸序列中的相应位置处仍为组氨酸。

[0366] 在其他实施例中,提供了具有除D10A和H840A之外的突变的dCas9变体,其例如导致核酸酶失活的Cas9(dCas9)。例如,此类突变包括在D10和H840处的其他氨基酸置换,或Cas9核酸酶结构域内的其他置换(例如HNH核酸酶子结构域和/或RuvC1子结构域中的置换)。在一些实施例中,提供了dCas9的变体或同源物,其至少约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。在一些实施例中,提供了dCas9的变体,其具有以下较短或较长氨基酸序列: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。

[0367] 在一些实施例中,所述Cas9结构域是Cas9切口酶。Cas9切口酶可以是仅能够切割双链核酸分子(例如双链DNA分子)的一条链的Cas9蛋白。在一些实施例中,所述Cas9切口酶切割双链核酸分子的目标链,意味着Cas9切口酶切割与结合至Cas9的gRNA(例如sgRNA)碱基配对(互补)的链。在一些实施例中,所述Cas9切口酶包含D10A突变并且在位置840处具有组氨酸。在一些实施例中,所述Cas9切口酶切割双链核酸分子的非目标、非碱基编辑链,意味着Cas9切口酶切割不与结合到Cas9的gRNA(例如sgRNA)碱基配对的链。在一些实施例中,所述Cas9切口酶包含H840A突变并且在位置10处具有天冬氨酸残基,或相应的突变。在一些实施例中,Cas9切口酶包含与本文提供的任何一种Cas9切口酶至少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%相同的氨基酸序列。基于本公开内容和本领域知识,其他合适的Cas9切口酶对于本领域技术人员来说是显而易见的,并且在本公开内容的范围内。

[0368] 示例性催化Cas9切口酶(nCas9)的氨基酸序列如下:

[0369] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKP
IREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0370] 在一些实施例中,所述Cas9是指来自古细菌(例如纳米古细菌)的Cas9,其构成单细胞原核微生物的域和界。在一些实施例中,所述核酸可编程DNA结合蛋白是指CasX或CasY,其已在例如Burstein等人的"New CRISPR-Cas systems from uncultivated microbes."中描述,Cell Res.2017Feb 21.doi:10.1038/cr.2017.21,其全部内容通过引用并入本文。使用基因组解析的宏基因组学,确定了许多CRISPR-Cas系统,包括在古细菌领域首次报道的Cas9。这种各有差异的Cas9蛋白是在鲜为人知的纳米古细菌中发现的,作为活性CRISPR-Cas系统的一部分。在细菌中,发现了两个以前未知的系统,CRISPR-CasX和CRISPR-CasY,它们是迄今为止发现的最紧凑的系统之一。在一些实施例中,在本文所述的

碱基编辑器系统中,Cas9被CasX或CasX的变体替代。在一些实施例中,在本文所述的碱基编辑器系统中,Cas9被CasY或CasY的变体替代。应当理解,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),并且在本公开的范围内。

[0371] 在一些实施例中,本文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。在一些实施例中,所述napDNAbp是CasX蛋白。在一些实施例中,napDNAbp是CasY蛋白。在一些实施例中,所述napDNAbp包含的氨基酸序列为至少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%与天然存在的CasX或CasY蛋白相同。在一些实施例中,所述可编程核苷酸结合蛋白是天然存在的CasX或CasY蛋白。在一些实施例中,所述可编程核苷酸结合蛋白包含与本文所述的任何CasX或CasY蛋白至少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%相同的氨基酸序列。应当理解,根据本公开也可以使用来自其他细菌物种的CasX和CasY。

[0372] 示例性的CasX((uniprot.org/uniprot/F0NN87;uniprot.org/uniprot/F0NH53)tr|F0NN87|F0NN87_SULIHCRISPR相关的Casx蛋白质OS＝冰岛硫化叶菌(菌株HVE10/4)GN＝SiH_0402PE＝4SV＝1)氨基酸序列如下:

[0373] MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAERRGKAKKKKGEEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKECEEVSAPSFVKPEFYEFGRSPGMVERTRRVKLEVEPHYLIIAAAGWVLTRLGKAKVSEGDYVGVNVFTPTRGILYSLIQNVNGIVPGIKPETAFGLWIARKVVSSVTNPNVSVVRIYTISDAVGQNPTTINGGFSIDLTKLLEKRYLLSERLEAIARNALSISSNMRERYIVLANYIYEYLTGSKRLEDLLYFANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG.

[0374] 示例性的CasX(>tr|F0NH53|F0NH53_SULIR CRISPR相关的蛋白质,Casx OS＝冰岛硫化叶菌(菌株REY15A)GN＝SiRe_0771PE＝4SV＝1)氨基酸序列如下:

[0375] MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAERRGKAKKKKGEEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKECEEVSAPSFVKPEFYKFGRSPGMVERTRRVKLEVEPHYLIMAAAGWVLTRLGKAKVSEGDYVGVNVFTPTRGILYSLIQNVNGIVPGIKPETAFGLWIARKVVSSVTNPNVSVVSIYTISDAVGQNPTTINGGFSIDLTKLLEKRDLLSERLEAIARNALSISSNMRERYIVLANYIYEYLTGSKRLEDLLYFANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG.

[0376] δ变形杆菌CasX

[0377] MEKRINKIRKKLSADNATKPVSRSGPMKTLLVRVMTDDLKKRLEKRRKKPEVMPQVISNNAANNLRMLLDDYTKMKEAILQVYWQEFKDDHVGLMCKFAQPASKKIDQNKLKPEMDEKGNLTTAGFACSQCGQPLFVYKLEQVSEKGKAYTNYFGRCNVAEHEKLILLAQLKPVKDSDEAVTYSLGKFGQRALDFYSIHVTKESTHPVKPLAQIAGNRYASGPVGKALSDACMGTIASFLSKYQDIIIEHQKVVKGNQKRLESLRELAGKENLEYPSVTLPPQPHTKEGVDfAYNEVIARVRMWVNLNLWQKLKLSRDDAKPLLRLKGFPSFPVVERRENEVDWWNTINEVKKLIDAKRDMGRVFWSGVTAEKRNTILEGYNYLPNENDHKKREGSLENPKKPAKRQFGDLLLYLEKKYAGDWGKVFDEAWERIDKKIAGLTSHIEREEARNAEDAQSKAVLTDWLRAKASFVLERLKEMDEKEFYACEIQLQKWYGDLRGNPFAVEAENRVVDISGFSIGSDGHSIQYRNLLAWKYLENGKREFYLLMNYGKKGRIRFTDGTDIKKSGKWQGLLYGGGKAKVIDLTFDPDDEQLIILPLAFGTRQGREFIWNDLLSLETGLIKLANGRVIEKTIYNKKIGRDEPALFVALTFERREVVDPSNIKPVNLIGVARGENIPAVIALTDPEGCPLPEFKDSSGGPTDILRIGEGYKEKQRAIQAAKEVEQRRAGGYSRKFASKSRNLADDMVRNSARDLFY

HAVTHDAVLVFANLSRGFGRQGKRTFMTERQYTKMEDWLTAKLAYEGLTSKTYLSKTLAQYTSKTCSNCGFTITYAD
MDVMLVRLKKTSDGWATTLNNKELKAEYQITYYNRYKRQTVEKELSAELDRLSEESGNNDISKWTKGRRDEALFLLK
KRFSHRPVQEQFVCLDCGHEVHAAEQAALNIARSWLFLNSNSTEFKSYKSGKQPFVGAWQAFYKRRLKEVWKPNA

[0378] 示例性的CasY((ncbi.nlm.nih.gov/protein/APG80656.1)>APG80656.1 CRISPR-相关的蛋白质CasY[未培养的副杆菌群细菌])氨基酸序列如下:

[0379] MSKRHPRISGVKGYRLHAQRLEYTGKSGAMRTIKYPLYSSPSGGRTVPREIVSAINDDYVGLYGLSNF
DDLYNAEKRNEEKVYSVLDFWYDCVQYGAVFSYTAPGLLKNVAEVRGGSYELTKTLKGSHLYDELQIDKVIKFLNK
KEISRANGSLDKLKKDIIDCFKAEYRERHKDQCNKLADDIKNAKKDAGASLGERQKKLFRDFFGISEQSENDKPSF
TNPLNLTCCLLPFDTVNNNRNRGEVLFNKLKEYAQKLDKNEGSLEMWEYIGIGNSGTAFSNFLGEGFLGRLRENKI
TELKKAMMDITDAWRGQEQEEELEKRLRILAALTIKLREPKFDNHWGGYRSDINGKLSSWLQNYINQTVKIKEDLK
GHKKDLKKAKEMINRFGESDTKEEAVVSSLLESIEKIVPDDSADDEKPDIPAIAIYRRFLSDGRLTLNRFVQREDV
QEALIKERLEAEKKKKPKKRKKKSDAEDEKETIDFKELFPHLAKPLKLVPNFYGDSKRELYKKYKNAAIYTDALWK
AVEKIYKSAFSSSLKNSFFDTDFDKDFFIKRLQKIFSVYRRFNTDKWKPIVKNSFAPYCDIVSLAENEVLYKPKQS
RSRKSAAIDKNRVRLPSTENIAKAGIALARELSVAGFDWKDLLKKEEHEEYIDLIELHKTALALLLAVTETQLDIS
ALDFVENGTVKDFMKTRDGNLVLEGRFLEMFSQSIVFSELRGLAGLMSRKEFITRSAIQTMNGKQAELLYIPHEFQ
SAKITTPKEMSRAFLDLAPAEFATSLEPESLSEKSLLKLKQMRYYPHYFGYELTRTGQGIDGGVAENALRLEKSPV
KKREIKCKQYKTLGRGQNKIVLYVRSSYYQTQFLEWFLHRPKNVQTDVAVSGSFLIDEKKVKTRWNYDALTVALEP
VSGSERVFVSQPFTIFPEKSAEEEGQRYLGIDIGEYGIAYTALEITGDSAKILDQNFISDPQLKTLREEVKGLKLD
QRRGTFAMPSTKIARIRESLVHSLRNRIHHLALKHKAKIVYELEVSRFEEGKQKIKKVYATLKKADVYSEIDADKN
LQTTVWGKLAVASEISASYTSQFCGACKKLWRAEMQVDETITTQELIGTVRVIKGGTLIDAIKDFMRPPIFDENDT
PFPKYRDFCDKHHISKKMRGNSCLFICPFCRANADADIQASQTIALLRYVKEEKKVEDYFERFRKLKNIKVLGQMK
KI.

[0380] Cas9核酸酶具有两个功能性核酸内切酶结构域:RuvC和HNH。Cas9在结合目标后发生构象变化,定位核酸酶结构域以切割目标DNA的相反链。Cas9介导的DNA切割的最终结果是目标DNA(PAM序列上游约3-4个核苷酸)内的双链断裂(DSB)。然后通过两种一般修复途径之一修复产生的DSB:(1)有效但容易出错的非同源末端连接(NHEJ)途径;或(2)效率较低但保真度高的同源定向修复(HDR)途径。

[0381] 非同源末端连接(NHEJ)和/或同源定向修复(HDR)的“效率”可以通过任何方便的方法计算。例如,在一些实施例中,效率可以用成功HDR的百分比来表示。例如,测量员核酸酶测定(Surveyor nuclease assay)可用于产生切割产物,并且产物与底物的比率可用于计算百分比。例如,可以使用测量员核酸酶(surveyor nuclease enzyme)直接切割含有作为成功HDR的结果的新整合的限制性序列的DNA。更多裂解的基质表明更高的HDR百分比(更高的HDR效率)。作为说明性示例,可以使用以下等式[(裂解产物)/(底物加裂解产物)]计算HDR的分数(百分比)(例如(b+c)/(a+b+c),其中“a”是DNA底物的条带强度,“b”和“c”是切割产物)。

[0382] 在一些实施例中,效率可以用成功NHEJ的百分比来表示。例如,T7核酸内切酶I测定可用于产生裂解产物,产物与底物的比率可用于计算NHEJ的百分比。T7核酸内切酶I切割由野生型和突变DNA链杂交产生的错配异源双链DNA(NHEJ在原始断裂位点产生小的随机插入缺失(indels))。更多的切割表明更高的NHEJ百分比(更高的NHEJ效率)。作为说明性示

例,NHEJ的分数(百分比)可以使用以下等式计算:(1-(1-(b+c)/(a+b+c))$^{1/2}$)×100,其中“a”是DNA底物的条带强度,“b”和“c”是切割产物(Ran等人,Cell.2013Sep.12;154(6):1380-9;和Ran等人,Nat Protoc.2013Nov.;8(11):2281-2308)。

[0383] NHEJ修复途径是最活跃的修复机制,它经常导致DSB位点的小核苷酸插入缺失(indels)。NHEJ介导的DSB修复的随机性具有重要的实际意义,因为表达Cas9和gRNA或向导多核苷酸的细胞群会导致多种突变。在一些实施例中,所述NHEJ在目标DNA中产生小的插入缺失,导致氨基酸缺失、插入或移码突变,从而导致目标基因的开放阅读框(ORF)内的过早终止密码子。理想的最终结果是目标基因内的功能丧失突变。

[0384] 虽然NHEJ介导的DSB修复通常会破坏基因的开放阅读框,但同源定向修复(HDR)可用于产生特定的核苷酸变化,范围从单个核苷酸变化到大插入,如添加荧光团或标签。为了利用HDR进行基因编辑,可以使用gRNA和Cas9或Cas9切口酶将包含所需序列的DNA修复模板递送到感兴趣的细胞类型中。所述修复模板可以包含所需的编辑以及紧邻目标上游和下游的其他同源序列(称为左右同源臂)。每个同源臂的长度取决于引入的变化的大小,更大的插入需要更长的同源臂。所述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。即使在表达Cas9、gRNA和外源修复模板的细胞中,HDR的效率通常也很低(<10%的修饰等位基因)。HDR的效率可以通过同步细胞来提高,因为HDR发生在细胞周期的S和G2阶段。NHEJ中涉及的化学或遗传抑制基因也可以增加HDR频率。

[0385] 在一些实施例中,Cas9是经修饰的Cas9。给定的gRNA靶向序列可以在整个基因组中具有额外的位点,其中存在部分同源性。这些位点称为脱靶位点,在设计gRNA时需要加以考虑。除了优化gRNA设计,还可以通过对Cas9的修改来提高CRISPR的特异性。Cas9通过两个核酸酶结构域RuvC和HNH的联合活性产生双链断裂(DSB)。Cas9切口酶是SpCas9的D10A突变体,保留一个核酸酶结构域并产生DNA切口而不是DSB。切口酶系统还可以与HDR介导的基因编辑相结合,以进行特定的基因编辑。

[0386] 在一些实施例中,Cas9是变体Cas9蛋白。变体Cas9多肽具有与野生型Cas9蛋白的氨基酸序列相比相差一个氨基酸的氨基酸序列(例如具有缺失、插入、取代、融合)。在一些情况下,所述变体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、插入或取代)。例如,在一些情况下,所述变体Cas9多肽具有与相应的野生型Cas9蛋白相比,少于50%、少于40%、少于30%、少于20%、少于10%、少于5%或少于1%的核酸酶的活性。在一些实施例中,所述变体Cas9蛋白没有实质性的核酸酶活性。当主题Cas9蛋白是没有实质性核酸酶活性的变体Cas9蛋白时,它可以被称为“dCas9”。

[0387] 在一些实施例中,变体Cas9蛋白具有降低的核酸酶活性。例如,变体Cas9蛋白表现出与野生型Cas9蛋白(例如野生型Cas9蛋白)相比,少于约20%、少于约15%、少于约10%、少于约5%、少于约1%或少于约0.1%的核酸内切酶活性。

[0388] 在一些实施例中,变体Cas9蛋白可以切割引导目标序列的互补链,但切割双链引导目标序列的非互补链的能力降低。例如,所述变体Cas9蛋白可以具有降低RuvC结构域功能的突变(氨基酸取代)。作为非限制性实例,在一些实施例中,所述变体Cas9蛋白具有D10A(在氨基酸位置10处天冬氨酸到丙氨酸)并且因此可以切割双链引导目标序列的互补链但切割非双链引导目标序列的互补链(因此当所述变体Cas9蛋白切割双链目标核酸时,导致单链断裂(SSB)而不是双链断裂(DSB))(见例如Jinek等人,Science.2012Aug.17;337

(6096):816-21)。

[0389] 在一些实施例中,变体Cas9蛋白可以切割双链引导目标序列的非互补链,但切割引导目标序列的互补链的能力降低。例如,所述变体Cas9蛋白可以具有降低HNH结构域(RuvC/HNH/RuvC结构域基序)功能的突变(氨基酸取代)。作为非限制性实例,在一些实施例中,所述变体Cas9蛋白具有H840A(在氨基酸位置840处组氨酸至丙氨酸)突变,因此可切割引导目标序列的非互补链,但切割引导目标序列的互补链的能力降低(因此,当变体Cas9蛋白切割双链引导目标序列时,会产生SSB而不是DSB)。此类Cas9蛋白切割引导目标序列(例如单链引导目标序列)的能力降低,但保留结合引导目标序列(例如单引导目标序列)的能力。

[0390] 在一些实施例中,变体Cas9蛋白切割双链目标DNA的互补链和非互补链的能力降低。作为非限制性实例,在一些实施例中,所述变体Cas9蛋白包含D10A和H840A突变两者,使得多肽切割双链目标DNA的互补链和非互补链的能力降低。此类Cas9蛋白切割目标DNA(例如单链目标DNA)的能力降低,但保留结合目标DNA(例如单链目标DNA)的能力。

[0391] 作为另一个非限制性实例,在一些实施例中,所述变体Cas9蛋白含有W476A和W1126A突变,使得多肽切割目标DNA的能力降低。此类Cas9蛋白切割目标DNA(例如单链目标DNA)的能力降低,但保留结合目标DNA(例如单链目标DNA)的能力。

[0392] 作为另一个非限制性实例,在一些实施例中,所述变体Cas9蛋白含有P475A、W476A、N477A、D1125A、W1126A和D1127A突变,使得多肽切割目标DNA的能力降低。此类Cas9蛋白切割目标DNA(例如单链目标DNA)的能力降低,但保留结合目标DNA(例如单链目标DNA)的能力。

[0393] 作为另一个非限制性实例,在一些实施例中,所述变体Cas9蛋白包含H840A、W476A和W1126A突变,使得多肽切割目标DNA的能力降低。此类Cas9蛋白切割目标DNA(例如单链目标DNA)的能力降低,但保留结合目标DNA(例如单链目标DNA)的能力。作为另一个非限制性实例,在一些实施例中,所述变体Cas9蛋白含有H840A、D10A、W476A和W1126A突变,使得多肽切割目标DNA的能力降低。此类Cas9蛋白切割目标DNA(例如单链目标DNA)的能力降低,但保留结合目标DNA(例如单链目标DNA)的能力。在一些实施例中,所述变体Cas9在Cas9 HNH结构域(A840H)中的位置840处恢复了催化性His残基。

[0394] 作为另一个非限制性实例,在一些实施例中,所述变体Cas9蛋白含有H840A、P475A、W476A、N477A、D1125A、W1126A和D1127A突变,使得多肽切割目标DNA的能力降低。此类Cas9蛋白切割目标DNA(例如单链目标DNA)的能力降低,但保留结合目标DNA(例如单链目标DNA)的能力。作为另一个非限制性实例,在一些实施例中,所述变体Cas9蛋白含有D10A、H840A、P475A、W476A、N477A、D1125A、W1126A和D1127A突变,使得多肽切割目标DNA的能力降低。此类Cas9蛋白切割目标DNA(例如单链目标DNA)的能力降低,但保留结合目标DNA(例如单链目标DNA)的能力。在一些实施例中,当变体Cas9蛋白包含W476A和W1126A突变或当变体Cas9蛋白包含P475A、W476A、N477A、D1125A、W1126A和D1127A突变时,所述变体Cas9蛋白不与PAM序列有效结合。因此,在一些此类实施例中,当此类变体Cas9蛋白用于结合方法时,该方法不需要PAM序列。换言之,在一些实施例中,当这种变体Cas9蛋白用于结合方法中时,该方法可以包括向导RNA,但是该方法可以在不存在PAM序列的情况下进行(并且结合的特异性是因此由向导RNA的靶向片段提供)。可以使其他残基突变以实现上述效果(即使一个或

其他核酸酶部分失活)。作为非限制性实例,残基D10、G12、G17、E762、H840、N854、N863、H982、H983、A984、D986和/或A987可以被改变(即被取代)。此外,丙氨酸取代以外的突变也是合适的。

[0395] 在一些实施例中,变体Cas9蛋白具有降低的催化活性(例如当Cas9蛋白具有D10、G12、G17、E762、H840、N854、N863、H982、H983、A984、D986和/或A987突变时,例如D10A、G12A、G17A、E762A、H840A、N854A、N863A、H982A、H983A、A984A和/或D986A),所述变体Cas9蛋白仍然可以以位点特异性方式结合目标DNA(因为它仍然是被向导RNA引导到目标DNA序列),只要它保留与向导RNA相互作用的能力。

[0396] 在一些实施例中,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。

[0397] 在一些实施例中,使用包括氨基酸取代D1135M、S1136Q、G1218K、E1219F、A1322R、D1332A、R1335E和T1337R(SpCas9-MQKFRAER)并且对改变的PAM 5'-NGC-3'具有特异性的修饰SpCas9。

[0398] 化脓性链球菌Cas9的替代品可以包括来自Cpf1家族的RNA引导的核酸内切酶,它们在哺乳动物细胞中显示出切割活性。普氏和弗朗西斯菌1(Prevotella and Francisella 1)(CRISPR/Cpf1)的CRISPR是一种类似于CRISPR/Cas9系统的DNA编辑技术。Cpf1是II类CRISPR/Cas系统的RNA引导的核酸内切酶。这种获得性免疫机制存在于普氏菌和弗朗西斯菌中。Cpf1基因与CRISPR基因座相关,编码内切核酸酶,使用向导RNA来寻找和切割病毒DNA。Cpf1是一种比Cas9更小、更简单的核酸内切酶,克服了CRISPR/Cas9系统的一些限制。与Cas9核酸酶不同,Cpf1介导的DNA切割的结果是具有短3'突出端的双链断裂。Cpf1的交错切割模式可以开辟定向基因转移的可能性,类似于传统的限制酶克隆,可以提高基因编辑的效率。与上述Cas9变体和直向同源物一样,所述Cpf1还可以将CRISPR可靶向的位点数量扩展到富含AT的区域或富含AT的基因组,这些区域缺乏SpCas9青睐的NGG PAM位点。Cpf1基因座包含一个混合的α/β结构域、一个RuvC-I后跟一个螺旋区域、一个RuvC-II和一个锌指样结构域。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没有HNH核酸内切酶结构域,并且Cpf1的N端没有Cas9的α螺旋识别叶。Cpf1 CRISPR-Cas结构域架构表明Cpf1在功能上是独一无二的,被归类为2类V型CRISPR系统。Cpf1基因座编码的Cas1、Cas2和Cas4蛋白与I型和III型系统比II型系统更相似。功能性Cpf1不需要反式激活CRISPR RNA(tracrRNA),因此,只需要CRISPR(crRNA)。这有利于基因组编辑,因为Cpf1不仅比Cas9小,而且它的sgRNA分子更小(大约是Cas9的一半核苷酸)。与Cas9靶向的富含G的PAM相比,Cpf1-crRNA复合物通过识别邻近基序5'-YTN-3'或5'-TTN-3'来切割目标DNA或RNA。在鉴定PAM后,Cpf1引入了一个粘端样DNA双链断裂,具有4或5个核苷酸的突出端。

[0399] 在一些实施例中,所述Cas9是对改变的PAM序列具有特异性的Cas9变体。在一些实施例中,额外的Cas9变体和PAM序列在Miller,S.M.等人,Continuous evolution of SpCas9 variants compatible with non-G PAMs,Nat.Biotechnol.(2020)中有所描述,其全部内容通过引用并入本文。在一些实施例中,Cas9变体没有特定的PAM要求。在一些实施例中,Cas9变体,例如SpCas9变体对NRNH PAM具有特异性,其中R是A或G并且H是A、C或T。在一些实施例中,所述SpCas9变体对PAM序列AAA、TAA、CAA、GAA、TAT、GAT或CAC具有特异性。在一些实施例中,所述SpCas9变体包含在SEQ ID NO:1中编号位置1114、1134、1135、1137、

1139、1151、1180、1188、1211、1218、1219、1221、1249、1256、1264、1290、1318、1317、1320、1321、1323、1332、1333、1335、1337、或1339或其相应位置处的氨基酸置换。在一些实施例中,所述SpCas9变体包含在SEQ ID NO:1中编号位置1114、1135、1218、1219、1221、1249、1320、1321、1323、1332、1333、1335、或1337或其相应位置处的氨基酸置换。在一些实施例中,所述SpCas9变体包含在如SEQ ID NO:1中编号位置1114、1134、1135、1137、1139、1151、1180、1188、1211、1219、1221、1256、1264、1290、1318、1317、1320、1323、1333或其相应位置处的氨基酸置换。在一些实施例中,所述SpCas9变体包含在SEQ ID NO:1中编号位置1114、1131、1135、1150、1156、1180、1191、1218、1219、1221、1227、1249、1253、1286、1293、1320、1321、1332、1335、1339或其相应位置处的氨基酸置换。在一些实施例中,所述SpCas9变体包含在SEQ ID NO:1中位置编号1114、1127、1135、1180、1207、1219、1234、1286、1301、1332、1335、1337、1338、1349或其相应位置处的氨基酸置换。SpCas9变体的示例性氨基酸置换和PAM特异性示于表1A-1D中。

[0400] 表1A.

[0401]

SpCas9	SpCas9 氨基酸位置												
	1114	1135	1218	1219	1221	1249	1320	1321	1323	1332	1333	1335	1337
	R	D	G	E	Q	P	A	P	A	D	R	R	T
AAA		N		V	H						G		
AAA		N		V	H						G		
AAA				V							G		
TAA	G	N		V							I		
TAA		N		V							I		A
TAA	G	N		V							I		A
CAA				V							K		
CAA		N		V							K		
CAA		N		V							K		

[0402]

SpCas9	SpCas9 氨基酸位置												
	1114	1135	1218	1219	1221	1249	1320	1321	1323	1332	1333	1335	1337
	R	D	G	E	Q	P	A	P	A	D	R	R	T
GAA				V	H		V				K		
GAA		N		V			V				K		
GAA				V	H		V				K		
TAT			S	V	H	S		S				L	
TAT			S	V	H	S		S				L	
TAT			S	V	H	S		S				L	
GAT				V							I		
GAT				V					D			Q	
GAT				V					D			Q	
CAC				V						N		Q	N
CAC		N		V								Q	N
CAC				V						N		Q	N

[0403] 表1B.

[0404]

SpCas9	SpCas9 氨基酸位置																		
	1114	1134	1135	1137	1139	1151	1180	1188	1211	1219	1221	1256	1264	1290	1318	1317	1320	1323	1333
	R	F	D	P	V	K	D	K	K	E	Q	Q	H	V	L	N	A	A	R
GAA										V	H						V		K
GAA			N	S						V							V	D	K
GAA			N							V	H		Y				V		K
CAA			N							V	H		Y				V		K
CAA	G		N	S						V	H		Y				V		K
CAA			N					R		V	H						V		K
CAA			N				G		R	V	H		Y				V		K
CAA			N							V	H		Y				V		K
AAA			N				G			V	H	R	Y				V	D	K
CAA	G		N				G			V	H		Y				V	D	K
CAA		L	N				G			V	H		Y			T	V	D	K
TAA	G		N				G			V	H		Y	G	S		V	D	K
TAA	G		N			E	G			V	H		Y		S		V		K
TAA	G		N				G			V	H		Y		S		V	D	K
TAA	G		N				G		R	V	H						V		K
TAA			N				G		R	V	H		Y				V		K
TAA	G		N		A		G			V	H						V		K
TAA	G		N							V	H						V		K

[0405] 表1C.

[0406]

SpCas9	SpCas9 氨基酸位置

[0407]

	1114	1131	1135	1150	1156	1180	1191	1218	1219	1221	1227	1249	1253	1286	1293	1320	1321	1332	1335	1339
	R	Y	D	E	K	D	K	G	E	Q	A	P	E	N	A	A	P	D	R	T
SacB. TAT			N				N		V	H						V	S		L	
SacB. TAT			N					S	V	H		S					S	G	L	
AAT			N					S	V	H	V	S		K	T		S	G	L	I
TAT	G		N			G		S	V	H		S	K				S	G	L	
TAT	G		N			G		S	V	H		S					S	G	L	
TAT	G	C	N			G		S	V	H		S					S	G	L	
TAT	G	C	N			G		S	V	H		S					S	G	L	
TAT	G	C	N			G		S	V	H		S					S	G	L	
TAT	G	C	N		E	G		S	V	H		S					S	G	L	
TAT	G	C	N	V		G		S	V	H		S					S	G	L	
TAT		C	N			G		S	V	H		S					S	G	L	
TAT	G	C	N			G		S	V	H		S					S	G	L	

[0408] 表1D.

[0409]

SpCas9	SpCas9 氨基酸位置													
	1114	1127	1135	1180	1207	1219	1234	1286	1301	1332	1335	1337	1338	1349
	R	D	D	D	E	E	N	N	P	D	R	T	S	H
SacB. CAC			N			V				N	Q	N		
AAC	G		N			V				N	Q	N		
AAC	G		N			V				N	Q	N		
TAC	G		N			V				N	Q	N		
TAC	G		N			V		H		N	Q	N		
TAC	G		N		G	V	D	H		N	Q	N		
TAC	G		N			V				N	Q	N		
TAC	G	G	N	E		V		H		N	Q	N		
TAC	G		N			V		H		N	Q	N		
TAC	G		N			V				N	Q	N	T	R

[0410] 在一些实施例中,Cas9是脑膜炎奈瑟菌Cas9(NmeCas9)或其变体。在一些实施例中,NmeCas9对NNNNGAYW PAM具有特异性,其中Y是C或T并且W是A或T。在一些实施例中,NmeCas9对NNNNGYTT PAM具有特异性,其中Y是C或T。在一些实施例中,NmeCas9对NNNNGTCT PAM具有特异性。在一些实施例中,NmeCas9是Nme1 Cas9。在一些实施例中,NmeCas9具有以下特异性:aNNNNGATT PAM、NNNNCCTA PAM、NNNNCCTC PAM、NNNNCCTT PAM、NNNNCCTG PAM、NNNNCCGT PAM、NNNNCCGGPAM、NNNNCCCA PAM、NNNNCCCT PAM、NNNNCCCC PAM、NNNNCCAT PAM、NNNNCCAG PAM、NNNNCCAT PAM、或NNNGATT PAM。在一些实施例中,Nme1Cas9对NNNNGATT PAM、NNNNCCTA PAM、NNNNCCTC PAM、NNNNCCTT PAM、或NNNNCCTG PAM具有特异性。在一些实施例中,NmeCas9对CAA PAM、CAAA PAM或CCA PAM具有特异性。在一些实施例中,NmeCas9是Nme2 Cas9。在一些实施例中,NmeCas9对NNNNCC(N4CC)PAM具有特异性,其中N是A、G、C或T中的任何一个。在一些实施例中,NmeCas9具有以下特异性:NNNNCCGT PAM、NNNNCCGGPAM、NNNNCCCA PAM、NNNNCCCT PAM、NNNNCCCC PAM、NNNNCCAT PAM、NNNNCCAG PAM、NNNNCCAT PAM、或NNNGATT PAM。在一些实施例中,NmeCas9是Nme3Cas9。在一些实施例中,NmeCas9对NNNNCAAA PAM、NNNNCC PAM或NNNNCNNN PAM具有特异性。额外的NmeCas9特征和PAM序列,如Edraki等人在Mol.Cell.(2019)73(4):714-726中所描述,通过引用全文并入本文。

[0411] 下面提供了Nme1Cas9的示例性氨基酸序列：

[0412] II型CRISPR RNA引导的核酸内切酶Cas9[脑膜炎奈瑟菌]WP_002235162.1

```
   1 maafkpnpin yilgldigia svgwamveid edenpiclid lgvrvferae vpktgdslam
  61 arrlarsvrr ltrrrahrll rarrllkreg vlqaadfden glikslpntp wqlraaaldr
 121 kltplewsav llhlikhrgy lsqrkneget adkelgallk gvadnahalq tgdfrtpael
 181 alnkfekesg hirnqrgdys htfsrkdlqa elillfekqk efgnphvsgg lkegietllm
 241 tqrpalsgda vqkmlghctf epaepkaakn tytaerfiwl tklnnlrile qgserpltdt
 301 eratlmdepy rkskltyaqa rkllgledta ffkglrygkd naeastlmem kayhaisral
 361 ekeglkdkks plnlspelqd eigtafslfk tdeditgrlk driqpeilea llkhisfdkf
 421 vqislkalrr ivplmeqgkr ydeacaeiyg dhygkkntee kiylppipad eirnpvvlra
 481 lsqarkving vvrrygspar ihietarevg ksfkdrkeie krqeenrkdr ekaaakfrey
```

[0413]

```
 541 fpnfvgepks kdilklrlye qqhgkclysg keinlgrlne kgyveidhal pfsrtwddsf
 601 nnkvlvlgse nqnkgnqtpy eyfngkdnsr ewqefkarve tsrfprskkq rillqkfded
 661 gfkernlndt ryvnrflcqf vadrmrltgk gkkrvfasng qitnllrgfw glrkvraend
 721 rhhaldavvv acstvamqqk itrfvrykem nafdgktidk etgevlhqkt hfpqpweffa
 781 qevmirvfgk pdgkpefeea dtpeklrtll aeklssrpea vheyvtplfv srapnrkmsg
 841 qghmetvksa krldegvsvl rvpltqlklk dlekmvnrer epklyealka rleahkddpa
 901 kafaepfyky dkagnrtqqv kavrveqvqk tgvwvrnhng iadnatmvrv dvfekgdkyy
 961 lvpiyswqva kgilpdravv qgkdeedwql iddsfnfkfs lhpndlvevi tkkarmfgyf
1021 aschrgtgni nirihdldhk igkngilegi gvktalsfqk yqidelgkei rpcrlkkrpp
1081 vr
```

[0414] 下面提供了Nme2Cas9的示例性氨基酸序列：

[0415] II型CRISPR RNA引导的核酸内切酶Cas9[脑膜炎奈瑟菌]WP_002230835.1

1 maafkpnpin yilgldigia svgwamveid eeenpirlid lgvrvferae vpktgdslam

61 arrlarsvrr ltrrrahrll rarrllkreg vlqaadfden glikslpntp wqlraaaldr

121 kltplewsav llhlikhrgy lsqrkneget adkelgallk gvannahalq tgdfrtpael

181 alnkfekesg hirnqrgdys htfsrkdlqa elillfekqk efgnphvsgg lkegietllm

241 tqrpalsgda vqkmlghctf epaepkaakn tytaerfiwl tklnnlrile qgserpltdt

301 eratlmdepy rkskltyaqa rkllgledta ffkglrygkd naeastlmem kayhaisral

361 ekeglkdkks plnlsselqd eigtafslfk tdeditgrlk drvqpeilea llkhisfdkf

421 vqislkalrr ivplmeqgkr ydeacaeiyg dhygkkntee kiylppipad eirnpvvlra

481 lsqarkving vvrrygspar ihietarevg ksfkdrkeie krqeenrkdr ekaaakfrey

[0416] 541 fpnfvgepks kdilklrlye qqhgkclysg keinlvrlne kgyveidhal pfsrtwddsf

601 nnkvlvlgse nqnkgnqtpy eyfngkdnsr ewqefkarve tsrfprskkq rillqkfded

661 gfkecnlndt ryvnrflcqf vadhilltgk gkrrvfasng qitnllrgfw glrkvraend

721 rhhaldavvv acstvamqqk itrfvrykem nafdgktidk etgkvlhqkt hfpqpweffa

781 qevmirvfgk pdgkpefeea dtpeklrtll aeklssrpea vheyvtplfv srapnrkmsg

841 ahkdtlrsak rfvkhnekis vkrvwlteik ladlenmvny kngreielye alkarleayg

901 gnakqafdpk dnpfykkggq lvkavrvekt qesgvllnkk naytiadngd mvrvdvfckv

961 dkkgknqyfi vpiyawqvae nilpdidckg yriddsytfc fslhkydlia fqkdekskve

1021 fayyincdss ngrfylawhd kgskeqqfri stqnlvliqk yqvnelgkei rpcrlkkrpp

1081 vr

[0417] 核碱基编辑器的Cas12结构域

[0418] 通常,微生物CRISPR-Cas系统分为1类和2类系统。1类系统具有多亚基效应复合物,而2类系统具有单个蛋白质效应器。例如,Cas9和Cpf1是2类效应器,尽管类型不同(分别为II型和V型)。除了Cpf1,2类V型CRISPR-Cas系统还包括Cas12a/Cpf1、Cas12b/C2c1、Cas12c/C2c3、Cas12d/CasY、Cas12e/CasX、Cas12g、Cas12h和Cas12i)。见例如Shmakov等人,“Discovery and Functional Characterization of Diverse Class 2CRISPR Cas Systems,”Mol.Cell,2015Nov.5;60(3):385-397;Makarova等人,“Classification and Nomenclature of CRISPR-Cas Systems:Where from Here?”CRISPR Journal,2018,1(5):325-336;和Yan等人,“Functionally Diverse Type V CRISPR-Cas Systems,”Science,2019Jan.4;363:88-91;其各自的全部内容通过引用并入本文。V型Cas蛋白包含一个RuvC(或RuvC样)核酸内切酶结构域。虽然成熟CRISPR RNA(crRNA)的生产通常不依赖于tracrRNA,但例如Cas12b/C2c1需要tracrRNA来生产crRNA。Cas12b/C2c1依赖crRNA和tracrRNA进行DNA切割。

[0419] 本发明中考虑的核酸可编程DNA结合蛋白包括归类为第2类V型(Cas12蛋白)的Cas蛋白。Cas 2类V型蛋白的非限制性实例包括Cas12a/Cpf1、Cas12b/C2c1、Cas12c/C2c3、Cas12d/CasY、Cas12e/CasX、Cas12g、Cas12h和Cas12i,其同源物或其修饰形式。如本文所用,所述Cas12蛋白也可称为Cas12核酸酶、Cas12结构域或Cas12蛋白结构域。在一些实施例中,本发明的所述Cas12蛋白包含被内部融合蛋白结构域例如脱氨酶结构域中断的氨基酸

序列。

[0420] 在一些实施例中,所述Cas12结构域是核酸酶无活性的Cas12结构域或Cas12切口酶。在一些实施例中,所述Cas12结构域是核酸酶活性结构域。例如,Cas12结构域可以是在双链核酸(例如双链DNA分子)的一条链上形成切口的Cas12结构域。在一些实施例中,Cas12结构域包含如本文所述的任一氨基酸序列。在一些实施例中,所述Cas12结构域包含与本文所述的任一氨基酸序列至少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%相同的氨基酸序列。在一些实施例中,所述Cas12结构域包含与本文所述的任一氨基酸序列相比,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个或更多个突变的氨基酸序列。在一些实施例中,所述Cas12结构域包含与本文列出的任一氨基酸序列相比,具有至少10、至少15、至少20、至少30、至少40、至少50、至少60、至少70、至少80、至少90、至少100、至少150、至少200、至少250、至少300、至少350、至少400、至少500、至少600、至少700、至少800、至少900、至少1000、至少1100或至少1200相同的连续氨基酸残基的氨基酸序列。

[0421] 在一些实施例中,提供了包含Cas12片段的蛋白质。例如,在一些实施例中,蛋白质包含两个Cas12域之一:(1)Cas12的gRNA结合结构域;或(2)Cas12的DNA切割结构域。在一些实施例中,包含Cas12或其片段的蛋白质被称为“Cas12变体”。Cas12变体与Cas12或其片段具有同源性。例如,Cas12变体与野生型Cas12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。在一些实施例中,所述Cas12变体与野生型Cas12相比,可具有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。在一些实施例中,所述Cas12变体包含Cas12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Cas12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。在一些实施例中,该片段是相应野生型Cas12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%相同、至少96%、至少97%、至少98%、至少99%或至少99.5%。在一些实施例中,该片段的长度为至少100个氨基酸。在一些实施例中,该片段是至少100、150、200、250、300、350、400、450、500、550、600、650、700、750、800、850、900、950、1000、1050、1100、1150、1200、1250或至少1300个氨基酸的长度。

[0422] 在一些实施例中,Cas12对应于,或部分或全部包含具有一个或多个改变所述Cas12核酸酶活性的突变的Cas12氨基酸序列。举例来说,此类突变包括Cas12的RuvC核酸酶结构域内的氨基酸置换。在一些实施例中,提供了Cas12的变体或同源物,其与野生型Cas12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。在一些实施例中,提供Cas12变体,其具有以下较短或较长氨基酸序列: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。

[0423] 在一些实施例中,本文提供的Cas12融合蛋白包含Cas12蛋白的全长氨基酸序列,例如本文提供的Cas12序列之一。然而,在其他实施例中,本文提供的融合蛋白不包含全长Cas12序列,而仅包含其一个或多个片段。本文提供了合适的Cas12结构域的示例性氨基酸序列,并且Cas12结构域和片段的其他合适的序列对本领域技术人员来说是显而易见的。

[0424] 通常,第2类V型Cas蛋白具有单个功能性RuvC核酸内切酶结构域(参见例如Chen等人,"CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity,"Science 360:436-439(2018)).在某些情况下,Cas12蛋白是Cas12b蛋白的变体。(见Strecker等人,Nature Communications,2019,10(1):Art.No.:212)。在一个实施例中,变体Cas12多肽与野生型Cas12蛋白的氨基酸序列相比,具有相差1、2、3、4、5个或更多个氨基酸的氨基酸序列(例如具有删除、插入、替换、融合)。在一些情况下,所述变体Cas12多肽具有降低所述Cas12多肽活性的氨基酸变化(例如缺失、插入或取代)。例如在一些情况下,所述变体Cas12是Cas12b多肽,其具有小于50%、小于40%、小于30%、小于20%、小于10%、小于5%或小于1的相应野生型Cas12b蛋白的切口酶活性。在某些情况下,所述变体Cas12b蛋白没有实质性的切口酶活性。

[0425] 在某些情况下,变体Cas12b蛋白的切口酶活性降低。例如变体Cas12b蛋白表现出小于约20%、小于约15%、小于约10%、小于约5%、小于约1%或小于约0.1%的野生型Cas12b蛋白的切口酶活性。

[0426] 在一些实施例中,Cas12蛋白包括来自Cas12a/Cpf1家族的RNA引导的核酸内切酶,其在哺乳动物细胞中表现出活性。普氏菌和弗朗西斯菌1(Prevotella and Francisella 1)(CRISPR/Cpf1)的CRISPR是一种类似于CRISPR/Cas9系统的DNA编辑技术。Cpf1是II类CRISPR/Cas系统的RNA引导的核酸内切酶。这种获得性免疫机制存在于普氏菌和弗朗西斯菌中。Cpf1基因与CRISPR基因座缔合,编码内切核酸酶,使用向导RNA来寻找和切割病毒DNA。Cpf1是一种比Cas9更小、更简单的核酸内切酶,克服了CRISPR/Cas9系统的一些限制。与Cas9核酸酶不同,Cpf1介导的DNA切割的结果是具有短3'突出端的双链断裂。Cpf1的交错切割模式可以开辟定向基因转移的可能性,类似于传统的限制酶克隆,可以提高基因编辑的效率。与上述Cas9变体和直向同源物一样,所述Cpf1还可以将CRISPR可靶向的位点数量扩展到富含AT的区域或富含AT的基因组,这些区域缺乏SpCas9青睐的NGG PAM位点。Cpf1基因座包含一个混合的α/β结构域、一个RuvC-I后跟一个螺旋区域、一个RuvC-II和一个锌指样结构域。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,与Cas9不同,Cpf1没有HNH核酸内切酶结构域,并且Cpf1的N-末端没有Cas9的α螺旋识别叶。Cpf1 CRISPR-Cas域架构表明Cpf1在功能上是独一无二的,被归类为2类V型CRISPR系统。Cpf1基因座编码的Cas1、Cas2和Cas4蛋白与I型和III型系统比II型系统更相似。功能性Cpf1不需要反式激活CRISPR RNA(tracrRNA),因此,只需要CRISPR(crRNA)。这有利于基因组编辑,因为Cpf1不仅比Cas9小,而且它的sgRNA分子更小(大约是Cas9的一半核苷酸)。与Cas9靶向的富含G的PAM相比,Cpf1-crRNA复合物通过识别邻近基序5'-YTN-3'或5'-TTTN-3'来切割目标DNA或RNA。在识别PAM后,Cpf1引入了一个粘性末端样DNA双链断裂,具有4或5个核苷酸的突出端。

[0427] 在本发明的一些方面,可以使用一种编码CRISPR酶的载体,该酶相对于相应的野

生型酶发生突变,使得突变的CRISPR酶缺乏切割含有目标序列的目标多核苷酸的一条或两条链的能力。Cas12可以指一种多肽,其具有与野生型示例性Cas12多肽(例如来自久氏芽孢杆菌的Cas12)至少或至少约50%、60%、70%、80%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、或100%的序列同一性和/或序列同源性。Cas12可以指一种多肽,其具有与野生型示例性Cas12多肽(例如来自久氏芽孢杆菌(BhCas12b)、芽孢杆菌V3-13(BvCas12b)和嗜酸脂环杆菌(AaCas12b))至多或至多约50%,60%、70%、80%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%的序列同一性和/或序列同源性。Cas12可以指Cas12蛋白的野生型或修饰形式,其可以包含氨基酸变化,例如缺失、插入、取代、变体、突变、融合、嵌合体或其任何组合。

[0428] 核酸可编程DNA结合蛋白

[0429] 本公开的一些方面提供了包含充当核酸可编程DNA结合蛋白的结构域的融合蛋白,其可用于将蛋白质,例如碱基编辑器引导至特定的核酸(例如DNA或RNA)序列。在特定实施例中,融合蛋白包含核酸可编程DNA结合蛋白结构域和脱氨酶结构域。核酸可编程DNA结合蛋白的非限制性实例包括:Cas9(例如dCas9和nCas9)、Cas12a/Cpf1、Cas12b/C2c1、Cas12c/C2c3、Cas12d/CasY、Cas12e/CasX、Cas12g、Cas12h和Cas12i。Cas酶的非限制性实例包括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,或其修饰或工程化版本。其他核酸可编程的DNA结合蛋白也在本公开的范围内,尽管它们可能未在本公开中具体列出。见例如Makarova等人,“Classification and Nomenclature of CRISPR-Cas Systems:Where from Here?”CRISPR J.2018Oct;1:325-336.doi:10.1089/crispr.2018.0033;Yan等人,“Functionally diverse type V CRISPR-Cas systems”Science.2019Jan 4;363(6422):88-91.doi:10.1126/science.aav7271,其各自的全部内容通过引用并入本文。

[0430] 具有与Cas9不同的PAM特异性的核酸可编程DNA结合蛋白的一个例子是来自普氏和弗朗西斯菌1(Cpf1)的成簇规则间隔短回文重复序列。与Cas9类似,Cpf1也是2类CRISPR效应器。已经表明,Cpf1介导强大的DNA干扰,其特征与Cas9不同。Cpf1是一种单一的RNA引导的内切核酸酶,缺乏tracrRNA,它利用富含T的原始间隔区相邻基序(TTN、TTTN或YTN)。此外,Cpf1通过交错的DNA双链断裂来切割DNA。在16个Cpf1家族蛋白中,来自氨基酸球菌属和毛螺菌科的两种酶被证明在人类细胞中具有有效的基因组编辑活性。Cpf1蛋白是本领域已知的并且之前已经描述过,例如Yamano等人,“Crystal structure of Cpf1 in complex with guide RNA and target DNA.”Cell(165)2016,p.949-962;其全部内容通过引用并入本文。

[0431] 在本组成物和方法中有用的是核酸酶无活性的Cpf1(dCpf1)变体,其可用作向导核苷酸序列可编程的DNA结合蛋白结构域。Cpf1蛋白具有与Cas9的RuvC结构域相似但不具

有HNH核酸内切酶结构域的RuvC样核酸内切酶结构域,且Cpf1的N-末端不具有Cas9的α-螺旋识别叶。它在Zetsche等人,Cell,163,759-771,2015中展示(通过引用并入本文),所述Cpf1的RuvC样结构域负责切割两条DNA链并且RuvC样结构域的失活使Cpf1核酸酶活性失活。例如与新泽西弗朗西斯菌(Francisella novicida)Cpf1中的D917A、E1006A或D1255A对应的突变会使Cpf1核酸酶活性失活。在一些实施例中,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。应当理解,根据本公开可以使用使Cpf1的RuvC结构域失活的任何突变,例如置换突变、缺失或插入。

[0432] 在一些实施例中,本文提供的任何融合蛋白的核酸可编程DNA结合蛋白(napDNAbp)可以是Cpf1蛋白。在一些实施例中,所述Cpf1蛋白是Cpf1切口酶(nCpf1)。在一些实施例中,Cpf1蛋白是核酸酶失活的Cpf1(dCpf1)。在一些实施例中,Cpf1、nCpf1或dCpf1包含与本文公开的Cpf1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%相同的氨基酸序列。在一些实施例中,所述dCpf1包含与本文公开的Cpf1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%相同的氨基酸序列,并且包含对应于D917A、E1006A、D1255A、D917A/E1006A、D917A/D1255A、E1006A/D1255A或D917A/E1006A/D1255A的突变。应当理解,根据本公开也可以使用来自其他细菌物种的Cpf1。

[0433] 野生型新泽西弗朗西斯菌(Francisella novicida)Cpf1(D917、E1006和D1255以粗体和下划线表示)

[0434] MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQF
FIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFK
NLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFK
GFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAE
ELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGI
NEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIA
AFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEY
ITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILA
NFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKL
KIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNF
ENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYK
LLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKF
IDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQ
GKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKK
ITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEI
NLLLKEKANDVHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAI
EKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**E**DLNFGFKRGRFKVE
KQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAG
FTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKG
KWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESD
KKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**D**ANGAY
HIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0435] 新泽西弗朗西斯菌Cpf1 D917A(A917、E1006和D1255为粗体和下划线)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQF
FIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFK
NLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFK
GFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAE
ELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGI
NEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIA
AFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEY
ITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILA
NFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKL
KIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNF
ENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYK
[0436] LLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKF
IDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQ
GKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKK
ITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEI
NLLLKEKANDVHILSI**A**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAI
EKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**E**DLNFGFKRGRFKVE
KQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAG
FTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKG
KWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESD
KKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**D**ANGAY
HIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0437] 新泽西弗朗西斯菌Cpf1 E1006A(D917、A1006和D1255为粗体和下划线)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQF
FIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFK
[0438] NLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFK
GFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAE
ELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGI
NEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIA

AFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEY
ITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILA
NFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKL
KIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNF
ENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYK
LLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKF
IDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQ
GKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKK
[0439] ITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEI
NLLLKEKANDVHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAI
EKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**A**DLNFGFKRGRFKVE
KQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAG
FTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKG
KWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESD
KKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**D**ANGAY
HIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0440] 新泽西弗朗西斯菌Cpf1 D1255A(D917、E1006和A1255为粗体和下划线)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQF
FIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFK
NLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFK
GFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAE
ELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGI
NEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIA

[0441] AFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEY
ITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILA
NFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKL
KIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNF
ENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYK
LLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKF
IDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQ

GKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKK
ITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEI
NLLLKEKANDVHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAI
EKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**E**DLNFGFKRGRFKVE

[0442] KQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAG
FTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKG
KWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESD
KKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**A**ANGAY
HIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0443] 新泽西弗朗西斯菌Cpf1 D917A/E1006A(A917、A1006和D1255以粗体和下划线表示)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQF
FIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFK
NLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFK
GFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAE
ELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGI
NEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIA
AFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEY
ITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILA
NFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKL

[0444] KIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNF
ENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYK
LLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKF
IDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQ
GKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKK
ITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEI
NLLLKEKANDVHILSI**A**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAI
EKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**A**DLNFGFKRGRFKVE
KQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAG
FTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKG
KWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESD

[0445] KKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**D**ANGAY
HIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0446] 新泽西弗朗西斯菌Cpf1 D917A/D1255A(A917、E1006和A1255以粗体和下划线表示)

[0447] MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQF
FIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFK
NLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFK
GFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAE
ELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGI
NEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIA
AFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEY
ITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILA
NFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKL
KIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNF
ENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYK
LLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKF
IDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQ
GKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKK
ITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEI
NLLLKEKANDVHILSI**A**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAI
EKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**E**DLNFGFKRGRFKVE
KQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAG
FTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKG
KWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESD
KKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**A**ANGAY
HIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0448] 新泽西弗朗西斯菌Cpf1 E1006A/D1255A(D917、A1006和A1255以粗体和下划线表示)

[0449] MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQF
FIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFK
NLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFK

GFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAE
ELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGI
NEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIA
AFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEY
ITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILA
NFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKL
KIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNF
ENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYK
LLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKF

[0450] IDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQ
GKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKK
ITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEI
NLLLKEKANDVHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAI
EKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**A**DLNFGFKRGRFKVE
KQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAG
FTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKG
KWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESD
KKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**A**ANGAY
HIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0451] 新泽西弗朗西斯菌Cpf1 D917A/E1006A/D1255A(A917、A1006和A1255为粗体和下划线)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQF
FIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFK
NLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFK
GFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAE

[0452] ELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGI
NEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIA
AFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEY
ITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILA
NFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKL
KIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNF

[0453] ENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYK
LLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKF
IDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQ
GKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKK
ITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEI
NLLLKEKANDVHILSIARGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAI
EKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVFADLNFGFKRGRFKVE
KQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAG
FTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKG
KWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESD
KKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDAAANGAY
HIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0454] 在一些实施例中,融合蛋白中存在的Cas9结构域之一可以被对PAM序列没有要求的向导核苷酸序列可编程的DNA结合蛋白结构域替换。

[0455] 在一些实施例中,Cas9结构域是来自金黄色葡萄球菌(SaCas9)的Cas9结构域。在一些实施例中,SaCas9结构域是核酸酶活性SaCas9、核酸酶无活性SaCas9(SaCas9d)或SaCas9切口酶(SaCas9n)。在一些实施例中,SaCas9包含N579A突变,或本文提供的任何氨基酸序列中的相应突变。

[0456] 在一些实施例中,所述SaCas9结构域、SaCas9d结构域或SaCas9n结构域可结合具有非规范PAM的核酸序列。在一些实施例中,所述SaCas9结构域、SaCas9d结构域或SaCas9n结构域可以结合具有NNGRRT或NNGRRT PAM序列的核酸序列。在一些实施例中,所述SaCas9结构域包含E781X、N967X和R1014X突变中的一个或多个,或本文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。在一些实施例中,所述SaCas9结构域包含E781K、N967K和R1014H突变中的一个或多个,或本文提供的任何氨基酸序列中的一个或多个相应突变。在一些实施例中,所述SaCas9结构域包含E781K、N967K或R1014H突变,或本文提供的任何氨基酸序列中的相应突变。

[0457] 示例性SaCas9序列

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRR
HRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNV
NEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAK
QLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFP
EELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAK
EILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSS
EDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRL
KLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELARE
KNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAI

[0458]

PLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEE**N**SKKGNRTPFQYLSSSDSKISY
ETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLR
SYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKL
DKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNR
ELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLK
LIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSR
NKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAE
FIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIA
SKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG

[0459] 上面加下划线和粗体的残基N579可以突变(例如突变到A579)以产生SaCas9切口酶。

[0460] 示例性SaCas9n序列

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRR
HRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNV
NEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAK
QLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFP
EELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAK

[0461]

EILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSS
EDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRL
KLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELARE
KNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAI
PLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEE**A**SKKGNRTPFQYLSSSDSKISY

[0462] ETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLR
SYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKL
DKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNR
ELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLK
LIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSR
NKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAE
FIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIA
SKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG

[0463] 上面的残基A579可以从N579突变以产生SaCas9切口酶,用下划线和粗体表示。

[0464] 示例性SaKKH Cas9

[0465] KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRR
HRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNV
NEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAK
QLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFP
EELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAK
EILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSS
EDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRL
KLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELARE
KNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAI
PLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEE**A**SKKGNRTPFQYLSSSDSKISY
ETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLR
SYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKL
DKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNR
*K*LINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLK
LIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSR
NKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAE
FIASFY*K*NDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPP*H*IIKTIA
SKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG.

[0466] 上面的残基A579可以从N579突变以产生SaCas9切口酶,用下划线和粗体表示。上面的残基K781、K967和H1014,可以从E781、N967和R1014突变以产生SaKKH Cas9,用下划线和斜体表示。

[0467] 在一些实施例中,napDNAbp是环状置换物。在以下序列中,纯文本表示腺苷脱氨酶

序列,粗体序列表示源自Cas9的序列,斜体序列表示接头序列,下划线序列表示双分核定位序列。

[0468] CP5(带有MSP“NGC”PID和“D10A”切口酶):

EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELE
NGRKRMLASAKFLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD
EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAFKYF
DTTIARKEYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD*GGSGGSGGSGGSGGSGGSG*
GM
DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF
GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFG
NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDA
ILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGY
[0469]
AGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGEL
HAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE
EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF
LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLK
IIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGW
GRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDS
LHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRE
RMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDN
LTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLK
SKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVR
KMIAKSEQEGADKRTADGSEFESPKKKRKV*

[0470] 在一些实施例中,所述核酸可编程DNA结合蛋白(napDNAbp)是微生物CRISPR-Cas系统的单一效应物。微生物CRISPR-Cas系统的单一效应器包括但不限于Cas9、Cpf1、Cas12b/C2c1和Cas12c/C2c3。通常,微生物CRISPR-Cas系统分为1类和2类系统。1类系统具有多亚基效应复合物,而2类系统具有单个蛋白质效应器。例如,Cas9和Cpf1是2类效应器。除了Cas9和Cpf1,Shmakov等人在“Discovery and Functional Characterization of Diverse Class2CRISPR Cas Systems”,Mol.Cell,2015Nov.5;60(3):385-397还描述了三

个不同的2类CRISPR-Cas系统(Cas12b/C2c1和Cas12c/C2c3),其全部内容通过引用并入本文。其中两个系统Cas12b/C2c1和Cas12c/C2c3的效应子包含与Cpf1相关的RuvC样核酸内切酶结构域。第三个系统包含一个具有两个预测HEPN RNase结构域的效应子。成熟CRISPR RNA的产生不依赖于tracrRNA,这与Cas12b/C2c1产生的CRISPR RNA不同。Cas12b/C2c1依赖CRISPR RNA和tracrRNA进行DNA切割。

[0471] 据报道,脂环酸芽胞杆菌(Alicyclobaccillus acidoterrastris)Cas12b/C2c1(AacC2c1)的晶体结构与嵌合单分子向导RNA(sgRNA)复合。见例如Liu等人,"C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism",Mol.Cell,2017Jan.19;65(2):310-322,其全部内容通过引用并入本文。在以三元复合物形式与目标DNA结合的脂环酸芽胞杆菌C2c1中也报道了所述晶体结构。见例如Yang等人,"PAM-dependent Target DNA Recognition and Cleavage by C2C1 CRISPR-Cas endonuclease",Cell,2016Dec.15;167(7):1814-1828,其全部内容通过引用并入本文。AacC2c1具有催化能力的构象,包括目标DNA链和非目标DNA链,已被独立地捕获在单个RuvC催化口袋中,Cas12b/C2c1介导的切割导致目标DNA的七核苷酸交错断裂。Cas12b/C2c1三元复合物与先前鉴定的Cas9和Cpf1对应物之间的结构比较证明了CRISPR-Cas9系统使用的机制的多样性。

[0472] 在一些实施例中,本文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。在一些实施例中,所述napDNAbp是Cas12b/C2c1蛋白。在一些实施例中,所述napDNAbp是Cas12c/C2c3蛋白。在一些实施例中,所述napDNAbp包含的氨基酸序列与天然存在的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。在一些实施例中,所述napDNAbp是天然存在的Cas12b/C2c1或Cas12c/C2c3蛋白。在一些实施例中,所述napDNAbp包含的氨基酸序列与本文提供的任何一个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。应当理解,根据本公开也可以使用来自其他细菌物种的Cas12b/C2c1或Cas12c/C2c3。

[0473] Cas12b/C2c1((uniprot.org/uniprot/T0D7A2#2)sp|T0D7A2|C2C1_ALIAG CRISPR相关核酸内切酶C2c1 OS=脂环酸芽胞杆菌(strain ATCC 49025/DSM 3922/CIP 106132/NCIMB 13137/GD3B)GN=c2c1 PE=1SV=1)氨基酸序列如下:

[0474] MAVKSIKVKLRLDDMPEIRAGLWKLHKEVNAGVRYYTEWLSLLRQENLYRRSPNGDGEQECDKTAEECKAELLERLRARQVENGHRGPAGSDDELLQLARQLYELLVPQAIGAKGDAQQIARKFLSPLADKDAVGGLGIAKAGNKPRWVRMREAGEPGWEEEKEKAETRKSADRTADVLRALADFGLKPLMRVYTDSEMSSVEWKPLRKGQAVRTWDRDMFQQAIERMMSWESWNQRVGQEYAKLVEQKNRFEQKNFVGQEHLVHLVNQLQQDMKEASPGLESKEQTAHYVTGRALRGSDKVFEKWGKLAPDAPFDLYDAEIKNVQRRNTRRFGSHDLFAKLAEPEYQALWREDASFLTRYAVYNSILRKLNHAKMFATFTLPDATAHPIWTRFDKLGGNLHQYTFLFNEFGERRHAIRFHKLLKVENGVAREVDDVTVPISMSEQLDNLLPRDPNEPIALYFRDYGAEQHFTGEFGGAKIQCRRDQLAHMHRRRGARDVYLNVSVRVQSQSEARGERRPPYAAVFRLVGDNHRAFVHFDKLSDYLAEHPDDGKLGSEGLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNSKGRVPFFFPIKGNDNLVAVHERSQLLKLPGETESKDLRAIREERQRTLRQLRTQLAYLRLLVRCGSEDVGRRERSWAKLIEQPVDAANHMTPDWREAFENELQKLKSLHGICSDKEWMDAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYAKDVVGGNSIEQIE

YLERQYKFLKSWSFFGKVSGQVIRAEKGSRFAITLREHIDHAKEDRLKKLADRIIMEALGYVYALDERGKGKWVAKY
PPCQLILLEELSEYQFNNDRPPSENNQLMQWSHRGVFQELINQAQVHDLLVGTMYAAFSSRFDARTGAPGIRCRRVP
ARCTQEHNPEPFPWWLNKFVVEHTLDACPLRADDLIPTGEGEIFVSPFSAEEGDFHQIHADLNAAQNLQQRLWSDFD
ISQIRLRCDWGEVDGELVLIPRLTGKRTADSYSNKVFYTNTGVTYYERERGKKRRKVFAQEKLSEEEAELLVEADEA
REKSVVLMRDPSGIINRGNWTRQKEFWSMVNQRIEGYLVKQIRSRVPLQDSACENTGDI

[0475] AacCas12b(脂环酸芽胞杆菌)-WP_067623834

[0476] MAVKSMKVKLRLDNMPEIRAGLWKLHTEVNAGVRYYTEWLSLLRQENLYRRSPNGDGEQECYKTAEECK
AELLERLRARQVENGHCGPAGSDDELLQLARQLYELLVPQAIGAKGDAQQIARKFLSPLADKDAVGGLGIAKAGNKP
RWVRMREAGEPGWEEEKAKAEARKSTDRTADVLRALADFGLKPLMRVYTDSDMSSVQWKPLRKGQAVRTWDRDMFQQ
AIERMMSWESWNQRVGEAYAKLVEQKSRFEQKNFVGQEHLVQLVNQLQQDMKEASHGLESKEQTAHYLTGRALRGSD
KVFEKWEKLDPDAPFDLYDTEIKNVQRRNTRRFGSHDLFAKLAEPKYQALWREDASFLTRYAVYNSIVRKLNHAKMF
ATFTLPDATAHPIWTRFDKLGGNLHQYTFLFNEFGEGRHAIRFQKLLTVEDGVAKEVDDVTVPISMSAQLDDLLPRD
PHELVALYFQDYGAEQHLAGEFGGAKIQYRRDQLNHLHARRGARDVYLNLSVRVQSQSEARGERRPPYAAVFRLVGD
NHRAFVHFDKLSDYLAEHPDDGKLGSEGLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNSEGRVPFCFPIEGNE
NLVAVHERSQLLKLPGETESKDLRAIREERQRTLRQLRTQLAYLRLLVRCGSEDVGRRERSWAKLIEQPMDANQMTP
DWREAFEDELQKLKSLYGICGDREWTEAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYQKDVVGGNSIEQIEY
LERQYKFLKSWSFFGKVSGQVIRAEKGSRFAITLREHIDHAKEDRLKKLADRIIMEALGYVYALDDERGKGKWVAKY
PPCQLILLEELSEYQFNNDRPPSENNQLMQWSHRGVFQELLNQAQVHDLLVGTMYAAFSSRFDARTGAPGIRCRRVP
ARCAREQNPEPFPWWLNKFVAEHKLDGCPLRADDLIPTGEGEFFVSPFSAEEGDFHQIHADLNAAQNLQRRLWSDFD
ISQIRLRCDWGEVDGEPVLIPRTTGKRTADSYGNKVFYTKTGVTYYERERGKKRRKVFAQEELSEEEAELLVEADEA
REKSVVLMRDPSGIINRGDWTRQKEFWSMVNQRIEGYLVKQIRSRVRLQESACENTGDI

[0477] BhCas12b(久氏芽孢杆菌)NCBI参考序列:WP_095142515

[0478] MAPKKKRKVGIHGVPAAATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYEHHEQD
PKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVEKKGEANQLSNKFLYPLVDPNSQ
SGKGTASSGRKPRWYNLKIAGDPSWEEEKKKWEEDKKKDPLAKILGKLAEYGLIPLFIPYTDSNEPIVKEIKWMEKS
RNQSVRRLDKDMFIQALERFLSWESWNLKVKEEYEKVEKEYKTLEERIKEDIQALKALEQYEKERQEQLLRDTLNTN
EYRLSKRGLRGWREIIQKWLKMDENEPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYATF
CEIDKKKKDAKQQATFTLADPINHPLWVRFEERSGSNLNKYRILTEQLHTEKLKKKLTVQLDRLIYPTESGGWEEKG
KVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGTLGGARVQFDRDHLRRYPHKVESGNVGRIYFNMTVN
IEPTESPVSKSLKIHRDDFPKVVNFKPKELTEWIKDSKGKKLKSGIESLEIGLRVMSIDLGQRQAAAASIFEVVDQK
PDIEGKLFFPIKGTELYAVHRASFNIKLPGETLVKSREVLRKAREDNLKLMNQKLNFLRNVLHFQQFEDITEREKRV
TKWISRQENSDVPLVYQDELIQIRELMYKPYKDWVAFLKQLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDE
IDRTRKFLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHLNALKEDRLKKMANTIIMHALGYCYDVRKKKWQAKNPACQ
IILFEDLSNYNPYEERSRFENSKLMKWSRREIPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKL
QDNRFFKNLQREGRLTLDKIAVLKEGDLYPDKGGEKFISLSKDRKCVTTHADINAAQNLQKRFWTRTHGFYKVYCKA
YQVDGQTVYIPESKDQKQKIIEEFGEGYFILKDGVYEWVNAGKLKIKKGSSKQSSSELVDSDILKDSFDLASELKGE
KLMLYRDPSGNVFPSDKWMAAGVFFGKLERILISKLTNQYSISTIEDDSSKQSMKRPAATKKAGQAKKKK

[0479] 包括称为BvCas12b V4的变体(S893R/K846R/E837G相对于上述的改变)。BhCas12b(V4)表示如下:5’mRNA Cap---5’UTR---bhCas12b---STOP序列---3’UTR---120polyA tail

5’UTR:GGGAAATAAGAGAGAAAAGAAGAGTAAGAAGAAATATAAGAGCCACC

[0480] 3’UTR(TriLink标准UTR)

[0481] GCTGGAGCCTCGGTGGCCATGCTTCTTGCCCCTTGGGCCTCCCCCCAGCCCCTCCTCCCCTTCCTGCAC
CCGTACCCCCGTGGTCTTTGAATAAAGTCTGA

[0482] bhCas12b(V4)的核酸序列

[0483] ATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGCCACCAGATCCTTCATC
CTGAAGATCGAGCCCAACGAGGAAGTGAAGAAAGGCCTCTGGAAAACCCACGAGGTGCTGAACCACGGAATCGCCTA
CTACATGAATATCCTGAAGCTGATCCGGCAAGAGGCCATCTACGAGCACCACGAGCAGGACCCCAAGAATCCCAAGA
AGGTGTCCAAGGCCGAGATCCAGGCCGAGCTGTGGGATTTCGTGCTGAAGATGCAGAAGTGCAACAGCTTCACACAC
GAGGTGGACAAGGACGAGGTGTTCAACATCCTGAGAGAGCTGTACGAGGAACTGGTGCCCAGCAGCGTGGAAAAGAA
GGGCGAAGCCAACCAGCTGAGCAACAAGTTTCTGTACCCTCTGGTGGACCCCAACAGCCAGTCTGGAAAGGGAACAG
CCAGCAGCGGCAGAAAGCCCAGATGGTACAACCTGAAGATTGCCGGCGATCCCTCCTGGGAAGAAGAGAAGAAGAAG
TGGGAAGAAGATAAGAAAAAGGACCCGCTGGCCAAGATCCTGGGCAAGCTGGCTGAGTACGGACTGATCCCTCTGTT
CATCCCCTACACCGACAGCAACGAGCCCATCGTGAAAGAAATCAAGTGGATGGAAAAGTCCCGGAACCAGAGCGTGC
GGCGGCTGGATAAGGACATGTTCATTCAGGCCCTGGAACGGTTCCTGAGCTGGGAGAGCTGGAACCTGAAAGTGAAA
GAGGAATACGAGAAGGTCGAGAAAGAGTACAAGACCCTGGAAGAGAGGATCAAAGAGGACATCCAGGCTCTGAAGGC
TCTGGAACAGTATGAGAAAGAGCGGCAAGAACAGCTGCTGCGGGACACCCTGAACACCAACGAGTACCGGCTGAGCA
AGAGAGGCCTTAGAGGCTGGCGGGAAATCATCCAGAAATGGCTGAAAATGGACGAGAACGAGCCCTCCGAGAAGTAC
CTGGAAGTGTTCAAGGACTACCAGCGGAAGCACCCTAGAGAGGCCGGCGATTACAGCGTGTACGAGTTCCTGTCCAA
GAAAGAGAACCACTTCATCTGGCGGAATCACCCTGAGTACCCCTACCTGTACGCCACCTTCTGCGAGATCGACAAGA
AAAAGAAGGACGCCAAGCAGCAGGCCACCTTCACACTGGCCGATCCTATCAATCACCCTCTGTGGGTCCGATTCGAG
GAAAGAAGCGGCAGCAACCTGAACAAGTACAGAATCCTGACCGAGCAGCTGCACACCGAGAAGCTGAAGAAAAAGCT
GACAGTGCAGCTGGACCGGCTGATCTACCCTACAGAATCTGGCGGCTGGGAAGAGAAGGGCAAAGTGGACATTGTGC
TGCTGCCCAGCCGGCAGTTCTACAACCAGATCTTCCTGGACATCGAGGAAAAGGGCAAGCACGCCTTCACCTACAAG
GATGAGAGCATCAAGTTCCCTCTGAAGGGCACACTCGGCGGAGCCAGAGTGCAGTTCGACAGAGATCACCTGAGAAG
ATACCCTCACAAGGTGGAAAGCGGCAACGTGGGCAGAATCTACTTCAACATGACCGTGAACATCGAGCCTACAGAGT
CCCCAGTGTCCAAGTCTCTGAAGATCCACCGGGACGACTTCCCCAAGGTGGTCAACTTCAAGCCCAAAGAACTGACC
GAGTGGATCAAGGACAGCAAGGGCAAGAAACTGAAGTCCGGCATCGAGTCCCTGGAAATCGGCCTGAGAGTGATGAG
CATCGACCTGGGACAGAGACAGGCCGCTGCCGCCTCTATTTTCGAGGTGGTGGATCAGAAGCCCGACATCGAAGGCA
AGCTGTTTTTCCCAATCAAGGGCACCGAGCTGTATGCCGTGCACAGAGCCAGCTTCAACATCAAGCTGCCCGGCGAG
ACACTGGTCAAGAGCAGAGAAGTGCTGCGGAAGGCCAGAGAGGACAATCTGAAACTGATGAACCAGAAGCTCAACTT
CCTGCGGAACGTGCTGCACTTCCAGCAGTTCGAGGACATCACCGAGAGAGAGAAGCGGGTCACCAAGTGGATCAGCA
GACAAGAGAACAGCGACGTGCCCCTGGTGTACCAGGATGAGCTGATCCAGATCCGCGAGCTGATGTACAAGCCTTAC
AAGGACTGGGTCGCCTTCCTGAAGCAGCTCCACAAGAGACTGGAAGTCGAGATCGGCAAAGAAGTGAAGCACTGGCG
GAAGTCCCTGAGCGACGGAAGAAAGGGCCTGTACGGCATCTCCCTGAAGAACATCGACGAGATCGATCGGACCCGGA
AGTTCCTGCTGAGATGGTCCCTGAGGCCTACCGAACCTGGCGAAGTGCGTAGACTGGAACCCGGCCAGAGATTCGCC
ATCGACCAGCTGAATCACCTGAACGCCCTGAAAGAAGATCGGCTGAAGAAGATGGCCAACACCATCATCATGCACGC
CCTGGGCTACTGCTACGACGTGCGGAAGAAGAAATGGCAGGCTAAGAACCCCGCCTGCCAGATCATCCTGTTCGAGG
ATCTGAGCAACTACAACCCCTACGAGGAAAGGTCCCGCTTCGAGAACAGCAAGCTCATGAAGTGGTCCAGACGCGAG

ATCCCCAGACAGGTTGCACTGCAGGGCGAGATCTATGGCCTGCAAGTGGGAGAAGTGGGCGCTCAGTTCAGCAGCAG
ATTCCACGCCAAGACAGGCAGCCCTGGCATCAGATGTAGCGTCGTGACCAAAGAGAAGCTGCAGGACAATCGGTTCT
TCAAGAATCTGCAGAGAGAGGGCAGACTGACCCTGGACAAAATCGCCGTGCTGAAAGAGGGCGATCTGTACCCAGAC
AAAGGCGGCGAGAAGTTCATCAGCCTGAGCAAGGATCGGAAGTGCGTGACCACACACGCCGACATCAACGCCGCTCA
GAACCTGCAGAAGCGGTTCTGGACAAGAACCCACGGCTTCTACAAGGTGTACTGCAAGGCCTACCAGGTGGACGGCC
AGACCGTGTACATCCCTGAGAGCAAGGACCAGAAGCAGAAGATCATCGAAGAGTTCGGCGAGGGCTACTTCATTCTG
AAGGACGGGGTGTACGAATGGGTCAACGCCGGCAAGCTGAAAATCAAGAAGGGCAGCTCCAAGCAGAGCAGCAGCGA
GCTGGTGGATAGCGACATCCTGAAAGACAGCTTCGACCTGGCCTCCGAGCTGAAAGGCGAAAAGCTGATGCTGTACA
GGGACCCCAGCGGCAATGTGTTCCCCAGCGACAAATGGATGGCCGCTGGCGTGTTCTTCGGAAAGCTGGAACGCATC
CTGATCAGCAAGCTGACCAACCAGTACTCCATCAGCACCATCGAGGACGACAGCAGCAAGCAGTCTATGAAAAGGCC
GGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAG

[0484] 在一些实施例中,所述Cas12b是BvCas12B。在一些实施例中,所述Cas12b包含氨基酸置换S893R、K846R和E837G,如下文提供的BvCas12B示例性序列中编号。

[0485] BvCas12b(芽孢梓菌屬V3-13)NCBI参考序列:WP_101661451.1

[0486] MAIRSIKLKMKTNSGTDSIYLRKALWRTHQLINEGIAYYMNLLTLYRQEAIGDKTKEAYQAELINIIRN
QQRNNGSSEEHGSDQEILALLRQLYELIIPSSIGESGDANQLGNKFLYPLVDPNSQSGKGTSNAGRKPRWKRLKEEG
NPDWELEKKKDEERKAKDPTVKIFDNLNKYGLLPLFPLFTNIQKDIEWLPLGKRQSVRKWDKDMFIQAIERLLSWES
WNRRVADEYKQLKEKTESYYKEHLTGGEEWIEKIRKFEKERNMELEKNAFAPNDGYFITSRQIRGWDRVYEKWSKLP
ESASPEELWKVVAEQQNKMSEGFGDPKVFSFLANRENRDIWRGHSERIYHIAAYNGLQKKLSRTKEQATFTLPDAIE
HPLWIRYESPGGTNLNLFKLEEKQKKNYYVTLSKIIWPSEEKWIEKENIEIPLAPSIQFNRQIKLKQHVKGKQEISF
SDYSSRISLDGVLGGSRIQFNRKYIKNHKELLGEGDIGPVFFNLVVDVAPLQETRNGRLQSPIGKALKVISSDFSKV
IDYKPKELMDWMNTGSASNSFGVASLLEGMRVMSIDMGQRTSASVSIFEVVKELPKDQEQKLFYSINDTELFAIHKR
SFLLNLPGEVVTKNNKQQRQERRKKRQFVRSQIRMLANVLRLETKKTPDERKKAIHKLMEIVQSYDSWTASQKEVWE
KELNLLTNMAAFNDEIWKESLVELHHRIEPYVGQIVSKWRKGLSEGRKNLAGISMWNIDELEDTRRLLISWSKRSRT
PGEANRIETDEPFGSSLLQHIQNVKDDRLKQMANLIIMTALGFKYDKEEKDRYKRWKETYPACQIILFENLNRYLFN
LDRSRRENSRLMKWAHRSIPRTVSMQGEMFGLQVGDVRSEYSSRFHAKTGAPGIRCHALTEEDLKAGSNTLKRLIED
GFINESELAYLKKGDIIPSQGGELFVTLSKRYKKDSDNNELTVIHADINAAQNLQKRFWQQNSEVYRVPCQLARMGE
DKLYIPKSQTETIKKYFGKGSFVKNNTEQEVYKWEKSEKMKIKTDTTFDLQDLDGFEDISKTIELAQEQQKKYLTMF
RDPSGYFFNNETWRPQKEYWSIVNNIIKSCLKKKILSNKVEL

[0487] 在一些实施例中,Cas12b是BTCas12b。BTCas12b(嗜热淀粉芽孢杆菌)NCBI参考序列:WP_041902512

[0488] MATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYEHHEQDPKNPKKVSKAEIQAEL
WDFVLKMQKCNSFTHEVDKDVVFNILRELYEELVPSSVEKKGEANQLSNKFLYPLVDPNSQSGKGTASSGRKPRWYN
LKIAGDPSWEEEKKKWEEDKKKDPLAKILGKLAEYGLIPLFIPFTDSNEPIVKEIKWMEKSRNQSVRRLDKDMFIQA
LERFLSWESWNLKVKEEYEKVEKEHKTLEERIKEDIQAFKSLEQYEKERQEQLLRDTLNTNEYRLSKRGLRGWREII
QKWLKMDENEPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYATFCEIDKKKKDAKQQATF
TLADPINHPLWVRFEERSGSNLNKYRILTEQLHTEKLKKKLTVQLDRLIYPTESGGWEEKGKVDIVLLPSRQFYNQI
FLDIEEKGKHAFTYKDESIKFPLKGTLGGARVQFDRDHLRRYPHKVESGNVGRIYFNMTVNIEPTESPVSKSLKIHR
DDFPKFVNFKPKELTEWIKDSKGKKLKSGIESLEIGLRVMSIDLGQRQAAAASIFEVVDQKPDIEGKLFFPIKGTEL

YAVHRASFNIKLPGETLVKSREVLRKAREDNLKLMNQKLNFLRNVLHFQQFEDITEREKRVTKWISRQENSDVPLVY
QDELIQIRELMYKPYKDWVAFLKQLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDEIDRTRKFLLRWSLRPT
EPGEVRRLEPGQRFAIDQLNHLNALKEDRLKKMANTIIMHALGYCYDVRKKKWQAKNPACQIILFEDLSNYNPYEER
SRFENSKLMKWSRREIPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKLQDNRFFKNLQREGRLT
LDKIAVLKEGDLYPDKGGEKFISLSKDRKLVTTHADINAAQNLQKRFWTRTHGFYKVYCKAYQVDGQTVYIPESKDQ
KQKIIEEFGEGYFILKDGVYEWGNAGKLKIKKGSSKQSSSELVDSDILKDSFDLASELKGEKLMLYRDPSGNVFPSD
KWMAAGVFFGKLERILISKLTNQYSISTIEDDSSKQSM

[0489] 在一些实施例中,napDNAbp是指Cas12c。在一些实施例中,Cas12c蛋白是Cas12c1或Cas12c1的变体。在一些实施例中,Cas12蛋白是Cas12c2或Cas12c2的变体。在一些实施例中,所述Cas12蛋白是来自嗜油菌属的HI0009(即OspCas12c)Cas12c蛋白或OspCas12c的变体。这些Cas12c分子已在Yan等人,“Functionally Diverse Type V CRISPR-Cas Systems,”Science,2019Jan.4;363:88-91;中进行了描述,其全部内容通过引用并入本文。在一些实施例中,所述napDNAbp包含的氨基酸序列与天然存在的Cas12c1、Cas12c2或OspCas12c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。在一些实施例中,所述napDNAbp是天然存在的Cas12c1、Cas12c2或OspCas12c蛋白。在一些实施例中,所述napDNAbp包含的氨基酸序列与本文所述的任何Cas12c1、Cas12c2或OspCas12c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。应当理解,根据本公开也可以使用来自其他细菌物种的Cas12c1、Cas12c2或OspCas12c。

[0490] Cas12c1

[0491] MQTKKTHLHLISAKASRKYRRTIACLSDTAKKDLERRKQSGAADPAQELSCLKTIKFKLEVPEGSKLPS
FDRISQIYNALETIEKGSLSYLLFALILSGFRIFPNSSAAKTFASSSCYKNDQFASQIKEIFGEMVKNFIPSELESI
LKKGRRKNNKDWTEENIKRVLNSEFGRKNSEGSSALFDSFLSKFSQELFRKFDSWNEVNKKYLEAAELLDSMLASYG
PFDSVCKMIGDSDSRNSLPDKSTIAFTNNAEITVDIESSVMPYMAIAALLREYRQSKSKAAPVAYVQSHLTTTNGNG
LSWFFKFGLDLIRKAPVSSKQSTSDGSKSLQELFSVPDDKLDGLKFIKEACEALPEASLLCGEKGELLGYQDFRTSF
AGHIDSWVANYVNRLFELIELVNQLPESIKLPSILTQKNHNLVASLGLQEAEVSHSLELFEGLVKNVRQTLKKLAGI
DISSSPNEQDIKEFYAFSDVLNRLGSIRNQIENAVQTAKKDKIDLESAIEWKEWKKLKKLPKLNGLGGGVPKQQELL
DKALESVKQIRHYQRIDFERVIQWAVNEHCLETVPKFLVDAEKKKINKESSTDFAAKENAVRFLLEGIGAAARGKTD
SVSKAAYNWFVVNNFLAKKDLNRYFINCQGCIYKPPYSKRRSLAFALRSDNKDTIEVVWEKFETFYKEISKEIEKFN
IFSQEFQTFLHLENLRMKLLLRRIQKPIPAEIAFFSLPQEYYDSLPPNVAFLALNQEITPSEYITQFNLYSSFLNGN
LILLRRSRSYLRAKFSWVGNSKLIYAAKEARLWKIPNAYWKSDEWKMILDSNVLVFDKAGNVLPAPTLKKVCEREGD
LRLFYPLLRQLPHDWCYRNPFVKSVGREKNVIEVNKEGEPKVASALPGSLFRLIGPAPFKSLLDDCFFNPLDKDLRE
CMLIVDQEISQKVEAQKVEASLESCTYSIAVPIRYHLEEPKVSNQFENVLAIDQGEAGLAYAVFSLKSIGEAETKPI
AVGTIRIPSIRRLIHSVSTYRKKKQRLQNFKQNYDSTAFIMRENVTGDVCAKIVGLMKEFNAFPVLEYDVKNLESGS
RQLSAVYKAVNSHFLYFKEPGRDALRKQLWYGGDSWTIDGIEIVTRERKEDGKEGVEKIVPLKVFPGRSVSARFTSK
TCSCCGRNVFDWLFTEKKAKTNKKFNVNSKGELTTADGVIQLFEADRSKGPKFYARRKERTPLTKPIAKGSYSLEEI
ERRVRTNLRRAPKSKQSRDTSQSQYFCVYKDCALHFSGMQADENAAINIGRRFLTALRKNRRSDFPSNVKISDRLLD
N

[0492] Cas12c2

[0493] MTKHSIPLHAFRNSGADARKWKGRIALLAKRGKETMRTLQFPLEMSEPEAAAINTTPFAVAYNAIEGTG
KGTLFDYWAKLHLAGFRFFPSGGAATIFRQQAVFEDASWNAAFCQQSGKDWPWLVPSKLYERFTKAPREVAKKDGSK
KSIEFTQENVANESHVSLVGASITDKTPEDQKEFFLKMAGALAEKFDSWKSANEDRIVAMKVIDEFLKSEGLHLPSL
ENIAVKCSVETKPDNATVAWHDAPMSGVQNLAIGVFATCASRIDNIYDLNGGKLSKLIQESATTPNVTALSWLFGKG
LEYFRTTDIDTIMQDFNIPASAKESIKPLVESAQAIPTMTVLGKKNYAPFRPNFGGKIDSWIANYASRLMLLNDILE
QIEPGFELPQALLDNETLMSGIDMTGDELKELIEAVYAWVDAAKQGLATLLGRGGNVDDAVQTFEQFSAMMDTLNGT
LNTISARYVRAVEMAGKDEARLEKLIECKFDIPKWCKSVPKLVGISGGLPKVEEEIKVMNAAFKDVRARMFVRFEEI
AAYVASKGAGMDVYDALEKRELEQIKKLKSAVPERAHIQAYRAVLHRIGRAVQNCSEKTKQLFSSKVIEMGVFKNPS
HLNNFIFNQKGAIYRSPFDRSRHAPYQLHADKLLKNDWLELLAEISATLMASESTEQMEDALRLERTRLQLQLSGLP
DWEYPASLAKPDIEVEIQTALKMQLAKDTVTSDVLQRAFNLYSSVLSGLTFKLLRRSFSLKMRFSVADTTQLIYVPK
VCDWAIPKQYLQAEGEIGIAARVVTESSPAKMVTEVEMKEPKALGHFMQQAPHDWYFDASLGGTQVAGRIVEKGKEV
GKERKLVGYRMRGNSAYKTVLDKSLVGNTELSQCSMIIEIPYTQTVDADFRAQVQAGLPKVSINLPVKETITASNKD
EQMLFDRFVAIDLGERGLGYAVFDAKTLELQESGHRPIKAITNLLNRTHHYEQRPNQRQKFQAKFNVNLSELRENTV
GDVCHQINRICAYYNAFPVLEYMVPDRLDKQLKSVYESVTNRYIWSSTDAHKSARVQFWLGGETWEHPYLKSAKDKK
PLVLSPGRGASGKGTSQTCSCCGRNPFDLIKDMKPRAKIAVVDGKAKLENSELKLFERNLESKDDMLARRHRNERAG
MEQPLTPGNYTVDEIKALLRANLRRAPKNRRTKDTTVSEYHCVFSDCGKTMHADENAAVNIGGKFIADIEK

[0494] OspCas12c

[0495] MTKLRHRQKKLTHDWAGSKKREVLGSNGKLQNPLLMPVKKGQVTEFRKAFSAYARATKGEMTDGRKNMF
THSFEPFKTKPSLHQCELADKAYQSLHSYLPGSLAHFLLSAHALGFRIFSKSGEATAFQASSKIEAYESKLASELAC
VDLSIQNLTISTLFNALTTSVRGKGEETSADPLIARFYTLLTGKPLSRDTQGPERDLAEVISRKIASSFGTWKEMTA
NPLQSLQFFEEELHALDANVSLSPAFDVLIKMNDLQGDLKNRTIVFDPDAPVFEYNAEDPADIIIKLTARYAKEAVI
KNQNVGNYVKNAITTTNANGLGWLLNKGLSLLPVSTDDELLEFIGVERSHPSCHALIELIAQLEAPELFEKNVFSDT
RSEVQGMIDSAVSNHIARLSSSRNSLSMDSEELERLIKSFQIHTPHCSLFIGAQSLSQQLESLPEALQSGVNSADIL
LGSTQYMLTNSLVEESIATYQRTLNRINYLSGVAGQINGAIKRKAIDGEKIHLPAAWSELISLPFIGQPVIDVESDL
AHLKNQYQTLSNEFDTLISALQKNFDLNFNKALLNRTQHFEAMCRSTKKNALSKPEIVSYRDLLARLTSCLYRGSLV
LRRAGIEVLKKHKIFESNSELREHVHERKHFVFVSPLDRKAKKLLRLTDSRPDLLHVIDEILQHDNLENKDRESLWL
VRSGYLLAGLPDQLSSSFINLPIITQKGDRRLIDLIQYDQINRDAFVMLVTSAFKSNLSGLQYRANKQSFVVTRTLS
PYLGSKLVYVPKDKDWLVPSQMFEGRFADILQSDYMVWKDAGRLCVIDTAKHLSNIKKSVFSSEEVLAFLRELPHRT
FIQTEVRGLGVNVDGIAFNNGDIPSLKTFSNCVQVKVSRTNTSLVQTLNRWFEGGKVSPPSIQFERAYYKKDDQIHE
DAAKRKIRFQMPATELVHASDDAGWTPSYLLGIDPGEYGMGLSLVSINNGEVLDSGFIHINSLINFASKKSNHQTKV
VPRQQYKSPYANYLEQSKDSAAGDIAHILDRLIYKLNALPVFEALSGNSQSAADQVWTKVLSFYTWGDNDAQNSIRK
QHWFGASHWDIKGMLRQPPTEKKPKPYIAFPGSQVSSYGNSQRCSCCGRNPIEQLREMAKDTSIKELKIRNSEIQLF
DGTIKLFNPDPSTVIERRRHNLGPSRIPVADRTFKNISPSSLEFKELITIVSRSIRHSPEFIAKKRGIGSEYFCAYS
DCNSSLNSEANAAANVAQKFQKQLFFEL

[0496] 在一些实施例中,所述napDNAbp是指Cas12g、Cas12h或Cas12i,其已在例如Yan等人,“Functionally Diverse Type VCRISPR-Cas Systems,”Science,2019Jan.4;363:88-91中描述,其各自的全部内容通过引用并入本文。通过聚合超过10TB的序列数据,确定了V型Cas蛋白的新分类,这些分类与先前表征的V类蛋白(包括Cas12g、Cas12h和Cas12i)表现

出微弱的相似性。在一些实施例中,所述Cas12蛋白是Cas12g或Cas12g的变体。在一些实施例中,所述Cas12蛋白是Cas12h或Cas12h的变体。在一些实施例中,所述Cas12蛋白是Cas12i或Cas12i的变体。应当理解,其他RNA引导的DNA结合蛋白可以用作napDNAbp,并且在本公开的范围内。在一些实施例中,所述napDNAbp包含的氨基酸序列与天然存在的Cas12g、Cas12h或Cas12i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。在一些实施例中,所述napDNAbp是天然存在的Cas12g、Cas12h或Cas12i蛋白。在一些实施例中,所述napDNAbp包含的氨基酸序列与本文所述的任何Cas12g、Cas12h或Cas12i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。应当理解,根据本公开也可以使用来自其他细菌物种的Cas12g、Cas12h或Cas12i。在一些实施例中,Cas12i是Cas12i1或Cas12i2。

[0497] Cas12g1

[0498] MAQASSTPAVSPRPRPRYREERTLVRKLLPRPGQSKQEFRENVKKLRKAFLQFNADVSGVCQWAIQFRP
RYGKPAEPTETFWKFFLEPETSLPPNDSRSPEFRRLQAFEAAAGINGAAALDDPAFTNELRDSILAVASRPKTKEAQ
RLFSRLKDYQPAHRMILAKVAAEWIESRYRRAHQNWERNYEEWKKEKQEWEQNHPELTPEIREAFNQIFQQLEVKEK
RVRICPAARLLQNKDNCQYAGKNKHSVLCNQFNEFKKNHLQGKAIKFFYKDAEKYLRCGLQSLKPNVQGPFREDWNK
YLRYMNLKEETLRGKNGGRLPHCKNLGQECEFNPHTALCKQYQQQLSSRPDLVQHDELYRKWRREYWREPRKPVFRY
PSVKRHSIAKIFGENYFQADFKNSVVGLRLDSMPAGQYLEFAFAPWPRNYRPQPGETEISSVHLHFVGTRPRIGFRF
RVPHKRSRFDCTQEELDELRSRTFPRKAQDQKFLEAARKRLLETFPGNAEQELRLLAVDLGTDSARAAFFIGKTFQQ
AFPLKIVKIEKLYEQWPNQKQAGDRRDASSKQPRPGLSRDHVGRHLQKMRAQASEIAQKRQELTGTPAPETTTDQAA
KKATLQPFDLRGLTVHTARMIRDWARLNARQIIQLAEENQVDLIVLESLRGFRPPGYENLDQEKKRRVAFFAHGRIR
RKVTEKAVERGMRVVTVPYLASSKVCAECRKKQKDNKQWEKNKKRGLFKCEGCGSQAQVDENAARVLGRVFWGEIEL
PTAIP

[0499] Cas12h1

[0500] MKVHEIPRSQLLKIKQYEGSFVEWYRDLQEDRKKFASLLFRWAAFGYAAREDDGATYISPSQALLERRL
LLGDAEDVAIKFLDVLFKGGAPSSSCYSLFYEDFALRDKAKYSGAKREFIEGLATMPLDKIIERIRQDEQLSKIPAE
EWLILGAEYSPEEIWEQVAPRIVNVDRSLGKQLRERLGIKCRRPHDAGYCKILMEVVARQLRSHNETYHEYLNQTHE
MKTKVANNLTNEFDLVCEFAEVLEEKNYGLGWYVLWQGVKQALKEQKKPTKIQIAVDQLRQPKFAGLLTAKWRALKG
AYDTWKLKKRLEKRKAFPYMPNWDNDYQIPVGLTGLGVFTLEVKRTEVVVDLKEHGKLFCSHSHYFGDLTAEKHPSR
YHLKFRHKLKLRKRDSRVEPTIGPWIEAALREITIQKKPNGVFYLGLPYALSHGIDNFQIAKRFFSAAKPDKEVING
LPSEMVVGAADLNLSNIVAPVKARIGKGLEGPLHALDYGYGELIDGPKILTPDGPRCGELISLKRDIVEIKSAIKEF
KACQREGLTMSEETTTWLSEVESPSDSPRCMIQSRIADTSRRLNSFKYQMNKEGYQDLAEALRLLDAMDSYNSLLES
YQRMHLSPGEQSPKEAKFDTKRASFRDLLRRRVAHTIVEYFDDCDIVFFEDLDGPSDSDSRNNALVKLLSPRTLLLY
IRQALEKRGIGMVEVAKDGTSQNNPISGHVGWRNKQNKSEIYFYEDKELLVMDADEVGAMNILCRGLNHSVCPYSFV
TKAPEKKNDEKKEGDYGKRVKRFLKDRYGSSNVRFLVASMGFVTVTTKRPKDALVGKRLYYHGGELVTHDLHNRMKD
EIKYLVEKEVLARRVSLSDSTIKSYKSFAHV

[0501] Cas12i1

[0502] MSNKEKNASETRKAYTTKMIPRSHDRMKLLGNFMDYLMDGTPIFFELWNQFGGGIDRDIISGTANKDKI
SDDLLLAVNWFKVMPINSKPQGVSPSNLANLFQQYSGSEPDIQAQEYFASNFDTEKHQWKDMRVEYERLLAELQLSR

SDMHHDLKLMYKEKCIGLSLSTAHYITSVMFGTGAKNNRQTKHQFYSKVIQLLEESTQINSVEQLASIILKAGDCDS
YRKLRIRCSRKGATPSILKIVQDYELGTNHDDEVNVPSLIANLKEKLGRFEYECEWKCMEKIKAFLASKVGPYYLGS
YSAMLENALSPIKGMTTKNCKFVLKQIDAKNDIKYENEPFGKIVEGFFDSPYFESDTNVKWVLHPHHIGESNIKTLW
EDLNAIHSKYEEDIASLSEDKKEKRIKVYQGDVCQTINTYCEEVGKEAKTPLVQLLRYLYSRKDDIAVDKIIDGITF
LSKKHKVEKQKINPVIQKYPSFNFGNNSKLLGKIISPKDKLKHNLKCNRNQVDNYIWIEIKVLNTKTMRWEKHHYAL
SSTRFLEEVYYPATSENPPDALAARFRTKTNGYEGKPALSAEQIEQIRSAPVGLRKVKKRQMRLEAARQQNLLPRYT
WGKDFNINICKRGNNFEVTLATKVKKKKEKNYKVVLGYDANIVRKNTYAAIEAHANGDGVIDYNDLPVKPIESGFVT
VESQVRDKSYDQLSYNGVKLLYCKPHVESRRSFLEKYRNGTMKDNRGNNIQIDFMKDFEAIADDETSLYYFNMKYCK
LLQSSIRNHSSQAKEYREEIFELLRDGKLSVLKLSSLSNLSFVMFKVAKSLIGTYFGHLLKKPKNSKSDVKAPPITD
EDKQKADPEMFALRLALEEKRLNKVKSKKEVIANKIVAKALELRDKYGPVLIKGENISDTTKKGKKSSTNSFLMDWL
ARGVANKVKEMVMMHQGLEFVEVNPNFTSHQDPFVHKNPENTFRARYSRCTPSELTEKNRKEILSFLSDKPSKRPTN
AYYNEGAMAFLATYGLKKNDVLGVSLEKFKQIMANILHQRSEDQLLFPSRGGMFYLATYKLDADATSVNWNGKQFWV
CNADLVAAYNVGLVDIQKDFKKK

[0503] Cas12i2

[0504] MSSAIKSYKSVLRPNERKNQLLKSTIQCLEDGSAFFFKMLQGLFGGITPEIVRFSTEQEKQQQDIALWC
AVNWFRPVSQDSLTHTIASDNLVEKFEEYYGGTASDAIKQYFSASIGESYYWNDCRQQYYDLCRELGVEVSDLTHDL
EILCREKCLAVATESNQNNSIISVLFGTGEKEDRSVKLRITKKILEAISNLKEIPKNVAPIQEIILNVAKATKETFR
QVYAGNLGAPSTLEKFIAKDGQKEFDLKKLQTDLKKVIRGKSKERDWCCQEELRSYVEQNTIQYDLWAWGEMFNKAH
TALKIKSTRNYNFAKQRLEQFKEIQSLNNLLVVKKLNDFFDSEFFSGEETYTICVHHLGGKDLSKLYKAWEDDPADP
ENAIVVLCDDLKNNFKKEPIRNILRYIFTIRQECSAQDILAAAKYNQQLDRYKSQKANPSVLGNQGFTWTNAVILPE
KAQRNDRPNSLDLRIWLYLKLRHPDGRWKKHHIPFYDTRFFQEIYAAGNSPVDTCQFRTPRFGYHLPKLTDQTAIRV
NKKHVKAAKTEARIRLAIQQGTLPVSNLKITEISATINSKGQVRIPVKFDVGRQKGTLQIGDRFCGYDQNQTASHAY
SLWEVVKEGQYHKELGCFVRFISSGDIVSITENRGNQFDQLSYEGLAYPQYADWRKKASKFVSLWQITKKNKKKEIV
TVEAKEKFDAICKYQPRLYKFNKEYAYLLRDIVRGKSLVELQQIRQEIFRFIEQDCGVTRLGSLSLSTLETVKAVKG
IIYSYFSTALNASKNNPISDEQRKEFDPELFALLEKLELIRTRKKKQKVERIANSLIQTCLENNIKFIRGEGDLSTT
NNATKKKANSRSMDWLARGVFNKIRQLAPMHNITLFGCGSLYTSHQDPLVHRNPDKAMKCRWAAIPVKDIGDWVLRK
LSQNLRAKNIGTGEYYHQGVKEFLSHYELQDLEEELLKWRSDRKSNIPCWVLQNRLAEKLGNKEAVVYIPVRGGRIY
FATHKVATGAVSIVFDQKQVWVCNADHVAAANIALTVKGIGEQSSDEENPDGSRIKLQLTS

[0505] 碱基编辑器的代表性核酸和蛋白质序列如下：

[0506] BhCas12b GGSGGS-ABE8-Xten20在P153

[0507]

GCCACCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGCC
ACCAGATCCTTCATCCTGAAGATCGAGCCCAACGAGGAAGTGAAGAAAGGCCTCTGGAAA
ACCCACGAGGTGCTGAACCACGGAATCGCCTACTACATGAATATCCTGAAGCTGATCCGG
CAAGAGGCCATCTACGAGCACCACGAGCAGGACCCCAAGAATCCCAAGAAGGTGTCCAAG
GCCGAGATCCAGGCCGAGCTGTGGGATTTCGTGCTGAAGATGCAGAAGTGCAACAGCTTC
ACACACGAGGTGGACAAGGACGAGGTGTTCAACATCCTGAGAGAGCTGTACGAGGAACTG
GTGCCCAGCAGCGTGGAAAAGAAGGGCGAAGCCAACCAGCTGAGCAACAAGTTTCTGTAC
CCTCTGGTGGACCCCAACAGCCAGTCTGGAAAGGGAACAGCCAGCAGCGGCAGAAAGCCC
AGATGGTACAACCTGAAGATTGCCGGCGATCCCggaggctctggaggaagcTCCGAAGTC
GAGTTTTCCCATGAGTACTGGATGAGACACGCATTGACTCTCGCAAAGAGGGCTCGAGAT
GAACGCGAGGTGCCCGTGGGGGCAGTACTCGTGCTCAACAATCGCGTAATCGGCGAAGGT
TGGAATAGGGCAATCGGACTCCACGACCCCACTGCACATGCGGAAATCATGGCCCTTCGA
CAGGGAGGGCTTGTGATGCAGAATTATCGACTTTATGATGCGACGCTGTACGTCACGTTT
GAACCTTGCGTAATGTGCGCGGGAGCTATGATTCACTCCCGCATTGGACGAGTTGTATTC
GGTGTTCGCAACGCCAAGACGGGTGCCGCAGGTTCACTGATGGACGTGCTGCATCATCCA
GGCATGAACCACCGGGTAGAAATCACAGAAGGCATATTGGCGGACGAATGTGCGGCGCTG
TTGTGTCGTTTTTTTCGCATGCCCAGGCGGGTCTTTAACGCCCAGAAAAAAGCACAATCC
TCTACTGACGGCTCTTCTGGATCTGAAACACCTGGCACAAGCGAGAGCGCCACCCCTGAG
AGCTCTGGCTCCTGGGAAGAAGAGAAGAAGAAGTGGGAAGAAGATAAGAAAAAGGACCCG
CTGGCCAAGATCCTGGGCAAGCTGGCTGAGTACGGACTGATCCCTCTGTTCATCCCCTAC

ACCGACAGCAACGAGCCCATCGTGAAAGAAATCAAGTGGATGGAAAAGTCCCGGAACCAG

AGCGTGCGGCGGCTGGATAAGGACATGTTCATTCAGGCCCTGGAACGGTTCCTGAGCTGG

GAGAGCTGGAACCTGAAAGTGAAAGAGGAATACGAGAAGGTCGAGAAAGAGTACAAGACC

CTGGAAGAGAGGATCAAAGAGGACATCCAGGCTCTGAAGGCTCTGGAACAGTATGAGAAA

GAGCGGCAAGAACAGCTGCTGCGGGACACCCTGAACACCAACGAGTACCGGCTGAGCAAG

AGAGGCCTTAGAGGCTGGCGGGAAATCATCCAGAAATGGCTGAAAATGGACGAGAACGAG

CCCTCCGAGAAGTACCTGGAAGTGTTCAAGGACTACCAGCGGAAGCACCCTAGAGAGGCC

GGCGATTACAGCGTGTACGAGTTCCTGTCCAAGAAAGAGAACCACTTCATCTGGCGGAAT

CACCCTGAGTACCCCTACCTGTACGCCACCTTCTGCGAGATCGACAAGAAAAAGAAGGAC

GCCAAGCAGCAGGCCACCTTCACACTGGCCGATCCTATCAATCACCCTCTGTGGGTCCGA

TTCGAGGAAAGAAGCGGCAGCAACCTGAACAAGTACAGAATCCTGACCGAGCAGCTGCAC

ACCGAGAAGCTGAAGAAAAAGCTGACAGTGCAGCTGGACCGGCTGATCTACCCTACAGAA

TCTGGCGGCTGGGAAGAGAAGGGCAAAGTGGACATTGTGCTGCTGCCCAGCCGGCAGTTC

TACAACCAGATCTTCCTGGACATCGAGGAAAAGGGCAAGCACGCCTTCACCTACAAGGAT

GAGAGCATCAAGTTCCCTCTGAAGGGCACACTCGGCGGAGCCAGAGTGCAGTTCGACAGA

[0508] GATCACCTGAGAAGATACCCTCACAAGGTGGAAAGCGGCAACGTGGGCAGAATCTACTTC

AACATGACCGTGAACATCGAGCCTACAGAGTCCCCAGTGTCCAAGTCTCTGAAGATCCAC

CGGGACGACTTCCCCAAGGTGGTCAACTTCAAGCCCAAAGAACTGACCGAGTGGATCAAG

GACAGCAAGGGCAAGAAACTGAAGTCCGGCATCGAGTCCCTGGAAATCGGCCTGAGAGTG

ATGAGCATCGACCTGGGACAGAGACAGGCCGCTGCCGCCTCTATTTTCGAGGTGGTGGAT

CAGAAGCCCGACATCGAAGGCAAGCTGTTTTTCCCAATCAAGGGCACCGAGCTGTATGCC

GTGCACAGAGCCAGCTTCAACATCAAGCTGCCCGGCGAGACACTGGTCAAGAGCAGAGAA

GTGCTGCGGAAGGCCAGAGAGGACAATCTGAAACTGATGAACCAGAAGCTCAACTTCCTG

CGGAACGTGCTGCACTTCCAGCAGTTCGAGGACATCACCGAGAGAGAGAAGCGGGTCACC

AAGTGGATCAGCAGACAAGAGAACAGCGACGTGCCCCTGGTGTACCAGGATGAGCTGATC

CAGATCCGCGAGCTGATGTACAAGCCTTACAAGGACTGGGTCGCCTTCCTGAAGCAGCTC

CACAAGAGACTGGAAGTCGAGATCGGCAAAGAAGTGAAGCACTGGCGGAAGTCCCTGAGC

GACGGAAGAAAGGGCCTGTACGGCATCTCCCTGAAGAACATCGACGAGATCGATCGGACC

CGGAAGTTCCTGCTGAGATGGTCCCTGAGGCCTACCGAACCTGGCGAAGTGCGTAGACTG

GAACCCGGCCAGAGATTCGCCATCGACCAGCTGAATCACCTGAACGCCCTGAAAGAAGAT

CGGCTGAAGAAGATGGCCAACACCATCATCATGCACGCCCTGGGCTACTGCTACGACGTG

[0509] CGGAAGAAGAAATGGCAGGCTAAGAACCCCGCCTGCCAGATCATCCTGTTCGAGGATCTG
AGCAACTACAACCCCTACGAGGAAAGGTCCCGCTTCGAGAACAGCAAGCTCATGAAGTGG
TCCAGACGCGAGATCCCCAGACAGGTTGCACTGCAGGGCGAGATCTATGGCCTGCAAGTG
GGAGAAGTGGGCGCTCAGTTCAGCAGCAGATTCCACGCCAAGACAGGCAGCCCTGGCATC
AGATGTAGCGTCGTGACCAAAGAGAAGCTGCAGGACAATCGGTTCTTCAAGAATCTGCAG
AGAGAGGGCAGACTGACCCTGGACAAAATCGCCGTGCTGAAAGAGGGCGATCTGTACCCA
GACAAAGGCGGCGAGAAGTTCATCAGCCTGAGCAAGGATCGGAAGTGCGTGACCACACAC
GCCGACATCAACGCCGCTCAGAACCTGCAGAAGCGGTTCTGGACAAGAACCCACGGCTTC
TACAAGGTGTACTGCAAGGCCTACCAGGTGGACGGCCAGACCGTGTACATCCCTGAGAGC
AAGGACCAGAAGCAGAAGATCATCGAAGAGTTCGGCGAGGGCTACTTCATTCTGAAGGAC
GGGGTGTACGAATGGGTCAACGCCGGCAAGCTGAAAATCAAGAAGGGCAGCTCCAAGCAG
AGCAGCAGCGAGCTGGTGGATAGCGACATCCTGAAAGACAGCTTCGACCTGGCCTCCGAG
CTGAAAGGCGAAAAGCTGATGCTGTACAGGGACCCCAGCGGCAATGTGTTCCCCAGCGAC
AAATGGATGGCCGCTGGCGTGTTCTTCGGAAAGCTGGAACGCATCCTGATCAGCAAGCTG
ACCAACCAGTACTCCATCAGCACCATCGAGGACGACAGCAGCAAGCAGTCTATGAAAAGG
CCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGGATCC*TACCCATACGAT*
GTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCC
*GACTATGCC*TAA

[0510] MAPKKKRKVGIHGVPAAATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYEHHEQD
PKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVEKKGEANQLSNKFLYPLVDPNSQ
SGKGTASSGRKPRWYNLKIAGDPGGSGGSSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAI
GLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGM
NHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSWEEEKKKWEEDKKKD
PLAKILGKLAEYGLIPLFIPYTDSNEPIVKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWESWNLKVKEEYEKVEK
EYKTLEERIKEDIQALKALEQYEKERQEQLLRDTLNTNEYRLSKRGLRGWREIIQKWLKMDENEPSEKYLEVFKDYQ
RKHPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYATFCEIDKKKKDAKQQATFTLADPINHPLWVRFEERSGSNLN
KYRILTEQLHTEKLKKKLTVQLDRLIYPTESGGWEEKGKVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPL
KGTLGGARVQFDRDHLRRYPHKVESGNVGRIYFNMTVNIEPTESPVSKSLKIHRDDFPKVVNFKPKELTEWIKDSKG
KKLKSGIESLEIGLRVMSIDLGQRQAAAASIFEVVDQKPDIEGKLFFPIKGTELYAVHRASFNIKLPGETLVKSREV
LRKAREDNLKLMNQKLNFLRNVLHFQQFEDITEREKRVTKWISRQENSDVPLVYQDELIQIRELMYKPYKDWVAFLK
QLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDEIDRTRKFLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHLN
ALKEDRLKKMANTIIMHALGYCYDVRKKKWQAKNPACQIILFEDLSNYNPYEERSRFENSKLMKWSRREIPRQVALQ
GEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKLQDNRFFKNLQREGRLTLDKIAVLKEGDLYPDKGGEKFIS
LSKDRKCVTTHADINAAQNLQKRFWTRTHGFYKVYCKAYQVDGQTVYIPESKDQKQKIIEEFGEGYFILKDGVYEWV
NAGKLKIKKGSSKQSSSELVDSDILKDSFDLASELKGEKLMLYRDPSGNVFPSDKWMAAGVFFGKLERILISKLTNQ

YSISTIEDDSSKQSMKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA

[0511] BhCas12b GGSGGS-ABE8-Xten20在K255

GCCACCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGCC
ACCAGATCCTTCATCCTGAAGATCGAGCCCAACGAGGAAGTGAAGAAAGGCCTCTGGAAA
ACCCACGAGGTGCTGAACCACGGAATCGCCTACTACATGAATATCCTGAAGCTGATCCGG
CAAGAGGCCATCTACGAGCACCACGAGCAGGACCCCAAGAATCCCAAGAAGGTGTCCAAG
GCCGAGATCCAGGCCGAGCTGTGGGATTTCGTGCTGAAGATGCAGAAGTGCAACAGCTTC
ACACACGAGGTGGACAAGGACGAGGTGTTCAACATCCTGAGAGAGCTGTACGAGGAACTG
GTGCCCAGCAGCGTGGAAAAGAAGGGCGAAGCCAACCAGCTGAGCAACAAGTTTCTGTAC
CCTCTGGTGGACCCCAACAGCCAGTCTGGAAAGGGAACAGCCAGCAGCGGCAGAAAGCCC
AGATGGTACAACCTGAAGATTGCCGGCGATCCCTCCTGGGAAGAAGAGAAGAAGAAGTGG

[0512] GAAGAAGATAAGAAAAAGGACCCGCTGGCCAAGATCCTGGGCAAGCTGGCTGAGTACGGA
CTGATCCCTCTGTTCATCCCCTACACCGACAGCAACGAGCCCATCGTGAAAGAAATCAAG
TGGATGGAAAAGTCCCGGAACCAGAGCGTGCGGCGGCTGGATAAGGACATGTTCATTCAG
GCCCTGGAACGGTTCCTGAGCTGGGAGAGCTGGAACCTGAAAGTGAAAGAGGAATACGAG
AAGGTCGAGAAAGAGTACAAGACCCTGGAAGAGAGGATCAAAggaggctctggaggaagc
TCCGAAGTCGAGTTTTCCCATGAGTACTGGATGAGACACGCATTGACTCTCGCAAAGAGG
GCTCGAGATGAACGCGAGGTGCCCGTGGGGGCAGTACTCGTGCTCAACAATCGCGTAATC
GGCGAAGGTTGGAATAGGGCAATCGGACTCCACGACCCCACTGCACATGCGGAAATCATG
GCCCTTCGACAGGGAGGGCTTGTGATGCAGAATTATCGACTTTATGATGCGACGCTGTAC
GTCACGTTTGAACCTTGCGTAATGTGCGCGGGAGCTATGATTCACTCCCGCATTGGACGA

GTTGTATTCGGTGTTCGCAACGCCAAGACGGGTGCCGCAGGTTCACTGATGGACGTGCTG

CATCATCCAGGCATGAACCACCGGGTAGAAATCACAGAAGGCATATTGGCGGACGAATGT

GCGGCGCTGTTGTGTCGTTTTTTTCGCATGCCCAGGCGGGTCTTTAACGCCCAGAAAAAA

GCACAATCCTCTACTGACGGCTCTTCTGGATCTGAAACACCTGGCACAAGCGAGAGCGCC

ACCCCTGAGAGCTCTGGCGAGGACATCCAGGCTCTGAAGGCTCTGGAACAGTATGAGAAA

GAGCGGCAAGAACAGCTGCTGCGGGACACCCTGAACACCAACGAGTACCGGCTGAGCAAG

AGAGGCCTTAGAGGCTGGCGGGAAATCATCCAGAAATGGCTGAAAATGGACGAGAACGAG

CCCTCCGAGAAGTACCTGGAAGTGTTCAAGGACTACCAGCGGAAGCACCCTAGAGAGGCC

GGCGATTACAGCGTGTACGAGTTCCTGTCCAAGAAAGAGAACCACTTCATCTGGCGGAAT

CACCCTGAGTACCCCTACCTGTACGCCACCTTCTGCGAGATCGACAAGAAAAAGAAGGAC

GCCAAGCAGCAGGCCACCTTCACACTGGCCGATCCTATCAATCACCCTCTGTGGGTCCGA

TTCGAGGAAAGAAGCGGCAGCAACCTGAACAAGTACAGAATCCTGACCGAGCAGCTGCAC

ACCGAGAAGCTGAAGAAAAAGCTGACAGTGCAGCTGGACCGGCTGATCTACCCTACAGAA

TCTGGCGGCTGGGAAGAGAAGGGCAAAGTGGACATTGTGCTGCTGCCCAGCCGGCAGTTC

TACAACCAGATCTTCCTGGACATCGAGGAAAAGGGCAAGCACGCCTTCACCTACAAGGAT

[0513] GAGAGCATCAAGTTCCCTCTGAAGGGCACACTCGGCGGAGCCAGAGTGCAGTTCGACAGA

GATCACCTGAGAAGATACCCTCACAAGGTGGAAAGCGGCAACGTGGGCAGAATCTACTTC

AACATGACCGTGAACATCGAGCCTACAGAGTCCCCAGTGTCCAAGTCTCTGAAGATCCAC

CGGGACGACTTCCCCAAGGTGGTCAACTTCAAGCCCAAAGAACTGACCGAGTGGATCAAG

GACAGCAAGGGCAAGAAACTGAAGTCCGGCATCGAGTCCCTGGAAATCGGCCTGAGAGTG

ATGAGCATCGACCTGGGACAGAGACAGGCCGCTGCCGCCTCTATTTTCGAGGTGGTGGAT

CAGAAGCCCGACATCGAAGGCAAGCTGTTTTTCCCAATCAAGGGCACCGAGCTGTATGCC

GTGCACAGAGCCAGCTTCAACATCAAGCTGCCCGGCGAGACACTGGTCAAGAGCAGAGAA

GTGCTGCGGAAGGCCAGAGAGGACAATCTGAAACTGATGAACCAGAAGCTCAACTTCCTG

CGGAACGTGCTGCACTTCCAGCAGTTCGAGGACATCACCGAGAGAGAGAAGCGGGTCACC

AAGTGGATCAGCAGACAAGAGAACAGCGACGTGCCCCTGGTGTACCAGGATGAGCTGATC

CAGATCCGCGAGCTGATGTACAAGCCTTACAAGGACTGGGTCGCCTTCCTGAAGCAGCTC

CACAAGAGACTGGAAGTCGAGATCGGCAAAGAAGTGAAGCACTGGCGGAAGTCCCTGAGC

GACGGAAGAAAGGGCCTGTACGGCATCTCCCTGAAGAACATCGACGAGATCGATCGGACC

CGGAAGTTCCTGCTGAGATGGTCCCTGAGGCCTACCGAACCTGGCGAAGTGCGTAGACTG

GAACCCGGCCAGAGATTCGCCATCGACCAGCTGAATCACCTGAACGCCCTGAAAGAAGAT

CGGCTGAAGAAGATGGCCAACACCATCATCATGCACGCCCTGGGCTACTGCTACGACGTG
CGGAAGAAGAAATGGCAGGCTAAGAACCCCGCCTGCCAGATCATCCTGTTCGAGGATCTG
AGCAACTACAACCCCTACGAGGAAAGGTCCCGCTTCGAGAACAGCAAGCTCATGAAGTGG
TCCAGACGCGAGATCCCCAGACAGGTTGCACTGCAGGGCGAGATCTATGGCCTGCAAGTG
GGAGAAGTGGGCGCTCAGTTCAGCAGCAGATTCCACGCCAAGACAGGCAGCCCTGGCATC
AGATGTAGCGTCGTGACCAAAGAGAAGCTGCAGGACAATCGGTTCTTCAAGAATCTGCAG
AGAGAGGGCAGACTGACCCTGGACAAAATCGCCGTGCTGAAAGAGGGCGATCTGTACCCA
GACAAAGGCGGCGAGAAGTTCATCAGCCTGAGCAAGGATCGGAAGTGCGTGACCACACAC
GCCGACATCAACGCCGCTCAGAACCTGCAGAAGCGGTTCTGGACAAGAACCCACGGCTTC

[0514] TACAAGGTGTACTGCAAGGCCTACCAGGTGGACGGCCAGACCGTGTACATCCCTGAGAGC
AAGGACCAGAAGCAGAAGATCATCGAAGAGTTCGGCGAGGGCTACTTCATTCTGAAGGAC
GGGGTGTACGAATGGGTCAACGCCGGCAAGCTGAAAATCAAGAAGGGCAGCTCCAAGCAG
AGCAGCAGCGAGCTGGTGGATAGCGACATCCTGAAAGACAGCTTCGACCTGGCCTCCGAG
CTGAAAGGCGAAAAGCTGATGCTGTACAGGGACCCCAGCGGCAATGTGTTCCCCAGCGAC
AAATGGATGGCCGCTGGCGTGTTCTTCGGAAAGCTGGAACGCATCCTGATCAGCAAGCTG
ACCAACCAGTACTCCATCAGCACCATCGAGGACGACAGCAGCAAGCAGTCTATGAAAAGG
CCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGGATCC*TACCCATACGAT*
GTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCC
*GACTATGCC*TAA

[0515] MAPKKKRKVGIHGVPAAATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYEHHEQD
PKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVEKKGEANQLSNKFLYPLVDPNSQ
SGKGTASSGRKPRWYNLKIAGDPSWEEEKKKWEEDKKKDPLAKILGKLAEYGLIPLFIPYTDSNEPIVKEIKWMEKS
RNQSVRRLDKDMFIQALERFLSWESWNLKVKEEYEKVEKEYKTLEERIKGGSGGSSEVEFSHEYWMRHALTLAKRAR
DEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIG
RVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGT
SESATPESSGEDIQALKALEQYEKERQEQLLRDTLNTNEYRLSKRGLRGWREIIQKWLKMDENEPSEKYLEVFKDYQ
RKHPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYATFCEIDKKKKDAKQQATFTLADPINHPLWVRFEERSGSNLN
KYRILTEQLHTEKLKKKLTVQLDRLIYPTESGGWEEKGKVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPL
KGTLGGARVQFDRDHLRRYPHKVESGNVGRIYFNMTVNIEPTESPVSKSLKIHRDDFPKVVNFKPKELTEWIKDSKG
KKLKSGIESLEIGLRVMSIDLGQRQAAAASIFEVVDQKPDIEGKLFFPIKGTELYAVHRASFNIKLPGETLVKSREV
LRKAREDNLKLMNQKLNFLRNVLHFQQFEDITEREKRVTKWISRQENSDVPLVYQDELIQIRELMYKPYKDWVAFLK
QLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDEIDRTRKFLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHLN
ALKEDRLKKMANTIIMHALGYCYDVRKKKWQAKNPACQIILFEDLSNYNPYEERSRFENSKLMKWSRREIPRQVALQ
GEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKLQDNRFFKNLQREGRLTLDKIAVLKEGDLYPDKGGEKFIS

LSKDRKCVTTHADINAAQNLQKRFWTRTHGFYKVYCKAYQVDGQTVYIPESKDQKQKIIEEFGEGYFILKDGVYEWV
NAGKLKIKKGSSKQSSSELVDSDILKDSFDLASELKGEKLMLYRDPSGNVFPSDKWMAAGVFFGKLERILISKLTNQ
YSISTIEDDSSKQSMKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA

[0516] BhCas12b GGSGGS-ABE8-Xten20在D306

GCCACCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGCC
ACCAGATCCTTCATCCTGAAGATCGAGCCCAACGAGGAAGTGAAGAAAGGCCTCTGGAAA
ACCCACGAGGTGCTGAACCACGGAATCGCCTACTACATGAATATCCTGAAGCTGATCCGG
CAAGAGGCCATCTACGAGCACCACGAGCAGGACCCCAAGAATCCCAAGAAGGTGTCCAAG
GCCGAGATCCAGGCCGAGCTGTGGGATTTCGTGCTGAAGATGCAGAAGTGCAACAGCTTC
ACACACGAGGTGGACAAGGACGAGGTGTTCAACATCCTGAGAGAGCTGTACGAGGAACTG
GTGCCCAGCAGCGTGGAAAAGAAGGGCGAAGCCAACCAGCTGAGCAACAAGTTTCTGTAC
CCTCTGGTGGACCCCAACAGCCAGTCTGGAAAGGGAACAGCCAGCAGCGGCAGAAAGCCC
AGATGGTACAACCTGAAGATTGCCGGCGATCCCTCCTGGGAAGAAGAGAAGAAGAAGTGG
[0517] GAAGAAGATAAGAAAAAGGACCCGCTGGCCAAGATCCTGGGCAAGCTGGCTGAGTACGGA
CTGATCCCTCTGTTCATCCCCTACACCGACAGCAACGAGCCCATCGTGAAAGAAATCAAG
TGGATGGAAAAGTCCCGGAACCAGAGCGTGCGGCGGCTGGATAAGGACATGTTCATTCAG
GCCCTGGAACGGTTCCTGAGCTGGGAGAGCTGGAACCTGAAAGTGAAAGAGGAATACGAG
AAGGTCGAGAAAGAGTACAAGACCCTGGAAGAGAGGATCAAAGAGGACATCCAGGCTCTG
AAGGCTCTGGAACAGTATGAGAAAGAGCGGCAAGAACAGCTGCTGCGGGACACCCTGAAC
ACCAACGAGTACCGGCTGAGCAAGAGAGGCCTTAGAGGCTGGCGGGAAATCATCCAGAAA
TGGCTGAAAATGGACggaggctctggaggaagcTCCGAAGTCGAGTTTTCCCATGAGTAC
TGGATGAGACACGCATTGACTCTCGCAAAGAGGGCTCGAGATGAACGCGAGGTGCCCGTG

GGGGCAGTACTCGTGCTCAACAATCGCGTAATCGGCGAAGGTTGGAATAGGGCAATCGGA
CTCCACGACCCCACTGCACATGCGGAAATCATGGCCCTTCGACAGGGAGGGCTTGTGATG
CAGAATTATCGACTTTATGATGCGACGCTGTACGTCACGTTTGAACCTTGCGTAATGTGC
GCGGGAGCTATGATTCACTCCCGCATTGGACGAGTTGTATTCGGTGTTCGCAACGCCAAG
ACGGGTGCCGCAGGTTCACTGATGGACGTGCTGCATCATCCAGGCATGAACCACCGGGTA
GAAATCACAGAAGGCATATTGGCGGACGAATGTGCGGCGCTGTTGTGTCGTTTTTTCGC
ATGCCCAGGCGGGTCTTTAACGCCCAGAAAAAAGCACAATCCTCTACTGACGGCTCTTCT
GGATCTGAAACACCTGGCACAAGCGAGAGCGCCACCCCTGAGAGCTCTGGCGAGAACGAG
CCCTCCGAGAAGTACCTGGAAGTGTTCAAGGACTACCAGCGGAAGCACCCTAGAGAGGCC
GGCGATTACAGCGTGTACGAGTTCCTGTCCAAGAAAGAGAACCACTTCATCTGGCGGAAT
CACCCTGAGTACCCCTACCTGTACGCCACCTTCTGCGAGATCGACAAGAAAAAGAAGGAC
GCCAAGCAGCAGGCCACCTTCACACTGGCCGATCCTATCAATCACCCTCTGTGGGTCCGA
TTCGAGGAAAGAAGCGGCAGCAACCTGAACAAGTACAGAATCCTGACCGAGCAGCTGCAC
ACCGAGAAGCTGAAGAAAAAGCTGACAGTGCAGCTGGACCGGCTGATCTACCCTACAGAA
TCTGGCGGCTGGGAAGAGAAGGGCAAAGTGGACATTGTGCTGCTGCCCAGCCGGCAGTTC
[0518] TACAACCAGATCTTCCTGGACATCGAGGAAAAGGGCAAGCACGCCTTCACCTACAAGGAT
GAGAGCATCAAGTTCCCTCTGAAGGGCACACTCGGCGGAGCCAGAGTGCAGTTCGACAGA
GATCACCTGAGAAGATACCCTCACAAGGTGGAAAGCGGCAACGTGGGCAGAATCTACTTC
AACATGACCGTGAACATCGAGCCTACAGAGTCCCCAGTGTCCAAGTCTCTGAAGATCCAC
CGGGACGACTTCCCCAAGGTGGTCAACTTCAAGCCCAAAGAACTGACCGAGTGGATCAAG
GACAGCAAGGGCAAGAAACTGAAGTCCGGCATCGAGTCCCTGGAAATCGGCCTGAGAGTG
ATGAGCATCGACCTGGGACAGAGACAGGCCGCTGCCGCCTCTATTTTCGAGGTGGTGGAT
CAGAAGCCCGACATCGAAGGCAAGCTGTTTTTCCCAATCAAGGGCACCGAGCTGTATGCC
GTGCACAGAGCCAGCTTCAACATCAAGCTGCCCGGCGAGACACTGGTCAAGAGCAGAGAA
GTGCTGCGGAAGGCCAGAGAGGACAATCTGAAACTGATGAACCAGAAGCTCAACTTCCTG
CGGAACGTGCTGCACTTCCAGCAGTTCGAGGACATCACCGAGAGAGAGAAGCGGGTCACC
AAGTGGATCAGCAGACAAGAGAACAGCGACGTGCCCCTGGTGTACCAGGATGAGCTGATC
CAGATCCGCGAGCTGATGTACAAGCCTTACAAGGACTGGGTCGCCTTCCTGAAGCAGCTC
CACAAGAGACTGGAAGTCGAGATCGGCAAAGAAGTGAAGCACTGGCGGAAGTCCCTGAGC
GACGGAAGAAAGGGCCTGTACGGCATCTCCCTGAAGAACATCGACGAGATCGATCGGACC
CGGAAGTTCCTGCTGAGATGGTCCCTGAGGCCTACCGAACCTGGCGAAGTGCGTAGACTG

[0519]

GAACCCGGCCAGAGATTCGCCATCGACCAGCTGAATCACCTGAACGCCCTGAAAGAAGAT
CGGCTGAAGAAGATGGCCAACACCATCATCATGCACGCCCTGGGCTACTGCTACGACGTG
CGGAAGAAGAAATGGCAGGCTAAGAACCCCGCCTGCCAGATCATCCTGTTCGAGGATCTG
AGCAACTACAACCCCTACGAGGAAAGGTCCCGCTTCGAGAACAGCAAGCTCATGAAGTGG
TCCAGACGCGAGATCCCCAGACAGGTTGCACTGCAGGGCGAGATCTATGGCCTGCAAGTG
GGAGAAGTGGGCGCTCAGTTCAGCAGCAGATTCCACGCCAAGACAGGCAGCCCTGGCATC
AGATGTAGCGTCGTGACCAAAGAGAAGCTGCAGGACAATCGGTTCTTCAAGAATCTGCAG
AGAGAGGGCAGACTGACCCTGGACAAAATCGCCGTGCTGAAAGAGGGCGATCTGTACCCA
GACAAAGGCGGCGAGAAGTTCATCAGCCTGAGCAAGGATCGGAAGTGCGTGACCACACAC
GCCGACATCAACGCCGCTCAGAACCTGCAGAAGCGGTTCTGGACAAGAACCCACGGCTTC
TACAAGGTGTACTGCAAGGCCTACCAGGTGGACGGCCAGACCGTGTACATCCCTGAGAGC
AAGGACCAGAAGCAGAAGATCATCGAAGAGTTCGGCGAGGGCTACTTCATTCTGAAGGAC
GGGGTGTACGAATGGGTCAACGCCGGCAAGCTGAAAATCAAGAAGGGCAGCTCCAAGCAG
AGCAGCAGCGAGCTGGTGGATAGCGACATCCTGAAAGACAGCTTCGACCTGGCCTCCGAG
CTGAAAGGCGAAAAGCTGATGCTGTACAGGGACCCCAGCGGCAATGTGTTCCCCAGCGAC
AAATGGATGGCCGCTGGCGTGTTCTTCGGAAAGCTGGAACGCATCCTGATCAGCAAGCTG
ACCAACCAGTACTCCATCAGCACCATCGAGGACGACAGCAGCAAGCAGTCTATGAAAAGG
CCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGGATCC*TACCCATACGAT*
GTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCC
*GACTATGCC*TAA

[0520] MAPKKKRKVGIHGVPAAATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYEHHEQD
PKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVEKKGEANQLSNKFLYPLVDPNSQ
SGKGTASSGRKPRWYNLKIAGDPSWEEEKKKWEEDKKKDPLAKILGKLAEYGLIPLFIPYTDSNEPIVKEIKWMEKS
RNQSVRRLDKDMFIQALERFLSWESWNLKVKEEYEKVEKEYKTLEERIKEDIQALKALEQYEKERQEQLLRDTLNTN
EYRLSKRGLRGWREIIQKWLKMDGGSGGSSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAI
GLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGM
NHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESATPESSGENEPSEKYLEVFKDYQ
RKHPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYATFCEIDKKKKDAKQQATFTLADPINHPLWVRFEERSGSNLN
KYRILTEQLHTEKLKKKLTVQLDRLIYPTESGGWEEKGKVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPL
KGTLGGARVQFDRDHLRRYPHKVESGNVGRIYFNMTVNIEPTESPVSKSLKIHRDDFPKVVNFKPKELTEWIKDSKG
KKLKSGIESLEIGLRVMSIDLGQRQAAAASIFEVVDQKPDIEGKLFFPIKGTELYAVHRASFNIKLPGETLVKSREV
LRKAREDNLKLMNQKLNFLRNVLHFQQFEDITEREKRVTKWISRQENSDVPLVYQDELIQIRELMYKPYKDWVAFLK
QLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDEIDRTRKFLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHLN
ALKEDRLKKMANTIIMHALGYCYDVRKKKWQAKNPACQIILFEDLSNYNPYEERSRFENSKLMKWSRREIPRQVALQ

GEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKLQDNRFFKNLQREGRLTLDKIAVLKEGDLYPDKGGEKFIS
LSKDRKCVTTHADINAAQNLQKRFWTRTHGFYKVYCKAYQVDGQTVYIPESKDQKQKIIEEFGEGYFILKDGVYEWV
NAGKLKIKKGSSKQSSSELVDSDILKDSFDLASELKGEKLMLYRDPSGNVFPSDKWMAAGVFFGKLERILISKLTNQ
YSISTIEDDSSKQSMKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA

[0521] BhCas12b GGSGGS-ABE8-Xten20在D980

[0522]

GCCACCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGCC
ACCAGATCCTTCATCCTGAAGATCGAGCCCAACGAGGAAGTGAAGAAAGGCCTCTGGAAA
ACCCACGAGGTGCTGAACCACGGAATCGCCTACTACATGAATATCCTGAAGCTGATCCGG
CAAGAGGCCATCTACGAGCACCACGAGCAGGACCCCAAGAATCCCAAGAAGGTGTCCAAG
GCCGAGATCCAGGCCGAGCTGTGGGATTTCGTGCTGAAGATGCAGAAGTGCAACAGCTTC
ACACACGAGGTGGACAAGGACGAGGTGTTCAACATCCTGAGAGAGCTGTACGAGGAACTG
GTGCCCAGCAGCGTGGAAAAGAAGGGCGAAGCCAACCAGCTGAGCAACAAGTTTCTGTAC
CCTCTGGTGGACCCCAACAGCCAGTCTGGAAAGGGAACAGCCAGCAGCGGCAGAAAGCCC
AGATGGTACAACCTGAAGATTGCCGGCGATCCCTCCTGGGAAGAAGAGAAGAAGAAGTGG
GAAGAAGATAAGAAAAAGGACCCGCTGGCCAAGATCCTGGGCAAGCTGGCTGAGTACGGA
CTGATCCCTCTGTTCATCCCCTACACCGACAGCAACGAGCCCATCGTGAAAGAAATCAAG
TGGATGGAAAAGTCCCGGAACCAGAGCGTGCGGCGGCTGGATAAGGACATGTTCATTCAG
GCCCTGGAACGGTTCCTGAGCTGGGAGAGCTGGAACCTGAAAGTGAAAGAGGAATACGAG
AAGGTCGAGAAAGAGTACAAGACCCTGGAAGAGAGGATCAAAGAGGACATCCAGGCTCTG
AAGGCTCTGGAACAGTATGAGAAAGAGCGGCAAGAACAGCTGCTGCGGGACACCCTGAAC
ACCAACGAGTACCGGCTGAGCAAGAGAGGCCTTAGAGGCTGGCGGGAAATCATCCAGAAA

TGGCTGAAAATGGACGAGAACGAGCCCTCCGAGAAGTACCTGGAAGTGTTCAAGGACTAC

CAGCGGAAGCACCCTAGAGAGGCCGGCGATTACAGCGTGTACGAGTTCCTGTCCAAGAAA

GAGAACCACTTCATCTGGCGGAATCACCCTGAGTACCCCTACCTGTACGCCACCTTCTGC

GAGATCGACAAGAAAAAGAAGGACGCCAAGCAGCAGGCCACCTTCACACTGGCCGATCCT

ATCAATCACCCTCTGTGGGTCCGATTCGAGGAAAGAAGCGGCAGCAACCTGAACAAGTAC

AGAATCCTGACCGAGCAGCTGCACACCGAGAAGCTGAAGAAAAAGCTGACAGTGCAGCTG

GACCGGCTGATCTACCCTACAGAATCTGGCGGCTGGGAAGAGAAGGGCAAAGTGGACATT

GTGCTGCTGCCCAGCCGGCAGTTCTACAACCAGATCTTCCTGGACATCGAGGAAAAGGGC

AAGCACGCCTTCACCTACAAGGATGAGAGCATCAAGTTCCCTCTGAAGGGCACACTCGGC

GGAGCCAGAGTGCAGTTCGACAGAGATCACCTGAGAAGATACCCTCACAAGGTGGAAAGC

GGCAACGTGGGCAGAATCTACTTCAACATGACCGTGAACATCGAGCCTACAGAGTCCCCA

GTGTCCAAGTCTCTGAAGATCCACCGGGACGACTTCCCCAAGGTGGTCAACTTCAAGCCC

AAAGAACTGACCGAGTGGATCAAGGACAGCAAGGGCAAGAAACTGAAGTCCGGCATCGAG

TCCCTGGAAATCGGCCTGAGAGTGATGAGCATCGACCTGGGACAGAGACAGGCCGCTGCC

GCCTCTATTTTCGAGGTGGTGGATCAGAAGCCCGACATCGAAGGCAAGCTGTTTTTCCCA

[0523] ATCAAGGGCACCGAGCTGTATGCCGTGCACAGAGCCAGCTTCAACATCAAGCTGCCCGGC

GAGACACTGGTCAAGAGCAGAGAAGTGCTGCGGAAGGCCAGAGAGGACAATCTGAAACTG

ATGAACCAGAAGCTCAACTTCCTGCGGAACGTGCTGCACTTCCAGCAGTTCGAGGACATC

ACCGAGAGAGAGAAGCGGGTCACCAAGTGGATCAGCAGACAAGAGAACAGCGACGTGCCC

CTGGTGTACCAGGATGAGCTGATCCAGATCCGCGAGCTGATGTACAAGCCTTACAAGGAC

TGGGTCGCCTTCCTGAAGCAGCTCCACAAGAGACTGGAAGTCGAGATCGGCAAAGAAGTG

AAGCACTGGCGGAAGTCCCTGAGCGACGGAAGAAAGGGCCTGTACGGCATCTCCCTGAAG

AACATCGACGAGATCGATCGGACCCGGAAGTTCCTGCTGAGATGGTCCCTGAGGCCTACC

GAACCTGGCGAAGTGCGTAGACTGGAACCCGGCCAGAGATTCGCCATCGACCAGCTGAAT

CACCTGAACGCCCTGAAAGAAGATCGGCTGAAGAAGATGGCCAACACCATCATCATGCAC

GCCCTGGGCTACTGCTACGACGTGCGGAAGAAGAAATGGCAGGCTAAGAACCCCGCCTGC

CAGATCATCCTGTTCGAGGATCTGAGCAACTACAACCCCTACGAGGAAAGGTCCCGCTTC

GAGAACAGCAAGCTCATGAAGTGGTCCAGACGCGAGATCCCCAGACAGGTTGCACTGCAG

GGCGAGATCTATGGCCTGCAAGTGGGAGAAGTGGGCGCTCAGTTCAGCAGCAGATTCCAC

GCCAAGACAGGCAGCCCTGGCATCAGATGTAGCGTCGTGACCAAAGAGAAGCTGCAGGAC

AATCGGTTCTTCAAGAATCTGCAGAGAGAGGGCAGACTGACCCTGGACAAAATCGCCGTG

[0524]

CTGAAAGAGGGCGATCTGTACCCAGACAAAGGCGGCGAGAAGTTCATCAGCCTGAGCAAG
GATCGGAAGTGCGTGACCACACACGCCGACATCAACGCCGCTCAGAACCTGCAGAAGCGG
TTCTGGACAAGAACCCACGGCTTCTACAAGGTGTACTGCAAGGCCTACCAGGTGGACgga
ggctctggaggaagcTCCGAAGTCGAGTTTTCCCATGAGTACTGGATGAGACACGCATTG
ACTCTCGCAAAGAGGGCTCGAGATGAACGCGAGGTGCCCGTGGGGGCAGTACTCGTGCTC
AACAATCGCGTAATCGGCGAAGGTTGGAATAGGGCAATCGGACTCCACGACCCCACTGCA
CATGCGGAAATCATGGCCCTTCGACAGGGAGGGCTTGTGATGCAGAATTATCGACTTTAT
GATGCGACGCTGTACGTCACGTTTGAACCTTGCGTAATGTGCGCGGGAGCTATGATTCAC
TCCCGCATTGGACGAGTTGTATTCGGTGTTCGCAACGCCAAGACGGGTGCCGCAGGTTCA
CTGATGGACGTGCTGCATCATCCAGGCATGAACCACCGGGTAGAAATCACAGAAGGCATA
TTGGCGGACGAATGTGCGGCGCTGTTGTGTCGTTTTTTTCGCATGCCCAGGCGGGTCTTT
AACGCCCAGAAAAAAGCACAATCCTCTACTGACGGCTCTTCTGGATCTGAAACACCTGGC
ACAAGCGAGAGCGCCACCCCTGAGAGCTCTGGCGGCCAGACCGTGTACATCCCTGAGAGC
AAGGACCAGAAGCAGAAGATCATCGAAGAGTTCGGCGAGGGCTACTTCATTCTGAAGGAC
GGGGTGTACGAATGGGTCAACGCCGGCAAGCTGAAAATCAAGAAGGGCAGCTCCAAGCAG
AGCAGCAGCGAGCTGGTGGATAGCGACATCCTGAAAGACAGCTTCGACCTGGCCTCCGAG
CTGAAAGGCGAAAAGCTGATGCTGTACAGGGACCCCAGCGGCAATGTGTTCCCCAGCGAC
AAATGGATGGCCGCTGGCGTGTTCTTCGGAAAGCTGGAACGCATCCTGATCAGCAAGCTG
ACCAACCAGTACTCCATCAGCACCATCGAGGACGACAGCAGCAAGCAGTCTATGAAAAGG
CCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGGATCC*TACCCATACGAT*
GTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCC
*GACTATGCC*TAA

[0525] MAPKKKRKVGIHGVPAAATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYEHHEQD
PKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVEKKGEANQLSNKFLYPLVDPNSQ
SGKGTASSGRKPRWYNLKIAGDPSWEEEKKKWEEDKKKDPLAKILGKLAEYGLIPLFIPYTDSNEPIVKEIKWMEKS
RNQSVRRLDKDMFIQALERFLSWESWNLKVKEEYEKVEKEYKTLEERIKEDIQALKALEQYEKERQEQLLRDTLNTN
EYRLSKRGLRGWREIIQKWLKMDENEPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYATF
CEIDKKKKDAKQQATFTLADPINHPLWVRFEERSGSNLNKYRILTEQLHTEKLKKKLTVQLDRLIYPTESGGWEEKG
KVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGTLGGARVQFDRDHLRRYPHKVESGNVGRIYFNMTVN
IEPTESPVSKSLKIHRDDFPKVVNFKPKELTEWIKDSKGKKLKSGIESLEIGLRVMSIDLGQRQAAAASIFEVVDQK
PDIEGKLFFPIKGTELYAVHRASFNIKLPGETLVKSREVLRKAREDNLKLMNQKLNFLRNVLHFQQFEDITEREKRV
TKWISRQENSDVPLVYQDELIQIRELMYKPYKDWVAFLKQLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDE
IDRTRKFLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHLNALKEDRLKKMANTIIMHALGYCYDVRKKKWQAKNPACQ
IILFEDLSNYNPYEERSRFENSKLMKWSRREIPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKL

QDNRFFKNLQREGRLTLDKIAVLKEGDLYPDKGGEKFISLSKDRKCVTTHADINAAQNLQKRFWTRTHGFYKVYCKA
YQVDGGSGGSSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGL
VMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALL
CRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESATPESSGGQTVYIPESKDQKQKIIEEFGEGYFILKDGVYEWV
NAGKLKIKKGSSKQSSSELVDSDILKDSFDLASELKGEKLMLYRDPSGNVFPSDKWMAAGVFFGKLERILISKLTNQ
YSISTIEDDSSKQSMKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA

[0526] BhCas12b GGSGGS-ABE8-Xten20在K1019

GCCACCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGCC
ACCAGATCCTTCATCCTGAAGATCGAGCCCAACGAGGAAGTGAAGAAAGGCCTCTGGAAA
ACCCACGAGGTGCTGAACCACGGAATCGCCTACTACATGAATATCCTGAAGCTGATCCGG
CAAGAGGCCATCTACGAGCACCACGAGCAGGACCCCAAGAATCCCAAGAAGGTGTCCAAG
GCCGAGATCCAGGCCGAGCTGTGGGATTTCGTGCTGAAGATGCAGAAGTGCAACAGCTTC
ACACACGAGGTGGACAAGGACGAGGTGTTCAACATCCTGAGAGAGCTGTACGAGGAACTG
GTGCCCAGCAGCGTGGAAAAGAAGGGCGAAGCCAACCAGCTGAGCAACAAGTTTCTGTAC
[0527] CCTCTGGTGGACCCCAACAGCCAGTCTGGAAAGGGAACAGCCAGCAGCGGCAGAAAGCCC
AGATGGTACAACCTGAAGATTGCCGGCGATCCCTCCTGGGAAGAAGAGAAGAAGAAGTGG
GAAGAAGATAAGAAAAAGGACCCGCTGGCCAAGATCCTGGGCAAGCTGGCTGAGTACGGA
CTGATCCCTCTGTTCATCCCCTACACCGACAGCAACGAGCCCATCGTGAAAGAAATCAAG
TGGATGGAAAAGTCCCGGAACCAGAGCGTGCGGCGGCTGGATAAGGACATGTTCATTCAG
GCCCTGGAACGGTTCCTGAGCTGGGAGAGCTGGAACCTGAAAGTGAAAGAGGAATACGAG
AAGGTCGAGAAAGAGTACAAGACCCTGGAAGAGAGGATCAAAGAGGACATCCAGGCTCTG

AAGGCTCTGGAACAGTATGAGAAAGAGCGGCAAGAACAGCTGCTGCGGGACACCCTGAAC

ACCAACGAGTACCGGCTGAGCAAGAGAGGCCTTAGAGGCTGGCGGGAAATCATCCAGAAA

TGGCTGAAAATGGACGAGAACGAGCCCTCCGAGAAGTACCTGGAAGTGTTCAAGGACTAC

CAGCGGAAGCACCCTAGAGAGGCCGGCGATTACAGCGTGTACGAGTTCCTGTCCAAGAAA

GAGAACCACTTCATCTGGCGGAATCACCCTGAGTACCCCTACCTGTACGCCACCTTCTGC

GAGATCGACAAGAAAAAGAAGGACGCCAAGCAGCAGGCCACCTTCACACTGGCCGATCCT

ATCAATCACCCTCTGTGGGTCCGATTCGAGGAAAGAAGCGGCAGCAACCTGAACAAGTAC

AGAATCCTGACCGAGCAGCTGCACACCGAGAAGCTGAAGAAAAAGCTGACAGTGCAGCTG

GACCGGCTGATCTACCCTACAGAATCTGGCGGCTGGGAAGAGAAGGGCAAAGTGGACATT

GTGCTGCTGCCCAGCCGGCAGTTCTACAACCAGATCTTCCTGGACATCGAGGAAAAGGGC

AAGCACGCCTTCACCTACAAGGATGAGAGCATCAAGTTCCCTCTGAAGGGCACACTCGGC

GGAGCCAGAGTGCAGTTCGACAGAGATCACCTGAGAAGATACCCTCACAAGGTGGAAAGC

GGCAACGTGGGCAGAATCTACTTCAACATGACCGTGAACATCGAGCCTACAGAGTCCCCA

GTGTCCAAGTCTCTGAAGATCCACCGGGACGACTTCCCCAAGGTGGTCAACTTCAAGCCC

AAAGAACTGACCGAGTGGATCAAGGACAGCAAGGGCAAGAAACTGAAGTCCGGCATCGAG

[0528] TCCCTGGAAATCGGCCTGAGAGTGATGAGCATCGACCTGGGACAGAGACAGGCCGCTGCC

GCCTCTATTTTCGAGGTGGTGGATCAGAAGCCCGACATCGAAGGCAAGCTGTTTTTCCCA

ATCAAGGGCACCGAGCTGTATGCCGTGCACAGAGCCAGCTTCAACATCAAGCTGCCCGGC

GAGACACTGGTCAAGAGCAGAGAAGTGCTGCGGAAGGCCAGAGAGGACAATCTGAAACTG

ATGAACCAGAAGCTCAACTTCCTGCGGAACGTGCTGCACTTCCAGCAGTTCGAGGACATC

ACCGAGAGAGAGAAGCGGGTCACCAAGTGGATCAGCAGACAAGAGAACAGCGACGTGCCC

CTGGTGTACCAGGATGAGCTGATCCAGATCCGCGAGCTGATGTACAAGCCTTACAAGGAC

TGGGTCGCCTTCCTGAAGCAGCTCCACAAGAGACTGGAAGTCGAGATCGGCAAAGAAGTG

AAGCACTGGCGGAAGTCCCTGAGCGACGGAAGAAAGGGCCTGTACGGCATCTCCCTGAAG

AACATCGACGAGATCGATCGGACCCGGAAGTTCCTGCTGAGATGGTCCCTGAGGCCTACC

GAACCTGGCGAAGTGCGTAGACTGGAACCCGGCCAGAGATTCGCCATCGACCAGCTGAAT

CACCTGAACGCCCTGAAAGAAGATCGGCTGAAGAAGATGGCCAACACCATCATCATGCAC

GCCCTGGGCTACTGCTACGACGTGCGGAAGAAGAAATGGCAGGCTAAGAACCCCGCCTGC

CAGATCATCCTGTTCGAGGATCTGAGCAACTACAACCCCTACGAGGAAAGGTCCCGCTTC

GAGAACAGCAAGCTCATGAAGTGGTCCAGACGCGAGATCCCCAGACAGGTTGCACTGCAG

GGCGAGATCTATGGCCTGCAAGTGGGAGAAGTGGGCGCTCAGTTCAGCAGCAGATTCCAC

[0529] GCCAAGACAGGCAGCCCTGGCATCAGATGTAGCGTCGTGACCAAAGAGAAGCTGCAGGAC
AATCGGTTCTTCAAGAATCTGCAGAGAGAGGGCAGACTGACCCTGGACAAAATCGCCGTG
CTGAAAGAGGGCGATCTGTACCCAGACAAAGGCGGCGAGAAGTTCATCAGCCTGAGCAAG
GATCGGAAGTGCGTGACCACACACGCCGACATCAACGCCGCTCAGAACCTGCAGAAGCGG
TTCTGGACAAGAACCCACGGCTTCTACAAGGTGTACTGCAAGGCCTACCAGGTGGACGGC
CAGACCGTGTACATCCCTGAGAGCAAGGACCAGAAGCAGAAGATCATCGAAGAGTTCGGC
GAGGGCTACTTCATTCTGAAGGACGGGGTGTACGAATGGGTCAACGCCGGCAAGggaggc
tctggaggaagcTCCGAAGTCGAGTTTTCCCATGAGTACTGGATGAGACACGCATTGACT
CTCGCAAAGAGGGCTCGAGATGAACGCGAGGTGCCCGTGGGGGCAGTACTCGTGCTCAAC
AATCGCGTAATCGGCGAAGGTTGGAATAGGGCAATCGGACTCCACGACCCCACTGCACAT
GCGGAAATCATGGCCCTTCGACAGGGAGGGCTTGTGATGCAGAATTATCGACTTTATGAT
GCGACGCTGTACGTCACGTTTGAACCTTGCGTAATGTGCGCGGGAGCTATGATTCACTCC
CGCATTGGACGAGTTGTATTCGGTGTTCGCAACGCCAAGACGGGTGCCGCAGGTTCACTG
ATGGACGTGCTGCATCATCCAGGCATGAACCACCGGGTAGAAATCACAGAAGGCATATTG
GCGGACGAATGTGCGGCGCTGTTGTGTCGTTTTTTTCGCATGCCCAGGCGGGTCTTTAAC
GCCCAGAAAAAAGCACAATCCTCTACTGACGGCTCTTCTGGATCTGAAACACCTGGCACA
AGCGAGAGCGCCACCCCTGAGAGCTCTGGCCTGAAAATCAAGAAGGGCAGCTCCAAGCAG
AGCAGCAGCGAGCTGGTGGATAGCGACATCCTGAAAGACAGCTTCGACCTGGCCTCCGAG
CTGAAAGGCGAAAAGCTGATGCTGTACAGGGACCCCAGCGGCAATGTGTTCCCCAGCGAC
AAATGGATGGCCGCTGGCGTGTTCTTCGGAAAGCTGGAACGCATCCTGATCAGCAAGCTG
ACCAACCAGTACTCCATCAGCACCATCGAGGACGACAGCAGCAAGCAGTCTATGAAAAGG
CCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGGATCC*TACCCATACGAT*
GTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCC
*GACTATGCC*TAA

[0530] MAPKKKRKVGIHGVPAAATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYEHHEQD
PKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVEKKGEANQLSNKFLYPLVDPNSQ
SGKGTASSGRKPRWYNLKIAGDPSWEEEKKKWEEDKKKDPLAKILGKLAEYGLIPLFIPYTDSNEPIVKEIKWMEKS
RNQSVRRLDKDMFIQALERFLSWESWNLKVKEEYEKVEKEYKTLEERIKEDIQALKALEQYEKERQEQLLRDTLNTN
EYRLSKRGLRGWREIIQKWLKMDENEPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYATF
CEIDKKKKDAKQQATFTLADPINHPLWVRFEERSGSNLNKYRILTEQLHTEKLKKKLTVQLDRLIYPTESGGWEEKG
KVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGTLGGARVQFDRDHLRRYPHKVESGNVGRIYFNMTVN
IEPTESPVSKSLKIHRDDFPKVVNFKPKELTEWIKDSKGKKLKSGIESLEIGLRVMSIDLGQRQAAAASIFEVVDQK
PDIEGKLFFPIKGTELYAVHRASFNIKLPGETLVKSREVLRKAREDNLKLMNQKLNFLRNVLHFQQFEDITEREKRV

TKWISRQENSDVPLVYQDELIQIRELMYKPYKDWVAFLKQLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDE
IDRTRKFLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHLNALKEDRLKKMANTIIMHALGYCYDVRKKKWQAKNPACQ
IILFEDLSNYNPYEERSRFENSKLMKWSRREIPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKL
QDNRFFKNLQREGRLTLDKIAVLKEGDLYPDKGGEKFISLSKDRKCVTTHADINAAQNLQKRFWTRTHGFYKVYCKA
YQVDGQTVYIPESKDQKQKIIEEFGEGYFILKDGVYEWVNAGKGGSGGSSEVEFSHEYWMRHALTLAKRARDEREVP
VGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGV
RNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESATP
ESSGLKIKKGSSKQSSSELVDSDILKDSFDLASELKGEKLMLYRDPSGNVFPSDKWMAAGVFFGKLERILISKLTNQ
YSISTIEDDSSKQSMKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA

[0531] 对于上面的序列,Kozak序列用粗体和下划线表示;标记N-末端核定位信号(NLS);小写字符表示GGGSGGS接头;

[0532] ＝＝＝＝＝标记编码ABE8的序列,未修改的序列编码BhCas12b;双下划线表示Xten20接头;单下划线表示C-末端NLS;GGATCC表示GS接头;斜体字符代表3x血凝素(HA)标签的编码序列。

[0533] 向导多核苷酸

[0534] 在一个实施例中,向导多核苷酸是向导RNA。RNA/Cas复合物可以帮助将Cas蛋白“引导”到目标DNA。Cas9/crRNA/tracrRNA核酸内切切割与间隔区互补的线性或环状dsDNA目标。与crRNA不互补的目标链首先被核酸内切切割,然后在3’-5’核酸外切。在自然界中,DNA结合和切割通常需要蛋白质和两种RNA。然而,可以对单向导RNA(“sgRNA”或简称为“gRNA”)进行工程改造,以便将crRNA和tracrRNA的各个方面整合到单个RNA种类中。参见例如Jinek M.等人,Science 337:816-821(2012),其全部内容通过引用并入本文。Cas9识别CRISPR重复序列(PAM或原型间隔区相邻基序)中的短基序,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众所周知的(参见例如“Complete genome sequence of an M1strain of Streptococcus pyogenes.”Ferretti,J.J.等人,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.等人,Nature 471:602-607(2011);和“Programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.”Jinek M.等人,Science 337:816-821(2012),其全部内容通过引用并入本文)。Cas9直向同源物已在各种物种中得到描述,包括但不限于化脓链球菌和嗜热链球菌。基于本公开内容,其他合适的Cas9核酸酶和序列对于本领域技术人员来说是显而易见的,并且此类Cas9核酸酶和序列包括来自Chylinski、Rhun和Charpentier“The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems”(2013)RNA Biology 10:5,726-737中公开的生物体和基因座的Cas9序列,其全部内容通过引用并入本文。在一些实施例中,Cas9核酸酶具有无活性(例如失活)DNA切割结构域,即Cas9是切口酶。

[0535] 在一些实施例中,所述向导多核苷酸是至少一种单向导RNA(“sgRNA”或“gRNA”)。在一些实施例中,所述向导多核苷酸是至少一种tracrRNA。在一些实施例中,所述向导多核苷酸不需要PAM序列来将多核苷酸可编程的DNA结合结构域(例如Cas9或Cpf1)引导至目标核苷酸序列。

[0536] 本文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CRISPR衍生结

构域)可以通过与向导多核苷酸缔合来识别目标多核苷酸序列。向导多核苷酸(例如gRNA)通常是单链的,并且可以被编程为位点特异性地结合(即通过互补碱基配对)多核苷酸的目标序列,从而将与向导核酸结合的碱基编辑器导向目标序列。向导多核苷酸可以是DNA。向导多核苷酸可以是RNA。在一些实施例中,向导多核苷酸包含天然核苷酸(例如腺苷)。在一些实施例中,向导多核苷酸包含非天然(或非天然)核苷酸(例如肽核酸或核苷酸类似物)。在一些实施例中,向导核酸序列的靶向区域在长度上可以是至少15、16、17、18、19、20、21、22、23、24、25、26、27、28、29或30个核苷酸。向导核酸的靶向区域的长度可以在10-30个核苷酸之间,或长度在15-25个核苷酸之间,或长度在15-20个核苷酸之间。

[0537] 在一些实施例中,向导多核苷酸包含两个或更多个单独的多核苷酸,它们可以通过例如互补碱基配对(例如双向导多核苷酸)彼此相互作用。例如,向导多核苷酸可包含CRISPR RNA(crRNA)和反式激活CRISPR RNA(tracrRNA)。例如,向导多核苷酸可包含一种或多种反式激活CRISPR RNA(tracrRNA)。

[0538] 在II型CRISPR系统中,CRISPR蛋白(例如Cas9)靶向核酸通常需要在包含识别目标序列的序列的第一个RNA分子(crRNA)和包含重复序列的第二个RNA分子(trRNA)之间进行互补碱基配对,所述重复序列形成稳定向导RNA-CRISPR蛋白复合物的支架区。此类双向导RNA系统可用作向导多核苷酸以将本文公开的碱基编辑器导向目标多核苷酸序列。

[0539] 在一些实施例中,本文提供的碱基编辑器利用单向导多核苷酸(例如gRNA)。在一些实施例中,本文提供的碱基编辑器利用双向导多核苷酸(例如双gRNA)。在一些实施例中,本文提供的碱基编辑器利用一种或多种向导多核苷酸(例如多个gRNA)。在一些实施例中,单向导多核苷酸用于本文所述的不同碱基编辑器。例如,单向导多核苷酸可用于腺苷碱基编辑器。

[0540] 在其他实施例中,向导多核苷酸可在单个分子(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。例如单分子向导多核苷酸可以是单向导RNA(sgRNA或gRNA)。在本文中,术语向导多核苷酸序列涵盖能够与碱基编辑器相互作用并将碱基编辑器引导至目标多核苷酸序列的任何单、双或多分子核酸。

[0541] 通常,向导多核苷酸(例如crRNA/trRNA复合物或gRNA)包含“多核苷酸靶向片段”,该片段包括能够识别和结合目标多核苷酸序列的序列,以及“蛋白质结合片段”,其将向导多核苷酸稳定在碱基编辑器的多核苷酸可编程核苷酸结合结构域组分内。在一些实施例中,所述向导多核苷酸的多核苷酸靶向区段识别并结合DNA多核苷酸,从而促进DNA中碱基的编辑。在其他实施例中,所述向导多核苷酸的多核苷酸靶向区段识别并结合RNA多核苷酸,从而促进RNA中碱基的编辑。在本文中,“区段”是指分子的区段或区域,例如,向导多核苷酸中的一段连续核苷酸。区段也可指复合物的区域/区段,使得区段可包含多于一种分子的区域。例如,当向导多核苷酸包含多个核酸分子时,蛋白质结合区段可以包括例如沿着互补区域杂交的多个单独分子的全部或一部分。在一些实施例中,包含两个单独分子的靶向DNA的RNA的蛋白质结合区段可包含(i)长度为100个碱基对的第一个RNA分子的40-75个碱基对;和(ii)长度为50个碱基对的第二个RNA分子的10-25个碱基对。除非在特定上下文中另有明确定义,否则“片段”的定义不限于特定总碱基对数,不限于来自给定RNA分子的任何特定碱基对数,不限于复合物中特定数量的单独分子,并且可以包括具有任何总长度的RNA分子区域并且可以包括与其他分子互补的区域。

[0542] 向导RNA或向导多核苷酸可包含两种或更多种RNA,例如CRISPR RNA(crRNA)和反式激活crRNA(tracrRNA)。在一些实施例中,向导RNA或向导多核苷酸包含单链RNA或通过crRNA和tracrRNA的一部分(例如功能部分)融合形成的单向导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可以与目标DNA杂交。

[0543] 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。通过用包含编码向导RNA和启动子的序列的分离的向导RNA或质粒DNA转染细胞,可以将向导RNA或向导多核苷酸转移到细胞中。向导RNA或向导多核苷酸也可以以其他方式转移到细胞中,例如使用病毒介导的基因递送。

[0544] 可以分离向导RNA或向导多核苷酸。例如,向导RNA可以以分离的RNA的形式转染到细胞或生物体中。可以使用本领域已知的任何体外转录系统通过体外转录来制备向导RNA。向导RNA可以以分离的RNA的形式而不是以包含向导RNA编码序列的质粒之形式转移到细胞中。

[0545] 向导RNA或向导多核苷酸可包含三个区域:5'端的第一区域可与染色体序列中的靶位点互补,第二内部区域可形成茎环结构,以及第三3'区域可以是单链的。每个向导RNA的第一区域也可以不同,使得每个向导RNA将融合蛋白导引至特定靶位点。此外,每个向导RNA的第二和第三区域在所有向导RNA中可以相同。

[0546] 向导RNA或向导多核苷酸的第一区域可与染色体序列中靶位点的序列互补,使得向导RNA的第一区域可与靶位点碱基配对。在一些实施例中,向导RNA的第一区域可包含或从约10个核苷酸到25个核苷酸(即,从10个核苷酸至约25个核苷酸;或从约10个核苷酸至约25个核苷酸;或从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。例如向导RNA的第一区域和染色体序列中的靶位点之间的碱基配对区域可以是或可以是约10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25或更多个核苷酸的长度。在一些实施例中,向导RNA的第一区域的长度可以是或可以是约19、20或21个核苷酸。

[0547] 向导RNA或向导多核苷酸还可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。环和茎的长度可以变化。例如,环的长度范围可为约3至10个核苷酸,而茎的长度范围可为约6至20个碱基对。茎可包含一个或多个1至10个或约10个核苷酸的凸起。第二区域的总长度可在约16至60个核苷酸的长度范围内。例如,环的长度可以是或可以是约4个核苷酸,茎可以是或可以是约12个碱基对。

[0548] 向导RNA或向导多核苷酸还可以在3'端包含基本上可以是单链的第三区域。例如,第三区域有时与目标细胞中的任何染色体序列不互补,有时与向导RNA的其余部分不互补。此外,第三区域的长度可以变化。第三区域的长度可以多于或多于约4个核苷酸。例如,第三区域的长度可以在约5至60个核苷酸的长度范围内。

[0549] 向导RNA或向导多核苷酸可以靶向基因目标的任何外显子或内含子。在一些实施例中,向导可以靶向基因的外显子1或2;在其他实施例中,向导可以靶向基因的外显子3或4。组成物可包含均靶向相同外显子的多个向导RNA,或在一些实施例中,可包含靶向不同外显子的多个向导RNA。可以靶向基因的外显子和内含子。

[0550] 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。目标核酸可以少于约20个核苷酸。目标核酸的长度可为至少约5、10、15、16、17、18、19、20、21、22、23、24、25、30或1-100个核苷酸之间的任何位置。目标核酸的长度可为至多约5、10、15、16、17、18、19、20、

21、22、23、24、25、30、40、50或1-100个核苷酸之间的任何位置。目标核酸序列可以是或可以是约20个碱基,其紧邻PAM的第一个核苷酸的5'处。向导RNA可以靶向核酸序列。目标核酸可以是至少或至少约1-10、1-20、1-30、1-40、1-50、1-60、1-70、1-80、1-90或1-100个核苷酸。

[0551] 向导多核苷酸,例如向导RNA,可以指可以与另一种核酸杂交的核酸,例如细胞基因组中的目标核酸或原型间隔区。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成位点特异性地结合核酸序列。向导多核苷酸可以包含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条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分子导入细胞。例如RNA分子可以在体外转录和/或可以化学合成。RNA可以从合成DNA分子转录,例如gBlocks®基因片段。然后可以将向导RNA作为RNA分子导入细胞中。向导RNA也可以以非RNA核酸分子例如DNA分子的形式引入细胞中。例如,编码向导RNA的DNA可以与启动子控制序列可操作地连接,以在感兴趣的细胞中表达向导RNA。RNA编码序列可以与被RNA聚合酶III(Pol III)识别的启动子序列可操作地连接。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。在一些实施例中,质粒载体(例如px333载体)可包含至少两个向导RNA编码的DNA序列。

[0552] 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,为了尽量减少核碱基编辑器系统中脱氨酶结构域潜在底物混杂的影响(例如AID结构域),可能无意中成为脱氨目标的残基数量(例如可能存在于目标核酸基因座内ssDNA上的脱靶C残基)可以被最小化。此外,软件工具可用于优化对应于目标核酸序列的gRNA,例如,最大限度地减少整个基因组的总脱靶活性。例如,对于使用化脓性链球菌Cas9的每个可能的靶向结构域选择,所有脱靶序列(在选定的PAM之前,例如NAG或NGG)都可以在基因组中被识别,其中包含多达特定数量(例如1、2、3、4、5、6、7、8、9或10)个错配的碱基对。可以识别与靶位点互补的gRNA的第一个区域,并且可以根据其总预测的脱靶分数对所有第一个区域(例如crRNA)进行排序;排名靠前的靶向结构域代表那些可能具有最大在靶和最少脱靶活性的结构域。可以使用本领域已知的和/或如本文所述的方法对候选靶向gRNA进行功能评估。

[0553] 作为非限制性实例,可使用DNA序列搜索算法来鉴定与Cas9一起使用的向导RNA的crRNA中的目标DNA杂交序列。gRNA设计可以使用基于公共工具cas-offinder的定制gRNA设计软件进行,如Bae S.、Park J.和Kim J.-S.于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。该软件在计算全基因组脱靶倾向后为向导评分。对于长度从17到24不等的向导,通常会考虑从完美匹配到7次不匹配的匹配。一旦通过计算确定了脱靶位点,就会为每个向导计算总分,并使用网络界面汇总在表格输出中。除了识别与PAM序列相邻的潜在靶位点外,该软件还识别与所选靶位点相差1、2、3或3个以上核苷酸的所有PAM相邻序列。可以获得目标核酸序列例如目标基因的基因组DNA序列,并且可以使用公开可用的工具例如RepeatMasker程序筛选重复组件。RepeatMasker在输入DNA序列中搜索重复要素和低复杂性区域。输出是给定查询序列中存在的重复的详细注释。

[0554] 在鉴定之后,向导RNA的第一区域,例如crRNA,可以根据它们与靶位点的距离、它们的正交性和5'核苷酸的存在进行分级,以便与相关的PAM序列(例如5'G基于人类基因组

中包含相关PAM的紧密匹配的鉴定,例如化脓性链球菌的NGG PAM、金黄色葡萄球菌的NNGRRT或NNGRRV PAM)。如本文所用,正交性是指人类基因组中包含最少数量的与目标序列的错配的序列的数量。例如,“高水平正交性”或“良好正交性”可以指在人类基因组中除了预期靶标之外没有相同序列的20聚体靶向结构域,也没有在目标序列中包含一两个错配。可以选择具有良好正交性的靶向结构域以最小化脱靶DNA切割。

[0555] 在一些实施例中,报告系统可用于检测碱基编辑活性和测试候选向导多核苷酸。在一些实施例中,报告系统可以包括基于报告基因的测定,其中碱基编辑活性导致报告基因的表达。例如,报告系统可包括包含失活起始密码子的报告基因,例如模板链上从3'-TAC-5'到3'-CAC-5'的突变。目标C成功脱氨基后,相应的mRNA将转录为5'-AUG-3'而不是5'-GUG-3',从而实现报告基因的翻译。合适的报告基因对本领域技术人员来说是显而易见的。报告基因的非限制性实例包括编码绿色荧光蛋白(GFP)、红色荧光蛋白(RFP)、荧光素酶、分泌性碱性磷酸酶(SEAP)或其表达对本领域技术人员来说是可检测的和显而易见的任何其他基因。报告系统可用于测试许多不同的gRNA,例如,确定相应脱氨酶将靶向目标DNA序列中的哪些残基。还可以测试靶向非模板链核苷酸残基的sgRNA,以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。在一些实施例中,此类gRNA可以被设计为使得突变的起始密码子不会与gRNA杂交。向导多核苷酸可包含标准核苷酸、经修饰的核苷酸(例如假尿苷)、核苷酸异构体和/或核苷酸类似物。在一些实施例中,向导多核苷酸可包含至少一种可检测标记。可检测标记可以是荧光团(例如FAM、TMR、Cy3、Cy5、德克萨斯红、俄勒冈绿、亚历克萨荧光(Alexa Fluors)、Halo标签或任何其他合适的荧光染料)、检测标签(例如生物素、地高辛等)、量子点或金粒子。

[0556] 向导多核苷酸可以化学和/或酶促合成。例如,可以使用基于亚磷酰胺的标准固相合成方法合成向导RNA。或者,可以通过将编码向导RNA的DNA与被噬菌体RNA聚合酶识别的启动子控制序列可操作地连接来体外合成向导RNA。合适的噬菌体启动子序列的例子包括T7、T3、SP6启动子序列或其变体。在向导RNA包含两个单独的分子(例如crRNA和tracrRNA)的实施例中,crRNA可以化学合成并且tracrRNA可以酶促合成。

[0557] 在一些实施例中,碱基编辑器系统可以包含多个向导多核苷酸,例如gRNA,其将碱基编辑器靶向一个或多个基因座(例如至少1个gRNA、至少2个gRNA、至少5个gRNA、至少10个gRNA、至少20gRNA、至少30g RNA或至少50gRNA)。在一些实施例中,多个gRNA序列可以串联排列在单个多核苷酸中。在一些实施例中,串联排列的gRNA序列被同向重复分开。

[0558] 编码向导RNA或向导多核苷酸的DNA序列也可以是载体的一部分。此外,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)、选择标记序列(例如GFP或抗生素抗性基因,例如嘌呤霉素)、复制起点等。编码向导RNA或向导多核苷酸的DNA分子可以是线性的或环状的。

[0559] 在一些实施例中,碱基编辑器系统的一个或多个组件可由DNA序列编码。此类DNA序列可以一起或单独引入表达系统,例如细胞。例如,可以将编码多核苷酸可编程核苷酸结合结构域和向导RNA的DNA序列引入细胞,每个DNA序列可以是单独分子的一部分(例如一个包含多核苷酸可编程核苷酸结合结构域编码序列的载体和第二个含有向导RNA编码序列的载体)或两者可以是同一分子的一部分(例如一种同时含有多核苷酸可编程核苷酸结合结构域和向导RNA的编码(和调控)序列的载体)。

[0560] 向导多核苷酸可以包含一个或多个修饰以提供具有新的或增强的特征的核酸。向导多核苷酸可包含核酸亲和标签。向导多核苷酸可包括合成核苷酸、合成核苷酸类似物、核苷酸衍生物和/或经修饰的核苷酸。

[0561] 在一些实施例中,gRNA或向导多核苷酸可包含修饰。可以在gRNA或向导多核苷酸的任何位置进行修饰。可以对单个gRNA或向导多核苷酸进行不止一种修饰。gRNA或向导多核苷酸可以在修饰后进行质量控制。在一些实施例中,质量控制可包括PAGE、HPLC、MS或其任何组合。

[0562] gRNA或向导多核苷酸的修饰可以是取代、插入、缺失、化学修饰、物理修饰、稳定化、纯化或其任何组合。

[0563] gRNA或向导多核苷酸也可以被5'腺苷酸、5'鸟苷-三磷酸帽、5'N7-甲基鸟苷-三磷酸帽、5'三磷酸帽、3'磷酸、3'硫代磷酸、5'磷酸、5'硫代磷酸、Cis-Syn胸苷二聚体、三聚体、C12间隔区、C3间隔区、C6间隔区、d间隔区、PC间隔区、r间隔区、间隔区18、间隔区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、胆固醇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、黑洞猝灭剂1、黑洞淬灭剂2、DABCYLSE、dT-DABCYL、IRDye QC-1、QSY-21、QSY-35、QSY-7、QSY-9、羧基接头、巯基接头、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、糖经修饰的类似物、摇摆/通用碱基、荧光染料标记、2'-氟RNA、2'-O-甲基RNA、甲基膦酸酯、磷酸二酯DNA、磷二酯RNA、硫代磷酸酯DNA、硫代磷酸酯RNA、UNA、假尿苷-5'-三磷酸、5'-甲基胞苷-5'-三磷酸或其任何组合修饰。

[0564] 在一些实施例中,修改是永久性的。在其他实施例中,修改是暂时的。在一些实施例中,对gRNA或向导多核苷酸进行了多次修饰。gRNA或向导多核苷酸修饰可以改变核苷酸的理化特性,例如它们的构象、极性、疏水性、化学反应性、碱基配对相互作用或其任何组合。

[0565] PAM序列可以是本领域已知的任何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。

[0566] 修饰也可以是硫代磷酸酯替代物。在一些实施例中,天然磷酸二酯键可易于被细胞核酸酶快速降解;使用硫代磷酸酯(PS)键替代物对核苷酸间键的修饰可以更稳定地通过细胞降解进行水解。修饰可以增加gRNA或向导多核苷酸的稳定性。修饰还可以增强生物活性。在一些实施例中,硫代磷酸酯增强的RNA gRNA可以抑制RNase A、RNase T1、小牛血清核酸酶或其任何组合。这些特性可以使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,可以在gRNA的5'-或3'-端的最后3-5个核苷酸之间引入硫代磷酸(PS)键,这可以抑制外切核酸酶降解。在一些实施例中,可以在整个gRNA中添加硫代磷酸酯键以减少核酸内切酶的攻击。

[0567] 原型间隔区相邻基序

[0568] 术语"原型间隔区相邻基序(protospacer adjacent motif,PAM)"或PAM样基序是指紧跟在CRISPR细菌适应性免疫系统中Cas9核酸酶靶向的DNA序列之后的2-6个碱基对DNA

序列。在一些实施例中,PAM可以是5’PAM(即位于原型间隔区5’端的上游)。在其他实施例中,PAM可以是3’PAM(即位于原型间隔区5’端的下游)。

[0569] PAM序列对于目标结合至关重要,但确切的序列取决于Cas蛋白的类型。

[0570] 本文提供的碱基编辑器可包含能够结合包含规范或非规范原型间隔区相邻基序(PAM)序列的核苷酸序列的CRISPR蛋白衍生结构域。PAM位点是靠近目标多核苷酸序列的核苷酸序列。本公开的一些方面提供包含具有不同PAM特异性的全部或部分CRISPR蛋白的碱基编辑器。

[0571] 例如,典型的Cas9蛋白,例如来自化脓性链球菌(spCas9)的Cas9,需要典型的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是鸟嘌呤。PAM可以是CRISPR蛋白质特异性的,并且在包含不同CRISPR蛋白质衍生结构域的不同碱基编辑器之间可以不同。PAM可以是目标序列的5’或3’。PAM可以位于目标序列的上游或下游。PAM的长度可以是1、2、3、4、5、6、7、8、9、10个或更多个核苷酸。通常,PAM的长度在2-6个核苷酸之间。下面的表2中描述了几种PAM变体。

[0572] 表2.Cas9蛋白和相应的PAM序列

[0573]

变体	PAM
spCas9	NGG
spCas9-VRQR	NGA
spCas9-VRER	NGCG
xCas9 (sp)	NGN
saCas9	NNGRRT

[0574]

saCas9-KKH	NNNRRT
spCas9-MQKSER	NGCG
spCas9-MQKSER	NGCN
spCas9-LRKIQK	NGTN
spCas9-LRVSQK	NGTN
spCas9-LRVSQL	NGTN
spCas9-MQKFRAER	NGC
Cpf1	5’ (TTTV)
SpyMac	5’-NAA-3’

[0575] 在一些实施例中,PAM是NGC。在一些实施例中,NGC PAM被Cas9变体识别。在一些实施例中,NGC PAM变体包括一个或多个选自D1135M、S1136Q、G1218K、E1219F、A1322R、D1332A、R1335E和T1337R(统称为“MQKFRAER”)的氨基酸取代。

[0576] 在一些实施例中,PAM是NGT。在一些实施例中,NGT PAM被Cas9变体识别。在一些实施例中,NGT PAM变体通过在一个或多个残基1335、1337、1135、1136、1218和/或1219处的靶向突变产生。在一些实施例中,NGT PAM变体通过在一个或多个残基1219、1335、1337、1218处的靶向突变产生。在一些实施例中,NGT PAM变体是通过在一个或多个残基1135、1136、1218、1219和1335处的靶向突变产生的。在一些实施例中,NGT PAM变体选自下表3A和3B中提供的一组靶向突变。

[0577] 表3A:残基1219、1335、1337、1218处的NGT PAM变体突变

[0578]

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

[0579]

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	

[0580] 表3B:残基1135、1136、1218、1219和1335处的NGT PAM变体突变

[0581]

变体	D1135L	S1136R	G1218S	E1219V	R1335Q
27	G				
28	V				

[0582]

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			

[0583] 在一些实施例中,NGT PAM变体选自表2和3中的变体5、7、28、31或36。在一些实施例中,变体具有改进的NGT PAM识别。

[0584] 在一些实施例中,NGT PAM变体在残基1219、1335、1337和/或1218处具有突变。在一些实施例中,从下表4中提供的变体中选择具有突变以提高识别的NGT PAM变体。

[0585] 表4:残基1219、1335、1337和1218处的NGT PAM变体突变

[0586]

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

[0587] 在一些实施例中,可以如下表5中提供的产生对NGT PAM具有特异性的碱基编辑器。

[0588] 表5.NGT PAM变体

[0589]

	NGTN variant	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

[0590]

变体6	LRSVRQL	L	R	S	V	R	Q	L

[0591] 在一些实施例中,所述NGTN变体是变体1。在一些实施例中,所述NGTN变体是变体2。在一些实施例中,所述NGTN变体是变体3。在一些实施例中,所述NGTN变体是变体4。在一些实施例中,所述NGTN变体是变体5。在一些实施例中,所述NGTN变体是变体6。

[0592] 在一些实施例中,所述Cas9结构域是来自化脓性链球菌(SpCas9)的Cas9结构域。在一些实施例中,所述SpCas9结构域是核酸酶活性SpCas9、核酸酶无活性SpCas9(SpCas9d)

或SpCas9切口酶(SpCas9n)。在一些实施例中,所述SpCas9包含D9X突变,或本文提供的任何氨基酸序列中的相应突变,其中X是除D之外的任何氨基酸。在一些实施例中,所述SpCas9包含D9A突变,或在本文提供的任何氨基酸序列中相应突变。在一些实施例中,所述SpCas9结构域、SpCas9d结构域或SpCas9n结构域可结合具有非规范PAM的核酸序列。在一些实施例中,所述SpCas9结构域、SpCas9d结构域或SpCas9n结构域可结合具有NGG、NGA或NGCG PAM序列的核酸序列。在一些实施例中,所述SpCas9结构域包含D1134X、R1335X和T1336X突变中的一个或多个,或本文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。在一些实施例中,所述SpCas9结构域包含D1134E、R1335Q和T1336R突变中的一个或多个,或本文提供的任何氨基酸序列中的相应突变。在一些实施例中,所述SpCas9结构域包含D1134E、R1335Q和T1336R突变,或本文提供的任何氨基酸序列中的相应突变。在一些实施例中,所述SpCas9结构域包含D1134X、R1335X和T1336X突变中的一个或多个,或本文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。在一些实施例中,所述SpCas9结构域包含D1134V、R1335Q和T1336R突变中的一个或多个,或本文提供的任何氨基酸序列中的相应突变。在一些实施例中,所述SpCas9结构域包含D1134V、R1335Q和T1336R突变,或本文提供的任何氨基酸序列中的相应突变。在一些实施例中,所述SpCas9结构域包含D1134X、G1217X、R1335X和T1336X突变中的一个或多个,或本文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。在一些实施例中,所述SpCas9结构域包含D1134V、G1217R、R1335Q和T1336R突变中的一个或多个,或本文提供的任何氨基酸序列中的相应突变。在一些实施例中,所述SpCas9结构域包含D1134V、G1217R、R1335Q和T1336R突变,或本文提供的任何氨基酸序列中的相应突变。

[0593] 在一些实施例中,本文提供的任何融合蛋白的Cas9结构域包含的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。在一些实施例中,本文提供的任何融合蛋白的Cas9结构域包含本文描述的任何Cas9多肽的氨基酸序列。在一些实施例中,本文提供的任何融合蛋白的Cas9结构域由本文描述的任何Cas9多肽的氨基酸序列组成。

[0594] 在一些实施例中,可将本文公开的碱基编辑器的CRISPR蛋白衍生结构域识别的PAM提供给与编码碱基编辑器的插入物(例如AAV插入物)不同的寡核苷酸上的细胞。在这样的实施例中,在单独的寡核苷酸上提供PAM可以允许切割目标序列,否则将无法切割,因为在与目标序列相同的多核苷酸上不存在相邻的PAM。

[0595] 在一个实施例中,化脓性链球菌Cas9(SpCas9)可用作基因组工程的CRISPR核酸内切酶。但是,也可以使用其他的。在一些实施例中,不同的核酸内切酶可用于靶向某些基因组目标。在一些实施例中,可以使用具有非NGG PAM序列的合成SpCas9衍生变体。此外,已经鉴定了来自不同物种的其他Cas9直向同源物,并且这些“非SpCas9”可以结合也可用于本公开的多种PAM序列。例如,相对较大的SpCas9(大约4kb编码序列)会导致携带SpCas9 cDNA的质粒无法在细胞中有效表达。相反,金黄色葡萄球菌Cas9(SaCas9)的编码序列比SpCas9短约1千碱基,可能使其在细胞中有效表达。与SpCas9类似,SaCas9核酸内切酶能够在体外和小鼠体内修饰哺乳动物细胞中的目標基因。在一些实施例中,Cas蛋白可以靶向不同的PAM序列。在一些实施例中,目標基因可以与例如Cas9 PAM、5’-NGG相邻。在其他实施例中,其他

Cas9直向同源物可具有不同的PAM要求。例如,其他PAM,例如嗜热链球菌(CRISPR1为5'-NNAGAA、CRISPR3为5'-NGGNG)和脑膜炎奈瑟菌(5'-NNNNGATT)的PAM也可以与目标基因相邻。

[0596] 在一些实施例中,对于化脓性链球菌系统,目标基因序列可以在(即5'到)5'-NGG PAM之前,并且20-nt向导RNA序列可以与相反链碱基配对以介导与PAM相邻的Cas9裂解。在一些实施例中,相邻切割可以是或可以是约PAM上游的3个碱基对。在一些实施例中,相邻切割可以是或可以是约PAM上游的10个碱基对。在一些实施例中,相邻切割可以是或可以是约PAM上游的0-20个碱基对。例如,相邻切割可以紧挨着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。相邻的切割也可以在PAM的下游1到30个碱基对。能够结合PAM序列的示例性SpCas9蛋白的序列如下:

[0597] 示例性结合PAM的SpCas9的氨基酸序列如下:

[0598] MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKP
IREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0599] 示例性的结合PAM的SpCas9n的氨基酸序列如下:

[0600] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS

DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKP
IREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0601] 示例性的结合PAM的SpEQR Cas9的氨基酸序列如下：

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESVLVEEDKKHERHPIFG
NIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD
VDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGN
LIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE
VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL
SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG

[0602] RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL
HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER
MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDH
IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK
MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF
ATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF**E**SPTVA
YSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE
QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA
PAAFKYFDTTIDRK**Q**Y**R**STKEVLDATLIHQSITGLYETRIDLSQLGGD

[0603] 在上述序列中,残基E1134、Q1334和R1336可以从D1134、R1335和T1336突变以产生SpEQR Cas9,用下划线和粗体表示。

[0604] 示例性的结合PAM的SpVQR Cas9的氨基酸序列如下:

[0605]
MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFG
NIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD
VDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGN
LIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE
VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL
SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL
HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER
MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDH
IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK
MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF
ATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF**V**SPTVA
YSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE
QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA
PAAFKYFDTTIDRK**Q**Y**R**STKEVLDATLIHQSITGLYETRIDLSQLGGD

[0606]

[0607] 在上述序列中,残基V1134、Q1334和R1336可以从D1134、R1335和T1336突变以产生SpVQR Cas9,用下划线和粗体表示。

[0608] 示例性的结合PAM的SpVRER Cas9的氨基酸序列如下:

[0609]
MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFG
NIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD
VDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGN
LIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE
VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL
SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL
HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER
MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDH
IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
[0610]
KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK
MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF
ATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF**V**SPTVA
YSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASA**R**ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE
QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA
PAAFKYFDTTIDRK**E**Y**R**STKEVLDATLIHQSITGLYETRIDLSQLGGD.

[0611] 在上述序列中,残基V1134、R1217、Q1334和R1336可以从D1134、G1217、R1335和T1336突变以产生SpVRER Cas9,用下划线和粗体表示。

[0612] 在一些实施例中,工程化的SpCas9变体能够识别侧翼为3'H(非G PAM)的原型间隔区相邻基序(PAM)序列(见表1A-1D;图24)。在一些实施例中,SpCas9变体识别NRNH PAM(其中R是A或G并且H是A、C或T)。在一些实施例中,非G PAM是NRRH、NRTH或NRCH(见例如Miller,S.M.等人,Continuous evolution of SpCas9 variants compatible with non-G PAMs,Nat.Biotechnol.(2020),其内容通过引用整体并入本文)。

[0613] 在一些实施例中,所述Cas9结构域是重组Cas9结构域。在一些实施例中,所述重组Cas9结构域是SpyMacCas9结构域。在一些实施例中,所述SpyMacCas9结构域是核酸酶活性的SpyMacCas9、核酸酶无活性的SpyMacCas9(SpyMacCas9d)或SpyMacCas9切口酶

(SpyMacCas9n)。在一些实施例中,所述SaCas9结构域、SaCas9d结构域或SaCas9n结构域可结合具有非规范PAM的核酸序列。在一些实施例中,所述SpyMacCas9结构域、SpCas9d结构域或SpCas9n结构域可以结合具有NAA PAM序列的核酸序列。

[0614] 具有天然5'-NAAN-3'PAM特异性的猕猴链球菌中Spy Cas9的示例性Cas9 A同系物的序列是本领域已知的,并且例如由Jakimo等人描述,(www.biorxiv.org/content/biorxiv/early/2018/09/27/429654.full.pdf),并在下面提供。

[0615] SpyMacCas9

[0616] MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQIYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRGMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGHSLHEQIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEIQTVGQNGGLFDDNPKSPLEVTPSKLVPLKKELNPKKYGGYQKPTTAYPVLLITDTKQLIPISVMNKKQFEQNPVKFLRDRGYQQVGKNDFIKLPKYTLVDIGDGIKRLWASSKEIHKGNQLVVSKKSQILLYHAHHLDSDLSNDYLQNHNQQFDVLFNEIISFSKKCKLGKEHIQKIENVYSNKKNSASIEELAESFIKLLGFTQLGATSPFNFLGVKLNQKQYKGKKDYILPCTEGTLIRQSITGLYETRVDLSKIGED.

[0617] 在一些实施例中,变体Cas9蛋白包含H840A、P475A、W476A、N477A、D1125A、W1126A和D1218A突变,使得多肽切割目标DNA或RNA的能力降低。此类Cas9蛋白切割目标DNA(例如单链目标DNA)的能力降低,但保留结合目标DNA(例如单链目标DNA)的能力。作为另一个非限制性实例,在一些实施例中,变体Cas9蛋白含有D10A、H840A、P475A、W476A、N477A、D1125A、W1126A和D1218A突变,使得多肽切割目标DNA的能力降低。此类Cas9蛋白切割靶DNA(例如单链目标DNA)的能力降低,但保留结合目标DNA(例如单链目标DNA)的能力。在一些实施例中,当变体Cas9蛋白包含W476A和W1126A突变或当变体Cas9蛋白包含P475A、W476A、N477A、D1125A、W1126A和D1218A突变时,变体Cas9蛋白不与PAM序列有效结合。因此,在一些此类情况下,当此类变体Cas9蛋白用于结合方法时,该方法不需要PAM序列。换言之,在一些实施例中,当这种变体Cas9蛋白用于结合方法中时,该方法可以包括向导RNA,但是该方法可以在不存在PAM序列的情况下进行(并且结合的特异性是因此由向导RNA的靶向片段提供)。可以突变其他残基以实现上述效果(即灭活一个或其他核酸酶部分)。作为非限制性实例,残基D10、G12、G17、E762、H840、N854、N863、H982、H983、A984、D986和/或A987处可以被改变(即被取代)。此外,丙氨酸取代以外的突变也是合适的。

[0618] 在一些实施例中,碱基编辑器的CRISPR蛋白衍生结构域可包含具有规范PAM序列(NGG)的Cas9蛋白的全部或部分。在其他实施例中,碱基编辑器的Cas9衍生结构域可采用非规范PAM序列。此类序列已在本领域中描述并且对技术人员来说是显而易见的。例如,结合非规范PAM序列的Cas9结构域已在Kleinstiver,B.P.等人,"Engineered CRISPR-Cas9 nucleases with altered PAM specificities"Nature 523,481-485(2015);和Kleinstiver,B.P.等人,"Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition"Nature Biotechnology 33,1293-1298(2015)中进行了描述,其各自的全部内容通过引用并入本文。

[0619] 降低PAM独占性的Cas9结构域

[0620] 通常,Cas9蛋白,例如来自化脓性链球菌(spCas9)的Cas9,需要典型的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)、或胞嘧啶(C),而G是鸟苷。这可能会限制编辑基因组内所需碱基的能力。在一些实施例中,本文提供的碱基编辑融合蛋白可能需要放置在精确位置,例如包含位于PAM上游的目标碱基的区域。见例如Komor,A.C.等人,"Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage"Nature533,420-424(2016),其全部内容通过引用并入本文。因此,在一些实施例中,本文提供的任何融合蛋白可以包含能够结合不包含规范(例如NGG)PAM序列的核苷酸序列的Cas9结构域。结合非规范PAM序列的Cas9结构域已在本领域中描述并且对技术人员来说是显而易见的。例如结合非规范PAM序列的Cas9结构域已在Kleinstiver,B.P.等人"Engineered CRISPR-Cas9nucleases with altered PAM specificities"Nature 523,481-485(2015);和Kleinstiver,B.P.等人,"Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition"NatureBiotechnology 33,1293-1298(2015)中进行了描述,其各自的全部内容通过引用并入本文。

[0621] 高保真Cas9结构域

[0622] 本公开的一些方面提供高保真Cas9結構域。在一些实施例中,高保真Cas9結構域是工程化的Cas9結構域,与相应的野生型Cas9結構域相比,其包含一个或多个突变,这些突变降低了Cas9結構域和DNA的糖-磷酸骨架之间的静电相互作用。不希望受任何特定理论的束缚,与DNA糖-磷酸骨架的静电相互作用降低的高保真Cas9结构域可能具有较少的脱靶效应。在一些实施例中,Cas9结构域(例如野生型Cas9结构域)包含一种或多种降低Cas9结构域与DNA糖-磷酸骨架之间缔合的突变。在一些实施例中,Cas9结构域包含一种或多种突变,其将Cas9结构域与DNA的糖-磷酸骨架之间的缔合降低至少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%。

[0623] 在一些实施例中,本文提供的任何Cas9融合蛋白包含N497X、R661X、Q695X和/或Q926X突变中的一个或多个,或本文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。在一些实施例中,本文提供的任何Cas9融合蛋白包含N497A、R661A、Q695A和/或Q926A突变中的一个或多个,或本文提供的任何氨基酸序列中的相应突变。在一些实施例中,Cas9结构域包含D10A突变,或本文提供的任何氨基酸序列中的相应突变。具有高保真度的Cas9结构域是本领域已知的并且对技术人员来说是显而易见的。例如,具有高保真度的Cas9结

构域已在Kleinstiver,B.P.等人,“High-fidelity CRISPR-Cas9nucleases with no detectable genome-wide off-target effects.”Nature529,490-495(2016);和Slaymaker,I.M.等人,“Rationally engineered Cas9 nucleases with improved specificity.”Science351,84-88(2015)中描述,其各自的全部内容通过引用并入本文。

[0624] 在一些实施例中,改良的Cas9是高保真Cas9酶。在一些实施例中,高保真Cas9酶是SpCas9(K855A)、eSpCas9(1.1)、SpCas9-HF1或超准确Cas9变体(HypaCas9)。改良的Cas9 eSpCas9(1.1)包含丙氨酸取代,削弱了HNH/RuvC凹槽与非目标DNA链之间的相互作用,防止链分离和在脱靶位点切割。同样,SpCas9-HF1通过破坏Cas9与DNA磷酸骨架相互作用的丙氨酸取代降低脱靶编辑。HypaCas9在REC3结构域中包含突变(SpCas9 N692A/M694A/Q695A/H698A),可增加Cas9校对和目标区分。与野生型Cas9相比,所有三种高保真酶产生的脱靶编辑更少。

[0625] 下面提供了示例性高保真Cas9。

[0626] 相对于Cas9的高保真Cas9结构域突变以粗体和下划线显示。

DKKYSIGL**A**IGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEA
TRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGN
IVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDV
DKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNL
IALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAIL
LSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAG
YIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHA
[0627] ILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEV
VDKGASAQSFIERMT**A**FDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLS
GEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKII
KDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG**A**
LSRKLINGIRDKQSGKTILDFLKSDGFANRNFM**A**LIHDDSLTFKEDIQKAQVSGQGDSLH
EHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERM
KRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHI
VPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT

[0628] KAERGGLSELDKAGFIKRQLVETRAITKHVAQILDSRMNTKYDENDKLIREVKVITLKSK
LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKM
IAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFA
TVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAY
SVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKY
SLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ
HKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAP
AAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0629] 包含核定位序列(NLS)的融合蛋白

[0630] 在一些实施例中,本文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)。在一个实施例中,使用双分NLS。在一些实施例中,NLS包含促进包含NLS的蛋白质输入细胞核(例如通过核转运)的氨基酸序列。在一些实施例中,本文提供的任何融合蛋白进一步包含核定位序列(NLS)。在一些实施例中,所述NLS与融合蛋白的N-末端融合。在一些实施例中,所述NLS与融合蛋白的C-末端融合。在一些实施例中,所述NLS与Cas9结构域的N-末端融合。在一些实施例中,所述NLS与nCas9结构域或dCas9结构域的C-末端融合。在一些实施例中,所述NLS与脱氨酶的N-末端融合。在一些实施例中,所述NLS与脱氨酶的C-末端融合。在一些实施例中,所述NLS通过一个或多个接头与融合蛋白融合。在一些实施例中,所述NLS在没有接头的情况下与融合蛋白融合。在一些实施例中,所述NLS包含本文提供或提及的任一NLS序列的氨基酸序列。额外的核定位序列是本领域已知的并且对技术人员来说是显而易见的。例如Plank等人,于PCT/EP2000/011690中描述了NLS序列,该文献内容通过引用并入本文,因为它们公开了示例性核定位序列。在一些实施例中,NLS包含选自以下的氨基酸序列:PKKKRKVEGADKRTADGSEFESPKKKRKV、KRTADGSEFESPKKKRKV、KRPAATKKAGQAKKKK、KKTELQTTNAENKTKKL、KRGINDRNFWRGENGRKTR、RKSGKIAAIVVKRPRKPKKKRKV、或MDSLLMNRRKFLYQFKNVRWAKGRRETYLC。

[0631] 在一些实施例中,NLS存在于接头中或NLS侧接接头,例如本文所述的接头。在一些实施例中,N-末端或C-末端NLS是双分NLS。双分NLS包含两个基本氨基酸簇,它们由相对较短的间隔序列分隔(因此双分-2部分,而单部分NLS不是)。核质蛋白的NLS KR[PAATKKAGQA]KKKK是无处不在的双分信号的原型:两个碱性氨基酸簇,由大约10个氨基酸的间隔区隔开。示例性双分NLS的序列如下:

[0632] PKKKRKVEGADKRTADGSEFESPKKKRKV

[0633] 在一些实施例中,包含腺苷脱氨酶、napDNAbp(例如Cas9结构域)和NLS的融合蛋白不包含接头序列。在一些实施例中,存在一个或多个结构域或蛋白质(例如腺苷脱氨酶、Cas9结构域或NLS)之间的接头序列。在一些实施例中,具有腺苷脱氨酶和Cas9结构域的示例性Cas9融合蛋白的一般结构包含以下结构中的任一种,其中NLS是核定位序列(例如本文提供的任何NLS),NH2是N-末端融合蛋白的C-末端,COOH是融合蛋白的C-末端:

[0634] NH_2-NLS-[腺苷脱氨酶]-[Cas9结构域]-COOH;

[0635] NH_2-NLS[Cas9结构域]-[腺苷脱氨酶]-COOH;

[0636] NH_2-[腺苷脱氨酶]-[Cas9结构域]-NLS-COOH;或

[0637] NH_2-[Cas9结构域]-[腺苷脱氨酶]-NLS-COOH。

[0638] 应当理解,本公开所述的融合蛋白可以包含一个或多个附加特征。例如,在一些实施例中,所述融合蛋白可包含抑制剂、细胞质定位序列、输出序列,例如核输出序列或其他定位序列,以及可用于溶解、纯化或检测所述融合蛋白的序列标签。本文提供的合适的蛋白质标签包括但不限于生物素羧化酶载体蛋白(BCCP)标签、myc标签、钙调蛋白标签、FLAG标签、血凝素(HA)标签、多组氨酸标签,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、Softags(例如Softag 1、Softag 3)、链标签、生物素连接酶标签、FlAsH标签、V5标签和SBP标签。其他合适的序列对本领域技术人员来说是显而易见的。在一些实施例中,所述融合蛋白包含一个或多个His标签。

[0639] 可以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,可以使用或使用约1、2、3、4、5、6、7、8、9、10个NLS。CRISPR酶可包含位于或靠近氨基末端的NLS,约或多于约1、2、3、4、5、6、7、8、9、10个位于或靠近羧基端的NLS,或任何这些的组合(例如在氨基末端的一个或多个NLS和在羧基末端的一个或多个NLS)。当存在一个以上的NLS时,每个NLS可以独立于其他选择,使得单个NLS可以存在于一个以上的副本中,和/或与一个或多个副本中存在的一个或多个其他NLS组合。

[0640] 方法中使用的CRISPR酶可包含约6个NLS。当与NLS最接近的氨基酸在距N-或C-末端的多肽链上约50个氨基酸内,例如在1、2、3、4、5、10、15、20、25、30、40或50个氨基酸内,NLS被认为靠近N或C端

[0641] 核碱基编辑结构域

[0642] 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)。所述碱基编辑器可以被编程,以通过与能够识别目标序列的向导多核苷酸相互作用来编辑目标多核苷酸序列中的一个或多个碱基。一旦识别了目标序列,所述碱基编辑器就锚定在要进行编辑的多核苷酸上,然后所述碱基编辑器的脱氨酶结构域组分可以编辑目标碱基。

[0643] 在一些实施例中,所述核碱基编辑结构域包括脱氨酶结构域。如本文特别描述的,所述脱氨酶结构域包括腺苷脱氨酶。在一些实施例中,术语“腺嘌呤脱氨酶”和“腺苷脱氨酶”可以互换使用。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)中有所描述,它们各自通过引用整体并入本文。另见Komor,A.C.等人,“Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage”Nature 533,420-424(2016);Gaudelli,N.M.等人,“Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage”Nature 551,464-471(2017);和Komor,A.C.等人,“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),其全部内容通过引用并入本文。

[0644] A到G编辑

[0645] 在一些实施例中,本文所述的碱基编辑器可包含脱氨酶结构域,其包括腺苷脱氨

酶。碱基编辑器的这种腺苷脱氨酶结构域可以通过将A脱氨基形成肌苷(I)来促进将腺嘌呤(A)核碱基编辑为鸟嘌呤(G)核碱基,肌苷(I)具有G的碱基配对特性。腺苷脱氨酶能够使脱氧核糖核酸(DNA)中脱氧腺苷残基的腺嘌呤脱氨基(即除去胺基团)。

[0646] 在一些实施例中,本文提供的核碱基编辑器可以通过将一个或多个蛋白质结构域融合在一起来制备,从而产生融合蛋白。在某些实施例中,本文提供的融合蛋白包含一种或多种改善融合蛋白的碱基编辑活性(例如效率、选择性和特异性)的特征。例如,本文提供的融合蛋白可以包含具有降低的核酸酶活性的Cas9结构域。在一些实施例中,本文提供的融合蛋白可具有不具核酸酶活性的Cas9结构域(dCas9),或切割双链DNA分子的一条链的Cas9结构域,称为Cas9切口酶(nCas9)。不希望受任何特定理论的束缚,催化残基(例如H840)的存在保持了Cas9切割含有与目标A相对的T的非编辑(例如非脱氨基)链的活性。Cas9催化残基(例如D10到A10)的突变可防止包含目标A残基的编辑链的切割。此类Cas9变体能够基于gRNA定义的目标序列在特定位置产生单链DNA断裂(缺口),从而修复非编辑链,最终导致非编辑链上的T到C变化。在一些实施例中,A-到-G碱基编辑器进一步包含肌苷碱基切除修复抑制剂,例如尿嘧啶糖基化酶抑制剂(UGI)结构域或无催化活性的肌苷特异性核酸酶。不希望受任何特定理论的束缚,UGI结构域或无催化活性的肌苷特异性核酸酶可以抑制或阻止脱氨基腺苷残基(例如肌苷)的碱基切除修复,这可以提高碱基编辑器的活性或效率。

[0647] 包含腺苷脱氨酶的碱基编辑器可作用于任何多核苷酸,包括DNA、RNA和DNA-RNA杂交体。在某些实施例中,包含腺苷脱氨酶的碱基编辑器可以使包含RNA的多核苷酸的目标A脱氨基。例如,所述碱基编辑器可以包含能够使RNA多核苷酸和/或DNA-RNA杂合多核苷酸的目标A脱氨基的腺苷脱氨酶结构域。在一个实施例中,掺入碱基编辑器的腺苷脱氨酶包含作用于RNA(ADAR,例如ADAR1或ADAR2)的全部或部分腺苷脱氨酶。在另一个实施例中,掺入碱基编辑器的腺苷脱氨酶包含作用于tRNA(ADAT)的全部或部分腺苷脱氨酶。包含腺苷脱氨酶结构域的碱基编辑器也能够使DNA多核苷酸的A核碱基脱氨基。在一个实施例中,碱基编辑器的腺苷脱氨酶结构域包含ADAT的全部或一部分,该ADAT包含一个或多个允许ADAT使DNA中的目标A脱氨基的突变。例如,所述碱基编辑器可包含来自大肠杆菌的ADAT(EcTadA)的全部或部分,其包含一个或多个以下突变:D108N、A106V、D147Y、E155V、L84F、H123Y、I156F,或另一种腺苷脱氨酶的相应突变。

[0648] 所述腺苷脱氨酶可以源自任何合适的生物体(例如大肠杆菌)。在一些实施例中,所述腺苷脱氨酶来自原核生物。在一些实施例中,所述腺苷脱氨酶来自细菌。在一些实施例中,所述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。在一些实施例中,所述腺苷脱氨酶来自大肠杆菌。在一些实施例中,所述腺嘌呤脱氨酶是天然存在的腺苷脱氨酶,其包括一个或多个对应于本文提供的任何突变(例如ecTadA中的突变)的突变。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测定来鉴定。因此,可以在任何天然存在的腺苷脱氨酶(例如与ecTadA具有同源性)中,产生对应于本文描述的任何突变(例如在ecTadA中鉴定的任何突变)的突变。

[0649] 腺苷脱氨酶

[0650] 在一些实施例中,本文所述的碱基编辑器可包含脱氨酶结构域,其包括腺苷脱氨酶。碱基编辑器的这种腺苷脱氨酶结构域可以通过将A脱氨基形成肌苷(I)来促进将腺嘌呤

(A)核碱基编辑为鸟嘌呤(G)核碱基,肌苷(I)具有G的碱基配对特性。腺苷脱氨酶能够使脱氧核糖核酸(DNA)中脱氧腺苷残基的腺嘌呤脱氨基(即除去胺基团)。

[0651] 在一些实施例中,本文提供的腺苷脱氨酶能够使腺嘌呤脱氨基。在一些实施例中,本文提供的腺苷脱氨酶能够使DNA脱氧腺苷残基中的腺嘌呤脱氨基。在一些实施例中,所述腺嘌呤脱氨酶是天然存在的腺苷脱氨酶,其包括一个或多个对应于本文提供的任何突变(例如ecTadA中的突变)的突变。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。因此,本领域技术人员将能够在任何天然存在的腺苷脱氨酶(例如与ecTadA具有同源性)中,产生对应于本文所述的任何突变(例如在ecTadA中鉴定的任何突变)的突变。在一些实施例中,所述腺苷脱氨酶来自原核生物。在一些实施例中,所述腺苷脱氨酶来自细菌。在一些实施例中,所述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。在一些实施例中,所述腺苷脱氨酶来自大肠杆菌。

[0652] 本发明提供了具有提高的效率(>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需碱基,并且不太可能编辑不打算改变的碱基(即“旁观者”)。

[0653] 在特定实施例中,TadA是PCT/US2017/045381(WO 2018/027078)中描述的TadA中的任一种,该专利通过引用整体并入本文。

[0654] 在一些实施例中,本发明的核碱基编辑器是包含以下序列改变的腺苷脱氨酶变体:

[0655] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD(也称为TadA*7.10)。

[0656] 在特定实施例中,所述融合蛋白包含单个(例如作为单体提供)TadA*8变体。在一些实施例中,TadA*8与Cas9切口酶连接。在一些实施例中,本发明所述的融合蛋白包含与TadA*8变体连接的野生型TadA(TadA(wt))的异源二聚体。在其他实施例中,本发明所述的融合蛋白包含与TadA*8变体连接的TadA*7.10的异源二聚体。在一些实施例中,所述碱基编辑器是包含TadA*8变体单体的ABE8。在一些实施例中,所述碱基编辑器是包含TadA*8变体和TadA(wt)的异源二聚体的ABE8。在一些实施例中,所述碱基编辑器是包含TadA*8变体和TadA*7.10的异源二聚体的ABE8。在一些实施例中,所述碱基编辑器是包含TadA*8变体的异源二聚体的ABE8。在一些实施例中,所述TadA*8变体选自表7。在一些实施例中,ABE8选自表7。相关序列如下:

[0657] 野生型TadA(TadA(wt))或“TadA参考序列”

[0658] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD(SEQ ID NO:2)

[0659] TadA*7.10:

[0660] MSEVEFSHEYW MRHALTLAKR ARDEREVPVGAVLVLNNRVI GEGWNRAIGL HDPTAHAEIM ALRQGGLVMQ NYRLIDATLY VTFEPCVMCA GAMIHSRIGR VVFGVRNAKT GAAGSLMDVL HYPGMNHRVE ITEGILADEC AALLCYFFRM PRQVFNAQKK AQSSTD

[0661] 在一些实施例中,所述腺苷脱氨酶包含氨基酸序列,其与本文提供的任何腺苷脱氨酶中列出的任何一个氨基酸序列至少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。应当理解,本文提供的腺苷脱氨酶可以包括一个或多个突变(例如本文提供的任何突变)。本公开提供具有特定百分比同一性的任何脱氨酶结构域,加上本文所述的任何突变或其组合。在一些实施例中,所述腺苷脱氨酶包含与参考序列或本文提供的任何腺苷脱氨酶相比,具有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、50或更多个突变的氨基酸序列。在一些实施例中,所述腺苷脱氨酶包含与本领域已知或本文所述的任一氨基酸序列相比,具有至少5个、至少10个、至少15个、至少20个、至少25个、至少30个、至少35个、至少40个、至少45个、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或至少有170个相同的连续氨基酸残基的氨基酸序列。

[0662] 在一些实施例中,TadA脱氨酶是全长大肠杆菌TadA脱氨酶。例如,在某些实施例中,所述腺苷脱氨酶包含氨基酸序列:

[0663] MRRAFITGVFFLSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD。

[0664] 然而,应当理解,可用于本申请的另外的腺苷脱氨酶对于技术人员是显而易见的并且在本公开的范围内。例如,所述腺苷脱氨酶可以是作用于tRNA的腺苷脱氨酶(ADAT)的同源物。非限制性地,示例性AD AT同源物的氨基酸序列包括以下:

[0665] 金黄色葡萄球菌TadA:

[0666] MGSHMTNDIYFMTLAIEEAKKAAQLGEVPIGAIITKDDEVIARAHNLRETLQQPTAHAEHIAIERAAKVLGSWRLEGCTLYVTLEPCVMCAGTIVMSRIPRVVYGADDPKGGCSGSLMNLLQQSNFNHRAIVDKGVLKEACSTLLTTFFKNLRANKKSTN

[0667] 枯草芽孢杆菌TadA:

[0668] MTQDELYMKEAIKEAKKAEEKGEVPIGAVLVINGEIIARAHNLRETEQRSIAHAEMLVIDEACKALGTWRLEGATLYVTLEPCPMCAGAVVLSRVEKVVFGAFDPKGGCSGTLMNLLQEERFNHQAEVVSGVLEEECGGMLSAFFRELRKKKKAARKNLSE

[0669] 鼠伤寒沙门氏菌(S.typhimurium)TadA:

[0670] MPPAFITGVTSLSDVELDHEYWMRHALTLAKRAWDEREVPVGAVLVHNHRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVLQNYRLLDTTLYVTLEPCVMCAGAMVHSRIGRVVFGARDAKTGAAGSLIDVLHHPGMNHRVEIIEGVLRDECATLLSDFFRMRRQEIKALKKADRAEGAGPAV

[0671] 腐败希瓦氏菌(S.putrefaciens)TadA:

[0672] MDEYWMQVAMQMAEKAEAAGEVPVGAVLVKDGQQIATGYNLSISQHDPTAHAEILCLRSAGKKLENYRLLDATLYITLEPCAMCAGAMVHSRIARVVYGARDEKTGAAGTVVNLLQHPAFNHQVEVTSGVLAEACSAQLSRFFKRRRDEKKALKLAQRAQQGIE

[0673] 流感嗜血杆菌F3031(H.influenzae)TadA:MDAAKVRSEFDEKMMRYALELADKAEALGEIPVGAVLVDDARNIIGEGWNLSIVQSDPTAHAEIIALRNGAKNIQNYRLLNSTLYVTLEPCTMCAGAILHSRIKRL

VFGASDYKTGAIGSRFHFFDDYKMNHTLEITSGVLAEECSQKLSTFFQKRREEKKIEKALLKSLSDK

[0674] 新月柄杆菌(C.crescentus)TadA:

[0675] MRTDESEDQDHRMMRLALDAARAAAEAGETPVGAVILDPSTGEVIATAGNGPIAAHDPTAHAEIAAMRA AAAKLGNYRLTDLTLVVTLEPCAMCAGAISHARIGRVVFGADDPKGGAVVHGPKFFAQPTCHWRPEVTGGVLADESA DLLRGFFRARRKAKI

[0676] 硫还原地杆菌(G.sulfurreducens)TadA:

[0677] MSSLKKTPIRDDAYWMGKAIREAAKAAARDEVPIGAVIVRDGAVIGRGHNLREGSNDPSAHAEMIAIRQ AARRSANWRLTGATLYVTLEPCLMCMGAIILARLERVVFGCYDPKGGAAGSLYDLSADPRLNHQVRLSPGVCQEECG TMLSDFFRDLRRRKKAKATPALFIDERKVPPEP

[0678] 大肠杆菌TadA(ecTadA)的一个实施例包括以下:MSEVEFSHEYWMRHALTLAKRARDERE VPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVF GVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

[0679] 在一些实施例中,所述腺苷脱氨酶来自原核生物。在一些实施例中,所述腺苷脱氨酶来自细菌。在一些实施例中,所述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。在一些实施例中,所述腺苷脱氨酶来自大肠杆菌。

[0680] 在一个实施例中,本发明的融合蛋白包含与TadA7.10连接的野生型TadA,其与Cas9切口酶连接。在特定实施例中,所述融合蛋白包含单个TadA7.10结构域(例如作为单体提供)。在其他实施例中,所述ABE7.10编辑器包含能够形成异源二聚体的TadA7.10和TadA(wt)。

[0681] 在一些实施例中,所述腺苷脱氨酶包含与本文提供的任何腺苷脱氨酶中列出的任何一个氨基酸序列至少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%相同的氨基酸序列。应当理解,本文提供的腺苷脱氨酶可以包括一个或多个突变(例如本文提供的任何突变)。本公开提供具有特定百分比同一性的任何脱氨酶结构域,加上本文所述的任何突变或其组合。在一些实施例中,腺苷脱氨酶包含与参考序列或本文提供的任何腺苷脱氨酶相比,具有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、50或更多个突变的氨基酸序列。在一些实施例中,所述腺苷脱氨酶包含与本领域已知或本文所述的任一氨基酸序列相比,具有至少5个、至少10个、至少15个、至少20个、至少25个、至少30个、至少35个、至少40个、至少45个、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或至少有170个相同的连续氨基酸残基的氨基酸序列。

[0682] 应当理解,本文提供的任何突变(例如基于TadA参考氨基酸序列)可以引入其他腺苷脱氨酶,例如大肠杆菌TadA(ecTadA)、金黄色葡萄球菌TadA(saTadA),或其他腺苷脱氨酶(如细菌腺苷脱氨酶)。对于本领域技术人员而言,如何鉴定相对于本文提供的TadA参考氨基酸序列与突变残基同源的序列将是显而易见的。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)中进行相对于TadA参考序列鉴定的任何突变。还应当理解,相对于TadA参考序列或另一种腺苷脱氨酶,本文提供的任何突变可以单独或以任何组合进

行。

[0683] 在一些实施例中,所述腺苷脱氨酶包含TadA参考序列中的D108X突变,或另一种腺苷脱氨酶(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之外的任何氨基酸。在一些实施例中,所述腺苷脱氨酶包含D108G、D108N、D108V、D108A或D108Y突变,或另一腺苷脱氨酶中的相应突变。

[0684] 在一些实施例中,所述腺苷脱氨酶包含TadA参考序列中的A106X突变,或另一种腺苷脱氨酶(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之外的任何氨基酸。在一些实施例中,所述腺苷脱氨酶包含TadA参考序列中的A106V突变,或另一种腺苷脱氨酶(例如野生型TadA或ecTadA)中的相应突变。

[0685] 在一些实施例中,所述腺苷脱氨酶包含TadA参考序列中的E155X突变,或另一种腺苷脱氨酶(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之外的任何氨基酸。在一些实施例中,腺苷脱氨酶包含TadA参考序列中的E155D、E155G或E155V突变,或另一种腺苷脱氨酶(例如ecTadA)中的相应突变。

[0686] 在一些实施例中,所述腺苷脱氨酶包含TadA参考序列中的D147X突变,或另一种腺苷脱氨酶(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之外的任何氨基酸。在一些实施例中,所述腺苷脱氨酶包含D147Y、TadA参考序列中的突变或另一种腺苷脱氨酶(例如ecTadA)中的相应突变。

[0687] 在一些实施例中,所述腺苷脱氨酶包含TadA参考序列中的A106X、E155X或D147X突变,或另一种腺苷脱氨酶(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之外的任何氨基酸。在一些实施例中,所述腺苷脱氨酶包含E155D、E155G或E155V突变。在一些实施例中,所述腺苷脱氨酶包含D147Y。

[0688] 例如,腺苷脱氨酶可含有相对于TadA参考序列的D108N、A106V、E155V和/或D147Y突变,或另一种腺苷脱氨酶(例如ecTadA)中的相应突变。在一些实施例中,所述腺苷脱氨酶包含TadA参考序列中的以下突变组(突变组由“;”分隔),或另一种腺苷脱氨酶(例如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;和D108N、A106V、E155V和D147Y。然而,应当理解,本文提供的相应突变的任何组合可以在腺苷脱氨酶(例如ecTadA)中进行。

[0689] 在一些实施例中,所述腺苷脱氨酶包含TadA参考序列中的以下一种或多种突变: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,或另一种腺苷脱氨酶(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之外的任何氨基酸。在一些实施例中,腺苷脱氨酶包含以下的一种或多种相对于TadA参考序列的突变:H8Y、T17S、L18E、W23L、L34S、W45L、R51H、A56E、or 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,或相对于另一种腺苷脱氨酶(例如ecTadA)的一个或多个相应突变。

[0690] 在一些实施例中,所述腺苷脱氨酶包含TadA参考序列中的一个或多个H8X、D108X

和/或N127X突变,或另一个腺苷脱氨酶(例如ecTadA)中的一个或多个相应突变,其中X表示存在任何氨基酸。在一些实施例中,所述腺苷脱氨酶包含一个或多个相对于TadA参考序列的H8Y、D108N和/或N127S突变,或一个或多个相对于另一种腺苷脱氨酶(例如ecTadA)的相应突变。

[0691] 在一些实施例中,腺苷脱氨酶包含以下的一种或多种相对于TadA参考序列的突变:H8X,R26X、M61X、L68X、M70X、A106X、D108X、A109X、N127X、D147X、R152X、Q154X、E155X、K161X、Q163X、和/或T166X,或相对于另一种腺苷脱氨酶(例如ecTadA)的一个或多个相应突变,其中X表示存在除野生型腺苷脱氨酶中的相应氨基酸之外的任何氨基酸。在一些实施例中,腺苷脱氨酶包含以下的一种或多种相对于TadA参考序列的突变:H8Y、R26W、M61I、L68Q、M70V、A106T、D108N、A109T、N127S、D147Y、R152C、Q154H或Q154R、E155G或E155V或E155D、K161Q、Q163H、和/或T166P,或相对于另一种腺苷脱氨酶(例如ecTadA)的一个或多个相应突变。

[0692] 在一些实施例中,腺苷脱氨酶相对于TadA参考序列,包含一个、两个、三个、四个、五个或六个选自由H8X、D108X、N127X、D147X、R152X和Q154X组成的组的突变,或相对于另一种腺苷脱氨酶的相应突变(例如ecTadA)的突变,其中X表示存在除野生型腺苷脱氨酶中的相应氨基酸之外的任何氨基酸。在一些实施例中,所述腺苷脱氨酶相对于TadA参考序列,包含一个、两个、三个、四个、五个、六个、七个或八个选自由H8X、M61X、M70X、D108X、N127X、Q154X、E155X和Q163X组成的组的突变,或相对于另一种腺苷脱氨酶(例如ecTadA)的相应突变,其中X表示存在除野生型腺苷脱氨酶中的相应氨基酸之外的任何氨基酸。在一些实施例中,所述腺苷脱氨酶相对于TadA参考序列,包含一个、两个、三个、四个或五个的选自由H8X、D108X、N127X、E155X和T166X组成的组的突变,或相对于另一种腺苷脱氨酶(例如ecTadA)的相应突变,其中X表示存在除野生型腺苷脱氨酶中的相应氨基酸之外的任何氨基酸。

[0693] 在一些实施例中,所述腺苷脱氨酶包含一个、两个、三个、四个、五个或六个选自由H8X、A106X、D108X的突变组成的组,相对于另一种腺苷脱氨酶的一个或多个突变,其中X表示存在任何除野生型腺苷脱氨酶中相应氨基酸以外的氨基酸。在一些实施例中,所述腺苷脱氨酶包含一个、两个、三个、四个、五个、六个、七个或八个选自由H8X、R26X、L68X、D108X、N127X、D147X和E155X组成的组的突变,或相应的突变或相对于另一种腺苷脱氨酶的突变,其中X表示存在除野生型腺苷脱氨酶中的相应氨基酸之外的任何氨基酸。在一些实施例中,所述腺苷脱氨酶相对于TadA参考序列,包含一个、两个、三个、四个或五个选自由H8X、D108X、A109X、N127X和E155X组成的组的突变,或相对于另一种腺苷脱氨酶(例如ecTadA)的相应突变,其中X表示存在除野生型腺苷脱氨酶中的相应氨基酸之外的任何氨基酸。

[0694] 在一些实施例中,所述腺苷脱氨酶相对于TadA参考序列包含一个、两个、三个、四个、五个或六个选自由H8Y、D108N、N127S、D147Y、R152C和Q154H组成的组的突变,或相对于另一种腺苷脱氨酶(例如ecTadA)的相应突变。在一些实施例中,所述腺苷脱氨酶相对于TadA参考序列包含一个、两个、三个、四个、五个、六个、七个或八个选自由H8Y、M61I、M70V、D108N、N127S、Q154R、E155G和Q163H组成的组的突变,或相对于另一种腺苷脱氨酶(例如ecTadA)的相应突变。在一些实施例中,所述腺苷脱氨酶包含一个、两个、三个、四个或五个相对于TadA参考序列的选自由H8Y、D108N、N127S、E155V和T166P组成的组的突变,或相对于另一种腺苷脱氨酶(例如ecTadA)的相应突变。在一些实施例中,所述腺苷脱氨酶相对于

TadA参考序列包含一个、两个、三个、四个、五个或六个选自由H8Y、A106T、D108N、N127S、E155D和K161Q组成的组的突变,,或相对于另一种腺苷脱氨酶(例如ecTadA)的相应突变。在一些实施例中,所述腺苷脱氨酶包含一个、两个、三个、四个、五个、六个、七个或八个突变,该突变选自H8Y、R26W、L68Q、D108N、N127S、D147Y和E155V,相对于TadA参考序列,或相对于另一种腺苷脱氨酶(例如ecTadA)的相应突变。在一些实施例中,所述腺苷脱氨酶包含一个、两个、三个、四个或五个相对于TadA参考序列的选自由H8Y、D108N、A109T、N127S和E155G组成的组的突变,或相对于另一种腺苷脱氨酶(例如ecTadA)的相应突变。

[0695] 本文提供的任何突变和任何额外的突变(例如基于ecTadA氨基酸序列)可以被引入到任何其他腺苷脱氨酶中。本文提供的任何突变可以单独或以任何组合在TadA参考序列或另一种腺苷脱氨酶(例如ecTadA)中进行。

[0696] A到G核碱基编辑蛋白的详细信息描述在国际PCT申请号PCT/2017/045381(WO2018/027078)和Gaudelli,NM等人,"Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage"Nature,551,464-471(2017),其全部内容通过引用并入本文。

[0697] 在一些实施例中,所述腺苷脱氨酶包含一个或多个相对于另一种腺苷脱氨酶(例如ecTadA)的相应突变。在一些实施例中,所述腺苷脱氨酶包含相对于TadA参考序列的D108N、D108G或D108V突变,或另一种腺苷脱氨酶(例如ecTadA)中的相应突变。在一些实施例中,所述腺苷脱氨酶包含相对于TadA参考序列的A106V和D108N突变,或相对于另一种腺苷脱氨酶(例如ecTadA)的相应突变。在一些实施例中,所述腺苷脱氨酶包含相对于TadA参考序列的R107C和D108N突变,或相对于另一种腺苷脱氨酶(例如ecTadA)的相应突变。在一些实施例中,所述腺苷脱氨酶包含相对于TadA参考序列的H8Y、D108N、N127S、D147Y和Q154H突变,或相对于另一种腺苷脱氨酶(例如ecTadA)的相应突变。在一些实施例中,所述腺苷脱氨酶包含相对于TadA参考序列的H8Y、D108N、N127S、D147Y和E155V突变,或相对于另一种腺苷脱氨酶(例如ecTadA)的相应突变。在一些实施例中,所述腺苷脱氨酶包含相对于TadA参考序列的D108N、D147Y和E155V突变,或相对于另一种腺苷脱氨酶(例如ecTadA)的相应突变。在一些实施例中,所述腺苷脱氨酶包含相对于TadA参考序列的H8Y、D108N和N127S突变,或相对于另一种腺苷脱氨酶(例如ecTadA)的相应突变。在一些实施例中,腺苷脱氨酶包含相对于TadA参考序列的A106V、D108N、D147Y和E155V突变,或相对于另一种腺苷脱氨酶(例如ecTadA)的相应突变。

[0698] 在一些实施例中,所述腺苷脱氨酶包含以下的一种或多种相对于TadA参考序列的突变:S2X、H8X、I49X、L84X、H123X、N127X、I156X和/或K160X,或相对于另一种腺苷脱氨酶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之外的任何氨基酸。在一些实施例中,所述腺苷脱氨酶包含以下的一种或多种相对于TadA参考序列的突变:S2A、H8Y、I49F、L84F、H123Y、N127S、I156F和/或K160S,或相对于另一种腺苷脱氨酶(例如ecTadA)的一个或多个相应突变。

[0699] 在一些实施例中,所述腺苷脱氨酶包含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之外的任何氨基酸。在一些实施例中,所述腺苷脱氨酶包含相对于TadA参考序列的L84F突变,或相对于另一种腺苷脱氨酶(例如ecTadA)的相应突变。

[0700] 在一些实施例中,所述腺苷脱氨酶包含相对于TadA参考序列的H123X突变,或相对

于另一种腺苷脱氨酶(例如ecTadA)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之外的任何氨基酸。在一些实施例中,所述腺苷脱氨酶包含相对于TadA参考序列的H123Y突变,或相对于另一种腺苷脱氨酶(例如ecTadA)的相应突变。

[0701] 在一些实施例中,所述腺苷脱氨酶包含相对于TadA参考序列的I156X突变,或相对于另一种腺苷脱氨酶(例如ecTadA)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之外的任何氨基酸。在一些实施例中,所述腺苷脱氨酶包含相对于TadA参考序列的I156F突变,或相对于另一种腺苷脱氨酶(例如ecTadA)的相应突变。

[0702] 在一些实施例中,所述腺苷脱氨酶相对于TadA参考序列包含一个、两个、三个、四个、五个、六个或七个选自由以下组成的组的突变:L84X、A106X、D108X、H123X、D147X、E155X、和I156X,或相对于另一种腺苷脱氨酶(例如ecTadA)的相应突变,其中X表示存在除野生型腺苷脱氨酶中的相应氨基酸之外的任何氨基酸。在一些实施例中,所述腺苷脱氨酶相对于TadA参考序列包含一个、两个、三个、四个、五个或六个选自由以下组成的组的突变:S2X、I49X、A106X,D108X、D147X和E155X,或相对于另一种腺苷脱氨酶(例如ecTadA)的相应突变,其中X表示存在除野生型腺苷脱氨酶中的相应氨基酸之外的任何氨基酸。在一些实施例中,所述腺苷脱氨酶包含一个、两个、三个、四个或五个相对于TadA参考序列的选自由H8X、A106X、D108X、N127X和K160X组成的组的突变,或相对于另一种腺苷脱氨酶(例如ecTadA)的相应突变,其中X表示存在除野生型腺苷脱氨酶中的相应氨基酸之外的任何氨基酸。

[0703] 在一些实施例中,所述腺苷脱氨酶包含一个、两个、三个、四个、五个、六个或七个相对于TadA参考序列的选自由L84F、A106V、D108N、H123Y、D147Y、E155V和I156F组成的组的突变,或相对于另一种腺苷脱氨酶(例如ecTadA)的相应突变。在一些实施例中,所述腺苷脱氨酶相对于TadA参考序列包含一个、两个、三个、四个、五个或六个选自由S2A、I49F、A106V、D108N、D147Y和E155V组成的组的突变,或相对于另一种腺苷脱氨酶的相应突变。

[0704] 在一些实施例中,所述腺苷脱氨酶包含一个、两个、三个、四个或五个相对于TadA参考序列的选自由H8Y、A106T、D108N、N127S和K160S组成的组的突变,或另一种腺苷脱氨酶(例如ecTadA)中的相应突变。

[0705] 在一些实施例中,所述腺苷脱氨酶包含相对于TadA参考序列的E25X、R26X、R107X、A142X和/或A143X突变中的一个或多个,或相对于另一种腺苷脱氨酶(例如ecTadA)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相应氨基酸之外的任何氨基酸。在一些实施例中,所述腺苷脱氨酶包含以下的一种或多种相对于TadA参考序列的突变: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,或相对于另一种腺苷脱氨酶(例如ecTadA)的一个或多个相应突变。在一些实施例中,所述腺苷脱氨酶包含一个或多个本文所述的对应于TadA参考序列的突变,或一个或多个相对于另一种腺苷脱氨酶(例如ecTadA)的对应突变。

[0706] 在一些实施例中,所述腺苷脱氨酶包含TadA参考序列中的E25X突变,或另一种腺苷脱氨酶(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之外的任何氨基酸。在一些实施例中,所述腺苷脱氨酶包含相对于TadA参考序列的E25M、E25D、E25A、E25R、E25V、E25S或E25Y突变,或相对于另一种腺苷脱氨酶(例如ecTadA)的相应

突变。

[0707] 在一些实施例中,所述腺苷脱氨酶包含相对于TadA参考序列的R26X突变,或相对于另一种腺苷脱氨酶(例如ecTadA)的相应突变,其中X表示除野生型中的相应氨基酸之外的任何氨基酸腺苷脱氨酶。在一些实施例中,所述腺苷脱氨酶包含相对于TadA参考序列的R26G、R26N、R26Q、R26C、R26L或R26K突变,或相对于另一种腺苷脱氨酶(例如ecTadA)的相应突变。

[0708] 在一些实施例中,所述腺苷脱氨酶包含相对于TadA参考序列的R107X突变,或相对于另一种腺苷脱氨酶(例如ecTadA)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之外的任何氨基酸。在一些实施例中,所述腺苷脱氨酶包含相对于TadA参考序列的R107P、R107K、R107A、R107N、R107W、R107H或R107S突变,或相对于另一种腺苷脱氨酶(例如ecTadA)的相应突变。

[0709] 在一些实施例中,所述腺苷脱氨酶包含TadA参考序列中的A142X突变,或另一种腺苷脱氨酶(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之外的任何氨基酸。在一些实施例中,所述腺苷脱氨酶包含TadA参考序列中的A142N、A142D、A142G突变,或另一种腺苷脱氨酶(例如ecTadA)中的相应突变。

[0710] 在一些实施例中,所述腺苷脱氨酶包含相对于TadA参考序列的A143X突变,或相对于另一种腺苷脱氨酶(例如ecTadA)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之外的任何氨基酸。在一些实施例中,所述腺苷脱氨酶包含相对于TadA参考序列的A143D、A143G、A143E、A143L、A143W、A143M、A143S、A143Q和/或A143R突变,或相对于另一种腺苷脱氨酶(例如TadaA)的相应突变.

[0711] 在一些实施例中,所述腺苷脱氨酶包含以下的一种或多种相对于TadA参考序列的突变:H36X、N37X、P48X、I49X、R51X、M70X、N72X、D77X、E134X、S146X、Q154X、K157X、和/或K161X,或相对于另一种腺苷脱氨酶(例如ecTadA)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之外的任何氨基酸。在一些实施例中,所述腺苷脱氨酶包含以下的一种或多种相对于TadA参考序列的突变:H36L、N37T、N37S、P48T、P48L、I49V、R51H、R51L、M70L、N72S、D77G、E134G、S146R、S146C、Q154H、K157N、和/或K161T,或相对于另一种腺苷脱氨酶(例如ecTadA)的一个或多个相应突变。

[0712] 在一些实施例中,所述腺苷脱氨酶包含相对于TadA参考序列的H36X突变,或相对于另一种腺苷脱氨酶(例如ecTadA)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之外的任何氨基酸。在一些实施例中,所述腺苷脱氨酶包含相对于TadA参考序列的H36L突变,或相对于另一种腺苷脱氨酶(例如ecTadA)的相应突变。

[0713] 在一些实施例中,所述腺苷脱氨酶包含相对于TadA参考序列的N37X突变,或相对于另一种腺苷脱氨酶(例如ecTadA)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之外的任何氨基酸。在一些实施例中,所述腺苷脱氨酶包含相对于TadA参考序列的N37T或N37S突变,或相对于另一种腺苷脱氨酶(例如ecTadA)的相应突变。

[0714] 在一些实施例中,所述腺苷脱氨酶包含相对于TadA参考序列的P48X突变,或相对于另一种腺苷脱氨酶(例如ecTadA)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之外的任何氨基酸。在一些实施例中,所述腺苷脱氨酶包含相对于TadA参考序列的P48T或P48L突变,或相对于另一种腺苷脱氨酶(例如ecTadA)的相应突变。

[0715] 在一些实施例中,所述腺苷脱氨酶包含相对于TadA参考序列的R51X突变,或相对于另一种腺苷脱氨酶的相应突变,其中X表示除野生型腺苷脱氨酶中的相应氨基酸之外的任何氨基酸。在一些实施例中,所述腺苷脱氨酶包含相对于TadA参考序列的R51H或R51L突变,或相对于另一种腺苷脱氨酶(例如ecTadA)的相应突变。

[0716] 在一些实施例中,所述腺苷脱氨酶包含相对于TadA参考序列的S146X突变,或相对于另一种腺苷脱氨酶(例如ecTadA)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之外的任何氨基酸。在一些实施例中,所述腺苷脱氨酶包含相对于TadA参考序列的S146R或S146C突变,或相对于另一种腺苷脱氨酶(例如ecTadA)的相应突变。

[0717] 在一些实施例中,所述腺苷脱氨酶包含相对于TadA参考序列的K157X突变,或相对于另一种腺苷脱氨酶(例如ecTadA)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之外的任何氨基酸。在一些实施例中,所述腺苷脱氨酶包含相对于TadA参考序列的K157N突变,或相对于另一种腺苷脱氨酶(例如ecTadA)的相应突变。

[0718] 在一些实施例中,所述腺苷脱氨酶包含相对于TadA参考序列的P48X突变,或相对于另一种腺苷脱氨酶(例如ecTadA)的相应突变,其中X表示除野生型腺苷脱氨酶中相应氨基酸之外的任何氨基酸。在一些实施例中,所述腺苷脱氨酶包含相对于TadA参考序列的P48S、P48T或P48A突变,或相对于另一种腺苷脱氨酶(例如ecTadA)的相应突变。

[0719] 在一些实施例中,所述腺苷脱氨酶包含相对于TadA参考序列的A142X突变,或相对于另一种腺苷脱氨酶(例如ecTadA)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之外的任何氨基酸。在一些实施例中,所述腺苷脱氨酶包含相对于TadA参考序列的A142N突变,或相对于另一种腺苷脱氨酶(例如ecTadA)的相应突变。

[0720] 在一些实施例中,所述腺苷脱氨酶包含相对于TadA参考序列的W23X突变,或相对于另一种腺苷脱氨酶(例如ecTadA)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之外的任何氨基酸。在一些实施例中,所述腺苷脱氨酶包含相对于TadA参考序列的W23R或W23L突变,或相对于另一种腺苷脱氨酶(例如ecTadA)的相应突变。

[0721] 在一些实施例中,所述腺苷脱氨酶包含相对于TadA参考序列的R152X突变,或相对于另一种腺苷脱氨酶(例如ecTadA)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之外的任何氨基酸。在一些实施例中,所述腺苷脱氨酶包含相对于TadA参考序列的R152P或R52H突变,或相对于另一种腺苷脱氨酶(例如ecTadA)的相应突变。

[0722] 在一个实施例中,所述腺苷脱氨酶可以包含相对于TadA参考序列的以下突变:H36L、R51L、L84F、A106V、D108N、H123Y、S146C、D147Y、E155V、I156F和K157N,或相对于另一种腺苷脱氨酶的相应突变。在一些实施例中,所述腺苷脱氨酶包含以下相对于TadA参考序列的突变组合,其中组合的每个突变由“_”分隔并且每个突变组合在括号之间:

[0723] (A106V_D108N),

[0724] (R107C_D108N),

[0725] (H8Y_D108N_N127S_D147Y_Q154H),

[0726] (H8Y_D108N_N127S_D147Y_E155V),

[0727] (D108N_D147Y_E155V),

[0728] (H8Y_D108N_N127S),

[0729] (H8Y_D108N_N127S_D147Y_Q154H),

[0730] (A106V_D108N_D147Y_E155V),

[0731] (D108Q_D147Y_E155V),

[0732] (D108M_D147Y_E155V),

[0733] (D108L_D147Y_E155V),

[0734] (D108K_D147Y_E155V),

[0735] (D108I_D147Y_E155V),

[0736] (D108F_D147Y_E155V),

[0737] (A106V_D108N_D147Y),

[0738] (A106V_D108M_D147Y_E155V),

[0739] (E59A_A106V_D108N_D147Y_E155V),

[0740] (E59A cat dead_A106V_D108N_D147Y_E155V),

[0741] (L84F_A106V_D108N_H123Y_D147Y_E155V_I156Y),

[0742] (L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),

[0743] (R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F),

[0744] (E25G_R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F),(E25D_R26G_L84F_A106V_R107K_D108N_H123Y_A142N_A143G_D147Y_E155V_I156F),

[0745] (R26Q_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F),

[0746] (E25M_R26G_L84F_A106V_R107P_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F),

[0747] (R26C_L84F_A106V_R107H_D108N_H123Y_A142N_D147Y_E155V_I156F),(L84F_A106V_D108N_H123Y_A142N_A143L_D147Y_E155V_I156F),

[0748] (R26G_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F),

[0749] (E25A_R26G_L84F_A106V_R107N_D108N_H123Y_A142N_A143E_D147Y_E155V_I156F),

[0750] (R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F),

[0751] (A106V_D108N_A142N_D147Y_E155V),

[0752] (R26G_A106V_D108N_A142N_D147Y_E155V),

[0753] (E25D_R26G_A106V_R107K_D108N_A142N_A143G_D147Y_E155V),

[0754] (R26G_A106V_D108N_R107H_A142N_A143D_D147Y_E155V),

[0755] (E25D_R26G_A106V_D108N_A142N_D147Y_E155V),

[0756] (A106V_R107K_D108N_A142N_D147Y_E155V),

[0757] (A106V_D108N_A142N_A143G_D147Y_E155V),

[0758] (A106V_D108N_A142N_A143L_D147Y_E155V),

[0759] (H36L_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N),

[0760] (N37T_P48T_M70L_L84F_A106V_D108N_H123Y_D147Y_I49V_E155V_I156F),

[0761] (N37S_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K161T),

[0762] (H36L_L84F_A106V_D108N_H123Y_D147Y_Q154H_E155V_I156F),

[0763] (N72S_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F),

[0764] (H36L_P48L_L84F_A106V_D108N_H123Y_E134G_D147Y_E155V_I156F),

[0765] (H36L_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K157N),(H36L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F),

[0766] (L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F_K161T),

[0767] (N37S_R51H_D77G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),

[0768] (R51L_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K157N),

[0769] (D24G_Q71R_L84F_H96L_A106V_D108N_H123Y_D147Y_E155V_I156F_K160E),

[0770] (H36L_G67V_L84F_A106V_D108N_H123Y_S146T_D147Y_E155V_I156F),

[0771] (Q71L_L84F_A106V_D108N_H123Y_L137M_A143E_D147Y_E155V_I156F),

[0772] (E25G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_Q159L),

[0773] (L84F_A91T_F104I_A106V_D108N_H123Y_D147Y_E155V_I156F),

[0774] (N72D_L84F_A106V_D108N_H123Y_G125A_D147Y_E155V_I156F),

[0775] (P48S_L84F_S97C_A106V_D108N_H123Y_D147Y_E155V_I156F),

[0776] (W23G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),

[0777] (D24G_P48L_Q71R_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_Q159L),

[0778] (L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F),

[0779] (H36L_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F_K157N),(N37S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_K161T),

[0780] (L84F_A106V_D108N_D147Y_E155V_I156F),

[0781] (R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K161T),

[0782] (L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K161T),

[0783] (L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K160E_K161T),

[0784] (L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K160E),

[0785] (R74Q_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),

[0786] (R74A_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),

[0787] (L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),

[0788] (R74Q_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),

[0789] (L84F_R98Q_A106V_D108N_H123Y_D147Y_E155V_I156F),

[0790] (L84F_A106V_D108N_H123Y_R129Q_D147Y_E155V_I156F),

[0791] (P48S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F),

[0792] (P48S_A142N),

[0793] (P48T_I49V_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_L157N),

[0794] (P48T_I49V_A142N),

[0795] (H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N),

[0796] (H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S146C_A142N_D147Y_E155V_I156F

[0797] (H36L_P48T_I49V_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N),

[0798] (H36L_P48T_I49V_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_

I156F_K157N),

[0799] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N),

[0800] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F_K157N),

[0801] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_A142N_D147Y_E155V_I156F_K157N),

[0802] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N),

[0803] (W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N),

[0804] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F_K161T),

[0805] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152H_E155V_I156F_K157N),

[0806] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F_K157N),

[0807] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F_K157N),

[0808] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142A_S146C_D147Y_E155V_I156F_K157N),

[0809] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142A_S146C_D147Y_R152P_E155V_I156F_K157N),

[0810] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F_K161T),

[0811] (W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F_K157N),

[0812] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_R152P_E155V_I156F_K157N)。

[0813] 在某些实施例中,本文提供的融合蛋白包含一种或多种改善融合蛋白的碱基编辑活性的特征。例如,本文提供的任何融合蛋白可以包含具有降低的核酸酶活性的Cas9结构域。在一些实施例中,本文提供的任何融合蛋白可具有不具核酸酶活性的Cas9结构域(dCas9),或切割双链DNA分子的一条链的Cas9结构域,称为Cas9切口酶(nCas9)。

[0814] 在一些实施例中,腺苷脱氨酶是TadA*7.10。在一些实施例中,TadA*7.10包括至少一种改变。在特定实施例中,TadA*7.10包括以下变化中的一种或多种:Y147T、Y147R、Q154S、Y123H、V82S、T166R和Q154R。改变Y123H在本文中也称为H123H(TadA*7.10中的改变H123Y回复回Y123H(wt))。在其他实施例中,TadA*7.10包括选自下组的改变的组合: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;和I76Y+V82S+Y123H+Y147R+Q154R。在特定的实施例中,腺苷脱氨酶变体包含从残基149、150、151、152、153、154、155、156和157处开始的C末端的缺失,相对于TadA*7.10、TadA参考序列,或另一个TadA中的相应突变。

[0815] 在其他实施例中,本发明的碱基编辑器是包含腺苷脱氨酶变体(例如TadA*8)的单体,所述变体包含一个或多个以下改变:Y147T、Y147R、Q154S、Y123H、V82S、T166R和/或Q154R,相对于TadA*7.10、TadA参考序列或另一个TadA中的相应突变。在其他实施例中,腺苷脱氨酶变体(TadA*8)是包含选自下组的改变组合的单体: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,相对于TadA*7.10、TadA参考序列或另一个TadA中的相应突变。在其他实施例中,碱基编辑器是包含野生型腺苷脱氨酶和腺苷脱氨酶变体(例如TadA*8)的异源二聚体,其包含一个或多个以下改变:Y147T、Y147R、Q154S、Y123H、V82S、T166R和/或Q154R,相对于TadA*7.10、TadA参考序列或另一个TadA中的相应突变。在其他实施例中,碱基编辑器是包含TadA*7.10结构域和腺苷脱氨酶变体结构域(例如TadA*8)的异源二聚体,其包含选自下组的改变的组合: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,相对于TadA*7.10、TadA参考序列或另一个TadA中的相应突变。

[0816] 在一个实施例中,腺苷脱氨酶是TadA*8,其包含或基本上由以下具有腺苷脱氨酶活性的序列或其片段组成:

[0817] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDV

[0818] LHYPGMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD

[0819] 在一些实施例中,所述TadA*8是截短的。在一些实施例中,相对于全长TadA*8,所述截短的TadA*8缺失1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个N-末端氨基酸残基。在一些实施例中,相对于全长TadA*8,所述截短的TadA*8缺失1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个C-末端氨基酸残基。在一些实施例中,所述腺苷脱氨酶变体是全长TadA*8。

[0820] 在一些实施例中,所述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、TadA*8.24。

[0821] 在一个实施例中,本发明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,其与Cas9切口酶连接。在特定实施例中,所述融合蛋白包含单个TadA*8结构域(例如作为单体提供)。在其他实施例中,所述碱基编辑器包含能够形成异源二聚体的TadA*8和TadA(wt)。示例性序列如下:

[0822] TadA(wt):

[0823] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD

[0824] TadA*7.10:

[0825] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

[0826] TadA*8:

[0827] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD。

[0828] 在一些实施例中,所述腺苷脱氨酶包含与本文提供的任何腺苷脱氨酶中列出的任何一个氨基酸序列至少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%相同的氨基酸序列。应当理解,本文提供的腺苷脱氨酶可以包括一个或多个突变(例如本文提供的任何突变)。本公开提供具有一定百分比同一性的任何脱氨酶结构域,加上本文所述的任何突变或其组合。在一些实施例中,与参考序列或本文提供的任何腺苷脱氨酶相比,所述腺苷脱氨酶包含具有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、50个或更多突变的氨基酸序列。在一些实施例中,与本领域已知或本文所述的任一氨基酸序列相比,所述腺苷脱氨酶包含具有至少5个、至少10个、至少15个、至少20个、至少25个、至少30个、至少35个、至少40个、至少45个、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的氨基酸序列。

[0829] 在特定实施例中,TadA*8在以下以粗体显示的任何位置处包含一个或多个突变。在其他实施例中,TadA*8在以下划线所示的任何位置处包含一个或多个突变:

[0830]
MSEVEFSHEY WMRHALTLAK RARDEREVPV GAVLVLNNRV IGEGWNRAIG 50
LHDPTAHAEI MALRQGGLVM QN**Y**RLIDATL Y**V**TFEPCVMC AGAMIHSRIG 100
RVVFGVRNAK TGAAGSLMDV LH**Y**PGMNHRV EITEGILADE CAALLC**Y**FFR 150
MPR**Q**VFNAQK KAQSS**T**D

[0831] 例如TadA*8包含单独或与以下Y147T、Y147R、Q154S、Y123H和/或Q154R中的任何一种或多种组合的氨基酸位置82和/或166处(例如V82S、T166R)的改变,相对于TadA*7.10、TadA参考序列或另一个TadA中的相应突变。在特定实施例中,改变的组合选自: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,相对于TadA*7.10、TadA参考序列或另一

个TadA中的相应突变。

[0832] 在一些实施例中,所述腺苷脱氨酶是TadA*8,其包含或基本上由以下具有腺苷脱氨酶活性的序列或其片段组成:

[0833] MSEVEFSHEY WMRHALTLAK RARDEREVPV GAVLVLNNRV IGEGWNRAIG LHDPTAHAEI MALRQGGLVM QNYRLIDATL YVTFEPCVMC AGAMIHSRIG RVVFGVRNAK TGAAGSLMDV LHYPGMNHRV EITEGILADE CAALLCTFFR MPRQVFNAQK KAQSSTD

[0834] 在一些实施例中,所述TadA*8被截短。在一些实施例中,相对于全长TadA*8,所述截短的TadA*8缺失1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个N-末端氨基酸残基。在一些实施例中,相对于全长TadA*8,所述截短的TadA*8缺失1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个C-末端氨基酸残基。在一些实施例中,所述腺苷脱氨酶变体是全长TadA*8。

[0835] 在一个实施例中,本发明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,其与Cas9切口酶连接。在特定实施例中,所述融合蛋白包含单个TadA*8结构域(例如作为单体提供)。在其他实施方式中,所述碱基编辑器包含能够形成异源二聚体的TadA*8和TadA(wt)。

[0836] 附加结构域

[0837] 本文所述的碱基编辑器可包括有助于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。在一些实施例中,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)和一个或多个附加结构域。在一些实施例中,所述附加结构域可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,或者是可能干扰所需碱基编辑结果的细胞机制(例如酶)的抑制剂。在一些实施例中,碱基编辑器可包含核酸酶、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、转录激活物或转录抑制结构域。

[0838] 在一些实施例中,碱基编辑器可包含尿嘧啶糖基化酶抑制剂(UGI)结构域。在一些实施例中,对U:G异源双链体DNA的存在的细胞DNA修复反应可能是细胞中核碱基编辑效率降低的原因。在此类实施例中,尿嘧啶DNA糖基化酶(UDG)可催化从细胞中的DNA去除U,这可启动碱基切除修复(BER),主要导致U:G对回复为C:G对。在这样的实施例中,可以在碱基编辑器中抑制BER,所述碱基编辑器包含结合单链、阻断编辑碱基、抑制UGI、抑制BER、保护编辑碱基和/或促进非编辑链修复的一个或多个结构域。因此,本公开内容考虑包含UGI结构域的碱基编辑器融合蛋白。

[0839] 在一些实施例中,碱基编辑器包含双链断裂(DSB)结合蛋白的全部或部分作为结构域。例如,DSB结合蛋白可以包括噬菌体Mu的Gam蛋白,其可以结合DSB的末端并且可以保护它们免于降解。见Komor,A.C.等人,“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),其全部内容通过引用并入本文。

[0840] 另外,在一些实施例中,Gam蛋白可以融合到碱基编辑器的N-末端。在一些实施例中,Gam蛋白可以融合到碱基编辑器的C-末端。噬菌体Mu的Gam蛋白可以与双链断裂(DSB)的末端结合并保护它们免于降解。在一些实施例中,使用Gam结合所述DSB的自由端可以减少

碱基编辑过程中插入缺失的形成。在一些实施例中,174个残基的Gam蛋白融合到碱基编辑器的N-末端。见Komor,A.C.等人,“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)。在一些实施例中,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删除至少一个氨基酸可以减少碱基编辑器的长度。在另一种情况下,一个或多个突变不会改变结构域相对于野生型结构域的长度。例如,任何结构域中的替换都不会/不会改变碱基编辑器的长度。

[0841] 在一些实施例中,碱基编辑器可包含核酸聚合酶(NAP)的全部或部分作为结构域。例如,碱基编辑器可以包含真核NAP的全部或一部分。在一些实施例中,整合到碱基编辑器中的NAP或其部分是DNA聚合酶。在一些实施例中,整合到碱基编辑器中的NAP或其部分具有转移损伤聚合酶活性。在一些实施例中,整合到碱基编辑器中的NAP或其部分是易位DNA聚合酶。在一些实施例中,整合到碱基编辑器中的NAP或其部分是Rev7、Rev1复合物、聚合酶iota、聚合酶kappa或聚合酶eta。在一些实施例中,整合到碱基编辑器中的NAP或其部分是真核聚合酶alpha、beta、gamma、epsilon、gamma、eta、iota、kappa、lambda、mu或nu组分。在一些实施例中,并入碱基编辑器中的NAP或其部分包含与核酸聚合酶(例如易位DNA聚合酶)至少75%、80%、85%、90%、95%、96%、97%、98%、99%或99.5%相同的氨基酸序列。

[0842] 碱基编辑系统

[0843] 本文提供的碱基编辑器系统的使用包括以下步骤:(a)将个体的多核苷酸(例如双链或单链DNA或RNA)的目标核苷酸序列与碱基编辑器系统接触,所述碱基编辑器系统包含核碱基编辑器(例如腺苷碱基编辑器)和向导多核酸(例如gRNA),其中目标核苷酸序列包含目标核碱基对;(b)诱导所述目标区域的链分离;(c)将目标区域的单链中的所述目标核碱基对的第一个核碱基转化为第二个核碱基;(d)切割不超过所述目标区域的一条链,其中与第一核碱基互补的第三核碱基被与第二核碱基互补的第四核碱基替换。应当理解,在一些实施例中,省略了步骤(b)。在一些实施例中,所述靶向核碱基对是一种或多种基因中的多个核碱基对。在一些实施例中,本文提供的碱基编辑器系统能够多重编辑一个或多个基因中的多个核碱基对。在一些实施例中,所述多个核碱基对位于同一基因中。在一些实施例中,所述多个核碱基对位于一个或多个基因中,其中至少一个基因位于不同的基因座中。

[0844] 在一些实施例中,所述切割的单链(切口链)与向导核酸杂交。在一些实施例中,所述切割的单链与包含第一核碱基的链相反。在一些实施例中,所述碱基编辑器包含Cas9结构域。在一些实施例中,第一个碱基是腺嘌呤,而第二个碱基不是G、C、A或T。在一些实施例中,第二个碱基是肌苷。

[0845] 本文提供的碱基编辑系统提供了一种新的基因组编辑方法,该方法使用一种包含催化缺陷化脓性链球菌Cas9、腺苷脱氨酶和碱基切除修复抑制剂的融合蛋白来诱导DNA中可编程的单核苷酸(C→T或A→G)变化,而不会产生双链DNA断裂,不需要供体DNA模板,也不会引起过量的随机插入和缺失。

[0846] 本文提供用于使用碱基编辑器系统编辑核碱基的系统、组成物和方法。在一些实施例中,所述碱基编辑器系统包含(1)碱基编辑器(BE),其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);(2)与多核苷酸可编

程核苷酸结合结构域结合的向导多核苷酸(例如向导RNA)。在一些实施例中,所述碱基编辑器系统包含腺苷碱基编辑器(ABE)。在一些实施例中,所述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。在一些实施例中,所述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。在一些实施例中,所述核碱基编辑结构域是脱氨酶结构域。在一些实施例中,所述脱氨酶结构域可以是腺嘌呤脱氨酶或腺苷脱氨酶。在一些实施例中,所述腺苷碱基编辑器可使DNA中的腺嘌呤脱氨基。在一些实施例中,ABE包括进化的TadA变体。

[0847] 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)中有所描述,它们各自通过引用整体并入本文。另见Komor,A.C.等人,"Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage"Nature 533,420-424(2016);Gaudelli,N.M.等人,"Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage"Nature 551,464-471(2017);和Komor,A.C.等人,"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),其全部内容通过引用并入本文。

[0848] 在一些实施例中,可利用单向导多核苷酸将脱氨酶靶向目标核酸序列。在一些实施例中,单对向导多核苷酸可用于将不同脱氨酶靶向目标核酸序列。

[0849] 所述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共价或非共价地彼此缔合。例如,在一些实施例中,所述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靶向目标核苷酸序列。在一些实施例中,多核苷酸可编程核苷酸结合结构域可融合或连接至脱氨酶结构域。在一些实施例中,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相互作用或缔合,将脱氨酶结构域靶向目标核苷酸序列。例如,在一些实施例中,所述核碱基编辑组分,例如脱氨酶组分,可包含额外异源部分或结构域,其能够与作为多核苷酸可编程核苷酸结合结构域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。在一些实施例中,所述额外异源部分可能能够与多肽结合、相互作用、缔合或形成复合物。在一些实施例中,所述额外异源部分可能能够与多核苷酸结合、相互作用、缔合或形成复合物。在一些实施例中,所述额外异源部分可能能够结合向导多核苷酸。在一些实施例中,所述额外异源部分可能能够结合多肽接头。在一些实施例中,所述额外异源部分能够结合多核苷酸接头。所述额外异源部分可以是蛋白质结构域。在一些实施例中,所述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 Com外壳蛋白结构域、无菌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。

[0850] 碱基编辑器系统可进一步包含向导多核苷酸组分。应当理解,所述碱基编辑器系统的组件分可以通过共价键、非共价相互作用、或其缔合和相互作用的任何组合彼此缔合。在一些实施例中,脱氨酶结构域可以通过向导多核苷酸靶向目标核苷酸序列。例如,在一些实施例中,所述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,如RNA或DNA结合蛋白),其能够与向导多核苷酸的部分或区段(例如多核苷酸基序)相互作用、缔合或能够与其形成复合物。在一些实施例中,

所述额外异源部分或结构域(例如多核苷酸结合结构域如RNA或DNA结合蛋白)可以与脱氨酶结构域融合或连接。在一些实施例中,所述额外异源部分可能能够与多肽结合、相互作用、缔合或形成复合物。在一些实施例中,所述额外异源部分可能能够与多核苷酸结合、相互作用、缔合或形成复合物。在一些实施例中,所述额外异源部分可能能够结合向导多核苷酸。在一些实施例中,所述额外异源部分可能能够结合多肽接头。在一些实施例中,所述额外异源部分能够结合多核苷酸接头。所述额外异源部分可以是蛋白质结构域。在一些实施例中,所述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 Com外壳蛋白结构域、无菌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。

[0851] 在一些实施例中,碱基编辑器系统可进一步包含碱基切除修复(BER)组分的抑制剂。应当理解,所述碱基编辑器系统的组分可以通过共价键、非共价相互作用或其缔合和相互作用的任何组合彼此缔合。所述BER组分的抑制剂可以包括碱基切除修复抑制剂。在一些实施例中,所述碱基切除修复抑制剂可以是尿嘧啶DNA糖基化酶抑制剂(UGI)。在一些实施例中,所述碱基切除修复抑制剂可以是肌苷碱基切除修复抑制剂。在一些实施例中,所述碱基切除修复抑制剂可以通过多核苷酸可编程核苷酸结合结构域靶向目标核苷酸序列。在一些实施例中,多核苷酸可编程核苷酸结合结构域可以与碱基切除修复抑制剂融合或连接。在一些实施例中,多核苷酸可编程核苷酸结合结构域可以与脱氨酶结构域和碱基切除修复抑制剂融合或连接。在一些实施例中,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抑制剂非共价相互作用或缔合,将碱基切除修复抑制剂靶向目标核苷酸序列。例如,在一些实施例中,所述碱基切除修复抑制剂组分可包含额外异源部分或结构域,该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。在一些实施例中,所述碱基切除修复抑制剂可以通过向导多核苷酸靶向目标核苷酸序列。例如,在一些实施例中,所述碱基切除修复抑制剂可以包含额外异源部分或结构域,其能够与向导多核苷酸的部分或区段(例如多核苷酸基序)相互作用、缔合或能够形成复合物。在一些实施例中,所述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,如RNA或DNA结合蛋白)可以与所述碱基切除修复抑制剂融合或连接。在一些实施例中,所述额外异源部分可能能够与多核苷酸结合、相互作用、缔合或形成复合物。在一些实施例中,所述额外异源部分可能能够结合向导多核苷酸。在一些实施例中,所述额外异源部分可能能够结合多肽接头。在一些实施例中,所述额外异源部分能够结合多核苷酸接头。所述额外的异源部分可以是蛋白质结构域。在一些实施例中,所述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 Com外壳蛋白结构域、无菌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。

[0852] 在一些实施例中,所述碱基编辑器抑制编辑链的碱基切除修复(BER)。在一些实施例中,所述碱基编辑器保护或结合非编辑链。在一些实施例中,所述碱基编辑器包含UGI活性。在一些实施例中,所述碱基编辑器包含无催化活性的肌苷特异性核酸酶。在一些实施例中,所述碱基编辑器包含切口酶活性。在一些实施例中,碱基对的预期编辑在PAM位点的上游。在一些实施例中,碱基对的预期编辑是PAM位点上游的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。在一些实施例中,碱基对的预期编辑在PAM位点的下

游。在一些实施例中,预期的编辑碱基对是PAM位点下游的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。

[0853] 在一些实施例中,该方法不需要规范的(例如NGG)PAM位点。在一些实施例中,所述核碱基编辑器包含接头或间隔区。在一些实施例中,所述接头或间隔物的长度为1-25个氨基酸。在一些实施例中,所述接头或间隔物的长度为5-20个氨基酸。在一些实施例中,所述接头或间隔区的长度为10、11、12、13、14、15、16、17、18、19或20个氨基酸。

[0854] 在一些实施例中,本文提供的碱基编辑融合蛋白需要位于精确位置,例如,目标碱基位于限定区域(例如"脱氨基窗口")内的位置。在一些实施例中,目标可以在4个碱基区域内。在一些实施例中,这种限定的目标区域可以是PAM上游的大约15个碱基。见Komor,A.C.等人,"Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage"Nature 533,420-424(2016);Gaudelli,N.M.等人,"Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage"Nature 551,464-471(2017);和Komor,A.C.等人,"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),其全部内容通过引用并入本文。

[0855] 在一些实施例中,所述目标区域包括目标窗口,其中目标窗口包括目标核碱基对。在一些实施例中,所述目标窗口包含1-10个核苷酸。在一些实施例中,所述目标窗口长度是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。在一些实施例中,碱基对的预期编辑在所述目标窗口内。在一些实施例中,所述目标窗口包括碱基对的预期编辑。在一些实施例中,使用本文提供的任何碱基编辑器来执行该方法。在一些实施例中,所述目标窗口是脱氨基窗口。脱氨基窗口可以是碱基编辑器作用于目标核苷酸并使其脱氨基的限定区域。在一些实施例中,所述脱氨基窗口在2、3、4、5、6、7、8、9或10个碱基区域内。在一些实施例中,所述脱氨基窗为PAM上游的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。

[0856] 本公开的碱基编辑器可以包含有助于编辑目标多核苷酸序列的任何结构域、特征或氨基酸序列。例如,在一些实施例中,所述碱基编辑器包含核定位序列(NLS)。在一些实施例中,所述碱基编辑器的NLS位于脱氨酶结构域和多核苷酸可编程核苷酸结合结构域之间。在一些实施例中,所述碱基编辑器的NLS位于多核苷酸可编程核苷酸结合结构域的C-末端。

[0857] 可存在于本文公开的碱基编辑器中的其他示例性特征是定位序列,例如细胞质定位序列、输出序列,例如核输出序列或其他定位序列,以及可用于增溶、纯化的序列标签,或融合蛋白的检测。本文提供的合适的蛋白质标签包括但不限于生物素羧化酶载体蛋白(BCCP)标签、myc标签、钙调蛋白标签、FLAG标签、血凝素(HA)标签、多组氨酸标签,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、Softags(例如Softag 1、Softag 3)、链标签、生物素连接酶标签、Flash标签、V5标签和SBP标签。其他合适的序列对本领域技术人员来说是显而易见的。在一些实施例中,所述融合蛋白包含一个或多个His标签。

[0858] 可包含在融合蛋白中的蛋白质结构域的非限制性实例包括脱氨酶结构域(例如腺苷脱氨酶)、尿嘧啶糖基化酶抑制剂(UGI)结构域、表位标签和报告基因序列。

[0859] 表位标签的非限制性实例包括组氨酸(His)标签、V5标签、FLAG标签、流感血凝素(HA)标签、Myc标签、VSV-G标签和硫氧还蛋白(Trx)标签。报告基因的例子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-葡萄糖醛酸酶、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)和自发荧光蛋白,包括蓝色荧光蛋白(BFP)。额外的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 ADNA结合结构域(DBD)融合、GAL4DNA结合结构域融合,和单纯疱疹病毒(HSV)BP16蛋白融合。

[0860] 在一些实施例中,所述腺苷碱基编辑器(ABE)可使DNA中的腺嘌呤脱氨基。在一些实施例中,ABE是通过用天然或工程化的大肠杆菌TadA、人ADAR2、小鼠ADA或人ADAT2替换BE3的APOBEC1组分而产生的。在一些实施例中,ABE包含进化的TadA变体。在一些实施例中,ABE是ABE 1.2(TadA*-XTEN-nCas9-NLS)。在一些实施例中,TadA*包含A106V和D108N突变。

[0861] 在一些实施例中,所述ABE是第二代ABE。在一些实施例中,所述ABE是ABE2.1,其包含TadA*(TadA*2.1)中的额外突变D147Y和E155V。在一些实施例中,所述ABE是ABE2.2,ABE2.1与无催化活性版本的人烷基腺嘌呤DNA糖基化酶(具有E125Q突变的AAG)融合。在一些实施例中,所述ABE是ABE2.3,ABE2.1融合到催化灭活版本的大肠杆菌Endo V(用D35A突变灭活)。在一些实施例中,所述ABE是ABE2.6,其接头长度(32个氨基酸,(SGGS)2-XTEN-$(SGGS)_2$)是ABE2.1中接头的两倍。在一些实施例中,所述ABE是ABE2.7,其是与另外的野生型TadA单体相连的ABE2.1。在一些实施例中,所述ABE是ABE2.8,其是与另外的TadA*2.1单体拴系的ABE2.1。在一些实施例中,所述ABE是ABE2.9,其是进化的TadA(TadA*2.1)与ABE2.1的N-末端的直接融合。在一些实施例中,所述ABE是ABE2.10,其是野生型TadA与ABE2.1的N末端的直接融合。在一些实施例中,所述ABE是ABE2.11,其是在TadA*单体的N-末端具有失活E59A突变的ABE2.9。在一些实施例中,所述ABE是ABE2.12,其是在内部TadA*单体中具有失活E59A突变的ABE2.9。

[0862] 在一些实施例中,所述ABE是第三代ABE。在一些实施例中,所述ABE是ABE3.1,其是具有三个额外TadA突变(L84F、H123Y和I156F)的ABE2.3。

[0863] 在一些实施例中,所述ABE是第四代ABE。在一些实施例中,所述ABE是ABE4.3,其是具有额外TadA突变A142N(TadA*4.3)的ABE3.1。

[0864] 在一些实施例中,所述ABE是第五代ABE。在一些实施例中,所述ABE是ABE5.1,其通过将来自存活克隆(H36L、R51L、S146C和K157N)的一组共有突变导入ABE3.1而产生。在一些实施例中,所述ABE是ABE5.3,其具有包含融合到内部进化的TadA*的野生型大肠杆菌TadA的异源二聚体构建体。在一些实施例中,所述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6所示。在一些实施例中,所述ABE是第六代ABE。在一些实施例中,所述ABE是ABE6.1、ABE6.2、ABE6.3、ABE6.4、ABE6.5或ABE6.6,如下表6所示。在一些实施例中,所述ABE是第七代ABE。在一些实施例中,所述ABE是ABE7.1、ABE7.2、ABE7.3、ABE7.4、ABE7.5、ABE7.6、ABE7.7、ABE7.8、ABE7.9或ABE7.10,如在下表6所示。

[0865] 表6.ABE的基因型

[0866]

	23	26	36	37	48	49	51	72	84	87	105	108	123	125	142	145	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

[0867]

	23	26	36	37	48	49	51	72	84	87	105	108	123	125	142	145	147	152	155	156	157	161
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	I	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
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	I	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

[0868]

	23	26	36	37	48	49	51	72	84	87	105	108	123	125	142	145	147	152	155	156	157	161
ABE7.3	I	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	I	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	I	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

[0869] 在一些实施例中,所述碱基编辑器是第八代ABE(ABE8)。在一些实施例中,所述ABE8包含TadA*8变体。在一些实施例中,所述ABE8具有包含TadA*8变体(“ABE8.x-m”)的单体构建体。在一些实施例中,所述ABE8是ABE8.1-m,其具有包含带有Y147T突变(TadA*8.1)的TadA*7.10的单体构建体。在一些实施例中,所述ABE8是ABE8.2-m,其具有包含带有Y147R突变(TadA*8.2)的TadA*7.10的单体构建体。在一些实施例中,所述ABE8是ABE8.3-m,其具有包含带有Q154S突变(TadA*8.3)的TadA*7.10的单体构建体。在一些实施例中,所述ABE8是ABE8.4-m,其具有包含带有Y123H突变(TadA*8.4)的TadA*7.10的单体构建体。在一些实施例中,所述ABE8是ABE8.5-m,其具有包含带有V82S突变(TadA*8.5)的TadA*7.10的单体构建体。在一些实施例中,所述ABE8是ABE8.6-m,其具有包含带有T166R突变(TadA*8.6)的TadA*7.10的单体构建体。在一些实施例中,所述ABE8是ABE8.7-m,其具有包含带有Q154R突变(TadA*8.7)的TadA*7.10的单体构建体。在一些实施例中,所述ABE8是ABE8.8-m,其具有包含带有Y147R、Q154R和Y123H突变(TadA*8.8)的TadA*7.10的单体构建体。在一些实施例中,所述ABE8是ABE8.9-m,其具有包含带有Y147R、Q154R和I76Y突变(TadA*8.9)的TadA*7.10的单体构建体。在一些实施例中,所述ABE8是ABE8.10-m,其具有包含带有Y147R、Q154R和T166R突变(TadA*8.10)的TadA*7.10的单体构建体。在一些实施例中,所述ABE8是ABE8.11-m,其具有包含带有Y147T和Q154R突变(TadA*8.11)的TadA*7.10的单体构建体。在一些实施例中,所述ABE8是ABE8.12-m,其具有包含带有Y147T和Q154S突变(TadA*8.12)的TadA*7.10的单体构建体。在一些实施例中,所述ABE8是ABE8.13-m,其具有包含带有Y123H(从H123Y回复的Y123H)、Y147R、Q154R和I76Y突变(TadA*8.13)的TadA*7.10的单体构建体。在一些实施例中,所述ABE8是ABE8.14-m,其具有包含带有I76Y和V82S突变(TadA*8.14)的TadA*7.10的单体构建体。在一些实施例中,所述ABE8是ABE8.15-m,其具有包含带有V82S和Y147R突变(TadA*8.15)的TadA*7.10的单体构建体。在一些实施例中,所述ABE8是ABE8.16-m,其具有包含带有V82S、Y123H(从H123Y回复的Y123H)和Y147R突变(TadA*8.16)的TadA*7.10的单体构建体。在一些实施例中,所述ABE8是ABE8.17-m,其具有包含带有V82S和Q154R突变(TadA*8.17)的TadA*7.10的单体构建体。在一些实施例中,所述ABE8是ABE8.18-m,其具有包含带有V82S、Y123H(从H123Y回复的Y123H)和Q154R突变(TadA*8.18)的TadA*7.10的单体构建体。在一些实施例中,所述ABE8是ABE8.19-m,其具有包含带有V82S、Y123H(从H123Y回复的Y123H)、Y147R和Q154R突变(TadA*8.19)的TadA*7.10的单体构建体。在一些实施例中,所述ABE8是ABE8.20-m,其具有包含带有I76Y、V82S、Y123H(Y123H从H123Y回复)、Y147R和Q154R突变(TadA*8.20)的TadA*7.10的单体构建体。在一些实施例中,所述ABE8是

ABE8.21-m,其具有包含带有Y147R和Q154S突变(TadA*8.21)的TadA*7.10的单体构建体。在一些实施例中,所述ABE8是ABE8.22-m,其具有包含带有V82S和Q154S突变(TadA*8.22)的TadA*7.10的单体构建体。在一些实施例中,所述ABE8是ABE8.23-m,其具有包含带有V82S和Y123H(从H123Y恢复的Y123H)突变(TadA*8.23)的TadA*7.10的单体构建体。在一些实施例中,所述ABE8是ABE8.24-m,其具有包含带有V82S、Y123H(Y123H从H123Y回复)和Y147T突变(TadA*8.24)的TadA*7.10的单体构建体。

[0870] 在一些实施例中,所述ABE8具有异源二聚体构建体,其包含与TadA*8变体("ABE8.x-d")融合的野生型大肠杆菌TadA。在一些实施例中,所述ABE8是ABE8.1-d,其具有异源二聚体构建体,该结构包含与带有Y147T突变(TadA*8.1)的TadA*7.10融合的野生型大肠杆菌TadA。在一些实施例中,所述ABE8是ABE8.2-d,其具有异源二聚体构建体,该结构包含与带有Y147R突变(TadA*8.2)的TadA*7.10融合的野生型大肠杆菌TadA。在一些实施例中,所述ABE8是ABE8.3-d,其具有异源二聚体构建体,该结构包含与带有Q154S突变(TadA*8.3)的TadA*7.10融合的野生型大肠杆菌TadA。在一些实施例中,所述ABE8是ABE8.4-d,其具有异源二聚体构建体,该结构包含与带有Y123H突变(TadA*8.4)的TadA*7.10融合的野生型大肠杆菌TadA。在一些实施例中,所述ABE8是ABE8.5-d,其具有异源二聚体构建体,该结构包含与带有V82S突变(TadA*8.5)的TadA*7.10融合的野生型大肠杆菌TadA。在一些实施例中,所述ABE8是ABE8.6-d,其具有异源二聚体构建体,该结构包含与带有T166R突变(TadA*8.6)的TadA*7.10融合的野生型大肠杆菌TadA。在一些实施例中,所述ABE8是ABE8.7-d,其具有异源二聚体构建体,该结构包含与带有Q154R突变(TadA*8.7)的TadA*7.10融合的野生型大肠杆菌TadA。在一些实施例中,所述ABE8是ABE8.8-d,其具有异源二聚体构建体,该结构包含与带有Y147R、Q154R和Y123H突变(TadA*8.8)的TadA*7.10融合的野生型大肠杆菌TadA。在一些实施例中,所述ABE8是ABE8.9-d,其具有异源二聚体构建体,该结构包含与带有Y147R、Q154R和I76Y突变(TadA*8.9)的TadA*7.10融合的野生型大肠杆菌TadA。在一些实施例中,所述ABE8是ABE8.10-d,其具有异源二聚体构建体,该结构包含与带有Y147R、Q154R和T166R突变(TadA*8.10)的TadA*7.10融合的野生型大肠杆菌TadA。在一些实施例中,所述ABE8是ABE8.11-d,其具有异源二聚体构建体,该结构包含与带有Y147T和Q154R突变(TadA*8.11)的TadA*7.10融合的野生型大肠杆菌TadA。在一些实施例中,所述ABE8是ABE8.12-d,其具有异源二聚体构建体,该结构包含与带有Y147T和Q154S突变(TadA*8.12)的TadA*7.10融合的野生型大肠杆菌TadA。在一些实施例中,所述ABE8是ABE8.13-d,其具有异源二聚体构建体,该结构包含与带有Y123H(从H123Y回复的Y123H)、Y147R、Q154R和I76Y突变(TadA*8.13)的TadA*7.10融合的野生型大肠杆菌TadA。在一些实施例中,所述ABE8是ABE8.14-d,其具有异源二聚体构建体,该结构包含与带有I76Y和V82S突变(TadA*8.14)的TadA*7.10融合的野生型大肠杆菌TadA。在一些实施例中,所述ABE8是ABE8.15-d,其具有异源二聚体构建体,该结构包含与带有V82S和Y147R突变(TadA*8.15)的TadA*7.10融合的野生型大肠杆菌TadA。在一些实施例中,所述ABE8是ABE8.16-d,其具有异源二聚体构建体,该结构包含与带有V82S、Y123H(从H123Y回复的Y123H)和Y147R突变(TadA*8.16)的TadA*7.10融合的野生型大肠杆菌TadA。在一些实施例中,所述ABE8是ABE8.17-d,其具有异源二聚体构建体,该结构包含与带有V82S和Q154R突变(TadA*8.17)的TadA*7.10融合的野生型大肠杆菌TadA。在一些实施例中,所述ABE8是ABE8.18-d,其具有异源二聚体构建体,该

结构包含与带有V82S、Y123H(从H123Y回复的Y123H)和Q154R突变(TadA*8.18)的TadA*7.10融合的野生型大肠杆菌TadA。在一些实施例中,所述ABE8是ABE8.19-d,其具有异源二聚体构建体,该结构包含与带有V82S、Y123H(从H123Y回复的Y123H)、Y147R和Q154R突变(TadA*8.19)的TadA*7.10融合的野生型大肠杆菌TadA。在一些实施例中,所述ABE8是ABE8.20-d,其具有异源二聚体构建体,该结构包含与带有I76Y、V82S、Y123H(从H123Y恢复的Y123H)、Y147R和Q154R突变(TadA*8.20)的TadA*7.10融合的野生型大肠杆菌TadA。在一些实施例中,所述ABE8是ABE8.21-d,其具有异源二聚体构建体,该结构包含与带有Y147R和Q154S突变(TadA*8.21)的TadA*7.10融合的野生型大肠杆菌TadA。在一些实施例中,所述ABE8是ABE8.22-d,其具有异源二聚体构建体,该结构包含与带有V82S和Q154S突变(TadA*8.22)的TadA*7.10融合的野生型大肠杆菌TadA。在一些实施例中,所述ABE8是ABE8.23-d,其具有异源二聚体构建体,该结构包含与带有V82S和Y123H(从H123Y恢复的Y123H)突变(TadA*8.23)的TadA*7.10融合的野生型大肠杆菌TadA。在一些实施例中,所述ABE8是ABE8.24-d,其具有异源二聚体构建体,该结构包含与带有V82S、Y123H(从H123Y回复的Y123H)和Y147T突变(TadA*8.24)的TadA*7.10融合的野生型大肠杆菌TadA。

[0871] 在一些实施例中,所述ABE8具有异源二聚体构建体,其包含与TadA*8变体("ABE8.x-7")融合的TadA*7.10。在一些实施例中,所述ABE8是ABE8.1-7,其具有包含TadA*7.10的异源二聚体构建体,该结构包含与带有Y147T突变(TadA*8.1)的TadA*7.10融合。在一些实施例中,所述ABE8是ABE8.2-7,其具有包含TadA*7.10的异源二聚体构建体,该结构与带有Y147R突变(TadA*8.2)的TadA*7.10融合。在一些实施例中,所述ABE8是ABE8.3-7,其具有包含TadA*7.10的异源二聚体构建体,该结构与带有Q154S突变(TadA*8.3)的TadA*7.10融合。在一些实施例中,所述ABE8是ABE8.4-7,其具有包含TadA*7.10的异源二聚体构建体,该结构与带有Y123H突变(TadA*8.4)的TadA*7.10融合。在一些实施例中,所述ABE8是ABE8.5-7,其具有包含TadA*7.10的异源二聚体构建体,该结构与带有V82S突变(TadA*8.5)的TadA*7.10融合。在一些实施例中,所述ABE8是ABE8.6-7,其具有包含TadA*7.10的异源二聚体构建体,该结构与带有T166R突变(TadA*8.6)的TadA*7.10融合。在一些实施例中,所述ABE8是ABE8.7-7,其具有包含TadA*7.10的异源二聚体构建体,该结构与带有Q154R突变(TadA*8.7)的TadA*7.10融合。在一些实施例中,所述ABE8是ABE8.8-7,其具有包含TadA*7.10的异源二聚体构建体,该结构与带有Y147R、Q154R和Y123H突变(TadA*8.8)的TadA*7.10融合。在一些实施例中,所述ABE8是ABE8.9-7,其具有包含TadA*7.10的异源二聚体构建体,该结构与带有Y147R、Q154R和I76Y突变(TadA*8.9)的TadA*7.10融合。在一些实施例中,所述ABE8是ABE8.10-7,其具有包含TadA*7.10的异源二聚体构建体,该结构与带有Y147R、Q154R和T166R突变(TadA*8.10)的TadA*7.10融合。在一些实施例中,所述ABE8是ABE8.11-7,其具有包含TadA*7.10的异源二聚体构建体,该结构与带有Y147T和Q154R突变(TadA*8.11)的TadA*7.10融合。在一些实施例中,所述ABE8是ABE8.12-7,其具有包含TadA*7.10的异源二聚体构建体,该结构与带有Y147T和Q154S突变(TadA*8.12)的TadA*7.10融合。在一些实施例中,所述ABE8是ABE8.13-7,其具有包含TadA*7.10的异源二聚体构建体,该结构与带有Y123H(从H123Y回复的Y123H)、Y147R、Q154R和I76Y突变(TadA*8.13)的TadA*7.10融合。在一些实施例中,所述ABE8是ABE8.14-7,其具有包含TadA*7.10的异源二聚体构建体,该结构与带有I76Y和V82S突变(TadA*8.14)的TadA*7.10融合。在一些实施例中,所述

ABE8是ABE8.15-7,其具有包含TadA*7.10的异源二聚体构建体,该结构与带有V82S和Y147R突变(TadA*8.15)的TadA*7.10融合。在一些实施例中,所述ABE8是ABE8.16-7,其具有包含TadA*7.10的异源二聚体构建体,该结构与带有V82S、Y123H(从H123Y回复的Y123H)和Y147R突变(TadA*8.16)的TadA*7.10融合。在一些实施例中,所述ABE8是ABE8.17-7,其具有包含TadA*7.10的异源二聚体构建体,该结构与带有V82S和Q154R突变(TadA*8.17)的TadA*7.10融合。在一些实施例中,所述ABE8是ABE8.18-7,其具有包含TadA*7.10的异源二聚体构建体,该结构与带有V82S、Y123H(从H123Y回复的Y123H)和Q154R突变(TadA*8.18)的TadA*7.10融合的TadA*7.10。在一些实施例中,所述ABE8是ABE8.19-7,其具有包含TadA*7.10的异源二聚体构建体,该结构与带有V82S、Y123H(从H123Y回复的Y123H)、Y147R和Q154R突变(TadA*8.19)的TadA*7.10融合。在一些实施例中,所述ABE8是ABE8.20-7,其具有包含TadA*7.10的异源二聚体构建体,该结构与带有I76Y、V82S、Y123H(Y123H从H123Y回复)、Y147R和Q154R突变(TadA*8.20)的TadA*7.10融合。在一些实施例中,所述ABE8是ABE8.21-7,其具有包含TadA*7.10的异源二聚体构建体,该结构与带有Y147R和Q154S突变(TadA*8.21)的TadA*7.10融合。在一些实施例中,所述ABE8是ABE8.22-7,其具有包含TadA*7.10的异源二聚体构建体,该结构与带有V82S和Q154S突变(TadA*8.22)的TadA*7.10融合。在一些实施例中,所述ABE8是ABE8.23-7,其具有包含TadA*7.10的异源二聚体构建体,该结构与带有V82S和Y123H(从H123Y回复的Y123H)突变(TadA*8.23)的TadA*7.10融合。在一些实施例中,所述ABE8是ABE8.24-7,其具有包含TadA*7.10的异源二聚体构建体,该结构与带有V82S、Y123H(从H123Y回复的Y123H)和Y147T突变(TadA*8.24)的TadA*7.10融合。

[0872] 在一些实施例中,所述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7所示。

[0873] 表7:腺苷脱氨酶碱基编辑器8(ABE8)

[0874]

ABE8	腺苷脱氨酶	腺苷脱氨酶描述
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)

[0875]

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)
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)

[0876] 在一些实施例中,碱基编辑器(例如ABE8)通过将腺苷脱氨酶变体(例如TadA*8)克隆到包括环状置换Cas9(例如CP5或CP6)和双分核定位序列的支架中来产生。在一些实施例中,所述碱基编辑器(例如ABE7.9、ABE7.10或ABE8)是NGC PAM CP5变体(化脓性链球菌Cas9或spVRQR Cas9)。在一些实施例中,所述碱基编辑器(例如ABE7.9、ABE7.10或ABE8)是AGA PAM CP5变体(化脓性链球菌Cas9或spVRQR Cas9)。在一些实施例中,所述碱基编辑器(例如ABE7.9、ABE7.10或ABE8)是NGC PAM CP6变体(化脓性链球菌Cas9或spVRQR Cas9)。在一些

实施例中,所述碱基编辑器(例如ABE7.9、ABE7.10或ABE8)是AGA PAM CP6变体(化脓性链球菌Cas9或spVRQR Cas9)。

[0877] 在一些实施例中,ABE具有如下表8所示的基因型。

[0878] 表8.ABE的基因型

[0879]

	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

[0880] 如下表9所示,描述了40个ABE8的基因型。指出了ABE进化的大肠杆菌TadA部分中的残留位置。当与ABE7.10突变不同时,显示了ABE8中的突变变化。在一些实施例中,所述ABE具有如下表9所示的ABE之一的基因型。

[0881] 表9.进化的TadA中的残基标识

[0882]

	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												

[0883]

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												
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						

[0884] 在一些实施例中,所述碱基编辑器是ABE8.1,其包含或基本上由以下具有腺苷脱氨酶活性的序列或其片段组成:

[0885] ABE8.1_Y147T_CP5_NGC PAM_单体

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEI

[0886] MALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDV

LHYPGMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD***SGGSSGGSSGSET***

[0887] *PGTSESATPESSGGSSGGS*EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGE
TGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP
KKYGGFMQPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK
EVKKDLIIKLPKYSLFELENGRKRMLASAKFLQKGNELALPSKYVNFLYLASHYEKLKGS
PEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAEN
IIHLFTLTNLGAPRAFKYFDTTIARKEYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD
*GGSGGSGGSGGSGGSGGSGGM*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDR
HSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH
MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSR
RLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLA
QIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVR
QQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNS
RFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV
YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEI
SGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIV
IEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGR
DMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMK
NYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMN
TKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEGADKRTADGSEFESPKKKRKV*

[0888] 上述序列中,纯文本表示腺苷脱氨酶序列,粗体序列表示源自Cas9的序列,斜体序列表示接头序列,下划线序列表示双分核定位序列。

[0889] 在一些实施例中,所述碱基编辑器是ABE8.1,其包含或基本上由以下具有腺苷脱氨酶活性的序列或其片段组成:

[0890] pNMG-B335 ABE8.1_Y147T_CP5_NGC PAM_单体

[0891] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEI
MALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDV

LHYPGMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTDSGGSSG*GSSGSET*

*PGTSESATPESSGGSSGGS*EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGE

TGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP

KKYGGFMQPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK

EVKKDLIIKLPKYSLFELENGRKRMLASAKFLQKGNELALPSKYVNFLYLASHYEKLKGS

PEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAEN

IIHLFTLTNLGAPRAFKYFDTTIARKEYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD

*GGSGGSGGSGGSGGSGGSGGM*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDR

HSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH

MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSR

RLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLA

[0892] QIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVR

QQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNS

RFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV

YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEI

SGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA

HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIV

IEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGR

DMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMK

NYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMN

TKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

YPKLESEFVYGDYKVYDVRKMIAKSEQEGADKRTADGSEFESPKKKRKV*

[0893] 上述序列中,纯文本表示腺苷脱氨酶序列,粗体序列表示源自Cas9的序列,斜体序列表示接头序列,下划线序列表示双分核定位序列。

[0894] 在一些实施例中,所述碱基编辑器是ABE8.14,其包含或基本上由以下具有腺苷脱氨酶活性的序列或其片段组成:

[0895] pNMG-357_ABE8.14与NGC PAM CP5

[0896] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEI

[0897] MALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDV
LHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTDGGSSGGS*SGSETP*
*GTSESATPESSGGSSGGS*MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIG
EGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRV
VFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKA
QSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGS*EIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESI
LPKRNSDKLIARKKDWDPKKYGGFMQPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIME
RSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAKFLQKGNELALPS
KYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKV
LSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYRSTKEVLDATLIHQ
SITGLYETRIDLSQLGGD*GGSGGSGGSGGSGGSGGSGGM*DKKYSIGLAIGTNSVGWAVIT
DEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYL
QEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKL
VDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPIN
ASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDA
KLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEK
MDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEK
ILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNL
PNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVK
QLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLT
LTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVK
VVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVE
NTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDK
NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLV
ETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYH
HAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEGADKRTADGSEFES
PKKKRKV*

[0898] 上述序列中,纯文本表示腺苷脱氨酶序列,粗体序列表示源自Cas9的序列,斜体序列表示接头序列,下划线序列表示双分核定位序列。

[0899] 在一些实施例中,所述碱基编辑器是ABE8.8-m,其包含或基本上由以下具有腺苷脱氨酶活性的序列或其片段组成:

[0900] ABE8.8-m

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEI
MALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDV
LHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSGSET*
*PGTSESATPESSGGSSGGS*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHS
IKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMI
KFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL
ENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQI
GDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQ
LPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYN
ELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISG

[0901] VEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHL
FDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSL
TFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDM
YVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNY
WRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTK
YDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYP
KLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPL
IETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIAR
KKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFL
EAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY
EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDL
SQLGGD<u>EGADKRTADGSEFESPKKKRKV</u>*

[0902] 上述序列中,纯文本表示腺苷脱氨酶序列,粗体序列表示源自Cas9的序列,斜体序列表示接头序列,下划线序列表示双分核定位序列,双下划线序列表示突变。

[0903] 在一些实施例中,所述碱基编辑器是ABE8.8-d,其包含或基本上由以下具有腺苷脱氨酶活性的序列或其片段组成:

[0904] ABE8.8-d

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEI
MALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDV
LHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD*SGGSSGGSSGSET*
*PGTSESATPESSGGSSGGS*SEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIG
EGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRV
VFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKA
QSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGS***DKKYSIGLAIGTNSVGWAVITDE**
YKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQE
IFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVD
STDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS
GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKL
QLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKR

[0905] **YDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD**
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKIL
TFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPN
EKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLT
LFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLK
SDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT
QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNR
GKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVET
RQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHA
HDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGG

FSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
[0906] ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD
ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
ATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[0907] 上述序列中,纯文本表示腺苷脱氨酶序列,粗体序列表示源自Cas9的序列,斜体序列表示接头序列,下划线序列表示双分核定位序列,双下划线序列表示突变。

[0908] 在一些实施例中,所述碱基编辑器是ABE8.13-m,其包含或基本上由以下具有腺苷脱氨酶活性的序列或其片段组成:

[0909] ABE8.13-m

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEI
MALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDV
LHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSGSET*
*PGTSESATPESSGGSSGGS*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHS
IKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMI
KFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL
ENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQI
GDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQ
LPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
[0910] RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYN
ELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISG
VEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHL
FDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSL
TFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDM
YVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNY
WRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTK
YDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYP

KLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPL
IETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIAR
KKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFL
[0911] EAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY
EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDL
SQLGGDEGADKRTADGSEFESPKKKRKV*

[0912] 上述序列中,纯文本表示腺苷脱氨酶序列,粗体序列表示源自Cas9的序列,斜体序列表示接头序列,下划线序列表示双分核定位序列,双下划线序列表示突变。

[0913] 在一些实施例中,所述碱基编辑器是ABE8.13-d,其包含或基本上由以下具有腺苷脱氨酶活性的序列或其片段组成:

[0914] ABE8.13-d

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEI
MALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDV
LHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD*SGGSSGGSSGSET*
*PGTSESATPESSGGSSGGS*SEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIG
EGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRV
VFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKA
QSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGS***DKKYSIGLAIGTNSVGWAVITDE**
YKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQE
IFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVD
[0915] **STDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS**
GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKL
QLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKR
YDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKIL
TFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPN
EKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLT
LFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLK

SDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT
QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNR
GKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVET
RQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHA
HDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
[0916] FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGG
FSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD
ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
ATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[0917] 上述序列中,纯文本表示腺苷脱氨酶序列,粗体序列表示源自Cas9的序列,斜体序列表示接头序列,下划线序列表示双分核定位序列,双下划线序列表示突变。

[0918] 在一些实施例中,所述碱基编辑器是ABE8.17-m,其包含或基本上由以下具有腺苷脱氨酶活性的序列或其片段组成:

[0919] ABE8.17-m

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEI
MALRQGGLVMQNYRLIDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDV
LHYPGMNHRVEITEGILADECAALLCYFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSGSET*
*PGTSESATPESSGGSSGGS*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHS
IKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMI
[0920] KFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL
ENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQI
GDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQ
LPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYN
ELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISG

[0921] **VEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHL FDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSL TFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDM YVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNY WRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTK YDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYP KLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPL IETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIAR KKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFL EAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDL SQLGGD**EGADKRTADGSEFESPKKKRKV*

[0922] 上述序列中,纯文本表示腺苷脱氨酶序列,粗体序列表示源自Cas9的序列,斜体序列表示接头序列,下划线序列表示双分核定位序列,双下划线序列表示突变。

[0923] 在一些实施例中,所述碱基编辑器是ABE8.17-d,其包含或基本上由以下具有腺苷脱氨酶活性的序列或其片段组成:

[0924] ABE8.17-d

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEI MALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDV LHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD*SGGSSGGSSGSET PGTSESATPESSGGSSGGS*SEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIG EGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYSTFEPCVMCAGAMIHSRIGRV

[0925] VFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRRVFNAQKKA QSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGS***DKKYSIGLAIGTNSVGWAVITDE YKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQE IFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVD STDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKL

QLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKR
YDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKIL
TFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPN
EKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLT
LFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLK
SDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT
[0926] QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNR
GKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVET
RQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHA
HDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGG
FSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD
ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
ATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[0927] 上述序列中,纯文本表示腺苷脱氨酶序列,粗体序列表示源自Cas9的序列,斜体序列表示接头序列,下划线序列表示双分核定位序列,双下划线序列表示突变。

[0928] 在一些实施例中,所述碱基编辑器是ABE8.20-m,其包含或基本上由以下具有腺苷脱氨酶活性的序列或其片段组成:

[0929] ABE8.20-m

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEI
MALRQGGLVMQNYRLYDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDV
LHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSGSET*
[0930] *PGTSESATPESSGGSSGGS*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHS
IKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMI

KFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL
ENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQI
GDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQ
LPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYN
ELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISG
VEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHL
FDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSL
TFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
[0931] MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDM
YVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNY
WRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTK
YDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYP
KLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPL
IETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIAR
KKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFL
EAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY
EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDL
SQLGGDEGADKRTADGSEFESPKKKRKV*

[0932] 上述序列中,纯文本表示腺苷脱氨酶序列,粗体序列表示源自Cas9的序列,斜体序列表示接头序列,下划线序列表示双分核定位序列,双下划线序列表示突变。

[0933] 在一些实施例中,所述碱基编辑器是ABE8.20-d,其包含或基本上由以下具有腺苷脱氨酶活性的序列或其片段组成:

[0934] ABE8.20-d

[0935] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEI
MALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDV
LHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD*SGGSSGGSSGSET*
*PGTSESATPESSGGSSGGS*SEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIG

[0936] EGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYSTFEPCVMCAGAMIHSRIGRV
VFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKA
QSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGS***DKKYSIGLAIGTNSVGWAVITDE**
YKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQE
IFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVD
STDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS
GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKL
QLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKR
YDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKIL
TFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPN
EKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLT
LFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLK
SDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT
QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNR
GKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVET
RQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHA
HDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGG
FSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD
ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
ATLIHQSITGLYETRIDLSQLGGD*EGADKRTADGSEFESPKKKRKV**

[0937] 上述序列中,纯文本表示腺苷脱氨酶序列,粗体序列表示源自Cas9的序列,斜体序列表示接头序列,下划线序列表示双分核定位序列,双下划线序列表示突变。

[0938] 在一些实施例中,本发明的ABE8选自以下序列:

[0939] 01.单(mono)ABE8.1_bpNLS+Y147T

[0940] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
TFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY

KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD
SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[0941] 02.单ABE8.1_bpNLS+Y147R

[0942] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
RFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD
SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK

QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[0943] 03.单ABE8.1_bpNLS+Q154S

[0944] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
YFFRMPRSVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD
SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[0945] 04.单ABE8.1_bpNLS+Y123H

[0946] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLC
YFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD
SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL

IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[0947] 05.单ABE8.1_bpNLS+V82S

[0948] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
YFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD
SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[0949] 06.单ABE8.1_bpNLS+T166R

[0950] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
YFFRMPRQVFNAQKKAQSSRDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL

KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD
SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[0951] 07.单ABE8.1_bpNLS+Q154R

[0952] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
YFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD
SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[0953] 08.单ABE8.1_bpNLS+Y147R_Q154R_Y123H

[0954] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLC
RFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS

DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD
SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[0955] 09.单ABE8.1_bpNLS+Y147R_Q154R_I76Y

[0956] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
RFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD
SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[0957] 10.单ABE8.1_bpNLS+Y147R_Q154R_T166R

[0958] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
RFFRMPRRVFNAQKKAQSSRDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD
SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[0959] 11.单ABE8.1_bpNLS+Y147T_Q154R

[0960] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
TFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD
SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG

GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[0961] 12.单ABE8.1_bpNLS+Y147T_Q154S

[0962] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
TFFRMPRSVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD
SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[0963] 13.单ABE8.1_bpNLS+H123Y123H_Y147R_Q154R_I76Y

[0964] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLC
RFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD

SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[0965] 14.单ABE8.1_bpNLS+V82S+Q154R

[0966] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
YFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD
SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[0967] 在一些实施例中,所述碱基编辑器是包含融合至核碱基编辑结构域(例如脱氨酶结构域的全部或部分)的多核苷酸可编程核苷酸结合结构域(例如Cas9衍生结构域)的融合蛋白。在某些实施例中,本文提供的融合蛋白包含一种或多种改善融合蛋白的碱基编辑活性的特征。例如,本文提供的任何融合蛋白可以包含具有降低的核酸酶活性的Cas9结构域。在一些实施例中,本文提供的任何融合蛋白可具有不具核酸酶活性的Cas9结构域(dCas9),或切割双链DNA分子的一条链的Cas9结构域,称为Cas9切口酶(nCas9)。

[0968] 在一些实施例中,所述碱基编辑器进一步包含一个包含全部或部分尿嘧啶糖基化酶抑制剂(UGI)的结构域。在一些实施例中,所述碱基编辑器包含一个包含全部或部分尿嘧

啶结合蛋白(UBP)的结构域,例如尿嘧啶DNA糖基化酶(UDG)。在一些实施例中,所述碱基编辑器包含一个包含全部或部分核酸聚合酶的结构域。在一些实施例中,掺入碱基编辑器的核酸聚合酶或其部分是易位DNA聚合酶。

[0969] 在一些实施例中,所述碱基编辑器的结构域可以包括多个结构域。例如,包含源自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瓣叶和NUC瓣叶的REC瓣叶和NUC瓣叶。在另一实例中,所述碱基编辑器可包含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。在一些实施例中,所述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多肽的野生型版本的突变(例如取代、插入、缺失)。例如,多核苷酸可编程DNA结合结构域的HNH结构域可包含H840A取代。在另一个实例中,多核苷酸可编程DNA结合结构域的RuvCI域可包含D10A取代。

[0970] 在使用或不使用一个或多个接头结构域(例如XTEN接头结构域)的情况下,本文公开的碱基编辑器的不同结构域(例如相邻结构域)可以彼此连接。在一些实施例中,接头结构域可以是键(例如共价键)、化学基团或连接两个分子或部分的分子,例如融合蛋白的两个结构域,例如第一结构域(例如Cas9衍生结构域)和第二结构域(例如腺苷脱氨酶结构域)。在一些实施例中,接头是共价键(例如碳-碳键、二硫键、碳-杂原子键等)。在某些实施例中,接头是酰胺键的碳氮键。在某些实施例中,接头是环状或无环、取代或未取代、支链或未支链的脂族或杂脂族接头。在某些实施例中,接头是聚合的(例如聚乙烯、聚乙二醇、聚酰胺、聚酯等)。在某些实施例中,接头包含氨基链烷酸的单体、二聚体或聚合物。在一些实施例中,接头包含氨基链烷酸(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)。在一些实施例中,接头包含氨基己酸(Ahx)的单体、二聚体或聚合物。在某些实施例中,接头基于碳环部分(例如环戊烷、环己烷)。在其他实施例中,接头包含聚乙二醇部分(PEG)。在某些实施例中,接头包含芳基或杂芳基部分。在某些实施例中,所述接头基于苯环。接头可以包括功能化部分以促进亲核剂(例如硫醇、氨基)从肽连接到接头。任何亲电子剂都可以用作接头的一部分。示例性的亲电子剂包括但不限于活化酯、活化酰胺、迈克尔受体、卤代烷、芳基卤、酰基卤和异硫氰酸酯。在一些实施例中,接头连接RNA可编程核酸酶的gRNA结合结构域,包括Cas9核酸酶结构域和核酸编辑蛋白的催化结构域。在一些实施例中,接头连接dCas9和第二结构域(例如UGI等)。

[0971] 通常,接头位于两个基团、分子或其他部分之间或两侧,并通过共价键与每个基团、分子或其他部分连接,从而将两者连接起来。在一些实施例中,接头是一个氨基酸或多个氨基酸(例如肽或蛋白质)。在一些实施例中,接头是有机分子、基团、聚合物或化学部分。在一些实施例中,接头的长度为2-100个氨基酸,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、30-35、35-40、40-45、45-50、50-60、60-70、70-80、80-90、90-100、100-150或150-200个氨基酸。在一些实施例中,接头的长度为约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。更长或更短的接头也被考虑。在一些实施例中,接头结构域包含氨基酸序列SGSETPGTSESATPES,其也可称为XTEN接头。可以使用任何连接融合蛋白结构域的方法(例如从非常柔性的接头形式(SGGS)n、(GGGS)n、(GGGGS)n和

(G)n,到更刚性的接头形式(EAAAK)n、(GGS)n、SGSETPGTSESATPES(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;全部内容通过引用并入本文)或$(XP)_n$基序)以达到所述核碱基编辑器活性的最佳长度。在一些实施例中,n是1、2、3、4、5、6、7、8、9、10、11、12、13、14或15。在一些实施例中,接头包含$(GGS)_n$基序,其中n为1、3或7。在一些实施例中,本文提供的融合蛋白的Cas9结构域通过包含氨基酸序列SGSETPGTSESATPES的接头融合。在一些实施例中,接头包含多个脯氨酸残基并且长度为5-21、5-14、5-9、5-7个氨基酸,例如PAPAP、PAPAPA、PAPAPAP、PAPAPAPA、$P(AP)_4$、$P(AP)_7$、$P(AP)_{10}$(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.2019Jan 25;10(1):439;全部内容通过引用并入本文)。这种富含脯氨酸的接头也称为"刚性"接头。

[0972] 本发明的融合蛋白包含核酸编辑结构域。在一些实施例中,所述脱氨酶是腺苷脱氨酶。在一些实施例中,所述脱氨酶是脊椎动物脱氨酶。在一些实施例中,所述脱氨酶是无脊椎动物脱氨酶。在一些实施例中,所述脱氨酶是人、黑猩猩、大猩猩、猴、牛、狗、大鼠或小鼠脱氨酶。在一些实施例中,所述脱氨酶是人脱氨酶。在一些实施例中,脱氨酶是大鼠脱氨酶。

[0973] 接头

[0974] 在某些实施例中,接头可用于连接本发明的任何肽或肽结构域。所述接头可以像共价键一样简单,或者它可以是多个原子长度的聚合接头。在某些实施例中,所述接头是多肽或基于氨基酸。在其他实施例中,所述接头不是肽样的。在某些实施例中,所述接头是共价键(例如碳-碳键、二硫键、碳-杂原子键等)。在某些实施例中,所述接头是酰胺键的碳-氮键。在某些实施例中,所述接头是环状或无环、取代或未取代、支链或未支链的脂族或杂脂族接头。在某些实施例中,所述接头是聚合的(例如聚乙烯、聚乙二醇、聚酰胺、聚酯等)。在某些实施例中,所述接头包含氨基链烷酸的单体、二聚体或聚合物。在某些实施例中,所述接头包含氨基链烷酸(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)。在某些实施例中,所述接头包含氨基己酸(Ahx)的单体、二聚体或聚合物。在某些实施例中,所述接头基于碳环部分(例如环戊烷、环己烷)。在其他实施例中,所述接头包含聚乙二醇部分(PEG)。在其他实施例中,所述接头包含氨基酸。在某些实施例中,所述接头包含肽。在某些实施例中,所述接头包含芳基或杂芳基部分。在某些实施例中,接头基于苯环。所述接头可以包括功能化部分以促进亲核剂(例如硫醇、氨基)从肽连接到接头。任何亲电子试剂都可以用作所述接头的一部分。示例性的亲电子剂包括但不限于活化酯、活化酰胺、迈克尔受体、卤代烷、芳基卤、酰基卤和异硫氰酸酯。

[0975] 在一些实施例中,所述接头是一个氨基酸或多个氨基酸(例如肽或蛋白质)。在一些实施例中,所述接头是键(例如共价键)、有机分子、基团、聚合物或化学部分。在一些实施例中,所述接头长度为约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。

[0976] 在一些实施例中,所述腺苷脱氨酶和napDNAbp通过长度为4、16、32或104个氨基酸

的接头融合。在一些实施例中,所述接头的长度为约3至约104个氨基酸。在一些实施例中,本文提供的任何融合蛋白包含通过接头彼此融合的腺苷脱氨酶和Cas9结构域。在脱氨酶结构域(例如工程化的ecTadA)和Cas9结构域之间,可以使用各种长度和柔性的接头(例如从非常柔性的$(GGGS)_n$、$(GGGGS)_n$和$(G)_n$形式的接头,到更刚性形式的接头$(EAAAK)_n$、$(SGGS)_n$、SGSETPGTSESATPES(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;全部内容通过引用并入本文)和$(XP)_n$以达到所述核碱基编辑器活性的最佳长度。在一些实施例中,n是1、2、3、4、5、6、7、8、9、10、11、12、13、14或15。在一些实施例中,接头包含$(GGS)_n$基序,其中n是1、3或7。在一些实施例中,本文提供的任何融合蛋白的腺苷脱氨酶和Cas9结构域通过包含氨基酸序列SGSETPGTSESATPES的接头(例如XTEN接头)融合。

[0977] Cas9与向导RNA复合

[0978] 本公开的一些方面提供包含本文提供的任何融合蛋白,和结合到融合蛋白的CAS9结构域(例如dCas9、核酸酶活性Cas9或Cas9切口酶)的向导RNA(例如靶向带有GSD1a可靶向突变的G6PC等位基因的向导)的复合物。可以使用任何连接融合蛋白结构域的方法(例如从非常柔性的接头形式$(EAAAK)_n$、$(SGGS)_n$、SGSETPGTSESATPES到更刚性的接头形式(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;全部内容通过引用并入本文)和$(XP)_n$以达到所述核碱基编辑器活性的最佳长度。在一些实施例中,n是1、2、3、4、5、6、7、8、9、10、11、12、13、14或15。在一些实施例中,所述接头包含$(GGS)_n$基序,其中n是1、3或7。在一些实施例中,本文提供的融合蛋白的Cas9结构域通过包含氨基酸序列SGSETPGTSESATPES的接头融合。

[0979] 在一些实施例中,所述向导核酸(例如向导RNA)的长度为15-100个核苷酸并且包含与目标序列互补的至少10个连续核苷酸的序列。在一些实施例中,所述向导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。在一些实施例中,所述向导RNA包含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个与目标序列互补的连续核苷酸。在一些实施例中,所述目标序列是DNA序列。在一些实施例中,所述目标序列是细菌、酵母、真菌、昆虫、植物或动物的基因组中的序列。在一些实施例中,所述目标序列是人类基因组中的序列。在一些实施例中,所述目标序列的3'端紧邻规范PAM序列(NGG)。在一些实施例中,所述目标序列的3'末端紧邻非规范PAM序列(例如表2或5'-NAA-3'中列出的序列)。在一些实施例中,所述向导核酸(例如向导RNA)与G6PC等位基因中带有GSD1a可靶向突变的序列互补。

[0980] 本公开的一些方面提供了使用本文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了包括使DNA分子与本文提供的任何融合蛋白和至少一种向导RNA接触的方法,其中所述向导RNA长约15-100个核苷酸并且包含至少与目标序列互补的10个连续核苷酸。在一些实施例中,所述目标序列的3'端紧邻AGC、GAG、TTT、GTG或CAA序列。在一些实施例中,所述目标序列的3'端紧邻NGA、NGCG、NGN、NNGRRT、NNNRRT、NGCG、NGCN、NGTN、NGTN、NGTN或5'(TTTV)序列。

[0981] 应当理解,各个序列中特定位置或残基的编号取决于所使用的特定蛋白质和编号方案。编号可能不同,例如,成熟蛋白质的前体和成熟蛋白质本身,物种之间的序列差异可能会影响编号。本领域技术人员将能够通过本领域众所周知的方法,例如通过序列比对和同源残基的测定,鉴定任何同源蛋白质和相应编码核酸中的相应残基。

[0982] 对于本领域技术人员显而易见的是,为了将本文公开的任何融合蛋白靶向靶位点,例如包含待编辑突变的位点,通常需要将融合蛋白与向导RNA共表达。如本文别处更详细解释的,向导RNA通常包含允许Cas9结合的tracrRNA框架和赋予Cas9:核酸编辑酶/结构域融合蛋白序列特异性的向导序列。或者,所述向导RNA和tracrRNA可以作为两个核酸分子分开提供。在一些实施例中,所述向导RNA包含一种结构,其中向导序列包含与目标序列互补的序列。所述向导序列的长度通常为20个核苷酸。基于本公开,用于将Cas9:核酸编辑酶/结构域融合蛋白靶向特定基因组靶位点的合适的向导RNA序列对于本领域技术人员来说将是显而易见的。此类合适的向导RNA序列通常包含与待编辑的目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将任何提供的融合蛋白靶向特定目标序列的示例性向导RNA序列。

[0983] Cas12与向导RNA复合

[0984] 本公开的一些方面提供包含本文提供的任何融合蛋白和向导RNA(例如靶向用于编辑的目標多核苷酸的向导)的复合物。

[0985] 在一些实施例中,所述向导核酸(例如向导RNA)的长度为15-100个核苷酸,并且包含与目标序列互补的至少10个连续核苷酸的序列。在一些实施例中,所述向导RNA是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。在一些实施例中,所述向导RNA包含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个与目标序列互补的连续核苷酸。在一些实施例中,所述目标序列是DNA序列。在一些实施例中,所述目标序列是细菌、酵母、真菌、昆虫、植物或动物的基因组中的序列。在一些实施例中,所述目标序列是人类基因组中的序列。在一些实施例中,所述目标序列的3'端紧邻规范PAM序列。在一些实施例中,所述目标序列的3'端紧邻非规范PAM序列。

[0986] 本公开的一些方面提供了使用本文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了包括使DNA分子与本文提供的任何融合蛋白和至少一种向导RNA接触的方法,其中所述向导RNA长约15-100个核苷酸并且包含至少与目标序列互补的10个连续核苷酸。在一些实施例中,所述目标序列的3'端紧邻例如TTN、DTTN、GTTN、ATTN、ATTC、DTTNT、WTTN、HATY、TTTN、TTTV、TTTC、TG、RTR或YTN PAM位点。

[0987] 应当理解,各个序列中特定位置或残基的编号取决于所使用的特定蛋白质和编号方案。编号可能不同,例如成熟蛋白质的前体和成熟蛋白质本身、物种之间的序列差异可能会影响编号。本领域技术人员将能够通过本领域众所周知的方法,例如通过序列比对和同源残基的测定,鉴定任何同源蛋白质和相应编码核酸中的相应残基。

[0988] 对于本领域技术人员显而易见的是,为了将本文公开的任何融合蛋白靶向靶位点,例如包含待编辑突变的位点,通常需要将融合蛋白与向导RNA共表达。如本文别处更详细解释的,向导RNA通常包含允许Cas12结合的tracrRNA框架和赋予Cas12:核酸编辑酶/结构域融合蛋白序列特异性的向导序列。或者,所述向导RNA和tracrRNA可以作为两个核酸分

子分开提供。在一些实施例中,所述向导RNA包含一种结构,其中向导序列包含与目标序列互补的序列。所述向导序列的长度通常为20个核苷酸。基于本公开,用于将Cas12:核酸编辑酶/结构域融合蛋白靶向特定基因组靶位点的合适的向导RNA序列对于本领域技术人员来说将是显而易见的。此类合适的向导RNA序列通常包含与待编辑的目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将任何提供的融合蛋白靶向特定目标序列的示例性向导RNA序列。

[0989] 本文公开的碱基编辑器的结构域可以以任何顺序排列,只要所述脱氨酶结构域内化在Cas12蛋白中即可。包含融合蛋白的碱基编辑器的非限制性实例包含例如Cas12结构域和脱氨酶结构域,可排列如下:

[0990] NH2-[Cas12结构域]-接头1-[ABE8]-接头2-[Cas12结构域]-COOH;

[0991] NH2-[Cas12结构域]-接头1-[ABE8]-[Cas12结构域]-COOH;

[0992] NH2-[Cas12结构域]-[ABE8]-接头2-[Cas12结构域]-COOH;

[0993] NH2-[Cas12结构域]-[ABE8]-[Cas12结构域]-COOH;

[0994] NH2-[Cas12结构域]-接头1-[ABE8]-接头2-[Cas12结构域]-[肌苷BER抑制剂]-COOH;

[0995] NH2-[Cas12结构域]-接头1-[ABE8]-[Cas12结构域]-[肌苷BER抑制剂]-COOH;

[0996] NH2-[Cas12结构域]-[ABE8]-接头2-[Cas12结构域]-[肌苷BER抑制剂]-COOH;;

[0997] NH2-[Cas12结构域]-[ABE8]-[Cas12结构域]-[肌苷BER抑制剂]-COOH;

[0998] NH2-[肌苷BER抑制剂]-[Cas12结构域]-接头1-[ABE8]-接头2-[Cas12结构域]-COOH;

[0999] NH2-[肌苷BER抑制剂]-[Cas12结构域]-接头1-[ABE8]-[Cas12结构域]-COOH;

[1000] NH2-[肌苷BER抑制剂]-[Cas12结构域]-[ABE8]-接头2-[Cas12结构域]-COOH;

[1001] NH2-[肌苷BER抑制剂]NH2-[Cas12结构域]-[ABE8]-[Cas12结构域]-COOH;

[1002] 此外,在某些情况下,Gam蛋白可以融合到碱基编辑器的N-末端。在某些情况下,Gam蛋白可以融合到碱基编辑器的C-末端。噬菌体Mu的Gam蛋白可以与双链断裂(DSB)的末端结合并保护它们免于降解。在一些实施例中,使用Gam结合DSB的自由端可以减少碱基编辑过程中插入缺失的形成。在一些实施例中,174个残基的Gam蛋白融合到碱基编辑器的N-末端。见Komor,A.C.等人,“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)。在某些情况下,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删除至少一个氨基酸可以减少所述碱基编辑器的长度。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,任何结构域中的替换都不会/不会改变所述碱基编辑器的长度。此类碱基编辑器的非限制性实例(其中所有结构域的长度与野生型结构域的长度相同)可包括:

[1003] NH2-[Cas12结构域]-接头1-[APOBEC1]-接头2-[Cas12结构域]-COOH;

[1004] NH2-[Cas12结构域]-接头1-[APOBEC1]-[Cas12结构域]-COOH;

[1005] NH2-[Cas12结构域]-[APOBEC1]-Linker2-[Cas12结构域]-COOH;

[1006] NH2-[Cas12结构域]-[APOBEC1]-[Cas12结构域]-COOH;

[1007] NH2-[Cas12结构域]-接头1-[APOBEC1]-接头2-[Cas12结构域]-[UGI]-COOH;

[1008] NH2-[Cas12结构域]-接头1-[APOBEC1]-[Cas12结构域]-[UGI]-COOH;

[1009] NH2-[Cas12结构域]-[APOBEC1]-接头2-[Cas12结构域]-[UGI]-COOH;

[1010] NH2-[Cas12结构域]-[APOBEC1]-[Cas12结构域]-[UGI]-COOH;

[1011] NH2-[UGI]-[Cas12结构域]-接头1-[APOBEC1]-接头2-[Cas12结构域]-COOH;

[1012] NH2-[UGI]-[Cas12结构域]-接头1-[APOBEC1]-[Cas12结构域]-COOH;

[1013] NH2-[UGI]-[Cas12结构域]-[APOBEC1]-接头2-[Cas12结构域]-COOH;

[1014] NH2-[UGI]-[Cas12结构域]-[APOBEC1]-[Cas12结构域]-COOH;

[1015] 在一些实施例中,本文提供的碱基编辑融合蛋白需要位于精确位置,例如,目标碱基位于限定区域(例如"脱氨基窗口")内的位置。在某些情况下,目标可以在4个碱基区域内。在某些情况下,这种定义的目标区域可以位于PAM上游大约15个碱基处。见Komor,A.C.等人,"Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage"Nature 533,420-424(2016);Gaudelli,N.M.等人,"Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage"Nature 551,464-471(2017);and Komor,A.C.等人,"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),其全部内容通过引用并入本文。

[1016] 定义的目标区域可以是脱氨基窗口。脱氨基窗口可以是碱基编辑器作用于目标核苷酸并使其脱氨基的限定区域。在一些实施例中,所述脱氨基窗口在2、3、4、5、6、7、8、9或10个碱基区域内。在一些实施例中,所述脱氨基窗口为PAM上游的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。

[1017] 本公开的碱基编辑器可以包含有助于编辑目标多核苷酸序列的任何结构域、特征或氨基酸序列。例如,在一些实施例中,所述碱基编辑器包含核定位序列(NLS)。在一些实施例中,所述碱基编辑器的NLS位于脱氨酶结构域和napDNAbp结构域之间。在一些实施例中,所述碱基编辑器的NLS位于napDNAbp结构域的C-末端。

[1018] 融合蛋白中包含的蛋白质结构域可以是异源功能域。可包含在融合蛋白中的蛋白质结构域的非限制性实例包括脱氨酶结构域(例如胞苷脱氨酶和/或腺苷脱氨酶)、尿嘧啶糖基化酶抑制剂(UGI)结构域、表位标签和报告基因序列。蛋白质结构域可以是异源功能域,例如,具有一种或多种以下活性:转录激活活性、转录抑制活性、转录释放因子活性、基因沉默活性、染色质修饰活性、表观遗传修饰活性、组蛋白修饰活性,RNA切割活性和核酸结合活性。此类异源功能域可赋予功能活性,例如修饰与目标DNA(例如组蛋白、DNA结合蛋白等)相关的目标多肽,导致例如组蛋白甲基化、组蛋白乙酰化、组蛋白泛素化等。赋予的其他功能和/或活性可以包括转座酶活性、整合酶活性、重组酶活性、连接酶活性、泛素连接酶活性、去泛素化活性、腺苷酸化活性、去腺苷酸化活性、SUMOylation活性、去SUMOylation活性,或上述的任何组合。

[1019] 可以用表位标签、报告蛋白、其他结合结构域检测或标记结构域。表位标签的非限制性实例包括组氨酸(His)标签、V5标签、FLAG标签、流感血凝素(HA)标签、Myc标签、VSV-G标签和硫氧还蛋白(Trx)标签。报告基因的例子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-葡萄糖醛酸酶、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)和自发荧光蛋白,包括蓝色荧光蛋白(BFP)。额外的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、GAL4 DNA结合结构域融合和单纯疱疹病毒(HSV)BP16蛋白融合。

[1020] 在一些实施例中,BhCas12b向导多核苷酸具有以下序列:

[1021] BhCas12b sgRNA支架(下划线)+20nt到23nt向导序列(用N_n表示)

[1022] 5'GUUCUGTCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUGCAGGGUGUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUACGAGGCAUUAGCACNNNNNNNNNNNNNNNNNNN-3'

[1023] 在一些实施例中,BvCas12b和AaCas12b向导多核苷酸具有以下序列:

[1024] BvCas12b sgRNA支架(下划线)+20nt到23nt向导序列(用N_n表示)

[1025] 5'GACCUAUAGGGUCAAUGAAUCUGUGCGUGUGCCAUAAGUAAUUAAAAAUUACCCACCACAGGAGCACCUGAAAACAGGUGCUUGGCACNNNNNNNNNNNNNNNNNNNN-3'

[1026] AaCas12b sgRNA支架(下划线)+20nt到23nt向导序列(用N_n表示)

[1027] 5'GUCUAAAGGACAGAAUUUUUCAACGGGUGUGCCAAUGGCCACUUUCCAGGUGGCAAAGCCCGUUGAACUUCUCAAAAAGAACGAUCUGAGAAGUGGCACNNNNNNNNNNNNNNNNNNNN-3'

[1028] 使用包含腺苷脱氨酶变体和Cas9结构域的融合蛋白的方法

[1029] 本公开的一些方面提供了使用本文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了包括使编码蛋白质的突变形式的DNA分子与本文提供的任何融合蛋白和至少一种向导RNA接触的方法,其中向导RNA长约15-100个核苷酸,并且包含与目标序列互补的至少10个连续核苷酸的序列。在一些实施例中,所述目标序列的3'端紧邻规范PAM序列(NGG)。在一些实施例中,所述目标序列的3'末端不与规范PAM序列(NGG)直接相邻。在一些实施例中,所述目标序列的3'端紧邻AGC、GAG、TTT、GTG或CAA序列。在一些实施例中,所述目标序列的3'端紧邻NGA、NGCG、NGN、NNGRRT、NNNRRT、NGCG、NGCN、NGTN、NGTN、NGTN或5'(TTTV)序列。

[1030] 应当理解,各个序列中特定位置或残基的编号取决于所使用的特定蛋白质和编号方案。编号可能不同,例如,成熟蛋白质的前体和成熟蛋白质本身,物种之间的序列差异可能会影响编号。本领域技术人员将能够通过本领域众所周知的方法,例如通过序列比对和同源残基的测定,鉴定任何同源蛋白质和相应编码核酸中的相应残基。

[1031] 对本领域技术人员来说显而易见的是,为了将包含Cas9结构域和腺苷脱氨酶变体(例如ABE8)的任何融合蛋白,如本文公开的,靶向至靶位点,例如包含对于要编辑的突变,通常需要将融合蛋白与向导RNA(例如sgRNA)共表达。如本文别处更详细解释的,所述向导RNA通常包含允许Cas9结合的tracrRNA框架和赋予Cas9:核酸编辑酶/结构域融合蛋白序列特异性的向导序列。或者,所述向导RNA和tracrRNA可以作为两个核酸分子分开提供。在一些实施例中,所述向导RNA包含一种结构,其中向导序列包含与目标序列互补的序列。所述向导序列通常为20个核苷酸长。基于本公开,用于将Cas9:核酸编辑酶/结构域融合蛋白靶向特定基因组靶位点的合适的向导RNA序列对于本领域技术人员来说将是显而易见的。此类合适的向导RNA序列通常包含与待编辑的目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将任何提供的融合蛋白靶向特定目标序列的示

例性向导RNA序列。

[1032] 碱基编辑器效率

[1033] CRISPR-Cas9核酸酶已被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,并在sgRNA序列指定的靶位点诱导双链DNA断裂(DSB)。细胞主要通过非同源末端连接(NHEJ)修复途径对这种DSB做出反应,这会导致随机插入缺失(indel),从而导致移码突变破坏基因。在存在与DSB侧翼序列高度同源的供体DNA模板的情况下,可以通过称为同源定向修复(HDR)的替代途径实现基因校正。不幸的是,在大多数非扰动条件下,HDR效率低下,取决于细胞状态和细胞类型,并且由更高频率的插入缺失主导。由于与人类疾病相关的大多数已知遗传变异都是点突变,因此需要能够更有效、更干净地进行精确点突变的方法。本文提供的碱基编辑系统提供了一种提供基因组编辑的新方法,无需产生双链DNA断裂,无需供体DNA模板,也无需诱导过量的随机插入和缺失。

[1034] 本发明的碱基编辑器有利地修饰编码包含突变的蛋白质的特定核苷酸碱基,而不产生显着比例的插入缺失。如本文所用,术语"插入缺失"是指核酸内核苷酸碱基的插入缺失。这种插入缺失可导致基因编码区内的移码突变。在一些实施例中,期望产生有效修饰(例如突变或脱氨基)核酸内的特定核苷酸而不在目标核酸序列中产生大量插入缺失(即indels)的碱基编辑器。在某些实施例中,本文提供的任何碱基编辑器与插入缺失相比能够产生更大比例的预期修饰(例如点突变或脱氨基)。

[1035] 在一些实施例中,本文提供的任何碱基编辑器系统导致在目标多核苷酸序列中小于50%、小于40%、小于30%、小于20%、小于19%、小于18%、小于17%、小于16%、小于15%、小于14%、小于13%、小于12%、小于11%、小于10%、小于9%、小于8%、小于7%、小于6%、小于5%、小于4%、小于3%、小于2%、小于1%、小于0.9%、小于0.8%、小于0.7%、小于0.6%、小于0.5%、小于0.4%、小于0.3%、小于0.2%、小于0.1%、小于0.09%、小于0.08%、小于0.07%、小于0.06%、小于0.05%、小于0.04%、小于0.03%、小于0.02%或小于0.01%的插入缺失形成。

[1036] 在一些实施例中,包含本文所述的ABE8碱基编辑器变体之一的碱基编辑器系统中的任一个导致在目标多核苷酸序列中小于50%、小于40%、小于30%、小于20%、小于19%、小于18%、小于17%、小于16%、小于15%、小于14%、小于13%、小于12%、小于11%、小于10%、小于9%、小于8%、小于7%、小于6%、小于5%、小于4%、小于3%、小于2%、小于1%、小于0.9%、小于0.8%、小于0.7%、小于0.6%、小于0.5%、小于0.4%、小于0.3%、小于0.2%、小于0.1%、小于0.09%、小于0.08%、小于0.07%、小于0.06%、小于0.05%、小于0.04%、小于0.03%、小于0.02%或小于0.01%的插入缺失形成。在一些实施例中,包含本文所述的ABE8碱基编辑器变体之一的任何碱基编辑器系统导致目标多核苷酸序列中小于0.8%的插入缺失形成。在一些实施例中,包含本文所述的ABE8碱基编辑器变体之一的任何碱基编辑器系统导致目标多核苷酸序列中至多0.8%的插入缺失形成。在一些实施例中,包含本文所述的ABE8碱基编辑器变体之一的任何碱基编辑器系统导致目标多核苷酸序列中小于0.3%的插入缺失形成。在一些实施例中,与包含ABE7碱基编辑器之一的碱基编辑器系统相比,包含所述ABE8碱基编辑器变体之一的任何碱基编辑器系统导致目标多核苷酸序列中较低的插入缺失形成。在一些实施例中,与包含ABE7.10的碱基编辑器系统相比,包含本

文所述的ABE8碱基编辑器变体之一的任何碱基编辑器系统导致目标多核苷酸序列中较低的插入缺失形成。

[1037] 在一些实施例中,与包含ABE7碱基编辑器之一的碱基编辑器系统相比,包含本文描述的ABE8碱基编辑器变体之一的任何碱基编辑器系统具有降低的插入/缺失频率。在一些实施例中,与包含ABE7碱基编辑器之一的碱基编辑器系统相比,包含本文所述的ABE8碱基编辑器变体之一的任何碱基编辑器系统,插入缺失频率降低至少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%、60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%或至少95%。在一些实施例中,与包含ABE7.10的碱基编辑器系统相比,包含本文所述的ABE8碱基编辑器变体之一的碱基编辑器系统,插入缺失频率降低至少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%。

[1038] 本发明提供了具有增加的效率和特异性的腺苷脱氨酶变体(例如ABE8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需碱基,并且不太可能编辑不打算改变的碱基(例如“旁观者”)。

[1039] 在一些实施例中,包含本文所述的ABE8碱基编辑器变体之一的任何碱基编辑系统具有减少的旁观者编辑或突变。在一些实施例中,非预期编辑或突变是旁观者突变或旁观者编辑,例如目标核苷酸序列的目标窗口中非预期或非目标位置中的目标碱基(例如A或C)的碱基编辑。在一些实施例中,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ABE8碱基编辑器变体之一的任何碱基编辑系统具有减少的旁观者编辑或突变。在一些实施例中,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ABE8碱基编辑器变体之一的任何碱基编辑系统已将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%。在一些实施例中,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ABE8碱基编辑器变体之一的任何碱基编辑系统已将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。

[1040] 在一些实施例中,包括本文所述的ABE8碱基编辑器变体之一的任何碱基编辑系统都减少了假编辑。在一些实施例中,非预期编辑或突变是假突变或假编辑,例如,在基因组的非预期或非目标区域中,对目标碱基(例如A或C)的非特异性编辑或不依赖于向导的编辑。在一些实施例中,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ABE8碱基编辑器变体之一的任何碱基编辑系统具有减少的假编辑。在一些实施例中,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ABE8碱基编辑器变体之一的任何碱基编辑系统已将假编辑减少至少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％。在一些实施例中，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ABE8碱基编辑器变体之一的任何碱基编辑系统已将假编辑减少至少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。

[1041] 本公开的一些方面基于以下认识：本文提供的任何碱基编辑器能够有效地在核酸(例如个体基因组内的核酸)中产生预期突变，例如点突变，而不会产生大量意外突变，例如意外点突变(即旁观者的突变)。在一些实施例中，本文提供的任何碱基编辑器能够产生至少0.01％的预期突变(即至少0.01％的碱基编辑效率)。在一些实施例中，本文提供的任何碱基编辑器能够产生至少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。

[1042] 在一些实施例中，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。在一些实施例中，所述碱基编辑效率可以通过计算细胞群中编辑的核碱基的百分比来测量。在一些实施例中，通过细胞群中编辑的核碱基来测量，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。

[1043] 在一些实施例中，与ABE7碱基编辑器相比相比，本文所述的任何ABE8碱基编辑器变体具有更高的碱基编辑效率。在一些实施例中，与ABE7碱基编辑器(例如ABE7.10)相比，本文所述的任何ABE8碱基编辑器变体具有至少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％、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％更高的碱基编辑效率

[1044] 在一些实施例中，与ABE7碱基编辑器例如ABE7.10相比，本文所述的任何ABE8碱基编辑器变体具有至少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倍更高的碱基编辑效率。

[1045] 在一些实施例中,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%的目标碱基编辑效率。在一些实施例中,通过细胞群中编辑的目标核碱基来测量,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%的在靶碱基编辑效率。

[1046] 在一些实施例中,与ABE7碱基编辑器相比,本文所述的任何ABE8碱基编辑器变体具有更高的在靶碱基编辑效率。在一些实施例中,与ABE7碱基编辑器(例如ABE7.10)相比,本文所述的任何ABE8碱基编辑器变体具有至少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%、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%更高的在靶碱基编辑效率。

[1047] 在一些实施例中,与ABE7碱基编辑器例如ABE7.10相比,本文所述的任何ABE8碱基编辑器变体具有至少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倍更高的在靶碱基编辑效率。

[1048] 本文所述的ABE8碱基编辑器变体可通过质粒、载体、LNP复合物或mRNA递送至宿主细胞。在一些实施例中,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。在一些实施例中,由编辑的核碱基测量,通过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%的在靶编辑效率。在一些实施例中,与由质粒或载体系统递送的ABE8碱基编辑器相比,由mRNA系统递送的ABE8碱基编辑器具有更高的碱基编辑效率。在一些实施例中,与由质粒或载体系统递送时相比,由mRNA系统递送时,本文所述的任何ABE8碱基编辑器变体具有至少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%、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%的在靶编辑效率。在一些实施例中,与由质粒或载体系统递送时相比,由mRNA系统递送时,本文所述的任何ABE8碱基编辑器变体具有至少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倍更高的在靶编辑效率。

[1049] 在一些实施例中,包含本文所述的ABE8碱基编辑器变体之一的碱基编辑器系统中的任一个导致少于50%、少于40%、少于30%、少于20%、少于19%、少于18%、少于17%、少于16%、少于15%、少于14%、少于13%、少于12%、少于11%、少于10%、少于9%、少于8%、少于7%、少于6%、少于5%、少于4%、少于3%、少于2%、少于1%、少于0.9%、少于0.8%、少于0.7%、少于0.6%、少于0.5%、少于0.4%、少于0.3%、少于0.2%、少于0.1%、少于0.09%、少于0.08%、少于0.07%、少于0.06%、少于0.05%、少于0.04%、少于0.03%、少于0.02%或少于0.01%的目标多核苷酸序列中的脱靶编辑。

[1050] 在一些实施例中,与由质粒或载体系统递送时相比,本文所述的任何ABE8碱基编辑器变体在由mRNA系统递送时具有更低的向导脱靶编辑效率。在一些实施例中,与由质粒或载体系统递送时相比,由mRNA系统递送时,本文所述的任何ABE8碱基编辑器变体具有至少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%较低的向导脱靶编辑效率。在一些实施例中,与由质粒或载体系统递送时相比,由mRNA系统递送时,本文所述的任何ABE8碱基编辑器变体具有至少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倍更低的向导脱靶编辑效率。在一些实施例中,与由质粒或载体系统递送时相比,本文所述的任何ABE8碱基编辑器变体在由mRNA系统递送时具有至少约2.2倍降低的向导脱靶编辑效率。

[1051] 在一些实施例中,与由质粒或载体系统递送时相比,本文所述的任何ABE8碱基编辑器变体在由mRNA系统递送时具有更低的不依赖于向导的脱靶编辑效率。在一些实施例中,与由质粒或载体系统递送时相比,由mRNA系统递送时,本文所述的任何ABE8碱基编辑器变体具有至少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％更低的不依赖于向导的脱靶编辑效率。在一些实施例中,与由质粒或载体系统递送时相比,由mRNA系统递送时,本文所述的任何ABE8碱基编辑器变体具有至少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倍更低的不依赖于向导的脱靶编辑效率。在一些实施例中,与由质粒或载体系统递送时相比,由mRNA系统递送时,本文所述的ABE8碱基编辑器变体具有134.0倍降低的不依赖于向导的脱靶编辑效率(例如假RNA脱氨基)。在一些实施例中,本文所述的ABE8碱基编辑器变体不增加跨基因组的不依赖于向导的突变率。

[1052] 本公开的一些方面基于以下认识:本文提供的任何碱基编辑器能够有效地在核酸(例如个体基因组内的核酸)中产生预期突变,例如点突变,而不会产生大量意外突变(例如假的脱靶编辑或旁观者编辑)。在一些实施例中,预期突变是由与gRNA结合的特定碱基编辑器产生的突变,专门设计用于改变或校正目标基因中的突变。本公开的一些方面基于以下认识:本文提供的任何碱基编辑器能够有效地在核酸(例如个体基因组内的核酸)中产生预期突变而不产生大量非预期突变。在一些实施例中,预期突变是由与gRNA结合的特定碱基编辑器产生的突变,专门设计用于改变或校正预期突变。在一些实施例中,预期突变是产生终止密码子的突变,例如基因编码区内的提前终止密码子。在一些实施例中,所述预期突变是消除终止密码子的突变。在一些实施例中,所述预期突变是改变基因剪接的突变。在一些实施例中,所述预期突变是改变基因(例如基因启动子或基因阻遏物)的调控序列的突变。

[1053] 在一些实施例中,本文提供的碱基编辑器能够产生大于1:1的预期突变与插入缺失(即非预期突变)比率。在一些实施例中,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或更多的预期突变与插入缺失比率。应当理解,本文描述的碱基编辑器的特征可以应用于任何融合蛋白,或使用本文提供的融合蛋白的方法。

[1054] 可以使用任何合适的方法确定预期突变和插入缺失的数量,例如,如国际PCT申请号PCT/2017/045381(WO2018/027078)和PCT/US2016/058344(WO2017/070632);Komor,A.C.等人,“Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage”Nature 533,420-424(2016);Gaudelli,N.M.等人,“Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage”Nature 551,464-471(2017);and Komor,A.C.等人,“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)中所描述;其全部内容通过引用并入本文。

[1055] 在一些实施例中,为了计算插入缺失频率,扫描测序读数以寻找与两个10-bp序列的精确匹配,所述两个10-bp序列位于可以发生插入缺失的窗口两侧。如果未找到完全匹配,则从分析中排除读取。如果此插入缺失窗口的长度与参考序列完全匹配,则读取被归类为不包含插入缺失。如果插入缺失窗口比参考序列长或短两个或更多碱基,则测序读数分别被归类为插入缺失。在一些实施例中,本文提供的碱基编辑器可以限制核酸区域中插入缺失的形成。在一些实施例中,该区域位于碱基编辑器靶向的核苷酸处或碱基编辑器靶向的核苷酸的2、3、4、5、6、7、8、9或10个核苷酸内的区域。

[1056] 在目标核苷酸区域形成的插入缺失的数量可取决于核酸(例如细胞基因组内的核酸)暴露于碱基编辑器的时间量。在一些实施例中,插入缺失的数量或比例是在将目标核苷酸序列(例如细胞基因组内的核酸)暴露于碱基编辑器后至少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天后决定。应当理解,如本文所述的碱基编辑器的特征可应用于任何融合蛋白或使用本文提供的融合蛋白的方法。

[1057] 在一些实施例中,本文提供的碱基编辑器能够限制核酸区域中插入缺失的形成。在一些实施例中,该区域位于碱基编辑器靶向的核苷酸处或碱基编辑器靶向的核苷酸的2、3、4、5、6、7、8、9或10个核苷酸内的区域。在一些实施例中,本文提供的任何碱基编辑器能够将核酸区域处插入缺失的形成限制为小于1%、小于1.5%、小于2%、小于2.5%、小于3%、小于3.5%、小于4%、小于4.5%、小于5%、小于6%、小于7%、小于8%、小于9%、小于10%、小于12%、小于15%或小于20%。在核酸区域形成的插入缺失的数量可取决于核酸(例如细胞基因组内的核酸)暴露于碱基编辑器的时间量。在一些实施例中,任何数量或比例的插入缺失是在将核酸(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天后决定。

[1058] 复合编辑

[1059] 在一些实施例中,本文提供的碱基编辑器系统能够复合编辑一个或多个基因中的多个核碱基对。在一些实施例中,所述多个核碱基对位于同一基因中。在一些实施例中,所述多个核碱基对位于一个或多个基因中,其中至少一个基因位于不同的基因座中。在一些实施例中,所述复合编辑可包含一种或多种向导多核苷酸。在一些实施例中,所述复合编辑可包括一个或多个碱基编辑器系统。在一些实施例中,所述复合编辑可以包括一个或多个具有单向导多核苷酸的碱基编辑器系统。在一些实施例中,所述复合编辑可以包括一个或多个具有多个向导多核苷酸的碱基编辑器系统。在一些实施例中,所述复合编辑可以包括一种或多种具有单碱基编辑器系统的向导多核苷酸。在一些实施例中,所述复合编辑可包含至少一种不需要PAM序列来靶向结合目标多核苷酸序列的向导多核苷酸。在一些实施例中,所述复合编辑可包含至少一种向导多核苷酸,其需要PAM序列靶向结合目标多核苷酸序列。在一些实施例中,所述复合编辑可以包括至少一种不需要PAM序列靶向结合目标多核苷酸序列的向导多核苷酸,和至少一种需要PAM序列靶向结合目标多核苷酸序列的向导多核苷酸的混合物。应当理解,使用本文所述的任何碱基编辑器的复合编辑的特征可以应用于使用本文提供的任何碱基编辑器的方法的任何组合。还应当理解,使用本文所述的任何碱基编辑器的复合编辑可以包括多个核碱基对的顺序编辑。

[1060] 在一些实施例中,所述多个核碱基对在一个或多个基因中。在一些实施例中,所述多个核碱基对在同一基因中。在一些实施例中,所述一个或多个基因中的至少一个基因位于不同基因座中。

[1061] 在一些实施例中,所述编辑是编辑至少一个蛋白质编码区中的所述多个核碱基对。在一些实施例中,所述编辑是编辑至少一个蛋白质非编码区中的所述多个核碱基对。在一些实施例中,所述编辑是编辑至少一个蛋白质编码区和至少一个蛋白质非编码区中的所述多个核碱基对。

[1062] 在一些实施例中,所述编辑与一种或多种向导多核苷酸结合。在一些实施例中,所述碱基编辑器系统可以包括一个或多个碱基编辑器系统。在一些实施例中,所述碱基编辑器系统可以包含一个或多个碱基编辑器系统与单个向导多核苷酸结合。在一些实施例中,所述碱基编辑器系统可包含一个或多个碱基编辑器系统与多个向导多核苷酸结合。在一些实施例中,所述编辑与具有单碱基编辑器系统的一种或多种向导多核苷酸结合。在一些实施例中,所述编辑与不需要PAM序列靶向结合目标多核苷酸序列的至少一种向导多核苷酸结合。在一些实施例中,所述编辑与需要PAM序列靶向结合目标多核苷酸序列的至少一种向导多核苷酸结合。在一些实施例中,所述编辑与不需要PAM序列靶向结合目标多核苷酸序列的至少一种向导多核苷酸,和需要PAM序列靶向结合目标多核苷酸序列的至少一种向导多核苷酸的混合物结合。应当理解,使用本文所述的任何碱基编辑器的复合编辑的特征可以应用于使用本文提供的任何碱基编辑器的方法的任何组合。还应当理解,所述编辑可以包括多个核碱基对的顺序编辑。

[1063] 在一些实施例中,能够对一个或多个基因中的多个核碱基对进行複合编辑的碱基编辑器系统包含本文所述的ABE8碱基编辑器变体之一。在一些实施例中,能够複合编辑一个或多个基因中的多个核碱基对的碱基编辑器系统包含ABE7碱基编辑器之一。在一些实施例中,与包含ABE7碱基编辑器之一的能够複合编辑的碱基编辑器系统相比,包含本文所述的ABE8碱基编辑器变体之一的能够複合编辑的碱基编辑器系统具有更高的複合编辑效率。在一些实施例中,包含本文所述的ABE8碱基编辑器变体之一的能够进行複合编辑的碱基编辑器系统,与包含ABE7碱基编辑器之一的能够进行複合编辑的碱基编辑器系统相比,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%、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%、至少高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%更高的複合编辑效率。在一些实施例中,包含本文所述的ABE8碱基编辑器变体之一的能够进行複合编辑的碱基编辑器系统,与包含ABE7碱基编辑器之一的能够进行複合编辑的碱基编辑器系统相比,具有至少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4.0倍、至少4.5倍、至少5.0倍、至少5.5倍或至少6.0倍更高的複合编辑效率。

[1064] 具有内部插入的融合蛋白

[1065] 本文提供了融合蛋白,其包含与核酸可编程核酸结合蛋白(例如napDNAbp)融合的异源多肽。异源多肽可以是在天然或野生型napDNAbp多肽序列中未发现的多肽。所述异源多肽可以在napDNAbp的C-末端、napDNAbp的N-末端融合至napDNAbp,或插入到napDNAbp的内部位置。在一些实施例中,所述异源多肽插入到napDNAbp的内部位置。

[1066] 在一些实施例中,所述异源多肽是脱氨酶或其功能片段。例如,融合蛋白可包含侧接Cas9或Cas12(例如Cas12b/C2c1)多肽的N-末端片段和C-末端片段的脱氨酶(例如腺苷脱氨酶)。融合蛋白中的脱氨酶可以是腺苷脱氨酶。在一些实施例中,所述腺苷脱氨酶是TadA(例如TadA7.10或TadA*8)。在一些实施例中,所述TadA是TadA*8。如本文所述的TadA序列(例如TadA7.10或TadA*8)是适用于上述融合蛋白的脱氨酶。

[1067] 所述脱氨酶可以是环状置换脱氨酶。例如,所述脱氨酶可以是环状排列的腺苷脱氨酶。在一些实施例中,所述脱氨酶是环状排列的TadA,在TadA参考序列中氨基酸残基编号116处环状排列。在一些实施例中,所述脱氨酶是环状排列的TadA,在TadA参考序列中氨基酸残基编号136处环状排列。在一些实施例中,所述脱氨酶是环状排列的TadA,在TadA参考序列中氨基酸残基编号65处环状排列。

[1068] 所述融合蛋白可包含多于一种脱氨酶。所述融合蛋白可包含例如1、2、3、4、5或更多种脱氨酶。在一些实施例中,所述融合蛋白包含一种脱氨酶。在一些实施例中,所述融合蛋白包含两种脱氨酶。融合蛋白中的所述两种或更多种脱氨酶可以是腺苷脱氨酶、胞苷脱氨酶或其组合。所述两种或更多种脱氨酶可以是同型二聚体。所述两种或更多种脱氨酶可以是异源二聚体。所述两种或多种脱氨酶可以串联插入到napDNAbp中。在一些实施例中,所述两种或更多种脱氨酶在napDNAbp中可能不串联。

[1069] 在一些实施例中,所述融合蛋白中的napDNAbp是Cas9多肽或其片段。所述Cas9多肽可以是变体Cas9多肽。在一些实施例中,所述Cas9多肽是Cas9切口酶(nCas9)多肽或其片段。在一些实施例中,所述Cas9多肽是核酸酶死亡的Cas9(dCas9)多肽或其片段。融合蛋白中的所述Cas9多肽可以是全长Cas9多肽。在一些情况下,融合蛋白中的所述Cas9多肽可能不是全长Cas9多肽。所述Cas9多肽可以在例如相对于天然存在的Cas9蛋白的N-末端或C-末端被截短。所述Cas9多肽可以是循环排列的Cas9蛋白。所述Cas9多肽可以是Cas9多肽的片段、部分或结构域,其仍然能够结合目标多核苷酸和向导核酸序列。

[1070] 在一些实施例中,所述Cas9多肽是化脓性链球菌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。

[1071] 融合蛋白的所述Cas9多肽可包含与天然存在的Cas9多肽至少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%相同的氨基酸序列。

[1072] 融合蛋白的所述Cas9多肽可包含与与下文阐述的Cas9氨基酸序列(下文称为“Cas9参考序列”)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%相同的氨基酸序列:

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFG
NIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD
VDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGN
LIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
[1073] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE
VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL
SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL
HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER
MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDH
IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
[1074] KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK
MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF
ATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVA
YSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE
QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA
[1075] PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (单下划线:HNH结构域;双下划线:RuvC结构域)

[1076] 在一些实施例中,融合蛋白中的所述napDNAbp是Cas12多肽,例如Cas12b/C2c1或其片段。所述Cas12多肽可以是变体Cas12多肽。

[1077] 所述异源多肽(例如脱氨酶)可以插入到napDNAbp(例如Cas9或Cas12(例如Cas12b/C2c1))的合适位置,例如,使得所述napDNAbp保持其结合目标多核苷酸和向导核酸的能力。脱氨酶(例如腺苷脱氨酶)可以插入到所述napDNAbp中而不损害脱氨酶(例如碱基编辑活性)或napDNAbp(例如结合目标核酸和向导核酸的能力)的功能。脱氨酶(例如腺苷脱氨酶)可以插入到所述napDNAbp中,例如无序区域或包含高温因子或B因子的区域,如晶体学研究所示。较不有序、无序或非结构化的蛋白质区域,例如溶剂暴露区域和环,可用于插入而不损害结构或功能。脱氨酶(例如腺苷脱氨酶)可以插入到所述napDNAbp的柔性环区或

溶剂暴露区中。在一些实施例中,所述脱氨酶(例如腺苷脱氨酶)插入所述Cas9或Cas12b/C2c1多肽的柔性环中。

[1078] 在一些实施例中,脱氨酶(例如腺苷脱氨酶)的插入位置通过对Cas9多肽的晶体结构的B因子分析来确定。在一些实施例中,所述脱氨酶(例如腺苷脱氨酶)插入包含高于平均B因子的所述Cas9多肽区域(例如与包含无序区域的总蛋白质或蛋白质结构域相比更高的B因子)。B因子或温度因子可以表示原子从其平均位置的波动(例如,由于温度相关的原子振动或晶格中的静态无序)。主链原子的高B因子(例如高于平均B因子)可以指示具有相对较高局部迁移率的区域。这样的区域可用于插入脱氨酶而不损害结构或功能。脱氨酶(例如腺苷脱氨酶)可以插入在具有Cα原子的残基的位置,其B因子比总蛋白质的平均B因子多50%、60%、70%、80%、90%、100%、110%、120、130%、140%、150%、160%、170%、180%、190%、200%或超过200%。脱氨酶(例如腺苷脱氨酶)可以插入在具有Cα原子的残基的位置,其B因子比包含所述残基的Cas9蛋白结构域的平均B因子多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B因子的Cas9多肽位置可以包括,例如,如上述Cas9参考序列中编号为768、792、1052、1015、1022、1026、1029、1067、1040、1054、1068、1246、1247和1248处的残基。包含高于平均B因子的Cas9多肽区域可以包括,例如,如上述Cas9参考序列中编号792-872、792-906和2-791处的残基。

[1079] 异源多肽(例如脱氨酶)可以插入到所述napDNAbp中的氨基酸残基处,该氨基酸残基选自:768、791、792、1015、1016、1022、1023、1026、1029、1040、1052、1054、1067、1068、1069、1246、1247和1248处,如在上述Cas9参考序列中编号,或另一个Cas9多肽中的相应氨基酸残基。在一些实施例中,所述异源多肽插入在下列氨基酸位置之间:768-769、791-792、792-793、1015-1016、1022-1023、1026-1027、1029-1030、1040-1041、1052-1053、1054-1055、1067-1068、1068-1069、1247-1248或1248-1249处,如在上述Cas9参考序列中编号,或其相应的氨基酸位置。在一些实施例中,所述异源多肽插入在下列氨基酸位置之间:769-770、792-793、793-794、1016-1017、1023-1024、1027-1028、1030-1031、1041-1042、1053-1054、1055-1056、1068-1069、1069-1070、1248-1249或1249-1250处,如在上述Cas9参考序列中编号,或其相应的氨基酸位置。在一些实施例中,异源多肽替换选自由以下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处,如在上述Cas9参考序列中编号,或另一个Cas9多肽中的相应氨基酸残基。应当理解,关于插入位置对上述Cas9参考序列的引用是为了说明的目的。如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多肽中相应位置的插入,例如Cas9切口酶(nCas9)、核酸酶死亡Cas9(dCas9)、Cas9变体缺少核酸酶结构域、截短的Cas9或缺少部分或完整HNH结构域的Cas9结构域。

[1080] 异源多肽(例如脱氨酶)可以插入到所述napDNAbp中的氨基酸残基处,该氨基酸残基选自:768、792、1022、1026、1040、1068和1247处,如在上述Cas9参考序列中编号,或另一Cas9多肽中的相应氨基酸残基。在一些实施例中,所述异源多肽插入在以下氨基酸位置之间:768-769、792-793、1022-1023、1026-1027、1029-1030、1040-1041、1068-1069或1247-1248,如上述Cas9参考序列中编号或其相应的氨基酸位置。在一些实施例中,所述异源多肽插入在以下氨基酸位置之间:769-770、793-794、1023-1024、1027-1028、1030-1031、1041-

1042、1069-1070或1248-1249,如上述Cas9参考序列中编号或其相应的氨基酸位置。在一些实施例中,所述异源多肽替换选自由以下组成的组的氨基酸残基:768、792、1022、1026、1040、1068和1247,如在上述Cas9参考序列中编号的,或另一Cas9多肽中的相应氨基酸残基。

[1081] 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多肽中的相应氨基酸残基处插入所述napDNAbp中。在一个实施例中,异源多肽(例如,脱氨酶)可以插入到所述napDNAbp中的氨基酸残基处,该氨基酸残基选自:1002、1003、1025、1052-1056、1242-1247、1061-1077、943-947、686-691、569-578、530-539和1060-1077,如在上述Cas9参考序列中编号的,或另一Cas9多肽中的相应氨基酸残基。所述脱氨酶(例如,腺苷脱氨酶)可插入所述残基的N-末端或C-末端或替换所述残基。在一些实施例中,陈述脱氨酶(例如腺苷脱氨酶)插入在陈述残基的C-末端。

[1082] 在一些实施例中,腺苷脱氨酶(例如TadA)插入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,如在上述Cas9参考序列中编号的,或另一Cas9多肽中的相应氨基酸残基。在一些实施例中,插入腺苷脱氨酶(例如TadA)以替换残基792-872、792-906或2-791,如以上Cas9参考序列中编号,或另一Cas9多肽中相应的氨基酸残基。在一些实施例中,所述腺苷脱氨酶插入在选自下组的氨基酸的N-末端:1015、1022、1029、1040、1068、1247、1054、1026、768、1067、1248、1052和1246,如以上Cas9参考序列中编号,或另一Cas9多肽中相应的氨基酸残基。在一些实施例中,所述腺苷脱氨酶插入在选自下组的氨基酸的C-末端:1015、1022、1029、1040、1068、1247、1054、1026、768、1067、1248、1052和1246,如以上Cas9参考序列中编号,或另一Cas9多肽中相应的氨基酸残基。在一些实施例中,插入所述腺苷脱氨酶以替换选自由以下组成的组的氨基酸:1015、1022、1029、1040、1068、1247、1054、1026、768、1067、1248、1052和1246,如以上Cas9参考序列中编号,或另一Cas9多肽中相应的氨基酸残基。

[1083] 在一些实施例中,所述脱氨酶(例如腺苷脱氨酶)插入在如以上Cas9参考序列中编号的氨基酸残基768处,或另一Cas9多肽中的相应氨基酸残基处。在一些实施例中,所述脱氨酶(例如腺苷脱氨酶)插入在如以上Cas9参考序列中编号的氨基酸残基768的N-末端,或另一Cas9多肽中的相应氨基酸残基。在一些实施例中,所述脱氨酶(例如腺苷脱氨酶)插入在如以上Cas9参考序列中编号的氨基酸残基768的C-末端,或另一Cas9多肽中的相应氨基酸残基。在一些实施例中,插入所述脱氨酶(例如腺苷脱氨酶)以替换如以上Cas9参考序列中编号的氨基酸残基768,或另一Cas9多肽中的相应氨基酸残基。

[1084] 在一些实施例中,所述脱氨酶(例如腺苷脱氨酶)插入在氨基酸残基791处或插入在氨基酸残基792处,如在以上Cas9参考序列中编号的,或在另一Cas9多肽中的相应氨基酸残基处。在一些实施例中,所述脱氨酶(例如腺苷脱氨酶)插入在氨基酸残基791的N-末端或插入在氨基酸792的N-末端,如在上述Cas9参考序列中编号的,或另一Cas9多肽中相应的氨基酸残基處。在一些实施例中,所述脱氨酶(例如腺苷脱氨酶)插入在氨基酸791的C-末端或插入在氨基酸792的N-末端,如在上述Cas9参考序列中编号的,或另一个Cas9多肽中相应的氨基酸残基。在一些实施例中,插入所述脱氨酶(例如腺苷脱氨酶)以替换氨基酸791,或插入以替换氨基酸792,如在上述Cas9参考序列中编号的,或另一Cas9多肽中的相应氨基酸残基。

[1085] 在一些实施例中,将脱氨酶(例如,腺苷脱氨素酶)插入如在上述Cas9参考序列中编号的氨基酸残基1016处,或另一Cas9多肽中的相应氨基酸残基。在一些实施例中,将所述脱氨酶(例如腺苷脱氨素酶)插入如在上述Cas9参考序列中编号的氨基酸残基1016的N-末端,或另一Cas9多肽中的相应氨基酸残基。在一些实施例中,所述脱氨酶(例如腺苷脱氨素酶)插入如在上述Cas9参考序列中编号的氨基酸残基1016的C-末端,或另一Cas9多肽中的相应氨基酸残基。在一些实施例中,插入所述脱氨酶(例如腺苷脱氨酶)以替换上述Cas9参考序列中编号的氨基酸残基1016,或另一Cas9多肽中的相应氨基酸残基。

[1086] 在一些实施例中,所述脱氨酶(例如腺苷脱氨酶)插入在氨基酸残基1022处,或插入在氨基酸残基1023处,如在以上Cas9参考序列中编号的,或另一Cas9多肽中的相应氨基酸残基。在一些实施例中,所述脱氨酶(例如腺苷脱氨酶)插入在氨基酸残基1022的N-末端或插入在氨基酸残基1023的N-末端,如在上述Cas9参考序列中编号,或另一Cas9多肽中相应的氨基酸残基。在一些实施例中,所述脱氨酶(例如腺苷脱氨酶)插入在氨基酸残基1022的C-末端或插入在氨基酸残基1023的C-末端,如在上述Cas9参考序列中编号,或另一Cas9多肽中相应的氨基酸残基。在一些实施例中,插入所述脱氨酶(例如腺苷脱氨酶)以替换氨基酸残基1022,或插入以替换氨基酸残基1023,如在上述Cas9参考序列中编号,或另一Cas9多肽中的相应氨基酸残基。

[1087] 在一些实施例中,将所述脱氨酶(例如腺苷脱氨酶)插入在氨基酸残基1026处,或者插入在氨基酸残基1029处,如在上述Cas9参考序列中编号,或在另一Cas9多肽中的相应氨基酸残基。在一些实施例中,将所述脱氨酶(例如腺苷脱氨酶)插入氨基酸残基1026的N-末端,或者插入氨基酸残基1029的N-末端,如在上述Cas9参考序列中,或另一Cas9多肽中相应的氨基酸残基。在一些实施例中,将所述脱氨酶(例如腺苷脱氨酶)插入氨基酸残基1026的C-末端,或者插入氨基酸残基1029的C-末端,如上述Cas9参考序列中编,或另一个Cas9多肽中的相应氨基酸残基。在一些实施例中,插入所述脱氨酶(例如腺苷脱氨酶)以替换氨基酸残基1026,或插入以替换氨基酸残基1029,如在上述Cas9参考序列中编号,或另一Cas9多肽中的相应氨基酸残基。

[1088] 在一些实施例中,将所述脱氨酶(例如腺苷脱氨酶)插入如在上述Cas9基准序列中编号的氨基酸残基1040处,或另一Cas9多肽中的相应氨基酸残基。在一些实施例中,将所述脱氨酶(例如腺苷脱氨素酶)插入如在上述Cas9基准序列中编号的氨基酸残基1040的N-末端,或另一个Cas9多肽中的相应氨基酸残基。在一些实施例中,将所述脱氨酶(例如腺苷脱氨酶)插入如在上述Cas9基准序列中编号的氨基酸残基1040的C-末端,或另一个Cas9多肽中的相应氨基酸残基。在一些实施例中,插入所述脱氨酶(例如腺苷脱氨酶)以替换上述Cas9基准序列中的编号的氨基酸残基1040,或另一Cas9多肽中的相应氨基酸残基。

[1089] 在一些实施例中,将所述脱氨酶(例如腺苷脱氨酶)插入氨基酸残基1052,或者在氨基酸残基1054处插入,如在上述Cas9基准序列中编号,或另一Cas9多肽中的相应氨基酸残基。在一些实施例中,将所述脱氨酶(例如腺苷脱氨酶)插入氨基酸残基1052的N-末端,或者插入氨基酸残基1054的N-末端,如在上述Cas9参考序列中,或相应的另一Cas9多肽中的氨基酸残基。在一些实施例中,将所述脱氨酶(例如腺苷脱氨素酶)插入氨基酸残基1052的C-末端,或者插入氨基酸残基1054的C-末端,如在上述Cas9参考序列中,或另一Cas9多肽中相应的氨基酸残基。在一些实施例中,将所述脱氨酶(例如腺苷脱氨酶)插入以替换氨基酸

残基1052,或插入以替换氨基酸残基1054,如在上述Cas9参考序列中,或另一Cas9多肽中的相应氨基酸残基。

[1090] 在一些实施例中,将所述脱氨酶(例如腺苷脱氨酶)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,如在上述Cas9参考序列中编号,或另一Cas9多肽中的相应氨基酸残基。在一些实施例中,将所述脱氨酶(例如,腺苷脱氨酶)插入氨基酸残基1067处的N-末端,或插入氨基酸残基1068处的N-末端,或插入氨基酸残基1069处的N-末端,如在上述Cas9参考序列中编号,或另一Cas9多肽中的相应氨基酸残基。在一些实施例中,所述脱氨酶(例如腺苷脱氨酶)插入在氨基酸残基1067处的C-末端,或插入在氨基酸残基1068处的C-末端,或插入在氨基酸残基1069处的C-末端,如在上述Cas9参考序列中编号的,或另一Cas9多肽中的相应氨基酸残基。在一些实施例中,插入所述脱氨酶(例如腺苷脱氨酶)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,如在上述Cas9参考序列中编号的,或另一个Cas9多肽中的相应氨基酸残基。

[1091] 在一些实施例中,所述脱氨酶(例如腺苷脱氨酶)插入在氨基酸残基1246处,或插入在氨基酸残基1247处,或插入在氨基酸残基1248处,如上述Cas9参考序列中编号的,或另一Cas9多肽中的相应氨基酸残基。在一些实施例中,所述脱氨酶(例如腺苷脱氨酶)插入在氨基酸残基1246处的N-末端,或插入在氨基酸残基1247处的N-末端,或插入在氨基酸残基1248处的N-末端,如上述Cas9参考序列中编号的,或另一Cas9多肽中的相应氨基酸残基处。在一些实施例中,所述脱氨酶(例如腺苷脱氨酶)插入在氨基酸残基1246处的C-末端,或插入在氨基酸残基1247处的C-末端,或插入在氨基酸残基1248处的C-末端,如在上述Cas9参考序列中编号的,或另一个Cas9多肽中的相应氨基酸残基处。在一些实施例中,所述脱氨酶(例如腺苷脱氨酶)被插入以替换氨基酸残基1246,或被插入以替换氨基酸残基1247,或被插入以替换氨基酸残基1248,如在上述Cas9参考序列中编号的,或另一个Cas9多肽中的相应氨基酸残基。

[1092] 在一些实施例中,异源多肽(例如脱氨酶)插入到Cas9多肽的柔性环中。所述柔性环部分可以从由以下组成的组中选择:530-537、569-570、686-691、943-947、1002-1025、1052-1077、1232-1247或1298-1300,如上述Cas9参考序列中编号的,或另一Cas9多肽中的相应氨基酸残基处。所述柔性环部分可以从由以下组成的组中选择:1-529、538-568、580-685、692-942、948-1001、1026-1051、1078-1231或1248-1297,如上述Cas9参考序列中编号的,或另一Cas9多肽中的相应氨基酸残基处。

[1093] 可以将异源多肽(例如腺嘌呤脱氨酶)插入以下对应于氨基酸残基的Cas9多肽区域:1017-1069、1242-1247、1052-1056、1060-1077、1002-1003、943-947、530-537、568-579、686-691、1242-1247、1298-1300、1066-1077、1052-1056或1060-1077,如上述Cas9参考序列中编号的,或另一Cas9多肽中的相应氨基酸残基处。

[1094] 可以插入异源多肽(例如腺嘌呤脱氨酶)代替Cas9多肽的缺失区域。所述缺失区域可以对应于Cas9多肽的N-末端或C-末端部分。在一些实施例中,所述缺失区域对应于如以上Cas9参考序列中编号的残基792-872,或另一Cas9多肽中的相应氨基酸残基处。在一些实施例中,所述缺失区域对应于如以上Cas9参考序列中编号的残基792-906,或另一Cas9多肽中的相应氨基酸残基处。在一些实施例中,所述缺失区域对应于上述Cas9参考序列中编号的残基2-791,或另一Cas9多肽中的对应氨基酸残基处。在一些实施例中,所述缺失区域对

应于上述Cas9参考序列中编号的残基1017-1069,或其相应的氨基酸残基处。

[1095] 示例性内部融合碱基编辑器在下表10A中提供:

[1096] 表10A:Cas9蛋白中的插入基因座

[1097]

BE ID	修饰	其他 ID
IBE001	Cas9 TadA 插入 1015	ISLAY01
IBE002	Cas9 TadA 插入 1022	ISLAY02
IBE003	Cas9 TadA 插入 1029	ISLAY03
IBE004	Cas9 TadA 插入 1040	ISLAY04
IBE005	Cas9 TadA 插入 1068	ISLAY05
IBE006	Cas9 TadA 插入 1247	ISLAY06
IBE007	Cas9 TadA 插入 1054	ISLAY07
IBE008	Cas9 TadA 插入 1026	ISLAY08
IBE009	Cas9 TadA 插入 768	ISLAY09
IBE020	delta HNH TadA 792	ISLAY20
IBE021	N-term 融合单 TadA 螺旋截短 165-端	ISLAY21

[1098]

IBE029	TadA-循环置换 116 插入 1067	ISLAY29
IBE031	TadA-循环置换 136 插入 1248	ISLAY31
IBE032	TadA-循环置换 136 插入 1052	ISLAY32
IBE035	delta 792-872 TadA 插入	ISLAY35
IBE036	delta 792-906 TadA 插入	ISLAY36
IBE043	TadA-循环置换 65 插入 1246	ISLAY43
IBE044	TadA ins C-端截短 2 791	ISLAY44

[1099] 异源多肽(例如脱氨酶)可以插入到Cas9多肽的结构域或功能域内。异源多肽(例如脱氨酶)可以插入在Cas9多肽的两个结构域或功能域之间。例如,在从Cas9多肽中删除结构域之后,可以插入异源多肽(例如脱氨酶)代替Cas9多肽的结构域或功能域。所述Cas9多肽的结构域或功能域可以包括例如RuvC I、RuvC II、RuvC III、Rec1、Rec2、PI或HNH。

[1100] 在一些实施例中,所述Cas9多肽缺乏一个或多个选自由以下组成的组的结构域:RuvC I、RuvC II、RuvC III、Rec1、Rec2、PI或HNH结构域。在一些实施例中,所述Cas9多肽缺乏核酸酶结构域。在一些实施例中,所述Cas9多肽缺乏HNH结构域。在一些实施例中,所述Cas9多肽缺少HNH结构域的一部分,使得Cas9多肽具有降低或消除的HNH活性。

[1101] 在一些实施例中,所述Cas9多肽包含核酸酶结构域的缺失,并且插入所述脱氨酶以替代核酸酶结构域。在一些实施例中,所述HNH结构域被删除并且所述脱氨酶被插入在其

位置上。在一些实施例中,一个或多个所述RuvC结构域被删除并且所述脱氨酶被插入在其位置上。

[1102] 包含异源多肽的融合蛋白的可以侧接napDNAbp的N-末端和C-末端片段。在一些实施例中,所述融合蛋白包含侧接Cas9多肽的N-末端片段和C-末端片段的脱氨酶。所述N-末端片段或C-末端片段可以结合目标多核苷酸序列。所述N-末端片段的C-末端或C-末端片段的N-末端可包含所述Cas9多肽的柔性环的一部分。所述N-末端片段的C-末端或C-末端片段的N-末端可包含所述Cas9多肽的α-螺旋结构的一部分。所述N-末端片段或C-末端片段可包含DNA结合结构域。所述N-末端片段或C-末端片段可包含RuvC结构域。所述N-末端片段或C-末端片段可包含HNH结构域。在一些实施例中,所述N-末端片段和C-末端片段均不包含HNH结构域。

[1103] 在一些实施例中,当所述融合蛋白使目标核碱基脱氨基时,所述N-末端Cas9片段的C-末端包含接近目标核碱基的氨基酸。在一些实施例中,当所述融合蛋白使目标核碱基脱氨基时,所述C-末端Cas9片段的N-末端包含接近目标核碱基的氨基酸。不同脱氨酶的插入位置可以不同,以使所述目标核碱基与所述N-末端Cas9片段的C-末端或C-末端Cas9片段的N-末端的氨基酸接近。例如,ABE的插入位置可以在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Cas9参考序列中所编号,或在另一个Cas9多肽中对应的氨基酸残基处。

[1104] 所述融合蛋白的N-末端Cas9片段(即所述融合蛋白中位于脱氨酶侧翼的N-末端Cas9片段)可包含Cas9多肽的N-末端。所述融合蛋白的N-末端Cas9片段可包含长度至少约:100、200、300、400、500、600、700、800、900、1000、1100、1200或1300个氨基酸。融合蛋白的N-末端Cas9片段可包含对应于以下氨基酸残基的序列:1-56、1-95、1-200、1-300、1-400、1-500、1-600、1-700、1-718、1-765、1-780、1-906、1-918或1-1100,如上述Cas9参考序列中所编号,或在另一个Cas9多肽中对应的氨基酸残基处。所述N-末端Cas9片段可包含的序列包含至少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%与下列氨基酸残基序列的同一性:1-56、1-95、1-200、1-300、1-400、1-500、1-600、1-700、1-718、1-765、1-780、1-906、1-918或1-1100,如上述Cas9参考序列中所编号,或在另一个Cas9多肽中对应的氨基酸残基处。

[1105] 所述融合蛋白的C-末端Cas9片段(即融合蛋白中位于脱氨酶侧翼的C端Cas9片段)可包含Cas9多肽的C-末端。所述融合蛋白的C-末端Cas9片段可包含长度至少约:100、200、300、400、500、600、700、800、900、1000、1100、1200或1300个氨基酸。融合蛋白的C-末端Cas9片段可包含对应于以下氨基酸残基的序列:1099-1368、918-1368、906-1368、780-1368、765-1368、718-1368、94-1368或56-1368,如上述Cas9参考序列中所编号,或在另一个Cas9多肽中对应的氨基酸残基处。所述N-末端Cas9片段可包含的序列包含至少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%与下列氨基酸残基序列的同一性:1099-1368、918-1368、906-1368、780-1368、765-1368、718-1368、94-1368或56-1368,如上述Cas9参考序列中所编号,或在另一个Cas9多肽中对应的氨基酸残基处。

[1106] 融合蛋白的N-末端Cas9片段和C-末端Cas9片段合在一起可能不对应于全长天然存在的Cas9多肽序列,例如,如上述Cas9参考序列中所述。

[1107] 本文所述的融合蛋白可以通过减少非靶位点(例如脱靶位点)的脱氨基作用,例如减少全基因组的假脱氨作用来实现靶向脱氨基作用。本文所述的融合蛋白可以在非靶位点处以减少的旁观者脱氨基作用实现靶向脱氨基作用。相较于例如包含融合至Cas9多肽N末端或C末端的脱氨酶的末端融合蛋白,非预期脱氨基或脱靶脱氨基可以减少至少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%。相较于例如包含融合至Cas9多肽N末端或C末端的脱氨酶的末端融合蛋白,非预期脱氨基或脱靶脱氨基可以减少至少一倍、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。

[1108] 在一些实施例中,融合蛋白的所述脱氨酶(例如腺苷脱氨酶)使R环范围内的不超过两个核碱基脱氨基。在一些实施例中,融合蛋白的所述脱氨酶使R环范围内不超过三个核碱基脱氨。在一些实施例中,融合蛋白的所述脱氨酶使R环范围内不超过2、3、4、5、6、7、8、9或10个核碱基脱氨基。R环是三链核酸结构,包括DNA:RNA杂交体、DNA:DNA或RNA:RNA互补结构以及与单链DNA相关的结构。如本文所用,当目标多核苷酸与CRISPR复合物或碱基编辑复合物接触时可形成R环,其中向导多核苷酸的一部分,例如,向导RNA,与目标多核苷酸的一部分杂交并置换,例如。目标DNA。在一些实施例中,R环包含间隔序列和目标DNA互补序列的杂交区域。R环区可以是约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个核碱基对长度。在一些实施例中,R环区的长度为约20个核碱基对。应当理解,如本文所用,R-环区不限于与向导多核苷酸杂交的目标DNA链。例如,R环区域内的目标核碱基的编辑可以针对包含与向导RNA互补链的DNA链,或者可以针对作为与向导RNA互补的链的相反链的DNA链。在一些实施例中,在所述R环区域中编辑包括将非互补链(原型间隔链)上的核碱基编辑为目标DNA序列中的向导RNA。

[1109] 本文所述的融合蛋白可以在不同于规范碱基编辑的编辑窗口中实现靶标脱氨基。在一些实施例中,目标核碱基位于目标多核苷酸序列中PAM序列上游约1至约20个碱基。在一些实施例中,目标核碱基位于目标多核苷酸序列中PAM序列上游的约2至约12个碱基。在一些实施例中,目标核碱基距离PAM序列或位于其上游约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个至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。在一些实施例中,目标核碱基位于AM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个碱基对。在一些实施例中,目标核碱基是PAM序列上游的约1、2、3、4、5、6、7、8或9个碱基对。在一些实施例中,目标核碱基是PAM序列上游约2、3、4或6个碱基对。

[1110] 所述融合蛋白可包含多于一种异源多肽。例如,所述融合蛋白可以另外包含一个或多个UGI结构域和/或一种或多种核定位信号。所述两个或多个异源结构域可以串联插入。所述两个或更多个异源结构域可以插入在这样的位置,使得它们在NapDNAbp中不串联。

[1111] 融合蛋白可包含所述脱氨酶和napDNAbp多肽之间的接头。所述接头可以是肽或非肽接头。例如所述接头可以是XTEN、(GGGS)n、(GGGGS)n、(G)n、(EAAAK)n、(GGS)n、SGSETPGTSESATPES。在一些实施例中,所述融合蛋白包含N-末端Cas9片段和脱氨酶之间的接头。在一些实施例中,所述融合蛋白包含C-末端Cas9片段和脱氨酶之间的接头。在一些实施例中,所述napDNAbp的N-末端和C-末端片段通过接头与脱氨酶连接。在一些实施例中,所述N-末端和C-末端片段在没有接头的情况下连接到脱氨酶结构域。在一些实施例中,所述融合蛋白包含N-末端Cas9片段与脱氨酶之间的接头,但不包含C-末端Cas9片段与脱氨酶之间的接头。在一些实施例中,所述融合蛋白包含C-末端Cas9片段与脱氨酶之间的接头,但不包含N-末端Cas9片段与脱氨酶之间的接头。

[1112] 在其他实施例中,所述Cas12多肽的N-末端或C-末端片段包含核酸可编程DNA结合结构域或RuvC结构域。在其他实施例中,所述融合蛋白在Cas12多肽和催化结构域之间含有接头。在其他实施例中,所述接头的氨基酸序列是GGSGGS或GSSGSETPGTSESATPESSG。在其他实施例中,所述接头是刚性接头。在上述方面的其他实施方式中,所述接头由GGAGGCTCTGGAGGAAGC或GGCTCTTCTGGATCTGAAACACCTGGCACAAGCGAGAGCGCCACCCCTGAGAGCTCTGGC编码。

[1113] 包含侧接有Cas9或Cas12多肽的N-末端和C-末端片段的异源催化结构域的融合蛋白也可用于本文所述方法中的碱基编辑。包含Cas9或Cas12和一个或多个脱氨酶结构域,例如腺苷脱氨酶,或包含侧接有Cas9或Cas12序列的腺苷脱氨酶结构域的融合蛋白也可用于目标序列的高度特异性和有效的碱基编辑。在一个实施例中,嵌合Cas9或Cas12融合蛋白包含插入Cas12多肽内的异源催化结构域。

[1114] 在各种实施例中,所述催化结构域具有DNA修饰活性(例如脱氨酶活性),例如腺苷脱氨酶活性。在一些实施例中,腺苷脱氨酶是TadA(例如,TadA7.10)。在一些实施例中,TadA是TadA*8。在其他实施例中,所述融合蛋白含有一个或多个催化结构域。在其他实施例中,一个或多个催化结构域中的至少一个插入所述Cas12多肽内或融合在Cas12 N-末端或C-末端。在其他实施例中,所述一个或多个催化结构域中的至少一个插入所述Cas12多肽的环、α螺旋区、非结构化部分或溶剂可及部分内。在其他实施例中,所述Cas12多肽是Cas12a、Cas12b、Cas12c、Cas12d、Cas12e、Cas12g、Cas12h或Cas12i。在其他实施例中,所述Cas12多肽与久氏芽孢杆菌Cas12b、嗜热淀粉芽孢杆菌Cas12b、芽孢杆菌属V3-13 Cas12b或嗜酸脂环杆菌Cas12b具有至少约85%的氨基酸序列同一性。在其他实施例中,所述Cas12多肽与久氏芽孢杆菌Cas12b、嗜热淀粉芽孢杆菌Cas12b、芽孢杆菌属V3-13 Cas12b或嗜酸脂环杆菌Cas12b具有至少约90%的氨基酸序列同一性。在其他实施例中,所述Cas12与久氏芽孢杆菌Cas12b、嗜热淀粉芽孢杆菌Cas12b、芽孢杆菌属V3-13 Cas12b或嗜酸脂环杆菌Cas12b具有至少约95%的氨基酸序列同一性。在其他实施例中,所述Cas12多肽包含或基本上由以下的

片段组成:久氏芽孢杆菌Cas12b、嗜热淀粉芽孢杆菌Cas12b、芽孢杆菌属V3-13Cas12b或嗜酸脂环杆菌Cas12b。

[1115] 在其他实施例中,所述催化结构域插入在BhCas12b的氨基酸位置153-154、255-256、306-307、980-981、1019-1020、534-535、604-605或344-345之间,或Cas12a、Cas12c、Cas12d、Cas12e、Cas12g、Cas12h或Cas12i的相应氨基酸残基之间。在其他实施例中,所述催化结构域插入在BhCas12b的氨基酸P153和S154之间。在其他实施例中,所述催化结构域插入在BhCas12b的氨基酸K255和E256之间。在其他实施例中,所述催化结构域插入在BhCas12b的氨基酸D980和G981之间。在其他实施例中,所述催化结构域插入在BhCas12b的氨基酸K1019和L1020之间。在其他实施例中,所述催化结构域插入在BhCas12b的氨基酸F534和P535之间。在其他实施例中,所述催化结构域插入在BhCas12b的氨基酸K604和G605之间。在其他实施例中,所述催化结构域插入在BhCas12b的氨基酸H344和F345之间。在其他实施例中,所述催化结构域插入在BvCas12b的氨基酸位置147和148、248和249、299和300、991和992或1031和1032之间或Cas12a、Cas12c、Cas12d、Cas12e、Cas12g、Cas12h或Cas12i的相应氨基酸残基之间。在其他实施例中,所述催化结构域插入在BvCas12b的氨基酸P147和D148之间。在其他实施例中,所述催化结构域插入在BvCas12b的氨基酸G248和G249之间。在其他实施例中,所述催化结构域插入在BvCas12b的氨基酸P299和E300之间。在其他实施例中,所述催化结构域插入在BvCas12b的氨基酸G991和E992之间。在其他实施例中,所述催化结构域插入在BvCas12b的氨基酸K1031和M1032之间。在其他实施例中,所述催化结构域插入在AaCas12b的氨基酸位置157和158、258和259、310和311、1008和1009或1044和1045之间,或Cas12a、Cas12c、Cas12d、Cas12e、Cas12g、Cas12h或Cas12i的相应氨基酸残基之间。在其他实施例中,所述催化结构域插入在AaCas12b的氨基酸P157和G158之间。在其他实施例中,所述催化结构域插入在AaCas12b的氨基酸V258和G259之间。在其他实施例中,所述催化结构域插入在AaCas12b的氨基酸D310和P311之间。在其他实施例中,所述催化结构域插入在AaCas12b的氨基酸G1008和E1009之间。在其他实施例中,所述催化结构域插入在AaCas12b的氨基酸G1044和K1045之间。

[1116] 在其他实施例中,所述融合蛋白包含核定位信号(例如双分核定位信号)。在其他实施例中,所述核定位信号的氨基酸序列是MAPKKKRKVGIHGVPAA。在上述方面的其他实施方式中,所述核定位信号由以下序列编码:

[1117] ATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCC。在其他实施例中,所述Cas12b多肽含有使RuvC结构域的催化活性沉默的突变。在其他实施例中,所述Cas12b多肽含有D574A、D829A和/或D952A突变。在其他实施例中,所述融合蛋白还包含标签(例如流感血凝素标签)。

[1118] 在一些实施例中,所述融合蛋白包含具有内部融合的核碱基编辑结构域(例如脱氨酶结构域的全部或部分,例如腺苷脱氨酶结构域)的napDNAbp结构域(例如Cas12衍生结构域)。在一些实施例中,所述napDNAbp是Cas12b。在一些实施例中,所述碱基编辑器包含一个BhCas12b结构域,在表10B中提供的基因座处插入了一个内部融合的TadA*8结构域。

[1119] 表10B:Cas12b蛋白中的插入基因座

[1120]

BhCas12b	插入部位	插入 aa 之间
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	**插入部位**	**插入 aa 之间**
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	**插入部位**	**插入 aa 之间**

[1121]

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

[1122] 作为非限制性实例,可以将腺苷脱氨酶(例如ABE8.13)插入到BhCas12b中以产生有效编辑核酸序列的融合蛋白(例如ABE8.13-BhCas12b)。

[1123] 示例性但非限制性的融合蛋白在美国临时申请号62/852,228和62/852,224中有所描述,其内容通过引用整体并入本文。

[1124] 编辑核酸的方法

[1125] 本公开的一些方面提供了用于编辑核酸的方法。在一些实施例中,该方法是用于编辑编码蛋白质的核酸分子的核碱基(例如双链DNA序列的碱基对)的方法。在一些实施例

中,该方法包括以下步骤:a)使核酸的目标区域(例如双链DNA序列)与包含碱基编辑器和向导核酸(例如gRNA)的复合物接触,b)诱导所述目标区域的链分离,c)将目标区域的单链中的所述目标核碱基对的第一个核碱基转化为第二个核碱基,和d)使用nCas9切割不超过一条的所述目标区域的链,其中与第一个核碱基互补的第三个核碱基被与第二个核碱基互补的第四个核碱基替换。在一些实施例中,该方法导致所述核酸中少于20%的插入缺失形成。应当理解,在一些实施例中,步骤b被省略。在一些实施例中,该方法导致插入缺失形成率低于19%、18%、16%、14%、12%、10%、8%、6%、4%、2%、1%、0.5%、0.2%或0.1%。在一些实施例中,该方法进一步包括用与第四个核碱基互补的第五个核碱基替换第二个核碱基,从而产生预期的编辑碱基对(例如G·C到A·T)。在一些实施例中,至少5%的预期碱基对被编辑。在一些实施例中,至少10%、15%、20%、25%、30%、35%、40%、45%或50%的预期碱基对被编辑。

[1126] 在一些实施例中,目标核苷酸中的预期产物与非预期产物的比率为至少2:1、5:1、10:1、20:1、30:1、40:1、50:1、60:1、70:1、80:1、90:1、100:1或200:1或更高。在一些实施例中,预期突变与插入缺失形成的比率大于1:1、10:1、50:1、100:1、500:1或1000:1或更高。在一些实施例中,所述切割的单链(切口链)与向导核酸杂交。在一些实施例中,所述切割的单链与包含第一核碱基的链相反。在一些实施例中,所述碱基编辑器包含dCas9结构域。在一些实施例中,所述碱基编辑器保护或结合非编辑链。在一些实施例中,所述预期的编辑碱基对在PAM位点的上游。在一些实施例中,预期的编辑碱基对是所述PAM位点上游的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。在一些实施例中,预期的编辑碱基对在PAM位点的下游。在一些实施例中,预期的编辑碱基对是所述PAM位点下游的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。在一些实施例中,该方法不需要规范的(例如NGG)PAM位点。在一些实施例中,所述核碱基编辑器包含接头。在一些实施例中,所述接头的长度为1-25个氨基酸。在一些实施例中,所述接头长度为5-20个氨基酸。在一些实施例中,接头的长度为10、11、12、13、14、15、16、17、18、19或20个氨基酸。在一个实施例中,所述接头的长度为32个氨基酸。在另一个实施例中,“长接头”的长度为至少约60个氨基酸。在其他实施例中,所述接头长度在约3-100个氨基酸之间。在一些实施例中,目标区域包括目标窗口,其中目标窗口包括目标核碱基对。在一些实施例中,所述目标窗口包含1-10个核苷酸。在一些实施例中,所述目标窗口的长度为1-9、1-8、1-7、1-6、1-5、1-4、1-3、1-2或1个核苷酸。在一些实施例中,所述目标窗口长度为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。在一些实施例中,预期的编辑碱基对在所述目标窗口内。在一些实施例中,所述目标窗口包括预期的编辑碱基对。在一些实施例中,使用本文提供的任何碱基编辑器来执行该方法。

[1127] 在一些实施例中,本公开提供了用于编辑核苷酸(例如编码蛋白质的基因中的SNP)的方法。在一些实施例中,本公开提供了一种用于编辑双链DNA序列的核碱基对的方法。在一些实施例中,该方法包括a)使双链DNA序列的目标区域与包含碱基编辑器和向导核酸(例如gRNA)的复合物接触,其中目标区域包含目标核碱基对,b)诱导所述目标区域的链分离,c)将目标区域的单链中的所述目标核碱基对的第一个核碱基转化为第二个核碱基,d)切割所述目标区域的不超过一条链,其中与第一个核碱基互补的第三个核碱基被与第二个核碱基互补的第四个核碱基替换,并且第二个核碱基被替换为与第四个核碱基互补的第

五个核碱基,从而产生预期编辑碱基对,其中产生预期编辑碱基对的效率至少为5%。应当理解,在一些实施例中,步骤b被省略。在一些实施例中,至少5%的预期碱基对被编辑。在一些实施例中,至少10%、15%、20%、25%、30%、35%、40%、45%或50%的预期碱基对被编辑。在一些实施例中,该方法导致少于19%、18%、16%、14%、12%、10%、8%、6%、4%、2%、1%、0.5%、0.2%或少于0.1%的插入缺失形成。在一些实施例中,目标核苷酸处的预期产物与非预期产物的比率为至少2:1、5:1、10:1、20:1、30:1、40:1、50:1、60:1、70:1、80:1、90:1、100:1或200:1或更高。在一些实施例中,预期突变与插入缺失形成的比率高于1:1、10:1、50:1、100:1、500:1或1000:1或更多。在一些实施例中,所述切割的单链与向导核酸杂交。在一些实施例中,所述切割的单链与包含第一核碱基的链相反。在一些实施例中,预期的编辑碱基对在PAM位点的上游。在一些实施例中,预期的编辑碱基对是所述PAM位点上游的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。在一些实施例中,预期的编辑碱基对在PAM位点的下游。在一些实施例中,预期的编辑碱基对是所述PAM位点下游的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。在一些实施例中,该方法不需要规范的(例如NGG)PAM位点。在一些实施例中,所述接头的长度为1-25个氨基酸。在一些实施例中,所述接头长度为5-20个氨基酸。在一些实施例中,所述接头的长度为10、11、12、13、14、15、16、17、18、19或20个氨基酸。在一些实施例中,目标区域包括目标窗口,其中目标窗口包括目标核碱基对。在一些实施例中,所述目标窗口包含1-10个核苷酸。在一些实施例中,所述目标窗口的长度为1-9、1-8、1-7、1-6、1-5、1-4、1-3、1-2或1个核苷酸。在一些实施例中,所述目标窗口长度为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。在一些实施例中,预期的编辑碱基对出现在所述目标窗口内。在一些实施例中,所述目标窗口包括预期的编辑碱基对。在一些实施例中,所述核碱基编辑器是本文提供的任何一种碱基编辑器。

[1128] 融合蛋白在宿主细胞中的表达

[1129] 本发明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实际上任何感兴趣的宿主细胞中表达,包括但不限于细菌、酵母、真菌、昆虫、植物和动物细胞。例如,编码本发明的腺苷脱氨酶的DNA可以通过基于cDNA序列为CDS的上游和下游设计合适的引物来克隆。所述克隆的DNA可以直接,或在需要时用限制酶消化后,或在添加合适的接头和/或核定位信号后,与编码碱基编辑系统的一种或多种额外组分的DNA连接。所述碱基编辑系统在宿主细胞中被翻译以形成复合体。

[1130] 编码本文所述的蛋白质结构域的DNA可以通过化学合成DNA获得,或者通过利用PCR法和Gibson组装法连接合成的部分重叠的寡聚DNA短链以构建编码其全长的DNA。通过化学合成或PCR方法或Gibson Assembly方法的组合构建全长DNA的优点是,可以根据引入DNA的宿主以CDS全长形式设计要使用的密码子。在异源DNA的表达中,通过将其DNA序列转化为宿主生物中频繁使用的密码子,预计蛋白质表达水平会增加可以使用作为所用宿主密码子使用频率的数据,例如在Kazusa DNA研究所的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),或者可以参考显示每个宿主中密码子使用频率的文件。参考获得的数据和要引入的DNA序列,可以将用于DNA序列的密码子中,在宿主中显示低使用频率的密码子转换为编码相同氨基酸并显示高使用频率的密码子。

[1131] 包含编码核酸序列识别模块和/或核酸碱基转化酶的DNA的表达载体可以,例如,

通过将DNA连接到合适表达载体中的启动子的下游来生产。

[1132] 作为表达载体,使用大肠杆菌来源的质粒(例如pBR322、pBR325、pUC12、pUC13);枯草芽孢杆菌来源的质粒(例如pUB110、pTP5、pC194);酵母来源的质粒(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);噬菌体如.lamda.phage等;昆虫病毒载体,例如杆状病毒等(例如BmNPV、AcNPV);动物病毒载体,例如逆转录病毒、痘苗病毒、腺病毒等。

[1133] 作为启动子,可以使用适合用于基因表达的宿主的任何启动子。在使用DSB的常规方法中,由于宿主细胞的存活率有时会因毒性而显着降低,因此期望通过使用诱导启动子在诱导开始时增加细胞数。然而,由于通过表达本发明的核酸修饰酶复合物也可以提供足够的细胞增殖,因此也可以没有限制地使用构成启动子。

[1134] 例如,当宿主是动物细胞时,使用SR.alpha.启动子、SV40启动子、LTR启动子、CMV(巨细胞病毒)启动子、RSV(劳斯肉瘤病毒)启动子、MoMuLV(莫罗尼鼠白血病病毒)LTR、HSV-TK(单纯疱疹病毒胸苷激酶)启动子等。其中,CMV启动子SR.alpha.启动子等是优选的。

[1135] 当宿主为大肠杆菌时,trp启动子、lac启动子、recA启动子、lamda.P.sub.L启动子、lpp启动子、T7启动子等是优选的。

[1136] 当宿主为芽孢杆菌属时,优选SPO1启动子、SPO2启动子、penP启动子等。

[1137] 当宿主为酵母菌时,优选Gal1/10启动子、PHO5启动子、PGK启动子、GAP启动子、ADH启动子等。

[1138] 当宿主是昆虫细胞时,多角体蛋白启动子、P10启动子等是优选的。

[1139] 当宿主是植物细胞时,CaMV35S启动子、CaMV19S启动子、NOS启动子等是优选的。

[1140] 作为表达载体,除上述载体外,还可以根据需要使用含有增强子、剪接信号、终止子、polyA添加信号、抗药性基因、营养缺陷型互补基因等选择标记、复制起点等。

[1141] 编码本文所述的蛋白质结构域的RNA可以通过以下方式制备,例如,以编码上述核酸序列识别模块的DNA和/或核酸碱基转化酶为模板,在本身已知的体外转录系统中转录为mRNA。

[1142] 本发明的融合蛋白可以通过将含有编码核酸序列识别模块和/或核酸碱基转化酶的DNA的表达载体导入宿主细胞,并培养宿主细胞而在细胞内表达。

[1143] 作为宿主,使用埃希氏菌属、芽孢杆菌属、酵母菌、昆虫细胞、昆虫、动物细胞等。

[1144] 作为大肠杆菌属,使用大肠杆菌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。

[1145] 作为芽孢杆菌属,使用枯草芽孢杆菌M1114[Gene,24,255(1983)]、枯草芽孢杆菌207-21[Journal of Biochemistry,95,87(1984)]等。

[1146] 作为酵母菌,使用酿酒酵母AH22,AH22R.sup.-、NA87-11A、DKD-5D、20B-12、粟酒裂殖酵母NCYC1913、NCYC2036、毕赤酵母KM71等。

[1147] 作为病毒为AcNPV时的昆虫细胞,白菜夜蛾幼虫来源的细胞建立系(草地贪夜蛾细胞;Sf细胞)、粉纹夜蛾中肠来源的MG1细胞、High Five.TM来源于粉纹夜蛾的卵的细胞、甘蓝夜蛾来源的细胞、Estigmena acrea来源的细胞等。当病毒为BmNPV时,家蚕来源的建立系

(家蚕N细胞;BmN细胞)等的细胞用作昆虫细胞。作为Sf细胞,例如使用Sf9细胞(ATCC CRL1711)、Sf21细胞[以上,In Vivo,13,213-217(1977)]等。

[1148] 作为昆虫,例如使用家蚕、果蝇、蟋蟀等的幼虫[Nature,315,592(1985)].

[1149] 作为动物细胞,细胞系如猴COS-7细胞、猴Vero细胞、中国仓鼠卵巢(CHO)细胞、dhfr基因缺陷型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、人类FL细胞等、使用人和其他哺乳动物的iPS细胞等多能干细胞,以及由各种组织制备的原代培养细胞。此外,还可以使用斑马鱼胚胎、非洲爪蟾卵母细胞等。

[1150] 作为植物细胞,使用由各种植物(例如水稻、小麦、玉米等谷物,西红柿、黄瓜、茄子等产品作物,如康乃馨、洋桔梗等园林植物,烟草、拟南芥等实验植物等)制备而成的悬浮培养细胞、愈伤组织、原生质体、叶段、根段等。

[1151] 上述所有宿主细胞都可以是单倍体(单倍体),也可以是多倍体(例如二倍体、三倍体、四倍体等)。在传统的突变引入方法中,突变原则上只引入一个同源染色体以产生异种基因类型。因此,除非发生显性突变,否则不会表达期望的表型,并且纯合子不方便地需要劳动和时间。相比之下,根据本发明,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,因此即使在隐性突变的情况下也可以在一代中表达期望的表型,这是非常有用的,因为可以解决传统方法的问题。

[1152] 根据宿主种类,表达载体可以通过已知的方法(例如溶菌酶法、感受态法、PEG法、CaCl2共沉淀法、电穿孔法、显微注射法、粒子枪法、脂质转染法、农杆菌法等)导入。

[1153] 大肠杆菌Escherichia coli可以根据以下描述的方法转化,例如Proc.Natl.Acad.Sci.USA,69,2110(1972),Gene,17,107(1982)等。

[1154] 芽孢杆菌属可以根据以下描述的方法引入载体中,例如Molecular&General Genetics,168,111(1979)等。

[1155] 酵母菌可以根据以下描述的方法导入载体中,例如Methods in Enzymology,194,182-187(1991),Proc.Natl.Acad.Sci.USA,75,1929(1978)等。

[1156] 根據例如Bio/Technology,6,47-55(1988)等描述的方法,可以将昆虫细胞和昆虫引入到载体中。

[1157] 动物细胞可以根据例如Cell Engineering additional volume 8,New Cell Engineering Experiment Protocol,263-267(1995)(published by Shujunsha),和Virology,52,456(1973)中描述的方法引入到载体中。

[1158] 导入了载体的细胞可以根据宿主的种类按照公知的方法进行培养。

[1159] 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用于培养的培养基。培养基优选含有转化体生长所需的碳源、氮源、无机物等。碳源的例子包括葡萄糖、糊精、可溶性淀粉、蔗糖等;氮源的例子包括无机或有机物质,例如铵盐、硝酸盐、玉米浆、蛋白胨、酪蛋白、肉提取物、豆饼、马铃薯提取物等。无机物的例子包括氯化钙、磷酸二氢钠、氯化镁等。培养基可以含有酵母提取物、维生素、生长促进因子等。培养基的pH值优选为约5-约8。

[1160] 作为培养大肠杆菌的培养基,优选例如含有葡萄糖、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]。必要时,例如可以将诸如3β-吲哚基丙烯酸之类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-约43℃下培养。必要时可进行

曝气和搅拌。

[1161] 芽孢杆菌属通常在约30-约40℃下培养。必要时可进行曝气和搅拌。

[1162] 用于培养酵母的培养基的例子包括Burkholder最小培养基[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-约8。培养通常在约20℃至约35℃下进行。必要时可进行曝气和搅拌。

[1163] 作为培养昆虫细胞或昆虫的培养基,使用例如适当地含有如灭活的10％牛血清等添加剂的Grace's昆虫培养基等[Nature,195,788(1962)]。培养基的pH值优选为约6.2至约6.4。培养通常在约27℃下进行。必要时可进行曝气和搅拌。

[1164] 作为用于培养动物细胞的培养基,例如使用含有约5-约20％胎牛血清的最低必需培养基(MEM)[Science,122,501(1952)]、Dulbecco's改良Eagle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-约8。培养通常在约30℃-约40℃下进行。必要时可进行曝气和搅拌。

[1165] 作为培养植物细胞的培养基,例如使用MS培养基、LS培养基、B5培养基等。培养基的pH优选为约5-约8。培养通常在约20℃-约30℃下进行。必要时可进行曝气和搅拌。

[1166] 当高等真核细胞,例如动物细胞、昆虫细胞、植物细胞等用作宿主细胞时,在诱导型启动子(例如金属硫蛋白启动子(由重金属离子诱导)、热休克蛋白启动子(由热休克诱导)、Tet-ON/Tet-OFF系统启动子(通过添加或去除四环素或其衍生物诱导),类固醇反应启动子(由类固醇激素或其衍生物诱导)等)的调控下,将编码本发明碱基编辑系统(例如包含腺苷脱氨酶变体)的DNA引入到宿主细胞中,在适当的阶段将诱导物质添加到培养基中(或从培养基中去除)以诱导核酸修饰酶复合物的表达,在给定时间内培养以进行碱基编辑,将突变引入目标基因,实现碱基编辑系统的暂时表达。

[1167] 原核细胞例如大肠杆菌等可以利用诱导型启动子。诱导型启动子的实例包括但不限于lac启动子(由IPTG诱导)、cspA启动子(由冷休克诱导)、araBAD启动子(由阿拉伯糖诱导)等。

[1168] 或者,当以动物细胞、昆虫细胞、植物细胞等高等真核细胞作为宿主细胞时,也可以利用上述诱导性启动子作为载体去除机制。即,载体安装有在宿主细胞中起作用的复制起点和编码复制所必需的蛋白质的核酸(例如SV40和大T抗原、oriP和EBNA-1等用于动物细胞),编码蛋白质的核酸的表达受上述诱导型启动子调控。因此,虽然载体在诱导物质存在的情况下可以自主复制,但当去除诱导物质时,自主复制不可用,载体自然随着细胞分裂而脱落(通过添加不能自主复制Tet-OFF系统载体中的四环素和强力霉素)。

[1169] 使用碱基编辑器的方法

[1170] 疾病相关基因和等位基因中点突变的校正为基因校正提供了新的策略,并应用于治疗学和基础研究。

[1171] 本公开提供了用于治疗被诊断患有与点突变相关或由其引起的疾病的个体的方法,所述点突变可以由本文提供的碱基编辑器系统校正。例如,在一些实施例中,提供了一种方法,其包括向患有此类疾病(例如由基因突变引起的疾病)的个体施用有效量的核碱基

编辑器(例如腺苷脱氨酶碱基编辑器)以校正疾病相关基因的点突变。本公开提供用于治疗与点突变相关或由其引起的GSD1a的方法,该点突变可以通过脱氨酶介导的基因编辑进行校正。基于本公开,可以用本文提供的策略和融合蛋白治疗的合适疾病对于本领域技术人员来说将是显而易见的。

[1172] 本文提供了使用碱基编辑器或碱基编辑器系统来编辑与疾病或病症相关的目标核苷酸序列中的核碱基的方法。在一些实施例中,所述碱基编辑器(例如包含腺苷脱氨酶和Cas9结构域)的活性导致点突变的校正。在一些实施例中,所述目标DNA序列包含与疾病或病症相关的G→A点突变,并且突变体A碱基的脱氨基产生与疾病或病症无关的序列。在一些实施例中,所述目标DNA序列包含与疾病或病症相关的T→C点突变,并且突变体C碱基的脱氨基产生与疾病或病症无关的序列。

[1173] 在一些实施例中,所述DNA序列编码蛋白质,并且点突变在密码子中,并且导致突变密码子编码的氨基酸与野生型密码子相比发生变化。在一些实施例中,所述突变体A的脱氨基导致突变体密码子编码的氨基酸发生变化。在一些实施例中,所述突变体A的脱氨基产生编码野生型氨基酸的密码子。在一些实施例中,所述突变体C的脱氨基导致突变体密码子编码的氨基酸发生变化。在一些实施例中,所述突变体C的脱氨基产生编码野生型氨基酸的密码子。在一些实施例中,所述个体已经或已经被诊断患有疾病或病症。

[1174] 在一些实施例中,本文提供的腺苷脱氨酶能够使DNA的脱氧腺苷残基脱氨基。本公开的其他方面提供了融合蛋白,其包含腺苷脱氨酶(例如,如本文所述的使DNA中的脱氧腺苷脱氨基的腺苷脱氨酶)和能够结合特定核苷酸序列的结构域(例如Cas9或Cpf1蛋白)。例如,所述腺苷可以转化为肌苷残基,其通常与胞嘧啶残基碱基配对。此类融合蛋白尤其可用于核酸序列的靶向编辑。此类融合蛋白可用于体外靶向编辑DNA,例如用于产生突变细胞或动物;用于引入靶向突变,例如用于校正离体细胞中的遗传缺陷,例如用于从个体获得的细胞中,所述细胞随后被重新引入相同或另一个体中;并且用于在体内引入靶向突变,例如,可以使用本文提供的核碱基编辑器处理遗传缺陷校正,或在G到A或T到C突变的疾病相关基因中引入失活突变。本公开提供了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、组成物、方法、试剂盒、系统等。

[1175] 使用核碱基编辑器靶向G6PC基因中的核苷酸

[1176] 如本文所述评估靶向G6PC基因中的核苷酸的核碱基编辑器的适用性。在一个实施例中,用编码本文所述的核碱基编辑器的一个或多个核酸分子连同少量编码报告基因(例如GFP)的载体转染、转导或以其他方式修饰感兴趣的单个细胞。这些细胞可以是永生化的人类细胞系,例如293T、K562或U20S。或者可以使用原代人类细胞。细胞还可以从个体或个体获得,例如从组织活检、手术、血液、血浆、血清或其他生物流体中获得。这些细胞可能与最终的细胞目标有关。

[1177] 如下文进一步描述的,可以使用病毒载体进行递送。在一个实施例中,转染可以使用脂质转染(例如Lipofectamine、Metafectamine或Fugene)或通过电穿孔进行。转染后,可以通过荧光显微镜或流式细胞术确定GFP的表达,以确认一致和高水平的转染。这些初步转染可以包含不同的核碱基编辑器,以确定哪种编辑器组合具有最大的活性。

[1178] 如本文所述评估核碱基编辑器的活性,即通过对目標基因测序以检测目標序列中的改变。对于桑格测序,纯化的PCR扩增子被克隆到质粒骨架中,转化、小量制备并用单一引

物测序。也可以使用下一代测序技术进行测序。使用下一代测序时,扩增子可能为300-500bp,预期切割位点不对称放置。PCR之后,可以将下一代测序接头和条形码(例如Illumina多重接头和索引)添加到扩增子的末端,例如用于高通量测序(例如在Illumina MiSeq上)。

[1179] 可以选择在初始测试中诱导最大水平目标特异性改变的融合蛋白进行进一步评估。

[1180] 在特定实施例中,所述核碱基编辑器用于靶向感兴趣的多核苷酸。在一个实施例中,将本发明的核碱基编辑器与用于靶向核酸序列例如含有GSD1a相关突变的G6PC多核苷酸的向导RNA一起递送至细胞(例如肝细胞),从而改变目标基因,即G6PC。

[1181] 在一些实施例中,碱基编辑器被向导RNA靶向以对感兴趣的基因(例如G6PC)的序列引入一个或多个编辑。在一些实施例中,将所述一种或多种改变引入葡萄糖-6-磷酸酶(G6PC)基因。在一些实施例中,所述一个或多个改变是R83C。在一些实施例中,所述一种或多种改变是Q347X。在一些实施例中,将所述改变引入代表性智人G6PC蛋白,见于NCBI参考序列号AAA16222.1,如下所示:

[1182]

```
  1 MEEGMNVLHD FGIQSTHYLQ VNYQDSQDWF ILVSVIADLR NAFYVLFPIW FHLQEAVGIK
 61 LLWVAVIGDW LNLVFKWILF GQRPYWWVLD TDYYSNTSVP LIKQFPVTCE TGPGSPSGHA
121 MGTAGVYYVM VTSTLSIFQG KIKPTYRFRC LNVILWLGFW AVQLNVCLSR IYLAAHFPHQ
181 VVAGVLSGIA VTETFSHIHS IYNASLKKYF LITFFLFSFA IGFYLLLKGL GVDLLWTLEK
241 AQRWCEQPEW VHIDTTPFAS LLKNLGTLFG LGLALNSSMY RESCKGKLSK WLPFRLSSIV
301 ASLVLLHVFD SLKPPSQVEL VFYVLSFCKS AVVPLASVSV IPYCLAQVLG QPHKKSL
```

[1183] 在一些实施例中,将所述改变引入代表性智人G6PC核酸序列,见于GenBank参考序列号U01120.1,如下提供:

```
   1 tagcagagca atcaccacca agcctggaat aactgcaagg gctctgctga catcttcctg
  61 aggtgccaag gaaatgagga tggaggaagg aatgaatgtt ctccatgact ttgggatcca
 121 gtcaacacat tacctccagg tgaattacca agactcccag gactggttca tcttggtgtc
 181 cgtgatcgca gacctcagga atgccttcta cgtcctcttc cccatctggt tccatcttca
 241 ggaagctgtg ggcattaaac tcctttgggt agctgtgatt ggagactggc tcaacctcgt
 301 ctttaagtgg attctctttg gacagcgtcc atactggtgg gttttggata ctgactacta
 361 cagcaacact tccgtgcccc tgataaagca gttccctgta acctgtgaga ctggaccagg
 421 gagcccctct ggccatgcca tgggcacagc aggtgtatac tacgtgatgg tcacatctac
 481 tctttccatc tttcagggaa agataaagcc gacctacaga tttcggtgct tgaatgtcat
 541 tttgtggttg ggattctggg ctgtgcagct gaatgtctgt ctgtcacgaa tctaccttgc
 601 tgctcatttt cctcatcaag ttgttgctgg agtcctgtca ggcattgctg ttacagaaac
 661 tttcagccac atccacagca tctataatgc cagcctcaag aaatattttc tcattacctt
 721 cttcctgttc agcttcgcca tcggatttta tctgctgctc aagggactgg gtgtagacct
 781 cctgtggact ctggagaaag cccagaggtg gtgcgagcag ccagaatggg tccacattga
 841 caccacaccc tttgccagcc tcctcaagaa cctgggcacg ctctttggcc tggggctggc
 901 tctcaactcc agcatgtaca gggagagctg caaggggaaa ctcagcaagt ggctcccatt
961 ccgcctcagc tctattgtag cctccctcgt cctcctgcac gtctttgact ccttgaaacc
1021 cccatcccaa gtcgagctgg tcttctacgt cttgtccttc tgcaagagtg cggtagtgcc
1081 cctggcatcc gtcagtgtca tcccctactg cctcgcccag gtcctgggcc agccgcacaa
1141 gaagtcgttg taagagatgt ggagtcttcg gtgtttaaag tcaacaacca tgccagggat
1201 tgaggaggac tactatttga agcaatgggc actggtattt ggagcaagtg acatgccatc
1261 cattctgccg tcgtggaatt aaatcacgga tggcagattg gagggtcgcc tggcttattc
1321 ccatgtgtga ctccagcctg ccctcagcac agactctttc agatggaggt gccatatcac
1381 gtacaccata tgcaagtttc ccgccaggag gtcctcctct ctctacttga atactctcac
1441 aagtagggag ctcactccca ctggaacagc ccattttatc tttgaatggt cttctgccag
1501 cccattttga ggccagaggt gctgtcagct caggtggtcc tcttttacaa tcctaatcat
1561 attgggtaat gtttttgaaa agctaatgaa gctattgaga aagacctgtt gctagaagtt
1621 gggttgttct ggattttccc ctgaagactt acttattctt ccgtcacata tacaaaagca
1681 agacttccag gtagggccag ctcacaagcc caggctggag atcctaactg agaattttct
1741 acctgtgttc attcttaccg agaaaaggag aaaggagctc tgaatctgat aggaaaagaa
1801 ggctgcctaa ggaggagttt ttagtatgtg gcgtatcatg caagtgctat gccaagccat
1861 gtctaaatgg ctttaattat atagtaatgc actctcagta atgggggacc agcttaagta
1921 taattaatag atggttagtg gggtaattct gcttctagta tttttttac tgtgcataca
```

[1184]

1981 tgttcatcgt atttccttgg atttctgaat ggctgcagtg acccagatat tgcactaggt
2041 caaaacattc aggtatagct gacatctcct ctatcacatt acatcatcct ccttataagc
2101 ccagctctgc tttttccaga ttcttccact ggctccacat ccacccccact ggatcttcag
2161 aaggctagag ggcgactctg gtggtgcttt tgtatgtttc aattaggctc tgaaatcttg
2221 ggcaaaatga caaggggagg gccaggattc ctctctcagg tcactccagt gttacttta
2281 attcctagag ggtaaatatg actcctttct ctatcccaag ccaaccaaga gcacattctt
2341 aaaggaaaag tcaacatctt ctctcttttt ttttttttt gagacagggt ctcactatgt
2401 tgcccaggct gctcttgaat tcctgggctc aagcagtcct cccaccctac cacagcgtcc
2461 cgcgtagctg gcatacaggt gcaagccact atgtccagct agccaactcc tccttgcctg
[1185] 2521 cttttctttt ttttttcttt tttgagacgg cgcacctatc acccaggctg gagtggagtg
2581 gcacgatctt ggctcactgc aacctcttcc tcctggttca agcgattctc atgtctcagc
2641 ctcctcagta gctaggacta ccggcgtgca ccaccatgcc aggctaattt ttatattttt
2701 agaattttag aagagatggg atttcatcat gttggccagg ctggtctcga actcctgacc
2761 tcaagtgatc cacctgcctt ggcctcccaa ggtgctagga ttacaggcat gagccaccgc
2821 accgggccct ccttgcctgt ttttcaatct catctgatat gcagagtatt tctgccccac
2881 ccacctaccc cccaaaaaaa gctgaagcct atttatttga aagtccttgt ttttgctact
2941 aattatatag tataccatac attatcattc aaaacaacca tcctgctcat aacatctttg
3001 aaaagaaaaa tatatatgtg cagtatttta ttaaagcaac attttattta agaataaagt
3061 cttgttaatt actatatttt agatgcaatg tgatc

[1186] 产生一个预期的突变

[1187] 在一些实施例中,本文提供的方法的目的是通过基因编辑恢复功能失调基因的功能。在一些实施例中,功能障碍基因的功能通过引入预期突变来恢复。在一些实施例中,本文提供的方法可用于破坏基因产物的正常功能。本文提供的核碱基编辑蛋白可以在体外验证基于基因编辑的人类治疗,例如,通过校正人类细胞培养物中的疾病相关突变。技术人员将理解,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)的融合蛋白可用于校正任何单点A到G或C到T的突变。在第一种情况下,所述突变体A到I的脱氨基校正了突变,在后一种情况下,与所述突变体T碱基配对的A脱氨基,然后进行一轮复制,校正了突变。

[1188] 在一些实施例中,本公开提供了碱基编辑器,其在核酸(例如个体基因组内的核酸)中有效产生预期突变,例如点突变,而不产生大量非预期突变,例如非预期点突变。在一些实施例中,预期突变是由与向导多核苷酸(例如gRNA)结合的特定碱基编辑器(例如腺苷碱基编辑器)产生的突变,所述向导多核苷酸(例如gRNA)专门设计用于产生预期突变。在一些实施例中,所述预期突变是与疾病或病症相关的突变。在一些实施例中,所述预期突变是与疾病或病症相关的腺嘌呤(A)到鸟嘌呤(G)的点突变。在一些实施例中,所述预期突变是与疾病或病症相关的胞嘧啶(C)到胸腺嘧啶(T)的点突变。在一些实施例中,所述预期突变是基因的编码区或非编码区内的腺嘌呤(A)到鸟嘌呤(G)的点突变。在一些实施例中,所述预期突变是基因的编码区或非编码区内的胞嘧啶(C)到胸腺嘧啶(T)的点突变。在一些实施例中,所述预期突变是产生终止密码子的点突变,例如基因编码区内的提前终止密码子。在一些实施例中,所述预期突变是消除终止密码子的突变。

[1189] 在一些实施例中,本文提供的任何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)。在一些实施例中,本文提供的任何碱基

编辑器能够产生预期突变与非预期突变的比率(例如预期的点突变:非预期的点突变)至少是1.5:1、至少是2:1、至少是2.5:1、至少是3:1、至少是3.5:1、至少是4:1、至少是4.5:1、至少是5:1、至少是5.5:1、至少是6:1、至少是6.5:1、至少是7:1、至少是7.5:1、至少是8:1、至少是10:1、至少是12:1、至少是15:1、至少是20:1、至少是25:1、至少是30:1、至少是40:1、至少是50:1、至少是100:1、至少是150:1、至少是200:1、至少是250:1、至少是500:1、或至少是1000:1或更多。

[1190] 碱基编辑器效率的细节在国际PCT申请号PCT/2017/045381(WO 2018/027078)和PCT/US2016/058344(WO 2017/070632)中进行了描述,它们各自通过引用整体并入本文。另见Komor,A.C.等人,"Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage"Nature 533,420-424(2016);Gaudelli,N.M.等人,"Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage"Nature 551,464-471(2017);and Komor,A.C.等人,"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),其全部内容以引用方式并入本文。

[1191] 在一些实施例中,使用本文提供的方法编辑一种或多种基因中的多个核碱基对导致形成至少一种预期突变。在一些实施例中,所述至少一种预期突变的所述形成导致致病突变的精确校正。应当理解,可以使用本文提供的任何方法或方法的组合来实现多重编辑。

[1192] 精确校正致病突变

[1193] 在一些实施例中,所述预期突变是对致病突变或致病突变的精确校正。所述致病突变可以是致病性单核苷酸多态性(SNP)或由SNP引起。例如,所述致病突变可以是基因编码的蛋白质中的氨基酸变化。在另一个例子中,所述致病突变可以是基因中的致病性SNP。所述精确校正可以是将所述致病突变恢复到其野生型状态。在一些实施例中,所述致病突变是与疾病或病症相关的G→A点突变,并且其中以A-to-G碱基编辑器(ABE)对突变体A碱基进行脱氨基,产生与疾病或病症不相关的序列。在一些实施例中,所述致病突变是C→T点突变。例如,可以通过将A-to-G碱基编辑器(ABE)靶向相反链并编辑致病性T核碱基的补体A来校正C→T点突变。碱基编辑器可以靶向致病性SNP或致病性SNP的互补序列。所述致病性或致病性突变和其他序列变异的描述命名法描述于den Dunnen,J.T.和Antonarakis,S.E."Mutation Nomenclature Extensions and Suggestions to Describe Complex Mutations:A Discussion."Human Mutation15:712(2000),其全部内容以引用方式并入本文。

[1194] 在一个特定的实施例中,所述疾病或病症是糖原贮积病1型(GSD1或冯吉尔克病)。在一些实施例中,所述疾病或病症是GSD1a。在一些实施例中,所述致病突变在G6PC基因中。在一些实施例中,所述致病突变是G6PC基因的Q347X。在一些实施例中,所述致病突变是G6PC基因的R83C。

[1195] 合成库

[1196] 本文提供融合蛋白库和使用其优化碱基编辑的方法,与规范碱基编辑器相比,该库允许替代的优选碱基编辑窗口。在一些实施例中,本公开提供了用于优化碱基编辑的蛋白库,其包含多种融合蛋白,其中所述多种融合蛋白中的每一种都包含侧接Cas9多肽的N-

末端片段和C-末端片段的脱氨酶,其中每一种所述融合蛋白中的N-末端片段不同于其余的多种融合蛋白的N-末端片段,或者其中每一种所述融合蛋白中的C-末端片段不同于其余的多种融合蛋白的C-末端片段,其中每一种所述融合蛋白的脱氨酶使目标核碱基在目标多核苷酸序列中的原始间隔区相邻基序(PAM)序列附近脱氨基,并且其中N-末端片段或C-末端片段结合目标多核苷酸序列。在一些实施例中,对于CRISPR R环内的每个核碱基,所述多种融合蛋白中的至少一种使核碱基脱氨基。在一些实施例中,对于距离PAM序列1至20个碱基对的目标多核苷酸内的每个核碱基,所述多种融合蛋白中的至少一种使核碱基脱氨基。在一些实施例中,本文提供包含允许优化碱基编辑的融合蛋白库的试剂盒。

[1197] 在一些实施例中,腺苷脱氨酶等位基因(例如TadA等位基因)的合成库可用于产生具有经修饰的碱基编辑效率和/或特异性的腺苷碱基编辑器。在一些实施例中,从合成文库产生的腺苷碱基编辑器包括更高的碱基编辑效率和/或特异性。在一些实施例中,与带有野生型TadA的腺苷碱基编辑器相比,从合成库生成的腺苷碱基编辑器表现出碱基编辑效率增加、碱基编辑特异性增加、脱靶编辑减少、旁观者编辑减少、插入缺失形成减少、和/或假编辑减少。在一些实施例中,与带有TadA*7.10的腺苷碱基编辑器相比,从合成库生成的腺苷碱基编辑器表现出碱基编辑效率增加、碱基编辑特异性增加、脱靶编辑减少、旁观者编辑减少、插入缺失形成减少、和/或假编辑减少。在一些实施例中,所述合成库包含ABE的随机化TadA部分。在一些实施例中,所述合成库在TadA的每个位置包含所有20个标准氨基酸取代。在一些实施例中,所述合成文库包含每个库成员1-2个核苷酸取代突变的平均频率。在一些实施例中,所述合成库包含在TadA*7.10中发现的背景突变。

[1198] 递送系统

[1199] 基于核酸的核碱基编辑器和gRNA的递送

[1200] 可以通过本领域已知的方法或如本文所述,在体外或体内将编码根据本公开的碱基编辑系统的核酸施用于个体或递送至细胞中。在一个实施例中,核碱基编辑器可以通过例如载体(例如病毒或非病毒载体)、基于非载体的方法(例如使用裸DNA、DNA复合物、脂质纳米颗粒)或其组合递送。

[1201] 编码核碱基编辑器的核酸可以以裸DNA或RNA的形式直接递送至细胞(例如造血细胞或其祖细胞、造血干细胞和/或诱导多能干细胞),例如通过转染或电穿孔,或者可以是与促进目标细胞摄取的分子(例如N-乙酰半乳糖胺)结合。也可以使用核酸载体,例如本文所述的载体。

[1202] 核酸载体可包含编码本文所述的融合蛋白的结构域的一个或多个序列。载体还可以包含编码信号肽的序列(例如用于核定位、核仁定位或线粒体定位),与编码蛋白质的序列缔合(例如插入或融合)。作为一个实例,核酸载体可以包括Cas9编码序列,其包括一个或多个核定位序列(例如来自SV40的核定位序列)和腺苷脱氨酶变体(例如ABE8)。

[1203] 核酸载体还可包括任何合适数量的调节/控制组件,例如启动子、增强子、内含子、聚腺苷酸化信号、Kozak共有序列或内部核糖体进入位点(IRES)。这些组件在本领域中是众所周知的。对于造血细胞,合适的启动子可以包括IFNβ或CD45。

[1204] 根据本公开的核酸载体包括重组病毒载体。示例性病毒载体在本文中阐述。也可以使用本领域已知的其他病毒载体。此外,病毒颗粒可用于递送核酸和/或肽形式的碱基编辑系统组分。例如,“空”病毒颗粒可以组装成包含任何合适的货物。病毒载体和病毒颗粒也

可以被设计成结合靶向配体来改变目标组织特异性。

[1205] 除了病毒载体,非病毒载体可用于递送编码根据本公开的基因组编辑系统的核酸。一类重要的非病毒核酸载体是纳米颗粒,它可以是有机的或无机的。纳米颗粒是本领域公知的。任何合适的纳米颗粒设计均可用于递送基因组编辑系统组分或编码此类组分的核酸。例如,有机(例如脂质和/或聚合物)纳米颗粒可适合用作本公开的某些实施例中的递送载体。用于纳米颗粒制剂和/或基因转移的示例性脂质示于表11(以下)。

[1206] 表11

[1207]

用于基因转移的脂质		
脂质	缩写	特征

[1208]

用于基因转移的脂质		
脂质	缩写	特征
1,2-二油酰基-sn-甘油-3-磷脂酰胆碱	DOPC	辅助者
1,2-二油酰基-sn-甘油-3-磷脂酰乙醇胺	DOPE	辅助者
胆固醇		辅助者
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
双十八烷基氨基甘醇精胺	DOGS	阳离子
N-(3-氨基丙基)-N,N-二甲基-2,3-双(十二烷氧基)-1-丙胺溴化物	GAP-DLRIE	阳离子
十六烷基三甲基溴化铵	CTAB	阳离子
6-鸟氨酸月桂氧己酯	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
2,3-二油氧基-N-[2(精胺甲酰胺乙基]-N,N-二甲基-1-三氟乙酸丙胺	DOSPA	阳离子
1,2-二油基-3-三甲基铵-丙烷	DOPA	阳离子
N-(2-羟乙基)-N,N-二甲基-2,3-双(十四烷氧基)-1-溴化丙胺	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'-二甲氨基乙烷)-氨基甲酰基]胆固醇	DC-Chol	阳离子
双胍盐胆固醇	BGTC	阳离子
1,3-双脱氧-2-(6-羧基-精基)-丙基酰胺	DOSPER	阳离子
二甲基十八烷基溴化铵	DDAB	阳离子
双十八烷基氨基甘氨酰精胺	DSL	阳离子
rac-[(2,3-二十八烷氧基丙基)(2-羟乙基)]-二甲基氯化铵	CLIP-1	阳离子
rac-[2(2,3-二十六烷基氧丙基-羟甲氧基)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
1,2-二硬脂氧基-N,N-二甲基-3-氨基丙烷	DSDMA	阳离子
1,2-二肉豆蔻酰-三甲基丙烷铵	DMTAP	阳离子
O,O'-二肉豆蔻基-N-赖氨酰天冬氨酸	DMKE	阳离子
1,2-二硬脂酰-sn-甘油-3-乙基磷酸胆碱	DSEPC	阳离子
N-软脂酰 D-赤型鞘氨醇氨基甲酰精胺	CCS	阳离子
N-t-叔丁基-NO-十四烷基-3-十四烷基氨基丙脒	diC14-amidine	阳离子
十八烯氧基[乙基-2-十七烯基-3 羟乙基]咪唑啉氯化物	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乙酰胺	RPR2091 20	阳离子

用于基因转移的脂质		
脂质	缩写	特征
1,2-二亚油氧基-3-二甲氨基丙烷	DLinDMA	阳离子
[1209] 2,2-二亚油基-4-二甲氨基乙基-[1,3]-二氧戊环	DLin-KC 2-DMA	阳离子
二亚油基-甲基-4-二甲氨基丁酸酯	DLin-MC 3-DMA	阳离子

[1210] 表12列出了用于基因转移和/或纳米颗粒制剂的示例性聚合物。

[1211] 表12

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(琥珀酰亚胺丙酸酯)	DSP
二甲基-3,3'-二硫代双丙亚氨酸酯	DTBP
聚(乙烯亚胺)双氨基甲酸酯	PEIC
聚(L-赖氨酸)	PLL
组氨酸经修饰的PLL	
聚(N-乙烯基吡咯烷酮)	PVP
聚(丙烯亚胺)	PPI
聚(酰氨基胺)	PAMAM
[1212]	
聚(酰胺基乙烯亚胺)	SS-PAEI
三亚乙基四胺	TETA
聚(β-氨基酯)	
聚(4-羟基-L-脯氨酸酯)	PHP
聚(烯丙胺)	
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-乙烯基溴化吡啶鎓)	
聚(磷腈)类	PPZ
聚(磷酸酯)类	PPE
聚(氨基磷酸酯)类	PPA

[1213]

用于基因转移的聚合物	
聚合物	缩写
聚(N-2-羟丙基甲基丙烯酰胺)	pHPMA
聚(2-(二甲氨基)乙基甲基丙烯酸酯)	pDMAEMA
聚(2-氨基乙基磷酸丙烯酯)	PPE-EA
壳聚糖	
半乳糖基化壳聚糖	
N-多酰化壳聚糖	
组蛋白	
胶原	
右旋精胺	D-SPM

[1214] 表13总结了编码本文所述融合蛋白的多核苷酸的递送方法。

[1215] 表13

[1216]

递送	载体/模式	递送至非分裂细胞	表达持续时间	基因组整合	递送的分子类型
物理	例如电穿孔、粒子枪、磷酸钙转染	是	短暂	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/否有修饰	RNA
	腺病毒	是	短暂	否	DNA
	腺相关病毒(AAV)	是	稳定	否	DNA
	痘苗病毒	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	短暂	取决于递送的内容	核酸和蛋白质

[1217]

递送	载体/模式	递送至非分裂细胞	表达持续时间	基因组整合	递送的分子类型
递送载体	聚合物纳米粒子	是	短暂	取决于递送的内容	核酸和蛋白质
	生物非病毒	是	短暂	否	核酸
	减毒细菌	是	短暂	否	核酸
	工程化噬菌体	是	短暂	否	核酸
	哺乳动物病毒样颗粒	是	短暂	否	核酸
	生物脂质体：红细胞幽灵和外泌体	是	短暂	否	核酸

[1218] 在另一方面,所述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,例如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,可以通过将核糖核蛋白(RNP)递送至细胞来实现。所述RNP包含与靶向gRNA复合的核酸结合蛋白,例如Cas9。可以使用已知方法将RNP递送至细胞,例如电穿孔、核转染或阳离子脂质介导的方法,例如Zuris,J.A.等人,2015,Nat.Biotechnology,33(1):73-80所报导。RNPs有利于在CRISPR基础编辑系统中使用,特别是对于难以转染的细胞,如原代细胞。此外,RNP还可以减轻细胞中蛋白质表达可能出现的困难,尤其是当真核启动子(例如CRISPR质粒中使用的CMV或EF1A)没有很好表达时。有利地,RNP的使用不需要将外源DNA递送到细胞中。此外,由于包含核酸结合蛋白和gRNA复合物的RNP会随时间降解,因此使用RNP有可能限制脱靶效应。以类似于基于质粒的技术的方式,RNP可用于递送结合蛋白(例如Cas9变体)和指导同源定向修复(HDR)。

[1219] 用于驱动碱基编辑器编码核酸分子表达的启动子可以包括AAV ITR。这有利于消除对额外启动子组件的需要,该组件会占据载体中的空间。释放的额外空间可用于驱动额外组件的表达,例如向导核酸或选择标记。ITR活性相对较弱,因此可用于降低因所选核酸酶过度表达而导致的潜在毒性。

[1220] 可以使用任何合适的启动子来驱动所述碱基编辑器和在适当情况下向导核酸的表达。对于到处可见的表达,可以使用的启动子包括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,合适的启动子可以包括:用于所有神经元的SynapsinI,用于兴奋性神经元的CaMKIIalpha,用于GABA能神经元的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。对于肺细胞表达,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。

[1221] 在一些实施例中,本公开的碱基编辑器具有足够小的尺寸,以允许单独的启动子

驱动所述碱基编辑器和相容的向导核酸在同一核酸分子内的表达。例如,载体或病毒载体可包含与编码所述碱基编辑器的核酸可操作连接的第一启动子,和可操作地连接至向导核酸的第二启动子。

[1222] 用于驱动向导核酸表达的启动子可以包括:Pol III启动子,例如U6或H1。使用Pol II启动子和内含子盒来表达gRNA腺相关病毒(AAV)。

[1223] 病毒载体

[1224] 因此,本文所述的碱基编辑器可以与病毒载体一起递送。在一些实施例中,本文公开的碱基编辑器可以在病毒载体中包含的核酸上编码。在一些实施例中,所述碱基编辑器系统的一种或多种组分可以在一种或多种病毒载体上编码。例如,碱基编辑器和向导核酸可以在单个病毒载体上编码。在其他实施例中,所述碱基编辑器和向导核酸在不同的病毒载体上编码。在任一情况下,所述碱基编辑器和向导核酸均可与启动子和终止子可操作地连接。病毒载体上编码的组分的组合可以由所选病毒载体的货物大小限制决定

[1225] 使用基于RNA或DNA病毒的系统来递送碱基编辑器,利用高度进化的过程将病毒靶向培养中或宿主中的特定细胞,并将病毒有效载荷运送到细胞核或宿主细胞基因组。病毒载体可以直接施用于培养中的细胞、患者(体内),或者它们可以用于体外处理细胞,并且可以任选地将经修饰的细胞施用于患者(离体)。常规的基于病毒的系统可以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载体。逆转录病毒、慢病毒和腺相关病毒基因转移方法可以整合到宿主基因组中,通常会导致插入的转基因长期表达。此外,在许多不同的细胞类型和目标组织中都观察到了高转导效率。

[1226] 病毒载体可包括慢病毒(例如基于HIV和FIV的载体)、腺病毒(例如AD100)、逆转录病毒(例如马洛尼鼠白血病病毒MML-V)、疱疹病毒载体(例如HSV-2)和腺相关病毒(AAV)或其他质粒或病毒载体类型,特别是使用来自例如美国专利第8,454,972号(腺病毒的制剂、剂量)、美国专利第8,404,658号(AAV的制剂、剂量)和美国专利第5,846,946号(DNA质粒的制剂、剂量)以及来自涉及慢病毒、AAV和腺病毒的临床试验的临床试验和出版物。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和涉及AAV的临床试验。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和涉及腺病毒的临床试验。对于质粒递送,给药途径、制剂和剂量可以如美国专利第5,846,946号和涉及质粒的临床研究。剂量可以基于或外推到平均70公斤的个体(例如成年人男性),并且可以针对不同体重和物种的患者、个体、哺乳动物进行调整。给药频率在医学或兽医从业者(例如医师、兽医)的范围内,这取决于通常的因素,包括患者或个体的年龄、性别、一般健康状况、其他状况以及正在解决的特定状况或症状。所述病毒载体可以注射到感兴趣的组织中。对于细胞类型特异性碱基编辑,所述碱基编辑器和可选向导核酸的表达可由细胞类型特异性启动子驱动。

[1227] 逆转录病毒的趋向性可以通过掺入外来包膜蛋白来改变,扩大目标细胞的潜在目标群体。慢病毒载体是能够转导或感染非分裂细胞并通常产生高病毒滴度的逆转录病毒载体。因此,逆转录病毒基因转移系统的选择将取决于目标组织。逆转录病毒载体由顺式作用的长末端重复序列组成,其具有高达6-10kb的外来序列的包装容量。最小的顺式作用LTR足以复制和包装载体,然后用于将治疗基因整合到目标细胞中以提供永久的转基因表达。广泛使用的逆转录病毒载体包括基于鼠白血病病毒(MuLV)、长臂猿白血病病毒(GaLV)、猿猴免疫缺陷病毒(SIV)、人类免疫缺陷病毒(HIV)及其组合的那些(见例如Buchscher等人,

J.Virol.66:2731-2739(1992);Johann等人,J.Virol.66:1635-1640(1992);Sommnerfelt等人,Virol.176:58-59(1990);Wilson等人,J.Virol.63:2374-2378(1989);Miller等人,J.Virol.65:2220-2224(1991);PCT/US94/05700).

[1228] 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整合到目标细胞中。例如,长度大于9kb的逆转录病毒载体与较小的病毒载体相比,会导致病毒滴度较低。在一些方面,本公开内容的碱基编辑器具有足够的大小,使其能够通过逆转录病毒载体有效地包装和递送到目标细胞中。在一些实施例中,碱基编辑器的大小使得即使在与向导核酸和/或可靶向核酸酶系统的其他组分一起表达时,也允许有效包装和递送。

[1229] 在首选暂时表达的应用中,可以使用基于腺病毒的系统。基于腺病毒的载体能够在许多细胞类型中具有非常高的转导效率并且不需要细胞分裂。使用这样的载体,已经获得了高滴度和表达水平。该载体可以在相对简单的系统中大量生产。腺相关病毒("AAV")载体也可用于用目标核酸转导细胞,例如在核酸和肽的体外生产中,以及用于体内和离体基因治疗程序(见例如West等人,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载体的构建在许多出版物中有所描述,包括美国专利号5,173,414;Tratschin等人,Mol.Cell.Biol.5:3251-3260(1985);Tratschin,等人,Mol.Cell.Biol.4:2072-2081(1984);Hermonat&Muzyczka,PNAS 81:6466-6470(1984);和Samulski等人,J.Virol.63:03822-3828(1989)。

[1230] AAV是一种小型的单链DNA依赖性病毒,属于细小病毒家族。4.7kb野生型(wt)AAV基因组由两个基因组成,分别编码四种复制蛋白和三种衣壳蛋白,两侧各接145bp反向末端重复序列(ITR)。病毒粒子由三种衣壳蛋白Vp1、Vp2和Vp3组成,它们以1:1:10的比例从相同的开放阅读框产生,但来自差异剪接(Vp1)和替代翻译起始位点(分别为Vp2和Vp3)。Vp3是病毒体中最丰富的亚基,参与定义病毒向性的细胞表面受体识别。已在Vp1的独特N末端鉴定出一个在病毒感染性中起作用的磷脂酶结构域。

[1231] 与wt AAV类似,重组AAV(rAAV)利用顺式作用的145bp ITR侧接载体转基因盒,提供高达4.5kb的外来DNA包装。感染后,rAAV可以表达本发明的融合蛋白,并且通过以环状头尾串联体的附加型存在而持续存在而不整合到宿主基因组中。尽管有许多使用该系统在体外和体内成功的rAAV实例,当基因编码序列的长度等于或大于wt AAV基因组时,有限的包装能力限制了AAV介导的基因递送的使用。

[1232] 可以根据应用选择病毒载体。例如,对于体内基因传递,AAV可能优于其他病毒载体。在一些实施例中,AAV允许低毒性,这可能是由于纯化方法不需要可以激活免疫反应的细胞颗粒的超速离心。在一些实施例中,AAV允许引起插入诱变的可能性很低,因为它不整合到宿主基因组中。腺病毒通常用作疫苗,因为它们诱导强烈的免疫原性反应。病毒载体的包装容量会限制可以包装到载体中的碱基编辑器的大小。

[1233] AAV的包装容量约为4.5Kb或4.75Kb,包括两个145碱基反向末端重复序列(ITR)。这意味着公开的碱基编辑器以及启动子和转录终止子可以适合单个病毒载体。大于4.5或4.75Kb的结构会导致病毒产量显着降低。比如SpCas9很大,基因本身就超过4.1Kb,很难打包成AAV。因此,本公开的实施例包括利用长度比常规碱基编辑器短的公开碱基编辑器。在

一些示例中,碱基编辑器小于4kb。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。在一些实施例中,所公开的碱基编辑器的长度为4.5kb或更小。

[1234] AAV可以是AAV1、AAV2、AAV5或其任意组合。可以根据要靶向的细胞选择AAV的类型;例如,可以选择AAV血清型1、2、5或混合衣壳AAV1、AAV2、AAV5或其任何组合用于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.等人,J.Virol.82:5887-5911(2008)中找到。

[1235] 慢病毒是复杂的逆转录病毒,具有在有丝分裂和有丝分裂后细胞中感染和表达其基因的能力。最常见的慢病毒是人类免疫缺陷病毒(HIV),它使用其他病毒的包膜糖蛋白来靶向广泛的细胞类型。

[1236] 慢病毒可以如下制备。克隆pCasES10(包含慢病毒转移质粒骨架)后,将低传代(p=5)的HEK293FT接种到T-75烧瓶中,在转染前一天在含有10%胎牛血清且不含抗生素的DMEM中达到50%汇合。20小时后,将培养基更换为OptiMEM(无血清)培养基,4小时后进行转染。用10μg慢病毒转移质粒(pCasES10)和以下包装质粒转染细胞:5μg pMD2.G(VSV-g假型)和7.5μg psPAX2(gag/pol/rev/tat)。可以在4mL OptiMEM中使用阳离子脂质递送剂(50μl Lipofectamine 2000和100ul Plus试剂)进行转染。6小时后,将培养基更换为含10%胎牛血清的不含抗生素的DMEM。这些方法在细胞培养过程中使用血清,但优选无血清方法。

[1237] 慢病毒可以如下纯化。48小时后收获病毒上清液。上清液首先清除碎屑,然后通过0.45μm低蛋白结合(PVDF)过滤器过滤。然后将它们在超速离心机中以24,000rpm的速度旋转2小时。病毒颗粒在50μl DMEM中于4℃下重悬过夜。然后等分并立即在-80℃下冷冻。

[1238] 在另一个实施例中,还考虑了基于马传染性贫血病毒(EIAV)的最小非灵长类慢病毒载体。在另一个实施例中,RetinoStat.RTM.,一种基于马传染性贫血病毒的慢病毒基因治疗载体,其表达预期通过视网膜下注射递送的血管抑制蛋白内皮抑制素和血管抑制素。在另一个实施例中,考虑使用自灭活的慢病毒载体。

[1239] 该系统的任何RNA,例如向导RNA或碱基编辑器编码的mRNA,都可以以RNA的形式递送。可以使用体外转录生成碱基编辑器编码mRNA。例如,可以使用包含以下组件的PCR盒合成核酸酶mRNA:T7启动子、可选的kozak序列(GCCACC)、核酸酶序列和3'UTR,例如来自β珠蛋白-polyA尾的3'UTR。该盒可用于T7聚合酶的转录。向导多核苷酸(例如gRNA)也可以使用体外转录从包含T7启动子的盒中转录,然后是序列"GG"和向导多核苷酸序列。

[1240] 为了增强表达并降低可能的毒性,可以修饰碱基编辑器编码序列和/或向导核酸以包括一种或多种经修饰的核苷,例如使用伪U或5-甲基-C。

[1241] AAV载体的小包装容量使得大量基因的传递和/或大型生理调控组件的使用具有挑战性。例如,可以通过将要递送的蛋白质分成两个或多个片段来解决这些挑战,其中N-末端片段与分裂的内含肽-N融合,C-末端片段与分裂的内含肽-C融合。然后将这些片段打包成两个或多个AAV载体。如本文所用,"内含肽"是指连接侧翼N-末端和C-末端外显子(例如要连接的片段)的自剪接蛋白质内含子(例如肽)。某些内含肽用于连接异源蛋白质片段的用途描述于例如Wood等人,J.Biol.Chem.289(21);14512-9(2014)中。例如当与分离的蛋白质片段融合时,内含肽IntN和IntC相互识别,将自身剪断并同时连接它们所融合的蛋白质

片段的侧翼N-和C-末端外显肽,从而从两个蛋白质片段重建全长蛋白质。其他合适的内含肽对本领域技术人员来说是显而易见的。

[1242] 本发明的融合蛋白片段的长度可以不同。在一些实施例中,蛋白质片段的长度为2个氨基酸至约1000个氨基酸。在一些实施例中,蛋白质片段的长度为约5个氨基酸至约500个氨基酸。在一些实施例中,蛋白质片段的长度范围为约20个氨基酸至约200个氨基酸。在一些实施例中,蛋白质片段的长度范围为约10个氨基酸至约100个氨基酸。其他长度的合适蛋白质片段对本领域技术人员来说是显而易见的。

[1243] 在一个实施例中,双AAV载体是通过将一个大的转基因表达盒分成两个独立的一半(5'和3'端,或头和尾)而产生的,其中该盒的每一半都被包装在一个AAV载体中(<5kb)。然后通过两种双AAV载体共感染同一细胞实现全长转基因表达盒的重新组装,然后:(1)5'和3'基因组之间的同源重组(HR)(双AAV重叠载体);(2)ITR介导的5'和3'基因组的尾对头串联(双AAV反式剪接载体);或(3)这两种机制的组合(双AAV混合载体)。在体内使用双AAV载体导致全长蛋白质的表达。双AAV载体平台的使用代表了一种有效且可行的基因转移策略,适用于大小大于4.7kb的转基因。

[1244] 内含肽

[1245] 在一些实施例中,核酸酶(例如Cas9)的一部分或片段与内含肽融合。所述核酸酶可以融合到内含肽的N-末端或C-末端。在一些实施例中,融合蛋白的一部分或片段与内含肽融合并与AAV衣壳蛋白融合。所述内含肽、核酸酶和衣壳蛋白可以任何排列融合在一起(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)。在一些实施例中,所述内含肽的N-末端与融合蛋白的C-末端融合,并且内含肽的C-末端与AAV衣壳蛋白的N-末端融合。

[1246] 内含肽(中间蛋白)是在各种不同生物体中发现的自动加工域,其执行称为蛋白质剪接的过程。蛋白质剪接是一个多步骤的生化反应,包括肽键的断裂和形成。虽然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,但内含肽也可用于化学操作几乎任何多肽骨架。

[1247] 在蛋白质剪接中,所述内含肽通过切割两个肽键将自身从前体多肽中切除,从而通过形成新的肽键连接侧翼外显肽(外部蛋白质)序列。这种重排发生在翻译后(或可能是共同翻译)。内含肽介导的蛋白质剪接自发发生,只需要内含肽结构域的折叠。

[1248] 大约5%的内含肽是分裂内含肽,它们被转录和翻译为两个独立的多肽,N-内含肽和C-内含肽,每个都融合到一个外显肽上。翻译后,内含肽片段自发地非共价组装成典型的内含肽结构以进行蛋白质反式剪接。所述蛋白质剪接的机制需要一系列酰基转移反应,导致内含肽-外显肽连接处的两个肽键断裂,并在N-和C-外显肽之间形成新的肽键。该过程通过激活连接N-外显肽和内含肽N-末端的肽键而启动。几乎所有内含肽在其N-末端都有一个半胱氨酸或丝氨酸,它们攻击C-末端N-外显肽残基的羰基碳。这种N到O/S酰基转移是由保守性的苏氨酸和组氨酸(称为TXXH基序)以及常见的天冬氨酸促进的,这导致形成线性(硫)酯中间体。接下来,该中间体通过第一个C-外显肽残基(+1)的亲核攻击进行反式(硫代)酯化,该残基是半胱氨酸、丝氨酸或苏氨酸。产生的支化(硫)酯中间体通过独特的转化得到分解:内含肽的高度保守性的C端天冬酰胺的环化。该过程由组氨酸(在高度保守的HNF基序中发现)和倒数第二个组氨酸促进,也可能涉及天冬氨酸。这种琥珀酰亚胺形成反应从反应复

合物中切除内含肽,并留下通过非肽键连接的外含肽。这种结构以独立于内含肽的方式迅速重排成稳定的肽键。

[1249] 在一些实施例中,碱基编辑器(例如ABE、CBE)的N-末端片段与分裂的内含肽-N融合并且C-末端片段与分裂的内含肽-C融合。然后将这些片段打包成两个或多个AAV载体。例如,某些内含肽用于连接异源蛋白质片段的用途描述于Wood等人,J.Biol.Chem.289(21);14512-9(2014)中。例如,当与分离的蛋白质片段融合时,所述内含肽IntN和IntC相互识别,将自身剪断并同时连接它们所融合的蛋白质片段的侧翼N-和C-末端外显肽,从而重建全长来自两个蛋白质片段的蛋白质。其他合适的内含肽对本领域技术人员来说是显而易见的。

[1250] 在一些实施例中,ABE在SpCas9的选定区域内的Ala、Ser、Thr或Cys残基处分裂成N-和C-末端片段。这些区域对应于由Cas9晶体结构分析确定的环区域。每个片段的N-末端与内含肽-N融合,每个片段的C-末端在氨基酸位置S303、T310、T313、S355、A456、S460、A463、T466、S469、T472、T474、C574、S577、A589和S590处与内含肽C融合,按以下顺序以粗体大写字母表示。

[1251]

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 llSdilrvnT eiTkaplsas mikrydehhq dltllkalvr qqlpekykei ffdqSkngya
361 gyidggasqe efykfikpil ekmdgteell vklnredllr kqrtfdngsi phqihlgelh
421 ailrrqedfy pflkdnreki ekiltfripy yvgplArgnS rfAwmTrkSe eTiTpwnfee
481 vvdkgasaqs fiermtnfdk nlpnekvlpk hsllyeyftv yneltkvkyv tegmrkpafl
541 sgeqkkaivd llfktnrkvt vkqlkedyfk kieCfdSvei sgvedrfnAS 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
[1252] 1261 qhkhyldeii eqisefskrv iladanldkv lsaynkhrdk pireqaenii hlftltnlga
1321 paafkyfdtt idrkrytstk evldatlihq sitglyetri dlsqlggd

[1253] 使用核碱基编辑器靶向突变

[1254] 如本文所述评估靶向突变的核碱基编辑器的适用性。在一个实施例中,用碱基编辑系统连同少量编码报告基因(例如GFP)的载体转导单个感兴趣的细胞。这些细胞可以是本领域已知的任何细胞系,包括永生化人细胞系,例如293T、K562或U20S。或者,可以使用原代细胞(例如人)。这样的细胞可能与最终的细胞目标相关。

[1255] 可以使用病毒载体进行递送。在一个实施例中,转染可以使用脂质转染(例如Lipofectamine或Fugene)或通过电穿孔进行。转染后,可以通过荧光显微镜或流式细胞术确定GFP的表达,以确认一致和高水平的转染。这些初步转染可以包含不同的核碱基编辑器,以确定哪种编辑器组合具有最大的活性。

[1256] 如本文所述评估核碱基编辑器的活性,即通过对细胞基因组进行测序以检测目标序列中的改变。对于桑格测序,纯化的PCR扩增子被克隆到质粒骨架中,转化、小量制备并用单一引物测序。也可以使用下一代测序技术进行测序。使用下一代测序时,扩增子可能为300-500bp,预期切割位点不对称放置。PCR之后,可以将下一代测序接头和条形码(例如Illumina多重接头和索引)添加到扩增子的末端,例如用于高通量测序(例如在Illumina MiSeq上)。

[1257] 可以选择在初始测试中诱导最大水平目标特异性改变的融合蛋白进行进一步评估。

[1258] 在特定实施例中,所述核碱基编辑器用于靶向感兴趣的多核苷酸。在一个实施例中,将本发明的核碱基编辑器与用于靶向细胞基因组内感兴趣的突变的向导RNA一起递送至细胞(例如造血细胞或其祖细胞、造血干细胞和/或诱导多能干细胞),从而改变突变。在一些实施例中,碱基编辑器被向导RNA靶向,以向感兴趣的基因的序列引入一个或多个编辑。

[1259] 该系统可以包括一种或多种不同的载体。在一个方面,所述碱基编辑器是密码子优化的以表达期望的细胞类型,优选真核细胞,优选哺乳动物细胞或人类细胞。

[1260] 一般而言,密码子优化是指一种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,所述过程通过使用在宿主细胞基因中更频繁或最频繁使用的密码子,替换原生序列至少一个密码子(例如约或多于约1、2、3、4、5、10、15、20、25、50或更多个密码子),同时保持原生氨基酸序列。各种物种对特定氨基酸的某些密码子表现出特定的偏向性。密码子偏向性(生物体之间密码子使用的差异)通常与信使RNA(mRNA)的翻译效率相关,而信使RNA(mRNA)的翻译效率又被认为取决于被翻译密码子的特性和特定转移RNA(tRNA)分子的可用性等。细胞中所选tRNA的优势通常反映了肽合成中最常用的密码子。因此,可以基于密码子优化来定制基因以在给定生物体中进行最佳基因表达。密码子使用表很容易获得,例如在www.kazusa.orjp/codon/上可用的“密码子使用数据库”(于2002年7月9日访问)中,这些表可以通过多种方式进行调整。见Nakamura,Y.等人″Codon usage tabulated from the international DNA sequence databases:status for the year 2000″Nucl.Acids Res.28:292(2000)。用于密码子优化特定序列以在特定宿主细胞中表达的计算机算法也是可用的,例如Gene Forge(Aptagen;Jacobus,PA)也是可用的。在一些实施例中,编码工程化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、或所有密码子)对应于对应于特定氨基酸最常用的密码子。

[1261] 包装细胞通常用于形成能够感染宿主细胞的病毒颗粒。这些细胞包括包装腺病毒的293细胞和包装逆转录病毒的psi.2细胞或PA317细胞。用于基因治疗的病毒载体通常是通过产生将核酸载体包装成病毒颗粒的细胞系来产生的。所述载体通常包含包装和随后整合到宿主中所需的最少病毒序列,其他病毒序列被要表达的多核苷酸的表达盒替换。缺失的病毒功能通常由包装细胞系反式供应。例如,用于基因治疗的AAV载体通常仅具有来自

AAV基因组的ITR序列,这些序列是包装和整合到宿主基因组中所需的。病毒DNA可以包装在一个细胞系中,该细胞系包含一个辅助者质粒编码其他AAV基因,即rep和cap,但缺少ITR序列。细胞系也可以用腺病毒作为辅助感染。所述辅助病毒可以促进AAV载体的复制和辅助质粒中AAV基因的表达。在某些情况下,由于缺乏ITR序列,所述辅助质粒没有大量包装。腺病毒的污染可以通过,例如,腺病毒比AAV更敏感的热处理来减少。

[1262] 医药组成物

[1263] 本公开的其他方面涉及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的医药组成物。如本文所用,术语“医药组成物”是指配制用于药物用途的组成物。在一些实施例中,所述医药组成物进一步包含药学上可接受的载体。在一些实施例中,所述医药组成物包含另外的药剂(例如用于特异性递送、增加半衰期或其他治疗化合物)。

[1264] 如本文所用,术语“药学上可接受的载体”是指药学上可接受的材料、组成物或赋形剂,例如液体或固体填充剂、稀释剂、赋形剂、制造助剂(例如润滑剂、滑石镁、硬脂酸钙或锌、或硬脂酸)或溶剂包封材料,参与将化合物从身体的一个部位(例如递送部位)运载或运输到另一部位(例如器官、组织或身体的一部分)。药学上可接受的载体在与制剂的其他成分相容并且对个体的组织没有伤害的意义上是“可接受的”(例如生理相容的、无菌的、生理pH等)。

[1265] 可用作药学上可接受的载体的材料的一些非限制性实例包括:(1)糖类,例如乳糖、葡萄糖和蔗糖;(2)淀粉,如玉米淀粉、马铃薯淀粉;(3)纤维素及其衍生物,如羧甲基纤维素钠、甲基纤维素、乙基纤维素、微晶纤维素、醋酸纤维素等;(4)黄蓍胶粉;(5)麦芽;(6)明胶;(7)润滑剂,如硬脂酸镁、十二烷基硫酸钠、滑石粉等;(8)可可脂、栓剂蜡等赋形剂;(9)油类,如花生油、棉籽油、红花油、香油、橄榄油、玉米油、豆油等;(10)二醇类,如丙二醇;(11)多元醇,如甘油、山梨糖醇、甘露糖醇和聚乙二醇(PEG);(12)油酸乙酯、月桂酸乙酯等酯类;(13)琼脂;(14)缓冲剂,如氢氧化镁、氢氧化铝等;(15)海藻酸;(16)无热原水;(17)等渗盐水;(18)林格氏溶液;(19)乙醇;(20)pH缓冲溶液;(21)聚酯、聚碳酸酯和/或聚酐;(22)填充剂,例如多肽和氨基酸。(23)血清醇,例如乙醇;(23)用于药物制剂的其他无毒兼容物质。润湿剂、着色剂、脱模剂、涂层剂、甜味剂、调味剂、加香剂、防腐剂和抗氧化剂也可存在于制剂中。诸如“赋形剂”、“载体”、“药学上可接受的载体”、“媒液”等术语在本文中可互换使用。

[1266] 医药组成物可包含一种或多种pH缓冲化合物以将制剂的pH维持在反映生理pH的预定水平,例如在约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混合物,例如组氨酸或氨基酸混合物,例如组氨酸和甘氨酸。在一些实施例中,pH缓冲化合物是将制剂的pH维持在预定水平,例如在约5.0至约8.0的范围内并且不螯合钙离子的试剂。这种pH缓冲化合物的说明性实例包括但不限于咪唑和乙酸根离子。所述pH缓冲化合物可以以适合将制剂的pH维持在预定水平的任何量存在。

[1267] 医药组成物还可包含一种或多种渗透调节剂,即,将制剂的渗透特性(例如渗透压、渗透压和/或渗透压)调节至接受者个体的血流和血细胞可接受的水平。所述渗透调节剂可以是不螯合钙离子的试剂。所述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渗透特性的任何化合物。本领域技术人员可以凭经验确定给定渗透调节剂用于本发

明制剂的适用性。合适类型的渗透调节剂的说明性实例包括但不限于:盐,例如氯化钠和乙酸钠;糖类,例如蔗糖、右旋糖和甘露醇;氨基酸,如甘氨酸;以及一种或多种这些药剂和/或药剂类型的混合物。所述渗透调节剂可以以足以调节制剂渗透特性的任何浓度存在。

[1268] 在一些实施例中,所述医药组成物被配制用于递送至个体,例如用于基因编辑。施用本文所述医药组成物的合适途径包括但不限于:局部、皮下、枕下、经皮、皮内、病灶内、关节内、腹膜内、膀胱内、经粘膜、牙龈、牙内、耳蜗内、经鼓膜、器官内、硬膜外、鞘内、肌内、静脉内、血管内、骨内、眼周、肿瘤内、脑内和脑室内给药。

[1269] 在一些实施例中,将本文所述的医药组成物局部施用至患病部位(例如,肿瘤部位)。在一些实施例中,本文所述的医药组成物通过注射、通过导管、通过栓剂或通过植入物给予个体,所述植入物为多孔、无孔或凝胶状材料,包括膜,例如唾液酸膜,或纤维。

[1270] 在其他实施例中,本文所述的医药组成物在控释系统中递送。在一个实施例中,可以使用泵(见例如Langer,1990,Science 249:1527-1533;Sefton,1989,CRC Crit.Ref.Biomed.Eng.14:201;Buchwald等人,1980,Surgery 88:507;Saudek等人,1989,N.Engl.J.Med.321:574)。在另一个实施例中,可以使用聚合材料。(见例如Medical Applications of Controlled Release(Langer和Wise编辑,CRC Press,Boca Raton,Fla.,1974);Controlled Drug Bioavailability,Drug Product Design and Performance(Smolen和Ball编辑,Wiley,New York,1984);Ranger和Peppas,1983,Macromol.Sci.Rev.Macromol.Chem.23:61。另见Levy等人,1985,Science 228:190;During等人,1989,Ann.Neurol.25:351;Howard等人,1989,J.Neurosurg.71:105.)Langer,同上,讨论了其他控释系统。

[1271] 在一些实施例中,根据常规程序将医药组成物配制成适于静脉内或皮下施用至个体例如人的组成物。在一些实施例中,用于通过注射给药的医药组成物是无菌等渗用途的溶液,用作增溶剂和局部麻醉剂例如利多卡因以缓解注射部位的疼痛。通常,成分以单位剂型单独提供或混合在一起提供,例如,作为指示活性剂量的密封容器如安瓿或小袋中的干燥冻干粉或无水浓缩物。当药物通过输液给药时,可以用装有无菌药用级水或盐水的输液瓶进行分配。当医药组成物通过注射给药时,可以提供无菌注射用水或盐水的安瓿,以便在给药前可以混合成分。

[1272] 用于全身给药的医药组成物可以是液体,例如无菌盐水、乳酸林格氏溶液或汉克氏溶液。此外,医药组成物可以是固体形式并在使用前立即重新溶解或悬浮。还考虑了冻干形式。医药组成物可包含在脂质颗粒或囊泡中,例如脂质体或微晶,其也适用于肠胃外给药。颗粒可以具有任何合适的结构,例如单层或多层,只要其中包含组成物。化合物可以被包裹在含有融合脂质二油酰磷脂酰乙醇胺(DOPE)、低水平(5-10mol%)阳离子脂质的“稳定质粒-脂质颗粒”(SPLP)中,并通过聚乙二醇(PEG)涂层稳定(Zhang Y.P.等人,Gene Ther.1999,6:1438-47)。带正电荷的脂质如N-[1-(2,3-二油酰氧基)丙基]-N,N,N-三甲基-甲基硫酸铵或“DOTAP”特别优选用于此类颗粒和囊泡。这种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;各自都通过引用并入本文。

[1273] 例如,本文所述的医药组成物可以作为单位剂量给药或包装。当用于本公开的医药组成物时,术语“单位剂量”是指适合作为个体的单位剂量的物理上离散的单位,每个单

位含有预定量的活性物质,经计算可产生所需的治疗效果,及所需的稀释剂;即,载体或媒液。

[1274] 此外,医药组成物可以作为药物试剂盒提供,其包含(a)含有冻干形式的本发明化合物的容器和(b)第二个容器,其包含药学上可接受的稀释剂(例如无菌用于重构或稀释本发明的冻干化合物。可选地与此类容器相关联的可以是由监管药物或生物产品制造、使用或销售的政府机构以规定形式发出的通知,该通知反映制造、使用或销售机构对人体管理的批准。

[1275] 在另一方面,包括含有可用于治疗上述疾病的材料的制品。在一些实施例中,所述制品包括容器和标签。合适的容器包括例如瓶子、小瓶、注射器和试管。所述容器可由多种材料制成,例如玻璃或塑料。在一些实施例中,所述容器容纳有效治疗本文所述疾病的组成物并且可以具有无菌进入口。例如,所述容器可以是静脉内溶液袋或具有可被皮下注射针刺穿的塞子的小瓶。所述组成物中的活性物质是本发明的化合物。在一些实施例中,所述容器上或与该容器相关的标签表明所述组成物用于治疗选择的疾病。所述制品可进一步包括第二容器,其包含药学上可接受的缓冲液,例如磷酸盐缓冲盐水、林格氏溶液或葡萄糖溶液。它还可以包括从商业和用户的角度来看所需的其他材料,包括其他缓冲液、稀释剂、过滤器、针头、注射器和带有使用说明的包装插页。

[1276] 在一些实施例中,本文描述的任何融合蛋白、gRNA和/或复合物作为医药组成物的一部分提供。在一些实施例中,所述医药组成物包含本文提供的任何融合蛋白。在一些实施例中,所述医药组成物包含本文提供的任何复合物。在一些实施例中,所述医药组成物包含核糖核蛋白复合物,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)。在一些实施例中,所述医药组成物包含gRNA、核酸可编程DNA结合蛋白、阳离子脂质和药学上可接受的赋形剂。医药组成物可任选地包含一种或多种额外的治疗活性物质。

[1277] 在一些实施例中,将本文提供的组成物施用于个体,例如施用于人类个体,以在所述个体内实现靶向基因组修饰。在一些实施例中,细胞获自所述个体并与本文提供的任何医药组成物接触。在一些实施例中,可选地在细胞中实现或检测到所需的基因组修饰之后,从个体体内取出并与医药组成物离体接触的细胞被重新引入所述个体体内。递送包含核酸酶的医药组成物的方法是已知的,并且描述于例如美国专利号6,453,242;6,503,717;6,534,261;6,599,692;6,607,882;6,689,558;6,824,978;6,933,113;6,979,539;7,013,219;和7,163,824,所有这些的公开内容通过引用整体并入本文。尽管本文提供的医药组成物的描述主要涉及适合于施用于人的医药组成物,但本领域技术人员将理解此类组成物通常适合于施用于各种动物或生物体,例如,兽医用。

[1278] 对适于施用于人的医药组成物的修饰以使所述组成物适于施用于各种动物是众所周知的,并且普通的兽医药理学家可以仅通过普通的实验(如果有的话)来设计和/或进行这种修饰。考虑给予所述医药组成物的个体包括但不限于人类和/或其他灵长类动物;哺乳动物、驯养动物、宠物和商业相关的哺乳动物,例如牛、猪、马、羊、猫、狗、小鼠和/或大鼠;和/或鸟类,包括商业相关的鸟类,例如鸡、鸭、鹅和/或火鸡。

[1279] 本文所述的医药组成物的制剂可以通过药理学领域中已知或以后开发的任何方法来制备。通常,此类制备方法包括将活性成分与赋形剂和/或一种或多种其他辅助成分结合的步骤,然后,如果需要和/或需要,将产品成型和/或包装成所需的单剂量或多剂量单

位。药物制剂可另外包含药学上可接受的赋形剂,如本文所用,其包括任何和所有溶剂、分散介质、稀释剂或其他液体载体、分散或悬浮助剂、表面活性剂、等渗剂、增稠剂或乳化剂、防腐剂、固体粘合剂、润滑剂等,适合于所需的特定剂型。Remington's The Science and Practice of Pharmacy,21st Edition,A.R.Gennaro(Lippincott,Williams&Wilkins,Baltimore,MD,2006;在此通过引用整体并入)公开了用于配制医药组成物的各种赋形剂及其制备的已知技术。另见PCT申请PCT/US2010/055131(公开号WO2011/053982A8,提交于2010年11月2日),其通过引用以其整体并入本文,用于生产包含核酸酶的医药组成物的其他合适的方法、试剂、赋形剂和溶剂。

[1280] 除非任何常规赋形剂介质与物质或其衍生物不兼容,例如通过产生任何不希望有的生物效应或以有害方式与所述医药组成物的任何其他成分相互作用,否则其用途被认为在本公开的范围。

[1281] 如上所述的组成物可以有效量给药。所述有效量将取决于给药方式、所治疗的特定病症和期望的结果。它还可能取决于病症的阶段、个体的年龄和身体状况、治疗的性质(如果有的话)以及医师公知的类似因素。对于治疗应用,该量足以达到医学上可取的结果。

[1282] 在一些实施例中,根据本公开的组成物可用于治疗多种疾病、病症和/或病症中的任一种。

[1283] 治疗糖原贮积病1a型(GSD1a)的方法

[1284] 还提供治疗1a型糖原贮积病(GSD1a)和/或引起GSD1a的G6PC基因突变的方法,包括向个体(例如哺乳动物,例如人)施用治疗有效量的医药组成物,其包含编码本文所述的碱基编辑器系统(例如腺苷脱氨酶碱基编辑器8(ABE8)和gRNA)的多核苷酸。在一些实施例中,所述碱基编辑器是包含多核苷酸可编程DNA结合结构域和腺苷脱氨酶结构域的融合蛋白。用所述碱基编辑器和一种或多种靶向碱基编辑器的向导多核苷酸转导所述个体的细胞,以实现包含G6PC基因突变的核酸序列的A·T至G·C改变。

[1285] 本文的方法包括向所述个体(包括被鉴定为需要此类治疗的个体,或怀疑处于疾病风险中并且需要此类治疗的个体)施用有效量的本文所述的组成物。识别需要此类治疗的个体可以由个体或医疗保健专业人员判断并且可以是主观的(例如意见)或客观的(例如可通过测试或诊断方法测量)。

[1286] 所述治疗方法通常包括施用治疗有效量的医药组成物,所述医药组成物包含例如编码碱基编辑器的载体和gRNA,其靶向有需要的个体(例如人类患者)的G6PC基因。此类治疗将适当地施用于患有、具有、易患GSD1a或处于GSD1a风险中的个体,特别是人类个体。本文所述的组成物也可用于治疗任何其他可能涉及GSD1a的病症。

[1287] 在一个实施例中,本发明提供了一种监测治疗进展的方法。该方法包括确定患有或易患与GSD1a相关的病症或其症状的个体的诊断标志物(Marker)(例如与GSD1a相关的SNP)或诊断测量(例如筛选、测定)的水平的步骤,其中已向所述个体施用足以治疗疾病或其症状的治疗量的本文组成物。在该方法中确定的标记水平可以与健康正常对照或其他患病患者中的已知标记水平进行比较,以确定所述个体的疾病状态。在优选的实施例中,在确定第一水平之后的时间点确定个体中标志物的第二水平,并且将这两个水平进行比较以监测疾病的进程或治疗的功效。在某些优选的实施例中,在开始根据本发明的治疗之前确定所述个体的治疗前标志物水平;然后可以将这种治疗前的标志物水平与治疗开始后所述个

体中的标志物水平进行比较,以确定治疗的功效。

[1288] 在一些实施例中,细胞获自个体并与本文提供的医药组成物接触。在一些实施例中,任选地在所需的基因组修饰已经在细胞中受到影响或检测到之后,从个体取出并离体与医药组成物接触的细胞被重新引入所述个体。递送包含核酸酶的医药组成物的方法描述于例如美国专利号6,453,242;6,503,717;6,534,261;6,599,692;6,607,882;6,689,558;6,824,978;6,933,113;6,979,539;7,013,219;和7,163,824,所有这些的公开内容通过引用整体并入本文。尽管本文提供的医药组成物的描述主要涉及适合于施用于人的医药组成物,但本领域技术人员应理解,此类组成物通常适合于施用于各种动物或生物体,例如,兽医用。

[1289] 试剂盒

[1290] 本公开的各个方面提供包含碱基编辑器系统的试剂盒。在一个实施例中,该试剂盒包含核酸结构,该核酸结构包含编码核碱基编辑器融合蛋白的核苷酸序列。所述融合蛋白包含脱氨酶(例如腺嘌呤脱氨酶)和核酸可编程DNA结合蛋白(napDNAbp)。在一些实施例中,所述试剂盒包含至少一种能够靶向感兴趣的核酸分子例如G6PC GSD1a相关突变的向导RNA。在一些实施例中,该试剂盒包含包含编码至少一种向导RNA的核苷酸序列的核酸结构。

[1291] 在一些实施例中,所述试剂盒提供了使用该试剂盒编辑一个或多个G6PC GSD1a相关突变的说明。说明书通常包括关于使用所述试剂盒编辑核酸分子的信息。在其他实施例中,说明包括以下至少一项:注意事项;警告;临床研究;和/或参考文献。说明可以直接印在容器上(如果有),或者作为贴在容器上的标签,或者作为单独的纸张、小册子、卡片或文件夹提供在容器中或随容器一起提供。在进一步的实施例中,试剂盒可以包括标签或单独的插页(包装插页)形式的用于合适的操作参数的说明。在又一个实施例中,所述试剂盒可以包括一个或多个容器,其中装有合适的阳性和阴性对照或对照样品,用作检测、校准或标准化的标准。所述试剂盒还可包括含有药学上可接受的缓冲液的第二容器,例如(无菌)磷酸盐缓冲盐水、林格氏溶液或葡萄糖溶液。它还可以包括从商业和用户的角度来看所需的其他材料,包括其他缓冲液、稀释剂、过滤器、针头、注射器和带有使用说明的包装插页。在某些实施例中,所述试剂盒可用于治疗患有糖原贮积病1a型(GSD1a)的个体。

[1292] 除非另有说明,否则本发明的实施采用分子生物学(包括重组技术)、微生物学、细胞生物学、生物化学和免疫学的常规技术,这些都在技术人员的能力范围内。这些技术在文献中得到了充分的解释,例如"Molecular Cloning:A Laboratory Manual",4th edition(Sambrook,2012);"Oligonucleotide Synthesis"(Gait,1984);"Animal Cell Culture"(Freshney,1987);"Methods in Enzymology""Handbook of Experimental Immunology"(Weir,1996);"Gene Transfer Vectors for Mammalian Cells"(Miller and Calos,1987);"Current Protocols in Molecular Biology"(Ausubel,1987);"PCR:The Polymerase Chain Reaction",(Mullis,1994);"Current Protocols in Immunology"(Coligan,1991)。这些技术适用于生产本发明的多核苷酸和多肽,因此,可以在制造和实践本发明时加以考虑。特定实施例的特别有用的技术将在以下部分中讨论。

[1293] 提出以下实施例是为了向本领域普通技术人员提供如何进行和使用本发明的测定、筛选和治疗方法的完整公开和描述,并非旨在限制发明人视为其发明的范围。

[1294] 实施例

[1295] 提供这些示例仅用于说明目的,而非用于限制本文提供的权利要求的范围。

[1296] 实施例1.基因编辑以校正HEK293T细胞中糖原贮积病1A型(冯吉尔克病)的Q347X突变

[1297] 实施例1.1Q347X突变校正的基础编辑策略。

[1298] GSD1a是由葡萄糖6-磷酸酶(G6PC)基因突变引起的,该基因影响了大约80%的GSD1患者。Q347X突变影响每年大约500名美国患者被诊断为GSD1a。这种突变是一个单碱基替换,它引入了一个终止密码子,在G6PC蛋白的347位(Q347X)提前终止。Q347X的精确校正可能会恢复G6PC的表达并使葡萄糖代谢正常化。

[1299] 腺苷碱基编辑器(ABE)采用Cas9部分和经过验证的原始间隔区相邻基序(PAM)序列偏好,通过在靶位点有效地转换A>G来评估它们校正Q347X突变的能力。代表性的G6PC核苷酸目标序列和相应的氨基酸序列示于图1。标明了用于校正Q347X突变的靶位点和旁观者位点“a”核碱基。该位点的精确校正将产生以下转换:TAG>CAG(终止密码子>谷氨酰胺)。

[1300] 实施例1.2在表达Q347X的HEK293T细胞中的精确校正。

[1301] 使用A·T到G·C DNA碱基编辑器(ABE)靶向Q347X突变以将其恢复为野生型序列,该碱基编辑器采用具有经过验证的原型间隔区相邻基序(PAM)序列偏好的Cas9部分。ABE碱基编辑器可用于靶向智人G6PC核酸序列中的腺苷(A)核碱基以校正Q347X突变。所述SNP处的A>G校正将G6PC多肽中位置347(Q347X)处的终止密码子更改为谷氨酰胺。

[1302] 为了确定哪种ABE-Cas9平台能够最有效和精确地校正目标Q347X突变,通过慢病毒转导将携带Q347X突变的G6PC等位基因基因组整合到HEK293T细胞中。

[1303] 指示在靶位点和旁观者位点“a”核碱基的G6PC目标/插入氨基酸序列示于图3A。G6PC gRNA序列与G6PC目标序列的互补序列杂交,如下所示:

[1304] GACCTAGGCGAGGCAGTAGG GGA

[1305] NGA PAM序列(即SpCas9-VRQR)在上方有下划线。

[1306] gRNA支架序列如下:

[1307] GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU

[1308] 使用ABE8变体分析精确在靶位点和旁观者校正GSD1a Q347X突变的百分比(图2A和2B)。变体2是使用腺苷脱氨酶TadA7.10 VRQR结构的阳性对照ABE。变体3是ABE8.5,它利用了TadA*8.5(TadA*7.10+V82S)(IVT MSP471)的单体构建体。变体4是ABE8.18,利用野生型TadA与TadA*8.18(TadA*7.10+V82S)(IVT MSP465)的异源二聚体构建体。

[1309] 在HEK293T细胞中观察到超过80%的在靶、精确校正的Q347突变(图2A和2B)。V348A的旁观者编辑在碱基编辑器变体中无法检测到。碱基编辑器变体的插入缺失水平低于3.5%。

[1310] 实施例1.3HEK293T细胞GSD1a Q347X突变校正的编辑器优化

[1311] 为了确定用于校正GSD1a中Q347X突变的最佳ABE碱基编辑器,将各种异源二聚体和单体ABE碱基编辑器电穿孔到HEK293T-Q347X细胞中。

[1312] Q347X突变的目靶标/间隔区序列示于图3A。目标/间隔序列显示在靶和旁观者“a”核碱基。可以使用NGA PAM变体(例如GGA)靶向Q347X突变。

[1313] 用ABE8变体靶向Q347X突变测试了向导RNA(gRNA)272(图3-4)。所述gRNA包含如本

文提供的、或基于技术人员的知识确定的、以及本领域技术人员将理解的疾病相关基因的支架序列和间隔序列(目标序列)。(见例如Komor,A.C.等人,“Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage”Nature 533,420-424(2016);Gaudelli,N.M.等人,“Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage”Nature 551,464-471(2017);Komor,A.C.等人,“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),和Rees,H.A.等人,“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)。

[1314] gRNA序列(#272)与G6PC DNA目标序列的互补序列杂交,如下所示:

[1315] GACCTAGGCGAGGCAGTAGG GGA

[1316] NGA PAM序列(即SpCas9)在上方有下划线

[1317] gRNA支架序列如下:

[1318] GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU

[1319] 相对于阳性对照ABE7.10 VRQR异源二聚体构建体(IVT464)、阳性对照单体构建体(IVTmsp468)和阴性对照GFP(图3B),评估使用各种ABE碱基编辑器变体的Q347X校正百分比。图3B是描绘使用ABE8单体和异源二聚体变体构建体校正GSD1a G6PC Q347X突变的百分比的图。所述单体ABE8变体构建体包括:利用TadA*8.1(TadA*7.10+Y147T)(IVT MSP469)单体构建体的ABE8.1、利用TadA*8.2(TadA*7.10+Y147R)(IVT MSP470)单体构建体的ABE8.2、利用TadA*8.3(TadA*7.10+Q154S)(IVT MSP473)单体构建体的ABE8.3、利用TadA*8.5(TadA*7.10+V82S)(IVT MSP471)单体构建体的ABE8.5、利用TadA*8.7(TadA*7.10+Q154R)(IVT MSP472)单体构建体的ABE8.7。所述异源二聚体ABE8变体构建体包括:利用带有TadA*8.14(TadA*7.10+Y147T)(IVT MSP463)的野生型TadA异源二聚体构建体的ABE8.14、利用带有TadA*8.15(TadA*7.10+Y147R)(IVT MSP464)的野生型TadA异源二聚体构建体的ABE8.15、利用带有TadA*8.16(TadA*7.10+Q154S)(IVT MSP467)的野生型TadA异源二聚体构建体的ABE8.16、利用带有TadA*8.18(TadA*7.10+V82S)(IVT MSP465)的野生型TadA异源二聚体构建体的ABE8.18、利用带有TadA*8.20(TadA*7.10+Q154R)(IVT MSP466)的野生型TadA异源二聚体构建体的ABE8.20。

[1320] Q347X突变的百分比校正显示,对于具有V82S突变的异源二聚体变体进行了>85%的编辑,对于具有V82S突变的单体变体进行了大约80%的编辑。对于所有测试的变体,旁观者活性可以忽略不计。

[1321] 实施例1.4用于校正HEK293T细胞中GSD1a Q347X突变的双突变编辑器优化。

[1322] 相对于阳性对照ABE7.10 VRQR异源二聚体构建体(IVT464)、阳性对照单体构建体(IVTmsp468)和阴性对照GFP(图4),评估了使用各种双突变ABE8变体对Q347X的校正百分比。图4是描绘使用双突变ABE8单体变体构建体校正GSD1a G6PC Q347X突变的百分比的图,比较在靶(A6)与旁观者(A2)A>G核碱基。所述单体ABE8变体构建体包括:利用TadA*8.5(TadA*7.10+V82S)(IVT MSP471)单体构建体的ABE8.5、利用TadA*8.28(TadA*7.10+V82S+

Y154S)单体构建体的ABE8.28(IVT MSP501)、利用TadA*8.29(TadA*7.10+V82S+Y147R)单体构建体(IVT MSP499)的ABE8.29、利用TadA*8.30(TadA*7.10+V82S+Y15)单体构建体(IVT MSP500)的ABE8.30、利用TadA*8.31(TadA*7.10+V82S+H123H)单体构建体(IVT MSP503)的ABE8.31、利用TadA*8.32(TadA*7.10+V82S+H12H)单体构建体Y147T)(IVT MSP502)的ABE8.32。

[1323] 所述双突变体ABE8变体的表现与单突变体(V82S单体)ABE8(ABE8.5)相似,碱基编辑效率约为70％至80％。对于所有测试的变体,旁观者活性可以忽略不计。

[1324] 实施例1.5用于校正患者来源B淋巴细胞中GSD1a Q347X突变的编辑器优化。

[1325] 为了确定用于校正GSD1a中Q347X突变的最佳ABE碱基编辑器,将异源二聚体和单体ABE碱基编辑器电穿孔到含有G6PC Q347X突变的患者来源的B淋巴细胞(Coriell Institute)中。

[1326] 用ABE8变体靶向Q347X突变测试了向导RNA(gRNA)272(图3-4)。所述gRNA包含如本文提供的、或基于技术人员的知识确定的、以及本领域技术人员将理解的疾病相关基因的支架序列和间隔序列(目标序列)。(见例如Komor,A.C.等人,“Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage”Nature 533,420-424(2016);Gaudelli,N.M.等人,“Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage”Nature 551,464-471(2017);Komor,A.C.等人,“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),和Rees,H.A.等人,“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).

[1327] gRNA序列(#272)与G6PC DNA目标序列的互补序列杂交,如下所示:

[1328] GACCTAGGCGAGGCAGTAGG GGA

[1329] NGA PAM序列(即SpCas9)在上面有下划线。

[1330] gRNA支架序列如下:

[1331] GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU

[1332] 相对于阳性对照ABE7.10 VRQR构建体(MSP565)和阴性对照GFP(图5),评估了使用各种ABE碱基编辑器变体对Q347X的校正百分比。图5是描绘使用ABE8单体和异源二聚体变体构建体对GSD1a G6PC Q347X突变的A>G校正百分比的图。这些构建体包括:利用带有TadA*8.2(TadA*7.10+Y147R)(MSP559)野生型TadA异源二聚体构建体的ABE8.2,利用带有TadA*8.18(TadA*7.10+V82S)(MSP560)野生型TadA异源二聚体构建体的ABE8.18,和利用TadA*8.5(TadA*7.10+V82S)(MSP561)单体构建体的ABE8.5。

[1333] Q347X突变的百分比校正显示ABE8变体的碱基编辑率约为50％-60％。对于所有测试的变体,旁观者活性可以忽略不计。

[1334] 实施例1.6杂合患者iPSc衍生的Q347X肝细胞中的精确校正。

[1335] 使用杂合的患者iPSc衍生的人类肝细胞(Definigen、Hep GSD1a批次493、507和518)测试了精确的在靶碱基编辑。所述GSD1a iPSc衍生的肝细胞是复合杂合子(Q347X/

G222R)并带有Q347X突变。

[1336] Q347X突变的目标/间隔区序列示于图6A。所述目标/间隔序列显示在靶和旁观者“a”核碱基。可以使用NGA PAM变体(例如GGA)靶向所述Q347X突变。

[1337] gRNA序列与G6PC DNA目标序列的互补序列杂交,如下所示:

[1338] GACCTAGGCGAGGCAGTAGG GGA

[1339] NGA PAM序列(即SpCas9)在上面有下划线。

[1340] gRNA支架序列如下:

[1341] GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU

[1342] 在杂合(Q347X、G222R)患者iPSc衍生的人肝细胞中使用ABE 7.10VRQR评估Q347X的在靶和旁观者A>G碱基编辑校正的百分比(图6B)。在靶碱基编辑对Q347X突变进行了高达15%的在靶、精确校正。没有可检测到的旁观者编辑和非常低水平的插入缺失被观察到。

[1343] 实施例1.7用于校正患者iPSc衍生肝细胞中GSD1a Q347X突变的编辑器优化。

[1344] 在iPSc衍生的人类肝细胞(Definigen、Hep GSD1a批次493、507和518)中使用体外转染方法测试了碱基编辑。所述GSD1a iPSc衍生的肝细胞是复合杂合子(Q347X/G222R)并带有Q347X突变。使用制造特定的铺板和维护程序以使细胞生长并进一步驱动其分化。在细胞接种后十二(12)天进行基于脂质转染的gRNA和mRNA复合物转染。细胞在转染后48小时被裂解并收获gDNA。

[1345] Q347X突变的目标/间隔区序列示于图7A。所述目标/间隔序列显示在靶和旁观者“a”核碱基。可以使用NGA PAM变体(例如GGA)靶向Q347X突变。

[1346] gRNA序列与G6PC DNA目标序列的互补序列杂交,如下所示:

[1347] GACCTAGGCGAGGCAGTAGG GGA

[1348] NGA PAM序列(即SpCas9)在上面有下划线。

[1349] gRNA支架序列如下:

[1350] GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU

[1351] 在患者iPSc衍生的肝细胞中使用ABE8变体,针对在靶A、转换为G、和插入缺失以校正GSD1a Q347X突变的碱基编辑效率显示在图7B中。相对于阳性对照TriLink pBxt464 ABE7.10 VRQR构建体和未处理的阴性对照,评估了各种ABE碱基编辑器变体的编辑效率(图7B)。生成的基础编辑器变体构建体包括:pUtR-TriLink-ABE7.10(Y147R)-VRQR 120A Bbsl;pUtR-TriLink-ABE7.10(V82S)-VRQR 120A Bbsl;pUtR-TriLink-mono-ABE7.10(V82S)-VRQR 120A Bbsl;pUtR-TriLink-VRQR-GeneArt 120A Bbsl;and pUtR-TriLink-Cas9-VRQR-nuclease-bpNLS120A Bbsl。TriLink构建体购自TriLink生物技术。

[1352] 在患者iPSc衍生的肝细胞中观察到G6PC Q347X的类似编辑,在iPSc衍生的肝细胞中优化的碱基编辑器变体的A>G编辑效率约为10%-12%。所述Cas9-VRQR核酸酶在目标区域得出与未处理对照具有相似序列特征的有效插入缺失。所有碱基编辑器都得出了从低至检测不出的插入缺失和旁观者V348A转换。

[1353] 实施例2.GSD1A突变的原代肝细胞共培养系统的体外转导实施例2.1共培养系统和转导方法。

[1354] 在原代肝细胞共培养系统中使用体外转导方法测试碱基编辑。为了产生共培养系统,将原代人肝细胞(PHH)(BioIVT)以每孔350k个细胞接种在I型胶原蛋白包被的24孔板(Corning,354408)中,并在37℃、5%二氧化碳下产生贴壁细胞单层。铺板后4小时,用CP培养基(BioIVT)清洗铺板的肝细胞以去除任何未粘附的细胞。将3T3-J2鼠胚胎成纤维细胞(Kerafast(分发自Howard Green(Harvard),波士顿)以95%肝细胞:5%成纤维细胞/孔的比例接种,并在37℃、5%二氧化碳下再培养12小时以形成共培养物。每2天更换一次培养基(每孔500μL)以进行持续维护。图8B所示的是来自两个人肝细胞供体RSE和TVR的转导的原代肝细胞的图像,所述肝细胞从两个不同的人类肝脏中分离出来。BioIVT(美国马里兰州)将它们分离、冷冻保存并分发出售以用于研究目的。

[1355] 为了进行转导,慢病毒被设计为将G6PC-R83C或G6PC-Q347X引入质粒载体,由VectorBuilder(en.vectorbuilder.com)制造和纯化。在肝细胞共培养形成后的第二天,将慢病毒以MOI 500滴加到每孔培养基中。在将培养基更换为新鲜CP培养基(BioIVT)之前,将共培养物转导16小时。每2天更换一次培养基(每孔500μL)以进行持续维护。在肝共培养物中蛋白质表达在7天期间发生。在转导后第7天,使用基于脂质转染的试剂转染共培养物,该试剂与gRNA和碱基编辑器mRNA共同配制。细胞在转染后48小时被裂解并收获gDNA。图8A提供了显示代表性时间点的肝细胞单层或肝细胞共培养物中体外转导时间表的时间线。

[1356] 实施例2.2使用慢病毒对用于GSD1a Q347X突变的原代肝细胞共培养系统进行体外转导。

[1357] 测试了体外转导。用设计有驱动G6PC Q347X突变表达的CMV启动子和具有半功能效价的3XFLAG-Tag的慢病毒载体成功转导了原代肝细胞共培养物。图9显示的是第6天在原代肝细胞共培养细胞中的GFP表达(GFP,亮视野,合并),和转导慢病毒载体与TriLink ABE7.10 VRQR(NGA PAM),以在30、100和300慢病毒多重感染(MOI)下校正GSD1a Q347X突变。

[1358] 实施例2.3针对GSD1a Q347X突变优化体外慢病毒转导的原代肝细胞共培养系统

[1359] 使用优化条件评估体外慢病毒转导的原代人肝细胞(PHH)共培养系统的编辑效率。来自人类供体RSE的PHH共培养系统在第2天用MOI 500慢病毒转导TBG-G6PC Q347X(图10A)。在第8天,转导的共培养物用TriLink ABE7.10 VRQR、gRNA 272(NGA PAM)转染。

[1360] gRNA序列(#272)与G6PC DNA靶序列的互补序列杂交,如下所示:

[1361] GACCTAGGCGAGGCAGTAGG GGA

[1362] NGA PAM序列(即SpCas9)在上面有下划线。

[1363] 相对于插入缺失评估了ABE碱基编辑器转染的共培养物的在靶编辑效率(图10B)。所述GSD1a Q347X突变在转导的原代肝细胞共培养物中的A>G碱基编辑效率约为11%-15%。如图10B所示,在动物模型中实现治疗益处需要大于5%的碱基编辑效率。

[1364] 在进一步的实验中,来自人类供体RSE的PHH共培养物在第2天在含有1% DMSO(含或不含4% PEG8000)的培养基中转导。转染后,共培养物用含有胶原酶III、胶原酶IV和透明质酸酶的0.5mg/ml溶液处理或保持不处理2分钟,以试图破坏分泌的细胞外基质(ECM),其被认为是转染试剂入口。

[1365] 相对于插入缺失评估了ABE碱基编辑器转染的共培养物的在靶编辑效率(图10C)。在转导的原代肝细胞共培养物中,位置6的GSD1a Q347X突变的A>G碱基编辑效率在治疗组

中约为11％-15％。处理过的PPH共培养物中的Q347X校正产生插入缺失和旁观者V348A转化。

[1366] 实施例2.4从GSD1a转基因小鼠模型分离的原代小鼠肝细胞中G6PC R83C突变的碱基编辑。

[1367] 使用原代小鼠肝细胞共培养系统测试了用于校正R83C的碱基编辑。从含有人G6PC R83C(V166L)突变的转基因小鼠模型中分离原代小鼠肝细胞(图17A)。为了确定用于校正GSD1a中的R83C突变的最佳ABE碱基编辑器,如实施例2.1中所述,将各种saCas9-ABE碱基编辑器构建体转染到原代小鼠肝细胞共培养系统中。

[1368] R83C突变的目标/间隔区核酸序列如下所示。所述目标/间隔区核酸序列显示在靶(粗体、斜体和下划线字体)、同义(斜体和下划线)和旁观者(斜体)“a”核碱基。原型间隔碱基用粗体表示,PAM用粗体和下划线表示,原型间隔之外的碱基用小写字母表示。可以使用NNGRRT PAM变体(例如GAGAAT)靶向R83C突变。

[1369] ccACCAGTATGGACACTGTCCAAAGAGAAT

[1370] 对应的氨基酸序列如下:

[1371] WWYPCQGFLI

[1372] gRNA序列(#820)与G6PC DNA靶序列的互补序列杂交,如下所示:

[1373] CCAGTATGGACaCTGTCCAAA GAGAAT

[1374] NNGRRT PAM序列(即SaCas9)在上面有下划线。

[1375] gRNA支架序列如下:

[1376] GUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCUACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUU

[1377] 测试使用各种saCas9切口酶-ABE碱基编辑器变体构建体的在靶(A12G)、同义(A10G)、旁观者(A6G)和插入缺失的编辑效率,以校正GSD1a R83C突变(图17B)。所述saCas9切口酶-ABE碱基编辑器变体构建体包括:pGL79 pUTR-TriLink-ABE8.1(ABE7.10+Y147T)-saCas9n;pGL80 pUTR-TriLink-ABE8.2(ABE7.10+Y147R)-saCas9n;pGL82 pUTR-TriLink-ABE8.7(ABE7.10+Q154R)-saCas9n;pGL83 pUTR-TriLink-ABE8.3(ABE7.10+Q154S)-saCas9n;pGL92 pUTR-TriLink-ABE7.10-saCas9n;pGL98pUTR-TriLink-monoTadA-ABE8.12(ABE7.10+Y147T+Q154S)-saCas9n。

[1378] 所述saCas9切口酶-ABE碱基编辑器变体构建体pGL79pUTR-TriLink-ABE8.1(ABE7.10+Y147T)-saCas9n和pGL80pUTR-TriLink-ABE8.2(ABE7.10+Y147R)-saCas9n达到最佳的目标编辑效率(图17B)。

[1379] 实施例3.基因编辑以校正HEK293T细胞中1A型糖原贮积病(冯吉尔克病)的R83C突变

[1380] 实施例3.1用于校正R83C突变的碱基编辑策略。

[1381] GSD1a是由葡萄糖6-磷酸酶(G6PC)基因突变引起的,该基因影响了大约80％的GSD1患者。R83C突变影响每年大约900名诊断为GSD1a的美国患者。该突变是在G6PC蛋白的83位(R83C)处引入半胱氨酸的单碱基替换。R83C的精确校正可能会恢复G6PC的表达并使葡萄糖代谢正常化。

[1382] 腺苷碱基编辑器(ABE)采用具有经过验证的原型间隔区相邻基序(PAM)序列偏好

的Cas9部分,评估了它们通过在目标位点有效转换A>G来校正R83C突变的能力。代表性的G6PC核苷酸目标序列和相应的氨基酸序列显示在靶和旁观者位点“a”核碱基上,用于校正R83C突变,如图11所示。该位点的精确校正将产生以下转换:TGT>CGT或TGT>CGC(半胱氨酸>精氨酸)。

[1383] 实施例3.2在表达Q347X的HEK293T细胞中的精确校正。

[1384] 使用A·T到G·C DNA碱基编辑器(ABE)靶向R83C突变以将其恢复为野生型序列,该碱基编辑器采用具有经过验证的原型间隔区相邻基序(PAM)序列偏好的Cas9部分。ABE碱基编辑器可用于靶向智人G6PC核酸序列中的腺苷(A)核碱基以校正R83C突变。所述SNP处的A>G校正将G6PC多肽中位置83(R83C)处的半胱氨酸改变为精氨酸。

[1385] 为了确定哪种ABE-Cas9平台能够最有效和准确地校正目标R83C突变,通过慢病毒转导将带有R83C突变的G6PC等位基因基因组整合到HEK293T细胞中。

[1386] G6PC目标/插入氨基酸序列指示靶位点、同义和旁观者位点“a”核碱基示于图12A。G6PC gRNA序列与G6PC靶序列的互补序列杂交,如下所示:

[1387] CCAGTTGGACACTGTCCAAA GAGAAT

[1388] NNGRRT PAM序列(即SaCas9)在上面有下划线。

[1389] gRNA支架序列如下:

[1390] GUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCUACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUU

[1391] 使用ABE8变体分析GSD1a Q347X突变的在靶和旁观者校正的校正百分比(图12B)。空质粒用作ABE碱基编辑器saABE7.10的阴性对照。变体2是使用单体TadA-SaCas9、pGL78构建体的阳性对照ABE。变体3是一种碱基编辑器,利用带有TadA*8.18(TadA*7.10+V82S)野生型TadA的异源二聚体构建体-SaCas9,pGL81。变体4是一个碱基编辑器,利用带有TadA*8.16(TadA*7.10+Q154S)野生型TadA的异源二聚体构建体-SaCas9,pGL83。

[1392] 在HEK293T细胞中观察到约30%的在靶、精确校正R83C突变(图12B)。用质粒观察到大约3%的A>G碱基编辑。变体1和变体2的旁观者编辑具有可比性。与旁观者校正相比,变体3于在靶校正方面表现出更高的水平。

[1393] 实施例3.3在HEK293T细胞中通过质粒转染对G6PC R83C突变进行碱基编辑。

[1394] 为了确定用于校正GSD1a中Q347X突变的最佳ABE碱基编辑器,将各种ABE碱基编辑器电穿孔到HEK293T-R83C细胞中。

[1395] 与G6PC DNA目标序列的互补序列杂交的向导RNA(gRNA)序列#820和#1121显示在图13A中。所述gRNA序列显示在靶、同义和旁观者“a”核碱基。使用NNGRRT PAM变体(例如gRNA#820GAGAAT PAM)(即SaCas9)或NGA PAM变体(例如gRNA#1121AGA PAM)(即SpCas9)靶向R83C突变。测试了gRNA#820和#1121用于校正GSD1a R83C突变。

[1396] 对于gRNA#1121,支架序列如下:

[1397] GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU

[1398] 对于gRNA#820,支架序列如下:

[1399] GUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCUACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUU

[1400] 所述ABE碱基编辑器构建体可能包括gRNA#820或#1121、碱基编辑器7.9或7.10,以及变体VRQR、VRQR CP5、VRQR CP6或saCas9-ABE(saABE)。所述ABE碱基编辑器构建体如图13B所示,包括:ABE7.9 VRQR gRNA#1121;ABE7.10 VRQR gRNA#1121、ABE7.9 VRQR CP5 gRNA#1121、ABE7.10 VRQR CP5 gRNA#1121、ABE7.9 VRQR CP6gRNA#1121、ABE7.10 VRQR CP6 gRNA#1121、ABE7.9 saABE gRNA#820和ABE7.10 saABE gRNA#820。相对于阴性对照评估使用各种ABE碱基编辑器变体的R83C校正百分比(图13B)。

[1401] 如图13B所示,与旁观者(Y85H)编辑相比,带有gRNA#820的ABE碱基编辑器saABE 7.10实现了对R83C突变的最佳A>G校正。使用NGA PAM的带有gRNA#1121的ABE碱基编辑器产生了不利的旁观者(Y85H)编辑,表明核碱基A10不是ABE VRQR的理想选择。环状置换子5和6的存在增加了核碱基A10处的A>G碱基编辑。

[1402] 实施例3.4用于校正HEK293T细胞中GSD1a R83C突变的编辑器优化。

[1403] 为了确定用于校正GSD1a中R83C突变的最佳ABE碱基编辑器,将各种异源二聚体saCas9-ABE(saABE)碱基编辑器电穿孔到HEK293T-R83C细胞中。所述碱基编辑器由oADE001/002cDNA PCR的mRNA IVT制备,使用2.5μg碱基编辑器和1μg向导RNA。所述HEK293T pLenti G6PC R83C细胞为p7-20,200k细胞/孔,一式三份。

[1404] R83C突变的目标/间隔区核酸序列如下所示。所述目标/间隔区核酸序列显示在靶(粗体、斜体和下划线字体)、同义(斜体和下划线)和旁观者(斜体)“a”核碱基。原型间隔碱基用粗体表示,PAM用粗体和下划线表示,原型间隔之外的碱基用小写字母表示。可以使用NNGRRT PAM变体(例如GAGAAT)靶向R83C突变。

[1405] ccACCAGTATGGACACTGTCCAAAGAGAAT

[1406] 对应的氨基酸序列如下:

[1407] WWYPCQGFLI

[1408] 与G6PC DNA目标序列的互补序列杂交的gRNA序列(#820)如下所示:

[1409] CCAGUAUGGACACUGUCCAAA GAGAAT

[1410] NNGRRT PAM序列(即SaCas9)在上面有下划线。

[1411] 对于上述gRNA序列,支架序列如下:

[1412] GUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCUACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUU

[1413] 使用各种saCas9-ABE(saABE)碱基编辑器异源二聚体变体构建体对在靶(A12G)、同义(A10G)、旁观者(A6G、A0G)和A6G+A10G和A6G+A12G的组合编辑的编辑效率进行了测试以校正GSD1a R83C突变(图14)。相对于单体TadA-saCas9阳性对照和GFP以及未处理的阴性对照,评估了使用各种ABE碱基编辑器变体对R83C的校正百分比(图14)。所述saABE碱基编辑器变体异源二聚体结构包括:TadA*8.14((WT)+(TadA7.10+Y147T))-saCas9;TadA*8.15((WT)+(TadA7.10+Y147R))-saCas9、TadA*8.18((WT)+(TadA7.10+V82S))-saCas9、TadA*8.20((WT)+(TadA7.10+Q154R))-saCas9和TadA*8.16((WT)+(TadA7.10+Q154S))-saCas9。

[1414] R83C突变的直接校正涉及原始间隔区内位置12G处的A>G转换。位于6G位的旁观者(Y85H(TAC>CAC))可能会降低G6PC的活性,因为突变位于酶的活性位点,然而,预计它的活性不低于R83C。10G时的转化是同义,即TGT>CGC(半胱氨酸>精氨酸)。

[1415] 实施例3.5用于校正HEK293T细胞中GSD1a R83C突变的双突变编辑器优化。

[1416] 评估了使用各种双突变单体和异源二聚体ABE8 saCas9切口酶(saCas9n)ABE(saABE)碱基编辑器变体对R83C的校正百分比。所述双突变碱基编辑器是用oADE001/002cDNA PCR的mRNA IVT制备的,使用2.5μg碱基编辑器和1μg向导RNA。所述HEK293T pLenti G6PC R83C细胞为p7-20,200k细胞/孔,一式三份。

[1417] R83C突变的目标/间隔区核酸序列如下所示。所述目标/间隔区核酸序列显示在靶(粗体、斜体和下划线字体)、同义(斜体和下划线)和旁观者(斜体)“a”核碱基。原型间隔碱基用粗体表示,PAM用粗体和下划线表示,原型间隔之外的碱基用小写字母表示。可以使用NNGRRT PAM变体(例如GAGAAT)靶向R83C突变。

[1418] ccACCAGTATGGACACTGTCCAAAGAGAAT

[1419] 对应的氨基酸序列如下:

[1420] WWYPCQGFLI

[1421] 与G6PC DNA靶序列的互补序列杂交的gRNA序列(#820)如下所示:

[1422] CCAGUAUGGACaCUGUCCAAA GAGAAT

[1423] NNGRRT PAM序列(即SaCas9)在上面有下划线。

[1424] 对于上述gRNA序列,支架序列如下:

[1425] GUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCUACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUU

[1426] 使用各种saCas9n-ABE(saABE)碱基编辑器双突变单体和异源二聚体变体构建体,对GSD1a R83C突变的在靶(A12G)、同义(A10G)、旁观者(A6G、A0G)和组合(A6G+A10G和A6G+A12G)校正的编辑效率进行测试(图15)。相对于单体TadA-saCas9n阳性对照和异源二聚体TadA-saCas9n阳性对照,评估使用各种双突变ABE碱基编辑器变体对R83C的校正百分比(图15)。所述saABE碱基编辑器变体单体构建体包括:TadA*8.1(TadA7.10+Y147T)-saCas9n;TadA*8.2(TadA7.10+Y147R)-saCas9n;TadA*8.3(TadA7.10+Q154S)-saCas9n;TadA*8.12(TadA7.10+Y147T+Q154S)-saCas9n;和TadA*8.27(TadA7.10+Y147R+Q154S)-saCas9n。

[1427] 所述saABE碱基编辑器变体异源二聚体结构包括:TadA*8.14((WT)+(TadA7.10+Y147T))-saCas9n;TadA*8.15((WT)+(TadA7.10+Y147R))-saCas9n;TadA*8.20((WT)+(TadA7.10+Q154R))-saCas9n;TadA*8.25((WT)+(TadA7.10+Y147T/Q154S))-saCas9n;和TadA*8.33((WT)+(TadA7.10+Y147R+Q154S))-saCas9n。

[1428] R83C突变的直接校正涉及原始间隔区内位置12G处的A>G转换。位于6G位的旁观者(Y85H(TAC>CAC))可能会降低G6PC的活性,因为突变位于酶的活性位点,然而,预计它的活性不低于R83C。10G处的转化是同义,即TGT>CGC(半胱氨酸>精氨酸)。

[1429] 通过在用于靶向R83C的TadA-SaCas9编辑器中包含两个突变而不是单个突变,进一步编辑器优化实验的目的,是为了实现更高水平的在靶编辑或进一步解开12G和6G的编辑。测试的双突变体中没有一个产生改进的在靶编辑,但它们确实保持了与测试的单突变体相似的编辑水平。如图15所示,TadA-SaCas9编辑器在靶位点产生30-40%的A>G转换,以及大约20%的旁观者编辑。测序表明,这些编辑主要以未链接的方式发生,在没有旁观者编辑的情况下,在靶编辑的频率很高。例如,二聚体TadA-SaCas9 pGL83产生如下表14所示的编辑。

[1430] 表14.TadA-SaCas9 A>G碱基编辑频率。

[1431]

二聚体 TadA-SaCas9 pGL83 测序	碱基编辑 (%)
CCAGT**A**TGG**A**C**A**CTGTCCAAA	35.79
CCAGT**A**TGG**A**C**G**CTGTCCAAA	33.27
CCAGT***G***TGG**A**C**A**CTGTCCAAA	19.47
CCAGT**A**TGG***G***C**A**CTGTCCAAA	4.87
CCAGT***G***TGG**A**C**G**CTGTCCAAA	3.43
CCAGT**A**TGG***G***C**G**CTGTCCAAA	1.10

[1432] 以粗体和下划线显示的是R83C在靶校正的A>G核碱基。以粗体和斜体显示的是Y85C的A>G核碱基旁观者校正。

[1433] 实施例3.6HEK293T细胞中GSD1a R83C校正优化编辑器的重现性研究。

[1434] 用优化的编辑器对GSD1a R83C校正的再现性进行了评估,如图16所示。

[1435] R83C突变的目标/间隔区核酸序列如下所示。所述目标/间隔区核酸序列显示在靶(粗体、斜体和下划线字体)、同义(斜体和下划线)和旁观者(斜体)“a”核碱基。原型间隔碱基用粗体表示,PAM用粗体和下划线表示,原型间隔之外的碱基用小写字母表示。可以使用NNGRRT PAM变体(例如GAGAAT)靶向R83C突变。

[1436] cc***A***CCAGT***A***TGG***A***C***A***CTGTCCAAAGAGAAT

[1437] 对应的氨基酸序列如下:

[1438] WWYPCQGFLI

[1439] 与G6PC DNA目标序列的互补序列杂交的gRNA序列(#820)如下所示:

[1440] CCAGUAUGGACaCUGUCCAAA GAGAAT

[1441] NNGRRT PAM序列(即SaCas9)在上面有下划线。

[1442] 对于上述gRNA序列,支架序列如下:

[1443] GUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCUACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUU

[1444] 测试了使用各种优化的ABE碱基编辑器对GSD1a R83C突变的在靶(A12G)、同义(A10G)和旁观者(A6G、A0G)校正的编辑效率(图16)。相对于pGL78单体ABE7.10阳性对照评估使用ABE碱基编辑器变体的R83C校正百分比(图15)。ABE碱基编辑器变体构建体包括:pGL97单体Q154S(TadA*8.3);pGL95单体Y147T(TadA*8.1);pGL98单体Y147T+Q154S(TadA*8.12);pGL83二聚体Q154S(TadA*8.16);pGL79二聚体Y147T(TadA*8.14);和pGL93二聚体Y147T+Q154S(TadA*8.25)。

[1445] 如图16所示,用于校正GSD1a R83C突变,与旁观者相比,优化的异源二聚体ABE碱基编辑器变体的在靶(12A)A>G碱基编辑表现出更好的效率。

[1446] 实施例3.7用于HEK293细胞中的GSD1a R83C校正的向导RNA截短研究。

[1447] 对GSD1a R83C校正进行了向导RNA截短研究,以了解改变向导长度对编辑的影响(例如改变在靶和脱靶活性)。在一项实验中,使用包含与SaCas9(pGL78)融合的单体TadA的ABE变体测试了靶向R83C位点的不同长度的向导:

[1448] CCAGUAUGGACaCUGUCCAAA(21nt)

[1449] CAGUAUGGACaCUGUCCAAA(20nt)

[1450] AGUAUGGACaCUGUCCAAA(19nt)

[1451] 简而言之,mRNA编码碱基编辑器在体外转录并通过片段分析仪确定为~70%纯度。带有G6PC R83C目标序列的慢病毒插入的HEK293T细胞(200K细胞/孔)以2.5μg编辑器mRNA、1μg向导RNA转染三次。令人惊讶的是,改变向导的靶向区域长度,改变了在12A下的在靶编辑(希望的)以及在6A下的脱靶旁观者编辑(不希望的)(图18)。特别是,19nt和20nt向导显示使用TadA-SaCas9碱基编辑器的在靶编辑增加,脱靶编辑减少。不受理论束缚,具有高在靶活性和高在靶编辑与脱靶编辑比率的向导RNA对于碱基编辑是理想的。

[1452] 在另一个实验中,使用20nt和21nt向导测试了各种ABE8(图19)。相较于与21nt向导一起使用时,当与20nt向导一起使用时,单体(TadA*8)和二聚体(TadA(wt)-TadA*8)变体显示出更高的在靶编辑(12A;希望的)与脱靶编辑(6A;不希望的)的比率。一般来说,与使用21nt向导相比,使用20nt向导的ABE编辑器也显示出更高的在靶编辑(12A;希望的)。与先前的碱基编辑器相比,单-R20A/K21A、单-V82G、双-R20A/K21A、双-V82G、双/单双突变变体与20nt向导一起,显示增加的在靶活性和减少的脱靶活性。另一项实验表明,与SaCas9融合的包含Y147T和Q1545突变的单体TadA7.10,或包含Y147T、Q1545和V82G突变的异源二聚体TadA(wt)-TadA7.10,与20nt向导RNA结合,在GSD1a R83C位点处提供了高水平的在靶编辑,伴随着低水平脱靶、旁观者编辑(图20)。

[1453] 还在本文所述的慢病毒转导的原代肝细胞共培养系统中测试了碱基编辑器-向导组合。转导的共培养物用diTadA-ABE7.10(Y147T/Q154S)-SaCas9或单TadA ABE7.10(Y147T/Q154S)-SaCas9(分别为MSP602或MSP603)和gRNA 820(20nt或21nt)转染。所有mRNA/gRNA组合在转导的肝细胞中产生特定的在靶(12G)编辑。用gRNA820-20nt转染提高了在靶编辑水平(图21)。在另一个实验中,测试一些先前测试的碱基编辑器-向导组合(即变体3-5条件),在原代人肝细胞模型中观察到显着水平的精确R83C校正(图22)。

[1454] 实施例4.杂合转基因GSD1A R83C小鼠体内的精确校正

[1455] 在GSD1a的人源化转基因小鼠模型中进行了用于精确校正G6PC R83C的概念研究的分子证明。在两项独立的研究中,向动物给药与编辑器mRNA和gRNA共同配制的LNP。研究1是用10-15周龄的动物进行的,并在给药后7天使其死亡(TD)。来自研究2的动物为8-11周龄,并在LNP给药后2周使其死亡。使其死亡后,分离肝脏、匀浆,并在靶位点及其周围进行PCR和下一代测序分析。

[1456] 转基因小鼠模型是在Applied StemCell(Miliptas,CA)通过使用单链DNA(ssDNA)供体(~1.2kb)敲除小鼠G6PC基因和敲入人G6PC cDNA生成的,设计用于在人类cDNA转录本(G6PC-201,ENST00000253801.6)中包含点突变R83C(CGT>TGT)。使用的编辑器是异源二聚体SaABE8.12(V82G),也列为异源二聚体SaABE8(Y147T、Q154S、V82G)。所述gRNA具有以下序列:CAGTATGGACACTGTCCAAA。

[1457] R83C突变的目标/间隔区核酸序列如下所示。

[1458] CCAGTATGG*AC***A**CTGTCCAAA

[1459] 所述目标/间隔区核酸序列显示在靶(粗体)、同义(斜体)和旁观者(下划线字体)“A”核碱基。可以使用GAGAAT PAM序列靶向R83C突变。

[1460] 如图23所示,携带具有R83C突变的人G6PC的转基因小鼠对R83C产生了约15-25%的在靶精确校正,而导致Y85H突变的旁观者编辑很少(约2-7%)。

[1461] 实施例5.材料和方法

[1462] 本文所述实施例中提供的结果是使用以下材料和方法获得的。

[1463] 本文描述了目标多核苷酸和gRNA的DNA序列以及使用的引物。对于gRNA,提供了以下支架序列:GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUUUU。该支架可用于NGG、NGA、NGC、NGT PAM(即SpCas9)。以下支架序列:GUUUUAGUAC UCUGUAAUGA AAAUUACAGA AUCUACUAAA ACAAGGCAAA AUGCCGUGUU UAUCUCGUCA ACUUGUUGGC GAGAUUUU可用于NNGRRT PAM(即SaCas9)。所述gRNA包含如本文提供的、或基于技术人员的知识确定的、以及本领域技术人员将理解的疾病相关基因的支架序列和间隔序列(目标序列)(例如表3A和3B)(见例如Komor,A.C.等人,"Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage"Nature 533,420-424(2016);Gaudelli,N.M.等人,"Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage"Nature 551,464-471(2017);Komor,A.C.等人,"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),和Rees,H.A.等人,"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)。

[1464] 使用VeraSeq ULtra DNA聚合酶(Enzymatics)或Q5热启动高保真DNA聚合酶(New England Biolabs)进行PCR。Base Editor(BE)质粒是使用USER克隆(New England Biolabs)构建的。脱氨酶基因合成为gBlocks基因片段(Integrated DNA Technologies)。使用的Cas9基因如下所列。Cas9基因是从先前报道的质粒中获得的。脱氨酶和融合基因被克隆到pCMV(哺乳动物密码子优化)或pET28b(大肠杆菌密码子优化)骨架中。sgRNA表达质粒是使用定点诱变构建的。

[1465] 简而言之,根据制造商的说明,使用T4多核苷酸激酶(New England Biolabs)将上文列出的引物进行5'磷酸化。接下来,根据制造商的说明,使用Q5热启动高保真聚合酶(New England Biolabs)和磷酸化引物和编码感兴趣基因的质粒作为模板进行PCR。PCR产物与DpnI(20U,New England Biolabs)在37℃下孵育1小时,在QIAprep离心柱(Qiagen)上纯化,并根据制造商的说明使用QuickLigase(New England Biolabs)进行连接。使用Mach1感受态细胞(ThermoFisher Scientific)进行DNA载体扩增。

[1466] 细胞培养

[1467] HEK293T(ATCC CRL-3216)和U2OS(ATCC HTB-96)在补充有10%(v/v)胎牛血清(FBS)的Dulbecco's改良Eagle's培养基(Dulbecco's Modified Eagle's Medium)加GlutaMax(ThermoFisher)中,在37℃和5%二氧化碳条件下维持。HCC1954细胞(ATCC CRL-2338)维持在RPMI-1640培养基(ThermoFisher Scientific)中,如上所述补充。含有目标基因(例如SERPINA1、G6PC、IDUA等)(aconic Biosciences)的永生化细胞在Dulbecco改良Eagle培养基加GlutaMax(ThermoFisher Scientific)中培养,并辅以10%(v/v)胎牛血清(FBS)和200μg ml^{-1}遗传霉素(ThermoFisher Scientific)。

[1468] HEK293T(293T)细胞系获自美国组织培养物保藏中心(ATCC)。293T细胞在补充有10％胎牛血清和1％青霉素/链霉素的DMEM中,在37℃和5％二氧化碳条件下维持。按照制造商的说明,用Lipofectamine 2000(Invitrogen)在24孔板中转染所有细胞系。用于脂质转染的DNA量为每孔1μg。293T细胞的转染效率通常高于80％,这是在递送对照GFP表达质粒后通过荧光显微镜确定的。

[1469] 转染。

[1470] HEK293T细胞接种在48孔胶原包被的BioCoat板(Corning)上,并以大约85％的汇合度进行转染。简而言之,根据制造商的方案,每孔使用1.5μl Lipofectamine 2000(ThermoFisher Scientific)转染750ng BE和250ng sgRNA表达质粒。根据制造商的说明(V试剂盒使用程序Q-001用于HEK293T细胞),使用适当的Amaxa Nucleofector II程序转染HEK293T细胞。

[1471] 对于质粒转染,使用Opti-MEM培养基和Lipofectamine 2000将HEK293T细胞接种并用250ng含有U6启动子并编码gRNA的表达质粒和750ng编码Cas9/ABE8变异碱基编辑器的表达质粒转染。ABE8使用的变体包括NGG PAM序列。将细胞在37℃和5％二氧化碳中维持5天,在转染后第3天更换药物。此后,细胞被裂解;分离基因组DNA并使用标准程序进行PCR,通常使用20-100ng模板DNA。添加接头(Illumina)后,对DNA进行深度测序。所需位点的碱基编辑通过MiSeq分析进行分析。

[1472] 基因组DNA样本的高通量DNA测序。

[1473] 3天后收获转染细胞,并根据制造商的说明使用Agencourt DNAdvance基因组DNA分离试剂盒(Beckman Coulter)分离基因组DNA。感兴趣的在靶和脱靶基因组区域通过PCR与侧翼高通量测序引物对进行扩增。根据制造商的说明,使用5ng基因组DNA作为模板,使用Phusion高保真DNA聚合酶(ThermoFisher)进行PCR扩增。对每个引物对分别确定循环数以确保反应在扩增的线性范围内停止。PCR产物使用RapidTips(Diffinity Genomics)纯化。使用含有测序接头的引物通过PCR扩增纯化的DNA。使用Quant-iT PicoGreen dsDNA检测试剂盒(ThermoFisher)和KAPA Library定量试剂盒-Illumina(KAPA Biosystems)对产物进行凝胶纯化和定量。如前所述(Pattanayak,Nature Biotechnol.31,839-843(2013))在Illumina MiSeq上对样品进行测序。

[1474] 对来自基因组DNA或从293T细胞重复转染中收获的RNA的PCR扩增子进行深度测序。通过凝胶电泳验证PCR产物的质量后,通过凝胶提取分离PCR产物,例如,使用Zymoclean Gel DNA Recovery试剂盒(Zymo Research)。Shotgun库是在没有剪切的情况下准备的。该库通过qPCR进行量化,并在一个MiSeq Nano流动槽上使用MiSeq500循环测序试剂盒版本2,从片段的每一端测序251个循环。使用bcl2fastq v2.17.1.14转换软件生成和解复用Fastq文件(Illumina)。

[1475] 其他实施例

[1476] 从前面的描述中,很明显可以对本文描述的本发明进行变化和修改以将其用于各种用途和条件。这样的实施例也在所附权利要求的范围内。

[1477] 在本文中对变量的任何定义中的要素列表的叙述,包括将该变量定义为任何单一要素或所列要素的组合(或子组合)。此处对实施例的叙述包括作为任何单个实施例或与任何其他实施例或其部分组合的实施例。

[1478] 本说明书中提及的所有出版物、专利和专利申请均以引用方式并入本文,其程度如同个别的出版物、专利或专利申请被具体地和分别地指示为以引用方式并入一样。除非另有说明,本说明书中提及的出版物、专利和专利申请通过引用整体并入本文。

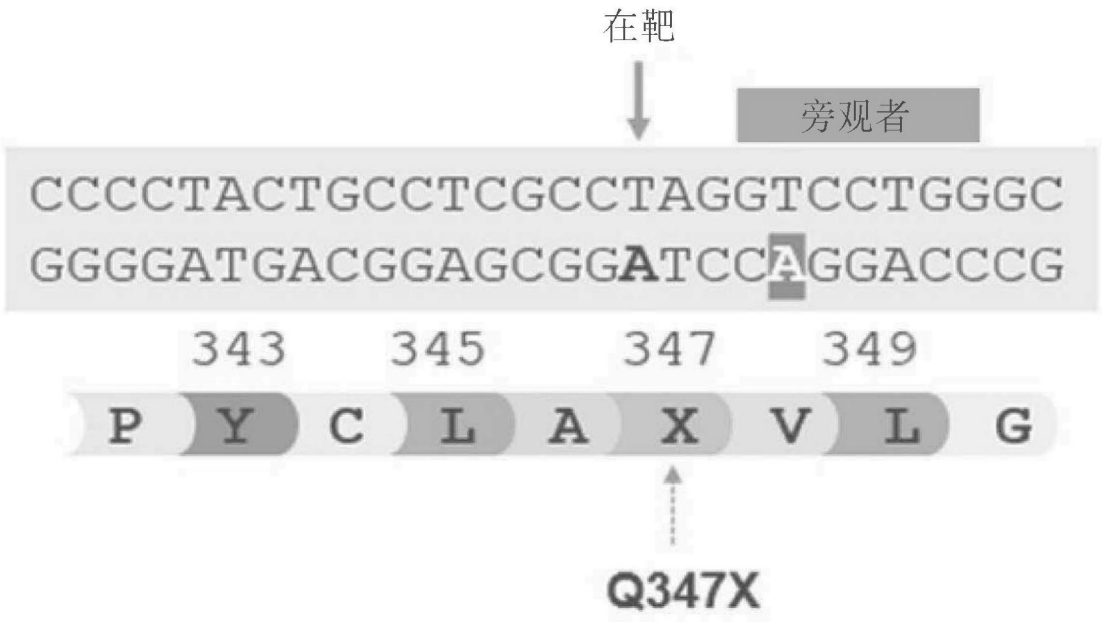

在靶
旁观者
CCCCTACTGCCTCGCCTAGGTCCTGGGC
GGGGATGACGGAGCGGATCCAGGACCCG
343
345
347
349
P Y C L A X V L G
Q347X

图1

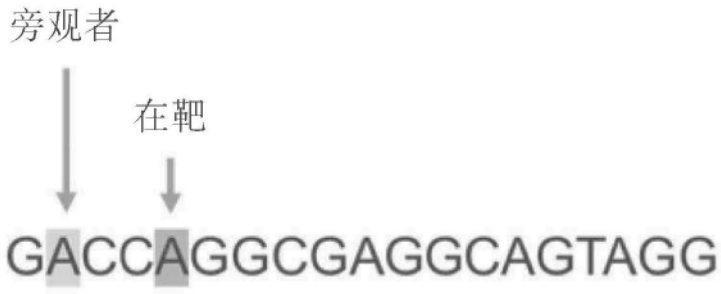

旁观者
在靶
GACCAGGCGAGGCAGTAGG

图2A

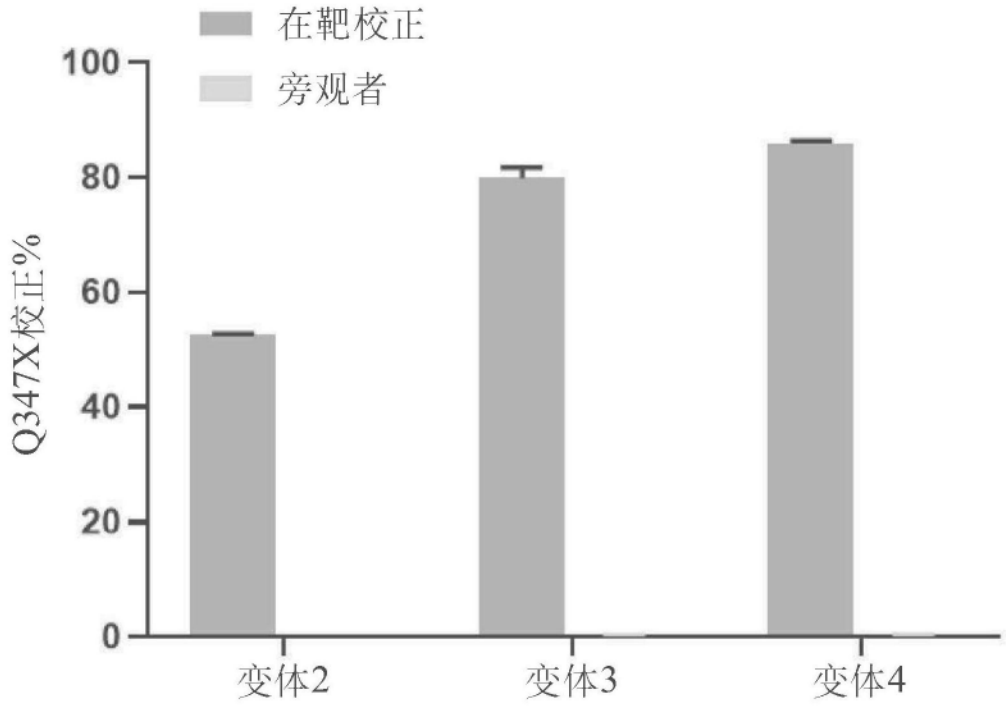

在靶校正
旁观者
100
80
60
40
20
0
Q347X校正%
变体2
变体3
变体4

图2B


旁观者
在靶
GACCTAGGCGAGGCAGTAGG GGA
V Q A L C Y PAM

图3A

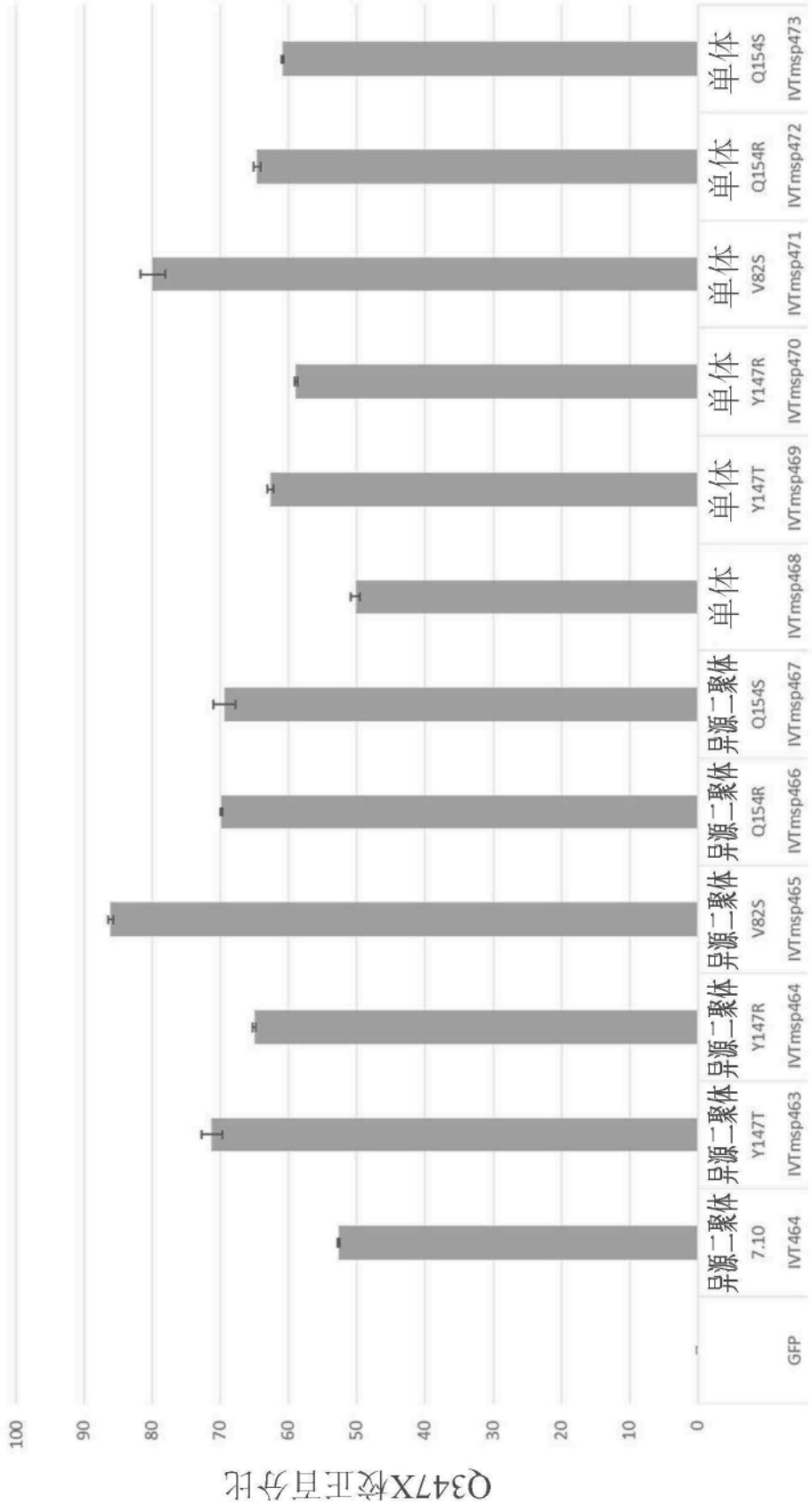

Q347X校正百分比
100
90
80
70
60
50
40
30
20
10
0
GFP
异源二聚体 7.10 IVT464
异源二聚体 Y147T IVTmsp463
异源二聚体 Y147R IVTmsp464
异源二聚体 V82S IVTmsp465
异源二聚体 Q154R IVTmsp466
异源二聚体 Q154S IVTmsp467
单体 IVTmsp468
单体 Y147T IVTmsp469
单体 Y147R IVTmsp470
单体 V82S IVTmsp471
单体 Q154R IVTmsp472
单体 Q154S IVTmsp473

图3B

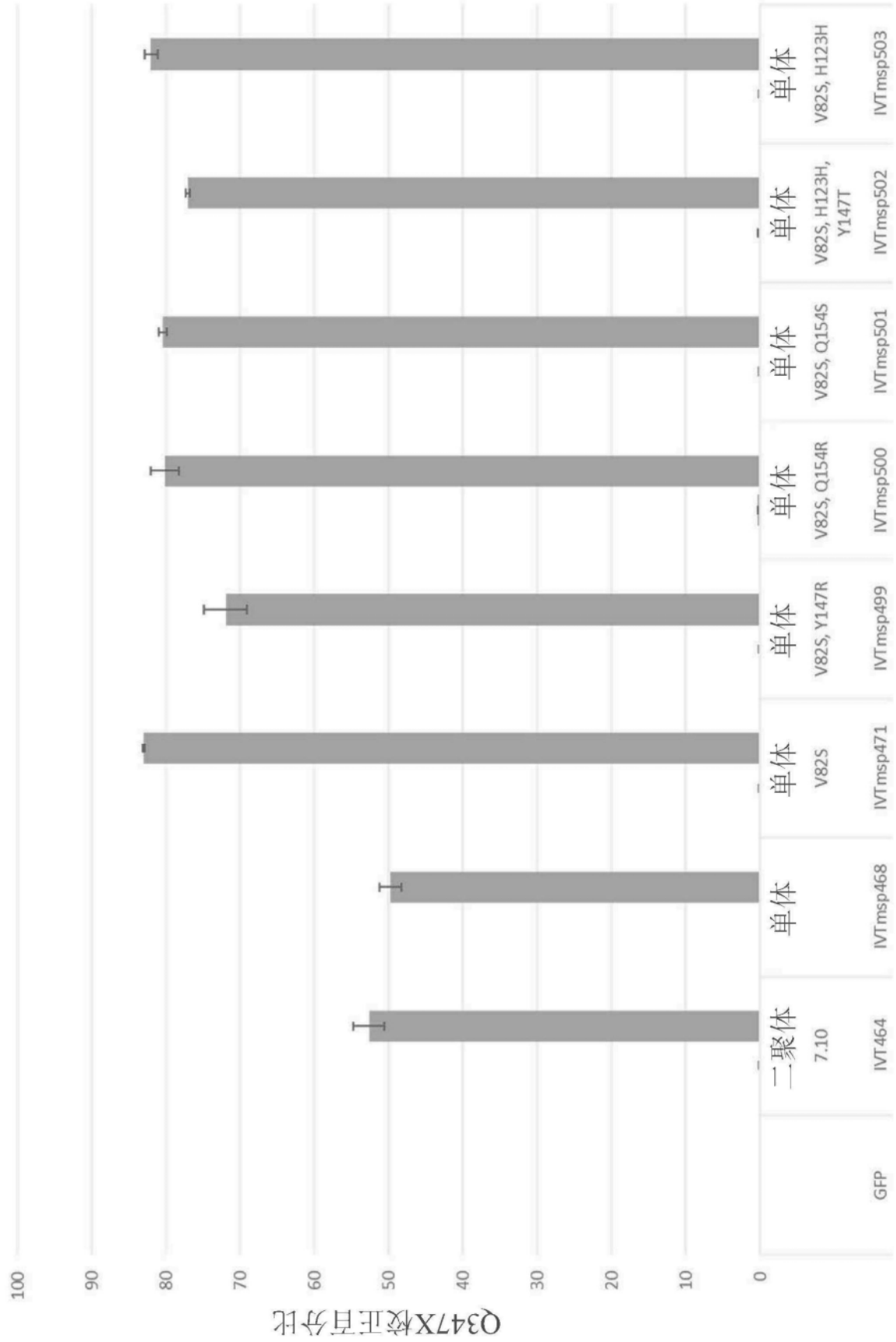

Q347X校正百分比
100
90
80
70
60
50
40
30
20
10
0
GFP
二聚体
7.10
IVT464
单体
IVTmsp468
单体
V82S
IVTmsp471
单体
V82S, Y147R
IVTmsp499
单体
V82S, Q154R
IVTmsp500
单体
V82S, Q154S
IVTmsp501
单体
V82S, H123H, Y147T
IVTmsp502
单体
V82S, H123H
IVTmsp503

图4

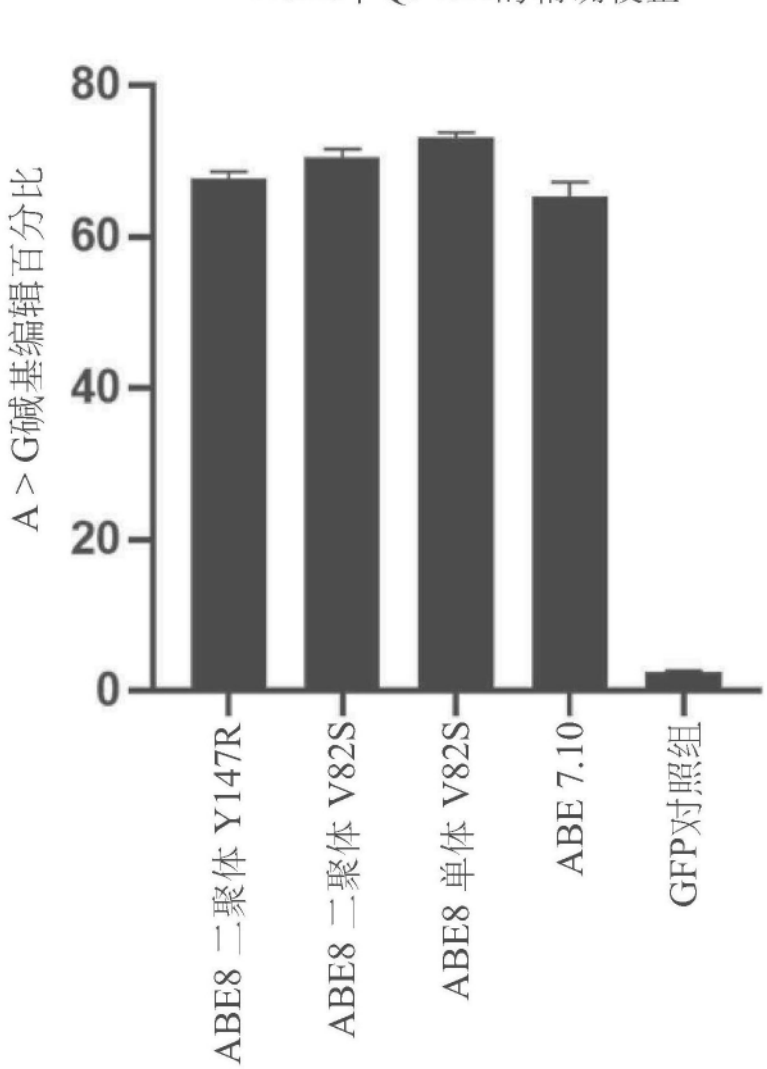

LCLs中Q347X的精确校正
A > G碱基编辑百分比
80
60
40
20
0
ABE8 二聚体 Y147R
ABE8 二聚体 V82S
ABE8 单体 V82S
ABE 7.10
GFP对照组

图5


旁观者
在靶
GACCTAGGCGAGGCAGTAGG GGA
V Q A L C Y PAM

图6A

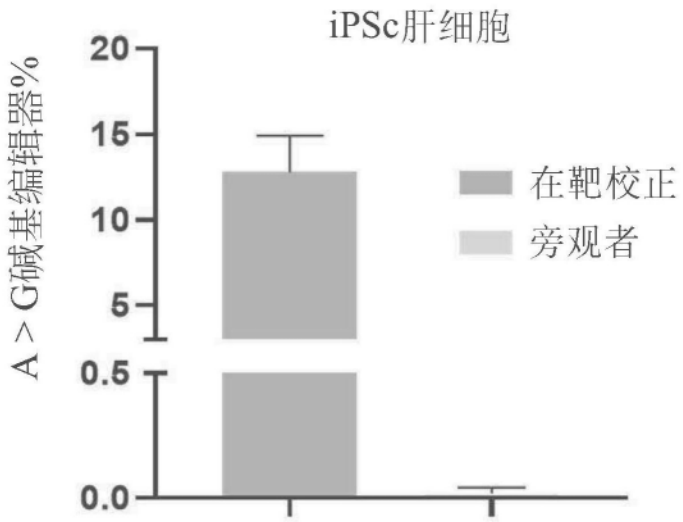

iPSc肝细胞
A > G碱基编辑器%
20
15
10
5
0.5
0.0
在靶校正
旁观者

图6B


旁观者
在靶
GA CCT AGG CGA GGC AGT AGG GGA
NGA

图7A

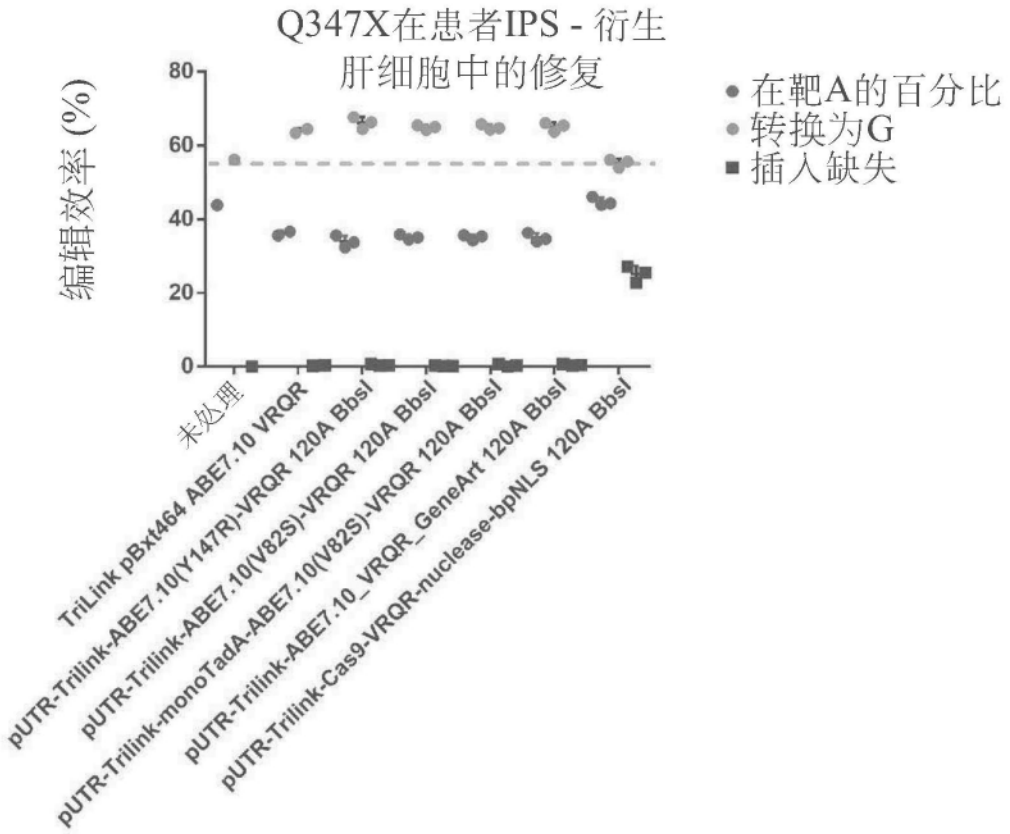

Q347X在患者IPS - 衍生
肝细胞中的修复
编辑效率 (%)
80
60
40
20
0
在靶A的百分比
转换为G
插入缺失
未处理
TriLink pBxt464 ABE7.10 VRQR
pUTR-Trilink-ABE7.10(Y147R)-VRQR 120A BbsI
pUTR-Trilink-ABE7.10(V82S)-VRQR 120A BbsI
pUTR-Trilink-monoTadA-ABE7.10(V82S)-VRQR 120A BbsI
pUTR-Trilink-ABE7.10_VRQR_GeneArt 120A BbsI
pUTR-Trilink-Cas9-VRQR-nuclease-bpNLS 120A BbsI

图7B

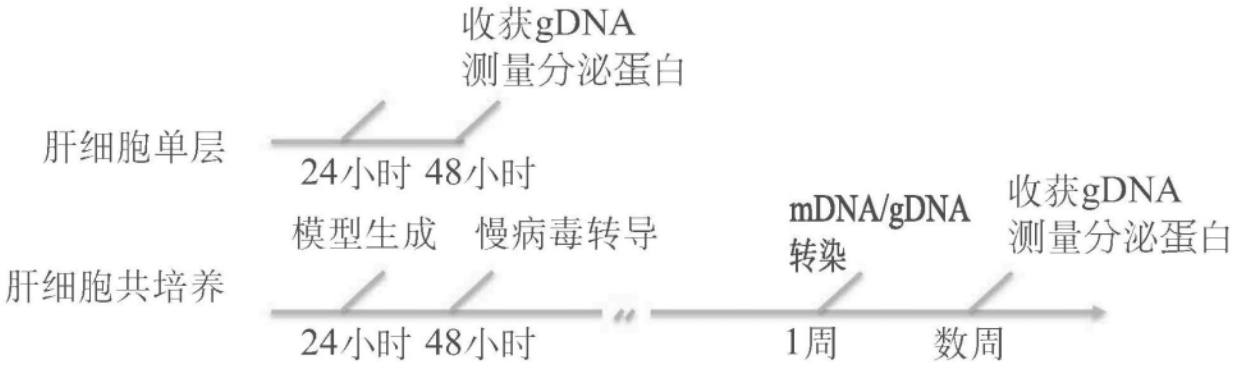

收获gDNA
测量分泌蛋白
肝细胞单层
24小时 48小时
模型生成 慢病毒转导
mDNA/gDNA
转染
收获gDNA
测量分泌蛋白
肝细胞共培养
24小时 48小时
1周
数周

图8A

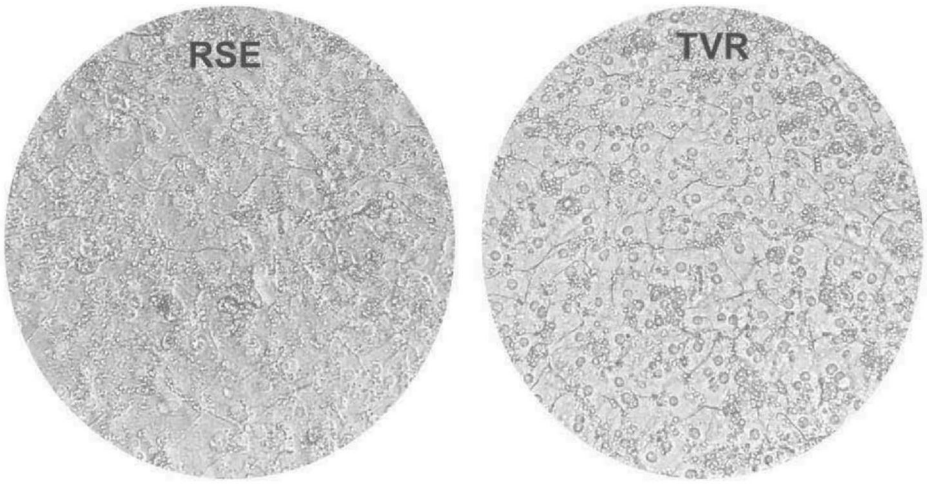


图8B

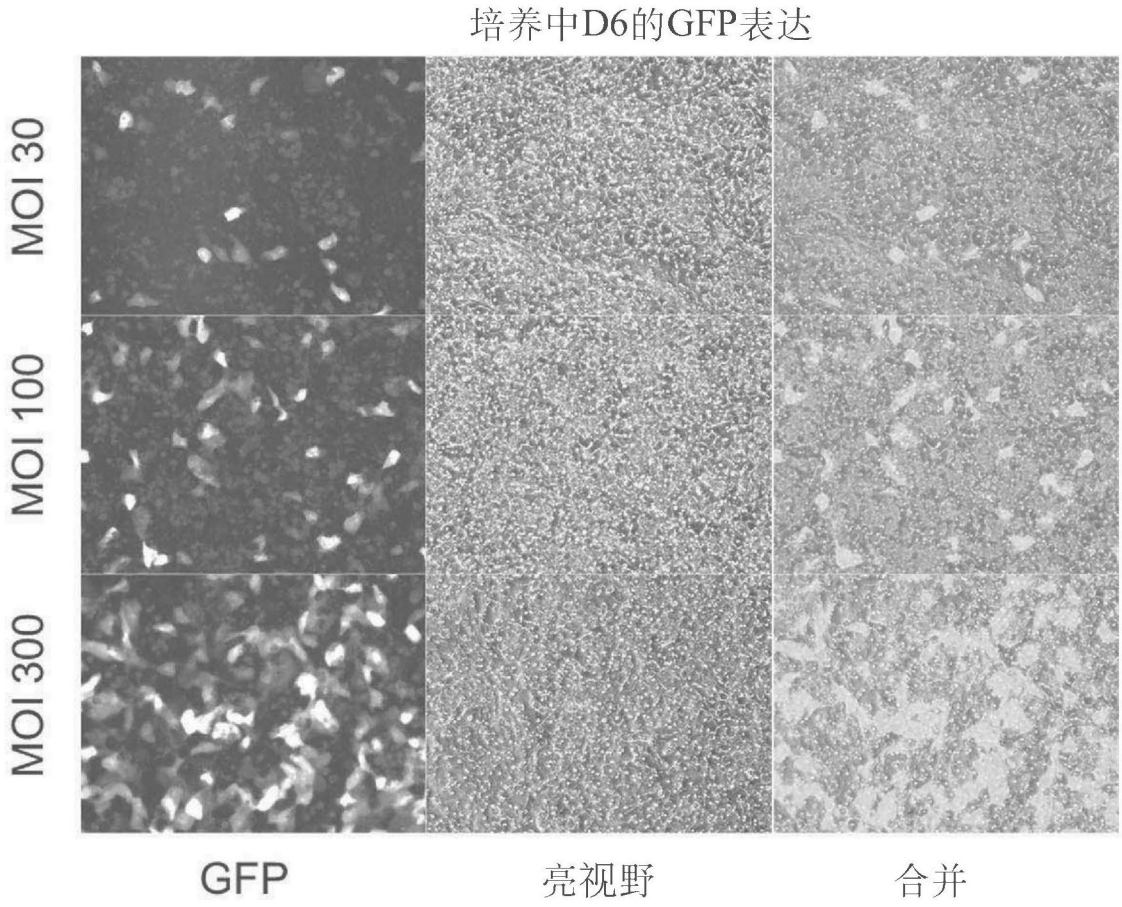


图9

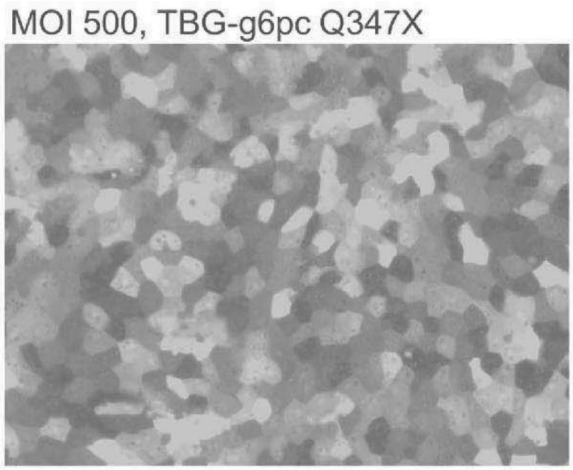

MOI 500, TBG-g6pc Q347X

图10A

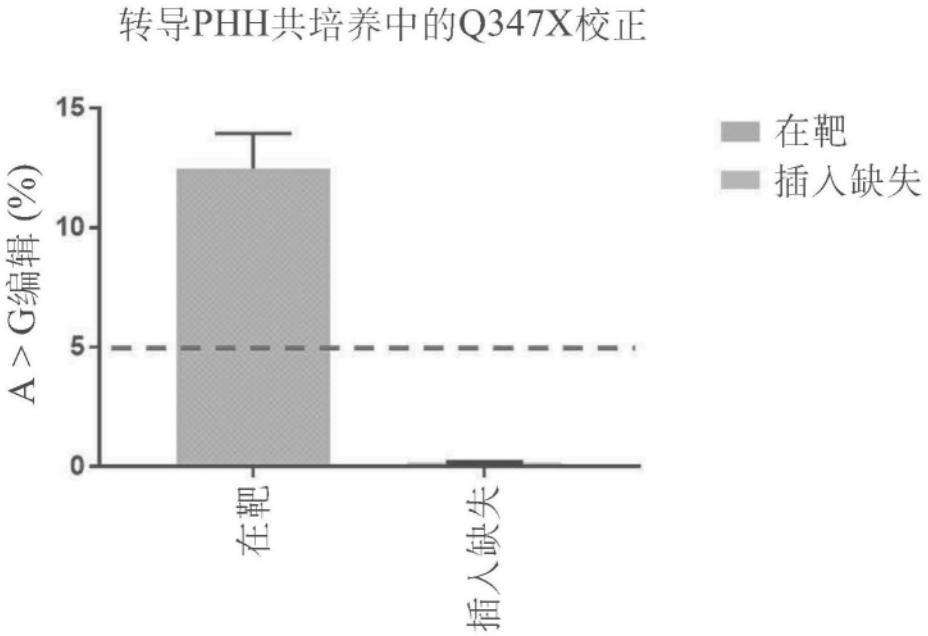

转导PHH共培养中的Q347X校正
A > G编辑 (%)
15
10
5
0
在靶
插入缺失
在靶
插入缺失

图10B

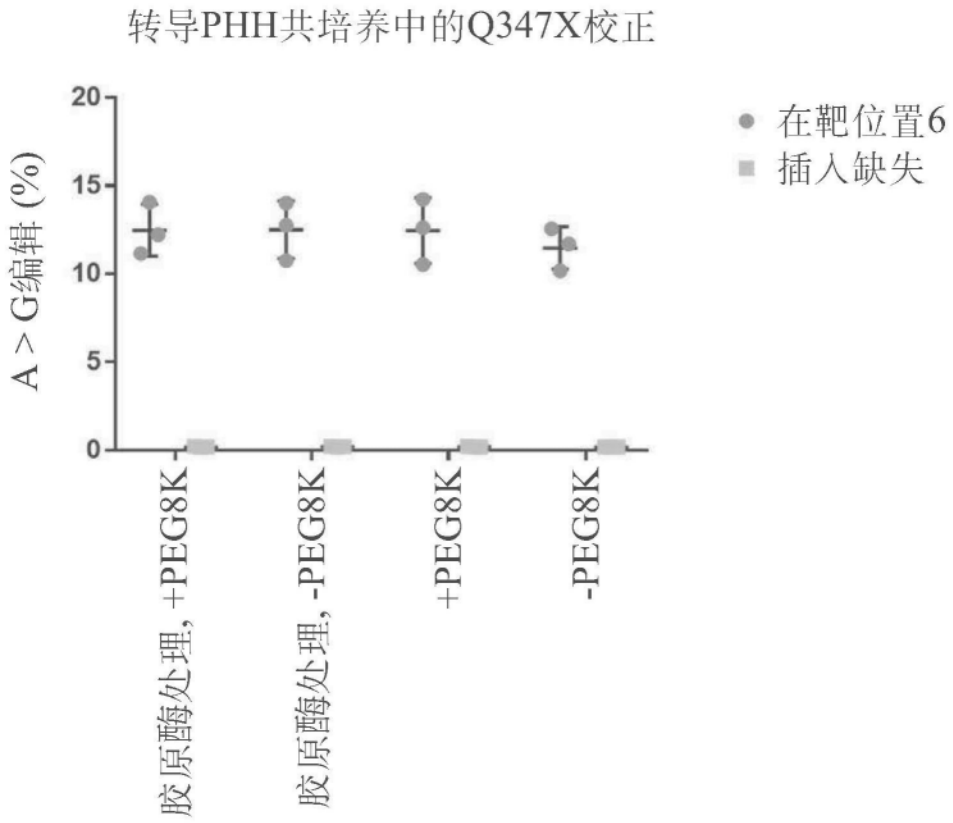

转导PHH共培养中的Q347X校正
A > G编辑 (%)
20
15
10
5
0
在靶位置6
插入缺失
胶原酶处理, +PEG8K
胶原酶处理, -PEG8K
+PEG8K
-PEG8K

图10C

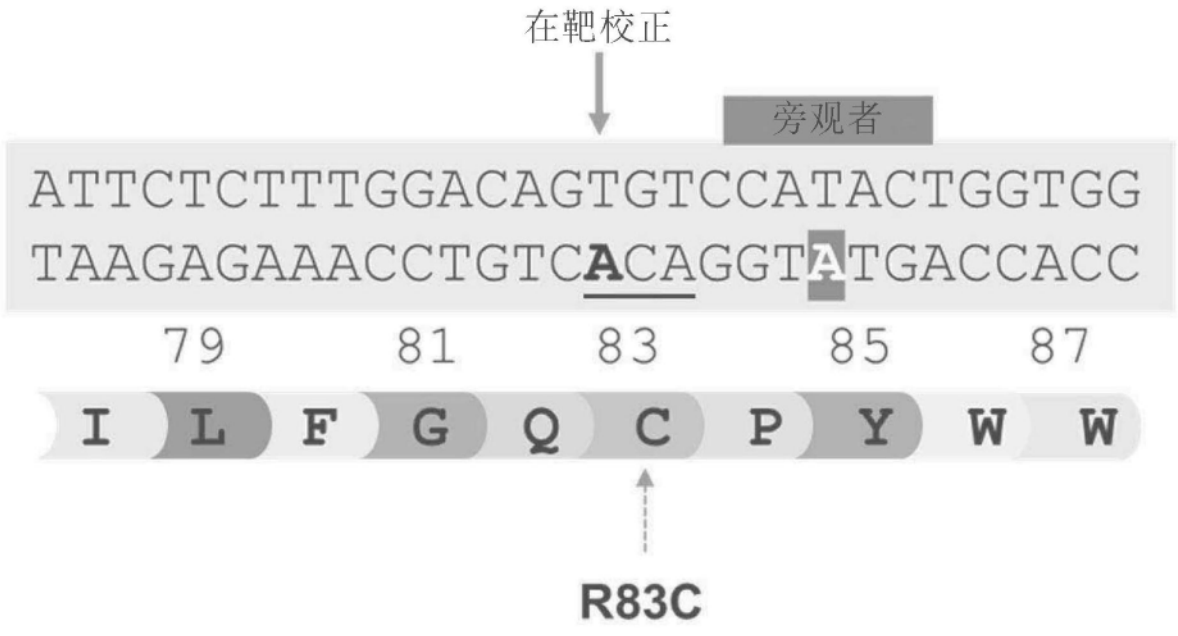

在靶校正
旁观者
ATTCTCTTTGGACAGTGTCCATACTGGTGG
TAAGAGAAACCTGTCACAGGTATGACCACC
79
81
83
85
87
I L F G Q C P Y W W
R83C

图11

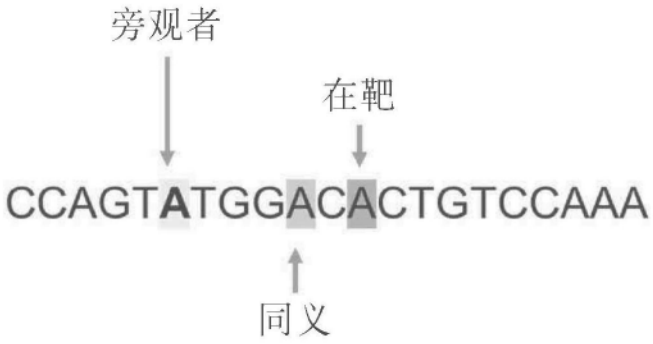


图12A

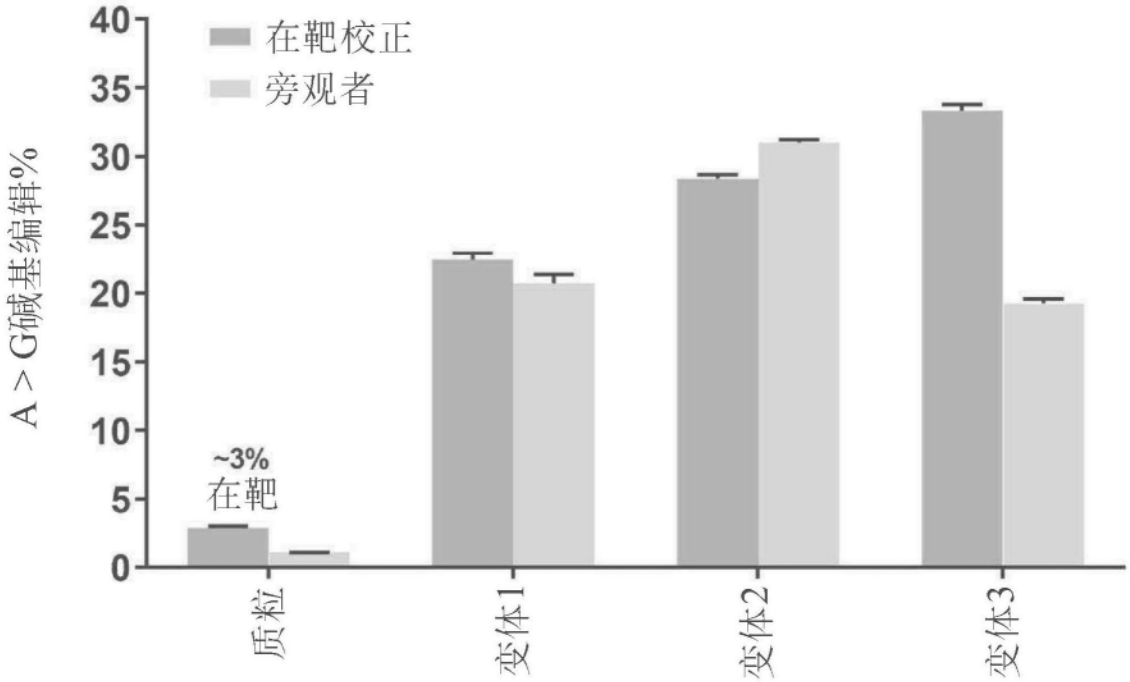


图12B

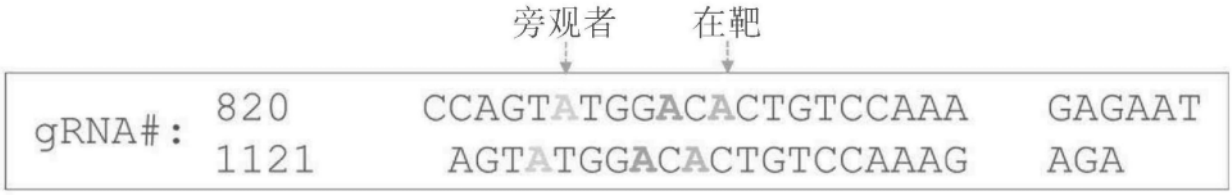


图13A

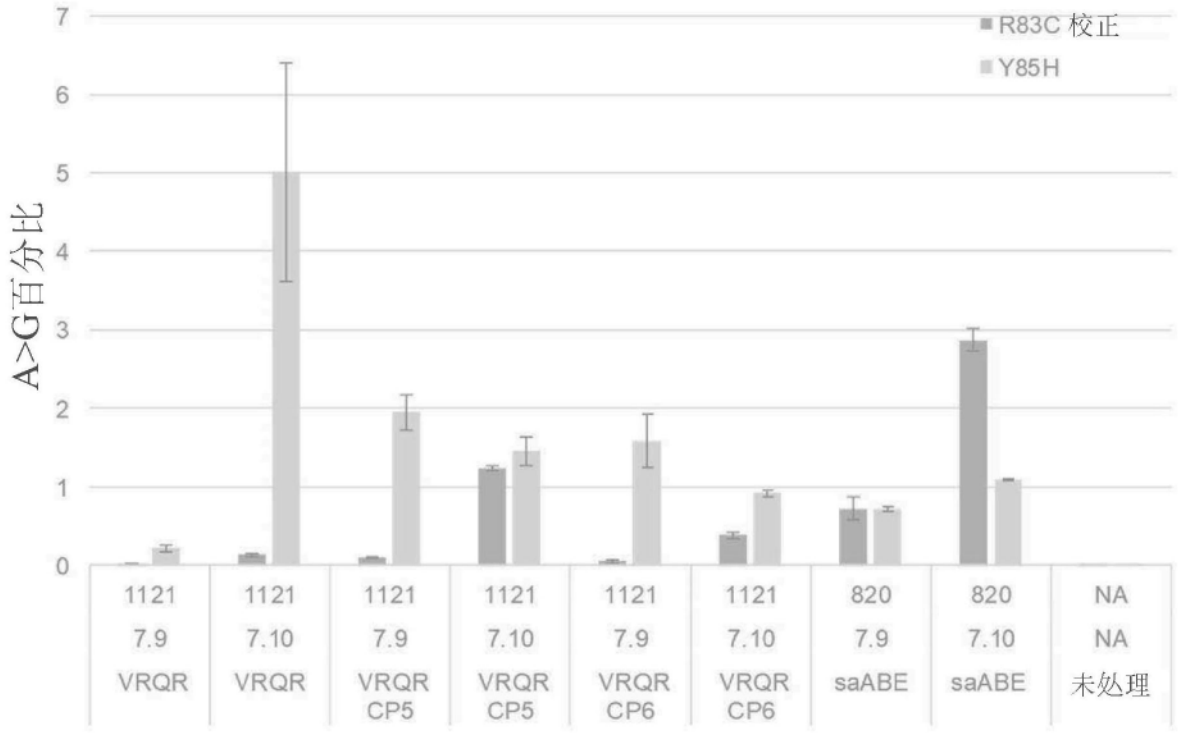

R83C 校正
Y85H
A>G百分比
0
1
2
3
4
5
6
7
1121 7.9 VRQR
1121 7.10 VRQR
1121 7.9 VRQR CP5
1121 7.10 VRQR CP5
1121 7.9 VRQR CP6
1121 7.10 VRQR CP6
820 7.9 saABE
820 7.10 saABE
NA NA 未处理

图13B

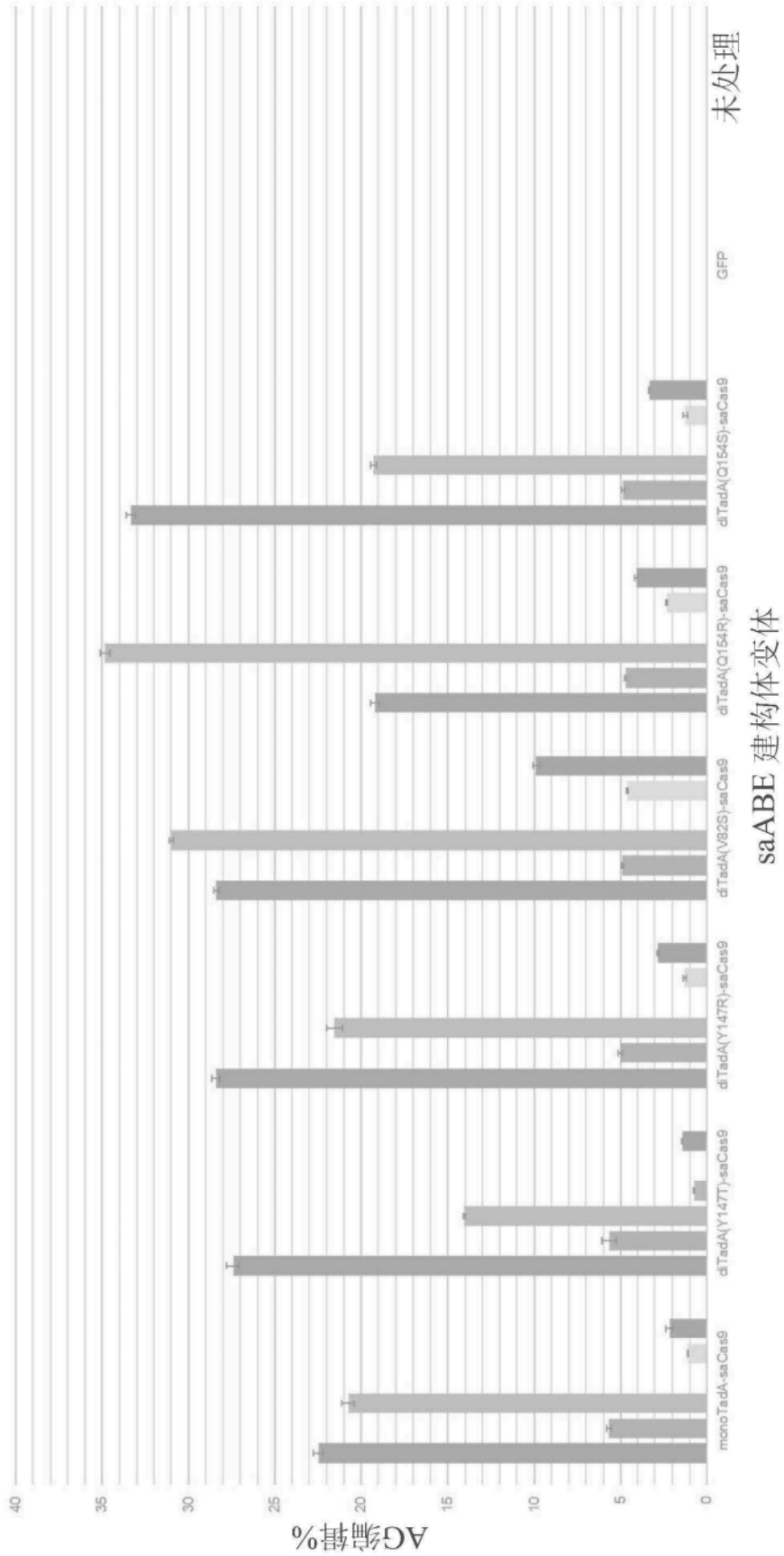

AG编辑%
0
5
10
15
20
25
30
35
40
monoTadA-saCas9
dTadA(Y147T)-saCas9
dTadA(Y147R)-saCas9
dTadA(V82S)-saCas9
dTadA(Q154R)-saCas9
dTadA(Q154S)-saCas9
GFP
未处理
saABE 建构体变体

图14

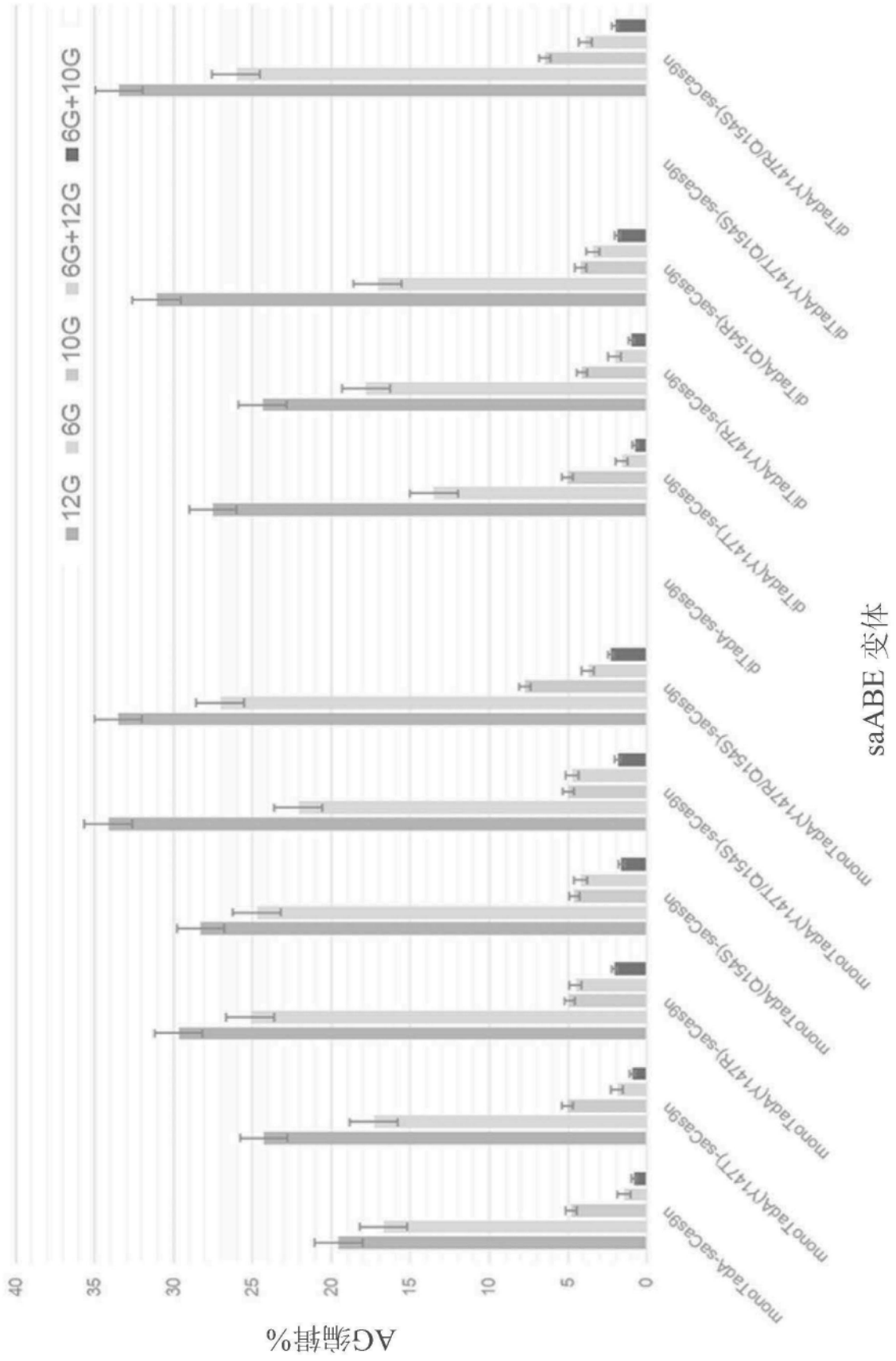

12G
6G
10G
6G+12G
6G+10G
AG编辑%
0
5
10
15
20
25
30
35
40
saABE 变体
monoTadA-saCas9n
monoTadA(Y147R/Q154S)-saCas9n
monoTadA(Y147T/Q154S)-saCas9n
monoTadA(Y147R)-saCas9n
monoTadA(Y147T)-saCas9n
monoTadA(Q154S)-saCas9n
monoTadA(Y147R/Q154S)-saCas9n
monoTadA(Y147T/Q154S)-saCas9n
dTadA-saCas9n
dTadA(Y147R/Q154S)-saCas9n
dTadA(Y147T/Q154S)-saCas9n
dTadA(Q154R)-saCas9n
dTadA(Y147R)-saCas9n
dTadA(Y147T)-saCas9n

图15

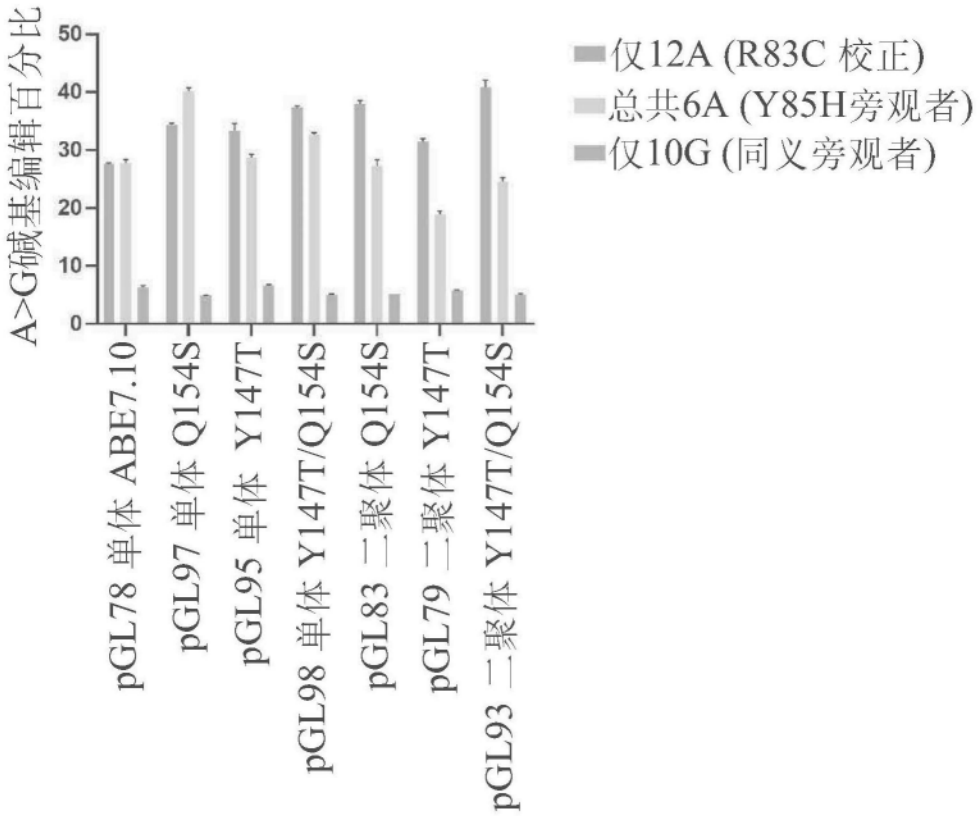

A>G碱基编辑百分比
50
40
30
20
10
0
仅12A (R83C 校正)
总共6A (Y85H旁观者)
仅10G (同义旁观者)
pGL78 单体 ABE7.10
pGL97 单体 Q154S
pGL95 单体 Y147T
pGL98 单体 Y147T/Q154S
pGL83 二聚体 Q154S
pGL79 二聚体 Y147T
pGL93 二聚体 Y147T/Q154S

图16

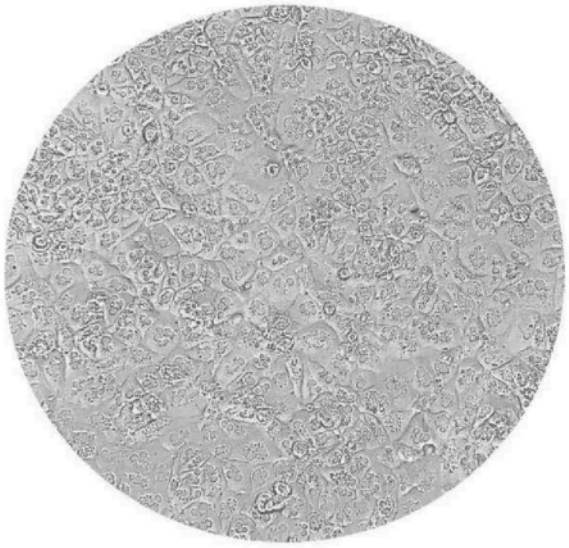


图17A

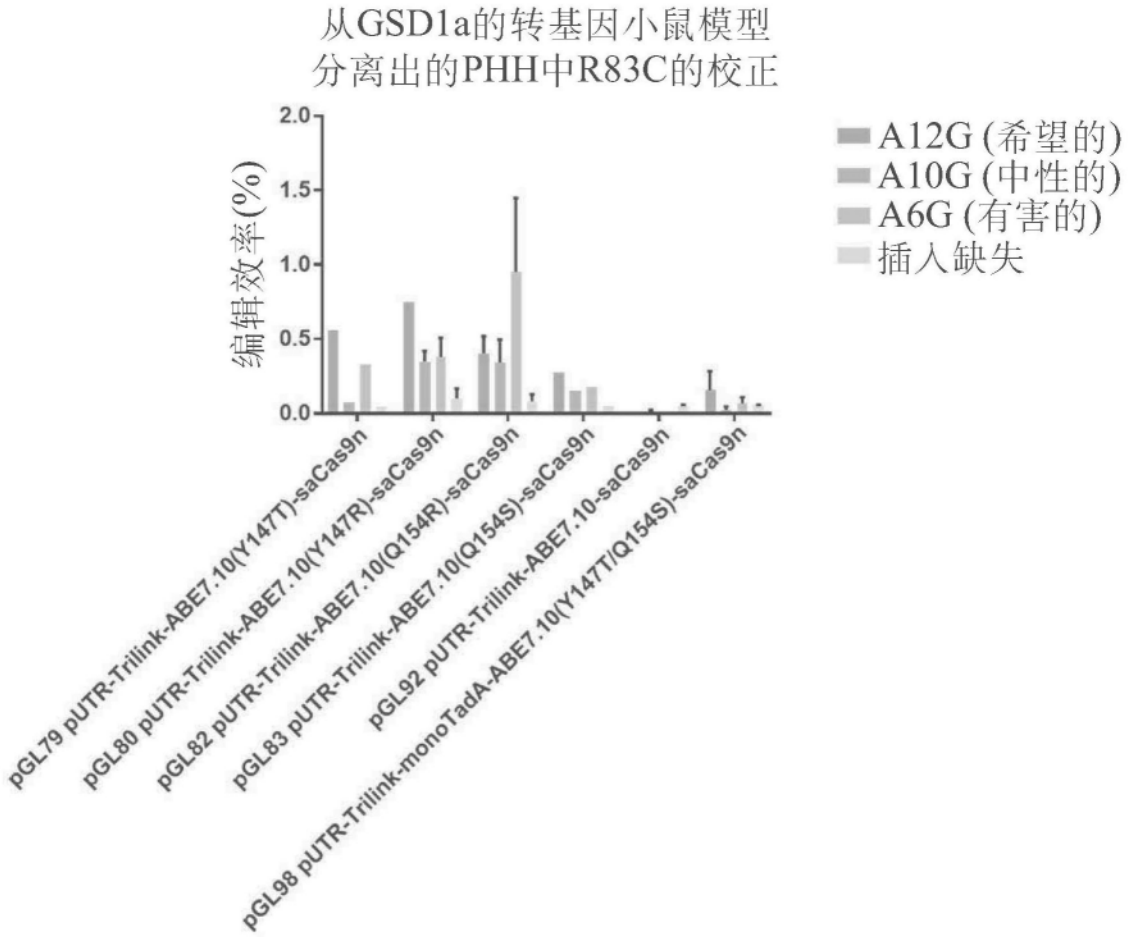
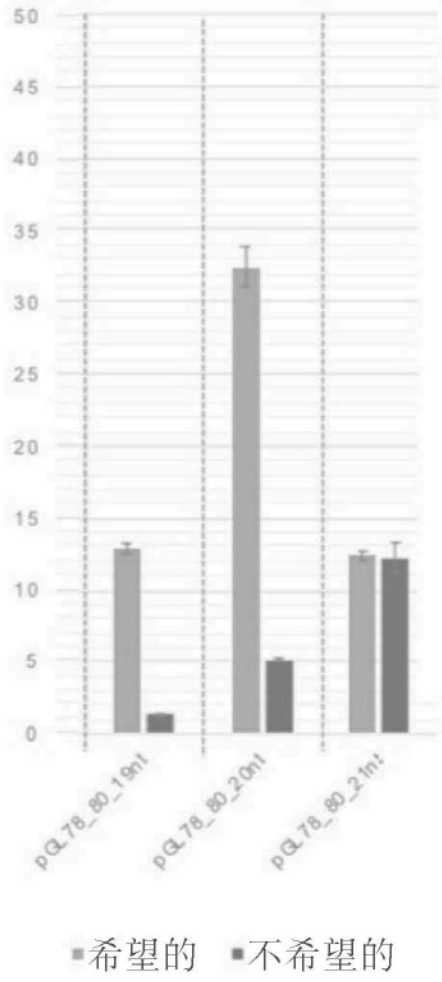

从GSD1a的转基因小鼠模型
分离出的PHH中R83C的校正
编辑效率(%)
2.0
1.5
1.0
0.5
0.0
A12G (希望的)
A10G (中性的)
A6G (有害的)
插入缺失
pGL79 pUTR-Trilink-ABE7.10(Y147T)-saCas9n
pGL80 pUTR-Trilink-ABE7.10(Y147R)-saCas9n
pGL82 pUTR-Trilink-ABE7.10(Q154R)-saCas9n
pGL83 pUTR-Trilink-ABE7.10(Q154S)-saCas9n
pGL92 pUTR-Trilink-ABE7.10-saCas9n
pGL98 pUTR-Trilink-monoTadA-ABE7.10(Y147T/Q154S)-saCas9n

图17B


50
45
40
35
30
25
20
15
10
5
0
pGL78_80_19nt
pGL78_80_20nt
pGL78_80_21nt
希望的
不希望的


说明:
PAM
原型间隔
原型间隔之外
靶位点:
1 4 7 10 13 16 19
CCA|CCA|GTA|TGG|ACA|CTG|TCC|AAA|GAG|AAT
AA序列:
W W Y P C Q G F L I
反向链靶位点:
CCAGTATGGACaCTGTCCAAAGAGAAT (gRNA820)
saCas9 使用 an NNGRRT PAM
编辑结果:
0G edit: W->R (不希望的)
3G edit: W->R (不希望的)
6G edit: Y->H (不希望的)
10G edit: C->C (中性的)
12G edit: C->R (希望的)
18T edit: G->R (不希望的)
gRNA截断研究
CCAGTATGGACaCTGTCCAAA (21nt)
CAGTATGGACaCTGTCCAAA (20nt)
AGTATGGACaCTGTCCAAA (19nt)

图18

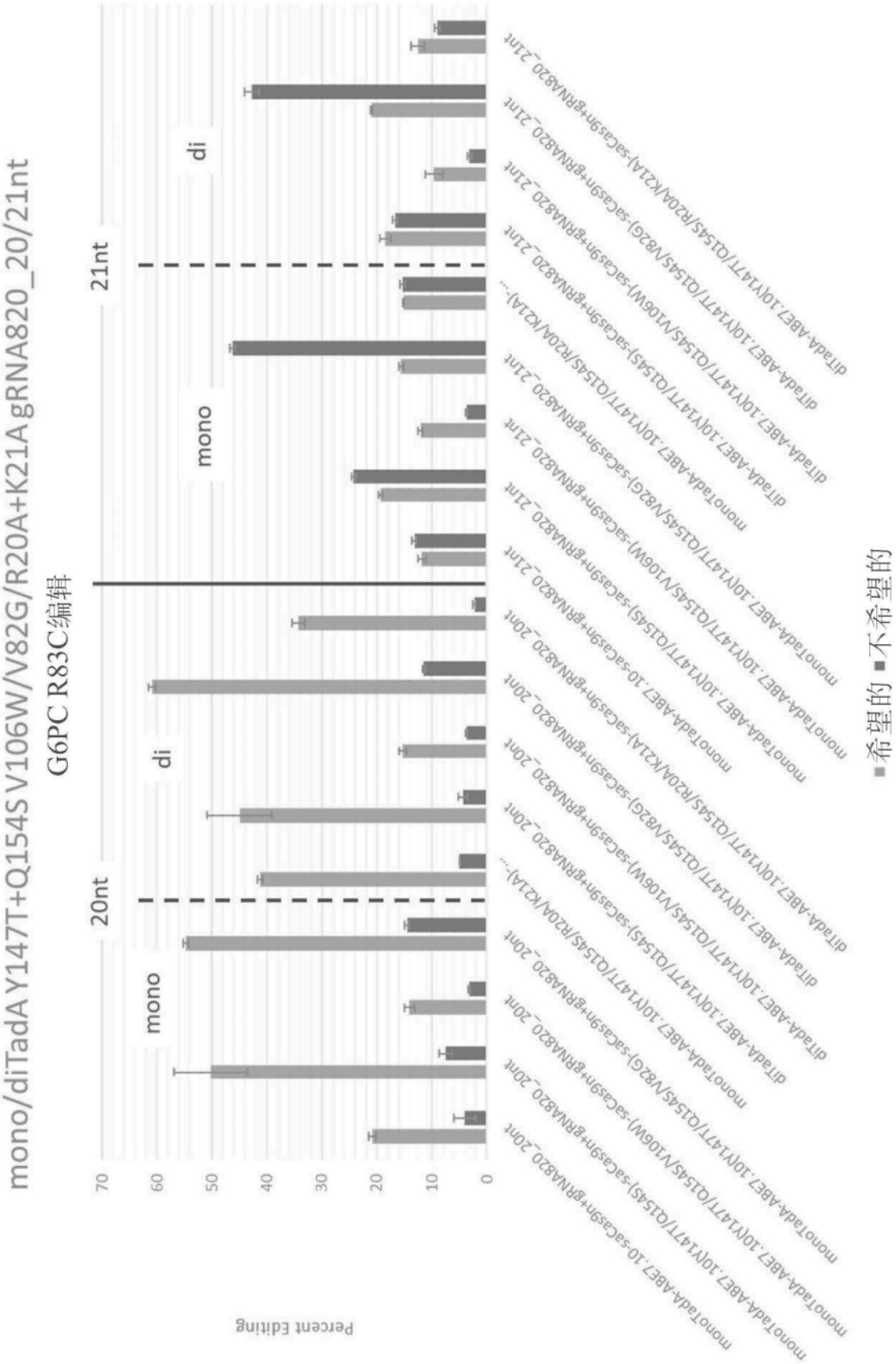
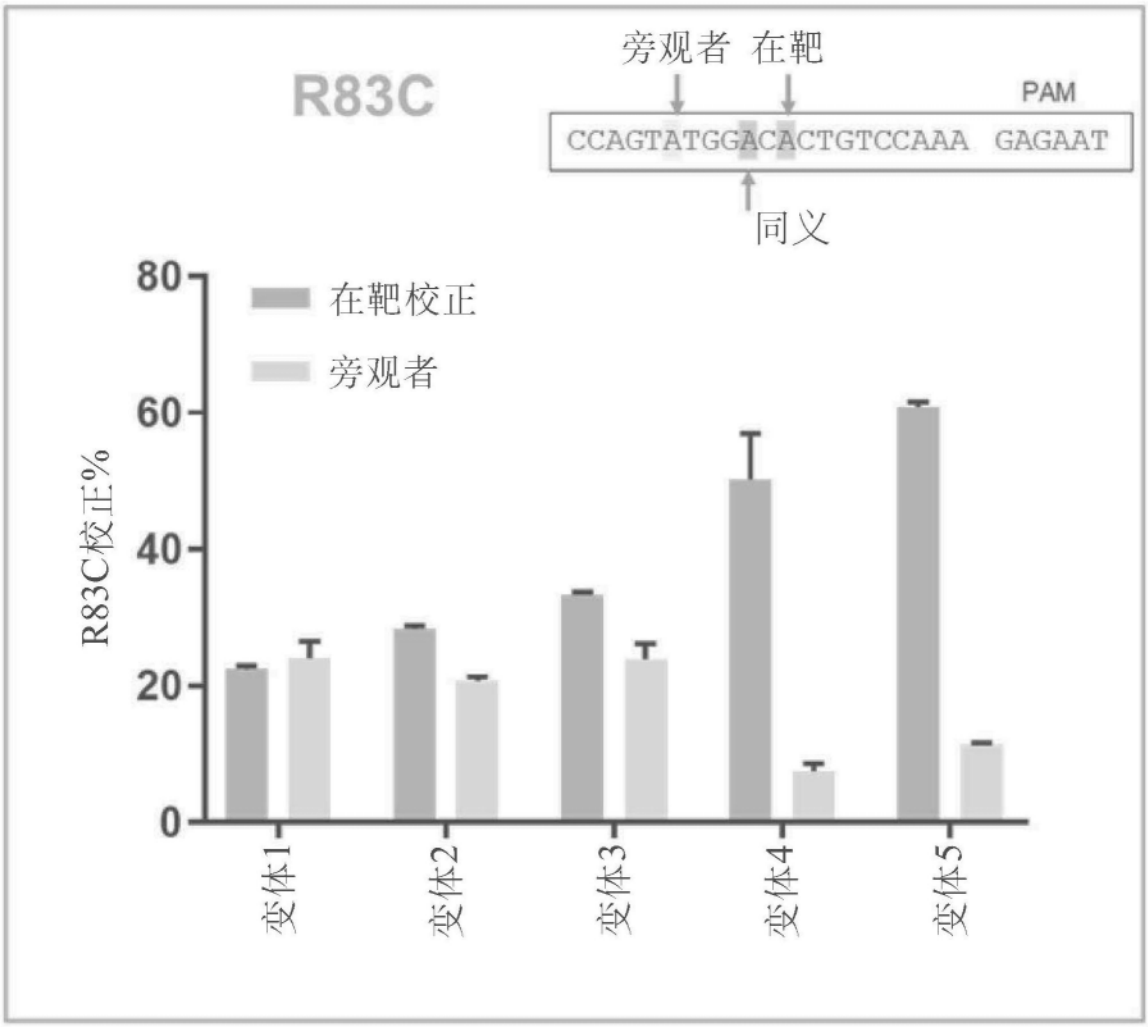
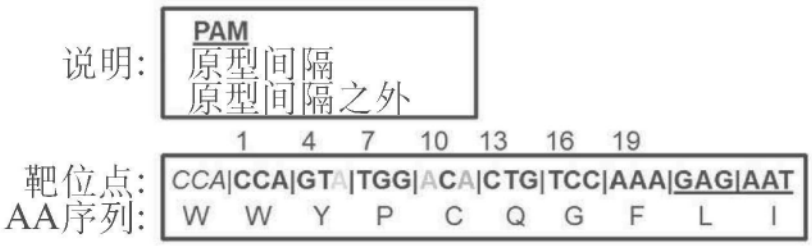

mono/diTadA Y147T+Q154S V106W/V82G/R20A+K21A gRNA820_20/21nt
G6PC R83C编辑
20nt
21nt
mono
di
Percent Editing
0
10
20
30
40
50
60
70
■希望的 ■不希望的

图19



	mRNA	gRNA
变体1	Mono TadA-SaCas9	21nt 820
变体2	Hetero TadA (Y147T) – SaCas9	21nt 820
变体3	Hetero TadA (Y147T, Q154S) – SaCas9	21nt 820
变体4	Mono TadA (Y147T, Q154S) – SaCas9	20nt 820
变体5	Hetero TadA (Y147T, Q154S, V82G) – SaCas9	20nt 820

图20


说明:
PAM
原型间隔
原型间隔之外
1 4 7 10 13 16 19
靶位点:
CCA|CCA|GTA|TGG|ACA|CTG|TCC|AAA|GAG|AAT
AA序列:
W W Y P C Q G F L I

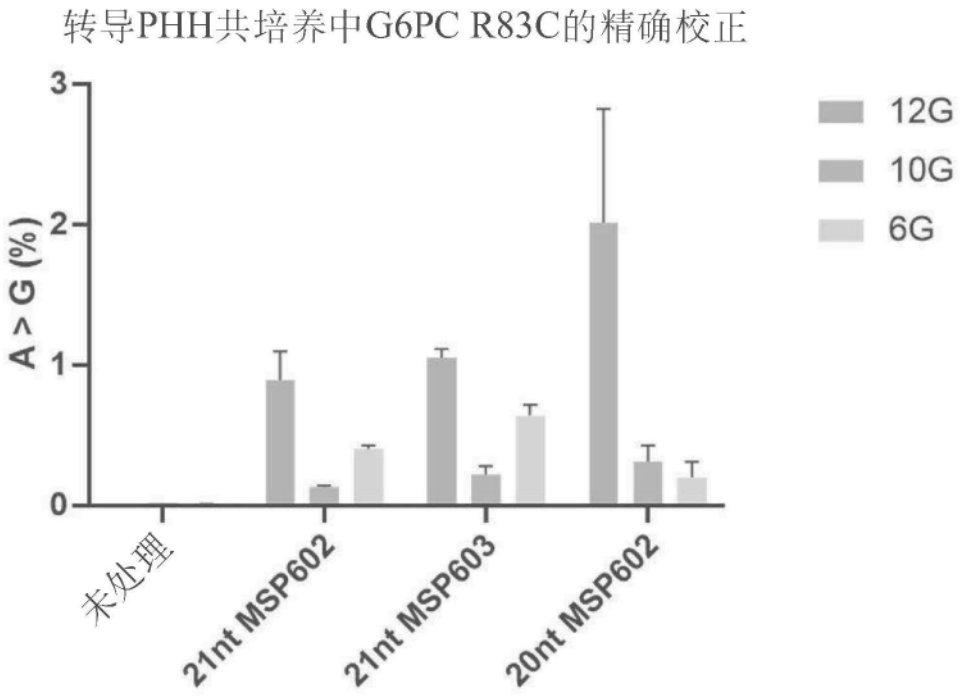
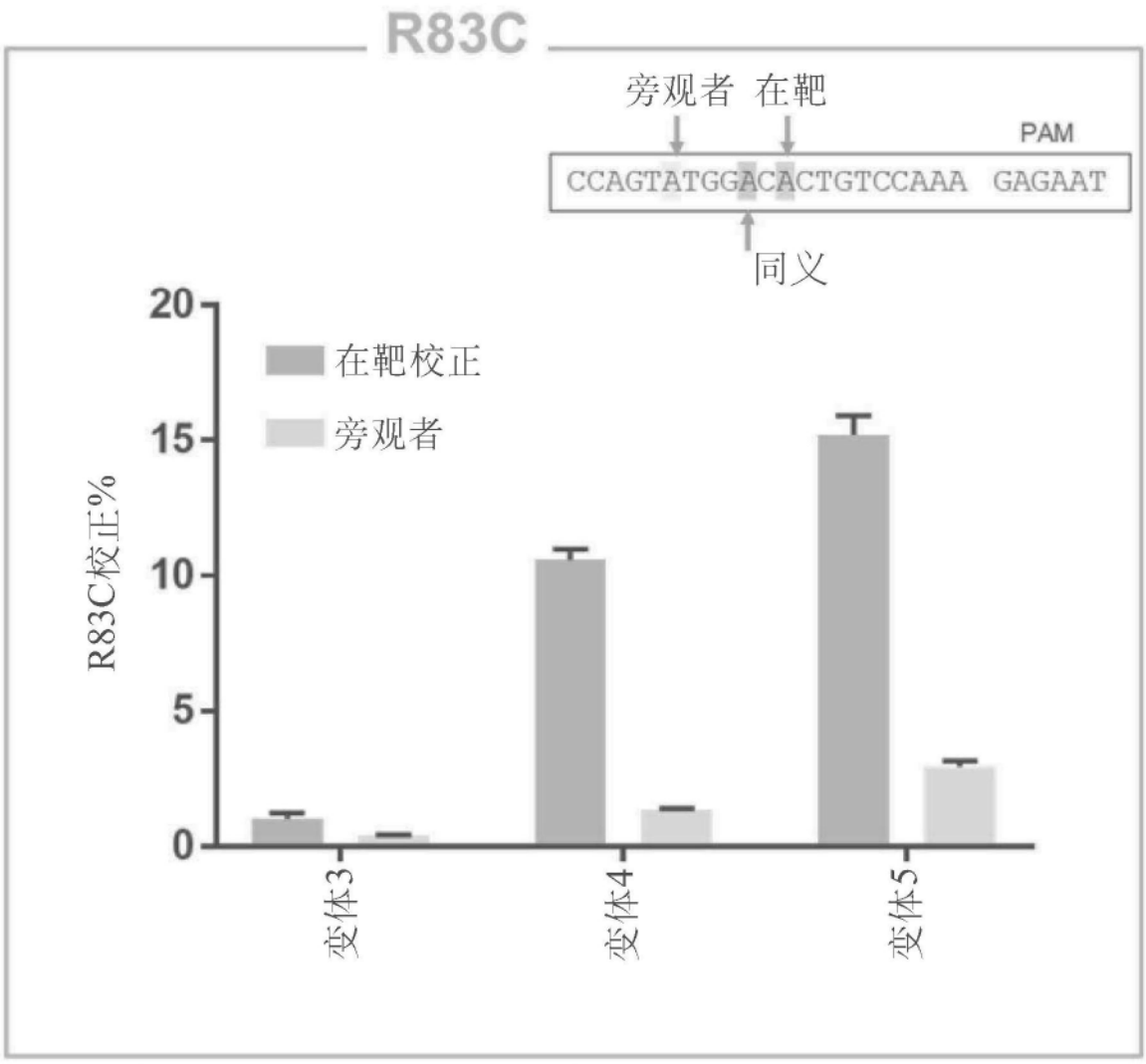

转导PHH共培养中G6PC R83C的精确校正
A > G (%)
3
2
1
0
12G
10G
6G
未处理
21nt MSP602
21nt MSP603
20nt MSP602

图21



	mRNA	gRNA
变体3	Hetero TadA (Y147T, Q154S) – SaCas9	21nt 820
变体4	Mono TadA (Y147T, Q154S) – SaCas9	20nt 820
变体5	Hetero TadA (Y147T, Q154S, V82G) – SaCas9	20nt 820

图22

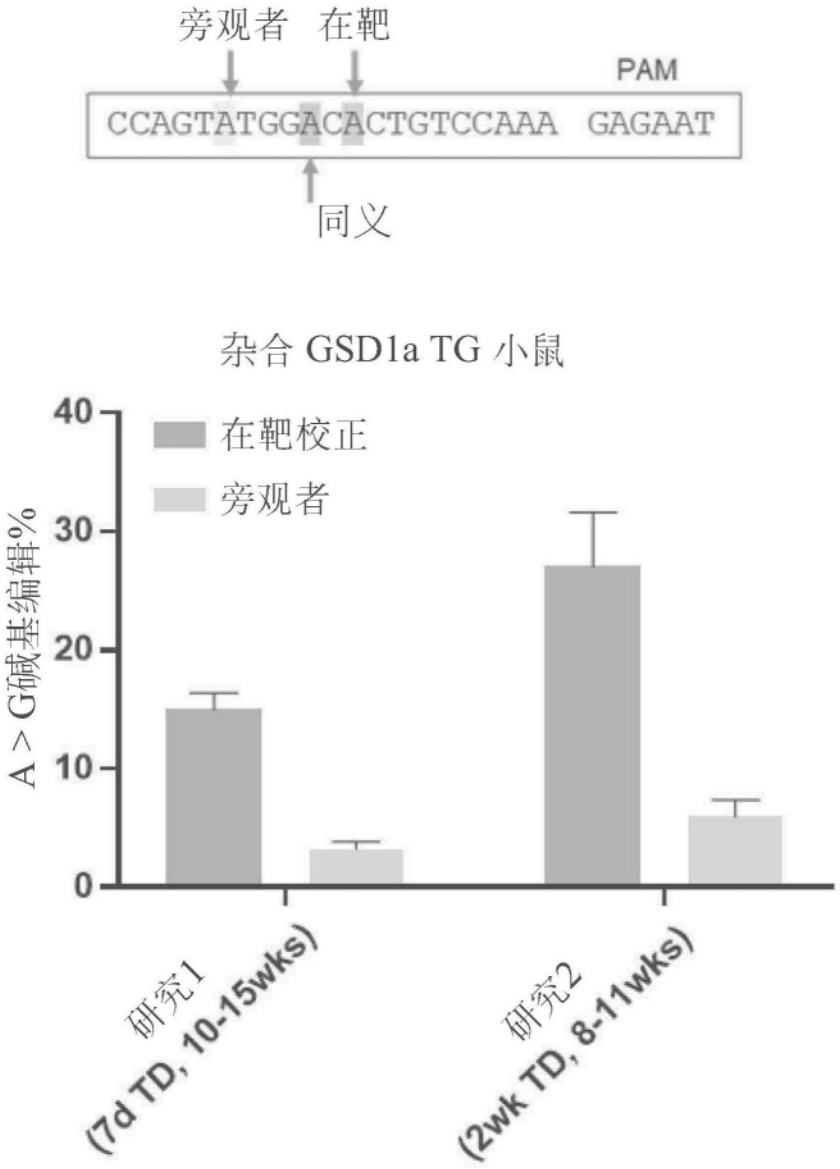

旁观者
在靶
PAM
CCAGTATGGACACTGTCCAAA GAGAAT
同义
杂合 GSD1a TG 小鼠
在靶校正
旁观者
A > G碱基编辑%
40
30
20
10
0
研究1
(7d TD, 10-15wks)
研究2
(2wk TD, 8-11wks)

图23

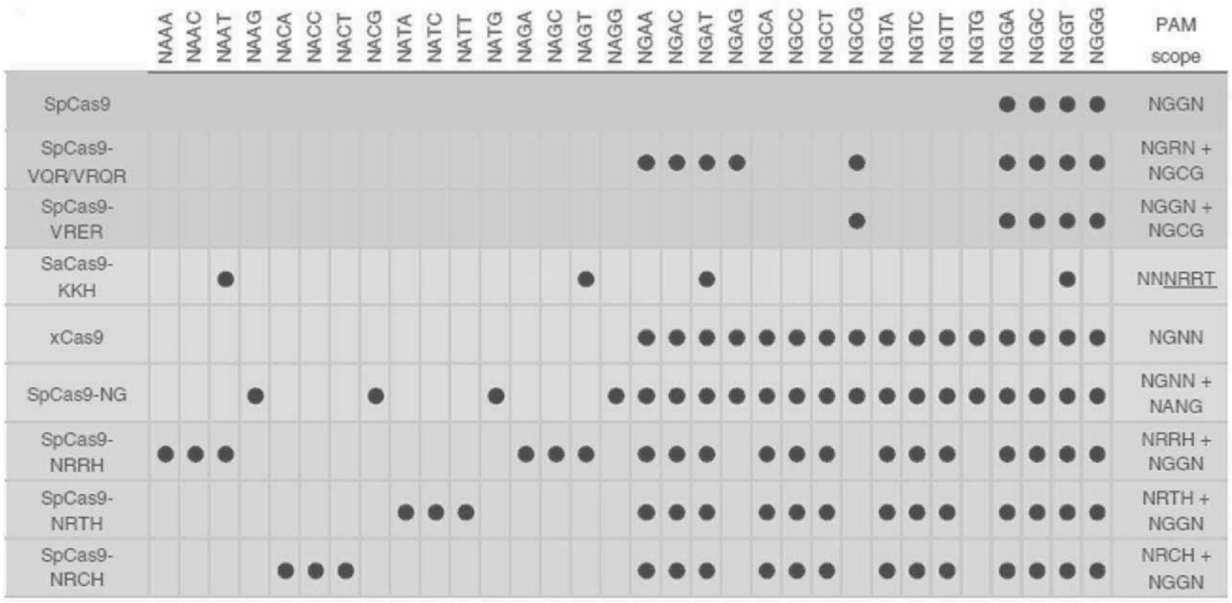

NAAA NAAC NAAT NAAG NACA NACC NACT NACG NATA NATC NATT NATG NAGA NAGC NAGT NAGG NGAA NGAC NGAT NGAG NGCA NGCC NGCT NGCG NGTA NGTC NGTT NGTG NGGA NGGC NGGT NGGG PAM scope
SpCas9 NGGN
SpCas9-VQR/VRQR NGRN + NGCG
SpCas9-VRER NGGN + NGCG
SaCas9-KKH NNNRRT
xCas9 NGNN
SpCas9-NG NGNN + NANG
SpCas9-NRRH NRRH + NGGN
SpCas9-NRTH NRTH + NGGN
SpCas9-NRCH NRCH + NGGN

图24