(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 114040970 B
(45)授权公告日 2024.09.27

(21)申请号 202080028186.5
(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 114040970 A
(43)申请公布日 2022.02.11
(30)优先权数据
62/805,271 2019.02.13 US
62/850,919 2019.05.21 US
62/852,228 2019.05.23 US
62/852,224 2019.05.23 US
62/873,138 2019.07.11 US
62/888,867 2019.08.19 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/US2020/018073 2020.02.13
(87)PCT国际申请的公布数据
WO2020/168051 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 I·斯雷梅克 N·戈代尔利
Y·于 B·蔡澈 D·A·玻恩
S-J·李 M·帕克 J·M·格尔克
N·彼得罗相 A·梅萨纳
S·贝尔科维奇
(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 9/22(2006.01)
C12N 9/78(2006.01)
(56)对比文件
WO 2018027078 A1,2018.02.08
审查员 许文彦

权利要求书9页 说明书385页
序列表（电子公布） 附图125页

(54)发明名称

使用腺苷脱氨酶碱基编辑器编辑疾病相关基因的方法,包括遗传性疾病的治疗

(57)摘要

本发明提供包括新颖可编程的腺苷碱基编辑器系统(例如,ABE8)的组合物,其提供通过向受试者给药具有提高的效率的可编程的腺苷碱基编辑器系统(例如,ABE8)来治疗受试者的疾病或病症(例如,帕金森氏病、贺勒氏症、蕾特氏症或斯塔加特氏病)的方法及使用这些腺苷脱氨酶变体编辑疾病相关基因的方法。



CN 114040970 B

1.一种组合物于制备用于治疗受试者的贺勒氏症的药物的用途,所述组合物包括(i)腺苷碱基编辑器或编码所述腺苷碱基编辑器的核酸序列和(ii)向导多核苷酸,其中所述腺苷碱基编辑器包括可编程的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相较于SEQ ID NO:20具有选自由以下任一氨基酸取代:

Y147T;

Y147R;

Q154S;

T166R;

Y123H、Y147R和Q154R;

I76Y、Y147R和Q154R;

Y147R、Q154R和T166R;

Y147T和Q154R;

Y147T和Q154S;或

I76Y、Y123H、Y147R和Q154R,

以及其中所述向导多核苷酸引导所述腺苷碱基编辑器在所述受试者的α-L-艾杜糖醛酸酶(IDUA)基因中实现A至G核碱基改变,从而治疗所述受试者的贺勒氏症,其中所述IDUA基因具有与贺勒氏症相关的SNP,所述SNP为相对SEQ ID NO:4发生W401X或W402X突变,其中X是终止密码子,且其中所述与贺勒氏症相关的SNP处的A至G改变将所述IDUA基因所编码的IDUA多肽中的终止密码子改变为色氨酸。

2.根据权利要求1所述的用途,其中所述组合物改善与贺勒氏症相关的至少一种症状。

3.根据权利要求1或2所述的用途,其中所述向导多核苷酸具有间隔,所述间隔具有核酸序列,所述核酸序列与具有与所述贺勒氏症相关的SNP的IDUA基因互补。

4.根据权利要求1或2所述的用途,其中所述腺苷碱基编辑器与单个向导RNA(sgRNA)形成复合物,所述sgRNA具有间隔,所述间隔具有核酸序列,所述核酸序列与具有与所述贺勒氏症相关的SNP的IDUA基因互补,其中所述间隔具有选自以下任一核酸序列:5'-GACUCUAGGCAGAGGUCUCAA-3'、5'-ACUCUAGGCAGAGGUCUCAA-3'、5'-CUCUAGGCGAAGUGUCGC-3'或5'-GCUCUAGGCCGAAGUGUCGC-3'。

5.一种组合物于制备用于治疗受试者的帕金森氏病的药物的用途,所述组合物包括:(i)腺苷碱基编辑器或编码所述腺苷碱基编辑器的核酸序列和(ii)向导多核苷酸,其中所述腺苷碱基编辑器包括可编程的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相较于SEQ ID NO:20具有选自以下任一氨基酸取代:

Y147T;

Y147R;

Q154S;

V82S;

T166R;

Y123H、Y147R和Q154R;

I76Y、Y147R和Q154R;

Y147R、Q154R和T166R;

Y147T和Q154R;

Y147T和Q154S;

I76Y、Y123H、Y147R和Q154R;或

V82S和Q154R,

以及其中所述向导多核苷酸引导所述腺苷碱基编辑器在所述受试者中影响富含亮氨酸的重复激酶-2(LRRK2)基因实现A至G核碱基改变,从而治疗受试者的帕金森氏病,其中所述LRRK2基因具有与帕金森氏病相关的SNP,所述SNP为相对于SEQ ID NO:3发生R1441C或G2019S突变。

6.根据权利要求5所述的用途,其中所述组合物改善与帕金森氏病相关的至少一种症状。

7.根据权利要求5所述的用途,其中所述向导多核苷酸具有间隔,所述间隔具有核酸序列,所述核酸序列与具有与所述帕金森氏病相关的SNP的LRRK2基因互补。

8.根据权利要求5所述的用途,其中所述腺苷碱基编辑器与sgRNA形成复合物,所述sgRNA具有间隔,所述间隔具有核酸序列,所述核酸序列与具有与帕金森氏病相关的SNP的LRRK2基因互补,其中所述间隔具有核酸序列:5'-AAGCGCAAGCCUGGAGGGAA-3';或5'-ACUACAGCAUUGCUCAGUAC-3'。

9.一种组合物于制备用于治疗受试者的蕾特氏症的药物的用途,所述组合物包括(i)腺苷碱基编辑器或编码所述腺苷碱基编辑器的核酸序列和(ii)向导多核苷酸,其中所述腺苷碱基编辑器包括可编程的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相较于SEQ ID NO:20具有选自以下任一氨基酸取代:

Y123H、Y147R和Q154R;

I76Y、Y147R和Q154R;或

I76Y、Y123H、Y147R和Q154R,

以及其中所述向导多核苷酸引导所述腺苷碱基编辑器在所述受试者中实现甲基CpG结合蛋白2(MECP2)基因中的A至G核碱基改变,从而治疗所述受试者的蕾特氏症,其中所述MECP2基因具有与蕾特氏症相关的SNP,所述SNP为相对于SEQ ID NO:5发生R106W或R255X突变,其中X是终止密码子,其中所述A至G核碱基改变在与所述蕾特氏症相关的SNP处,且其中所述A至G核碱基改变将所述与蕾特氏症相关的SNP改变为野生型核碱基。

10.根据权利要求9所述的用途,其中所述组合物改善与根据蕾特氏症相关的至少一种症状。

11.根据权利要求9所述的用途,其中所述向导多核苷酸具有间隔,所述间隔具有核酸序列,所述核酸序列与具有与蕾特氏症相关的SNP的MECP2基因互补。

12.根据权利要求9所述的用途,其中所述腺苷碱基编辑器与sgRNA形成复合物,所述sgRNA具有间隔,所述间隔具有核酸序列,所述核酸序列与具有与蕾特氏症相关的SNP的MECP2基因互补,其中所述间隔具有选自以下任一核酸序列:5'-CUUUUCACUUCCUGCCGGGG-3'、5'-AGCUUCCAUGUCCAGCCUUC-3'、5'-ACCAUGAAGUCAAAAUCAUU-3'或5'-GCUUUCAGCCCCGUUUCUUG-3'。

13.一种组合物于制备用于治疗受试者的斯塔加特氏病的药物的用途,所述组合物包括(i)腺苷碱基编辑器或编码所述腺苷碱基编辑器的核酸序列和(ii)向导多核苷酸,其中

所述腺苷碱基编辑器包括可编程的DNA结合结构域和腺苷脱氨酶结构域，

其中腺苷脱氨酶结构域相较于SEQ ID NO:20具有V82S的氨基酸取代，

以及其中所述向导多核苷酸引导所述腺苷碱基编辑器在所述受试者中实现ATP结合盒亚家族成员4(ABCA4)基因中的A至G核碱基改变，从而治疗所述受试者中的斯塔加特氏病，其中所述ABCA4基因具有与斯塔加特氏病相关的SNP，所述SNP为相对于SEQ ID NO:6发生G1961E突变，其中所述A至G核碱基改变在与斯塔加特氏病相关的SNP处，其中所述A至G核碱基改变将所述与斯塔加特氏病相关的SNP改变为野生型核碱基。

14.如权利要求13所述的用途，其中所述组合物改善至少一种与斯塔加特氏病相关的症状。

15.根据权利要求13所述用途，其中所述向导多核苷酸具有间隔，所述间隔具有核酸序列，所述核酸序列与具有与所述斯塔加特氏病相关的SNP的ABCA4基因互补。

16.根据权利要求13所述用途，其中所述腺苷碱基编辑器与sgRNA形成复合物，所述sgRNA具有间隔，所述间隔具有核酸序列，所述核酸序列与具有与斯塔加特氏相关的SNP的ABCA4基因互补，其中所述间隔具有序列5'-CUCCAGGGCGAACUUCGACACACAGC-3'。

17.一种组合物于制备用于编辑具有与帕金森氏病相关的SNP的富含亮氨酸的重复激酶-2(LRRK2)基因的药物的用途，所述编辑富含亮氨酸的重复激酶-2(LRRK2)基因包括使LRRK2基因或其调控组件与包括以下的组合物接触(i)腺苷碱基编辑器或编码所述腺苷碱基编辑器的核酸序列和(ii)向导多核苷酸，其中所述腺苷碱基编辑器包括可编程的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相较于SEQ ID NO:20具有选自以下任一氨基酸取代：

Y147T；

Y147R；

Q154S；

V82S；

T166R；

Y123H、Y147R和Q154R；

I76Y、Y147R和Q154R；

Y147R、Q154R和T166R；

Y147T和Q154R；

Y147T和Q154S；

I76Y、Y123H、Y147R和Q154R；或

V82S和Q154R，

以及其中所述向导多核苷酸引导所述腺苷碱基编辑器在LRRK2基因中实现A至G核碱基改变，其中所述A至G核碱基改变在与所述帕金森氏病相关的SNP处，且其中所述与帕金森氏病相关的SNP为相较于SEQ ID NO:3发生R1441C或G2019S突变。

18.根据权利要求17所述的用途，其中所述向导多核苷酸具有间隔，所述间隔具有核酸序列，所述核酸序列与具有与所述帕金森氏病相关的SNP的LRRK2基因互补。

19.根据权利要求17所述的用途，其中所述腺苷碱基编辑器与sgRNA形成复合物，所述sgRNA具有间隔，所述间隔具有核酸序列，所述核酸序列与具有与帕金森氏病相关的SNP的

LRRK2基因互补,其中所述间隔具有核酸序列:5'-AAGCGCAAGCCUGGAGGGAA-3';或5'-ACUACAGCAUUGCUCAGUAC-3'。

20.一种组合物于制备用于编辑α-L-艾杜糖醛酸酶(IDUA)基因的药物的用途,所述编辑α-L-艾杜糖醛酸酶(IDUA)基因包括使所述IDUA基因与包括以下的组合物接触(i)腺苷碱基编辑器或编码所述腺苷碱基编辑器的核酸序列和(ii)向导多核苷酸,其中所述腺苷碱基编辑器包括可编程的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相较于SEQ ID NO:20具有选自以下任一氨基酸取代:

Y147T;

Y147R;

Q154S;

T166R;

Y123H、Y147R和Q154R;

I76Y、Y147R和Q154R;

Y147R、Q154R和T166R;

Y147T和Q154R;

Y147T和Q154S;或

I76Y、Y123H、Y147R和Q154R,

以及其中所述向导多核苷酸引导所述腺苷碱基编辑器在IDUA基因中实现A至G核碱基改变,其中所述IDUA基因具有与贺勒氏症相关的SNP,所述SNP为相较于SEQ ID NO:4发生W402X或W401X突变,其中X是终止密码子,其中所述A至G核碱基改变在与贺勒氏症相关的SNP处,且其中所述与贺勒氏症相关的SNP处的A至G改变将由所述IDUA基因所编码的IDUA多肽中的终止密码子改变为色氨酸。

21.根据权利要求20所述的用途,其中所述向导多核苷酸具有间隔,所述间隔具有核酸序列,所述核酸序列与包括与贺勒氏症相关的SNP的IDUA基因互补。

22.根据权利要求20所述的用途,其中所述腺苷碱基编辑器与sgRNA形成复合物,所述sgRNA具有间隔,所述间隔具有核酸序列,所述核酸序列与包括与贺勒氏症相关的SNP的IDUA基因互补,其中所述间隔具有选自以下任一的核酸序列:5'-GACUCUAGGCAGAGGUCUCAA-3'、5'-ACUCUAGGCAGAGGUCUCAA-3'、5'-CUCUAGGCCGAAGUGUCGC-3'或5'-GCUCUAGGCCGAAGUGUCGC-3'。

23.一种组合物于制备用于编辑甲基CpG结合蛋白2(MECP2)基因的药物的用途,所述组合物包括(i)腺苷碱基编辑器或编码所述腺苷碱基编辑器的核酸序列和(ii)向导多核苷酸,其中所述腺苷碱基编辑器包括可编程的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相较于SEQ ID NO:20具有选自以下任一氨基酸取代:

Y123H、Y147R和Q154R;

I76Y、Y147R和Q154R;或

I76Y、Y123H、Y147R和Q154R,

以及其中所述向导多核苷酸引导所述腺苷碱基编辑器在MECP2基因中实现A至G核碱基改变,其中所述MECP2基因具有与蕾特氏症相关的SNP,所述SNP为相较于SEQ ID NO:5发生R106W或R255X突变,其中X是终止密码子,其中所述A至G核碱基改变在与所述蕾特氏症相关

的SNP处,且其中所述A至G核碱基改变将所述与蕾特氏症相关的SNP改变为野生型核碱基。

24.根据权利要求23所述的用途,其中所述向导多核苷酸具有间隔,所述间隔具有核酸序列,所述核酸序列与具有所述与蕾特氏症相关的SNP的MECP2基因互补。

25.根据权利要求23所述的用途,其中所述腺苷碱基编辑器与sgRNA形成复合物,所述sgRNA具有间隔,所述间隔具有核酸序列,所述核酸序列与具有与蕾特氏症相关的SNP的MECP2基因互补,其中所述间隔具有选自以下任一核酸序列:5'-CUUUUCACUUCCUGCCGGGG-3'、5'-AGCUUCCAUGUCCAGCCUUC-3'、5'-ACCAUGAAGUCAAAAUCAUU-3'或5'-GCUUUCAGCCCCGUUUCUUG-3'。

26.一种组合物于制备用于编辑ATP结合盒亚家族成员4(ABCA4)基因的药物的用途,所述编辑ATP结合盒亚家族成员4(ABCA4)基因包括使ABCA4基因或其调控组件与包括以下的组合物接触(i)腺苷碱基编辑器或编码所述腺苷碱基编辑器的核酸序列和(ii)向导多核苷酸,其中所述腺苷碱基编辑器包括可编程的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相较于SEQ ID NO:20具有V82S的氨基酸取代,

以及其中所述向导多核苷酸引导所述腺苷碱基编辑器在ABCA4基因中实现A至G核碱基改变,其中所述ABCA4基因具有与斯塔加特氏病相关的SNP,所述SNP为相较于SEQ ID NO:6发生G1961E突变,其中所述A至G核碱基改变在与所述斯塔加特氏病相关的SNP处,且其中所述A至G核碱基改变将所述与斯塔加特氏病相关的SNP改变为野生型核碱基。

27.根据权利要求26所述的用途,其中所述组合物改善至少一种与斯塔加特氏病相关的症状。

28.根据权利要求26所述的用途,其中所述向导多核苷酸具有间隔,所述间隔具有核酸序列,所述核酸序列与具有与斯塔加特氏病相关的SNP的ABCA4基因互补。

29.根据权利要求26所述的用途,其中所述腺苷碱基编辑器与sgRNA形成复合物,所述sgRNA具有间隔,所述间隔具有核酸序列,所述核酸序列与具有与斯塔加特氏病相关的SNP的ABCA4基因互补,其中所述间隔具有序列5'-CUCCAGGGCGAACUUCGACACACAGC-3'。

30.根据权利要求17、20和26中任一项所述的用途,其中所述接触在细胞中进行。

31.根据权利要求30所述的用途,其中所述接触导致所述细胞基因组中小于10%的插入缺失,其中插入缺失率通过在单核苷酸修饰侧边的序列与未修饰序列之间的错配频率来测量。

32.根据权利要求30所述的用途,其中所述接触导致细胞中基因组中小于5%的插入缺失,其中插入缺失率通过在单核苷酸修饰侧边的序列与未修饰序列之间的错配频率来测量。

33.根据权利要求30所述用途,其中所述接触导致所述细胞基因组中小于1%的插入缺失,其中插入缺失率通过在单核苷酸修饰侧边的序列与未修饰序列之间的错配频率来测量。

34.根据权利要求30所述的用途,其中所述细胞是神经元。

35.根据权利要求17、20和26中任一项所述的用途,其中所述接触是在细胞群中。

36.根据权利要求35所述的用途,其中在所述接触步骤之后,所述接触导致细胞群的至少40%中的A至G核碱基改变。

37.根据权利要求35所述的用途,其中在所述接触步骤之后,所述接触导致细胞群的至

少50％中的A至G核碱基改变。

38.根据权利要求35所述的用途,其中在所述接触步骤之后,所述接触导致细胞群的至少70％中的A至G核碱基改变。

39.如权利要求35所述的用途,其中至少90％的细胞在接触步骤之后是有活力的。

40.根据权利要求35所述的用途,其中所述细胞群在所述接触步骤之后未被富集。

41.根据权利要求35所述的用途,其中所述细胞群是神经元。

42.根据权利要求30所述的用途,其中所述接触是体内或离体的。

43.根据权利要求1、5、9、13、17、20、23和26中任一项所述的用途,其中所述可编程的DNA结合结构域是Cas9。

44.根据权利要求43所述的用途,其中所述Cas9是SpCas9、SaCas9或其变体。

45.根据权利要求43所述的用途,其中所述可编程的DNA结合结构域包括具有改变的前间隔邻近基序(PAM)特异性的修饰SpCas9。

46.根据权利要求45所述的用途,其中所述Cas9对选自以下任一PAM序列具有特异性:NGG、NGA、NGCG、NGN、NNGRRT、NNNRRT、NGCN、NGTN和NGC;其中N是A、G、C或T;以及其中R是A或G。

47.根据权利要求43所述的用途,其中所述可编程的DNA结合结构域是核酸酶失活变体。

48.根据权利要求43所述的用途,其中所述可编程的DNA结合结构域是切口酶变体。

49.根据权利要求1、5、9、13、17、20、23和26中任一项所述的用途,其中所述腺苷碱基编辑器包括腺苷脱氨酶单体。

50.根据权利要求1、5、9、13、17、20、23和26中任一项所述的用途,其中所述腺苷碱基编辑器包括腺苷脱氨酶二聚体。

51.一种碱基编辑器系统,其包括(i)腺苷碱基编辑器或编码所述腺苷碱基编辑器的核酸序列和(ii)向导多核苷酸,其中所述腺苷碱基编辑器包括可编程的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相较于SEQ ID NO:20具有选自以下任一氨基酸取代:

Y147T;

Y147R;

Q154S;

V82S;

T166R;

Y123H、Y147R和Q154R;

I76Y、Y147R和Q154R;

Y147R、Q154R和T166R;

Y147T和Q154R;

Y147T和Q154S;

I76Y、Y123H、Y147R和Q154R;或

V82S和Q154R,

以及其中所述向导多核苷酸引导所述腺苷碱基编辑器在LRRK2基因中实现A至G核碱基

改变,其中所述LRRK2基因具有与帕金森氏病相关的SNP,所述SNP为相较于SEQ ID NO:3发生R1441C或G2019S突变。

52.根据权利要求51所述的碱基编辑器系统,其中所述向导多核苷酸具有间隔,所述间隔具有核酸序列,所述核酸序列与具有与帕金森氏病相关的SNP的LRRK2基因互补。

53.根据权利要求51所述的碱基编辑器系统,其中所述腺苷碱基编辑器与sgRNA形成复合物,所述sgRNA具有间隔,所述间隔具有核酸序列,所述核酸序列与具有与帕金森氏病相关的SNP的LRRK2基因互补,其中所述间隔具有核酸序列:5'-AAGCGCAAGCCUGGAGGGAA-3';或5'-ACUACAGCAUUGCUCAGUAC-3'。

54.一种碱基编辑器系统,其包括(i)腺苷碱基编辑器或编码所述腺苷碱基编辑器的核酸序列和(ii)向导多核苷酸,其中所述腺苷碱基编辑器包括可编程的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相较于SEQ ID NO:20具有选自以下任一氨基酸取代:

Y147T;

Y147R;

Q154S;

T166R;

Y123H、Y147R和Q154R;

I76Y、Y147R和Q154R;

Y147R、Q154R和T166R;

Y147T和Q154R;

Y147T和Q154S;或

I76Y、Y123H、Y147R和Q154R,

并且其中所述向导多核苷酸引导所述腺苷碱基编辑器在α-L-艾杜糖醛酸酶(IDUA)基因中实现A至G核碱基改变,其中所述IDUA基因具有与贺勒氏症相关的SNP,所述SNP为相较于SEQ ID NO:4发生W402X或W401X突变,其中X是终止密码子,其中所述A至G核碱基改变在与贺勒氏症相关的SNP处,且其中所述与贺勒氏症相关的SNP处的A至G改变将所述IDUA基因所编码的IDUA多肽中的终止密码子改变为色氨酸。

55.根据权利要求54所述的碱基编辑器系统,其中所述向导多核苷酸具有间隔,所述间隔具有核酸序列,所述核酸序列与具有与贺勒氏症相关的SNP的IDUA基因互补。

56.根据权利要求54所述的碱基编辑器系统,其中所述腺苷碱基编辑器与sgRNA形成复合物,所述sgRNA具有间隔,所述间隔具有核酸序列,所述核酸序列与具有与贺勒氏症相关的SNP的IDUA基因互补,其中所述间隔具有选自以下任一核酸序列:5'-GACUCUAGGCAGAGGUCUCAA-3'、5'-ACUCUAGGCAGAGGUCUCAA-3'、5'-CUCUAGGCCGAAGUGUCGC-3'或5'-GCUCUAGGCCGAAGUGUCGC-3'。

57.一种碱基编辑器系统,其包括(i)腺苷碱基编辑器或编码所述腺苷碱基编辑器的核酸序列和(ii)向导多核苷酸,其中所述腺苷碱基编辑器包括可编程的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相较于SEQ ID NO:20具有选自以下任一氨基酸取代:

Y123H、Y147R和Q154R;

I76Y、Y147R和Q154R;或

I76Y、Y123H、Y147R和Q154R,

以及其中所述向导多核苷酸引导所述腺苷碱基编辑器在甲基CpG结合蛋白2(MECP2)基因中实现A至G核碱基改变,其中所述MECP2基因具有与蕾特氏症相关的SNP,所述SNP为相较于SEQ ID NO:5发生R106W或R255X突变,其中X是终止密码子,其中所述A至G核碱基改变在所述与蕾特氏症相关的SNP处,且其中所述A至G核碱基改变将所述与蕾特氏症相关的SNP改变为野生型核碱基。

58.根据权利要求57所述的碱基编辑器系统,其中所述向导多核苷酸具有间隔,所述间隔具有核酸序列,所述核酸序列与具有与所述蕾特氏症相关的SNP的MECP2基因互补。

59.根据权利要求57所述的碱基编辑器系统,其中所述腺苷碱基编辑器与sgRNA形成复合物,所述sgRNA具有间隔,所述间隔具有核酸序列,所述核酸序列与具有与所述蕾特氏症相关的SNP的MECP2基因互补,其中所述间隔具有选自由以下任一核酸序列:5'-CUUUUCACUUCCUGCCGGGG-3'、5'-AGCUUCCAUGUCCAGCCUUC-3'、5'-ACCAUGAAGUCAAAAUCAUU-3'或5'-GCUUUCAGCCCCGUUUCUUG-3'。

60.一种碱基编辑器系统,其包括(i)腺苷碱基编辑器或编码所述腺苷碱基编辑器的核酸序列和(ii)向导多核苷酸,其中所述腺苷碱基编辑器包括可编程的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相较于SEQ ID NO:20具有V82S的氨基酸取代,

以及其中所述向导多核苷酸引导所述腺苷碱基编辑器在ATP结合盒亚家族成员4(ABCA4)基因中实现A至G核碱基改变,其中所述ABCA4基因具有与斯塔加特氏病相关的SNP,所述SNP为相较于SEQ IN NO:6发生G1961E突变,其中所述A至G核碱基改变在与所述斯塔加特氏病相关的SNP处,且其中所述A至G核碱基改变将所述与斯塔加特氏病相关的SNP改变为野生型核碱基。

61.根据权利要求60所述的碱基编辑器系统,其中所述向导多核苷酸具有间隔,所述间隔具有核酸序列,所述核酸序列与具有与斯塔加特氏病相关的SNP的ABCA4基因互补。

62.根据权利要求60所述的碱基编辑器系统,其中所述腺苷碱基编辑器与sgRNA形成复合物,所述sgRNA具有间隔,所述间隔具有核酸序列,所述核酸序列与具有与斯塔加特氏病相关的SNP的ABCA4基因互补,其中所述间隔具有序列5'-CUCCAGGGCGAACUUCGACACACAGC-3'。

63.根据权利要求51、54、57和60中任一项所述的碱基编辑器系统,其中所述可编程的DNA结合结构域是Cas9。

64.根据权利要求63所述的碱基编辑器系统,其中所述Cas9是SpCas9、SaCas9或其变体。

65.根据权利要求63所述的碱基编辑器系统,其中所述可编程的DNA结合结构域包括具有改变的前间隔邻近基序(PAM)特异性的修饰SpCas9。

66.根据权利要求65所述的碱基编辑器系统,其中所述Cas9对选自以下任一PAM序列具有特异性:NGG、NGA、NGCG、NGN、NNGRRT、NNNRRT、NGCN、NGTN或NGC,其中N是A、G、C或T,以及其中R是A或G。

67.根据权利要求63所述的碱基编辑器系统,其中所述可编程的DNA结合结构域是核酸

酶失活变体。

68.根据权利要求63所述碱基编辑器系统,其中所述可编程的DNA结合结构域是切口酶变体。

69.根据权利要求51、54、57和60中任一项所述的碱基编辑器系统,其中所述腺苷碱基编辑器包括腺苷脱氨酶单体。

70.根据权利要求51、54、57和60中任一项所述的碱基编辑器系统,其中所述腺苷碱基编辑器包括腺苷脱氨酶二聚体。

71.一种载体,其包括编码权利要求51至70中任一项所述的腺苷碱基编辑器的核酸序列。

72.一种载体,其包括编码根据权利要求51至70中任一项所述的腺苷碱基编辑器的核酸序列和根据权利要求51至70中任一项所述的碱基编辑器系统的所述向导多核苷酸。

73.根据权利要求71或72所述的载体,其中所述载体是病毒载体。

74.根据权利要求73所述的载体,其中所述病毒载体是慢病毒载体或AAV载体。

75.一种细胞,其包括根据权利要求51至70中任一项所述的碱基编辑器系统或根据权利要求71至74中任一项所述的载体。

76.根据权利要求75所述的细胞,其中所述细胞是中枢神经系统细胞。

77.根据权利要求75所述的细胞,其中所述细胞是神经元。

78.根据权利要求75所述的细胞,其中所述细胞是光感受器。

79.根据权利要求75至78中任一项所述的细胞,其中所述细胞是体外或体内的。

80.一种药物组合物,其包括根据权利要求51至70任一项所述的碱基编辑器系统的碱基编辑器、根据权利要求71至74任一项所述的载体或根据权利要求75至79任一项所述的细胞及药学上可接受的载体。

81.根据权利要求80的药物组合物,还包括脂质。

82.根据权利要求80的药物组合物,还包括病毒。

83.一种试剂盒,其包括根据权利要求51至70任一项所述的碱基编辑器系统的碱基编辑器或根据权利要求71至74任一项所述的载体。

使用腺苷脱氨酶碱基编辑器编辑疾病相关基因的方法,包括遗传性疾病的治疗

[0001] 相关申请的交叉引用

[0002] 本申请请求保护于2019年2月13日提交的美国临时申请第62/805,271号;2019年5月23日提交的美国临时申请第62/852,228号;2019年5月23日提交的美国临时申请第62/852,224号;2019年7月11日提交的美国临时申请第62/873,138号;2019年8月19日提交的美国临时申请第62/888,867号;2019年11月6日提交的美国临时申请第62/931,722号;2019年11月27日提交的美国临时申请第62/941,569号;2020年1月27日提交的美国临时申请第62/966,526号的权益,其公开内容特此通过引用方式整体并入本文。

[0003] 通过引用方式并入

[0004] 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,其程度就如同每个单独的出版物、专利或专利申请被具体地和单独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方式整体并入本文。

背景技术

[0005] 核酸序列的靶向编辑,例如基因组DNA的靶向切割或靶向修饰,是基因功能研究的一种很有前景的方法,也有为人类遗传性疾病提供新的治法的可能。当前可用的碱基编辑器包含将目标C·G碱基对转换为T·A的胞苷碱基编辑器(例如,BE4)和将A·T转换为G·C的腺嘌呤碱基编辑器(例如,ABE7.10)。本领域需要能够以更高的特异性和效率在目标序列内诱导修饰的改进的碱基编辑器。

发明内容

[0006] 本发明提供包括具有提高的效率的新颖腺嘌呤碱基编辑器(例如ABE8)的组合物和使用包括腺苷脱氨酶变体的碱基编辑器来编辑目标序列的方法。

[0007] 在一些方面,本文提供一种治疗受试者的神经系统失调的方法,所述方法包括:向受试者给药(i)腺苷碱基编辑器或编码所述腺苷碱基编辑器的核酸序列和(ii)向导多核苷酸或编码所述向导多核苷酸的核酸序列,其中所述腺苷碱基编辑器包括可编程的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包括在SEQ ID NO:2中编号为氨基酸位置82或166或其相应位置的氨基酸取代,并且其中所述向导多核苷酸向导腺苷碱基编辑器在与受试者的神经系统失调相关的目标基因或其调控组件中实现A至G核碱基改变,从而治疗受试者的神经系统失调。在所述方面的另一个实施方案中,目标基因是α-L-艾杜糖醛酸酶(IDUA)基因并且神经系统疾病是贺勒氏症(Hurler syndrome)。在所述方面的一个实施方案中,目标基因是富含亮氨酸的重复激酶-2(leucine-rich repeat kinase-2,LRRK2)基因,并且神经系统疾病是帕金森氏病。在所述方面的一个实施方案中,目标基因是甲基CpG结合蛋白2(MECP2)基因并且神经系统疾病是蕾特氏症(Rett syndrome)。在所述方面的另一个实施方案中,目标基因是ATP结合盒亚家族成员4(ABCA4)基因,并且神经疾病是

斯塔加特氏病(Stargardt disease)。

[0008] 在一些方面,本文提供一种治疗受试者的贺勒氏症的方法,所述方法包括向受试者给药(i)腺苷碱基编辑器或编码所述腺苷碱基编辑器的核酸序列和(ii)向导多核苷酸或编码所述向导多核苷酸的核酸序列,其中所述腺苷碱基编辑器包括可编程的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包括在SEQ ID NO:2中编号为氨基酸位置82或166或其相应位置的氨基酸取代,并且其中所述向导多核苷酸向导腺苷碱基编辑器在受试者中实现α-L-艾杜糖醛酸酶(IDUA)基因或其调控组件中的A至G核碱基改变,从而治疗受试者的贺勒氏症(Hurler syndrome)。

[0009] 在一些实施方案中,给药改善与贺勒氏症相关的至少一种症状。在一些实施方案中,相较于以所述腺苷脱氨酶中没有所述氨基酸取代的碱基编辑器的治疗,给药导致与贺勒氏症相关的至少一种症状的更快改善。

[0010] 在一些实施方案中,IDUA基因或其调控组件包括与贺勒氏症相关的SNP。在一些实施方案中,A至G核碱基改变在与贺勒氏症相关的SNP处。在一些实施方案中,与贺勒氏症相关的SNP导致IDUA基因编码的SEQ ID NO:4中编号为IDUA多肽或其变体中的W402X或W401X氨基酸突变,其中X是终止密码子。在一些实施方案中,A至G核碱基改变将与贺勒氏症相关的SNP改变为野生型核碱基。在一些实施方案中,A至G核碱基改变将与贺勒氏症相关的SNP改变为非野生型核碱基,导致贺勒氏症的一种或多种改善的症状。在一些实施方案中,与贺勒氏症相关的SNP处的A至G改变将IDUA基因编码的IDUA多肽中的终止密码子改变为色氨酸。

[0011] 在一些实施方案中,向导多核苷酸包括与包括与贺勒氏症相关的SNP的IDUA基因或其调控组件互补的核酸序列。在一些实施方案中,腺苷碱基编辑器与单个向导RNA(sgRNA)形成复合物,所述单个向导RNA(sgRNA)包括与包括与贺勒氏症相关的SNP的IDUA基因或其调控组件互补的核酸序列。在一些实施方案中,sgRNA包括选自由以下所组成的群组的核酸序列:5'-GACUCUAGGCAGAGGUCUCAA-3'、5'-ACUCUAGGCAGAGGUCUCAA-3'、5'-CUCUAGGCCGAAGUGUCGC-3'及5'-GCUCUAGGCCGAAGUGUCGC-3'。

[0012] 在一些方面,本文提供一种治疗受试者的帕金森氏病的方法,所述方法包括:向受试者给药(i)腺苷碱基编辑器或编码所述腺苷碱基编辑器的核酸序列和(ii)向导多核苷酸或编码所述向导多核苷酸的核酸序列,其中所述腺苷碱基编辑器包括可编程的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包括在SEQ ID NO:2中编号的氨基酸位置82或166或其相应位置的氨基酸取代,并且其中所述向导多核苷酸向导腺苷碱基编辑器在受试者中实现富含亮氨酸的重复激酶-2(LRRK2)基因或其调控组件的A至G核碱基改变,从而治疗受试者的帕金森氏病。

[0013] 在一些实施方案中,给药改善与帕金森氏病相关的至少一种症状。在一些实施方案中,相较于以所述腺苷脱氨酶中没有所述氨基酸取代的碱基编辑器的治疗,给药导致与帕金森氏病相关的至少一种症状的更快改善。

[0014] 在一些实施方案中,LRRK2基因或其调控组件包括与帕金森氏病相关的SNP。在一些实施方案中,A至G核碱基改变在与帕金森氏病相关的SNP处。在一些实施方案中,与帕金森氏病相关的SNP导致在由LRRK2基因编码的SEQ ID NO:3中编号的LRRK2多肽或其变体中的A419V、R1441C、R1441H或G2019S氨基酸突变。

[0015] 在一些实施方案中,A至G核碱基改变将与帕金森氏病相关的SNP改变为野生型核碱基。在一些实施方案中,A至G核碱基改变将与帕金森氏病相关的SNP改变为非野生型核碱基,从而导致帕金森氏病的一种或多种改善的症状。在一些实施方案中,A至G核碱基改变将LRRK2基因编码的LRRK2多肽中的半胱氨酸或组氨酸改变为精氨酸。在一些实施方案中,A至G改变将LRRK2基因编码的LRRK2多肽中的丝氨酸改变为甘氨酸。在一些实施方案中,A至G改变在由LRRK2基因编码的SEQ ID NO:3中编号的LRRK2多肽或其变体的位置144用精氨酸(R)置换半胱氨酸(C)或组氨酸(H)或在位置144用用甘氨酸(G)置换丝氨酸。

[0016] 在一些方面,本文提供一种治疗受试者的帕金森氏病的方法,所述方法包括:向受试者给药(i)腺苷碱基编辑器或编码所述腺苷碱基编辑器的核酸序列和(ii)向导多核苷酸或编码所述向导多核苷酸的核酸序列,其中所述腺苷碱基编辑器包括可编程的DNA结合结构域和腺苷脱氨酶结构域,并且其中所述向导多核苷酸向导腺苷碱基编辑器在与帕金森氏病相关的LRRK2基因中在SNP处实现A至G核碱基改变,其中所述SNP不编码在SEQ ID NO:3中编号的LRRK2多肽或其变体中的G2019S突变。

[0017] 在一些实施方案中,腺苷脱氨酶结构域包括在SEQ ID NO:2中编号的氨基酸位置82或166处或其相应位置的氨基酸取代。在一些实施方案中,向导多核苷酸包括与包括与帕金森氏病相关的SNP的LRRK2基因或其调控组件互补的核酸序列。在一些实施方案中,腺苷碱基编辑器与单个向导RNA(sgRNA)形成复合物,所述单个向导RNA(sgRNA)包括与包括与帕金森氏病相关的SNP的LRRK2基因或其调控组件互补的核酸序列。在一些实施方案中,sgRNA包括核酸序列:5'-AAGCGCAAGCCUGGAGGGAA-3';或5'-ACUACAGCAUUGCUCAGUAC-3'。

[0018] 在一些方面,本文提供一种治疗受试者的蕾特氏症的方法,所述方法包括向受试者给药(i)腺苷碱基编辑器或编码所述腺苷碱基编辑器的核酸序列和(ii)向导多核苷酸或编码所述向导多核苷酸的核酸序列,其中所述腺苷碱基编辑器包括可编程的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包括在SEQ ID NO:2中编号的氨基酸位置82或166或其相应位置的氨基酸取代:,并且其中所述向导多核苷酸向导腺苷碱基编辑器在受试者中实现甲基CpG结合蛋白2(MECP2)基因或其调控组件中的A至G核碱基改变,从而治疗受试者的蕾特氏症(Rett syndrome)。

[0019] 在一些实施方案中,给药改善与蕾特氏症相关的至少一种症状。在一些实施方案中,相较于以所述腺苷脱氨酶中没有所述氨基酸取代的碱基编辑器的治疗,给药导致与蕾特氏症相关的至少一种症状的更快改善。在一些实施方案中,MECP2基因或其调控组件包括与蕾特氏症相关的SNP。在一些实施方案中,A至G核碱基改变发生在与蕾特氏症相关的SNP处。在一些实施方案中,与蕾特氏症相关的SNP导致所述MECP2基因所编码的在SEQ ID NO:5中编号的MECP2多肽或其变体中的R106W或T158M氨基酸突变。在一些实施方案中,与蕾特氏症相关的SNP导致MECP2基因编码的MECP2多肽中的R255X或R270X氨基酸突变,其中X是终止密码子。

[0020] 在一些实施方案中,A至G核碱基改变将与蕾特氏症相关的SNP改变为野生型核碱基。在一些实施方案中,A至G核碱基改变将与蕾特氏症相关的SNP改变为非野生型核碱基,导致蕾特氏症症状改善。在一些实施方案中,与蕾特氏症相关的SNP处的A至G核碱基改变将MECP2多肽中的终止密码子改变为色氨酸。

[0021] 在一些实施方案中,向导多核苷酸包括与包括与蕾特氏症相关的SNP的MECP2基因

或其调控组件互补的核酸序列蕾特氏症。在一些实施方案中,腺苷碱基编辑器与单个向导RNA(sgRNA)形成复合物,其包括与包括与蕾特氏症相关的SNP的MECP2基因或其调控组件互补的核酸序列。在一些实施方案中,向导多核苷酸包括选自由以下所组成的群组的核酸序列:5'-CUUUUCACUUCCUGCCGGGG-3',5'-AGCUUCCAUGUCCAGCCUUC-3'、5'-ACCAUGAAGUCAAAAUCAUU-3'及5'-GCUUUCAGCCCCGUUUCUUG-3'。

[0022] 在一些方面,本文提供一种治疗受试者的斯塔加特氏病的方法,所述方法包括向受试者给药(i)腺苷碱基编辑器或编码所述腺苷碱基编辑器的核酸序列和(ii)向导多核苷酸或编码所述向导多核苷酸的核酸序列,其中所述腺苷碱基编辑器包括可编程的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包括在SEQ ID NO:2中编号的氨基酸位置82或166或其相应位置的氨基酸取代,并且其中所述向导多核苷酸向导腺苷碱基编辑器在受试者的ATP结合盒亚家族成员4(ABCA4)基因或其调控组件中实现A至G核碱基改变,从而治疗受试者的斯塔加特氏病。

[0023] 在一些实施方案中,给药改善与斯塔加特氏病相关的至少一种症状。在一些实施方案中,相较于以所述腺苷脱氨酶中没有所述氨基酸取代的碱基编辑器的治疗,给药导致与斯塔加特氏病相关的至少一种症状的更快改善。

[0024] 在一些实施方案中,ABCA4基因包括与斯塔加特氏病相关的SNP。在一些实施方案中,A至G核碱基改变在与斯塔加特氏病相关的SNP处。在一些实施方案中,与斯塔加特氏病相关的SNP导致由ABCA4所述基因所编码的在SEQ ID NO:6中编号的ABCA4多肽或其变体中的A1038V或G1961E氨基酸突变。在一些实施方案中,与斯塔加特氏病相关的SNP导致在SEQ ID NO:6中编号的ABCA4多肽或其变体中的G1961E氨基酸突变。

[0025] 在一些实施方案中,A至G核碱基改变将与斯塔加特氏病相关的SNP改变为野生型核碱基。在一些实施方案中,A至G核碱基改变将与斯塔加特氏病相关的SNP改变为非野生型核碱基,其导致斯塔加特氏病的一种或多种改善的症状。在一些实施方案中,向导多核苷酸包括与包括与斯塔加特氏病相关的SNP的ABCA4基因或其调控组件互补的核酸序列。

[0026] 在一些实施方案中,腺苷碱基编辑器与单个向导RNA(sgRNA)形成复合物,其包括与包括与斯塔加特氏病相关的SNP的ABCA4基因或其调控组件互补的核酸序列。在一些实施方案中,sgRNA包括序列5'-CUCCAGGGCGAACUUCGACACACAGC-3'。

[0027] 在各个方面,与使用包括没有氨基酸取代的腺苷脱氨酶结构域的碱基编辑器的治疗相比,本文所述治疗导致神经系统失调的症状改善。

[0028] 在一些方面,本文提供一种编辑与神经系统失调相关的目标基因或其调控组件的方法,所述方法包括使目标基因或其调控组件与(i)腺苷碱基编辑器和(ii)向导多核苷酸,其中所述腺苷碱基编辑器包括可编程的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包括在SEQ ID NO:2中编号的氨基酸位置82或166或其相应位置的氨基酸取代,其中所述向导多核苷酸向导腺苷碱基编辑器在与神经系统失调相关的目标基因或其调控组件中实现A至G核碱基改变。在所述方面的一个实施方案中,目标基因是富含亮氨酸的重复激酶-2(LRRK2)基因,并且神经系统疾病是帕金森氏病。在所述方面的另一个实施方案中,目标基因是α-L-艾杜糖醛酸酶(IDUA)基因,并且神经系统疾病是贺勒氏症。在所述方面的一个实施方案中,目标基因是甲基CpG结合蛋白2(MECP2)基因,并且神经系统疾病是蕾特氏症。在所述方面的另一个实施方案中,目标基因是ATP结合盒亚家族成员4(ABCA4)基

因,并且神经系统疾病是斯塔加特氏病。

[0029] 在一些方面,本文提供一种编辑富含亮氨酸的重复激酶-2(LRRK2)基因或其调控组件的方法,所述方法包括将LRRK2基因或其调控组件与(i)腺苷碱基编辑器或编码所述腺苷碱基编辑器的核酸序列和(ii)向导多核苷酸或编码所述向导多核苷酸的核酸序列,其中腺苷碱基编辑器包括可编程的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包括在SEQ ID NO:2中编号的氨基酸位置82或166或其相应位置的氨基酸取代,并且其中所述向导多核苷酸向导腺苷碱基编辑器在LRRK基因其调控组件中中实现A至G核碱基改变。

[0030] 在一些实施方案中,A至G核碱基改变在与帕金森氏病相关的SNP处。在一些实施方案中,与帕金森氏病相关的SNP导致在由LRRK2基因编码的SEQ ID NO:3中编号的LRRK2多肽或其变体中的A419V、R1441C、R1441H或G2019S氨基酸突变。在一些实施方案中,A至G核碱基改变将与帕金森氏病相关的SNP改变为野生型核碱基。在一些实施方案中,A至G核碱基改变将与帕金森氏病相关的SNP改变为非野生型核碱基,导致帕金森氏病的一种或多种改善的症状。

[0031] 在一些实施方案中,A至G核碱基改变将由LRRK2基因编码的LRRK2多肽中的半胱氨酸或组氨酸改变为精氨酸。在一些实施方案中,A至G的改变将由LRRK2基因编码的LRRK2多肽中的丝氨酸改变为甘氨酸。在一些实施方案中,A至G改变在由LRRK2基因编码的SEQ ID NO:3中编号为LRRK2多肽或其变体的位置144用精氨酸(R)置换半胱氨酸(C)或组氨酸(H)或在位置2019用甘氨酸(G)替换丝氨酸。

[0032] 在一些方面,本文提供一种编辑富含亮氨酸的重复激酶-2(LRRK2)基因或其调控组件的方法,所述方法包括将LRRK基因或其调控组件与(i)腺苷碱基编辑器或编码所述腺苷碱基编辑器的核酸序列和(ii)向导多核苷酸或编码所述向导多核苷酸的核酸序列,其中所述腺苷碱基编辑器包括可编程的DNA结合结构域和腺苷脱氨酶结构域,并且其中所述向导多核苷酸向导腺苷碱基编辑器在LRRK2基因中的SNP处实现A至G核碱基改变,其中所述SNP不编码在SEQ ID NO:3中编号的LRRK2多肽或其变体中的G2019S突变。

[0033] 在一些实施方案中,腺苷脱氨酶结构域包括在SEQ ID NO:2中编号的氨基酸位置82或166或其相应位置氨基酸取代。在一些实施方案中,向导多核苷酸包括与包括与帕金森氏病相关的SNP的LRRK2基因或其调控组件互补的核酸序列。

[0034] 在一些实施方案中,腺苷碱基编辑器与单个向导RNA(sgRNA)形成复合物,其包括与包括与帕金森氏病相关的SNP的LRRK2基因或其调控组件互补的核酸序列。在一些实施方案中,sgRNA包括核酸序列:5'-AAGCGCAAGCCUGGAGGGAA-3';或5'-ACUACAGCAUUGCUCAGUAC-3'。

[0035] 在一些方面,本文提供一种编辑α-L-艾杜糖醛酸酶(IDUA)基因或其调控组件的方法,所述方法包括将IDUA基因或其调控组件与(i)腺苷碱基编辑器或编码所述腺苷碱基编辑器的核酸序列和(ii)向导多核苷酸或编码所述向导多核苷酸的核酸序列,其中所述腺苷碱基编辑器包括可编程的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包括氨基酸在SEQ ID NO:2中编号为氨基酸位置82或166或其相应位置的氨基酸取代,并且其中所述向导多核苷酸向导腺苷碱基编辑器在IDUA基因或其调节组件中实现A至G核碱基改变。

[0036] 在一些实施方案中,IDUA基因或其调控组件包括与贺勒氏症相关的SNP。在一些实施方案中,A至G核碱基改变在与贺勒氏症相关的SNP处。在一些实施方案中,与贺勒氏症相关的SNP导致由IDUA基因编码的在SEQ ID NO:4中编号为IDUA多肽或其变体中的W402X或W401X氨基酸突变,其中X是终止密码子。

[0037] 在一些实施方案中,A至G核碱基的改变将与贺勒氏症相关的SNP改变为野生型核碱基。在一些实施方案中,A至G核碱基改变将与贺勒氏症相关的SNP改变为非野生型核碱基,导致贺勒氏症的一种或多种改善的症状。在一些实施方案中,与贺勒氏症相关的SNP处的A至G改变将由IDUA基因编码的IDUA多肽中的终止密码子改变为色氨酸。

[0038] 在一些实施方案中,向导多核苷酸包括与包括与贺勒氏症相关的SNP的IDUA基因或其调控组件互补的核酸序列。在一些实施方案中,腺苷碱基编辑器与单个向导RNA(sgRNA)形成复合物,所述向导RNA(sgRNA)包括与包括与贺勒氏症相关的SNP的IDUA基因或其调控组件互补的核酸序列。在一些实施方案中,sgRNA包括选自由以下所组成的群组的核酸序列:5'-GACUCUAGGCAGAGGUCUCAA-3'、5'-ACUCUAGGCAGAGGUCUCAA-3'、5'-CUCUAGGCCGAAGUGUCGC-3'及5'-GCUCUAGGCCGAAGUGUCGC-3'。

[0039] 在一些方面,本文提供一种编辑甲基CpG结合蛋白2(MECP2)基因或其调控组件的方法,所述方法包括向受试者给药(i)腺苷碱基编辑器或编码所述腺苷碱基编辑器的核酸序列和(ii)向导多核苷酸或编码所述向导多核苷酸的核酸序列,其中所述腺苷碱基编辑器包括可编程的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包括在SEQ ID NO:2中编号为氨基酸位置82或166或其相应位置的氨基酸取代,并且其中所述向导多核苷酸向导腺苷碱基编辑器在MECP2基因或其调控组件中实现A至G核碱基改变。

[0040] 在一些实施方案中,MECP2基因或其调控组件包括与蕾特氏症相关的SNP。在一些实施方案中,A至G核碱基改变发生在与蕾特氏症相关的SNP处。在一些实施方案中,与蕾特氏症相关的SNP导致在由MECP2基因编码的SEQ ID NO:5中编号为MECP2多肽或其变体中的R106W或T158M氨基酸突变。在一些实施方案中,与蕾特氏症相关的SNP导致由MECP2基因编码的MECP2多肽中的R255X或R270X氨基酸突变,其中X是终止密码子。

[0041] 在一些实施方案中,A至G核碱基改变将与蕾特氏症相关的SNP改变为野生型核碱基。在一些实施方案中,A至G核碱基改变将与蕾特氏症相关的SNP改变为非野生型核碱基,导致蕾特氏症的一种或多种改善的症状。在一些实施方案中,与蕾特氏症相关的SNP处的A至G核碱基改变将MECP2多肽中的终止密码子改变为色氨酸。

[0042] 在一些实施方案中,向导多核苷酸包括与包括与蕾特氏症相关的SNP的MECP2基因或其调控组件互补的核酸序列。在一些实施方案中,腺苷碱基编辑器与单个向导RNA(sgRNA)形成复合物,其包括与包括与蕾特氏症相关的SNP的MECP2基因或其调控组件互补的核酸序列。在一些实施方案中,向导多核苷酸包括选自由以下所组成的群组的核酸序列:5'-CUUUUCACUUCCUGCCGGGG-3'、5'-AGCUUCCAUGUCCAGCCUUC-3',5'-ACCAUGAAGUCAAAAUCAUU-3'及5'-GCUUUCAGCCCCGUUUCUUG-3'。

[0043] 在一些方面,本文提供一种编辑ATP结合盒亚家族成员4(ABCA4)基因或其调控组件的方法,所述方法包括使ABCA4基因或其调控组件与(i)腺苷碱基编辑器或核酸编码所述腺苷碱基编辑器的核酸序列和(ii)向导多核苷酸或编码所述向导多核苷酸的核酸序列,其中所述腺苷碱基编辑器包括可编程的DNA结合结构域和腺苷脱氨酶结构域,其中腺苷脱氨

酶结构域包括氨基酸在SEQ ID NO:2中编号的氨基酸位置82或166或其相应位置的取代,并且其中所述向导多核苷酸向导腺苷碱基编辑器在ABCA4基因或调控组件中实现A至G核碱基改变。

[0044] 在一些实施方案中,给药改善与斯塔加特氏病相关的至少一种症状。在一些实施方案中,相较于以所述腺苷脱氨酶中没有所述氨基酸取代的碱基编辑器的治疗,给药导致与斯塔加特氏病相关的至少一种症状的更快改善。

[0045] 在一些实施方案中,ABCA4基因包括与斯塔加特氏病相关的SNP。在一些实施方案中,A至G核碱基改变在与斯塔加特氏病相关的SNP处。在一些实施方案中,与斯塔加特氏病相关的SNP导致在由ABCA4基因编码的SEQ ID NO:6中编号为ABCA4多肽或其变体中的A1038V或G1961E氨基酸突变。在一些实施方案中,与斯塔加特氏病相关的SNP导致在SEQ ID NO:6中编号的ABCA4多肽或其变体中的G1961E氨基酸突变。

[0046] 在一些实施方案中,A至G核碱基改变将与斯塔加特氏病相关的SNP改变为野生型核碱基。在一些实施方案中,A至G核碱基改变将与斯塔加特氏病相关的SNP改变为非野生型核碱基,导致斯塔加特氏病的一种或多种改善的症状。在一些实施方案中,向导多核苷酸包括与包括与斯塔加特氏病相关的SNP的ABCA4基因或其调控组件互补的核酸序列。

[0047] 在一些实施方案中,腺苷碱基编辑器与单个向导RNA(sgRNA)形成复合物,其包括与包括与斯塔加特氏病相关的SNP的ABCA4基因或其调控组件互补的核酸序列。在一些实施方案中,sgRNA包括序列5'-CUCCAGGGCGAACUUCGACACACAGC-3'。

[0048] 在上述方面的各种实施方案中,接触是在细胞中。在一些实施方案中,接触导致细胞中基因组中小于10%的插入缺失,其中插入缺失率通过在单核苷酸修饰侧边的序列与未修饰序列之间的错配频率来测量。在一些实施方案中,接触导致细胞中基因组中小于5%的插入缺失,其中插入缺失率通过在单核苷酸修饰侧边的序列与未修饰序列之间的错配频率来测量。在一些实施方案中,接触导致细胞中基因组中小于1%的插入缺失,其中插入缺失率通过在单核苷酸修饰侧边的序列与未修饰序列之间的错配频率来测量。

[0049] 在上述方面的各种实施方案中,细胞是神经元。在一些实施方案中,接触是在细胞群中。在一些实施方案中,接触导致接触步骤后至少40%的细胞群中的A至G核碱基改变。在一些实施方案中,接触导致接触步骤后至少50%的细胞群中的A至G核碱基改变。在一些实施方案中,接触导致接触步骤后至少70%的细胞群中的A至G核碱基改变。在一些实施方案中,至少90%的细胞在接触步骤之后是存活的。在一些实施方案中,细胞群在接触步骤后没有富集。在一些实施方案中,细胞群是神经元。在一些实施方案中,接触是体内或离体的。

[0050] 在上述各个方面和实施方案中,多核苷酸可编程的DNA结合结构域是Cas9。在一些实施方案中,Cas9是SpCas9、SaCas9或其变体。在一些实施方案中,多核苷酸可编程的DNA结合结构域包括具有改变的前间隔邻近基序(protospacer-adjacent motif,PAM)特异性的修饰的SpCas9。在一些实施方案中,Cas9对选自由NGG、NGA、NGCG、NGN、NNGRRT、NNNRRT、NGCG、NGCN、NGTN及NGC所组成的群组的PAM序列具有特异性;其中N是A、G、C或T;及其中R是A或G。在一些实施方案中,多核苷酸可编程的DNA结合结构域是核酸酶失活变体。在一些实施方案中,多核苷酸可编程的DNA结合结构域是切口酶变体。在一些实施方案中,切口酶变体包括氨基酸取代D10A或其相应的氨基酸取代。在本文所提供的各个方面和实施方案中,腺苷脱氨酶结构域包括TadA结构域。在一些实施方案中,腺苷脱氨酶包括TadA脱氨酶,其包括

V82S改变和/或T166R改变。

[0051] 在上述各个方面和实施方案中,腺苷脱氨酶还包括一种或多种以下改变:Y147T、Y147R、Q154S、Y123H、Q154R或其组合。在本文所提供的各个方面和实施方案中,腺苷脱氨酶包括选自由以下组成的群组的改变的组合:Y147R+Q154R+Y123H;Y147R+Q154R+I76Y;Y147R+Q154R+T166R;Y147T+Q154R;Y147T+Q154S;及Y123H+Y147R+Q154R+I76Y。在本文所提供的各个方面和实施方案中,腺苷碱基编辑器结构域包括腺苷脱氨酶单体。在本文所提供的各个方面和实施方案中,腺苷碱基编辑器包括腺苷脱氨酶二聚体。在一些实施方案中,TadA脱氨酶是TadA*8变体。在一些实施方案中,TadA*8变体是选自: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及TadA*8.13。在一些实施方案中,腺苷碱基编辑器是选自由以下组成的群组的ABE8碱基编辑器:ABE8.1、ABE8.2、ABE8.3、ABE8.4、ABE8.5、ABE8.6、ABE8.7、ABE8.8、ABE8.9、ABE8.10、ABE8.11、ABE8.12及ABE8.13。

[0052] 在一些方面,本文提供通过本文公开的各个方面和实施方案中描述的方法产生的细胞。在一些方面,本文提供通过本文公开的各个方面和实施方案中描述的方法产生的细胞群。

[0053] 在一些方面,本文提供一种碱基编辑器系统,其包括(i)腺苷碱基编辑器或编码所述腺苷碱基编辑器的核酸序列和(ii)向导多核苷酸或编码所述向导多核苷酸的核酸序列,其中所述腺苷碱基编辑器包括可编程的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在SEQ ID NO:2中编号的氨基酸位置82或166或其相应位置包括氨基酸取代,并且其中所述向导多核苷酸向导腺苷碱基编辑器在与神经系统失调相关的目标基因或其调控组件中实现A至G核碱基改变。在上述方面的一个实施方案中,目标基因是富含亮氨酸的重复激酶-2(LRRK2)基因,并且神经系统疾病是帕金森氏病。在所述方面的另一个实施方案中,目标基因是α-L-艾杜糖醛酸酶(IDUA)基因,并且神经系统疾病是贺勒氏症。在所述方面的一个实施方案中,目标基因是甲基CpG结合蛋白2(MECP2)基因,并且神经系统疾病是蕾特氏症。在所述方面的另一个实施方案中,目标基因是ATP结合盒亚家族成员4(ABCA4)基因,并且神经系统疾病是斯塔加特氏病。

[0054] 在一些方面,本文提供一种碱基编辑器系统,其包括(i)腺苷碱基编辑器或编码所述腺苷碱基编辑器的核酸序列和(ii)向导多核苷酸或编码所述向导多核苷酸的核酸序列,其中腺苷碱基编辑器包括可编程的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包括在SEQ ID NO:2编号的氨基酸位置82或166或其相应位置的氨基酸取代,并且其中所述向导多核苷酸向导腺苷碱基编辑器在LRRK基因及其调控组件中实现A至G核碱基的改变。

[0055] 在一些实施方案中,A至G核碱基改变位于LRRK2基因或其调控组件中与帕金森氏病相关的SNP处。在一些实施方案中,与帕金森氏病相关的SNP导致在由LRRK2基因编码的SEQ ID NO:3中编号的LRRK2多肽或其变体中的A419V、R1441C、R1441H或G2019S氨基酸突变。

[0056] 在一些实施方案中,A至G核碱基改变将与帕金森氏病相关的SNP改变为野生型核碱基。在一些实施方案中,A至G核碱基改变将与帕金森氏病相关的SNP改变为非野生型核碱基,导致帕金森症状的改善。在一些实施方案中,A至G核碱基改变将LRRK2基因编码的LRRK2

多肽中的半胱氨酸或组氨酸改变为精氨酸。在一些实施方案中,A至G的改变将LRRK2基因编码的LRRK2多肽中的丝氨酸改变为甘氨酸。在一些实施方案中,A至G改变在由LRRK2基因编码的SEQ ID NO:3中编号为LRRK2多肽或其变体的位置144用精氨酸(R)置换半胱氨酸(C)或组氨酸(H)或在位置2019用用甘氨酸(G)置换丝氨酸。在一些实施方案中,腺苷脱氨酶结构域包括在SEQ ID NO:2中编号为氨基酸位置82或166或其相应位置的氨基酸取代。

[0057] 在一些实施方案中,向导多核苷酸包括与包括与帕金森氏病相关的SNP的LRRK2基因或其调控组件互补的核酸序列。在一些实施方案中,腺苷碱基编辑器与单个向导RNA(sgRNA)形成复合物,其包括与包括与帕金森氏病相关的SNP的LRRK2基因或其调控组件互补的核酸序列。在一些实施方案中,sgRNA包括核酸序列:5'-AAGCGCAAGCCUGGAGGGAA-3';或5'-ACUACAGCAUUGCUCAGUAC-3'。

[0058] 在一些方面,本文提供一种碱基编辑器系统,其包括(i)腺苷碱基编辑器或编码所述腺苷碱基编辑器的核酸序列和(ii)向导多核苷酸或编码所述向导多核苷酸的核酸序列,其中所述腺苷碱基编辑器包括可编程的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包括在SEQ ID NO:2中编号为氨基酸位置82或166或其相应位置的氨基酸取代,并且其中所述向导多核苷酸向导腺苷碱基编辑器在α-L-艾杜糖醛酸酶(IDUA)基因或其调控组件中实现A至G核碱基的改变。

[0059] 在一些实施方案中,IDUA基因或其调控组件包括与贺勒氏症相关的SNP。在一些实施方案中,A至G核碱基改变在与贺勒氏症相关的SNP处。在一些实施方案中,与贺勒氏症相关的SNP导致在由IDUA基因编码的SEQ ID NO:4中编号为IDUA多肽或其变体中的W402X或W401X氨基酸突变,其中X是终止密码子。

[0060] 在一些实施方案中,A至G核碱基的改变将与贺勒氏症相关的SNP改变为野生型核碱基。在一些实施方案中,A至G核碱基改变将与贺勒氏症相关的SNP改变为非野生型核碱基,导致贺勒氏症的一种或多种改善的症状。在一些实施方案中,与贺勒氏症相关的SNP处的A至G改变将由IDUA基因编码的IDUA多肽中的终止密码子改变为色氨酸。

[0061] 在一些实施方案中,向导多核苷酸包括与包括与贺勒氏症相关的SNP的IDUA基因或其调控组件互补的核酸序列。在一些实施方案中,腺苷碱基编辑器与单个向导RNA(sgRNA)形成复合物,其包括与包括与贺勒氏症相关的SNP的IDUA基因或其调控组件互补的核酸序列。在一些实施方案中,sgRNA包括选自由以下组成的群组的核酸序列:5'-GACUCUAGGCAGAGGUCUCAA-3',5'-ACUCUAGGCAGAGGUCUCAA-3'、5'-CUCUAGGCCGAAGUGUCGC-3'及5'-GCUCUAGGCCGAAGUGUCGC-3'。

[0062] 在一些方面,本文提供一种碱基编辑器系统,其包括(i)腺苷碱基编辑器或编码所述腺苷碱基编辑器的核酸序列和(ii)向导多核苷酸或编码所述向导多核苷酸的核酸序列,其中所述腺苷碱基编辑器包括可编程的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包括在SEQ ID NO:2中编号为氨基酸位置82或166或其相应位置的氨基酸取代,并且其中所述向导多核苷酸向导腺苷碱基编辑器在甲基CpG结合蛋白2(MECP2)基因或其调控组件中实现A至G核碱基的改变。

[0063] 在一些实施方案中,MECP2基因或其调控组件包括与蕾特氏症相关的SNP。在一些实施方案中,A至G核碱基改变发生在与蕾特氏症相关的SNP处。在一些实施方案中,与蕾特氏症相关的SNP导致在由MECP2基因编码的SEQ ID NO:5中编号为MECP2多肽或其变体中的

R106W或T158M氨基酸突变。在一些实施方案中,与蕾特氏症相关的SNP导致在由MECP2基因编码的MECP2多肽中的R255X或R270X氨基酸突变,其中X是终止密码子。

[0064] 在一些实施方案中,A至G核碱基改变将与蕾特氏症相关的SNP改变为野生型核碱基。在一些实施方案中,A至G核碱基改变将与蕾特氏症相关的SNP改变为非野生型核碱基,导致蕾特氏症的一种或多种改善的症状。在一些实施方案中,与蕾特氏症相关的SNP处的A至G核碱基改变将MECP2多肽中的终止密码子改变为色氨酸。

[0065] 在一些实施方案中,向导多核苷酸包括与包括与蕾特氏症相关的SNP的MECP2基因或其调控组件互补的核酸序列。在一些实施方案中,腺苷碱基编辑器与单个向导RNA(sgRNA)形成复合物,其包括与包括与蕾特氏症相关的SNP的MECP2基因或其调控组件互补的核酸序列。在一些实施方案中,向导多核苷酸包括选自由以下所组成的群组的核酸序列:5'-CUUUUCACUUCCUGCCGGGG-3'、5'-AGCUUCCAUGUCCAGCCUUC-3'、5'-ACCAUGAAGUCAAAAUCAUU-3'及5'-GCUUUCAGCCCCGUUUCUUG-3'。

[0066] 在一些方面,本文提供一种碱基编辑器系统,其包括接触(i)腺苷碱基编辑器或编码所述腺苷碱基编辑器的核酸序列和(ii)向导多核苷酸或编码所述向导多核苷酸的核酸序列,其中所述腺苷碱基编辑器包括可编程的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包括在SEQ ID NO:2中编号为氨基酸位置82或166或其相应位置的氨基酸取代,并且其中所述向导多核苷酸向导腺苷碱基编辑器在ATP结合盒亚家族成员4(ABCA4)基因或其调控组件中实现A至G核碱基的改变。

[0067] 在一些实施方案中,给药改善与斯塔加特氏病相关的至少一种症状。在一些实施方案中,相较于以所述腺苷脱氨酶中没有所述氨基酸取代的碱基编辑器的治疗,给药导致与斯塔加特氏病相关的至少一种症状的更快改善。在一些实施方案中,ABCA4基因包括与斯塔加特氏病相关的SNP。在一些实施方案中,A至G核碱基改变在与斯塔加特氏病相关的SNP处。在一些实施方案中,与斯塔加特氏病相关的SNP导致在由ABCA4基因编码的SEQ ID NO:6中编号为ABCA4多肽或其变体中的A1038V或G1961E氨基酸突变。在一些实施方案中,与斯塔加特氏病相关的SNP导致在SEQ ID NO:6中编号为ABCA4多肽或其变体中的G1961E氨基酸突变。

[0068] 在一些实施方案中,A至G核碱基改变将与斯塔加特氏病相关的SNP改变为野生型核碱基。在一些实施方案中,A至G核碱基改变将与斯塔加特氏病相关的SNP改变为非野生型核碱基,导致改善的斯塔加特氏病症状。在一些实施方案中,向导多核苷酸包括与包括与斯塔加特氏病相关的SNP的ABCA4基因或其调控组件互补的核酸序列。

[0069] 在一些实施方案中,腺苷碱基编辑器与单个向导RNA(sgRNA)形成复合物,其包括与包括与斯塔加特氏病相关的SNP的ABCA4基因或其调控组件互补的核酸序列。在一些实施方案中,sgRNA包括序列5'-CUCCAGGGCGAACUUCGACACACAGC-3'。

[0070] 在本文所提供的各个方面和实施方案中,多核苷酸可编程的DNA结合结构域是Cas9。在一些实施方案中,Cas9是SpCas9、SaCas9或其变体。在一些实施方案中,多核苷酸可编程的DNA结合结构域包括具有改变的前间隔邻近基序(PAM)特异性的修饰SpCas9。在一些实施方案中,Cas9对选自由NGG、NGA、NGCG、NGN、NNGRRT、NNNRRT、NGCG、NGCN、NGTN和NGC所组成的群组的PAM序列具有特异性,其中N是A、G、C或T,且其中R是A或G。在一些实施方案中,多核苷酸可编程的DNA结合结构域是核酸酶失活变体。在一些实施方案中,多核苷酸可编程的

DNA结合结构域是切口酶变体。在一些实施方案中,切口酶变体包括氨基酸取代D10A或其相应的氨基酸取代。

[0071] 在本文所提供的各个方面和实施方案中,腺苷脱氨酶结构域包括TadA结构域。在一些实施方案中,腺苷脱氨酶包括TadA脱氨酶,其包括V82S改变和/或T166R改变。

[0072] 在本文所提供的各个方面和实施方案中,腺苷脱氨酶进一步包括一种或多种以下改变:Y147T、Y147R、Q154S、Y123H、Q154R或其组合。在本文所提供的各个方面和实施方案中,腺苷脱氨酶包括选自由以下所组成的群组的改变的组合:Y147R+Q154R+Y123H;Y147R+Q154R+I76Y;Y147R+Q154R+T166R;Y147T+Q154R;Y147T+Q154S;及Y123H+Y147R+Q154R+I76Y。在一些实施方案中,腺苷碱基编辑器结构域包括腺苷脱氨酶单体。在一些实施方案中,腺苷碱基编辑器包括腺苷脱氨酶二聚体。

[0073] 在本文所提供的各个方面和实施方案中,TadA脱氨酶是TadA*8变体。在一些实施方案中,TadA*8变体是选自由以下所组成的群组: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及TadA*8.13。在一些实施方案中,腺苷碱基编辑器是选自由以下所组成的群组的ABE8碱基编辑器:ABE8.1、ABE8.2、ABE8.3、ABE8.4、ABE8.5、ABE8.6、ABE8.7、ABE8.8、ABE8.9、ABE8.10、ABE8.11、ABE8.12及ABE8.13。

[0074] 在一些方面,本文提供一种包括编码本文所述腺苷碱基编辑器的核酸序列的载体。在一些方面,本文提供一种包括编码本文所述腺苷碱基编辑器和向导多核苷酸的核酸序列的载体。在一些实施方案中,载体是病毒载体、慢病毒载体或AAV载体。

[0075] 在一些方面,本文提供一种包括本文所述的碱基编辑器系统或载体的细胞。在一些实施方案中,细胞是中枢神经系统细胞。在一些实施方案中,细胞是神经元。在一些实施方案中,细胞是光感受器。在一些实施方案中,细胞是体外、体内或离体的。

[0076] 在一些方面,本文提供一种药物组合物,其包括本文所述碱基编辑器、载体或细胞以及药学上可接受的载体。在一个实施方案中,本文所述药物组合物进一步包括脂质。在另一个实施方案中,本文所述药物组合物进一步包括病毒。

[0077] 在一些方面,本文提供包括本文所述碱基编辑器或载体的试剂盒。

[0078] 在本文所述方法的各种实施方案中,向导多核苷酸的至少一个核苷酸包括非天然存在的修饰。在本文所述方法的各种实施方案中,核酸序列的至少一个核苷酸包括非天然存在的修饰。在各种实施方案中,碱基编辑器系统的核酸序列的至少一个核苷酸包括非天然存在的修饰。在一些实施方案中,非天然存在的修饰是化学修饰。在一些实施方案中,化学修饰是2'-O-甲基化。在一些实施方案中,核酸序列包括硫代磷酸酯。

[0079] 此处的描述和实施例详细说明了本公开的实施方案。应当理解,本公开不限于本文所述的特定实施方案并且因此可以变化。本领域技术人员将认识到本公开内容存在多种变化和修改,而这些变化和修改包括在其范畴内。

[0080] 除非另外指明,否则本文公开的一些实施方案的实践采用免疫学、生物化学、化学、分子生物学、微生物学、细胞生物学、基因组学和重组DNA的常规技术,这些技术在本领域技术范畴内。参见例如Sambrook and Green,Molecular Cloning:A Laboratory Manual,4th Edition(2012);the series Current Protocols in Molecular Biology(F.M.Ausubel,et al.eds.);the series Methods In Enzymology(Academic Press,

Inc.)、PCR 2:A Practical Approach(M.J.MacPherson,B.D.Hames and G.R.Taylor eds.(1995))、Harlow and Lane,eds.(1988)Antibodies,A Laboratory Manual,and Culture of Animal Cells:A Manual of Basic Technique and Specialized Applications,6th Edition(R.I.Freshney,ed.(2010))。

[0081] 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述主题。

[0082] 尽管可以在单个实施方案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方案的上下文中在本文中描述本公开,本公开也可以在单个实施方案中实施。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述主题。

[0083] 本公开的特征在所附权利要求中特别阐述。通过参考以下阐述性实施方案的详细描述,其中利用了本公开的原理,并鉴于如下所述附图,将获得对本发明的特征和优点的更好的理解。

[0084] 定义

[0085] 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,本文描述优选的材料和方法。因此,本文中使用的术语仅用于描述特定实施方案的目的,并不旨在进行限制。

[0086] 除非另有定义,否则本文使用的所有技术和科学术语均具有本领域技术人员通常理解的含义。以下参考文献为本领域技术人员提供本发明中使用的许多术语的一般定义:Singleton et al.,Dictionary of Microbiology and Molecular Biology(2nd ed.1994);The Cambridge Dictionary of Science and Technology(Walker ed.,1988);The Glossary of Genetics,5th Ed.,R.Rieger et al.(eds.),Springer Verlag(1991);及Hale&Marham,The Harper Collins Dictionary of Biology(1991)。

[0087] 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”、“一个”和“所述”包含复数参考,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。此外,术语“包含”以及诸如“包括”、“含有”和“所含(included)”等其他形式的使用不是限制性的。

[0088] 如在本说明书和权利要求中使用的,词语“包括”(以及任何形式的包括,例如“包括”和“包含”)、“具有”(以及任何形式的具有,例如“具有”和“有”)、“包含”(以及任何形式的包含,例如“包括”和“含”)或“含有”(以及任何形式的含有,例如“含”和“含有”)是包容性或开放式,并且不排除额外的、未提及的组件或方法步骤。预期本说明书中讨论的任何实施方案可关于本公开的任何方法或组合实施,反之亦然。此外,本公开的组合物可用于实现本公开的方法。

[0089] 术语“大约”或“约”是指在由本领域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所述值,即测量系统。例如,根据本领域的实践,“大约”可以表示在1个标准偏差以内或超过1个标准偏差。或者,“大约”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“大约”的含义应假定在特定值的可接受误差范围内。

[0090] 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的范围被理解为包含来自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。

[0091] 说明书中对“一些实施方案”、“实施方案”、“一个实施方案”或“其他实施方案”的引用是指结合实施方案描述的特定特征、结构或特性包括在至少一些实施方案中,但不包括必须是本公开的所有实施方案。

[0092] “无碱基编辑器”是指能够切除核碱基并插入DNA核碱基(A、T、C或G)的试剂。无碱基编辑器包括核酸糖基化酶多肽或其片段。在一个实施方案中,核酸糖基化酶是突变的人类尿嘧啶DNA糖基化酶,其包括以下序列中的氨基酸204处的Asp(例如,置换氨基酸204处的Asn),或尿嘧啶DNA糖基化酶中的相应位置,并且具有胞嘧啶-DNA糖基化酶活性或其活性片段。在一个实施方案中,核酸糖基化酶是突变的人类尿嘧啶DNA糖基化酶,其包括以下序列中氨基酸147处的Ala、Gly、Cys或Ser(例如,置换氨基酸147处的Tyr),或尿嘧啶DNA糖基化酶中的相应位置,并具有胸腺嘧啶-DNA糖基化酶活性或其活性片段。实施例性人类尿嘧啶-DNA糖基化酶异形体1的序列如下:

[0093] 1mgvfclgpwg lgrklrtpgk gplqllsrlc gdhlqaipak kapagqeepg tppssplsae

[0094] 61qldriqrnka aallrlaarn vpvgfgeswk khlsgefgkp yfiklmgfva eerkhytvyp

[0095] 121pphqvftwtq mcdikdvkvv ilgqdpyhgp nqahglcfsv qrpvpppps l eniykelstd

[0096] 181 iedfvhpghg dlsgwakqgv lllnavltvr ahqanshker gweqftdavv swlnqnsngl

[0097] 241 vfllwgsyaq kkgsaidrkr hhvlqtahps plsvyrgffg crhfsktnel lqksgkkpid

[0098] 301 wkel

[0099] 人类尿嘧啶-DNA糖基化酶异形体2的序列如下:

[0100]
```
  1 migqktlysf fspsparkrh apspepavqg tgvagvpees gdaaaipakk apagqeepgt
 61 ppssplsaeq ldriqrnkaa allrlaarnv pvgfgeswkk hlsgefgkpy fiklmgfvae
121 erkhytvypp phqvftwtqm cdikdvkvvi lgqdpyhgpn qahglcfsvq rpvpppps le
181 niykelstdi edfvhpghgd lsgwakqgvl llnavltvra hqanshkerg weqftdavvs
241 wlnqnsnglv fllwgsyaqk kgsaidrkrh hvlqtahpsp lsvyrgffgc rhfsktnell
301 qksgkkpidw kel
```

[0101] 在其他实施方案中,无碱基编辑器是PCT/JP2015/080958和US20170321210中描述的任意一种无碱基编辑器,其通过引用方式并入本文。在特定实施方案中,无碱基编辑器包括在以上序列中以粗体显示的位置的突变并具有底线,或在本领域中已知的任何其他无碱基编辑器或尿嘧啶去糖基酶中的相应的氨基酸处。在一个实施方案中,无碱基编辑器包括在Y147、N204、L272和/或R276或相应位置的突变。在另一个实施方案中,无碱基编辑器包括Y147A或Y147G突变或相应的突变。在另一个实施方案中,无碱基编辑器包括N204D突变或相应的突变。在另一个实施方案中,无碱基编辑器包括L272A突变或相应的突变。在另一个实施方案中,无碱基编辑器包括R276E或R276C突变或相应的突变。

[0102] “腺苷脱氨酶”是指能够催化腺嘌呤或腺苷水解脱氨的多肽或其片段。在一些实施方案中,脱氨酶或脱氨酶结构域是催化腺苷水解脱氨为肌苷或脱氧腺苷水解脱氨为脱氧肌苷的腺苷脱氨酶。在一些实施方案中,腺苷脱氨酶催化脱氧核糖核酸(DNA)中腺嘌呤或腺苷的水解脱氨。本文所提供的腺苷脱氨酶(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。

[0103] 在一些实施方案中,腺苷脱氨酶是TadA脱氨酶。在一些实施方案中,TadA脱氨酶是TadA变体。在一些实施方案中,TadA变体是TadA*8。在一些实施方案中,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。在一些实施方案中,脱氨酶或脱氨酶结构域在自然界中不存在。例如,在一些实施方案中,脱氨酶或脱氨酶结构域为至少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%与天然存在的脱氨酶。例如,国际PCT申请第PCT/2017/045381号(WO 2018/027078)和国际PCT申请第PCT/US2016/058344号(WO 2017/070632)中描述了脱氨酶结构域,各自通过引用整体并入本文。另参见Komor,A.C.,et al.,"Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage"Nature 533,420-424(2016);Gaudelli,N.M.,et al.,"Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage"Nature 551,464-471(2017);Komor,A.C.,et al.,"Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity"Science Advances 3:eaao4774(2017)),以及Rees,H.A.,et al.,"Base editing:precision chemistry on the genome and transcriptome of living cells."Nat Rev Genet.2018Dec;19(12):770-788.doi:10.1038/s41576-018-0059-1,其全部内容通过引用方式并入本文。

[0104] 野生型TadA(wt)腺苷脱氨酶具有以下序列(也称为TadA参考序列):

[0105] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD(SEQ ID NO:2)。

[0106] 在一些实施方案中,腺苷脱氨酶包括以下序列的改变:

[0107] MSEVEFSHEY WMRHALTLAK RARDEREVPV GAVLVLNNRV IGEGWNRAIG LHDPTAHAEI MALRQGGLVM QNYRLIDATL YVTFEPCVMC AGAMIHSRIG RVVFGVRNAK TGAAGSLMDV LHYPGMNHRV EITEGILADE CAALLCYFFR MPRQVFNAQK KAQSSTD

[0108] (也称为TadA*7.10)。

[0109] 在一些实施方案中,TadA*7.10包括至少一种改变。在一些实施方案中,TadA*7.10包括氨基酸82和/或166处的改变。在特定实施方案中,上述序列的变体包括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中的改变H123Y回复为Y123H(野生型))。在其他实施方案中,TadA*7.10序列的变体包括选自由以下所组成的群组的改变的组合: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;及I76Y+V82S+Y123H+Y147R+Q154R。

[0110] 在其他实施方案中,本发明提供包含缺失例如TadA*8的腺苷脱氨酶变体,其包括

从残基149、150、151、152、153、154、155、156或157开始的C端的缺失。在其他实施方案中,腺苷脱氨酶变体是包括一个或多个以下改变的TadA(例如TadA*8)单体:Y147T、Y147R、Q154S、Y123H、V82S、T166R和/或Q154R。在其他实施方案中,腺苷脱氨酶变体是TadA(例如,TadA*8),一种包括选自由以下所组成的群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;及I76Y+V82S+Y123H+Y147R+Q154R。

[0111] 在其他实施方案中,腺苷脱氨酶变体是包括两个腺苷脱氨酶结构域(例如,TadA*8)的同源二聚体,各个所述腺苷脱氨酶结构域具有一个或多个以下改变Y147T、Y147R、Q154S、Y123H、V82S、T166R、和/或Q154R。在其他实施方案中,腺苷脱氨酶变体是包括两个腺苷脱氨酶结构域(例如,TadA*8)的同源二聚体,每个结构域具有选自由以下所组成的群组的改变的组合: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;及I76Y+V82S+Y123H+Y147R+Q154R。

[0112] 在其他实施方案中,腺苷脱氨酶变体是包括野生型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,其包括一种或多种以下改变Y147T、Y147R、Q154S、Y123H、V82S、T166R、和/或Q154R。在其他实施方案中,腺苷脱氨酶变体是包括野生型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,所述腺苷脱氨酶变体结构域包括选自由以下所组成的群组的改变的组合: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;及I76Y+V82S+Y123H+Y147R+Q154R。

[0113] 在其他实施方案中,腺苷脱氨酶变体是包括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,所述腺苷脱氨酶变体结构域包括一种或多种以下改变Y147T、Y147R、Q154S、Y123H、V82S、T166R、和/或Q154R。在其他实施方案中,腺苷脱氨酶变体是包括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,所述腺苷脱氨酶变体结构域包括一种或多种以下改变的组合: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。

[0114] 在一个实施方案中,腺苷脱氨酶是TadA*8,其包括或基本上由以下具有腺苷脱氨酶活性的序列或其片段所组成:

[0115] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD。

[0116] 在一些实施方案中,TadA*8是截短的。在一些实施方案中,截短的TadA*8相对于全长的TadA*8缺少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个N端氨基酸残基。在一些实施方案中,截短的TadA*8相对于全长的TadA*8缺少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、6、17、18、19或20个C端氨基酸残基。在一些实施方案中,腺苷脱氨酶变体是全长的TadA*8。

[0117] 在特定实施方案中,腺苷脱氨酶异源二聚体包括选自由以下之一者的TadA*8结构域和腺苷脱氨酶结构域:

[0118] 在特定实施方案中,腺苷脱氨酶异源二聚体包括选自由以下之一者的TadA*8结构域和腺苷脱氨酶结构域:

[0119] 大肠杆菌(Escherichia coli)TadA:

[0120] MRRAFITGVFFLSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD

[0121] 大肠杆菌(E.coli)TadA(N端是截短的):

[0122] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD

[0123] 金黄色葡萄球菌(Staphylococcus aureus(S.aureus))TadA:

[0124] MGSHMTNDIYFMTLAIEEAKKAAQLGEVPIGAIITKDDEVIARAHNLRETLQQPTAHAEHIAIERAAKVLGSWRLEGCTLYVTLEPCVMCAGTIVMSRIPRVVYGADDPKGGCSGSLMNLLQQSNFNHRAIVDKGVLKEACSTLLTTFFKNLRANKKSTN

[0125] 枯草芽孢杆菌(Bacillus subtilis(B.subtilis))TadA:

[0126] MTQDELYMKEAIKEAKKAEEKGEVPIGAVLVINGEIIARAHNLRETEQRSIAHAEMLVIDEACKALGTWRLEGATLYVTLEPCPMCAGAVVLSRVEKVVFGAFDPKGGCSGTLMNLLQEERFNHQAEVVSGVLEEECGGMLSAFFRELRKKKKAARKNLSE

[0127] 鼠伤寒沙门氏菌(Salmonella typhimurium(S.typhimurium))TadA:

[0128] MPPAFITGVTSLSDVELDHEYWMRHALTLAKRAWDEREVPVGAVLVHNHRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVLQNYRLLDTTLYVTLEPCVMCAGAMVHSRIGRVVFGARDAKTGAAGSLIDVLHHPGMNHRVEIIEGVLRDECATLLSDFFRMRRQEIKALKKADRAEGAGPAV

[0129] 腐败希瓦氏菌(Shewanella putrefaciens(S.putrefaciens))TadA:

[0130] MDEYWMQVAMQMAEKAEAAGEVPVGAVLVKDGQQIATGYNLSISQHDPTAHAEILCLRSAGKKLENYRLLDATLYITLEPCAMCAGAMVHSRIARVVYGARDEKTGAAGTVVNLLQHPAFNHQVEVTSGVLAEACSAQLSRFFKRRRDEKKALKLAQRAQQGIE

[0131] 流感嗜血杆菌F3031(Haemophilus influenzae F3031(H.influenzae))TadA:

[0132] MDAAKVRSEFDEKMMRYALELADKAEALGEIPVGAVLVDDARNIIGEGWNLSIVQSDPTAHAEIIALRNGAKNIQNYRLLNSTLYVTLEPCTMCAGAILHSRIKRLVFGASDYKTGAIGSRFHFFDDYKMNHTLEITSGVLAEECSQKLSTFFQKRREEKKIEKALLKSLSDK

[0133] 新月柄杆菌(Caulobacter crescentus(C.crescentus))TadA:

MRTDESEDQDHRMMRLALDAARAAAEAGETPVGAVILDPSTGEVIATAGNGPIAAHDPTAHAEIAAM

[0134] RAAAAKLGNYRLTDLTLVVTLEPCAMCAGAISHARIGRVVFGADDPKGGAVVHGPKFFAQPTCHWRP

EVTGGVLADESADLLRGFFRARRKAKI

[0135] 硫还原地杆菌(Geobacter sulfurreducens(G.sulfurreducens))TadA:

[0136] MSSLKKTPIRDDAYWMGKAIREAAKAAARDEVPIGAVIVRDGAVIGRGHNLREGSNDPSAHAEMIAIRQ AARRSANWRLTGATLYVTLEPCLMCMGAIILARLERVVFGCYDPKGGAAGSLYDLSADPRLNHQVRLSPGVCQEECG TMLSDFFRDLRRRKKAKATPALFIDERKVPPEP

[0137] TadA*7.10

[0138] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAE

[0139] IMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLM

[0140] DVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

[0141] 预期作为具有TadA*8的异源二聚体组分的其他TadA7.10或TadA7.10变体包括:

[0142] GSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRA

[0143] IGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAG

[0144] SLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

[0145] TadA7.10 CP65

[0146] TAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVL

[0147] HYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGS

[0148] EVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDP

[0149] TadA7.10 CP83

[0150] YRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILAD

[0151] ECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAK

[0152] RARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQN

[0153] TadA7.10 CP136

[0154] MNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEF

[0155] SHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNY

[0156] RLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPG

[0157] TadA7.10 C-截短

[0158] GSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRA

[0159] IGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAG

[0160] SLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFN

[0161] TadA7.10 C-截短2

[0162] GSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRA

[0163] IGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAG

[0164] SLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQ

[0165] TadA7.10Δ59-66+C-截短

[0166] GSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRA

[0167] HAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHY

[0168] PGMNHRVEITEGILADECAALLCYFFRMPRQVFN

[0169] TadA7.10Δ59-66

[0170] GSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRV

[0171] IGEGWNRAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVR

[0172] NAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD。

[0173] 在一些实施方案中,腺苷脱氨酶变体包括TadA7.10的改变。在一些实施方案中,TadA7.10包括在氨基酸82或166处的改变。在特定实施方案中,上述序列中的变体包括以下改变中的一个或多个:Y147T、Y147R、Q154S、Y123H、V82S、T166R、和Q154R。在其他实施方案中,腺苷脱氨酶变体包括选自由Y147R+Q154R+Y123H;Y147R+Q154R+I76Y;Y147R+Q154R+T166R;Y147T+Q154R;Y147T+Q154S;和Y123H+Y147R+Q154R+I76Y所组成的群组的改变的组合。

[0174] 在其他实施方案中,本发明提供包含缺失的腺苷脱氨酶变体,例如包括从残基149、150、151、152、153、154、155、156或157处开始的C端缺失的TadA7.10。在一些实施方案中,腺苷脱氨酶变体是包括一个或多个以下改变的TadA单体:Y147T、Y147R、Q154S、Y123H、V82S、T166R、Q154R。在其他实施方案中,腺苷脱氨酶变体是包括以下改变的单体:Y147R+Q154R+Y123H;Y147R+Q154R+I76Y;Y147R+Q154R+T166R;Y147T+Q154R;Y147T+Q154S;及Y123H+Y147R+Q154R+I76Y。在又其他实施方案中,腺苷脱氨酶变体是包括两个腺苷脱氨酶结构域的同源二聚体,各个所述腺苷脱氨酶结构域具有一个或多个以下改变Y147T、Y147R、Q154S、Y123H、V82S、T166R、Q154R。在其他实施方案中,腺苷脱氨酶变体是包括野生型腺苷脱氨酶结构域或TadA7.10结构域和腺苷脱氨酶变体结构域的异源二聚体,所述腺苷脱氨酶变体结构域包括一种或多种以下改变Y147T、Y147R、Q154S、Y123H、V82S、T166R,Q154R。在其他实施方案中,腺苷脱氨酶变体是包括TadA7.10结构域和TadA7.10的腺苷脱氨酶变体的异源二聚体,该TadA7.10的腺苷脱氨酶变体包括以下改变:Y147R+Q154R+Y123H;Y147R+Q154R+I76Y;Y147R+Q154R+T166R;Y147T+Q154R;Y147T+Q154S;及Y123H+Y147R+Q154R+I76Y。

[0175] “给药”在本文中是指向患者或受试者提供本文所述一种或多种组合物。例如但不限于,组合物给药(例如注射),其可以通过静脉内(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.)注射进行。可以采用一种或多种这样的途径。肠胃外给药可以是,例如,通过推注或随时间逐渐灌注。在一些实施方案中,肠胃外给药包括血管内、静脉内、肌肉内、动脉内、鞘内、瘤内、皮内、腹膜内、经气管、皮下、胶质层下(subcuticularly)、关节内、囊下、蛛网膜下腔和胸骨内输注或注射。或者,或同时,可以通过口服途径给药。

[0176] “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。

[0177] “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,诸如本文所述那些方法检测的。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10%变化、25%变化、40%变化、50%变化或更大。

[0178] “改善”是指减少、抑制、减弱、减少、阻止或稳定疾病的发展或进展。

[0179] “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,同时具有相对于天然存在的多核苷酸或多肽增强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。

[0180] “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试剂。在各种实施方案中,碱基编辑器包括核碱基修饰多肽(例如,脱氨酶)和与向导多核苷酸(例如,向导RNA)结合的核酸可编程的核苷酸结合结构域。在各种实施方案中,所述药剂是包括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,DNA)内的碱基(例如,A、T、C、G或U)的结构域。在一些实施方案中,多核苷酸可编程的DNA结合结构域与脱氨酶结构域融合或连接。在一个实施方案中,所述药剂是包括具有碱基编辑活性的结构域的融合蛋白。在另一个实施方案中,具有碱基编辑活性的蛋白质结构域与向导RNA连接(例如,经由向导RNA上的RNA结合基序和与脱氨酶融合的RNA结合结构域)。在一些实施方案中,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。在一些实施方案中,碱基编辑器能够使DNA分子内的一个或多个碱基脱氨。在一些实施方案中,碱基编辑器能够使DNA内的腺苷(A)脱氨。在一些实施方案中,碱基编辑器是腺苷碱基编辑器(ABE)。

[0181] “胞苷脱氨酶”是指能够催化将氨基转换为羰基的脱氨反应的多肽或其片段。在一些实施方案中,胞苷脱氨酶与APOBEC或AID具有至少约85%的同一性。在一个实施方案中,胞苷脱氨酶将胞嘧啶转换为尿嘧啶或将5-甲基胞嘧啶转换为胸腺嘧啶。PmCDA1(衍生自海洋鳃鳗(Petromyzon marinus)(海七鳃鳗胞嘧啶脱氨酶1,“PmCDA1”)、AID(激活诱导胞苷脱氨酶;(activation-induced cytidine deaminase,AICDA))(衍生自哺乳动物(例如,人类、猪、牛、马、猴等)和APOBEC是实施例性的胞苷脱氨酶。

[0182] 在一些实施方案中,碱基编辑器是与脱氨酶(例如,腺苷脱氨酶或胞苷脱氨酶)融合的可重编程的碱基编辑器。在一些实施方案中,碱基编辑器是与脱氨酶(例如,腺苷脱氨酶或胞苷脱氨酶)融合的Cas9。在一些实施方案中,碱基编辑器是与脱氨酶(例如,腺苷脱氨酶或胞苷脱氨酶)融合的无核酸酶的Cas9(dCas9)。在一些实施方案中,Cas9是环状排列的Cas9(例如,spCas9或saCas9)。环状排列的Cas9是本领域已知的并且描述于例如Oakes et al.,Cell 176,254-267,2019。在一些实施方案中,碱基编辑器与碱基切除修复抑制剂融合,例如,UGI结构域或dISN结构域。在一些实施方案中,融合蛋白包括与脱氨酶和碱基切除修复抑制剂诸如UGI或dISN结构域融合的Cas9切口酶。在其他实施方案中,碱基编辑器是无碱基碱基编辑器。

[0183] 在一些实施方案中,碱基编辑器是腺苷碱基编辑器(ABE)。在一些实施方案中,腺苷脱氨酶由TadA进化而来。在一些实施方案中,本发明的碱基编辑器包括具有内部融合的催化性(例如脱氨酶)结构域的napDNAbp结构域。在一些实施方案中,napDNAbp是具有内部融合的脱氨酶结构域的Cas12a(Cpf1)。在一些实施方案中,napDNAbp是具有内部融合的脱氨酶结构域的Cas12b(c2c1)。在一些实施方案中,napDNAbp是具有内部融合的脱氨酶结构域的Cas12c(c2c3)。在一些实施方案中,napDNAbp是具有内部融合的脱氨酶结构域的Cas12d(CasX)。在一些实施方案中,napDNAbp是具有内部融合的脱氨酶结构域的Cas12e(CasY)。在一些实施方案中,napDNAbp是具有内部融合的脱氨酶结构域的Cas12g。在一些实施方案中,napDNAbp是具有内部融合的脱氨酶结构域的Cas12h。在一些实施方案中,napDNAbp是具有内部融合的脱氨酶结构域的Cas12i。在一些实施方案中,碱基编辑器是与脱氨酶结构域融合的催化性死亡的Cas12(dCas12)。在一些实施方案中,碱基编辑器是与脱

氨酶结构域融合的Cas12切口酶(nCas12)。

[0184] 在一些实施方案中,碱基编辑器是通过将腺苷脱氨酶变体(例如,TadA*8)克隆到包括环状排列的Cas9(例如,spCAS9或saCAS9)和二分核定位序列的支架中来产生(例如,ABE8)。环状排列的Cas9是本领域已知的并且描述于例如Oakes et al.,Cell 176,254-267,2019。实施例性环状排列物如下,其中粗体序列表示衍生自Cas9的序列,斜体序列表示连接子序列,底线序列表示二分核定位序列。

[0185] CP5(具有MSP"NGC=具有突变的Pam变体常规Cas9如同NGG"PID=蛋白质相互作用结构域(Protein Interacting Domain)和"D10A"切口酶):

[0186] **EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNI**
VKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPTVAYSVLVVAKVEKGKSKKLKSVK
ELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAKFLQKGNELAL
PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYN
KHRDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYRSTKEVLDATLIHQSITGLYETRIDL
SQLGGD*GGSGGSGGSGGSGGSGGSGGM***DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRH**
SIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVE
EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLI
ALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVN
TEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFI
KPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKI
LTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPK
HSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS
VEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKA
QVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSR
ERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQS
FLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELD
KAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREIN
NYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEGADKRTADGSEFESPKKK
RKV*

[0187] 在一些实施方案中,ABE8是选自下表6至9、13或14的碱基编辑器。在一些实施方案中,ABE8含有从TadA进化而来的腺苷脱氨酶变体。在一些实施方案中,ABE8的腺苷脱氨酶变体是如下表7、9、13或14中所述TadA*8变体。在一些实施方案中,腺苷脱氨酶变体是包括一个或多个选自Y147T、Y147R、Q154S、Y123H、V82S、T166R、和/或Q154R的改变的TadA*7.10变体(例如TadA*8)。在各种实施方案中,ABE8包括具有选自由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所组成的群组的改变的组合的TadA*7.10变体(例如TadA*8)。在一些实

施方案中,ABE8是单体构建体。在一些实施方案中,ABE8是异源二聚体构建体。在一些实施方案中,ABE8碱基编辑器包括以下序列:

[0188] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD。

[0189] 在一些实施方案中,多核苷酸可编程的DNA结合结构域是CRISPR相关(例如,Cas或Cpf1)酶。在一些实施方案中,碱基编辑器是与脱氨酶结构域融合的催化性死亡的Cas9(dCas9)。在一些实施方案中,碱基编辑器是与脱氨酶结构域融合的Cas9切口酶(nCas9)。在一些实施方案中,碱基编辑器与碱基切除修复(base excision repair,BER)抑制剂融合。在一些实施方案中,碱基切除修复抑制剂是尿嘧啶DNA糖基化酶抑制剂(UGI)。在一些实施方案中,碱基切除修复抑制剂是肌苷碱基切除修复抑制剂。

[0190] 碱基编辑器的细节在国际PCT申请第PCT/2017/045381(WO 2018/027078)号和国际PCT申请第PCT/US2016/058344(WO 2017/070632)号中进行了描述,它们各自通过引用整体并入本文。另参见Komor,A.C.et al.,“无Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage”Nature533,420-424(2016);Gaudelli,N.M.,et al.,“Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage”Nature 551,464-471(2017);Komor,A.C.,et al.,“Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity”Science Advances 3:eaao4774(2017),以及Rees,H.A.,et al.,“Base editing:precision chemistry on the genome and transcriptome of living cells.”Nat Rev Genet.2018Dec;19(12):770-788.doi:10.1038/s41576-018-0059-1,其全部内容特此引入作为参考。

[0191] 举例来说,本文所述的碱基编辑组合物、系统和方法中使用的胞苷碱基编辑器具有以下核酸序列(8877个碱基对),(Addgene,Watertown,MA.;Komor AC,et al.,2017,Sci Adv.,30;3(8):eaao4774.doi:10.1126/sciadv.aao4774)如下。还包括与BE4核酸序列具有至少95%或更高同一性的多核苷酸序列。

[0192]

1 atatgccaag tacgcccccct attgacgtca atgacggtaa atggcccgcc tggcattatg

61 cccagtacat gaccttatgg gactttccta cttggcagta catctacgta ttagtcatcg

121 ctattaccat ggtgatgcgg ttttggcagt acatcaatgg gcgtggatag cggtttgact

181 cacggggatt tccaagtctc caccccattg acgtcaatgg gagtttgttt tggcaccaaa

241 atcaacggga ctttccaaaa tgtcgtaaca actccgcccc attgacgcaa atgggcggta

301 ggcgtgtacg gtgggaggtc tatataagca gagctggttt agtgaaccgt cagatccgct

361 agagatccgc ggccgctaat acgactcact atagggagag ccgccaccat gagctcagag

421 actggcccag tggctgtgga ccccacattg agacggcgga tcgagcccca tgagtttgag

481 gtattcttcg atccgagaga gctccgcaag gagacctgcc tgctttacga aattaattgg

541 gggggccggc actccatttg gcgacataca tcacagaaca ctaacaagca cgtcgaagtc

601 aacttcatcg agaagttcac gacagaaaga tatttctgtc cgaacacaag gtgcagcatt

661 acctggtttc tcagctggag cccatgcggc gaatgtagta gggccatcac tgaattcctg

721 tcaaggtatc cccacgtcac tctgtttatt tacatcgcaa ggctgtacca ccacgctgac

781 ccccgcaatc gacaaggcct gcgggatttg atctcttcag gtgtgactat ccaaattatg

841 actgagcagg agtcaggata ctgctggaga aactttgtga attatagccc gagtaatgaa

901 gcccactggc ctaggtatcc ccatctgtgg gtacgactgt acgttcttga

actgtactgc

961 atcatactgg gcctgcctcc ttgtctcaac attctgagaa ggaagcagcc acagctgaca

1021 ttctttacca tcgctcttca gtcttgtcat taccagcgac tgcccccaca cattctctgg

1081 gccaccgggt tgaaatctgg tggttcttct ggtggttcta gcggcagcga gactcccggg

1141 acctcagagt ccgccacacc cgaaagttct ggtggttctt ctggtggttc tgataaaaag

1201 tattctattg gtttagccat cggcactaat tccgttggat gggctgtcat aaccgatgaa

1261 tacaaagtac cttcaaagaa atttaaggtg ttggggaaca cagaccgtca ttcgattaaa

1321 aagaatctta tcggtgccct cctattcgat agtggcgaaa cggcagaggc gactcgcctg

1381 aaacgaaccg ctcggagaag gtatacacgt cgcaagaacc gaatatgtta cttacaagaa

1441 atttttagca atgagatggc caaagttgac gattctttct ttcaccgttt ggaagagtcc

[0193]

1501 ttccttgtcg aagaggacaa gaaacatgaa cggcacccca tctttggaaa catagtagat

1561 gaggtggcat atcatgaaaa gtacccaacg atttatcacc tcagaaaaaa gctagttgac

1621 tcaactgata aagcggacct gaggttaatc tacttggctc ttgcccatat gataaagttc

1681 cgtgggcact ttctcattga gggtgatcta aatccggaca actcggatgt cgacaaactg

1741 ttcatccagt tagtacaaac ctataatcag ttgtttgaag agaaccctat aaatgcaagt

1801 ggcgtggatg cgaaggctat tcttagcgcc cgcctctcta aatcccgacg gctagaaaac

1861 ctgatcgcac aattacccgg agagaagaaa aatgggttgt tcggtaacct tatagcgctc

1921 tcactaggcc tgacaccaaa ttttaagtcg aacttcgact tagctgaaga tgccaaattg

1981 cagcttagta aggacacgta cgatgacgat ctcgacaatc tactggcaca aattggagat

2041 cagtatgcgg acttattttt ggctgccaaa aaccttagcg atgcaatcct

cctatctgac

2101 atactgagag ttaatactga gattaccaag gcgccgttat ccgcttcaat gatcaaaagg

2161 tacgatgaac atcaccaaga cttgacactt ctcaaggccc tagtccgtca gcaactgcct

2221 gagaaatata aggaaatatt ctttgatcag tcgaaaaacg ggtacgcagg ttatattgac

2281 ggcggagcga gtcaagagga attctacaag tttatcaaac ccatattaga gaagatggat

2341 gggacggaag agttgcttgt aaaactcaat cgcgaagatc tactgcgaaa gcagcggact

2401 ttcgacaacg gtagcattcc acatcaaatc cacttaggcg aattgcatgc tatacttaga

2461 aggcaggagg atttttatcc gttcctcaaa gacaatcgtg aaaagattga gaaaatccta

2521 acctttcgca taccttacta tgtgggaccc ctggcccgag ggaactctcg gttcgcatgg

2581 atgacaagaa agtccgaaga aacgattact ccatggaatt ttgaggaagt tgtcgataaa

[0194]

2641 ggtgcgtcag ctcaatcgtt catcgagagg atgaccaact ttgacaagaa tttaccgaac

2701 gaaaaagtat tgcctaagca cagtttactt tacgagtatt tcacagtgta caatgaactc

2761 acgaaagtta agtatgtcac tgagggcatg cgtaaacccg cctttctaag cggagaacag

2821 aagaaagcaa tagtagatct gttattcaag accaaccgca aagtgacagt taagcaattg

2881 aaagaggact actttaagaa aattgaatgc ttcgattctg tcgagatctc cggggtagaa

2941 gatcgattta atgcgtcact tggtacgtat catgacctcc taaagataat taaagataag

3001 gacttcctgg ataacgaaga gaatgaagat atcttagaag atatagtgtt gactcttacc

3061 ctctttgaag atcgggaaat gattgaggaa agactaaaaa catacgctca cctgttcgac

3121 gataaggtta tgaaacagtt aaagaggcgt cgctatacgg gctggggacg attgtcgcgg

3181 aaacttatca acgggataag agacaagcaa agtggtaaaa ctattctcga

ttttctaaag

3241 agcgacggct tcgccaatag gaactttatg cagctgatcc atgatgactc tttaaccttc

3301 aaagaggata tacaaaaggc acaggtttcc ggacaagggg actcattgca cgaacatatt

3361 gcgaatcttg ctggttcgcc agccatcaaa aagggcatac tccagacagt caaagtagtg

3421 gatgagctag ttaaggtcat gggacgtcac aaaccggaaa acattgtaat cgagatggca

3481 cgcgaaaatc aaacgactca gaaggggcaa aaaaacagtc gagagcggat gaagagaata

3541 gaagagggta ttaaagaact gggcagccag atcttaaagg agcatcctgt ggaaaatacc

3601 caattgcaga acgagaaact ttacctctat tacctacaaa atggaaggga catgtatgtt

3661 gatcaggaac tggacataaa ccgtttatct gattacgacg tcgatcacat tgtaccccaa

3721 tccttttga aggacgattc aatcgacaat aaagtgctta cacgctcgga taagaaccga

[0195]

3781 gggaaaagtg acaatgttcc aagcgaggaa gtcgtaaaga aaatgaagaa ctattggcgg

3841 cagctcctaa atgcgaaact gataacgcaa agaaagttcg ataacttaac taaagctgag

3901 aggggtggct tgtctgaact tgacaaggcc ggatttatta aacgtcagct cgtggaaacc

3961 cgccaaatca caaagcatgt tgcacagata ctagattccc gaatgaatac gaaatacgac

4021 gagaacgata agctgattcg ggaagtcaaa gtaatcactt taaagtcaaa attggtgtcg

4081 gacttcagaa aggattttca attctataaa gttagggaga taaataacta ccaccatgcg

4141 cacgacgctt atcttaatgc cgtcgtaggg accgcactca ttaagaaata cccgaagcta

4201 gaaagtgagt ttgtgtatgg tgattacaaa gtttatgacg tccgtaagat gatcgcgaaa

4261 agcgaacagg agataggcaa ggctacagcc aaatacttct tttattctaa cattatgaat

4321 ttctttaaga cggaaatcac tctggcaaac ggagagatac gcaaacgacc

tttaattgaa

4381 accaatgggg agacaggtga aatcgtatgg gataagggcc gggacttcgc gacggtgaga

4441 aaagttttgt ccatgcccca agtcaacata gtaaagaaaa ctgaggtgca gaccggaggg

4501 ttttcaaagg aatcgattct tccaaaaagg aatagtgata agctcatcgc tcgtaaaaag

4561 gactgggacc cgaaaaagta cggtggcttc gatagcccta cagttgccta ttctgtccta

4621 gtagtggcaa aagttgagaa gggaaaatcc aagaaactga agtcagtcaa agaattattg

4681 gggataacga ttatggagcg ctcgtctttt gaaaagaacc ccatcgactt ccttgaggcg

4741 aaaggttaca aggaagtaaa aaaggatctc ataattaaac taccaaagta tagtctgttt

4801 gagttagaaa atggccgaaa acggatgttg gctagcgccg gagagcttca aaaggggaac

4861 gaactcgcac taccgtctaa atacgtgaat ttcctgtatt tagcgtccca ttacgagaag

[0196]

4921 ttgaaaggtt cacctgaaga taacgaacag aagcaacttt ttgttgagca gcacaaacat

4981 tatctcgacg aaatcataga gcaaatttcg gaattcagta agagagtcat cctagctgat

5041 gccaatctgg acaaagtatt aagcgcatac aacaagcaca gggataaacc catacgtgag

5101 caggcggaaa atattatcca tttgtttact cttaccaacc tcggcgctcc agccgcattc

5161 aagtattttg acacaacgat agatcgcaaa cgatacactt ctaccaagga ggtgctagac

5221 gcgacactga ttcaccaatc catcacggga ttatatgaaa ctcggataga tttgtcacag

5281 cttgggggtg actctggtgg ttctggagga tctggtggtt ctactaatct gtcagatatt

5341 attgaaaagg agaccggtaa gcaactggtt atccaggaat ccatcctcat gctcccagag

5401 gaggtggaag aagtcattgg gaacaagccg gaaagcgata tactcgtgca caccgcctac

5461 gacgagagca ccgacgagaa tgtcatgctt ctgactagcg acgcccctga

atacaagcct

5521 tgggctctgg tcatacagga tagcaacggt gagaacaaga ttaagatgct ctctggtggt

5581 tctggaggat ctggtggttc tactaatctg tcagatatta ttgaaaagga gaccggtaag

5641 caactggtta tccaggaatc catcctcatg ctcccagagg aggtggaaga agtcattggg

5701 aacaagccgg aaagcgatat actcgtgcac accgcctacg acgagagcac cgacgagaat

5761 gtcatgcttc tgactagcga cgcccctgaa tacaagcctt gggctctggt catacaggat

5821 agcaacggtg agaacaagat taagatgctc tctggtggtt ctcccaagaa gaagaggaaa

5881 gtctaaccgg tcatcatcac catcaccatt gagtttaaac ccgctgatca gcctcgactg

5941 tgccttctag ttgccagcca tctgttgttt gcccctcccc cgtgccttcc ttgaccctgg

6001 aaggtgccac tcccactgtc ctttcctaat aaaatgagga aattgcatcg cattgtctga

[0197]

6061 gtaggtgtca ttctattctg gggggtgggg tggggcagga cagcaagggg gaggattggg

6121 aagacaatag caggcatgct ggggatgcgg tgggctctat ggcttctgag gcggaaagaa

6181 ccagctgggg ctcgataccg tcgacctcta gctagagctt ggcgtaatca tggtcatagc

6241 tgtttcctgt gtgaaattgt tatccgctca caattccaca caacatacga gccggaagca

6301 taaagtgtaa agcctagggt gcctaatgag tgagctaact cacattaatt gcgttgcgct

6361 cactgcccgc tttccagtcg ggaaacctgt cgtgccagct gcattaatga atcggccaac

6421 gcgcggggag aggcggtttg cgtattgggc gctcttccgc ttcctcgctc actgactcgc

6481 tgcgctcggt cgttcggctg cggcgagcgg tatcagctca ctcaaaggcg gtaatacggt

6541 tatccacaga atcaggggat aacgcaggaa agaacatgtg agcaaaaggc cagcaaaagg

6601 ccaggaaccg taaaaaggcc gcgttgctgg cgtttttcca taggctccgc

cccctgacg

6661 agcatcacaa aaatcgacgc tcaagtcaga ggtggcgaaa cccgacagga ctataaagat

6721 accaggcgtt tccccctgga agctccctcg tgcgctctcc tgttccgacc ctgccgctta

6781 ccggatacct gtccgccttt ctcccttcgg gaagcgtggc gctttctcat agctcacgct

6841 gtaggtatct cagttcggtg taggtcgttc gctccaagct gggctgtgtg cacgaacccc

6901 ccgttcagcc cgaccgctgc gccttatccg gtaactatcg tcttgagtcc aacccggtaa

6961 gacacgactt atcgccactg gcagcagcca ctggtaacag gattagcaga gcgaggtatg

7021 taggcggtgc tacagagttc ttgaagtggt ggcctaacta cggctacact agaagaacag

7081 tatttggtat ctgcgctctg ctgaagccag ttaccttcgg aaaaagagtt ggtagctctt

7141 gatccggcaa acaaaccacc gctggtagcg gtggtttttt tgtttgcaag cagcagatta

[0198]

7201 cgcgcagaaa aaaaggatct caagaagatc ctttgatctt ttctacgggg tctgacgctc

7261 agtggaacga aaactcacgt taagggattt tggtcatgag attatcaaaa aggatcttca

7321 cctagatcct tttaaattaa aaatgaagtt ttaaatcaat ctaaagtata tatgagtaaa

7381 cttggtctga cagttaccaa tgcttaatca gtgaggcacc tatctcagcg atctgtctat

7441 ttcgttcatc catagttgcc tgactccccg tcgtgtagat aactacgata cgggagggct

7501 taccatctgg ccccagtgct gcaatgatac cgcgagaccc acgctcaccg gctccagatt

7561 tatcagcaat aaaccagcca gccggaaggg ccgagcgcag aagtggtcct gcaactttat

7621 ccgcctccat ccagtctatt aattgttgcc gggaagctag agtaagtagt tcgccagtta

7681 atagtttgcg caacgttgtt gccattgcta caggcatcgt ggtgtcacgc tcgtcgtttg

7741 gtatggcttc attcagctcc ggttcccaac gatcaaggcg agttacatga

tcccccatgt

7801 tgtgcaaaaa agcggttagc tccttcggtc ctccgatcgt tgtcagaagt aagttggccg

7861 cagtgttatc actcatggtt atggcagcac tgcataattc tcttactgtc atgccatccg

7921 taagatgctt ttctgtgact ggtgagtact caaccaagtc attctgagaa tagtgtatgc

7981 ggcgaccgag ttgctcttgc ccggcgtcaa tacgggataa taccgcgcca catagcagaa

8041 ctttaaaagt gctcatcatt ggaaaacgtt cttcggggcg aaaactctca aggatcttac

8101 cgctgttgag atccagttcg atgtaaccca ctcgtgcacc caactgatct tcagcatctt

8161 ttactttcac cagcgtttct gggtgagcaa aaacaggaag gcaaaatgcc gcaaaaaagg

8221 gaataagggc gacacggaaa tgttgaatac tcatactctt cctttttcaa tattattgaa

8281 gcatttatca gggttattgt ctcatgagcg gatacatatt tgaatgtatt tagaaaaata

8341 aacaaatagg ggttccgcgc acatttcccc gaaaagtgcc acctgacgtc gacggatcgg

8401 gagatcgatc tcccgatccc ctagggtcga ctctcagtac aatctgctct gatgccgcat

8461 agttaagcca gtatctgctc cctgcttgtg tgttggaggt cgctgagtag tgcgcgagca

8521 aaatttaagc tacaacaagg caaggcttga ccgacaattg catgaagaat ctgcttaggg

8581 ttaggcgttt tgcgctgctt cgcgatgtac gggccagata tacgcgttga cattgattat

8641 tgactagtta ttaatagtaa tcaattacgg ggtcattagt tcatagccca tatatggagt

8701 tccgcgttac ataacttacg gtaaatggcc cgcctggctg accgcccaac gacccccgcc

8761 cattgacgtc aataatgacg tatgttccca tagtaacgcc aatagggact ttccattgac

8821 gtcaatgggt ggagtattta cggtaaactg cccacttggc agtacatcaa gtgtatc

[0199]

[0200] BE4氨基酸序列：

[0201] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFI

[0202] EKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDL

[0203] ISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQP

[0204] QLTFFTIALQSCHYQRLPPHILWATGLKSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIG

[0205] LAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR

[0206] KNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKL

[0207] VDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAK

[0208] AILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLD

[0209] NLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL

[0210] PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI

[0211] PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWN

[0212] FEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQ

[0213] KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEE

[0214] NEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKT

[0215] ILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVD

[0216] ELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLY

[0217] LYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKM

[0218] KNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEN

[0219] DKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDY

[0220] KVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF

[0221] ATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA

[0222] KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML

[0223] ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVI

[0224] LADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLI

[0225] HQSITGLYETRIDLSQLGGDSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPE

[0226] SDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGSTNLSDIIEKET

[0227] GKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGEN

[0228] KIKMLSGGSPKKKRK

[0229] 举例来说,本文所述的碱基编辑组合物、系统和方法中使用的腺嘌呤碱基编辑器(ABE)具有核酸序列(8877个碱基对),(Addgene,Watertown,MA.;Gaudelli NM,et al.,Nature.2017Nov23;551(7681):464-471.doi:10.1038/nature24644;Koblan LW,et al.,Nat Biotechnol.2018Oct;36(9):843-846.doi:10.1038/nbt.4172.)如下。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。

[0230] ATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCA

[0231] GTACAT

[0232] GACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTG

[0233] ATGCGG

[0234] TTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACC

[0235] CCATTG

[0236] ACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTC

[0237] CGCCCC

[0238] ATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTG

[0239] AACCGT

[0240] CAGATCCGCTAGAGATCCGCGGCCGCTAATACGACTCACTATAGGGAGAGCCGCCACCATGAAA

[0241] CGGACA

[0242] GCCGACGGAAGCGAGTTCGAGTCACCAAAGAAGAAGCGGAAAGTCTCTGAAGTCGAGTTTAGCC

[0243] ACGAGT

[0244] ATTGGATGAGGCACGCACTGACCCTGGCAAAGCGAGCATGGGATGAAAGAGAAGTCCCCGTGGG

[0245] CGCCGT

[0246] GCTGGTGCACAACAATAGAGTGATCGGAGAGGGATGGAACAGGCCAATCGGCCGCCACGACCCT

[0247] ACCGCA

[0248] CACGCAGAGATCATGGCACTGAGGCAGGGAGGCCTGGTCATGCAGAATTACCGCCTGATCGATG

[0249] CCACCC

[0250] TGTATGTGACACTGGAGCCATGCGTGATGTGCGCAGGAGCAATGATCCACAGCAGGATCGGAAG

[0251] AGTGGT

[0252] GTTCGGAGCACGGGACGCCAAGACCGGCGCAGCAGGCTCCCTGATGGATGTGCTGCACCACCCC

[0253] GGCATG

[0254] AACCACCGGGTGGAGATCACAGAGGGAATCCTGGCAGACGAGTGCGCCGCCCTGCTGAGCGATT

[0255] TCTTTA

[0256] GAATGCGGAGACAGGAGATCAAGGCCCAGAAGAAGGCACAGAGCTCCACCGACTCTGGAGGATC

[0257] TAGCGG

[0258] AGGATCCTCTGGAAGCGAGACACCAGGCACAAGCGAGTCCGCCACACCAGAGAGCTCCGGCGGC

[0259] TCCTCC

[0260] GGAGGATCCTCTGAGGTGGAGTTTTCCCACGAGTACTGGATGAGACATGCCCTGACCCTGGCCA

[0261] AGAGGG

[0262] CACGCGATGAGAGGGAGGTGCCTGTGGGAGCCGTGCTGGTGCTGAACAATAGAGTGATCGGCGA

[0263] GGGCTG

[0264] GAACAGAGCCATCGGCCTGCACGACCCAACAGCCCATGCCGAAATTATGGCCCTGAGACAGGGC

[0265] GGCCTG

[0266] GTCATGCAGAACTACAGACTGATTGACGCCACCCTGTACGTGACATTCGAGCCTTGCGTGATGT

[0267] GCGCCG

[0268] GCGCCATGATCCACTCTAGGATCGGCCGCGTGGTGTTTGGCGTGAGGAACGCAAAAACCGGCGC

[0269] CGCAGG

[0270] CTCCCTGATGGACGTGCTGCACTACCCCGGCATGAATCACCGCGTCGAAATTACCGAGGGAATC

[0271] CTGGCA

[0272] GATGAATGTGCCGCCCTGCTGTGCTATTTCTTTCGGATGCCTAGACAGGTGTTCAATGCTCAGA

[0273] AGAAGG

[0274] CCCAGAGCTCCACCGACTCCGGAGGATCTAGCGGAGGCTCCTCTGGCTCTGAGACACCTGGCAC

[0275] AAGCGA

[0276] GAGCGCAACACCTGAAAGCAGCGGGGGCAGCAGCGGGGGGTCAGACAAGAAGTACAGCATCGGC

[0277] CTGGCC

[0278] ATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAAT

[0279] TCAAGG

[0280] TGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAG

[0281] CGGCGA

[0282] AACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGG

[0283] ATCTGC

[0284] TATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGG

[0285] AAGAGT

[0286] CCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGA

[0287] GGTGGC

[0288] CTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAG

[0289] GCCGAC

[0290] CTGCGGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGG

[0291] GCGACC

[0292] TGAACCCCGACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCT

[0293] GTTCGA

[0294] GGAAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAG

[0295] AGCAGA

[0296] CGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGA

[0297] TTGCCC

[0298] TGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCA

[0299] GCTGAG

[0300] CAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGAC

[0301] CTGTTT

[0302] CTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGA

[0303] TCACCA

[0304] AGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCT

[0305] GAAAGC

[0306] TCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGC

[0307] TACGCC

[0308] GGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAA

[0309] AGATGG

[0310] ACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTT

[0311] CGACAA

[0312] CGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGAT

[0313] TTTTAC

[0314] CCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACG

[0315] TGGGCC

[0316] CTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCC

[0317] CTGGAA

[0318] CTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTC

[0319] GATAAG

[0320] AACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATA

[0321] ACGAGC

[0322] TGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCAGAA

[0323] AAAGGC

[0324] CATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTAC

[0325] TTCAAG

[0326] AAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGG

[0327] GCACAT

[0328] ACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACAT

[0329] TCTGGA

[0330] AGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACC

[0331] TATGCC

[0332] CACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGC

[0333] TGAGCC

[0334] GGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTC

[0335] CGACGG

[0336] CTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATC

[0337] CAGAAA

[0338] GCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCG

[0339] CCATTA

[0340] AGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAA

[0341] GCCCGA

[0342] GAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGC

[0343] GAGAGA

[0344] ATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGG

[0345] AAAACA

[0346] CCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGA

[0347] CCAGGA

[0348] ACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAG

[0349] GACGAC

[0350] TCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCT

[0351] CCGAAG

[0352] AGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAG

[0353] AAAGTT

[0354] CGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAG

[0355] AGACAG

[0356] CTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTA

[0357] AGTACG

[0358] ACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGA

[0359] TTTCCG

[0360] GAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTAC

[0361] CTGAAC

[0362] GCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCG

[0363] ACTACA

[0364] AGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAA

[0365] GTACTT

[0366] CTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGG

[0367] AAGCGG

[0368] CCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCA

[0369] CCGTGC

[0370] GGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTT

[0371] CAGCAA

[0372] AGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCT

[0373] AAGAAG

[0374] TACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGG

[0375] GCAAGT

[0376] CCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGA

[0377] GAAGAA

[0378] TCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTG

[0379] CCTAAG

[0380] TACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGA

[0381] AGGGAA

[0382] ACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCT

[0383] GAAGGG

[0384] CTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAG

[0385] ATCATC

[0386] GAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGT

[0387] CCGCCT

[0388] ACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCT

[0389] GACCAA

[0390] TCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGC

[0391] ACCAAA

[0392] GAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACC

[0393] TGTCTC

[0394] AGCTGGGAGGTGACTCTGGCGGCTCAAAAAGAACCGCCGACGGCAGCGAATTCGAGCCCAAGAA

[0395] GAAGAG

[0396] GAAAGTCTAACCGGTCATCATCACCATCACCATTGAGTTTAAACCCGCTGATCAGCCTCGACTG

[0397] TGCCTT

[0398] CTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCAC

[0399] TCCCAC

[0400] TGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTG

[0401] GGGGGT

[0402] GGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGG

[0403] TGGGCT

[0404] CTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCGATACCGTCGACCTCTAGCTAGAGCTT

[0405] GGCGTA

[0406] ATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGA

[0407] GCCGGA

[0408] AGCATAAAGTGTAAAGCCTAGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCT

[0409] CACTGC

[0410] CCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAG

[0411] AGGCGG

[0412] TTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTG

[0413] CGGCGA

[0414] GCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAA

[0415] AGAACA

[0416] TGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCA

[0417] TAGGCT

[0418] CCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGA

[0419] CTATAA

[0420] AGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTA

[0421] CCGGAT

[0422] ACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCT

[0423] CAGTTC

[0424] GGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGC

[0425] GCCTTA

[0426] TCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCA

[0427] CTGGTA

[0428] ACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTA

[0429] CGGCTA

[0430] CACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTT

[0431] GGTAGC

[0432] TCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTA

[0433] CGCGCA

[0434] GAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACACTCAGTGGAACGA

[0435] AAACTC

[0436] ACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAA

[0437] AAATGA

[0438] AGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCA

[0439] GTGAGG

[0440] CACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGAT

[0441] AACTAC

[0442] GATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCG

[0443] GCTCCA

[0444] GATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTAT

[0445] CCGCCT

[0446] CCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCG

[0447] CAACGT

[0448] TGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCC

[0449] GGTTCC

[0450] CAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTC

[0451] CTCCGA

[0452] TCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTC

[0453] TCTTAC

[0454] TGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAA

[0455] TAGTGT

[0456] ATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAA

[0457] CTTTAA

[0458] AAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAG

[0459] ATCCAG

[0460] TTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCT

[0461] GGGTGA

[0462] GCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATAC

[0463] TCATAC

[0464] TCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATT

[0465] TGAATG

[0466] TATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTC

[0467] GACGGA

[0468] TCGGGAGATCGATCTCCCGATCCCCTAGGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCAT

[0469] AGTTAA

[0470] GCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGC

[0471] TACAAC

[0472] AAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTT

[0473] CGCGAT

[0474] GTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGG

[0475] GGTCAT

[0476] TAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTG

[0477] ACCGCC

[0478] CAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACT

[0479] TTCCAT

[0480] TGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATC

[0481] “碱基编辑活性”是指用于化学改变多核苷酸内的碱基。在一个实施方案中,将第一个碱基转换为第二个碱基。在一个实施方案中,碱基编辑活性是胞苷脱氨酶活性,例如将目标C·G转换为T·A。在另一个实施方案中,碱基编辑活性是腺苷或腺嘌呤脱氨酶活性,例如将A·T转换为G·C。在另一个实施方案中,碱基编辑活性是胞苷脱氨酶活性,例如将目标C·G转换为T·A和腺苷或腺嘌呤脱氨酶活性,例如将A·T转换为G·C。在一些实施方案中,碱基编辑活动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,通过桑格测序或下一代测序。在一些实施方案中,碱基编辑效率通过具有受碱基编辑器影响的核碱基转换的总测序读数的百分比来测量,例如,具有转换为G.C碱基对的目标A.T碱基对的总测序读数的百分比。在一些实施方案中,当在细胞群中进行碱基编辑时,碱基编辑效率通过具有受碱基编辑器实现的核碱基转换的总细胞的百分比来测量。

[0482] 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。在各种实施方案中,碱基编辑器系统包括(1)多核苷酸可编程的核苷酸结合结构域(例如Cas9);(2)用于使所述核碱基脱氨的脱氨酶结构域(例如,腺苷脱氨酶或胞苷脱氨酶);(3)一种或多种向导多核苷酸(例如向导RNA)。在一些实施方案中,多核苷酸可编程的核苷酸结合结构域是多核苷酸可编程的DNA结合结构域。在一些实施方案中,碱基编辑器是腺嘌呤或腺苷碱基编辑器(ABE)。在一些实施方案中,碱基编辑器系统是ABE8。

[0483] 在一些实施方案中,碱基编辑器系统可以包括超过一个碱基编辑组分。例如,碱基编辑器系统可包括超过一种脱氨酶。在一些实施方案中,碱基编辑器系统可包括一种或多种腺苷脱氨酶。在一些实施方案中,可利用单个向导多核苷酸将不同的脱氨酶靶向目标核酸序列。在一些实施方案中,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。

[0484] 碱基编辑器系统的脱氨酶结构域和多核苷酸可编程的核苷酸结合组分可以共价或非共价地彼此连结,或其连结和相互作用的任何组合。例如,在一些实施方案中,脱氨酶结构域可以通过多核苷酸可编程的核苷酸结合结构域靶向目标核苷酸序列。在一些实施方案中,多核苷酸可编程的核苷酸结合结构域可融合或连接至脱氨酶结构域。在一些实施方案中,多核苷酸可编程的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连结将脱氨酶结构域靶向目标核苷酸序列。例如,在一些实施方案中,脱氨酶结构域可包括额外的异源部分或结构域,其能够与作为多核苷酸可编程的核苷酸结合结构域的一部分的额外的异源部分或结构域相互作用、连结或形成复合物。在一些实施方案中,额外的异源部分可能能够与多肽结合、相互作用、连结或形成复合物。在一些实施方案中,额外的异源部分

可能能够与多核苷酸结合、相互作用、连结或形成复合物。在一些实施方案中,额外的异源部分可能能够结合向导多核苷酸。在一些实施方案中,额外的异源部分可能能够结合多肽连接子。在一些实施方案中,额外的异源部分能够结合多核苷酸连接子。额外的异源部分可以是蛋白质结构域。在一些实施方案中,额外的异源部分可以是K同源性(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、不育α基序、端粒酶Ku结合基序和Ku蛋白、端粒酶Sm7结合基序和Sm7蛋白、或RNA识别基序。

[0485] 碱基编辑器系统可进一步包括向导多核苷酸组分。应当理解,碱基编辑器系统的组分可以经由共价键、非共价相互作用或其关联和相互作用的任何组合彼此关联。在一些实施方案中,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,在一些实施方案中,脱氨酶结构域可包括能够与以下物质相互作用、连结或能够与向导多核苷酸的一部分或区段(例如,多核苷酸基序)形成复合物的另外的异源部分或结构域(例如,多核苷酸结合结构域,例如RNA或DNA结合蛋白)。在一些实施方案中,额外的异源部分或结构域(例如,多核苷酸结合结构域,诸如RNA或DNA结合蛋白)可以与脱氨酶结构域融合或连接。在一些实施方案中,额外的异源部分可能能够与多肽结合、相互作用、连结或形成复合物。在一些实施方案中,额外的异源部分可能能够与多核苷酸结合、相互作用、连结或形成复合物。在一些实施方案中,额外的异源部分可能能够结合向导多核苷酸。在一些实施方案中,额外的异源部分可能能够结合多肽连接子。在一些实施方案中,额外的异源部分能够结合多核苷酸连接子。额外的异源部分可以是蛋白质结构域。在一些实施方案中,额外的异源部分可以是K同源性(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、不育α基序、端粒酶Ku结合基序和Ku蛋白、端粒酶Sm7结合基序和Sm7蛋白、或RNA识别基序。

[0486] 在一些实施方案中,碱基编辑器系统可进一步包括碱基切除修复(BER)组分的抑制剂。应当理解,碱基编辑器系统的组分可以经由共价键、非共价相互作用或其关联和相互作用的任何组合彼此关联。BER成分的抑制剂可以包括BER抑制剂。在一些实施方案中,BER抑制剂可以是尿嘧啶DNA糖基化酶抑制剂(UGI)。在一些实施方案中,BER抑制剂可以是肌苷BER抑制剂。在一些实施方案中,BER抑制剂可以通过多核苷酸可编程的核苷酸结合结构域靶向目标核苷酸序列。在一些实施方案中,多核苷酸可编程的核苷酸结合结构域可以与BER抑制剂融合或连接。在一些实施方案中,多核苷酸可编程的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。在一些实施方案中,多核苷酸可编程的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连结来将BER抑制剂靶向目标核苷酸序列。例如,在一些实施方案中,BER组分的抑制剂可包括额外的异源部分或结构域,其能够与作为可编程的多核苷酸结合结构域一部分的额外的异源部分或结构域相互作用、连结或形成复合物。

[0487] 在一些实施方案中,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,在一些实施方案中,BER抑制剂可以包括能够与具有向导多核苷酸的部分或区段(例如,多核苷酸基序)相互作用、连结或能够形成复合物的另外的异源部分或结构域(例如,多核苷酸结合结构域,诸如RNA或DNA结合蛋白)。在一些实施方案中,向导多核苷酸的另外的异源部分或结构域(例如,多核苷酸结合结构域,诸如RNA或DNA结合蛋白)可以与BER抑制剂融合或连接。在一些实施方案中,额外的异源部分可能能够与多核苷酸结合、相互作用、连结

或形成复合物。在一些实施方案中,额外的异源部分可能能够结合向导多核苷酸。在一些实施方案中,额外的异源部分可能能够结合多肽连接子。在一些实施方案中,额外的异源部分能够结合多核苷酸连接子。额外的异源部分可以是蛋白质结构域。在一些实施方案中,额外的异源部分可以是K同源性(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、不育α基序、端粒酶Ku结合基序和Ku蛋白、端粒酶Sm7结合基序和Sm7蛋白、或RNA识别基序。

[0488] 术语“Cas9”或“Cas9结构域”是指包括Cas9蛋白或其片段(例如,包括Cas9、和/或gRNA的活性、失活或部分活性的DNA切割结构域的蛋白)的RNA向导的核酸酶Cas9的结合结构域)。Cas9核酸酶有时也称为Casn1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,其针对移动遗传组件(病毒、转座组件和接合质粒)提供保护。CRISPR簇包括间隔、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类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的向导。随后,Cas9/crRNA/tracrRNA核酸内切方式切割与间隔互补的线性或环状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3'-5'核酸外切方式修剪。在自然界中,DNA结合和切割通常需要蛋白质和两种RNA。然而,可以对单个向导RNA(“sgRNA”,或简称为“gNRA”)进行改造,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.,et al.,Science337:816-821(2012),其全部内容通过引用方式并入本文。Cas9识别CRISPR重复序列(PAM或前间隔邻近基序)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 M1 strain of Streptococcus pyogenes.” Ferretti et al.,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.,et al.,Nature 471:602-607(2011);以及“A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.”Jinek M.,et al.,Science 337:816-821(2012),各个通过引用方式整体并入本文)。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,and Charpentier,“The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems”(2013)RNA Biology 10:5,726-737中公开的生物体和位点;其以引用方式整体并入本文。

[0489] 实施例性的Cas9是化脓链球菌Cas9(spCas9),其氨基酸序列提供如下:

[0490]
MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETAEATRLKRT
ARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPT
IYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQIYNQLFEENPI
NASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSK
DTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLL
KALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS
EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRK
PAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDLLKIIKD
KDFLDNEENEDILEDIVLTLTLFEDRGMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
RDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGHSLHEQIANLAGSPAIKKGI
LQTVKIVDELVKVMGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQ
LQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRM
[0491]
NTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLES
EFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIV
WDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVA
YSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELE
NGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQIS
EFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKE
VLDATLIHQSITGLYETRIDLSQLGGD

[0492] (单底线:HNH结构域;双底线:RuvC域)

[0493] 核酸酶失活的Cas9蛋白可互换地称为“dCas9”蛋白(对于核酸酶而言-“死”的Cas9)或催化性失活的Cas9。用于产生具有失活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 et al.,Science.337:816-821(2012);Qi et al.,“Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression”(2013)Cell.28;152(5):1173-83,各个以引用方式整体并入本文)。例如,已知Cas9的DNA切割域包括两个子域,HNH核酸酶子域和RuvC1子域。HNH子域切割与gRNA互补的链,而RuvC1子域切割非互补链。这些子域内的突变可以使Cas9的核酸酶活性沉默。例如,突变的D10A和H840A完全使化脓链球菌Cas的核酸酶活性失活(Jinek et al.,Science.337:816-821(2012);Qi et al.,Cell.28;152(5):1173-83(2013))。在一些实施方案中,Cas9核酸酶具有失活(例如,失活的)的DNA切割结构域,即,Cas9是切口酶,称为“nCas9”蛋白(对于“切口酶”Cas9)。在一些实施方案中,提供包括Cas9片段的蛋白质。例如,在一些实施方案中,蛋白质包括两个Cas9结构域之一:(1)Cas9的gRNA结合结构域;或(2)Cas9的DNA切割域。在一些实施方案中,包括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变体在与野生型Cas9具有至少约70%相同、至少约80%相同、至少约90%相同、至少约95%相同、至少约96%相同、至少约97%相同、至少约98%相同、至少约99%相同、至少约99.5%相同或至少约99.9%相同。在一些实施方案中,与野生型Cas9相比,Cas9变体可具有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个或更多个的氨基酸变化。在一些实施方案中,Cas9变体包括Cas9的片段(例如,gRNA结合结构域或DNA切割域),使得所述片段至少约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%与野生型Cas9的相应片段相同。在一些实施方案中,所述片段是至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%与对应野生型Cas9氨基酸长度相同。

[0494] 在一些实施方案中,所述片段的长度为至少100个氨基酸。在一些实施方案中,所述片段的长度是至少100、150、200、250、300、350、400、450、500、550、600、650、700、750、800、850、900、1000、1050、1100、1150、1200、1250或至少1300个氨基酸。

[0495] 在一些实施方案中,野生型Cas9对应于来自化脓链球菌的Cas9(NCBI参考序列:NC_017053.1,核苷酸和氨基酸序列如下)。

[0496] ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGATCACTG

[0497] ATGATTATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTATCAAAAA

[0498] AAATCTTATAGGGGCTCTTTTATTTGGCAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACA

[0499] GCTCGTAGAAGGTATACACGTCGGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGA

[0500] TGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGTCTTTTTTGGTGGAAGAAGACAAGAA

[0501] GCATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAAGTTGCTTATCATGAGAAATATCCAACT

[0502] ATCTATCATCTGCGAAAAAAATTGGCAGATTCTACTGATAAAGCGGATTTGCGCTTAATCTATTTGG

[0503] CCTTAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAGATTTAAATCCTGATAATAG

[0504] TGATGTGGACAAACTATTTATCCAGTTGGTACAAATCTACAATCAATTATTTGAAGAAAACCCTATT

[0505] AACGCAAGTAGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAA

[0506] ATCTCATTGCTCAGCTCCCCGGTGAGAAGAGAAATGGCTTGTTTGGGAATCTCATTGCTTTGTCATT

[0507] GGGATTGACCCCTAATTTTAAATCAAATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAA

[0508] GATACTTACGATGATGATTTAGATAATTTATTGGCGCAAATTGGAGATCAATATGCTGATTTGTTTT

[0509] TGGCAGCTAAGAATTTATCAGATGCTATTTTACTTTCAGATATCCTAAGAGTAAATAGTGAAATAAC

[0510] TAAGGCTCCCCTATCAGCTTCAATGATTAAGCGCTACGATGAACATCATCAAGACTTGACTCTTTTA

[0511] AAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTATAAAGAAATCTTTTTTGATCAATCAAAAAACG

[0512] GATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAAATTTATCAAACCAATTTT

[0513] AGAAAAAATGGATGGTACTGAGGAATTATTGGTGAAACTAAATCGTGAAGATTTGCTGCGCAAGCAA

[0514] CGGACCTTTGACAACGGCTCTATTCCCCATCAAATTCACTTGGGTGAGCTGCATGCTATTTTGAGAA

[0515] GACAAGAAGACTTTTATCCATTTTTAAAAGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCG

[0516] AATTCCTTATTATGTTGGTCCATTGGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCT

[0517] GAAGAAACAATTACCCCATGGAATTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTTA

[0518] TTGAACGCATGACAAACTTTGATAAAAATCTTCCAAATGAAAAAGTACTACCAAAACATAGTTTGCT

[0519] TTATGAGTATTTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTGAGGGAATGCGAAAA

[0520] CCAGCATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAACAAATCGAAAAG

[0521] TAACCGTTAAGCAATTAAAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGAAATTTC

[0522] AGGAGTTGAAGATAGATTTAATGCTTCATTAGGCGCCTACCATGATTTGCTAAAAATTATTAAAGAT

[0523] AAAGATTTTTTGGATAATGAAGAAAATGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTAT

[0524] TTGAAGATAGGGGGATGATTGAGGAAAGACTTAAAACATATGCTCACCTCTTTGATGATAAGGTGAT

[0525] GAAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTTTGTCTCGAAAATTGATTAATGGTATT

[0526] AGGGATAAGCAATCTGGCAAAACAATATTAGATTTTTTGAAATCAGATGGTTTTGCCAATCGCAATT

[0527] TTATGCAGCTGATCCATGATGATAGTTTGACATTTAAAGAAGATATTCAAAAAGCACAGGTGTCTGG

[0528] ACAAGGCCATAGTTTACATGAACAGATTGCTAACTTAGCTGGCAGTCCTGCTATTAAAAAAGGTATT

[0529] TTACAGACTGTAAAAATTGTTGATGAACTGGTCAAAGTAATGGGGCATAAGCCAGAAAATATCGTTA

[0530] TTGAAATGGCACGTGAAAATCAGACAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACG

[0531] AATCGAAGAAGGTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAATACTCAA

[0532] TTGCAAAATGAAAAGCTCTATCTCTATTATCTACAAAATGGAAGAGACATGTATGTGGACCAAGAAT

[0533] TAGATATTAATCGTTTAAGTGATTATGATGTCGATCACATTGTTCCACAAAGTTTCATTAAAGACGA

[0534] TTCAATAGACAATAAGGTACTAACGCGTTCTGATAAAAATCGTGGTAAATCGGATAACGTTCCAAGT

[0535] GAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAACTTCTAAACGCCAAGTTAATCACTCAAC

[0536] GTAAGTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTGAACTTGATAAAGCTGGTTTTAT

[0537] CAAACGCCAATTGGTTGAAACTCGCCAAATCACTAAGCATGTGGCACAAATTTTGGATAGTCGCATG

[0538] AATACTAAATACGATGAAAATGATAAACTTATTCGAGAGGTTAAAGTGATTACCTTAAAATCTAAAT

[0539] TAGTTTCTGACTTCCGAAAAGATTTCCAATTCTATAAAGTACGTGAGATTAACAATTACCATCATGC

[0540] CCATGATGCGTATCTAAATGCCGTCGTTGGAACTGCTTTGATTAAGAAATATCCAAAACTTGAATCG

[0541] GAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTCGTAAAATGATTGCTAAGTCTGAGCAAGAAA

[0542] TAGGCAAAGCAACCGCAAAATATTTCTTTTACTCTAATATCATGAACTTCTTCAAAACAGAAATTAC

[0543] ACTTGCAAATGGAGAGATTCGCAAACGCCCTCTAATCGAAACTAATGGGGAAACTGGAGAAATTGTC

[0544] TGGGATAAAGGGCGAGATTTTGCCACAGTGCGCAAAGTATTGTCCATGCCCCAAGTCAATATTGTCA

[0545] AGAAAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAATTTTACCAAAAAGAAATTCGGACAA

[0546] GCTTATTGCTCGTAAAAAAGACTGGGATCCAAAAAAATATGGTGGTTTTGATAGTCCAACGGTAGCT

[0547] TATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGAAATCGAAGAAGTTAAAATCCGTTAAAGAGT

[0548] TACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAAATCCGATTGACTTTTTAGAAGCTAA

[0549] AGGATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACCTAAATATAGTCTTTTTGAGTTAGAA

[0550] AACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAATTACAAAAAGGAAATGAGCTGGCTCTGCCAA

[0551] GCAAATATGTGAATTTTTTATATTTAGCTAGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAA

[0552] CGAACAAAAACAATTGTTTGTGGAGCAGCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGT

[0553] GAATTTTCTAAGCGTGTTATTTTAGCAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAAC

[0554] ATAGAGACAAACCAATACGTGAACAAGCAGAAAATATTATTCATTTATTTACGTTGACGAATCTTGG

[0555] AGCTCCCGCTGCTTTTAAATATTTTGATACAACAATTGATCGTAAACGATATACGTCTACAAAAGAA

[0556] GTTTTAGATGCCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGTC

[0557] AGCTAGGAGGTGACTGA

[0558]
MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD
NSDVDKLFIQLVQIYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNG
LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKN
LSDAILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ
SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQ
IHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET
ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNAS
LGAYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRGMIEERLKTYAHLFDDKVMK
QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDI
QKAQVSGQGHSLHEQIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQ
TTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQE
LDINRLSDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQL
LNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDE
NDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKL
ESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLI
ETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIAR
KKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDF
LEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS
HYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRD
KPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYET
RIDLSQLGGD

[0559] (单底线:HNH结构域;双底线:RuvC域)

[0560] 在一些实施方案中,野生型Cas9对应于或包括以下核苷酸和/或氨基酸序列:

[0561] ATGGATAAAAAGTATTCTATTGGTTTAGACATCGGCACTAATTCCGTTGGATGGGCTGT

[0562] CATAACCGATGAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACC

[0563] GTCATTCGATTAAAAAGAATCTTATCGGTGCCCTCCTATTCGATAGTGGCGAAACGGCA

[0564] GAGGCGACTCGCCTGAAACGAACCGCTCGGAGAAGGTATACACGTCGCAAGAACCGAAT

[0565] ATGTTACTTACAAGAAATTTTTAGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTTC

[0566] ACCGTTTGGAAGAGTCCTTCCTTGTCGAAGAGGACAAGAAACATGAACGGCACCCCATC

[0567] TTTGGAAACATAGTAGATGAGGTGGCATATCATGAAAAGTACCCAACGATTTATCACCT

[0568] CAGAAAAAAGCTAGTTGACTCAACTGATAAAGCGGACCTGAGGTTAATCTACTTGGCTC

[0569] TTGCCCATATGATAAAGTTCCGTGGGCACTTTCTCATTGAGGGTGATCTAAATCCGGAC

[0570] AACTCGGATGTCGACAAACTGTTCATCCAGTTAGTACAAACCTATAATCAGTTGTTTGA

[0571] AGAGAACCCTATAAATGCAAGTGGCGTGGATGCGAAGGCTATTCTTAGCGCCCGCCTCT

[0572] CTAAATCCCGACGGCTAGAAAACCTGATCGCACAATTACCCGGAGAGAAGAAAAATGGG

[0573] TTGTTCGGTAACCTTATAGCGCTCTCACTAGGCCTGACACCAAATTTTAAGTCGAACTT

[0574] CGACTTAGCTGAAGATGCCAAATTGCAGCTTAGTAAGGACACGTACGATGACGATCTCG

[0575] ACAATCTACTGGCACAAATTGGAGATCAGTATGCGGACTTATTTTTGGCTGCCAAAAAC

[0576] CTTAGCGATGCAATCCTCCTATCTGACATACTGAGAGTTAATACTGAGATTACCAAGGC

[0577] GCCGTTATCCGCTTCAATGATCAAAAGGTACGATGAACATCACCAAGACTTGACACTTC

[0578] TCAAGGCCCTAGTCCGTCAGCAACTGCCTGAGAAATATAAGGAAATATTCTTTGATCAG

[0579] TCGAAAAACGGGTACGCAGGTTATATTGACGGCGGAGCGAGTCAAGAGGAATTCTACAA

[0580] GTTTATCAAACCCATATTAGAGAAGATGGATGGGACGGAAGAGTTGCTTGTAAAACTCA

[0581] ATCGCGAAGATCTACTGCGAAAGCAGCGGACTTTCGACAACGGTAGCATTCCACATCAA

[0582] ATCCACTTAGGCGAATTGCATGCTATACTTAGAAGGCAGGAGGATTTTTATCCGTTCCT

[0583] CAAAGACAATCGTGAAAAGATTGAGAAAATCCTAACCTTTCGCATACCTTACTATGTGG

[0584] GACCCCTGGCCCGAGGGAACTCTCGGTTCGCATGGATGACAAGAAAGTCCGAAGAAACG

[0585] ATTACTCCATGGAATTTTGAGGAAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTTCAT

[0586] CGAGAGGATGACCAACTTTGACAAGAATTTACCGAACGAAAAAGTATTGCCTAAGCACA

[0587] GTTTACTTTACGAGTATTTCACAGTGTACAATGAACTCACGAAAGTTAAGTATGTCACT

[0588] GAGGGCATGCGTAAACCCGCCTTTCTAAGCGGAGAACAGAAGAAAGCAATAGTAGATCT

[0589] GTTATTCAAGACCAACCGCAAAGTGACAGTTAAGCAATTGAAAGAGGACTACTTTAAGA

[0590] AAATTGAATGCTTCGATTCTGTCGAGATCTCCGGGGTAGAAGATCGATTTAATGCGTCA

[0591] CTTGGTACGTATCATGACCTCCTAAAGATAATTAAAGATAAGGACTTCCTGGATAACGA

[0592] AGAGAATGAAGATATCTTAGAAGATATAGTGTTGACTCTTACCCTCTTTGAAGATCGGG

[0593] AAATGATTGAGGAAAGACTAAAAACATACGCTCACCTGTTCGACGATAAGGTTATGAAA

[0594] CAGTTAAAGAGGCGTCGCTATACGGGCTGGGGACGATTGTCGCGGAAACTTATCAACGG

[0595] GATAAGAGACAAGCAAAGTGGTAAAACTATTCTCGATTTTCTAAAGAGCGACGGCTTCG

[0596] CCAATAGGAACTTTATGCAGCTGATCCATGATGACTCTTTAACCTTCAAAGAGGATATA

[0597] CAAAAGGCACAGGTTTCCGGACAAGGGGACTCATTGCACGAACATATTGCGAATCTTGC

[0598] TGGTTCGCCAGCCATCAAAAAGGGCATACTCCAGACAGTCAAAGTAGTGGATGAGCTAG

[0599] TTAAGGTCATGGGACGTCACAAACCGGAAAACATTGTAATCGAGATGGCACGCGAAAAT

[0600] CAAACGACTCAGAAGGGGCAAAAAAACAGTCGAGAGCGGATGAAGAGAATAGAAGAGGG

[0601] TATTAAAGAACTGGGCAGCCAGATCTTAAAGGAGCATCCTGTGGAAAATACCCAATTGC

[0602] AGAACGAGAAACTTTACCTCTATTACCTACAAAATGGAAGGGACATGTATGTTGATCAG

[0603] GAACTGGACATAAACCGTTTATCTGATTACGACGTCGATCACATTGTACCCCAATCCTT

[0604] TTTGAAGGACGATTCAATCGACAATAAAGTGCTTACACGCTCGGATAAGAACCGAGGGA

[0605] AAAGTGACAATGTTCCAAGCGAGGAAGTCGTAAAGAAAATGAAGAACTATTGGCGGCAG

[0606] CTCCTAAATGCGAAACTGATAACGCAAAGAAAGTTCGATAACTTAACTAAAGCTGAGAG

[0607] GGGTGGCTTGTCTGAACTTGACAAGGCCGGATTTATTAAACGTCAGCTCGTGGAAACCC

[0608] GCCAAATCACAAAGCATGTTGCACAGATACTAGATTCCCGAATGAATACGAAATACGAC

[0609] GAGAACGATAAGCTGATTCGGGAAGTCAAAGTAATCACTTTAAAGTCAAAATTGGTGTC

[0610] GGACTTCAGAAAGGATTTTCAATTCTATAAAGTTAGGGAGATAAATAACTACCACCATG

[0611] CGCACGACGCTTATCTTAATGCCGTCGTAGGGACCGCACTCATTAAGAAATACCCGAAG

[0612] CTAGAAAGTGAGTTTGTGTATGGTGATTACAAAGTTTATGACGTCCGTAAGATGATCGC

[0613] GAAAAGCGAACAGGAGATAGGCAAGGCTACAGCCAAATACTTCTTTTATTCTAACATTA

[0614] TGAATTTCTTTAAGACGGAAATCACTCTGGCAAACGGAGAGATACGCAAACGACCTTTA

[0615] ATTGAAACCAATGGGGAGACAGGTGAAATCGTATGGGATAAGGGCCGGGACTTCGCGAC

[0616] GGTGAGAAAAGTTTTGTCCATGCCCCAAGTCAACATAGTAAAGAAAACTGAGGTGCAGA

[0617] CCGGAGGGTTTTCAAAGGAATCGATTCTTCCAAAAAGGAATAGTGATAAGCTCATCGCT

[0618] CGTAAAAAGGACTGGGACCCGAAAAAGTACGGTGGCTTCGATAGCCCTACAGTTGCCTA

[0619] TTCTGTCCTAGTAGTGGCAAAAGTTGAGAAGGGAAAATCCAAGAAACTGAAGTCAGTCA

[0620] AAGAATTATTGGGGATAACGATTATGGAGCGCTCGTCTTTTGAAAAGAACCCCATCGAC

[0621] TTCCTTGAGGCGAAAGGTTACAAGGAAGTAAAAAAGGATCTCATAATTAAACTACCAAA

[0622] GTATAGTCTGTTTGAGTTAGAAAATGGCCGAAAACGGATGTTGGCTAGCGCCGGAGAGC

[0623] TTCAAAAGGGGAACGAACTCGCACTACCGTCTAAATACGTGAATTTCCTGTATTTAGCG

[0624] TCCCATTACGAGAAGTTGAAAGGTTCACCTGAAGATAACGAACAGAAGCAACTTTTTGT

[0625] TGAGCAGCACAAACATTATCTCGACGAAATCATAGAGCAAATTTCGGAATTCAGTAAGA

[0626] GAGTCATCCTAGCTGATGCCAATCTGGACAAAGTATTAAGCGCATACAACAAGCACAGG

[0627] GATAAACCCATACGTGAGCAGGCGGAAAATATTATCCATTTGTTTACTCTTACCAACCT

[0628] CGGCGCTCCAGCCGCATTCAAGTATTTTGACACAACGATAGATCGCAAACGATACACTT

[0629] CTACCAAGGAGGTGCTAGACGCGACACTGATTCACCAATCCATCACGGGATTATATGAA

[0630] ACTCGGATAGATTTGTCACAGCTTGGGGGTGACGGATCCCCCAAGAAGAAGAGGAAAGT

[0631] CTCGAGCGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGG

[0632] ATGACGATGACAAGGCTGCAGGA

[0633]
MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD
NSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG
LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKN
LSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ
SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQ
IHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET
ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNAS
LGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMK
QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDI
QKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMAREN
QTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQ
LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD
ENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK
LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPL
IETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA
RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLA
SHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR
DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYE
TRIDLSQLGGD

[0634] (单底线:HNH结构域;双底线:RuvC域)

[0635] 在一些实施方案中,野生型Cas9对应于来自化脓链球菌的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)。

[0636] ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGT

[0637] GATCACTGATGAATATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACC

[0638] GCCACAGTATCAAAAAAAATCTTATAGGGGCTCTTTTATTTGACAGTGGAGAGACAGCG

[0639] GAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTAT

[0640] TTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTC

[0641] ATCGACTTGAAGAGTCTTTTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATT

[0642] TTTGGAAATATAGTAGATGAAGTTGCTTATCATGAGAAATATCCAACTATCTATCATCT

[0643] GCGAAAAAAATTGGTAGATTCTACTGATAAAGCGGATTTGCGCTTAATCTATTTGGCCT

[0644] TAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAGATTTAAATCCTGAT

[0645] AATAGTGATGTGGACAAACTATTTATCCAGTTGGTACAAACCTACAATCAATTATTTGA

[0646] AGAAAACCCTATTAACGCAAGTGGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGA

[0647] GTAAATCAAGACGATTAGAAAATCTCATTGCTCAGCTCCCCGGTGAGAAGAAAAATGGC

[0648] TTATTTGGGAATCTCATTGCTTTGTCATTGGGTTTGACCCCTAATTTTAAATCAAATTT

[0649] TGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTAG

[0650] ATAATTTATTGGCGCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAAT

[0651] TTATCAGATGCTATTTTACTTTCAGATATCCTAAGAGTAAATACTGAAATAACTAAGGC

[0652] TCCCCTATCAGCTTCAATGATTAAACGCTACGATGAACATCATCAAGACTTGACTCTTT

[0653] TAAAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTATAAAGAAATCTTTTTTGATCAA

[0654] TCAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAA

[0655] ATTTATCAAACCAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTGGTGAAACTAA

[0656] ATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACAACGGCTCTATTCCCCATCAA

[0657] ATTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTT

[0658] AAAAGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTG

[0659] GTCCATTGGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACA

[0660] ATTACCCCATGGAATTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTTAT

[0661] TGAACGCATGACAAACTTTGATAAAAATCTTCCAAATGAAAAAGTACTACCAAAACATA

[0662] GTTTGCTTTATGAGTATTTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACT

[0663] GAAGGAATGCGAAAACCAGCATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTT

[0664] ACTCTTCAAAACAAATCGAAAAGTAACCGTTAAGCAATTAAAAGAAGATTATTTCAAAA

[0665] AAATAGAATGTTTTGATAGTGTTGAAATTTCAGGAGTTGAAGATAGATTTAATGCTTCA

[0666] TTAGGTACCTACCATGATTTGCTAAAAATTATTAAAGATAAAGATTTTTTGGATAATGA

[0667] AGAAAATGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAAGATAGGG

[0668] AGATGATTGAGGAAAGACTTAAAACATATGCTCACCTCTTTGATGATAAGGTGATGAAA

[0669] CAGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTTTGTCTCGAAAATTGATTAATGG

[0670] TATTAGGGATAAGCAATCTGGCAAAACAATATTAGATTTTTTGAAATCAGATGGTTTTG

[0671] CCAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGACATTTAAAGAAGACATT

[0672] CAAAAAGCACAAGTGTCTGGACAAGGCGATAGTTTACATGAACATATTGCAAATTTAGC

[0673] TGGTAGCCCTGCTATTAAAAAAGGTATTTTACAGACTGTAAAAGTTGTTGATGAATTGG

[0674] TCAAAGTAATGGGGCGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACGTGAAAAT

[0675] CAGACAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAATCGAAGAAGG

[0676] TATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAATACTCAATTGC

[0677] AAAATGAAAAGCTCTATCTCTATTATCTCCAAAATGGAAGAGACATGTATGTGGACCAA

[0678] GAATTAGATATTAATCGTTTAAGTGATTATGATGTCGATCACATTGTTCCACAAAGTTT

[0679] CCTTAAAGACGATTCAATAGACAATAAGGTCTTAACGCGTTCTGATAAAAATCGTGGTA

[0680] AATCGGATAACGTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAA

[0681] CTTCTAAACGCCAAGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACG

[0682] TGGAGGTTTGAGTGAACTTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTC

[0683] GCCAAATCACTAAGCATGTGGCACAAATTTTGGATAGTCGCATGAATACTAAATACGAT

[0684] GAAAATGATAAACTTATTCGAGAGGTTAAAGTGATTACCTTAAAATCTAAATTAGTTTC

[0685] TGACTTCCGAAAAGATTTCCAATTCTATAAAGTACGTGAGATTAACAATTACCATCATG

[0686] CCCATGATGCGTATCTAAATGCCGTCGTTGGAACTGCTTTGATTAAGAAATATCCAAAA

[0687] CTTGAATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTCGTAAAATGATTGC

[0688] TAAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAAAATATTTCTTTTACTCTAATATCA

[0689] TGAACTTCTTCAAAACAGAAATTACACTTGCAAATGGAGAGATTCGCAAACGCCCTCTA

[0690] ATCGAAACTAATGGGGAAACTGGAGAAATTGTCTGGGATAAAGGGCGAGATTTTGCCAC

[0691] AGTGCGCAAAGTATTGTCCATGCCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAGA

[0692] CAGGCGGATTCTCCAAGGAGTCAATTTTACCAAAAAGAAATTCGGACAAGCTTATTGCT

[0693] CGTAAAAAAGACTGGGATCCAAAAAAATATGGTGGTTTTGATAGTCCAACGGTAGCTTA

[0694] TTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGAAATCGAAGAAGTTAAAATCCGTTA

[0695] AAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAAATCCGATTGAC

[0696] TTTTTAGAAGCTAAAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACCTAA

[0697] ATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAAT

[0698] TACAAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATTTAGCT

[0699] AGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTTGT

[0700] GGAGCAGCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGC

[0701] GTGTTATTTTAGCAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGA

[0702] GACAAACCAATACGTGAACAAGCAGAAAATATTATTCATTTATTTACGTTGACGAATCT

[0703] TGGAGCTCCCGCTGCTTTTAAATATTTTGATACAACAATTGATCGTAAACGATATACGT

[0704] CTACAAAAGAAGTTTTAGATGCCACTCTTATCCATCAATCCATCACTGGTCTTTATGAA

[0705] ACACGCATTGATTTGAGTCAGCTAGGAGGTGACTGA

[0706] MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD
NSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG
LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKN
LSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ
SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQ
IHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET
ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNAS
LGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMK
QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDI
QKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMAREN
QTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQ
LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD
ENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK
LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPL
IETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA
RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLA
SHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR
DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYE
TRIDLSQLGGD (SEQ ID NO：1)

[0707] (单底线:HNH结构域;双底线:RuvC域)

[0708] 在一些实施方案中,Cas9是指来自:溃疡棒状杆菌(Corynebacterium ulcerans)(NCBI参考序列:NC_015683.1、NC_017317.1)的Cas9;白喉棒状杆菌(Corynebacterium diphtheria)(NCBI参考序列:NC_016782.1、NC_016786.1);食蚜蝇(Spiroplasma syrphidicola)(NCBI参考序列:NC_021284.1);中间普氏菌(Prevotella intermedia)(NCBI参考序列:NC_017861.1);Spiroplasma taiwanense(NCBI参考序列:NC_021846.1);鱼型链球菌(Streptococcus iniae)(NCBI参考序列:NC_021314.1);波罗海贝尔氏菌(Belliella baltica)(NCBI参考序列:NC_018010.1);冷弯曲菌(Psychroflexus Torquisi)(NCBI参考序列:NC_018721.1);嗜热链球菌(NCBI参考序列:YP_820832.1)、无害李斯特菌(Listeria innocua)(NCBI参考序列:NP_472073.1)、空肠弯曲菌(Campylobacter jejuni)(NCBI参考序列:YP_002344900.1)或脑膜炎奈瑟菌(Neisseria meningitidis)(NCBI参考序列:YP_002342100.1)或来自任何其他生物体的Cas9。

[0709] 在一些实施方案中,dCas9对应于或部分或全部包括具有一个或多个使Cas9核酸酶活性失活的突变的Cas9氨基酸序列。例如,在一些实施方案中,dCas9结构域包括在SEQ ID NO:1中编号为D10A和H840A突变或另一个Cas9中的相应的突变。在一些实施方案中,dCas9包括dCas9(D10A和H840A)的氨基酸序列:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD
NSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG
LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKN
LSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ
SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQ
IHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET
ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT

[0710] EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNAS
LGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMK
QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDI
QKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMAREN
QTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQ
LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD
ENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK
LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPL
IETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA

[0711] RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLA
SHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR
DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYE
TRIDLSQLGGD

[0712] (单底线:HNH结构域;双底线:RuvC域)。

[0713] 在一些实施方案中,Cas9结构域包括D10A突变,而位置840的残基在上文提供的氨基酸序列中或在本文所提供的任何氨基酸序列中的相应位置仍为组氨酸。

[0714] 在其他实施方案中,提供具有除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处的其他氨基酸取代,或Cas9核酸酶结构域内的其他取代(例如,HNH核酸酶子域和/或RuvC1子域中的取代)。在一些实施方案中,提供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。在一些实施方案中,提供具有短或长约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。

[0715] 在一些实施方案中,本文所提供的Cas9融合蛋白包括Cas9蛋白的全长氨基酸序列,例如本文所提供的Cas9序列之一。然而,在其他实施方案中,本文所提供的融合蛋白不包括全长Cas9序列,而仅包括其一个或多个片段。本文提供合适的Cas9结构域和Cas9片段的实施例性氨基酸序列,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。

[0716] 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,是在本公开的范围内。实施例性的Cas9蛋白包括但不限于那些以下提供者。在一些实施方案中,Cas9蛋白是核酸酶死亡的Cas9(dCas9)。在一些实施方案中,Cas9蛋白是Cas9切口酶(nCas9)。在一些实施方案中,Cas9蛋白是核酸酶活性的Cas9。

[0717] 实施例性催化失活的Cas9(dCas9):

[0718] DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA

[0719] RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI

[0720] YHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPIN

[0721] ASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKD

[0722] TYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLK

[0723] ALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQR

[0724] TFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSE

[0725] ETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKP

[0726] AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK

[0727] DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIR

[0728] DKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGIL

[0729] QTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQ

[0730] LQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS

[0731] EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRM

[0732] NTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLES

[0733] EFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIV

[0734] WDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVA

[0735] YSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELE

[0736] NGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQIS

[0737] EFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKE

[0738] VLDATLIHQSITGLYETRIDLSQLGGD

[0739] 实施例性催化的Cas9切口酶(nCas9):

[0740] DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA

[0741] RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI

[0742] YHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPIN

[0743] ASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKD

[0744] TYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLK

[0745] ALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQR

[0746] TFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSE

[0747] ETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKP

[0748] AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK

[0749] DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIR

[0750] DKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGIL

[0751] QTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQ

[0752] LQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS

[0753] EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRM

[0754] NTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLES

[0755] EFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIV

[0756] WDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVA

[0757] YSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELE

[0758] NGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQIS

[0759] EFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKE

[0760] VLDATLIHQSITGLYETRIDLSQLGGD

[0761] 实施例性催化活性的Cas9：

[0762] DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA

[0763] RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI

[0764] YHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPIN

[0765] ASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKD

[0766] TYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLK

[0767] ALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQR

[0768] TFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSE

[0769] ETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKP

[0770] AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK

[0771] DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIR

[0772] DKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGIL

[0773] QTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQ

[0774] LQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS

[0775] EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRM

[0776] NTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLES

[0777] EFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIV

[0778] WDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVA

[0779] YSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELE

[0780] NGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQIS

[0781] EFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEV

LDATLIHQSITGLYETRIDLSQLGGD.

[0782] 在一些实施方案中,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。在一些实施方案中,Cas9是指CasX或CasY,其已描述在例如Burstein et al.,"New CRISPR-Cas systems from uncultivated microbes."Cell Res.2017Feb 21.doi:10.1038/cr.2017.21,其特此以引入方式整体并入。使用基因组解析的宏基因组学,识别了许多CRISPR-Cas系统,包括在古细菌领域首次报道的Cas9。这种发散的Cas9蛋白在很少被研究的纳米古细菌中发现,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。在一些实施方案中,Cas9是指CasX或CasX的变体。在一些实施方案中,Cas9是指CasY或CasY的变体。应当理解,其他RNA向导的DNA结合蛋白可以用作核酸可编程的DNA结合蛋白(napDNAbp),并且在本公开的范围内。

[0783] 在特定实施方案中,可用于本发明的方法中的napDNAbp包括本领域已知的并且例如由Oakes et al.,Cell 176,254-267,2019描述的环状排列物。实施例性环状排列物如下粗体序列表示衍生自Cas9的序列,斜体序列表示连接子序列,底线序列表示二分核定位序列,CP5(具有MSP"NGC＝具有突变的Pam变体常规Cas9如同NGG"PID＝蛋白质相互作用结构域和"D10A"切口酶):

[0784] **EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNI**
VKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPTVAYSVLVVAKVEKGKSKKLKSVK
ELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAKFLQKGNELAL
PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYN
KHRDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYRSTKEVLDATLIHQSITGLYETRIDL
SQLGGD*GGSGGSGGSGGSGGSGGSGGM***DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRH**
SIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVE
EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLI
ALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVN
TEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFI
KPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKI
LTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPK
HSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS
VEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKA
QVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSR
ERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQS
FLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELD
KAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREIN
NYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEGADKRTADGSEFESPKKK
RKV*

[0785] 可并入碱基编辑器的多核苷酸可编程的核苷酸结合结构域的非限制性实例包括

衍生自CRISPR蛋白结构域、限制性核酸酶、大范围核酸酶、TAL核酸酶(TALEN)和锌指核酸酶(ZFN)。

[0786] 在一些实施方案中,本文所提供的任何融合蛋白的核酸可编程的DNA结合蛋白(napDNAbp)可以是CasX或CasY蛋白。在一些实施方案中,napDNAbp是CasX蛋白。在一些实施方案中,napDNAbp是CasY蛋白。在一些实施方案中,napDNAbp包括的氨基酸序列为与天然存在的CasX或CasY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%蛋白相同。在一些实施方案中,napDNAbp是天然存在的CasX或CasY蛋白。在一些实施方案中,napDNAbp包括的氨基酸序列为与本文所述任何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。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、CasX和CasY。

[0787] Cas12b/C2c1(uniprot.org/uniprot/T0D7A2#2)

[0788] sp|T0D7A2|C2C1_ALIAG CRISPR相关的内切核酸酶C2c1 OS=酸土脂环酸芽孢杆菌(Alicyclobacillus acido-terrestris)(菌株ATCC 49025/DSM 3922/CIP 106132/NCIMB 13137/GD3B)GN=c2c1PE=1SV=1

[0789] MAVKSIKVKLRLDDMPEIRAGLWKLHKEVNAGVRYYTEWLSLLRQENLYRRSPNGDGEQECDKTAEE

[0790] CKAELLERLRARQVENGHRGPAGSDDELLQLARQLYELLVPQAIGAKGDAQQIARKFLSPLADKDAV

[0791] GGLGIAKAGNKPRWVRMREAGEPGWEEEKEKAETRKSADRTADVLRALADFGLKPLMRVYTDSEMSS

[0792] VEWKPLRKGQAVRTWDRDMFQQAIERMMSWESWNQRVGQEYAKLVEQKNRFEQKNFVGQEHLVHLVN

[0793] QLQQDMKEASPGLESKEQTAHYVTGRALRGSDKVFEKWGKLAPDAPFDLYDAEIKNVQRRNTRRFGS

[0794] HDLFAKLAEPEYQALWREDASFLTRYAVYNSILRKLNHAKMFATFTLPDATAHPIWTRFDKLGGNLH

[0795] QYTFLFNEFGERRHAIRFHKLLKVENGVAREVDDVTVPISMSEQLDNLLPRDPNEPIALYFRDYGAE

[0796] QHFTGEFGGAKIQCRRDQLAHMHRRRGARDVYLNVSVRVQSQSEARGERRPPYAAVFRLVGDNHRAF

[0797] VHFDKLSDYLAEHPDDGKLGSEGLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNSKGRVPFFFP

[0798] IKGNDNLVAVHERSQLLKLPGETESKDLRAIREERQRTLRQLRTQLAYLRLLVRCGSEDVGRRERSW

[0799] AKLIEQPVDAANHMTPDWREAFENELQKLKSLHGICSDKEWMDAVYESVRRVWRHMGKQVRDWRKDV

[0800] RSGERPKIRGYAKDVVGGNSIEQIEYLERQYKFLKSWSFFGKVSGQVIRAEKGSRFAITLREHIDHA

[0801] KEDRLKKLADRIIMEALGYVYALDERGKGKWVAKYPPCQLILLEELSEYQFNNDRPPSENNQLMQWS

[0802] HRGVFQELINQAQVHDLLVGTMYAAFSSRFDARTGAPGIRCRRVPARCTQEHNPEPFPWWLNKFVVE

[0803] HTLDACPLRADDLIPTGEGEIFVSPFSAEEGDFHQIHADLNAAQNLQQRLWSDFDISQIRLRCDWGE

[0804] VDGELVLIPRLTGKRTADSYSNKVFYTNTGVTYYERERGKKRRKVFAQEKLSEEEAELLVEADEARE

[0805] KSVVLMRDPSGIINRGNWTRQKEFWSMV NQRIEGYLVKQIRSRVPLQDSACENTGDI

[0806] CasX(uniprot.org/uniprot/F0NN87;uniprot.org/uniprot/F0NH53)

[0807] >tr|F0NN87|F0NN87_SULIH CRISPR相关的Casx蛋白OS=冰岛硫化叶菌(Sulfolobus islandicus)(菌株HVE10/4)GN=SiH_0402PE=4SV=1

[0808] MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAERRGKAKKKK

[0809] GEEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKECEEVSAPSFVKPEFYEF

[0810] GRSPGMVERTRRVKLEVEPHYLIIAAAGWVLTRLGKAKVSEGDYVGVNVFTPTRGILYSLIQNVNGI

[0811] VPGIKPETAFGLWIARKVVSSVTNPNVSVVRIYTISDAVGQNPTTINGGFSIDLTKLLEKRYLLSER

[0812] LEAIARNALSISSNMRERYIVLANYIYEYLTG SKRLEDLLYFANRDLIMNLNSDDGKVRDLKLISA

[0813] YVNGELIRGEG

[0814] >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

[0815] MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAERRGKAKKKK

[0816] GEEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKECEEVSAPSFVKPEFYKF

[0817] GRSPGMVERTRRVKLEVEPHYLIMAAAGWVLTRLGKAKVSEGDYVGVNVFTPTRGILYSLIQNVNGI

[0818] VPGIKPETAFGLWIARKVVSSVTNPNVSVVSIYTISDAVGQNPTTINGGFSIDLTKLLEKRDLLSER

[0819] LEAIARNALSISSNMRERYIVLANYIYEYLTGSKRLEDLLYFANRDLIMNLNSDDGKVRDLKLISAY

[0820] VNGELIRGEG

[0821] 变形杆菌CasX

[0822] MEKRINKIRKKLSADNATKPVSRSGPMKTLLVRVMTDDLKKRLEKRRKKPEVMPQVISNNAANNLRM

[0823] LLDDYTKMKEAILQVYWQEFKDDHVGLMCKFAQPASKKIDQNKLKPEMDEKGNLTTAGFACSQCGQP

[0824] LFVYKLEQVSEKGKAYTNYFGRCNVAEHEKLILLAQLKPVKDSDEAVTYSLGKFGQRALDFYSIHVT

[0825] KESTHPVKPLAQIAGNRYASGPVGKALSDACMGTIASFLSKYQDIIIEHQKVVKGNQKRLESLRELA

[0826] GKENLEYPSVTLPPQPHTKEGVDAYNEVIARVRMWVNLNLWQKLKLSRDDAKPLLRLKGFPSFPVVE

[0827] RRENEVDWWNTINEVKKLIDAKRDMGRVFWSGVTAEKRNTILEGYNYLPNENDHKKREGSLENPKKP

[0828] AKRQFGDLLLYLEKKYAGDWGKVFDEAWERIDKKIAGLTSHIEREEARNAEDAQSKAVLTDWLRAKA

[0829] SFVLERLKEMDEKEFYACEIQLQKWYGDLRGNPFAVEAENRVVDISGFSIGSDGHSIQYRNLLAWKY

[0830] LENGKREFYLLMNYGKKGRIRFTDGTDIKKSGKWQGLLYGGGKAKVIDLTFDPDDEQLIILPLAFGT

[0831] RQGREFIWNDLLSLETGLIKLANGRVIEKTIYNKKIGRDEPALFVALTFERREVVDPSNIKPVNLIG

[0832] VARGENIPAVIALTDPEGCPLPEFKDSSGGPTDILRIGEGYKEKQRAIQAAKEVEQRRAGGYSRKFA

[0833] SKSRNLADDMVRNSARDLFYHAVTHDAVLVFANLSRGFGRQGKRTFMTERQYTKMEDWLTAKLAYEG

[0834] LTSKTYLSKTLAQYTSKTCSNCGFTITYADMDVMLVRLKKTSDGWATTLNNKELKAEYQITYYNRYK

[0835] RQTVEKELSAELDRLSEESGNNDISKWTKGRRDEALFLLKKRFSHRPVQEQFVCLDCGHEVHAAEQA

[0836] ALNIARSWLFLNSNSTEFKSYKSGKQPFVGAWQAFYKRRLKEVWKPNA

[0837] CasY(ncbi.nlm.nih.gov/protein/APG80656.1)

[0838] >APG80656.1 CRISPR相关的蛋白CasY[未培养的俭菌门(Parcubacteria)细菌]

[0839] MSKRHPRISGVKGYRLHAQRLEYTGKSGAMRTIKYPLYSSPSGGRTVPREIVSAINDDYVGLYGLSN

[0840] FDDLYNAEKRNEEKVYSVLDFWYDCVQYGAVFSYTAPGLLKNVAEVRGGSYELTKTLKGSHLYDELQ

[0841] IDKVIKFLNKKEISRANGSLDKLKKDIIDCFKAEYRERHKDQCNKLADDIKNAKKDAGASLGERQKK

[0842] LFRDFFGISEQSENDKPSFTNPLNLTCCLLPFDTVNNNRNRGEVLFNKLKEYAQKLDKNEGSLEMWE

[0843] YIGIGNSGTAFSNFLGEGFLGRLRENKITELKKAMMDITDAWRGQEQEEELEKRLRILAALTIKLRE

[0844] PKFDNHWGGYRSDINGKLSSWLQNYINQTVKIKEDLKGHKKDLKKAKEMINRFGESDTKEEAVVSSL

[0845] LESIEKIVPDDSADDEKPDIPAIAIYRRFLSDGRLTLNRFVQREDVQEALIKERLEAEKKKKPKKRK

[0846] KKSDAEDEKETIDFKELFPHLAKPLKLVPNFYGDSKRELYKKYKNAAIYTDALWKAVEKIYKSAFSS

[0847] SLKNSFFDTDFDKDFFIKRLQKIFSVYRRFNTDKWKPIVKNSFAPYCDIVSLAENEVLYKPKQSRSR

[0848] KSAAIDKNRVRLPSTENIAKAGIALARELSVAGFDWKDLLKKEEHEEYIDLIELHKTALALLLAVTE

[0849] TQLDISALDFVENGTVKDFMKTRDGNLVLEGRFLEMFSQSIVFSELRGLAGLMSRKEFITRSAIQTM

[0850] NGKQAELLYIPHEFQSAKITTPKEMSRAFLDLAPAEFATSLEPESLSEKSLLKLKQMRYYPHYFGYE

[0851] LTRTGQGIDGGVAENALRLEKSPVKKREIKCKQYKTLGRGQNKIVLYVRSSYYQTQFLEWFLHRPKN

[0852] VQTDVAVSGSFLIDEKKVKTRWNYDALTVALEPVSGSERVFVSQPFTIFPEKSAEEEGQRYLGIDIG

[0853] EYGIAYTALEITGDSAKILDQNFISDPQLKTLREEVKGLKLDQRRGTFAMPSTKIARIRESLVHSLR

[0854] NRIHHLALKHKAKIVYELEVSRFEEGKQKIKKVYATLKKADVYSEIDADKNLQTTVWGKLAVASEIS

[0855] ASYTSQFCGACKKLWRAEMQVDETITTQELIGTVRVIKGGTLIDAIKDFMRPPIFDENDTPFPKYRD

[0856] FCDKHHISKKMRGNSCLFICPFCRANADADIQASQTIALLRYVKEEKKVEDYFERFRKLKNIKVLGQ

[0857] MKKI

[0858] 术语“Cas12”或“Cas12结构域”是指包括Cas12蛋白或其片段(例如,包括Cas12的活性、失活或部分活性的DNA切割域的蛋白和/或Cas12的gRNA结合结构域)的RNA向导的核酸酶。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)相关的核酸酶。以下提供实施例性久氏芽孢杆菌(Bacillus hisashii) Cas 12b(BhCas12b)Cas 12结构域的序列:MAPKKKRKVGIHGVPAAATRSFILKIEPNEEVKKGLWKTH EVLNHGIAYYMNILKLIRQEAIYEHHEQDPKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYE ELVPSSVEKKGEANQLSN KFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEEKKKWEEDKKKDPLAKIL GKLAEYGLIPLFIPYTDSNEPIVKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWESWNLKVKEEYEKVEKEYKTL EERIKEDIQALKALEQYEKERQEQLLRDTLNTNEYRLSKRGLRGWREIIQKWLKMDENEPSEKYLEVFKDYQRKHP REAGDYSVYEFLSKKENHFIWRNHPEYPYLYATFCEIDKKKKDAKQQATFTLADPINHPLWVRFEERSGSNLNKYR ILTEQLHTEKLKKKLTVQLDRLIYPTESGGWEEKGKVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKG TLGGARVQFDRDHLRRYPHKVESGNVGRIYFNMTVNIEPTESPVSKSLKIHRDDFPKVVNFKPKELTEWIKDSKGK KLKSGIESLEIGLRVMSIDLGQRQAAAAASIFEVVDQKPDIEGKLFFPIKGTELYAVHRASFNIKLPGETLVKSREV LRKAREDNLKLMNQKLNFLRNVLHFQQFEDITEREKRVTKWISRQENSDVPLVYQDELIQIRELMYKPYKDWVAFL KQLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDEIDRTRKFLLRWSLRPTEPGEVRRLEPGQRFAIDQLNH LNALKEDRLKKMANTIIMHALGYCYDVRKKKWQAKNPACQIILFEDLSNYNPYEERSRFENSKLMKWSRREIPRQV ALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKLQDNRFFKNLQREGRLTLDKIAVLKEGDLYPDKGGE KFISLSKDRKCVTTHADINAAQNLQKRFWTRTHGFYKVYCKAYQVDGQTVYIPESKDQKQKIIEEFGEGYFILKDG VYEWVNAGKLKIKKGSSKQSSSELVDSDILKDSFDLASELKGEKLMLYRDPSGNVFPSDKWMAAGVFFGKLERILI SKLTNQYSISTIEDDSSKQSMKRPAATKKAGQAKKKK。

[0859] 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。

[0860] “胞苷脱氨酶”是指能够催化将氨基转换为羰基的脱氨反应的多肽或其片段。在一个实施方案中,胞苷脱氨酶将胞嘧啶转换为尿嘧啶或将5-甲基胞嘧啶转换为胸腺嘧啶。PmCDA1(衍生自海七鳃鳗(海七鳃鳗胞嘧啶脱氨酶1,“PmCDA1”))、AID(激活诱导胞苷脱氨酶;AICDA)(衍生自哺乳动物(例如,人类、猪、牛、马、猴等))和APOBEC是实施例性的胞苷脱氨酶。

[0861] 术语“保守性氨基酸取代”或“保守性突变”是指一个氨基酸被具有共同特性的另一个氨基酸取代。定义个别氨基酸之间共同特性的一种功能方法是分析同源生物体的相应蛋白质之间氨基酸变化的标准化频率(Schulz,G.E.and Schirmer,R.H.,Principles of Protein Structure,Springer-Verlag,New York(1979))。根据这样的分析,可以定义氨基

酸群组为其中群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.and Schirmer,R.H.,同上)。保守性突变的非限制性实例包括氨基酸的氨基酸取代,例如以赖氨酸取代精氨酸和反之亦然,从而可以保持正电荷;以谷氨酸取代天冬氨酸和反之亦然,以保持负电荷;以丝氨酸取代苏氨酸,使得可以保持一个游离的-OH;和以谷氨酰胺取代天冬酰胺,使得可以保持游离的-NH_2。

[0862] 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,而在靠近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。

[0863] 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。在一些实施方案中,脱氨酶是腺苷脱氨酶,其催化腺嘌呤水解脱氨为次黄嘌呤。在一些实施方案中,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)。在一些实施方案中,脱氨酶或脱氨酶结构域是分别催化腺苷或脱氧腺苷水解脱氨为肌苷或脱氧肌苷的腺苷脱氨酶。在一些实施方案中,腺苷脱氨酶催化脱氧核糖核酸(DNA)中腺苷的水解脱氨。本文所提供的腺苷脱氨酶(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。在一些实施方案中,腺苷脱氨酶来自细菌,例如大肠杆菌、金黄色葡萄球菌、鼠伤寒沙门氏菌、腐败希瓦氏菌、流感嗜血杆菌或新月芽孢杆菌。

[0864] 在一些实施方案中,腺苷脱氨酶是TadA脱氨酶。在一些实施方案中,TadA脱氨酶是TadA变体。在一些实施方案中,TadA变体是TadA*8。在一些实施方案中,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。在一些实施方案中,脱氨酶或脱氨酶结构域在自然界中不存在。例如,在一些实施方案中,脱氨酶或脱氨酶结构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申请第PCT/2017/045381(WO 2018/027078)号和第PCT/US2016/058344(WO 2017/070632)号中描述了脱氨酶结构域,各自通过引用方式整体并入本文。另请参见Komor,A.C.,et al.,“Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage”Nature 533,420-424(2016);Gaudelli,N.M.,et al.,“Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage”Nature 551,464-471(2017);Komor,A.C.,et al.,“Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity”Science Advances 3:eaao4774(2017)),以及Rees,H.A.,et al.,“Base editing:precision chemistry on the genome and transcriptome of living cells.”Nat Rev Genet.2018Dec;19(12):770-788.doi:10.1038/s41576-018-0059-1,其特此以引入方式整体并入。

[0865] “检测”是指识别待检测分析物的存在、不存在或量。在一个实施方案中,检测多核苷酸或多肽中的序列改变。在另一个实施方案中,检测插入缺失的存在。

[0866] “可检测标记”是指一种组合物,当与感兴趣的分子连接时,经由光谱、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,通常用于酶联免疫吸附测定(ELISA))、生物素、地高辛,或半抗原。

[0867] "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况或病症。

[0868] 如本文所用,术语"有效量"是指足以引发所需生物反应的生物活性剂的量。用于实施本发明以治疗疾病的活性剂的有效量取决于给药方式、受试者的年龄、体重和一般健康状况而变化。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。在一个实施方案中,有效量是足以在细胞(例如,体外或体内细胞)中的感兴趣基因中引入改变的本发明的碱基编辑器(例如,包括可编程的DNA结合蛋白、核碱基编辑器及gRNA的融合蛋白)的量。在一些实施方案中,本文所提供的融合蛋白,例如包括nCas9结构域和脱氨酶结构域(例如,腺苷脱氨酶或胞苷脱氨酶)的核碱基编辑器的有效量可以指足以使诱导编辑由核碱基编辑器特异性地绑定和编辑的目标位点。在一个实施方案中,有效量是达成治疗效果(例如,减轻或控制疾病或其症状或病况)所需的碱基编辑器的量。这种治疗效果不需要足以改变受试者、组织或器官的所有细胞中的感兴趣基因,而仅改变约存在于受试者、组织或器官中的1%、5%、10%、25%、50%、75%或更多的感兴趣基因。

[0869] 在一些实施方案中,本文所提供的融合蛋白,例如包括nCas9结构域和脱氨酶结构域(例如,腺苷脱氨酶或胞苷脱氨酶)的核碱基编辑器的有效量是指足以诱导由本文所述核碱基编辑器特异性地结合和编辑的目标位点的量。本领域技术人员将理解,有效量的药剂,例如融合蛋白、核酸酶、杂合蛋白、蛋白二聚体、蛋白(或蛋白二聚体)和多核苷酸的复合物,或多核苷酸,可能会因各种因素而异,例如,所需的生物反应,例如在欲编辑的特定等位基因、基因组或目标位点,欲靶向的细胞或组织、和/或所使用的药剂。

[0870] "片段"是指多肽或核酸分子的一部分。所述部分包括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%。片段可包括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。

[0871] "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的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。在一个实施方案中,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可称为单个向导RNA(sgRNA),但"gRNA"可互换使用以指以单个分子或两个或更多个分子的复合物形式存在的向导RNA。通常,作为单个RNA种类存在的gRNA包括两个结构域:(1)与目标核酸具有同源性的结构域(例如,向导Cas9复合物与目标的结合);(2)结合Cas9蛋白的结构域。在一些实施方案中,结构域(2)对应于称为tracrRNA的序列,并且包括茎环结构。例如,在一些实施方案中,结构域(2)与Jinek et al.,Science337:816-821(2012)中提供的tracrRNA相同或同源,其通过引用方式整体内容并入本文。gRNA的其他实施例(例如,那些包括结构域2者)可以于2013年9月6日提交的美国临时专利申请U.S.S.N.61/874,682(题为"可切换的Cas9核酸酶及其用途(Switchable Cas9Nucleases and Uses Thereof)")和于2013年9月6日提交的美国临时专利申请U.S.S.N.61/874,746题为"功能性核酸酶递送系统(Delivery System For Functional Nucleases)"),各自特此通过引用方式整体内容并入本文。在一些实施方式中,gRNA包括结构域(1)和(2)中的两个或更多个,并且可以被称为"延伸的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包括与目标

位点互补的核苷酸序列,其介导核酸酶/RNA复合物与所述目标位点的结合,提供核酸酶:RNA复合物的序列特异性。

[0872] “杂交”是指互补核碱基之间的氢键,其可以是沃森-克里克、胡斯坦或反向胡斯坦氢键。例如,腺嘌呤和胸腺嘧啶是通过形成氢键配对的互补的核碱基。

[0873] 术语“碱基修复抑制剂”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。在一些实施方案中,IBR是肌苷碱基切除修复的抑制剂。碱基修复的实施例性抑制剂包括APE1、Endo III、Endo IV、Endo V、Endo VIII、Fpg、hOGG1、hNEIL1、T7 Endo1、T4PDG、UDG、hSMUG1及hAAG的抑制剂。在一些实施方案中,IBR是Endo V或hAAG的抑制剂。在一些实施方案中,IBR是催化失活的EndoV或催化失活的hAAG。在一些实施方案中,碱基修复抑制剂是Endo V或hAAG的抑制剂。在一些实施方案中,碱基修复抑制剂是催化失活的EndoV或催化失活的hAAG。

[0874] 在一些实施方案中,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。在一些实施方案中,UGI结构域包括野生型UGI或野生型UGI的片段。在一些实施方案中,本文所提供的UGI蛋白质包括UGI的片段和与UGI或UGI片段同源的蛋白质。在一些实施方案中,碱基修复抑制剂是肌苷碱基切除修复抑制剂。在一些实施方案中,碱基修复抑制剂是“催化失活的肌苷特异性核酸酶”或“死亡的肌苷特异性核酸酶”。不希望受任何特定理论的束缚,催化失活的肌苷糖基化酶(例如,烷基腺嘌呤糖基化酶(AAG))可以结合肌苷,但不能产生无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。在一些实施方案中,催化失活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。非限制性实施例性催化失活肌苷特异性核酸酶包括催化失活的烷基腺苷糖基化酶(AAG核酸酶),例如来自人,和催化失活内切核酸酶V(EndoV核酸酶),例如来自大肠杆菌。在一些实施方案中,催化失活的AAG核酸酶包括E125Q突变或另一种AAG核酸酶中的相应的突变。

[0875] “增加”是指至少10%、25%、50%、75%或100%的正向变化。

[0876] “内含肽”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)。内含子也称为“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”。在一些实施方案中,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是由两个独立的基因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”。

[0877] 还可以使用其他内含肽系统。例如,已经描述了基于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(例如,在Stevens et al.,J Am Chem Soc.2016Feb.24;138(7):2162-5,通过引用方式并入本文)。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,于美国专利No.8,394,604中所述,其通过引用方式并入本文。

[0878] 提供内含肽的实施例性核苷酸和氨基酸序列。

[0879] DnaE内含肽-N DNA：

[0880] TGCCTGTCATACGAAACCGAGATACTGACAGTAGAATATGGCCTTCTGCCAATCGGGAAGATTGTGG

[0881] AGAAACGGATAGAATGCACAGTTTACTCTGTCGATAACAATGGTAACATTTATACTCAGCCAGTTGC

[0882] CCAGTGGCACGACCGGGGAGAGCAGGAAGTATTCGAATACTGTCTGGAGGATGGAAGTCTCATTAGG

[0883] GCCACTAAGGACCACAAATTTATGACAGTCGATGGCCAGATGCTGCCTATAGACGAAATCTTTGAGC

[0884] GAGAGTTGGACCTCATGCGAGTTGACAACCTTCCTAAT

[0885] DnaE内含肽-N蛋白:

[0886] CLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDR

[0887] GEQEVFEYCLEDGSLIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPN

[0888] DnaE内含肽-C DNA:

[0889] ATGATCAAGATAGCTACAAGGAAGTATCTTGGCAAACAAAACGTTTATGA

[0890] TATTGGAGTCGAAAGAGATCACAACTTTGCTCTGAAGAACGGATTCATAGCTTCTAAT

[0891] 内含肽-C:

[0892] MIKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN

[0893] Cfa-N DNA：

[0894] TGCCTGTCTTATGATACCGAGATACTTACCGTTGAATATGGCTTCTTGCCTATTGGAAAGATTGTCG

[0895] AAGAGAGAATTGAATGCACAGTATATACTGTAGACAAGAATGGTTTCGTTTACACACAGCCCATTGC

[0896] TCAATGGCACAATCGCGGCGAACAAGAAGTATTTGAGTACTGTCTCGAGGATGGAAGCATCATACGA

[0897] GCAACTAAAGATCATAAATTCATGACCACTGACGGGCAGATGTTGCCAATAGATGAGATATTCGAGC

[0898] GGGGCTTGGATCTCAAACAAGTGGATGGATTGCCA

[0899] Cfa-N蛋白：

[0900] CLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCLEDGSIIR

[0901] ATKDHKFMTTDGQMLPIDEIFERGLDLKQVDGLP

[0902] Cfa-C DNA：

[0903] ATGAAGAGGACTGCCGATGGATCAGAGTTTGAATCTCCCAAGAAGAAGAGGAAAGTAAAGATAATAT

[0904] CTCGAAAAAGTCTTGGTACCCAAAATGTCTATGATATTGGAGTGGAGAAAGATCACAACTTCCTTCT

[0905] CAAGAACGGTCTCGTAGCCAGCAAC

[0906] Cfa-C蛋白：

[0907] MKRTADGSEFESPKKKRKVKIISRKSLGTQNVYDIGVEKDHNFLLKNGLVASN

[0908] 内含肽-N和内含肽-C可以分别与分裂Cas9的N端部分和分裂Cas9的C端部分融合，以用于连接分裂Cas9的N端部分和分裂Cas9的C端部分。例如，在一些实施方案中，将内含肽-N与分裂Cas9的N端部分的C端融合，即形成N--[分裂Cas9的N端部分的结构]-[内含肽-N]--C。在一些实施方案中，内含肽-C与分裂的Cas9的C端部分的N端融合，即形成N-[内含肽-C]--[C端的结构Cas9]-C的一部分。内含肽介导的蛋白质剪接用于连接与内含肽融合的蛋白质(例如，分裂Cas9)的机制是本领域已知的，例如，在Shah et al.,Chem Sci.2014;5(1):446-461中所述，其通过引用方式并入本文。用于设计和使用内含肽的方法在本领域中是已知的，并且描述于例如WO2014004336、WO2017132580、US20150344549和US20180127780中，各自通过引用方式整体并入本文。

[0909] 术语“单离的”、“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。“单离”表示与原始来源或周围环境的分离程度。“纯化”表示高于单离程度的分离程度。“纯化的”或“生物纯的”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特性或引起其他不利后果。亦即,若本发明的核酸或肽在通过重组DNA技术生产时实质上不含细胞材料、病毒材料或培养基,或者在化学合成时实质上不含化学前体或其他化学品,则所述核酸或肽被纯化。纯度和均一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。

[0910] “单离的多核苷酸”是指在本发明的核酸分子所衍生自的生物体的天然存在的基因组中不含基因的核酸(例如,DNA),所述基因在所述基因的侧边。因此,所述术语包括,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;或作为与其他序列无关的个别分子(例如,通过PCR或限制性内切核酸酶酶切产生的cDNA或基因组或cDNA片段)存在。此外,所述术语包括从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。

[0911] “单离的多肽”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,多肽就被单离出来。优选地,所述制剂是至少75重量%,更优选至少90重量%,并且最优选至少99重量%的本发明多肽。本发明的分离的多肽可以,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获得。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。

[0912] 如本文所用,术语“连接子”可指共价连接子(例如,共价键)、非共价连接子、化学基团或连接两个分子或部分(例如,蛋白质复合物或核糖核复合物的两个组分或融合蛋白的两个结构域)的分子,例如多核苷酸可编程的DNA结合结构域(例如dCas9)和脱氨酶结构域(例如腺苷脱氨酶、胞苷脱氨酶或腺苷脱氨酶和胞苷脱氨酶)或napDNAbp结构域(例如Cas12b)和脱氨酶结构域(例如腺苷脱氨酶或胞苷脱氨酶)。在特定实施方案中,连接子在插入Cas蛋白或其片段内的脱氨酶结构域的侧边。连接子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,在一些实施方案中,连接子可以连接多核苷酸可编程的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域。在一些实施方案中,连接子可以连接CRISPR多肽和脱氨酶。在一些实施方案中,连接子可以连接Cas9和脱氨酶。在一些实施方案中,连接子可以连接dCas9和脱氨酶。在一些实施方案中,连接子可以连接nCas9和脱氨酶。例如,在一些实施方案中,连接子可以连接Cas12a/Cpf1、Cas12b/C2c1、Cas12c/C2c3、Cas12d/CasY、Cas12e/CasX、Cas12g、Cas12h或Cas12i及脱氨酶。在一些实施方案中,连接子可以连接向导多核苷酸和脱氨酶。在一些实施方案中,连接子可以连接碱基编辑器系统的脱氨组分和多核苷酸可编程的核苷酸结合组分。在一些实施方案中,连接子可以连接碱基编辑器系统的脱氨组分和napDNAbp组分的RNA结合部分。在一些实施方案中,连接子可以连接碱基编辑器系统的脱氨组分和多核苷酸可编程的核苷酸结合组分的RNA结合部分。在一些实施方案中,连接子可以连接碱基编辑器系统的脱氨组分的RNA结合部分和多核苷酸可编程的核苷酸结合组分的RNA结合部分。连接子可以位于两个基团、分子或其他部分之间或侧边,并经由共价键或非共价相互作用连接到各者,因此连接两者。在一些实施方案中,连

接子可以是有机分子、基团、聚合物或化学部分。在一些实施方案中,连接子可以是多核苷酸。在一些实施方案中,连接子可以是DNA连接子。在一些实施方案中,连接子可以是RNA连接子。在一些实施方案中,连接子可包括能够结合配体的适体。在一些实施方案中,配体可以是碳水化合物、肽、蛋白质或核酸。在一些实施方案中,连接子可包括可衍生自核糖开关的适体。衍生自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Q核苷1(pre-queosine1,PreQ1)核糖开关。在一些实施方案中,连接子可包括与多肽或蛋白质结构域诸如多肽配体结合的适体。在一些实施方案中,多肽配体可以是K同源性(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、不育α基序、端粒酶Ku结合基序和Ku蛋白、端粒酶Sm7结合基序和Sm7蛋白、或RNA识别基序。在一些实施方案中,多肽配体可以是碱基编辑器系统组分的一部分。例如,核碱基编辑组分可包括脱氨酶结构域和RNA识别基序。

[0913] 在一些实施方案中,连接子可以是氨基酸或多个氨基酸(例如,肽或蛋白质)。在一些实施方案中,连接子的长度可为约5至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至30、30至40、40至50、50至60、60至70、70至80、80至90、或90至100个氨基酸的长度。在一些实施方案中,连接子的长度可为约100至150、150至200、200至250、250至300、300至350、350至400、400至450或450至500个氨基酸。也可以考虑更长或更短的连接子。

[0914] 在一些实施方案中,连接子连接RNA可编程的核酸酶的gRNA结合结构域,包括Cas9核酸酶结构域和核酸编辑蛋白(例如,胞苷或腺苷脱氨酶)的催化域。在一些实施方案中,连接子连接dCas9和核酸编辑蛋白。例如,连接子位于两个基团、分子或其他部分之间或侧边,并经由共价键与各者连接,因此连接两者。在一些实施方案中,连接子是一个氨基酸或多个氨基酸(例如,肽或蛋白质)。在一些实施方案中,连接子是有机分子、基团、聚合物或化学部分。在一些实施方案中,连接子的长度为5至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个氨基酸的长度。也预期更长或更短的连接子。

[0915] 在一些实施方案中,核碱基编辑器的结构域通过包括SGGSSGSETPGTSESATPESSGGS、SGGSSGGSSGSETPGTSESATPESSGGSSGGS、或GGSGGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSGG的氨基酸序列的连接子融合。在一些实施方案中,核碱基编辑器的结构域经由包括氨基酸序列SGSETPGTSESATPES的连接子融合,其也可称为XTEN连接子。在一些实施方案中,连接子包括氨基酸序列SGGS。在一些实施方案中,连接子包括$(SGGS)_n$、$(GGGS)_n$、$(GGGGS)_n$、$(G)_n$、$(EAAAK)_n$、$(GGS)_n$、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。在一些实施方案中,n是1、2、3、4、5、6、7、8、9、10、11、12、13、14或15。

[0916] 在一些实施方案中,连接子的长度为24个氨基酸。在一些实施方案中,连接子包括氨基酸序列SGGSSGGSSGSETPGTSESATPES。在一些实施方案中,连接子的长度为40个氨基酸。

在一些实施方案中,连接子包括氨基酸序列SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGS。在一些实施方案中,连接子的长度为64个氨基酸。在一些实施方案中,连接子包括氨基酸序列SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGSSGSETPGTSESATPESSGGS SGGS。在一些实施方案中,连接子的长度为92个氨基酸。在一些实施方案中,连接子包括氨基酸序列PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATS。

[0917] “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。

[0918] 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识别原始残基随后是所述残基在序列中的位置以及通过新取代残基的身份来描述。用于进行本文所提供的氨基酸取代(突变)的各种方法是本领域公知的,并且由例如Green and Sambrook,Molecular Cloning:A Laboratory Manual(4th ed.,Cold Spring Harbor Laboratory Press,Cold Spring Harbor,N.Y.(2012))提供。在一些实施方案中,当前公开的碱基编辑器可以有效地在核酸(例如,受试者基因组内的核酸)中产生“预期突变”,诸如点突变,而不会产生大量非预期突变,诸如意外的点突变。在一些实施方案中,预期突变是由特定碱基编辑器(例如,胞苷碱基编辑器或腺苷碱基编辑器)与向导多核苷酸(例如,gRNA)结合而产生的突变,特别设计用于产生预期突变。

[0919] 通常,在序列(例如,如本文所述氨基酸序列)中产生或识别的突变相对于参考(或野生型)序列(即不包括突变的序列)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。

[0920] 术语“非保守性突变”涉及不同群组之间的氨基酸取代,例如,以赖氨酸取代色氨酸,或以丝氨酸取代苯丙氨酸等。在这种情况下,非保守性氨基酸取代优选不干扰,或抑制功能性变体的生物活性。非保守性氨基酸取代可增强功能性变体的生物活性,使得与野生型蛋白质相比,功能性变体的生物活性增加。

[0921] 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于2000年11月23日提交的Plank等人的国际PCT申请PCT/EP2000/011690,其在2001年5月31日公布为WO/2001/038547公布,其内容以引用方式并入本文以用于其对实施例性核定位序列的公开。在其他实施方案中,NLS是优化的NLS,例如由Koblan et al.,Nature Biotech.2018doi:10.1038/nbt.4172所述。在一些实施方案中,NLS包括氨基酸序列KRTADGSEFESPKKKRKV、KRPAATKKAGQAKKKK、KKTELQTTNAENKTKKL、KRGINDRNFWRGENGRKTR、RKSGKIAAIVVKRPRK、PKKKRKV、或MDSLLMNRRKFLYQFKNVRWAKGRRETYLC。\

[0922] 如本文所用,术语“核酸”和“核酸分子”是指包括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,例如包括三个或更多个核苷酸的核酸分子是线性分子,其中相邻的核苷酸通过磷酸二酯键相互连接。在一些实施方案中,“核酸”是指个别核酸残基(例如,核苷酸和/或核苷)。在一些实施方案中,“核酸”是指包括三个或更多个个别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,至少三个核苷酸的串)。在一些实施方案中,“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中。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,核酸可以包括核苷类似物,诸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。在一些实施方案中,核酸是或包括天然核苷(例如,腺苷、胸苷、鸟苷、胞苷、尿苷、脱氧腺苷、脱氧胸苷、脱氧鸟苷及脱氧胞苷);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)-甲基鸟苷及2-硫代胞苷);化学修饰的碱基;生物修饰碱基(例如甲基化碱基);插入的碱基;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/或修饰的磷酸基团(例如,硫代磷酸酯和5’-N-亚磷酰胺键)。

[0923] 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)连结的蛋白质,所述核酸是诸如将napDNAbp向导至特定的核酸序列的向导核酸或向导多核苷酸(例如,gRNA)。在一些实施方案中,多核苷酸可编程的核苷酸结合结构域是多核苷酸可编程的DNA结合结构域。在一些实施方案中,多核苷酸可编程的核苷酸结合结构域是多核苷酸可编程的RNA结合结构域。在一些实施方案中,多核苷酸可编程的核苷酸结合结构域是Cas9蛋白。Cas9蛋白可以与向导RNA连结,所述向导RNA将Cas9蛋白向导至与向导RNA互补的特定DNA序列。在一些实施方案中,napDNAbp是Cas9结构域,例如核酸酶活性Cas9、Cas9切口酶(nCas9)或核酸酶失活Cas9(dCas9)。核酸可编程的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及Cas12i和。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(也以知为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类Cas效应蛋白、V型Cas效应蛋白、VI类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。参见,例如,Makarova et al.“Classification and Nomenclature of CRISPR-Cas Systems:Where from Here?”CRISPR J.2018Oct;1:325-336.doi:10.1089/crispr.2018.0033;Yan et al.,“Functionally diverse type V CRISPR-Cas systems”Science.2019Jan 4;363(6422):88-91.doi:10.1126/science.aav7271,各自特此通过引用方式并入。

[0924] 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,而核苷是核苷酸的组分。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)及尿嘧啶(U)—被称为原始(primary)或权威(canonical)。腺嘌呤和鸟嘌呤衍生自嘌呤,而胞嘧啶、尿嘧啶和胸腺嘧啶衍生自嘧啶。DNA和RNA还可以包括其他(非主要)修饰的碱基。非限制性实施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)及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,两者皆通过脱氨(用羰基置换胺基团)产生。次黄嘌呤可以由腺嘌呤修饰而成。黄嘌呤可以由鸟嘌呤修饰。尿嘧啶可由胞嘧啶脱氨产生。“核苷”是由核碱基和五碳糖(核糖或脱氧核糖)所组成。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、脱氧尿苷及脱氧胞苷。具有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)及假尿苷(Ψ)。“核苷酸”是由核碱基、五碳糖(核糖或脱氧核糖)和至少一个磷酸基团所组成。

[0925] 术语“核酸可编程的DNA结合蛋白”或“napDNAbp”是指与核酸(例如DNA或RNA)连结的蛋白质,例如向导napDNAbp至特定核酸序列的向导核酸。例如,Cas12蛋白可以与向导Cas12蛋白连接与向导RNA互补的特定DNA序列的向导RNA。在一些实施方案中,napDNAbp是Cas12结构域,例如核酸酶活性Cas12结构域。napDNAbps的例子包括Cas12a/Cpf1、Cas12b/C2c1、Cas12c/C2c3、Cas12d/CasY、Cas12e/CasX、Cas12g、Cas12h及Cas12i。其他napDNAbps也在本公开的范围内,尽管它们可能没有在本公开中具体列出。参见,例如,Makarova et al.“Classification and Nomenclature of CRISPR-Cas Systems:Where from Here?”CRISPR J.2018Oct;1:325-336.doi:10.1089/crispr.2018.0033;Yan et al.,“Functionally diverse type V CRISPR-Cas systems”Science.2019Jan 4;363(6422):88-91.doi:10.1126/science.aav7271,各自特此通过引用方式全部内容并入。

[0926] 如本文所用,术语“核苷碱基编辑结构域”或“核苷碱基编辑蛋白”是指可催化RNA或DNA中的核苷碱基修饰的蛋白质或酶,诸如胞嘧啶(或胞苷)成尿嘧啶(或尿苷)或胸腺嘧啶(或胸苷)及腺嘌呤(或腺苷)成次黄嘌呤(或肌苷)脱氨,以及非模板化的核苷酸添加和插入。在一些实施方案中,核碱基编辑结构域是脱氨酶结构域(例如,腺嘌呤脱氨酶或腺苷脱氨酶;或胞苷脱氨酶或胞嘧啶脱氨酶)。在一些实施方案中,核碱基编辑结构域是超过一个脱氨酶结构域(例如,腺嘌呤脱氨酶或腺苷脱氨酶和胞苷或胞嘧啶脱氨酶)。在一些实施方案中,核碱基编辑结构域可以是天然存在的核碱基编辑结构域。在一些实施方案中,核碱基编辑结构域可以是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、人类、黑猩猩、大猩猩、猴、牛、狗、大鼠或小鼠。例如,核碱基编辑蛋白在国际PCT申请号PCT/2017/045381(WO 2018/027078)和PCT/US2016/058344(WO 2017/070632)中有所描述,各自通过引用整体并入本文。另参阅Komor,A.C.,et al.,“Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage”Nature 533,420-424(2016);Gaudelli,N.M.,et al.,“Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage”Nature 551,464-471(2017);以及Komor,A.C.,et al.,“Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity”Science Advances3:

eaao4774(2017),其全部内容特此通过引入方式并入。

[0927] 如本文所用,“获得药剂”中的“获得”包括合成、购买或以其他方式获得所述药剂。

[0928] 如本文所用,“患者”或“受试者”是指被诊断患有、有患有的风险或发展、或怀疑患有或发展疾病或病症的哺乳动物受试者或个体。在一些实施方案中,术语“患者”是指具有高于平均发展疾病或病症的可能性的哺乳动物受试者。实施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。实施例性的人类患者可以是男性和/或女性。

[0929] “有需要的患者”或“有需要的受试者”在本文中是指被诊断患有、有风险或患有、预定患有或怀疑患有疾病或病症的患者。

[0930] 术语“致病突变”、“致病变异”、“疾病外壳突变”、“致病变异”、“致病突变”或“易感突变”是指增加个体对某种疾病或病症的易感性或素质。在一些实施方案中,致病性突变包括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。

[0931] 术语“药学上可接受的载体”是指药学上可接受的材料、组合物或赋形剂,诸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、硬脂酸钙或硬脂酸锌或硬脂酸)或溶剂包封材料,涉及将化合物从身体的一个部位(例如,递送部位)携带或运输到另一部位(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兼容并且对受试者的组织没有伤害的意义上是“可接受的”(例如,生理兼容的、无菌的、生理pH等)。诸如“赋形剂”、“载体”、“药学上可接受的载体”、“载体”等术语在本文中可互换使用。

[0932] 术语“药物组合物”可以指配制用于药物用途的组合物。

[0933] 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、肽或多肽可以指个别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,诸如碳水化合物基团、羟基、磷酸基、法呢基、异法呢基、脂肪酸基团、用于缀合、功能化或其他修饰的连接子等。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”是指包括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端)部分或羧基末端(C端)蛋白,因此分别形成氨基-末端融合蛋白或羧基-末端融合蛋白。蛋白质可以包括不同的结构域,例如,核酸结合结构域(例如,引导蛋白质与目标位点结合的Cas9的gRNA结合结构域)和核酸切割域,或核酸编辑蛋白的催化域。在一些实施方案中,蛋白质包括蛋白质部分(例如构成核酸结合结构域的氨基酸序列)和有机化合物(例如可以作用为核酸切割剂的化合物)。在一些实施方案中,蛋白质与核酸例如RNA或DNA形成复合物或连结。本文所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,本文所提供的蛋白质可以经由重组蛋白质表达和纯化产生,这尤其适用于包括肽连接子的融合蛋白质。重组蛋白表达和纯化的方法是众所周知的,包括Green and Sambrook,Molecular Cloning:A Laboratory Manual(4th ed.,Cold Spring Harbor Laboratory Press,Cold Spring Harbor,NY(2012)),其全部内容其通过引用方式并入本文。

[0934] 本文公开的多肽和蛋白质(包括其功能性部分和功能性变体)可包括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--羟基脯氨酸和反式-4-羟基脯氨酸、4-氨基苯丙氨酸、4-硝基苯丙氨酸、4-氯苯丙氨酸、4-羧基苯丙氨酸、β-苯基丝氨酸、β-羟基苯丙氨酸、苯基甘氨酸、α-萘丙氨酸、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、高苯丙氨酸及α-叔丁基甘氨酸。多肽和蛋白质可以与多肽构建体的一个或多个氨基酸的翻译后修饰相关。翻译后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、糖基磷脂酰肌醇化、脂酰化和碘化。

[0935] 术语“多核苷酸可编程的核苷酸结合结构域”或“核酸可编程的DNA结合蛋白(napDNAbp)”是指与核酸(例如DNA或RNA)连结的蛋白质,诸如将多核苷酸可编程的核苷酸结合结构域向导至特定核酸序列的向导多核苷酸(例如向导RNA)。在一些实施方案中,多核苷酸可编程的核苷酸结合结构域是多核苷酸可编程的DNA结合结构域。在一些实施方案中,多核苷酸可编程的核苷酸结合结构域是多核苷酸可编程的RNA结合结构域。在一些实施方案中,多核苷酸可编程的核苷酸结合结构域是Cas12蛋白。

[0936] 如本文在蛋白质或核酸的上下文中使用的术语“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,在一些实施方案中,重组蛋白或核酸分子包括氨基酸或核苷酸序列,所述序列包括至少一个、至少两个、至少三个、至少四个、至少五个、至少六个或至少与任何自然发生的序列相比,有七个突变。

[0937] “减少”是指至少10%、25%、50%、75%或100%的负变化。

[0938] “参考”是指标准或对照条件。在一个实施方案中,参考是野生型或健康细胞。在其他实施方案中且不限于,参照是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/或不包括感兴趣的多核苷酸的对照载体的未处理细胞。

[0939] “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将通常为至少约16个氨基酸、至少约20个氨基酸、至少约25个氨基酸、约35个氨基酸、约50个氨基酸或约100个氨基酸。对于核酸,参考核酸序列的长度通常为至少约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。在一些实施方案中,参考序列是目的蛋白质的野生型序列。在其他实施方案中,参考序列是编码野生型蛋白质的多核苷酸序列。

[0940] 术语“RNA可编程的核酸酶”和“RNA向导的核酸酶”与一种或多种不是切割目标的RNA一起使用(例如,结合或连结)。在一些实施方案中,当与RNA形成复合物时,RNA可编程的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可以作为两个或更多个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可称为单个向导RNA(sgRNA),尽管“gRNA”可互换使用以指以单个分子或两个或更多个

分子的复合物形式存在的向导RNA。通常,作为单个RNA种类存在的gRNA包括两个结构域:(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);和(2)结合Cas9蛋白的结构域。在一些实施方案中,结构域(2)对应于已知为tracrRNA的序列,并且包括茎环结构。例如,在一些实施方案中,结构域(2)与如Jinek et al.,Science337:816-821(2012)中提供的tracrRNA相同或同源,其全部内容通过引用方式并入本文。gRNA的其他实例(例如,那些包含结构域2者)可以在于2013年9月6日提交的美国临时专利申请U.S.S.N.61/874,682(题为"可切换的Cas9核酸酶及其用途(Switchable Cas9Nucleases and Uses Thereof"))以及于2013年9月6日提交的美国临时专利申请U.S.S.N.61/874,746(题为"功能性核酸酶递送(Delivery System For Functional Nucleases"))中找到,各自的全部内容通过引用方式并入本文。在一些实施方案中,gRNA包括结构域(1)和(2)中的两个或更多个,并且可以被称为"延长的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包括与目标位点互补的核苷酸序列,其介导核酸酶/RNA复合物与所述目标位点的结合,提供核酸酶:RNA复合物的序列特异性。

[0941] 在一些实施方案中,RNA可编程的核酸酶是(CRISPR相关的系统)Cas9核酸内切酶,例如来自化脓链球菌的Cas9(Casnl)(参见,例如,"Complete genome sequence of an M1 strain of Streptococcus pyogenes."Ferretti J.J.,et al.,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.,et al.,Nature471:602-607(2011)。

[0942] 由于RNA可编程的核酸酶(例如Cas9)使用RNA:DNA杂交来靶向DNA切割位点,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可编程的核酸酶(诸如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.et ak,m Multiplex genome engineering using CRISPR/Cas systems.Science 339,819-823(2013);Mali,P.et al.,RNA-guided human genome engineering via Cas9.Science 339,823-826(2013);Hwang,W.Y.et al.,Efficient genome editing in zebrafish using a CRISPR-Cas system.Nature biotechnology 31,227-229(2013);Jinek,M.et al.,RNA-programmed genome editing in human cells.eLife 2,e00471(2013);Dicarlo,J.E.et al.,Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems.Nucleic acids research(2013);Jiang,W.et al.,RNA-guided editing of bacterial genomes using CRISPR-Cas systems.Nature biotechnology 31,233-239(2013);其全部内容通过引用方式并入本文)。

[0943] 术语"单核苷酸多态性(SNP)"是发生在基因组中特定位置的单个核苷酸的变异,其中各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,所述位置被A占据。这意味着在所述特定位置存在SNP,并且两个可能的核苷酸变异C或A被称为所述位置的等位基因。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)。在一些实施方案中,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,而非同义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(single nucleotide variant,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。

[0944] "特异性结合"是指识别并结合以下多肽和/或核酸分子的核酸分子、多肽或其复合物(例如,核酸可编程的DNA结合结构域和向导核酸)、化合物或分子本发明,但基本上不识别和结合样本中的其他分子,例如生物样本。

[0945] 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"实质同一性"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,但将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,本文所述基因)或其部分之间配对以形成双链分子。(参见,例如Wahl,G.M.and S.L.Berger(1987)Methods Enzymol.152:399;Kimmel,A.R.(1987)Methods Enzymol.152:507).

[0946] 例如,严格盐浓度通常低于约750mM NaCl和75mM柠檬酸三钠,优选低于约500mM NaCl和50mM柠檬酸三钠,更优选低于约250mM NaCl和25mM柠檬酸三钠。可以在不存在有机溶剂例如甲酰胺的情况下获得低严格杂交,而可以在存在至少约35%甲酰胺,更优选至少约50%甲酰胺的情况下获得高严格杂交。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。改变附加参数,例如杂交时间、去污剂浓度、例如,十二烷基硫酸钠(SDS),以及载体DNA的包括或排除是本领域技术人员公知的。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。在一个实施方案中,杂交将在30℃在750mM NaCl、75mM柠檬酸三钠和1% SDS中发生。在另一个实施方案中,杂交将在37℃在500mM NaCl、50mM柠檬酸三钠、1% SDS、35%甲酰胺及100μg/ml变性鲑鱼精子DNA(ssDNA)中发生。在另一个实施方案中,杂交将在42℃在250mM NaCl、25mM柠檬酸三钠、1% SDS、50%甲酰胺及200μg/ml ssDNA中发生。对于本领域技术人员来说,这些条件的有用变化将是显而易见的。

[0947] 对于大多数应用,杂交后的洗涤步骤在严格性方面也将有所不同。洗涤严格条件可以通过盐浓度和温度来定义。如上,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 NaCl和3mM柠檬酸三钠,最优选小于约15mM NaCl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。在一个实施方案中,将发生洗涤步骤在25℃,在30mM NaCl、3mM柠檬酸三钠及0.1%SDS中。在更优选的实施方案中,洗涤步骤将在42℃在15mM NaCl、1.5mM柠檬酸三钠和0.1% SDS中进行。在更优选的实施方案中,洗涤步骤将在68℃在15mM NaCl、1.5mM柠檬酸三钠及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

and Davis(Science 196:180,1977);Grunstein and Hogness(Proc.Natl.Acad.Sci.,USA 72:3961,1975);Ausubel et al.(Current Protocols in Molecular Biology,Wiley Interscience,New York,2001);Berger and Kimmel(Guide to Molecular Cloning Techniques,1987,Academic Press,New York);及Sambrook et al.,Molecular Cloning: A Laboratory Manual,Cold Spring Harbor Laboratory Press,New York。

[0948] “分裂”是指分成两个或更多个片段。

[0949] “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端片段和C端片段提供的Cas9蛋白。对应于Cas9蛋白的N端部分和C端部分的多肽可以被剪接以形成“重建的”Cas9蛋白。在特定实施方案中,Cas9蛋白质在蛋白质的无序区域内被分成两个片段,例如,在Nishimasu et al.,Cell,Volume 156,Issue 5,pp.935-949,2014中所述,或如在Jiang et al.(2016)Science 351:867-871.PDB file:5F9R中所述,各自通过引用方式并入本文。在一些实施方案中,所述蛋白质在SpCas9区域内约氨基酸A292至G364、F445至K483或E565至T637之间的任何C、T、A或S处,或在任何其他Cas9、Cas9变体(例如,nCas9、dCas9)或其他napDNAbp处分成两个片段。在一些实施方案中,蛋白质在SpCas9 T310、T313、A456、S469或C574处被分成两个片段。在一些实施方案中,将蛋白质分成两个片段的过程称为“分裂”的蛋白质。

[0950] 在其他实施方案中,Cas9蛋白的N端部分包括氨基酸1至573或1至637化脓链球菌Cas9野生型(SpCas9)(NCBI参考序列:NC_002737.2,Uniprot参考序列:Q99ZW2)和Cas9蛋白的C端部分包括SpCas9野生型的氨基酸574至1368或638至1368的一部分。

[0951] 分裂的Cas9的C端部分可以与分裂的Cas9的N端部分连接以形成完整的Cas9蛋白。在一些实施方案中,Cas9蛋白的C端部分从Cas9蛋白的N端部分结束的地方开始。因此,在一些实施方案中,分裂的Cas9的C端部分包括spCas9的氨基酸(551-651)-1368的一部分。“(551-651)-1368”是指起始于氨基酸551至651(包含的)之间的氨基酸并终止于氨基酸1368。例如,分裂的Cas9的C端部分可以包括spCas9之任何一个氨基酸551至1368、552至1368、553至1368、554至1368、555至1368、556至1368、557至1368、558至1368、559至1368、560至1368、561至1368、562至1368、563至1368、564至1368、565至1368、566至1368、567至1368、568至1368、569至1368、570至1368、571至1368、572至1368、573至1368、574至1368、575至1368、576至1368、577至1368、578至1368、579至1368、580至1368、581至1368、582至1368、583至1368、584至1368、585至1368、586至1368、587至1368、588至1368、589至1368、590至1368、591至1368、592至1368、593至1368、594至1368、595至1368、596至1368、597至1368、598至1368、599至1368、600至1368、601至1368、602至1368、603至1368、604至1368、605至1368、606至1368、607至1368、608至1368、609至1368、610至1368、611至1368、612至1368、613至1368、614至1368、615至1368、616至1368、617至1368、618至1368、619至1368、620至1368、621至1368、622至1368、623至1368、624至1368、625至1368、626至1368、627至1368、628至1368、629至1368、630至1368、631至1368、632至1368、633至1368、634至1368、635至1368、636至1368、637至1368、638至1368、639至1368、640至1368、641至1368、642至1368、643至1368、644至1368、645至1368、646至1368、647至1368、648至1368、649至1368、650至1368、或651至1368的一部分。在一些实施方案中,分裂的Cas9蛋白的C端部分包括SpCas9之氨基酸574至1368或638至1368。

[0952] “受试者”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。受试者包括家畜、饲养以生产劳动和提供商品诸如食物的驯养动物,包括但不限于牛、山羊、鸡、马、猪、兔及绵羊。

[0953] “实质上相同”是指多肽或核酸分子与参考氨基酸序列(例如,本文所载任一个氨基酸序列)或核酸序列(例如,本文所载任一个氨基酸序列)表现出至少50%的同一性。在一个实施方案中,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。

[0954] 序列同一性通常使用序列分析软件(例如,遗传学计算机组的序列分析软件包(Sequence Analysis Software Package of the Genetics Computer Group),University of Wisconsin Biotechnology Center,1710University Avenue,Madison,Wis.53705、序列相似性搜索、BESTFIT、GAP或PILEUP/PRETTYBOX程序)。此类软件通过为各种取代、删除、和/或其他修饰分配同源性程度来匹配相同或相似的序列。保守取代通常包括以下群组内的取代: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实施例性方法中,可以使用序列相似性搜索程序,其中e^{-3}和e^{-100}之间的概率分数表示密切相关的序列。

[0955] 例如,COBALT与以下参数一起使用:

[0956] a)对齐参数:间隙罚分-11、-1和端间隙罚分-5、-1,

[0957] b)CDD参数:使用RPS序列相似性搜索(BLAST);序列相似性搜索E-值0.003;查找保守列并重新计算,以及

[0958] c)查询集群参数:使用查询集群;字号4;最大集群距离0.8;字母常规。

[0959] 例如,使用EMBOSS Needle时具有以下参数:

[0960] a)矩阵:BLOSUM62;

[0961] b)缺口开放:10;

[0962] c)缺口延长:0.5;

[0963] d)输出格式:对;

[0964] e)末端缺口罚分:错误;

[0965] f)末端缺口开放:10;以及

[0966] g)末端缺口延长:0.5。

[0967] 术语“目标位点”是指被核碱基编辑器修饰的核酸分子内的序列。在一个实施方案中,目标点被脱氨酶或包括脱氨酶(例如,胞苷或腺嘌呤脱氨酶)的融合蛋白脱氨。

[0968] 如本文所用,术语“治疗(treat)”、“治疗(treating)”、“治疗(treatment)”等是指减少或改善病症和/或与其相关的症状或获得期望的药理效果和/或生理效果。将理解尽管不排除,治疗病症或病症并不要求完全消除病症、病况或与其相关的症状。在一些实施方案中,所述作用是治疗性的,即不限于所述作用部分或完全减少、减弱、消除、减缓、减轻、降低疾病和/或可归因于所述疾病的不利症状的强度或治愈所述疾病和/或不利症状。在一些实施方案中,该作用是预防性的,即所述作用保护或防止疾病或病况的发生或复发。为此,目前公开的方法包括给药治疗有效量的如本文所述组合物。

[0969] “尿嘧啶糖基化酶抑制剂”或“UGI”是指抑制尿嘧啶-切除修复系统的药剂。在一个

实施方案中,试剂是结合宿主尿嘧啶-DNA糖基化酶并防止从DNA去除尿嘧啶残基的蛋白质或其片段。在一个实施方案中,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。在一些实施方案中,UGI结构域包括野生型UGI或其修饰形式。在一些实施方案中,UGI结构域包括下文阐述的实施例性氨基酸序列的片段。在一些实施方案中,UGI片段包括的氨基酸序列,其包括至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%的至少下文提供的实施例性UGI序列。在一些实施方案中,UGI包括下述与实施例性UGI氨基酸序列或其片段同源的氨基酸序列。在一些实施方案中,UGI或其一部分为下述与野生型UGI或UGI序列或其部分至少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。

[0970] 实施例性的UGI包括如下氨基酸序列:

[0971] >sp|P14739|UNGI_BPPB2尿嘧啶-DNA糖基化酶抑制剂MTNLSDIIEKETGKQLVIQESIL MLPEEVEEVIGNKPESDILVHTAYDES TDENVMLLTSD APEYKPWALVIQDSNGENKIKML。

[0972] 术语“载体”是指将核酸序列引入细胞中而产生转换细胞的手段。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”是包括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。

[0973] 本文所提供的任何组合物或方法可以与本文所提供的任何其他组合物和方法中的一种或多种组合。

[0974] 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品结果,产生复杂的遗传产品群体。虽然可以通过同源性定向修复(homology directed repair,HDR)途径实现精确的、用户定义的修复结果,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,所述途径相对于竞争性、易出错的非同源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,防止有丝分裂后细胞中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。

[0975] 本文所提供的任何组合物或方法可以与本文所提供的任何其他组合物和方法中的一种或多种组合。

[0976] 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品结果,产生复杂的遗传产品群体。虽然可以通过同源性定向修复(HDR)途径实现精确的、用户定义的修复结果,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,所述途径相对于竞争性、易出错的非同源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,防止有丝分裂后细胞中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

[0977] 图1A-1C描绘了质粒。图1A是编码TadA7.10-dCas9碱基编辑器的表达载体。图1B是包括编码赋予氯霉素抗性(CamR)和奇霉素抗性(SpectR)的蛋白质的核酸分子的质粒。所述质粒还包括被两个点突变不作用的卡那霉素抗性基因。图1C是包括编码赋予氯霉素抗性(CamR)和奇霉素抗性(SpectR)的蛋白质的核酸分子的质粒。所述质粒还包括被三点突变不作用的卡那霉素抗性基因。

[0978] 图2是用图1中描绘的表达载体转导的细菌菌落的影像,其中包括有缺陷的卡那霉素抗性基因。载体包括使用易错PCR产生的ABE7.10变体。使用增加浓度的卡那霉素,选择表达这些“进化的”ABE7.10变体的细菌细胞以获得卡那霉素抗性。表达具有腺苷脱氨酶活性的ABE7.10变体的细菌能够校正引入卡那霉素抗性基因的突变,从而恢复卡那霉素抗性。选择卡那霉素抗性细胞用于进一步分析。

[0979] 图3A和3B图示了血红蛋白亚基γ(HGB1)基因座的调节区的编辑,其是胎儿血红蛋白上调的治疗相关位点。图3A是HGB1基因调节区的一部分的图。图3B量化了腺苷脱氨酶变体的效率和特异性。在HEK293T细胞中的血红蛋白亚基γ1(HGB1)位点检测编辑,其为胎儿血红蛋白上调的治疗相关位点。上图描绘了HGB1基因调控序列目标区域中的核苷酸残基。A5、A8、A9和A11表示HGB1中经过编辑的腺苷残基。

[0980] 图4说明了包括识别非规范PAM序列的dCas9的腺苷碱基编辑器的相对有效性。上图描述了血红蛋白亚基的编码序列。下图展示了腺苷脱氨酶变异碱基编辑器与不同长度的向导RNA的效率。

[0981] 图5是说明ABE8碱基编辑器的效率和特异性的图。对预期目标核苷酸和非预期目标核苷酸(旁观者)的编辑百分比进行量化。

[0982] 图6是说明ABE8碱基编辑器的效率和特异性的图。对预期目标核苷酸和非预期目标核苷酸(旁观者)的编辑百分比进行量化。

[0983] 图7A-7D描绘了第八代腺嘌呤碱基编辑器在人类细胞中介导卓越的A•T至G•C的转换。图7A说明了腺嘌呤碱基编辑的概述:i)ABE8在基因组中的sgRNA靶向位点处创建R环;ii)TadA*脱氨酶通过R环的ss-DNA部分的水解脱氨将腺嘌呤化学转换为肌苷;iii)Cas9的D10A切口酶切割与含有肌苷的链相反的链;iv)含有肌苷的链可用作DNA复制过程中的模板;v)在DNA聚合酶的背景下,肌苷优先与胞嘧啶碱基配对;及vi)复制后,肌苷可能被鸟苷替代。图7B图示ABE8.x-m和ABE8.x-d的架构。图7C图示与和tRNAArg2复合的金黄色葡萄球菌TadA(未显示)(PDB 2B3J)对齐的大肠杆菌TadA脱氨酶(PDB 1Z3A)的三个视角。在第八轮进化中确定的突变被突出显示。图7D是描绘横跨八个基因组位点的Hek293T细胞中核心ABE8构建体相对于ABE7.10构建体的A•T至G•C碱基编辑效率的图。值和误差线反映在不同日期进行的三个独立的生物重复的平均值和标准差。

[0984] 图8A至8C描绘了Cas9 PAM-变异ABE8和催化性死亡的Cas9 ABE8变体比人类细胞中的相应ABE7.10变体介导更高的A•T至G•C转换。值和误差线反映在不同日期进行的三个独立的生物重复的平均值和标准差。图8A是描绘具有NG-Cas9 ABE8(-NG PAM)的Hek293T细胞中A•T至G•C的转换的图。图8B是描绘具有Sa-Cas9 ABE8(-NNGRRT PAM)的Hek293T细胞中A•T至G•C的转换的图。图8C是描绘具有催化性失活的dCas9-ABE8(化脓链球菌Cas9中的D10A、H840A)的Hek293T细胞中A•T到G•C的转换的图。

[0985] 图9A至9E描绘了在Hek293T细胞中ABE7.10、ABEmax和ABEmax与一个BPNLS之间的靶向和脱靶编辑频率之间的比较。显示了对于在不同的日子进行的n＝3个独立的生物重复的个别数据点,误差线代表标准差。。图9A和9B是描绘在目标DNA编辑频率的图。图9B和9C是描绘sgRNA向导的DNA脱靶编辑频率的图。图9E是描绘RNA脱靶编辑频率的图。

[0986] 图10A至10B描绘HEK293T细胞中TadA、C端α-螺旋截断ABE构建体的中值A•T至G•C转换和相应的INDEL形成。图10A是描绘横跨8个基因组位点的A•T至G•C中值编辑转换的热图。图10B是描绘INDEL形成的热图。Δ残基值对应于TadA中的缺失位置。从n＝3个生物重复产生中值。

[0987] 图11是描绘横跨8个基因组位点的HEK293T细胞中40个ABE8构建体的中值A•T至G•C转换的热图。中值由两个或更多生物重复确定。

[0988] 图12是描绘横跨8个基因组位点的HEK293T细胞中40个ABE8构建体的中值INDEL％的热图。中值由两个或更多个生物重复确定。

[0989] 图13是描绘编辑中的倍数变化ABE8:ABE7的图。Hek293T细胞中八个不同基因组位点目标内所有A位置的平均ABE8:ABE7 A•T至G•C编辑的表示。位置2至12表示目标腺嘌呤在20核苷酸前间隔内的位置,位置20直接位于-NGG PAM的5’。

[0990] 图14描绘ABE8的树状图。选择用于进一步研究的核心ABE8构建体以黑色突出显示。

[0991] 图15是描绘横跨8个基因组位点的HEK293T细胞中核心8个ABE8构建体的中值A•T到G•C转换的热图。从三个或更多生物重复测定中值。

[0992] 图16是描绘在HEK293T细胞中的8个基因组位点测试的核心8个ABE8的中值INDEL频率的热图。

[0993] 图17是描绘在HEK293T细胞中的六个基因组位点处核心NG-ABE8构建体9(-NG PAM)的中值A•T至G•C转换的热图。从n＝3个生物重复产生中值。

[0994] 图18是描绘在HEK293T细胞中的六个基因组位点测试的核心NG-ABE8的中值INDEL频率的热图。从n＝3个生物重复产生的中值。

[0995] 图19是描绘HEK293T细胞中六个基因组位点处核心Sa-ABE8构建体(-NNGRRT PAM)的中值A•T至G•C转换的热图。位点位置在22核苷酸前间隔中编号为-2至20(5’到3’)。位置20是位于NNGRRT PAM的5’。从n＝3个生物重复产生中值。

[0996] 图20是描绘在HEK293T细胞中的8个基因组位点测试的核心Sa-ABE8的中值INDEL频率的热图。从n＝3个生物重复产生中值。

[0997] 图21是描绘HEK293T细胞中八个基因组位点处核心dC9-ABE8-m构建体的中值A•T至G•C转换的热图。死亡的Cas9(dC9)被定义为化脓链球菌Cas9中的D10A和H840A突变。n≥3个生物重复产生中值。

[0998] 图22是描绘HEK293T细胞中八个基因组位点处核心dC9-ABE8-d构建体的中值A•T至G•C转换的热图。死亡的Cas9(dC9)被定义为化脓链球菌Cas9中的D10A和H840A突变。n≥3个生物重复产生中值。

[0999] 图23A和23B描绘在HEK293T细胞中的8个基因组位点测试的核心dC9-ABE8的中值INDEL频率。n≥3个生物重复产生中值。图23A是描绘dC9-ABE8-m变体相对于ABE7.10的插入/缺失频率的热图。图23B是描绘dC9-ABE8-d变体相对于ABE7.10的插入/缺失频率的热

图。

[1000] 图24是描绘用ABE8和ABE7.10处理的Hek293T细胞进行C·G至T·A编辑的图。每个位点的编辑频率在目标内的所有C位置上平均。前间隔中的胞嘧啶用阴影表示。

[1001] 图25A至25H描绘通过ABE8构建体和具有改善对DNA的特异性的TadA突变的ABE8构建体进行靶上DNA和sgRNA依赖性DNA脱靶编辑。显示了对于n＝3个独立的在不同的日子进行单个数据点,误差线代表标准差生物重复。图25A和25B是描绘与ABE7相比核心ABE8构建体的靶上DNA编辑频率的图。图25C和25D是描绘具有改善RNA脱靶编辑的突变的ABE8的靶上DNA编辑频率的图。图25E和25F是描绘与ABE7相比核心ABE 8构建体的sgRNA向导的DNA脱靶编辑频率的图。图25G和25H是描绘具有改善RNA脱靶编辑的突变的ABE 8构建体的sgRNA向导的DNA脱靶编辑频率的图。

[1002] 图26是描绘人类细胞中12个先前识别的sgRNA依赖性Cas9脱靶基因座处的插入缺失频率的图。显示了对于在不同的日子进行的n＝3个独立的生物重复的个别数据点且误差条代表标准偏差。,。

[1003] 图27A和27B描绘原代细胞中A·T到G·C的转换和表型结果。图27A是描绘用来自两个不同供体的ABE处理的CD34+细胞中-198HBG1/2位点的A·T至G·C的转换的图。在处理后48和144小时进行NGS分析。-198HBG1/2目标序列,其中A7突出显示。绘制A7的A·T至G·C的百分比。图27B是描绘形成为α-球蛋白的分数的γ-球蛋白的百分比的图。显示来自两个不同供体的值,ABE治疗后和红细胞分化。

[1004] 图28A和28B描绘在HBG1/2上游的-198启动子位点处用ABE8处理的CD34+细胞从A·T至G·C的转换。图28A是描绘在编辑器处理后48和144小时来自两个供体的CD34+细胞中ABE8的A至G编辑频率的热图,其中供体2对于镰状细胞病是杂合的。图28B是含有仅A7编辑或组合的(A7+A8)编辑的总测序读数分布的图形表示。

[1005] 图29是描绘在γ-球蛋白启动子的-198位点处用ABE8处理的CD34+细胞的INDEL频率的热图。在48小时和144小时时间点显示来自两个供体的频率。

[1006] 图30描绘未处理的分化CD34+细胞(供体1)的UHPLC UV-Vis迹线(220nm)和球蛋白链水平的整合。

[1007] 图31描绘用ABE7.10-m(供体1)处理的分化的CD34+细胞的UHPLC UV-Vis迹线(220nm)和球蛋白链水平的整合。

[1008] 图32描绘用ABE7.10-d(供体1)处理的分化的CD34+细胞的UHPLC UV-Vis迹线(220nm)和球蛋白链水平的整合。

[1009] 图33描绘用ABE8.8-m(供体1)处理的分化的CD34+细胞的UHPLC UV-Vis迹线(220nm)和球蛋白链水平的整合。

[1010] 图34描绘用ABE8.8-d(供体1)处理的分化的CD34+细胞的UHPLC UV-Vis迹线(220nm)和球蛋白链水平的整合。

[1011] 图35描绘用ABE8.13-m(供体1)处理的分化的CD34+细胞的UHPLC UV-Vis迹线(220nm)和球蛋白链水平的整合。

[1012] 图36描绘用ABE8.13-d(供体1)处理的分化的CD34+细胞的UHPLC UV-Vis迹线(220nm)和球蛋白链水平的整合。

[1013] 图37描绘用ABE8.17-m(供体1)处理的分化的CD34+细胞的UHPLC UV-Vis迹线

(220nm)和球蛋白链水平的整合。

[1014] 图38描绘用ABE8.17-d(供体1)处理的分化的CD34+细胞的UHPLC UV-Vis迹线(220nm)和球蛋白链水平的整合。

[1015] 图39描绘用ABE8.20-m(供体1)处理的分化的CD34+细胞的UHPLC UV-Vis迹线(220nm)和球蛋白链水平的整合。

[1016] 图40描绘用ABE8.20-d处理的分化的CD34+细胞(供体1)的UHPLC UV-Vis迹线(220nm)和球蛋白链水平的整合。

[1017] 图41描绘未处理的分化的CD34+细胞(供体2)的UHPLC UV-Vis迹线(220nm)和球蛋白链水平的整合。注意:供体2是镰状细胞病的杂合子。

[1018] 图42描绘用ABE7.10-m处理的分化的CD34+细胞(供体2)的UHPLC UV-Vis迹线(220nm)和球蛋白链水平的整合。注意:供体2对于镰状细胞病是杂合的。

[1019] 图43描绘用ABE7.10-d处理的分化的CD34+细胞(供体2)的UHPLC UV-Vis迹线(220nm)和球蛋白链水平的整合。注意:供体2对于镰状细胞病是杂合的。

[1020] 图44描绘用ABE8.8-m处理的分化的CD34+细胞(供体2)的UHPLC UV-Vis迹线(220nm)和球蛋白链水平的整合。注意:供体2对于镰状细胞病是杂合的。

[1021] 图45描绘用ABE8.8-d处理的分化的CD34+细胞(供体2)的UHPLC UV-Vis迹线(220nm)和球蛋白链水平的整合。注意:供体2对于镰状细胞病是杂合的。

[1022] 图46描绘用ABE8.13-m处理的分化的CD34+细胞(供体2)的UHPLC UV-Vis迹线(220nm)和球蛋白链水平的整合。注意:供体2对于镰状细胞病是杂合的。

[1023] 图47描绘用ABE8.13-d处理的分化的CD34+细胞(供体2)的UHPLC UV-Vis迹线(220nm)和球蛋白链水平的整合。注意:供体2对于是镰状细胞病是杂合的。

[1024] 图48描绘用ABE8.17-m处理的分化的CD34+细胞(供体2)的UHPLC UV-Vis迹线(220nm)和球蛋白链水平的整合。注意:供体2对于是镰状细胞病是杂合的。

[1025] 图49描绘了用ABE8.17-d处理的分化的CD34+细胞(供体2)的UHPLC UV-Vis迹线(220nm)和球蛋白链水平的整合。注意:供体2对于是镰状细胞病是杂合的。

[1026] 图50描绘用ABE8.20-m处理的分化的CD34+细胞(供体2)的UHPLC UV-Vis迹线(220nm)和球蛋白链水平的整合。注意:供体2对于镰状细胞病是杂合的。

[1027] 图51A至51E描绘在两个独立位点用ABE8.8编辑在去核前在红细胞分化后第11天达到超过90％的编辑,在红细胞分化后第18天约60％的γ球蛋白超过α球蛋白或总β家族球蛋白。图51A是描绘2个独立实验中2个健康供体的ABE8.8编辑平均值的图。使用区分HBG1和HBG2的引物测量编辑效率。图51B是描绘2个独立实验中1个健康供体的平均值的图。使用识别HBG1和HBG2的引物测量编辑效率。图51C是描绘具有杂合E6V突变的供体中ABE8.8的编辑的图。图51D和51E是描绘ABE8.8编辑细胞中γ球蛋白增加的图。

[1028] 图52A和52B描绘使用ABE变体校正镰状细胞突变的百分比编辑。图52A是描绘在SCD患者成纤维细胞中具有约70％编辑的不同编辑器变体的屏幕的图。图52B是描绘来自用先导ABE变体编辑的健康供体的CD34细胞,以靶向位于编辑窗口内的相邻的脯氨酸中的同义突变A13并作为编辑SCD突变的代理的图。ABE8变体在代理A13上的平均编辑频率约为40％。

[1029] 图53A和53B描绘RNA扩增子测序以检测与ABE治疗相关的RNA中的细胞A至I编辑。

对于在不同的日子进行的n＝3个独立的生物重复显示了单个数据点,误差线代表标准差。,。图53A是描绘与ABE7和Cas9(D10A)切口酶对照相比,核心ABE 8构建体的靶向RNA扩增子中的A至I编辑频率的图。图53B是描绘具有已据报道改善RNA脱靶编辑的突变的ABE8的靶向RNA扩增子中的A至I编辑频率的图。

[1030] 图54是说明由帕金森氏病中多巴胺能神经元损失引起的多巴胺损失的示意图。

[1031] 图55是显示用于校正与帕金森氏病相关的LRRK2中的R1441C和R1441H突变的向导RNA和目标序列的示意图。

[1032] 图56是显示用于校正与帕金森氏病相关的LRRK2中的Y1699C、G2019S和I2020突变的目标序列的示意图。

[1033] 图57A至57C提供图表、示意图和表格。图57A量化LRRK2目标序列的核酸位置7处A至G的转换百分比。所使用的编辑器被指定为PV1至PV14,以下提供对此的描述。pCMV表示CMV启动子;bpNLS表示二分核定位信号;monoABE8.1表示ABE8.1碱基编辑器的单体形式。图57B描绘用于校正与帕金森氏病相关的LRRK2中的R1441C突变的目标序列和向导RNA。图57C显示LRRK2目标序列的核酸位置7处A至G的转换百分比。编辑器PV1至14用以编辑LRRKR1441C。编辑器(15至28)用以编辑G2109。用于校正LRRK突变的编辑器(PV1-28)如下:

[1034] PV1(也称为PV15).pCMV_monoABE8.1_bpNLS+Y147T

[1035] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD

[1036] PV2(也称为PV16).pCMV_monoABE8.1_bpNLS+Y147R

[1037] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCRFFRMPRQVFNAQKKAQSSTD

[1038] PV3(也称为PV17).pCMV_monoABE8.1_bpNLS+Q154S

[1039] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRSVFNAQKKAQSSTD

[1040] PV4(也称为PV18).pCMV_monoABE8.1_bpNLS+Y123H

[1041] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

[1042] PV5(也称为PV19).pCMV_monoABE8.1_bpNLS+V82S

[1043] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

[1044] PV6(也称为PV20).pCMV_monoABE8.1_bpNLS+T166R

[1045] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSRD

[1046] PV7(也称为PV21).pCMV_monoABE8.1_bpNLS+Q154R

[1047] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC YFFRMPRRVFNAQKKAQSSTD

[1048] PV8(也称为PV22).pCMV_monoABE8.1_bpNLS+Y147R_Q154R_Y123H

[1049] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLC RFFRMPRRVFNAQKKAQSSTD

[1050] PV9(也称为PV23).pCMV_monoABE8.1_bpNLS+Y147R_Q154R_I76Y

[1051] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV MQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC RFFRMPRRVFNAQKKAQSSTD

[1052] PV10(也称为PV24).pCMV_monoABE8.1_bpNLS+Y147R_Q154R_T166R

[1053] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC RFFRMPRRVFNAQKKAQSSRD

[1054] PV11(也称为PV25).pCMV_monoABE8.1_bpNLS+Y147T_Q154R

[1055] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC TFFRMPRRVFNAQKKAQSSTD

[1056] PV12(也称为PV26).pCMV_monoABE8.1_bpNLS+Y147T_Q154S

[1057] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC TFFRMPRSVFNAQKKAQSSTD

[1058] PV13(也称为PV27).pCMV_monoABE8.1_bpNLS+H123Y123H_Y147R_Q154R_I76Y

[1059] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV MQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLC RFFRMPRRVFNAQKKAQSSTD

[1060] PV14(也称为PV28).pCMV_monoABE8.1_bpNLS+V82S+Q154R

[1061] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV MQNYRLIDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC YFFRMPRRVFNAQKKAQSSTD

[1062] 图58A至58C提供图表、示意图和表格。图58A量化LRRK2目标序列的核酸位置6处A至G的转换百分比。所使用的编辑器被指定为PV15至PV28,以上提供对此的描述。pCMV表示CMV启动子;bpNLS表示二分核定位信号;monoABE8.1表示ABE8.1碱基编辑器的单体形式。图58B描绘用于校正与帕金森氏病相关的LRRK2中的G2019S突变的目标序列和向导RNA。图58C显示LRRK2目标序列的核酸位置4和6处A至G的转换百分比。位置4处的A至G转换是旁观者效应。

[1063] 图59A至59L描绘编码R1441C的LRRK2目标序列的位置7处的A至G转换的序列读数

(参见图57A至57C)。显示编辑器(PV1至14)。在图56中提供PV1至28的描述。

[1064] 图60A至60W描绘编码G2019S的LRRK2目标序列的位置4和6处的A至G转换的序列读数(参见图58A至58C)。

[1065] 图61A提供描述用于校正LRRK2中的致病突变A419V的目标序列的示意图,其由反义链G>A突变编码。使用用对TGG PAM具有特异性的SpCas9变体的向12位A的ABE校正突变。

[1066] 图61B提供描绘用于校正LRRK2中与帕金森氏病相关的致病性突变L1114L的目标序列的示意图。所述突变是使用具有胞苷脱氨酶活性(CBE)的碱基编辑器进行校正的反义链T>C。

[1067] 图61C提供描绘用于校正LRRK2中与帕金森氏病相关的致病性突变I1122V的目标序列的示意图。所述突变是使用具有胞苷脱氨酶活性(CBE)的碱基编辑器进行校正的反义链T>C。

[1068] 图61D提供描述用于校正LRRK2中与帕金森氏病相关的致病性突变M1869V的目标序列的示意图。所述突变是使用具有胞苷脱氨酶活性(CBE)的碱基编辑器进行校正的反义链T>C。

[1069] 图62A和62B描绘在HEK293T细胞中的小家鼠IDUA W401X突变精确碱基编辑校正。图62A是描绘使用用21个核苷酸向导RNA的ABE8碱基编辑器变体对小家鼠IDUA W401X突变进行碱基编辑的百分比的图。图62B是描绘使用21个核苷酸向导RNA的ABE8碱基编辑器变体的插入缺失百分比的图。

[1070] 图63是描绘使用用20个核苷酸向导RNA或21个核苷酸向导RNA的ABE8碱基编辑器变体对小家鼠IDUA W401X突变进行碱基编辑的百分比的图。

[1071] 图64描绘作为用以校正W402X突变的A至G核苷酸碱基编辑的目标的智人IDUA基因组核酸和氨基酸序列的图解说明。图中还显示了相应的向导RNA(gRNA)的核酸序列。图中指出的是IDUA核酸序列中的目标腺苷(A)核碱基(框内)。

[1072] 图65A和65B描绘HEK293T细胞中的智人IDUA W402X突变的精确碱基编辑校正。图65A是描绘使用ABE8碱基编辑器变体使用用20个核苷酸的向导RNA对智人IDUA W402X突变进行碱基编辑的百分比的图。图65B是描绘使用20个核苷酸的向导RNA的ABE8碱基编辑器变体的插入缺失百分比的图。

[1073] 图66A至660是描述如在其中已发生碱基编辑的细胞中的基因组DNA的PCR之后通过深度测序(MySeq)所检测,使用ABE8碱基编辑器变体的IDUA核酸序列中A至G核苷酸变化百分比的效率的表。图66A至66M分别描绘了使用各ABE8碱基编辑器变体ABE8.1至ABE8.13的三个样本在IDUA核酸目标位点中的位置6处的A至G碱基编辑的百分比。图66N描绘使用阳性对照碱基编辑器ABE7.10的三个样本在IDUA核酸目标位点中的位置6处的A至G碱基编辑的百分比。图660描绘使用两个阴性对照样本在IDUA核酸目标位点中的位置6处A至G碱基编辑的百分比。

[1074] 图67图示Rett/MECP2:突变校正。MECP2功能丧失—可能由许多不同的重新(de novo)突变引起。X连锁:XX患者因MECP2丢失而嵌合;XY通常会导致婴儿死亡。

[1075] 图68图示前3个向导序列的蕾特氏症R106W突变校正。

[1076] 图69图示使用具有NGTT PAM优化的编辑器的蕾特氏症R255X突变校正。

[1077] 图70A至70C:贺勒(Hurler)/IDUA突变校正。图70A图示IDUA W402X突变校正的实

验设计。图70B图示各编辑器构造的编辑百分比。图70C图示编辑和未编辑的构建体的比活性(nmol/mg/h)。

[1078] 图71描绘使用ABE 8.8的体内碱基编辑。从左到右实施例个样本:向导11(AAV9)、向导12(AAV9)、向导11(PHP.eB)、向导12(PHP.eB)及控制。

[1079] 图72A至72B。模型细胞系中ABCA4 G1961E等位基因上ABE7.10和ABE8变体的A·T至G·C转换。图72A:在21个核苷酸个核苷酸间隔sgRNA和碱基编辑器变体的质粒脂质转染后,HEK293T细胞中整合的疾病等位基因和ABCA4 G1961E密码子摆动碱基处的A·T到G·C转换。细胞在脂质转染后培育5天,然后进行编辑评估。图72B:包含ABCA4 G1961E疾病等位基因、密码子的摆动碱基和21个核苷酸间隔sgRNA使用的-NGG PAM的感兴趣的位点的DNA序列。误差棒代表三个重复标准差。。在每个数据集中,疾病等位基因在左侧,摆动碱基在右侧。

[1080] 图73.模型细胞系中ABCA4G1961E等位基因的sgRNA间隔长度变体将A·T转换为G·C。在对不同间隔长度和ABE7.10的sgRNA进行质粒脂质转染后,HEK293T细胞中整合疾病等位基因和ABCA4G1961E密码子摆动碱基处的A·T至G·C转换。细胞在脂质转染后培育5天,然后进行编辑评估。hRz=在sgRNA的5'端包括自切割锤头核酶。误差棒代表三个重复的标准差。在每个数据集中,疾病等位基因在左侧,摆动碱基在右侧。

[1081] 图74.使用拆分内含肽重构的分裂碱基编辑器的双AAV递送示意图。两个AAV粒子与碱基编辑所需的组分分开包装。一种病毒编码碱基编辑器C端区域与N端分裂内含肽融合,且互补病毒编码碱基编辑器N端区域与C端分裂内含肽融合以及sgRNA。在互补病毒共转导后,sgRNA被转录,且碱基编辑器的各半通过经由分裂内含肽的蛋白质反式剪接表达和重组。

[1082] 图75A至75B。通过双AAV递送在野生型细胞中ABCA4 G1961处的分裂ABE变体实现A·T至G·C的转换。图75A:野生型ARPE-19细胞中野生型ABCA4 G1961目标位点的A·T至G·C和C·G至T·A的转换,其中位置8A的编辑作为这些细胞中编辑的替代靶点。以每个细胞每个病毒5E+4个病毒基因组的MOI感染的细胞。细胞在感染后培育2周,然后进行编辑评估。误差棒代表六个重复的标准差。对于每个数据点,用位置8(A>G)替代位点处理的样本显示在左侧,和位置5(C>T)替代位点处理的样本显示在右侧。

[1083] 图75B:野生型目标位点的DNA序列,包含ABCA4 G1961等位基因和靶向野生型序列的21个核苷酸间隔sgRNA使用的-NGG PAM。

[1084] 图76A至76B。野生型ARPE-19细胞中的脱靶碱基编辑感染表达分裂ABE7.10的AAV2和靶向ABCA4G1961E疾病等位基因的sgRNA。图76A:与未处理的对照(灰色)相比,在与双AAV(蓝绿色)共感染2周后,目标或脱靶前间隔的最大A·T到G·C转换。图76B:与未处理的对照(灰色)相比,在与双AAV(蓝绿色)共感染2周后,靶或脱靶前间隔的最大非A·T到G·C转换。对于每个数据点,用野生型(wt)ARPE-19细胞处理的样本显示在左侧,而未经处理的wt ARPE-19细胞显示在右侧。

[1085] 如图77.由于野生型ARPE-19细胞中的碱基编辑而形成的插入缺失,所述细胞被表达分裂ABE7.10的AAV2和靶向ABCA4 G1961E的疾病等位基因的sgRNA双重感染。与未处理的对照(灰色)相比,在与双AAV(蓝绿色)共感染2周后,在目标或脱靶前间隔内或附近形成的插入缺失的百分比。对于每个数据点,用野生型(wt)ARPE-19细胞处理的样本显示在左侧,

而未经处理的wt ARPE-19细胞显示在右侧。

[1086] 图78:灵长类动物视网膜完整性和GFP在培养后第22天的表达。切片在4℃用抗视紫红质、抗GFP和生物素化花生凝集素抗体免疫标记过夜。Anc80L65.hGRK.eGFP显示GFP仅在含有光感受器的外核层(ONL)中观察到,证实了GRK启动子的光感受器特异性活性。顶行是第0天,未转导。第二行是第22天,未转导。第三行是第22天,GRK。第四行是第22天,CMB。列未染色(第1列)、DAPI(第2列)、GFP(第3列)、PNA(第4列)及视紫质(第5列)。

[1087] 图79:NHP中的Cas9表达。早在培养后第6天就在灵长类动物视网膜中检测到Cas9表达。结果显示为ABE7.10(第1和第2列)、ABE8.5(第2和第3列)和ABE8.9(第3和第4列)。顶行:培养后第6天。底行:培养后第17天。结果表示AA系统递送表达Cas9的分裂内含肽。比例尺:100μm。

[1088] 主要组件符号说明

[1089] 无。

具体实施方式

[1090] 本发明提供包括具有提高的效率的新颖腺嘌呤碱基编辑器(例如ABE8)的组合物和使用它们靶上核碱基序列中产生修饰的方法。

[1091] 核碱基编辑器

[1092] 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。本文所述的是包括多核苷酸可编程的核苷酸结合结构域(例如,Cas9)和核碱基编辑结构域(例如,腺苷脱氨酶)的核碱基编辑器或碱基编辑器。当多核苷酸可编程的核苷酸结合结构域(例如,Cas9)与结合的向导多核苷酸(例如,gRNA)结合时,可以特异性地结合目标多核苷酸序列(即,经由结合的向导核酸和目标多核苷酸序列的碱基的互补碱基配对),且从而将碱基编辑器定位到需要编辑的目标核酸序列。在一些实施方案中,目标多核苷酸序列包括单链DNA或双链DNA。在一些实施方案中,目标多核苷酸序列包括RNA。在一些实施方案中,目标多核苷酸序列包括DNA-RNA杂交体。

[1093] 多核苷酸可编程的核苷酸结合结构域

[1094] 应当理解,多核苷酸可编程的核苷酸结合结构域还可以包括结合RNA的核酸可编程的蛋白质。例如,多核苷酸可编程的核苷酸结合结构域可以与将多核苷酸可编程的核苷酸结合结构域向导至RNA的核酸相关联。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。

[1095] 碱基编辑器的多核苷酸可编程的核苷酸结合结构域本身可以包括一个或多个域。例如,多核苷酸可编程的核苷酸结合结构域可包括一个或多个核酸酶结构域。在一些实施方案中,多核苷酸可编程的核苷酸结合结构域的核酸酶结构域可包括核酸内切酶或核酸外切酶。在本文中,术语“外切核酸酶”是指能够从游离端酶切核酸(例如,RNA或DNA)的蛋白质或多肽,并且术语“内切核酸酶”是指能够催化(例如,切割)内部区域的蛋白质或多肽。在一些实施方案中,核酸内切酶可以切割双链核酸的单链。在一些实施方案中,内切核酸酶可以切割双链核酸分子的两条链。在一些实施方案中,多核苷酸可编程的核苷酸结合结构域可以是脱氧核糖核酸酶。在一些实施方案中,多核苷酸可编程的核苷酸结合结构域可以是核糖核酸酶。

[1096] 在一些实施方案中,多核苷酸可编程的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。在一些实施方案中,多核苷酸可编程的核苷酸结合结构域可包括切口酶结构域。在本文中,术语“切口酶”是指包括核酸酶结构域的多核苷酸可编程的核苷酸结合结构域,其包括能够仅切割双链核酸分子(例如,DNA)中的两条链中的一条链。在一些实施方案中,切口酶可以通过将一个或多个突变引入活性多核苷酸可编程的核苷酸结合结构域来衍生自完全催化活性(例如,天然)形式的多核苷酸可编程的核苷酸结合结构域。例如,当多核苷酸可编程的核苷酸结合结构域包括衍生自Cas9的切口酶结构域时,衍生自Cas9的切口酶结构域可包括D10A突变和位置840的组氨酸。特此类实施方案中,残基H840保留催化活性并从而可切割核酸双链体的单链。在另一个实施方案中,衍生自Cas9的切口酶结构域可包括H840A突变,而位置10的氨基酸残基仍为D。在一些实施方案中,切口酶可通过去除切口酶活性不需要的全部或部分核酸酶结构域,衍生自完全催化活性(例如,天然)形式的多核苷酸可编程的核苷酸结合结构域。例如,在多核苷酸可编程的核苷酸结合结构域包括衍生自Cas9的切口酶结构域的情况下,衍生自Cas9的切口酶结构域可包括RuvC域或HNH结构域的全部或部分的缺失。

[1097] 实施例性催化活性Cas9的氨基酸序列如下:

[1098] MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRT

[1099] ARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPT

[1100] IYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPI

[1101] NASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSK

[1102] DTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLL

[1103] KALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ

[1104] RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS

[1105] EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRK

[1106] PAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKD

[1107] KDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI

[1108] RDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGI

[1109] LQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT

[1110] QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVP

[1111] SEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSR

[1112] MNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLE

[1113] SEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI

[1114] VWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTV

[1115] AYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFEL

[1116] ENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQI

[1117] SEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTK

[1118] EVLDATLIHQSITGLYETRIDLSQLGGD。

[1119] 包括包含切口酶结构域的多核苷酸可编程的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)。在一些实施方案中,被包括切口酶结构域(例如,衍生自Cas9的切口酶结构

域)的碱基编辑器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括要编辑的碱基的链相反)。在其他实施方案中,包括切口酶结构域(例如,衍生自Cas9的切口酶结构域)的碱基编辑器可以切割被靶向进行编辑的DNA分子的链。特此类实施方案中,非靶向链未被切割。

[1120] 本文还提供碱基编辑器,其包括催化性死亡的多核苷酸可编程的核苷酸结合结构域(即,不能切割目标多核苷酸序列)。在本文中,术语“催化性死亡”和“核酸酶死亡”可互换使用,以指具有导致无法切割核酸的链的多核苷酸可编程的核苷酸结合结构域的一个或多个突变和/或缺失。在一些实施方案中,催化性死亡的多核苷酸可编程的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,在碱基编辑器包括Cas9结构域的情况下,Cas9可以包括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。在其他实施方案中,催化性死亡的多核苷酸可编程的核苷酸结合结构域可包括催化结构域(例如,RuvC1和/或HNH结构域)的全部或部分的一个或多个缺失。在进一步的实施方案中,催化性死亡的多核苷酸可编程的核苷酸结合结构域包括点突变(例如,D10A或H840A)以及全部或部分核酸酶结构域的缺失。

[1121] 本文还考虑了能够从多核苷酸可编程的核苷酸结合结构域的先前功能性版本产生催化性死亡的多核苷酸可编程的核苷酸结合结构域的突变。例如,在催化性死亡的Cas9(“dCas9”)的情况下,提供具有导致核酸酶失活的Cas9的除了D10A和H840A以外的突变的变体。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取代(例如,HNH核酸酶子域和/或RuvC1子域中的取代)。基于本公开内容和本领域的知识,其他合适的无核酸酶活性的dCas9结构域对于本领域技术人员来说是显而易见的,并且在本公开内容的范围内。此类额外的实施例性合适的核酸酶失活Cas9结构域包括但不限于D10A/H840A、D10A/D839A/H840A及D10A/D839A/H840A/N863A突变域(参见,Prashant et al.,CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering.Nature Biotechnology.2013;31(9):833-838,其全部内容通过引用方式并入本文)。

[1122] 可并入碱基编辑器的多核苷酸可编程的核苷酸结合结构域的非限制性实例包括衍生自CRISPR蛋白的结构域、限制性核酸酶、大范围核酸酶、TAL核酸酶(TALEN)及锌指核酸酶(ZFN)。在一些实施方案中,碱基编辑器包括多核苷酸可编程的核苷酸结合结构域,其包括天然或修饰的蛋白质或其部分,其经由结合的向导核酸能够在CRISPR(即,成簇的规则间隔的短回文重复序列)期间结合核酸序列介导的核酸修饰。这种蛋白质在本文中称为“CRISPR蛋白质”。因此,本文公开包括包含CRISPR蛋白的全部或部分的多核苷酸可编程的核苷酸结合结构域的碱基编辑器(即,包括CRISPR蛋白的全部或部分作为结构域的碱基编辑器,也称为“衍生自CRISPR蛋白的结构域”)。与CRISPR蛋白的野生型或天然版本相比,可以对掺入碱基编辑器的CRISPR蛋白源结构域进行修饰。例如,如下所述,衍生自CRISPR蛋白的结构域可包括一个或多个相对于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。

[1123] CRISPR是一种适应性免疫系统,其针对移动性遗传组件(病毒、转座组件及接合质粒)提供保护。CRISPR簇含有间隔、与先行移动性组件互补的序列及目标入侵核酸。CRISPR簇被转录并加工成CRISPR RNA(crRNA)。在II类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)及Cas9蛋白。tracrRNA作为核糖核酸酶3辅助处理的pre-crRNA的向导。随后,Cas9/crRNA/tracrRNA核酸内切切割与间隔互补的线性或环状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"gNRA")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.,et al.,Science 337:816-821(2012),其全部内容通过引用方式并入本文。Cas9识别CRISPR重复序列(PAM或前间隔邻近基序)中的短基序,以帮助区分自我与非自我。

[1124] 在一些实施方案中,本文所述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个核苷酸间隔所组成,所述间隔定义了欲修改的基因组目标。因此,本领域技术人员可以改变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。

[1125] 在一些实施方案中,gRNA支架序列如下:GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUUUU。

[1126] 在一些实施方案中,掺入碱基编辑器的衍生自CRISPR蛋白的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)。在一些实施方案中,掺入碱基编辑器的衍生自CRISPR蛋白的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。在一些实施方案中,掺入碱基编辑器的衍生自CRISPR蛋白的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性死亡结构域。在一些实施方案中,由碱基编辑器的衍生自CRISPR蛋白的结构域结合的目标多核苷酸是DNA。在一些实施方案中,由碱基编辑器的衍生自CRISPR蛋白的结构域结合的目标多核苷酸是RNA。

[1127] 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(也已知为Csn1或Csx12)、Cas10、Csy1、Csy2、Csy3、Csy4、Cse1、Cse2、Cse3、Cse4、Cse5e、Csc1、Csc2、Csa5、Csn1、Csn2、Csm1、Csm2、Cmr3、Cmr4、Cmr5、Cmr6、Cmr1、Cmr3、Cmr4、Cmr5、Cmr6、Csb1、Csb2、Csb3、Csx17、Csx14、Csx10、Csx16、CsaX、Csx3、Csx1、Csx1S、Csf1、Csf2、Csf0、Csf4、Csd1、Csd2、Cst1、Cst2、Csh1、Csh2、Csa1、Csa2、Csa3、Csa4、Csa5、Cas12a/Cpf1、Cas12b/C2c1、Cas12c/C2c3、Cas12d/CasY、Cas12e/CasX、Cas12g、Cas12h及Cas12i、CARF、DinG、其同系物、或修饰版本。未修饰的CRISPR酶可以具有DNA切割活性,例如Cas9,其具有两个功能性核酸内切酶结构域:RuvC和HNH。CRISPR酶可以向导切割目标序列处的一条或两条链,诸如在目标序列内和/或在目标序列的互补序列内。例如,CRISPR酶可以向导切割一条或两条链的从目标序列的第一个或最后一个核苷酸的约1、2、3、4、5、6、7、8、9、10、15、20、25、50、100、200、500或更多个碱基对。

[1128] 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可指与野生型实施例性Cas9多肽(例如,来自化脓链球菌的Cas9)具有至少或至少约50%、60%、70%、80%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%的序列同一性和/或序列同源性。Cas9可以指与野生型实施例性Cas9多肽(例如,来自化脓链球菌)具有

最多或最多约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。Cas9可指Cas9蛋白的野生型或修饰形式,其可包括氨基酸变化,诸如缺失、插入、取代、变异、突变、融合、嵌合体或其任何组合。

[1129] 在一些实施方案中,碱基编辑器的衍生自CRISPR蛋白的结构域可包括来自溃疡棒状杆菌溃疡棒状杆菌(NCBI参考序列:NC_015683.1、NC_017317.1)的Cas9;白喉棒状杆菌(NCBI参考序列:NC_016782.1、NC_016786.1);食蚜蝇(NCBI参考序列:NC_021284.1);中间普氏菌(NCBI参考序列:NC_017861.1);Spiroplasma taiwanense(NCBI参考序列:NC_021846.1);鱼型链球菌(NCBI参考序列:NC_021314.1);波罗海贝尔氏菌(NCBI参考序列:NC_018010.1);冷弯曲菌(NCBI参考序列:NC_018721.1);嗜热链球菌(NCBI参考序列:YP_820832.1)、无害李斯特菌(NCBI参考序列:NP_472073.1)、空肠弯曲菌(NCBI参考序列:YP_002344900.1)或脑膜炎奈瑟菌(NCBI参考序列:YP_002342100.1)、化脓链球菌或金黄色葡萄球菌的Cas9的全部或部分。

[1130] 核碱基编辑器的Cas9域

[1131] 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 M1 strain of Streptococcus pyogenes.”Ferretti et al.,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.,et al.,Nature 471:602-607(2011);及“A programmabledual-RNA-guided DNA endonuclease in adaptive bacterial immunity.”Jinek M.,et al.,Science 337:816-821(2012),各自的全部内容通过引用方式并入本文)。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,and Charpentier,“The tracrRNA and Cas9 families of type IICRISPR-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;其全部内容以引用方式并入本文。

[1132] 在一些实施方案中,核酸可编程的DNA结合蛋白(napDNAbp)是Cas9结构域。本文提供非限制性的实施例性Cas9域。Cas9结构域可以是核酸酶活性Cas9结构域、核酸酶失活Cas9结构域(dCas9)或Cas9切口酶(nCas9)。在一些实施方案中,Cas9结构域是核酸酶活性结构域。例如,Cas9结构域可以是切割双链核酸的两条链(例如双链DNA分子的两条链)的Cas9结构域。在一些实施方案中,Cas9结构域包括如本文所载任一个氨基酸序列。在一些实施方案中,Cas9结构域包括与本文所载的任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。在一些实施方案中,Cas9结构域包括与具有本文所载任一个氨基酸序列相比1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、21、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50或更多个或更多个突变。在一些实施方案中,Cas9结构域包括具有与本文所载任一个氨基酸序列相比至少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。

[1133] 在一些实施方案中,提供包括Cas9片段的蛋白质。例如,在一些实施方案中,蛋白质包括两个Cas9结构域之一:(1)Cas9的gRNA结合结构域;或(2)Cas9的DNA切割域。在一些实施方案中,包括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。在一些实施方案中,与野生型Cas9相比,Cas9变体可具有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、4546、47、48、49、50个或更多个氨基酸变化。在一些实施方案中,Cas9变体包括Cas9的片段(例如,gRNA结合结构域或DNA切割域),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。在一些实施方案中,所述片段是至少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%或至少对应野生型Cas9氨基酸长度的至少99.5%。在一些实施方案中,所述片段的长度为至少100个氨基酸。在一些实施方案中,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。

[1134] 在一些实施方案中,本文所提供的Cas9融合蛋白包括Cas9蛋白的全长氨基酸序列,例如本文所提供的Cas9序列之一。然而,在其他实施方案中,本文所提供的融合蛋白不包括全长Cas9序列,而仅包括其一个或多个片段。本文提供合适的Cas9结构域和Cas9片段的实施例性氨基酸序列,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。

[1135] Cas9蛋白可以与向导RNA相关联,所述向导RNA将Cas9蛋白向导至与向导RNA互补的特定DNA序列。在一些实施方案中,多核苷酸可编程的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)或核酸酶失活Cas9(dCas9)。核酸可编程的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及Cas12c/C2C3。

[1136] 在一些实施方案中,野生型Cas9对应于来自化脓链球菌的Cas9(NCBI参考序列:NC_017053.1,核苷酸和氨基酸序列如下)。

[1137] ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGATCACTGATGATTATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTATCAAAAAAAATCTTATAGGGGCTCTTTTATTTGGCAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGTCTTTTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAAGTTGCTTATCATGAGAAATATCCAACTATCTATCATCTGCGAAAAAAATTGGCAGATTCTACTGATAAAGCGGATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAGATTTAAATCCTGATAATAGTGATGTGGACAAACTATTTATCCAGTTGGTACAAATCTACAATCAATTATTTGAAGAAAACCCTATTAACGCAAGTAGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCA

TTGCTCAGCTCCCCGGTGAGAAGAGAAATGGCTTGTTTGGGAATCTCATTGCTTTGTCATTGGGATTGACCCCTAA
TTTTAAATCAAATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTAGAT
AATTTATTGGCGCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGATGCTATTTTAC
TTTCAGATATCCTAAGAGTAAATAGTGAAATAACTAAGGCTCCCCTATCAGCTTCAATGATTAAGCGCTACGATGA
ACATCATCAAGACTTGACTCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTATAAAGAAATCTTTTTT
GATCAATCAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAAATTTATCAAAC
CAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTGGTGAAACTAAATCGTGAAGATTTGCTGCGCAAGCAACG
GACCTTTGACAACGGCTCTATTCCCCATCAAATTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGAC
TTTTATCCATTTTTAAAAGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTC
CATTGGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCATGGAATTTTGA
AGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACAAACTTTGATAAAAATCTTCCAAAT
GAAAAAGTACTACCAAAACATAGTTTGCTTTATGAGTATTTTACGGTTTATAACGAATTGACAAAGGTCAAATATG
TTACTGAGGGAATGCGAAAACCAGCATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAAC
AAATCGAAAAGTAACCGTTAAGCAATTAAAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGAAATT
TCAGGAGTTGAAGATAGATTTAATGCTTCATTAGGCGCCTACCATGATTTGCTAAAAATTATTAAAGATAAAGATT
TTTTGGATAATGAAGAAAATGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAAGATAGGGGGAT
GATTGAGGAAAGACTTAAAACATATGCTCACCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTAT
ACTGGTTGGGGACGTTTGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTAGATT
TTTTGAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGACATTTAAAGAAGA
TATTCAAAAAGCACAGGTGTCTGGACAAGGCCATAGTTTACATGAACAGATTGCTAACTTAGCTGGCAGTCCTGCT
ATTAAAAAAGGTATTTTACAGACTGTAAAAATTGTTGATGAACTGGTCAAAGTAATGGGGCATAAGCCAGAAAATA
TCGTTATTGAAATGGCACGTGAAAATCAGACAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAAT
CGAAGAAGGTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAATACTCAATTGCAAAATGAA
AAGCTCTATCTCTATTATCTACAAAATGGAAGAGACATGTATGTGGACCAAGAATTAGATATTAATCGTTTAAGTG
ATTATGATGTCGATCACATTGTTCCACAAAGTTTCATTAAAGACGATTCAATAGACAATAAGGTACTAACGCGTTC
TGATAAAAATCGTGGTAAATCGGATAACGTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAA
CTTCTAAACGCCAAGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTGAAC
TTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAAATCACTAAGCATGTGGCACAAATTTTGGA
TAGTCGCATGAATACTAAATACGATGAAAATGATAAACTTATTCGAGAGGTTAAAGTGATTACCTTAAAATCTAAA
TTAGTTTCTGACTTCCGAAAAGATTTCCAATTCTATAAAGTACGTGAGATTAACAATTACCATCATGCCCATGATG
CGTATCTAAATGCCGTCGTTGGAACTGCTTTGATTAAGAAATATCCAAAACTTGAATCGGAGTTTGTCTATGGTGA
TTATAAAGTTTATGATGTTCGTAAAATGATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAAAATATTTC
TTTTACTCTAATATCATGAACTTCTTCAAAACAGAAATTACACTTGCAAATGGAGAGATTCGCAAACGCCCTCTAA
TCGAAACTAATGGGGAAACTGGAGAAATTGTCTGGGATAAAGGGCGAGATTTTGCCACAGTGCGCAAAGTATTGTC
CATGCCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAATTTTACCAAAA
AGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGATCCAAAAAAATATGGTGGTTTTGATAGTCCAACGG
TAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGAAATCGAAGAAGTTAAAATCCGTTAAAGAGTTACT
AGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAAATCCGATTGACTTTTTAGAAGCTAAAGGATATAAGGAA
GTTAAAAAAGACTTAATCATTAAACTACCTAAATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGG

CTAGTGCCGGAGAATTACAAAAAGGAAATGAGCTGGCTCTGCCAA GCAAATATGTGAATTTTTTATATTTAGCTA
GTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTTGTGGAGCAGCATAAGCATTAT
TTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCGTGTTATTTTAGCAGATGCCAATTTAGATAAAGTTCT
TAGTGCATATAACAAACATAGAGACAAACCAATACGTGAACAAGCAGAAAATATTATTCATTTATTTACGTTGACGA
ATCTTGGAGCTCCCGCTGCTTTTAAATATTTTGATACAACAATTGATCGTAAACGATATACGTCTACAAAAGAAGTT
TTAGATGCCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAGCTAGGAGGTGA
CTGA

MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD
NSDVDKLFIQLVQIYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNG
LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKN
LSDAILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ
SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQ
IHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET
ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNAS
LGAYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRGMIEERLKTYAHLFDDKVMK
QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDI
[1138] QKAQVSGQGHSLHEQIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQ
TTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQE
LDINRLSDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQL
LNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDE
NDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKL
ESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLI
ETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIAR
KKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDF
LEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS
HYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRD
KPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYET
RIDLSQLGGD

[1139] (单底线:HNH结构域;双底线:RuvC域)

[1140] 在一些实施方案中,野生型Cas9对应于或包括以下核苷酸和/或氨基酸序列:

[1141] ATGGATAAAAAGTATTCTATTGGTTTAGACATCGGCACTAATTCCGTTGGATGGGCTGT

[1142] CATAACCGATGAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACC

[1143] GTCATTCGATTAAAAAGAATCTTATCGGTGCCCTCCTATTCGATAGTGGCGAAACGGCA

[1144] GAGGCGACTCGCCTGAAACGAACCGCTCGGAGAAGGTATACACGTCGCAAGAACCGAAT

[1145] ATGTTACTTACAAGAAATTTTTAGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTTC

[1146] ACCGTTTGGAAGAGTCCTTCCTTGTCGAAGAGGACAAGAAACATGAACGGCACCCCATC

[1147] TTTGGAAACATAGTAGATGAGGTGGCATATCATGAAAAGTACCCAACGATTTATCACCT

[1148] CAGAAAAAAGCTAGTTGACTCAACTGATAAAGCGGACCTGAGGTTAATCTACTTGGCTC

[1149] TTGCCCATATGATAAAGTTCCGTGGGCACTTTCTCATTGAGGGTGATCTAAATCCGGAC

[1150] AACTCGGATGTCGACAAACTGTTCATCCAGTTAGTACAAACCTATAATCAGTTGTTTGA

[1151] AGAGAACCCTATAAATGCAAGTGGCGTGGATGCGAAGGCTATTCTTAGCGCCCGCCTCT

[1152] CTAAATCCCGACGGCTAGAAAACCTGATCGCACAATTACCCGGAGAGAAGAAAAATGGG

[1153] TTGTTCGGTAACCTTATAGCGCTCTCACTAGGCCTGACACCAAATTTTAAGTCGAACTT

[1154] CGACTTAGCTGAAGATGCCAAATTGCAGCTTAGTAAGGACACGTACGATGACGATCTCG

[1155] ACAATCTACTGGCACAAATTGGAGATCAGTATGCGGACTTATTTTTGGCTGCCAAAAAC

[1156] CTTAGCGATGCAATCCTCCTATCTGACATACTGAGAGTTAATACTGAGATTACCAAGGC

[1157] GCCGTTATCCGCTTCAATGATCAAAAGGTACGATGAACATCACCAAGACTTGACACTTC

[1158] TCAAGGCCCTAGTCCGTCAGCAACTGCCTGAGAAATATAAGGAAATATTCTTTGATCAG

[1159] TCGAAAAACGGGTACGCAGGTTATATTGACGGCGGAGCGAGTCAAGAGGAATTCTACAA

[1160] GTTTATCAAACCCATATTAGAGAAGATGGATGGGACGGAAGAGTTGCTTGTAAAACTCA

[1161] ATCGCGAAGATCTACTGCGAAAGCAGCGGACTTTCGACAACGGTAGCATTCCACATCAA

[1162] ATCCACTTAGGCGAATTGCATGCTATACTTAGAAGGCAGGAGGATTTTTATCCGTTCCT

[1163] CAAAGACAATCGTGAAAAGATTGAGAAAATCCTAACCTTTCGCATACCTTACTATGTGG

[1164] GACCCCTGGCCCGAGGGAACTCTCGGTTCGCATGGATGACAAGAAAGTCCGAAGAAACG

[1165] ATTACTCCATGGAATTTTGAGGAAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTTCAT

[1166] CGAGAGGATGACCAACTTTGACAAGAATTTACCGAACGAAAAAGTATTGCCTAAGCACA

[1167] GTTTACTTTACGAGTATTTCACAGTGTACAATGAACTCACGAAAGTTAAGTATGTCACT

[1168] GAGGGCATGCGTAAACCCGCCTTTCTAAGCGGAGAACAGAAGAAAGCAATAGTAGATCT

[1169] GTTATTCAAGACCAACCGCAAAGTGACAGTTAAGCAATTGAAAGAGGACTACTTTAAGA

[1170] AAATTGAATGCTTCGATTCTGTCGAGATCTCCGGGGTAGAAGATCGATTTAATGCGTCA

[1171] CTTGGTACGTATCATGACCTCCTAAAGATAATTAAAGATAAGGACTTCCTGGATAACGA

[1172] AGAGAATGAAGATATCTTAGAAGATATAGTGTTGACTCTTACCCTCTTTGAAGATCGGG

[1173] AAATGATTGAGGAAAGACTAAAAACATACGCTCACCTGTTCGACGATAAGGTTATGAAA

[1174] CAGTTAAAGAGGCGTCGCTATACGGGCTGGGGACGATTGTCGCGGAAACTTATCAACGG

[1175] GATAAGAGACAAGCAAAGTGGTAAAACTATTCTCGATTTTCTAAAGAGCGACGGCTTCG

[1176] CCAATAGGAACTTTATGCAGCTGATCCATGATGACTCTTTAACCTTCAAAGAGGATATA

[1177] CAAAAGGCACAGGTTTCCGGACAAGGGGACTCATTGCACGAACATATTGCGAATCTTGC

[1178] TGGTTCGCCAGCCATCAAAAAGGGCATACTCCAGACAGTCAAAGTAGTGGATGAGCTAG

[1179] TTAAGGTCATGGGACGTCACAAACCGGAAAACATTGTAATCGAGATGGCACGCGAAAAT

[1180] CAAACGACTCAGAAGGGGCAAAAAAACAGTCGAGAGCGGATGAAGAGAATAGAAGAGGG

[1181] TATTAAAGAACTGGGCAGCCAGATCTTAAAGGAGCATCCTGTGGAAAATACCCAATTGC

[1182] AGAACGAGAAACTTTACCTCTATTACCTACAAAATGGAAGGGACATGTATGTTGATCAG

[1183] GAACTGGACATAAACCGTTTATCTGATTACGACGTCGATCACATTGTACCCCAATCCTT

[1184] TTTGAAGGACGATTCAATCGACAATAAAGTGCTTACACGCTCGGATAAGAACCGAGGGA

[1185] AAAGTGACAATGTTCCAAGCGAGGAAGTCGTAAAGAAAATGAAGAACTATTGGCGGCAG

[1186] CTCCTAAATGCGAAACTGATAACGCAAAGAAAGTTCGATAACTTAACTAAAGCTGAGAG

[1187] GGGTGGCTTGTCTGAACTTGACAAGGCCGGATTTATTAAACGTCAGCTCGTGGAAACCC

[1188] GCCAAATCACAAAGCATGTTGCACAGATACTAGATTCCCGAATGAATACGAAATACGAC

[1189] GAGAACGATAAGCTGATTCGGGAAGTCAAAGTAATCACTTTAAAGTCAAAATTGGTGTC

[1190] GGACTTCAGAAAGGATTTTCAATTCTATAAAGTTAGGGAGATAAATAACTACCACCATG

[1191] CGCACGACGCTTATCTTAATGCCGTCGTAGGGACCGCACTCATTAAGAAATACCCGAAG

[1192] CTAGAAAGTGAGTTTGTGTATGGTGATTACAAAGTTTATGACGTCCGTAAGATGATCGC

[1193] GAAAAGCGAACAGGAGATAGGCAAGGCTACAGCCAAATACTTCTTTTATTCTAACATTA

[1194] TGAATTTCTTTAAGACGGAAATCACTCTGGCAAACGGAGAGATACGCAAACGACCTTTA

[1195] ATTGAAACCAATGGGGAGACAGGTGAAATCGTATGGGATAAGGGCCGGGACTTCGCGAC

[1196] GGTGAGAAAAGTTTTGTCCATGCCCCAAGTCAACATAGTAAAGAAAACTGAGGTGCAGA

[1197] CCGGAGGGTTTTCAAAGGAATCGATTCTTCCAAAAAGGAATAGTGATAAGCTCATCGCT

[1198] CGTAAAAAGGACTGGGACCCGAAAAAGTACGGTGGCTTCGATAGCCCTACAGTTGCCTA

[1199] TTCTGTCCTAGTAGTGGCAAAAGTTGAGAAGGGAAAATCCAAGAAACTGAAGTCAGTCA

[1200] AAGAATTATTGGGGATAACGATTATGGAGCGCTCGTCTTTTGAAAAGAACCCCATCGAC

[1201] TTCCTTGAGGCGAAAGGTTACAAGGAAGTAAAAAAGGATCTCATAATTAAACTACCAAA

[1202] GTATAGTCTGTTTGAGTTAGAAAATGGCCGAAAACGGATGTTGGCTAGCGCCGGAGAGC

[1203] TTCAAAAGGGGAACGAACTCGCACTACCGTCTAAATACGTGAATTTCCTGTATTTAGCG

[1204] TCCCATTACGAGAAGTTGAAAGGTTCACCTGAAGATAACGAACAGAAGCAACTTTTTGT

[1205] TGAGCAGCACAAACATTATCTCGACGAAATCATAGAGCAAATTTCGGAATTCAGTAAGA

[1206] GAGTCATCCTAGCTGATGCCAATCTGGACAAAGTATTAAGCGCATACAACAAGCACAGG

[1207] GATAAACCCATACGTGAGCAGGCGGAAAATATTATCCATTTGTTTACTCTTACCAACCT

[1208] CGGCGCTCCAGCCGCATTCAAGTATTTTGACACAACGATAGATCGCAAACGATACACTT

[1209] CTACCAAGGAGGTGCTAGACGCGACACTGATTCACCAATCCATCACGGGATTATATGAA

[1210] ACTCGGATAGATTTGTCACAGCTTGGGGGTGACGGATCCCCCAAGAAGAAGAGGAAAGT

[1211] CTCGAGCGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGG

[1212] ATGACGATGACAAGGCTGCAGGA

[1213]
MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD
NSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG
LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKN
LSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ
SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQ
IHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET
ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNAS
LGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMK
QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDI
QKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMAREN
QTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQ
LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD
ENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK
LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPL
IETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA
RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLA
SHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR
DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYE
TRIDLSQLGGD

[1214] (单底线:HNH结构域;双底线:RuvC域)。

[1215] 在一些实施方案中,野生型Cas9对应于来自化脓链球菌的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):

[1216] ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGT

[1217] GATCACTGATGAATATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACC

[1218] GCCACAGTATCAAAAAAAATCTTATAGGGGCTCTTTTATTTGACAGTGGAGAGACAGCG

[1219] GAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTAT

[1220] TTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTC

[1221] ATCGACTTGAAGAGTCTTTTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATT

[1222] TTTGGAAATATAGTAGATGAAGTTGCTTATCATGAGAAATATCCAACTATCTATCATCT

[1223] GCGAAAAAAATTGGTAGATTCTACTGATAAAGCGGATTTGCGCTTAATCTATTTGGCCT

[1224] TAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAGATTTAAATCCTGAT

[1225] AATAGTGATGTGGACAAACTATTTATCCAGTTGGTACAAACCTACAATCAATTATTTGA

[1226] AGAAAACCCTATTAACGCAAGTGGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGA

[1227] GTAAATCAAGACGATTAGAAAATCTCATTGCTCAGCTCCCCGGTGAGAAGAAAAATGGC

[1228] TTATTTGGGAATCTCATTGCTTTGTCATTGGGTTTGACCCCTAATTTTAAATCAAATTT

[1229] TGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTAG

[1230] ATAATTTATTGGCGCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAAT

[1231] TTATCAGATGCTATTTTACTTTCAGATATCCTAAGAGTAAATACTGAAATAACTAAGGC

[1232] TCCCCTATCAGCTTCAATGATTAAACGCTACGATGAACATCATCAAGACTTGACTCTTT

[1233] TAAAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTATAAAGAAATCTTTTTTGATCAA

[1234] TCAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAA

[1235] ATTTATCAAACCAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTGGTGAAACTAA

[1236] ATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACAACGGCTCTATTCCCCATCAA

[1237] ATTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTT

[1238] AAAAGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTG

[1239] GTCCATTGGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACA

[1240] ATTACCCCATGGAATTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTTAT

[1241] TGAACGCATGACAAACTTTGATAAAAATCTTCCAAATGAAAAAGTACTACCAAAACATA

[1242] GTTTGCTTTATGAGTATTTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACT

[1243] GAAGGAATGCGAAAACCAGCATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTT

[1244] ACTCTTCAAAACAAATCGAAAAGTAACCGTTAAGCAATTAAAAGAAGATTATTTCAAAA

[1245] AAATAGAATGTTTTGATAGTGTTGAAATTTCAGGAGTTGAAGATAGATTTAATGCTTCA

[1246] TTAGGTACCTACCATGATTTGCTAAAAATTATTAAAGATAAAGATTTTTTGGATAATGA

[1247] AGAAAATGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAAGATAGGG

[1248] AGATGATTGAGGAAAGACTTAAAACATATGCTCACCTCTTTGATGATAAGGTGATGAAA

[1249] CAGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTTTGTCTCGAAAATTGATTAATGG

[1250] TATTAGGGATAAGCAATCTGGCAAAACAATATTAGATTTTTTGAAATCAGATGGTTTTG

[1251] CCAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGACATTTAAAGAAGACATT

[1252] CAAAAAGCACAAGTGTCTGGACAAGGCGATAGTTTACATGAACATATTGCAAATTTAGC

[1253] TGGTAGCCCTGCTATTAAAAAAGGTATTTTACAGACTGTAAAAGTTGTTGATGAATTGG

[1254] TCAAAGTAATGGGGCGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACGTGAAAAT

[1255] CAGACAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAATCGAAGAAGG

[1256] TATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAATACTCAATTGC

[1257] AAAATGAAAAGCTCTATCTCTATTATCTCCAAAATGGAAGAGACATGTATGTGGACCAA

[1258] GAATTAGATATTAATCGTTTAAGTGATTATGATGTCGATCACATTGTTCCACAAAGTTT

[1259] CCTTAAAGACGATTCAATAGACAATAAGGTCTTAACGCGTTCTGATAAAAATCGTGGTA

[1260] AATCGGATAACGTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAA

[1261] CTTCTAAACGCCAAGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACG

[1262] TGGAGGTTTGAGTGAACTTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTC

[1263] GCCAAATCACTAAGCATGTGGCACAAATTTTGGATAGTCGCATGAATACTAAATACGAT

[1264] GAAAATGATAAACTTATTCGAGAGGTTAAAGTGATTACCTTAAAATCTAAATTAGTTTC

[1265] TGACTTCCGAAAAGATTTCCAATTCTATAAAGTACGTGAGATTAACAATTACCATCATG

[1266] CCCATGATGCGTATCTAAATGCCGTCGTTGGAACTGCTTTGATTAAGAAATATCCAAAA

[1267] CTTGAATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTCGTAAAATGATTGC

[1268] TAAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAAAATATTTCTTTTACTCTAATATCA

[1269] TGAACTTCTTCAAAACAGAAATTACACTTGCAAATGGAGAGATTCGCAAACGCCCTCTA

[1270] ATCGAAACTAATGGGGAAACTGGAGAAATTGTCTGGGATAAAGGGCGAGATTTTGCCAC

[1271] AGTGCGCAAAGTATTGTCCATGCCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAGA

[1272] CAGGCGGATTCTCCAAGGAGTCAATTTTACCAAAAAGAAATTCGGACAAGCTTATTGCT

[1273] CGTAAAAAAGACTGGGATCCAAAAAAATATGGTGGTTTTGATAGTCCAACGGTAGCTTA

[1274] TTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGAAATCGAAGAAGTTAAAATCCGTTA

[1275] AAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAAATCCGATTGAC

[1276] TTTTTAGAAGCTAAAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACCTAA

[1277] ATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAAT

[1278] TACAAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATTTAGCT

[1279] AGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTTGT

[1280] GGAGCAGCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGC

[1281] GTGTTATTTTAGCAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGA

[1282] GACAAACCAATACGTGAACAAGCAGAAAATATTATTCATTTATTTACGTTGACGAATCT

[1283] TGGAGCTCCCGCTGCTTTTAAATATTTTGATACAACAATTGATCGTAAACGATATACGT

[1284] CTACAAAAGAAGTTTTAGATGCCACTCTTATCCATCAATCCATCACTGGTCTTTATGAA

[1285] ACACGCATTGATTTGAGTCAGCTAGGAGGTGACTGA

[1286] MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI

FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD
NSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG
LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKN
LSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ
SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQ
IHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET
ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNAS
LGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMK
QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDI

[1287] QKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMAREN
QTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQ
LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD
ENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK
LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPL
IETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA
RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLA
SHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR
DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYE

[1288] TRIDLSQLGGD(单底线:HNH结构域;双底线:RuvC域)

[1289] 在一些实施方案中,Cas9是指来自以下的Cas9:溃疡棒状杆菌溃疡棒状杆菌(NCBI参考序列:NC_015683.1、NC_017317.1)的Cas9;白喉棒状杆菌(NCBI参考序列:NC_016782.1、NC_016786.1);食蚜蝇(NCBI参考序列:NC_021284.1);中间普氏菌(NCBI参考序列:NC_017861.1);Spiroplasma taiwanense(NCBI参考序列:NC_021846.1);鱼型链球菌(NCBI参考序列:NC_021314.1);波罗海贝尔氏菌(NCBI参考序列:NC_018010.1);冷弯曲菌(NCBI参考序列:NC_018721.1);嗜热链球菌(NCBI参考序列:YP_820832.1)、无害李斯特菌(NCBI参考序列:NP_472073.1)、空肠弯曲菌(NCBI参考序列:YP_002344900.1)或脑膜炎奈瑟菌(NCBI参考序列:YP_002342100.1)、化脓链球菌或金黄色葡萄球菌。

[1290] 应当理解,额外的Cas9蛋白(例如,核酸酶死亡的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。实施例性的Cas9蛋白包括但不限于那些以下提供者。在一些实施方案中,Cas9蛋白是核酸酶死亡的Cas9(dCas9)。在一些实施方案中,Cas9蛋白是Cas9切口酶(nCas9)。在一些实施方案中,Cas9蛋白是核酸酶活性的Cas9。

[1291] 在一些实施方案中,Cas9结构域是无核酸酶活性的Cas9结构域(dCas9)。例如,dCas9结构域可以结合双链核酸分子(例如,经由gRNA分子),而不切割双链核酸分子的任何

一条链。在一些实施方案中,核酸酶失活的dCas9结构域包括本文所列氨基酸序列的D10X突变和H840X突变,或本文所提供的任何氨基酸序列中的相应的突变,其中X是任何氨基酸变化。在一些实施方案中,核酸酶失活的dCas9结构域包括本文所述氨基酸序列的D10A突变和H840A突变,或本文所提供的任何氨基酸序列中的相应的突变。作为一个实例,无核酸酶活性的Cas9结构域包括在克隆载体pPlatTET-gRNA2(登录号BAV54124)中列出的氨基酸序列。

[1292] 实施例性催化失活的Cas9(dCas9)的氨基酸序列如下:

[1293] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRT

[1294] ARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPT

[1295] IYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPI

[1296] NASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSK

[1297] DTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLL

[1298] KALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ

[1299] RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS

[1300] EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRK

[1301] PAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKD

[1302] KDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI

[1303] RDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGI

[1304] LQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT

[1305] QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVP

[1306] SEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSR

[1307] MNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLE

[1308] SEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI

[1309] VWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTV

[1310] AYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFEL

[1311] ENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQI

[1312] SEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTK

[1313] EVLDATLIHQSITGLYETRIDLSQLGGD

[1314] (参见,例如,Qi et al.,“Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression.”Cell.2013;152(5):1173-83,其全部内容通过引用方式并入本文)。

[1315] 基于本公开内容和本领域知识,其他合适的核酸酶失活的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畴内。此类额外的实施例性合适的核酸酶失活的Cas9结构域包括但不限于D10A/H840A、D10A/D839A/H840A和D10A/D839A/H840A/N863A突变域(参见,例如Prashant et al.,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方式并入本文)。

[1316] 在一些实施方案中,Cas9核酸酶具有失活(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)。核酸酶失活的Cas9蛋白可互换地称为

“dCas9”蛋白(对于核酸酶-“死亡”的Cas9)或催化失活的Cas9。用于产生具有失活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 et al.,Science.337:816-821(2012);Qi et al.,“Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression”(2013)Cell.28;152(5):1173-83,各自的全部内容通过引用方式并入本文)。例如,已知Cas9的DNA切割域包括两个子域,HNH核酸酶子域和RuvC1子域。HNH子域切割与gRNA互补链,而RuvC1子域切割非互补链。这些子域内的突变可以使Cas9的核酸酶活性沉默。例如,突变D10A和H840A使化脓链球菌Cas9的核酸酶活性完全失活(Jinek et al.,Science.337:816-821(2012);Qi et al.,Cell.28;152(5):1173-83(2013))。

[1317] 在一些实施方案中,dCas9结构域包括本文所提供的dCas9结构域的任一个至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少与至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相同。在一些实施方案中本文所载氨基酸序列中的任一个相比,Cas9结构域包括具有1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、21、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50或更多个或更多个突变。在一些实施方案中,与本文所载氨基酸序列中的任一个相比,Cas9结构域包括具有至少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。

[1318] 在一些实施方案中,dCas9对应于或部分或全部包括具有一个或多个使Cas9核酸酶活性失活的突变的Cas9氨基酸序列。例如,在一些实施方案中,dCas9结构域包括D10A和H840A突变或另一个Cas9中的相应的突变。

[1319] 在一些实施方案中,dCas9包括dCas9(D10A和H840A)的氨基酸序列:

[1320]
MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD
NSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG
LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKN
LSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ
SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQ
IHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET
ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNAS
LGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMK
QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDI

QKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (单底线:HNH结构域;双底线:RuvC域)。

[1321]

[1322]

[1323] 在一些实施方案中,Cas9结构域包括D10A突变,而位置840的残基在上文提供的氨基酸序列中或在本文所提供的任何氨基酸序列中的相应位置仍为组氨酸。

[1324] 在其他实施方案中,提供具有除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核酸酶子域和/或RuvC1子域中的置换)。在一些实施方案中,提供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。在一些实施方案中,提供具有短或长约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个氨基酸或更多的胺基酸序列的dCas9变体。

[1325] 在一些实施方案中,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。在一些实施方案中,Cas9切口酶切割双链核酸分子的目标链,表示Cas9切口酶切割与结合Cas9的gRNA(例如,sgRNA)碱基配对(互补)的链。在一些实施方案中,Cas9切口酶包括D10A突变并在位置840具有组氨酸。在一些实施方案中,Cas9切口酶切割双链核酸分子的非目标、非碱基编辑链,表示Cas9切口酶切割不与结合Cas9的gRNA(例如,sgRNA)碱基配对的链。在一些实施方案中,Cas9切口酶包括H840A突变并且在位置10具有天冬氨酸残基,或相应的突变。在一些实施方案中,Cas9切口酶包括与本文所提供的任何一种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%与本文所提供的任何一种Cas9切口相同的氨基酸序列。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将是显而易见的,并且在本公开内容的范围内。

[1326] 实施例性催化Cas9切口酶(nCas9)的氨基酸序列如下:

[1327] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE

NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKP
IREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1328] 在一些实施方案中,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。在一些实施方案中,可编程的核苷酸结合蛋白可以是CasX或CasY蛋白,其已在例如Burstein et al.,"New CRISPR-Cas systems from uncultivated microbes."Cell Res.2017Feb 21.doi:10.1038/cr.2017.21中描述,其全部内容特此通过引入方式作为参考。使用基因组解析的宏基因组学,识别许多CRISPR-Cas系统,包括在古细菌领域首次报道的Cas9。这种发散的Cas9蛋白是在为活性CRISPR-Cas系统的一部分的鲜为人知的纳米古细菌中发现的。在细菌中,发现了两个以前未知的系统CRISPR-CasX和CRISPR-CasY,其为目前为止发现的最紧凑的系统之一。在一些实施方案中,在本文所述碱基编辑器系统中,Cas9被CasX或CasX的变体置换。在一些实施方案中,在本文所述碱基编辑器系统中,Cas9被CasY或CasY的变体置换。应当理解,其他RNA向导的DNA结合蛋白可以用作核酸可编程的DNA结合蛋白(napDNAbp),并且在本公开的范畴内。

[1329] 在一些实施方案中,本文所提供的任何融合蛋白的核酸可编程的DNA结合蛋白(napDNAbp)可以是CasX或CasY蛋白。在一些实施方案中,napDNAbp是CasX蛋白。在一些实施方案中,napDNAbp是CasY蛋白。在一些实施方案中,napDNAbp包括与天然存在的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%相同的氨基酸序列。在一些实施方案中,可编程的核苷酸结合蛋白是天然存在的CasX或CasY蛋白。在一些实施方案中,可编程的核苷酸结合蛋白包括本文所述任何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%的氨基酸序列、至少96%、至少97%、至少98%、至少99%或至少99.5%与相同。应当理解,根据本公开也可以使用来自其他细菌物种的CasX和CasY。

[1330] 在一些实施方案中,Cas9是脑膜炎奈瑟菌Cas9(NmeCas9)或其变体。Edraki et al.Mol.Cell.(2019)73(4):714-726中所述NmeCas9特征和PAM序列是通过引用方式整体并入本文。

[1331] 以下提供Nme1Cas9的实施例性氨基酸序列:

[1332] II类CRISPR RNA向导的核酸内切酶Cas9[脑膜炎奈瑟菌]WP_002235162.1

[1333]
```
  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
```

[1334]
```
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
```

[1335] 以下提供Nme2Cas9的实施例性氨基酸序列：

[1336] II类CRISPR RNA向导的核酸内切酶Cas9[脑膜炎奈瑟菌]WP_002230835.1

[1337]
```
  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
```

[1338] 在一些实施方案中,Cas蛋白是CasX或CasY。一个规范的CasX((uniprot.org/uniprot/F0NN87;uniprot.org/uniprot/F0NH53)tr|F0NN87|F0NN87_SULIHCRISPR相关的Casx蛋白OS＝冰岛硫化叶菌(菌株HVE10/4)GN＝SiH_0402PE＝4SV＝1)氨基酸序列如下：

[1339] MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAERRGKAKKKKG EEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKECEEVSAPSFVKPEFYEFGRSPGMVERT RRVKLEVEPHYLIIAAAGWVLTRLGKAKVSEGDYVGVNVFTPTRGILYSLIQNVNGIVPGIKPETAFGLWIARKVV SSVTNPNVSVVRIYTISDAVGQNPTTINGGFSIDLTKLLEKRYLLSERLEAIARNALSISSNMRERYIVLANYIYE YLTG SKRLEDLLYFANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG.

[1340] 实施例性CasX(>tr|F0NH53|F0NH53_SULIR CRISPR相关的蛋白Casx OS＝冰岛硫化叶菌(菌株REY15A)GN＝SiRe_0771PE＝4SV＝1)氨基酸序列如下：

[1341] MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAER

[1342] RGKAKKKKGEEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKEC

[1343] EEVSAPSFVKPEFYKFGRSPGMVERTRRVKLEVEPHYLIMAAAGWVLTRLGKAKVSEGD

[1344] YVGVNVFTPTRGILYSLIQNVNGIVPGIKPETAFGLWIARKVVSSVTNPNVSVVSIYTI

[1345] SDAVGQNPTTINGGFSIDLTKLLEKRDLLSERLEAIARNALSISSNMRERYIVLANYIY

[1346] EYLTGSKRLEDLLYFANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG。

[1347] 变形杆菌CasX

[1348] MEKRINKIRKKLSADNATKPVSRSGPMKTLLVRVMTDDLKKRLEKRRKKPEVMPQVISN

[1349] NAANNLRMLLDDYTKMKEAILQVYWQEFKDDHVGLMCKFAQPASKKIDQNKLKPEMDEK

[1350] GNLTTAGFACSQCGQPLFVYKLEQVSEKGKAYTNYFGRCNVAEHEKLILLAQLKPVKDS

[1351] DEAVTYSLGKFGQRALDFYSIHVTKESTHPVKPLAQIAGNRYASGPVGKALSDACMGTI

[1352] ASFLSKYQDIIIEHQKVVKGNQKRLESLRELAGKENLEYPSVTLPPQPHTKEGVDfAYN

[1353] EVIARVRMWVNLNLWQKLKLSRDDAKPLLRLKGFPSFPVVERRENEVDWWNTINEVKKL

[1354] IDAKRDMGRVFWSGVTAEKRNTILEGYNYLPNENDHKKREGSLENPKKPAKRQFGDLLL

[1355] YLEKKYAGDWGKVFDEAWERIDKKIAGLTSHIEREEARNAEDAQSKAVLTDWLRAKASF

[1356] VLERLKEMDEKEFYACEIQLQKWYGDLRGNPFAVEAENRVVDISGFSIGSDGHSIQYRN

[1357] LLAWKYLENGKREFYLLMNYGKKGRIRFTDGTDIKKSGKWQGLLYGGGKAKVIDLTFDP

[1358] DDEQLIILPLAFGTRQGREFIWNDLLSLETGLIKLANGRVIEKTIYNKKIGRDEPALFV

[1359] ALTFERREVVDPSNIKPVNLIGVARGENIPAVIALTDPEGCPLPEFKDSSGGPTDILRI

[1360] GEGYKEKQRAIQAAKEVEQRRAGGYSRKFASKSRNLADDMVRNSARDLFYHAVTHDAVL

[1361] VFANLSRGFGRQGKRTFMTERQYTKMEDWLTAKLAYEGLTSKTYLSKTLAQYTSKTCSN

[1362] CGFTITYADMDVMLVRLKKTSDGWATTLNNKELKAEYQITYYNRYKRQTVEKELSAELD

[1363] RLSEESGNNDISKWTKGRRDEALFLLKKRFSHRPVQEQFVCLDCGHEVHAAEQAALNIA

[1364] RSWLFLNSNSTEFKSYKSGKQPFVGAWQAFYKRRLKEVWKPNA

[1365] 实施例性CasY((ncbi.nlm.nih.gov/protein/APG80656.1)>APG80656.1 CRISPR-相关的蛋白CasY[未培养的俭菌门菌])氨基酸序列如下：

[1366] MSKRHPRISGVKGYRLHAQRLEYTGKSGAMRTIKYPLYSSPSGGRTVPREIVSAINDDY

[1367] VGLYGLSNFDDLYNAEKRNEEKVYSVLDFWYDCVQYGAVFSYTAPGLLKNVAEVRGGSY

[1368] ELTKTLKGSHLYDELQIDKVIKFLNKKEISRANGSLDKLKKDIIDCFKAEYRERHKDQC

[1369] NKLADDIKNAKKDAGASLGERQKKLFRDFFGISEQSENDKPSFTNPLNLTCCLLPFDTV

[1370] NNNRNRGEVLFNKLKEYAQKLDKNEGSLEMWEYIGIGNSGTAFSNFLGEGFLGRLRENK

[1371] ITELKKAMMDITDAWRGQEQEEELEKRLRILAALTIKLREPKFDNHWGGYRSDINGKLS

[1372] SWLQNYINQTVKIKEDLKGHKKDLKKAKEMINRFGESDTKEEAVVSSLLESIEKIVPDD

[1373] SADDEKPDIPAIAIYRRFLSDGRLTLNRFVQREDVQEALIKERLEAEKKKKPKKRKKKS

[1374] DAEDEKETIDFKELFPHLAKPLKLVPNFYGDSKRELYKKYKNAAIYTDALWKAVEKIYK

[1375] SAFSSSLKNSFFDTDFDKDFFIKRLQKIFSVYRRFNTDKWKPIVKNSFAPYCDIVSLAE

[1376] NEVLYKPKQSRSRKSAAIDKNRVRLPSTENIAKAGIALARELSVAGFDWKDLLKKEEHE

[1377] EYIDLIELHKTALALLLAVTETQLDISALDFVENGTVKDFMKTRDGNLVLEGRFLEMFS

[1378] QSIVFSELRGLAGLMSRKEFITRSAIQTMNGKQAELLYIPHEFQSAKITTPKEMSRAFL

[1379] DLAPAEFATSLEPESLSEKSLLKLKQMRYYPHYFGYELTRTGQGIDGGVAENALRLEKS

[1380] PVKKREIKCKQYKTLGRGQNKIVLYVRSSYYQTQFLEWFLHRPKNVQTDVAVSGSFLID

[1381] EKKVKTRWNYDALTVALEPVSGSERVFVSQPFTIFPEKSAEEEGQRYLGIDIGEYGIAY

[1382] TALEITGDSAKILDQNFISDPQLKTLREEVKGLKLDQRRGTFAMPSTKIARIRESLVHS

[1383] LRNRIHHLALKHKAKIVYELEVSRFEEGKQKIKKVYATLKKADVYSEIDADKNLQTTVW

[1384] GKLAVASEISASYTSQFCGACKKLWRAEMQVDETITTQELIGTVRVIKGGTLIDAIKDF

[1385] MRPPIFDENDTPFPKYRDFCDKHHISKKMRGNSCLFICPFCRANADADIQASQTIALLR

[1386] YVKEEKKVEDYFERFRKLKNIKVLGQMKKI。

[1387] Cas9核酸酶具有两个功能性核酸内切酶结构域:RuvC和HNH。Cas9在目标结合后发生构象变化,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至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端连接(non-homologous end joining,NHEJ)途径;或(2)效率较低但保真度高的同源性定向修复(HDR)途径。

[1388] 非同源端连接(NHEJ)和/或同源性定向修复(HDR)的“效率”可以通过任何方便的方法计算。例如,在一些实施方案中,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。更多裂解的底物表示更高的HDR百分比(更高的HDR效率)。作为说明性实例,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)。

[1389] 在一些实施方案中,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或缺失(indel))。更多的切割表示更高的NHEJ百分比(更高的NHEJ效率)。作为说明性实施例,NHEJ的分数(百分比)可以使用以下等式计算:$(1-(1-(b+c)/(a+b+c))^{1/2})\times 100$,其中“a”是DNA底物的条带强度,“b”和“c”是切割产物(Ran et.al.,Cell.2013Sep.12;154(6):1380-9;以及Ran et al.,Nat Protoc.2013Nov.;8(11):2281-2308)。

[1390] NHEJ修复途径是最活跃的修复机制,其经常导致DSB位点的小核苷酸插入或缺失(indel)。NHEJ介导的DSB修复的随机性具有重要的实际意义,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。在大多数实施方案中,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,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。

[1391] 虽然NHEJ介导的DSB修复通常破坏基因的开放阅读框,同源性定向修复(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。

[1392] 为了利用HDR进行基因编辑,可以使用一个或多个gRNA和Cas9或Cas9切口酶将包括所需序列的DNA修复模板递送到感兴趣的细胞类型中。修复模板可以包括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,其中更大的插入需要更长的同源臂。修复模板可以是单链寡核苷酸、双链寡核苷酸或双链DNA质粒。即使在表达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。涉及NHEJ的化学或基因抑制基因也可以增加HDR频率。

[1393] 在一些实施方案中,Cas9是修饰的Cas9。给定的gRNA靶向序列可以在整个基因组中具有其他存在部分同源性的位点。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。

[1394] 在一些实施方案中,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,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,变体Cas9多肽具有少于50%、少于40%、少于30%、少于20%、少于10%、少于5%或少于1%的核酸酶相应的野生型Cas9蛋白的活性。在一些实施方案中,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”。

[1395] 在一些实施方案中,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Cas9蛋白(例如野生型Cas9蛋白)的核酸内切酶活性。。

[1396] 在一些实施方案中,变体Cas9蛋白可以切割向导目标序列的互补链,但切割双链向导目标序列的非互补链的能力降低。例如,变体Cas9蛋白可以具有降低RuvC结构域功能的突变(氨基酸取代)。作为非限制性实例,在一些实施方案中,变体Cas9蛋白具有D10A(在氨基酸位置10的天冬氨酸成丙氨酸),并且因此可切割双链向导目标序列的互补链,但切割非双链向导目标序列的互补链(因此当变体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))(参见,例如,Jinek et al.,Science.2012Aug.17;337(6096):816-21)。

[1397] 在一些实施方案中,变体Cas9蛋白可以切割双链向导目标序列的非互补链,但切割向导目标序列的互补链的能力降低。例如,变体Cas9蛋白可以具有降低HNH结构域(RuvC/HNH/RuvC结构域基序)功能的突变(氨基酸取代)。作为非限制性实例,在一些实施方案中,变体Cas9蛋白具有H840A(在氨基酸位置840的组氨酸成丙氨酸)突变,因此可切割向导目标序列的非互补链,但切割向导目标序列的互补链的能力降低(因此,当变体Cas9蛋白切割双链向导目标序列时,产生SSB而不是DSB)。此类Cas9蛋白具有切割向导目标序列(例如,单链向导目标序列)的能力降低,但保留结合向导目标序列(例如,单链向导目标序列)的能力。

[1398] 在一些实施方案中,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,在一些实施方案中,变体Cas9蛋白包括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,单链标DNA)的能力降低,但保留结合标DNA(例如,单链目标DNA)的能力。

[1399] 作为另一个非限制性实例,在一些实施方案中,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。

[1400] 作为另一个非限制性实例,在一些实施方案中,变体Cas9蛋白含有P475A、W476A、

N477A、D1125A、W1126A和D1127A突变,使得多肽具有切割目标DNA的能力降低。此类Cas9蛋白具有切割目标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。

[1401] 作为另一个非限制性实例,在一些实施方案中,变体Cas9蛋白包括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,在一些实施方案中,变体Cas9蛋白含有H840A、D10A、W476A和W1126A突变,使得多肽具有切割目标DNA的能力降低。此类Cas9蛋白具有切割目标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。在一些实施方案中,变体Cas9在Cas9 HNH结构域(A840H)中的位置840恢复了催化性His残基。

[1402] 作为另一个非限制性实例,在一些实施方案中,变体Cas9蛋白包括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蛋白具有切割目标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,在一些实施方案中,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蛋白具有切割目标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。在一些实施方案中,当变体Cas9蛋白包括W476A和W1126A突变或当变体Cas9蛋白包括P475A、W476A、N477A、D1125A、W1126A和D1127A突变时,变体Cas9蛋白不与PAM序列有效结合。因此,在一些此类实施方案中,当此类变体Cas9蛋白用于结合方法时,所述方法不需要PAM序列。换言之,在一些实施方案中,当这种变体Cas9蛋白用于结合方法中时,所述方法可以包括向导RNA,但是所述方法可以在不存在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)。其他残基可以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。

[1403] 在一些实施方案中,具有降低的催化活性的变体Cas9蛋白(例如,当Cas9蛋白具有D10、G12、G17、E762、H840、N854、N863、H982、H983、A984、D986、和/或A987突变时,例如,D10A、G12A、G17A、E762A、H840A、N854A、N863A、H982A、H983A、A984A、和/或D986A),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向导到目标DNA序列),只要它保留与向导RNA相互作用的能力。

[1404] 在一些实施方案中,变体Cas蛋白可以是spCas9、spCas9-VRQR、spCas9-VRER、xCas9(sp)、saCas9、saCas9-KKH、spCas9-MQKSER、spCas9-LRKIQK或spCas9-LRVSQL。

[1405] 在一些实施方案中,使用包含氨基酸取代D1135M、S1136Q、G1218K、E1219F、A1322R、D1332A、R1335E和T1337R(SpCas9-MQKFRAER)并且对改变的PAM 5'-NGC具有特异性的修饰的SpCas9。

[1406] 化脓链球菌Cas9的替代品可以包括来自Cpf1家族的在哺乳动物细胞中显示出切割活性RNA向导的核酸内切酶。普雷沃氏菌属(Prevotella)和弗朗西斯氏菌属1(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向导的核酸内切酶。这种获得性免疫机制存在于普雷沃

氏菌和弗朗西斯氏菌中。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,其类似于传统的限制性酶克隆,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,Cpf1还可以将CRISPR可靶向的位点数量扩大到乏SpCas9青睐的NGG PAM位点的富含AT的区域或富含AT的基因组。Cpf1基因座包括一个混合的α/β结构域、RuvC-I、随后为螺旋区域、RuvC-II及锌指样结构域。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结构域架构显示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蛋白更类似于I类和III类,而不是来自II类系统。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复合物通过识别前间隔邻近基序5'-YTN-3'来切割目标DNA或RNA。识别PAM后,Cpf1引入了一个粘性端样DNA双链断裂,有4或5个核苷酸突出端。

[1407] 核碱基编辑器的Cas12结构域

[1408] 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类和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及Cas12i)。参见,例如,Shmakov et al.,"Discovery and Functional Characterization of Diverse Class 2CRISPR Cas Systems,"Mol.Cell,2015Nov.5;60(3):385-397;Makarova et al.,"Classification and Nomenclature of CRISPR-Cas Systems:Where from Here?"CRISPR Journal,2018,1(5):325-336;以及Yan et al.,"Functionally Diverse Type V CRISPR-Cas Systems,"Science,2019Jan.4;363:88-91;其各自全部内容特此通过引用方式并入。V型Cas蛋白含有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,例如Cas12b/C2c1需要tracrRNA来生产crRNA。Cas12b/C2c1依赖crRNA和tracrRNA进行DNA切割。

[1409] 本发明中考虑的核酸可编程的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及Cas12i、其同源物或其修饰形式。如本文所用,Cas12蛋白也可称为Cas12核酸酶、Cas12结构域或Cas12蛋白结构域。在一些实施方案中,本发明的Cas12蛋白包括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。

[1410] 在一些实施方案中,Cas12结构域是无核酸酶活性的Cas12结构域或Cas12切口酶。在一些实施方案中,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。在一些实施方案中,Cas12结构域包括如本文所载任一个氨基酸序列。在一些实施方案中,Cas12结构域包括与本文所载任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。在一些实施方案

中,Cas12结构域包括与本文所载任一个氨基酸序列相比,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或更多个突变。在一些实施方案中,Cas12结构域包括与本文所载任一个氨基酸序列相比,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。

[1411] 在一些实施方案中,提供包括Cas12片段的蛋白质。例如,在一些实施方案中,蛋白质包括两个Cas12域之一:(1)Cas12的gRNA结合结构域;或(2)Cas12的DNA切割域。在一些实施方案中,包括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、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。在一些实施方案中,与野生型Cas12相比,Cas12变体可具有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、4546、47、48、49、50或更多的氨基酸变化。在一些实施方案中,Cas12变体包括Cas12的片段(例如,gRNA结合结构域或DNA切割域),使得所述片段与野生型Cas12的相应片段至少约70%相同、至少约80%相同、至少约90%相同、至少约95%相同、至少约96%相同、至少约97%相同、至少约98%相同、至少约99%相同、至少约99.5%相同或至少约99.9%相同。在一些实施方案中,所述片段是至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相应野生型Cas12的氨基酸长度。在一些实施方案中,所述片段的长度为至少100个氨基酸。在一些实施方案中,所述片段是至少100、150、200、250、300、350、400、450、500、550、600、650、700、750、800、850、900、950、1000、1050、1150、1200、1250或至少1300个氨基酸的长度。

[1412] 在一些实施方案中,Cas12对应于或部分或全部包括具有一个或多个改变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取代。在一些实施方案中,提供Cas12的变体或同源物与野生型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。在一些实施方案中,提供具有短或长约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个氨基酸或更多个氨基酸序列的Cas12。

[1413] 在一些实施方案中,本文所提供的Cas12融合蛋白包括Cas12蛋白的全长氨基酸序列,例如本文所提供的Cas12序列之一。然而,在其他实施方案中,本文所提供的融合蛋白不包括全长Cas12序列,而仅包括其一个或多个片段。本文提供合适的Cas12域的实施例性氨基酸序列,并且Cas12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。

[1414] 通常,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 et al.,“CRISPR-Cas12a target binding unleashes indiscriminate-stranded DNase activity,”Science360:436-439(2018))。在某些情况下,Cas12蛋白是Cas12b蛋白的变

体。(参见Strecker et al.,Nature Communications,2019,10(1):Art.No.:212)。在一个实施方案中,当与野生型Cas12蛋白氨基酸序列相比时,变体Cas12多肽具有相差1、2、3、4、5个或更多个氨基酸的氨基酸序列(例如,具有缺失、插入、取代、融合)。在一些情况下,变体Cas12多肽具有降低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,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蛋白的切口酶活性的百分比。在某些情况下,变体Cas12b蛋白没有实质性的切口酶活性。

[1415] 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。

[1416] 在一些实施方案中,Cas12蛋白包括来自Cas12a/Cpf1家族的RNA向导的核酸内切酶,其在哺乳动物细胞中表现出活性。普雷沃氏菌和弗朗西斯氏菌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向导的核酸内切酶。这种获得性免疫机制存在普雷沃氏菌和弗朗西斯氏菌中。Cpf1基因与CRISPR基因座相关,编码使用向导RNA来寻找和切割病毒DNA的内切核酸酶。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,类似于传统的限制性酶克隆,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,Cpf1还可以将CRISPR可靶向的位点数量扩大到缺乏SpCas9青睐的NGG PAM位点的富含AT的区域或富含AT的基因组。Cpf1基因座包括一个混合的α/β结构域、RuvC-I随后的螺旋区域、RuvC-II及锌指样结构域。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端没有Cas9的α螺旋识别叶。Cpf1 CRISPR-Cas域架构表显示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类和III类系统相比与II类系统更相似。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引入了具有4或5个核苷酸的突出端的粘性端样DNA双链断裂,。

[1417] 在本发明的一些方面,载体编码一种CRISPR酶,其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可以使用。Cas12可指与野生型实施例性Cas12多肽(例如,来自久氏芽孢杆菌的Cas12)具有至少或至少约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可以指与野生型实施例性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))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。Cas12可以指Cas12蛋白的野生型或修饰形式,其可以包括氨基酸变化,例如缺失、插入、取代、变异、突变、融合、嵌合体或其任何组合。

[1418] 核酸可编程的DNA结合蛋白

[1419] 本公开的一些方面提供包括作用为核酸可编程的DNA结合蛋白的结构域的融合蛋白,其可用于将蛋白质(诸如碱基编辑器)向导至特定的核酸(例如,DNA或RNA)序列。在特定实施方案中,融合蛋白包括核酸可编程的DNA结合蛋白结构域和脱氨酶结构域。核酸可编程的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及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(也已知为Csn1或Csx12)、Cas10、Cas10d、Cas12a/Cpf1、Cas12b/C2c1、Cas12c/C2c3、Cas12d/CasY、Cas12e/CasX、Cas12g、Cas12h、Cas12i、Csy1、Csy2、Csy3、Csy4、Css1、Css2、Cse5e、Csc2、Csa5、Csn1、Csn2、Csm1、Csm2、Csm3、Csm4、Csm5、Csm6、Cmr1、Cmr2、Cmr3、Cmr4、Cmr5、Cmr6、Csb1、Csb2、Csb3、Csx17、Csx14、Csx10、Csx16、Csx1、Csx1S、Csx11、Csf1、Csf2、Cs0、Csf4、Csd1、Csd2、Cst1、Cst2、Csh1、Csh2、Csa1、Csa2、Csa3、Csa4、Csa5、II类Cas效应蛋白、V型Cas效应蛋白、VI类蛋白质、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。参见,例如,Makarova et al.“Classification and Nomenclature of CRISPR-Cas Systems:Where from Here?”CRISPR J.2018Oct;1:325-336.doi:10.1089/crispr.2018.0033;Yan et al.,“Functionally diverse type V CRISPR-Cas systems”Science.2019Jan 4;363(6422):88-91.doi:10.1126/science.aav7271,各自的全部内容特此通过引用方式并入。

[1420] 具有与Cas9不同的PAM特异性的核酸可编程的DNA结合蛋白的一个例子是来自普雷沃氏菌和弗朗西斯氏菌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。已经显示Cpf1介导特征与Cas9不同的强大的DNA干扰。Cpf1是缺乏tracrRNA的单个RNA向导的内切核酸酶,并且其利用富含T的前间隔邻近基序(TTN、TTTN或YTN)。此外,Cpf1经由交错的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(Acidaminococcus)和毛螺菌科(Lachnospiraceae)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,并且之前已经在例如Yamano等人,“Cpf1与向导RNA和靶DNA复合的晶体结构”。细胞(165)2016,p.949-962;其全部内容特此通过引入方式作为参考。

[1421] 在本组合物和方法中有用的是核酸酶失活的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端不具有Cas9的α-螺旋识别叶。Zetsche et al.,Cell,163,759-771,2015(通过引用方式并入本文)显示Cpf1的RuvC样结构域负责切割两条DNA链并使RuvC样结构域失活让Cpf1核酸酶活性失活。例如,与弗朗西斯氏菌Cpf1中的D917A、E1006A或D1255A对应的突变会使Cpf1核酸酶活性失活。在一些实施方案中,本公开的dCpf1包括对应于D917A、E1006A、D1255A、D917A/E1006A、D917A/D1255A、E1006A/D1255A或D917A/E1006A/D122的突变。应当理解,根据本公开可以使用使Cpf的RuvC结构域失活的任何突变,例如取代突变、缺失或插入。

[1422] 在一些实施方案中,本文所提供的任何融合蛋白的核酸可编程的DNA结合蛋白(napDNAbp)可以是Cpf1蛋白。在一些实施方案中,Cpf1蛋白是Cpf1切口酶(nCpf1)。在一些

实施方案中,Cpf1蛋白是核酸酶失活的Cpf1(dCpf1)。在一些实施方案中,Cpf1、nCpf1或dCpf1包括与本文公开的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。在一些实施方案中,dCpf1包括与本文公开的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,并且包括对应于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。

[1423] 野生型弗朗西斯菌Cpf1(D917、E1006和D1255以粗体和底线表示)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDK
YHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEY
IKDSEKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEI
IKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDK
APEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITK
FNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKS
FVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIY
FKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKA

[1424] KYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISI
KYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEH
FYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPD
NTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKV
FFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSI
SKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLY
LFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKK
ITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKF
NDEINLLLKEKANDVHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTN
YHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**E**DL
NFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETF
KKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLD

[1425] KGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEK
LLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISP
VADVNGNFFDSRQAPKNMPQDA**D**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKN
EEYFEFVQNRNN

[1426] 弗朗西斯菌Cpf1 D917A(A917、E1006和D1255为粗体和底线)

[1427]

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDK
YHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEY
IKDSEKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEI
IKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDK
APEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITK
FNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKS
FVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIY
FKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKA
KYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISI
KYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEH
FYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPD
NTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKV
FFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSI
SKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLY
LFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKK
ITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKF
NDEINLLLKEKANDVHILSI**A**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTN
YHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**E**DL
NFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETF
KKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLD
KGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEK
LLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISP
VADVNGNFFDSRQAPKNMPQDA**D**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKN
EEYFEFVQNRNN

[1428] 弗朗西斯菌Cpf1 E1006A(D917、A1006和D1255为粗体和底线)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDK
YHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEY
IKDSEKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEI
IKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDK
APEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITK
FNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKS
FVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIY
FKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKA
KYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISI
KYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEH
FYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPD
NTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKV
[1429] FFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSI
SKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLY
LFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKK
ITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKF
NDEINLLLKEKANDVHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTN
YHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**A**DL
NFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETF
KKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLD
KGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEK
LLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISP
VADVNGNFFDSRQAPKNMPQDA**D**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKN
EEYFEFVQNRNN

[1430] 弗朗西斯菌Cpf1 D1255A(D917、E1006和A1255为粗体和底线)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDK
YHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEY
IKDSEKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEI
[1431] IKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDK
APEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITK
FNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKS
FVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIY

FKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKA
KYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISI
KYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEH
FYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPD
NTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKV
FFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSI
SKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLY
LFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKK

[1432] ITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKF
NDEINLLLKEKANDVHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTN
YHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**E**DL
NFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETF
KKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLD
KGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEK
LLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISP
VADVNGNFFDSRQAPKNMPQDA**A**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKN
EEYFEFVQNRNN

[1433] 弗朗西斯菌Cpf1 D917A/E1006A(A917、A1006和D1255为粗体和底线)
MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDK
YHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEY
IKDSEKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEI
IKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDK
APEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITK
FNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKS
FVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIY

[1434] FKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKA
KYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISI
KYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEH
FYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPD
NTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKV
FFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSI
SKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLY
LFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKK

ITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKF
NDEINLLLKEKANDVHILSI**A**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTN
YHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**A**DL
NFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETF

[1435] KKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLD
KGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEK
LLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISP
VADVNGNFFDSRQAPKNMPQDA**D**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKN
EEYFEFVQNRNN

[1436] 弗朗西斯菌Cpf1 D917A/D1255A(A917、E1006和A1255为粗体和底线)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDK
YHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEY
IKDSEKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEI
IKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDK
APEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITK
FNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKS
FVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIY
FKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKA
KYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISI
KYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEH
FYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPD

[1437] NTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKV
FFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSI
SKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLY
LFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKK
ITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKF
NDEINLLLKEKANDVHILSI**A**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTN
YHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**E**DL
NFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETF
KKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLD
KGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEK
LLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISP
VADVNGNFFDSRQAPKNMPQDA**A**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKN

[1438] EEYFEFVQNRNN

[1439] 弗朗西斯菌Cpf1 E1006A/D1255A(D917、A1006和A1255为粗体和底线)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDK
YHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEY
IKDSEKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEI
IKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDK
APEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITK
FNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKS
FVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIY
FKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKA
KYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISI
KYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEH
FYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPD
NTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKV
[1440]
FFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSI
SKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLY
LFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKK
ITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKF
NDEINLLLKEKANDVHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTN
YHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**A**DL
NFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETF
KKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLD
KGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEK
LLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISP
VADVNGNFFDSRQAPKNMPQDA**A**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKN
EEYFEFVQNRNN

[1441] 弗朗西斯菌Cpf1 D917A/E1006A/D1255A(A917、A1006和A1255为粗体和底线)

[1442]

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDK
YHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEY
IKDSEKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEI
IKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDK
APEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITK
FNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKS
FVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIY
FKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKA
KYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISI
KYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEH
FYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPD
NTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKV
FFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSI
SKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLY
LFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKK
ITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKF
NDEINLLLKEKANDVHILSI**A**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTN
YHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**A**DL
NFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETF
KKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLD
KGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEK
LLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISP
VADVNGNFFDSRQAPKNMPQDA**A**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKN
EEYFEFVQNRNN

[1443] 在一些实施方案中,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。

[1444] 在一些实施方案中,Cas9结构域是来自金黄色葡萄球菌(SaCas9)的2Cas9结构域。在一些实施方案中,SaCas9结构域是核酸酶活性SaCas9、核酸酶失活的SaCas9(SaCas9d)或SaCas9切口酶(SaCas9n)。在一些实施方案中,SaCas9包括N579A突变,或本文所提供的任何氨基酸序列中的相应的突变。

[1445] 在一些实施方案中,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。在一些实施方案中,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。在一些实施方案中,SaCas9结构域包括E781X、N967X和R1014X突变中的一个或多个,或本文所提供的任何氨基酸序列中的相应的突变,其中X是任何氨基酸。在一些实施方案中,SaCas9结构域包括E781K、N967K和R1014H突变中的一个或多个,或本文所提供的任何氨基酸序列中的一个或多个相应的突

变。在一些实施方案中,SaCas9结构域包括E781K、N967K或R1014H突变,或本文所提供的任何氨基酸序列中的相应的突变。

[1446] 实施例性SaCas9序列

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVK
KLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTK
EQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYI
DLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITR
DENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDIT
ARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLI
LDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGL
PNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYS
LEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEE**N**SKKGNRTPFQYLSSSDSKISYETF
KKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDV
KVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQA
ESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNL
NGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKD
NGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENY
YEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLE
NMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG

[1447]

[1448] 以上加底线和粗体的残基N579可以突变(例如,突变成A579)以产生SaCas9切口酶。

[1449] 实施例性SaCas9n序列

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVK
KLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTK
EQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYI
DLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITR
DENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDIT
ARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLI
LDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGL
PNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYS
LEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEE**A**SKKGNRTPFQYLSSSDSKISYETF
KKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDV
KVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQA
ESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNL
NGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKD
NGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENY
YEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLE
NMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG

[1450]

[1451]

[1452] 以上的残基A579可以从N579突变以产生SaCas9切口酶,用底线和粗体表示。

[1453] 实施例性SaKKH Cas9

[1454] KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVK
KLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTK
EQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYI
DLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITR
DENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDIT
ARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLI
LDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGL
PNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYS
LEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEE**A**SKKGNRTPFQYLSSSDSKISYETF
KKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDV
KVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQA
ESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNR*K*LINDTLYSTRKDDKGNTLIVNNL
NGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKD
NGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENY
YEVNSKCYEEAKKLKKISNQAEFIASFY*K*NDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLE
NMNDKRPP*H*IIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG。

[1455] 以上可以从N579突变以产生SaCas9切口酶的的残基A579用底线和粗体表示。以上可以从E781、N967和R1014突变以产生SaKKH Cas9的残基K781、K967和H1014用底线和斜体表示。

[1456] 在一些实施方案中,napDNAbp是环状排列物。在以下序列中,纯文本表示腺苷脱氨酶序列,粗体序列表示衍生自Cas9的序列,斜体序列表示连接子序列,底线序列表示二分核定位序列。

[1457] CP5(具有MSP“NGC”PID和“D10A”切口酶):

[1458] **EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNI**

VKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPTVAYSVLVVAKVEKGKSKKLKSVK
ELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAKFLQKGNELAL
PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYN
KHRDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYRSTKEVLDATLIHQSITGLYETRIDL
SQLGGD*GGSGGSGGSGGSGGSGGSGGM*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRH
SIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVE
EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLI
ALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVN
TEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFI
[1459] KPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKI
LTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPK
HSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS
VEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKA
QVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSR
ERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQS
FLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELD
KAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREIN
NYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEGADKRTADGSEFESPKKK
RKV*

[1460] 在一些实施方案中,核酸可编程的DNA结合蛋白(napDNAbp)是微生物CRISPR-Cas系统的单个效应器。微生物CRISPR-Cas系统的单个效应器包括但不限于Cas9、Cpf1、Cas12b/C2c1及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器复合物,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之外,Shmakov et al.,“Discovery and Functional Characterization of Diverse Class 2CRISPR Cas Systems”,Mol.Cell,2015Nov.5;60(3):385-397中描述了三个不同的2类CRISPR-Cas系统(Cas12b/C2c1和Cas12c/C2c3),其全部内容特此通过引入方式并入。其中两个系统Cas12b/C2c1和Cas12c/C2c3的效应器包括与Cpf1相关的RuvC样核酸内切酶结构域。第三个系统包括一个具有两个预测HEPN RNase结构域的效应器。与Cas12b/C2c1产生的CRISPR RNA不同,成熟CRISPR RNA的产生不依赖于CRISPR RNA和tracrRNA进行DNA切割。

[1461] 据报道,酸土脂环酸芽孢杆菌Cas12b/C2c1(AacC2c1)的晶体结构与嵌合单分子向导RNA(sgRNA)形成复合物。参见例如,Liu et al.,“C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism”,Mol.Cell,2017Jan.19;65(2):310-322,其全部内容特此通过引入方式并入。在以三元复合物形式与目标DNA结合的酸土脂环酸芽孢杆菌C2c1中也报道了晶体结构。参见例如,Yang et al.,“PAM-dependent Target DNA Recognition and Cleavage by C2C1CRISPR-Cas endonuclease”,Cell,2016Dec.15;167(7):1814-1828,其全部内容特此通过引入方式并入。AacC2c1的具有催化能力的构象(具有

目标DNA链和非目标DNA链)已独立地位于在单个RuvC催化口袋中,其中Cas12b/C2c1介导的切割导致目标DNA的七个核苷酸交错断裂。Cas12b/C2c1三元复合物与先前识别的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。

[1462] 在一些实施方案中,本文所提供的任何融合蛋白的核酸可编程的DNA结合蛋白(napDNAbp)可以是Cas12b/C2c1或Cas12c/C2c3蛋白。在一些实施方案中,napDNAbp是Cas12b/C2c1蛋白。在一些实施方案中,napDNAbp是Cas12c/C2c3蛋白。在一些实施方案中,napDNAbp包括的氨基酸序列为与天然存在的Cas12b/C2c1或Cas12c/C2c3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。在一些实施方案中,napDNAbp是天然存在的Cas12b/C2c1或Cas12c/C2c3蛋白。在一些实施方案中,napDNAbp包括的氨基酸序列为与本文所提供的任何一个napDNAbp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。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。

[1463] Cas12b/C2c1((uniprot.org/uniprot/T0D7A2#2)

[1464] sp|T0D7A2|C2C1_ALIAG CRISPR相关的核酸内切酶C2c1 OS=酸土脂环酸芽孢杆菌(ATCC 49025/DSM 3922/CIP31B 3926/CIP31B31G/CIP3131B3100 c2c1 PE=1SV=1)氨基酸序列如下:

[1465] MAVKSIKVKLRLDDMPEIRAGLWKLHKEVNAGVRYYTEWLSLLRQENLYRRSPNGDGEQ

[1466] ECDKTAEECKAELLERLRARQVENGHRGPAGSDDELLQLARQLYELLVPQAIGAKGDAQ

[1467] QIARKFLSPLADKDAVGGLGIAKAGNKPRWVRMREAGEPGWEEEKEKAETRKSADRTAD

[1468] VLRALADFGLKPLMRVYTDSEMSSVEWKPLRKGQAVRTWDRDMFQQAIERMMSWESWNQ

[1469] RVGQEYAKLVEQKNRFEQKNFVGQEHLVHLVNQLQQDMKEASPGLESKEQTAHYVTGRA

[1470] LRGSDKVFEKWGKLAPDAPFDLYDAEIKNVQRRNTRRFGSHDLFAKLAEPEYQALWRED

[1471] ASFLTRYAVYNSILRKLNHAKMFATFTLPDATAHPIWTRFDKLGGNLHQYTFLFNEFGE

[1472] RRHAIRFHKLLKVENGVAREVDDVTVPISMSEQLDNLLPRDPNEPIALYFRDYGAEQHF

[1473] TGEFGGAKIQCRRDQLAHMHRRRGARDVYLNVSVRVQSQSEARGERRPPYAAVFRLVGD

[1474] NHRAFVHFDKLSDYLAEHPDDGKLGSEGLLSGLRVMSVDLGLRTSASISVFRVARKDEL

[1475] KPNSKGRVPFFFPIKGNDNLVAVHERSQLLKLPGETESKDLRAIREERQRTLRQLRTQL

[1476] AYLRLLVRCGSEDVGRRERSWAKLIEQPVDAANHMTPDWREAFENELQKLKSLHGICSD

[1477] KEWMDAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYAKDVVGGNSIEQIEYLERQ

[1478] YKFLKSWSFFGKVSGQVIRAEKGSRFAITLREHIDHAKEDRLKKLADRIIMEALGYVYA

[1479] LDERGKGKWVAKYPPCQLILLEELSEYQFNNDRPPSENNQLMQWSHRGVFQELINQAQV

[1480] HDLLVGTMYAAFSSRFDARTGAPGIRCRRVPARCTQEHNPEPFPWWLNKFVVEHTLDAC

[1481] PLRADDLIPTGEGEIFVSPFSAEEGDFHQIHADLNAAQNLQQRLWSDFDISQIRLRCDW

[1482] GEVDGELVLIPRLTGKRTADSYSNKVFYTNTGVTYYERERGKKRRKVFAQEKLSEEEAE

[1483] LLVEADEAREKSVVLMRDPSGIINRGNWTRQKEFWSMV NQRIEGYLVKQIRSRVPLQD

[1484] SACENTGDI。

[1485] AacCas12b(嗜酸脂环杆菌)-WP_067623834

[1486] MAVKSMKVKLRLDNMPEIRAGLWKLHTEVNAGVRYYTEWLSLLRQENLYRRSPNGDGEQECYKTAEECK

AELLERLRARQVENGHCGPAGSDDELLQLARQLYELLVPQAIGAKGDAQQIARKFLSPLADKDAVGGLGIAKAGNKP
RWVRMREAGEPGWEEEKAKAEARKSTDRTADVLRALADFGLKPLMRVYTDSDMSSVQWKPLRKGQAVRTWDRDMFQQ
AIERMMSWESWNQRVGEAYAKLVEQKSRFEQKNFVGQEHLVQLVNQLQQDMKEASHGLESKEQTAHYLTGRALRGSD
KVFEKWEKLDPDAPFDLYDTEIKNVQRRNTRRFGSHDLFAKLAEPKYQALWREDASFLTRYAVYNSIVRKLNHAKMF
ATFTLPDATAHPIWTRFDKLGGNLHQYTFLFNEFGEGRHAIRFQKLLTVEDGVAKEVDDVTVPISMSAQLDDLLPRD
PHELVALYFQDYGAEQHLAGEFGGAKIQYRRDQLNHLHARRGARDVYLNLSVRVQSQSEARGERRPPYAAVFRLVGD
NHRAFVHFDKLSDYLAEHPDDGKLGSEGLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNSEGRVPFCFPIEGNE
NLVAVHERSQLLKLPGETESKDLRAIREERQRTLRQLRTQLAYLRLLVRCGSEDVGRRERSWAKLIEQPMDANQMTP
DWREAFEDELQKLKSLYGICGDREWTEAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYQKDVVGGNSIEQIEY
LERQYKFLKSWSFFGKVSGQVIRAEKGSRFAITLREHIDHAKEDRLKKLADRIIMEALGYVYALDDERGKGKWVAKY
PPCQLILLEELSEYQFNNDRPPSENNQLMQWSHRGVFQELLNQAQVHDLLVGTMYAAFSSRFDARTGAPGIRCRRVP
ARCAREQNPEPFPWWLNKFVAEHKLDGCPLRADDLIPTGEGEFFVSPFSAEEGDFHQIHADLNAAQNLQRRLWSDFD
ISQIRLRCDWGEVDGEPVLIPRTTGKRTADSYGNKVFYTKTGVTYYERERGKKRRKVFAQEELSEEEAELLVEADEA
REKSVVLMRDPSGIINRGDWTRQKEFWSMVNQRIEGYLVKQIRSRVRLQESACENTGDI

[1487] BhCas12b(久石芽孢杆菌)NCBI参考序列:WP_095142515MAPKKKRKVGIHGVPAAATRS
FILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYEHHEQDPKNPKKVSKAEIQAELWDFVLKMQKCNSF
THEVDKDEVFNILRELYEELVPSSVEKKGEANQLSNKFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEEK
KKWEEDKKKDPLAKILGKLAEYGLIPLFIPYTDSNEPIVKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWESWNLK
VKEEYEKVEKEYKTLEERIKEDIQALKALEQYEKERQEQLLRDTLNTNEYRLSKRGLRGWREIIQKWLKMDENEPSE
KYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYATFCEIDKKKKDAKQQATFTLADPINHPLWVR
FEERSGSNLNKYRILTEQLHTEKLKKKLTVQLDRLIYPTESGGWEEKGKVDIVLLPSRQFYNQIFLDIEEKGKHAFT
YKDESIKFPLKGTLGGARVQFDRDHLRRYPHKVESGNVGRIYFNMTVNIEPTESPVSKSLKIHRDDFPKVVNFKPKE
LTEWIKDSKGKKLKSGIESLEIGLRVMSIDLGQRQAAAAASIFEVVDQKPDIEGKLFFPIKGTELYAVHRASFNIKLP
GETLVKSREVLRKAREDNLKLMNQKLNFLRNVLHFQQFEDITEREKRVTKWISRQENSDVPLVYQDELIQIRELMYK
PYKDWVAFLKQLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDEIDRTRKFLLRWSLRPTEPGEVRRLEPGQR
FAIDQLNHLNALKEDRLKKMANTIIMHALGYCYDVRKKKWQAKNPACQIILFEDLSNYNPYEERSRFENSKLMKWSR
REIPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKLQDNRFFKNLQREGRLTLDKIAVLKEGDLY
PDKGGEKFISLSKDRKCVTTHADINAAQNLQKRFWTRTHGFYKVYCKAYQVDGQTVYIPESKDQKQKIIEEFGEGYF
ILKDGVYEWVNAGKLKIKKGSSKQSSSELVDSDILKDSFDLASELKGEKLMLYRDPSGNVFPSDKWMAAGVFFGKLE
RILISKLTNQYSISTIEDDSSKQSMKRPAATKKAGQAKKKK

[1488] BvCas12b V4(相对于野生型的变化S893R/K846R/E837G)是表达如下:5'mRNA Cap---5'UTR---bhCas12b---STOP序列---3'UTR---120多聚苷酸尾5'UTR:GGGAAATAAGAGAGAAAAGAAGAGTAAGAA GAAATATAAGAGCCACC

[1489] 3'UTR(TriLink标准UTR)

[1490] GCTGGAGCCTCGGTGGCCATGCTTCTTGCCCCTTGGGCCTCCCCCCAGCCCCTCCTCCCCTTCCTGCAC CCGTACCCCCGTGGTCTTTGAATAAAGTCTGA

[1491] bhCas12b(V4)的核酸序列

[1492] ATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGCCACCAGATCCTTCA

[1493] TCCTGAAGATCGAGCCCAACGAGGAAGTGAAGAAAGGCCTCTGGAAAACCCACGAGGTGCTGAACCA

[1494] CGGAATCGCCTACTACATGAATATCCTGAAGCTGATCCGGCAAGAGGCCATCTACGAGCACCACGAG

[1495] CAGGACCCCAAGAATCCCAAGAAGGTGTCCAAGGCCGAGATCCAGGCCGAGCTGTGGGATTTCGTGC

[1496] TGAAGATGCAGAAGTGCAACAGCTTCACACACGAGGTGGACAAGGACGAGGTGTTCAACATCCTGAG

[1497] AGAGCTGTACGAGGAACTGGTGCCCAGCAGCGTGGAAAAGAAGGGCGAAGCCAACCAGCTGAGCAAC

[1498] AAGTTTCTGTACCCTCTGGTGGACCCCAACAGCCAGTCTGGAAAGGGAACAGCCAGCAGCGGCAGAA

[1499] AGCCCAGATGGTACAACCTGAAGATTGCCGGCGATCCCTCCTGGGAAGAAGAGAAGAAGAAGTGGGA

[1500] AGAAGATAAGAAAAAGGACCCGCTGGCCAAGATCCTGGGCAAGCTGGCTGAGTACGGACTGATCCCT

[1501] CTGTTCATCCCCTACACCGACAGCAACGAGCCCATCGTGAAAGAAATCAAGTGGATGGAAAAGTCCC

[1502] GGAACCAGAGCGTGCGGCGGCTGGATAAGGACATGTTCATTCAGGCCCTGGAACGGTTCCTGAGCTG

[1503] GGAGAGCTGGAACCTGAAAGTGAAAGAGGAATACGAGAAGGTCGAGAAAGAGTACAAGACCCTGGAA

[1504] GAGAGGATCAAAGAGGACATCCAGGCTCTGAAGGCTCTGGAACAGTATGAGAAAGAGCGGCAAGAAC

[1505] AGCTGCTGCGGGACACCCTGAACACCAACGAGTACCGGCTGAGCAAGAGAGGCCTTAGAGGCTGGCG

[1506] GGAAATCATCCAGAAATGGCTGAAAATGGACGAGAACGAGCCCTCCGAGAAGTACCTGGAAGTGTTC

[1507] AAGGACTACCAGCGGAAGCACCCTAGAGAGGCCGGCGATTACAGCGTGTACGAGTTCCTGTCCAAGA

[1508] AAGAGAACCACTTCATCTGGCGGAATCACCCTGAGTACCCCTACCTGTACGCCACCTTCTGCGAGAT

[1509] CGACAAGAAAAAGAAGGACGCCAAGCAGCAGGCCACCTTCACACTGGCCGATCCTATCAATCACCCT

[1510] CTGTGGGTCCGATTCGAGGAAAGAAGCGGCAGCAACCTGAACAAGTACAGAATCCTGACCGAGCAGC

[1511] TGCACACCGAGAAGCTGAAGAAAAAGCTGACAGTGCAGCTGGACCGGCTGATCTACCCTACAGAATC

[1512] TGGCGGCTGGGAAGAGAAGGGCAAAGTGGACATTGTGCTGCTGCCCAGCCGGCAGTTCTACAACCAG

[1513] ATCTTCCTGGACATCGAGGAAAAGGGCAAGCACGCCTTCACCTACAAGGATGAGAGCATCAAGTTCC

[1514] CTCTGAAGGGCACACTCGGCGGAGCCAGAGTGCAGTTCGACAGAGATCACCTGAGAAGATACCCTCA

[1515] CAAGGTGGAAAGCGGCAACGTGGGCAGAATCTACTTCAACATGACCGTGAACATCGAGCCTACAGAG

[1516] TCCCCAGTGTCCAAGTCTCTGAAGATCCACCGGGACGACTTCCCCAAGGTGGTCAACTTCAAGCCCA

[1517] AAGAACTGACCGAGTGGATCAAGGACAGCAAGGGCAAGAAACTGAAGTCCGGCATCGAGTCCCTGGA

[1518] AATCGGCCTGAGAGTGATGAGCATCGACCTGGGACAGAGACAGGCCGCTGCCGCCTCTATTTTCGAG

[1519] GTGGTGGATCAGAAGCCCGACATCGAAGGCAAGCTGTTTTTCCCAATCAAGGGCACCGAGCTGTATG

[1520] CCGTGCACAGAGCCAGCTTCAACATCAAGCTGCCCGGCGAGACACTGGTCAAGAGCAGAGAAGTGCT

[1521] GCGGAAGGCCAGAGAGGACAATCTGAAACTGATGAACCAGAAGCTCAACTTCCTGCGGAACGTGCTG

[1522] CACTTCCAGCAGTTCGAGGACATCACCGAGAGAGAGAAGCGGGTCACCAAGTGGATCAGCAGACAAG

[1523] AGAACAGCGACGTGCCCCTGGTGTACCAGGATGAGCTGATCCAGATCCGCGAGCTGATGTACAAGCC

[1524] TTACAAGGACTGGGTCGCCTTCCTGAAGCAGCTCCACAAGAGACTGGAAGTCGAGATCGGCAAAGAA

[1525] GTGAAGCACTGGCGGAAGTCCCTGAGCGACGGAAGAAAGGGCCTGTACGGCATCTCCCTGAAGAACA

[1526] TCGACGAGATCGATCGGACCCGGAAGTTCCTGCTGAGATGGTCCCTGAGGCCTACCGAACCTGGCGA

[1527] AGTGCGTAGACTGGAACCCGGCCAGAGATTCGCCATCGACCAGCTGAATCACCTGAACGCCCTGAAA

[1528] GAAGATCGGCTGAAGAAGATGGCCAACACCATCATCATGCACGCCCTGGGCTACTGCTACGACGTGC

[1529] GGAAGAAGAAATGGCAGGCTAAGAACCCCGCCTGCCAGATCATCCTGTTCGAGGATCTGAGCAACTA

[1530] CAACCCCTACGAGGAAAGGTCCCGCTTCGAGAACAGCAAGCTCATGAAGTGGTCCAGACGCGAGATC

[1531] CCCAGACAGGTTGCACTGCAGGGCGAGATCTATGGCCTGCAAGTGGGAGAAGTGGGCGCTCAGTTCA

[1532] GCAGCAGATTCCACGCCAAGACAGGCAGCCCTGGCATCAGATGTAGCGTCGTGACCAAAGAGAAGCT

[1533] GCAGGACAATCGGTTCTTCAAGAATCTGCAGAGAGAGGGCAGACTGACCCTGGACAAAATCGCCGTG

[1534] CTGAAAGAGGGCGATCTGTACCCAGACAAAGGCGGCGAGAAGTTCATCAGCCTGAGCAAGGATCGGA

[1535] AGTGCGTGACCACACACGCCGACATCAACGCCGCTCAGAACCTGCAGAAGCGGTTCTGGACAAGAAC

[1536] CCACGGCTTCTACAAGGTGTACTGCAAGGCCTACCAGGTGGACGGCCAGACCGTGTACATCCCTGAG

[1537] AGCAAGGACCAGAAGCAGAAGATCATCGAAGAGTTCGGCGAGGGCTACTTCATTCTGAAGGACGGGG

[1538] TGTACGAATGGGTCAACGCCGGCAAGCTGAAAATCAAGAAGGGCAGCTCCAAGCAGAGCAGCAGCGA

[1539] GCTGGTGGATAGCGACATCCTGAAAGACAGCTTCGACCTGGCCTCCGAGCTGAAAGGCGAAAAGCTG

[1540] ATGCTGTACAGGGACCCCAGCGGCAATGTGTTCCCCAGCGACAAATGGATGGCCGCTGGCGTGTTCT

[1541] TCGGAAAGCTGGAACGCATCCTGATCAGCAAGCTGACCAACCAGTACTCCATCAGCACCATCGAGGA

[1542] CGACAGCAGCAAGCAGTCTATGAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAA

[1543] AAG

[1544] 在一些实施方案中,Cas12b是BvCas12B,其是BhCas12b的变体并且相对于BhCas12B包括以下变化:S893R、K846R和E837G。

[1545] BvCas12b(芽孢杆菌V3-13)NCBI参考序列:WP_101661451.1MAIRSIKLKMKTNSGTDSIYLRKALWRTHQLINEGIAYYMNLLTLYRQEAIGDKTKEAYQAELINIIRNQQRNNGSSEEHGSDQEILALLRQLYELIIPSSIGESGDANQLGNKFLYPLVDPNSQSGKGTSNAGRKPRWKRLKEEGNPDWELEKKKDEERKAKDPTVKIFDNLNKYGLLPLFPLFTNIQKDIEWLPLGKRQSVRKWDKDMFIQAIERLLSWESWNRRVADEYKQLKEKTESYYKEHLTGGEEWIEKIRKFEKERNMELEKNAFAPNDGYFITSRQIRGWDRVYEKWSKLPESASPEELWKVVAEQQNKMSEGFGDPKVFSFLANRENRDIWRGHSERIYHIAAYNGLQKKLSRTKEQATFTLPDAIEHPLWIRYESPGGTNLNLFKLEEKQKKNYYVTLSKIIWPSEEKWIEKENIEIPLAPSIQFNRQIKLKQHVKGKQEISFSDYSSRISLDGVLGGSRIQFNRKYIKNHKELLGEGDIGPVFFNLVVDVAPLQETRNGRLQSPIGKALKVISSDFSKVIDYKPKELMDWMNTGSASNSFGVASLLEGMRVMSIDMGQRTSASVSIFEVVKELPKDQEQKLFYSINDTELFAIHKRSFLLNLPGEVVTKNNKQQRQERRKKRQFVRSQIRMLANVLRLETKKTPDERKKAIHKLMEIVQSYDSWTASQKEVWEKELNLLTNMAAFNDEIWKESLVELHHRIEPYVGQIVSKWRKGLSEGRKNLAGISMWNIDELEDTRRLLISWSKRSRTPGEANRIETDEPFGSSLLQHIQNVKDDRLKQMANLIIMTALGFKYDKEEKDRYKRWKETYPACQIILFENLNRYLFNLDRSRRENSRLMKWAHRSIPRTVSMQGEMFGLQVGDVRSEYSSRFHAKTGAPGIRCHALTEEDLKAGSNTLKRLIEDGFINESELAYLKKGDIIPSQGGELFVTL SKRYKKDSDNNELTVIHADINAAQNLQKRFWQQNSEVYRVPCQLARMGEDKLYIPKSQTETIKKYFGKGSFVKNNTEQEVYKWEKSEKMKIKTDTTFDLQDLDGFEDISKTIELAQEQQKKYLTMFRDPSGYFFNNETWRPQKEYWSIVNNIIKSCLKKKILSNKVEL

[1546] 在一些实施方案中,Cas12b是BTCas12b。BTCas12b(嗜热淀粉芽孢杆菌(Bacillus thermoamylovorans))NCBI参考序列:WP_041902512

[1547] MATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYEHHEQDPKNPKKV

[1548] SKAEIQAELWDFVLKMQKCNSFTHEVDKDVVFNILRELYEELVPSSVEKKGEANQLSNKF

[1549] LYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEEKKKWEEDKKKDPLAKILGKLAE

[1550] YGLIPLFIPFTDSNEPIVKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWESWNLKVKEE

[1551] YEKVEKEHKTLEERIKEDIQAFKSLEQYEKERQEQLLRDTLNTNEYRLSKRGLRGWREII

[1552] QKWLKMDENEPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYAT

[1553] FCEIDKKKKDAKQQATFTLADPINHPLWVRFEERSGSNLNKYRILTEQLHTEKLKKKLTV

[1554] QLDRLIYPTESGGWEEKGKVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGT

[1555] LGGARVQFDRDHLRRYPHKVESGNVGRIYFNMTVNIEPTESPVSKSLKIHRDDFPKFVNF

[1556] KPKELTEWIKDSKGKKLKSGIESLEIGLRVMSIDLGQRQAAAASIFEVVDQKPDIEGKLF

[1557] FPIKGTELYAVHRASFNIKLPGETLVKSREVLRKAREDNLKLMNQKLNFLRNVLHFQQFE

[1558] DITEREKRVTKWISRQENSDVPLVYQDELIQIRELMYKPYKDWVAFLKQLHKRLEVEIGK

[1559] EVKHWRKSLSDGRKGLYGISLKNIDEIDRTRKFLLRWSLRPTEPGEVRRLEPGQRFAIDQ

[1560] LNHLNALKEDRLKKMANTIIMHALGYCYDVRKKKWQAKNPACQIILFEDLSNYNPYEERS

[1561] RFENSKLMKWSRREIPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKL

[1562] QDNRFFKNLQREGRLTLDKIAVLKEGDLYPDKGGEKFISLSKDRKLVTTHADINAAQNLQ

[1563] KRFWTRTHGFYKVYCKAYQVDGQTVYIPESKDQKQKIIEEFGEGYFILKDGVYEWGNAGK

[1564] LKIKKGSSKQSSSELVDSDILKDSFDLASELKGEKLMLYRDPSGNVFPSDKWMAAGVFFG

[1565] KLERILISKLTNQYSISTIEDDSSKQSM

[1566] 在一些实施方案中,napDNAbp是指Cas12c。在一些实施方案中,Cas12c蛋白是Cas12c1或Cas12c1的变体。在一些实施方案中,Cas12蛋白是Cas12c2或Cas12c2的变体。在一些实施方案中,Cas12蛋白是来自嗜油菌属HI0009(即OspCas12c)或OspCas12c的变体的Cas12c蛋白。这些Cas12c分子已在Yan et al.,“Functionally Diverse Type V CRISPR-Cas Systems,”Science,2019Jan.4;363:88-91中描述;其全部内容特此通过引入方式作为参考。在一些实施方案中,napDNAbp包括的氨基酸序列为与天然存在的Cas12c1、Cas12c2或OspCas12c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。在一些实施方案中,napDNAbp是天然存在的Cas12c1、Cas12c2或OspCas12c蛋白。在一些实施方案中,napDNAbp包括的氨基酸序列为与本文所述任何Cas12c1、Cas12c2或OspCas12c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。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。

[1567] Cas12c1

[1568] MQTKKTHLHLISAKASRKYRRTIACLSDTAKKDLERRKQSGAADPAQELS

[1569] CLKTIKFKLEVPEGSKLPSFDRISQIYNALETIEKGSLSYLLFALILSGF

[1570] RIFPNSSAAKTFASSSCYKNDQFASQIKEIFGEMVKNFIPSELESILKKG

[1571] RRKNNKDWTEENIKRVLNSEFGRKNSEGSSALFDSFLSKFSQELFRKFDS

[1572] WNEVNKKYLEAAELLDSMLASYGPFDSVCKMIGDSDSRNSLPDKSTIAFT

[1573] NNAEITVDIESSVMPYMAIAALLREYRQSKSKAAPVAYVQSHLTTTNGNG

[1574] LSWFFKFGLDLIRKAPVSSKQSTSDGSKSLQELFSVPDDKLDGLKFIKEA

[1575] CEALPEASLLCGEKGELLGYQDFRTSFAGHIDSWVANYVNRLFELIELVN

[1576] QLPESIKLPSILTQKNHNLVASLGLQEAEVSHSLELFEGLVKNVRQTLKK

[1577] LAGIDISSSPNEQDIKEFYAFSDVLNRLGSIRNQIENAVQTAKKDKIDLE

[1578] SAIEWKEWKKLKKLPKLNGLGGGVPKQQELLDKALESVKQIRHYQRIDFE

[1579] RVIQWAVNEHCLETVPKFLVDAEKKKINKESSTDFAAKENAVRFLLEGIG

[1580] AAARGKTDSVSKAAYNWFVVNNFLAKKDLNRYFINCQGCIYKPPYSKRRS

[1581] LAFALRSDNKDTIEVVWEKFETFYKEISKEIEKFNIFSQEFQTFLHLENL

[1582] RMKLLLRRIQKPIPAEIAFFSLPQEYYDSLPPNVAFLALNQEITPSEYIT

[1583] QFNLYSSFLNGNLILLRRSRSYLRAKFSWVGNSKLIYAAKEARLWKIPNA

[1584] YWKSDEWKMILDSNVLVFDKAGNVLPAPTLKKVCEREGDLRLFYPLLRQL

[1585] PHDWCYRNPFVKSVGREKNVIEVNKEGEPKVASALPGSLFRLIGPAPFKS

[1586] LLDDCFFNPLDKDLRECMLIVDQEISQKVEAQKVEASLESCTYSIAVPIR

[1587] YHLEEPKVSNQFENVLAIDQGEAGLAYAVFSLKSIGEAETKPIAVGTIRI

[1588] PSIRRLIHSVSTYRKKKQRLQNFKQNYDSTAFIMRENVTGDVCAKIVGLM

[1589] KEFNAFPVLEYDVKNLESGSRQLSAVYKAVNSHFLYFKEPGRDALRKQLW

[1590] YGGDSWTIDGIEIVTRERKEDGKEGVEKIVPLKVFPGRSVSARFTSKTCS

[1591] CCGRNVFDWLFTEKKAKTNKKFNVNSKGELTTADGVIQLFEADRSKGPKF

[1592] YARRKERTPLTKPIAKGSYSLEEIERRVRTNLRRAPKSKQSRDTSQSQYF

[1593] CVYKDCALHFSGMQADENAAINIGRRFLTALRKNRRSDFPSNVKISDRLL

[1594] DN

[1595] Cas12c2

[1596] MTKHSIPLHAFRNSGADARKWKGRIALLAKRGKETMRTLQFPLEMSEPEA

[1597] AAINTTPFAVAYNAIEGTGKGTLFDYWAKLHLAGFRFFPSGGAATIFRQQ

[1598] AVFEDASWNAAFCQQSGKDWPWLVPSKLYERFTKAPREVAKKDGSKKSIE

[1599] FTQENVANESHVSLVGASITDKTPEDQKEFFLKMAGALAEKFDSWKSANE

[1600] DRIVAMKVIDEFLKSEGLHLPSLENIAVKCSVETKPDNATVAWHDAPMSG

[1601] VQNLAIGVFATCASRIDNIYDLNGGKLSKLIQESATTPNVTALSWLFGKG

[1602] LEYFRTTDIDTIMQDFNIPASAKESIKPLVESAQAIPTMTVLGKKNYAPF

[1603] RPNFGGKIDSWIANYASRLMLLNDILEQIEPGFELPQALLDNETLMSGID

[1604] MTGDELKELIEAVYAWVDAAKQGLATLLGRGGNVDDAVQTFEQFSAMMDT

[1605] LNGTLNTISARYVRAVEMAGKDEARLEKLIECKFDIPKWCKSVPKLVGIS

[1606] GGLPKVEEEIKVMNAAFKDVRARMFVRFEEIAAYVASKGAGMDVYDALEK

[1607] RELEQIKKLKSAVPERAHIQAYRAVLHRIGRAVQNCSEKTKQLFSSKVIE

[1608] MGVFKNPSHLNNFIFNQKGAIYRSPFDRSRHAPYQLHADKLLKNDWLELL

[1609] AEISATLMASESTEQMEDALRLERTRLQLQLSGLPDWEYPASLAKPDIEV

[1610] EIQTALKMQLAKDTVTSDVLQRAFNLYSSVLSGLTFKLLRRSFSLKMRFS

[1611] VADTTQLIYVPKVCDWAIPKQYLQAEGEIGIAARVVTESSPAKMVTEVEM

[1612] KEPKALGHFMQQAPHDWYFDASLGGTQVAGRIVEKGKEVGKERKLVGYRM

[1613] RGNSAYKTVLDKSLVGNTELSQCSMIIEIPYTQTVDADFRAQVQAGLPKV

[1614] SINLPVKETITASNKDEQMLFDRFVAIDLGERGLGYAVFDAKTLELQESG

[1615] HRPIKAITNLLNRTHHYEQRPNQRQKFQAKFNVNLSELRENTVGDVCHQI

[1616] NRICAYYNAFPVLEYMVPDRLDKQLKSVYESVTNRYIWSSTDAHKSARVQ

[1617] FWLGGETWEHPYLKSAKDKKPLVLSPGRGASGKGTSQTCSCCGRNPFDLI

[1618] KDMKPRAKIAVVDGKAKLENSELKLFERNLESKDDMLARRHRNERAGMEQ

[1619] PLTPGNYTVDEIKALLRANLRRAPKNRRTKDTTVSEYHCVFSDCGKTMHA

[1620] DENAAVNIGGKFIADIEK

[1621] OspCas12c

[1622] MTKLRHRQKKLTHDWAGSKKREVLGSNGKLQNPLLMPVKKGQVTEFRKAFSAYARATKGEMTDGRKN

[1623] MFTHSFEPFKTKPSLHQCELADKAYQSLHSYLPGSLAHFLLSAHALGFRIFSKSGEATAFQASSKIE

[1624] AYESKLASELACVDLSIQNLTISTLFNALTTSVRGKGEETSADPLIARFYTLLTGKPLSRDTQGPER

[1625] DLAEVISRKIASSFGTWKEMTANPLQSLQFFEEELHALDANVSLSPAFDVLIKMNDLQGDLKNRTIV

[1626] FDPDAPVFEYNAEDPADIIIKLTARYAKEAVIKNQNVGNYVKNAITTTNANGLGWLLNKGLSLLPVS

[1627] TDDELLEFIGVERSHPSCHALIELIAQLEAPELFEKNVFSDTRSEVQGMIDSAVSNHIARLSSSRNS

[1628] LSMDSEELERLIKSFQIHTPHCSLFIGAQSLSQQLESLPEALQSGVNSADILLGSTQYMLTNSLVEE

[1629] SIATYQRTLNRINYLSGVAGQINGAIKRKAIDGEKIHLPAAWSELISLPFIGQPVIDVESDLAHLKN

[1630] QYQTLSNEFDTLISALQKNFDLNFNKALLNRTQHFEAMCRSTKKNALSKPEIVSYRDLLARLTSCLY

[1631] RGSLVLRRAGIEVLKKHKIFESNSELREHVHERKHFVFVSPLDRKAKKLLRLTDSRPDLLHVIDEIL

[1632] QHDNLENKDRESLWLVRSGYLLAGLPDQLSSSFINLPIITQKGDRRLIDLIQYDQINRDAFVMLVTS

[1633] AFKSNLSGLQYRANKQSFVVTRTLSPYLGSKLVYVPKDKDWLVPSQMFEGRFADILQSDYMVWKDAG

[1634] RLCVIDTAKHLSNIKKSVFSSEEVLAFLRELPHRTFIQTEVRGLGVNVDGIAFNNGDIPSLKTFSNC

[1635] VQVKVSRTNTSLVQTLNRWFEGGKVSPPSIQFERAYYKKDDQIHEDAAKRKIRFQMPATELVHASDD

[1636] AGWTPSYLLGIDPGEYGMGLSLVSINNGEVLDSGFIHINSLINFASKKSNHQTKVVPRQQYKSPYAN

[1637] YLEQSKDSAAGDIAHILDRLIYKLNALPVFEALSGNSQSAADQVWTKVLSFYTWGDNDAQNSIRKQH

[1638] WFGASHWDIKGMLRQPPTEKKPKPYIAFPGSQVSSYGNSQRCSCCGRNPIEQLREMAKDTSIKELKI

[1639] RNSEIQLFDGTIKLFNPDPSTVIERRRHNLGPSRIPVADRTFKNISPSSLEFKELITIVSRSIRHSP

[1640] EFIAKKRGIGSEYFCAYSDCNSSLNSEANAAANVAQKFQKQLFFEL

[1641] 在一些实施方案中,napDNAbp是指Cas12g、Cas12h或Cas12i,其已在例如Yan et al.,“Functionally Diverse Type V CRISPR-Cas Systems,”Science,2019Jan.4;363:88-91中描述;其各自的全部内容特此通过引用方式并入。通过聚合超过10TB的序列数据识别了V型Cas蛋白的新分类,其与先前表征的V类蛋白(包括Cas12g、Cas12h和Cas12i)表现出弱相似性。在一些实施方案中,Cas12蛋白是Cas12g或Cas12g的变体。在一些实施方案中,Cas12蛋白是Cas12h或Cas12h的变体。在一些实施方案中,Cas12蛋白是Cas12i或Cas12i的变体。应当理解,其他RNA向导的DNA结合蛋白可以用作napDNAbp,并且在本公开的范畴内。在一些实施方案中,napDNAbp包括的氨基酸序列为与天然存在的Cas12g、Cas12h或Cas12i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。在一些实施方案中,napDNAbp是天然存在的Cas12g、Cas12h或Cas12i蛋白。在一些实施方案中,napDNAbp包括的氨基酸序列为与本文所述任何Cas12g、Cas12h或Cas12i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。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。在一些实施方案中,Cas12i是Cas12i1或Cas12i2。

[1642] Cas12g1

[1643] MAQASSTPAVSPRPRPRYREERTLVRKLLPRPGQSKQEFRENVKKLRKAFLQFNADVSGVCQWAIQF

[1644] RPRYGKPAEPTETFWKFFLEPETSLPPNDSRSPEFRRLQAFEAAAGINGAAALDDPAFTNELRDSIL

[1645] AVASRPKTKEAQRLFSRLKDYQPAHRMILAKVAAEWIESRYRRAHQNWERNYEEWKKEKQEWEQNHP

[1646] ELTPEIREAFNQIFQQLEVKEKRVRICPAARLLQNKDNCQYAGKNKHSVLCNQFNEFKKNHLQGKAI

[1647] KFFYKDAEKYLRCGLQSLKPNVQGPFREDWNKYLRYMNLKEETLRGKNGGRLPHCKNLGQECEFNPH

[1648] TALCKQYQQQLSSRPDLVQHDELYRKWRREYWREPRKPVFRYPSVKRHSIAKIFGENYFQADFKNSV

[1649] VGLRLDSMPAGQYLEFAFAPWPRNYRPQPGETEISSVHLHFVGTRPRIGFRFRVPHKRSRFDCTQEE

[1650] LDELRSRTFPRKAQDQKFLEAARKRLLETFPGNAEQELRLLAVDLGTDSARAAFFIGKTFQQAFPLK

[1651] IVKIEKLYEQWPNQKQAGDRRDASSKQPRPGLSRDHVGRHLQKMRAQASEIAQKRQELTGTPAPETT

[1652] TDQAAKKATLQPFDLRGLTVHTARMIRDWARLNARQIIQLAEENQVDLIVLESLRGFRPPGYENLDQ

[1653] EKKRRVAFFAHGRIRRKVTEKAVERGMRVVTVPYLASSKVCAECRKKQKDNKQWEKNKKRGLFKCEG

[1654] CGSQAQVDENAARVLGRVFWGEIELPTAIP

[1655] Cas12h1

[1656] MKVHEIPRSQLLKIKQYEGSFVEWYRDLQEDRKKFASLLFRWAAFGYAAREDDGATYISPSQALLER

[1657] RLLLGDAEDVAIKFLDVLFKGGAPSSSCYSLFYEDFALRDKAKYSGAKREFIEGLATMPLDKIIERI

[1658] RQDEQLSKIPAEEWLILGAEYSPEEIWEQVAPRIVNVDRSLGKQLRERLGIKCRRPHDAGYCKILME

[1659] VVARQLRSHNETYHEYLNQTHEMKTKVANNLTNEFDLVCEFAEVLEEKNYGLGWYVLWQGVKQALKE

[1660] QKKPTKIQIAVDQLRQPKFAGLLTAKWRALKGAYDTWKLKKRLEKRKAFPYMPNWDNDYQIPVGLTG

[1661] LGVFTLEVKRTEVVVDLKEHGKLFCSHSHYFGDLTAEKHPSRYHLKFRHKLKLRKRDSRVEPTIGPW

[1662] IEAALREITIQKKPNGVFYLGLPYALSHGIDNFQIAKRFFSAAKPDKEVINGLPSEMVVGAADLNLS

[1663] NIVAPVKARIGKGLEGPLHALDYGYGELIDGPKILTPDGPRCGELISLKRDIVEIKSAIKEFKACQR

[1664] EGLTMSEETTTWLSEVESPSDSPRCMIQSRIADTSRRLNSFKYQMNKEGYQDLAEALRLLDAMDSYN

[1665] SLLESYQRMHLSPGEQSPKEAKFDTKRASFRDLLRRRVAHTIVEYFDDCDIVFFEDLDGPSDSDSRN

[1666] NALVKLLSPRTLLLYIRQALEKRGIGMVEVAKDGTSQNNPISGHVGWRNKQNKSEIYFYEDKELLVM

[1667] DADEVGAMNILCRGLNHSVCPYSFVTKAPEKKNDEKKEGDYGKRVKRFLKDRYGSSNVRFLVASMGF

[1668] VTVTTKRPKDALVGKRLYYHGGELVTHDLHNRMKDEIKYLVEKEVLARRVSLSDSTIKSYKSFAHV

[1669] Cas12i1

[1670] MSNKEKNASETRKAYTTKMIPRSHDRMKLLGNFMDYLMDGTPIFFELWNQFGGGIDRDIISGTANKD

[1671] KISDDLLLAVNWFKVMPINSKPQGVSPSNLANLFQQYSGSEPDIQAQEYFASNFDTEKHQWKDMRVE

[1672] YERLLAELQLSRSDMHHDLKLMYKEKCIGLSLSTAHYITSVMFGTGAKNNRQTKHQFYSKVIQLLEE

[1673] STQINSVEQLASIILKAGDCDSYRKLRIRCSRKGATPSILKIVQDYELGTNHDDEVNVPSLIANLKE

[1674] KLGRFEYECEWKCMEKIKAFLASKVGPYYLGSYSAMLENALSPIKGMTTKNCKFVLKQIDAKNDIKY

[1675] ENEPFGKIVEGFFDSPYFESDTNVKWVLHPHHIGESNIKTLWEDLNAIHSKYEEDIASLSEDKKEKR

[1676] IKVYQGDVCQTINTYCEEVGKEAKTPLVQLLRYLYSRKDDIAVDKIIDGITFLSKKHKVEKQKINPV

[1677] IQKYPSFNFGNNSKLLGKIISPKDKLKHNLKCNRNQVDNYIWIEIKVLNTKTMRWEKHHYALSSTRF

[1678] LEEVYYPATSENPPDALAARFRTKTNGYEGKPALSAEQIEQIRSAPVGLRKVKKRQMRLEAARQQNL

[1679] LPRYTWGKDFNINICKRGNNFEVTLATKVKKKKEKNYKVVLGYDANIVRKNTYAAIEAHANGDGVID

[1680] YNDLPVKPIESGFVTVESQVRDKSYDQLSYNGVKLLYCKPHVESRRSFLEKYRNGTMKDNRGNNIQI

[1681] DFMKDFEAIADDETSLYYFNMKYCKLLQSSIRNHSSQAKEYREEIFELLRDGKLSVLKLSSLSNLSF

[1682] VMFKVAKSLIGTYFGHLLKKPKNSKSDVKAPPITDEDKQKADPEMFALRLALEEKRLNKVKSKKEVI

[1683] ANKIVAKALELRDKYGPVLIKGENISDTTKKGKKSSTNSFLMDWLARGVANKVKEMVMMHQGLEFVE

[1684] VNPNFTSHQDPFVHKNPENTFRARYSRCTPSELTEKNRKEILSFLSDKPSKRPTNAYYNEGAMAFLA

[1685] TYGLKKNDVLGVSLEKFKQIMANILHQRSEDQLLFPSRGGMFYLATYKLDADATSVNWNGKQFWVCN

[1686] ADLVAAYNVGLVDIQKDFKKK

[1687] Cas12i2

[1688] MSSAIKSYKSVLRPNERKNQLLKSTIQCLEDGSAFFFKMLQGLFGGITPEIVRFSTEQEKQQQDIAL

[1689] WCAVNWFRPVSQDSLTHTIASDNLVEKFEEYYGGTASDAIKQYFSASIGESYYWNDCRQQYYDLCRE

[1690] LGVEVSDLTHDLEILCREKCLAVATESNQNNSIISVLFGTGEKEDRSVKLRITKKILEAISNLKEIP

[1691] KNVAPIQEIILNVAKATKETFRQVYAGNLGAPSTLEKFIAKDGQKEFDLKKLQTDLKKVIRGKSKER

[1692] DWCCQEELRSYVEQNTIQYDLWAWGEMFNKAHTALKIKSTRNYNFAKQRLEQFKEIQSLNNLLVVKK

[1693] LNDFFDSEFFSGEETYTICVHHLGGKDLSKLYKAWEDDPADPENAIVVLCDDLKNNFKKEPIRNILR

[1694] YIFTIRQECSAQDILAAAKYNQQLDRYKSQKANPSVLGNQGFTWTNAVILPEKAQRNDRPNSLDLRI

[1695] WLYLKLRHPDGRWKKHHIPFYDTRFFQEIYAAGNSPVDTCQFRTPRFGYHLPKLTDQTAIRVNKKHV

[1696] KAAKTEARIRLAIQQGTLPVSNLKITEISATINSKGQVRIPVKFDVGRQKGTLQIGDRFCGYDQNQT

[1697] ASHAYSLWEVVKEGQYHKELGCFVRFISSGDIVSITENRGNQFDQLSYEGLAYPQYADWRKKASKFV

[1698] SLWQITKKNKKKEIVTVEAKEKFDAICKYQPRLYKFNKEYAYLLRDIVRGKSLVELQQIRQEIFRFI

[1699] EQDCGVTRLGSLSLSTLETVKAVKGIIYSYFSTALNASKNNPISDEQRKEFDPELFALLEKLELIRT

[1700] RKKKQKVERIANSLIQTCLENNIKFIRGEGDLSTTNNATKKKANSRSMDWLARGVFNKIRQLAPMHN

[1701] ITLFGCGSLYTSHQDPLVHRNPDKAMKCRWAAIPVKDIGDWVLRKLSQNLRAKNIGTGEYYHQGVKE

[1702] FLSHYELQDLEEELLKWRSDRKSNIPCWVLQNRLAEKLGNKEAVVYIPVRGGRIYFATHKVATGAVS

[1703] IVFDQKQVWVCNADHVAAANIALTVKGIGEQSSDEENPDGSRIKLQLTS

[1704] 碱基编辑器的代表性核酸和蛋白质序列如下：

[1705] 在P153的BhCas12b GGSGGS-ABE8-Xten20

GCCACCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGCCACCAGAT
CCTTCATCCTGAAGATCGAGCCCAACGAGGAAGTGAAGAAAGGCCTCTGGAAAACCCACGAGGTGCT
GAACCACGGAATCGCCTACTACATGAATATCCTGAAGCTGATCCGGCAAGAGGCCATCTACGAGCAC

[1706] CACGAGCAGGACCCCAAGAATCCCAAGAAGGTGTCCAAGGCCGAGATCCAGGCCGAGCTGTGGGATT
TCGTGCTGAAGATGCAGAAGTGCAACAGCTTCACACACGAGGTGGACAAGGACGAGGTGTTCAACAT
CCTGAGAGAGCTGTACGAGGAACTGGTGCCCAGCAGCGTGGAAAAGAAGGGCGAAGCCAACCAGCTG
AGCAACAAGTTTCTGTACCCTCTGGTGGACCCCAACAGCCAGTCTGGAAAGGGAACAGCCAGCAGCG

[1707]

GCAGAAAGCCCAGATGGTACAACCTGAAGATTGCCGGCGATCCCggaggctctggaggaagcTCCGA
AGTCGAGTTTTCCCATGAGTACTGGATGAGACACGCATTGACTCTCGCAAAGAGGGCTCGAGATGAA
CGCGAGGTGCCCGTGGGGGCAGTACTCGTGCTCAACAATCGCGTAATCGGCGAAGGTTGGAATAGGG
CAATCGGACTCCACGACCCCACTGCACATGCGGAAATCATGGCCCTTCGACAGGGAGGGCTTGTGAT
GCAGAATTATCGACTTTATGATGCGACGCTGTACGTCACGTTTGAACCTTGCGTAATGTGCGCGGGA
GCTATGATTCACTCCCGCATTGGACGAGTTGTATTCGGTGTTCGCAACGCCAAGACGGGTGCCGCAG
GTTCACTGATGGACGTGCTGCATCATCCAGGCATGAACCACCGGGTAGAAATCACAGAAGGCATATT
GGCGGACGAATGTGCGGCGCTGTTGTGTCGTTTTTTTCGCATGCCCAGGCGGGTCTTTAACGCCCAG
AAAAAAGCACAATCCTCTACTGACGGCTCTTCTGGATCTGAAACACCTGGCACAAGCGAGAGCGCCA
CCCCTGAGAGCTCTGGCTCCTGGGAAGAAGAGAAGAAGAAGTGGGAAGAAGATAAGAAAAAGGACCC
GCTGGCCAAGATCCTGGGCAAGCTGGCTGAGTACGGACTGATCCCTCTGTTCATCCCCTACACCGAC
AGCAACGAGCCCATCGTGAAAGAAATCAAGTGGATGGAAAAGTCCCGGAACCAGAGCGTGCGGCGGC
TGGATAAGGACATGTTCATTCAGGCCCTGGAACGGTTCCTGAGCTGGGAGAGCTGGAACCTGAAAGT
GAAAGAGGAATACGAGAAGGTCGAGAAAGAGTACAAGACCCTGGAAGAGAGGATCAAAGAGGACATC
CAGGCTCTGAAGGCTCTGGAACAGTATGAGAAAGAGCGGCAAGAACAGCTGCTGCGGGACACCCTGA
ACACCAACGAGTACCGGCTGAGCAAGAGAGGCCTTAGAGGCTGGCGGGAAATCATCCAGAAATGGCT
GAAAATGGACGAGAACGAGCCCTCCGAGAAGTACCTGGAAGTGTTCAAGGACTACCAGCGGAAGCAC
CCTAGAGAGGCCGGCGATTACAGCGTGTACGAGTTCCTGTCCAAGAAAGAGAACCACTTCATCTGGC
GGAATCACCCTGAGTACCCCTACCTGTACGCCACCTTCTGCGAGATCGACAAGAAAAAGAAGGACGC
CAAGCAGCAGGCCACCTTCACACTGGCCGATCCTATCAATCACCCTCTGTGGGTCCGATTCGAGGAA
AGAAGCGGCAGCAACCTGAACAAGTACAGAATCCTGACCGAGCAGCTGCACACCGAGAAGCTGAAGA
AAAAGCTGACAGTGCAGCTGGACCGGCTGATCTACCCTACAGAATCTGGCGGCTGGGAAGAGAAGGG
CAAAGTGGACATTGTGCTGCTGCCCAGCCGGCAGTTCTACAACCAGATCTTCCTGGACATCGAGGAA
AAGGGCAAGCACGCCTTCACCTACAAGGATGAGAGCATCAAGTTCCCTCTGAAGGGCACACTCGGCG
GAGCCAGAGTGCAGTTCGACAGAGATCACCTGAGAAGATACCCTCACAAGGTGGAAAGCGGCAACGT
GGGCAGAATCTACTTCAACATGACCGTGAACATCGAGCCTACAGAGTCCCCAGTGTCCAAGTCTCTG
AAGATCCACCGGGACGACTTCCCCAAGGTGGTCAACTTCAAGCCCAAAGAACTGACCGAGTGGATCA
AGGACAGCAAGGGCAAGAAACTGAAGTCCGGCATCGAGTCCCTGGAAATCGGCCTGAGAGTGATGAG
CATCGACCTGGGACAGAGACAGGCCGCTGCCGCCTCTATTTTCGAGGTGGTGGATCAGAAGCCCGAC
ATCGAAGGCAAGCTGTTTTTCCCAATCAAGGGCACCGAGCTGTATGCCGTGCACAGAGCCAGCTTCA
ACATCAAGCTGCCCGGCGAGACACTGGTCAAGAGCAGAGAAGTGCTGCGGAAGGCCAGAGAGGACAA
TCTGAAACTGATGAACCAGAAGCTCAACTTCCTGCGGAACGTGCTGCACTTCCAGCAGTTCGAGGAC
ATCACCGAGAGAGAGAAGCGGGTCACCAAGTGGATCAGCAGACAAGAGAACAGCGACGTGCCCCTGG
TGTACCAGGATGAGCTGATCCAGATCCGCGAGCTGATGTACAAGCCTTACAAGGACTGGGTCGCCTT
CCTGAAGCAGCTCCACAAGAGACTGGAAGTCGAGATCGGCAAAGAAGTGAAGCACTGGCGGAAGTCC
CTGAGCGACGGAAGAAAGGGCCTGTACGGCATCTCCCTGAAGAACATCGACGAGATCGATCGGACCC
GGAAGTTCCTGCTGAGATGGTCCCTGAGGCCTACCGAACCTGGCGAAGTGCGTAGACTGGAACCCGG
CCAGAGATTCGCCATCGACCAGCTGAATCACCTGAACGCCCTGAAAGAAGATCGGCTGAAGAAGATG
GCCAACACCATCATCATGCACGCCCTGGGCTACTGCTACGACGTGCGGAAGAAGAAATGGCAGGCTA
AGAACCCCGCCTGCCAGATCATCCTGTTCGAGGATCTGAGCAACTACAACCCCTACGAGGAAAGGTC

CCGCTTCGAGAACAGCAAGCTCATGAAGTGGTCCAGACGCGAGATCCCCAGACAGGTTGCACTGCAG
GGCGAGATCTATGGCCTGCAAGTGGGAGAAGTGGGCGCTCAGTTCAGCAGCAGATTCCACGCCAAGA
CAGGCAGCCCTGGCATCAGATGTAGCGTCGTGACCAAAGAGAAGCTGCAGGACAATCGGTTCTTCAA
GAATCTGCAGAGAGAGGGCAGACTGACCCTGGACAAAATCGCCGTGCTGAAAGAGGGCGATCTGTAC
CCAGACAAAGGCGGCGAGAAGTTCATCAGCCTGAGCAAGGATCGGAAGTGCGTGACCACACACGCCG
ACATCAACGCCGCTCAGAACCTGCAGAAGCGGTTCTGGACAAGAACCCACGGCTTCTACAAGGTGTA
CTGCAAGGCCTACCAGGTGGACGGCCAGACCGTGTACATCCCTGAGAGCAAGGACCAGAAGCAGA

[1708] AGATCATCGAAGAGTTCGGCGAGGGCTACTTCATTCTGAAGGACGGGGTGTACGAATGGGTCAACGC
CGGCAAGCTGAAAATCAAGAAGGGCAGCTCCAAGCAGAGCAGCAGCGAGCTGGTGGATAGCGACATC
CTGAAAGACAGCTTCGACCTGGCCTCCGAGCTGAAAGGCGAAAAGCTGATGCTGTACAGGGACCCCA
GCGGCAATGTGTTCCCCAGCGACAAATGGATGGCCGCTGGCGTGTTCTTCGGAAAGCTGGAACGCAT
CCTGATCAGCAAGCTGACCAACCAGTACTCCATCAGCACCATCGAGGACGACAGCAGCAAGCAGTCT
ATGAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGGATCC*TACCCATACG*
ATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTA
*TGCC*TAA

[1709] MAPKKKRKVGIHGVPAAATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYEHHE
[1710] QDPKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVEKKGEANQLSN
[1711] KFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPGGSGGSSEVEFSHEYWMRHALTLAKRARDERE
[1712] VPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAM
[1713] IHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKK
[1714] AQSSTDGSSGSETPGTSESATPESSGSWEEEKKKWEEDKKKDPLAKILGKLAEYGLIPLFIPYTDSN
[1715] EPIVKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWESWNLKVKEEYEKVEKEYKTLEERIKEDIQA
[1716] LKALEQYEKERQEQLLRDTLNTNEYRLSKRGLRGWREIIQKWLKMDENEPSEKYLEVFKDYQRKHPR
[1717] EAGDYSVYEFLSKKENHFIWRNHPEYPYLYATFCEIDKKKKDAKQQATFTLADPINHPLWVRFEERS
[1718] GSNLNKYRILTEQLHTEKLKKKLTVQLDRLIYPTESGGWEEKGKVDIVLLPSRQFYNQIFLDIEEKG
[1719] KHAFTYKDESIKFPLKGTLGGARVQFDRDHLRRYPHKVESGNVGRIYFNMTVNIEPTESPVSKSLKI
[1720] HRDDFPKVVNFKPKELTEWIKDSKGKKLKSGIESLEIGLRVMSIDLGQRQAAAASIFEVVDQKPDIE
[1721] GKLFFPIKGTELYAVHRASFNIKLPGETLVKSREVLRKAREDNLKLMNQKLNFLRNVLHFQQFEDIT
[1722] EREKRVTKWISRQENSDVPLVYQDELIQIRELMYKPYKDWVAFLKQLHKRLEVEIGKEVKHWRKSLS
[1723] DGRKGLYGISLKNIDEIDRTRKFLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHLNALKEDRLKKMAN
[1724] TIIMHALGYCYDVRKKKWQAKNPACQIILFEDLSNYNPYEERSRFENSKLMKWSRREIPRQVALQGE
[1725] IYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKLQDNRFFKNLQREGRLTLDKIAVLKEGDLYPD
[1726] KGGEKFISLSKDRKCVTTHADINAAQNLQKRFWTRTHGFYKVYCKAYQVDGQTVYIPESKDQKQKII
[1727] EEFGEGYFILKDGVYEWVNAGKLKIKKGSSKQSSSELVDSDILKDSFDLASELKGEKLMLYRDPSGN
[1728] VFPSDKWMAAGVFFGKLERILISKLTNQYSISTIEDDSSKQSMKRPAATKKAGQAKKKKGSYPYDVP
[1729] DYAYPYDVPDYAYPYDVPDYA
[1730] 在K255的BhCas12b GGSGGS-ABE8-Xten20

[1731]

GCCACCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGCCACCAGAT
CCTTCATCCTGAAGATCGAGCCCAACGAGGAAGTGAAGAAAGGCCTCTGGAAAACCCACGAGGTGCT
GAACCACGGAATCGCCTACTACATGAATATCCTGAAGCTGATCCGGCAAGAGGCCATCTACGAGCAC
CACGAGCAGGACCCCAAGAATCCCAAGAAGGTGTCCAAGGCCGAGATCCAGGCCGAGCTGTGGGATT
TCGTGCTGAAGATGCAGAAGTGCAACAGCTTCACACACGAGGTGGACAAGGACGAGGTGTTCAACAT
CCTGAGAGAGCTGTACGAGGAACTGGTGCCCAGCAGCGTGGAAAAGAAGGGCGAAGCCAACCAGCTG
AGCAACAAGTTTCTGTACCCTCTGGTGGACCCCAACAGCCAGTCTGGAAAGGGAACAGCCAGCAGCG
GCAGAAAGCCCAGATGGTACAACCTGAAGATTGCCGGCGATCCCTCCTGGGAAGAAGAGAAGAAGAA
GTGGGAAGAAGATAAGAAAAAGGACCCGCTGGCCAAGATCCTGGGCAAGCTGGCTGAGTACGGACTG
ATCCCTCTGTTCATCCCCTACACCGACAGCAACGAGCCCATCGTGAAAGAAATCAAGTGGATGGAAA
AGTCCCGGAACCAGAGCGTGCGGCGGCTGGATAAGGACATGTTCATTCAGGCCCTGGAACGGTTCCT
GAGCTGGGAGAGCTGGAACCTGAAAGTGAAAGAGGAATACGAGAAGGTCGAGAAAGAGTACAAGACC
CTGGAAGAGAGGATCAAAggaggctctggaggaagcTCCGAAGTCGAGTTTTCCCATGAGTACTGGA
TGAGACACGCATTGACTCTCGCAAAGAGGGCTCGAGATGAACGCGAGGTGCCCGTGGGGGCAGTACT
CGTGCTCAACAATCGCGTAATCGGCGAAGGTTGGAATAGGGCAATCGGACTCCACGACCCCACTGCA
CATGCGGAAATCATGGCCCTTCGACAGGGAGGGCTTGTGATGCAGAATTATCGACTTTATGATGCGA
CGCTGTACGTCACGTTTGAACCTTGCGTAATGTGCGCGGGAGCTATGATTCACTCCCGCATTGGACG
AGTTGTATTCGGTGTTCGCAACGCCAAGACGGGTGCCGCAGGTTCACTGATGGACGTGCTGCATCAT
CCAGGCATGAACCACCGGGTAGAAATCACAGAAGGCATATTGGCGGACGAATGTGCGGCGCTGTTGT
GTCGTTTTTTTCGCATGCCCAGGCGGGTCTTTAACGCCCAGAAAAAAGCACAATCCTCTACTGACGG
CTCTTCTGGATCTGAAACACCTGGCACAAGCGAGAGCGCCACCCCTGAGAGCTCTGGCGAGGACATC
CAGGCTCTGAAGGCTCTGGAACAGTATGAGAAAGAGCGGCAAGAACAGCTGCTGCGGGACACCCTGA
ACACCAACGAGTACCGGCTGAGCAAGAGAGGCCTTAGAGGCTGGCGGGAAATCATCCAGAAATGGCT
GAAAATGGACGAGAACGAGCCCTCCGAGAAGTACCTGGAAGTGTTCAAGGACTACCAGCGGAAGCAC
CCTAGAGAGGCCGGCGATTACAGCGTGTACGAGTTCCTGTCCAAGAAAGAGAACCACTTCATCTGGC
GGAATCACCCTGAGTACCCCTACCTGTACGCCACCTTCTGCGAGATCGACAAGAAAAAGAAGGACGC
CAAGCAGCAGGCCACCTTCACACTGGCCGATCCTATCAATCACCCTCTGTGGGTCCGATTCGAGGAA
AGAAGCGGCAGCAACCTGAACAAGTACAGAATCCTGACCGAGCAGCTGCACACCGAGAAGCTGAAGA
AAAAGCTGACAGTGCAGCTGGACCGGCTGATCTACCCTACAGAATCTGGCGGCTGGGAAGAGAAGGG
CAAAGTGGACATTGTGCTGCTGCCCAGCCGGCAGTTCTACAACCAGATCTTCCTGGACATCGAGGAA
AAGGGCAAGCACGCCTTCACCTACAAGGATGAGAGCATCAAGTTCCCTCTGAAGGGCACACTCGGCG
GAGCCAGAGTGCAGTTCGACAGAGATCACCTGAGAAGATACCCTCACAAGGTGGAAAGCGGCAACGT
GGGCAGAATCTACTTCAACATGACCGTGAACATCGAGCCTACAGAGTCCCCAGTGTCCAAGTCTCTG
AAGATCCACCGGGACGACTTCCCCAAGGTGGTCAACTTCAAGCCCAAAGAACTGACCGAGTGGATCA
AGGACAGCAAGGGCAAGAAACTGAAGTCCGGCATCGAGTCCCTGGAAATCGGCCTGAGAGTGATGAG
CATCGACCTGGGACAGAGACAGGCCGCTGCCGCCTCTATTTTCGAGGTGGTGGATCAGAAGCCCGAC
ATCGAAGGCAAGCTGTTTTTCCCAATCAAGGGCACCGAGCTGTATGCCGTGCACAGAGCCAGCTTCA
ACATCAAGCTGCCCGGCGAGACACTGGTCAAGAGCAGAGAAGTGCTGCGGAAGGCCAGAGAGGACAA
TCTGAAACTGATGAACCAGAAGCTCAACTTCCTGCGGAACGTGCTGCACTTCCAGCAGTTCGAGGAC
ATCACCGAGAGAGAGAAGCGGGTCACCAAGTGGATCAGCAGACAAGAGAACAGCGACGTGCCCCTGG

[1732]

TGTACCAGGATGAGCTGATCCAGATCCGCGAGCTGATGTACAAGCCTTACAAGGACTGGGTCGCCTT
CCTGAAGCAGCTCCACAAGAGACTGGAAGTCGAGATCGGCAAAGAAGTGAAGCACTGGCGGAAGTCC
CTGAGCGACGGAAGAAAGGGCCTGTACGGCATCTCCCTGAAGAACATCGACGAGATCGATCGGACCC
GGAAGTTCCTGCTGAGATGGTCCCTGAGGCCTACCGAACCTGGCGAAGTGCGTAGACTGGAACCCGG
CCAGAGATTCGCCATCGACCAGCTGAATCACCTGAACGCCCTGAAAGAAGATCGGCTGAAGAAGATG
GCCAACACCATCATCATGCACGCCCTGGGCTACTGCTACGACGTGCGGAAGAAGAAATGGCAGGCTA
AGAACCCCGCCTGCCAGATCATCCTGTTCGAGGATCTGAGCAACTACAACCCCTACGAGGAAAGGTC
CCGCTTCGAGAACAGCAAGCTCATGAAGTGGTCCAGACGCGAGATCCCCAGACAGGTTGCACTGCAG
GGCGAGATCTATGGCCTGCAAGTGGGAGAAGTGGGCGCTCAGTTCAGCAGCAGATTCCACGCCAAGA
CAGGCAGCCCTGGCATCAGATGTAGCGTCGTGACCAAAGAGAAGCTGCAGGACAATCGGTTCTTCAA
GAATCTGCAGAGAGAGGGCAGACTGACCCTGGACAAAATCGCCGTGCTGAAAGAGGGCGATCTGTAC
CCAGACAAAGGCGGCGAGAAGTTCATCAGCCTGAGCAAGGATCGGAAGTGCGTGACCACACACGCCG
ACATCAACGCCGCTCAGAACCTGCAGAAGCGGTTCTGGACAAGAACCCACGGCTTCTACAAGGTGTA
CTGCAAGGCCTACCAGGTGGACGGCCAGACCGTGTACATCCCTGAGAGCAAGGACCAGAAGCAGAAG
ATCATCGAAGAGTTCGGCGAGGGCTACTTCATTCTGAAGGACGGGGTGTACGAATGGGTCAACGCCG
GCAAGCTGAAAATCAAGAAGGGCAGCTCCAAGCAGAGCAGCAGCGAGCTGGTGGATAGCGACATCCT
GAAAGACAGCTTCGACCTGGCCTCCGAGCTGAAAGGCGAAAAGCTGATGCTGTACAGGGACCCCAGC
GGCAATGTGTTCCCCAGCGACAAATGGATGGCCGCTGGCGTGTTCTTCGGAAAGCTGGAACGCATCC
TGATCAGCAAGCTGACCAACCAGTACTCCATCAGCACCATCGAGGACGACAGCAGCAAGCAGTCTAT
GAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGGATCC*TACCCATACGAT*
GTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATG
*CC*TAA

[1733] MAPKKKRKVGIHGVPAAATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYEHHE

[1734] QDPKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVEKKGEANQLSN

[1735] KFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEEKKKWEEDKKKDPLAKILGKLAEYGLIP

[1736] LFIPYTDSNEPIVKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWESWNLKVKEEYEKVEKEYKTLE

[1737] ERIKGGSGGSSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHA

[1738] EIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPG

[1739] MNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESATPESSGEDIQA

[1740] LKALEQYEKERQEQLLRDTLNTNEYRLSKRGLRGWREIIQKWLKMDENEPSEKYLEVFKDYQRKHPR

[1741] EAGDYSVYEFLSKKENHFIWRNHPEYPYLYATFCEIDKKKKDAKQQATFTLADPINHPLWVRFEERS

[1742] GSNLNKYRILTEQLHTEKLKKKLTVQLDRLIYPTESGGWEEKGKVDIVLLPSRQFYNQIFLDIEEKG

[1743] KHAFTYKDESIKFPLKGTLGGARVQFDRDHLRRYPHKVESGNVGRIYFNMTVNIEPTESPVSKSLKI

[1744] HRDDFPKVVNFKPKELTEWIKDSKGKKLKSGIESLEIGLRVMSIDLGQRQAAAASIFEVVDQKPDIE

[1745] GKLFFPIKGTELYAVHRASFNIKLPGETLVKSREVLRKAREDNLKLMNQKLNFLRNVLHFQQFEDIT

[1746] EREKRVTKWISRQENSDVPLVYQDELIQIRELMYKPYKDWVAFLKQLHKRLEVEIGKEVKHWRKSLS

[1747] DGRKGLYGISLKNIDEIDRTRKFLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHLNALKEDRLKKMAN

[1748] TIIMHALGYCYDVRKKKWQAKNPACQIILFEDLSNYNPYEERSRFENSKLMKWSRREIPRQVALQGE

[1749] IYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKLQDNRFFKNLQREGRLTLDKIAVLKEGDLYPD

[1750] KGGEKFISLSKDRKCVTTHADINAAQNLQKRFWTRTHGFYKVYCKAYQVDGQTVYIPESKDQKQKII

[1751] EEFGEGYFILKDGVYEWVNAGKLKIKKGSSKQSSSELVDSDILKDSFDLASELKGEKLMLYRDPSGN

[1752] VFPSDKWMAAGVFFGKLERILISKLTNQYSISTIEDDSSKQSMKRPAATKKAGQAKKKKGSYPYDVP

[1753] DYAYPYDVPDYAYPYDVPDYA

[1754] 在D306的BhCas12b GGSGGS-ABE8-Xten20

GCCACCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGCCACCAGAT
CCTTCATCCTGAAGATCGAGCCCAACGAGGAAGTGAAGAAAGGCCTCTGGAAAACCCACGAGGTGCT
GAACCACGGAATCGCCTACTACATGAATATCCTGAAGCTGATCCGGCAAGAGGCCATCTACGAGCAC
CACGAGCAGGACCCCAAGAATCCCAAGAAGGTGTCCAAGGCCGAGATCCAGGCCGAGCTGTGGGATT
TCGTGCTGAAGATGCAGAAGTGCAACAGCTTCACACACGAGGTGGACAAGGACGAGGTGTTCAACAT
CCTGAGAGAGCTGTACGAGGAACTGGTGCCCAGCAGCGTGGAAAAGAAGGGCGAAGCCAACCAGCTG
AGCAACAAGTTTCTGTACCCTCTGGTGGACCCCAACAGCCAGTCTGGAAAGGGAACAGCCAGCAGCG
GCAGAAAGCCCAGATGGTACAACCTGAAGATTGCCGGCGATCCCTCCTGGGAAGAAGAGAAGAAGAA
GTGGGAAGAAGATAAGAAAAAGGACCCGCTGGCCAAGATCCTGGGCAAGCTGGCTGAGTACGGACTG
ATCCCTCTGTTCATCCCCTACACCGACAGCAACGAGCCCATCGTGAAAGAAATCAAGTGGATGGAAA
AGTCCCGGAACCAGAGCGTGCGGCGGCTGGATAAGGACATGTTCATTCAGGCCCTGGAACGGTTCCT
GAGCTGGGAGAGCTGGAACCTGAAAGTGAAAGAGGAATACGAGAAGGTCGAGAAAGAGTACAAGACC
CTGGAAGAGAGGATCAAAGAGGACATCCAGGCTCTGAAGGCTCTGGAACAGTATGAGAAAGAGCGGC
AAGAACAGCTGCTGCGGGACACCCTGAACACCAACGAGTACCGGCTGAGCAAGAGAGGCCTTAGAGG
CTGGCGGGAAATCATCCAGAAATGGCTGAAAATGGACggaggctctggaggaagcTCCGAAGTCGAG
TTTTCCCATGAGTACTGGATGAGACACGCATTGACTCTCGCAAAGAGGGCTCGAGATGAACGCGAGG
[1755] TGCCCGTGGGGGCAGTACTCGTGCTCAACAATCGCGTAATCGGCGAAGGTTGGAATAGGGCAATCGG
ACTCCACGACCCCACTGCACATGCGGAAATCATGGCCCTTCGACAGGGAGGGCTTGTGATGCAGAAT
TATCGACTTTATGATGCGACGCTGTACGTCACGTTTGAACCTTGCGTAATGTGCGCGGGAGCTATGA
TTCACTCCCGCATTGGACGAGTTGTATTCGGTGTTCGCAACGCCAAGACGGGTGCCGCAGGTTCACT
GATGGACGTGCTGCATCATCCAGGCATGAACCACCGGGTAGAAATCACAGAAGGCATATTGGCGGAC
GAATGTGCGGCGCTGTTGTGTCGTTTTTTTCGCATGCCCAGGCGGGTCTTTAACGCCCAGAAAAAAG
CACAATCCTCTACTGACGGCTCTTCTGGATCTGAAACACCTGGCACAAGCGAGAGCGCCACCCCTGA
GAGCTCTGGCGAGAACGAGCCCTCCGAGAAGTACCTGGAAGTGTTCAAGGACTACCAGCGGAAGCAC
CCTAGAGAGGCCGGCGATTACAGCGTGTACGAGTTCCTGTCCAAGAAAGAGAACCACTTCATCTGGC
GGAATCACCCTGAGTACCCCTACCTGTACGCCACCTTCTGCGAGATCGACAAGAAAAAGAAGGACGC
CAAGCAGCAGGCCACCTTCACACTGGCCGATCCTATCAATCACCCTCTGTGGGTCCGATTCGAGGAA
AGAAGCGGCAGCAACCTGAACAAGTACAGAATCCTGACCGAGCAGCTGCACACCGAGAAGCTGAAGA
AAAAGCTGACAGTGCAGCTGGACCGGCTGATCTACCCTACAGAATCTGGCGGCTGGGAAGAGAAGGG
CAAAGTGGACATTGTGCTGCTGCCCAGCCGGCAGTTCTACAACCAGATCTTCCTGGACATCGAGGAA
AAGGGCAAGCACGCCTTCACCTACAAGGATGAGAGCATCAAGTTCCCTCTGAAGGGCACACTCGGCG
GAGCCAGAGTGCAGTTCGACAGAGATCACCTGAGAAGATACCCTCACAAGGTGGAAAGCGGCAACGT
GGGCAGAATCTACTTCAACATGACCGTGAACATCGAGCCTACAGAGTCCCCAGTGTCCAAGTCTCTG

AAGATCCACCGGGACGACTTCCCCAAGGTGGTCAACTTCAAGCCCAAAGAACTGACCGAGTGGATCA
AGGACAGCAAGGGCAAGAAACTGAAGTCCGGCATCGAGTCCCTGGAAATCGGCCTGAGAGTGATGAG
CATCGACCTGGGACAGAGACAGGCCGCTGCCGCCTCTATTTTCGAGGTGGTGGATCAGAAGCCCGAC
ATCGAAGGCAAGCTGTTTTTCCCAATCAAGGGCACCGAGCTGTATGCCGTGCACAGAGCCAGCTTCA
ACATCAAGCTGCCCGGCGAGACACTGGTCAAGAGCAGAGAAGTGCTGCGGAAGGCCAGAGAGGACAA
TCTGAAACTGATGAACCAGAAGCTCAACTTCCTGCGGAACGTGCTGCACTTCCAGCAGTTCGAGGAC
ATCACCGAGAGAGAGAAGCGGGTCACCAAGTGGATCAGCAGACAAGAGAACAGCGACGTGCCCCTGG
TGTACCAGGATGAGCTGATCCAGATCCGCGAGCTGATGTACAAGCCTTACAAGGACTGGGTCGCCTT
CCTGAAGCAGCTCCACAAGAGACTGGAAGTCGAGATCGGCAAAGAAGTGAAGCACTGGCGGAAGTCC
CTGAGCGACGGAAGAAAGGGCCTGTACGGCATCTCCCTGAAGAACATCGACGAGATCGATCGGACCC
GGAAGTTCCTGCTGAGATGGTCCCTGAGGCCTACCGAACCTGGCGAAGTGCGTAGACTGGAACCCGG
CCAGAGATTCGCCATCGACCAGCTGAATCACCTGAACGCCCTGAAAGAAGATCGGCTGAAGAAGATG
GCCAACACCATCATCATGCACGCCCTGGGCTACTGCTACGACGTGCGGAAGAAGAAATGGCAGGCTA
AGAACCCCGCCTGCCAGATCATCCTGTTCGAGGATCTGAGCAACTACAACCCCTACGAGGAAAGGTC
[1756] CCGCTTCGAGAACAGCAAGCTCATGAAGTGGTCCAGACGCGAGATCCCCAGACAGGTTGCACTGCAG
GGCGAGATCTATGGCCTGCAAGTGGGAGAAGTGGGCGCTCAGTTCAGCAGCAGATTCCACGCCAAGA
CAGGCAGCCCTGGCATCAGATGTAGCGTCGTGACCAAAGAGAAGCTGCAGGACAATCGGTTCTTCAA
GAATCTGCAGAGAGAGGGCAGACTGACCCTGGACAAAATCGCCGTGCTGAAAGAGGGCGATCTGTAC
CCAGACAAAGGCGGCGAGAAGTTCATCAGCCTGAGCAAGGATCGGAAGTGCGTGACCACACACGCCG
ACATCAACGCCGCTCAGAACCTGCAGAAGCGGTTCTGGACAAGAACCCACGGCTTCTACAAGGTGTA
CTGCAAGGCCTACCAGGTGGACGGCCAGACCGTGTACATCCCTGAGAGCAAGGACCAGAAGCAGAAG
ATCATCGAAGAGTTCGGCGAGGGCTACTTCATTCTGAAGGACGGGGTGTACGAATGGGTCAACGCCG
GCAAGCTGAAAATCAAGAAGGGCAGCTCCAAGCAGAGCAGCAGCGAGCTGGTGGATAGCGACATCCT
GAAAGACAGCTTCGACCTGGCCTCCGAGCTGAAAGGCGAAAAGCTGATGCTGTACAGGGACCCCAGC
GGCAATGTGTTCCCCAGCGACAAATGGATGGCCGCTGGCGTGTTCTTCGGAAAGCTGGAACGCATCC
TGATCAGCAAGCTGACCAACCAGTACTCCATCAGCACCATCGAGGACGACAGCAGCAAGCAGTCTAT
GAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGGATCC*TACCCATACGAT*
GTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATG
*CC*TAA

[1757] MAPKKKRKVGIHGVPAAATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYEHHE
[1758] QDPKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVEKKGEANQLSN
[1759] KFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEEKKKWEEDKKKDPLAKILGKLAEYGLIP
[1760] LFIPYTDSNEPIVKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWESWNLKVKEEYEKVEKEYKTLE
[1761] ERIKEDIQALKALEQYEKERQEQLLRDTLNTNEYRLSKRGLRGWREIIQKWLKMDGGSGGSSEVEFS
[1762] HEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYR
[1763] LYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADEC
[1764] AALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESATPESSGENEPSEKYLEVFKDYQRKHPR
[1765] EAGDYSVYEFLSKKENHFIWRNHPEYPYLYATFCEIDKKKKDAKQQATFTLADPINHPLWVRFEERS
[1766] GSNLNKYRILTEQLHTEKLKKKLTVQLDRLIYPTESGGWEEKGKVDIVLLPSRQFYNQIFLDIEEKG
[1767] KHAFTYKDESIKFPLKGTLGGARVQFDRDHLRRYPHKVESGNVGRIYFNMTVNIEPTESPVSKSLKI

[1768] HRDDFPKVVNFKPKELTEWIKDSKGKKLKSGIESLEIGLRVMSIDLGQRQAAAASIFEVVDQKPDIE

[1769] GKLFFPIKGTELYAVHRASFNIKLPGETLVKSREVLRKAREDNLKLMNQKLNFLRNVLHFQQFEDIT

[1770] EREKRVTKWISRQENSDVPLVYQDELIQIRELMYKPYKDWVAFLKQLHKRLEVEIGKEVKHWRKSLS

[1771] DGRKGLYGISLKNIDEIDRTRKFLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHLNALKEDRLKKMAN

[1772] TIIMHALGYCYDVRKKKWQAKNPACQIILFEDLSNYNPYEERSRFENSKLMKWSRREIPRQVALQGE

[1773] IYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKLQDNRFFKNLQREGRLTLDKIAVLKEGDLYPD

[1774] KGGEKFISLSKDRKCVTTHADINAAQNLQKRFWTRTHGFYKVYCKAYQVDGQTVYIPESKDQKQKII

[1775] EEFGEGYFILKDGVYEWVNAGKLKIKKGSSKQSSSELVDSDILKDSFDLASELKGEKLMLYRDPSGN

[1776] VFPSDKWMAAGVFFGKLERILISKLTNQYSISTIEDDSSKQSMKRPAATKKAGQAKKKKGSYPYDVP

[1777] DYAYPYDVPDYAYPYDVPDYA

[1778] 在D980的BhCas12b GGSGGS-ABE8-Xten20

[1779]
GCCACCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGCCACCAGAT
CCTTCATCCTGAAGATCGAGCCCAACGAGGAAGTGAAGAAAGGCCTCTGGAAAACCCACGAGGTGCT
GAACCACGGAATCGCCTACTACATGAATATCCTGAAGCTGATCCGGCAAGAGGCCATCTACGAGCAC
CACGAGCAGGACCCCAAGAATCCCAAGAAGGTGTCCAAGGCCGAGATCCAGGCCGAGCTGTGGGATT
TCGTGCTGAAGATGCAGAAGTGCAACAGCTTCACACACGAGGTGGACAAGGACGAGGTGTTCAACAT
CCTGAGAGAGCTGTACGAGGAACTGGTGCCCAGCAGCGTGGAAAAGAAGGGCGAAGCCAACCAGCTG
AGCAACAAGTTTCTGTACCCTCTGGTGGACCCCAACAGCCAGTCTGGAAAGGGAACAGCCAGCAGCG
GCAGAAAGCCCAGATGGTACAACCTGAAGATTGCCGGCGATCCCTCCTGGGAAGAAGAGAAGAAGAA
GTGGGAAGAAGATAAGAAAAAGGACCCGCTGGCCAAGATCCTGGGCAAGCTGGCTGAGTACGGACTG
ATCCCTCTGTTCATCCCCTACACCGACAGCAACGAGCCCATCGTGAAAGAAATCAAGTGGATGGAAA
AGTCCCGGAACCAGAGCGTGCGGCGGCTGGATAAGGACATGTTCATTCAGGCCCTGGAACGGTTCCT
GAGCTGGGAGAGCTGGAACCTGAAAGTGAAAGAGGAATACGAGAAGGTCGAGAAAGAGTACAAGACC
CTGGAAGAGAGGATCAAAGAGGACATCCAGGCTCTGAAGGCTCTGGAACAGTATGAGAAAGAGCGGC
AAGAACAGCTGCTGCGGGACACCCTGAACACCAACGAGTACCGGCTGAGCAAGAGAGGCCTTAGAGG
CTGGCGGGAAATCATCCAGAAATGGCTGAAAATGGACGAGAACGAGCCCTCCGAGAAGTACCTGGAA
GTGTTCAAGGACTACCAGCGGAAGCACCCTAGAGAGGCCGGCGATTACAGCGTGTACGAGTTCCTGT
CCAAGAAAGAGAACCACTTCATCTGGCGGAATCACCCTGAGTACCCCTACCTGTACGCCACCTTCTG
CGAGATCGACAAGAAAAAGAAGGACGCCAAGCAGCAGGCCACCTTCACACTGGCCGATCCTATCAAT
CACCCTCTGTGGGTCCGATTCGAGGAAAGAAGCGGCAGCAACCTGAACAAGTACAGAATCCTGACCG
AGCAGCTGCACACCGAGAAGCTGAAGAAAAAGCTGACAGTGCAGCTGGACCGGCTGATCTACCCTAC
AGAATCTGGCGGCTGGGAAGAGAAGGGCAAAGTGGACATTGTGCTGCTGCCCAGCCGGCAGTTCTAC
AACCAGATCTTCCTGGACATCGAGGAAAAGGGCAAGCACGCCTTCACCTACAAGGATGAGAGCATCA
AGTTCCCTCTGAAGGGCACACTCGGCGGAGCCAGAGTGCAGTTCGACAGAGATCACCTGAGAAGATA
CCCTCACAAGGTGGAAAGCGGCAACGTGGGCAGAATCTACTTCAACATGACCGTGAACATCGAGCCT
ACAGAGTCCCCAGTGTCCAAGTCTCTGAAGATCCACCGGGACGACTTCCCCAAGGTGGTCAACTTCA
AGCCCAAAGAACTGACCGAGTGGATCAAGGACAGCAAGGGCAAGAAACTGAAGTCCGGCATCGAGTC

CCTGGAAATCGGCCTGAGAGTGATGAGCATCGACCTGGGACAGAGACAGGCCGCTGCCGCCTCTATT
TTCGAGGTGGTGGATCAGAAGCCCGACATCGAAGGCAAGCTGTTTTTCCCAATCAAGGGCACCGAGC
TGTATGCCGTGCACAGAGCCAGCTTCAACATCAAGCTGCCCGGCGAGACACTGGTCAAGAGCAGAGA
AGTGCTGCGGAAGGCCAGAGAGGACAATCTGAAACTGATGAACCAGAAGCTCAACTTCCTGCGGAAC
GTGCTGCACTTCCAGCAGTTCGAGGACATCACCGAGAGAGAGAAGCGGGTCACCAAGTGGATCAGCA
GACAAGAGAACAGCGACGTGCCCCTGGTGTACCAGGATGAGCTGATCCAGATCCGCGAGCTGATGTA
CAAGCCTTACAAGGACTGGGTCGCCTTCCTGAAGCAGCTCCACAAGAGACTGGAAGTCGAGATCGGC
AAAGAAGTGAAGCACTGGCGGAAGTCCCTGAGCGACGGAAGAAAGGGCCTGTACGGCATCTCCCTGA
AGAACATCGACGAGATCGATCGGACCCGGAAGTTCCTGCTGAGATGGTCCCTGAGGCCTACCGAACC
TGGCGAAGTGCGTAGACTGGAACCCGGCCAGAGATTCGCCATCGACCAGCTGAATCACCTGAACGCC
CTGAAAGAAGATCGGCTGAAGAAGATGGCCAACACCATCATCATGCACGCCCTGGGCTACTGCTACG
ACGTGCGGAAGAAGAAATGGCAGGCTAAGAACCCCGCCTGCCAGATCATCCTGTTCGAGGATCTGAG
CAACTACAACCCCTACGAGGAAAGGTCCCGCTTCGAGAACAGCAAGCTCATGAAGTGGTCCAGACGC
GAGATCCCCAGACAGGTTGCACTGCAGGGCGAGATCTATGGCCTGCAAGTGGGAGAAGTGGGCGCTC
AGTTCAGCAGCAGATTCCACGCCAAGACAGGCAGCCCTGGCATCAGATGTAGCGTCGTGACCAAAGA
GAAGCTGCAGGACAATCGGTTCTTCAAGAATCTGCAGAGAGAGGGCAGACTGACCCTGGACAAAATC
GCCGTGCTGAAAGAGGGCGATCTGTACCCAGACAAAGGCGGCGAGAAGTTCATCAGCCTGAGCAAGG
ATCGGAAGTGCGTGACCACACACGCCGACATCAACGCCGCTCAGAACCTGCAGAAGCGGTTCTGGAC

[1780]

AAGAACCCACGGCTTCTACAAGGTGTACTGCAAGGCCTACCAGGTGGACggaggctctggaggaagc
TCCGAAGTCGAGTTTTCCCATGAGTACTGGATGAGACACGCATTGACTCTCGCAAAGAGGGCTCGAG
ATGAACGCGAGGTGCCCGTGGGGGCAGTACTCGTGCTCAACAATCGCGTAATCGGCGAAGGTTGGAA
TAGGGCAATCGGACTCCACGACCCCACTGCACATGCGGAAATCATGGCCCTTCGACAGGGAGGGCTT
GTGATGCAGAATTATCGACTTTATGATGCGACGCTGTACGTCACGTTTGAACCTTGCGTAATGTGCG
CGGGAGCTATGATTCACTCCCGCATTGGACGAGTTGTATTCGGTGTTCGCAACGCCAAGACGGGTGC
CGCAGGTTCACTGATGGACGTGCTGCATCATCCAGGCATGAACCACCGGGTAGAAATCACAGAAGGC
ATATTGGCGGACGAATGTGCGGCGCTGTTGTGTCGTTTTTTTCGCATGCCCAGGCGGGTCTTTAACG
CCCAGAAAAAAGCACAATCCTCTACTGACGGCTCTTCTGGATCTGAAACACCTGGCACAAGCGAGAG
CGCCACCCCTGAGAGCTCTGGCGGCCAGACCGTGTACATCCCTGAGAGCAAGGACCAGAAGCAGAAG
ATCATCGAAGAGTTCGGCGAGGGCTACTTCATTCTGAAGGACGGGGTGTACGAATGGGTCAACGCCG
GCAAGCTGAAAATCAAGAAGGGCAGCTCCAAGCAGAGCAGCAGCGAGCTGGTGGATAGCGACATCCT
GAAAGACAGCTTCGACCTGGCCTCCGAGCTGAAAGGCGAAAAGCTGATGCTGTACAGGGACCCCAGC
GGCAATGTGTTCCCCAGCGACAAATGGATGGCCGCTGGCGTGTTCTTCGGAAAGCTGGAACGCATCC
TGATCAGCAAGCTGACCAACCAGTACTCCATCAGCACCATCGAGGACGACAGCAGCAAGCAGTCTAT
GAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGGATCC*TACCCATACGAT*
GTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATG
*CC*TAA

[1781] MAPKKKRKVGIHGVPAAATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYEHHE

[1782] QDPKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVEKKGEANQLSN

[1783] KFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEEKKKWEEDKKKDPLAKILGKLAEYGLIP

[1784] LFIPYTDSNEPIVKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWESWNLKVKEEYEKVEKEYKTLE

[1785] ERIKEDIQALKALEQYEKERQEQLLRDTLNTNEYRLSKRGLRGWREIIQKWLKMDENEPSEKYLEVF

[1786] KDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYATFCEIDKKKKDAKQQATFTLADPINHP

[1787] LWVRFEERSGSNLNKYRILTEQLHTEKLKKKLTVQLDRLIYPTESGGWEEKGKVDIVLLPSRQFYNQ

[1788] IFLDIEEKGKHAFTYKDESIKFPLKGTLGGARVQFDRDHLRRYPHKVESGNVGRIYFNMTVNIEPTE

[1789] SPVSKSLKIHRDDFPKVVNFKPKELTEWIKDSKGKKLKSGIESLEIGLRVMSIDLGQRQAAAASIFE

[1790] VVDQKPDIEGKLFFPIKGTELYAVHRASFNIKLPGETLVKSREVLRKAREDNLKLMNQKLNFLRNVL

[1791] HFQQFEDITEREKRVTKWISRQENSDVPLVYQDELIQIRELMYKPYKDWVAFLKQLHKRLEVEIGKE

[1792] VKHWRKSLSDGRKGLYGISLKNIDEIDRTRKFLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHLNALK

[1793] EDRLKKMANTIIMHALGYCYDVRKKKWQAKNPACQIILFEDLSNYNPYEERSRFENSKLMKWSRREI

[1794] PRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKLQDNRFFKNLQREGRLTLDKIAV

[1795] LKEGDLYPDKGGEKFISLSKDRKCVTTHADINAAQNLQKRFWTRTHGFYKVYCKAYQVDGGSGGSSE

[1796] VEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVM

[1797] QNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGIL

[1798] ADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESATPESSGGQTVYIPESKDQKQKII

[1799] EEFGEGYFILKDGVYEWVNAGKLKIKKGSSKQSSSELVDSDILKDSFDLASELKGEKLMLYRDPSGN

[1800] VFPSDKWMAAGVFFGKLERILISKLTNQYSISTIEDDSSKQSMKRPAATKKAGQAKKKKGSYPYDVP

[1801] DYAYPYDVPDYAYPYDVPDYA

[1802] 在K1019的BhCas12b GGSGGS-ABE8-Xten20

GCCACCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGCCACCAGAT
CCTTCATCCTGAAGATCGAGCCCAACGAGGAAGTGAAGAAAGGCCTCTGGAAAACCCACGAGGTGCT
GAACCACGGAATCGCCTACTACATGAATATCCTGAAGCTGATCCGGCAAGAGGCCATCTACGAGCAC
CACGAGCAGGACCCCAAGAATCCCAAGAAGGTGTCCAAGGCCGAGATCCAGGCCGAGCTGTGGGATT
TCGTGCTGAAGATGCAGAAGTGCAACAGCTTCACACACGAGGTGGACAAGGACGAGGTGTTCAACAT
CCTGAGAGAGCTGTACGAGGAACTGGTGCCCAGCAGCGTGGAAAAGAAGGGCGAAGCCAACCAGCTG
AGCAACAAGTTTCTGTACCCTCTGGTGGACCCCAACAGCCAGTCTGGAAAGGGAACAGCCAGCAGCG
GCAGAAAGCCCAGATGGTACAACCTGAAGATTGCCGGCGATCCCTCCTGGGAAGAAGAGAAGAAGAA
GTGGGAAGAAGATAAGAAAAAGGACCCGCTGGCCAAGATCCTGGGCAAGCTGGCTGAGTACGGACTG

[1803] ATCCCTCTGTTCATCCCCTACACCGACAGCAACGAGCCCATCGTGAAAGAAATCAAGTGGATGGAAA
AGTCCCGGAACCAGAGCGTGCGGCGGCTGGATAAGGACATGTTCATTCAGGCCCTGGAACGGTTCCT
GAGCTGGGAGAGCTGGAACCTGAAAGTGAAAGAGGAATACGAGAAGGTCGAGAAAGAGTACAAGACC
CTGGAAGAGAGGATCAAAGAGGACATCCAGGCTCTGAAGGCTCTGGAACAGTATGAGAAAGAGCGGC
AAGAACAGCTGCTGCGGGACACCCTGAACACCAACGAGTACCGGCTGAGCAAGAGAGGCCTTAGAGG
CTGGCGGGAAATCATCCAGAAATGGCTGAAAATGGACGAGAACGAGCCCTCCGAGAAGTACCTGGAA
GTGTTCAAGGACTACCAGCGGAAGCACCCTAGAGAGGCCGGCGATTACAGCGTGTACGAGTTCCTGT
CCAAGAAAGAGAACCACTTCATCTGGCGGAATCACCCTGAGTACCCCTACCTGTACGCCACCTTCTG
CGAGATCGACAAGAAAAAGAAGGACGCCAAGCAGCAGGCCACCTTCACACTGGCCGATCCTATCAAT
CACCCTCTGTGGGTCCGATTCGAGGAAAGAAGCGGCAGCAACCTGAACAAGTACAGAATCCTGACCG

AGCAGCTGCACACCGAGAAGCTGAAGAAAAAGCTGACAGTGCAGCTGGACCGGCTGATCTACCCTAC
AGAATCTGGCGGCTGGGAAGAGAAGGGCAAAGTGGACATTGTGCTGCTGCCCAGCCGGCAGTTCTAC
AACCAGATCTTCCTGGACATCGAGGAAAAGGGCAAGCACGCCTTCACCTACAAGGATGAGAGCATCA
AGTTCCCTCTGAAGGGCACACTCGGCGGAGCCAGAGTGCAGTTCGACAGAGATCACCTGAGAAGATA
CCCTCACAAGGTGGAAAGCGGCAACGTGGGCAGAATCTACTTCAACATGACCGTGAACATCGAGCCT
ACAGAGTCCCCAGTGTCCAAGTCTCTGAAGATCCACCGGGACGACTTCCCCAAGGTGGTCAACTTCA
AGCCCAAAGAACTGACCGAGTGGATCAAGGACAGCAAGGGCAAGAAACTGAAGTCCGGCATCGAGTC
CCTGGAAATCGGCCTGAGAGTGATGAGCATCGACCTGGGACAGAGACAGGCCGCTGCCGCCTCTATT
TTCGAGGTGGTGGATCAGAAGCCCGACATCGAAGGCAAGCTGTTTTTCCCAATCAAGGGCACCGAGC
TGTATGCCGTGCACAGAGCCAGCTTCAACATCAAGCTGCCCGGCGAGACACTGGTCAAGAGCAGAGA
AGTGCTGCGGAAGGCCAGAGAGGACAATCTGAAACTGATGAACCAGAAGCTCAACTTCCTGCGGAAC
GTGCTGCACTTCCAGCAGTTCGAGGACATCACCGAGAGAGAGAAGCGGGTCACCAAGTGGATCAGCA
GACAAGAGAACAGCGACGTGCCCCTGGTGTACCAGGATGAGCTGATCCAGATCCGCGAGCTGATGTA
CAAGCCTTACAAGGACTGGGTCGCCTTCCTGAAGCAGCTCCACAAGAGACTGGAAGTCGAGATCGGC
AAAGAAGTGAAGCACTGGCGGAAGTCCCTGAGCGACGGAAGAAAGGGCCTGTACGGCATCTCCCTGA
AGAACATCGACGAGATCGATCGGACCCGGAAGTTCCTGCTGAGATGGTCCCTGAGGCCTACCGAACC
TGGCGAAGTGCGTAGACTGGAACCCGGCCAGAGATTCGCCATCGACCAGCTGAATCACCTGAACGCC
CTGAAAGAAGATCGGCTGAAGAAGATGGCCAACACCATCATCATGCACGCCCTGGGCTACTGCTACG
ACGTGCGGAAGAAGAAATGGCAGGCTAAGAACCCCGCCTGCCAGATCATCCTGTTCGAGGATCTGAG
CAACTACAACCCCTACGAGGAAAGGTCCCGCTTCGAGAACAGCAAGCTCATGAAGTGGTCCAGACGC
[1804]
GAGATCCCCAGACAGGTTGCACTGCAGGGCGAGATCTATGGCCTGCAAGTGGGAGAAGTGGGCGCTC
AGTTCAGCAGCAGATTCCACGCCAAGACAGGCAGCCCTGGCATCAGATGTAGCGTCGTGACCAAAGA
GAAGCTGCAGGACAATCGGTTCTTCAAGAATCTGCAGAGAGAGGGCAGACTGACCCTGGACAAAATC
GCCGTGCTGAAAGAGGGCGATCTGTACCCAGACAAAGGCGGCGAGAAGTTCATCAGCCTGAGCAAGG
ATCGGAAGTGCGTGACCACACACGCCGACATCAACGCCGCTCAGAACCTGCAGAAGCGGTTCTGGAC
AAGAACCCACGGCTTCTACAAGGTGTACTGCAAGGCCTACCAGGTGGACGGCCAGACCGTGTACATC
CCTGAGAGCAAGGACCAGAAGCAGAAGATCATCGAAGAGTTCGGCGAGGGCTACTTCATTCTGAAGG
ACGGGGTGTACGAATGGGTCAACGCCGGCAAGggaggctctggaggaagcTCCGAAGTCGAGTTTTC
CCATGAGTACTGGATGAGACACGCATTGACTCTCGCAAAGAGGGCTCGAGATGAACGCGAGGTGCCC
GTGGGGGCAGTACTCGTGCTCAACAATCGCGTAATCGGCGAAGGTTGGAATAGGGCAATCGGACTCC
ACGACCCCACTGCACATGCGGAAATCATGGCCCTTCGACAGGGAGGGCTTGTGATGCAGAATTATCG
ACTTTATGATGCGACGCTGTACGTCACGTTTGAACCTTGCGTAATGTGCGCGGGAGCTATGATTCAC
TCCCGCATTGGACGAGTTGTATTCGGTGTTCGCAACGCCAAGACGGGTGCCGCAGGTTCACTGATGG
ACGTGCTGCATCATCCAGGCATGAACCACCGGGTAGAAATCACAGAAGGCATATTGGCGGACGAATG
TGCGGCGCTGTTGTGTCGTTTTTTTCGCATGCCCAGGCGGGTCTTTAACGCCCAGAAAAAAGCACAA
TCCTCTACTGACGGCTCTTCTGGATCTGAAACACCTGGCACAAGCGAGAGCGCCACCCCTGAGAGCT
CTGGCCTGAAAATCAAGAAGGGCAGCTCCAAGCAGAGCAGCAGCGAGCTGGTGGATAGCGACATCCT
GAAAGACAGCTTCGACCTGGCCTCCGAGCTGAAAGGCGAAAAGCTGATGCTGTACAGGGACCCCAGC
GGCAATGTGTTCCCCAGCGACAAATGGATGGCCGCTGGCGTGTTCTTCGGAAAGCTGGAACGCATCC
TGATCAGCAAGCTGACCAACCAGTACTCCATCAGCACCATCGAGGACGACAGCAGCAAGCAGTCTAT

GAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGGATCCTACCCATACGAT

[1805] GTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATG CCTAA

[1806] MAPKKKRKVGIHGVPAAATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYEHHE

[1807] QDPKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVEKKGEANQLSN

[1808] KFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEEKKKWEEDKKKDPLAKILGKLAEYGLIP

[1809] LFIPYTDSNEPIVKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWESWNLKVKEEYEKVEKEYKTLE

[1810] ERIKEDIQALKALEQYEKERQEQLLRDTLNTNEYRLSKRGLRGWREIIQKWLKMDENEPSEKYLEVF

[1811] KDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYATFCEIDKKKKDAKQQATFTLADPINHP

[1812] LWVRFEERSGSNLNKYRILTEQLHTEKLKKKLTVQLDRLIYPTESGGWEEKGKVDIVLLPSRQFYNQ

[1813] IFLDIEEKGKHAFTYKDESIKFPLKGTLGGARVQFDRDHLRRYPHKVESGNVGRIYFNMTVNIEPTE

[1814] SPVSKSLKIHRDDFPKVVNFKPKELTEWIKDSKGKKLKSGIESLEIGLRVMSIDLGQRQAAAASIFE

[1815] VVDQKPDIEGKLFFPIKGTELYAVHRASFNIKLPGETLVKSREVLRKAREDNLKLMNQKLNFLRNVL

[1816] HFQQFEDITEREKRVTKWISRQENSDVPLVYQDELIQIRELMYKPYKDWVAFLKQLHKRLEVEIGKE

[1817] VKHWRKSLSDGRKGLYGISLKNIDEIDRTRKFLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHLNALK

[1818] EDRLKKMANTIIMHALGYCYDVRKKKWQAKNPACQIILFEDLSNYNPYEERSRFENSKLMKWSRREI

[1819] PRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKLQDNRFFKNLQREGRLTLDKIAV

[1820] LKEGDLYPDKGGEKFISLSKDRKCVTTHADINAAQNLQKRFWTRTHGFYKVYCKAYQVDGQTVYIPE

[1821] SKDQKQKIIEEFGEGYFILKDGVYEWVNAGKGGSGGSSEVEFSHEYWMRHALTLAKRARDEREVPVG

[1822] AVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSR

[1823] IGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSS

[1824] TDGSSGSETPGTSESATPESSGLKIKKGSSKQSSSELVDSDILKDSFDLASELKGEKLMLYRDPSGN

[1825] VFPSDKWMAAGVFFGKLERILISKLTNQYSISTIEDDSSKQSMKRPAATKKAGQAKKKKGSYPYDVP

[1826] DYAYPYDVPDYAYPYDVPDYA

[1827] 对于以上的序列,Kozak序列用粗体和底线表示;标记N端核定位信号(NLS);小写字母表示GGGSGGS连接子;标记编码ABE8的序列,未修改的序列编码BhCas12b;双底线表示Xten20连接子;单底线表示C端NLS;GGATCC表示GS连接子;斜体字母代表3x血凝素(HA)标签的编码序列。

[1828] 向导多核苷酸

[1829] 在一个实施方案中,向导多核苷酸是向导RNA。RNA/Cas复合物可以帮助将Cas蛋白“向导”到目标DNA。Cas9/crRNA/tracrRNA核酸内切切割与间隔互补的线性或环状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“gNRA”)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.et al.,Science 337:816-821(2012),其全部内容通过引用方式并入本文。Cas9识别CRISPR重复序列(PAM或前间隔邻近基序)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,“Complete genome sequence of an M1 strain of Streptococcus pyogenes.”Ferretti,J.J.et al.,Natl.Acad.Sci.U.S.A.98:4658-4663(2001);“CRISPR RNA maturation by trans-

encoded small RNA and host factor RNase III."Deltcheva E.et al.,Nature 471:602-607(2011);以及"Programmabledual-RNA-guided DNA endonuclease in adaptive bacterial immunity."Jinek M.et al,Science 337:816-821(2012),各自的全部内容通过引用方式并入本文)。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,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The tracrRNA and Cas9 families of type IICRISPR-Cas immunity systems"(2013)RNA Biology 10:5,726-737;其全部内容通过引用方式并入本文。在一些实施方案中,Cas9核酸酶具有失活(inactive)(例如,失活(inactivated))的DNA切割结构域,即Cas9是切口酶。

[1830] 在一些实施方案中,向导多核苷酸是至少一种单个向导RNA("sgRNA"或"gNRA")。在一些实施方案中,向导多核苷酸是至少一种tracrRNA。在一些实施方案中,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)向导至目标核苷酸序列。

[1831] 本文公开的碱基编辑器的多核苷酸可编程的核苷酸结合结构域(例如,衍生自CRISPR的结构域)可以通过与向导多核苷酸关联来辨识目标多核苷酸序列。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,经由互补碱基配对)多核苷酸的目标序列,从而引导与向导核酸结合的碱基编辑器核酸到目标序列。向导多核苷酸可以是DNA。向导多核苷酸可以是RNA。在一些实施方案中,向导多核苷酸包括天然核苷酸(例如,腺苷)。在一些实施方案中,向导多核苷酸包括非天然(或非天然)核苷酸(例如,肽核酸或核苷酸类似物)。在一些实施方案中,向导核酸序列的靶向区域可以长度为是至少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至30个核苷酸之间,或长度在15至25个核苷酸之间,或长度在15至20个核苷酸之间。

[1832] 在一些实施方案中,向导多核苷酸包括两个或更多个单独的多核苷酸,其可以经由例如互补碱基配对(例如,双向导多核苷酸)彼此相互作用。例如,向导多核苷酸可包括CRISPR RNA(crRNA)和反式激活CRISPR RNA(tracrRNA)。例如,向导多核苷酸可包括一种或多种反式激活CRISPR RNA(tracrRNA)。

[1833] 在II类CRISPR系统中,CRISPR蛋白(例如,Cas9)靶向核酸通常需要包括辨识目标序列的序列的第一个RNA分子(crRNA)和包括识别重复序列的第二个RNA分子(trRNA)之间的互补碱基配对,其形成稳定向导RNA-CRISPR蛋白复合物的支架区域。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向导至目标多核苷酸序列。

[1834] 在一些实施方案中,本文所提供的碱基编辑器利用单个向导多核苷酸(例如,gRNA)。在一些实施方案中,本文所提供的碱基编辑器利用双向导多核苷酸(例如,双gRNA)。在一些实施方案中,本文所提供的碱基编辑器利用一种或多种向导多核苷酸(例如,多个gRNA)。在一些实施方案中,单个向导多核苷酸用于本文所述不同碱基编辑器。例如,单个向导多核苷酸可用于胞苷碱基编辑器和腺苷碱基编辑器。

[1835] 在其他实施方案中,向导多核苷酸可在单个分子(即,单分子向导核酸)中包括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

基编辑器向导至目标多核苷酸序列的任何单、双或多分子核酸。

[1836] 通常,向导多核苷酸(例如,crRNA/trRNA复合物或gRNA)包括包含能够辨识和结合目标多核苷酸序列的序列的“多核苷酸-靶向区段”,以及稳定在碱基编辑器的多核苷酸可编程的核苷酸结合结构域组分内的向导多核苷酸的“蛋白质结合区段”。在一些实施方案中,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。在其他实施方案中,向导多核苷酸的多核苷酸靶向区段辨识并结合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/段,使得区段可包括超过例如,当向导多核苷酸包括多个个别核酸分子时,蛋白质结合区段可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。在一些实施方案中,-包括两个单独分子的靶向DNA的RNA的蛋白质结合区段可包括(i)长度为100个碱基对的第一个RNA分子的40至75个碱基对;和(ii)长度为50个碱基对的第二个RNA分子的10至25个碱基对。除非在特定上下文中另有明确定义,否则“区段”的定义不限于总碱基对的特定数目,不限于给定的RNA分子的任何特定数目的碱基对,不限于复合物内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。

[1837] 向导RNA或向导多核苷酸可包括两个或更多个RNA,例如CRISPR RNA(crRNA)和反式激活crRNA(tracrRNA)。向导RNA或向导多核苷酸有时可包括单链RNA或通过crRNA和tracrRNA的一部分(例如,功能性部分)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括crRNA和tracrRNA的双RNA。此外,crRNA可以与目标DNA杂交。

[1838] 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可以是包括编码向导RNA的序列的载体。通过用包括编码向导RNA和启动子的序列的单离的向导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。

[1839] 可以单离向导RNA或向导多核苷酸。例如,向导RNA可以以单离的RNA的形式转染到细胞或生物体中。可以通过使用本领域已知的任何体外转录系统体外转录来制备向导RNA。向导RNA可以以单离的RNA的形式而不是以包括向导RNA的编码序列的质粒的形式转移到细胞中。

[1840] 向导RNA或向导多核苷酸可包括三个区域:5’端的可与染色体序列中的目标位点互补的第一区域、可形成茎环结构的第二内部区域以及’可以是单链的第三3区域。各向导RNA的第一区域也可以不同,使得每个向导RNA将融合蛋白向导至特定目标位点。此外,各向导RNA的第二和第三区域在所有向导RNA中可以相同。

[1841] 向导RNA或向导多核苷酸的第一区域可与染色体序列中的目标位点的序列互补,使得向导RNA的第一区域可与目标位点碱基配对。在一些实施方案中,向导RNA的第一区域可包括或从约10个核苷酸到25个核苷酸(即,从10个核苷酸到约25个核苷酸;或从约10个核苷酸到约25个核苷酸;或从约10个核苷酸到约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目标位点之间的碱基配对区域可以是或可以是长度为约10、11、12、13、14、15、16、17、18、19、20、22、23、24、25或更多个核苷酸。有时,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。

[1842] 向导RNA或向导多核苷酸还可包括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括茎(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。茎可包括一个或多个1至10个或约10个核苷酸的凸起物。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。

[1843] 向导RNA或向导多核苷酸还可以在3'端包括基本上可以是单链的第三区域。例如,第三个区域有时与感兴趣的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。

[1844] 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。在一些实施方案中,向导可以靶向基因的外显子1或2;在其他实施方案中,向导可以靶向基因的外显子3或4。组合物可包括均靶向相同外显子的多个向导RNA,或在一些实施方案中,可包括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。

[1845] 向导RNA或向导多核苷酸可以靶向20个核苷酸或约20个核苷酸的核酸序列。目标核酸可以少于或少于约20个核苷酸。目标核酸的长度可为至少或至少约5、10、15、16、17、18、19、20、21、22、23、24、25、30或1至100个核苷酸之间的任何位置。目标核酸可以是至多或至多约5、10、15、16、17、18、19、20、21、22、23、24、25、30、40、50或1至100之间的任何位置核苷酸长度。目标核酸序列可以是或可以是紧邻PAM的第一个核苷酸的5'处的约20个碱基。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至10、1至20、1至30、1至40、1至50、1至60、1至70、1至80、1至90或1至100个核苷酸。

[1846] 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前间隔。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括多核苷酸链,并且可以称为单个向导多核苷酸。向导多核苷酸可以包括两条多核苷酸链,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)辨识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。在一些实施方案中,质粒载体(例如,px333载体)可包括至少两个向导RNA编码的DNA序列。

[1847] 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,为了使核碱基编辑器系统中脱氨酶结构域(例如AID结构域)的潜在底物混杂的影响最小化,可能无意中使成为脱氨目标的残基数量(例如,可能潜在驻留在目标核酸基因座内的ssDNA上)可最小化。此外,软件工具可用于优化对应于目标核酸序列的gRNA,例如,使整个基因组的总脱靶活性最小化。例如,对于使用化脓链球菌Cas9的各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,其中包括多达特定数量(例如,1、2、3、4、5、6、7、8、9或10)个错配的碱基对。可以识别与目标位点互补的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;排名靠前的目标域代表那些可能具有最大的

目标和最少的脱靶活动的域。可以使用本领域已知的和/或如本文所载方法对候选靶向gRNA进行功能评估。

[1848] 作为非限制性实例,可使用DNA序列搜索算法来识别与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.,&Kim J.-S.Cas-OFFinder:A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases.Bioinformatics 30,1473-1475(2014)中所述。所述软件在计算其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次错配的匹配。一旦通过计算确定了脱靶位点,就会为每次向导计算总分,并使用网络界面汇总在表格输出中。除了识别与PAM序列相邻的潜在目标位点外,所述软件还可以识别与所选目标位点相差1、2、3或多于3个核苷酸的所有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搜索输入DNA序列的重复组件和低复杂性区域。输出是给定查询序列中存在的重复的详细注释。

[1849] 在识别之后,向导RNA的第一区域(例如crRNA)可以根据其与目标位点的距离、其正交性和5'核苷酸的存在进行分级,以便与相关的PAM序列(例如,基于含有相关PAM的人类基因组中的紧密匹配的识别的5'G,例如,化脓链球菌的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括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目标之外没有相同序列的20聚体靶向结构域,也没有在目标中含有一或两个错配的任何序列。可以选择具有良好正交性的靶向结构域,以使脱靶DNA切割最小化。

[1850] 在一些实施方案中,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。在一些实施方案中,报告系统可以包括基于报告基因的测定,其中碱基编辑活动导致报告基因的表达。例如,报告系统可包含包括去激活的起始密码子的报告基因,例如模板链上从3'-TAC-5'至3'-CAC-5'的突变。目标C成功脱氨后,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(green fluorescence protein,GFP)、红色荧光蛋白(red fluorescence protein,RFP)、荧光素酶、分泌的碱性磷酸酶(secreted alkaline phosphate,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的基因。报告系统可用于测试许多不同的gRNA,例如,以确定相应脱氨酶将针对目标DNA序列的哪个或哪些残基。还可以测试靶向非模板链的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。在一些实施方案中,此类gRNA可以被设计为使得突变的起始密码子将不与gRNA碱基配对。向导多核苷酸可包括标准核糖核苷酸、修饰的核糖核苷酸(例如假尿苷)、核糖核苷酸异构体、和/或核糖核苷酸类似物。在一些实施方案中,向导多核苷酸可包括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(Texas Red)、俄勒冈绿(Oregon green)、Alexa Fluors、Halo标签或合适的荧光染料)、检测标签(例如,生物素、地高辛等)、量子点或金粒子。

[1851] 向导多核苷酸可以化学合成、酶促合成或其组合。例如,可以使用标准基于亚磷酰胺的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。在其中向导RNA包括两个单独的分子(例如,crRNA和tracrRNA)的实施方案中,crRNA可以化学合成并且tracrRNA可以酶促合成。

[1852] 在一些实施方案中,碱基编辑器系统可包括多个向导多核苷酸,例如gRNA。例如,gRNA可以靶向包括在碱基编辑器系统中的一个或多个目标基因座(例如,至少1个gRNA、至少2个gRNA、至少5个gRNA、至少10个gRNA、至少20个gRNA、至少30个gRNA、至少50gRNA)。多个gRNA序列可以串联排列并且优选地由同向重复分开。

[1853] 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,载体可以包括额外的表达控制序列(例如,增强子序列、Kozak序列、聚腺苷酸化序列、转录终止序列等)、可选择的标记序列(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的DNA分子也可以是线性的。编码向导RNA或向导多核苷酸的DNA分子也可以是环状的。

[1854] 在一些实施方案中,碱基编辑器系统的一个或多个组分可由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的核苷酸结合结构域和向导RNA的DNA序列引入细胞,各DNA序列可以是单独分子的一部分(例如,一个含有多核苷酸可编程的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,一个含有多核苷酸可编程的核苷酸结合结构域和向导RNA的编码(和调节)序列的载体)。

[1855] 向导多核苷酸可以包括一个或多个修饰以提供具有新的或增强的特征的核酸。向导多核苷酸可包括核酸亲和性标签。向导多核苷酸可包括合成核苷酸、合成核苷酸类似物、核苷酸源物、和/或修饰的核苷酸。

[1856] 在一些实施方案中,gRNA或向导多核苷酸可包括修饰。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。在一些实施方案中,质量控制可包含PAGE、HPLC、MS或其任何组合。

[1857] 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。

[1858] 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'磷酸酯、3'硫代磷酸酯、5'磷酸酯、5'修饰硫代磷酸酯、Cis-Syn胸苷二聚体、三聚体、C12间隔、C3间隔、C6间隔、dSpacer、PC间隔、rSpacer、间隔18、间隔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、黑洞猝灭剂2、DABCYL SE、dT-DABCYL、IRDye QC-1、QSY-21、QSY-35、QSY-7、QSY-9、羧基连接子、巯基连接子、2'-脱氧核糖核苷类似物嘌呤、2'-脱氧核糖核苷类似物嘧啶、核糖核苷类似物、2'-O-甲基核糖核苷类似物、糖修饰的类似物、摆动/通用碱基、荧光染料标记、2'-氟RNA、2'-O-甲基RNA、甲基膦酸酯、磷酸二酯DNA、磷酸二酯RNA、硫代磷酸酯DNA、硫代磷酸酯RNA、UNA、假尿苷-5'-三磷酸、5'-甲基胞苷-5'-三磷酸或其任何组合。

[1859] 在一些实施方案中,修改是永久性的。在其他实施方案中,修改是暂时的。在一些

实施方案中,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。

[1860] 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。

[1861] 修饰也可以是硫代磷酸酯替代物。在一些实施方案中,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。在一些实施方案中,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。例如,可以在gRNA的5'-或"-端的最后3至5个核苷酸之间引入硫代磷酸酯(PS)键,其可以抑制外切核酸酶降解。在一些实施方案中,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。

[1862] 不同的Cas12b直向同源物(例如BhCas12b、BvCas12b和AaCas12b)使用不同的支架序列(也称为tracrRNA)。在一些实施方案中,支架序列被优化用于与BhCas12b蛋白一起使用,并且具有以下序列:(其中在实际gRNA中,T被尿苷(U)置换)。

[1863] BhCas12b sgRNA支架(底线)+20核苷酸至23核苷酸向导序列(用N表示)。

[1864] 5'GTTCTGTCTTTTGGTCAGGACAACCGTCTAGCTATAAGTGCTGCAGGGTGTGAGAAACTCCTATTGCTGGACGATGTCTCTTACGAGGCATTAGCACNNNNNNNNNNNNNNNNNNNN-3'

[1865] 在一些实施方案中,支架序列被优化以与BvCas12b蛋白一起使用,并且具有以下序列:(其中在实际gRNA中T被尿苷(U)替代)。

[1866] BvCas12b sgRNA支架(底线)+20核苷酸到23核苷酸向导序列(用N表示)

[1867] 5'GACCTATAGGGTCAATGAATCTGTGCGTGTGCCATAAGTAATTAAAAATTACCCACCACAGGAGCACCTGAAAACAGGTGCTTGGCACNNNNNNNNNNNNNNNNNNNN-3'

[1868] 在一些实施方案中,支架序列被优化用于与AaCas12b蛋白一起使用,并且具有以下序列:(其中在实际gRNA中,T被尿苷(U)置换)。

[1869] AaCas12b sgRNA支架(底线)+20核苷酸至23核苷酸向导序列(用N表示)

[1870] 5'GTCTAAAGGACAGAATTTTTCAACGGGTGTGCCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAACTTCTCAAAAAGAACGATCTGAGAAGTGGCACNNNNNNNNNNNNNNNNNNNN-3'

[1871] 因此,本领域技术人员可以改变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。

[1872] 前间隔序列邻近基序

[1873] 术语"前间隔邻近基序(PAM)"或PAM样基序是指紧邻在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至6个碱基对DNA序列。在一些实施方案中,PAM可以是5'PAM(即,位于前间隔5'端的上游)。在其他实施方式中,PAM可以是3'PAM(即,位于前间隔5'端的下游)。

[1874] PAM序列对于目标结合关重要,但确切的序列取决于Cas蛋白的类型。

[1875] 本文所提供的碱基编辑器可包括能够结合包括规范或非规范前间隔邻近基序(PAM)序列的核苷酸序列的衍生自CRISPR蛋白的结构域。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括具有不同PAM特异性的全部或部分CRISPR蛋白的碱基编辑器。

[1876] 例如,规范的Cas9蛋白(诸如来自化脓链球菌(spCas9)的Cas9)需要规范的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蛋白特异性的,并且在包括不同衍生自CRISPR源的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至6个核苷酸之间。以下的表1描述了几种PAM变体。

[1877] 表1.Cas9蛋白和相应的PAM序列

[1878]

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

[1879] 在一些实施方案中,PAM是NGC。在一些实施方案中,NGC PAM被Cas9变体识别。在一些实施方案中,NGC PAM变体包括一个或多个选自D1135M、S1136Q、G1218K、E1219F、A1322R、D1332A、R1335E和T1337R(统称为"MQKFRAER")的氨基酸取代。

[1880] 在一些实施方案中,PAM是NGT。在一些实施方案中,NGT PAM被Cas9变体识别。在一些实施方案中,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。在一些实施方案中,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。在一些实施方案中,NGT PAM变体通过在一个或多个残基1135、1136、1218、1219和1335产生。在一些实施方案中,NGT PAM变体选自下表2和表3提供的一组靶向突变。

[1881] 表2:在残基1219、1335、1337、1218处的NGT PAM变体突变

[1882]

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
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	

[1883] 表3:在残基1135、1136、1218、1219和1335处的NGT PAM变体突变

[1884]

变体	D1135L	S1136R	G1218S	E1219V	R1335Q
27	G				
28	V				
29	I				
30		A			
31		W			
32		H			
33		K			
34			K		
35			R		
36			Q		
37			T		
38			N		
39				I	

40				A	
41				N	
42				Q	
43				G	
44				L	
45				S	
46				T	
47					L
48					I
49					V
50					N
51					S
52					T
53					F
54					Y
55	N1286Q	I1331F			

[1885] 在一些实施方案中,NGT PAM变体选自表2和3中的变体5、7、28、31或36。在一些实施方案中,变体具有改进的NGT PAM识别。

[1886] 在一些实施方案中,NGT PAM变体在残基1219、1335、1337、和/或1218处具有突变。在一些实施方案中,从下表4中提供的变体中选择具有用于改进辨识的突变的NGT PAM变体。

[1887] 表4:在残基1219、1335、1337和1218处的NGT PAM变体突变

[1888]

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

[1889] 在一些实施方案中,可以如下表5提供的产生具有NGT PAM特异性的碱基编辑器。

[1890] 表5A.NGT PAM变体

[1891]

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

变体5	LRSVRQK	L	R	S	V	R	Q	K
变体6	LRSVRQL	L	R	S	V	R	Q	L

[1892] 在一些实施方案中,NGTN变体是变体1。在一些实施方案中,NGTN变体是变体2。在一些实施方案中,NGTN变体是变体3。在一些实施方案中,NGTN变体是变体4。在一些实施方案中,NGTN变体是变体5。在一些实施方案中,NGTN变体是变体6。

[1893] 在一些实施方案中,Cas9结构域是来自化脓链球菌(SpCas9)的Cas9结构域。在一些实施方案中,SpCas9结构域是核酸酶活性SpCas9、核酸酶失活的SpCas9(SpCas9d)或SpCas9切口酶(SpCas9n)。在一些实施方案中,SpCas9包括D10X突变,或本文所提供的任何氨基酸序列中的相应的突变,如SEQ ID NO:1中编号者,其中X是除D之外的任何氨基酸。在一些实施方案中,SpCas9包括在SEQ ID NO:1中编号为D10A突变,或本文所提供的任何氨基酸序列中的相应的突变。在一些实施方案中,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。在一些实施方案中,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。在一些实施方案中,SpCas9结构域包括在SEQ ID NO:1中编号为D1135X、R1335X和T1337X突变中的一个或多个,或本文所提供的任何氨基酸序列中的相应的突变,其中X是任何氨基酸。在一些实施方案中,SpCas9结构域包括在SEQ ID NO:1中编号为D1136E、R1335Q和T1337R突变中的一个或多个,或本文所提供的任何氨基酸序列中的相应的突变。在一些实施方案中,SpCas9结构域包括在SEQ ID NO:1中编号为D1135E、R1335Q和T1337R突变,或本文所提供的任何氨基酸序列中的相应的突变。在一些实施方案中,SpCas9结构域包括在SEQ ID NO:1中编号为D1135X、R1335X和T1337X突变中的一个或多个,或本文所提供的任何氨基酸序列中的相应的突变,其中X是任何氨基酸。在一些实施方案中,SpCas9结构域包括在SEQ ID NO:1中编号为D1135V、R1335Q和T1337R突变中的一个或多个,或本文所提供的任何氨基酸序列中的相应的突变。在一些实施方案中,SpCas9结构域包括在SEQ ID NO:1中编号为D1135V、R1335Q和T1337R突变,或本文所提供的任何氨基酸序列中的相应的突变。在一些实施方案中,SpCas9结构域包括在SEQ ID NO:1中编号为D1135X、G1218X、R1335X和T1337X突变中的一个或多个,或本文所提供的任何氨基酸序列中的相应的突变,其中X是任何氨基酸。在一些实施方案中,SpCas9结构域包括在SEQ ID NO:1中编号为D1135V、G1218R、R1335Q和T1337R突变中的一个或多个,或本文所提供的任何氨基酸序列中的相应的突变。在一些实施方案中,SpCas9结构域包括在SEQ ID NO:1中编号为D1135V、G1218R、R1335Q和T1337R突变,或本文所提供的任何氨基酸序列中的相应的突变。

[1894] 在一些实施方案中,Cas9是对改变的PAM序列具有特异性的Cas9变体。在一些实施方案中,额外的Cas9变体和PAM序列在Miller et al.,Continuous evolution of SpCas9 variants compatible with non-G PAMs.Nat Biotechnol(2020).https://doi.org/10.1038/s41587-020-0412-8中描述,其全部内容通过引用方式并入本文。在一些实施方案中,Cas9变体没有特定的PAM要求。在一些实施方案中,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。在一些实施方案中,SpCas9变体对PAM序列AAA、TAA、CAA、GAA、TAT、GAT或CAC具有特异性。在一些实施方案中,SpCas9变体包括在SEQ ID NO:1编号为位置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如或其

相应位置的氨基酸取代。在一些实施方案中,SpCas9变体包括在SEQ ID NO:1中编号为1114、1135、1218、1219、1221、1249、1320、1321、1323、1332、1333、1335或1337位置或其相应位置的氨基酸取代。在一些实方案中,SpCas9变体包括在SEQ ID NO:1编号为位置1114、1134、1135、1137、1139、1151、1180、1188、1211、1219、1221、1256、1264、1290、1318、1317、1320、1323、1333或其相应位置的氨基酸取代。在一些实施方案中,SpCas9变体包括在SEQ ID NO:1编号为位置1114、1131、1135、1150、1156、1180、1191、1218、1219、1221、1227、1249、1253、1286、1293、1320、1321、1332、1335、1339或其相应位置的氨基酸取代。在一些实施方案中,SpCas9变体包括在SEQ ID NO:1编号为位置1114、1127、1135、1180、1207、1219、1234、1286、1301、1332、1335、1337、1398、1349或其对应位置编的氨基酸取代。SpCas9变体的实施例性氨基酸取代和PAM特异性显示在下表5B、5C、5D和5E。

[1895] 表5B.额外的变异突变和PAM

[1896]

SpCas9 /PAM		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		
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

[1897] 表5C.额外的变异突变和PAM

[1898]

SpCas9 /PAM	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

[1899] 表5D.额外的变异突变和PAM

[1900]

SpCas9/ PAM	SpCas9 氨基酸位置																			
	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	

[1901] 表5E.额外的变异突变和PAM。

[1902]

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

[1903] 在一些实施方案中,Cas9是脑膜炎奈瑟菌Cas9(NmeCas9)或其变体。在一些实施方案中,NmeCas9对NNNNGAYW PAM具有特异性,其中Y是C或T并且W是A或T。在一些实施方案中,NmeCas9对NNNNGYTT PAM具有特异性,其中Y是C或T。在一些实施方案中,NmeCas9对NNNNGTCT PAM具有特异性。在一些实施方案中,NmeCas9是Nme1 Cas9。在一些实施方案中,NmeCas9对NNNNGATT 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具有特异性PAM、NNNNCCAG PAM、NNNNCCAT PAM或NNNGATT PAM。在一些实施方案中,Nme1Cas9对NNNNGATT PAM、NNNNCCTA PAM、NNNNCCTC PAM、NNNNCCTT PAM或NNNNCCTG PAM具有特异性。在一些实施方案中,NmeCas9对CAA PAM、CAAA PAM或CCA PAM具有特异性。在一些实施方案中,NmeCas9是Nme2 Cas9。在一些实施方案中,NmeCas9对NNNNCC(N4CC)PAM具有特异性,其中N是A、G、C或T中的任一个。在一些实施方案中,NmeCas9对NNNNCCGT PAM、NNNNCCGGPAM、NNNNCCCA PAM、NNNNCCCT PAM、NNNNCCCC PAM、NNNNCCAT PAM、NNNNCCAG PAM、NNNNCCAT PAM或NNNGATT PAM具有特异性。在一些实施方案中,NmeCas9是Nme3Cas9。在一些实施方案中,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 et al.Mol.Cell.(2019)73(4):714-726是通过引用方式整体并入本文。

[1904] 以下提供Nme1Cas9的实施例性氨基酸序列:

[1905] II类CRISPR RNA向导的核酸内切酶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
 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
```

[1906]

[1907] 以下提供Nme2Cas9的实施例性氨基酸序列：

[1908] II类CRISPR RNA向导的核酸内切酶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
 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
```

[1909]

[1910] 在一些实施方案中,本文所提供的任何融合蛋白的Cas9结构域包括的氨基酸序列为与本文所述Cas9多肽至少60％、至少65％、至少70％、至少75％、至少80％、至少85％、至少90％、至少95％、至少96％、至少97％、至少98％、至少99％或至少99.5％相同。在一些实施方案中,本文所提供的任何融合蛋白的Cas9结构域包括本文所述的任何Cas9多肽的氨基酸序列。在一些实施方案中,本文所提供的任何融合蛋白的Cas9结构域是由本文所述的任何Cas9多肽的氨基酸序列组成。

[1911] 在一些实施方案中,可将本文公开的碱基编辑器的衍生自CRISPR蛋白的结构域辨识的PAM提供给与编码碱基编辑器的插入物(例如,AAV插入物)单独的寡核苷酸上的细胞。在这样的实施方案中,在单独的寡核苷酸上提供PAM可以允许切割通常不能被切割的目标序列,因为在与目标序列相同的多核苷酸上不存在相邻的PAM。

[1912] 在一个实施方案中,化脓链球菌Cas9(SpCas9)可用作基因组工程的CRISPR核酸内切酶。然而,也可以使用其他者。在一些实施方案中,不同的核酸内切酶可用以靶向某些基

因组目标。在一些实施方案中,可以使用具有非NGG PAM序列的合成SpCas9源变体。此外,已经识别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)可导致携带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核酸内切酶能够在哺乳动物细胞体外和小鼠体内修饰的目标基因。在一些实施方案中,Cas蛋白可以靶向不同的PAM序列。在一些实施方案中,目标基因可以与例如Cas9 PAM、5’-NGG相邻。在其他实施方案中,其他Cas9直向同源物可具有不同的PAM要求。例如,其他PAM,诸如嗜热链球菌(CRISPR1为5’-NNAGAA及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。

[1913] 在一些实施方案中,对于化脓链球菌系统,目标基因序列可以在(即,5’至)5’-NGG PAM之前,并且20个核苷酸向导RNA序列可以与相反链碱基配对以介导与PAM相邻的Cas9裂解。在一些实施方案中,相邻切割可以是PAM上游的3个碱基对或可以是PAM上游的约3个碱基对。在一些实施方案中,相邻切割可以是或可以是PAM上游的约10个碱基对。在一些实施方案中,相邻切割可以是PAM上游的0至20个碱基对或可以是约PAM上游的0至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序列的实施例性SpCas9蛋白的序列如下:

[1914] 实施例性结合PAM的SpCas9的氨基酸序列如下:

[1915] MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRT

[1916] ARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPT

[1917] IYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPI

[1918] NASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSK

[1919] DTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLL

[1920] KALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ

[1921] RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS

[1922] EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRK

[1923] PAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKD

[1924] KDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI

[1925] RDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGI

[1926] LQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT

[1927] QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVP

[1928] SEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSR

[1929] MNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLE

[1930] SEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI

[1931] VWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTV

[1932] AYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFEL

[1933] ENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQI

[1934] SEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTK

[1935] EVLDATLIHQSITGLYETRIDLSQLGGD

[1936] 实施例性的结合PAM的SpCas9n的氨基酸序列如下:MDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD
SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
RYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1937] 实施例性的结合PAM的SpEQR Cas9的氨基酸序列如下:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRT
ARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPT
IYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPI
NASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSK
DTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLL
KALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS
EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRK
PAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKD
KDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
[1938] RDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGI
LQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT
QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVP
SEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSR
MNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLE
SEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI
VWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF**E**SPTV
AYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFEL
ENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQI
SEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK**Q**Y**R**STK
EVLDATLIHQSITGLYETRIDLSQLGGD

[1939] 在以上序列中,用底线和粗体表示可从D1134、R1334和T1336突变以产生SpEQR Cas9的残基E1134、Q1334和R1336。

[1940] 实施例性的结合PAM的SpVQR Cas9的氨基酸序列如下:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRT
ARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPT
IYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPI
NASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSK
DTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLL
KALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS
EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRK
PAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKD
KDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
[1941] RDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGI
LQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT
QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVP
SEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSR
MNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLE
SEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI
VWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF**V**SPTV
AYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFEL
ENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQI
SEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK**Q**Y**R**STK
EVLDATLIHQSITGLYETRIDLSQLGGD

[1942] 在以上序列中,用底线和粗体表示可以从D1134、R1334和T1336突变以产生SpVQR Cas9的残基V1134、Q1334和R1336。

[1943] 实施例性的结合PAM的SpVRER Cas9的氨基酸序列如下:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRT
ARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPT
IYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPI
NASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSK
DTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLL
KALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
[1944] RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS
EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRK
PAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKD
KDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
RDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGI
LQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT
QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVP

[1945] SEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSR
MNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLE
SEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI
VWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF**V**SPTV
AYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFEL
ENGRKRMLASA**R**ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQI
SEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK**E**Y**R**STK
EVLDATLIHQSITGLYETRIDLSQLGGD.

[1946] 在以上序列中,用底线和粗体表示可以从D1134、G1217、R1334和T1336突变以产生SpVRER Cas9的残基V1134、R1217、Q1334和R1336。

[1947] 在一些实施方案中,Cas9结构域是重组Cas9结构域。在一些实施方案中,重组Cas9结构域是SpyMacCas9结构域。在一些实施方案中,SpyMacCas9结构域是核酸酶活性的SpyMacCas9、核酸酶失活的SpyMacCas9(SpyMacCas9d)或SpyMacCas9切口酶(SpyMacCas9n)。在一些实施方案中,SaCas9结构域、SaCas9d结构域或SaCas9n结构域可结合具有非规范PAM的核酸序列。在一些实施方案中,SpyMacCas9结构域、SpCas9d结构域或SpCas9n结构域可以结合具有NAA PAM序列的核酸序列。

[1948] 具有天然5'-NAAN-3'PAM特异性的猕猴链球菌中Spy Cas9的实施例性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),并在以下提供。

[1949] SpyMacCas9

[1950] MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETAE

[1951] ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFG

[1952] NIVDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD

[1953] VDKLFIQLVQIYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLFGN

[1954] LIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[1955] LLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA

[1956] GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH

[1957] AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE

[1958] VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL

[1959] SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDLLKI

[1960] IKDKDFLDNEENEDILEDIVLTLTLFEDRGMIEERLKTYAHLFDDKVMKQLKRRRYTGWG

[1961] RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGHSL

[1962] HEQIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQTTQKGQKNSRERM

[1963] KRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHI

[1964] VPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT

[1965] KAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSK

[1966] LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKM

[1967] IAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFA

[1968] TVRKVLSMPQVNIVKKTEIQTVGQNGGLFDDNPKSPLEVTPSKLVPLKKELNPKKYGGYQ

[1969] KPTTAYPVLLITDTKQLIPISVMNKKQFEQNPVKFLRDRGYQQVGKNDFIKLPKYTLVDI

[1970] GDGIKRLWASSKEIHKGNQLVVSKKSQILLYHAHHLDSDLSNDYLQNHNQQFDVLFNEII

[1971] SFSKKCKLGKEHIQKIENVYSNKKNSASIEELAESFIKLLGFTQLGATSPFNFLGVKLNQKQYKGKKD YILPCTEGTLIRQSITGLYETRVDLSKIGED。

[1972] 在一些实施方案中,变体Cas9蛋白包括H840A、P475A、W476A、N477A、D1125A、W1126A及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,在一些实施方案中,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。在一些实施方案中,当变体Cas9蛋白包括W476A和W1126A突变或当变体Cas9蛋白包括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,所述方法不需要PAM序列。换言之,在一些实施方案中,当这种变体Cas9蛋白用于结合方法中时,所述方法可以包括向导RNA,但是所述方法可以在不存在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。

[1973] 在一些实施方案中,碱基编辑器的衍生自CRISPR蛋白的结构域可包括具有规范PAM序列(NGG)的Cas9蛋白的全部或部分。在其他实施方案中,碱基编辑器的Cas9源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本领域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.,et al.,“Engineered CRISPR-Cas9 nucleases with altered PAM specificities”Nature 523,481-485(2015);以及Kleinstiver,B.P.,et al.,“Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition”Nature Biotechnology 33,1293-1298(2015)中描述;其各自的全部内容特此通过引用方式并入。

[1974] 降低PAM排他性的Cas9域

[1975] 通常,Cas9蛋白,例如来自化脓链球菌(spCas9)的Cas9,需要规范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),G是鸟苷。这可能会限制编辑基因组内所需碱基的能力。在一些实施方案中,本文所提供的碱基编辑融合蛋白可能需要放置在精确位置,例如包括位于PAM上游的目标碱基的区域。参见例如Komor,A.C.,et al.,“Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage”Nature 533,420-424(2016),其全部内容通过引用方式并入本文。因此,在一些实施方案中,本文所提供的任何融合蛋白可含有能够结合不含有规范(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本领域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.,et al.,“Engineered CRISPR-Cas9nucleases with altered PAM specificities”Nature 523,481-485(2015);及Kleinstiver,B.P.,et al.,“Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by

modifying PAM recognition"Nature Biotechnology 33,1293-1298(2015)中描述;其各自的全部内容特此通过引用方式并入。

[1976] 高保真的Cas9结构域

[1977] 本公开的一些方面提供高保真的Cas9结构域。在一些实施方案中,高保真Cas9结构域是工程化的Cas9结构域,其包括与相应的野生型Cas9结构域相比减少Cas9结构域和DNA的糖-磷酸骨架之间的静电相互作用的一个或多个突变。不希望受任何特定理论的束缚,与DNA的糖-磷酸酯骨架的静电相互作用减少的高保真Cas9结构域可能具有较少的脱靶效应。在一些实施方案中,Cas9结构域(例如,野生型Cas9结构域)包括降低Cas9结构域与DNA的糖-磷酸酯骨架之间的关联的一种或多种突变。在一些实施方案中,Cas9结构域包括将Cas9结构域与DNA的糖-磷酸酯骨架之间的关联降低至少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%。

[1978] 在一些实施方案中,本文所提供的任何Cas9融合蛋白包括N497X、R661X、Q695X、和/或Q926X突变或本文所提供的任何氨基酸序列中的相应的突变中的一个或多个,其中X是任何氨基酸。在一些实施方案中,本文所提供的任何Cas9融合蛋白包括N497A、R661A、Q695A、和/或Q926A突变中的一个或多个,或本文所提供的任何氨基酸序列中的相应的突变。在一些实施方案中,Cas9结构域包括D10A突变,或本文所提供的任何氨基酸序列中的相应的突变。具有高保真度的Cas9域是本领域已知的并且对本领域技术人员来说是显而易见的。例如,Kleinstiver,B.P.,et al."High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects."Nature 529,490-495(2016);以及Slaymaker,I.M.,et al."Rationally engineered Cas9 nucleases with improved specificity."Science351,84-88(2015);;其各自的全部内容通过引用方式并入本文。

[1979] 在一些实施方案中,修饰的Cas9是高保真Cas9酶。在一些实施方案中,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)包括丙氨酸取代,其削弱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磷酸酯骨架相互作用的丙氨酸取代降低脱靶编辑。HypaCas9在REC3结构域中包括可增加Cas9校对和目标区分的突变(SpCas9N692A/M694A/Q695A/H698A)。与野生型Cas9相比,所有三种高保真酶产生的脱靶编辑更少。

[1980] 以下提供实施例性高保真度Cas9。

[1981] 相对于Cas9的高保真Cas9域突变以粗体和底线显示。

[1982]
DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEV
AYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQ
TYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKS
NFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS
ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEK
MDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTF
RIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTAFDKNLPNEKVLPK
HSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIEC
FDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKT
YAHLFDDKVMKQLKRRRYTGWGALSRKLINGIRDKQSGKTILDFLKSDGFANRNFMALIHDDSL
TFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARE
NQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI
NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQ
RKFDNLTKAERGGLSELDKAGFIKRQLVETRAITKHVAQILDSRMNTKYDENDKLIREVKVITL
KSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMI
AKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKV
LSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE
KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSK
RVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEV
LDATLIHQSITGLYETRIDLSQLGGD

[1983] 包括Cas9结构域和胞苷脱氨酶和/或腺苷脱氨酶的融合蛋白

[1984] 本公开的一些方面提供包括napDNAbp(例如,Cas9结构域)和一个或多个腺苷脱氨酶、胞苷脱氨酶结构域、和/或DNA糖基化酶结构域的融合蛋白。在一些实施方案中,融合蛋白包括Cas9结构域和腺苷脱氨酶结构域(例如,TadA*A)。应当理解,Cas9结构域可以是本文所提供的任何Cas9结构域或Cas9蛋白(例如,dCas9或nCas9)。在一些实施方案中,本文所提供的任何Cas9结构域或Cas9蛋白(例如,dCas9或nCas9)可以与本文所提供的任何胞苷脱氨酶和/或腺苷脱氨酶(例如,TadA*A)融合。例如但不限于,在一些实施方案中,融合蛋白包括以下结构:

[1985] NH_2-[胞苷脱氨酶]-[Cas9结构域]-[腺苷脱氨酶]-COOH;

[1986] NH_2-[腺苷脱氨酶]-[Cas9结构域]-[胞苷脱氨酶]-COOH;

[1987] NH_2-[腺苷脱氨酶]-[胞苷脱氨酶]-[Cas9结构域]-COOH;

[1988] NH_2-[胞苷脱氨酶]-[腺苷脱氨酶]-[Cas9结构域]-COOH;

[1989] NH_2-[Cas9结构域]-[腺苷脱氨酶]-[胞苷脱氨酶]-COOH;

[1990] NH_2-[Cas9结构域]-[胞苷脱氨酶]-[腺苷脱氨酶]-COOH;

[1991] NH_2-[腺苷脱氨酶]-[Cas9结构域]-COOH;

[1992] NH_2-[Cas9结构域]-[腺苷脱氨酶]-COOH;

[1993] NH_2-[胞苷脱氨酶]-[Cas9结构域]-COOH;或

[1994] NH_2-[Cas9结构域]-[胞苷脱氨酶]-COOH。

[1995] 在一些实施方案中,包括胞苷脱氨酶、无碱基编辑器和腺苷脱氨酶及napDNAbp(例如Cas9结构域)的融合蛋白不包括连接子序列。在一些实施方案中,在胞苷脱氨酶和/或腺

苷脱氨酶结构域与napDNAbp之间存在连接子。在一些实施方案中,以上通用架构中使用的“-”表示可任选连接子的存在。在一些实施方案中,胞苷脱氨酶和腺苷脱氨酶及napDNAbp经由本文所提供的任何连接子融合。例如,在一些实施方案中,胞苷脱氨酶和/或腺苷脱氨酶和napDNAbp经由本文所提供的任何连接子融合。

[1996] 包括核定位序列的融合蛋白(NLS)

[1997] 在一些实施方案中,本文所提供的融合蛋白进一步包括一个或多个(例如,2、3、4、5个)核靶向序列,例如核定位序列(NLS)。在一个实施方案中,使用二分NLS。在一些实施方案中,NLS包括促进包括NLS的蛋白质输入细胞核(例如,通过核转运)的氨基酸序列。在一些实施方案中,本文所提供的任何融合蛋白进一步包括核定位序列(NLS)。在一些实施方案中,NLS与融合蛋白的N端融合。在一些实施方案中,NLS与融合蛋白的C端融合。在一些实施方案中,NLS与Cas9结构域的N端融合。在一些实施方案中,NLS与nCas9结构域或dCas9结构域的C端融合。在一些实施方案中,NLS与脱氨酶的N端融合。在一些实施方案中,NLS与脱氨酶的C端融合。在一些实施方案中,NLS经由一个或多个连接子与融合蛋白融合。在一些实施方案中,NLS在没有连接子的情况下与融合蛋白融合。在一些实施方案中,NLS包括本文所提供或提及的NLS序列的任一个的氨基酸序列。额外的核定位序列是本领域已知的并且对本领域技术人员来说是显而易见的。例如,Plank等人,PCT/EP2000/011690中描述了NLS序列,其内容通过引用方式并入本文中,因为它们公开了实施例性核定位序列。在一些实施方案中,NLS包括氨基酸序列PKKKRKVEGADKRTADGSEFESPKKKRKV、KRTADGSEFESPKKKRKV、KRPAATKKAGQAKKKK、KKTELQTTNAENKTKKL、KRGINDRNFWRGENGRKTR、RKSGKIAAIVVKRPRKPKKKRKV或MDSLLMNRRKFLYQFKNVRWAKGRRETYLC。

[1998] 在一些实施方案中,NLS存在于连接子中或连接子是在NLS的侧边,例如本文所述连接子。在一些实施方案中,N端或C端NLS是二分NLS。二分NLS包括两个基本氨基酸簇,其是由相对短的间隔序列分隔(因此二分-2部分,而单部分NLS不是)。核质蛋白的NLS、KR[PAATKKAGQA]KKKK是无处不在的二分信号的原型:由大约10个氨基酸的间隔隔开的两个碱性氨基酸簇。实施例性二分NLS的序列如下:

[1999] PKKKRKVEGADKRTADGSEFESPKKKRKV

[2000] 在一些实施方案中,本发明的融合蛋白不包括连接子序列。在一些实施方案中,存在一个或多个结构域或蛋白质之间的连接子序列。在一些实施方案中,具有腺苷脱氨酶或胞苷脱氨酶和Cas9结构域的实施例性Cas9融合蛋白的一般结构包括以下结构中的任一种,其中NLS是核定位序列(例如,本文所提供的任何NLS),NH_2是融合蛋白的N端,COOH是融合蛋白的C端:

[2001] NH_2-NLS-[腺苷脱氨酶]-[Cas9结构域]-COOH;

[2002] NH_2-NLS[Cas9结构域]-[腺苷脱氨酶]-COOH;

[2003] NH_2-[腺苷脱氨酶]-[Cas9结构域]-NLS-COOH;

[2004] NH_2-[Cas9结构域]-[腺苷脱氨酶]-NLS-COOH;

[2005] NH_2-NLS-[胞苷脱氨酶]-[Cas9结构域]-COOH;

[2006] NH_2-NLS[Cas9结构域]-[胞苷脱氨酶]-COOH;

[2007] NH_2-[胞苷脱氨酶]-[Cas9结构域]-NLS-COOH;或

[2008] NH_2-[Cas9结构域]-[胞苷脱氨酶]-NLS-COOH。

[2009] 应当理解,本公开的融合蛋白可以包括一个或多个附加特征。例如,在一些实施方案中,融合蛋白可包括抑制剂、细胞质定位序列、输出序列(诸如核输出序列)或其他定位序列,以及可用于溶解、纯化或检测融合蛋白的序列标签。本文所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(biotin carboxylase carrier protein,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(maltose binding protein,MBP)标签、nus标签、谷胱甘肽-S-转移酶(GST)-标签、绿色荧光蛋白(GFP)标签、硫氧还蛋白标签、S-标签、Softag(例如,Softag 1、Softag 3)、链标签(strep-tag)、生物素连接酶标签、FlAsH标签、V5标签和SBP标签。额外合适的序列对本领域技术人员来说是显而易见的。在一些实施方案中,融合蛋白包括一个或多个His标签。

[2010] 可以使用编码包括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酶可包括位于或靠近氨端的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和在羧基端的一个或多个NLS)。当存在超过一个的NLS时,各NLS可以独立于其他选择,使得单个NLS可以存在于超过一个的副本中和/或与一个或多个其他NLS存在于一个或多个副本中。

[2011] 方法中使用的CRISPR酶可包括约6个NLS。当与NLS最接近的氨基酸在沿着自N端或C端算起的多肽链约50个氨基酸内,例如在1、2、3、4、5、10、15、20、25、30、40或50个氨基酸内时,NLS被认为接近N端或C端。

[2012] 核碱基编辑结构域

[2013] 本文描述包括融合蛋白的碱基编辑器,所述融合蛋白包括多核苷酸可编程的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)。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,碱基编辑器就锚定在要进行编辑的多核苷酸上,然后碱基编辑器的脱氨酶结构域组分可以编辑目标碱基。

[2014] 在一些实施方案中,核碱基编辑结构域包括脱氨酶结构域。如本文特别描述的,脱氨酶结构域包括胞嘧啶脱氨酶或腺苷脱氨酶。在一些实施方案中,术语“胞嘧啶脱氨酶”和“胞苷脱氨酶”可以互换使用。在一些实施方案中,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)中有描述,其各自通过引用方式整体并入本文。另参见Komor,A.C.,et al.,“Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage”Nature 533,420-424(2016);Gaudelli,N.M.,et al.,“Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage”Nature 551,464-471(2017);以及Komor,A.C.,et al.,“Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity”Science Advances 3:eaao4774(2017),其整体内容通过引用方式并入本文。

[2015] A至G编辑

[2016] 在一些实施方案中,本文所述碱基编辑器可包括包含腺苷脱氨酶的脱氨酶结构

域。碱基编辑器的这种腺苷脱氨酶结构域可以通过将A脱氨形成的具有G的碱基配对特性肌苷(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)。腺苷脱氨酶能够将脱氧核糖核酸(DNA)中的脱氧腺苷残基的腺嘌呤脱氨(即,除去胺基团)。

[2017] 在一些实施方案中,本文所提供的核碱基编辑器可以通过将一个或多个蛋白质结构域融合在一起来制备,从而产生融合蛋白。在某些实施方案中,本文所提供的融合蛋白包括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,本文所提供的融合蛋白可包括核酸酶活性降低的Cas9结构域。在一些实施方案中,本文所提供的融合蛋白可具有不具有核酸酶活性的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切割含有与靶向的A相对的T的非编辑(例如,非脱氨)链的活性。Cas9的催化残基(例如,D10至A10)的突变防止切割包括靶向的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),导致修复未编辑的链,最终导致非编辑链上的T到C变化。在一些实施方案中,A至G碱基编辑器进一步包括肌苷碱基切除修复抑制剂,例如尿嘧啶糖基化酶抑制剂(UGI)结构域或催化失活的肌苷特异性核酸酶。不希望受任何特定理论的束缚,UGI结构域或催化失活的肌苷特异性核酸酶可以抑制或阻止脱氨腺苷残基(例如肌苷)的碱基切除修复,其可以提高碱基编辑器的活性或效率。

[2018] 包括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及DNA-RNA杂交体。在某些实施方案中,包括腺苷脱氨酶的碱基编辑器可以使包括RNA的多核苷酸的靶向的A脱氨。例如,碱基编辑器可以包括能够使RNA多核苷酸和/或DNA-RNA杂合多核苷酸的靶向的A脱氨的腺苷脱氨酶结构域。在一个实施方案中,掺入碱基编辑器的腺苷脱氨酶包括作用于RNA(ADAR,例如ADAR1或ADAR2)的全部或部分腺苷脱氨酶。在另一个实施方案中,掺入碱基编辑器的腺苷脱氨酶包括作用于tRNA(ADAT)的全部或部分腺苷脱氨酶。包括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。在一个实施方案中,碱基编辑器的腺苷脱氨酶结构域包括ADAT的全部或一部分,所述ADAT包括一个或多个允许使其对DNA中的目标A脱氨的突变。例如,碱基编辑器可包括来自大肠杆菌的ADAT(EcTadA)的全部或部分,其包括一个或多个以下突变:D108N、A106V、D147Y、E155V、L84F、H123Y、I156F,或另一腺苷脱氨酶。

[2019] 腺苷脱氨酶可以衍生自任何合适的生物体(例如,大肠杆菌)。在一些实施方案中,腺嘌呤脱氨酶是天然存在的腺苷脱氨酶,其包括对应于本文所提供的任何突变(例如,ecTadA中的突变)的一个或多个突变。任何同源蛋白质中的相应残基可以通过例如序列比对和同源残基的测定来识别。可以相应地产生对应于本文所述的任何突变(例如,在ecTadA中识别的任何突变)的任何天然存在的腺苷脱氨酶(例如,与ecTadA具有同源性)中的突变。

[2020] 腺苷脱氨酶

[2021] 在一些实施方案中,本文所述碱基编辑器可包括脱氨酶结构域,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脱氨以形成显示G的碱基配对特性肌苷(I)来促进将腺嘌呤(A)核碱基编辑为鸟嘌呤(G)核碱基。腺苷脱氨酶能够将脱氧核糖核酸(DNA)中的脱氧腺苷残基的腺嘌呤脱氨(即,除去胺基团)。

[2022] 在一些实施方案中,本文所提供的腺苷脱氨酶能够使腺嘌呤脱氨。在一些实施方

案中,本文所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。在一些实施方案中,腺嘌呤脱氨酶是天然存在的腺苷脱氨酶,其包括对应于本文所提供的任何突变的一个或多个突变(例如,ecTadA中的突变)的突变。本领域技术人员将能够识别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任何突变(例如,在ecTadA中识别的任何突变)的突变。在一些实施方案中,腺苷脱氨酶来自原核生物。在一些实施方案中,腺苷脱氨酶来自细菌。在一些实施方案中,腺苷脱氨酶来自大肠杆菌、金黄色葡萄球菌、伤寒沙门氏菌、腐败希瓦氏菌、流感嗜血杆菌、新月柄杆菌或枯草芽孢杆菌。在一些实施方案中,腺苷脱氨酶来自大肠杆菌。

[2023] 本发明提供具有提高的效率(>50至60%)和特异性的腺苷脱氨酶变体。特别地,本文所述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意欲改变的碱基(即,“旁观者”)。

[2024] 在特定实施方案中,TadA是PCT/US2017/045381(WO 2018/027078)中描述的TadA中的任一个,其通过引用方式整体并入本文。

[2025] 在一些实施方案中,本发明的核碱基编辑器是包括以下序列改变的腺苷脱氨酶变体:

[2026] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD(也称为TadA*7.10)。

[2027] 在特定实施方案中,融合蛋白包括单个(例如,作为单体提供)TadA*8变体。在一些实施方案中,TadA*8与Cas9切口酶连接。在一些实施方案中,本发明的融合蛋白包括与TadA*8变体连接的野生型TadA(TadA(wt))的异源二聚体。在其他实施方案中,本发明的融合蛋白包括与TadA*8变体连接的TadA*7.10的异源二聚体。在一些实施方案中,碱基编辑器是包括TadA*8变体单体的ABE8。在一些实施方案中,碱基编辑器是包括TadA*8变体和TadA(wt)的异源二聚体的ABE8。在一些实施方案中,碱基编辑器是包括TadA*8变体和TadA*7.10的异源二聚体的ABE8。在一些实施方案中,碱基编辑器是包括TadA*8变体的异源二聚体的ABE8。在一些实施方案中,TadA*8变体是选自表7。在一些实施方案中,ABE8是选自表7。相关序列如下:

[2028] 野生型TadA(TadA(wt))或“TadA参考序列”

[2029] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTA HAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGS LMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD(SEQ ID NO:2)

[2030] TadA*7.10:

[2031] MSEVEFSHEYW MRHALTLAKR ARDEREVPVGAVLVLNNRVI GEGWNRAIGL HDPTAHAEIM ALRQGGLVMQ NYRLIDATLY VTFEPCVMCA GAMIHSRIGR VVFGVRNAKT GAAGSLMDVL HYPGMNHRVE ITEGILADEC AALLCYFFRM PRQVFNAQKK AQSSTD

[2032] 在一些实施方案中,腺苷脱氨酶包括至少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%与本文所提供的任何腺苷脱氨酶中所载的任一个氨基酸序列相同。应当理解,本文

所提供的腺苷脱氨酶可以包括一个或多个突变(例如,本文所提供的任何突变)。本公开提供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任何突变或其组合。在一些实施方案中,腺苷脱氨酶包括与参考序列或本文所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、21、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50或更多个突变。在一些实施方案中,与本领域已知或本文所述任一个氨基酸序列相比,腺苷脱氨酶包括具有至少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。

[2033] 在一些实施方案中,TadA脱氨酶是全长大肠杆菌TadA脱氨酶。例如,在某些实施方案中,腺苷脱氨酶包括氨基酸序列:

[2034] MRRAFITGVFFLSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD.

[2035] 然而,应当理解,可用于本申请的额外腺苷脱氨酶对于本领域技术人员是显而易见的并且在本公开的范畴内。例如,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,实施例性AD AT同源物的氨基酸序列包括以下者:

[2036] 金黄色葡萄球菌TadA:

[2037] MGSHMTNDIYFMTLAIEEAKKAAQLGEVPIGAIITKDDEVIARAHNLRETLQQPTAHAEHIAIERAAKVLGSWRLEGCTLYVTLEPCVMCAGTIVMSRIPRVVYGADDPKGGCSGSLMNLLQQSNFNHRAIVDKGVLKEACSTLLTTFFKNLRANKKSTN

[2038] 枯草芽孢杆菌TadA:

[2039] MTQDELYMKEAIKEAKKAEEKGEVPIGAVLVINGEIIARAHNLRETEQRSIAHAEMLVIDEACKALGTWRLEGATLYVTLEPCPMCAGAVVLSRVEKVVFGAFDPKGGCSGTLMNLLQEERFNHQAEVVSGVLEEECGGMLSAFFRELRKKKKAARKNLSE

[2040] 鼠伤寒沙门氏菌TadA:

[2041] MPPAFITGVTSLSDVELDHEYWMRHALTLAKRAWDEREVPVGAVLVHNHRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVLQNYRLLDTTLYVTLEPCVMCAGAMVHSRIGRVVFGARDAKTGAAGSLIDVLHHPGMNHRVEIIEGVLRDECATLLSDFFRMRRQEIKALKKADRAEGAGPAV

[2042] 腐败希瓦氏菌TadA:

[2043] MDEYWMQVAMQMAEKAEAAGEVPVGAVLVKDGQQIATGYNLSISQHDPTAHAEILCLRSAGKKLENYRLLDATLYITLEPCAMCAGAMVHSRIARVVYGARDEKTGAAGTVVNLLQHPAFNHQVEVTSGVLAEACSAQLSRFFKRRRDEKKALKLAQRAQQGIE

[2044] 流感嗜血杆菌F3031TadA:

[2045] MDAAKVRSEFDEKMMRYALELADKAEALGEIPVGAVLVDDARNIIGEGWNLSIVQSDPTAHAEIIALRNGAKNIQNYRLLNSTLYVTLEPCTMCAGAILHSRIKRLVFGASDYKTGAIGSRFHFFDDYKMNHTLEITSGVLAEECSQKLSTFFQKRREEKKIEKALLKSLSDK

[2046] 新月柄杆菌TadA:

[2047] MRTDESEDQDHRMMRLALDAARAAAEAGETPVGAVILDPSTGEVIATAGNGPIAAHDPTAHAEIAAMRA

AAAKLGNYRLTDLTLVVTLEPCAMCAGAISHARIGRVVFGADDPKGGAVVHGPKFFAQPTCHWRPEVTGGVLADESA DLLRGFFRARRKAKI

[2048] 硫还原地杆菌(Geobacter sulfurreducens(G.sulfurreducens))TadA:

[2049] MSSLKKTPIRDDAYWMGKAIREAAKAAARDEVPIGAVIVRDGAVIGRGHNLREGSNDPSAHAEMIAIRQ AARRSANWRLTGATLYVTLEPCLMCMGAIILARLERVVFGCYDPKGGAAGSLYDLSADPRLNHQVRLSPGVCQEECG TMLSDFFRDLRRRKKAKATPALFIDERKVPPEP

[2050] 大肠杆菌TadA(ecTadA)的一个实施方案包括以下者:

[2051] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC YFFRMPRQVFNAQKKAQSSTD

[2052] 在一些实施方案中,腺苷脱氨酶来自原核生物。在一些实施方案中,腺苷脱氨酶来自细菌。在一些实施方案中,腺苷脱氨酶来自大肠杆菌、金黄色葡萄球菌、伤寒沙门氏菌、腐败希瓦氏菌、流感嗜血杆菌、新月柄杆菌或枯草芽孢杆菌。在一些实施方案中,腺苷脱氨酶来自大肠杆菌。

[2053] 在一个实施方案中,本发明的融合蛋白包括与TadA7.10连接的野生型TadA,其与Cas9切口酶连接。在特定实施方案中,融合蛋白包括单个TadA7.10结构域(例如,作为单体提供)。在其他实施方案中,ABE7.10编辑器包括能够形成异源二聚体的TadA7.10和TadA(wt)。

[2054] 在一些实施方案中,腺苷脱氨酶包括本文所提供的任何腺苷脱氨酶中所载任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%与相同。应当理解,本文所提供的腺苷脱氨酶可以包括一个或多个突变(例如,本文所提供的任何突变)。本公开提供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任何突变或其组合。在一些实施方案中,腺苷脱氨酶包括与参考序列或本文所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、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有个或更多个突变。在一些实施方案中,腺苷脱氨酶包括与本领域已知或本文所述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。

[2055] 应当理解,本文所提供的任何突变(例如,基于TadA参考序列)可以引入其他腺苷脱氨酶,诸如大肠杆菌TadA(ecTadA)、金黄色葡萄球菌TadA(saTadA)或其他腺苷脱氨酶(例如,细菌腺苷脱氨酶)。对于本领域技术人员显而易见的是,可以类似地比对额外的脱氨酶以识别可以如本文所提供的突变的同源氨基酸残基。因此,可以在具有同源氨基酸残基的其他腺苷脱氨酶(例如,ecTadA)中进行在TadA参考序列中识别的任何突变。还应当理解,本文所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中进行。应当理解,TadA变体中的氨基酸取代是在TadA参考序列(SEQ ID NO:2)中编号,并且可以是在具有同源氨基酸残基的任何其他TadA变体中的相应的氨基酸取代或位置。应当理解,各别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;与TadA参考序列

共享同源性的TadA变体中的编号可能不同,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)识别任何同源蛋白质和各别编码核酸中的各别残基。

[2056] 在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的D108X突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变,其中X表示除了野生型腺苷脱氨酶中的相应的氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包括D108G、D108N、D108V、D108A或D108Y突变,或另一腺苷脱氨酶中的相应的突变。

[2057] 在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的A106X突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变,其中X表示除了野生型腺苷脱氨酶中的相应的氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包括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)中的相应的突变。

[2058] 在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的E155X突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变,其中X的存在表示除了野生型中的相应的氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包括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)中的相应的突变。

[2059] 在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的D147X突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变,其中X的存在表示除了野生型中的相应的氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的D147Y突变或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变。

[2060] 在一些实施方案中,腺苷脱氨酶包括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)中的相应的突变,其中X表示除了野生型腺苷脱氨酶中的相应的氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包括E155D、E155G或E155V突变。在一些实施方案中,腺苷脱氨酶包括D147Y。

[2061] 例如,腺苷脱氨酶可以在TadA参考序列中包括D108N、A106V、E155V、和/或D147Y突变,或在另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的以下群组的突变(群组的突变";"分隔),或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变: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D 147Y;和D108N、A106V、E155V及D147Y。然而,应当理解,本文所提供的相应的突变的任何组合可以在腺苷脱氨酶(例如,ecTadA)中进行。

[2062] 在一些实施方案中,腺苷脱氨酶包括以下TadA参考序列中的中的H8X、T17X、L18X、W23X、L34X、W45X、R51X、A56X、E59X、E85X、M94X、I95X、V102X、F104X、A106X、R107 X、D108X、K110X、M118X、N127X、A138X、F149X、M151X、R153X、Q154X、I156X和/或K157X突变中的一个或多个,或另一种腺苷脱氨酶(例如ecTadA)中的一个或多个相应的突变,其中X的存在表示除了野生型腺苷脱氨酶中的相应的氨基酸的任何氨基酸。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的H8Y、T17S、L18E、W23L、L34S、W45L、R51H、A56E或A56S、E59G、E85K或E85G、M94L、I95L、V102A、F104L、A106V、R107C或R107H、或R107P、D108G、或D108N、或D108V、或D108A、或D108Y、K110I、M118K、N127S、A138V、F149Y、M151V、R153C、Q1565D和/或K157R突变中的一个或多个、或另一种腺苷脱氨酶(例如,ecTadA)中的一个或多个相应的突变。

[2063] 在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的H8X、D108X、和/或N127X突变中的一个或多个,或另一种腺苷脱氨酶(例如,ecTadA)中的一个或多个相应的突变,其中X表示存在任何氨基酸。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的一个或多个H8Y、D108N、和/或N127S突变,或另一个腺苷脱氨酶(例如,ecTadA)中的一个或多个相应的突变。

[2064] 在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的H8X、R26X、M61X、L68X、M70X、A106X、D108X、A109X、N127X、D147X、R152X、Q154X、E155X、K161X、Q161X、Q161X、Q161X、Q161X/T中的一个或多个突变,或另一种腺苷脱氨酶(例如,ecTadA)中的一个或多个相应的突变,其中X表示除了存在除野生型腺苷脱氨酶中的相应的氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的H8Y、R26W、M61I、L68Q、M70V、A106T、D108N、A109T、N127S、D147Y、R152C、Q154H或Q154R、E155G、或E155V、或E155D、K161Q、Q163H和/或T166P突变中的一个或多个,或另一种腺苷脱氨酶(例如ecTadA)中的一个或多个相应的突变。

[2065] 在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的选自由H8X、D108X、N127X、D147X、R152X及Q154X组成的群组的一个、两个、三个、四个、五个或六个突变或在另一种腺苷脱氨酶(例如,ecTadA)中一个或多个相应的突变,其中X表示存在除了野生型腺苷脱氨酶中的相应的氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包括选自由TadA参考序列中的H8X、M61X、M70X、D108X、N127X、Q154X、E155X及Q163X所组成的群组的一个、两个、三个、四个、五个、六个、七个或八个突变是,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变或突变,其中X表示存在除了野生型腺苷脱氨酶中的相应的氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶TadA参考序列中包括一个、两个、三个、四个或五个选自由H8X、D108X、N127X、E155X和T166X所组成的群组的突变,或另一腺苷脱氨酶中的一个或多个相应的突变脱氨酶(例如,ecTadA),其中X表示存在除了野生型腺苷脱氨酶中的相应的氨基酸之外的任何氨基酸。

[2066] 在一些实施方案中,腺苷脱氨酶包括选自由H8X、A106X、D108X所组成的群组的一个、两个、三个、四个、五个或六个突变、另一种腺苷脱氨酶中的一个或多个突变,其中X表示除了野生型腺苷脱氨酶中的相应的氨基酸的任何氨基酸。在一些实施方案中,腺苷脱氨酶包括选自由H8X、R26X、L68X、D108X、N127X、D147X和E155X所组成的群组的一个、两个、三个、四个、五个、六个、七个或八个突变,或另一种腺苷脱氨酶中的一个或多个相应的突变,其中X表示存在除了野生型腺苷脱氨酶中相应的氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的选自由H8X、D108X、A109X、N127X和E155X所组成的群组的一个、两个、三个、四个或五个突变,或另一种腺苷脱氨酶中的一个或多个相应的突变脱氨酶(例如,ecTadA),其中X表示存在除了野生型腺苷脱氨酶中的相应的氨基酸之外的任何氨基酸。

[2067] 在一些实施方案中,腺苷脱氨酶包括TadA参考序列的选自由H8Y、D108N、N127S、D147Y、R152C和Q154H所组成的组中的一个、两个、三个、四个、五个或六个突变,或在另一种腺苷脱氨酶(例如,ecTadA)中的一个或多个相应的突变。在一些实施方案中,腺苷脱氨酶在TadA参考序列中包括选自由H8Y、M61I、M70V、D108N、N127S、Q154R、E155G和Q163H组成的群组的一个、两个、三个、四个、五个、六个、七个或八个突变,或另一种腺苷脱氨酶(例如,

ecTadA)中的一个或多个相应的突变。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的选自由H8Y、D108N、N127S、E155V和T166P所组成的群组的一个、两个、三个、四个或五个突变,或另一种脱氨酶腺苷(例如,ecTadA)中的一个或多个相应的突变。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的选自由H8Y、A106T、D108N、N127S、E155D和K161Q所组成的群组中的一个、两个、三个、四个、五个或六个突变,或在另一种腺苷脱氨酶(例如,ecTadA)中的相应的一个或多个突变。在一些实施方案中,腺苷脱氨酶在TadA参考序列中包括选自由H8Y、R26W、L68Q、D108N、N127S、D147Y和E155V所组成的群组中的一个、两个、三个、四个、五个、六个、七个或八个突变,或另一种腺苷脱氨酶(例如,ecTadA)中的一个或多个相应的突变。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的选自由H8Y、D108N、A109T、N127S和E155G所组成的群组中的一个、两个、三个、四个或五个突变,或另一种腺苷脱氨酶(例如,ecTadA)中的一个或多个相应的突变。

[2068] 本文所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。本文所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。

[2069] A至G核碱基编辑蛋白的详细信息在国际PCT申请号PCT/2017/045381(WO2018/027078)和Gaudelli,N.M.,et al.,"Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage"Nature,551,464-471(2017),其全部内容通过引用方式并入本文。

[2070] 在一些实施方案中,腺苷脱氨酶包括另一种腺苷脱氨酶(例如,ecTadA)中的一个或多个相应的突变。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的D108N、D108G或D108V突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的A106V和D108N突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的R107C和D108N突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的H8Y、D108N、N127S、D147Y和Q154H突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的H8Y、D108N、N127S、D147Y和E155V突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的D108N、D147Y和E155V突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的H8Y、D108N和N127S突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的A106V、D108N、D147Y和E155V突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变。

[2071] 在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的S2X、H8X、I49X、L84X、H123X、N127X、I156X和/或K160X突变中的一个或多个,或另一种腺苷脱氨酶中的一个或多个相应的突变,其中存在X表示除了野生型腺苷脱氨酶中相应的氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的S2A、H8Y、I49F、L84F、H123Y、N127S、I156F和/或K160S突变中的一个或多个,或另一种腺苷脱氨酶(例如,ecTadA)中的一个或多个相应的突变。

[2072] 在一些实施方案中,腺苷脱氨酶包括L84X突变腺苷脱氨酶,其中X表示除了野生型

腺苷脱氨酶中的相应的氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的L84F突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变。

[2073] 在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的H123X突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变,其中X表示除野生型腺苷脱氨酶中的相应的氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的H123Y突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变。

[2074] 在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的I156X突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变,其中X表示除野生型腺苷脱氨酶中的相应的氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的I156F突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变。

[2075] 在一些实施方案中,腺苷脱氨酶在TadA参考序列中包括选自由L84X、A106X、D108X、H123X、D147X、E155X和I156X组成的群组的一个、两个、三个、四个、五个、六个或七个突变,或另一种腺苷脱氨酶(例如ecTadA)中的一个或多个相应的突变,其中X表示存在除了野生型腺苷脱氨酶中的相应的氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶在TadA参考序列中包括选自由S2X、I49X、A106X、D108X、D147X和E155X所组成的群组的一个、两个、三个、四个、五个或六个突变,或在另一种腺苷脱氨酶(例如,ecTadA)中的一个或多个相应的突变,其中X表示存在除了野生型腺苷脱氨酶中的相应的氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的选自由H8X、A106X、D108X、N127X及K160X所组成的群组中的一个、两个、三个、四个或五个突变,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个相应的突变,其中X表示存在除野生型腺苷脱氨酶中的相应的氨基酸之外的任何氨基酸。

[2076] 在一些实施方案中,腺苷脱氨酶在TadA参考序列中包括选自由L84F、A106V、D108N、H123Y、D147Y、E155V及I156F所组成的群组中的一个、两个、三个、四个、五个、六个或七个突变,或另一种腺苷脱氨酶(例如,ecTadA)中的一个或多个相应的突变。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的选自由S2A、I49F、A106V、D108N、D147Y和E155V所组成的群组中的一个、两个、三个、四个、五个或六个突变。

[2077] 在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的选自由H8Y、A106T、D108N、N127S和K160S所组成的群组中的一个、两个、三个、四个或五个突变,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个相应的突变。

[2078] 在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的E25X、R26X、R107X、A142X、和/或A143X突变中的一个或多个,或另一种腺苷脱氨酶(例如,ecTadA)中的一个或多个相应的突变,其中存在X表示除了野生型腺苷脱氨酶中相应的氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的E25M、E25D、E25A、E25R、E25V、E25S、E25Y、R26G、R26N、R26Q、R26C、R26L、R26K、R107P、R107K、R107W、R107H、R107S、A142N、A142D、A142G、A143D、A143G、A143E、A143L、A143W、A143M、A143S、A143Q和/或A143R突变中的一个或多个,或另一种腺苷脱氨酶(例如,ecTadA)中的一个或多个相应的突变。在一些实施方案中,腺苷脱氨酶包括一个或多个本文所述相应于TadA参考序列的突变,或另一种腺苷脱氨酶(例如,ecTadA)中的一个或多个相应的突变。

[2079] 在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的E25X突变,或另一种腺苷

脱氨酶(例如,ecTadA)中的相应的突变,其中X表示除野生型腺苷脱氨酶中的相应的氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的E25M、E25D、E25A、E25R、E25V、E25S或E25Y突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变。

[2080] 在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的R26X突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变,其中X表示除了野生型腺苷脱氨酶中的相应的氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的R26G、R26N、R26Q、R26C、R26L或R26K突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变。

[2081] 在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的R107X突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变,其中X表示除了野生型腺苷脱氨酶中的相应的氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的R107P、R107K、R107A、R107N、R107W、R107H或R107S突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变。

[2082] 在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的A142X突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变,其中X表示除了野生型腺苷脱氨酶中的相应的氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的A142N、A142D、A142G突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变。

[2083] 在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的A143X突变,或另一腺苷脱氨酶(例如,ecTadA)中的相应的突变,其中X表示除了野生型腺苷脱氨酶中的相应的氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的A143D、A143G、A143E、A143L、A143W、A143M、A143S、A143Q和/或A143R突变,或另一种腺苷脱氨酶(例如ecTadA)中的相应的突变。

[2084] 在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的H36X、N37X、P48X、I49X、R51X、M70X、N72X、D77X、E134X、S146X、Q154X、K157X、和/或K161X的一个或多个,或另一种腺苷脱氨酶(例如,ecTadA)中的一个或多个相应的突变,其中X表示存在除了野生型腺苷脱氨酶中的相应的氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的H36L、N37T、N37S、P48T、P48L、I49V、R51H、R51L、M70L、N72S、D77G、E134G、S146R、S146C、Q154H、、K157H、和/或K161T突变中的一个或多个,或另一种腺苷脱氨酶(例如,ecTadA)中的一个或多个相应的突变。

[2085] 在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的H36X突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变,其中X表示除了野生型腺苷脱氨酶中的相应的氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的H36L突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变。

[2086] 在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的N37X突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变,其中X表示除了野生型腺苷脱氨酶中的相应的氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的N37T或N37S突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变。

[2087] 在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的P48X突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变,其中X表示除了野生型腺苷脱氨酶中的相应的氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的P48T或P48L

突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变。

[2088] 在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的R51X突变,或另一种腺苷脱氨酶中的相应的突变,其中X表示除了野生型腺苷脱氨酶中的相应的氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的R51H或R51L突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变。

[2089] 在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的S146X突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变,其中X表示除了野生型腺苷脱氨酶中的相应的氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的S146R或S146C突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变。

[2090] 在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的K157X突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变,其中X表示除了野生型腺苷脱氨酶中的相应的氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的K157N突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变。

[2091] 在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的P48X突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变,其中X表示除了野生型腺苷脱氨酶中的相应的氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的P48S、P48T或P48A突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变。

[2092] 在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的A142X突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变,其中X表示除了野生型腺苷脱氨酶中的相应的氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的A142N突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变。

[2093] 在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的W23X突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变,其中X表示除了野生型腺苷脱氨酶中的相应的氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的W23R或W23L突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变。

[2094] 在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的R152X突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变,其中X表示除了野生型腺苷脱氨酶中的相应的氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的R152P或R52H突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应的突变。

[2095] 在一个实施方案中,腺苷脱氨酶可包括突变H36L、R51L、L84F、A106V、D108N、H123Y、S146C、D147Y、E155V、I156F和K157N。在一些实施方案中,腺苷脱氨酶包括以下相对于TadA参考序列的突变组合,其中每个突变组合是由“_”分隔,并且每个突变组合在括号之间:

[2096] (A106V_D108N)、

[2097] (R107C_D108N)、

[2098] (H8Y_D108N_N127S_D147Y_Q154H)、

[2099] (H8Y_D108N_N127S_D147Y_E155V)、

[2100] (D108N_D147Y_E155V)、

[2101] (H8Y_D108N_N127S)、

[2102] (H8Y_D108N_N127S_D147Y_Q154H)、

[2103] (A106V_D108N_D147Y_E155V)，

[2104] (D108Q_D147Y_E155V)，

[2105] (D108M_D147Y_E155V)、

[2106] (D108L_D147Y_E155V)、

[2107] (D108K_D147Y_E155V)、

[2108] (D108I_D147Y_E155V)，

[2109] (D108F_D147Y_E155V)、

[2110] (A106V_D108N_D147Y)、

[2111] (A106V_D108M_D147Y_E155V)，

[2112] (E59A_A106V_D108N_D147Y_E155V)、

[2113] (E59A cat dead_A106V_D108N_D147Y_E155V)、

[2114] (L84F_A106V_D108N_H123Y_D147Y_E155V_I156Y)、

[2115] (L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)、

[2116] (R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F)，

[2117] (E25G_R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V

[2118] _I156F)、(E25D_R26G_L84F_A106V_R107K_D108N_H123Y_A142N_A143G_D147Y_E155V_

[2119] I156F)、

[2120] (R26Q_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F)、

[2121] (E25M_R26G_L84F_A106V_R107P_D108N_H123Y_A142N_A143D_D147Y_E155V

[2122] _I156F)、

[2123] (R26C_L84F_A106V_R107H_D108N_H123Y_A142N_D147Y_E155V_I156F)、(L84F_A106V_D108N_H123Y_A142N_A143L_D147Y_E155V_I156F)、

[2124] (R26G_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F)、

[2125] (E25A_R26G_L84F_A106V_R107N_D108N_H123Y_A142N_A143E_D147Y_E155V

[2126] _I156F)、

[2127] (R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F)、

[2128] (A106V_D108N_A142N_D147Y_E155V)、

[2129] (R26G_A106V_D108N_A142N_D147Y_E155V)、

[2130] (E25D_R26G_A106V_R107K_D108N_A142N_A143G_D147Y_E155V)、

[2131] (R26G_A106V_D108N_R107H_A142N_A143D_D147Y_E155V)、

[2132] (E25D_R26G_A106V_D108N_A142N_D147Y_E155V)、

[2133] (A106V_R107K_D108N_A142N_D147Y_E155V)、

[2134] (A106V_D108N_A142N_A143G_D147Y_E155V)、

[2135] (A106V_D108N_A142N_A143L_D147Y_E155V)、

[2136] (H36L_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N)、

[2137] (N37T_P48T_M70L_L84F_A106V_D108N_H123Y_D147Y_I49V_E155V_I156F)、

[2138] (N37S_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K161T)、

[2139] (H36L_L84F_A106V_D108N_H123Y_D147Y_Q154H_E155V_I156F)、

[2140] (N72S_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F)、

[2141] (H36L_P48L_L84F_A106V_D108N_H123Y_E134G_D147Y_E155V_I156F)、

[2142] (H36L_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K157N)、(H36L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F)、

[2143] (L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F_K161T)、

[2144] (N37S_R51H_D77G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)、

[2145] (R51L_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K157N)、

[2146] (D24G_Q71R_L84F_H96L_A106V_D108N_H123Y_D147Y_E155V_I156F_K160E)、

[2147] (H36L_G67V_L84F_A106V_D108N_H123Y_S146T_D147Y_E155V_I156F)、

[2148] (Q71L_L84F_A106V_D108N_H123Y_L137M_A143E_D147Y_E155V_I156F)、

[2149] (E25G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_Q159L)、

[2150] (L84F_A91T_F104I_A106V_D108N_H123Y_D147Y_E155V_I156F)、

[2151] (N72D_L84F_A106V_D108N_H123Y_G125A_D147Y_E155V_I156F)、

[2152] (P48S_L84F_S97C_A106V_D108N_H123Y_D147Y_E155V_I156F)、

[2153] (W23G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)、

[2154] (D24G_P48L_Q71R_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_Q159L),

[2155] (L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F)、

[2156] (H36L_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F

[2157] _K157N)、(N37S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_K161T),

[2158] (L84F_A106V_D108N_D147Y_E155V_I156F)、

[2159] (R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K161T),

[2160] (L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K161T)、

[2161] (L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K160E_K161T),

[2162] (L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K160E)、

[2163] (R74Q_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)、

[2164] (R74A_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)、

[2165] (L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)、

[2166] (R74Q_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)、

[2167] (L84F_R98Q_A106V_D108N_H123Y_D147Y_E155V_I156F)、

[2168] (L84F_A106V_D108N_H123Y_R129Q_D147Y_E155V_I156F)、

[2169] (P48S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F)、

[2170] (P48S_A142N)、

[2171] (P48T_I49V_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_L157N),

[2172] (P48T_I49V_A142N)、

[2173] (H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N)、

[2174] (H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S146C_A142N_D147Y_E155V_I156F

[2175] (H36L_P48T_I49V_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_

K157N)、

[2176] (H36L_P48T_I49V_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F_K157N)、

[2177] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N)、

[2178] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F_K157N)、

[2179] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_A142N_D147Y_E155V_I156F_K157N)、

[2180] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N)、

[2181] (W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N)、

[2182] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F_K161T)、

[2183] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152H_E155V_I156F_K157N)、

[2184] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F_K157N)、

[2185] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F_K157N)、

[2186] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142A_S146C_D147Y_E155V

[2187] _I156F_K157N)、

[2188] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142A_S146C_D147Y_R152P_E155V_I156F_K157N)、

[2189] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F_K161T)、

[2190] (W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F_K157N)、

[2191] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_R152P_E155V

[2192] _I156F_K157N)。

[2193] 在某些实施方案中,本文所提供的融合蛋白包括一种或多种改善融合蛋白的碱基编辑活性的特征。例如,本文所提供的任何融合蛋白可以包括具有降低的核酸酶活性的Cas9结构域。在一些实施方案中,本文所提供的任何融合蛋白可具有不具有核酸酶活性的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))。

[2194] 在一些实施方案中,腺苷脱氨酶是TadA*7.10。在一些实施方案中,TadA*7.10包括至少一种改变。在特定实施方案中,TadA*7.10包括对TadA*7.10的以下一个或多个改变或额外改变:Y147T、Y147R、Q154S、Y123H、V82S、T166R及Q154R。改变Y123H在本文中也称为

H123H(TadA*7.10中的改变H123Y回复成Y123H(wt))。在其他实施方案中,TadA*7.10包括选自由以下群组的改变的组合: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;及I76Y+V82S+Y123H+Y147R+Q154R。在特定实施方案中,腺苷脱氨酶变体包括从残基149、150、151、152、153、154、155、156和157开始的C端缺失。

[2195] 在一些实施方案中,TadA变体包括至少一个相对于TadA7.10的至少一个改变。在一些实施方案中,TadA变体包括至少一个相对于野生型TadA的至少一个改变。TadA变体中的氨基酸改变可以是本文所述相对于TadA7.10或野生型TadA的任何一种氨基酸取代。在一些实施方案中,TadA变体(例如TadA8)在氨基酸位置23、26、36、37、48、49、51、72、84、87、105、108、123、125、142、145、147、152、152、1 16、157、161或其任何组合包括氨基酸改变。在一些实施方案中,TadA变体包括相对于TadA7.10的氨基酸改变V82X,其中X是除了V之外的任何氨基酸。在一些实施方案中,TadA变体包括相对于TadA7.10的V82S改变。在一些实施方案中,氨基酸X是酸性氨基酸、碱性氨基酸或中性氨基酸。在一些实施方案中,TadA变体包括相对于TadA7.10的氨基酸改变T166X,其中X是除了T之外的任何氨基酸。在一些实施方案中,氨基酸X是酸性氨基酸、碱性氨基酸或中性氨基酸酸。在一些实施方案中,TadA变体包括相对于TadA7.10氨基酸改变V82X、Y147X、Q154X、I76X、Y123X、R23X、L36X、A48X、L51X、F84X、V106X、N108X、Y123X、C146X、Y147X、P152X、Q154X、V155X、F156X、N157X、T166X或其任何组合,其中X是除了TadA7.10中的氨基酸之外的任何氨基酸。在一些实施方案中,X是酸性氨基酸、碱性氨基酸或中性氨基酸。在一些实施方案中,X将TadA参考序列中的氨基酸回复成野生型氨基酸。

[2196] 在其他实施方案中,本发明的碱基编辑器是包括腺苷脱氨酶变体(例如TadA*8)的单体,所述腺苷脱氨酶变体包括与TadA*7.10或参考TadA序列相比一个或多个以下改变:Y147T、Y147R、Q154S、Y123H、V82S、T166R、和/或Q154R。在其他实施方案中,腺苷脱氨酶变体(TadA*8)是包括选自由以下群组的改变的单体: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;及I76Y+V82S+Y123H+Y147R+Q154R。在其他实施方案中,碱基编辑器是包括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,所述腺苷脱氨酶变体包括一个或多个以下改变Y147T、Y147R、Q154S、Y123H、V82S、T166R、和/或Q154R。在其他实施方案中,碱基编辑器是包括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,所述腺苷脱氨酶变体包括选自由下组的改变的组合: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;及I76Y+V82S+Y123H+Y147R+Q154R。

[2197] 在一个实施方案中,腺苷脱氨酶是TadA*8,其包括或基本上由以下具有腺苷脱氨酶活性的序列或其片段所组成:

[2198] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAE

[2199] IMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLM

[2200] DVLHYPGMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD

[2201] 在一些实施方案中,TadA*8是截短的。在一些实施方案中,截短的TadA*8相对于全长TadA*8缺少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、6、17、18、19或20个N端氨基酸残基。在一些实方案中,截短的TadA*8相对于全长TadA*8缺少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、6、17、18、19或20个C端氨基酸残基。在一些实施方案中,腺苷脱氨酶变体是全长TadA*8。

[2202] 在一些实施方案中,TadA*8是TadA*8.1、TadA*8.2、TadA*8.3、TadA*8.4、TadA*8.5、TadA*8.6、TadA*8.7、TadA*8.8、TadA*8.9、TadA*8.10、TadA*8.11、TadA*8.12、TadA*8.13、TadA*8.14、TadA*8.15、TadA*8.16、TadA*8.17、TadA*8.18、TadA*8.19、TadA*8.20、TadA*8.21、TadA*8.22、TadA*8.23、TadA*8.24。

[2203] 在一个实施方案中,本发明的融合蛋白包括与与Cas9切口酶连接的本文所述腺苷脱氨酶变体(例如,TadA*8)连接的野生型TadA。在特定实施方案中,融合蛋白包括单个TadA*8结构域(例如,作为单体提供)。在其他实施方案中,碱基编辑器包括能够形成异源二聚体的TadA*8和TadA(wt)。实施例性序列如下:

[2204] TadA(野生型):

[2205] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGG

[2206] LVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITE

[2207] GILADECAALLSDFFRMRRQEIKAQKKAQSSTD

[2208] TadA*7.10:

[2209] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG

[2210] LVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE

[2211] GILADECAALLCYFFRMPRQVFNAQKKAQSSTD

[2212] TadA*8:

[2213] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG

[2214] LVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD.

[2215] 在一些实施方案中,腺苷脱氨酶包括与本文所提供的任何腺苷脱氨酶中所载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,本文所提供的腺苷脱氨酶可以包括一个或多个突变(例如,本文所提供的任何突变)。本公开提供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任何突变或其组合。在一些实施方案中,腺苷脱氨酶包括具有与参考序列或本文所提供的任何腺苷脱氨酶相比1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、21、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50或更多个突变。在一些实施方案中,腺苷脱氨酶包括具有与本领域已知或本文所述任一氨基酸序列相比至少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。

[2216] 在特定实施方案中,TadA*8在以下以粗体显示的任何位置包括一个或多个突变。在其他实施方案中,TadA*8在以底线所示的任何位置包括一个或多个突变:

MSEVEFSHEY WMRHALTLAK RARDEREVPV GAVLVLNNRV IGEGWNRAIG ^{50} LHDPTAHAEI

[2217] MALRQGGLVM QNYRLIDATL YVTFEPCVMC AGAMIHSRIG ^{100} RVVFGVRNAK TGAAGSLMDV

LHYPGMNHRV EITEGILADE CAALLCYFFR ^{150} MPRQVFNAQK KAQSSTD

[2218] 例如,TadA*8包括单独的氨基酸位置82和/或166(例如V82S、T166R)或其与以下Y147T、Y147R、Q154S、Y123H、和/或Q154R中的任何一个或多个的组合的改变。在特定实施方案中,改变的组合是选自由以下所组成的群组: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;及I76Y+V82S+Y123H+Y147R+Q154R。

[2219] 在一些实施方案中,腺苷脱氨酶是TadA*8,其包括或基本上由以下具有腺苷脱氨酶活性的序列或其片段所组成:

[2220] MSEVEFSHEY WMRHALTLAK RARDEREVPV GAVLVLNNRV IGEGWNRAIG LHDPTAHAEI MALRQGGLVM QNYRLIDATL YVTFEPCVMC AGAMIHSRIG

[2221] RVVFGVRNAK TGAAGSLMDV LHYPGMNHRV EITEGILADE CAALLCTFFR MPRQVFNAQK KAQSSTD

[2222] 在一些实施方案中,TadA*8是截短的。在一些实施方案中,截短的TadA*8相对于全长TadA*8缺少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、6、17、18、19或20个N端氨基酸残基。在一些实施方案中,截短的TadA*8相对于全长TadA*8缺少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、6、17、18、19或,20个C端氨基酸残基。在一些实施方案中,腺苷脱氨酶变体是全长TadA*8。

[2223] 在一个实施方案中,本发明的融合蛋白包括与与Cas9切口酶连接的本文所述腺苷脱氨酶变体(例如,TadA*8)连接的野生型TadA。在特定实施方案中,融合蛋白包括单个TadA*8结构域(例如,作为单体提供)。在其他实施方式中,碱基编辑器包括能够形成异源二聚体的TadA*8和TadA(wt)。

[2224] 在一些实施方案中,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。在一些实施方案中,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。在一些实施方案中,与具有野生型TadA腺苷碱基编辑器相比,从合成文库产生的腺苷碱基编辑器表现出增加的碱基编辑效率、增加的碱基编辑特异性、减少的脱靶编辑、减少的旁观者编辑、减少的插入缺失形成、和/或减少的虚假编辑。在一些实施方案中,与具有TadA*7.10腺苷碱基编辑器相比,从合成文库产生的腺苷碱基编辑器表现出增加的碱基编辑效率、增加的碱基编辑特异性、减少的脱靶编辑、减少的旁观者编辑、减少的插入缺失形成、和/或减少的虚假编辑。在一些实施方案中,合成文库包括ABE的随机化TadA部分。在一些实施方案中,合成文库在TadA的每个位置包括所有20个标准氨基酸取代。在一些实施方案中,合成文库包括每个文库成员1至2个核苷酸取代突变的平均频率。在一些实施方案中,合成文库包括在TadA*7.10中发现

的背景突变。

[2225] 在一些实施方案中,本文所述碱基编辑系统包括具有插入Cas9中的TadA的ABE。提供具有插入Cas9的TadA的相关ABE的序列。

[2226] 101Cas9 TadAins 1015

[2227] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[2228] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[2229] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[2230] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[2231] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[2232] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[2233] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[2234] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[2235] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[2236] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[2237] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[2238] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[2239] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[2240] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[2241] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[2242] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[2243] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[2244] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[2245] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[2246] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[2247] YPKLESEFVYGDYKVGSSGSETPGTSESATPESSGSEVEFSHEYWMRHAL

[2248] TLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQG

[2249] GLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGS

[2250] LMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSST

[2251] DYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE

[2252] TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR

[2253] NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL

[2254] GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML

[2255] ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH

[2256] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[2257] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[2258] LGGD

[2259] 102 Cas9 TadAins 1022

[2260] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[2261] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[2262] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[2263] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[2264] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[2265] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[2266] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[2267] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[2268] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[2269] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[2270] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[2271] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[2272] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[2273] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[2274] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[2275] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[2276] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[2277] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[2278] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[2279] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[2280] YPKLESEFVYGDYKVYDVRKMIGSSGSETPGTSESATPESSGSEVEFSHE

[2281] YWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAE

[2282] IMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNA

[2283] KTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ

[2284] KKAQSSTDAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE

[2285] TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR

[2286] NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL

[2287] GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML

[2288] ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH

[2289] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[2290] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[2291] LGGD

[2292] 103 Cas9 TadAins 1029

[2293] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[2294] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[2295] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[2296] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[2297] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[2298] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[2299] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[2300] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[2301] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[2302] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[2303] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[2304] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[2305] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[2306] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[2307] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[2308] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[2309] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[2310] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[2311] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[2312] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[2313] YPKLESEFVYGDYKVYDVRKMIAKSEQEIGSSGSETPGTSESATPESSGS

[2314] EVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLH

[2315] DPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRV

[2316] VFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMP

[2317] RQVFNAQKKAQSSTDGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE

[2318] TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR

[2319] NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL

[2320] GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML

[2321] ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH

[2322] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[2323] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[2324] LGGD

[2325] 103 Cas9 TadAins 1040

[2326] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[2327] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[2328] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[2329] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[2330] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[2331] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[2332] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[2333] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[2334] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[2335] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[2336] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[2337] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[2338] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[2339] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[2340] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[2341] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[2342] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[2343] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[2344] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[2345] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[2346] YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSGSSGSETPGT

[2347] SESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVI

[2348] GEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCA

[2349] GAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADEC

[2350] AALLCYFFRMPRQVFNAQKKAQSSTDNIMNFFKTEITLANGEIRKRPLIE

[2351] TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR

[2352] NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL

[2353] GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML

[2354] ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH

[2355] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[2356] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[2357] LGGD

[2358] 105 Cas9 TadAins 1068

[2359] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[2360] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[2361] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[2362] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[2363] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[2364] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[2365] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[2366] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[2367] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[2368] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[2369] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[2370] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[2371] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[2372] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[2373] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[2374] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[2375] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[2376] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[2377] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[2378] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[2379] YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI

[2380] TLANGEIRKRPLIETNGEGSSGSETPGTSESATPESSGSEVEFSHEYWMR

[2381] HALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMAL

[2382] RQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGA

[2383] AGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQ

[2384] SSTDTGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR

[2385] NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL

[2386] GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML

[2387] ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH

[2388] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[2389] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[2390] LGGD

[2391] 106 Cas9 TadAins 1247

[2392] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[2393] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[2394] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[2395] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[2396] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[2397] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[2398] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[2399] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[2400] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[2401] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[2402] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[2403] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[2404] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[2405] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[2406] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[2407] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[2408] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[2409] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[2410] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[2411] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[2412] YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI

[2413] TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV

[2414] QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE

[2415] KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK

[2416] YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGGSS

[2417] GSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVL

[2418] VLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTF

[2419] EPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE

[2420] GILADECAALLCYFFRMPRQVFNAQKKAQSSTDSPEDNEQKQLFVEQHKH

[2421] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[2422] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[2423] LGGD

[2424] 107 Cas9 TadAins 1054

[2425] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[2426] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[2427] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[2428] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[2429] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[2430] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[2431] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[2432] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[2433] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[2434] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[2435] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[2436] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[2437] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[2438] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[2439] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[2440] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[2441] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[2442] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[2443] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[2444] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[2445] YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI

[2446] TLANGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDERE

[2447] VPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLID

[2448] ATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMN

[2449] HRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGEIRKRPLIE

[2450] TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR

[2451] NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL

[2452] GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML

[2453] ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH

[2454] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[2455] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[2456] LGGD

[2457] 108 Cas9 TadAins 1026

[2458] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[2459] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[2460] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[2461] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[2462] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[2463] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[2464] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[2465] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[2466] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[2467] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[2468] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[2469] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[2470] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[2471] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[2472] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[2473] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[2474] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[2475] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[2476] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[2477] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[2478] YPKLESEFVYGDYKVYDVRKMIAKSEGSSGSETPGTSESATPESSGSEVE

[2479] FSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPT

[2480] AHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFG

[2481] VRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQV

[2482] FNAQKKAQSSTDQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE

[2483] TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR

[2484] NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL

[2485] GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML

[2486] ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH

[2487] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[2488] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[2489] LGGD

[2490] 109 Cas9 TadAins 768

[2491] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[2492] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[2493] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[2494] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[2495] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[2496] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[2497] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[2498] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[2499] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[2500] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[2501] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[2502] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[2503] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[2504] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[2505] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[2506] MGRHKPENIVIEMARENQGSSGSETPGTSESATPESSGSEVEFSHEYWMR

[2507] HALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMAL

[2508] RQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGA

[2509] AGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRTTQKGQKNSR

[2510] ERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL

[2511] DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK

[2512] MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQIT

[2513] KHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREI

[2514] NNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQE

[2515] IGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGR

[2516] DFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP

[2517] KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNP

[2518] IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELAL

[2519] PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSK

[2520] RVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFD

[2521] TTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[2522] 110.1 Cas9 TadAins 1250

[2523] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF

IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAV
LVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAK
TGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPREDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILA
DANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI
DLSQLGGD

[2524] 110.2 Cas9 TadAins 1250

[2525] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPGSSGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPV
GAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVR
NAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPREDNEQKQLFVEQHKHYLDEIIEQISEFSKRV
ILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYE
TRIDLSQLGGD

[2526] 110.3 Cas9 TadAins 1250

[2527] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT

EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPGSSGSSGSETPGTSESATPESGSSSGSEVEFSHEYWMRHALTLAKRARDERE
VPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVF
GVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPREDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITG
LYETRIDLSQLGGD

[2528] 110.4 Cas9 TadAins 1250

[2529] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPGSSGSSGSETPGTSESATPESGSSSGSEVEFSHEYWMRHALTLAKRARDERE
VPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVF
GVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMRREDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITG
LYETRIDLSQLGGD

[2530] 110.5 Cas9 TadAins 1249

[2531] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST

DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSGSSGSSGSETPGTSESATPESGSSSGSEVEFSHEYWMRHALTLAKRARDEREV
PVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFG
VRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMRRPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITG
LYETRIDLSQLGGD

[2532] 110.5 Cas9 TadAins Δ59-66 1250

[2533] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[2534] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[2535] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[2536] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[2537] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[2538] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[2539] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[2540] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[2541] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[2542] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[2543] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[2544] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[2545] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[2546] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[2547] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[2548] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[2549] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[2550] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[2551] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[2552] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[2553] YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI

[2554] TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV

[2555] QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE

[2556] KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK

[2557] YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPG

[2558] SSGSSGSETPGTSESATPESGSSGSEVEFSHEYWMRHALTLAKRARDERE

[2559] VPVGAVLVLNNRVIGEGWNRAHAEIMALRQGGLVMQNYRLIDATLYVTFE

[2560] PCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEG

[2561] ILADECAALLCYFFRMPRQVFNAQKKAQSSTDEDNEQKQLFVEQHKHYLD

[2562] EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTN

[2563] LGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGG

[2564] D

[2565] 110.6 Cas9 TadAins 1251

[2566] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPEGSSGSSGSETPGTSESATPESGSSSGSEVEFSHEYWMRHALTLAKRARDER
EVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVV
FGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMRRDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITG
LYETRIDLSQLGGD

[2567] 110.7 Cas9 TadAins 1252

[2568] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE

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

[2569] 110.8Cas9 TadAins Δ59-66C-截短1250

[2570] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAV
LVLNNRVIGEGWNRAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLM
DVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKV
LSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGG

D

[2571] 111.1Cas9 TadAins 997

[2572] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[2573] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[2574] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[2575] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[2576] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[2577] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[2578] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[2579] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[2580] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[2581] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[2582] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[2583] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[2584] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[2585] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[2586] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[2587] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[2588] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[2589] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[2590] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[2591] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALSHE

[2592] YWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAE

[2593] IMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNA

[2594] KTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ

[2595] KKAQSSTDGSSGSETPGTSESATPESSGIKKYPKLESEFVYGDYKVYDVR

[2596] KMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGET

[2597] GEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKL

[2598] IARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIM

[2599] ERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGE

[2600] LQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI

[2601] IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLG

[2602] APAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[2603] 111.2 Cas9 TadAins 997

[2604] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[2605] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[2606] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[2607] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[2608] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[2609] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[2610] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[2611] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[2612] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[2613] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[2614] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[2615] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[2616] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[2617] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[2618] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[2619] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[2620] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[2621] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[2622] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[2623] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALSHE

[2624] YWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAE

[2625] IMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNA

[2626] KTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ

[2627] KKAQSSTDGSSGSSGSETPGTSESATPESSGGSSIKKYPKLESEFVYGDY

[2628] KVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLI

[2629] ETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPK

[2630] RNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKEL

[2631] LGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRM

[2632] LASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHK

[2633] HYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLF

[2634] TLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLS

[2635] QLGGD

[2636] 112ΔHNH TadA

[2637] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[2638] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[2639] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[2640] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[2641] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[2642] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[2643] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[2644] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[2645] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[2646] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[2647] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[2648] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[2649] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[2650] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[2651] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[2652] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSEVEFSHE

[2653] YWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAE

[2654] IMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNA

[2655] KTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ

[2656] KKAQSSTDGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEND

[2657] KLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAL

[2658] IKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFK

[2659] TEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKK

[2660] TEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA

[2661] KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIK

[2662] LPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKG

[2663] SPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH

[2664] RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATL

[2665] IHQSITGLYETRIDLSQLGGD

[2666] 113N-终止的单TadA螺旋截断165-端

[2667] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
YFFRMPRSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTD
RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYN
QLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDT
YDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILR
RQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKN
LPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSV
EISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRR
YTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNE
KLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQL
LNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLV
SDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS
NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSD
KLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDL
IIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIE
QISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLI

HQSITGLYETRIDLSQLGGD

[2668] 114N-终止的单TadA螺旋截短的165-端Δ59-65

[2669] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRTAHAEIMALRQGGLVMQNYRLI
DATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPR
SGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKN
LIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIV
DEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN
LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ
SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYP
FLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVL
PKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRL
SRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQ
TVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYL
QNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLIT
QRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDF
QFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFK
TEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKK
DWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKY
SLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSK
RVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGL
YETRIDLSQLGGD

[2670] 115.1 Cas9 TadAins1004

[2671] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[2672] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[2673] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[2674] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[2675] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[2676] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[2677] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[2678] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[2679] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[2680] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[2681] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[2682] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[2683] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[2684] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[2685] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[2686] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[2687] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[2688] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[2689] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[2690] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[2691] YPKGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREV

[2692] PVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDA

[2693] TLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNH

[2694] RVEITEGILADECAALLCYFFRMPRQLESEFVYGDYKVYDVRKMIAKSEQ

[2695] EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG

[2696] RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD

[2697] PKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKN

[2698] PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELA

[2699] LPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS

[2700] KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF

[2701] DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[2702] 115.2 Cas9 TadAins1005

[2703] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[2704] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[2705] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[2706] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[2707] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[2708] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[2709] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[2710] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[2711] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[2712] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[2713] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[2714] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[2715] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[2716] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[2717] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[2718] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[2719] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[2720] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[2721] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[2722] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[2723] YPKLGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDERE

[2724] VPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLID

[2725] ATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMN

[2726] HRVEITEGILADECAALLCYFFRMPRQESEFVYGDYKVYDVRKMIAKSEQ

[2727] EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG

[2728] RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD

[2729] PKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKN

[2730] PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELA

[2731] LPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS

[2732] KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF

[2733] DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[2734] 115.3 Cas9 TadAins1006

[2735] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[2736] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[2737] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[2738] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[2739] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[2740] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[2741] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[2742] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[2743] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[2744] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[2745] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[2746] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[2747] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[2748] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[2749] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[2750] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[2751] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[2752] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[2753] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[2754] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[2755] YPKLEGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDER

[2756] EVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLI

[2757] DATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGM

[2758] NHRVEITEGILADECAALLCYFFRMPRQSEFVYGDYKVYDVRKMIAKSEQ

[2759] EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG

[2760] RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD

[2761] PKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKN

[2762] PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELA

[2763] LPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS

[2764] KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF

[2765] DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[2766] 115.4 Cas9 TadAins1007

[2767] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[2768] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[2769] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[2770] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[2771] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[2772] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[2773] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[2774] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[2775] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[2776] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[2777] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[2778] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[2779] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[2780] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[2781] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[2782] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[2783] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[2784] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[2785] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[2786] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[2787] YPKLESGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDE

[2788] REVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRL

[2789] IDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPG

[2790] MNHRVEITEGILADECAALLCYFFRMPRQEFVYGDYKVYDVRKMIAKSEQ

[2791] EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG

[2792] RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD

[2793] PKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKN

[2794] PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELA

[2795] LPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS

[2796] KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF

[2797] DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[2798] 116.1Cas9 TadAins C端的截短2 792

[2799] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[2800] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[2801] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[2802] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[2803] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[2804] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[2805] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[2806] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[2807] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[2808] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[2809] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[2810] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[2811] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[2812] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[2813] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[2814] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGGSSGSETP

[2815] GTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNR

[2816] VIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVM

[2817] CAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILAD

[2818] ECAALLCYFFRMPRQSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQE

[2819] LDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVK

[2820] KMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI

[2821] TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVRE

[2822] INNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQ

[2823] EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG

[2824] RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD

[2825] PKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKN

[2826] PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELA

[2827] LPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS

[2828] KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF

[2829] DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[2830] 116.2Cas9 TadAins C端的截短2 791

[2831] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[2832] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[2833] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[2834] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[2835] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[2836] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[2837] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[2838] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[2839] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[2840] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[2841] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[2842] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[2843] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[2844] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[2845] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[2846] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSSGSETPG

[2847] TSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRV

[2848] IGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMC

[2849] AGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADE

[2850] CAALLCYFFRMPRQGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQE

[2851] LDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVK

[2852] KMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI

[2853] TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVRE

[2854] INNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQ

[2855] EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG

[2856] RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD

[2857] PKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKN

[2858] PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELA

[2859] LPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS

[2860] KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF

[2861] DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[2862] 116.3Cas9 TadAins C端的截短2 790

[2863] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[2864] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[2865] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[2866] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[2867] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[2868] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[2869] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[2870] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[2871] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[2872] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[2873] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[2874] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[2875] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[2876] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[2877] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[2878] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKEGSSGSETPGT

[2879] SESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVI

[2880] GEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCA

[2881] GAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADEC

[2882] AALLCYFFRMPRQLGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQE

[2883] LDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVK

[2884] KMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI

[2885] TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVRE

[2886] INNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQ

[2887] EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG

[2888] RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD

[2889] PKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKN

[2890] PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELA

[2891] LPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS

[2892] KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF

[2893] DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[2894] 117Cas9Δ1017-1069

[2895] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[2896] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[2897] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[2898] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[2899] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[2900] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[2901] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[2902] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[2903] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[2904] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[2905] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[2906] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[2907] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[2908] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[2909] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[2910] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[2911] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[2912] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[2913] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[2914] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[2915] YPKLESEFVYGDYKVYSSGSEVEFSHEYWMRHALTLAKRARDEREVPVGA

[2916] VLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYV

[2917] TFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEI

[2918] TEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGEIVWDKGRDFATVR

[2919] KVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF

[2920] DSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEA

[2921] KGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVN

[2922] FLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD

[2923] ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK

[2924] RYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[2925] 118 Cas9 TadA-CP116ins 1067

[2926] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[2927] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[2928] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[2929] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[2930] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[2931] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[2932] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[2933] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[2934] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[2935] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[2936] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[2937] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[2938] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[2939] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[2940] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[2941] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[2942] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[2943] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[2944] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[2945] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[2946] YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI

[2947] TLANGEIRKRPLIETNMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ

[2948] KKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRAR

[2949] DEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNY

[2950] RLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHY

[2951] PGGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR

[2952] NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL

[2953] GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML

[2954] ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH

[2955] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[2956] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[2957] LGGD

[2958] 119 Cas9 TadAins 701

[2959] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[2960] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[2961] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[2962] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[2963] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[2964] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[2965] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[2966] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[2967] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[2968] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[2969] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[2970] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[2971] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[2972] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[2973] SGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPV

[2974] GAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATL

[2975] YVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRV

[2976] EITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDLTFKEDIQKAQVS

[2977] GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMA

[2978] RENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLY

[2979] YLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNR

[2980] GKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKA

[2981] GFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVS

[2982] DFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYK

[2983] VYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE

[2984] TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR

[2985] NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL

[2986] GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML

[2987] ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH

[2988] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[2989] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[2990] LGGD

[2991] 120 Cas9 TadACP136ins 1248

[2992] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[2993] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[2994] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[2995] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[2996] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[2997] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[2998] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[2999] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[3000] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[3001] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[3002] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[3003] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[3004] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[3005] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[3006] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[3007] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[3008] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[3009] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[3010] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[3011] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[3012] YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI

[3013] TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV

[3014] QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE

[3015] KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK

[3016] YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSMN

[3017] HRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGT

[3018] SESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVI

[3019] GEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCA

[3020] GAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGPEDNEQKQLFVEQHKH

[3021] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[3022] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[3023] LGGD

[3024] 121 Cas9 TadACP136ins 1052

[3025] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[3026] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[3027] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[3028] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[3029] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[3030] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[3031] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[3032] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[3033] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[3034] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[3035] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[3036] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[3037] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[3038] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[3039] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[3040] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[3041] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[3042] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[3043] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[3044] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[3045] YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI

[3046] TLAMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGS

[3047] ETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVL

[3048] NNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEP

[3049] CVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGNGEIRKRPLIE

[3050] TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR

[3051] NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL

[3052] GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML

[3053] ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH

[3054] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[3055] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[3056] LGGD

[3057] 122 Cas9 TadACP136ins 1041

[3058] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[3059] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[3060] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[3061] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[3062] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[3063] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[3064] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[3065] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[3066] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[3067] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[3068] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[3069] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[3070] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[3071] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[3072] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[3073] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[3074] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[3075] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[3076] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[3077] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[3078] YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSMNHRVEITEG

[3079] ILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPES

[3080] SGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAI

[3081] GLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRI

[3082] GRVVFGVRNAKTGAAGSLMDVLHYPGNIMNFFKTEITLANGEIRKRPLIE

[3083] TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR

[3084] NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL

[3085] GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML

[3086] ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH

[3087] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[3088] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[3089] LGGD

[3090] 123 Cas9 TadACP139ins 1299

[3091] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[3092] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[3093] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[3094] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[3095] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[3096] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[3097] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[3098] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[3099] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[3100] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[3101] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[3102] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[3103] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[3104] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[3105] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[3106] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[3107] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[3108] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[3109] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[3110] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[3111] YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI

[3112] TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV

[3113] QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE

[3114] KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK

[3115] YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPE

[3116] DNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRMN

[3117] HRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGT

[3118] SESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVI

[3119] GEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCA

[3120] GAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGDKPIREQAENIIHLFT

[3121] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[3122] LGGD

[3123] 124 Cas9Δ792-872 TadAins

[3124] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[3125] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[3126] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[3127] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[3128] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[3129] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[3130] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[3131] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[3132] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[3133] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[3134] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[3135] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[3136] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[3137] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[3138] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[3139] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSEVEFSHE

[3140] YWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAE

[3141] IMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNA

[3142] KTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ

[3143] KKAQSSTDEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKA

[3144] GFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVS

[3145] DFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYK

[3146] VYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE

[3147] TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR

[3148] NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL

[3149] GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML

[3150] ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH

[3151] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[3152] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[3153] LGGD

[3154] 125 Cas9Δ792-906 TadAins

[3155] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[3156] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[3157] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[3158] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[3159] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[3160] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[3161] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[3162] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[3163] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[3164] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[3165] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[3166] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[3167] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[3168] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[3169] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[3170] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSEVEFSHE

[3171] YWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAE

[3172] IMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNA

[3173] KTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ

[3174] KKAQSSTDGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDK

[3175] LIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALI

[3176] KKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKT

[3177] EITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKT

[3178] EVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK

[3179] VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKL

[3180] PKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGS

[3181] PEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR

[3182] DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLI

[3183] HQSITGLYETRIDLSQLGGD

[3184] 126 TadA CP65ins 1003

[3185] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[3186] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[3187] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[3188] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[3189] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[3190] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[3191] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[3192] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[3193] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[3194] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[3195] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[3196] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[3197] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[3198] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[3199] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[3200] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[3201] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[3202] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[3203] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[3204] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[3205] YPKTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGR

[3206] VVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRM

[3207] PRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHA

[3208] LTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPLESEFVYGDYK

[3209] VYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE

[3210] TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR

[3211] NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL

[3212] GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML

[3213] ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH

[3214] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[3215] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[3216] LGGD

[3217] 127 TadA CP65ins 1016

[3218] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[3219] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[3220] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[3221] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[3222] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[3223] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[3224] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[3225] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[3226] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[3227] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[3228] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[3229] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[3230] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[3231] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[3232] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[3233] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[3234] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[3235] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[3236] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[3237] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[3238] YPKLESEFVYGDYKVTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVM

[3239] CAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILAD

[3240] ECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSE

[3241] VEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHD

[3242] PYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE

[3243] TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR

[3244] NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL

[3245] GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML

[3246] ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH

[3247] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[3248] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[3249] LGGD

[3250] 128 TadA CP65ins 1022

[3251] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[3252] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[3253] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[3254] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[3255] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[3256] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[3257] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[3258] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[3259] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[3260] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[3261] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[3262] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[3263] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[3264] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[3265] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[3266] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[3267] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[3268] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[3269] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[3270] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[3271] YPKLESEFVYGDYKVYDVRKMITAHAEIMALRQGGLVMQNYRLIDATLYV

[3272] TFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEI

[3273] TEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESAT

[3274] PESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWN

[3275] RAIGLHDPAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE

[3276] TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR

[3277] NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL

[3278] GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML

[3279] ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH

[3280] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[3281] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[3282] LGGD

[3283] 129 TadA CP65ins 1029

[3284] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[3285] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[3286] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[3287] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[3288] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[3289] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[3290] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[3291] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[3292] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[3293] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[3294] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[3295] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[3296] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[3297] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[3298] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[3299] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[3300] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[3301] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[3302] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[3303] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[3304] YPKLESEFVYGDYKVYDVRKMIAKSEQEITAHAEIMALRQGGLVMQNYRL

[3305] IDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPG

[3306] MNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETP

[3307] GTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNR

[3308] VIGEGWNRAIGLHDPGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE

[3309] TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR

[3310] NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL

[3311] GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML

[3312] ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH

[3313] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[3314] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[3315] LGGD

[3316] 130 TadA CP65ins 1041

[3317] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[3318] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[3319] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[3320] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[3321] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[3322] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[3323] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[3324] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[3325] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[3326] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[3327] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[3328] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[3329] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[3330] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[3331] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[3332] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[3333] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[3334] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[3335] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[3336] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[3337] YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSTAHAEIMALR

[3338] QGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAA

[3339] GSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQS

[3340] STDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREV

[3341] PVGAVLVLNNRVIGEGWNRAIGLHDPNIMNFFKTEITLANGEIRKRPLIE

[3342] TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR

[3343] NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL

[3344] GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML

[3345] ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH

[3346] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[3347] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[3348] LGGD

[3349] 131 TadA CP65ins 1054

[3350] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[3351] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[3352] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[3353] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[3354] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[3355] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[3356] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[3357] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[3358] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[3359] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[3360] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[3361] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[3362] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[3363] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[3364] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[3365] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[3366] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[3367] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[3368] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[3369] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[3370] YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI

[3371] TLANTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIG

[3372] RVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFR

[3373] MPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRH

[3374] ALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPGEIRKRPLIE

[3375] TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR

[3376] NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL

[3377] GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML

[3378] ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH

[3379] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[3380] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[3381] LGGD

[3382] 132 TadA CP65ins 1246

[3383] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[3384] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[3385] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[3386] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[3387] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[3388] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[3389] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[3390] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[3391] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[3392] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[3393] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[3394] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[3395] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[3396] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[3397] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[3398] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[3399] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[3400] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[3401] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[3402] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[3403] YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI

[3404] TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV

[3405] QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE

[3406] KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK

[3407] YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGTAH

[3408] AEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVR

[3409] NAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFN

[3410] AQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKR

[3411] ARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPSPEDNEQKQLFVEQHKH

[3412] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[3413] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[3414] LGGD

[3415] 在一些实施方案中,相应地,产生腺苷脱氨酶碱基编辑器以将TadA或其变体插入Cas9多肽中所识别的位置。

[3416] 在一些实施方案中,腺苷脱氨酶等位基因(例如TadA等位基因)的合成文库可用于产生具有修饰的碱基编辑效率和/或特异性的腺苷碱基编辑器。在一些实施方案中,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。在一些实施方案中,与腺苷碱基编辑器相比,从合成文库产生的具有野生型TadA的腺苷碱基编辑器表现出增加的的碱基编辑效率、增加的碱基编辑特异性、减少的脱靶编辑、减少的旁观者编辑、减少的插入缺失形成、和/或减少的虚假编辑。在一些实施方案中,与具有TadA*7.10的腺苷碱基编辑器相比,从合成文库产生的腺苷碱基编辑器表现出增加的碱基编辑效率、增加的碱基编辑特异性、减少的脱靶编辑、减少的旁观者编辑、减少的插入缺失形成、和/或减少的虚假编辑。在一些实施方案中,合成文库包括ABE的随机化TadA部分。在一些实施方案中,合成文库在TadA的每个位置包括所有20个标准氨基酸取代。在一些实施方案中,合成文库包括每个

文库成员1-2个核苷酸取代突变的平均频率。在一些实施方案中,合成文库包括在TadA*7.10中发现的背景突变。

[3417] C至T编辑

[3418] 在一些实施方案中,本文公开的碱基编辑器包括融合蛋白,所述融合蛋白包括能够使多核苷酸的目标胞苷(C)碱基脱氨以产生有胸腺嘧啶的碱基配对特性的尿苷(U)的胞苷脱氨酶。在一些实施方案中,例如,当多核苷酸是双链(例如,DNA)时,尿苷碱基经胸苷碱基(例如,通过细胞修复机制)取代以产生C:G至T:A转换。在其他实施方案中,碱基编辑器将核酸中的C脱氨为U无法伴随将U取代为T。

[3419] 多核苷酸中目标C的脱氨以产生U是可由本文所述碱基编辑器执行的碱基编辑类型的非限制性实例。在另一个实例中,包括胞苷脱氨酶结构域的碱基编辑器可以介导胞苷(C)碱基转换成鸟嘌呤(G)碱基。例如,可以通过碱基切除修复机制(例如,通过尿嘧啶DNA糖基化酶(UDG)结构域)从多核苷酸切除由碱基编辑器的胞苷脱氨酶结构域对胞苷脱氨而产生的多核苷酸的U,产生一个无碱基位点。然后与无碱基位点相对的核碱基可以被另一个碱基(例如C)取代(例如,通过碱基修复机制),例如通过跨损伤聚合酶。尽管与无碱基位点相对的核碱基通常被C取代,但其他取代(例如A、G或T)也可能发生。

[3420] 因此,在一些实施方案中,本文所述碱基编辑器包括能够将多核苷酸中的目标C脱氨成U的脱氨结构域(例如,胞苷脱氨酶结构域)。此外,如下所述,在一些实施方案中,碱基编辑器可包括促进由脱氨产生的U转换为T或G的额外结构域。例如,包括胞苷脱氨酶结构域的碱基编辑器可进一步包括尿嘧啶糖基化酶抑制剂(UGI)结构域以介导由T取代U,完成C至T碱基编辑事件。在另一个实例中,碱基编辑器可以掺入跨损伤聚合酶以提高C至G碱基编辑的效率,因为跨损伤聚合酶可以促进与无碱基位点相对的C的掺入(即,导致G在无碱基位点的掺入,并完成C至G基础编辑事件)。在一些实施方案中,附加结构域与脱氨酶结构域在内部融合。

[3421] 包括胞苷脱氨酶作为结构域的碱基编辑器可以使任何多核苷酸中的目标C脱氨,包括DNA、RNA和DNA-RNA杂交体。通常,胞苷脱氨酶催化位于多核苷酸单链部分上下文中的C核碱基。在一些实施方案中,包括目标C的完整多核苷酸可以是单链的。例如,掺入碱基编辑器的胞苷脱氨酶可以使单链RNA多核苷酸中的目标C脱氨。在其他实施方案中,包括胞苷脱氨酶结构域的碱基编辑器可作用于双链多核苷酸,但目标C可位于多核苷酸的在脱氨反应时处于单链状态的部分。例如,在其中napDNAbp结构域包括Cas12结构域的实施方案中,在形成Cas12-gRNA-目标DNA复合物的过程中可以留下几个核苷酸不配对,导致形成Cas12“R-环复合物”。这些未配对的核苷酸可形成单链DNA气泡,可作为单链特异性核苷酸脱氨酶(例如胞苷脱氨酶)的底物。

[3422] 在一些实施方案中,碱基编辑器的胞苷脱氨酶可包括载脂蛋白B mRNA编辑复合物(APOBEC)家族脱氨酶的全部或部分。APOBEC家族包括进化上保守的胞苷脱氨酶。所述家族的成员是C至U编辑酶。APOBEC样蛋白的N端结构域是催化结构域,而C端结构域是假催化结构域。更具体地,催化结构域是锌依赖性胞苷脱氨酶结构域并且对于胞苷脱氨很重要。APOBEC家族成员包括APOBEC1、APOBEC2、APOBEC3A、APOBEC3B、APOBEC3C、APOBEC3D(现称为“APOBEC3E”)、APOBEC3F、APOBEC3G、APOBEC3H、APOBEC4及激活诱导(胞苷)脱氨酶(AID)。许多包括修饰的胞苷脱氨酶的碱基编辑器是可商购的,包括但不限于SaBE3、SaKKH-BE3、VQR-

BE3、EQR-BE3、VRER-BE3、YE1-BE3、EE-BE3、YE2-BE3及YEE-BE3,其可从Addgene(质粒85169、85170、85171、85172、85173、85174、85175、85176、85177)获得。在一些实施方案中,掺入碱基编辑器的脱氨酶包括APOBEC1脱氨酶的全部或部分。在一些实施方案中,掺入碱基编辑器的脱氨酶包括APOBEC2脱氨酶的全部或部分。在一些实施方案中,掺入碱基编辑器的脱氨酶包括APOBEC3脱氨酶的全部或部分。在一些实施方案中,掺入碱基编辑器的脱氨酶包括APOBEC3A脱氨酶的全部或部分。在一些实施方案中,掺入碱基编辑器的脱氨酶包括APOBEC3B脱氨酶的全部或部分。在一些实施方案中,掺入碱基编辑器的脱氨酶包括APOBEC3C脱氨酶的全部或部分。在一些实施方案中,掺入碱基编辑器的脱氨酶包括APOBEC3D脱氨酶的全部或部分。在一些实施方案中,掺入碱基编辑器的脱氨酶包括APOBEC3E脱氨酶的全部或部分。在一些实施方案中,掺入碱基编辑器的脱氨酶包括APOBEC3F脱氨酶的全部或部分。在一些实施方案中,掺入碱基编辑器的脱氨酶包括APOBEC3G脱氨酶的全部或部分。在一些实施方案中,掺入碱基编辑器的脱氨酶包括APOBEC3H脱氨酶的全部或部分。在一些实施方案中,掺入碱基编辑器的脱氨酶包括APOBEC4脱氨酶的全部或部分。在一些实施方案中,掺入碱基编辑器的脱氨酶包括全部或部分激活诱导脱氨酶(AID)。

[3423] 在一些实施方案中,掺入碱基编辑器的脱氨酶包括全部或部分胞苷脱氨酶1(CDA1)。应将理解,碱基编辑器可以包括来自任何合适的生物体(例如,人类或大鼠)的脱氨酶。在一些实施方案中,碱基编辑器的脱氨酶结构域来自人类、黑猩猩、大猩猩、猴、牛、狗、大鼠或小鼠。在一些实施方案中,碱基编辑器的脱氨酶结构域衍生自大鼠(例如,大鼠APOBEC1)。在一些实施方案中,碱基编辑器的脱氨酶结构域是人类APOBEC1。在一些实施方案中,碱基编辑器的脱氨酶结构域是pmCDA1。

[3424] PmCDA1的碱基序列和氨基酸序列以及人类AID的CDS的碱基序列和氨基酸序列如下所示。

[3425] >tr|A5H718|A5H718_PETMA胞嘧啶脱氨酶OS=海七鳃鳗OX=7757PE=2SV=1

[3426] MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQS

[3427] GTERGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHT

[3428] LKIWACKLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWL

[3429] EKTLKRAEKRRSELSIMIQVKILHTTKSPAV

[3430] >EF094822.1海七鳃鳗单离物PmCDA.21胞嘧啶脱氨酶mRNA,完整

[3431] cds

[3432] TGACACGACACAGCCGTGTATATGAGGAAGGGTAGCTGGATGGGGGGGGGGGGAATACG

[3433] TTCAGAGAGGACATTAGCGAGCGTCTTGTTGGTGGCCTTGAGTCTAGACACCTGCAGAC

[3434] ATGACCGACGCTGAGTACGTGAGAATCCATGAGAAGTTGGACATCTACACGTTTAAGAA

[3435] ACAGTTTTTCAACAACAAAAAATCCGTGTCGCATAGATGCTACGTTCTCTTTGAATTAA

[3436] AACGACGGGGTGAACGTAGAGCGTGTTTTTGGGGCTATGCTGTGAATAAACCACAGAGC

[3437] GGGACAGAACGTGGAATTCACGCCGAAATCTTTAGCATTAGAAAAGTCGAAGAATACCT

[3438] GCGCGACAACCCCGGACAATTCACGATAAATTGGTACTCATCCTGGAGTCCTTGTGCAG

[3439] ATTGCGCTGAAAAGATCTTAGAATGGTATAACCAGGAGCTGCGGGGGAACGGCCACACT

[3440] TTGAAAATCTGGGCTTGCAAACTCTATTACGAGAAAAATGCGAGGAATCAAATTGGGCT

[3441] GTGGAACCTCAGAGATAACGGGGTTGGGTTGAATGTAATGGTAAGTGAACACTACCAAT

[3442] GTTGCAGGAAAATATTCATCCAATCGTCGCACAATCAATTGAATGAGAATAGATGGCTT

[3443] GAGAAGACTTTGAAGCGAGCTGAAAAACGACGGAGCGAGTTGTCCATTATGATTCAGGT

[3444] AAAAATACTCCACACCACTAAGAGTCCTGCTGTTTAAGAGGCTATGCGGATGGTTTTC

[3445] >tr|Q6QJ80|Q6QJ80_人类激活诱导的胞苷脱氨酶OS＝智人OX＝9606GN＝AICDA PE＝2SV＝1

[3446] MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVEL

[3447] LFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCED

[3448] RKAEPEGLRRLHRAGVQIAIMTFKAPV

[3449] >在12号染色体上的NG_011588.1:5001-15681智人激活诱导的胞苷脱氨酶(AICDA),RefSeqGene(LRG_17)

[3450] AGAGAACCATCATTAATTGAAGTGAGATTTTTCTGGCCTGAGACTTGCAGGGAGGCAAG

[3451] AAGACACTCTGGACACCACTATGGACAGGTAAAGAGGCAGTCTTCTCGTGGGTGATTGC

[3452] ACTGGCCTTCCTCTCAGAGCAAATCTGAGTAATGAGACTGGTAGCTATCCCTTTCTCTC

[3453] ATGTAACTGTCTGACTGATAAGATCAGCTTGATCAATATGCATATATATTTTTTGATCT

[3454] GTCTCCTTTTCTTCTATTCAGATCTTATACGCTGTCAGCCCAATTCTTTCTGTTTCAGA

[3455] CTTCTCTTGATTTCCCTCTTTTTCATGTGGCAAAAGAAGTAGTGCGTACAATGTACTGA

[3456] TTCGTCCTGAGATTTGTACCATGGTTGAAACTAATTTATGGTAATAATATTAACATAGC

[3457] AAATCTTTAGAGACTCAAATCATGAAAAGGTAATAGCAGTACTGTACTAAAAACGGTAG

[3458] TGCTAATTTTCGTAATAATTTTGTAAATATTCAACAGTAAAACAACTTGAAGACACACT

[3459] TTCCTAGGGAGGCGTTACTGAAATAATTTAGCTATAGTAAGAAAATTTGTAATTTTAGA

[3460] AATGCCAAGCATTCTAAATTAATTGCTTGAAAGTCACTATGATTGTGTCCATTATAAGG

[3461] AGACAAATTCATTCAAGCAAGTTATTTAATGTTAAAGGCCCAATTGTTAGGCAGTTAAT

[3462] GGCACTTTTACTATTAACTAATCTTTCCATTTGTTCAGACGTAGCTTAACTTACCTCTT

[3463] AGGTGTGAATTTGGTTAAGGTCCTCATAATGTCTTTATGTGCAGTTTTTGATAGGTTAT

[3464] TGTCATAGAACTTATTCTATTCCTACATTTATGATTACTATGGATGTATGAGAATAACA

[3465] CCTAATCCTTATACTTTACCTCAATTTAACTCCTTTATAAAGAACTTACATTACAGAAT

[3466] AAAGATTTTTTAAAAATATATTTTTTTGTAGAGACAGGGTCTTAGCCCAGCCGAGGCTG

[3467] GTCTCTAAGTCCTGGCCCAAGCGATCCTCCTGCCTGGGCCTCCTAAAGTGCTGGAATTA

[3468] TAGACATGAGCCATCACATCCAATATACAGAATAAAGATTTTTAATGGAGGATTTAATG

[3469] TTCTTCAGAAAATTTTCTTGAGGTCAGACAATGTCAAATGTCTCCTCAGTTTACACTGA

[3470] GATTTTGAAAACAAGTCTGAGCTATAGGTCCTTGTGAAGGGTCCATTGGAAATACTTGT

[3471] TCAAAGTAAAATGGAAAGCAAAGGTAAAATCAGCAGTTGAAATTCAGAGAAAGACAGAA

[3472] AAGGAGAAAAGATGAAATTCAACAGGACAGAAGGGAAATATATTATCATTAAGGAGGAC

[3473] AGTATCTGTAGAGCTCATTAGTGATGGCAAAATGACTTGGTCAGGATTATTTTTAACCC

[3474] GCTTGTTTCTGGTTTGCACGGCTGGGGATGCAGCTAGGGTTCTGCCTCAGGGAGCACAG

[3475] CTGTCCAGAGCAGCTGTCAGCCTGCAAGCCTGAAACACTCCCTCGGTAAAGTCCTTCCT

[3476] ACTCAGGACAGAAATGACGAGAACAGGGAGCTGGAAACAGGCCCCTAACCAGAGAAGGG

[3477] AAGTAATGGATCAACAAAGTTAACTAGCAGGTCAGGATCACGCAATTCATTTCACTCTG

[3478] ACTGGTAACATGTGACAGAAACAGTGTAGGCTTATTGTATTTTCATGTAGAGTAGGACC

[3479] CAAAAATCCACCCAAAGTCCTTTATCTATGCCACATCCTTCTTATCTATACTTCCAGGA

[3480] CACTTTTTCTTCCTTATGATAAGGCTCTCTCTCTCTCCACACACACACACACACACACA

[3481] CACACACACACACACACACACACAAACACACACCCCGCCAACCAAGGTGCATGTAAAAA

[3482] GATGTAGATTCCTCTGCCTTTCTCATCTACACAGCCCAGGAGGGTAAGTTAATATAAGA

[3483] GGGATTTATTGGTAAGAGATGATGCTTAATCTGTTTAACACTGGGCCTCAAAGAGAGAA

[3484] TTTCTTTTCTTCTGTACTTATTAAGCACCTATTATGTGTTGAGCTTATATATACAAAGG

[3485] GTTATTATATGCTAATATAGTAATAGTAATGGTGGTTGGTACTATGGTAATTACCATAA

[3486] AAATTATTATCCTTTTAAAATAAAGCTAATTATTATTGGATCTTTTTTAGTATTCATTT

[3487] TATGTTTTTTATGTTTTTGATTTTTTAAAAGACAATCTCACCCTGTTACCCAGGCTGGA

[3488] GTGCAGTGGTGCAATCATAGCTTTCTGCAGTCTTGAACTCCTGGGCTCAAGCAATCCTC

[3489] CTGCCTTGGCCTCCCAAAGTGTTGGGATACAGTCATGAGCCACTGCATCTGGCCTAGGA

[3490] TCCATTTAGATTAAAATATGCATTTTAAATTTTAAAATAATATGGCTAATTTTTACCTT

[3491] ATGTAATGTGTATACTGGCAATAAATCTAGTTTGCTGCCTAAAGTTTAAAGTGCTTTCC

[3492] AGTAAGCTTCATGTACGTGAGGGGAGACATTTAAAGTGAAACAGACAGCCAGGTGTGGT

[3493] GGCTCACGCCTGTAATCCCAGCACTCTGGGAGGCTGAGGTGGGTGGATCGCTTGAGCCC

[3494] TGGAGTTCAAGACCAGCCTGAGCAACATGGCAAAACGCTGTTTCTATAACAAAAATTAG

[3495] CCGGGCATGGTGGCATGTGCCTGTGGTCCCAGCTACTAGGGGGCTGAGGCAGGAGAATC

[3496] GTTGGAGCCCAGGAGGTCAAGGCTGCACTGAGCAGTGCTTGCGCCACTGCACTCCAGCC

[3497] TGGGTGACAGGACCAGACCTTGCCTCAAAAAAATAAGAAGAAAAATTAAAAATAAATGG

[3498] AAACAACTACAAAGAGCTGTTGTCCTAGATGAGCTACTTAGTTAGGCTGATATTTTGGT

[3499] ATTTAACTTTTAAAGTCAGGGTCTGTCACCTGCACTACATTATTAAAATATCAATTCTC

[3500] AATGTATATCCACACAAAGACTGGTACGTGAATGTTCATAGTACCTTTATTCACAAAAC

[3501] CCCAAAGTAGAGACTATCCAAATATCCATCAACAAGTGAACAAATAAACAAAATGTGCT

[3502] ATATCCATGCAATGGAATACCACCCTGCAGTACAAAGAAGCTACTTGGGGATGAATCCC

[3503] AAAGTCATGACGCTAAATGAAAGAGTCAGACATGAAGGAGGAGATAATGTATGCCATAC

[3504] GAAATTCTAGAAAATGAAAGTAACTTATAGTTACAGAAAGCAAATCAGGGCAGGCATAG

[3505] AGGCTCACACCTGTAATCCCAGCACTTTGAGAGGCCACGTGGGAAGATTGCTAGAACTC

[3506] AGGAGTTCAAGACCAGCCTGGGCAACACAGTGAAACTCCATTCTCCACAAAAATGGGAA

[3507] AAAAAGAAAGCAAATCAGTGGTTGTCCTGTGGGGAGGGGAAGGACTGCAAAGAGGGAAG

[3508] AAGCTCTGGTGGGGTGAGGGTGGTGATTCAGGTTCTGTATCCTGACTGTGGTAGCAGTT

[3509] TGGGGTGTTTACATCCAAAAATATTCGTAGAATTATGCATCTTAAATGGGTGGAGTTTA

[3510] CTGTATGTAAATTATACCTCAATGTAAGAAAAAATAATGTGTAAGAAAACTTTCAATTC

[3511] TCTTGCCAGCAAACGTTATTCAAATTCCTGAGCCCTTTACTTCGCAAATTCTCTGCACT

[3512] TCTGCCCCGTACCATTAGGTGACAGCACTAGCTCCACAAATTGGATAAATGCATTTCTG

[3513] GAAAAGACTAGGGACAAAATCCAGGCATCACTTGTGCTTTCATATCAACCATGCTGTAC

[3514] AGCTTGTGTTGCTGTCTGCAGCTGCAATGGGGACTCTTGATTTCTTTAAGGAAACTTGG

[3515] GTTACCAGAGTATTTCCACAAATGCTATTCAAATTAGTGCTTATGATATGCAAGACACT

[3516] GTGCTAGGAGCCAGAAAACAAAGAGGAGGAGAAATCAGTCATTATGTGGGAACAACATA

[3517] GCAAGATATTTAGATCATTTTGACTAGTTAAAAAAGCAGCAGAGTACAAAATCACACAT

[3518] GCAATCAGTATAATCCAAATCATGTAAATATGTGCCTGTAGAAAGACTAGAGGAATAAA

[3519] CACAAGAATCTTAACAGTCATTGTCATTAGACACTAAGTCTAATTATTATTATTAGACA

[3520] CTATGATATTTGAGATTTAAAAAATCTTTAATATTTTAAAATTTAGAGCTCTTCTATTT

[3521] TTCCATAGTATTCAAGTTTGACAATGATCAAGTATTACTCTTTCTTTTTTTTTTTTTT

[3522] TTTTTTTTTTGAGATGGAGTTTTGGTCTTGTTGCCCATGCTGGAGTGGAATGGCATGAC

[3523] CATAGCTCACTGCAACCTCCACCTCCTGGGTTCAAGCAAAGCTGTCGCCTCAGCCTCCC

[3524] GGGTAGATGGGATTACAGGCGCCCACCACCACACTCGGCTAATGTTTGTATTTTTAGTA

[3525] GAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCAGAGGATCC

[3526] ACCTGCCTCAGCCTCCCAAAGTGCTGGGATTACAGATGTAGGCCACTGCGCCCGGCCAA

[3527] GTATTGCTCTTATACATTAAAAAACAGGTGTGAGCCACTGCGCCCAGCCAGGTATTGCT

[3528] CTTATACATTAAAAAATAGGCCGGTGCAGTGGCTCACGCCTGTAATCCCAGCACTTTGG

[3529] GAAGCCAAGGCGGGCAGAACACCCGAGGTCAGGAGTCCAAGGCCAGCCTGGCCAAGATG

[3530] GTGAAACCCCGTCTCTATTAAAAATACAAACATTACCTGGGCATGATGGTGGGCGCCTG

[3531] TAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGGATCCGCGGAGCCTGGCAGATCTGCC

[3532] TGAGCCTGGGAGGTTGAGGCTACAGTAAGCCAAGATCATGCCAGTATACTTCAGCCTGG

[3533] GCGACAAAGTGAGACCGTAACAAAAAAAAAAAAATTTAAAAAAAGAAATTTAGATCAAG

[3534] ATCCAACTGTAAAAAGTGGCCTAAACACCACATTAAAGAGTTTGGAGTTTATTCTGCAG

[3535] GCAGAAGAGAACCATCAGGGGGTCTTCAGCATGGGAATGGCATGGTGCACCTGGTTTTT

[3536] GTGAGATCATGGTGGTGACAGTGTGGGGAATGTTATTTTGGAGGGACTGGAGGCAGACA

[3537] GACCGGTTAAAAGGCCAGCACAACAGATAAGGAGGAAGAAGATGAGGGCTTGGACCGAA

[3538] GCAGAGAAGAGCAAACAGGGAAGGTACAAATTCAAGAAATATTGGGGGGTTTGAATCAA

[3539] CACATTTAGATGATTAATTAAATATGAGGACTGAGGAATAAGAAATGAGTCAAGGATGG

[3540] TTCCAGGCTGCTAGGCTGCTTACCTGAGGTGGCAAAGTCGGGAGGAGTGGCAGTTTAGG

[3541] ACAGGGGGCAGTTGAGGAATATTGTTTTGATCATTTTGAGTTTGAGGTACAAGTTGGAC

[3542] ACTTAGGTAAAGACTGGAGGGGAAATCTGAATATACAATTATGGGACTGAGGAACAAGT

[3543] TTATTTTATTTTTTGTTTCGTTTTCTTGTTGAAGAACAAATTTAATTGTAATCCCAAGT

[3544] CATCAGCATCTAGAAGACAGTGGCAGGAGGTGACTGTCTTGTGGGTAAGGGTTTGGGGT

[3545] CCTTGATGAGTATCTCTCAATTGGCCTTAAATATAAGCAGGAAAAGGAGTTTATGATGG

[3546] ATTCCAGGCTCAGCAGGGCTCAGGAGGGCTCAGGCAGCCAGCAGAGGAAGTCAGAGCAT

[3547] CTTCTTTGGTTTAGCCCAAGTAATGACTTCCTTAAAAAGCTGAAGGAAAATCCAGAGTG

[3548] ACCAGATTATAAACTGTACTCTTGCATTTTCTCTCCCTCCTCTCACCCACAGCCTCTTG

[3549] ATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCCGCTGGGCTAAGGGTCGGCG

[3550] TGAGACCTACCTGTGCTACGTAGTGAAGAGGCGTGACAGTGCTACATCCTTTTCACTGG

[3551] ACTTTGGTTATCTTCGCAATAAGGTATCAATTAAAGTCGGCTTTGCAAGCAGTTTAATG

[3552] GTCAACTGTGAGTGCTTTTAGAGCCACCTGCTGATGGTATTACTTCCATCCTTTTTTGG

[3553] CATTTGTGTCTCTATCACATTCCTCAAATCCTTTTTTTTATTTCTTTTTCCATGTCCAT

[3554] GCACCCATATTAGACATGGCCCAAAATATGTGATTTAATTCCTCCCCAGTAATGCTGGG

[3555] CACCCTAATACCACTCCTTCCTTCAGTGCCAAGAACAACTGCTCCCAAACTGTTTACCA

[3556] GCTTTCCTCAGCATCTGAATTGCCTTTGAGATTAATTAAGCTAAAAGCATTTTTATATG

[3557] GGAGAATATTATCAGCTTGTCCAAGCAAAAATTTTAAATGTGAAAAACAAATTGTGTCT

[3558] TAAGCATTTTTGAAAATTAAGGAAGAAGAATTTGGGAAAAAATTAACGGTGGCTCAATT

[3559] CTGTCTTCCAAATGATTTCTTTTCCCTCCTACTCACATGGGTCGTAGGCCAGTGAATAC

[3560] ATTCAACATGGTGATCCCCAGAAAACTCAGAGAAGCCTCGGCTGATGATTAATTAAATT

[3561] GATCTTTCGGCTACCCGAGAGAATTACATTTCCAAGAGACTTCTTCACCAAAATCCAGA

[3562] TGGGTTTACATAAACTTCTGCCCACGGGTATCTCCTCTCTCCTAACACGCTGTGACGTC

[3563] TGGGCTTGGTGGAATCTCAGGGAAGCATCCGTGGGGTGGAAGGTCATCGTCTGGCTCGT

[3564] TGTTTGATGGTTATATTACCATGCAATTTTCTTTGCCTACATTTGTATTGAATACATCC

[3565] CAATCTCCTTCCTATTCGGTGACATGACACATTCTATTTCAGAAGGCTTTGATTTTATC

[3566] AAGCACTTTCATTTACTTCTCATGGCAGTGCCTATTACTTCTCTTACAATACCCATCTG

[3567] TCTGCTTTACCAAAATCTATTTCCCCTTTTCAGATCCTCCCAAATGGTCCTCATAAACT

[3568] GTCCTGCCTCCACCTAGTGGTCCAGGTATATTTCCACAATGTTACATCAACAGGCACTT

[3569] CTAGCCATTTTCCTTCTCAAAAGGTGCAAAAAGCAACTTCATAAACACAAATTAAATCT

[3570] TCGGTGAGGTAGTGTGATGCTGCTTCCTCCCAACTCAGCGCACTTCGTCTTCCTCATTC

[3571] CACAAAAACCCATAGCCTTCCTTCACTCTGCAGGACTAGTGCTGCCAAGGGTTCAGCTC

[3572] TACCTACTGGTGTGCTCTTTTGAGCAAGTTGCTTAGCCTCTCTGTAACACAAGGACAAT

[3573] AGCTGCAAGCATCCCCAAAGATCATTGCAGGAGACAATGACTAAGGCTACCAGAGCCGC

[3574] AATAAAAGTCAGTGAATTTTAGCGTGGTCCTCTCTGTCTCTCCAGAACGGCTGCCACGT

[3575] GGAATTGCTCTTCCTCCGCTACATCTCGGACTGGGACCTAGACCCTGGCCGCTGCTACC

[3576] GCGTCACCTGGTTCACCTCCTGGAGCCCCTGCTACGACTGTGCCCGACATGTGGCCGAC

[3577] TTTCTGCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCCTCTACTTCTG

[3578] TGAGGACCGCAAGGCTGAGCCCGAGGGGCTGCGGCGGCTGCACCGCGCCGGGGTGCAAA

[3579] TAGCCATCATGACCTTCAAAGGTGCGAAAGGGCCTTCCGCGCAGGCGCAGTGCAGCAGC

[3580] CCGCATTCGGGATTGCGATGCGGAATGAATGAGTTAGTGGGGAAGCTCGAGGGGAAGAA

[3581] GTGGGCGGGGATTCTGGTTCACCTCTGGAGCCGAAATTAAAGATTAGAAGCAGAGAAAA

[3582] GAGTGAATGGCTCAGAGACAAGGCCCCGAGGAAATGAGAAAATGGGGCCAGGGTTGCTT

[3583] CTTTCCCCTCGATTTGGAACCTGAACTGTCTTCTACCCCCATATCCCCGCCTTTTTTTC

[3584] CTTTTTTTTTTTTTGAAGATTATTTTTACTGCTGGAATACTTTTGTAGAAAACCACGAA

[3585] AGAACTTTCAAAGCCTGGGAAGGGCTGCATGAAAATTCAGTTCGTCTCTCCAGACAGCT

[3586] TCGGCGCATCCTTTTGGTAAGGGGCTTCCTCGCTTTTTAAATTTTCTTTCTTTCTCTAC

[3587] AGTCTTTTTTGGAGTTTCGTATATTTCTTATATTTTCTTATTGTTCAATCACTCTCAGT

[3588] TTTCATCTGATGAAAACTTTATTTCTCCTCCACATCAGCTTTTTCTTCTGCTGTTTCAC

[3589] CATTCAGAGCCCTCTGCTAAGGTTCCTTTTCCCTCCCTTTTCTTTCTTTTGTTGTTTCA

[3590] CATCTTTAAATTTCTGTCTCTCCCCAGGGTTGCGTTTCCTTCCTGGTCAGAATTCTTTT

[3591] CTCCTTTTTTTTTTTTTTTTTTTTTTTTTAAACAAACAAACAAAAAACCCAAAAAAA

[3592] CTCTTTCCCAATTTACTTTCTTCCAACATGTTACAAAGCCATCCACTCAGTTTAGAAGA

[3593] CTCTCCGGCCCCACCGACCCCCAACCTCGTTTTGAAGCCATTCACTCAATTTGCTTCTC

[3594] TCTTTCTCTACAGCCCCTGTATGAGGTTGATGACTTACGAGACGCATTTCGTACTTTGG

[3595] GACTTTGATAGCAACTTCCAGGAATGTCACACACGATGAAATATCTCTGCTGAAGACAG

[3596] TGGATAAAAAACAGTCCTTCAAGTCTTCTCTGTTTTTATTCTTCAACTCTCACTTTCTT

[3597] AGAGTTTACAGAAAAAATATTTATATACGACTCTTTAAAAAGATCTATGTCTTGAAAAT

[3598] AGAGAAGGAACACAGGTCTGGCCAGGGACGTGCTGCAATTGGTGCAGTTTTGAATGCAA

[3599] CATTGTCCCCTACTGGGAATAACAGAACTGCAGGACCTGGGAGCATCCTAAAGTGTCAA

[3600] CGTTTTTCTATGACTTTTAGGTAGGATGAGAGCAGAAGGTAGATCCTAAAAAGCATGGT

[3601] GAGAGGATCAAATGTTTTTATATCAACATCCTTTATTATTTGATTCATTTGAGTTAACA

[3602] GTGGTGTTAGTGATAGATTTTTCTATTCTTTTCCCTTGACGTTTACTTTCAAGTAACAC

[3603] AAACTCTTCCATCAGGCCATGATCTATAGGACCTCCTAATGAGAGTATCTGGGTGATTG

[3604] TGACCCCAAACCATCTCTCCAAAGCATTAATATCCAATCATGCGCTGTATGTTTTAATC

[3605] AGCAGAAGCATGTTTTTATGTTTGTACAAAAGAAGATTGTTATGGGTGGGGATGGAGGT

[3606] ATAGACCATGCATGGTCACCTTCAAGCTACTTTAATAAAGGATCTTAAAATGGGCAGGA

[3607] GGACTGTGAACAAGACACCCTAATAATGGGTTGATGTCTGAAGTAGCAAATCTTCTGGA

[3608] AACGCAAACTCTTTTAAGGAAGTCCCTAATTTAGAAACACCCACAAACTTCACATATCA

[3609] TAATTAGCAAACAATTGGAAGGAAGTTGCTTGAATGTTGGGGAGAGGAAAATCTATTGG

[3610] CTCTCGTGGGTCTCTTCATCTCAGAAATGCCAATCAGGTCAAGGTTTGCTACATTTTGT

[3611] ATGTGTGTGATGCTTCTCCCAAAGGTATATTAACTATATAAGAGAGTTGTGACAAAACA

[3612] GAATGATAAAGCTGCGAACCGTGGCACACGCTCATAGTTCTAGCTGCTTGGGAGGTTGA

[3613] GGAGGGAGGATGGCTTGAACACAGGTGTTCAAGGCCAGCCTGGGCAACATAACAAGATC

[3614] CTGTCTCTCAAAAAAAAAAAAAAAAAAAAGAAAGAGAGAGGGCCGGGCGTGGTGGCTCA

[3615] CGCCTGTAATCCCAGCACTTTGGGAGGCCGAGCCGGGCGGATCACCTGTGGTCAGGAGT

[3616] TTGAGACCAGCCTGGCCAACATGGCAAAACCCCGTCTGTACTCAAAATGCAAAAATTAG

[3617] CCAGGCGTGGTAGCAGGCACCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAAT

[3618] CGCTTGAACCCAGGAGGTGGAGGTTGCAGTAAGCTGAGATCGTGCCGTTGCACTCCAGC

[3619] CTGGGCGACAAGAGCAAGACTCTGTCTCAGAAAAAAAAAAAAAAAAAGAGAGAGAGAGAG

[3620] AAAGAGAACAATATTTGGGAGAGAAGGATGGGGAAGCATTGCAAGGAAATTGTGCTTTA

[3621] TCCAACAAAATGTAAGGAGCCAATAAGGGATCCCTATTTGTCTCTTTTGGTGTCTATTT

[3622] GTCCCTAACAACTGTCTTTGACAGTGAGAAAAATATTCAGAATAACCATATCCCTGTGC

[3623] CGTTATTACCTAGCAACCCTTGCAATGAAGATGAGCAGATCCACAGGAAAACTTGAATG

[3624] CACAACTGTCTTATTTTAATCTTATTGTACATAAGTTTGTAAAAGAGTTAAAAATTGTT

[3625] ACTTCATGTATTCATTTATATTTTATATTATTTTGCGTCTAATGATTTTTTATTAACAT

[3626] GATTTCCTTTTCTGATATATTGAAATGGAGTCTCAAAGCTTCATAAATTTATAACTTTA

[3627] GAAATGATTCTAATAACAACGTATGTAATTGTAACATTGCAGTAATGGTGCTACGAAGC

[3628] CATTTCTCTTGATTTTTAGTAAACTTTTATGACAGCAAATTTGCTTCTGGCTCACTTTC

[3629] AATCAGTTAAATAAATGATAAATAATTTTGGAAGCTGTGAAGATAAAATACCAAATAAA

[3630] ATAATATAAAAGTGATTTATATGAAGTTAAAATAAAAAATCAGTATGATGGAATAAACT

[3631] TG

[3632] 以下提供可用于本发明方法的其他胞苷脱氨酶。

[3633] rAPOBEC-1大鼠(Rattus norvegicus)

[3634] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVE

[3635] VNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRN

[3636] RQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLP

[3637] PCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK

[3638] mAPOBEC-1小家鼠(Mus musculus)

[3639] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSVWRHTSQNTSN

[3640] HVEVNFLEKFTTERYFRPNTRCSITWFLSWSPCGECSRAITEFLSRHPYVTLFIYIARLY

[3641] HHTDQRNRQGLRDLISSGVTIQIMTEQEYCYCWRNFVNYPPSNEAYWPRYPHLWVKLYVLE

[3642] LYCIILGLPPCLKILRRKQPQLTFFTITLQTCHYQRIPPHLLWATGLK

[3643] maAPOBEC-1黄金仓鼠(Mesocricetus auratus)

[3644] MSSETGPVVVDPTLRRRIEPHEFDAFFDQGELRKETCLLYEIRWGGRHNIWRHTGQNTSRHV

[3645] EINFIEKFTSERYFYPSTRCSIVWFLSWSPCGECSKAITEFLSGHPNVTLFIYAARLY

[3646] HHTDQRNRQGLRDLISRGVTIRIMTEQEYCYCWRNFVNYPPSNEVYWPRYPNLWMRLYALE

[3647] LYCIHLGLPPCLKIKRRHQYPLTFFRLNLQSCHYQRIPPHILWATGFI

[3648] hAPOBEC-1智人

[3649] MTSEKGPSTGDPTLRRRIEPWEFDVFYDPRELRKEACLLYEIKWGMSRKIWRSSGKNTTNH

[3650] VEVNFIKKFTSERDFHPSMSCSITWFLSWSPCWECSQAIREFLSRHPGVTLVIYVARLF

[3651] WHMDQQNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLY

[3652] ALELHCIILSLPPCLKISRRWQNHLTFFRLHLQNCHYQTIPPHILLATGLIHPSVAWR

[3653] ppAPOBEC-1婆罗洲猩猩(Pongo pygmaeus)

[3654] MTSEKGPSTGDPTLRRRIESWEFDVFYDPRELRKETCLLYEIKWGMSRKIWRSSGKNTTNHV

[3655] EVNFIKKFTSERRFHSSISCSITWFLSWSPCWECSQAIREFLSQHPGVTLVIYVARLF

[3656] WHMDQRNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLY

[3657] ALELHCIILSLPPCLKISRRWQNHLAFFRLHLQNCHYQTIPPHILLATGLIHPSVTWR

[3658] ocAPOBEC1穴兔(Oryctolagus cuniculus)

[3659] MASEKGPSNKDYTLRRRIEPWEFEVFFDPQELRKEACLLYEIKWGASSKTWRSSGKNTTNH

[3660] VEVNFLEKLTSEGRLGPSTCCSITWFLSWSPCWECSMAIREFLSQHPGVTLIIFVARLF

[3661] QHMDRRNRQGLKDLVTSGVTVRVMSVSEYCYCWENFVNYPPGKAAQWPRYPPRWMLMY

[3662] ALELYCIILGLPPCLKISRRHQKQLTFFSLTPQYCHYKMIPPYILLATGLLQPSVPWR

[3663] mdAPOBEC-1灰短尾负鼠(Monodelphis domestica)

[3664] MNSKTGPSVGDATLRRRIKPWEFVAFFNPQELRKETCLLYEIKWGNQNIWRHSNQNTSQHA

[3665] EINFMEKFTAERHFNSSVRCSITWFLSWSPCWECSKAIRKFLDHYPNVTLAIFISRLYWHMDQ

[3666] QHRQGLKELVHSGVTIQIMSYSEYHYCWRNFVDYPQGEEDYWPKYPYLWIMLYVLELHCII

[3667] LGLPPCLKISGSHSNQLALFSLDLQDCHYQKIPYNVLVATGLVQPFVTWR

[3668] mAPOBEC-2小家鼠

[3669] MAQKEEAAEAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFRNVEYS

[3670] SGRNKTFLCYVVEVQSKGGQAQATQGYLEDEHAGAHAEEAFFNTILPAFDPALKYNVTWY

[3671] VSSSPCAACADRILKTLSKTKNLRLLILVSRLFMWEEPEVQAALKKLKEAGCKLRIMKPQDF

[3672] EYIWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADILK

[3673] hAPOBEC-2智人

[3674] MAQKEEAAVATEAASQNGEDLENLDDPEKLKELIELPPFEIVTGERLPANFFKFQFRNVE

[3675] YSSGRNKTFLCYVVEAQGKGGQVQASRGYLEDEHAAAHAEEAFFNTILPAFDPALRYNVT

[3676] WYVSSSPCAACADRIIKTLSKTKNLRLLILVGRLFMWEEPEIQAALKKLKEAGCKLRIMKPQ

[3677] DFEYVWQNFVEQEEGESKAFQPWEDIQENFLYYEEKLADILK

[3678] ppAPOBEC-2婆罗洲猩猩

[3679] MAQKEEAAAATEAASQNGEDLENLDDPEKLKELIELPPFEIVTGERLPANFFKFQFRNVE

[3680] YSSGRNKTFLCYVVEAQGKGGQVQASRGYLEDEHAAAHAEEAFFNTILPAFDPALRYNVT

[3681] WYVSSSPCAACADRIIKTLSKTKNLRLLILVGRLFMWEELEIQDALKKLKEAGCKLRIMKPQ

[3682] DFEYVWQNFVEQEEGESKAFQPWEDIQENFLYYEEKLADILK

[3683] btAPOBEC-2牛(Bos Taurus)

[3684] MAQKEEAAAAAEPASQNGEEVENLEDPEKLKELIELPPFEIVTGERLPAHYFKFQFRNVE

[3685] YSSGRNKTFLCYVVEAQSKGGQVQASRGYLEDEHATNHAEEAFFNSIMPTFDPALRYMVT

[3686] WYVSSSPCAACADRIVKTLNKTKNLRLLILVGRLFMWEEPEIQAALRKLKEAGCRLRIMKP

[3687] QDFEYIWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADILK

[3688] mAPOBEC-3小家鼠

[3689] MQPQRLGPRAGMGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLGYAKGRKDTFLCYEVTR

[3690] KDCDSPVSLHHGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQI

[3691] VRFLATHHNLSLDIFSSRLYNVQDPETQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNG

[3692] GRRFRPWKRLLTNFRYQDSKLQEILRPCYISVPSSSSSTLSNICLTKGLPETRFWVEGRRMDPL

[3693] SEEEFYSQFYNQRVKHLCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKGKQHAEILFLDKI

[3694] RSMELSQVTITCYLTWSPCPNCAWQLAAFKRDRPDLILHIYTSRLYFHWKRPFQKGLCSLWQ

[3695] SGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLRRIKESWGLQDLVNDFG

[3696] NLQLGPPMS

[3697] hAPOBEC-3A智人

[3698] MEASPASGPRHLMDPHIFTSNFNNGIGRHKTYLCYEVERLDNGTSVKMDQHRGFLHNQAK

[3699] NLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGEVRAFLQENTHVRL

[3700] RIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFKHCWDTFVDHQGCPFQPWDGLDEH

[3701] SQALSGRLRAILQNQGN

[3702] hAPOBEC-3B智人

[3703] MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGQVYF

[3704] KPQYHAEMCFLSWFCGNQLPAYKCFQITWFVSWTPCPDCVAKLAEFLSEHPNVTLTISAARL

[3705] YYYWERDYRRALCRLSQAGARVTIMDYEEFAYCWENFVYNEGQQFMPWYKFDENYAFLH

[3706] RTLKEILRYLMDPDTFTFNFNNDPLVLRRRQTYLCYEVERLDNGTWVLMDQHMGFLCNEA

[3707] KNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGEVRAFLQENTHV

[3708] RLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVYRQGCPFQPWDGLEE

[3709] HSQALSGRLRAILQNQGN

[3710] hAPOBEC-3C智人

[3711] MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFRNQV

[3712] DSETHCHAERCFLSWFCDDILSPNTKYQVTWYTSWSPCPDCAGEVAEFLARHSNVNLTIFTA

[3713] RLYYFQYPCYQEGLRSLSQEGVAVEIMDYEDFKYCWENFVYNDNEPFKPWKGLKTNFRLL

[3714] KRRLRESLQ

[3715] hAPOBEC-3D智人

[3716] MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGPVLP

[3717] KRQSNHRQEVYFRFENHAEMCFLSWFCGNRLPANRRFQITWFVSWNPCLPCVVKVTKFLA

[3718] EHPNVTLTISAARLYYYRDRDWRWVLLRLHKAGARVKIMDYEDFAYCWENFVCNEGQPFM

[3719] PWYKFDDNYASLHRTLKEILRNPMEAMYPHIFYFHFKNLLKACGRNESWLCFTMEVTKHH

[3720] SAVFRKRGVFRNQVDPETHCHAERCFLSWFCDDILSPNTNYEVTWYTSWSPCPECAGEVAE

[3721] FLARHSNVNLTIFTARLCYFWDTDYQEGLCSLSQEGASVKIMGYKDFVSCWKNFVYSDDEP

[3722] FKPWKGLQTNFRLLKRRLREILQ

[3723] hAPOBEC-3F智人

[3724] MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPRLDAKIFRGQVYSQ

[3725] PEHHAEMCFLSWFCGNQLPAYKCFQITWFVSWTPCPDCVAKLAEFLAEHPNVTLTISAARLY

[3726] YYWERDYRRALCRLSQAGARVKIMDDEEFAYCWENFVYSEGQPFMPWYKFDDNYAFLHR

[3727] TLKEILRNPMEAMYPHIFYFHFKNLRKAYGRNESWLCFTMEVVKHHSPVSWKRGVFRNQV

[3728] DPETHCHAERCFLSWFCDDILSPNTNYEVTWYTSWSPCPECAGEVAEFLARHSNVNLTIFTA

[3729] RLYYFWDTDYQEGLRSLSQEGASVEIMGYKDFKYCWENFVYNDDEPFKPWKGLKYNFLFL

[3730] DSKLQEILE

[3731] hAPOBEC-3G智人

[3732] MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQVYSE

[3733] LKYHPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPKVTLTIFVAR

[3734] LYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEPWNNLPKYY

[3735] ILLHIMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLC

[3736] NQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKH

[3737] VSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVDHQGCPFQPWDGLDE

[3738] HSQDLSGRLRAILQNQEN

[3739] hAPOBEC-4智人

[3740] MEPIYEEYLANHGTIVKPYYWLSFSLDCSNCPYHIRTGEEARVSLTEFCQIFGFPYGTTF

[3741] PQTKHLTFYELKTSSGSLVQKGHASSCTGNYIHPESMLFEMNGYLDSAIYNNDSIRHIIL

[3742] YSNNSPCNEANHCCISKMYNFLITYPGITLSIYFSQLYHTEMDFPASAWNREALRSLASL

[3743] WPRVVLSPISGGIWHSVLHSFISGVSGSHVFQPILTGRALADRHNAYEINAITGVKPYFT

[3744] DVLLQTKRNPNTKAQEALESYPLNNAFPGQFFQMPSGQLQPNLPPDLRAPVVFVLVPLRDLP

[3745] PMHMGQNPNKPRNIVRHLNMPQMSFQETKDLGRLPTGRSVEIVEITEQFASSKEADEKKKK

[3746] KGKK

[3747] mAPOBEC-4小家鼠

[3748] MDSLLMKQKKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSCSLDFGHLRNKSGCHVELL

[3749] FLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVAEFLRWNPNLSLRIFTARLYFCEDRKAE

[3750] PEGLRRLHRAGVQIGIMTFKDYFYCWNTFVENRERTFKAWEGLHENSVRLTRQLRRILLPLY

[3751] EVDDLRDAFRMLGF

[3752] rAPOBEC-4大鼠

[3753] MEPLYEEYLTHSGTIVKPYYWLSVSLNCTNCPYHIRTGEEARVPYTEFHQTFGFPWSTYP

[3754] QTKHLTFYELRSSSGNLIQKGLASNCTGSHTHPESMLFERDGYLDSLIFHDSNIRHIILY

[3755] SNNSPCDEANHCCISKMYNFLMNYPEVTLSVFFSQLYHTENQFPTSAWNREALRGLASLWP

[3756] QVTLSAISGGIWQSILETFVSGISEGLTAVRPFTAGRTLTDRYNAYEINCITEVKPYFT

[3757] DALHSWQKENQDQKVWAASENQPLHNTTPAQWQPDMSQDCRTPAVFMLVPYRDLPPIHVN

[3758] PSPQKPRTVVRHLNTLQLSASKVKALRKSPSGRPVKKEEARKGSTRSQEANETNKSKWKKQ

[3759] TLFIKSNICHLLEREQKKIGILSSWSV

[3760] mfAPOBEC-4食蟹猕猴(Macaca fascicularis)

[3761] MEPTYEEYLANHGTIVKPYYWLSFSLDCSNCPYHIRTGEEARVSLTEFCQIFGFPYGTTY

[3762] PQTKHLTFYELKTSSGSLVQKGHASSCTGNYIHPESMLFEMNGYLDSAIYNNDSIRHIIL

[3763] YCNNSPCNEANHCCISKVYNFLITYPGITLSIYFSQLYHTEMDFPASAWNREALRSLASL

[3764] WPRVVLSPISGGIWHSVLHSFVSGVSGSHVFQPILTGRALTDRYNAYEINAITGVKPFFT

[3765] DVLLHTKRNPNTKAQMALESYPLNNAFPGQSFQMTSGIPPDLRAPVVFVLLPLRDLPPMHM

[3766] GQDPNKPRNIIRHLNMPQMSFQETKDLERLPTRRSVETVEITERFASSKQAEEKTKKKKGKKhAID智人

[3767] MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELL

[3768] FLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAE

[3769] PEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLY

[3770] EVDDLRDAFRTLGL

[3771] clAID家犬(Canis lupus familiaris)

[3772] MDSLLMKQRKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLDFGHLRNKSGCHVELL

[3773] FLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFAARLYFCEDRKAE

[3774] PEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENREKTFKAWEGLHENSVRLSRQLRRILLPLY

[3775] EVDDLRDAFRTLGL

[3776] btAID牛

[3777] MDSLLKKQRQFLYQFKNVRWAKGRHETYLCYVVKRRDSPTSFSLDFGHLRNKAGCHVELL

[3778] FLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFTARLYFCDKERKA

[3779] EPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPL

[3780] YEVDDLRDAFRTLGL

[3781] mAID小家鼠

[3782] MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELL

[3783] FLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAE

[3784] PEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLY

[3785] EVDDLRDAFRTLGL

[3786] pmCDA-1海七鳃鳗

[3787] MAGYECVRVSEKLDFDTFEFQFENLHYATERHRTYVIFDVKPQSAGGRSRRLWGYIINNPNV

[3788] CHAELILMSMIDRHLESNPGVYAMTWYMSWSPCANCSSKLNPWLKNLLEEQGHTLTMHFS

[3789] RIYDRDREGDHRGLRGLKHVSNSFRMGVVGRAEVKECLAEYVEASRRTLTWLDTTESMAA

[3790] KMRRKLFCILVRCAGMRESGIPLHLFTLQTPLLSGRVVWWRV

[3791] pmCDA-2海七鳃鳗

[3792] MELREVVDCALASCVRHEPLSRVAFLRCFAAPSQKPRGTVILFYVEGAGRGVTGGHAVNYN

[3793] KQGTSIHAEVLLLSAVRAALLRRRRCEDGEEATRGCTLHCYSTYSPCRDCVEYIQEFGASTG

[3794] VRVVIHCCRLYELDVNRRRSEAEGVLRSLSRLGRDFRLMGPRDAIALLLGGRLANTADGES

[3795] GASGNAWVTETNVVEPLVDMTGFGDEDLHAQVQRNKQIREAYANYASAVSLMLGELHVDP

[3796] DKFPFLAEFLAQTSVEPSGTPRETRGRPRGASSRGPEIGRQRPADFERALGAYGLFLHPRIVSR

[3797] EADREEIKRDLIVVMRKHNYQGP

[3798] pmCDA-5海七鳃鳗

[3799] MAGDENVRVSEKLDFDTFEFQFENLHYATERHRTYVIFDVKPQSAGGRSRRLWGYIINNPNV

[3800] CHAELILMSMIDRHLESNPGVYAMTWYMSWSPCANCSSKLNPWLKNLLEEQGHTLMMHF

[3801] SRIYDRDREGDHRGLRGLKHVSNSFRMGVVGRAEVKECLAEYVEASRRTLTWLDTTESMA

[3802] AKMRRKLFCILVRCAGMRESGMPLHLFT

[3803] yCD酿酒酵母菌(Saccharomyces cerevisiae)

[3804] MVTGGMASKWDQKGMDIAYEEAALGYKEGGVPIGGCLINNKDGSVLGRGHNMRFQKGSA

[3805] TLHGEISTLENCGRLEGKVYKDTTLYTTLSPCDMCTGAIIMYGIPRCVVGENVNFKSKGEKY

[3806] LQTRGHEVVVVDDERCKKIMKQFIDERPQDWFEDIGE

[3807] rAPOBEC-1(Δ177-186)

[3808] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVE

[3809] VNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRN

[3810] RQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRGLPPCLNILRRKQ

[3811] PQLTFFTIALQSCHYQRLPPHILWATGLK

[3812] rAPOBEC-1(Δ202-213)

[3813] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTS

[3814] QNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHV

[3815] TLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEA

[3816] HWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQHYQRLPPHILWATGLK

[3817] 人类AID:

[3818] MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGY
LRNKNGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVAD
FLRGNPNLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDY
FYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDA
FRTLGL

[3819] (底线:核定位序列;双底线:核输出信号)

[3820] 小鼠AID:

[3821] MDSLLMKQKKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSCSLDFGHLRNKSGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVAEFLRWNPNLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIGIMTFKDYFYCWNTFVENRERTFKAWEGLHENSVRLTRQLRRILLPLYEVDDLRDAFRMLGF

[3822] (底线:核定位序列;双底线:核输出信号)

[3823] 犬AID:

[3824] MDSLLMKQRKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLDFGHLRNKSGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFAARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENREKTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL

[3825] (底线:核定位序列;双底线:核输出信号)

[3826] 牛AID:

[3827] MDSLLKKQRQFLYQFKNVRWAKGRHETYLCYVVKRRDSPTSFSLDFGHLRNKAGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFTARLYFCDKERKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL

[3828] (底线:核定位序列;双底线:核输出信号)

[3829] 大鼠AID

[3830] MAVGSKPKAALVGPHWERERIWCFLCSTGLGTQQTGQTSRWLRPAATQDPVSPPRSLLMKQRKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLDFGYLRNKSGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLTGWGALPAGLMSPARPSDYFYCWNTFVENHERTFKAWEGLHENSVRLSRRLRRILLPLYEVDDLRDAFRTLGL

[3831] (底线:核定位序列;双底线:核输出信号)

[3832] 小鼠APOBEC-3

[3833] MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLHHGVFKNKDNI*HAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFEC*AEQIVRFLATHHNLSLDIFSSRLYNVQDPETQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWKRLLTNFRYQDSKLQEILRPCYIPVPSSSSSTLSNICLTKGLPETRFCVEGRRMDPLSEEEFYSQFYNQRVKHLCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKGKQ*HAEILFLDKIRSMELSQVTITCYLTWSPCPNC*AWQLAAFKRDRPDLILHIYTSRLYFHWKRPFQKGLCSLWQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLRRIKESWGLQDLVNDFGNLQLGPPMS

[3834] (斜体:核酸编辑结构域)

[3835] 大鼠APOBEC-3:

MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNRLRYAIDRKDTFLCYEVTRKDCDSPVS
LHHGVFKNKDNI*HAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFEC*AEQVLRFLA
THHNLSLDIFSSRLYNIRDPENQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRF
RPWKKLLTNFRYQDSKLQEILRPCYIPVPSSSSSTLSNICLTKGLPETRFCVERRRVHL
[3836] LSEEEFYSQFYNQRVKHLCYYHGVKPYLCYQLEQFNGQAPLKGCLLSEKGKQ*HAEILFL*
*DKIRSMELSQVIITCYLTWSPCPNC*AWQLAAFKRDRPDLILHIYTSRLYFHWKRPFQKG
LCSLWQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLHRIKESWGL
QDLVNDFGNLQLGPPMS

[3837] (斜体:核酸编辑结构域)

[3838] 猕猴APOBEC-3G:

MVEPMDPRTFVSNFNNRPILSGLNTVWLCCEVKTKDPSGPPLDAKIFQGKVYSKAKY*HP*
*EMRFLRWFHKWRQLHHDQEYKVTWYVSWSPCTRC*ANSVATFLAKDPKVTLTIFVARLYY
FWKPDYQQALRILCQKRGGPHATMKIMNYNEFQDCWNKFVDGRGKPFKPRNNLPKHYTL
[3839] LQATLGELLRHLMDPGTFTSNFNNKPWVSGQHETYLCYKVERLHNDTWVPLNQHRGFLR
NQAPNIHGFPKGR*HAELCFLDLIPFWKLDGQQYRVTCFTSWSPCFSC*AQEMAKFISNNE
HVSLCIFAARIYDDQGRYQEGLRALHRDGAKIAMMNYSEFEYCWDTFVDRQGRPFQPWD
GLDEHSQALSGRLRAI

[3840] (斜体:核酸编辑结构域;底线:细胞质定位信号)黑猩猩APOBEC-3G:

MKPHFRNPVERMYQDTFSDNFYNRPILSHRNTVWLCYEVKTKGPSRPPLDAKIFRGQVY
SKLKY*HPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKC*TRDVATFLAEDPKVTLTI
FVARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEPWNN
[3841] LPKYYILLHIMLGEILRHSMDPPTFTSNFNNELWVRGRHETYLCYEVERLHNDTWVLLN
QRRGFLCNQAPHKHGFLEGR*HAELCFLDVIPFWKLDLHQDYRVTCFTSWSPCFSC*AQEM
AKFISNNKHVSLCIFAARIYDDQGRCQEGLRTLAKAGAKISIMTYSEFKHCWDTFVDHQ
GCPFQPWDGLEEHSQALSGRLRAILQNQGN

[3842] (斜体:核酸编辑结构域;底线:细胞质定位信号)绿猴APOBEC-3G:

MNPQIRNMVEQMEPDIFVYYFNNRPILSGRNTVWLCYEVKTKDPSGPPLDANIFQGKLY
PEAKD*HPEMKFLHWFRKWRQLHRDQEYEVTWYVSWSPCTRC*ANSVATFLAEDPKVTLTI
[3843] FVARLYYFWKPDYQQALRILCQERGGPHATMKIMNYNEFQHCWNEFVDGQGKPFKPRKN
LPKHYTLLHATLGELLRHVMDPGTFTSNFNNKPWVSGQRETYLCYKVERSHNDTWVLLN
QHRGFLRNQAPDRHGFPKGR*HAELCFLDLIPFWKLDDQQYRVTCFTSWSPCFSC*AQKMA
[3844] KFISNNKHVSLCIFAARIYDDQGRCQEGLRTLHRDGAKIAVMNYSEFEYCWDTFVDRQG
RPFQPWDGLDEHSQALSGRLRAI

[3845] (斜体:核酸编辑结构域;底线:细胞质定位信号)人类APOBEC-3G:

MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQVY

SELKY*HPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKC*TRDMATFLAEDPKVTLTI

FVARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEPWNN

[3846] LPKYYILLHIMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLN

QRRGFLCNQAPHKHGFLEGR*HAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSC*AQEM

AKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVDHQ

GCPFQPWDGLDEHSQDLSGRLRAILQNQEN

[3847] (斜体:核酸编辑结构域;底线:细胞质定位信号)人类APOBEC-3F:

MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPRLDAKIFRGQVY

SQPEH*HAEMCFLSWFCGNQLPAYKCFQITWFVSWTPCPDC*VAKLAEFLAEHPNVTLTIS

AARLYYYWERDYRRALCRLSQAGARVKIMDDEEFAYCWENFVYSEGQPFMPWYKFDDNY

[3848] AFLHRTLKEILRNPMEAMYPHIFYFHFKNLRKAYGRNESWLCFTMEVVKHHSPVSWKRG

VFRNQVDPETH*CHAERCFLSWFCDDILSPNTNYEVTWYTSWSPCPEC*AGEVAEFLARHS

NVNLTIFTARLYYFWDTDYQEGLRSLSQEGASVEIMGYKDFKYCWENFVYNDDEPFKPW

KGLKYNFLFLDSKLQEILE

[3849] (斜体:核酸编辑结构域)

[3850] 人类APOBEC-3B:

MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGQV

YFKPQY*HAEMCFLSWFCGNQLPAYKCFQITWFVSWTPCPDC*VAKLAEFLSEHPNVTLTI

SAARLYYYWERDYRRALCRLSQAGARVTIMDYEEFAYCWENFVYNEGQQFMPWYKFDEN

[3851] YAFLHRTLKEILRYLMDPDTFTFNFNNDPLVLRRRQTYLCYEVERLDNGTWVLMDQHMG

FLCNEAKNLLCGFY*GRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGC*AGEVRA

FLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVYRQGC

PFQPWDGLEEHSQALSGRLRAILQNQGN

[3852] (斜体:核酸编辑结构域)

[3853] 大鼠APOBEC-3B:

[3854] MQPQGLGPNAGMGPVCLGCSHRRPYSPIRNPLKKLYQQTFYFHFKNVRYAWGRKNNFLC

[3855] YEVNGMDCALPVPLRQGVFRKQGHIHAELCFIYWFHDKVLRVLSPMEEFKVTWYMSWSP

[3856] CSKCAEQVARFLAAHRNLSLAIFSSRLYYYLRNPNYQQKLCRLIQEGVHVAAMDLPEFK

[3857] KCWNKFVDNDGQPFRPWMRLRINFSFYDCKLQEIFSRMNLLREDVFYLQFNNSHRVKPV

[3858] QNRYYRRKSYLCYQLERANGQEPLKGYLLYKKGEQHVEILFLEKMRSMELSQVRITCYL

[3859] TWSPCPNCARQLAAFKKDHPDLILRIYTSRLYFWRKKFQKGLCTLWRSGIHVDVMDLPQ

[3860] FADCWTNFVNPQRPFRPWNELEKNSWRIQRRLRRIKESWGL

[3861] 牛APOBEC-3B:

[3862] DGWEVAFRSGTVLKAGVLGVSMTEGWAGSGHPGQGACVWTPGTRNTMNLLREVLFKQQF

[3863] GNQPRVPAPYYRRKTYLCYQLKQRNDLTLDRGCFRNKKQRHAERFIDKINSLDLNPSQS

[3864] YKIICYITWSPCPNCANELVNFITRNNHLKLEIFASRLYFHWIKSFKMGLQDLQNAGIS

[3865] VAVMTHTEFEDCWEQFVDNQSRPFQPWDKLEQYSASIRRRLQRILTAPI

[3866] 黑猩猩APOBEC-3B：

[3867] MNPQIRNPMEWMYQRTFYYNFENEPILYGRSYTWLCYEVKIRRGHSNLLWDTGVFRGQM

[3868] YSQPEHHAEMCFLSWFCGNQLSAYKCFQITWFVSWTPCPDCVAKLAKFLAEHPNVTLTI

[3869] SAARLYYYWERDYRRALCRLSQAGARVKIMDDEEFAYCWENFVYNEGQPFMPWYKFDDN

[3870] YAFLHRTLKEIIRHLMDPDTFTFNFNNDPLVLRRHQTYLCYEVERLDNGTWVLMDQHMG

[3871] FLCNEAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGQVRA

[3872] FLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVYRQGC

[3873] PFQPWDGLEEHSQALSGRLRAILQVRASSLCMVPHRPPPPPQSPGPCLPLCSEPPLGSL

[3874] LPTGRPAPSLPFLLTASFSFPPPASLPPLPSLSLSPGHLPVPSFHSLTSCSIQPPCSSR

[3875] IRETEGWASVSKEGRDLG

[3876] 人类POBEC-3C：

[3877] MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFRNQV DSETH*CHAERCFLSWFCDDILSPNTKYQVTWYTSWSPCPDC*AGEVAEFLARHSNVNLTI FTARLYYFQYPCYQEGLRSLSQEGVAVEIMDYEDFKYCWENFVYNDNEPFKPWKGLKTN FRLLKRRLRESLQ

[3878] (斜体:核酸编辑结构域)

[3879] 大猩猩APOBEC-3C3C

[3880] MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFRNQV DSETH*CHAERCFLSWECDDILSPNTNYQVTWYTSWSPCPEC*AGEVAEFLARHSNVNLTI

[3881] FTARLYYFQDTDYQEGLRSLSQEGVAVKIMDYKDFKYCWENFVYNDDEPFKPWKGLKYN FRFLKRRLQEILE

[3882] 人类APOBEC-3A：

[3883] MEASPASGPRHLMDPHIFTSNFNNGIGRHKTYLCYEVERLDNGTSVKMDQHRGFLHNQA KNLLCGFYGR*HAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGC*AGEVRAFLQENT HVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFKHCWDTFVDHQGCPFQPWD GLDEHSQALSGRLRAILQNQGN

[3884] (斜体:核酸编辑结构域)

[3885] 猕猴APOBEC-3A：

[3886] MDGSPASRPRHLMDPNTFTFNFNNDLSVRGRHQTYLCYEVERLDNGTWVPMDERRGFLC NKAKNVPCGDYGC*HVELRFLCEVPSWQLDPAQTYRVTWFISWSPC*FRRGCAGQVRVFLQ ENKHVRLRIFAARIYDYDPLYQEALRTLRDAGAQVSIMTYEEFKHCWDTFVDRQGRPFQ PWDGLDEHSQALSGRLRAILQNQGN

[3887] (斜体:核酸编辑结构域)

[3888] 牛APOBEC-3A3A：

[3889] MDEYTFTENFNNQGWPSKTYLCYEMERLDGDATIPLDEYKGFVRNKGLDQPEKPC*HAEL YFLGKIHSWNLDRNQHYRLTCFISWSPC*YDCAQKLTTFLKENHHISLHILASRIYTHNR FGCHQSGLCELQAAGARITIMTFEDFKHCWETFVDHKGKPFQPWEGLNVKSQALCTELQ AILKTQQN

[3890] (斜体:核酸编辑结构域)

[3891] 人类APOBEC-3H:

[3892] MALLTAETFRLQFNNKRRLRRPYYPRKALLCYQLTPQNGSTPTRGYFENKKKC*HAEICF INEIKSMGLDETQCYQVTCYLTWSPCSSC*AWELVDFIKAHDHLNLGIFASRLYYHWCKP QQKGLRLLCGSQVPVEVMGFPKFADCWENFVDHEKPLSFNPYKMLEELDKNSRAIKRRL ERIKIPGVRAQGRYMDILCDAEV

[3893] (斜体:核酸编辑结构域)

[3894] 猕猴APOBEC-3H:

[3895] MALLTAKTFSLQFNNKRRVNKPYYPRKALLCYQLTPQNGSTPTRGHLKNKKKDHAEIRF

[3896] INKIKSMGLDETQCYQVTCYLTWSPCPSCAGELVDFIKAHRHLNLRIFASRLYYHWRPN

[3897] YQEGLLLLCGSQVPVEVMGLPEFTDCWENFVDHKEPPSFNPSEKLEELDKNSQAIKRRL

[3898] ERIKSRSVDVLENGLRSLQLGPVTPSSSIRNSR

[3899] 人类APOBEC-3D:

[3900] MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGPV LPKRQSNHRQEVYFRFEN*HAEMCFLSWFCGNRLPANRRFQITWFVSWNPC*LPCVVKVTK FLAEHPNVTLTISAARLYYYRDRDWRWVLLRLHKAGARVKIMDYEDFAYCWENFVCNEG QPFMPWYKFDDNYASLHRTLKEILRNPMEAMYPHIFYFHFKNLLKACGRNESWLCFTME VTKHHSAVFRKRGVFRNQVDPETHC*HAERCFLSWFCDDILSPNTNYEVTWYTSWSPCPE C*AGEVAEFLARHSNVNLTIFTARLCYFWDTDYQEGLCSLSQEGASVKIMGYKDFVSCWK NFVYSDDEPFKPWKGLQTNFRLLKRRLREILQ

[3901] (斜体:核酸编辑结构域)

[3902] 人类APOBEC-1:

[3903] MTSEKGPSTGDPTLRRRIEPWEFDVFYDPRELRKEACLLYEIKWGMSRKIWRSSGKNTT

[3904] NHVEVNFIKKFTSERDFHPSMSCSITWFLSWSPCWECSQAIREFLSRHPGVTLVIYVAR

[3905] LFWHMDQQNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMML

[3906] YALELHCIILSLPPCLKISRRWQNHLTFFRLHLQNCHYQTIPPHILLATGLIHPSVAWR小鼠APOBEC-1:

[3907] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSVWRHTSQNTS

[3908] NHVEVNFLEKFTTERYFRPNTRCSITWFLSWSPCGECSRAITEFLSRHPYVTLFIYIAR

[3909] LYHHTDQRNRQGLRDLISSGVTIQIMTEQEYCYCWRNFVNYPPSNEAYWPRYPHLWVKL

[3910] YVLELYCIILGLPPCLKILRRKQPQLTFFTITLQTCHYQRIPPHLLWATGLK

[3911] 大鼠APOBEC-1:

[3912] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTN

[3913] KHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIAR

[3914] LYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRL

[3915] YVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK

[3916] 人类APOBEC-2：

[3917] MAQKEEAAVATEAASQNGEDLENLDDPEKLKELIELPPFEIVTGERLPANFFKFQFRNV

[3918] EYSSGRNKTFLCYVVEAQGKGGQVQASRGYLEDEHAAAHAEEAFFNTILPAFDPALRYN

[3919] VTWYVSSSPCAACADRIIKTLSKTKNLRLLILVGRLFMWEEPEIQAALKKLKEAGCKLR

[3920] IMKPQDFEYVWQNFVEQEEGESKAFQPWEDIQENFLYYEEKLADILK

[3921] 小鼠APOBEC-2：

[3922] MAQKEEAAEAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFRNV

[3923] EYSSGRNKTFLCYVVEVQSKGGQAQATQGYLEDEHAGAHAEEAFFNTILPAFDPALKYN

[3924] VTWYVSSSPCAACADRILKTLSKTKNLRLLILVSRLFMWEEPEVQAALKKLKEAGCKLR

[3925] IMKPQDFEYIWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADILK

[3926] 大鼠APOBEC-2：

[3927] MAQKEEAAEAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFRNV

[3928] EYSSGRNKTFLCYVVEAQSKGGQVQATQGYLEDEHAGAHAEEAFFNTILPAFDPALKYN

[3929] VTWYVSSSPCAACADRILKTLSKTKNLRLLILVSRLFMWEEPEVQAALKKLKEAGCKLR

[3930] IMKPQDFEYLWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADILK

[3931] 牛APOBEC-2：

[3932] MAQKEEAAAAAEPASQNGEEVENLEDPEKLKELIELPPFEIVTGERLPAHYFKFQFRNV

[3933] EYSSGRNKTFLCYVVEAQSKGGQVQASRGYLEDEHATNHAEEAFFNSIMPTFDPALRYM

[3934] VTWYVSSSPCAACADRIVKTLNKTKNLRLLILVGRLFMWEEPEIQAALRKLKEAGCRLR

[3935] IMKPQDFEYIWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADILK

[3936] 海七鳃鳗CDA1(pmCDA1)

[3937] MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRAC

[3938] FWGYAVNKPQSGTERGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSP

[3939] CADCAEKILEWYNQELRGNGHTLKIWACKLYYEKNARNQIGLWNLRDN

[3940] GVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKRRSELS

[3941] FMIQVKILHTTKSPAV

[3942] 人类APOBEC3G链A

[3943] MDPPTFTFNFNNEPWWGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQ

[3944] APHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQ

[3945] EMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISFTYSE

[3946] FKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQ

[3947] 以下提供根据本公开内容的方面可以在Cas12的氨基酸序列内内部融合的其他实施例性脱氨酶。应当理解,在一些实施方案中,可以使用相应序列的活性结构域,例如,没有定位信号(例如,核定位序列、核输出信号或细胞质定位信号)的结构域。

[3948] C至T核碱基编辑蛋白的详细信息在国际PCT申请号PCT/US2016/058344(WO 2017/070632)和Komor,A.C.,et al.,“Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage”Nature 533,420-424(2016),其全部内容

通过引用方式并入本文。

[3949] 胞苷脱氨酶

[3950] 在一个实施方案中,本发明的融合蛋白包括胞苷脱氨酶。在一些实施方案中,本文所提供的胞苷脱氨酶能够将胞嘧啶或5-甲基胞嘧啶脱氨为尿嘧啶或胸腺嘧啶。在一些实施方案中,本文所提供的胞嘧啶脱氨酶能够使DNA中的胞嘧啶脱氨。胞苷脱氨酶可以衍生自任何合适的生物体。在一些实施方案中,胞苷脱氨酶来自原核生物。在一些实施方案中,胞苷脱氨酶来自细菌。在一些实施方案中,胞苷脱氨酶来自哺乳动物(例如人类)。

[3951] 在一些实施方案中,胞苷脱氨酶包括与本文所述任一胞苷脱氨酶氨基酸序列至少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。

[3952] 本发明的融合蛋白包括核酸编辑结构域。在一些实施方案中,核酸编辑结构域可以催化C至U碱基的变化。在一些实施方案中,核酸编辑结构域是脱氨酶结构域。在一些实施方案中,脱氨酶是胞苷脱氨酶或腺苷脱氨酶。在一些实施方案中,脱氨酶是载脂蛋白B mRNA-编辑复合物(APOBEC)家族脱氨酶。在一些实施方案中,脱氨酶是APOBEC1脱氨酶。在一些实施方案中,脱氨酶是APOBEC2脱氨酶。在一些实施方案中,脱氨酶是APOBEC3脱氨酶。在一些实施方案中,脱氨酶是APOBEC3A脱氨酶。在一些实施方案中,脱氨酶是APOBEC3B脱氨酶。在一些实施方案中,脱氨酶是APOBEC3C脱氨酶。在一些实施方案中,脱氨酶是APOBEC3D脱氨酶。在一些实施方案中,脱氨酶是APOBEC3E脱氨酶。在一些实施方案中,脱氨酶是APOBEC3F脱氨酶。在一些实施方案中,脱氨酶是APOBEC3G脱氨酶。在一些实施方案中,脱氨酶是APOBEC3H脱氨酶。在一些实施方案中,脱氨酶是APOBEC4脱氨酶。在一些实施方案中,脱氨酶是激活诱导的脱氨酶(AID)。在一些实施方案中,脱氨酶是脊椎动物脱氨酶。在一些实施方案中,脱氨酶是无脊椎动物脱氨酶。在一些实施方案中,脱氨酶是人类、黑猩猩、大猩猩、猴、牛、狗、大鼠或小鼠脱氨酶。在一些实施方案中,脱氨酶是人类脱氨酶。在一些实施方案中,脱氨酶是大鼠脱氨酶,例如rAPOBEC1。在一些实施方案中,脱氨酶是海七鳗胞苷脱氨酶1(pmCDA1)。在一些实施方案中,脱氨酶是人APOBEC3G。在一些实施方案中,脱氨酶是人类APOBEC3G的片段。在一些实施方案中,核酸编辑结构域是与本文所述任何脱氨酶的脱氨酶结构域至少80%、至少85%、至少90%、至少92%、至少95%、至少96%、至少97%、至少98%、至少99%,或至少99.5%相同。

[3953] 额外的结构域

[3954] 本文所述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。在一些实施方案中,碱基编辑器包括多核苷酸可编程的核苷酸结合结构域(例如,Cas9)、核碱基编辑结构域(例如,脱氨酶结构域)和一个或多个额外的结构域。在一些实施方案中,额外的结构域可以促进碱基编辑器的酶促或催化功能、碱基编辑器的结合功能,或者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。在一些实施方案中,碱基编辑器可包括核酸酶、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。

[3955] 在一些实施方案中,碱基编辑器可包括尿嘧啶糖基化酶抑制剂(UGI)结构域。UGI结构域可以例如通过抑制由C脱氨形成的U转换回到C核碱基来提高包括胞苷脱氨酶结构域的碱基编辑器的效率。在一些实施方案中,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。特此类实施方案中,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。在这样的实施方案中,可以在碱基编辑器中抑制BER,所述碱基编辑器包括结合单链、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,本公开内容考虑包括UGI结构域的碱基编辑器融合蛋白。

[3956] 在一些实施方案中,碱基编辑器包括双链断裂(DSB)结合蛋白的全部或部分作为结构域。例如,DSB结合蛋白可以包含噬菌体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。参见Komor,A.C.,et al.,"Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity"Science Advances3:eaao4774(2017),其全部内容特此以参考方式并入。

[3957] 另外,在一些实施方案中,Gam蛋白可以与碱基编辑器的N端融合。在一些实施方案中,Gam蛋白可以与碱基编辑器的C端融合。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。在一些实施方式中,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。在一些实施方案中,174个残基的Gam蛋白与碱基编辑器的N端融合。参见,例如Komor,A.C.,et al.,"Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity"Science Advances 3:eaao4774(2017)。在一些实施方案中,一个或多个突变可以改变碱基编辑器域相对于野生型结构域的长度。例如,至少一个结构域中的至少一个氨基酸的缺失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,任何结构域中的一个或多个取代都不会改变碱基编辑器的长度。

[3958] 在一些实施方案中,碱基编辑器可包括核酸聚合酶(NAP)的全部或部分作为结构域。例如,碱基编辑器可以包括真核NAP的全部或一部分。在一些实施方案中,并入碱基编辑器中的NAP或其部分是DNA聚合酶。在一些实施方案中,并入碱基编辑器中的NAP或其部分具有跨损伤聚合酶活性。在一些实施方案中,并入碱基编辑器中的NAP或其部分是跨损伤DNA聚合酶。在一些实施方案中,并入碱基编辑器中的NAP或其部分是Rev7、Rev1复合物、聚合酶ι、聚合酶κ或聚合酶η。在一些实施方案中,并入碱基编辑器中的NAP或其部分是真核聚合酶α、β、γ、δ、ε、ζ、η、ι、κ、λ、μ或ν组分。在一些实施方案中,并入碱基编辑器中的NAP或其部分包括与核酸聚合酶(例如,跨损伤DNA聚合酶)至少75%、80%、85%、90%、95%、96%、97%、98%、99%、或99.5%相同。

[3959] 其他核碱基编辑器

[3960] 本发明提供模块化多效应核碱基编辑器,其中本领域已知的几乎任何核碱基编辑器都可以插入本文所述融合蛋白中或替换为胞苷脱氨酶或腺苷脱氨酶。在一个实施方案中,本发明的特征在于包括无碱基核碱基编辑器结构域的多效应器核碱基编辑器。无碱基核碱基编辑器是本领域已知的并且描述于例如Kavli et al.,EMBO J.15:3442-3447,1996,其通过引用方式并入本文。

[3961] 在一个实施方案中,多效应核碱基编辑器包括以下域A至C、A至D或A至E:

[3962] NH_2-[A-B-C]-COOH,

[3963] NH_2-[A-B-C-D]-COOH或

[3964] NH_2-[A-B-C-D-E]-COOH

[3965] 其中A和C或A、C和E各自包括以下一种或多种:腺苷脱氨酶结构域或其活性片段、胞苷脱氨酶结构域或其活性片段、DNA糖基化酶结构域或其活性片段;并且其中B或B和D,各自包括一个或多个具有核酸序列特异性结合活性的结构域。

[3966] 在一个实施方案中,多效应核碱基编辑器包括NH_2-[A_n-B_o-C_n]-COOH,

[3967] NH_2-[A_n-B_o-C_n-D_o]-COOH,或

[3968] NH_2-[A_n-B_o-C_p-D_o-Eq]-COOH;

[3969] 其中A和C或A、C和E各自包括以下一种或多种:腺苷脱氨酶结构域或其活性片段、胞苷脱氨酶结构域或其活性片段和DNA糖基化酶结构域或其活性片段;其中n为整数:1、2、3、4或5,且其中p为整数:0、1、2、3、4或5;B或B和D各自包括具有核酸序列特异性结合活性的结构域;其中o是整数:1、2、3、4或5。

[3970] 基础编辑器系统

[3971] 本文所提供的碱基编辑器系统的使用包括以下步骤:(a)将受试者的多核苷酸(例如,双链或单链DNA或RNA)的目标核苷酸序列与包括核碱基编辑器(例如,双链或单链DNA或RNA)的碱基编辑器系统接触,腺苷碱基编辑器)和向导多核酸(例如,gRNA),其中所述目标核苷酸序列包括靶核碱基对;(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换为第二个核碱基;(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置换。应当理解,在一些实施方案中,省略了步骤(b)。在一些实施方案中,所述靶向的核碱基对是一种或多种基因中的多个核碱基对。在一些实施方案中,本文所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。在一些实施方案中,多个核碱基对位于同一基因中。在一些实施方案中,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。

[3972] 在一些实施方案中,切割的单链(切口链)与向导核酸杂交。在一些实施方案中,切割的单链与包括第一核碱基的链相反。在一些实施方案中,碱基编辑器包括Cas9结构域。在一些实施方案中,第一个碱基是腺嘌呤,而第二个碱基不是G、C、A或T。在一些实施方案中,第二个碱基是肌苷。

[3973] 本文所提供的碱基编辑系统提供一种基因组编辑的新方法,所述方法使用包括催化缺陷的化脓链球菌Cas9、胞苷脱氨酶及碱基切除修复抑制剂的融合蛋白来诱导可编程的单核苷酸(C→T或A→G)改变DNA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。

[3974] 本文所提供用于使用碱基编辑器系统编辑核碱基的系统、组合物和方法。在一些实施方案中,碱基编辑器系统包括(1)碱基编辑器(BE),其包括多核苷酸可编程的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)与多核苷酸可编程的核苷酸结合结构域结合的向导多核苷酸(例如,向导RNA)。在一些实施方案中,碱基编辑器系统包括腺苷碱基编辑器(ABE)。在一些实施方案中,多核苷酸可编程的核苷酸结合结构域是多核苷酸可编程的DNA结合结构域。在一些实施方案中,多核苷酸可编程的核苷酸结合结构域是多核苷酸可编程的RNA结合结构域。在一些实施方案中,核碱基编辑结构域是

脱氨酶结构域。在一些实施方案中,脱氨酶结构域可以是腺嘌呤脱氨酶或腺苷脱氨酶。在一些实施方案中,腺苷碱基编辑器可使DNA中的腺嘌呤脱氨。在一些实施方案中,ABE包括进化的TadA变体。

[3975] 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,它们各自通过引用方式整体并入本文。另参见Komor,A.C.,et al.,"Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage"Nature 533,420-424(2016);Gaudelli,N.M.,et al.,"Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage"Nature 551,464-471(2017);以及Komor,A.C.,et al.,"Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity"Science Advances 3:eaao4774(2017),其全部内容通过引用方式并入本文。

[3976] 在一些实施方案中,可利用单个向导多核苷酸将脱氨酶靶向目标核酸序列。在一些实施方案中,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。

[3977] 碱基编辑器系统的核碱基成分和多核苷酸可编程的核苷酸结合成分可以共价或非共价地彼此连结。例如,在一些实施方案中,脱氨酶结构域可以通过多核苷酸可编程的核苷酸结合结构域靶向目标核苷酸序列。在一些实施方案中,多核苷酸可编程的核苷酸结合结构域可以与脱氨酶结构域融合或连接。在一些实施方案中,多核苷酸可编程的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连结将脱氨酶结构域靶向目标核苷酸序列。例如,在一些实施方案中,核碱基编辑成分(例如脱氨酶成分)可以包括能够与另外的异源部分或结构域相互作用、连结或形成复合物的作为多核苷酸可编程的核苷酸结合结构域的一部分额外的异源部分或结构域。在一些实施方案中,额外的异源部分可能能够与多肽结合、相互作用、连结或形成复合物。在一些实施方案中,额外的异源部分可能能够与多核苷酸结合、相互作用、连结或形成复合物。在一些实施方案中,额外的异源部分可能能够结合向导多核苷酸。在一些实施方案中,额外的异源部分可能能够结合多肽连接子。在一些实施方案中,额外的异源部分能够结合多核苷酸连接子。额外的异源部分可以是蛋白质结构域。在一些实施方案中,额外的异源部分可以是K同源性(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、不育α基序、端粒酶Ku结合基序和Ku蛋白、端粒酶Sm7结合基序和Sm7蛋白、或RNA识别基序。

[3978] 碱基编辑器系统可进一步包括向导多核苷酸组分。应当理解,碱基编辑器系统的组分可以经由共价键、非共价相互作用或其连结和相互作用的任何组合彼此关联。在一些实施方案中,脱氨酶结构域可以经由向导多核苷酸靶向目标核苷酸序列。例如,在一些实施方案中,碱基编辑器系统的核碱基编辑组分(例如脱氨酶组分)可以包括额外的异源部分或结构域(例如,多核苷酸结合结构域,例如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、连结或能够与向导多核苷酸的部分或区段(例如,多核苷酸基序)形成复合物。在一些实施方案中,额外的异源部分或结构域(例如,多核苷酸结合结构域,诸如RNA或DNA结合蛋白)可以与脱氨酶结构域融合或连接。在一些实施方案中,额外的异源部分可能能够与多肽结合、相互作用、连结或形成复合物。在一些实施方案中,额外的异源部分可能能够与多核苷酸结合、相互作用、连结或形成复合物。在一些实

施方案中,额外的异源部分可能能够结合向导多核苷酸。在一些实施方案中,额外的异源部分可能能够结合多肽连接子。在一些实施方案中,额外的异源部分能够结合多核苷酸连接子。额外的异源部分可以是蛋白质结构域。在一些实施方案中,额外的异源部分可以是K同源性(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、不育α基序、端粒酶Ku结合基序及Ku蛋白,端粒酶Sm7结合基序和Sm7蛋白,或RNA识别基序。

[3979] 在一些实施方案中,碱基编辑器系统可进一步包括碱基切除修复(BER)组分的抑制剂。应当理解,碱基编辑器系统的组分可以经由共价键、非共价相互作用或其连结和相互作用的任何组合彼此连结。BER组分的抑制剂可以包括碱基切除修复抑制剂。在一些实施方案中,碱基切除修复抑制剂可以是尿嘧啶DNA糖基化酶抑制剂(UGI)。在一些实施方案中,碱基切除修复抑制剂可以是肌苷碱基切除修复抑制剂。在一些实施方案中,碱基切除修复抑制剂可以通过多核苷酸可编程的核苷酸结合结构域靶向目标核苷酸序列。在一些实施方案中,多核苷酸可编程的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。在一些实施方案中,多核苷酸可编程的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。在一些实施方案中,多核苷酸可编程的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连结来将碱基切除修复抑制剂靶向目标核苷酸序列。例如,在一些实施方案中,碱基切除修复抑制剂组分可包括额外的异源部分或结构域,所述部分或结构域能够与作为多核苷酸可编程的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、连结或形成复合物。在一些实施方案中,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,在一些实施方案中,碱基切除修复抑制剂可包括能够与相互作用、连结或能够形成的额外的异源部分或结构域(例如,多核苷酸结合结构域,例如RNA或DNA结合蛋白)与向导多核苷酸的部分或区段(例如,多核苷酸基序)的复合物。在一些实施方案中,向导多核苷酸的额外异源部分或结构域(例如,多核苷酸结合结构域如RNA或DNA结合蛋白)可以与碱基切除修复抑制剂融合或连接。在一些实施方案中,额外的异源部分可能能够与多核苷酸结合、相互作用、连结或形成复合物。在一些实施方案中,额外的异源部分可能能够结合向导多核苷酸。在一些实施方案中,额外的异源部分可能能够结合多肽连接子。在一些实施方案中,额外的异源部分能够结合多核苷酸连接子。额外的异源部分可以是蛋白质结构域。在一些实施方案中,额外的异源部分可以是K同源性(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。

[3980] 在一些实施方案中,碱基编辑器抑制编辑链的碱基切除修复(BER)。在一些实施方案中,碱基编辑器保护或结合非编辑链。在一些实施方案中,碱基编辑器包括UGI活性。在一些实施方案中,碱基编辑器包括催化失活的肌苷特异性核酸酶。在一些实施方案中,碱基编辑器包括切口酶活性。在一些实施方案中,碱基对的预期编辑在PAM位点的上游。在一些实施方案中,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。在一些实施方案中,碱基对的预期编辑在PAM位点的下游。在一些实施方案中,预期的编辑碱基对是PAM位点下游的核苷酸流的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。

[3981] 在一些实施方案中,所述方法不需要规范的(例如,NGG)PAM位点。在一些实施方案中,核碱基编辑器包括连接子或间隔。在一些实施方案中,连接子或间隔物的长度为1至25

个氨基酸。在一些实施方案中,连接子或间隔物的长度为5至20个氨基酸。在一些实施方案中,连接子或间隔的长度为10、11、12、13、14、15、16、17、18、19或20个氨基酸。

[3982] 在一些实施方案中,本文所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,“脱氨窗口”)内的位置。在一些实施方案中,目标可以在4个碱基区域内。在一些实施方案中,这种限定的目标区域可以是PAM上游的大约15个碱基。参见Komor,A.C.,et al.,“Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage”Nature 533,420-424(2016);Gaudelli,N.M.,et al.,“Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage”Nature551,464-471(2017);以及Komor,A.C.,et al.,“Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity”Science Advances 3:eaao4774(2017),其全部内容特此通过引入方式并入。

[3983] 在一些实施方案中,目标区域包括目标窗口,其中所述目标窗口包括目标核碱基对。在一些实施方案中,目标窗口包括1至10个核苷酸。在一些实施方案中,目标窗口的长度是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。在一些实施方案中,碱基对的预期编辑是在目标窗口内。在一些实施方案中,目标窗口包括碱基对的预期编辑。在一些实施方案中,使用本文所提供的任何碱基编辑器来执行所述方法。在一些实施方案中,目标窗口是脱氨窗口。脱氨窗口可以是其中碱基编辑器作用于目标核苷酸并使其脱氨的限定区域。在一些实施方案中,脱氨窗口是在2、3、4、5、6、7、8、9或10个碱基区域内。在一些实施方案中,脱氨窗为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。

[3984] 本公开的碱基编辑器可以包括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,在一些实施方案中,碱基编辑器包括核定位序列(NLS)。在一些实施方案中,碱基编辑器的NLS位于脱氨酶结构域和多核苷酸可编程的核苷酸结合结构域之间。在一些实施方案中,碱基编辑器的NLS位于多核苷酸可编程的核苷酸结合结构域的C端。

[3985] 可存在于本文公开的碱基编辑器中的其他实施例性特征是定位序列,例如细胞质定位序列、输出序列(诸如核输出序列)或其他定位序列,以及可用于增溶、纯化或检测融合蛋白的序列标签。本文所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标签及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。在一些实施方案中,融合蛋白包括一个或多个His标签。

[3986] 可包括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、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。

[3987] 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-半乳糖基化酶、

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、GAL4 DNA结合结构域融合和单纯疱疹病毒(HSV)BP16蛋白融合。

[3988] 碱基编辑器或碱基编辑器系统可以是本文所述任何碱基编辑器。在一些实施方案中,碱基编辑器包括如本文所述腺苷脱氨酶单体或腺苷脱氨酶二聚体。在一些实施方案中,碱基编辑器包括具有插入Cas9多肽的柔性环中的脱氨酶(例如腺苷脱氨酶)的Cas9。在一些实施方案中,碱基编辑器通过表达各自编码Cas9-分裂内含肽融合蛋白的N端片段和C端片段的两个多核苷酸产生。碱基编辑器可经由RNP、载体、病毒载体或核酸如mRNA递送至宿主细胞。在一些实施方案中,编码碱基编辑器的多核苷酸构建体包括结构pMRNA-Trilink-ISLAY3-monoTadA-ABE7.10(V82S)-MQKFRAER 120A BbsI、pMRNA-Trilink-ISLAY3-ABE7.10(V82S、Y147T、Q154S)MQKFRAER 120A BbsI、或pMRNA-Trilink-ISLAY3-ABE7.10(V82T、Y147T、Q154S)-MQKFRAER 120A BbsI。向导RNA可以被化学修饰。在一些实施方案中,向导RNA包括一种或多种化学修饰的核碱基,诸如2'-O-甲基(2'-OMe)、2'-脱氧(2'-H)、2'-OC1-3烷基-O-C1-3烷基,诸如2'-甲氧基乙基("2'-MOE")、2'-氟("2'-F")、2'-氨基("2'-NH2")、2'-阿拉伯糖基("2'-阿拉伯")核苷酸、2'-F-阿拉伯糖基("2'-F-阿拉伯")核苷酸、2'-锁定的核酸("LNA")核苷酸、2'-解锁的核酸("ULNA")")核苷酸、L形式的糖("L-糖")、4'-硫代核糖核苷酸或如本文所述任何化学修饰。在一些实施方案中,向导RNA包括核苷酸间键修饰例如硫代磷酸酯"P(S)"(P(S))、膦酰基羧酸酯(P(CH2)nCOOR)诸如膦酰基乙酸酯"PACE"(P(CH2COO-))、硫代膦酸羧酸盐((S)P(CH2)nCOOR)诸如硫代膦酰乙酸盐"thioPACE"((S)P(CH2)nCOO-))、烷基膦酸盐(P(C1-3烷基)诸如甲基膦酸盐-P(CH3)、硼膦酸盐(P(BH3))及二硫代磷酸酯(P(S)2)。在一些实施方案中,向导RNA包括核碱基化学修饰,诸如2-硫尿嘧啶("2-硫代U")、2-硫代胞嘧啶("2-硫代C")、4-硫尿嘧啶("4-硫代U")、6-硫鸟嘌呤("6-硫代G")、2-氨基腺嘌呤("2-氨基A")、2-氨基嘌呤、假尿嘧啶、次黄嘌呤、7-脱氮鸟嘌呤、7-脱氮杂-8-氮杂鸟嘌呤、7-脱氮杂腺嘌呤、7-脱氮-8-氮杂腺嘌呤、5-甲基胞嘧啶("5-甲基C")、5-甲基尿嘧啶("5-甲基U")、5-羟甲基胞嘧啶、5-羟甲基尿嘧啶、5,6-脱氢尿嘧啶、5-丙炔基胞嘧啶、5-丙炔基尿嘧啶、5-乙炔基胞嘧啶、5-乙炔基尿嘧啶、5-烯丙基尿嘧啶("5-烯丙基U")、5-烯丙基胞嘧啶("5-烯丙基C")、5-氨基烯丙基尿嘧啶("5-氨基烯丙基U")、5-氨基烯丙基-胞嘧啶("5-氨基烯丙基C")、无碱基核苷酸、Z碱基、P碱基、非结构化核酸("UNA")、异鸟嘌呤("异G")、异胞嘧啶("异C")。在一些实施方案中,向导RNA在核苷酸糖、核碱基、磷酸二酯键和/或核苷酸磷酸酯上包括一种或多种同位素修饰。此类修饰包括一个或多个15N、13C、14C、氘、3H、32P、125I、131I原子或其其他原子或元素的核苷酸。

[3989] 在一些实施方案中,腺苷碱基编辑器(ABE)可使DNA中的腺嘌呤脱氨。在一些实施方案中,ABE是通过用天然的或工程化的大肠杆菌TadA、人类ADAR2、小鼠ADA或人类ADAT2替换BE3的APOBEC组分而产生的。在一些实施方案中,ABE包括进化的TadA变体。在一些实施方案中,ABE是ABE 1.2(TadA*-XTEN-nCas9-NLS)。在一些实施方案中,TadA*包括A106V和D108N突变。

[3990] 在一些实施方案中,ABE是第二代ABE。在一些实施方案中,ABE是ABE2.1,其包括

TadA*(TadA*2.1)中的额外突变D147Y和E155V。在一些实施方案中,ABE是与催化失活版本的人烷基腺嘌呤DNA糖基化酶(具有E125Q突变的AAG)融合的ABE2.2,ABE2.1。在一些实施方案中,ABE是与催化灭活版本的大肠杆菌Endo V(用D35A突变灭活)融合的ABE2.3,ABE2.1。在一些实施方案中,ABE是ABE2.6,其连接子长度(32个氨基酸,$(SGGS)_2$-XTEN-$(SGGS)_2$)是ABE2.1中连接子的两倍。在一些实施方案中,ABE是ABE2.7,其是与额外的野生型TadA单体相连的ABE2.1。在一些实施方案中,ABE是ABE2.8,其是与额外的TadA*2.1单体拴系的ABE2.1。在一些实施方案中,ABE是ABE2.9,其是进化的TadA(TadA*2.1)与ABE2.1的N端的直接融合。在一些实施方案中,ABE是ABE2.10,其是野生型TadA与ABE2.1的N端的直接融合。在一些实施方案中,ABE是ABE2.11,其是在TadA*单体的N端具有失活的E59A突变的ABE2.9。在一些实施方案中,ABE是ABE2.12,其是在内部TadA*单体中具有失活的E59A突变的ABE2.9。

[3991] 在一些实施例中,ABE是第三代ABE。在一些实施方案中,ABE是ABE3.1,其是具有三个额外的TadA突变(L84F、H123Y和I156F)的ABE2.3。

[3992] 在一些实施例中,ABE是第四代ABE。在一些实施方案中,ABE是ABE4.3,其是具有额外TadA突变A142N(TadA*4.3)的ABE3.1。

[3993] 在一些实施例中,ABE是第五代ABE。在一些实施方案中,ABE是ABE5.1,其通过将来自存活的克隆(H36L、R51L、S146C和K157N)的一组共有突变导入ABE3.1而产生。在一些实施方案中,ABE是ABE5.3,其具有包括融合内部进化的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,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,ABE是第六代ABE。在一些实施例中,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,ABE是第七代ABE。在一些实施例中,ABE是ABE7.1、ABE7.2、ABE7.3、ABE7.4、ABE7.5、ABE7.6、ABE7.7、ABE7.8、ABE 7.9或ABE7.10,如下表6所示。

[3994] 表6.ABE的基因型

[3995]

	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

[3996]

	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

[3997]

	23	26	36	37	48	49	51	72	84	87	105	108	123	125	142	145	147	152	155	156	157	161
ABE7.2	W	R	L	N	A		L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
ABE7.3	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

[3998] 在一些实施例中,碱基编辑器是腺苷碱基编辑器。在一些实施方案中,腺苷碱基编辑器是第八代ABE(ABE8)。在一些实施方案中,ABE8包括TadA*8变体。在一些实施例中,ABE8具有包括TadA*8变体(“ABE8.x-m”)的单体构建体。在一些实施方案中,ABE8是ABE8.1-m,其具有含有Y147T突变(TadA*8.1)的TadA*7.10的单体构建体。在一些实施方案中,ABE8是ABE8.2-m,其具有含有Y147R突变(TadA*8.2)的TadA*7.10的单体构建体。在一些实施方案中,ABE8是ABE8.3-m,其具有含有Q154S突变(TadA*8.3)的TadA*7.10的单体构建体。在一些实施方案中,ABE8是ABE8.4-m,其具有含有Y123H突变(TadA*8.4)的TadA*7.10的单体构建体。在一些实施方案中,ABE8是ABE8.5-m,其具有含有V82S突变(TadA*8.5)的TadA*7.10的单体构建体。在一些实施方案中,ABE8是ABE8.6-m,其具有含有T166R突变(TadA*8.6)的TadA*7.10的单体构建体。在一些实施方案中,ABE8是ABE8.7-m,其具有含有Q154R突变(TadA*8.7)的TadA*7.10的单体构建体。在一些实施方案中,ABE8是ABE8.8-m,其具有含有Y147R、Q154R和Y123H突变(TadA*8.8)的TadA*7.10的单体构建体。在一些实施方案中,ABE8是ABE8.9-m,其具有含有Y147R、Q154R和I76Y突变(TadA*8.9)的TadA*7.10的单体构建体。在一些实施方案中,ABE8是ABE8.10-m,其具有含有Y147R、Q154R和T166R突变(TadA*8.10)的TadA*7.10的单体构建体。在一些实施方案中,ABE8是ABE8.11-m,其具有含有Y147T和Q154R突变(TadA*8.11)的TadA*7.10的单体构建体。在一些实施方案中,ABE8是ABE8.12-m,其具有含有Y147T和Q154S突变(TadA*8.12)的TadA*7.10的单体构建体。在一些实施方案中,ABE8是ABE8.13-m,其具有含有TadA*7.10和Y123H(从H123Y回复的Y123H)、Y147R、Q154R和I76Y突变(TadA*8.13)的单体构建体。在一些实施方案中,ABE8是ABE8.14-m,其具有含有I76Y和V82S突变(TadA*8.14)的TadA*7.10的单体构建体。在一些实施方案中,ABE8是ABE8.15-m,其具有含有V82S和Y147R突变(TadA*8.15)的TadA*7.10的单体构建体。在一些实施方案中,ABE8是ABE8.16-m,其具有含有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。在一些实施方案中,ABE8是ABE8.17-m,其具有含有V82S和Q154R突变(TadA*8.17)的TadA*7.10的单体构建体。在一些实施方案中,ABE8是ABE8.18-m,其具有含有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。在一些实施方案中,ABE8是ABE8.19-m,其具有含有TadA*7.10和V82S、Y123H(从H123Y回复的Y123H)、Y147R和Q154R突变(TadA*8.19)的单体构建体。在一些实施方案中,ABE8是ABE8.20-m,其具有含有TadA*7.10和I76Y、V82S、Y123H(Y123H从H123Y回复)、Y147R和Q154R突变(TadA*8.20)的单体构建体。在一些实施方案中,ABE8是ABE8.21-

m,其具有含有Y147R和Q154S突变(TadA*8.21)的TadA*7.10的单体构建体。在一些实施方案中,ABE8是ABE8.22-m,其具有含有V82S和Q154S突变(TadA*8.22)的TadA*7.10的单体构建体。在一些实施方案中,ABE8是ABE8.23-m,其具有含有V82S和Y123H(从H123Y恢复的Y123H)突变(TadA*8.23)的TadA*7.10的单体构建体。在一些实施方案中,ABE8是ABE8.24-m,其具有含有TadA*7.10和V82S、Y123H(Y123H从H123Y回复)和Y147T突变(TadA*8.24)的单体构建体。

[3999] 在一些实施方案中,ABE8具有含有与TadA*8变体("ABE8.x-d")融合的野生型大肠杆菌TadA的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.1-d,其具有含有与具有Y147T突变(TadA*8.1)的TadA*7.10融合的野生型大肠杆菌TadA的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.2-d,其具有含有与具有Y147R突变(TadA*8.2)的TadA*7.10融合的野生型大肠杆菌TadA的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.3-d,其具有含有与具有Q154S突变(TadA*8.3)的TadA*7.10融合的野生型大肠杆菌TadA的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.4-d,其具有含有与具有Y123H突变(TadA*8.4)的TadA*7.10融合的野生型大肠杆菌TadA的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.5-d,其具有含有与具有V82S突变(TadA*8.5)的TadA*7.10融合的野生型大肠杆菌TadA的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.6-d,其具有含有与具有T166R突变(TadA*8.6)的TadA*7.10融合的野生型大肠杆菌TadA的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.7-d,其具有含有与具有Q154R突变(TadA*8.7)的TadA*7.10融合的野生型大肠杆菌TadA的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.8-d,其具有含有与具有Y147R、Q154R和Y123H突变(TadA*8.8)的TadA*7.10融合的野生型大肠杆菌TadA的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.9-d,其具有含有与具有Y147R、Q154R和I76Y突变(TadA*8.9)的TadA*7.10融合的野生型大肠杆菌TadA的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.10-d,其具有含有与具有Y147R、Q154R和T166R突变(TadA*8.10)的TadA*7.10融合的野生型大肠杆菌TadA的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.11-d,其具有含有与具有Y147T和Q154R突变(TadA*8.11)的TadA*7.10融合的野生型大肠杆菌TadA的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.12-d,其具有含有与具有Y147T和Q154S突变(TadA*8.12)的TadA*7.10融合的野生型大肠杆菌TadA的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.13-d,其具有含有与具有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的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.14-d,其具有含有具有I76Y和V82S突变的TadA*7.10(TadA*8.14)的野生型大肠杆菌TadA融合的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.15-d,其具有含有具有V82S和Y147R突变(TadA*8.15)的TadA*7.10的野生型大肠杆菌TadA融合的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.16-d,其具有含有具有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融合的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.17-d,其具有含有具有V82S和Q154R突变(TadA*8.17)的TadA*7.10融合的野生型大肠杆菌TadA的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.18-d,其具有含有具有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的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.19-d,其具有含有

具有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的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.20-d,其具有含有具有I76Y、V82S、Y123H(从H123Y恢复的Y123H)、Y147R和Q154R突变(TadA)融合TadA*7.10的野生型大肠杆菌TadA的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.21-d,其具有含有具有Y147R和Q154S突变(TadA*8.21)的TadA*7.10融合的野生型大肠杆菌TadA。在一些实施方案中,ABE8是ABE8.22-d,其具有含有具有V82S和Q154S突变的TadA*7.10(TadA*8.22)的野生型大肠杆菌TadA融合的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.23-d,其具有含有具有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融合的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.24-d,其具有含有具有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融合的异源二聚体构建体。

[4000] 在一些实施方案中,ABE8具有含有与TadA*8变体("ABE8.x-7")融合的TadA*7.10的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.1-7,其具有含有与具有Y147T突变(TadA*8.1)的TadA*7.10融合的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.2-7,其具有含有与具有Y147R突变(TadA*8.2)的TadA*7.10融合的TadA*7.10的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.3-7,其具有含有与具有Q154S突变(TadA*8.3)的TadA*7.10融合的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.4-7,其具有含有与具有Y123H突变(TadA*8.4)的TadA*7.10融合的TadA*7.10的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.5-7,其具有含有与具有V82S突变(TadA*8.5)的TadA*7.10融合的TadA*7.10的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.6-7,其具有含有与具有T166R突变(TadA*8.6)的TadA*7.10融合的TadA*7.10的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.7-7,其具有含有与具有Q154R突变(TadA*8.7)的TadA*7.10融合的TadA*7.10的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.8-7,其具有含有与具有Y147R、Q154R和Y123H突变(TadA*8.8)的TadA*7.10融合的TadA*7.10融合的的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.9-7,其具有含有与具有Y147R、Q154R和I76Y突变(TadA*8.9)的TadA*7.10融合的TadA*7.10的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.10-7,其具有含有与具有Y147R、Q154R和T166R突变(TadA*8.10)的TadA*7.10融合的TadA*7.10的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.11-7,其具有含有与具有Y147T和Q154R突变(TadA*8.11)的TadA*7.10融合的TadA*7.10的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.12-7,其具有包括与具有Y147T和Q154S突变(TadA*8.12)的TadA*7.10融合的TadA*7.10融合的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.13-7,其具有包括与具有Y123H(从H123Y回复的Y123H)、Y147R、Q154R和I76Y突变(TadA*8.13)的TadA*7.10融合的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.14-7,其具有包括与具有I76Y和V82S突变(TadA*8.14)的TadA*7.10融合的TadA*7.10的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.15-7,其具有包括与具有V82S和Y147R突变(TadA*8.15)的TadA*7.10融合的TadA*7.10融合的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.16-7,其具有含有TadA*7.10其具有包括与具有V82S、Y123H(从H123Y回复的Y123H)和Y147R突变(TadA*8.16)的TadA*7.10融合的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.17-7,其具有含有与具有V82S和Q154R突变(TadA*8.17)的TadA*7.10融合

的TadA*7.10融合的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.18-7,其具有含有与具有V82S、Y123H(从H123Y回复的Y123H)和Q154R突变(TadA*8.18)的TadA*7.10融合的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.19-7,其具有含有与具有V82S、Y123H(从H123Y回复的Y123H)、Y147R和Q154R突变(TadA*8.19)的TadA*7.10融合的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.20-7,其具有含有与具有I76Y、V82S、Y123H(Y123H从H123Y回复)、Y147R和Q154R突变(TadA*8.20)的TadA*7.10融合的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.21-7,其具有含有与具有Y147R和Q154S突变(TadA*8.21)的TadA*7.10融合的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.22-7,其具有含有与具有V82S和Q154S突变(TadA*8.22)的TadA*7.10融合的TadA*7.10的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.23-7,其具有含有与具有V82S和Y123H(从H123Y恢复的Y123H)突变(TadA*8.23)的TadA*7.10融合的异源二聚体构建体。在一些实施方案中,ABE8是ABE8.24-7,其具有含有与具有V82S、Y123H(从H123Y回复的Y123H)和Y147T突变(TadA*8.24)的TadA*7.10融合的异源二聚体构建体。

[4001] 在一些实施方案中,ABE是ABE8.1-m、ABE8.2-m、ABE8.3-m、ABE8.4-m、ABE8.5-m、ABE8.6-m、ABE8.7-m、ABE8.8-m、ABE8.9-m、ABE8.10-m、ABE8.11-m、ABE8.12-m、ABE8.13-m、ABE8.14-m、ABE8.15-m、ABE8.16-m、ABE8.17-m、ABE8.18-m、ABE8.19-m、ABE8.20-m、ABE8.21-m、ABE8.22-m、ABE8.23-m、ABE8.24-m、ABE8.1-d、ABE8.2-d、ABE8.3-d、ABE8.4-d、ABE8.5-d、ABE8.6-d、ABE8.7-d、ABE8.8-d、ABE8.9-d、ABE8.10-d、ABE8.11-d、ABE8.12-d、ABE8.13-d、ABE8.14-d、ABE8.15-d、ABE8.16-d、ABE8.17-d、ABE8.18-d、ABE8.19-d、ABE8.20-d、ABE8.21-d、ABE8.22-d、ABE8.23-d或ABE8.24-d,如下表7所示。

[4002] 表7:ABE8碱基编辑器

[4003]

ABE8 碱基编辑器	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)
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)

[4004]

ABE8碱基编辑器	腺苷脱氨酶	腺苷脱氨酶描述
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)

[4005] 在一些实施方案中,碱基编辑器(例如,ABE8)是通过将腺苷脱氨酶变体(例如,TadA*8)克隆到包括环状置换的Cas9(例如,CP5或CP6)和二分核定位序列的支架中来产生。在一些实施方案中,碱基编辑器(例如,ABE7.9、ABE7.10或ABE8)是NGC PAM CP5变体(化脓链球菌Cas9或spVRQR Cas9)。在一些实施方案中,碱基编辑器(例如,ABE7.9、ABE7.10或ABE8)是AGA PAM CP5变体(化脓链球菌Cas9或spVRQR Cas9)。在一些实施方案中,碱基编辑器(例如,ABE7.9、ABE7.10或ABE8)是NGC PAM CP6变体(化脓链球菌Cas9或spVRQR Cas9)。在一些实施方案中,碱基编辑器(例如ABE7.9、ABE7.10或ABE8)是AGA PAM CP6变体(化脓链球菌Cas9或spVRQR Cas9)。

[4006] 在一些实施方案中,ABE具有如下表8所示的基因型。

[4007] 表8.ABE的基因型

[4008]

	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

[4009] 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。在一些实施方案中,ABE具有如下表9所示的ABE之一的基因型。

[4010] 表9.进化的TadA中的残基标识

[4011]

	23	36	48	51	76	82	84	106	108	123	146	147	152	154	155	156	157	166
ABE7.10	**R**	**L**	**A**	**L**	**I**	**V**	**F**	**V**	**N**	**Y**	**C**	**Y**	**P**	**Q**	**V**	**F**	**N**	**T**
ABE8.1-m												T						
ABE8.2-m												R						
ABE8.3-m														S				
ABE8.4-m										H								
ABE8.5-m						S												
ABE8.6-m																		R
ABE8.7-m														R				
ABE8.8-m										H		R		R				
ABE8.9-m					Y							R		R				
ABE8.10-m												R		R				R
ABE8.11-m												T		R				
ABE8.12-m												T		S				
ABE8.13-m					Y					H		R		R				
ABE8.14-m					Y	S												
ABE8.15-m						S						R						
ABE8.16-m						S				H		R						
ABE8.17-m						S								R				
ABE8.18-m						S				H				R				
ABE8.19-m						S				H		R		R				
ABE8.20-m					Y	S				H		R		R				
ABE8.21-m												R		S				
ABE8.22-m						S								S				
ABE8.23-m						S				H								
ABE8.24-m						S				H		T						
ABE8.1-d												T						
ABE8.2-d												R						
ABE8.3-d														S				
ABE8.4-d										H								

[4012]

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						

[4013] 在一些实施方案中,碱基编辑器是ABE8.1,其包括或基本上由以下具有腺苷脱氨酶活性的序列或其片段所组成:

[4014] ABE8.1_Y147T_CP5_NGC PAM_单体

[4015]

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG
LVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE
GILADECAALLCTFFRMPRQVFNAQKKAQSSTD***SGGSSGGSSGSETPGTSESATPESSGGSSGGS***EI
GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVK
KTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPTVAYSVLVVAKVEKGKSKKLKSVKEL
LGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAKFLQKGNELALPS
KYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYRSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD*GGSGGSGGSGGSGGSGGSGGM*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSI
KKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEED
KKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD
NSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIAL
SLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTE
ITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILT
FRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHS

LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVE
ISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDK
VMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQV
SGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER

[4016] MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFL
KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKA
GFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNY
HHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEGADKRTADGSEFESPKKKRK
V*

[4017] 以上序列中,纯文本表示腺苷脱氨酶序列,粗体序列表示衍生自Cas9的序列,斜体序列表示连接子序列,底线序列表示二分核定位序列。

[4018] 在一些实施方案中,碱基编辑器是ABE8.1,其包括或基本上由以下具有腺苷脱氨酶活性的序列或其片段所组成:

[4019] pNMG-B335 ABE8.1_Y147T_CP5_NGC PAM_单体

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG
LVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE
GILADECAALLCTFFRMPRQVFNAQKKAQSSTDSGGSSG*GSSGSETPGTSESATPESSGGSSGGS*EI
GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVK
KTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPTVAYSVLVVAKVEKGKSKKLKSVKEL
LGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAKFLQKGNELALPS
KYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYRSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD*GGSGGSGGSGGSGGSGGSGGM*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSI
KKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEED
KKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD

[4020] NSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIAL
SLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTE
ITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILT
FRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHS
LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVE
ISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDK
VMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQV
SGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER
MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFL
KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKA
GFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNY
HHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEGADKRTADGSEFESPKKKRK

[4021] V*

[4022] 以上序列中,纯文本表示腺苷脱氨酶序列,粗体序列表示衍生自Cas9的序列,斜体序列表示连接子序列,底线序列表示二分核定位序列。

[4023] 在一些实施方案中,碱基编辑器是ABE8.14,其包括或基本上由以下具有腺苷脱氨酶活性的序列或其片段所组成:

[4024] 具有NGC PAM的CP5pNMG-357_ABE8.14

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGG
LVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITE
GILADECAALLSDFFRMRRQEIKAQKKAQSSTDGGSSGGS*SGSETPGTSESATPESSGGSSGGS*MSE
VEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVM
QNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGIL
ADECAALLCTFFRMPRQVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGS***EIGKA**
TAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTE
VQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAKFLQKGNELALPSKYV
NFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDK
PIREQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYRSTKEVLDATLIHQSITGLYETRIDLSQLGG
D*GGSGGSGGSGGSGGSGGSGGM***DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKN**
LIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKH
[4025] **ERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD**
VDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLG
LTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITK
APLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILE
KMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLY
EYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISG
VEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMK
QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ
GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKR
IEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFI
KRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHA
HDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEGADKRTADGSEFESPKKKRKV*

[4026] 以上序列中,纯文本表示腺苷脱氨酶序列,粗体序列表示衍生自Cas9的序列,斜体序列表示连接子序列,底线序列表示二分核定位序列。

[4027] 在一些实施方案中,碱基编辑器是ABE8.8-m,其包括或基本上由以下具有腺苷脱氨酶活性的序列或其片段所组成:

[4028] ABE8.8-m

[4029]

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG
LVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITE
GILADECAALLCRFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGS*DK
KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR
RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH
LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS
GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY
DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL
VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET
ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF
LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF
LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK
QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT
VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ
NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT
KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF
VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD
KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYS
VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG
RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF
SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL
DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[4030] 以上序列中,纯文本表示腺苷脱氨酶序列,粗体序列表示衍生自Cas9的序列,斜体序列表示连接子序列,底线序列表示双核定位序列,及双底线序列表示突变。

[4031] 在一些实施方案中,碱基编辑器是ABE8.8-d,其包括或基本上由以下具有腺苷脱氨酶活性的序列或其片段所组成:

[4032] ABE8.8-d

[4033]

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGG
LVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITE
GILADECAALLSDFFRMRRQEIKAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGS*SE

[4034]

VEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVM
QNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGIL
ADECAALLCRFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGS*DKKYS
IGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYT
RRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRK
KLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVD
AKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDD
LDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQ
QLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG
SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP
WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSG
EQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDN
EENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV
VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEK
LYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVK
KMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD
ENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYG
DYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGR
DFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLV
VAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKR
MLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKR
VILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[4035] 以上序列中,纯文本表示腺苷脱氨酶序列,粗体序列表示衍生自Cas9的序列,斜体序列表示连接子序列,底线序列表示双核定位序列,及双底线序列表示突变。

[4036] 在一些实施方案中,碱基编辑器是ABE8.13-m,其包括或基本上由以下具有腺苷脱氨酶活性的序列或其片段所组成:

[4037] ABE8.13-m

[4038]

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG
LVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITE
GILADECAALLCRFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGS*DK
KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR
RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH
LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS
GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY
DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL

[4039]

VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET
ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF
LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF
LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK
QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT
VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ
NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT
KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF
VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD
KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYS
VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG
RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF
SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL
DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[4040] 以上序列中,纯文本表示腺苷脱氨酶序列,粗体序列表示衍生自Cas9的序列,斜体序列表示连接子序列,底线序列表示双核定位序列,双底线序列表示突变。

[4041] 在一些实施方案中,碱基编辑器是ABE8.13-d,其包括或基本上由以下具有腺苷脱氨酶活性的序列或其片段所组成:

[4042] ABE8.13-d

[4043]

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGG
LVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITE
GILADECAALLSDFFRMRRQEIKAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGS*SE
VEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVM
QNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGIL
ADECAALLCRFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGS***DKKYS**
IGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYT
RRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRK
KLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVD
AKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDD
LDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQ
QLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG
SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP
WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSG
EQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDN
EENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG

KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV
VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEK
LYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVK
KMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD
ENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYG

[4044] DYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGR
DFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLV
VAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKR
MLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKR
VILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[4045] 以上序列中,纯文本表示腺苷脱氨酶序列,粗体序列表示衍生自Cas9的序列,斜体序列表示连接子序列,底线序列表示双核定位序列,双底线序列表示突变。

[4046] 在一些实施方案中,碱基编辑器是ABE8.17-m,其包括或基本上由以下具有腺苷脱氨酶活性的序列或其片段所组成:

[4047] ABE8.17-m

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG
LVMQNYRLIDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE
GILADECAALLCYFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGS*DK
KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR
RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH
LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS
GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY
DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL
VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET

[4048] ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF
LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF
LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK
QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT
VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ
NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT
KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF
VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD
KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYS
VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG

RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF

[4049] **SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL**

DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[4050] 以上序列中,纯文本表示腺苷脱氨酶序列,粗体序列表示衍生自Cas9的序列,斜体序列表示连接子序列,底线序列表示双核定位序列,双底线序列表示突变。

[4051] 在一些实施方案中,碱基编辑器是ABE8.17-d,其包括或基本上由以下具有腺苷脱氨酶活性的序列或其片段所组成:

[4052] ABE8.17-d

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTA

HAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTG

AAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD*S*

*GGSSGGSSGSETPGTSESATPESSGGSSGGS*SEVEFSHEYWMRHALTLAKRARDER

EVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYS

TFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILA

DECAALLCYFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGG*

*SSGGS***DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALL**

FDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVE

EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKF

RGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSR

RLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLD

[4053] **NLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDL**

TLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEE

LLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKIL

TFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTN

RKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEEN

EDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLIN

GIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHI

ANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER

MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDY

DVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLI

TQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDK

LIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKL

ESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKR

PLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS

DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS

[4054] **SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELAL**

PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD

ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTK

EVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[4055] 以上序列中,纯文本表示腺苷脱氨酶序列,粗体序列表示衍生自Cas9的序列,斜体序列表示连接子序列,底线序列表示双核定位序列,双底线序列表示突变。

[4056] 在一些实施方案中,碱基编辑器是ABE8.20-m,其包括或基本上由以下具有腺苷脱氨酶活性的序列或其片段所组成:

[4057] ABE8.20-m

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAE

IMALRQGGLVMQNYRLYDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLM

DVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSG*

*SETPGTSESATPESSGGSSGGS***DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNT**

DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSF

FHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYL

ALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSAR

LSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDD

LDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT

LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVK

[4058] **LNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYY**

VGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPK

HSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYF

KKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFED

REMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG

FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDE

LVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQ

LQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNR

GKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVE

TRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYH

HAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSN

[4059] **IMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKS VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRY TSTKEVLDATLIHQSITGLYETRIDLSQLGGD**EGADKRTADGSEFESPKKKRKV*

[4060] 以上序列中,纯文本表示腺苷脱氨酶序列,粗体序列表示衍生自Cas9的序列,斜体序列表示连接子序列,底线序列表示双核定位序列,双底线序列表示突变。

[4061] 在一些实施方案中,碱基编辑器是ABE8.20-d,其包括或基本上由以下具有腺苷脱氨酶活性的序列或其片段所组成:

[4062] ABE8.20-d

[4063] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTA HAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTG AAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD*S GGSSGGSSGSETPGTSESATPESSGGSSGGS*SEVEFSHEYWMRHALTLAKRARDER EVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYS TFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILA DECAALLCRFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGG SSGGS***DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALL FDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVE EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKF RGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSR RLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLD NLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDL TLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEE LLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKIL TFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTN RKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEEN EDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLIN GIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHI ANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDY

[4064] **DVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLI TQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDK LIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKL ESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKR PLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELAL PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTK EVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV***

[4065] 以上序列中,纯文本表示腺苷脱氨酶序列,粗体序列表示衍生自Cas9的序列,斜体序列表示连接子序列,底线序列表示双核定位序列,双底线序列表示突变。

[4066] 在一些实施方案中,本发明的ABE8是选自由以下序列:

[4067] 01.monoABE8.1_bpNLS+Y147T

[4068] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG

[4069] LVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE

[4070] GILADECAALLCTFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDK

[4071] KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

[4072] RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH

[4073] LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

[4074] GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY

[4075] DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL

[4076] VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF

[4077] DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET

[4078] ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF

[4079] LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF

[4080] LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

[4081] QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT

[4082] VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ

[4083] NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE

[4084] VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT

[4085] KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF

[4086] VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD

[4087] KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS

[4088] VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG

[4089] RKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF

[4090] SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL

[4091] DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[4092] 02.monoABE8.1_bpNLS+Y147R

[4093] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG

[4094] LVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE

[4095] GILADECAALLCRFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDK

[4096] KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

[4097] RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH

[4098] LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

[4099] GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY

[4100] DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL

[4101] VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF

[4102] DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET

[4103] ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF

[4104] LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF

[4105] LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

[4106] QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT

[4107] VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ

[4108] NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE

[4109] VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT

[4110] KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF

[4111] VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD

[4112] KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS

[4113] VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG

[4114] RKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF

[4115] SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL

[4116] DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[4117] 03.monoABE8.1_bpNLS+Q154S

[4118] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG

[4119] LVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE

[4120] GILADECAALLCYFFRMPRSVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDK

[4121] KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

[4122] RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH

[4123] LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

[4124] GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY

[4125] DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL

[4126] VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF

[4127] DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET

[4128] ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF

[4129] LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF

[4130] LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

[4131] QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT

[4132] VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ

[4133] NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE

[4134] VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT

[4135] KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF

[4136] VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD

[4137] KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS

[4138] VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG

[4139] RKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF

[4140] SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL

[4141] DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[4142] 04.monoABE8.1_bpNLS+Y123H

[4143] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG

[4144] LVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITE

[4145] GILADECAALLCYFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDK

[4146] KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

[4147] RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH

[4148] LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

[4149] GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY

[4150] DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL

[4151] VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF

[4152] DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET

[4153] ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF

[4154] LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF

[4155] LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

[4156] QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT

[4157] VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ

[4158] NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE

[4159] VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT

[4160] KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF

[4161] VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD

[4162] KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS

[4163] VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG

[4164] RKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF

[4165] SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL

[4166] DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[4167] 05.monoABE8.1_bpNLS+V82S

[4168] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG

[4169] LVMQNYRLIDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE

[4170] GILADECAALLCYFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDK

[4171] KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

[4172] RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH

[4173] LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

[4174] GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY

[4175] DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL

[4176] VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF

[4177] DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET

[4178] ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF

[4179] LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF

[4180] LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

[4181] QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT

[4182] VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ

[4183] NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE

[4184] VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT

[4185] KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF

[4186] VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD

[4187] KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS

[4188] VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG

[4189] RKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF

[4190] SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL

[4191] DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[4192] 06.monoABE8.1_bpNLS+T166R

[4193] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG

[4194] LVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE

[4195] GILADECAALLCYFFRMPRQVFNAQKKAQSSRDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDK

[4196] KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

[4197] RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH

[4198] LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

[4199] GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY

[4200] DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL

[4201] VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF

[4202] DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET

[4203] ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF

[4204] LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF

[4205] LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

[4206] QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT

[4207] VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ

[4208] NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE

[4209] VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT

[4210] KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF

[4211] VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD

[4212] KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS

[4213] VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG

[4214] RKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF

[4215] SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL

[4216] DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[4217] 07.monoABE8.1_bpNLS+Q154R

[4218] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG

[4219] LVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE

[4220] GILADECAALLCYFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDK

[4221] KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

[4222] RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH

[4223] LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

[4224] GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY

[4225] DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL

[4226] VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF

[4227] DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET

[4228] ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF

[4229] LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF

[4230] LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

[4231] QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT

[4232] VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ

[4233] NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE

[4234] VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT

[4235] KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF

[4236] VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD

[4237] KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS

[4238] VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG

[4239] RKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF

[4240] SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL

[4241] DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[4242] 08.monoABE8.1_bpNLS+Y147R_Q154R_Y123H

[4243] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG

[4244] LVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITE

[4245] GILADECAALLCRFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDK

[4246] KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

[4247] RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH

[4248] LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

[4249] GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY

[4250] DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL

[4251] VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF

[4252] DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET

[4253] ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF

[4254] LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF

[4255] LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

[4256] QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT

[4257] VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ

[4258] NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE

[4259] VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT

[4260] KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF

[4261] VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD

[4262] KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS

[4263] VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG

[4264] RKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF

[4265] SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL

[4266] DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[4267] 09.monoABE8.1_bpNLS+Y147R_Q154R_I76Y

[4268] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG

[4269] LVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE

[4270] GILADECAALLCRFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDK

[4271] KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

[4272] RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH

[4273] LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

[4274] GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY

[4275] DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL

[4276] VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF

[4277] DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET

[4278] ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF

[4279] LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF

[4280] LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

[4281] QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT

[4282] VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ

[4283] NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE

[4284] VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT

[4285] KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF

[4286] VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD

[4287] KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS

[4288] VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG

[4289] RKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF

[4290] SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL

[4291] DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[4292] 10.monoABE8.1_bpNLS+Y147R_Q154R_T166R

[4293] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG

[4294] LVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE

[4295] GILADECAALLCRFFRMPRRVFNAQKKAQSSRDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDK

[4296] KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

[4297] RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH

[4298] LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

[4299] GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY

[4300] DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL

[4301] VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF

[4302] DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET

[4303] ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF

[4304] LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF

[4305] LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

[4306] QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT

[4307] VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ

[4308] NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE

[4309] VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT

[4310] KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF

[4311] VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD

[4312] KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS

[4313] VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG

[4314] RKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF

[4315] SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL

[4316] DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[4317] 11.monoABE8.1_bpNLS+Y147T_Q154R

[4318] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG

[4319] LVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE

[4320] GILADECAALLCTFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDK

[4321] KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

[4322] RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH

[4323] LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

[4324] GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY

[4325] DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL

[4326] VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF

[4327] DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET

[4328] ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF

[4329] LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF

[4330] LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

[4331] QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT

[4332] VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ

[4333] NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE

[4334] VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT

[4335] KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF

[4336] VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD

[4337] KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS

[4338] VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG

[4339] RKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF

[4340] SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL

[4341] DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[4342] 12.monoABE8.1_bpNLS+Y147T_Q154S

[4343] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG

[4344] LVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE

[4345] GILADECAALLCTFFRMPRSVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDK

[4346] KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

[4347] RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH

[4348] LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

[4349] GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY

[4350] DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL

[4351] VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF

[4352] DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET

[4353] ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF

[4354] LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF

[4355] LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

[4356] QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT

[4357] VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ

[4358] NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE

[4359] VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT

[4360] KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF

[4361] VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD

[4362] KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS

[4363] VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG

[4364] RKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF

[4365] SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL

[4366] DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[4367] 13.monoABE8.1_bpNLS+H123Y123H_Y147R_Q154R_I76Y

[4368] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLC
RFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD
SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[4369] 14.monoABE8.1_bpNLS+V82S+Q154R

[4370] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG

[4371] LVMQNYRLIDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE

[4372] GILADECAALLCYFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDK

[4373] KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

[4374] RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH

[4375] LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

[4376] GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY

[4377] DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL

[4378] VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF

[4379] DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET

[4380] ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF

[4381] LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF

[4382] LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

[4383] QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT

[4384] VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ

[4385] NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE

[4386] VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT

[4387] KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF

[4388] VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD

[4389] KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS

[4390] VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG

[4391] RKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF

[4392] SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL

[4393] DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[4394] 在一些实施方案中,碱基编辑器是包括与核碱基编辑结构域(例如,脱氨酶结构域的全部或部分)融合的多核苷酸可编程的核苷酸结合结构域(例如,Cas9源域)的融合蛋白。在某些实施方案中,本文所提供的融合蛋白包括一种或多种改善融合蛋白的碱基编辑活性的特征。例如,本文所提供的任何融合蛋白可以包括具有降低的核酸酶活性的Cas9结构域。在一些实施方案中,本文所提供的任何融合蛋白可具有不具有核酸酶活性的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))。

[4395] 在一些实施方案中,碱基编辑器进一步包括包含全部或部分尿嘧啶糖基化酶抑制剂(UGI)的结构域。在一些实施方案中,碱基编辑器包括包含尿嘧啶结合蛋白(UBP),诸如尿嘧啶DNA糖基化酶(UDG)的全部或部分的结构域。在一些实施方案中,碱基编辑器包括包含全部或部分核酸聚合酶的结构域。在一些实施方案中,掺入碱基编辑器的核酸聚合酶或其部分是跨损伤DNA聚合酶。

[4396] 在一些实施方案中,碱基编辑器的域可以包括多个结构域。例如,包括衍生自Cas9的多核苷酸可编程的核苷酸结合结构域的碱基编辑器可以包括对应于野生型或天然Cas9的REC叶和NUC叶的REC叶和NUC叶。在另一个实例中,碱基编辑器可包括RuvCI结构域、BH结构域、REC1结构域、REC2结构域、RuvCII结构域、L1结构域、HNH结构域、L2结构域、RuvCIII结构域、WED结构域、TOPO结构域或CTD结构域中的一个或多个。在一些实施方案中,碱基编辑器的一个或多个结构域包括相对于包括所述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,多核苷酸可编程的DNA结合结构域的HNH结构域可包括H840A取代。在另一个实例中,多核苷酸可编程的DNA结合结构域的RuvCI域可包括D10A取代。

[4397] 在使用或不使用一个或多个连接子结构域(例如,XTEN连接子结构域)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。在一些实施方案中,连接子结构域可以是键(例如共价键)、化学基团或连接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,衍生自Cas9的结构域)和第二结构域(例如,腺苷脱氨酶结构域)。在一些实施方案中,连接子是共价键(例如,碳-碳键、二硫键、碳-杂原子键等)。在某些实施方案中,连接子是酰胺键的碳氮键。在某些实施方案中,连接子是环状或无环、经取代或未经取代、支链或无支链的脂族或杂脂族连接子。在某些实施方案中,连接子是聚合的(例如,聚乙烯、聚乙二醇、聚酰胺、聚酯等)。在某些实施方案中,连接子包括氨基链烷酸的单体、二聚体或聚合物。在一些实施方案中,连接子包括氨基链烷酸(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)。在一些实施方案中,连

接子包括氨基己酸(Ahx)的单体、二聚体或聚合物。在某些实施方案中,连接子是基于碳环部分(例如,环戊烷、环己烷)。在其他实施方案中,连接子包括聚乙二醇部分(PEG)。在某些实施方案中,连接子包括芳基或杂芳基部分。在某些实施方案中,连接子基于苯环。连接子可以包括功能化部分以促进亲核试剂(例如,硫醇、氨基)从肽连接到连接子。任何亲电子试剂都可以用作连接子的一部分。实施例性的亲电子试剂包括但不限于激活酯、激活酰胺、迈克尔受体(Michael receptor)、卤代烷、芳基卤、酰基卤和异硫氰酸酯。在一些实施方案中,连接子连接RNA可编程的核酸酶的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域。在一些实施方案中,连接子连接dCas9和第二结构域(例如,UGI等)。

[4398] 通常,连接子位于两个基团、分子或其他部分之间或侧边,并经由共价键与每个基团、分子或其他部分连接,从而将两者连接起来。在一些实施方案中,连接子是一个氨基酸或多个氨基酸(例如,肽或蛋白质)。在一些实施方案中,连接子是有机分子、基团、聚合物或化学部分。在一些实施方案中,连接子的长度为2至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至35、35至40、40至45、45至50、50至60、60至70、70至80、80至90、90至100、100至150或150至200个氨基酸的长度。在一些实施方案中,连接子为约3至约104(例如,5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、48、49、50、55、60、65、70、75、80、85、90、95或100)个氨基酸长度。更长或更短的连接子也被考虑。在一些实施方案中,连接子域包括氨基酸序列SGSETPGTSESATPES,其也可称为XTEN连接子。可以使用任何连接融合蛋白结构域的方法(例如,从柔性的(SGGS)n、(GGGS)n、(GGGGS)n及(G)n形式的连接子到更刚性的连接子形式(EAAAK)n、(GGS)n、SGSETPGTSESATPES(参见,例如Guilinger JP,Thompson DB,Liu DR.Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification.Nat.Biotechnol.2014;32(6):577-82;其全部内容通过引用方式并入本文)或$(XP)_n$基序,以便实现核碱基编辑器活性的最佳长度。在一些实施方案中,n是1、2、3、4、5、6、7、8、9、10、11、12、13、14或15。在一些实施方案中,连接子包括(GGS)n基序,其中n是1、3或7。在一些实施方案中,本文所提供的融合蛋白的Cas9结构域经由包括氨基酸序列SGSETPGTSESATPES的连接子融合。在一些实施方案中,连接子包括多个脯氨酸残基并且长度是5至21、5至14、5至9、5至7种氨基酸,例如PAPAP、PAPAPA、PAPAPAP、PAPAPAPA、$P(AP)_4$、$P(AP)_7$、$P(AP)_{10}$(参见,例如Tan J,Zhang F,Karcher D,Bock R.Engineering of high-precision base editors for site-specific single nucleotide replacement.Nat Commun.2019Jan25;10(1):439;其全部内容通过引用方式并入本文)。这种富含脯氨酸的连接子也称为“刚性”连接子。

[4399] 本发明的融合蛋白包括核酸编辑结构域。在一些实施方案中,脱氨酶是腺苷脱氨酶。在一些实施方案中,脱氨酶是脊椎动物脱氨酶。在一些实施方案中,脱氨酶是无脊椎动物脱氨酶。在一些实施方案中,脱氨酶是人类、黑猩猩、大猩猩、猴、牛、狗、大鼠或小鼠脱氨酶。在一些实施方案中,脱氨酶是人类脱氨酶。在一些实施方案中,脱氨酶是大鼠脱氨酶。

[4400] 如本文所用,“异源二聚体”可以指包括野生型TadA结构域和变异TadA*7.10结构域或两个变异TadA结构域(例如,具有Y147T和Q154S改变的TadA7.10和TadA7.10)的融合蛋白改动。

[4401] 在一些实施方案中,碱基编辑器包括融合蛋白,所述融合蛋白包括内部融合或插入napDNAbp的异源多肽。异源多肽(例如,脱氨酶)可以在合适的位置插入napDNAbp(例如,Cas9),例如,使得napDNAbp保持其结合目标多核苷酸和向导核酸的能力。脱氨酶可以插入napDNAbp而不损害脱氨酶的功能(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)。如晶体学研究所示,脱氨酶可以插入napDNAbp,例如无序区域或包括高温因子或B因子的区域。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。脱氨酶可以插入napDNAbp的柔性环区域或溶剂暴露区域中。在一些实施方案中,将脱氨酶插入Cas9多肽的柔性环。

[4402] 在一些实施方案中,脱氨酶的插入位置通过Cas9多肽的晶体结构的B因子分析确定。在一些实施方案中,脱氨酶插入Cas9多肽的包括高于平均B因子的区域中(例如,与包括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)。骨架原子的高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。脱氨酶可以插入具有Cα原子的的残基的位置,所述位置具有B因子比总蛋白质的平均B因子高50%、60%、70%、80%、90%、100%、110%、120%、130%、140%、150%、160%、170%、180%、190%、200%或大于200%。脱氨酶可以插入具有Cα原子的残基的位置,所述位置具有B因子比包括残基的Cas9蛋白结构域的平均B因子高50%、60%、70%、80%、90%、100%、110%、120%、130%、140%、150%、160%、170%、180%、190%、200%或超过200%。包括高于平均B因子的Cas9多肽位置可以包括,例如,如编号为SEQ ID No:1中编号为残基768、792、1052、1015、1022、1026、1029、1067、1040、1054、1068、1246、12248。包括高于平均B因子的Cas9多肽区域可包括例如SEQ ID No:1中编号为残基792至872、792至906及2至791。

[4403] 异源多肽(例如脱氨酶)可以插入napDNAbp的氨基酸残基处,所述氨基酸残基是选自由以下所组成的群组:在SEQ ID NO:1中编号为768、791、792、1015、1016、1022、1023、1026、1029、1040、1052、1054、1067、1068、1069、1246、1247及1248,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,异源多肽插入在SEQ ID NO:1中编号为氨基酸位置768至769、791至792、792至793、1015至1016、1022至1023、1026至1027、1029至1030、1040至1041、1052至1053、1067至1068、1068至1069、1247至1248或1248至1249如或其相应的氨基酸位置。在一些实施方案中,异源多肽插入在在SEQ ID NO:1中编号为氨基酸位置769至770、792至793、793至794、1016至1017、1023至1024、1027至1028、1030至1031、1041至1041、1052至1053、1055至1056、1068至1069、1069至1070、1248至1249、或1249至1250、1068至1069、1069至1070、1248至1249或1249至1250或其相应的氨基酸位置之间。在一些实施方案中,异源多肽置换选自由以下所组成的群组的氨基酸残基:在SEQ ID NO:1中编号为768、791、792、1015、1016、1022、1023、1026、1029、1040、1052、1054、1067、1068、1069、1246、1247和1248,或另一个Cas9多肽中的相应的氨基酸残基。应当理解,关于插入位置对SEQ ID NO:1的引用是为了说明的目的。本文所讨论的插入不限于SEQ ID NO:1的Cas9多肽序列,而是包括在变异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。

[4404] 异源多肽(例如,脱氨酶)可以插入napDNAbp中的氨基酸残基处,所述氨基酸残基选自以下所组成的群组:在SEQ ID NO:1中编号为768、792、1022、1026、1040、1068和1247,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,异源多肽插入在SEQ ID NO:1中编号为768至769、792至793、1022至1023、1026至1027、1029至1030、1040至1041、1068至1069或1248至1249或其相应的氨基酸位置之间。在一些实施方案中,异源多肽插入在SEQ ID NO:1编号为SEQ中的第769至770、793至794、1023至1024、1027至1028、1030至1031、1041至1042、1069至1070或1248至1249或其相应的氨基酸位置之间。在一些实施方案中,异源多肽置换选自由以下所组成的群组的氨基酸残基:在SEQ ID NO:1中编号为768、792、1022、1026、1040、1068和1247,或另一个Cas9中的相应的氨基酸残基Cas9。

[4405] 在一些实施方案中,ABE(例如,TadA)插入在选自由以下所组成的群组的氨基酸残基处:在SEQ ID NO:1中编号为1015、1022、1029、1040、1068、1247、1054、1026、768、1067、1248、1052及1246,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,ABE(例如,TadA)被插入以代替在SEQ ID NO:1中编号为残基792至872、792至906或2至791,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,ABE插入在选自由以下所组成的群组的氨基酸的N端:在SEQ ID NO:1中编号为1015、1022、1029、1040、1068、1247、1054、1026、768、1067、1248、1052及1246,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,ABE插入在选自由以下所组成的群组的氨基酸的C端:在SEQ ID NO:1中编号为1015、1022、1029、1040、1068、1247、1054、1026、768、1067、1248、1052及1246,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,插入ABE以替换选自由以下所组成的群组的氨基酸:在SEQ ID NO:1中编号为1015、1022、1029、1040、1068、1247、1054、1026、768、1067、1248、1052及1246,或另一个Cas9多肽中的相应的氨基酸残基。

[4406] 在一些实施方案中,CBE(例如,APOBEC1)插入在选自由以下所组成的群组的氨基酸残基处:如在SEQ ID NO:1中编号为1016、1023、1029、1040、1069及1247,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,ABE插入在选自由以下所组成的群组的氨基酸的N端:在SEQ ID NO:1中编号为1016、1023、1029、1040、1069及1247,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,ABE插入在选自由以下所组成的群组的氨基酸的C端:如在SEQ ID NO:1中编号为1016、1023、1029、1040、1069和1247,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,插入ABE以替换选自由以下所组成的群组的氨基酸:如在SEQ ID NO:1中编号为1016、1023、1029、1040、1069及1247,或另一个Cas9多肽中的相应的氨基酸残基。

[4407] 在一些实施方案中,脱氨酶插入在SEQ ID NO:1中编号为氨基酸残基768处,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,ABE插入在SEQ ID NO:1中编号为氨基酸768的N端,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,ABE插入在SEQ ID NO:1中编号为氨基酸768的C端,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,插入ABE以置换在SEQ ID NO:1中编号为氨基酸768处,或另一个Cas9多肽中的相应的氨基酸残基。

[4408] 在一些实施方案中,脱氨酶插入在SEQ ID NO:1中编号为氨基酸残基791处,或另一个Cas9多肽中的相应的氨基酸残基处。在一些实施方案中,ABE插入在SEQ ID NO:1中编号为氨基酸791,或另一个Cas9多肽中的相应的氨基酸残基的N端。在一些实施方案中,ABE

插入在SEQ ID NO:1中编号为氨基酸791的C端,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,插入ABE以置换在SEQ ID NO:1中编号为氨基酸791处,或另一个Cas9多肽中的相应的氨基酸残基。

[4409] 在一些实施方案中,脱氨酶插入在SEQ ID NO:1中编号为氨基酸残基792处,或另一个Cas9多肽中的相应的氨基酸残基处。在一些实施方案中,ABE插入在SEQ ID NO:1中编号为氨基酸792的N端,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,ABE插入在SEQ ID NO:1中编号为氨基酸792的C端,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,插入ABE以置换在SEQ ID NO:1中编号为氨基酸792处,或另一个Cas9多肽中的相应的氨基酸残基。

[4410] 在一些实施方案中,脱氨酶插入在SEQ ID NO:1中编号为氨基酸残基1016处,或另一个Cas9多肽中的相应的氨基酸残基处。在一些实施方案中,ABE插入在SEQ ID NO:1中编号为氨基酸1016的N端,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,ABE插入在SEQ ID NO:1中编号为氨基酸1016的C端,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,插入ABE以置换在SEQ ID NO:1中编号为氨基酸1016处,或另一个Cas9多肽中的相应的氨基酸残基。

[4411] 在一些实施方案中,脱氨酶插入在SEQ ID NO:1中编号为氨基酸残基1022处,或另一个Cas9多肽中的相应的氨基酸残基处。在一些实施方案中,ABE插入在SEQ ID NO:1中编号为氨基酸1022的N端,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,ABE插入在SEQ ID NO:1中编号为氨基酸1022的C端,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,插入ABE以置换在SEQ ID NO:1中编号的氨基酸1022,或另一个Cas9多肽中的相应的氨基酸残基。

[4412] 在一些实施方案中,脱氨酶插入在SEQ ID NO:1中编号为氨基酸残基1023处,或另一个Cas9多肽中的相应的氨基酸残基处。在一些实施方案中,ABE插入在SEQ ID NO:1中编号为氨基酸1023的N端,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,ABE插入在SEQ ID NO:1中编号为氨基酸1023的C端,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,插入ABE以置换在SEQ ID NO:1中编号的氨基酸1023处,或另一个Cas9多肽中的相应的氨基酸残基。

[4413] 在一些实施方案中,脱氨酶插入在SEQ ID NO:1中编号为氨基酸残基1026处,或另一个Cas9多肽中的相应的氨基酸残基处。在一些实施方案中,ABE插入在SEQ ID NO:1中编号为氨基酸1026的N端,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,ABE插入在SEQ ID NO:1中编号为氨基酸1026的C端,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,插入ABE以置换在SEQ ID NO:1中编号为氨基酸1026处,或另一个Cas9多肽中的相应的氨基酸残基。

[4414] 在一些实施方案中,脱氨酶插入在SEQ ID NO:1中编号为氨基酸残基1029处,或另一个Cas9多肽中的相应的氨基酸残基处。在一些实施方案中,ABE插入在SEQ ID NO:1中编号为氨基酸1029的N端,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,ABE插入在SEQ ID NO:1中编号为氨基酸1029的C端,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,插入ABE以置换在SEQ ID NO:1中编号为氨基酸1029处,或另一个Cas9多肽中的相应的氨基酸残基。

[4415] 在一些实施方案中,脱氨酶插入在SEQ ID NO:1中编号为氨基酸残基1040处,或另一个Cas9多肽中的相应的氨基酸残基处。在一些实施方案中,ABE插入在SEQ ID NO:1中编号为氨基酸140的N端,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,ABE插入在SEQ ID NO:1中编号为氨基酸1040的C端,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,插入ABE以置换在SEQ ID NO:1中编号为氨基酸1040处,或另一个Cas9多肽中的相应的氨基酸残基。

[4416] 在一些实施方案中,脱氨酶插入在SEQ ID NO:1中编号为氨基酸残基1052处,或另一个Cas9多肽中的相应的氨基酸残基处。在一些实施方案中,ABE插入在SEQ ID NO:1中编号为氨基酸1052的N端,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,ABE插入在SEQ ID NO:1中编号为氨基酸1052的C端,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,插入ABE以置换在SEQ ID NO:1中编号为氨基酸1052处,或另一个Cas9多肽中的相应的氨基酸残基。

[4417] 在一些实施方案中,脱氨酶插入在SEQ ID NO:1中编号为氨基酸残基1054处或另一个Cas9多肽中的相应的氨基酸残基处。在一些实施方案中,ABE插入在SEQ ID NO:1中编号为氨基酸1054的N端,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,ABE插入在SEQ ID NO:1中编号为氨基酸1054的C端,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,插入ABE以置换在SEQ ID NO:1中编号为氨基酸1054处,或另一个Cas9多肽中的相应的氨基酸残基。

[4418] 在一些实施方案中,脱氨酶插入在SEQ ID NO:1中编号为氨基酸残基1067处,或另一个Cas9多肽中的相应的氨基酸残基处。在一些实施方案中,ABE插入在SEQ ID NO:1中编号为氨基酸1067的N端,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,ABE插入在SEQ ID NO:1中编号为氨基酸1067的C端,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,插入ABE以置换在SEQ ID NO:1中编号为氨基酸1067处,或另一个Cas9多肽中的相应的氨基酸残基。

[4419] 在一些实施方案中,脱氨酶插入在SEQ ID NO:1中编号为氨基酸残基1068处,或另一个Cas9多肽中的相应的氨基酸残基处。在一些实施方案中,ABE插入在SEQ ID NO:1中编号为氨基酸1068的N端,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,ABE插入在SEQ ID NO:1中编号为氨基酸1068的C端,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,插入ABE以置换在SEQ ID NO:1中编号为氨基酸1068处,或另一个Cas9多肽中的相应的氨基酸残基。

[4420] 在一些实施方案中,脱氨酶插入在SEQ ID NO:1中编号为氨基酸残基1069处,或另一个Cas9多肽中的相应的氨基酸残基处。在一些实施方案中,ABE插入在SEQ ID NO:1中编号为氨基酸1069的N端,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,ABE插入在SEQ ID NO:1中编号为氨基酸1069的C端,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,插入ABE以置换在SEQ ID NO:1中编号为氨基酸1069处,或另一个Cas9多肽中的相应的氨基酸残基。

[4421] 在一些实施方案中,脱氨酶插入在SEQ ID NO:1中编号为氨基酸残基1246处,或另一个Cas9多肽中的相应的氨基酸残基处。在一些实施方案中,ABE插入在SEQ ID NO:1中编号为氨基酸1246的N端,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,ABE

插入在SEQ ID NO:1中编号为氨基酸1246的C端,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,插入ABE以置换在SEQ ID NO:1中编号为氨基酸1246处,或另一个Cas9多肽中的相应的氨基酸残基。

[4422] 在一些实施方案中,脱氨酶插入在SEQ ID NO:1中编号为氨基酸残基1247处,或另一个Cas9多肽中的相应的氨基酸残基处。在一些实施方案中,ABE插入在SEQ ID NO:1中编号为氨基酸1247的N端,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,ABE插入在SEQ ID NO:1中编号为氨基酸1247的C端,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,插入ABE以置换在SEQ ID NO:1中编号为氨基酸1247处,或另一个Cas9多肽中的相应的氨基酸残基。

[4423] 在一些实施方案中,脱氨酶插入在SEQ ID NO:1中编号为氨基酸残基1248处,或另一个Cas9多肽中的相应的氨基酸残基处。在一些实施方案中,ABE插入在SEQ ID NO:1中编号为氨基酸1248或另一个Cas9多肽中的相应的氨基酸残基的N端。在一些实施方案中,ABE插入在SEQ ID NO:1中编号为氨基酸1248的C端,或另一个Cas9多肽中的相应的氨基酸残基。在一些实施方案中,插入ABE以置换在SEQ ID NO:1中编号为氨基酸1248处,或另一个Cas9多肽中的相应的氨基酸残基。

[4424] 在一些实施方案中,异源多肽(例如,脱氨酶)被插入Cas9多肽的柔性环中。柔性环部分可以选自如SEQ ID NO:1中编号为530至537、569至570、686至691、943至947、1002至1025、1052至1077、1232至1247、或1298至1300所组成的群组,或另一个Cas9多肽中的相应的氨基酸残基。柔性环部分可以选自由以下所组成的群组:在SEQ ID NO:1中编号为1至529、538至568、580至685、692至942、948至1001、1026至1051、1078至1231或1248至1297,或另一个Cas9多肽中的相应的氨基酸残基。

[4425] 可以将异源多肽(例如脱氨酶)插入对应于以下氨基酸残基的Cas9多肽区域:在SEQ ID NO:1中编号为1017至1069、1242至1247、1052至1056、1060至1077、1002至1003、943至947、530至537、568至579、686至691、1242至1247、1298至1300、1066至1077、1052至1056或1066至1077,或另一个Cas9多肽中的相应的氨基酸残基。

[4426] 可以插入异源多肽(例如脱氨酶)代替Cas9多肽的缺失区域。缺失区域可以对应于Cas9多肽的N端或C端部分。在一些实施方案中,缺失区域对应于在SEQ ID NO:1中编号为残基792至872,或另一个Cas9多肽中的对应氨基酸残基。在一些实施方案中,缺失区对应于在SEQ ID NO:1中编号为残基792至906,或另一个Cas9多肽中的对应氨基酸残基。在一些实施方案中,缺失区域对应于在SEQ ID NO:1中编号为残基2至791,或另一个Cas9多肽中的对应氨基酸残基。

[4427] 异源多肽(例如,脱氨酶)可以插入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。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。

[4428] 在一些实施方案中,Cas9多肽缺乏一个或多个选自由以下所组成的群组的结构域:RuvC I、RuvC II、RuvC III、Rec1、Rec2、PI或HNH结构域。在一些实施方案中,Cas9多肽缺乏核酸酶结构域。在一些实施方案中,Cas9多肽缺乏HNH结构域。在一些实施方案中,Cas9多肽缺少HNH结构域的一部分,使得Cas9多肽具有降低或消除的HNH活性。

[4429] 在一些实施方案中,Cas9多肽包括核酸酶结构域的缺失并且插入脱氨酶以替代核酸酶结构域。在一些实施方案中,HNH结构域被删除并且脱氨酶被插入在其位置上。在一些实施方案中,删除一个或多个RuvC结构域,并且脱氨酶被插入在其位置上。

[4430] napDNAbp的N端和C端片段可在包括异源多肽的融合蛋白的侧边。在一些实施方案中,融合蛋白包括N端片段和C端片段在侧边脱氨酶的Cas9多肽。N端片段或C端片段可以结合目标多核苷酸序列。N端片段的C端或C端片段的N端可包括Cas9多肽的柔性环的一部分。N端片段的C端或C端片段的N端可包括Cas9多肽的α-螺旋结构的一部分。N端片段或C端片段可包括DNA结合结构域。N端片段或C端片段可包括RuvC结构域。N端片段或C端片段可包括HNH结构域。在一些实施方案中,N端片段和C端片段均不包括HNH结构域。

[4431] 在一些实施方案中,当融合蛋白使目标核碱基脱氨时,N端Cas9片段的C端包括接近目标核碱基的氨基酸。在一些实施方案中,当融合蛋白使目标核碱基脱氨时,C端Cas9片段的N端包括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,以使目标核碱基与N端Cas9片段的C端或C端Cas9片段的N端的氨基酸接近。例如,ABE的插入位置可以在选自由以下所组成的群组的氨基酸残基处:在SEQ ID NO:1中编号为1015、1022、1029、1040、1068、1247、1054、1026、768、1067、1248、10546,或另一个Cas9多肽中的相应的氨基酸残基。CBE的合适插入位置可以是选自由以下所组成的群组的氨基酸残基:在SEQ ID NO:1中编号为1016、1023、1029、1040、1069和1247,或另一个Cas9中的相应的氨基酸残基多肽。在某些实施方案中,ABE的插入可以插入以上列出的氨基酸残基中的任一个的N端或C端。在一些实施方案中,可以插入ABE的插入以替换以上列出的氨基酸残基中的任何一个。

[4432] 融合蛋白的N端Cas9片段(即在融合蛋白中的脱氨酶侧边N端Cas9片段)可包括Cas9多肽的N端。融合蛋白的N端Cas9片段可包括至少约:100、200、300、400、500、600、700、800、900、1000、1100、1200或1300个氨基酸的长度。融合蛋白的N端Cas9片段可包括对应于以下氨基酸残基的序列:在SEQ ID NO:1中编号为1至56、1至95、1至200、1至300、1至400、1至500、1至600、1至700、1至718、1至765、1至780、1至906、1至918或1至1100,或另一个Cas9多肽中的相应的氨基酸残基。N端Cas9片段可包括的序列包括与在SEQ ID NO:1中编号为氨基酸残基:1至56、1至95、1至200、1至300、1至400、1至500、1至600、1至700、1至718、1至765、1至780、1至906、1至918或1至1100或在另一个Cas9多肽中的相应的氨基酸残基至少: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%的序列同一性与。

[4433] 融合蛋白的C端Cas9片段(在融合蛋白中的脱氨酶侧边N端Cas9片段)可包括Cas9多肽的C端。融合蛋白的C端Cas9片段可包括至少约:100、200、300、400、500、600、700、800、900、1000、1100、1200或1300个氨基酸的长度。融合蛋白的C端Cas9片段可包括对应于以下氨基酸残基的序列:在SEQ ID NO:1中编号为1099至1368、918至1368、906至1368、780至1368、765至1368、718至1368、94至1368或56至1368或另一个Cas9多肽中的相应的氨基酸残基。N端Cas9片段可包括的序列包括与在SEQ ID NO:1中编号为氨基酸残基:1099至1368、918至1368、906至1368、780至1368、765至1368、718至1368、94至1368或56至1368或另一个Cas9多肽中的相应的氨基酸残基至少: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%的序列同一性。

[4434] 融合蛋白的N端Cas9片段和C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,在SEQ ID NO:1中所示者。

[4435] 本文所述融合蛋白可以通过减少非目标位点(例如,脱靶位点)的脱氨,例如减少全基因组的假脱氨来实现靶向脱氨。本文所述融合蛋白可以在非目标位点处以减少的旁观者脱氨实现靶向脱氨。不希望的脱氨或脱靶脱氨可以减少至少30%、至少40%、至少50%、至少60%、至少70%、至少80%、至少90%、至少95%或至少99%与例如包括与Cas9多肽的N端或C端融合的脱氨酶的末端融合蛋白相比。不需要的脱氨或脱靶脱氨可以减少至少一倍、至少两倍、至少三倍、至少四倍、至少五倍、至少十倍、至少十五倍、至少二十倍、至少三十倍、至少四十倍、至少五十倍、至少60倍、至少70倍、至少80倍、至少90倍或至少一百倍,相比之下,例如,包括与Cas9多肽的N端或C端融合的脱氨酶的末端融合蛋白。

[4436] 在一些实施方案中,融合蛋白的脱氨酶使R环范围内不超过两个核碱基脱氨。在一些实施方案中,融合蛋白的脱氨酶使R环范围内不超过三个核碱基脱氨。在一些实施方案中,融合蛋白的脱氨酶使R环范围内不超过2、3、4、5、6、7、8、9或10个核碱基脱氨。R环是三链核酸结构,包含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,可以形成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。在一些实施方案中,R-环包括间隔序列和目标DNA互补序列的杂交区域。R环区可以是长度为约5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49或50个碱基对。在一些实施方案中,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括与向导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。在一些实施方案中,在R环区域中编辑包括将非互补链(前间隔链)上的核碱基编辑为目标DNA序列中的向导RNA。

[4437] 本文所述融合蛋白可以在不同于规范碱基编辑的编辑窗口中实现目标脱氨。在一些实施方案中,目标核碱基在目标多核苷酸序列中PAM序列上游约1至约20个碱基。在一些实施方案中,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。在一些实施方案中,目标核碱基远离在目标多核苷酸序列中PAM序列上游为约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、大约10到14个碱基对、大约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、大约17至19个碱基对、约18到20个碱基对。在一些实施方案中,目标核碱基是约1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20或更多碱基对远离或上游PAM序列。在一些实施方案中,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。在一些实施方案中,目标核碱基是PAM序列上游约2、3、4或6个碱基对。

[4438] 因此,本文还提供融合蛋白文库和使用其优化碱基编辑的方法,与经典碱基编辑器相比,允许替代的优选碱基编辑窗口,例如BE4。在一些实施方案中,本公开提供用于优化碱基编辑的蛋白质文库,其包括多个融合蛋白,其中所述多个融合蛋白中的每一个都包括脱氨酶的N端片段和C端片段在侧边的Cas9多肽,其中所述融合蛋白中的每一种的N端片段不同于多种融合蛋白的其余部分的N端片段,或者其中所述融合蛋白中的每一种的C端片段不同于C端多种融合蛋白的其余片段,其中每个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前间隔邻近基序(PAM)序列附近脱氨,并且其中N端片段或C端片段结合目标多核苷酸序列。在一些实施方案中,对于CRISPR R环内的每个核碱基,多个融合蛋白中的至少一个使核碱基脱氨。在一些实施方案中,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,多个融合蛋白中的至少一个使核碱基脱氨。在一些实施方案中,本文所提供包括允许优化碱基编辑的融合蛋白文库的试剂盒。

[4439] 融合蛋白可包括多于一种异源多肽。例如,融合蛋白可以另外包括一个或多个UGI结构域和/或一个或多个核定位信号。两个或多个异源域可以串联插入。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。

[4440] 在一些实施方案中,碱基编辑器包括融合蛋白,所述融合蛋白包括napDNAbp结构域(例如Cas12源结构域)和内部融合的核碱基编辑结构域(例如脱氨酶结构域的全部或部分)。在一些实施方案中,napDNAbp是Cas12b。在一些实施方案中,碱基编辑器包括具有插入在下表A中提供的基因座处的内部融合的TadA*8结构域的BhCas12b结构域。

[4441] 表A:Cas12b蛋白中的插入位点

[4442]

BhCas12b	插入位点	插入氨基酸之间
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	插入位点	插入氨基酸之间
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	插入位点	插入氨基酸之间
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

[4443] 在一些实施方案中,碱基编辑器可以包括多个结构域。例如,包括衍生自Cas12蛋白的napDNAbp结构域的碱基编辑器可包括对应于野生型或天然Cas12的REC叶和NUC叶的REC叶和NUC叶。在另一个实例中,碱基编辑器可以包括一个或多个RuvC结构域WED结构域。在一些实施方案中,碱基编辑器的一个或多个结构域包括相对于包括所述结构域的多肽的野生型版本的突变(例如,取代、插入、缺失)。

[4444] 本发明的融合蛋白包括核酸编辑结构域。在一些实施方案中,核酸编辑结构域可以催化C至U碱基的变化。在一些实施方案中,核酸编辑结构域是脱氨酶结构域。在一些实施方案中,脱氨酶是胞苷脱氨酶或腺苷脱氨酶。在一些实施方案中,脱氨酶是载脂蛋白B mRNA-编辑复合物(APOBEC)家族脱氨酶。在一些实施方案中,脱氨酶是APOBEC1脱氨酶。在一些实施方案中,脱氨酶是APOBEC2脱氨酶。在一些实施方案中,脱氨酶是APOBEC3脱氨酶。在一些实施方案中,脱氨酶是APOBEC3A脱氨酶。在一些实施方案中,脱氨酶是APOBEC3B脱氨酶。在一些实施方案中,脱氨酶是APOBEC3C脱氨酶。在一些实施方案中,脱氨酶是APOBEC3D脱氨酶。在一些实施方案中,脱氨酶是APOBEC3E脱氨酶。在一些实施方案中,脱氨酶是APOBEC3F脱氨酶。在一些实施方案中,脱氨酶是APOBEC3G脱氨酶。在一些实施方案中,脱氨酶是APOBEC3H脱氨酶。在一些实施方案中,脱氨酶是APOBEC4脱氨酶。

[4445] 在一些实施方案中,脱氨酶是激活诱导的脱氨酶(AID)。

[4446] 在一些实施方案中,脱氨酶是脊椎动物脱氨酶。在一些实施方案中,脱氨酶是无脊椎动物脱氨酶。在一些实施方案中,脱氨酶是人类、黑猩猩、大猩猩、猴、牛、狗、大鼠或小鼠脱氨酶。在一些实施方案中,脱氨酶是人类脱氨酶。在一些实施方案中,脱氨酶是大鼠脱氨酶,例如,rAPOBEC1。在一些实施方案中,脱氨酶是海七鳗胞苷脱氨酶1(pmCDA1)。在一些实施方案中,脱氨酶是人类APOBEC3G。在一些实施方案中,脱氨酶是人类APOBEC3G的片段。在一些实施方案中,脱氨酶是包括D316R和D317R突变的人类APOBEC3G变体。在一些实施方案中,脱氨酶是人类APOBEC3G的片段,并且包括对应于D316R和D317R突变的突变。在一些实施方案中,核酸编辑结构域是与本文所述任何脱氨酶的脱氨酶结构域至少80%、至少85%、至少90%、至少92%、至少95%、至少96%、至少97%、至少98%、至少99%,或至少99.5%相同。

[4447] 连接子

[4448] 在某些实施方案中,连接子可用于连接本发明的任何肽或肽结构域。连接子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连接子。在某些实施方案中,连接子是多肽或基于氨基酸。在其他实施方案中,连接子不是肽样的。在某些实施方案中,连接子是共价键(例如,碳-碳键、二硫键、碳-杂原子键等)。在某些实施方案中,连接子是酰胺键的碳-氮键。在某些实施方案中,连接子是环状或无环、取代或未取代、支链或未支链的脂族或

杂脂族连接子。在某些实施方案中,连接子是聚合的(例如,聚乙烯、聚乙二醇、聚酰胺、聚酯等)。在某些实施方案中,连接子包括氨基链烷酸的单体、二聚体或聚合物。在某些实施方案中,连接子包括氨基链烷酸(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)。在某些实施方案中,连接子包括氨基己酸(Ahx)的单体、二聚体或聚合物。在某些实施方案中,连接子基于碳环部分(例如,环戊烷、环己烷)。在其他实施方案中,连接子包括聚乙二醇部分(PEG)。在其他实施方案中,连接子包括氨基酸。在某些实施方案中,连接子包括肽。在某些实施方案中,连接子包括芳基或杂芳基部分。在某些实施方案中,连接子基于苯环。连接子可以包括功能化部分以促进亲核试剂(例如,硫醇、氨基)从肽连接到连接子。任何亲电子试剂都可以用作连接子的一部分。实施例性的亲电子试剂包括但不限于激活酯、激活酰胺、迈克尔受体、卤代烷、芳基卤、酰基卤和异硫氰酸酯。

[4449] 在一些实施方案中,连接子是一个氨基酸或多个氨基酸(例如,肽或蛋白质)。在一些实施方案中,连接子是键(例如共价键)、有机分子、基团、聚合物或化学部分。在一些实施方案中,连接子的长度为约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、48、49、50、55、60、65、70、75、80、85、90、95或100)个氨基酸。

[4450] 在一些实施方案中,腺苷脱氨酶和napDNAbp通过长度为4、16、32或104个氨基酸的连接子融合。在一些实施方案中,连接子的长度为约3至约104个氨基酸。在一些实施方案中,本文所提供的任何融合蛋白包括经由连接子彼此融合的腺苷脱氨酶和Cas9结构域。可以采用脱氨酶结构域(例如,工程化的ecTadA)和Cas9结构域之间的各种连接子长度和灵活性(例如,范围是从$(GGGS)_n$、$(GGGGS)_n$及$(G)_n$形式的非常灵活的连接子到$(EAAAK)_n$、$(SGGS)_n$、SGSETPGTSESATPES形式的更刚性连接子(参见,例如Guilinger JP,Thompson DB,Liu DR.Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification.Nat.Biotechnol.2014;32(6):577-82;其全部内容通过引用方式并入本文)和$(XP)_n$)以实现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,连接子包括$(GGS)_n$基序,其中n是1、3或7。在一些实施方案中,本文所提供的任何融合蛋白的腺苷脱氨酶和Cas9结构域经由包括氨基酸序列SGSETPGTSESATPES的连接子(例如,XTEN连接子)融合。

[4451] Cas9与向导RNA的复合物

[4452] 本公开的一些方面提供包括本文所提供的任何融合蛋白和结合融合蛋白的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A\突变的向导)。这些复合物也称为核糖核蛋白(RNP)。可以使用任何连接融合蛋白结构域的方法(例如,范围是从非常灵活的$(GGGS)_n$、$(GGGGS)_n$和$(G)_n$形式的连接子到更刚性的$(EAAAK)_n$、$(SGGS)_n$、SGSETPGTSESATPES形式的连接子(参见,例如Guilinger JP,Thompson DB,Liu DR.Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification.Nat.Biotechnol.2014;32(6):577-82;其全部内容以引用方式并入本文)和(XP)n)以实现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。在一些实施方案中,连接子包括(GGS)n基序,其中n是1、3或7。在一些实施方案中,本文所提供的融合蛋白的Cas9结构域通过包括氨基酸序列的连接子融合SGSETPGTSESATPES。

[4453] 在一些实施方案中,向导核酸(例如向导RNA)的长度为15至100个核苷酸并且包括与目标序列互补的至少10个连续核苷酸的序列。在一些实施方案中,向导RNA的长度是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。在一些实施方案中,向导RNA包括与目标序列互补的15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39或40个连续核苷酸。在一些实施方案中,目标序列是DNA序列。在一些实施方案中,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。在一些实施方案中,目标序列是人类基因组中的序列。在一些实施方案中,目标序列的3'端紧邻规范PAM序列(NGG)。在一些实施方案中,目标序列的3'端紧邻非规规范PAM序列(例如,表1列出的序列或5'-NAA-3')。在一些实施方案中,向导核酸(例如,向导RNA)与感兴趣的基因(例如,与疾病或病症相关的基因)中的序列互补。

[4454] 本公开的一些方面提供使用本文所提供的融合蛋白或复合物的方法。例如,本公开的一些方面提供包括使DNA分子与本文所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约15至100个核苷酸长并且包括至少与目标序列互补的10个连续核苷酸。在一些实施方案中,目标序列的3'端紧邻规范PAM序列(NGG)。在一些实施方案中,目标序列的3'端不与规范PAM序列(NGG)紧邻。在一些实施方案中,目标序列的3'端紧邻AGC、GAG、TTT、GTG或CAA序列。在一些实施方案中,目标序列的3'端紧邻NGA、NGCG、NGN、NNGRRT、NNNRRT、NGCG、NGCN、NGTN、NGTN、NGTN或5'(TTTV)序列。

[4455] 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,识别任何同源蛋白质和相应编码核酸中的相应残基。

[4456] 对于本领域技术人员将是显而易见的是,为了将本文公开的任何融合蛋白靶向目标位点,例如包括待编辑突变的位点,通常需要共表达融合蛋白和向导RNA。如本文别处更详细解释的,向导RNA通常包括允许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,向导RNA和tracrRNA可以作为两个核酸分子分开提供。在一些实施方案中,向导RNA包括一种结构,其中向导序列包括与目标序列互补的序列。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目标位点的合适的向导RNA的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序列通常包括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。本文提供一些适用于将任何提供的融合蛋白靶向特定目标序列的实施例性向导RNA序列。

[4457] 使用包括腺苷脱氨酶变体和Cas9结构域的融合蛋白的方法

[4458] 本公开的一些方面提供使用本文所提供的融合蛋白或复合物的方法。例如,本公开的一些方面提供包括使编码蛋白质的突变形式的DNA分子与本文所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约15至100个核苷酸长,并且包括与目标序列互补的至少10个连续核苷酸的序列。在一些实施方案中,目标序列的3'端紧邻规范PAM序列(NGG)。在一些实施方案中,目标序列的3'端不与规范PAM序列(NGG)紧邻。在一些实施方案中,目标序列的3'端紧邻AGC、GAG、TTT、GTG或CAA序列。在一些实施方案中,目标序列的

3’端紧邻NGA、NGCG、NGN、NNGRRT、NNNRRT、NGCG、NGCN、NGTN、NGTN、NGTN或5’(TTTV)序列。

[4459] 应将理解,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,识别任何同源蛋白质和相应编码核酸中的相应残基。

[4460] 对本领域技术人员来说将是显而易见的是,为了将如本文公开的包括Cas9结构域和腺苷脱氨酶变体(例如,ABE8)的任何融合蛋白靶向目标位点,例如,包括对于要编辑的突变,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,向导RNA通常包括允许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,向导RNA和tracrRNA可以作为两个核酸分子分开提供。在一些实施方案中,向导RNA包括一种结构,其中向导序列包括与目标序列互补的序列。向导序列通常为20个核苷酸长。基于本公开,用于将Cas9:核酸编辑酶/结构域融合蛋白靶向特定基因组目标位点的合适的向导RNA的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序列通常包括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。本文提供一些适用于将任何提供的融合蛋白靶向特定目标序列的实施例性向导RNA序列。

[4461] Cas12与向导RNA的复合物

[4462] 本公开的一些方面提供包括本文所提供的任何融合蛋白和向导RNA(例如,靶向用于编辑的目标多核苷酸的向导)的复合物。

[4463] 在一些实施方案中,向导核酸(例如向导RNA)为15至100个核苷酸的长,并且包括与目标序列互补的至少10个连续核苷酸的序列。在一些实施方案中,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。在一些实施方案中,向导RNA包括与目标序列互补的15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39或40个连续核苷酸。在一些实施方案中,目标序列是DNA序列。在一些实施方案中,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。在一些实施方案中,目标序列是人类基因组中的序列。在一些实施方案中,目标序列的3’端紧邻规范PAM序列。在一些实施方案中,目标序列的3’端紧邻非规规范PAM序列。

[4464] 本公开的一些方面提供使用本文所提供的融合蛋白或复合物的方法。例如,本公开的一些方面提供包括使DNA分子与本文所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约15至100个核苷酸长,并且包括至少与目标序列互补的10个连续核苷酸。在一些实施方案中,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。

[4465] 应将理解,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,识别任何同源蛋白质和相应编码核酸中的相应残基。

[4466] 对于本领域技术人员将是显而易见的是,为了将本文公开的任何融合蛋白靶向目标位点,例如包括待编辑突变的位点,通常需要共表达融合蛋白和向导RNA。如本文别处更

详细解释的,向导RNA通常包括允许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,向导RNA和tracrRNA可以作为两个核酸分子分开提供。在一些实施方案中,向导RNA包括一种结构,其中向导序列包括与目标序列互补的序列。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目标位点的合适的向导RNA的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序列通常包括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。本文提供一些适用于将任何提供的融合蛋白靶向特定目标序列的实施例性向导RNA序列。

[4467] 本文公开的碱基编辑器的结构域可以以任何顺序排列,只要脱氨酶结构域内化在Cas12蛋白中即可。包括融合蛋白的碱基编辑器的非限制性实例可以排列如下:

[4468] NH2-[Cas12结构域]-连接子1-[ABE8]-连接子2-[Cas12结构域]-COOH;

[4469] NH2-[Cas12结构域]-连接子1-[ABE8]-[Cas12结构域]-COOH;

[4470] NH2-[Cas12结构域]-[ABE8]-连接子2-[Cas12结构域]-COOH;

[4471] NH2-[Cas12结构域]-[ABE8]-[Cas12结构域]-COOH;

[4472] NH2-[Cas12结构域]-连接子1-[ABE8]-连接子2-[Cas12域]-[肌苷BER抑制剂]-COOH;

[4473] NH2-[Cas12结构域]-连接子1-[ABE8]-[Cas12结构域]-[肌苷BER抑制剂]-COOH;

[4474] NH2-[Cas12结构域]-[ABE8]-连接子2-[Cas12结构域]-[肌苷BER抑制剂]-COOH;

[4475] NH2-[Cas12结构域]-[ABE8]-[Cas12结构域]-[肌苷BER抑制剂]-COOH;

[4476] NH2-[肌苷BER抑制剂]-[Cas12结构域]-连接子1-[ABE8]-连接子2-[Cas12结构域]-COOH;

[4477] NH2-[肌苷BER抑制剂]-[Cas12结构域]-连接子1-[ABE8]-[Cas12结构域]-COOH;

[4478] NH2-[肌苷BER抑制剂]-[Cas12结构域]-[ABE8]-连接子2-[Cas12结构域]-COOH;

[4479] NH2-[肌苷BER抑制剂]NH2-[Cas12结构域]-[ABE8]-[Cas12结构域]-COOH;

[4480] 此外,在某些情况下,Gam蛋白可以与碱基编辑器的N端融合。在某些情况下,Gam蛋白可以与碱基编辑器的C端融合。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。在一些实施方案中,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。在一些实施方案中,174个残基的Gam蛋白与碱基编辑器的N端融合。看。参见,Komor,A.C.,et al.,“Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity”Science Advances 3:eaao4774(2017)。在某些情况下,一个或多个突变可以改变碱基编辑器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,一个或多个突变不改变域相对于野生型结构域的长度。例如,任何结构域中的取代都不会改变碱基编辑器的长度。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)可包含:

[4481] NH2-[Cas12结构域]-连接子1-[APOBEC1]-连接子2-[Cas12结构域]-COOH;

[4482] NH2-[Cas12结构域]-连接子1-[APOBEC1]-[Cas12结构域]-COOH;

[4483] NH2-[Cas12结构域]-[APOBEC1]-连接子2-[Cas12结构域]-COOH;

[4484] NH2-[Cas12结构域]-[APOBEC1]-[Cas12结构域]-COOH;

[4485] NH2-[Cas12结构域]-连接子1-[APOBEC1]-连接子2-[Cas12域]-[UGI]-COOH;

[4486] NH2-[Cas12结构域]-连接子1-[APOBEC1]-[Cas12结构域]-[UGI]-COOH;

[4487] NH2-[Cas12结构域]-[APOBEC1]-连接子2-[Cas12结构域]-[UGI]-COOH;

[4488] NH2-[Cas12结构域]-[APOBEC1]-[Cas12结构域]-[UGI]-COOH;

[4489] NH2-[UGI]-[Cas12结构域]-连接子1-[APOBEC1]-连接子2-[Cas12结构域]-COOH;

[4490] NH2-[UGI]-[Cas12结构域]-连接子1-[APOBEC1]-[Cas12结构域]-COOH;

[4491] NH2-[UGI]-[Cas12结构域]-[APOBEC1]-连接子2-[Cas12结构域]-COOH;

[4492] NH2-[UGI]-[Cas12结构域]-[APOBEC1]-[Cas12结构域]-COOH;

[4493] 在一些实施方案中,本文所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。参见Komor,A.C.,et al.,“Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage”Nature 533,420-424(2016);Gaudelli,N.M.,et al.,“Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage”Nature 551,464-471(2017);以及Komor,A.C.,et al.,“Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity”Science Advances 3:eaao4774(2017),其全部内容特此引入作为参考。

[4494] 定义的目标区域可以是脱氨窗口。脱氨窗口可以是其中碱基编辑器作用于目标核苷酸并使其脱氨的限定区域。在一些实施例中,脱氨窗口在2、3、4、5、6、7、8、9或10个碱基区域内。在一些实施例中,脱氨窗为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。

[4495] 本公开的碱基编辑器可以包括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,碱基编辑器包括核定位序列(NLS)。在一些实施方案中,碱基编辑器的NLS位于脱氨酶结构域和napDNAbp结构域之间。在一些实施方案中,碱基编辑器的NLS位于napDNAbp结构域的C端。

[4496] 融合蛋白中包括的蛋白质结构域可以是异源功能域。可包括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、SUMO化活性、去SUMO化活性,或上述的任何组合。

[4497] 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-半乳糖基化酶、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)及包括蓝色荧光蛋白(BFP)的自发荧光蛋白。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。

[4498] 在一些实施方案中,BhCas12b向导多核苷酸具有以下序列:

[4499] BhCas12b sgRNA支架(底线)+20核苷酸至23核苷酸向导序列(用N_n表示)

[4500] 5'GTTCTGTCTTTTGGTCAGGACAACCGTCTAGCTATAAGTGCTGCAGGGTGTGAGAAACTCCTATTGCTGGACGATGTCTCTTACGAGGCATTAGCACNNNNNNNNNNNNNNNNNNN-3'

[4501] 在一些实施方案中,BvCas12b和AaCas12b向导多核苷酸具有以下序列:

[4502] BvCas12b sgRNA支架(底线)+20核苷酸至23核苷酸向导序列(用N_n表示)

[4503] 5'GACCTATAGGGTCAATGAATCTGTGCGTGTGCCATAAGTAATTAAAAATTACCCACCACAGGAGCACCTGAAAACAGGTGCTTGGCACNNNNNNNNNNNNNNNNNNN-3'

[4504] AaCas12b sgRNA支架(底线)+20核苷酸至23核苷酸向导序列(用N_n表示)

[4505] 5'GTCTAAAGGACAGAATTTTTCAACGGGTGTGCCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAACTTCTCAAAAAGAACGATCTGAGAAGTGGCACNNNNNNNNNNNNNNNNNNN-3'

[4506] 碱基编辑器效率

[4507] CRISPR-Cas9核酸酶已被广泛用于介导靶向的基因组编辑。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序列指定的目标位点诱导双链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,这会导致随机插入或缺失(indel),从而导致移码突变破坏基因。在存在与DSB侧边序列高度同源的供体DNA模板的情况下,可以通过称为同源性定向修复(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。本文所提供的碱基编辑系统提供一种提供基因组编辑的新方法,所述方法无需产生双链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。

[4508] 本发明的融合蛋白有利地修饰编码包括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,"插入缺失"是指核酸内核苷酸碱基的插入或缺失。这种插入或缺失可导致基因编码区内的移码突变。在一些实施方案中,期望产生有效修饰(例如,突变或脱氨)核酸内的特定核苷酸而不在核酸中产生大量插入或缺失(即,插入缺失)的碱基编辑器。在某些实施方案中,本文所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,突变或脱氨)。

[4509] 在一些实施方案中,本文所提供的任何碱基编辑器系统导致在目标多核苷酸序列中形成小于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。

[4510] 在一些实施方案中,包括本文所述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形成小于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。在一些实施方案中,包括本文所述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小于0.8%的插入缺失形成。在一些实施方案中,包括本文所述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至多0.8%的插入缺失形成。在一些实施方案中,包括本文所述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小于0.3%的插入缺失形成。在一些实施方案中,与包括ABE7碱基编辑器之一的碱基编辑器系统相比,包括所述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。在一些实施方案中,与包括ABE7.10的碱基编辑器系统相比,包括本文所述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。

[4511] 在一些实施方案中,与包括ABE7碱基编辑器之一的碱基编辑器系统相比,包括本文所述的ABE8碱基编辑器变体之一的任何碱基编辑器系统具有降低的插入/缺失频率。在一些实施方案中,与包括ABE7碱基编辑器之一的碱基编辑器系统相比,包括本文所述ABE8碱基编辑器变体之一的任何碱基编辑器系统具有至少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%或至少95%的插入缺失频率降低。在一些实施方案中,与包括ABE7.10的碱基编辑器系统相比,包括本文所述ABE8碱基编辑器变体之一的碱基编辑器系统具有至少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%、至少至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%或至少95%的插入/缺失频率降低。

[4512] 本发明提供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在碱基编辑窗口中改变的碱基(例如,“旁观者”)。

[4513] 在一些实施方案中,包括本文所述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。在一些实施方案中,非预期编辑或突变是旁观者突变或旁观者编辑,例如目标核苷酸序列的靶窗口中非预期或非靶位置中的目标碱基(例如,A或C)的碱基编辑。在一些实施方案中,与包括ABE7碱基编辑器(例如ABE7.10)的碱基编辑器系统相比,包括本文所述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。在一些实施方案中,与包括ABE7碱基编辑器(例如ABE7.10)的碱基编辑器系统相

比,包括本文所述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%。在一些实施方案中,与包括ABE7碱基编辑器例如ABE7.10的碱基编辑器系统相比,包括本文所述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。

[4514] 在一些实施方案中,包括本文所述ABE8碱基编辑器变体之一的任何碱基编辑系统已减少了虚假编辑。在一些实施方案中,非预期编辑或突变是虚假突变或虚假编辑,例如,非特异性编辑或向导独立编辑基因组的非预期或非目标区域中的目标碱基(例如,A或C)。在一些实施方案中,与包括ABE7碱基编辑器(例如ABE7.10)的碱基编辑器系统相比,包括本文所述ABE8碱基编辑器变体之一的任何碱基编辑系统具有减少的虚假编辑。在一些实施方案中,与包括ABE7碱基编辑器的系统(例如ABE7.10)的碱基编辑器相比,包括本文所述ABE8碱基编辑器变体之一的任何碱基编辑系统已将虚假编辑减少至少1%、至少2%、至少3%、至少4%、至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%或至少99%。在一些实施方案中,与包括ABE7碱基编辑器例如ABE7.10的碱基编辑器系统相比,包括本文所述ABE8碱基编辑器变体之一的任何碱基编辑系统已将虚假编辑减少至少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。

[4515] 本公开的一些方面基于以下认识:本文所提供的任何碱基编辑器能够有效地在核酸(例如,受试者基因组内的核酸)中产生预期突变,例如点突变不会产生大量意外突变,例如意外点突变(即旁观者的突变)。在一些实施方案中,本文所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)。在一些实施方案中,本文所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。

[4516] 在一些实施方案中,本文所述任何ABE8碱基编辑器变体具有至少0.01%、至少1%、至少2%、至少3%、至少4%、至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%或至少99%的碱基编辑效率。在一些实施方案中,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。在一些实施方案中,如通过细胞群中编辑的核碱基所测量,本文所述任何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%。

[4517] 在一些实施方案中,与ABE7碱基编辑器(例如ABE7.10)相比,本文所述任何ABE8碱

基编辑器变体具有更高的碱基编辑效率。在一些实施方案中,本文所述任何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%。

[4518] 在一些实施方案中,与ABE7碱基编辑器例如ABE7.10相比本文所述任何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倍高的碱基编辑效率。

[4519] 在一些实施方案中,本文所述任何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%的靶向碱基编辑效率。在一些实施方案中,如由细胞群中的编辑的目标核碱基所测量,本文所述任何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%的靶上碱基编辑效率。

[4520] 在一些实施方案中,与ABE7碱基编辑器相比,本文所述任何ABE8碱基编辑器变体具有更高的目标碱基编辑效率。在一些实施方案中,与ABE7碱基编辑器(例如ABE7.10)相比,本文所述任何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%,至少高390%、至少400%、至少450%、或至少500%的靶上碱基编辑效率。

[4521] 在一些实施方案中,与ABE7碱基编辑器(例如ABE7.10)相比,本文所述任何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倍高的靶上碱基编辑效率。

[4522] 本文所述ABE8碱基编辑器变体可经由质粒、载体、LNP复合物或mRNA递送至宿主细胞。在一些实施方案中,本文所述任何ABE8碱基编辑器变体作为mRNA递送至宿主细胞。在一些实施方案中,如由编辑的核碱基所测量,经由基于核酸的递送系统(例如mRNA)递送的ABE8碱基编辑器具有至少1%、至少2%、至少3%、至少4%、至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%或至少99%的靶上编辑效率。在一些实施方案中,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。在一些实施方案中,与由质粒或载体系统递送时相比,本文所述任何ABE8碱基编辑器变体当由mRNA系统递送时具有至少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%的靶上编辑效率。在一些实施方案中,与由质粒或载体系统递送时相比,本文所述任何ABE8碱基编辑器变体当由mRNA系统递送时具有至少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倍高的靶上编辑效率。

[4523] 在一些实施方案中,包括本文所述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的小于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%的脱靶编辑。

[4524] 在一些实施方案中,与由质粒或载体系统递送时相比,本文所述任何ABE8碱基编

辑器变体当由mRNA系统递送时具有更低的向导脱靶编辑效率。在一些实施方案中,当由mRNA系统递送时与由质粒或载体系统递送时相比,本文所述任何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%低的向导脱靶编辑效率。在一些实施方案中,与由质粒或载体系统递送时相比,本文所述任何ABE8碱基编辑器变体当由mRNA系统递送时具有至少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倍低的导脱靶编辑效率。在一些实施方案中,当与由质粒或载体系统递送时相比,本文所述任何ABE8碱基编辑器变体当由mRNA系统递送时具有至少约2.2倍低的向导脱靶编辑效率。

[4525] 在一些实施方案中,与由质粒或载体系统递送时相比,本文所述任何ABE8碱基编辑器变体当由mRNA系统递送时具有更低的不依赖于向导的脱靶编辑效率。在一些实施方案中,当由质粒或载体系统递送时,本文所述任何ABE8碱基编辑器变体当由mRNA系统递送时具有至少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%的更低的与向导无关的脱靶编辑效率。在一些实施方案中,与由质粒或载体系统递送时相比,本文所述任何ABE8碱基编辑器变体当mRNA系统递送时具有至少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倍低的不依赖于向导的脱靶编辑效率。在一些实施方案中,与由质粒或载体系统递送时相比,本文所述ABE8碱基编辑器变体当mRNA系统递送时具有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)。在一些实施方案中,本文所述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。

[4526] 在一些实施方案中,本文所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的比率。在一些实施方案中,本文所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、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的比率。

[4527] 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)中所述;Komor,A.C.,et al.,"Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage"Nature 533,420-424(2016);Gaudelli,N.M.,et al.,"Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage"Nature 551,464-471(2017);以及Komor,A.C.,et al.,"Improved base

excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity"Science Advances 3: eaao4774(2017);其全部内容特此通过引用方式并入。

[4528] 在一些实施方案中,为了计算插入缺失频率,扫描测序读数以寻找与两个10个碱基序列的精确匹配,所述两个10个碱基序列在可以发生插入缺失的窗口侧边。若未找到完全匹配,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括插入缺失。如果插入缺失窗口比参考序列长或短两个或更多个碱基,则测序读数分别被归类为插入或缺失。在一些实施方案中,本文所提供的碱基编辑器可以限制核酸区域中插入缺失的形成。在一些实施方案中,所述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。

[4529] 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。在一些实施方案中,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天后测定插入缺失的数量或比例。应当理解,如本文所述碱基编辑器的特征可应用于任何融合蛋白或使用本文所提供的融合蛋白的方法。

[4530] 在一些实施方案中,本文所提供的碱基编辑器能够限制核酸区域中插入缺失的形成。在一些实施方案中,所述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。在一些实施方案中,本文所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。在一些实施方案中,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天后测定插入缺失的数量或比例。

[4531] 本公开的一些方面基于本文所提供的任何碱基编辑器能够有效地在核酸(例如,受试者基因组内的核酸)中产生预期突变而不产生大量非预期突变的认知。突变(例如,虚假的脱靶编辑或旁观者编辑)。在一些实施方案中,预期突变是由专门设计用于改变或校正目标基因中的突变的与gRNA结合的特定碱基编辑器产生的突变。在一些实施方案中,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HBG突变。

[4532] 在一些实施方案中,本文所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(例如,预期突变:非预期突变)的比率。在一些实施方案中,本文所提供的任何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、至至少3.5:1、至少4:1、至少4.5:1、至少5:1、至少5.5:1、至少6:1、至少6.5:1、至少7:1、至少7.5:1,至少8:1,至少10:1,至少12:1,至少15:1,至少20:1,至少25:1,至少30:1,至少40:1、至少50:1、至少100:1、至少150:1、至少200:1、至少250:1、至少500:1或至少1000:1或更多的预期突变与非预期突变的比率。应当理解,本文所述的碱基编辑器的特征可以应用于任何融合蛋白,或使用本文所提供的融合蛋白的方法。

[4533] 多重编辑

[4534] 在一些实施方案中,本文所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。在一些实施方案中,多个核碱基对位于同一基因中。在一些实施方案中,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。在一些实施方案中,多重编辑可包括一种或多种向导多核苷酸。在一些实施方案中,多重编辑可包括一个或多个碱基编辑器系统。在一些实施方案中,多重编辑可以包括一个或多个具有单个向导多核苷酸的碱基编辑器系统。在一些实施方案中,多重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。在一些实施方案中,多重编辑可以包括一个或多个具有单碱基编辑器系统的向导多核苷酸。在一些实施方案中,多重编辑可包括至少一个不需要PAM序列来靶向结合目标多核苷酸序列的向导多核苷酸。在一些实施方案中,多重编辑可包括至少一个向导多核苷酸,其需要PAM序列靶向结合目标多核苷酸序列。在一些实施方案中,多重编辑可以包括至少一个不需要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个需要PAM序列靶向结合目标多核苷酸的向导多核苷酸的混合。应当理解,使用本文所述的任何碱基编辑器的多重编辑的特征可以应用于使用本文所提供的任何碱基编辑器的方法的任何组合。还应当理解,使用本文所述任何碱基编辑器的多重编辑可以包括多个核碱基对的顺序编辑。

[4535] 在一些实施方案中,多个核碱基对在一个或多个基因中。在一些实施方案中,多个核碱基对在同一基因中。在一些实施方案中,一个或多个基因中的至少一个基因位于不同基因座中。

[4536] 在一些实施方案中,编辑是编辑至少一个蛋白质编码区中的多个核碱基对。在一些实施方案中,编辑是编辑至少一个蛋白质非编码区中的多个核碱基对。在一些实施方案中,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多个核碱基对。

[4537] 在一些实施方案中,编辑与一个或多个向导多核苷酸结合。在一些实施方案中,碱基编辑器系统可以包括一个或多个碱基编辑器系统。在一些实施方案中,碱基编辑器系统可以包括一个或多个碱基编辑器系统与单个向导多核苷酸结合。在一些实施方案中,碱基编辑器系统可包括一个或多个碱基编辑器系统与多个向导多核苷酸结合。在一些实施方案中,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。在一些实施方案中,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个向导多核苷酸结合。在一些实施方案中,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。在一些实施方案中,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结合。应当理解,使用本文所述任何碱基编辑器的多重编辑的特征可以应用于使用本文所提供的任何碱基编辑器的方法的任何组合。还应当理解,编辑可以包括多个核碱基对的顺序编辑。

[4538] 在一些实施方案中,能够对一种或多种基因中的多个核碱基对进行多重编辑的碱基编辑器系统包括本文所述ABE8碱基编辑器变体之一。在一些实施方案中,能够多重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括ABE7碱基编辑器之一。在一些实施方案中,与包括ABE7碱基编辑器之一的能够多重编辑的碱基编辑器系统相比,包括本文所述的ABE8碱基编辑器变体之一的能够多重编辑的碱基编辑器系统具有更高的多重编辑效率。在一些实施方案中,与包括ABE7碱基编辑器之一的能够多重编辑的碱基编辑器系统

相比,能够进行多重编辑的碱基编辑器系统包括本文所述ABE8碱基编辑器变体之一具有至少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%更高的多重编辑效率。在一些实施方案中,与包括ABE7碱基编辑器之一的能够进行多重编辑的碱基编辑器系统相比,包括本文所述ABE8碱基编辑器变体之一的能够多重编辑的碱基编辑器系统具有至少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倍高的多重编辑效率。

[4539] 编辑核酸的方法

[4540] 本公开的一些方面提供用于编辑核酸的方法。在一些实施方案中,所述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。在一些实施方案中,所述方法包括以下步骤:a)使核酸的目标区域(例如,双链DNA序列)与包括碱基编辑器(例如,与腺苷脱氨酶融合的Cas9结构域)的复合物接触和向导核酸(例如,gRNA),b)诱导所述靶区域的链分离,c)将目标区域的单链中的所述靶核碱基对的第一个核碱基转换为第二个核碱基,及d)使用nCas9切割不超过目标区域中的一个链中,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。在一些实施方案中,所述方法导致核酸中少于20%的插入缺失形成。应当理解,在一些实施方案中,省略步骤b。在一些实施方案中,所述方法导致小于19%、18%、16%、14%、12%、10%、8%、6%、4%、2%、1%、0.5%、0.2%或0.1%的插入缺失形成率。在一些实施方案中,所述方法进一步包括用与第四个核碱基互补的第五个核碱基置换第二个核碱基,从而产生预期的编辑碱基对(例如,G·C至A·T)。在一些实施方案中,编辑至少5%的预期碱基对。在一些实施方案中,编辑至少10%、15%、20%、25%、30%、35%、40%、45%或50%的预期碱基对。

[4541] 在一些实施方案中,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。在一些实施方案中,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。在一些实施方案中,切割的单链(切口链)与向导核酸杂交。在一些实施方案中,切割的单链与包括第一核碱基的链相反。在一些实施方案中,碱基编辑器包括dCas9结构域。在一些实施方案中,碱基编辑器保护或结合非编辑链。在一些实施方案中,预期的编辑碱基对在PAM位点的上游。在一些实施方案中,预期的编辑碱基对是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。在一些实施方案中,预期的编辑碱基对在PAM位点的下游。在一些实施方案中,预期的编辑碱基对是在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。在一些实施方案中,所述方法不需要规范的(例如,NGG)PAM位点。在一些实施方案中,核碱基编辑器包括连接子。在一些实施方案中,连接子的长度为1至25个氨基酸。在一些实施方案中,连接子的长度为5至20个氨基酸。在一些实施方案中,连接子的长度为10、11、12、13、14、15、16、17、18、19或20个氨基酸。在一个实施方案中,连接子的长度为32个氨基酸。在另一个实施方案中,“长连接子”的长度为至少约60个氨基酸。在其他实施方案中,连接子长度在约3至100个氨基酸之间。在一些实施方案中,目标区域包括目标窗口,其中所述目标窗口包括目标核碱基对。在一些实施方案中,目标窗口包括1至10个核苷酸。在一些实施方案中,目标窗口的长度为1至9、1至8、1至7、1至6、1至5、1至4、1至3、1至2或1个核苷酸。在一些实施方案中,目标窗口的长度为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。在一些实施方案中,预期的编辑碱基对在目标窗口内。在一些实施方案中,目标窗口包括预期的编辑碱基对。在一些实施方案中,使用本文所提供的任何碱基编辑器来执行所述方法。在一些实施方案中,目标窗口是甲基化窗口。

[4542] 在一些实施方案中,本公开提供用于编辑核苷酸(例如,编码蛋白质的基因中的SNP)的方法。在一些实施方案中,本公开提供一种用于编辑双链DNA序列的核碱基对的方法。在一些实施方案中,所述方法包括a)使双链DNA序列的目标区域与包括碱基编辑器和向导核酸(例如gRNA)的复合物接触,其中所述目标区域包括目标核碱基对,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换为第二个核碱基,d)切割不超过一条所述目标区域的链,其中与第一核碱基互补的第三个核碱基被与第二核碱基互补的第四核碱基置换,并且第二核碱基被与第四核碱基互补的第五核碱基替换,从而产生预期的编辑碱基对,其中产生预期的碱基对的效率编辑的碱基对至少为5%。应当理解,在一些实施方案中,省略步骤b。在一些实施方案中,至少5%的预期碱基对被编辑。在一些实施方案中,至少10%、15%、20%、25%、30%、35%、40%、45%或50%的预期碱基对被编辑。在一些实施方案中,所述方法导致小于19%、18%、16%、14%、12%、10%、8%、6%、4%、2%、1%、0.5%、0.2%或更少0.1%插入缺失形成。在一些实施方案中,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。在一些实施方案中,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。在一些实施方案中,切割的单链与向导核酸杂交。在一些实施方案中,切割的单链与包括第一核碱基的链相反。在一些实施方案中,预期的编辑碱基对在PAM位点的上游。在一些实施方案中,预期的编辑碱基对是在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。在一些实施方案中,预期的编辑碱基对在PAM位点的下游。在一些实施方案中,预期的编辑碱基对是在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。在一些实施方案中,所述方法不需要规范的(例如,NGG)PAM位点。在一些实施方案中,连接子的长度为1至25个氨基酸。在一些实施方案中,连接子的长度为5至20个氨基酸。在一些实施方案中,连接子的长度为10、11、12、13、14、15、16、17、18、19或20个氨基酸。在一些实施方案中,目标区域包括目标窗口,其中所述目标窗口包括目标核碱基对。在一些实施方案中,目标窗口包括1至10个核苷酸。在一些实施方案中,目标窗口的长度为1至9、1至8、1至7、1至6、1至5、1至4、1至3、1至2或1个核苷酸。在一些实施方案中,目标窗口的长度是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。在一些实施方案中,预期的编辑碱基对出现在目标窗口内。在一些实施方案中,目标窗口包括预期的编辑碱基对。在一些实施方案中,核碱基编辑器是本文所提供的任何一种碱基编辑器。

[4543] 融合蛋白在宿主细胞中的表达

[4544] 本发明的包括腺苷脱氨酶变体的融合蛋白可以使用本领域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,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编码本发明的腺苷脱氨酶的DNA。克隆的DNA可以直接,或在需要时用限制性酶酶切后,或在添加合适的连接子和/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。碱基编辑系统在宿主细胞中被翻译以形成复合物。

[4545] 融合蛋白通过将编码一个或多个具有核碱基修饰活性的结构域(例如腺苷脱氨酶)的一个或多个多核苷酸可操作地连接到编码naDNAbp的多核苷酸以制备编码本发明的融合蛋白的多核苷酸来产生。在一些实施方案中,编码napDNAbp的多核苷酸和编码具有核碱基修饰活性的结构域的DNA可以各自与编码结合结构域或其结合配偶体的DNA融合,或者两种DNA都可以与编码分离内含肽的DNA融合,从而由此核酸序列识别转换模块和核酸碱基转换酶在宿主细胞中翻译形成复合物。在这些情况下,当需要时,连接子和/或核定位信号可连接至DNA之一或两者的合适位置。

[4546] 编码本文所述蛋白质结构域的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短链构建编码其全长的DNA。通过化学合成或PCR方法或Gibson组装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序列转换为宿主生物中频繁使用的密码子,而预计蛋白质表达水平会增加。可以使用作为待使用的宿主中密码子使用频率的数据,例如在Kazusa DNA Research Institute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),也可以参考显示每个宿主中密码子使用频率的文件。参考获得的数据和待引入的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。

[4547] 包括编码核酸序列辨识模块和/或核酸碱基转换酶的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产生。

[4548] 作为表达载体,来源于大肠杆菌的质粒(例如,pBR322、pBR325、pUC12、pUC13);枯草芽孢杆菌源的质粒(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);噬菌体诸如λ噬菌体等;昆虫病毒载体,例如杆状病毒等(例如,BmNPV、AcNPV);使用动物病毒载体,例如逆转录病毒、痘苗病毒、腺病毒等。

[4549] 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,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,也可以没有限制地使用构成启动子。

[4550] 例如,当宿主是动物细胞时,可使用SRα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、SRα启动子等是优选的。当宿主为大肠杆菌时,trp启动子、lac启动子、recA启动子、λP_L启动子、lpp启动子、T7启动子等是优选的。当宿主为芽孢杆菌属时,SPO1启动子、SPO2启动子、penP启动子等是优选的。当宿主为酵母时,Gal1/10启动子、PHO5启动子、PGK启动子、GAP启动子、ADH启动子等是优选的。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。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。

[4551] 在一些实施方式中,表达载体可按需求括有增强子、剪接信号、终止子、多聚苷酸添加信号、选择标记如耐药基因、营养缺陷型互补基因、复制起点等。

[4552] 编码本文所述蛋白质结构域的RNA可以通过例如在本身已知的体外转录系统中通过使用编码上述核酸序列辨识模块和/或核酸碱基转换酶为模板的编码DNA的载体转录为mRNA来制备。

[4553] 本发明的融合蛋白可以通过将含有编码核酸序列辨识模块和/或核酸碱基转换酶的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。

[4554] 可用于本发明的宿主细胞包括细菌细胞、酵母、昆虫细胞、动物细胞等。

[4555] 大肠杆菌属包括大肠杆菌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(JA22 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。

[4556] 芽孢杆菌属包括枯草芽孢杆菌M1114(Gene,24,255(1983))、枯草芽孢杆菌207-21(Journal of Biochemistry,95,87(1984))等。

[4557] 可用于表达本发明融合蛋白的酵母菌包括酿酒酵母菌AH22、AH22R$^-$、NA87-11A、DKD-5D、20B-12、粟酒裂殖酵母菌(Saccharomyces pombe)NCYC1913、NCYC2036、毕赤酵母菌(Pichia pastoris)KM71等。

[4558] 融合蛋白在昆虫细胞中表达,例如使用病毒载体,如AcNPV。昆虫宿主细胞包括以下任何细胞系:衍生自甘蓝粘虫幼虫的已建立的细胞系(草地夜蛾细胞;Sf细胞)、衍生自粉纹夜蛾(Trichoplusia ni)、High Five的中肠的MG1细胞。使用衍生自源粉纹夜蛾的卵的细胞、衍生自甘蓝夜蛾(Mamestra brasicae)的细胞、衍生自盐泽灯蛾(Estigmena acrea)的细胞等。当病毒为BmNPV时,衍生自家蚕的已建立的细胞系(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,Sf9细胞(ATCC CRL1711)、Sf21细胞(以上,In Vivo,13,213-217(1977))等。

[4559] 作为昆虫,例如家蚕、果蝇、蟋蟀等的幼虫用于表达本发明的融合蛋白(Nature,315,592(1985))。

[4560] 哺乳动物细胞系可用于表达融合蛋白。此类细胞系包括猴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细胞等及人类和其他哺乳动物的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。

[4561] 可以使用本领域技术人员公知的方法将植物细胞保持在培养物中。植物细胞培养

包括由各种植物(如水稻、小麦、玉米等谷物,西红柿、黄瓜、茄子、康乃馨、洋桔梗、烟草、拟南芥等产品作物)制备而成的悬浮培养的细胞、愈伤组织、原生质体、叶段、根段。

[4562] 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,否则不会表达所需的表型,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所需的表型,这是非常有用的,因为常规方法可以解决。

[4563] 使用任何转染方法(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)将编码本发明融合蛋白的表达载体导入宿主细胞方法)。根据要转染的宿主细胞选择转染方法。

[4564] 大肠杆菌可以根据,例如,Proc.Natl.Acad.Sci.USA,69,2110(1972),Gene,17,107(1982)等中所述方法转换。

[4565] 芽孢杆菌属可以根据例如Molecular&General Genetics,168,111(1979)等中描述的方法引入载体中。

[4566] 酵母细胞可以根据,例如,Methods in Enzymology,194,182-187(1991),Proc.Natl.Acad.Sci.USA,75,1929(1978)等中所述方法引入载体中。

[4567] 昆虫细胞可以根据例如在,Bio/Technology,6,47-55(1988)等中所述方法引入载体中。

[4568] 哺乳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),以及Virology,52,456(1973)中所述方法引入载体中。

[4569] 根据已知方法培养包括本发明表达载体的细胞,其根据宿主而变化。

[4570] 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。培养基优选含有转换体生长所需的碳源、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至约8。

[4571] 作为用于培养大肠杆菌的培养基,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)是优选的。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促进剂的有效功能。大肠杆菌通常在约15℃至约43℃下培养。必要时可进行曝气和搅拌。

[4572] 芽孢杆菌属通常在约30至约40℃培养。必要时可进行曝气和搅拌。

[4573] 用于培养酵母的培养基的例子包括伯克霍尔德最小培养基(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至约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。

[4574] 作为用于培养昆虫细胞或昆虫的培养基,例如,使用适当地含有诸如灭活的10%

牛血清等添加剂的Grace昆虫培养基(Grace's Insect Medium)(Nature,195,788(1962))等。培养基的pH值优选为约6.2至约6.4。培养通常在约27℃进行。必要时可进行曝气和搅拌。

[4575] 使用作为用于培养动物细胞的培养基,例如,含有约5至约20%胎牛血清的最低必需培养基(MEM)(Science,122,501(1952))、改良伊格尔氏培养基(Dulbecco's modified Eagle medium(DMEM))(Virology,8,396(1959)),RPMI 1640培养基(The Journal of the American Medical Association,199,519(1967)),199培养基(Proceeding of the Society for the Biological Medicine,73,1(1950))等。培养基的pH优选为约6至约8。培养通常在约30℃至约40℃进行。必要时可进行曝气和搅拌。

[4576] 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至约8。培养通常在约20℃至约30℃进行。必要时可进行曝气和搅拌。

[4577] 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,将编码本发明碱基编辑系统(例如,包括腺苷脱氨酶变体)的DNA引入到在诱导型启动子(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诱导)、类固醇反应启动子(由类固醇激素或其源物诱导)等)调控下的宿主细胞或其源物,在适当的阶段将诱导物质加入培养基中(或从培养基中去除)以诱导核酸的表达-修饰酶复合物,培养一定时间进行碱基编辑,将突变引入目标基因,以实现碱基编辑系统的瞬时表达。

[4578] 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。

[4579] 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)。

[4580] 使用基础编辑器的方法

[4581] 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新的策略,并应用于治疗学和基础研究。

[4582] 本公开提供用于治疗被诊断患有与点突变相关或由其引起的疾病的受试者的方法,所述点突变可以由本文所提供的碱基编辑器系统校正。例如,在一些实施方案中,提供一种方法,其包括向患有此类疾病(例如由基因突变引起的疾病)的受试者给药有效量的核碱基编辑器(例如,腺苷脱氨酶碱基编辑器或胞苷脱氨酶碱基编辑器)以用于校正疾病相关基因中的点突变。本公开提供用于治疗与可以通过脱氨酶介导的基因编辑校正的点突变相关或由其引起的疾病的方法。基于本公开,可以用本文所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。本文提供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。在一些实施方案中,碱基编辑器(例如,包括腺苷脱氨酶和Cas12结构域)的活性导致点突变的校正。在一些实施

方案中,目标DNA序列包括与疾病或病症相关的G→A点突变,并且突变体A碱基的脱氨产生与疾病或病症无关的序列。在一些实施方案中,目标DNA序列包括与疾病或病症相关的T→C点突变,并且突变体C碱基的脱氨产生与疾病或病症无关的序列。

[4583] 在一些实施方案中,目标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。在一些实施方案中,突变体A的脱氨导致突变体密码子编码的氨基酸发生变化。在一些实施方案中,突变体A的脱氨产生编码野生型氨基酸的密码子。在一些实施方案中,突变体C的脱氨导致突变体密码子编码的氨基酸发生变化。在一些实施方案中,突变体C的脱氨产生编码野生型氨基酸的密码子。在一些实施方案中,受试者已经或已经被诊断患有疾病或病症。

[4584] 在一些实施方案中,本文所提供的腺苷脱氨酶能够使DNA的脱氧腺苷残基脱氨。本公开的其他方面提供融合蛋白,其包括腺苷脱氨酶(例如,如本文所述使DNA中的脱氧腺苷脱氨的腺苷脱氨酶)和能够结合特定核苷酸序列的结构域(例如,Cas12)。例如,腺苷可以转换为肌苷残基,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的DNA的靶向编辑,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,例如在从受试者获得的细胞中随后被重新引入相同或另一受试者中;以及对于体内靶向突变的引入,例如,可以使用本文所提供的核碱基编辑器处理遗传缺陷的校正或在疾病相关基因中G至A或T至C突变中失活突变的引入。本公开提供利用脱氨酶和核碱基编辑器的脱氨酶、融合蛋白、核酸、载体、细胞、组合物、方法、试剂盒、系统等。

[4585] 产生预期的突变

[4586] 在一些实施方案中,本文所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。在一些实施方案中,功能障碍基因的功能通过引入预期突变来恢复。在一些实施方案中,本文所提供的方法可用于破坏基因产物的正常功能。本文所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。本领域技术人员将理解,本文所提供的核碱基编辑蛋白,例如包括napDNAbp结构域(例如,Cas12)和核碱基编辑结构域(例如,腺苷脱氨酶结构域或胞苷脱氨酶结构域)的融合蛋白可以是用于校正任何单个点A至G或C到T突变。在第一种情况下,突变体A至I的脱氨校正了突变,在后一种情况下,与突变体T碱基配对的A脱氨,然后进行一轮复制而校正了突变。

[4587] 在一些实施方案中,本公开提供在核酸(例如,受试者基因组内的核酸)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的碱基编辑器(例如作为意外的点突变)。在一些实施方案中,预期突变是由特定碱基编辑器(例如,胞苷碱基编辑器或腺苷碱基编辑器)与特别设计用于产生预期突变的向导多核苷酸(例如,gRNA)结合而产生的突变。在一些实施方案中,预期突变是与疾病或病症相关的突变。在一些实施方案中,预期突变是与疾病或病症相关的腺嘌呤(A)至鸟嘌呤(G)的点突变。在一些实施方案中,预期突变是与疾病或病症相关的胞嘧啶(C)至胸腺嘧啶(T)的点突变。在一些实施方案中,预期突变是基因的编码区或非编码区内的腺嘌呤(A)至鸟嘌呤(G)的点突变。在一些实施方案中,预期突变是基因的编码区或非编码区内的胞嘧啶(C)至胸腺嘧啶(T)的点突变。在一些实施方案中,预期突变是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

方案中,预期突变是消除终止密码子的突变。

[4588] 在一些实施方案中,本文所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)。在一些实施方案中,任何碱基本文所提供的编辑器能够产生至少1.5:1、至少2:1、至少2.5:1、至少3:1、至少3.5:1、至少4:1、至少4.5:1、至少5:1、至少5.5:1、至少6:1、至少6.5:1、至少7:1、至少7.5:1、至少8:1、至少10:1、至少12:1、至少15:1、至少20:1、至少25:1、至少30:1、至少40:1、至少50:1、至少100:1、至少150:1、至少200:1、至少250:1、至少500:1,或至少1000:1,或更高的预期突变与非预期突变(例如,预期点突变:

[4589] 非预期点突变)的比率。

[4590] 碱基编辑器效率的细节在国际PCT申请编号第PCT/2017/045381(WO 2018/027078)和PCT/US2016/058344(WO 2017/070632)中进行了描述,其各自的全部内容通过引用方式并入本文。另参见Komor,A.C.,et al.,"Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage"Nature 533,420-424(2016);Gaudelli,N.M.,et al.,"Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage"Nature 551,464-471(2017);以及Komor,A.C.,et al.,"Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity"Science Advances3:eaao4774(2017),其全部内容通过引用方式并入本文。

[4591] 在一些实施方案中,使用本文所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。在一些实施方案中,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,可以使用本文所提供的任何方法或方法的组合来实现多重编辑。

[4592] 输送系统

[4593] 基于核酸的核碱基编辑器和gRNA的递送

[4594] 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内容的碱基编辑系统的核酸给药于受试者或递送至细胞中。在一个实施方案中,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、mRNA、脂质纳米颗粒)或其组合递送。

[4595] 编码核碱基编辑器的核酸可以以裸露的DNA或RNA,例如mRNA,例如通过转染或电穿孔的方式直接递送至细胞(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),或者可以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,例如本文所述载体。

[4596] 核酸载体可包括编码本文所述融合蛋白的结构域的一个或多个序列。载体还可以包括编码信号肽的序列(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相关联(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)。

[4597] 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。

[4598] 根据本公开的核酸载体包括重组病毒载体。实施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括任何合适的货物。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。

[4599] 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方案中的递送载体。用于纳米颗粒制剂、和/或基因转移的实施例性脂质示于表10(以下)。

[4600] 表10

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	帮手
1,2-二油酰基-sn-甘油-3-磷脂酰乙醇胺	DOPE	帮手
胆固醇		帮手
N-[1-(2,3-二油酰氧基)丙机]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-月桂氧基己酯鸟氨酸	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-丙胺溴化物	MDRIE	阳离子
二肉豆蔻氧基丙基二甲基羟乙基溴化铵	DMRI	阳离子阳
3β-[N-(N',N'-二甲氨基乙烷)-氨基甲酰基]胆固醇	DC-Chol	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
外消旋-[(2,3-[(2,3-二十八烷氧基丙基)(2-羟乙基羟乙	CLIP-1	阳离子

[4601]

[4602]

用于基因转移的脂质		
脂质	缩写	特征
基)]- 二甲基氯化铵		
外消旋-[2(2,3-二十六烷氧基丙基-氧基甲氧基) 乙基]三甲基溴化铵	CLIP-6	阳离子
乙基二肉豆蔻酰磷脂酰胆碱	DSDMA	阳离子
1,2-二肉豆蔻酰基-三甲基丙烷铵	DMTAP	阳离子
0,0'-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-乙基磷胆碱	DSEPC	阳离子
N-棕榈酰 D-赤型鞘氨醇氨基甲酰精胺	CCS	阳离子
N-叔丁基-NO-十四烷基-3-十四烷基氨基丙脒	二 C14-	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	RPR209120	阳离子
1,2-二亚油氧基-3-二甲氨基丙烷	DLinDMA	阳离子
2,2-二亚油基-4-二甲氨基乙基-[1,3]-二氧戊环	DLin-KC2-DMA	阳离子
二亚油基-甲基-4-二甲氨基丁酸酯	DLin-KC3-DM	阳离子

[4603] 表11列出用于基因转移和/或纳米颗粒制剂的实施例性聚合物。

[4604] 表11

[4605]

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
组氨酸修饰的 PLL	

[4606]

用于基因转移的聚合物	
聚合物	缩写
聚(N-乙烯基吡咯烷酮)	PVP
聚(丙烯亚胺)	PPI
聚(酰氨基胺)	PAMAM
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
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

[4607] 表12总结编码本文所述融合蛋白的多核苷酸的递送方法。

[4608] 表12

[4609]

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	提供的分子类型
身体的	(例如，电穿孔、粒子	是	短暂的	不	核酸和蛋白质

[4610]

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	提供的分子类型
病毒性	枪、磷酸钙转染逆转录病毒	不是	稳定的	是	核糖核酸
	慢病毒	是	稳定的	是/否，有修饰	核糖核酸
	腺病毒	是	短暂的	不	脱氧核糖核酸
	腺相关病毒 (AAV)	是	稳定的	不	脱氧核糖核酸
	痘苗病毒	是	非常瞬态	不	脱氧核糖核酸
	单纯疱疹病毒	是	稳定的	不	脱氧核糖核酸
非病毒	阳离子脂质体	是	短暂的	取决于递送的内容	核酸和蛋白质
	聚合物纳米粒子	是	短暂的	取决于递送的内容	核酸和蛋白质
生物非病毒	减毒细菌	是	短暂的	不	核酸
递送载体	工程化噬菌体	是	短暂的	不	核酸
身体的	哺乳动物病毒样颗粒	是	短暂的	不	核酸
病毒性的	生物脂质体：红细胞幽灵和外泌体	是	短暂的	不	核酸

[4611] 在另一方面,基因组编辑系统组分或编码此类组分的核酸(例如核酸结合蛋白,诸如例如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。RNP包括与靶向gRNA形成复合物的核酸结合蛋白(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.et al.,2015,Nat.Biotechnology,33(1):73-80中所报导者。RNP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,由于包括核酸结合蛋白和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)和向导同源性定向修复(HDR)。

[4612] 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,所述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。

[4613] 可以使用任何合适的启动子来驱动碱基编辑器和在适当情况下向导核酸的表达。对于无处不在的表达,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:用于所有神经元的突触素I、用于兴奋性神经元的CaMKIIalpha 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。

[4614] 在一些实施方案中,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括与编码碱基编辑器的核酸可操作连接的第一启动子和与向导核酸可操作连接的第二启动子。

[4615] 用于驱动向导核酸表达的启动子可以包括:Pol III启动子,诸如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)。

[4616] 病毒载体

[4617] 因此,本文所述碱基编辑器可以与病毒载体一起递送。在一些实施方案中,本文公开的碱基编辑器可以在病毒载体中包括的核酸上编码。在一些实施方案中,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。在其他实施方案中,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,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通过所选病毒载体的货物大小限制来确定。

[4618] 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。病毒载体可以直接给药于培养中的细胞、患者(体内),或者可以用于体外处理细胞,并且可以任选地将修饰的细胞给药于患者(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。

[4619] 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)和腺病毒载体。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号8,454,972(腺病毒的制剂、剂量)、美国专利号8,404,658(AAV的制剂、剂量)的制剂和剂量)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号5,846,946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(例如人类成年男性),并且可以针对不同体重和物种的患者、受试者、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,包括患者或受试者的年龄、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。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,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。

[4620] 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,其包装容量高达6至10kb的外源序列。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(参见,例如,Buchscher et al.,J.Virol.66:2731-2739(1992);Johann et al.,J.Virol.66:1635-1640(1992);Sommnerfelt et al.,Virol.176:58-59(1990);Wilson et al.,J.Virol.63:2374-2378(1989);Miller et al.,J.Virol.65:2220-2224(1991);PCT/US94/05700)。

[4621] 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以使得能够通过逆转录病毒载体有效包装和递送到目标细胞中。在一些实施方案中,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。

[4622] 在首选瞬时表达的应用中,可以使用基于腺病毒的系统。基于腺病毒的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。所述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(参见,例如,West et al.,Virology160:38-47(1987);U.S.Patent No.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 et al.,Mol.Cell.Biol.5:3251-3260(1985);Tratschin,et al.,Mol.Cell.Biol.4:2072-2081(1984);Hermonat&Muzyczka,PNAS 81:6466-6470(1984);以及Samulski et al.,J.Virol.63:03822-3828(1989)。

[4623] 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,两侧各有145碱基对反向末端重复序列(ITR)。病毒粒子由三种衣壳蛋白Vp1、Vp2和Vp3所组成,它们以1:1:10的比率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,并且参与定义病毒向性的细胞表面受体识别。已在Vp1的独特N端识别出一个在病毒感染性中起作用的磷脂酶结构域。

[4624] 与野生型AAV类似,重组AAV(rAAV)利用顺式作用的145碱基对ITR位于载体转基因盒的侧边,提供高达4.5kb的外源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。尽管有许多使用所述系统在体外和体内成功的rAAV例子,但当基因编码序列的长度等于或大于AAV介导

的基因递送时,有限的包装能力限制了AAV介导的基因递送的使用。wt AAV基因组。

[4625] 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。在一些实施方案中,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。在一些实施方案中,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。

[4626] AAV的包装容量约为4.5Kb或4.75Kb,包含两个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的构建体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,很难包装成AAV。因此,本公开的实施方案包括利用长度比常规碱基编辑器短的公开碱基编辑器。在一些实施例中,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、3.0kb、2.9kb、2.8kb、2.7kb、2.6kb、2.5kb、2kb或1.5kb。在一些实施方案中,所公开的碱基编辑器的长度为4.5kb或更小。

[4627] 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.et al,J.Virol.82:5887-5911(2008)中发现)。

[4628] 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。

[4629] 慢病毒可以如下制备。克隆pCasES10(其含有慢病毒转移质粒骨架)后,将低传代(p=5)的HEK293FT接种到T-75烧瓶中至50%汇合,前一天在含有10%胎儿牛血清且不含抗生素的DMEM中转染。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的pMD2.G(VSV-g假型)和7.5μg的psPAX2(gag/pol/rev/tat)。可以在4mL OptiMEM中使用阳离子脂质递送剂(50μl Lipofectamine2000和100ul Plus试剂)进行转染。6小时后,将培养基更换为含10%胎儿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。

[4630] 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℃重悬过夜。然后等分并立即在-80℃下冷冻。

[4631] 在另一个实施方案中,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。在另一个实施方案中,RetinoStat.RTM.是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。在另一个实施方案中,考虑使用自我失活的慢病毒载体。

[4632] 所述系统的任何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。可以使用体外转录产生碱基编辑器编码mRNA。例如,可以使用包括以下组件的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球

蛋白-多聚苷酸尾的3'UTR)。所述盒可用于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括T7启动子的盒中转录,然后是序列"GG"和向导多核苷酸序列。

[4633] 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修饰的核苷,例如使用伪U或5-甲基-C。

[4634] 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,可以通过将一个或多个待递送的蛋白质分成两个或多个片段来解决这些挑战,其中所述N端片段与分裂的内含物肽-N融合,及所述C端片段与分裂的内含肽-C融合。然后将这些片段包装成两个或多个AAV载体。如本文所用,"内含肽"是指连接侧边N端和C端外显子(例如,待连接的片段)的自剪接蛋白质内含子(例如,肽)。例如,Wood et al., J.Biol.Chem.289(21);14512-9(2014)中所述。例如,当与分离的蛋白质片段融合时,含内肽IntN和IntC相互识别,将自身剪接出来,同时连接它们所融合的蛋白质片段的侧边N端和C端外显肽,从而重新构建一个来自两个蛋白质片段的全长蛋白质。其他合适的内含肽对本领域技术人员来说将是显而易见的。

[4635] 本发明的融合蛋白片段的长度可以不同。在一些实施方案中,蛋白质片段的长度为2个氨基酸至约1000个氨基酸。在一些实施方案中,蛋白质片段的长度为约5个氨基酸至约500个氨基酸。在一些实施方案中,蛋白质片段的长度范围为约20个氨基酸至约200个氨基酸。在一些实施方案中,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。

[4636] 在一个实施方案中,双AAV载体是通过将一个大的转基因表达盒分成两半(5'和3'端,或头和尾)而产生的,其中盒的每一半都被包装在一个AAV载体(<5KB)。然后通过两种双AAV载体共感染同一细胞,然后实现全长转基因表达盒的重新组装,随后:(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。

[4637] 内含肽

[4638] 在一些实施方案中,核酸酶(例如,Cas9)的一部分或片段融合内含肽上。核酸酶可以与内含肽的N端或C端融合。在一些实施方案中,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。内含肽、核酸酶和衣壳蛋白可以任意排列融合在一起(如核酸酶-内含袖-衣壳、内含袖-核酸酶-衣壳、衣壳-内含肽-核酸酶等)。在一些实施方案中,内含肽的N端与融合蛋白的C端融合,内含肽的C端与AAV衣壳蛋白的N端融合。

[4639] 内含蛋白(intervening protein)是存在于各种不同生物体中的自动加工结构域,它们进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内含肽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。

[4640] 在蛋白质剪接中,内含肽通过切割两个肽键将自身从前体多肽中分离出来,从而通过形成新的肽键连接侧边外显肽(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是自发发生的,只需要内含肽结构域的折叠。

[4641] 约5%的内含肽是分裂的内含肽,它们被转录和翻译为两个独立的多肽(N-内含臂

和C-内含臂),每一个都融合为一个外显肽。翻译时,内含袖片段自发地非共价组装成规范的内含袖结构,进行蛋白质反式剪接。蛋白质剪接的机制需要一系列酰基转移反应,导致在内含肽-外显肽连接处的两个肽键断裂,并且在N-和C-外显肽之间形成新的肽键。这个过程是通过激活连接内含肽的N-外显肽和N端的肽键启动的。几乎所有的内含含物在它们的N端都有一个半胱氨酸或丝氨酸,它们攻击C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,所述中间体通过第一个C-外显肽残基(+1)的亲核攻击进行反式(硫)酯化,所述残基是半胱氨酸、丝氨酸或苏氨酸。由此产生的支化(硫)酯中间体通过独特的转换解析:内含酚的高度保守的C端天冬酰胺的环化。所述过程由组氨酸(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,留下通过非肽键连接的外肽。这种结构以不依赖于内含链的方式迅速重排成稳定的肽键。

[4642] 在一些实施方案中,碱基编辑器(例如,ABE、CBE)的N端片段与分裂的内含肽-N融合,并且C端片段与内含肽-C融合。然后将这些片段包装成两个或多个AAV载体。例如,Wood et al.,J.Biol.Chem.289(21);14512-9(2014)中所述。例如,当与分离的蛋白质片段融合时,含内肽IntN和IntC相互识别,将自身剪接出来,同时连接它们所融合的蛋白质片段的侧边N端和C端外显肽,从而重新构建一个来自两个蛋白质片段的全长蛋白质。其他合适的内含含乳对本领域技术人员来说将是显而易见的。

[4643] 在一些实施方案中,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端与一个内含肽-N融合,并且每个片段的C端与在氨基酸位置S303、T310、T313、S355、A456、S460、A463、T466、S469、T472、T474、C574、S577、A589及S590内含肽N融合,在以下的顺序中以粗体大写字母表示。

[4644]
```
  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
```

[4645]
```
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
1261 qhkhyldeii eqisefskrv iladanldkv lsaynkhrdk pireqaenii hlftltnlga
1321 paafkyfdtt idrkrytstk evldatlihq sitglyetri dlsqlggd
```

[4646] 使用核碱基编辑器靶向突变

[4647] 如本文所述评估靶向突变的核碱基编辑器的适用性。在一个实施方案中,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,包括永生化人类细胞系,例如293T、K562或U20S。或者,可以使用原代细胞(例如,人类)。这样的细胞可能与最终的细胞目标相关。

[4648] 可以使用病毒载体进行递送。在一个实施方案中,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括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。

[4649] 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Sanger测序,纯化的PCR扩增子被克隆到质粒骨架中,转换、小量制备并用单个引物测序。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,可以将下一代测序连接子和条形码(例如Illumina多重连接子和索引)添加到扩增子的末端,例如用于高通量测序(例如在Illumina MiSeq上)。

[4650] 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。

[4651] 在特定实施方案中,核碱基编辑器用于靶向感兴趣的多核苷酸。在一个实施方案中,将本发明的核碱基编辑器连同用于靶向感兴趣突变的向导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)在细胞的基因组内,从而改变突变。在一些实施方案中,碱基编辑器被向导RNA靶向以向感兴趣基因的序列引入一个或多个编辑。

[4652] 在一个实施方案中,核碱基编辑器用于靶向调控序列,包括但不限于剪接位点、增强子和转录调控组件。然后使用本领域已知的任何方法测定改变对受调控组件控制的基因表达的影响。

[4653] 在又其他实施方案中,本发明的核碱基编辑器用于靶向生物体基因组内的感兴趣的多核苷酸。在一个实施方案中,将本发明的核碱基编辑器连同用于平铺细胞基因组内的

多种序列的向导RNA文库一起递送至细胞,从而系统地改变整个基因组中的序列。

[4654] 所述系统可以包括一种或多种不同的载体。在一个方面,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。

[4655] 一般而言,密码子优化是指通过替换至少一个密码子(例如约或多于约1、2、3、4、5、10、15、20、25、50或更多密码子)与在所述宿主细胞的基因中更频繁或最频繁使用的密码子同时保持天然氨基酸序列。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。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(Codon Usage Database)"(2002年7月9日浏览),这些表可以通过多种方式进行调整。参见Nakamura,Y.,et al."Codon usage tabulated from the international DNA sequence databases:status for the year 2000"Nucl.Acids Res.28:292(2000)。用于密码子优化用于在特定宿主细胞中表达的特定序列的计算机算法也是可用的,例如Gene Forge(Aptagen;Jacobus,PA)也是可用的。在一些实施方案中,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)对应于最常使用的特定氨基酸的密码子。

[4656] 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个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。载体通常包括包装和随后整合到宿主中所需的最少病毒序列,其他病毒序列被用于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提供。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可以包装在细胞系中,所述细胞系包括编码其他AAV基因(即rep和cap)的辅助质粒,但缺少ITR序列。细胞系也可以用腺病毒作为辅助感染。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,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。

[4657] 多效应核碱基编辑器的应用

[4658] 多效应核碱基编辑器可用于靶向感兴趣的多核苷酸以产生改变蛋白质表达的改变。在一个实施方案中,多效应核碱基编辑器用于修饰非编码或调控序列,包括但不限于剪接位点、增强子及转录调控组件。然后使用本领域已知的任何方法测定改变对受调控组件控制的基因表达的影响。在一个特定实施方案中,多效应核碱基编辑器能够显着改变调节序列,从而消除其调节基因表达的能力。有利地,与其他RNA可编程的核酸酶相比,这可以在基因组目标序列中不产生双链断裂的情况下完成。

[4659] 多效应核碱基编辑器可用于靶向感兴趣的多核苷酸以产生改变蛋白质活性的改变。例如,在诱变的背景下,多效应核碱基编辑器比容易出错的PCR和其他基于聚合酶的方法具有许多优势。因为本发明的多效应核碱基编辑器在目标区域的多个碱基处产生改变,所以相对于通过易错PCR引入的突变,此类突改变有可能在蛋白质水平表达,后者不太可能在蛋白质水平表达。考虑到密码子中的单个核苷酸变化可能仍编码相同的氨基酸(例如,密

码子简并性)。与在整个多核苷酸中诱导随机改变的易错PCR不同,本发明的多效应核碱基编辑器可用于靶向感兴趣蛋白质的小区域或限定区域内的特定氨基酸。

[4660] 在其他实施方案中,本发明的多效应核碱基编辑器用于靶向生物体基因组内的感兴趣的多核苷酸。在一个实施方案中,生物体是微生物组(microbiome)的细菌(例如,拟杆菌门(Bacteriodetes)、疣微菌门(Verrucomicrobia)、厚壁菌门(Firmicutes);γ-变形菌门(Gammaproteobacteria)、α变形菌门(Alphaproteobacteria)、拟杆菌门、梭状芽孢杆菌(Clostridia)、红皮杆菌门(Erysipelotrichia)、芽孢杆菌;肠杆菌门(Enterobacteriales)、拟杆菌门、疣状微菌、梭菌目(Clostridiales)、红皮杆菌门、乳杆菌门(Lactobacillales)、肠杆菌科(Enterobacteriaceae)、拟杆菌科(Bacteriodacease)、红皮杆菌门、普雷沃氏菌属(Prevotellaceae)、科里杆菌科(Coriobacteriaceae)和产碱菌科(Alcaligenaceae)、埃希氏菌属(Escherichia)、拟杆菌属、另枝菌属(Alistipes)、艾克曼菌(Akkermansia)、杆菌属(Clostridium)、乳杆菌属)。在另一个实施方案中,生物体是农业上重要的动物(例如,牛、绵羊、山羊、马、鸡、火鸡)或植物(例如,大豆、小麦、玉米、水稻、烟草、苹果、葡萄、桃子、李子、樱桃)。在一个实施方案中,将本发明的多效应核碱基编辑器连同用于平铺细胞基因组内的多种序列的向导RNA文库一起递送至细胞,从而系统地改变整个基因组中的序列。

[4661] 可以在多种蛋白质中的任一种中进行突变以促进结构功能分析或改变蛋白质的内源活性。例如,可以在酶(例如激酶、磷酸酶、羧化酶、磷酸二酯酶)或酶底物、受体或其配体以及抗体及其抗原中进行突变。在一个实施方案中,多效应核碱基编辑器靶向编码酶活性位点、受体配体结合位点或抗体互补决定区(CDR)的核酸分子。在酶的情况下,在活性位点诱导突变可以增加、减少或消除酶的活性。突变对酶的影响在酶活性测定中表征,包括本领域技术人员已知和/或显而易见的多种测定中的任一种。在受体的情况下,在配体结合位点发生的突变可以增加、减少或消除受体对其配体的亲和力。在受体/配体结合测定法中测定此类突变的影响,包括本领域技术人员已知和/或显而易见的多种测定法中的任一种。

[4662] 药物组合物

[4663] 本公开的其他方面涉及包括本文所述任何碱基编辑器、融合蛋白或融合蛋白-向导多核苷酸复合物的药物组合物。在一些实施方案中,药物组合物进一步包括药学上可接受的载体。在一些实施方案中,药物组合物包括另外的药剂(例如,用于特异性递送、增加半衰期或其他治疗化合物)。

[4664] 合适的药学上可接受的载体通常包括惰性物质,这些惰性物质有助于将药物组合物给予受试者,有助于将药物组合物加工成可递送的制剂,或有助于在给药前储存药物组合物。药学上可接受的载体可以包括能够稳定、优化或以其他方式改变制剂的形式、稠度、粘度、pH、药代动力学、溶解度的试剂。

[4665] 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,诸如乳糖、葡萄糖及蔗糖;(2)淀粉,诸如玉米淀粉和马铃薯淀粉;(3)纤维素及其源物,诸如羧甲基纤维素钠、甲基纤维素、乙基纤维素、微晶纤维素及醋酸纤维素;(4)黄蓍胶粉;(5)麦芽;(6)明胶;(7)润滑剂,诸如硬脂酸镁、十二烷基硫酸钠、及滑石粉;(8)赋形剂,诸如可可脂和栓剂蜡;(9)油类,诸如花生油、棉籽油、红花油、香油、橄榄油、玉米油及豆油;(10)二醇类,诸如丙二醇;(11)多元醇,诸如甘油、山梨糖醇、甘露糖醇及聚乙二醇(PEG);(12)酯类,诸如油

酸乙酯、月桂酸乙酯;(13)琼脂;(14)缓冲剂,诸如氢氧化镁和氢氧化铝;(15)海藻酸;(16)无热原水;(17)等渗盐水;(18)林格液;(19)乙醇;(20)pH缓冲溶液;(21)聚酯、聚碳酸酯和/或聚酐;(22)填充剂,诸如多肽和氨基酸;(23)血清醇,诸如乙醇;及(23)用于药物制剂的其他无毒兼容物质。缓冲剂、润湿剂、乳化剂、稀释剂、包封剂、皮肤渗透促进剂、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以存在于制剂中。例如,载体可包含但不限于盐水、缓冲盐水、葡萄糖、精氨酸、蔗糖、水、甘油、乙醇、山梨糖醇、葡聚糖、羧甲基纤维素钠及其组合。

[4666] 药物组合物可包括一种或多种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。或者,pH缓冲化合物优选是将制剂的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。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。

[4667] 药物组合物还可包括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)调节至血流和血细胞可接受的水平的化合物接收个体。渗透调节剂可以是不螯合钙离子的试剂。渗透调节剂可以是本领域技术人员已知或可获得的调节制剂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,诸如氯化钠和乙酸钠;糖类,诸如蔗糖、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。一种或多种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。

[4668] 在一些实施方案中,药物组合物被配制用于递送至受试者,例如用于基因编辑。在一些实施方案中,本文考虑的药物组合物的给药可以使用常规技术进行,包括但不限于输注、输注或肠胃外。在一些实施方案中,肠胃外给药包括血管内、静脉内、肌肉内、动脉内、鞘内、瘤内、皮内、腹膜内、经气管、皮下、表皮下、关节内、囊下、蛛网膜下和胸骨内输注或注射。在一些实施方案中,本文所述药物组合物的合适给药途径包括但不限于:局部、皮下、经皮、皮内、病灶内、关节内、腹膜内、膀胱内、经粘膜、牙龈、齿内、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、瘤内、脑内和脑室内给药。

[4669] 在一些实施方案中,将本文所述药物组合物局部给药至患病部位(例如,肿瘤部位)。在一些实施方案中,本文所述药物组合物通过注射、通过导管、通过栓剂或通过植入物给予受试者,植入物为多孔、无孔或凝胶状材料,包括膜,诸如唾液酸膜,或纤维。

[4670] 在其他实施方案中,本文所述药物组合物在控释系统中递送。在一个实施方案中,可以使用泵(参见,例如,Langer,1990,Science 249:1527-1533;Sefton,1989,CRC Crit.Ref.Biomed.Eng.14:201;Buchwald et al.,1980,Surgery 88:507;Saudek et al,1989,N.Engl.J.Med.321:574)。在另一个实施方案中,可以使用聚合材料。(参见,例如,Medical Applications of Controlled Release(Langer and Wise eds.,CRC Press,Boca Raton,Fla.,1974);Controlled Drug Bioavailability,Drug Product Design and Performance(Smolen and Balleds.,Wiley,New York,1984);Ranger and Peppas,1983,Macromol.Sci.Rev.Macromol.Chem.23:61.See also Levy et al.,1985,Science 228:190;During et al.,1989,Ann.Neurol.25:351;Howard et ah,1989,J.Neurosurg.71:

105。)其他控释系统在例如Langer,supra中有讨论。

[4671] 在一些实施方案中,根据常规程序将药物组合物配制成适于静脉内或皮下给药至受试者例如人的组合物。在一些实施方案中,用于通过注射给药的药物组合物是无菌等渗用途的溶液,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。当药物组合物通过注射给药时,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。

[4672] 用于全身给药的药物组合物可以是液体,例如无菌盐水、乳酸林格氏液或汉克氏液。此外,药物组合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。药物组合物可包括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,只要其中包括组合物。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至10摩尔%)阳离子脂质的"稳定质粒-脂质颗粒"(stable plasmid-lipid particle,SPLP)中,并通过聚乙二醇(PEG)涂层稳定(Z Zhang Y.P.et ah,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。参见,例如,美国专利号4,880,635;4,906,477;4,911,928;4,917,951;4,920,016;及4,921,757;其中各者通过引用方式并入本文。

[4673] 例如,本文所述药物组合物可以作为单位剂量给药或包装。当用于本公开的药物组合物时,术语"单位剂量"是指适合作为受试者的单位剂量的物理上离散的单位,每个单位包括经计算以产生期望的治疗效果的预定量的活性物质。所需的稀释剂;即,承运人或车辆。

[4674] 此外,药物组合物可以作为药物试剂盒提供,其包括(a)含有冻干形式的本发明化合物的容器和(b)含有药学上可接受的稀释剂(例如,用于重构或稀释所述稀释剂的无菌容器)的第二个容器。本发明的冻干化合物。任选地与此类容器相关联的可以是由管理药物或生物产品的制造、使用或销售的政府机构规定的形式的通知,所述通知反映了制造、使用机构的批准或出售用于人类管理。

[4675] 在另一方面,包括含有可用于治疗上述疾病的材料的制品。在一些实施方案中,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。容器可由多种材料制成,例如玻璃或塑料。在一些实施方案中,容器容纳有效治疗本文所述疾病的组合物并且可以具有无菌进入口。例如,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。组合物中的活性剂是本发明的化合物。在一些实施方案中,容器上或与容器相关的标签表示组合物用于治疗选择的疾病。制品可进一步包括第二容器,其包括药学上可接受的缓冲液,例如磷酸盐缓冲盐水、林格氏液或葡萄糖溶液。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。

[4676] 在一些实施方案中,本文所述的任何融合蛋白、gRNA、和/或复合物作为药物组合物的一部分提供。在一些实施方案中,药物组合物包括本文所提供的任何融合蛋白。在一些实施方案中,药物组合物包括本文所提供的任何复合物。在一些实施方案中,药物组合物包

括核糖核蛋白复合物,所述复合物包括与gRNA和阳离子脂质形成复合物的RNA向导的核酸酶(例如Cas9)。在一些实施方案中,药物组合物包括gRNA、核酸可编程的DNA结合蛋白、阳离子脂质和药学上可接受的赋形剂。药物组合物可任选地包括一种或多种额外的治疗活性物质。

[4677] 在一些实施方案中,将本文所提供的组合物给药于受试者,例如对人类给药类受试者,以在受试者内实现靶向基因组修饰。在一些实施方案中,细胞获自受试者并与本文所提供的任何药物组合物接触。在一些实施方案中,任选地在细胞中实现或检测到所需基因组修饰之后,从受试者取出并离体与药物组合物接触的细胞重新引入受试者。递送包括核酸酶的药物组合物的方法是已知的,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;及7,163,824,其内容通过引用整体并入本文。尽管本文所提供的药物组合物的描述主要涉及适合于对人类给药的药物组合物,但本领域技术人员将理解此类组合物通常适合于给药于各种动物或生物体,例如,兽医用。

[4678] 为使组合物适合于对各种动物给药而对适合对人给药的药物组合物进行修饰是众所周知的,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。考虑给予药物组合物的受试者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,诸如牛、猪、马、羊、猫、狗、小鼠、和/或大鼠;和/或鸟类,包括商业相关鸟类,诸如鸡、鸭、鹅、和/或火鸡。

[4679] 本文所述药物组合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。药物制剂可另外包括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;其全部内容通过引用方式并入本文)公开了用于配制药物组合物的各种赋形剂及其制备的已知技术。还参见PCT申请PCT/US2010/055131(公开号WO2011/053982A8,2010年11月2日提交),其整体内容通过引用并入本文,用于生产包括核酸酶的药物组合物的其他合适的方法、试剂、赋形剂和溶剂。

[4680] 除非任何常规赋形剂介质与物质或其源物不兼容,例如通过产生任何不希望的生物效应或以有害方式与药物组合物的任何其他成分相互作用,否则其用途被认为是在本公开的范围。

[4681] 如上所述组合物可以有效量给药。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、受试者的年龄和身体状况、同时治疗的性质(如果有的话)以及医师公知的类似因素。对于治疗应用,所述量足以达到医学上所需的结果。

[4682] 在一些实施方案中,根据本公开的组合物可用于治疗多种疾病、障碍、和/或病症中的任一种。

[4683] 治疗方法

[4684] 还提供治疗疾病或病症和/或引起疾病或病症的基因突变的方法。这些方法包括向受试者(例如,哺乳动物,诸如人类)给药治疗有效量的药物组合物,所述药物组合物包括编码本文所述碱基编辑器系统(例如,碱基编辑器和gRNA)的多核苷酸。在一些实施方案中,碱基编辑器是包括napDNAbp结构域和腺苷脱氨酶结构域或胞苷脱氨酶结构域的融合蛋白。受试者的细胞用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以实现A·T至G·C的改变(若细胞用腺苷脱氨酶结构域转导)或C·G至U·A改变(若细胞用胞苷脱氨酶结构域转导)含有感兴趣基因突变的核酸序列。

[4685] 本文的方法包括向受试者(包括识别为需要此类治疗的受试者,或怀疑处于疾病风险中并需要此类治疗的受试者)给药有效量的本文所述组合物。识别需要此类治疗的受试者可以由受试者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)。

[4686] 治疗方法通常包括给药治疗有效量的药物组合物,所述药物组合物包括例如编码碱基编辑器的载体和靶向有需要的受试者(例如人类患者)的感兴趣基因的gRNA。其中。此类治疗将适当地给药于患有、患有、易患所述疾病或病症或处于所述疾病或病症风险中的受试者,特别是人类受试者。

[4687] 在一个实施方案中,本发明提供一种监测治疗进展的方法。所述方法包括确定患有或易患疾病、病症或病症的受试者中诊断标志物(Marker)(例如,与疾病或病症相关的SNP)或诊断测量(例如,筛选、测定)的水平的步骤。其中受试者已给予足以治疗疾病或其症状的治疗量的本文组合物。在所述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,以确定受试者的疾病状态。在优选的实施方案中,在确定第一水平之后的时间点确定受试者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。在某些优选的实施方案中,在开始根据本发明的治疗之前确定受试者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受试者中的标志物水平进行比较,以确定治疗的功效。

[4688] 在一些实施方案中,细胞获自受试者并与本文所提供的药物组合物接触。在一些实施方案中,任选地在细胞中实现或检测到所需基因组修饰之后,从受试者取出并离体与药物组合物接触的细胞重新引入受试者。递送包括核酸酶的药物组合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;及7,163,824,其所有公开内容通过引用整体并入本文。尽管本文所提供的药物组合物的描述主要涉及适合于对人类给药的药物组合物,本领域技术人员将理解此类组合物通常适合于给药于各种动物或生物体,例如,兽医用途。

[4689] 本文提供用本文所述组合物或系统,例如碱基编辑器系统或碱基编辑器蛋白质治疗疾病或病症的方法。任选地,本文所述碱基编辑器系统可以与一种或多种其他处理组合。一方面,本文提供一种通过给药本文所述碱基编辑器来治疗有需要的受试者的斯塔加特氏病(SD)的方法。一方面,本文提供一种通过给药本文所述碱基编辑器来治疗有需要的受试者的帕金森氏病(PD)的方法。一方面,本文提供一种通过给药本文所述碱基编辑器来治疗有需要的受试者的蕾特氏症(RTT)的方法。一方面,本文提供一种通过给药本文所述碱基编辑器来治疗有需要的受试者的贺勒氏症(HS)的方法。

[4690] 个体受试者的反应可以表征为完全反应、部分反应或稳定的疾病。在一些实施方

案中,响应是部分响应(partial response,PR)。在一些实施方案中,响应是完全响应(complete response,CR)。在一些实施方案中,响应导致受试者的无进展存活(例如,稳定的疾病)。

[4691] 在一些实施方案中,如果人类受试者没有用化合物治疗,例如,与人类受试者的预期存活时间相比,治疗导致人类受试者的存活时间增加。在一些实施方案中,与用包括ABE7(例如,ABE7.10)的碱基编辑器治疗的受试者相比,延长的存活时间包括疾病进展更慢。

[4692] 在一些实施方案中,待用所述方法治疗的人类受试者是儿童(例如,0至18岁)。在其他实施方案中,待用所述方法治疗的人类受试者是成年人(例如,18岁以上)。

[4693] 斯塔加特氏病(SD)

[4694] 斯塔加特氏病(也称为斯塔加特氏病黄斑营养不良、青少年黄斑变性或黄色斑点沉着(fundus flavimaculatus))是视网膜(即,眼睛后部感知光线的组织)的一种遗传性疾病。斯塔加特氏病是导致黄斑变性的几种遗传性疾病之一。这种疾病通常会导致儿童期或青春期视力下降;尽管在某些情况下可能直到成年后期才注意到视力丧失。这种疾病很少发展为完全失明。通常,随着时间的推移,随着黄斑的进行性损伤(退化),视力丧失会随着时间的推移缓慢进展至20/200或更糟。在一种情况下,要用本文所述方法治疗的斯塔加特氏病包括幼年斯塔加特氏病。在另一种情况下,要用本文所述方法治疗的斯塔加特氏病包括迟发的斯塔加特氏病。在另一种情况下,要用本文所述方法治疗的斯塔加特氏病包括斯塔加特型显性黄斑营养不良。在另一种情况下,待用本文所述方法治疗的斯塔加特氏病包括显性斯塔加特样黄斑营养不良。

[4695] 斯塔加特氏病的症状进展可能因每位患者而异。较早发病的患者通常会出现更快的视力丧失。视力丧失最初可能会缓慢下降,然后迅速恶化,直至趋于平稳。大多数斯塔加特氏病患者的视力最终将达到20/200或更差。随着年龄的增长,患有斯塔加特氏病的人也可能开始失去一些周边(侧面)视力。

[4696] 在一些实施方案中,致病性SNP与斯塔加特氏病相关;任选地,致病性SNP位于ABCA4基因中;任选地,致病突变包括A1038V、L541P、G1961E或其组合。在一些实施方案中,致病性SNP与弹性假黄瘤(Pseudoxanthoma elasticum)相关;任选地,致病性SNP位于ABCC6基因中;并且任选地,致病性突变包括R1141*(无义突变)。在一些实施方案中,致病性SNP与中链酰基辅酶A脱氢酶缺乏症相关;任选地,致病性SNP位于ACADM基因中;并且任选地,致病突变包括K329E。在一些实施方案中,致病性SNP与严重联合免疫缺陷相关;任选地,致病性SNP位于ADA基因中;并且任选地,致病突变包括G216R、Q3*或其组合。

[4697] ABCA4多肽的实施例性氨基酸序列提供如下:

[4698] >sp|P78363|ABCA4_人类视网膜特异性磷脂转运ATPase ABCA4 OS=智人OX=9606GN=ABCA4 PE=1SV=3

[4699] MGFVRQIQLLLWKNWTLRKRQKIRFVVELVWPLSLFLVLIWLRNANPLYSHHECHFPNKAMPSAGMLPWLQGIFCNVNNPCFQSPTPGESPGIVSNYNNSILARVYRDFQELLMNAPESQHLGRIWTELHILSQFMDTLRTHPERIAGRGIRIRDILKDEETLTLFLIKNIGLSDSVVYLLINSQVRPEQFAHGVPDLALKDIACSEALLERFIIFSQRRGAKTVRYALCSLSQGTLQWIEDTLYANVDFFKLFRVLPTLLDSRSQGINLRSWGGILSDMSPRIQEFIHRPSMQDLLWVTRPLMQNGGPETFTKLMGILSDLLCGYPEGGGSRVLSFNWYEDNNYKAFLGIDSTRKDPIYSYDRRTTSFCNALIQSLESNPLTKIAWRAAKPLLMGKILYTPDSPAARRILKNANSTFEELEHVRKLVKAWEEVGPQIWYFFDNSTQM

NMIRDTLGNPTVKDFLNRQLGEEGITAEAILNFLYKGPRESQADDMANFDWRDIFNITDRTLRLVNQYLECLVLDK
FESYNDETQLTQRALSLLEENMFWAGVVFPDMYPWTSSLPPHVKYKIRMDIDVVEKTNKIKDRYWDSGPRADPVED
FRYIWGGFAYLQDMVEQGITRSQVQAEAPVGIYLQQMPYPCFVDDSFMIILNRCFPIFMVLAWIYSVSMTVKSIVL
EKELRLKETLKNQGVSNAVIWCTWFLDSFSIMSMSIFLLTIFIMHGRILHYSDPFILFLFLLAFSTATIMLCFLLS
TFFSKASLAAACSGVIYFTLYLPHILCFAWQDRMTAELKKAVSLLSPVAFGFGTEYLVRFEEQGLGLQWSNIGNSP
TEGDEFSFLLSMQMMLLDAAVYGLLAWYLDQVFPGDYGTPLPWYFLLQESYWLGGEGCSTREERALEKTEPLTEET
EDPEHPEGIHDSFFEREHPGWVPGVCVKNLVKIFEPCGRPAVDRLNITFYENQITAFLGHNGAGKTTTLSILTGLL
PPTSGTVLVGGRDIETSLDAVRQSLGMCPQHNILFHHLTVAEHMLFYAQLKGKSQEEAQLEMEAMLEDTGLHHKRN
EEAQDLSGGMQRKLSVAIAFVGDAKVVILDEPTSGVDPYSRRSIWDLLLKYRSGRTIIMSTHHMDEADLLGDRIAI
IAQGRLYCSGTPLFLKNCFGTGLYLTLVRKMKNIQSQRKGSEGTCSCSSKGFSTTCPAHVDDLTPEQVLDGDVNEL
MDVVLHHVPEAKLVECIGQELIFLLPNKNFKHRAYASLFRELEETLADLGLSSFGISDTPLEEIFLKVTEDSDSGP
LFAGGAQQKRENVNPRHPCLGPREKAGQTPQDSNVCSPGAPAAHPEGQPPPEPECPGPQLNTGTQLVLQHVQALLV
KRFQHTIRSHKDFLAQIVLPATFVFLALMLSIVIPPFGEYPALTLHPWIYGQQYTFFSMDEPGSEQFTVLADVLLN
KPGFGNRCLKEGWLPEYPCGNSTPWKTPSVSPNITQLFQKQKWTQVNPSPSCRCSTREKLTMLPECPEGAGGLPPP
QRTQRSTEILQDLTDRNISDFLVKTYPALIRSSLKSKFWVNEQRYGGISIGGKLPVVPITGEALVGFLSDLGRIMN
VSGGPITREASKEIPDFLKHLETEDNIKVWFNNKGWHALVSFLNVAHNAILRASLPKDRSPEEYGITVISQPLNLT
KEQLSEITVLTTSVDAVVAICVIFSMSFVPASFVLYLIQERVNKSKHLQFISGVSPTTYWVTNFLWDIMNYSVSAG
LVVGIFIGFQKKAYTSPENLPALVALLLLYGWAVIPMMYPASFLFDVPSTAYVALSCANLFIGINSSAITFILELF
ENNRTLLRFNAVLRKLLIVFPHFCLGRGLIDLALSQAVTDVYARFGEEHSANPFHWDLIGKNLFAMVVEGVVYFLL
TLLVQRHFFLSQWIAEPTKEPIVDEDDDVAEERQRIITGGNKTDILRLHELTKIYPGTSSPAVDRLCVGVRPGECF
GLLGVNGAGKTTTFKMLTGDTTVTSGDATVAGKSILTNISEVHQNMGYCPQFDAIDELLTGREHLYLYARLRGVPA
EEIEKVANWSIKSLGLTVYADCLAGTYSGGNKRKLSTAIALIGCPPLVLLDEPTTGMDPQARRMLWNVIVSIIREG
RAVVLTSHSMEECEALCTRLAIMVKGAFRCMGTIQHLKSKFGDGYIVTMKIKSPKDDLLPDLNPVEQFFQGNFPGS
VQRERHYNMLQFQVSSSSLARIFQLLLSHKDSLLIEEYSVTQTTLDQVFVNFAKQQTESHDLPLHPRAAGASRQAQD
(SEQ ID NO:6)靶向在GCTGTGTGTCGAAGTTCGCCCTGGAGAGGTG或GCTGTGTGTCGGAGTTCGCCCTGG
AGAGGTGABCA4基因的序列的向导RNA序列,其中PAM序列带有底线。向导RNA包括一个序列
CACCUCUCCAGGGCGAACUUCGACACACAGC或CACCUCUCCAGGGCGAACUCCGACACACAGC。

[4700] 斯塔加特氏病的一种或多种症状包括但不限于双眼中心视力变化缓慢的灰色、黑色或朦胧斑点;从明暗环境移动时,眼睛需要比平时更长的时间来适应;眼睛可能对强光更敏感;疾病后期出现色盲、视网膜色素上皮(RPE)细胞中诸如A2E的有毒脂褐素的积累、光感受器死亡、11-顺式-视黄醛(11cRAL或视黄醛)的合成增加、视紫质再生增加、脂褐素积累、脂褐素色素的形成、视网膜变性、废物的产生、A2E(和A2E相关分子)的形成、A2E(和A2E相关分子)的积累、脉络膜新生血管形成、脉络膜视网膜萎缩或其组合。受试者可能表现出斯塔加特氏病的一种或多种症状的改善。在一个实施方案中,一种或多种症状的改善至少为5%。在另一个实施方案中,一种或多种症状的改善是至少10%。在另一个实施方案中,一种或多种症状的改善是至少15%。在另一个实施方案中,一种或多种症状的改善是至少20%。在另一个实施方案中,一种或多种症状的改善是至少25%。在另一个实施方案中,一种或多种症状的改善是至少30%。在另一个实施方案中,一种或多种症状的改善是至少35%。在另一个实施方案中,一种或多种症状的改善是至少40%。在另一个实施方案中,一种或多种症

状的改善是至少50％。在另一个实施方案中,一种或多种症状的改善是至少60％。在另一个实施方案中,一种或多种症状的改善是至少70％。在另一个实施方案中,一种或多种症状的改善是至少75％。在另一个实施方案中,一种或多种症状的改善是至少80％。在另一个实施方案中,一种或多种症状的改善是至少85％。在另一个实施方案中,一种或多种症状的改善是至少90％。在另一个实施方案中,一种或多种症状的改善是至少95％。

[4701] 帕金森氏病(PD)

[4702] 帕金森氏病(PD)是最常见的运动障碍,影响全球超过600万人。PD可表现为幼年或早发,但主要累及55岁以上的个体,65岁以后发病率急剧上升。PD的临床特征包括运动迟缓、姿势不稳、静止性震颤和僵硬等。主要与黑质(substantia nigra,SN)致密部中多巴胺能神经元的逐渐丧失有关。据信,在正常衰老过程中,所述区域每年大约有0.1至0.2％的多巴胺能神经元丢失,但这一速度在PD患者中会大大加快,并且当这些神经元丢失约70至80％时就会出现症状。PD的另一个病理特点是在剩余的多巴胺能神经元中存在α-突触核蛋白蛋白包涵体,已知为路易小体(Lewis body,LB)。

[4703] 尽管大多数PD病例是特发性的,但约有10％的病例报告有家族史,并且越来越多的突变与所述疾病的家族性和散发性形式有关。富含亮氨酸的重复激酶2(LRRK2)的常染色体显性G2019S突变是家族性和散发性PD患者最常见的已知原因。

[4704] LRRK2多肽的实施例性氨基酸序列提供如下:

[4705] >sp|Q5S007|LRRK2_人类富含亮氨酸的重复丝氨酸/苏氨酸蛋白激酶2OS＝智人OX＝9606GN＝LRRK2 PE＝1SV＝2

[4706] MASGSCQGCEEDEETLKKLIVRLNNVQEGKQIETLVQILEDLLVFTYSERASKLFQGKNIHVPLLIVL
DSYMRVASVQQVGWSLLCKLIEVCPGTMQSLMGPQDVGNDWEVLGVHQLILKMLTVHNASVNLSVIGLKTLDLLLT
SGKITLLILDEESDIFMLIFDAMHSFPANDEVQKLGCKALHVLFERVSEEQLTEFVENKDYMILLSALTNFKDEEE
IVLHVLHCLHSLAIPCNNVEVLMSGNVRCYNIVVEAMKAFPMSERIQEVSCCLLHRLTLGNFFNILVLNEVHEFVV
KAVQQYPENAALQISALSCLALLTETIFLNQDLEEKNENQENDDEGEEDKLFWLEACYKALTWHRKNKHVQEAACW
ALNNLLMYQNSLHEKIGDEDGHFPAHREVMLSMLMHSSSKEVFQASANALSTLLEQNVNFRKILLSKGIHLNVLEL
MQKHIHSPEVAESGCKMLNHLFEGSNTSLDIMAAVVPKILTVMKRHETSLPVQLEALRAILHFIVPGMPEESREDT
EFHHKLNMVKKQCFKNDIHKLVLAALNRFIGNPGIQKCGLKVISSIVHFPDALEMLSLEGAMDSVLHTLQMYPDDQ
EIQCLGLSLIGYLITKKNVFIGTGHLLAKILVSSLYRFKDVAEIQTKGFQTILAILKLSASFSKLLVHHSFDLVIF
HQMSSNIMEQKDQQFLNLCCKCFAKVAMDDYLKNVMLERACDQNNSIMVECLLLLGADANQAKEGSSLICQVCEKE
SSPKLVELLLNSGSREQDVRKALTISIGKGDSQIISLLLRRLALDVANNSICLGGFCIGKVEPSWLGPLFPDKTSN
LRKQTNIASTLARMVIRYQMKSAVEEGTASGSDGNFSEDVLSKFDEWTFIPDSSMDSVFAQSDDLDSEGSEGSFLV
KKKSNSISVGEFYRDAVLQRCSPNLQRHSNSLGPIFDHEDLLKRKRKILSSDDSLRSSKLQSHMRHSDSISSLASE
REYITSLDLSANELRDIDALSQKCCISVHLEHLEKLELHQNALTSFPQQLCETLKSLTHLDLHSNKFTSFPSYLLK
MSCIANLDVSRNDIGPSVVLDPTVKCPTLKQFNLSYNQLSFVPENLTDVVEKLEQLILEGNKISGICSPLRLKELK
ILNLSKNHISSLSENFLEACPKVESFSARMNFLAAMPFLPPSMTILKLSQNKFSCIPEAILNLPHLRSLDMSSNDI
QYLPGPAHWKSLNLRELLFSHNQISILDLSEKAYLWSRVEKLHLSHNKLKEIPPEIGCLENLTSLDVSYNLELRSF
PNEMGKLSKIWDLPLDELHLNFDFKHIGCKAKDIIRFLQQRLKKAVPYNRMKLMIVGNTGSGKTTLLQQLMKTKKS
DLGMQSATVGIDVKDWPIQIRDKRKRDLVLNVWDFAGREEFYSTHPHFMTQRALYLAVYDLSKGQAEVDAMKPWLF
NIKARASSSPVILVGTHLDVSDEKQRKACMSKITKELLNKRGFPAIRDYHFVNATEESDALAKLRKTIINESLNFK

IRDQLVVGQLIPDCYVELEKIILSERKNVPIEFPVIDRKRLLQLVRENQLQLDENELPHAVHFLNESGVLLHFQDP
ALQLSDLYFVEPKWLCKIMAQILTVKVEGCPKHPKGIISRRDVEKFLSKKRKFPKNYMSQYFKLLEKFQIALPIGE
EYLLVPSSLSDHRPVIELPHCENSEIIIRLYEMPYFPMGFWSRLINRLLEISPYMLSGRERALRPNRMYWRQGIYL
NWSPEAYCLVGSEVLDNHPESFLKITVPSCRKGCILLGQVVDHIDSLMEEWFPGLLEIDICGEGETLLKKWALYSF
NDGEEHQKILLDDLMKKAEEGDLLVNPDQPRLTIPISQIAPDLILADLPRNIMLNNDELEFEQAPEFLLGDGSFGS
VYRAAYEGEEVAVKIFNKHTSLRLLRQELVVLCHLHHPSLISLLAAGIRPRMLVMELASKGSLDRLLQQDKASLTR
TLQHRIALHVADGLRYLHSAMIIYRDLKPHNVLLFTLYPNAAIIAKIADYGIAQYCCRMGIKTSEGTPGFRAPEVA
RGNVIYNQQADVYSFGLLLYDILTTGGRIVEGLKFPNEFDELEIQGKLPDPVKEYGCAPWPMVEKLIKQCLKENPQ
ERPTSAQVFDILNSAELVCLTRRILLPKNVIVECMVATHHNSRNASIWLGCGHTDRGQLSFLDLNTEGYTSEEVAD
SRILCLALVHLPVEKESWIVSGTQSGTLLVINTEDGKKRHTLEKMTDSVTCLYCNSFSKQSKQKNFLLVGTADGKL
AIFEDKTVKLKGAAPLKILNIGNVSTPLMCLSESTNSTERNVMWGGCGTKIFSFSNDFTIQKLIETRTSQLFSYAA
FSDSNIITVVVDTALYIAKQNSPVVEVWDKKTEKLCGLIDCVHFLREVMVKENKESKHKMSYSGRVKTLCLQKNTA
LWIGTGGGHILLLDLSTRRLIRVIYNFCNSVRVMMTAQLGSLKNVMLVLGYNRKNTEGTQKQKEIQSCLTVWDINL
PHEVQNLEKHIEVRKELAEKMRRTSVE(SEQ ID NO:3)帕金森氏病是一种影响运动的进行性神经系统疾病。症状通常是逐渐开始的,有时仅用一只手就开始几乎不明显的震颤。震颤可能很常见,但这种疾病也通常会导致僵硬或运动减慢。在帕金森氏病的早期阶段,受试者的面部可能很少或没有表情。行走时,患者的手臂不得摆动。言语可能会变得柔和或含糊不清。帕金森氏病的症状通常会随着时间的推移而恶化。

[4707] 帕金森氏病的体征和症状因患者而异。早期迹象可能是轻微的,不会引起注意。症状通常从身体的一侧开始,并且通常会在所述侧变得更糟,即使在症状开始影响两侧之后也是如此。帕金森氏病的体征和症状包括但不限于一种或多种震颤、运动缓慢(运动迟缓);僵硬的肌肉;姿势和平衡受损;失去自动运动(例如,进行无意识运动的能力下降,包含眨眼、微笑或走路时摆动手臂);说话的变化(例如,说话轻声、快速、含糊不清或在说话前犹豫;说话可能更单调,而不是通常的语调);书写变化(例如,书写可能变得困难,书写可能看起来很小);血压变化(例如,由于血压突然下降(直立性低血压),患者站立时可能会感到晕眩或头晕);嗅觉功能障碍(例如,患者可能会遇到嗅觉问题);疲劳(例如,许多帕金森氏病患者会失去精力并感到疲劳,尤其是在当天晚些时候);疼痛(例如,一些帕金森氏病患者会在身体的特定部位或整个身体部位出现疼痛);性功能障碍(例如,一些帕金森氏病患者会注意到性欲或性能力下降);或其组合。

[4708] 受试者可能表现出帕金森氏病的一种或多种症状的改善。在一个实施方案中,一种或多种症状的改善至少为5%。在另一个实施方案中,一种或多种症状的改善是至少10%。在另一个实施方案中,一种或多种症状的改善是至少15%。在另一个实施方案中,一种或多种症状的改善是至少20%。在另一个实施方案中,一种或多种症状的改善是至少25%。在另一个实施方案中,一种或多种症状的改善是至少30%。在另一个实施方案中,一种或多种症状的改善是至少35%。在另一个实施方案中,一种或多种症状的改善是至少40%。在另一个实施方案中,一种或多种症状的改善是至少50%。在另一个实施方案中,一种或多种症状的改善是至少60%。在另一个实施方案中,一种或多种症状的改善是至少70%。在另一个实施方案中,一种或多种症状的改善是至少75%。在另一个实施方案中,一种或多种症状的改善是至少80%。在另一个实施方案中,一种或多种症状的改善是至少

85％。在另一个实施方案中，一种或多种症状的改善是至少90％。在另一个实施方案中，一种或多种症状的改善是至少95％。

[4709] 还提供治疗PD和/或LRRK中导致PD的基因突变的方法，包括向受试者(例如哺乳动物，例如人类)给药治疗有效量的药物组合物，所述药物组合物包括编码的多核苷酸本文所述的碱基编辑器系统(例如，碱基编辑器和gRNA)。在一些实施方案中，碱基编辑器是包括多核苷酸可编程的DNA结合结构域和腺苷脱氨酶结构域或胞苷脱氨酶结构域的融合蛋白。受试者的细胞用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以实现含有LRRK基因突变的核酸序列的A·T至G·C的改变(若细胞用腺苷脱氨酶结构域转导)或C·G U·A改变(细胞若用胞苷脱氨酶结构域转导)。

[4710] 本文的方法包括向受试者(包括识别为需要此类治疗的受试者，或怀疑处于疾病风险中并需要此类治疗的受试者)给药有效量的本文所述组合物。识别需要此类治疗的受试者可以由受试者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)。

[4711] 治疗方法通常包括给药治疗有效量的药物组合物，所述药物组合物包括例如编码碱基编辑器的载体和靶向有需要的受试者(例如人类患者)的LRRK2基因的gRNA。此类治疗将适当地给药于患有、具有、易患或有患PD的风险的受试者，特别是人类受试者。本文的组合物还可用于治疗可能涉及PD的任何其他病症。

[4712] 在一些实施方案中，向导RNA在目标序列处靶向LRRK基因GCTCGCCCTTCTTCTTCCCCTGTGA、GTCTTTCCCTCCAGGCTCGCCCTTCTTCTTCCCCTGTGA、TCACAGGGGAAGAAGAAGGGCGAGC、或TCACAGGGGAAGAAGAAGGGCGAGCCTGGAGGGAAAGAC。在一些实施方案中，向导RNA包括序列GCUCGCCCUUCUUCUUCCCCUGUGA、GUCUUUCCCUCCAGGCUCGCCCUUCUUCUUCCCCUGUGA、UCACAGGGGAAGAAGAAGGGCGAGC、或UCACAGGGGAAGAAGAAGGGCGAGCCUGGAGGGAAAGAC。在一些实施方案中，向导RNA包括序列UCCGACUAUAUGAAAUGCCUUAUUUUCCAAUGGGAUUUUGG或UUGCAAAGAUUGCUGACUAGGGCAUUGCUCAGUACUGCUGUAGAAUGG。

[4713] 在一个实施方案中，提供一种监测治疗进展的方法。所述方法包含确定患有或易患与PD相关的病症或其症状的受试者的诊断标志物(例如，与PD相关的SNP)或诊断测量(例如，筛选、测定)的水平的步骤。已向受试者给药足以治疗疾病或其症状的治疗量的本文组合物。在所述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，以确定受试者的疾病状态。在优选的实施方案中，在确定第一水平之后的时间点确定受试者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。在某些优选的实施方案中，在开始根据本发明的治疗之前确定受试者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受试者中的标志物水平进行比较，以确定治疗的功效。

[4714] 在一些实施方案中，将本文所提供的组合物给药于受试者，例如对人类给药类受试者，以在受试者内实现靶向基因组修饰。在一些实施方案中，细胞获自受试者并与本文所提供的任何药物组合物接触。在一些实施方案中，任选地在细胞中实现或检测到所需基因组修饰之后，从受试者取出并离体与药物组合物接触的细胞重新引入受试者。递送包括核酸酶的药物组合物的方法是已知的，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;及7,163,824,其全部公开内容通过引用并入本文。尽管本文所提供的药物组合物的描述主要涉及适合于对人类给药类的药物组合物,但本领域技术人员将理解此类组合物通常适合于给药于各种动物或生物体,例如,兽医用途。

[4715] 贺勒氏症(HS)

[4716] 贺勒氏症是1型粘多糖贮积症(MPS1)的最严重形式,这是一种罕见的常染色体隐性溶酶体贮积症,大约每200,000名新生儿中就有一人发生。MPS1的特点是骨骼异常、认知障碍、心脏病、呼吸系统问题、肝脏和脾脏肿大以及预期寿命缩短。MPS1是由α-L-艾杜糖醛酸酶(IDUA)基因突变引起的,导致对于溶酶体中的糖胺聚糖分解重要的α-L-艾杜糖醛酸酶的缺乏。贺勒氏症患者的标准治疗是骨髓移植,但这种治疗不能校正已经造成的任何损伤,尤其是中枢神经系统,并且存在与移植相关的死亡风险。因此,需要用于治疗贺勒氏症患者的新颖组合物和方法。

[4717] 本公开提供用于治疗与点突变相关或由其引起的贺勒氏症的方法,所述点突变可以通过碱基编辑器介导的基因编辑来校正和/或症状可以治疗或改善。

[4718] 还提供治疗贺勒氏症和/或导致贺勒氏症的IDUA基因中的基因突变的方法,包括向受试者(例如,哺乳动物,例如人类)给药治疗有效量的本文所述包括多核苷酸编码碱基编辑器系统(例如,ABE8碱基编辑器和gRNA)的药物组合物。在一些实施方案中,碱基编辑器是包括多核苷酸可编程的DNA结合结构域和腺苷脱氨酶结构域的融合蛋白。用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受试者的细胞,以实现包括IDUA基因突变的核酸序列的A•T至G•C改变。

[4719] IDUA多肽的实施例性氨基酸序列提供如下:

[4720] >sp|P35475|IDUA_人类α-L-艾杜糖醛酸酶OS=智人OX=9606GN=IDUA PE=1SV=2

[4721] MRPLRPRAALLALLASLLAAPPVAPAEAPHLVHVDAARALWPLRRFWRSTGFCPPLPHSQADQYVLSWDQQLNLAYVGAVPHRGIKQVRTHWLLELVTTRGSTGRGLSYNFTHLDGYLDLLRENQLLPGFELMGSASGHFTDFEDKQQVFEWKDLVSSLARRYIGRYGLAHVSKWNFETWNEPDHHDFDNVSMTMQGFLNYYDACSEGLRAASPALRLGGPGDSFHTPPRSPLSWGLLRHCHDGTNFFTGEAGVRLDYISLHRKGARSSISILEQEKVVAQQIRQLFPKFADTPIYNDEADPLVGWSLPQPWRADVTYAAMVVKVIAQHQNLLLANTTSAFPYALLSNDNAFLSYHPHPFAQRTLTARFQVNNTRPPHVQLLRKPVLTAMGLLALLDEEQLWAEVSQAGTVLDSNHTVGVLASAHRPQGPADAWRAAVLIYASDDTRAHPNRSVAVTLRLRGVPPGPGLVYVTRYLDNGLCSPDGEWRRLGRPVFPTAEQFRRMRAAEDPVAAAPRPLPAGGRLTLRPALRLPSLLLVHVCARPEKPPGQVTRLRALPLTQGQLVLVWSDEHVGSKCLWTYEIQFSQDGKAYTPVSRKPSTFNLFVFSPDTGAVSGSYRVRALDYWARPGPFSDPVPYLEVPVPRGPPSPGNP(SEQ ID NO:4)。

[4722] 本文的方法包括向受试者(包括识别为需要此类治疗的受试者,或怀疑处于疾病风险中并需要此类治疗的受试者)给药有效量的本文所述组合物。识别需要此类治疗的受试者可以由受试者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)。

[4723] 治疗方法通常包括给药治疗有效量的药物组合物,所述药物组合物包括例如编码碱基编辑器的载体和靶向有需要的受试者(例如人类患者)的IDUA基因的gRNA。此类治疗将适当地给药于患有、具有、易患或处于贺勒氏症风险中的受试者,特别是人类受试者。本文的组合物也可用于治疗可能涉及贺勒氏症的任何其他疾病。向导RNA序列可以包括间隔序

列CTTTTCACTTTTCCTGCCGGGG(R255X)、AGCTTCCATGTCCAGCCTTC(R106W)、ACCATGAAGTCAAAATCATT(T158M)或GCTTTCAGCCCCGTTTCTTG(R270X)。

[4724] 在一个实施方案中,本发明提供一种监测治疗进展的方法。所述方法包括确定患有或易患与Hurler相关的病症或其症状的受试者的诊断标志物(例如,与贺勒氏症相关的SNP)或诊断测量(例如,筛选、测定)的水平的步骤。已向受试者给药足以治疗疾病或其症状的治疗量的本文组合物的综合征。在所述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,以确定受试者的疾病状态。在优选的实施方案中,在确定第一水平之后的时间点确定受试者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。在某些优选的实施方案中,在开始根据本发明的治疗之前确定受试者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受试者中的标志物水平进行比较,以确定治疗的功效。

[4725] 在一些实施方案中,细胞获自受试者并与本文所提供的药物组合物接触。在一些实施方案中,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,从受试者取出并离体与药物组合物接触的细胞被重新引入受试者。递送包括核酸酶的药物组合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,其公开内容通过引用整体并入本文。尽管本文所提供的药物组合物的描述主要涉及适合于对人类给药的药物组合物,但本领域技术人员应理解,此类组合物通常适合于给药于各种动物或生物体,例如,兽医用途。

[4726] 待治疗的贺勒氏症的一种或多种症状包括但不限于粗糙的面部特征、角膜混浊、肝肿大、后凸畸形/驼背、疝气、气道相关症状(诸如睡眠障碍/打鼾)、脾肿大、心脏瓣膜异常、认知障碍、多发性营养不良、舌头扩大、关节挛缩、扁桃体扩大或其组合。

[4727] 受试者可能表现出贺勒氏症的一种或多种症状的改善。在一个实施方案中,一种或多种症状的改善至少为5%。在另一个实施方案中,一种或多种症状的改善是至少10%。在另一个实施方案中,一种或多种症状的改善是至少15%。在另一个实施方案中,一种或多种症状的改善是至少20%。在另一个实施方案中,一种或多种症状的改善是至少25%。在另一个实施方案中,一种或多种症状的改善是至少30%。在另一个实施方案中,一种或多种症状的改善是至少35%。在另一个实施方案中,一种或多种症状的改善是至少40%。在另一个实施方案中,一种或多种症状的改善是至少50%。在另一个实施方案中,一种或多种症状的改善是至少60%。在另一个实施方案中,一种或多种症状的改善是至少70%。在另一个实施方案中,一种或多种症状的改善是至少75%。在另一个实施方案中,一种或多种症状的改善是至少80%。在另一个实施方案中,一种或多种症状的改善是至少85%。在另一个实施方案中,一种或多种症状的改善是至少90%。在另一个实施方案中,一种或多种症状的改善是至少95%。

[4728] 蕾特氏症(RTT)

[4729] 还提供治疗蕾特氏症(RTT)和/或导致RTT的Mecp2基因突变的方法,包括向受试者(例如哺乳动物,例如人类)给药治疗有效量的药物组合物,所述药物组合物包括编码本文所述的碱基编辑器系统(例如,碱基编辑器和gRNA)的多核苷酸。在一些实施方案中,碱基编辑器是包括多核苷酸可编程的DNA结合结构域和腺苷脱氨酶结构域或胞苷脱氨酶结构域的

融合蛋白。受试者的细胞用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以实现含有Mecp2基因突变的核酸序列的A·T至G·C的改变(若细胞用腺苷脱氨酶结构域转导)或C·G U·A改变(若细胞用胞苷脱氨酶结构域转导)。

[4730] MECP2多肽的实施例性氨基酸序列提供如下:

[4731] >sp|P51608|MECP2_人类甲基CpG结合蛋白2OS=智人OX=9606GN=MECP2 PE=1SV=1

[4732] MVAGMLGLREEKSEDQDLQGLKDKPLKFKKVKKDKKEEKEGKHEPVQPSAHHSAEPAEAGKAETSEGSGSAPAVPEASASPKQRRSIIRDRGPMYDDPTLPEGWTRKLKQRKSGRSAGKYDVYLINPQGKAFRSKVELIAYFEKVGDTSLDPNDFDFTVTGRGSPSRREQKPPKKPKSPKAPGTGRGRGRPKGSGTTRPKAATSEGVQVKRVLEKSPGKLLVKMPFQTSPGGKAEGGGATTSTQVMVIKRPGRKRKAEADPQAIPKKRGRKPGSVVAAAAAEAKKKAVKESSIRSVQETVLPIKKRKTRETVSIEVKEVVKPLLVSTLGEKSGKGLKTCKSPGRKSKESSPKGRSSSASSPPKKEHHHHHHHSESPKAPVPLLPPLPPPPPEPESSEDPTSPPEPQDLSSSVCKEEKMPRGGSLESDGCPKEPAKTQPAVATAATAAEKYKHRGEGERKDIVSSSMPRPNREEPVDSRTPVTERVS(SEQ ID NO:5)。

[4733] 向导RNA序列可以包括间隔

[4734] CTTTTCACTTTTCCTGCCGGGG(R255X、AGCTTCCATGTCCAGCCTTC(R106W)、ACCATGAAGTCAAAATCATT(T158M)、或GCTTTCAGCCCCGTTTCTTG(R270X)。

[4735] 本文的方法包括向受试者(包括识别为需要此类治疗的受试者,或怀疑处于疾病风险中并需要此类治疗的受试者)给药有效量的本文所述组合物。识别需要此类治疗的受试者可以由受试者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)。

[4736] 治疗方法通常包括给药治疗有效量的药物组合物,所述药物组合物包括例如编码碱基编辑器的载体和靶向有需要的受试者(例如人类患者)的Mecp2基因的gRNA。此类治疗将适当地给药于患有、具有、易患RTT或处于RTT风险中的受试者,特别是人类受试者。本文的组合物还可用于治疗任何其他可能涉及RTT的疾病。

[4737] 在一个实施方案中,本发明提供一种监测治疗进展的方法。所述方法包括确定患有或易患与RTT相关的病症或其症状的受试者的诊断标志物(例如,与RTT相关的SNP)或诊断测量(例如,筛选、测定)的水平的步骤。已向受试者给药足以治疗疾病或其症状的治疗量的本文组合物。在所述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,以确定受试者的疾病状态。在优选的实施方案中,在确定第一水平之后的时间点确定受试者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。在某些优选的实施方案中,在开始根据本发明的治疗之前确定受试者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受试者中的标志物水平进行比较,以确定治疗的功效。

[4738] 在一些实施方案中,细胞获自受试者并与本文所提供的药物组合物接触。在一些实施方案中,任选地在细胞中实现或检测到所需基因组修饰之后,从受试者取出并离体与药物组合物接触的细胞重新引入受试者。递送包括核酸酶的药物组合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;及7,163,824,其公开内容通过引用整体并入本文。尽管本文所提供的药物组合物的描述主要涉及适合于对人类给药的药物组合物,但本领域

技术人员应理解,此类组合物通常适合于给药于各种动物或生物体,例如,兽医用途。

[4739] 蕾特氏症的一种或多种症状包括但不限于就寝时间抵抗、入睡延迟、睡眠持续时间、睡眠焦虑、夜醒、异态睡眠、睡眠呼吸障碍、白天嗜睡、手部功能、行走、语言和非语言沟通、理解、注意力、行为问题、情绪、癫痫活动、行为和情绪特征,或其组合。

[4740] 受试者可能表现出蕾特氏症的一种或多种症状的改善。在一个实施方案中,一种或多种症状的改善至少为5%。在另一个实施方案中,一种或多种症状的改善是至少10%。在另一个实施方案中,一种或多种症状的改善是至少15%。在另一个实施方案中,一种或多种症状的改善是至少20%。在另一个实施方案中,一种或多种症状的改善是至少25%。在另一个实施方案中,一种或多种症状的改善是至少30%。在另一个实施方案中,一种或多种症状的改善是至少35%。在另一个实施方案中,一种或多种症状的改善是至少40%。在另一个实施方案中,一种或多种症状的改善是至少50%。在另一个实施方案中,一种或多种症状的改善是至少60%。在另一个实施方案中,一种或多种症状的改善是至少70%。在另一个实施方案中,一种或多种症状的改善是至少75%。在另一个实施方案中,一种或多种症状的改善是至少80%。在另一个实施方案中,一种或多种症状的改善是至少85%。在另一个实施方案中,一种或多种症状的改善是至少90%。在另一个实施方案中,一种或多种症状的改善是至少95%。

[4741] 试剂盒

[4742] 本公开的各个方面提供包括碱基编辑器系统的试剂盒。在一个实施方案中,所述试剂盒包括核酸构建体,所述核酸构建体包括编码核碱基编辑器融合蛋白的核苷酸序列。融合蛋白包括脱氨酶(例如胞苷脱氨酶或腺嘌呤脱氨酶)和核酸可编程的DNA结合蛋白(napDNAbp)。在一些实施方案中,试剂盒包括至少一种能够靶向感兴趣的核酸分子的向导RNA。在一些实施方案中,所述试剂盒包括包含编码至少一种向导RNA的核苷酸序列的核酸构建体。

[4743] 在一些实施方案中,所述试剂盒提供使用所述试剂盒编辑一个或多个突变的说明。说明书通常包括关于使用试剂盒编辑核酸分子的信息。在其他实施方案中,说明包括以下至少一项:注意事项;警告;临床研究;和/或参考。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。在进一步的实施方案中,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。在又一个实施方案中,试剂盒可以包括具有合适的阳性和阴性对照或对照样本的一个或多个容器,以用作检测、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。

[4744] 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,这些都在本领域技术人员的能力范围内。此类技术在文献中得到充分解释,例如,本公开的各个方面提供包括碱基编辑器系统的试剂盒。在一个实施方案中,所述试剂盒包括核酸构建体,所述核酸构建体包括编码核碱基编辑器融合蛋白的核苷酸序列。融合蛋白包括脱氨酶(例如胞苷脱氨酶或腺嘌呤脱氨酶)和核酸可编程的DNA结合蛋白(napDNAbp)。在一些实施方案中,试剂盒包括至少一种能够靶向感兴趣的

核酸分子的向导RNA。在一些实施方案中,所述试剂盒包括包含编码至少一种向导RNA的核苷酸序列的核酸构建体。

[4745] 在一些实施方案中,所述试剂盒提供使用所述试剂盒编辑一个或多个突变的说明。说明书通常包括关于使用试剂盒编辑核酸分子的信息。在其他实施方案中,说明包括以下至少一项:注意事项;警告;临床研究;和/或参考。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。在进一步的实施方案中,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。在又一个实施方案中,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本,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。

[4746] 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,这些都在本领域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",second edition(Sambrook,1989);"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。特定实施方案的特别有用的技术将在以下部分中讨论。

[4747] 提出以下实施例是为了向本领域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,并且不旨在限制本发明的范围发明者认为是他们的发明。

[4748] 实施例

[4749] 实施例1:具有增加的编辑效率的腺苷碱基编辑器

[4750] 包含Tad7.10-dCas9融合蛋白的碱基编辑系统能够以大约10至20%的效率编辑目标多核苷酸,但对于需要更高效率的用途,其使用可能会受到限制。为了识别具有提高的效率和特异性的腺嘌呤碱基编辑器,包括腺苷脱氨酶TadA 7.10的构建体通过易错PCR进行诱变,随后克隆到与编码dCas9(一种核酸可编程的DNA结合蛋白)的核酸序列相邻的表达载体中(图1A)。将包括腺苷脱氨酶变体的表达载体与编码氯霉素抗性(CamR)和奇霉素抗性(SpectR)的选择质粒共转换到感受态细菌细胞中,并具有通过两个点突变变得无功能的卡那霉素抗性基因(第7轮进化策略)(图1B)。选择细胞以恢复卡那霉素抗性,这是腺苷脱氨酶活性的读数。在随后的几轮选择中,将表达载体与编码氯霉素抗性(CamR)和奇霉素抗性(SpectR)的质粒共转换到感受态细胞中,并具有通过三点突变变得无功能的卡那霉素抗性基因(第8轮进化策略))(图1C)。灭活的卡那霉素抗性基因核酸序列提供如下:

ccggaattgccagctggggcgccctctggtaaggttgggaagccctg
caaagtaaactggatggctttcttgccgccaaggatctgatggcgcaggg
gatcaagatctgatcaagagacaggatgaggatcct**ttcgcATGATCGAA**
TAAGATGGATTGCACGCAGGTTCTCCGGCC**GCTTAGGTGGAGCGCCTATT**
CGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGT
TCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTG
TCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCT
GGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAG
CGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTG

[4751] TCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAAT
GCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAG
CGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTC
GATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACT
GTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGA
CCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTT
TCTGGA*TTCATTAACTGTGGCCGGCT*GGGTGTGGCGGACCGCTATCAGGA
CATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGG
CTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGC
ATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTAA

[4752] 上述序列中,小写字母表示卡那霉素抗性启动子区,粗体序列表示靶向失活部分(Q4*和W15*),斜体序列表示卡那霉素抗性基因(D208N)靶向失活位点,底线序列表示PAM序列。

[4753] 再次将细胞接种到一系列卡那霉素浓度增加的琼脂糖板上。如图2所示,具有有效碱基编辑活性的腺苷脱氨酶变体能够校正卡那霉素抗性基因中存在的突变,并选择用于进一步分析。在表13中描述了在细菌细胞中显示有效碱基编辑的腺苷脱氨酶变体碱基编辑器。产生了编码包括所选腺苷脱氨酶变体的碱基编辑器的哺乳动物表达载体。

[4754] 表达与镰状细胞病相关的β-球蛋白的Hek293T细胞含有E6V(也称为E7V)突变,用于测试腺苷脱氨酶变体的编辑效率(图3A和3B)。这些被称为"Hek293T/HBBE6V"细胞的细胞使用表达碱基编辑系统的慢病毒载体转导,所述达碱基编辑系统包括包含表13中列出的ABE8碱基编辑器的融合蛋白。ABE8碱基编辑器是通过将腺苷脱氨酶变体克隆到支架中而产生的包括一个环状置换Cas9和一个二分核定位序列。环状置换Cas9是本领域已知的并且描述于例如Oakes et al.,Cell 176,254-267,2019。这些序列在下文中提供。

[4755] 胎儿血红蛋白的上调是克服镰状细胞病的一种治疗方法。图3A显示了胎儿血红蛋白上调的治疗相关位点。编辑残基5和8的腺苷可以显着降低BCL11A结合,从而增加胎儿血红蛋白的表达。参考图3A,ABE8碱基编辑器的碱基编辑活动比ABE7.10碱基编辑器高约2至3倍。

[4756] 表13:新颖腺嘌呤碱基编辑器ABE8

[4757]

质粒ID	描述	功能
280	ABE8.1	单体_TadA*7.10+Y147T
281	ABE8.2	单体_TadA*7.10+Y147R
282	ABE8.3	单体_TadA*7.10+Q154S
283	ABE8.4	单体_TadA*7.10+Y123H
284	ABE8.5	单体_TadA*7.10+V82S
285	ABE8.6	单体_TadA*7.10+T166R
286	ABE8.7	单体_TadA*7.10+Q154R
287	ABE8.8	单体_Y147R_Q154R_Y123H
288	ABE8.9	单体_Y147R_Q154R_I76Y
289	ABE8.10	单体_Y147R_Q154R_T166R
290	ABE8.11	单体_Y147T_Q154R
291	ABE8.12	单体_Y147T_Q154S
292	ABE8.13	单体_H123Y123H_Y147R_Q154R_I76Y
293	ABE8.14	异源二聚体_TadA*7.10+Y147T
294	ABE8.15	异源二聚体_TadA*7.10+Y147R
295	ABE8.16	异源二聚体_TadA*7.10+Q154S
296	ABE8.17	异源二聚体_TadA*7.10+Y123H
297	ABE8.18	异源二聚体_TadA*7.10+V82S
298	ABE8.19	异源二聚体_TadA*7.10+T166R
299	ABE8.20	异源二聚体_TadA*7.10+Q154R
300	ABE8.21	异源二聚体_Y147R_Q154R_Y123H
301	ABE8.22	异源二聚体_Y147R_Q154R_I76Y
302	ABE8.23	异源二聚体_Y147R_Q154R_T166R
303	ABE8.24	异源二聚体_Y147T_Q154R
304	ABE8.25	异源二聚体_Y147T_Q154S

[4758] 参考图4所示,ABE8碱基编辑器与靶向编码HBB E6V的多核苷酸的18、19、20、21或22个核苷酸向导RNA一起被引入Hek293T/HBBE6V细胞。当与循环置换(Cp)-Cas9融合时,ABE8编辑器显示出更高的编辑效率。总共测试了40种不同的ABE8构建体(表14)和三种ABE7.10构建体在Hek293T/HBBE6V细胞中的编辑活性。例示性构建体的序列如下。为了评估编辑的特异性,监测目标突变和非预期突变或旁观者突变(图5)。密码子5中腺苷的意外编辑是沉默的。然而,密码子9的非预期编辑导致丝氨酸到脯氨酸突变。再次参考图5,与ABE7.10编辑器相比,多个ABE8碱基编辑器显示出更高的编辑效率和特异性,并且没有一个编辑器具有导致丝氨酸至脯氨酸错义突变的显着旁观者编辑。

[4759] 在含有镰状细胞突变的成纤维细胞中对选定的ABE8碱基编辑器和ABE7.10碱基编辑器对照进行了进一步分析。如图6所示,与ABE7.10相比,ABE8编辑器增加了碱基编辑活动。ABE8.18显示大约70%的效率。选定的ABE8编辑器也显示出前所未有的特异性。重要的是,所有ABE8编辑器的平均插入缺失形成低于0.1%。

[4760] 表14:

[4761]

质粒 ID	描述	功能
335	NGC PAM CP5 变体(化脓链球菌 Cas9)_ABE8.1	单体_TadA*7.10+Y147T
336	NGC PAM CP5 variant (化脓链球菌 Cas9)_ABE8.2	单体_TadA*7.10+Y147R
337	NGC PAM CP5 变体(化脓链球菌 Cas9)_ABE8.3	单体_TadA*7.10+Q154S
338	NGC PAM CP5 变体(化脓链球菌 Cas9)_ABE8.4	单体_TadA*7.10+Y123H
339	NGC PAM CP5 变体(化脓链球菌 Cas9)_ABE8.5	单体_TadA*7.10+V82S
340	NGC PAM CP5 变体(化脓链球菌 Cas9)_ABE8.6	单体_TadA*7.10+T166R
341	NGC PAM CP5 变体(化脓链球菌 Cas9)_ABE8.7	单体_TadA*7.10+Q154R
342	NGC PAM CP5 变体(化脓链球菌 Cas9)_ABE8.8	单体_ Y147R_Q154R_Y123H
343	NGC PAM CP5 变体(化脓链球菌 Cas9)_ABE8.9	单体_ Y147R_Q154R_I76Y
344	NGC PAM CP5 变体(化脓链球菌 Cas9)_ABE8.10	单体_ Y147R_Q154R_T166R

[4762]

345	NGC PAM CP5 变体(化脓链球菌 Cas9)_ABE8.11	单体_ Y147T_Q154R
346	NGC PAM CP5 变体(化脓链球菌 Cas9)_ABE8.12	单体_ Y147T_Q154S
347	NGC PAM CP5 变体(化脓链球菌 Cas9)_ABE8.13	单体_H123Y123H_ Y147R_Q154R_I76Y
348	NGC PAM CP5 变体(化脓链球菌 Cas9)_ABE7.10	在 TadA7.10*残基 149 的单体_缺失
349	NGC PAM CP5 变体(化脓链球菌 Cas9)_ABE7.10	在 TadA7.10*残基 150 的单体_缺失
350	NGC PAM CP5 变体(化脓链球菌 Cas9)_ABE7.10	在 TadA7.10*残基 151 的单体_缺失
351	NGC PAM CP5 变体(化脓链球菌 Cas9)_ABE7.10	在 TadA7.10*残基 152 的单体_缺失
352	NGC PAM CP5 变体(化脓链球菌 Cas9)_ABE7.10	在 TadA7.10*残基 153 的单体_缺失
353	NGC PAM CP5 变体(化脓链球菌 Cas9)_ABE7.10	在 TadA7.10*残基 154 的单体_缺失
354	NGC PAM CP5 变体(化脓链球菌 Cas9)_ABE7.10	在 TadA7.10*残基 155 的单体_缺失
355	NGC PAM CP5 变体(化脓链球菌 Cas9)_ABE7.10	在 TadA7.10*残基 156 的单体_缺失
356	NGC PAM CP5 变体(化脓链球菌 Cas9)_ABE7.10	在 TadA7.10*残基 157 的单体_缺失
357	NGC PAM CP5 变体(化脓链球菌 Cas9)_ABE8.14	异源二聚体_TadA*7.10+Y147T
358	NGC PAM CP5 变体(化脓链球菌 Cas9)_ABE8.15	异源二聚体_TadA*7.10+Y147R
359	NGC PAM CP5 变体(化脓链球菌 Cas9)_ABE8.16	异源二聚体_TadA*7.10+Q154S
360	NGC PAM CP5 变体(化脓链球菌 Cas9)_ABE8.17	异源二聚体_TadA*7.10+Y123H
361	NGC PAM C P5 变体(化脓链球菌 Cas9)_ABE8.18	异源二聚体_TadA*7.10+V82S
362	NGC PAM CP5 变体(化脓链球菌	异源二聚体_TadA*7.10+T166R

[4763]

	Cas9)_ABE8.19	
363	NGC PAM CP5变体(化脓链球菌 Cas9)_ABE8.20	异源二聚体_TadA*7.10+Q154R
364	NGC PAM CP5变体(化脓链球菌 Cas9)_ABE8.21	异源二聚体_ Y147R_Q154R_Y123H
365	NGC PAM CP5变体(化脓链球菌 Cas9)_ABE8.22	异源二聚体_ Y147R_Q154R_I76Y
366	NGC PAM CP5变体(化脓链球菌 Cas9)_ABE8.23	异源二聚体_ Y147R_Q154R_T166R
367	NGC PAM CP5变体(化脓链球菌 Cas9)_ABE8.24	异源二聚体_ Y147T_Q154R
368	NGC PAM CP5变体(化脓链球菌 Cas9)_ABE8.25	异源二聚体_ Y147T_Q154S
369	NGC PAM CP5变体(化脓链球菌 Cas9)_ABE8.26	异源二聚体_H123Y123H_ Y147R_Q154R_I76Y
370	NGC PAM CP5变体(化脓链球菌 Cas9)_ABE7.10	在TadA7.10*残基的149异源二聚体_缺失
371	NGC PAM CP5变体(化脓链球菌 Cas9)_ABE7.10	在TadA7.10*残基的150异源二聚体_缺失
372	NGC PAM CP5变体(化脓链球菌 Cas9)_ABE7.10	在TadA7.10*残基的151异源二聚体_缺失
373	NGC PAM CP5变体(化脓链球菌 Cas9)_ABE7.10	在TadA7.10*残基的152异源二聚体_缺失
374	NGC PAM CP5变体(化脓链球菌 Cas9)_ABE7.10	在TadA7.10*残基的153异源二聚体_缺失
375	NGC PAM CP5变体(化脓链球菌 Cas9)_ABE7.10	在TadA7.10*残基的154异源二聚体_缺失
376	NGC PAM CP5变体(化脓链球菌 Cas9)_ABE7.10	在TadA7.10*残基的155异源二聚体_缺失
377	NGC PAM CP5变体(化脓链球菌 Cas9)_ABE7.10	在TadA7.10*残基的156异源二聚体_缺失
378	NGC PAM CP5变体(化脓链球菌 Cas9)_ABE7.10	在TadA7.10*残基的157异源二聚体_缺失

[4764] 实施例2:ABE8设计的密码子优化和NLS选择

[4765] 已经确定Cas9密码子使用和核定位序列可以显着改变真核生物的基因组编辑效率(参见,例如,Kim,S.et al.,Rescue of high-specificity Cas9 variants using sgRNAs with matched 5'nucleotides.Genome Biol 18,218,doi:10.1186/s13059-017-1355-3(2017);Mikami,M.et al.,Comparison of CRISPR/Cas9 expression constructs

for efficient targeted mutagenesis in rice.Plant Mol Biol 88,561-572,doi:10.1007/s11103-015-0342-x(2015);Jinek,M.et al.,RNA-programmed genome editing in human cells.Elife2,e00471,doi:10.7554/eLife.00471(2013))。碱基编辑器的原始Cas9n组分包括六个潜在的多聚腺苷酸化位点,导致在真核生物中的表达不佳(参见,例如,Kim,S.et al.,Rescue of high-specificity Cas9 variants using sgRNAs with matched 5'nucleotides.Genome Biol 18,218,doi:10.1186/s13059-017-1355-3(2017);Komor,A.C.et al.,Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage.Nature 533,420-424,doi:10.1038/nature17946(2016);Gaudelli,N.M.et al.Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage.Nature 551,464-471,doi:10.1038/nature24644(2017))。用广泛优化的密码子序列置换可以提高碱基编辑效率(参见,例如Cong,L.et al.Multiplex genome engineering using CRISPR/Cas systems.Science 339,819-823,doi:10.1126/science.1231143(2013);Koblan,L.W.et al.Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction.Nat Biotechnol,doi:10.1038/nbt.4172(2018);Zafra,M.P.et al.Optimized base editors enable efficient editing in cells,organoids and mice.Nat Biotechnol,doi:10.1038/nbt.4194(2018))。

[4766] 评估了与四种ABE构建体相关的DNA靶上(图9A、9B)、DNA脱靶(图9C、9D)和RNA脱靶(图9E)碱基编辑频率,所有这些都含有密码子-优化的Cas9(D10A):i)ABE7.10,它具有单个C端BP-SV40 NLS;ii)缺少ABE7.10的5'TadA野生型部分的monoABE7.10;iii)ABEmax,其含有密码子优化的TadA区域和两个BPNLS序列;及iv)ABEmax(-BPNLS),其具有作为ABEmax的TadA密码子优化,但含有单个C端BP-SV40 NLS。

[4767] 所有四种构建体都显示出非常相似的靶向编辑效率,表示NLS结构和TadA密码子优化并不能决定靶向编辑效率(图9A、9B)。脱靶谱也高度相似,但与ABE7.10(图9C、9D)相比,ABEmax在一个位点显示出显着更大的DNA脱靶编辑(p=0.00027,学生的双尾T检验)。ABEmax(-NBPNLS)显示出比ABE7.10高1.6倍的RNA脱靶编辑平均频率(图9E)。

[4768] 实施例3:具有扩展靶向范围的卓越腺嘌呤碱基编辑器

[4769] ABE是一种分子机器,包括进化的大肠杆菌tRNAARG修饰酶TadA,所述酶与催化受损的Cas9蛋白(D10A切口酶Cas9、nCas9)共价融合(图7A和7B)。为了克服先前腺嘌呤碱基编辑器的局限性,通过设计ABE来提高细菌选择系统的严格性,所述ABE必须进行三个并发的A·T至G·C回复编辑才能在抗生素选择中存活下来。在之前的ABE进化中,TadA文库是通过容易出错的PCR创建的。相比之下,使用的TadA等位基因的合成文库在TadA的每个位置包括所有20个规范氨基酸取代,每个文库成员的平均频率为1至2个核苷酸替换突变。所述化学文库能够进入比容易出错的PCR技术更大的序列空间。

[4770] 约300个克隆被分离出来并随后被测序。从得到的测序数据中,在TadA*内识别了八个突变,这些突变以高频率富集(表7和9)。八个已识别的氨基酸突变中有六个需要每个密码子至少两个核碱基变化,这是以前的TadA容易出错的文库未观察到的。两个富集的突变改变了靠近腺嘌呤脱氨活性位点的残基(I76和V82)(图7C)。除了先前报道的TadA*7.10的C端α螺旋中的四个突变之外,在相同的α螺旋(Y147R和Q154R)内观察到了两个新的突变

(图7C)。这种高度突变的α-螺旋对于强大的产品形成是必要的,因为在截断时碱基编辑效率显着降低(图10A和10B)。

[4771] 为了测试哺乳动物细胞中TadA*变体的活性,BE密码子优化和NLS方向与最有利的靶上和脱靶特征一起使用(参见实施例2;图9A至9E)。8个富集的TadA*突变以各种组合被整合到ABE7.10中,产生了40个新的ABE8变体(表7和9)。制作其中ABE的TadA区域是非活性(野生型)和活性(进化)TadA*原体的异源二聚体融合或工程化TadA*的单个原体的ABE8构建体,产生约500个碱基-配对较小的编辑器。这些架构变体分别称为ABE8.x-d和ABE8.x-m(表7和9)。

[4772] 首先,评估了这40个构建体相对于ABE7.10横跨8个基因组位点的靶向DNA编辑效率,这些位点包括位于2到20位(其中NGG PAM=21、22、23位)的8个基因组位点。规范的20个核苷酸个核苷酸化脓链球菌前间隔(图11)。N端野生型TadA构建体对于使用ABE8进行稳健的DNA编辑不是必需的。事实上,包括N端野生型TadA(ABE8.xd)的构建体在编辑窗口偏好、总DNA编辑结果和相对于其经济化架构(ABE8.xm)的插入缺失频率方面表现相似(图7D、图11、图12)。尽管构建体内TadA(野生型):TadA*8二聚化可能不是ABE8活性所必需的,但并不排除在ABE8表达的碱基编辑器之间发生反式TadA*8:TadA*8二聚化的可能性。

[4773] 在所有测试的位点中,与ABE7.10相比,ABE8在前间隔的规范位置(A5至A7)处的编辑率提高了约1.5倍,在非规范位点(A3至A4、A8至A10)处的编辑率提高了约3.2倍(图13)。目标序列、目标窗口内“A”的位置和ABE8构建体身份之间的倍数差异不同(图7D、图11、图13)。总体而言,相对于ABE7.10(范围1.34至4.49),所有测试位点中所有位置的编辑变化中位数是1.94倍。

[4774] 接下来,从四十个构建体的大型ABE8池中,选择ABE8构建体的子集(ABE8.8-m、ABE8.13-m、ABE8.17-m、ABE8.20-m、ABE8.8-d、ABE8.13-m、ABE8.17-d和ABE8.20-d)以进行更详细的评估。如通过分层聚类分析确定,这些构建体代表了在8个基因组位点之间在编辑性能方面具有明显差异的ABE8的(图14)。这些ABE8在所有测试的基因组位点都显着优于ABE7.10(P值=0.0006871,双尾Wilcoxon秩和检验),并且包括从ABE8定向进化活动中识别的各种突变组合(图15和图16)。

[4775] 尽管已经描述了识别非NGG PAM的ABE变体,与使用靶向NGG PAM序列的化脓链球菌Cas9观察到的结果相比,这些构建体的编辑效率在许多情况下会降低(参见例如Huang,T.P.et al.Circularly permuted and PAM-modified Cas9 variants broaden the targeting scope of base editors.Nat Biotechnol 37,626-631,doi:10.1038/s41587-019-0134-y(2019);Hua,K.et al.,Expanding the base editing scope in rice by using Cas9 variants.Plant Biotechnol J,doi:10.1111/pbi.12993(2018);Yang,L.et al.,Increasing targeting scope of adenosine base editors in mouse and rat embryos through fusion of TadA deaminase with Cas9 variants.Protein Cell 9,814-819,doi:10.1007/s13238-018-0568-x(2018))。为了确定进化的脱氨酶是否也提高了带有非NGG PAM的目标位点的编辑效率,创建了用工程化的化脓链球菌变体NG-Cas9(PAM:NG)(Nishimasu,H.et al.Engineered CRISPR-Cas9 nuclease with expanded targeting space.Science(2018))或金黄色葡萄球菌Cas9(SaCas9,PAM:NNGRRT)(Ran,F.A.et al.In vivo genome editing using Staphylococcus aureus Cas9.Nature 520,186-191,doi:

10.1038/nature14299(2015))置换化脓链球菌Cas9的ABE8编辑器。当将ABE7.10和ABE8构建体进行SpCas9-NG(NG-ABE8.xm/d)和SaCas9(Sa-ABE8.xm/d)的比较时,分别观察到A·T至G·C编辑频率的中位数分别增加了1.6倍和2.0倍(图8A、8B和17至20)。与SpCas9-ABE8类似,ABE7.10和ABE8构建体之间对于非NGG PAM变体的编辑效率差异最大的是在位于编辑窗口中首选位置(化脓链球菌:位置4至8;金黄色葡萄球菌:位置6至13;另参见Rees,H.A.&Liu,D.R.,Base editing:precision chemistry on the genome and transcriptome of living cells.Nat Rev Genet 19,770-788,doi:10.1038/s41576-018-0059-1(2018))。利用非NGG PAM的ABE8直向同源物拓宽了高效A碱基编辑的靶向范围。

[4776] 对于需要最大限度减少插入缺失形成的应用,用催化"死亡"版本的Cas9(D10A+H840A)置换Cas9的催化受损的D10A切口酶突变体的效果(参见Rees,H.A.&Liu,D.R.,Base editing:precision chemistry on the genome and transcriptome of living cells.Nat Rev Genet 19,770-788,doi:10.1038/s41576-018-0059-1(2018))在核心8个ABE8构建体("dC9-ABE8.xm/d")中探索。通过在ABE中用死亡的Cas9替换切口酶,观察到dC9-ABE8变体相对于ABE7.10的插入缺失频率降低了>90%,同时保持了显着更高(2.1倍)的靶上DNA编辑效率(图2、8C、21、22、23A和23B)。尽管观察到高于背景的插入缺失,在测试位点的频率范围仅为0.3至0.8%。令人鼓舞的是,与规范ABE8相比,dC9-ABE8变体的靶上DNA编辑效率仅降低了14%。

[4777] 在目标基因座上的另一类非所欲的ABE介导的基因组编辑是ABE依赖性胞嘧啶至尿嘧啶(C·G至T·A)的转换(参见Grunewald,J.et al.,CRISPR DNA base editors with reduced RNA off-target and self-editing activities.Nat Biotechnol 37,1041-1048,doi:10.1038/s41587-019-0236-6(2019);Lee,C.et al.CRISPR-Pass:Gene Rescue of Nonsense Mutations Using Adenine Base Editors.Mol Ther 27,1364-1371,doi:10.1016/j.ymthe.2019.05.013(2019))。在测试的8个目标位点,C至T编辑的第95个百分位数被测量为0.45%具有ABE8变体和0.15%具有ABE7.10-d或-m,表示可能发生用ABE的靶上胞嘧啶脱氨,但频率通常非常低(图24)。总之,这些数据表示,与其他通常非所欲的副产品相比,ABE8保留了A至G转换的高特异性。

[4778] 实施例4:通过ABE8构建体进行的DNA靶上编辑和sgRNA依赖性DNA脱靶编辑提高对DNA的特异性

[4779] 与所有碱基编辑器一样,ABE8具有在基因组和转录组中的脱靶位点处起作用的可能(参见,例如Gaudelli,N.M.et al.Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage.Nature 551,464-471,doi:10.1038/nature24644(2017);Komor,A.C.,et al.,Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage.Nature 533,420-424,doi:10.1038/nature17946(2016);Grunewald,J.et al.CRISPR DNA base editors with reduced RNA off-target and self-editing activities.Nat Biotechnol 37,1041-1048,doi:10.1038/s41587-019-0236-6(2019);Rees,H.A.,et al.,Analysis and minimization of cellular RNA editing by DNA adenine base editors.Sci Adv 5,eaax5717,doi:10.1126/sciadv.aax5717(2019);Rees,H.A.et al.Improving the DNA specificity and applicability of base editing through protein engineering and protein

delivery.Nat Commun 8,15790,doi:10.1038/ncomms15790(2017);Jin,S.et al.Cytosine,but not adenine,base editors induce genome-wide off-target mutations in rice.Science 364,292-295,doi:10.1126/science.aaw7166(2019);Zuo,E.et al.Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos.Science 364,289-292,doi:10.1126/science.aav9973(2019);Lee,H.K.,et al.,Cytosine but not adenine base editor generates mutations in mice.Biorxiv,doi:https://doi.org/10.1101/731927(2019);Grunewald,J.et al.Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors.Nature 569,433-437,doi:10.1038/s41586-019-1161-z(2019);Zhou,C.et al.Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis.Nature 571,275-278,doi:10.1038/s41586-019-1314-0(2019))。

[4780] 在基因组DNA中4个靶上点(图25A和25B)和12个先前识别的sgRNA相关脱靶位点(Tsai,S.Q.et al.GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases.Nat Biotechnol 33,187-197,doi:10.1038/nbt.3117(2015))(图25E和25F),所有这些都被证实是HEK293T细胞中真正的Cas9脱靶基因座(图26)。正如从其在靶上基因座上增加的活性所预期,ABE8构建体的DNA脱靶编辑频率比ABE7.10高3至6倍。虽然这是使用ABE8构建体的警告,但仔细选择和分析sgRNA可以大大减少sgRNA依赖的脱靶编辑(参见Tsai,S.Q.et al.GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases.Nat Biotechnol 33,187-197,doi:10.1038/nbt.3117(2015);Yeh,W.H.,et al.,In vivo base editing of post-mitotic sensory cells.Nat Commun 9,2184,doi:10.1038/s41467-018-04580-3(2018))。对于需要使用混杂sgRNA的应用,安装DNA和RNA特异性增强V106W突变(Rees,H.A.,Wilson,C.,Doman,J.L.&Liu,D.R.Analysis and minimization of cellular RNA editing by DNA adenine base editors.Sci Adv 5,eaax5717,doi:10.1126/sciadv.aax5717(2019))进入ABE8.17m的TadA结构域可以将DNA脱靶编辑减少2.6倍,同时保持超过ABE7.10的脱靶编辑水平(图25C、25D、25G和25H)。

[4781] 为了测量ABE8的不依赖于sgRNA的脱靶活性,在用ABE处理的HEK293T细胞中进行了细胞RNA的靶向扩增和高通量测序(参见Gaudelli,N.M.et al.Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage.Nature 551,464-471,doi:10.1038/nature24644(2017);Rees,H.A.,et al.,Analysis and minimization of cellular RNA editing by DNA adenine base editors.Sci Adv 5,eaax5717,doi:10.1126/sciadv.aax5717(2019))。在所述测定中,与ABE7.10相比,ABE显示出高2.3至5.3倍的细胞RNA腺苷脱氨的平均频率(图25A)。

[4782] 为了减轻虚假的RNA脱靶编辑,先前发表的突变(Grunewald,J.et al.CRISPR DNA base editors with reduced RNA off-target and self-editing activities.Nat Biotechnol 37,1041-1048,doi:10.1038/s41587-019-0236-6(2019);Rees,H.A.,et al.,Analysis and minimization of cellular RNA editing by DNA adenine base editors.Sci Adv 5,eaax5717,doi:10.1126/sciadv.aax5717(2019);Grunewald,J.et

al.Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors.Nature 569,433-437,doi:10.1038/s41586-019-1161-z(2019);Zhou,C.et al.Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis.Nature 571,275-278,doi:10.1038/s41586-019-1314-0(2019))将脱氨酶的TadA部分安装到ABE8.17-m,以评估脱靶编辑频率的降低。所有这些突变都不同程度地降低了ABE8.17-m的靶上编辑频率,其中V106W和F148A对ABE8的损害最小(图25C和25D)。其中,只有V106W能够显着降低脱靶RNA和DNA编辑的水平(图25B)。因此,将V106W突变包含在ABE8中适用于必须避免细胞转录组瞬时扰动的情况,或与混杂的sgRNA使用。

[4783] 实施例5:用于治疗血液病症的腺嘌呤碱基编辑器

[4784] 在人类造血干细胞(HSC)中评估了ABE8构建体。在作为细胞疗法给予患者之前,对HSC进行体外操作和/或编辑是治疗血液病的一种很有前景的方法。之前已经证明ABE可以在HBG1/2启动子区域的-198位置引入T至C的取代(Gaudelli,N.M.et al.Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage.Nature 551,464-471,doi:10.1038/nature24644(2017))。这种自然发生的等位基因产生胎儿血红蛋白遗传性持久性(hereditary persistence of feal hemoglobin,HPFH),导致γ-球蛋白水平增加到成年期,这可以减轻镰状细胞病和β-地中海贫血中β-球蛋白的缺陷(Wienert,B.et al.KLF1 drives the expression of fetal hemoglobin in British HPFH.Blood 130,803-807,doi:10.1182/blood-2017-02-767400(2017))。为了再现HPFH表型和评估ABE8的临床相关性,从两个供体中单离出CD34+造血干细胞,并用编码ABE8编辑器的mRNA和将目标A置于前间隔的第7位的末端修饰的sgRNA转染,。

[4785] -198HBG1/2启动子目标位点的平均ABE8编辑效率在早期时间点(48小时)比任一ABE7.10构建体高2至3倍,在后期比任一ABE7.10高1.3至2倍(144h)(图27A、图28、图29)。这些动力学差异对于离体疗法在临床上很重要,在这种疗法中,在给予细胞疗法之前必须将细胞培养保持在最低限度。

[4786] 接下来,通过UPLC对ABE处理和红细胞分化后产生的γ-球蛋白的量进行定量(图30至50)。当将ABE8.13-d与使用ABE7.10-m/d时,与模拟处理细胞相比,观察到来自ABE8处理组的红细胞中%γ-球蛋白/α-球蛋白表达的平均增加3.5倍(图27B)。据预测,改善镰状细胞病的症状需要≥20%的HbF,而β-地中海贫血患者可能需要甚至更高的最低水平(参见例如,Canver,M.C.&Orkin,S.H.Customizing the genome as therapy for the beta-hemoglobinopathies.Blood 127,2536-2545,doi:10.1182/blood-2016-01-678128(2016);Fitzhugh,C.D.et al.At least 20%donor myeloid chimerism is necessary to reverse the sickle phenotype after allogeneic HSCT.Blood 130,1946-1948,doi:10.1182/blood-2017-03-772392(2017))。在ABE8处理后观察到的γ-球蛋白水平超过了这个阈值。

[4787] 总体而言,ABE8s在γ球蛋白基因HBG1和HBG2的启动子处重建了胎儿血红蛋白(HPFH)等位基因的天然遗传持久性,在人类CD34+细胞培养物中实现高达60%的编辑效率和在分化的红细胞中的相应的伽马球蛋白表达上调。

[4788] 实施例6:用于治疗镰状细胞病和β地中海贫血的互补碱基编辑方法

[4789] 镰状细胞病(SCD)和β地中海贫血是β球蛋白产生和功能障碍,可导致严重贫血和

多种器官系统的严重疾病并发症。通过上调胎儿血红蛋白(HbF)或校正β球蛋白基因而设计的造血干细胞的自体移植有可能减轻β血红蛋白病患者的疾病负担。碱基编辑是最近开发的一项技术,可以在不引入双链DNA断裂的情况下精确修改基因组。

[4790] 用胞嘧啶和腺嘌呤碱基编辑器(ABE)对γ球蛋白基因启动子进行了全面筛选,以识别会去抑制HbF的改变。确定了三个显着上调HbF的区域,并且最有效的核苷酸残基转换得到了遗传性胎儿血红蛋白(HPFH)持久性患者的自然变异的支持。已开发ABE,其在HBG1和HBG2启动子内的关键调控基序的核苷酸转换后,可显着增加HbF的水平。CD34+造血干细胞和祖细胞(HSPC)在临床规模上进行纯化,并使用旨在保持自我更新能力的过程进行编辑。在具有不同ABE的两个独立位点的编辑达到94%,并通过UPLC产生高达63%的γ球蛋白(图51A至51E)。根据HPFH的临床观察和将较高的HbF剂量与较轻的疾病联系起来的非干预治疗,观察到的HbF水平应该为大多数SCD和β地中海贫血患者提供保护(Ngo et al.,2011Brit J Hem;Musallam et al.,2012Blood)。

[4791] 直接校正SCD的Glu6Val突变一直是为SCD人群设计的基因疗法的近期目标。当前的碱基编辑技术还不能转换像镰状β球蛋白中的A至T颠换导致的突变;然而,ABE变体已被设计为辨识和编辑缬氨酸的相反链腺嘌呤残基。这导致缬氨酸转换为丙氨酸,并产生一种称为Hb G-望加锡(Makassar)的天然变异体。在所述位置具有丙氨酸的β球蛋白无助于聚合物形成,并且Hb G-望加锡患者的血液学参数和红细胞形态正常。用这些ABE变体编辑的SCD患者成纤维细胞实现了高达70%的靶腺嘌呤转换(图52A)。然后使用先导ABE变体编辑来自健康供体的CD34细胞,针对位于编辑窗口内的相邻脯氨酸中的同义突变,并作为编辑SCD突变的代理。平均编辑频率为40%(图52B)。在同种异体移植环境中记录的这些水平的供体骨髓嵌合现象超过了逆转镰状表型所需的20%(颠换导致的突变)。

[4792] 实施例7:材料和方法

[4793] 一般方法:

[4794] 所有克隆均通过USER酶(New England Biolabs)克隆方法进行(参见Geu-Flores et al.,USER fusion:a rapid and efficient method for simultaneous fusion and cloning of multiple PCR products.Nucleic Acids Res 35,e55,doi:10.1093/nar/gkm106(2007)),并且用于PCR扩增的模板是作为细菌或哺乳动物密码子优化的基因片段(GeneArt)购买的。将创建的载体转换到Mach T1R感受态细胞(Thermo Fisher Scientific)中并保持在-80℃长期储存。本工作中使用的所有引物均购自Integrated DNA Technologies,并使用Phusion U DNA Polymerase Green MultiPlex PCR Master Mix(ThermoFisher)或Q5 Hot Start High-Fidelity 2x Master Mix(New England Biolabs)进行PCRS。这项工作中使用的所有质粒都是使用ZymoPURE Plasmid Midiprep(Zymo Research Corporation)从50mL的Mach1培养物中新鲜制备的,其中涉及内毒素去除程序。分子生物学级Hyclone水(GE Healthcare Life Sciences)用于所有分析、转染和PCR反应,以确保排除DNAse活性。

[4795] 用于Hek293T哺乳动物细胞转染的sgRNA的氨基酸序列在下表15中提供。20个核苷酸个核苷酸目标前区隔以粗体显示。当目标DNA序列不以“G”开头时,会在引物的5'端添加一个“G”,因为已经确定人类U6启动子在转录起始位点优选“G”(参见Cong,L.et al.,Multiplex genome engineering using CRISPR/Cas systems.Science339,819-823,doi:

10.1126/science.1231143(2013))。前面描述的pFYF sgRNA质粒用作PCR扩增的模板。

[4796] 表15:用于Hek293T哺乳动物细胞转染的sgRNA序列。

[4797]

位点	RNA 前间隔序列	Cas9 支架	PAM
1	GAACACAAAGCAUAGACUGC	化脓链球菌	NGG
2	GGGAAAGACCCAGCAUCCGU	化脓链球菌	NGG
3	GCUCCCAUCACAUCAACCGG	化脓链球菌	NGG
4	GGUGAGUGAGUGUGUGCGUG	化脓链球菌	NGG
5	GGCUUCAGGUUCUAAAUGAG	化脓链球菌	NGG
6	GCAGAGAGUCGCCGUCUCCA	化脓链球菌	NGG
7	GUGUAAGACCUCAAAAGCAC	化脓链球菌	NGG
8	GAUGAGAAGGAGAAGUUCUU	化脓链球菌	NGG
9	GAGGACAAAGUACAAACGGC	化脓链球菌	AGA
10	GCCACCACAGGGAAGCUGGG	化脓链球菌	TGA
11	GCUCUCAGGCCCUGUCCGCA	化脓链球菌	CGT
12	GAGCAAAUACCAGAGAUAAG	化脓链球菌	AGA
13	GAUCAGGAAAUAGAGCCACA	化脓链球菌	GGC
14	GCCCAUCCCUGAGUCCAGCG	化脓链球菌	AGC
15	GAACACGAAGACAUCUGAAGGUA	金黄色葡萄球菌	TTGAAT
16	GAUUUACAGCCUGGCCUUUGGGG	金黄色葡萄球菌	TCGGGT
17	GGAGAGAAAGAGAAGUUGAUUG	金黄色葡萄球菌	ATGGGT
18	GAGGGUGAGGGAUGAGAUAAUG	金黄色葡萄球菌	ATGAGT
19	GGUGGAGGAGGGUGCAUGGGGU	金黄色葡萄球菌	CAGAAT
20	GCUGUUGCAUGAGGAAAGGGAC	金黄色葡萄球菌	TAGAGT
HEK2	GAACACAAAGCAUAGACUGC	化脓链球菌	CGG
HEK3	GGCCCAGACUGAGCACGUGA	化脓链球菌	TGG
HEK4	GGCACUGCGGCUGGAGGUGG	化脓链球菌	GGG
LDLR	GCAGAGCACUGGAAUUCGUCA	化脓链球菌	GGG

[4798] sgRNA支架序列如下:

[4799] 化脓链球菌:

[4800] GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC

[4801] 金黄色葡萄球菌:

[4802] GUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCUACUAAAACAAGGC AAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGA

[4803] 产生用于定向进化的输入细菌TadA*文库

[4804] TadA*8.0文库旨在编码TadA*7.10开放阅读框中每个氨基酸位置的所有20个氨基酸(Gaudelli,N.M.et al.,Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage.Nature551,464-471,doi:10.1038/nature24644(2017))。每个TadA*8.0文库成员包括大约1至2个新的编码突变并且是化学合成的,购自Ranomics Inc(加拿大多伦多)。TadA*8.0文库使用Phusion U Green MultiPlex PCR Master Mix进行

PCR扩增,用户组装成针对ABE引导进化优化的细菌载体(Gaudelli,N.M.et al.,Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage.Nature551,464-471,doi:10.1038/nature24644(2017))。

[4805] TadaA变体的细菌进化

[4806] 包括TadA*8文库的ABE的定向进化按照之前的描述进行(Gaudelli,NM等人,在基因组DNA中进行A*T到G*C的可编程的碱基编辑,无需DNA切割。Nature 551,464-471,doi:10.1038/nature24644(2017))有以下变化:i)大肠杆菌10beta(新英格兰生物实验室)被用作进化宿主;ii)卡那霉素的存活依赖于三种基因失活成分的校正(例如,存活需要逆转卡那霉素中的两个终止突变和一个活性位点突变)。卡那霉素抗性基因序列包括ABE8进化的选择突变。在10个β宿主细胞中将选择质粒和编辑器共培养过夜后,将文库培养物铺在2xYT琼脂培养基上,所述培养基补充有质粒维持抗生素和增加浓度的选择抗生素卡那霉素(64-512μg/mL)。让细菌生长1天,富集后对存活克隆的TadA*8部分进行Sanger测序。然后通过USER组装将经识别的感兴趣的TadA*8突变整合到哺乳动物表达载体中。

[4807] 一般HEK293T和RPMI-8226哺乳动物培养条件

[4808] 细胞在37℃、5% CO_2条件下培养。HEK293T细胞[CLBTx013,美国典型细胞培养物保藏中心(American Type Cell Culture Collection,ATCC)]在Dulbecco改良Eagles培养基加Glutamax(10566-016,Thermo Fisher Scientific)和10%(v/v)胎儿牛血清(A31606-02,Thermo Fisher科学)。RPMI-8226(CCL-155,ATCC)细胞在含有10%(v/v)胎牛血清(Gibco)的RPMI-1640培养基(Gibco)中培养。从供应商处收到后,细胞被检测为支原体阴性。

[4809] Hek293T质粒转染和gDNA提取

[4810] HEK293T细胞以35,000个细胞/孔的密度接种到48孔聚-D-赖氨酸处理的BioCoat板(Corning)上,并在铺板后18至24小时进行转染。使用NucleoCounter NC-200(Chemometec)对细胞进行计数。向这些细胞中加入750ng碱基编辑器或核酸酶对照、250ng sgRNA和10ng GFP-max质粒(Lonza),在Opti-MEM还原血清培养基(ThermoFisher Scientific)中稀释至12.5μL总体积。将溶液与11μL Opti-MEM还原血清培养基中的1.5μL Lipofectamine 2000(ThermoFisher)混合,并在室温下静置15分钟。然后将整个25μL混合物转移到预先接种的Hek293T细胞中,并使其培育约120小时。培育后,吸出培养基并用250μL 1x PBS溶液(ThermoFisher Scientific)洗涤细胞两次,并加入100μL新鲜制备的裂解缓冲液(100mM Tris-HCl,pH 7.0,0.05% SDS,25μg/mL蛋白酶K(Thermo Fisher Scientific)含有裂解缓冲液的转染板在37℃下培育1小时,然后将混合物转移到96孔PCR板中,并在80℃下加热30分钟。

[4811] 分析ABE架构和ABE8构建体的DNA和RNA脱靶编辑

[4812] 在脂质转染前16至20小时,将HEK293T细胞以每孔30,000个细胞的密度接种在不含抗生素的DMEM+Glutamax培养基(Thermo Fisher Scientific)中的48孔聚D-赖氨酸涂层板(Corning)上。750ng切口酶或碱基编辑器表达质粒DNA与250ng sgRNA表达质粒DNA混合在15μl OPTIMEM+Glutamax中。将其与10μl脂质混合物组合,每孔包括1.5μl Lipofectamine 2000和8.5μl OPTIMEM+Glutamax。转染后3天收获细胞并收获DNA或RNA。对于DNA分析,将细胞在1X PBS中洗涤一次,然后根据制造商的说明在100μl

QuickExtractTMBuffer(Lucigen)中裂解。对于RNA收获,根据制造商的说明,将MagMAXTMmirVanaTMTotal RNA Isolation Kit(Thermo Fisher Scientific)与KingFisherTMFlex纯化系统一起使用。

[4813] 靶向的RNA测序主要按照之前的描述进行(参见Rees,H.A.et al.,Analysis and minimization ofcellular RNA editoring by DNA adenine base editors.Sci Adv 5,eaax5717,doi:10.1126/sciadv.aax5717(2019))。根据制造商的说明,使用SuperScript IV一步法RT-PCR系统和EZDNase(Thermo Fisher Scientific)从分离的RNA中制备cDNA。使用以下程序:58℃12分钟;98℃2分钟;随后进行PCR循环,其因扩增子而异:对于CTNNB1和IP90:32个循环[98℃持续10秒;60℃10秒;72℃持续30秒]和RSL1D1 35个循环[98℃持续10秒;58℃10秒;72℃30秒]。没有与样本同时运行RT对照。在组合的RT-PCR之后,使用如上所述Illumina Miseq对扩增子进行条形码标记和测序。每个扩增子中的第一个125个核苷酸,从每个扩增子中正向引物末端后的第一个碱基开始,与参考序列比对并用于分析每个扩增子中的平均和最大A-to-I频率(图53A)和53B)。

[4814] 脱靶DNA测序是使用列于下表16中先前发表的引物进行的(参见Komor,A.C.et al.,Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage.Nature 533,420-424,doi:10.1038/nature17946(2016);Rees,H.A.et al.,Analysis and minimization of cellular RNA editing by DNA adenine base editors.Sci Adv 5,eaax5717,doi:10.1126/sciadv.aax5717(2019)),使用两步PCR和条形码方法以制备用于使用上述Illumina Miseq测序仪测序的样本。

[4815] 表16:用于扩增基因组位点的HTS引物:

[4816]

引物名称	序列
fwd_位点_1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCAGCCCCATCTGTCAAACT
rev_位点_1	TGGAGTTCAGACGTGTGCTCTTCCGATCTTGAATGGATTCCTTGGAAACAATGA

[4817]

引物名称	序列
fwd_位点_2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTGAGGGAGAGCCGTGTAGTT
rev_位点_2	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCCTCTCAAAGTGCTGGGAT
fwd_位点_3	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCATCAGGCTCTCAGCTCAG
rev_位点_3	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTCGTGGGTTTGTGGTTGC
fwd_位点_4	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGCCCATTCCCTCTTTAGCCA
rev_位点_4	TGGAGTTCAGACGTGTGCTCTTCCGATCTGAGCCGTTCCCTCTTTGCTA
fwd_位点_5	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAACCTGTGTGACACTTGGCA
rev_位点_5	TGGAGTTCAGACGTGTGCTCTTCCGATCTGTCTGGCCCAAGATCACACA
fwd_位点_6	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNCACGGATAAAGACGCTGGGA
rev_位点_6	TGGAGTTCAGACGTGTGCTCTTCCGATCTGGGGTCCCAGGTGCTGAC
fwd_位点_7	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNTTGATTGTCTCCTTTGCCGC
rev_位点_7	TGGAGTTCAGACGTGTGCTCTTCCGATCTTGACCCAGTGTTTGATAGATCAGT
fwd_位点_8	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNCACCCCTTCAGTCCATGCTT
rev_位点_8	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCTGATGGGGAGGAACGAGT
fwd_位点_9	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCAGCTCAGCCTGAGTGTTGA
rev_位点_9	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCCCACCCTAGTCATTGGAG
fwd_位点_10	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTCAGAGGGACACACTGTGG
rev_位点_10	TGGAGTTCAGACGTGTGCTCTTCCGATCTCACACTCACTCACCCACACA
fwd_位点_11	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTGTGTGGGTGAGTGAGTGTG
rev_位点_11	TGGAGTTCAGACGTGTGCTCTTCCGATCTCACCAAGGTTCACAGCCTGA
fwd_位点_12	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTGTCTCTGCCTGTAGCTGC
rev_位点_12	TGGAGTTCAGACGTGTGCTCTTCCGATCTCGCTCTGGGCTTCATCTTCA
fwd_位点_13	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTGGGATTATGGGTGTGAGCC
rev_位点_13	TGGAGTTCAGACGTGTGCTCTTCCGATCTTGCCTTCCTCCTCTCTCTCC
fwd_位点_14	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTGCAGACCAGATTCGGAGAA

[4818]

引物名称	序列
rev_位点_14	TGGAGTTCAGACGTGTGCTCTTCCGATCTGTTCAGTTTCCAGGGGGTCC
fwd_位点_15	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCCGCACAGCCTTAGTTCAA
rev_位点_15	TGGAGTTCAGACGTGTGCTCTTCCGATCTAACTTGAAGAGACGGCAGCA
fwd_位点_16	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCCCCAGCTACAGAAAGGTC
rev_位点_16	TGGAGTTCAGACGTGTGCTCTTCCGATCTATTTCCACCGCAAAATGGCC
fwd_位点_17	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCACTTCAGCCCAGGAGTAT
rev_位点_17	TGGAGTTCAGACGTGTGCTCTTCCGATCTTGTGTATGGTGAGAGGTAGGGA
fwd_位点_18	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTCTGAGGTCACACAGTGGG
rev_位点_18	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTGAGAGCAGGGACCACATC
fwd_位点_19	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGGGAGGTGGAGAGAGGATGT
rev_位点_19	TGGAGTTCAGACGTGTGCTCTTCCGATCTACTCTTCCTGAGGTCTAGGAACCCG
fwd_位点_20	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCCTGTTCCTAAAGCCCACC
rev_位点_20	TGGAGTTCAGACGTGTGCTCTTCCGATCTACTCTCTGGTTCTGTTTGTGGCCA
fwd_CTNNB1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNATTTGATGGAGTTGGACATGGCC
rev_CTNNB1	TGGAGTTCAGACGTGTGCTCTCCAGCTACTTGTTCTTGAGTGAAGG
fwd_RSL1D1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTGGCTTTCCAAATCAGTGGGTC
rev_RSL1D1	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTCATAAGCTTAGACCAACAAGC
fwd_IP90	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCTGGTTGACCAATCTGTGGTG
rev_IP90	TGGAGTTCAGACGTGTGCTCTCTGCGTCTGGATCAGGTACG
fwd_HEK293_位点 2_off1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTGTGGAGAGTGAGTAAGCCA
rev_HEK293_位点 2_off1	TGGAGTTCAGACGTGTGCTCTTCCGATCTACGGTAGGATGATTTCAGGCA
fwd_HEK293_位点 2_off2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCACAAAGCAGTGTAGCTCAGG
rev_HEK293_位点 2_off2	TGGAGTTCAGACGTGTGCTCTTCCGATCTTTTTTGGTACTCGAGTGTTATTCAG
fwd_HEK293_位点 3_off1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCCCCTGTTGACCTGGAGAA
rev_HEK293_位点 3_off1	TGGAGTTCAGACGTGTGCTCTTCCGATCTCACTGTACTTGCCCTGACCA

[4819]

引物名称	序列
fwd_HEK293_位点3_off2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTGGTGTTGACAGGGAGCAA
rev_HEK293_位点3_off2	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTGAGATGTGGGCAGAAGGG
fwd_HEK293_位点3_off3	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTGAGAGGGAACAGAAGGGCT
rev_HEK293_位点3_off3	TGGAGTTCAGACGTGTGCTCTTCCGATCTGTCCAAAGGCCCAAGAACCT
fwd_HEK293_位点3_off4	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCCTAGCACTTTGGAAGGTCG
rev_HEK293_位点3_off4	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCTCATCTTAATCTGCTCAGCC
fwd_HEK293_位点3_off5	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAAAGGAGCAGCTCTTCCTGG
rev_HEK293_位点3_off5	TGGAGTTCAGACGTGTGCTCTTCCGATCTGTCTGCACCATCTCCCACAA
fwd_HEK293_位点4_off1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGGCATGGCTTCTGAGACTCA
rev_HEK293_位点4_off1	TGGAGTTCAGACGTGTGCTCTTCCGATCTGTCTCCCTTGCACTCCCTGTCTTT
fwd_HEK293_位点4_off2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTTGGCAATGGAGGCATTGG
rev_HEK293_位点4_off2	TGGAGTTCAGACGTGTGCTCTTCCGATCTGAAGAGGCTGCCCATGAGAG
fwd_HEK293_位点4_off3	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGGTCTGAGGCTCGAATCCTG
rev_HEK293_位点4_off3	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTGTGGCCTCCATATCCCTG
fwd_HEK293_位点4_off4	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTTCCACCAGAACTCAGCCC
rev_HEK293_位点4_off4	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCTCGGTTCCTCCACAACAC
fwd_HEK293_位点4_off5	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCACGGGAAGGACAGGAGAAG
rev_HEK293_位点4_off5	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCAGGGGAGGGATAAAGCAG
fwd_HEK_位点_3	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGGAAACGCCCATGCAATTAGTC
rev_HEK_位点_3	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTTGTCAACCAGTATCCCGGTG
fwd_HEK_位点_2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTGAATGGATTCCTTGGAAACAATG
rev_HEK_位点_2	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCAGCCCCATCTGTCAAACT
fwd_HEK_位点_4	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCCTTTCAACCCGAACGGAG

[4820]

引物名称	序列
rev_HEK_位点_4	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGCTGGTCTTCTTTCCCCTCC
fwd_LDLR	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGCCCTGCTTCTTTTTCTCTGGT
rev_LDLR	TGGAGTTCAGACGTGTGCTCTTCCGATCTACCATTAACGCAGCCAACTTCA
fwd_TRAC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCATGAGGTCTATGGACTTCAAGAGCAA
Rev_TRAC	TGGAGTTCAGACGTGTGCTCTTCCGATCTCATCATTGACCAGAGCTCTGGGCAGAA
fwd_CBLB	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCACTTACCAGCATTACTTCCTAAACC
Rev_CBLB	TGGAGTTCAGACGTGTGCTCTTCCGATCTATGGGCTCCACTTTTCAGCTCTGTAA
fwd_CD7	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCAGTTCAGGCACATGTAGGAGGGA
Rev_CD7	TGGAGTTCAGACGTGTGCTCTTCCGATCTACCGCCTGCAGCTGTCGGACACTGGCA
fwd_B2M	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAAAAGATGAGTATGCCTGCCGTG
Rev_B2M	TGGAGTTCAGACGTGTGCTCTTCCGATCTCAGATTGTTTATATCAGATGGGATGGG
fwd_CIITA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTATGCAAGTTTGGTCCTGAGCCCTCCC
Rev_CIITA	TGGAGTTCAGACGTGTGCTCTTCCGATCTGATGTGGGTTCCCTGCGCTCTGCA
fwd_PDCD1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCAGGGACTGAGGGTGGAAGGTCC
Rev_PDCD1	TGGAGTTCAGACGTGTGCTCTTCCGATCTACCTCCGCCTGAGCAGTGGAGAA

[4821] CD34+细胞中使用的ABE编辑器的mRNA生产

[4822] 编辑器被克隆到编码dT7启动子的质粒中,然后是5'UTR、Kozak序列、ORF和3'UTR。dT7启动子在T7启动子内携带一个失活点突变,可防止从环状质粒转录。所述质粒以PCR反应(Q5 Hot Start 2X Master Mix)为模板,其中正向引物校正T7启动子内的SNP,并且反向引物将120A尾部附加到3'UTR。所得PCR产物在Zymo Research 25μg DCC柱上纯化,并用作后续体外转录中的mRNA模板。按照说明手册使用NEB HiScribe High-Yield Kit,但用N1-甲基-假尿苷完全取代尿苷,并用CleanCap AG(Trilink)进行共转录加帽。通过氯化锂沉淀进行反应净化。用于扩增的引物见表17。

[4823] 表17:用于ABE8 T7体外转录反应的引物

[4824]

名称	序列
fwd_IVT	TCGAGCTCGGTACCTAATACGACTCAC
rev_IVT	TTTTTT

孔前48小时,将CD34+细胞解冻并放入含有1% Glutamax(Gibco)、100ng/mL TPO(Peprotech)、SCF(Peprotech)和Flt-3(Peprotech)的X-VIVO 10(Lonza)中

[4827] CD34+细胞的电穿孔

[4828] 解冻后48小时,将细胞离心以去除X-VIVO 10培养基,并在含有0.1% HSA(Akron Biotechnologies)的MaxCyte缓冲液(HyClone)中洗涤。然后将细胞以每毫升1,250,000个细胞的浓度重新悬浮在冷的MaxCyte缓冲液中,并分成多个20μL等分试样。然后根据实验条件将ABE mRNA(0.15μM)和-198HBG1/2sgRNA(4.05μM)等分,并在MaxCyte缓冲液中升至5μL。将20μL细胞以3份为一组添加到5μL RNA混合物中,并装入OC25x3 MaxCyte比色皿的每个室中进行电穿孔。接收电荷后,从腔室中收集25μL并放置在24孔未处理培养板的孔中心。细胞在培养箱(37℃,5% CO_2)中恢复20分钟。20分钟恢复后,将含有1% Glutamax、100ng/mL TPO、SCF和Flt-3的X-VIVO 10添加到细胞中,浓度为1,000,000个细胞/mL。然后将细胞留在培养箱(37℃,5% CO_2)中进一步恢复48小时。

[4829] ABE电穿孔后的红细胞分化

[4830] 电穿孔后静置48小时(培养第0天)后,将细胞离心并以每毫升20,000个细胞移至含有5%人类血清、330μg/mL转铁蛋白(Sigma)、10μg/mL人类胰岛素(Sigma)、2U/mL肝素钠(Sigma)、3U/mL EPO(Peprotech)、100ng/mL SCF(Peprotech)、5μg/mL IL3以及50μM氢化可的松(Sigma)的“第1阶段”IMDM培养基(ATCC)。在培养的第4天,将细胞加入4倍体积的相同培养基。在第7天,将细胞离心并以200,000个细胞/mL移至含有5%人类血清(Sigma)、330μg/mL转铁蛋白、10μg/mL人类胰岛素、2U/mL肝素钠、3U/mL EPO及100ng/mL SCF的“第2阶段”IMDM培养基。在第11天,将细胞离心并以每毫升1,000,000个细胞移至含有5%人类血清、330μg/mL转铁蛋白、10μg/mL人类胰岛素、2U/mL肝素钠和3U/mL EPO的“第3阶段”IMDM培养基中。在第14天,将细胞离心并重新悬浮在与第11天相同的培养基中,但以每毫升5,000,000个细胞加入。在第18天,将分化的红细胞收集成500,000个细胞等分试样,在500μL DPBS(Gibco)中洗涤一次并在UHPLC处理前在-80℃冷冻24小时。

[4831] 用于UHPLC分析的红细胞样本的制备

[4832] 冷冻的红细胞沉淀在室温下解冻。使用ACK裂解缓冲液将沉淀物稀释至5x 104个细胞/μL的最终浓度。用移液管混合样本并在室温下培育5分钟。然后将样本在-80℃冷冻5分钟使其解冻,并在以6,700g离心10分钟之前用移液器混合。小心去除上清液(不干扰细胞碎片沉淀),转移到新板中并且用超纯水稀释至5x 10^3个细胞/μL以用于UHPLC分析。

[4833] 超高效液相色谱(UHPLC)分析

[4834] 在配备二元泵和UV检测器(Thermo Fisher Scientific,Vanquish Horizon)的UHPLC系统上进行球蛋白链的反相分离。固定相是柱温均为60℃由ACQUITY Peptide BEH C18色谱柱(2.1x 150mm,1.7m珠子,300A孔)和AQUITY Peptide BEH C18 VanGuard预柱(2.1x 5mm,1.7μm珠子,300A孔)(皆为Waters Corp)后所组成。将0.1%三氟乙酸(TFA)水溶液(A)和0.08% TFA乙腈溶液(B)以0.25mL/min的流速用以进行洗脱。球蛋白链的分离是使用40至52%B的线性梯度0至10分钟;52至40%B的线性梯度10至10.5分钟;40%B 12分钟实现的。进样量为10μL,在整个分析过程中收集波长为220nm和数据速率为5Hz的紫外光谱。通过血红蛋白标准品的LC/MS分析确认了球蛋白链的身份。

[4835] CD34+细胞的基因组DNA提取

[4836] ABE电穿孔后(48小时后),将等分细胞培养在含有1% Glutamax(Gibco)、100ng/mL TPO(Peprotech)、SCF(Peprotech)及Flt-3(Peprotech))的X-VIVO 10培养基(Lonza)中。在培养后48小时和144小时后,收集并离心100,000个细胞。将50μL Quick Extract(Lucigen)添加到细胞沉淀中,并将细胞混合物转移到96孔PCR板(Bio-Rad)中。裂解物在65℃下加热15分钟,然后在98℃下加热10分钟。细胞裂解物储存在-20℃。

[4837] 序列

[4838] 在以下序列中,小写字母表示卡那霉素抗性启动子区域,粗体序列表示靶向失活部分(Q4*和W15*),斜体序列表示卡那霉素抗性基因(D208N)靶向失活位点,及底线序列表示PAM序列。

[4839] 失活的卡那霉素抗性基因:

[4840]
ccggaattgccagctggggcgccctctggtaaggttgggaagccctgcaaagtaaactggatggctt
tcttgccgccaaggatctgatggcgcaggggatcaagatctgatcaagagacaggatgaggatcct**t**
tcgcATGATCGAATAAGATGGATTGCACGCAGGTTCTCCGGCC**GCTTAGGTGGAGCGCCTATT**CGGC
TATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGC
GCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCG
GCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGA
AGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCG
AGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATT
CGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAG
GATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCA
TGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAA
TGGCCGCTTTTCTGGA*TTCATTAACTGTGGCCGGCT*GGGTGTGGCGGACCGCTATCAGGACATAGCG
TTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACG
GTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTAA

[4841] 在以下序列中,纯文本表示腺苷脱氨酶序列,粗体序列表示衍生自Cas9的序列,斜体序列表示连接子序列,底线序列表示双核定位序列,及双底线序列表示突变。

[4842] CP5(具有MSP“NGC”PID和“D10A”切口酶):

EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNI
VKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPTVAYSVLVVAKVEKGKSKKLKSVK
ELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAKFLQKGNELAL
PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYN
KHRDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYRSTKEVLDATLIHQSITGLYETRIDL
SQLGGD*GGSGGSGGSGGSGGSGGSGGM***DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRH**
SIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVE
EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLI
ALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVN
TEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFI
[4843] **KPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKI**
LTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPK
HSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS
VEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKA
QVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSR
ERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQS
FLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELD
KAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREIN
NYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEGADKRTADGSEFESPKKK
RKV*

[4844] ABE8.1_Y147T_CP5_NGC PAM_单体

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG
LVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE
GILADECAALLCTFFRMPRQVFNAQKKAQSSTD***SGGSSGGSSGSETPGTSESATPESSGGSSGGS***EI
GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVK
KTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPTVAYSVLVVAKVEKGKSKKLKSVKEL
LGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAKFLQKGNELALPS
[4845] **KYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH**
RDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYRSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD*GGSGGSGGSGGSGGSGGSGGM***DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSI**
KKNLIGALLFDSGETAEATRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEED
KKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD
NSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIAL
SLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTE

ITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILT
FRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHS
LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVE
ISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDK
VMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQV
[4846] **SGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER**
MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFL
KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKA
GFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNY
HHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEGADKRTADGSEFESPKKKRK
V*

[4847] pNMG-B335 ABE8.1_Y147T_CP5_NGC PAM_单体

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG
LVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE
GILADECAALLCTFFRMPRQVFNAQKKAQSSTDSGGSSG*GSSGSETPGTSESATPESSGGSSGGS*EI
GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVK
KTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPTVAYSVLVVAKVEKGKSKKLKSVKEL
LGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAKFLQKGNELALPS
KYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYRSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD*GGSGGSGGSGGSGGSGGSGGM***DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSI**
KKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEED
KKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD
NSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIAL
[4848] **SLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTE**
ITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILT
FRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHS
LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVE
ISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDK
VMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQV
SGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER
MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFL
KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKA
GFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNY
HHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEGADKRTADGSEFESPKKKRK
V*

[4849] 具有NGC PAM CP5的pNMG-357_ABE8.14

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGG
LVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITE
GILADECAALLSDFFRMRRQEIKAQKKAQSSTDGGSSGGS*SGSETPGTSESATPESSGGSSGGS*MSE
VEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVM
QNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGIL
ADECAALLCTFFRMPRQVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGS*
EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNI
VKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPTVAYSVLVVAKVEKGKSKKLKSVK
ELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAKFLQKGNELAL
PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYN
KHRDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYRSTKEVLDATLIHQSITGLYETRIDL
SQLGGD*GGSGGSGGSGGSGGSGGSGGM***DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRH**
SIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVE
EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
[4850] **PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLI**
ALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVN
TEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFI
KPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKI
LTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPK
HSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS
VEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKA
QVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSR
ERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQS
FLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELD
KAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREIN
NYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEGADKRTADGSEFESPKKK
RKV*

[4851] ABE8.8-m

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG
LVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITE
GILADECAALLCRFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGS***DK**
[4852] **KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR**
RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH
LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY
DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL
VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET
ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF
LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF
LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK
QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT
VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ
[4853] **NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE**
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT
KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF
VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD
KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYS
VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG
RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF
SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL
DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[4854] ABE8.8-d

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGG
LVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITE
GILADECAALLSDFFRMRRQEIKAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGS*SE
VEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVM
QNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGIL
ADECAALLCRFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGS*DKKYS
IGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYT
RRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRK
KLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVD
AKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDD
[4855] **LDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQ**
QLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG
SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP
WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSG
EQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDN
EENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV
VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEK
LYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVK
KMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD

ENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYG
DYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGR
DFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLV
[4856] **VAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKR**
MLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKR
VILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[4857] ABE8.13-m

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG
LVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITE
GILADECAALLCRFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGS***DK**
KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR
RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH
LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS
GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY
DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL
VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET
ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF
[4858] **LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF**
LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK
QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT
VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ
NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT
KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF
VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD
KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYS
VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG
RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF
SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL
DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[4859] ABE8.13-d

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGG
LVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITE
[4860] GILADECAALLSDFFRMRRQEIKAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGS*SE
VEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVM
QNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGIL

ADECAALLCRFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGS***DKKYS**
IGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYT
RRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRK
KLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVD
AKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDD
LDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQ
QLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG
SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP
WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSG
EQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDN
EENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
[4861] **KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV**
VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEK
LYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVK
KMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD
ENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYG
DYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGR
DFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLV
VAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKR
MLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKR
VILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[4862] ABE8.17-m

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG
LVMQNYRLIDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE
GILADECAALLCYFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGS***DK**
KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR
RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH
LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS
GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY
[4863] **DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL**
VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET
ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF
LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF
LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK
QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT
VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ

NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT
KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF
VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD

[4864] **KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYS**
VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG
RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF
SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL
DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[4865] ABE8.17-d

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGG
LVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITE
GILADECAALLSDFFRMRRQEIKAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGS*SE
VEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVM
QNYRLIDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGIL
ADECAALLCYFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGS***DKKYS**
IGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYT
RRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRK
KLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVD
AKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDD
LDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQ
QLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG
SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP

[4866] **WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSG**
EQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDN
EENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV
VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEK
LYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVK
KMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD
ENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYG
DYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGR
DFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLV
VAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKR
MLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKR
VILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[4867] ABE8.20-m

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG
LVMQNYRLYDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITE
GILADECAALLCRFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGS*DK
KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR
RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH
LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS
GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY
DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL
VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET
ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF
LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF
[4868]
LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK
QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT
VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ
NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT
KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF
VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD
KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYS
VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG
RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF
SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL
DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[4869] ABE8.20-d

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGG
LVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITE
GILADECAALLSDFFRMRRQEIKAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGS*SE
VEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVM
QNYRLYDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGIL
ADECAALLCRFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGS*DKKYS
[4870]
IGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYT
RRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRK
KLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVD
AKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDD
LDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQ
QLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG

SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP
WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSG
EQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDN
EENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV
VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEK
LYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVK
[4871] KMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD
ENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYG
DYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGR
DFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLV
VAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKR
MLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKR
VILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[4872] 01.monoABE8.1_bpNLS+Y147T

[4873] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG
[4874] LVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE
[4875] GILADECAALLCTFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDK
[4876] KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR
[4877] RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH
[4878] LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS
[4879] GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY
[4880] DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL
[4881] VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
[4882] DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET
[4883] ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF
[4884] LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF
[4885] LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK
[4886] QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT
[4887] VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ
[4888] NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
[4889] VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT
[4890] KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF
[4891] VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD
[4892] KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS
[4893] VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG
[4894] RKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF
[4895] SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL

[4896] DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[4897] 02.monoABE8.1_bpNLS+Y147R

[4898] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG

[4899] LVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE

[4900] GILADECAALLCRFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDK

[4901] KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

[4902] RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH

[4903] LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

[4904] GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY

[4905] DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL

[4906] VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF

[4907] DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET

[4908] ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF

[4909] LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF

[4910] LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

[4911] QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT

[4912] VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ

[4913] NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE

[4914] VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT

[4915] KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF

[4916] VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD

[4917] KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS

[4918] VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG

[4919] RKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF

[4920] SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL

[4921] DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[4922] 03.monoABE8.1_bpNLS+Q154S

[4923] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG

[4924] LVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE

[4925] GILADECAALLCYFFRMPRSVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDK

[4926] KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

[4927] RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH

[4928] LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

[4929] GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY

[4930] DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL

[4931] VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF

[4932] DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET

[4933] ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF

[4934] LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF

[4935] LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

[4936] QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT

[4937] VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ

[4938] NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE

[4939] VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT

[4940] KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF

[4941] VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD

[4942] KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS

[4943] VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG

[4944] RKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF

[4945] SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL

[4946] DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[4947] 04.monoABE8.1_bpNLS+Y123H

[4948] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG

[4949] LVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITE

[4950] GILADECAALLCYFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDK

[4951] KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

[4952] RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH

[4953] LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

[4954] GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY

[4955] DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL

[4956] VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF

[4957] DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET

[4958] ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF

[4959] LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF

[4960] LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

[4961] QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT

[4962] VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ

[4963] NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE

[4964] VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT

[4965] KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF

[4966] VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD

[4967] KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS

[4968] VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG

[4969] RKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF

[4970] SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL

[4971] DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[4972] 05.monoABE8.1_bpNLS+V82S

[4973] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG

[4974] LVMQNYRLIDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE

[4975] GILADECAALLCYFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDK

[4976] KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

[4977] RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH

[4978] LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

[4979] GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY

[4980] DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL

[4981] VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF

[4982] DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET

[4983] ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF

[4984] LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF

[4985] LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

[4986] QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT

[4987] VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ

[4988] NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE

[4989] VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT

[4990] KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF

[4991] VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD

[4992] KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS

[4993] VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG

[4994] RKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF

[4995] SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL

[4996] DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[4997] 06.monoABE8.1_bpNLS+T166R

[4998] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG

[4999] LVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE

[5000] GILADECAALLCYFFRMPRQVFNAQKKAQSSRDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDK

[5001] KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

[5002] RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH

[5003] LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

[5004] GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY

[5005] DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL

[5006] VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF

[5007] DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET

[5008] ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF

[5009] LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF

[5010] LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

[5011] QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT

[5012] VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ

[5013] NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE

[5014] VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT

[5015] KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF

[5016] VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD

[5017] KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS

[5018] VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG

[5019] RKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF

[5020] SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL

[5021] DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[5022] 07.monoABE8.1_bpNLS+Q154R

[5023] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG

[5024] LVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE

[5025] GILADECAALLCYFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDK

[5026] KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

[5027] RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH

[5028] LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

[5029] GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY

[5030] DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL

[5031] VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF

[5032] DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET

[5033] ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF

[5034] LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF

[5035] LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

[5036] QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT

[5037] VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ

[5038] NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE

[5039] VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT

[5040] KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF

[5041] VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD

[5042] KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS

[5043] VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG

[5044] RKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF

[5045] SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL

[5046] DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[5047] 08.monoABE8.1_bpNLS+Y147R_Q154R_Y123H

[5048] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG

[5049] LVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITE

[5050] GILADECAALLCRFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDK

[5051] KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

[5052] RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH

[5053] LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

[5054] GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY

[5055] DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL

[5056] VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF

[5057] DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET

[5058] ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF

[5059] LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF

[5060] LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

[5061] QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT

[5062] VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ

[5063] NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE

[5064] VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT

[5065] KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF

[5066] VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD

[5067] KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS

[5068] VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG

[5069] RKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF

[5070] SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL

[5071] DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[5072] 09.monoABE8.1_bpNLS+Y147R_Q154R_I76Y

[5073] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG

[5074] LVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE

[5075] GILADECAALLCRFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDK

[5076] KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

[5077] RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH

[5078] LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

[5079] GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY

[5080] DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL

[5081] VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF

[5082] DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET

[5083] ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF

[5084] LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF

[5085] LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

[5086] QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT

[5087] VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ

[5088] NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE

[5089] VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT

[5090] KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF

[5091] VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD

[5092] KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS

[5093] VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG

[5094] RKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF

[5095] SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL

[5096] DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[5097] 10.monoABE8.1_bpNLS+Y147R_Q154R_T166R

[5098] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG

[5099] LVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE

[5100] GILADECAALLCRFFRMPRRVFNAQKKAQSSRDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDK

[5101] KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

[5102] RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH

[5103] LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

[5104] GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY

[5105] DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL

[5106] VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF

[5107] DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET

[5108] ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF

[5109] LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF

[5110] LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

[5111] QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT

[5112] VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ

[5113] NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE

[5114] VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT

[5115] KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF

[5116] VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD

[5117] KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS

[5118] VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG

[5119] RKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF

[5120] SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL

[5121] DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[5122] 11.monoABE8.1_bpNLS+Y147T_Q154R

[5123] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG

[5124] LVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE

[5125] GILADECAALLCTFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDK

[5126] KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

[5127] RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH

[5128] LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

[5129] GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY

[5130] DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL

[5131] VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF

[5132] DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET

[5133] ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF

[5134] LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF

[5135] LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

[5136] QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT

[5137] VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ

[5138] NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE

[5139] VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT

[5140] KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF

[5141] VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD

[5142] KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS

[5143] VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG

[5144] RKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF

[5145] SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL

[5146] DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[5147] 12.monoABE8.1_bpNLS+Y147T_Q154S

[5148] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG

[5149] LVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE

[5150] GILADECAALLCTFFRMPRSVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDK

[5151] KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

[5152] RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH

[5153] LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

[5154] GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY

[5155] DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL

[5156] VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF

[5157] DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET

[5158] ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF

[5159] LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF

[5160] LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

[5161] QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT

[5162] VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ

[5163] NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE

[5164] VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT

[5165] KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF

[5166] VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD

[5167] KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS

[5168] VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG

[5169] RKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF

[5170] SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL

[5171] DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[5172] 13.monoABE8.1_bpNLS+H123Y123H_Y147R_Q154R_I76Y

[5173] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLC
RFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD
SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[5174] 14.monoABE8.1_bpNLS+V82S+Q154R

[5175] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG

[5176] LVMQNYRLIDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE

[5177] GILADECAALLCYFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDK

[5178] KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

[5179] RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH

[5180] LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

[5181] GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY

[5182] DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL

[5183] VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF

[5184] DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET

[5185] ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF

[5186] LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF

[5187] LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

[5188] QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT

[5189] VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ

[5190] NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE

[5191] VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT

[5192] KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF

[5193] VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD

[5194] KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS

[5195] VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG

[5196] RKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF

[5197] SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL

[5198] DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[5199] 实施例8.帕金森氏病

[5200] 材料和方法

[5201] 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。

[5202] 实施例中使用的ABE序列如下：

[5203] 01.monoABE8.1_bpNLS+Y147T

[5204] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG

[5205] LVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE

[5206] GILADECAALLCTFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDK

[5207] KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

[5208] RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH

[5209] LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

[5210] GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY

[5211] DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL

[5212] VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF

[5213] DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET

[5214] ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF

[5215] LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF

[5216] LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

[5217] QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT

[5218] VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ

[5219] NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE

[5220] VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT

[5221] KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF

[5222] VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD

[5223] KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS

[5224] VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG

[5225] RKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF

[5226] SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL

[5227] DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[5228] 02.monoABE8.1_bpNLS+Y147R

[5229] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG

[5230] LVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE

[5231] GILADECAALLCRFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDK

[5232] KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

[5233] RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH

[5234] LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

[5235] GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY

[5236] DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL

[5237] VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF

[5238] DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET

[5239] ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF

[5240] LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF

[5241] LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

[5242] QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT

[5243] VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ

[5244] NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE

[5245] VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT

[5246] KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF

[5247] VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD

[5248] KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS

[5249] VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG

[5250] RKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF

[5251] SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL

[5252] DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[5253] 03.monoABE8.1_bpNLS+Q154S

[5254] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG

[5255] LVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE

[5256] GILADECAALLCYFFRMPRSVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDK

[5257] KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

[5258] RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH

[5259] LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

[5260] GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY

[5261] DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL

[5262] VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF

[5263] DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET

[5264] ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF

[5265] LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF

[5266] LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

[5267] QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT

[5268] VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ

[5269] NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE

[5270] VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT

[5271] KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF

[5272] VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD

[5273] KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS

[5274] VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG

[5275] RKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF

[5276] SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL

[5277] DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[5278] 04.monoABE8.1_bpNLS+Y123H

[5279] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG

[5280] LVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITE

[5281] GILADECAALLCYFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDK

[5282] KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

[5283] RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH

[5284] LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

[5285] GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY

[5286] DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL

[5287] VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF

[5288] DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET

[5289] ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF

[5290] LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF

[5291] LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

[5292] QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT

[5293] VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ

[5294] NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE

[5295] VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT

[5296] KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF

[5297] VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD

[5298] KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS

[5299] VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG

[5300] RKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF

[5301] SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL

[5302] DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[5303] 05.monoABE8.1_bpNLS+V82S

[5304] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG

[5305] LVMQNYRLIDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE

[5306] GILADECAALLCYFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDK

[5307] KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

[5308] RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH

[5309] LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

[5310] GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY

[5311] DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL

[5312] VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF

[5313] DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET

[5314] ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF

[5315] LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF

[5316] LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

[5317] QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT

[5318] VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ

[5319] NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE

[5320] VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT

[5321] KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF

[5322] VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD

[5323] KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS

[5324] VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG

[5325] RKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF

[5326] SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL

[5327] DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[5328] 06.monoABE8.1_bpNLS+T166R

[5329] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG

[5330] LVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE

[5331] GILADECAALLCYFFRMPRQVFNAQKKAQSSRDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDK

[5332] KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

[5333] RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH

[5334] LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

[5335] GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY

[5336] DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL

[5337] VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF

[5338] DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET

[5339] ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF

[5340] LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF

[5341] LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

[5342] QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT

[5343] VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ

[5344] NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE

[5345] VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT

[5346] KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF

[5347] VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD

[5348] KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS

[5349] VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG

[5350] RKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF

[5351] SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL

[5352] DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[5353] 07.monoABE8.1_bpNLS+Q154R

[5354] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG

[5355] LVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE

[5356] GILADECAALLCYFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDK

[5357] KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

[5358] RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH

[5359] LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

[5360] GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY

[5361] DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL

[5362] VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF

[5363] DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET

[5364] ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF

[5365] LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF

[5366] LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

[5367] QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT

[5368] VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ

[5369] NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE

[5370] VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT

[5371] KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF

[5372] VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD

[5373] KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS

[5374] VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG

[5375] RKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF

[5376] SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL

[5377] DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[5378] 08.monoABE8.1_bpNLS+Y147R_Q154R_Y123H

[5379] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG

[5380] LVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITE

[5381] GILADECAALLCRFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDK

[5382] KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

[5383] RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH

[5384] LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

[5385] GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY

[5386] DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL

[5387] VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF

[5388] DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET

[5389] ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF

[5390] LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF

[5391] LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

[5392] QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT

[5393] VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ

[5394] NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE

[5395] VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT

[5396] KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF

[5397] VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD

[5398] KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS

[5399] VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG

[5400] RKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF

[5401] SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL

[5402] DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[5403] 09.monoABE8.1_bpNLS+Y147R_Q154R_I76Y

[5404] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG

[5405] LVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE

[5406] GILADECAALLCRFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDK

[5407] KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

[5408] RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH

[5409] LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

[5410] GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY

[5411] DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL

[5412] VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF

[5413] DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET

[5414] ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF

[5415] LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF

[5416] LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

[5417] QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT

[5418] VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ

[5419] NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE

[5420] VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT

[5421] KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF

[5422] VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD

[5423] KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS

[5424] VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG

[5425] RKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF

[5426] SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL

[5427] DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[5428] 10.monoABE8.1_bpNLS+Y147R_Q154R_T166R

[5429] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG

[5430] LVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE

[5431] GILADECAALLCRFFRMPRRVFNAQKKAQSSRDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDK

[5432] KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

[5433] RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH

[5434] LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

[5435] GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY

[5436] DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL

[5437] VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF

[5438] DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET

[5439] ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF

[5440] LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF

[5441] LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

[5442] QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT

[5443] VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ

[5444] NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE

[5445] VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT

[5446] KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF

[5447] VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD

[5448] KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS

[5449] VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG

[5450] RKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF

[5451] SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL

[5452] DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[5453] 11.monoABE8.1_bpNLS+Y147T_Q154R

[5454] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG

[5455] LVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE

[5456] GILADECAALLCTFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDK

[5457] KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

[5458] RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH

[5459] LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

[5460] GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY

[5461] DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL

[5462] VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF

[5463] DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET

[5464] ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF

[5465] LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF

[5466] LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

[5467] QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT

[5468] VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ

[5469] NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE

[5470] VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT

[5471] KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF

[5472] VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD

[5473] KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS

[5474] VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG

[5475] RKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF

[5476] SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL

[5477] DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[5478] 12.monoABE8.1_bpNLS+Y147T_Q154S

[5479] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG

[5480] LVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE

[5481] GILADECAALLCTFFRMPRSVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDK

[5482] KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

[5483] RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH

[5484] LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

[5485] GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY

[5486] DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL

[5487] VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF

[5488] DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET

[5489] ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF

[5490] LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF

[5491] LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

[5492] QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT

[5493] VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ

[5494] NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE

[5495] VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT

[5496] KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF

[5497] VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD

[5498] KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS

[5499] VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG

[5500] RKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF

[5501] SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL

[5502] DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[5503] 13.monoABE8.1_bpNLS+H123Y123H_Y147R_Q154R_I76YMSEVEFSHEYWMRHALTLAKRA
RDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRI
GRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDSGGSSGGSS
GSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDS
GETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY
PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAK
AILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQY
ADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYI
DGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKI
EKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY
FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY
HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIR
DKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELV
KVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVD
QELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTK
AERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREIN
NYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGE
IRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGG
FVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGR
KRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANL
DKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSITGLYETRIDLSQ
LGGDEGADKRTADGSEFESPKKKRKV

[5504] 14.monoABE8.1_bpNLS+V82S+Q154R

[5505] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG

[5506] LVMQNYRLIDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE

[5507] GILADECAALLCYFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDK

[5508] KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

[5509] RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH

[5510] LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

[5511] GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY

[5512] DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL

[5513] VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF

[5514] DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET

[5515] ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF

[5516] LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF

[5517] LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

[5518] QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT

[5519] VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ

[5520] NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE

[5521] VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT

[5522] KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF

[5523] VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD

[5524] KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYS

[5525] VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG

[5526] RKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF

[5527] SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL

[5528] DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[5529] 包含氨基酸置换D1135M、S1136Q、G1218K、E1219F、A1322R、D1332A、R1335E及T1337R(MQKFRAER)且对改变的PAM 5'-NGC-3'具有特异性的修饰的SpCas9是用于校正G。对改变的PAM 5'-NGA-3具有特异性的改良SpCas9-VRQR用于校正R1441C。

[5530] 克隆.

[5531] 本文描述目标多核苷酸和gRNA的DNA序列以及使用的引物。对于gRNA,提供以下支架序列:GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU AAGGCUAGUC CGUUAUCAAC UUGAAAAAAGU GGCACCGAGU CGGUGCUUUU。所述支架用于图57A至C和图58A至中描述的PAM(例如,分别为NGA和NGC PAM)。gRNA涵盖LRRK2基因的支架序列和间隔序列(目标序列),其包括如本文所述或根据本领域技术人员的知识确定的并且如本领域技术人员所理解的。(参见,例如,Komor,A.C.,et al.,"Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage"Nature 533,420-424(2016);Gaudelli,N.M.,et al.,"Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage"Nature 551,464-471(2017);Komor,A.C.,et al.,"Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity"Science Advances3:eaao4774(2017)及Rees,H.A.,et al.,"Base editing:precision chemistry on the genome and transcriptome of living cells."Nat Rev Genet.2018Dec;19(12):770-788.doi:10.1038/s41576-018-0059-1)。

[5532] 将VeraSeq ULtraDNA聚合酶(Enzymatics)或Q5热启动高保真DNA聚合酶(New England Biolabs)用以进行PCR。碱基编辑器(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。将脱氨酶和融合基克隆到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。

[5533] 简而言之,根据制造商的说明,使用T4多核苷酸激酶(New England Biolabs)将引物5'磷酸化。为了扩增向导1和2的区域,使用了以下引物:

[5534] 向导1引物(oAM129;对于5'-AAGCGCAAGCCTGGAGGGAA-3'):

[5535] 5'-GAAGCGCAAGCCTGGAGGGAAGTTTTAGAGCTAGAAATAGCA-3';

[5536] 向导2引物(oAM130;5’-ACTACAGCATTGCTCAGTAC-3’)：

[5537] 5’-GACTACAGCATTGCTCAGTACGTTTTAGAGCTAGAAATAGCA-3’；

[5538] 共同引物(oAM95)：

[5539] 5’-GGTGTTTCGTCCTTTCCACAAG-3’。

[5540] 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)与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℃培育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。

[5541] ssDNA的体外脱氨酶测定。

[5542] 以下提供所有ssDNA底物的序列。所有Cy3标记的底物均从Integrated DNA Technologies(IDT)获得。使用TNT T7快速偶联转录/翻译试剂盒(Promega)根据制造商的说明使用1μg质粒在体外表达脱氨酶。蛋白质表达后,将5μl裂解物与35μl ssDNA(1.8μM)和USER酶(1单位)在CutSmart缓冲液(新英格兰生物实验室)(50mM醋酸钾、29mM Tris-醋酸盐、10mM醋酸镁)中混合,100μg ml-1BSA,pH 7.9)并在37℃下培育2小时。在10％ TBE-尿素凝胶(Bio-Rad)上从全长未修饰的底物中分离出裂解的含U底物。

[5543] His6-ABE8/PV1-28-连接子-dCas9融合体的表达和纯化。

[5544] 用质粒(例如编码pET28b-His6-ABE8/PV1-28-连接子-dCas9的质粒)转换大肠杆菌BL21 STAR(DE3)感受态细胞(ThermoFisher Scientific)。所得表达菌株在含有100μg ml-1卡那霉素的Luria-Bertani(LB)肉汤中在37℃下生长过夜。将细胞以1:100稀释到相同的生长培养基中,并在37℃生长至OD600＝～0.6。在2小时内将培养物冷却至4℃,并添加0.5mM异丙基-β-d-1-硫代吡喃半乳糖苷(IPTG)以诱导蛋白质表达。约16小时后,通过以4,000g离心收集细胞并重悬于裂解缓冲液(50mM三(羟甲基)-氨基甲烷(Tris)-HCl(pH 7.5)、1M NaCl、20％甘油、10mM三(2-羧乙基)膦(TCEP,Soltec Ventures))。通过超声处理(20秒脉冲开启,20秒脉冲关闭,总共8分钟,输出功率为6瓦)裂解细胞,并在以25,000g离心15分钟后分离裂解液上清液。将裂解物与His-Pur镍-氨基乙酸(镍-NTA)树脂(ThermoFisher Scientific)在4℃培育1小时以捕获带有His标签的融合蛋白。将树脂转移到柱上并用40ml裂解缓冲液洗涤。His标记的融合蛋白在补充有285mM咪唑的裂解缓冲液中洗脱,并通过超滤(Amicon-Millipore,100-kDa分子量截止)浓缩至1ml总体积。将蛋白质在含有50mM三(羟甲基)-氨基甲烷(Tris)-HCl(pH7.0)、0.1M NaCl、20％甘油、10mM TCEP的低盐纯化缓冲液中稀释至20ml,并加载到SP Sepharose Fast Flow上树脂(GE Life Sciences)。用40ml这种低盐缓冲液洗涤树脂,然后用5ml活性缓冲液洗脱蛋白质,其中含有50mM三(羟甲基)-氨基甲烷(Tris)-HCl(pH 7.0)、0.5M NaCl、20％甘油、10mM TCEP。洗脱的蛋白质通过SDS-PAGE进行定量。

[5545] sgRNA的体外转录.

[5546] 根据制造商的说明,使用TranscriptAid T7高产量转录试剂盒(ThermoFisher Scientific)在体外转录含有T7启动子和20个碱基对的sgRNA目标序列的线性DNA片段。根据制造商的说明使用MEGAclear试剂盒(ThermoFisher Scientific)纯化sgRNA产物,并通过紫外吸亮度进行定量。

[5547] Cy3偶联的dsDNA底物的制备.

[5548] 通常,未标记的序列链(例如80个核苷酸未标记链的序列)订购为来自IDT的PAGE纯化的寡核苷酸。与每个80个核苷酸底物的3'端互补的25个核苷酸Cy3标记引物订购自IDT的HPLC纯化的寡核苷酸。为了产生Cy3标记的dsDNA底物,将80nt链(5μl 100μM溶液)与Cy3标记的引物(5μl 100μM溶液)在NEBuffer 2(38.25μl的50mM NaCl、10mM Tris-HCl、10mM $MgCl_2$、1mM DTT、pH 7.9溶液,New England Biolabs)和dNTP(0.75μl的100mM溶液)并加热至95℃5分钟,然后以每秒0.1℃的速率逐渐冷却至45℃。在此退火期后,添加Klenow exo-(5U,New England Biolabs)并将反应在37℃培育1小时。将溶液用缓冲液PB(250μl,Qiagen)和异丙醇(50μl)稀释,并在QIAprep离心柱(Qiagen)上纯化,及用50μl Tris缓冲液洗脱。在dsDNA上的脱氨酶测定。将纯化的融合蛋白(20μl的活性缓冲液中的1.9μM)与1当量的适当sgRNA混合,并在环境温度下培育5分钟。加入Cy3标记的dsDNA底物至终浓度为125nM,所得溶液在37℃培育2小时。通过添加缓冲液PB(100μl,Qiagen)和异丙醇(25μl)将dsDNA从融合体中分离出来,并在EconoSpin微量离心柱(Epoch Life Science)上纯化,用20μl CutSmart缓冲液(New England Biolabs)。将USER酶(1U,New England Biolabs)添加到纯化、编辑的dsDNA中,并在37℃下培育1小时。通过将5μl反应溶液与15μl基于DMSO的上样缓冲液(5mM Tris、0.5mM EDTA、12.5%甘油、0.02%溴酚蓝、80% DMSO)混合,Cy3标记的链从其补体中完全变性。在10% TBE-尿素凝胶(Bio-Rad)上将全长含C底物与任何切割的含U编辑底物分离,并在GE Amersham Typhoon成像仪上成像。

[5549] 用于高通量测序的体外编辑dsDNA的制备.

[5550] 寡核苷酸获自IDT。互补序列在Tris缓冲液中组合(5μl的100μM溶液)并通过加热至95℃5分钟进行退火,然后以每秒0.1℃的速率逐渐冷却至45℃以产生60碱基对的dsDNA底物。将纯化的融合蛋白(20μl的1.9μM活性缓冲液)与1当量的适当sgRNA混合,并在环境温度培育5分钟。加入60聚体dsDNA底物至终浓度为125nM,所得溶液在37℃培育2小时。通过添加缓冲液PB(100μl,Qiagen)和异丙醇(25μl)将dsDNA与融合体分离,并在EconoSpin微量离心柱(Epoch Life Science)上纯化,用20μl Tris缓冲液洗脱。根据制造商的说明,通过使用高通量测序引物对和VeraSeq Ultra(Enzymatics)的PCR扩增得到编辑的DNA(1μl用作模板),进行13个扩增循环。PCR反应产物使用RapidTips(Diffinity Genomics)纯化,纯化的DNA使用含有测序连接子的引物通过PCR扩增,纯化,并如前所述在MiSeq高通量DNA测序仪(Illumina)上测序。

[5551] 细胞培养.

[5552] HEK293T(ATCC CRL-3216)和U2OS(ATCC HTB-96)维持在Dulbecco's Modified Eagle's Medium plus GlutaMax(ThermoFisher)中,添加10%(v/v)胎儿牛血清(FBS),37℃,添加5% CO_2。HCC1954细胞(ATCC CRL-2338)维持在如上所述补充的RPMI-1640培养基(ThermoFisher Scientific)中。含有LRRK2)(Taconic Biosciences)的永生化细胞在补充有10%(v/v)胎儿牛血清(FBS)和200μg ml-1遗传素(ThermoFisher Scientific)的Dulbecco's Modified Eagle's Medium plus GlutaMax(ThermoFisher Scientific)中培养。

[5553] 转染.

[5554] 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,使用适当的Amaxa Nucleofector II程序(使用程序Q-001用于HEK293T细胞的V试剂盒)转染HEK293T细胞。

[5555] 基因组DNA样本的高通量DNA测序.

[5556] 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)单离基因组DNA。靶上和脱靶目标基因组区域通过PCR与侧边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。使用含有测序连接子的引物通过PCR扩增纯化的DNA。将Quant-iT PicoGreen dsDNA Assay Kit(ThermoFisher)和KAPA Library Quantification Kit-Illumina(KAPA Biosystems)用以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上对样本进行测序。

[5557] 数据分析.

[5558] 将MiSeq Reporter(Illumina)自动解复(demultiplexed)用测序读数,并使用定制的Matlab分析个别FASTQ档案。将Smith-Waterman算法用以将每个读数与适当的参考序列成对比对。Q分数低于31的碱基频率用被N置换,因此在计算核苷酸频率时被排除在外。这种处理产生的MiSeq碱基识别错误率大约为千分之一。读数和参考序列不包括空位的比对序列存储在比对表中,从中可以为每个基因座列出碱基频率。插入缺失频率使用先前描述的标准(Zuris,et al.,Nature Biotechnol.33,73-80(2015))通过自定义Matlab脚本进行量化。扫描测序读数是否与位于两侧的其中可能发生插入缺失的窗口的两个10碱基对序列完全匹配。若没有精确匹配,则读取被排除在分析之外。若这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括插入缺失。若插入缺失窗口是两个或更多个比参考序列更长或更短的碱基,则测序读数分别被归类为插入或缺失。

[5559] 碱基编辑器中的PAM变体验证

[5560] 新颖CRISPR系统和PAM变体使碱基编辑器(例如PV1至PV28)能够对LRRK多核苷酸中存在的目标SNP进行精确校正。已经评估和验证了几种新的PAM变体。PAM评估和碱基编辑器的细节描述于例如国际PCT申请号PCT/2017/045381(WO2018/027078);PCT/US2016/058344(WO2017/070632);Kleinstiver,B.P.,et al.,“Engineered CRISPR-Cas9 nucleases with altered PAM specificities”Nature523,481-485(2015);以及Kleinstiver,B.P.,et al.,“Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition”Nature Biotechnology33,1293-1298(2015),其各自通过引用方式整体并入本文。亦参见Komor,A.C.,et al.,“Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage”Nature 533,420-424(2016);Gaudelli,N.M.,et al.,“Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage”Nature551,464-471(2017);及Komor,A.C.,et al.,“Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher

efficiency and product purity"Science Advances 3:eaao4774(2017),其各自全部内容通过引用方式并入本文。

[5561] 基因编辑校正帕金森氏病突变

[5562] LRRK2中的致病突变R1441C和R1441H与帕金森氏病有关。如图55所示,R1441C突变与LRRK2基因反义链中的G>A突变相关,并且使用具有腺苷脱氨酶活性和AGA PAM特异性的碱基编辑器校正。如图55所示,RRK中的R1441H由LRRK2基因中的G>A突变编码,并且使用具有腺苷脱氨酶活性和对目标序列位置3具有TGA PAM特异性或位置5具有TGT PAM特异性的碱基编辑器校正。

[5563] 图56是显示用于校正与帕金森氏病相关的LRRK2中的Y1699C、G2019S及I2020T突变的目标序列的示意图。Y1699C突变与LRRK2基因反义链上的T>C突变相关,所述突变使用具有胞苷脱氨酶活性的碱基编辑器进行校正。G2019S突变与LRRK2基因反义链上的G>A突变相关,所述突变使用具有腺苷脱氨酶活性的碱基编辑器进行校正。I2020T是由LRRK2基因中的T>C突变编码,所述突变使用具有胞苷脱氨酶活性和TGC PAM特异性的碱基编辑器进行校正。

[5564] 如图57A至57C所示,编辑器PV 1至14使用具有图57B所示序列的向导RNA使用编辑器编辑LRRK2 R1441C,但目标序列中的所有胸苷(T)都被尿苷(U)取代(向导RNA1:5'-AAGCGCAAGCCUGGAGGGAA-3')。A至G的转换百分比示于图57A。例示性序列读数显示于图57。

[5565] 如图58A至58C中所示,编辑器PV 15至28使用具有图58B所示序列的向导RNA使用编辑器编辑LRRK2 G2019S,但目标序列中的所有胸苷(T)都被尿苷(U)取代取代(向导RNA2:5'-ACUACAGCAUUGCUCAGUAC-3')。靶上位点和脱靶位点的A至G的转换百分比显示在图58C中。例示性序列读数显示于图58C。编辑器(PV15至28)用于编辑G2019S。

[5566] 用于校正LRRK突变的编辑器(PV15-28)的描述如下:

[5567] PV1(也称为PV15).pCMV_monoABE8.1_bpNLS+Y147T

[5568] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD

[5569] PV2(也称为PV16).pCMV_monoABE8.1_bpNLS+Y147R

[5570] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCRFFRMPRQVFNAQKKAQSSTD

[5571] PV3(也称为PV17).pCMV_monoABE8.1_bpNLS+Q154S

[5572] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRSVFNAQKKAQSSTD

[5573] PV4(也称为PV18).pCMV_monoABE8.1_bpNLS+Y123H

[5574] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

[5575] PV5(也称为PV19).pCMV_monoABE8.1_bpNLS+V82S

[5576] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
YFFRMPRQVFNAQKKAQSSTD

[5577] PV6(也称为PV20).pCMV_monoABE8.1_bpNLS+T166R

[5578] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
YFFRMPRQVFNAQKKAQSSRD

[5579] PV7(也称为PV21).pCMV_monoABE8.1_bpNLS+Q154R

[5580] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
YFFRMPRRVFNAQKKAQSSTD

[5581] PV8(也称为PV22).pCMV_monoABE8.1_bpNLS+Y147R_Q154R_Y123H

[5582] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLC
RFFRMPRRVFNAQKKAQSSTD

[5583] PV9(也称为PV23).pCMV_monoABE8.1_bpNLS+Y147R_Q154R_I76Y

[5584] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
RFFRMPRRVFNAQKKAQSSTD

[5585] PV10(也称为PV24).pCMV_monoABE8.1_bpNLS+Y147R_Q154R_T166R

[5586] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
RFFRMPRRVFNAQKKAQSSRD

[5587] PV11(也称为PV25).pCMV_monoABE8.1_bpNLS+Y147T_Q154R

[5588] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
TFFRMPRRVFNAQKKAQSSTD

[5589] PV12(也称为PV26).pCMV_monoABE8.1_bpNLS+Y147T_Q154S

[5590] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
TFFRMPRSVFNAQKKAQSSTD

[5591] PV13(也称为PV27).pCMV_monoABE8.1_bpNLS+H123Y123H_Y147R_Q154R_I76Y

[5592] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLC
RFFRMPRRVFNAQKKAQSSTD

[5593] PV14(也称为PV28).pCMV_monoABE8.1_bpNLS+V82S+Q154R

[5594] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
YFFRMPRRVFNAQKKAQSSTD

[5595] 图59A至59L提供编码R1441C的LRRK2目标序列的位置7处的A至G转变的实施例性序列读数。编辑器显示为(PV1至14)。

[5596] 图60A至60W描绘了编码G2019S的LRRK2目标序列的位置4和6处的A至G转变的序列读数。编辑器显示为(PV15至28)。

[5597] 使用类似策略校正LRRK2中与帕金森氏病相关的其他致病性突变(图61A至61D)。

[5598] 实施例9.用于校正小鼠α-L-艾杜糖基化酶(IDUA)基因中的W401X突变的贺勒氏病碱基编辑。

[5599] 贺勒氏症是1型粘多糖贮积症(MPS1)中最严重的一种。MPS1是由α-L-艾杜糖醛酸酶(IDUA)基因突变引起的。目前,还没有包括人类IDUA基因的转基因小鼠模型。然而,小家鼠IDUA蛋白质氨基酸序列和智人IDUA蛋白质氨基酸序列之间存在高度保守性。在人类中,与严重贺勒氏症表型相关的MPS1的常见突变是W402X。所述突变是在IDUA蛋白的位置402(W402X)引入终止密码子的单碱基取代,并且与纯合子中极其严重的临床表型相关。在小鼠中,IDUA蛋白的等效突变是W401X。ABE可用于通过在目标位点有效地转换A>G来校正W401X突变,而校正小家鼠IDUA基因。SNP处的A>G校正将IDUA多肽中位置401(W401X)的终止密码子更改为色氨酸。

[5600] W401X突变的目标是使用A·T至G·C DNA碱基编辑器(ABE)恢复为野生型序列,所述编辑器采用具有经过验证的前间隔邻近基序(PAM)序列偏好的Cas9部分。为了确定哪种向导RNA(gRNA)和ABE8-Cas9平台能够最有效和精确地校正靶向IDUA突变,通过慢病毒转导将携带贺勒氏症W401X靶向突变的小家鼠IDUA等位基因基因组整合到HEK293T细胞中。如上所述,使用Opti-MEM培养基和Lipofectamine2000转染HEK293T细胞,并以每孔30,000个细胞接种具有250ng gRNA和750ng ABE8变异碱基编辑器表达质粒的48孔板。ABE8碱基编辑器变体含有NGG PAM序列(即,SpCas9)。细胞在转染后5天裂解并准备进行测序(在转染后第3天更换培养基),并通过miSeq分析在所需位点分析碱基编辑。

[5601] 小家鼠IDUA的DNA靶向/插入序列如下所示,对应于在NCBI参考序列号NM_008325.4发现的代表性小家鼠IDUA基因序列的核酸1077至1358。

[5602]
CTCCCAGCGCACACTTACTGCTCGATTCCAGGTCAACAATACTCACC
CACCCCACGTGCAGTTGCTGCGAAAGCCAGTACTCACAGTCATGGGGCTC
ATGGCCCTGTTGGATGGAGAACAACTCT**A**GGCAGAGGTCTCAAAGGCTGG
GGCTGTGTTGGACAGCAATCATACAGTGGGTGTCCTGGCCAGCACCCATC
ACCCTGAAGGCTCCGCAGCGGCCTGGAGTACCACAGTCCTCATCTACACT
AGTGATGACACCCACGCACACCCCAACCACAGTAT

[5603] 上述W401X突变的小家鼠DNA靶/插入序列包括一个“A”核碱基(以粗体和底线显示),而小家鼠IDUA基因序列在位置1202包括一个“G”核碱基。

[5604] 测试了两种向导RNA对小家鼠IDUA W401X突变的碱基编辑。gRNA涵盖如本文所提供的或基于熟练从业者的知识确定的以及如本领域熟练的从业者将理解的疾病相关基因的支架序列和间隔序列(目标序列)。(参见,例如,Komor,A.C.,et al.,“Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage”Nature 533,420-424(2016);Gaudelli,N.M.,et al.,“Programmable base editing of A·T to G·C in genomic DNA without DNA cleavage”Nature 551,464-471(2017);

Komor,A.C.,et al.,“Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity”Science Advances 3:eaao4774(2017);及Rees,H.A.,et al.,“Base editing:precision chemistry on the genome and transcriptome of living cells.”Nat Rev Genet.2018Dec;19(12):770-788.doi:10.1038/s41576-018-0059-1)。

[5605] 测试了靶向IDUA W401X突变和ABE8碱基编辑器变体的21个核苷酸向导RNA(gRNA)(图62A和62B)。与上述DNA目标序列的互补序列杂交的21个核苷酸的gRNA序列如下所示:gACTCTAGGCAGAGGTCTCAAAGG.NGG PAM序列(即SpCas9)在以上有底线。gRNA序列中的小写“g”表示其中聚合酶(例如Pol III)必须启动转录的序列中的错配。向导RNA包括序列UUGAGACCUCUGCCUAGAGU。

[5606] 对于上述gRNA序列,支架序列如下:

[5607] GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT。

[5608] 使用的ABE碱基编辑器包括ABE8单体变体:ABE8.1、ABE8.12、ABE8.13、ABE8.4、ABE8.5、ABE8.6、ABE8.7、ABE8.8、ABE8.9、ABE8.10、ABE8.11、ABE8.12和ABE8.13。阳性对照碱基编辑器ABE7.10和阴性对照也用于比较。

[5609] 与ABE7.10阳性对照相比,ABE8碱基编辑器变体具有相当或增加的碱基编辑活性,其中W401X具有约40%的碱基编辑校正(图62A)。如图62B所示,插入缺失形成的百分比与约0.4至0.6%的ABE7.10阳性对照相当。

[5610] 还测试了靶向具有ABE8碱基编辑器变体的IDUA W401X突变的20个核苷酸向导RNA(gRNA),并与上述21个核苷酸gRNA进行比较(图63)。与上述DNA目标序列的互补序列杂交的20个核苷酸的gRNA序列如下所示:ACTCTAGGCAGAGGTCTCAA AGG.NGG PAM序列(即SpCas9)在以上有底线。

[5611] 对于上述gRNA序列,支架序列如下:

[5612] GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT。

[5613] 使用的ABE碱基编辑器包括ABE8变体:ABE8.1、ABE8.12、ABE8.13、ABE8.4、ABE8.5、ABE8.6、ABE8.7、ABE8.8、ABE8.9、ABE8.10、ABE8.11、ABE8.12及ABE8.13。阳性对照碱基编辑器ABE7.10和阴性对照也用于比较。

[5614] ABE8碱基编辑器变体使用20个核苷酸gRNA或21个核苷酸gRNA具有约W401X 40%的碱基编辑校正(图63)。

[5615] 实施例10.用于校正人类α-L-艾杜糖基化酶(IDUA)基因中的W402X突变的贺勒氏症碱基编辑。

[5616] 贺勒氏症是1型粘多糖贮积症(MPS1)中最严重的一种MPS1是由α-L-艾杜糖醛酸酶(IDUA)基因突变引起的。在人类中,与严重贺勒氏症表型相关的MPS1的常见突变是W402X。所述突变是单碱基替换,在IDUA蛋白的402位(W402X)引入终止密码子,并且与纯合子中极其严重的临床表型相关。ABE可用于通过在目标位点有效地转换A>G来校正W402X突变以校正人IDUA基因。

[5617] W402X突变的目标是使用A·T至G·C DNA碱基编辑器(ABE)恢复为野生型序列,所

述编辑器采用具有经过验证的前间隔邻近基序(PAM)序列偏好的Cas9部分。如图64所示,ABE碱基编辑器和向导RNA(gRNA)可用于靶向智人IDUA核酸序列中的腺苷(A)核碱基(盒装)以校正W402X突变。SNP处的A>G校正将IDUA多肽中位置402(W402X)的终止密码子更改为色氨酸。

[5618] 为了确定哪个ABE8-Cas9平台能够最有效和精确地校正靶向IDUA突变,通过慢病毒转导将带有贺勒氏症W402X可靶向突变的智人IDUA等位基因基因组整合到HEK293T细胞中。如上所述,使用Opti-MEM培养基和Lipofectamine 2000转染HEK293T细胞,并以每孔30,000个细胞接种具有250ng gRNA和750ng ABE8变异碱基编辑器表达质粒的48孔板。ABE8碱基编辑器变体含有NGG PAM序列(即SpCas9)。细胞在转染后5天裂解并准备进行测序(在转染后第3天更换培养基),并通过miSeq分析在所需位点分析碱基编辑。

[5619] 智人IDUA多核苷酸序列中的DNA靶向/插入序列如下所示,对应于在NCBI参考序列号NM_000203.5发现的代表性智人IDUA基因序列的核酸1076至1358。

[5620]
CCTTCGCGCAGCGCACGCTCACCGCGCGCTTCCAGGTCAACAACACCCGCCCGCCGCACGTGCAGCT
GTTGCGCAAGCCGGTGCTCACGGCCATGGGGCTGCTGGCGCTGCTGGATGAGGAGCAGCTCT**A**GGCC
GAAGTGTCGCAGGCCGGGACCGTCCTGGACAGCAACCACACGGTGGGCGTCCTGGCCAGCGCCCACC
GCCCCCAGGGCCCGGCCGACGCCTGGCGCGCCGCGGTGCTGATCTACGCGAGCGACGACACCCGCGC

[5621] CCACCCCAACCGCAG

[5622] 上述智人目标/插入序列包括“A”核碱基(以粗体和底线显示),而智人IDUA基因序列在IDUA序列的位置1205含有“G”核碱基。

[5623] 用ABE8碱基编辑器变体测试了靶向W402X突变的20核苷酸向导RNA(gRNA)(图65A)。gRNA涵盖如本文所提供的或基于熟练从业者的知识确定的以及如本领域熟练的从业者将理解的疾病相关基因的支架序列和间隔序列(目标序列)。(参见,Komor,A.C.,et al.,“Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage”Nature 533,420-424(2016);Gaudelli,N.M.,et al.,“Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage”Nature 551,464-471(2017);Komor,A.C.,et al.,“Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity”Science Advances 3:eaao4774(2017);及Rees,H.A.,et al.,“Base editing:precision chemistry on the genome and transcriptome of living cells.”Nat Rev Genet.2018Dec;19(12):770-788.doi:10.1038/s41576-018-0059-1)。

[5624] 与上述DNA目标序列的互补序列杂交的gRNA序列为:GCTCTAGGCCGAAGTGTCGCAGG.NGG PAM序列(即SpCas9)以上有底线。

[5625] 对于上述gRNA序列,支架序列是如下:

[5626] GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATC AACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT

[5627] 使用的ABE碱基编辑器包括ABE8变体:ABE8.1、ABE8.2、ABE8.3、ABE8.4、ABE8.5、ABE8.6、ABE8.7、ABE8.8、ABE8.9、ABE8.10、ABE8.11、ABE8.12及ABE8.13。阳性对照碱基编辑器ABE7.10和阴性对照也用于比较。

[5628] 与ABE7.10阳性对照相比,ABE8碱基编辑器变体具有相当或增加的碱基编辑活性,其中W402X具有约30至40%的碱基编辑校正(图65A)。如图65B所示,插入缺失形成的百分比与ABE7.10阳性对照相当或低于约0.2至0.5%。

[5629] 通过PCR产物的深度测序检测到的IDUA核酸序列中位置6的目标"A"核碱基的A至G碱基编辑的效率呈现在图66A至66O。下表18总结了与ABE7.10阳性对照(图66N)和阴性对照(图66O)相比,ABE8碱基编辑器变体(图66A至66M)在IDUA核酸目标位点中的位置6处实现的A至G碱基编辑的百分比。

[5630] 表18.具有基础编辑器变体的IDUA目标位点基础编辑百分比。

[5631]

样本	ABE	A 至 G 碱基编辑(%)
NMG280-1	ABE8.1	31.17%
NMG280-2	ABE8.1	31.65%
NMG280-3	ABE8.1	33.61%
NMG281-1	ABE8.2	41.89%
NMG281-2	ABE8.2	40.96%
NMG281-3	ABE8.2	35.02%
NMG282-1	ABE8.3	34.07%
NMG282-2	ABE8.3	31.86%
NMG282-3	ABE8.3	30.78%
NMG283-1	ABE8.4	32.72%
NMG283-2	ABE8.4	31.87%
NMG283-3	ABE8.4	27.74%
NMG284-1	ABE8.5	8.11%
NMG284-2	ABE8.5	10.02%
NMG284-3	ABE8.5	9.5%
NMG285-1	ABE8.6	39%
NMG285-2	ABE8.6	31.94%
NMG285-3	ABE8.6	37.55%
NMG286-1	ABE8.7	39.84%
NMG286-2	ABE8.7	32.14%
NMG286-3	ABE8.7	40.85%
NMG287-1	ABE8.8	38.92%
NMG287-2	ABE8.8	38.87%
NMG287-3	ABE8.8	35.53%
NMG288-1	ABE8.9	36.73%
NMG288-2	ABE8.9	38.97%
NMG288-3	ABE8.9	28.23%
NMG289-1	ABE8.10	34.1%

[5632]

样本	ABE	A至G碱基编辑(%)
NMG289-2	ABE8.10	33.7%
NMG289-3	ABE8.10	39.92%
NMG290-1	ABE8.11	41.41%
NMG290-2	ABE8.11	36.78%
NMG290-3	ABE8.11	39.83%
NMG291-1	ABE8.12	37.87%
NMG291-2	ABE8.12	45.05%
NMG291-3	ABE8.12	35.94%
NMG292-1	ABE8.13	34.44%
NMG292-2	ABE8.13	33.73%
NMG292-3	ABE8.13	39.7%
NP20-1	ABE7.10	33.73%
NP20-2	ABE7.10	33.93%
NP20-3	ABE7.10	26.89%
Negct-1	阴性对照	0.04%
Negct-2	阴性对照	0.03%

[5633] 如表18和图66A至66M所示,使用ABE8碱基编辑器变体在IDUA核酸序列的位置6(靶向的“A”核碱基位点)实现了平均约33.9%的A至G碱基编辑。这与阳性对照ABE7.10(图66N)相当。

[5634] A)实施例11.细胞培养和转染.

[5635] 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% CO2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。

[5636] 对于质粒转染,将HEK293T细胞铺板,并用250ng含有U6启动子并编码gRNA的表达质粒,以及使用Opti-MEM培养基和Lipofectamine 2000的编码Cas9/ABE8变体碱基编辑器的750ng表达质粒转染。使用的ABE8碱基编辑器变体包含NGG PAM序列。将细胞在37℃和5% CO_2中维持5天,在转染后第3天更换药物。此后,细胞裂解;分离基因组DNA并使用标准程序进行PCR,通常使用20至100ng模板DNA。添加连接子(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。

[5637] 对来自从293T细胞重复转染中收获的基因组DNA或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 Kit(Zymo Research)。鸟枪法(shotgun)文库是在没有剪切的情况下准备的。所述文库通过qPCR进行量化,并在一个MiSeq Nano流通池上使用MiSeq 500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转换软件产生和

解复用Fastq档案(Illumina)。

[5638] A)实施例12:蕾特氏症突变的碱基编辑校正

[5639] 产生了一个含有单拷贝MECP2的慢病毒HEK系,其中包括6个蕾特(Rett)突变,包括R106W和R255X,并用于筛选向导RNA序列和ABE变体。将用于向导和ABE表达的质粒转染到慢病毒HEK系中,并且收集基因组DNA并通过NextGen测序分析目标位点的突变。参考图67和68,向导1显示出整体最高的编辑,其中ABE 8.8、ABE 8.9和ABE 8.13与向导1和向导2都能良好运作。向导RNA序列包括一个间隔CTTTTCACTTTTCCTGCCGGGG(R255X)、AGCTTCCATGTCCAGCCTTC(R106W)、ACCATGAAGTCAAAATCATT(T158M)、或GCTTTCAGCCCCGTTTCTTG(R270X)。

[5640] 测试了具有不同PAM识别特异性的碱基编辑器变体;对应于PAM改变的Cas9取代显示在表19中。*变体5对应于R255X的25%编辑。这些方法类似于上述R106W的方法,但据以其他组产生的ABE突变。参考图69和下表19,促成变体5的突变在该特定基因座处显示出最高的编辑量。

[5641] 表19

[5642]

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
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	

[5643] 实施例13:贺勒/IDUA突变校正

[5644] 产生了含有单拷贝MECP2和贺勒突变的慢病毒HEK系,并用于筛选如上所述向导RNA序列和变体。在图70A至70C中,使用碱基编辑器变体检查了遗传的IDUA功能丧失(W402X)突变的校正。来自两个W402X纯合勒氏症患者(GM06214和GM00798)以及未受影响的杂合亲本(GM00799)的成纤维细胞获自Coriell。使用ABE 8.8的mRNA和人类W402X向导对患者来源的和BJ成纤维细胞进行电穿孔。使用QuickExract裂解液提取基因组DNA,并使用NextGen测序进行测序以评估A至G编辑。通过分光亮度法测定艾杜糖基化酶活性。裂解物在酸性缓冲液中与4-甲基伞形酮-艾杜糖醛酸一起培育2小时,然后用碱性溶液淬灭。4-甲基伞形酮裂解通过荧光(365nm激发,445nm发射)测量。在患者成纤维细胞中观察到高度编辑,这导致酶活性增加,与未受影响的杂合子GM00799中观察到的酶活性相当。

[5645] 实施例14:使用ABE 8.8进行体内碱基编辑

[5646] 参考图70,编码分裂内含肽ABE8.8的病毒基因组和鼠ROSA26基因座的gRNA被包装到AAV9和PHP.eB衣壳中。向C57BL/6小鼠侧脑室注射9e11总vg AAV9(每条内含含臂分裂4.5e11)和7.5e11总vg PHP.eB(每条内含肽分裂3.75e11)。6周后,收获并解剖大脑和脊髓以进行基因组DNA提取和测序。AAV9转导在最接近侧脑室的结构海马体中最高,编辑率高达13%。

[5647] 实施例15:ABE8变体碱基编辑

[5648] 为了确定在ABCA4中恢复G1961E突变的最佳碱基编辑器,将四十种独特的ABE8变体与ABE7.10进行比较,以使用具有我们之前证明所述序列是所述目标位点上的最佳间隔长度的具有21个核苷酸间隔序列(图73)的sgRNA,在慢病毒敲入模型细胞系中进行疾病等位基因(图72A至72B)的A至G碱基转换。所有变体都在疾病等位基因和摆动碱基上提供可测量的A至G编辑。在摆动碱基处进行编辑会导致沉默突变并且无害。随后对六个表现最佳的变体进行了密码子优化,并将其整合到拆分AAV系统中以进行进一步验证。由于编码碱基编辑器、sgRNA和表达调控组件的DNA序列的大小超过了单个AAV颗粒的包装限制,可以将所需部分分成两个AAV颗粒并共同递送。在所述递送方法中,编码碱基编辑器的基因在两种病毒之间分裂,并且在共感染后使用分裂的内含肽重构全长蛋白质(图74)。缺少野生型TadA结构域(ABE8-m)的分裂ABE8变体被包装成成对的AAV2载体,其中一个还编码靶向野生型ABCA4目标位点的sgRNA的单个副本。在所述实验中,在ABCA4 G1961密码子的摆动碱基处评估了编辑作为野生型细胞中不存在的疾病等位基因的替代物。野生型ARPE-19细胞与双AAV共转导,并对感兴趣的21个核苷酸目标序列(图75A至75B)评估碱基编辑率。ABE变体7.9、7.10、8.5-m、8.8-m、8.9-m和8.18-m在转换替代位点8A方面同样有效,然而,变体8.8-m、8.9-m及8.18-m也催化了不良反应位置5C处的C至T转换。与亲本ABE7.10变体相比,去除了野生型TadA结构域的ABE7.10变体(ABE7.10-m)的活性明显降低了50%。这些结果表示变体ABE8.5-m是也缺乏野生型TadA结构域的目标位点上最有效的编辑器,这减少碱基编辑器的总尺寸594个碱基对的DNA或198个氨基酸残基。

[5649] 向导RNA序列靶向在序列GCTGTGTGTCGAAGTTCGCCCTGGAGAGGTG或GCTGTGTGTCGGAGTTCGCCCTGGAGAGGTG的ABCA4基因,其中PAM序列具有底线。向导RNA包括一个序列CACCUCUCCAGGGCGAACUUCGACACACAGC或CACCUCUCCAGGGCGAACUCCGACACACAGC。

[5650] 评估了使用靶向ABCA4 G1961E基因座的21个核苷酸间隔长度sgRNA在人类基因组内进行脱靶碱基编辑的潜力。通过计算扫描人类参考基因组(GRCh38)中所有与ABCA4

G1961E gRNA前间隔序列的不完美匹配,然后是与SpCas9 NGG PAM匹配的3'序列,以对基因组内潜在脱靶位点进行计算机模拟预测。评估了包括多达5个错配和单个RNA或DNA凸起的所有序列。潜在的脱靶位点优先用于实验评估,其是基于(a)错配数量少,以及(b)与编码外显子重叠(由GENCODE转录本注释确定)和癌症相关基因(如COSMIC癌症基因普查中报告的)。通过计算机分析没有发现基因组中与21个核苷酸间隔序列有三个或更少错配的预测脱靶。我们使用双AAV系统将ABE7.10和靶向ABCA4 G1961E疾病等位基因的21个核苷酸间隔sgRNA共同递送到野生型ARPE-19细胞中,并通过计算机预测的脱靶位点的28个靶向扩增子测序评估编辑。与未处理的细胞相比,在处理的细胞中没有预测的脱靶位点被显着碱基编辑(图76A至76B)并且在脱靶位点或在目标位点处没有发现显着的插入缺失(图77)。正如所预期,观察到的唯一显着编辑发生在处理细胞中的ABCA4 G1961摆动碱基处,因为靶向ABCA4 G1961E疾病等位基因的sgRNA仅包括与这些细胞中存在的野生型等位基因的单个错配碱基对。这些结果表示sgRNA不会在任何被评估的计算机预测脱靶位点促进脱靶DNA编辑。

[5651] 实施例20:灵长类动物视网膜实施例

[5652] 非人灵长类动物的眼睛在死后1至2小时收获并在死后4至8小时之间进行培养。使用6mm活检穿孔器从整个神经视网膜上进行穿孔。将感光面朝下的视网膜置于6孔组织培养板中的核孔膜顶部。将载体(10ul,1.26E+12vg/ml)移液到神经视网膜和膜之间以在视网膜组织下形成一个小泡。每3天更换一次培养基,并将组织培育0至22天。在不同时间点收集组织,并固定在10%中性缓冲福尔马林中并进行组织学处理。

[5653] '灵长类动物视网膜完整性'

[5654] 切片在4℃下用抗视紫红质、抗GFP及生物素化花生凝集素抗体免疫标记过夜。在PBS中洗涤后,将样本与二抗在室温下培育1小时。载玻片在PBS中洗涤并用含有DAPI的甘油基液体封固剂封固。

[5655] 在第0天(D)和第22天收获来自非人类灵长类动物的视网膜外植体。比较D0和D22未转导的视网膜外植体的组织学染色显示,当培养长达22天时,视网膜的细胞类型得以保留。在D22,与光感受器特异性GFP表达载体(Anc80L65.hGRK.eGFP)相比,暴露于Anc80L65.CMV.eGFP的视网膜培养物中的GFP表达定性更亮。用Anc80L65.hGRK1.eGFP转导视网膜外植体证明GFP仅存在于含光感受器的外核层(ONL),证实了hGRK1启动子的光感受器特异性活性。参见图78。

[5656] 'NHP中的Cas9表达'

[5657] 切片在4℃下用小鼠和兔单克隆Cas9抗体免疫标记过夜。在PBS中洗涤后,将样本与二抗在室温下培育1小时。载玻片在PBS中洗涤并用含有DAPI的甘油基液体封固剂封固。

[5658] 在非人类灵长类动物视网膜外植体上测试了编码优化分裂ABE(ABE7.10、8.5、8.9)碱基编辑器的双AAV2粒子。为了测试基于全长的编辑器的重建,在被表达每个碱基编辑器的AAV颗粒共感染-分裂内含肽一半(half)后,在不同时间点收集组织并对Cas9 N端和C端进行染色。我们早在第6天就观察到Cas9 N(绿色染色)和C(红色染色)末端的表达,并在感染后保持到第17天,这暗示碱基编辑器可能存在编辑活动窗口。这些结果表示双AAV分裂内含碱基编辑器在非人类灵长类动物视网膜外植体中表达Cas9。参见图79。

[5659] 其他实施方案

[5660] 从前面的描述中,很明显可以对这里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。这样的实施方案也在所附权利要求的范畴内。

[5661] 在本文中对变量的任何定义中的组件列表的叙述包括将所述变量定义为任何单个组件或所列组件的组合(或子组合)。此处对实施方案的叙述包括作为任何单个实施方案或与任何其他实施方案或其部分组合的实施方案。

[5662] 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,其程度就如同每个单独的出版物、专利或专利申请被具体地和单独地指示为以引用方式并入一样。除非另有说明,否则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。

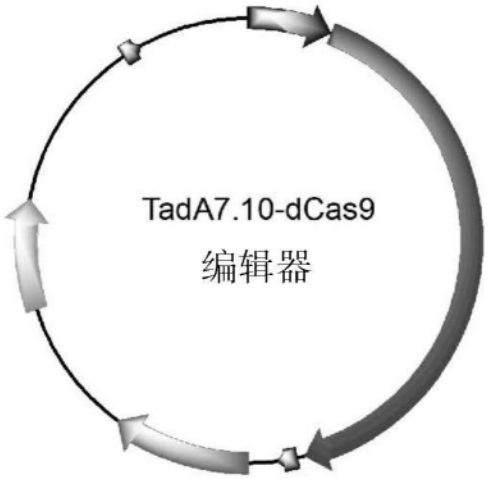


图1A

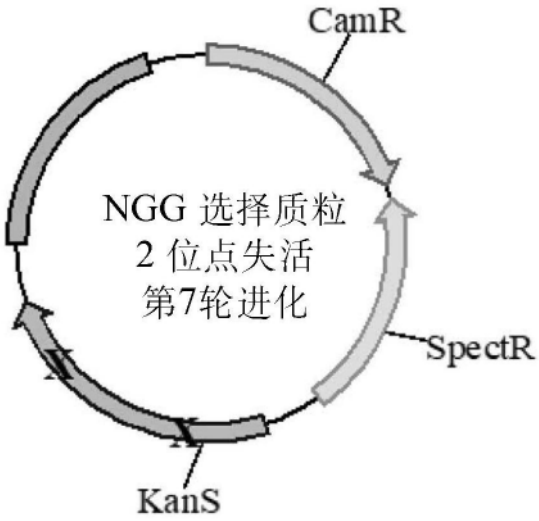


图1B

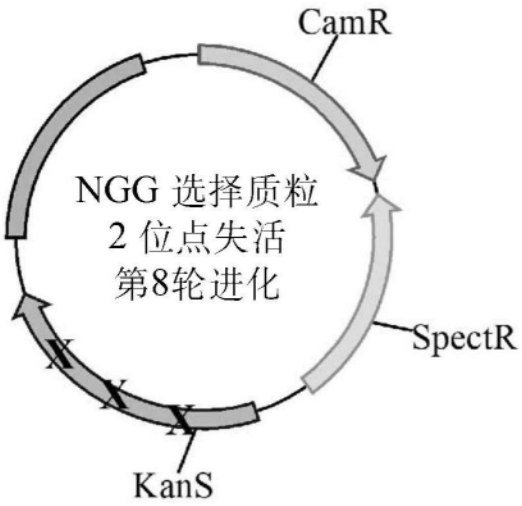

CamR
NGG 选择质粒
2 位点失活
第8轮进化
SpectR
KanS

图1C

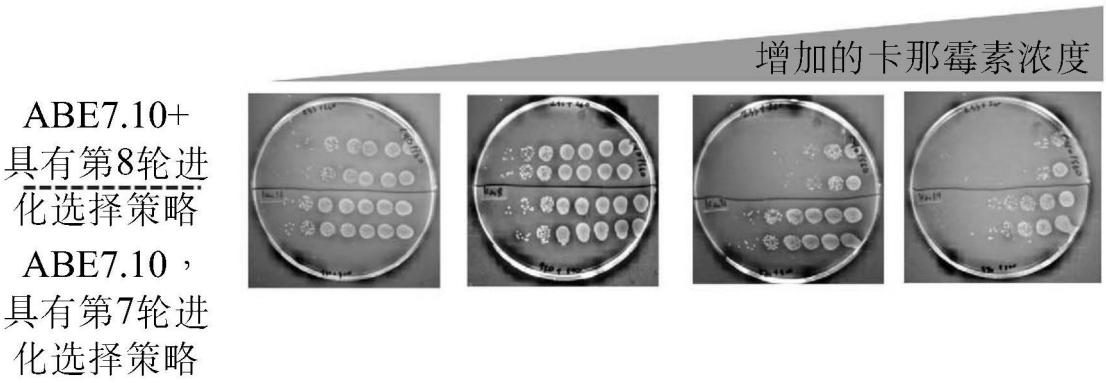

增加的卡那霉素浓度
ABE7.10+
具有第8轮进
化选择策略
ABE7.10，
具有第7轮进
化选择策略

图2

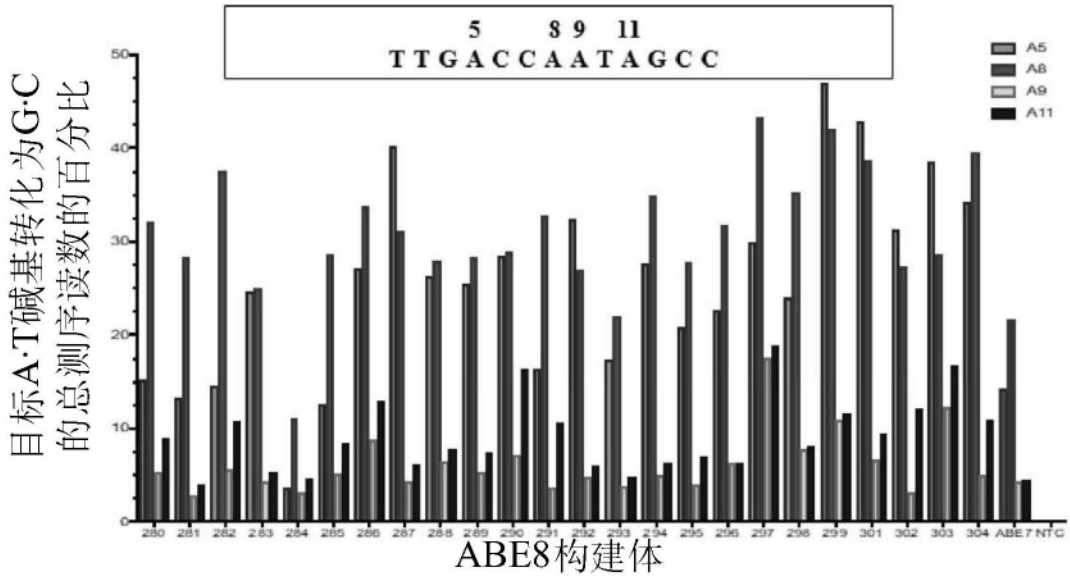

5 8 9 11
TTGACCAATAGCC
A5
A8
A9
A11
目标A·T碱基转化为G·C的总测序读数的百分比
ABE8构建体

图3A

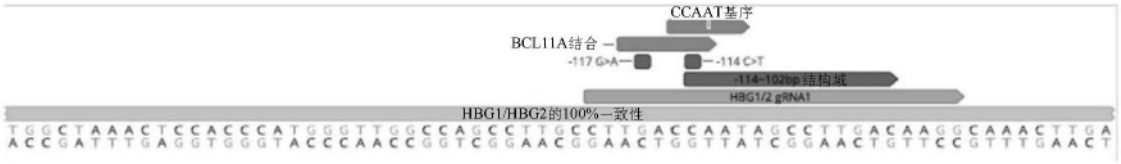

CCAAT基序
BCL11A结合
-117 G>A
-114 C>T
-114~102bp 结构域
HBG1/2 gRNA1
HBG1/HBG2的100%一致性

图3B

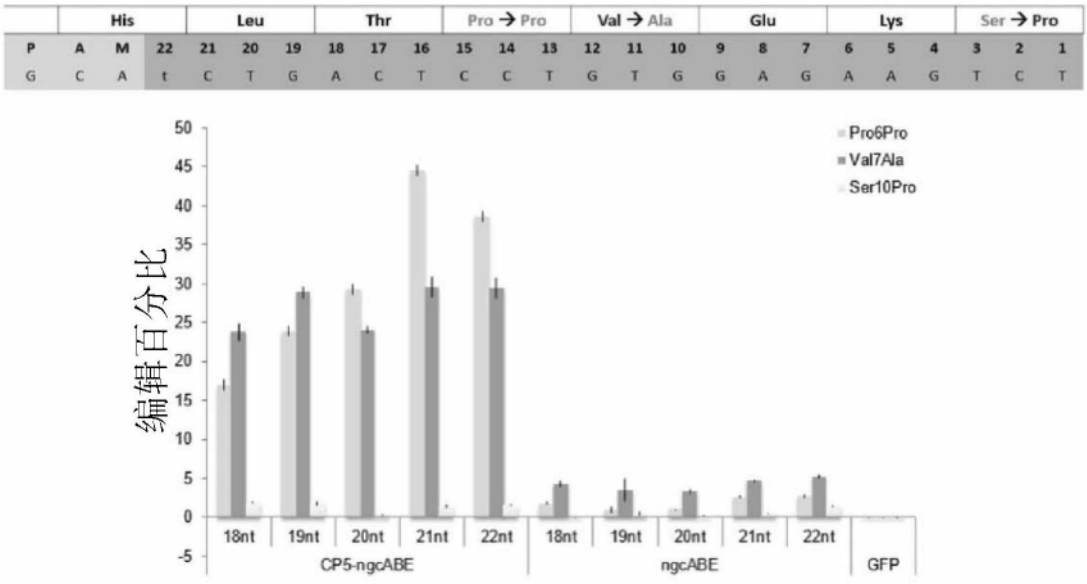

His
Leu
Thr
Pro → Pro
Val → Ala
Glu
Lys
Ser → Pro
Pro6Pro
Val7Ala
Ser10Pro
编辑百分比
18nt
19nt
20nt
21nt
22nt
CP5-ngcABE
ngcABE
GFP

图4

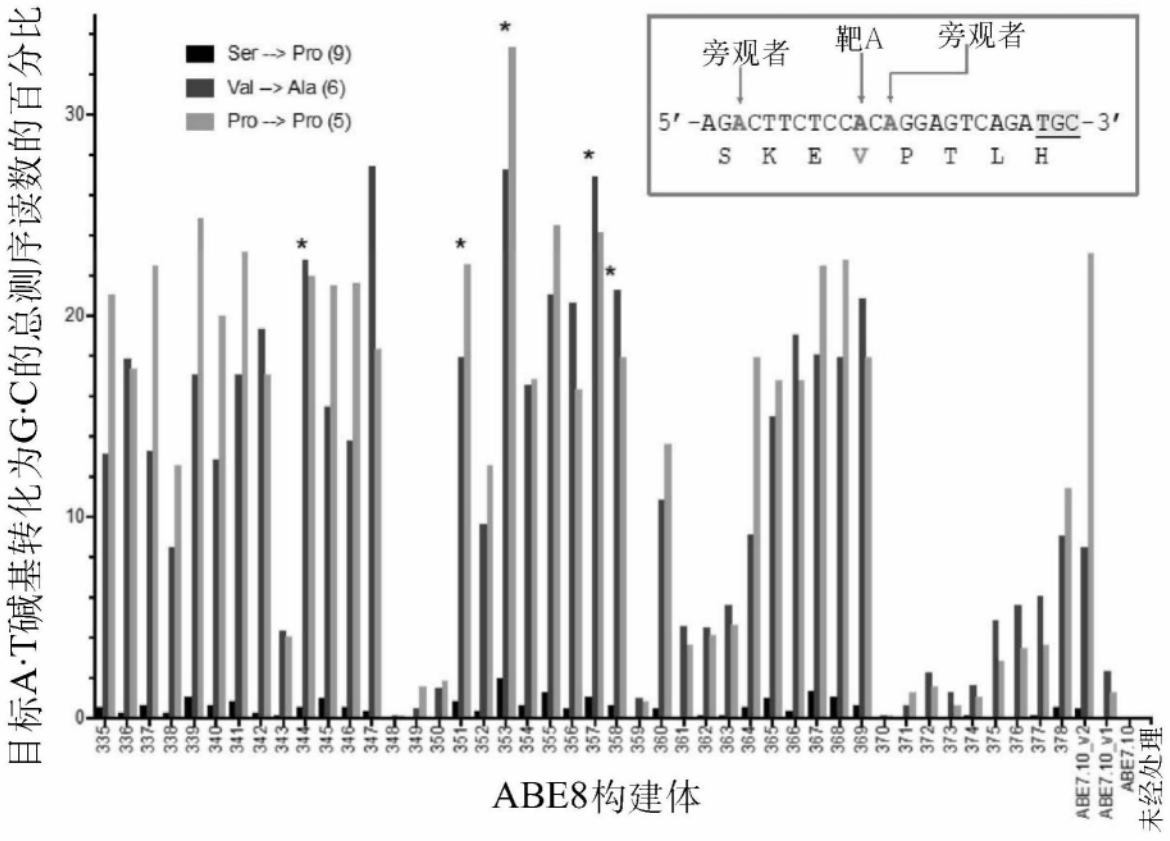

目标A·T碱基转化为G·C的总测序读数的百分比
Ser --> Pro (9)
Val --> Ala (6)
Pro --> Pro (5)
旁观者
靶A
旁观者
5′-AGACTTCTCCACAGGAGTCAGATGC-3′
S K E V P T L H
ABE8构建体
ABE7.10_v2
ABE7.10_v1
ABE7.10
未经处理

图5

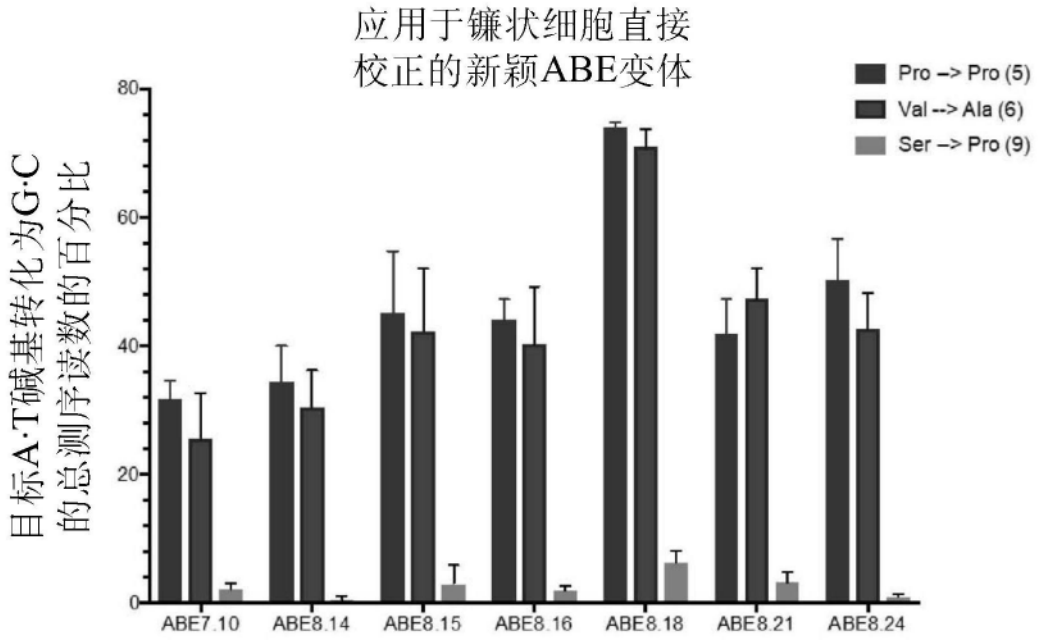

应用于镰状细胞直接
校正的新颖ABE变体
目标A·T碱基转化为G·C
的总测序读数的百分比
Pro --> Pro (5)
Val --> Ala (6)
Ser --> Pro (9)
ABE7.10
ABE8.14
ABE8.15
ABE8.16
ABE8.18
ABE8.21
ABE8.24

图6

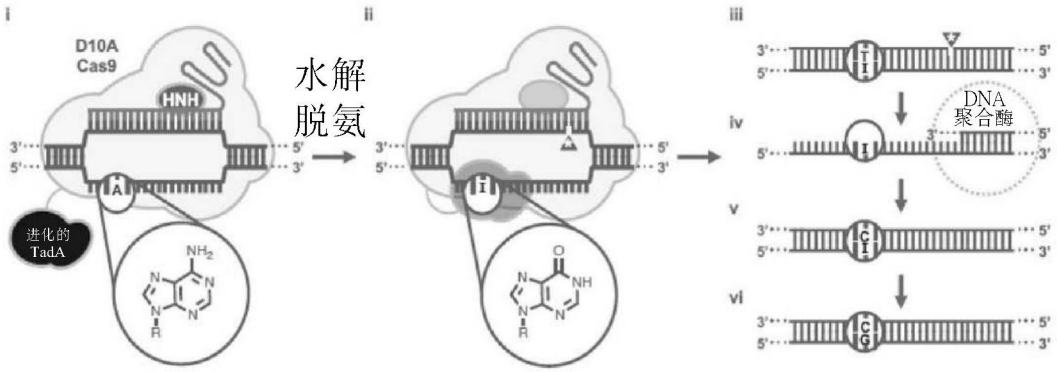

i
ii
iii
D10A
Cas9
HNH
水解
脱氨
进化的
TadA
DNA
聚合酶
iv
v
vi

图7A

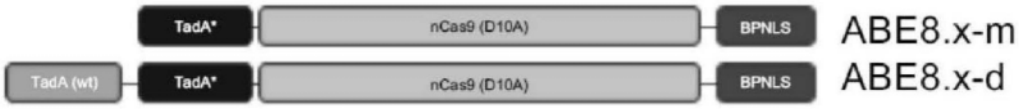

TadA*
nCas9 (D10A)
BPNLS
ABE8.x-m
TadA (wt)
TadA*
nCas9 (D10A)
BPNLS
ABE8.x-d

图7B

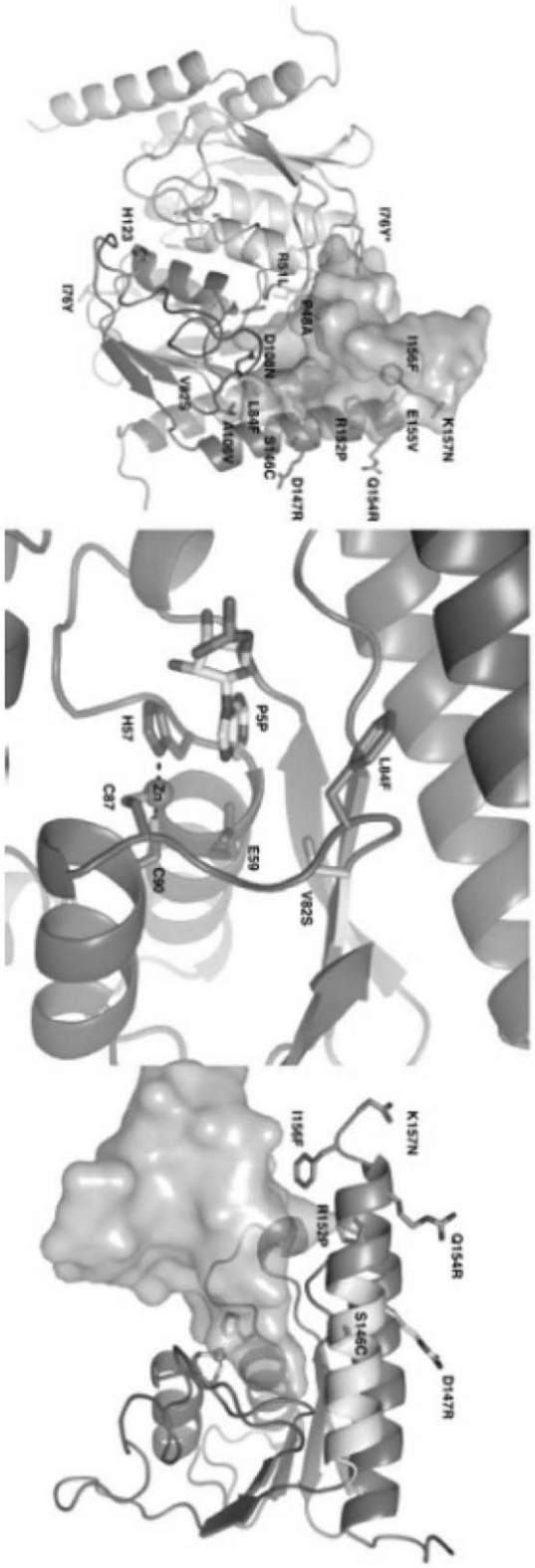

H123
I76Y*
R91L
I76Y
Q48A
D150N
V23S
L84F
A168V
S146C
D147R
R152P
E155V
I156F
K157N
Q154R
H57
P9P
Zn
C37
C90
E59
V82S
L84F
I156F
K157N
R152P
Q154R
S146C
D147R

图7C

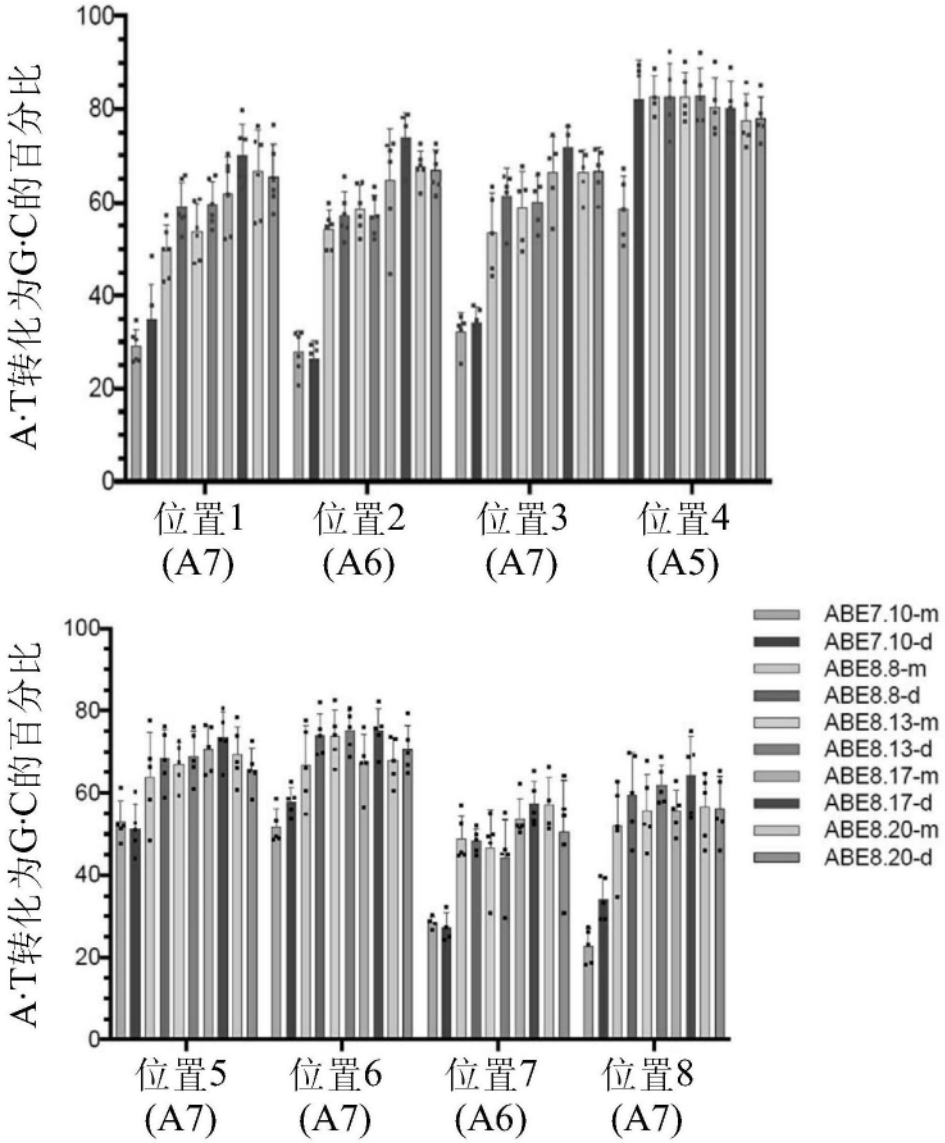

A·T转化为G·C的百分比
100
80
60
40
20
0
位置1
(A7)
位置2
(A6)
位置3
(A7)
位置4
(A5)
位置5
(A7)
位置6
(A7)
位置7
(A6)
位置8
(A7)
ABE7.10-m
ABE7.10-d
ABE8.8-m
ABE8.8-d
ABE8.13-m
ABE8.13-d
ABE8.17-m
ABE8.17-d
ABE8.20-m
ABE8.20-d

图7D

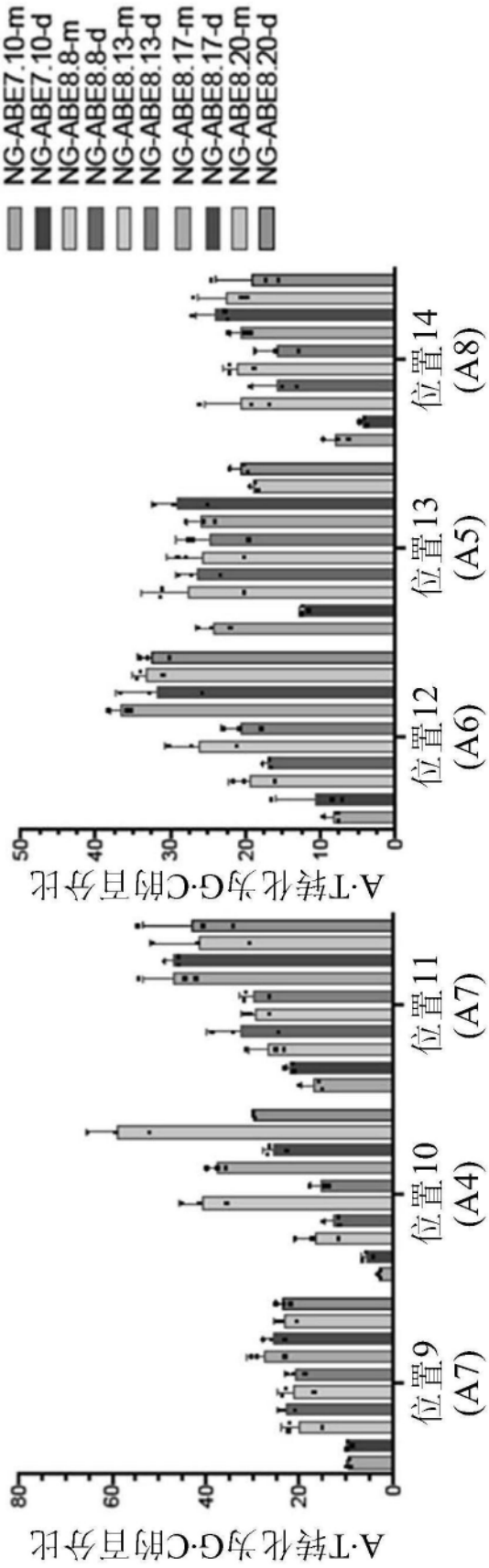

NG-ABE7.10-m
NG-ABE7.10-d
NG-ABE8.8-m
NG-ABE8.8-d
NG-ABE8.13-m
NG-ABE8.13-d
NG-ABE8.17-m
NG-ABE8.17-d
NG-ABE8.20-m
NG-ABE8.20-d
A·T转化为G·C的百分比
0
20
40
60
80
位置9
(A7)
位置10
(A4)
位置11
(A7)
A·T转化为G·C的百分比
0
10
20
30
40
50
位置12
(A6)
位置13
(A5)
位置14
(A8)

图8A

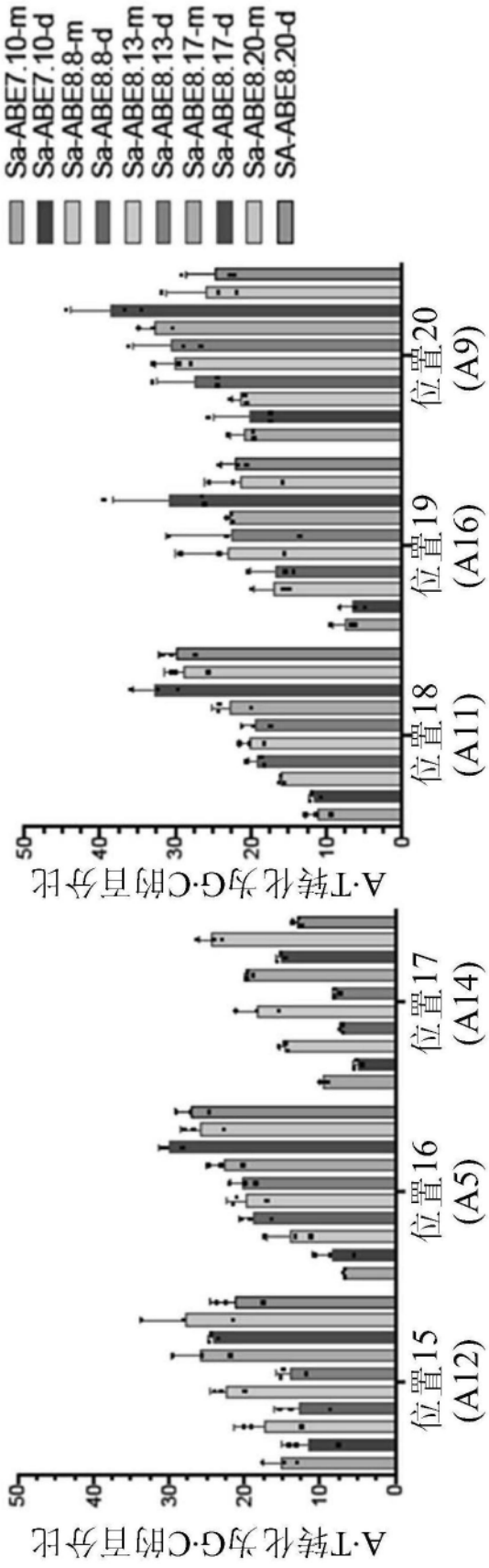

Sa-ABE7.10-m
Sa-ABE7.10-d
Sa-ABE8.8-m
Sa-ABE8.8-d
Sa-ABE8.13-m
Sa-ABE8.13-d
Sa-ABE8.17-m
Sa-ABE8.17-d
Sa-ABE8.20-m
SA-ABE8.20-d
A·T转化为G·C的百分比
0
10
20
30
40
50
位置15 (A12)
位置16 (A5)
位置17 (A14)
位置18 (A11)
位置19 (A16)
位置20 (A9)

图8B

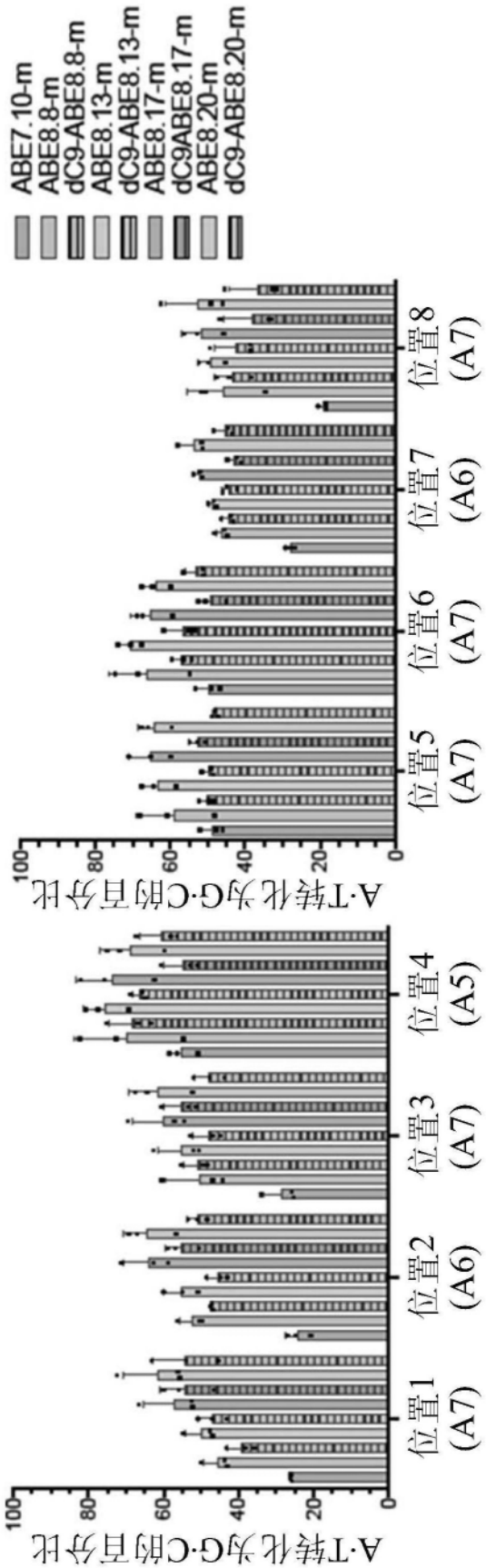

ABE7.10-m
ABE8.8-m
dC9-ABE8.8-m
ABE8.13-m
dC9-ABE8.13-m
ABE8.17-m
dC9ABE8.17-m
ABE8.20-m
dC9-ABE8.20-m
A·T转化为G·C的百分比
0
20
40
60
80
100
位置1 (A7)
位置2 (A6)
位置3 (A7)
位置4 (A5)
位置5 (A7)
位置6 (A7)
位置7 (A6)
位置8 (A7)

图8C

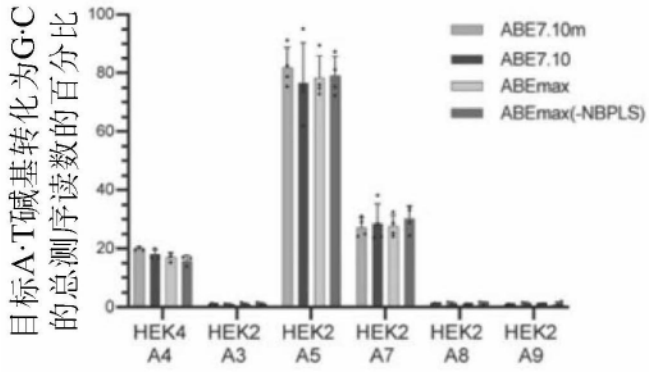

目标A·T碱基转化为G·C的总测序读数的百分比
100
80
60
40
20
0
ABE7.10m
ABE7.10
ABEmax
ABEmax(-NBPLS)
HEK4 A4
HEK2 A3
HEK2 A5
HEK2 A7
HEK2 A8
HEK2 A9

图9A

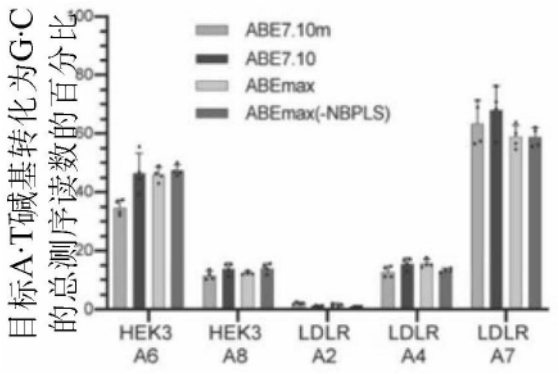

目标A·T碱基转化为G·C的总测序读数的百分比
ABE7.10m
ABE7.10
ABEmax
ABEmax(-NBPLS)
HEK3 A6
HEK3 A8
LDLR A2
LDLR A4
LDLR A7

图9B

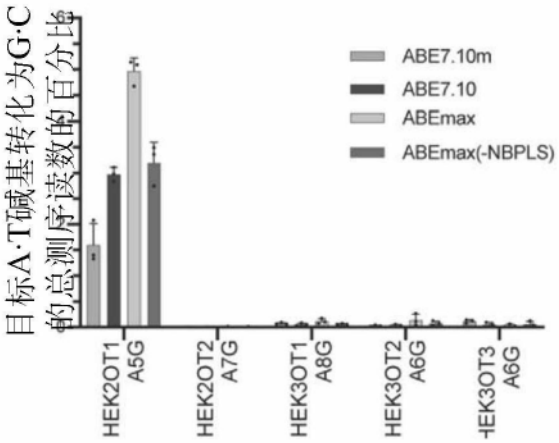

目标A·T碱基转化为G·C的总测序读数的百分比
ABE7.10m
ABE7.10
ABEmax
ABEmax(-NBPLS)
HEK2OT1 A5G
HEK2OT2 A7G
HEK3OT1 A8G
HEK3OT2 A6G
HEK3OT3 A6G

图9C

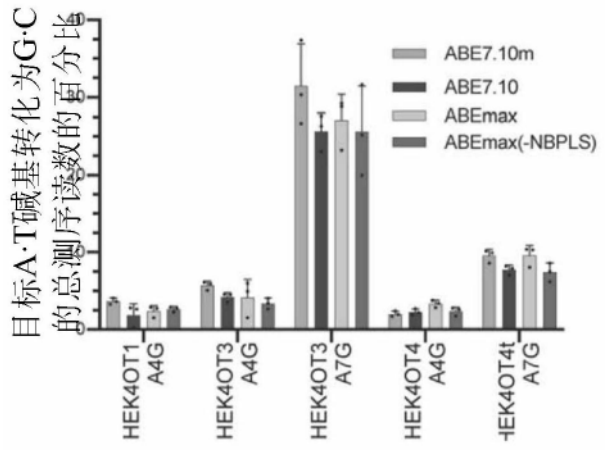

目标A·T碱基转化为G·C的总测序读数的百分比
ABE7.10m
ABE7.10
ABEmax
ABEmax(-NBPLS)
HEK4OT1 A4G
HEK4OT3 A4G
HEK4OT3 A7G
HEK4OT4 A4G
HEK4OT4t A7G

图9D

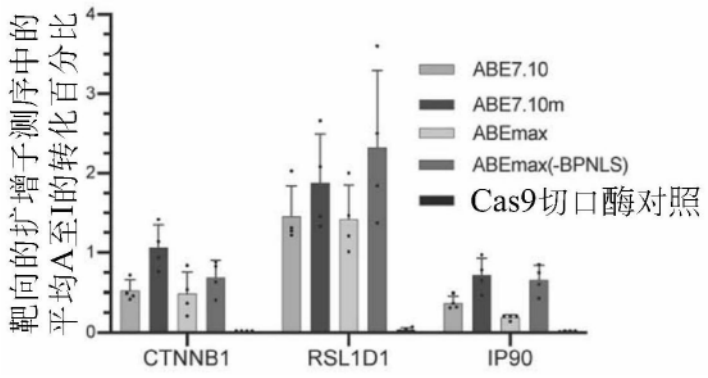

靶向的扩增子测序中的平均A至I的转化百分比
0
1
2
3
4
ABE7.10
ABE7.10m
ABEmax
ABEmax(-BPNLS)
Cas9切口酶对照
CTNNB1
RSL1D1
IP90

图9E

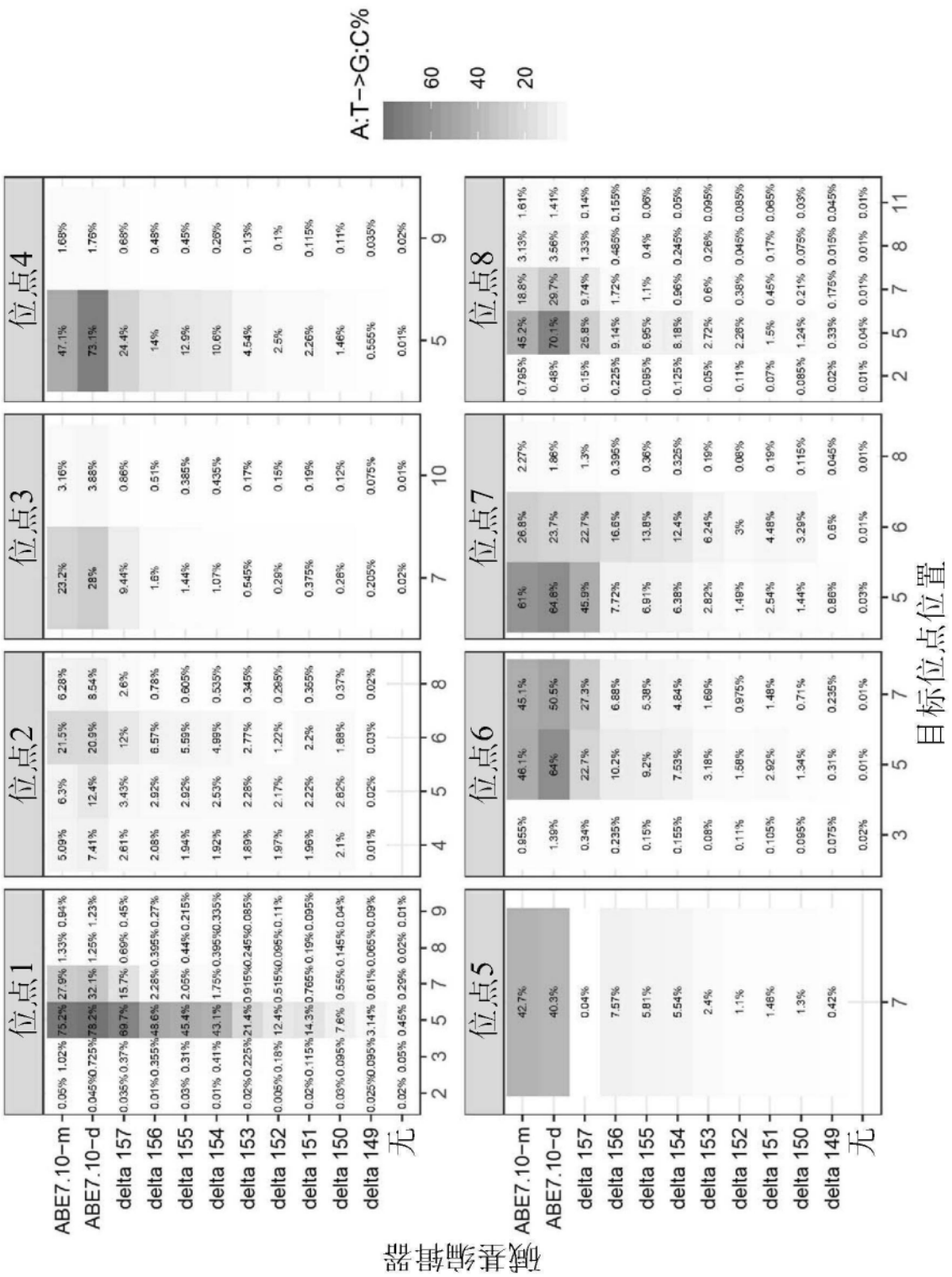

A:T->G:C%
60
40
20
位点1
位点2
位点3
位点4
位点5
位点6
位点7
位点8
碱基编辑器
ABE7.10-m
ABE7.10-d
delta 157
delta 156
delta 155
delta 154
delta 153
delta 152
delta 151
delta 150
delta 149
无
目标位点位置

图10A

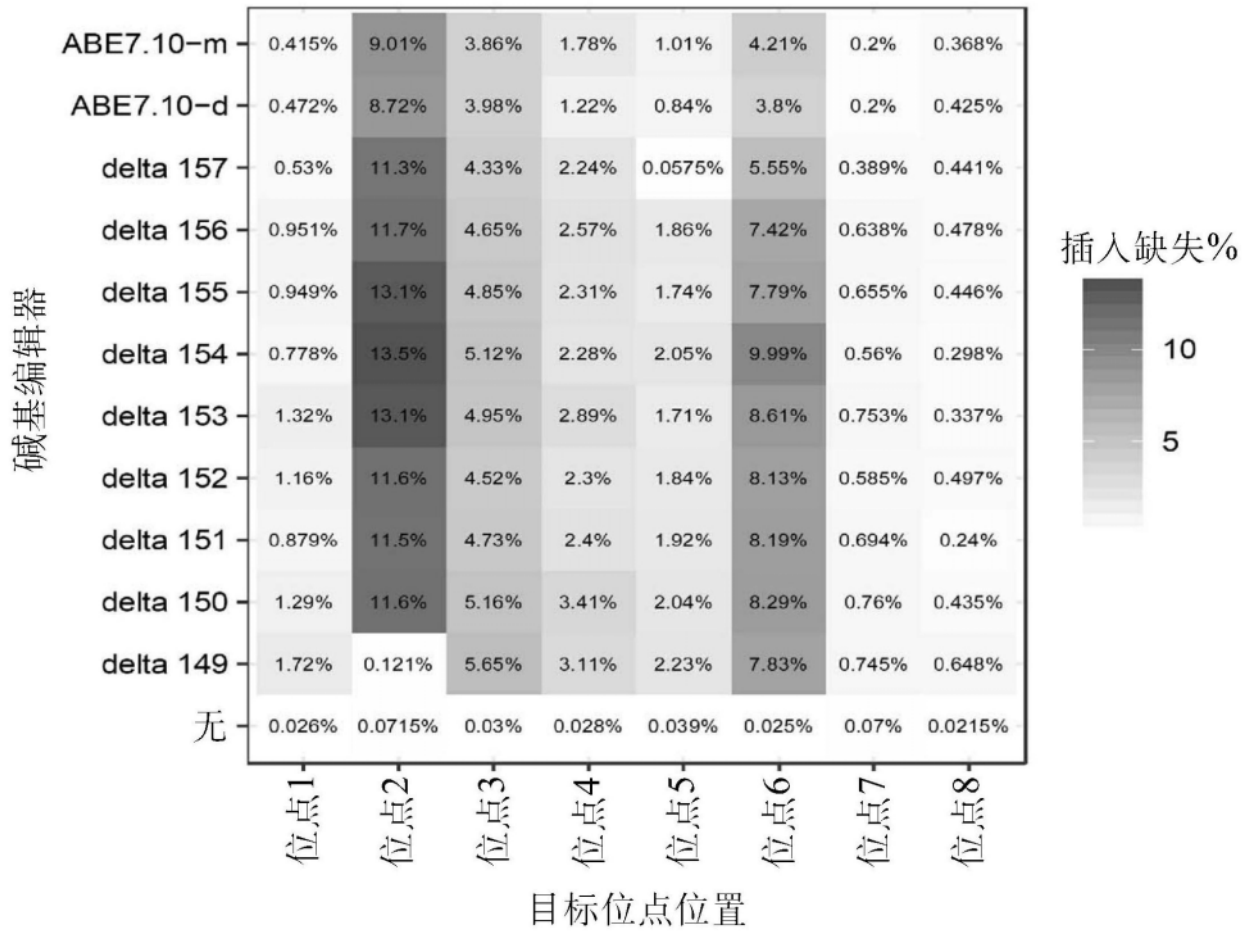

碱基编辑器
ABE7.10-m
ABE7.10-d
delta 157
delta 156
delta 155
delta 154
delta 153
delta 152
delta 151
delta 150
delta 149
无
0.415% 9.01% 3.86% 1.78% 1.01% 4.21% 0.2% 0.368%
0.472% 8.72% 3.98% 1.22% 0.84% 3.8% 0.2% 0.425%
0.53% 11.3% 4.33% 2.24% 0.0575% 5.55% 0.389% 0.441%
0.951% 11.7% 4.65% 2.57% 1.86% 7.42% 0.638% 0.478%
0.949% 13.1% 4.85% 2.31% 1.74% 7.79% 0.655% 0.446%
0.778% 13.5% 5.12% 2.28% 2.05% 9.99% 0.56% 0.298%
1.32% 13.1% 4.95% 2.89% 1.71% 8.61% 0.753% 0.337%
1.16% 11.6% 4.52% 2.3% 1.84% 8.13% 0.585% 0.497%
0.879% 11.5% 4.73% 2.4% 1.92% 8.19% 0.694% 0.24%
1.29% 11.6% 5.16% 3.41% 2.04% 8.29% 0.76% 0.435%
1.72% 0.121% 5.65% 3.11% 2.23% 7.83% 0.745% 0.648%
0.026% 0.0715% 0.03% 0.028% 0.039% 0.025% 0.07% 0.0215%
位点1
位点2
位点3
位点4
位点5
位点6
位点7
位点8
目标位点位置
插入缺失%
10
5

图10B

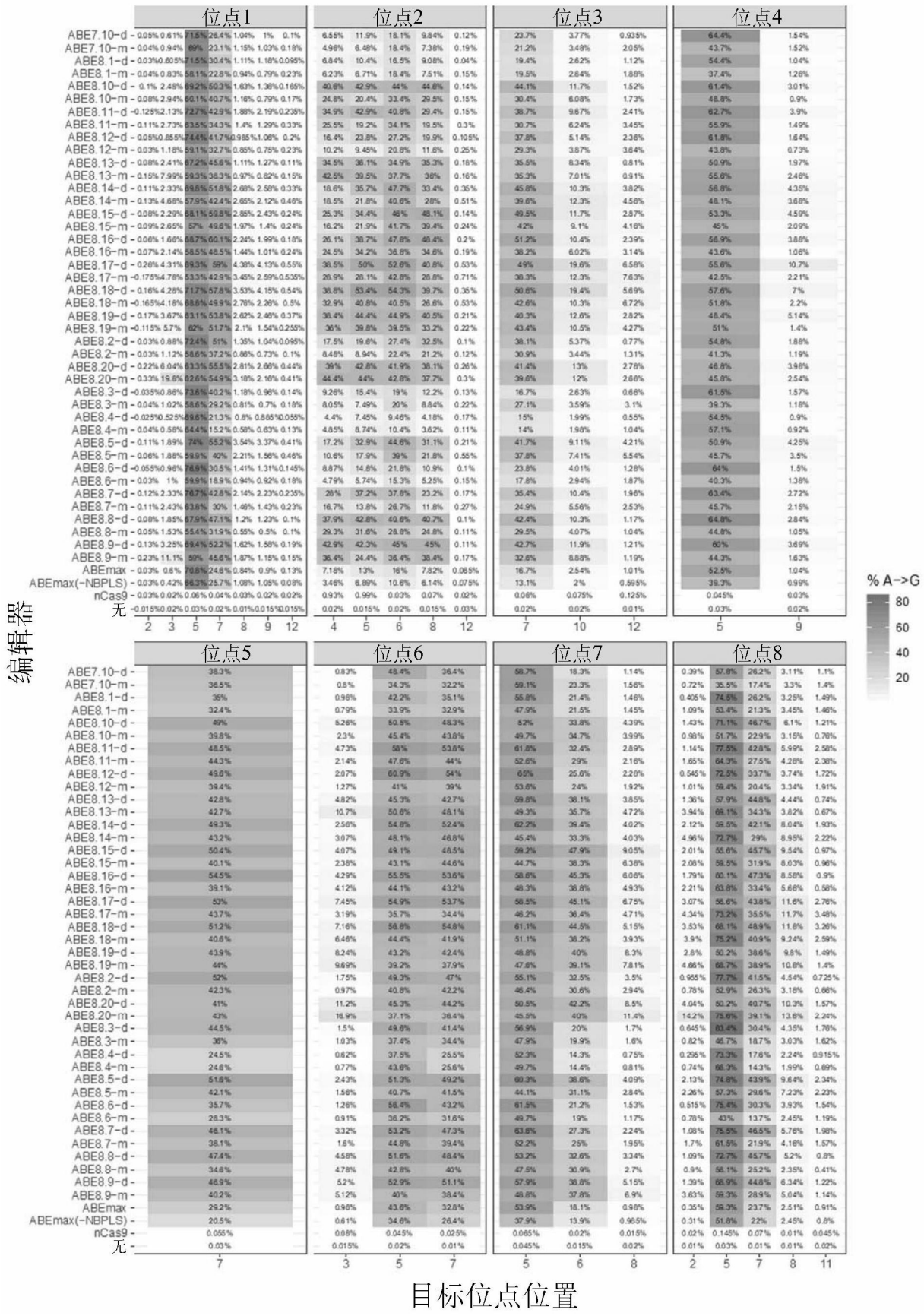

位点1
位点2
位点3
位点4
位点5
位点6
位点7
位点8
编辑器
ABE7.10-d
ABE7.10-m
ABE8.1-d
ABE8.1-m
ABE8.10-d
ABE8.10-m
ABE8.11-d
ABE8.11-m
ABE8.12-d
ABE8.12-m
ABE8.13-d
ABE8.13-m
ABE8.14-d
ABE8.14-m
ABE8.15-d
ABE8.15-m
ABE8.16-d
ABE8.16-m
ABE8.17-d
ABE8.17-m
ABE8.18-d
ABE8.18-m
ABE8.19-d
ABE8.19-m
ABE8.2-d
ABE8.2-m
ABE8.20-d
ABE8.20-m
ABE8.3-d
ABE8.3-m
ABE8.4-d
ABE8.4-m
ABE8.5-d
ABE8.5-m
ABE8.6-d
ABE8.6-m
ABE8.7-d
ABE8.7-m
ABE8.8-d
ABE8.8-m
ABE8.9-d
ABE8.9-m
ABEmax
ABEmax(-NBPLS)
nCas9
无
% A->G
80
60
40
20
目标位点位置

图11

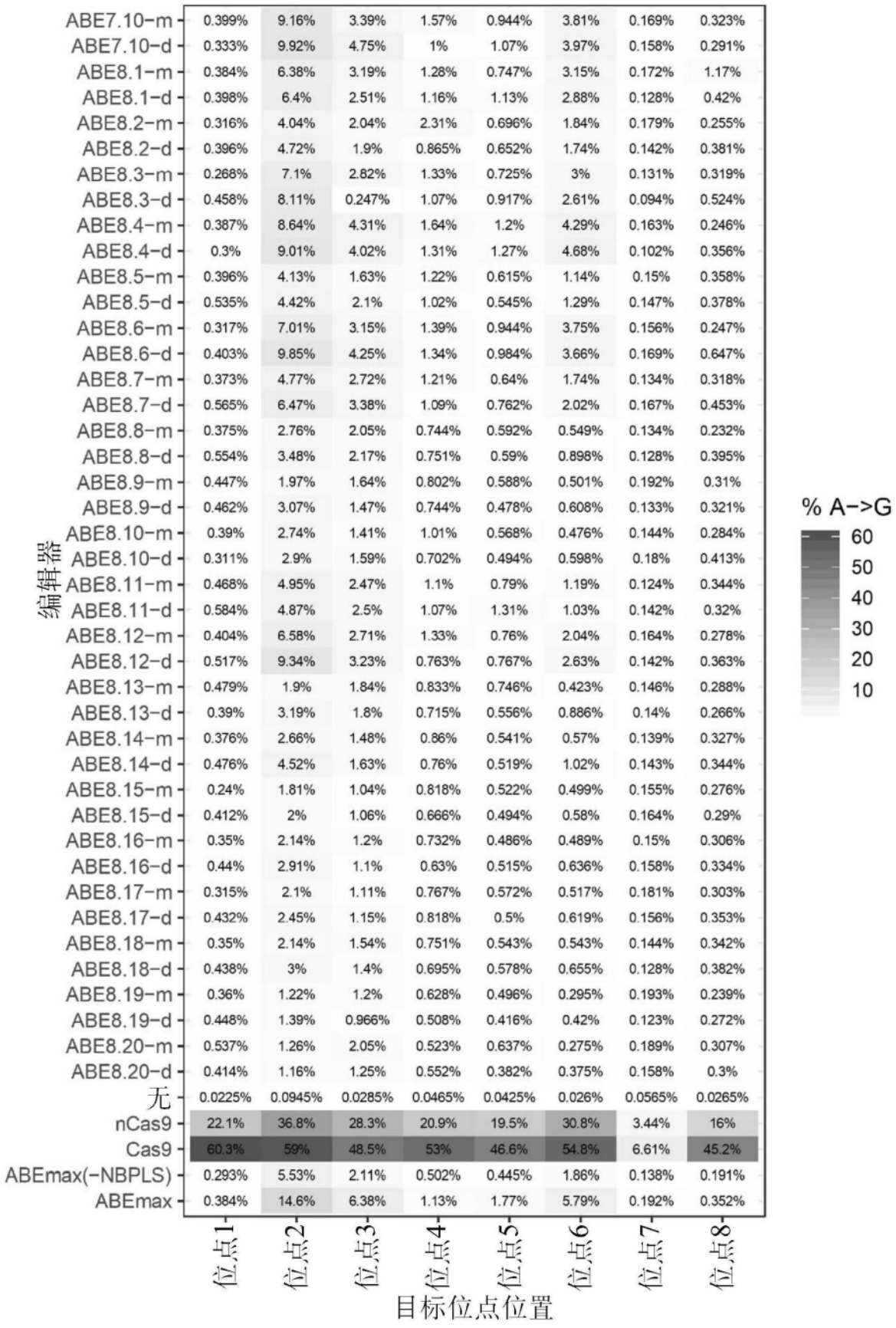

编辑器
ABE7.10-m 0.399% 9.16% 3.39% 1.57% 0.944% 3.81% 0.169% 0.323%
ABE7.10-d 0.333% 9.92% 4.75% 1% 1.07% 3.97% 0.158% 0.291%
ABE8.1-m 0.384% 6.38% 3.19% 1.28% 0.747% 3.15% 0.172% 1.17%
ABE8.1-d 0.398% 6.4% 2.51% 1.16% 1.13% 2.88% 0.128% 0.42%
ABE8.2-m 0.316% 4.04% 2.04% 2.31% 0.696% 1.84% 0.179% 0.255%
ABE8.2-d 0.396% 4.72% 1.9% 0.865% 0.652% 1.74% 0.142% 0.381%
ABE8.3-m 0.268% 7.1% 2.82% 1.33% 0.725% 3% 0.131% 0.319%
ABE8.3-d 0.458% 8.11% 0.247% 1.07% 0.917% 2.61% 0.094% 0.524%
ABE8.4-m 0.387% 8.64% 4.31% 1.64% 1.2% 4.29% 0.163% 0.246%
ABE8.4-d 0.3% 9.01% 4.02% 1.31% 1.27% 4.68% 0.102% 0.356%
ABE8.5-m 0.396% 4.13% 1.63% 1.22% 0.615% 1.14% 0.15% 0.358%
ABE8.5-d 0.535% 4.42% 2.1% 1.02% 0.545% 1.29% 0.147% 0.378%
ABE8.6-m 0.317% 7.01% 3.15% 1.39% 0.944% 3.75% 0.156% 0.247%
ABE8.6-d 0.403% 9.85% 4.25% 1.34% 0.984% 3.66% 0.169% 0.647%
ABE8.7-m 0.373% 4.77% 2.72% 1.21% 0.64% 1.74% 0.134% 0.318%
ABE8.7-d 0.565% 6.47% 3.38% 1.09% 0.762% 2.02% 0.167% 0.453%
ABE8.8-m 0.375% 2.76% 2.05% 0.744% 0.592% 0.549% 0.134% 0.232%
ABE8.8-d 0.554% 3.48% 2.17% 0.751% 0.59% 0.898% 0.128% 0.395%
ABE8.9-m 0.447% 1.97% 1.64% 0.802% 0.588% 0.501% 0.192% 0.31%
ABE8.9-d 0.462% 3.07% 1.47% 0.744% 0.478% 0.608% 0.133% 0.321%
ABE8.10-m 0.39% 2.74% 1.41% 1.01% 0.568% 0.476% 0.144% 0.284%
ABE8.10-d 0.311% 2.9% 1.59% 0.702% 0.494% 0.598% 0.18% 0.413%
ABE8.11-m 0.468% 4.95% 2.47% 1.1% 0.79% 1.19% 0.124% 0.344%
ABE8.11-d 0.584% 4.87% 2.5% 1.07% 1.31% 1.03% 0.142% 0.32%
ABE8.12-m 0.404% 6.58% 2.71% 1.33% 0.76% 2.04% 0.164% 0.278%
ABE8.12-d 0.517% 9.34% 3.23% 0.763% 0.767% 2.63% 0.142% 0.363%
ABE8.13-m 0.479% 1.9% 1.84% 0.833% 0.746% 0.423% 0.146% 0.288%
ABE8.13-d 0.39% 3.19% 1.8% 0.715% 0.556% 0.886% 0.14% 0.266%
ABE8.14-m 0.376% 2.66% 1.48% 0.86% 0.541% 0.57% 0.139% 0.327%
ABE8.14-d 0.476% 4.52% 1.63% 0.76% 0.519% 1.02% 0.143% 0.344%
ABE8.15-m 0.24% 1.81% 1.04% 0.818% 0.522% 0.499% 0.155% 0.276%
ABE8.15-d 0.412% 2% 1.06% 0.666% 0.494% 0.58% 0.164% 0.29%
ABE8.16-m 0.35% 2.14% 1.2% 0.732% 0.486% 0.489% 0.15% 0.306%
ABE8.16-d 0.44% 2.91% 1.1% 0.63% 0.515% 0.636% 0.158% 0.334%
ABE8.17-m 0.315% 2.1% 1.11% 0.767% 0.572% 0.517% 0.181% 0.303%
ABE8.17-d 0.432% 2.45% 1.15% 0.818% 0.5% 0.619% 0.156% 0.353%
ABE8.18-m 0.35% 2.14% 1.54% 0.751% 0.543% 0.543% 0.144% 0.342%
ABE8.18-d 0.438% 3% 1.4% 0.695% 0.578% 0.655% 0.128% 0.382%
ABE8.19-m 0.36% 1.22% 1.2% 0.628% 0.496% 0.295% 0.193% 0.239%
ABE8.19-d 0.448% 1.39% 0.966% 0.508% 0.416% 0.42% 0.123% 0.272%
ABE8.20-m 0.537% 1.26% 2.05% 0.523% 0.637% 0.275% 0.189% 0.307%
ABE8.20-d 0.414% 1.16% 1.25% 0.552% 0.382% 0.375% 0.158% 0.3%
无 0.0225% 0.0945% 0.0285% 0.0465% 0.0425% 0.026% 0.0565% 0.0265%
nCas9 22.1% 36.8% 28.3% 20.9% 19.5% 30.8% 3.44% 16%
Cas9 60.3% 59% 48.5% 53% 46.6% 54.8% 6.61% 45.2%
ABEmax(-NBPLS) 0.293% 5.53% 2.11% 0.502% 0.445% 1.86% 0.138% 0.191%
ABEmax 0.384% 14.6% 6.38% 1.13% 1.77% 5.79% 0.192% 0.352%
位点1
位点2
位点3
位点4
位点5
位点6
位点7
位点8
目标位点位置
% A->G
60
50
40
30
20
10

图12

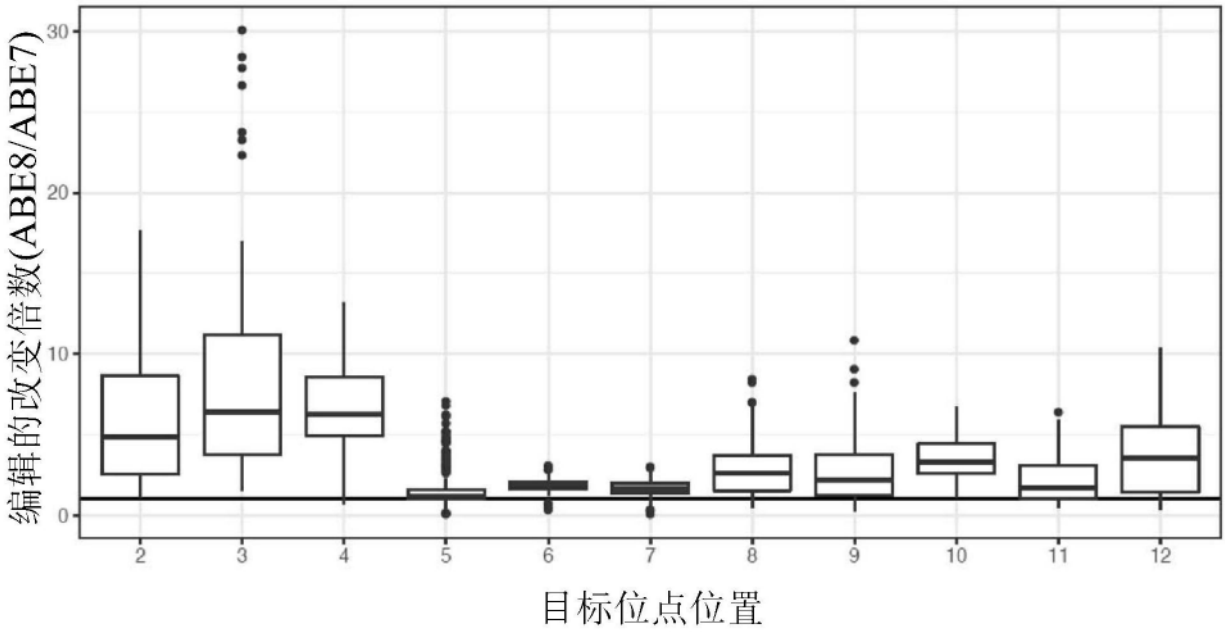


图13

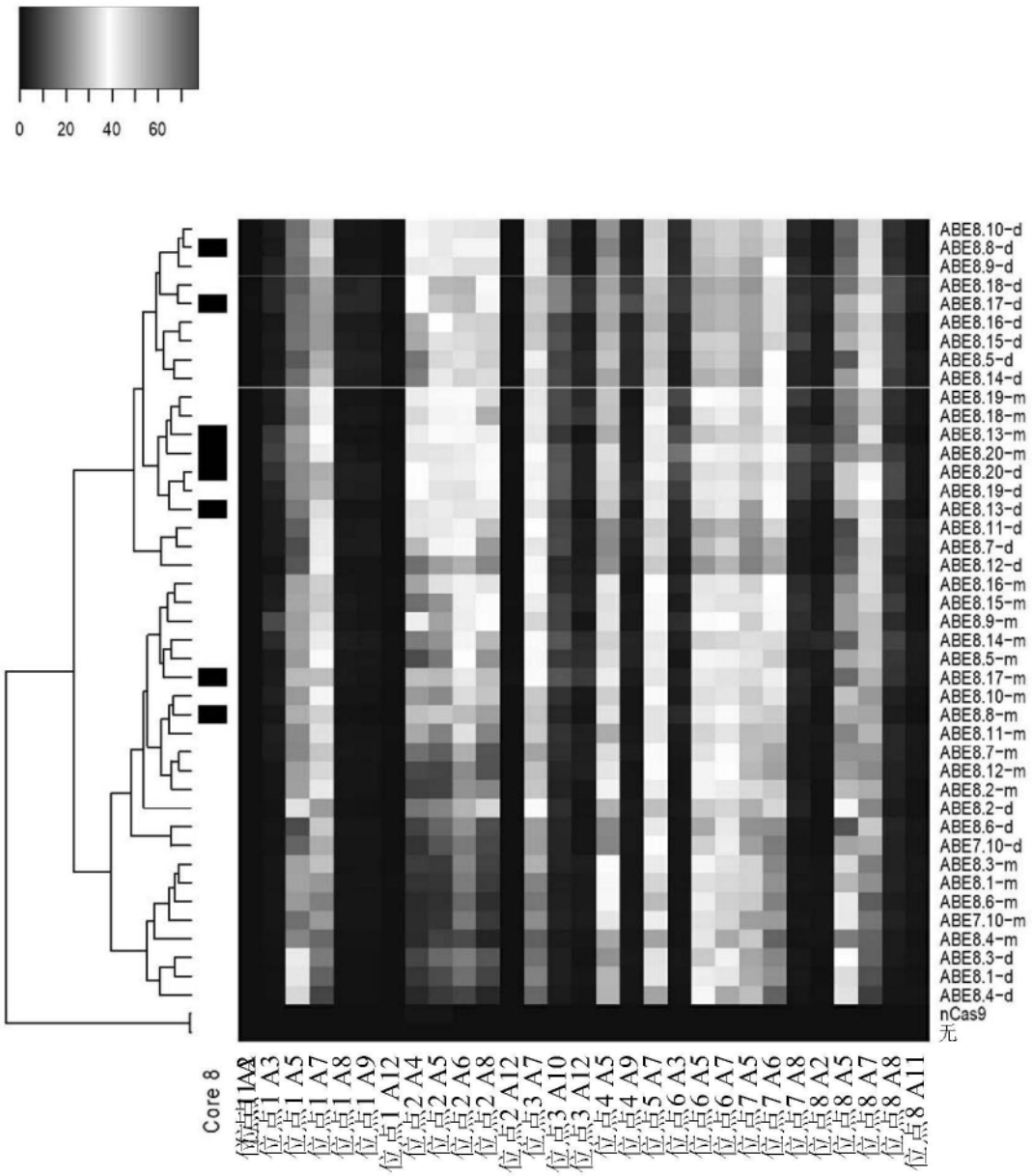

0
20
40
60
ABE8.10-d
ABE8.8-d
ABE8.9-d
ABE8.18-d
ABE8.17-d
ABE8.16-d
ABE8.15-d
ABE8.5-d
ABE8.14-d
ABE8.19-m
ABE8.18-m
ABE8.13-m
ABE8.20-m
ABE8.20-d
ABE8.19-d
ABE8.13-d
ABE8.11-d
ABE8.7-d
ABE8.12-d
ABE8.16-m
ABE8.15-m
ABE8.9-m
ABE8.14-m
ABE8.5-m
ABE8.17-m
ABE8.10-m
ABE8.8-m
ABE8.11-m
ABE8.7-m
ABE8.12-m
ABE8.2-m
ABE8.2-d
ABE8.6-d
ABE7.10-d
ABE8.3-m
ABE8.1-m
ABE8.6-m
ABE7.10-m
ABE8.4-m
ABE8.3-d
ABE8.1-d
ABE8.4-d
nCas9
无
Core 8

图14

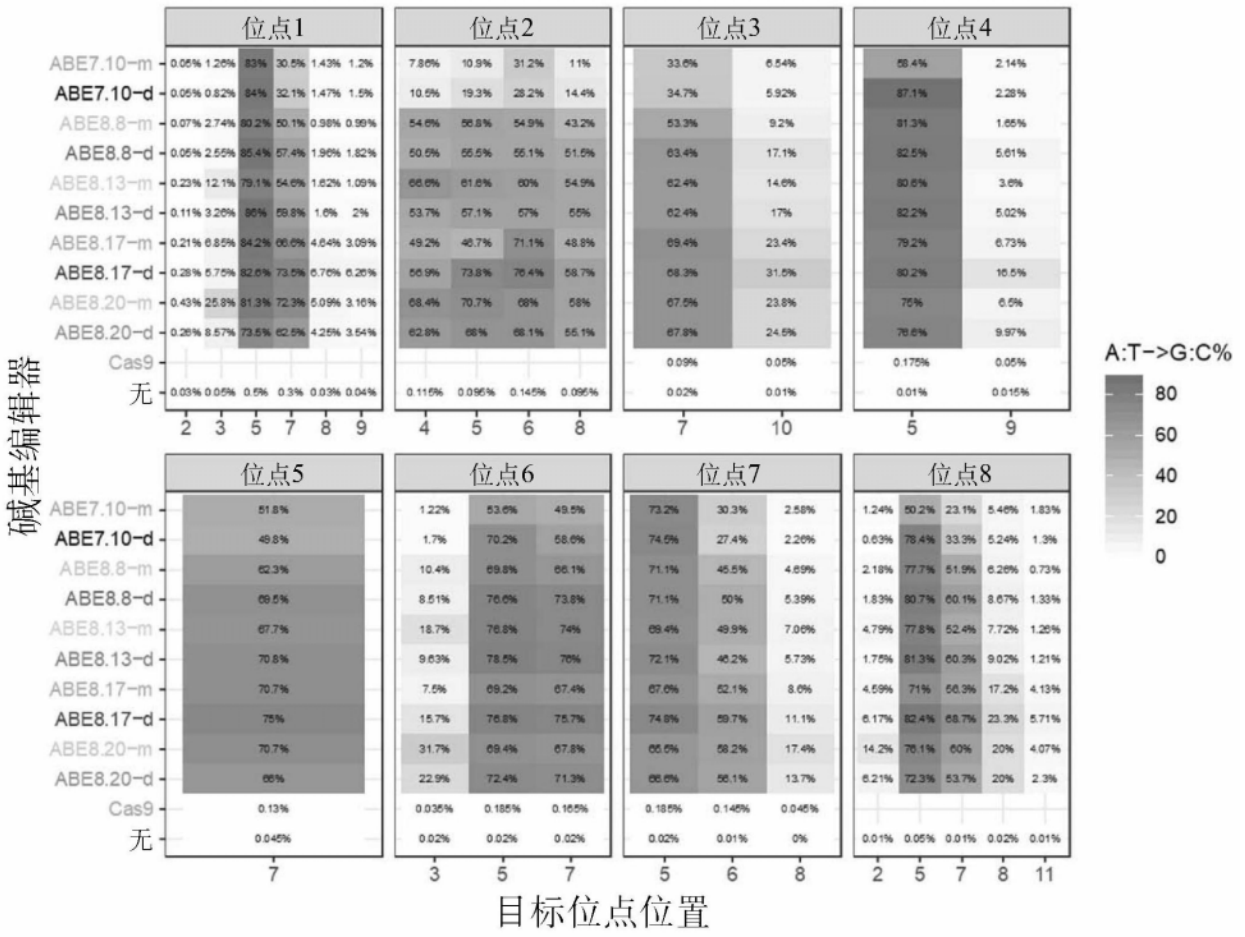

位点1
位点2
位点3
位点4
位点5
位点6
位点7
位点8
碱基编辑器
ABE7.10-m
ABE7.10-d
ABE8.8-m
ABE8.8-d
ABE8.13-m
ABE8.13-d
ABE8.17-m
ABE8.17-d
ABE8.20-m
ABE8.20-d
Cas9
无
A:T->G:C%
80
60
40
20
0
目标位点位置

图15

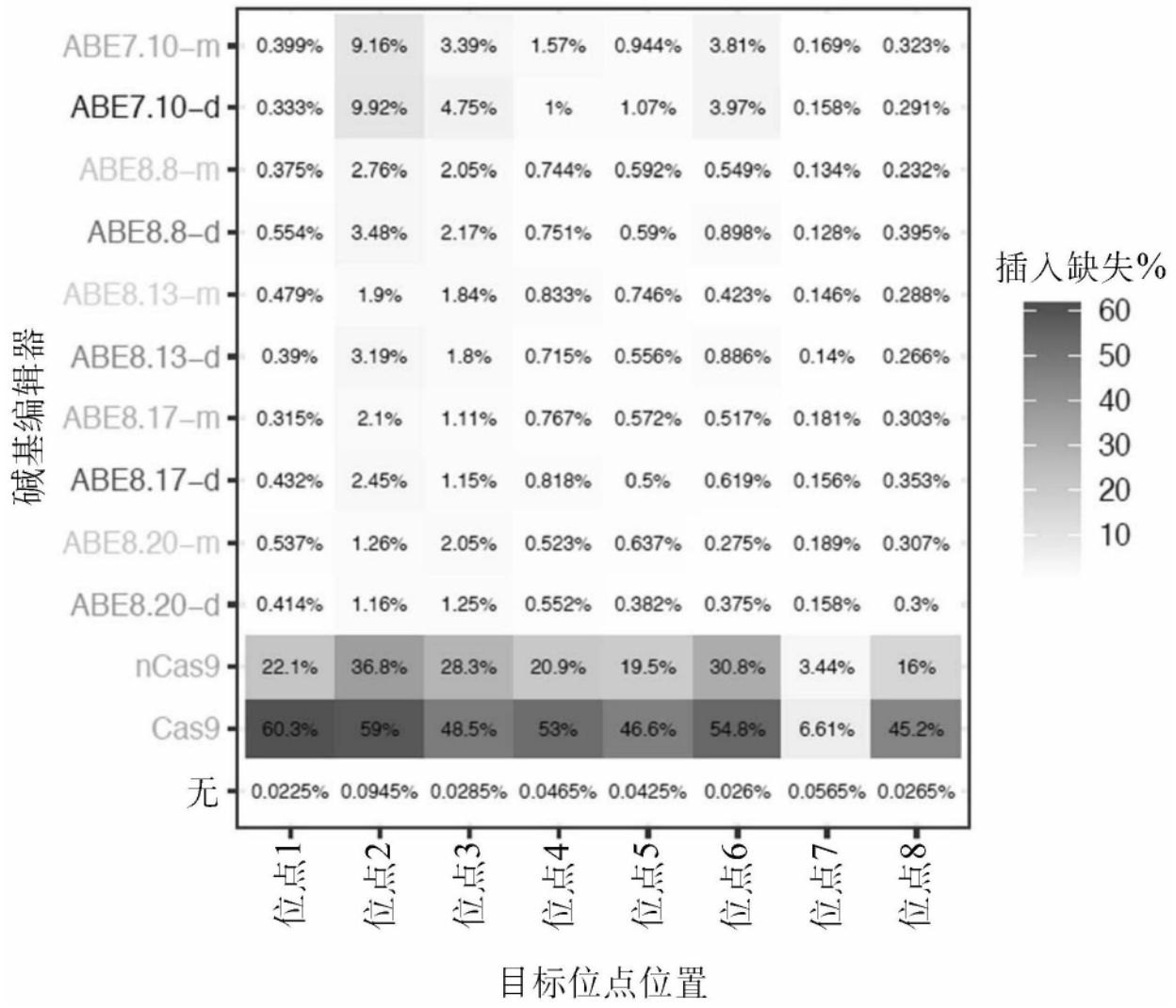

碱基编辑器
ABE7.10-m 0.399% 9.16% 3.39% 1.57% 0.944% 3.81% 0.169% 0.323%
ABE7.10-d 0.333% 9.92% 4.75% 1% 1.07% 3.97% 0.158% 0.291%
ABE8.8-m 0.375% 2.76% 2.05% 0.744% 0.592% 0.549% 0.134% 0.232%
ABE8.8-d 0.554% 3.48% 2.17% 0.751% 0.59% 0.898% 0.128% 0.395%
ABE8.13-m 0.479% 1.9% 1.84% 0.833% 0.746% 0.423% 0.146% 0.288%
ABE8.13-d 0.39% 3.19% 1.8% 0.715% 0.556% 0.886% 0.14% 0.266%
ABE8.17-m 0.315% 2.1% 1.11% 0.767% 0.572% 0.517% 0.181% 0.303%
ABE8.17-d 0.432% 2.45% 1.15% 0.818% 0.5% 0.619% 0.156% 0.353%
ABE8.20-m 0.537% 1.26% 2.05% 0.523% 0.637% 0.275% 0.189% 0.307%
ABE8.20-d 0.414% 1.16% 1.25% 0.552% 0.382% 0.375% 0.158% 0.3%
nCas9 22.1% 36.8% 28.3% 20.9% 19.5% 30.8% 3.44% 16%
Cas9 60.3% 59% 48.5% 53% 46.6% 54.8% 6.61% 45.2%
无 0.0225% 0.0945% 0.0285% 0.0465% 0.0425% 0.026% 0.0565% 0.0265%
位点1
位点2
位点3
位点4
位点5
位点6
位点7
位点8
目标位点位置
插入缺失%
60
50
40
30
20
10

图16

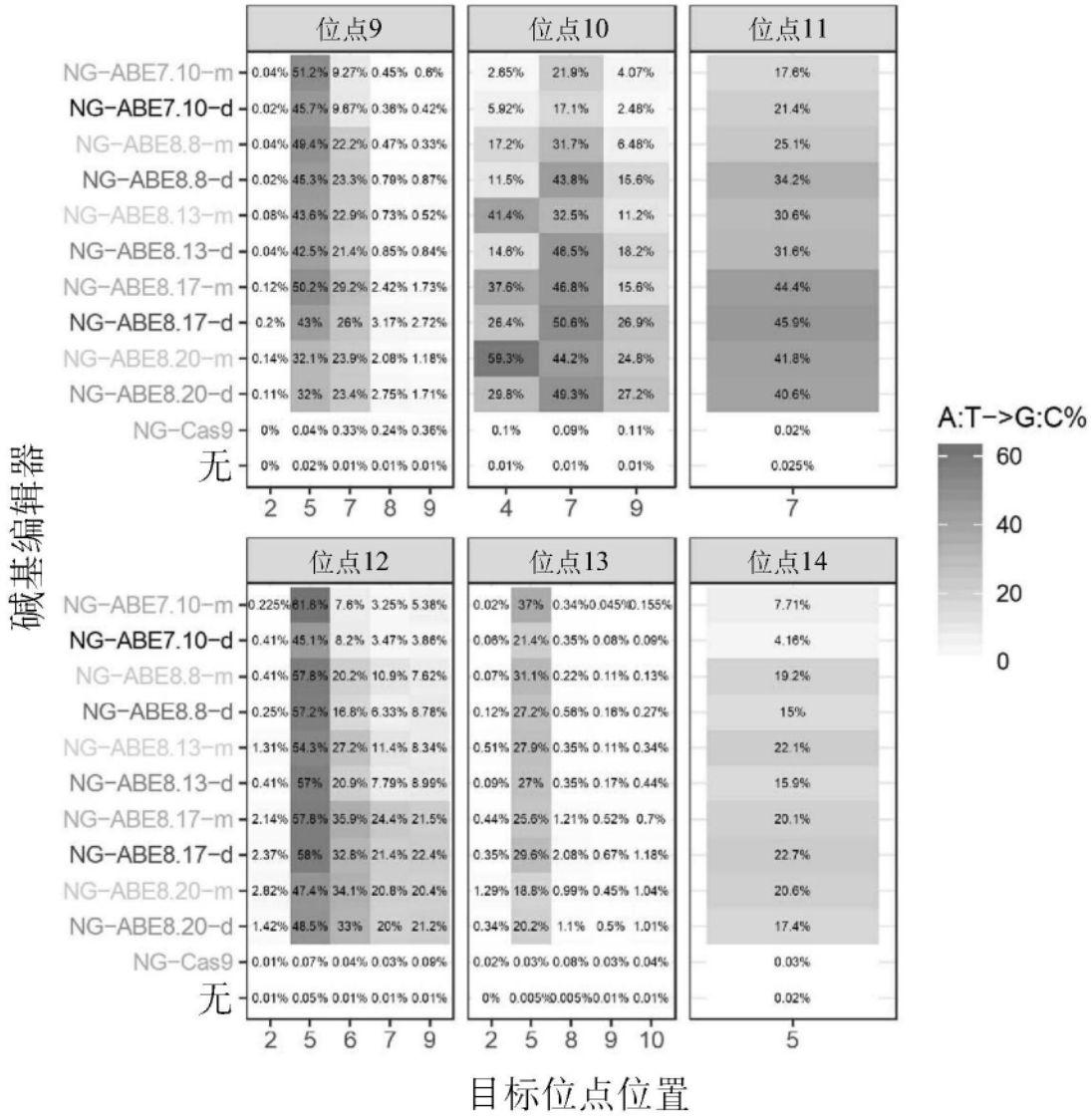

位点9
位点10
位点11
位点12
位点13
位点14
NG-ABE7.10-m
NG-ABE7.10-d
NG-ABE8.8-m
NG-ABE8.8-d
NG-ABE8.13-m
NG-ABE8.13-d
NG-ABE8.17-m
NG-ABE8.17-d
NG-ABE8.20-m
NG-ABE8.20-d
NG-Cas9
无
碱基编辑器
目标位点位置
A:T->G:C%
60
40
20
0

图17

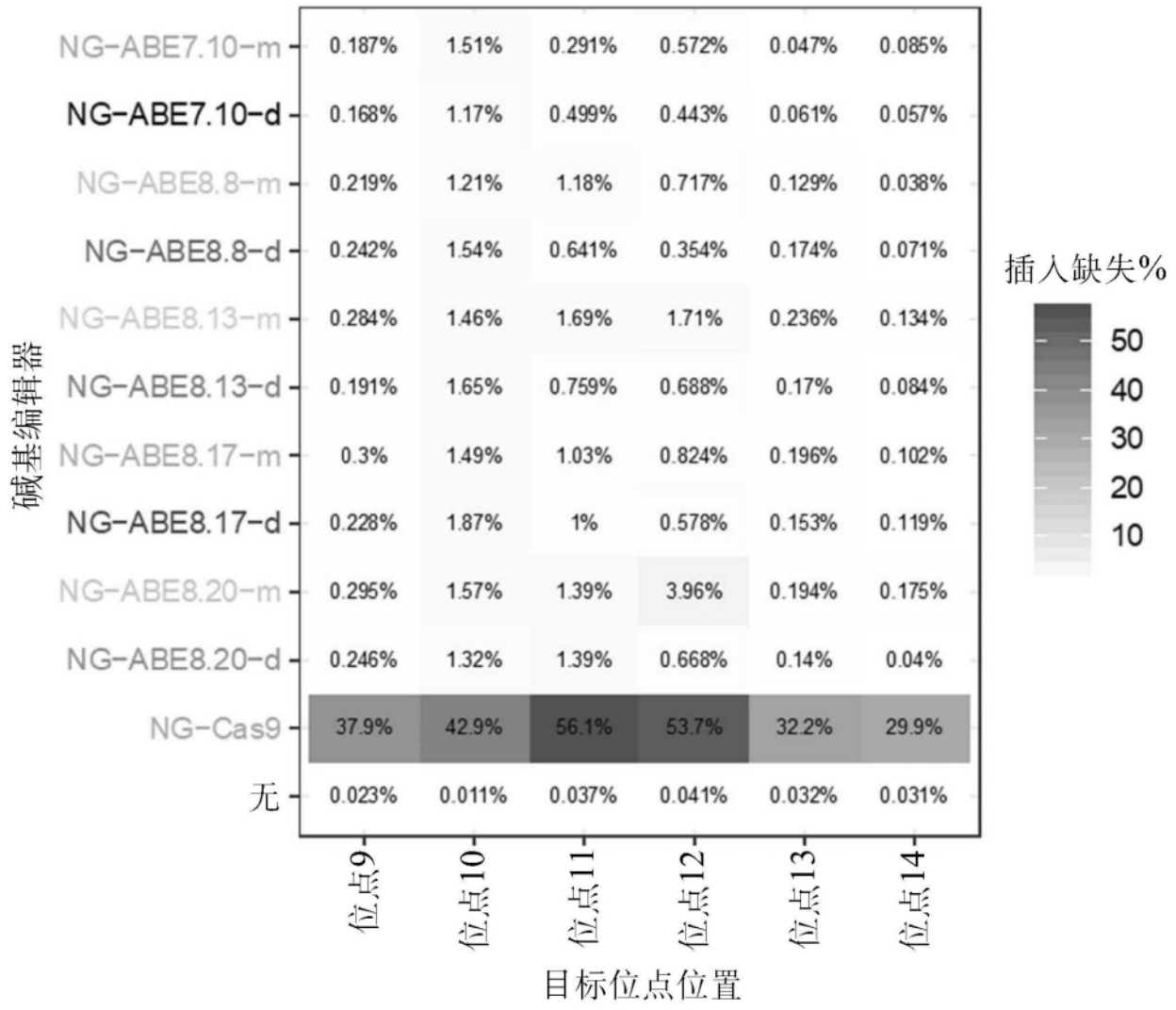

碱基编辑器
NG-ABE7.10-m
NG-ABE7.10-d
NG-ABE8.8-m
NG-ABE8.8-d
NG-ABE8.13-m
NG-ABE8.13-d
NG-ABE8.17-m
NG-ABE8.17-d
NG-ABE8.20-m
NG-ABE8.20-d
NG-Cas9
无
0.187% 1.51% 0.291% 0.572% 0.047% 0.085%
0.168% 1.17% 0.499% 0.443% 0.061% 0.057%
0.219% 1.21% 1.18% 0.717% 0.129% 0.038%
0.242% 1.54% 0.641% 0.354% 0.174% 0.071%
0.284% 1.46% 1.69% 1.71% 0.236% 0.134%
0.191% 1.65% 0.759% 0.688% 0.17% 0.084%
0.3% 1.49% 1.03% 0.824% 0.196% 0.102%
0.228% 1.87% 1% 0.578% 0.153% 0.119%
0.295% 1.57% 1.39% 3.96% 0.194% 0.175%
0.246% 1.32% 1.39% 0.668% 0.14% 0.04%
37.9% 42.9% 56.1% 53.7% 32.2% 29.9%
0.023% 0.011% 0.037% 0.041% 0.032% 0.031%
位点9
位点10
位点11
位点12
位点13
位点14
目标位点位置
插入缺失%
50
40
30
20
10

图18

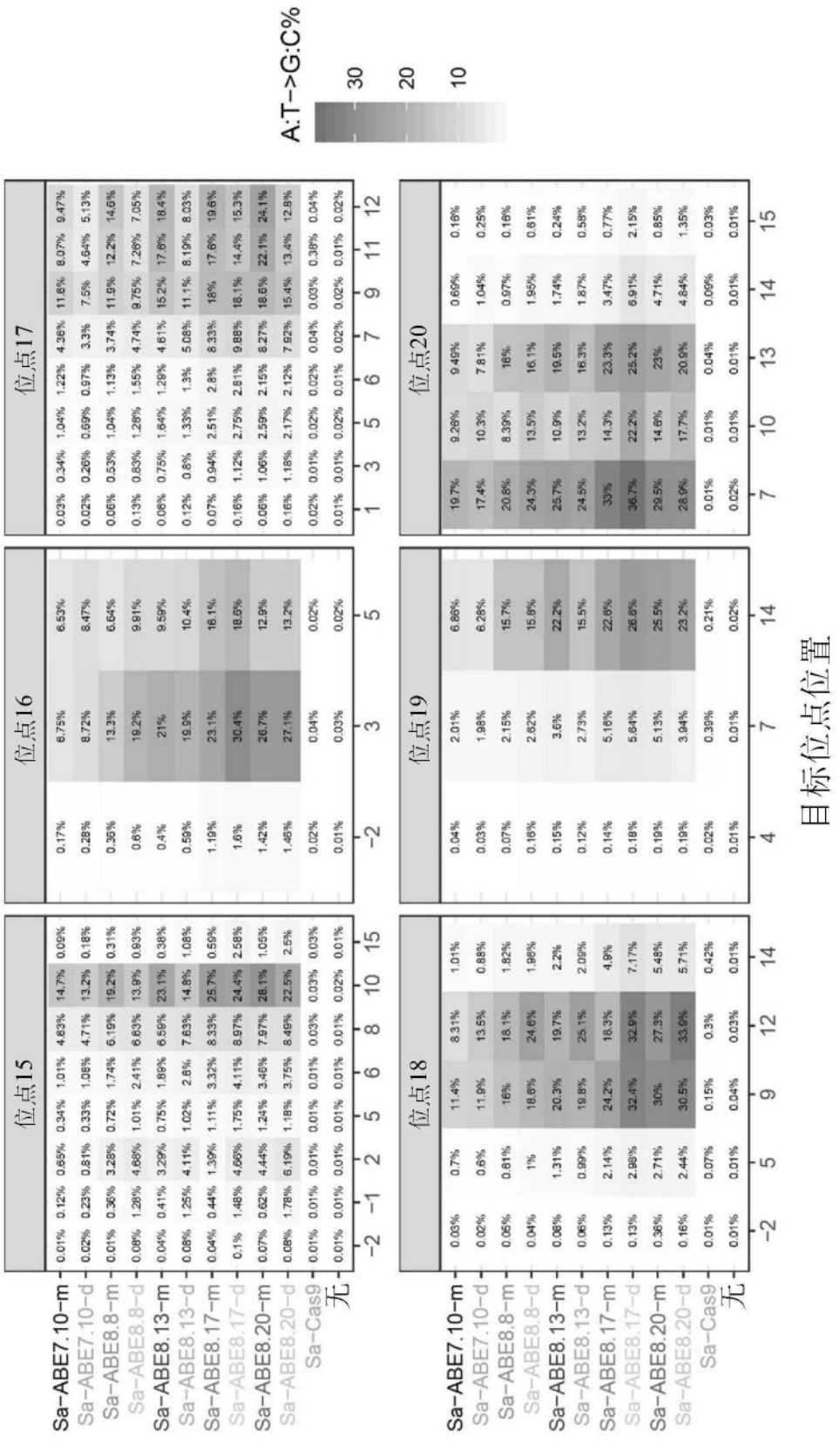

A:T->G:C%
30
20
10
位点15
位点16
位点17
位点18
位点19
位点20
Sa-ABE7.10-m
Sa-ABE7.10-d
Sa-ABE8.8-m
Sa-ABE8.8-d
Sa-ABE8.13-m
Sa-ABE8.13-d
Sa-ABE8.17-m
Sa-ABE8.17-d
Sa-ABE8.20-m
Sa-ABE8.20-d
Sa-Cas9
无
碱基编辑器
目标位点位置

图19

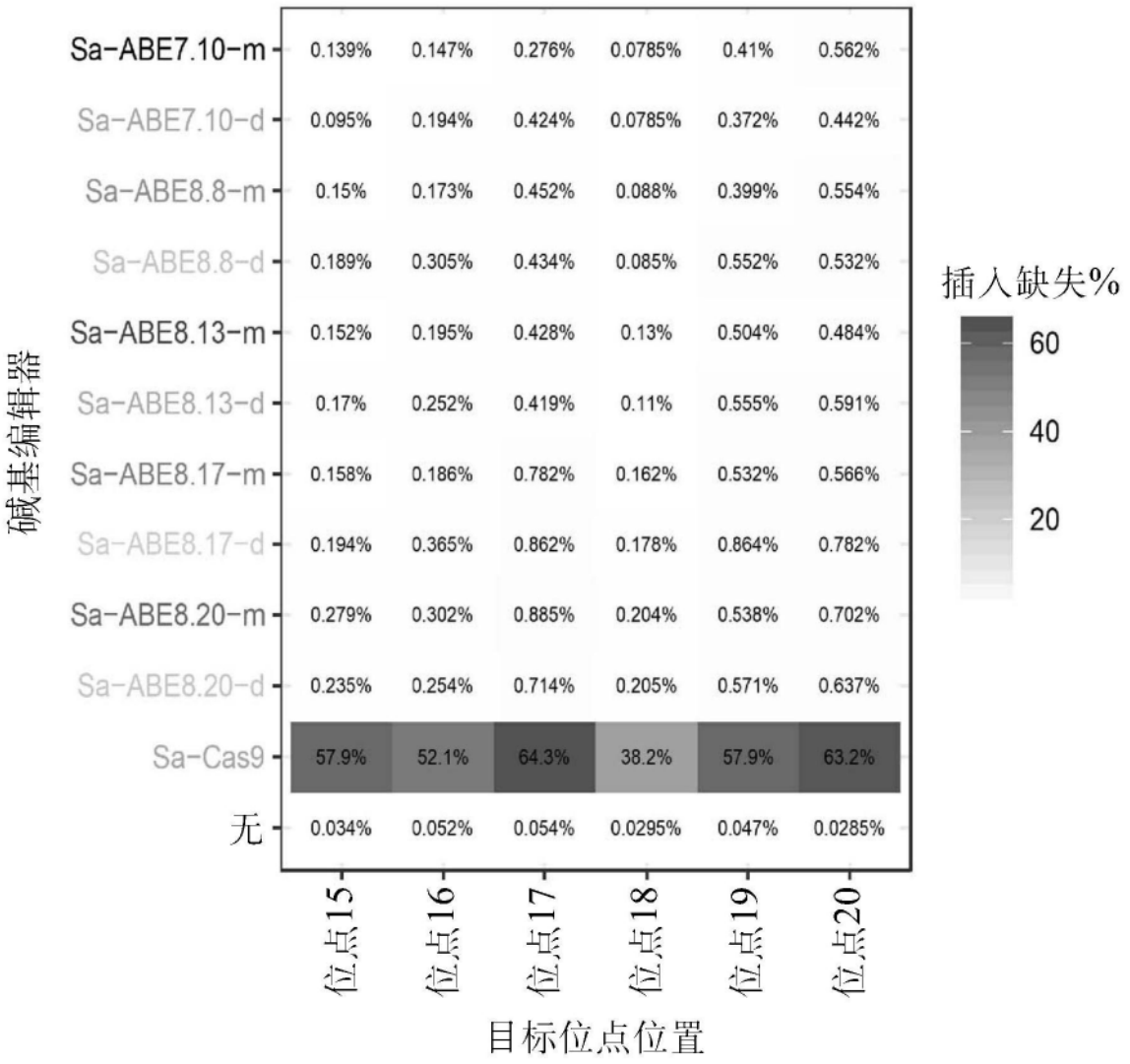

碱基编辑器
Sa-ABE7.10-m
Sa-ABE7.10-d
Sa-ABE8.8-m
Sa-ABE8.8-d
Sa-ABE8.13-m
Sa-ABE8.13-d
Sa-ABE8.17-m
Sa-ABE8.17-d
Sa-ABE8.20-m
Sa-ABE8.20-d
Sa-Cas9
无
0.139% 0.147% 0.276% 0.0785% 0.41% 0.562%
0.095% 0.194% 0.424% 0.0785% 0.372% 0.442%
0.15% 0.173% 0.452% 0.088% 0.399% 0.554%
0.189% 0.305% 0.434% 0.085% 0.552% 0.532%
0.152% 0.195% 0.428% 0.13% 0.504% 0.484%
0.17% 0.252% 0.419% 0.11% 0.555% 0.591%
0.158% 0.186% 0.782% 0.162% 0.532% 0.566%
0.194% 0.365% 0.862% 0.178% 0.864% 0.782%
0.279% 0.302% 0.885% 0.204% 0.538% 0.702%
0.235% 0.254% 0.714% 0.205% 0.571% 0.637%
57.9% 52.1% 64.3% 38.2% 57.9% 63.2%
0.034% 0.052% 0.054% 0.0295% 0.047% 0.0285%
位点15
位点16
位点17
位点18
位点19
位点20
目标位点位置
插入缺失%
60
40
20

图20

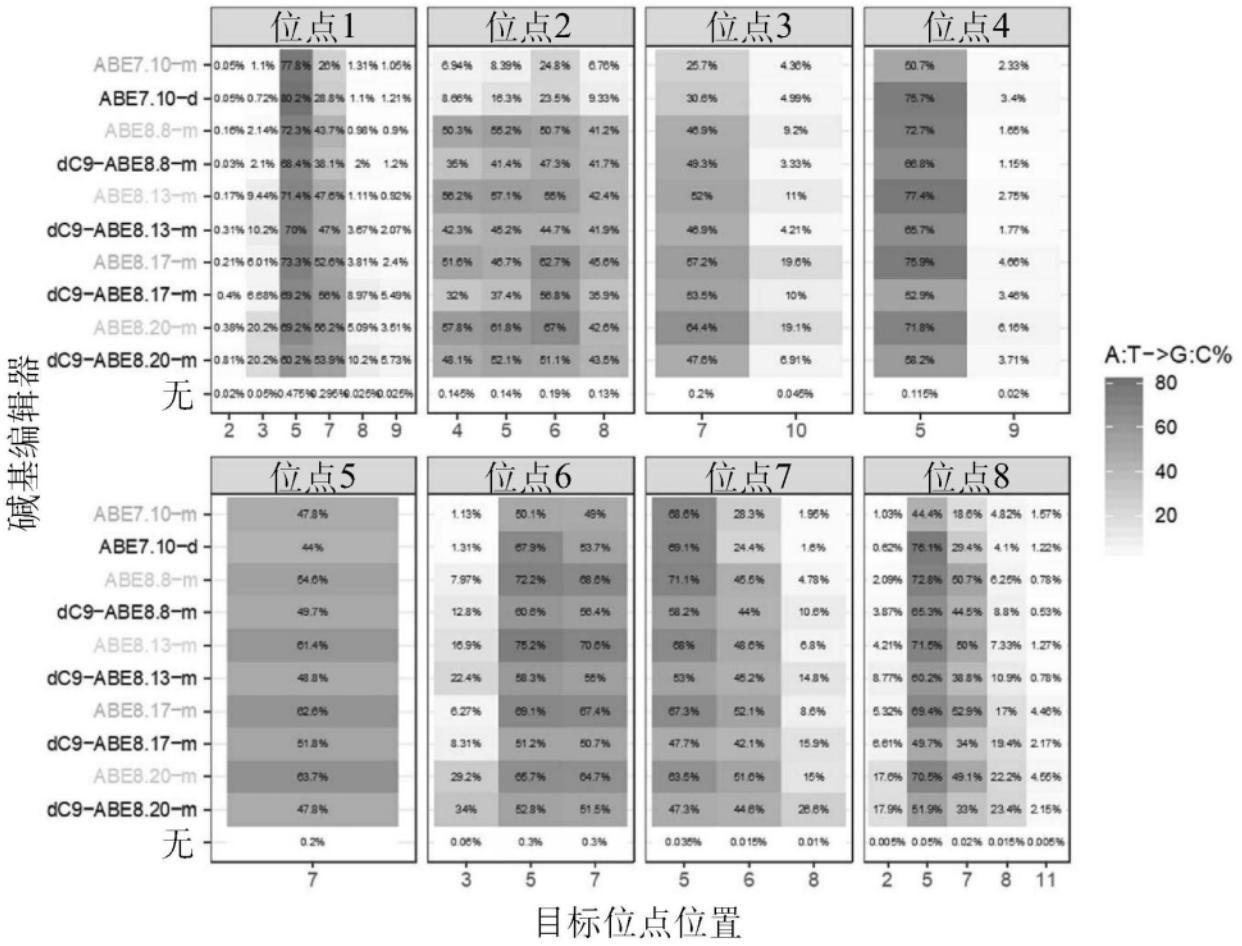

位点1
位点2
位点3
位点4
位点5
位点6
位点7
位点8
碱基编辑器
ABE7.10−m
ABE7.10−d
ABE8.8−m
dC9−ABE8.8−m
ABE8.13−m
dC9−ABE8.13−m
ABE8.17−m
dC9−ABE8.17−m
ABE8.20−m
dC9−ABE8.20−m
无
目标位点位置
A:T−>G:C%
80
60
40
20

图21

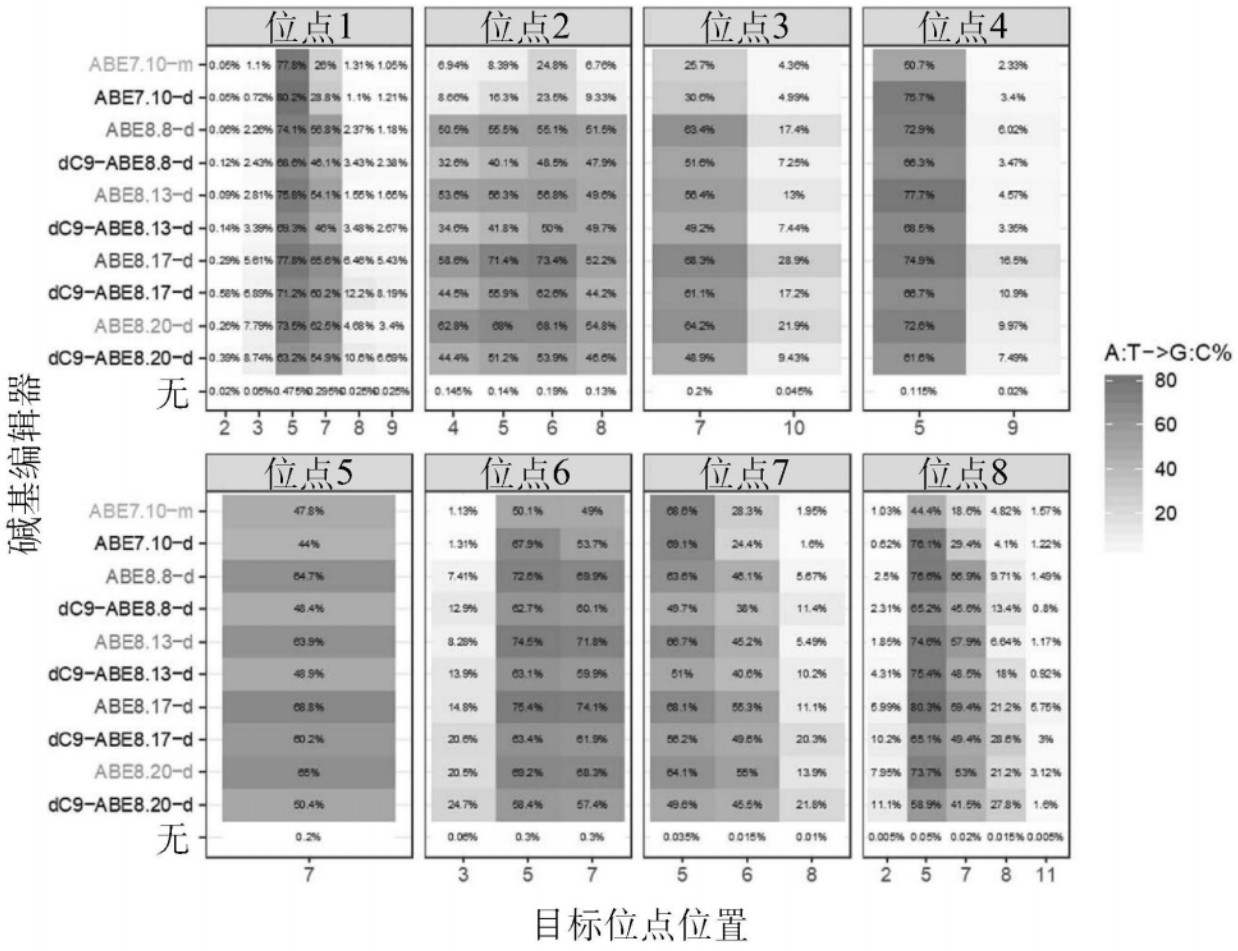

位点1
位点2
位点3
位点4
位点5
位点6
位点7
位点8
碱基编辑器
ABE7.10-m
ABE7.10-d
ABE8.8-d
dC9-ABE8.8-d
ABE8.13-d
dC9-ABE8.13-d
ABE8.17-d
dC9-ABE8.17-d
ABE8.20-d
dC9-ABE8.20-d
无
目标位点位置
A:T->G:C%
80
60
40
20

图22

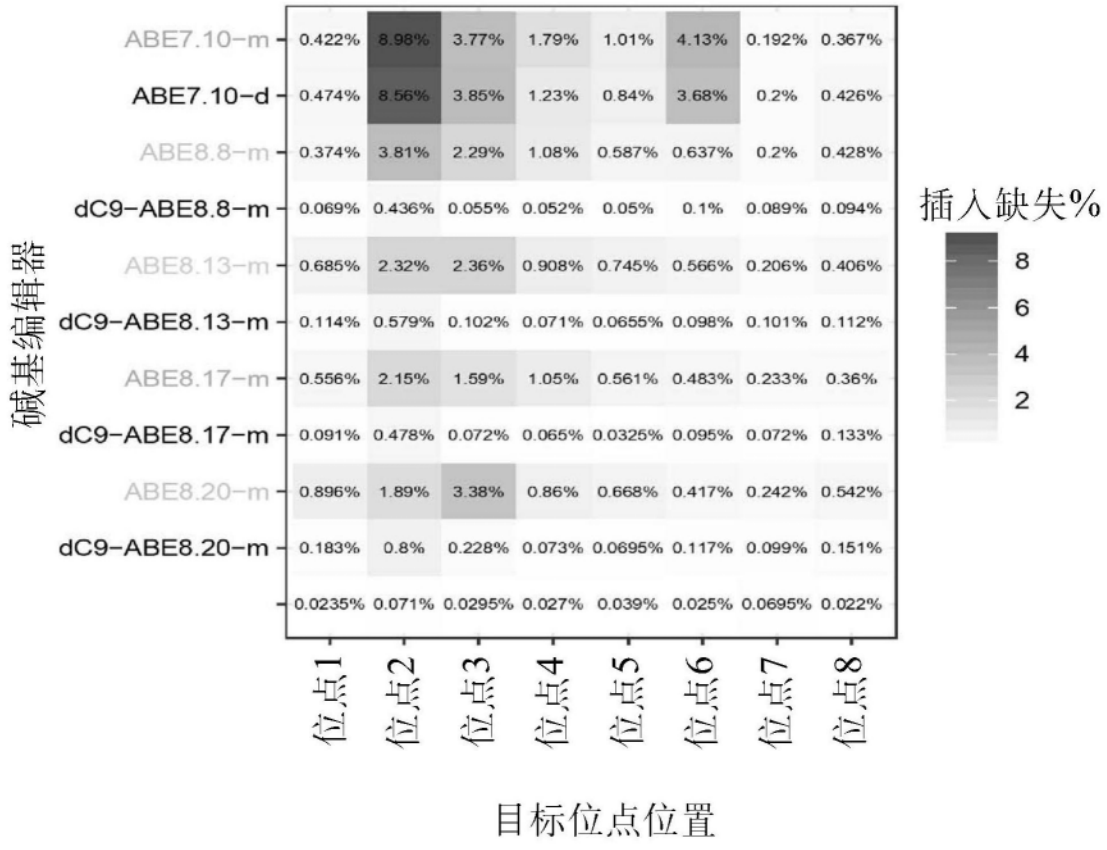

碱基编辑器
ABE7.10-m 0.422% 8.98% 3.77% 1.79% 1.01% 4.13% 0.192% 0.367%
ABE7.10-d 0.474% 8.56% 3.85% 1.23% 0.84% 3.68% 0.2% 0.426%
ABE8.8-m 0.374% 3.81% 2.29% 1.08% 0.587% 0.637% 0.2% 0.428%
dC9-ABE8.8-m 0.069% 0.436% 0.055% 0.052% 0.05% 0.1% 0.089% 0.094%
ABE8.13-m 0.685% 2.32% 2.36% 0.908% 0.745% 0.566% 0.206% 0.406%
dC9-ABE8.13-m 0.114% 0.579% 0.102% 0.071% 0.0655% 0.098% 0.101% 0.112%
ABE8.17-m 0.556% 2.15% 1.59% 1.05% 0.561% 0.483% 0.233% 0.36%
dC9-ABE8.17-m 0.091% 0.478% 0.072% 0.065% 0.0325% 0.095% 0.072% 0.133%
ABE8.20-m 0.896% 1.89% 3.38% 0.86% 0.668% 0.417% 0.242% 0.542%
dC9-ABE8.20-m 0.183% 0.8% 0.228% 0.073% 0.0695% 0.117% 0.099% 0.151%
0.0235% 0.071% 0.0295% 0.027% 0.039% 0.025% 0.0695% 0.022%
位点1
位点2
位点3
位点4
位点5
位点6
位点7
位点8
目标位点位置
插入缺失%
8
6
4
2

图23A

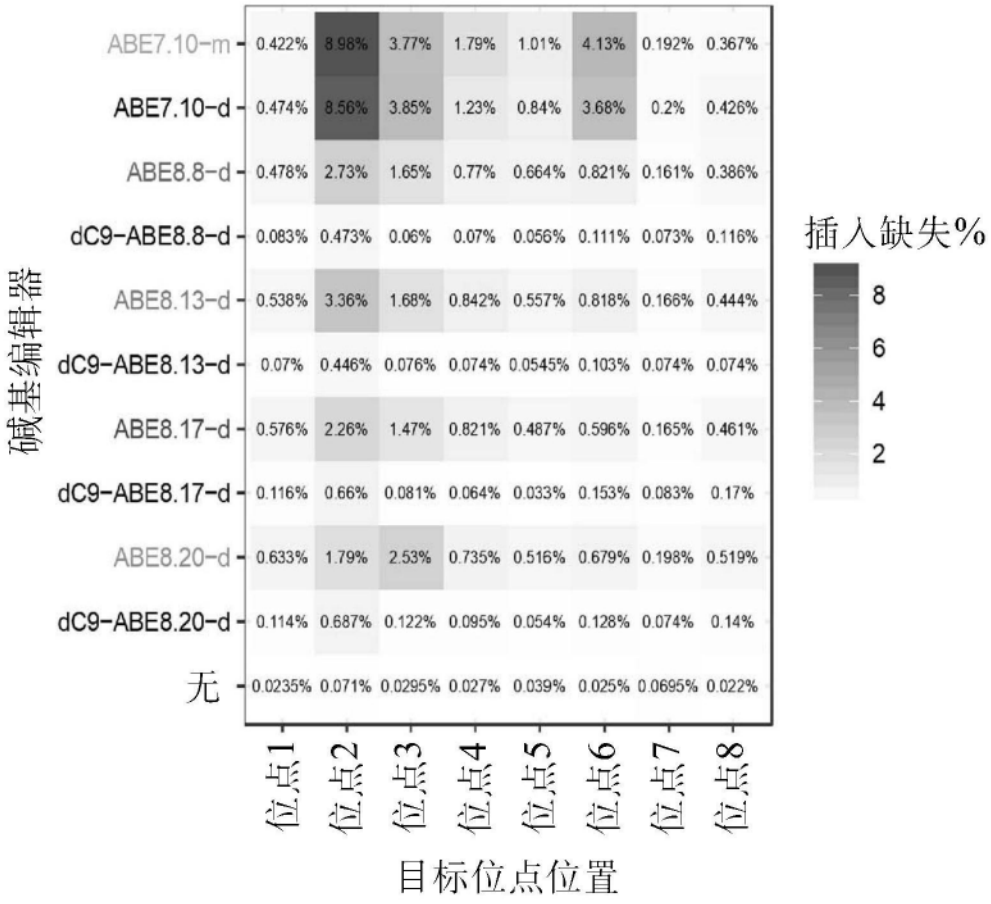
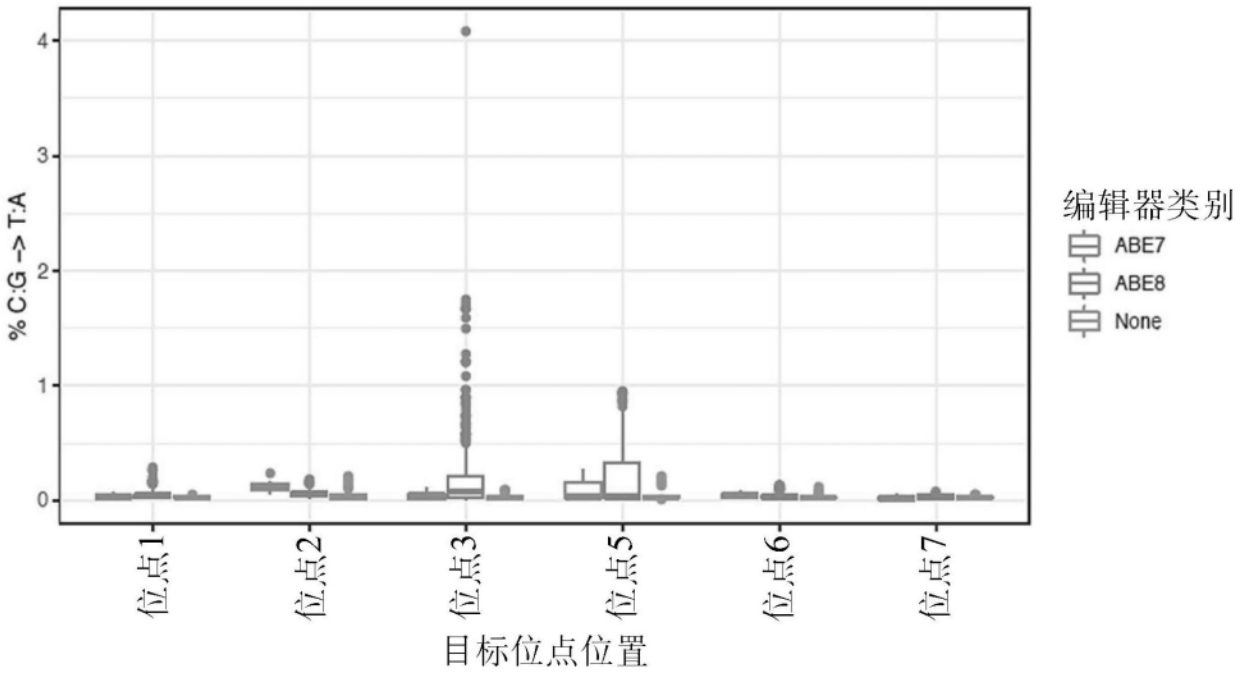

碱基编辑器
ABE7.10-m
ABE7.10-d
ABE8.8-d
dC9-ABE8.8-d
ABE8.13-d
dC9-ABE8.13-d
ABE8.17-d
dC9-ABE8.17-d
ABE8.20-d
dC9-ABE8.20-d
无
0.422% 8.98% 3.77% 1.79% 1.01% 4.13% 0.192% 0.367%
0.474% 8.56% 3.85% 1.23% 0.84% 3.68% 0.2% 0.426%
0.478% 2.73% 1.65% 0.77% 0.664% 0.821% 0.161% 0.386%
0.083% 0.473% 0.06% 0.07% 0.056% 0.111% 0.073% 0.116%
0.538% 3.36% 1.68% 0.842% 0.557% 0.818% 0.166% 0.444%
0.07% 0.446% 0.076% 0.074% 0.0545% 0.103% 0.074% 0.074%
0.576% 2.26% 1.47% 0.821% 0.487% 0.596% 0.165% 0.461%
0.116% 0.66% 0.081% 0.064% 0.033% 0.153% 0.083% 0.17%
0.633% 1.79% 2.53% 0.735% 0.516% 0.679% 0.198% 0.519%
0.114% 0.687% 0.122% 0.095% 0.054% 0.128% 0.074% 0.14%
0.0235% 0.071% 0.0295% 0.027% 0.039% 0.025% 0.0695% 0.022%
位点1
位点2
位点3
位点4
位点5
位点6
位点7
位点8
目标位点位置
插入缺失%
8
6
4
2

图23B



位点	目标DNA序列
1	GAACACAAAGCATAGACTGC
2	GGGAAAGACCCAGCATCCGT
3	GCTCCCATCACATCAACCGG
5	GGCTTCAGGTTCTAAATGAG
6	GCAGAGAGTCGCCGTCTCCA
7	GTGTAAGACCTCAAAAGCAC

图24

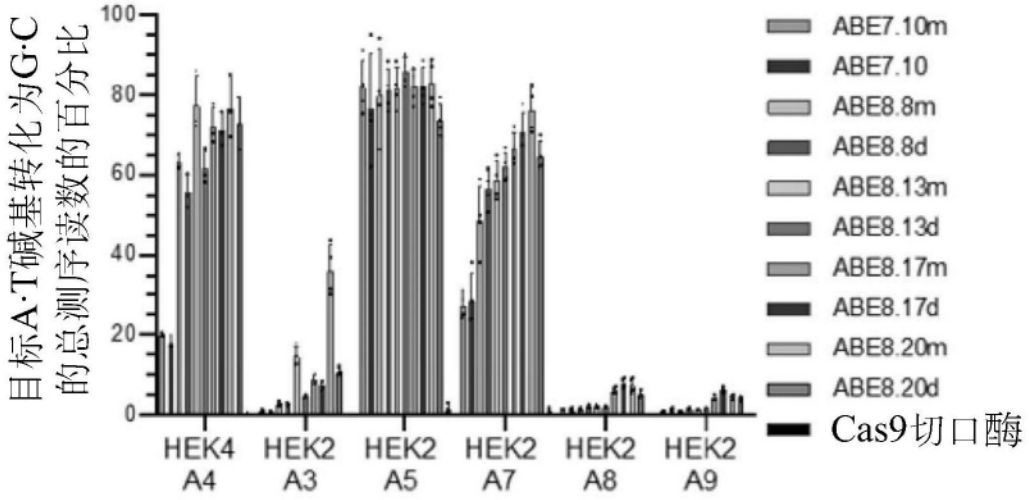


图25A

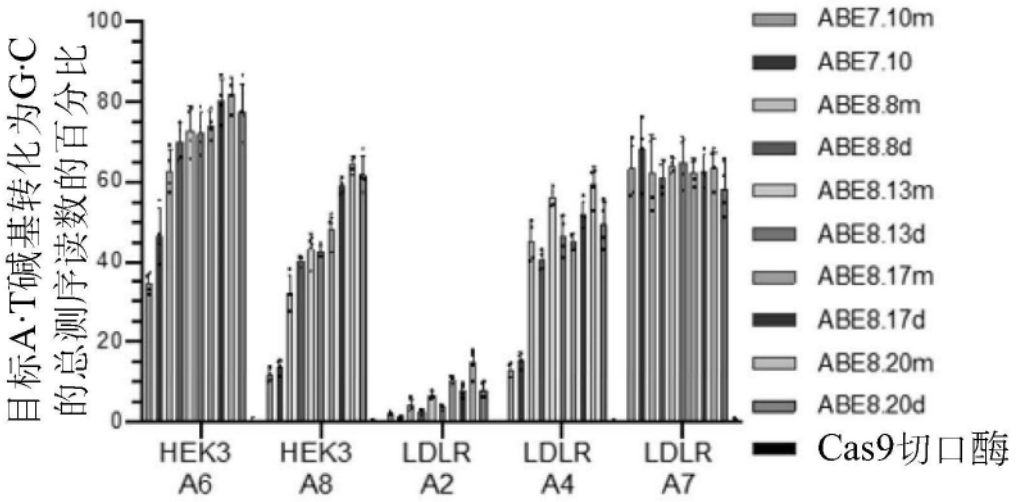

目标A·T碱基转化为G·C的总测序读数的百分比
100
80
60
40
20
0
HEK3 A6
HEK3 A8
LDLR A2
LDLR A4
LDLR A7
ABE7.10m
ABE7.10
ABE8.8m
ABE8.8d
ABE8.13m
ABE8.13d
ABE8.17m
ABE8.17d
ABE8.20m
ABE8.20d
Cas9切口酶

图25B

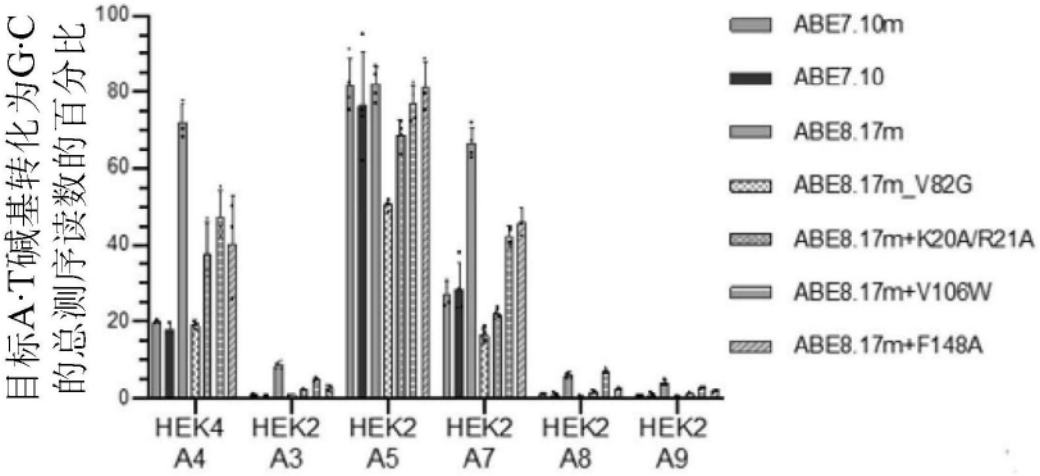

目标A·T碱基转化为G·C的总测序读数的百分比
100
80
60
40
20
0
HEK4 A4
HEK2 A3
HEK2 A5
HEK2 A7
HEK2 A8
HEK2 A9
ABE7.10m
ABE7.10
ABE8.17m
ABE8.17m_V82G
ABE8.17m+K20A/R21A
ABE8.17m+V106W
ABE8.17m+F148A

图25C

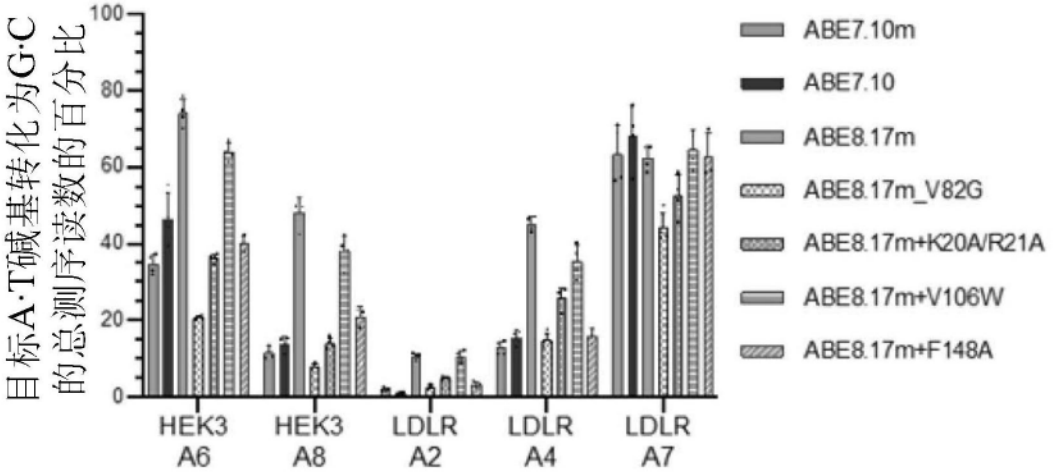

目标A·T碱基转化为G·C的总测序读数的百分比
100
80
60
40
20
0
HEK3 A6
HEK3 A8
LDLR A2
LDLR A4
LDLR A7
ABE7.10m
ABE7.10
ABE8.17m
ABE8.17m_V82G
ABE8.17m+K20A/R21A
ABE8.17m+V106W
ABE8.17m+F148A

图25D

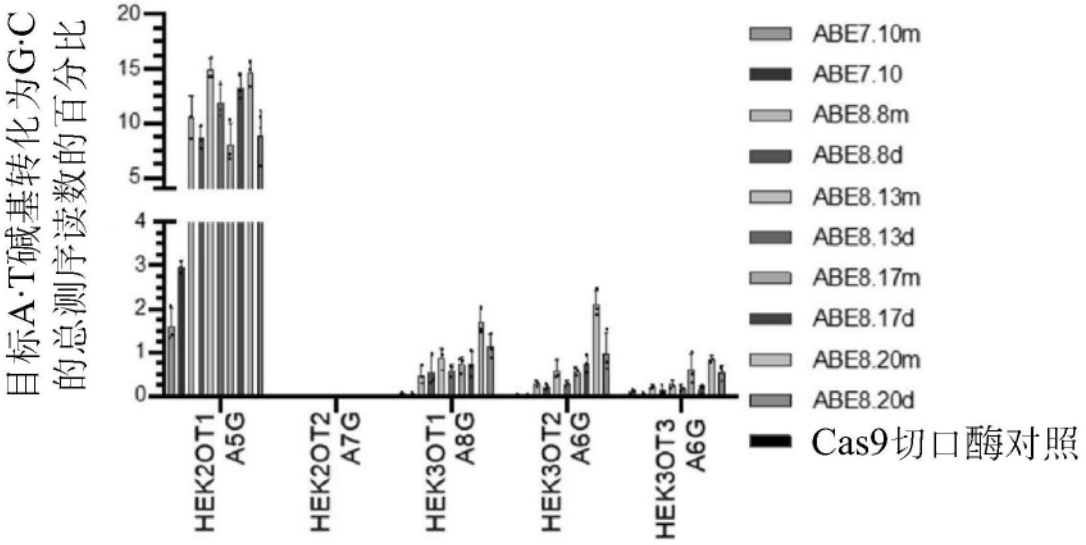

目标A·T碱基转化为G·C的总测序读数的百分比
20
15
10
5
4
3
2
1
0
HEK2OT1 A5G
HEK2OT2 A7G
HEK3OT1 A8G
HEK3OT2 A6G
HEK3OT3 A6G
ABE7.10m
ABE7.10
ABE8.8m
ABE8.8d
ABE8.13m
ABE8.13d
ABE8.17m
ABE8.17d
ABE8.20m
ABE8.20d
Cas9切口酶对照

图25E

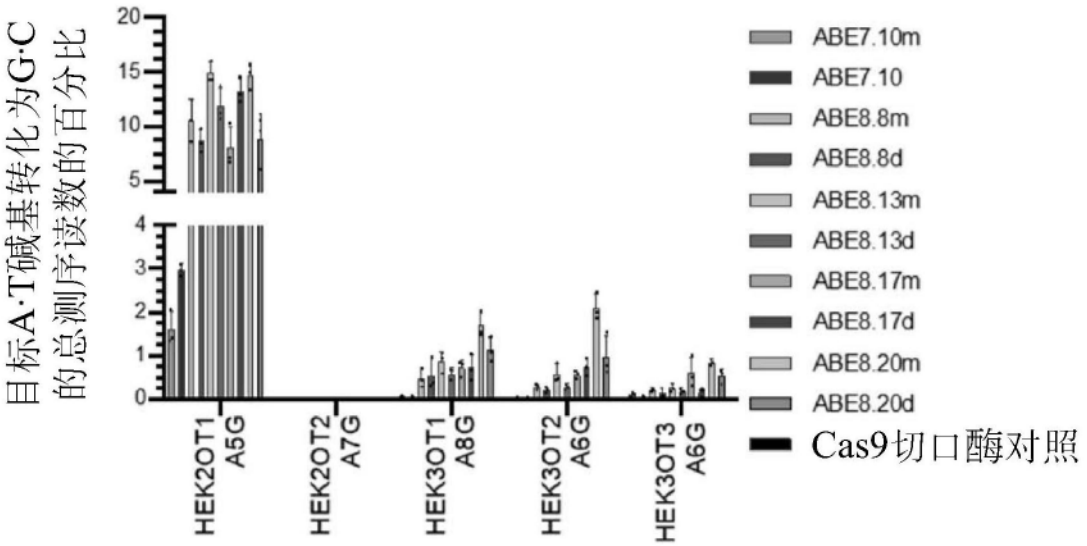

目标A·T碱基转化为G·C的总测序读数的百分比
20
15
10
5
4
3
2
1
0
HEK2OT1 A5G
HEK2OT2 A7G
HEK3OT1 A8G
HEK3OT2 A6G
HEK3OT3 A6G
ABE7.10m
ABE7.10
ABE8.8m
ABE8.8d
ABE8.13m
ABE8.13d
ABE8.17m
ABE8.17d
ABE8.20m
ABE8.20d
Cas9切口酶对照

图25F

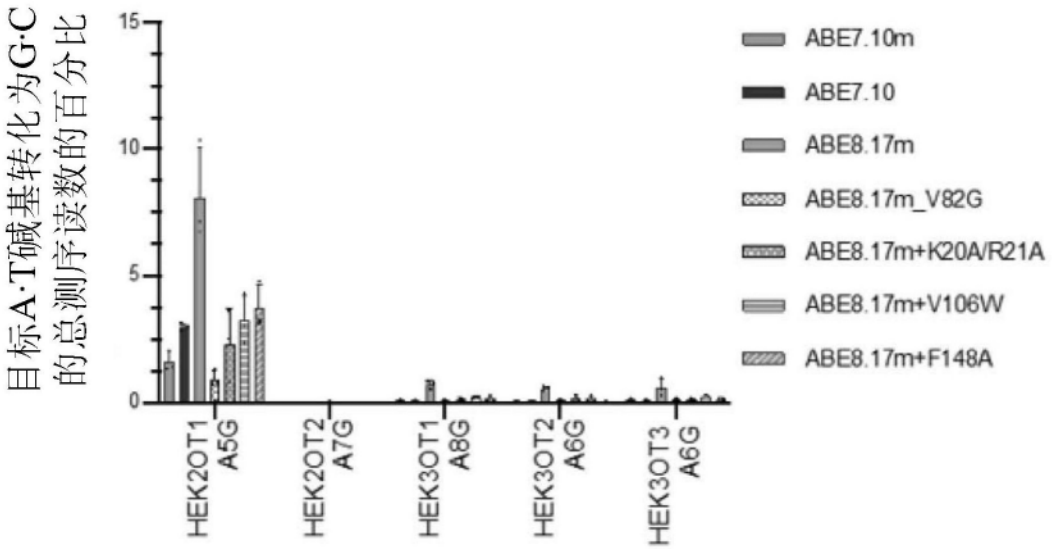

目标A·T碱基转化为G·C的总测序读数的百分比
15
10
5
0
HEK2OT1 A5G
HEK2OT2 A7G
HEK3OT1 A8G
HEK3OT2 A6G
HEK3OT3 A6G
ABE7.10m
ABE7.10
ABE8.17m
ABE8.17m_V82G
ABE8.17m+K20A/R21A
ABE8.17m+V106W
ABE8.17m+F148A

图25G

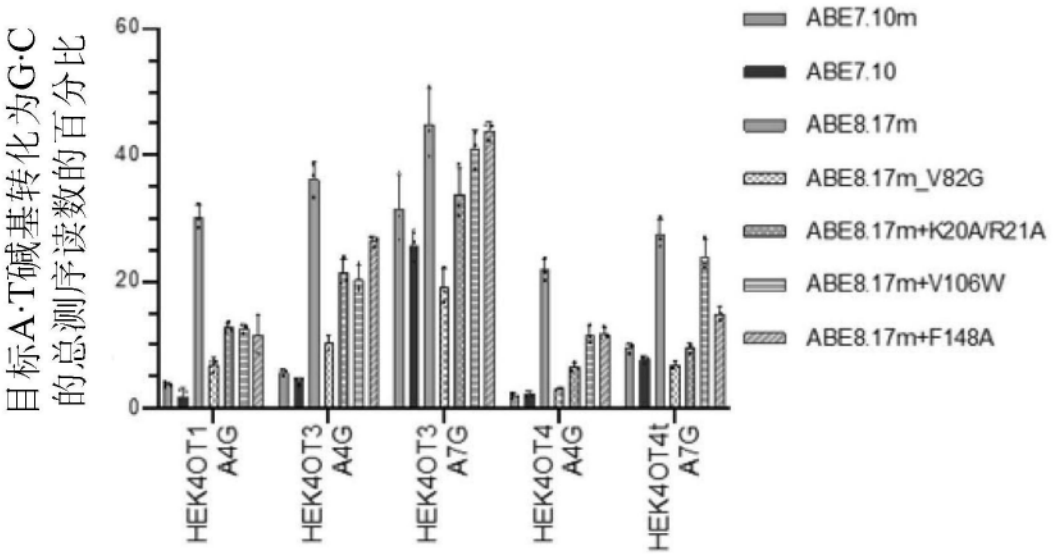

目标A·T碱基转化为G·C的总测序读数的百分比
60
40
20
0
ABE7.10m
ABE7.10
ABE8.17m
ABE8.17m_V82G
ABE8.17m+K20A/R21A
ABE8.17m+V106W
ABE8.17m+F148A
HEK4OT1 A4G
HEK4OT3 A4G
HEK4OT3 A7G
HEK4OT4 A4G
HEK4OT4t A7G

图25H

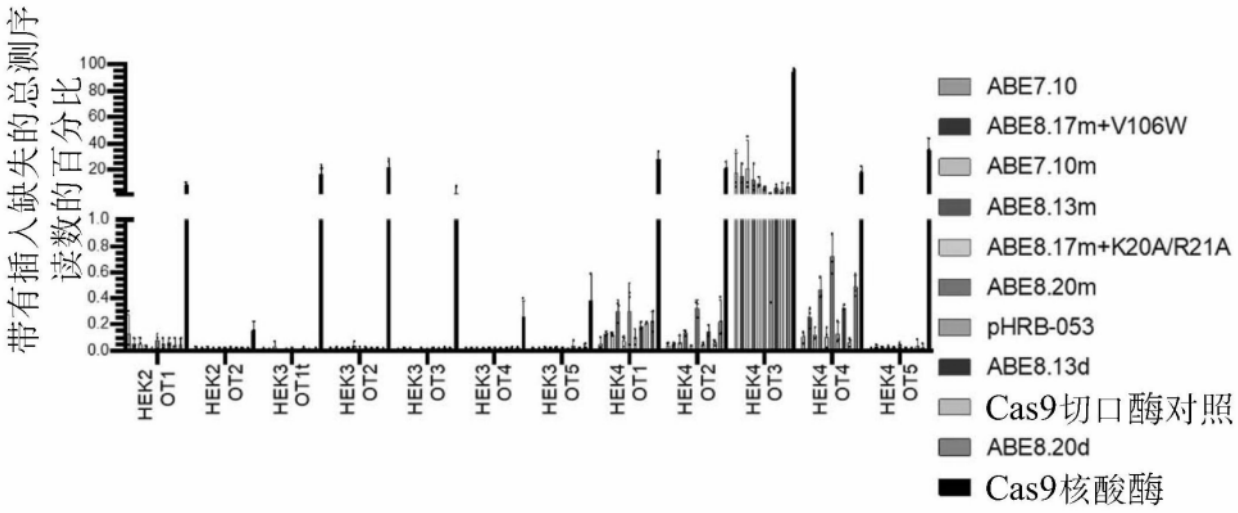

带有插入缺失的总测序读数的百分比
100
80
60
40
20
1.0
0.8
0.6
0.4
0.2
0.0
ABE7.10
ABE8.17m+V106W
ABE7.10m
ABE8.13m
ABE8.17m+K20A/R21A
ABE8.20m
pHRB-053
ABE8.13d
Cas9切口酶对照
ABE8.20d
Cas9核酸酶
HEK2 OT1
HEK2 OT2
HEK3 OT1t
HEK3 OT2
HEK3 OT3
HEK3 OT4
HEK3 OT5
HEK4 OT1
HEK4 OT2
HEK4 OT3
HEK4 OT4
HEK4 OT5

图26

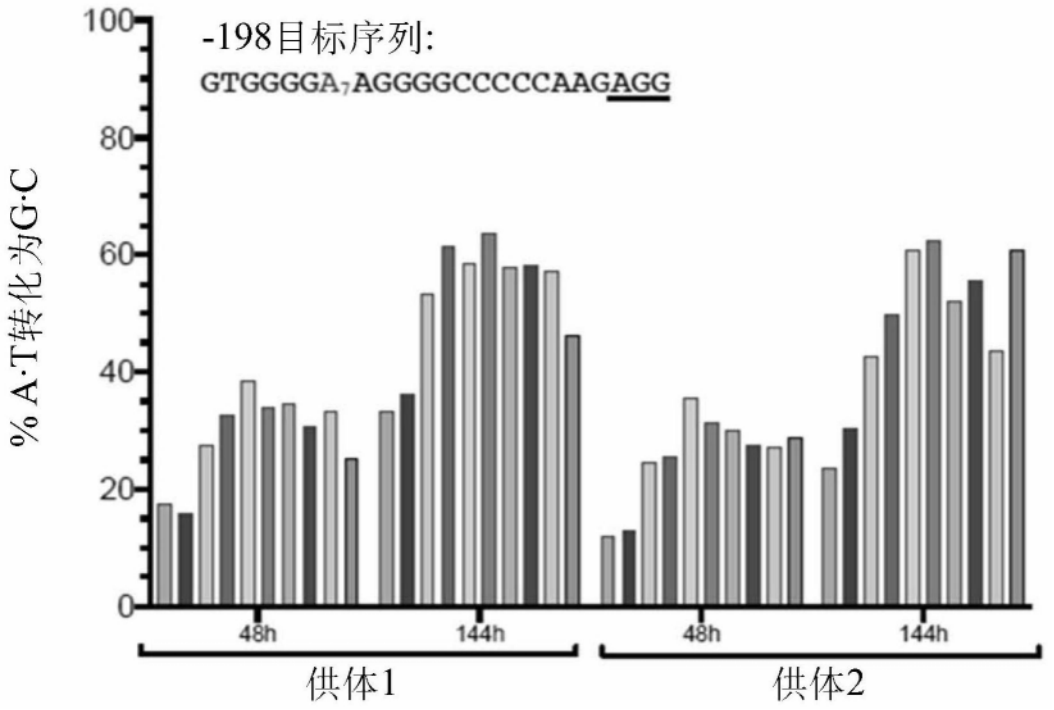

-198目标序列:
GTGGGGA7AGGGGCCCCCAAGAGG
% A·T转化为G·C
100
80
60
40
20
0
48h
144h
48h
144h
供体1
供体2

图27A

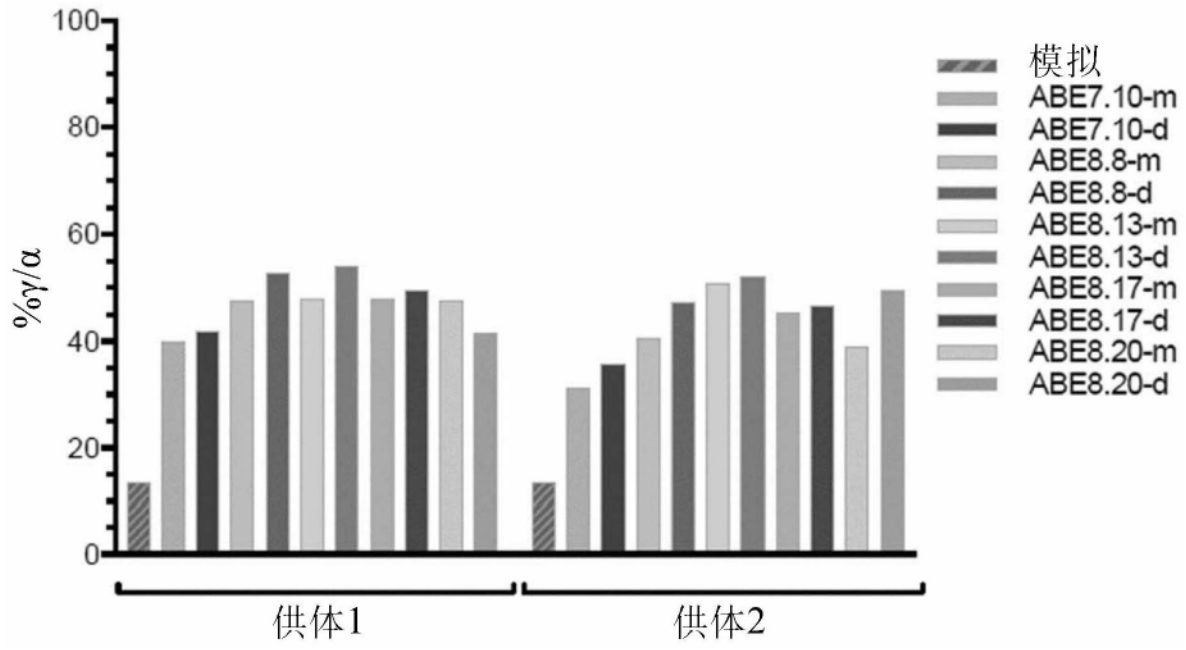

%γ/α
100
80
60
40
20
0
模拟
ABE7.10-m
ABE7.10-d
ABE8.8-m
ABE8.8-d
ABE8.13-m
ABE8.13-d
ABE8.17-m
ABE8.17-d
ABE8.20-m
ABE8.20-d
供体1
供体2

图27B


供体1
供体1
供体2
供体2
48h
144h
48h
144h
处理
ABE7.10-m
ABE7.10-d
ABE8.8-m
ABE8.8-d
ABE8.13-m
ABE8.13-d
ABE8.17-m
ABE8.17-d
ABE8.20-m
ABE8.20-d
模拟
15.7% 1.34% 1.48%
14.6% 0.96% 0.69%
24.9% 1.96% 0.36%
26.3% 7.17% 1.21%
32.3% 5.12% 0.3%
24.5% 7.82% 0.09%
29.4% 2.79% 0.35%
22.3% 7.6% 0.22%
29.5% 3.59% 0.28%
20.3% 4.21% 0.18%
0.08% 0% 0.01%
28.8% 3.69% 2.63%
31.5% 4.1% 1.34%
44.4% 7.73% 0.51%
39.7% 16.4% 1.6%
41.5% 14.9% 0.16%
35.6% 19.5% 0.24%
48.7% 11.1% 0.45%
35.1% 20.9% 0.3%
49.3% 12.6% 0.26%
32.1% 12.4% 0.24%
0.09% 0.01% 0.01%
10.8% 0.92% 1%
11.6% 1.04% 0.43%
22.7% 1.59% 0.24%
23.6% 5.69% 1.2%
30.4% 4.39% 0.37%
21.2% 5.06% 0.15%
22.8% 2.3% 0.25%
20.5% 6.17% 0.29%
28.1% 2.66% 0.26%
24.1% 3.88% 0.26%
0.07% 0.01% 0.01%
20.8% 2.2% 2.01%
26.9% 2.86% 1.01%
36.7% 5.05% 0.26%
36.8% 13.4% 1.63%
45.6% 13.4% 0.33%
30.8% 11.6% 0.19%
40.4% 7.97% 0.46%
36.1% 17.6% 0.39%
50.5% 10.3% 0.49%
44.2% 14.7% 0.31%
0.08% 0.02% 0.01%
7G 7G+8G 8G
A:T->G:C%
50
40
30
20
10
0
目标位点位置

图28A

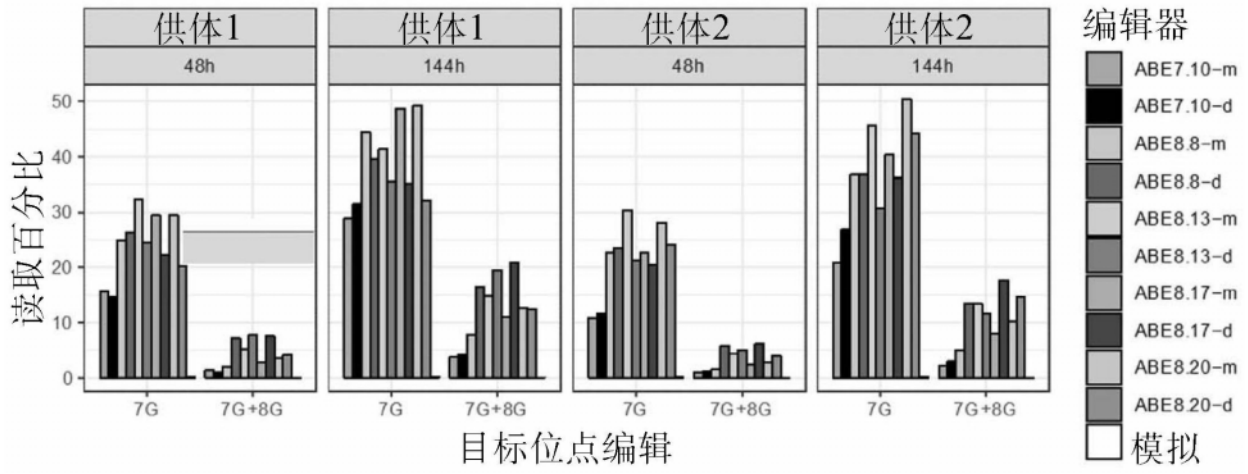

供体1
供体1
供体2
供体2
48h
144h
48h
144h
读取百分比
50
40
30
20
10
0
7G 7G+8G
目标位点编辑
编辑器
ABE7.10-m
ABE7.10-d
ABE8.8-m
ABE8.8-d
ABE8.13-m
ABE8.13-d
ABE8.17-m
ABE8.17-d
ABE8.20-m
ABE8.20-d
模拟

图28B

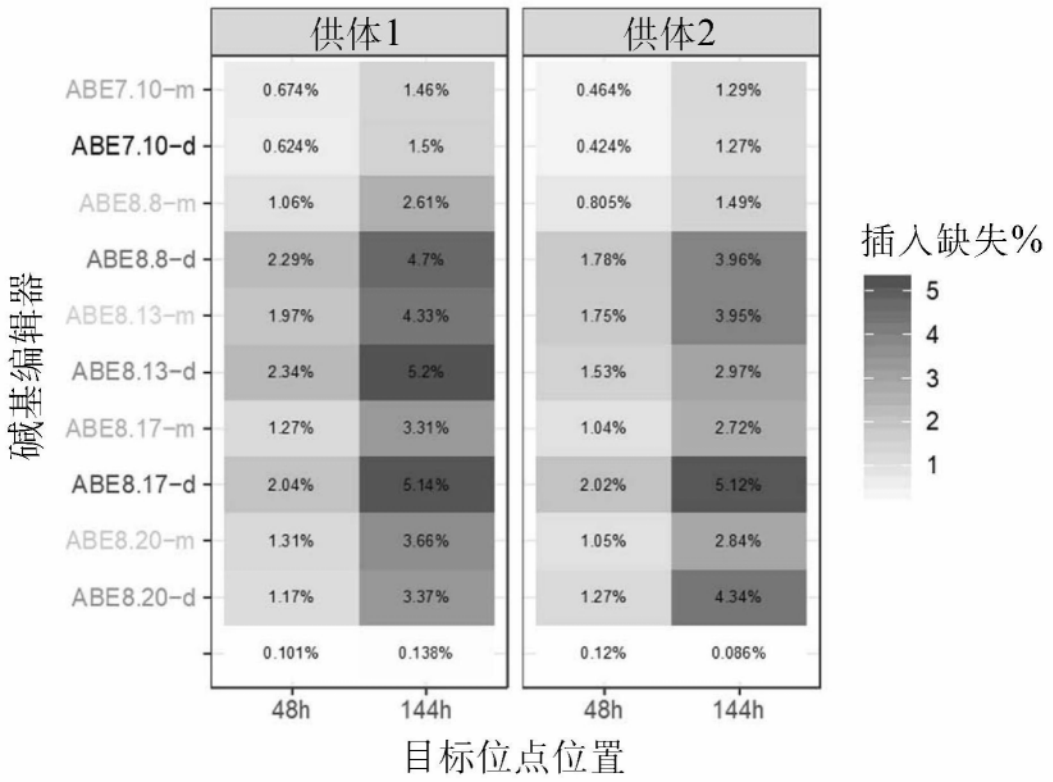
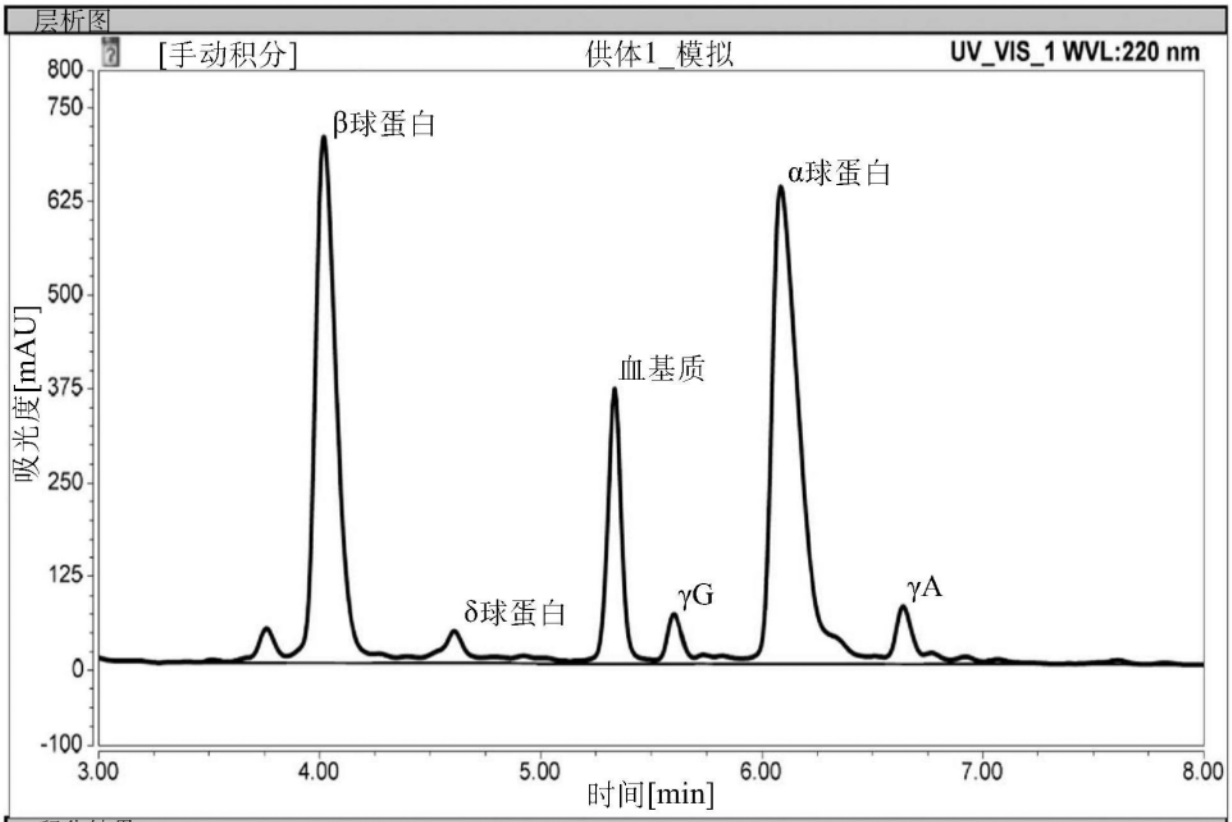

供体1
供体2
碱基编辑器
ABE7.10-m
ABE7.10-d
ABE8.8-m
ABE8.8-d
ABE8.13-m
ABE8.13-d
ABE8.17-m
ABE8.17-d
ABE8.20-m
ABE8.20-d
0.674%
1.46%
0.464%
1.29%
0.624%
1.5%
0.424%
1.27%
1.06%
2.61%
0.805%
1.49%
2.29%
4.7%
1.78%
3.96%
1.97%
4.33%
1.75%
3.95%
2.34%
5.2%
1.53%
2.97%
1.27%
3.31%
1.04%
2.72%
2.04%
5.14%
2.02%
5.12%
1.31%
3.66%
1.05%
2.84%
1.17%
3.37%
1.27%
4.34%
0.101%
0.138%
0.12%
0.086%
48h
144h
48h
144h
目标位点位置
插入缺失%
5
4
3
2
1

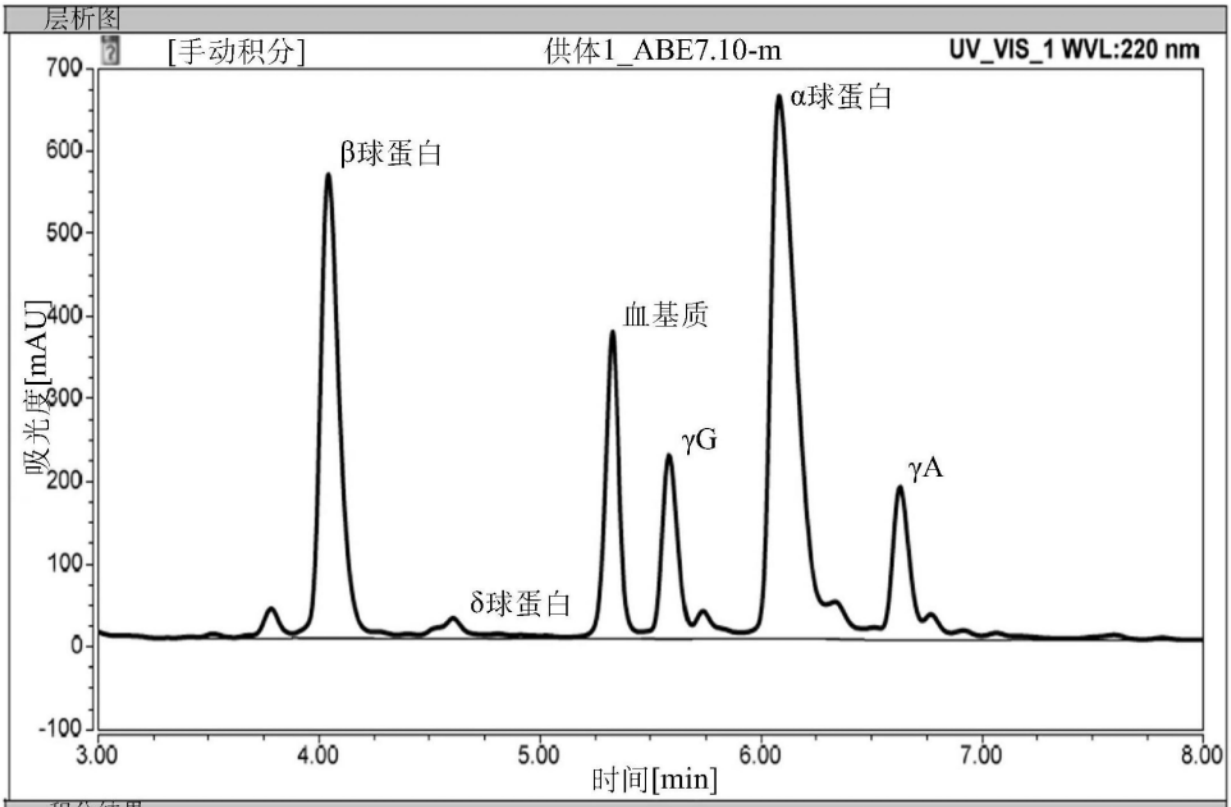
图29



积分结果

No.	波峰名称	滞留时间 min	面积 mAU*min	高度 mAU	相对面积 %	相对高度 %
6	β球蛋白	4.020	77.631	701.634	33.61	32.98
n.a.	Hbs	n.a.	n.a.	n.a.	n.a.	n.a.
10	δ球蛋白	4.610	3.780	42.943	1.64	2.02
14	血基质	5.337	27.191	366.542	11.77	17.23
15	γG	5.603	5.248	66.622	2.27	3.13
18	α球蛋白	6.083	88.477	636.811	38.31	29.94
21	γA	6.637	6.751	77.427	2.92	3.64

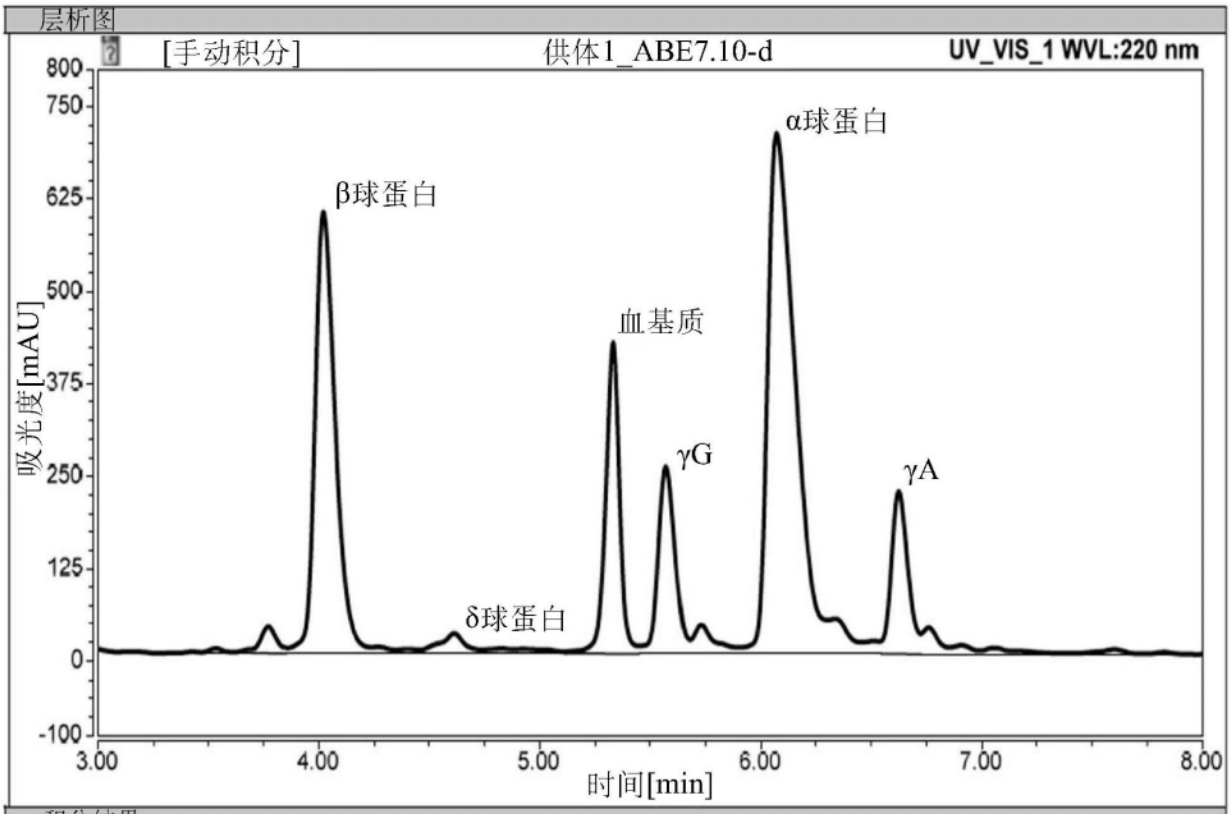
图30



积分结果

No.	波峰名称	滞留时间 min	面积 mAU*min	高度 mAU	相对面积 %	相对高度 %
5	β球蛋白	4.043	56.404	561.657	24.21	24.80
n.a.	Hbs	n.a.	n.a.	n.a.	n.a.	n.a.
9	δ球蛋白	4.607	2.463	24.382	1.06	1.08
13	血基质	5.330	27.334	371.802	11.73	16.42
14	γG	5.583	18.764	222.410	8.06	9.82
16	α球蛋白	6.080	88.236	657.695	37.88	29.04
19	γA	6.627	16.470	185.055	7.07	8.17

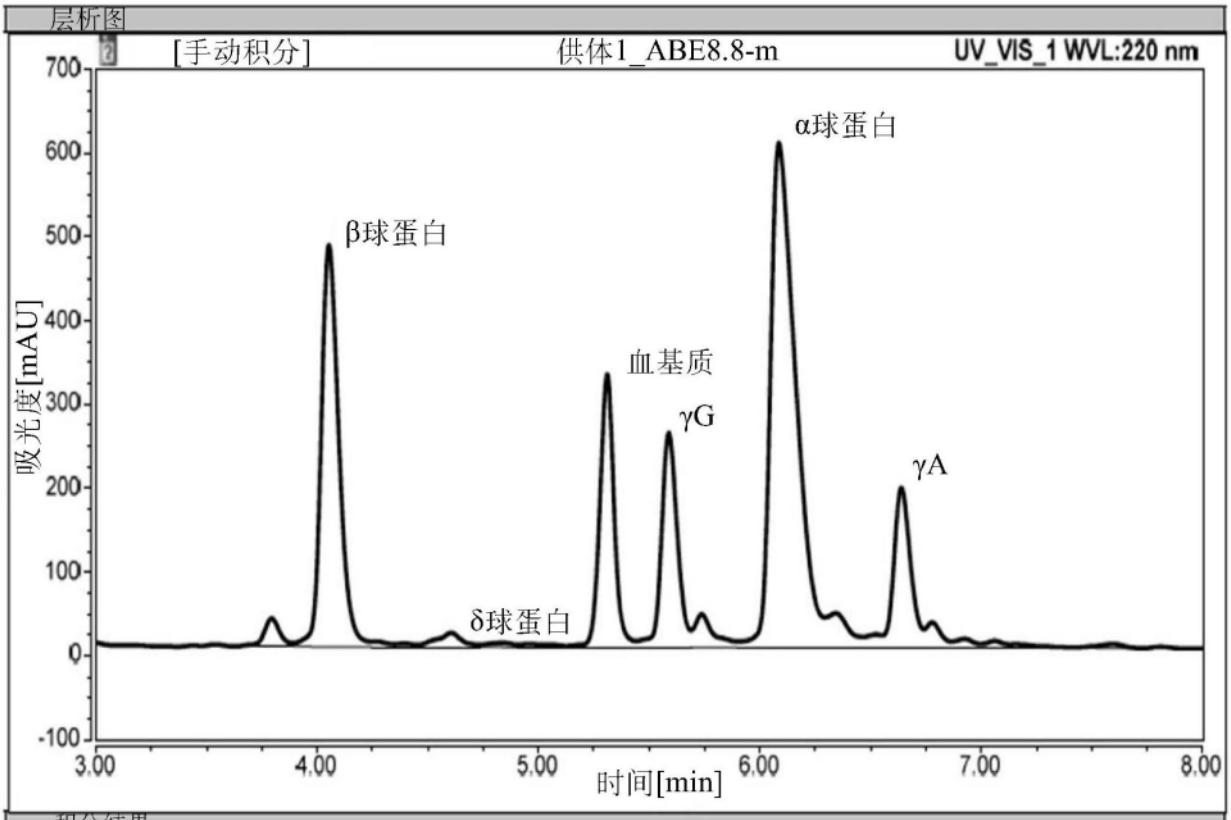
图31



积分结果

No.	波峰名称	滞留时间 min	面积 mAU*min	高度 mAU	相对面积 %	相对高度 %
6	β球蛋白	4.020	63.540	596.818	23.88	23.74
n.a.	Hbs	n.a.	n.a.	n.a.	n.a.	n.a.
10	δ球蛋白	4.613	2.822	27.354	1.06	1.09
14	血基质	5.333	31.163	421.735	11.71	16.77
15	γG	5.570	22.558	253.399	8.48	10.08
18	α球蛋白	6.070	100.612	704.910	37.81	28.03
21	γA	6.623	19.220	220.846	7.22	8.78

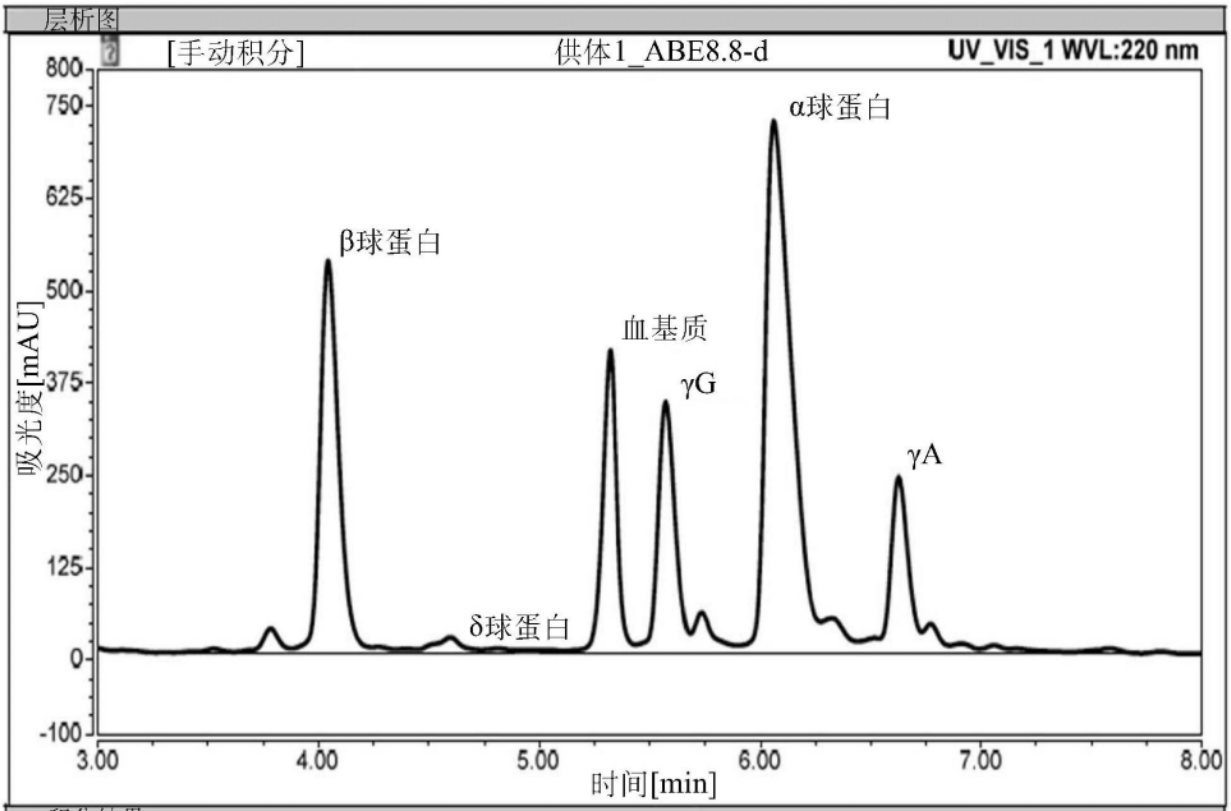
图32



积分结果

No.	波峰名称	滞留时间 min	面积 mAU*min	高度 mAU	相对面积 %	相对高度 %
5	β球蛋白	4.053	44.227	479.648	21.03	22.78
n.a.	Hbs	n.a.	n.a.	n.a.	n.a.	n.a.
9	δ球蛋白	4.607	1.704	17.056	0.81	0.81
13	血基质	5.310	24.221	326.068	11.52	15.48
14	γG	5.590	21.346	256.462	10.15	12.18
16	α球蛋白	6.087	79.706	602.792	37.90	28.62
19	γA	6.640	16.631	192.016	7.91	9.12

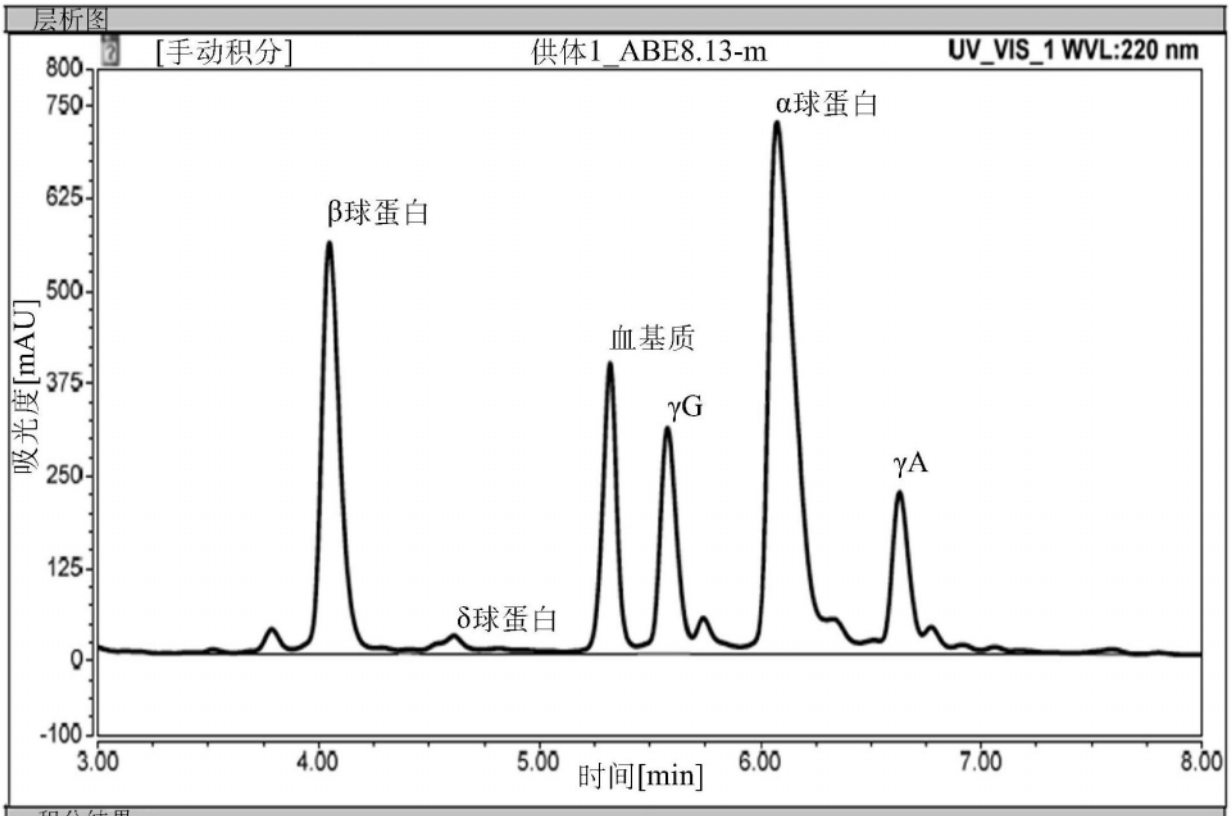
图33



积分结果

No.	波峰名称	滞留时间 min	面积 mAU*min	高度 mAU	相对面积 %	相对高度 %
5	β球蛋白	4.043	51.802	530.669	19.60	20.80
n.a.	Hbs	n.a.	n.a.	n.a.	n.a.	n.a.
9	δ球蛋白	4.597	2.269	21.040	0.86	0.82
13	血基质	5.323	30.637	410.481	11.59	16.09
14	γG	5.573	30.747	340.883	11.63	13.36
16	α球蛋白	6.060	98.417	721.350	37.24	28.28
19	γA	6.627	21.595	239.647	8.17	9.39

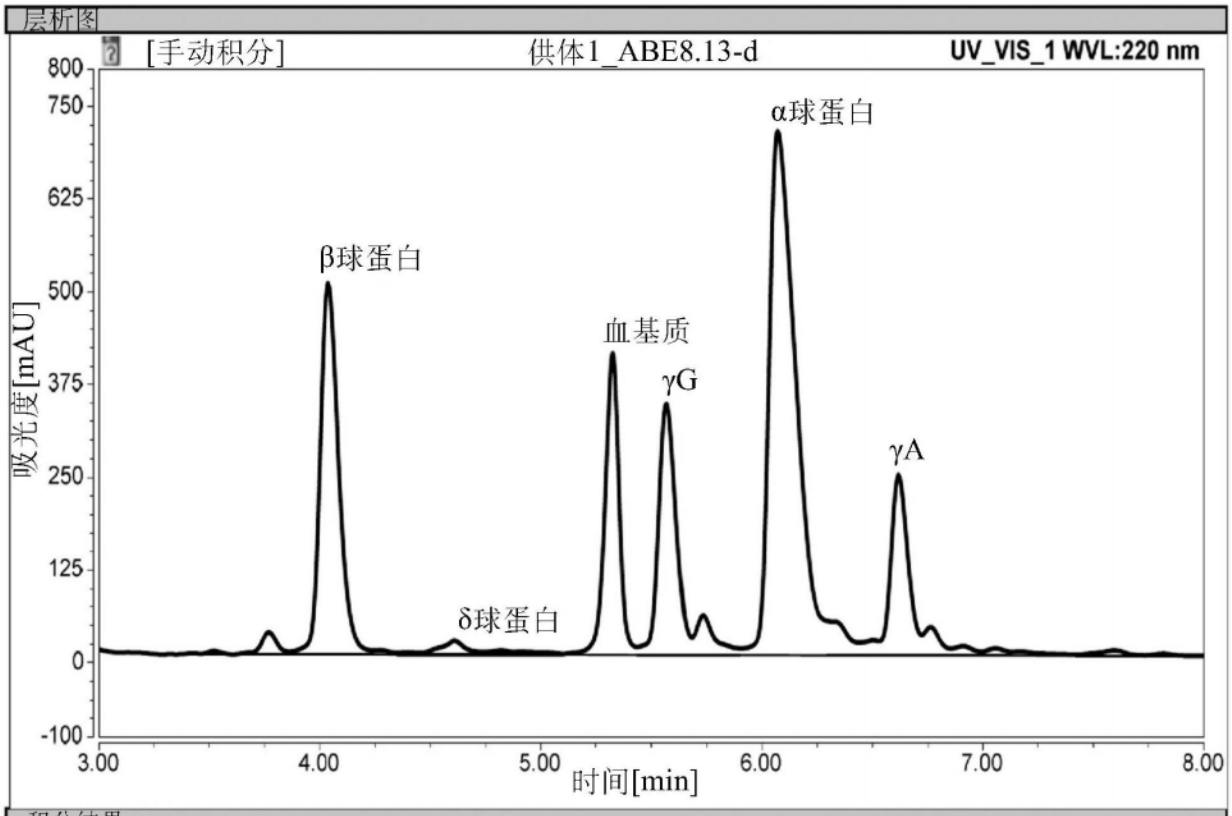
图34



积分结果

No.	波峰名称	滞留时间 min	面积 mAU*min	高度 mAU	相对面积 %	相对高度 %
5	β球蛋白	4.047	55.084	554.598	21.23	22.22
n.a.	Hbs	n.a.	n.a.	n.a.	n.a.	n.a.
9	δ球蛋白	4.613	2.417	24.372	0.93	0.98
13	血基质	5.320	29.500	393.066	11.37	15.75
14	γG	5.580	27.177	306.655	10.48	12.29
16	α球蛋白	6.073	98.048	719.982	37.79	28.85
19	γA	6.630	19.949	219.869	7.69	8.81

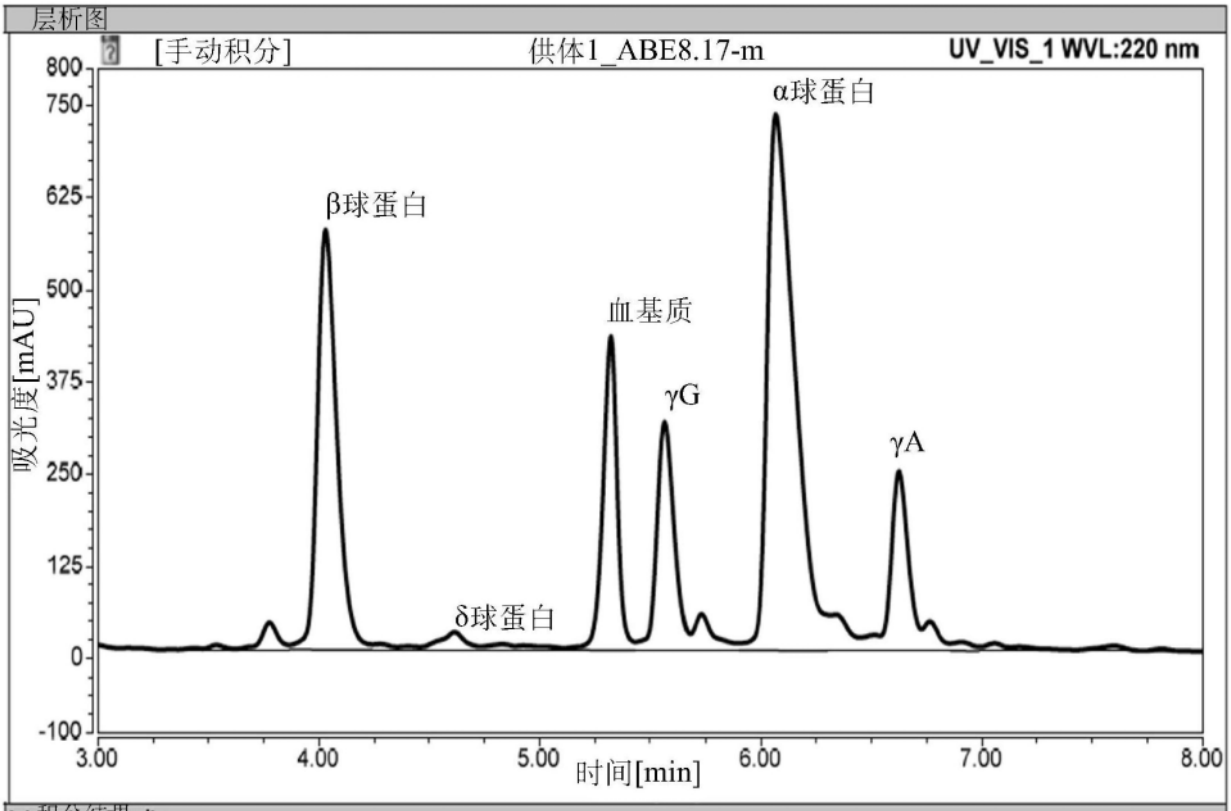
图35



积分结果

No.	波峰名称	滞留时间 min	面积 mAU*min	高度 mAU	相对面积 %	相对高度 %
5	β球蛋白	4.037	49.284	501.306	19.04	20.14
n.a.	Hbs	n.a.	n.a.	n.a.	n.a.	n.a.
9	δ球蛋白	4.610	1.960	18.855	0.76	0.76
13	血基质	5.327	30.274	407.530	11.69	16.37
14	γG	5.570	31.059	339.163	12.00	13.63
16	α球蛋白	6.070	98.109	707.464	37.89	28.42
19	γA	6.617	21.869	243.715	8.45	9.79

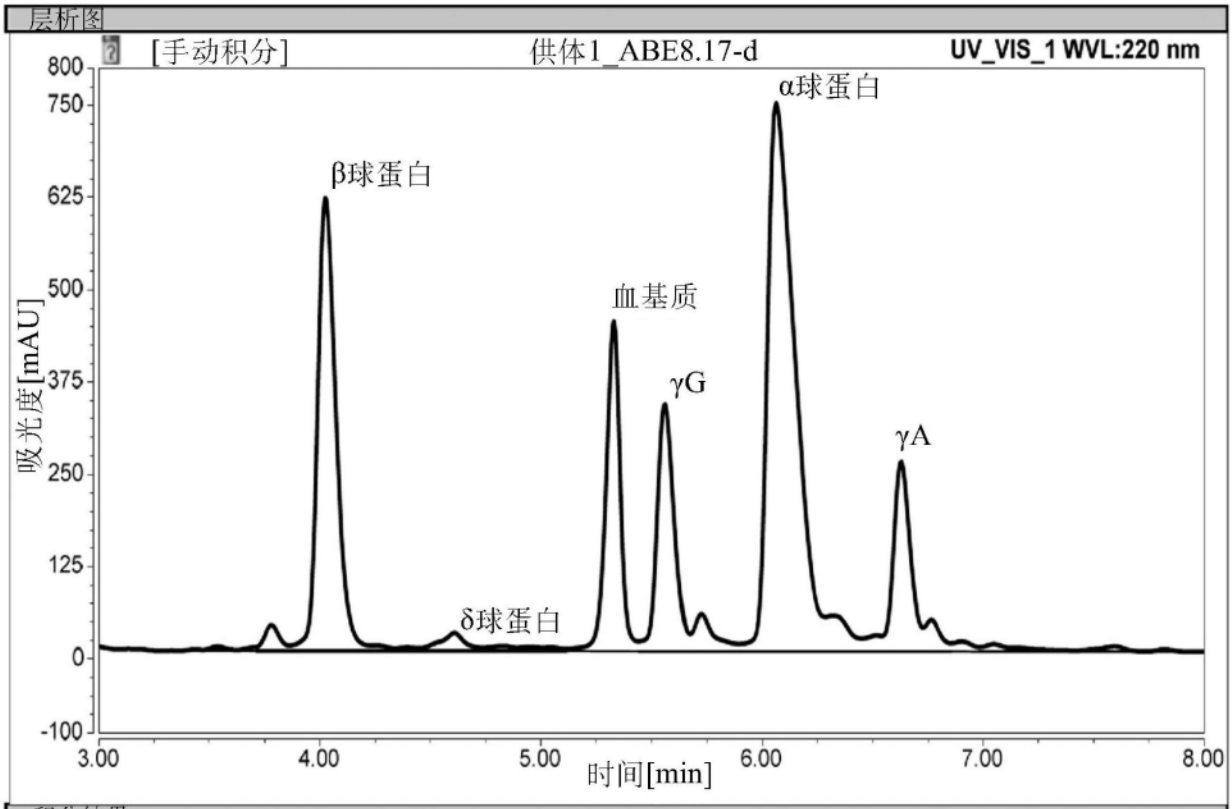
图36



Integration Results 积分结果

No.	波峰名称	滞留时间 min	面积 mAU*min	高度 mAU	相对面积 %	相对高度 %
5	β球蛋白	4.030	59.000	569.475	21.35	21.91
n.a.	Hbs	n.a.	n.a.	n.a.	n.a.	n.a.
9	δ球蛋白	4.617	2.483	24.660	0.90	0.95
13	血基质	5.323	31.792	426.360	11.50	16.40
14	γG	5.567	28.258	310.274	10.22	11.94
16	α球蛋白	6.067	104.586	728.120	37.84	28.01
19	γA	6.627	21.512	243.710	7.78	9.37

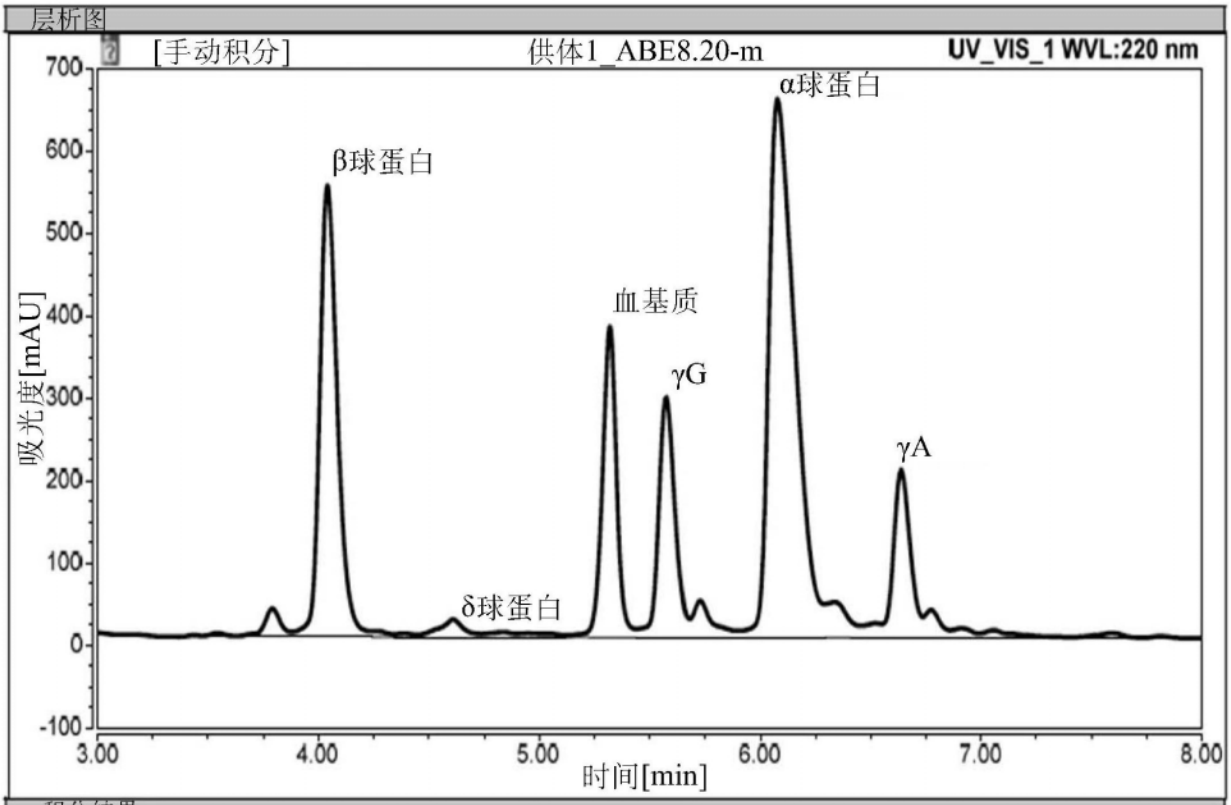
图37



积分结果

No.	波峰名称	滞留时间 min	面积 mAU*min	高度 mAU	相对面积 %	相对高度 %
6	β球蛋白	4.023	60.046	613.639	21.18	22.63
n.a.	Hbs	n.a.	n.a.	n.a.	n.a.	n.a.
10	δ球蛋白	4.610	2.491	24.450	0.88	0.90
14	血基质	5.330	33.236	446.899	11.72	16.48
15	γG	5.560	29.840	335.709	10.52	12.38
17	α球蛋白	6.063	106.077	742.712	37.41	27.38
20	γA	6.627	22.570	257.472	7.96	9.49

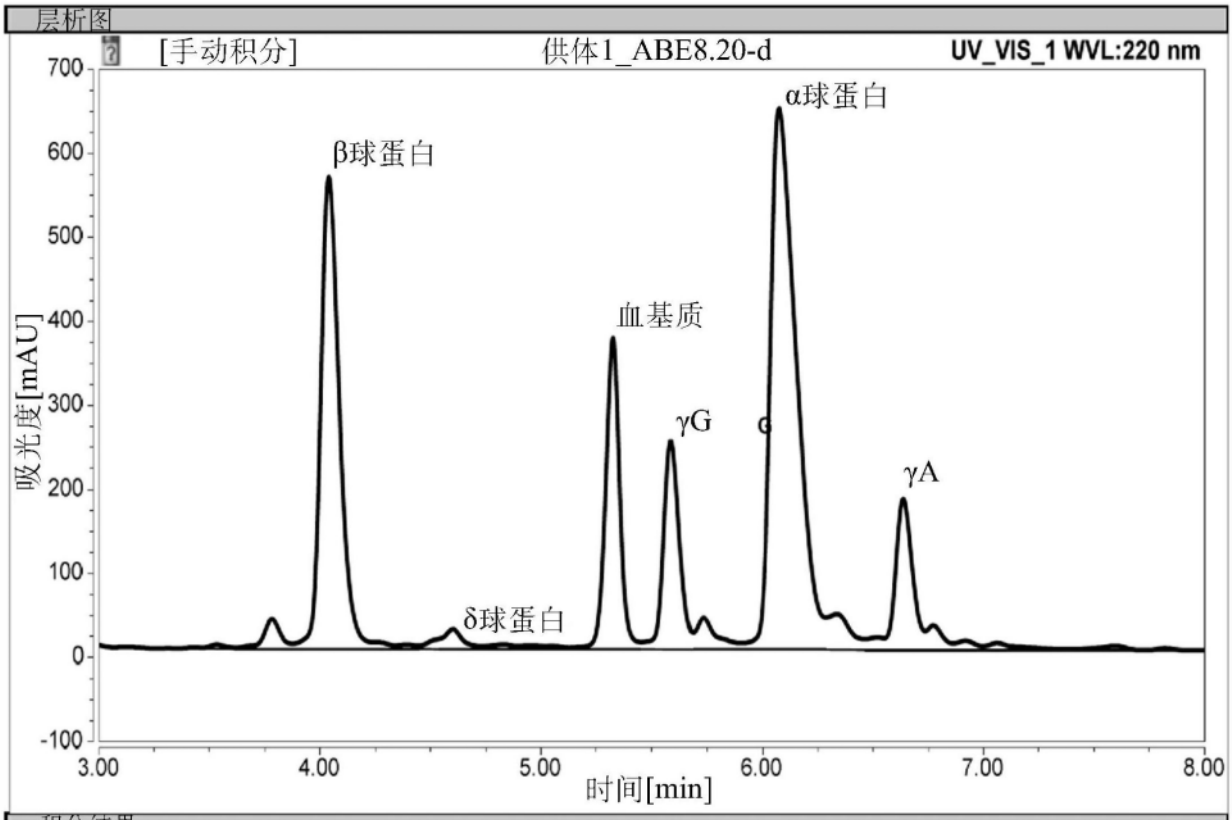
图38



积分结果

No.	波峰名称	滞留时间 min	面积 mAU*min	高度 mAU	相对面积 %	相对高度 %
5	β球蛋白	4.040	51.610	548.050	21.57	23.31
n.a.	Hbs	n.a.	n.a.	n.a.	n.a.	n.a.
9	δ球蛋白	4.610	2.233	21.329	0.93	0.91
13	血基质	5.320	27.916	377.668	11.67	16.06
14	γG	5.573	24.931	291.767	10.42	12.41
16	α球蛋白	6.077	90.099	653.942	37.66	27.81
19	γA	6.637	17.798	203.835	7.44	8.67

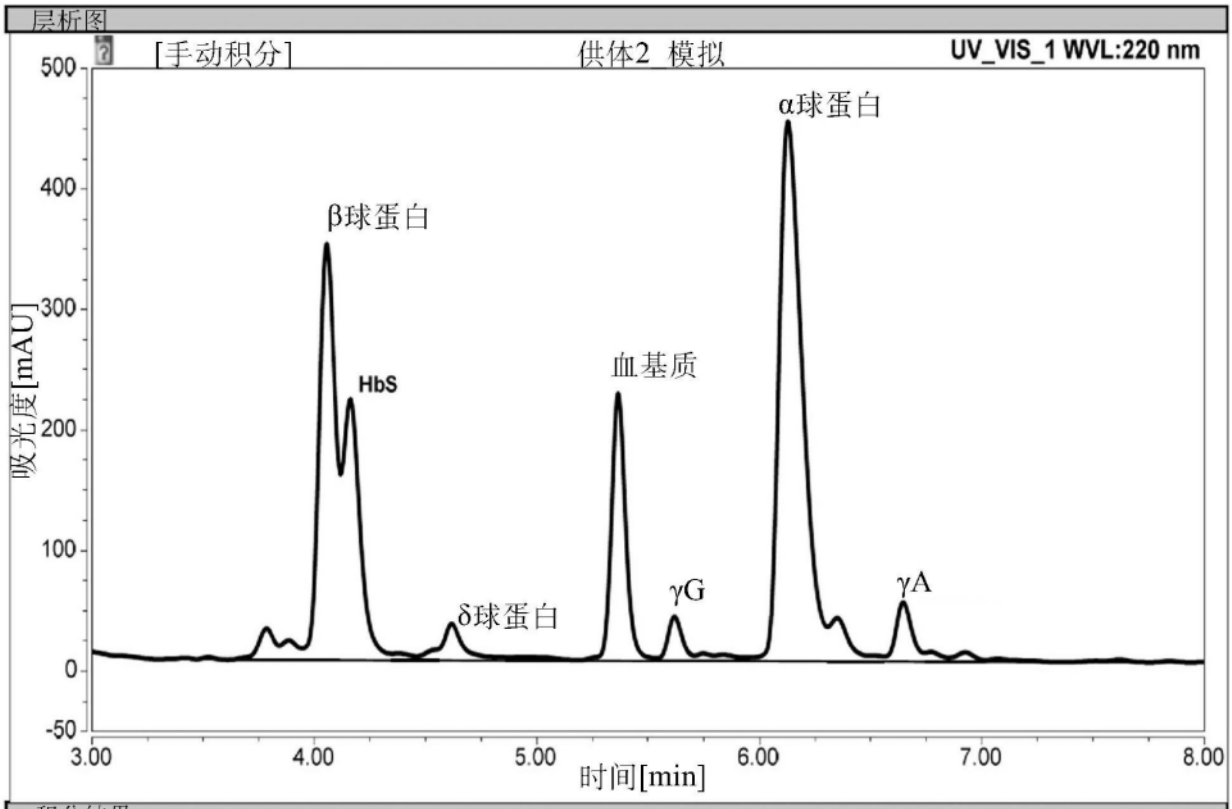
图39



积分结果

No.	波峰名称	滞留时间 min	面积 mAU*min	高度 mAU	相对面积 %	相对高度 %
6	β球蛋白	4.040	55.490	562.369	23.82	24.71
n.a.	Hbs	n.a.	n.a.	n.a.	n.a.	n.a.
10	δ球蛋白	4.600	2.399	23.713	1.03	1.04
14	血基质	5.327	27.308	371.229	11.72	16.31
15	γG	5.587	20.808	248.715	8.93	10.93
17	α球蛋白	6.073	87.517	645.669	37.56	28.38
20	γA	6.637	15.700	180.637	6.74	7.94

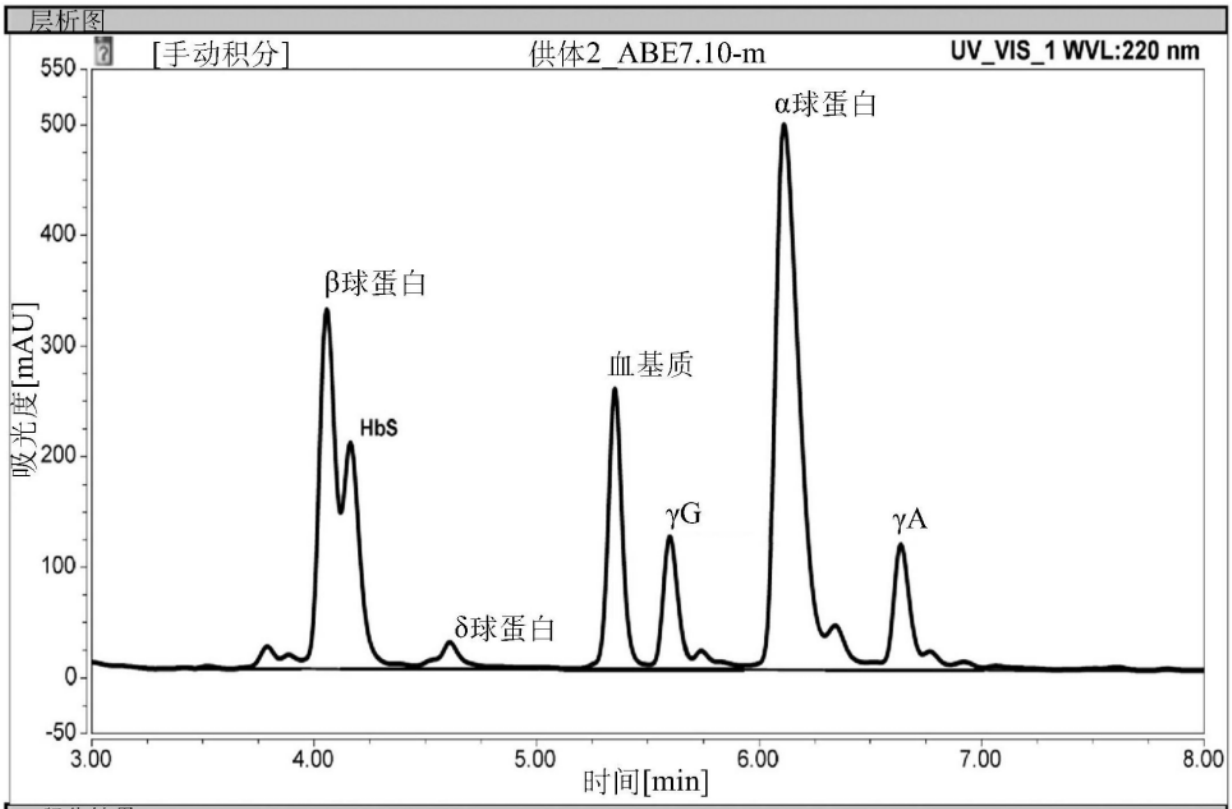
图40



积分结果

No.	波峰名称	滞留时间 min	面积 mAU*min	高度 mAU	相对面积 %	相对高度 %
5	β球蛋白	4.057	28.943	345.410	20.25	23.19
6	Hbs	4.160	19.794	216.879	13.85	14.56
9	δ球蛋白	4.617	3.170	30.593	2.22	2.05
11	血基质	5.367	16.498	222.155	11.54	14.91
12	γG	5.620	2.805	37.002	1.96	2.48
15	α球蛋白	6.127	54.386	448.385	38.05	30.10
18	γA	6.647	4.170	49.520	2.92	3.32

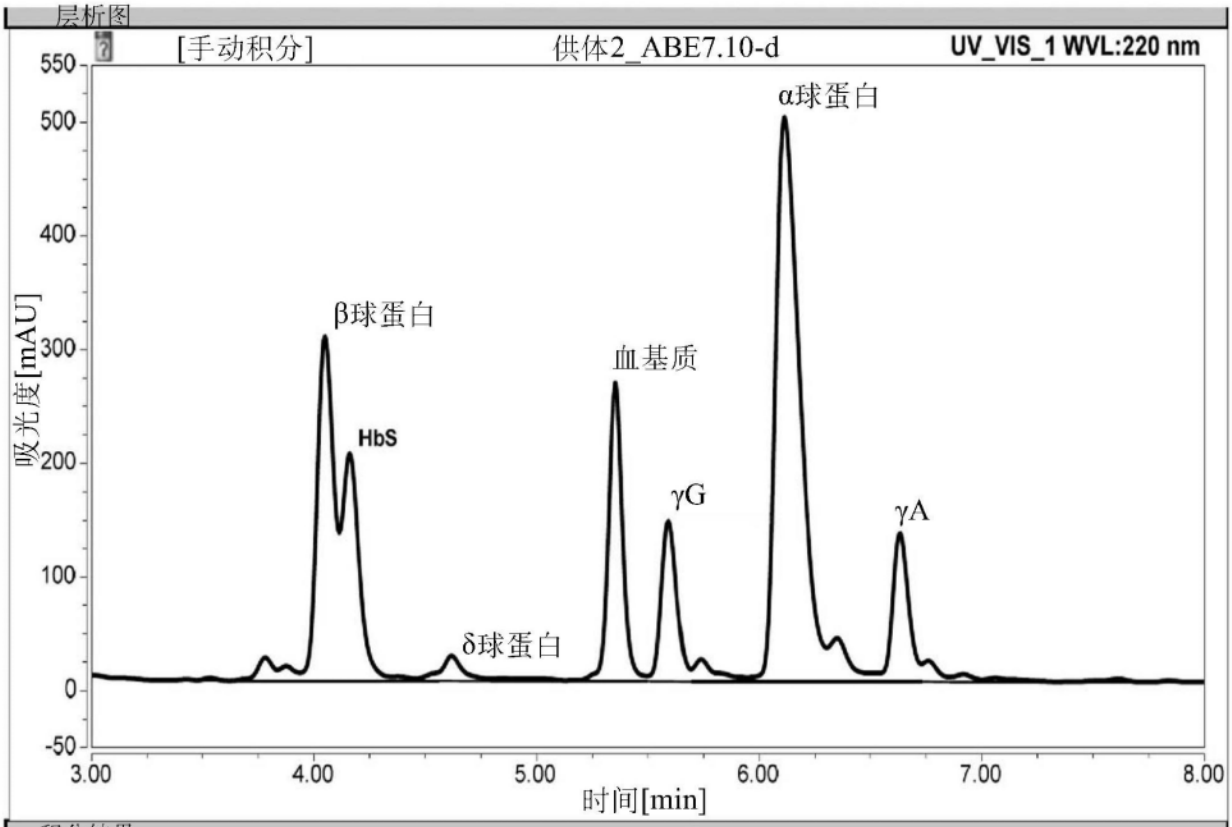
图41



积分结果

No.	波峰名称	滞留时间 min	面积 mAU*min	高度 mAU	相对面积 %	相对高度 %
5	β球蛋白	4.057	27.210	325.305	16.96	19.17
6	Hbs	4.163	18.038	204.594	11.24	12.05
9	δ球蛋白	4.610	2.458	24.147	1.53	1.42
14	血基质	5.353	18.788	253.663	11.71	14.94
15	γG	5.600	9.475	120.558	5.91	7.10
18	α球蛋白	6.110	60.688	493.180	37.83	29.06
21	γA	6.637	9.526	113.714	5.94	6.70

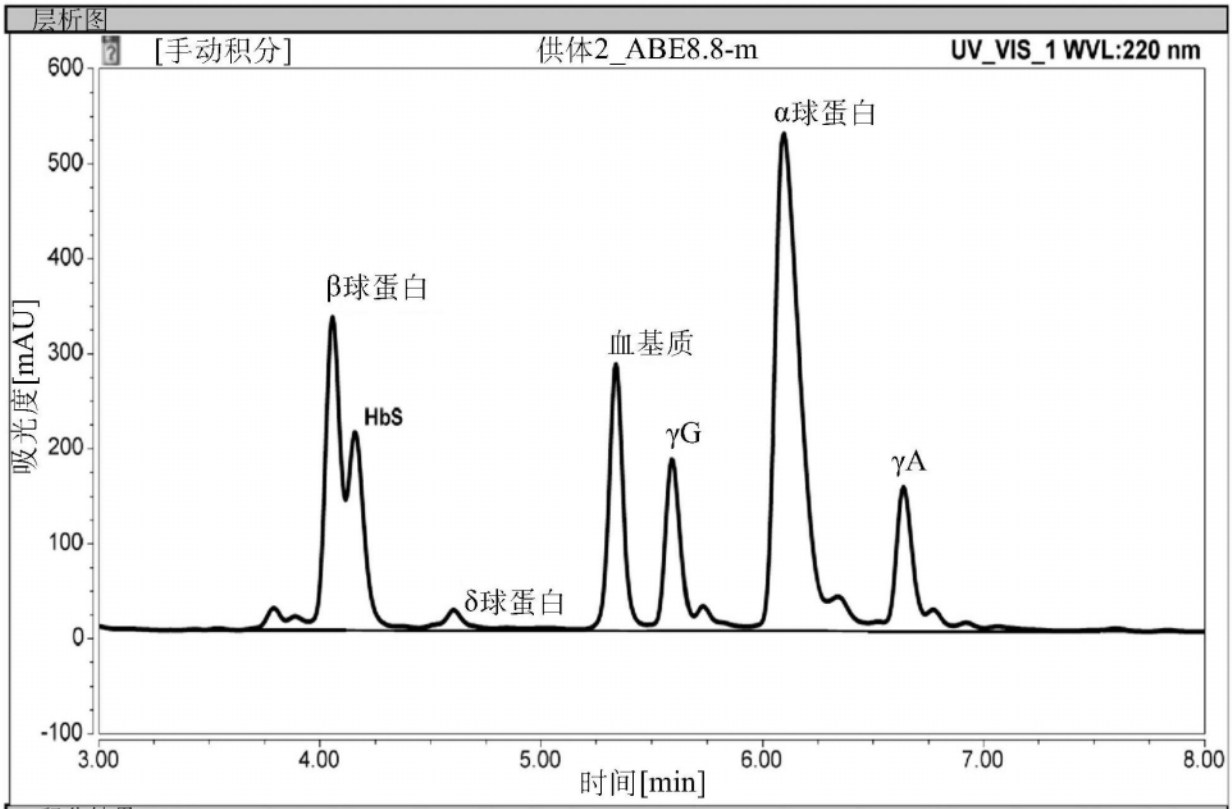
图42



积分结果

No.	波峰名称	滞留时间 min	面积 mAU*min	高度 mAU	相对面积 %	相对高度 %
5	β球蛋白	4.050	26.525	303.461	15.99	17.61
6	Hbs	4.160	17.741	200.547	10.70	11.64
9	δ球蛋白	4.620	2.209	22.450	1.33	1.30
14	血基质	5.357	19.493	262.669	11.75	15.24
15	γG	5.593	11.528	141.729	6.95	8.23
18	α球蛋白	6.113	62.859	497.019	37.90	28.85
21	γA	6.630	10.827	131.062	6.53	7.61

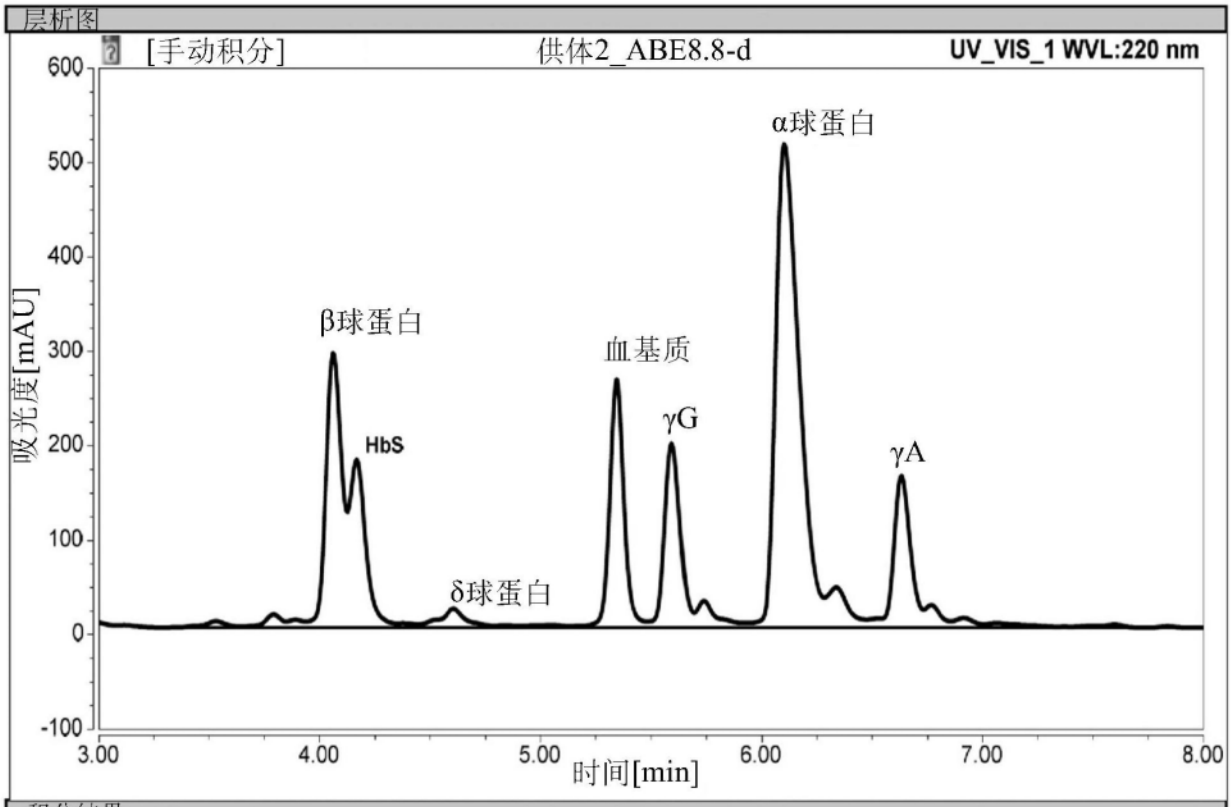
图43



积分结果

No.	波峰名称	滞留时间 min	面积 mAU*min	高度 mAU	相对面积 %	相对高度 %
5	β球蛋白	4.057	26.492	330.353	14.76	17.48
6	Hbs	4.160	17.650	209.154	9.84	11.07
9	δ球蛋白	4.603	2.190	21.674	1.22	1.15
14	血基质	5.340	20.930	280.930	11.66	14.87
15	γG	5.593	14.357	180.573	8.00	9.56
18	α球蛋白	6.097	67.616	524.013	37.68	27.73
21	γA	6.637	12.935	152.286	7.21	8.06

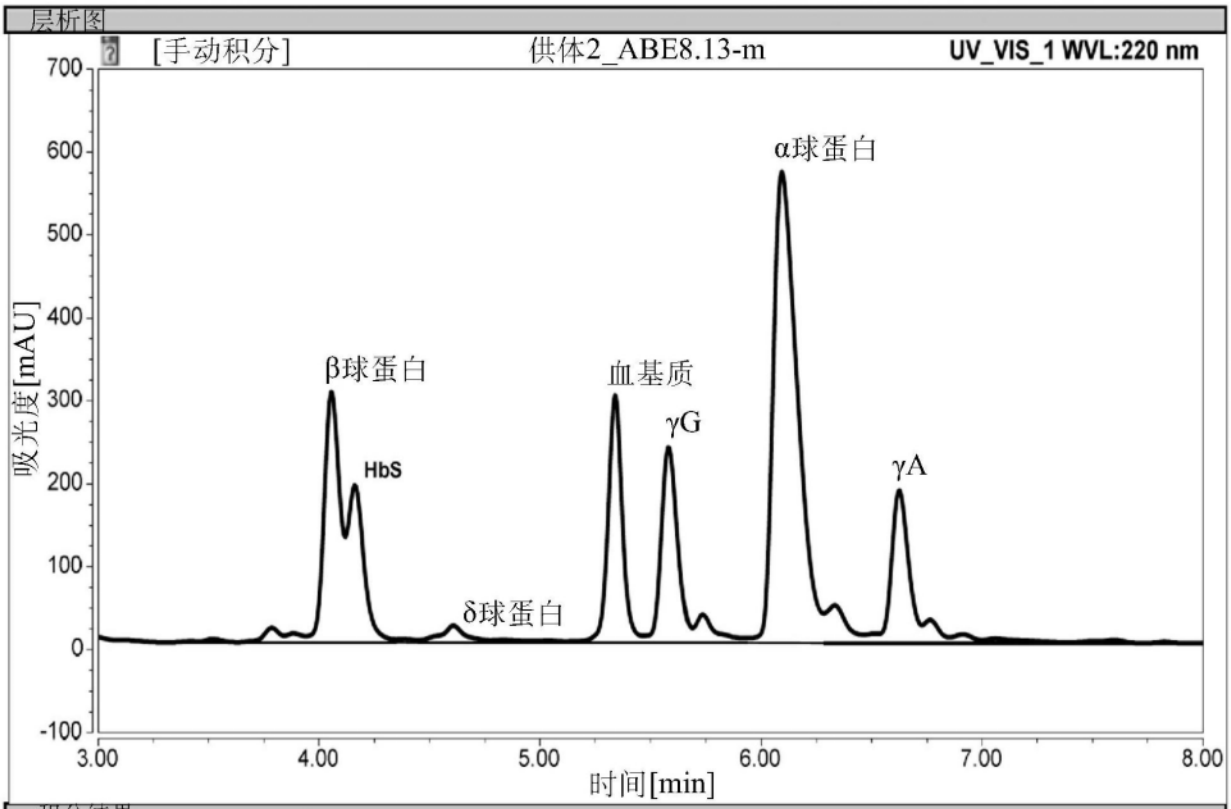
图44



积分结果

No.	波峰名称	滞留时间 min	面积 mAU*min	高度 mAU	相对面积 %	相对高度 %
6	β球蛋白	4.063	23.888	291.371	13.90	16.06
7	Hbs	4.167	15.512	178.139	9.02	9.82
10	δ球蛋白	4.607	2.249	20.296	1.31	1.12
15	血基质	5.343	19.534	262.940	11.36	14.49
16	γG	5.593	16.034	195.308	9.33	10.77
19	α球蛋白	6.103	63.055	512.839	36.68	28.27
22	γA	6.633	13.860	161.449	8.06	8.90

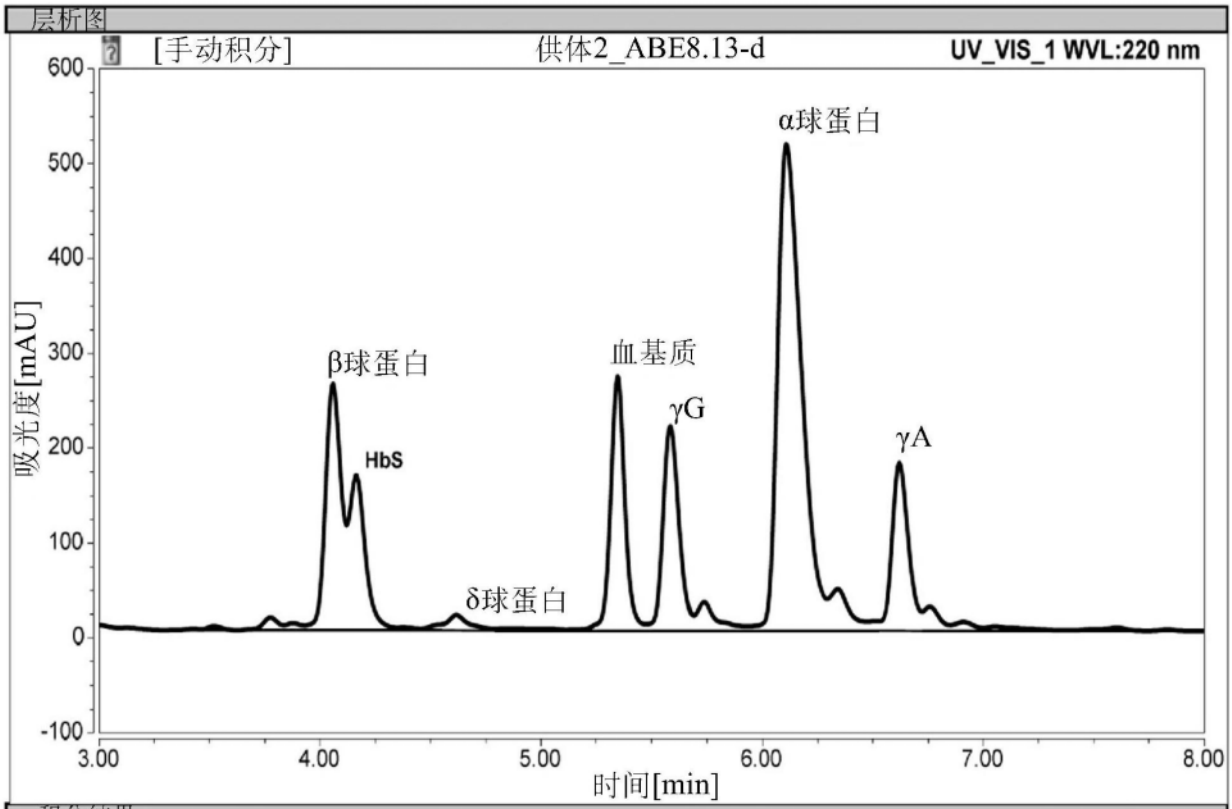
图45



积分结果

No.	波峰名称	滞留时间 min	面积 mAU*min	高度 mAU	相对面积 %	相对高度 %
6	β球蛋白	4.057	24.649	303.680	12.77	15.05
7	Hbs	4.163	16.705	190.174	8.65	9.43
10	δ球蛋白	4.607	2.114	20.478	1.10	1.01
15	血基质	5.340	22.158	298.385	11.48	14.79
16	γG	5.583	20.107	236.242	10.41	11.71
19	α球蛋白	6.093	71.587	568.640	37.08	28.18
22	γA	6.623	16.259	184.774	8.42	9.16

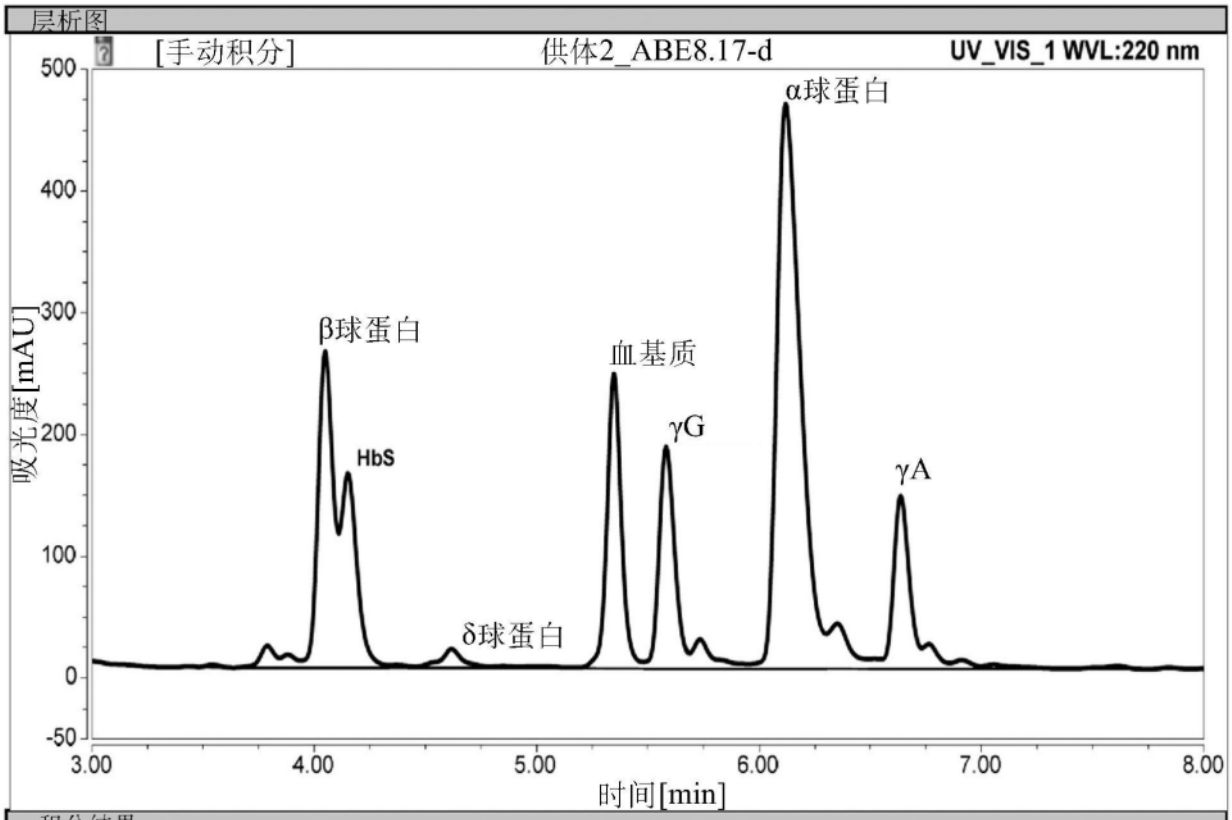
图46



积分结果

No.	波峰名称	滞留时间 min	面积 mAU*min	高度 mAU	相对面积 %	相对高度 %
5	β球蛋白	4.057	21.912	260.274	12.79	14.49
6	Hbs	4.163	13.772	163.531	8.04	9.10
9	δ球蛋白	4.617	1.827	16.318	1.07	0.91
13	血基质	5.347	19.932	268.078	11.64	14.92
14	γG	5.587	18.026	215.798	10.52	12.01
17	α球蛋白	6.110	64.064	513.833	37.40	28.61
20	γA	6.620	15.335	177.074	8.95	9.86

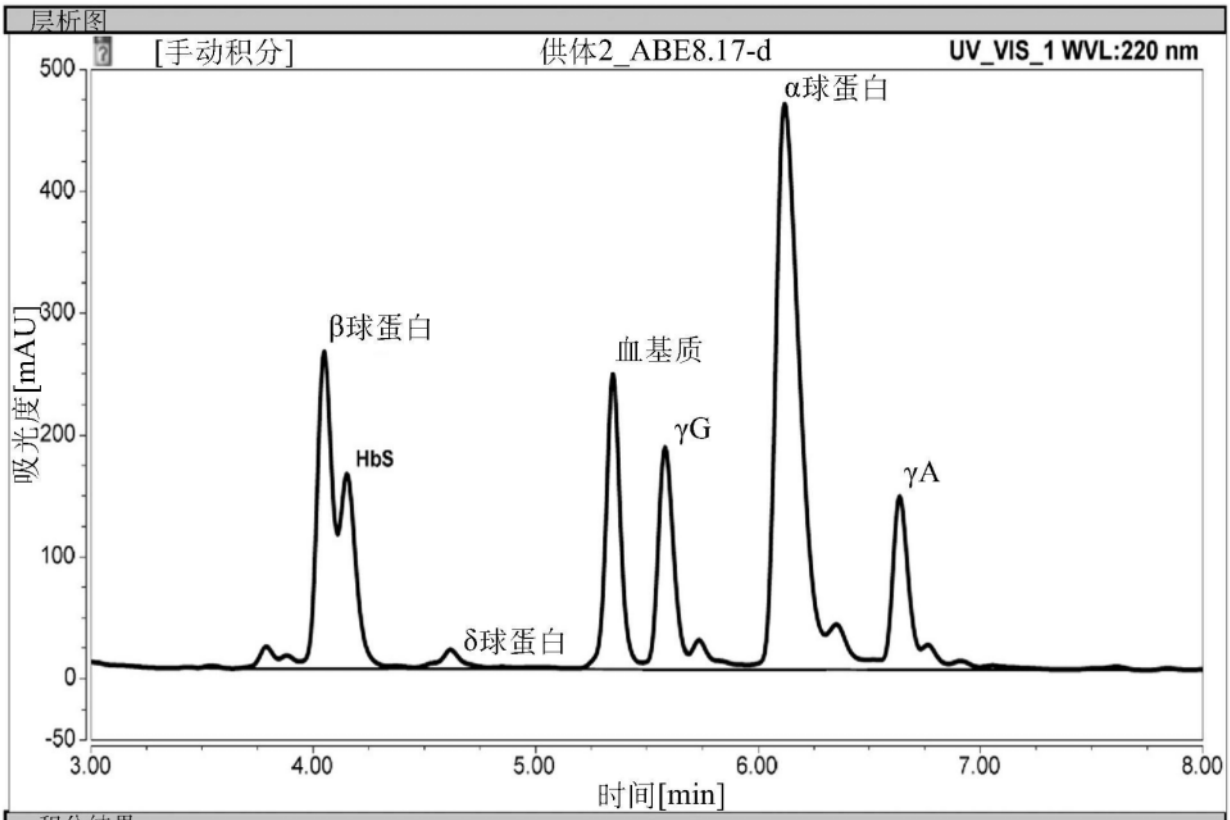
图47



积分结果

No.	波峰名称	滞留时间 min	面积 mAU*min	高度 mAU	相对面积 %	相对高度 %
5	β球蛋白	4.050	20.688	260.391	13.65	15.97
6	Hbs	4.150	13.530	159.994	8.93	9.81
9	δ球蛋白	4.617	1.590	15.794	1.05	0.97
14	血基质	5.350	18.069	242.095	11.92	14.85
15	γG	5.583	14.825	182.475	9.78	11.19
18	α球蛋白	6.117	57.055	464.210	37.65	28.48
21	γA	6.637	11.691	142.851	7.72	8.76

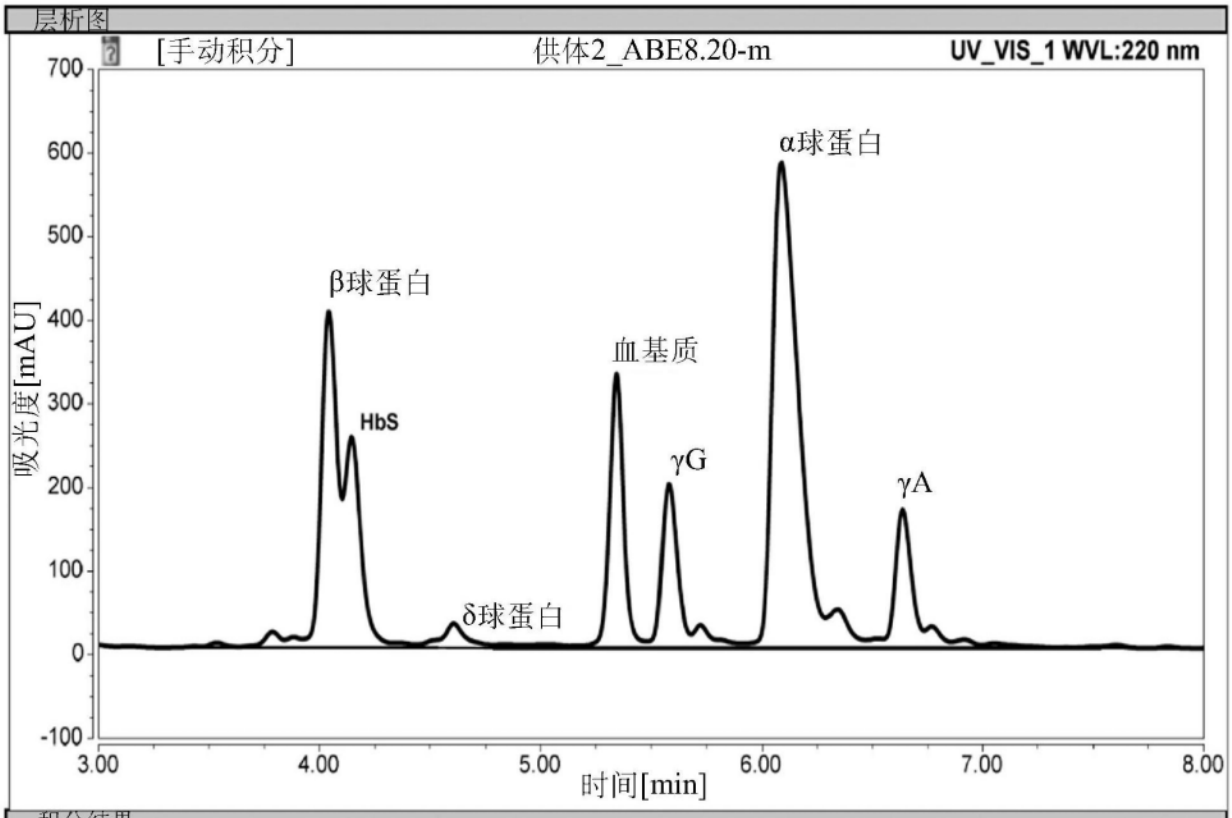
图48



积分结果

No.	波峰名称	滞留时间 min	面积 mAU*min	高度 mAU	相对面积 %	相对高度 %
5	β球蛋白	4.050	20.688	260.391	13.65	15.97
6	Hbs	4.150	13.530	159.994	8.93	9.81
9	δ球蛋白	4.617	1.590	15.794	1.05	0.97
14	血基质	5.350	18.069	242.095	11.92	14.85
15	γG	5.583	14.825	182.475	9.78	11.19
18	α球蛋白	6.117	57.055	464.210	37.65	28.48
21	γA	6.637	11.691	142.851	7.72	8.76

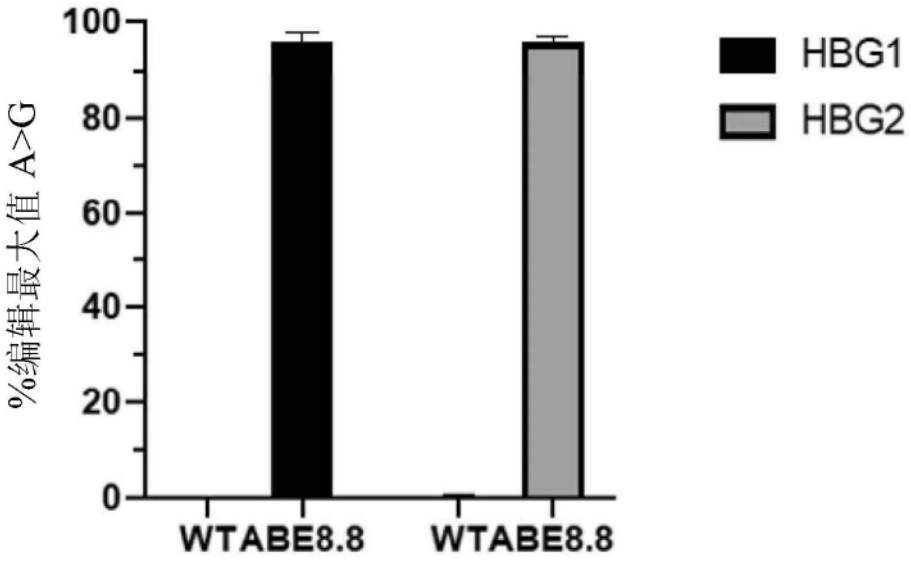
图49



积分结果

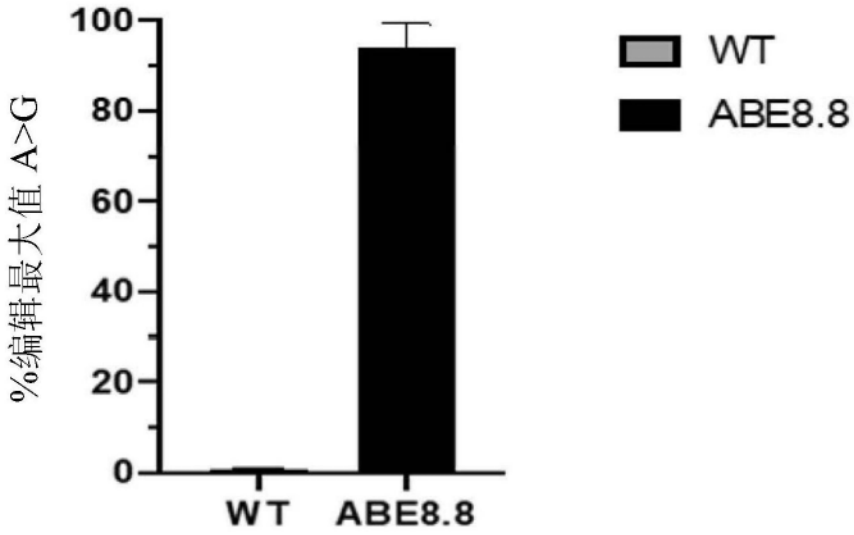
No.	波峰名称	滞留时间 min	面积 mAU*min	高度 mAU	相对面积 %	相对高度 %
6	β球蛋白	4.040	33.339	402.775	15.75	18.52
7	Hbs	4.143	22.030	252.466	10.41	11.61
10	δ球蛋白	4.607	3.256	29.080	1.54	1.34
14	血基质	5.343	24.399	328.265	11.53	15.10
15	γG	5.580	15.959	197.001	7.54	9.06
18	α球蛋白	6.087	78.149	581.487	36.93	26.74
21	γA	6.637	14.315	166.369	6.76	7.65

图50



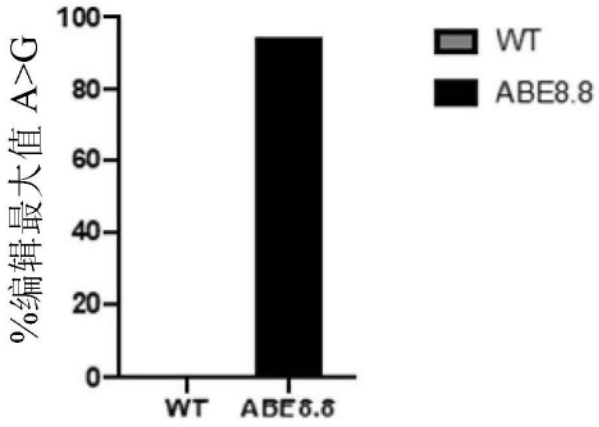
N =2健康供体，2实验

图51A



N =1健康供体，2实验

图51B



N =1健康供体，1实验
E6V杂合

图51C

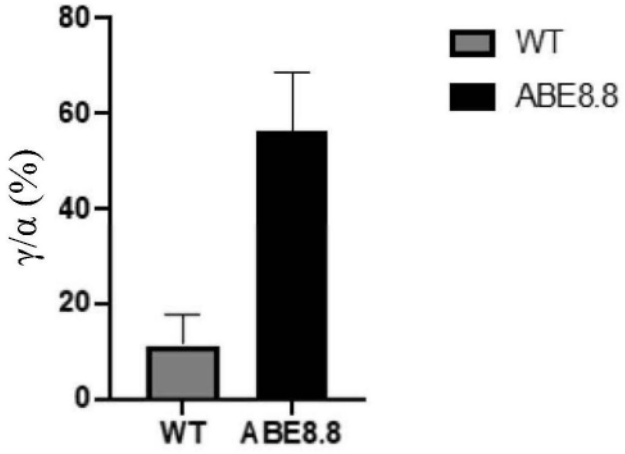


图51D

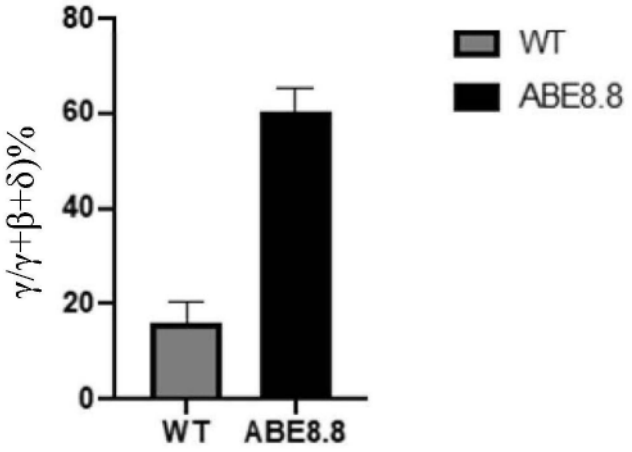

80
60
40
20
0
γ/γ+β+δ)%
WT
ABE8.8
WT
ABE8.8

图51E

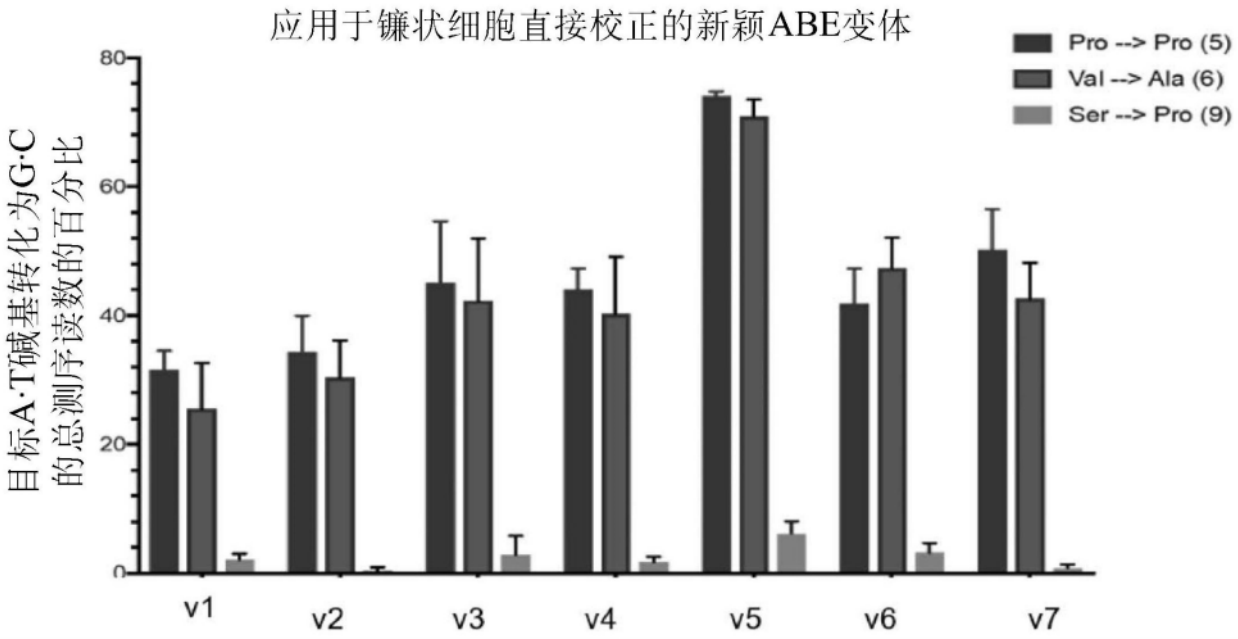

应用于镰状细胞直接校正的新颖ABE变体
目标A·T碱基转化为G·C
的总测序读数的百分比
Pro --> Pro (5)
Val --> Ala (6)
Ser --> Pro (9)
80
60
40
20
0
v1
v2
v3
v4
v5
v6
v7

图52A

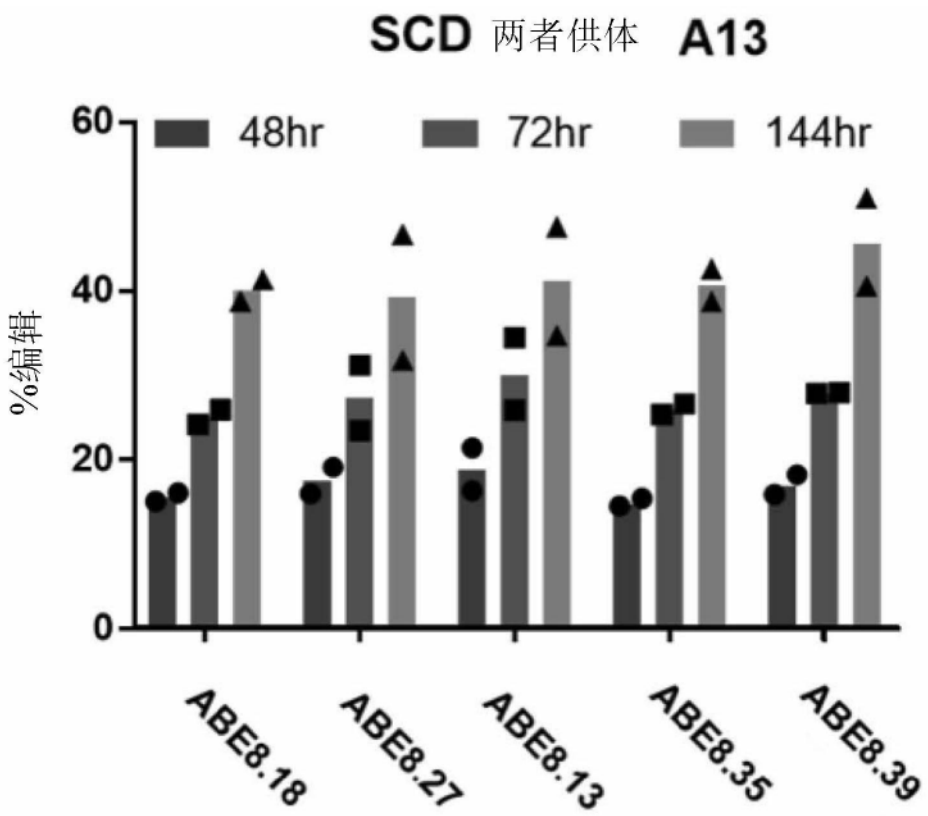

SCD 两者供体 A13
60
40
20
0
%编辑
48hr
72hr
144hr
ABE8.18
ABE8.27
ABE8.13
ABE8.35
ABE8.39

图52B

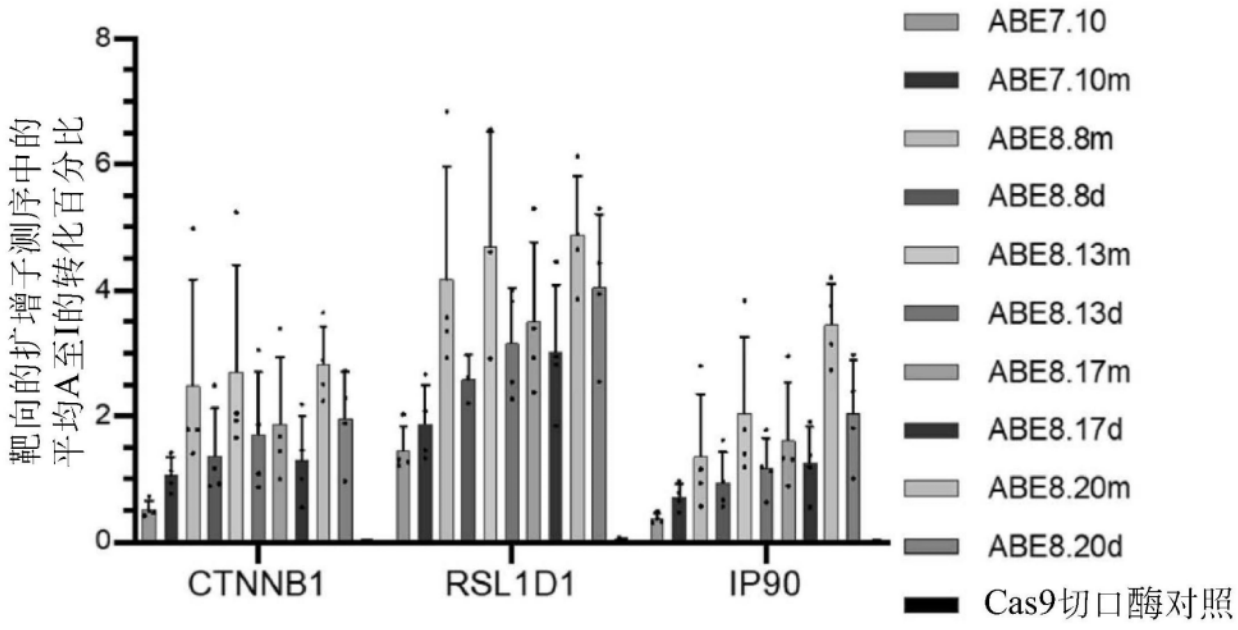

8
6
4
2
0
靶向的扩增子测序中的
平均A至I的转化百分比
CTNNB1
RSL1D1
IP90
ABE7.10
ABE7.10m
ABE8.8m
ABE8.8d
ABE8.13m
ABE8.13d
ABE8.17m
ABE8.17d
ABE8.20m
ABE8.20d
Cas9切口酶对照

图53A

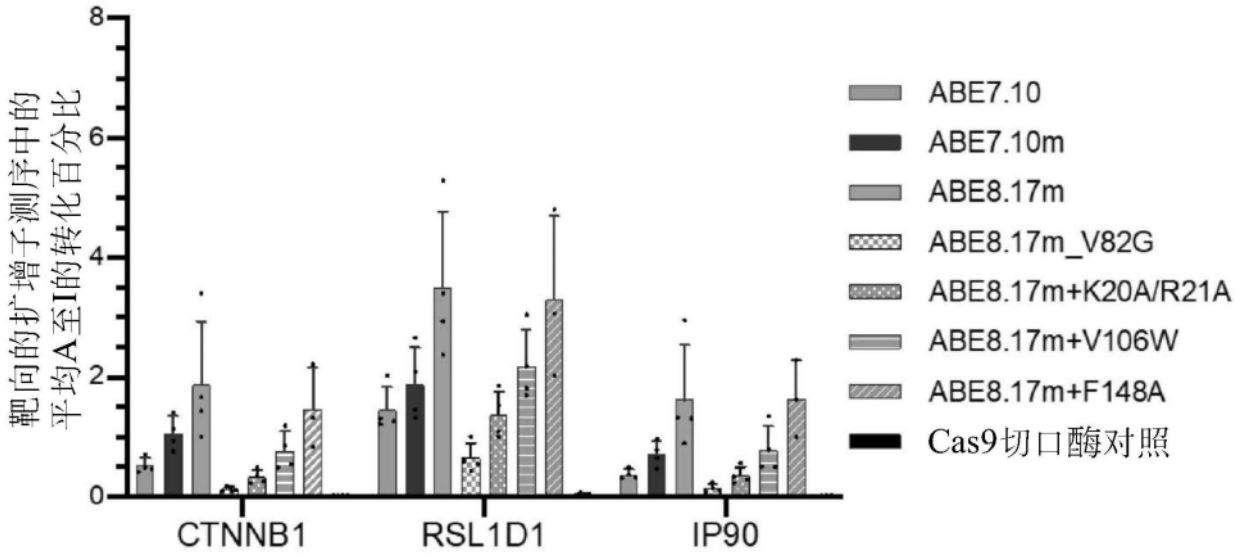

靶向的扩增子测序中的
平均A至I的转化百分比
8
6
4
2
0
CTNNB1
RSL1D1
IP90
ABE7.10
ABE7.10m
ABE8.17m
ABE8.17m_V82G
ABE8.17m+K20A/R21A
ABE8.17m+V106W
ABE8.17m+F148A
Cas9切口酶对照

图53B

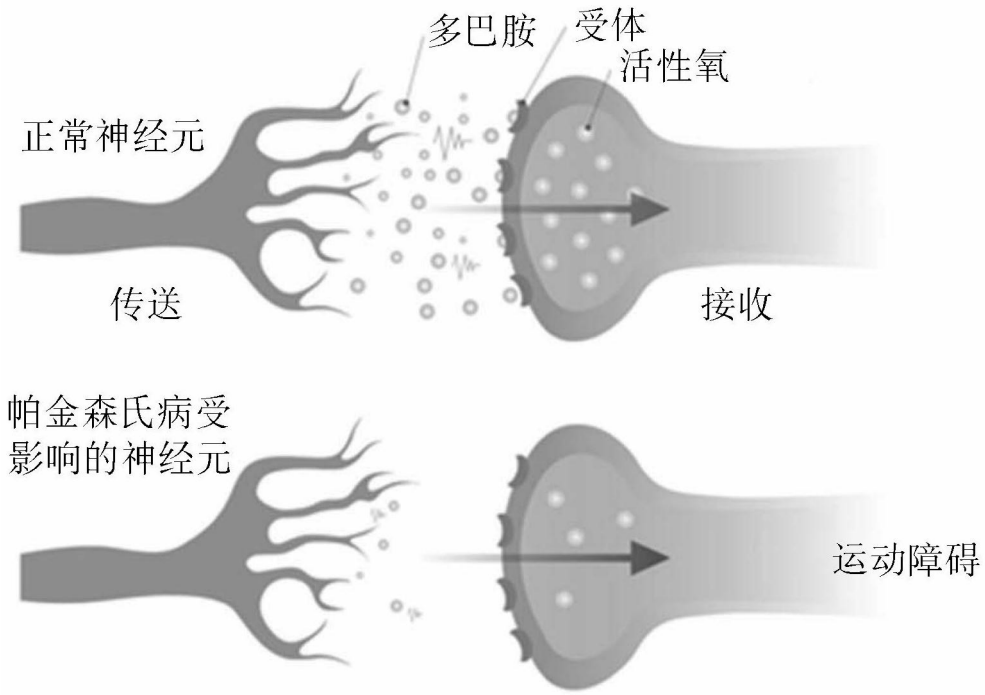

多巴胺
受体
活性氧
正常神经元
传送
接收
帕金森氏病受
影响的神经元
运动障碍

图54

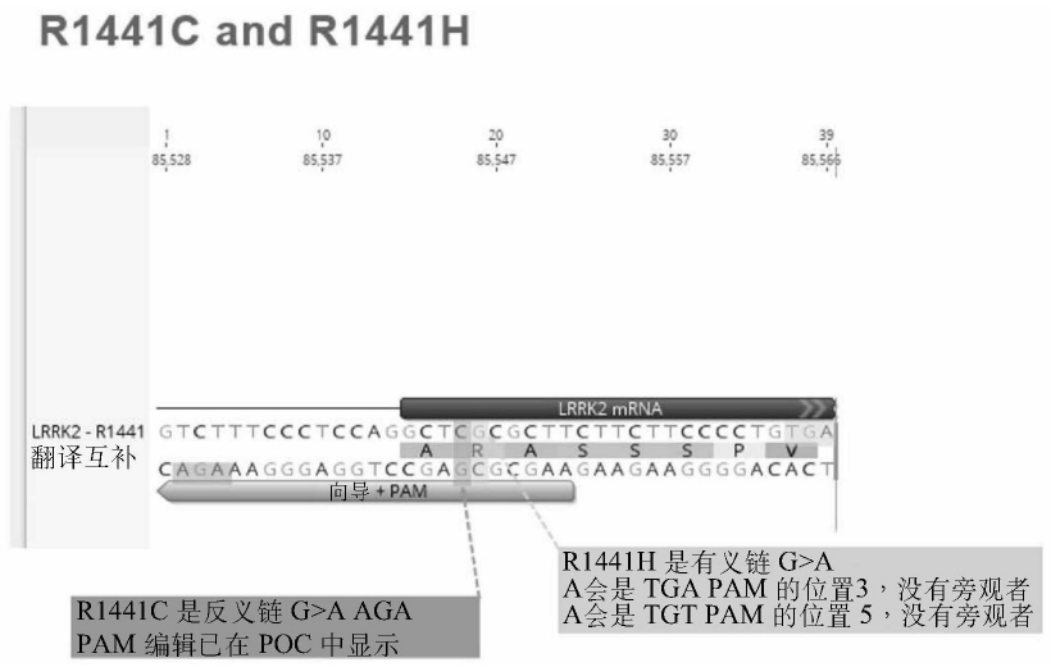

R1441C and R1441H
LRRK2 mRNA
LRRK2 - R1441
翻译互补
向导 + PAM
R1441C 是反义链 G>A AGA
PAM 编辑已在 POC 中显示
R1441H 是有义链 G>A
A会是 TGA PAM 的位置3，没有旁观者
A会是 TGT PAM 的位置 5，没有旁观者

图55

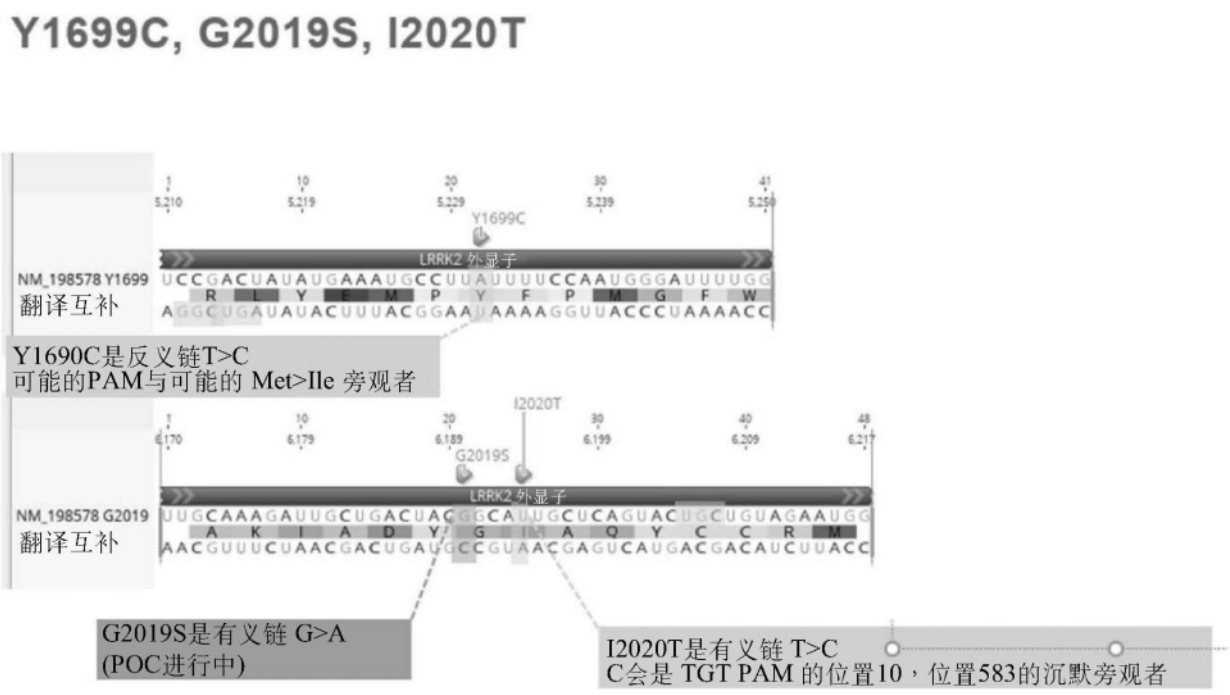

Y1699C, G2019S, I2020T
Y1699C
LRRK2 外显子
NM_198578 Y1699
翻译互补
Y1690C是反义链T>C
可能的PAM与可能的 Met>Ile 旁观者
I2020T
G2019S
LRRK2 外显子
NM_198578 G2019
翻译互补
G2019S是有义链 G>A
(POC进行中)
I2020T是有义链 T>C
C会是 TGT PAM 的位置10，位置583的沉默旁观者

图56

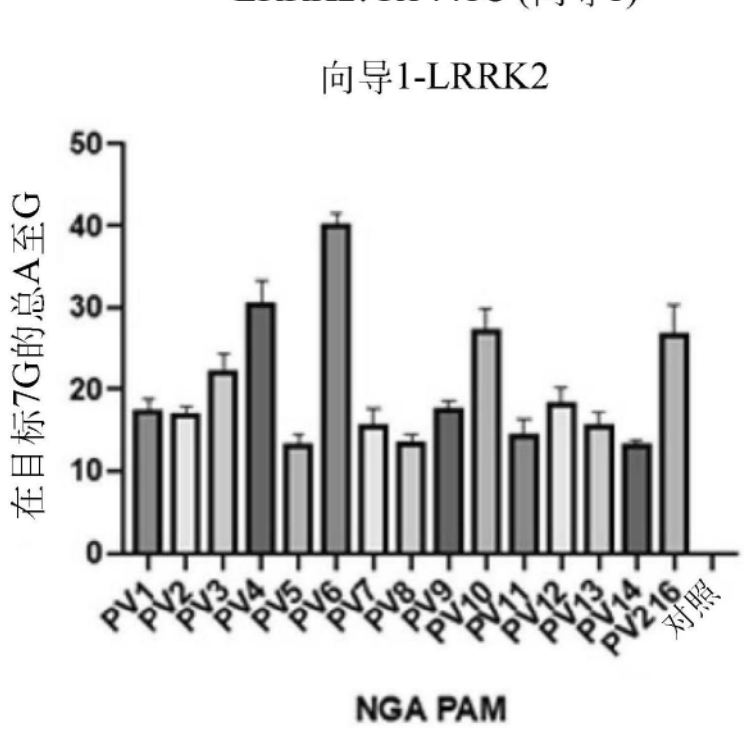

LRRK2: R1441C (向导1)
向导1-LRRK2
在目标7G的总A至G
50
40
30
20
10
0
PV1
PV2
PV3
PV4
PV5
PV6
PV7
PV8
PV9
PV10
PV11
PV12
PV13
PV14
PV216
对照
NGA PAM

图57A

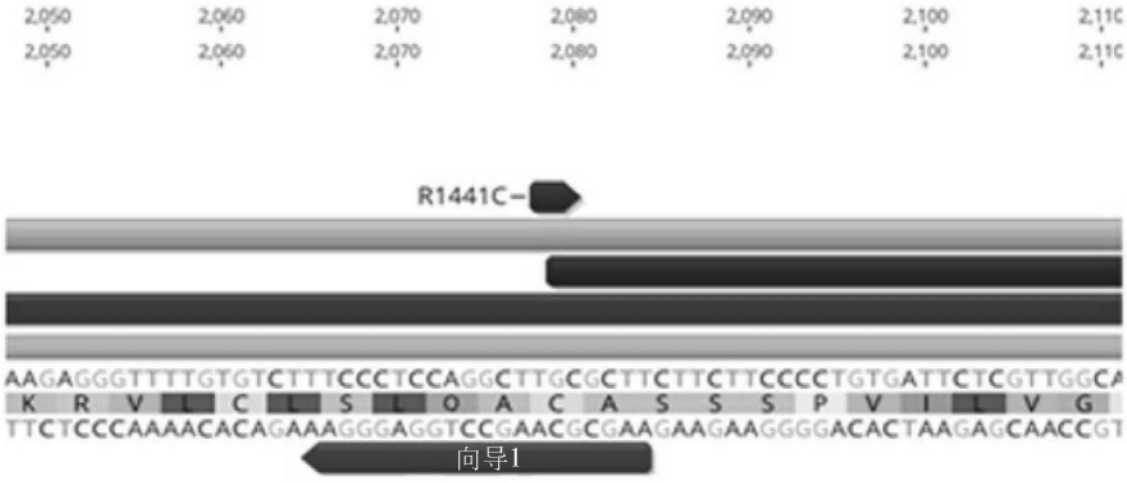

2,050
2,060
2,070
2,080
2,090
2,100
R1441C
AAGAGGGTTTTGTGTCTTTCCCTCCAGGCTTGCGCTTCTTCTTCCCCTGTGATTCTCGTTGGCA
K R V L C L S L Q A C A S S S P V I L V G
TTCTCCCAAAACACAGAAAGGGAGGTCCGAACGCGAAGAAGAAGGGGACACTAAGAGCAACCGT
向导1

图57B

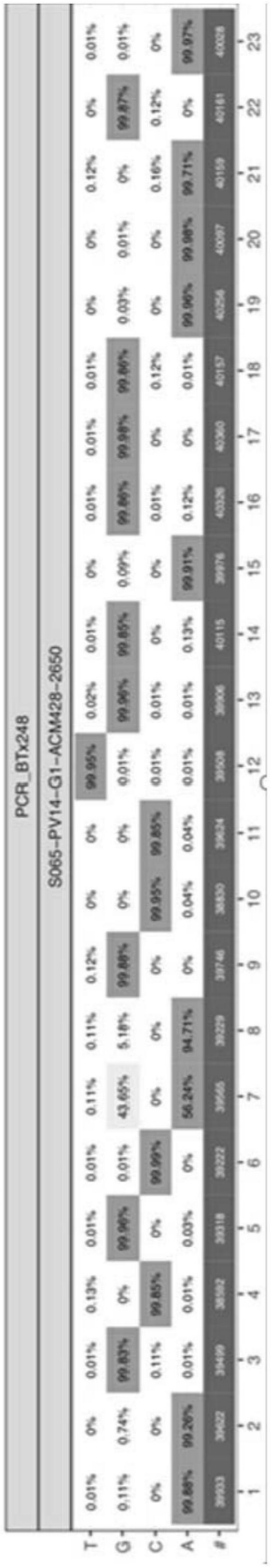

PCR_BTx248
S065-PV14-G1-ACM428-2650
T
G
C
A
#
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

图57C

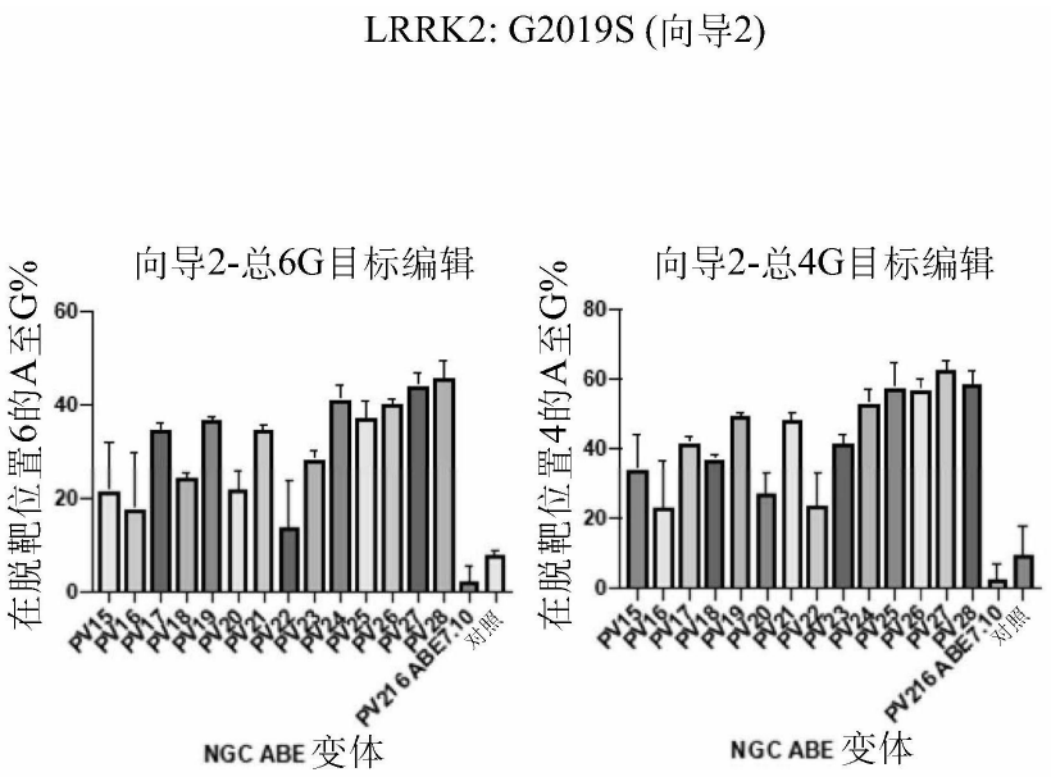


图58A

- 第一点:具有“向导2”的目标G2019S
- 在位置4的旁观者造成酪氨酸至半胱氨酸的转化

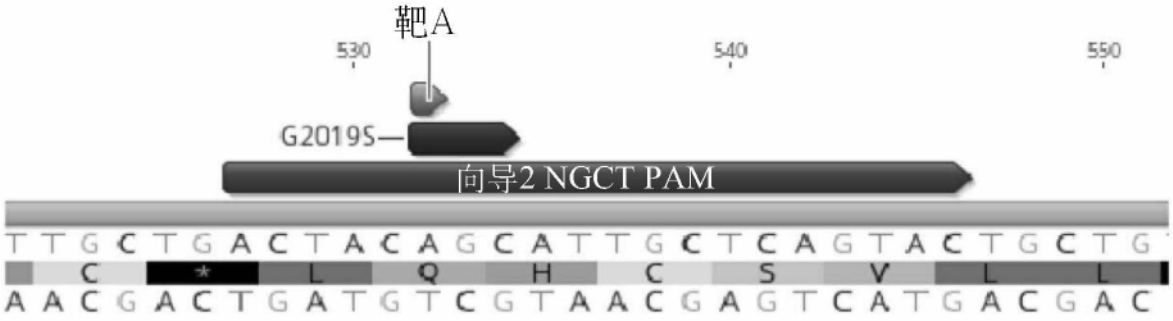


图58B

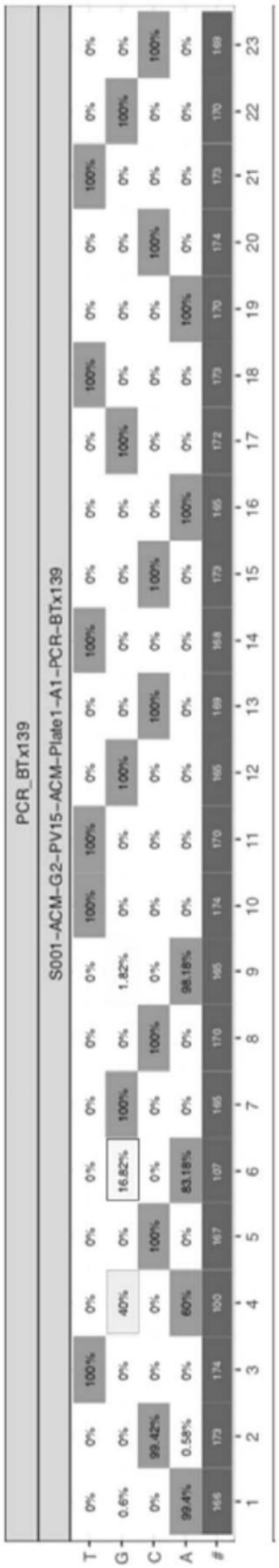

PCR_BTx139
S001-ACM-G2-PV15-ACM-Plate1-A1-PCR-BTx139
T
G
C
A
#
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

图58C

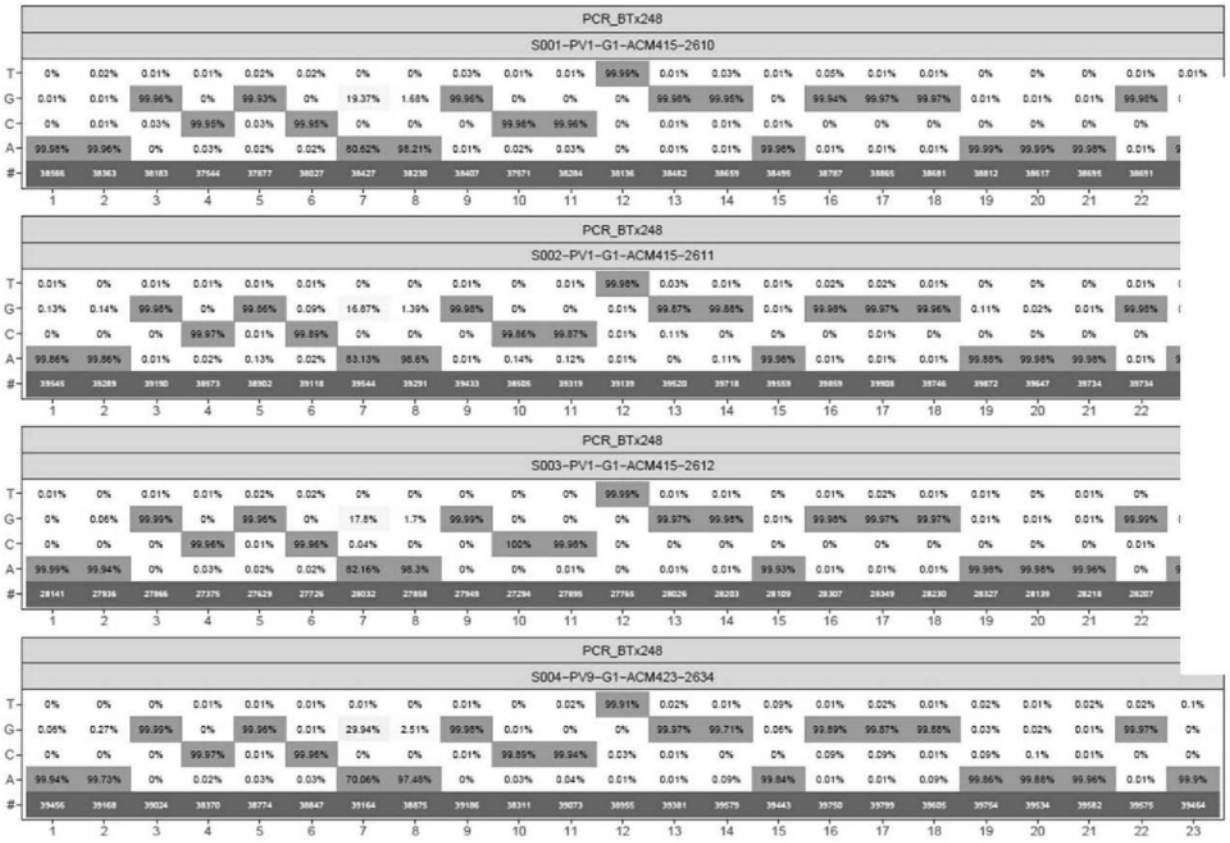

PCR_BTx248
S001-PV1-G1-ACM415-2610
PCR_BTx248
S002-PV1-G1-ACM415-2611
PCR_BTx248
S003-PV1-G1-ACM415-2612
PCR_BTx248
S004-PV9-G1-ACM423-2634

图59A

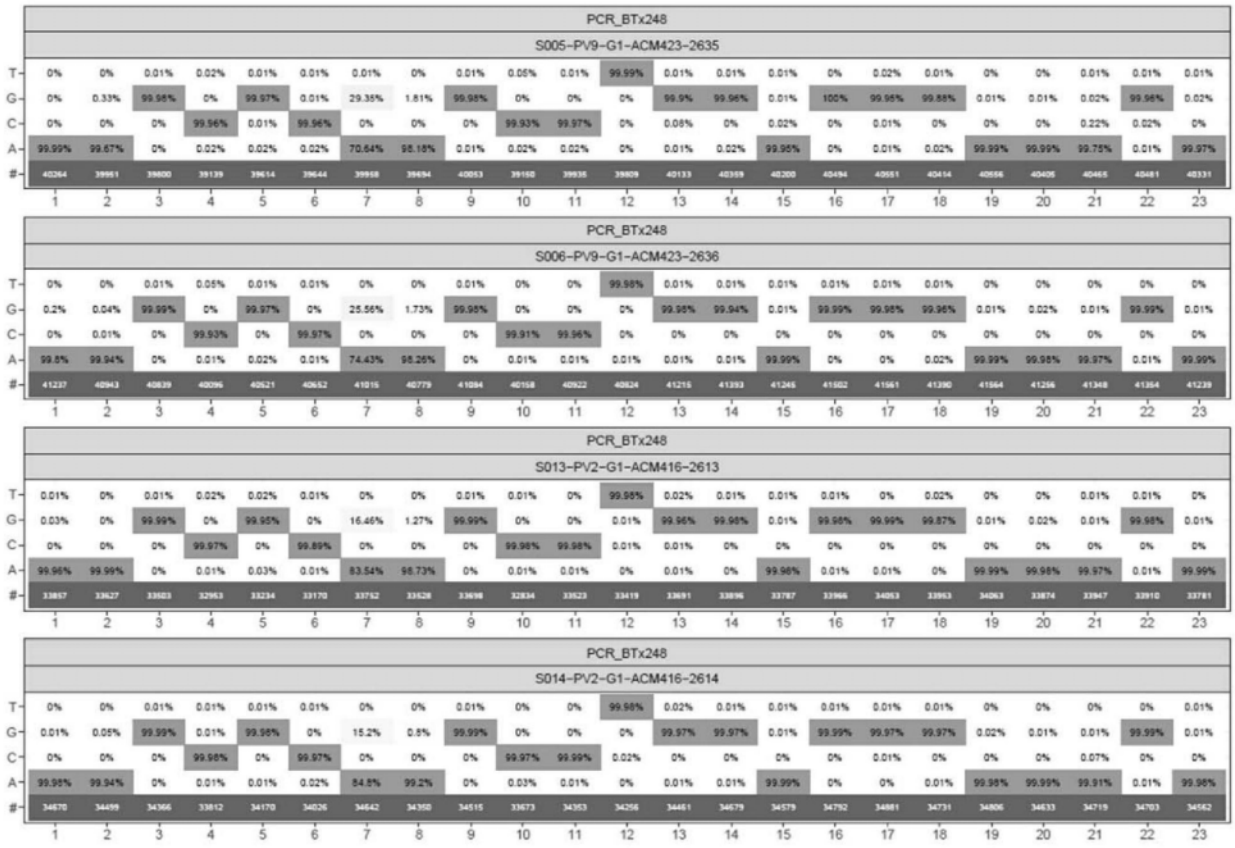

PCR_BTx248
S005-PV9-G1-ACM423-2635
PCR_BTx248
S006-PV9-G1-ACM423-2636
PCR_BTx248
S013-PV2-G1-ACM416-2613
PCR_BTx248
S014-PV2-G1-ACM416-2614
T
G
C
A
#
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

图59B

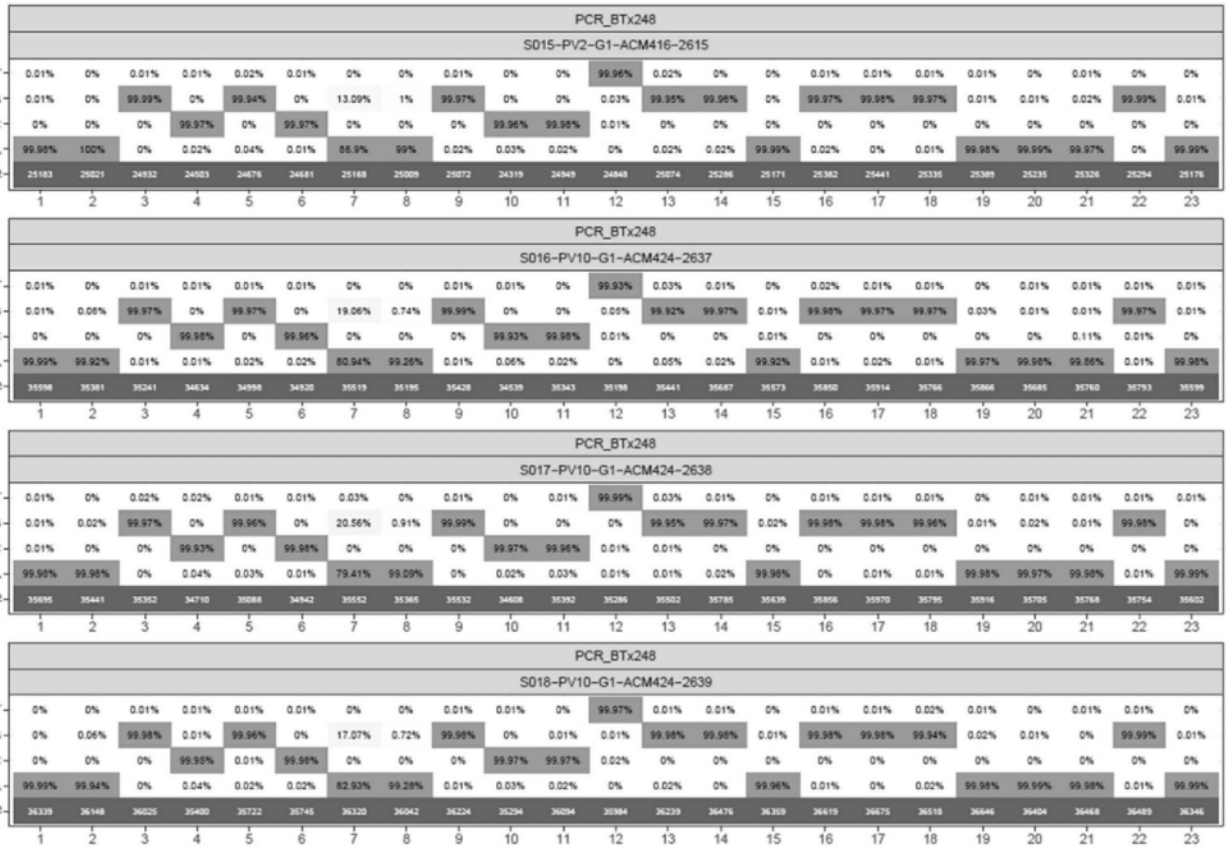

PCR_BTx248
S015-PV2-G1-ACM416-2615
PCR_BTx248
S016-PV10-G1-ACM424-2637
PCR_BTx248
S017-PV10-G1-ACM424-2638
PCR_BTx248
S018-PV10-G1-ACM424-2639

图59C

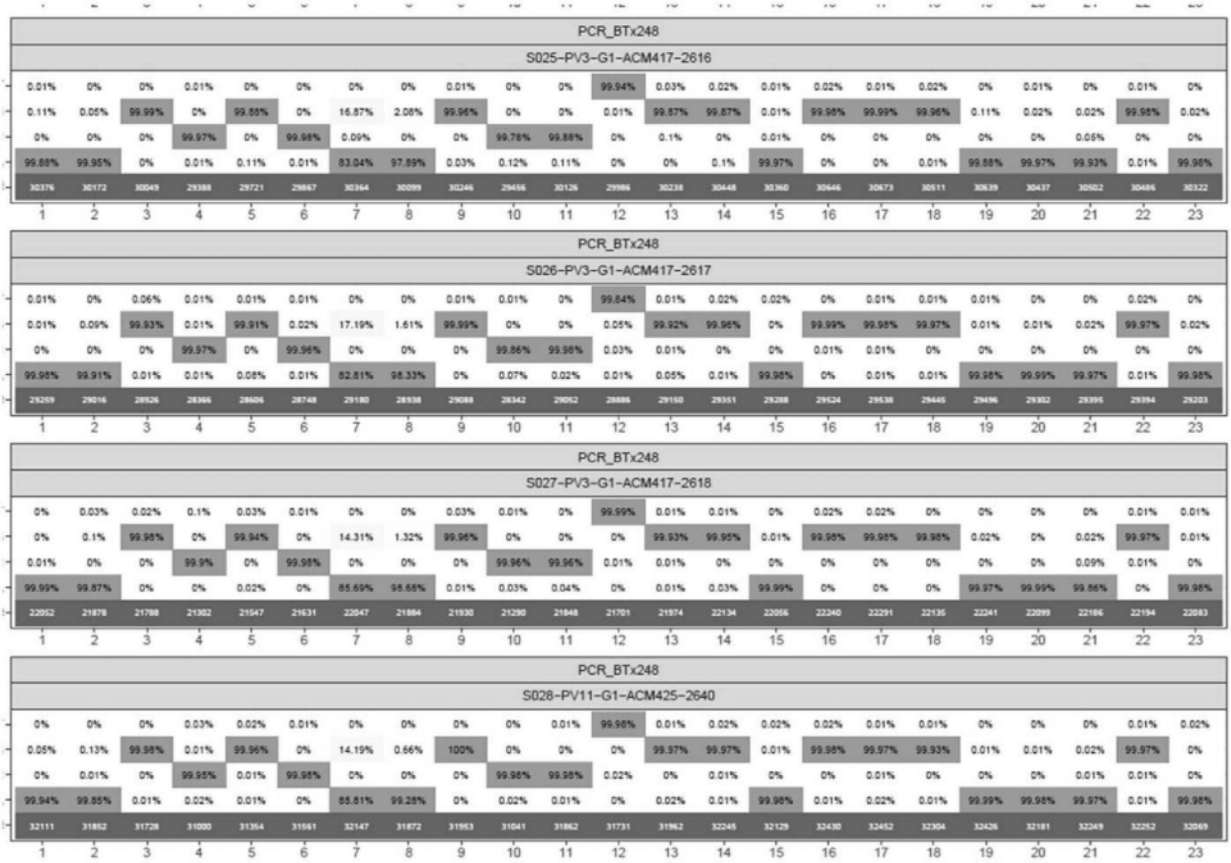

PCR_BTx248
S025-PV3-G1-ACM417-2616
PCR_BTx248
S026-PV3-G1-ACM417-2617
PCR_BTx248
S027-PV3-G1-ACM417-2618
PCR_BTx248
S028-PV11-G1-ACM425-2640

图59D

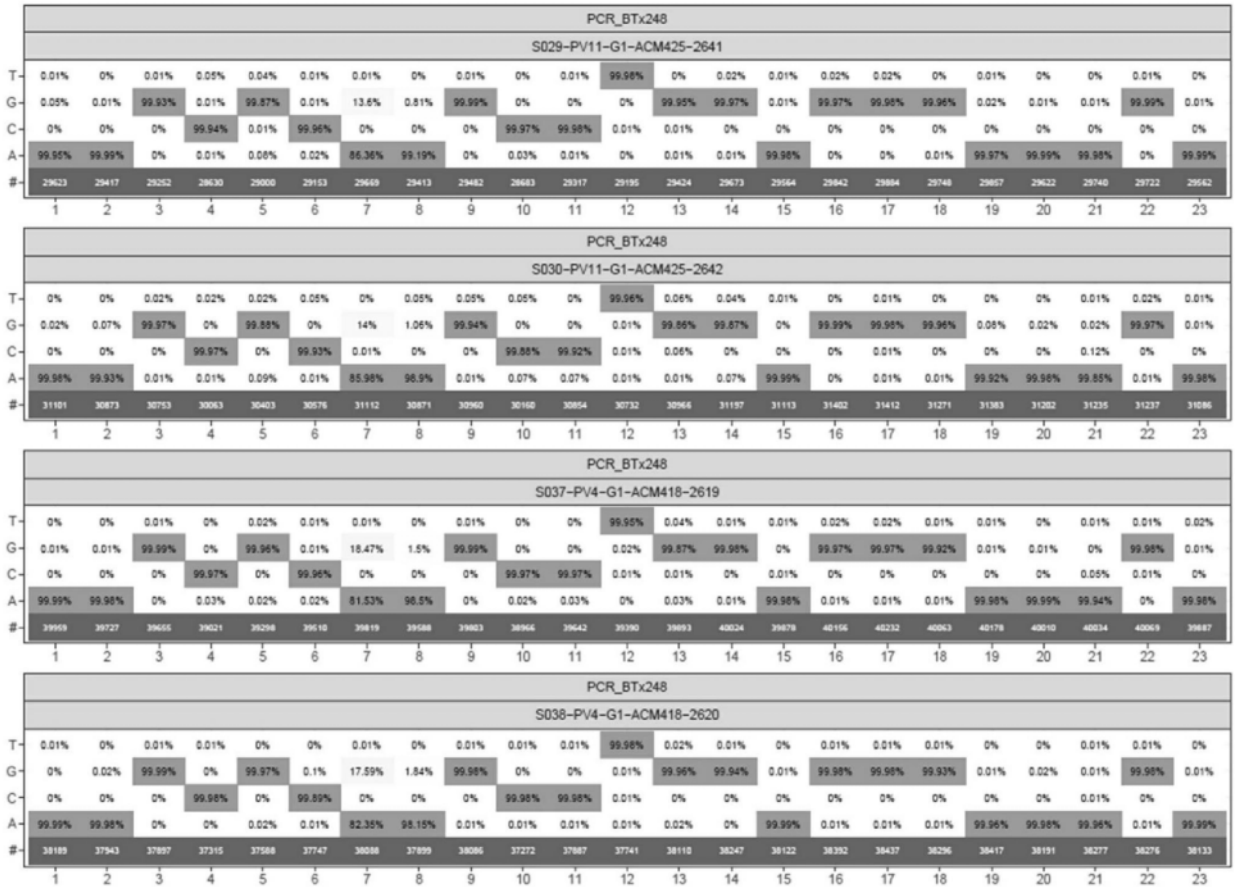

PCR_BTx248
S029-PV11-G1-ACM425-2641
PCR_BTx248
S030-PV11-G1-ACM425-2642
PCR_BTx248
S037-PV4-G1-ACM418-2619
PCR_BTx248
S038-PV4-G1-ACM418-2620
T
G
C
A
#
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

图59E

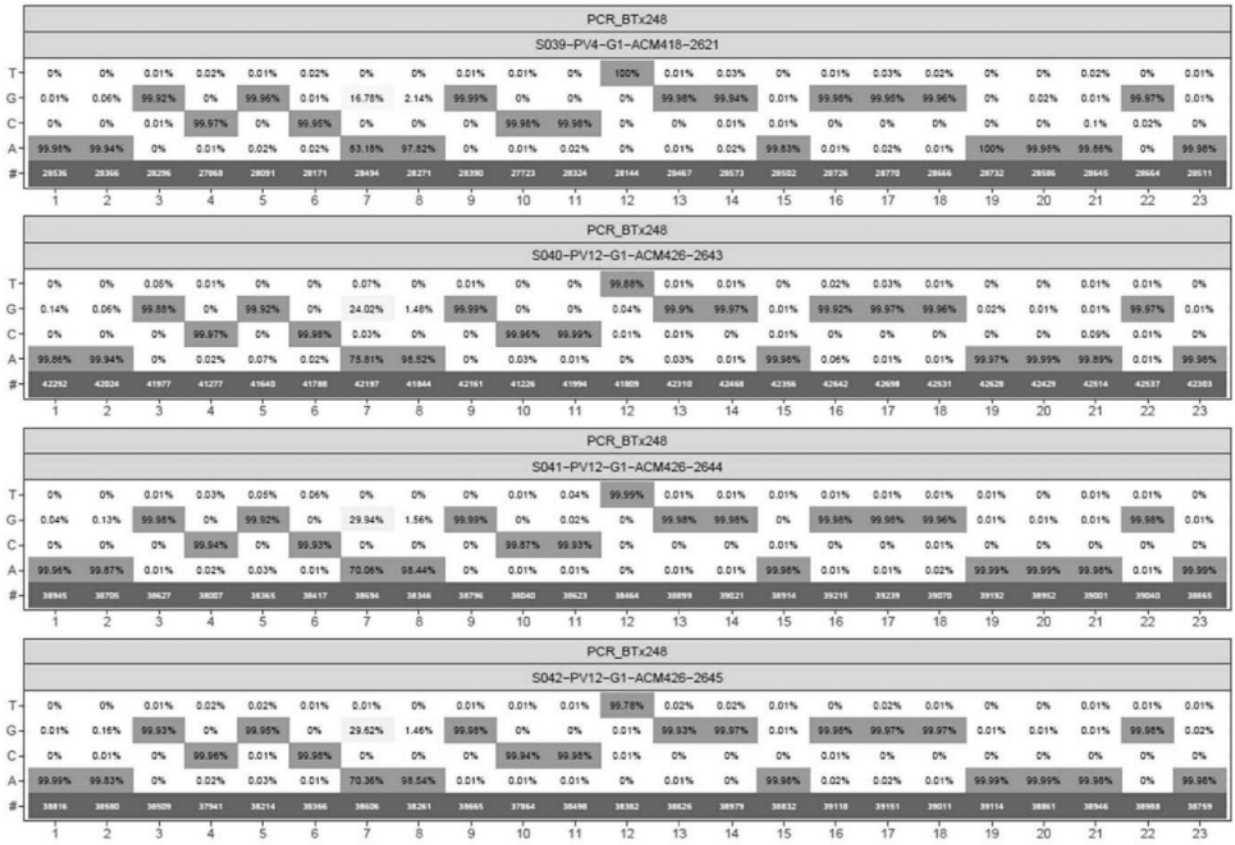

PCR_BTx248
S039-PV4-G1-ACM418-2621
PCR_BTx248
S040-PV12-G1-ACM426-2643
PCR_BTx248
S041-PV12-G1-ACM426-2644
PCR_BTx248
S042-PV12-G1-ACM426-2645

图59F

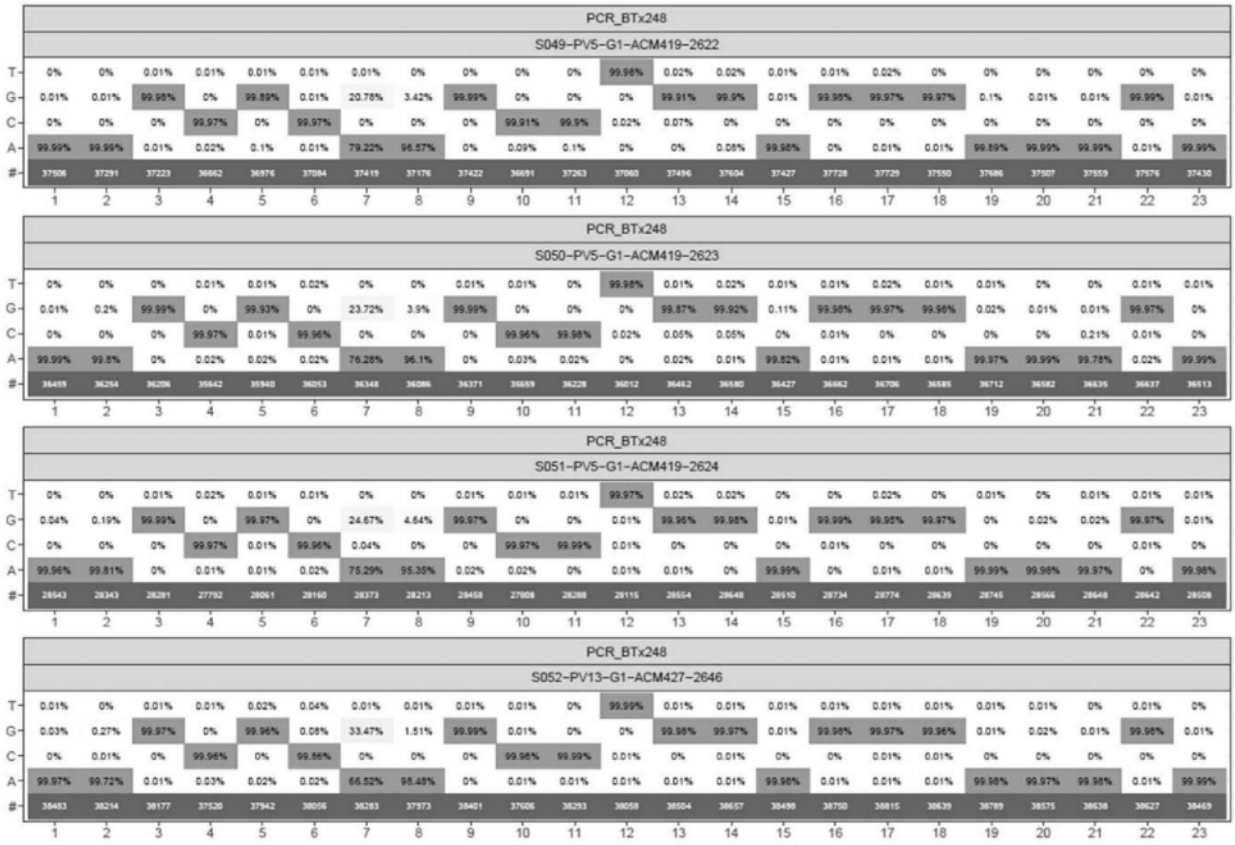

PCR_BTx248
S049-PV5-G1-ACM419-2622
PCR_BTx248
S050-PV5-G1-ACM419-2623
PCR_BTx248
S051-PV5-G1-ACM419-2624
PCR_BTx248
S052-PV13-G1-ACM427-2646

图59G

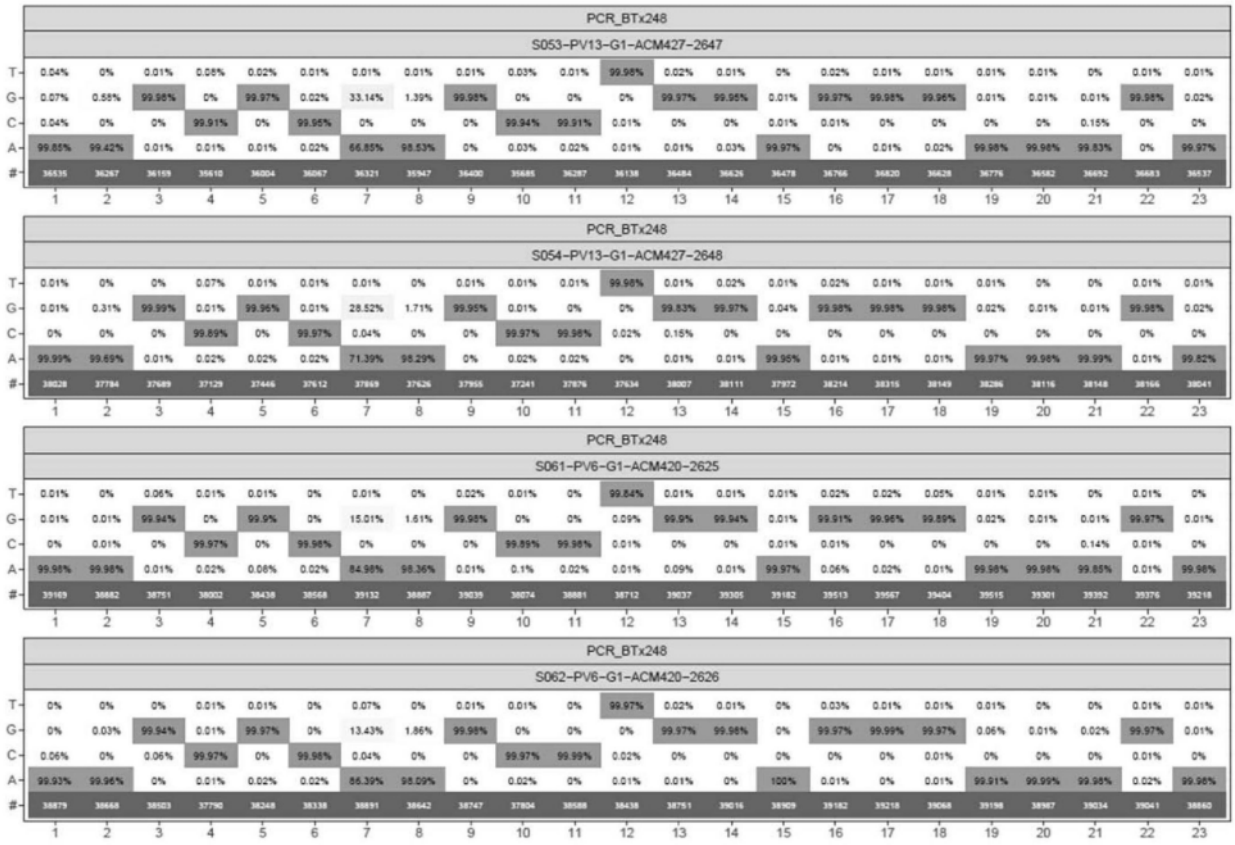

PCR_BTx248
S053-PV13-G1-ACM427-2647
PCR_BTx248
S054-PV13-G1-ACM427-2648
PCR_BTx248
S061-PV6-G1-ACM420-2625
PCR_BTx248
S062-PV6-G1-ACM420-2626
T
G
C
A
#
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

图59H

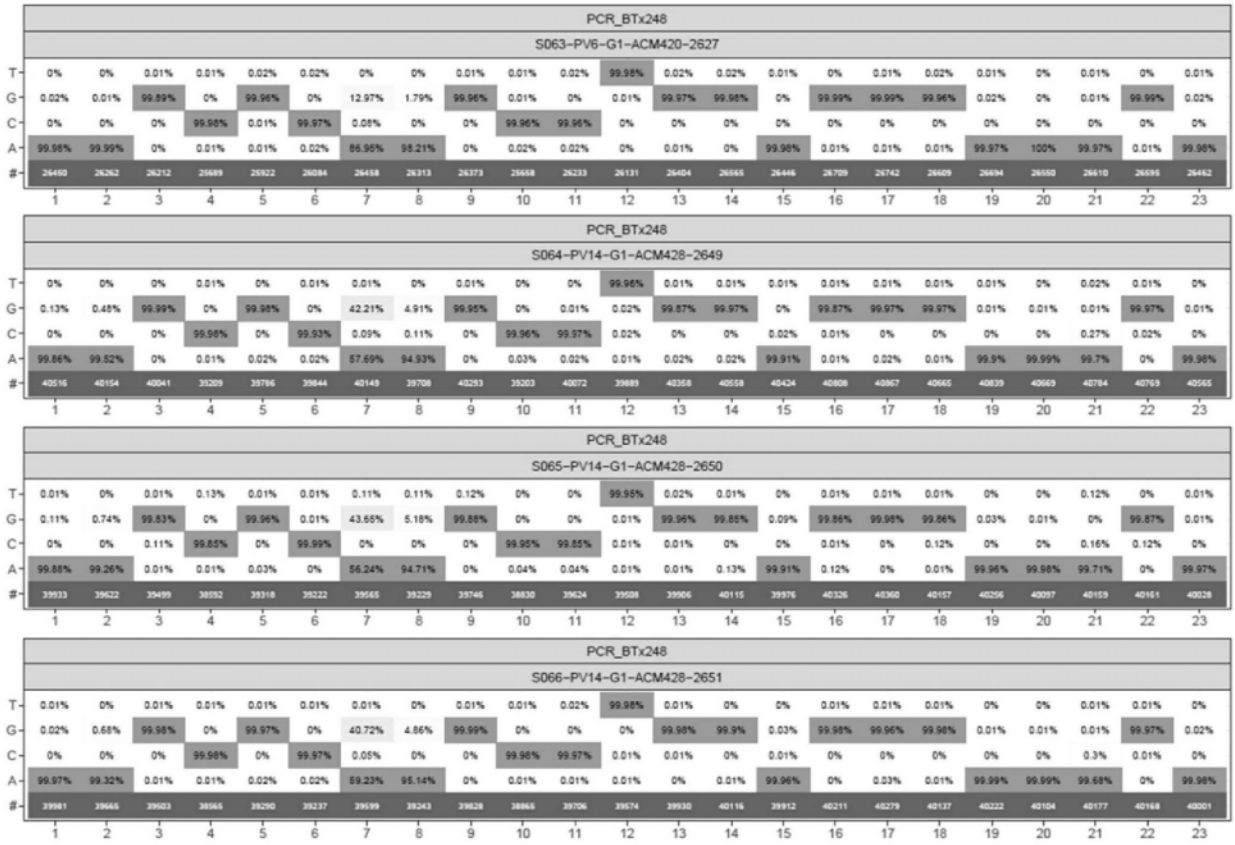

PCR_BTx248
S063-PV6-G1-ACM420-2627
PCR_BTx248
S064-PV14-G1-ACM428-2649
PCR_BTx248
S065-PV14-G1-ACM428-2650
PCR_BTx248
S066-PV14-G1-ACM428-2651
T
G
C
A
#
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

图59I

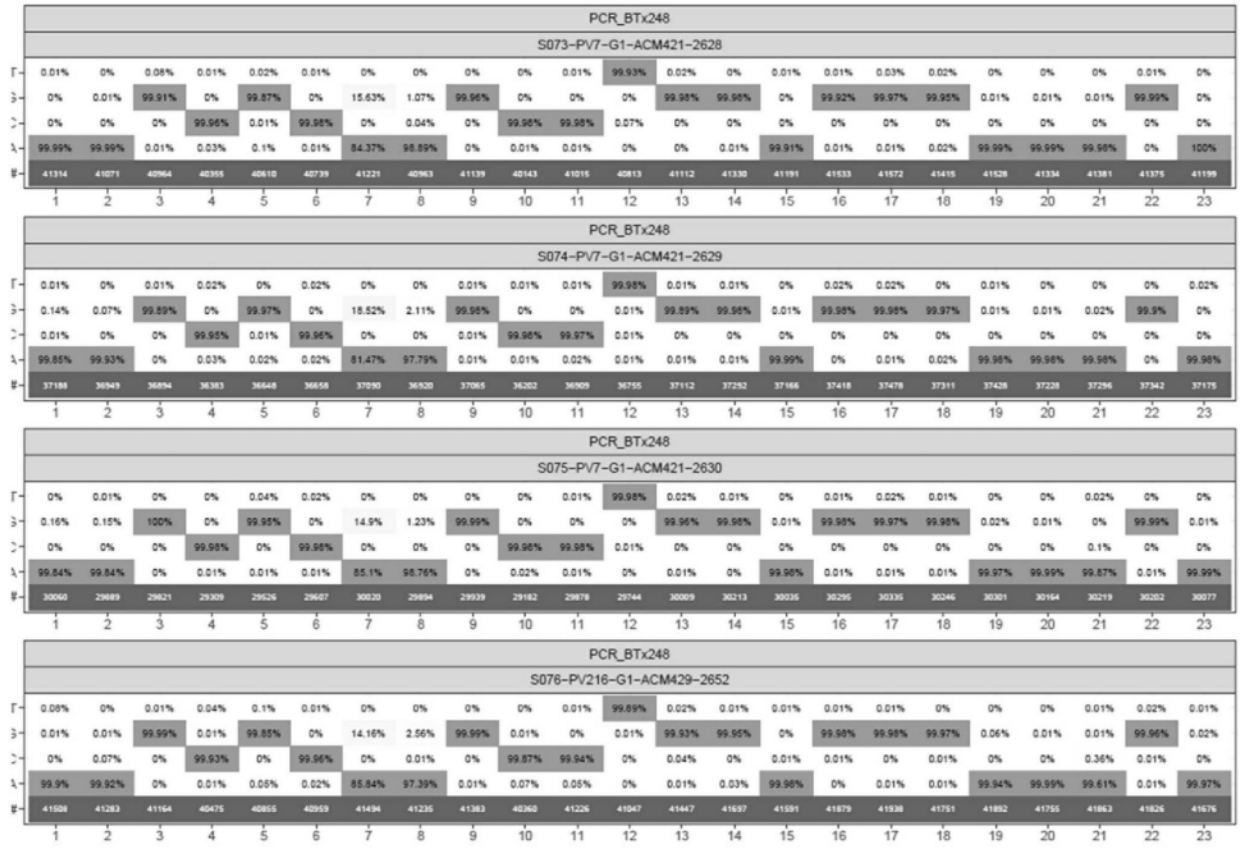

PCR_BTx248
S073-PV7-G1-ACM421-2628
PCR_BTx248
S074-PV7-G1-ACM421-2629
PCR_BTx248
S075-PV7-G1-ACM421-2630
PCR_BTx248
S076-PV216-G1-ACM429-2652

图59J

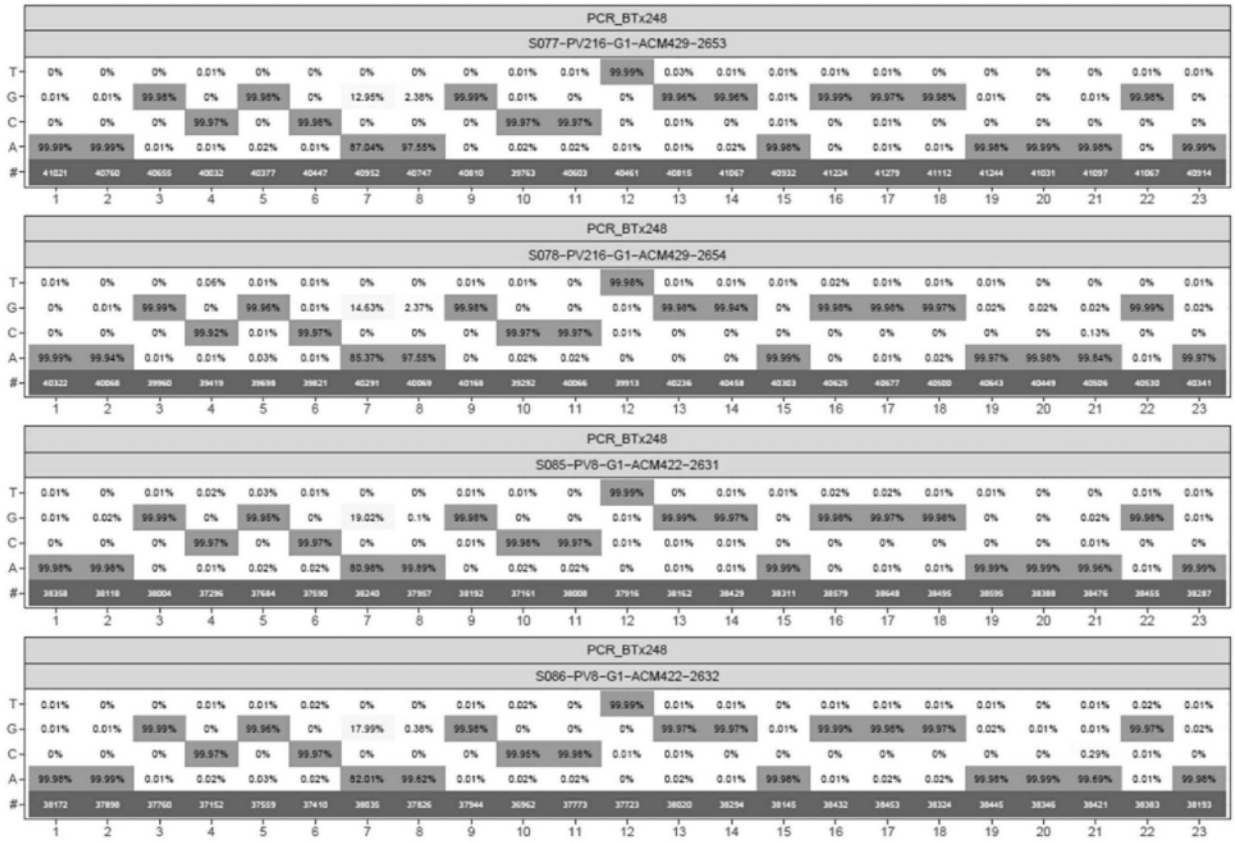

PCR_BTx248
S077-PV216-G1-ACM429-2653
PCR_BTx248
S078-PV216-G1-ACM429-2654
PCR_BTx248
S085-PV8-G1-ACM422-2631
PCR_BTx248
S086-PV8-G1-ACM422-2632
T
G
C
A
#
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

图59K

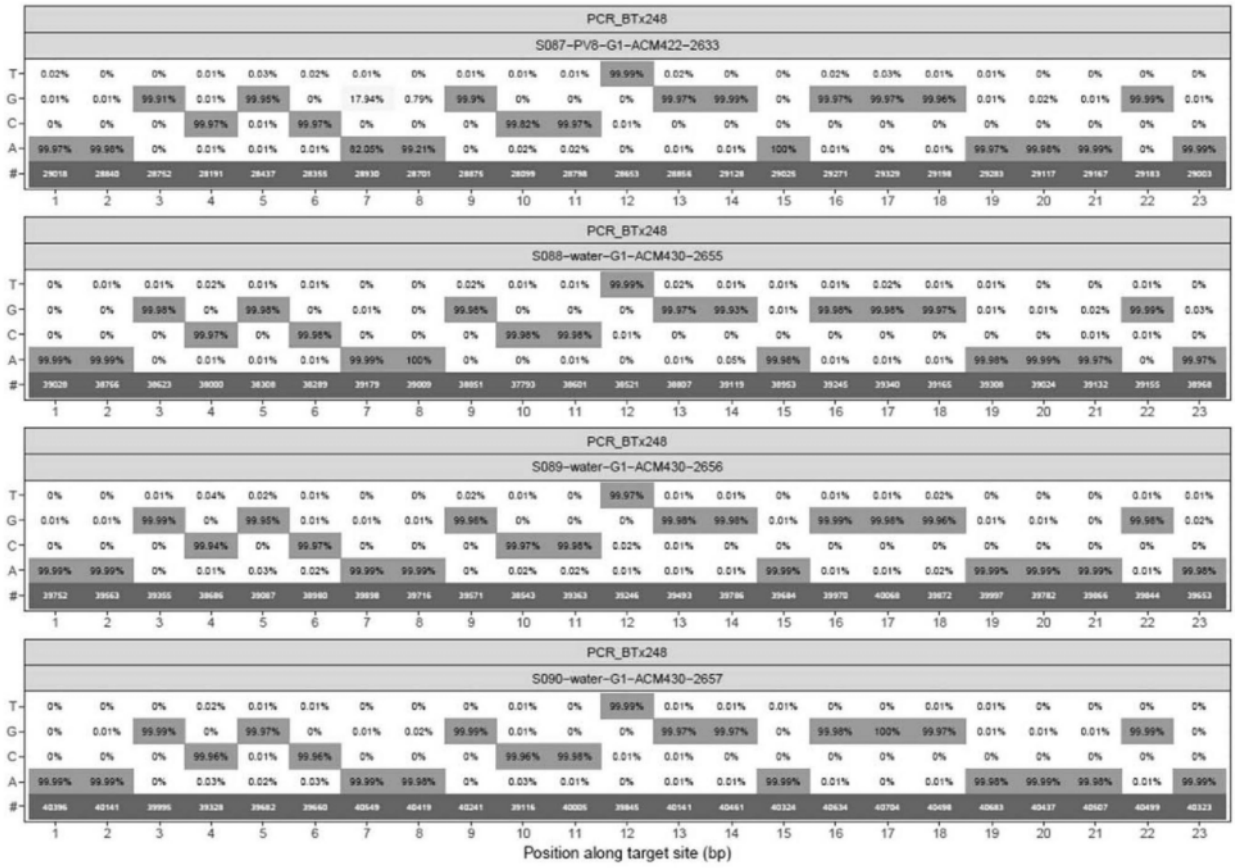

PCR_BTx248
S087-PV8-G1-ACM422-2633
PCR_BTx248
S088-water-G1-ACM430-2655
PCR_BTx248
S089-water-G1-ACM430-2656
PCR_BTx248
S090-water-G1-ACM430-2657
T
G
C
A
#
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Position along target site (bp)

图59L

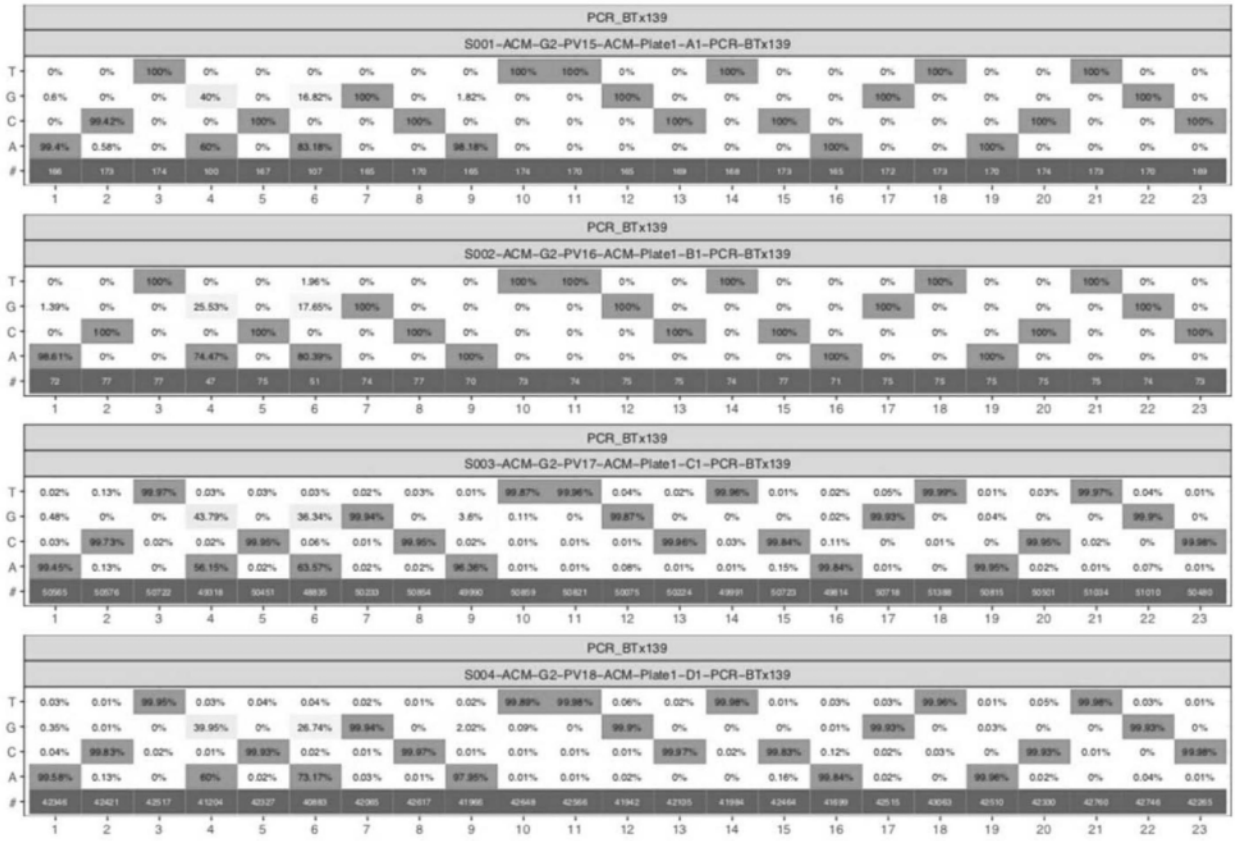

PCR_BTx139
S001-ACM-G2-PV15-ACM-Plate1-A1-PCR-BTx139
PCR_BTx139
S002-ACM-G2-PV16-ACM-Plate1-B1-PCR-BTx139
PCR_BTx139
S003-ACM-G2-PV17-ACM-Plate1-C1-PCR-BTx139
PCR_BTx139
S004-ACM-G2-PV18-ACM-Plate1-D1-PCR-BTx139

图60A

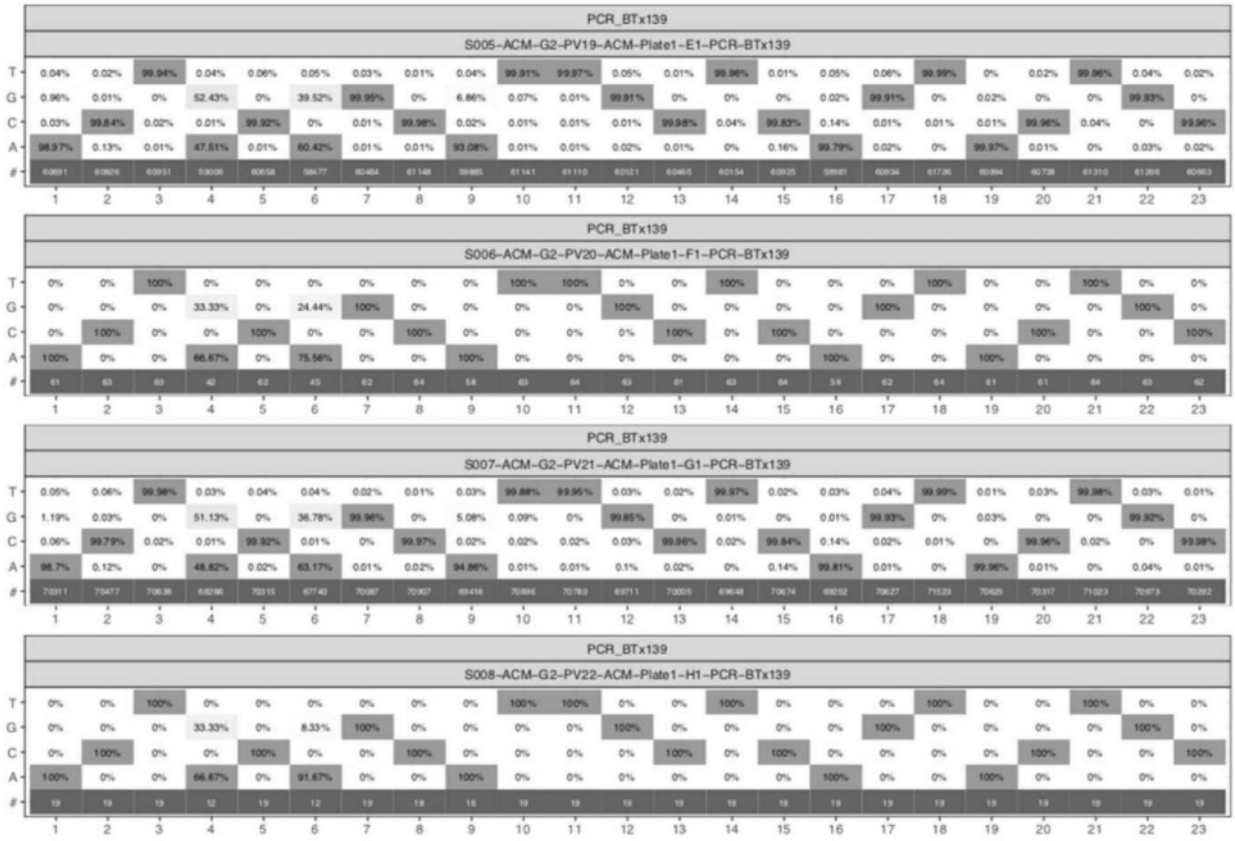

PCR_BTx139
S005-ACM-G2-PV19-ACM-Plate1-E1-PCR-BTx139
PCR_BTx139
S006-ACM-G2-PV20-ACM-Plate1-F1-PCR-BTx139
PCR_BTx139
S007-ACM-G2-PV21-ACM-Plate1-G1-PCR-BTx139
PCR_BTx139
S008-ACM-G2-PV22-ACM-Plate1-H1-PCR-BTx139

图60B

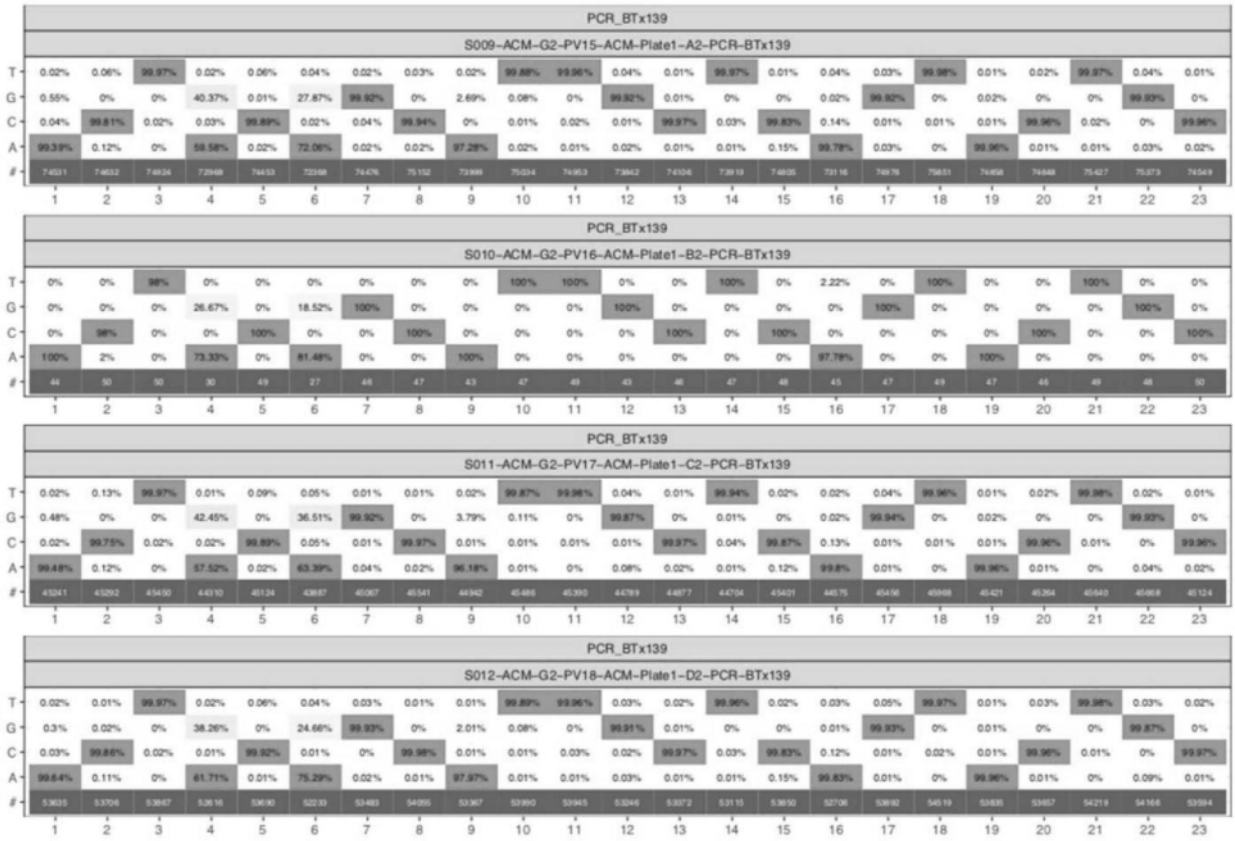

PCR_BTx139
S009-ACM-G2-PV15-ACM-Plate1-A2-PCR-BTx139
PCR_BTx139
S010-ACM-G2-PV16-ACM-Plate1-B2-PCR-BTx139
PCR_BTx139
S011-ACM-G2-PV17-ACM-Plate1-C2-PCR-BTx139
PCR_BTx139
S012-ACM-G2-PV18-ACM-Plate1-D2-PCR-BTx139

图60C

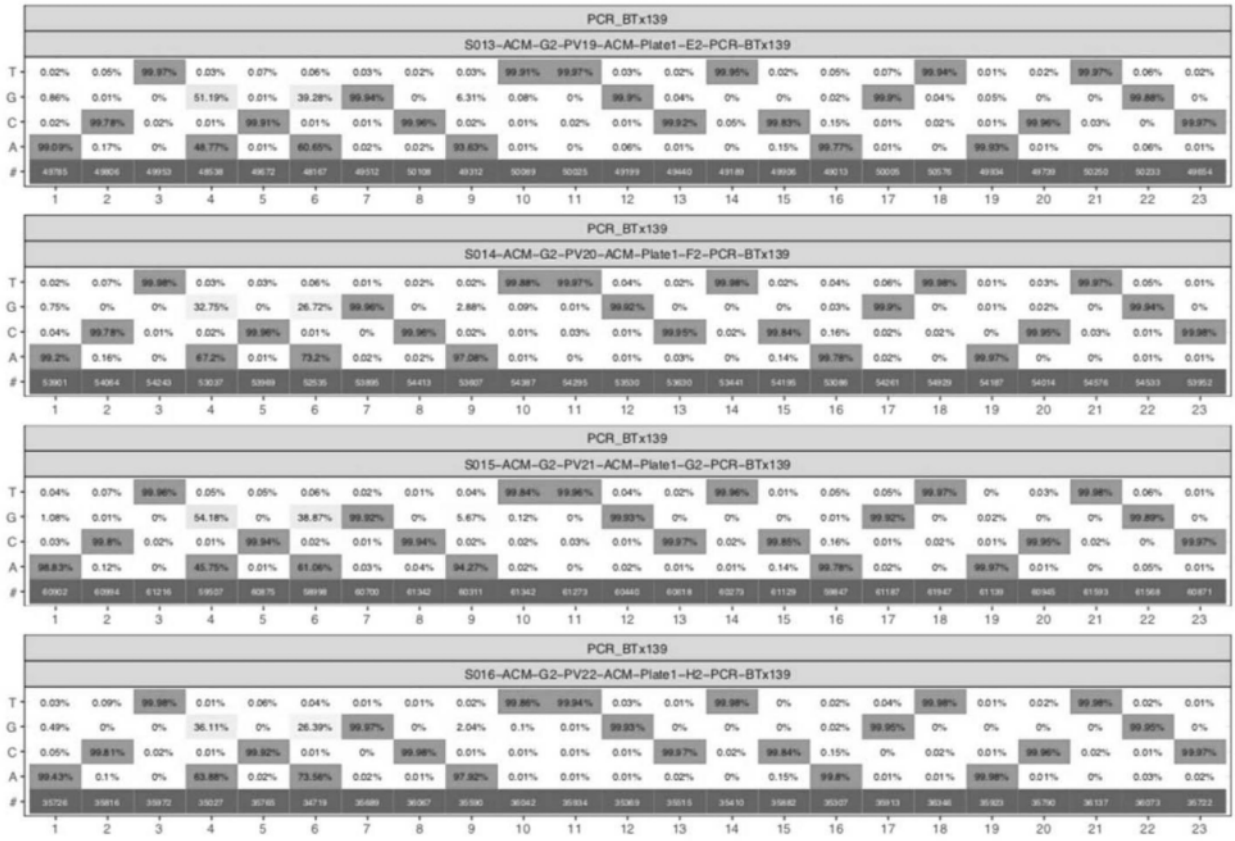

PCR_BTx139
S013-ACM-G2-PV19-ACM-Plate1-E2-PCR-BTx139
PCR_BTx139
S014-ACM-G2-PV20-ACM-Plate1-F2-PCR-BTx139
PCR_BTx139
S015-ACM-G2-PV21-ACM-Plate1-G2-PCR-BTx139
PCR_BTx139
S016-ACM-G2-PV22-ACM-Plate1-H2-PCR-BTx139
T
G
C
A
#
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

图60D

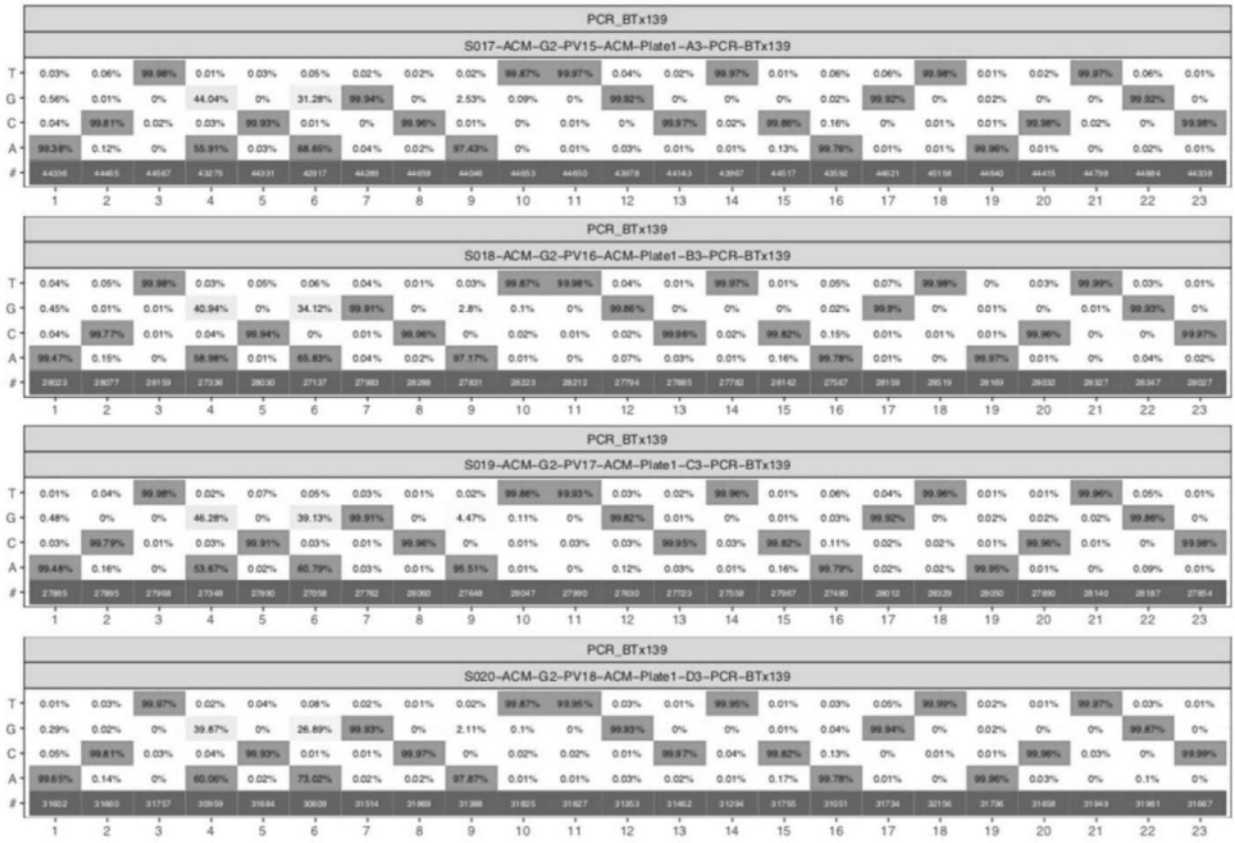

PCR_BTx139
S017-ACM-G2-PV15-ACM-Plate1-A3-PCR-BTx139
PCR_BTx139
S018-ACM-G2-PV16-ACM-Plate1-B3-PCR-BTx139
PCR_BTx139
S019-ACM-G2-PV17-ACM-Plate1-C3-PCR-BTx139
PCR_BTx139
S020-ACM-G2-PV18-ACM-Plate1-D3-PCR-BTx139
T
G
C
A
#
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

图60E

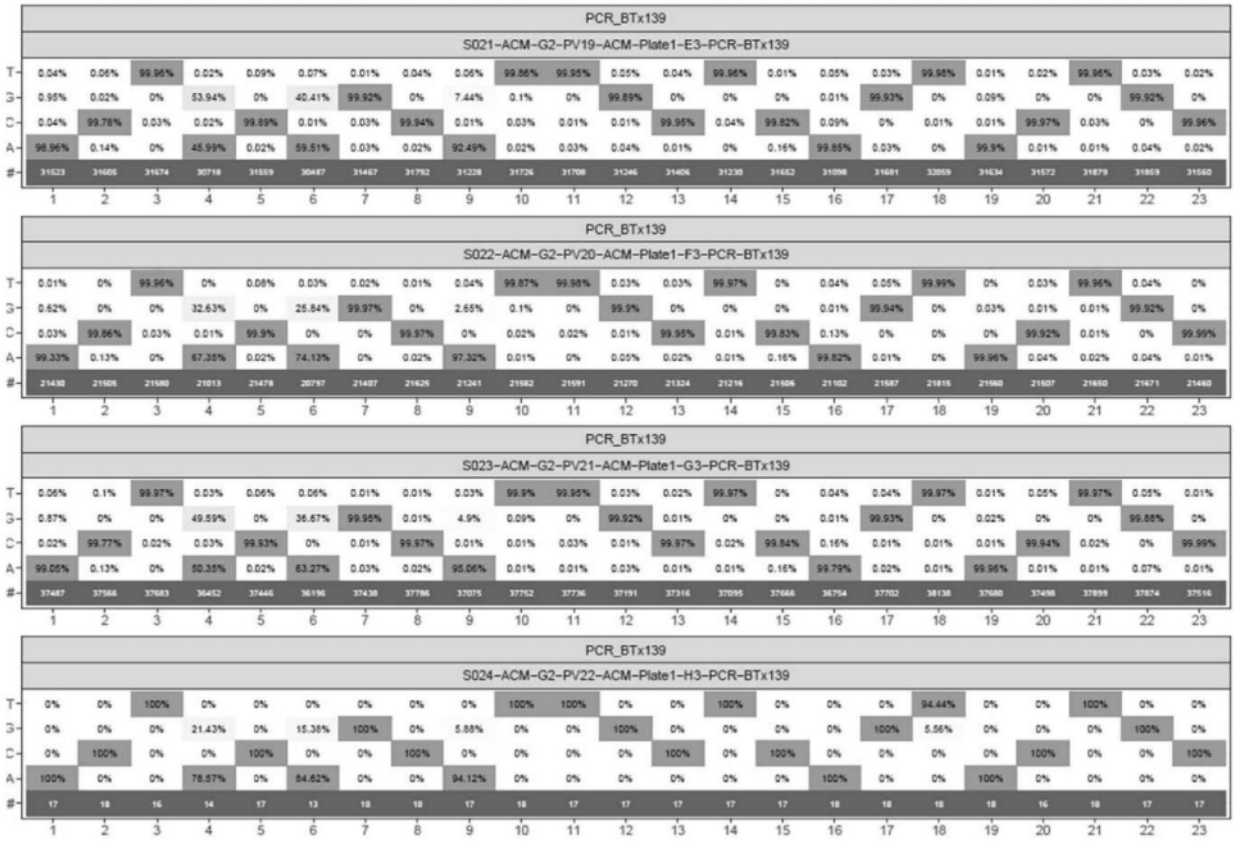

PCR_BTx139
S021-ACM-G2-PV19-ACM-Plate1-E3-PCR-BTx139
PCR_BTx139
S022-ACM-G2-PV20-ACM-Plate1-F3-PCR-BTx139
PCR_BTx139
S023-ACM-G2-PV21-ACM-Plate1-G3-PCR-BTx139
PCR_BTx139
S024-ACM-G2-PV22-ACM-Plate1-H3-PCR-BTx139

图60F

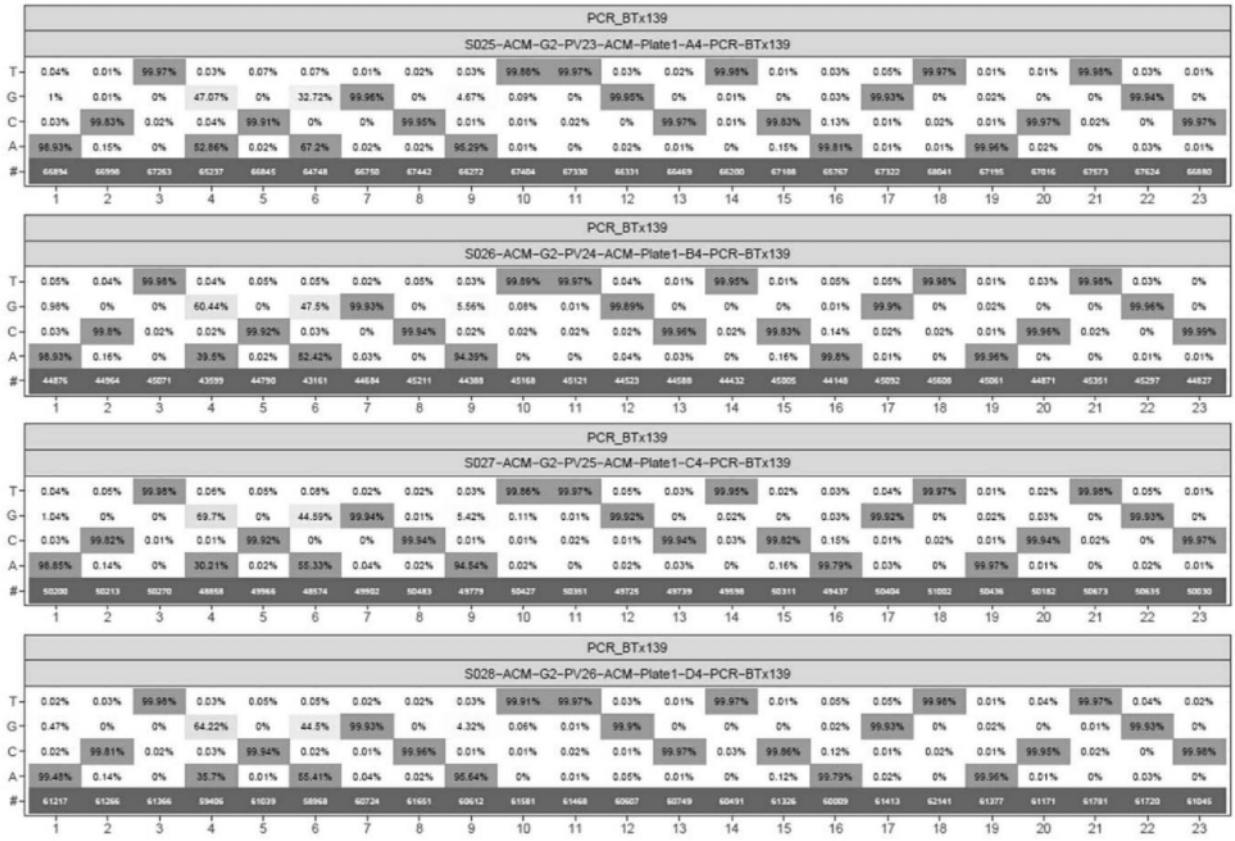

PCR_BTx139
S025-ACM-G2-PV23-ACM-Plate1-A4-PCR-BTx139
PCR_BTx139
S026-ACM-G2-PV24-ACM-Plate1-B4-PCR-BTx139
PCR_BTx139
S027-ACM-G2-PV25-ACM-Plate1-C4-PCR-BTx139
PCR_BTx139
S028-ACM-G2-PV26-ACM-Plate1-D4-PCR-BTx139
T
G
C
A
#
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

图60G

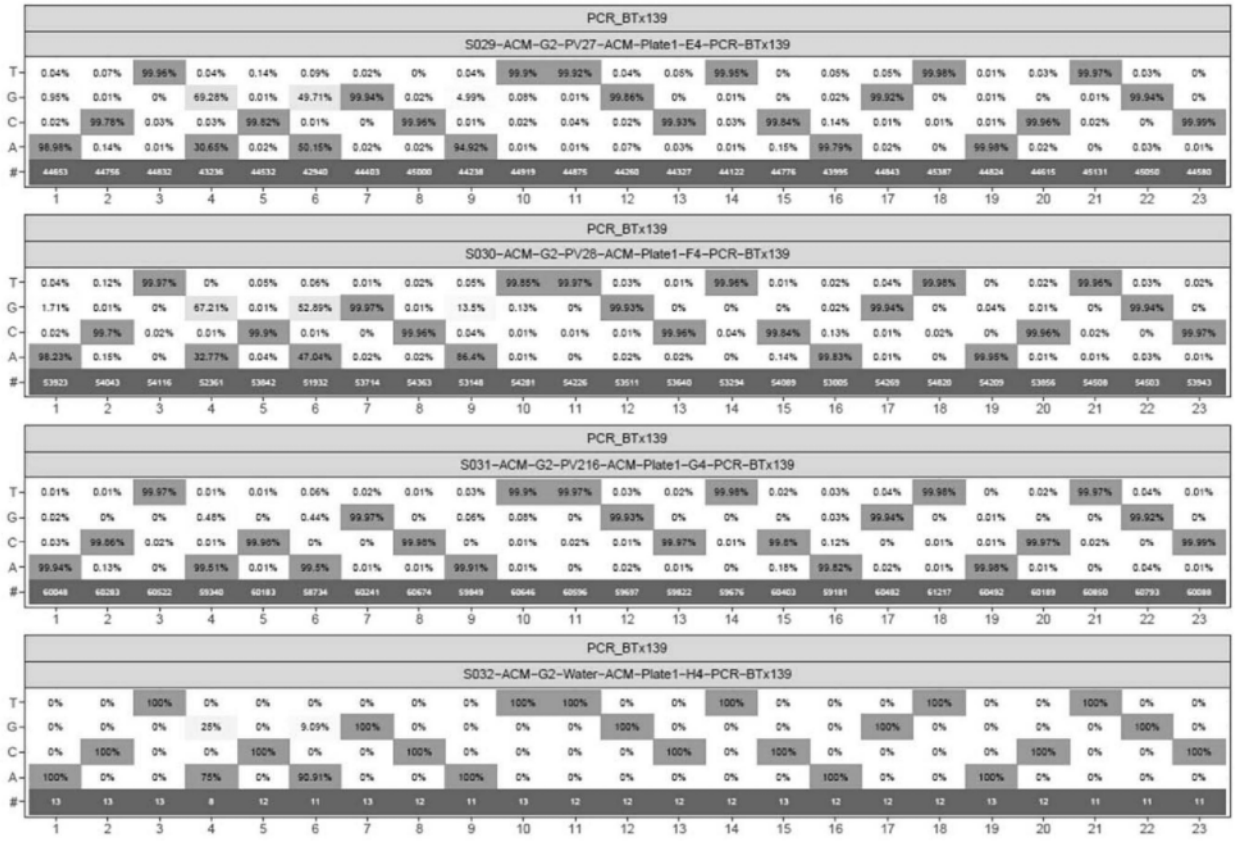

PCR_BTx139
S029-ACM-G2-PV27-ACM-Plate1-E4-PCR-BTx139
PCR_BTx139
S030-ACM-G2-PV28-ACM-Plate1-F4-PCR-BTx139
PCR_BTx139
S031-ACM-G2-PV216-ACM-Plate1-G4-PCR-BTx139
PCR_BTx139
S032-ACM-G2-Water-ACM-Plate1-H4-PCR-BTx139
T
G
C
A
#
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

图60H

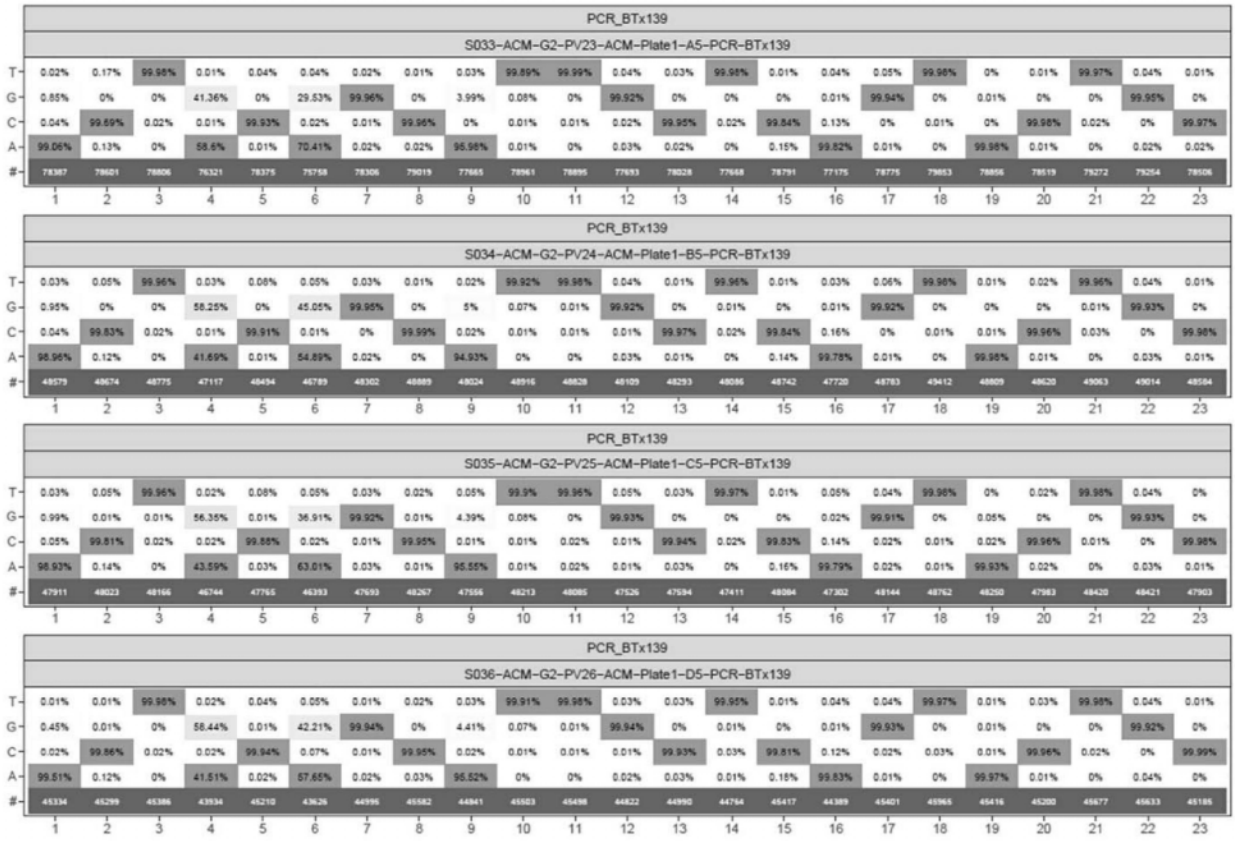

PCR_BTx139
S033-ACM-G2-PV23-ACM-Plate1-A5-PCR-BTx139
PCR_BTx139
S034-ACM-G2-PV24-ACM-Plate1-B5-PCR-BTx139
PCR_BTx139
S035-ACM-G2-PV25-ACM-Plate1-C5-PCR-BTx139
PCR_BTx139
S036-ACM-G2-PV26-ACM-Plate1-D5-PCR-BTx139
T
G
C
A
#
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

图60I

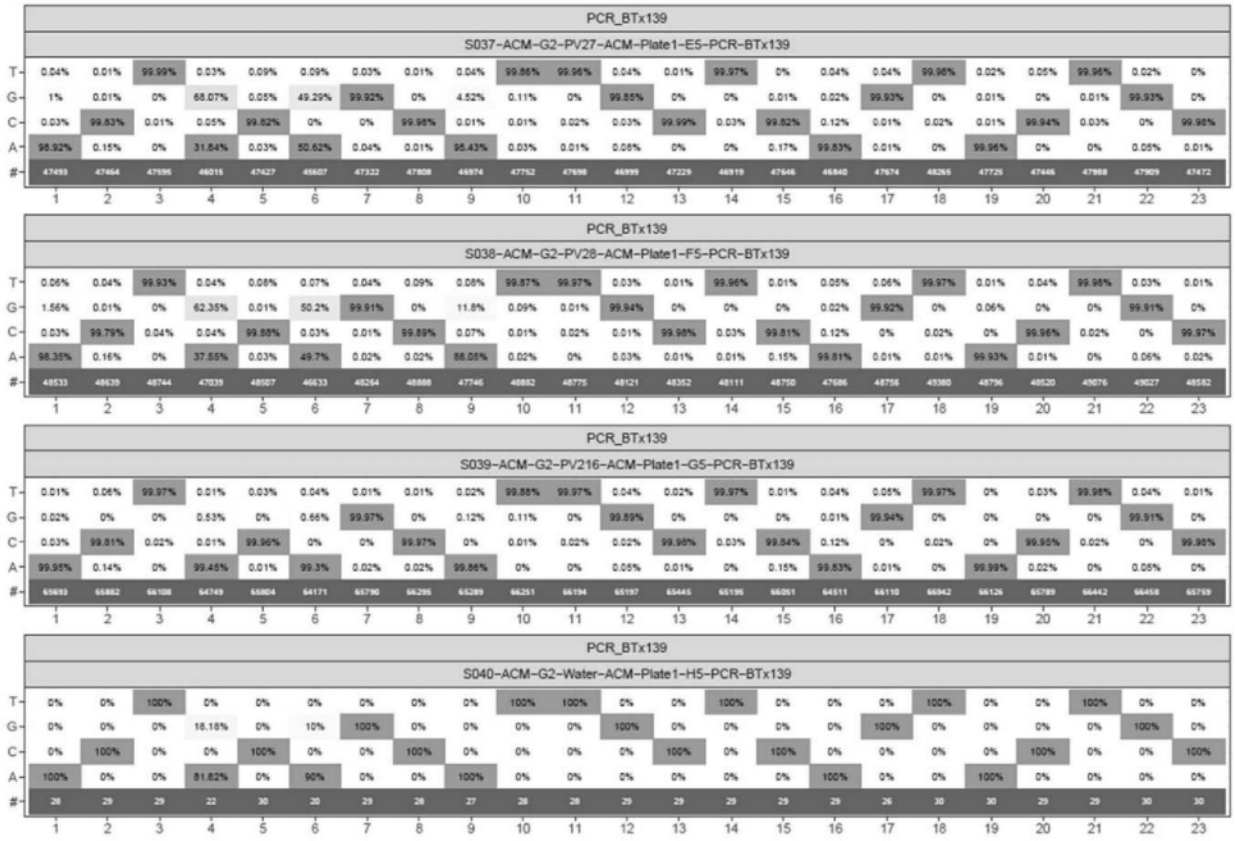

PCR_BTx139
S037-ACM-G2-PV27-ACM-Plate1-E5-PCR-BTx139
PCR_BTx139
S038-ACM-G2-PV28-ACM-Plate1-F5-PCR-BTx139
PCR_BTx139
S039-ACM-G2-PV216-ACM-Plate1-G5-PCR-BTx139
PCR_BTx139
S040-ACM-G2-Water-ACM-Plate1-H5-PCR-BTx139

图60J

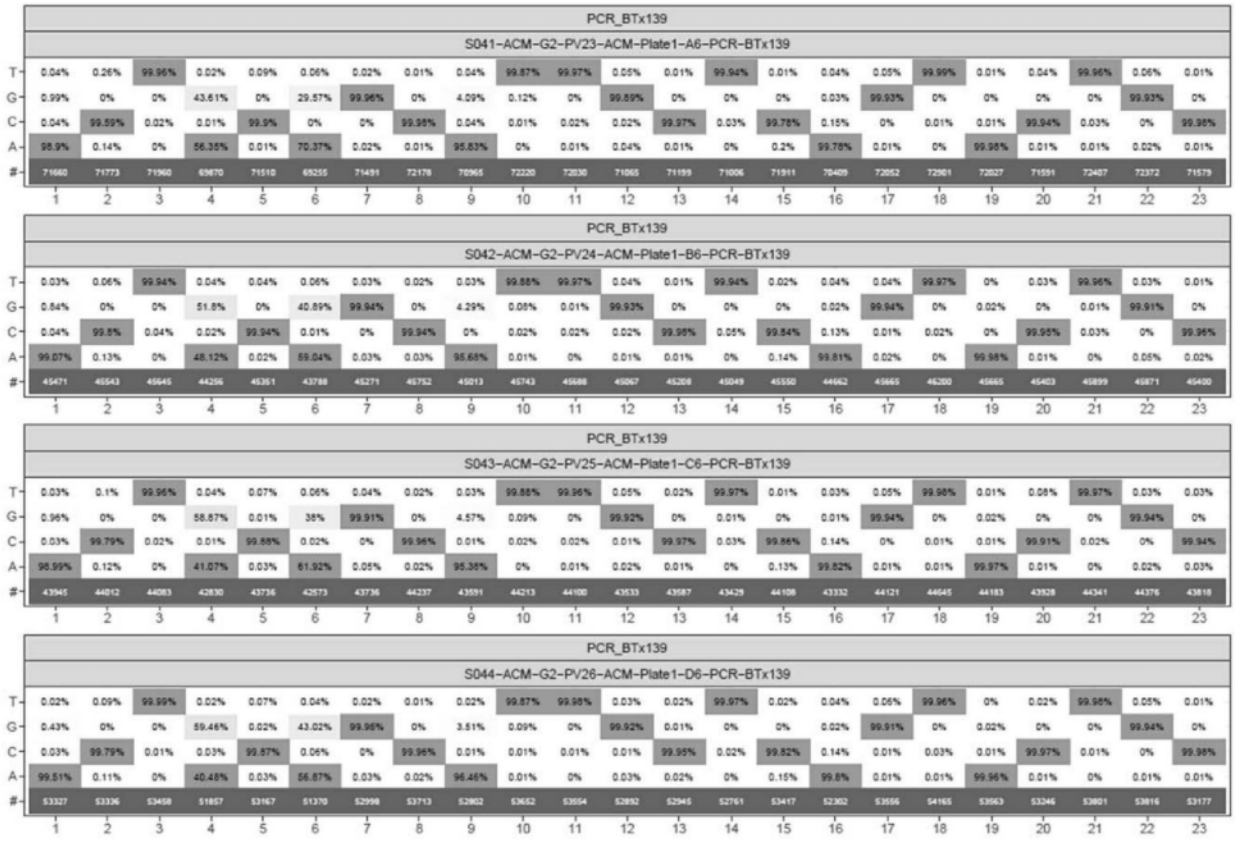

PCR_BTx139
S041-ACM-G2-PV23-ACM-Plate1-A6-PCR-BTx139
PCR_BTx139
S042-ACM-G2-PV24-ACM-Plate1-B6-PCR-BTx139
PCR_BTx139
S043-ACM-G2-PV25-ACM-Plate1-C6-PCR-BTx139
PCR_BTx139
S044-ACM-G2-PV26-ACM-Plate1-D6-PCR-BTx139
T
G
C
A
#
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

图60K

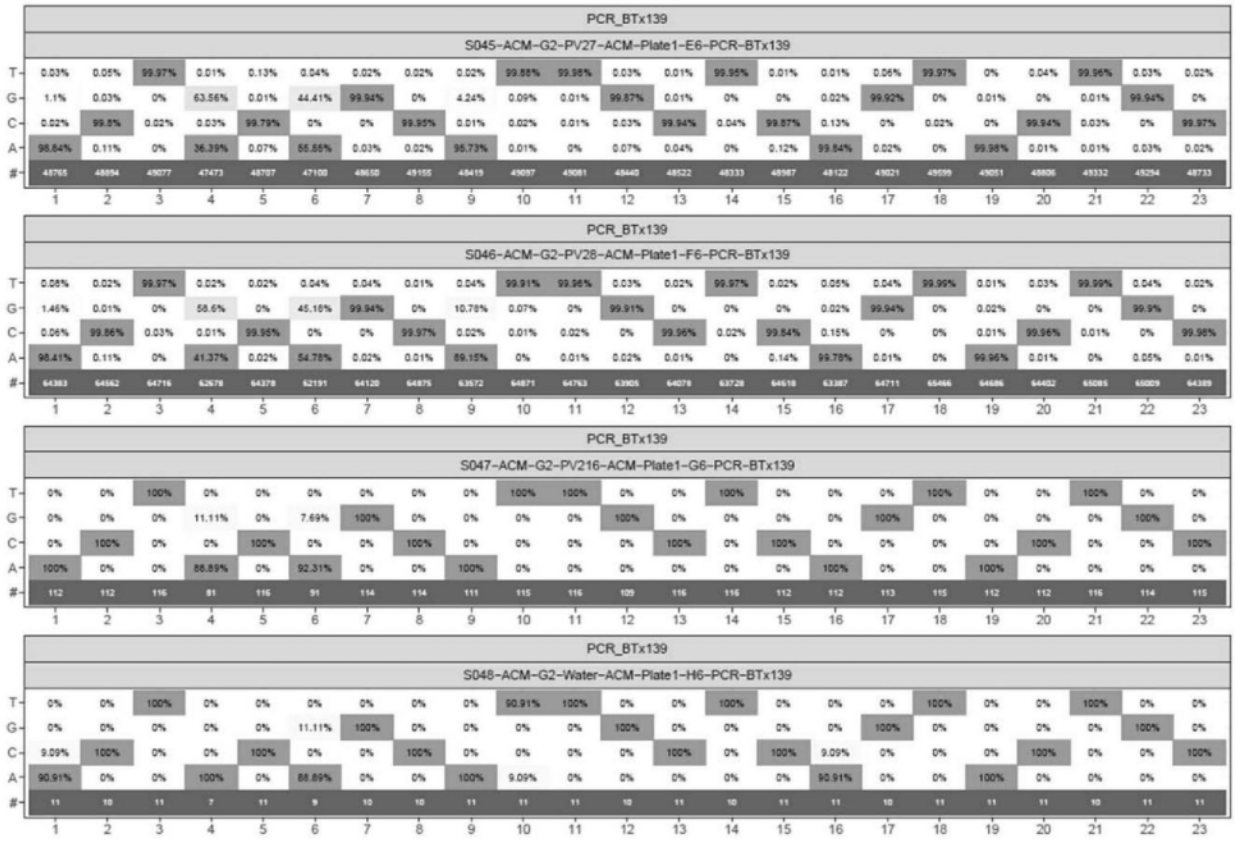

PCR_BTx139
S045-ACM-G2-PV27-ACM-Plate1-E6-PCR-BTx139
PCR_BTx139
S046-ACM-G2-PV28-ACM-Plate1-F6-PCR-BTx139
PCR_BTx139
S047-ACM-G2-PV216-ACM-Plate1-G6-PCR-BTx139
PCR_BTx139
S048-ACM-G2-Water-ACM-Plate1-H6-PCR-BTx139
T
G
C
A
#
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

图60L

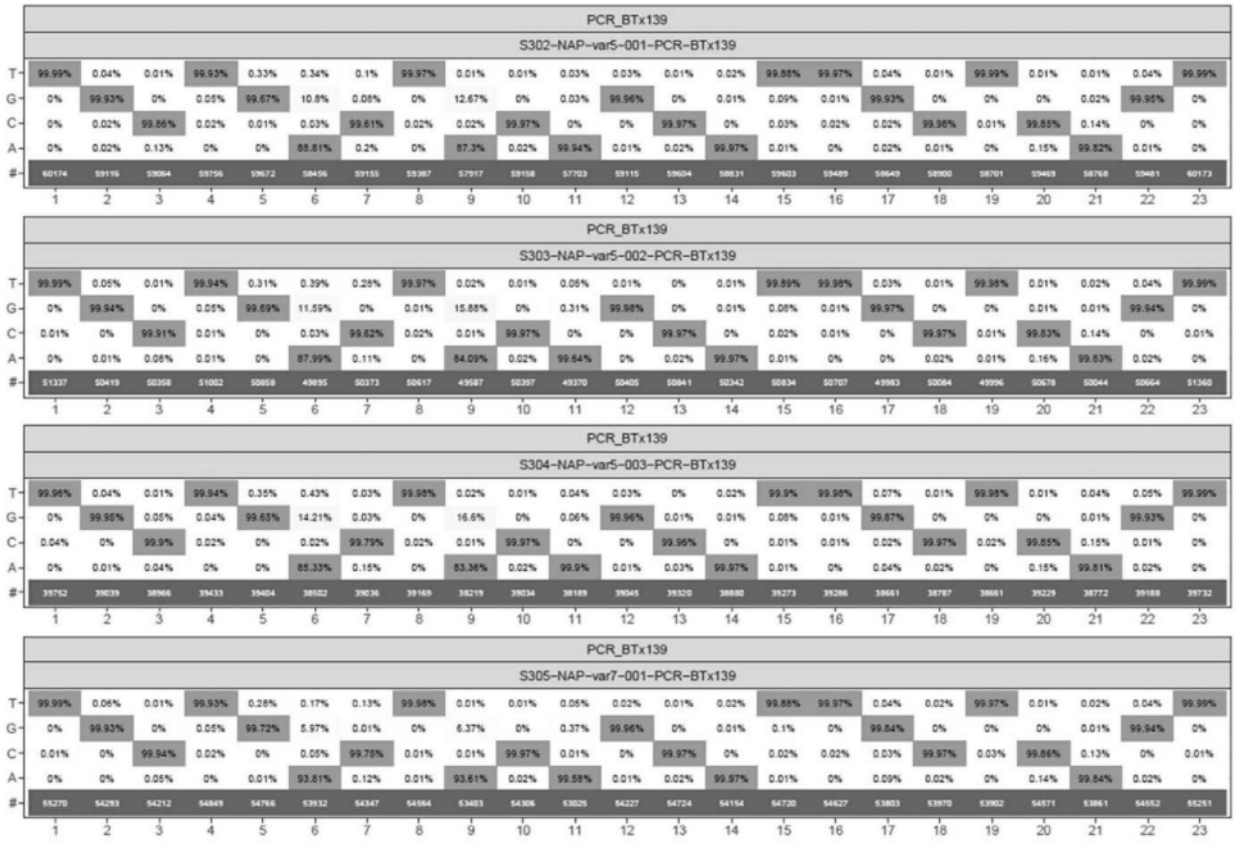

PCR_BTx139
S302-NAP-var5-001-PCR-BTx139
PCR_BTx139
S303-NAP-var5-002-PCR-BTx139
PCR_BTx139
S304-NAP-var5-003-PCR-BTx139
PCR_BTx139
S305-NAP-var7-001-PCR-BTx139
T
G
C
A
#
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

图60M

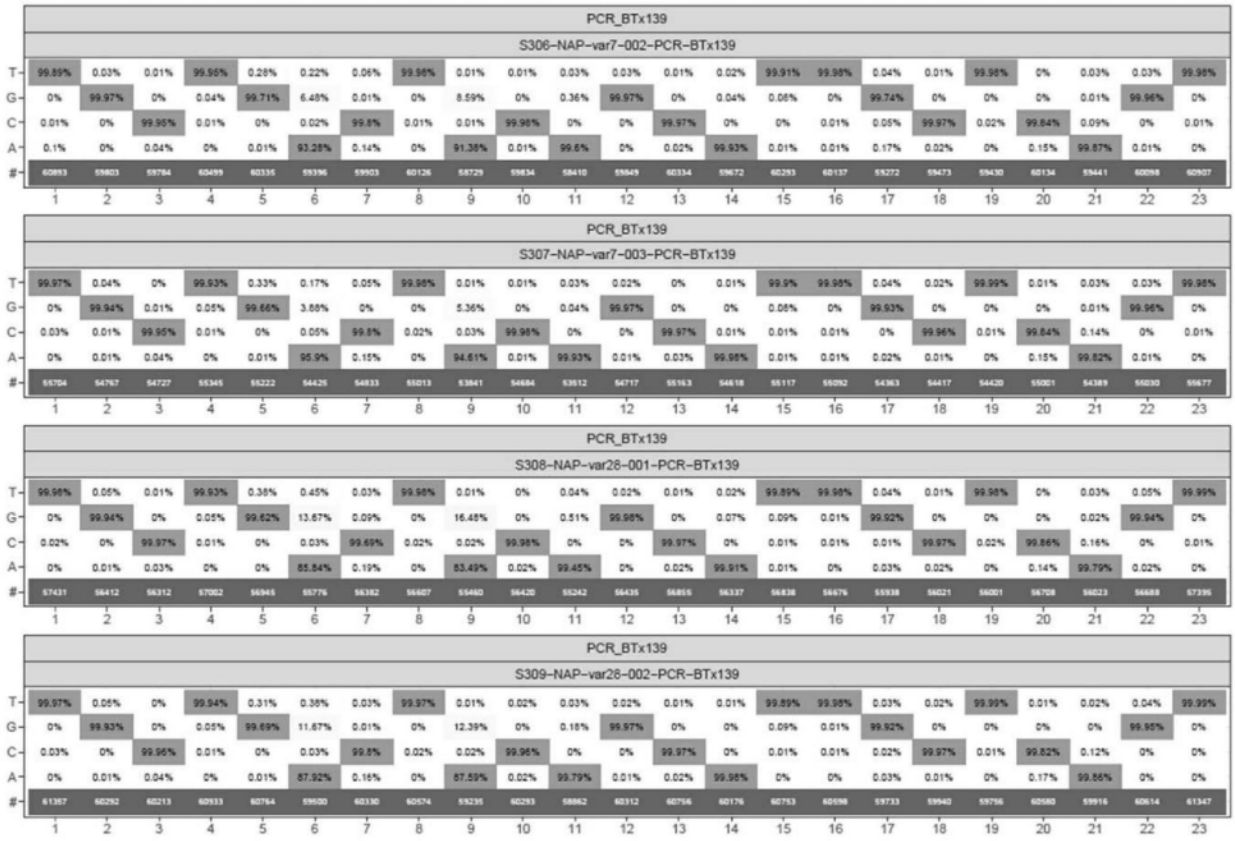

PCR_BTx139
S306-NAP-var7-002-PCR-BTx139
PCR_BTx139
S307-NAP-var7-003-PCR-BTx139
PCR_BTx139
S308-NAP-var28-001-PCR-BTx139
PCR_BTx139
S309-NAP-var28-002-PCR-BTx139

图60N

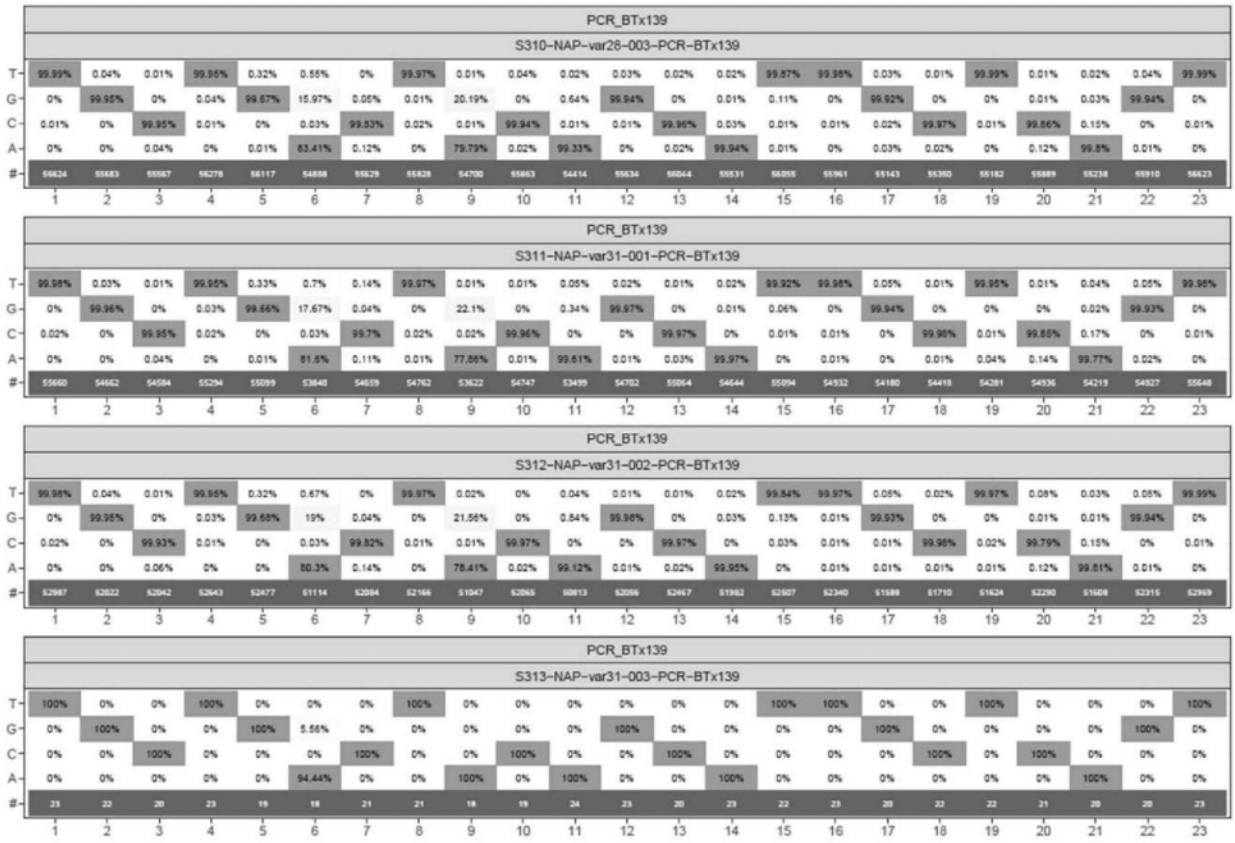

PCR_BTx139
S310-NAP-var28-003-PCR-BTx139
PCR_BTx139
S311-NAP-var31-001-PCR-BTx139
PCR_BTx139
S312-NAP-var31-002-PCR-BTx139
PCR_BTx139
S313-NAP-var31-003-PCR-BTx139

图600

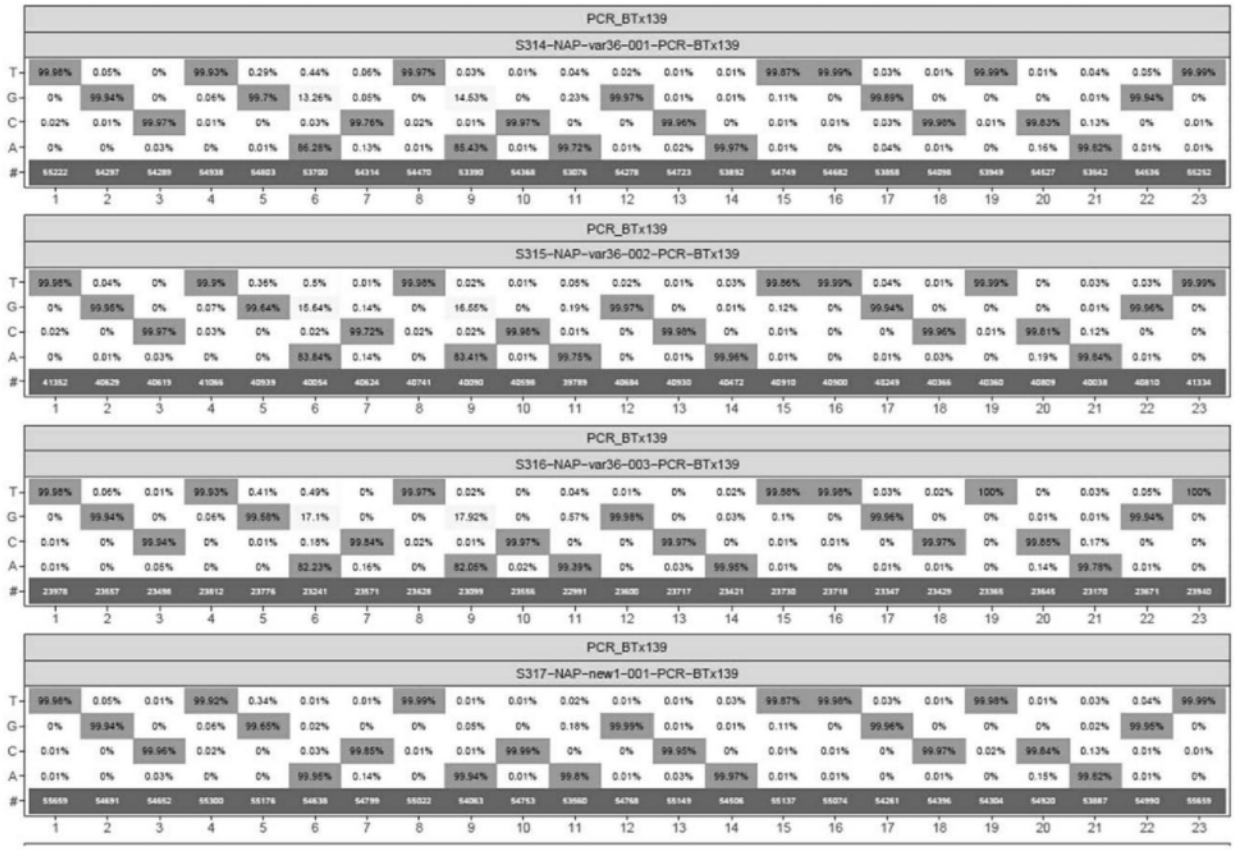

PCR_BTx139
S314-NAP-var36-001-PCR-BTx139
PCR_BTx139
S315-NAP-var36-002-PCR-BTx139
PCR_BTx139
S316-NAP-var36-003-PCR-BTx139
PCR_BTx139
S317-NAP-new1-001-PCR-BTx139

图60P

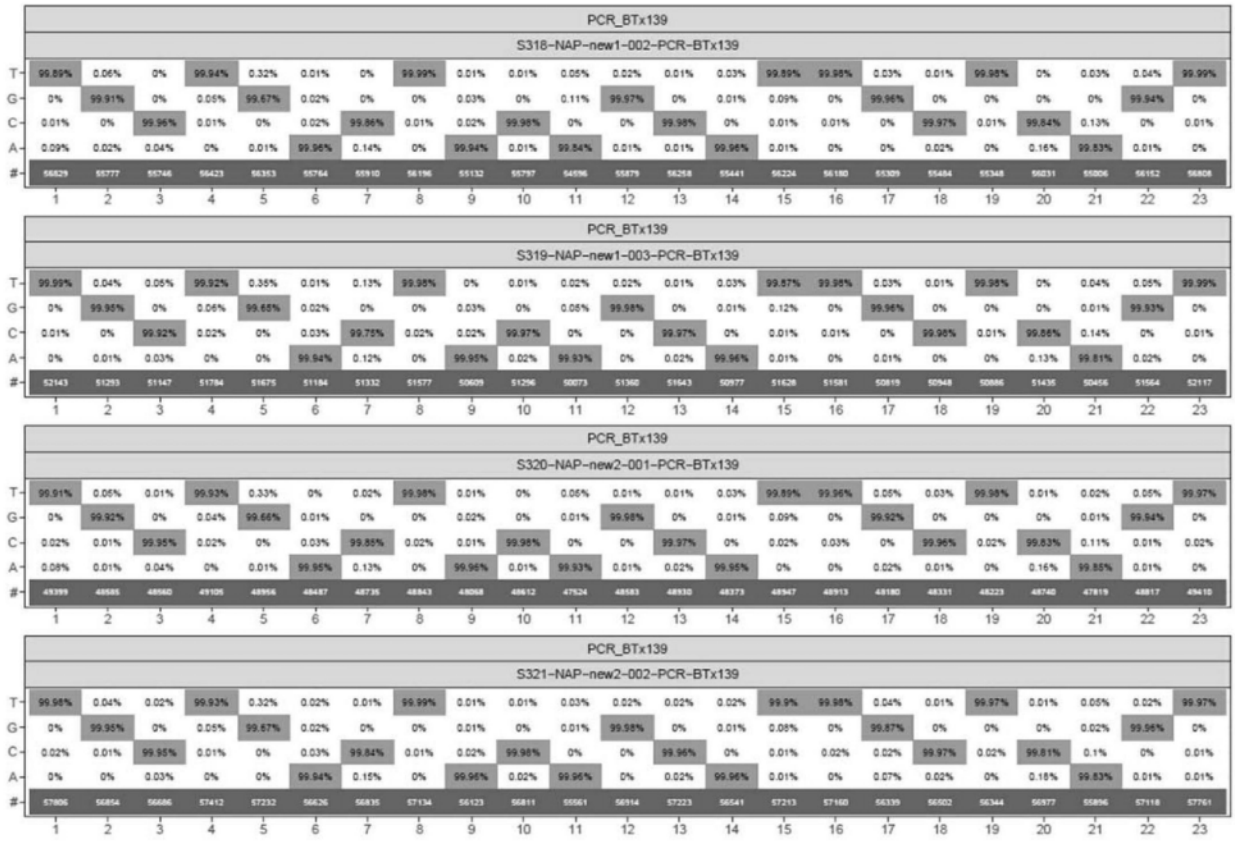

PCR_BTx139
S318-NAP-new1-002-PCR-BTx139
PCR_BTx139
S319-NAP-new1-003-PCR-BTx139
PCR_BTx139
S320-NAP-new2-001-PCR-BTx139
PCR_BTx139
S321-NAP-new2-002-PCR-BTx139
T
G
C
A
#
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

图60Q


PCR_BTx139
S322-NAP-new2-003-PCR-BTx139
PCR_BTx139
S323-NAP-new3-001-PCR-BTx139
PCR_BTx139
S324-NAP-new3-002-PCR-BTx139
PCR_BTx139
S325-NAP-new3-003-PCR-BTx139
PCR_BTx139
S326-NAP-new4-001-PCR-BTx139
PCR_BTx139
S327-NAP-new4-002-PCR-BTx139
PCR_BTx139
S328-NAP-new4-003-PCR-BTx139
PCR_BTx139
S329-NAP-new5-001-PCR-BTx139

图60R


PCR_BTx139
S334-NAP-new6-003-PCR-BTx139
PCR_BTx139
S335-NAP-new7-001-PCR-BTx139
PCR_BTx139
S336-NAP-new7-002-PCR-BTx139
PCR_BTx139
S337-NAP-new7-003-PCR-BTx139
PCR_BTx139
S330-NAP-new5-002-PCR-BTx139
PCR_BTx139
S331-NAP-new5-003-PCR-BTx139
PCR_BTx139
S332-NAP-new6-001-PCR-BTx139
PCR_BTx139
S333-NAP-new6-002-PCR-BTx139

图60S


PCR_BTx139
S342-NAP-new9-002-PCR-BTx139
PCR_BTx139
S343-NAP-new9-003-PCR-BTx139
PCR_BTx139
S344-NAP-new10-001-PCR-BTx139
PCR_BTx139
S345-NAP-new10-002-PCR-BTx139
PCR_BTx139
S338-NAP-new8-001-PCR-BTx139
PCR_BTx139
S339-NAP-new8-002-PCR-BTx139
PCR_BTx139
S340-NAP-new8-003-PCR-BTx139
PCR_BTx139
S341-NAP-new9-001-PCR-BTx139

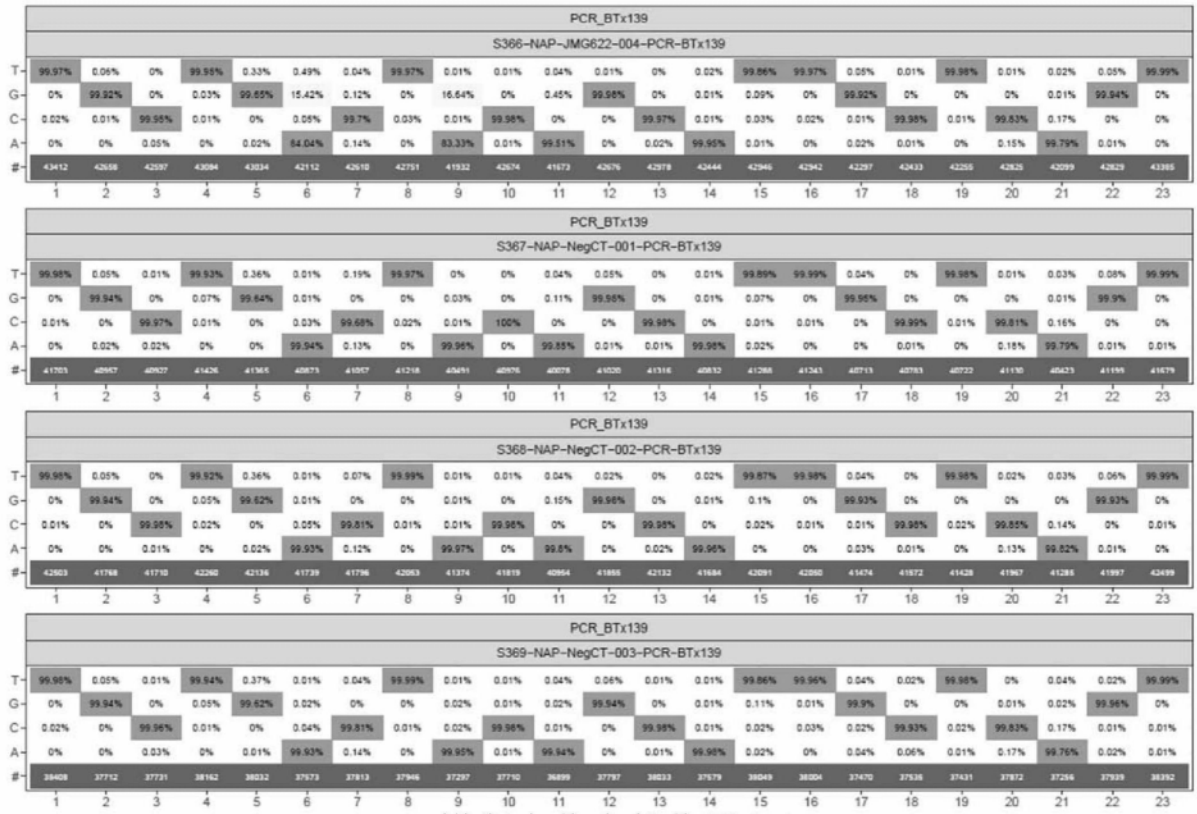
图60T


PCR_BTx139
S350-NAP-new12-001-PCR-BTx139
PCR_BTx139
S351-NAP-new12-002-PCR-BTx139
PCR_BTx139
S352-NAP-new12-003-PCR-BTx139
PCR_BTx139
S353-NAP-new13-001-PCR-BTx139
PCR_BTx139
S346-NAP-new10-003-PCR-BTx139
PCR_BTx139
S347-NAP-new11-001-PCR-BTx139
PCR_BTx139
S348-NAP-new11-002-PCR-BTx139
PCR_BTx139
S349-NAP-new11-003-PCR-BTx139

图60U


PCR_BTx139
S358-NAP-new14-003-PCR-BTx139
PCR_BTx139
S359-NAP-NP237-001-PCR-BTx139
PCR_BTx139
S360-NAP-NP237-002-PCR-BTx139
PCR_BTx139
S361-NAP-NP237-003-PCR-BTx139
PCR_BTx139
S354-NAP-new13-002-PCR-BTx139
PCR_BTx139
S355-NAP-new13-003-PCR-BTx139
PCR_BTx139
S356-NAP-new14-001-PCR-BTx139
PCR_BTx139
S357-NAP-new14-002-PCR-BTx139

图60V


PCR_BTx139
S366-NAP-JMG622-004-PCR-BTx139
PCR_BTx139
S367-NAP-NegCT-001-PCR-BTx139
PCR_BTx139
S368-NAP-NegCT-002-PCR-BTx139
PCR_BTx139
S369-NAP-NegCT-003-PCR-BTx139
沿靶向位点的位置(bp)

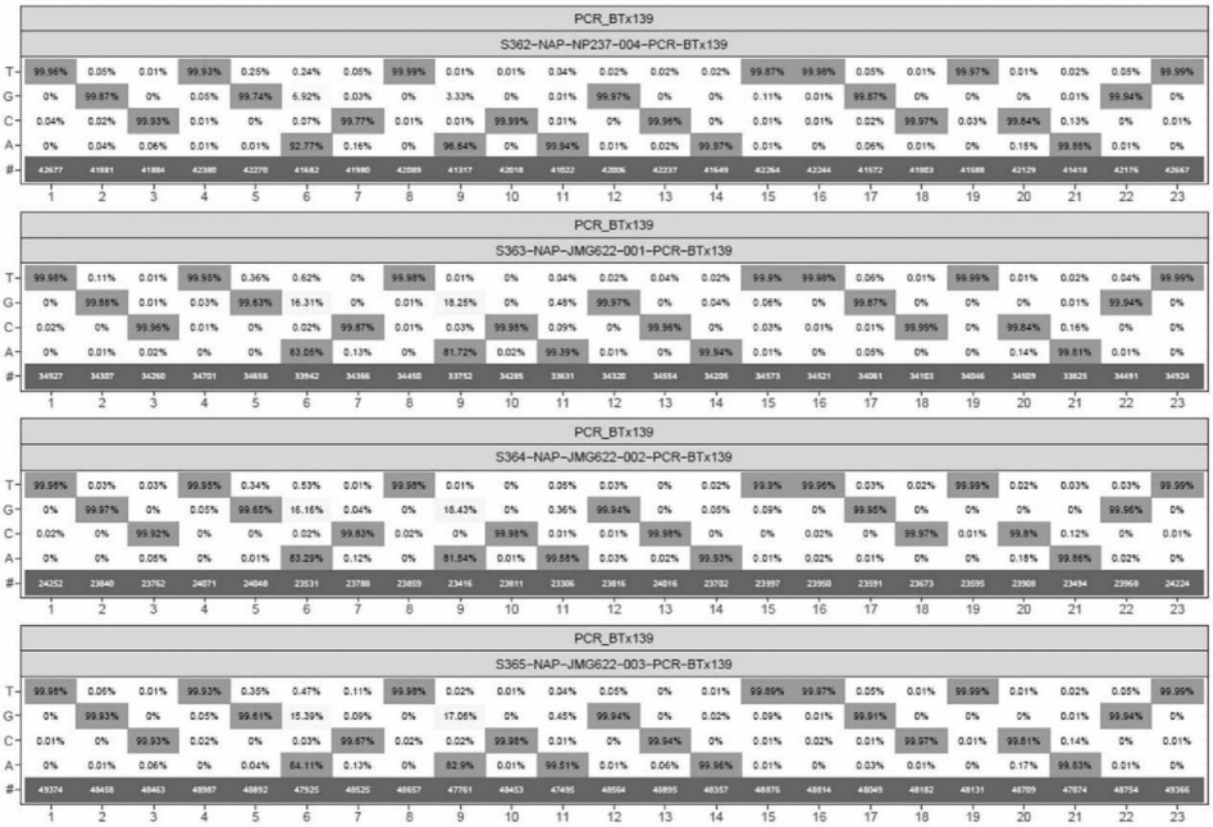

PCR_BTx139
S362-NAP-NP237-004-PCR-BTx139
PCR_BTx139
S363-NAP-JMG622-001-PCR-BTx139
PCR_BTx139
S364-NAP-JMG622-002-PCR-BTx139
PCR_BTx139
S365-NAP-JMG622-003-PCR-BTx139

图60W

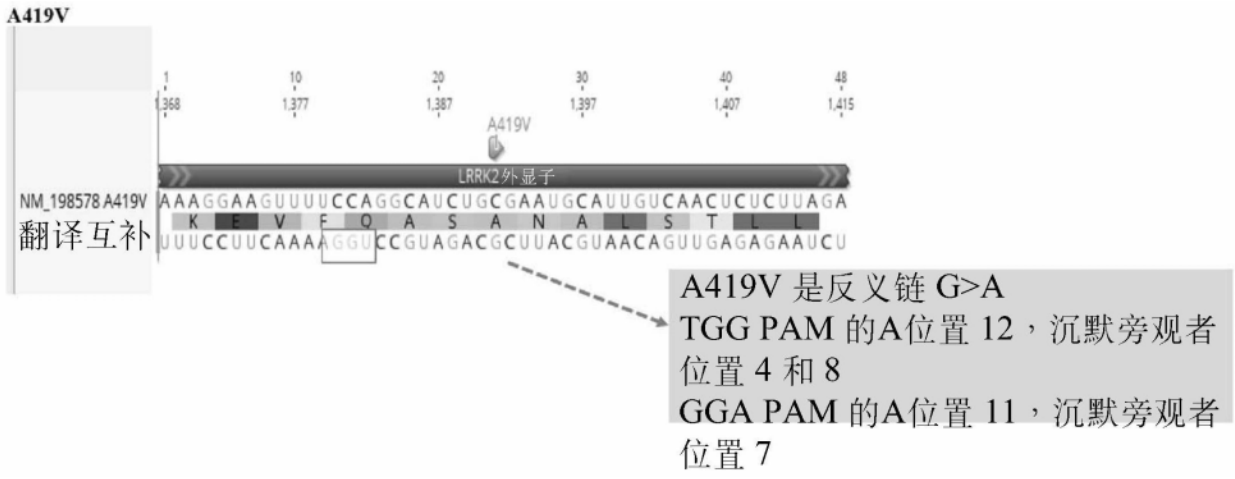

A419V
1
10
20
30
40
48
1,368
1,377
1,387
1,397
1,407
1,415
A419V
LRRK2外显子
NM_198578 A419V
翻译互补
A419V 是反义链 G>A
TGG PAM 的A位置 12，沉默旁观者
位置 4 和 8
GGA PAM 的A位置 11，沉默旁观者
位置 7

图61A

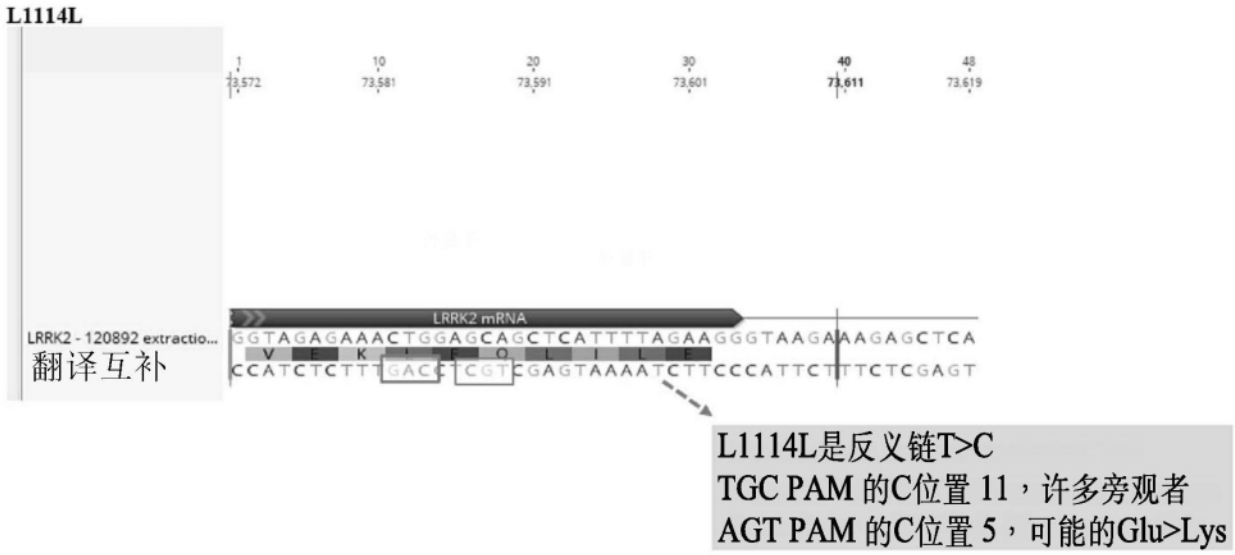

L1114L
1
10
20
30
40
48
73,572
73,581
73,591
73,601
73,611
73,619
LRRK2 mRNA
LRRK2 - 120892 extractio...
翻译互补
L1114L是反义链T>C
TGC PAM 的C位置 11，许多旁观者
AGT PAM 的C位置 5，可能的Glu>Lys

图61B

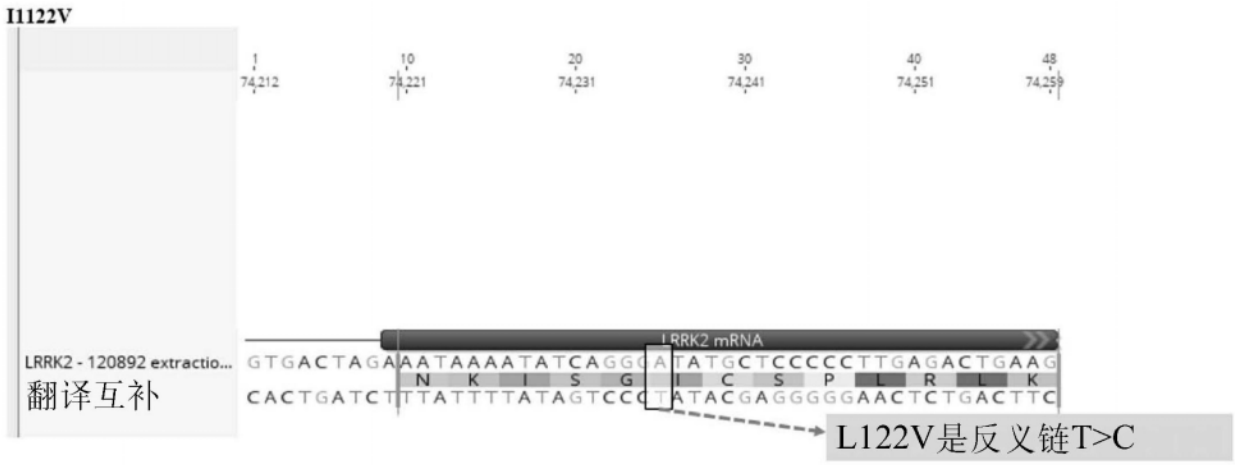


图61C

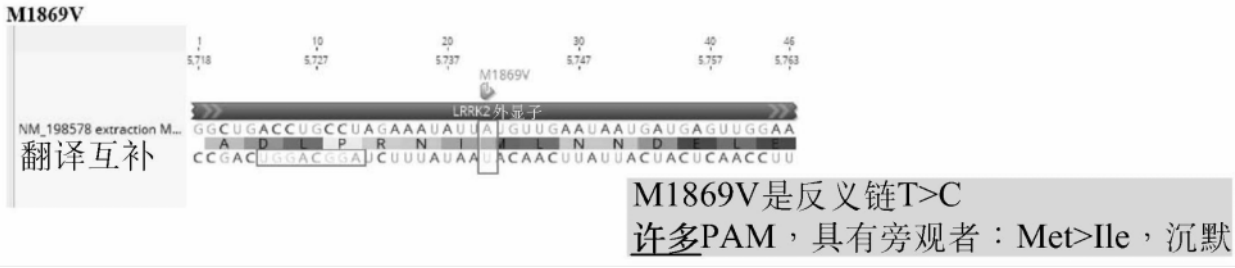


图61D

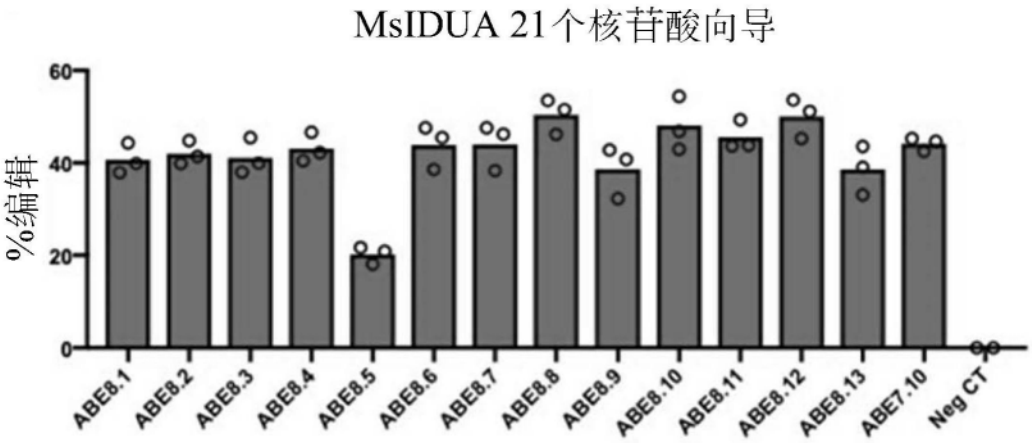


图62A

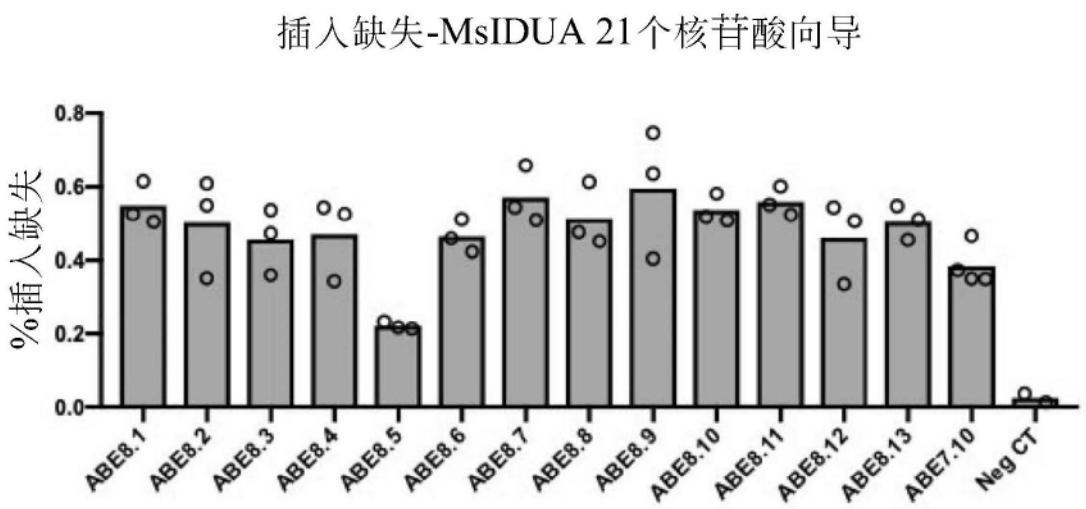

插入缺失-MsIDUA 21个核苷酸向导
%插入缺失
0.8
0.6
0.4
0.2
0.0
ABE8.1
ABE8.2
ABE8.3
ABE8.4
ABE8.5
ABE8.6
ABE8.7
ABE8.8
ABE8.9
ABE8.10
ABE8.11
ABE8.12
ABE8.13
ABE7.10
Neg CT

图62B

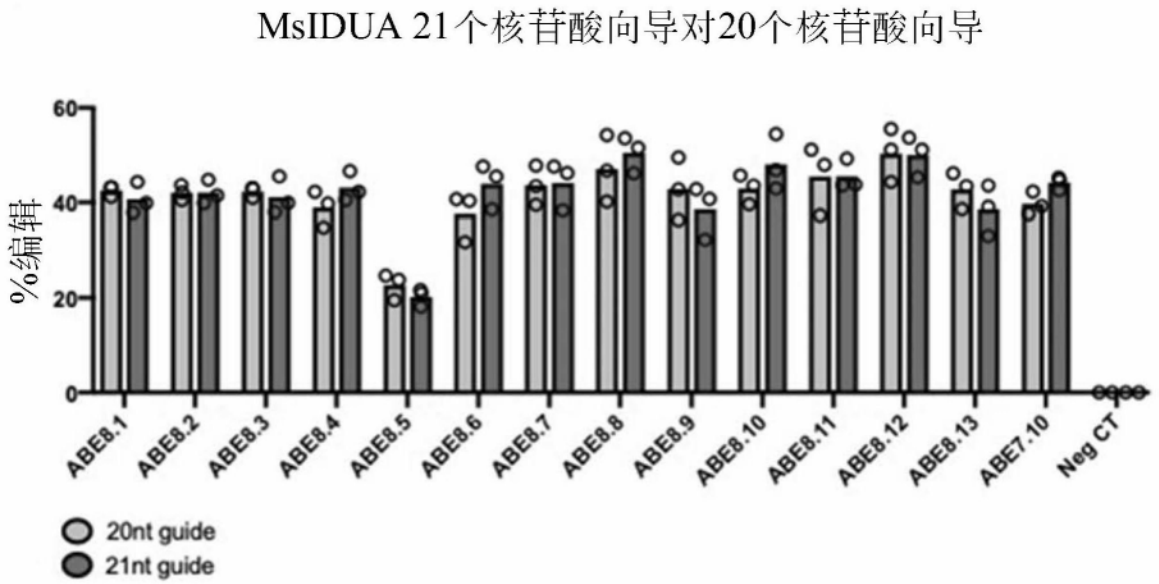

MsIDUA 21个核苷酸向导对20个核苷酸向导
%编辑
60
40
20
0
ABE8.1
ABE8.2
ABE8.3
ABE8.4
ABE8.5
ABE8.6
ABE8.7
ABE8.8
ABE8.9
ABE8.10
ABE8.11
ABE8.12
ABE8.13
ABE7.10
Neg CT
20nt guide
21nt guide

图63

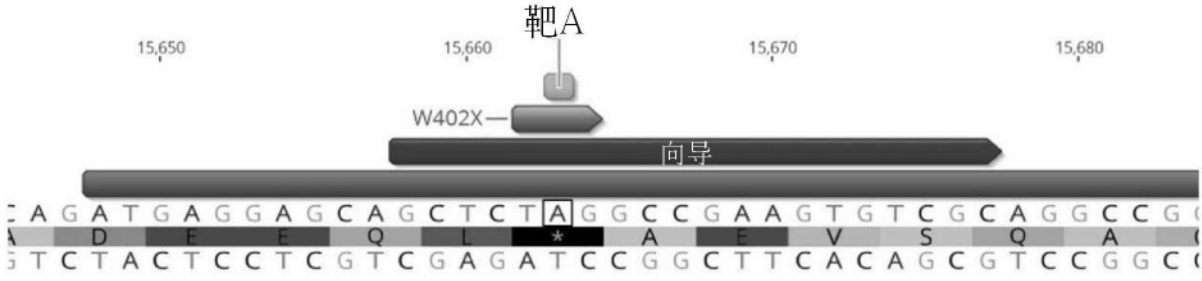

靶A
15,650
15,660
15,670
15,680
W402X
向导

图64

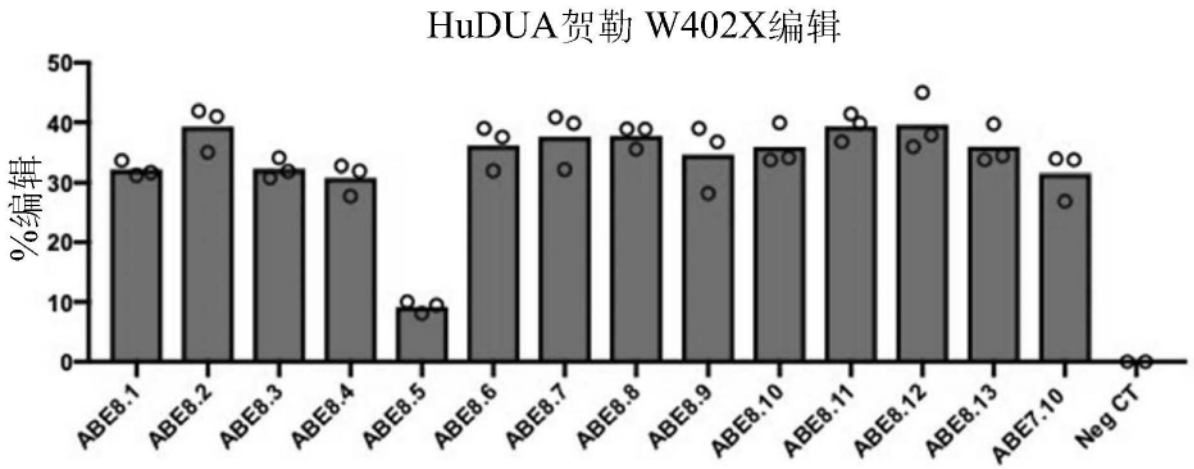

HuDUA贺勒 W402X编辑
%编辑
50
40
30
20
10
0
ABE8.1
ABE8.2
ABE8.3
ABE8.4
ABE8.5
ABE8.6
ABE8.7
ABE8.8
ABE8.9
ABE8.10
ABE8.11
ABE8.12
ABE8.13
ABE7.10
Neg CT

图65A

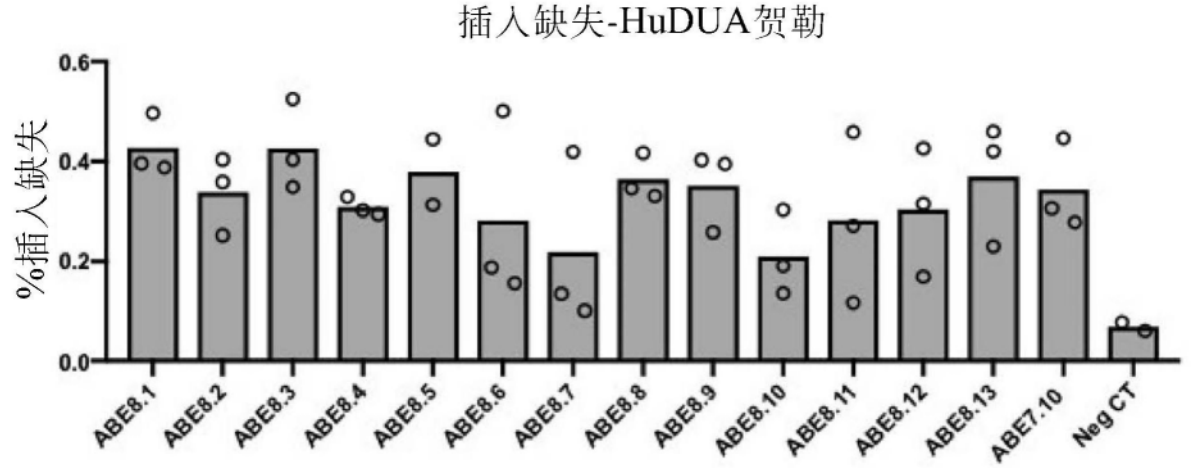

插入缺失-HuDUA贺勒
%插入缺失
0.6
0.4
0.2
0.0
ABE8.1
ABE8.2
ABE8.3
ABE8.4
ABE8.5
ABE8.6
ABE8.7
ABE8.8
ABE8.9
ABE8.10
ABE8.11
ABE8.12
ABE8.13
ABE7.10
Neg CT

图65B

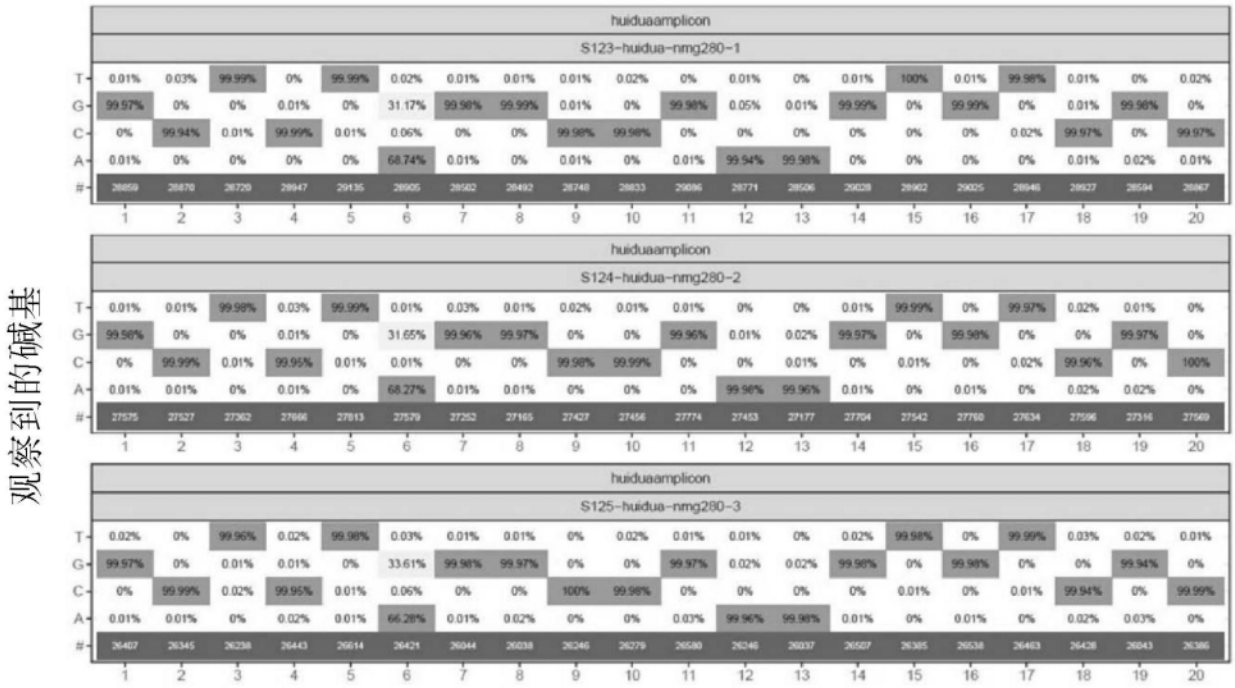

huiduaamplicon
S123-huidua-nmg280-1
huiduaamplicon
S124-huidua-nmg280-2
huiduaamplicon
S125-huidua-nmg280-3
观察到的碱基
沿著目标位点的位置(碱基对)

图66A

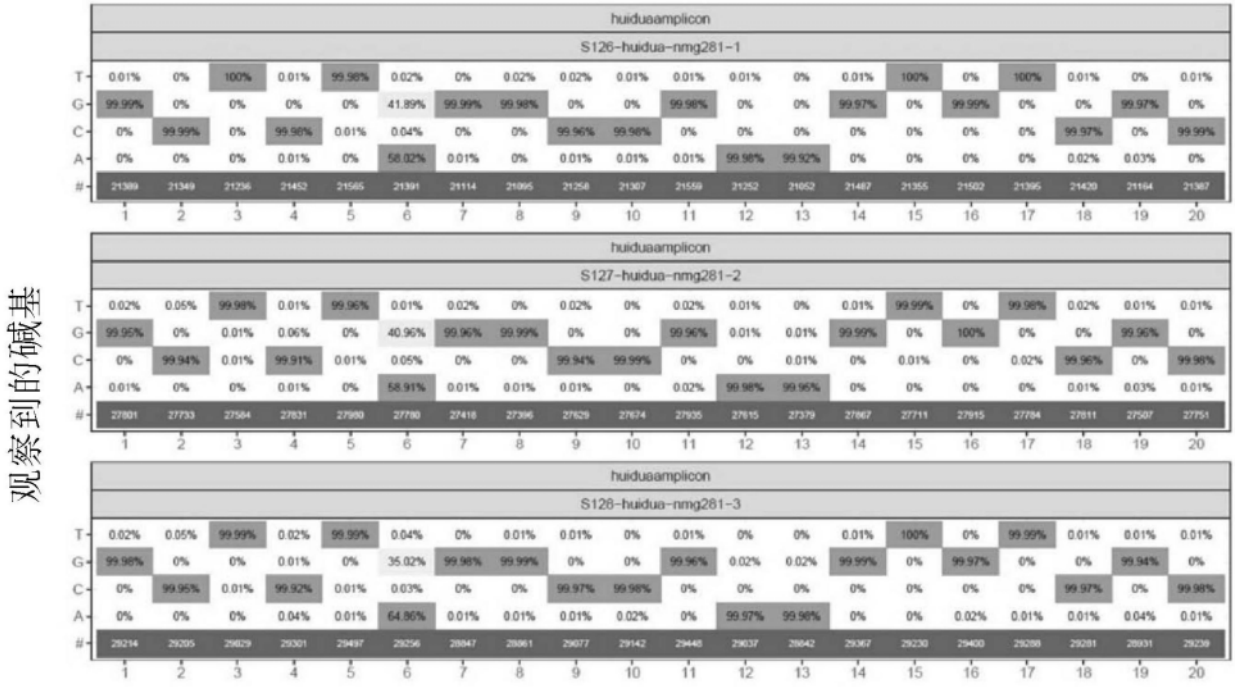

huiduaamplicon
S126-huidua-nmg281-1
huiduaamplicon
S127-huidua-nmg281-2
huiduaamplicon
S128-huidua-nmg281-3
观察到的碱基
沿著目标位点的位置(碱基对)

图66B

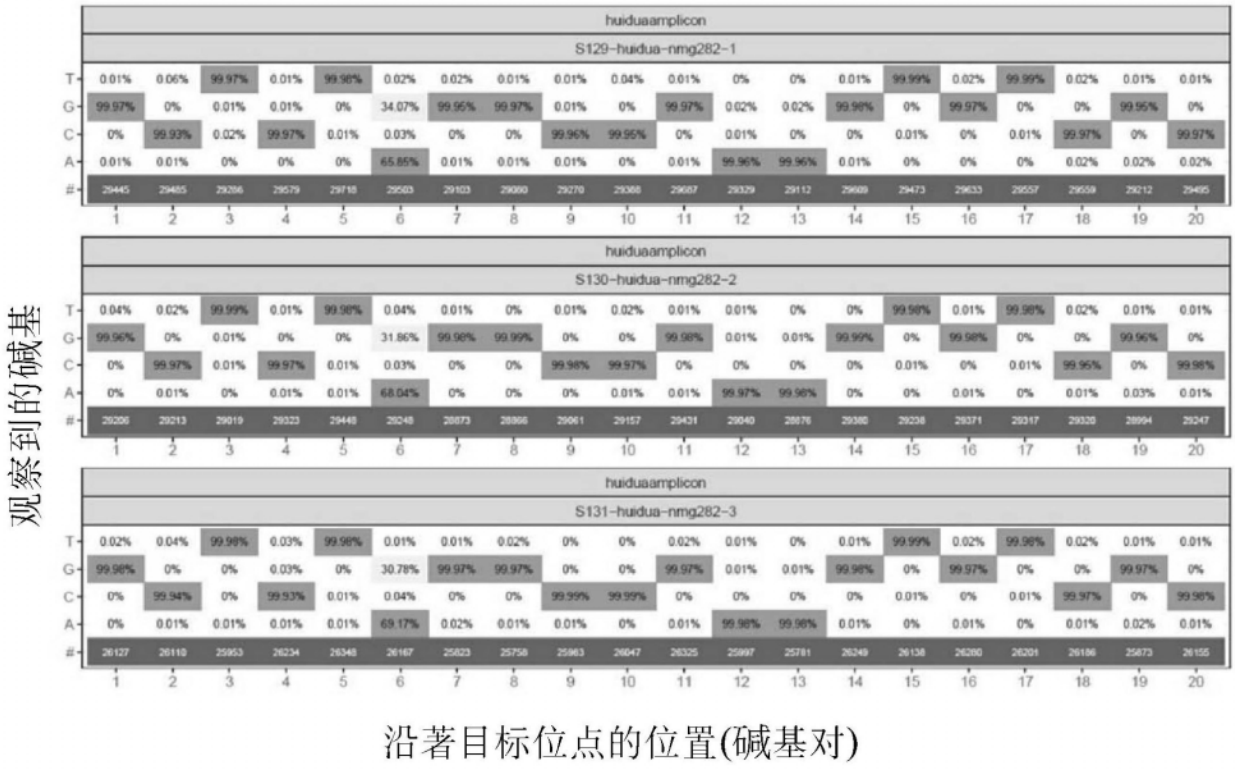

huiduaamplicon
S129-huidua-nmg282-1
huiduaamplicon
S130-huidua-nmg282-2
huiduaamplicon
S131-huidua-nmg282-3
观察到的碱基
沿著目标位点的位置(碱基对)

图66C

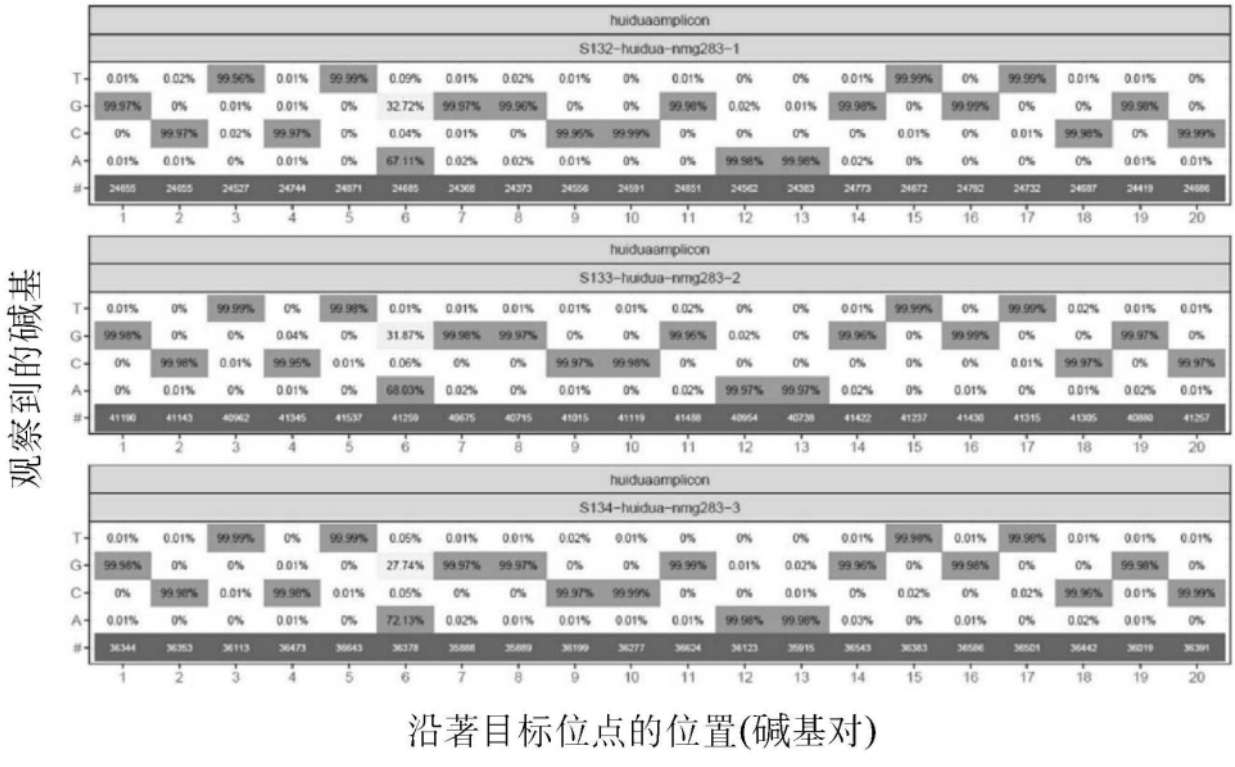

huiduaamplicon
S132-huidua-nmg283-1
huiduaamplicon
S133-huidua-nmg283-2
huiduaamplicon
S134-huidua-nmg283-3
观察到的碱基
沿著目标位点的位置(碱基对)

图66D

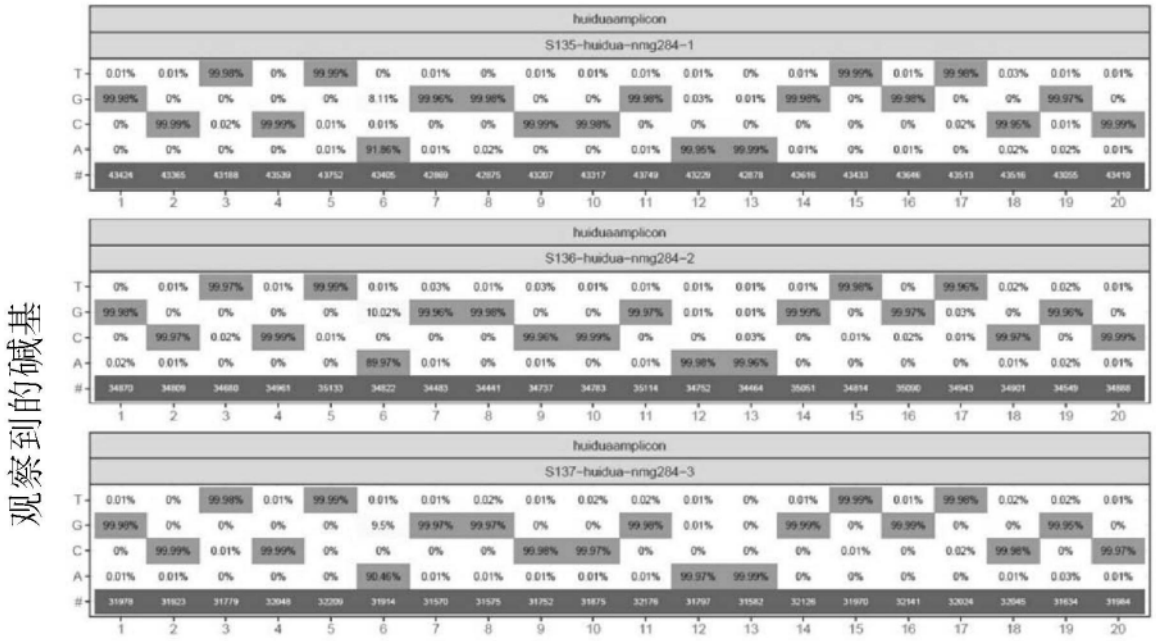

huiduaamplicon
S135-huidua-nmg284-1
huiduaamplicon
S136-huidua-nmg284-2
huiduaamplicon
S137-huidua-nmg284-3
观察到的碱基
沿著目标位点的位置(碱基对)

图66E

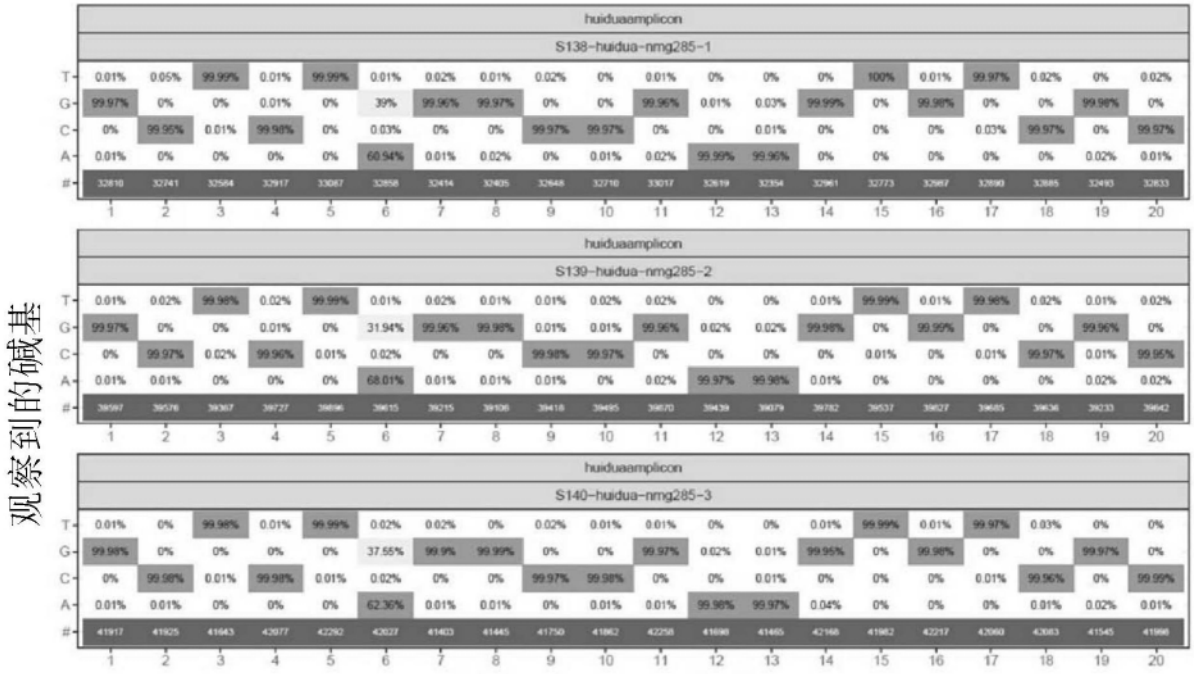

huiduaamplicon
S138-huidua-nmg285-1
huiduaamplicon
S139-huidua-nmg285-2
huiduaamplicon
S140-huidua-nmg285-3
观察到的碱基
沿著目标位点的位置(碱基对)

图66F

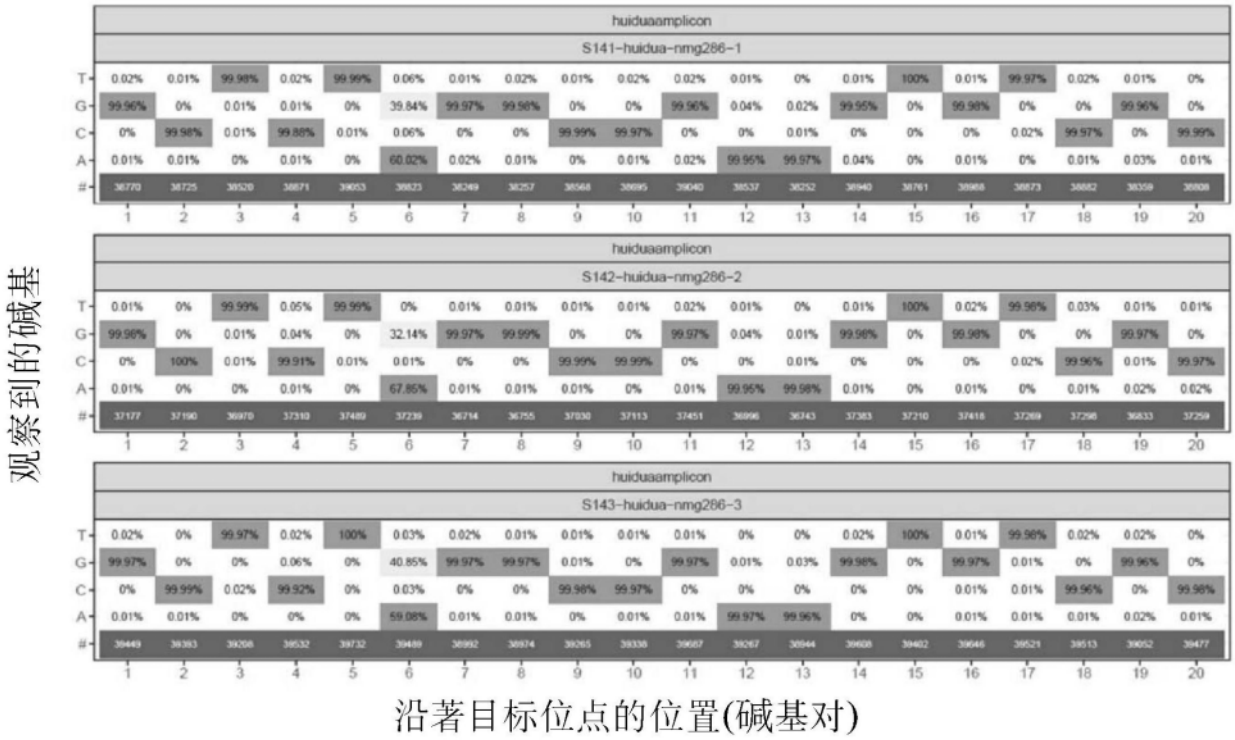

huiduaamplicon
S141-huidua-nmg286-1
huiduaamplicon
S142-huidua-nmg286-2
huiduaamplicon
S143-huidua-nmg286-3
观察到的碱基
沿著目标位点的位置(碱基对)

图66G

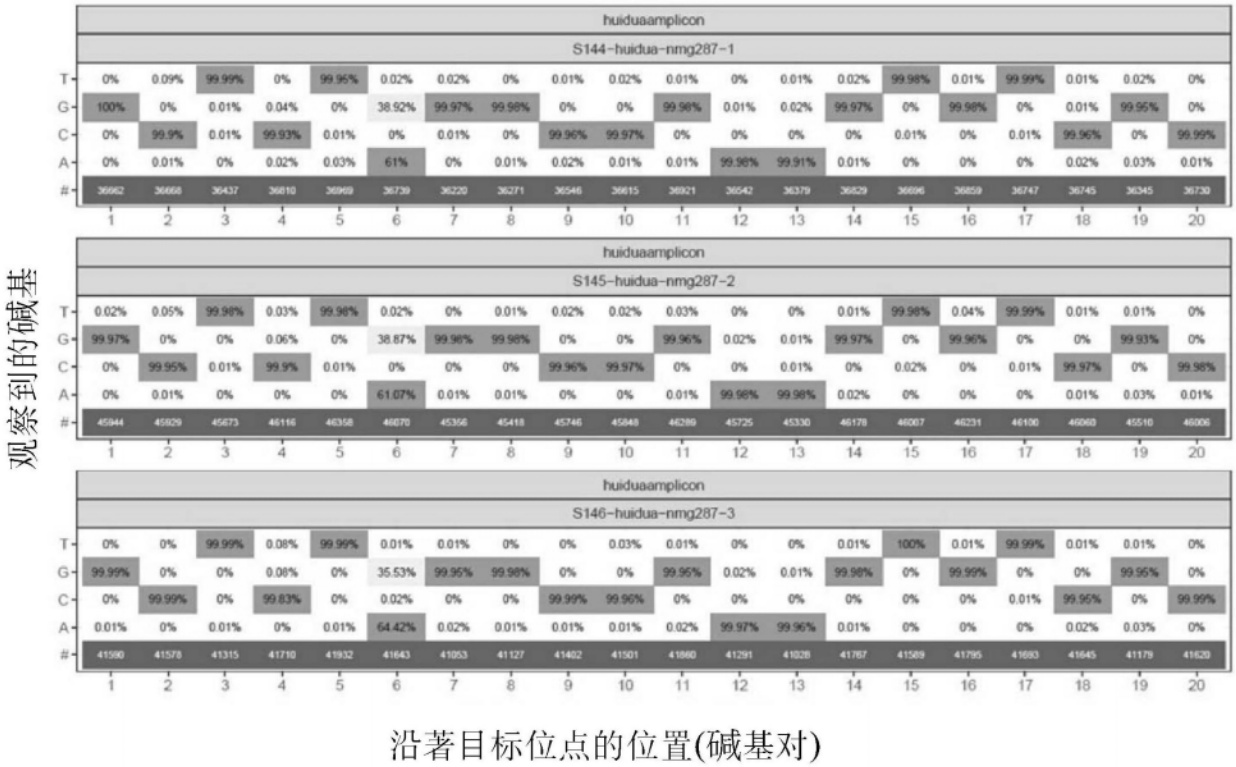

huiduaamplicon
S144-huidua-nmg287-1
huiduaamplicon
S145-huidua-nmg287-2
huiduaamplicon
S146-huidua-nmg287-3
观察到的碱基
沿著目标位点的位置(碱基对)

图66H

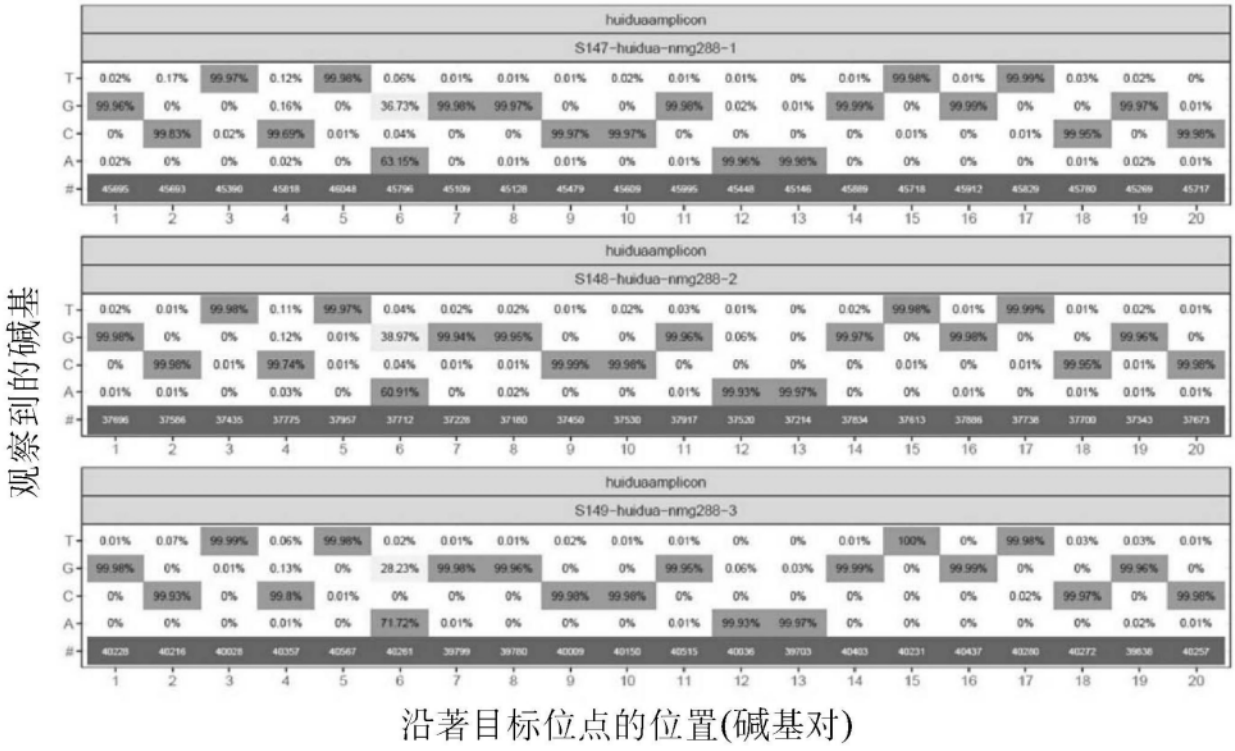

huiduaamplicon
S147-huidua-nmg288-1
huiduaamplicon
S148-huidua-nmg288-2
huiduaamplicon
S149-huidua-nmg288-3
观察到的碱基
沿著目标位点的位置(碱基对)

图66I


huiduaamplicon
S150-huidua-nmg289-1
huiduaamplicon
S151-huidua-nmg289-2
huiduaamplicon
S152-huidua-nmg289-3
观察到的碱基
沿著目标位点的位置(碱基对)

图66J

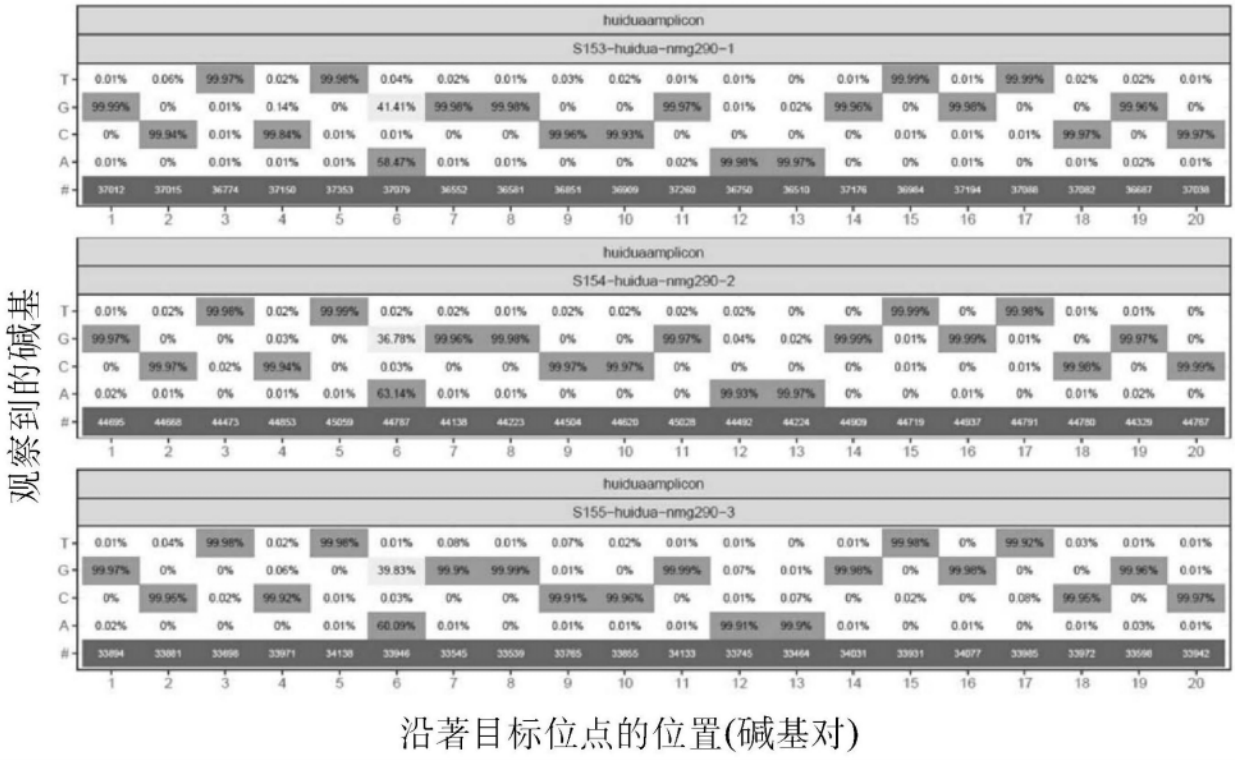

huiduaamplicon
S153-huidua-nmg290-1
huiduaamplicon
S154-huidua-nmg290-2
huiduaamplicon
S155-huidua-nmg290-3
观察到的碱基
沿著目标位点的位置(碱基对)

图66K

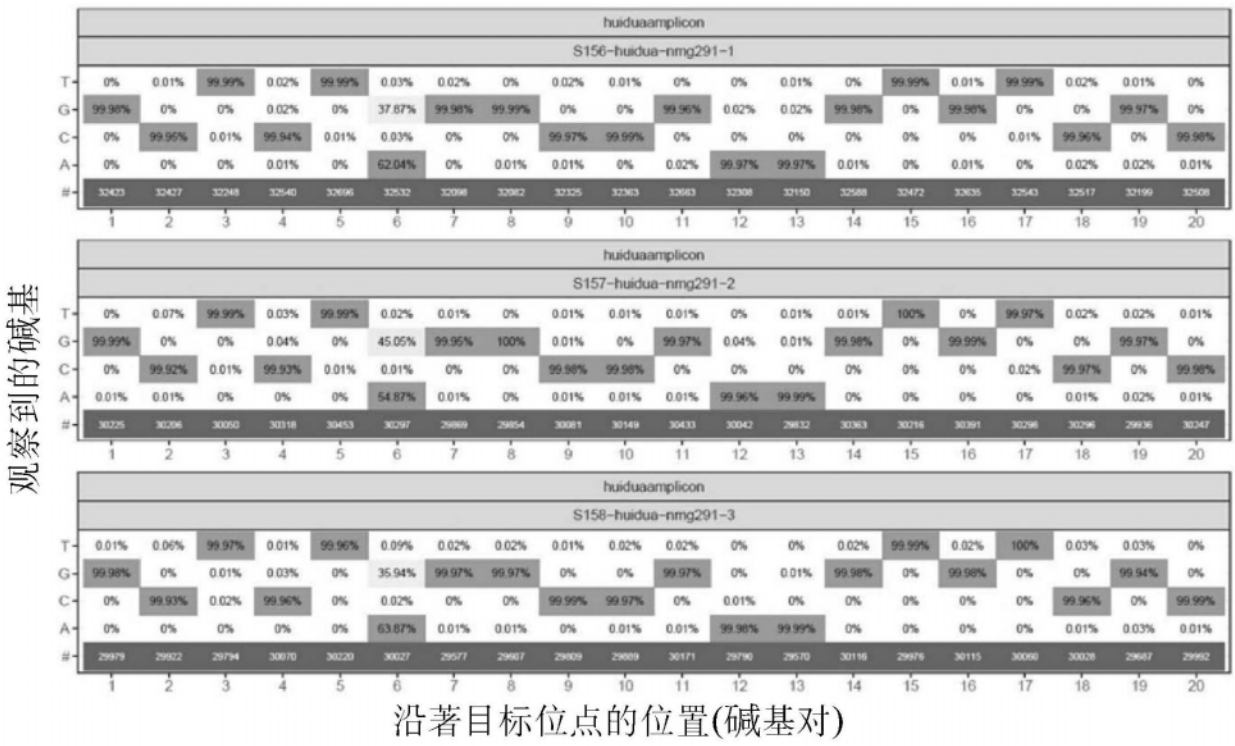

huiduaamplicon
S156-huidua-nmg291-1
huiduaamplicon
S157-huidua-nmg291-2
huiduaamplicon
S158-huidua-nmg291-3
观察到的碱基
沿著目标位点的位置(碱基对)

图66L

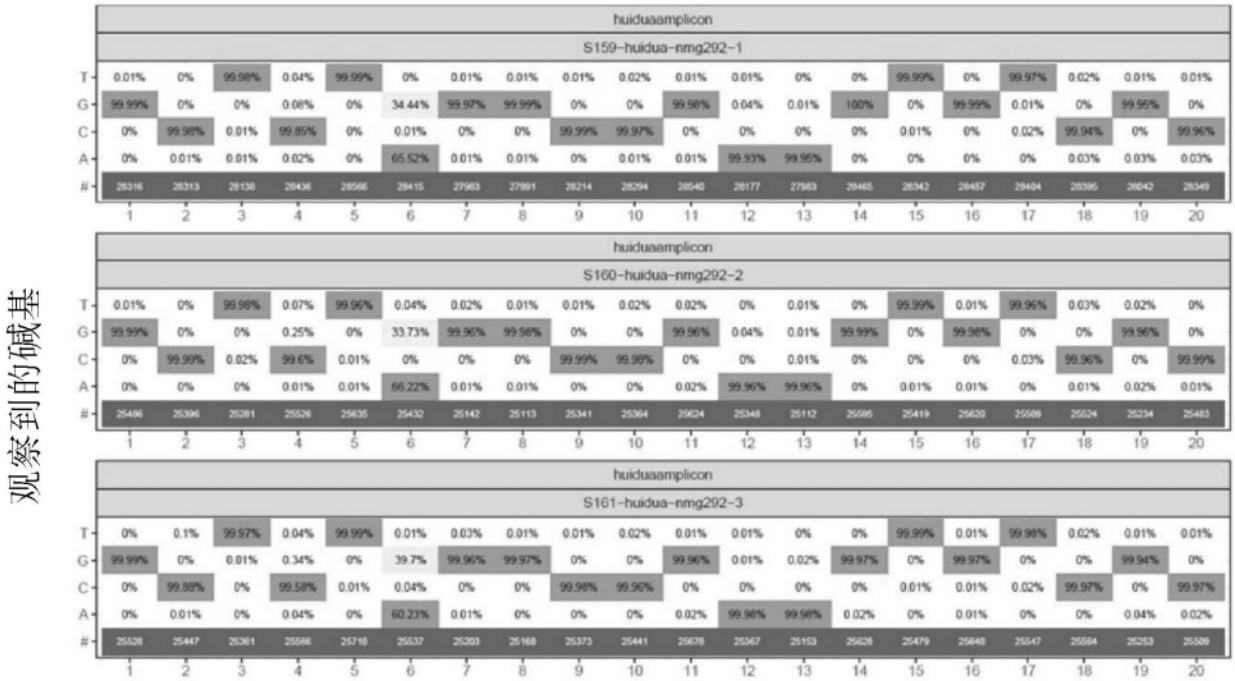

huiduaamplicon
S159-huidua-nmg292-1
huiduaamplicon
S160-huidua-nmg292-2
huiduaamplicon
S161-huidua-nmg292-3
观察到的碱基
沿著目标位点的位置(碱基对)

图66M

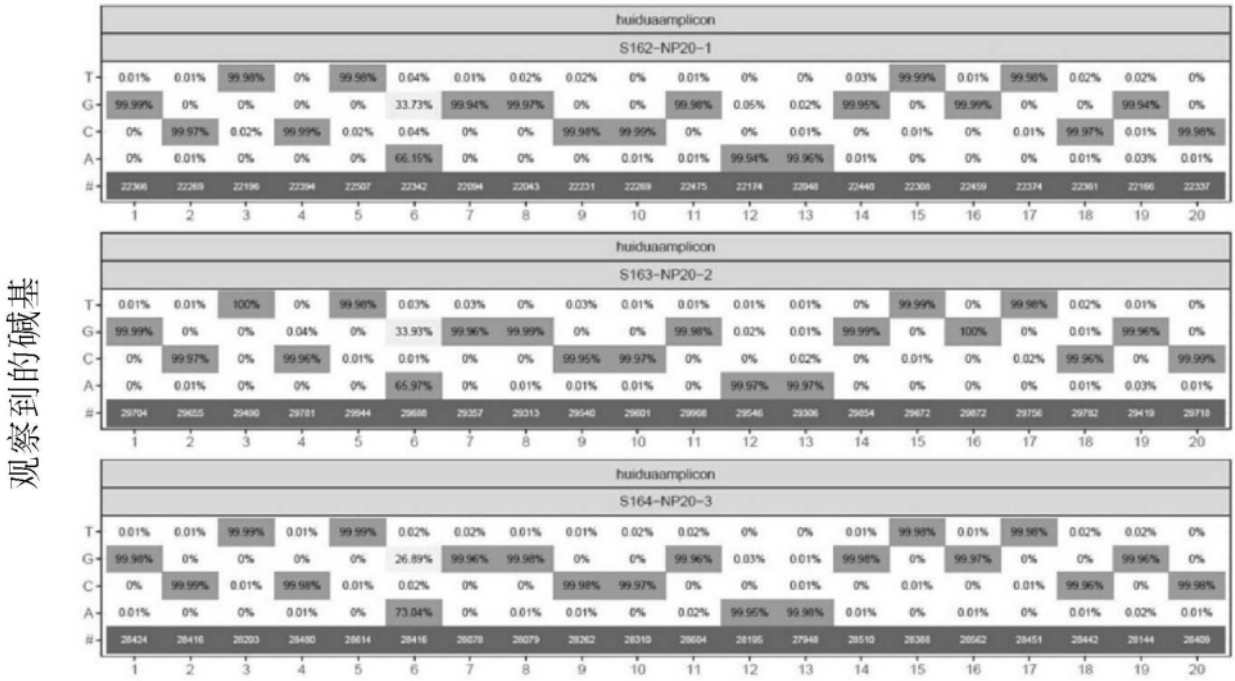

huiduaamplicon
S162-NP20-1
huiduaamplicon
S163-NP20-2
huiduaamplicon
S164-NP20-3
观察到的碱基
沿著目标位点的位置(碱基对)

图66N

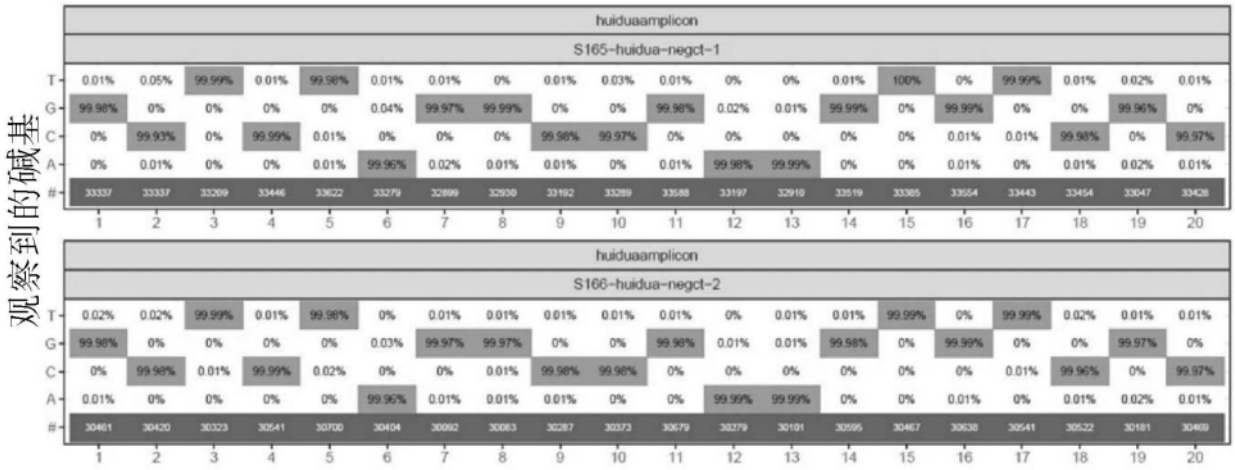


图660

突变	间隔	PAM	Lenti-HEK筛选
R255X	CTTTTCACTTCCTGCCGGGG	NGTT	第一低编辑 第二(优化)：新变体 高达 35%
R106W	AGCTTCCATGTCCAGCCTTC	NGG	高达35%的编辑, ABE8.9
T158M	ACCATGAAGTCAAAATCATT	NGG	低目标，高旁观者
R270X	GCTTTCAGCCCCGTTTCTTG	NGG	没有编辑，设计问题?

图67

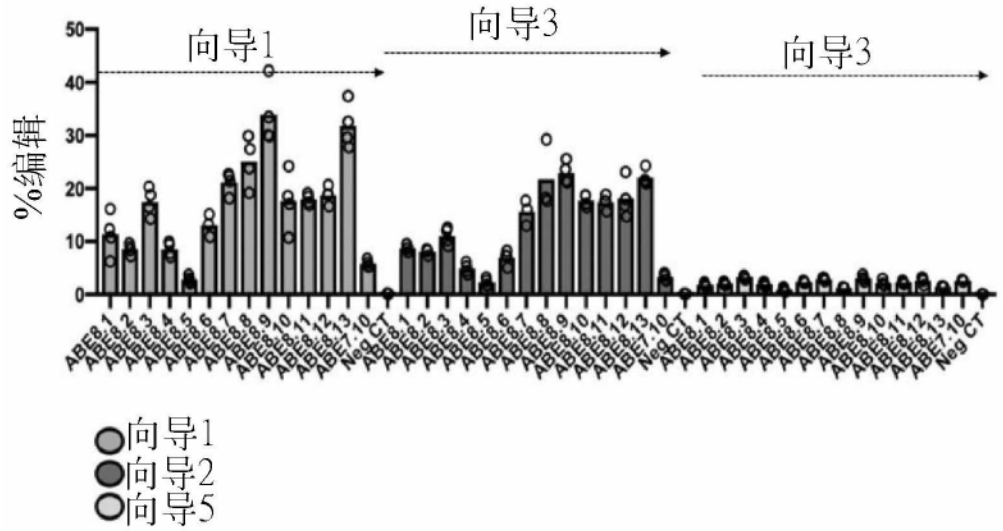


图68

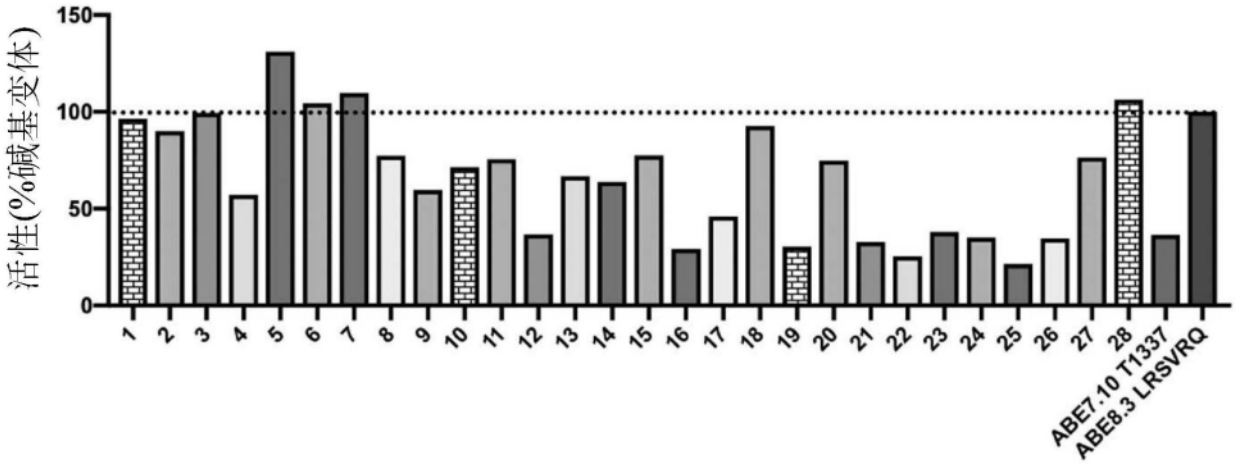


图69

突变	间隔	PAM	Lenti-HEK筛选	患者成纤维细胞	小鼠模型
W402X	CTCTAGGCCGAAGTGTCGC	NGG	高达40%具有多个ABE8.X	mRNA EP 后高达 80% 的校正和恢复的酶活性	在应用干细胞中进行中(2 F1幼崽)

图70A

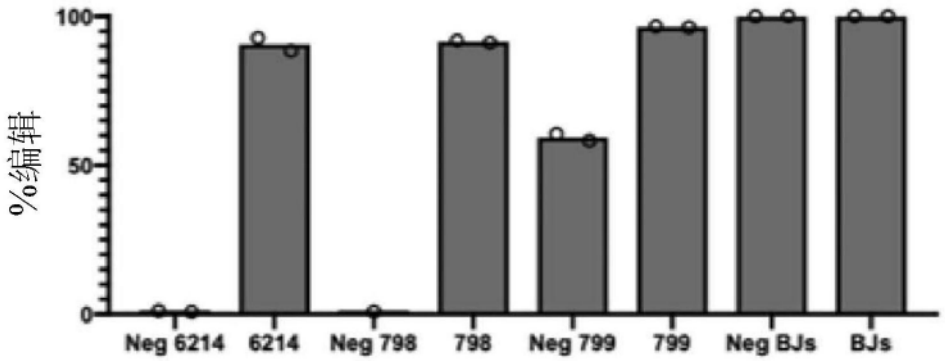

%编辑
100
50
0
Neg 6214
6214
Neg 798
798
Neg 799
799
Neg BJs
BJs

图70B

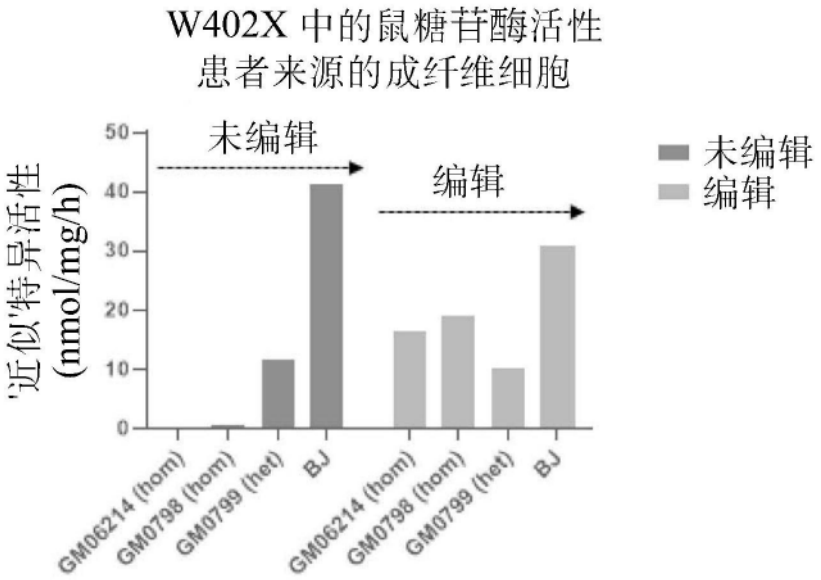

W402X 中的鼠糖苷酶活性
患者来源的成纤维细胞
'近似'特异活性
(nmol/mg/h)
50
40
30
20
10
0
未编辑
编辑
未编辑
编辑
GM06214 (hom)
GM0798 (hom)
GM0799 (het)
BJ
GM06214 (hom)
GM0798 (hom)
GM0799 (het)
BJ

图70C

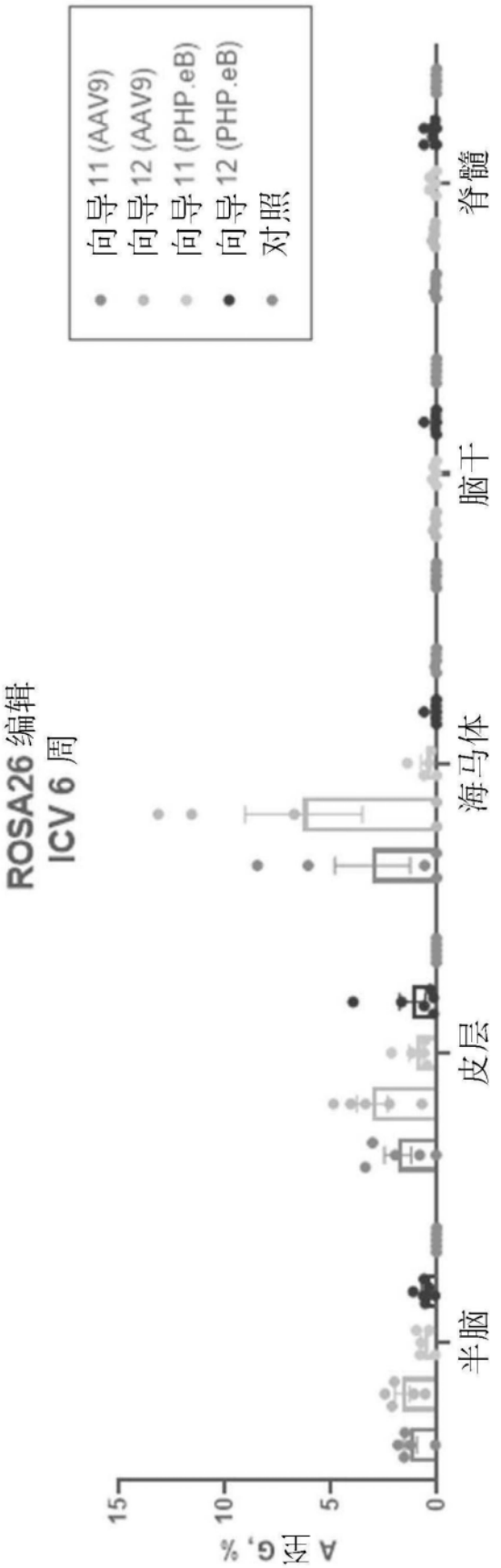

向导11 (AAV9)
向导12 (AAV9)
向导11 (PHP.eB)
向导12 (PHP.eB)
对照
ROSA26 编辑
ICV 6 周
A 至 G, %
15
10
5
0
半脑
皮层
海马体
脑干
脊髓

图71

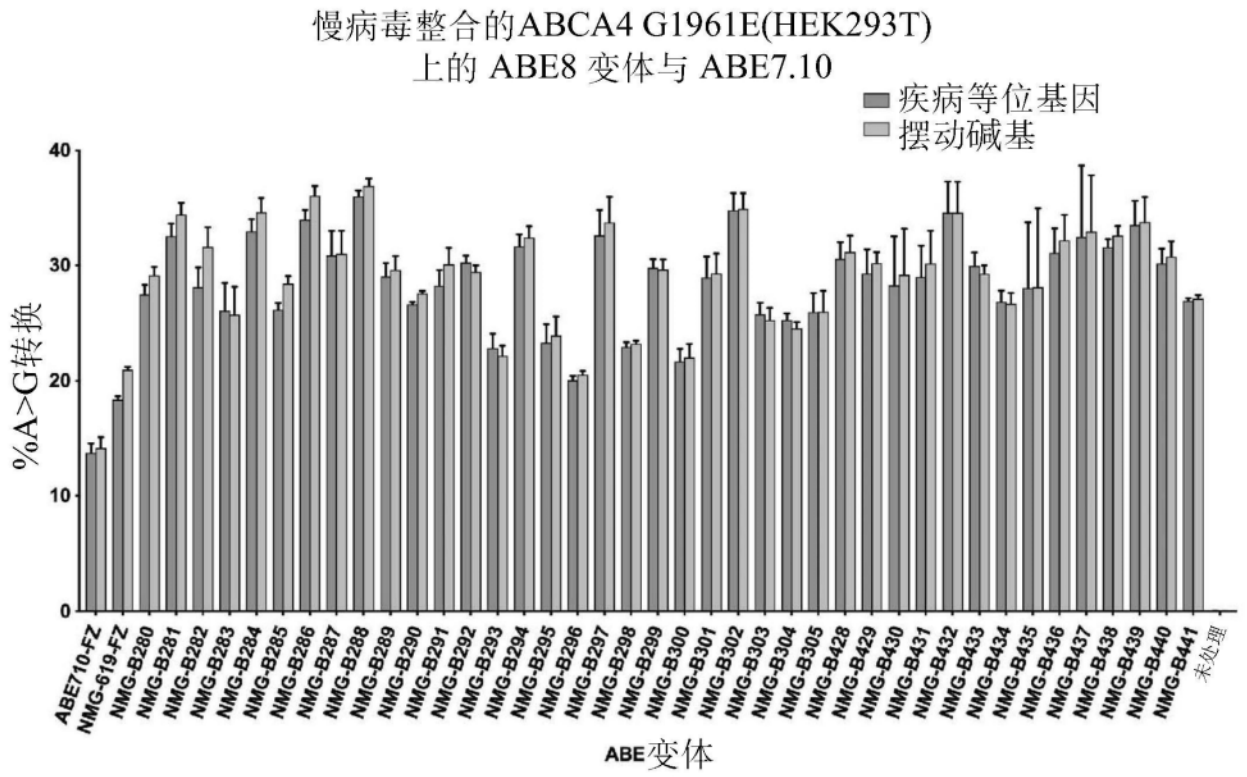

慢病毒整合的ABCA4 G1961E(HEK293T)
上的 ABE8 变体与 ABE7.10
疾病等位基因
摆动碱基
%A>G转换
40
30
20
10
0
ABE710-FZ
NMG-619-FZ
NMG-B280
NMG-B281
NMG-B282
NMG-B283
NMG-B284
NMG-B285
NMG-B286
NMG-B287
NMG-B288
NMG-B289
NMG-B290
NMG-B291
NMG-B292
NMG-B293
NMG-B294
NMG-B295
NMG-B296
NMG-B297
NMG-B298
NMG-B299
NMG-B300
NMG-B301
NMG-B302
NMG-B303
NMG-B304
NMG-B305
NMG-B428
NMG-B429
NMG-B430
NMG-B431
NMG-B432
NMG-B433
NMG-B434
NMG-B435
NMG-B436
NMG-B437
NMG-B438
NMG-B439
NMG-B440
NMG-B441
未处理
ABE变体

图72A

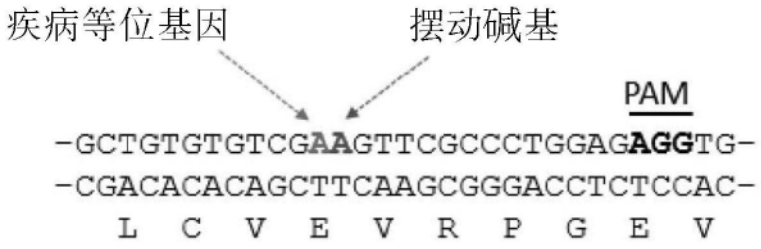

疾病等位基因
摆动碱基
PAM
-GCTGTGTGTCGAAGTTCGCCCTGGAGAGGTG-
-CGACACACAGCTTCAAGCGGGACCTCTCCAC-
L C V E V R P G E V

图72B

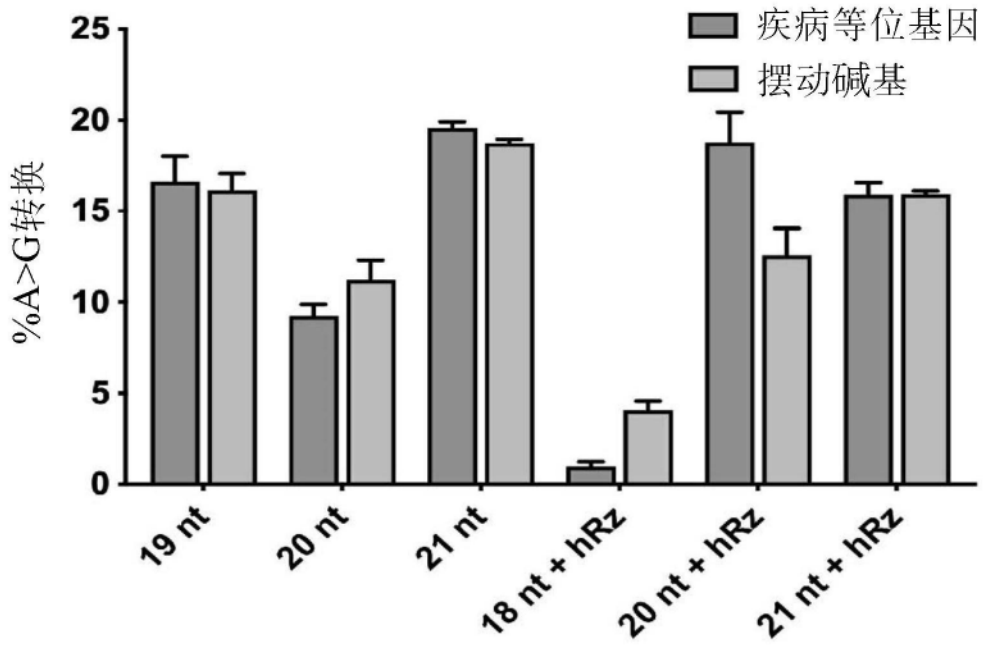

疾病等位基因
摆动碱基
%A>G转换
25
20
15
10
5
0
19 nt
20 nt
21 nt
18 nt + hRz
20 nt + hRz
21 nt + hRz

图73

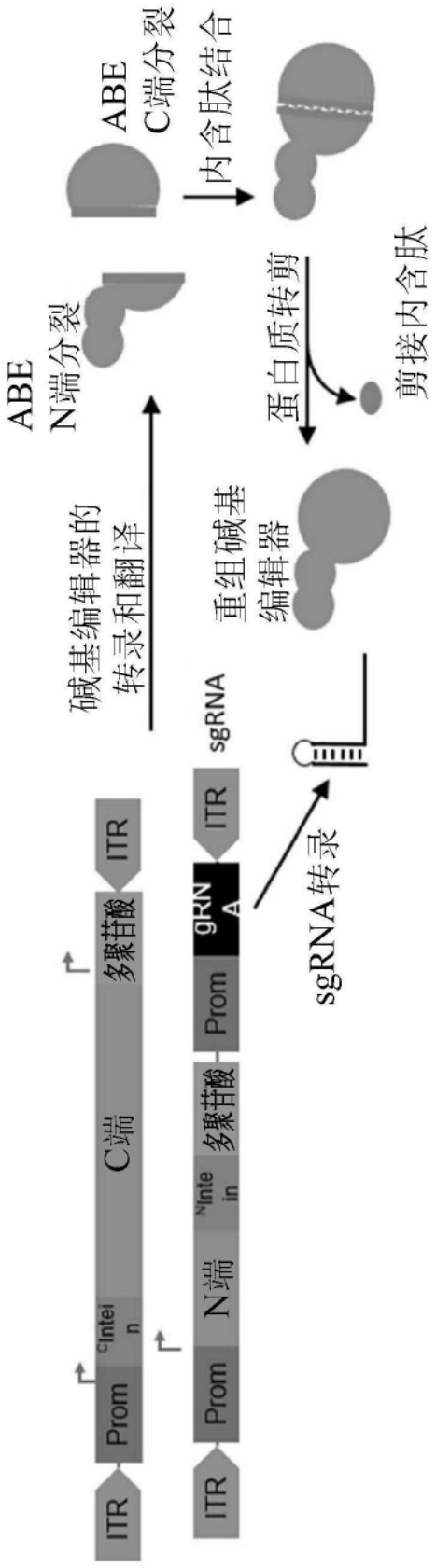

ITR
Prom
cIntein
C端
多聚苷酸
ITR
Prom
N端
NIntein
多聚苷酸
Prom
gRNA
ITR
sgRNA
sgRNA转录
碱基编辑器的转录和翻译
ABE
N端分裂
重组碱基编辑器
蛋白质转剪
剪接内含肽
ABE
C端分裂
内含肽结合

图74

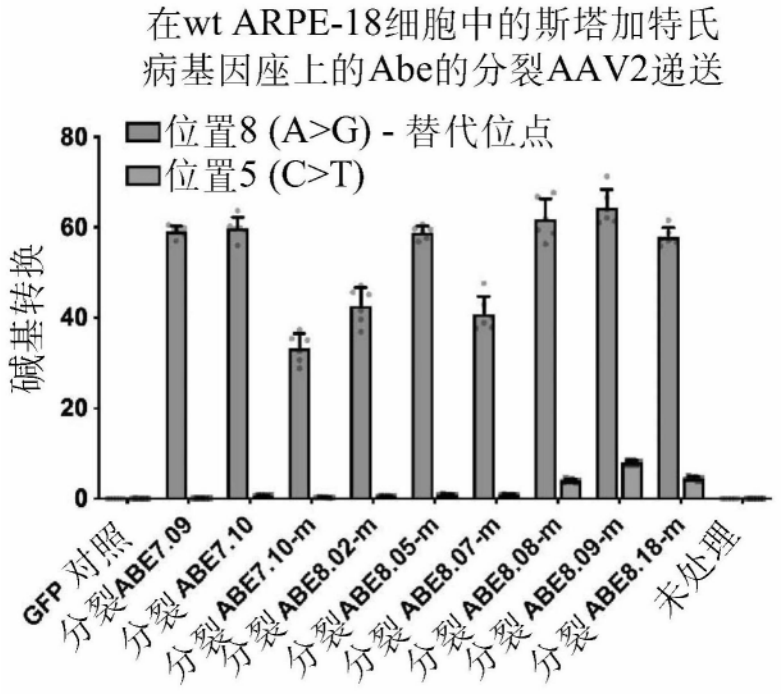

在wt ARPE-18细胞中的斯塔加特氏
病基因座上的Abe的分裂AAV2递送
位置8 (A>G) - 替代位点
位置5 (C>T)
碱基转换
80
60
40
20
0
GFP 对照
分裂ABE7.09
分裂ABE7.10
分裂ABE7.10-m
分裂ABE8.02-m
分裂ABE8.05-m
分裂ABE8.07-m
分裂ABE8.08-m
分裂ABE8.09-m
分裂ABE8.18-m
未处理

图75A

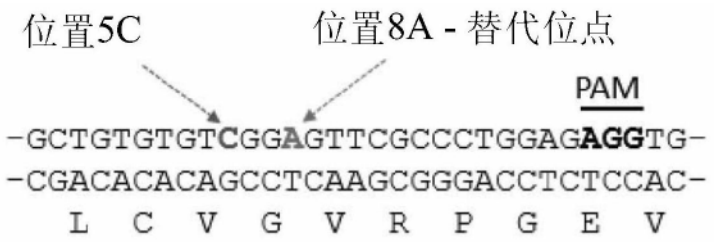

位置5C
位置8A - 替代位点
PAM
-GCTGTGTGTCGGAGTTCGCCCTGGAGAGGTG-
-CGACACACAGCCTCAAGCGGGACCTCTCCAC-
L C V G V R P G E V

图75B

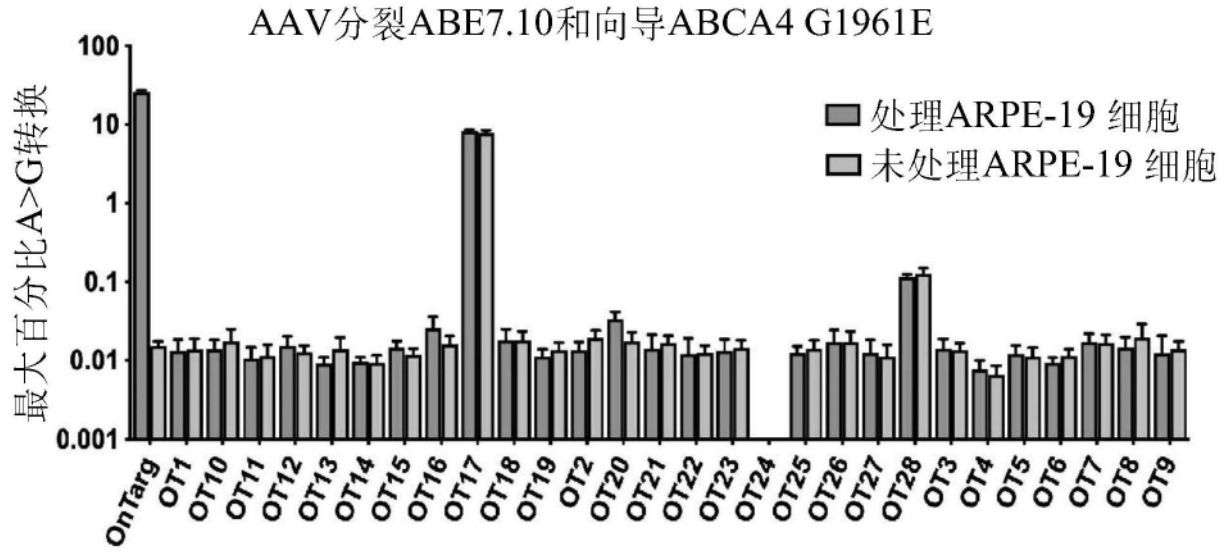

AAV分裂ABE7.10和向导ABCA4 G1961E
最大百分比A>G转换
处理ARPE-19 细胞
未处理ARPE-19 细胞
100
10
1
0.1
0.01
0.001
OnTarg
OT1
OT10
OT11
OT12
OT13
OT14
OT15
OT16
OT17
OT18
OT19
OT2
OT20
OT21
OT22
OT23
OT24
OT25
OT26
OT27
OT28
OT3
OT4
OT5
OT6
OT7
OT8
OT9

图76A

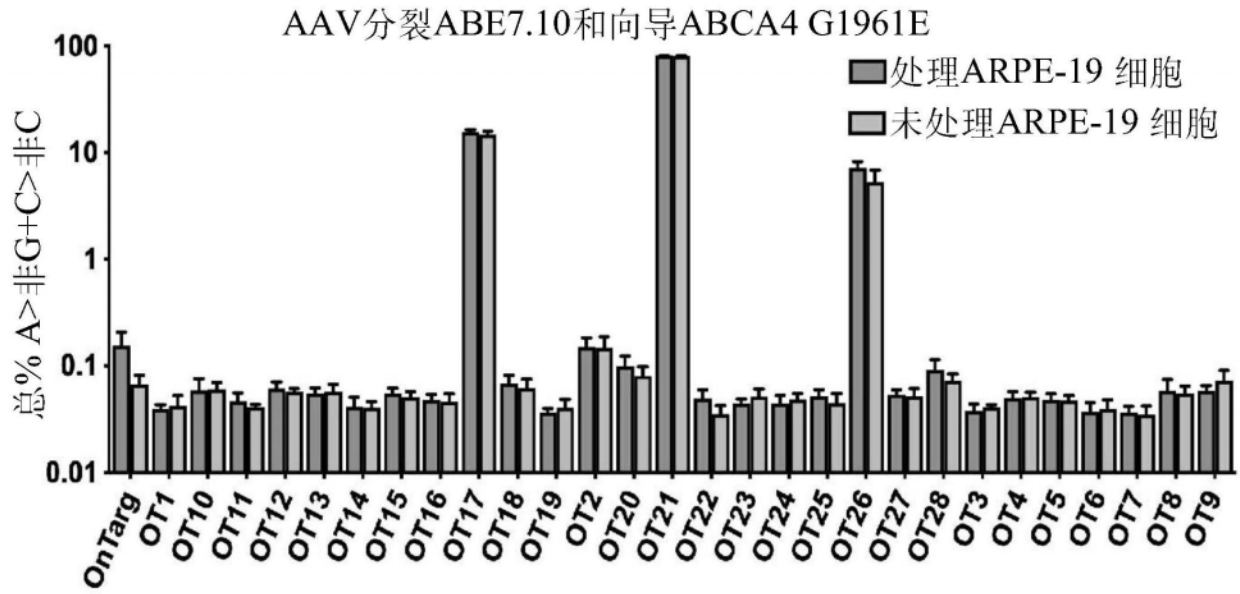

AAV分裂ABE7.10和向导ABCA4 G1961E
总% A>非G+C>非C
处理ARPE-19 细胞
未处理ARPE-19 细胞
100
10
1
0.1
0.01
OnTarg
OT1
OT10
OT11
OT12
OT13
OT14
OT15
OT16
OT17
OT18
OT19
OT2
OT20
OT21
OT22
OT23
OT24
OT25
OT26
OT27
OT28
OT3
OT4
OT5
OT6
OT7
OT8
OT9

图76B

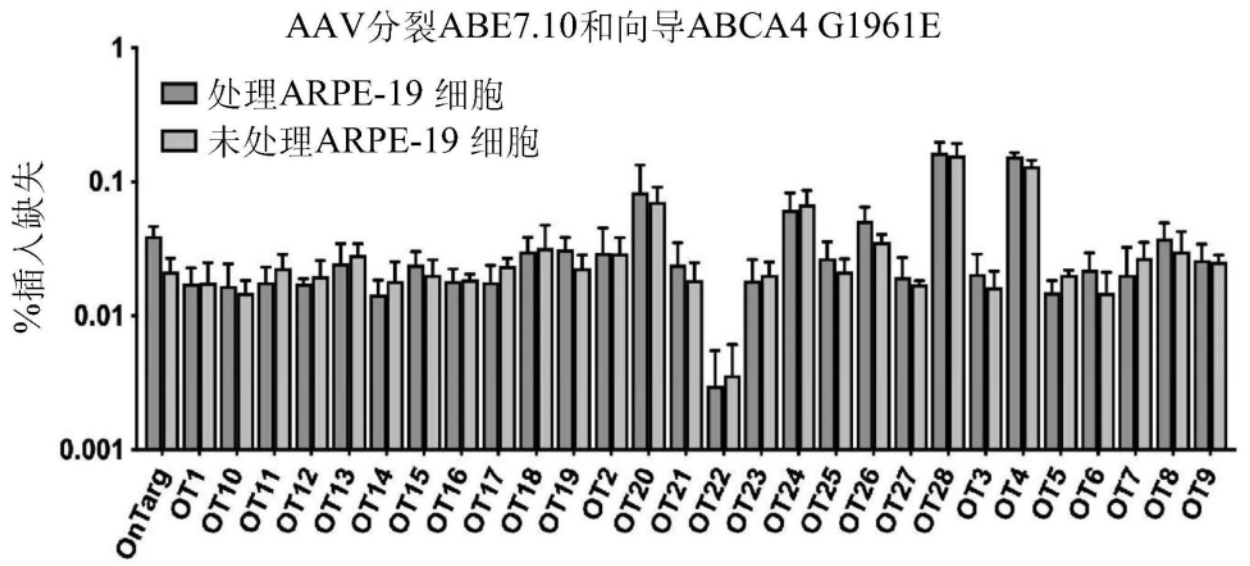

AAV分裂ABE7.10和向导ABCA4 G1961E
处理ARPE-19 细胞
未处理ARPE-19 细胞
%插入缺失
1
0.1
0.01
0.001
OnTarg
OT1
OT10
OT11
OT12
OT13
OT14
OT15
OT16
OT17
OT18
OT19
OT2
OT20
OT21
OT22
OT23
OT24
OT25
OT26
OT27
OT28
OT3
OT4
OT5
OT6
OT7
OT8
OT9

图77

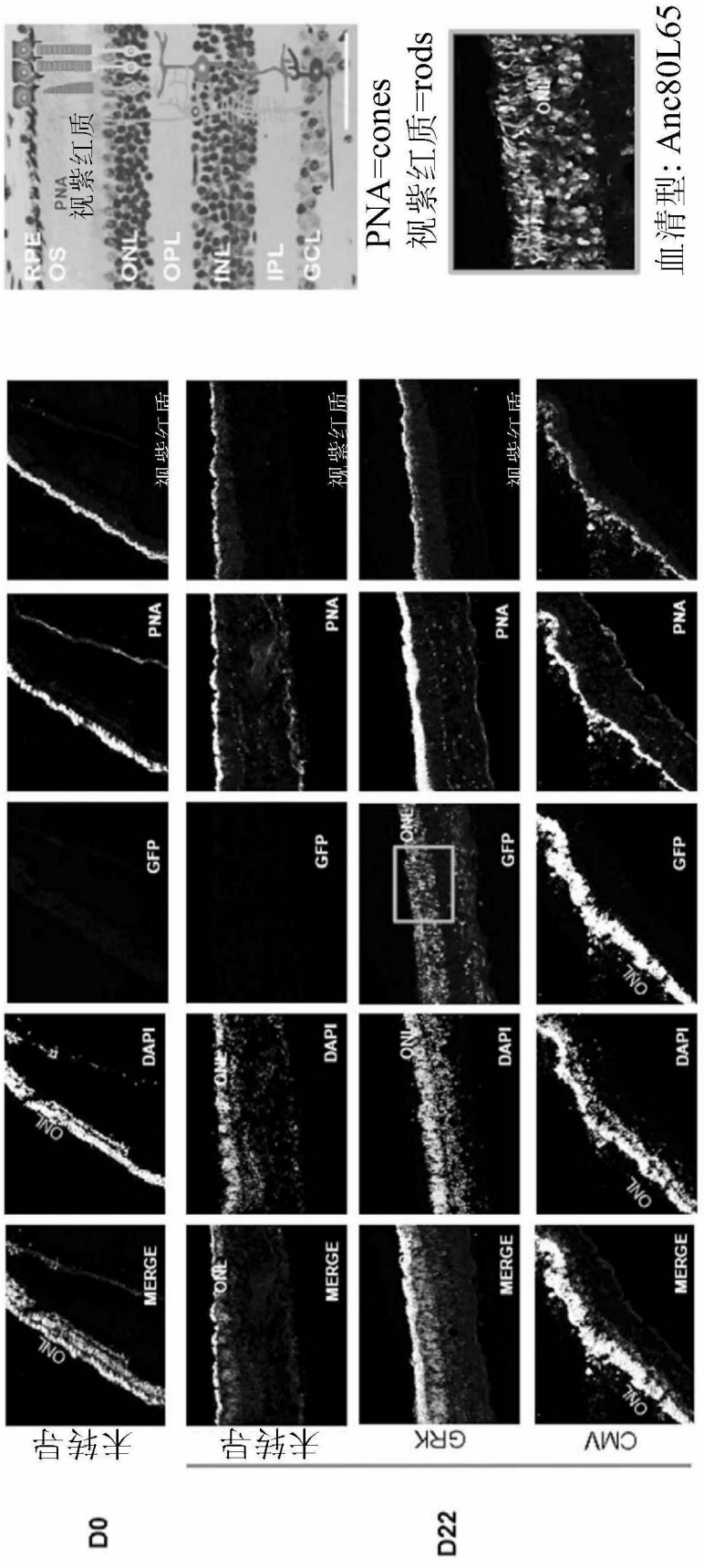
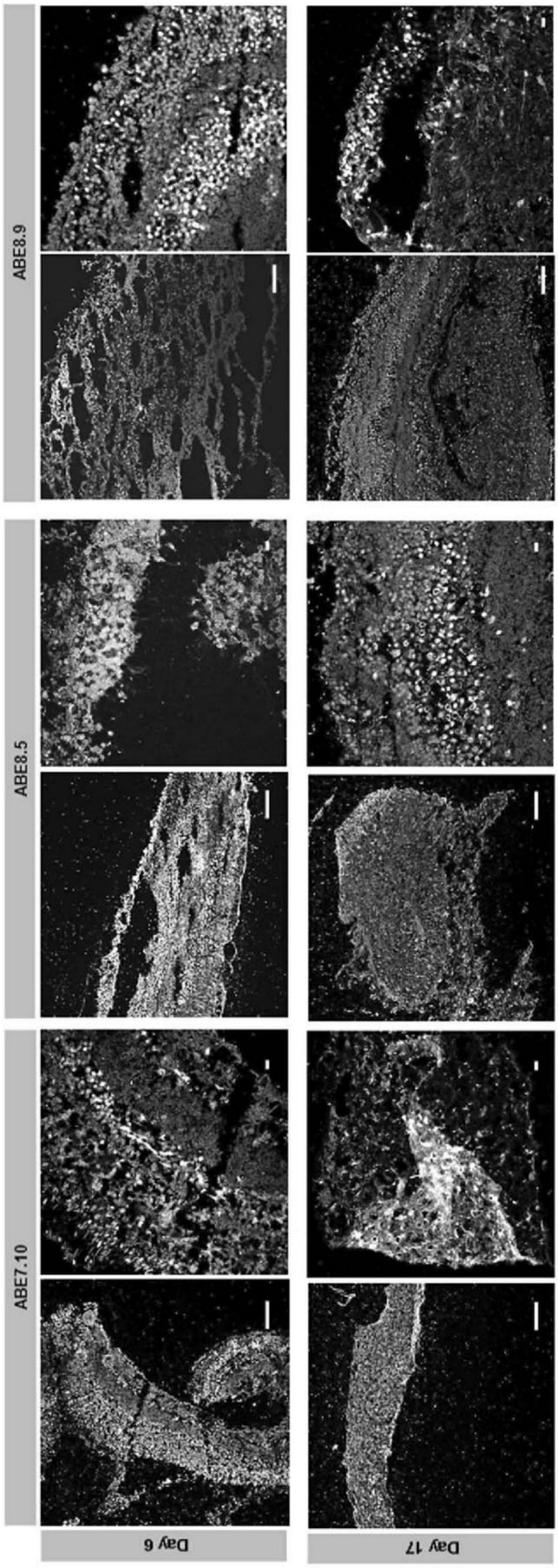

RPE
OS
PNA
视紫红质
ONL
OPL
INL
IPL
GCL
PNA=cones
视紫红质=rods
血清型：Anc80L65
D0
D22
未转导
未转导
GRK
CMV
视紫红质
PNA
GFP
DAPI
MERGE
ONL

图78



我们的双AAV系统提供表达Cas9的分裂内含肽

图79