(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 115279398 B
(45)授权公告日 2025.07.04

(21)申请号 202080082229.8
(22)申请日 2020.09.25
(65)同一申请的已公布的文献号
申请公布号 CN 115279398 A
(43)申请公布日 2022.11.01
(30)优先权数据
62/907,254 2019.09.27 US
(85)PCT国际申请进入国家阶段日
2022.05.25
(86)PCT国际申请的申请数据
PCT/US2020/052822 2020.09.25
(87)PCT国际申请的公布数据
WO2021/062227 EN 2021.04.01
(73)专利权人 比姆医疗股份有限公司
地址 美国马萨诸塞州
专利权人 UCL商业有限责任公司

(72)发明人 J·M·格尔克 A·D·爱德华兹
R·默里 W·卡西姆

(74)专利代理机构 北京戈程知识产权代理有限公司 11314
专利代理师 程伟
(51)Int.Cl.
A61K 39/00(2006.01)
C07K 16/46(2006.01)
C12N 5/0783(2010.01)
A61P 35/00(2006.01)

(56)对比文件
CN 107075483 A,2017.08.18
WO 2018165631 A1,2018.09.13
CN 109996811 A,2019.07.09

审查员 林薇

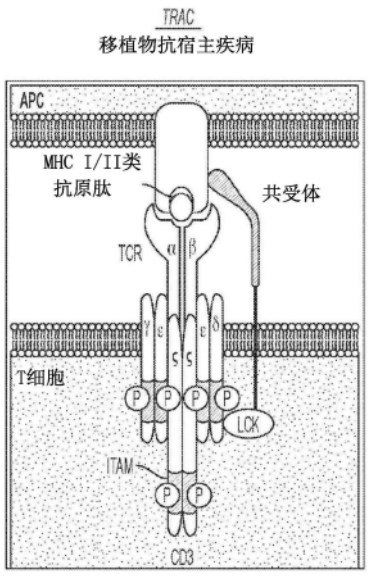
权利要求书2页 说明书433页
序列表（电子公布） 附图72页

(54)发明名称

治疗液态癌症的组合物和方法

(57)摘要

如下所述,本发明提出经基因修饰的免疫细胞,所述免疫细胞具有增强的抗瘤形成活性、抗免疫抑制性和降低的引起移植物抗宿主反应的风险,或其组合物。本发明还提出用于生产和使用这些经修饰的免疫效应细胞的方法。



CN 115279398 B

1.一种用于生产具有降低的免疫原性的表达CAR的免疫细胞或表达CAR的免疫细胞群的离体方法,所述方法包括:

a)将免疫细胞或免疫细胞群中的靶标多核苷酸与

1)包含向导多核苷酸和碱基编辑器的碱基编辑器系统接触,所述碱基编辑器包含含有核酸可编程DNA结合蛋白(napDNAbp)和TadA*8脱氨酶结构域的融合蛋白;或者

2)包含向导多核苷酸、napDNAbp和TadA*8脱氨酶结构域的碱基编辑器系统接触,

其中所述TadA*8脱氨酶结构域的氨基酸序列如SEQ ID NO:8所示或如SEQ ID NO:8的位置2至167所示,以及所述TadA*8脱氨酶结构域的氨基酸序列具有参考SEQ ID NO:8的氨基酸改变,

其中所述氨基酸改变由以下组成:

Y147T;

Y147R;

Q154S;

Y123H;

V82S;

T166R;

Q154R;

Y147R、Q154R和Y123H;

Y147R、Q154R和I76Y;

Y147R、Q154R和T166R;

Y147T和Q154R;

Y147T和Q154S;

I76Y、Y123H、Y147R和Q154R;

I76Y和V82S;

V82S和Y147R;

V82S、Y123H和Y147R;

V82S和Q154R;

V82S、Y123H和Q154R;

V82S、Y123H、Y147R和Q154R;

I76Y、V82S、Y123H、Y147R和Q154R;

Y147R和Q154S;

V82S和Q154S;

V82S和Y123H;或

V82S、Y123H和Y147T;

b)通过核碱基修饰引入中断CD5基因和/或CD7基因中的剪接位点的突变,从而降低或消除选自由CD5和CD7所组成的群组的抗原的表达,其中所述CD5基因编码SEQ ID NO:27的氨基酸序列,以及所述CD7基因编码SEQ ID NO:29的氨基酸序列;

c)通过多重编辑引入中断B2M基因、CD52基因、PD1基因和/或TRAC基因中的剪接位点的突变至所述免疫细胞或免疫细胞群内,从而减少或消除由所述基因编码的以下群组的多肽

的表达,其中所述B2M基因编码SEQ ID NO:17的氨基酸序列,所述CD52基因编码SEQ ID NO:33的氨基酸序列,所述D1基因编码SEQ ID NO:91的氨基酸序列,以及所述TRAC基因编码SEQ ID NO:93的氨基酸序列;其中所述多肽的群组是:

i)B2M、TRAC和PD1,或

ii)PD1、TRAC和CD52;以及

d)于所述免疫细胞或免疫细胞群中表达靶向具有减少或消除的表达的CD5或CD7抗原的嵌合抗原受体,从而生产具有降低的免疫原性的表达CAR的免疫细胞或表达CAR的免疫细胞群,

其中通过所述方法生产的所述表达CAR的免疫细胞或表达CAR的免疫细胞群不包含可检测的易位,以及其中

1)所述方法包含减少或消除CD7抗原的表达,且其中所述CAR靶向所述CD7抗原且由SEQ ID NO:113所示的氨基酸序列组成;或

2)所述CAR是CD5嵌合抗原受体,其由选自SEQ ID NO:3至7和112的氨基酸序列组成。

2.根据权利要求1所述的方法,其中所述向导多核苷酸包含选自CUCUUACCUGUACCAUAACC、CCUACCUGUCACCAGGACCA和CACCUACCUAAGAACCAUCC的核酸序列。

3.根据权利要求1所述的方法,其中与相对应的对照细胞相比,通过所述方法生产的所述表达CAR的免疫细胞或表达CAR的免疫细胞群表现出抗同类相杀性和/或增加的抗瘤形成活性。

4.根据权利要求1所述的方法,其中通过所述方法生产的所述表达CAR的免疫细胞或表达CAR的免疫细胞群不包含可检测的易位。

5.根据权利要求1所述的方法,其中通过电穿孔、核转染、阳离子脂质介导的方法、病毒转导或其组合将所述碱基编辑器和一个或多个向导核酸序列引入所述免疫细胞或免疫细胞群中。

6.根据权利要求1所述的方法,进一步包括从所述表达CAR的免疫细胞或表达CAR的免疫细胞群中耗尽TCRα/β+细胞。

治疗液态癌症的组合物和方法

[0001] 相关申请的交叉引用

[0002] 本申请是国际PCT申请,主张2019年9月27日提交的美国临时专利申请62/907,254的优先权和权益,该临时专利申请的整体内容通过引用而整体并入本文。

技术领域

[0003] 本申请涉及经基因修饰的免疫细胞,具有增强的抗瘤形成活性、抗免疫抑制性、降低的引起移植物抗宿主反应的风险或其组合。本申请还涉及生产和使用该经基因修饰的免疫细胞的方法。

背景技术

[0004] 自体和同种异体免疫疗法是肿瘤治疗方法,在该方法中,将表达嵌合抗原受体的免疫细胞给药至受试者。为了生成表达嵌合抗原受体(CAR)的免疫细胞,首先从受试者(自体)或与接受治疗的受试者分开的供体(同种异体)收集免疫细胞,并经基因修饰以表达嵌合抗原受体。所得细胞在其细胞表面上表达嵌合抗原受体(例如,CAR T细胞),并且当给药至受试者时,嵌合抗原受体结合至由肿瘤细胞表达的标志物。这种与肿瘤标志物的相互作用激活CAR-T细胞,随后这些细胞杀死肿瘤细胞。但对于使自体或同种异体细胞疗法有效或高效,必须减少或避免显著的条件和细胞应答诸如T细胞信号传导抑制。对于同种异体细胞疗法,移植物抗宿主疾病和宿主对CAR-T细胞的排异可造成额外的挑战。编辑这些过程中牵涉到的基因可增强CAR-T细胞功能和对免疫阻遏或抑制的抗性,当目前用于制作此类编辑的方法具有在CAR-T细胞中诱导大的基因组重排的潜力,从而对其功效造成负面影响。因此,对于更精确地修饰免疫细胞尤其是CAR-T细胞的技术存在显著需求。本申请涉及这种和其他重要需求。

发明内容

[0005] 如下所述,本发明特征在于经基因修饰的免疫细胞,其具有增强的抗瘤形成活性、对免疫阻遏的抗性,以及降低的引起移植物抗宿主反应或宿主抗移植物反应的风险,其中宿主CD8+ T细胞将移植物识别为非自我(例如,其中移植物接受者生成针对被移植器官的免疫应答),或其组合。在一种实施方案中,向患有移植物抗宿主疾病(GVHD)或具有发展出GVHD倾向的受试者给药CAR-T细胞,该细胞缺乏或具有降低水平的功能性TRAC。在一种实施方案中,向患有宿主抗移植物疾病(HVGD)或具有发展出HVGD倾向的受试者给药CAR-T细胞,该细胞缺乏或具有降低水平的β2微球蛋白(B2M)。本发明还提出用于生产和使用这些经修饰的免疫细胞的方法。

[0006] 一方面,本发明提供一种组合物,其包含两种或更多种免疫细胞,每种免疫细胞包含a)不同的嵌合抗原受体,其靶向选自由CD5、CD7、CD3、CD33和CD123所组成的群组的抗原,其中该免疫细胞包含减少或消除所靶向抗原的表达的突变;和b)一个或多个突变,其减少或消除选自由TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和PD1所组成的群组的免疫

原性多肽的表达。在一些实施方案中,免疫细胞中的一种包含靶向CD5的嵌合抗原受体,且另一种免疫细胞包含靶向选自由CD7、CD3、CD33和CD123所组成的群组的抗原的嵌合抗原受体。在一些实施方案中,组合物包含至少三种免疫细胞,每一种包含靶向不同抗原的嵌合抗原受体,其中被靶向抗原是CD3、CD5和CD7。在一些实施方案中,免疫细胞中的一种表达靶向CD7的嵌合抗原受体,且另一种免疫细胞表达靶向选自由CD3、CD33和CD123所组成的群组的抗原的嵌合抗原受体。在一些实施方案中,免疫细胞中的一种表达靶向CD3的嵌合抗原受体,且另一种免疫细胞表达靶向选自由CD33和CD123所组成的群组的抗原的嵌合抗原受体。在一些实施方案中,免疫细胞中的一种表达靶向CD33的嵌合抗原受体,且另一种免疫细胞表达靶向CD123的嵌合抗原受体。在一些实施方案中,一种免疫细胞表达两种、三种、四种或更多种不同的嵌合抗原受体。在一些实施方案中,嵌合抗原受体中的一种靶向CD5,且另一嵌合抗原受体靶向选自由CD7、CD3、CD33和CD123所组成的群组的抗原。在一些实施方案中,嵌合抗原受体中的一种靶向CD7,且另一嵌合抗原受体靶向选自由CD3、CD33和CD123所组成的群组的抗原。在一些实施方案中,嵌合抗原受体中的一种靶向CD3,且另一嵌合抗原受体靶向选自由CD33和CD123所组成的群组的抗原。在一些实施方案中,嵌合抗原受体中的一种靶向CD33,且另一嵌合抗原受体靶向CD123。

[0007] 另一方面,本发明提供一种组合物,其包含至少三种免疫细胞,每一种包含不同的嵌合抗原受体,其中一种嵌合抗原受体靶向CD3,一种靶向CD5,且第三种靶向CD7,其中细胞进一步包含减少或消除被靶向抗原的表达的突变;并且进一步包含一个或多个突变,所述突变减少或消除选自由TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和PD1所组成的群组的多肽的表达。

[0008] 又一方面,本发明提供一种组合物,其包含至少两种免疫细胞,每一种包含不同的嵌合抗原受体,其中一种嵌合抗原受体靶向CD3,且另一种靶向CD7,其中细胞进一步包含减少或消除被靶向抗原的表达的突变;并且进一步包含一个或多个突变,所述突变减少或消除选自由TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和PD1所组成的群组的多肽的表达。

[0009] 一方面,本发明提供一种组合物,其包含至少两种免疫细胞,每一种包含不同的嵌合抗原受体,其中一种嵌合抗原受体靶向CD5,且另一种靶向CD7,其中细胞进一步包含减少或消除被靶向抗原的表达的突变;并且进一步包含一个或多个突变,所述突变减少或消除选自由TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和PD1所组成的群组的多肽的表达。

[0010] 另一方面,本发明提供一种组合物,其包含至少两种免疫细胞,每一种包含不同的嵌合抗原受体,其中一种嵌合抗原受体靶向CD3,且另一种靶向CD5,其中细胞进一步包含减少或消除被靶向抗原的表达的突变;并且进一步包含一个或多个突变,所述突变减少或消除选自由TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和PD1所组成的群组的多肽的表达。

[0011] 又一方面,本发明提供一种组合物,其包含至少两种免疫细胞,每一种包含不同的嵌合抗原受体,其中一种嵌合抗原受体靶向CD33,且另一种靶向CD123,其中细胞进一步包含减少或消除被靶向抗原的表达的突变;并且进一步包含一个或多个突变,所述突变减少或消除选自由TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和PD1所组成的群组的多肽的表达。

[0012] 在一些实施方案中,突变是C至T或A至G突变,该突变沉默所述基因或将终止密码子引入所述基因中。在一些实施方案中,突变引入提前终止密码子或改变剪接供体或受体位点。在一些实施方案中,突变通过包含脱氨酶结构域的碱基编辑器生成。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,碱基编辑器是BE4。在一些实施方案中,突变将编码多肽的表达相对于缺少所述突变的对应对照细胞减少约50%或更多。在一些实施方案中,组合物包含免疫细胞群。在一些实施方案中,群的至少50%包含一个或多个突变,所述突变减少或消除所述被靶向抗原和/或所述免疫原性多肽的表达。在一些实施方案中,免疫细胞是抗同类相杀的。在一些实施方案中,免疫细胞具有增加的抗瘤形成活性。在一些实施方案中,免疫细胞不包含可检测的易位。在一些实施方案中,免疫细胞包含少于1%的插入缺失。在一些实施方案中,免疫细胞是哺乳动物细胞。在一些实施方案中,免疫细胞是人类或啮齿动物细胞。在一些实施方案中,免疫细胞是细胞毒性T细胞、调节性T细胞、T辅助细胞、树突细胞、B细胞或NK细胞,或其祖细胞。在一些实施方案中,祖细胞是造血干细胞。

[0013] 一方面,本发明提供一种药物组合物,其包含有效量的任何本文所提供的组合物和药学上可接受的赋形剂。

[0014] 另一方面,本发明提供一种碱基编辑器系统,其包含融合蛋白,所述融合蛋白包含核酸可编程DNA结合蛋白(napDNAbp)和脱氨酶结构域,以及至少两个向导多核苷酸,每个向导多核苷酸靶向不同的抗原,其中抗原选自由CD5、CD7、CD3、CD33和CD123所组成的群组。在一些实施方案中,一种向导多核苷酸靶向CD5,且另一种靶向选自由CD7、CD3、CD33和CD123所组成的群组。在一些实施方案中,系统包含三个向导多核苷酸,每个向导多核苷酸靶向抗原CD3、CD5和CD7中的一者。在一些实施方案中,一种向导多核苷酸靶向CD7,且另一种靶向选自由CD3、CD33和CD123所组成的群组。在一些实施方案中,一种向导多核苷酸靶向CD3,且另一种靶向选自由CD33和CD123所组成的群组。在一些实施方案中,一种你想到多核苷酸靶向CD33,且另一种靶向CD123。

[0015] 在一些实施方案中,向导多核苷酸各自包含选自表26的核酸序列。在一些实施方案中,向导多核苷酸包含选自AGCGACUGCAGAAAGAAGAG或CAUACCAGCUGAGCCGUCCG的核酸序列。在一些实施方案中,融合蛋白进一步包含一个或多个尿嘧啶糖基化酶抑制剂(UGI)和/或一个或多个核定位序列(NLS)。在一些实施方案中,napDNAbp包含Cas9、Cas12a/Cpf1、Cas12b/C2c1、Cas12c/C2c3、Cas12d/CasY、Cas12e/CasX、Cas12g、Cas12h、Cas12i或Cas12j/CasΦ多肽或其部分。在一些实施方案中,napDNAbp包含Cas12多肽或其片段。在一些实施方案中,napDNAbp包含Cas9多肽或其片段。在一些实施方案中,Cas9是死亡的Cas9(dCas9)或Cas9切口酶(nCas9)。在一些实施方案中,Cas9是经修饰的金黄色葡萄球菌Cas9(SaCas9)、嗜热链球菌1Cas9(St1Cas9)、或经修饰的酿脓链球菌Cas9(SpCas9)。在一些实施方案中,Cas9包含改变的前间区序列邻近基序(PAM)特异性。在一些实施方案中,改变的PAM对核酸序列5'-NGC-3'具有特异性。

[0016] 在一些实施方案中,脱氨酶结构域能够将胞苷或腺苷脱氨基。在一些实施方案中,脱氨酶结构域是腺苷脱氨酶结构域。在一些实施方案中,胞苷脱氨酶是APOBEC脱氨酶。在一些实施方案中,腺苷脱氨酶是TadA变体。在一些实施方案中,TadA变体是TadA*8变体。在一些实施方案中,TadA*8变体是TadA*8.1、TadA*8.2、TadA*8.3、TadA*8.4、TadA*8.5、TadA*8.6、TadA*8.7、TadA*8.8、TadA*8.9、TadA*8.10、TadA*8.11、TadA*8.12、TadA*8.13、TadA*

8.14、TadA*8.15、TadA*8.16、TadA*8.17、TadA*8.18、TadA*8.19、TadA*8.20、TadA*8.21、TadA*8.22、TadA*8.23或TadA*8.24。在一些实施方案中,TadA变体是TadA*9变体。

[0017] 一方面,本发明提供一种药物组合物,其包含本文所提供的任何碱基编辑器系统。

[0018] 另一方面,本发明提供一种多核苷酸,其编码本文所提供的任何碱基编辑器系统和向导多核苷酸。

[0019] 又一方面,本发明提供一种载体,其包含本文所提供的任何多核苷酸。在一些实施方案中,载体是病毒载体。在一些实施方案中,病毒载体是逆转录病毒载体、腺病毒载体、慢病毒载体、疱疹病毒载体或腺相关病毒载体(AAV)。

[0020] 一方面,本发明提供一种药物组合物,其包含本文所提供的任何多核苷酸或任何载体。

[0021] 一方面,本发明提供一种用于生产具有降低的免疫原性的表达CAR的免疫细胞的方法。在一些实施方案中,该方法包括在表达CAR的免疫细胞中表达碱基编辑器系统,该碱基编辑器系统包含融合蛋白和两个向导多核苷酸,该融合蛋白包含核酸可编程DNA结合蛋白(napDNAbp)和脱氨酶结构域,该向导多核苷酸各自靶向编码不同抗原的多核苷酸,其中抗原选自由CD5、CD7、CD3、CD33和CD123所组成的群组,从而生产具有降低的免疫原性的表达CAR的免疫细胞。在一些实施方案中,免疫细胞表达或接触向导多核苷酸,所述向导多核苷酸靶向编码选自由TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和PD1所组成的群组的多肽的多核苷酸。

[0022] 在一些实施方案中,该方法包括(a)在表达CAR的免疫细胞中表达碱基编辑器系统,该碱基编辑器系统包含融合蛋白,该融合蛋白包含核酸可编程DNA结合蛋白(napDNAbp)和脱氨酶结构域,和(b)使该表达CAR的免疫细胞与至少两个向导多核苷酸接触,每个向导多核苷酸靶向编码不同抗原的多核苷酸,其中所述抗原选自由CD5、CD7、CD3、CD33和CD123所组成的群组,从而生产具有降低的免疫原性的表达CAR的免疫细胞。

[0023] 在一些实施方案中,该方法进一步包括使表达CAR的免疫细胞与向导多核苷酸接触,所述向导多核苷酸靶向编码选自由TRAC、LAG-3、FAS、CIITA、TRB在一些实施方案中,免疫细胞表达嵌合抗原受体(CAR),所述CAR靶向选自由CD5、CD7、CD3、CD33和CD123所组成的群组的抗原。在一些实施方案中,该方法进一步包括将突变引入免疫细胞内,所述突变减少或消除选自由TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和PD1所组成的群组的至少一种多肽的表达,从而生产具有降低的免疫原性的表达CAR的免疫细胞群。

[0024] 一方面,本发明提供一种用于生产具有降低的免疫原性的表达CAR的免疫细胞群的方法。在一些实施方案中,该方法包括a)将减少或消除选自由CD5、CD7、CD3、CD33和CD123所组成的群组的抗原的表达的突变引入一种免疫细胞内并将所述抗原中的一个中的不同突变引入第二种免疫细胞内;及b)将突变引入所述免疫细胞内,所述突变减少或消除选自由TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和PD1所组成的群组的至少一种多肽的表达,从而生产具有降低的免疫原性的表达CAR的免疫细胞群。在一些实施方案中,由免疫细胞表达的嵌合抗原受体靶向选自由CD3、CD5、CD7、CD33和CD123所组成的群组的抗原。在一些实施方案中,通过该方法生产的免疫细胞表达靶向CD3、CD5和/或CD7抗原的嵌合抗原受体,但未能表达或表达降低的水平的CD3、CD5和/或CD7抗原。在一些实施方案中,通过该方法生产的免疫细胞表达靶向CD33和CD123抗原的嵌合抗原受体,但未能表达或表达降低

的水平的CD33和CD123抗原。在一些实施方案中,CAR是CD5嵌合抗原受体(CAR)。在一些实施方案中,CD5 CAR由表28中呈现的CD5 CAR构建体编码。

[0025] 另一方面,本发明提供一种用于生产具有降低的免疫原性的免疫细胞的方法。在一些实施方案中,该方法包括a)将突变引入内源性CD5基因序列中或调节元件中,该突变减少或消除CD5的表达;以及b)在所述细胞中表达表28中呈现的CD5 CAR构建体。在一些实施方案中,CD5 CAR构建体编码CD5 CAR多肽,所述多肽包含选自下列的氨基酸序列或由其组成:

[0026] a)

1 MEFGLSWLFLVAILKGVQCIDAMGNIQLVQSGPELKKPGETVKISCKASGYTFTNYGMNW 60
61 VKQAPGKGLRWMGWINTHTGEPTYADDFKGRFAFSLETSASTAYLQINNLKNEDTATYFC
120
121 TRRGYDWYFDVWGAGTTVTVSSGGGGSGGGGSGGGGSDIKMTQSPSSMYASLGERVTITC
180
181 KASQDINSYLSWFHHKPGKSPKTLIYRANRLVDGVPSRFSGSGSGQDYSLTISSLDYEDM
[0027] 240
241 GIYYCQQYDESPWTFGGGTKLEMKGSGDPAEPKSPDKTHTCPGQPREPQVYTLPPSRDEL
300
301 TKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQ
360
361 QGNVFSCSVMHEALHNHYTQKSLSLSPGKKDPKFWVLVVVGGVLACYSLLVTVAFIIFWV
420
421 RSKRSRLLHSDYMNMTPRRPGPTRKHYQPYAPPRDFAAYRSRVKFSRSADAPAYQQGQNQ
480
[0028] 481 LYNELNLGRREEYDVLDKRRGRDPEMGGKPRRKNPQEGLYNELQKDKMAEAYSEIGMKGE
540
541 RRRGKGHDGLYQGLSTATKDTYDALHMQALPPR 573;

[0029] b)

1 MEFGLSWLFLVAILKGVQCIDAMGNIQLVQSGPELKKPGETVKISCKASGYTFTNYGMNW
60
61 VKQAPGKGLRWMGWINTHTGEPTYADDFKGRFAFSLETSASTAYLQINNLKNEDTATYFC
120
121 TRRGYDWYFDVWGAGTTVTVSSGGGGSGGGGSGGGGSDIKMTQSPSSMYASLGERVTITC
180
181 KASQDINSYLSWFHHKPGKSPKTLIYRANRLVDGVPSRFSGSGSGQDYSLTISSLDYEDM
240
[0030] 241 GIYYCQQYDESPWTFGGGTKLEMKGSGDPATTTPAPRPPTPAPTIASQPLSLRPEACRPA
300
301 AGGAVHTRGLDFACDFWVLVVVGGVLACYSLLVTVAFIIFWVRSKRSRLLHSDYMNMTPR
360
361 RPGPTRKHYQPYAPPRDFAAYRSRVKFSRSADAPAYQQGQNQLYNELNLGRREEYDVLDK
420
421 RRGRDPEMGGKPRRKNPQEGLYNELQKDKMAEAYSEIGMKGERRRGKGHDGLYQGLSTAT
480
481 KDTYDALHMQALPPR 495;

[0031] c)

1 MEFGLSWLFLVAILKGVQCIDAMGNIQLVQSGPELKKPGETVKISCKASGYTFTNYGMNW
60
61 VKQAPGKGLRWMGWINTHTGEPTYADDFKGRFAFSLETSASTAYLQINNLKNEDTATYFC
120
121 TRRGYDWYFDVWGAGTTVTVSSGGGGSGGGGSGGGGSDIKMTQSPSSMYASLGERVTITC
180
181 KASQDINSYLSWFHHKPGKSPKTLIYRANRLVDGVPSRFSGSGSGQDYSLTISSLDYEDM
240

[0032] 241 GIYYCQQYDESPWTFGGGTKLEMKGSGDPAIEVMYPPPYLDNEKSNGTIIHVKGKHLCPS
300
301 PLFPGPSKPFWVLVVVGGVLACYSLLVTVAFIIFWVRSKRSRLLHSDYMNMTPRRPGPTR
360
361 KHYQPYAPPRDFAAYRSRVKFSRSADAPAYQQGQNQLYNELNLGRREEYDVLDKRRGRDP
420
421 EMGGKPRRKNPQEGLYNELQKDKMAEAYSEIGMKGERRRGKGHDGLYQGLSTATKDTYDA
480
481 LHMQALPPR 489;

[0033] d)

[0034] 1 MEFGLSWLFLVAILKGVQCIDAMGNIQLVQSGPELKKPGETVKISCKASGYTFTNYGMNW
60

61 VKQAPGKGLRWMGWINTHTGEPTYADDFKGRFAFSLETSASTAYLQINNLKNEDTATYFC
120
121 TRRGYDWYFDVWGAGTTVTVSSGGGGSGGGGSGGGGSDIKMTQSPSSMYASLGERVTITC
180
181 KASQDINSYLSWFHHKPGKSPKTLIYRANRLVDGVPSRFSGSGSGQDYSLTISSLDYEDM
240
241 GIYYCQQYDESPWTFGGGTKLEMKGSGDPAEPKSPDKTHTCPGQPREPQVYTLPPSRDEL
300
301 TKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQ
[0035] 360
361 QGNVFSCSVMHEALHNHYTQKSLSLSPGKKDPKPTTTPAPRPPTPAPTIASQPLSLRPEA
420
421 CRPAAGGAVHTRGLDFACDIYIWAPLAGTCGVLLLSLVITLYCKRGRKKLLYIFKQPFMR
480
481 PVQTTQEEDGCSCRFPEEEEGGCELRVKFSRSADAPAYQQGQNQLYNELNLGRREEYDVL
540
541 DKRRGRDPEMGGKPRRKNPQEGLYNELQKDKMAEAYSEIGMKGERRRGKGHDGLYQGLST
600
601 ATKDTYDALHMQALPPR 617；或

[0036] e)

1 MEFGLSWLFLVAILKGVQCIDAMGNIQLVQSGPELKKPGETVKISCKASGYTFTNYGMNW
60
61 VKQAPGKGLRWMGWINTHTGEPTYADDFKGRFAFSLETSASTAYLQINNLKNEDTATYFC
120
121 TRRGYDWYFDVWGAGTTVTVSSGGGGSGGGGSGGGGSDIKMTQSPSSMYASLGERVTITC
180
181 KASQDINSYLSWFHHKPGKSPKTLIYRANRLVDGVPSRFSGSGSGQDYSLTISSLDYEDM
240
[0037] 241 GIYYCQQYDESPWTFGGGTKLEMKGSGDPATTTPAPRPPTPAPTIASQPLSLRPEACRPA
300
301 AGGAVHTRGLDFACDIYIWAPLAGTCGVLLLSLVITLYCKRGRKKLLYIFKQPFMRPVQT
360
361 TQEEDGCSCRFPEEEEGGCELRVKFSRSADAPAYQQGQNQLYNELNLGRREEYDVLDKRR
420
421 DPEMGGKPRRKNPQEGLYNELQKDKMAEAYSEIGMKGERRRGKGHDGLYQGLSTATKD
480
481 DALHMQALPPRX 494.

[0038] 在一些实施方案中，与相对应的对照细胞相比，通过该方法生产的免疫细胞表现

出抗同类相杀性和/或增加的抗瘤形成活性。在一些实施方案中,该方法在体内或间接体内进行。在一些实施方案中,通过该方法生成的免疫细胞不包含可检测的易位。在一些实施方案中,通过该方法生成的免疫细胞包含少于1%的插入缺失。在一些实施方案中,通过该方法生成的免疫细胞包含少于5%的非靶标编辑。在一些实施方案中,通过该方法生成的免疫细胞包含少于5%的脱靶编辑。在一些实施方案中,通过核碱基修饰来生成突变。在一些实施方案中,突变在外显子中。在一些实施方案中,突变导致减少或消除蛋白质表达的提前终止密码子。在一些实施方案中,突变在剪接供体位点或剪接受体位点中。在一些实施方案中,一个或多个突变通过使靶标多核苷酸与碱基编辑器系统接触来生成,所述碱基编辑器系统包含具有核酸可编程DNA结合蛋白(napDNAbp)的融合蛋白、脱氨酶和一个或多个向导多核苷酸。

[0039] 在一些实施方案中,脱氨酶是腺苷或胞苷脱氨酶。在一些实施方案中,胞苷脱氨酶是BE4。在一些实施方案中,突变将编码多肽的表达相对于缺少所述突变的对应对照细胞减少至少约50%或更多。在一些实施方案中,向导多核苷酸包含选自表26处所提供的那些的序列。在一些实施方案中,向导多核苷酸包含选自AGCGACUGCAGAAAGAAGAG或CAUACCAGCUGAGCCGUCCG的核酸序列。在一些实施方案中,一个或多个核酸序列各自将napDNAbp靶向至CD5、FAS、LAG-3、CD52、TRAC、B2M、CIITA、TRBC1、TRBC2和/或PDC1/PD-1基因或调节元件。在一些实施方案中,通过电穿孔、核转染、阳离子脂质介导的方法、病毒转导或其组合将碱基编辑器和一个或多个向导核酸序列引入免疫细胞中。在一些实施方案中,该方法进一步包括将免疫细胞在培养物中扩张以生成免疫细胞群。在一些实施方案中,在至少约50%的免疫细胞群中,抗原或多肽的表达减少。在一些实施方案中,该方法进一步包括从所述经修饰的免疫细胞群中耗尽TCRα/β+细胞。

[0040] 一方面,本发明提供通过本文所提供的任何方法生产的具有降低的免疫原性的表达CAR的免疫细胞。

[0041] 另一方面,本发明提供一种药物组合物,其包含本文所提供的任何免疫细胞。在一些实施方案中,药物组合物包括药学上可接受的赋形剂。

[0042] 又一方面,本发明提供一种用于杀死肿瘤细胞的方法。在一些实施方案中,该方法包括使表达选自由CD5、CD7、CD3、CD33和CD123所组成的群组的抗原的肿瘤细胞与两种或更多种免疫细胞接触,每种免疫细胞表达不同的嵌合抗原受体,所述嵌合抗原受体靶向该细胞所表达的抗原中的两者,其中免疫细胞包含一个或多个减少或消除被靶向抗原的表达的突变,和一个或多个减少或消除选自由TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和PD1所组成的群组的多肽的表达的突变。在一些实施方案中,该方法在体外或体内进行。在一些实施方案中,肿瘤细胞源自瘤形成。

[0043] 一方面,本发明提供用于治疗受试者的瘤形成的方法。在一些实施方案中,该方法包括向受试者给药两种或更多种免疫细胞,每种免疫细胞表达不同的嵌合抗原受体,所述嵌合抗原受体靶向由受试者的肿瘤细胞表达的抗原,所述抗原选自由CD5、CD7、CD3、CD33和CD123所组成的群组,其中免疫细胞包含一个或多个减少s或消除所述抗原的表达的突变,且免疫细胞各自进一步包含一个或多个减少或消除选自由TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和PD1所组成的群组的多肽的表达的突变。

[0044] 在一些实施方案中,瘤形成是血液学癌症。在一些实施方案中,瘤形成是液态癌

症。在一些实施方案中,血液学癌症是白血病、骨髓瘤和/或淋巴瘤。在一些实施方案中,血液学癌症是B细胞癌症。在一些实施方案中,血液学癌症选自以下至少一项:T细胞急性成淋巴细胞性白血病(T-ALL)、蕈样肉芽肿(MF)、Sézary综合征(SS)、外周T/NK细胞淋巴瘤、间变性大细胞淋巴瘤ALK$^+$、原发性皮肤T细胞淋巴瘤、T细胞大颗粒淋巴细胞型白血病、血管免疫母细胞性T/NK细胞淋巴瘤、肝脾T细胞淋巴瘤、原发性皮肤CD30$^+$淋巴增生性疾病、结外NK/T细胞淋巴瘤、成人T细胞白血病/淋巴瘤、T细胞幼淋巴细胞白血病、皮下脂膜炎样T细胞淋巴瘤、原发性皮肤γδT细胞淋巴瘤、侵袭性NK细胞白血病和肠病相关T细胞淋巴瘤。在一些实施方案中,血液学癌症是T细胞急性成淋巴细胞性白血病(T-ALL)细胞。在一些实施方案中,血液学癌症是T细胞急性骨髓性白血病(AML)。

[0045] 另一方面,本发明提供用于治疗选定受试者的瘤形成的方法。在一些实施方案中,该方法包括向选定受试者给药两种或更多种免疫细胞,每种免疫细胞表达不同的嵌合抗原受体,所述嵌合抗原受体靶向由受试者的肿瘤细胞表达的抗原,其中抗原选自由CD5、CD7、CD3、CD33和CD123所组成的群组,其中免疫细胞包含一个或多个减少或消除所述抗原的表达的突变,且免疫细胞各自进一步包含一个或多个减少或消除选自由TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和PD1所组成的群组的多肽的表达的突变,其中受试者经选择为患有表达选自由CD5、CD7、CD3、CD33和CD123所组成的群组的抗原的瘤形成。在一些实施方案中,血液学癌症是T细胞急性成淋巴细胞性白血病(T-ALL)细胞。在一些实施方案中,血液学癌症是T细胞急性骨髓性白血病(AML)。在一些实施方案中,两种或更多种表达不同嵌合抗原受体的免疫细胞依次给药。在一些实施方案中,两种或更多种表达不同嵌合抗原受体的免疫细胞同时给药。

[0046] 又一方面,本发明提供一种用于抗原依赖性地杀死受试者肿瘤细胞的方法。在一些实施方案中,该方法包括向患有表达选自由CD5、CD7、CD3、CD33和CD123所组成的群组的抗原的瘤形成的受试者给药两种或更多种免疫细胞,每种免疫细胞表达不同的嵌合抗原受体,所述嵌合抗原受体靶向所述抗原中的一者,其中免疫细胞包含一个或多个减少或消除被靶向抗原的表达的突变,和一个或多个减少或消除选自由TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和PD1所组成的群组的多肽的表达的突变。

[0047] 一方面,本发明提供一种用于抗原依赖性地杀死受试者急性骨髓性白血病(AML)细胞的方法。在一些实施方案中,该方法包括向患有表达CD33和CD123抗原的AML的受试者给药两种或更多种免疫细胞,每种免疫细胞表达不同的嵌合抗原受体,所述嵌合抗原受体靶向所述抗原中的一者,其中所述免疫细胞包含一个或多个减少或消除所述被靶向抗原的表达的突变,和一个或多个减少或消除选自由TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和PD1所组成的群组的多肽的表达的突变。

[0048] 一方面,本发明提供一种用于抗原依赖性地杀死受试者T细胞急性成淋巴细胞性白血病(T-ALL)细胞的方法。在一些实施方案中,该方法包括向患有表达CD3、CD5和CD7抗原的T-ALL的受试者给药至少三种免疫细胞,每种免疫细胞表达不同的嵌合抗原受体,所述嵌合抗原受体靶向所述抗原中的一者,其中所述免疫细胞包含一个或多个减少或消除所述被靶向抗原的表达的突变,和一个或多个减少或消除选自由TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和PD1所组成的群组的多肽的表达的突变。

[0049] 另一方面,本发明提供用于治疗选定受试者的癌症的方法。在一些实施方案中,该

方法包括向受试者给药至少两种免疫细胞,每种免疫细胞表达靶向CD33或CD123抗原的嵌合抗原受体,其中免疫细胞包含一个或多个减少或消除所述被靶向抗原的表达的突变,和一个或多个减少或消除选自由TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和PD1所组成的群组的多肽的表达的突变,其中受试者通过将癌症表征为表达CD33和CD123抗原来选择。在一些实施方案中,表达CD33和CD123抗原的癌症是AML。

[0050] 又一方面,本发明提供用于治疗选定受试者的癌症的方法。在一些实施方案中,该方法包括向受试者给药三种或更多种免疫细胞,每种免疫细胞表达靶向CD3、CD5和CD7抗原的不同嵌合抗原受体,其中免疫细胞包含一个或多个减少或消除所述被靶向抗原的表达的突变,和一个或多个减少或消除选自由TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和PD1所组成的群组的多肽的表达的突变,其中所述受试者通过将所述癌症表征为表达CD3、CD5和CD7抗原来选择。在一些实施方案中,表达CD3、CD5和CD7抗原的癌症是T-ALL。

[0051] 在一些实施方案中,免疫细胞是细胞毒性T细胞、调节性T细胞、T辅助细胞、树突细胞、B细胞或NK细胞。在一些实施方案中,受试者先前已经用淋巴细胞删除进行治疗。在一些实施方案中,淋巴细胞删除包括给药环磷酰胺、氟达拉滨和/或阿仑单抗(Cy/Flu/Campath)。在一些实施方案中,受试者对于化疗是难治性的或具有高肿瘤负荷。在一些实施方案中,受试者随后用同种异体造血干细胞移植(allo-HSCT)进行治疗。在一些实施方案中,免疫细胞源自单个人类供体。在一些实施方案中,免疫细胞对于受试者是自体的。在一些实施方案中,免疫细胞对于受试者是同种异体的。在一些实施方案中,受试者是哺乳动物受试者。在一些实施方案中,受试者是人类或啮齿动物。在一些实施方案中,受试者是人类儿童受试者。

[0052] 一方面,本发明提供一种试剂盒,其包括本文所提供的任何组合物,用于治疗癌症。两一方面,本发明提供一种试剂盒,其包括本文所提供的任何碱基编辑器系统,用于生成具有降低的免疫原性的表达CAR的免疫细胞。在一些实施方案中,本文所提供的任何试剂盒包括用于使用该试剂盒的书面使用说明。

[0053] 本文的说明和实施例详细示出了本公开的实施方案。应理解,本公开不限于本文所述的特定实施方案并因此可变。本领域技术人员将认识到,存在大量涵盖于本公开范畴内的对本公开的变更和修饰。

[0054] 除非另外指明,否则本文所公开的一些实施方案的实践采用免疫学、生物化学、化学、分子生物学、微生物学、细胞生物学、基因学和重组DNA的常规技术,这些技术是本领域所能做到的。参见,例如,Sambrook and Green,Molecular Cloning:A Laboratory Manual,第四版(2012);Current Protocols in Molecular Biology系列(F.M.Ausubel,等人eds.);Methods In Enzymology系列(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;和Culture of Animal Cells:AManual of Basic Technique and Specialized Applications,第六版(R.I.Freshney,ed.(2010)。

[0055] 尽管本公开的各种特征可以在单一实施方案的情境中描述,但这些特征也可以单独提供或以任何组合提供。相反,尽管为了清楚起见,可以在单独实施方案的情境中描述本发明,但本发明也可以在单一实施方案中实现。本文所用的章节标题仅用于组织目的,而不视为对所描述的主旨的限制。

[0056] 本公开的特征在所附权利要求书中以特殊性详述。通过参考以下详述示例性实施方案的具体实施方式(其中使用了本公开的原理)以及下文所述的附图,可获得对本公开的特征和优点的更佳理解。

[0057] 定义

[0058] 以下定义是对本领域定义的补充并针对本申请,且不适用于任何相关或不相关的案例,例如,任何重复授权的专利或申请。尽管与本文所述类似或等效的任何方法和材料可用于实践或测试本公开,但优选的材料和方法如本文所述。据此,本文使用术语仅用于描述具体实施方案的目的,而不试图限制。

[0059] 除非另做定义,否则本文使用的全部科技术语均具有本发明所属领域技术人员通常理解的意义。下述参考文献向技术人员提供很多本发明所用术语的一般定义:《微生物学和分子生物学词典(第二版)》(Singleton等人,Dictionary of Microbiology and Molecular Biology(2nd ed.1994));《剑桥科技词典》(he Cambridge Dictionary of Science and Technology(Walker ed.,1988));《遗传性专业词典(第五版)》(The Glossary of Genetics,5th Ed.,R.Rieger等人(eds.),Springer Verlag(1991));以及《哈珀·柯林斯生物学词典》(Hale&Marham,The Harper Collins Dictionary of Biology(1991))。

[0060] 在本申请中,除非具体地另做指定,否则单数的使用包括复数。必须注意的是,如本说明书中所用,除非语境中明确地另外指明,否则单数形式“一”和“该”包括多个所指对象。在本申请中,除非另做指定“或”的使用意为“和/或”,并且理解为是包含性的。此外,术语“包括”以及其他形式诸如动词形式、主动或被动形式的使用没有限制。

[0061] 如本说明书和权利要求书中所用,词语“包含”(和任何形式的包含,诸如单数和复数形式)、“具有”(和任何形式的具有,诸如单数和复数形式)、“包括”(和任何形式的包括,诸如单数和复数形式)或“含有”(和任何形式的含有,诸如单数和复数形式)是包含性的或开放的,并且不排除额外的、未引用的元件或方法步骤。设想本说明书中讨论的任何实施方案可以关于本公开的任何方法或组合物来实施,反之亦然。此外,本公开的组合物可用于实现本公开的方法。

[0062] 术语“约”或“大约”意为处于由本领域一般技术人员确定的特定值的可接受的误差范围内,将部分地取决于该值如何测量或测定,即,测量系统的限制。例如,“约”可以意为,根据本领域的实践,处于一(1)个或超过一(1)个标准偏差内。或者,“约”可以意为给定至的至多20%、至多10%、至多5%、或至多1%的范围。或者,特别是关于生物系统或过长,该术语可以意为处于一定数量级内,诸如值的5倍内或2倍内。当在申请书和权利要求书中描述特定值时,除非另外指定,否则应假定术语“约”意为处于特定值的可接受的误差范围内。

[0063] 本文提供的范围理解为是该范围内所有值的简写。例如,1至50的范围理解为包括来自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所组成组的任何数字、数字组合或子范围。

[0064] 说明书中的“一些实施方案”、“实施方案”、“一种实施方案”或“其他实施方案”意为关于实施方案描述的特定特点、结构或特征包括在本公开的至少一些实施方案中,但不

必包括在全部实施方案中。

[0065] “腺苷脱氨酶”意为能够催化腺嘌呤或腺苷的水解脱氨基的多肽或其片段。在一些实施方案中,脱氨酶或脱氨酶结构域是催化腺苷到肌苷或脱氧腺苷到脱氧肌苷的水解脱氨基。在一些实施方案中,腺苷脱氨酶催化脱氧核糖核酸(DNA)中的腺嘌呤或腺苷的水解脱氨基。本文提供的腺苷脱氨酶(例如,工程改造的腺苷脱氨酶、进化的腺苷脱氨酶)可以来自任何生物体诸如细菌。

[0066] 在一些实施方案中,脱氨酶或脱氨酶结构域是来自生物体诸如人类、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,腺苷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为至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%、至少99.1%、至少99.2%、至少99.3%、至少99.4%、至少99.5%、至少99.6%、至少99.7%、至少99.8%或至少99.9%相同。在一些实施方案中,腺苷脱氨酶来自细菌,诸如大肠杆菌(E.coli)、金黄色葡萄球菌(S.Aureus)、枯草芽孢杆菌(B.subtilis)、伤寒沙门氏菌(S.typhi)、腐败希瓦氏菌(S.putrefaciens)、流感嗜血杆菌(H.influenzae)、新月柄杆菌(C.crescentus)或硫还原地杆菌(G.sulfurreducens)。在一些施方案中,腺苷脱氨酶是TadA脱氨酶。在一些实施方案中,TadA脱氨酶是大肠杆菌TadA(ecTadA)脱氨酶或其片段。

[0067] 在一些实施方案中,ecTadA胞苷脱氨酶是ecTadA。例如,截短的ecTadA可以相对于全长ecTadA缺少一个或多个N端氨基酸。在一些实施方案中,相对于全长ecTadA,截短的ecTadA8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,相对于全长ecTadA,截短的ecTadA8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,ecTadA脱氨酶不包含N端甲硫氨酸。在一些实施方案中,TadA脱氨酶是N端截短的TadA。在特定实施方案中,TadA是PCT/US2017/045381中描述的TadA中的任一种,该专利通过引用并入本文。

[0068] 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*7.10。在一些实施方案中,TadA变体是TadA*8。在一些实施方案中,TadA*8是TadA*8.1、TadA*8.2、TadA*8.3、TadA*8.4、TadA*8.5、TadA*8.6、TadA*8.7、TadA*8.8、TadA*8.9、TadA*8.10、TadA*8.11、TadA*8.12、TadA*8.13、TadA*8.14、TadA*8.15、TadA*8.16、TadA*8.17、TadA*8.18、TadA*8.19、TadA*8.20、TadA*8.21、TadA*8.22、TadA*8.23或TadA*8.24。在一些实施方案中,TadA*8是TadA*8a、TadA*8b、TadA*8c、TadA*8d或TadA*8e。在一些实施方案中,TadA*8是TadA*8e。在一些实施方案中,TadA变体是TadA*9。

[0069] “腺苷脱氨酶碱基编辑器8(ABE8)多肽”或“ABE8”意为本文所定义的包含腺苷脱氨酶变体的碱基编辑器,该腺苷脱氨酶变体包含在以下参考序列的氨基酸位置82和/或166处的改变:MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

[0070] 在一些实施方案中,ABE8包含相对于参考序列的如本文所述的进一步改变。

[0071] “腺苷脱氨酶碱基编辑器8(ABE8)多肽”意为编码ABE8的多核苷酸。

[0072] “腺苷脱氨酶碱基编辑器9(ABE9)多肽”或“ABE9”意为本文所定义的包含腺苷脱氨酶变体的碱基编辑器,该腺苷脱氨酶变体包含以下改变中的一者或多者:R21N、R23H、E25F、N38G、L51W、P54C、M70V、Q71M、N72K、Y73S、V82T、M94V、P124W、T133K、D139L、D139M、C146R和A158K,这些改变在以下参考序列中:

[0073]

```
         10         20         30         40         50
MSEVEFSHEY WMRHALTLAK RARDEREVPV GAVLVLNNRV IGEGWNRAIG
         60         70         80         90        100
LHDPTAHAEI MALRQGGLVM QNYRLIDATL YVTFEPCVMC AGAMIHSRIG
        110        120        130        140        150
RVVFGVRNAK TGAAGSLMDV LHYPGMNHRV EITEGILADE CAALLCYFFR
        160
MPRQVFNAQK KAQSSTD
```

[0074] 在参考序列中改变的相关碱基以下划线和加粗字体显示。在一些实施方案中,ABE9包含相对于参考序列的如本文所述的进一步改变。ABE9碱基编辑器的细节在国际PCT申请号PCT/2020/049975中描述,其通过引用以其整体并入本文。

[0075] “腺苷脱氨酶碱基编辑器9(ABE9)多肽”意为编码ABE9的多核苷酸。

[0076] 本文中,“给药”指代向患者或受试者提供一种或多种本文所述的组合物。例如但不限于,组合物给药(例如,注射)可以通过静脉内(i.v.)注射、皮下(s.c.)注射、皮内(i.d.)注射、腹腔内(i.p.)注射或肌肉内(i.m.)注射执行。可采用一种或多种此类途径。肠胃外给药可以通过例如积存注射或通过随时间推移的渐进输液进行。在一些实施方案中,肠胃外给药包括血管内、静脉内、肌肉内、动漫、鞘内、肿瘤内、皮内、腹膜内、透过器官、皮下、表皮下、关节内、包膜下、蛛网膜下和胸骨内输液或注射。或者或同时,给药可以通过口服途径进行。

[0077] “药剂”意为任何小分子化学化合物、抗体、核酸分子或多肽,或其片段。

[0078] “改变”意为通过该领域的标准已知方法例如本文所述的那些方法检测的基因或多肽的结构、表达水平或活性的变化(例如增加或减少)。如本文所用,改变(例如增加或减少)包括多核苷酸或多肽的变化或表达水平的变化,诸如10%的变化、25%的变化、40%的变化、或50%的变化或更高。

[0079] 如本文所用,“同种异体的”是指在基因上与被比较细胞有所不同的同一物种的细胞。

[0080] “缓解”意为减少、压制、衰减、消除、阻止疾病的发展或进展或令疾病的发展或进展稳定化。

[0081] “类似物”意为不相同但具有类似的功能或结构特征的分子。例如,多核苷酸或多肽类似物保留对应的天然存在的多核苷酸或多肽的生物学活性,同时具有某些修饰,这些修饰将所述类似物的功能相对于天然存在的多核苷酸或多肽增强。此类修饰可增加所述类似物对DNA的亲和性、效率、特异性、蛋白酶或蛋白酶耐性、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。

[0082] “抗瘤形成活性”意为预防或抑制肿瘤的突变和/或增殖。

[0083] 如本文所用,“自体的”是指来自同一受试者的细胞。

[0084] 如本文所用,术语“抗体”是指一种免疫球蛋白分子,其特异性地结合至特定抗原或具有与特定抗原的免疫学反应性,并且包括多克隆、单克隆、基因工程改造的和其他修饰形式的抗体,包括但不限于嵌合抗体、人源化抗体、异源缀合物抗体(例如,双特异性抗体、三特异性抗体、四特异性抗体、双体抗体、三体抗体和四体抗体)和抗体的抗原结合片段,包括例如Fab'、F(ab')2、Fab、Fv、rlgG和scFv片段。除非另外指出,否则术语“单克隆抗体”(mAb)意为包括完整分子以及能够特异性地结合至靶标蛋白的抗体片段(包括例如Fab和F(ab')2片段)两者。如本文所用,Fab和F(ab')2片段是指缺乏完整抗体的Fc片段的抗体片段。这些抗体片段示例在本文中描述。

[0085] “B细胞成熟抗原,或肿瘤坏死因子受体超家族成员17多肽,(BCMA)”意为一种在成熟B淋巴细胞上表达的蛋白质,其与NCBI登录号NP_001183或其片段具有至少约85%的氨基酸序列一致性。示例性BCMA多肽序列提供如下。

[0086] >NP_001183.2肿瘤坏死因子受体超家族成员17[智人(Homo sapiens)]

[0087] MLQMAGQCSQNEYFDSLLHACIPCQLRCSSNTPPLTCQRYCNASVTNSVKGTNAILWTCLGLSLIISLA
VFVLMFLLRKINSEPLKDEFKNTGSGLLGMANIDLEKSRTGDEIILPRGLEYTVEECTCEDCIKSKPKVDSDHCFPL
PAMEEGATILVTTKTNDYCKSLPAALSATEIEKSISAR

[0088] 在复发性或难治性多发性骨髓瘤和其他血液肿瘤疗法中可以靶向该抗原。

[0089] “B细胞成熟抗原,或肿瘤坏死因子受体超家族成员17,(BCMA)多核苷酸”意为编码BCMA多肽的核酸分子。BCMA基因编码细胞表面抗原,该细胞表面抗原识别B细胞激活因子。示例性B2M多核苷酸序列提供如下。

[0090] >NM_001192.2智人(Homo sapiens)TNF受体超家族成员17(TNFRSF17),mRNA

[0091] AAGACTCAAACTTAGAAACTTGAATTAGATGTGGTATTCAAATCCTTAGCTGCCGCGAAGACACAGACA
GCCCCCGTAAGAACCCACGAAGCAGGCGAAGTTCATTGTTCTCAACATTCTAGCTGCTCTTGCTGCATTTGCTCTGG
AATTCTTGTAGAGATATTACTTGTCCTTCCAGGCTGTTCTTTCTGTAGCTCCCTTGTTTTCTTTTTGTGATCATGTT
GCAGATGGCTGGGCAGTGCTCCCAAAATGAATATTTTGACAGTTTGTTGCATGCTTGCATACCTTGTCAACTTCGAT
GTTCTTCTAATACTCCTCCTCTAACATGTCAGCGTTATTGTAATGCAAGTGTGACCAATTCAGTGAAAGGAACGAAT
GCGATTCTCTGGACCTGTTTGGGACTGAGCTTAATAATTTCTTTGGCAGTTTTCGTGCTAATGTTTTTGCTAAGGAA
GATAAACTCTGAACCATTAAAGGACGAGTTTAAAAACACAGGATCAGGTCTCCTGGGCATGGCTAACATTGACCTGG
AAAAGAGCAGGACTGGTGATGAAATTATTCTTCCGAGAGGCCTCGAGTACACGGTGGAAGAATGCACCTGTGAAGAC
TGCATCAAGAGCAAACCGAAGGTCGACTCTGACCATTGCTTTCCACTCCCAGCTATGGAGGAAGGCGCAACCATTCT
TGTCACCACGAAAACGAATGACTATTGCAAGAGCCTGCCAGCTGCTTTGAGTGCTACGGAGATAGAGAAATCAATTT
CTGCTAGGTAATTAACCATTTCGACTCGAGCAGTGCCACTTTAAAAATCTTTTGTCAGAATAGATGATGTGTCAGAT
CTCTTTAGGATGACTGTATTTTTCAGTTGCCGATACAGCTTTTTGTCCTCTAACTGTGGAAACTCTTTATGTTAGAT
ATATTTCTCTAGGTTACTGTTGGGAGCTTAATGGTAGAAACTTCCTTGGTTTCATGATTAAACTCTTTTTTTTCCTG
A

[0092] “碱基编辑器(BE)”或“核碱基编辑器(NBE)”意为结合多核苷酸并具有核碱基修饰活性的药剂。在一种实施方案中,该药剂使用核酸可编程DNA结合蛋白在特异性序列处结合多核苷酸。在另一实施方案中,碱基编辑器是能够修饰核酸分子(例如DNA)内的胞苷碱基的酶。在一些实施方案中,碱基编辑器能够将核酸分子内的碱基脱氨基。在一些实施方案中,碱基编辑器能够将DNA分子内的碱基脱氨基。在一些实施方案中,碱基编辑器能够将DNA内

的胞苷脱氨基。在一些实施方案中,碱基编辑器是包含胞苷脱氨酶或腺苷脱氨酶的融合蛋白。在一些实施方案中,碱基编辑器是融合至胞苷脱氨酶或腺苷脱氨酶的Cas9蛋白。在一些实施方案中,碱基编辑器是融合至胞苷脱氨酶或腺苷脱氨酶的Cas9切口酶(nCas9)。在一些实施方案中,碱基编辑器融合至碱基切除修复的抑制剂,例如,UGI结构域。在一些实施方案中,融合蛋白包含Cas9切口酶,该切口酶融合至脱氨酶和碱基切除修复的抑制剂,诸如UG结构域。

[0093] 在一些实施方案中,胞苷脱氨酶或腺苷脱氨酶核碱基编辑器是包含以下结构域A-B的多肽:NH_2-[A-B]-COOH,其中A包含胞苷脱氨酶结构域、腺苷脱氨酶结构域或其活性片段,并且其中B包含一个或多个具有核酸序列特异性结合活性的结构域。在一种实施方案中,先前方面的胞苷或腺苷脱氨酶核碱基编辑器多肽含有:NH_2-[A_n-B_o]-COOH,其中A包含胞苷脱氨酶结构域、腺苷脱氨酶结构域或其活性片段,其中n是整数:1、2、3、4或5;并且其中B包含具有核酸序列特异性结合活性的结构域;并且其中o是整数:1、2、3、4或5。在一种实施方案中,多肽含有一个或多个核定位序列。在一种实施方案中,多肽含有位于N端或C端的至少一个所述核定位序列。在一种实施方案中,多肽含有作为二分体核定位信号的核定位信号。在一种实施方案中,多肽含有一个或多个通过连接子链接的结构域。

[0094] 在一些实施方案中,碱基编辑器是胞苷碱基编辑器(CBE)。在一些实施方案中,碱基编辑器是腺苷碱基编辑器(ABE)。在一些实施方案中,碱基编辑器是腺苷碱基编辑器(ABE)和胞苷碱基编辑器(CBE)。在一些实施方案中,碱基编辑器是融合至腺苷脱氨酶的无核酸酶活性的Cas9(dCas9)。在一些实施方案中,Cas9是环状排列的(circular permutant)Cas9(例如,spCas9或saCas9)。环状排列的Cas9是本领域中已知的并且在例如Oakes等人,Cell 176,254-267,2019中描述。在一些实施方案中,碱基编辑器融合至碱基切除修复的抑制剂,例如,UGI结构域或dISN结构域。在一些实施方案中,融合蛋白包含Cas9切口酶,该切口酶融合至脱氨酶和碱基切除修复的抑制剂,诸如UGI或dISN结构域。在其他实施方案中,碱基编辑器是无碱基的碱基编辑器。

[0095] 在一些实施方案中,腺苷脱氨酶从TadA进化而来。在一些实施方案中,多核苷酸可编程DNA结合结构域是CRISPR相关(例如,Cas或Cpf1)酶。在一些实施方案中,碱基编辑器是融合至脱氨酶结构域的催化上死亡的Cas9(dCas9)。在一些实施方案中,碱基编辑器是融合至脱氨酶结构域的Cas9切口酶(nCas9)。在一些实施方案中,碱基编辑器融合至碱基切除修复(BER)的抑制剂。在一些实施方案中,碱基切除修复的抑制剂是尿嘧啶DNA糖基化酶抑制剂(UGI)。在一些实施方案中,碱基切除修复的抑制剂是肌苷碱基切除修复抑制剂。

[0096] 在一些实施方案中,通过将腺苷脱氨酶变体(例如,TadA*7.10)克隆到支架中来生成碱基编辑器,该支架包括环状排列的Cas9(例如,spCAS9)和二分体核定位序列。在一些实施方案中,通过将腺苷脱氨酶变体(例如,TadA*8)克隆到支架中来生成碱基编辑器(例如,ABE8),该支架包括环状排列的Cas9(例如,spCAS9或saCAS9)和二分体核定位序列。环状排列的Cas9是本领域中已知的并且在例如Oakes等人,Cell 176,254-267,2019中描述。

[0097] 在一些实施方案中,多核苷酸可编程DNA结合结构域是CRISPR相关(例如,Cas或Cpf1)酶。在一些实施方案中,碱基编辑器是融合至脱氨酶结构域的催化上死亡的Cas9(dCas9)。在一些实施方案中,碱基编辑器是融合至脱氨酶结构域的Cas9切口酶(nCas9)。在一些实施方案中,碱基编辑器融合至碱基切除修复(BER)的抑制剂。在一些实施方案中,碱

基切除修复的抑制剂是尿嘧啶DNA糖基化酶抑制剂(UGI)。在一些实施方案中,碱基切除修复的抑制剂是肌苷碱基切除修复抑制剂。

[0098] 碱基编辑器的细节在国际PCT申请号PCT/2017/045381(WO 2018/027078)和PCT/US2016/058344(WO 2017/070632)中描述,其各自通过引用以其整体并入本文。也参见Komor,A.C.,等人,“Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage”Nature 533,420-424(2016);Gaudelli,N.M.,等人,“Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage”Nature 551,464-471(2017);Komor,A.C.,等人,“Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:Abase editors with higher efficiency and product purity”Science Advances 3:eaao4774(2017);和Rees,H.A.,等人,“Base editing:precision chemistry on the genome and transcriptome of living cells.”Nat Rev Genet.2018Dec;19(12):770-788.doi:10.1038/s41576-018-0059-1,其整体内容通过引用并入本文。

[0099] 例如,用于本文所述的碱基编辑组合物、系统和方法中的腺苷碱基编辑器(ABE)具有如下所述的核酸序列(8877个碱基对)(Addgene,Watertown,MA.;Gaudelli NM,等人,Nature.2017Nov23;551(7681):464-471.doi:10.1038/nature24644;Koblan LW,等人,Nat Biotechnol.2018Oct;36(9):843-846.doi:10.1038/nbt.4172.)。也涵盖与ABE核酸序列具有至少95%或更高一致性的多核苷酸序列。

[0100] ATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTAGAGATCCGCGGCCGCTAATACGACTCACTATAGGGAGAGCCGCCACCATGAAACGGACAGCCGACGGAAGCGAGTTCGAGTCACCAAAGAAGAAGCGGAAAGTCTCTGAAGTCGAGTTTAGCCACGAGTATTGGATGAGGCACGCACTGACCCTGGCAAAGCGAGCATGGGATGAAAGAGAAGTCCCCGTGGGCGCCGTGCTGGTGCACAACAATAGAGTGATCGGAGAGGGATGGAACAGGCCAATCGGCCGCCACGACCCTACCGCACACGCAGAGATCATGGCACTGAGGCAGGGAGGCCTGGTCATGCAGAATTACCGCCTGATCGATGCCACCCTGTATGTGACACTGGAGCCATGCGTGATGTGCGCAGGAGCAATGATCCACAGCAGGATCGGAAGAGTGGTGTTCGGAGCACGGGACGCCAAGACCGGCGCAGCAGGCTCCCTGATGGATGTGCTGCACCACCCCGGCATGAACCACCGGGTGGAGATCACAGAGGGAATCCTGGCAGACGAGTGCGCCGCCCTGCTGAGCGATTTCTTTAGAATGCGGAGACAGGAGATCAAGGCCCAGAAGAAGGCACAGAGCTCCACCGACTCTGGAGGATCTAGCGGAGGATCCTCTGGAAGCGAGACACCAGGCACAAGCGAGTCCGCCACACCAGAGAGCTCCGGCGGCTCCTCCGGAGGATCCTCTGAGGTGGAGTTTTCCCACGAGTACTGGATGAGACATGCCCTGACCCTGGCCAAGAGGGCACGCGATGAGAGGGAGGTGCCTGTGGGAGCCGTGCTGGTGCTGAACAATAGAGTGATCGGCGAGGGCTGGAACAGAGCCATCGGCCTGCACGACCCAACAGCCCATGCCGAAATTATGGCCCTGAGACAGGGCGGCCTGGTCATGCAGAACTACAGACTGATTGACGCCACCCTGTACGTGACATTCGAGCCTTGCGTGATGTGCGCCGGCGCCATGATCCACTCTAGGATCGGCCGCGTGGTGTTTGGCGTGAGGAACGCAAAAACCGGCGCCGCAGGCTCCCTGATGGACGTGCTGCACTACCCCGGCATGAATCACCGCGTCGAAATTACCGAGGGAATCCTGGCAGATGAATGTGCCGCCCTGCTGTGCTATTTCTTTCGGATGCCTAGACAGGTGTTCAATGCTCAGAAGAAGGCCCAGAGCTCCACCGACTCCGGAGGATCTAGCGGAGGCTCCTCTGGCTCTGAGACACCTGGCACAAGCG

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

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

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

[0101] 在一些实施方案中,碱基编辑器是腺苷脱氨酶碱基编辑器8(ABE8)。在一些实施方案中,ABE8选自来自下文表13、14或16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。在一些实施方案中,ABE8的腺苷脱氨酶变体是如下文表11、13或14中所述的TadA*8变体。在一些实施方案中,腺苷脱氨酶是TadA*7.10变体(例如,TadA*8),其包含选自由以下组的改变中的一者或多者:Y147T、Q154S、Y123H、V82S、T166R和/或Q154R。在各种实施方案中,ABE8包含TadA*7.10变体(例如,TadA*8),该变体具有选自以下组的改变的组合:Y147T+Q154R;Y147T+Q154S;Y147R+Q154S;V82S+Q154S;V82S+Y147R;V82S+Q154R;V82S+Y123H;I76Y+V82S;V82S+Y123H+Y147T;V82S+Y123H+Y147R;V82S+Y123H+Q154R;Y147R+Q154R+Y123H;Y147R+Q154R+I76Y;Y147R+Q154R+T166R;Y123H+Y147R+Q154R+I76Y;V82S+Y123H+Y147R+Q154R;和I76Y+V82S+Y123H+Y147R+Q154R。在一些实施方案中,ABE8是单体构建体。在一些实施方案中,ABE8是异二聚构建体。在一些实施方案中,ABE8碱基编辑器包含序列:

[0102] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAAL
LCTFFRMPRQVFNAQKKAQSSTD。

[0103] 例如,如在本文所述的碱基编辑组合物、系统和方法中使用的胞苷碱基编辑器(CBE)具有如下提供的以下核酸序列(8877个碱基对)(Addgene,Watertown,MA.;Komor AC,等人,2017,Sci Adv.,30;3(8):eaao4774.doi:10.1126/sciadv.aao4774)。也涵盖与BE4核酸序列具有至少95%或更高一致性的多核苷酸序列。

[0104]

```
1 ATATGCCAAG TACGCCCCCT 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
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
```

[0105]

```
2581 ATGACAAGAA AGTCCGAAGA AACGATTACT CCATGGAATT TTGAGGAAGT TGTCGATAAA
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 TCCTTTTTGA AGGACGATTC AATCGACAAT AAAGTGCTTA CACGCTCGGA TAAGAACCGA
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
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
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 CCCCCTGACG
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
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
[0106] 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

[0107] 在一些实施方案中,胞苷碱基编辑器是具有选自以下中的一个的核酸序列的BE4:

[0108] 原始BE4核酸序列:

[0109] ATGagctcagagactggcccagtggctgtggaccccacattgagacggcggatcgagccccatgagtt
tgaggtattcttcgatccgagagagctccgcaaggagacctgcctgctttacgaaattaattgggggggccggcac
tccatttggcgacatacatcacagaacactaacaagcacgtcgaagtcaacttcatcgagaagttcacgacagaaa
gatatttctgtccgaacacaaggtgcagcattacctggtttctcagctggagccgcgaatgtagtagggccatcac
tgaattcctgtcaaggtatccccacgtcactctgtttatttacatcgcaaggctgtaccaccacgctgacccccgc
aatcgacaaggcctgcgggatttgatctcttcaggtgtgactatccaaattatgactgagcaggagtcaggatact
gctggagaaactttgtgaattatagcccgagtaatgaagcccactggcctaggtatccccatctgtgggtacgact
gtacgttcttgaactgtactgcatcatactgggcctgcctccttgtctcaacattctgagaaggaagcagccacag
ctgacattctttaccatcgctcttcagtcttgtcattaccagcgactgcccccacacattctctgggccaccgggt
tgaaatctggtggttcttctggtggttctagcggcagcgagactcccgggacctcagagtccgccacacccgaaag

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

ccaccatgcgcacgacgcttatcttaatgccgtcgtagggaccgcactcattaagaaatacccgaagctagaaagt
gagtttgtgtatggtgattacaaagtttatgacgtccgtaagatgatcgcgaaaagcgaacaggagataggcaagg
ctacagccaaatacttcttttattctaacattatgaatttctttaagacggaaatcactctggcaaacggagagat
acgcaaacgacctttaattgaaaccaatggggagacaggtgaaatcgtatgggataagggccgggacttcgcgacg
gtgagaaaagttttgtccatgccccaagtcaacatagtaaagaaaactgaggtgcagaccggagggttttcaaagg
aatcgattcttccaaaaaggaatagtgataagctcatcgctcgtaaaaaggactgggacccgaaaaagtacggtgg
cttcgatagccctacagttgcctattctgtcctagtagtggcaaaagttgagaagggaaaatccaagaaactgaag
tcagtcaaagaattattggggataacgattatggagcgctcgtctttgaaaagaaccccatcgacttccttgagg
cgaaaggttacaaggaagtaaaaaaggatctcataattaaactaccaaagtatagtctgtttgagttagaaaatgg
ccgaaaacggatgttggctagcgccggagagcttcaaaaggggaacgaactcgcactaccgtctaaatacgtgaat
ttcctgtatttagcgtcccattacgagaagttgaaaggttcacctgaagataacgaacagaagcaacttttgttg
agcagcacaaacattatctcgacgaaatcatagagcaaatttcggaattcagtaagagagtcatcctagctgatgc
caatctggacaaagtattaagcgcatacaacaagcacagggataaacccatacgtgagcaggcggaaaatattatc
catttgtttactcttaccaacctcggcgctccagccgcattcaagtattttgacacaacgatagatcgcaaacgat
acacttctaccaaggaggtgctagacgcgacactgattcaccaatccatcacgggattatatgaaactcggataga
tttgtcacagcttgggggtgactctggtggttctggaggatctggtggttctactaatctgtcagatattattgaa
aaggagaccggtaagcaactggttatccaggaatccatcctcatgctcccagaggaggtggaagaagtcattggga
acaagccggaaagcgatatactcgtgcacaccgcctacgacgagagcaccgacgagaatgtcatgcttctgactag
cgacgcccctgaatacaagccttgggctctggtcatacaggatagcaacggtgagaacaagattaagatgctctct
ggtggttctggaggatctggtggttctactaatctgtcagatattattgaaaaggagaccggtaagcaactggtta
tccaggaatccatcctcatgctcccagaggaggtggaagaagtcattgggaacaagccggaaagcgatatactcgt
gcacaccgcctacgacgagagcaccgacgagaatgtcatgcttctgactagcgacgcccctgaatacaagccttgg
gctctggtcatacaggatagcaacggtgagaacaagattaagatgctctctggtggttctAAAAGGACGGCGGACG
GATCAGAGTTCGAGAGTCCGAAAAAAAAACGAAAGGTCGAAtaa

[0110] BE4密码子优化1核酸序列：

[0111] ATGTCATCCGAAACCGGGCCAGTGGCCGTAGACCCAACACTCAGGAGGCGGATAGAACCCCATGAGTT
TGAAGTGTTCTTCGACCCCAGAGAGCTGCGCAAAGAGACTTGCCTCCTGTATGAAATAAATTGGGGGGGTCGCCAT
TCAATTTGGAGGCACACTAGCCAGAATACTAACAAACACGTGGAGGTAAATTTTATCGAGAAGTTTACCACCGAAA
GATACTTTTGCCCCAATACACGGTGTTCAATTACCTGGTTTCTGTCATGGAGTCCATGTGGAGAATGTAGTAGAGC
GATAACTGAGTTCCTGTCTCGATATCCTCACGTCACGTTGTTTATATACATCGCTCGGCTTTATCACCATGCGGAC
CCGCGGAACAGGCAAGGTCTTCGGGACCTCATATCCTCTGGGGTGACCATCCAGATAATGACGGAGCAAGAGAGCG
GATACTGCTGGCGAAACTTTGTTAACTACAGCCCAAGCAATGAGGCACACTGGCCTAGATATCCGCATCTCTGGGT
TCGACTGTATGTCCTTGAACTGTACTGCATAATTCTGGGACTTCCGCCATGCTTGAACATTCTGCGGCGGAAACAA
CCACAGCTGACCTTTTTCACGATTGCTCTCCAAAGTTGTCACTACCAGCGATTGCCACCCCACATCTTGTGGGCTA
CTGGACTCAAGTCTGGAGGAAGTTCAGGCGGAAGCAGCGGGTCTGAAACGCCCGGAACCTCAGAGAGCGCAACGCC
CGAAAGCTCTGGAGGGTCAAGTGGTGGTAGTGATAAGAAATACTCCATCGGCCTCGCCATCGGTACGAATTCTGTC
GGTTGGGCCGTTATCACCGATGAGTACAAGGTCCCTTCTAAGAAATTCAAGGTTTTGGGCAATACAGACCGCCATT
CTATAAAAAAAAACCTGATCGGCGCCCTTTTGTTTGACAGTGGTGAGACTGCTGAAGCGACTCGCCTGAAGCGAAC
TGCCAGGAGGCGGTATACGAGGCGAAAAAACCGAATTTGTTACCTCCAGGAGATTTTCTCAAATGAAATGGCCAAG

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

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

[0112] BE4密码子优化2核酸序列：

[0113] ATGAGCAGCGAGACAGGCCCTGTGGCTGTGGATCCTACACTGCGGAGAAGAATCGAGCCCCACGAGTT
CGAGGTGTTCTTCGACCCCAGAGAGCTGCGGAAAGAGACATGCCTGCTGTACGAGATCAACTGGGGCGGCAGACAC
TCTATCTGGCGGCACACAAGCCAGAACACCAACAAGCACGTGGAAGTGAACTTTATCGAGAAGTTTACGACCGAGC
GGTACTTCTGCCCCAACACCAGATGCAGCATCACCTGGTTTCTGAGCTGGTCCCCTTGCGGCGAGTGCAGCAGAGC
CATCACCGAGTTTCTGTCCAGATATCCCCACGTGACCCTGTTCATCTATATCGCCCGGCTGTACCACCACGCCGAT
CCTAGAAATAGACAGGGACTGCGCGACCTGATCAGCAGCGGAGTGACCATCCAGATCATGACCGAGCAAGAGAGCG
GCTACTGCTGGCGGAACTTCGTGAACTACAGCCCCAGCAACGAAGCCCACTGGCCTAGATATCCTCACCTGTGGGT
CCGACTGTACGTGCTGGAACTGTACTGCATCATCCTGGGCCTGCCTCCATGCCTGAACATCCTGAGAAGAAAGCAG
CCTCAGCTGACCTTCTTCACAATCGCCCTGCAGAGCTGCCACTACCAGAGACTGCCTCCACACATCCTGTGGGCCA
CCGGACTTAAGAGCGGAGGATCTAGCGGCGGCTCTAGCGGATCTGAGACACCTGGCACAAGCGAGTCTGCCACACC
TGAGAGTAGCGGCGGATCTTCTGGCGGCTCCGACAAGAAGTACTCTATCGGACTGGCCATCGGCACCAACTCTGTT
GGATGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACA
GCATCAAGAAGAATCTGATCGGCGCCCTGCTGTTCGACTCTGGCGAAACAGCCGAAGCCACCAGACTGAAGAGAAC
CGCCAGGCGGAGATACACCCGGCGGAAGAACCGGATCTGCTACCTGCAAGAGATCTTCAGCAACGAGATGGCCAAG
GTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGACAAGAAGCACGAGCGGCACCCCA
TCTTCGGCAACATCGTGGATGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGT
GGACAGCACCGACAAGGCCGACCTGAGACTGATCTACCTGGCTCTGGCCCACATGATCAAGTTCCGGGGCCACTTT
CTGATCGAGGGCGATCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACC

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

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

[0114] “碱基编辑活性”意为发挥作用以在化学上改变多核苷酸内的碱基。在一种实施方案中,第一碱基被转变为第二碱基。在一种实施方案中,碱基编辑活性是胞苷脱氨酶活性,例如,将靶标C·G转变为T·A。在另一实施方案中,碱基编辑活性是腺苷或腺嘌呤脱氨酶活性,例如,将靶标A·T转变为G·C。在另一实施方案中,碱基编辑活性是胞苷脱氨酶活性,例如将靶标C·G转变为T·A,和腺苷或腺嘌呤脱氨酶活性,例如,将靶标A·T转变为G·C。

[0115] 在一些实施方案中,通过编辑效率评估碱基编辑活性。可以通过任何合适手段,例如,通过测序或下一代测序来测量碱基编辑效率。在一些实施方案中,通过具有由碱基编辑器实现的核碱基转变的总测序读数的百分比,例如,具有靶标A·T碱基对转变为G·C碱基对或者靶标C·G碱基对转变为T·A碱基对的总测序读数的百分比,来测量碱基编辑效率。在一些实施方案中,通过当在细胞群中执行碱基编辑时具有由碱基编辑器实现的核碱基转变的总细胞的百分比来测量碱基编辑效率。

[0116] 术语“碱基编辑器系统”是指用于编辑靶标核苷酸序列的核碱基的系统。在各种实施方案中,碱基编辑器系统包含(1)多核苷酸可编程核苷酸结合结构域(例如Cas9);(2)用于将所述核碱基脱氨基的脱氨酶结构域(例如腺苷脱氨酶或胞苷脱氨酶);和(3)一个或多个向导多核苷酸(例如,向导RNA)。在一些实施方案中,多核苷酸可编程核苷酸结合结构域是多核苷酸可编程DNA结合结构域。在一些实施方案中,碱基编辑器是胞苷碱基编辑器(CBE)。在一些实施方案中,碱基编辑器是腺嘌呤或腺苷碱基编辑器(ABE)。在一些实施方案中,碱基编辑器系统是腺苷脱氨酶碱基编辑器8(ABE8)。

[0117] 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。在一些实施方案中,ABE8是异源构建体。在一些实施方案中,ABE8是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。

[0118] 在一些实施方案中,碱基编辑器系统可包含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个脱氨酶。在一些实施方案中,碱基编辑器酮可包括一个或多个胞苷脱氨酶。在一些实施方案中,碱基编辑器酮可包括一个或多个腺苷脱氨酶。在一些实施方案中,可利用单向导多核苷酸将不同的脱氨酶靶向至靶标核酸序列。在一些实施方案中,可利用一对向导多核苷酸将不同的脱氨酶靶向至靶标核酸序列。

[0119] 碱基编辑器系统核碱基组分和多核苷酸可编程核苷酸结合组分可以共价地或非共价地彼此缔合,或其缔合和相互作用的任意组合。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核苷酸结合结构域被靶向至靶标核苷酸序列。在一些实施方案中,多核苷酸可编程核苷酸结合结构域可融合或链接至脱氨酶结构域。在一些实施方案中,多核苷酸可编程核苷酸结合结构域可通过与脱氨酶结构域非共价地相互作用或缔合而将脱氨酶结构域靶向至靶标核苷酸序列。例如,在一些实施方案中,核碱基编辑组分,例如,脱氨酶结构域可包含额外的异源部分或结构域,该异源部分或结构域能够与作为多核苷酸可编程核苷酸结合结构域的一部分的额外异源部分或结构域相互作用、缔合或形成复合物。在一些实施方案中,额外的异源部分可能能够与多肽结合、相互作用、缔合或形成复合物。在一些实施方案中,额外的异源部分可能能够与多核苷酸结合、相互作用、缔合或形成复合物。在一些实施方案中,额外的异源部分可能能够结合至向导多核苷酸。在一些实施方案中,额外的异源部分可能能够结合至多肽连接子。在一些实施方案中,额外的异源部分可能能够结合至多核苷酸连接子。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(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。

[0120] 碱基编辑器系统可进一步包含向导多核苷酸组分。应了解,碱基编辑器系统的组分可以经由共价键、非共价相互作用、或其缔合和相互作用的任意组合彼此缔合。在一些实施方案中,脱氨结构域可以通过向导多核苷酸被靶向至靶标核苷酸序列。例如,在一些实施方案中,碱基编辑器系统的核碱基编辑组分,例如脱氨酶结构域,可包含额外的异源部分或结构域(例如,多核苷酸结合结构域诸如RNA或DNA结合蛋白),该额外的异源部分或结构域能够与向导多核苷酸的部分或链段(例如,多核苷酸基序)相互作用、缔合或形成复合物。在一些实施方案中,额外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)可融合或链接至脱氨酶结构域。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、缔合或形成复合物。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、缔合或形成复合物。在一些实施方案中,额外的异源部分可能能够结合至向导多核苷酸。在一些实施方案中,额外的异源部分可能能够结合至多肽连接子。在一些实施方案中,额外的异源部分可能能够结合至多核苷酸连接子。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(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。

[0121] 在一些实施方案中,碱基编辑器系统可进一步包含碱基切除修复(BER)的抑制剂组分。应了解,碱基编辑器系统的组分可以经由共价键、非共价相互作用、或其缔合和相互作用的任意组合彼此缔合。BER的抑制剂组分可包含BER抑制剂。在一些实施方案中,BER的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的抑制剂可以是肌苷BER抑制剂。在一些实施方案中,BER的抑制剂可以通过多核苷酸可编程核苷酸结合结构域被靶向至靶标核苷酸序列。在一些实施方案中,多核苷酸可编程核苷酸结合结构域可融合或链接至BER的抑制剂。在一些实施方案中,多核苷酸可编程核苷酸结合结构域可融合或链接至脱氨酶结构域和BER的抑制剂。在一些实施方案中,多核苷酸可编程核苷酸结合结构域可通过与BER的抑制剂非共价地相互作用或缔合而将BER的抑制剂靶向至靶标核苷酸序列。例如,在一些实施方案中,BER的抑制剂组分可包含额外的异源部分或结构域,该异源部分或结构域能够与作为多核苷酸可编程核苷酸结合结构域的一部分的额外异源部分或结构域相互作用、缔合或形成复合物。

[0122] 在一些实施方案中,BER的抑制剂可以通过向导多核苷酸被靶向至靶标核苷酸序列。例如,在一些实施方案中,碱基编辑器系统的BER的抑制剂组分,可包含额外的异源部分或结构域(例如,多核苷酸结合结构域诸如RNA或DNA结合蛋白),该额外的异源部分或结构域能够与向导多核苷酸的部分或链段(例如,多核苷酸基序)相互作用、缔合或形成复合物。在一些实施方案中,向导多核苷酸的额外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)可融合或链接至BER的抑制剂。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、缔合或形成复合物。在一些实施方案中,额外的异源部分可能能够结合至向导多核苷酸。在一些实施方案中,额外的异源部分可能能够结合至多肽连接子。在一些实施方案中,额外的异源部分可能能够结合至多核苷酸连接子。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(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。

[0123] “β-2微球蛋白(B2M)多肽”意为一种蛋白质,其与UniProt登录号P61769或其片段具有至少约85%的氨基酸序列一致性并且具有免疫调节活性。示例性B2M多肽序列提供如下。

[0124] >sp|P61769|B2MG_HUMANβ-2-微球蛋白OS=智人(Homo sapiens)OX=9606GN=B2M PE=1SV=1

[0125] MSRSVALAVLALLSLSGLEAIQRTPKIQVYSRHPAENGKSNFLNCYVSGFHPSDIEVDLLKNGERIEKVEHSDLSFSKDWSFYLLYYTEFTPTEKDEYACRVNHVTLSQPKIVKWDRDM

[0126] “β-2微球蛋白(B2M)多核苷酸”意为一种编码B2M多肽的核酸分子。β-2微球蛋白基因编码与主要组织相容性复合体缔合的血清蛋白。B2M牵涉到由宿主CD8+ T细胞进行的非自我识别中。示例性B2M多核苷酸序列提供如下。

[0127] >DQ217933.1智人(Homo sapiens)β-2微球蛋白(B2M)基因,完全编码序列

[0128] CATGTCATAAATGGTAAGTCCAAGAAAAATACAGGTATTCCCCCCCAAAGAAAACTGTAAAATCGACTTTTTTCTATCTGTACTGTTTTTTATTGGTTTTTAAATTGGTTTTCCAAGTGAGTAAATCAGAATCTATCTGTAATGGATTTTAAATTTAGTGTTTCTCTGTGATGTAGTAAACAAGAAACTAGAGGCAAAAATAGCCCTGTCCCTTGCTAAACTT

CTAAGGCACTTTTCTAGTACAACTCAACACTAACATTTCAGGCCTTTAGTGCCTTATATGAGTTTTTAAAAGGGGGA
AAAGGGAGGGAGCAAGAGTGTCTTAACTCATACATTTAGGCATAACAATTATTCTCATATTTTAGTTATTGAGAGGG
CTGGTAGAAAAACTAGGTAAATAATATTAATAATTATAGCGCTTATTAAACACTACAGAACACTTACTATGTACCAG
GCATTGTGGGAGGCTCTCTCTTGTGCATTATCTCATTTCATTAGGTCCATGGAGAGTATTGCATTTTCTTAGTTTAG
GCATGGCCTCCACAATAAAGATTATCAAAAGCCTAAAAATATGTAAAAGAAACCTAGAAGTTATTTGTTGTGCTCCT
TGGGGAAGCTAGGCAAATCCTTTCAACTGAAAACCATGGTGACTTCCAAGATCTCTGCCCCTCCCCATCGCCATGGT
CCACTTCCTCTTCTCACTGTTCCTCTTAGAAAAGATCTGTGGACTCCACCACCACGAAATGGCGGCACCTTATTTAT
GGTCACTTTAGAGGGTAGGTTTTCTTAATGGGTCTGCCTGTCATGTTTAACGTCCTTGGCTGGGTCCAAGGCAGATG
CAGTCCAAACTCTCACTAAAATTGCCGAGCCCTTTGTCTTCCAGTGTCTAAAATATTAATGTCAATGGAATCAGGCC
AGAGTTTGAATTCTAGTCTCTTAGCCTTTGTTTCCCCTGTCCATAAAATGAATGGGGGTAATTCTTTCCTCCTACAG
TTTATTTATATATTCACTAATTCATTCATTCATCCATCCATTCGTTCATTCGGTTTACTGAGTACCTACTATGTGCC
AGCCCCTGTTCTAGGGTGGAAACTAAGAGAATGATGTACCTAGAGGGCGCTGGAAGCTCTAAAGCCCTAGCAGTTAC
TGCTTTTACTATTAGTGGTCGTTTTTTTCTCCCCCCCCGCCCCCCGACAAATCAACAGAACAAAGAAAATTACCTAAA
CAGCAAGGACATAGGGAGGAACTTCTTGGCACAGAACTTTCCAAACACTTTTTCCTGAAGGGATACAAGAAGCAAGA
AAGGTACTCTTTCACTAGGACCTTCTCTGAGCTGTCCTCAGGATGCTTTTGGGACTATTTTTCTTACCCAGAGAATG
GAGAAACCCTGCAGGGAATTCCCAAGCTGTAGTTATAAACAGAAGTTCTCCTTCTGCTAGGTAGCATTCAAAGATCT
TAATCTTCTGGGTTTCCGTTTTCTCGAATGAAAAATGCAGGTCCGAGCAGTTAACTGGCTGGGGCACCATTAGCAAG
TCACTTAGCATCTCTGGGGCCAGTCTGCAAAGCGAGGGGGCAGCCTTAATGTGCCTCCAGCCTGAAGTCCTAGAATG
AGCGCCCGGTGTCCCAAGCTGGGGCGCGCACCCCAGATCGGAGGGCGCCGATGTACAGACAGCAAACTCACCCAGTC
TAGTGCATGCCTTCTTAAACATCACGAGACTCTAAGAAAAGGAAACTGAAAACGGGAAAGTCCCTCTCTCTAACCTG
GCACTGCGTCGCTGGCTTGGAGACAGGTGACGGTCCCTGCGGGCCTTGTCCTGATTGGCTGGGCACGCGTTTAATAT
AAGTGGAGGCGTCGCGCTGGCGGGCATTCCTGAAGCTGACAGCATTCGGGCCGAGATGTCTCGCTCCGTGGCCTTAG
CTGTGCTCGCGCTACTCTCTCTTTCTGGCCTGGAGGCTATCCAGCGTGAGTCTCTCCTACCCTCCCGCTCTGGTCCT
TCCTCTCCCGCTCTGCACCCTCTGTGGCCCTCGCTGTGCTCTCTCGCTCCGTGACTTCCCTTCTCCAAGTTCTCCTT
GGTGGCCCGCCGTGGGGCTAGTCCAGGGCTGGATCTCGGGGAAGCGGCGGGGTGGCCTGGGAGTGGGGAAGGGGGTG
CGCACCCGGGACGCGCGCTACTTGCCCCTTTCGGCGGGGAGCAGGGGAGACCTTTGGCCTACGGCGACGGGAGGGTC
GGGACAAAGTTTAGGGCGTCGATAAGCGTCAGAGCGCCGAGGTTGGGGGAGGGTTTCTCTTCCGCTCTTTCGCGGGG
CCTCTGGCTCCCCCAGCGCAGCTGGAGTGGGGGACGGGTAGGCTCGTCCCAAAGGCGCGGCGCTGAGGTTTGTGAAC
GCGTGGAGGGGCGCTTGGGGTCTGGGGGAGGCGTCGCCCGGGTAAGCCTGTCTGCTGCGGCTCTGCTTCCCTTAGAC
TGGAGAGCTGTGGACTTCGTCTAGGCGCCCGCTAAGTTCGCATGTCCTAGCACCTCTGGGTCTATGTGGGGCCACAC
CGTGGGGAGGAAACAGCACGCGACGTTTGTAGAATGCTTGGCTGTGATACAAAGCGGTTTCGAATAATTAACTTATT
TGTTCCCATCACATGTCACTTTTAAAAAATTATAAGAACTACCCGTTATTGACATCTTTCTGTGTGCCAAGGACTTT
ATGTGCTTTGCGTCATTTAATTTTGAAAACAGTTATCTTCCGCCATAGATAACTACTATGGTTATCTTCTGCCTCTC
ACAGATGAAGAAACTAAGGCACCGAGATTTTAAGAAACTTAATTACACAGGGGATAAATGGCAGCAATCGAGATTGA
AGTCAAGCCTAACCAGGGCTTTTGCGGGAGCGCATGCCTTTTGGCTGTAATTCGTGCATTTTTTTTTAAGAAAAACG
CCTGCCTTCTGCGTGAGATTCTCCAGAGCAAACTGGGCGGCATGGGCCCTGTGGTCTTTTCGTACAGAGGGCTTCCT
CTTTGGCTCTTTGCCTGGTTGTTTCCAAGATGTACTGTGCCTCTTACTTTCGGTTTTGAAAACATGAGGGGGTTGGG
CGTGGTAGCTTACGCCTGTAATCCCAGCACTTAGGGAGGCCGAGGCGGGAGGATGGCTTGAGGTCCGTAGTTGAGAC
CAGCCTGGCCAACATGGTGAAGCCTGGTCTCTACAAAAAATAATAACAAAAATTAGCCGGGTGTGGTGGCTCGTGCC

TGTGGTCCCAGCTGCTCCGGTGGCTGAGGCGGGAGGATCTCTTGAGCTTAGGCTTTTGAGCTATCATGGCGCCAGTG
CACTCCAGCGTGGGCAACAGAGCGAGACCCTGTCTCTCAAAAAAGAAAAAAAAAAAAAAAGAAAGAGAAAAGAAAAG
AAAGAAAGAAGTGAAGGTTTGTCAGTCAGGGGAGCTGTAAAACCATTAATAAAGATAATCCAAGATGGTTACCAAGA
CTGTTGAGGACGCCAGAGATCTTGAGCACTTTCTAAGTACCTGGCAATACACTAAGCGCGCTCACCTTTTCCTCTGG
CAAAACATGATCGAAAGCAGAATGTTTTGATCATGAGAAAATTGCATTTAATTTGAATACAATTTATTTACAACATA
AAGGATAATGTATATATCACCACCATTACTGGTATTTGCTGGTTATGTTAGATGTCATTTTAAAAAATAACAATCTG
ATATTTAAAAAAAAATCTTATTTTGAAAATTTCCAAAGTAATACATGCCATGCATAGACCATTTCTGGAAGATACCA
CAAGAAACATGTAATGATGATTGCCTCTGAAGGTCTATTTTCCTCCTCTGACCTGTGTGTGGGTTTTGTTTTTGTTT
TACTGTGGGCATAAATTAATTTTTCAGTTAAGTTTTGGAAGCTTAAATAACTCTCCAAAAGTCATAAAGCCAGTAAC
TGGTTGAGCCCAAATTCAAACCCAGCCTGTCTGATACTTGTCCTCTTCTTAGAAAAGATTACAGTGATGCTCTCACA
AAATCTTGCCGCCTTCCCTCAAACAGAGAGTTCCAGGCAGGATGAATCTGTGCTCTGATCCCTGAGGCATTTAATAT
GTTCTTATTATTAGAAGCTCAGATGCAAAGAGCTCTCTTAGCTTTTAATGTTATGAAAAAAATCAGGTCTTCATTAG
ATTCCCCAATCCACCTCTTGATGGGGCTAGTAGCCTTTCCTTAATGATAGGGTGTTTCTAGAGAGATATATCTGGTC
AAGGTGGCCTGGTACTCCTCCTTCTCCCCACAGCCTCCCAGACAAGGAGGAGTAGCTGCCTTTTAGTGATCATGTAC
CCTGAATATAAGTGTATTTAAAAGAATTTTATACACATATATTTAGTGTCAATCTGTATATTTAGTAGCACTAACAC
TTCTCTTCATTTTCAATGAAAAATATAGAGTTTATAATATTTTCTTCCCACTTCCCCATGGATGGTCTAGTCATGCC
TCTCATTTTGGAAAGTACTGTTTCTGAAACATTAGGCAATATATTCCCAACCTGGCTAGTTTACAGCAATCACCTGT
GGATGCTAATTAAAACGCAAATCCCACTGTCACATGCATTACTCCATTTGATCATAATGGAAAGTATGTTCTGTCCC
ATTTGCCATAGTCCTCACCTATCCCTGTTGTATTTTATCGGGTCCAACTCAACCATTTAAGGTATTTGCCAGCTCTT
GTATGCATTTAGGTTTTGTTTCTTTGTTTTTTAGCTCATGAAATTAGGTACAAAGTCAGAGAGGGGTCTGGCATATA
AAACCTCAGCAGAAATAAAGAGGTTTTGTTGTTTGGTAAGAACATACCTTGGGTTGGTTGGGCACGGTGGCTCGTGC
CTGTAATCCCAACACTTTGGGAGGCCAAGGCAGGCTGATCACTTGAAGTTGGGAGTTCAAGACCAGCCTGGCCAACA
TGGTGAAATCCCGTCTCTACTGAAAATACAAAAATTAACCAGGCATGGTGGTGTGTGCCTGTAGTCCCAGGAATCAC
TTGAACCCAGGAGGCGGAGGTTGCAGTGAGCTGAGATCTCACCACTGCACACTGCACTCCAGCCTGGGCAATGGAAT
GAGATTCCATCCCAAAAAATAAAAAAATAAAAAAATAAAGAACATACCTTGGGTTGATCCACTTAGGAACCTCAGAT
AATAACATCTGCCACGTATAGAGCAATTGCTATGTCCCAGGCACTCTACTAGACACTTCATACAGTTTAGAAAATCA
GATGGGTGTAGATCAAGGCAGGAGCAGGAACCAAAAAGAAAGGCATAAACATAAGAAAAAAAATGGAAGGGGTGGAA
ACAGAGTACAATAACATGAGTAATTTGATGGGGGCTATTATGAACTGAGAAATGAACTTTGAAAAGTATCTTGGGGC
CAAATCATGTAGACTCTTGAGTGATGTGTTAAGGAATGCTATGAGTGCTGAGAGGGCATCAGAAGTCCTTGAGAGCC
TCCAGAGAAAGGCTCTTAAAAATGCAGCGCAATCTCCAGTGACAGAAGATACTGCTAGAAATCTGCTAGAAAAAAAA
CAAAAAAGGCATGTATAGAGGAATTATGAGGGAAAGATACCAAGTCACGGTTTATTCTTCAAAATGGAGGTGGCTTG
TTGGGAAGGTGGAAGCTCATTTGGCCAGAGTGGAAATGGAATTGGGAGAAATCGATGACCAAATGTAAACACTTGGT
GCCTGATATAGCTTGACACCAAGTTAGCCCCAAGTGAAATACCCTGGCAATATTAATGTGTCTTTTCCCGATATTCC
TCAGGTACTCCAAAGATTCAGGTTTACTCACGTCATCCAGCAGAGAATGGAAAGTCAAATTTCCTGAATTGCTATGT
GTCTGGGTTTCATCCATCCGACATTGAAGTTGACTTACTGAAGAATGGAGAGAGAATTGAAAAAGTGGAGCATTCAG
ACTTGTCTTTCAGCAAGGACTGGTCTTTCTATCTCTTGTACTACACTGAATTCACCCCCACTGAAAAAGATGAGTAT
GCCTGCCGTGTGAACCATGTGACTTTGTCACAGCCCAAGATAGTTAAGTGGGGTAAGTCTTACATTCTTTTGTAAGC
TGCTGAAAGTTGTGTATGAGTAGTCATATCATAAAGCTGCTTTGATATAAAAAAGGTCTATGGCCATACTACCCTGA
ATGAGTCCCATCCCATCTGATATAAACAATCTGCATATTGGGATTGTCAGGGAATGTTCTTAAAGATCAGATTAGTG

GCACCTGCTGAGATACTGATGCACAGCATGGTTTCTGAACCAGTAGTTTCCCTGCAGTTGAGCAGGGAGCAGCAGCA
GCACTTGCACAAATACATATACACTCTTAACACTTCTTACCTACTGGCTTCCTCTAGCTTTTGTGGCAGCTTCAGGT
ATATTTAGCACTGAACGAACATCTCAAGAAGGTATAGGCCTTTGTTTGTAAGTCCTGCTGTCCTAGCATCCTATAAT
CCTGGACTTCTCCAGTACTTTCTGGCTGGATTGGTATCTGAGGCTAGTAGGAAGGGCTTGTTCCTGCTGGGTAGCTC
TAAACAATGTATTCATGGGTAGGAACAGCAGCCTATTCTGCCAGCCTTATTTCTAACCATTTTAGACATTTGTTAGT
ACATGGTATTTTAAAAGTAAAACTTAATGTCTTCCTTTTTTTTCTCCACTGTCTTTTTCATAGATCGAGACATGTAA
GCAGCATCATGGAGGTAAGTTTTTGACCTTGAGAAAATGTTTTTGTTTCACTGTCCTGAGGACTATTTATAGACAGC
TCTAACATGATAACCCTCACTATGTGGAGAACATTGACAGAGTAACATTTTAGCAGGGAAAGAAGAATCCTACAGGG
TCATGTTCCCTTCTCCTGTGGAGTGGCATGAAGAAGGTGTATGGCCCCAGGTATGGCCATATTACTGACCCTCTACA
GAGAGGGCAAAGGAACTGCCAGTATGGTATTGCAGGATAAAGGCAGGTGGTTACCCACATTACCTGCAAGGCTTTGA
TCTTTCTTCTGCCATTTCCACATTGGACATCTCTGCTGAGGAGAGAAAATGAACCACTCTTTTCCTTTGTATAATGT
TGTTTTATTCTTCAGACAGAAGAGAGGAGTTATACAGCTCTGCAGACATCCCATTCCTGTATGGGGACTGTGTTTGC
CTCTTAGAGGTTCCCAGGCCACTAGAGGAGATAAAGGGAAACAGATTGTTATAACTTGATATAATGATACTATAATA
GATGTAACTACAAGGAGCTCCAGAAGCAAGAGAGAGGGAGGAACTTGGACTTCTCTGCATCTTTAGTTGGAGTCCAA
AGGCTTTTCAATGAAATTCTACTGCCCAGGGTACATTGATGCTGAAACCCCATTCAAATCTCCTGTTATATTCTAGA
ACAGGGAATTGATTTGGGAGAGCATCAGGAAGGTGGATGATCTGCCCAGTCACACTGTTAGTAAATTGTAGAGCCAG
GACCTGAACTCTAATATAGTCATGTGTTACTTAATGACGGGGACATGTTCTGAGAAATGCTTACACAAACCTAGGTG
TTGTAGCCTACTACACGCATAGGCTACATGGTATAGCCTATTGCTCCTAGACTACAAACCTGTACAGCCTGTTACTG
TACTGAATACTGTGGGCAGTTGTAACACAATGGTAAGTATTTGTGTATCTAAACATAGAAGTTGCAGTAAAAATATG
CTATTTTAATCTTATGAGACCACTGTCATATATACAGTCCATCATTGACCAAAACATCATATCAGCATTTTTTCTTC
TAAGATTTTGGGAGCACCAAAGGGATACACTAACAGGATATACTCTTTATAATGGGTTTGGAGAACTGTCTGCAGCT
ACTTCTTTTAAAAAGGTGATCTACACAGTAGAAATTAGACAAGTTTGGTAATGAGATCTGCAATCCAAATAAAATAA
ATTCATTGCTAACCTTTTTCTTTTCTTTTCAGGTTTGAAGATGCCGCATTTGGATTGGATGAATTCCAAATTCTGCT
TGCTTGCTTTTTAATATTGATATGCTTATACACTTACACTTTATGCACAAAATGTAGGGTTATAATAATGTTAACAT
GGACATGATCTTCTTTATAATTCTACTTTGAGTGCTGTCTCCATGTTTGATGTATCTGAGCAGGTTGCTCCACAGGT
AGCTCTAGGAGGGCTGGCAACTTAGAGGTGGGGAGCAGAGAATTCTCTTATCCAACATCAACATCTTGGTCAGATTT
GAACTCTTCAATCTCTTGCACTCAAAGCTTGTTAAGATAGTTAAGCGTGCATAAGTTAACTTCCAATTTACATACTC
TGCTTAGAATTTGGGGGAAAATTTAGAAATATAATTGACAGGATTATTGGAAATTTGTTATAATGAATGAAACATTT
TGTCATATAAGATTCATATTTACTTCTTATACATTTGATAAAGTAAGGCATGGTTGTGGTTAATCTGGTTTATTTTT
GTTCCACAAGTTAAATAAATCATAAAACTTGATGTGTTATCTCTTATATCTCACTCCCACTATTACCCCTTTATTTT
CAAACAGGGAAACAGTCTTCAAGTTCCACTTGGTAAAAAATGTGAACCCCTTGTATATAGAGTTTGGCTCACAGTGT
AAAGGGCCTCAGTGATTCACATTTTCCAGATTAGGAATCTGATGCTCAAAGAAGTTAAATGGCATAGTTGGGGTGAC
ACAGCTGTCTAGTGGGAGGCCAGCCTTCTATATTTTAGCCAGCGTTCTTTCCTGCGGGCCAGGTCATGAGGAGTATG
CAGACTCTAAGAGGGAGCAAAAGTATCTGAAGGATTTAATATTTTAGCAAGGAATAGATATACAATCATCCCTTGGT
CTCCCTGGGGGATTGGTTTCAGGACCCCTTCTTGGACACCAAATCTATGGATATTTAAGTCCCTTCTATAAAATGGT
ATAGTATTTGCATATAACCTATCCACATCCTCCTGTATACTTTAAATCATTTCTAGATTACTTGTAATACCTAATAC
AATGTAAATGCTATGCAAATAGTTGTTATTGTTTAAGGAATAATGACAAGAAAAAAAAGTCTGTACATGCTCAGTAA
AGACACAACCATCCCTTTTTTTCCCCAGTGTTTTTGATCCATGGTTTGCTGAATCCACAGATGTGGAGCCCCTGGAT
ACGGAAGGCCCGCTGTACTTTGAATGACAAATAACAGATTTAAA

[0129] 术语“Cas9”或“Cas9结构域”是指RNA引导的核酸酶或其片段,其包含Cas9蛋白(例如,包含Cas9的活性、无活性或部分活性的DNA切割结构域和/或Cas9的gRNA结合结构域的蛋白质)。Cas9核酸酶有时是指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,其提供针对移动基因元件(病毒、转座因子和接合质粒)的保护。CRISPR簇含有间隔序列、与前驱移动元件互补的序列和靶标侵入核酸。CRISPR簇被转录并处理为CRISPR RNA(crRNA)。在II型CRISPR系统中,pre-crRNA的正确处理需要反向编码的小RNA(tracrRNA)、内源性核糖核酸酶3(rnc)和Cas9蛋白。tracrRNA充当向导用于核糖核酸酶3辅助的pre-crRNA处理。随后,Cas9/crRNA/tracrRNA核酸内切地切割与间隔序列互补的线性或环状dsDNA靶标。不与crRNA互补的靶标链首先被核酸内切地切断,然后被3’-5’核酸外切地修剪。在自然界,DNA结合和切割通常需要蛋白质和两种RNA。但是,单向导RNA(“sgRNA”或简写为“gRNA”)可以经工程改造以将多种crRNA和tracrRNA并入单一RNA物质中。参见,例如,Jinek M.等人,Charpentier E.Science 337:816-821(2012),其整体内容通过引用并入本文。Cas9识别CRISPR重复序列中的短基序(PAM或前间区序列邻近基序)以帮助区分自我与非自我。Cas9核酸酶序列和结构是本领域技术人员周知的(参见,例如,“Complete genome sequence of an M1 strain of Streptococcus pyogenes.”Ferretti,J.J.等人,Natl.Acad.Sci.U.S.A.98:4658-4663(2001);“CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.”Deltcheva E.等人,Nature 471:602-607(2011);和“A programmabledual-RNA-guided DNA endonuclease in adaptive bacterial immunity.”Jinek M.等人,Science 337:816-821(2012),其各自的整体内容通过引用并入本文)。Cas9直系同源物已经在各种物种中描述,包括酿脓链球菌(S.pyogenes)和嗜热链球菌(S.thermophilus)。基于本公开,其他合适的Cas9核酸酶和序列对于本领域技术人员将是显而易见的,并且此类Cas9核酸酶和序列包括来自生物体和基因座的Cas9序列,如Chylinski,Rhun,and Charpentier,“The tracrRNA and Cas9 families of type IICRISPR-Cas immunity systems”(2013)RNA Biology 10:5,726-737中所公开的;其整体内容通过引用并入本文。

[0130] 无核酸酶活性的Cas9蛋白可以互换地称为“dCas9”蛋白(对于核酸酶“死亡的”Cas9)或无催化活性的Cas9。用于生成具有无活性的DNA切割结构域的Cas9蛋白(或其片段)是已知的(参见,例如,Jinek等人,Science.337:816-821(2012);Qi等人,“Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression”(2013)Cell.28;152(5):1173-83,其各自的整体内容通过引用并入本文)。例如,已知Cas9的DNA切割结构域包括两个次结构域:HNH核酸酶次结构域和RuvC1次结构域。HNH次结构域切割与gRNA互补的链,而RuvC1次结构域切割非互补链。这些次结构域中的突变可以沉默Cas9的核酸酶活性。例如,突变D10A和H840A将酿脓链球菌Cas9的核酸酶活性完全灭活(Jinek等人,Science.337:816-821(2012);Qi等人,Cell.28;152(5):1173-83(2013))。在一些实施方案中,dCas9对应于或者部分或整体地包含具有一个或多个突变的Cas9氨基酸序列,该突变将Cas9核酸酶活性灭活。在一些实施方案中,dCas9结构域包含D10A和H840A突变或另一Cas9中的对应突变。在一些实施方案中,Cas9核酸酶具有失活的(例如,灭活的)DNA切割结构域,换言之,Cas9是切口酶,称为“nCas9”蛋白(对于“切口酶”Cas9)。应了解,其他Cas9蛋白(例如,核酸酶死亡的Cas9(dCas9)、Cas9切口酶(nCas9)、或核

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

[0131] 在一些实施方案中,提供包含Cas9的片段的蛋白质。例如,在一些实施方案中,蛋白质包含一个或两个Cas9结构域:(1)Cas9的gRNA结合结构域;或(2)Cas9的DNA切割结构域。在一些实施方案中,包含Cas9或其片段的蛋白质称为"Cas9变体"。Cas9变体与Cas9或其片段共享同源性。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、45、46、47、48、49、50或更多个氨基酸变化。在一些实施方案中,Cas9变体包含Cas9的片段(例如,gRNA结合结构域或DNA切割结构域),使得该片段与野生型Cas9的对应片段至少约70%相同、至少约80%相同、至少约90%相同、至少约95%相同、至少约96%相同、至少约97%相同、至少约98%相同、至少约99%相同、至少约99.5%相同、或至少约99.9%相同。在一些实施方案中,该片段是对应的野生型Cas9的氨基酸长度的至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%相同、至少96%、至少97%、至少98%、至少99%、或至少99.5%。

[0132] 在一些实施方案中,该片段是至少100个氨基酸的长度。在一些实施方案中,该片段是至少100、150、200、250、300、350、400、450、500、550、600、650、700、750、800、850、900、950、1000、1050、1100、1150、1200、1250、或至少1300个氨基酸的长度。

[0133] 在一些实施方案中,Cas9是指来自以下的Cas9:溃疡棒状杆菌(Corynebacterium ulcerans)(NCBI Refs:NC_015683.1,NC_017317.1)、白喉棒状杆菌(Corynebacterium diphtheria)(NCBI Refs:NC_016782.1,NC_016786.1)、螺原体(Spiroplasmasyrphidicola)(NCBI Ref:NC_021284.1)、中间普雷沃菌(Prevotellaintermedia)(NCBI Ref:NC_017861.1)、螺原体(Spiroplasmataiwanense)(NCBI Ref:NC_021846.1)、海豚链球菌(Streptococcusiniae)(NCBI Ref:NC_021314.1)、Belliella baltica(NCBI Ref:NC_018010.1)、扭曲冷弯曲菌(Psychroflexus torquis)(NCBI Ref:NC_018721.1)、嗜热链球菌(Streptococcus thermophilus)(NCBI Ref:YP_820832.1)、无害利斯特菌(Listeria innocua)(NCBI Ref:NP_472073.1)、空肠弯曲杆菌(Campylobacter jejuni)(NCBI Ref:YP_002344900.1)、或脑膜炎奈瑟氏菌(Neisseria meningitidis)(NCBI Ref:YP_002342100.1);或是指来自任何其他生物体的Cas9。

[0134] 在一些实施方案中,Cas9来自脑膜炎奈瑟氏菌(Neisseria meningitidis)(Nme)。在一些实施方案中,Cas9是Nme1、Nme2或Nme3。在一些实施方案中,用于Nme1、Nme2或Nme3的PAM相互作用结构域分别是N_4GAT、N_4CC和N_4CAAA(参见,例如,Edraki,A.,等人,A Compact, High-Accuracy Cas9 with a Dinucleotide PAM for In Vivo Genome Editing, Molecular Cell(2018))。

[0135] 在一些实施方案中,如本文所提供的Cas9融合蛋白包含Cas9蛋白(例如,本文所提

供的Cas9序列中的一者)的全长氨基酸序列。然而,在其他实施方案中,本文所提供的融合蛋白不包含全长Cas9序列,而仅包含其一个或多个片段。例如,在一些实施方案中,本文所提供的Cas9融合蛋白包含Cas9片段,其中该片段结合crRNA和tracrRNA或sgRNA,但不包含功能性核酸酶结构域,例如,在它仅包含截短版本的核酸酶结构域或根本不包含核酸酶结构域的情况下。

[0136] 合适的Cas9结构域和Cas9片段的示例性氨基酸序列提供在本文中,且Cas9结构域和片段的其他合适的序列对于本领域技术人员将是显而易见的。

[0137] 在一些实施方案中,Cas9是指来自古生菌(例如,纳米古生菌)的Cas9,其构成单细胞原核微生物的领域和王国。在一些实施方案中,Cas9是指CasX或CasY,其已经描述于例如Burstein等人,"New CRISPR-Cas systems from uncultivated microbes."Cell Res.2017Feb 21.doi:10.1038/cr.2017.21中,其整体内容通过引用并入本文。使用基因组解析的宏基因组学,鉴定了大量CRISPR-Cas系统,包括在古生菌生命领域首次报告的Cas9。这种不同的Cas9蛋白是在很少研究的纳米古生菌中作为活性CRISPR-Cas系统发现的。在细菌中发现了两种先前未知的系统,CRISPR-CasX和CRISPR-CasY,它们是迄今发现的最为紧凑的系统中的一个。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。应了解,其他RNA引导的DNA结合蛋白可以用作核酸可编程DNA结合蛋白(napDNAbp),并且出于本公开的范畴内。

[0138] 在特定实施方案中,可用于本发明方法中的napDNAbp包括环状排列物,它们是本领域中已知的并且由例如Oakes等人,Cell 176,254-267,2019描述。

[0139] 可并入碱基编辑器的多核苷酸可编程核苷酸结合结构域的非限制性示例包括衍生自CRISPR蛋白的结构域、限制性核酸酶、巨核酸酶、TAL核酸酶(TALEN)和锌指核酸酶(ZFN)。

[0140] 在一些实施方案中,本文所提供的核酸可编程DNA结合蛋白(napDNAbp)或任何融合蛋白可以是CasX或CasY蛋白。在一些实施方案中,napDNAbp是CasX蛋白。在一些实施方案中,napDNAbp是CasY蛋白。在一些实施方案中,napDNAbp包含氨基酸序列,该氨基酸序列与天然存在的CasX或CasY蛋白至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%、或至少99.5%相同。在一些实施方案中,napDNAbp是天然存在的CasX或CasY蛋白。在一些实施方案中,napDNAbp包含氨基酸序列,该氨基酸序列与本文所述的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。应了解,来自其他细菌物种的CasX和CasY也可根据本公开使用。

[0141] 术语"Cas12b"或"Cas12b结构域"是指RNA引导的核酸酶或其片段,其包含Cas12b、C2c1蛋白(例如,包含Cas12b的活性、无活性或部分活性的DNA切割结构域和/或Cas12b的gRNA结合结构域的蛋白质)。其各自的内容通过引用并入本文。Cas12b同源物已经在各种物种中描述,包括但不限于,抗酸土脂环酸芽孢杆菌(Alicyclobacillus acidoterrestris)、嗜酸脂环酸芽孢杆菌(Alicyclobacillus acidophilus)(Teng等人,Cell Discov.2018Nov 27;4:63)、外村尚芽孢杆菌(Bacillus hisashi)和芽孢杆菌属(Bacillus sp.)V3-13。基于本发明,其他合适的Cas12b核酸酶序列对于本领域技术人员将是显而易见的。

[0142] 在一些实施方案中,包含Cas12b或其片段的蛋白质称为"Cas12b变体"。Cas12b变

体与Cas12b或其片段共享同源性。例如,Cas12b变体与野生型Cas12b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Cas12b相比,Cas12b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、45、46、47、48、49、50或更多个氨基酸变化。在一些实施方案中,Cas12b变体包含Cas12b的片段(例如,gRNA结合结构域或DNA切割结构域),使得该片段与野生型Cas12b的对应片段至少约70％相同、至少约80％相同、至少约90％相同、至少约95％相同、至少约96％相同、至少约97％相同、至少约98％相同、至少约99％相同、至少约99.5％相同、或至少约99.9％相同。在一些实施方案中,该片段是对应的野生型Cas12b的氨基酸长度的至少30％、至少35％、至少40％、至少45％、至少50％、至少55％、至少60％、至少65％、至少70％、至少75％、至少80％、至少85％、至少90％、至少95％相同、至少96％、至少97％、至少98％、至少99％、或至少99.5％。示例性Cas12b多肽在本文中列出。

[0143] “Cbl原癌基因B(CBLB)多肽”意为一种蛋白质,其与GenBank登录号ABC86700.1或其片段具有至少约85％的氨基酸序列一致性并且牵涉到免疫应答的调节中。示例性CBLB多肽序列提供如下。

[0144] >ABC86700.1 CBL-B[智人(Homo sapiens)]

[0145] MANSMNGRNPGGRGGNPRKGRILGIIDAIQDAVGPPKQAAADRRTVEKTWKLMDKVVRLCQNPKLQLKN
SPPYILDILPDTYQHLRLILSKYDDNQKLAQLSENEYFKIYIDSLMKKSKRAIRLFKEGKERMYEEQSQDRRNLTKL
SLIFSHMLAEIKAIFPNGQFQGDNFRITKADAAEFWRKFFGDKTIVPWKVFRQCLHEVHQISSGLEAMALKSTIDLT
CNDYISVFEFDIFTRLFQPWGSILRNWNFLAVTHPGYMAFLTYDEVKARLQKYSTKPGSYIFRLSCTRLGQWAIGYV
TGDGNILQTIPHNKPLFQALIDGSREGFYLYPDGRSYNPDLTGLCEPTPHDHIKVTQEQYELYCEMGSTFQLCKICA
ENDKDVKIEPCGHLMCTSCLTAWQESDGQGCPFCRCEIKGTEPIIVDPFDPRDEGSRCCSIIDPFGMPMLDLDDDDD
REESLMMNRLANVRKCTDRQNSPVTSPGSSPLAQRRKPQPDPLQIPHLSLPPVPPRLDLIQKGIVRSPCGSPTGSPK
SSPCMVRKQDKPLPAPPPPLRDPPPPPPERPPPIPPDNRLSRHIHHVESVPSRDPPMPLEAWCPRDVFGTNQLVGCR
LLGEGSPKPGITASSNVNGRHSRVGSDPVLMRKHRRHDLPLEGAKVFSNGHLGSEEYDVPPRLSPPPPVTTLLPSIK
CTGPLANSLSEKTRDPVEEDDDEYKIPSSHPVSLNSQPSHCHNVKPPVRSCDNGHCMLNGTHGPSSEKKSNIPDLSI
YLKGDVFDSASDPVPLPPARPPTRDNPKHGSSLNRTPSDYDLLIPPLGEDAFDALPPSLPPPPPPARHSLIEHSKPP
GSSSRPSSGQDLFLLPSDPFVDLASGQVPLPPARRLPGENVKTNRTSQDYDQLPSCSDGSQAPARPPKPRPRRTAPE
IHHRKPHGPEAALENVDAKIAKLMGEGYAFEEVKRALEIAQNNVEVARSILREFAFPPPVSPRLNL

[0146] “Cbl原癌基因B(CBLB)多核苷酸”意为编码CBLB多肽的核酸分子。CBLB基因编码E3泛素连接酶。示例性CBLB核酸序列提供如下。

[0147] >DQ349203.1智人(Homo sapiens)CBL-B mRNA,完全编码序列ATGGCAAACTCAATGA
ATGGCAGAAACCCTGGTGGTCGAGGAGGAAATCCCCGAAAAGGTCGAATTTTGGGTATTATTGATGCTATTCAGGAT
GCAGTTGGACCCCCTAAGCAAGCTGCCGCAGATCGCAGGACCGTGGAGAAGACTTGGAAGCTCATGGACAAAGTGGT
AAGACTGTGCCAAAATCCCAAACTTCAGTTGAAAAATAGCCCACCATATATACTTGATATTTTGCCTGATACATATC
AGCATTTACGACTTATATTGAGTAAATATGATGACAACCAGAAACTTGCCCAACTCAGTGAGAATGAGTACTTTAAA
ATCTACATTGATAGCCTTATGAAAAAGTCAAAACGGGCAATAAGACTCTTTAAAGAAGGCAAGGAGAGAATGTATGA
AGAACAGTCACAGGACAGACGAAATCTCACAAAACTGTCCCTTATCTTCAGTCACATGCTGGCAGAAATCAAAGCAA

TCTTTCCCAATGGTCAATTCCAGGGAGATAACTTTCGTATCACAAAAGCAGATGCTGCTGAATTCTGGAGAAAGTTT
TTTGGAGACAAAACTATCGTACCATGGAAAGTATTCAGACAGTGCCTTCATGAGGTCCACCAGATTAGCTCTGGCCT
GGAAGCAATGGCTCTAAAATCAACAATTGATTTAACTTGCAATGATTACATTTCAGTTTTTGAATTTGATATTTTTA
CCAGGCTGTTTCAGCCTTGGGGCTCTATTTTGCGGAATTGGAATTTCTTAGCTGTGACACATCCAGGTTACATGGCA
TTTCTCACATATGATGAAGTTAAAGCACGACTACAGAAATATAGCACCAAACCCGGAAGCTATATTTTCCGGTTAAG
TTGCACTCGATTGGGACAGTGGGCCATTGGCTATGTGACTGGGGATGGGAATATCTTACAGACCATACCTCATAACA
AGCCCTTATTTCAAGCCCTGATTGATGGCAGCAGGGAAGGATTTTATCTTTATCCTGATGGGAGGAGTTATAATCCT
GATTTAACTGGATTATGTGAACCTACACCTCATGACCATATAAAAGTTACACAGGAACAATATGAATTATATTGTGA
AATGGGCTCCACTTTTCAGCTCTGTAAGATTTGTGCAGAGAATGACAAAGATGTCAAGATTGAGCCTTGTGGGCATT
TGATGTGCACCTCTTGCCTTACGGCATGGCAGGAGTCGGATGGTCAGGGCTGCCCTTTCTGTCGTTGTGAAATAAAA
GGAACTGAGCCCATAATCGTGGACCCCTTTGATCCAAGAGATGAAGGCTCCAGGTGTTGCAGCATCATTGACCCCTT
TGGCATGCCGATGCTAGACTTGGACGACGATGATGATCGTGAGGAGTCCTTGATGATGAATCGGTTGGCAAACGTCC
GAAAGTGCACTGACAGGCAGAACTCACCAGTCACATCACCAGGATCCTCTCCCCTTGCCCAGAGAAGAAAGCCACAG
CCTGACCCACTCCAGATCCCACATCTAAGCCTGCCACCCGTGCCTCCTCGCCTGGATCTAATTCAGAAAGGCATAGT
TAGATCTCCCTGTGGCAGCCCAACGGGTTCACCAAAGTCTTCTCCTTGCATGGTGAGAAAACAAGATAAACCACTCC
CAGCACCACCTCCTCCCTTAAGAGATCCTCCTCCACCGCCACCTGAAAGACCTCCACCAATCCCACCAGACAATAGA
CTGAGTAGACACATCCATCATGTGGAAAGCGTGCCTTCCAGAGACCCGCCAATGCCTCTTGAAGCATGGTGCCCTCG
GGATGTGTTTGGGACTAATCAGCTTGTGGGATGTCGACTCCTAGGGGAGGGCTCTCCAAAACCTGGAATCACAGCGA
GTTCAAATGTCAATGGAAGGCACAGTAGAGTGGGCTCTGACCCAGTGCTTATGCGGAAACACAGACGCCATGATTTG
CCTTTAGAAGGAGCTAAGGTCTTTTCCAATGGTCACCTTGGAAGTGAAGAATATGATGTTCCTCCCCGGCTTTCTCC
TCCTCCTCCAGTTACCACCCTCCTCCCTAGCATAAAGTGTACTGGTCCGTTAGCAAATTCTCTTTCAGAGAAAACAA
GAGACCCAGTAGAGGAAGATGATGATGAATACAAGATTCCTTCATCCCACCCTGTTTCCCTGAATTCACAACCATCT
CATTGTCATAATGTAAAACCTCCTGTTCGGTCTTGTGATAATGGTCACTGTATGCTGAATGGAACACATGGTCCATC
TTCAGAGAAGAAATCAAACATCCCTGACTTAAGCATATATTTAAAGGGAGATGTTTTTGATTCAGCCTCTGATCCCG
TGCCATTACCACCTGCCAGGCCTCCAACTCGGGACAATCCAAAGCATGGTTCTTCACTCAACAGGACGCCCTCTGAT
TATGATCTTCTCATCCCTCCATTAGGTGAAGATGCTTTTGATGCCCTCCCTCCATCTCTCCCACCTCCCCCACCTCC
TGCAAGGCATAGTCTCATTGAACATTCAAAACCTCCTGGCTCCAGTAGCCGGCCATCCTCAGGACAGGATCTTTTTC
TTCTTCCTTCAGATCCCTTTGTTGATCTAGCAAGTGGCCAAGTTCCTTTGCCTCCTGCTAGAAGGTTACCAGGTGAA
AATGTCAAAACTAACAGAACATCACAGGACTATGATCAGCTTCCTTCATGTTCAGATGGTTCACAGGCACCAGCCAG
ACCCCCTAAACCACGACCGCGCAGGACTGCACCAGAAATTCACCACAGAAAACCCCATGGGCCTGAGGCGGCATTGG
AAAATGTCGATGCAAAAATTGCAAAACTCATGGGAGAGGGTTATGCCTTTGAAGAGGTGAAGAGAGCCTTAGAGATA
GCCCAGAATAATGTCGAAGTTGCCCGGAGCATCCTCCGAGAATTTGCCTTCCCTCCTCCAGTATCCCCACGTCTAAA
TCTATAG

[0148] “嵌合抗原受体”或“CAR”意为合成的或工程改造的受体,其包含与一个或多个细胞内信号传导结构域(例如,T细胞信号传导结构域)接合的细胞外抗原结合结构域,该细胞外抗原结合结构域赋予免疫效应细胞以抗原特异性。在一些实施方案中,CAR包括跨膜结构域。

[0149] “嵌合抗原受体T细胞”或“CAR-T细胞”意为表达CAR的T细胞,其具有由抗体来源的靶向CAR的结构域所决定的抗原特异性。如本文所用“CAR-T细胞”包括T细胞或NK细胞。如本

文所用,“CAR-T细胞”包括经工程改造以表达CAR或T细胞受体(TCR)的细胞。在一些实施方案中,CAR-T细胞可以是T辅助CD4+和/或T效应CD8+细胞,任选以所定义比例存在。制备CAR(例如,用于治疗癌症)的方法可公开获得(参见例如,Park等人,Trends Biotechnol.,29:550-557,2011;Grupp等人,N Engl J Med.,368:1509-1518,2013;Han等人,J.Hematol Oncol.6:47,2013;Haso等人,(2013)Blood,121,1165-1174;PCT公开WO2012/079000、WO2013/059593和美国专利公开2012/0213783,其各自通过引用以其整体并入本文)。

[0150] “II类主要组织相容性复合体反式激活因子(CIITA)”意为一种蛋白质,其与NCBI登录号NP_001273331.1或其片段具有至少约85%的氨基酸序列一致性且具有免疫调节活性。示例性氨基酸序列提供如下。

[0151] >NP_001273331.1II类MHC反式激活因子亚型1[智人(Homo sapiens)]

[0152] MRCLAPRPAGSYLSEPQGSSQCATMELGPLEGGYLELLNSDADPLCLYHFYDQMDLAGEEEIELYSEPD
TDTINCDQFSRLLCDMEGDEETREAYANIAELDQYVFQDSQLEGLSKDIFIEHIGPDEVIGESMEMPAEVGQKSQKR
PFPEELPADLKHWKPAEPPTVVTGSLLVGPVSDCSTLPCLPLPALFNQEPASGQMRLEKTDQIPMPFSSSSLSCLNL
PEGPIQFVPTISTLPHGLWQISEAGTGVSSIFIYHGEVPQASQVPPPSGFTVHGLPTSPDRPGSTSPFAPSATDLPS
MPEPALTSRANMTEHKTSPTQCPAAGEVSNKLPKWPEPVEQFYRSLQDTYGAEPAGPDGILVEVDLVQARLERSSSK
SLERELATPDWAERQLAQGGLAEVLLAAKEHRRPRETRVIAVLGKAGQGKSYWAGAVSRAWACGRLPQYDFVFSVPC
HCLNRPGDAYGLQDLLFSLGPQPLVAADEVFSHILKRPDRVLLILDGFEELEAQDGFLHSTCGPAPAEPCSLRGLLA
GLFQKKLLRGCTLLLTARPRGRLVQSLSKADALFELSGFSMEQAQAYVMRYFESSGMTEHQDRALTLLRDRPLLLSH
SHSPTLCRAVCQLSEALLELGEDAKLPSTLTGLYVGLLGRAALDSPPGALAELAKLAWELGRRHQSTLQEDQFPSAD
VRTWAMAKGLVQHPPRAAESELAFPSFLLQCFLGALWLALSGEIKDKELPQYLALTPRKKRPYDNWLEGVPRFLAGL
IFQPPARCLGALLGPSAAASVDRKQKVLARYLKRLQPGTLRARQLLELLHCAHEAEEAGIWQHVVQELPGRLSFLGT
RLTPPDAHVLGKALEAAGQDFSLDLRSTGICPSGLGSLVGLSCVTRFRAALSDTVALWESLQQHGETKLLQAAEEKF
TIEPFKAKSLKDVEDLGKLVQTQRTRSSSEDTAGELPAVRDLKKLEFALGPVSGPQAFPKLVRILTAFSSLQHLDLD
ALSENKIGDEGVSQLSATFPQLKSLETLNLSQNNITDLGAYKLAEALPSLAASLLRLSLYNNCICDVGAESLARVLP
DMVSLRVMDVQYNKFTAAGAQQLAASLRRCPHVETLAMWTPTIPFSVQEHLQQQDSRISLR

[0153] “II类主要组织相容性复合体反式激活因子(CIITA)”意为编码CIITA多肽的核酸。示例性CIITA核酸序列提供如下。

[0154] >NM_001286402.1智人(Homo sapiens)II类主要组织相容性复合体反式激活因子(CIITA),转录物变体1,mRNA

[0155] GGTTAGTGATGAGGCTAGTGATGAGGCTGTGTGCTTCTGAGCTGGGCATCCGAAGGCATCCTTGGGGAA
GCTGAGGGCACGAGGAGGGGCTGCCAGACTCCGGGAGCTGCTGCCTGGCTGGGATTCCTACACAATGCGTTGCCTGG
CTCCACGCCCTGCTGGGTCCTACCTGTCAGAGCCCCAAGGCAGCTCACAGTGTGCCACCATGGAGTTGGGGCCCCTA
GAAGGTGGCTACCTGGAGCTTCTTAACAGCGATGCTGACCCCCTGTGCCTCTACCACTTCTATGACCAGATGGACCT
GGCTGGAGAAGAAGAGATTGAGCTCTACTCAGAACCCGACACAGACACCATCAACTGCGACCAGTTCAGCAGGCTGT
TGTGTGACATGGAAGGTGATGAAGAGACCAGGGAGGCTTATGCCAATATCGCGGAACTGGACCAGTATGTCTTCCAG
GACTCCCAGCTGGAGGGCCTGAGCAAGGACATTTTCATAGAGCACATAGGACCAGATGAAGTGATCGGTGAGAGTAT
GGAGATGCCAGCAGAAGTTGGGCAGAAAAGTCAGAAAAGACCCTTCCCAGAGGAGCTTCCGGCAGACCTGAAGCACT
GGAAGCCAGCTGAGCCCCCCACTGTGGTGACTGGCAGTCTCCTAGTGGGACCAGTGAGCGACTGCTCCACCCTGCCC
TGCCTGCCACTGCCTGCGCTGTTCAACCAGGAGCCAGCCTCCGGCCAGATGCGCCTGGAGAAAACCGACCAGATTCC

CATGCCTTTCTCCAGTTCCTCGTTGAGCTGCCTGAATCTCCCTGAGGGACCCATCCAGTTTGTCCCCACCATCTCCA
CTCTGCCCCATGGGCTCTGGCAAATCTCTGAGGCTGGAACAGGGGTCTCCAGTATATTCATCTACCATGGTGAGGTG
CCCCAGGCCAGCCAAGTACCCCCTCCCAGTGGATTCACTGTCCACGGCCTCCCAACATCTCCAGACCGGCCAGGCTC
CACCAGCCCCTTCGCTCCATCAGCCACTGACCTGCCCAGCATGCCTGAACCTGCCCTGACCTCCCGAGCAAACATGA
CAGAGCACAAGACGTCCCCCACCCAATGCCCGGCAGCTGGAGAGGTCTCCAACAAGCTTCCAAAATGGCCTGAGCCG
GTGGAGCAGTTCTACCGCTCACTGCAGGACACGTATGGTGCCGAGCCCGCAGGCCCGGATGGCATCCTAGTGGAGGT
GGATCTGGTGCAGGCCAGGCTGGAGAGGAGCAGCAGCAAGAGCCTGGAGCGGGAACTGGCCACCCCGGACTGGGCAG
AACGGCAGCTGGCCCAAGGAGGCCTGGCTGAGGTGCTGTTGGCTGCCAAGGAGCACCGGCGGCCGCGTGAGACACGA
GTGATTGCTGTGCTGGGCAAAGCTGGTCAGGGCAAGAGCTATTGGGCTGGGGCAGTGAGCCGGGCCTGGGCTTGTGG
CCGGCTTCCCCAGTACGACTTTGTCTTCTCTGTCCCCTGCCATTGCTTGAACCGTCCGGGGGATGCCTATGGCCTGC
AGGATCTGCTCTTCTCCCTGGGCCCACAGCCACTCGTGGCGGCCGATGAGGTTTTCAGCCACATCTTGAAGAGACCT
GACCGCGTTCTGCTCATCCTAGACGGCTTCGAGGAGCTGGAAGCGCAAGATGGCTTCCTGCACAGCACGTGCGGACC
GGCACCGGCGGAGCCCTGCTCCCTCCGGGGGCTGCTGGCCGGCCTTTTCCAGAAGAAGCTGCTCCGAGGTTGCACCC
TCCTCCTCACAGCCCGGCCCCGGGGCCGCCTGGTCCAGAGCCTGAGCAAGGCCGACGCCCTATTTGAGCTGTCCGGC
TTCTCCATGGAGCAGGCCCAGGCATACGTGATGCGCTACTTTGAGAGCTCAGGGATGACAGAGCACCAAGACAGAGC
CCTGACGCTCCTCCGGGACCGGCCACTTCTTCTCAGTCACAGCCACAGCCCTACTTTGTGCCGGGCAGTGTGCCAGC
TCTCAGAGGCCCTGCTGGAGCTTGGGGAGGACGCCAAGCTGCCCTCCACGCTCACGGGACTCTATGTCGGCCTGCTG
GGCCGTGCAGCCCTCGACAGCCCCCCCGGGGCCCTGGCAGAGCTGGCCAAGCTGGCCTGGGAGCTGGGCCGCAGACA
TCAAAGTACCCTACAGGAGGACCAGTTCCCATCCGCAGACGTGAGGACCTGGGCGATGGCCAAAGGCTTAGTCCAAC
ACCCACCGCGGGCCGCAGAGTCCGAGCTGGCCTTCCCCAGCTTCCTCCTGCAATGCTTCCTGGGGGCCCTGTGGCTG
GCTCTGAGTGGCGAAATCAAGGACAAGGAGCTCCCGCAGTACCTAGCATTGACCCCAAGGAAGAAGAGGCCCTATGA
CAACTGGCTGGAGGGCGTGCCACGCTTTCTGGCTGGGCTGATCTTCCAGCCTCCCGCCCGCTGCCTGGGAGCCCTAC
TCGGGCCATCGGCGGCTGCCTCGGTGGACAGGAAGCAGAAGGTGCTTGCGAGGTACCTGAAGCGGCTGCAGCCGGGG
ACACTGCGGGCGCGGCAGCTGCTGGAGCTGCTGCACTGCGCCCACGAGGCCGAGGAGGCTGGAATTTGGCAGCACGT
GGTACAGGAGCTCCCCGGCCGCCTCTCTTTTCTGGGCACCCGCCTCACGCCTCCTGATGCACATGTACTGGGCAAGG
CCTTGGAGGCGGCGGGCCAAGACTTCTCCCTGGACCTCCGCAGCACTGGCATTTGCCCCTCTGGATTGGGGAGCCTC
GTGGGACTCAGCTGTGTCACCCGTTTCAGGGCTGCCTTGAGCGACACGGTGGCGCTGTGGGAGTCCCTGCAGCAGCA
TGGGGAGACCAAGCTACTTCAGGCAGCAGAGGAGAAGTTCACCATCGAGCCTTTCAAAGCCAAGTCCCTGAAGGATG
TGGAAGACCTGGGAAAGCTTGTGCAGACTCAGAGGACGAGAAGTTCCTCGGAAGACACAGCTGGGGAGCTCCCTGCT
GTTCGGGACCTAAAGAAACTGGAGTTTGCGCTGGGCCCTGTCTCAGGCCCCCAGGCTTTCCCCAAACTGGTGCGGAT
CCTCACGGCCTTTTCCTCCCTGCAGCATCTGGACCTGGATGCGCTGAGTGAGAACAAGATCGGGGACGAGGGTGTCT
CGCAGCTCTCAGCCACCTTCCCCCAGCTGAAGTCCTTGGAAACCCTCAATCTGTCCCAGAACAACATCACTGACCTG
GGTGCCTACAAACTCGCCGAGGCCCTGCCTTCGCTCGCTGCATCCCTGCTCAGGCTAAGCTTGTACAATAACTGCAT
CTGCGACGTGGGAGCCGAGAGCTTGGCTCGTGTGCTTCCGGACATGGTGTCCCTCCGGGTGATGGACGTCCAGTACA
ACAAGTTCACGGCTGCCGGGGCCCAGCAGCTCGCTGCCAGCCTTCGGAGGTGTCCTCATGTGGAGACGCTGGCGATG
TGGACGCCCACCATCCCATTCAGTGTCCAGGAACACCTGCAACAACAGGATTCACGGATCAGCCTGAGATGATCCCA
GCTGTGCTCTGGACAGGCATGTTCTCTGAGGACACTAACCACGCTGGACCTTGAACTGGGTACTTGTGGACACAGCT
CTTCTCCAGGCTGTATCCCATGAGCCTCAGCATCCTGGCACCCGGCCCCTGCTGGTTCAGGGTTGGCCCCTGCCCGG
CTGCGGAATGAACCACATCTTGCTCTGCTGACAGACACAGGCCCGGCTCCAGGCTCCTTTAGCGCCCAGTTGGGTGG

ATGCCTGGTGGCAGCTGCGGTCCACCCAGGAGCCCCGAGGCCTTCTCTGAAGGACATTGCGGACAGCCACGGCCAGG
CCAGAGGGAGTGACAGAGGCAGCCCCATTCTGCCTGCCCAGGCCCCTGCCACCCTGGGGAGAAAGTACTTCTTTTTT
TTTATTTTTAGACAGAGTCTCACTGTTGCCCAGGCTGGCGTGCAGTGGTGCGATCTGGGTTCACTGCAACCTCCGCC
TCTTGGGTTCAAGCGATTCTTCTGCTTCAGCCTCCCGAGTAGCTGGGACTACAGGCACCCACCATCATGTCTGGCTA
ATTTTTCATTTTTAGTAGAGACAGGGTTTTGCCATGTTGGCCAGGCTGGTCTCAAACTCTTGACCTCAGGTGATCCA
CCCACCTCAGCCTCCCAAAGTGCTGGGATTACAAGCGTGAGCCACTGCACCGGGCCACAGAGAAAGTACTTCTCCAC
CCTGCTCTCCGACCAGACACCTTGACAGGGCACACCGGGCACTCAGAAGACACTGATGGGCAACCCCCAGCCTGCTA
ATTCCCCAGATTGCAACAGGCTGGGCTTCAGTGGCAGCTGCTTTTGTCTATGGGACTCAATGCACTGACATTGTTGG
CCAAAGCCAAAGCTAGGCCTGGCCAGATGCACCAGCCCTTAGCAGGGAAACAGCTAATGGGACACTAATGGGGCGGT
GAGAGGGGAACAGACTGGAAGCACAGCTTCATTTCCTGTGTCTTTTTTCACTACATTATAAATGTCTCTTTAATGTC
ACAGGCAGGTCCAGGGTTTGAGTTCATACCCTGTTACCATTTTGGGGTACCCACTGCTCTGGTTATCTAATATGTAA
CAAGCCACCCCAAATCATAGTGGCTTAAAACAACACTCACATTTA

[0156] 本公开中,“包含”、“包含有”、“含有”和“具有”等可具有美国专利法所规定的意义,并且可意为“包括”、“包括在内”等;“主要由...组成”或“基本由...组成”等具有美国专利法所规定的意义,并且该术语是开放性的,允许超过其所引用者的存在,只要超过其所引用者的存在不改变其所引用者的基本特征或新颖特征即可,但不包括先前技术实施方案。

[0157] “细胞毒T淋巴细胞相关蛋白4(CTLA-4)多肽”意为一种蛋白质,其与NCBI登录号EAW70354.1或其片段具有至少约85%的序列一致性。示例性氨基酸序列提供如下:

[0158] >EAW70354.1细胞毒T淋巴细胞相关蛋白4[智人(Homo sapiens)]

[0159] MACLGFQRHKAQLNLATRTWPCTLLFFLLFIPVFCKAMHVAQPAVVLASSRGIASFVCEYASPGKATEV
RVTVLRQADSQVTEVCAATYMMGNELTFLDDSICTGTSSGNQVNLTIQGLRAMDTGLYICKVELMYPPPYYLGIGNG
TQIYVIDPEPCPDSDFLLWILAAVSSGLFFYSFLLTAVSLSKMLKKRSPLTTGVYVKMPPTEPECEKQFQPYFIPIN

[0160] “细胞毒T淋巴细胞相关蛋白4(CTLA-4)多核苷酸”意为编码CTLA-4多肽的核酸分子。CTLA-4基因编码免疫球蛋白超家族并且编码将抑制信号传输到T细胞的蛋白质。示例性CTLA-4核酸序列提供如下。

[0161] >BC074842.2智人(Homo sapiens)细胞毒T淋巴细胞相关蛋白4,mRNA(cDNA clone MGC:104099IMAGE:30915552),完全编码序列

[0162] GACCTGAACACCGCTCCCATAAAGCCATGGCTTGCCTTGGATTTCAGCGGCACAAGGCTCAGCTGAACC
TGGCTACCAGGACCTGGCCCTGCACTCTCCTGTTTTTTCTTCTCTTCATCCCTGTCTTCTGCAAAGCAATGCACGTG
GCCCAGCCTGCTGTGGTACTGGCCAGCAGCCGAGGCATCGCCAGCTTTGTGTGTGAGTATGCATCTCCAGGCAAAGC
CACTGAGGTCCGGGTGACAGTGCTTCGGCAGGCTGACAGCCAGGTGACTGAAGTCTGTGCGGCAACCTACATGATGG
GGAATGAGTTGACCTTCCTAGATGATTCCATCTGCACGGGCACCTCCAGTGGAAATCAAGTGAACCTCACTATCCAA
GGACTGAGGGCCATGGACACGGGACTCTACATCTGCAAGGTGGAGCTCATGTACCCACCGCCATACTACCTGGGCAT
AGGCAACGGAACCCAGATTTATGTAATTGATCCAGAACCGTGCCCAGATTCTGACTTCCTCCTCTGGATCCTTGCAG
CAGTTAGTTCGGGGTTGTTTTTTTATAGCTTTCTCCTCACAGCTGTTTCTTTGAGCAAAATGCTAAAGAAAAGAAGC
CCTCTTACAACAGGGGTCTATGTGAAAATGCCCCCAACAGAGCCAGAATGTGAAAAGCAATTTCAGCCTTATTTTAT
TCCCATCAATTGAGAAACCATTATGAAGAAGAGAGTCCATATTTCAATTTCCAAGAGCTGAGG

[0163] “分化簇2(CD2)多肽”意为一种蛋白质,其与NCBI登录号NP_001758.2或其片段具有至少约85%的氨基酸序列一致性并且具有免疫调节活性。示例性氨基酸序列提供如下。

[0164] >NP_001758.2T细胞表面抗原CD2亚型2前体[智人(Homo sapiens)]

[0165]
```
  1 MSFPCKFVAS FLLIFNVSSK GAVSKEITNA LETWGALGQD INLDIPSFQM SDDIDDIKWE
 61 KTSDKKKIAQ FRKEKETFKE KDTYKLFKNG TLKIKHLKTD DQDIYKVSIY DTKGKNVLEK
121 IFDLKIQERV SKPKISWTCI NTTLTCEVMN GTDPELNLYQ DGKHLKLSQR VITHKWTTSL
181 SAKFKCTAGN KVSKESSVEP VSCPEKGLDI YLIIGICGGG SLLMVFVALL VFYITKRKKQ
241 RSRRNDEELE TRAHRVATEE RGRKPHQIPA STPQNPATSQ HPPPPPGHRS QAPSHRPPPP
301  GHRVQHQPQK RPPAPSGTQV HQQKGPPLPR PRVQPKPPHG AAENSLSPSS N
```

[0166] CD2细胞质结构域(氨基酸残基235至351)以粗体显示。来自智人(Homo Sapiens)的示例性CD2多肽的架构显示在图4中。

[0167] “分化簇2(CD2)多核苷酸”意为编码CD2多肽的核酸。示例性CD2核酸序列提供如下。>NM_001767.5智人(Homo sapiens)CD2分子(CD2),转录物变体2,RNA

[0168]
```
   1 agtctcactt cagttccttt tgcatgaaga gctcagaatc aaaagaggaa accaacccct
  61 aagatgagct ttccatgtaa atttgtagcc agcttccttc tgattttcaa tgtttcttcc
 121 aaaggtgcag tctccaaaga gattacgaat gccttggaaa cctggggtgc cttgggtcag
 181 gacatcaact tggacattcc tagttttcaa atgagtgatg atattgacga tataaaatgg
 241 gaaaaaactt cagacaagaa aaagattgca caattcagaa aagagaaaga gactttcaag
 301 gaaaaagata catataagct atttaaaaat ggaactctga aaattaagca tctgaagacc
 361 gatgatcagg atatctacaa ggtatcaata tatgatacaa aaggaaaaaa tgtgttggaa
 421 aaaatatttg atttgaagat tcaagagagg gtctcaaaac caaagatctc ctggacttgt
 481 atcaacacaa ccctgacctg tgaggtaatg aatggaactg accccgaatt aaacctgtat
 541 caagatggga aacatctaaa actttctcag agggtcatca cacacaagtg gaccaccagc
 601 ctgagtgcaa aattcaagtg cacagcaggg aacaaagtca gcaaggaatc cagtgtcgag
 661 cctgtcagct gtccagagaa aggtctggac atctatctca tcattggcat atgtggagga
 721 ggcagcctct tgatggtctt tgtggcactg ctcgttttct atatcaccaa aaggaaaaaa
 781 cagaggagtc ggagaaatga tgaggagctg gagacaagag cccacagagt agctactgaa
 841 gaaaggggcc ggaagcccca ccaaattcca gcttcaaccc ctcagaatcc agcaacttcc
 901 caacatcctc ctccaccacc tggtcatcgt tcccaggcac ctagtcatcg tcccccgcct
 961 cctggacacc gtgttcagca ccagcctcag aagaggcctc ctgctccgtc gggcacacaa
1021 gttcaccagc agaaaggccc gcccctcccc agacctcgag ttcagccaaa acctccccat
1081 ggggcagcag aaaactcatt gtccccttcc tctaattaaa aaagatagaa actgtctttt
1141 tcaataaaaa gcactgtgga tttctgccct cctgatgtgc atatccgtac ttccatgagg
1201 tgttttctgt gtgcagaaca ttgtcacctc ctgaggctgt gggccacagc cacctctgca
1261 tcttcgaact cagccatgtg gtcaacatct ggagtttttg gtctcctcag agagctccat
1321 cacaccagta aggagaagca atataagtgt gattgcaaga atggtagagg accgagcaca
1381 gaaatcttag agatttcttg tcccctctca ggtcatgtgt agatgcgata aatcaagtga
1441 ttggtgtgcc tgggtctcac tacaagcagc ctatctgctt aagagactct ggagtttctt
1501 atgtgccctg gtggacactt gcccaccatc ctgtgagtaa aagtgaaata aaagctttga
1561 ctaga
```

[0169] “分化簇5(CD5)多肽”意为一种蛋白质,其与NCBI登录号NP_001333385.1或其片段具有至少约85％的氨基酸序列一致性并且具有免疫调节活性。示例性氨基酸序列提供如下。

[0170] >NP_001333385.1T细胞表面糖蛋白CD5亚型2[智人(Homo sapiens)]

[0171] MVCSQSWGRSSKQWEDPSQASKVCQRLNCGVPLSLGPFLVTYTPQSSIICYGQLGSFSNCSHSRNDMCHSLGLTCLEPQKTTPPTTRPPPTTTPEPTAPPRLQLVAQSGGQHCAGVVEFYSGSLGGTISYEAQDKTQDLENFLCNNLQCGSFLKHLPETEAGRAQDPGEPREHQPLPIQWKIQNSSCTSLEHCFRKIKPQKSGRVLALLCSGFQPKVQSRLVGGSSICEGTVEVRQGAQWAALCDSSSARSSLRWEEVCREQQCGSVNSYRVLDAGDPTSRGLFCPHQKLSQCHELWERNSYCKKVFVTCQDPNPAGLAAGTVASIILALVLLVVLLVVCGPLAYKKLVKKFRQKKQRQWIGPTGMNQNMSFHRNHTATVRSHAENPTASHVDNEYSQPPRNSHLSAYPALEGALHRSSMQPDNSSDSDYDLHGAQRL

[0172] “分化簇5(CD5)多核苷酸”意为编码CD5多肽的核酸。示例性CD5核酸序列提供如下。

[0173] >NM_001346456.1智人(Homo sapiens)CD5分子(CD5),转录物变体2,mRNA

```
   1 gagtcttgct gatgctcccg gctgaataaa ccccttcctt ctttaacttg gtgtctgagg
  61 ggttttgtct gtggcttgtc ctgctacatt tcttggttcc ctgaccagga agcaaagtga
 121 ttaacggaca gttgaggcag ccccttaggc agcttaggcc tgccttgtgg agcatccccg
 181 cggggaactc tggccagctt gagcgacacg gatcctcaga gcgctcccag gtaggcaatt
 241 gccccagtgg aatgcctcgt cagagcagtg catggcaggc ccctgtggag gatcaacgca
 301 gtggctgaac acagggaagg aactggcact tggagtccgg acaactgaaa cttgtcgctt
 361 cctgcctcgg acggctcagc tggtatgacc cagatttcca ggcaaggctc acccgttcca
 421 actcgaagtg ccagggccag ctggaggtct acctcaagga cggatggcac atggtttgca
 481 gccagagctg gggccggagc tccaagcagt gggaggaccc cagtcaagcg tcaaaagtct
 541 gccagcggct gaactgtggg gtgcccttaa gccttggccc cttccttgtc acctacacac
 601 ctcagagctc aatcatctgc tacggacaac tgggctcctt ctccaactgc agccacagca
 661 gaaatgacat gtgtcactct ctgggcctga cctgcttaga accccagaag acaacacctc
 721 caacgacaag gccccgccc accacaactc cagagcccac agctcctccc aggctgcagc
 781 tggtggcaca gtctggcggc cagcactgtg ccggcgtggt ggagttctac agcggcagcc
 841 tgggggggtac catcagctat gaggcccagg acaagaccca ggacctggag aacttcctct
 901 gcaacaacct ccagtgtggc tccttcttga agcatctgcc agagactgag gcaggcagag
 961 cccaagaccc aggggagcca cgggaacacc agcccttgcc aatccaatgg aagatccaga
1021 actcaagctg tacctccctg gagcattgct tcaggaaaat caagccccag aaaagtggcc
1081 gagttcttgc cctcctttgc tcaggtttcc agcccaaggt gcagagccgt ctggtggggg
1141 gcagcagcat ctgtgaaggc accgtggagg tgcgccaggg ggctcagtgg gcagccctgt
1201 gtgacagctc ttcagccagg agctcgctgc ggtgggagga ggtgtgccgg gagcagcagt
1261 gtggcagcgt caactcctat cgagtgctgg acgctggtga cccaacatcc cgggggctct
1321 tctgtcccca tcagaagctg tcccagtgcc acgaactttg ggagagaaat tcctactgca
1381 agaaggtgtt tgtcacatgc caggatccaa accccgcagg cctggccgca ggcacggtgg
1441 caagcatcat cctggccctg gtgctcctgg tggtgctgct ggtcgtgtgc ggccccttg
1501 cctacaagaa gctagtgaag aaattccgcc agaagaagca gcgccagtgg attggcccaa
1561 cgggaatgaa ccaaaacatg tctttccatc gcaaccacac ggcaaccgtc cgatcccatg
1621 ctgagaaccc cacagcctcc cacgtggata acgaatacag ccaacctccc aggaactccc
1681 acctgtcagc ttatccagct ctggaagggg ctctgcatcg ctcctccatg cagcctgaca
1741 actcctccga cagtgactat gatctgcatg gggctcagag gctgtaaaga actgggatcc
1801 atgagcaaaa agccgagagc cagacctgtt tgtcctgaga aaactgtccg ctcttcactt
1861 gaaatcatgt ccctatttct accccggcca gaacatggac agaggccaga agccttccgg
1921 acaggcgctg ctgccccgag tggcaggcca gctcacactc tgctgcacaa cagctcggcc
1981 gcccctccac ttgtggaagc tgtggtgggc agagccccaa aacaagcagc cttccaacta
2041 gagactcggg ggtgtctgaa gggggccccc tttccctgcc cgctggggag cggcgtctca
2101 gtgaaatcgg ctttctcctc agactctgtc cctggtaagg agtgacaagg aagctcacag
2161 ctgggcgagt gcattttgaa tagttttttg taagtagtgc ttttcctcct tcctgacaaa

2221 tcgagcgctt tggcctcttc tgtgcagcat ccacccctgc ggatccctct ggggaggaca
2281 ggaaggggac tcccggagac ctctgcagcc gtggtggtca gaggctgctc acctgagcac
2341 aaagacagct ctgcacattc accgcagctg ccagccaggg gtctgggtgg gcaccaccct
2401 gacccacagc gtcaccccac tccctctgtc ttatgactcc cctccccaac cccctcatct
2461 aaagacacct tcctttccac tggctgtcaa gcccacaggg caccagtgcc acccagggcc
2521 cggcacaaag gggcgcctag taaaccttaa ccaacttggt tttttgcttc acccagcaat
2581 taaaagtccc aagctgaggt agtttcagtc catcacagtt catcttctaa cccaagagtc
2641 agagatgggg ctggtcatgt tcctttggtt tgaataactc ccttgacgaa aacagactcc
2701 tctagtactt ggagatcttg gacgtacacc taatcccatg gggcctcggc ttccttaact
2761 gcaagtgaga agaggaggtc tacccaggag cctcgggtct gatcaaggga gaggccaggc
2821 gcagctcact gcggcggctc cctaagaagg tgaagcaaca tgggaacaca tcctaagaca
2881 ggtcctttct ccacgccatt tgatgctgta tctcctggga gcacaggcat caatggtcca
2941 agccgcataa taagtctgga agagcaaaag ggagttacta ggatatgggg tgggctgctc
3001 ccagaatctg ctcagctttc tgcccccacc aacaccctcc aaccaggcct tgccttctga
3061 gagcccccgt ggccaagccc aggtcacaga tcttcccccg accatgctgg gaatccagaa
3121 acagggaccc catttgtctt cccatatctg gtggaggtga gggggctcct caaaagggaa
3181 ctgagaggct gctcttaggg agggcaaagg ttcgggggca gccagtgtct cccatcagtg
3241 ccttttttaa taaaagctct ttcatctata gtttggccac catacagtgg cctcaaagca
3301 accatggcct acttaaaaac caaaccaaaa ataaagagtt tagttgagga gaaaaaaaaa
3361 aaaaaaaaaa aaaaaa
```

[0174]

[0175]

[0176] “分化簇7(CD7)多肽”意为一种蛋白质,其与NCBI参考序列NP_006128.1或其片段

具有至少约85％的氨基酸序列一致性并且具有免疫调节活性。示例性氨基酸序列提供如下。

[0177] >NP_006128.1T细胞抗原前体[智人(Homo sapiens)]

[0178]
```
  1 MAGPPRLLLL PLLLALARGL PGALAAQEVQ QSPHCTTVPV GASVNITCST SGGLRGIYLR
 61 QLGPQPQDII YYEDGVVPTT DRRFRGRIDF SGSQDNLTIT MHRLQLSDTG TYTCQAITEV
121 NVYGSGTLVL VTEEQSQGWH RCSDAPPRAS ALPAPPTGSA LPDPQTASAL PDPPAASALP
181 AALAVISFLL GLGLGVACVL ARTQIKKLCS WRDKNSAACV VYEDMSHSRC NTLSSPNQYQ
```

[0179] “分化簇7(CD7)多核苷酸”意为编码CD7多肽的核酸分子。示例性CD7核酸序列提供如下。

[0180] >NM_006137.7智人(Homo sapiens)CD7分子(CD7),mRNA

[0181]
```
   1 ctctctgagc tctgagcgcc tgcggtctcc tgtgtgctgc tctctgtggg gtcctgtaga
  61 cccagagagg ctcagctgca ctcgcccggc tgggagagct gggtgtgggg aacatggccg
 121 ggcctccgag gctcctgctg ctgcccctgc ttctggcgct ggctcgcggc ctgcctgggg
 181 ccctggctgc ccaagaggtg cagcagtctc cccactgcac gactgtcccc gtgggagcct
 241 ccgtcaacat cacctgctcc accagcgggg gcctgcgtgg gatctacctg aggcagctcg
 301 ggccacagcc ccaagacatc atttactacg aggacggggt ggtgcccact acggacagac
 361 ggttccgggg ccgcatcgac ttctcagggt cccaggacaa cctgactatc accatgcacc
 421 gcctgcagct gtcggacact ggcacctaca cctgccaggc catcacggag gtcaatgtct
 481 acggctccgg caccctggtc ctggtgacag aggaacagtc ccaaggatgg cacagatgct
 541 cggacgcccc accaagggcc tctgccctcc ctgccccacc gacaggctcc gccctccctg
 601 acccgcagac agcctctgcc ctccctgacc cgccagcagc ctctgccctc cctgcggccc
 661 tggcggtgat ctccttcctc ctcgggctgg gcctgggggt ggcgtgtgtg ctggcgagga
 721 cacagataaa gaaactgtgc tcgtggcggg ataagaattc ggcggcatgt gtggtgtacg
 781 aggacatgtc gcacagccgc tgcaacacgc tgtcctcccc caaccagtac cagtgaccca
 841 gtgggccccc gcacgtcccg cctgtggtcc ccccagcacc ttccctgccc caccatgccc
 901 cccaccctgc cacacccctc accctgctgt cctcccacgg ctgcagcaga gtttgaaggg
 961 cccagccgtg cccagctcca agcagacaca caggcagtgg ccaggcccca cggtgcttct
1021 cagtggacaa tgatgcctcc tccgggaagc cttccctgcc cagcccacgc cgccaccggg
1081 aggaagcctg actgtccttt ggctgcatct cccgaccatg gccaaggagg gcttttctgt
1141 gggatgggcc tgggcacgcg gccctctcct gtcagtgccg gcccacccac cagcaggccc
1201 ccaaccccca ggcagcccgg cagaggacgg gaggagacca gtcccccacc cagccgtacc
1261 agaaataaag gcttctgtgc ttcc
```

[0182] “分化簇33(CD33)多肽”意为一种蛋白质,其与NCBI参考序列NP_001763.3或其片段具有至少约85％的氨基酸序列一致性。CD33也称为Siglec-3。示例性氨基酸序列提供如下。

[0183] >NP_001763.3髓样细胞表面抗原CD33亚型1前体[智人(Homo sapiens)]

[0184]
```
  1 MPLLLLLPLL WAGALAMDPN FWLQVQESVT VQEGLCVLVP CTFFHPIPYY DKNSPVHGYW
 61 FREGAIISRD SPVATNKLDQ EVQEETQGRF RLLGDPSRNN CSLSIVDARR RDNGSYFFRM
121 ERGSTKYSYK SPQLSVHVTD LTHRPKILIP GTLEPGHSKN LTCSVSWACE QGTPPIFSWL
181 SAAPTSLGPR TTHSSVLIIT PRPQDHGTNL TCQVKFAGAG VTTERTIQLN VTYVPQNPTT
241 GIFPGDGSGK QETRAGVVHG AIGGAGVTAL LALCLCLIFF IVKTHRRKAA RTAVGRNDTH
301 PTTGSASPKH QKKSKLHGPT ETSSCSGAAP TVEMDEELHY ASLNFHGMNP SKDTSTEYSE
361 VRTQ
```

[0185] “分化簇33(CD33)多核苷酸”意为编码CD33多肽的核酸分子。示例性CD33核酸序列提供如下。

[0186] >NM_001772.4智人(Homo sapiens)CD33分子(CD33),转录物变体1,mRNA

[0187]
```
   1 ctgctcacac aggaagccct ggaagctgct tcctcagaca tgccgctgct gctactgctg
  61 cccctgctgt gggcaggggc cctggctatg gatccaaatt tctggctgca agtgcaggag
 121 tcagtgacgg tacaggaggg tttgtgcgtc ctcgtgccct gcactttctt ccatcccata
 181 ccctactacg acaagaactc cccagttcat ggttactggt tccgggaagg agccattata
 241 tccagggact ctccagtggc cacaaacaag ctagatcaag aagtacagga ggagactcag
 301 ggcagattcc gcctccttgg ggatcccagt aggaacaact gctccctgag catcgtagac
 361 gccaggagga gggataatgg ttcatacttc tttcggatgg agagaggaag taccaaatac
 421 agttacaaat ctccccagct ctctgtgcat gtgacagact tgacccacag gcccaaaatc
 481 ctcatccctg gcactctaga acccggccac tccaaaaacc tgacctgctc tgtgtcctgg
 541 gcctgtgagc agggaacacc cccgatcttc tcctggttgt cagctgcccc cacctccctg
 601 ggccccagga ctactcactc ctcggtgctc ataatcaccc cacggcccca ggaccacggc
 661 accaacctga cctgtcaggt gaagttcgct ggagctggtg tgactacgga gagaaccatc
 721 cagctcaacg tcacctatgt tccacagaac ccaacaactg gtatctttcc aggagatggc
 781 tcagggaaac aagagaccag agcaggagtg gttcatgggg ccattggagg agctggtgtt
 841 acagccctgc tcgctctttg tctctgcctc atcttcttca tagtgaagac ccacaggagg
 901 aaagcagcca ggacagcagt gggcaggaat gacacccacc ctaccacagg gtcagcctcc
 961 ccgaaacacc agaagaagtc caagttacat ggccccactg aaacctcaag ctgttcaggt
1021 gccgccccta ctgtggagat ggatgaggag ctgcattatg cttccctcaa ctttcatggg
1081 atgaatcctt ccaaggacac ctccaccgaa tactcagagg tcaggaccca gtgaggaacc
1141 cacaagagca tcaggctcag ctagaagatc cacatcctct acaggtcggg gaccaaaggc
1201 tgattcttgg agatttaaca ccccacaggc aatgggttta tagacattat gtgagtttcc
1261 tgctatatta acatcatctt agactttgca agcagagagt cgtggaatca aatctgtgct
1321 ctttcatttg ctaagtgtat gatgtcacac aagctcctta accttccatg tctccatttt
1381 cttctctgtg aagtaggtat aagaagtcct atctcatagg gatgctgtga gcattaaata
1441 aaggtacaca tggaaaacac ca
```

[0188] “分化簇52(CD52)多肽”意为一种蛋白质,其与NCBI参考序列NP_001794.2或其片段具有至少约85％的氨基酸序列一致性。CD52也称为CAMPATH-1。示例性氨基酸序列提供如下。

[0189] >NP_001794.2CAMPATH-1抗原前体[智人(Homo sapiens)]

[0190] 1 MKRFLFLLLTISLLVMVQIQTGLSGQNDTSQTSSPSASSNISGGIFLFFVANAIIHLFCF

[0191] 61 S

[0192] “分化簇52(CD52)多核苷酸”意为编码CD52多肽的核酸分子。示例性CD52核酸序列提供如下。

[0193] >NM_001803.3智人(Homosapiens)CD52分子(CD52),mRNA

[0194]
```
   1 agacagccct gagatcacct aaaaagctgc taccaagaca gccacgaaga tcctaccaaa
  61 atgaagcgct tcctcttcct cctactcacc atcagcctcc tggttatggt acagatacaa
 121 actggactct caggacaaaa cgacaccagc caaaccagca gcccctcagc atccagcaac
 181 ataagcggag gcatttttcct tttcttcgtg gccaatgcca taatccacct cttctgcttc
 241 agttgaggtg acacgtctca gccttagccc tgtgcccccct gaaacagctg ccaccatcac
 301 tcgcaagaga atcccctcca tctttgggag gggttgatgc cagacatcac caggttgtag
 361 aagttgacag gcagtgccat gggggcaaca gccaaaatag gggggtaatg atgtaggggc
 421 caagcagtgc ccagctgggg gtcaataaag ttacccttgt acttgca
```

[0195] “分化簇123(CD123)多肽”意为一种蛋白质,其与NCBI参考序列NP_002174.1或其片段具有至少约85％的氨基酸序列一致性。CD123也称为白介素-3受体。示例性氨基酸序列提供如下。

[0196] >NP_002174.1白介素-3受体亚基α亚型1前体[智人(Homo sapiens)]

[0197]
```
  1 MVLLWLTLLL IALPCLLQTK EDPNPPITNL RMKAKAQQLT WDLNRNVTDI ECVKDADYSM
 61 PAVNNSYCQF GAISLCEVTN YTVRVANPPF STWILFPENS GKPWAGAENL TCWIHDVDFL
121 SCSWAVGPGA PADVQYDLYL NVANRRQQYE CLHYKTDAQG TRIGCRFDDI SRLSSGSQSS
181 HILVRGRSAA FGIPCTDKFV VFSQIEILTP PNMTAKCNKT HSFMHWKMRS HFNRKFRYEL
241 QIQKRMQPVI TEQVRDRTSF QLLNPGTYTV QIRARERVYE FLSAWSTPQR FECDQEEGAN
301 TRAWRTSLLI ALGTLLALVC VFVICRRYLV MQRLFPRIPH MKDPIGDSFQ NDKLVVWEAG
361 KAGLEECLVT EVQVVQKT
```

[0198] “分化簇123(CD123)多核苷酸”意为编码CD123多肽的核酸分子。示例性CD123核酸序列提供如下。

[0199] >NM_002183.4智人(Homosapiens)白介素3受体亚基α(IL3RA),转录物变体1,mRNA

[0200]
```
   1 cttcggtttc tcttcgggga aagctgcttt cagcgcacac gggaagatat cagaaacatc
  61 ctaggatcag gacaccccag atcttctcaa ctggaaccac gaaggctgtt tcttccacac
 121 agtactttga tctccattta agcaggcacc tctgtcctgc gttccggagc tgcgttcccg
 181 atggtcctcc tttggctcac gctgctcctg atcgccctgc cctgtctcct gcaaacgaag
 241 gaagatccaa acccaccaat cacgaaccta aggatgaaag caaaggctca gcagttgacc
 301 tgggacctta acagaaatgt gaccgatatc gagtgtgtta aagacgccga ctattctatg
 361 ccggcagtga acaatagcta ttgccagttt ggagcaattt ccttatgtga agtgaccaac
 421 tacaccgtcc gagtggccaa cccaccattc tccacgtgga tcctcttccc tgagaacagt
 481 gggaagcctt gggcaggtgc ggagaatctg acctgctgga ttcatgacgt ggatttcttg
 541 agctgcagct gggcggtagg cccgggggcc cccgcggacg tccagtacga cctgtacttg
 601 aacgttgcca acaggcgtca acagtacgag tgtcttcact acaaaacgga tgctcaggga
 661 acacgtatcg ggtgtcgttt cgatgacatc tctcgactct ccagcggttc tcaaagttcc
 721 cacatcctgg tgcggggcag gagcgcagcc ttcggtatcc cctgcacaga taagtttgtc
 781 gtcttttcac agattgagat attaactcca cccaacatga ctgcaaagtg taataagaca
 841 cattccttta tgcactggaa aatgagaagt catttcaatc gcaaatttcg ctatgagctt
 901 cagatacaaa agagaatgca gcctgtaatc acagaacagg tcagagacag aacctccttc
 961 cagctactca atcctggaac gtacacagta caaataagag cccgggaaag agtgtatgaa
1021 ttcttgagcg cctggagcac cccccagcgc ttcgagtgcg accaggagga gggcgcaaac
1081 acacgtgcct ggcggacgtc gctgctgatc gcgctgggga cgctgctggc cctggtctgt
1141 gtcttcgtga tctgcagaag gtatctggtg atgcagagac tctttccccg catccctcac
1201 atgaaagacc ccatcggtga cagcttccaa aacgacaagc tggtggtctg ggaggcgggc
1261 aaagccggcc tggaggagtg tctggtgact gaagtacagg tcgtgcagaa aacttgagac
1321 tggggttcag ggcttgtggg ggtctgcctc aatctccctg gccgggccag gcgcctgcac
1381 agactggctg ctggacctgc gcacgcagcc caggaatgga cattcctaac gggtggtggg
1441 catgggagat gcctgtgtaa tttcgtccga agctgccagg aagaagaaca gaactttgtg
1501 tgtttatttc atgataaagt gatttttttt ttttttaaccc a
```

[0201] “分化簇137(CD137)多肽”意为一种蛋白质,其与NCBI参考序列NP_001552.2或其片段具有至少约85%的氨基酸序列一致性。Cd137也称为4-1BB。示例性氨基酸序列提供如下。

[0202] >NP_001552.2肿瘤坏死因子受体超家族成员9前体[智人(Homo sapiens)]

[0203]
```
  1 MGNSCYNIVA TLLLVLNFER TRSLQDPCSN CPAGTFCDNN RNQICSPCPP NSFSSAGGQR
 61 TCDICRQCKG VFRTRKECSS TSNAECDCTP GFHCLGAGCS MCEQDCKQGQ ELTKKGCKDC
121 CFGTFNDQKR GICRPWTNCS LDGKSVLVNG TKERDVVCGP SPADLSPGAS SVTPPAPARE
181 PGHSPQIISF FLALTSTALL FLLFFLTLRF SVVKRGRKKL LYIFKQPFMR PVQTTQEEDG
241 CSCRFPEEEE GGCEL
```

[0204] “分化簇137(CD137)多核苷酸”意为编码CD137多肽的核酸分子。示例性CD137核酸序列提供如下。

[0205] >NM_001561.6智人(Homo sapiens)TNF受体超家族成员9(TNFRSF9),mRNA

[0206]

```
   1 gcagaagcct gaagaccaag gagtggaaag ttctccggca gccctgagat ctcaagagtg
  61 acatttgtga gaccagctaa tttgattaaa attctcttgg aatcagcttt gctagtatca
 121 tacctgtgcc agatttcatc atgggaaaca gctgttacaa catagtagcc actctgttgc
 181 tggtcctcaa ctttgagagg acaagatcat tgcaggatcc ttgtagtaac tgcccagctg
 241 gtacattctg tgataataac aggaatcaga tttgcagtcc ctgtcctcca aatagtttct
 301 ccagcgcagg tggacaaagg acctgtgaca tatgcaggca gtgtaaaggt gttttcagga
 361 ccaggaagga gtgttcctcc accagcaatg cagagtgtga ctgcactcca gggtttcact
 421 gcctgggggc aggatgcagc atgtgtgaac aggattgtaa acaaggtcaa gaactgacaa
 481 aaaaaggttg taaagactgt tgctttggga catttaacga tcagaaacgt ggcatctgtc
 541 gaccctggac aaactgttct ttggatggaa agtctgtgct tgtgaatggg acgaaggaga
 601 gggacgtggt ctgtggacca tctccagccg acctctctcc gggagcatcc tctgtgaccc
 661 cgcctgcccc tgcgagagag ccaggacact ctccgcagat catctccttc tttcttgcgc
 721 tgacgtcgac tgcgttgctc ttcctgctgt tcttcctcac gctccgtttc tctgttgtta
 781 aacggggcag aaagaaactc ctgtatatat tcaaacaacc atttatgaga ccagtacaaa
 841 ctactcaaga ggaagatggc tgtagctgcc gatttccaga agaagaagaa ggaggatgtg
 901 aactgtgaaa tggaagtcaa tagggctgtt gggactttct tgaaaagaag caaggaaata
 961 tgagtcatcc gctatcacag ctttcaaaag caagaacacc atcctacata atacccagga
1021 ttcccccaac acacgttctt ttctaaatgc caatgagttg gcctttaaaa atgcaccact
1081 ttttttttct ttttgacagg gtctcactct gtcacccagg ctggagtgca gtggcaccac
1141 catggctctc tgcagccttg acctctggga gctcaagtga tcctcctgcc tcagtctcct
1201 gagtagctgg aactacaagg aagggccacc acacctgact aacttttttg ttttttgttt
1261 ggtaaagatg gcatttcacc atgttgtaca ggctggtctc aaactcctag gttcactttg
1321 gcctcccaaa gtgctgggat tacagacatg aactgccagg cccggccaaa ataatgcacc
1381 acttttaaca gaacagacag atgaggacag agctggtgat aaaaaaaaaa aaaaaaaagc
1441 attttctaga taccacttaa caggtttgag ctagtttttt tgaaatccaa agaaaattat
1501 agtttaaatt caattacata gtccagtggt ccaactataa ttataatcaa aatcaatgca
1561 ggtttgtttt ttggtgctaa tatgacatat gacaataagc cacgaggtgc agtaagtacc
1621 cgactaaagt ttccgtgggt tctgtcatgt aacacgacat gctccaccgt caggggggag
1681 tatgagcaga gtgcctgagt ttagggtcaa ggacaaaaaa cctcaggcct ggaggaagtt
1741 ttggaaagag ttcaagtgtc tgtatatcct atggtcttct ccatcctcac accttctgcc
1801 tttgtcctgc tccctttaa gccaggttac attctaaaaa ttcttaactt ttaacataat
1861 attttatacc aaagccaata aatgaactgc atatgatagg tatgaagtac agtgagaaaa
1921 ttaacacctg tgagctcatt gtcctaccac agcactagag tgggggccgc caaactccca
1981 tggccaaacc tggtgcacca tttgcctttg tttgtctgtt ggtttgcttg agacagtctt
2041 gctctgttgc ccaggctgga atggagtggc tattcacagg cacaatcata gcacacttta
2101 gccttaaact cctgggctca agtgatccac ccgcctcagt ctcccaagta gctgggatta
2161 caggtgcaaa cctggcatgc ctgccattgt ttggcttatg atctaaggat agctttttaa
2221 attttattca ttttattttt ttttgagaca gtgtctcact ctgtctccca ggctggagta
2281 cagtggtaca atcttggatc accgcctccc agtttcaagt gatctccctg cctcagcctc
```

2341 ctaagtagct gggactacag gtatgtgcca ccacgcctgg ctaattttta tatttttagt
2401 agagacgggg tttcaccatg ttgtccaggc tggtctcaaa ctcctgacct caggtgatct
2461 gcccacctct gcctcccaaa gtgctgggat tacaggcatg agccaccatg cctggccatt
2521 tcttacactt ttgtatgaca tgcctattgc aagcttgcgt gcctctgtcc catgttattt
2581 tactctggga tttaggtgga gggagcagct tctatttgga acattggcca tcgcatggca
2641 aatgggtatc tgtcacttct gctcctattt agttggttct actataacct ttagagcaaa
2701 tcctgcagcc aagccaggca tcaatagggc agaaaagtat attctgtaaa taggggtgag
2761 gagaagatat ttctgaacaa tagtctactg cagtaccaaa ttgcttttca aagtggctgt
2821 tctaatgtac tcccgtcagt catataagtg tcatgtaagt atcccattga tccacatcct
2881 tgctaccctc tggtactatc aggtgccctt aattttgcca agccagtggg tatagaatga
2941 gatctcactg tggtcttagt ttgcatttgc ttggttactg atgagcacct tgtcaaatat
3001 ttatatacca tttgtgttta ttttttttaaa taaaatgctt gctcatgctt ttttgcccat
3061 ttgcaaaaaa acttggggcc gggtgcagtg gctcatgcct gtagtcccag ctctttggga
3121 ggccaaggtg ggcagatcgc ttgagcccag gagttcgaga ccagccttgg caacatggcg
3181 aaaccctgtc tttacaaaaa atacaaaaat tagccgggtg tggtggtgtg cacctgaagt
3241 cccagctact cagtaggttc gctttgagcc tgggaggcag aggttgcagt gagctgggac
3301 cgcatcacta cacttcagcc tgggcaacag agaaaaacct tttctcagaa acaaacaaac
3361 ccaaatgtgg ttgtttgtcc tgattcctaa aaggtcttta tgtattctag ataataatct
3421 ttggtcagtt atatgtgtta aaaaatatct tctttgtggc caggcacggt agctcacacc
3481 tgtaatccca gcactttgcg gggctgaggt gggtggatca tctgaggtca agagttcaag
3541 atcagcctgg ccaacacagt gaaaccccat ctctactaaa catgtacaaa acttagctgg
3601 gtatggtggc gggtgcctgt aaccccagct gctccagagg ctgtggcaga agaatcgctt
3661 gaacccagga ggcagaggtt gcagcgagcc aagattgtgc cattgcactc cagactgggt
3721 gacaagagtg aaattctgcc tatctatcta tctatctatc tatatctata tatatatata
3781 tatatatcct ttgtaattta tttttccctt tttaaaattt tttataaaat tctttttat
3841 ttttattttt agcagaggtg aggtttctga ggtttcatta tgttgcccag gctggtcttg
3901 aactcctgag ctcaagtgat cctcccacct cagccttcca aagtgctgga attgcagaca
3961 tgagccaccg cgcccctcct gtttttctct aattaatggt gtctttcttt gtctttctgg
4021 taataagcaa aaagttcttc atttgatttg gttaaattta taactgtttt ctcatatggt
[0207] 4081 taacattttt tcttgcctgg ctaaagaaat ccttttctgc ccaatactat aaagaggttt
4141 gcccacattt tattccaaaa gttttaagtt ttgtctttca tcttgaagtc taatgtatca
4201 ggaactggct tttgtgcctg ttgggaggta gtgatccaat tccatgtctt gcatgtaggt
4261 aaccactggt ccctgcgcca tgtattcaat acgtcgtctt tctcctgcgg gtctgcaatc
4321 tcacctacca tccatcaagt ttccataggg ccatgggtct gcttctgggc tccctgttct
4381 gttccattgt caatttgtct atcctgtgcc agtatcacac tgtgtttatt acaatagctt
4441 tgtaacagct ctcgatatcc ggtaggacat ctccctccac cttctttttc tacttcagaa
4501 gtgtcttagc taggtcaggc acggtggctc acgcctgtaa tcccagcact ttgggaggcc
4561 gacgcggatg gatcacctga ggtcaggagt tttgagacag cctggccaac atggtgaaac
4621 cccatctcta ctaaaaaata caaaaattag tcaggcatgg tggcatgtgc ctgtaatccc
4681 agctatttgg gaggctgagg ccggagaatt gcttgaaccc gggggggcgga ggttgcagtg
4741 agccgagatc gtaccattgc actccagcct gggtgacaga gcgaaactct gtctcaggaa
4801 aaaaaagaaa agagatgtct tggttattct tggttcttta ttattcaata taaattttag
4861 aagctgaatt tgaaaagatt tggattggaa tttcattaaa tctacaggtc aatttaggga
4921 gagttgataa ttttacagaa ttgagtcatc tggtgttcca ataagaataa gagaacaatt
4981 attggctgta caattcttgc caaatagtag gcaaagcaaa gcttaggaag tatactggtg
5041 ccatttcagg aacaaagcta ggtgcgaata tttttgtctt tctgaatcat gatgctgtaa
5101 gttctaaagt gatttctcct cttggctttg gacacatggt gtttaattac ctactgctga
5161 ctatccacaa acagaaagag actggtcatg ccccacaggg ttggggtatc caagataatg
5221 gagcgaggct ctcatgtgtc ctaggttaca caccgaaaat ccacagttta ttctgtgaag
5281 aaaggaggct atgtttatga tacagactgt gatattttta tcatagccta ttctggtatc
5341 atgtgcaaaa gctataaatg aaaaacacag gaacttggca tgtgagtcat tgctcccccc
5401 aaatgacaat taataaggaa ggaacattga gacagaataa aatgatcccc ttctgggttt
5461 aatttagaaa gttccataat taggtttaat agaaataaat gtaaatttct atgattaaaa
5521 ataaattagc acatttaggg atacacaaat tataaatcat tttctaaatg ctaaaaacaa
5581 gctcaggttt ttttcagaag aaagttttaa tttttttttct ttagtggaag atatcactct
5641 gacggaaagt tttgatgtga ggggcggatg actataaagt gggcatcttc ccccacagga
5701 agatgtttcc atctgtgggt gagaggtgcc caccgcagct agggcaggtt acatgtgccc
5761 tgtgtgtggt aggacttgga gagtgatctt tatcaacgtt tttatttaaa agactatcta
5821 ataaaacaca aaactatgat gttcacagga aaaaaagaat aagaaaaaaa ga

[0208] “分化簇247(CD247)多肽”意为一种蛋白质,其与NCBI参考序列NP_932170.1或其片段具有至少约85%的氨基酸序列一致性。CD137也称为CD3ζ。示例性氨基酸序列提供如

下。

[0209] >NP_932170.1T细胞表面糖蛋白CD3ζ链亚型1前体[智人(Homo sapiens)]

[0210]
```
  1 mkwkalftaa ilqaqlpite aqsfglldpk lcylldgilf iygviltalf lrvkfsrsad
 61 apayqqgqnq lynelnlgrr eeydvldkrr grdpemggkp qrrknpqegl ynelqkdkma
121 eayseigmkg errrgkghdg lyqglstatk dtydalhmqa lppr
```

[0211] “分化簇247(CD247)多核苷酸”意为编码CD247多肽的核酸分子。示例性CD247核酸序列提供如下。

[0212] >NM_NM_198053.3智人(Homo sapiens)CD247分子(CD247),转录物变体1,mRNA

[0213]
```
   1 aaccgtcccg gccaccgctg cctcagcctc tgcctcccag cctctttctg agggaaagga
  61 caagatgaag tggaaggcgc ttttcaccgc ggccatcctg caggcacagt tgccgattac
 121 agaggcacag agctttggcc tgctggatcc caaactctgc tacctgctgg atggaatcct
 181 cttcatctat ggtgtcattc tcactgcctt gttcctgaga gtgaagttca gcaggagcgc
 241 agacgccccc gcgtaccagc agggccagaa ccagctctat aacgagctca atctaggacg
 301 aagagaggag tacgatgttt tggacaagag acgtggccgg gaccctgaga tgggggaaaa
 361 gccgcagaga aggaagaacc ctcaggaagg cctgtacaat gaactgcaga aagataagat
 421 ggcggaggcc tacagtgaga ttgggatgaa aggcgagcgc cggaggggca aggggcacga
 481 tggcctttac cagggtctca gtacagccac caaggacacc tacgacgccc ttcacatgca
 541 ggccctgccc cctcgctaac agccagggga tttcaccact caaaggccag acctgcagac
 601 gcccagatta tgagacacag gatgaagcat ttacaacccg gttcactctt ctcagccact
 661 gaagtattcc cctttatgta caggatgctt tggttatatt tagctccaaa ccttcacaca
 721 cagactgttg tccctgcact ctttaaggga gtgtactccc agggcttacg gccctggcct
 781 tgggccctct ggtttgccgg tggtgcaggt agacctgtct cctggcggtt cctcgttctc
 841 cctgggaggc gggcgcactg cctctcacag ctgagttgtt gagtctgttt tgtaaagtcc
 901 ccagagaaag cgcagatgct agcacatgcc ctaatgtctg tatcactctg tgtctgagtg
 961 gcttcactcc tgctgtaaat ttggcttctg ttgtcacctt cacctccttt caaggtaact
1021 gtactgggcc atgttgtgcc tccctggtga gagggccggg cagaggggca gatggaaagg
1081 agcctaggcc aggtgcaacc agggagctgc aggggcatgg gaaggtgggc gggcagggga
1141 gggtcagcca gggcctgcga gggcagcggg agcctccctg cctcaggcct ctgtgccgca
1201 ccattgaact gtaccatgtg ctacaggggc cagaagatga acagactgac cttgatgagc
1261 tgtgcacaaa gtggcataaa aaacatgtgg ttacacagtg tgaataaagt gctgcggagc
1321 aagaggaggc cgttgattca cttcacgctt tcagcgaatg acaaaatcat ctttgtgaag
1381 gcctcgcagg aagacccaac acatgggacc tataactgcc cagcggacag tggcaggaca
1441 ggaaaaaccc gtcaatgtac taggatactg ctgcgtcatt acagggcaca ggccatggat
1501 ggaaaacgct ctctactctg ctttttttct actgttttaa tttatactgg catgctaaag
1561 ccttcctatt ttgcataata aatgcttcag tgaaaatgca
```

[0214] “共同给药”或“共给药”是指在疗程期间给药两种或更多种治疗剂或药物组合物。该共同给药可以是同时给药或依次给药。较晚给药的治疗剂或药物组合物的依次给药可能在给药第一种药物组合物或治疗剂后在疗程期间的任何时间发生。

[0215] 术语“保守氨基酸置换”或“保守突变”是指一个氨基酸被具有共有特性的另一氨基酸替换。定义个体氨基酸之间的共有特性的功能性途径是分析同源生物体的对应蛋白质之间的氨基酸变化的标准化频率(Schulz,G.E.and Schirmer,R.H.,Principles of Protein Structure,Springer-Verlag,New York(1979))。根据此类分析,可定义多组氨基酸,其中组内氨基酸优先彼此交换,并因此在它们对整体蛋白质结构的影响方面彼此最为近似(Schulz,G.E.and Schirmer,R.H.,同上)。保守突变的非限制性示例包括氨基酸的氨基酸置换,例如,赖氨酸置换精氨酸,反之亦然,使得正电荷可得以保持;谷氨酸置换天冬氨酸,反之亦然,使得负电荷可得以保持;丝氨酸替换苏氨酸,使得游离-OH可得以保持;以及谷氨酰胺替换天冬酰胺,使得游离-NH_2可得以保持。

[0216] 本文中可互换使用的术语“编码序列”或“蛋白质编码序列”是指编码蛋白质的多核苷酸链段。该区域或序列在靠近5’末端结合有起始密码子且在靠近3’末端结合有终止密

码子。编码序列也可以称为开放阅读框。

[0217] “密码子优化”意为,为了增强在感兴趣的宿主细胞内的表达,通过将天然序列至少一个密码子(例如,约或超过约1、2、3、4、5、10、15、20、25、50或更多密码子)替换为更频繁或最频繁用于宿主细胞的基因中同时维持天然氨基酸序列的密码子来修饰核酸序列的过程。对于特定氨基酸的某些密码子,各种物种均表现出特定偏倚。密码子偏倚(生物体之间密码子使用的差异)经常与信使RNA(mRNA)的翻译效率相关,而信使RNA的翻译效率据信取决于正在被翻译的密码子的性质和特定转移RNA(tRNA)分子的可用性。细胞中选定tRNA的优势通常反映了肽合成中最常用的密码子。据此,可以基于密码子优化在给定生物体内裁剪基因以进行最优基因表达。密码子使用表可轻易获得,例如,在C可在www.kazusa.orjp/codon/(2002年7月9日访问)获得的“密码子使用数据库”处获得,且这些表可以多种方式调适。参见,Nakamura,Y.,等人“Codon usage tabulated from the international DNA sequence databases:status for the year 2000”Nucl.Acids Res.28:292(2000)。用于对特定序列进行密码子优化以在特定宿主细胞中表达的计算机算法也是可获得的,诸如Gene Forge(Aptagen;Jacobus,Pa.)。在一些实施方案中,编码经工程改造的核酸酶的序列中的一个或多个密码子(例如,1、2、3、4、5、10、15、20、25、50、或更多、或全部密码子)对应于特定氨基酸最频繁使用的密码子。

[0218] “胞苷脱氨酶”意为能够催化将氨基转变为羰基的脱氨基反应的多肽或其片段。在一种实施方案中,胞苷脱氨酶将胞嘧啶转变为尿嘧啶或将5-甲基胞嘧啶转变为胸腺嘧啶。源自海七鳃鳗(Petromyzon marinus)的PmCDA1(海七鳃鳗胞嘧啶脱氨酶1,(“PmCDA1”))、或源自哺乳动物(例如,人类、猪、牛、马、猴等)的AID(激活诱导的胞苷脱氨酶;(“AICDA”))和APOBEC是示例性的胞苷脱氨酶。

[0219] PmCDA1的碱基序列和氨基酸序列以及人类AID的碱基序列和氨基酸序列显示如下。

[0220] >tr|A5H718|A5H718_PETMA胞嘧啶脱氨酶OS=海七鳃鳗(Petromyzon marinus)OX=7757PE=2SV=1

[0221] MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQSGTERGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLKIWACKLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKRRSELSIMIQVKILHTTKSPAV

[0222] >EF094822.1海七鳃鳗分离PmCDA.21胞嘧啶脱氨酶mRNA,完全编码序列

[0223] TGACACGACACAGCCGTGTATATGAGGAAGGGTAGCTGGATGGGGGGGGGGGGAATACGTTCAGAGAGGACATTAGCGAGCGTCTTGTTGGTGGCCTTGAGTCTAGACACCTGCAGACATGACCGACGCTGAGTACGTGAGAATCCATGAGAAGTTGGACATCTACACGTTTAAGAAACAGTTTTTCAACAACAAAAAATCCGTGTCGCATAGATGCTACGTTCTCTTTGAATTAAAACGACGGGGTGAACGTAGAGCGTGTTTTTGGGGCTATGCTGTGAATAAACCACAGAGCGGGACAGAACGTGGAATTCACGCCGAAATCTTTAGCATTAGAAAAGTCGAAGAATACCTGCGCGACAACCCCGGACAATTCACGATAAATTGGTACTCATCCTGGAGTCCTTGTGCAGATTGCGCTGAAAAGATCTTAGAATGGTATAACCAGGAGCTGCGGGGGAACGGCCACACTTTGAAAATCTGGGCTTGCAAACTCTATTACGAGAAAAATGCGAGGAATCAAATTGGGCTGTGGAACCTCAGAGATAACGGGGTTGGGTTGAATGTAATGGTAAGTGAACACTACCAATGTTGCAGGAAAATATTCATCCAATCGTCGCACAATCAATTGAATGAGAATAGATGGCTTGAGAAGACTTTGAAGCGAGCTGAAAAACGACGGAGCGAGTTGTCCATTATGATTCAGGTAAAAATACTCCACACCACTAAGAGTCCTGCTGTTTAAGAGGCTATGCGGATGGT

TTTC

[0224] >tr|Q6QJ80|Q6QJ80_人类激活诱导型胞苷脱氨酶OS＝智人(Homo sapiens)OX＝9606GN＝AICDA PE＝2SV＝1

[0225] MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKAPV

[0226] >NG_011588.1:5001-15681智人(Homo sapiens)激活诱导型胞苷脱氨酶(AICDA),RefSeqGene(LRG_17),在12号染色体上AGAGAACCATCATTAATTGAAGTGAGATTTTTCTGGCCTGAGACTTGCAGGGAGGCAAGAAGACACTCTGGACACCACTATGGACAGGTAAAGAGGCAGTCTTCTCGTGGGTGATTGCACTGGCCTTCCTCTCAGAGCAAATCTGAGTAATGAGACTGGTAGCTATCCCTTTCTCTCATGTAACTGTCTGACTGATAAGATCAGCTTGATCAATATGCATATATATTTTTTGATCTGTCTCCTTTTCTTCTATTCAGATCTTATACGCTGTCAGCCCAATTCTTTCTGTTTCAGACTTCTCTTGATTTCCCTCTTTTTCATGTGGCAAAAGAAGTAGTGCGTACAATGTACTGATTCGTCCTGAGATTTGTACCATGGTTGAAACTAATTTATGGTAATAATATTAACATAGCAAATCTTTAGAGACTCAAATCATGAAAAGGTAATAGCAGTACTGTACTAAAAACGGTAGTGCTAATTTTCGTAATAATTTTGTAAATATTCAACAGTAAAACAACTTGAAGACACACTTTCCTAGGGAGGCGTTACTGAAATAATTTAGCTATAGTAAGAAAATTTGTAATTTTAGAAATGCCAAGCATTCTAAATTAATTGCTTGAAAGTCACTATGATTGTGTCCATTATAAGGAGACAAATTCATTCAAGCAAGTTATTTAATGTTAAAGGCCCAATTGTTAGGCAGTTAATGGCACTTTTACTATTAACTAATCTTTCCATTTGTTCAGACGTAGCTTAACTTACCTCTTAGGTGTGAATTTGGTTAAGGTCCTCATAATGTCTTTATGTGCAGTTTTTGATAGGTTATTGTCATAGAACTTATTCTATTCCTACATTTATGATTACTATGGATGTATGAGAATAACACCTAATCCTTATACTTTACCTCAATTTAACTCCTTTATAAAGAACTTACATTACAGAATAAAGATTTTTTAAAAATATATTTTTTTGTAGAGACAGGGTCTTAGCCCAGCCGAGGCTGGTCTCTAAGTCCTGGCCCAAGCGATCCTCCTGCCTGGGCCTCCTAAAGTGCTGGAATTATAGACATGAGCCATCACATCCAATATACAGAATAAAGATTTTTAATGGAGGATTTAATGTTCTTCAGAAAATTTTCTTGAGGTCAGACAATGTCAAATGTCTCCTCAGTTTACACTGAGATTTTGAAAACAAGTCTGAGCTATAGGTCCTTGTGAAGGGTCCATTGGAAATACTTGTTCAAAGTAAAATGGAAAGCAAAGGTAAAATCAGCAGTTGAAATTCAGAGAAAGACAGAAAAGGAGAAAAGATGAAATTCAACAGGACAGAAGGGAAATATATTATCATTAAGGAGGACAGTATCTGTAGAGCTCATTAGTGATGGCAAAATGACTTGGTCAGGATTATTTTTAACCCGCTTGTTTCTGGTTTGCACGGCTGGGGATGCAGCTAGGGTTCTGCCTCAGGGAGCACAGCTGTCCAGAGCAGCTGTCAGCCTGCAAGCCTGAAACACTCCCTCGGTAAAGTCCTTCCTACTCAGGACAGAAATGACGAGAACAGGGAGCTGGAAACAGGCCCCTAACCAGAGAAGGGAAGTAATGGATCAACAAAGTTAACTAGCAGGTCAGGATCACGCAATTCATTTCACTCTGACTGGTAACATGTGACAGAAACAGTGTAGGCTTATTGTATTTTCATGTAGAGTAGGACCCAAAAATCCACCCAAAGTCCTTTATCTATGCCACATCCTTCTTATCTATACTTCCAGGACACTTTTTCTTCCTTATGATAAGGCTCTCTCTCTCTCCACACACACACACACACACACACACACACACACACACACACACACACACACAAACACACACCCCGCCAACCAAGGTGCATGTAAAAAGATGTAGATTCCTCTGCCTTTCTCATCTACACAGCCCAGGAGGGTAAGTTAATATAAGAGGGATTTATTGGTAAGAGATGATGCTTAATCTGTTTAACACTGGGCCTCAAAGAGAGAATTTCTTTTCTTCTGTACTTATTAAGCACCTATTATGTGTTGAGCTTATATATACAAAGGGTTATTATATGCTAATATAGTAATAGTAATGGTGGTTGGTACTATGGTAATTACCATAAAAATTATTATCCTTTTAAAATAAAGCTAATTATTATTGGATCTTTTTTAGTATTCATTTTATGTTTTTTATGTTTTTGATTTTTTAAAAGACAATCTCACCCTGTTACCCAGGCTGGAGTGCAGTGGTGCAATCATAGCTTTCTGCAGTCTTGAACTCCTGGGCTCAAGCAATCCTCCTGCCTTGGCCTCCCAAAGTGTTGGGATACAGTCATGAGCCACTGCATCTGGCCTAGGATCCATTTAGATTAAAATATGCATTTTAAATTTTAAAATAATATGGCTAATTTTTACCTTATGTAATGTGTATACTGGCAATAAATCTAGTTTGCTGCCTAAAGTTTAAAGTGCTTTCCAGTAAGCTTCATGTACGTGAGGGGAGACA

TTTAAAGTGAAACAGACAGCCAGGTGTGGTGGCTCACGCCTGTAATCCCAGCACTCTGGGAGGCTGAGGTGGGTGGA
TCGCTTGAGCCCTGGAGTTCAAGACCAGCCTGAGCAACATGGCAAAACGCTGTTTCTATAACAAAAATTAGCCGGGC
ATGGTGGCATGTGCCTGTGGTCCCAGCTACTAGGGGGCTGAGGCAGGAGAATCGTTGGAGCCCAGGAGGTCAAGGCT
GCACTGAGCAGTGCTTGCGCCACTGCACTCCAGCCTGGGTGACAGGACCAGACCTTGCCTCAAAAAAATAAGAAGAA
AAATTAAAAATAAATGGAAACAACTACAAAGAGCTGTTGTCCTAGATGAGCTACTTAGTTAGGCTGATATTTTGGTA
TTTAACTTTTAAAGTCAGGGTCTGTCACCTGCACTACATTATTAAAATATCAATTCTCAATGTATATCCACACAAAG
ACTGGTACGTGAATGTTCATAGTACCTTTATTCACAAAACCCCAAAGTAGAGACTATCCAAATATCCATCAACAAGT
GAACAAATAAACAAAATGTGCTATATCCATGCAATGGAATACCACCCTGCAGTACAAAGAAGCTACTTGGGGATGAA
TCCCAAAGTCATGACGCTAAATGAAAGAGTCAGACATGAAGGAGGAGATAATGTATGCCATACGAAATTCTAGAAAA
TGAAAGTAACTTATAGTTACAGAAAGCAAATCAGGGCAGGCATAGAGGCTCACACCTGTAATCCCAGCACTTTGAGA
GGCCACGTGGGAAGATTGCTAGAACTCAGGAGTTCAAGACCAGCCTGGGCAACACAGTGAAACTCCATTCTCCACAA
AAATGGGAAAAAAAGAAAGCAAATCAGTGGTTGTCCTGTGGGGAGGGGAAGGACTGCAAAGAGGGAAGAAGCTCTGG
TGGGGTGAGGGTGGTGATTCAGGTTCTGTATCCTGACTGTGGTAGCAGTTTGGGGTGTTTACATCCAAAAATATTCG
TAGAATTATGCATCTTAAATGGGTGGAGTTTACTGTATGTAAATTATACCTCAATGTAAGAAAAAATAATGTGTAAG
AAAACTTTCAATTCTCTTGCCAGCAAACGTTATTCAAATTCCTGAGCCCTTTACTTCGCAAATTCTCTGCACTTCTG
CCCCGTACCATTAGGTGACAGCACTAGCTCCACAAATTGGATAAATGCATTTCTGGAAAAGACTAGGGACAAAATCC
AGGCATCACTTGTGCTTTCATATCAACCATGCTGTACAGCTTGTGTTGCTGTCTGCAGCTGCAATGGGGACTCTTGA
TTTCTTTAAGGAAACTTGGGTTACCAGAGTATTTCCACAAATGCTATTCAAATTAGTGCTTATGATATGCAAGACAC
TGTGCTAGGAGCCAGAAAACAAAGAGGAGGAGAAATCAGTCATTATGTGGGAACAACATAGCAAGATATTTAGATCA
TTTTGACTAGTTAAAAAAGCAGCAGAGTACAAAATCACACATGCAATCAGTATAATCCAAATCATGTAAATATGTGC
CTGTAGAAAGACTAGAGGAATAAACACAAGAATCTTAACAGTCATTGTCATTAGACACTAAGTCTAATTATTATTAT
TAGACACTATGATATTTGAGATTTAAAAAATCTTTAATATTTTAAAATTTAGAGCTCTTCTATTTTTCCATAGTATT
CAAGTTTGACAATGATCAAGTATTACTCTTTCTTTTTTTTTTTTTTTTTTTTTTTTGAGATGGAGTTTTGGTCTTG
TTGCCCATGCTGGAGTGGAATGGCATGACCATAGCTCACTGCAACCTCCACCTCCTGGGTTCAAGCAAAGCTGTCGC
CTCAGCCTCCCGGGTAGATGGGATTACAGGCGCCCACCACCACACTCGGCTAATGTTTGTATTTTTAGTAGAGATGG
GGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCAGAGGATCCACCTGCCTCAGCCTCCCAAAGTGCT
GGGATTACAGATGTAGGCCACTGCGCCCGGCCAAGTATTGCTCTTATACATTAAAAAACAGGTGTGAGCCACTGCGC
CCAGCCAGGTATTGCTCTTATACATTAAAAAATAGGCCGGTGCAGTGGCTCACGCCTGTAATCCCAGCACTTTGGGA
AGCCAAGGCGGGCAGAACACCCGAGGTCAGGAGTCCAAGGCCAGCCTGGCCAAGATGGTGAAACCCCGTCTCTATTA
AAAATACAAACATTACCTGGGCATGATGGTGGGCGCCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGGATCC
GCGGAGCCTGGCAGATCTGCCTGAGCCTGGGAGGTTGAGGCTACAGTAAGCCAAGATCATGCCAGTATACTTCAGCC
TGGGCGACAAAGTGAGACCGTAACAAAAAAAAAAAAATTTAAAAAAAGAAATTTAGATCAAGATCCAACTGTAAAAA
GTGGCCTAAACACCACATTAAAGAGTTTGGAGTTTATTCTGCAGGCAGAAGAGAACCATCAGGGGGTCTTCAGCATG
GGAATGGCATGGTGCACCTGGTTTTTGTGAGATCATGGTGGTGACAGTGTGGGGAATGTTATTTTGGAGGGACTGGA
GGCAGACAGACCGGTTAAAAGGCCAGCACAACAGATAAGGAGGAAGAAGATGAGGGCTTGGACCGAAGCAGAGAAGA
GCAAACAGGGAAGGTACAAATTCAAGAAATATTGGGGGGTTTGAATCAACACATTTAGATGATTAATTAAATATGAG
GACTGAGGAATAAGAAATGAGTCAAGGATGGTTCCAGGCTGCTAGGCTGCTTACCTGAGGTGGCAAAGTCGGGAGGA
GTGGCAGTTTAGGACAGGGGGCAGTTGAGGAATATTGTTTTGATCATTTTGAGTTTGAGGTACAAGTTGGACACTTA
GGTAAAGACTGGAGGGGAAATCTGAATATACAATTATGGGACTGAGGAACAAGTTTATTTTATTTTTTGTTTCGTTT

TCTTGTTGAAGAACAAATTTAATTGTAATCCCAAGTCATCAGCATCTAGAAGACAGTGGCAGGAGGTGACTGTCTTG
TGGGTAAGGGTTTGGGGTCCTTGATGAGTATCTCTCAATTGGCCTTAAATATAAGCAGGAAAAGGAGTTTATGATGG
ATTCCAGGCTCAGCAGGGCTCAGGAGGGCTCAGGCAGCCAGCAGAGGAAGTCAGAGCATCTTCTTTGGTTTAGCCCA
AGTAATGACTTCCTTAAAAAGCTGAAGGAAAATCCAGAGTGACCAGATTATAAACTGTACTCTTGCATTTTCTCTCC
CTCCTCTCACCCACAGCCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCCGCTGGGCTAAGGGT
CGGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGGCGTGACAGTGCTACATCCTTTTCACTGGACTTTGGTTATCT
TCGCAATAAGGTATCAATTAAAGTCGGCTTTGCAAGCAGTTTAATGGTCAACTGTGAGTGCTTTTAGAGCCACCTGC
TGATGGTATTACTTCCATCCTTTTTTGGCATTTGTGTCTCTATCACATTCCTCAAATCCTTTTTTTTATTTCTTTTT
CCATGTCCATGCACCCATATTAGACATGGCCCAAAATATGTGATTTAATTCCTCCCCAGTAATGCTGGGCACCCTAA
TACCACTCCTTCCTTCAGTGCCAAGAACAACTGCTCCCAAACTGTTTACCAGCTTTCCTCAGCATCTGAATTGCCTT
TGAGATTAATTAAGCTAAAAGCATTTTTATATGGGAGAATATTATCAGCTTGTCCAAGCAAAAATTTTAAATGTGAA
AAACAAATTGTGTCTTAAGCATTTTTGAAAATTAAGGAAGAAGAATTTGGGAAAAAATTAACGGTGGCTCAATTCTG
TCTTCCAAATGATTTCTTTTCCCTCCTACTCACATGGGTCGTAGGCCAGTGAATACATTCAACATGGTGATCCCCAG
AAAACTCAGAGAAGCCTCGGCTGATGATTAATTAAATTGATCTTTCGGCTACCCGAGAGAATTACATTTCCAAGAGA
CTTCTTCACCAAAATCCAGATGGGTTTACATAAACTTCTGCCCACGGGTATCTCCTCTCTCCTAACACGCTGTGACG
TCTGGGCTTGGTGGAATCTCAGGGAAGCATCCGTGGGGTGGAAGGTCATCGTCTGGCTCGTTGTTTGATGGTTATAT
TACCATGCAATTTTCTTTGCCTACATTTGTATTGAATACATCCCAATCTCCTTCCTATTCGGTGACATGACACATTC
TATTTCAGAAGGCTTTGATTTTATCAAGCACTTTCATTTACTTCTCATGGCAGTGCCTATTACTTCTCTTACAATAC
CCATCTGTCTGCTTTACCAAAATCTATTTCCCCTTTTCAGATCCTCCCAAATGGTCCTCATAAACTGTCCTGCCTCC
ACCTAGTGGTCCAGGTATATTTCCACAATGTTACATCAACAGGCACTTCTAGCCATTTTCCTTCTCAAAAGGTGCAA
AAAGCAACTTCATAAACACAAATTAAATCTTCGGTGAGGTAGTGTGATGCTGCTTCCTCCCAACTCAGCGCACTTCG
TCTTCCTCATTCCACAAAAACCCATAGCCTTCCTTCACTCTGCAGGACTAGTGCTGCCAAGGGTTCAGCTCTACCTA
CTGGTGTGCTCTTTTGAGCAAGTTGCTTAGCCTCTCTGTAACACAAGGACAATAGCTGCAAGCATCCCCAAAGATCA
TTGCAGGAGACAATGACTAAGGCTACCAGAGCCGCAATAAAAGTCAGTGAATTTTAGCGTGGTCCTCTCTGTCTCTC
CAGAACGGCTGCCACGTGGAATTGCTCTTCCTCCGCTACATCTCGGACTGGGACCTAGACCCTGGCCGCTGCTACCG
CGTCACCTGGTTCACCTCCTGGAGCCCCTGCTACGACTGTGCCCGACATGTGGCCGACTTTCTGCGAGGGAACCCCA
ACCTCAGTCTGAGGATCTTCACCGCGCGCCTCTACTTCTGTGAGGACCGCAAGGCTGAGCCCGAGGGGCTGCGGCGG
CTGCACCGCGCCGGGGTGCAAATAGCCATCATGACCTTCAAAGGTGCGAAAGGGCCTTCCGCGCAGGCGCAGTGCAG
CAGCCCGCATTCGGGATTGCGATGCGGAATGAATGAGTTAGTGGGGAAGCTCGAGGGGAAGAAGTGGGCGGGGATTC
TGGTTCACCTCTGGAGCCGAAATTAAAGATTAGAAGCAGAGAAAAGAGTGAATGGCTCAGAGACAAGGCCCCGAGGA
AATGAGAAAATGGGGCCAGGGTTGCTTCTTTCCCCTCGATTTGGAACCTGAACTGTCTTCTACCCCCATATCCCCGC
CTTTTTTTCCTTTTTTTTTTTTTGAAGATTATTTTTACTGCTGGAATACTTTTGTAGAAAACCACGAAAGAACTTTC
AAAGCCTGGGAAGGGCTGCATGAAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCCTTTTGGTAAGGGGCTT
CCTCGCTTTTTAAATTTTCTTTCTTTCTCTACAGTCTTTTTTGGAGTTTCGTATATTTCTTATATTTTCTTATTGTT
CAATCACTCTCAGTTTTCATCTGATGAAAACTTTATTTCTCCTCCACATCAGCTTTTTCTTCTGCTGTTTCACCATT
CAGAGCCCTCTGCTAAGGTTCCTTTTCCCTCCCTTTTCTTTCTTTTGTTGTTTCACATCTTTAAATTTCTGTCTCTC
CCCAGGGTTGCGTTTCCTTCCTGGTCAGAATTCTTTTCTCCTTTTTTTTTTTTTTTTTTTTTTTTTTAAACAAACA
AACAAAAAACCCAAAAAAACTCTTTCCCAATTTACTTTCTTCCAACATGTTACAAAGCCATCCACTCAGTTTAGAAG
ACTCTCCGGCCCCACCGACCCCCAACCTCGTTTTGAAGCCATTCACTCAATTTGCTTCTCTCTTTCTCTACAGCCCC

TGTATGAGGTTGATGACTTACGAGACGCATTTCGTACTTTGGGACTTTGATAGCAACTTCCAGGAATGTCACACACG
ATGAAATATCTCTGCTGAAGACAGTGGATAAAAAACAGTCCTTCAAGTCTTCTCTGTTTTTATTCTTCAACTCTCAC
TTTCTTAGAGTTTACAGAAAAAATATTTATATACGACTCTTTAAAAAGATCTATGTCTTGAAAATAGAGAAGGAACA
CAGGTCTGGCCAGGGACGTGCTGCAATTGGTGCAGTTTTGAATGCAACATTGTCCCCTACTGGGAATAACAGAACTG
CAGGACCTGGGAGCATCCTAAAGTGTCAACGTTTTTCTATGACTTTTAGGTAGGATGAGAGCAGAAGGTAGATCCTA
AAAAGCATGGTGAGAGGATCAAATGTTTTTATATCAACATCCTTTATTATTTGATTCATTTGAGTTAACAGTGGTGT
TAGTGATAGATTTTTCTATTCTTTTCCCTTGACGTTTACTTTCAAGTAACACAAACTCTTCCATCAGGCCATGATCT
ATAGGACCTCCTAATGAGAGTATCTGGGTGATTGTGACCCCAAACCATCTCTCCAAAGCATTAATATCCAATCATGC
GCTGTATGTTTTAATCAGCAGAAGCATGTTTTTATGTTTGTACAAAAGAAGATTGTTATGGGTGGGGATGGAGGTAT
AGACCATGCATGGTCACCTTCAAGCTACTTTAATAAAGGATCTTAAAATGGGCAGGAGGACTGTGAACAAGACACCC
TAATAATGGGTTGATGTCTGAAGTAGCAAATCTTCTGGAAACGCAAACTCTTTTAAGGAAGTCCCTAATTTAGAAAC
ACCCACAAACTTCACATATCATAATTAGCAAACAATTGGAAGGAAGTTGCTTGAATGTTGGGGAGAGGAAAATCTAT
TGGCTCTCGTGGGTCTCTTCATCTCAGAAATGCCAATCAGGTCAAGGTTTGCTACATTTTGTATGTGTGTGATGCTT
CTCCCAAAGGTATATTAACTATATAAGAGAGTTGTGACAAAACAGAATGATAAAGCTGCGAACCGTGGCACACGCTC
ATAGTTCTAGCTGCTTGGGAGGTTGAGGAGGGAGGATGGCTTGAACACAGGTGTTCAAGGCCAGCCTGGGCAACATA
ACAAGATCCTGTCTCTCAAAAAAAAAAAAAAAAAAAAGAAAGAGAGAGGGCCGGGCGTGGTGGCTCACGCCTGTAAT
CCCAGCACTTTGGGAGGCCGAGCCGGGCGGATCACCTGTGGTCAGGAGTTTGAGACCAGCCTGGCCAACATGGCAAA
ACCCCGTCTGTACTCAAAATGCAAAAATTAGCCAGGCGTGGTAGCAGGCACCTGTAATCCCAGCTACTTGGGAGGCT
GAGGCAGGAGAATCGCTTGAACCCAGGAGGTGGAGGTTGCAGTAAGCTGAGATCGTGCCGTTGCACTCCAGCCTGGG
CGACAAGAGCAAGACTCTGTCTCAGAAAAAAAAAAAAAAAAAGAGAGAGAGAGAGAGAAAGAGAACAATATTTGGGAGAG
AAGGATGGGGAAGCATTGCAAGGAAATTGTGCTTTATCCAACAAAATGTAAGGAGCCAATAAGGGATCCCTATTTGT
CTCTTTTGGTGTCTATTTGTCCCTAACAACTGTCTTTGACAGTGAGAAAAATATTCAGAATAACCATATCCCTGTGC
CGTTATTACCTAGCAACCCTTGCAATGAAGATGAGCAGATCCACAGGAAAACTTGAATGCACAACTGTCTTATTTTA
ATCTTATTGTACATAAGTTTGTAAAAGAGTTAAAAATTGTTACTTCATGTATTCATTTATATTTTATATTATTTTGC
GTCTAATGATTTTTTATTAACATGATTTCCTTTTCTGATATATTGAAATGGAGTCTCAAAGCTTCATAAATTTATAA
CTTTAGAAATGATTCTAATAACAACGTATGTAATTGTAACATTGCAGTAATGGTGCTACGAAGCCATTTCTCTTGAT
TTTTAGTAAACTTTTATGACAGCAAATTTGCTTCTGGCTCACTTTCAATCAGTTAAATAAATGATAAATAATTTTGG
AAGCTGTGAAGATAAAATACCAAATAAAATAATATAAAAGTGATTTATATGAAGTTAAAATAAAAAATCAGTATGAT
GGAATAAACTTG

[0227] 载脂蛋白B mRNA编辑酶,催化性多肽样(APOBEC)是一个进化上保守的胞苷脱氨酶家族。这一家族的成员是C到U编辑酶。APOBEC样蛋白的N端结构域是催化性结构域,而C端结构域是假催化性结构域。更具体地,催化性结构域是锌依赖性胞苷脱氨酶结构域,且对于胞苷脱氨基而言非常重要。APOBEC家族成员包括APOBEC1、APOBEC2、APOBEC3A、APOBEC3B、APOBEC3C、APOBEC3D(现在称为“APOBEC3E”)、APOBEC3F、APOBEC3G、APOBEC3H、APOBEC4和激活诱导型(胞苷)脱氨酶。很多经修饰的胞苷脱氨酶是可商购的,包括但不限于,SaBE3、SaKKH-BE3、VQR-BE3、EQR-BE3、VRER-BE3、YE1-BE3、EE-BE3、YE2-BE3和YEE-BE3,它们可以从Addgene获得(质粒85169、85170、85171、85172、85173、85174、85175、85176、85177)。

[0228] 可根据本公开的方面融合至Cas9的其他示例性脱氨酶提供于本文中。应理解,在一些实施方案中,可使用相应序列的活性结构域,例如灭有定位信号的结构域(核定位序

列,没有核输出信号、细胞质定位信号)。

[0229] 如本文所用,术语“脱氨酶”或“脱氨酶结构域”是指催化脱氨基反应的蛋白质或酶。在一些实施方案中,脱氨酶或脱氨酶结构域是胞苷脱氨酶,其催化胞苷或者脱氧胞苷分别至尿苷或脱氧尿苷的水解性脱氨基。在一些实施方案中,脱氨酶或脱氨酶结构域是胞嘧啶脱氨酶,其催化胞嘧啶至尿嘧啶的水解性脱氨基。在一些实施方案中,脱氨酶是腺苷脱氨酶,其催化腺嘌呤至次黄嘌呤的水解性脱氨基。在一些实施方案中,脱氨酶是腺苷脱氨酶,其催化腺苷或腺嘌呤(A)至肌苷(I)的水解性脱氨基。在一些实施方案中,脱氨酶或脱氨酶结构域是腺苷脱氨酶,其催化腺苷或脱氧腺苷分别至肌苷或脱氧肌苷的水解性脱氨基。在一些实施方案中,腺苷脱氨酶催化脱氧核糖核酸(DNA)中的腺苷的水解脱氨基。本文提供的腺苷脱氨酶(例如,工程改造的腺苷脱氨酶、进化的腺苷脱氨酶)可以来自任何生物体诸如细菌。在一些实施方案中,腺苷脱氨酶来自细菌,诸如大肠杆菌(Escherichia coli)、金黄色葡萄球菌(Staphylococcus aureus)、伤寒沙门氏菌(Salmonella typhimurium)、腐败希瓦氏菌(Shewanella putrefaciens)、流感嗜血杆菌(Haemophilus influenzae)或新月柄杆菌(Caulobacter crescentus)。

[0230] 在一些实施方案中,脱氨酶或脱氨酶结构域是来自生物体诸如人类、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,腺苷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为至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%、至少99.1%、至少99.2%、至少99.3%、至少99.4%、至少99.5%、至少99.6%、至少99.7%、至少99.8%或至少99.9%相同。

[0231] 例如,碱基编辑器在国际PCT申请号PCT/2017/045381(WO 2018/027078)和PCT/US2016/058344(WO 2017/070632)中描述,其各自通过引用以其整体并入本文。也参见Komor,A.C.,等人,“Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage”Nature 533,420-424(2016);Gaudelli,N.M.,等人,“Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage”Nature 551,464-471(2017);Komor,A.C.,等人,“Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:Abase editors with higher efficiency and product purity”Science Advances 3:eaao4774(2017);和Rees,H.A.,等人,“Base editing:precision chemistry on the genome and transcriptome of living cells.”Nat Rev Genet.2018Dec;19(12):770-788.doi:10.1038/s41576-018-0059-1,其整体内容通过引用并入本文。

[0232] “检测”是指鉴定待检测分析物的存在、不存在或量。

[0233] “可检测标签”意为一种组合物,其链接至感兴趣的分子,使得后者可经由光谱、光化学、生物化学、免疫化学或化学手段检测。例如,可用的标签包括放射活性同位素、磁珠、金属性珠、胶体颗粒、荧光染料、电子密集型试剂、酶(例如,如酶联免疫吸附测定(ELISA)中常用者)、生物素、地高辛或半抗原。

[0234] “疾病”意为损害或干扰细胞、组织或器官正常功能的任何病症或病变。在一种实施方案中,疾病是瘤形成或癌症。在一些实施方案中,疾病是血液学癌症。“血液学癌症”意

为免疫系统细胞的恶性病变。在一些实施方案中,血液学癌症是白血病、骨髓瘤和/或淋巴瘤。淋巴瘤和白血病是“液态癌症”或者存在于血液中的癌症的示例,并且源自骨髓中造血前体或血液中成熟造血细胞的转化。白血病可以是淋巴性的或骨髓性的以及急性的或慢性的。在骨髓瘤的情况下,转化的细胞是完全分化的浆细胞,可能作为恶性细胞的分散集合或作为骨髓中的固体团块的形式存在。在淋巴瘤的情况下,在次级淋巴组织中的转化的淋巴细胞生成固体团块。淋巴瘤归类为霍奇金淋巴瘤(HL)或非霍奇金淋巴瘤(NHL)。

[0235] 在一些实施方案中,血液学癌症是B细胞癌症。在一些实施方案中,B细胞癌症是淋巴瘤或白血病。在一些情况下,白血病包括白血病前期。在一些情况下,白血病是急性白血病。急性白血病包括,例如,急性骨髓性白血病(AML)。急性白血病也包括,例如,急性淋巴性白血病或急性淋巴细胞性白血病(ALL);ALL包括B谱系ALL、T谱系ALL和T细胞急性淋巴细胞性白血病(T-ALL)。

[0236] 疾病的非限制性示例包括T细胞急性成淋巴细胞性白血病(T-ALL)、蕈样肉芽肿(MF)、Sézary综合征(SS)、外周T/NK细胞淋巴瘤、间变性大细胞淋巴瘤ALK+、原发性皮肤T细胞淋巴瘤、T细胞大颗粒淋巴细胞型白血病、血管免疫母细胞性T/NK细胞淋巴瘤、肝脾T细胞淋巴瘤、原发性皮肤CD30+淋巴增生性疾病、结外NK/T细胞淋巴瘤、成人T细胞白血病/淋巴瘤、T细胞幼淋巴细胞白血病、皮下脂膜炎样T细胞淋巴瘤、原发性皮肤γδT细胞淋巴瘤、侵袭性NK细胞白血病和肠病相关T细胞淋巴瘤。在一些实施方案中,疾病是液态肿瘤。在一些实施方案中,疾病是T细胞急性成淋巴细胞性白血病(T-ALL)。在一些实施方案中,疾病是T细胞急性骨髓性白血病(AML)。

[0237] 如本文所用,术语“有效量”是指生物活性剂的足以引起所希望的生物学应答的量。在一些实施方案中,有效量是相对于未治疗的患者缓解疾病症状所需的量。用来实践本发明的用于对疾病进行治疗性处理的活性剂的有效量依据给药方式以及受试者的年龄、体重和一般健康状况而变。最终,主治医师或兽医将决定适宜的量和剂量方案。此量称为“有效”量。在一种实施方案中,有效量是本发明的碱基编辑器(例如,包含可编程DNA结合蛋白、核碱基编辑器和gRNA的融合蛋白)的量,该量足以将改变引入细胞(例如,体外细胞或体内细胞)内的感兴趣的基因中。在一种实施方案中,有效量是实现治疗效果(例如,减轻或控制疾病或其症状或病状)所需的碱基编辑器的量。此类治疗效果不必足以改变受试者、组织或器官的全部细胞中的感兴趣基因,而可以仅改变受试者、组织或器官中存在的约1%、5%、10%、25%、50%、75%或更多细胞中的感兴趣基因。

[0238] 在一些实施方案中,本文所提供的融合蛋白的有效量,例如,包含nCas9结构域和脱氨酶结构域(例如,腺苷脱氨酶结构域或胞苷脱氨酶结构域)的核碱基编辑器的有效量,是指融合蛋白的量,该量足以诱导对由本文所述的核碱基编辑器特异性地结合并编辑的靶标位点的编辑。如本领域技术人员将会了解的,药剂(例如,融合蛋白、核酸酶、杂交蛋白、蛋白二聚体、蛋白质(或蛋白二聚体)与多核苷酸的复合物、或多核苷酸)的量可依据各种因素而变,例如,依据所希望的生物学应答而变,例如,依据待编辑的特异性等位基因、基因组或靶标位点而变,依据待靶向的细胞或组织而不,和/或依据所使用的药剂而变。在CAR-T细胞的情境中,“有效量”是指给药至患者以实现治疗性应答所必需的细胞的数量。

[0239] 如本文所用,术语“表位”意为抗原决定簇。表位是抗原分子的一部分,通过其结构来决定江湖识别并结合它的特异性抗体。

[0240] “Fas细胞表面死亡受体(FAS)多肽”意为一种蛋白质,其与NCBI登录号NP_000034.1或其片段具有至少约85%的氨基酸序列一致性。示例性氨基酸序列提供如下。

[0241] >NP_000034.1肿瘤坏死因子受体超家族成员6亚型1前体[智人(Homo sapiens)]

```
        1 mlgiwtllpl vltsvarlss ksvnaqvtdi nskglelrkt vttvetqnle
glhhdgqfch
       61   kpcppgerka   rdctvngdep   dcvpcqegke   ytdkahfssk
crrcrlcdeg hgleveinct
      121   rtqntkcrck   pnffcnstvc   ehcdpctkce   hgiikectlt
```

[0242]

```
sntkckeegs rsnlgwlcll
      181   llpiplivwv   krkevqktcr   khrkenqgsh   esptlnpetv
ainlsdvdls kyittiagvm
      241   tlsqvkgfvr   kngvneakid   eikndnvqdt   aeqkvqllrn
whqlhgkkea ydtlikdlkk
      301 anlctlaeki qtiilkdits dsensnfrne iqslv
```

[0243] “Fas细胞表面死亡受体(FAS)多核苷酸”意为编码FAS多肽的核酸。示例性FAS核酸序列提供如下。

[0244] >NM_000043.6智人(Homo sapiens)Fas细胞表面死亡受体(FAS),转录物变体1,mRNA

```
        1 ctcttctccc gcgggttggt ggacccgctc agtacggagt tggggaagct
ctttcacttc
       61   ggaggattgc   tcaacaacca   tgctgggcat   ctggaccctc
ctacctctgg ttcttacgtc
      121   tgttgctaga   ttatcgtcca   aaagtgttaa   tgcccaagtg
actgacatca actccaaggg
      181   attggaattg   aggaagactg   ttactacagt   tgagactcag
aacttggaag gcctgcatca
```

[0245]

```
      241   tgatggccaa   ttctgccata   agccctgtcc   tccaggtgaa
aggaaagcta gggactgcac
      301   agtcaatggg   gatgaaccag   actgcgtgcc   ctgccaagaa
gggaaggagt acacagacaa
      361   agcccatttt   tcttccaaat   gcagaagatg   tagattgtgt
gatgaaggac atggcttaga
      421   agtggaaata   aactgcaccc   ggacccagaa   taccaagtgc
agatgtaaac caaacttttt
      481   ttgtaactct   actgtatgtg   aacactgtga   cccttgcacc
```

[0246]

```
aaatgtgaac atggaatcat
     541   caaggaatgc   acactcacca   gcaacaccaa   gtgcaaagag
gaaggatcca gatctaactt
     601   ggggtggctt   tgtcttcttc   ttttgccaat   tccactaatt
gtttgggtga agagaaagga
     661   agtacagaaa   acatgcagaa   agcacagaaa   ggaaaaccaa
ggttctcatg aatctccaac
     721   tttaaatcct   gaaacagtgg   caataaattt   atctgatgtt
gacttgagta aatatatcac
     781   cactattgct   ggagtcatga   cactaagtca   agttaaaggc
tttgttcgaa agaatggtgt
     841   caatgaagcc   aaaatagatg   agatcaagaa   tgacaatgtc
caagacacag cagaacagaa
     901   agttcaactg   cttcgtaatt   ggcatcaact   tcatggaaag
aaagaagcgt atgacacatt
     961   gattaaagat   ctcaaaaaag   ccaatctttg   tactcttgca
gagaaaattc agactatcat
    1021   cctcaaggac   attactagtg   actcagaaaa   ttcaaacttc
agaaatgaaa tccaaagctt
    1081   ggtctagagt   gaaaaacaac   aaattcagtt   ctgagtatat
gcaattagtg tttgaaaaga
    1141   ttcttaatag   ctggctgtaa   atactgcttg   gtttttttact
gggtacattt tatcatttat
    1201   tagcgctgaa   gagccaacat   atttgtagat   ttttaatatc
tcatgattct gcctccaagg
    1261   atgtttaaaa   tctagttggg   aaaacaaact   tcatcaagag
taaatgcagt ggcatgctaa
    1321   gtacccaaat   aggagtgtat   gcagaggatg   aaagattaag
attatgctct ggcatctaac
    1381   atatgattct   gtagtatgaa   tgtaatcagt   gtatgttagt
acaaatgtct atccacaggc
    1441   taaccccact   ctatgaatca   atagaagaag   ctatgacctt
ttgctgaaat atcagttact
    1501   gaacaggcag   gccactttgc   ctctaaatta   cctctgataa
ttctagagat tttaccatat
    1561   ttctaaactt   tgtttataac   tctgagaaga   tcatatttat
gtaaagtata tgtatttgag
    1621   tgcagaattt   aaataaggct   ctacctcaaa   gacctttgca
cagtttattg gtgtcatatt
    1681   atacaatatt   tcaattgtga   attcacatag   aaaacattaa
attataatgt ttgactatta
    1741   tatatgtgta   tgcattttac   tggctcaaaa   ctacctactt
ctttctcagg catcaaaagc
    1801   attttgagca   ggagagtatt   actagagctt   tgccacctct
ccatttttgc cttggtgctc
    1861   atcttaatgg   cctaatgcac   ccccaaacat   ggaaatatca
ccaaaaaata cttaatagtc
    1921   caccaaaagg   caagactgcc   cttagaaatt   ctagcctggt
ttggagatac taactgctct
    1981   cagagaaagt   agctttgtga   catgtcatga   acccatgttt
gcaatcaaag atgataaaat
    2041   agattcttat   ttttccccca   cccccgaaaa   tgttcaataa
```

tgtcccatgt aaaacctgct
2101 acaaatggca gcttatacat agcaatggta aaatcatcat
ctggatttag gaattgctct
2161 tgtcataccc ccaagtttct aagatttaag attctcctta
ctactatcct acgtttaaat
2221 atctttgaaa gtttgtatta aatgtgaatt ttaagaaata
atatttatat ttctgtaaat
2281 gtaaactgtg aagatagtta taaactgaag cagatacctg
gaaccaccta aagaacttcc
2341 atttatggag gatttttttg ccccttgtgt ttggaattat
aaaatatagg taaaagtacg
2401 taattaaata atgttttgg tatttctggt tttctctttt
ttggtagggg cttgcttttt
2461 ggttttgtct tccttttctc taactgatgc taaatataac
ttgtctttaa tgcttcttgg
2521 atcccttaga aggtacttcc tttttaacct taacccttt
agtagttaaa taattatttc
2581 cataggttgc tattgccaag aagacctctt ccaaacagca
catgattatt cgtcaaacag
2641 tttcgtattc cagatactgg aatgtggata agaaagtata
catttcaagg ggtaggtttt
2701 attattaaga aagccaaatg aggattttga aatattcttt
cctgcatatt atccattcta
2761 gctacatgct ggccagtggg ccacctttct tttctgcaat
ttaatgctag taatatattc
2821 tatttaaccc atgagtccca aagtattagc atttcaacat
[0247] gtaagcatgt cggtaagata
2881 gttgtgcttt gcttagggtt ccctcctgtg ttatggtctg
gaaagtgtct ttaggcagaa
2941 agtctgagtg atcacagggt tcactcatta atttctcttt
tctgagccat catagtctgt
3001 gctgtctgct ctccagtttt ctatttctag acagaagtag
ggcaagttag gtactagtta
3061 ttcttcatgg ccagaagtgc aagttctact ttgcaagaca
agattaagtt agagaacacc
3121 ctattccact ttggtgaact cagagcaaga actttgagtt
cctttgggag gaagacagtg
3181 gagaagtctt tgtacttggt gatgtggttt ttttcctcat
ggcttcacct agtggcccca
3241 agcatgactt ctcccatgtc aatgagcaca gccacattcc
cgagttgagg tgaccccacg
3301 gtccagaatc atcctcattc tggtgaacct ggttctcttt
gtggtgggca tactgggtag
3361 gagaatcacc caaaggtcac ccatgagctg cagaaaaaaa
ggctatttgc agaaggagct
3421 cacagatcac attgaaagca ttgcatattc aaacatcttg
gtcttcttta ttggcatgcc
3481 cacagggtct tctgacctct gattagatca gacacttttt
agatattgaa tcatcagttt
3541 ctgtacaact atctgaataa ggtatataat caatgaaatt
tagaattttt ttctatgctt
3601 actcctgatt ggtaatttgt ttgggtttag aattctatac
aaggccattt gtaattttcc
[0248] 3661 tcagcacttt aaaaatatta aaccatgttt tcttaa

[0249] “片段”意为多肽或核酸分子的一部分。这一部分含有参考核酸分子或多肽整个长度的至少约10％、20％、30％、40％、50％、60％、70％、80％或90％。片段可含有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。

[0250] “同类相杀”意为免疫细胞被其他免疫细胞杀死,包括自身抗原驱动的免疫细胞杀死。在某些实施方案中,本发明的免疫细胞经基因修饰以防止或减少被表达嵌合抗原受体(CAR)的免疫细胞识别的抗原的表达,从而防止或减少同类相杀。在各种实施方案中,同类相杀可以在体内(例如,在受试者体内)或间接体内(例如,在免疫细胞制剂中)发生。

[0251] “移植物抗宿主疾病”(GVHD)是指一种病理状况,其中所移植的供体细胞生成针对宿主细胞的免疫应答。

[0252] “向导RNA”或“gRNA”意为一种多核苷酸,其对于靶标序列可能具有特异性并且可以与多核苷酸可编程核苷酸结合结构域蛋白(例如,Cas9或Cpf1)形成复合物。在一种实施方案中,向导多核苷酸是向导RNA(gRNA)。gRNA可以作为两种或更多种RNA的复合物存在,或作为单一RNA分子存在。作为单一RNA分子存在的gRNA可以称为单向导RNA(sgRNA),但“gRNA”可互换地使用以指代作为单一分子或作为两个或更多个分子的复合物存在的向导RNA。典型地,作为单一RNA物质存在的gRNA包含两个结构域:(1)与靶标核酸共享同源性的结构域(例如,并且引导Cas9复合物至靶标的结合);和(2)结合Cas9蛋白的结构域。在一些实施方案中,结构域(2)对应于称为tracrRNA的序列并且包含茎环结构。例如,在一些实施方案中,结构域(2)与Jinek等人,Science337:816-821(2012)中提供的tracrRNA相同或同源,该文献的整体内容通过引用并入本文。gRNA的其他示例(例如,那些包括结构域2的)可见于题为“Switchable Cas9 Nucleases and Uses Thereof”的US20160208288和题为“Delivery System For Functional Nucleases”的US 9,737,604中,各专利的整体内容通过引用而以其整体并入本文。在一些实施方案中,gRNA包含两个或更多个结构域(1)和(2)并且可以成为“扩展gRNA”。扩展gRNA将会结合两个或更多个Cas9蛋白并在两个或更多个不同区域处结合靶标核酸,如本文所述。gRNA包含与靶标位点互补的核苷酸序列,其介导核酸酶/RNA复合物至所述靶标位点的结合,提供核酸酶:RNA复合物的序列特异性。如本领域技术人员将会了解的,RNA多核苷酸序列,例如,gRNA序列,包括核碱基尿嘧啶(U)(一种嘧啶衍生物)而非核碱基胸腺嘧啶(T),胸腺嘧啶包括在DNA多核苷酸序列中。在RNA中,尿嘧啶与腺嘌呤进行碱基配对并且在DNA转录期间替换胸腺嘧啶。

[0253] “异二聚体”意为一种融合蛋白,其包含两个结构域,诸如野生型TadA结构域和TadA结构域的变体(例如,TadA*8),或两个变体TadA及头晕(例如,TadA*7.10和TadA*8,或者两个TadA*8结构域)。

[0254] “宿主抗移植物疾病”(HVGD)是指一种病理状况,其中宿主的免疫系统生成针对所移植的供体细胞的免疫应答。

[0255] “杂交”意为互补核碱基之间的氢键键合,该氢键键合可以是Watson-Crick、Hoogsteen或反向Hoogsteen氢键键合。例如,腺嘌呤和胸腺嘧啶是通过形成氢键进行配对的互补核碱基。

[0256] “免疫细胞”意为能够生成免疫应答的免疫系统的细胞。

[0257] “免疫效应细胞”意为一种淋巴细胞,其一旦被激活就能够对于靶标细胞实现免疫

应答。在一些实施方案中,免疫效应细胞是效应子T细胞。在一些实施方案中,效应子T细胞是天然CD8+ T细胞、细胞毒T细胞、天然杀手T(NKT)细胞、天然杀手(NK)细胞或调节性T(Treg)细胞。在一些实施方案中,免疫效应细胞是效应子NK细胞。在一些实施方案中,效应子T细胞是胸腺细胞、不成熟的T淋巴细胞、成熟的T淋巴细胞、静息T淋巴细胞或激活的T淋巴细胞。在一些实施方案中,免疫效应细胞是CD4+CD8+ T细胞或CD4-CD8-T细胞。在一些实施方案中,免疫效应细胞是T辅助细胞。在一些实施方案中,T辅助细胞是T辅助1(Th1)、T辅助2(Th2)细胞或表达CD4的辅助T细胞(CD4+T细胞)。

[0258] “免疫应答调节基因”或“免疫应答调节因子”意为一种基因,其编码牵涉入免疫应答调节中的多肽。免疫应答调节基因可以以多种机制或在不同层面上调节免疫应答。例如,免疫应答调节基因可以抑制或促进免疫细胞例如T细胞的激活。免疫应答调节基因可以增加或降低免疫细胞的激活阈值。在一些实施方案中,免疫应答调节基因正向调节免疫细胞信号转导途径。在一些实施方案中,免疫应答调节基因负向调节免疫细胞信号转导途径。在一些实施方案中,免疫应答调节基因编码抗原、抗体、细胞因子或神经内分泌。

[0259] “免疫原性基因”意为一种基因,其编码能引起免疫应答的多肽。例如,免疫原性基因可以编码引起免疫应答的免疫原。在一些实施方案中,免疫原性基因编码细胞表面蛋白。在一些实施方案中,免疫原性基因编码细胞表面抗原或细胞表面标志物。在一些实施方案中,细胞表面标志物是T细胞标志物或B细胞标志物。在一些实施方案中,免疫原性基因编码CD2、CD3e、CD3 delta、CD3 gamma、TRAC、TRBC1、TRBC2、CD4、CD5、CD7、CD8、CD19、CD23、CD27、CD28、CD30、CD33、CD52、CD70、CD127、CD122、CD130、CD132、CD38、CD69、CD11a、CD58、CD99、CD103、CCR4、CCR5、CCR6、CCR9、CCR10、CXCR3、CXCR4、CLA、CD161、B2M或CIITA多肽。

[0260] 术语“碱基修复的抑制剂”或“IBR”是指一种蛋白质,其能够抑制核酸修复酶(例如,碱基切除修复(BER)酶)的活性。在一些实施方案中,IBR是肌苷碱基切除修复的抑制剂。示例性的剪辑修复抑制剂包括APE1、Endo III、Endo IV、Endo V、Endo VIII、Fpg、hOGG1、hNEIL1、T7 Endo1、T4PDG、UDG、hSMUG1和hAAG的抑制剂。在一些实施方案中,IBR是Endo V或hAAG的抑制剂。在一些实施方案中,IBR是无催化活性的EndoV或无催化活性的hAAG。在一些实施方案中,碱基修复抑制剂是Endo V或hAAG的抑制剂。在一些实施方案中,碱基修复抑制剂是无催化活性的EndoV或无催化活性的hAAG。

[0261] 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结构域包含野生型UGI或野生型UGI的片段。在一些实施方案中,本文所提供的UGI蛋白包括UGI的片段和与UGI或UGI片段同源的蛋白质。在一些实施方案中,碱基修复抑制剂是肌苷碱基切除修复的抑制剂。在一些实施方案中,碱基修复抑制剂是“无催化活性的肌苷特异性核酸酶”或“死亡的肌苷特异性核酸酶”。不欲受缚于特定理论,无催化活性的肌苷糖基化酶(例如,烷基腺嘌呤糖基化酶(AAG))可以结合肌苷,但不能创建无碱基位点或去除该肌苷,从而在空间上将新形成的肌苷部分从DNA损伤/修复机制中阻隔开来。在一些实施方案中,物催化活性的肌苷特异性核酸酶可能能够结合核酸中的肌苷但不切割该核酸。非限制性的示例性的无催化活性的肌苷特异性核酸酶包括无催化活性的烷基腺苷糖基化酶(AAG核酸酶),例如,来自人类;以及无催化活性的核酸内切酶V(EndoV核酸酶),例如,来自大肠杆菌。在一些实施方案中,无催化活性的AAG核酸酶包含E125Q突变或另一AAG核酸酶中的对应突变。

[0262] “增加”意为至少10％、25％、50％、75％或100％的正向改变。

[0263] “内含肽”是蛋白质的片段,其能够在称为蛋白质剪接的过程中切除其自身并且用肽键将剩余片段(外显肽)接合。内含肽也称为“蛋白质内含子”。内含肽切除其自身并接合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(在内含肽介导的蛋白质剪接之前含有内含肽的蛋白质)的内含肽来自两个基因。本文中,这种内含肽称为分体内含肽(例如,分体内含肽-N和分体内含肽-C)。例如,在蓝藻中,DnaE(即,DNA聚合酶III的催化亚基)由两个独立的基因(即dnaE-n和dnaE-c)编码。本文中,由dnaE-n基因编码的内含肽可称为“内含肽-N”。本文中,由dnaE-c基因编码的内含肽可称为“内含肽-C”。

[0264] 也可使用其他内含肽系统。例如,已经描述了基于dnaE内含肽的合成内含肽,即Cfa-N(例如,分体内含肽-N)和Cfa-C(例如,分体内含肽-C)内含肽对(例如,在Stevens等人,J Am Chem Soc.2016Feb.24;138(7):2162-5中,该文献通过引用并入本文)。可根据本公开使用的内含肽对的非限制性示例包括:Cfa DnaE内含肽、Ssp GyrB内含肽、Ssp DnaX内含肽、Ter DnaE3内含肽、Ter ThyX内含肽、Rma DnaB内含肽和Cne Prp8内含肽(例如,如美国专利号8,394,604中所述,该专利通过引用并入本文)。

[0265] 下文提供示例性的内含肽的核苷酸和氨基酸序列。

[0266] DnaE内含肽-N DNA:

[0267] TGCCTGTCATACGAAACCGAGATACTGACAGTAGAATATGGCCTTCTGCCAATCGGGAAGATTGTGGAGAAACGGATAGAATGCACAGTTTACTCTGTCGATAACAATGGTAACATTTATACTCAGCCAGTTGCCCAGTGGCACGACCGGGGAGAGCAGGAAGTATTCGAATACTGTCTGGAGGATGGAAGTCTCATTAGGGCCACTAAGGACCACAAATTTATGACAGTCGATGGCCAGATGCTGCCTATAGACGAAATCTTTGAGCGAGAGTTGGACCTCATGCGAGTTGACAACCTTCCTAAT

[0268] DnaE内含肽-N蛋白:

[0269] CLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYCLEDGSLIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPN

[0270] DnaE内含肽-C DNA:

[0271] ATGATCAAGATAGCTACAAGGAAGTATCTTGGCAAACAAAACGTTTATGATATTGGAGTCGAAAGAGATCACAACTTTGCTCTGAAGAACGGATTCATAGCTTCTAAT

[0272] 内含肽-C:MIKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN

[0273] Cfa-N DNA:

[0274] TGCCTGTCTTATGATACCGAGATACTTACCGTTGAATATGGCTTCTTGCCTATTGGAAAGATTGTCGAAGAGAGAATTGAATGCACAGTATATACTGTAGACAAGAATGGTTTCGTTTACACACAGCCCATTGCTCAATGGCACAATCGCGGCGAACAAGAAGTATTTGAGTACTGTCTCGAGGATGGAAGCATCATACGAGCAACTAAAGATCATAAATTCATGACCACTGACGGGCAGATGTTGCCAATAGATGAGATATTCGAGCGGGGCTTGGATCTCAAACAAGTGGATGGATTGCCA

[0275] Cfa-N蛋白:

[0276] CLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCLEDGSIIRATKDHKFMTTDGQMLPIDEIFERGLDLKQVDGLP

[0277] Cfa-C DNA:

[0278] ATGAAGAGGACTGCCGATGGATCAGAGTTTGAATCTCCCAAGAAGAAGAGGAAAGTAAAGATAATATCTCGAAAAAGTCTTGGTACCCAAAATGTCTATGATATTGGAGTGGAGAAAGATCACAACTTCCTTCTCAAGAACGGTCTCGTAGCCAGCAAC

[0279] Cfa-C蛋白:MKRTADGSEFESPKKKRKVKIISRKSLGTQNVYDIGVEKDHNFLLKNGLVASN

[0280] 内含肽-N和内含肽-C可以分别融合至分体Cas9的N端部分和分体Cas9的C端部分,用于将分体Cas9的N端部分与分体Cas9的C端部分接合。例如,在一些实施方案中,内含肽-N融合至分体Cas9的N端部分的C端,即,以形成N--[分体Cas9的N端部分]-[内含肽-N]--C的结构。在一些实施方案中,内含肽-C融合至分体Cas9的C端部分的N端,即,以形成N--[内含肽-C]-[分体Cas9的C端部分]--C的结构。用于接合内含肽所融合至的蛋白质(例如,分体Cas9)的内含肽介导的蛋白质剪接机制是本领域中已知的,例如,如Shah等人,Chem Sci.2014;5(1):446-461中所述,该文献通过引用并入本文。用于设计和使用内含肽的方法是本领域中已知的并且由例如WO2014004336、WO2017132580、US20150344549和US20180127780描述,这些专利各自通过引用而以其整体并入本文。

[0281] 术语“分离的”、“纯化的”或“生物学纯的”是指在不同程度上不含通常伴随它的成分(如在其原生状态下所见)的材料。“分离”表示与原始来源或周围物质分隔的程度。“纯化”表示高于分离的分割程度。“纯化的”或“生物学纯的”蛋白质足够不含其他材料,使得任何杂质在材料层面不影响该蛋白质的生物学性质或造成其他负面后果。换句话说,当通过重组DNA技术生产时,如果本发明的核酸或肽基本上不含细胞材料、病毒材料或培养基,则它是纯化的;或者当化学合成时,如果不含化学前体或其他化学物质,则它是纯化的。典型使用分析化学技术例如聚丙烯酰胺凝胶电泳或高效液相色谱测定纯度及同质性。术语“纯化的”可表示核酸或蛋白质在电泳凝胶中产生实质上的一个带。对于可进行修饰例如磷酸化或糖基化的蛋白质,不同的修饰可产生不同的分离蛋白质,这些蛋白质可独立纯化。

[0282] “分离的多核苷酸”意为一种核酸(例如,DNA),其不含在本发明核酸分子所来源的有机体的天然基因组中位于该基因侧翼的基因。因此该术语包括,举例而言,重组DNA,其被并入载体内、自主复制质粒或病毒内、或并入原核生物或真核生物的基因组DNA内;或作为独立于其他序列的单独分子存在(例如,通过PCR或限制性核酸内切酶消化产生的cDNA或基因组或cDNA片段)。此外,该术语包括从DNA分子转录的RNA分子,以及作为编码附加多肽序列的杂交基因的一部分的重组DNA。

[0283] “分离的多肽”意为已经与其天然伴随组分分隔的本发明的多肽。典型地,当多肽不含至少60重量%的天然与其相关的蛋白质和天然有机分子。优选地,该制剂是至少75重量%、更优选至少90重量%、且最优选至少99重量%的本发明的多肽。本发明的分离多肽可以例如通过从天然来源萃取、通过表达编码此多肽的重组核酸、或通过化学合成该蛋白质而获得。可通过任何适宜的方法,例如柱色谱、聚丙烯酰胺凝胶电泳或HPLC分析,来测量纯度。

[0284] “前导肽“意为一种短氨基酸序列(例如,大约16至30个氨基酸的长度),其将新合成的分泌蛋白或膜蛋白引导至膜并穿过膜(例如,内质网膜)。前导肽典型定位在多肽的N端,并且可以在该多肽已经跨过膜后通过信号肽酶移除。前导肽序列典型含有三个共有结构特征:N端极性碱性区(n区)、疏水性核心和亲水性c区。在一些实施方案中,本发明的CAR包括前导肽序列(例如,抗原结合结构域的N端)。示例性前导肽氨基酸序列是:

METDTLLLWVLLLWVPGSTG。

[0285] 如本文中所用,术语“连接子”是指的键(例如,共价键)、化学基团或分子,其链接两个分子或两个部分,例如,蛋白复合物或核糖核酸复合物的两个组分或者融合蛋白的两个结构域,例如,多核苷酸可编程DNA结合结构域(例如,dCas9)和脱氨酶结构域(例如,腺苷脱氨酶、胞苷脱氨酶)。连接子可以接合碱基编辑器系统的不同组分或多个组分的不同部分。例如,在一些实施方案中,连接子可以接合多核苷酸可编程核苷酸结合结构域的向导多核苷酸结合结构域与脱氨酶的催化性结构域。在一些实施方案中,连接子可以接合CRISPR多肽与脱氨酶。在一些实施方案中,连接子可以接合Cas9与脱氨酶。在一些实施方案中,连接子可以接合dCas9与脱氨酶。在一些实施方案中,连接子可以接合nCas9与脱氨酶。在一些实施方案中,连接子可以接合向导多核苷酸与脱氨酶。在一些实施方案中,连接子可以接合碱基编辑器系统的脱氨基组分与多核苷酸可编程核苷酸结合组分。在一些实施方案中,连接子可以接合碱基编辑器系统的脱氨基组分的RNA结合部分与多核苷酸可编程核苷酸结合组分。在一些实施方案中,连接子可以接合碱基编辑器系统的脱氨基组分的RNA结合部分与多核苷酸可编程核苷酸结合组分的RNA结合部分。连接子可以位于两个基团、分子或其他部分之间或连接子侧翼具有两个基团、分子或其他部分,并且彼此经由共价键或非共价相互作用联接,从而将两者联接。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,连接子可包含能够结合至配体的适体。在一些实施方案中,配体可以是碳水化合物、肽、蛋白质或核酸。在一些实施方案中,连接子可包含适体,该适体可源自核糖开关。适体所源自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苷前体1(PreQ1)核糖开关。在一些实施方案中,连接子可包含结合至多肽或蛋白质结构域诸如多肽配体的适体。在一些实施方案中,多肽配体可以是K同源(KH)结构域、MS2外壳蛋白结构域、PP7外壳蛋白结构域、SfMu 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。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含脱氨酶结构域和RNA识别基序。

[0286] 在一些实施方案中,连接子可以是一个氨基酸或多个氨基酸(例如,肽或蛋白质)。在一些实施方案中,连接子可以是约5至100个氨基酸的长度,例如,约5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、20至30、30至40、40至50、50至60、60至70、70至80、80至90、或90至100个氨基酸的长度。在一些实施方案中,连接子可以是约100至150、150至200、200至250、250至300、300至350、350至400、400至450、或450至500个氨基酸的长度。也可以考虑更长或更短的连接子。

[0287] 在一些实施方案中,连接子接合RNA可编程核酸酶的gRNA结合结构域(包括Cas9核酸酶结构域)与核酸编辑蛋白的催化性结构域(例如,胞苷或腺苷脱氨酶)。在一些实施方案中,连接子接合dCas9与核酸编辑蛋白。例如,连接子位于两个基团、分子或其他部分之间或连接子侧翼具有两个基团、分子或其他部分,并且彼此经由共价键联接,从而将两者联接。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,连接子是5至200个氨基酸的长度,例如,5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、35、45、50、55、60、60、65、70、70、75、80、85、90、90、95、100、101、102、103、104、105、110、120、130、140、150、160、175、180、190或200个氨基酸的长度。也考虑更长或更短的连接子。在一些实施方案中,连接子包含SGSETPGTSESATPES的氨基酸序列,也可以称为XTEN连接子。在一些实施方案中,连接子包含氨基酸序列SGGS。在一些实施方案中,连接子包含(SGGS)n、(GGGS)n、(GGGGS)n、(G)n、(EAAAK)n、(GGS)n、SGSETPGTSESATPES或(XP)n基序,或这些中任意者的组合,其中n独立地是1到30之间的整数,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。在一些实施方案中,连接子包含多个脯氨酸残基并且是5至21、5至14、5至9、5至7个氨基酸的长度,例如,PAPAP、PAPAPA、PAPAPAP、PAPAPAPA、P(AP)4、P(AP)7、P(AP)10。此类富含脯氨酸的连接子也称为“刚性”连接子。

[0288] 在一些实施方案中,嵌合抗原受体包含至少一个连接子。该至少一个连接子将可变重(VH)区接合或连接至嵌合抗原受体的细胞外结合结构域的恒定重(CH)区。连接子也可以将可变轻(VL)区连接至该细胞外结结合结构域的的可变恒定(VC)区。

[0289] 在一些实施方案中,碱基编辑器的结构域经由连接子融合,该连接子包含以下氨基酸序列:

[0290] SGGSSGSETPGTSESATPESSGGS、SGGSSGGSSGSETPGTSESATPESSGGSSGGS或GGSGGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSGGS。

[0291] 在一些实施方案中,碱基编辑器的结构域经由连接子融合,该连接子包含SGSETPGTSESATPES的氨基酸序列,也可以称为XTEN连接子。在一些实施方案中,连接子包含氨基酸序列SGGS。在一些实施方案中,连接子是24个氨基酸的长度。在一些实施方案中,连接子包含氨基酸序列SGGSSGGSSGSETPGTSESATPES。在一些实施方案中,连接子是40个氨基酸的长度。在一些实施方案中,连接子包含氨基酸序列SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGS。在一些实施方案中,连接子是64个氨基酸的长度。在一些实施方案中,连接子包含氨基酸序列SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGSSGSETPGTSESATPESSGGSSGGS。在一些实施方案中,连接子是92个氨基酸的长度。在一些实施方案中,连接子包含氨基酸序列PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATS。

[0292] 如本文所用,术语“液态癌症”是指存在于体液例如血液、淋巴和骨髓中的癌细胞。液态癌症包括但不限于,白血病、骨髓瘤和液态淋巴瘤。如本文所用,液态癌症不包括实体瘤诸如肉瘤和癌,或不含囊肿或液态区域的实体淋巴瘤。“液态癌症”可以是复发性的、难治性的或转移性的。待使用本文所述方法治疗的液态癌症可以是,例如,也爱淋巴瘤;液态淋巴瘤包括含有囊肿或液态区域的淋巴瘤。

[0293] “淋巴细胞激活基因3(LAG-3)多肽”意为一种蛋白质,其与NCBI登录号NP_002277.4或其片段具有至少约85%的氨基酸序列一致性并且具有免疫调节活性。示例性氨基酸序列提供如下。

[0294] >NP_002277.4淋巴细胞激活基因3蛋白质前体[智人(Homo sapiens)]

```
          1 mweaqflgll flqplwvapv kplqpgaevp vvwaqegapa qlpcsptipl
qdlsllrrag
         61   vtwqhqpdsg   ppaaapghpl   apgphpaaps   swgprprryt
vlsvgpgglr sgrlplqprv
        121   qldergrqrg   dfslwlrpar   radageyraa   vhlrdralsc
rlrlrlgqas mtasppgslr
        181   asdwvilncs   fsrpdrpasv   hwfrnrgqgr   vpvresphhh
laesflflpq vspmdsgpwg
[0295]  241   ciltyrdgfn   vsimynltvl   glepptpltv   yagagsrvgl
pcrlpagvgt rsfltakwtp
        301   pgggpdllvt   gdngdftlrl   edvsqaqagt   ytchihlqeq
qlnatvtlai itvtpksfgs
        361   pgslgkllce   vtpvsgqerf   vwssldtpsq   rsfsgpwlea
qeaqllsqpw qcqlyqgerl
        421   lgaavyftel   sspgaqrsgr   apgalpaghl   llflilgvls
llllvtgafg fhlwrrqwrp
        481 rrfsaleqgi hppqaqskie eleqepepep epepepepep epeql
```

[0296] “淋巴细胞激活基因3(LAG-3)多核苷酸”意为编码LAG-3多肽的核酸。示例性LAG-3核酸序列提供如下。

[0297] >NM_002286.6智人(Homo sapiens)淋巴细胞激活3(LAG3),mRNA

```
          1 agagaccagc agaacggcat cccagccacg acggccactt tgctctgtct
gctctccgcc
         61   acggccctgc   tctgttccct   gggacacccc   cgcccccacc
tcctcaggct gcctgatctg
        121   cccagctttc   cagctttcct   ctggattccg   gcctctggtc
atccctcccc accctctctc
        181   caaggccctc   tcctggtctc   ccttcttcta   gaacccttc
[0298]  ctccacctcc ctctctgcag
        241   aacttctcct   ttacccccca   ccccccacca   ctgccccctt
tccttttctg acctcctttt
        301   ggagggctca   gcgctgccca   gaccatagga   gagatgtggg
aggctcagtt cctgggcttg
        361   ctgtttctgc   agccgctttg   ggtggctcca   gtgaagcctc
tccagccagg ggctgaggtc
        421   ccggtggtgt   gggcccagga   gggggctcct   gcccagctcc
```

```
cctgcagccc cacaatcccc
     481   ctccaggatc   tcagccttct   gcgaagagca   ggggtcactt
ggcagcatca gccagacagt
     541   ggcccgcccg   ctgccgcccc   cggccatccc   ctggcccccg
gccctcaccc ggcggcgccc
     601   tcctcctggg   ggcccaggcc   ccgccgctac   acggtgctga
gcgtgggtcc cggaggcctg
     661   cgcagcggga   ggctgccccc   gcagccccgc   gtccagctgg
atgagcgcgg ccggcagcgc
     721   ggggacttct   cgctatggct   gcgcccagcc   cggcgcgcgg
acgccggcga gtaccgcgcc
     781   gcggtgcacc   tcagggaccg   cgccctctcc   tgccgcctcc
gtctgcgcct gggccaggcc
     841   tcgatgactg   ccagcccccc   aggatctctc   agagcctccg
actgggtcat tttgaactgc
     901   tccttcagcc   gccctgaccg   cccagcctct   gtgcattggt
tccggaaccg gggccagggc
     961   cgagtccctg   tccgggagtc   cccccatcac   cacttagcgg
aaagcttcct cttcctgccc
    1021   caagtcagcc   ccatggactc   tgggccctgg   ggctgcatcc
tcacctacag agatggcttc
    1081   aacgtctcca   tcatgtataa   cctcactgtt   ctgggtctgg
agcccccaac tcccttgaca
    1141   gtgtacgctg   gagcaggttc   cagggtgggg   ctgccctgcc
gcctgcctgc tggtgtgggg
    1201   acccggtctt   tcctcactgc   caagtggact   cctcctgggg
gaggccctga cctcctggtg
    1261   actggagaca   atggcgactt   taccccttcga   ctagaggatg
tgagccaggc ccaggctggg
    1321   acctacacct   gccatatcca   tctgcaggaa   cagcagctca
atgccactgt cacattggca
    1381   atcatcacag   tgactcccaa   atcctttggg   tcacctggat
ccctggggaa gctgctttgt
    1441   gaggtgactc   cagtatctgg   acaagaacgc   tttgtgtgga
gctctctgga caccccatcc
    1501   cagaggagtt   tctcaggacc   ttggctggag   gcacaggagg
cccagctcct ttcccagcct
    1561   tggcaatgcc   agctgtacca   gggggagagg   cttcttggag
cagcagtgta cttcacagag
    1621   ctgtctagcc   caggtgccca   acgctctggg   agagccccag
gtgccctccc agcaggccac
    1681   ctcctgctgt   ttctcatcct   tggtgtcctt   tctctgctcc
ttttggtgac tggagccttt
    1741   ggctttcacc   tttggagaag   acagtggcga   ccaagacgat
tttctgcctt agagcaaggg
    1801   attcaccctc   cgcaggctca   gagcaagata   gaggagctgg
agcaagaacc ggagccggag
    1861   ccggagccgg   aaccggagcc   cgagcccgag   cccgagccgg
agcagctctg acctggagct
    1921   gaggcagcca   gcagatctca   gcagcccagt   ccaaataaac
tccctgtcag cagcaa
```

[0299]

[0300] “标志物”意为在表达水平或活性上具有与疾病或病变相关的改变的任何蛋白质或多核苷酸。

[0301] 如本文所用,术语“突变”是指序列(例如,核酸或氨基酸序列)内的残基被置换为另一残基,或序列内一个或多个残基的缺失或插入。本文中,突变通常描述为:原始残基的身份,然后是该残基在序列中的位置,然后是新置换的残基的身份。各种用于作出本文所提供的氨基酸置换(突变)的方法是本领域中周知的,并且由例如《分子克隆:实验室手册(第四版)》(Green and Sambrook,Molecular Cloning:ALaboratory Manual(4th ed.,Cold Spring Harbor Laboratory Press,Cold Spring Harbor,N.Y.(2012)))提供。在一些实施方案中,本文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)中生成“期望突变”诸如点突变,而不生成大量的非期望突变诸如非期望的点突变。在一些实施方案中,期望突变是通过结合至向导多核苷酸(例如,gRNA)的经专门设计以生成期望突变的特异性碱基编辑器(例如,胞苷碱基编辑器或腺苷碱基编辑器)生成的突变。

[0302] 通常,在序列(例如,本文所述的氨基酸序列)中作出或鉴别的突变相对于参考(或野生型)序列(即,不含有该突变的序列)编号。本领域技术人员将会轻易地理解如何相对于参考序列确定氨基酸和核酸序列中的突变的位置。

[0303] “瘤形成”是指表现出异常生长或增值的细胞或组织。术语瘤形成涵盖癌症、液态和实体肿瘤。在一些实施方案中,瘤形成是实体肿瘤。在其他实施方案中,瘤形成是液态肿瘤。在一些实施方案中,瘤形成是血液学癌症。在一些实施方案中,血液学癌症是白血病、骨髓瘤和/或淋巴瘤。在一些实施方案中,血液学癌症是B细胞癌症。在一些实施方案中,B细胞癌症是淋巴瘤或白血病。在一些情况下,白血病包括白血病前期。在一些情况下,白血病是急性白血病。急性白血病包括,例如,急性骨髓性白血病(AML)。急性白血病也包括,例如,急性淋巴性白血病或急性淋巴细胞性白血病(ALL);ALL包括B谱系ALL、T谱系ALL和T细胞急性淋巴细胞性白血病(T-ALL)。

[0304] 瘤形成的非限制性示例包括T细胞急性成淋巴细胞性白血病(T-ALL)、蕈样肉芽肿(MF)、Sézary综合征(SS)、外周T/NK细胞淋巴瘤、间变性大细胞淋巴瘤ALK+、原发性皮肤T细胞淋巴瘤、T细胞大颗粒淋巴细胞型白血病、血管免疫母细胞性T/NK细胞淋巴瘤、肝脾T细胞淋巴瘤、原发性皮肤CD30+淋巴增生性疾病、结外NK/T细胞淋巴瘤、成人T细胞白血病/淋巴瘤、T细胞幼淋巴细胞白血病、皮下脂膜炎样T细胞淋巴瘤、原发性皮肤γδT细胞淋巴瘤、侵袭性NK细胞白血病和肠病相关T细胞淋巴瘤。在一些实施方案中,瘤形成是T细胞急性成淋巴细胞性白血病(T-ALL)。在一些实施方案中,瘤形成是T细胞急性骨髓性白血病(AML)。

[0305] “激活的T细胞核因子1(NFATc1)多肽”意为一种蛋白质,其与NCBI登录号NM_172390.2或其片段具有至少约85%的序列一致性,并且是激活的T细胞DNA结合转录复合体的组分。示例性氨基酸序列提供如下。

[0306] >NP_765978.1激活的T细胞核因子,细胞质1亚型A[智人(Homo sapiens)]

[0307] MPSTSFPVPSKFPLGPAAAVFGRGETLGPAPRAGGTMKSAEEEHYGYASSNVSPALPLPTAHSTLPAPC
HNLQTSTPGIIPPADHPSGYGAALDGGPAGYFLSSGHTRPDGAPALESPRIEITSCLGLYHNNNQFFHDVEVEDVLP
SSKRSPSTATLSLPSLEAYRDPSCLSPASSLSSRSCNSEASSYESNYSYPYASPQTSPWQSPCVSPKTTDPEEGFPR
GLGACTLLGSPRHSPSTSPRASVTEESWLGARSSRPASPCNKRKYSLNGRQPPYSPHHSPTPSPHGSPRVSVTDDSW
LGNTTQYTSSAIVAAINALTTDSSLDLGDGVPVKSRKTTLEQPPSVALKVEPVGEDLGSPPPPADFAPEDYSSFQHI
RKGGFCDQYLAVPQHPYQWAKPKPLSPTSYMSPTLPALDWQLPSHSGPYELRIEVQPKSHHRAHYETEGSRGAVKAS
AGGHPIVQLHGYLENEPLMLQLFIGTADDRLLRPHAFYQVHRITGKTVSTTSHEAILSNTKVLEIPLLPENSMRAVI

DCAGILKLRNSDIELRKGETDIGRKNTRVRLVFRVHVPQPSGRTLSLQVASNPIECSQRSAQELPLVEKQSTDSYPV
VGGKKMVLSGHNFLQDSKVIFVEKAPDGHHVWEMEAKTDRDLCKPNSLVVEIPPFRNQRITSPVHVSFYVCNGKRKR
SQYQRFTYLPANGNAIFLTVSREHERVGCFF

[0308] “激活的T细胞核因子1(NFATc1)多核苷酸”意为编码NFATc1多肽的核酸。NFATc1基因编码一种蛋白质,该蛋白质牵涉到细胞因子基因尤其是IL-2和IL-4在T细胞中的诱导型表达。示例性核酸序列提供如下。

[0309] >NM_172390.2智人(Homo sapiens)激活的T细胞核因子1(NFATC1),转录物变体1,mRNA

[0310] GGCGGGCGCTCGGCGACTCGTCCCCGGGGCCCCGCGCGGGCCCGGGCAGCAGGGGCGTGATGTCACGG
CAGGGAGGGGGCGCGGGAGCCGCCGGGCCGGCGGGGAGGCGGGGGAGGTGTTTTCCAGCTTTAAAAAGGCAGGAGG
CAGAGCGCGGCCCTGCGTCAGAGCGAGACTCAGAGGCTCCGAACTCGCCGGCGGAGTCGCCGCGCCAGATCCCAGC
AGCAGGGCGCGGGCACCGGGGCGCGGGCAGGGCTCGGAGCCACCGCGCAGGTCCTAGGGCCGCGGCCGGGCCCCGC
CACGCGCGCACACGCCCCTCGATGACTTTCCTCCGGGGCGCGCGGCGCTGAGCCCGGGGCGAGGGCTGTCTTCCCG
GAGACCCGACCCCGGCAGCGCGGGGCGGCCGCTTCTCCTGTGCCTCCGCCCGCCGCTCCACTCCCCGCCGCCGCCG
CGCGGATGCCAAGCACCAGCTTTCCAGTCCCTTCCAAGTTTCCACTTGGCCCTGCGGCTGCGGTCTTCGGGAGAGG
AGAAACTTTGGGGCCCGCGCCGCGCGCCGGCGGCACCATGAAGTCAGCGGAGGAAGAACACTATGGCTATGCATCC
TCCAACGTCAGCCCCGCCCTGCCGCTCCCCACGGCGCACTCCACCCTGCCGGCCCCGTGCCACAACCTTCAGACCT
CCACACCGGGCATCATCCCGCCGGCGGATCACCCCTCGGGGTACGGAGCAGCTTTGGACGGTGGGCCCGCGGGCTA
CTTCCTCTCCTCCGGCCACACCAGGCCTGATGGGGCCCCTGCCCTGGAGAGTCCTCGCATCGAGATAACCTCGTGC
TTGGGCCTGTACCACAACAATAACCAGTTTTTCCACGATGTGGAGGTGGAAGACGTCCTCCCTAGCTCCAAACGGT
CCCCCTCCACGGCCACGCTGAGTCTGCCCAGCCTGGAGGCCTACAGAGACCCCTCGTGCCTGAGCCCGGCCAGCAG
CCTGTCCTCCCGGAGCTGCAACTCAGAGGCCTCCTCCTACGAGTCCAACTACTCGTACCCGTACGCGTCCCCCCAG
ACGTCGCCATGGCAGTCTCCCTGCGTGTCTCCCAAGACCACGGACCCCGAGGAGGGCTTTCCCCGCGGGCTGGGGG
CCTGCACACTGCTGGGTTCCCCGCGGCACTCCCCCTCCACCTCGCCCCGCGCCAGCGTCACTGAGGAGAGCTGGCT
GGGTGCCCGCTCCTCCAGACCCGCGTCCCCTTGCAACAAGAGGAAGTACAGCCTCAACGGCCGGCAGCCGCCCTAC
TCACCCCACCACTCGCCCACGCCGTCCCCGCACGGCTCCCCGCGGGTCAGCGTGACCGACGACTCGTGGTTGGGCA
ACACCACCCAGTACACCAGCTCGGCCATCGTGGCCGCCATCAACGCGCTGACCACCGACAGCAGCCTGGACCTGGG
AGATGGCGTCCCTGTCAAGTCCCGCAAGACCACCCTGGAGCAGCCGCCCTCAGTGGCGCTCAAGGTGGAGCCCGTC
GGGGAGGACCTGGGCAGCCCCCCGCCCCCGGCCGACTTCGCGCCCGAAGACTACTCCTCTTTCCAGCACATCAGGA
AGGGCGGCTTCTGCGACCAGTACCTGGCGGTGCCGCAGCACCCCTACCAGTGGGCGAAGCCCAAGCCCCTGTCCCC
TACGTCCTACATGAGCCCGACCCTGCCCGCCCTGGACTGGCAGCTGCCGTCCCACTCAGGCCCGTATGAGCTTCGG
ATTGAGGTGCAGCCCAAGTCCCACCACCGAGCCCACTACGAGACGGAGGGCAGCCGGGGGGCCGTGAAGGCGTCGG
CCGGAGGACACCCCATCGTGCAGCTGCATGGCTACTTGGAGAATGAGCCGCTGATGCTGCAGCTTTTCATTGGGAC
GGCGGACGACCGCCTGCTGCGCCCGCACGCCTTCTACCAGGTGCACCGCATCACAGGGAAGACCGTGTCCACCACC
AGCCACGAGGCCATCCTCTCCAACACCAAAGTCCTGGAGATCCCACTCCTGCCGGAGAACAGCATGCGAGCCGTCA
TTGACTGTGCCGGAATCCTGAAACTCAGAAACTCCGACATTGAACTTCGGAAAGGAGAGACGGACATCGGGAGGAA
GAACACACGGGTACGGCTGGTGTTCCGCGTTCACGTCCCGCAACCCAGCGGCCGCACGCTGTCCCTGCAGGTGGCC
TCCAACCCCATCGAATGCTCCCAGCGCTCAGCTCAGGAGCTGCCTCTGGTGGAGAAGCAGAGCACGGACAGCTATC
CGGTCGTGGGCGGGAAGAAGATGGTCCTGTCTGGCCACAACTTCCTGCAGGACTCCAAGGTCATTTTCGTGGAGAA

AGCCCCAGATGGCCACCATGTCTGGGAGATGGAAGCGAAAACTGACCGGGACCTGTGCAAGCCGAATTCTCTGGTG
GTTGAGATCCCGCCATTTCGGAATCAGAGGATAACCAGCCCCGTTCACGTCAGTTTCTACGTCTGCAACGGGAAGA
GAAAGCGAAGCCAGTACCAGCGTTTCACCTACCTTCCCGCCAACGGTAACGCCATCTTTCTAACCGTAAGCCGTGA
ACATGAGCGCGTGGGGTGCTTTTTCTAAAGACGCAGAAACGACGTCGCCGTAAAGCAGCGTGGCGTGTTGCACATT
TAACTGTGTGATGTCCCGTTAGTGAGACCGAGCCATCGATGCCCTGAAAAGGAAAGGAAAAGGGAAGCTTCGGATG
CATTTTCCTTGATCCCTGTTGGGGGTGGGGGGCGGGGGTTGCATACTCAGATAGTCACGGTTATTTTGCTTCTTGC
GAATGTATAACAGCCAAGGGGAAAACATGGCTCTTCTGCTCCAAAAAACTGAGGGGGTCCTGGTGTGCATTTGCAC
CCTAAAGCTGCTTACGGTGAAAAGGCAAATAGGTATAGCTATTTTGCAGGCACCTTTAGG AATAAACTTTGCTTT
TAAGCCTGTAAAAAAAAAAAAAA

[0311] 术语“非保守突变”包括不同组之间的氨基酸置换,例如,赖氨酸替换色氨酸,或苯丙氨酸替换丝氨酸等。在这种情况下,优选非保守氨基酸置换不干扰或抑制功能性变体的生物学活性。非保守氨基酸置换可增强功能性变体的生物学活性,使得该功能性变体的生物学活性相较于野生型蛋白质有所增加。

[0312] 术语“核定位序列”、“核定位信号”或“NLS”是指一种氨基酸序列,其促进蛋白质进入细胞核。核定位序列是本领域中已知的,并且在例如2000年11月23日提交的Plank等人的国际PCT申请PCT/EP2000/011690并且在2001年5月31日公开的WO/2001/038547中有所描述,其内容通过引用其关于示例性核定位序列的公开内容而并入本文。在其他实施方案中,NLS是由例如Koblan等人,Nature Biotech.2018doi:10.1038/nbt.4172描述的优化NLS。在一些实施方案中,NLS包含氨基酸序列PKKKRKVEGADKRTADGSEFESPKKKRKV、KRTADGSEFESPKKKRKV、KRPAATKKAGQAKKKK、KKTELQTTNAENKTKKL、KRGINDRNFWRGENGRKTR、RKSGKIAAIVVKRPRK、PKKKRKV或MDSLLMNRRKFLYQFKNVRWAKGRRETYLC。

[0313] 如本文所用,术语“核酸”和“核酸分子”是指一种化合物,其包含核碱基和酸性部分,例如,核苷、核苷酸或核苷酸的聚合物。典型地,聚合性核酸,例如包含三个或更多个核苷酸的核酸分子,是线性分子,在该分子中,相邻核苷酸经由磷酸二酯链结链接至彼此。在一些实施方案中,“核酸”是指单个核酸残基(例如,核苷酸和/或核苷)。在一些实施方案中,“核酸”是指寡核苷酸链,其包含三个或更多个单个核苷酸残基。如本文所用,术语“寡核苷酸”和“多核苷酸”可互换使用,是指核苷酸的聚合物(例如,一串至少三个核苷酸)。在一些实施方案中,“核酸”涵盖RNA以及单链和/或双链DNA。核酸可以是天然存在的,例如,在转录物、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基因组的情境中。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、髯公染色体、工程改造的基因组、或其片段,或者合成的DNA、RNA、DNA/RNA杂交物,或者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,具有除磷酸二酯主链之外的类似物。核酸可以从天然来源纯化,使用重组表达系统生产并任选地纯化,化学合成等。若适用,例如,在化学合成分子的情况下,核酸可包含核苷类似物诸如具有经化学修饰的碱基或糖的类似物以及主链修饰。除非另外指明,否则核酸序列以5’到3’方向呈现。在一些实施方案中,核酸是或包含天然核苷(例如,腺苷、胸苷、鸟苷、胞苷、尿苷、脱氧腺苷、脱氧胸苷、脱氧鸟苷和脱氧胞苷);核苷类似物(例如,2-氨基腺苷、2-巯基胸苷、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-去氮

腺苷、7-去氮鸟苷、8-氧代腺苷、8-氧代鸟苷、O(6)-甲基鸟苷和2-巯基胞苷);经化学修饰的碱基;经生物修饰的碱基(例如,甲基化碱基);插置碱基;经修饰的糖(2'-例如氟核糖、核糖、2'-脱氧核糖、阿拉伯糖和己糖);和/或经修饰的磷酸酯基团(例如,硫代磷酸酯和5'-N-亚磷酰胺键结)。术语"核酸可编程DNA结合蛋白"或"napDNAbp"可以与"多核苷酸可编程核苷酸结合结构域"互换使用,是指与核酸(例如,DNA或RNA)诸如向导核酸或向导多核苷酸(例如,gRNA)缔合的蛋白质,该核酸将napDNAbp引导至特异性核酸序列。在一些实施方案中,多核苷酸可编程核苷酸结合结构域是多核苷酸可编程DNA结合结构域。在一些实施方案中,多核苷酸可编程核苷酸结合结构域是多核苷酸可编程RNA结合结构域。在一些实施方案中,多核苷酸可编程核苷酸结合结构域是Cas9蛋白。Cas9蛋白可与向导RNA缔合,该向导RNA将Cas9蛋白引导至与该向导RNA互补的特异性DNA序列。在一些实施方案中,napDNAbp是Cas9结构域,例如,核酸酶活性的Cas9、Cas9切口酶(nCas9)、或无核酸酶活性的Cas9(dCas9)。核酸可编程DNA结合蛋白的非限制性示例包括Cas9(例如,dCas9和nCas9)、Cas12a/Cpf1、Cas12b/C2c1、Cas12c/C2c3、Cas12d/CasY、Cas12e/CasX、Cas12g、Cas12h、Cas12i和Cas12j/CasΦ。Cas酶的非限制性示例包括Cas1、Cas1B、Cas2、Cas3、Cas4、Cas5、Cas5d、Cas5t、Cas5h、Cas5a、Cas6、Cas7、Cas8、Cas8a、Cas8b、Cas8c、Cas9(也称为Csn1或Csx12)、Cas10、Cas10d、Cas12a/Cpf1、Cas12b/C2c1、Cas12c/C2c3、Cas12d/CasY、Cas12e/CasX、Cas12g、Cas12h、Cas12i、Cas12j/CasΦ、Csy1、Csy2、Csy3、Csy4、Cse1、Cse2、Cse3、Cse4、Cse5e、Csc1、Csc2、Csa5、Csn1、Csn2、Csm1、Csm2、Csm3、Csm4、Csm5、Csm6、Cmr1、Cmr3、Cmr4、Cmr5、Cmr6、Csb1、Csb2、Csb3、Csx17、Csx14、Csx10、Csx16、CsaX、Csx3、Csx1、Csx1S、Csx11、Csf1、Csf2、Cs0、Csf4、Csd1、Csd2、Cst1、Cst2、Csh1、Csh2、Csa1、Csa2、Csa3、Csa4、Csa5、II型Cas效应蛋白、V型Cas效应蛋白、VI型Cas效应蛋白、CARF、DinG、其同源、或其经修饰或工程改造的版本。其他核酸可编程DNA结合蛋白也处于本公开范畴内,但它们可以未在本公开中具体列出。参见,例如,Makarova等人"Classification and Nomenclature of CRISPR-Cas Systems:Where from Here?"CRISPR J.2018Oct;1:325-336.doi:10.1089/crispr.2018.0033;Yan等人,"Functionally diverse type V CRISPR-Cas systems"Science.2019Jan 4;363(6422):88-91.doi:10.1126/science.aav7271,各自的整体内容通过引用并入本文。

[0314] 本文中可互换使用的术语"核碱基"、"含氮碱基"或"碱基"是指含氮的生物化合物,其形成核苷,核苷进而作为核苷酸的组分。核碱基形成碱基对的能力和一个堆栈在另一个上的能力直接导致长链螺旋结构,诸如核糖核酸(RNA)和脱氧核糖核酸(DNA)。五种核碱基,即腺嘌呤(A)、胞嘧啶(C)、鸟嘌呤(G)、胸腺嘧啶(T)和尿嘧啶(U),称为主要核碱基或标准核碱基。腺嘌呤和鸟嘌呤衍生自嘌呤,而胞嘧啶、尿嘧啶和胸腺嘧啶衍生自嘧啶。DNA和RNA也可含有经修饰的其他(非主要)碱基。非限制性的示例性经修饰的核碱基可包括次黄嘌呤、黄嘌呤、7-甲基鸟嘌呤、5,6-二氢尿嘧啶、5-甲基胞嘧啶(m5C)和5-羟甲基胞嘧啶。次黄嘌呤和黄嘌呤两者均可通过诱变剂存在下的脱氨基(氨基替换为羰基)而产生。次黄嘌呤可自腺嘌呤修饰而来。黄嘌呤可在鸟嘌呤修饰而来。尿嘧啶可以来自于胞嘧啶的脱氨基反应。"核苷"由核碱基和五碳糖(核糖或脱氧核糖)组成。核苷的示例包括腺苷、鸟苷、尿苷、胞苷、5-甲基尿苷(m5U)、脱氧腺苷、脱氧鸟苷、胸苷、脱氧尿苷和脱氧胞苷。具有经修饰的核碱基的核苷的示例包括肌苷(I)、黄苷(X)、7-甲基鸟苷(m7G)、二氢尿苷(D)、5-甲基胞苷(m5C)

和假尿苷(Ψ)。“核苷酸”由核碱基、五碳糖(核糖或脱氧核糖)和至少一个磷酸酯基团组成。

[0315] 如本文所用,术语“核碱基编辑结构域”或“核碱基编辑蛋白”是指一种蛋白质或酶,其催化RNA或DNA中的核碱基修饰,诸如胞嘧啶(或胞苷)到尿嘧啶(或尿苷)或胸腺嘧啶(或胸苷),和腺嘌呤(或腺苷)到次黄嘌呤(或肌苷)脱氨基,以及非模板化核苷酸加成和插入。在一些实施方案中,核碱基编辑结构域是脱氨酶结构域(例如,腺嘌呤脱氨酶或腺苷脱氨酶;或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从天然存在的核碱基编辑结构域工程改造的或进化的核碱基编辑结构域。核碱基编辑结构域可以来自任何生物体,诸如细菌、人类、黑猩猩、大猩猩、猴、牛、狗、大鼠或小鼠。

[0316] 如本文所用,“获得试剂”中的“获得”包括合成、购买或其它获取该试剂的途径。如本文所用,“患者”或“受试者”是指被诊断为患有或易患或易发展出疾病或疾患或者处于患有或发展出疾病或疾患风险下的哺乳动物受试者或个体。在一些实施方案中,术语“患者”是指具有高于平均值的发展出疾病或疾患的可能性的哺乳动物受试者。示例性患者可以是人类、非人类灵长动物、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、沙鼠或豚鼠)和其他可能受益于本文所公开的疗法的哺乳动物。示例性人类患者可以是男性和/或女性。

[0317] 本文中,“有此需要的患者”或“有此需要的受试者”是指被诊断为患有、患有、预期患有或易患疾病或疾患(例如,T或NK细胞恶性病变)或者处于该疾病或疾患风险下的患者。

[0318] 术语“致病突变”、“致病变体”、“造成疾病的突变”、“造成疾病的变体”、“有害突变”或“诱因性突变”是指一种基因改变或突变,其增加个体对某种疾病或疾患的易感性或倾向性。在一些实施方案中,致病突变在由基因所编码的蛋白质中包含至少一个野生型氨基酸被置换为至少一个致病氨基酸。

[0319] 术语“药学上可接受的载体”意为药学上可接受的材料、组合物或媒介物,诸如液体或固体填料、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,牵涉到将化合物从身体的一个位点(例如,递送位点)携带或运输至另一位点(例如,器官、组织或身体的部分)中。就与制剂的其他成分相容且不损伤受试者组织(例如,生理上相容的、无菌的、生理pH等)意义而言,药学上可接受的载体是“可接受的”。术语诸如“赋形剂”、“载体”、“药学上可接受的载体”、“媒介物”等在本文中可互换地使用。

[0320] 术语“药物组合物”意为经配制用于药学用途的组合物。在一些实施方案中,药物组合物进一步包含药学上可接受的载体。在一些实施方案中,药物组合物包含附加试剂(例如,用于特异性递送、增加半衰期,或其他治疗性化合物)。

[0321] “程序性细胞死亡受体1(PDCD1或PD-1)多肽”意为一种蛋白质,其与NCBI登录号AJS10360.1或其片段具有至少约85%的氨基酸序列一致性。PD-1蛋白被认为牵涉到免疫反应期间的T细胞功能调节中并且牵涉到耐受条件中。示例性B2M多肽序列提供如下。

[0322] >AJS10360.1程序性细胞死亡1蛋白[智人(Homo sapiens)]MQIPQAPWPVVWAVLQLG
WRPGWFLDSPDRPWNPPTFSPALLVVTEGDNATFTCSFSNTSESFVLNWYRMSPSNQTDKLAAFPEDRSQPGQDCRF

RVTQLPNGRDFHMSVVRARRNDSGTYLCGAISLAPKAQIKESLRAELRVTERRAEVPTAHPSPSPRPAGQFQTLVVG
VVGGLLGSLVLLVWVLAVICSRAARGTIGARRTGQPLKEDPSAVPVFSVDYGELDFQWREKTPEPPVPCVPEQTEYA
TIVFPSGMGTSSPARRGSADGPRSAQPLRPEDGHCSWPL

[0323] “程序性细胞死亡1(PDCD1或PD-1)多核苷酸”意为编码PD-1多肽的核酸分子。PDCD1基因编码一种抑制性细胞表面受体,其以抗原特异性方式抑制T细胞效应子功能。示例性PDCD1核酸序列提供如下。

[0324] >AY238517.1智人(Homo sapiens)程序性细胞死亡1(PDCD1)mRNA,完全编码序列

[0325] ATGCAGATCCCACAGGCGCCCTGGCCAGTCGTCTGGGCGGTGCTACAACTGGGCTGGCGGCCAGGATGG
TTCTTAGACTCCCCAGACAGGCCCTGGAACCCCCCCACCTTCTCCCCAGCCCTGCTCGTGGTGACCGAAGGGGACAA
CGCCACCTTCACCTGCAGCTTCTCCAACACATCGGAGAGCTTCGTGCTAAACTGGTACCGCATGAGCCCCAGCAACC
AGACGGACAAGCTGGCCGCCTTCCCCGAGGACCGCAGCCAGCCCGGCCAGGACTGCCGCTTCCGTGTCACACAACTG
CCCAACGGGCGTGACTTCCACATGAGCGTGGTCAGGGCCCGGCGCAATGACAGCGGCACCTACCTCTGTGGGGCCAT
CTCCCTGGCCCCCAAGGCGCAGATCAAAGAGAGCCTGCGGGCAGAGCTCAGGGTGACAGAGAGAAGGGCAGAAGTGC
CCACAGCCCACCCCAGCCCCTCACCCAGGCCAGCCGGCCAGTTCCAAACCCTGGTGGTTGGTGTCGTGGGCGGCCTG
CTGGGCAGCCTGGTGCTGCTAGTCTGGGTCCTGGCCGTCATCTGCTCCCGGGCCGCACGAGGGACAATAGGAGCCAG
GCGCACCGGCCAGCCCCTGAAGGAGGACCCCTCAGCCGTGCCTGTGTTCTCTGTGGACTATGGGGAGCTGGATTTCC
AGTGGCGAGAGAAGACCCCGGAGCCCCCCGTGCCCTGTGTCCCTGAGCAGACGGAGTATGCCACCATTGTCTTTCCT
AGCGGAATGGGCACCTCATCCCCCGCCCGCAGGGGCTCAGCTGACGGCCCTCGGAGTGCCCAGCCACTGAGGCCTGA
GGATGGACACTGCTCTTGGCCCCTCTGA

[0326] 术语“蛋白质”、“肽”、“多肽”及其语法等效物在本文中可互换使用,并且是指通过肽(酰胺)键链接在一起的氨基酸残基的聚合物。该术语是指任何尺寸、结构或功能的蛋白质、肽或多肽。典型地,蛋白质、肽或多肽的长度将会是至少三个氨基酸。蛋白质、肽或多肽可以指单个蛋白质或蛋白质的集合。蛋白质、肽或多肽中的一个或多个氨基酸可以例如,通过以下进行修饰:化学实体诸如碳水化合物基团、羟基、磷酸酯基团、法尼基、异法尼基、脂肪酸基团的加成;用于缀合的连接子;官能化;或其他修饰等。蛋白质、肽或多肽也可以是单一分子或可以是多分子复合物。蛋白质、肽或多肽可以仅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、或其任意组合。如本文所用,术语“融合蛋白”是指一种杂交多肽,其包含来自至少两种不同蛋白质的蛋白结构域。一种蛋白质可以定位在融合蛋白的氨基端(N端)部分或在羧基端(C端)蛋白处,因此分别形成氨基端融合蛋白或羧基端融合蛋白。蛋白质可包含不同的结构域,例如,核酸结合结构域(例如,Cas9的gRNA结合结构域,其引导蛋白质到靶标位点的结合)和核酸切割结构域,或核酸编辑蛋白的催化性结构域。在一些实施方案中,蛋白质包含蛋白质部分(例如,构建核酸结合结构域的氨基酸序列)和有机化合物(例如,可充当核酸切割剂的化合物)。在一些实施方案中,蛋白质与核酸(例如,RNA或DNA)复合或缔合。本文所提供的任何蛋白质可通过本领域已知的任何方法生产。例如,本文所提供的蛋白质可经由重组蛋白表达和纯化来生产,其尤其适用于包含肽连接子的融合蛋白。用于重组蛋白表达和纯化的方法是周知的,并且包括由《分子克隆:实验室手册》(Green and Sambrook,Molecular Cloning:A Laboratory Manual(4th ed.,Cold Spring Harbor Laboratory Press,Cold Spring Harbor,N.Y.(2012)))描述的那些,该文献的整体内容通过引用并入本文。

[0327] 本文所公开的多肽和蛋白质(包括其功能性部分和功能性变体)可包含合成氨基酸代替一个或多个天然存在的氨基酸。此类合成氨基酸是本领域中已知的,并且包括,例如,氨基环己烷甲酸、正亮氨酸、α-氨基正癸酸、高丝氨酸、S-乙酰基氨基甲基-半胱氨酸、反式-3-和反式-4-羟基脯氨酸、4-氨基苯丙氨酸、4-硝基苯丙氨酸、4-氯苯丙氨酸、4-羧基苯丙氨酸、β-苯基丝氨酸、β-羟基苯丙氨酸、苯基甘氨酸、α-萘基丙氨酸、环己基丙氨酸、环己基甘氨酸、吲哚啉-2-甲酸、1,2,3,4-四氢异喹啉-3-甲酸、氨基丙二酸、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-氨基环戊烷甲酸、α-氨基环己烷甲酸、α-氨基环庚烷甲酸、α-(2-氨基-2-正亮氨酸)-甲酸、α,γ-二氨基丁酸、α,β-二氨基丙酸、高丙氨酸和α-叔丁基甘氨酸。多肽和蛋白质可与多肽构建体的一个或多个氨基酸的翻译后修饰缔合。翻译后修饰的非限制性示例包括磷酸化、酰化(包括乙酰化和甲酰化)、糖基化(包括N-联和O-联)、酰胺化、羟基化、烷基化(包括甲基化和乙基化)、泛素化、吡咯烷酮甲酸的加成、二硫键桥的形成、硫酸化、肉豆蔻酰化、棕榈酰化、异戊二烯化、法尼基化、香叶基化、糖基磷脂酰肌醇化、脂化和碘化。

[0328] "启动子"意为核酸控制序列的阵列,其引导核酸的转录。启动子包括接近转录起始位点的必需核酸序列。启动子也任选地包括远端增强子或抑制子序列元件。"组成型启动子"是一种持续活跃且不受外部信号或分子调节的启动子。相反,"诱导型启动子"的活性由外部信号或分子(例如,转录因子)调节。举例而言,启动子可以是CMV启动子。

[0329] 如本文所用,在蛋白质或核酸语境中的术语"重组"是指在自然界不存在但作为人类工程改造产物的蛋白质或核酸。例如,在一些实施方案中,重组蛋白质或核酸分子包含氨基酸或核苷酸序列,与任何天然存在的序列相比,该序列包含一个、至少两个、至少三个、至少四个、至少五个、至少六个或至少七个突变。

[0330] "降低"意为至少10%、25%、50%、75%或100%的反向改变。

[0331] "参考"意为标准或对照条件。在一种实施方案中,才可是野生型或健康细胞。在其他实施方案中并且不限于,参考是为处理的细胞,该细胞未经历测试条件或经历了安慰剂或生理盐水、媒介、缓冲液和/或不携带感兴趣的多核苷酸的对照载体。

[0332] "参考序列"是用作序列比对基础的定义序列。参考序列可以是特定序列的子集或整体,例如,全长度cDNA或基因序列的节段,或完整的cDNA或基因序列。对于多肽,参考多肽序列的长度通常将会是至少约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将会是至少约50个核苷酸,至少约60个核苷酸,至少约75个核苷酸,和约100个核苷酸或约300个核苷酸或其附近或期间任意整数个核苷酸。在一些实施方案中,参考序列是感兴趣的蛋白质的野生型序列。在其他实施方案中,参考序列是编码野生型蛋白质的多核苷酸序列。

[0333] 术语"RNA可编程核酸酶"和"RNA引导的核酸酶"与并非切割靶点的一种或多种RNA合用(例如,结合或缔合)。在一些实施方案中,当RNA可编程核酸酶与RNA复合时,可以指代为核酸酶:RNA复合物。典型地,结合的RNA指代为向导RNA(gRNA)。gRNA可以作为两种或更多种RNA的复合物存在,或作为单一RNA分子存在。作为单一RNA分子存在的gRNA可以称为单向导RNA(sgRNA),但"gRNA"可互换地使用以指代作为单一分子或作为两个或更多个分子的复合物存在的向导RNA。典型地,作为单一RNA物质存在的gRNA包含两个结构域:(1)与靶标核酸共享同源性的结构域(例如,并且引导Cas9复合物至靶标的结合);和(2)结合Cas9蛋白的

结构域。在一些实施方案中,结构域(2)对应于称为tracrRNA的序列并且包含茎环结构。例如,在一些实施方案中,结构域(2)与Jinek等人,Science 337:816-821(2012)中提供的tracrRNA相同或同源,该文献的整体内容通过引用并入本文。gRNA的其他实例(例如,那些包括结构域2的)可见于2013年9月6日提交的名为“Switchable Cas9 Nucleases and Uses Thereof”的美国专利申请号61/874,682和2013年9月6日提交的名为“Delivery System For Functional Nucleases”的美国专利临时申请号61/874,746,其各自的完整内容通过引用以其整体并入本文。在一些实施方案中,gRNA包含两个或更多个结构域(1)和(2)并且可以成为“扩展gRNA”。例如,扩展gRNA将会结合两个或更多个Cas9蛋白并在两个或更多个不同区域处结合靶标核酸,如本文所述。gRNA包含与靶标位点互补的核苷酸序列,其介导核酸酶/RNA复合物至所述靶标位点的结合,提供核酸酶:RNA复合物的序列特异性。

[0334] 在一些实施方案中,RNA可编程核酸酶是(CRISPR相关系统)Cas9核酸内切酶,例如,来自酿脓链球菌(Streptococcus pyogenes)的Cas9(Csn1)(参见例如,“Complete genome sequence of an M1 strain of Streptococcus pyogenes.”Ferretti J.J.,等人,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.,等人,Nature 471:602-607(2011))。因为RNA可编程核碱基(例如,Cas9)使用RNA:DNA杂交以靶向DNA切割位点,这些蛋白质原则上能够被靶向至向导RNA特异性的任何序列。使用RNA可编程核碱基诸如Cas9进行位点特异性切割(例如,以修饰基因组)的方法是本领域中已知的(参见,例如,Cong,L.et al,Multiplex genome engineering using CRISPR/Cas systems.Science 339,819-823(2013);Mali,P.et al,RNA-guided human genome engineering via Cas9.Science 339,823-826(2013);Hwang,W.Y.et al,Efficient genome editing in zebrafish using a CRISPR-Cas system.Nature 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 ah RNA-guided editing of bacterial genomes using CRISPR-Cas systems.Nature biotechnology 31,233-239(2013);其各自的整体内容通过引用并入本文)。

[0335] “信号传导结构域”意为在T细胞内表达的蛋白质的细胞内部分,其转导T细胞效应子功能信号(例如,激活信号)并且引导该T细胞执行专业功能。T细胞激活可以由多种因素诱导,包括抗原与T细胞表面上的T细胞受体的结合以及同源配体与T细胞表面上的共刺激分子的结合。T细胞共刺激分子是T细胞上的同源结合配偶体,其与共刺激配体特异性地结合,从而介导T细胞的共刺激应答,诸如但不限于,增殖。共刺激分子包括但不限于MHC I类分子。在一些实施方案中,共刺激结构域是CD2细胞质结构域。T细胞的激活导致免疫应答,诸如T细胞增殖和分化(参见,例如,Smith-Garvin等人,Annu.Rev.Immunol.,27:591-619,2009)。示例性T细胞信号传导结构域是本领域中已知的。非限制性示例包括CD2、CD3ζ、CD8、CD28、CD27、CD154、GITR(TNFRSF18)、CD134(OX40)和CD137(4-1BB)信号传导结构域。

[0336] “单链抗体”或“scFv”意为经基因工程改造的分子,其含有通过核实的肽连接子链接为经基因融合的单链分子的一个或多个抗体的VH和VL结构域(参见,例如,Bird等人,Science,242:423-426,1988;Huston等人,Proc.Natl.Acad.Sci.,85:5879-5883,1988;

Ahmad等人,Clin.Dev.Immunol.,2012,doi:10.1155/2012/980250:Marbry,IDrugs,13:543-549,2010)。在一些实施方案中,scFv中VH结构域和VL结构域的分子内取向是VH结构域-连接子结构域-VL结构域。在一些实施方案中,scFv中VH结构域和VL结构域的分子内取向是VL结构域-连接子结构域-VH结构域。

[0337] 术语“单一核苷酸多态性(SNP)”是单一核苷酸中的变异,该变异出现在基因组中的特异性位置处,其中每种变异以合适的程度存在于群体中(例如,>1%)。例如,在人类基因组的特异性碱基位置处,C核苷酸可以出现在大多数个体中,但在少数个体中,该位置被A占据。这意味着在这一特异性位置存在SNP,并且两个可能的核苷酸变异C或A被称为这一位置的等位基因。SNP是疾病易感性的差异的基础。疾病的严重程度和我们的身体对于治疗的应答也是基因变异的体现。SNP可落入基因的编码区内、基因的非编码区内或基因间区(基因之间的区域)内。在一种实施方案中,由于遗传密码的简并性,编码序列内的SNP不一定改变所产生的蛋白质的氨基酸序列。编码区内的SNP有两种类型:同义和非同义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。不在蛋白质编码区内的SNP仍然可能影响基因剪接、转录因子结合、信使RNA降解、或非编码RNA的序列。受这一类型的SNP影响的基因表达称为eSNP(表达SNP)且可以在该基因的上游或下游。单一核苷酸变体(SNV)是单一核苷酸中的变异而没有任何频率显著,并且可出现在体细胞中。体细胞单一核苷酸变异(例如,与癌症相关)也可称为单核苷酸改变。

[0338] “特异性地结合”意为核酸分子、多肽、或其复合物(例如,核酸可编程DNA结合蛋白、向导核酸和嵌合抗原受体)、化合物或分子识别并结合本发明的多肽和/或核酸分子,但基本上不识别并结合样品(例如,生物样品)中的其他分子。例如,嵌合抗原受体特异性地结合至表达在细胞表面上的特定标志物,但不结合至该细胞表面上的其他多肽、碳水化合物、脂质或任何其他化合物。

[0339] 可用于本发明方法中的核酸分子包括任何编码本发明多肽或其片段的核酸分子。此类核酸分子无需与内源性核酸序列100%一致,但典型将表现出基本一致性。具有与内源性序列的“基本一致性”的多核苷酸典型能与双链核酸分子的至少一条链杂交。可用于本发明方法中的核酸分子包括任何编码本发明多肽或其片段的核酸分子。此类核酸分子无需与内源性核酸序列100%一致,但典型将表现出基本一致性。具有与内源性序列的“基本一致性”的多核苷酸典型能与双链核酸分子的至少一条链杂交。“杂交”意为在各种严格条件下在互补多核苷酸序列(例如,本文所述的基因)或其片段之间配对以形成双链分子。(见,例如,Wahl,G.M.and S.L.Berger(1987)Methods Enzymol.152:399;Kimmel,A.R.(1987)Methods Enzymol.152:507)。

[0340] 例如,严格的盐浓度一般将低于约750mM NaCl和75mM柠檬酸三钠,优选低于500mM NaCl和50mM柠檬酸三钠,且更优选约250mM NaCl和25mM柠檬酸三钠。在有机溶剂例如甲酰胺的缺席下可获得低严格度杂交,而在至少约35%甲酰胺且更优选至少约50%甲酰胺的存在下可获得高严格度杂交。严格的温度条件一般将包括至少约30℃、更优选至少约37℃且最优选至少约42℃的温度。不同的其他参数例如杂交时间、表面活性剂例如十二烷基硫酸钠(SDS)的浓度、以及载剂DNA的包含或不包含是该领域技术人员所熟知的。通过按需要将这些条件组合,实现各种水平的严格度。在一种实施方案中,杂交将在30℃下在750mM NaCl、75mM柠檬酸三钠和1% SDS中发生。在另一实施方案中,杂交将在37℃下在500mM

NaCl、50mM柠檬酸三钠、1％ SDS、35％甲酰胺和100μg/ml变性鲑鱼精DNA(ssDNA)中发生。在另一实施方案中,杂交将在42℃下在250mM NaCl、25mM柠檬酸三钠、1％ SDS、50％甲酰胺和200μg/ml ssDNA中发生。这些条件的可用改变对于该领域技术人员是显而易见的。

[0341] 对于大多数应用,杂交后的洗涤步骤也将改变严格度。洗涤严格度条件可通过盐浓度并且通过温度界定。如上,可通过减低盐浓度或升高温度来增加洗涤严格度。例如,用于洗涤步骤的严格的盐浓度优选将低于约30mM NaCl和3mM柠檬酸三钠,且最优选约15mM NaCl和1.5mM柠檬酸三钠。用于洗涤步骤的严格的温度条件一般将包括至少约25℃、更优选至少约42℃、甚至更优选至少约68℃的温度。一种实施方案中,洗涤步骤将在25℃下在30mM NaCl、3mM柠檬酸三钠和0.1％ SDS中发生。在一种更优选的实施方案中,洗涤步骤将在42℃下在15mM NaCl、1.5mM柠檬酸三钠和0.1％ SDS中发生。在一种更优选的实施方案中,洗涤步骤将在68℃下在15mM NaCl、1.5mM柠檬酸三钠和0.1％ SDS中发生。这些条件的其他改变对于该领域技术人员是显而易见的。杂交技术是该领域技术人员所熟知的,并且在例如Benton and Davis(Science 196:180,1977);Grunstein and Hogness(Proc.Natl.Acad.Sci.,USA 72:3961,1975);《分子生物学现代方法》(Ausubel等人(Current Protocols in Molecular Biology,Wiley Interscience,New York,2001));《分子克隆技术指南》(Berger and Kimmel(Guide to Molecular Cloning Techniques,1987,Academic Press,New York));和《分子克隆:实验室手册》(Sambrook等人,Molecular Cloning:A Laboratory Manual,Cold Spring Harbor Laboratory Press,New York)中有所描述。

[0342] “分体”意为分为两个或更多个片段。

[0343] “分体Cas9蛋白”或“分体Cas9”是指一种Cas9蛋白,其提供为由两个独立的核苷酸序列编码的N端片段和C端片段。对应于Cas9蛋白的N端部分和C端部分的多肽可以经剪接以形成“重建的”Cas9蛋白。在特定实施方案中,Cas9蛋白被分为位于蛋白质的无序区内的两个片段,例如,如Nishimasu等人,Cell,Volume 156,Issue 5,pp.935-949,2014中所述或如Jiang等人(2016)Science 351:867-871.PDB file:5F9R中所述,该文献各自通过引用并入本文。在一些实施方案中,蛋白质被分为位于SpCas9的氨基酸A292至G364之间、F445至K483之间或E565至T637之间的区域内的任何C、T、A或S处或位于任何其他Cas9、Cas9变体(例如,nCas9、dCas9)或其他napDNAbp中对应位置处的两个片段。在一些实施方案中,蛋白质被分为位于SpCas9 T310、T313、A456、S469或C574处的两个片段。在一些实施方案中,将蛋白质分为两个片段的过程称为“分裂”蛋白质。

[0344] 在其他实施方案中,Cas9蛋白的N端部分包含氨基酸1至573或1至637酿脓链球菌Cas9野生型(SpCas9)(NCBI参考序列:NC_002737.2,Uniprot参考序列:Q99ZW2),且Cas9蛋白的C端部分包含SpCas9野生型的氨基酸574至1368或638至1368的部分或其对应位置。

[0345] 分体Cas9的C端部分可以与分体Cas9的N端部分接合以形成完整的Cas9蛋白。在一些实施方案中,Cas9蛋白的C端部分起始于Cas9蛋白的N端部分结束之处。因此,在一些实施方案中,分体Cas9的C端部分包含spCas9的氨基酸(551至651)至1368的部分。“(551至651)至1368”意为在氨基酸551至651(含)之间的氨基酸处开始并且在氨基酸1368处结束。例如,分体Cas9的C端部分可包含spCas9的氨基酸551至1368、552至1368、553至1368、554至1368、555至1368、556至1368、557至1368、558至1368、559至1368、560至1368、561至1368、562至

1368、563至1368、564至1368、565至1368、566至1368、567至1368、568至1368、569至1368、570至1368、571至1368、572至1368、573至1368、574至1368、575至1368、576至1368、577至1368、578至1368、579至1368、580至1368、581至1368、582至1368、583至1368、584至1368、585至1368、586至1368、587至1368、588至1368、589至1368、590至1368、591至1368、592至1368、593至1368、594至1368、595至1368、596至1368、597至1368、598至1368、599至1368、600至1368、601至1368、602至1368、603至1368、604至1368、605至1368、606至1368、607至1368、608至1368、609至1368、610至1368、611至1368、612至1368、613至1368、614至1368、615至1368、616至1368、617至1368、618至1368、619至1368、620至1368、621至1368、622至1368、623至1368、624至1368、625至1368、626至1368、627至1368、628至1368、629至1368、630至1368、631至1368、632至1368、633至1368、634至1368、635至1368、636至1368、637至1368、638至1368、639至1368、640至1368、641至1368、642至1368、643至1368、644至1368、645至1368、646至1368、647至1368、648至1368、649至1368、650至1368或651至1368中的任一者。在一些实施方案中,分体Cas9蛋白的C端部分包含SpCas9的氨基酸574至1368或638至1368的部分。

[0346] “受试者”意为哺乳动物,包括但不限于,人类和非人类哺乳动物,例如牛、马、犬、羊或猫。受试者包括家畜、用来提供劳动力和提供日用品诸如食物的驯养动物,包括但不限于,牛、山羊、鸡、马、猪、兔和绵羊。

[0347] “基本上相同”意为多肽或核酸分子表现出至少50%的与参考氨基酸序列(例如,本文所述氨基酸序列中的任何一个)或核酸分子(例如,本文所述核酸序列中的任何一个)的一致性。在一种实施方案中,该序列在氨基酸水平上或核酸暖水瓶上与用于比较的核酸至少60%、80%或85%、90%、95%或甚至99%相同。

[0348] 典型使用序列分析软件(例如,威斯康星大学生物技术中心遗传学计算机组(Genetics Computer Group,University of Wisconsin Biotechnology Center,1710University Avenue,Madison,Wis.53705)的序列分析软件包(Sequence Analysis Software Package),BLAST、BESTFIT、COBALT、EMBOSS Needle、GAP或PILEUP/PRETTYBOX程序)测量序列一致性。此类软件通过对不同的替换、删除和/或其他修饰指定同源性程度来匹配相同或相似的序列。保守替换典型包括系列各组的组内替换:甘氨酸、丙氨酸;缬氨酸、异亮氨酸、亮氨酸;天冬氨酸、谷氨酸、天冬酰胺、谷氨酰胺;丝氨酸、苏氨酸;赖氨酸、精氨酸;以及苯丙氨酸、酪氨酸。在测定一致性程度的一个示例性方法中,可使用BLAST程序,e-3和e-100之间的概率评分指示密切相关的序列。

[0349] 使用COBALT,例如,采用以下参数:

[0350] a)比对参数:空位罚分-11、-1,且末端空位罚分-5、-1,

[0351] b)CDD参数:使用RPS BLAST打开;Blast E值0.003;Find Conserved columns and Recompute打开,和

[0352] c)查询聚类参数:使用query clusters打开;字长4;最大簇间距离0.8;Alphabet Regular。

[0353] 使用EMBOSS Needle,例如,使用以下参数:

[0354] a)矩阵:BLOSUM62;

[0355] b)GAP OPEN:10;

[0356] c)GAP EXTEND:0.5;

[0357] d)OUTPUT FORMAT;对;

[0358] e)END GAP PENALTY:假;

[0359] f)END GAP OPEN:10;和

[0360] g)END GAP EXTEND:0.5.

[0361] 术语“靶标位点”是指核酸分子内的被核碱基编辑器修饰的序列。在一种实施方案中,靶标位点由脱氨酶或包含脱氨酶(例如,胞苷或腺苷脱氨酶)的融合蛋白决定。

[0362] “T细胞受体α恒定区(TRAC)多肽”意为一种蛋白质,其与NCBI登录号P01848.2或其片段具有至少约85%的氨基酸序列一致性并且具有免疫调节活性。示例性氨基酸序列提供如下。

[0363] >sp|P01848.2|TRAC_HUMAN RecName:Full=T细胞受体α恒定区IQNPDPAVYQLRDSKSSDKSVCLFTDFDSQTNVSQSKDSDVYITDKTVLDMRSMDFKSNSAVAWSNKSDFACANAFNNSIIPEDTFFPSPESSCDVKLVEKSFETDTNLNFQNLSVIGFRILLLKVAGFNLLMTLRLWSS

[0364] “T细胞受体α恒定区(TRAC)多核苷酸”意为编码TRAC多肽的核酸。示例性TRAC核酸序列提供如下。

[0365] UCSC人类基因组数据库,基因ENSG00000277734.8人T细胞受体α链(TCR-α)

[0366] catgctaatcctccggcaaacctctgtttcctcctcaaaaggcaggaggtcggaaagaataaacaatgagagtcacattaaaaacacaaaatcctacggaaatactgaagaatgagtctcagcactaaggaaaagcctccagcagctcctgctttctgagggtgaaggatagacgctgtggctctgcatgactcactagcactctatcacggccatattctggcagggtcagtggctccaactaacatttgtttggtactttacagtttattaaatagatgtttatatggagaagctctcatttctttctcagaagagcctggctaggaaggtggatgaggcaccatattcattttgcaggtgaaattcctgagatgtaaggagctgctgtgacttgctcaaggccttatatcgagtaaacggtagtgctggggcttagacgcaggtgttctgatttatagttcaaaacctctatcaatgagagagcaatctcctggtaatgtgatagatttcccaacttaatgccaacataccataaacctcccattctgctaatgcccagcctaagttggggagaccactccagattccaagatgtacagtttgctttgctgggccttttcccatgcctgcctttactctgccagagttatattgctggggttttgaagaagatcctattaaataaaagaataagcagtattattaagtagccctgcatttcaggtttccttgagtggcaggccaggcctggccgtgaacgttcactgaaatcatggcctcttggccaagattgatagcttgtgcctgtccctgagtcccagtccatcacgagcagctggtttctaagatgctatttcccgtataaagcatgagaccgtgacttgccagccccacagagcccccgcccttgtccatcactggcatctggactccagcctgggttggggcaaagagggaaatgagatcatgtcctaacccctgatcctcttgtcccacagATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACAAGTCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGATTCTGATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCTATGGACTTCAAGAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGCCTTCAACAACAGCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGgtaagggcagctttggtgccttcgcaggctgtttccttgcttcaggaatggccaggttctgcccagagctctggtcaatgatgtctaaaactcctctgattggtggtctcggccttatccattgccaccaaaaccctcttttactaagaaacagtgagccttgttctggcagtccagagaatgacacgggaaaaaagcagatgaagagaaggtggcaggagagggcacgtggcccagcctcagtctctccaactgagttcctgcctgcctgcctttgctcagactgtttgccccttactgctcttctaggcctcattctaagccccttctccaagttgcctctccttatttctccctgtctgccaaaaaatctttcccagctcactaagtcagtctcacgcagtcactcattaacccaccaatcactgattgtgccggcacatgaatgcaccaggtgttgaagtgg

aggaattaaaaagtcagatgaggggtgtgcccagaggaagcaccattctagttgggggagcccatctgtcagctgg
gaaaagtccaaataacttcagattggaatgtgttttaactcagggttgagaaaacagctaccttcaggacaaaagt
cagggaagggctctctgaagaaatgctacttgaagataccagccctaccaagggcagggagaggaccctatagagg
cctgggacaggagctcaatgagaaaggagaagagcagcaggcatgagttgaatgaaggaggcagggccgggtcaca
gggccttctaggccatgagagggtagacagtattctaaggacgccagaaagctgttgatcggcttcaagcagggga
gggacacctaatttgcttttcttttttttttttttttttttttttttttttgagatggagttttgctcttgttgcc
caggctggagtgcaatggtgcatcttggctcactgcaacctccgcctcccaggttcaagtgattctcctgcctcag
cctcccgagtagctgagattacaggcacccgccaccatgcctggctaatttttttgtatttttagtagagacagggt
ttcactatgttggccaggctggtctcgaactcctgacctcaggtgatccacccgcttcagcctcccaaagtgctgg
gattacaggcgtgagccaccacacccggcctgcttttcttaaagatcaatctgagtgctgtacggagagtgggttg
taagccaagagtagaagcagaaagggagcagttgcagcagagagatgatggaggcctgggcagggtggtggcaggg
aggtaaccaacaccattcaggtttcaaaggtagaaccatgcagggatgagaaagcaaagaggggatcaaggaaggc
agctggattttggcctgagcagctgagtcaatgatagtgccgtttactaagaagaaaccaaggaaaaaatttgggg
tgcagggatcaaaacttttggaacatatgaaagtacgtgtttatactctttatggcccttgtcactatgtatgcc
tcgctgcctccattggactctagaatgaagccaggcaagagcagggtctatgtgtgatggcacatgtggccagggt
catgcaacatgtactttgtacaaacagtgtatattgagtaaatagaaatggtgtccaggagccgaggtatcggtcc
tgccagggccaggggctctccctagcaggtgctcatatgctgtaagttccctccagatctctccacaaggaggcat
ggaaaggctgtagttgttcacctgcccaagaactaggaggtctggggtgggagagtcagcctgctctggatgctga
aagaatgtctgtttttccttttagAAAGTTCCTGTGATGTCAAGCTGGTCGAGAAAAGCTTTGAAACAGgtaagac
aggggtctagcctgggtttgcacaggattgcggaagtgatgaacccgcaataaccctgcctggatgagggagtggg
aagaaattagtagatgtgggaatgaatgatgaggaatggaaacagcggttcaagacctgcccagagctgggtgggg
tctctcctgaatccctctcaccatctctgactttccattctaagcactttgaggatgagtttctagcttcaataga
ccaaggactctctcctaggcctctgtattcctttcaacagctccactgtcaagagagccagagagagcttctgggt
ggcccagctgtgaaatttctgagtcccttagggatagccctaaacgaaccagatcatcctgaggacagccaagagg
ttttgccttctttcaagacaagcaacagtactcacataggctgtgggcaatggtcctgtctctcaagaatcccctg
ccactcctcacacccaccctgggcccatattcatttccatttgagttgttcttattgagtcatccttcctgtggta
gcggaactcactaaggggcccatctggacccgaggtattgtgatgataaattctgagcacctaccccatccccaga
agggctcagaaataaaataagagccaagtctagtcggtgtttcctgtcttgaaacacaatactgttggccctggaa
gaatgcacagaatctgtttgtaaggggatatgcacagaagctgcaagggacaggaggtgcaggagctgcaggcctc
ccccacccagcctgctctgccttggggaaaaccgtgggtgtgtcctgcaggccatgcaggcctgggacatgcaagc
ccataaccgctgtggcctcttggttttacagATACGAACCTAAACTTTCAAAACCTGTCAGTGATTGGGTTCCGAA
TCCTCCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATGACGCTGCGGCTGTGGTCCAGCTGAGgtgaggggccttg
aagctgggagtggggtttagggacgcgggtctctgggtgcatcctaagctctgagagcaaacctccctgcagggtc
ttgcttttaagtccaaagcctgagcccaccaaactctcctacttcttcctgttacaaattcctcttgtgcaataat
aatggcctgaaacgctgtaaaatatcctcatttcagccgcctcagttgcacttctcccctatgaggtaggaagaac
agttgtttagaaacgaagaaactgaggccccacagctaatgagtggaggaagagagacacttgtgtacaccacatg
ccttgtgttgtacttctctcaccgtgtaacctcctcatgtcctctctccccagtacggctctcttagctcagtaga
aagaagacattacactcatattacaccccaatcctggctagagtctccgcaccctcctcccccagggtccccagtc
gtcttgctgacaactgcatcctgttccatcaccatcaaaaaaaaactccaggctgggtgcgggggctcacacctgt

aatcccagcactttgggaggcagaggcaggaggagcacaggagctggagaccagcctgggcaacacagggagaccc
cgcctctacaaaaagtgaaaaaattaaccaggtgtggtgctgcacacctgtagtcccagctacttaagaggctgag
atgggaggatcgcttgagccctggaatgttgaggctacaatgagctgtgattgcgtcactgcactccagcctggaa
gacaaagcaagatcctgtctcaaataataaaaaaaataagaactccagggtacatttgctcctagaactctaccac
atagccccaaacagagccatcaccatcacatccctaacagtcctgggtcttcctcagtgtccagcctgacttctgt
tcttcctcattccagATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTCCTTCCTCTGCATTGCCCCTCTTC
TCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCAAGGAGGTGAAAGCTGCTACCACCTCTGTGCCCCCCCGGCAAT
GCCACCAACTGGATCCTACCCGAATTTATGATTAAGATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCTGT
TCCCTTATTGCTGCTTGTCACTGCCTGACATTCACGGCAGAGGCAAGGCTGCTGCAGCCTCCCCTGGCTGTGCACAT
TCCCTCCTGCTCCCCAGAGACTGCCTCCGCCATCCCACAGATGATGGATCTTCAGTGGGTTCTCTTGGGCTCTAGGT
CCTGCAGAATGTTGTGAGGGGTTTATTTTTTTTTAATAGTGTTCATAAAGAAATACATAGTATTCTTCTTCTCAAGA
CGTGGGGGGAAATTATCTCATTATCGAGGCCCTGCTATGCTGTGTATCTGGGCGTGTTGTATGTCCTGCTGCCGATG
CCTTCATTAAAATGATTTGGAAGAGCAGA

[0367] 上文中以小写字母表示的核苷酸是非翻译区或内含子,且以大写字母表示的核苷酸是外显子。

[0368] >X02592.1人类T细胞α链(TCR-α)的mRNA

[0369] TTTTGAAACCCTTCAAAGGCAGAGACTTGTCCAGCCTAACCTGCCTGCTGCTCCTAGCTCCTGAGGCTC
AGGGCCCTTGGCTTCTGTCCGCTCTGCTCAGGGCCCTCCAGCGTGGCCACTGCTCAGCCATGCTCCTGCTGCTCGTC
CCAGTGCTCGAGGTGATTTTTACCCTGGGAGGAACCAGAGCCCAGTCGGTGACCCAGCTTGGCAGCCACGTCTCTGT
CTCTGAAGGAGCCCTGGTTCTGCTGAGGTGCAACTACTCATCGTCTGTTCCACCATATCTCTTCTGGTATGTGCAAT
ACCCCAACCAAGGACTCCAGCTTCTCCTGAAGTACACATCAGCGGCCACCCTGGTTAAAGGCATCAACGGTTTTGAG
GCTGAATTTAAGAAGAGTGAAACCTCCTTCCACCTGACGAAACCCTCAGCCCATATGAGCGACGCGGCTGAGTACTT
CTGTGCTGTGAGTGATCTCGAACCGAACAGCAGTGCTTCCAAGATAATCTTTGGATCAGGGACCAGACTCAGCATCC
GGCCAAATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAGACTCTAAATCCAGTGACAAGTCTGTCTGCCTA
TTCACCGATTTTGATTCTCAAACAAATGTGTCACAAAGTAAGGATTCTGATGTGTATATCACAGACAAAACTGTGCT
AGACATGAGGTCTATGGACTTCAAGAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAAACG
CCTTCAACAACAGCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCCTGTGATGTCAAGCTGGTCGAG
AAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAACCTGTCAGTGATTGGGTTCCGAATCCTCCTCCTGAAAGT
GGCCGGGTTTAATCTGCTCATGACGCTGCGGCTGTGGTCCAGCTGAGATCTGCAAGATTGTAAGACAGCCTGTGCTC
CCTCGCTCCTTCCTCTGCATTGCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCAAGGAGGTGAAA
GCTGCTACCACCTCTGTGCCCCCCCGGTAATGCCACCAACTGGATCCTACCCGAATTTATGATTAAGATTGCTGAAG
AGCTGCCAAACACTGCTGCCACCCCCTCTGTTCCCTTATTGCTGCTTGTCACTGCCTGACATTCACGGCAGAGGCAA
GGCTGCTGCAGCCTCCCCTGGCTGTGCACATTCCCTCCTGCTCCCCAGAGACTGCCTCCGCCATCCCACAGATGATG
GATCTTCAGTGGGTTCTCTTGGGCTCTAGGTCCTGGAGAATGTTGTGAGGGGTTTATTTTTTTTTAATAGTGTTCAT
AAAGAAATACATAGTATTCTTCTTCTCAAGACGTGGGGGGAAATTATCTCATTATCGAGGCCCTGCTATGCTGTGTG
TCTGGGCGTGTTGTATGTCCTGCTGCCGATGCCTTCATTAAAATGATTTGGAA

[0370] “T细胞受体β恒定区1多肽(TRBC1)”意为一种蛋白质,其与NCBI登录号P01850或其片段具有至少约85%的氨基酸序列一致性并且具有免疫调节活性。示例性氨基酸序列提供如下。

[0371] >sp|P01850|TRBC1_HUMAN T细胞受体β恒定区1OS＝智人(Homo sapiens)OX＝9606GN＝TRBC1 PE＝1SV＝4

[0372] DLNKVFPPEVAVFEPSEAEISHTQKATLVCLATGFFPDHVELSWWVNGKEVHSGVSTDPQPLKEQPALNDSRYCLSSRLRVSATFWQNPRNHFRCQVQFYGLSENDEWTQDRAKPVTQIVSAEAWGRADCGFTSVSYQQGVLSATILYEILLGKATLYAVLVSALVLMAMVKRKDF

[0373] “T细胞受体β恒定区1多核苷酸(TRBC1)”意为编码TRBC1多肽的核酸。示例性TRBC1核酸序列提供如下。

[0374] >X00437.1

[0375] CTGGTCTAGAATATTCCACATCTGCTCTCACTCTGCCATGGACTCCTGGACCTTCTGCTGTGTGTCCCTTTGCATCCTGGTAGCGAAGCATACAGATGCTGGAGTTATCCAGTCACCCCGCCATGAGGTGACAGAGATGGGACAAGAAGTGACTCTGAGATGTAAACCAATTTCAGGCCACAACTCCCTTTTCTGGTACAGACAGACCATGATGCGGGGACTGGAGTTGCTCATTTACTTTAACAACAACGTTCCGATAGATGATTCAGGGATGCCCGAGGATCGATTCTCAGCTAAGATGCCTAATGCATCATTCTCCACTCTGAAGATCCAGCCCTCAGAACCCAGGGACTCAGCTGTGTACTTCTGTGCCAGCAGTTTCTCGACCTGTTCGGCTAACTATGGCTACACCTTCGGTTCGGGGACCAGGTTAACCGTTGTAGAGGACCTGAACAAGGTGTTCCCACCCGAGGTCGCTGTGTTTGAGCCATCAGAAGCAGAGATCTCCCACACCCAAAAGGCCACACTGGTGTGCCTGGCCACAGGCTTCTTCCCCGACCACGTGGAGCTGAGCTGGTGGGTGAATGGGAAGGAGGTGCACAGTGGGGTCAGCACAGACCCGCAGCCCCTCAAGGAGCAGCCCGCCCTCAATGACTCCAGATACTGCCTGAGCAGCCGCCTGAGGGTCTCGGCCACCTTCTGGCAGAACCCCCGCAACCACTTCCGCTGTCAAGTCCAGTTCTACGGGCTCTCGGAGAATGACGAGTGGACCCAGGATAGGGCCAAACCCGTCACCCAGATCGTCAGCGCCGAGGCCTGGGGTAGAGCAGACTGTGGCTTTACCTCGGTGTCCTACCAGCAAGGGGTCCTGTCTGCCACCATCCTCTATGAGATCCTGCTAGGGAAGGCCACCCTGTATGCTGTGCTGGTCAGCGCCCTTGTGTTGATGGCCATGGTCAAGAGAAAGGATTTCTGAAGGCAGCCCTGGAAGTGGAGTTAGGAGCTTCTAACCCGTCATGGTTCAATACACATTCTTCTTTTGCCAGCGCTTCTGAAGAGCTGCTCTCACCTCTCTGCATCCCAATAGATATCCCCCTATGTGCATGCACACCTGCACACTCACGGCTGAAATCTCCCTAACCCAGGGGGAC

[0376] “T细胞受体β恒定区2多肽(TRBC2)”意为一种蛋白质,其与NCBI登录号A0A5B9或其片段具有至少约85％的氨基酸序列一致性并且具有免疫调节活性。示例性氨基酸序列提供如下。

[0377] >sp|A0A5B9|TRBC2_HUMAN T细胞受体β恒定区1OS＝智人(Homo sapiens)OX＝9606GN＝TRBC2 PE＝2SV＝2

[0378] DLKNVFPPKVAVFEPSEAEISHTQKATLVCLATGFYPDHVELSWWVNGKEVHSGVSTDPQPLKEQPALNDSRYCLSSRLRVSATFWQNPRNHFRCQVQFYGLSENDEWTQDRAKPVTQIVSAEAWGRADCGFTSESYQQGVLSATILYEILLGKATLYAVLVSALVLMAMVKRKDSRG

[0379] “T细胞受体β恒定区2多核苷酸(TRBC2)”意为编码TRAC多肽的核酸。示例性TRBC2核酸序列提供如下。

[0380] >NG_001333.2:655095-656583智人(Homo sapiens)染色体7上的T细胞受体β基因组(TRB)AGGACCTGAAAAACGTGTTCCCACCCGAGGTCGCTGTGTTTGAGCCATCAGAAGCAGAGATCTCCCACACCCCAAAAGGCCACACTGGTATGCCTGGCCACAGGCTTCTACCCCGACCACGTGGAGCTGAGCTGGTGGGTGAATGGGAAGGAGGTGCACAGTGGGGTCAGCACAGACCCGCAGCCCCTCAAGGAGCAGCCCGCCCTCAATGACTCCAGATACTGCCTGAGCAGCCGCCTGAGGGTCTCGGCCACCTTCTGGCAGAACCCCCGCAACCACTTCCGCTGTCAAGTCCAGTTCT

ACGGGCTCTCGGAGAATGACGAGTGGACCCAGGATAGGGCCAAACCCGTCACCCAGATCGTCAGCGCCGAGGCCTGG
GGTAGAGCAGGTGAGTGGGGCCTGGGGAGATGCCTGGAGGAGATTAGGTGAGACCAGCTACCAGGGAAAATGGAAAG
ATCCAGGTAGCGGACAAGACTAGATCCAGAAGAAAGCCAGAGTGGACAAGGTGGGATGATCAAGGTTCACAGGGTCA
GCAAAGCACGGTGTGCACTTCCCCCACCAAGAAGCATAGAGGCTGAATGGAGCACCTCAAGCTCATTCTTCCTTCAG
ATCCTGACACCTTAGAGCTAAGCTTTCAAGTCTCCCTGAGGACCAGCCATACAGCTCAGCATCTGAGTGGTGTGCAT
CCCATTCTCTTCTGGGGTCCTGGTTTCCTAAGATCATAGTGACCACTTCGCTGGCACTGGAGCAGCATGAGGGAGAC
AGAACCAGGGCTATCAAAGGAGGCTGACTTTGTACTATCTGATATGCATGTGTTTGTGGCCTGTGAGTCTGTGATGT
AAGGCTCAATGTCCTTACAAAGCAGCATTCTCTCATCCATTTTTCTTCCCCTGTTTTCTTTCAGACTGTGGCTTCAC
CTCCGGTAAGTGAGTCTCTCCTTTTTCTCTCTATCTTTCGCCGTCTCTGCTCTCGAACCAGGGCATGGAGAATCCAC
GGACACAGGGGCGTGAGGGAGGCCAGAGCCACCTGTGCACAGGTGCCTACATGCTCTGTTCTTGTCAACAGAGTCTT
ACCAGCAAGGGGTCCTGTCTGCCACCATCCTCTATGAGATCTTGCTAGGGAAGGCCACCTTGTATGCCGTGCTGGTC
AGTGCCCTCGTGCTGATGGCCATGGTAAGGAGGAGGGTGGGATAGGGCAGATGATGGGGGCAGGGGATGGAACATCA
CACATGGGCATAAAGGAATCTCAGAGCCAGAGCACAGCCTAATATATCCTATCACCTCAATGAAACCATAATGAAGC
CAGACTGGGGAGAAAATGCAGGGAATATCACAGAATGCATCATGGGAGGATGGAGACAACCAGCGAGCCCTACTCAA
ATTAGGCCTCAGAGCCCGCCTCCCCTGCCCTACTCCTGCTGTGCCATAGCCCCTGAAACCCTGAAAATGTTCTCTCT
TCCACAGGTCAAGAGAAAGGATTCCAGAGGCTAG

[0381] “tet甲基胞嘧啶双加氧酶2(TET2)多肽”意为一种蛋白质,其与NCBI登录号FM992369.1或其片段具有至少约85%的氨基酸序列一致性并且具有将甲基胞嘧啶转化为5-羟甲基胞嘧啶的催化活性。该基因中的缺陷已经与骨髓增值性疾患相关联,并且该酶将胞嘧啶甲基化的能力有贡献于转录调节。示例性TET2氨基酸序列提供如下。

[0382] >CAX30492.1 tet致癌基因家族成员2[智人(Homo sapiens)]

[0383] MEQDRTNHVEGNRLSPFLIPSPPICQTEPLATKLQNGSPLPERAHPEVNGDTKWHSFKSYYGIPCMKGS
QNSRVSPDFTQESRGYSKCLQNGGIKRTVSEPSLSGLLQIKKLKQDQKANGERRNFGVSQERNPGESSQPNVSDLSD
KKESVSSVAQENAVKDFTSFSTHNCSGPENPELQILNEQEGKSANYHDKNIVLLKNKAVLMPNGATVSASSVEHTHG
ELLEKTLSQYYPDCVSIAVQKTTSHINAINSQATNELSCEITHPSHTSGQINSAQTSNSELPPKPAAVVSEACDADD
ADNASKLAAMLNTCSFQKPEQLQQQKSVFEICPSPAENNIQGTTKLASGEEFCSGSSSNLQAPGGSSERYLKQNEMN
GAYFKQSSVFTKDSFSATTTPPPPSQLLLSPPPPLPQVPQLPSEGKSTLNGGVLEEHHHYPNQSNTTLLREVKIEGK
PEAPPSQSPNPSTHVCSPSPMLSERPQNNCVNRNDIQTAGTMTVPLCSEKTRPMSEHLKHNPPIFGSSGELQDNCQQ
LMRNKEQEILKGRDKEQTRDLVPPTQHYLKPGWIELKAPRFHQAESHLKRNEASLPSILQYQPNLSNQMTSKQYTGN
SNMPGGLPRQAYTQKTTQLEHKSQMYQVEMNQGQSQGTVDQHLQFQKPSHQVHFSKTDHLPKAHVQSLCGTRFHFQQ
RADSQTEKLMSPVLKQHLNQQASETEPFSNSHLLQHKPHKQAAQTQPSQSSHLPQNQQQQQKLQIKNKEEILQTFPH
PQSNNDQQREGSFFGQTKVEECFHGENQYSKSSEFETHNVQMGLEEVQNINRRNSPYSQTMKSSACKIQVSCSNNTH
LVSENKEQTTHPELFAGNKTQNLHHMQYFPNNVIPKQDLLHRCFQEQEQKSQQASVLQGYKNRNQDMSGQQAAQLAQ
QRYLIHNHANVFPVPDQGGSHTQTPPQKDTQKHAALRWHLLQKQEQQQTQQPQTESCHSQMHRPIKVEPGCKPHACM
HTAPPENKTWKKVTKQENPPASCDNVQQKSIIETMEQHLKQFHAKSLFDHKALTLKSQKQVKVEMSGPVTVLTRQTT
AAELDSHTPALEQQTTSSEKTPTKRTAASVLNNFIESPSKLLDTPIKNLLDTPVKTQYDFPSCRCVEQIIEKDEGPF
YTHLGAGPNVAAIREIMEERFGQKGKAIRIERVIYTGKEGKSSQGCPIAKWVVRRSSSEEKLLCLVRERAGHTCEAA
VIVILILVWEGIPLSLADKLYSELTETLRKYGTLTNRRCALNEERTCACQGLDPETCGASFSFGCSWSMYYNGCKFA
RSKIPRKFKLLGDDPKEEEKLESHLQNLSTLMAPTYKKLAPDAYNNQIEYEHRAPECRLGLKEGRPFSGVTACLDFC

AHAHRDLHNMQNGSTLVCTLTREDNREFGGKPEDEQLHVLPLYKVSDVDEFGSVEAQEEKKRSGAIQVLSSFRRKVR
MLAEPVKTCRQRKLEAKKAAAEKLSSLENSSNKNEKEKSAPSRTKQTENASQAKQLAELLRLSGPVMQQSQQPQPLQ
KQPPQPQQQQRPQQQQPHHPQTESVNSYSASGSTNPYMRRPNPVSPYPNSSHTSDIYGSTSPMNFYSTSSQAAGSYL
NSSNPMNPYPGLLNQNTQYPSYQCNGNLSVDNCSPYLGSYSPQSQPMDLYRYPSQDPLSKLSLPPIHTLYQPRFGNS
QSFTSKYLGYGNQNMQGDGFSSCTIRPNVHHVGKLPPYPTHEMDGHFMGATSRLPPNLSNPNMDYKNGEHHSPSHII
HNYSAAPGMFNSSLHALHLQNKENDMLSHTANGLSKMLPALNHDRTACVQGGLHKLSDANGQEKQPLALVQGVASGA
EDNDEVWSDSEQSFLDPDIGGVAVAPTHGSILIECAKRELHATTPLKNPNRNHPTRISLVFYQHKSMNEPKHGLALW
EAKMAEKAREKEEECEKYGPDYVPQKSHGKKVKREPAEPHETSEPTYLRFIKSLAERTMSVTTDSTVTTSPYAFTRV
TGPYNRYI

[0384] “tet甲基胞嘧啶双加氧酶2(TET2)多核苷酸”意为编码TET2多肽的核酸分子。TET多肽编码甲基胞嘧啶双加氧酶并且具有转录调节活性。示例性TET2核酸序列呈现如下。

[0385] >FM992369.1智人(Homo sapiens)tet致癌基因家族成员2(TET2基因)的mRNACCG
TGCCATCCCAACCTCCCACCTCGCCCCCAACCTTCGCGCTTGCTCTGCTTCTTCTCCCAGGGGTGGAGACCCGCCGA
GGTCCCCGGGGTTCCCGAGGGCTGCACCCTTCCCCGCGCTCGCCAGCCCTGGCCCCTACTCCGCGCTGGTCCGGGCG
CACCACTCCCCCCGCGCCACTGCACGGCGTGAGGGCAGCCCAGGTCTCCACTGCGCGCCCCGCTGTACGGCCCCAGG
TGCCGCCGGCCTTTGTGCTGGACGCCCGGTGCGGGGGGCTAATTCCCTGGGAGCCGGGGCTGAGGGCCCCAGGGCGG
CGGCGCAGGCCGGGGCGGAGCGGGAGGAGGCCGGGGCGGAGCAGGAGGAGGCCCGGGCGGAGGAGGAGAGCCGGCGG
TAGCGGCAGTGGCAGCGGCGAGAGCTTGGGCGGCCGCCGCCGCCTCCTCGCGAGCGCCGCGCGCCCGGGTCCCGCTC
GCATGCAAGTCACGTCCGCCCCCTCGGCGCGGCCGCCCCGAGACGCCGGCCCCGCTGAGTGATGAGAACAGACGTCA
AACTGCCTTATGAATATTGATGCGGAGGCTAGGCTGCTTTCGTAGAGAAGCAGAAGGAAGCAAGATGGCTGCCCTTT
AGGATTTGTTAGAAAGGAGACCCGACTGCAACTGCTGGATTGCTGCAAGGCTGAGGGACGAGAACGAGGCTGGCAAA
CATTCAGCAGCACACCCTCTCAAGATTGTTTACTTGCCTTTGCTCCTGTTGAGTTACAACGCTTGGAAGCAGGAGAT
GGGCTCAGCAGCAGCCAATAGGACATGATCCAGGAAGAGCAAATTCAACTAGAGGGCAGCCTTGTGGATGGCCCCGA
AGCAAGCCTGATGGAACAGGATAGAACCAACCATGTTGAGGGCAACAGACTAAGTCCATTCCTGATACCATCACCTC
CCATTTGCCAGACAGAACCTCTGGCTACAAAGCTCCAGAATGGAAGCCCACTGCCTGAGAGAGCTCATCCAGAAGTA
AATGGAGACACCAAGTGGCACTCTTTCAAAAGTTATTATGGAATACCCTGTATGAAGGGAAGCCAGAATAGTCGTGT
GAGTCCTGACTTTACACAAGAAAGTAGAGGGTATTCCAAGTGTTTGCAAAATGGAGGAATAAAACGCACAGTTAGTG
AACCTTCTCTCTCTGGGCTCCTTCAGATCAAGAAATTGAAACAAGACCAAAAGGCTAATGGAGAAAGACGTAACTTC
GGGGTAAGCCAAGAAAGAAATCCAGGTGAAAGCAGTCAACCAAATGTCTCCGATTTGAGTGATAAGAAAGAATCTGT
GAGTTCTGTAGCCCAAGAAAATGCAGTTAAAGATTTCACCAGTTTTTCAACACATAACTGCAGTGGGCCTGAAAATC
CAGAGCTTCAGATTCTGAATGAGCAGGAGGGGAAAAGTGCTAATTACCATGACAAGAACATTGTATTACTTAAAAAC
AAGGCAGTGCTAATGCCTAATGGTGCTACAGTTTCTGCCTCTTCCGTGGAACACACACATGGTGAACTCCTGGAAAA
AACACTGTCTCAATATTATCCAGATTGTGTTTCCATTGCGGTGCAGAAAACCACATCTCACATAAATGCCATTAACA
GTCAGGCTACTAATGAGTTGTCCTGTGAGATCACTCACCCATCGCATACCTCAGGGCAGATCAATTCCGCACAGACC
TCTAACTCTGAGCTGCCTCCAAAGCCAGCTGCAGTGGTGAGTGAGGCCTGTGATGCTGATGATGCTGATAATGCCAG
TAAACTAGCTGCAATGCTAAATACCTGTTCCTTTCAGAAACCAGAACAACTACAACAACAAAAATCAGTTTTTGAGA
TATGCCCATCTCCTGCAGAAAATAACATCCAGGGAACCACAAAGCTAGCGTCTGGTGAAGAATTCTGTTCAGGTTCC
AGCAGCAATTTGCAAGCTCCTGGTGGCAGCTCTGAACGGTATTTAAAACAAAATGAAATGAATGGTGCTTACTTCAA
GCAAAGCTCAGTGTTCACTAAGGATTCCTTTTCTGCCACTACCACACCACCACCATCACAATTGCTTCTTTCTC

CCCCTCCTCCTCTTCCACAGGTTCCTCAGCTTCCTTCAGAAGGAAAAAGCACTCTGAATGGTGGAGTTTTAGAAGAA
CACCACCACTACCCCAACCAAAGTAACACAACACTTTTAAGGGAAGTGAAAATAGAGGGTAAACCTGAGGCACCACC
TTCCCAGAGTCCTAATCCATCTACACATGTATGCAGCCCTTCTCCGATGCTTTCTGAAAGGCCTCAGAATAATTGTG
TGAACAGGAATGACATACAGACTGCAGGGACAATGACTGTTCCATTGTGTTCTGAGAAAACAAGACCAATGTCAGAA
CACCTCAAGCATAACCCACCAATTTTTGGTAGCAGTGGAGAGCTACAGGACAACTGCCAGCAGTTGATGAGAAACAA
AGAGCAAGAGATTCTGAAGGGTCGAGACAAGGAGCAAACACGAGATCTTGTGCCCCCAACACAGCACTATCTGAAAC
CAGGATGGATTGAATTGAAGGCCCCTCGTTTTCACCAAGCGGAATCCCATCTAAAACGTAATGAGGCATCACTGCCA
TCAATTCTTCAGTATCAACCCAATCTCTCCAATCAAATGACCTCCAAACAATACACTGGAAATTCCAACATGCCTGG
GGGGCTCCCAAGGCAAGCTTACACCCAGAAAACAACACAGCTGGAGCACAAGTCACAAATGTACCAAGTTGAAATGA
ATCAAGGGCAGTCCCAAGGTACAGTGGACCAACATCTCCAGTTCCAAAAACCCTCACACCAGGTGCACTTCTCCAAA
ACAGACCATTTACCAAAAGCTCATGTGCAGTCACTGTGTGGCACTAGATTTCATTTTCAACAAAGAGCAGATTCCCA
AACTGAAAAACTTATGTCCCCAGTGTTGAAACAGCACTTGAATCAACAGGCTTCAGAGACTGAGCCATTTTCAAACT
CACACCTTTTGCAACATAAGCCTCATAAACAGGCAGCACAAACACAACCATCCCAGAGTTCACATCTCCCTCAAAAC
CAGCAACAGCAGCAAAAATTACAAATAAAGAATAAAGAGGAAATACTCCAGACTTTTCCTCACCCCCAAAGCAACAA
TGATCAGCAAAGAGAAGGATCATTCTTTGGCCAGACTAAAGTGGAAGAATGTTTTCATGGTGAAAATCAGTATTCAA
AATCAAGCGAGTTCGAGACTCATAATGTCCAAATGGGACTGGAGGAAGTACAGAATATAAATCGTAGAAATTCCCCT
TATAGTCAGACCATGAAATCAAGTGCATGCAAAATACAGGTTTCTTGTTCAAACAATACACACCTAGTTTCAGAGAA
TAAAGAACAGACTACACATCCTGAACTTTTTGCAGGAAACAAGACCCAAAACTTGCATCACATGCAATATTTTCCAA
ATAATGTGATCCCAAAGCAAGATCTTCTTCACAGGTGCTTTCAAGAACAGGAGCAGAAGTCACAACAAGCTTCAGTT
CTACAGGGATATAAAAATAGAAACCAAGATATGTCTGGTCAACAAGCTGCGCAACTTGCTCAGCAAAGGTACTTGAT
ACATAACCATGCAAATGTTTTTCCTGTGCCTGACCAGGGAGGAAGTCACACTCAGACCCCTCCCCAGAAGGACACTC
AAAAGCATGCTGCTCTAAGGTGGCATCTCTTACAGAAGCAAGAACAGCAGCAAACACAGCAACCCCAAACTGAGTCT
TGCCATAGTCAGATGCACAGGCCAATTAAGGTGGAACCTGGATGCAAGCCACATGCCTGTATGCACACAGCACCACC
AGAAAACAAAACATGGAAAAAGGTAACTAAGCAAGAGAATCCACCTGCAAGCTGTGATAATGTGCAGCAAAAGAGCA
TCATTGAGACCATGGAGCAGCATCTGAAGCAGTTTCACGCCAAGTCGTTATTTGACCATAAGGCTCTTACTCTCAAA
TCACAGAAGCAAGTAAAAGTTGAAATGTCAGGGCCAGTCACAGTTTTGACTAGACAAACCACTGCTGCAGAACTTGA
TAGCCACACCCCAGCTTTAGAGCAGCAAACAACTTCTTCAGAAAAGACACCAACCAAAAGAACAGCTGCTTCTGTTC
TCAATAATTTTATAGAGTCACCTTCCAAATTACTAGATACTCCTATAAAAAATTTATTGGATACACCTGTCAAGACT
CAATATGATTTCCCATCTTGCAGATGTGTAGAGCAAATTATTGAAAAAGATGAAGGTCCTTTTTATACCCATCTAGG
AGCAGGTCCTAATGTGGCAGCTATTAGAGAAATCATGGAAGAAAGGTTTGGACAGAAGGGTAAAGCTATTAGGATTG
AAAGAGTCATCTATACTGGTAAAGAAGGCAAAAGTTCTCAGGGATGTCCTATTGCTAAGTGGGTGGTTCGCAGAAGC
AGCAGTGAAGAGAAGCTACTGTGTTTGGTGCGGGAGCGAGCTGGCCACACCTGTGAGGCTGCAGTGATTGTGATTCT
CATCCTGGTGTGGGAAGGAATCCCGCTGTCTCTGGCTGACAAACTCTACTCGGAGCTTACCGAGACGCTGAGGAAAT
ACGGCACGCTCACCAATCGCCGGTGTGCCTTGAATGAAGAGAGAACTTGCGCCTGTCAGGGGCTGGATCCAGAAACC
TGTGGTGCCTCCTTCTCTTTTGGTTGTTCATGGAGCATGTACTACAATGGATGTAAGTTTGCCAGAAGCAAGATCCC
AAGGAAGTTTAAGCTGCTTGGGGATGACCCAAAAGAGGAAGAGAAACTGGAGTCTCATTTGCAAAACCTGTCCACTC
TTATGGCACCAACATATAAGAAACTTGCACCTGATGCATATAATAATCAGATTGAATATGAACACAGAGCACCAGAG
TGCCGTCTGGGTCTGAAGGAAGGCCGTCCATTCTCAGGGGTCACTGCATGTTTGGACTTCTGTGCTCATGCCCACAG
AGACTTGCACAACATGCAGAATGGCAGCACATTGGTATGCACTCTCACTAGAGAAGACAATCGAGAATTTGGAGGAA

AACCTGAGGATGAGCAGCTTCACGTTCTGCCTTTATACAAAGTCTCTGACGTGGATGAGTTTGGGAGTGTGGAAGCT
CAGGAGGAGAAAAAACGGAGTGGTGCCATTCAGGTACTGAGTTCTTTTCGGCGAAAAGTCAGGATGTTAGCAGAGCC
AGTCAAGACTTGCCGACAAAGGAAACTAGAAGCCAAGAAAGCTGCAGCTGAAAAGCTTTCCTCCCTGGAGAACAGCT
CAAATAAAAATGAAAAGGAAAAGTCAGCCCCATCACGTACAAAACAAACTGAAAACGCAAGCCAGGCTAAACAGTTG
GCAGAACTTTTGCGACTTTCAGGACCAGTCATGCAGCAGTCCCAGCAGCCCCAGCCTCTACAGAAGCAGCCACCACA
GCCCCAGCAGCAGCAGAGACCCCAGCAGCAGCAGCCACATCACCCTCAGACAGAGTCTGTCAACTCTTATTCTGCTT
CTGGATCCACCAATCCATACATGAGACGGCCCAATCCAGTTAGTCCTTATCCAAACTCTTCACACACTTCAGATATC
TATGGAAGCACCAGCCCTATGAACTTCTATTCCACCTCATCTCAAGCTGCAGGTTCATATTTGAATTCTTCTAATCC
CATGAACCCTTACCCTGGGCTTTTGAATCAGAATACCCAATATCCATCATATCAATGCAATGGAAACCTATCAGTGG
ACAACTGCTCCCCATATCTGGGTTCCTATTCTCCCCAGTCTCAGCCGATGGATCTGTATAGGTATCCAAGCCAAGAC
CCTCTGTCTAAGCTCAGTCTACCACCCATCCATACACTTTACCAGCCAAGGTTTGGAAATAGCCAGAGTTTTACATC
TAAATACTTAGGTTATGGAAACCAAAATATGCAGGGAGATGGTTTCAGCAGTTGTACCATTAGACCAAATGTACATC
ATGTAGGGAAATTGCCTCCTTATCCCACTCATGAGATGGATGGCCACTTCATGGGAGCCACCTCTAGATTACCACCC
AATCTGAGCAATCCAAACATGGACTATAAAAATGGTGAACATCATTCACCTTCTCACATAATCCATAACTACAGTGC
AGCTCCGGGCATGTTCAACAGCTCTCTTCATGCCCTGCATCTCCAAAACAAGGAGAATGACATGCTTTCCCACACAG
CTAATGGGTTATCAAAGATGCTTCCAGCTCTTAACCATGATAGAACTGCTTGTGTCCAAGGAGGCTTACACAAATTA
AGTGATGCTAATGGTCAGGAAAAGCAGCCATTGGCACTAGTCCAGGGTGTGGCTTCTGGTGCAGAGGACAACGATGA
GGTCTGGTCAGACAGCGAGCAGAGCTTTCTGGATCCTGACATTGGGGGAGTGGCCGTGGCTCCAACTCATGGGTCAA
TTCTCATTGAGTGTGCAAAGCGTGAGCTGCATGCCACAACCCCTTTAAAGAATCCCAATAGGAATCACCCCACCAGG
ATCTCCCTCGTCTTTTACCAGCATAAGAGCATGAATGAGCCAAAACATGGCTTGGCTCTTTGGGAAGCCAAAATGGC
TGAAAAAGCCCGTGAGAAAGAGGAAGAGTGTGAAAAGTATGGCCCAGACTATGTGCCTCAGAAATCCCATGGCAAAA
AAGTGAAACGGGAGCCTGCTGAGCCACATGAAACTTCAGAGCCCACTTACCTGCGTTTCATCAAGTCTCTTGCCGAA
AGGACCATGTCCGTGACCACAGACTCCACAGTAACTACATCTCCATATGCCTTCACTCGGGTCACAGGGCCTTACAA
CAGATATATATGAAGATATATATGATATCACCCCCTTTTGTTGGTTACCTCACTTGAAAAGACCACAACCAACCTGT
CAGTAGTATAGTTCTCATGACGTGGGCAGTGGGGAAAGGTCACAGTATTCATGACAAATGTGGTGGGAAAAACCTCA
GCTCACCAGCAACAAAAGAGGTTATCTTACCATAGCACTTAATTTTCACTGGCTCCCAAGTGGTCACAGATGGCATC
TAGGAAAAGACCAAAGCATTCTATGCAAAAAGAAGGTGGGGAAGAAAGTGTTCCGCAATTTACATTTTTAAACACTG
GTTCTATTATTGGACGAGATGATATGTAAATGTGATCCCCCCCCCCCGCTTACAACTCTACACATCTGTGACCACTT
TTAATAATATCAAGTTTGCATAGTCATGGAACACAAATCAAACAAGTACTGTAGTATTACAGTGACAGGAATCTTAA
AATACCATCTGGTGCTGAATATATGATGTACTGAAATACTGGAATTATGGCTTTTTGAAATGCAGTTTTTACTGTAA
TCTTAACTTTTATTTATCAAAATAGCTACAGGAAACATGAATAGCAGGAAAACACTGAATTTGTTTGGATGTTCTAA
GAAATGGTGCTAAGAAAATGGTGTCTTTAATAGCTAAAAATTTAATGCCTTTATATCATCAAGATGCTATCAGTGTA
CTCCAGTGCCCTTGAATAATAGGGGTACCTTTTCATTCAAGTTTTTATCATAATTACCTATTCTTACACAAGCTTAG
TTTTTAAAATGTGGACATTTTAAAGGCCTCTGGATTTTGCTCATCCAGTGAAGTCCTTGTAGGACAATAAACGTATA
TATGTACATATATACACAAACATGTATATGTGCACACACATGTATATGTATAAATATTTTAAATGGTGTTTTAGAAG
CACTTTGTCTACCTAAGCTTTGACAACTTGAACAATGCTAAGGTACTGAGATGTTTAAAAAACAAGTTTACTTTCAT
TTTAGAATGCAAAGTTGATTTTTTTAAGGAAACAAAGAAAGCTTTTAAAATATTTTTGCTTTTAGCCATGCATCTGC
TGATGAGCAATTGTGTCCATTTTTAACACAGCCAGTTAAATCCACCATGGGGCTTACTGGATTCAAGGGAATACGTT
AGTCCACAAAACATGTTTTCTGGTGCTCATCTCACATGCTATACTGTAAAACAGTTTTATACAAAATTGTATGACAA

GTTCATTGCTCAAAAATGTACAGTTTTAAGAATTTTCTATTAACTGCAGGTAATAATTAGCTGCATGCTGCAGACTC
AACAAAGCTAGTTCACTGAAGCCTATGCTATTTTATGGATCATAGGCTCTTCAGAGAACTGAATGGCAGTCTGCCTT
TGTGTTGATAATTATGTACATTGTGACGTTGTCATTTCTTAGCTTAAGTGTCCTCTTTAACAAGAGGATTGAGCAGA
CTGATGCCTGCATAAGATGAATAAACAGGGTTAGTTCCATGTGAATCTGTCAGTTAAAAAGAAACAAAAACAGGCAG
CTGGTTTGCTGTGGTGGTTTTAAATCATTAATTTGTATAAAGAAGTGAAAGAGTTGTATAGTAAATTAAATTGTAAA
CAAAACTTTTTTAATGCAATGCTTTAGTATTTTAGTACTGTAAAAAAATTAAATATATACATATATATATATATATA
TATATATATATATATGAGTTTGAAGCAGAATTCACATCATGATGGTGCTACTCAGCCTGCTACAAATATATCATAAT
GTGAGCTAAGAATTCATTAAATGTTTGAGTGATGTTCCTACTTGTCATATACCTCAACACTAGTTTGGCAATAGGAT
ATTGAACTGAGAGTGAAAGCATTGTGTACCATCATTTTTTTCCAAGTCCTTTTTTTTATTGTTAAAAAAAAAAGCAT
ACCTTTTTTCAATACTTGATTTCTTAGCAAGTATAACTTGAACTTCAACCTTTTTGTTCTAAAAATTCAGGGATATT
TCAGCTCATGCTCTCCCTATGCCAACATGTCACCTGTGTTTATGTAAAATTGTTGTAGGTTAATAAATATATTCTTT
GTCAGGGATTTAACCCTTTTATTTTGAATCCCTTCTATTTTACTTGTACATGTGCTGATGTAACTAAAACTAATTTT
GTAAATCTGTTGGCTCTTTTTATTGTAAAGAAAAGCATTTTAAAAGTTTGAGGAATCTTTTGACTGTTTCAAGCAGG
AAAAAAAAATTACATGAAAATAGAATGCACTGAGTTGATAAAGGGAAAAATTGTAAGGCAGGAGTTTGGCAAGTGGC
TGTTGGCCAGAGACTTACTTGTAACTCTCTAAATGAAGTTTTTTTGATCCTGTAATCACTGAAGGTACATACTCCAT
GTGGACTTCCCTTAAACAGGCAAACACCTACAGGTATGGTGTGCAACAGATTGTACAATTACATTTTGGCCTAAATA
CATTTTTGCTTACTAGTATTTAAAATAAATTCTTAATCAGAGGAGGCCTTTGGGTTTTATTGGTCAAATCTTTGTAA
GCTGGCTTTTGTCTTTTTAAAAAATTTCTTGAATTTGTGGTTGTGTCCAATTTGCAAACATTTCCAAAAATGTTTGC
TTTGCTTACAAACCACATGATTTTAATGTTTTTTGTATACCATAATATCTAGCCCCAAACATTTGATTACTACATGT
GCATTGGTGATTTTGATCATCCATTCTTAATATTTGATTTCTGTGTCACCTACTGTCATTTGTTAAACTGCTGGCCA
ACAAGAACAGGAAGTATAGTTTGGGGGGTTGGGGAGAGTTTACATAAGGAAGAGAAGAAATTGAGTGGCATATTGTA
AATATCAGATCTATAATTGTAAATATAAAACCTGCCTCAGTTAGAATGAATGGAAAGCAGATCTACAATTTGCTAAT
ATAGGAATATCAGGTTGACTATATAGCCATACTTGAAAATGCTTCTGAGTGGTGTCAACTTTACTTGAATGAATTTT
TCATCTTGATTGACGCACAGTGATGTACAGTTCACTTCTGAAGCTAGTGGTTAACTTGTGTAGGAAACTTTTGCAGT
TTGACACTAAGATAACTTCTGTGTGCATTTTTCTATGCTTTTTTAAAAACTAGTTTCATTTCATTTTCATGAGATGT
TTGGTTTATAAGATCTGAGGATGGTTATAAATACTGTAAGTATTGTAATGTTATGAATGCAGGTTATTTGAAAGCTG
TTTATTATTATATCATTCCTGATAATGCTATGTGAGTGTTTTTAATAAAATTTATATTTATTTAATGCACTCTAAGT
GTTGTCTTCCT

[0386] “转化生长因子受体2(TGFBRII)多肽”意为一种蛋白质,其与NCBI登录号ABG65632.1或其片段具有至少约85%的氨基酸序列一致性并且具有免疫抑制活性。示例性氨基酸序列提供如下。

[0387] >ABG65632.1转化生长因子β受体II[智人(Homo sapiens)]MGRGLLRGLWPLHIVLWT
RIASTIPPHVQKSVNNDMIVTDNNGAVKFPQLCKFCDVRFSTCDNQKSCMSNCSITSICEKPQEVCVAVWRKNDENI
TLETVCHDPKLPYHDFILEDAASPKCIMKEKKKPGETFFMCSCSSDECNDNIIFSEEYNTSNPDLLLVIFQVTGISL
LPPLGVAISVIIIFYCYRVNRQQKLSSTWETGKTRKLMEFSEHCAIILEDDRSDISSTCANNINHNTELLPIELDTL
VGKGRFAEVYKAKLKQNTSEQFETVAVKIFPYEEYASWKTEKDIFSDINLKHENILQFLTAEERKTELGKQYWLITA
FHAKGNLQEYLTRHVISWEDLRKLGSSLARGIAHLHSDHTPCGRPKMPIVHRDLKSSNILVKNDLTCCLCDFGLSLR
LDPTLSVDDLANSGQVGTARYMAPEVLESRMNLENVESFKQTDVYSMALVLWEMTSRCNAVGEVKDYEPPFGSKVRE
HPCVESMKDNVLRDRGRPEIPSFWLNHQGIQMVCETLTECWDHDPEARLTAQCVAERFSELEHLDRLSGRSCSEEKI

PEDGSLNTTK

[0388] “转化生长因子受体2(TGFBRII)多核苷酸”意为编码TGFBRII多肽的核酸分子。TGFBRII基因编码具有丝氨酸/苏氨酸激酶活性的跨膜蛋白。示例性TGFBRII核酸序列提供如下。

[0389] >M85079.1人类TGF-βII型受体mRNA,完全编码序列

[0390] GTTGGCGAGGAGTTTCCTGTTTCCCCCGCAGCGCTGAGTTGAAGTTGAGTGAGTCACTCGCGCGCACGG
AGCGACGACACCCCCGCGCGTGCACCCGCTCGGGACAGGAGCCGGACTCCTGTGCAGCTTCCCTCGGCCGCCGGGGG
CCTCCCCGCGCCTCGCCGGCCTCCAGGCCCCTCCTGGCTGGCGAGCGGGCGCCACATCTGGCCCGCACATCTGCGCT
GCCGGCCCGGCGCGGGGTCCGGAGAGGGCGCGGCGCGGAGCGCAGCCAGGGGTCCGGGAAGGCGCCGTCCGTGCGCT
GGGGGCTCGGTCTATGACGAGCAGCGGGGTCTGCCATGGGTCGGGGGCTGCTCAGGGGCCTGTGGCCGCTGCACATC
GTCCTGTGGACGCGTATCGCCAGCACGATCCCACCGCACGTTCAGAAGTCGGTTAATAACGACATGATAGTCACTGA
CAACAACGGTGCAGTCAAGTTTCCACAACTGTGTAAATTTTGTGATGTGAGATTTTCCACCTGTGACAACCAGAAAT
CCTGCATGAGCAACTGCAGCATCACCTCCATCTGTGAGAAGCCACAGGAAGTCTGTGTGGCTGTATGGAGAAAGAAT
GACGAGAACATAACACTAGAGACAGTTTGCCATGACCCCAAGCTCCCCTACCATGACTTTATTCTGGAAGATGCTGC
TTCTCCAAAGTGCATTATGAAGGAAAAAAAAAAAGCCTGGTGAGACTTTCTTCATGTGTTCCTGTAGCTCTGATGAGT
GCAATGACAACATCATCTTCTCAGAAGAATATAACACCAGCAATCCTGACTTGTTGCTAGTCATATTTCAAGTGACA
GGCATCAGCCTCCTGCCACCACTGGGAGTTGCCATATCTGTCATCATCATCTTCTACTGCTACCGCGTTAACCGGCA
GCAGAAGCTGAGTTCAACCTGGGAAACCGGCAAGACGCGGAAGCTCATGGAGTTCAGCGAGCACTGTGCCATCATCC
TGGAAGATGACCGCTCTGACATCAGCTCCACGTGTGCCAACAACATCAACCACAACACAGAGCTGCTGCCCATTGAG
CTGGACACCCTGGTGGGGAAAGGTCGCTTTGCTGAGGTCTATAAGGCCAAGCTGAAGCAGAACACTTCAGAGCAGTT
TGAGACAGTGGCAGTCAAGATCTTTCCCTATGAGGAGTATGCCTCTTGGAAGACAGAGAAGGACATCTTCTCAGACA
TCAATCTGAAGCATGAGAACATACTCCAGTTCCTGACGGCTGAGGAGCGGAAGACGGAGTTGGGGAAACAATACTGG
CTGATCACCGCCTTCCACGCCAAGGGCAACCTACAGGAGTACCTGACGCGGCATGTCATCAGCTGGGAGGACCTGCG
CAAGCTGGGCAGCTCCCTCGCCCGGGGGATTGCTCACCTCCACAGTGATCACACTCCATGTGGGAGGCCCAAGATGC
CCATCGTGCACAGGGACCTCAAGAGCTCCAATATCCTCGTGAAGAACGACCTAACCTGCTGCCTGTGTGACTTTGGG
CTTTCCCTGCGTCTGGACCCTACTCTGTCTGTGGATGACCTGGCTAACAGTGGGCAGGTGGGAACTGCAAGATACAT
GGCTCCAGAAGTCCTAGAATCCAGGATGAATTTGGAGAATGCTGAGTCCTTCAAGCAGACCGATGTCTACTCCATGG
CTCTGGTGCTCTGGGAAATGACATCTCGCTGTAATGCAGTGGGAGAAGTAAAAGATTATGAGCCTCCATTTGGTTCC
AAGGTGCGGGAGCACCCCTGTGTCGAAAGCATGAAGGACAACGTGTTGAGAGATCGAGGGCGACCAGAAATTCCCAG
CTTCTGGCTCAACCACCAGGGCATCCAGATGGTGTGTGAGACGTTGACTGAGTGCTGGGACCACGACCCAGAGGCCC
GTCTCACAGCCCAGTGTGTGGCAGAACGCTTCAGTGAGCTGGAGCATCTGGACAGGCTCTCGGGGAGGAGCTGCTCG
GAGGAGAAGATTCCTGAAGACGGCTCCCTAAACACTACCAAATAGCTCTTATGGGGCAGGCTGGGCATGTCCAAAGA
GGCTGCCCCTCTCACCAAA

[0391] “具有Ig和ITIM结构域的T细胞免疫受体(TIGIT)多肽”意为一种蛋白质,其与NCBI登录号ACD74757.1或其片段具有至少约85%的氨基酸序列一致性并且具有免疫调节活性。示例性TIGIT氨基酸序列提供如下。

[0392] >ACD74757.1具有Ig和ITIM结构域的T细胞免疫受体[智人(Homo sapiens)]

[0393] MRWCLLLIWAQGLRQAPLASGMMTGTIETTGNISAEKGGSIILQCHLSSTTAQVTQVNWEQQDQLLAIC
NADLGWHISPSFKDRVAPGPGLGLTLQSLTVNDTGEYFCIYHTYPDGTYTGRIFLEVLESSVAEHGARFQIPLLGAM

AATLVVICTAVIVVVALTRKKKALRIHSVEGDLRRKSAGQEEWSPSAPSPPGSCVQAEAAPAGLCGEQRGEDCAELH
DYFNVLSYRSLGNCSFFTETG

[0394] “具有Ig和ITIM结构域的T细胞免疫受体(TIGIT)多核苷酸”意为编码TIGIT多肽的核酸。TIGIT基因编码与瘤形成和T细胞衰竭相关的抑制性免疫受体。示例性核酸序列提供如下。

[0395] >EU675310.1具有Ig和ITIM结构域的智人(Homo sapiens)T细胞免疫受体(TIGIT)mRNA,完全编码序列

[0396] CGTCCTATCTGCAGTCGGCTACTTTCAGTGGCAGAAGAGGCCACATCTGCTTCCTGTAGGCCCTCTGGG
CAGAAGCATGCGCTGGTGTCTCCTCCTGATCTGGGCCCAGGGGCTGAGGCAGGCTCCCCTCGCCTCAGGAATGATGA
CAGGCACAATAGAAACAACGGGGAACATTTCTGCAGAGAAAGGTGGCTCTATCATCTTACAATGTCACCTCTCCTCC
ACCACGGCACAAGTGACCCAGGTCAACTGGGAGCAGCAGGACCAGCTTCTGGCCATTTGTAATGCTGACTTGGGGTG
GCACATCTCCCCATCCTTCAAGGATCGAGTGGCCCCAGGTCCCGGCCTGGGCCTCACCCTCCAGTCGCTGACCGTGA
ACGATACAGGGGAGTACTTCTGCATCTATCACACCTACCCTGATGGGACGTACACTGGGAGAATCTTCCTGGAGGTC
CTAGAAAGCTCAGTGGCTGAGCACGGTGCCAGGTTCCAGATTCCATTGCTTGGAGCCATGGCCGCGACGCTGGTGGT
CATCTGCACAGCAGTCATCGTGGTGGTCGCGTTGACTAGAAAGAAGAAAGCCCTCAGAATCCATTCTGTGGAAGGTG
ACCTCAGGAGAAAATCAGCTGGACAGGAGGAATGGAGCCCCAGTGCTCCCTCACCCCCAGGAAGCTGTGTCCAGGCA
GAAGCTGCACCTGCTGGGCTCTGTGGAGAGCAGCGGGGAGAGGACTGTGCCGAGCTGCATGACTACTTCAATGTCCT
GAGTTACAGAAGCCTGGGTAACTGCAGCTTCTTCACAGAGACTGGTTAGCAACCAGAGGCATCTTCTGG

[0397] 如本文所用,“转导”意为经由病毒载体将基因或遗传物质转移到细胞。

[0398] 如本文所用,“转化”是指通过引入外源性核酸将基因改变引入细胞内的过程。

[0399] “转染”是指经由化学或物理手段将基因或遗传物质转移到细胞。

[0400] “易位”意为核酸链段在非同源染色体之间的重排。

[0401] “跨膜结构域”意为一种氨基酸序列,其插入到脂质双层,诸如细胞或病毒或病毒样颗粒的脂质双层内。跨膜结构域可用于将感兴趣的蛋白质(例如,CAR)锚定至膜。跨膜结构域可以源自天然来源或源自合成来源。若来源是天然的,则该结构域可以源自任何膜结合蛋白或跨膜蛋白。用于所公开的CAR的跨膜结构域可包括T细胞受体的α、β或ζ链、CD28、CD3ε、CD45、CD4、CD5、CD8、CD9、CD16、CD22、CD33,CD37,CD64,CD80,CD86,CD134,CD137,CD154的至少跨膜区。在一些实施方案中,跨膜结构域源自CD4、CD8α、CD28和CD3ζ。在一些实施方案中,跨膜结构域是CD8α铰链和跨膜结构域。

[0402] 如本文所用,术语“处理”、“治疗”等是指减轻或缓解疾病或疾患和/或与其有关的症状或者获得所希望的药理学和/或生理学效果。应理解,尽管没有排除,但治疗病变或病症并不需要完全消除该病变、病症或与之相关的症状。在一些实施方案中,该效果是治疗性的,即,不限于,该效果部分地或完全地减轻、消除、废除、削弱、减缓、降低疾病和/或可归因于该疾病的负面症候的强度或治愈疾病和/或可归因于该疾病的负面症候。在一些实施方案中,该效果是预防性的,即,该效果保护或阻止疾病或病症的出现或复发。为此,本公开的方法包括给药治疗有效量的本文所述的组合物。

[0403] 术语“尿嘧啶糖基化酶抑制剂”或“UGI”意为抑制尿嘧啶切除修复系统的药剂。在一种实施方案中,药剂是一种蛋白质或其片段,其结合宿主尿嘧啶-DNA糖苷酶并阻止尿嘧啶残基从DNA去除。在一种实施方案中,UGI是能够抑制尿嘧啶-DNA糖苷酶碱基切除修复酶

的蛋白质、其片段或结构域。在一些实施方案中,UGI结构域包含野生型UGI或其经修饰的版本。在一些实施方案中,UGI结构域包含下文详述的示例性氨基酸序列的片段。在一些实施方案中,UGI片段包含氨基酸序列,该氨基酸序列包含下文提供的示例性UGI序列的至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%。在一些实施方案中,UGI包含氨基酸序列,该氨基酸序列与下文详述的示例性UGI氨基酸序列或其片段同源。在一种实施方案中,UGI或其部分与下文详述的野生型UGI或UGI序列或其片段至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%、至少99.5%、至少99.9%或100%相同。示例性UGI包含如下的氨基酸序列:

[0404] >splP14739IUNGI_BPPB2尿嘧啶-DNA糖基化酶抑制剂

[0405] MTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVM LLTSDAPEYKPWALVIQDSNGENKIKML。术语“载体”是指将核酸序列引入细胞内以得到转化细胞的手段。载体包括质粒、转座子、噬菌体、病毒V脂质体和附加体。“表达载体”是包含待在接纳者细胞内表达的核酸序列的核酸序列。表达载体可包括额外的核酸序列以促进及或促使所引入的序列诸如起始子、终止子、启动子和分泌序列的表达。

[0406] “T细胞受体ζ链相关蛋白激酶70(ZAP70)多肽”意为一种蛋白质,其与NCBI登录号AAH53878.1具有至少约85%的氨基酸序列一致性并且具有激酶活性。示例性氨基酸序列提供如下。

[0407] >AAH53878.1ζ链(TCR)相关蛋白激酶70kDa[智人(Homo sapiens)]

[0408] MPDPAAHLPFFYGSISRAEAEEHLKLAGMADGLFLLRQCLRSLGGYVLSLVHDVRFHHFPIERQLNGTYAIAGGKAHCGPAELCEFYSRDPDGLPCNLRKPCNRPSGLEPQPGVFDCLRDAMVRDYVRQTWKLEGEALEQAIISQAPQVEKLIATTAHERMPWYHSSLTREEAERKLYSGAQTDGKFLLRPRKEQGTYALSLIYGKTVYHYLISQDKAGKYCIPEGTKFDTLWQLVEYLKLKADGLIYCLKEACPNSSASNASGAAAPTLPAHPSTLTHPQRRIDTLNSDGYTPEPARITSPDKPRPMPMDTSVYESPYSDPEELKDKKLFLKRDNLLIADIELGCGNFGSVRQGVYRMRKKQIDVAIKVLKQGTEKADTEEMMREAQIMHQLDNPYIVRLIGVCQAEALMLVMEMAGGGPLHKFLVGKREEIPVSNVAELLHQVSMGMKYLEEKNFVHRDLAARNVLLVNRHYAKISDFGLSKALGADDSYYTARSAGKWPLKWYAPECINFRKFSSRSDVWSYGVTMWEALSYGQKPYKKMKGPEVMAFIEQGKRMECPPECPPELYALMSDCWIYKWEDRPDFLTVEQRMRACYYSLASKVEGPPGSTQKAEAACA

[0409] “T细胞受体ζ链相关蛋白激酶70(ZAP70)”意为编码ZAP70多肽的核酸。ZAP70基因编码牵涉到T细胞发育和淋巴细胞激活中的酪氨酸激酶。功能性ZAP10的缺失可导致以CD8+T细胞的缺乏为特征的重症综合性免疫缺陷。示例性ZAP70核酸序列提供如下。

[0410] >BC053878.1智人(Homo sapiens)ζ链(TCR)相关蛋白激酶70kDa,mRNA(cDNA克隆体MGC:61743IMAGE:5757161),完全编码序列

[0411] GCTTGCCGGAGCTCAGCAGACACCAGGCCTTCCGGGCAGGCCTGGCCCACCGTGGGCCTCAGAGCTGCTGCTGGGGCATTCAGAACCGGCTCTCCATTGGCATTGGGACCAGAGACCCCGCAAGTGGCCTGTTTGCCTGGACATCCACCTGTACGTCCCCAGGTTTCGGGAGGCCCAGGGGCGATGCCAGACCCCGCGGCGCACCTGCCCTTCTTCTACGGCAGCATCTCGCGTGCCGAGGCCGAGGAGCACCTGAAGCTGGCGGGCATGGCGGACGGGCTCTTCCTGCTGCGCCAGTGCCTGCGCTCGCTGGGCGGCTATGTGCTGTCGCTCGTGCACGATGTGCGCTTCCACCACTTTCCCATCGAGCGCCAGCTCAACGGCACCTACGCCATTGCCGGCGGCAAAGCGCACTGTGGACCGGCAGAGCTCTGCGAGTTCTACTCGCGCGACC

CCGACGGGCTGCCCTGCAACCTGCGCAAGCCGTGCAACCGGCCGTCGGGCCTCGAGCCGCAGCCGGGGGTCTTCGAC
TGCCTGCGAGACGCCATGGTGCGTGACTACGTGCGCCAGACGTGGAAGCTGGAGGGCGAGGCCCTGGAGCAGGCCAT
CATCAGCCAGGCCCCGCAGGTGGAGAAGCTCATTGCTACGACGGCCCACGAGCGGATGCCCTGGTACCACAGCAGCC
TGACGCGTGAGGAGGCCGAGCGCAAACTTTACTCTGGGGCGCAGACCGACGGCAAGTTCCTGCTGAGGCCGCGGAAG
GAGCAGGGCACATACGCCCTGTCCCTCATCTATGGGAAGACGGTGTACCACTACCTCATCAGCCAAGACAAGGCGGG
CAAGTACTGCATTCCCGAGGGCACCAAGTTTGACACGCTCTGGCAGCTGGTGGAGTATCTGAAGCTGAAGGCGGACG
GGCTCATCTACTGCCTGAAGGAGGCCTGCCCCAACAGCAGTGCCAGCAACGCCTCAGGGGCTGCTGCTCCCACACTC
CCAGCCCACCCATCCACGTTGACTCATCCTCAGAGACGAATCGACACCCTCAACTCAGATGGATACACCCCTGAGCC
AGCACGCATAACGTCCCCAGACAAACCGCGGCCGATGCCCATGGACACGAGCGTGTATGAGAGCCCCTACAGCGACC
CAGAGGAGCTCAAGGACAAGAAGCTCTTCCTGAAGCGCGATAACCTCCTCATAGCTGACATTGAACTTGGCTGCGGC
AACTTTGGCTCAGTGCGCCAGGGCGTGTACCGCATGCGCAAGAAGCAGATCGACGTGGCCATCAAGGTGCTGAAGCA
GGGCACGGAGAAGGCAGACACGGAAGAGATGATGCGCGAGGCGCAGATCATGCACCAGCTGGACAACCCCTACATCG
TGCGGCTCATTGGCGTCTGCCAGGCCGAGGCCCTCATGCTGGTCATGGAGATGGCTGGGGGCGGGCCGCTGCACAAG
TTCCTGGTCGGCAAGAGGGAGGAGATCCCTGTGAGCAATGTGGCCGAGCTGCTGCACCAGGTGTCCATGGGGATGAA
GTACCTGGAGGAGAAGAACTTTGTGCACCGTGACCTGGCGGCCCGCAACGTCCTGCTGGTTAACCGGCACTACGCCA
AGATCAGCGACTTTGGCCTCTCCAAAGCACTGGGTGCCGACGACAGCTACTACACTGCCCGCTCAGCAGGGAAGTGG
CCGCTCAAGTGGTACGCACCCGAATGCATCAACTTCCGCAAGTTCTCCAGCCGCAGCGATGTCTGGAGCTATGGGGT
CACCATGTGGGAGGCCTTGTCCTACGGCCAGAAGCCCTACAAGAAGATGAAAGGGCCGGAGGTCATGGCCTTCATCG
AGCAGGGCAAGCGGATGGAATGCCCACCAGAGTGTCCACCCGAACTGTACGCACTCATGAGTGACTGCTGGATCTAC
AAGTGGGAGGATCGCCCCGACTTCCTGACCGTGGAGCAGCGCATGCGAGCCTGTTACTACAGCCTGGCCAGCAAGGT
GGAAGGGCCCCCAGGCAGCACACAGAAGGCTGAGGCTGCCTGTGCCTGAGCTCCCGCTGCCCAGGGGAGCCCTCCAC
ACCGGCTCTTCCCCACCCTCAGCCCCACCCCAGGTCCTGCAGTCTGGCTGAGCCCTGCTTGGTTGTCTCCACACACA
GCTGGGCTGTGGTAGGGGGTGTCTCAGGCCACACCGGCCTTGCATTGCCTGCCTGGCCCCCTGTCCTCTCTGGCTGG
GGAGCAGGGAGGTCCGGGAGGGTGCGGCTGTGCAGCCTGTCCTGGGCTGGTGGCTCCCGGAGGGCCCTGAGCTGAGG
GCATTGCTTACACGGATGCCTTCCCCTGGGCCCTGACATTGGAGCCTGGGCATCCTCAGGTGGTCAGGCGTAGATCA
CCAGAATAAACCCAGCTTCCCTCTTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAA

[0412] 本文中的任何变量定义中对一系列化学基团的叙述包括该变量作为任何单一基团或所列基团的组合的定义。本文中对于变量或方面的实施方案的叙述包括该实施方案作为任何单一实施方案或与其他实施方案或其部分组合。

[0413] 本文所提供的任何组合物或方法可与本文所提供的任何其他组合物和方法中的一者或多者合用。

[0414] DNA编辑已成为一种可行的手段,通过在基因水平上校正致病突变来改变疾病状态。直到最近,全部DNA编辑平台的功能都是在特定的基因组位点诱导DNA双链断裂(DSB),并依靠内源性DNA修复途径以半随机方式确定产物结果,从而产生复杂的遗传产物群体。虽然通过同源定向修复(HDR)途径可以实现精确的、用户定义的修复结果,但在治疗上相关的细胞类型中,使用HDR进行高效修复面临许多挑战。在实践中,与竞争性、易出错的非同源末端连接途径相比,这一途径效率低下。再者,HDR严格限制在细胞周期的G1期和S期,从而阻止有丝分裂后细胞中DSB的精确修复。因此,在这些群体中,以用户定义的、可编程的方式高

效地改变基因组序列已被证明是困难的或不可能的。

[0415] 通过引用并入

[0416] 本说明书中提及的全部出版物、专利和专利申请通过引用以与各独立出版物、专利或专利申请被具体且独立地指明为通过引用并入相同的程度并入本文。

附图说明

[0417] 本发明的新颖特征在所附权利要求书中以特殊性详述。通过参考以下详述示例性实施方案的具体实施方式(其中使用了本发明的原理)以及附图,可获得对本发明的特征和优点的更佳理解:

[0418] 图1A至图1B是影响T细胞功能的三种蛋白质的例示性说明。图1A是TRAC蛋白的例示性说明,该蛋白是移植物抗宿主疾病的关键成分。图1B是B2M蛋白的例示性说明,该蛋白是存在于有核细胞上的MHC 1类抗原呈递复合体的成分,其可以被宿主的CD8+ T细胞识别。图1C是T细胞信号传导的例示性说明,该信号传导导致PDCD1基因的表达,并且所得PD-1蛋白发挥作用以抑制T细胞传导。

[0419] 图2是在碱基编辑后具有敲低的靶标基因表达的细胞的百分比的图。“EP”表示电穿孔。

[0420] 图3是在未经转导的细胞或经BE4碱基编辑系统或Cas9核酸酶转导的细胞中观察到的基因修饰类型的百分比的图。

[0421] 图4是描述通过细胞的百分比测得的靶标核苷酸修饰百分比的图,该细胞通过流式细胞术(FC)在经BE4和将BE4引导至剪接位点受体(SA)或供体(SD)的sgRNA转导的细胞中测定的靶标蛋白表达为阴性,或该细胞生成STOP密码子。对照细胞经虚拟电穿孔(EP)。

[0422] 图5是BE4系统中断剪接位点受体(SA)、剪接供体(SD)或生成STOP密码子的示意图。

[0423] 图6是总结用来中断靶标基因的sgRNA的脱靶结合位点的图表。

[0424] 图7是总结用BE4或Cas9编辑的表现出减少的蛋白质表达的细胞的百分比的流式细胞术(FC)数据的图。细胞被选通至B2M或CD3,后者是TRAC表达的指标。

[0425] 图8A是未经编辑的对照细胞的FACS数据的散点图。图8B是已经在B2M、TRAC和PD1基因座处编辑的细胞的FACS数据的散点图。

[0426] 图9是例示性说明本文所述的碱基编辑技术在修饰可能对CAR-T免疫疗法产生负面影响的特异性基因中的有效性的图。

[0427] 图10是描述用以检测并定量基因修饰和易位的微滴式数字PCR(ddPCR)方案的示意图。

[0428] 图11呈现两张图,显示从对使用BE4系统或Cas9系统编辑的细胞的下一代测序(NGS)分析或ddPCR生成的数据。

[0429] 图12是例示性说明Cbl-b在阻遏T细胞激活中充当的角色的原理图。

[0430] 图13是描述通过中断剪接位点进行Cbl-b敲低的效率的图。SA=剪接受体;SD=剪接供体;STOP-生成STOP密码子;2°Only=仅二抗;C373是指功能损失变体(C373R);RL1-A::APC-A=激光;ICS=细胞内染色。

[0431] 图14是例示性说明在T细胞中的GRIN2B和DNMT1基因中的Cas12b介导的插入缺失

率。EP表示电穿孔。

[0432] 图15是总结经由用bvCas12b和对于TRAC、GRIN2B和DNMT1具有特异性的向导RNA进行电穿孔(EP)来转导且针对CD3进行选通的细胞的荧光辅助细胞分选(FACS)数据的图。

[0433] 图16是经CAR-P2A-mCherry慢病毒转导的细胞的荧光辅助细胞分选数据的散点图,表明了CAR表达。

[0434] 图17是荧光辅助分选数据的散点图,表明了在经聚(1,8-辛二醇柠檬酸酯)(POC)慢病毒载体转导的细胞中的CAR表达。

[0435] 图18是显示BE4产生有效、持久的基因敲除且具有高产物纯度的图。

[0436] 图19A是代表性FACS分析,显示由于通过BE4或spCas9进行基因敲除导致的蛋白质表明表达的损失。图19B是显示通过BE4或spCas9进行的基因敲除造成B2M表面表达损失的图。

[0437] 图20是描述B2M、TRAC和PD-1靶标位点的定位的示意图。当B2M、TRAC和PD-1序列重组时,可检测到易位。

[0438] 图21是显示多重碱基编辑不显著损害细胞扩增的图。

[0439] 图22是显示BE4以与spCas9类似的上靶编辑效率和细胞表型生成经三重编辑的T细胞的图。

[0440] 图23描述流式细胞术分析,其显示经三重编辑的CD3-、B2M-、PD1-T细胞的生成。

[0441] 图24描述流式细胞术分析,其显示经BE4和Cas9编辑的细胞中的CAR表达。

[0442] 图25是显示CAR-T细胞杀死或抗原阳性细胞的图。

[0443] 图26是显示Cas12b和BE4可经配对以在T细胞中进行有效的多重编辑的图。

[0444] 图27是显示Cas12b可以通过在Cas12核酸酶和靶向基因座的sgRNA的存在下将编码CAR的双链DNA模板引入细胞而将嵌合抗原受体(CAR)的插入引导至该基因座内。

[0445] 图28A和图28B是显示用来沉默基因以工程化改造CAR-T细胞的碱基编辑路径的说明。图28A示出了用于CAR-T编辑的靶标。图28B显示使用碱基编辑器进行沉默的两个策略:用CBE创建终止密码子和用CBE中断剪接。

[0446] 图29是例示性漫画表示,在一些情况下,解决AML中的克隆性需要多种CAR-T。

[0447] 图30是例示性漫画表示,在一些情况下,消除T-ALL CAR-T组合的同类相杀需要多重编辑。

[0448] 图31A和图31B例示性说明测试对于T-ALL的四种同时碱基编辑的实验的结果;与核酸酶相比,CAR-T不影响产率。图31A例示性说明用于以下情况的四处T-ALL编辑竞赛中的理论产率:无电穿孔(无EP;深色圆圈)、仅EP(深色方块)、CBE变体1(浅色、正三角)、CBE变体2(浅色、倒三角)和Cas9(浅色菱形)。图31B例示性说明电穿孔后的细胞存活率。对于在24小时(hr)、48hr、72hr、96hr和168hr的每一组结果,结果从右至左显示为无EP、仅EP、CBE变体1、CBE变体2和Cas9。

[0449] 图32例示性说明超过90%的四处敲除是由BE4造成的。电穿孔以5M细胞规模执行,并且四重编辑在单个电穿孔(EP)步骤中进行。对于全部4个靶标,使用rBE4均实现了大于90%的编辑效率。对于每个结果组,数据从左至右为:PD1、CD7、TRAC和CD52。

[0450] 图33例示性说明碱基编辑未造成细胞产率的差异。数据为:无P(圆圈)、仅EP(方块)、rBE4(三角形)、ppBE4(倒三角)和Cas9(菱形)。

[0451] 图34例示性说明经编辑的CAR-T靶向模型肿瘤细胞上的CD3和CD7。数据为:UTD 1:1(圆圈)、7CAR8 1:1(方块)和3CAR2 1:1(三角)。

[0452] 图35是用于靶向T-AL肿瘤细胞的示例性CD7 CAR-T细胞的示意图。

[0453] 图36是流程图,描述用于治疗具有CD7 CAR-T细胞的患者的临床方案。

[0454] 图37A至图37C描述CD7 CAR-T细胞生产。图37A是流程图,描述用于生产TALL017异源融化的CD7 CAR-T细胞的方案。图37B是流程图,描述用于生产TALL083 CD7 CAR-T细胞的方案。图37C是荧光辅助细胞分选数据的散点图,表明TALL017 CAR-T细胞在融化后被高度激活。

[0455] 图38是描述通过下一代测序(NGS)测得的TALL017和TALL038 CD7CAR-T细胞中的总编辑数的图。

[0456] 图39A和图39B描述在融化后24小时,TALL017和TALL038 CD7CAR-T细胞中的TCRα/β、CD7和CD52的表达。图39A是TALL017和TALL038 CD7 CAR-T细胞的荧光辅助细胞分选数据的散点图。图39B是描述经由FACS测得的残余蛋白质表达的图。

[0457] 图40描述用于身份组的选通方案。

[0458] 图41是描述作为CD2+/-与CD56+/-细胞的混合物的最终细胞产物身份的图。

[0459] 图42是描述融化后24小时和48小时的CD7 CAR-T细胞表达。

[0460] 图43包括多张图,其描述在融化后,TALL038 CD7 CAR-T细胞中的CD25表达比TALL017 CD7 CAR-T细胞中低。

[0461] 图44是流程图,描述用于体外表征的基于CD7 CAR-T珠的效力方案。

[0462] 图45A和图45B描述TALL038 CD7 CAR-T细胞应答CD7抗原而释放IFNγ、TNFα和IL-2。图45A是描述TALL038 CD7 CAR-T细胞释放IFNγ的图。图45B是描述TALL038 CD7 CAR-T细胞释放TNFα、IL-10和IL-2的图。

[0463] 图46是示意图,描述用于测量肿瘤细胞的CAR导向T细胞杀死的共培养方案。

[0464] 图47A和图47B描述,与TALL017 CAR T相比,TALL038 CD7 CAR-T细胞在再次挑战时表现出增加的CCRF。图47A是描述主要挑战的图。图47B是描述次要挑战的图。

[0465] 图48是描述在CCRF移植后第10天/CD7 CAR-T治疗第-1天的生物发光辐射数据(均值,SEM)的图。总通量在线性尺度上测量。

[0466] 图49A和图49B描述在CCRF移植后第19天/CD7 CAR-T治疗后第8天的生物发光辐射数据(均值,SEM)。图49A是描述平均肿瘤负担的图。总通量在线性尺度上测量。图49B是描述平均小鼠重量的图。

[0467] 图50A和图50B描述在CCRF移植后第38天/CD7 CAR-T治疗后第27天的生物发光辐射数据(均值,SEM)。图50A是描述平均肿瘤负担的图。总通量在线性尺度上测量。图50B是描述平均肿瘤负担的图。总通量在对数尺度上测量。

[0468] 图51A和图51B描述在CCRF移植后第38天/CD7 CAR-T治疗后第27天的生物发光辐射数据(个体小鼠)。图51A是描述平均肿瘤负担的图。总通量在线性尺度上测量。图51B是描述平均肿瘤负担的图。总通量在对数尺度上测量。

[0469] 图52A和图52B描述在CCRF移植后第38天/CD7 CAR-T治疗后第27天的生物发光辐射数据(个体小鼠)。图52A包括描述平均肿瘤负担的图。总通量在线性尺度(上图)上和对数尺度(下)上测量。图52B包括描述平均肿瘤负担的图。总通量在线性尺度(上图)上和对数尺

度(下)上测量。

[0470] 图53是描述CD5 gRNA候选物的编辑效率的图。

[0471] 图54是描述经由CD5 gRNA候选物g103和g104与BE4组合的NGS取得的编辑效率的图。

[0472] 图55是描述示例性CD5 CAR构建体的生产的示意图。

[0473] 图56是描述来自使用CD5 gRNA候选物g103和g104进行的慢病毒载体(LVV)筛选的CAR表达的图。

[0474] 图57是描述基于铰链序列的CAR LVV表达的图。

[0475] 图58A至图58D是描述在活体成像细胞杀死测定中,经编辑的和未经编辑的构建体对于CD5+CCRF是完全细胞毒性的图。

[0476] 图59A和图59B描述使用sgRNA 103或sgRNA 104在CD5+CCRF-CEM白血病细胞系的存在下通过CD5 CAR LVV构建体生产细胞因子(IFNγ)的图。图59A(左)是描述以使用sgRNA 103编辑过的经CD5 CAR LVV(LV63-69)转导的T细胞单独地或与CCRF细胞组合地生产IFNγ的图。图59A(右)是描述以使用sgRNA 104编辑的经CD5 CAR LVV(LV63-69)转导的T细胞单独地或与CCRF细胞组合地生产IFNγ的图。图59A(下)是描述使用未经编辑的经CD5 CAR LVV(LV63-69)转导的T细胞或CCRF细胞生产IFNγ的图。图59B(上)是描述以未经编辑的或使用sgRNA 103或104编辑过的经CD5 CAR LVV(LV63-69)转导的T细胞单独地生产IFNγ的图。图59B(下)是描述以未经编辑的或使用sgRNA 103或104编辑过的经CD5 CAR LVV(LV63-69)转导的CCRF细胞生产IFNγ的图。

具体实施方式

[0477] 本发明提出经基因修饰的免疫细胞,所述免疫细胞具有增强的抗瘤形成活性、抗免疫抑制性和降低的引起移植物抗宿主反应或宿主抗移植物反应的风险,或其组合物。本发明还提出用于生产和使用这些经修饰的免疫效应细胞(例如,免疫效应细胞,诸如T细胞)的方法。

[0478] 在一种实施方案中,向患有移植物抗宿主疾病(GVHD)或具有发展出GVHD倾向的受试者给药CAR-T细胞,该细胞缺乏或具有降低水平的功能性TRAC。在一种实施方案中,向患有宿主抗移植物疾病(HVGD)或具有发展出HVGD倾向的受试者给药CAR-T细胞,该细胞缺乏或具有降低水平的β2微球蛋白(B2M)。

[0479] 使用如本文所述的包含胞苷脱氨酶或腺苷脱氨酶的碱基编辑器系统来实施对免疫效应细胞的修饰以表达嵌合抗原受体以及敲除或敲低特异性基因,从而减少其表达可能对免疫细胞功能造成的负面影响。

[0480] 自体的、患者来源的嵌合抗原受体-T细胞(CAR-T)疗法已经在治疗一些血液学癌症中显示了显著的疗效。虽然这些产物已经导致患者的显著临床受益,但对于生成个性化疗法的需求带来了巨大的制造挑战和经济负担。同种异体基因CAR-T疗法是为应对这些挑战而开发的潜在解决方案,具有与自体产品类似的临床功效,同时治疗很多具有源自单个健康供体的患者,从而实质性地降低商品成本和批次间的差异。

[0481] 大多数第一代同种异体基因CAR-T使用核酸酶将两个或更多个被靶向的基因组DNA双链断裂(DSB)引入靶标T细胞群体中,依靠易错的DNA修复来生成突变,这些突变以半

随机模式敲除靶标基因。此类基于核酸酶的基因敲除策略辅助降低CAR-T的移植物抗宿主疾病和宿主排异的风险。然而,同时引入多个DSB导致最终细胞产物含有大量基因组重排诸如经平衡和未经平衡的易位,和相对高峰度的局部重排(包括倒位和大的缺失)。再者,随着通过所引入的DSB作出的同时基因修饰数量增加,在经处理的细胞群中观察到可观的基因毒性。这有可能显著降低每次生产过程中的细胞扩张潜能,从而减少每个健康供体可接受治疗的患者数量。

[0482] 碱基编辑器(BE)是一类新兴基因编辑试剂,其能够高效地、用户定义地修饰靶标基因组DNA而不产生DSB。本文中,通过使用碱基编辑技术来减少或消除可检测的基因组重排并同时还改善细胞扩张,提出了一种生产同种异体CAR-T细胞的方法。如本文所示,与仅使用核酸酶的编辑策略相反,通过碱基编辑来同时修饰一个或多个(例如,一个、两个、三个、四个、五个、六个、七个、八个、九个、十个或更多个基因的基因座产生高效基因敲除而没有可检测的易位事件。

[0483] 在一些实施方案中,使用本文提供的碱基编辑组合物和方法修饰免疫细胞内的至少一个或多个基因或其调节元件。在一些实施方案中,至少一个或多个基因或其调节元件选自ACAT1、ACLY、ADORA2A、AXL、B2M、BATF、BCL2L11、BTLA、CAMK2D、cAMP、CASP8、CBLB、CCR5、CD2、CD3D、CD3E、CD3G、CD4、CD5、CD7、CD8A、CD33、CD38、CD52、CD70、CD82、CD86、CD96、CD123、CD160、CD244、CD276、CDK8、CDKN1B、Chi311、CIITA、CISH、CSF2CSK、CTLA-4、CUL3、Cyp11a1、DCK、DGKA、DGKZ、DHX37、ELOB(TCEB2)、ENTPD1(CD39)、FADD、FAS、GATA3、IL6、IL6R、IL10、IL10RA、IRF4、IRF8、JUNB、Lag3、、LAIR-1(CD305)、LDHA、LIF、LYN、MAP4K4、MAPK14、MCJ、MEF2D、MGAT5、NR4A1、NR4A2、NR4A3、NT5E(CD73)、ODC1、OTULINL(FAM105A)、PAG1、PDCD1、PDIA3、PHD1(EGLN2)、PHD2(EGLN1)、PHD3(EGLN3)、PIK3CD、PIKFYVE、PPARa、PPARd、PRDMI1、PRKACA、PTEN、PTPN2、PTPN6、PTPN11、PVRIG(CD112R)、RASA2、RFXANK、SELPG/PSGL1、SIGLEC15、SLA、SLAMF7、SOCS1、Spry1、Spry2、STK4、SUV39、H1TET2、TGFbRII、TIGIT、Tim-3、TMEM222、TNFAIP3、TNFRSF8(CD30)、TNFRSF10B、TOX、TOX2、、TRAC、TRBC1、TRBC2、UBASH3A、VHL、VISTA、XBP1、YAP1和ZC3H12A。在一些实施方案中,至少一个或多个基因或其调节元件选自CD3、CD5、CD7、CD33、CD123、TRAC、LAG-3、FAS、CD52、TRBC1、TRBC2、B2M、以及CIITA和PD-1。在一些实施方案中,经修饰的免疫细胞包含在CD5和选自TRAC、LAG-3、FAS、CD52、TRBC1、TRBC2、B2M、以及CIITA和PD-1的至少一个或多个基因或其调节元件中的修饰。在一些实施方案中,经修饰的免疫细胞包含在CD7和选自TRAC、LAG-3、FAS、CD52、TRBC1、TRBC2、B2M、以及CIITA和PD-1的至少一个或多个基因或其调节元件中的修饰。在一些实施方案中,经修饰的免疫细胞包含在CD33和选自TRAC、LAG-3、FAS、CD52、TRBC1、TRBC2、B2M、以及CIITA和PD-1的至少一个或多个基因或其调节元件中的修饰。在一些实施方案中,经修饰的免疫细胞包含在CD3和选自TRAC、LAG-3、FAS、CD52、TRBC1、TRBC2、B2M、以及CIITA和PD-1的至少一个或多个基因或其调节元件中的修饰。在一些实施方案中,经修饰的免疫细胞包含在CD123和选自TRAC、LAG-3、FAS、CD52、TRBC1、TRBC2、B2M、以及CIITA和PD-1的至少一个或多个基因或其调节元件中的修饰。基因的多重编辑可用于创建具有改善的治疗特性的CAR-T细胞疗法。该方法解决了经多重编辑T细胞产品的已知限制并且是朝着下一代精确细胞疗法的有前景的发展。

[0484] 一方面,本发明提供一种通用CAR-T细胞。在一些实施方案中,本文所述的CAR-T细

胞是同种异体细胞。在一些实施方案中,通用CAR-T细胞是同种异体T细胞,其可用来表达所希望的CAR,并且可以普遍适用,无论捐赠者和接受者的免疫原性兼容性如何。同种异体免疫细胞可以源自一个或多个供体。在某些实施方案中,同种异体免疫细胞源自单个人类供体。例如,同种异体T细胞可以源自单个健康人类供体的PBMC。在某些实施方案中,同种异体免疫细胞源自多个人类供体。在一些实施方案中,如本文所述,通用CAR-T细胞可以通过以下生成:使用基因修饰将同步修饰引入多个基因座中,例如,三个、四个、五个、六个、七个、八个、九个、十个或更多个基因座中。如本文所述的修饰或同步修饰可以是通过碱基编辑器生成的基因编辑,诸如碱基编辑。碱基编辑器可以C碱基编辑器或A碱基编辑器。如本文所讨论,碱基编辑可用来实现基因中断,使得该基因不被表达。通过碱基编辑进行的修饰可用来实现基因表达的减少。在一些实施方案中,碱基编辑器可用来引入基因修饰,使得经编辑的基因不生成机构是或功能上可行的蛋白质产物。在一些实施方案中,修饰,诸如本文所述的同步修饰,可包含基因编辑,诸如碱基编辑,使得该基因产物的表达或功能性被以任何方式改变。例如,基因产物的表达可以相较于基线表达水平增强或上调。在一些实施方案中,基因产物的活性或功能性可以作为碱基编辑或协同作用的多个碱基编辑事件的结果而上调。

[0485] 在一些实施方案中,通用CAR-T细胞的生成可能比自体T细胞(CAR-T)有利,后者可能难以生成以供急用。与自体细胞制剂相比,同种异体方法更适合于与工程化改造自体T细胞表达CAR的不确定性相关的许多情况,并最终在医疗紧急情况下获得移植所需的细胞产物。然而,对于同种异体T细胞或“成品”T细胞,重要的是小心地穿越宿主对CAR-T细胞的反应性(HVGD)以及同种异体T细胞对于宿主细胞的潜在敌意(GVHD)。在这种设想下,碱基编辑可以成功地用来生成多个同时基因编辑事件,使得(a)可能减少或下调抗原的表达以生成同类相杀抗性免疫细胞;(b)可能生成平台细胞类型,其完全没有或表达少量的内源性T细胞受体,例如,TCRα链(诸如经由TRAC的碱基编辑),或TCRβ链(诸如通过TRBC1/TRBC2的碱基编辑);和/或(c)可减少或下调可能与宿主组织系统不相容的抗原的表达,且反之亦然。

[0486] 在一些实施方案中,本文所述的方法可用来生成表达CAR-T的自体T细胞。在一些实施方案中,可以在单个电穿孔事件中实施多个碱基编辑事件,从而减少电穿孔事件相关毒性。任何用于将外源性遗传物质并入细胞内的已知方法都可以用来替换电穿孔,且因此设想将本领域中已知的此类方法用于任何本文所述方法中。

[0487] 在一些方案中,向患有瘤形成(例如,T或NK细胞恶性病变)或具有发展出瘤形成潜力的受试者给药有效量的经修饰的免疫效应细胞(例如,CAR-T细胞),该细胞缺乏或具有降低水平的CD2并且表达CD2共刺激结构域的CD2嵌合抗原受体。在一些实施方案中,给药至受试者的经CD2修饰的免疫细胞在一个或多个基因或其调节元件(例如,CD52、TRAC、PD-1)中用本文所提供的碱基编辑组合物和方法进一步修饰。

[0488] 如本文所示,碱基编辑与CAR插入的组合是用于生成具有最低限度基因组重排的同类相杀抗性同种异体T细胞的可用策略。基因的多重编辑也可用于创建具有改善的治疗特性的CAR-T细胞疗法。该方法解决了CAR-T疗法的已知限制并且是朝着下一代精确细胞疗法的有前景的发展。

[0489] 在一个实验中,碱基编辑器BE4展示了对T细胞中的三种细胞表面靶标(TRAC、B2M和PD-1)的高效多重碱基编辑,在一次电穿孔中分别敲除95%、95%和88%的基因表达,以生成具有高百分比的B2M和CD3蛋白表达降低的细胞的细胞群。编辑这些基因中的每一种可

用于创建具有改善的治疗特性的CAR-T细胞疗法。这些基因中的每一种可以通过单一靶向的碱基变化(C到T)沉默而不产生双链断裂。因此,经BE4处理的细胞也未显示任何可测量的易位(大规模基因组重排),而用核酸酶进行相同三重编辑的细胞确实显示出可检测的基因组重排。

[0490] 因此,将基于核酸酶的TRAC基因敲除与同时进行的两个额外基因的BE介导敲除偶合,产生具有最低限度基因组重排的同种异体基因T细胞群,从而能够在TRAC基因座有针对性地插入CAR转基因。综上所述,这表明,与单独使用核酸酶的方法相比,单独使用碱基编辑或与单一核酸酶敲除和CAR插入相结合是一种有用的策略,可以产生具有最低限度基因组重排的同种异体T细胞。该方法解决了经多重编辑T细胞产品的已知限制并且是朝着下一代精确细胞疗法的有前景的发展。

[0491] 嵌合抗原受体和CAR-T细胞

[0492] 本发明提供使用本文所述的核碱基编辑器修饰的免疫细胞,其表达嵌合抗原受体(CAR)。修饰免疫细胞以表达嵌合抗原受体可增强免疫细胞的免疫反应活性,其中嵌合抗原受体具有对抗原上的表位的亲和性,其中抗原与经改变的生物体适应性相关。例如,嵌合抗原受体可具有对于在肿瘤细胞中表达的蛋白质上的表位的亲和性。因为CAR-T细胞可以独立于主要组织相容性复合体(MHC)发挥作用,激活的CAR-T细胞可以杀死表达抗原的肿瘤细胞。CAR-T细胞的直接作用避开了肿瘤细胞防御机制,这些机制是作为对MHC向免疫细胞呈递抗原的应答而进化的。

[0493] 在一些实施方案中,本发明提供表达嵌合抗原受体的免疫效应细胞,该嵌合抗原受体靶向牵涉入自体免疫应答的B细胞(例如,受试者的表达针对受试者自身组织生成的抗体的B细胞)。

[0494] 一些实施方案包含自体免疫细胞免疫疗法,其中免疫细胞获自具有以表达表面标志物的癌性或其他经改变的细胞为特征的疾病或经改变的适应性的受试者。所获免疫细胞经基因修饰以表达嵌合抗原受体并针对特异性抗原而有效地重新定向。因此,在一些实施方案中,免疫细胞获自需要CAR-T免疫疗法的受试者。在一些实施方案中,在从受试者获得这些自体免疫细胞后,这些细胞在短时间内经培养和修饰。在其他实施方案中,获得自体细胞并随后储存以供进一步使用。这种做法可能适用于将来可能会经历将会减少免疫细胞计数的平行治疗的个体。在同种异体免疫细胞免疫疗法中,免疫细胞可以从将会接受治疗的受试者以外的供体获得。在一些实施方案中,免疫细胞从健康受试者或供体获得,并且经基因修饰以表达嵌合抗原受体并且针对特异性抗原而有效地重新定向。免疫细胞在经修饰以表达嵌合抗原受体后被给药至受试者,以用于治疗瘤形成(例如,白血病)。在一些实施方案中,待修饰以表达嵌合抗原受体的免疫细胞可以从现有的储备免疫细胞培养物获得。

[0495] 免疫细胞和/或免疫效应细胞可以从使用本领域中已知的标准技术从受试者或供体收集的样品分离或纯化。例如,免疫效应细胞可以通过裂解红细胞并通过离心去除外周单核血细胞而从全血样品分离或纯化。免疫效应细胞可以使用选择性纯化方法进一步分离或纯化,该方法基于细胞特异性标志物诸如CD25、CD3、CD4、CD8、CD28、CD45RA或CD45RO来分离免疫效应细胞。在一种实施方案中,CD25+用作标志物来选择调节性T细胞。在一种实施方案中,CD4+用作标志物来选择T细胞。在一种实施方案中,CD8+用作标志物来选择T细胞。在一种实施方案中,CD4+和CD8+用作标志物来选择T细胞。在一种实施方案中,CD4+和CD25+用

作标志物来选择T细胞。

[0496] 在其他实施方案中,本发明提供在TCR恒定区(TRAC)处具有靶向基因敲除的T细胞,其负责TCRαβ表面表达。TCRαβ缺陷性CAR T细胞与同种异体免疫疗法相容(Qasim等人,Sci.Transl.Med.9,eaaj2013(2017);Valton等人,Mol Ther.2015Sep;23(9):1507-1518)。若需要,使用CliniMACS磁珠耗尽去除残余TCRαβT细胞,从而最小化GVHD风险。在另一实施方案中,本发明提供了供体T细胞,其经间接体内选择以识别表达在接纳者造血细胞上的最小组织相容性抗原,从而最小化移植物抗宿主疾病(GVHD)的风险,而GVHD是移植后发病率和死亡率的主要原因(Warren等人,Blood2010;115(19):3869-3878)。用于分离或纯化免疫效应细胞的另一技术是流式细胞术。在荧光激活细胞分选中,使用对于免疫效应细胞标志物具有亲和性的经荧光标记的抗体来标记样品中的免疫效应细胞。使用适用于表达该标志物的细胞的选通策略来隔离细胞。例如,可以通过使用例如对免疫效应细胞标志物(例如,CD4、CD8、CD28、CD45)具有特异性的经荧光标记的抗体和对应选通策略将T细胞与样品中的其他细胞分离。在一种实施方案中,采用CD45选通策略。在一些实施方案中,采用一种用于其他对于免疫效应细胞为特异性的选通策略来代替CD45选通策略或与其组合。在一种实施方案中,采用CD4选通策略。在一种实施方案中,采用CD8选通策略。在一种实施方案中,采用CD25选通策略。在一种实施方案中,采用CD4和CD8选通策略。在一种实施方案中,采用CD4和CD26选通策略。在一些实施方案中,采用一种用于其他对于免疫效应细胞为特异性的选通策略来代替CD4、CD25和/或CD8选通策略或与其组合。在一些实施方案中,采用图40中提供的选通策略。

[0497] 本发明预期的免疫效应细胞是效应T细胞。在一些实施方案中,效应子T细胞是天然CD8+ T细胞、细胞毒T细胞、天然杀手T(NKT)细胞、天然杀手(NK)细胞或调节性T(Treg)细胞。在一些实施方案中,效应子T细胞是胸腺细胞、不成熟的T淋巴细胞、成熟的T淋巴细胞、静息T淋巴细胞或激活的T淋巴细胞。在一些实施方案中,免疫效应细胞是CD4+CD8+ T细胞或CD4-CD8-T细胞。在一些实施方案中,免疫效应细胞是T辅助细胞。在一些实施方案中,T辅助细胞是T辅助1(Th1)、T辅助2(Th2)细胞或表达CD4的辅助T细胞(CD4+ T细胞)。在一些实施方案中,免疫效应细胞是效应子NK细胞。在一些实施方案中,免疫效应细胞是其他T细胞亚群。经修饰的免疫效应细胞除了表达嵌合抗原受体之外,还可以表达外源性细胞因子、不同的嵌合受体或将会增强免疫效应细胞信号传导或功能的任何其他药剂。例如,嵌合抗原受体和细胞因子的共表达可增强CAR-T细胞裂解靶标细胞的能力。

[0498] 本发明预期的嵌合抗原受体包含细胞外结构域、跨膜结构域和细胞内结构域。抗原结合至细胞外结合结构域可激活CAR-T细胞并生成效应子应答,该效应子应答包括CAR-T细胞增殖、细胞因子生产和其他导致抗原表达细胞死亡的过程。在本发明的一些实施方案中,嵌合抗原受体进一步包含连接子。在一些实施方案中,连接子是(GGGGS)n连接子。在一些实施方案中,连接子是(GGGGS)3连接子。在一些实施方案中,本发明的CAR包括前导肽序列(例如,抗原结合结构域的N端)。示例性前导肽氨基酸序列是:METDTLLLWVLLLWVPGSTG。

[0499] 本文预期的嵌合抗原受体的细胞外结合结构域包含抗体或其抗原结合片段的氨基酸序列,其具有对于特异性抗原的亲和性。在各种实施方案中,CAR特异性地结合5T4。示例性抗5T4 CAR包括但不限于CART-5T4(Oxford BioMedica plc)和UCART-5T4(Cellectis SA)。

[0500] 在各种实施方案中,CAR特异性地结合甲胎蛋白。示例性抗甲胎蛋白CAR包括但不限于ET-1402(Eureka Therapeutics Inc)。

[0501] 在各种实施方案中,CAR特异性地结合Ax1。示例性抗Axl CAR包括但不限于CCT-301-38(F1 Oncology Inc)。

[0502] 在各种实施方案中,CAR特异性地结合B7H6。示例性抗B7H6 CAR包括但不限于CYAD-04(Celyad SA)。

[0503] 在各种实施方案中,CAR特异性地结合BCMA。示例性抗BCMACAR包括但不限于ACTR-087+SEA-BCMA(Seattle Genetics Inc)、ALLO-715(Cellectis SA)、ARI-0002(Institut d'Investigacions Biomediques August Pi I Sunyer)、bb-2121(bluebird bio Inc)、bb-21217(bluebird bio Inc)、CART-BCMA(University of Pennsylvania)、CT-053(Carsgen Therapeutics Ltd)、Descartes-08(Cartesian Therapeutics)、FCARH-143(Juno Therapeutics Inc)、ICTCAR-032(Innovative Cellular Therapeutics Co Ltd)、IM21 CART(Beijing Immunochina Medical Science&Technology Co Ltd)、JCARH-125(Memorial Sloan-Kettering Cancer Center)、KITE-585(Kite Pharma Inc)、LCAR-B38M(Nanj ing Legend Biotech Co Ltd)、LCAR-B4822M(Nanjing Legend Biotech Co Ltd)、MCARH-171(Memorial Sloan-Kettering Cancer Center)、P-BCMA-101(Poseida Therapeutics Inc)、P-BCMA-ALL01(Poseida Therapeutics Inc)、spCART-269(Shanghai Unicar-Therapy Bio-medicine Technology Co Ltd)和BCMA02/bb2121(bluebird bio Inc)。BCMA02/bb2121 CAR的多肽序列提供如下:

[0504] MALPVTALLLPLALLLHAARPDIVLTQSPPSLAMSLGKRATISCRASESVTILGSHLIHWYQQKPGQPPTLLIQLASNVQTGVPARFSGSGSRTDFTLTIDPVEEDDVAVYYCLQSRTIPRTFGGGTKLEIKGSTSGSGKPGSGEGSTKGQIQLVQSGPELKKPGETVKISCKASGYTFTDYSINWVKRAPGKGLKWMGWINTETREPAYAYDFRGRFAFSLETSASTAYLQINNLKYEDTATYFCALDYSYAMDYWGQGTSVTVSSAAATTTPAPRPPTPAPTIASQPLSLRPEACRPAAGGAVHTRGLDFACDIYIWAPLAGTCGVLLLSLVITLYCKRGRKKLLYIFKQPFMRPVQTTQEEDGCSCRFPEEEEGGCELRVKFSRSADAPAYQQGQNQLYNELNLGRREEYDVLDKRRGRDPEMGGKPRRKNPQEGLYNELQKDKMAEAYSEIGMKGERRRGKGHDGLYQGLSTATKDTYDALHMQALPPR

[0505] 在各种实施方案中,CAR特异性地结合CCK2R。示例性抗CCK2R CAR包括但不限于抗CCK2R CAR-T衔接子分子(CAM)+抗FITC CAR T细胞疗法(癌症)、Endocyte/Purdue(Purdue University)。

[0506] 在各种实施方案中,CAR特异性地结合CD抗原。示例性抗CD抗原CAR包括但不限于VM-802(ViroMed Co Ltd)。

[0507] 在各种实施方案中,CAR特异性地结合CD123。示例性抗CD123 CAR包括但不限于MB-102(Fortress Biotech Inc)、RNA CART123(University of Pennsylvania)、SFG-iMC-CD123ζ(Bellicum Pharmaceuticals Inc)和UCART-123(Cellectis SA)。

[0508] 在各种实施方案中,CAR特异性地结合CD133。示例性抗CD133 CAR包括但不限于KD-030(Nanjing Kaedi Biotech Inc)。

[0509] 在各种实施方案中,CAR特异性地结合CD138。示例性抗CD138 CAR包括但不限于ATLCAR.CD138(UNC Lineberger Comprehensive Cancer Center)和CART-138中(中国人民解放军总医院)。

[0510] 在各种实施方案中,CAR特异性地结合CD171。示例性抗CD171 CAR包括但不限于JCAR-023(Juno Therapeutics Inc)。

[0511] 在各种实施方案中,CAR特异性地结合CD19。示例性抗CD19 CAR包括但不限于1928z-41BBL(Memorial Sloan-Kettering Cancer Center)、1928z-E27(Memorial Sloan-Kettering Cancer Center)、19-28z-T2(广州生物医药与健康研究院)、4G7-CARD(University College London)、4SCAR19(深圳市免疫基因治疗研究院)、ALLO-501(Pfizer Inc)、ATA-190(QIMR Berghofer Medical Research Institute)、AUTO-1(University College London)、AVA-008(Avacta Ltd)、axicabtagene ciloleucel(Kite Pharma Inc)、BG-T19(广州生物基因科技有限公司)、BinD-19(Shenzhen BinDeBio Ltd.)、BPX-401(Bellicum Pharmaceuticals Inc)、CAR19h28TM41BBz(Westmead Institute for Medical Research)、C-CAR-011(中国人民解放军总医院)、CD19CART(Innovative Cellular Therapeutics Co Ltd)、CIK-CAR.CD19(Formula Pharmaceuticals Inc)、CLIC-1901(Ottawa Hospital Research Institute)、CSG-CD19(Carsgen Therapeutics Ltd)、CTL-119(University of Pennsylvania)、CTX-101(CRISPR Therapeutics AG)、DSCAR-01(Shanghai Hrain Biotechnology)、ET-190(Eureka Therapeutics Inc)、FT-819(Memorial Sloan-Kettering Cancer Center)、ICAR-19(Immune Cell Therapy Inc)、IM19 CAR-T(Beijing Immunochina Medical Science&Technology Co Ltd)、JCAR-014(Juno Therapeutics Inc)、JWCAR-029(MingJu Therapeutics(Shanghai)Co.、Ltd)、KD-C-19(Nanjing Kaedi Biotech Inc)、LinCART19(iCell Gene Therapeutics)、lisocabtagene maraleucel(Juno Therapeutics Inc)、MatchCART(Shanghai Hrain Biotechnology)、MB-CART19.1(上海儿童医学中心(Shanghai Children's Medical Center))、PBCAR-0191(Precision BioSciences Inc)、PCAR-019(PersonGen Biomedicine(Suzhou)Co Ltd)、pCAR-19B(重庆精密生物科技有限公司)、PZ-01(Pinze Lifetechnology Co Ltd)、RB-1916(Refuge Biotechnologies Inc)、SKLB-083019(成都银河生物医药有限公司)、spCART-19(上海优康生物医药科技有限公司)、TBI-1501(Takara Bio Inc)、TC-110(TCR2 Therapeutics Inc)、TI-1007(Timmune Biotech Inc)、tisagenlecleucel(Abramson Cancer Center of the University of Pennsylvania)、U-CART(Shanghai Bioray Laboratory Inc)、UCART-19(Wugen Inc)、UCART-19(Cellectis SA)、vadacabtagene leraleucel(Memorial Sloan-Kettering Cancer Center)、XLCART-001(Nanjing Medical University)、和yinnuokati-19(深圳市创新免疫科技有限公司)。

[0512] 在各种实施方案中,CAR特异性地结合CD2。示例性抗CD2 CAR包括但不限于UCART-2(Wugen Inc)。

[0513] 在各种实施方案中,CAR特异性地结合CD20。示例性抗CD20 CAR包括但不限于ACTR-087(National University of Singapore)、ACTR-707(Unum Therapeutics Inc),CBM-C20.1(中国人民解放军总医院)、MB-106(Fred Hutchinson Cancer Research Center)和MB-CART20.1(Miltenyi Biotec GmbH)。

[0514] 在各种实施方案中,CAR特异性地结合CD22。示例性抗CD22 CAR包括但不限于抗CD22 CAR T细胞疗法(B细胞急性成淋巴细胞性白血病)、宾夕法尼亚大学(University of Pennsylvania)、CD22-CART(上海优康生物医药科技有限公司)、JCAR-018(Opus Bio Inc)、

MendCART(Shanghai Hrain Biotechnology)和UCART-22(Cellectis SA)。

[0515] 在各种实施方案中,CAR特异性地结合CD30。示例性抗CD30 CAR包括但不限于ATLCAR.CD30(UNC Lineberger Comprehensive Cancer Center)、CBM-C30.1(中国人民解放军总医院)和Hu30-CD28zeta(National Cancer Institute)。

[0516] 在各种实施方案中,CAR特异性地结合CD33。示例性抗CD33 CAR包括但不限于抗CD33 CAR γδT细胞疗法(急性骨髓性白血病)、TC BioPharm/伦敦学院大学(University College London),CAR33VH(Opus Bio Inc)、CART-33(中国人民解放军总医院)、CIK-CAR.CD33(Formula Pharmaceuticals Inc)、UCART-33(Cellectis SA)和VOR-33(Columbia University)。

[0517] 在各种实施方案中,CAR特异性地结合CD38。示例性抗CD38 CAR包括但不限于UCART-38(Cellectis SA)。

[0518] 在各种实施方案中,CAR特异性地结合CD38 A2。示例性抗CD38 A2CAR包括但不限于T-007(TNK Therapeutics Inc)。

[0519] 在各种实施方案中,CAR特异性地结合CD4。示例性抗CD4 CAR包括但不限于CD4CAR(iCell Gene Therapeutics)。

[0520] 在各种实施方案中,CAR特异性地结合CD44。示例性抗CD44 CAR包括但不限于CAR-CD44v6(Istituto Scientifico H San Raffaele)。

[0521] 在各种实施方案中,CAR特异性地结合CD5。示例性抗CD5 CAR包括但不限于CD5CAR(iCell Gene Therapeutics)。示例性CD5 CAR核酸序列提供如下:

>5CAR-CH3-CD28TM-CD28-CD3Z

1 MEFGLSWLFLVAILKGVQCIDAMGNIQLVQSGPELKKPGETVKISCKASGYTFTNYGMNW 60

61 VKQAPGKGLRWMGWINTHTGEPTYADDFKGRFAFSLETSASTAYLQINNLKNEDTATYFC 120

121 TRRGYDWYFDVWGAGTTVTVSSGGGGSGGGGSGGGGSDIKMTQSPSSMYASLGERVTITC 180

181 KASQDINSYLSWFHHKPGKSPKTLIYRANRLVDGVPSRFSGSGSGQDYSLTISSLDYEDM 240

241 GIYYCQQYDESPWTFGGGTKLEMKGSGDPAEPKSPDKTHTCPGQPREPQVYTLPPSRDEL 300

301 TKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQ 360

361 QGNVFSCSVMHEALHNHYTQKSLSLSPGKKDPKFWVLVVVGGVLACYSLLVTVAFIIFWV 420

421 RSKRSRLLHSDYMNMTPRRPGPTRKHYQPYAPPRDFAAYRSRVKFSRSADAPAYQQGQNQ 480

[0522] 481 LYNELNLGRREEYDVLDKRRGRDPEMGGKPRRKNPQEGLYNELQKDKMAEAYSEIGMKGE 540

541 RRRGKGHDGLYQGLSTATKDTYDALHMQALPPR 573

>5CAR-CD8aH-CD28TM-CD28-CD3Z

1 MEFGLSWLFLVAILKGVQCIDAMGNIQLVQSGPELKKPGETVKISCKASGYTFTNYGMNW 60

61 VKQAPGKGLRWMGWINTHTGEPTYADDFKGRFAFSLETSASTAYLQINNLKNEDTATYFC 120

121 TRRGYDWYFDVWGAGTTVTVSSGGGGSGGGGSGGGGSDIKMTQSPSSMYASLGERVTITC 180

181 KASQDINSYLSWFHHKPGKSPKTLIYRANRLVDGVPSRFSGSGSGQDYSLTISSLDYEDM 240

241 GIYYCQQYDESPWTFGGGTKLEMKGSGDPATTTPAPRPPTPAPTIASQPLSLRPEACRPA 300

301 AGGAVHTRGLDFACDFWVLVVVGGVLACYSLLVTVAFIIFWVRSKRSRLLHSDYMNMTPR 360

361 RPGPTRKHYQPYAPPRDFAAYRSRVKFSRSADAPAYQQGQNQLYNELNLGRREEYDVLDK 420

421 RRGRDPEMGGKPRRKNPQEGLYNELQKDKMAEAYSEIGMKGERRRGKGHDGLYQGLSTAT
480

481 KDTYDALHMQALPPR 495

>5CAR-CD28H-CD28TM-CD28-CD3Z
1 MEFGLSWLFLVAILKGVQCIDAMGNIQLVQSGPELKKPGETVKISCKASGYTFTNYGMNW
60

61 VKQAPGKGLRWMGWINTHTGEPTYADDFKGRFAFSLETSASTAYLQINNLKNEDTATYFC
120

121 TRRGYDWYFDVWGAGTTVTVSSGGGGSGGGGSGGGGSDIKMTQSPSSMYASLGERVTITC
180

181 KASQDINSYLSWFHHKPGKSPKTLIYRANRLVDGVPSRFSGSGSGQDYSLTISSLDYEDM
240

241 GIYYCQQYDESPWTFGGGTKLEMKGSGDPAIEVMYPPPYLDNEKSNGTIIHVKGKHLCPS
300

301 PLFPGPSKPFWVLVVVGGVLACYSLLVTVAFIIFWVRSKRSRLLHSDYMNMTPRRPGPTR
360

361 KHYQPYAPPRDFAAYRSRVKFSRSADAPAYQQGQNQLYNELNLGRREEYDVLDKRRGRDP
420

[0523]

421 EMGGKPRRKNPQEGLYNELQKDKMAEAYSEIGMKGERRRGKGHDGLYQGLSTATKDTYDA
480

481 LHMQALPPR 489

>5CAR-CH3-CD8aTM-41BB-CD3Z
1 MEFGLSWLFLVAILKGVQCIDAMGNIQLVQSGPELKKPGETVKISCKASGYTFTNYGMNW
60

61 VKQAPGKGLRWMGWINTHTGEPTYADDFKGRFAFSLETSASTAYLQINNLKNEDTATYFC
120

121 TRRGYDWYFDVWGAGTTVTVSSGGGGSGGGGSGGGGSDIKMTQSPSSMYASLGERVTITC
180

181 KASQDINSYLSWFHHKPGKSPKTLIYRANRLVDGVPSRFSGSGSGQDYSLTISSLDYEDM
240

241 GIYYCQQYDESPWTFGGGTKLEMKGSGDPAEPKSPDKTHTCPGQPREPQVYTLPPSRDEL
300

301 TKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQ
360

361 QGNVFSCSVMHEALHNHYTQKSLSLSPGKKDPKPTTTPAPRPPTPAPTIASQPLSLRPEA
420

421 CRPAAGGAVHTRGLDFACDIYIWAPLAGTCGVLLLSLVITLYCKRGRKKLLYIFKQPFMR
480

481 PVQTTQEEDGCSCRFPEEEEGGCELRVKFSRSADAPAYQQGQNQLYNELNLGRREEYDVL
540

541 DKRRGRDPEMGGKPRRKNPQEGLYNELQKDKMAEAYSEIGMKGERRRGKGHDGLYQGLST
600

601 ATKDTYDALHMQALPPR 617

>5CAR-CD8aH-CD8aTM-41BB-CD3Z
1 MEFGLSWLFLVAILKGVQCIDAMGNIQLVQSGPELKKPGETVKISCKASGYTFTNYGMNW
60

61 VKQAPGKGLRWMGWINTHTGEPTYADDFKGRFAFSLETSASTAYLQINNLKNEDTATYFC
120

121 TRRGYDWYFDVWGAGTTVTVSSGGGGSGGGGSGGGGSDIKMTQSPSSMYASLGERVTITC
180

[0524]

181 KASQDINSYLSWFHHKPGKSPKTLIYRANRLVDGVPSRFSGSGSGQDYSLTISSLDYEDM
240

241 GIYYCQQYDESPWTFGGGTKLEMKGSGDPATTTPAPRPPTPAPTIASQPLSLRPEACRPA
300

301 AGGAVHTRGLDFACDIYIWAPLAGTCGVLLLSLVITLYCKRGRKKLLYIFKQPFMRPVQT
360

361 TQEEDGCSCRFPEEEEGGCELRVKFSRSADAPAYQQGQNQLYNELNLGRREEYDVLDKRR
420

421 GRDPEMGGKPRRKNPQEGLYNELQKDKMAEAYSEIGMKGERRRGKGHDGLYQGLSTATKD
480

481 TYDALHMQALPPRX 494

[0525] 在各种实施方案中,CAR特异性地结合CD7。示例性抗CD7 CAR包括但不限于CAR-pNK(PersonGen Biomedicine(Suzhou)Co Ltd)和CD7.CAR/28ζCAR T细胞(Baylor College of Medicine)、UCART7(Washington University in St Louis)。世袭CD7 CAR氨基酸序列如下:

>7CAR8

1 MALPVTALLLPLALLLHAARPGSDIELTQSPAIMSASLGEEITLTCSASSSVSYMHWYQQ
60

61 KSGTSPKLLIYSTSNLASGVPSRFSGSGSGTFYSLTISSVEAEDAADYYCHQWSSYTFGG
[0526] 120

121 GTKLEIKRGGGGSGGGGSGGGGSQVKLQESGGGLVKPGGSLKLSCAASGFTFSSYAMSWV
180

181 RQTPEKRLEWVATISSGGSYTYYPDSVKGRFTISRDNAKNTLYLQMSSLRSEDTAMYYCA
240

241 RQDGYYPGWFANWGQGTTVTVSSAAATTTPAPRPPTPAPTIASQPLSLRPEACRPAAGGA

[0527]
```
300
301 VHTRGLDFACDIYIWAPLAGTCGVLLLSLVITLYCRSKRSRLLHSDYMNMTPRRPGPTRK
360
361 HYQPYAPPRDFAAYRSRVKFSRSADAPAYQQGQNQLYNELNLGRREEYDVLDKRRGRDPE
420
421 MGGKPRRKNPQEGLYNELQKDKMAEAYSEIGMKGERRRGKGHDGLYQGLSTATKDTYDAL
480
481 HMQALPPR 488
```

[0528] 在各种实施方案中,CAR特异性地结合CDH17。示例性抗CDH17 CAR包括但不限于ARB-001.T(Arbele Ltd)。

[0529] 在各种实施方案中,CAR特异性地结合CEA。示例性抗CEA CAR包括但不限于HORC-020(HumOrigin Inc)。

[0530] 在各种实施方案中,CAR特异性地结合嵌合TGF-β受体(CTBR)。示例性抗嵌合TGF-β受体(CTBR)CAR包括但不限于CAR-CTBR T细胞(bluebird bio Inc)。

[0531] 在各种实施方案中,CAR特异性地结合Claudin18.2。示例性抗Claudin18.2 CAR包括但不限于CAR-CLD18 T细胞(Carsgen Therapeutics Ltd)和KD-022(Nanj ing Kaedi Biotech Inc)。

[0532] 在各种实施方案中,CAR特异性地结合CLL1。示例性抗CLL1 CAR包括但不限于KITE-796(Kite Pharma Inc)。

[0533] 在各种实施方案中,CAR特异性地结合DLL3。示例性抗DLL3 CAR包括但不限于AMG-119(Amgen Inc)。

[0534] 在各种实施方案中,CAR特异性地结合双重BCMA/TACI(APRIL)。示例性抗双重BCMA/TACI(APRIL)CAR包括但不限于AUTO-2(Autolus Therapeutics Limited)。

[0535] 在各种实施方案中,CAR特异性地结合双重CD19/CD22。示例性抗双重CD19/CD22 CAR包括但不限于AUTO-3(Autolus Therapeutics Limited)和LCAR-L10D(Nanj ing Legend Biotech Co Ltd)。

[0536] 在各种实施方案中,CAR特异性地结合CD19。

[0537] 在各种实施方案中,CAR特异性地结合双重CLL1/CD33。示例性抗双重CLL1/CD33 CAR包括但不限于ICG-136(iCell Gene Therapeutics)。

[0538] 在各种实施方案中,CAR特异性地结合双重EpCAM/CD3。示例性抗双重EpCAM/CD3 CAR包括但不限于IKT-701(Icell Kealex Therapeutics)。

[0539] 在各种实施方案中,CAR特异性地结合双重ErbB/4ab。示例性抗双重ErbB/4ab CAR包括但不限于LEU-001(King's College London)。

[0540] 在各种实施方案中,CAR特异性地结合双重FAP/CD3。示例性抗双重FAP/CD3 CAR包括但不限于IKT-702(Icell Kealex Therapeutics)。

[0541] 在各种实施方案中,CAR特异性地结合EBV。示例性抗EBV CAR包括但不限于TT-18(Tessa Therapeutics Pte Ltd)。

[0542] 在各种实施方案中,CAR特异性地结合EGFR。示例性抗EGFR CAR包括但不限于抗EGFR CAR T细胞疗法(CBLB MegaTAL,癌症),蓝鸟生物(bluebird bio Inc);表达CTLA-4检查点抑制剂+PD-1检查点抑制剂mAbs的抗EGFR CAR T细胞疗法(EGFR阳性晚期实体瘤),上海细胞疗法研究院(Shanghai Cell Therapy Research Institute);CSG-EGFR(Carsgen

Therapeutics Ltd)和EGFR-IL12-CART(Pregene(Shenzhen)Biotechnology Co Ltd)。

[0543] 在各种实施方案中,CAR特异性地结合EGFRvIII。示例性抗EGFRvIIICAR包括但不限于KD-035(Nanjing Kaedi Biotech Inc)和UCART-EgfrVIII(Cellectis SA)。

[0544] 在各种实施方案中,CAR特异性地结合Flt3。示例性抗Flt3 CAR包括但不限于ALLO-819(Pfizer Inc)和AMG-553(Amgen Inc)。

[0545] 在各种实施方案中,CAR特异性地结合叶酸受体。示例性抗叶酸受体CAR包括但不限于EC17/CAR T(Endocyte Inc)。

[0546] 在各种实施方案中,CAR特异性地结合G250。示例性抗G250 CAR包括但不限于自体T淋巴细胞疗法(G250-scFV转导的,肾细胞癌),Erasmus Medical Center(Daniel den Hoed Cancer Center)。

[0547] 在各种实施方案中,CAR特异性地结合GD2。示例性抗GD2 CAR包括但不限于1RG-CART(University College London)、4SCAR-GD2(Shenzhen Geno-Immune Medical Institute)、C7R-GD2.CART细胞(Baylor College of Medicine)、CMD-501(Baylor College of Medicine)、CSG-GD2(Carsgen Therapeutics Ltd)、GD2-CART01(Bambino Gesu Hospital and Research Institute)、GINAKIT细胞(Baylor College of Medicine)、iC9-GD2-CAR-IL-15T细胞(UNC Lineberger Comprehensive Cancer Center)和IKT-703(Icell Kealex Therapeutics)。

[0548] 在各种实施方案中,CAR特异性地结合GD2和MUC1。示例性抗GD2/MUC1 CAR包括但不限于PSMACAR-T(University of Pennsylvania)。

[0549] 在各种实施方案中,CAR特异性地结合GPC3。示例性抗GPC3 CAR包括但不限于ARB-002.T(Arbele Ltd)、CSG-GPC3(Carsgen Therapeutics Ltd)、GLYCAR(Baylor College of Medicine)和TT-14(Tessa Therapeutics Pte Ltd)。

[0550] 在各种实施方案中,CAR特异性地结合Her2。示例性抗Her2 CAR包括但不限于ACTR-087+曲妥珠单抗(Unum Therapeutics Inc)、ACTR-707+曲妥珠单抗(Unum Therapeutics Inc)、CIDeCAR(Bellicum Pharmaceuticals Inc)、MB-103(Mustang Bio Inc)、RB-H21(Refuge Biotechnologies Inc)和TT-16(Baylor College of Medicine)。

[0551] 在各种实施方案中,CAR特异性地结合IL13R。示例性抗IL13R CAR包括但不限于MB-101(City of Hope)和YYB-103(YooYoung Pharmaceuticals Co Ltd)。

[0552] 在各种实施方案中,CAR特异性地结合整合素β-7。示例性抗整合素β-7CAR包括但不限于MMG49 CAR T细胞疗法(Osaka University)。

[0553] 在各种实施方案中,CAR特异性地结合LC抗原。示例性抗LC抗原CAR包括但不限于VM-803(ViroMed Co Ltd)和VM-804(ViroMed Co Ltd)。

[0554] 在各种实施方案中,CAR特异性地结合间皮素。示例性抗间皮素CAR包括但不限于CARMA-hMeso(Johns Hopkins University)、CSG-MESO(Carsgen Therapeutics Ltd)、iCasp9M28z(Memorial Sloan-Kettering Cancer Center)、KD-021(Nanjing Kaedi Biotech Inc)、m-28z-T2(Guangzhou Institutes of Biomedicine and Health)、MesoCART(University of Pennsylvania)、meso-CAR-T+PD-78(MirImmune LLC),RB-M1(Refuge Biotechnologies Inc)和TC-210(TCR2 Therapeutics Inc)。

[0555] 在各种实施方案中,CAR特异性地结合MUC1。示例性抗MUC1 CAR包括但不限于

anti-MUC1 CAR T-cell therapy+PD-1knockout T cell therapy(esophageal cancer/NSCLC)、Guangzhou Anjie Biomedical Technology/University of Technology Sydney(Guangzhou Anjie Biomedical Technology Co LTD)、ICTCAR-043(Innovative Cellular Therapeutics Co Ltd)、ICTCAR-046(Innovative Cellular Therapeutics Co Ltd)、P-MUC1C-101(Poseida Therapeutics Inc)和TAB-28z(OncoTab Inc)。

[0556] 在各种实施方案中,CAR特异性地结合MUC16。示例性抗MUC16 CAR包括但不限于4H1128Z-E27(Eureka Therapeutics Inc)和JCAR-020(Memorial Sloan-Kettering Cancer Center)。

[0557] 在各种实施方案中,CAR特异性地结合nfP2X7。示例性抗nfP2X7CAR包括但不限于BIL-022c(Biosceptre International Ltd)。

[0558] 在各种实施方案中,CAR特异性地结合PSCA。示例性抗PSCA CAR包括但不限于BPX-601(Bellicum Pharmaceuticals Inc)。

[0559] 在各种实施方案中,CAR特异性地结合PSMA。CIK-CAR.PSMA(Formula Pharmaceuticals Inc)和P-PSMA-101(Poseida Therapeutics Inc)。

[0560] 在各种实施方案中,CAR特异性地结合ROR1。示例性抗ROR1 CAR包括但不限于JCAR-024(Fred Hutchinson Cancer Research Center)。

[0561] 在各种实施方案中,CAR特异性地结合ROR2。示例性抗ROR2 CAR包括但不限于CCT-301-59(F1 Oncology Inc)。

[0562] 在各种实施方案中,CAR特异性地结合SLAMF7。示例性抗SLAMF7CAR包括但不限于UCART-CS1(Cellectis SA)。

[0563] 在各种实施方案中,CAR特异性地结合TRBC1。示例性抗TRBC1 CAR包括但不限于AUTO-4(Autolus Therapeutics Limited)。

[0564] 在各种实施方案中,CAR特异性地结合TRBC2。示例性抗TRBC2 CAR包括但不限于AUTO-5(Autolus Therapeutics Limited)。

[0565] 在各种实施方案中,CAR特异性地结合TSHR。示例性抗TSHR CAR包括但不限于ICTCAT-023(Innovative Cellular Therapeutics Co Ltd)。

[0566] 在各种实施方案中,CAR特异性地结合VEGFR-1。示例性抗VEGFR-1CAR包括但不限于SKLB-083017(Sichuan University)。

[0567] 在各种实施方案中,CAR是AT-101(AbClon Inc);AU-101、AU-105和AU-180(Aurora Biopharma Inc);CARMA-0508(Carisma Therapeutics);CAR-T(Fate Therapeutics Inc);CAR-T(Cell Design Labs Inc);CM-CX1(Celdara Medical LLC);CMD-502、CMD-503和CMD-504(Baylor College of Medicine);CSG-002和CSG-005(Carsgen Therapeutics Ltd);ET-1501、ET-1502和ET-1504(Eureka Therapeutics Inc);FT-61314(Fate Therapeutics Inc);GB-7001(Shanghai GeneChem Co Ltd);IMA-201(Immatics Biotechnologies GmbH);IMM-005和IMM-039(Immunome Inc);ImmuniCAR(TC BioPharm Ltd);NT-0004和NT-0009(BioNTech Cell和Gene Therapies GmbH)、OGD-203(OGD2 Pharma SAS)、PMC-005B(PharmAbcine)、和TI-7007(Timmune Biotech Inc)。

[0568] 本文还提供编码本文所述嵌合抗原受体的核酸。在一些实施方案中,核酸是分离的或纯化的。核酸的间接体内递送可以使用本领域已知方法实施。例如,可以用编码嵌合抗

原受体的核酸载体转化从受试者获得的免疫细胞。然后,该载体可以用来转化接纳者免疫细胞,使得这些细胞将会随后表达嵌合抗原受体。转化免疫细胞的有效手段包括转染和转导。此类方法是本领域中周知的。例如,适用于递送编码嵌合抗原受体的核酸分子(和编码碱基编辑器的核酸)的方法可以见于国际专利申请号PCT/US2009/040040和美国专利号8,450,112、9,132,153和9,669,058,其各自以其整体并入本文。此外,本文所述的用于递送编码碱基编辑器的那些方法和载体适用于递送编码嵌合抗原受体的核酸。

[0569] 细胞外结合结构域

[0570] 本发明的嵌合抗原受体包括细胞外结合结构域。本文预期的嵌合抗原受体的细胞外结合结构域包含抗体或其抗原结合片段的氨基酸序列,其具有对于特异性抗原的亲和性。在一些实施方案中,抗原是CD3。在一些实施方案中,抗原是CD5。在一些实施方案中,抗原是CD7。在一些实施方案中,抗原是CD33。在一些实施方案中,抗原是CD123。

[0571] 在一些实施方案中,嵌合抗原受体包含抗体的氨基酸序列。在一些实施方案中,嵌合抗原受体包含抗体的抗原结合片段的氨基酸序列。细胞外结合结构域的抗体(或其片段)部分识别并结合至抗原的表位。在一些实施方案中,嵌合抗原受体的抗体片段部分是单链可变片段(scFv)。scFv包含单克隆抗体的轻和可变片段。在其他实施方案中,嵌合抗原受体的抗体片段部分是多链可变片段,其可包含超过一个细胞外结合结构域并因此同时结合至超过一个抗原。在多链可变片段实施方案中,铰链区可以分隔不同的可变片段,提供必要的空间排列和柔性。

[0572] 在其他实施方案中,嵌合抗原受体的抗体部分包含至少一条重链和至少一条轻链。在一些实施方案中,嵌合抗原受体的抗体部分包含通过二硫桥接合的两条重链以及两条轻链,其中轻链各自通过二硫桥接合至重链中的一条。在一些实施方案中,轻链包含恒定区和可变区。位于抗体可变区内的互补决定区负责抗体对于特定抗原的亲和性。因此,识别不同抗原的抗体包含不同的互补决定区。互补决定区位于细胞外结合结构域的可变结构域内,且可变结构域(即,可变重和可变轻)可以使用连接子连接,或者在一些实施方案中使用二硫桥连接。在一些实施方案中,可变重链和可变轻链通过(GGGGS)n连接子连接,其中n是1至10的整数。在一些实施方案中,连接子是(GGGGS)3连接子。

[0573] 在一些实施方案中,被细胞外结构域识别并结合的抗原是蛋白质或肽、核酸、脂质或多糖。抗原可以是异源的,诸如在致病性细菌或病毒中表达的那些。抗原也可以是合成的;例如,一些个体对合成乳胶极端过敏,且暴露于该抗原可能导致极端免疫反应。在一些实施方案中,抗原是自体的,并且在患病或者以其他方式改变的细胞上表达。例如,在一些实施方案中,抗原在肿瘤细胞中表达。在一些实施方案中,肿瘤细胞是实体瘤细胞。在其他实施方案中,肿瘤细胞是液态瘤细胞。在其他实施方案中,肿瘤细胞是血液学癌症,诸如B细胞癌症。在一些实施方案中,B细胞癌症是淋巴瘤或白血病。

[0574] 待使用本文所述方法治疗的液态癌症可以是例如白血病。在一些情况下,白血病包括白血病前期。在一些情况下,白血病是急性白血病。急性白血病包括,例如,急性骨髓性白血病(AML)。急性白血病也包括,例如,急性淋巴性白血病或急性淋巴细胞性白血病(ALL);ALL包括B谱系ALL、T谱系ALL和T细胞急性淋巴细胞性白血病(T-ALL)。

[0575] 瘤形成的非限制性示例包括T细胞急性成淋巴细胞性白血病(T-ALL)、蕈样肉芽肿(MF)、Sézary综合征(SS)、外周T/NK细胞淋巴瘤、间变性大细胞淋巴瘤ALK+、原发性皮肤T细

胞淋巴瘤、T细胞大颗粒淋巴细胞型白血病、血管免疫母细胞性T/NK细胞淋巴瘤、肝脾T细胞淋巴瘤、原发性皮肤CD30+淋巴增生性疾病、结外NK/T细胞淋巴瘤、成人T细胞白血病/淋巴瘤、T细胞幼淋巴细胞白血病、皮下脂膜炎样T细胞淋巴瘤、原发性皮肤γδT细胞淋巴瘤、侵袭性NK细胞白血病和肠病相关T细胞淋巴瘤。在一些实施方案中,瘤形成是T细胞急性成淋巴细胞性白血病(T-ALL)。在一些实施方案中,瘤形成是T细胞急性骨髓性白血病(AML)。

[0576] 抗体-抗原相互作用是非共价相互作用,由氢键键合、静电或疏水相互作用或范德华力所致。嵌合抗原受体的细胞外结合结构域对于抗原的亲和性可使用下式计算:

[0577] KA=[抗体-抗原]/[抗体][抗原],其中

[0578] [Ab]=抗体上未被占据的结合位点的摩尔浓度;

[0579] [Ag]=抗原上未被占据的结合位点的摩尔浓度;并且

[0580] [Ab-Ag]=抗体-抗原复合物的摩尔浓度。

[0581] 抗体-抗原相互作用也可以基于抗原从抗体解离来表征。解离常数(KD)是缔合速率与解离速率的比率,并且与亲和常数成反比。因此,KD=1/KA。本领域技术人员将熟悉这些概念并将会知道传统方法诸如ELISA测定可用来计算这些常数。

[0582] 跨膜结构域

[0583] 本发明的嵌合抗原受体包括跨膜结构域。本文所述嵌合抗原受体的跨膜结构域横跨CAR-T细胞脂质双层细胞膜并分隔细胞外结合结构域与细胞内信号传导结构域。在一些实施方案中,这一结构域源自其他具有跨膜结构域的受体,而在其他实施方案中,这一结构域是合成的。在一些实施方案中,跨膜结构域可以源自非人类跨膜结构域,并且在一些实施方案中可以是人源化的。"人源化的"意为具有编码跨膜结构域的核酸的序列,其经优化以使得更容易或更有效地在人类受试者中表达。在一些实施方案中,跨膜结构域源自在人类免疫效应细胞中表达的另一跨膜结构域。此类蛋白质的示例包括但不限于,T细胞受体(TCR)复合物的亚基PD1、或任何分化簇蛋白、或其他在免疫效应细胞中表达的其他蛋白质和具有跨膜结构域的蛋白质。在一些实施方案中,跨膜结构域将是合成的,并且此类序列将包含很多疏水性残基。

[0584] 用于所公开的CAR的跨膜结构域可包括T细胞受体的α、β或ζ链、CD28、CD3ε、CD45、CD4、CD5、CD8、CD9、CD16、CD22、CD33、CD37、CD64、CD80、CD86、CD134、CD137、CD154的至少跨膜区。在一些实施方案中,跨膜结构域源自CD4、CD8α、CD28或CD3ζ。在一些实施方案中,跨膜结构域是CD28跨膜结构域。在一些实施方案中,跨膜结构域是CD8α跨膜结构域。

[0585] 在一些实施方案中,跨膜结构域是CD8α铰链和跨膜结构域。在一些实施方案中,CD8α铰链和跨膜结构域与下文提供的示例性氨基酸序列至少85%、至少90%、至少95%、至少98%、至少99%或100%相同:

[0586] SDPTTTPAPRPPTPAPTIASQPLSLRPEACRPAAGGAVHTRGLDFACDIYIWAPLAGTCGVLLLSLVITLYC

[0587] 在一些实施方案中,嵌合抗原受体经设计以包含位于跨膜结构域与细胞外结构域、细胞内结构域或两者之间的间隔序列。此类间隔序列可以是1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个氨基酸的长度。在一些实施方案中,间隔序列可以是20、30、40、50、60、70、80、90或100个氨基酸的长度。在又其他实施方案中,间隔序列可以是100到500个氨基酸之间的长度。间隔序列可以是任何多肽,其将一个结构域与另一个结构

域链接并且用来定位此类经连接的结构域,以增强或优化嵌合抗原受体功能。在一些实施方案中,铰链/间隔序列选自CH3、CD8α或CD28。

[0588] 细胞内信号传导结构域

[0589] 本发明的嵌合抗原受体包括细胞内信号传导结构域。细胞内信号传导结构域是在T细胞内表达的蛋白质的细胞内部分,其转导T细胞效应子功能信号(例如,激活信号)并且引导该T细胞执行专业功能。T细胞激活可以由多种因素诱导,包括抗原与T细胞表面上的T细胞受体的结合以及同源配体与T细胞表面上的共刺激分子的结合。T细胞共刺激分子是T细胞上的同源结合配偶体,其与共刺激配体特异性地结合,从而介导T细胞的共刺激应答,诸如但不限于,增殖。共刺激分子包括但不限于MHC I类分子。T细胞的激活导致免疫应答,诸如T细胞增殖和分化(参见,例如,Smith-Garvin等人,Annu.Rev.Immunol.,27:591-619,2009)。示例性T细胞信号传导结构域是本领域中已知的。非限制性示例包括CD3ζ、CD8、CD28、CD27、CD154、GITR(TNFRSF18)、CD134(OX40)和CD137(4-1BB)信号传导结构域。

[0590] 本文设想的嵌合抗原受体的细胞内信号传导结构域包含主要信号传导结构域。在一些实施方案中,嵌合抗原受体包含主要信号传导结构域和词语或共刺激信号传导结构域。

[0591] 在一些实施方案中,主要信号传导结构域包含一个或多个免疫受体酪氨酸激活基序或ITAM。在一些实施方案中,主要信号传导结构域包含超过一个ITAM。并入嵌合抗原受体中的ITAM可以源自来自其他细胞受体的ITAM。在一些实施方案中,包含ITAM的主要信号传导结构域可以源自TCR复合物的亚基,诸如CD3γ、CD3ε、CD3ζ或CD3δ(参见图1A)。在一些实施方案中,包含ITAM的主要信号传导结构域可以源自FcRγ、FcRβ、CD5、CD22、CD79a、CD79b或CD66d。

[0592] 在一些实施方案中,主要信号传导结构域选自CD8、CD28、CD134(OX40)、CD137(4-1BB)和CD3ζ所组成的群组。在一些实施方案中,主要信号传导结构域是CD3ζ信号传导结构域。在一些实施方案中,CD3ζ信号传导结构域与下文提供的示例性氨基酸序列至少85%、至少90%、至少95%、至少98%、至少99%或100%相同:

[0593] RVKFSRSADAPAYQQGQNQLYNELNLGRREEYDVLDKRRGRDPEMGGKPRRKNPQEGLYNELQKDKMAEAYSEIGMKGERRRGKGHDGLYQGLSTATKDTYDALHMQALPPR

[0594] 在一些实施方案中,主要信号传导结构域是CD134(OX40)信号传导结构域。在一些实施方案中,CD134(OX40)信号传导结构域与下文提供的示例性氨基酸序列至少85%、至少90%、至少95%、至少98%、至少99%或100%相同:

[0595] rrdqrlppdahkppgggsfrtpiqeeqadahstlaki

[0596] 在一些实施方案中,次要或共刺激信号传导结构域源自4-1BB、CD2、CD4、CD28、CDS、CD8α、CD83、CD134、CD137、ICOS或CD154。在一些实施方案中,次要信号传导结构域是CD28信号传导结构域。在一些实施方案中,CD28信号传导结构域与下文提供的示例性氨基酸序列至少85%、至少90%、至少95%、至少98%、至少99%或100%相同:

[0597] SKRSRLLHSDYMNMTPRRPGPTRKHYQPYAPPRDFAAYRS

[0598] 在一些实施方案中,次要信号传导结构域是CD137(4-1BB)信号传导结构域。在一些实施方案中,CD137(4-1BB)信号传导结构域与下文提供的示例性氨基酸序列至少85%、至少90%、至少95%、至少98%、至少99%或100%相同:

[0599] KRGRKKLLYIFKQPFMRPVQTTQEEDGCSCRFPEEEEGGCEL

[0600] 在一些实施方案中,CD137(4-1BB)信号传导结构域与下文提供的示例性氨基酸序列至少85%、至少90%、至少95%、至少98%、至少99%或100%相同:

[0601] RFSVVKRGRKKLLYIFKQPFMRPVQTTQEEDGCSCRFPEEEEGGCEL.

[0602] 在一些实施方案中,CAR包含一个或多个信号传导结构域。在一些实施方案中,CAR包含4-1BB信号传导结构域和CD3ζ信号传导结构域。在一些实施方案中,CAR包含CD28信号传导结构域和CD3ζ信号传导结构域。

[0603] 免疫细胞中靶标基因的编辑

[0604] 本发明提供免疫细胞,其包含嵌合抗原(CAR)和一个或多个经编辑的基因、一个或多个其调节元件或其组合,其中经编辑的基因的表达被敲除或敲低。在一些实施方案中,免疫细胞包含CAR和改变的内源性基因,其提供对同类相杀的抗性,增强免疫细胞功能、对免疫阻遏或抑制的抗性或其组合。在一些实施方案中,免疫细胞是T细胞。在一些实施方案中,免疫细胞是CAR-T细胞。在一些实施方案中,免疫细胞是NK细胞。在一些实施方案中,每个经编辑的基因可包含单个碱基编辑。在一些实施方案中,每个经编辑的基因可包含位于基因不同区的多个碱基编辑。

[0605] 在一些实施方案中,单个修饰事件(诸如电穿孔)可以引入一个或多个碱基编辑。在一些实施方案中,可以同时将至少一重、两重、三重、四重、五重、六重、七重、八重、九重、十重、十一重、十二重、十三重、十四重、十五重、十六重、十七重、十八重、十九重、二十重或更多重编辑引入一重或多重基因中。在一些实施方案中,免疫细胞,包括但不限于任何包含选自任何前述基因编辑的经编辑的基因的免疫细胞,可以经编辑以在其他基因中生成突变,该突变增强CAR-T的功能或降低免疫抑制或对细胞的抑制。

[0606] 在一些实施方案中,CAR-T细胞具有与类似的但不进一步具有本文所述一重或多重编辑的CAR-T细胞相比增加的同类相杀抗性。在一些实施方案中,CAR-T细胞具有与类似的但不进一步具有本文所述一重或多重编辑的CAR-T细胞相比降低的免疫原性。在一些实施方案中,CAR-T细胞具有与类似的但不进一步具有本文所述一重或多重编辑的CAR-T细胞相比更低的激活阈值。在一些实施方案中,CAR-T细胞具有与类似的但不进一步具有本文所述一重或多重编辑的CAR-T细胞相比增加的抗瘤形成活性。

[0607] 在一些实施方案中,本文提供一种免疫细胞,其具有至少一个在内源性基因中或其调节元件中的修饰。在一些实施方案中,免疫细胞可以包含在至少一个、两个、三个、四个、五个、六个、七个、八个、九个、十个、十一个、十二个、十三个、十四个、十五个、十六个、十七个、十八个、十九个、二十个或更多个内源性基因或其调节元件中的进一步修饰。

[0608] 在一些实施方案中,该一个或多个基因、或一个或多个其调节元件、或其组合可选自由以下所组成的群组:CD3抗原(CD3);CD5抗原(CD5);CD7抗原(CD7);CD33抗原(CD33);CD52抗原(CD52);CD123抗原(CD123);T细胞受体α恒定区(TRAC);程序性细胞死亡1(PDCD1或PD-1);Fas细胞表面死亡受体(FAS);淋巴细胞激活基因3(LAG-3);II类主要组织相容性复合体反式激活因子(CIITA);T细胞受体β恒定区1(TRBC1);T细胞受体β恒定区2(TRBC2);和β-2微球蛋白(B2M)。在一些实施方案中,CD3、CD5、CD7、CD33或CD123被编辑。在一些实施方案中,免疫细胞包含经编辑的CD3基因,和额外的至少一个经编辑的基因。在一些实施方案中,免疫细胞包含经编辑的CD5基因,和额外的至少一个经编辑的基因。在一些实施方案

中,免疫细胞包含经编辑的CD7基因,和额外的至少一个经编辑的基因。在一些实施方案中,免疫细胞包含经编辑的CD33基因,和额外的至少一个经编辑的基因。在一些实施方案中,免疫细胞包含经编辑的CD123基因,和额外的至少一个经编辑的基因。该至少一个经编辑的基因可以选自前述段落中提及的一系列基因。在一些实施方案中,CD3、CD5、CD7、CD33或CD123与TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和PD1中的一者或多者组合被编辑。

[0609] 在各种实施方案中,经修饰的免疫细胞包含在CD5、TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和PD1中的一者或多者或其组合中的突变。在各种实施方案中,经修饰的免疫细胞包含在CD5、TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和PD1中的两者、三者、四者、五者、六者、七者、八者、九者或十者中的突变。在各种实施方案中,经修饰的免疫细胞包含在CD5中的突变以及在TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和PD1中的一者或多者或其组合中的突变。在一种实施方案中,经修饰的免疫细胞包含在CD5、TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和PD1中的突变。

[0610] 在各种实施方案中,经修饰的免疫细胞包含在CD7、TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和PD1中的一者或多者或其组合中的突变。在各种实施方案中,经修饰的免疫细胞包含在CD7、TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和PD1中的两者、三者、四者、五者、六者、七者、八者、九者或十者中的突变。在各种实施方案中,经修饰的免疫细胞包含在CD7中的突变以及在TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和PD1中的一者或多者或其组合中的突变。在一种实施方案中,经修饰的免疫细胞包含在CD7、TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和PD1中的突变。

[0611] 在各种实施方案中,经修饰的免疫细胞包含在CD3、TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和PD1中的一者或多者或其组合中的突变。在各种实施方案中,经修饰的免疫细胞包含在CD3、TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和PD1中的两者、三者、四者、五者、六者、七者、八者、九者或十者中的突变。在各种实施方案中,经修饰的免疫细胞包含在CD3中的突变以及在TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和PD1中的一者或多者或其组合中的突变。在一种实施方案中,经修饰的免疫细胞包含在CD3、TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和PD1中的突变。

[0612] 在各种实施方案中,经修饰的免疫细胞包含在CD33、TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和PD1中的一者或多者或其组合中的突变。在各种实施方案中,经修饰的免疫细胞包含在CD33、TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和PD1中的两者、三者、四者、五者、六者、七者、八者、九者或十者中的突变。在各种实施方案中,经修饰的免疫细胞包含在CD33中的突变以及在TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和PD1中的一者或多者或其组合中的突变。在一种实施方案中,经修饰的免疫细胞包含在CD33、TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和PD1中的突变。

[0613] 在各种实施方案中,经修饰的免疫细胞包含在CD123、TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和PD1中的一者或多者或其组合中的突变。在各种实施方案中,经修饰的免疫细胞包含在CD123、TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和PD1中的两者、三者、四者、五者、六者、七者、八者、九者或十者中的突变。在各种实施方案中,经修饰的免疫细胞包含在CD123中的突变以及在TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和PD1中的一者或多者或其组合中的突变。在一种实施方案中,经修饰的免疫细胞包含在

CD123、TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和PD1中的突变。

[0614] 在一些实施方案中,至少一个修饰是单个核碱基修饰。在一些实施方案中,改变的内源性基因可以通过碱基编辑创建。在一些实施方案中,碱基编辑可以减少或削弱基因表达。在一些实施方案中,碱基编辑可以减少或削弱基因激活。在一些实施方案中,碱基编辑可以减少或削弱基因产物的功能性。在一些其他实施方案中,碱基编辑可以激活或提升基因表达。在一些实施方案中,碱基编辑可以增加基因产物的功能性。

[0615] 表达内源性免疫细胞受体以及嵌合抗原受体的同种异体免疫细胞可以识别并粘附宿主细胞,这种情况称为移植物抗宿主疾病(GVHD)。免疫细胞受体复合物的α组分由TRAC基因编码,并且在一些实施方案中,该基因被编辑,使得TCR复合物的α亚基无功能或不存在。因为这一亚基是内源性免疫细胞信号传导所必需的,编辑这一基因可降低由同种异体免疫细胞造成的移植物抗宿主疾病的风险。

[0616] 在一些实施方案中,用来提供同类相杀抗性、增强免疫细胞功能或降低免疫阻遏或抑制的基因编辑可以在免疫细胞被转化以表达嵌合抗原受体之前在该细胞中发生。另一方面,用来增强免疫细胞功能或降低免疫阻遏或抑制的基因编辑可以在CAR-T细胞中发生,即,在免疫细胞已经被这暖和以表达嵌合抗原受体之后发生。

[0617] 在本发明的一些实施方案中,CAR-T细胞中的CD5经编辑以敲除或敲低表达。然后,CAR-T经转化以表达具有CD5 scFv的嵌合抗原受体。通过敲除或敲低CD5基因的表达,经修饰的CAR-T细胞进行同类相杀的可能性较小。

[0618] 在本发明的一些实施方案中,CAR-T细胞中的CD7经编辑以敲除或敲低表达。然后,CAR-T经转化以表达具有CD7 scFv的嵌合抗原受体。通过敲除或敲低CD7基因的表达,经修饰的CAR-T细胞进行同类相杀的可能性较小。

[0619] 在本发明的一些实施方案中,CAR-T细胞中的CD33经编辑以敲除或敲低表达。然后,CAR-T经转化以表达具有CD33 scFv的嵌合抗原受体。通过敲除或敲低CD33基因的表达,经修饰的CAR-T细胞进行同类相杀的可能性较小。

[0620] 在本发明的一些实施方案中,CAR-T细胞中的CD3经编辑以敲除或敲低表达。然后,CAR-T经转化以表达具有CD3 scFv的嵌合抗原受体。通过敲除或敲低CD3基因的表达,经修饰的CAR-T细胞进行同类相杀的可能性较小。

[0621] 在本发明的一些实施方案中,CAR-T细胞中的CD123经编辑以敲除或敲低表达。然后,CAR-T经转化以表达具有CD123 scFv的嵌合抗原受体。通过敲除或敲低CD123基因的表达,经修饰的CAR-T细胞进行同类相杀的可能性较小。

[0622] 宿主免疫细胞可以潜在地将同种异体CAR-T细胞识别为非自我并且引起免疫应答以去除非自我细胞。B2M在几乎全部有核细胞中表达并且于MHC I类复合物缔合(图1B)。循环宿主CD8+ T细胞可以将这一B2M蛋白识别为非自我并且杀死同种异体细胞。为了克服这一移植物排异,在一些实施方案中,B2M基因经编辑以敲除或敲低表达。在一些实施方案中,本文提供一种免疫细胞,其具有经编辑的B2M基因,使得该免疫细胞不表达内源功能性B2M。在一些实施方案中,本文提供一种CAR-T细胞,其具有经编辑的B2M基因,使得该CAR-T细胞表现出减少的或可忽略的或没有B2M的表达。

[0623] 在一些实施方案中,免疫细胞包含嵌合抗原受体和一个或多个经编辑的基因、其调节元件或其组合。经编辑的基因可以是免疫应答调节基因、免疫原基因、检查点抑制剂基

因、牵涉到免疫应答中的基因、细胞表面标志物例如T细胞表面标志物、或其任意组合。在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的基因,该经编辑的基因与激活的T细胞增殖、α-βT细胞激活、γ-δT细胞激活、T细胞增殖的正向调节、T辅助细胞增殖或分化的负向调节有关,或其调节元件或其组合。在一些实施方案中,经编辑的基因可以是检查点抑制剂基因,例如PD1基因、PDC1基因、与其形成或激活途径相关或调节该途径的成员。在一些实施方案中,经编辑的基因是TRAC基因。在一些实施方案中,经编辑的基因是CD5基因。在一些实施方案中,经编辑的基因是CD7基因。在一些实施方案中,经编辑的基因是CD33基因。在一些实施方案中,经编辑的基因是CD3基因。在一些实施方案中,经编辑的基因是CD123基因。在一些实施方案中,经编辑的基因是B2M基因。在一些实施方案中,经编辑的基因是CIITA基因。在一些实施方案中,经编辑的基因是TRBC1/2基因。在一些实施方案中,经编辑的基因是CD5基因。在一些实施方案中,经编辑的基因是CD7基因。在一些实施方案中,经编辑的基因是CD52基因。在一些实施方案中,至少一个经编辑的基因选自PD-1、CD2、CD3、CD5、CD7、CD52、B2M、TRBC1/2、CIITA和TRAC,或其组合。在一些实施方案中,PD-1、CD2、CD52和TRAC基因被编辑。在一些实施方案中,PD-1、CD2、CD52、B2M、TRBC1/2、CIITA和TRAC基因被编辑。在一些实施方案中,PD-1、CD5、CD52和TRAC基因被编辑。在一些实施方案中,PD-1、CD3、CD7和CD52基因被编辑。

[0624] 在一些实施方案中,内源性基因的编辑降低了该基因的表达。在一些实施方案中,与没有修饰的对照细胞相比,内源性基因的编辑将该基因的表达减少至少50%。在一些实施方案中,与没有修饰的对照细胞相比,内源性基因的编辑将该基因的表达减少至少60%。在一些实施方案中,与没有修饰的对照细胞相比,内源性基因的编辑将该基因的表达减少至少70%。在一些实施方案中,与没有修饰的对照细胞相比,内源性基因的编辑将该基因的表达减少至少80%。在一些实施方案中,与没有修饰的对照细胞相比,内源性基因的编辑将该基因的表达减少至少90%。在一些实施方案中,与没有修饰的对照细胞相比,内源性基因的编辑将该基因的表达减少至少100%。在一些实施方案中,内源性基因的编辑消除基因表达。

[0625] 在本发明的一些实施方案中,CAR-T细胞中的PDCD1经编辑以敲除或敲低表达。PDCD1基因编码细胞表面受体PD-1(一种在免疫细胞中表达的免疫系统检查点),并且牵涉到通过促进抗原特异性免疫细胞的细胞凋亡而降低免疫性中。通过敲除或敲低PDCD1基因的表达,经修饰的CAR-T细胞凋亡的坑能效较低,增殖的可能性较高,并且可以从程序性细胞死亡免疫检查点逃逸。

[0626] CBLB编码E3泛素连接酶,该酶在抑制免疫效应细胞激活中发挥显著作用。参照图1C,CBLB蛋白有利于导致免疫效应细胞耐受的信号传导途径并且积极抑制导致免疫效应细胞激活的信号传导途径。因为免疫效应细胞激活是CAR-T细胞进行移植后体内增殖所必需的,在本发明的一些实施方案中,CBLB基因被编辑以敲除或敲低表达。

[0627] 在一些实施方案中,本发明提供一种免疫细胞,其具有经编辑的TRAC基因(其中TRAC基因可包含一个、两个、三个、四个、五个、六个、七个、八个、九个或十个或更多个碱基编辑),使得免疫细胞不表达内源功能性T细胞受体链。在一些实施方案中,免疫细胞是表达嵌合抗原受体的T细胞(CAR-T细胞)。在一些实施方案中,本文提供一种CAR-T细胞,其具有在TRAC基因中的碱基编辑,使得CAR-T细胞具有减少的或可忽略的或没有内源性T细胞受体

α蛋白的表达。

[0628] 在一些实施方案中,本文提供一种免疫细胞,其具有经编辑的CIITA基因,使得该免疫细胞不表达内源功能性II类主要组织相容性复合物反向激活因子。在一些实施方案中,本文提供一种CAR-T细胞,其具有经编辑的CIITA基因,使得该CAR-T细胞表现出减少的或可忽略的或没有内源性II类主要组织相容性复合物反向激活因子的表达。

[0629] 在一些实施方案中,本文提供一种免疫细胞,其具有经编辑的TRBC1或TRBC2基因,使得该免疫细胞不表达内源功能性T细胞受体β链。在一些实施方案中,本文提供一种CAR-T细胞,其具有经编辑的TRBC1/TRBC2基因,使得该CAR-T细胞表现出减少的或可忽略的或没有内源性T细胞受体β链的表达。

[0630] 在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的TRAC、B2M、PDCD1、CBLB基因或其组合,其中经编辑的基因的表达被敲除或敲低。在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的TRAC基因,其中经编辑的基因的表达被敲除或敲低。在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的TRAC和B2M基因,其中经编辑的基因的表达要么被敲除或要么敲低。在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的TRAC和PDCD1基因,其中经编辑的基因的表达要么被敲除或要么敲低。在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的TRAC和CBLB基因,其中经编辑的基因的表达要么被敲除或要么敲低。在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的TRAC、B2M和PDCD1基因,其中经编辑的基因的表达要么被敲除或要么敲低。在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的TRAC、B2M和CBLB基因,其中经编辑的基因的表达要么被敲除或要么敲低。在一些实施方案中,免疫细胞或免疫效应细胞包含嵌合抗原受体和经编辑的TRAC、PDCD1和CBLB基因,其中经编辑的基因的表达要么被敲除或要么敲低。在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的TRAC、B2M、PDCD1和CBLB基因,其中经编辑的基因的表达要么被敲除或要么敲低。在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的B2M基因,其中经编辑的基因的表达被敲除或敲低。在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的B2M和PDCD1基因,其中经编辑的基因的表达要么被敲除或要么敲低。在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的B2M和CBLB基因,其中经编辑的基因的表达要么被敲除或要么敲低。在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的B2M、PDCD1和CBLB基因,其中经编辑的基因的表达要么被敲除或要么敲低。在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的PDCD基因,其中经编辑的基因的表达被敲除或敲低。在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的PDCD和CBLB基因,其中经编辑的基因的表达要么被敲除或要么敲低。在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的CBLB基因,其中经编辑的基因的表达被敲除或敲低。

[0631] 在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的CD5基因,其中经编辑的基因的表达被敲除或敲低。在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的CD7基因,其中经编辑的基因的表达被敲除或敲低。在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的CD33基因,其中经编辑的基因的表达被敲除或敲低。在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的CD3基因,其中经编辑的基因的表达被敲除或敲低。在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的CD123基因,其中经编辑的基因的表达被敲除或敲低。在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的FAS

基因,其中经编辑的基因的表达被敲除或敲低。在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的LAG-3基因,其中经编辑的基因的表达被敲除或敲低。在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的CIITA基因,其中经编辑的基因的表达被敲除或敲低。在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的TRBC1基因,其中经编辑的基因的表达被敲除或敲低。在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的TRBC2基因,其中经编辑的基因的表达被敲除或敲低。在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的CD52基因,其中经编辑的基因的表达被敲除或敲低。在一些实施方案中,免疫细胞包含嵌合抗原受体和一个或多个经编辑的CD3、CD5、CD7、CD33、CD123、TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和/或PD1基因,其中经编辑的基因的表达被敲除或敲低。

[0632] 碱基编辑可以位于基因的任何合适位置处,或在该基因的调节元件中。因此,预期在起始密码子处的单碱基编辑例如可以完全废止该基因的表达。在一些实施方案中,改变的内源性基因可以在外显子、内含子、外显子-内含子拼接点、或其调节元件中经修饰或编辑。修饰可以是对基因或其调节元件中单个核碱基的编辑。修饰可以在外显子、超过一个外显子、内含子或超过一个内含子中,或其组合。修饰可以在基因的开放阅读框中。修饰可以在基因的未截短区内,例如,在3'-UTR或5'-UTR中。在一些实施方案中,修饰在内源性基因的调节元件中。在一些实施方案中,修饰在启动子、增强子、算子、沉默子、绝缘子、终止子、转录起始序列、翻译起始序列(例如,Kozak序列)中,或其组合。在一些实施方案中,碱基编辑可以将提前STOP密码子引入外显子中,导致翻译产物的缺乏或截短,这些产物可能错误折叠并因此通过降解被消除,或者碱基编辑可以产生容易降解的不稳定mRNA。

[0633] 在一些实施方案中,碱基编辑可以例如在人类PDC1/PD-1基因的外显子1、外显子2、或外显子3、或外显子4或外显子5上执行。在一些实施方案中,人类PDC1/PD-1基因中的碱基编辑在外显子1内的位点处执行。在一些实施方案中,人类PDC1/PD-1基因中的碱基编辑在外显子2内的位点处执行。在一些实施方案中,人类PDC1/PD-1基因中的碱基编辑在外显子3内的位点处执行。在一些实施方案中,人类PDC1/PD-1基因中的碱基编辑在外显子4内的位点处执行。在一些实施方案中,人类PDC1/PD-1基因中的碱基编辑在外显子5内的位点处执行。在一些实施方案中,一个或多个碱基编辑动作可以在人类PDC1/PD-1基因上在外显子1、外显子2、外显子3、外显子4、外显子5或其任意组合处执行。

[0634] 在一些实施方案中,人类PDC1/PD-1A基因中的碱基编辑通过编辑靶向外显子1的向导RNA间隔序列的位置4、6、7、8或9来执行。在一些实施方案中,人类PDC1/PD-1A基因中的碱基编辑通过编辑靶向外显子1的向导RNA间隔序列的位置4、6、7、8或9来执行。在一些实施方案中,人类PDC1/PD-1A基因中的碱基编辑通过编辑靶向外显子2的向导RNA间隔序列的位置7、8或9来执行。在一些实施方案中,人类PDC1/PD-1A基因中的碱基编辑通过编辑靶向外显子3的向导RNA间隔序列的位置5、7或8来执行。在一些实施方案中,人类PDC1/PD-1A基因中的碱基编辑通过编辑靶向外显子5的向导RNA间隔序列的位置5或8来执行。

[0635] 在一些实施方案中,碱基编辑可以例如在人类CD7基因的外显子1、外显子2或外显子3上执行。在一些实施方案中,人类CD7基因中的碱基编辑在外显子1内的位点处执行。在一些实施方案中,人类CD7基因中的碱基编辑在外显子2内的位点处执行。在一些实施方案中,人类CD7基因中的碱基编辑在外显子3内的位点处执行。在一些实施方案中,一个或多个

碱基编辑动作可以在人类CD7基因上在外显子1、外显子2、外显子3或其任意组合处执行。在一些实施方案中,人类CD7基因中的碱基编辑在外显子1内的位置4、8、9处执行。在一些实施方案中,人类CD7基因中的碱基编辑通过编辑靶向外显子2的向导RNA间隔序列的位置5、6、7、8或9来执行。在一些实施方案中,人类CD7基因中的碱基编辑通过编辑靶向外显子3的向导RNA间隔序列的位置4或9来执行。

[0636] 在一些实施方案中,碱基编辑可以例如在人类LAG-3基因的外显子1、外显子2、外显子3、外显子4、外显子5、外显子6、外显子7或外显子8上执行。在一些实施方案中,人类LAG-3基因中的碱基编辑在外显子1内的位点处执行。在一些实施方案中,人类LAG-3基因中的碱基编辑在外显子2内的位点处执行。在一些实施方案中,人类LAG-3基因中的碱基编辑在外显子3内的位点处执行。在一些实施方案中,人类LAG-3基因中的碱基编辑在外显子4内的位点处执行。在一些实施方案中,人类LAG-3基因中的碱基编辑在外显子5内的位点处执行。在一些实施方案中,人类LAG-3基因中的碱基编辑在外显子6内的位点处执行。在一些实施方案中,人类LAG-3基因中的碱基编辑在外显子7内的位点处执行。在一些实施方案中,人类LAG-3基因中的碱基编辑在外显子8内的位点处执行。在一些实施方案中,一个或多个碱基编辑动作可以在人类LAG-3基因上在外显子1、外显子2、外显子3、外显子4、外显子5、外显子6、外显子7、外显子8或其任意组合处执行。在一些实施方案中,人类LAG-3基因中的碱基编辑通过编辑靶向外显子1的向导RNA间隔序列的位置4或8来执行。在一些实施方案中,人类LAG-3基因中的碱基编辑通过编辑靶向外显子2的向导RNA间隔序列的位置4、6或8来执行。在一些实施方案中,人类LAG-3基因中的碱基编辑通过编辑靶向外显子3的向导RNA间隔序列的位置4、5、6或7来执行。在一些实施方案中,人类LAG-3基因中的碱基编辑通过编辑靶向外显子4的向导RNA间隔序列的位置4、8或9来执行。在一些实施方案中,人类LAG-3基因中的碱基编辑通过编辑靶向外显子5的向导RNA间隔序列的位置8或9来执行。在一些实施方案中,人类LAG-3基因中的碱基编辑通过编辑靶向外显子6的向导RNA间隔序列的位置4、6、7或8来执行。在一些实施方案中,人类LAG-3基因中的碱基编辑通过编辑靶向外显子7的向导RNA间隔序列的位置4、6或7来执行。在一些实施方案中,人类LAG-3基因中的碱基编辑通过编辑靶向外显子8的向导RNA间隔序列的位置8来执行。

[0637] 在一些实施方案中,碱基编辑可以例如在人类CD33基因的外显子1、外显子2、外显子3、外显子4、外显子5或外显子6上执行。在一些实施方案中,人类CD33基因中的碱基编辑在外显子1内的位点处执行。在一些实施方案中,人类CD33基因中的碱基编辑在外显子2内的位点处执行。在一些实施方案中,人类CD33基因中的碱基编辑在外显子3内的位点处执行。在一些实施方案中,人类CD33基因中的碱基编辑在外显子4内的位点处执行。在一些实施方案中,人类CD33基因中的碱基编辑在外显子5内的位点处执行。在一些实施方案中,人类CD33基因中的碱基编辑在外显子6内的位点处执行。在一些实施方案中,一个或多个碱基编辑动作可以在人类CD33基因上在外显子1、外显子2、外显子3、外显子4、外显子5、外显子6或其任意组合处执行。在一些实施方案中,人类CD33基因中的碱基编辑通过编辑靶向外显子9的向导RNA间隔序列的位置7、8或1来执行。在一些实施方案中,人类CD33基因中的碱基编辑通过编辑靶向外显子2的向导RNA间隔序列的位置4、5、6或8来执行。在一些实施方案中,人类CD33基因中的碱基编辑通过编辑靶向外显子3的向导RNA间隔序列的位置4、5、6或7来执行。在一些实施方案中,人类CD33基因中的碱基编辑通过编辑靶向外显子4的向导RNA

间隔序列的位置6或7来执行。在一些实施方案中,人类CD33基因中的碱基编辑通过编辑靶向外显子5的向导RNA间隔序列的位置7或8来执行。在一些实施方案中,人类CD33基因中的碱基编辑通过编辑靶向外显子6的向导RNA间隔序列的位置4、5或6来执行。

[0638] 在一些实施方案中,碱基编辑可以例如在人类CD123基因的外显子1、外显子2、外显子3、外显子4、外显子5、外显子7、外显子8、外显子10或外显子11上执行。在一些实施方案中,人类CD123基因中的碱基编辑在外显子1内的位点处执行。在一些实施方案中,人类CD123基因中的碱基编辑在外显子2内的位点处执行。在一些实施方案中,人类CD123基因中的碱基编辑在外显子3内的位点处执行。在一些实施方案中,人类CD123基因中的碱基编辑在外显子4内的位点处执行。在一些实施方案中,人类CD123基因中的碱基编辑在外显子5内的位点处执行。在一些实施方案中,人类CD123基因中的碱基编辑在外显子7内的位点处执行。在一些实施方案中,人类CD123基因中的碱基编辑在外显子8内的位点处执行。在一些实施方案中,人类CD123基因中的碱基编辑在外显子10内的位点处执行。在一些实施方案中,人类CD123基因中的碱基编辑在外显子11内的位点处执行。在一些实施方案中,一个或多个碱基编辑动作可以在人类CD123基因上在外显子1、外显子2、外显子3、外显子4、外显子5、外显子7、外显子8、外显子10、外显子11或其任意组合处执行。在一些实施方案中,人类CD123基因中的碱基编辑通过编辑靶向外显子6的向导RNA间隔序列的位置1来执行。在一些实施方案中,人类CD123基因中的碱基编辑通过编辑靶向外显子2的向导RNA间隔序列的位置4、6或8来执行。在一些实施方案中,人类CD123基因中的碱基编辑通过编辑靶向外显子3的向导RNA间隔序列的位置5、6或8来执行。在一些实施方案中,人类CD123基因中的碱基编辑通过编辑靶向外显子4的向导RNA间隔序列的位置5或6来执行。在一些实施方案中,人类CD123基因中的碱基编辑通过编辑靶向外显子5的向导RNA间隔序列的位置4或5来执行。在一些实施方案中,人类CD123基因中的碱基编辑通过编辑靶向外显子7的向导RNA间隔序列的位置5来执行。在一些实施方案中,人类CD123基因中的碱基编辑通过编辑靶向外显子8的向导RNA间隔序列的位置5、6、7或8来执行。在一些实施方案中,人类CD123基因中的碱基编辑通过编辑靶向外显子10的向导RNA间隔序列的位置4、7或8来执行。在一些实施方案中,人类CD123基因中的碱基编辑通过编辑靶向外显子11的向导RNA间隔序列的位置5或8来执行。

[0639] 在一些实施方案中,碱基编辑可以例如在人类FAS基因的外显子1、或外显子3、或外显子4、或外显子5、或外显子6、或外显子7或、外显子8或外显子9上执行。在一些实施方案中,人类FAS基因中的碱基编辑在外显子1内的位点处执行。在一些实施方案中,人类FAS基因中的碱基编辑在外显子3内的位点处执行。在一些实施方案中,人类FAS基因中的碱基编辑在外显子4内的位点处执行。在一些实施方案中,人类FAS基因中的碱基编辑在外显子5内的位点处执行。在一些实施方案中,人类FAS中的碱基编辑在外显子6内的位点处执行。在一些实施方案中,人类FAS中的碱基编辑在外显子7内的位点处执行。在一些实施方案中,人类FAS中的碱基编辑在外显子8内的位点处执行。在一些实施方案中,人类FAS中的碱基编辑在外显子9内的位点处执行。在一些实施方案中,一个或多个碱基编辑动作可以在人类FAS基因上在外显子1、外显子3、外显子4、外显子5、外显子6、外显子7、外显子8、外显子9或其任意组合处执行。

[0640] 在一些实施方案中,人类FAS基因中的碱基编辑通过编辑靶向外显子1的向导RNA间隔序列的位置9来执行。在一些实施方案中,人类FAS基因中的碱基编辑通过编辑靶向外

显子3的向导RNA间隔序列的位置6来执行。在一些实施方案中,人类FAS基因中的碱基编辑通过编辑靶向外显子4的向导RNA间隔序列的位置7来执行。在一些实施方案中,人类FAS基因中的碱基编辑通过编辑靶向外显子5的向导RNA间隔序列的位置5来执行。在一些实施方案中,人类FAS基因中的碱基编辑通过编辑靶向外显子6的向导RNA间隔序列的位置4或7来执行。在一些实施方案中,人类FAS基因中的碱基编辑通过编辑靶向外显子7的向导RNA间隔序列的位置8来执行。在一些实施方案中,人类FAS基因中的碱基编辑通过编辑靶向外显子8的向导RNA间隔序列的位置8来执行。在一些实施方案中,人类FAS基因中的碱基编辑通过编辑靶向外显子9的向导RNA间隔序列的位置5或6来执行。

[0641] 在一些实施方案中,碱基编辑可以例如在人类CD52基因的外显子1、外显子2或外显子3上执行。在一些实施方案中,人类CD52基因中的碱基编辑在外显子1内的位点处执行。在一些实施方案中,人类CD52基因中的碱基编辑在外显子2内的位点处执行。在一些实施方案中,人类CD7基因中的碱基编辑在外显子3内的位点处执行。在一些实施方案中,一个或多个碱基编辑动作可以在人类CD52基因上在外显子1、外显子2、外显子3或其任意组合处执行。在一些实施方案中,人类CD52基因中的碱基编辑通过编辑靶向外显子1的向导RNA间隔序列的位置4或7来执行。在一些实施方案中,人类CD52基因中的碱基编辑通过编辑靶向外显子7的向导RNA间隔序列的位置5、6或2来执行。

[0642] 在一些实施方案中,碱基编辑可以例如在人类CD5基因的外显子1、或外显子2、或外显子3、或外显子4、或外显子5、或外显子6、或外显子7或、或外显子8、或外显子9或外显子10上执行。在一些实施方案中,人类CD5基因中的碱基编辑在外显子1内的位点处执行。在一些实施方案中,人类CD5基因中的碱基编辑在外显子2内的位点处执行。在一些实施方案中,人类CD5基因中的碱基编辑在外显子3内的位点处执行。在一些实施方案中,人类CD5基因中的碱基编辑在外显子4内的位点处执行。在一些实施方案中,人类CD5基因中的碱基编辑在外显子5内的位点处执行。在一些实施方案中,人类CD5基因中的碱基编辑在外显子6内的位点处执行。在一些实施方案中,人类CD5基因中的碱基编辑在外显子7内的位点处执行。在一些实施方案中,人类CD5基因中的碱基编辑在外显子8内的位点处执行。在一些实施方案中,人类CD5基因中的碱基编辑在外显子9内的位点处执行。在一些实施方案中,人类CD5基因中的碱基编辑在外显子10内的位点处执行。在一些实施方案中,一个或多个碱基编辑动作可以在人类CD5基因上在外显子1、外显子2、外显子3、外显子4、外显子5、外显子6、外显子7、外显子8、外显子9、外显子10或其任意组合处执行。

[0643] 在一些实施方案中,人类CD5基因中的碱基编辑通过编辑靶向外显子1的向导RNA间隔序列的位置6来执行。在一些实施方案中,人类CD5基因中的碱基编辑通过编辑靶向外显子1的向导RNA间隔序列的位置6来执行。在一些实施方案中,人类CD5基因中的碱基编辑通过编辑靶向外显子2的向导RNA间隔序列的位置5和/或6来执行。在一些实施方案中,人类CD5基因中的碱基编辑通过编辑靶向外显子2的向导RNA间隔序列的位置5和/或6来执行。在一些实施方案中,人类CD5基因中的碱基编辑通过编辑靶向外显子3的向导RNA间隔序列的位置5、6、8和/或9来执行。在一些实施方案中,人类CD5基因中的碱基编辑通过编辑靶向外显子4的向导RNA间隔序列的位置4或5来执行。在一些实施方案中,人类CD5基因中的碱基编辑通过编辑靶向外显子5的向导RNA间隔序列的位置4、5、7、8或9来执行。在一些实施方案中,人类CD5基因中的碱基编辑通过编辑靶向外显子6的向导RNA间隔序列的位置4、5、6、7、8

和/或9来执行。在一些实施方案中,人类CD5基因中的碱基编辑通过编辑靶向外显子7的向导RNA间隔序列的位置4来执行。在一些实施方案中,人类CD5基因中的碱基编辑通过编辑靶向外显子8的向导RNA间隔序列的位置4、5或7来执行。在一些实施方案中,人类CD5基因中的碱基编辑通过编辑靶向外显子9的向导RNA间隔序列的位置6或8来执行。在一些实施方案中,人类CD5基因中的碱基编辑通过编辑靶向外显子10的向导RNA间隔序列的位置9来执行。

[0644] 在一些实施方案中,碱基编辑可以例如在人类TRAC基因(UCSC基因组数据库ENSG00000277734.8)的外显子1或外显子2或外显子3或外显子4上执行。在一些实施方案中,人类TRAC基因中的碱基编辑在外显子1内的位点处执行。在一些实施方案中,人类TRAC基因中的碱基编辑在外显子2内的位点处执行。在一些实施方案中,人类TRAC基因中的碱基编辑在外显子3内的位点处执行。在一些实施方案中,人类TRAC基因中的碱基编辑在外显子4内的位点处执行。在一些实施方案中,一个或多个碱基编辑动作可以在人类TRAC基因上在外显子1、外显子2、外显子3、外显子4或其任意组合处执行。在一些实施方案中,人类TRAC基因中的碱基编辑通过编辑靶向外显子1内的向导RNA间隔序列的位置5、6或9来执行。

[0645] 在一些实施方案中,碱基编辑可以例如在人类B2M基因(染色体15,NC_000015.10,44711492-44718877;示例性mRNA序列NM_004048)的外显子1、外显子2、外显子3或外显子4上执行。在一些实施方案中,人类B2M基因中的碱基编辑在外显子1内的位点处执行。在一些实施方案中,人类B2M基因中的碱基编辑在外显子2内的位点处执行。在一些实施方案中,人类B2M基因中的碱基编辑在外显子3内的位点处执行。在一些实施方案中,人类B2M基因中的碱基编辑在外显子4内的位点处执行。在一些实施方案中,一个或多个碱基编辑动作可以在人类B2M基因上在外显子1、外显子2、外显子3、外显子4或其任意组合处执行。在一些实施方案中,人类B2M基因中的碱基编辑通过编辑靶向外显子1的向导RNA间隔序列的位置5来执行。在一些实施方案中,人类B2M基因中的碱基编辑通过编辑靶向外显子2的向导RNA间隔序列的位置4、6或9来执行。

[0646] 在一些实施方案中,碱基编辑可以例如在人类CIITA基因的外显子1、外显子2、外显子3、外显子4、外显子5、外显子6、外显子7、外显子8、外显子9、外显子10、外显子11、外显子12、外显子13、外显子14、外显子15、外显子16、外显子17、外显子18或外显子19处执行。在一些实施方案中,人类CD52基因中的碱基编辑在外显子1内的位点处执行。在一些实施方案中,人类CIITA基因中的碱基编辑在外显子2内的位点处执行。在一些实施方案中,人类CIITA基因中的碱基编辑在外显子3内的位点处执行。在一些实施方案中,人类CIITA基因中的碱基编辑在外显子4内的位点处执行。在一些实施方案中,人类CIITA基因中的碱基编辑在外显子5内的位点处执行。在一些实施方案中,人类CIITA基因中的碱基编辑在外显子6内的位点处执行。在一些实施方案中,人类CIITA基因中的碱基编辑在外显子7内的位点处执行。在一些实施方案中,人类CIITA基因中的碱基编辑在外显子8内的位点处执行。在一些实施方案中,人类CIITA基因中的碱基编辑在外显子9内的位点处执行。在一些实施方案中,人类CIITA基因中的碱基编辑在外显子10内的位点处执行。在一些实施方案中,人类CIITA基因中的碱基编辑在外显子11内的位点处执行。在一些实施方案中,人类CIITA基因中的碱基编辑在外显子12内的位点处执行。在一些实施方案中,人类CIITA基因中的碱基编辑在外显子13内的位点处执行。在一些实施方案中,人类CIITA基因中的碱基编辑在外显子14内的位点处执行。在一些实施方案中,人类CIITA基因中的碱基编辑在外显子15内的位点处执行。

在一些实施方案中,人类CIITA基因中的碱基编辑在外显子16内的位点处执行。在一些实施方案中,人类CIITA基因中的碱基编辑在外显子17内的位点处执行。在一些实施方案中,人类CIITA基因中的碱基编辑在外显子18内的位点处执行。在一些实施方案中,人类CIITA基因中的碱基编辑在外显子19内的位点处执行。在一些实施方案中,一个或多个碱基编辑动作可以在人类CIITA基因上在外显子1、外显子2、外显子3、外显子4、外显子5、外显子6、外显子7、外显子8、外显子9、外显子10、外显子11、外显子12、外显子13、外显子14、外显子15、外显子16、外显子17、外显子18、外显子19或其任意组合处执行。

[0647] 在一些实施方案中,人类CIITA基因中的碱基编辑通过编辑靶向外显子1的向导RNA间隔序列的位置6或7来执行。在一些实施方案中,人类CIITA基因中的碱基编辑通过编辑靶向外显子2的向导RNA间隔序列的位置7或8来执行。在一些实施方案中,人类CIITA基因中的碱基编辑通过编辑靶向外显子4的向导RNA间隔序列的位置8来执行。在一些实施方案中,人类CIITA基因中的碱基编辑通过编辑靶向外显子7的向导RNA间隔序列的位置4、7或8来执行。在一些实施方案中,人类CIITA基因中的碱基编辑通过编辑靶向外显子8的向导RNA间隔序列的位置8来执行。在一些实施方案中,人类CIITA基因中的碱基编辑通过编辑靶向外显子9的向导RNA间隔序列的位置4、6或7来执行。在一些实施方案中,人类CIITA基因中的碱基编辑通过编辑靶向外显子10的向导RNA间隔序列的位置4、5或7来执行。在一些实施方案中,人类CIITA基因中的碱基编辑通过编辑靶向外显子11的向导RNA间隔序列的位置4、5、6、7或8来执行。在一些实施方案中,人类CIITA基因中的碱基编辑通过编辑靶向外显子12的向导RNA间隔序列的位置6来执行。在一些实施方案中,人类CIITA基因中的碱基编辑通过编辑靶向外显子14的向导RNA间隔序列的位置4或5来执行。在一些实施方案中,人类CIITA基因中的碱基编辑通过编辑靶向外显子15的向导RNA间隔序列的位置4、7或8来执行。在一些实施方案中,人类CIITA基因中的碱基编辑通过编辑靶向外显子16的向导RNA间隔序列的位置5、7或8来执行。在一些实施方案中,人类CIITA基因中的碱基编辑通过编辑靶向外显子17的向导RNA间隔序列的位置7或8来执行。在一些实施方案中,人类CIITA基因中的碱基编辑通过编辑靶向外显子18的向导RNA间隔序列的位置5来执行。在一些实施方案中,人类CIITA基因中的碱基编辑通过编辑靶向外显子19的向导RNA间隔序列的位置5、6、7或8来执行。

[0648] 在一些实施方案中,碱基编辑可以例如在人类TRBC1基因的外显子1、外显子2或外显子3上执行。在一些实施方案中,人类TRBC1基因中的碱基编辑在外显子1内的位点处执行。在一些实施方案中,人类TRBC1基因中的碱基编辑在外显子2内的位点处执行。在一些实施方案中,人类TRBC1基因中的碱基编辑在外显子3内的位点处执行。在一些实施方案中,一个或多个碱基编辑动作可以在人类TRBC1基因上在外显子1、外显子2、外显子3或其任意组合处执行。在一些实施方案中,人类TRBC1基因中的碱基编辑通过编辑靶向外显子1的向导RNA间隔序列的位置5、6、7或8来执行。在一些实施方案中,人类TRBC1基因中的碱基编辑通过编辑靶向外显子2的向导RNA间隔序列的位置8来执行。在一些实施方案中,人类TRBC1基因中的碱基编辑通过编辑靶向外显子3的向导RNA间隔序列的位置4或5来执行。

[0649] 在一些实施方案中,碱基编辑可以例如在人类TRBC2基因的外显子1、外显子2或外显子3上执行。在一些实施方案中,人类TRBC2基因中的碱基编辑在外显子1内的位点处执行。在一些实施方案中,人类TRBC2基因中的碱基编辑在外显子2内的位点处执行。在一些实施方案中,人类TRBC2基因中的碱基编辑在外显子3内的位点处执行。在一些实施方案中,一

个或多个碱基编辑动作可以在人类TRBC2基因上在外显子1、外显子2、外显子3或其任意组合处执行。在一些实施方案中,人类TRBC2基因中的碱基编辑通过编辑靶向外显子1的向导RNA间隔序列的位置5、6、7或8来执行。在一些实施方案中,人类TRBC2基因中的碱基编辑通过编辑靶向外显子2的向导RNA间隔序列的位置7或8来执行。在一些实施方案中,人类TRBC2基因中的碱基编辑通过编辑靶向外显子3的向导RNA间隔序列的位置4来执行。

[0650] 在一些实施方案中,碱基编辑可以在内含子上执行。例如,碱基编辑可以在内含子上执行。在一些实施方案中,碱基编辑可以在内含子内的位点处执行。在一些实施方案中,碱基编辑可以在一个或多个内含子内的位点处执行。在一些实施方案中,碱基编辑可以在基因中多个内含子的任何外显子处执行。在一些实施方案中,一个或多个碱基编辑可以在外显子上、内含子上、或外显子与内含子的任意组合处执行。

[0651] 在一些实施方案中,修饰或碱基编辑可以在启动子位点内。在一些实施方案中,可以将碱基编辑引入替代性启动子位点内。在一些实施方案中,碱基编辑可以在5'调节元件诸如增强子内。在一些实施方案中,可以引入碱基编辑以中断核酸结合蛋白的结合位点。示例性核酸结合蛋白可以是聚合酶、核酸酶、旋转酶、拓扑异构酶、甲基化酶或甲基转移酶、转录因子、增强子、PABP、锌指蛋白等。

[0652] 在一些实施方案中,碱基编辑可以用于剪接中断以沉默靶标蛋白表达。在一些实施方案中,碱基编辑可以生成剪接受体-剪接供体(SA-SD)位点。生成SA-SD或SA-SD位点的被靶向碱基编辑可导致减少的基因表达。在一些实施方案中,碱基编辑器(例如,ABE、CBE)用来靶向构建剪接受体和剪接供体位点的二核苷酸基序,而该位点是每个内含子的最前和最后的两个核苷酸。例如,可以靶向CD2的外显子3剪接供体(SD)位点进行碱基编辑。在一些实施方案中,使用腺苷碱基编辑器(ABE)实现剪接中断。在一些实施方案中,使用胞苷碱基编辑器(CBE)实现剪接中断。在一些实施方案中,碱基编辑器(例如,ABE,CBE)用来通过创建STOP密码子来编辑外显子。提前终止密码子可以在外显子、内含子或未翻译区内生成。在一些实施方案中,碱基编辑可用来将超过一个STOP密码子引入到一个或多个替代性阅读框中。例如,提前STOP密码子可以引入到外显子2内的位置8处、外显子3内的位置4处、外显子3内的位置6处、外显子3内的位置9处、外显子4内的位置4处、外显子4内的位置5处、或外显子5内的位置4处。在一些实施方案中,通过腺苷碱基编辑器(ABE)生成终止密码子。在一些实施方案中,通过胞苷碱基编辑器(CBE)生成终止密码子。在一些实施方案中,CBE生成以下编辑(以下划线字体显示)中的一者以生成终止子(STOP)密码子:CAG→TAG;CAA→TAA;CGA→TGA;TGG→TGA;TGG→TAG;或TGG→TAA。

[0653] 在一些实施方案中,可以将修饰/碱基编辑引入3'-UTR处,例如,在多腺苷酸化(poly-A)位点中。在一些实施方案中,碱基编辑可以在5'-UTR区上执行。

[0654] 在一些实施方案中,免疫细胞,包括但不限于任何包含任何前述基因编辑的免疫细胞,可以经编辑以在其他基因中生成突变,该突变增强CAR-T的功能或降低免疫抑制或对细胞的抑制。例如,在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的TGFBR2、ZAP70、NFATc1、TET2基因或其组合,其中经编辑的基因的表达要么被敲除要么被敲低。在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的TGFBR2基因,其中经编辑的基因的表达被敲除或敲低。在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的TGFBR2和ZAP70基因,其中经编辑的基因的表达要么被敲除或要么敲低。在一些实施方案中,免疫细

胞包含嵌合抗原受体和经编辑的TGFBR2和ZAP70基因,其中经编辑的基因的表达要么被敲除或要么敲低。在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的TGFBR2和NFATC1基因,其中经编辑的基因的表达要么被敲除或要么敲低。在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的TGFBR2和TET2基因,其中经编辑的基因的表达要么被敲除或要么敲低。在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的TGFBR2、ZAP70和NFATC1基因,其中经编辑的基因的表达要么被敲除或要么敲低。在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的TGFBR2、ZAP70和TET2基因,其中经编辑的基因的表达要么被敲除或要么敲低。在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的TGFBR2、NFATC1和TET2基因,其中经编辑的基因的表达要么被敲除或要么敲低。在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的TGFBR2、ZAP70、NFATC1和TET2基因,其中经编辑的基因的表达要么被敲除或要么敲低。在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的ZAP70基因,其中经编辑的基因的表达被敲除或敲低。在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的ZAP70和NFATC1基因,其中经编辑的基因的表达要么被敲除或要么敲低。在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的ZAP70和TET2基因,其中经编辑的基因的表达要么被敲除或要么敲低。在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的ZAP70、PDCD1和TET2基因,其中经编辑的基因的表达要么被敲除或要么敲低。在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的PDCD1基因,其中经编辑的基因的表达被敲除或敲低。在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的PDCD1和TET2基因,其中经编辑的基因的表达要么被敲除或要么敲低。在一些实施方案中,免疫细胞包含嵌合抗原受体和经编辑的TET2基因,其中经编辑的基因的表达被敲除或敲低。

[0655] 在一些实施方案中,嵌合抗原受体插入到TRAC基因内。这具有优势。首先,因为TRAC在免疫细胞内被高度表达,当其构建体被设计为将嵌合抗原受体插入TRAC基因内以使得该受体的表达由TRAC启动子驱动时,该嵌合抗原受体将会被类似地表达。其次,将嵌合抗原受体插入TRAC基因内将敲除TRAC表达。在一些实施方案中,本文所述的基因编辑系统可用来将嵌合抗原受体插入到TRAC基因座内。对于TRAC基因座具有特异性的gRNA可以将基因编辑系统引导至该基因座并启动双链DNA切割。在特定实施方案中,gRNA与Cas12b协同使用。在各种实施方案中,基因编辑系统与具有编码CAR受体的序列的核酸协同使用。示例性向导RNA提供在下表1A中。

[0656] 表1A向导RNA

[0657]

GUUCUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCU GCAGGGUGUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUAC GAGGCAUUAGCACAGAGUCUCUCAGCUGGUACAC	ATTN	BhCas12b 核酸酶	TRAC KO gRNA 1(外显子 2)
GUUCUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCU GCAGGGUGUGAGGAAACUCCUAUUGCUGGACGAUGUCUCUUA CGAGGCAUUAGCACACCGAUUUUGAUUCUCAAACA	ATTN	BhCas12b 核酸酶	TRAC KO gRNA 2(外显子 2)
GUUCUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCU GCAGGGUGUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUAC GAGGCAUUAGCACUCAAACAAAUGUGCACAAAG	ATTN	BhCas12b 核酸酶	TRAC KO gRNA 3(外显子 2)
GUUCUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCU GCAGGGUGUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUAC GAGGCAUUAGCACUCAAACAAAUGUGUCACAAAG	ATTN	BhCas12b 核酸酶	TRAC KO gRNA 4(外显子 2)

[0658]

GUUCUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCU GCAGGGUGUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUAC GAGGCAUUAGCACUUUGAGAAUCAAAAUCGGUA	ATTN	BhCas12b 核酸酶	TRAC KO gRNA 5(外显子 2)
GUUCUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCU GCAGGGUGUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUAC GAGGCAUUAGCACUGAUGUGUAUAUCACAGACAA	ATTN	BhCas12b 核酸酶	TRAC KO gRNA 6(外显子 2)
GUUCUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCU GCAGGGUGUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUAC GAGGCAUUAGCAGUUGCUCCAGGCCACAGCAU	ATTN	BhCas12b 核酸酶	TRAC KO gRNA 7(外显子 2)
GUUCUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCU GCAGGGUGUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUAC GAGGCAUUAGCACUUCCAGAAGACACCUUCUUCC	ATTN	BhCas12b 核酸酶	TRAC KO gRNA 8(外显子 2)
GUUCUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCU GCAGGGUGUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUAC GAGGCAUUAGCACCAGAAGACACCUUCUUCCCCA	ATTN	BhCas12b 核酸酶	TRAC KO gRNA 9(外显子 2)
GUUCUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCU GCAGGGUGUGAGAGAAACUCCUAUUGCUGGACGAUGUCUCUU ACGAGGCAUUAGCACGGUUCCGAAUCCUCCUGA	ATTN	BhCas12b 核酸酶	TRAC KO gRNA 10(外显子 4)
GUUCUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCU GCAGGGUGUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUAC GAGGCAUUAGCACGGAACCCAAUCACUGACAGGU	ATTN	BhCas12b 核酸酶	TRAC KO gRNA 11(外显子 4)

[0659] DNA构建体编码嵌合抗原受体和含有位于gRNA靶向序列侧翼的TRAC DNA的扩展延伸段的核酸。不受缚于理论,构建体结合至互补TRAC序列,然后驻留在构建体上TRAC序列附近的嵌合抗原受体DNA被插入到损害位点处,有效地敲除TRAC基因并敲入嵌合抗原受体核酸。表1B提供用于TRAC基因的向导RNA,其可以将碱基编辑结构引导至TRAC基因座,使得能够进行嵌合抗原受体核酸的插入。前11个gRNA用于BhCas12b核酸酶。第二组11个用于BvCas12b核酸酶。这些全部用于通过创建双链断裂而将CAR插入到TRAC处,而非用于碱基编辑。

[0660] 表1B:TRAC向导RNA

[0661]

向导 RNA	基因外显子
GTTCTGTCTTTTGGTCAGGACAACCGTCTAGCTATAAGTGCTGCAGGGTGTGAGAA ACTCCTATTGCTGGACGATGTCTCTTACGAGGCATTAGCACAGAGTCTCTCAGCTG GTACA	TRAC KO gRNA 1
GTTCTGTCTTTTGGTCAGGACAACCGTCTAGCTATAAGTGCTGCAGGGTGTGAGAA ACTCCTATTGCTGGACGATGTCTCTTACGAGGCATTAGCACACCGATTTTGATTCT CAAAC	TRAC KO gRNA 2
GTTCTGTCTTTTGGTCAGGACAACCGTCTAGCTATAAGTGCTGCAGGGTGTGAGAA ACTCCTATTGCTGGACGATGTCTCTTACGAGGCATTAGCACTGATTCTCAAACAAA TGTGT	TRAC KO gRNA 3

[0662]

向导 RNA	基因外显子
GTTCTGTCTTTTGGTCAGGACAACCGTCTAGCTATAAGTGCTGCAGGGTGTGAGAA ACTCCTATTGCTGGACGATGTCTCTTACGAGGCATTAG**CAC**TCAAACAAATGTGTC ACAAA	TRAC KO gRNA 4
GTTCTGTCTTTTGGTCAGGACAACCGTCTAGCTATAAGTGCTGCAGGGTGTGAGAA ACTCCTATTGCTGGACGATGTCTCTTACGAGGCATTAGCACGTTTGAGAATCAAAA TCGGT	TRAC KO gRNA 5
GTTCTGTCTTTTGGTCAGGACAACCGTCTAGCTATAAGTGCTGCAGGGTGTGAGAA ACTCCTATTGCTGGACGATGTCTCTTACGAGGCATTAGCACTGATGTGTATATCAC AGACA	TRAC KO gRNA 6
GTTCTGTCTTTTGGTCAGGACAACCGTCTAGCTATAAGTGCTGCAGGGTGTGAGAA ACTCCTATTGCTGGACGATGTCTCTTACGAGGCATTAGCACGTTGCTCCAGGCCAC AGCAC	TRAC KO gRNA 7
GTTCTGTCTTTTGGTCAGGACAACCGTCTAGCTATAAGTGCTGCAGGGTGTGAGAA ACTCCTATTGCTGGACGATGTCTCTTACGAGGCATTAGCACTTCCAGAAGACACCT TCTTC	TRAC KO gRNA 8
GTTCTGTCTTTTGGTCAGGACAACCGTCTAGCTATAAGTGCTGCAGGGTGTGAGAA ACTCCTATTGCTGGACGATGTCTCTTACGAGGCATTAGCACCAGAAGACACCTTCT TCCCC	TRAC KO gRNA 9
GTTCTGTCTTTTGGTCAGGACAACCGTCTAGCTATAAGTGCTGCAGGGTGTGAGAA ACTCCTATTGCTGGACGATGTCTCTTACGAGGCATTAGCACGGTTCCGAATCCTCC TCCTG	TRAC KO gRNA 10
GTTCTGTCTTTTGGTCAGGACAACCGTCTAGCTATAAGTGCTGCAGGGTGTGAGAA ACTCCTATTGCTGGACGATGTCTCTTACGAGGCATTAGCACGGAACCCAATCACTG ACAGG	TRAC KO gRNA 11
GACCTATAGGGTCAATGAATCTGTGCGTGTGCCATAAGTAATTAAAAATTACCCAC CACAGGAGCACCTGAAAACAGGTGCTTGGCACAGAGTCTCTCAGCTGGTACA	TRAC KO gRNA 1
GACCTATAGGGTCAATGAATCTGTGCGTGTGCCATAAGTAATTAAAAATTACCCAC CACAGGAGCACCTGAAAACAGGTGCTTGGCACACCGATTTTGATTCTCAAAC	TRAC KO gRNA 2
GACCTATAGGGTCAATGAATCTGTGCGTGTGCCATAAGTAATTAAAAATTACCCAC CACAGGAGCACCTGAAAACAGGTGCTTGGCACTGATTCTCAAACAAATGTGT	TRAC KO gRNA 3

[0663] 表1B:续

[0664]

向导 RNA	基因外显子
GACCTATAGGGTCAATGAATCTGTGCGTGTGCCATAAGTAATTAAAAATTACCCAC CACAGGAGCACCTGAAAACAGGTGCTTGGCACTCAAACAAATGTGTCACAAA	TRAC KO gRNA 4
GACCTATAGGGTCAATGAATCTGTGCGTGTGCCATAAGTAATTAAAAATTACCCAC CACAGGAGCACCTGAAAACAGGTGCTTGGCACGTTTGAGAATCAAAATCGGT	TRAC KO gRNA 5
GACCTATAGGGTCAATGAATCTGTGCGTGTGCCATAAGTAATTAAAAATTACCCAC CACAGGAGCACCTGAAAACAGGTGCTTGGCACTGATGTGTATATCACAGACA	TRAC KO gRNA 6
GACCTATAGGGTCAATGAATCTGTGCGTGTGCCATAAGTAATTAAAAATTACCCAC CACAGGAGCACCTGAAAACAGGTGCTTGGCACGTTGCTCCAGGCCACAGCAC	TRAC KO gRNA 7
GACCTATAGGGTCAATGAATCTGTGCGTGTGCCATAAGTAATTAAAAATTACCCAC CACAGGAGCACCTGAAAACAGGTGCTTGGCACTTCCAGAAGACACCTTCTTC	TRAC KO gRNA 8

[0665]

向导 RNA	基因外显子
GACCTATAGGGTCAATGAATCTGTGCGTGTGCCATAAGTAATTAAAAATTACCCAC CACAGGAGCACCTGAAAACAGGTGCTTGGCACCAGAAGACACCTTCTTCCCC	TRAC KO gRNA 9
GACCTATAGGGTCAATGAATCTGTGCGTGTGCCATAAGTAATTAAAAATTACCCAC CACAGGAGCACCTGAAAACAGGTGCTTGGCACGGTTCCGAATCCTCCTCCTG	TRAC KO gRNA 10
GACCTATAGGGTCAATGAATCTGTGCGTGTGCCATAAGTAATTAAAAATTACCCAC CACAGGAGCACCTGAAAACAGGTGCTTGGCACGGAACCCAATCACTGACAGG	TRAC KO gRNA 11

[0666] 前11个gRNA用于BhCas12b核酸酶。第二组11个gRNA用于BvCas12b核酸酶。在第一种情况下,支架序列加粗显示。

[0667] 在一些实施方案中,可以使用BE4碱基编辑器将本发明的编码嵌合抗原受体的核酸靶向至TRAC基因座。在一些实施方案中,使用CRISPR/Cas9碱基编辑系统将嵌合抗原受体靶向至TRAC基因座。

[0668] 为了产生上述碱基编辑,从受试者采集免疫细胞并使其与两个或更多个向导RNA以及包含核酸可编程DNA结合蛋白(napDNAbp)和胞苷脱氨酶或腺苷脱氨酶的核碱基编辑器多肽接触。在一些实施方案中,使所采集的免疫细胞与至少一种核酸接触,其中该至少一种核酸编码两个或更多个向导RNA以及包含核酸可编程DNA结合蛋白(napDNAbp)和胞苷脱氨酶的核碱基编辑器多肽。在一些实施方案中,gRNA包含核苷酸类似物。这些核苷酸类似物可抑制gRNA被细胞过程降解。表2提供待用于gRNA的靶标序列。

[0669]

表 2：示例性靶标序列

靶标蛋白	靶标残基	gRNA 靶标	碱基编辑器	密码子变化	残基功能
NFATC1	R118	CTCGATGCGAGGACTCTCCA	BE	CGC>CAC	钙调磷酸酶结合
	I119	TCTCGATGCGAGGACTCTCC	ABE	ATC>ACC	钙调磷酸酶结合
	E120	CATCGAGATAACCTCGTGCT	ABE	GAG>GGG	钙调磷酸酶结合
	S172	TGGCCGGGCTCAGGCACGAG	BE	AGC>AAC	磷酸化
	W396	GCCCACTGGTAGGGGTGCTG	ABE	TGG>CGG	钙调磷酸酶结合
	R439	TGGGCTCGGTGGTGGGACTT	BE	CGA>CAA	DNA 结合
	H441	CGAGCCCACTACGAGACGGA	ABE	CAC>CGC	DNA 结合
	Y442	CTCGTAGTGGGCTCGGTGGT	ABE	TAC>CAC	DNA 结合
	K452	GCCGTGAAGGCGTCGGCCGG	ABE	AAG>GGG	DNA 结合
	R540	GTTTCTGAGTTTCAGGATTC	BE	AGA>AAA	DNA 结合
	R555	CATCGGGAGGAAGAACACAC	ABE	AGG>GGG	DNA 结合
	K556	GGAGGAAGAACACACGGGTA	ABE	AAG>GGG	DNA 结合
	Q589	GAGCGCTGGGCTGCATCAGA	BE	CAG>CAT	DNA 结合
NFATC2	E114	TGATCTCGATCCGAGGGCTC	BE	GAG>AAA	钙调磷酸酶结合
	I115	ACGGAGTGATCTCGATCCGA	ABE	ATC>ACC	钙调磷酸酶结合
	R253	GCGGAGGCATTCGTGCGCCG	ABE	AGG>GGG	NLS

[0670]

靶标蛋白	靶标残基	gRNA 靶标	碱基编辑器	密码子变化	残基功能
	S99	GCCGCGCTCAGAAACTTCTG	BE	AGC>AAC	磷酸化
	S107	GGGCCTCGGGCCTGAGCCCT	BE	TCG>TTG	磷酸化
	S148	CCTCGGGCTGGCGGCCACCC	BE	AGC>AAC	磷酸化
	S236	CCACTCGCCCGTGCCCCGTC	BE	TCG>TTG	磷酸化
	S255	GCATTCGTGCGCCGAGGCCT	BE	TCG>TTG	磷酸化
	S268	GAGCCTCACCCCAGCGCTCC	BE	TCA>TTA	磷酸化
	S274	GAGGGGCTCCGGGAGCGCTG	BE	AGC>AAC	磷酸化
	S326	AGGGCTGGTCTTTCCACATCT	BE	AGC>AAC	磷酸化
NFATC4	S213	GCGGGGAGCCCAGGCCAAAG	ABE	TCC>CCC	磷酸化
AKT1	T305	GCCACCATGAAGACCTTTTG	BE	ACC>ATT	磷酸化
	T312	TTGCGGCACACCTGAGTACC	BE	ACA>ATA	磷酸化
	S473	GTAGGAGAACTGGGGGAAGT	ABE	TCC>CCC	磷酸化
	Y474	CTCCTACTCGGCCAGCGGCA	ABE	TAC>TGC	磷酸化
AKT2	T309	GAAAACCTTCTGTGGGACCC	BE	ACC>ATT	磷酸化

[0671]

靶标蛋白	靶标残基	gRNA 靶标	碱基编辑器	密码子变化	残基功能
	S474	AGTAGGAGAACTGGGGGAAG	ABE	TCC>CCC	磷酸化
BLIMP1	C608 (ZF2)	GTTGCAAGTCTGACATTTGA	ABE	TGC>CGC	DNA 结合
	C608 (ZF2)	GTTGCAAGTCTGACATTTGA	BE	TGC>TAC	DNA 结合
	H621 (ZF2)	GAAACACTACCTGGTACACA	BE	CAC>TAT	DNA 结合
	C636 (ZF3)	TGTGGCAGACCTACAGTGTA	BE	TGC>TAC	DNA 结合
	C664 (ZF4)	GGGCACACCTTGCATTGGTA	ABE	TGC>CGC	DNA 结合
	剪接位点 1	CTGCGCACCTGGCATTCATG	BE		
GCN2 切口酶（IDO 途径）	外显子 1 SD	CCTACCGGTCCGCAAGCGTC	BE		敲除
	外显子 2 SD	ACTCACACATCTGGATAGGT	BE		敲除
	外显子 5 SD	GACTTACCTAGACCTTCCTG	BE		敲除
CBL-B	C373	AATCTTACAGAGCTGAAAAG	BE	TGT>TAT	E3 泛素连接酶
	Y665.1	CATCATATTCTTCACTTCCA	ABE	TAT>CAC	

[0672]

靶标蛋白	靶标残基	gRNA 靶标	碱基编辑器	密码子变化	残基功能
	Y665.2	AAGAATATGATGTTCCTCCC	ABE	TAT>TGT	
	K907	CCCCTAAACCACGACCGCGC	ABE	AAA>GGG	
	R911	TCCTGCGCGGTCGTGGTTTA	BE	CGC>CAC	
SHP1	Y377	CCCTACTCTGTGACCAACTG	ABE	TAC>TGC	
IRF4	R96	CGCAGGCGCGTCTTCCAGGT	BE	CGC>CAC	DNA 结合
	R98	GCACCGCAGGCGCGTCTTCC	BE	CGG>CAG	DNA 结合
	K103	GAACAAGAGCAATGACTTTG	ABE	AAG>GGG	DNA 结合
PD1	外显子 1 STOP	CACCTACCTAAGAACCATCC	BE		敲除
	外显子 2 STOP	GGGGTTCCAGGGCCTGTCTG	BE		敲除
TET2	H1386	GACTTGCACAACATGCAGAA	BE	CAC>TAC	DNA 结合
	R1302	TTGCCAGAAGCAAGATCCCA	ABE	AGA>GGG	DNA 结合
	S1290	CCATGAACAACCAAAAGAGA	ABE	TCA>CCA	DNA 结合

[0673]

靶标蛋白	靶标残基	gRNA 靶标	碱基编辑器	密码子变化	残基功能
SMARCA4	T353	TCACCCCCATCCAGAAGCCG	BE	ACC>ATT	磷酸化
	S610	ATCTGGCTGGTCTCGTCCAG	BE	AGC>ATC	磷酸化
	S613	GATGAGCGACCTCCCGGTGA	ABE	AGC>GGC	磷酸化
	S695	AGACAGCGATGACGTCTCTG	ABE	AGC>GGC	磷酸化
	S699	ACGTCTCTGAGGTGGACGCG	BE	TCT>TTT	磷酸化
	S1452	TTAGGGGAGAGTTTCTCGGC	ABE	TCC>CCC	磷酸化
	S1575	GGAGAGTGAGGAGGAGGAAG	ABE	AGT>GGT	磷酸化
	S1586	AAGGCTCCGAATCCGAATCT	BE	TCC>TTT	磷酸化
	S1627	ATCGTCACTCACGACCGGCT	BE	AGT>AAT	磷酸化
	S1631	TGACAGTGAGGAGGAACAAG	ABE	AGT>GGT	磷酸化
CDK4	P173	CACCCGTGGTTGTTACACTC	BE	CCC>CTT	
ZAP70	S144	CATCAGCCAGGCCCCCGCAGG	ABE	AGC>TGC	磷酸化
	Y292	GGTGTATCCATCTGAGTTGA	ABE	TAC>CAC	磷酸化
	Y292	GGGTGTATCCATCTGAGTTG	ABE	TAC>CAC	磷酸化

[0674]

靶标蛋白	靶标残基	gRNA 靶标	碱基编辑器	密码子变化	残基功能
	R360	GCGCAAGAAGCAGATCGACG	BE	CGC>TGC	超等位基因活性
	Y598	TTACTACAGCCTGGCCAGCA	ABE	TAC>TGC	磷酸化

[0675] 本发明使用的胞苷和腺苷脱氨酶核碱基编辑器可以作用于DNA,包括单链DNA。呈现了使用它们在免疫细胞内的靶标核碱基序列中生成修饰的方法。

[0676] 在某些实施方案中,本文所提供的融合蛋白包含一个或多个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包含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没有核酸酶活性的Cas9结构域(dCas9),或切断双螺旋DNA分子的一条链的Cas9结构域,称为Cas9切口酶(nCas9)。不与受缚于特定例如,催化性残基(例如,H840)的存在维持Cas9裂解与被靶向的核碱基相对的未经编辑的(例如,未经甲基化的)链的活性。催化性残基的突变(例如,D10到A10)阻止含有被靶向的A残基的经编辑链的裂解。此类Cas9变体可以在基于gRNA定义的靶标序列的特异性位置处生成单链DNA断裂(切口),导致非编辑链的修复,最终导致非编辑链上的核碱基改变。

[0677] 核碱基编辑器

[0678] 本文公开一种碱基编辑器或核碱基编辑器,其用于编辑、修饰或改变多核苷酸中的靶标核苷酸序列。本文描述一种核碱基编辑器或碱基编辑器,其包含多核苷酸可编程核苷酸结合结构域和核碱基编辑结构域(例如,腺苷脱氨酶、胞苷脱氨酶)。当与结合的向导多核苷酸(例如,gRNA)协同作用时,多核苷酸可编程核苷酸结合结构域可特异性地结合至靶标多核苷酸序列(即,经由结合的向导核酸的碱基与靶标多核苷酸序列的碱基之间的互补碱基配对),并因此将碱基编辑器定位至待编辑的所希望的靶标核酸序列。在一些实施方案中,靶标多核苷酸序列包含单链DNA或双链DNA。在一些实施方案中,靶标多核苷酸序列包含RNA。在一些实施方案中,靶标多核苷酸序列包含DNA-RNA杂交物。

[0679] 多核苷酸可编程核苷酸结合结构域

[0680] 应了解,多核苷酸可编程核苷酸结合结构域也可以包括结合RNA的核酸可编程蛋白质。例如,多核苷酸可编程核苷酸结合结构域可以与核酸缔合,该核酸将多核苷酸可编程核苷酸结合结构域引导至RNA。其他核酸可编程DNA结合蛋白也处于本公开范畴内,但它们未在本公开中具体列出。

[0681] 碱基编辑器的多核苷酸可编程核苷酸结合结构域自身可包含一个或多个结构域。例如,多核苷酸可编程核苷酸结合结构域可包含一个或多个核酸酶结构域。在一些实施方案中,多核苷酸可编程核苷酸结合结构域的核酸酶结构域可包含核酸内切酶或核酸外切酶。本文中,术语“核酸外切酶”是指能够从游离末端消化核酸(例如,RNA或DNA)的蛋白质或多肽,且术语“核酸内切酶”是指能够催化(例如,切割)核酸(例如,DNA或RNA)中的内部区域的蛋白质或多肽。在一些实施方案中,核酸内切酶可切割双链核酸的单链。在一些实施方案中,核酸内切酶可切割双链核酸分子的两条链。在一些实施方案中,多核苷酸可编程核苷酸结合结构域可以是脱氧核糖核酸酶。在一些实施方案中,多核苷酸可编程核苷酸结合结构域可以是核糖核酸酶。

[0682] 在一些实施方案中,多核苷酸可编程核苷酸结合结构域的核酸酶结构域可以将靶标多核苷酸的零条、一条或两条链切口。在一些实施方案中,多核苷酸可编程核苷酸结合结构域可包含切口酶结构域。本文中,“切口酶”是指一种包含核酸酶结构域的多核苷酸可编程核苷酸结合结构域,其能够仅切割双螺旋核酸分子(例如,DNA)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核苷酸结合结构域而衍生自完全催化活性(例如,天然)形式的多核苷酸可编程核苷酸结合结构域。例如,若多核苷酸可编程核苷酸结合结构域包含衍生自Cas9的切口酶,则衍生自Cas9的切口

酶结构域可包括D10A突变和在位置840处的组氨酸。在这些实施方案中,残基H840保持催化活性,并因此可切割核酸双螺旋的一条链。在另一示例中,衍生自Cas9的切口酶结构域可包含H840A突变,而在位置10处的氨基酸残基保持为D。在一些实施方案中,切口酶可以通过去除切口酶活性所不需要的核酸酶的全部或一部分而衍生自完全催化活性(例如,天然)形式的多核苷酸可编程核苷酸结合结构域。例如,若多核苷酸可编程核苷酸结合结构域包含衍生自Cas9的切口酶,则衍生自Cas9的切口酶结构域可包含RuvC结构域或HNH结构域中的全部或一部分的缺失。

[0683] 示例性的催化活性Cas9的氨基酸序列如下:

[0684] MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD。

[0685] 因此,包含具有切口酶结构域的多核苷酸可编程核苷酸结合结构域的碱基编辑器能够在特异性多核苷酸靶标序列(例如,通过结合的向导核酸的互补序列确定)处生成单链DNA断裂(切口)。在一些实施方案中,被包含切口酶结构域(例如,衍生自Cas9的切口酶结构域)的碱基编辑器切割的核酸双螺旋靶标多核苷酸序列的链是未经碱基编辑器编辑的链(即,被碱基编辑器切割的链与包含待编辑的碱基的链相反)。在其他实施方案中,包含切口酶结构域(例如,衍生自Cas9的切口酶结构域)的碱基编辑器可以切割待被靶向以进行编辑的DNA分子的链。在这些实施方案中,非靶向链未经切割。

[0686] 本文还提供包含催化上死亡(即,不能切割靶标多核苷酸序列)的多核苷酸可编程核苷酸结合结构域的碱基编辑器。本文中,术语“催化上死亡的”和“核酸酶死亡的”互换使用,是指一种多核苷酸可编程核苷酸结合结构域,其具有一个或多个导致其失去切割核酸链的能力的突变和/或缺失。在一些实施方案中,催化上死亡的多核苷酸可编程核苷酸结合结构域碱基编辑器可能缺乏核酸酶活性,这是在一个或多个核酸酶结构域中的特异性点突

变的结果。例如,在碱基编辑器包含Cas9结构域的情况下,Cas9可包含D10A突变和H840A突变两者。此类突变将两个核酸酶结构域失活,从而导致核酸酶活性的丧失。在其他实施方案中,催化上死亡的多核苷酸可编程核苷酸结合结构域可包含对话结构域(例如,RuvC1和/或HNH结构域)的全部或一部分的一个或多个缺失。在进一步的实施方案中,催化上死亡的多核苷酸可编程核苷酸结合结构域包含点突变(例如,D10A或H840A)以及核酸酶结构域的全部或一部分的缺失。

[0687] 本文也设想能够从先前功能性版本的多核苷酸可编程核苷酸结合结构域生成催化上死亡的多核苷酸可编程核苷酸结合结构域的突变。例如,在催化上死亡的Cas9("dCas9")的情况下,提供了具有除D10A和H840A之外的突变的变体,其导致无核酸酶活性的Cas9。举例而言,此类突变包括在D10和H840处的其他氨基酸置换,或Cas9的核酸酶结构域内的其他置换(例如,HNH核酸酶次结构域和/或RuvC1次结构域中的置换)。基于本公开和本领域知识,其他合适的无核酸酶活性的dCas9结构域对于本领域技术人员是显而易见的并且出于本公开的范畴内。此等其他示例性合适的无核酸酶活性的Cas9结构域包括但不限于,D10A/H840A、D10A/D839A/H840A和D10A/D839A/H840A/N863A突变结构域(参见,例如,Prashant等人,CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering.Nature Biotechnology.2013;31(9):833-838,其整体内容通过引用并入本文)。

[0688] 可并入碱基编辑器的多核苷酸可编程核苷酸结合结构域的非限制性示例包括衍生自CRISPR蛋白的结构域、限制性核酸酶、巨核酸酶、TAL核酸酶(TALEN)和锌指核酸酶(ZFN)。在一些实施方案中,碱基编辑器包含具有天然或修饰的蛋白质或其部分的多核苷酸可编程核苷酸结合结构域,经由结合的向导多核苷酸,该多核苷酸可编程核苷酸结合结构域能够在CRISPR(即,成簇的规则间隔短回文重复序列)介导的核酸修饰期间结合至核酸序列。本文中,该蛋白质称为"CRISPR蛋白"。据此,本文公开一种包含具有CRISPR蛋白的全部或一部分的多核苷酸可编程核苷酸结合结构域的碱基编辑器(即,包含CRISPR蛋白的全部或一部分作为结构域(也称为衍生自CRISPR蛋白的结构域)的碱基编辑器)。与野生型或天然版本的CRISPR蛋白相比,并入碱基编辑器中的衍生自CRISPR蛋白的结构域可以经修饰。例如,如下所述,相对于野生型或天然版本的CRISPR蛋白,衍生自CRISPR蛋白的结构域可包含一种或多种突变、插入、缺失、重排和/或重组。

[0689] CRISPR是一种适应性免疫系统,其提供针对移动基因元件(病毒、转座因子和接合质粒)的保护。CRISPR簇含有间隔序列、与前驱移动元件互补的序列和靶标侵入核酸。CRISPR簇被转录并处理为CRISPR RNA(crRNA)。在II型CRISPR系统中,pre-crRNA的正确处理需要反向编码的小RNA(tracrRNA)、内源性核糖核酸酶3(rnc)和Cas9蛋白。tracrRNA充当向导用于核糖核酸酶3辅助的pre-crRNA处理。随后,Cas9/crRNA/tracrRNA核酸内切地切割与间隔序列互补的线性或环状dsDNA靶标。不与crRNA互补的靶标链首先被核酸内切地切断,然后被3'-5'核酸外切地修剪。在自然界,DNA结合和切割通常需要蛋白质和两种RNA。但是,单向导RNA("sgRNA"或简写为"gRNA")可以经工程改造以将多种crRNA和tracrRNA并入单一RNA物质中。参见,例如,Jinek M.等人,Science 337:816-821(2012),其整体内容通过引用并入本文。。Cas9识别CRISPR重复序列中的短基序(PAM或前间区序列邻近基序)以帮助区分自我与非自我。

[0690] 在一些实施方案中,本文所述的方法可利用经工程改造的Cas蛋白。向导RNA(gRNA)是一种短合成RNA,由Cas结合所必需的支架序列和限定待修饰的基因组靶标的使用者定义的~20个核苷酸的间隔序列。因此,技术人员可改变Cas蛋白的基因组靶标,特异性部分地通过gRNA靶向序列对于基因组靶标的特异性与对于该基因剩余部分的特异性如何来确定。

[0691] 在一些实施方案中,gRNA支架序列如下:

[0692] GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU。

[0693] 在一种实施方案中,RNA支架包含茎环。在一种实施方案中,RNA支架包含核酸序列:

[0694] GUUUUUGUACUCUCAAGAUUUAAGUAACUGUACAACGAAACUUACACAGUUACUUAAAUCUUGCAGAAGCUACAAAGAUAAGGCUUCAUGCCGAAAUCAACACCCUGUCAUUUUAUGGCAGGGUG。

[0695] 在一种实施方案中,RNA支架包含标准茎环。在一种实施方案中,RNA支架包含核酸序列:

[0696] GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCU*mU*mU*mU

[0697] 其中m=2'-O-甲基修饰且*=3'硫代磷酸酯核苷酸见链接(即,第一个3'端RNA残基在这里显示)。

[0698] 在一种实施方案中,RNA支架包含核酸序列:

[0699] GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU

[0700] 其中m=2'-O-甲基修饰且*=3'硫代磷酸酯核苷酸见链接(即,第一个3'端RNA残基在这里显示)。

[0701] 在一种实施方案中,酿脓链球菌sgRNA支架多核苷酸序列如下:

[0702] GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC。

[0703] 在一种实施方案中,金黄色葡萄球菌sgRNA支架多核苷酸序列如下:

[0704] GUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCUACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGA。

[0705] 在一种实施方案中,BhCas12b sgRNA支架具有以下多核苷酸序列:

[0706] GUUCUGTCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUGCAGGGUGUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUACGAGGCAUUAGCAC。

[0707] 在一种实施方案中,BvCas12b sgRNA支架具有以下多核苷酸序列:

[0708] GACCUAUAGGGUCAAUGAAUCUGUGCGUGUGCCAUAAGUAAUUAAAAAUUACCCACCACAGGAGCACCUGAAAACAGGUGCUUGGCAC。

[0709] 在一种实施方案中,RNA支架包含非标准序列。在一种实施方案中,RNA支架包含核酸序列:

[0710] GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGGACCGAGUCGGUGCU*mU*mU*mU

[0711] 其中m＝2'-O-甲基修饰且*＝3'硫代磷酸酯核苷酸见链接(即,第一个3'端RNA残基在这里显示)。

[0712] 在一些实施方案中,并入碱基编辑器中的衍生自CRISPR蛋白的结构域是核酸内切酶(例如,脱氧核糖核酸酶或核糖核酸酶),其在与结合的向导核酸协同作用时能够结合靶标多核苷酸。在一些实施方案中,并入碱基编辑器中的衍生自CRISPR蛋白的结构域是切口酶,其在与结合的向导核酸协同作用时能够结合靶标多核苷酸。在一些实施方案中,并入碱基编辑器中的衍生自CRISPR蛋白的结构域是催化上死亡的结构域,其在与结合的向导核酸协同作用时能够结合靶标多核苷酸。在一些实施方案中,被碱基编辑器的衍生自CRISPR蛋白的结构域结合的靶标多核苷酸是DNA。在一些实施方案中,被碱基编辑器的衍生自CRISPR蛋白的结构域结合的靶标多核苷酸是RNA。

[0713] 本文中可使用的Cas蛋白包括1类和2类。Cas蛋白的非限制性示例包括Cas1、Cas1B、Cas2、Cas3、Cas4、Cas5、Cas5d、Cas5t、Cas5h、Cas5a、Cas6、Cas7、Cas8、Cas9(也称为Csn1或Csx12)、Cas10、Csy1、Csy2、Csy3、Csy4、Cse1、Cse2、Cse3、Cse4、Cse5e、Csc1、Csc2、Csa5、Csn1、Csn2、Csm1、Csm2、Csm3、Csm4、Csm5、Csm6、Cmr1、Cmr3、Cmr4、Cmr5、Cmr6、Csb1、Csb2、Csb3、Csx17、Csx14、Csx10、Csx16、CsaX、Csx3、Csx1、Csx1S、Csf1、Csf2、Cs0、Csf4、Csd1、Csd2、Cst1、Cst2、Csh1、Csh2、Csa1、Csa2、Csa3、Csa4、Csa5、Cas12a/Cpf1、Cas12b/C2c1、Cas12c/C2c3、Cas12d/CasY、Cas12e/CasX、Cas12g、Cas12h、Cas12i、和Cas12j/CasΦ、CARF、DinG、其同源物、或其经修饰的版本。未修饰的CRISPR酶可具有DNA切割活性,诸如Cas9,其具有两个核酸内切酶结构域:RuvC和HNH。CRISPR酶可引导靶标序列的一条或两条链的切割,诸如在靶标序列内和/或在靶标序列的互补序列内。例如,CRISPR酶可引导从靶标序列的第一个或最后一个核苷酸起约1、2、3、4、5、6、7、8、9、10、15、20、25、50、100、200、500或更多个碱基对内的一条或两条链的切割。

[0714] 可使用以下载体,该载体编码相对于对应的野生型酶突变的CRISPR酶,使得突变的CRISPR酶缺乏切割含有靶标序列的靶标多核苷酸的一条或两条链的能力。Cas9可以指一种多肽,其与野生型示例性Cas9多肽(例如,来自酿脓链球菌的Cas9)具有至少或至少约50%、60%、70%、80%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%的序列一致性和/或序列同源性。Cas9可以指一种多肽,其与野生型示例性Cas9多肽(例如,来自酿脓链球菌)具有至多或至多约50%、60%、70%、80%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%的序列一致性和/或序列同源性。Cas9可以指野生型或经修饰形式的Cas9蛋白,该经修饰形式的Cas9蛋白可以包含氨基酸变化诸如缺失、插入、置换、变体、突变、融合、嵌合或其任意组合。

[0715] 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以下的Cas9的全部或一部分:溃疡棒状杆菌(Corynebacterium ulcerans)(NCBI Refs:NC_015683.1,NC_017317.1);白喉棒状杆菌(Corynebacteriumdiphtheria)(NCBI Refs:NC_016782.1,NC_016786.1);螺原体(Spiroplasma syrphidicola)(NCBI Ref:NC_021284.1);中间普雷沃菌(Prevotella intermedia)(NCBI Ref:NC_017861.1);螺原体(Spiroplasma taiwanense)(NCBI Ref:NC_021846.1);海豚链球菌(Streptococcus iniae)(NCBI Ref:NC_021314.1);Belliella baltica(NCBI Ref:NC_018010.1);扭曲冷弯曲菌(Psychroflexus torquis)(NCBI Ref:NC_018721.1);嗜热链球菌(Streptococcus

thermophilus)(NCBI Ref:YP_820832.1);无害利斯特菌(Listeria innocua)(NCBI Ref:NP_472073.1);空肠弯曲杆菌(Campylobacter jejuni)(NCBI Ref:YP_002344900.1);脑膜炎奈瑟氏菌(Neisseria meningitidis)(NCBI Ref:YP_002342100.1);酿脓链球菌(Streptococcus pyogenes)或金黄色葡萄球菌(Staphylococcus aureus)。

[0716] 核碱基编辑器的Cas9结构域

[0717] Cas9核酸酶序列和结构是本领域技术人员周知的(参见,例如,"Complete genome sequence of an M1 strain of Streptococcus pyogenes."Ferretti,J.J.等人,Natl.Acad.Sci.U.S.A.98:4658-4663(2001);"CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III."Deltcheva E.等人,Nature471:602-607(2011);和"A programmabledual-RNA-guided DNA endonuclease in adaptive bacterial immunity."Jinek M.等人,Science 337:816-821(2012),其各自的整体内容通过引用并入本文)。Cas9直系同源物已经在各种物种中描述,包括酿脓链球菌(S.pyogenes)和嗜热链球菌(S.thermophilus)。基于本公开,其他合适的Cas9核酸酶和序列对于本领域技术人员将是显而易见的,并且此类Cas9核酸酶和序列包括来自生物体和基因座的Cas9序列,如Chylinski,Rhun,and Charpentier,"The tracrRNA and Cas9families of type IICRISPR-Cas immunity systems"(2013)RNA Biology 10:5,726-737中所公开的;其整体内容通过引用并入本文。

[0718] 在一些实施方案中,核酸可编程DNA结合结构域(napDNAbp)是Cas9结构域。非限制性的示例性Cas9结构域在本文中提供。Cas9结构域可以是核酸酶活性的Cas9结构域、无核酸酶活性的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的结构域。例如,Cas9结构域可以是切断双螺旋核酸的两条链(例如,双螺旋DNA分子的两条链)的Cas9结构域。在一些实施方案中,Cas9结构域包含本文中详述的氨基酸序列中的任一者。在一些实施方案中,Cas9结构域包含氨基酸序列,该氨基酸序列与本文详述的氨基酸序列中的任一者至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相同。在一些实施方案中,Cas9结构域包含氨基酸序列,与本文详述的氨基酸序列中的任一者相比,该氨基酸序列具有1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、21、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50个或更多个突变。在一些实施方案中,Cas9结构域包含氨基酸序列,与本文详述的氨基酸序列中的任一者相比,该氨基酸序列具有至少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。

[0719] 在一些实施方案中,提供包含Cas9的片段的蛋白质。例如,在一些实施方案中,蛋白质包含一个或两个Cas9结构域:(1)Cas9的gRNA结合结构域;或(2)Cas9的DNA切割结构域。在一些实施方案中,包含Cas9或其片段的蛋白质称为"Cas9变体"。Cas9变体与Cas9或其片段共享同源性。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、45、46、47、48、49、50或更多个氨基酸变化。在一些实施方案中,Cas9变体包含Cas9的片段(例如,gRNA结合结构域或DNA切割结构域),使得该片段与野生型Cas9的对应片段至少约70%相同、至少约80%相同、至少约90%相同、至少约95%相同、至少约96%相同、至少约97%相同、至少约98%相同、至少约99%相同、至少约99.5%相同、或至少约99.9%相同。在一些实施方案中,该片段是对应的野生型Cas9的氨基酸长度的至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%相同、至少96%、至少97%、至少98%、至少99%、或至少99.5%。在一些实施方案中,该片段是至少100个氨基酸的长度。在一些实施方案中,该片段是至少100、150、200、250、300、350、400、450、500、550、600、650、700、750、800、850、900、950、1000、1050、1100、1150、1200、1250、或至少1300个氨基酸的长度。

[0720] 在一些实施方案中,如本文所提供的Cas9融合蛋白包含Cas9蛋白(例如,本文所提供的Cas9序列中的一者)的全长氨基酸序列。然而,在其他实施方案中,本文所提供的融合蛋白不包含全长Cas9序列,而仅包含其一个或多个片段。合适的Cas9结构域和Cas9片段的示例性氨基酸序列提供在本文中,且Cas9结构域和片段的其他合适的序列对于本领域技术人员将是显而易见的。

[0721] Cas9蛋白可与向导RNA缔合,该向导RNA将Cas9蛋白引导至与该向导RNA具有互补性的特异性DNA序列。在一些实施方案中,多核苷酸可编程核苷酸结合结构域是Cas9结构域,例如,核酸酶活性的Cas9、Cas9切口酶(nCas9)、或无核酸酶活性的Cas9(dCas9)。核酸可编程DNA结合蛋白的示例包括而不限于,Cas9(例如,dCas9和nCas9)、Cas12a/Cpf1、Cas12b/C2c1、Cas12c/C2c3、Cas12d/CasY、Cas12e/CasX、Cas12g、Cas12h、Cas12i和Cas12j/CasΦ。Cas酶的非限制性示例包括Cas1、Cas1B、Cas2、Cas3、Cas4、Cas5、Cas5d、Cas5t、Cas5h、Cas5a、Cas6、Cas7、Cas8、Cas8a、Cas8b、Cas8c、Cas9(也称为Csn1或Csx12)、Cas10、Cas10d、Cas12a/Cpf1、Cas12b/C2c1、Cas12c/C2c3、Cas12d/CasY、Cas12e/CasX、Cas12g、Cas12h、Cas12i、Cas12j/CasΦ、Csy1、Csy2、Csy3、Csy4、Cse1、Cse2、Cse3、Cse4、Cse5e、Csc1、Csc2、Csa5、Csn1、Csn2、Csm1、Csm2、Csm3、Csm4、Csm5、Csm6、Cmr1、Cmr3、Cmr4、Cmr5、Cmr6、Csb1、Csb2、Csb3、Csx17、Csx14、Csx10、Csx16、CsaX、Csx3、Csx1、Csx1S、Csx11、Csf1、Csf2、Cs0、Csf4、Csd1、Csd2、Cst1、Cst2、Csh1、Csh2、Csa1、Csa2、Csa3、Csa4、Csa5、II型Cas效应蛋白、V型Cas效应蛋白、VI型Cas效应蛋白、CARF、DinG、其同源、或其经修饰或工程改造的版本。

[0722] 在一些实施方案中,野生型Cas9对应于来自酿脓链球菌(Streptococcus pyogenes)的Cas9(NCBI参考序列:NC_017053.1)。示例性酿脓链球菌(Streptococcus pyogenes)Cas9(spCas9)核酸序列提供如下:

[0723] ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGATCACTGATGATTATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTATCAAAAAAAATCTTATAGGGGCTCTTTTATTTGGCAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGTCTTTTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAAGTTGCTTATCATGAGAAATATCCAACTATCTATCATCTGCGAAAAAAATTGGCAGATTCTACTGATAAAGCGGATTTGC

GCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAGATTTAAATCCTGAT
AATAGTGATGTGGACAAACTATTTATCCAGTTGGTACAAATCTACAATCAATTATTTGAAGAAAACCCTATTAACGC
AAGTAGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGCTCAGC
TCCCCGGTGAGAAGAGAAATGGCTTGTTTGGGAATCTCATTGCTTTGTCATTGGGATTGACCCCTAATTTTAAATCA
AATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTAGATAATTTATTGGC
GCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGATGCTATTTTACTTTCAGATATCC
TAAGAGTAAATAGTGAAATAACTAAGGCTCCCCTATCAGCTTCAATGATTAAGCGCTACGATGAACATCATCAAGAC
TTGACTCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTATAAAGAAATCTTTTTTGATCAATCAAAAAA
CGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAAATTTATCAAACCAATTTTAGAAAAAA
TGGATGGTACTGAGGAATTATTGGTGAAACTAAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACAACGGC
TCTATTCCCCATCAAATTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTTAAA
AGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTCCATTGGCGCGTGGCAATA
GTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCATGGAATTTTGAAGAAGTTGTCGATAAAGGT
GCTTCAGCTCAATCATTTATTGAACGCATGACAAACTTTGATAAAAATCTTCCAAATGAAAAAGTACTACCAAAACA
TAGTTTGCTTTATGAGTATTTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTGAGGGAATGCGAAAAC
CAGCATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAACAAATCGAAAAGTAACCGTTAAG
CAATTAAAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGAAATTTCAGGAGTTGAAGATAGATTTAA
TGCTTCATTAGGCGCCTACCATGATTTGCTAAAAATTATTAAAGATAAAGATTTTTTGGATAATGAAGAAAATGAAG
ATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAAGATAGGGGGATGATTGAGGAAAGACTTAAAACATAT
GCTCACCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTTTGTCTCGAAA
ATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTAGATTTTTTGAAATCAGATGGTTTTGCCAATC
GCAATTTTATGCAGCTGATCCATGATGATAGTTTGACATTTAAAGAAGATATTCAAAAAGCACAGGTGTCTGGACAA
GGCCATAGTTTACATGAACAGATTGCTAACTTAGCTGGCAGTCCTGCTATTAAAAAAGGTATTTTACAGACTGTAAA
AATTGTTGATGAACTGGTCAAAGTAATGGGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACGTGAAAATCAGA
CAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAATCGAAGAAGGTATCAAAGAATTAGGAAGTCAG
ATTCTTAAAGAGCATCCTGTTGAAAATACTCAATTGCAAAATGAAAAGCTCTATCTCTATTATCTACAAAATGGAAG
AGACATGTATGTGGACCAAGAATTAGATATTAATCGTTTAAGTGATTATGATGTCGATCACATTGTTCCACAAAGTT
TCATTAAAGACGATTCAATAGACAATAAGGTACTAACGCGTTCTGATAAAAATCGTGGTAAATCGGATAACGTTCCA
AGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAACTTCTAAACGCCAAGTTAATCACTCAACGTAAGTT
TGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTGAACTTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTG
AAACTCGCCAAATCACTAAGCATGTGGCACAAATTTTGGATAGTCGCATGAATACTAAATACGATGAAAATGATAAA
CTTATTCGAGAGGTTAAAGTGATTACCTTAAAATCTAAATTAGTTTCTGACTTCCGAAAAGATTTCCAATTCTATAA
AGTACGTGAGATTAACAATTACCATCATGCCCATGATGCGTATCTAAATGCCGTCGTTGGAACTGCTTTGATTAAGA
AATATCCAAAACTTGAATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTCGTAAAATGATTGCTAAGTCT
GAGCAAGAAATAGGCAAAGCAACCGCAAAATATTTCTTTTACTCTAATATCATGAACTTCTTCAAAACAGAAATTAC
ACTTGCAAATGGAGAGATTCGCAAACGCCCTCTAATCGAAACTAATGGGGAAACTGGAGAAATTGTCTGGGATAAAG
GGCGAGATTTTGCCACAGTGCGCAAAGTATTGTCCATGCCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAGACA
GGCGGATTCTCCAAGGAGTCAATTTTACCAAAAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGATCC
AAAAAAATATGGTGGTTTTGATAGTCCAACGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGAAAT

CGAAGAAGTTAAAATCCGTTAAAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAAATCCGATT
GACTTTTTAGAAGCTAAAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACCTAAATATAGTCTTTTTGA
GTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAATTACAAAAAGGAAATGAGCTGGCTCTGCCAAGCA
AATATGTGAATTTTTTATATTTAGCTAGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAA
TTGTTTGTGGAGCAGCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCGTGTTATTTT
AGCAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAGACAAACCAATACGTGAACAAGCAGAAA
ATATTATTCATTTATTTACGTTGACGAATCTTGGAGCTCCCGCTGCTTTTAAATATTTTGATACAACAATTGATCGT
AAACGATATACGTCTACAAAAGAAGTTTTAGATGCCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACACG
CATTGATTTGAGTCAGCTAGGAGGTGACTGA

[0724] 示例性酿脓链球菌(Streptococcus pyogenes)Cas9(spCas9)氨基酸序列提供如下:

MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETA

[0725] EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD
NSDVDKLFIQLVQIYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNG
LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKN
LSDAILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ
SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQ
IHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET
ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNAS
LGAYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRGMIEERLKTYAHLFDDKVMK
QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDI
QKAQVSGQGHSLHEQIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQ

[0726] TTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQE
LDINRLSDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQL
LNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDE
NDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKL
ESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLI
ETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIAR
KKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDF
LEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS
HYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRD
KPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYET
RIDLSQLGGD

[0727] (单下划线:HNH结构域;双下划线:RuvC结构域)

[0728] 在一些实施方案中,野生型Cas9对应于或包含以下核苷酸序列:

[0729] ATGGATAAAAAGTATTCTATTGGTTTAGACATCGGCACTAATTCCGTTGGATGGGCTGTCATAACCGAT
GAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACCGTCATTCGATTAAAAAGAATCTTATCGG
TGCCCTCCTATTCGATAGTGGCGAAACGGCAGAGGCGACTCGCCTGAAACGAACCGCTCGGAGAAGGTATACACGTC
GCAAGAACCGAATATGTTACTTACAAGAAATTTTTAGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTTCACCGT
TTGGAAGAGTCCTTCCTTGTCGAAGAGGACAAGAAACATGAACGGCACCCCATCTTTGGAAACATAGTAGATGAGGT
GGCATATCATGAAAAGTACCCAACGATTTATCACCTCAGAAAAAAGCTAGTTGACTCAACTGATAAAGCGGACCTGA
GGTTAATCTACTTGGCTCTTGCCCATATGATAAAGTTCCGTGGGCACTTTCTCATTGAGGGTGATCTAAATCCGGAC
AACTCGGATGTCGACAAACTGTTCATCCAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGAACCCTATAAATGC
AAGTGGCGTGGATGCGAAGGCTATTCTTAGCGCCCGCCTCTCTAAATCCCGACGGCTAGAAAACCTGATCGCACAAT
TACCCGGAGAGAAGAAAAATGGGTTGTTCGGTAACCTTATAGCGCTCTCACTAGGCCTGACACCAAATTTTAAGTCG
AACTTCGACTTAGCTGAAGATGCCAAATTGCAGCTTAGTAAGGACACGTACGATGACGATCTCGACAATCTACTGGC
ACAAATTGGAGATCAGTATGCGGACTTATTTTTGGCTGCCAAAAACCTTAGCGATGCAATCCTCCTATCTGACATAC
TGAGAGTTAATACTGAGATTACCAAGGCGCCGTTATCCGCTTCAATGATCAAAAGGTACGATGAACATCACCAAGAC
TTGACACTTCTCAAGGCCCTAGTCCGTCAGCAACTGCCTGAGAAATATAAGGAAATATTCTTTGATCAGTCGAAAAA
CGGGTACGCAGGTTATATTGACGGCGGAGCGAGTCAAGAGGAATTCTACAAGTTTATCAAACCCATATTAGAGAAGA
TGGATGGGACGGAAGAGTTGCTTGTAAAACTCAATCGCGAAGATCTACTGCGAAAGCAGCGGACTTTCGACAACGGT
AGCATTCCACATCAAATCCACTTAGGCGAATTGCATGCTATACTTAGAAGGCAGGAGGATTTTTATCCGTTCCTCAA
AGACAATCGTGAAAAGATTGAGAAAATCCTAACCTTTCGCATACCTTACTATGTGGGACCCCTGGCCCGAGGGAACT
CTCGGTTCGCATGGATGACAAGAAAGTCCGAAGAAACGATTACTCCATGGAATTTTGAGGAAGTTGTCGATAAAGGT
GCGTCAGCTCAATCGTTCATCGAGAGGATGACCAACTTTGACAAGAATTTACCGAACGAAAAAGTATTGCCTAAGCA
CAGTTTACTTTACGAGTATTTCACAGTGTACAATGAACTCACGAAAGTTAAGTATGTCACTGAGGGCATGCGTAAAC
CCGCCTTTCTAAGCGGAGAACAGAAGAAAGCAATAGTAGATCTGTTATTCAAGACCAACCGCAAAGTGACAGTTAAG
CAATTGAAAGAGGACTACTTTAAGAAAATTGAATGCTTCGATTCTGTCGAGATCTCCGGGGTAGAAGATCGATTTAA
TGCGTCACTTGGTACGTATCATGACCTCCTAAAGATAATTAAAGATAAGGACTTCCTGGATAACGAAGAGAATGAAG
ATATCTTAGAAGATATAGTGTTGACTCTTACCCTCTTTGAAGATCGGGAAATGATTGAGGAAAGACTAAAAACATAC
GCTCACCTGTTCGACGATAAGGTTATGAAACAGTTAAAGAGGCGTCGCTATACGGGCTGGGGACGATTGTCGCGGAA
ACTTATCAACGGGATAAGAGACAAGCAAAGTGGTAAAACTATTCTCGATTTTCTAAAGAGCGACGGCTTCGCCAATA
GGAACTTTATGCAGCTGATCCATGATGACTCTTTAACCTTCAAAGAGGATATACAAAAGGCACAGGTTTCCGGACAA
GGGGACTCATTGCACGAACATATTGCGAATCTTGCTGGTTCGCCAGCCATCAAAAAGGGCATACTCCAGACAGTCAA
AGTAGTGGATGAGCTAGTTAAGGTCATGGGACGTCACAAACCGGAAAACATTGTAATCGAGATGGCACGCGAAAATC
AAACGACTCAGAAGGGGCAAAAAAACAGTCGAGAGCGGATGAAGAGAATAGAAGAGGGTATTAAAGAACTGGGCAGC
CAGATCTTAAAGGAGCATCCTGTGGAAAATACCCAATTGCAGAACGAGAAACTTTACCTCTATTACCTACAAAATGG
AAGGGACATGTATGTTGATCAGGAACTGGACATAAACCGTTTATCTGATTACGACGTCGATCACATTGTACCCCAAT
CCTTTTTGAAGGACGATTCAATCGACAATAAAGTGCTTACACGCTCGGATAAGAACCGAGGGAAAAGTGACAATGTT
CCAAGCGAGGAAGTCGTAAAGAAAATGAAGAACTATTGGCGGCAGCTCCTAAATGCGAAACTGATAACGCAAAGAAA
GTTCGATAACTTAACTAAAGCTGAGAGGGGTGGCTTGTCTGAACTTGACAAGGCCGGATTTATTAAACGTCAGCTCG
TGGAAACCCGCCAAATCACAAAGCATGTTGCACAGATACTAGATTCCCGAATGAATACGAAATACGACGAGAACGAT
AAGCTGATTCGGGAAGTCAAAGTAATCACTTTAAAGTCAAAATTGGTGTCGGACTTCAGAAAGGATTTTCAATTCTA

TAAAGTTAGGGAGATAAATAACTACCACCATGCGCACGACGCTTATCTTAATGCCGTCGTAGGGACCGCACTCATTA
AGAAATACCCGAAGCTAGAAAGTGAGTTTGTGTATGGTGATTACAAAGTTTATGACGTCCGTAAGATGATCGCGAAA
AGCGAACAGGAGATAGGCAAGGCTACAGCCAAATACTTCTTTTATTCTAACATTATGAATTTCTTTAAGACGGAAAT
CACTCTGGCAAACGGAGAGATACGCAAACGACCTTTAATTGAAACCAATGGGGAGACAGGTGAAATCGTATGGGATA
AGGGCCGGGACTTCGCGACGGTGAGAAAAGTTTTGTCCATGCCCCAAGTCAACATAGTAAAGAAAACTGAGGTGCAG
ACCGGAGGGTTTTCAAAGGAATCGATTCTTCCAAAAAGGAATAGTGATAAGCTCATCGCTCGTAAAAAGGACTGGGA
CCCGAAAAAGTACGGTGGCTTCGATAGCCCTACAGTTGCCTATTCTGTCCTAGTAGTGGCAAAAGTTGAGAAGGGAA
AATCCAAGAAACTGAAGTCAGTCAAAGAATTATTGGGGATAACGATTATGGAGCGCTCGTCTTTTGAAAAGAACCCC
ATCGACTTCCTTGAGGCGAAAGGTTACAAGGAAGTAAAAAAGGATCTCATAATTAAACTACCAAAGTATAGTCTGTT
TGAGTTAGAAAATGGCCGAAAACGGATGTTGGCTAGCGCCGGAGAGCTTCAAAAGGGGAACGAACTCGCACTACCGT
CTAAATACGTGAATTTCCTGTATTTAGCGTCCCATTACGAGAAGTTGAAAGGTTCACCTGAAGATAACGAACAGAAG
CAACTTTTTGTTGAGCAGCACAAACATTATCTCGACGAAATCATAGAGCAAATTTCGGAATTCAGTAAGAGAGTCAT
CCTAGCTGATGCCAATCTGGACAAAGTATTAAGCGCATACAACAAGCACAGGGATAAACCCATACGTGAGCAGGCGG
AAAATATTATCCATTTGTTTACTCTTACCAACCTCGGCGCTCCAGCCGCATTCAAGTATTTTGACACAACGATAGAT
CGCAAACGATACACTTCTACCAAGGAGGTGCTAGACGCGACACTGATTCACCAATCCATCACGGGATTATATGAAAC
TCGGATAGATTTGTCACAGCTTGGGGGTGACGGATCCCCCAAGAAGAAGAGGAAAGTCTCGAGCGACTACAAAGACC
ATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGGCTGCAGGA

[0730] 在一些实施方案中,野生型Cas9对应于或包含以下氨基酸序列:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD

[0731] NSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG
LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKN
LSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ
SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQ

IHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET
ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNAS
LGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMK
QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDI
QKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMAREN
QTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQ
[0732] LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD
ENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK
LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPL
IETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA
RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLA
SHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR
DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYE
TRIDLSQLGGD

[0733] (单下划线:HNH结构域;双下划线:RuvC结构域)。

[0734] 在一些实施方案中,野生型Cas9对应于来自酿脓链球菌(Streptococcus pyogenes)的Cas9(NCBI参考序列:NC_002737.2(核苷酸序列如下):

[0735] ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGATCACTGAT
GAATATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTATCAAAAAAAATCTTATAGG
GGCTCTTTTATTTGACAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTC
GGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGA
CTTGAAGAGTCTTTTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAAGT
TGCTTATCATGAGAAATATCCAACTATCTATCATCTGCGAAAAAAATTGGTAGATTCTACTGATAAAGCGGATTTGC
GCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAGATTTAAATCCTGAT
AATAGTGATGTGGACAAACTATTTATCCAGTTGGTACAAACCTACAATCAATTATTTGAAGAAAACCCTATTAACGC
AAGTGGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGCTCAGC
TCCCCGGTGAGAAGAAAAATGGCTTATTTGGGAATCTCATTGCTTTGTCATTGGGTTTGACCCCTAATTTTAAATCA
AATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTAGATAATTTATTGGC
GCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGATGCTATTTTACTTTCAGATATCC
TAAGAGTAAATACTGAAATAACTAAGGCTCCCCTATCAGCTTCAATGATTAAACGCTACGATGAACATCATCAAGAC
TTGACTCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTATAAAGAAATCTTTTTTGATCAATCAAAAAA
CGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAAATTTATCAAACCAATTTTAGAAAAAA
TGGATGGTACTGAGGAATTATTGGTGAAACTAAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACAACGGC
TCTATTCCCCATCAAATTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTTAAA

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

[0736] 在一些实施方案中,野生型Cas9对应于来自酿脓链球菌(Streptococcus pyogenes)的Cas9(Uniprot参考序列:Q99ZW2(氨基酸序列如下)):

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD
NSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG

[0737] LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKN
LSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ
SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQ
IHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET
ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNAS
LGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMK
QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDI
QKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMAREN
QTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQ
LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD

[0738] ENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK
LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPL
IETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA
RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLA
SHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR
DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYE
TRIDLSQLGGD

[0739] (单下划线:HNH结构域;双下划线:RuvC结构域)

[0740] 在一些实施方案中,Cas9是指来自以下的Cas9:溃疡棒状杆菌(Corynebacterium ulcerans)(NCBI Refs:NC_015683.1,NC_017317.1)、白喉棒状杆菌(Corynebacterium diphtheria)(NCBI Refs:NC_016782.1,NC_016786.1)、螺原体(Spiroplasma syrphidicola)(NCBI Ref:NC_021284.1)、中间普雷沃菌(Prevotella intermedia)(NCBI Ref:NC_017861.1)、螺原体(Spiroplasma taiwanense)(NCBI Ref:NC_021846.1)、海豚链球菌(Streptococcus iniae)(NCBI Ref:NC_021314.1)、Belliella baltica(NCBI Ref:NC_018010.1)、扭曲冷弯曲菌(Psychroflexus torquis)(NCBI Ref:NC_018721.1)、嗜热链球菌(Streptococcus thermophilus)(NCBI Ref:YP_820832.1)、无害利斯特菌(Listeria innocua)(NCBI Ref:NP_472073.1)、空肠弯曲杆菌(Campylobacter jejuni)(NCBI Ref:P_002344900.1)、或脑膜炎奈瑟氏菌(Neisseria meningitidis)(NCBI Ref:YP_002342100.1);或是指来自任何其他生物体的Cas9。

[0741] 应了解,其他Cas9蛋白(例如,核酸酶死亡的Cas9(dCas9)、Cas9切口酶(nCas9)、或核酸酶活性的Cas9),包括其变体和同源物,处于本公开的范畴内。示例性Cas9蛋白包括但不限于以下提供的那些。在一些实施方案中,Cas9蛋白是核酸酶死亡的Cas9(dCas9)。在一些实施方案中,Cas9蛋白是Cas9切口酶(nCas9)。在一些实施方案中,Cas9蛋白是核酸酶活性的Cas9。

[0742] 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结构域可结合至双螺旋核酸分子(例如,经由gRNA分子)而不切割该双螺旋核酸分子的任一条链。在一些实施方案中,无核酸酶活性的dCas9结构域包含本文详述的氨基酸序列的D10X突变和H840X突变,或本文所提供的任何氨基酸序列中的对应突变,其中X是任何氨基酸变化。在一些实施方案中,无核酸酶活性的dCas9结构域包含本文详述的氨基酸序列的D10A突变和H840A突变,或本文所提供的任何氨基酸序列中的对应突变。作为一个示例,无核酸酶活性的Cas9结构域包含克隆载体pPlatTET-gRNA2(登录号BAV54124)中详述的氨基酸序列。

[0743] 示例性的无催化活性的Cas9(dCas9)的氨基酸序列如下:MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD(参见,例如,Qi等人,“Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression.”Cell.2013;152(5):1173-83,其整体内容通过引用并入本文)。

[0744] 基于本公开和本领域知识,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将是显而易见的并且出于本公开的范畴内。此等其他示例性合适的无核酸酶活性的Cas9结构域包括但不限于,D10A/H840A、D10A/D839A/H840A和D10A/D839A/H840A/N863A突变结构域(参见,例如,Prashant等人,CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome

engineering.Nature Biotechnology.2013;31(9):833-838,其整体内容通过引用并入本文)。

[0745] 在一些实施方案中,Cas9核酸酶具有失活的(例如,灭活的)DNA切割结构域,换言之,Cas9是切口酶,称为“nCas9”蛋白(对于“切口酶”Cas9)。无核酸酶活性的Cas9蛋白可以互换地称为“dCas9”蛋白(对于核酸酶“死亡的”Cas9)或无催化活性的Cas9。用于生成具有无活性的DNA切割结构域的Cas9蛋白(或其片段)是已知的(参见,例如,Jinek等人,Science.337:816-821(2012);Qi等人,“Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression”(2013)Cell.28;152(5):1173-83,其各自的整体内容通过引用并入本文)。例如,已知Cas9的DNA切割结构域包括两个次结构域:HNH核酸酶次结构域和RuvC1次结构域。HNH次结构域切割与gRNA互补的链,而RuvC1次结构域切割非互补链。这些次结构域中的突变可以沉默Cas9的核酸酶活性。例如,突变D10A和H840A将酿脓链球菌Cas9的核酸酶活性完全灭活(Jinek等人,Science.337:816-821(2012);Qi等人,Cell.28;152(5):1173-83(2013))。

[0746] 在一些实施方案中,dCas9结构域包含氨基酸序列,该氨基酸序列与本文所提供的dCas9结构域中的任一者至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相同。在一些实施方案中,Cas9结构域包含氨基酸序列,与本文详述的氨基酸序列中的任一者相比,该氨基酸序列具有1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、21、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50个或更多个突变。在一些实施方案中,Cas9结构域包含氨基酸序列,与本文详述的氨基酸序列中的任一者相比,该氨基酸序列具有至少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。

[0747] 在一些实施方案中,dCas9对应于或者部分或整体地包含具有一个或多个突变的Cas9氨基酸序列,该突变将Cas9核酸酶活性灭活。在一些实施方案中,无核酸酶活性的dCas9结构域包含本文详述的氨基酸序列的D10X突变和H840X突变,或本文所提供的任何氨基酸序列中的对应突变,其中X是任何氨基酸变化。在一些实施方案中,无核酸酶活性的dCas9结构域包含本文详述的氨基酸序列的D10A突变和H840A突变,或本文所提供的任何氨基酸序列中的对应突变。在一些实施方案中,无核酸酶活性的Cas9结构域包含克隆载体pPlatTET-gRNA2(登录号BAV54124)中详述的氨基酸序列。

[0748] 在一些实施方案中,dCas9包含dCas9(D10A和H840A)的氨基酸序列:

[0749]
```
MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD
NSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG
LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKN
LSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ
```

SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQ
IHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET
ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNAS
LGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMK
QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDI
QKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMAREN
QTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQ
[0750] LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD
ENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK
LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPL
IETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA
RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLA
SHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR
DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYE
TRIDLSQLGGD

[0751] (单下划线:HNH结构域;双下划线:RuvC结构域)。

[0752] 在一些实施方案中,示例性的无催化活性的Cas9(dCas9)的氨基酸序列如下:

[0753] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
AIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKP

IREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0754] (参见,例如,Qi等人,“Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression.”Cell.2013;152(5):1173-83,其整体内容通过引用并入本文)。

[0755] 在一些实施方案中,示例性的无催化活性的Cas9(dCas9)的氨基酸序列如下:

[0756] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
AIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKP
IREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0757] 在一些实施方案中,Cas9结构域包含D10A突变,而上文提供的氨基酸序列中的位置840处或本文提供的任何氨基酸序列中对应位置处的残基保持为组氨酸。

[0758] 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,这利润导致无核酸酶活性的Cas9(dCas9)。举例而言,此类突变包括在D10和H840处的其他氨基酸置换,或Cas9的核酸酶结构域内的其他置换(例如,HNH核酸酶次结构域和/或RuvC1次结构域中的置换)。在一些实施方案中,提供dCas9的变体或同源物,它们至少约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,提供具有氨基酸序列的dCas9变体,该氨基酸序列更短或更长,相异约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。

[0759] 基于本公开和本领域知识,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将是显而易见的并且出于本公开的范畴内。此等其他示例性合适的无核酸酶活性的Cas9结构域包括但不限于,D10A/H840A、D10A/D839A/H840A和D10A/D839A/H840A/N863A突变结构域(参见,例如,Prashant等人,CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome

engineering.Nature Biotechnology.2013;31(9):833-838,其整体内容通过引用并入本文)。

[0760] 在一些实施方案中,Cas9核酸酶具有失活的(例如,灭活的)DNA切割结构域,换言之,Cas9是切口酶,称为“nCas9”蛋白(对于“切口酶”Cas9)。Cas9切口酶可以是能够仅切割双螺旋核酸分子(例如,双螺旋DNA分子)的一条链的Cas9蛋白。在一些实施方案中,Cas9切口酶切割双螺旋核酸分子的靶标链,意味着Cas9切口酶切割与结合至Cas9的gRNA(例如,sgRNA)碱基配对的链。在一些实施方案中,Cas9切口酶包含D10A突变并且在位置840处具有组氨酸。在一些实施方案中,Cas9切口酶切割双螺旋核酸分子的非靶标、非碱基编辑链,意味着Cas9切口酶切割不与结合至Cas9的gRNA(例如,sgRNA)碱基配对的链。在一些实施方案中,Cas9切口酶包含H840A突变并且在位置10处具有天冬氨酸残基,或对应的突变。在一些实施方案中,Cas9切口酶包含氨基酸序列,该氨基酸序列与本文所提供的Cas9切口酶中的任一者至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相同。基于本公开和本领域知识,其他合适的Cas9切口酶对于本领域技术人员将是显而易见的并且出于本公开的范畴内。

[0761] 示例性的催化Cas9切口酶(nCas9)的氨基酸序列如下:

[0762] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0763] 在一些实施方案中,Cas9是指来自古生菌(例如,纳米古生菌)的Cas9,其构成单细胞原核微生物的领域和王国。在一些实施方案中,核酸可编程DNA结合蛋白可以是asX或CasY蛋白,其已经描述于例如Burstein等人,“New CRISPR-Cas systems from uncultivated microbes.”Cell Res.2017Feb 21.doi:10.1038/cr.2017.21中,其整体内容通过引用并入本文。使用基因组解析的宏基因组学,鉴定了大量CRISPR-Cas系统,包括在

古生菌生命领域首次报告的Cas9。这种不同的Cas9蛋白是在很少研究的纳米古生菌中作为活性CRISPR-Cas系统发现的。在细菌中发现了两种先前未知的系统,CRISPR-CasX和CRISPR-CasY,它们是迄今发现的最为紧凑的系统中的一个。在一些实施方案中,在本文所述的碱基编辑器系统中,Cas9被CasX或CasX的变体替换。在一些实施方案中,在本文所述的碱基编辑器系统中,Cas9被CasY或CasY的变体替换。应了解,其他RNA引导的DNA结合蛋白可以用作核酸可编程DNA结合蛋白(napDNAbp),并且出于本公开的范畴内。

[0764] 在一些实施方案中,本文所提供的任何融合蛋白的核酸可编程DNA结合蛋白(napDNAbp)可以是CasX或CasY蛋白。在一些实施方案中,napDNAbp是CasX蛋白。在一些实施方案中,napDNAbp是CasY蛋白。在一些实施方案中,可编程核苷酸结合蛋白包含氨基酸序列,该氨基酸序列与本文所述的CasX或CasY蛋白至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%、或至少99.5%相同。在一些实施方案中,可编程核苷酸结合蛋白是天然存在的CasX或CasY蛋白。在一些实施方案中,napDNAbp包含氨基酸序列,该氨基酸序列与天然存在的CasX或CasY蛋白至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%、或至少99.5%相同。应了解,来自其他细菌物种的CasX和CasY也可根据本公开使用。

[0765] 示例性CasX((uniprot.org/uniprot/F0NN87;uniprot.org/uniprot/F0NH53)tr|F0NN87|F0NN87_SULIHCRISPR相关Casx蛋白OS＝冰岛硫化叶菌(Sulfolobus islandicus)(菌株HVE10/4)GN＝SiH_0402PE＝4SV＝1)氨基酸序列如下:

[0766] MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAERRGKAKKKKGEEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKECEEVSAPSFVKPEFYEFGRSPGMVERTRRVKLEVEPHYLIIAAAGWVLTRLGKAKVSEGDYVGVNVFTPTRGILYSLIQNVNGIVPGIKPETAFGLWIARKVVSSVTNPNVSVVRIYTISDAVGQNPTTINGGFSIDLTKLLEKRYLLSERLEAIARNALSISSNMRERYIVLANYIYEYLTG SKRLEDLLYFANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG.

[0767] 示例性CasX(>tr|F0NH53|F0NH53_SULIR CRISPR相关蛋白,Casx OS＝冰岛硫化叶菌(菌株REY15A)GN＝SiRe_0771PE＝4SV＝1)氨基酸序列如下:

[0768] MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAERRGKAKKKKGEEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKECEEVSAPSFVKPEFYKFGRSPGMVERTRRVKLEVEPHYLIMAAAGWVLTRLGKAKVSEGDYVGVNVFTPTRGILYSLIQNVNGIVPGIKPETAFGLWIARKVVSSVTNPNVSVVSIYTISDAVGQNPTTINGGFSIDLTKLLEKRDLLSERLEAIARNALSISSNMRERYIVLANYIYEYLTGSKRLEDLLYFANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG.

[0769] δ变形菌纲CasX

[0770] MEKRINKIRKKLSADNATKPVSRSGPMKTLLVRVMTDDLKKRLEKRRKKPEVMPQVISNNAANNLRMLLDDYTKMKEAILQVYWQEFKDDHVGLMCKFAQPASKKIDQNKLKPEMDEKGNLTTAGFACSQCGQPLFVYKLEQVSEKGKAYTNYFGRCNVAEHEKLILLAQLKPVKDSDEAVTYSLGKFGQRALDFYSIHVTKESTHPVKPLAQIAGNRYASGPVGKALSDACMGTIASFLSKYQDIIIEHQKVVKGNQKRLESLRELAGKENLEYPSVTLPPQPHTKEGVDfAYNEVIARVRMWVNLNLWQKLKLSRDDAKPLLRLKGFPSFPVVERRENEVDWWNTINEVKKLIDAKRDMGRVFWSGVTAEKRNTILEGYNYLPNENDHKKREGSLENPKKPAKRQFGDLLLYLEKKYAGDWGKVFDEAWERIDKKIAGLTSHIEREEARNAEDAQSKAVLTDWLRAKASFVLERLKEMDEKEFYACEIQLQKWYGDLRGNPFAVEAENRVVDISGFSIGSDGHSI

QYRNLLAWKYLENGKREFYLLMNYGKKGRIRFTDGTDIKKSGKWQGLLYGGGKAKVIDLTFDPDDEQLIILPLAFGT
RQGREFIWNDLLSLETGLIKLANGRVIEKTIYNKKIGRDEPALFVALTFERREVVDPSNIKPVNLIGVARGENIPAV
IALTDPEGCPLPEFKDSSGGPTDILRIGEGYKEKQRAIQAAKEVEQRRAGGYSRKFASKSRNLADDMVRNSARDLFY
HAVTHDAVLVFANLSRGFGRQGKRTFMTERQYTKMEDWLTAKLAYEGLTSKTYLSKTLAQYTSKTCSNCGFTITYAD
MDVMLVRLKKTSDGWATTLNNKELKAEYQITYYNRYKRQTVEKELSAELDRLSEESGNNDISKWTKGRRDEALFLLK
KRFSHRPVQEQFVCLDCGHEVHAAEQAALNIARSWLFLNSNSTEFKSYKSGKQPFVGAWQAFYKRRLKEVWKPNA

[0771] 示例性CasY

[0772] ((ncbi.nlm.nih.gov/protein/APG80656.1)>APG80656.1 CRISPR相关蛋白CasY[未培养的Parcubacteria组细菌])氨基酸序列如下:

[0773] MSKRHPRISGVKGYRLHAQRLEYTGKSGAMRTIKYPLYSSPSGGRTVPREIVSAINDDYVGLYGLSNF
DDLYNAEKRNEEKVYSVLDFWYDCVQYGAVFSYTAPGLLKNVAEVRGGSYELTKTLKGSHLYDELQIDKVIKFLNK
KEISRANGSLDKLKKDIIDCFKAEYRERHKDQCNKLADDIKNAKKDAGASLGERQKKLFRDFFGISEQSENDKPSF
TNPLNLTCCLLPFDTVNNNRNRGEVLFNKLKEYAQKLDKNEGSLEMWEYIGIGNSGTAFSNFLGEGFLGRLRENKI
TELKKAMMDITDAWRGQEQEEELEKRLRILAALTIKLREPKFDNHWGGYRSDINGKLSSWLQNYINQTVKIKEDLK
GHKKDLKKAKEMINRFGESDTKEEAVVSSLLESIEKIVPDDSADDEKPDIPAIAIYRRFLSDGRLTLNRFVQREDV
QEALIKERLEAEKKKKPKKRKKKSDAEDEKETIDFKELFPHLAKPLKLVPNFYGDSKRELYKKYKNAAIYTDALWK
AVEKIYKSAFSSSLKNSFFDTDFDKDFFIKRLQKIFSVYRRFNTDKWKPIVKNSFAPYCDIVSLAENEVLYKPKQS
RSRKSAAIDKNRVRLPSTENIAKAGIALARELSVAGFDWKDLLKKEEHEEYIDLIELHKTALALLLAVTETQLDIS
ALDFVENGTVKDFMKTRDGNLVLEGRFLEMFSQSIVFSELRGLAGLMSRKEFITRSAIQTMNGKQAELLYIPHEFQ
SAKITTPKEMSRAFLDLAPAEFATSLEPESLSEKSLLKLKQMRYYPHYFGYELTRTGQGIDGGVAENALRLEKSPV
KKREIKCKQYKTLGRGQNKIVLYVRSSYYQTQFLEWFLHRPKNVQTDVAVSGSFLIDEKKVKTRWNYDALTVALEP
VSGSERVFVSQPFTIFPEKSAEEEGQRYLGIDIGEYGIAYTALEITGDSAKILDQNFISDPQLKTLREEVKGLKLD
QRRGTFAMPSTKIARIRESLVHSLRNRIHHLALKHKAKIVYELEVSRFEEGKQKIKKVYATLKKADVYSEIDADKN
LQTTVWGKLAVASEISASYTSQFCGACKKLWRAEMQVDETITTQELIGTVRVIKGGTLIDAIKDFMRPPIFDENDT
PFPKYRDFCDKHHISKKMRGNSCLFICPFCRANADADIQASQTIALLRYVKEEKKVEDYFERFRKLKNIKVLGQMK
KI。

[0774] Cas9核酸酶具有两个功能性核酸内切酶结构域:RuvC和HNH。Cas9在靶标结合时经历构象变化,将核酸酶结构域定位以切割靶标DNA的相反链。The end result of Cas9介导的DNA切割的最终结果是靶标DNA(PAM序列上游的～3-4核苷酸)内的双链断裂(DSB)。然后,所得DSB通过两种通用修复途径的一种修复:(1)有效但易错的非同源末端接合(NHEJ)途径;或(2)效率较低但高保真度同源性引导修复(HDR)途径。

[0775] 非同源末端接合(NHEJ)和/或同源性引导修复(HDR)的“效率”可以通过任何便利的方法计算。例如,在一些实施方案中,效率可以表达为成功HDR的百分比。例如,surveyor核酸酶测定可用于生成切割产物,并且产物与底物的比率可以用于计算该百分比。例如,可使用surveyor核酸酶,该酶直接切割含有新整合的限制性序列的DNA,而该DNA作为成功HDR的结果。经切割的底物越多,表明HDR的百分比越高(HDR的效率越高)。作为说明性示例,HDR的分数(百分比)可以使用以下等式计算:[(经切割的产物)/(底物加上经切割的产物)](例如,(b+c)/(a+b+c),其中“a”是DNA底物的带强度,且“b”和“c”是经切割的产物)。

[0776] 在一些实施方案中,效率可以表达为成功NHEJ的百分比。例如,T7核酸内切酶I测

定可用于生成切割产物,并且产物与底物的比率可以用于计算NHEJ百分比。T7核酸内切酶I切割由野生型和突变体DNA链杂交而来的错配杂双螺旋DNA(NHEJ在原始断裂的位点处生成小随机插入或缺失(插入缺失))。切割越多,表明NHEJ的百分比越高(NHEJ的效率越高)。作为说明性示例,NHEJ的分数(百分比)可以使用以下等式计算:$(1-(1-(b+c)/(a+b+c))^{1/2})\times 100$,其中"a"是DNA底物的带强度,且"b"和"c"是切割产物(Ran等人,Cell.2013Sep.12;154(6):1380-9;和Ran等人,Nat Protoc.2013Nov.;8(11):2281-2308)。

[0777] NHEJ修复途径是活性最高的修复途径,并且其频繁地在DSB位点处造成小核苷酸插入或缺失(插入缺失)。NHEJ介导的DSB修复的随机性具有重要的实践意义,因为表达Cas9和gRNA或向导多核苷酸的细胞群可以导致不同配置的突变。在大多数实施方案中,NHEJ在靶标DNA中产生小插入缺失,导致氨基酸缺失、插入或框移突变,造成所靶向基因的开放阅读框(ORF)内的过早终止密码子。理想的最终结果是所靶向基因内的功能丧失突变。

[0778] 尽管NHEJ介导的DSB修复经常中断该基因的开放阅读框,但同源性引导修复(HDR)可用于生成从单核苷酸变化到大插入范围内的特异性核苷酸变化,如荧光团或标签的添加。

[0779] 为了利用HDR进行基因编辑,可以使用gRNA和Cas9或Cas9切口酶将含有所希望的序列的DNA修复模板递送至感兴趣的细胞类型中。修复模板可含有所希望的编辑以及位于靶标上游和下游紧邻该靶标的同源序列(称为左右同源臂)。每个同源臂的长度可以取决于被引入的变化的尺寸,且插入越大,所需的同源臂越长。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和外源性修复模板的细胞内,HDR的效率通常也较低(<10%的经修饰的等位基因)。HDR的效率可以通过将细胞同步化而增强,因为HDR在细胞周期的S期和G2期发生。化学或遗传学地抑制牵涉入NHEJ的基因也可增加HDR频率。

[0780] 在一些实施方案中,Cas9是经修饰的Cas9。给定的gRNA靶向序列可在部分同源性存在之处的整个基因组中具有额外位点。这些位点称为脱靶并且在涉及gRNA时应予以考虑。除了优化gRNA设计之外,CRISPR特异性也可以通过修饰Cas9而增加。Cas9通过两个核酸酶结构域RuvC和HNH的组合活性产生双链断裂(DSB)。Cas9切口酶,SpCas9的D10A突变体,保持一个核酸酶结构域并产生DNA切口而非DSB。切口酶系统也可以与HDR介导的编辑组合以进行特异性基因编辑。

[0781] 在一些实施方案中,Cas9是变体Cas9蛋白。变体Cas9多肽具有氨基酸序列,该氨基酸序列较之于野生型Cas9蛋白的氨基酸序列相异一个氨基酸(例如,具有缺失、插入、置换、融合)。在一些实例中,变体Cas9多肽具有氨基酸变化(例如,缺失、插入或置换),该变化降低Cas9多肽的核酸酶活性。例如,在一些实例中,变体Cas9多肽具有低于50%、低于40%、低于30%、低于20%、低于10%、低于5%、或低于1%的对应野生型Cas9蛋白的核酸酶活性。在一些实施方案中,变体Cas9蛋白基本上没有核酸酶活性。当受试Cas9蛋白是基本上没有核酸酶活性的变体Cas9蛋白时,它可以称为"dCas9"。

[0782] 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蛋白表现出低于约20%、低于约15%、低于约10%、低于约5%、低于约1%、或低于约0.1%的野生型Cas9蛋白(例如,野生型Cas9蛋白)的核酸内切酶活性。

[0783] 在一些实施方案中,变体Cas9蛋白可切割向导靶标核酸的互补链,但切割双链向

导靶标序列的非互补链的能力降低。例如,变体Cas9蛋白可具有降低RuvC结构域的功能的突变(氨基酸置换)。作为非限制性示例,在一些实施方案中,变体Cas9蛋白具有D10A(在氨基酸位置10处用天冬氨酸代替丙氨酸),并且因此可切割双链向导靶标序列的互补链但切割双链向导靶标序列的非互补链的能力降低(因此当变体Cas9蛋白切割双链靶标核酸时导致单链断裂(SSB)而非双链断裂(DSB))(参见,例如,Jinek等人,Science.2012Aug.17;337(6096):816-21)。

[0784] 在一些实施方案中,变体Cas9蛋白可切割双链向导靶标核酸的非互补链,但切割向导靶标序列的互补链的能力降低。例如,变体Cas9蛋白可具有降低HNH结构域(RuvC/HNH/RuvC结构域基序)的功能的突变(氨基酸置换)。作为非限制性示例,在一些实施方案中,变体Cas9蛋白具有H840A(在氨基酸位置840处用组氨酸代替丙氨酸)突变,并且因此可切割向导靶标序列的非互补链但切割向导靶标序列的互补链的能力降低(因此当变体Cas9蛋白切割双链向导靶标序列时导致SSB而非DSB)。这种Cas9蛋白切割向导靶标序列(例如,单链向导靶标序列)的能力降低,但保持结合向导靶标序列(例如,单链向导靶标序列)的能力。

[0785] 在一些实施方案中,变体Cas9蛋白切割双链靶标DNA的互补链和非互补链两者的能力降低。作为非限制性示例,在一些实施方案中,变体Cas9蛋白携带D10A突变和H840A突变两者,使得该多肽切割双链靶标DNA的互补链和非互补链两者的能力降低。这种Cas9蛋白切割靶标DNA(例如,单链靶标DNA)的能力降低,但保持结合靶标DNA(例如,单链靶标DNA)的能力。

[0786] 作为另一非限制性示例,在一些实施方案中,变体Cas9蛋白携带W476A和W1126A突变,使得该多肽切割靶标DNA的能力降低。这种Cas9蛋白切割靶标DNA(例如,单链靶标DNA)的能力降低,但保持结合靶标DNA(例如,单链靶标DNA)的能力。

[0787] 作为另一非限制性示例,在一些实施方案中,变体Cas9蛋白携带P475A、W476A、N477A、D1125A、W1126A和D1127A突变,使得该多肽切割靶标DNA的能力降低。这种Cas9蛋白切割靶标DNA(例如,单链靶标DNA)的能力降低,但保持结合靶标DNA(例如,单链靶标DNA)的能力。

[0788] 作为另一非限制性示例,在一些实施方案中,变体Cas9蛋白携带H840A、W476A和W1126A突变,使得该多肽切割靶标DNA的能力降低。这种Cas9蛋白切割靶标DNA(例如,单链靶标DNA)的能力降低,但保持结合靶标DNA(例如,单链靶标DNA)的能力。作为另一非限制性示例,在一些实施方案中,变体Cas9蛋白携带H840A、D10A、W476A和W1126A突变,使得该多肽切割靶标DNA的能力降低。这种Cas9蛋白切割靶标DNA(例如,单链靶标DNA)的能力降低,但保持结合靶标DNA(例如,单链靶标DNA)的能力。在系实施方案中,变体Cas9在Cas9 HNH结构域中的位置840处具有恢复的催化His残基(A840H)。

[0789] 作为另一非限制性示例,在一些实施方案中,变体Cas9蛋白携带H840A、P475A、W476A、N477A、D1125A、W1126A和D1127A突变,使得该多肽切割靶标DNA的能力降低。这种Cas9蛋白切割靶标DNA(例如,单链靶标DNA)的能力降低,但保持结合靶标DNA(例如,单链靶标DNA)的能力。作为另一非限制性示例,在一些实施方案中,变体Cas9蛋白携带D10A、H840A、P475A、W476A、N477A、D1125A、W1126A和D1127A突变,使得该多肽切割靶标DNA的能力降低。这种Cas9蛋白切割靶标DNA(例如,单链靶标DNA)的能力降低,但保持结合靶标DNA(例如,单链靶标DNA)的能力。在一些实施方案中,当变体Cas9蛋白携带W476A和W1126A突变时

或当变体Cas9蛋白携带P475A、W476A、N477A、D1125A、W1126A和D1127A突变时,变体Cas9蛋白不有效地结合至PAM序列。因此,在一些此类实施方案中,当这种变体Cas9蛋白在结合方法中使用时,该方法不需要PAM序列。换言之,在一些实施方案中,当这种变体Cas9蛋白在结合方法中使用时,该方法可包括向导RNA,但该方法可在PAM序列不存在下执行(并且结合特异性因此由向导RNA的靶向链段提供)。其他残基可经突变以实现上述效果(即,将一个或其他核酸酶蛋白灭活)。作为非限制性示例,残基D10、G12、G17、E762、H840、N854、N863、H982、H983、A984、D986和/或A987可改变(即,置换)。而且,除丙氨酸置换之外的突变是合适的。

[0790] 在一些实施方案中,变体Cas9蛋白具有降低的催化活性(即,当Cas9蛋白具有D10、G12、G17、E762、H840、N854、N863、H982、H983、A984、D986和/或A987突变,例如,D10A、G12A、G17A、E762A、H840A、N854A、N863A、H982A、H983A、A984A和/或D986A时),该变体Cas9蛋白仍可以位点特异性的方式结合至靶标DNA(因为它仍然被向导RNA引导至靶标DNA序列),只要它保持与向导RNA相互作用的能力即可。

[0791] 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。

[0792] 在一些实施方案中,使用包括氨基酸置换D1135M、S1136Q、G1218K、E1219F、A1322R、D1332A、R1335E和T1337R且对于改变的PAM 5'-NGC-3'具有特异性的SpCas9(SpCas9-MQKFRAER)。

[0793] 酿脓链球菌Cas9的替代物可以包括来自Cpf1家族的RNA引导的核酸内切酶,其在哺乳动物细胞中表现出切割活性。来自普雷沃菌属(Prevotella)和弗朗西丝氏菌属(Francisella 1)(CRISPR/Cpf1)的CRISPR是类似于CRISPR/Cas9系统的DNA编辑技术。Cpf1是II类CRISPR/Cas系统的RNA引导的核酸内切酶。这种获得性免疫机制见于普雷沃菌属和弗朗西丝氏菌属细菌。Cpf1基因与CRISPR基因座相关,编码使用向导RNA来发现并切割病毒DNA的核酸内切酶。Cpf1是一种比Cas9更小且更简单的核酸内切酶,克服了CRISPR/Cas9系统的一些限制。不同于Cas9核酸酶,Cpf1介导的DNA切割的结果是具有短3'突出部分的双链断裂。Cpf1的交错切割模式可以产生定向基因转移的可能性,类似于传统的限制酶克隆,这可增加基因编辑的效率。类似于上述Cas9变体和同源物,Cpf1也可以扩展可以被CRISPR靶向至缺乏SpCas9偏好的NGG PAM位点的富含AT的区域或富含AT的基因组的位点数。Cpf1基因座含有混合的α/β结构域,RuvC-I,之后是螺旋区,RuvC-II和锌指样结构域。Cpf1蛋白具有类似于Cas9的RuvC结构域的RuvC样核酸内切酶结构域。

[0794] 此外,不同于Cas9,Cpf1没有HNH核酸内切酶结构域,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在功能上是独特的,被归类为2类V型CRISPR系统。Cpf1基因座编码Cas1、Cas2和Cas4蛋白,相较于II型系统,它们更类似于I型和III型系统。功能性Cpf1不需要反向激活CRISPR RNA(tracrRNA),因此仅需要CRISPR(crRNA)。这有利于基因组编辑,因为TCpf1不仅比Cas9小,而且它具有更小的sgRNA分子(核苷酸数是Cas9中核苷酸数的大约一半)。与被Cas9靶向的富含G的PAM相反,Cpf1-crRNA复合物通过鉴定前间区序列邻近基序5'-YTN-3'或5'-TTN-3'来切割靶标DNA或RNA。在鉴定PAM后,Cpf1引入具有4或5个核苷酸突出部分的粘端样DNA双链断裂。

[0795] 在一些实施方案中,Cas9是Cas9变体,该变体对于改变的PAM序列具有特异性。在一些实施方案中,额外的Cas9变体和PAM序列描述于Miller,S.M.,等人Continuous

evolution of SpCas9 variants compatible with non-G PAMs,Nat.Biotechnol.(2020)中其整体内容通过引用并入本文。在一些实施方案中,Cas9没有特异性PAM要求。在一些实施方案中,Cas9变体,例如,SpCas9变体,对于NRNH PAM具有特异性,其中R是A或G且H是A、C或T。在一些实施方案中,SpCas9变体具有对于PAM序列AAA、TAA、CAA、GAA、TAT、GAT或CAC的特异性。在一些实施方案中,SpCas9变体包含在位置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处或其对应位置处的氨基酸置换。在一些实施方案中,SpCas9变体包含在位置1114、1135、1218、1219、1221、1249、1320、1321、1323、1332、1333、1335或1337处或其对应位置处的氨基酸置换。在一些实施方案中,SpCas9变体包含在位置1114、1134、1135、1137、1139、1151、1180、1188、1211、1219、1221、1256、1264、1290、1318、1317、1320、1323、1333处或其对应位置处的氨基酸置换。在一些实施方案中,SpCas9变体包含在位置1114、1131、1135、1150、1156、1180、1191、1218、1219、1221、1227、1249、1253、1286、1293、1320、1321、1332、1335、1339处或其对应位置处的氨基酸置换。在一些实施方案中,SpCas9变体包含在位置1114、1127、1135、1180、1207、1219、1234、1286、1301、1332、1335、1337、1338、1349处或其对应位置处的氨基酸置换。SpCas9变体的示例性氨基酸置换和PAM特异性显示在表3A至表3D中。

[0796] 表3A SpCas9变体

[0797]

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

[0798]

表 3B

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

[0799]

表 3C

SpCas9	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	

[0800]

表 3D

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

[0801] 在一些实施方案中,Cas9是脑膜炎奈瑟氏菌(Neisseria meningitidis)Cas9(NmeCas9)或其变体。在一些实施方案中,NmeCas9对于NNNNGAYW PAM具有特异性,其中Y是C或T且W是A或T。在一些实施方案中,NmeCas9对于NNNNGYTT PAM具有特异性,其中Y是C或T。在一些实施方案中,NmeCas9对于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具有对于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的特异性。

在一些实施方案中,Nme1Cas9具有对于NNNNGATT PAM、NNNNCCTA PAM、NNNNCCTC PAM、NNNNCCTT PAM或NNNNCCTG PAM的特异性。在一些实施方案中,NmeCas9具有对于CAA PAM、CAAA PAM或CCA PAM的特异性。在一些实施方案中,NmeCas9是Nme2 Cas9。在一些实施方案中,NmeCas9具有对于NNNNCC(N4CC)PAM的特异性,其中N是A、G、C或T中的任一个。在一些实施方案中,NmeCas9具有对于NNNNCCGT PAM、NNNNCCGGPAM、NNNNCCCA PAM、NNNNCCCT PAM、NNNNCCCC PAM、NNNNCCAT PAM、NNNNCCAG PAM、NNNNCCAT PAM或NNNGATT PAM的特异性。在一些实施方案中,NmeCas9是Nme3Cas9。在一些实施方案中,NmeCas9具有对于NNNNCAAA PAM、NNNNCC PAM或NNNNCNNN PAM的特异性。如Edraki等人Mol.Cell.(2019)73(4):714-726中所述的其他NmeCas9特征和PAM序列通过引用以其整体并入本文。

[0802] 示例性脑膜炎奈瑟氏菌(Neisseria meningitidis)Cas9蛋白,Nme1Cas9,(NCBI参考:WP_002235162.1;II型CRISPR RNA引导的核酸内切酶Cas9)具有以下氨基酸序列:

[0803]
```
   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
```

[0804] 示例性脑膜炎奈瑟氏菌(Neisseria meningitidis)Cas9蛋白,Nme2Cas9,(NCBI参考:WP_002230835;II型CRISPR RNA引导的核酸内切酶Cas9)具有以下氨基酸序列:

[0805]
```
   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
```

[0806] 核碱基编辑器的Cas12结构域

[0807] 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子复合物,而2类系统具有单一蛋白质效应子。例如,Cas9和Cpf1是2类效应子,但属于不同类型(分别为II型和V型)。除了Cpf1以外,2类V型CRISPR-Cas系统也包括Cas12a/Cpf1、Cas12b/C2c1、Cas12c/C2c3、Cas12d/CasY、Cas12e/CasX、Cas12g、Cas12h、Cas12i和Cas12j/CasΦ。参见,例如,Shmakov等人,"Discovery and Functional Characterization of Diverse Class 2CRISPR Cas Systems,"Mol.Cell,2015Nov.5;60(3):385-397;Makarova等人,"Classification and Nomenclature of CRISPR-Cas Systems:Where from Here?"CRISPR Journal,2018,1(5):325-336;和Yan等人,"Functionally Diverse Type V CRISPR-Cas Systems,"Science,2019Jan.4;363:88-91;各自的整体内容通过引用并入本文。V型Cas蛋白含有RuvC(或RuvC样)核酸内切酶结构域。尽管成熟CRISPR RNA(crRNA)的生产通常不依赖于tracrRNA,但Cas12b/C2c1例如需要tracrRNA来生产crRNA。Cas12b/C2c1依据crRNA和tracrRNA两者进行DNA切割。

[0808] 本发明设想的核酸可编程DNA结合蛋白包括归类为2类V型的Cas蛋白(Cas12蛋白)。2类V型Cas蛋白的示例包括Cas12a/Cpf1、Cas12b/C2c1、Cas12c/C2c3、Cas12d/CasY、Cas12e/CasX、Cas12g、Cas12h、Cas12i和Cas12j/CasΦ,其同源物或其经修饰的版本。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蛋白包含被内部融合蛋白结构域诸如脱氨酶结构域中断的氨基酸序列。

[0809] 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的结构域。例如,Cas12结构域可以是将双螺旋核酸(例如,双螺旋DNA分子)的一条链切口的Cas12结构域。在一些实施方案中,Cas12结构域包含本文中详述的氨基酸序列中的任一者。在一些实施方案中,Cas12结构域包含氨基酸序列,该氨基酸序列与本文详述的氨基酸序列中的任一者至少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结构域包含氨基酸序列,与本文详述的氨基酸序列中的任一者相比,该氨基酸序列具有1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、21、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50个或更多个突变。在一些实施方案中,Cas12结构域包含氨基酸序列,与本文详述的氨基酸序列中的任一者相比,该氨基酸序列具有至少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。

[0810] 在一些实施方案中,提供包含Cas12的片段的蛋白质。例如,在一些实施方案中,蛋白质包含一个或两个Cas12结构域:(1)Cas12的gRNA结合结构域;或(2)Cas12的DNA切割结构域。在一些实施方案中,包含Cas12或其片段的蛋白质称为"Cas12变体"。Cas12变体与Cas12或其片段共享同源性。例如,Cas12变体与野生型Cas12至少约70%相同、至少约80%相同、至少约90%相同、至少约95%相同、至少约96%相同、至少约97%相同、至少约98%相同、至少约99%相同、至少约99.5%相同、或至少约99.9%相同。在一些实施方案中,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、45、46、47、48、49、50或更多个氨基酸变化。在一些实施方案中,Cas12变体包含Cas12的片段(例如,gRNA结合结构域或DNA切割结构域),使得该片段与野生型Cas12的对应片段至少约70%相同、至少约80%相同、至少约90%相同、至少约95%相同、至少约96%相同、至少约97%相同、至少约98%相同、至少约99%相同、至少约99.5%相同、或至少约99.9%相同。在一些实施方案中,该片段是对应的野生型Cas12的氨基酸长度的至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%相同、至少96%、至少97%、至少98%、至少99%、或至少99.5%。在一些实施方案中,该片段是至少100个氨基酸的长度。在一些实施方案中,该片段是至少100、150、200、250、300、350、400、450、500、550、600、650、700、750、800、850、900、950、1000、1050、1100、1150、1200、1250、或至少1300个氨基酸的长度。

[0811] 在一些实施方案中,Cas12对应于或者部分或整体地包含具有一个或多个突变的Cas12氨基酸序列,该突变将Cas12核酸酶活性灭活。此类突变,举例而言,包括在Cas12的RuvC核酸酶结构域内的氨基酸置换。在一些实施方案中,提供Cas12的变体或同源物,它们与野生型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,提供具有氨基酸序列的Cas12变体,该氨基酸序列更短或更长,相异约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。

[0812] 在一些实施方案中,如本文所提供的Cas12融合蛋白包含Cas12蛋白(例如,本文所提供的Cas12序列中的一者)的全长氨基酸序列。然而,在其他实施方案中,本文所提供的融合蛋白不包含全长Cas12序列,而仅包含其一个或多个片段。合适的Cas12结构域的示例性氨基酸序列提供在本文中,且Cas12结构域和片段的其他合适的序列对于本领域技术人员将是显而易见的。

[0813] 通常,2类V型Cas蛋白具有单一功能性RuvC核酸内切酶结构域(参见,例如,Chen等人,"CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity,"Science360:436-439(2018))。在一些情况下,Cas12蛋白是变体Cas12b蛋白。(参见Strecker等人,Nature Communications,2019,10(1):Art.No.:212)。在一种实施方案中,变体Cas12多肽具有氨基酸序列,该氨基酸序列较之于野生型Cas12蛋白的氨基酸序列相异1、2、3、4、5或更多个氨基酸(例如,具有缺失、插入、置换、融合)。在一些实例中,变体Cas12多肽具有氨基酸变化(例如,缺失、插入或置换),该变化降低Cas12多肽的活性。例如,在一些实例中,变体Cas12是Cas12b多肽,其具有低于50%、低于40%、低于30%、低于20%、低于10%、低于5%、或低于1%的对应野生型Cas12b蛋白的切口酶活性。在一些情况下,变体Cas12b蛋白基本上没有切口酶活性。

[0814] 在一些情况下,变体Cas12b蛋白具有降低的切口酶活性。例如,变体Cas12b蛋白表现出低于约20%、低于约15%、低于约10%、低于约5%、低于约1%、或低于约0.1%的野生型Cas12b蛋白的切口酶活性。

[0815] 在一些实施方案中,Cas12蛋白包括来自在哺乳动物细胞内显示活性的Cas12a/Cpf1家族的RNA引导的核酸内切酶。来自普雷沃菌属(Prevotella)和弗朗西丝氏菌属

(Francisella 1)(CRISPR/Cpf1)的CRISPR是类似于CRISPR/Cas9系统的DNA编辑技术。Cpf1是II类CRISPR/Cas系统的RNA引导的核酸内切酶。这种获得性免疫机制见于普雷沃菌属和弗朗西丝氏菌属细菌。Cpf1基因与CRISPR基因座相关,编码使用向导RNA来发现并切割病毒DNA的核酸内切酶。Cpf1是一种比Cas9更小且更简单的核酸内切酶,克服了CRISPR/Cas9系统的一些限制。不同于Cas9核酸酶,Cpf1介导的DNA切割的结果是具有短3'突出部分的双链断裂。Cpf1的交错切割模式可以产生定向基因转移的可能性,类似于传统的限制酶克隆,这可增加基因编辑的效率。类似于上述Cas9变体和同源物,Cpf1也可以扩展可以被CRISPR靶向至缺乏SpCas9偏好的NGG PAM位点的富含AT的区域或富含AT的基因组的位点数。Cpf1基因座含有混合的α/β结构域,RuvC-I,之后是螺旋区,RuvC-II和锌指样结构域。Cpf1蛋白具有类似于Cas9的RuvC结构域的RuvC样核酸内切酶结构域。此外,不同于Cas9,Cpf1没有HNH核酸内切酶结构域,且Cpf1的N端没有Cas9的α-螺旋识别叶。Cpf1CRISPR-Cas结构域架构显示,Cpf1在功能上是独特的,被归类为2类V型CRISPR系统。Cpf1基因座编码Cas1、Cas2和Cas4蛋白,相较于II型系统,它们更类似于I型和III型系统。功能性Cpf1不需要反向激活CRISPR RNA(tracrRNA),因此仅需要CRISPR(crRNA)。这有利于基因组编辑,因为TCpf1不仅比Cas9小,而且它具有更小的sgRNA分子(核苷酸数是Cas9中核苷酸数的大约一半)。与被Cas9靶向的富含G的PAM相反,Cpf1-crRNA复合物通过鉴定前间区序列邻近基序5'-YTN-3'或5'-TTTN-3'来切割靶标DNA或RNA。在鉴定PAM后,Cpf1引入具有4或5个核苷酸突出部分的粘端样DNA双链断裂。

[0816] 在本发明的一些方面,可使用以下载体,该载体编码相对于对应的野生型酶突变的CRISPR酶,使得突变的CRISPR酶缺乏切割含有靶标序列的靶标多核苷酸的一条或两条链的能力。Cas12可以指一种多肽,其与野生型示例性Cas12多肽(例如,来自外村尚芽孢杆菌的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%的序列一致性和/或序列同源性。Cas12可以指一种多肽,其与野生型示例性Cas12多肽(例如,来自外村尚芽孢杆菌(BhCas12b)、芽孢杆菌属V3-13(BvCas12b)和嗜酸性脂环杆菌(Alicyclobacillus acidiphilus)(AaCas12b))具有至多或至多约50%、60%、70%、80%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%的序列一致性和/或序列同源性。Cas12可以指野生型或经修饰形式的Cas9蛋白,该经修饰形式的Cas12蛋白可以包含氨基酸变化诸如缺失、插入、置换、变体、突变、融合、嵌合或其任意组合。

[0817] 在一些实施方案中,BhCas12b向导多核苷酸具有以下序列:

[0818] BhCas12b sgRNA支架(加下划线)+20nt至23nt向导序列(由N_n表示)

[0819] 5'GUUCUGTCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUGCAGGGUGUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUACGAGGCAUUAGCACNNNNNNNNNNNNNNNNNNNN-3'

[0820] 在一些实施方案中,BvCas12b和AaCas12b向导多核苷酸具有以下序列:

[0821] BvCas12b sgRNA支架(加下划线)+20nt至23nt向导序列(由N_n表示)

[0822] 5'GACCUAUAGGGUCAAUGAAUCUGUGCGUGUGCCAUAAGUAAUUAAAAAUUACCCACCACAGGAGCACCUGAAAACAGGUGCUUGGCACNNNNNNNNNNNNNNNNNNNN-3'

[0823] AaCas12b sgRNA支架(加下划线)+20nt至23nt向导序列(由N_n表示)

[0824] 5'GUCUAAAGGACAGAAUUUUUCAACGGGUGUGCCAAUGGCCACUUUCCAGGUGGCAAAGCCCGUUGA

ACUUCUCAAAAAGAACGAUCUGAGAAGUGGCACNNNNNNNNNNNNNNNNNNNN-3’

[0825] 核酸可编程DNA结合蛋白

[0826] 本公开的一些方面提供包含充当核酸可编程DNA结合蛋白的结构域的融合蛋白，其可用于将蛋白质诸如碱基编辑器引导至特异性核酸(例如,DNA或RNA)序列。在特定实施方案中,融合蛋白包含核酸可编程DNA结合蛋白结构域和脱氨酶结构域。核酸可编程DNA结合蛋白的非限制性示例包括Cas9(例如,dCas9和nCas9)、Cas12a/Cpf1、Cas12b/C2c1、Cas12c/C2c3、Cas12d/CasY、Cas12e/CasX、Cas12g、Cas12h、Cas12i和Cas12j/CasΦ。Cas酶的非限制性示例包括Cas1、Cas1B、Cas2、Cas3、Cas4、Cas5、Cas5d、Cas5t、Cas5h、Cas5a、Cas6、Cas7、Cas8、Cas8a、Cas8b、Cas8c、Cas9(也称为Csn1或Csx12)、Cas10、Cas10d、Cas12a/Cpf1、Cas12b/C2c1、Cas12c/C2c3、Cas12d/CasY、Cas12e/CasX、Cas12g、Cas12h、Cas12i、Cas12j/CasΦ、Csy1、Csy2、Csy3、Csy4、Cse1、Cse2、Cse3、Cse4、Cse5e、Csc1、Csc2、Csa5、Csn1、Csn2、Csm1、Csm2、Csm3、Csm4、Csm5、Csm6、Cmr1、Cmr3、Cmr4、Cmr5、Cmr6、Csb1、Csb2、Csb3、Csx17、Csx14、Csx10、Csx16、CsaX、Csx3、Csx1、Csx1S、Csx11、Csf1、Csf2、Cs0、Csf4、Csd1、Csd2、Cst1、Cst2、Csh1、Csh2、Csa1、Csa2、Csa3、Csa4、Csa5、II型Cas效应蛋白、V型Cas效应蛋白、VI型Cas效应蛋白、CARF、DinG、其同源、或其经修饰或工程改造的版本。其他核酸可编程DNA结合蛋白也处于本公开范畴内,但它们可以未在本公开中具体列出。参见,例如,Makarova等人“Classification and Nomenclature of CRISPR-Cas Systems:Where from Here?”CRISPR J.2018Oct;1:325-336.doi:10.1089/crispr.2018.0033;Yan等人,“Functionally diverse type V CRISPR-Cas systems”Science.2019Jan 4;363(6422):88-91.doi:10.1126/science.aav7271,各自的整体内容通过引用并入本文。

[0827] 具有不同于Cas9的PAM特异性的核酸可编程DNA结合蛋白的一个示例是来自普雷沃菌属和弗朗西丝氏菌属1(Cpf1)的成簇的规则间隔短回文重复序列。类似于Cas9,Cpf1也是2类CRISPR效应子。已经证明,Cpf1以不同于Cas9的特征来介导稳健的DNA干扰。Cpf1是缺乏tracrRNA的单一RNA引导的核酸内切酶,并且它利用富含T的前间区序列邻近基序(TTN、TTN或YTN)。此外,Cpf1经由交错的DNA双链断裂来切割DNA。16种Cpf1家族蛋白质中,来自氨基酸球菌属(Acidaminococcus)和毛螺菌科(Lachnospiraceae)的两种酶显示在人类细胞中具有有效的基因组编辑活性。Cpf1蛋白是本领域中已知的并且先前已经有所描述,例如Yamano等人,“Crystal structure of Cpflin complex with guide RNA and target DNA.”Cell(165)2016,p.949-962;其整体内容通过引用并入本文。

[0828] 可用于本发明组合物和方法中的是无核酸酶活性的Cpf1(dCpf1)变体,其可用作向导核苷酸序列可编程DNA结合蛋白结构域。Cpf1蛋白具有类似于Cas9的RuvC结构域的RuvC样核酸内切酶结构域但没有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。Zetsche等人,Cell,163,759-771,2015(其通过引用并入本文)中显示,Cpf1的RuvC样结构域负责切割两条DNA链,并且RuvC样结构域的灭活将Cpf1核酸酶活性灭活。例如,对应于新凶手弗朗西丝氏菌(Francisella novicida)Cpf1中的D917A、E1006A或D1255A的突变将Cpf1核酸酶活性灭活。在一些实施方案中,本公开的dCpf1包含对应于D917A、E1006A、D1255A、D917A/E1006A、D917A/D1255A、E1006A/D1255A或D917A/E1006A/D1255A的突变。应理解,可根据本公开使用任何将Cpf1的RuvC结构域灭活的突变,例如,置换突变、缺失或插入。

[0829] 在一些实施方案中,本文所提供的任何融合蛋白的核酸可编程DNA结合蛋白(napDNAbp)可以是Cpf1蛋白。在一些实施方案中,Cpf1蛋白是Cpf1切口酶(nCpf1)。在一些实施方案中,Cpf1蛋白是无核酸酶活性的Cpf1(dCpf1)。在一些实施方案中,Cpf1、nCpf1或dCpf1包含氨基酸序列,该氨基酸序列与本文所公开的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包含氨基酸序列,该氨基酸序列与本文所公开的Cpf1至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%、或至少99.5%相同;并且包含对应于D917A、E1006A、D1255A、D917A/E1006A、D917A/D1255A、E1006A/D1255A或D917A/E1006A/D1255A的突变。应了解,来自其他细菌物种的Cpf1也可根据本公开使用。

[0830] 野生型新凶手弗朗西丝氏菌(Francisella novicida)Cpf1(D917、E1006和D1255加粗并加下划线)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQ
FFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEK
FKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTT
YFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKK
DLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENT
KRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQS

[0831] FYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVI
GTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQ
CRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLD
QTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKP
YSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKEN
KGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYE
KFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFE
NISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNG
EAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPIT
INFKSSGANKFNDEINLLLKEKANDVHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFNII
GNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIV

[0832] VF**E**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFE
TFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDK
GYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKD
YSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGN
FFDSRQAPKNMPQDA**D**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNR
NN

[0833] 新凶手弗朗西丝氏菌(Francisella novicida)Cpf1 D917A(A917、E1006和D1255加粗并加下划线)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQ
FFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEK
FKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTT
YFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKK
DLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENT
KRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQS
FYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVI
GTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQ
CRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLD
QTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKP

[0834] YSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKEN
KGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYE
KFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFE
NISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNG
EAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPIT
INFKSSGANKFNDEINLLLKEKANDVHILSI**A**RGERHLAYYTLVDGKGNIIKQDTFNII
GNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIV
VF**E**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFE
TFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDK
GYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKD
YSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGN
FFDSRQAPKNMPQDA**D**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNR

[0835] NN

[0836] 新凶手弗朗西丝氏菌(Francisella novicida)Cpf1 E1006A(D917、A1006和D1255加粗并加下划线)

[0837]

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEE
ILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAK
KGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTS
IIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFE
IANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDT
ESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSL
TDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKH
RDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQT
NNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNF
ENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANK
MLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEW
KDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRP
NLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLI
KDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSI**D**RGERHLAYYTLVDGKGNI
IKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAI
VVF**A**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGK
QTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDK
AAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKF
FAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDADANGAYHIGLKGLMLL
GRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0838] 新凶手弗朗西丝氏菌(Francisella novicida)Cpf1 D1255A(D917、E1006和A1255加粗并加下划线)

[0839]

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQ
FFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEK
FKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTT
YFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKK
DLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENT
KRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQS
FYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVI
GTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQ
CRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLD
QTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKP
YSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKEN
KGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYE
KFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFE
NISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNG

EAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPIT
INFKSSGANKFNDEINLLLKEKANDVHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFNII
GNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIV
VF**E**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFE

[0840] TFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDK
GYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKD
YSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGN
FFDSRQAPKNMPQDA**A**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNR
NN

[0841] 新凶手弗朗西丝氏菌(Francisella novicida)Cpf1 D917A/E1006A(A917、A1006和D1255加粗并加下划线)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQ
FFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEK
FKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTT
YFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKK
DLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENT
KRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQS
FYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVI
GTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQ
CRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLD
QTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKP
YSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKEN

[0842] KGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYE
KFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFE
NISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNG
EAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPIT
INFKSSGANKFNDEINLLLKEKANDVHILSI**A**RGERHLAYYTLVDGKGNIIKQDTFNII
GNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIV
VF**A**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFE
TFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDK
GYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKD
YSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGN
FFDSRQAPKNMPQDA**D**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNR
NN

[0843] 新凶手弗朗西丝氏菌(Francisella novicida)Cpf1 D917A/D1255A(A917、E1006和A1255加粗并加下划线)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQ
FFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEK
FKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTT
YFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKK
DLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENT
KRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQS
FYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVI
GTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQ
CRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLD
QTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKP
YSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKEN

[0844] KGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYE
KFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFE
NISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNG
EAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPIT
INFKSSGANKFNDEINLLLKEKANDVHILSI**A**RGERHLAYYTLVDGKGNIIKQDTFNII
GNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIV
VF**E**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFE
TFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDK
GYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKD
YSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGN
FFDSRQAPKNMPQDA**A**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNR
NN

[0845] 新凶手弗朗西丝氏菌(Francisella novicida)Cpf1E1006A/D1255A(D917、A1006和A1255加粗并加下划线)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQ
FFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEK
FKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTT

[0846] YFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKK
DLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENT
KRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQS
FYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVI

[0847]

GTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQ
CRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLD
QTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKP
YSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKEN
KGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYE
KFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFE
NISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNG
EAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPIT
INFKSSGANKFNDEINLLLKEKANDVHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFNII
GNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIV
VF**A**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFE
TFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDK
GYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKD
YSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGN
FFDSRQAPKNMPQDA**A**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNR
NN

[0848] 新凶手弗朗西丝氏菌(Francisella novicida)Cpf1D917A/E1006A/D1255A(A917、A1006和A1255加粗并加下划线)

[0849]

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEE
ILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAK
KGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTS
IIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFE
IANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDT
ESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSL
TDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKH
RDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQT
NNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNF
ENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANK
MLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEW
KDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRP
NLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLI
KDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSI**A**RGERHLAYYTLVDGKGNI
IKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAI
VVF**A**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGK
QTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDK
AAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKF
FAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**A**ANGAYHIGLKGLMLL
GRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0850] 在一些实施方案中,融合蛋白中存在的Cas9结构域中的一者可以替换为不需要PAM序列的向导核苷酸序列可编程DNA结合蛋白结构域。

[0851] 在一些实施方案中,Cas9结构域是来自金黄色葡萄球菌的Cas9结构域(SaCas9)。在一些实施方案中,SaCas9结构域是核酸酶活性的SaCas9、无核酸酶活性的SaCas9(SaCas9d)、或SaCas9切口酶(SaCas9n)。在一些实施方案中,SaCas9结构域包含N579A突变,或本文所提供的任何氨基酸序列中的对应突变。

[0852] 在一些实施方案中,SaCas9结构域、SaCas9d结构域或SaCas9n结构域可结合至具有非标准PAM的核酸序列。在一些实施方案中,SaCas9结构域、SaCas9d结构域或SaCas9n结构域可结合至具有NNGRRT或NNGRRT PAM序列的核酸序列。在一些实施方案中,SaCas9结构域包含E781X、N967X和R1014X突变中的一者或多者,或本文所提供的任何氨基酸序列中的对应突变,其中X是任何氨基酸。在一些实施方案中,SaCas9结构域包含E781K、N967K和R1014H突变中的一者或多者,或本文所提供的任何氨基酸序列中的一个或多个对应突变。在一些实施方案中,SaCas9结构域包含E781K、N967K或R1014H突变,或本文所提供的任何氨基酸序列中的对应突变。

[0853] 示例性SaCas9序列

[0854]
KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQ
RVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNEL
STKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFID
TYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLV
ITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIK
DITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAI
NLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKK
YGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKC
LYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEE**N**SKKGNRTPFQYLSSSDSKISY
ETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNN
LDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEE
KQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIV
NNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYS
KKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKK
ENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYRE
YLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG

[0855] 上述加下划线并加粗的残基N579可以经突变(例如,至A579)以获得SaCas9切口酶。

[0856] 示例性SaCas9n序列

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRR
RHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVH
NVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVK
EAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHC
TYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTL
KQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKIL
TIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQ
IAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPND
IIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEG

[0857]

KCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEE**A**SKKGNRTPFQY
LSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTR
YATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIA
NADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKD
YKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKL
LMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNK
LNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSK
CYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREY
LENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG

[0858] 上述残基A579,其可以从N579突变以获得SaCas9切口酶,加下划线并加粗。

[0859] 示例性SaKKH Cas9

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRR
RHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVH
NVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVK
EAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHC
TYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTL
KQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKIL

[0860]

TIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQ
IAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPND
IIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEG
KCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEE**A**SKKGNRTPFQY
LSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTR
YATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIA
NADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKD

YKYSHRVDKKPNR*K*LINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKL

LMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNK

[0861] LNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSK

CYEEAKKLKKISNQAEFIASFY*K*NDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREY

LENMNDKRPP*H*IIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG。

[0862] 上述残基A579,其可以从N579突变以获得SaCas9切口酶,加下划线并加粗。上述残基K781、K967和H1014,其可以从E781、N967和R1014突变以获得SaKKH Cas9,加下划线并以斜体显示。

[0863] 在一些实施方案中,napDNAbp是环状排列的。在以下序列中,纯文本表示腺苷脱氨酶结构域,粗体序列指示源自Cas9的序列,斜体序列表示连接子序列,加下划线序列表示二分体核定位序列,且加双下划线序列指示突变。星号(*)表示STOP密码子。

[0864] CP5(其中MSP“NGC=具有常规Cas9 Pam如NGG的突变的变体”PID=蛋白质相互作用结构域,且“D10A”切口酶):

EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKV

LSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPTVAYSVLV

VAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLF

ELENGRKRMLASAKFLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHK

HYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPR

AFKYFDTTIARKEYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD*GGSGGSGGSGGSG*

*GSGGSGGM***DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALL**

FDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDK

KHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLI

[0865] **EGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQL**

PGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYAD

LFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY

KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFD

NGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWM

TRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNEL

TKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGV

EDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHL

FDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDS

LTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIV

IEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNG

RDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK

[0866] **MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDS**

RMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTA

LIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEGADKRTADGSEFESPKKKRKV*

[0867] 在一些实施方案中,核酸可编程DNA结合蛋白(napDNAbp)是微生物CRISPR-Cas系统的单一效应子。微生物CRISPR-Cas系统的单一效应子包括而不限于,Cas9、Cpf1、Cas12b/C2c1和Cas12c/C2c3。通常,微生物CRISPR-Cas系统分为1类和2类系统。1类系统具有多亚基效应子复合物,而2类系统具有单一蛋白质效应子。例如,Cas9和Cpf1是2类效应子。除了Cas9和Cpf1之外,三种不同的2类CRISPR-Cas系统(Cas12b/C2c1和Cas12c/C2c3)已经由Shmakov等人,"Discovery and Functional Characterization of Diverse Class 2CRISPR Cas Systems",Mol.Cell,2015Nov.5;60(3):385-397描述,其整体内容通过引用并入本文。该系统中两者的效应子,Cas12b/C2c1和Cas12c/C2c3,含有关于Cpf1的RuvC样核酸内切酶结构域。第三种系统含有具有两个阐明的HEPN RNase结构域的效应子。不同于通过Cas12b/C2c1生产CRISPR RNA,成熟CRISPR RNA的产生时tracrRNA依赖性的。Cas12b/C2c1依据CRISPR RNA和tracrRNA两者进行DNA切割。

[0868] 酸土脂环酸芽孢杆菌(Alicyclobaccillus acidoterrastris)Cas12b/C2c1(AacC2c1)的晶体结构已经在与嵌合单分子向导RNA(sgRNA)的复合物中报告。参见,例如,Liu等人,"C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism",Mol.Cell,2017Jan.19;65(2):310-322,其整体内容通过引用并入本文。晶体结构也在结合至靶标DNA的酸土脂环酸芽孢杆菌C2c1中作为三元复合物报告。参见,例如,Yang等人,"PAM-dependent Target DNA Recognition and Cleavage by C2C1 CRISPR-Cas endonuclease",Cell,2016Dec.15;167(7):1814-1828,其整体内容通过引用并入本文。AacC2c1的催化感受态构象,其中靶标和非靶标两条DNA链已经独立地捕获定位在单一RuvC催化袋内,且Cas12b/C2c1介导的切割导致靶标DNA的交错七核苷酸断裂。Cas12b/C2c1三元复合物与先前鉴定的Cas9和Cpf1对应物之间的结构比较证明了CRISPR-Cas9系统所使用的机制的多样性。

[0869] 在一些实施方案中,本文所提供的任何融合蛋白的核酸可编程DNA结合蛋白(napDNAbp)可以是Cas12b/C2c1或Cas12c/C2c3蛋白。在一些实施方案中,napDNAbp是Cas12b/C2c1蛋白。在一些实施方案中,napDNAbp是Cas12c/C2c3蛋白。在一些实施方案中,napDNAbp包含氨基酸序列,该氨基酸序列与天然存在的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包含氨基酸序列,该氨基酸序列与本文所提供的napDNAbp序列中的任一者至少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。应了解,来自其他细菌物种的Cas12b/C2c1和Cas12c/C2c3也可根据本公开使用。

[0870] Cas12b/C2c1((uniprot.org/uniprot/T0D7A2#2)

[0871] sp|T0D7A2|C2C1_ALIAG CRISPR相关核酸内切酶C2c1 OS＝酸土脂环酸芽孢杆菌(菌株ATCC 49025/DSM 3922/CIP 106132/NCIMB

[0872] 13137/GD3B)GN＝c2c1 PE＝1SV＝1)氨基酸序列如下：

[0873] MAVKSIKVKLRLDDMPEIRAGLWKLHKEVNAGVRYYTEWLSLLRQENLYRRSPNGDGEQECDKTAEECK
AELLERLRARQVENGHRGPAGSDDELLQLARQLYELLVPQAIGAKGDAQQIARKFLSPLADKDAVGGLGIAKAGNKP
RWVRMREAGEPGWEEEKEKAETRKSADRTADVLRALADFGLKPLMRVYTDSEMSSVEWKPLRKGQAVRTWDRDMFQQ
AIERMMSWESWNQRVGQEYAKLVEQKNRFEQKNFVGQEHLVHLVNQLQQDMKEASPGLESKEQTAHYVTGRALRGSD
KVFEKWGKLAPDAPFDLYDAEIKNVQRRNTRRFGSHDLFAKLAEPEYQALWREDASFLTRYAVYNSILRKLNHAKMF
ATFTLPDATAHPIWTRFDKLGGNLHQYTFLFNEFGERRHAIRFHKLLKVENGVAREVDDVTVPISMSEQLDNLLPRD
PNEPIALYFRDYGAEQHFTGEFGGAKIQCRRDQLAHMHRRRGARDVYLNVSVRVQSQSEARGERRPPYAAVFRLVGD
NHRAFVHFDKLSDYLAEHPDDGKLGSEGLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNSKGRVPFFFPIKGND
NLVAVHERSQLLKLPGETESKDLRAIREERQRTLRQLRTQLAYLRLLVRCGSEDVGRRERSWAKLIEQPVDAANHMT
PDWREAFENELQKLKSLHGICSDKEWMDAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYAKDVVGGNSIEQIE
YLERQYKFLKSWSFFGKVSGQVIRAEKGSRFAITLREHIDHAKEDRLKKLADRIIMEALGYVYALDERGKGKWVAKY
PPCQLILLEELSEYQFNNDRPPSENNQLMQWSHRGVFQELINQAQVHDLLVGTMYAAFSSRFDARTGAPGIRCRRVP
ARCTQEHNPEPFPWWLNKFVVEHTLDACPLRADDLIPTGEGEIFVSPFSAEEGDFHQIHADLNAAQNLQQRLWSDFD
ISQIRLRCDWGEVDGELVLIPRLTGKRTADSYSNKVFYTNTGVTYYERERGKKRRKVFAQEKLSEEEAELLVEADEA
REKSVVLMRDPSGIINRGNWTRQKEFWSMVNQRIEGYLVKQIRSRVPLQDSACENTGDI

[0874] AacCas12b(嗜热脂环酸芽胞杆菌(Alicyclobacillus acidiphilus))-WP_067623834

[0875] MAVKSMKVKLRLDNMPEIRAGLWKLHTEVNAGVRYYTEWLSLLRQENLYRRSPNGDGEQECYKTAEECK
AELLERLRARQVENGHCGPAGSDDELLQLARQLYELLVPQAIGAKGDAQQIARKFLSPLADKDAVGGLGIAKAGNKP
RWVRMREAGEPGWEEEKAKAEARKSTDRTADVLRALADFGLKPLMRVYTDSDMSSVQWKPLRKGQAVRTWDRDMFQQ
AIERMMSWESWNQRVGEAYAKLVEQKSRFEQKNFVGQEHLVQLVNQLQQDMKEASHGLESKEQTAHYLTGRALRGSD
KVFEKWEKLDPDAPFDLYDTEIKNVQRRNTRRFGSHDLFAKLAEPKYQALWREDASFLTRYAVYNSIVRKLNHAKMF
ATFTLPDATAHPIWTRFDKLGGNLHQYTFLFNEFGEGRHAIRFQKLLTVEDGVAKEVDDVTVPISMSAQLDDLLPRD
PHELVALYFQDYGAEQHLAGEFGGAKIQYRRDQLNHLHARRGARDVYLNLSVRVQSQSEARGERRPPYAAVFRLVGD
NHRAFVHFDKLSDYLAEHPDDGKLGSEGLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNSEGRVPFCFPIEGNE
NLVAVHERSQLLKLPGETESKDLRAIREERQRTLRQLRTQLAYLRLLVRCGSEDVGRRERSWAKLIEQPMDANQMTP
DWREAFEDELQKLKSLYGICGDREWTEAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYQKDVVGGNSIEQIEY
LERQYKFLKSWSFFGKVSGQVIRAEKGSRFAITLREHIDHAKEDRLKKLADRIIMEALGYVYALDDERGKGKWVAKY
PPCQLILLEELSEYQFNNDRPPSENNQLMQWSHRGVFQELLNQAQVHDLLVGTMYAAFSSRFDARTGAPGIRCRRVP
ARCAREQNPEPFPWWLNKFVAEHKLDGCPLRADDLIPTGEGEFFVSPFSAEEGDFHQIHADLNAAQNLQRRLWSDFD
ISQIRLRCDWGEVDGEPVLIPRTTGKRTADSYGNKVFYTKTGVTYYERERGKKRRKVFAQEELSEEEAELLVEADEA
REKSVVLMRDPSGIINRGDWTRQKEFWSMVNQRIEGYLVKQIRSRVRLQESACENTGDI

[0876] BhCas12b(外村尚芽孢杆菌(Bacillus hisashii))NCBI参考序列:WP_095142515
MAPKKKRKVGIHGVPAAATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYEHHEQDPKNPKKVS
KAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVEKKGEANQLSNKFLYPLVDPNSQSGKGTASS
GRKPRWYNLKIAGDPSWEEEKKKWEEDKKKDPLAKILGKLAEYGLIPLFIPYTDSNEPIVKEIKWMEKSRNQSVRRL

DKDMFIQALERFLSWESWNLKVKEEYEKVEKEYKTLEERIKEDIQALKALEQYEKERQEQLLRDTLNTNEYRLSKRG
LRGWREIIQKWLKMDENEPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYATFCEIDKKKK
DAKQQATFTLADPINHPLWVRFEERSGSNLNKYRILTEQLHTEKLKKKLTVQLDRLIYPTESGGWEEKGKVDIVLLP
SRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGTLGGARVQFDRDHLRRYPHKVESGNVGRIYFNMTVNIEPTESPV
SKSLKIHRDDFPKVVNFKPKELTEWIKDSKGKKLKSGIESLEIGLRVMSIDLGQRQAAAASIFEVVDQKPDIEGKLF
FPIKGTELYAVHRASFNIKLPGETLVKSREVLRKAREDNLKLMNQKLNFLRNVLHFQQFEDITEREKRVTKWISRQE
NSDVPLVYQDELIQIRELMYKPYKDWVAFLKQLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDEIDRTRKFL
LRWSLRPTEPGEVRRLEPGQRFAIDQLNHLNALKEDRLKKMANTIIMHALGYCYDVRKKKWQAKNPACQIILFEDLS
NYNPYEERSRFENSKLMKWSRREIPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKLQDNRFFKN
LQREGRLTLDKIAVLKEGDLYPDKGGEKFISLSKDRKCVTTHADINAAQNLQKRFWTRTHGFYKVYCKAYQVDGQTV
YIPESKDQKQKIIEEFGEGYFILKDGVYEWVNAGKLKIKKGSSKQSSSELVDSDILKDSFDLASELKGEKLMLYRDP
SGNVFPSDKWMAAGVFFGKLERILISKLTNQYSISTIEDDSSKQSMKRPAATKKAGQAKKKK

[0877] 在一些实施方案中,Cas12b是BvCas12b(V4),其是BhCas12b的变体并且包含以下相对于BhCas12b的变化:S893R、K846R和E837G。BhCas12b(V4)表示如下:5'mRNA Cap---5'UTR---bhCas12b---STOP序列---3'UTR---120polyA尾部。

[0878] 5'UTR:GGGAAATAAGAGAGAAAAGAAGAGTAAGAAGAAATATAAGAGCCACC

[0879] 3'UTR(TriLink标准UTR)

[0880] GCTGGAGCCTCGGTGGCCATGCTTCTTGCCCCTTGGGCCTCCCCCCAGCCCCTCCTCCCCTTCCTGCAC
CCGTACCCCCGTGGTCTTTGAATAAAGTCTGA

[0881] bhCas12b(V4)的核酸序列

[0882] ATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGCCACCAGATCCTTCATC
CTGAAGATCGAGCCCAACGAGGAAGTGAAGAAAGGCCTCTGGAAAACCCACGAGGTGCTGAACCACGGAATCGCCTA
CTACATGAATATCCTGAAGCTGATCCGGCAAGAGGCCATCTACGAGCACCACGAGCAGGACCCCAAGAATCCCAAGA
AGGTGTCCAAGGCCGAGATCCAGGCCGAGCTGTGGGATTTCGTGCTGAAGATGCAGAAGTGCAACAGCTTCACACAC
GAGGTGGACAAGGACGAGGTGTTCAACATCCTGAGAGAGCTGTACGAGGAACTGGTGCCCAGCAGCGTGGAAAAGAA
GGGCGAAGCCAACCAGCTGAGCAACAAGTTTCTGTACCCTCTGGTGGACCCCAACAGCCAGTCTGGAAAGGGAACAG
CCAGCAGCGGCAGAAAGCCCAGATGGTACAACCTGAAGATTGCCGGCGATCCCTCCTGGGAAGAAGAGAAGAAGAAG
TGGGAAGAAGATAAGAAAAAGGACCCGCTGGCCAAGATCCTGGGCAAGCTGGCTGAGTACGGACTGATCCCTCTGTT
CATCCCCTACACCGACAGCAACGAGCCCATCGTGAAAGAAATCAAGTGGATGGAAAAGTCCCGGAACCAGAGCGTGC
GGCGGCTGGATAAGGACATGTTCATTCAGGCCCTGGAACGGTTCCTGAGCTGGGAGAGCTGGAACCTGAAAGTGAAA
GAGGAATACGAGAAGGTCGAGAAAGAGTACAAGACCCTGGAAGAGAGGATCAAAGAGGACATCCAGGCTCTGAAGGC
TCTGGAACAGTATGAGAAAGAGCGGCAAGAACAGCTGCTGCGGGACACCCTGAACACCAACGAGTACCGGCTGAGCA
AGAGAGGCCTTAGAGGCTGGCGGGAAATCATCCAGAAATGGCTGAAAATGGACGAGAACGAGCCCTCCGAGAAGTAC
CTGGAAGTGTTCAAGGACTACCAGCGGAAGCACCCTAGAGAGGCCGGCGATTACAGCGTGTACGAGTTCCTGTCCAA
GAAAGAGAACCACTTCATCTGGCGGAATCACCCTGAGTACCCCTACCTGTACGCCACCTTCTGCGAGATCGACAAGA
AAAAGAAGGACGCCAAGCAGCAGGCCACCTTCACACTGGCCGATCCTATCAATCACCCTCTGTGGGTCCGATTCGAG
GAAAGAAGCGGCAGCAACCTGAACAAGTACAGAATCCTGACCGAGCAGCTGCACACCGAGAAGCTGAAGAAAAAGCT
GACAGTGCAGCTGGACCGGCTGATCTACCCTACAGAATCTGGCGGCTGGGAAGAGAAGGGCAAAGTGGACATTGTGC
TGCTGCCCAGCCGGCAGTTCTACAACCAGATCTTCCTGGACATCGAGGAAAAGGGCAAGCACGCCTTCACCTACAAG

GATGAGAGCATCAAGTTCCCTCTGAAGGGCACACTCGGCGGAGCCAGAGTGCAGTTCGACAGAGATCACCTGAGAAG
ATACCCTCACAAGGTGGAAAGCGGCAACGTGGGCAGAATCTACTTCAACATGACCGTGAACATCGAGCCTACAGAGT
CCCCAGTGTCCAAGTCTCTGAAGATCCACCGGGACGACTTCCCCAAGGTGGTCAACTTCAAGCCCAAAGAACTGACC
GAGTGGATCAAGGACAGCAAGGGCAAGAAACTGAAGTCCGGCATCGAGTCCCTGGAAATCGGCCTGAGAGTGATGAG
CATCGACCTGGGACAGAGACAGGCCGCTGCCGCCTCTATTTTCGAGGTGGTGGATCAGAAGCCCGACATCGAAGGCA
AGCTGTTTTTCCCAATCAAGGGCACCGAGCTGTATGCCGTGCACAGAGCCAGCTTCAACATCAAGCTGCCCGGCGAG
ACACTGGTCAAGAGCAGAGAAGTGCTGCGGAAGGCCAGAGAGGACAATCTGAAACTGATGAACCAGAAGCTCAACTT
CCTGCGGAACGTGCTGCACTTCCAGCAGTTCGAGGACATCACCGAGAGAGAGAAGCGGGTCACCAAGTGGATCAGCA
GACAAGAGAACAGCGACGTGCCCCTGGTGTACCAGGATGAGCTGATCCAGATCCGCGAGCTGATGTACAAGCCTTAC
AAGGACTGGGTCGCCTTCCTGAAGCAGCTCCACAAGAGACTGGAAGTCGAGATCGGCAAAGAAGTGAAGCACTGGCG
GAAGTCCCTGAGCGACGGAAGAAAGGGCCTGTACGGCATCTCCCTGAAGAACATCGACGAGATCGATCGGACCCGGA
AGTTCCTGCTGAGATGGTCCCTGAGGCCTACCGAACCTGGCGAAGTGCGTAGACTGGAACCCGGCCAGAGATTCGCC
ATCGACCAGCTGAATCACCTGAACGCCCTGAAAGAAGATCGGCTGAAGAAGATGGCCAACACCATCATCATGCACGC
CCTGGGCTACTGCTACGACGTGCGGAAGAAGAAATGGCAGGCTAAGAACCCCGCCTGCCAGATCATCCTGTTCGAGG
ATCTGAGCAACTACAACCCCTACGAGGAAAGGTCCCGCTTCGAGAACAGCAAGCTCATGAAGTGGTCCAGACGCGAG
ATCCCCAGACAGGTTGCACTGCAGGGCGAGATCTATGGCCTGCAAGTGGGAGAAGTGGGCGCTCAGTTCAGCAGCAG
ATTCCACGCCAAGACAGGCAGCCCTGGCATCAGATGTAGCGTCGTGACCAAAGAGAAGCTGCAGGACAATCGGTTCT
TCAAGAATCTGCAGAGAGAGGGCAGACTGACCCTGGACAAAATCGCCGTGCTGAAAGAGGGCGATCTGTACCCAGAC
AAAGGCGGCGAGAAGTTCATCAGCCTGAGCAAGGATCGGAAGTGCGTGACCACACACGCCGACATCAACGCCGCTCA
GAACCTGCAGAAGCGGTTCTGGACAAGAACCCACGGCTTCTACAAGGTGTACTGCAAGGCCTACCAGGTGGACGGCC
AGACCGTGTACATCCCTGAGAGCAAGGACCAGAAGCAGAAGATCATCGAAGAGTTCGGCGAGGGCTACTTCATTCTG
AAGGACGGGGTGTACGAATGGGTCAACGCCGGCAAGCTGAAAATCAAGAAGGGCAGCTCCAAGCAGAGCAGCAGCGA
GCTGGTGGATAGCGACATCCTGAAAGACAGCTTCGACCTGGCCTCCGAGCTGAAAGGCGAAAAGCTGATGCTGTACA
GGGACCCCAGCGGCAATGTGTTCCCCAGCGACAAATGGATGGCCGCTGGCGTGTTCTTCGGAAAGCTGGAACGCATC
CTGATCAGCAAGCTGACCAACCAGTACTCCATCAGCACCATCGAGGACGACAGCAGCAAGCAGTCTATGAAAAGGCC
GGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAG

[0883] 在一些实施方案中,Cas12b是BvCas12B。在一些实施方案中,Cas12b包含氨基酸置换S893R、K846R和E837G,如下文提供的BvCas12b示例性序列中编号。

[0884] BvCas12b(芽孢杆菌属(Bacillussp.V3-13)NCBI参考序列:WP_101661451.1

[0885] MAIRSIKLKMKTNSGTDSIYLRKALWRTHQLINEGIAYYMNLLTLYRQEAIGDKTKEAYQAELINIIRN
QQRNNGSSEEHGSDQEILALLRQLYELIIPSSIGESGDANQLGNKFLYPLVDPNSQSGKGTSNAGRKPRWKRLKEEG
NPDWELEKKKDEERKAKDPTVKIFDNLNKYGLLPLFPLFTNIQKDIEWLPLGKRQSVRKWDKDMFIQAIERLLSWES
WNRRVADEYKQLKEKTESYYKEHLTGGEEWIEKIRKFEKERNMELEKNAFAPNDGYFITSRQIRGWDRVYEKWSKLP
ESASPEELWKVVAEQQNKMSEGFGDPKVFSFLANRENRDIWRGHSERIYHIAAYNGLQKKLSRTKEQATFTLPDAIE
HPLWIRYESPGGTNLNLFKLEEKQKKNYYVTLSKIIWPSEEKWIEKENIEIPLAPSIQFNRQIKLKQHVKGKQEISF
SDYSSRISLDGVLGGSRIQFNRKYIKNHKELLGEGDIGPVFFNLVVDVAPLQETRNGRLQSPIGKALKVISSDFSKV
IDYKPKELMDWMNTGSASNSFGVASLLEGMRVMSIDMGQRTSASVSIFEVVKELPKDQEQKLFYSINDTELFAIHKR
SFLLNLPGEVVTKNNKQQRQERRKKRQFVRSQIRMLANVLRLETKKTPDERKKAIHKLMEIVQSYDSWTASQKEVWE
KELNLLTNMAAFNDEIWKESLVELHHRIEPYVGQIVSKWRKGLSEGRKNLAGISMWNIDELEDTRRLLISWSKRSRT

PGEANRIETDEPFGSSLLQHIQNVKDDRLKQMANLIIMTALGFKYDKEEKDRYKRWKETYPACQIILFENLNRYLFN
LDRSRRENSRLMKWAHRSIPRTVSMQGEMFGLQVGDVRSEYSSRFHAKTGAPGIRCHALTEEDLKAGSNTLKRLIED
GFINESELAYLKKGDIIPSQGGELFVTLSKRYKKDSDNNELTVIHADINAAQNLQKRFWQQNSEVYRVPCQLARMGE
DKLYIPKSQTETIKKYFGKGSFVKNNTEQEVYKWEKSEKMKIKTDTTFDLQDLDGFEDISKTIELAQEQQKKYLTMF
RDPSGYFFNNETWRPQKEYWSIVNNIIKSCLKKKILSNKVEL

[0886] 在一些实施方案中,Cas12b是BTCas12b.BTCas12b(热噬淀粉芽孢杆菌(Bacillus thermoamylovorans))NCBI参考序列:WP_041902512MATRSFILKIEPNEEVKKGLWKTHEVLNHGIA
YYMNILKLIRQEAIYEHHEQDPKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDVVFNILRELYEELVPSSVEK
KGEANQLSNKFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEEKKKWEEDKKKDPLAKILGKLAEYGLIPL
FIPFTDSNEPIVKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWESWNLKVKEEYEKVEKEHKTLEERIKEDIQAFK
SLEQYEKERQEQLLRDTLNTNEYRLSKRGLRGWREIIQKWLKMDENEPSEKYLEVFKDYQRKHPREAGDYSVYEFLS
KKENHFIWRNHPEYPYLYATFCEIDKKKKDAKQQATFTLADPINHPLWVRFEERSGSNLNKYRILTEQLHTEKLKKK
LTVQLDRLIYPTESGGWEEKGKVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGTLGGARVQFDRDHLR
RYPHKVESGNVGRIYFNMTVNIEPTESPVSKSLKIHRDDFPKFVNFKPKELTEWIKDSKGKKLKSGIESLEIGLRVM
SIDLGQRQAAAASIFEVVDQKPDIEGKLFFPIKGTELYAVHRASFNIKLPGETLVKSREVLRKAREDNLKLMNQKLN
FLRNVLHFQQFEDITEREKRVTKWISRQENSDVPLVYQDELIQIRELMYKPYKDWVAFLKQLHKRLEVEIGKEVKHW
RKSLSDGRKGLYGISLKNIDEIDRTRKFLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHLNALKEDRLKKMANTIIMH
ALGYCYDVRKKKWQAKNPACQIILFEDLSNYNPYEERSRFENSKLMKWSRREIPRQVALQGEIYGLQVGEVGAQFSS
RFHAKTGSPGIRCSVVTKEKLQDNRFFKNLQREGRLTLDKIAVLKEGDLYPDKGGEKFISLSKDRKLVTTHADINAA
QNLQKRFWTRTHGFYKVYCKAYQVDGQTVYIPESKDQKQKIIEEFGEGYFILKDGVYEWGNAGKLKIKKGSSKQSSS
ELVDSDILKDSFDLASELKGEKLMLYRDPSGNVFPSDKWMAAGVFFGKLERILISKLTNQYSISTIEDDSSKQSM

[0887] 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(Oleiphilus)HI0009的Cas12c(即OspCas12c)或OspCas12c的变体。这些Cas12c分子已经在Yan等人,“Functionally Diverse Type V CRISPR-Cas Systems,”Science,2019Jan.4;363:88-91中描述,其整体内容通过引用并入本文。在一些实施方案中,napDNAbp包含氨基酸序列,该氨基酸序列与天然存在的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包含氨基酸序列,该氨基酸序列与本文所述的任何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。应了解,来自其他细菌物种的Cas12c1、Cas12c2或OspCas12c也可根据本公开使用。

[0888] Cas12c1

[0889] MQTKKTHLHLISAKASRKYRRTIACLSDTAKKDLERRKQSGAADPAQELSCLKTIKFKLEVPEGSKLPS
FDRISQIYNALETIEKGSLSYLLFALILSGFRIFPNSSAAKTFASSSCYKNDQFASQIKEIFGEMVKNFIPSELESI
LKKGRRKNNKDWTEENIKRVLNSEFGRKNSEGSSALFDSFLSKFSQELFRKFDSWNEVNKKYLEAAELLDSMLASYG
PFDSVCKMIGDSDSRNSLPDKSTIAFTNNAEITVDIESSVMPYMAIAALLREYRQSKSKAAPVAYVQSHLTTTNGNG

LSWFFKFGLDLIRKAPVSSKQSTSDGSKSLQELFSVPDDKLDGLKFIKEACEALPEASLLCGEKGELLGYQDFRTSF
AGHIDSWVANYVNRLFELIELVNQLPESIKLPSILTQKNHNLVASLGLQEAEVSHSLELFEGLVKNVRQTLKKLAGI
DISSSPNEQDIKEFYAFSDVLNRLGSIRNQIENAVQTAKKDKIDLESAIEWKEWKKLKKLPKLNGLGGGVPKQQELL
DKALESVKQIRHYQRIDFERVIQWAVNEHCLETVPKFLVDAEKKKINKESSTDFAAKENAVRFLLEGIGAAARGKTD
SVSKAAYNWFVVNNFLAKKDLNRYFINCQGCIYKPPYSKRRSLAFALRSDNKDTIEVVWEKFETFYKEISKEIEKFN
IFSQEFQTFLHLENLRMKLLLRRIQKPIPAEIAFFSLPQEYYDSLPPNVAFLALNQEITPSEYITQFNLYSSFLNGN
LILLRRSRSYLRAKFSWVGNSKLIYAAKEARLWKIPNAYWKSDEWKMILDSNVLVFDKAGNVLPAPTLKKVCEREGD
LRLFYPLLRQLPHDWCYRNPFVKSVGREKNVIEVNKEGEPKVASALPGSLFRLIGPAPFKSLLDDCFFNPLDKDLRE
CMLIVDQEISQKVEAQKVEASLESCTYSIAVPIRYHLEEPKVSNQFENVLAIDQGEAGLAYAVFSLKSIGEAETKPI
AVGTIRIPSIRRLIHSVSTYRKKKQRLQNFKQNYDSTAFIMRENVTGDVCAKIVGLMKEFNAFPVLEYDVKNLESGS
RQLSAVYKAVNSHFLYFKEPGRDALRKQLWYGGDSWTIDGIEIVTRERKEDGKEGVEKIVPLKVFPGRSVSARFTSK
TCSCCGRNVFDWLFTEKKAKTNKKFNVNSKGELTTADGVIQLFEADRSKGPKFYARRKERTPLTKPIAKGSYSLEEI
ERRVRTNLRRAPKSKQSRDTSQSQYFCVYKDCALHFSGMQADENAAINIGRRFLTALRKNRRSDFPSNVKISDRLLD
N

[0890] Cas12c2

[0891] MTKHSIPLHAFRNSGADARKWKGRIALLAKRGKETMRTLQFPLEMSEPEAAAINTTPFAVAYNAIEGTG
KGTLFDYWAKLHLAGFRFFPSGGAATIFRQQAVFEDASWNAAFCQQSGKDWPWLVPSKLYERFTKAPREVAKKDGSK
KSIEFTQENVANESHVSLVGASITDKTPEDQKEFFLKMAGALAEKFDSWKSANEDRIVAMKVIDEFLKSEGLHLPSL
ENIAVKCSVETKPDNATVAWHDAPMSGVQNLAIGVFATCASRIDNIYDLNGGKLSKLIQESATTPNVTALSWLFGKG
LEYFRTTDIDTIMQDFNIPASAKESIKPLVESAQAIPTMTVLGKKNYAPFRPNFGGKIDSWIANYASRLMLLNDILE
QIEPGFELPQALLDNETLMSGIDMTGDELKELIEAVYAWVDAAKQGLATLLGRGGNVDDAVQTFEQFSAMMDTLNGT
LNTISARYVRAVEMAGKDEARLEKLIECKFDIPKWCKSVPKLVGISGGLPKVEEEIKVMNAAFKDVRARMFVRFEEI
AAYVASKGAGMDVYDALEKRELEQIKKLKSAVPERAHIQAYRAVLHRIGRAVQNCSEKTKQLFSSKVIEMGVFKNPS
HLNNFIFNQKGAIYRSPFDRSRHAPYQLHADKLLKNDWLELLAEISATLMASESTEQMEDALRLERTRLQLQLSGLP
DWEYPASLAKPDIEVEIQTALKMQLAKDTVTSDVLQRAFNLYSSVLSGLTFKLLRRSFSLKMRFSVADTTQLIYVPK
VCDWAIPKQYLQAEGEIGIAARVVTESSPAKMVTEVEMKEPKALGHFMQQAPHDWYFDASLGGTQVAGRIVEKGKEV
GKERKLVGYRMRGNSAYKTVLDKSLVGNTELSQCSMIIEIPYTQTVDADFRAQVQAGLPKVSINLPVKETITASNKD
EQMLFDRFVAIDLGERGLGYAVFDAKTLELQESGHRPIKAITNLLNRTHHYEQRPNQRQKFQAKFNVNLSELRENTV
GDVCHQINRICAYYNAFPVLEYMVPDRLDKQLKSVYESVTNRYIWSSTDAHKSARVQFWLGGETWEHPYLKSAKDKK
PLVLSPGRGASGKGTSQTCSCCGRNPFDLIKDMKPRAKIAVVDGKAKLENSELKLFERNLESKDDMLARRHRNERAG
MEQPLTPGNYTVDEIKALLRANLRRAPKNRRTKDTTVSEYHCVFSDCGKTMHADENAAVNIGGKFIADIEK

[0892] OspCas12c

[0893] MTKLRHRQKKLTHDWAGSKKREVLGSNGKLQNPLLMPVKKGQVTEFRKAFSAYARATKGEMTDGRKNMF
THSFEPFKTKPSLHQCELADKAYQSLHSYLPGSLAHFLLSAHALGFRIFSKSGEATAFQASSKIEAYESKLASELAC
VDLSIQNLTISTLFNALTTSVRGKGEETSADPLIARFYTLLTGKPLSRDTQGPERDLAEVISRKIASSFGTWKEMTA
NPLQSLQFFEEELHALDANVSLSPAFDVLIKMNDLQGDLKNRTIVFDPDAPVFEYNAEDPADIIIKLTARYAKEAVI
KNQNVGNYVKNAITTTNANGLGWLLNKGLSLLPVSTDDELLEFIGVERSHPSCHALIELIAQLEAPELFEKNVFSDT
RSEVQGMIDSAVSNHIARLSSSRNSLSMDSEELERLIKSFQIHTPHCSLFIGAQSLSQQLESLPEALQSGVNSADIL
LGSTQYMLTNSLVEESIATYQRTLNRINYLSGVAGQINGAIKRKAIDGEKIHLPAAWSELISLPFIGQPVIDVESDL

AHLKNQYQTLSNEFDTLISALQKNFDLNFNKALLNRTQHFEAMCRSTKKNALSKPEIVSYRDLLARLTSCLYRGSLV
LRRAGIEVLKKHKIFESNSELREHVHERKHFVFVSPLDRKAKKLLRLTDSRPDLLHVIDEILQHDNLENKDRESLWL
VRSGYLLAGLPDQLSSSFINLPIITQKGDRRLIDLIQYDQINRDAFVMLVTSAFKSNLSGLQYRANKQSFVVTRTLS
PYLGSKLVYVPKDKDWLVPSQMFEGRFADILQSDYMVWKDAGRLCVIDTAKHLSNIKKSVFSSEEVLAFLRELPHRT
FIQTEVRGLGVNVDGIAFNNGDIPSLKTFSNCVQVKVSRTNTSLVQTLNRWFEGGKVSPPSIQFERAYYKKDDQIHE
DAAKRKIRFQMPATELVHASDDAGWTPSYLLGIDPGEYGMGLSLVSINNGEVLDSGFIHINSLINFASKKSNHQTKV
VPRQQYKSPYANYLEQSKDSAAGDIAHILDRLIYKLNALPVFEALSGNSQSAADQVWTKVLSFYTWGDNDAQNSIRK
QHWFGASHWDIKGMLRQPPTEKKPKPYIAFPGSQVSSYGNSQRCSCCGRNPIEQLREMAKDTSIKELKIRNSEIQLF
DGTIKLFNPDPSTVIERRRHNLGPSRIPVADRTFKNISPSSLEFKELITIVSRSIRHSPEFIAKKRGIGSEYFCAYS
DCNSSLNSEANAAANVAQKFQKQLFFEL

[0894] 在一些实施方案中,napDNAbp是指Cas12g、Cas12h或Cas12i,其已经在例如Yan等人,“Functionally Diverse Type V CRISPR-Cas Systems,”Science,2019Jan.4;363:88-91中描述,各自的整体内容通过引用并入本文。通过聚集超过10TB的序列数据,发现了V型Cas蛋白质的新分类,这些蛋白质与之前描述的V类蛋白质(包括Cas12g、Cas12h和Cas12i)表现出微弱的相似性。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。应了解,其他RNA引导的DNA结合蛋白可以用作napDNAbp,并且出于本公开的范畴内。在一些实施方案中,napDNAbp包含氨基酸序列,该氨基酸序列与天然存在的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包含氨基酸序列,该氨基酸序列与本文所述的任何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。应了解,来自其他细菌物种的Cas12g、Cas12h或Cas12i也可根据本公开使用。在一些实施方案中,Cas12i是Cas12i1或Cas12i2。

[0895] Cas12g1

[0896] MAQASSTPAVSPRPRPRYREERTLVRKLLPRPGQSKQEFRENVKKLRKAFLQFNADVSGVCQWAIQFRP
RYGKPAEPTETFWKFFLEPETSLPPNDSRSPEFRRLQAFEAAAGINGAAALDDPAFTNELRDSILAVASRPKTKEAQ
RLFSRLKDYQPAHRMILAKVAAEWIESRYRRAHQNWERNYEEWKKEKQEWEQNHPELTPEIREAFNQIFQQLEVKEK
RVRICPAARLLQNKDNCQYAGKNKHSVLCNQFNEFKKNHLQGKAIKFFYKDAEKYLRCGLQSLKPNVQGPFREDWNK
YLRYMNLKEETLRGKNGGRLPHCKNLGQECEFNPHTALCKQYQQQLSSRPDLVQHDELYRKWRREYWREPRKPVFRY
PSVKRHSIAKIFGENYFQADFKNSVVGLRLDSMPAGQYLEFAFAPWPRNYRPQPGETEISSVHLHFVGTRPRIGFRF
RVPHKRSRFDCTQEELDELRSRTFPRKAQDQKFLEAARKRLLETFPGNAEQELRLLAVDLGTDSARAAFFIGKTFQQ
AFPLKIVKIEKLYEQWPNQKQAGDRRDASSKQPRPGLSRDHVGRHLQKMRAQASEIAQKRQELTGTPAPETTTDQAA
KKATLQPFDLRGLTVHTARMIRDWARLNARQIIQLAEENQVDLIVLESLRGFRPPGYENLDQEKKRRVAFFAHGRIR
RKVTEKAVERGMRVVTVPYLASSKVCAECRKKQKDNKQWEKNKKRGLFKCEGCGSQAQVDENAARVLGRVFWGEIEL
PTAIP

[0897] Cas12h1

[0898] MKVHEIPRSQLLKIKQYEGSFVEWYRDLQEDRKKFASLLFRWAAFGYAAREDDGATYISPSQALLERRL

LLGDAEDVAIKFLDVLFKGGAPSSSCYSLFYEDFALRDKAKYSGAKREFIEGLATMPLDKIIERIRQDEQLSKIPAE
EWLILGAEYSPEEIWEQVAPRIVNVDRSLGKQLRERLGIKCRRPHDAGYCKILMEVVARQLRSHNETYHEYLNQTHE
MKTKVANNLTNEFDLVCEFAEVLEEKNYGLGWYVLWQGVKQALKEQKKPTKIQIAVDQLRQPKFAGLLTAKWRALKG
AYDTWKLKKRLEKRKAFPYMPNWDNDYQIPVGLTGLGVFTLEVKRTEVVVDLKEHGKLFCSHSHYFGDLTAEKHPSR
YHLKFRHKLKLRKRDSRVEPTIGPWIEAALREITIQKKPNGVFYLGLPYALSHGIDNFQIAKRFFSAAKPDKEVING
LPSEMVVGAADLNLSNIVAPVKARIGKGLEGPLHALDYGYGELIDGPKILTPDGPRCGELISLKRDIVEIKSAIKEF
KACQREGLTMSEETTTWLSEVESPSDSPRCMIQSRIADTSRRLNSFKYQMNKEGYQDLAEALRLLDAMDSYNSLLES
YQRMHLSPGEQSPKEAKFDTKRASFRDLLRRRVAHTIVEYFDDCDIVFFEDLDGPSDSDSRNNALVKLLSPRTLLLY
IRQALEKRGIGMVEVAKDGTSQNNPISGHVGWRNKQNKSEIYFYEDKELLVMDADEVGAMNILCRGLNHSVCPYSFV
TKAPEKKNDEKKEGDYGKRVKRFLKDRYGSSNVRFLVASMGFVTVTTKRPKDALVGKRLYYHGGELVTHDLHNRMKD
EIKYLVEKEVLARRVSLSDSTIKSYKSFAHV

[0899] Cas12i1

[0900] MSNKEKNASETRKAYTTKMIPRSHDRMKLLGNFMDYLMDGTPIFFELWNQFGGGIDRDIISGTANKDKI
SDDLLLAVNWFKVMPINSKPQGVSPSNLANLFQQYSGSEPDIQAQEYFASNFDTEKHQWKDMRVEYERLLAELQLSR
SDMHHDLKLMYKEKCIGLSLSTAHYITSVMFGTGAKNNRQTKHQFYSKVIQLLEESTQINSVEQLASIILKAGDCDS
YRKLRIRCSRKGATPSILKIVQDYELGTNHDDEVNVPSLIANLKEKLGRFEYECEWKCMEKIKAFLASKVGPYYLGS
YSAMLENALSPIKGMTTKNCKFVLKQIDAKNDIKYENEPFGKIVEGFFDSPYFESDTNVKWVLHPHHIGESNIKTLW
EDLNAIHSKYEEDIASLSEDKKEKRIKVYQGDVCQTINTYCEEVGKEAKTPLVQLLRYLYSRKDDIAVDKIIDGITF
LSKKHKVEKQKINPVIQKYPSFNFGNNSKLLGKIISPKDKLKHNLKCNRNQVDNYIWIEIKVLNTKTMRWEKHHYAL
SSTRFLEEVYYPATSENPPDALAARFRTKTNGYEGKPALSAEQIEQIRSAPVGLRKVKKRQMRLEAARQQNLLPRYT
WGKDFNINICKRGNNFEVTLATKVKKKKEKNYKVVLGYDANIVRKNTYAAIEAHANGDGVIDYNDLPVKPIESGFVT
VESQVRDKSYDQLSYNGVKLLYCKPHVESRRSFLEKYRNGTMKDNRGNNIQIDFMKDFEAIADDETSLYYFNMKYCK
LLQSSIRNHSSQAKEYREEIFELLRDGKLSVLKLSSLSNLSFVMFKVAKSLIGTYFGHLLKKPKNSKSDVKAPPITD
EDKQKADPEMFALRLALEEKRLNKVKSKKEVIANKIVAKALELRDKYGPVLIKGENISDTTKKGKKSSTNSFLMDWL
ARGVANKVKEMVMMHQGLEFVEVNPNFTSHQDPFVHKNPENTFRARYSRCTPSELTEKNRKEILSFLSDKPSKRPTN
AYYNEGAMAFLATYGLKKNDVLGVSLEKFKQIMANILHQRSEDQLLFPSRGGMFYLATYKLDADATSVNWNGKQFWV
CNADLVAAYNVGLVDIQKDFKKK

[0901] Cas12i2

[0902] MSSAIKSYKSVLRPNERKNQLLKSTIQCLEDGSAFFFKMLQGLFGGITPEIVRFSTEQEKQQQDIALWC
AVNWFRPVSQDSLTHTIASDNLVEKFEEYYGGTASDAIKQYFSASIGESYYWNDCRQQYYDLCRELGVEVSDLTHDL
EILCREKCLAVATESNQNNSIISVLFGTGEKEDRSVKLRITKKILEAISNLKEIPKNVAPIQEIILNVAKATKETFR
QVYAGNLGAPSTLEKFIAKDGQKEFDLKKLQTDLKKVIRGKSKERDWCCQEELRSYVEQNTIQYDLWAWGEMFNKAH
TALKIKSTRNYNFAKQRLEQFKEIQSLNNLLVVKKLNDFFDSEFFSGEETYTICVHHLGGKDLSKLYKAWEDDPADP
ENAIVVLCDDLKNNFKKEPIRNILRYIFTIRQECSAQDILAAAKYNQQLDRYKSQKANPSVLGNQGFTWTNAVILPE
KAQRNDRPNSLDLRIWLYLKLRHPDGRWKKHHIPFYDTRFFQEIYAAGNSPVDTCQFRTPRFGYHLPKLTDQTAIRV
NKKHVKAAKTEARIRLAIQQGTLPVSNLKITEISATINSKGQVRIPVKFDVGRQKGTLQIGDRFCGYDQNQTASHAY
SLWEVVKEGQYHKELGCFVRFISSGDIVSITENRGNQFDQLSYEGLAYPQYADWRKKASKFVSLWQITKKNKKKEIV
TVEAKEKFDAICKYQPRLYKFNKEYAYLLRDIVRGKSLVELQQIRQEIFRFIEQDCGVTRLGSLSLSTLETVKAVKG
IIYSYFSTALNASKNNPISDEQRKEFDPELFALLEKLELIRTRKKKQKVERIANSLIQTCLENNIKFIRGEGDLSTT

NNATKKKANSRSMDWLARGVFNKIRQLAPMHNITLFGCGSLYTSHQDPLVHRNPDKAMKCRWAAIPVKDIGDWVLRK
LSQNLRAKNIGTGEYYHQGVKEFLSHYELQDLEEEELLKWRSDRKSNIPCWVLQNRLAEKLGNKEAVVYIPVRGGRIY
FATHKVATGAVSIVFDQKQVWVCNADHVAAANIALTVKGIGEQSSDEENPDGSRIKLQLTS

[0903] 碱基编辑器的代表性核酸和蛋白序列如下：

[0904] BhCas12b GGSGGS-ABE8-Xten20在P153处

[0905] GCCACCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGC
CACCAGATCCTTCATCCTGAAGATCGAGCCCAACGAGGAAGTGAAGAAAGGCCTCTGGA

AAACCCACGAGGTGCTGAACCACGGAATCGCCTACTACATGAATATCCTGAAGCTGATC

CGGCAAGAGGCCATCTACGAGCACCACGAGCAGGACCCCAAGAATCCCAAGAAGGTGTC

CAAGGCCGAGATCCAGGCCGAGCTGTGGGATTTCGTGCTGAAGATGCAGAAGTGCAACA

GCTTCACACACGAGGTGGACAAGGACGAGGTGTTCAACATCCTGAGAGAGCTGTACGAG

GAACTGGTGCCCAGCAGCGTGGAAAAGAAGGGCGAAGCCAACCAGCTGAGCAACAAGTT

TCTGTACCCTCTGGTGGACCCCAACAGCCAGTCTGGAAAGGGAACAGCCAGCAGCGGCA

GAAAGCCCAGATGGTACAACCTGAAGATTGCCGGCGATCCCggaggctctggaggaagc

TCCGAAGTCGAGTTTTCCCATGAGTACTGGATGAGACACGCATTGACTCTCGCAAAGAG

GGCTCGAGATGAACGCGAGGTGCCCGTGGGGGCAGTACTCGTGCTCAACAATCGCGTAA

TCGGCGAAGGTTGGAATAGGGCAATCGGACTCCACGACCCCACTGCACATGCGGAAATC

ATGGCCCTTCGACAGGGAGGGCTTGTGATGCAGAATTATCGACTTTATGATGCGACGCT

GTACGTCACGTTTGAACCTTGCGTAATGTGCGCGGGAGCTATGATTCACTCCCGCATTG

GACGAGTTGTATTCGGTGTTCGCAACGCCAAGACGGGTGCCGCAGGTTCACTGATGGAC

GTGCTGCATCATCCAGGCATGAACCACCGGGTAGAAATCACAGAAGGCATATTGGCGGA

CGAATGTGCGGCGCTGTTGTGTCGTTTTTTTCGCATGCCCAGGCGGGTCTTTAACGCCC

AGAAAAAAGCACAATCCTCTACTGACGGCTCTTCTGGATCTGAAACACCTGGCACAAGC

GAGAGCGCCACCCCTGAGAGCTCTGGCTCCTGGGAAGAAGAGAAGAAGAAGTGGGAAGA

[0906] AGATAAGAAAAAGGACCCGCTGGCCAAGATCCTGGGCAAGCTGGCTGAGTACGGACTGA

TCCCTCTGTTCATCCCCTACACCGACAGCAACGAGCCCATCGTGAAAGAAATCAAGTGG

ATGGAAAAGTCCCGGAACCAGAGCGTGCGGCGGCTGGATAAGGACATGTTCATTCAGGC

CCTGGAACGGTTCCTGAGCTGGGAGAGCTGGAACCTGAAAGTGAAAGAGGAATACGAGA

AGGTCGAGAAAGAGTACAAGACCCTGGAAGAGAGGATCAAAGAGGACATCCAGGCTCTG

AAGGCTCTGGAACAGTATGAGAAAGAGCGGCAAGAACAGCTGCTGCGGGACACCCTGAA

CACCAACGAGTACCGGCTGAGCAAGAGAGGCCTTAGAGGCTGGCGGGAAATCATCCAGA

AATGGCTGAAAATGGACGAGAACGAGCCCTCCGAGAAGTACCTGGAAGTGTTCAAGGAC

TACCAGCGGAAGCACCCTAGAGAGGCCGGCGATTACAGCGTGTACGAGTTCCTGTCCAA

GAAAGAGAACCACTTCATCTGGCGGAATCACCCTGAGTACCCCTACCTGTACGCCACCT

TCTGCGAGATCGACAAGAAAAAGAAGGACGCCAAGCAGCAGGCCACCTTCACACTGGCC

GATCCTATCAATCACCCTCTGTGGGTCCGATTCGAGGAAAGAAGCGGCAGCAACCTGAA

CAAGTACAGAATCCTGACCGAGCAGCTGCACACCGAGAAGCTGAAGAAAAAGCTGACAG

TGCAGCTGGACCGGCTGATCTACCCTACAGAATCTGGCGGCTGGGAAGAGAAGGGCAAA

GTGGACATTGTGCTGCTGCCCAGCCGGCAGTTCTACAACCAGATCTTCCTGGACATCGA

GGAAAAGGGCAAGCACGCCTTCACCTACAAGGATGAGAGCATCAAGTTCCCTCTGAAGG

GCACACTCGGCGGAGCCAGAGTGCAGTTCGACAGAGATCACCTGAGAAGATACCCTCAC

AAGGTGGAAAGCGGCAACGTGGGCAGAATCTACTTCAACATGACCGTGAACATCGAGCC

TACAGAGTCCCCAGTGTCCAAGTCTCTGAAGATCCACCGGGACGACTTCCCCAAGGTGG

TCAACTTCAAGCCCAAAGAACTGACCGAGTGGATCAAGGACAGCAAGGGCAAGAAACTG

AAGTCCGGCATCGAGTCCCTGGAAATCGGCCTGAGAGTGATGAGCATCGACCTGGGACA

GAGACAGGCCGCTGCCGCCTCTATTTTCGAGGTGGTGGATCAGAAGCCCGACATCGAAG

GCAAGCTGTTTTTCCCAATCAAGGGCACCGAGCTGTATGCCGTGCACAGAGCCAGCTTC

AACATCAAGCTGCCCGGCGAGACACTGGTCAAGAGCAGAGAAGTGCTGCGGAAGGCCAG

AGAGGACAATCTGAAACTGATGAACCAGAAGCTCAACTTCCTGCGGAACGTGCTGCACT

TCCAGCAGTTCGAGGACATCACCGAGAGAGAGAAGCGGGTCACCAAGTGGATCAGCAGA

CAAGAGAACAGCGACGTGCCCCTGGTGTACCAGGATGAGCTGATCCAGATCCGCGAGCT

GATGTACAAGCCTTACAAGGACTGGGTCGCCTTCCTGAAGCAGCTCCACAAGAGACTGG

AAGTCGAGATCGGCAAAGAAGTGAAGCACTGGCGGAAGTCCCTGAGCGACGGAAGAAAG

GGCCTGTACGGCATCTCCCTGAAGAACATCGACGAGATCGATCGGACCCGGAAGTTCCT

GCTGAGATGGTCCCTGAGGCCTACCGAACCTGGCGAAGTGCGTAGACTGGAACCCGGCC

AGAGATTCGCCATCGACCAGCTGAATCACCTGAACGCCCTGAAAGAAGATCGGCTGAAG

AAGATGGCCAACACCATCATCATGCACGCCCTGGGCTACTGCTACGACGTGCGGAAGAA

GAAATGGCAGGCTAAGAACCCCGCCTGCCAGATCATCCTGTTCGAGGATCTGAGCAACT

[0907] ACAACCCCTACGAGGAAAGGTCCCGCTTCGAGAACAGCAAGCTCATGAAGTGGTCCAGA

CGCGAGATCCCCAGACAGGTTGCACTGCAGGGCGAGATCTATGGCCTGCAAGTGGGAGA

AGTGGGCGCTCAGTTCAGCAGCAGATTCCACGCCAAGACAGGCAGCCCTGGCATCAGAT

GTAGCGTCGTGACCAAAGAGAAGCTGCAGGACAATCGGTTCTTCAAGAATCTGCAGAGA

GAGGGCAGACTGACCCTGGACAAAATCGCCGTGCTGAAAGAGGGCGATCTGTACCCAGA

CAAAGGCGGCGAGAAGTTCATCAGCCTGAGCAAGGATCGGAAGTGCGTGACCACACACG

CCGACATCAACGCCGCTCAGAACCTGCAGAAGCGGTTCTGGACAAGAACCCACGGCTTC

TACAAGGTGTACTGCAAGGCCTACCAGGTGGACGGCCAGACCGTGTACATCCCTGA

GAGCAAGGACCAGAAGCAGAAGATCATCGAAGAGTTCGGCGAGGGCTACTTCATTCTGA

AGGACGGGGTGTACGAATGGGTCAACGCCGGCAAGCTGAAAATCAAGAAGGGCAGCTCC

AAGCAGAGCAGCAGCGAGCTGGTGGATAGCGACATCCTGAAAGACAGCTTCGACCTGGC

CTCCGAGCTGAAAGGCGAAAAGCTGATGCTGTACAGGGACCCCAGCGGCAATGTGTTCC

CCAGCGACAAATGGATGGCCGCTGGCGTGTTCTTCGGAAAGCTGGAACGCATCCTGATC

AGCAAGCTGACCAACCAGTACTCCATCAGCACCATCGAGGACGACAGCAGCAAGCAGTC

TATGAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGGATCCT

ACCCATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCA

*TATGATGTCCCCGACTATGCC*TAA

MAPKKKRKVGIHGVPAAATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQ
EAIYEHHEQDPKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEEL
VPSSVEKKGEANQLSNKFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPGGSGGSSE
VEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVL
HHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSES
ATPESSGSWEEEKKKWEEDKKKDPLAKILGKLAEYGLIPLFIPYTDSNEPIVKEIKWME
KSRNQSVRRLDKDMFIQALERFLSWESWNLKVKEEYEKVEKEYKTLEERIKEDIQALKA
LEQYEKERQEQLLRDTLNTNEYRLSKRGLRGWREIIQKWLKMDENEPSEKYLEVFKDYQ
RKHPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYATFCEIDKKKKDAKQQATFTLADP
INHPLWVRFEERSGSNLNKYRILTEQLHTEKLKKKLTVQLDRLIYPTESGGWEEKGKVD
IVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGTLGGARVQFDRDHLRRYPHKV
[0908] ESGNVGRIYFNMTVNIEPTESPVSKSLKIHRDDFPKVVNFKPKELTEWIKDSKGKKLKS
GIESLEIGLRVMSIDLGQRQAAAASIFEVVDQKPDIEGKLFFPIKGTELYAVHRASFNI
KLPGETLVKSREVLRKAREDNLKLMNQKLNFLRNVLHFQQFEDITEREKRVTKWISRQE
NSDVPLVYQDELIQIRELMYKPYKDWVAFLKQLHKRLEVEIGKEVKHWRKSLSDGRKGL
YGISLKNIDEIDRTRKFLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHLNALKEDRLKKM
ANTIIMHALGYCYDVRKKKWQAKNPACQIILFEDLSNYNPYEERSRFENSKLMKWSRRE
IPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKLQDNRFFKNLQREG
RLTLDKIAVLKEGDLYPDKGGEKFISLSKDRKCVTTHADINAAQNLQKRFWTRTHGFYK
VYCKAYQVDGQTVYIPESKDQKQKIIEEFGEGYFILKDGVYEWVNAGKLKIKKGSSKQS
SSELVDSDILKDSFDLASELKGEKLMLYRDPSGNVFPSDKWMAAGVFFGKLERILISKL
TNQYSISTIEDDSSKQSMKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDV
PDYA

[0909] BhCas12b GGSGGS-ABE8-Xten20在K255处

GCCACCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGC
CACCAGATCCTTCATCCTGAAGATCGAGCCCAACGAGGAAGTGAAGAAAGGCCTCTGGA
AAACCCACGAGGTGCTGAACCACGGAATCGCCTACTACATGAATATCCTGAAGCTGATC
CGGCAAGAGGCCATCTACGAGCACCACGAGCAGGACCCCAAGAATCCCAAGAAGGTGTC
CAAGGCCGAGATCCAGGCCGAGCTGTGGGATTTCGTGCTGAAGATGCAGAAGTGCAACA
GCTTCACACACGAGGTGGACAAGGACGAGGTGTTCAACATCCTGAGAGAGCTGTACGAG
GAACTGGTGCCCAGCAGCGTGGAAAAGAAGGGCGAAGCCAACCAGCTGAGCAACAAGTT
TCTGTACCCTCTGGTGGACCCCAACAGCCAGTCTGGAAAGGGAACAGCCAGCAGCGGCA

GAAAGCCCAGATGGTACAACCTGAAGATTGCCGGCGATCCCTCCTGGGAAGAAGAGAAG

AAGAAGTGGGAAGAAGATAAGAAAAAGGACCCGCTGGCCAAGATCCTGGGCAAGCTGGC

TGAGTACGGACTGATCCCTCTGTTCATCCCCTACACCGACAGCAACGAGCCCATCGTGA

AAGAAATCAAGTGGATGGAAAAGTCCCGGAACCAGAGCGTGCGGCGGCTGGATAAGGAC

ATGTTCATTCAGGCCCTGGAACGGTTCCTGAGCTGGGAGAGCTGGAACCTGAAAGTGAA

AGAGGAATACGAGAAGGTCGAGAAAGAGTACAAGACCCTGGAAGAGAGGATCAAAggag

gctctggaggaagcTCCGAAGTCGAGTTTTCCCATGAGTACTGGATGAGACACGCATTG

ACTCTCGCAAAGAGGGCTCGAGATGAACGCGAGGTGCCCGTGGGGGCAGTACTCGTGCT

CAACAATCGCGTAATCGGCGAAGGTTGGAATAGGGCAATCGGACTCCACGACCCCACTG

CACATGCGGAAATCATGGCCCTTCGACAGGGAGGGCTTGTGATGCAGAATTATCGACTT

TATGATGCGACGCTGTACGTCACGTTTGAACCTTGCGTAATGTGCGCGGGAGCTATGAT

TCACTCCCGCATTGGACGAGTTGTATTCGGTGTTCGCAACGCCAAGACGGGTGCCGCAG

GTTCACTGATGGACGTGCTGCATCATCCAGGCATGAACCACCGGGTAGAAATCACAGAA

GGCATATTGGCGGACGAATGTGCGGCGCTGTTGTGTCGTTTTTTTCGCATGCCCAGGCG

GGTCTTTAACGCCCAGAAAAAAGCACAATCCTCTACTGACGGCTCTTCTGGATCTGAAA

CACCTGGCACAAGCGAGAGCGCCACCCCTGAGAGCTCTGGCGAGGACATCCAGGCTCTG

AAGGCTCTGGAACAGTATGAGAAAGAGCGGCAAGAACAGCTGCTGCGGGACACCCTGAA

[0910] CACCAACGAGTACCGGCTGAGCAAGAGAGGCCTTAGAGGCTGGCGGGAAATCATCCAGA

AATGGCTGAAAATGGACGAGAACGAGCCCTCCGAGAAGTACCTGGAAGTGTTCAAGGAC

TACCAGCGGAAGCACCCTAGAGAGGCCGGCGATTACAGCGTGTACGAGTTCCTGTCCAA

GAAAGAGAACCACTTCATCTGGCGGAATCACCCTGAGTACCCCTACCTGTACGCCACCT

TCTGCGAGATCGACAAGAAAAAGAAGGACGCCAAGCAGCAGGCCACCTTCACACTGGCC

GATCCTATCAATCACCCTCTGTGGGTCCGATTCGAGGAAAGAAGCGGCAGCAACCTGAA

CAAGTACAGAATCCTGACCGAGCAGCTGCACACCGAGAAGCTGAAGAAAAAGCTGACAG

TGCAGCTGGACCGGCTGATCTACCCTACAGAATCTGGCGGCTGGGAAGAGAAGGGCAAA

GTGGACATTGTGCTGCTGCCCAGCCGGCAGTTCTACAACCAGATCTTCCTGGACATCGA

GGAAAAGGGCAAGCACGCCTTCACCTACAAGGATGAGAGCATCAAGTTCCCTCTGAAGG

GCACACTCGGCGGAGCCAGAGTGCAGTTCGACAGAGATCACCTGAGAAGATACCCTCAC

AAGGTGGAAAGCGGCAACGTGGGCAGAATCTACTTCAACATGACCGTGAACATCGAGCC

TACAGAGTCCCCAGTGTCCAAGTCTCTGAAGATCCACCGGGACGACTTCCCCAAGGTGG

TCAACTTCAAGCCCAAAGAACTGACCGAGTGGATCAAGGACAGCAAGGGCAAGAAACTG

AAGTCCGGCATCGAGTCCCTGGAAATCGGCCTGAGAGTGATGAGCATCGACCTGGGACA

GAGACAGGCCGCTGCCGCCTCTATTTTCGAGGTGGTGGATCAGAAGCCCGACATCGAAG

GCAAGCTGTTTTTCCCAATCAAGGGCACCGAGCTGTATGCCGTGCACAGAGCCAGCTTC

AACATCAAGCTGCCCGGCGAGACACTGGTCAAGAGCAGAGAAGTGCTGCGGAAGGCCAG

AGAGGACAATCTGAAACTGATGAACCAGAAGCTCAACTTCCTGCGGAACGTGCTGCACT

TCCAGCAGTTCGAGGACATCACCGAGAGAGAGAAGCGGGTCACCAAGTGGATCAGCAGA

CAAGAGAACAGCGACGTGCCCCTGGTGTACCAGGATGAGCTGATCCAGATCCGCGAGCT

GATGTACAAGCCTTACAAGGACTGGGTCGCCTTCCTGAAGCAGCTCCACAAGAGACTGG

AAGTCGAGATCGGCAAAGAAGTGAAGCACTGGCGGAAGTCCCTGAGCGACGGAAGAAAG

GGCCTGTACGGCATCTCCCTGAAGAACATCGACGAGATCGATCGGACCCGGAAGTTCCT

GCTGAGATGGTCCCTGAGGCCTACCGAACCTGGCGAAGTGCGTAGACTGGAACCCGGCC

AGAGATTCGCCATCGACCAGCTGAATCACCTGAACGCCCTGAAAGAAGATCGGCTGAAG

AAGATGGCCAACACCATCATCATGCACGCCCTGGGCTACTGCTACGACGTGCGGAAGAA

GAAATGGCAGGCTAAGAACCCCGCCTGCCAGATCATCCTGTTCGAGGATCTGAGCAACT

ACAACCCCTACGAGGAAAGGTCCCGCTTCGAGAACAGCAAGCTCATGAAGTGGTCCAGA

CGCGAGATCCCCAGACAGGTTGCACTGCAGGGCGAGATCTATGGCCTGCAAGTGGGAGA

AGTGGGCGCTCAGTTCAGCAGCAGATTCCACGCCAAGACAGGCAGCCCTGGCATCAGAT

[0911] GTAGCGTCGTGACCAAAGAGAAGCTGCAGGACAATCGGTTCTTCAAGAATCTGCAGAGA

GAGGGCAGACTGACCCTGGACAAAATCGCCGTGCTGAAAGAGGGCGATCTGTACCCAGA

CAAAGGCGGCGAGAAGTTCATCAGCCTGAGCAAGGATCGGAAGTGCGTGACCACACACG

CCGACATCAACGCCGCTCAGAACCTGCAGAAGCGGTTCTGGACAAGAACCCACGGCTTC

TACAAGGTGTACTGCAAGGCCTACCAGGTGGACGGCCAGACCGTGTACATCCCTGAGAG

CAAGGACCAGAAGCAGAAGATCATCGAAGAGTTCGGCGAGGGCTACTTCATTCTGAAGG

ACGGGGTGTACGAATGGGTCAACGCCGGCAAGCTGAAAATCAAGAAGGGCAGCTCCAAG

CAGAGCAGCAGCGAGCTGGTGGATAGCGACATCCTGAAAGACAGCTTCGACCTGGCCTC

CGAGCTGAAAGGCGAAAAGCTGATGCTGTACAGGGACCCCAGCGGCAATGTGTTCCCCA

GCGACAAATGGATGGCCGCTGGCGTGTTCTTCGGAAAGCTGGAACGCATCCTGATCAGC

AAGCTGACCAACCAGTACTCCATCAGCACCATCGAGGACGACAGCAGCAAGCAGTCTAT

GAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGGATCC*TACC*

CATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATAT

*GATGTCCCCGACTATGCC*TAA

[0912] MAPKKKRKVGIHGVPAAATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYEHHEQD
PKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVEKKGEANQLSNKFLYPLVDPNSQ
SGKGTASSGRKPRWYNLKIAGDPSWEEEKKKWEEDKKKDPLAKILGKLAEYGLIPLFIPYTDSNEPIVKEIKWMEKS
RNQSVRRLDKDMFIQALERFLSWESWNLKVKEEYEKVEKEYKTLEERIKGGSGGSSEVEFSHEYWMRHALTLAKRAR
DEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIG
RVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGT
SESATPESSGEDIQALKALEQYEKERQEQLLRDTLNTNEYRLSKRGLRGWREIIQKWLKMDENEPSEKYLEVFKDYQ
RKHPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYATFCEIDKKKKDAKQQATFTLADPINHPLWVRFEERSGSNLN
KYRILTEQLHTEKLKKKLTVQLDRLIYPTESGGWEEKGKVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPL

KGTLGGARVQFDRDHLRRYPHKVESGNVGRIYFNMTVNIEPTESPVSKSLKIHRDDFPKVVNFKPKELTEWIKDSKG
KKLKSGIESLEIGLRVMSIDLGQRQAAAASIFEVVDQKPDIEGKLFFPIKGTELYAVHRASFNIKLPGETLVKSREV
LRKAREDNLKLMNQKLNFLRNVLHFQQFEDITEREKRVTKWISRQENSDVPLVYQDELIQIRELMYKPYKDWVAFLK
QLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDEIDRTRKFLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHLN
ALKEDRLKKMANTIIMHALGYCYDVRKKKWQAKNPACQIILFEDLSNYNPYEERSRFENSKLMKWSRREIPRQVALQ
GEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKLQDNRFFKNLQREGRLTLDKIAVLKEGDLYPDKGGEKFIS
LSKDRKCVTTHADINAAQNLQKRFWTRTHGFYKVYCKAYQVDGQTVYIPESKDQKQKIIEEFGEGYFILKDGVYEWV
NAGKLKIKKGSSKQSSSELVDSDILKDSFDLASELKGEKLMLYRDPSGNVFPSDKWMAAGVFFGKLERILISKLTNQ
YSISTIEDDSSKQSMKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA

[0913] BhCas12bGGSGGS-ABE8-Xten20在D306处

GCCACCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGC
CACCAGATCCTTCATCCTGAAGATCGAGCCCAACGAGGAAGTGAAGAAAGGCCTCTGGA
AAACCCACGAGGTGCTGAACCACGGAATCGCCTACTACATGAATATCCTGAAGCTGATC
CGGCAAGAGGCCATCTACGAGCACCACGAGCAGGACCCCAAGAATCCCAAGAAGGTGTC
CAAGGCCGAGATCCAGGCCGAGCTGTGGGATTTCGTGCTGAAGATGCAGAAGTGCAACA
GCTTCACACACGAGGTGGACAAGGACGAGGTGTTCAACATCCTGAGAGAGCTGTACGAG
GAACTGGTGCCCAGCAGCGTGGAAAAGAAGGGCGAAGCCAACCAGCTGAGCAACAAGTT

[0914] TCTGTACCCTCTGGTGGACCCCAACAGCCAGTCTGGAAAGGGAACAGCCAGCAGCGGCA
GAAAGCCCAGATGGTACAACCTGAAGATTGCCGGCGATCCCTCCTGGGAAGAAGAGAAG
AAGAAGTGGGAAGAAGATAAGAAAAAGGACCCGCTGGCCAAGATCCTGGGCAAGCTGGC
TGAGTACGGACTGATCCCTCTGTTCATCCCCTACACCGACAGCAACGAGCCCATCGTGA
AAGAAATCAAGTGGATGGAAAAGTCCCGGAACCAGAGCGTGCGGCGGCTGGATAAGGAC
ATGTTCATTCAGGCCCTGGAACGGTTCCTGAGCTGGGAGAGCTGGAACCTGAAAGTGAA
AGAGGAATACGAGAAGGTCGAGAAAGAGTACAAGACCCTGGAAGAGAGGATCAAAGAGG
ACATCCAGGCTCTGAAGGCTCTGGAACAGTATGAGAAAGAGCGGCAAGAACAGCTGCTG

CGGGACACCCTGAACACCAACGAGTACCGGCTGAGCAAGAGAGGCCTTAGAGGCTGGCG

GGAAATCATCCAGAAATGGCTGAAAATGGACggaggctctggaggaagcTCCGAAGTCG

AGTTTTCCCATGAGTACTGGATGAGACACGCATTGACTCTCGCAAAGAGGGCTCGAGAT

GAACGCGAGGTGCCCGTGGGGGCAGTACTCGTGCTCAACAATCGCGTAATCGGCGAAGG

TTGGAATAGGGCAATCGGACTCCACGACCCCACTGCACATGCGGAAATCATGGCCCTTC

GACAGGGAGGGCTTGTGATGCAGAATTATCGACTTTATGATGCGACGCTGTACGTCACG

TTTGAACCTTGCGTAATGTGCGCGGGAGCTATGATTCACTCCCGCATTGGACGAGTTGT

ATTCGGTGTTCGCAACGCCAAGACGGGTGCCGCAGGTTCACTGATGGACGTGCTGCATC

ATCCAGGCATGAACCACCGGGTAGAAATCACAGAAGGCATATTGGCGGACGAATGTGCG

GCGCTGTTGTGTCGTTTTTTTCGCATGCCCAGGCGGGTCTTTAACGCCCAGAAAAAAGC

ACAATCCTCTACTGACGGCTCTTCTGGATCTGAAACACCTGGCACAAGCGAGAGCGCCA

CCCCTGAGAGCTCTGGCGAGAACGAGCCCTCCGAGAAGTACCTGGAAGTGTTCAAGGAC

TACCAGCGGAAGCACCCTAGAGAGGCCGGCGATTACAGCGTGTACGAGTTCCTGTCCAA

GAAAGAGAACCACTTCATCTGGCGGAATCACCCTGAGTACCCCTACCTGTACGCCACCT

TCTGCGAGATCGACAAGAAAAAGAAGGACGCCAAGCAGCAGGCCACCTTCACACTGGCC

GATCCTATCAATCACCCTCTGTGGGTCCGATTCGAGGAAAGAAGCGGCAGCAACCTGAA

CAAGTACAGAATCCTGACCGAGCAGCTGCACACCGAGAAGCTGAAGAAAAAGCTGACAG

[0915] TGCAGCTGGACCGGCTGATCTACCCTACAGAATCTGGCGGCTGGGAAGAGAAGGGCAAA

GTGGACATTGTGCTGCTGCCCAGCCGGCAGTTCTACAACCAGATCTTCCTGGACATCGA

GGAAAAGGGCAAGCACGCCTTCACCTACAAGGATGAGAGCATCAAGTTCCCTCTGAAGG

GCACACTCGGCGGAGCCAGAGTGCAGTTCGACAGAGATCACCTGAGAAGATACCCTCAC

AAGGTGGAAAGCGGCAACGTGGGCAGAATCTACTTCAACATGACCGTGAACATCGAGCC

TACAGAGTCCCCAGTGTCCAAGTCTCTGAAGATCCACCGGGACGACTTCCCCAAGGTGG

TCAACTTCAAGCCCAAAGAACTGACCGAGTGGATCAAGGACAGCAAGGGCAAGAAACTG

AAGTCCGGCATCGAGTCCCTGGAAATCGGCCTGAGAGTGATGAGCATCGACCTGGGACA

GAGACAGGCCGCTGCCGCCTCTATTTTCGAGGTGGTGGATCAGAAGCCCGACATCGAAG

GCAAGCTGTTTTTCCCAATCAAGGGCACCGAGCTGTATGCCGTGCACAGAGCCAGCTTC

AACATCAAGCTGCCCGGCGAGACACTGGTCAAGAGCAGAGAAGTGCTGCGGAAGGCCAG

AGAGGACAATCTGAAACTGATGAACCAGAAGCTCAACTTCCTGCGGAACGTGCTGCACT

TCCAGCAGTTCGAGGACATCACCGAGAGAGAGAAGCGGGTCACCAAGTGGATCAGCAGA

CAAGAGAACAGCGACGTGCCCCTGGTGTACCAGGATGAGCTGATCCAGATCCGCGAGCT

GATGTACAAGCCTTACAAGGACTGGGTCGCCTTCCTGAAGCAGCTCCACAAGAGACTGG

AAGTCGAGATCGGCAAAGAAGTGAAGCACTGGCGGAAGTCCCTGAGCGACGGAAGAAAG

GGCCTGTACGGCATCTCCCTGAAGAACATCGACGAGATCGATCGGACCCGGAAGTTCCT

GCTGAGATGGTCCCTGAGGCCTACCGAACCTGGCGAAGTGCGTAGACTGGAACCCGGCC

[0916]

AGAGATTCGCCATCGACCAGCTGAATCACCTGAACGCCCTGAAAGAAGATCGGCTGAAG

AAGATGGCCAACACCATCATCATGCACGCCCTGGGCTACTGCTACGACGTGCGGAAGAA

GAAATGGCAGGCTAAGAACCCCGCCTGCCAGATCATCCTGTTCGAGGATCTGAGCAACT

ACAACCCCTACGAGGAAAGGTCCCGCTTCGAGAACAGCAAGCTCATGAAGTGGTCCAGA

CGCGAGATCCCCAGACAGGTTGCACTGCAGGGCGAGATCTATGGCCTGCAAGTGGGAGA

AGTGGGCGCTCAGTTCAGCAGCAGATTCCACGCCAAGACAGGCAGCCCTGGCATCAGAT

GTAGCGTCGTGACCAAAGAGAAGCTGCAGGACAATCGGTTCTTCAAGAATCTGCAGAGA

GAGGGCAGACTGACCCTGGACAAAATCGCCGTGCTGAAAGAGGGCGATCTGTACCCAGA

CAAAGGCGGCGAGAAGTTCATCAGCCTGAGCAAGGATCGGAAGTGCGTGACCACACACG

CCGACATCAACGCCGCTCAGAACCTGCAGAAGCGGTTCTGGACAAGAACCCACGGCTTC

TACAAGGTGTACTGCAAGGCCTACCAGGTGGACGGCCAGACCGTGTACATCCCTGAGAG

CAAGGACCAGAAGCAGAAGATCATCGAAGAGTTCGGCGAGGGCTACTTCATTCTGAAGG

ACGGGGTGTACGAATGGGTCAACGCCGGCAAGCTGAAAATCAAGAAGGGCAGCTCCAAG

CAGAGCAGCAGCGAGCTGGTGGATAGCGACATCCTGAAAGACAGCTTCGACCTGGCCTC

CGAGCTGAAAGGCGAAAAGCTGATGCTGTACAGGGACCCCAGCGGCAATGTGTTCCCCA

GCGACAAATGGATGGCCGCTGGCGTGTTCTTCGGAAAGCTGGAACGCATCCTGATCAGC

AAGCTGACCAACCAGTACTCCATCAGCACCATCGAGGACGACAGCAGCAAGCAGTCTAT

GAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGGATCC*TACC*

CATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATAT

*GATGTCCCCGACTATGCC*TAA

[0917] MAPKKKRKVGIHGVPAAATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYEHHEQD
PKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVEKKGEANQLSNKFLYPLVDPNSQ
SGKGTASSGRKPRWYNLKIAGDPSWEEEKKKWEEDKKKDPLAKILGKLAEYGLIPLFIPYTDSNEPIVKEIKWMEKS
RNQSVRRLDKDMFIQALERFLSWESWNLKVKEEYEKVEKEYKTLEERIKEDIQALKALEQYEKERQEQLLRDTLNTN
EYRLSKRGLRGWREIIQKWLKMDGGSGGSSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAI
GLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGM
NHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESATPESSGENEPSEKYLEVFKDYQ
RKHPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYATFCEIDKKKKDAKQQATFTLADPINHPLWVRFEERSGSNLN
KYRILTEQLHTEKLKKKLTVQLDRLIYPTESGGWEEKGKVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPL
KGTLGGARVQFDRDHLRRYPHKVESGNVGRIYFNMTVNIEPTESPVSKSLKIHRDDFPKVVNFKPKELTEWIKDSKG
KKLKSGIESLEIGLRVMSIDLGQRQAAAASIFEVVDQKPDIEGKLFFPIKGTELYAVHRASFNIKLPGETLVKSREV
LRKAREDNLKLMNQKLNFLRNVLHFQQFEDITEREKRVTKWISRQENSDVPLVYQDELIQIRELMYKPYKDWVAFLK
QLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDEIDRTRKFLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHLN
ALKEDRLKKMANTIIMHALGYCYDVRKKKWQAKNPACQIILFEDLSNYNPYEERSRFENSKLMKWSRREIPRQVALQ
GEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKLQDNRFFKNLQREGRLTLDKIAVLKEGDLYPDKGGEKFIS
LSKDRKCVTTHADINAAQNLQKRFWTRTHGFYKVYCKAYQVDGQTVYIPESKDQKQKIIEEFGEGYFILKDGVYEWV
NAGKLKIKKGSSKQSSSELVDSDILKDSFDLASELKGEKLMLYRDPSGNVFPSDKWMAAGVFFGKLERILISKLTNQ

YSISTIEDDSSKQSMKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA

[0918] BhCas12bGGSGGS-ABE8-Xten20在D980处

GCCACCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGC

CACCAGATCCTTCATCCTGAAGATCGAGCCCAACGAGGAAGTGAAGAAAGGCCTCTGGA

AAACCCACGAGGTGCTGAACCACGGAATCGCCTACTACATGAATATCCTGAAGCTGATC

CGGCAAGAGGCCATCTACGAGCACCACGAGCAGGACCCCAAGAATCCCAAGAAGGTGTC

CAAGGCCGAGATCCAGGCCGAGCTGTGGGATTTCGTGCTGAAGATGCAGAAGTGCAACA

GCTTCACACACGAGGTGGACAAGGACGAGGTGTTCAACATCCTGAGAGAGCTGTACGAG

GAACTGGTGCCCAGCAGCGTGGAAAAGAAGGGCGAAGCCAACCAGCTGAGCAACAAGTT

TCTGTACCCTCTGGTGGACCCCAACAGCCAGTCTGGAAAGGGAACAGCCAGCAGCGGCA

GAAAGCCCAGATGGTACAACCTGAAGATTGCCGGCGATCCCTCCTGGGAAGAAGAGAAG

AAGAAGTGGGAAGAAGATAAGAAAAAGGACCCGCTGGCCAAGATCCTGGGCAAGCTGGC

TGAGTACGGACTGATCCCTCTGTTCATCCCCTACACCGACAGCAACGAGCCCATCGTGA

[0919]

AAGAAATCAAGTGGATGGAAAAGTCCCGGAACCAGAGCGTGCGGCGGCTGGATAAGGAC

ATGTTCATTCAGGCCCTGGAACGGTTCCTGAGCTGGGAGAGCTGGAACCTGAAAGTGAA

AGAGGAATACGAGAAGGTCGAGAAAGAGTACAAGACCCTGGAAGAGAGGATCAAAGAGG

ACATCCAGGCTCTGAAGGCTCTGGAACAGTATGAGAAAGAGCGGCAAGAACAGCTGCTG

CGGGACACCCTGAACACCAACGAGTACCGGCTGAGCAAGAGAGGCCTTAGAGGCTGGCG

GGAAATCATCCAGAAATGGCTGAAAATGGACGAGAACGAGCCCTCCGAGAAGTACCTGG

AAGTGTTCAAGGACTACCAGCGGAAGCACCCTAGAGAGGCCGGCGATTACAGCGTGTAC

GAGTTCCTGTCCAAGAAAGAGAACCACTTCATCTGGCGGAATCACCCTGAGTACCCCTA

CCTGTACGCCACCTTCTGCGAGATCGACAAGAAAAAGAAGGACGCCAAGCAGCAGGCCA

CCTTCACACTGGCCGATCCTATCAATCACCCTCTGTGGGTCCGATTCGAGGAAAGAAGC

GGCAGCAACCTGAACAAGTACAGAATCCTGACCGAGCAGCTGCACACCGAGAAGCTGAA

GAAAAAGCTGACAGTGCAGCTGGACCGGCTGATCTACCCTACAGAATCTGGCGGCTGGG

AAGAGAAGGGCAAAGTGGACATTGTGCTGCTGCCCAGCCGGCAGTTCTACAACCAGATC

TTCCTGGACATCGAGGAAAAGGGCAAGCACGCCTTCACCTACAAGGATGAGAGCATCAA

GTTCCCTCTGAAGGGCACACTCGGCGGAGCCAGAGTGCAGTTCGACAGAGATCACCTGA

GAAGATACCCTCACAAGGTGGAAAGCGGCAACGTGGGCAGAATCTACTTCAACATGACC

GTGAACATCGAGCCTACAGAGTCCCCAGTGTCCAAGTCTCTGAAGATCCACCGGGACGA

CTTCCCCAAGGTGGTCAACTTCAAGCCCAAAGAACTGACCGAGTGGATCAAGGACAGCA

AGGGCAAGAAACTGAAGTCCGGCATCGAGTCCCTGGAAATCGGCCTGAGAGTGATGAGC

ATCGACCTGGGACAGAGACAGGCCGCTGCCGCCTCTATTTTCGAGGTGGTGGATCAGAA

GCCCGACATCGAAGGCAAGCTGTTTTTCCCAATCAAGGGCACCGAGCTGTATGCCGTGC

ACAGAGCCAGCTTCAACATCAAGCTGCCCGGCGAGACACTGGTCAAGAGCAGAGAAGTG

CTGCGGAAGGCCAGAGAGGACAATCTGAAACTGATGAACCAGAAGCTCAACTTCCTGCG

GAACGTGCTGCACTTCCAGCAGTTCGAGGACATCACCGAGAGAGAGAAGCGGGTCACCA

AGTGGATCAGCAGACAAGAGAACAGCGACGTGCCCCTGGTGTACCAGGATGAGCTGATC

CAGATCCGCGAGCTGATGTACAAGCCTTACAAGGACTGGGTCGCCTTCCTGAAGCAGCT

CCACAAGAGACTGGAAGTCGAGATCGGCAAAGAAGTGAAGCACTGGCGGAAGTCCCTGA

GCGACGGAAGAAAGGGCCTGTACGGCATCTCCCTGAAGAACATCGACGAGATCGATCGG

[0920] ACCCGGAAGTTCCTGCTGAGATGGTCCCTGAGGCCTACCGAACCTGGCGAAGTGCGTAG

ACTGGAACCCGGCCAGAGATTCGCCATCGACCAGCTGAATCACCTGAACGCCCTGAAAG

AAGATCGGCTGAAGAAGATGGCCAACACCATCATCATGCACGCCCTGGGCTACTGCTAC

GACGTGCGGAAGAAGAAATGGCAGGCTAAGAACCCCGCCTGCCAGATCATCCTGTTCGA

GGATCTGAGCAACTACAACCCCTACGAGGAAAGGTCCCGCTTCGAGAACAGCAAGCTCA

TGAAGTGGTCCAGACGCGAGATCCCCAGACAGGTTGCACTGCAGGGCGAGATCTATGGC

CTGCAAGTGGGAGAAGTGGGCGCTCAGTTCAGCAGCAGATTCCACGCCAAGACAGGCAG

CCCTGGCATCAGATGTAGCGTCGTGACCAAAGAGAAGCTGCAGGACAATCGGTTCTTCA

AGAATCTGCAGAGAGAGGGCAGACTGACCCTGGACAAAATCGCCGTGCTGAAAGAGGGC

GATCTGTACCCAGACAAAGGCGGCGAGAAGTTCATCAGCCTGAGCAAGGATCGGAAGTG

CGTGACCACACACGCCGACATCAACGCCGCTCAGAACCTGCAGAAGCGGTTCTGGACAA

GAACCCACGGCTTCTACAAGGTGTACTGCAAGGCCTACCAGGTGGACggaggctctgga

ggaagcTCCGAAGTCGAGTTTTCCCATGAGTACTGGATGAGACACGCATTGACTCTCGC

AAAGAGGGCTCGAGATGAACGCGAGGTGCCCGTGGGGGCAGTACTCGTGCTCAACAATC

GCGTAATCGGCGAAGGTTGGAATAGGGCAATCGGACTCCACGACCCCACTGCACATGCG

GAAATCATGGCCCTTCGACAGGGAGGGCTTGTGATGCAGAATTATCGACTTTATGATGC

GACGCTGTACGTCACGTTTGAACCTTGCGTAATGTGCGCGGGAGCTATGATTCACTCCC

GCATTGGACGAGTTGTATTCGGTGTTCGCAACGCCAAGACGGGTGCCGCAGGTTCACTG

ATGGACGTGCTGCATCATCCAGGCATGAACCACCGGGTAGAAATCACAGAAGGCATATT
GGCGGACGAATGTGCGGCGCTGTTGTGTCGTTTTTTTCGCATGCCCAGGCGGGTCTTTA
ACGCCCAGAAAAAAGCACAATCCTCTACTGACGGCTCTTCTGGATCTGAAACACCTGGC
ACAAGCGAGAGCGCCACCCCTGAGAGCTCTGGCGGCCAGACCGTGTACATCCCTGAGAG
CAAGGACCAGAAGCAGAAGATCATCGAAGAGTTCGGCGAGGGCTACTTCATTCTGAAGG
ACGGGGTGTACGAATGGGTCAACGCCGGCAAGCTGAAAATCAAGAAGGGCAGCTCCAAG

[0921] CAGAGCAGCAGCGAGCTGGTGGATAGCGACATCCTGAAAGACAGCTTCGACCTGGCCTC
CGAGCTGAAAGGCGAAAAGCTGATGCTGTACAGGGACCCCAGCGGCAATGTGTTCCCCA
GCGACAAATGGATGGCCGCTGGCGTGTTCTTCGGAAAGCTGGAACGCATCCTGATCAGC
AAGCTGACCAACCAGTACTCCATCAGCACCATCGAGGACGACAGCAGCAAGCAGTCTAT
GAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGGATCC*TACC*
CATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATAT
*GATGTCCCCGACTATGCC*TAA

[0922] MAPKKKRKVGIHGVPAAATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYEHHEQD
PKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVEKKGEANQLSNKFLYPLVDPNSQ
SGKGTASSGRKPRWYNLKIAGDPSWEEEKKKWEEDKKKDPLAKILGKLAEYGLIPLFIPYTDSNEPIVKEIKWMEKS
RNQSVRRLDKDMFIQALERFLSWESWNLKVKEEYEKVEKEYKTLEERIKEDIQALKALEQYEKERQEQLLRDTLNTN
EYRLSKRGLRGWREIIQKWLKMDENEPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYATF
CEIDKKKKDAKQQATFTLADPINHPLWVRFEERSGSNLNKYRILTEQLHTEKLKKKLTVQLDRLIYPTESGGWEEKG
KVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGTLGGARVQFDRDHLRRYPHKVESGNVGRIYFNMTVN
IEPTESPVSKSLKIHRDDFPKVVNFKPKELTEWIKDSKGKKLKSGIESLEIGLRVMSIDLGQRQAAAASIFEVVDQK
PDIEGKLFFPIKGTELYAVHRASFNIKLPGETLVKSREVLRKAREDNLKLMNQKLNFLRNVLHFQQFEDITEREKRV
TKWISRQENSDVPLVYQDELIQIRELMYKPYKDWVAFLKQLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDE
IDRTRKFLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHLNALKEDRLKKMANTIIMHALGYCYDVRKKKWQAKNPACQ
IILFEDLSNYNPYEERSRFENSKLMKWSRREIPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKL
QDNRFFKNLQREGRLTLDKIAVLKEGDLYPDKGGEKFISLSKDRKCVTTHADINAAQNLQKRFWTRTHGFYKVYCKA
YQVDGGSGGSSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGL
VMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALL
CRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESATPESSGGQTVYIPESKDQKQKIIEEFGEGYFILKDGVYEWV
NAGKLKIKKGSSKQSSSELVDSDILKDSFDLASELKGEKLMLYRDPSGNVFPSDKWMAAGVFFGKLERILISKLTNQ
YSISTIEDDSSKQSMKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA

[0923] BhCas12b GGSGGS-ABE8-Xten20在K1019处

GCCACCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGC

CACCAGATCCTTCATCCTGAAGATCGAGCCCAACGAGGAAGTGAAGAAAGGCCTCTGGA

AAACCCACGAGGTGCTGAACCACGGAATCGCCTACTACATGAATATCCTGAAGCTGATC

CGGCAAGAGGCCATCTACGAGCACCACGAGCAGGACCCCAAGAATCCCAAGAAGGTGTC

CAAGGCCGAGATCCAGGCCGAGCTGTGGGATTTCGTGCTGAAGATGCAGAAGTGCAACA

GCTTCACACACGAGGTGGACAAGGACGAGGTGTTCAACATCCTGAGAGAGCTGTACGAG

GAACTGGTGCCCAGCAGCGTGGAAAAGAAGGGCGAAGCCAACCAGCTGAGCAACAAGTT

TCTGTACCCTCTGGTGGACCCCAACAGCCAGTCTGGAAAGGGAACAGCCAGCAGCGGCA

GAAAGCCCAGATGGTACAACCTGAAGATTGCCGGCGATCCCTCCTGGGAAGAAGAGAAG

AAGAAGTGGGAAGAAGATAAGAAAAAGGACCCGCTGGCCAAGATCCTGGGCAAGCTGGC

TGAGTACGGACTGATCCCTCTGTTCATCCCCTACACCGACAGCAACGAGCCCATCGTGA

AAGAAATCAAGTGGATGGAAAAGTCCCGGAACCAGAGCGTGCGGCGGCTGGATAAGGAC

ATGTTCATTCAGGCCCTGGAACGGTTCCTGAGCTGGGAGAGCTGGAACCTGAAAGTGAA

AGAGGAATACGAGAAGGTCGAGAAAGAGTACAAGACCCTGGAAGAGAGGATCAAAGAGG

[0924] ACATCCAGGCTCTGAAGGCTCTGGAACAGTATGAGAAAGAGCGGCAAGAACAGCTGCTG

CGGGACACCCTGAACACCAACGAGTACCGGCTGAGCAAGAGAGGCCTTAGAGGCTGGCG

GGAAATCATCCAGAAATGGCTGAAAATGGACGAGAACGAGCCCTCCGAGAAGTACCTGG

AAGTGTTCAAGGACTACCAGCGGAAGCACCCTAGAGAGGCCGGCGATTACAGCGTGTAC

GAGTTCCTGTCCAAGAAAGAGAACCACTTCATCTGGCGGAATCACCCTGAGTACCCCTA

CCTGTACGCCACCTTCTGCGAGATCGACAAGAAAAAGAAGGACGCCAAGCAGCAGGCCA

CCTTCACACTGGCCGATCCTATCAATCACCCTCTGTGGGTCCGATTCGAGGAAAGAAGC

GGCAGCAACCTGAACAAGTACAGAATCCTGACCGAGCAGCTGCACACCGAGAAGCTGAA

GAAAAAGCTGACAGTGCAGCTGGACCGGCTGATCTACCCTACAGAATCTGGCGGCTGGG

AAGAGAAGGGCAAAGTGGACATTGTGCTGCTGCCCAGCCGGCAGTTCTACAACCAGATC

TTCCTGGACATCGAGGAAAAGGGCAAGCACGCCTTCACCTACAAGGATGAGAGCATCAA

GTTCCCTCTGAAGGGCACACTCGGCGGAGCCAGAGTGCAGTTCGACAGAGATCACCTGA

GAAGATACCCTCACAAGGTGGAAAGCGGCAACGTGGGCAGAATCTACTTCAACATGACC

GTGAACATCGAGCCTACAGAGTCCCCAGTGTCCAAGTCTCTGAAGATCCACCGGGACGA

CTTCCCCAAGGTGGTCAACTTCAAGCCCAAAGAACTGACCGAGTGGATCAAGGACAGCA

AGGGCAAGAAACTGAAGTCCGGCATCGAGTCCCTGGAAATCGGCCTGAGAGTGATGAGC

ATCGACCTGGGACAGAGACAGGCCGCTGCCGCCTCTATTTTCGAGGTGGTGGATCAGAA

GCCCGACATCGAAGGCAAGCTGTTTTTCCCAATCAAGGGCACCGAGCTGTATGCCGTGC

ACAGAGCCAGCTTCAACATCAAGCTGCCCGGCGAGACACTGGTCAAGAGCAGAGAAGTG

CTGCGGAAGGCCAGAGAGGACAATCTGAAACTGATGAACCAGAAGCTCAACTTCCTGCG

GAACGTGCTGCACTTCCAGCAGTTCGAGGACATCACCGAGAGAGAGAAGCGGGTCACCA

AGTGGATCAGCAGACAAGAGAACAGCGACGTGCCCCTGGTGTACCAGGATGAGCTGATC

CAGATCCGCGAGCTGATGTACAAGCCTTACAAGGACTGGGTCGCCTTCCTGAAGCAGCT

CCACAAGAGACTGGAAGTCGAGATCGGCAAAGAAGTGAAGCACTGGCGGAAGTCCCTGA

GCGACGGAAGAAAGGGCCTGTACGGCATCTCCCTGAAGAACATCGACGAGATCGATCGG

ACCCGGAAGTTCCTGCTGAGATGGTCCCTGAGGCCTACCGAACCTGGCGAAGTGCGTAG

ACTGGAACCCGGCCAGAGATTCGCCATCGACCAGCTGAATCACCTGAACGCCCTGAAAG

AAGATCGGCTGAAGAAGATGGCCAACACCATCATCATGCACGCCCTGGGCTACTGCTAC

GACGTGCGGAAGAAGAAATGGCAGGCTAAGAACCCCGCCTGCCAGATCATCCTGTTCGA

GGATCTGAGCAACTACAACCCCTACGAGGAAAGGTCCCGCTTCGAGAACAGCAAGCTCA

TGAAGTGGTCCAGACGCGAGATCCCCAGACAGGTTGCACTGCAGGGCGAGATCTATGGC

CTGCAAGTGGGAGAAGTGGGCGCTCAGTTCAGCAGCAGATTCCACGCCAAGACAGGCAG

[0925] CCCTGGCATCAGATGTAGCGTCGTGACCAAAGAGAAGCTGCAGGACAATCGGTTCTTCA

AGAATCTGCAGAGAGAGGGCAGACTGACCCTGGACAAAATCGCCGTGCTGAAAGAGGGC

GATCTGTACCCAGACAAAGGCGGCGAGAAGTTCATCAGCCTGAGCAAGGATCGGAAGTG

CGTGACCACACACGCCGACATCAACGCCGCTCAGAACCTGCAGAAGCGGTTCTGGACAA

GAACCCACGGCTTCTACAAGGTGTACTGCAAGGCCTACCAGGTGGACGGCCAGACCGTG

TACATCCCTGAGAGCAAGGACCAGAAGCAGAAGATCATCGAAGAGTTCGGCGAGGGCTA

CTTCATTCTGAAGGACGGGGTGTACGAATGGGTCAACGCCGGCAAGggaggctctggag

gaagcTCCGAAGTCGAGTTTTCCCATGAGTACTGGATGAGACACGCATTGACTCTCGCA

AAGAGGGCTCGAGATGAACGCGAGGTGCCCGTGGGGGCAGTACTCGTGCTCAACAATCG

CGTAATCGGCGAAGGTTGGAATAGGGCAATCGGACTCCACGACCCCACTGCACATGCGG

AAATCATGGCCCTTCGACAGGGAGGGCTTGTGATGCAGAATTATCGACTTTATGATGCG

ACGCTGTACGTCACGTTTGAACCTTGCGTAATGTGCGCGGGAGCTATGATTCACTCCCG

CATTGGACGAGTTGTATTCGGTGTTCGCAACGCCAAGACGGGTGCCGCAGGTTCACTGA

TGGACGTGCTGCATCATCCAGGCATGAACCACCGGGTAGAAATCACAGAAGGCATATTG

GCGGACGAATGTGCGGCGCTGTTGTGTCGTTTTTTTCGCATGCCCAGGCGGGTCTTTAA

CGCCCAGAAAAAAGCACAATCCTCTACTGACGGCTCTTCTGGATCTGAAACACCTGGCA

CAAGCGAGAGCGCCACCCCTGAGAGCTCTGGCCTGAAAATCAAGAAGGGCAGCTCCAAG

CAGAGCAGCAGCGAGCTGGTGGATAGCGACATCCTGAAAGACAGCTTCGACCTGGCCTC

[0926] CGAGCTGAAAGGCGAAAAGCTGATGCTGTACAGGGACCCCAGCGGCAATGTGTTCCCCA
GCGACAAATGGATGGCCGCTGGCGTGTTCTTCGGAAAGCTGGAACGCATCCTGATCAGC
AAGCTGACCAACCAGTACTCCATCAGCACCATCGAGGACGACAGCAGCAAGCAGTCTAT
GAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGGATCC*TACC*
CATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATAT
*GATGTCCCCGACTATGCC*TAA

[0927] MAPKKKRKVGIHGVPAAATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYEHHEQD
PKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVEKKGEANQLSNKFLYPLVDPNSQ
SGKGTASSGRKPRWYNLKIAGDPSWEEEKKKWEEDKKKDPLAKILGKLAEYGLIPLFIPYTDSNEPIVKEIKWMEKS
RNQSVRRLDKDMFIQALERFLSWESWNLKVKEEYEKVEKEYKTLEERIKEDIQALKALEQYEKERQEQLLRDTLNTN
EYRLSKRGLRGWREIIQKWLKMDENEPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYATF
CEIDKKKKDAKQQATFTLADPINHPLWVRFEERSGSNLNKYRILTEQLHTEKLKKKLTVQLDRLIYPTESGGWEEKG
KVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGTLGGARVQFDRDHLRRYPHKVESGNVGRIYFNMTVN
IEPTESPVSKSLKIHRDDFPKVVNFKPKELTEWIKDSKGKKLKSGIESLEIGLRVMSIDLGQRQAAAASIFEVVDQK
PDIEGKLFFPIKGTELYAVHRASFNIKLPGETLVKSREVLRKAREDNLKLMNQKLNFLRNVLHFQQFEDITEREKRV
TKWISRQENSDVPLVYQDELIQIRELMYKPYKDWVAFLKQLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDE
IDRTRKFLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHLNALKEDRLKKMANTIIMHALGYCYDVRKKKWQAKNPACQ
IILFEDLSNYNPYEERSRFENSKLMKWSRREIPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKL
QDNRFFKNLQREGRLTLDKIAVLKEGDLYPDKGGEKFISLSKDRKCVTTHADINAAQNLQKRFWTRTHGFYKVYCKA
YQVDGQTVYIPESKDQKQKIIEEFGEGYFILKDGVYEWVNAGKGGSGGSSEVEFSHEYWMRHALTLAKRARDEREVP
VGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGV
RNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESATP
ESSGLKIKKGSSKQSSSELVDSDILKDSFDLASELKGEKLMLYRDPSGNVFPSDKWMAAGVFFGKLERILISKLTNQ
YSISTIEDDSSKQSMKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYAYPYDVPDYA

[0928] 对于上述序列,Kozak序列加粗并加下划线;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,未经修饰的序列编码BhCas12b;双下划线表示Xten20连接子;单下划线表示C端NLS;GGATCC表示GS连接子;且斜体字母代表3x血细胞凝集素(HA)标签的编码序列。

[0929] 在一些实施方案中,本文所提供的任何融合蛋白的核酸可编程DNA结合蛋白(napDNAbp)可以是Cas12j/CasΦ蛋白。Cas12j/CasΦ在Pausch等人,"CRISPR-CasΦ from huge phages is a hypercompact genome editor,"Science,17July 2020,Vol.369,Issue 6501,pp.333-337之描述,其通过引用以其整体并入本文。在一些实施方案中,napDNAbp包含氨基酸序列,该氨基酸序列与天然存在的Cas12j/CasΦ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Cas12j/CasΦ蛋白。在一些实施方案中,napDNAbp是无核酸酶活性的("死亡的")Cas12j/CasΦ蛋白。应了解,来自其他物种的Cas12j/CasΦ也可根据本公开使用。

[0930] 示例性Cas12j/CasΦ氨基酸序列如下:

[0931] >CasΦ-1

[0932] MADTPTLFTQFLRHHLPGQRFRKDILKQAGRILANKGEDATIAFLRGKSEESPPDFQPPVKCPIIACS
RPLTEWPIYQASVAIQGYVYGQSLAEFEASDPGCSKDGLLGWFDKTGVCTDYFSVQGLNLIFQNARKRYIGVQTKV
TNRNEKRHKKLKRINAKRIAEGLPELTSDEPESALDETGHLIDPPGLNTNIYCYQQVSPKPLALSEVNQLPTAYAG
YSTSGDDPIQPMVTKDRLSISKGQPGYIPEHQRALLSQKKHRRMRGYGLKARALLVIVRIQDDWAVIDLRSLLRNA
YWRRIVQTKEPSTITKLLKLVTGDPVLDATRMVATFTYKPGIVQVRSAKCLKNKQGSKLFSERYLNETVSVTSIDL
GSNNLVAVATYRLVNGNTPELLQRFTLPSHLVKDFERYKQAHDTLEDSIQKTAVASLPQGQQTEIRMWSMYGFREA
QERVCQELGLADGSIPWNVMTATSTILTDLFLARGGDPKKCMFTSEPKKKKNSKQVLYKIRDRAWAKMYRTLLSKE
TREAWNKALWGLKRGSPDYARLSKRKEELARRCVNYTISTAEKRAQCGRTIVALEDLNIGFFHGRGKQEPGWVGLF
TRKKENRWLMQALHKAFLELAHHRGYHVIEVNPAYTSQTCPVCRHCDPDNRDQHNREAFHCIGCGFRGNADLDVAT
HNIAMVAITGESLKRARGSVASKTPQPLAAE*

[0933] >CasΦ-2

[0934] MPKPAVESEFSKVLKKHFPGERFRSSYMKRGGKILAAQGEEAVVAYLQGKSEEEPPNFQPPAKCHVVTK
SRDFAEWPIMKASEAIQRYIYALSTTERAACKPGKSSESHAAWFAATGVSNHGYSHVQGLNLIFDHTLGRYDGVLKK
VQLRNEKARARLESINASRADEGLPEIKAEEEEVATNETGHLLQPPGINPSFYVYQTISPQAYRPRDEIVLPPEYAG
YVRDPNAPIPLGVVRNRCDIQKGCPGYIPEWQREAGTAISPKTGKAVTVPGLSPKKNKRMRRYWRSEKEKAQDALLV
TVRIGTDWVVIDVRGLLRNARWRTIAPKDISLNALLDLFTGDPVIDVRRNIVTFTYTLDACGTYARKWTLKGKQTKA
TLDKLTATQTVALVAIDLGQTNPISAGISRVTQENGALQCEPLDRFTLPDDLLKDISAYRIAWDRNEEELRARSVEA
LPEAQQAEVRALDGVSKETARTQLCADFGLDPKRLPWDKMSSNTTFISEALLSNSVSRDQVFFTPAPKKGAKKKAPV
EVMRKDRTWARAYKPRLSVEAQKLKNEALWALKRTSPEYLKLSRRKEELCRRSINYVIEKTRRRTQCQIVIPVIEDL
NVRFFHGSGKRLPGWDNFFTAKKENRWFIQGLHKAFSDLRTHRSFYVFEVRPERTSITCPKCGHCEVGNRDGEAFQC
LSCGKTCNADLDVATHNLTQVALTGKTMPKREEPRDAQGTAPARKTKKASKSKAPPAEREDQTPAQEPSQTS

[0935] >CasΦ-3

[0936] MEKEITELTKIRREFPNKKFSSTDMKKAGKLLKAEGPDAVRDFLNSCQEIIGDFKPPVKTNIVSISRP
FEEWPVSMVGRAIQEYYFSLTKEELESVHPGTSSEDHKSFFNITGLSNYNYTSVQGLNLIFKNAKAIYDGTLVKAN
NKNKKLEKKFNEINHKRSLEGLPIITPDFEEPFDENGHLNNPPGINRNIYGYQGCAAKVFVPSKHKMVSLPKEYEG
YNRDPNLSLAGFRNRLEIPEGEPGHVPWFQRMDIPEGQIGHVNKIQRFNFVHGKNSGKVKFSDKTGRVKRYHHSKY
KDATKPYKFLEESKKVSALDSILAIITIGDDWVVFDIRGLYRNVFYRELAQKGLTAVQLLDLFTGDPVIDPKKGVV
TFSYKEGVVPVFSQKIVPRFKSRDTLEKLTSQGPVALLSVDLGQNEPVAARVCSLKNINDKITLDNSCRISFLDDY
KKQIKDYRDSLDELEIKIRLEAINSLETNQQVEIRDLDVFSADRAKANTVDMFDIDPNLISWDSMSDARVSTQISD
LYLKNGGDESRVYFEINNKRIKRSDYNISQLVRPKLSDSTRKNLNDSIWKLKRTSEEYLKLSKRKLELSRAVVNYT
IRQSKLLSGINDIVIILEDLDVKKKFNGRGIRDIGWDNFFSSRKENRWFIPAFHKAFSELSSNRGLCVIEVNPAWT
SATCPDCGFCSKENRDGINFTCRKCGVSYHADIDVATLNIARVAVLGKPMSGPADRERLGDTKKPRVARSRKTMKR
KDISNSTVEAMVTA*

[0937] >CasΦ-4

[0938] MYSLEMADLKSEPSLLAKLLRDRFPGKYWLPKYWKLAEKKRLTGGEEAACEYMADKQLDSPPPNFRPP
ARCVILAKSRPFEDWPVHRVASKAQSFVIGLSEQGFAALRAAPPSTADARRDWLRSHGASEDDLMALEAQLLETIM
GNAISLHGGVLKKIDNANVKAAKRLSGRNEARLNKGLQELPPEQEGSAYGADGLLVNPPGLNLNIYCRKSCCPKPV
KNTARFVGHYPGYLRDSDSILISGTMDRLTIIEGMPGHIPAWQREQGLVKPGGRRRRLSGSESNMRQKVDPSTGPR

RSTRSGTVNRSNQRTGRNGDPLLVEIRMKEDWVLLDARGLLRNLRWRESKRGLSCDHEDLSLSGLLALFSGDPVID
PVRNEVVFLYGEGIIPVRSTKPVGTRQSKKLLERQASMGPLTLISCDLGQTNLIAGRASAISLTHGSLGVRSSVRI
ELDPEIIKSFERLRKDADRLETEILTAAKETLSDEQRGEVNSHEKDSPQTAKASLCRELGLHPPSLPWGQMGPSTT
FIADMLISHGRDDDAFLSHGEFPTLEKRKKFDKRFCLESRPLLSSETRKALNESLWEVKRTSSEYARLSQRKKEMA
RRAVNFVVEISRRKTGLSNVIVNIEDLNVRIFHGGGKQAPGWDGFFRPKSENRWFIQAIHKAFSDLAAHHGIPVIE
SDPQRTSMTCPECGHCDSKNRNGVRFLCKGCGASMDADFDAACRNLERVALTGKPMPKPSTSCERLLSATTGKVCS
DHSLSHDAIEKAS*

[0939] >CasΦ-5

[0940] MSSLPTPLELLKQKHADLFKGLQFSSKDNKMAGKVLKKDGEEAALAFLSERGVSRGELPNFRPPAKTL
VVAQSRPFEEFPIYRVSEAIQLYVYSLSVKELETVPSGSSTKKEHQRFFQDSSVPDFGYTSVQGLNKIFGLARGIY
LGVITRGENQLQKAKSKHEALNKKRRASGEAETEFDPTPYEYMTPERKLAKPPGVNHSIMCYVDISVDEFDFRNPD
GIVLPSEYAGYCREINTAIEKGTVDRLGHLKGGPGYIPGHQRKESTTEGPKINFRKGRIRRSYTALYAKRDSRRVR
QGKLALPSYRHHMMRLNSNAESAILAVIFFGKDWVVFDLRGLLRNVRWRNLFVDGSTPSTLLGMFGDPVIDPKRGV
VAFCYKEQIVPVVSKSITKMVKAPELLNKLYLKSEDPLVLVAIDLGQTNPVGVGVYRVMNASLDYEVVTRFALESE
LLREIESYRQRTNAFEAQIRAETFDAMTSEEQEEITRVRAFSASKAKENVCHRFGMPVDAVDWATMGSNTIHIAKW
VMRHGDPSLVEVLEYRKDNEIKLDKNGVPKKVKLTDKRIANLTSIRLRFSQETSKHYNDTMWELRRKHPVYQKLSK
SKADFSRRVVNSIIRRVNHLVPRARIVFIIEDLKNLGKVFHGSGKRELGWDSYFEPKSENRWFIQVLHKAFSETGK
HKGYYIIECWPNWTSCTCPKCSCCDSENRHGEVFRCLACGYTCNTDFGTAPDNLVKIATTGKGLPGPKKRCKGSSK
GKNPKIARSSETGVSVTESGAPKVKKSSPTQTSQSSSQSAP*

[0941] >CasΦ-6

[0942] MNKIEKEKTPLAKLMNENFAGLRFPFAIIKQAGKKLLKEGELKTIEYMTGKGSIEPLPNFKPPVKCLIV
AKRRDLKYFPICKASCEIQSYVYSLNYKDFMDYFSTPMTSQKQHEEFFKKSGLNIEYQNVAGLNLIFNNVKNTYNGV
ILKVKNRNEKLKKKAIKNNYEFEEIKTFNDDGCLINKPGINNVIYCFQSISPKILKNITHLPKEYNDYDCSVDRNII
QKYVSRLDIPESQPGHVPEWQRKLPEFNNTNNPRRRRKWYSNGRNISKGYSVDQVNQAKIEDSLLAQIKIGEDWIIL
DIRGLLRDLNRRELISYKNKLTIKDVLGFFSDYPIIDIKKNLVTFCYKEGVIQVVSQKSIGNKKSKQLLEKLIENKP
IALVSIDLGQTNPVSVKISKLNKINNKISIESFTYRFLNEEILKEIEKYRKDYDKLELKLINEA

[0943] >CasΦ-7

[0944] MSNTAVSTREHMSNKTTPPSPLSLLLRAHFPGLKFESQDYKIAGKKLRDGGPEAVISYLTGKGQAKLK
DVKPPAKAFVIAQSRPFIEWDLVRVSRQIQEKIFGIPATKGRPKQDGLSETAFNEAVASLEVDGKSKLNEETRAAF
YEVLGLDAPSLHAQAQNALIKSAISIREGVLKKVENRNEKNLSKTKRRKEAGEEATFVEEKAHDERGYLIHPPGVN
QTIPGYQAVVIKSCPSDFIGLPSGCLAKESAEALTDYLPHDRMTIPKGQPGYVPEWQHPLLNRRKNRRRRDWYSAS
LNKPKATCSKRSGTPNRKNSRTDQIQSGRFKGAIPVLMRFQDEWVIIDIRGLLRNARYRKLLKEKSTIPDLLSLFT
GDPSIDMRQGVCTFIYKAGQACSAKMVKTKNAPEILSELTKSGPVVLVSIDLGQTNPIAAKVSRVTQLSDGQLSHE
TLLRELLSNDSSDGKEIARYRVASDRLRDKLANLAVERLSPEHKSEILRAKNDTPALCKARVCAALGLNPEMIAWD
KMTPYTEFLATAYLEKGGDRKVATLKPKNRPEMLRRDIKFKGTEGVRIEVSPEAAEAYREAQWDLQRTSPEYLRLS
TWKQELTKRILNQLRHKAAKSSQCEVVVMAFEDLNIKMMHGNGKWADGGWDAFFIKKRENRWFMQAFHKSLTELGA
HKGVPTIEVTPHRTSITCTKCGHCDKANRDGERFACQKCGFVAHADLEIATDNIERVALTGKPMPKPESERSGDAK
KSVGARKAAFKPEEDAEAAE*

[0945] >CasΦ-8

[0946] MIKPTVSQFLTPGFKLIRNHSRTAGLKLKNEGEEACKKFVRENEIPKDECPNFQGGPAIANIIAKSRE
FTEWEIYQSSLAIQEVIFTLPKDKLPEPILKEEWRAQWLSEHGLDTVPYKEAAGLNLIIKNAVNTYKGVQVKVDNK
NKNNLAKINRKNEIAKLNGEQEISFEEIKAFDDKGYLLQKPSPNKSIYCYQSVSPKPFITSKYHNVNLPEEYIGYY
RKSNEPIVSPYQFDRLRIPIGEPGYVPKWQYTFLSKKENKRRKLSKRIKNVSPILGIICIKKDWCVFDMRGLLRTN
HWKKYHKPTDSINDLFDYFTGDPVIDTKANVVRFRYKMENGIVNYKPVREKKGKELLENICDQNGSCKLATVDVGQ
NNPVAIGLFELKKVNGELTKTLISRHPTPIDFCNKITAYRERYDKLESSIKLDAIKQLTSEQKIEVDNYNNNFTPQ
NTKQIVCSKLNINPNDLPWDKMISGTHFISEKAQVSNKSEIYFTSTDKGKTKDVMKSDYKWFQDYKPKLSKEVRDA
LSDIEWRLRRESLEFNKLSKSREQDARQLANWISSMCDVIGIENLVKKNNFFGGSGKREPGWDNFYKPKKENRWWI
NAIHKALTELSQNKGKRVILLPAMRTSITCPKCKYCDSKNRNGEKFNCLKCGIELNADIDVATENLATVAITAQSM
PKPTCERSGDAKKPVRARKAKAPEFHDKLAPSYTVVLREAV*

[0947] >CasΦ-9

[0948] MRSSREIGDKILMRQPAEKTAFQVFRQEVIGTQKLSGGDAKTAGRLYKQGKMEAAREWLLKGARDDVP
PNFQPPAKCLVVAVSHPFEEWDISKTNHDVQAYIYAQPLQAEGHLNGLSEKWEDTSADQHKLWFEKTGVPDRGLPV
QAINKIAKAAVNRAFGVVRKVENRNEKRRSRDNRIAEHNRENGLTEVVREAPEVATNADGFLLHPPGIDPSILSYA
SVSPVPYNSSKHSFVRLPEEYQAYNVEPDAPIPQFVVEDRFAIPPGQPGYVPEWQRLKCSTNKHRRMRQWSNQDYK
PKAGRRAKPLEFQAHLTRERAKGALLVVMRIKEDWVVFDVRGLLRNVEWRKVLSEEAREKLTLKGLLDLFTGDPVI
DTKRGIVTFLYKAEITKILSKRTVKTKNARDLLLRLTEPGEDGLRREVGLVAVDLGQTHPIAAAIYRIGRTSAGAL
ESTVLHRQGLREDQKEKLKEYRKRHTALDSRLRKEAFETLSVEQQKEIVTVSGSGAQITKDKVCNYLGVDPSTLPW
EKMGSYTHFISDDFLRRGGDPNIVHFDRQPKKGKVSKKSQRIKRSDSQWVGRMRPRLSQETAKARMEADWAAQNEN
EEYKRLARSKQELARWCVNTLLQNTRCITQCDEIVVVIEDLNVKSLHGKGAREPGWDNFFTPKTENRWFIQILHKT
FSELPKHRGEHVIEGCPLRTSITCPACSYCDKNSRNGEKFVCVACGATFHADFEVATYNLVRLATTGMPMPKSLER
QGGGEKAGGARKARKKAKQVEKIVVQANANVTMNGASLHSP*

[0949] >CasΦ-10

[0950] MDMLDTETNYATETPAQQQDYSPKPPKKAQRAPKGFSKKARPEKKPPKPITLFTQKHFSGVRFLKRVI
RDASKILKLSESRTITFLEQAIERDGSAPPDVTPPVHNTIMAVTRPFEEWPEVILSKALQKHCYALTKKIKIKTWP
KKGPGKKCLAAWSARTKIPLIPGQVQATNGLFDRIGSIYDGVEKKVTNRNANKKLEYDEAIKEGRNPAVPEYETAY
NIDGTLINKPGYNPNLYITQSRTPRLITEADRPLVEKILWQMVEKKTQSRNQARRARLEKAAHLQGLPVPKFVPEK
VDRSQKIEIRIIDPLDKIEPYMPQDRMAIKASQDGHVPYWQRPFLSKRRNRRVRAGWGKQVSSIQAWLTGALLVIV
RLGNEAFLADIRGALRNAQWRKLLKPDATYQSLFNLFTGDPVVNTRTNHLTMAYREGVVNIVKSRSFKGRQTREHL
LTLLGQGKTVAGVSFDLGQKHAAGLLAAHFGLGEDGNPVFTPIQACFLPQRYLDSLTNYRNRYDALTLDMRRQSLL
ALTPAQQQEFADAQRDPGGQAKRACCLKLNLNPDEIRWDLVSGISTMISDLYIERGGDPRDVHQQVETKPKGKRKS
EIRILKIRDGKWAYDFRPKIADETRKAQREQLWKLQKASSEFERLSRYKINIARAIANWALQWGRELSGCDIVIPV
LEDLNVGSKFFDGKGKWLLGWDNRFTPKKENRWFIKVLHKAVAELAPHRGVPVYEVMPHRTSMTCPACHYCHPTNR
EGDRFECQSCHVVKNTDRDVAPYNILRVAVEGKTLDRWQAEKKPQAEPDRPMILIDNQES*

[0951] 上述序列中的星号(*)表示STOP密码子。或者,CasΦ-1也称为Cas12j直系同源物1。因此,CasΦ-1至CasΦ-10也可以分别称为Cas12j直系同源物1至10。

[0952] 向导多核苷酸

[0953] 在一种实施方案中,向导多核苷酸是向导RNA。如本文所用,术语“向导RNA(gRNA)”及其语法等效物可以是指一种RNA,其可能对于靶标DNA是特异性的并且可以与Cas蛋白形

成复合物。RNA/Cas复合物可有助于将Cas蛋白“引导”至靶标DNA。Cas9/crRNA/tracrRNA核酸内切地切割与间隔序列互补的线性或环状dsDNA靶标。不与crRNA互补的靶标链首先被核酸内切地切断,然后被3’-5’核酸外切地修剪。在自然界,DNA结合和切割通常需要蛋白质和两种RNA。但是,单向导RNA(“sgRNA”或简写为“gRNA”)可以经工程改造以将多种crRNA和tracrRNA并入单一RNA物质中。参见,例如,JinekM.等人,Science337:816-821(2012),其整体内容通过引用并入本文。。Cas9识别CRISPR重复序列中的短基序(PAM或前间区序列邻近基序)以帮助区分自我与非自我。Cas9核酸酶序列和结构是本领域技术人员周知的(参见,例如,“Complete genome sequence of an M1 strain of Streptococcus pyogenes.”Ferretti,J.J.等人,Natl.Acad.Sci.U.S.A.98:4658-4663(2001);“CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.”Deltcheva E.等人,Nature471:602-607(2011);和“Programmabledual-RNA-guided DNA endonuclease in adaptive bacterial immunity.”Jinek M.等人,Science 337:816-821(2012),其各自的整体内容通过引用并入本文)。Cas9直系同源物已经在各种物种中描述,包括酿脓链球菌(S.pyogenes)和嗜热链球菌(S.thermophilus)。基于本公开,其他合适的Cas9核酸酶和序列对于本领域技术人员可以是显而易见的,并且此类Cas9核酸酶和序列包括来自生物体和基因座的Cas9序列,如Chylinski,Rhun,and Charpentier,“The tracrRNA and Cas9families of type IICRISPR-Cas immunity systems”(2013)RNA Biology 10:5,726-737中所公开的;其整体内容通过引用并入本文。在一些实施方案中,Cas9核酸酶具有失活的(例如,灭活的)DNA切割结构域,换言之,Cas9是切口酶。

[0954] 在一些实施方案中,向导多核苷酸是至少一个单向导RNA(“sgRNA”或“gRNA”)。在一些实施方案中,向导多核苷酸是至少一个tracrRNA。在一些实施方案中,向导多核苷酸不需要PAM序列来将多核苷酸可编程DNA结合结构域(例如,Cas9或Cpf1)引导至靶标核苷酸序列。

[0955] 本文所公开的碱基编辑器的多核苷酸可编程核苷酸结合结构域(例如,源自CRISPR的结构域)可通过与向导多核苷酸缔合来识别靶标多核苷酸序列。向导多核苷酸(例如,gRNA)通常是单链的并且可编程以位点特异性地(即,经由互补碱基配对)结合至多核苷酸靶标序列,从而将与该向导多核苷酸协同作用的碱基编辑器引导至该靶标序列。向导多核苷酸可以是DNA。向导多核苷酸可以是RNA。本领域一般技术人员将会了解,在向导多核苷酸序列中,尿嘧啶(U)替换序列中的胸腺嘧啶(T)。在一些实施方案中,向导多核苷酸包含天然核苷酸(例如,腺苷)。在一些实施方案中,向导多核苷酸包含非天然(或非自然)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是至少15、16、17、18、19、20、21、22、23、24、25、26、27、28、29或30个核苷酸的长度。向导核酸的靶向区域可以是10至30个核苷酸之间的长度,或15至25个核苷酸之间的长度,或15至20个核苷酸之间的长度。在一些实施方案中,向导多核苷酸可以截短1、2、3、4个等核苷酸,尤其在5’末端。作为非限制性示例,长度为20个核苷酸的向导多核苷酸可以截短1、2、3、4个等核苷酸,尤其在5’末端。

[0956] 在一些实施方案中,向导多核苷酸包含两个或更多个单个多核苷酸,这些单个多核苷酸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,向导多核苷酸可包含CRISPR RNA(crRNA)和反向活化CRISPR RNA(tracrRNA)。例如,向导多核苷酸可

包含一个或多个反向活化CRISPR RNA(tracrRNA)。

[0957] 在II型CRISPR系统中,通过CRISPR蛋白(例如,Cas9)靶向核酸通常需要在包含识别靶标序列的序列的第一RNA分子(crRNA)与包含形成稳定化向导RNA-CRISPR蛋白复合物的支架区的重复序列的第二RNA分子(trRNA)之间的互补碱基配对。此类双向导RNA系统可用作向导多核苷酸,以将本文所公开的碱基编辑器引导至靶标多核苷酸序列。

[0958] 在一些实施方案中,本文所提供的碱基编辑器利用单向导多核苷酸(例如,sgRNA)。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,两个gRNA)。在一些实施方案中,本文所提供的碱基编辑器利用一个或多个向导多核苷酸(例如,多个gRNA)。在一些实施方案中,单向导多核苷酸用于本文所述的不同碱基编辑器。例如,单向导多核苷酸可用于胞苷碱基编辑器和腺苷碱基编辑器。

[0959] 在其他实施方案中,向导多核苷酸可在一个分子(即,单分子向导核酸)中包含靶向核酸的一部分的多核苷酸和该核酸的支架部分两者。例如,单分子向导多核苷酸可以是单向导RNA(sgRNA或gRNA)。本文中,术语向导多核苷酸序列任何能够与碱基编辑器相互作用并将碱基编辑器引导至靶标多核苷酸序列的单分子、双分子或多分子核酸。

[0960] 通常,向导多核苷酸(例如,crRNA/trRNA复合物或gRNA)包含:“多核苷酸靶链段”,该链段包括能够识别并结合至靶标多核苷酸序列的序列;和“蛋白结合链段”,该链段稳定化碱基编辑器的多核苷酸可编程核苷酸结合结构域组分内的向导多核苷酸。在一些实施方案中,向导多核苷酸的多核苷酸靶向链段识别并结合至DNA多核苷酸,从而促进对DNA中的碱基的编辑。在其他实施方案中,向导多核苷酸的多核苷酸靶向链段识别并结合至RNA多核苷酸,从而促进对RNA中的碱基的编辑。本文中,“链段”是指分子的节段或区域,例如,向导多核苷酸中的核苷酸的连续延伸段。链段也可以指复合物的区域/节段,使得链段可包含超过一个分子的区域。例如,若向导多核苷酸包含多个核酸分子,则蛋白结合链段可包括例如沿着互补性区域杂交的多个独立分子中的全部或一部分。在一些实施方案中,包含两个独立分子的靶向DNA的RNA的蛋白结合链段包含(i)长度为100个碱基对的第一RNA的碱基对40至75;和(ii)长度为50个碱基对的第二RNA分子的碱基对10至25。除非在特定语境中具体地另做定义,否则“链段”的定义不限于具体数目的总碱基对,不限于任何来自给定RNA分子的特定数目的碱基对,不限于复合物内的特定数目的独立分子,且可包括RNA分子的区域,这些区域可以是任何总长度并且可包括与其他分子具有互补性的区域。

[0961] 向导RNA或向导多核苷酸可包含两个或更多个RNA,例如,CRISPR RNA(crRNA)和反向活化crRNA(tracrRNA)。有时,向导RNA或向导多核苷酸可包含单链RNA,或通过将crRNA的一部分(例如,功能性部分)与tracrRNA融合形成的单向导RNA(sgRNA)。向导RNA或向导多核苷酸也可以是包含crRNA和tracrRNA的双RNA。此外,crRNA可以与靶标DNA杂交。

[0962] 如上所述,向导RNA或向导多核苷酸可以是表达产物。例如,编码向导RNA的DNA可以是包含编码向导RNA的序列的载体。可通过用分离的向导RNA或包含编码向导RNA的序列和启动子的质粒DNA转染细胞而将向导RNA或向导多核苷酸转移至该细胞内。向导RNA或向导多核苷酸也可以用其他方式转移至细胞内,诸如使用病毒介导的基因递送。

[0963] 向导RNA或向导多核苷酸可以是分离的。例如,向导RNA可以以分离的RNA的兴衰转移至细胞或生物体内。向导RNA可以通过体外转录使用本领域中已知的任何体外转录系统制备。向导RNA可以以分离的RNA的形式而非以包含向导RNA的编码序列的质粒的形式转移

至细胞内。

[0964] 向导RNA或向导多核苷酸可包含三个区域:位于5'末端处的第一区域,其可以与染色体序列中的靶标位点互补;第二内部区域,其可以形成茎环结构;和第三3'区域,其可以是单链的。每个向导RNA的第一区域也可以是不同的,使得每个向导RNA将融合蛋白引导至特异性靶标位点。再者,每个向导RNA的第二和第三区域在全部向导RNA中可以是相同的。

[0965] 向导RNA或向导多核苷酸的第一区域可以与染色体序列中的靶标位点处的序列互补,使得向导RNA的第一区域可以与靶标位点进行碱基配对。在一些实施方案中,向导RNA可包含约10个核苷酸至25个核苷酸(即,10个核苷酸至25个核苷酸;或约10个核苷酸至约25个核苷酸;或10个核苷酸至约25个核苷酸;或约10个核苷酸至25个核苷酸)或更多.例如,向导RNA的第一区域与染色体序列中靶标位点之间的碱基配对区域可以是(约)10、11、12、13、14、15、16、17、18、19、20、22、23、24、25个或更多个核苷酸的长度。在一些实施方案中,向导RNA的第一区域可以是(约)19、20或21个核苷酸的长度。

[0966] 向导RNA或向导多核苷酸也可包含形成二级结构的第二区域。例如,由向导RNA形成的二级结构可包含茎(或发夹)和环。环和茎的长度可变。例如,环可以在约3至10个核苷酸的长度范围内,且茎可以在约6至20个碱基对的长度范围内。茎可包含一个或多个1至10个或约10个核苷酸的凸起。第二区域的整体长度可以在(约)16至60个核苷酸的长度范围内。例如,环可以是(约)4个核苷酸的长度,且茎可以是(约)12个碱基对。

[0967] 向导RNA或向导多核苷酸也可在3'末端包含第三区域,该第三区域可以基本上是单链的。例如,第三区域有时不与感兴趣的细胞内的任何染色体序列互补并且有时不与向导RNA的剩余部分互补。再者,第三区域的长度可变。第三区域可以是超过(约)4个核苷酸的长度。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。

[0968] 向导RNA或向导多核苷酸可以靶向基因靶标的任何外显子或内含子。在一些实施方案中,向导可靶向基因的外显子1或2,在其他实施方案中,向导可靶向基因的外显子3或4。组合物可包含多个向导RNA,这些向导RNA全部靶向相同的外显子,或在一些实施方案中,多个向导RNA可靶向不同的外显子。可靶向基因的外显子和内含子。

[0969] 向导RNA或向导多核苷酸可靶向(约)20个核苷酸的核酸序列。靶标核酸可以少于(约)20个核苷酸。靶标核酸可以是至少或至少约5、10、15、16、17、18、19、20、21、22、23、24、25、30个或1至100个核苷酸之间任一者的长度。靶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个或1至100个核苷酸之间任一者的长度。靶标核酸序列可以是紧邻PAM的5'第一个核苷酸的的(约)20个碱基。向导RNA可靶向核酸序列。靶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。

[0970] 向导多核苷酸,例如,向导RNA,可以指一种核酸,其可以与另一种核酸(例如,细胞基因组中的靶标核酸或前间区序列)杂交。向导多核苷酸可以是RNA。向导多核苷酸可以是DNA。向导多核苷酸可以经编程或设计以位点特异性地结合至核酸的序列。向导多核苷酸可包含一条多核苷酸链且可称为单向导多核苷酸。向导多核苷酸可包含两条多核苷酸链且可称为双向导多核苷酸。向导RNA可以作为RNA分子被引入细胞内。例如,RNA分子可以是体外转录的且/或可以是化学合成的。RNA可以从合成DNA分子(例如,gBlocks®基因片段)转录。然后,向导RNA可以作为RNA分子被引入细胞内。向导RNA也可以以非RNA核酸分子(例如,

DNA分子)的形式被引入细胞内。例如,可以将编码向导RNA的DNA可操作地链接至启动子控制序列,以在感兴趣的细胞内进行向导RNA的表达。可以将RNA编码序列可操作地链接至被RNA聚合酶III(Pol III)识别的启动子序列。可用于表达向导RNA的质粒载体包括但不限于,px330载体和px333载体。在一些实施方案中,质粒载体(例如,px333载体)可包含至少两个编码向导RNA的DNA序列。

[0971] 用于选择、涉及和验证向导多核苷酸(例如,向导)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,为了将核碱基编辑器系统中的脱氨酶结构域(例如,AID结构域)的潜在底物混杂的影响降至最低,可以将无意中成为脱氨基靶标的残基(例如,会潜在地驻留在靶标核酸基因座内ssDNA上的脱靶C残基)数降至最低。此外,可使用软件工具来优化对应于靶标核酸序列的gRNA,例如,以将跨基因组的总脱靶活性降至最低。例如,对于使用酿脓链球菌Cas9的每种可能的靶向结构与选择,可以跨基因组鉴定全部脱靶序列(在前选定的PAM,例如,NAG或NGG),该基因组含有最多某个数目(例如,1、2、3、4、5、6、7、8、9或10)的误配碱基对。可以鉴定与靶标位点互补的gRNA的第一区域,并且可将根据其总预测脱靶得分将全部第一区域(例如,crRNA)排名;名次最高的靶向结构域表示似乎具有最高上靶活性和最低脱靶活性的那些。候选靶向gRNA可以通过使用本领域中已知和/或如本文所述的方法功能性地评定。

[0972] 作为非限制性示例,用于与Cas9合用的向导RNA的crRNA中的靶标DNA杂交序列可以使用DNA序列检索算法鉴定。可使用定制gRNA设计软件基于公共工具cas-offinder进行gRNA设计,如Bae S.,Park J.,&Kim J.-S.Cas-OFFinder:A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases.Bioinformatics 30,1473-1475(2014)中所述。这一软件在计算向导的全基因组脱靶倾向后将这些向导评分。通常,对于长度在17至24范围内的向导,考虑完美匹配到7个误配范围内的匹配。一旦通过计算确定脱靶位点,就计算每个向导的总得分并使用网络界面总结到表格输出中。除了鉴定与PAM序列相邻的潜在靶标位点之外,该软件还鉴定全部PAM相邻序列,这些序列与所选择的靶标位点相异1、2、3或超过3个核苷酸。可获得用于靶标核酸序列(例如,靶标基因)的基因组DNA序列,并且可使用可公共获得的工具(例如,RepeatMasker程序)来筛选重复元件。RepeatMasker对输入的DNA序列进行重复元件和低复杂度区域的检索。输出是对给定查询序列中存在的重复序列的详细注释。

[0973] 鉴定后,可基于向导RNA(例如,crRNA)的第一区域到靶标位点的距离、其正交性和用于与相关PAM序列进行密切匹配的5'核苷酸(例如,基于含有相关PAM(例如,对于酿脓链球菌为NGG PAM,对于金黄色葡萄球菌为NNGRRT或NNGRRV PAM)的人类基因组中密切匹配的5'G)的存在对这些第一区域进行排名。如本文所用,正交性是指含有最低数目的靶标序列误配的人类基因组中的序列数目。"高水平的正交性"或"良好正交性"可以指,例如,20-mer靶向结构域,其既没有除了所希望的靶标之外的人类基因组中的相同序列,也没有任何含有一个或两个靶标序列中的误配的序列。可以选择具有良好正交性的靶向结构域以将脱靶DNA切割降至最低。

[0974] 在一些实施方案中,报告基因系统可以用于检测碱基编辑活性并测试候选向导多核苷酸。在一些实施方案中,报告基因系统可包含基于报告基因的测定,其中碱基编辑活性导致报告基因的表达。例如,报告基因系统可包括报告基因,该报告基因包含去活化的起始

密码子,例如,从3'-TAC-5'到3'-CAC-5'的模板链上的突变。在靶标C成功脱氨基后,对应的mRNA将会转录为5'-AUG-3'而非5'-GUG-3',使得能够进行报告基因的翻译。合适的报告基因对于本领域技术人员将是显而易见的。报告基因的非限制性示例包括编码以下的基因:绿色荧光蛋白(GFP)、红色荧光蛋白(RFP)、荧光素酶、分泌型碱性磷酸酶(SEAP)、或其表达可检测且对于本领域技术人员是显而易见的任何其他基因。报告基因系统可用来测试多种不同的gRNA,例如,为了确定各脱氨酶将会靶向靶标DNA序列的哪个(些)核苷酸残基。为了评估特异性碱基编辑蛋白(例如,Cas9脱氨酶融合蛋白),也可以测试靶向非模板链核苷酸残基的sgRNA。在一些实施方案中,此类gRNA可设计为使得突变的起始密码子不会与gRNA进行碱基配对。向导多核苷酸可以包含标准核苷酸、经修饰的核苷酸(例如,假尿苷)、核苷酸异构体和/或核苷酸类似物。在一些实施方案中,向导多核苷酸可包含至少一个可检测的标记。可检测的标记可以是荧光团(例如,FAM、TMR、Cy3、Cy5、德克萨斯红、俄勒冈绿、Alexa Fluors、Halo标签或任何其他合适的荧光染料)、检测标签(例如,生物素、地高辛等)、量子点或金颗粒。

[0975] 向导多核苷酸可化学合成和/或酶促合成。例如,向导RNA可使用基于亚磷酰胺的标准固相合成方法合成。或者,通过将编码向导RNA的DNA可操作地链接至被噬菌体、RNA聚合酶识别的启动子控制序列,可在体外合成向导RNA。合适的噬菌体启动子序列的示例包括T7、T3、SP6启动子序列,或其变体。在其中向导RNA包含两个独立分子(例如,crRNA和tracrRNA)的实施方案中,crRNA可化学合成且tracrRNA可酶促合成。

[0976] 在一些实施方案中,碱基编辑器系统可包含多个向导多核苷酸,例如,gRNA。例如,gRNA可以将碱基编辑器靶向至一个或多个靶标基因座(例如,至少一(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或至少50个gRNA)。在一些实施方案中,多个gRNA序列可串联排列,且优选通过直接重复序列分离。

[0977] 编码向导RNA或向导多核苷酸的DNA也可以是载体的一部分。在一些实施方案中,载体包含额外的表达控制序列(例如,增强子序列、Kozak序列、聚腺苷酸化序列、转录终止序列等)、可选择的标记物序列(例如,GFP或耐抗生素基因诸如嘌呤霉素)、复制起点等。编码向导RNA或向导多核苷酸的DNA分子可以是环状的或线性的。

[0978] 在一些实施方案中,碱基编辑器系统的一个或多个组分可以由DNA序列编码。可以将此类DNA序列一起或独立地引入表达系统(例如,细胞)内。例如,可以将编码多核苷酸可编程核苷酸结合结构域和向导RNA的DNA序列引入细胞内,每个DNA序列可以是独立分子的一部分(例如,一个载体含有多核苷酸可编程核苷酸结合结构域编码序列,而第二载体含有向导RNA编码序列)或两者可以是同一分子的一部分(例如,一个载体含有用于多核苷酸可编程核苷酸结合结构域和向导RNA的编码(和调节)序列)。

[0979] 向导多核苷酸可包含一个或多个修饰以提供具有新特征或增强特征的核酸。向导多核苷酸可包含核酸亲和性标签。向导多核苷酸可包含合成核苷酸、合成核苷酸类似物、核苷酸衍生物和/或经修饰的核苷酸。

[0980] 在一些实施方案中,gRNA或向导多核苷酸可包含修饰。修饰可以在gRNA或向导多核苷酸的任何位置处作出。可对单一gRNA或向导多核苷酸作出超过一个修饰。gRNA或向导多核苷酸可在修饰后经历质量控制。在一些实施方案中,质量控制可包括PAGE、HPLC、MS或其任意组合。

[0981] gRNA或向导多核苷酸的修饰可以是置换、插入、缺失、化学修饰、物理修饰、稳定化、纯化或其任意组合。

[0982] gRNA或向导多核苷酸也可以通过以下修饰:5'腺苷酸化、5'鸟苷-三磷酸酯封端、5'N7-甲基鸟苷-三磷酸酯封端、5'三磷酸酯封端、3'磷酸酯、3'硫代磷酸酯、5'磷酸酯、5'硫代磷酸酯和、Cis-Syn胸苷二聚体、三聚体、C12间隔序列、C3间隔序列、C6间隔序列、dSpacer、PC间隔序列、rSpacer、间隔序列18、间隔序列9、3'-3'修饰、5'-5'修饰、无碱基、吖啶、偶氮苯、生物素、生物素BB、生物素TEG、胆固醇基TEG、脱硫生物素TEG、DNP TEG、DNP-X、DOTA、dT-生物素、双生物素、PC生物素、补骨脂素C2、补骨脂素C6、TINA、3'DABCYL、黑洞淬灭剂1、黑洞淬灭剂2、DABCYL SE、dT-DABCYL、IRDye QC-1、QSY-21、QSY-35、QSY-7、QSY-9、羧基连接子、硫醇连接子、2'-脱氧核糖核苷类似物嘌呤、2'-脱氧核糖核苷类似物嘧啶、核糖核苷类似物、2'-O-甲基核糖核苷类似物、糖修饰类似物、wobble/通用碱基、荧光染料标记、2'-氟RNA、2'-O-甲基RNA、甲基磷酸酯、磷酸二酯DNA、磷酸二酯RNA、硫代磷酸酯DNA、硫代磷酸酯RNA、UNA、脚鸟苷-5'-三磷酸酯、5'-甲基胞苷-5'-三磷酸酯,或其任意组合。

[0983] 在一些实施方案中,修饰是永久的。在一些实施方案中,修饰是暂时的。在一些实施方案中,对gRNA或向导多核苷酸作出多个修饰。gRNA或向导多核苷酸修饰可改变核苷酸的生理化学性质,诸如其构象、极性、疏水性、化学反应性、碱基配对相互反应,或其任意组合。

[0984] 修饰也可以是硫代磷酸酯替代物。在一些实施方案中,天然磷酸二酯键可能容易被细胞核酸酶快速降解;并且,使用硫代磷酸酯(PS)键替代物进行核苷酸间链结的修饰可能更加稳定二不易被细胞降解所水解。修饰可以在gRNA或向导多核苷酸中增加稳定性。修饰也可以增强生物学活性。在一些实施方案中,硫代磷酸酯增强的RNA gRNA可抑制RNase A、RNase T1、小牛血清核酸酶,或其任意组合。这些性质可允许使用待用于特定应用的PS-RNA gRNA,在该应用中,暴露于核酸酶是体内或体外的高概率事件。例如,硫代磷酸酯(PS)键可以被引入gRNA的5'-或‘'-末端处的最末端3至5个核苷酸之间,这可以抑制核酸外切酶降解。在一些实施方案中,硫代磷酸酯键可添加到整个gRNA各处以减少被核酸外切酶攻击。

[0985] 在一些实施方案中,向导RNA经设计以中断剪接位点(即,剪接受体(SA)或剪接供体(SD))。在一些实施方案中,向导RNA经设计,使得碱基编辑导致提前STOP密码子。表4A至表4B提供旨在中断剪接位点或导致提前STOP密码子的gRNA靶标序列的非详尽列表。

[0986]

表 4A：gRNA：剪接位点和 STOP 密码子

基因	说明	靶向序列	gRNA 间隔序列
VISTA	外显子 1 SD (pos6)	CCTTACCTAGGGACGCAGCC	CCUUACCUAGGGACGCAGCC
	外显子 1 STOP (pos7)	GGATCCCCAGCGCCAGCTGC	GGAUCCCCAGCGCCAGCUGC
	外显子 1 STOP (pos5)	AGCGCCAGCTGCCGGCCTCC	AGCGCCAGCUGCCGGCCUCC
	外显子 1 STOP (pos4)	GCGCCAGCTGCCGGCCTCCA	GCGCCAGCUGCCGGCCUCCA
	外显子 2 STOP (pos8)	CCTGGCTCAGCGCCACGGGC	CCUGGCUCAGCGCCACGGGC
	外显子 2 STOP (pos5)	GCTGCAGGTGCAGACAGGTG	GCUGCAGGUGCAGACAGGUG
	外显子 2 STOP (pos7)	GCGGTACCACGTCTTGTAGA	GCGGUACCACGUCUUGUAGA
	外显子 3 SA (pos4)	TGCCTGTGGGAACAAACAGA	UGCCUGUGGGAACAAACAGA
	外显子 3 SD (pos5)	CTTACTTTCACTATCCTGGG	CUUACUUUCACUAUCCUGGG
	外显子 3 SD (pos8)	TCCCTTACTTTCACTATCCT	UCCCUUACUUUCACUAUCCU
	外显子 3 STOP (pos5)	CTCCCAGGATAGTGAAAGTA	CUCCCAGGAUAGUGAAAGUA
	外显子 4 SA (pos7)	TGATGTCTGAAAGGGCAGAG	UGAUGUCUGAAAGGGCAGAG
	外显子 5 STOP (pos5)	TGCCCAGGAGCTGGTGCGGA	UGCCCAGGAGCUGGUGCGGA
	外显子 6 SA (pos4)	TTGCTGCCACAGAACCAGAA	UUGCUGCCACAGAACCAGAA
	外显子 6 STOP (pos4)	ATTCAAGGGATTGAAAACCC	AUUCAAGGGAUUGAAAACCC
	外显子 6 STOP (pos8)	ACCTGCCCAGGGGATACCCG	ACCUGCCCAGGGGAUACCCG
	外显子 6 STOP (pos7)	CAGCGGCAGCCTTCTGAGTC	CAGCGGCAGCCUUCUGAGUC
TRAC	外显子 1 STOP 1 (pos5)	GCTACAAACAAGCTCATCTT	GCUACAAACAAGCUCAUCUU

[0987]

基因	说明	靶向序列	gRNA 间隔序列
	外显子 1 STOP 2 (pos6)	CCAGCCAAGTACGTAAGTAG	CCAGCCAAGUACGUAAGUAG
	外显子 2 SA (pos9)	CTGGATATCTGTGGGACAAG	CUGGAUAUCUGUGGGACAAG
	外显子 2 SD	CTTACCTGGGCTGGGGAAGA	CUUACCUGGGCUGGGGAAGA
	外显子 4 SA	TTCGTATCTGTAAAACCAAG	UUCGUAUCUGUAAAACCAAG
	外显子 4 STOP	TTTCAAAACCTGTCAGTGAT	UUUCAAAACCUGUCAGUGAU
	外显子 4 STOP	TTCAAAACCTGTCAGTGATT	UUCAAAACCUGUCAGUGAUU
Tim-3	外显子 2 SA (pos6)	GGACCCTGCATAGAGAGAGA	GGACCCUGCAUAGAGAGAGA
	外显子 2 STOP (pos5)	TGCCCCAGCAGACGGGCACG	UGCCCCAGCAGACGGGCACG
	外显子 3 SD (pos5)	GTTACCTGGGCCATGTCCCC	GUUACCUGGGCCAUGUCCCC
	外显子 4 SD (pos5)	CTTACTGTTAGATTTATATC	CUUACUGUUAGAUUUAUAUC
	外显子 4 SD (pos4)	TTACTGTTAGATTTATATCA	UUACUGUUAGAUUUAUAUCA
	外显子 5 SA (pos5)	TTTGCTATGGAAACACAAAC	UUUGCUAUGGAAACACAAAC
	外显子 5 STOP (pos8)	TCCATAGCAAATATCCACAT	UCCAUAGCAAAUAUCCACAU
	外显子 7 STOP (pos5)	GCAGCAACCCTCACAACCTT	GCAGCAACCCUCACAACCUU
	外显子 7 STOP (pos 4)	CAGCAACCCTCACAACCTTT	CAGCAACCCUCACAACCUUU
TIGIT	外显子 1 STOP (pos4)	AGGCAGGCTCCCCTCGCCTC	AGGCAGGCUCCCCUCGCCUC
	外显子 2 STOP (5&8)	GGAGCAGCAGGACCAGCTTC	GGAGCAGCAGGACCAGCUUC
	外显子 2 SD (pos9)	CAGGAATACCTGAGCTTTCT	CAGGAAUACCUGAGCUUUCU
	外显子 3 STOP (pos7)	AGGTTCCAGATTCCATTGCT	AGGUUCCAGAUUCCAUUGCU

[0988]

基因	说明	靶向序列	gRNA 间隔序列
	外显子 1 STOP	CTGGGCCCAGGGGCTGAGGC	CUGGGCCCAGGGGCUGAGGC
	外显子 2 STOP	GATCGAGTGGCCCCAGGTCC	GAUCGAGUGGCCCCAGGUCC
TGFbRII	外显子 1 SD (JMG79)	TCACCCGACTTCTGAACGTG	UCACCCGACUUCUGAACGUG
	外显子 3 SD (JMG83)	TTACCTGCCCACTGTTAGCC	UUACCUGCCCACUGUUAGCC
	外显子 2 STOP (JMG80)	GAAGCCACAGGAAGTCTGTG	GAAGCCACAGGAAGUCUGUG
	外显子 3 STOP (JMG81)	ACTCCAGTTCCTGACGGCTG	ACUCCAGUUCCUGACGGCUG
	外显子 3 STOP (JMG82)	ACCTACAGGAGTACCTGACG	ACCUACAGGAGUACCUGACG
	外显子 4 STOP (JMG84)	TTCCCAGAGCACCAGAGCCA	UUCCCAGAGCACCAGAGCCA
	外显子 1 STOP (JMG85)	ACGTTCAGAAGTCGGGTGAG	ACGUUCAGAAGUCGGGUGAG
	外显子 3 STOP (pos8)	TTCAGAGCAGTTTGAGACAG	UUCAGAGCAGUUUGAGACAG
RFXANK	外显子 2 SA (JMG8)	CCTGCTGGGAAACAGACAAC	CCUGCUGGGAAACAGACAAC
	外显子 2 SD (JMG9)	CACTCACAGTCTAGGGTGGC	CACUCACAGUCUAGGGUGGC
	外显子 2 STOP (pos8)	CAACCGGCAGCGAGGGAACG	CAACCGGCAGCGAGGGAACG
	外显子 3 SA (pos7)	ACAGGGCTGGGGCAGGACAG	ACAGGGCUGGGGCAGGACAG
	外显子 3 STOP (pos8)	CATCCACCAGCTCGCAGCAC	CAUCCACCAGCUCGCAGCAC
	外显子 3 STOP (pos7)	ATCCACCAGCTCGCAGCACA	AUCCACCAGCUCGCAGCACA
	外显子 3 STOP (pos6)	TCCACCAGCTCGCAGCACAG	UCCACCAGCUCGCAGCACAG
	外显子 3 STOP (pos5)	CCACCAGCTCGCAGCACAGG	CCACCAGCUCGCAGCACAGG
	外显子 4 SA (JMG10)	TGTCACCTGGCAGGAGGAGG	UGUCACCUGGCAGGAGGAGG

[0989]

基因	说明	靶向序列	gRNA 间隔序列
	外显子 4 SA (pos6)	GTCACCTGGCAGGAGGAGGC	GUCACCUGGCAGGAGGAGGC
	外显子 5 SA (pos7)	GGCACCCTGCAGGGAGAAGA	GGCACCCUGCAGGGAGAAGA
	外显子 5 SA (JMG11)	GCACCCTGCAGGGAGAAGAA	GCACCCUGCAGGGAGAAGAA
	外显子 6 SA (pos4)	ATTCTGTCGTGGGTAGGGGC	AUUCUGUCGUGGGUAGGGGC
	外显子 6 SA (JMG12)	CTCCATTCTGTCGTGGGTAG	CUCCAUUCUGUCGUGGGUAG
	外显子 7 SA (pos8)	CCTCGGGCTGCAAAGGAGAG	CCUCGGGCUGCAAAGGAGAG
	外显子 7 SA (pos5)	CGGGCTGCAAAGGAGAGGGG	CGGGCUGCAAAGGAGAGGGG
	外显子 7 SD (pos6)	GCTGACCTTTCCGGTATCCC	GCUGACCUUUCCGGUAUCCC
	外显子 7 SD (pos5)	CTGACCTTTCCGGTATCCCA	CUGACCUUUCCGGUAUCCCA
	外显子 8 SA (pos8)	TGTTGCACTGAGATGGGGCA	UGUUGCACUGAGAUGGGGCA
	外显子 8 SA (pos9)	CTGTTGCACTGAGATGGGGC	CUGUUGCACUGAGAUGGGGC
PVRIG (CD112R)	外显子 1 STOP (pos7)	GCCCTGCAGCCCCCCAGAACC	GCCCUGCAGCCCCCCAGAACC
	外显子 1 SD (pos5)	CTCACCCGCAGTGACACACA	CUCACCCGCAGUGACACACA
	外显子 1 STOP (pos8)	GCAGCACCCAGGGCAGGACC	GCAGCACCCAGGGCAGGACC
	外显子 1 STOP (pos7)	CAGCACCCAGGGCAGGACCA	CAGCACCCAGGGCAGGACCA
	外显子 2 SA (pos5)	GTCCCTGTGGAACAGCAGCA	GUCCCUGUGGAACAGCAGCA
	外显子 2 STOP (pos8)	GTGGGTTCAAGTTCGGATGG	GUGGGUUCAAGUUCGGAUGG
	外显子 2 SD (pos 7)	GCCCCACCTGGGTCTGAGCT	GCCCCACCUGGGUCUGAGCU
	外显子 2 SD (pos8)	GGCCCCACCTGGGTCTGAGC	GGCCCCACCUGGGUCUGAGC

[0990]

基因	说明	靶向序列	gRNA 间隔序列
	外显子 2 SD (pos4)	CCACCTGGGTCTGAGCTGGG	CCACCUGGGUCUGAGCUGGG
	外显子 2 STOP (pos8)	AGGCCTCCCAGGAGCCCTCA	AGGCCUCCCAGGAGCCCUCA
	外显子 2 STOP (pos4)	CTCCCAGGAGCCCTCAGGGA	CUCCCAGGAGCCCUCAGGGA
	外显子 2 STOP (pos4)	CCCCCAGCTCACAGTCACCA	CCCCCAGCUCACAGUCACCA
	外显子 3 SD (pos8)	GGTCTCACCGGTGCTTATGT	GGUCUCACCGGUGCUUAUGU
	外显子 3 STOP (pos9)	TGCTGCGCCGACATAAGCAC	UGCUGCGCCGACAUAAGCAC
	外显子 4 SA (pos8)	GGCAGGGCTGGGAGAGAGCA	GGCAGGGCUGGGAGAGAGCA
	外显子 4 STOP (pos9)	CGAGAGCACGAGCATGGGTG	CGAGAGCACGAGCAUGGGUG
	外显子 4 STOP (pos6)	GAGCACGAGCATGGGTGAGG	GAGCACGAGCAUGGGUGAGG
	外显子 4 STOP (pos5)	AGCACGAGCATGGGTGAGGA	AGCACGAGCAUGGGUGAGGA
	外显子 4 STOP (pos4)	GCACGAGCATGGGTGAGGAG	GCACGAGCAUGGGUGAGGAG
	外显子 4 SD (pos5)	CTCACCCATGCTCGTGCTCT	CUCACCCAUGCUCGUGCUCU
	外显子 5 SA (pos6)	GGTGCCTGCGCGGGGGAAGG	GGUGCCUGCGCGGGGGAAGG
	外显子 5 SA (pos5)	GTGCCTGCGCGGGGGAAGGA	GUGCCUGCGCGGGGGAAGGA
	外显子 5 SA (pos9)	CTTGGTGCCTGCGCGGGGGA	CUUGGUGCCUGCGCGGGGGA
	外显子 5 STOP (pos6)	GGCCCCAGGGCCCTGCCGCC	GGCCCCAGGGCCCUGCCGCC
	外显子 5 STOP (pos9)	TCTACGCTCAGGCAGGGGAG	UCUACGCUCAGGCAGGGGAG
	外显子 5 STOP (pos4)	CCACCAGGACGGCCCCCCAT	CCACCAGGACGGCCCCCCAU
	外显子 5 STOP (pos5)	AGGCCCAGGCGGCAGGGCCC	AGGCCCAGGCGGCAGGGCCC

[0991]

基因	说明	靶向序列	gRNA 间隔序列
	外显子 5 STOP (pos4)	GGCCCAGGCGGCAGGGCCCT	GGCCCAGGCGGCAGGGCCCU
PDCD1	外显子 1 STOP 2 (pos9)	ACGACTGGCCAGGGCGCCTG	ACGACUGGCCAGGGCGCCUG
	外显子 1 STOP 4 (pos7)	CACCGCCCAGACGACTGGCC	CACCGCCCAGACGACUGGCC
	外显子 1 STOP (pos4)	CTACAACTGGGCTGGCGGCC	CUACAACUGGGCUGGCGGCC
	外显子 1 SD	CACCTACCTAAGAACCATCC	CACCUACCUAAGAACCAUCC
	外显子 2 SA	GGAGTCTGAGAGATGGAGAG	GGAGUCUGAGAGAUGGAGAG
	外显子 2 STOP 1 (pos8)	CAGCAACCAGACGGACAAGC	CAGCAACCAGACGGACAAGC
	外显子 2 STOP 2 (pos9)	GTGTCACACAACTGCCCAAC	GUGUCACACAACUGCCCAAC
	外显子 3 STOP 1 (pos8)	AGCCGGCCAGTTCCAAACCC	AGCCGGCCAGUUCCAAACCC
	外显子 3 STOP (pos7)	CAGTTCCAAACCCTGGTGGT	CAGUUCCAAACCCUGGUGGU
	外显子 3 STOP 2 (pos5)	CGGCCAGTTCCAAACCCTGG	CGGCCAGUUCCAAACCCUGG
	外显子 3 STOP (pos5)	GGACCCAGACTAGCAGCACC	GGACCCAGACUAGCAGCACC
	外显子 3 SD	GACGTTACCTCGTGCGGCCC	GACGUUACCUCGUGCGGCCC
	外显子 4 SA	TCCCTGCAGAGAAACACACT	UCCCUGCAGAGAAACACACU
	外显子 4 SD	GAGACTCACCAGGGGCTGGC	GAGACUCACCAGGGGCUGGC
	外显子 5 SA	CCTCCTTCTTTGAGGAGAAA	CCUCCUUCUUUGAGGAGAAA
	外显子 2 STOP (pos 7)	GGGGTTCCAGGGCCTGTCTG	GGGGUUCCAGGGCCUGUCUG
	外显子 3 SA	TTCTCTCTGGAAGGGCACAA	UUCUCUCUGGAAGGGCACAA
	外显子 5 STOP 1 (pos 8)	CCAGTGGCGAGAGAAGACCC	CCAGUGGCGAGAGAAGACCC

[0992]

基因	说明	靶向序列	gRNA 间隔序列
	外显子 5 STOP 2 (pos 5)	TGCCCAGCCACTGAGGCCTG	UGCCCAGCCACUGAGGCCUG
	外显子 1 STOP 1 (pos8)	CGACTGGCCAGGGCGCCTGT	CGACUGGCCAGGGCGCCUGU
	外显子 1 STOP 3 (pos6)	ACCGCCCAGACGACTGGCCA	ACCGCCCAGACGACUGGCCA
Lag3	外显子 1 STOP (pos8)	GTTTCTGCAGCCGCTTTGGG	GUUUCUGCAGCCGCUUUGGG
	外显子 1 SD (pos4)	TTACCTGGAGCCACCCAAAG	UUACCUGGAGCCACCCAAAG
	外显子 2 SA (pos4)	TCACTAGGTGAGCAAAAGAG	UCACUAGGUGAGCAAAAGAG
	外显子 2 STOP (pos8)	GCCTCTCCAGCCAGGGGCTG	GCCUCUCCAGCCAGGGGCUG
	外显子 2 STOP (pos 6)	CTTGGCAGCATCAGCCAGAC	CUUGGCAGCAUCAGCCAGAC
	外显子 3 SA (pos4)	CCACTGGGCGGGAAAGAGAA	CCACUGGGCGGGAAAGAGAA
	外显子 3 SD (pos6)	ACATACTCGAGGCCTGGCCC	ACAUACUCGAGGCCUGGCCC
	外显子 3 STOP (pos5)	CCTGCAGCCCCGCGTCCAGC	CCUGCAGCCCCGCGUCCAGC
	外显子 3 STOP (pos7)	CGCGTCCAGCTGGATGAGCG	CGCGUCCAGCUGGAUGAGCG
	外显子 3 STOP (pos6)	TGGGCCAGGCCTCGAGTATG	UGGGCCAGGCCUCGAGUAUG
	外显子 4 SD (pos4)	GGGAGTTACCCAGAACAGTG	GGGAGUUACCCAGAACAGUG
	外显子 4 STOP (pos8)	CCTGCCCCAAGTCAGCCCCA	CCUGCCCCAAGUCAGCCCCA
	外显子 4 STOP (pos9)	GCCAGGGCCGAGTCCCTGTC	GCCAGGGCCGAGUCCCUGUC
	外显子 4 STOP (pos8)	CCAGGGCCGAGTCCCTGTCC	CCAGGGCCGAGUCCCUGUCC
	外显子 4 STOP (pos4)	GCCCCAGGGCCCAGAGTCCA	GCCCCAGGGCCCAGAGUCCA
	外显子 5 STOP (pos9)	ATGTGAGCCAGGCCCAGGCT	AUGUGAGCCAGGCCCAGGCU

[0993]

基因	说明	靶向序列	gRNA 间隔序列
	外显子 5 STOP (pos 8)	GAGGAGTCCACTTGGCAGTG	GAGGAGUCCACUUGGCAGUG
	外显子 6 SA (pos7)	GAGTCACTGAAAAGAGTAGA	GAGUCACUGAAAAGAGUAGA
	外显子 6 STOP (pos6)	CTGGACAAGAACGCTTTGTG	CUGGACAAGAACGCUUUGUG
	外显子 6 STOP (pos7)	CCATCCCAGAGGAGTTTCTC	CCAUCCCAGAGGAGUUUCUC
	外显子 6 STOP (pos4)	TGGCAATGCCAGCTGTACCA	UGGCAAUGCCAGCUGUACCA
	外显子 6 STOP (pos4)	TACCAGGGGGAGAGGCTTCT	UACCAGGGGGAGAGGCUUCU
	外显子 6 STOP (pos8)	GGCATTGCCAAGGCTGGGAA	GGCAUUGCCAAGGCUGGGAA
	外显子 7 SA (pos6)	GGCACCTATGGAGAAAGTAC	GGCACCUAUGGAGAAAGUAC
	外显子 7 STOP (pos4)	AGACAGGTGAGCCAGGGACA	AGACAGGUGAGCCAGGGACA
	外显子 7 SD (pos7)	GGCTCACCTGTCTTCTCCAA	GGCUCACCUGUCUUCUCCAA
	外显子 8 SA (pos8)	GTCGCCACTGTGAGAAGAGA	GUCGCCACUGUGAGAAGAGA
	外显子 8 STOP (pos8)	GCAGGCTCAGAGCAAGATAG	GCAGGCUCAGAGCAAGAUAG
	外显子 8 STOP (pos8)	GCTGGAGCAAGAACCGGAGC	GCUGGAGCAAGAACCGGAGC
CTLA-4	外显子 1 SD (pos 6)	ACTCACCTTTGCAGAAGACA	ACUCACCUUUGCAGAAGACA
	外显子 1 SD	CACTCACCTTTGCAGAAGAC	CACUCACCUUUGCAGAAGAC
	外显子 1 STOP (pos5)	AGGGCCAGGTCCTGGTAGCC	AGGGCCAGGUCCUGGUAGCC
	外显子 2 STOP	GGCCCAGCCTGCTGTGGTAC	GGCCCAGCCUGCUGUGGUAC
	外显子 2 STOP (pos 8)	GCTTCGGCAGGCTGACAGCC	GCUUCGGCAGGCUGACAGCC
	外显子 2 STOP	TATCCAAGGACTGAGGGCCA	UAUCCAAGGACUGAGGGCCA

[0994]

基因	说明	靶向序列	gRNA 间隔序列
	外显子 2 STOP	GGAACCCAGATTTATGTAAT	GGAACCCAGAUUUAUGUAAU
	外显子 2 SD	GCTCACCAATTACATAAATC	GCUCACCAAUUACAUAAAUC
	外显子 2 SD	CTCACCAATTACATAAATCT	CUCACCAAUUACAUAAAUCU
	外显子 1 STOP	CTCAGCTGAACCTGGCTACC	CUCAGCUGAACCUGGCUACC
Chi311	外显子 1 STOP (pos8)	GGCGTCTCAAACAGGTATCT	GGCGUCUCAAACAGGUAUCU
	外显子 1 SA (pos7)	CAAAGCCTGAAGAGAAATCC	CAAAGCCUGAAGAGAAAUCC
	外显子 3 SA (pos6)	AGAGCCTGAAGGAGAAGTCT	AGAGCCUGAAGGAGAAGUCU
	外显子 3 STOP (pos4)	TCCCAGTACCGGGAAGGCGA	UCCCAGUACCGGGAAGGCGA
	外显子 4 SA (pos6)	GGTTCCTGTGGAGCACAGGG	GGUUCCUGUGGAGCACAGGG
	外显子 4 SA (pos9)	TGGGGTTCCTGTGGAGCACA	UGGGGUUCCUGUGGAGCACA
	外显子 6 SA (pos8)	TCATTTCCTAGATGGGAGAC	UCAUUUCCUAGAUGGGAGAC
	外显子 6 SA (pos4)	TTCCTAGATGGGAGACAGGC	UUCCUAGAUGGGAGACAGGC
	外显子 8 SA (pos9)	CCAGGTGTCTGAGGAGGAAG	CCAGGUGUCUGAGGAGGAAG
	外显子 8 SA (pos5)	GTGTCTGAGGAGGAAGGGGA	GUGUCUGAGGAGGAAGGGGA
	外显子 9 SA (pos6)	TAGTCCTGGGTGGGGTAGGG	UAGUCCUGGGUGGGGUAGGG
	外显子 9 SA (pos5)	AGTCCTGGGTGGGGTAGGGT	AGUCCUGGGUGGGGUAGGGU
	外显子 9 SD (pos6)	CATTACCTCATAGTAGGCAA	CAUUACCUCAUAGUAGGCAA
	外显子 9 SD (pos7)	CCATTACCTCATAGTAGGCA	CCAUUACCUCAUAGUAGGCA
	外显子 10 SA (pos7)	ACAGATCTGAGCAGATAACA	ACAGAUCUGAGCAGAUAACA

[0995]

基因	说明	靶向序列	gRNA 间隔序列
	外显子 10 STOP (pos 7)	TCCTACCCACTGGTTGCCCT	UCCUACCCACUGGUUGCCCU
	外显子 11 STOP (pos7)	AGGTGCAGTACCTGAAGGAC	AGGUGCAGUACCUGAAGGAC
	外显子 11 STOP (pos5)	CAGGCAGCTGGCGGGCGCCA	CAGGCAGCUGGCGGGCGCCA
	外显子 11 STOP (pos7)	GACTTCCAGGGCTCCTTCTG	GACUUCCAGGGCUCCUUCUG
CD96	外显子 1 STOP (pos5)	CATCCAGATACATTTTGTCA	CAUCCAGAUACAUUUUGUCA
	外显子 2 STOP (pos5)	ACCTGCCAAACACAGACAGT	ACCUGCCAAACACAGACAGU
	外显子 2 STOP (pos7)	CGTGCAGATGCAATGGTCCA	CGUGCAGAUGCAAUGGUCCA
	外显子 3 SA (pos6)	TGTAACTGTAACAAAACATA	UGUAACUGUAACAAAACAUA
	外显子 3 SD (pos6)	ACTTACCACCGACCATGCAT	ACUUACCACCGACCAUGCAU
	外显子 5 SD (pos5)	CTTACCAAAAACCTTGACTG	CUUACCAAAAACCUUGACUG
	外显子 5 STOP (pos6)	CCAGTCCAAATCTTCGATGA	CCAGUCCAAAUCUUCGAUGA
	外显子 5 STOP (pos7)	CAGTCCAAATCTTCGATGAT	CAGUCCAAAUCUUCGAUGAU
	外显子 7 STOP (pos4)	AAACCATGTGATATTTGCTT	AAACCAUGUGAUAUUUGCUU
	外显子 8 STOP (pos6)	ATGTTCCACACTTTATTTCC	AUGUUCCACACUUUAUUUCC
	外显子 10 SD (pos4)	TCACGTTGAGGAGTGGTGTT	UCACGUUGAGGAGUGGUGUU
	外显子 13 SA (pos7)	CATTGTCTAGGGATATAAAG	CAUUGUCUAGGGAUAUAAAG
	外显子 13 SA (pos8)	ACATTGTCTAGGGATATAAA	ACAUUGUCUAGGGAUAUAAA
	外显子 13 SA (pos9)	GACATTGTCTAGGGATATAA	GACAUUGUCUAGGGAUAUAA
	外显子 14 STOP (pos4)	TGGCCAGGACATTCCATCTT	UGGCCAGGACAUUCCAUCUU

[0996]

基因	说明	靶向序列	gRNA 间隔序列
	外显子 15 SA (pos6)	CCATTCTAGGAACAAAATAT	CCAUUCUAGGAACAAAAUAU
Cblb	外显子 1 STOP	GAGCTTCCAAGTCTTCTCCA	GAGCUUCCAAGUCUUCUCCA
	外显子 1 STOP (JMG44)	TCCCCGAAAAGGTCGAATTT	UCCCCGAAAAGGUCGAAUUU
	外显子 2 STOP	ATGAAGAACAGTCACAGGAC	AUGAAGAACAGUCACAGGAC
	外显子 3 SA	GATTTCGTCTGTAGGCACAA	GAUUUCGUCUGUAGGCACAA
	外显子 4 SD	TAAACTTACCTGAAACAGCC	UAAACUUACCUGAAACAGCC
	外显子 4 STOP	ATTCAGACAGTGCCTTCATG	AUUCAGACAGUGCCUUCAUG
	外显子 6 STOP	GTTGCACTCGATTGGGACAG	GUUGCACUCGAUUGGGACAG
	外显子 6 STOP	TTATTTCAAGCCCTGATTGA	UUAUUUCAAGCCCUGAUUGA
	外显子 7 SD	TTACCTGTGTAACTTTTATA	UUACCUGUGUAACUUUUAUA
	外显子 8 SA (pos8)	ATTGTTCCTGGAATTTGGGG	AUUGUUCCUGGAAUUUGGGG
	外显子 8 SD (JMG48)	ATTATACCTGCCATGCCGTA	AUUAUACCUGCCAUGCCGUA
	外显子 8 SA (pos 5) (JMG46)	GTTCCTGGAATTTGGGGAGG	GUUCCUGGAAUUUGGGGAGG
	外显子 8 STOP (JMG47)	CTGCCATGCCGTAAGGCAAG	CUGCCAUGCCGUAAGGCAAG
	外显子 10 SD (JMG49)	TCTACCTTTGGTGAACCCGT	UCUACCUUUGGUGAACCCGU
	外显子 11 SD (JMG50)	CTTACCTTAGCTCCTTCTAA	CUUACCUUAGCUCCUUCUAA
	外显子 11 STOP	GGGATGTCGACTCCTAGGGG	GGGAUGUCGACUCCUAGGGG
	外显子 11 STOP	CGAGGGCACCATGCTTCAAG	CGAGGGCACCAUGCUUCAAG
	外显子 12 SD	AAACTCACTTTATGCTAGGG	AAACUCACUUUAUGCUAGGG

[0997]

基因	说明	靶向序列	gRNA 间隔序列
	外显子 12 SD (JMG51)	CTCACTTTATGCTAGGGAGG	CUCACUUUAUGCUAGGGAGG
	外显子 16 SA (JMG52)	CTTCACCTGCATTTAAAGAA	CUUCACCUGCAUUUAAAGAA
	外显子 4 STOP (JMG45)	CCACCAGATTAGCTCTGGCC	CCACCAGAUUAGCUCUGGCC
	外显子 10 SD (pos4)	CTACCTTTGGTGAACCCGTT	CUACCUUUGGUGAACCCGUU
BTLA	外显子 1 STOP (pos6)	ATGTTCCAGATGTCCAGATA	AUGUUCCAGAUGUCCAGAUA
	外显子 1 STOP (pos5)	TGTTCCAGATGTCCAGATAT	UGUUCCAGAUGUCCAGAUAU
	外显子 2 STOP (pos8)	AGATAGACAAACAAGTTGGA	AGAUAGACAAACAAGUUGGA
	外显子 2 STOP (pos9)	AGCTTGCACCAAGTCACATG	AGCUUGCACCAAGUCACAUG
	外显子 3 SD (pos6)	ACCCACCTTGGTGCCTTCTC	ACCCACCUUGGUGCCUUCUC
B2M (BE)	外显子 1 SD	ACTCACGCTGGATAGCCTCC	ACUCACGCUGGAUAGCCUCC
	外显子 2 SA (pos9)	TGGAGTACCTGAGGAATATC	UGGAGUACCUGAGGAAUAUC
	外显子 2 STOP (pos6)	TTACCCCACTTAACTATCTT	UUACCCCACUUAACUAUCUU
	外显子 3 SA	TCGATCTATGAAAAGACAG	UCGAUCUAUGAAAAGACAG
	外显子 2 STOP	TACCCCACTTAACTATCT	UACCCCACUUAACUAUCU
B2M (ABE)	外显子 1 SD 1 (pos 5)	ACTCACGCTGGATAGCCTCC	ACUCACGCUGGAUAGCCUCC
	外显子 2 SA (pos 4)	CTCAGGTACTCCAAAGATTC	CUCAGGUACUCCAAAGAUUC
	外显子 2 SD (pos 4)	CTTACCCCACTTAACTATCT	CUUACCCCACUUAACUAUCU
TET2	外显子 1 STOP 1 (pos 8)	CATTTGCCAGACAGAACCTC	CAUUUGCCAGACAGAACCUC
	外显子 1 STOP 2 (pos 4)	AAACAAGACCAAAAGGCTAA	AAACAAGACCAAAAGGCUAA

[0998]

基因	说明	靶向序列	gRNA 间隔序列
	外显子 1 STOP 3 (pos 7)	GTAAGCCAAGAAAGAAATCC	GUAAGCCAAGAAAGAAAUCC
	外显子 1 STOP 4 (pos 5)	GCTTCAGATTCTGAATGAGC	GCUUCAGAUUCUGAAUGAGC
	外显子 1 STOP 5 (pos 7)	TTAAAACAAAATGAAATGAA	UUAAAACAAAAUGAAAUGAA
	外显子 1 STOP 6 (pos 7)	GTTCCTCAGCTTCCTTCAGA	GUUCCUCAGCUUCCUUCAGA
	外显子 1 STOP 7 (pos 8)	CAAAGAGCAAGAGATTCTGA	CAAAGAGCAAGAGAUUCUGA
	外显子 1 STOP 8 (pos 7)	AAAGAGCAAGAGATTCTGAA	AAAGAGCAAGAGAUUCUGAA
	外显子 1 STOP 9 (pos 4)	ACACAGCACTATCTGAAACC	ACACAGCACUAUCUGAAACC
	外显子 1 STOP 10 (pos 5)	CACCCAGAAAACAACACAGC	CACCCAGAAAACAACACAGC
	外显子 1 STOP 11 (pos 4)	TACCAAGTTGAAATGAATCA	UACCAAGUUGAAAUGAAUCA
	外显子 1 STOP 12 (pos 7)	ATGAATCAAGGGCAGTCCCA	AUGAAUCAAGGGCAGUCCCA
	外显子 1 STOP 13 (pos 5)	AGGGCAGTCCCAAGGTACAG	AGGGCAGUCCCAAGGUACAG
	外显子 1 STOP 14 (pos 5)	GTTCCAAAAACCCTCACACC	GUUCCAAAAACCCUCACACC
	外显子 1 STOP 15 (pos 5)	GAAACAGCACTTGAATCAAC	GAAACAGCACUUGAAUCAAC
	外显子 1 STOP 16 (pos 5)	ATTACAAATAAAGAATAAAG	AUUACAAAUAAAGAAUAAAG
	外显子 1 STOP 17 (pos 8)	TAATGTCCAAATGGGACTGG	UAAUGUCCAAAUGGGACUGG
	外显子 1 STOP 18 (pos 6)	CAAAGCAAGATCTTCTTCAC	CAAAGCAAGAUCUUCUUCAC
	外显子 1 STOP 19 (pos 5)	ACAACAAGCTTCAGTTCTAC	ACAACAAGCUUCAGUUCUAC
	外显子 1 STOP 20 (pos 6)	CTGCGCAACTTGCTCAGCAA	CUGCGCAACUUGCUCAGCAA
	外显子 1 STOP 21 (pos 5)	CACTCAGACCCCTCCCCAGA	CACUCAGACCCCUCCCCAGA

[0999]

基因	说明	靶向序列	gRNA 间隔序列
	外显子 1 STOP 22 (pos 6)	TTTTTCCATGTTTTGTTTTC	UUUUUCCAUGUUUUGUUUUC
	外显子 1 SD (pos 4)	TTACCTACACATCTGCAAGA	UUACCUACACAUCUGCAAGA
	外显子 3 SD (pos 8)	ACACTTACCCACTTAGCAAT	ACACUUACCCACUUAGCAAU
	外显子 7 STOP (pos 5)	CATGCAGAATGGCAGCACAT	CAUGCAGAAUGGCAGCACAU
	外显子 8 STOP 1 (pos 6)	AAGCTCAGGAGGAGAAAAAA	AAGCUCAGGAGGAGAAAAAA
	外显子 8 STOP 2 (pos 8)	CGCAAGCCAGGCTAAACAGT	CGCAAGCCAGGCUAAACAGU
	外显子 9 STOP 1 (pos 8)	TTCTCCCCAGTCTCAGCCGA	UUCUCCCCAGUCUCAGCCGA
	外显子 9 STOP 2 (pos 5)	TGGTCAGGAAAAGCAGCCAT	UGGUCAGGAAAAGCAGCCAU
	外显子 9 STOP 3 (pos 7)	CTAGTCCAGGGTGTGGCTTC	CUAGUCCAGGGUGUGGCUUC
Spry1	外显子 1 STOP 1	CCCCAAAATCAACATGGCAG	CCCCAAAAUCAACAUGGCAG
	外显子 1 STOP 2	TGTGATCCAGCAGCCTTCTT	UGUGAUCCAGCAGCCUUCUU
	外显子 1 STOP 3	GACCAGATCAAGGCCATAAG	GACCAGAUCAAGGCCAUAAG
	外显子 1 STOP 4	CAAGACAAGAAAAGCATGAA	CAAGACAAGAAAAGCAUGAA
	外显子 1 STOP 5	CTGAACAGGGACTGTTAGGA	CUGAACAGGGACUGUUAGGA
Spry2	外显子 1 STOP 1	CCAGAGCTCAGAGTGGCAAC	CCAGAGCUCAGAGUGGCAAC
	外显子 1 STOP 2	TTGCTGCAGACGCCCCGTGA	UUGCUGCAGACGCCCCGUGA
	外显子 1 STOP 3	CTGCAGACGCCCCGTGACGG	CUGCAGACGCCCCGUGACGG
	外显子 1 STOP 4	CGACAAGCAGTGCCTTTGCT	CGACAAGCAGUGCCUUUGCU
	外显子 1 STOP 5	GCCCAGAACGTGATTGACTA	GCCCAGAACGUGAUUGACUA

[1000]

基因	说明	靶向序列	gRNA 间隔序列
	外显子 1 STOP 6	TGTGCCAGGGGGTGTTATGAC	UGUGCCAGGGGUGUUAUGAC
	外显子 1 STOP 7	CAGATCCAGTCTGATGGCAG	CAGAUCCAGUCUGAUGGCAG
	外显子 1 STOP 8	TGTACACGATGGTCAGCCAT	UGUACACGAUGGUCAGCCAU
CIITA	外显子 1 SD (pos 6)	TTTTACCTTGGGGCTCTGAC	UUUUACCUUGGGGCUCUGAC
	外显子 1 STOP 1 (pos 6)	AGCCCCAAGGTAAAAAGGCC	AGCCCCAAGGUAAAAAGGCC
	外显子 1 STOP 2 (pos 7)	GAGCCCCAAGGTAAAAAGGC	GAGCCCCAAGGUAAAAAGGC
	外显子 2 STOP 1 (pos 8)	CAGCTCACAGTGTGCCACCA	CAGCUCACAGUGUGCCACCA
	外显子 2 STOP 2 (pos 7)	TATGACCAGATGGACCTGGC	UAUGACCAGAUGGACCUGGC
	外显子 4 STOP 1 (pos 8)	ACTGGACCAGTATGTCTTCC	ACUGGACCAGUAUGUCUUCC
	外显子 4 STOP 2 (pos 8)	TGTCTTCCAGGACTCCCAGC	UGUCUUCCAGGACUCCCAGC
	外显子 7 STOP 1 (pos 7)	TTCAACCAGGAGCCAGCCTC	UUCAACCAGGAGCCAGCCUC
	外显子 7 STOP 2 (pos 4)	GACCAGATTCCCAGTATGTT	GACCAGAUUCCCAGUAUGUU
	外显子 7 SD (pos 8)	TAACATACTGGGAATCTGGT	UAACAUACUGGGAAUCUGGU
	外显子 8 SA (pos 8)	AAAGGCACTGCAAGAGACAA	AAAGGCACUGCAAGAGACAA
	外显子 8 STOP (pos 8)	CTCTGGCAAATCTCTGAGGC	CUCUGGCAAAUCUCUGAGGC
	外显子 9 STOP 1 (pos 4)	AGCCAAGTACCCCCTCCCAG	AGCCAAGUACCCCCUCCCAG
	外显子 9 STOP 2 (pos 7)	ACCTCCCGAGCAAACATGAC	ACCUCCCGAGCAAACAUGAC
	外显子 9 SD (pos 6)	CCTTACCTGTCATGTTTGCT	CCUUACCUGUCAUGUUUGCU
	外显子 10 SA (pos 5)	TGCTCTGGAGATGGAGAAGC	UGCUCUGGAGAUGGAGAAGC

[1001]

基因	说明	靶向序列	gRNA 间隔序列
	外显子 10 STOP 1 (pos 7)	CCCACCCAATGCCCGGCAGC	CCCACCCAAUGCCCGGCAGC
	外显子 10 STOP 2 (pos 4)	AGGCCATTTTGGAAGCTTGT	AGGCCAUUUUGGAAGCUUGU
	外显子 11 SA (pos 8)	ACCGGCTCTGCAAAGGCCAG	ACCGGCUCUGCAAAGGCCAG
	外显子 11 STOP 1 (pos 6)	TGGTGCAGGCCAGGCTGGAG	UGGUGCAGGCCAGGCUGGAG
	外显子 11 STOP 3 (pos 7)	GAACGGCAGCTGGCCCAAGG	GAACGGCAGCUGGCCCAAGG
	外显子 11 STOP 4 (pos 5)	GGCCCAAGGAGGCCTGGCTG	GGCCCAAGGAGGCCUGGCUG
	外显子 11 STOP 5 (pos 5)	GACACGAGTGATTGCTGTGC	GACACGAGUGAUUGCUGUGC
	外显子 11 STOP 5 (pos 6)	CTGGTCAGGGCAAGAGCTAT	CUGGUCAGGGCAAGAGCUAU
	外显子 11 STOP 5 (pos 8)	GGGCCCACAGCCACTCGTGG	GGGCCCACAGCCACUCGUGG
	外显子 11 STOP 6 (pos 4)	TTCCAGAAGAAGCTGCTCCG	UUCCAGAAGAAGCUGCUCCG
	外显子 11 STOP 7 (pos 8)	CCTGGTCCAGAGCCTGAGCA	CCUGGUCCAGAGCCUGAGCA
	外显子 11 STOP 8 (pos 8)	CAGACATCAAAGTACCCTAC	CAGACAUCAAAGUACCCUAC
	外显子 11 STOP 9 (pos 5)	ACATCAAAGTACCCTACAGG	ACAUCAAAGUACCCUACAGG
	外显子 11 STOP 10 (pos 4)	CGCCCAGGTCCTCACGTCTG	CGCCCAGGUCCUCACGUCUG
	外显子 11 STOP 11 (pos 8)	CTTAGTCCAACACCCACCGC	CUUAGUCCAACACCCACCGC
	外显子 11 STOP 12 (pos 8)	CCTCCTGCAATGCTTCCTGG	CCUCCUGCAAUGCUUCCUGG
	外显子 11 STOP 13 (pos 8)	GAGCCAGCCACAGGGCCCCC	GAGCCAGCCACAGGGCCCCC
	外显子 11 STOP 14 (pos 6)	GGAAGCAGAAGGTGCTTGCG	GGAAGCAGAAGGUGCUUGCG
	外显子 11 STOP 15 (pos 6)	GGCTGCAGCCGGGGACACTG	GGCUGCAGCCGGGGACACUG

[1002]

基因	说明	靶向序列	gRNA 间隔序列
	外显子 11 STOP 16 (pos 4)	CTGCCAAATTCCAGCCTCCT	CUGCCAAAUUCCAGCCUCCU
	外显子 11 STOP 17 (pos 8)	GGCGGGCCAAGACTTCTCCC	GGCGGGCCAAGACUUCUCCC
	外显子 12 STOP 1 (pos 6)	AGACTCAGAGGTGAGAGGAG	AGACUCAGAGGUGAGAGGAG
	外显子 14 SA (pos 4)	AGCCTAGGAGGCAAAGAGCA	AGCCUAGGAGGCAAAGAGCA
	外显子 14 STOP 1 (pos 5)	CCCCCAGGCTTTCCCCAAAC	CCCCCAGGCUUUCCCCAAAC
	外显子 14 SD (pos 4)	TCACTCCAGATGCTGCAGGG	UCACUCCAGAUGCUGCAGGG
	外显子 15 SA (pos 4)	AGGCTGCAGGTGGAATCAGA	AGGCUGCAGGUGGAAUCAGA
	外显子 15 STOP 1 (pos 8)	CTTCCCCCAGCTGAAGTCCT	CUUCCCCCAGCUGAAGUCCU
	外显子 15 SD (pos 7)	CACTCACTTGAGGGTTTCCA	CACUCACUUGAGGGUUUCCA
	外显子 16 SA (pos 5)	CAGACTGCGGGGACACAGTG	CAGACUGCGGGGACACAGUG
	外显子 16 SD 1 (pos 8)	CCACTCACCTTAGCCTGAGC	CCACUCACCUUAGCCUGAGC
	外显子 16 SD 2 (pos 7)	CACTCACCTTAGCCTGAGCA	CACUCACCUUAGCCUGAGCA
	外显子 17 SA (pos 8)	GTACAAGCTGTCGGAAACAG	GUACAAGCUGUCGGAAACAG
	外显子 17 SD 1 (pos 8)	ACACTCACTCCATCACCCGG	ACACUCACUCCAUCACCCGG
	外显子 17 SD 2 (pos 7)	CACTCACTCCATCACCCGGA	CACUCACUCCAUCACCCGGA
	外显子 18 STOP (pos 5)	CGTCCAGTACAACAAGTTCA	CGUCCAGUACAACAAGUUCA
	外显子 19 SA 1 (pos 8)	CCACATCCTGCAAGGGGGGA	CCACAUCCUGCAAGGGGGGA
	外显子 19 SA 2 (pos 7)	CACATCCTGCAAGGGGGGAT	CACAUCCUGCAAGGGGGGAU
	外显子 19 STOP 1 (pos 8)	TGGGCGTCCACATCCTGCAA	UGGGCGUCCACAUCCUGCAA

[1003]

基因	说明	靶向序列	gRNA 间隔序列
	外显子 19 STOP 2 (pos 7)	GGGCGTCCACATCCTGCAAG	GGGCGUCCACAUCCUGCAAG
	外显子 19 STOP 3 (pos 6)	GGCGTCCACATCCTGCAAGG	GGCGUCCACAUCCUGCAAGG
	外显子 19 STOP 4 (pos 5)	GCGTCCACATCCTGCAAGGG	GCGUCCACAUCCUGCAAGGG
CD7	外显子 1 STOP (pos 4)	GCCCAAGGTAAGAGCTTCCC	GCCCAAGGUAAGAGCUUCCC
	外显子 1 SD 1 (pos 8)	GCTCTTACCTTGGGCAGCCA	GCUCUUACCUUGGGCAGCCA
	外显子 1 SD 2 (pos 9)	AGCTCTTACCTTGGGCAGCC	AGCUCUUACCUUGGGCAGCC
	外显子 2 SA 1 (pos 8)	TGCACCTCTGGGGAGGACCT	UGCACCUCUGGGGAGGACCU
	外显子 2 SA 2 (pos 9)	CTGCACCTCTGGGGAGGACC	CUGCACCUCUGGGGAGGACC
	外显子 2 STOP 1 (pos 7)	CGCCTGCAGCTGTCGGACAC	CGCCUGCAGCUGUCGGACAC
	外显子 2 STOP 2 (pos 8)	CACCTGCCAGGCCATCACGG	CACCUGCCAGGCCAUCACGG
	外显子 2 SD 1 (pos 6)	CCCTACCTGTCACCAGGACC	CCCUACCUGUCACCAGGACC
	外显子 2 SD 2 (pos 5)	CCTACCTGTCACCAGGACCA	CCUACCUGUCACCAGGACCA
	外显子 3 SA (pos 4)	CCTCTGAGAAGGAAAAAAGA	CCUCUGAGAAGGAAAAAAGA
	外显子 3 STOP 1 (pos9)	CAGAGGAACAGTCCCAAGGA	CAGAGGAACAGUCCCAAGGA
CD33	外显子 1 SD 1 (pos 7)	CACTCACCTGCCCACAGCAG	CACUCACCUGCCCACAGCAG
	外显子 1 SD 2 (pos 8)	CCACTCACCTGCCCACAGCA	CCACUCACCUGCCCACAGCA
	外显子 1 SD (pos 9)	GCCACTCACCTGCCCACAGC	GCCACUCACCUGCCCACAGC
	外显子 2 SA 1 (pos 8)	AGGGCCCCTGTGGGGAAACG	AGGGCCCCUGUGGGGAAACG
	外显子 2 SA 2 (pos 7)	GGGCCCCTGTGGGGAAACGA	GGGCCCCUGUGGGGAAACGA

[1004]

基因	说明	靶向序列	gRNA 间隔序列
	外显子 2 STOP 1 (pos 8)	GCAAGTGCAGGAGTCAGTGA	GCAAGUGCAGGAGUCAGUGA
	外显子 2 STOP 2 (pos 6)	CGGAACCAGTAACCATGAAC	CGGAACCAGUAACCAUGAAC
	外显子 2 STOP 3 (pos 5)	GGAACCAGTAACCATGAACT	GGAACCAGUAACCAUGAACU
	外显子 2 STOP 4 (pos 4)	GAACCAGTAACCATGAACTG	GAACCAGUAACCAUGAACUG
	外显子 2 STOP 5 (pos 8)	GCTAGATCAAGAAGTACAGG	GCUAGAUCAAGAAGUACAGG
	外显子 2 STOP 6 (pos 8)	AGAAGTACAGGAGGAGACTC	AGAAGUACAGGAGGAGACUC
	外显子 3 SA 1 (pos 6)	CAAGTCTAGTGAGGAGAAAG	CAAGUCUAGUGAGGAGAAAG
	外显子 3 SA 2 (pos 5)	AAGTCTAGTGAGGAGAAAGA	AAGUCUAGUGAGGAGAAAGA
	外显子 3 SA 3 (pos 4)	AGTCTAGTGAGGAGAAAGAG	AGUCUAGUGAGGAGAAAGAG
	外显子 3 STOP 1 (pos 7)	ACAGGCCCAGGACACAGAGC	ACAGGCCCAGGACACAGAGC
	外显子 3 STOP 2 (pos 7)	ACCTGTCAGGTGAAGTTCGC	ACCUGUCAGGUGAAGUUCGC
	外显子 3 SD 1 (pos 6)	ACTTACAGGTGACGTTGAGC	ACUUACAGGUGACGUUGAGC
	外显子 4 SA 1 (pos 6)	AACATCTAGGAGAGGAAGAG	AACAUCUAGGAGAGGAAGAG
	外显子 4 STOP 1 (pos 7)	GTTCCACAGAACCCAACAAC	GUUCCACAGAACCCAACAAC
	外显子 4 SD 1 (pos 7)	TTCCTACCTGAGCCATCTCC	UUCCUACCUGAGCCAUCUCC
	外显子 5 SD (pos 8)	ATGCTCACATGAAGAAGATG	AUGCUCACAUGAAGAAGAUG
	外显子 5 STOP 1 (pos 7)	GGGAAACAAGAGACCAGAGC	GGGAAACAAGAGACCAGAGC
	外显子 6 SA 1 (pos 6)	TCACTCTGATGGGAGACACC	UCACUCUGAUGGGAGACACC
	外显子 6 SA 2 (pos 5)	CACTCTGATGGGAGACACCA	CACUCUGAUGGGAGACACCA

[1005]

基因	说明	靶向序列	gRNA 间隔序列
	外显子 6 SA 1 (pos 4)	TTTCTTATGGAGAGGAAAGA	UUUCUUAUGGAGAGGAAAGA
CD52	外显子 1 STOP (pos 4)	GTACAGGTAAGAGCAACGCC	GUACAGGUAAGAGCAACGCC
	外显子 1 SD (pos7)	CTCTTACCTGTACCATAACC	CUCUUACCUGUACCAUAACC
	外显子 1 SD (pos 4)	TTACCTGTACCATAACCAGG	UUACCUGUACCAUAACCAGG
	外显子 2 SA (pos 6)	TGTATCTGTAGGAGGAGAAG	UGUAUCUGUAGGAGGAGAAG
	外显子 2 SA (pos 5)	GTATCTGTAGGAGGAGAAGT	GUAUCUGUAGGAGGAGAAGU
	外显子 2 STOP (pos 7)	CAGATACAAACTGGACTCTC	CAGAUACAAACUGGACUCUC
CD123	外显子 1 SD (pos 6)	TCTTACCTTCCTTCGTTTGC	UCUUACCUUCCUUCGUUUGC
	外显子 2 SA 1 (pos 8)	TTTGGATCTAAAACGGTGAC	UUUGGAUCUAAAACGGUGAC
	外显子 2 SA 2 (pos 4)	GATCTAAAACGGTGACAGGT	GAUCUAAAACGGUGACAGGU
	外显子 2 STOP 1 (pos 8)	AAAGGCTCAGCAGTTGACCT	AAAGGCUCAGCAGUUGACCU
	外显子 2 SD (pos 6)	ATTTACCGGCATAGAATAGT	AUUUACCGGCAUAGAAUAGU
	外显子 3 SA (pos 8)	TCACTGCCTAAGAGAGACAT	UCACUGCCUAAGAGAGACAU
	外显子 3 STOP 1 (pos 6)	AGGATCCACGTGGAGAATGG	AGGAUCCACGUGGAGAAUGG
	外显子 3 STOP 2 (pos 5)	GGATCCACGTGGAGAATGGT	GGAUCCACGUGGAGAAUGGU
	外显子 3 SD (pos 6)	TCTCACTGTTCTCAGGGAAG	UCUCACUGUUCUCAGGGAAG
	外显子 4 STOP 1 (pos 6)	CCTGCCCAAGGCTTCCCACC	CCUGCCCAAGGCUUCCCACC
	外显子 4 STOP 2 (pos 5)	CTGCCCAAGGCTTCCCACCT	CUGCCCAAGGCUUCCCACCU
	外显子 5 SA 1 (pos 6)	GCCTGCTGCGGTAAGCGGTA	GCCUGCUGCGGUAAGCGGUA

[1006]

基因	说明	靶向序列	gRNA 间隔序列
	外显子 5 STOP 1 (pos 7)	GATGCTCAGGGAACACGTAT	GAUGCUCAGGGAACACGUAU
	外显子 5 STOP 2 (pos 5)	TTCTCAAAGTTCCCACATCC	UUCUCAAAGUUCCCACAUCC
	外显子 5 STOP 3 (pos 4)	TCACAGATTGGTGAGTAGCC	UCACAGAUUGGUGAGUAGCC
	外显子 7 SD (pos 5)	CTCACCTGTTCTGTGATTAC	CUCACCUGUUCUGUGAUUAC
	外显子 8 STOP 1 (pos 7)	TCCTTCCAGCTACTCAATCC	UCCUUCCAGCUACUCAAUCC
	外显子 8 STOP 2 (pos 8)	CACAGTACAAATAAGAGCCC	CACAGUACAAAUAAGAGCCC
	外显子 8 STOP 3 (pos 6)	CCCCCCAGCGCTTCGGTGAG	CCCCCCAGCGCUUCGGUGAG
	外显子 8 STOP 4 (pos 5)	CCCCCAGCGCTTCGGTGAGT	CCCCCAGCGCUUCGGUGAGU
	外显子 8 SD (pos 8)	CCACTCACCGAAGCGCTGGG	CCACUCACCGAAGCGCUGGG
	外显子 10 SA (pos 4)	TACCTCGGAGGAAAGAGAAA	UACCUCGGAGGAAAGAGAAA
	外显子 10 STOP (pos 8)	CAGCTTCCAAAACGACAAGC	CAGCUUCCAAAACGACAAGC
	外显子 10 SD (pos 7)	AACATACCAGCTTGTCGTTT	AACAUACCAGCUUGUCGUUU
	外显子 11 SA 1 (pos 8)	AGACCACCTGCAGAGACGAG	AGACCACCUGCAGAGACGAG
	外显子 11 SA 2 (pos 5)	CCACCTGCAGAGACGAGAGG	CCACCUGCAGAGACGAGAGG
TRBC1	外显子 1 STOP 1 (pos 8)	CCACACCCAAAAGGCCACAC	CCACACCCAAAAGGCCACAC
	外显子 1 STOP 2 (pos 5)	CCCACCAGCTCAGCTCCACG	CCCACCAGCUCAGCUCCACG
	外显子 1 STOP 3 (pos 7)	CGCTGTCAAGTCCAGTTCTA	CGCUGUCAAGUCCAGUUCUA
	外显子 1 STOP 4 (pos 6)	GCTGTCAAGTCCAGTTCTAC	GCUGUCAAGUCCAGUUCUAC
	外显子 1 STOP 5 (pos 5)	CACCCAGATCGTCAGCGCCG	CACCCAGAUCGUCAGCGCCG

[1007]

基因	说明	靶向序列	gRNA 间隔序列
	外显子 1 SD (pos 8)	CCACTCACCTGCTCTACCCC	CCACUCACCUGCUCUACCCC
	外显子 2 SA (pos 8)	CCACAGTCTGAAAGAAAGCA	CCACAGUCUGAAAGAAAGCA
	外显子 3 SA (pos 5)	GACACTGTTGGCACGGAGGA	GACACUGUUGGCACGGAGGA
	外显子 3 SD (pos 4)	TTACCATGGCCATCAACACA	UUACCAUGGCCAUCAACACA
TRBC2	外显子 1 STOP 1 (pos 8)	CCACACCCAAAAGGCCACAC	CCACACCCAAAAGGCCACAC
	外显子 1 STOP 2 (pos 5)	CCCACCAGCTCAGCTCCACG	CCCACCAGCUCAGCUCCACG
	外显子 1 STOP 3 (pos 7)	CGCTGTCAAGTCCAGTTCTA	CGCUGUCAAGUCCAGUUCUA
	外显子 1 STOP 4 (pos 6)	GCTGTCAAGTCCAGTTCTAC	GCUGUCAAGUCCAGUUCUAC
	外显子 1 STOP 5 (pos 5)	CACCCAGATCGTCAGCGCCG	CACCCAGAUCGUCAGCGCCG
	外显子 2 SA (pos 8)	CCACAGTCTGAAAGAAAACA	CCACAGUCUGAAAGAAAACA
	外显子 2 SA (pos 7)	CACAGTCTGAAAGAAAACAG	CACAGUCUGAAAGAAAACAG
	外显子 3 SD (pos 4)	TTACCATGGCCATCAGCACG	UUACCAUGGCCAUCAGCACG
	外显子 1 SD (pos 8)	CCACTCACCTGCTCTACCCC	CCACUCACCUGCUCUACCCC
CISH	外显子 1 STOP	TCTGCGTTCAGGGGTAAGCG	UCUGCGUUCAGGGGUAAGCG
	外显子 1 SD	GCGCTTACCCCTGAACGCAG	GCGCUUACCCCUGAACGCAG
	外显子 2 STOP 2	GACTGGGCAGCGGCCCCTGT	GACUGGGCAGCGGCCCCUGU
	外显子 2 STOP 1	GGACTGGGCAGCGGCCCCTG	GGACUGGGCAGCGGCCCCUG
	外显子 2 STOP 3	GTCATGCAGCCCTTGCCTGC	GUCAUGCAGCCCUUGCCUGC
	外显子 2 STOP 4	TCATGCAGCCCTTGCCTGCT	UCAUGCAGCCCUUGCCUGCU

[1008]

基因	说明	靶向序列	gRNA 间隔序列
	外显子 2 STOP 5	CATGCAGCCCTTGCCTGCTG	CAUGCAGCCCUUGCCUGCUG
	外显子 2 SD 1	CTCACCAGATTCCCGAAGGT	CUCACCAGAUUCCCGAAGGU
	外显子 2 SD 2	CAGACTCACCAGATTCCCGA	CAGACUCACCAGAUUCCCGA
	外显子 3 SA 1 (pos 4)	AGCCTAGGCAAGTGCAGAGG	AGCCUAGGCAAGUGCAGAGG
	外显子 3 SA 2 (pos 5)	CAGCCTAGGCAAGTGCAGAG	CAGCCUAGGCAAGUGCAGAG
	外显子 3 SA 3 (pos 7)	ACCAGCCTAGGCAAGTGCAG	ACCAGCCUAGGCAAGUGCAG
	外显子 3 STOP 1 (pos 8)	TGGAACCCCAATACCAGCCT	UGGAACCCCAAUACCAGCCU
	外显子 3 STOP 2 (pos 7)	CACCTGCAGAAGATGCCAGA	CACCUGCAGAAGAUGCCAGA
ACAT1	外显子 1 SD 1 (pos 7)	CGCTCACCTGCACCAGCCTC	CGCUCACCUGCACCAGCCUC
	外显子 3 SA (pos 5)	CTTCCTGGCAAGACACAAGA	CUUCCUGGCAAGACACAAGA
	外显子 3 STOP (pos 5)	AATTCAGGGAGCCATTGAAA	AAUUCAGGGAGCCAUUGAAA
	外显子 3 SD (pos 8)	CTACTGACCTGCCTTTTCAA	CUACUGACCUGCCUUUUCAA
	外显子 5 STOP (pos 7)	GCCTCTCAAAGTCTTATGTG	GCCUCUCAAAGUCUUAUGUG
	外显子 7 STOP (pos 4)	TTCCCATGCTGCTTTACTTC	UUCCCAUGCUGCUUUACUUC
	外显子 8 STOP (pos 8)	TTTAGGTCAACCAGATGTAG	UUUAGGUCAACCAGAUGUAG
	外显子 9 SA (pos 9)	TGTGCCTGAAAGCAAAAATG	UGUGCCUGAAAGCAAAAAUG
	外显子 9 SD (pos 4)	TTACCTACTATTCTTGCCAG	UUACCUACUAUUCUUGCCAG
	外显子 10 SA (pos 6)	AAATGCTGTTTAAAAAAGG	AAAUGCUGUUUAAAAAAGG
	外显子 11 STOP (pos 4)	CCCCAAAAAGTGAATATCAA	CCCCAAAAAGUGAAUAUCAA

[1009]

基因	说明	靶向序列	gRNA 间隔序列
Cyp11a1	外显子 1 STOP 1 (pos 4)	GTCCAGAATTTCCAGAAGTA	GUCCAGAAUUUCCAGAAGUA
	外显子 2 SA 1 (pos 4)	TCCCTGGAGGGGTGGGGGAG	UCCCUGGAGGGGUGGGGGAG
	外显子 2 SD 1 (pos 4)	TCACTTCAACAGGACTCCTA	UCACUUCAACAGGACUCCUA
	外显子 3 SD 1 (pos 6)	CCTTACACTCAAAGGCAAAG	CCUUACACUCAAAGGCAAAG
	外显子 4 SA (pos 5)	ATGGCTGCAGGGAGAGGAAG	AUGGCUGCAGGGAGAGGAAG
	外显子 4 STOP 1 (pos 8)	GGAGCGCCAGGGGATGCTGG	GGAGCGCCAGGGGAUGCUGG
	外显子 4 STOP 2 (pos 8)	TCACGTCCCATGCAGCCACA	UCACGUCCCAUGCAGCCACA
	外显子 6 SA (pos 8)	TGGACGTCTGGTGGGGAGTA	UGGACGUCUGGUGGGGAGUA
	外显子 8 STOP 1(pos 6)	ACTCACATTGATGAGGAAGA	ACUCACAUUGAUGAGGAAGA
	外显子 9 SA (pos 7)	CAGCATCTGAGAAAGGCAGA	CAGCAUCUGAGAAAGGCAGA
	外显子 9 STOP 1 (pos 5)	AATCCAACACCTCAGCGATG	AAUCCAACACCUCAGCGAUG
	外显子 9 STOP 2 (pos 4)	ATCCAACACCTCAGCGATGT	AUCCAACACCUCAGCGAUGU
GATA3	外显子 1 STOP 1 (pos 8)	CGCGGCGCAGTACCCGCTGC	CGCGGCGCAGUACCCGCUGC
	外显子 1 SD 1 (pos 7)	CACTCACCGTGGTGGGTCGG	CACUCACCGUGGUGGGUCGG
	外显子 1 SD 2 (pos 6)	ACTCACCGTGGTGGGTCGGA	ACUCACCGUGGUGGGUCGGA
	外显子 2 SA 1 (pos 8)	TGGCTCCCTGTGGGCAACG	UGGCUCCCUGUGGGCAACG
	外显子 2 STOP 2 (pos 5)	GATTCCAGGGGGAGGCGGTG	GAUUCCAGGGGGAGGCGGUG
	外显子 2 SD 1 (pos 8)	GCTCCTACCTGTGCTGGACC	GCUCCUACCUGUGCUGGACC
	外显子 3 STOP 1 (pos 7)	TCGCCGCCACAGTGGGGTCG	UCGCCGCCACAGUGGGGUCG

[1010]

基因	说明	靶向序列	gRNA 间隔序列
	外显子 4 SA (pos 5)	CAGACTGAGAGTGGGGAGAG	CAGACUGAGAGUGGGGAGAG
	外显子 4 STOP 1 (pos 7)	CCTCCTCCAGAGTGTGGTTG	CCUCCUCCAGAGUGUGGUUG
NR4A1	外显子 1 STOP 1 (pos 8)	AGCCATCCCAGGGAGAGAGC	AGCCAUCCCAGGGAGAGAGC
	外显子 1 STOP 2 (pos 7)	GCCATCCCAGGGAGAGAGCT	GCCAUCCCAGGGAGAGAGCU
	外显子 1 STOP 3 (pos 6)	CCATCCCAGGGAGAGAGCTG	CCAUCCCAGGGAGAGAGCUG
	外显子 1 STOP 4 (pos 5)	CTCACAGGCCACCCACCAGC	CUCACAGGCCACCCACCAGC
	外显子 2 STOP 1 (pos 8)	CCGCTTCCAGAAGTGCCTGG	CCGCUUCCAGAAGUGCCUGG
	外显子 2 STOP 2 (pos 5)	CTTCCAGAAGTGCCTGGCGG	CUUCCAGAAGUGCCUGGCGG
	外显子 3 SA 1 (pos 5)	ACAACTGCAAAGGAATGGGT	ACAACUGCAAAGGAAUGGGU
	外显子 3 SA 2 (pos 4)	CAACTGCAAAGGAATGGGTA	CAACUGCAAAGGAAUGGGUA
	外显子 4 SA (pos 4)	GAACTAGGAAGACGGTCCAG	GAACUAGGAAGACGGUCCAG
	外显子 4 STOP 1 (pos 8)	GGCTGACCAGGACCTGTTGC	GGCUGACCAGGACCUGUUGC
	外显子 4 SD 1 (pos 5)	CTCACCTGTACGCCAGGCGG	CUCACCUGUACGCCAGGCGG
	外显子 4 SD 2 (pos 8)	GCTCTCACCTGTACGCCAGG	GCUCUCACCUGUACGCCAGG
	外显子 5 SA (pos 8)	CTTAGACCTGGCAGGCAGAT	CUUAGACCUGGCAGGCAGAU
	外显子 5 STOP 1 (pos 5)	CAATCCAGTCCCCGAAGCCA	CAAUCCAGUCCCCGAAGCCA
	外显子 5 STOP 2 (pos 4)	AATCCAGTCCCCGAAGCCAC	AAUCCAGUCCCCGAAGCCAC
	外显子 5 SD 1 (pos 6)	ACTCACCGGTGATGAGGACA	ACUCACCGGUGAUGAGGACA
	外显子 5 SD 2 (pos 5)	CTCACCGGTGATGAGGACAA	CUCACCGGUGAUGAGGACAA

[1011]

基因	说明	靶向序列	gRNA 间隔序列
	外显子 6 SA (pos 6)	CCGGTCTGCGGGAAGGGTAC	CCGGUCUGCGGGAAGGGUAC
	外显子 6 STOP 1 (pos 8)	TGGGCTGCAGGAGCCGCGGC	UGGGCUGCAGGAGCCGCGGC
NR4A2	外显子 1 STOP 1 (pos 7)	TTGTACCAAATGCCCCTGTC	UUGUACCAAAUGCCCCUGUC
	外显子 1 STOP 2 (pos 8)	CGGACAGCAGTCCTCCATTA	CGGACAGCAGUCCUCCAUUA
	外显子 1 STOP 3 (pos 6)	AGGTGCAGCACAGCCCCATG	AGGUGCAGCACAGCCCCAUG
	外显子 1 STOP 4 (pos 5)	GGTGCAGCACAGCCCCATGT	GGUGCAGCACAGCCCCAUGU
	外显子 1 STOP 5 (pos 7)	AGTTGCCAGATGCGCTTCGA	AGUUGCCAGAUGCGCUUCGA
	外显子 1 STOP 6 (pos 6)	GTTGCCAGATGCGCTTCGAC	GUUGCCAGAUGCGCUUCGAC
	外显子 1 STOP 7 (pos 5)	GTCTCAGCTGCTCGACACGC	GUCUCAGCUGCUCGACACGC
	外显子 3 SD (pos 7)	TTCTTACCCTGGAATAGTCC	UUCUUACCCUGGAAUAGUCC
	外显子 4 SD (pos 5)	ATTACCTGTATGCTAATCGA	AUUACCUGUAUGCUAAUCGA
	外显子 5 STOP 1 (pos 4)	TTGCAATGCGTTCGTGGCTT	UUGCAAUGCGUUCGUGGCUU
	外显子 5 SD (pos 6)	ACTGACCTGTGACCATAGCC	ACUGACCUGUGACCAUAGCC
NR4A3	外显子 2 SA (pos 4)	TATCTGCAGGGACAGAGAAA	UAUCUGCAGGGACAGAGAAA
	外显子 2 STOP 1 (pos 8)	TGCGGCGCAGACATACAGCT	UGCGGCGCAGACAUACAGCU
	外显子 2 STOP 2 (pos 6)	CCCCGCAGGCGGGGGCGTTA	CCCCGCAGGCGGGGGCGUUA
	外显子 3 STOP 1 (pos 4)	TTTCAGAAGTGTCTCAGTGT	UUUCAGAAGUGUCUCAGUGU
	外显子 5 SD (pos 5)	ATTACCTGATGGAAAGTCTG	AUUACCUGAUGGAAAGUCUG
	外显子 6 STOP 1 (pos 4)	CTTCAGTGCCTTCGTGGATT	CUUCAGUGCCUUCGUGGAUU

[1012]

基因	说明	靶向序列	gRNA 间隔序列
	外显子 7 SA (pos 4)	TTTCTGCAGAGGGATAGAGA	UUUCUGCAGAGGGAUAGAGA
	外显子 7 STOP 1 (pos 8)	AGACCACCAGAGTAAGGGAC	AGACCACCAGAGUAAGGGAC
MCJ	外显子 1 STOP (pos 6)	ACTTGCAGCCCTCGGCCAAA	ACUUGCAGCCCUCGGCCAAA
FAS	外显子 1 SD (pos 9)	AGGGCTCACCAGAGGTAGGA	AGGGCUCACCAGAGGUAGGA
	外显子 3 SA (pos 6)	TTCACCTGCCCAAGGAAAAA	UUCACCUGCCCAAGGAAAAA
	外显子 4 SA (pos 7)	CTAAGCCTAGAAAATCAGTT	CUAAGCCUAGAAAAUCAGUU
	外显子 5 SA (pos 5)	ACATCTAGAAAAAAAAATAC	ACAUCUAGAAAAAAAAAUAC
	外显子 5 SD (pos 5)	ATTACCTTCCTCTTTGCACT	AUUACCUUCCUCUUUGCACU
	外显子 6 SA (pos 5)	GATCCTGTAGGTTGGAACAT	GAUCCUGUAGGUUGGAACAU
	外显子 6 STOP 1 (pos 4)	AAGCCACCCCAAGTTAGATC	AAGCCACCCCAAGUUAGAUC
	外显子 6 SD (pos 7)	AACTTACCCCAAACAATTAG	AACUUACCCCAAACAAUUAG
	外显子 7 SD (pos 8)	ATACCTACAGGATTTAAAGT	AUACCUACAGGAUUUAAAGU
	外显子 8 SA (pos 8)	GTTTCCTAGAAAGCAAAAAA	GUUUCCUAGAAAGCAAAAAA
	外显子 9 STOP 1 (pos 6)	AAGTTCAACTGCTTCGTAAT	AAGUUCAACUGCUUCGUAAU
	外显子 9 STOP (pos 5)	AATTCAGACTATCATCCTCA	AAUUCAGACUAUCAUCCUCA
SELPG/PSGL1	外显子 1 STOP 1 (pos 6)	GCTTGCAGCTGTGGGACACC	GCUUGCAGCUGUGGGACACC
	外显子 2 STOP 1 (pos 8)	GACCACTCAACCAGTGCCCA	GACCACUCAACCAGUGCCCA
	外显子 3 STOP 1 (pos 8)	GGAGGCACAGACCACTCCAC	GGAGGCACAGACCACUCCAC
	外显子 4 STOP 1 (pos 5)	GGCACAGACAACTCGACTGA	GGCACAGACAACUCGACUGA

[1013]

基因	说明	靶向序列	gRNA 间隔序列
	外显子 5 STOP 1 (pos 8)	GGAGGCACAGACCACTCCAC	GGAGGCACAGACCACUCCAC
	外显子 6 STOP 1 (pos 4)	GCACAGACCACTCAACCCAC	GCACAGACCACUCAACCCAC
	外显子 7 STOP 1 (pos 8)	GACCACTCAACCCACAGGCC	GACCACUCAACCCACAGGCC
	外显子 8 STOP 1 (pos 8)	GACCACTCAAACCACAGCCA	GACCACUCAAACCACAGCCA
	外显子 9 STOP 1 (pos 8)	GACCACTCAACCCACAGCCA	GACCACUCAACCCACAGCCA
	外显子 10 STOP 1 (pos 8)	GGAGGCACAGACCACTCCAC	GGAGGCACAGACCACUCCAC
	外显子 11 STOP 1 (pos 8)	GACCACTCAACCAGCAGCCA	GACCACUCAACCAGCAGCCA

[1014]

表 2B

基因	gRNA名称	gRNA靶	取向	靶标	预测结果
PDCD1	Ex. 1 SD	CACCTACCTAAGAACCATCC	反义	C7	剪接供体中断： GT → AT
PDCD1	Ex. 2 SA	GGAGTCTGAGAGATGGAGAG	反义	C6	剪接受体中断： AG → AA
PDCD1	Ex. 3 SA	TTCTCTCTGGAAGGGCACAA	反义	C7	剪接受体中断： AG → AA
PDCD1	Ex. 3 SD	GACGTTACCTCGTGCGGCCC	反义	C8	剪接受体中断： GT → AT
PDCD1	Ex. 4 SA	CCTGCAGAGAAACACACTTG	反义	C2	剪接受体中断： AG → AA
PDCD1	Ex. 2 pmSTOP	GGGGTTCCAGGGCCTGTCTG	反义	C7, C8	pmSTOP 中断：TGG (Trp) → TAG, TGA, TAA
PDCD1	Ex. 3 pmSTOP_1	CAGTTCCAAACCCTGGTGGT	有义	C7	pmSTOP 中断：CAA (Gln) → TAA
PDCD1	Ex. 3 pmSTOP_2	GGACCCAGACTAGCAGCACC	反义	C5, C6	pmSTOP 中断：TGG (Trp) → TAG, TGA, TAA
TRAC	Ex. 1 SD	CTTACCTGGGCTGGGAAGA	反义	C5	剪接供体中断： GT → AT
TRAC	Ex. 3 SA	TTCGTATCTGTAAAACCAAG	反义	C8	剪接受体中断： AG → AA
TRAC	Ex. 3 pmSTOP_1	TTTCAAAACCTGTCAGTGAT	有义	C4	pmSTOP 中断：CAA (Gln) → TAA
TRAC	Ex. 3 pmSTOP_2	TTCAAAACCTGTCAGTGATT	有义	C3	pmSTOP 中断：CAA (Gln) → TAA

[1015] 前间区序列邻近基序

[1016] 术语“前间区序列邻近基序(PAM)”或“PAM样基序”是指紧随在由CRISPR细菌适应性免疫系统中的Cas9核酸酶所靶向的DNA序列之后的2至6个间接对DNA序列。在一些实施方案中,PAM可以是5’PAM(即,定位在前间区序列5’末端的上游)。在其他实施方案中,PAM可以是3’PAM(即,定位在前间区序列5’末端的下游)。PAM序列对于靶标结合而言至关重要,但确切序列取决于Cas蛋白的类型。PAM序列可以是本领域中已知的任何PAM序列。合适的PAM序列包括但不限于,NGG、NGA、NGC、NGN、NGT、NGTT、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。

[1017] 本文所提供的碱基编辑器可包含衍生自CRISPR蛋白的结构域,该结构域能够结合

含有标准或非标准前间区序列邻近基序(PAM)序列的核苷酸序列。PAM位点是接近靶标多核苷酸序列的核苷酸序列。本公开的一些方面提供碱基编辑器,这些碱基编辑器包含具有不同PAM特异性的CRISPR蛋白中的全部或一部分。

[1018] 例如,Cas9蛋白,诸如来自酿脓链球菌的Cas9(spCas9),典型需要标准NGG PAM序列以结合特定核酸区域,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),且G是鸟嘌呤。PAM可以是CRISPR蛋白特异性的,并且可以在包含不同的衍生自CRISPR的结构域的不同碱基编辑器之间存在差异。PAM可以是靶标序列的5'或3'。PAM可以是靶标序列的上游或下游。PAM可以是1、2、3、4、5、6、7、8、9、10个或更多个核苷酸的长度。PAM经常是2至6个核苷酸之间的长度。

[1019] 在一些实施方案中,PAM是"NRN"PAM,其中"NRN"中的"N"是腺嘌呤(A)、胸腺嘧啶(T)、鸟嘌呤(G)或胞嘧啶(C),且R是腺嘌呤(A)或鸟嘌呤(G);或者PAM是"NYN"PAM,其中"NYN"中的"N"是腺嘌呤(A)、胸腺嘧啶(T)、鸟嘌呤(G)或胞嘧啶(C),且Y是胞苷(C)或胸腺嘧啶(T),例如,如R.T.Walton等人,2020,Science,10.1126/science.aba8853(2020)中所述,其完整内容以其整体并入本文。

[1020] 几种PAM变体在下表5中描述。

[1021] 表5.Cas9蛋白和对应的PAM序列

[1022]

变体	PAM
spCas9	NGG
spCas9-VRQR	NGA

[1023]

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’

[1024] 在一些实施方案中,PAM是NGC。在一些实施方案中,NGC PAM被Cas9变体识别。在一些实施方案中,NGC PAM变体包括选自D1135M、S1136Q、G1218K、E1219F、A1322R、D1332A、R1335E和T1337R(统称为“MQKFRAER”)的一个或多个氨基酸置换。

[1025] 在一些实施方案中,PAM是NGT。在一些实施方案中,NGT PAM被Cas9变体识别。在一些实施方案中,通过在一个或个残基1335、1337、1135、1136、1218和/或1219处的被靶向的突变产生NGT PAM变体。在一些实施方案中,通过在一个或个残基1219、1335、1337、1218处的被靶向的突变产生NGT PAM变体。在一些实施方案中,通过在一个或个残基1135、1136、1218、1219和1335处的被靶向的突变产生NGT PAM变体。在一些实施方案中,NGT PAM变体选自下表6A和表6B中提供的一组被靶向的突变。

[1026] 表6A:在残基1219、1335、1337、1218处的NGT PAM变体突变

[1027]

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	

[1028] 表6B:在残基1135、1136、1218、1219和1335处的NGT PAM变体突变

[1029]

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

[1030]

54					Y
55	N1286Q	I1331F			

[1031] 在一些实施方案中,NGT PAM变体选自表6A和表6B中的变体5、7、28、31或36。在一些实施方案中,变体具有改善的NGT PAM识别。

[1032] 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7中提供的变体中选择具有用于改善识别的突变的NGT PAM变体。

[1033] 表7:在残基1219、1335、1337和1218处的NGT PAM变体突变

[1034]

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

[1035] 在一些实施方案中,从下表5中提供的变体中选择NGT PAM。

[1036] 表8.NGT PAM变体

[1037]

	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

[1038]

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

[1039] 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。

[1040] 在一些实施方案中,Cas9结构域是来自酿脓链球菌的Cas9结构域(SpCas9)。在一些实施方案中,SpCas9结构域是核酸酶活性的SpCas9、无核酸酶活性的SpCas9(SpCas9d)、或SpCas9切口酶(SpCas9n)。在一些实施方案中,SpCas9包含D9X突变或本文所提供的任何氨基酸序列中的对应突变,其中X是除D之外的任何氨基酸。在一些实施方案中,SpCas9结构域包含D9A突变,或本文所提供的任何氨基酸序列中的对应突变。在一些实施方案中,SpCas9结构域、SpCas9d结构域或SpCas9n结构域可结合至具有非标准PAM的核酸序列。在一些实施方案中,SpCas9结构域、SpCas9d结构域或SpCas9n结构域可结合至具有NGG、NGA或NGCG PAM序列的核酸序列。

[1041] 在一些实施方案中,SpCas9结构域包含D1135X、R1335X和T1337X突变中的一者或多者,或本文所提供的任何氨基酸序列中的对应突变,其中X是任何氨基酸。在一些实施方案中,SpCas9结构域包含D1135E、R1335Q和T1337R突变中的一者或多者,或本文所提供的任何氨基酸序列中的对应突变。在一些实施方案中,SpCas9结构域包含D1135E、R1335Q和T1337R突变,或本文所提供的任何氨基酸序列中的对应突变。在一些实施方案中,SpCas9结构域包含D1135X、R1335X和T1337X突变中的一者或多者,或本文所提供的任何氨基酸序列中的对应突变,其中X是任何氨基酸。在一些实施方案中,SpCas9结构域包含D1135V、R1335Q和T1337R突变中的一者或多者,或本文所提供的任何氨基酸序列中的对应突变。在一些实施方案中,SpCas9结构域包含D1135V、R1335Q和T1337R突变,或本文所提供的任何氨基酸序列中的对应突变。在一些实施方案中,SpCas9结构域包含D1135X、G1218X、R1335X和T1337X突变中的一者或多者,或本文所提供的任何氨基酸序列中的对应突变,其中X是任何氨基酸。在一些实施方案中,SpCas9结构域包含D1135V、G1218R、R1335Q和T1337R突变中的一者或多者,或本文所提供的任何氨基酸序列中的对应突变。在一些实施方案中,SpCas9结构域包含D1135V、G1218R、R1335Q和T1337R突变,或本文所提供的任何氨基酸序列中的对应突变。

[1042] 在一些实施方案中,本文所提供的任何融合蛋白的Cas9结构域包含氨基酸序列,该氨基酸序列与本文所述的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结构域包含本文所述的任何Cas9多肽的氨基酸序列。在一些实施方案中,本文所提供的任何融合蛋白的Cas9结构域由本文所述的任何Cas9多肽的氨基酸序列组成。

[1043] 在一些实施例中,可以将被本文所公开的碱基编辑器的基于CRISPR蛋白的结构域所识别的PAM在独立寡核苷酸上提供至编码该碱基编辑器的插入物(例如,AAV插入物),从而将其提供至细胞。在此类实施方案中,提供在独立寡核苷酸上的PAM可允许切割不能以其他方式被切割的靶标序列,因为在与靶标序列相同的多核苷酸上不存在邻近PAM。

[1044] 在一种实施方案中,酿脓链球菌Cas9(SpCas9)可用作CRISPR核酸内切酶进行基因组工程改造。然而,可使用其他的。在一些实施方案中,不同的核酸内切酶可用来靶向某些基因组靶标。在一些实施方案中,可使用具有非NGG PAM序列的衍生自SpCas9的合成变体。

此外,已经鉴定了来自各种物种的Cas9直系同源物,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,相对大尺寸的SpCas9(大约4kb编码序列)可导致携带不能在细胞内有效表达的SpCas9cDNA的质粒。相反,金黄色葡萄球菌Cas9(SaCas9)的编码序列比SpCas9短大约1千碱基,可能允许其在细胞内有效地表达。类似于SpCas9,SaCas9核酸内切酶能够在体外修饰哺乳动物细胞中的靶标基因并且能够在体内修饰小鼠细胞中的靶标基因。在一些实施方案中,Cas蛋白可靶向不同的PAM序列。在一些实施方案中,例如,靶标基因可能与Cas9 PAM即5'-NGG相邻。在其他实施方案中,其他Cas9直系同源物可具有不同的PAM要求。例如,其他PAM诸如嗜热链球菌(S.thermophilus)的那些(对于CRISPR1为5'-NNAGAA,且对于CRISPR3为5'-NGGNG)和脑膜炎奈瑟氏菌(Neisseria meningitidis)的那些(5'-NNNNGATT)也可以发现与靶标基因相邻。

[1045] 在一些实施方案中,对于酿脓链球菌系统,靶标基因序列可在5'-NGG PAM之前(即,处于其5'),并且20-nt向导RNA序列可以与相反链进行碱基配对以寄到与PAM相邻的Cas9切割。在一些实施方案中,相邻切断可以在PAM上游的(约)3个碱基对处。在一些实施方案中,相邻切断可以在PAM上游的(约)10个碱基对处。在一些实施方案中,相邻切断可以在PAM上游的(约)0至20个碱基对处。例如,相邻切断可以在紧邻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个碱基对处。相邻切断也可以在PAM下游1至30个碱基对处。能够结合PAM的示例性SpCas9蛋白的序列如下:

[1046] 示例性的PAM结合SpCas9的氨基酸序列如下:

[1047] MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKP
IREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1048] 示例性的PAM结合SpCas9n的氨基酸序列如下:

[1049] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR

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

[1050] 示例性的PAM结合SpEQR Cas9的氨基酸序列如下：

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESVLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD
NSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG
LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKN
LSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ
SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQ
IHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET
ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNAS
LGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMK
QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDI
[1051] QKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMAREN
QTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQ
LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD
ENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK
LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPL
IETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA
RKKDWDPKKYGGF**E**SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLA
SHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR
DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK**Q**Y**R**STKEVLDATLIHQSITGLYE
TRIDLSQLGGD

[1052] 在上述序列中,残基E1135、Q1335和R1337,其可以从D1135、R1335和T1337突变以获得SpEQR Cas9,加下划线并以粗体显示。

[1053] 示例性的PAM结合SpVQR Cas9的氨基酸序列如下:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
[1054] FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD
NSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG

LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKN
LSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ
SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQ
IHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET
ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNAS
LGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMK
QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDI
QKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMAREN
QTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
[1055] ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQ
LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD
ENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK
LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPL
IETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA
RKKDWDPKKYGGF**V**SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLA
SHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR
DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK**Q**Y**R**STKEVLDATLIHQSITGLYE
TRIDLSQLGGD

[1056] 在上述序列中,残基V1135、Q1335和R1337,其可以从D1135、R1335和T1337突变以获得SpVQR Cas9,加下划线并以粗体显示。

[1057] 示例性的PAM结合SpVRER Cas9的氨基酸序列如下:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD
NSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG
LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKN
[1058] LSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ
SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQ
IHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET
ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNAS
LGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMK

QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDI
QKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMAREN
QTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQ
LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD
ENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK

[1059] LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPL
IETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA
RKKDWDPKKYGGF**V**SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA**R**ELQKGNELALPSKYVNFLYLA
SHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR
DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK**E**Y**R**STKEVLDATLIHQSITGLYE
TRIDLSQLGGD。

[1060] 在上述序列中,残基V1135、R1218、Q1335和R1337,其可以从D1135、G1218、R1335和T1337突变以获得SpVRER Cas9,加下划线并以粗体显示。

[1061] 在一些实施方案中,经工程改造的SpCas9变体能够识别侧翼具有3'H的前间区序列邻近基序(non-G PAM)(参见表3A至3D)。在一些实施方案中,SpCas9变体识别NRNH PAM(其中R是A或G,且H是A、C或T)。在一些实施方案中,non-G PAM是NRRH、NRTH或NRCH(参见例如,Miller,S.M.,等人Continuous evolution of SpCas9 variants compatible with non-G PAMs,Nat.Biotechnol.(2020),其内容通过引用以其整体并入本文)。

[1062] 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、无核酸酶活性的SpyMacCas9(SpyMacCas9d)、或SpyMacCas9切口酶(SpyMacCas9n)。在一些实施方案中,SaCas9结构域、SaCas9d结构域或SaCas9n结构域可结合至具有非标准PAM的核酸序列。在一些实施方案中,SpyMacCas9结构域、SpCas9d结构域、或SpCas9n构域可结合至具有NAA PAM序列的核酸序列。

[1063] 具有天然5'-NAAN-3'PAM特异性的猴链球菌(Streptococcus macacae)中SpyCas9的示例性Cas9 A同源物的序列是本领域中已知的并且由例如Jakimo等人,(www.biorxiv.org/content/biorxiv/early/2018/09/27/429654.full.pdf)描述,且提供如下。

[1064] SpyMacCas9

[1065] MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQIYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI

PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNEL
TKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDRGMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGHSLHEQIANLAGSPAIKKGILQTVKIVDELVKVMG
HKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELD
INRLSDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAER
GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNY
HHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI
RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEIQTVGQNGGLFDDNPKSPLEVTPSKLVPLKKELN
PKKYGGYQKPTTAYPVLLITDTKQLIPISVMNKKQFEQNPVKFLRDRGYQQVGKNDFIKLPKYTLVDIGDGIKRLW
ASSKEIHKGNQLVVSKKSQILLYHAHHLDSDLSNDYLQNHNQQFDVLFNEIISFSKKCKLGKEHIQKIENVYSNKK
NSASIEELAESFIKLLGFTQLGATSPFNFLGVKLNQKQYKGKKDYILPCTEGTLIRQSITGLYETRVDLSKIGED.

[1066] 在一些实施方案中,变体Cas9蛋白携带H840A、P475A、W476A、N477A、D1125A、W1126A和D1218A突变,使得该多肽切割靶标DNA或RNA的能力降低。这种Cas9蛋白切割靶标DNA(例如,单链靶标DNA)的能力降低,但保持结合靶标DNA(例如,单链靶标DNA)的能力。作为另一非限制性示例,在一些实施方案中,变体Cas9蛋白携带D10A、H840A、P475A、W476A、N477A、D1125A、W1126A和D1218A突变,使得该多肽切割靶标DNA的能力降低。这种Cas9蛋白切割靶标DNA(例如,单链靶标DNA)的能力降低,但保持结合靶标DNA(例如,单链靶标DNA)的能力。在一些实施方案中,当变体Cas9蛋白携带W476A和W1126A突变时或当变体Cas9蛋白携带P475A、W476A、N477A、D1125A、W1126A和D1218A突变时,变体Cas9蛋白不有效地结合至PAM序列。因此,在一些此类情况下,当这种变体Cas9蛋白在结合方法中使用时,该方法不需要PAM序列。换言之,在一些实施方案中,当这种变体Cas9蛋白在结合方法中使用时,该方法可包括向导RNA,但该方法可在PAM序列不存在下执行(并且结合特异性因此由向导RNA的靶向链段提供)。其他残基可经突变以实现上述效果(即,将一个或其他核酸酶蛋白灭活)。作为非限制性示例,残基D10、G12、G17、E762、H840、N854、N863、H982、H983、A984、D986和/或A987可改变(即,置换)。而且,除丙氨酸置换之外的突变是合适的。

[1067] 在一些实施方案中,碱基编辑器的衍生自CRISPR蛋白的结构域可包含具有标准PAM序列(NGG)的Cas9蛋白的全部或一部分。在其他实施方案中,碱基编辑器的衍生自Cas9的结构域可采用非标准PAM序列。此类序列已经在本领域中有所描述,并且对于熟练技术人员将会是显而易见的。例如,结合非标准PAM序列的Cas9结构域已经描述于以下文献中:Kleinstiver,B.P.,等人,"Engineered CRISPR-Cas9nucleases with altered PAM specificities"Nature 523,481-485(2015);和Kleinstiver,B.P.,等人,"Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition"Nature Biotechnology 33,1293-1298(2015);R.T.Walton等人"Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants"Science10.1126/science.aba8853(2020);Hu等人"Evolved Cas9 variants with broad PAM compatibility and high DNA specificity,"Nature,2018Apr.5,556(7699),57-63;Miller等人,"Continuous evolution of SpCas9 variants compatible with non-G PAMs"Nat.Biotechnol.,2020Apr;38(4):471-481;各自的整体内容通过引用

并入本文。

[1068] 具有降低的排他性的Cas9结构域

[1069] 典型地,Cas9蛋白,诸如来自酿脓链球菌的Cas9(spCas9),需要标准NGG PAM序列以结合特定核酸区域,其中"NGG"中的"N"是腺苷(A)、胸苷(T)或胞苷(C),且G是鸟苷。这可能限制编辑基因组内所希望的碱基的能力。在一些实施方案中,本文所提供的碱基编辑融合蛋白可能需要放置在精确位置处,例如,包含处于PAM上游的靶标甲基的区域。参见例如,Komor,A.C.,等人,"Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage"Nature 533,420-424(2016),其整体内容通过引用并入本文。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Cas9结构域,该结构域能够结合部含有标准PAm序列(例如,NGG)的核苷酸序列。结合至非标准PAM序列的Cas9结构域已经在本领域中有所描述,并且对于熟练技术人员将会是显而易见的。例如,结合非标准PAM序列的Cas9结构域已经在Kleinstiver,B.P.,等人,"Engineered CRISPR-Cas9 nucleases with altered PAM specificities"Nature 523,481-485(2015)和Kleinstiver,B.P.,等人,"Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition"NatureBiotechnology33,1293-1298(2015)中描述,各自的整体内容通过引用并入本文。

[1070] 具有内部插入的融合蛋白

[1071] 本文提供融合蛋白,该融合蛋白包含融合至核酸可编程核酸结合结构域(例如,napDNAbp)的异源多肽。异源多肽可以是未见于天然或野生型napDNAbp多肽序列中的多肽。异源多肽可以在napDNAbp的C末端处、在napDNAbp的N末端处融合至napDNAbp,或插入在napDNAbp的内部位置处。在一些实施方案中,异源多肽插入在napDNAbp的内部位置处。在一些实施方案中,医院多肽是脱氨酶或其功能性片段。例如,融合多肽可包含位于Cas9或Cas12(例如,Cas12b/C2c1)多肽的N端片段和C端片段侧翼的脱氨酶。融合蛋白中的脱氨酶可以是胞苷脱氨酶。在一些实施方案中,胞苷脱氨酶是APOBEC脱氨酶(例如,APOBEC1)。融合蛋白中的脱氨酶可以是腺苷脱氨酶。在一些实施方案中,腺苷脱氨酶是TadA(例如,TadA*7.10或TadA*8)。在一些实施方案中,TadA是TadA*8或TadA*9。如本文所述的TadA序列(例如,TadA7.10或TadA*8)是适用于上述融合蛋白的脱氨酶。

[1072] 在一些实施方案中,融合蛋白包含以下结构:

[1073] NH2-[napDNAbp的N端片段]-[脱氨酶]-[napDNAbp的C端片段]-COOH;

[1074] NH2-[Cas9的N端片段]-[腺苷脱氨酶]-[Cas9的C端片段]-COOH;

[1075] NH2-[Cas12的N端片段]-[腺苷脱氨酶]-[Cas12的C端片段]-COOH;

[1076] NH2-[Cas9的N端片段]-[胞苷脱氨酶]-[Cas9的C端片段]-COOH;

[1077] NH2-[Cas12的N端片段]-[胞苷脱氨酶]-[Cas12的C端片段]-COOH;

[1078] 其中,"]-["的每个实例是任选的连接子。

[1079] 脱氨酶可以是环状排列的脱氨酶。例如,脱氨酶可以是环状排列的腺苷脱氨酶。在一些实施方案中,脱氨酶是环状排列的TadA,在如TadA参考序列中编号的氨基酸残基116处环状排列。在一些实施方案中,脱氨酶是环状排列的TadA,在如TadA参考序列中编号的氨基酸残基136处环状排列。在一些实施方案中,脱氨酶是环状排列的TadA,在如TadA参考序列中编号的氨基酸残基65处环状排列。

[1080] 融合蛋白可包含超过一个脱氨酶。融合蛋白可包含,例如,1、2、3、4、5个或更多脱氨酶。在一些实施方案中,融合蛋白包含一个脱氨酶。在一些实施方案中,融合蛋白包含两个脱氨酶。融合蛋白中的两个或更多个脱氨酶可以是腺苷脱氨酶、胞苷脱氨酶、或其组合。两个或更多个脱氨酶可以是同二聚体。两个或更多个脱氨酶可以是异二聚体。两个或更多个脱氨酶可串联插入到apDNAbp中。在一些实施方案中,两个或更多个脱氨酶在apDNAbp中可以不是串联方式。

[1081] 在一些实施方案中,融合蛋白中的napDNAbp是Cas9多肽或其片段。Cas9多肽可以是变体Cas9多肽。在一些实施方案中,Cas9多肽是Cas9切口酶(nCas9)多肽或其片段。在一些实施方案中,Cas9多肽是核酸酶死亡的Cas9(dCas9)多肽或其片段。融合蛋白中的Cas9多肽可以是全长Cas9多肽。在一些情况下,融合蛋白中的Cas9多肽可以不是全长Cas9多肽。Cas9多肽可以在例如N端或C端相对于天然存在的Cas9蛋白截短。Cas9多肽可以是环状排列的Cas9蛋白。Cas9多肽可以是Cas9多肽的片段、部分或结构域,其仍能够结合靶标核苷酸和向导核酸序列。

[1082] 在一些实施方案中,Cas9多肽是酿脓链球菌Cas9(SpCas9)、金黄色葡萄球菌Cas9(SaCas9)、嗜热链球菌1Cas9(St1Cas9),或其片段或变体。

[1083] 融合蛋白的Cas9多肽可包含氨基酸序列,该氨基酸序列与天然存在的Cas9多肽至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%、或至少99.5%相同。

[1084] 融合蛋白的Cas9多肽可包含氨基酸序列,该氨基酸序列与下文详述的Cas9氨基酸序列(下文中称为“Cas9参考序列”)至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%、或至少99.5%相同:

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD
NSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG
LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKN
LSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ
SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQ
IHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET
ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNAS
LGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMK
[1085] QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDI
QKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMAREN
QTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQ
LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD
ENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK
LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPL
IETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA
RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLA
SHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR
DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYE
TRIDLSQLGGD (单下划线:HNH结构域;双下划线:RuvC结构域)。

[1086] 包含侧翼为Cas9多肽的N端和C端的异源催化性结构域的融合蛋白也可用于在本文所述的方法中用于碱基编辑。包含Cas9和一个或多个脱氨酶结构域(例如,腺苷脱氨酶)或者包含侧翼具有Cas9序列的腺苷脱氨酶结构域的融合蛋白也可以用于靶标序列的高特异性且高效的碱基编辑。在一种实施方案中,嵌合Cas9融合蛋白含有插入到Cas9多肽内的异源催化性结构域(例如,腺苷脱氨酶、胞苷脱氨酶、或腺苷脱氨酶和胞苷脱氨酶)。在一些实施方案中,融合蛋白包含插入到Cas9内的腺苷脱氨酶结构域和胞苷脱氨酶结构域。在一些实施方案中,腺苷脱氨酶融合到Cas9内,并且胞苷脱氨酶融合至C端。在一些实施方案中,腺苷脱氨酶融合到Cas9内,并且胞苷脱氨酶融合至N端。在一些实施方案中,胞苷脱氨酶融合到Cas9内,并且腺苷脱氨酶融合至C端。在一些实施方案中,胞苷脱氨酶融合到Cas9内,并且腺苷脱氨酶融合至N端。

[1087] 具有腺苷脱氨酶和胞苷脱氨酶以及Cas9的融合蛋白的示例性结构提供如下:

[1088] NH_2-[Cas9(腺苷脱氨酶)]-[胞苷脱氨酶]-COOH;

[1089] NH_2-[胞苷脱氨酶]-[Cas9(腺苷脱氨酶)]-COOH;

[1090] NH_2-[Cas9(胞苷脱氨酶)]-[腺苷脱氨酶]-COOH;或

[1091] NH_2-[腺苷脱氨酶]-[Cas9(胞苷脱氨酶)]-COOH。

[1092] 在一些实施方案中上述通用架构中所用的“-”表示任选的连接子的存在。

[1093] 在各种实施方案中,催化性结构域具有DNA修饰活性(例如,脱氨酶活性),诸如腺苷脱氨酶活性。在一些实施方案中,腺苷脱氨酶是TadA(例如,TadA*7.10)。在一些实施方案中,TadA是TadA*8。在一些实施方案中,TadA*8融合到Cas9内,并且胞苷脱氨酶融合至C端。在一些实施方案中,TadA*8融合到Cas9内,并且胞苷脱氨酶融合至N端。在一些实施方案中,胞苷脱氨酶融合到Cas9内,并且TadA*8融合至C端。在一些实施方案中,胞苷脱氨酶融合到Cas9内,并且TadA*8融合至N端。具有TadA*8和胞苷脱氨酶以及Cas9的融合蛋白的示例性结构提供如下:

[1094] NH_2-[Cas9(TadA*8)]-[胞苷脱氨酶]-COOH;

[1095] NH_2-[胞苷脱氨酶]-[Cas9(TadA*8)]-COOH;

[1096] NH_2-[Cas9(TadA*8)]-[TadA*8]-COOH;或

[1097] NH_2-[TadA*8]-[Cas9(TadA*8)]-COOH。

[1098] 在一些实施方案中上述通用架构中所用的“-”表示任选的连接子的存在。

[1099] 异源多肽(例如,脱氨酶)可以在合适位置插入到napDNAbp(例如,Cas9或Cas12(例如,Cas12b/C2c1))内,例如,使得napDNAbp保持其结合靶标多核苷酸和向导核酸的能力。脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或腺苷脱氨酶和胞苷脱氨酶)可以插入到napDNAbp内而不破坏该脱氨酶的功能(例如,碱基编辑活性)或该napDNAbp的功能(例如,结合至靶标核酸和向导核酸的能力)。脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或腺苷脱氨酶颚胞苷脱氨酶)可以插入到napDNAbp内,插入位置在例如如晶体学研究所示的无序区处或包含高温因子或B因子的区域处。蛋白质的有序性较差的区域、无序区或结构凌乱的区域(例如,暴露于溶剂的区域和环)可以用于插入而不破坏结构或功能。脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或腺苷脱氨酶和胞苷脱氨酶)可以插入到napDNAbp内的柔性环区域或暴露于溶剂的区域中。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或腺苷脱氨酶和胞苷脱氨酶)插入到Cas9或Cas12b/C2c1多肽的柔性环内。

[1100] 在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或腺苷脱氨酶和胞苷脱氨酶)的插入位置通过Cas9多肽的晶体结构的B因子分析来确定。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或腺苷脱氨酶和胞苷脱氨酶)插入到Cas9多肽的区域中,该区域包含高于平均值的B因子(例如,与总蛋白或包含无序区的蛋白结构域相比,更高的B因子)。B因子或温度因子可表示原子自其平均位置的波动(例如,作为晶格中温度依赖性原子振动或静态无序的结果)。骨架原子的高B因子(例如,高于平均B因子)可能是具有相对高的局部移动性的区域的标志。这种区域可用于插入脱氨酶而不破坏结构或功能。脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或腺苷脱氨酶和胞苷脱氨酶)可以插入在具有以下残基的位置处,该残基具有Cα原子,其该原子的B因子比总蛋白的平均B因子高50%、60%、70%、80%、90%、100%、110%、120%、130%、140%、150%、160%、170%、180%、190%、200%、或多于200%。脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或腺苷脱氨酶和胞苷脱氨酶)可以插入在具有以下残基的位置处,该残基具有Cα原子,其该原子的B因子比包含该残基的Cas9蛋

白结构域的平均B因子高50％、60％、70％、80％、90％、100％、110％、120％、130％、140％、150％、160％、170％、180％、190％、200％、或多于200％。包含高于平均值的B因子的Cas9多肽位置可包括，例如，氨基酸残基768、792、1052、1015、1022、1026、1029、1067、1040、1054、1068、1246、1247和1248，如上述Cas9参考序列中编号。包含高于平均值的B因子的Cas9多肽区域可包括，例如，氨基酸残基792至872、792至906、和2至791，如上述Cas9参考序列中编号。

[1101] 异源多肽(例如，脱氨酶)可在选自由以下所组成的群组的氨基酸残基处插入到napDNAbp中：768、791、792、1015、1016、1022、1023、1026、1029、1040、1052、1054、1067、1068、1069、1246、1247和1248，如上述Cas9参考序列中编号，或另一Cas9多肽中的对应氨基酸残基。在一些实施方案中，异源多肽插入到以下氨基酸位置之间：768至769、791至792、792至793、1015至1016、1022至1023、1026至1027、1029至1030、1040至1041、1052至1053、1054至1055、1067至1068、1068至1069、1247至1248、或1248至1249，如上述Cas参考序列中编号或其对应氨基酸位置。在一些实施方案中，异源多肽插入到以下氨基酸位置之间：769至770、792至793、793至794、1016至1017、1023至1024、1027至1028、1030至1031、1041至1042、1053至1054、1055至1056、1068至1069、1069至1070、1248至1249、或1249至1250，如上述Cas参考序列中编号或其对应氨基酸位置。在一些实施方案中，异源多肽代替选自由以下所组成的群组的氨基酸残基：768、791、792、1015、1016、1022、1023、1026、1029、1040、1052、1054、1067、1068、1069、1246、1247和1248，如上述Cas9参考序列中编号，或另一Cas9多肽中的对应氨基酸残基。应理解，对于上述Cas9参考序列的关于插入位置的参考用于说明性目的。本文所讨论的插入不限于上述Cas9参考序列的Cas9多肽序列，而是包括在变体Cas9多肽(例如，Cas9切口酶(nCas9)、核酸酶死亡的Cas9(dCas9)、缺少核酸酶结构域的Cas9变体、截短的Cas9、或缺少部分或完整HNH结构域的Cas9结构域)中的对应位置处的插入。

[1102] 异源多肽(例如，脱氨酶)可在选自由以下所组成的群组的氨基酸残基处插入到napDNAbp中：768、792、1022、1026、1040、1068和1247，如上述Cas9参考序列中编号，或另一Cas9多肽中的对应氨基酸残基。在一些实施方案中，异源多肽插入到以下氨基酸位置之间：768至769、792至793、1022至1023、1026至1027、1029至1030、1040至1041、1068至1069、或1247至1248，如上述Cas参考序列中编号或其对应氨基酸位置。在一些实施方案中，异源多肽插入到以下氨基酸位置之间：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，如上述Cas参考序列中编号或其对应氨基酸位置。在一些实施方案中，异源多肽代替选自由以下所组成的群组的氨基酸残基：768、792、1022、1026、1040、1068和1247，如上述Cas9参考序列中编号，或另一Cas9多肽中的对应氨基酸残基。

[1103] 异源多肽(例如，脱氨酶)可在如本文所述的氨基酸残基处插入到napDNAbp内，或在另一Cas9多肽的对应氨基酸残基处。在一种实施方案中，异源多肽(例如，脱氨酶)可在选自由以下所组成的群组的氨基酸残基处插入到napDNAbp中：1002、1003、1025、1052至1056、1242至1247、1061之1077、943至947、686至691、569至578、530至539和1060至1077，如上述Cas9参考序列中编号，或另一Cas9多肽中的对应氨基酸残基。脱氨酶(例如，腺苷脱氨酶、胞苷脱氨酶、或腺苷脱氨酶和胞苷脱氨酶)可插入在残基的N端或C端或替代该残基。在一些实施方案中，脱氨酶(例如，腺苷脱氨酶、胞苷脱氨酶、或腺苷脱氨酶和胞苷脱氨酶)插入到残

基的C端处。

[1104] 在一些实施方案中,腺苷脱氨酶(例如,TadA)插入在选自由以下所组成的群组的氨基酸残基处:1015、1022、1029、1040、1068、1247、1054、1026、768、1067、1248、1052和1246,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基。在一些实施方案中,腺苷脱氨酶(例如,TadA)插入残基792至872、792至906、或2至791的位置处,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基。在一些实施方案中,腺苷脱氨酶插入在选自由以下所组成的群组的氨基酸的N端处:1015、1022、1029、1040、1068、1247、1054、1026、768、1067、1248、1052和1246,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基。在一些实施方案中,腺苷脱氨酶插入在选自由以下所组成的群组的氨基酸的C端处:1015、1022、1029、1040、1068、1247、1054、1026、768、1067、1248、1052和1246,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基。在一些实施方案中,腺苷脱氨酶插入以替换选自由以下所组成的群组的氨基酸:1015、1022、1029、1040、1068、1247、1054、1026、768、1067、1248、1052和1246,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基。

[1105] 在一些实施方案中,胞苷脱氨酶(例如,APOBEC1)插入在选自由以下所组成的群组的氨基酸残基处:1016、1023、1029、1040、1069和1247,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基。在一些实施方案中,胞苷脱氨酶插入在选自由以下所组成的群组的氨基酸的N端处:1016、1023、1029、1040、1069和1247,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基。在一些实施方案中,胞苷脱氨酶插入在选自由以下所组成的群组的氨基酸的C端处:1016、1023、1029、1040、1069和1247,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基。在一些实施方案中,胞苷脱氨酶插入以替换选自由以下所组成的群组的氨基酸:1016、1023、1029、1040、1069和1247,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基。

[1106] 在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入在氨基酸残基768处,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入在氨基酸残基768的N端处,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入在氨基酸残基768的C端处,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入以替换氨基酸残基768,如上述Cas9参考序列中编号或另一Cas9多肽中的对应氨基酸残基。

[1107] 在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入在氨基酸残基791处或插入在氨基酸残基792处,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入在氨基酸残基791的N端处或插入在氨基酸792的N端处,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入在氨基酸791的C端处或插入在氨基酸792的N端处,如上述Cas9参考序列中编号,或另

一Cas9多肽中的对应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入以替换氨基酸791或插入以替换氨基酸792,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基。

[1108] 在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入在氨基酸残基1016处,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入在氨基酸残基1016的N端处,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入在氨基酸残基1016的C端处,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入以替换氨基酸残基1016,如上述Cas9参考序列中编号或另一Cas9多肽中的对应氨基酸残基。

[1109] 在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入在氨基酸残基1022处或插入在氨基酸残基1023处,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入在氨基酸残基1022的N端处或插入在氨基酸残基1023的N端处,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入在氨基酸残基1022的C端处或插入在氨基酸残基1023的C端处,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入以替换氨基酸残基1022或插入以替换氨基酸残基1023,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基。

[1110] 在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入在氨基酸残基1026处或插入在氨基酸残基1029处,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入在氨基酸残基1026的N端处或插入在氨基酸残基1029的N端处,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入在氨基酸残基1026的C端处或插入在氨基酸残基1029的C端处,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入以替换氨基酸残基1026或插入以替换氨基酸残基1029,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基。

[1111] 在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入在氨基酸残基1040处,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入在氨基酸残基1040的N端处,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨

酶、或者腺苷脱氨酶和胞苷脱氨酶)插入在氨基酸残基1040的C端处,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入以替换氨基酸残基1040,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基。

[1112] 在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入在氨基酸残基1052处或插入在氨基酸残基1054处,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入在氨基酸残基1052的N端处或插入在氨基酸残基1054的N端处,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入在氨基酸残基1052的C端处或插入在氨基酸残基1054的C端处,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入以替换氨基酸残基1052或插入以替换氨基酸残基1054,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基。

[1113] 在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入在氨基酸残基1067处、或插入在氨基酸残基1068处、或插入在氨基酸残基1069处,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入在氨基酸残基1067的N端处、或插入在氨基酸残基1068的N端处、或插入在氨基酸残基1069的N端处,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入在氨基酸残基1067的C端处、或插入在氨基酸残基1068的C端处、或插入在氨基酸残基1069的C端处,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入以替换氨基酸残基1067、或插入以替换氨基酸残基1068、或插入以替换氨基酸残基1069,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基。

[1114] 在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入在氨基酸残基1246处、或插入在氨基酸残基1247处、或插入在氨基酸残基1248处,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入在氨基酸残基1246的N端处、或插入在氨基酸残基1247的N端处、或插入在氨基酸残基1248的N端处,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入在氨基酸残基1246的C端处、或插入在氨基酸残基1247的C端处、或插入在氨基酸残基1248的C端处,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)插入以替换氨基酸残基1246、或插入以替换氨基酸残基1247、或插入以替换氨基酸残基1248,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基。

[1115] 在一些实施方案中,异源多肽(例如,脱氨酶)插入Cas9多肽的柔性环内。柔性环部分可以选自由以下项所组成的群组: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,如上述Cas9参考序列中编号,或另一Cas9多肽的对应氨基酸残基。柔性环部分可选自由以下项所组成的群组:1至529、538至568、580至685、692至942、948至1001、1026至1051、1078至1231、或1248至1297,如上述Cas9参考序列中编号,或另一Cas9多肽的对应氨基酸残基。

[1116] 异源多肽(例如,腺苷脱氨酶)可插入对应于以下氨基酸残基的Cas9多肽区内:1017至1069、1242至1247、1052至1056、1060至1077、1002-1003、943至947、530至537、568至579、686至691、1242至1247、1298-1300、1066至1077、1052至1056、或1060至1077,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基。

[1117] 异源多肽(例如,腺苷脱氨酶)可插入Cas9多肽的缺失区处。缺失区可对应于Cas9多肽的N端或C端部分。在一些实施方案中,缺失区对应于残基792至872,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基。在一些实施方案中,缺失区对应于残基792至906,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基。在一些实施方案中,缺失区对应于残基2至791,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基。在一些实施方案中,缺失区对应于残基1017至1069,如上述Cas9参考序列中编号,或其对应氨基酸残基。

[1118] 示例性的内部融合蛋白碱基编辑器提供在下表9中:

[1119] 表9:Cas9蛋白中的插入基因座

[1120]

BE ID	修饰	其他ID
IBE001	Cas9 TadA ins 1015	ISLAY01
IBE002	Cas9 TadA ins 1022	ISLAY02
IBE003	Cas9 TadA ins 1029	ISLAY03
IBE004	Cas9 TadA ins 1040	ISLAY04
IBE005	Cas9 TadA ins 1068	ISLAY05
IBE006	Cas9 TadA ins 1247	ISLAY06
IBE007	Cas9 TadA ins 1054	ISLAY07
IBE008	Cas9 TadA ins 1026	ISLAY08
IBE009	Cas9 TadA ins 768	ISLAY09
IBE020	δHNH TadA 792	ISLAY20
IBE021	N-端融合单TadA螺旋截短的165末端	ISLAY21
IBE029	TadA-环状排列116ins1067	ISLAY29
IBE031	TadA-环状排列136ins1248	ISLAY31
IBE032	TadA-环状排列136ins 1052	ISLAY32
IBE035	δ792-872TadA ins	ISLAY35
IBE036	δ792-906TadA ins	ISLAY36
IBE043	TadA-环状排列65ins1246	ISLAY43
IBE044	TadA ins C端截短2 791	ISLAY44

[1121] 异源多肽(例如,脱氨酶)可插入Cas9多肽的结构性或功能性结构域内。异源多肽

(例如,脱氨酶)可插入Cas9多肽的两个结构性或功能性结构域之间。异源多肽(例如,脱氨酶)可插Cas9多肽的结构性或功能性结构域处,例如,在从Cas9多肽删除该结构域后。Cas9多肽的结构性或功能性结构域可包括,例如,RuvC I、RuvC II、RuvCIII、Rec1、Rec2、PI或HNH。

[1122] 在一些实施方案中,Cas9多肽缺少选自以下所组成的群组的一个或多个结构域:RuvC I、RuvC II、RuvC III、Rec1、Rec2、PI或HNH结构域。在一些实施方案中,Cas9多肽缺少核酸酶结构域。在一些实施方案中,Cas9多肽缺少HNH结构域。在一些实施方案中,Cas9多肽缺少HNH结构域的一部分,使得该Cas9多肽具有降低或消除的HNH活性。在一些实施方案中,Cas9多肽包含核酸酶结构域的缺失,并且脱氨酶插入以替换该核酸酶结构域。在一些实施方案中,HNH结构域被删除,并且脱氨酶被插入在该处。在一些实施方案中,一个或多个RuvC结构域被删除,并且脱氨酶被插入在该处。

[1123] 包含异源多肽的融合蛋白侧翼可以是napDNAbp的N端片段和C端片段。在一些实施方案中,融合蛋白包含脱氨酶,该脱氨酶的侧翼是Cas9多肽的N端片段和C端片段。N端片段或C端片段可结合靶标多核苷酸序列。N端片段的C端或C端片段的N端可包含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端可包含Cas9多肽的α螺旋结构的一部分。N端片段或C端片段可包含DNA结合结构域。N端片段或C端片段可包含RuvC结构域。N端片段或C端片段可包含HNH结构域。在一些实施方案中,N端片段或C端片段均不包含HNH结构域。

[1124] 在一些实施方案中,N端Cas9片段的C端包含在融合蛋白将靶标核碱基脱氨基时接近该靶标核碱基的氨基酸。在一些实施方案中,C端Cas9片段的N端包含在融合蛋白将靶标核碱基脱氨基时接近该靶标核碱基的氨基酸。不同脱氨酶的插入位置可以不同,以便在靶标核碱基与N端Cas9片段的C端或C端Cas9片段的N端中的氨基酸之间具有接近性。例如,脱氨酶的插入位置可以在选自以下所组成的群组的氨基酸残基处:1015、1022、1029、1040、1068、1247、1054、1026、768、1067、1248、1052和1246,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基。

[1125] 融合蛋白的N端Cas9片段(即,在融合蛋白中位于脱氨酶侧翼的N端Cas9片段)可包含Cas9多肽的N端。融合蛋白的N端Cas9片段可包含至少约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片段可包含对应于以下氨基酸残基的序列: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,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基。N端Cas9片段可包含序列,该序列与以下氨基酸残基具有至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%、或至少99.5%序列一致性: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,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基。

[1126] 融合蛋白的C端Cas9片段(即,在融合蛋白中位于脱氨酶侧翼的C端Cas9片段)可包含Cas9多肽的C端。融合蛋白的C端Cas9片段可包含至少约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片段可包含对应于以下氨基酸残基的序列:1099至1368、918至1368、906至1368、780至1368、765至1368、718至1368、94至1368、或56至1368,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基

酸残基。N端Cas9片段可包含序列,该序列与以下氨基酸残基具有至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%、或至少99.5%序列一致性:1099至1368、918至1368、906至1368、780至1368、765至1368、718至1368、94至1368、或56至1368,如上述Cas9参考序列中编号,或另一Cas9多肽中的对应氨基酸残基。

[1127] 融合蛋白的N端Cas9片段和C端Cas9片段结合在一起可能不对应于天然存在的全长Cas9多肽序列,例如,如上述Cas9参考序列中所详述的。

[1128] 本文所述的融合蛋白可实现靶向脱氨基,且在非靶标位点(例如,脱靶位点)处的脱氨基减少,诸如减少全基因组谬误脱氨基。本文所述的融合蛋白可实现靶向脱氨基,且减少在非靶标位点处的旁观者脱氨基。与例如包含融合至Cas9多肽的N端或C端的脱氨酶的末端融合蛋白相比,不希望的脱氨基或脱靶脱氨基作用可以减少至少30%,至少40%,至少50%,至少60%,至少70%,至少80%,至少90%、至少95%、或至少99%。与例如包含融合至Cas9多肽的N端或C端的脱氨酶的末端融合蛋白相比,不希望的脱氨基或脱靶脱氨基作用可以减少至少一倍、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。

[1129] 在一些实施方案中,融合蛋白的脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶、或者腺苷脱氨酶和胞苷脱氨酶)将R环范围内的不超过两个核碱基脱氨基。在一些实施方案中,融合蛋白的脱氨酶将R环范围内的不超过三个核碱基脱氨基。在一些实施方案中,融合蛋白的脱氨酶将R环范围内的不超过2、3、4、5、6、7、8、9、或10个核碱基脱氨基。R环是三链核酸结构,包括DNA:RNA杂交物、DNA:DNA、或RNA:RNA互补结构并且与单链DNA缔合。如本文所用,R环可以在靶标多核苷酸与CRISPR复合物或碱基编辑复合物接触时形成,其中向导多核苷酸(例如,向导RNA)的一部分与靶标多核苷酸(例如,靶标DNA)的一部分杂交并替代后者。在一些实施方案中,R环包含间隔序列与靶标DNA互补序列的杂交区。R环区可以是约5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49或50个核碱基对的长度。在一些实施方案中,R环区为约20个核碱基对的长度。应理解,如本文所用,R环区不限于与向导多核苷酸杂交的靶标DNA链。例如,对R环区内的靶标核碱基的编辑可以是对包含与向导RNA互补的链的DNA链的编辑,或者是对作为与向导RNA的互补链相反的链的DNA链的编辑。在一些实施方案中,在R环区内的编辑包含编辑靶标DNA序列中的与向导RNA非互补的链(前间区序列链)上的核碱基。

[1130] 本文所述的融合蛋白可在不同于标准碱基编辑的编辑窗口内实现靶标脱氨基。在一些实施方案中,靶标核碱基是靶标多核苷酸序列中的PAM序列的上游约1至约20个碱基。在一些实施方案中,靶标核碱基是靶标多核苷酸序列中的PAM序列的上游约2至约12个碱基。在一些实施方案中,靶标核碱基是距离PAM序列或是PAM序列上游约1至9个碱基对、约2至10个碱基对、约3至11个碱基对、约4至12个碱基对、约5至13个碱基对、约6至14个碱基对、约7至15个碱基对、约8至16个碱基对、约9至17个碱基对、约10至18个碱基对、约11至19个碱基对、约12至20个碱基对、约1至7个碱基对、约2至8个碱基对、约3至9个碱基对、约4至10个碱基对、约5至11个碱基对、约6至12个碱基对、约7至13个碱基对、约8至14个碱基对、约9至

15个碱基对、约10至16个碱基对、约11至17个碱基对、约12至18个碱基对、约13至19个碱基对、约14至20个碱基对、约1至5个碱基对、约2至6个碱基对、约3至7个碱基对、约4至8个碱基对、约5至9个碱基对、约6至10个碱基对、约7至11个碱基对、约8至12个碱基对、约9至13个碱基对、约10至14个碱基对、约11至15个碱基对、约12至16个碱基对、约13至17个碱基对、约14至18个碱基对、约15至19个碱基对、约16至20个碱基对、约1至3个碱基对、约2至4个碱基对、约3至5个碱基对、约4至6个碱基对、约5至7个碱基对、约6至8个碱基对、约7至9个碱基对、约8至10个碱基对、约9至11个碱基对、约10至12个碱基对、约11至13个碱基对、约12至14个碱基对、约13至15个碱基对、约14至16个碱基对、约15至17个碱基对、约16至18个碱基对、约17至19个碱基对、约18至20个碱基对。在一些实施方案中,靶标核碱基是距离PAM序列或是PAM序列上游约1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20或更多个碱基对。在一些实施方案中,靶标核碱基是PAM序列上游约1、2、3、4、5、6、7、8或9个碱基对。在一些实施方案中,靶标核碱基是PAM序列上游约1、2、3、4或6个碱基对。

[1131] 融合蛋白可包含超过一个异源多肽。例如,融合多肽可额外包含一个或多个UGI结构域和/或一个或多个核定位信号。两个或更多个异源结构域可以串联插入。两个或更多个异源结构域可以插入在使得它们在NapDNAbp中不串联的位置处。

[1132] 融合蛋白可包含位于脱氨酶与napDNAbp多肽之间的连接子。连接基可以是肽或非肽连接子。例如,连接子可以是XTEN、(GGGS)n、(GGGGS)n、(G)n、(EAAAK)n、(GGS)n、SGSETPGTSESATPES。在一些实施方案中,融合蛋白包含位于N端Cas9片段与脱氨酶之间的连接子。在一些实施方案中,融合蛋白包含位于C端Cas9片段与脱氨酶之间的连接子。在一些实施方案中,napDNAbp的N端和C端片段用连接子连结至脱氨酶。在一些实施方案中,N端片段和C端片段接合至脱氨酶结构域而无需连接子。在一些实施方案中,融合蛋白包含位于N端Cas9片段与脱氨酶之间的连接子,而不包含位于C端Cas9片段与脱氨酶之间的连接子。在一些实施方案中,融合蛋白包含位于C端Cas9片段与脱氨酶之间的连接子,而不包含位于N端Cas9片段与脱氨酶之间的连接子。

[1133] 在一些实施方案中,融合蛋白中的napDNAbp是Cas12多肽,例如,Cas12b/C2c1,或其片段。Cas12多肽可以是变体Cas12多肽。在其他实施方案中,Cas12多肽的N端或C端片段包含核酸可编程DNA结合结构域或RuvC结构域。在其他实施方案中,融合蛋白含有位于Cas12多肽与催化性结构域之间的连接子。在其他实施方案中,连接子的氨基酸序列是GGSGGS或GSSGSETPGTSESATPESSG。在其他实施方案中,连接子是刚性连接子。在上述方面的其他实施方案中,连接子由GGAGGCTCTGGAGGAAGC或GGCTCTTCTGGATCTGAAACACCTGGCACAAGCGAGAGCGCCACCCCTGAGAGCTCTGGC编码。

[1134] 包含侧翼为Cas12多肽的N端和C端的异源催化性结构域的融合蛋白也可用于在本文所述的方法中用于碱基编辑。包含Cas12和一个或多个脱氨酶结构域(例如,腺苷脱氨酶)或者包含侧翼具有Cas12序列的腺苷脱氨酶结构域的融合蛋白也可以用于靶标序列的高特异性且高效的碱基编辑。在一种实施方案中,嵌合Cas12融合蛋白含有插入到Cas12多肽内的异源催化性结构域(例如,腺苷脱氨酶、胞苷脱氨酶、或腺苷脱氨酶和胞苷脱氨酶)。在一些实施方案中,融合蛋白包含插入到Cas12内的腺苷脱氨酶结构域和胞苷脱氨酶结构域。在一些实施方案中,腺苷脱氨酶融合到Cas12内,并且胞苷脱氨酶融合至C端。在一些实施方案中,腺苷脱氨酶融合到Cas12内,并且胞苷脱氨酶融合至N端。在一些实施方案中,胞苷脱氨

酶融合到Cas12内,并且腺苷脱氨酶融合至C端。在一些实施方案中,胞苷脱氨酶融合到Cas12内,并且腺苷脱氨酶融合至N端。具有腺苷脱氨酶和胞苷脱氨酶以及Cas12的融合蛋白的示例性结构提供如下:

[1135] NH_2-[Cas12(腺苷脱氨酶)]-[胞苷脱氨酶]-COOH;

[1136] NH_2-[胞苷脱氨酶]-[Cas12(腺苷脱氨酶)]-COOH;

[1137] NH_2-[Cas12(胞苷脱氨酶)]-[腺苷脱氨酶]-COOH;或

[1138] NH_2-[腺苷脱氨酶]-[Cas12(胞苷脱氨酶)]-COOH;

[1139] 在一些实施方案中上述通用架构中所用的“-”表示任选的连接子的存在。

[1140] 在各种实施方案中,催化性结构域具有DNA修饰活性(例如,脱氨酶活性),诸如腺苷脱氨酶活性。在一些实施方案中,腺苷脱氨酶是TadA(例如,TadA*7.10)。在一些实施方案中,TadA是TadA*8。在一些实施方案中,TadA*8融合到Cas12内,并且胞苷脱氨酶融合至C端。在一些实施方案中,TadA*8融合到Cas12内,并且胞苷脱氨酶融合至N端。在一些实施方案中,胞苷脱氨酶融合到Cas12内,并且TadA*8融合至C端。在一些实施方案中,胞苷脱氨酶融合到Cas12内,并且TadA*8融合至N端。具有TadA*8和胞苷脱氨酶以及Cas12的融合蛋白的示例性结构提供如下:

[1141] N-[Cas12(TadA*8)]-[把颇感脱氨酶]-C;

[1142] N-[胞苷脱氨酶]-[Cas12(TadA*8)]-C;

[1143] N-[Cas12(胞苷脱氨酶)]-[TadA*8]-C;或

[1144] N-[TadA*8]-[Cas12(胞苷脱氨酶)]-C。

[1145] 在一些实施方案中上述通用架构中所用的“-”表示任选的连接子的存在。

[1146] 在其他实施方案中,融合蛋白含有一个或多个催化性结构域。在其他实施方案中,一个或多个催化性结构域中的至少一个插入到Cas12多肽内或融合在Cas12 N端或C端。在其他实施方案中,一个或多个催化性结构域中的至少一个插入到Cas12多肽的环、α螺旋、未结构化部分、或溶剂可及部分内。在其他实施方案中,Cas12多肽是Cas12a、Cas12b、Cas12c、Cas12d、Cas12e、Cas12g、Cas12h、Cas12i或Cas12j/CasΦ。在其他实施方案中,Cas12多肽与外村尚芽孢杆菌(Bacillus hisashii)Cas12b、热噬淀粉芽孢杆菌(Bacillus thermoamylovorans)Cas12b、芽孢杆菌属(Bacillus sp.V3-13)Cas12b或嗜热脂环酸芽孢杆菌(Alicyclobacillus acidiphilus)Cas12b具有至少约85%的氨基酸序列一致性。在其他实施方案中,Cas12多肽与外村尚芽孢杆菌Cas12b、热噬淀粉芽孢杆菌Cas12b、芽孢杆菌属V3-13 Cas12b或嗜热脂环酸芽孢杆菌Cas12b具有至少约90%的氨基酸序列一致性。在其他实施方案中,Cas12多肽与外村尚芽孢杆菌Cas12b、热噬淀粉芽孢杆菌Cas12b、芽孢杆菌属V3-13 Cas12b或嗜热脂环酸芽孢杆菌Cas12b具有至少约95%的氨基酸序列一致性。在其他实施方案中,Cas12多肽含有外村尚芽孢杆菌Cas12b、热噬淀粉芽孢杆菌Cas12b、芽孢杆菌属V3-13 Cas12b或嗜热脂环酸芽孢杆菌Cas12b,或由其组成。

[1147] 在其他实施方案中,催化性结构域插入在BhCas12b的氨基酸位置153至154、255至256、306至307、980至981、1019至1020、534至535、604至605、或344至345之间,或Cas12a、Cas12c、Cas12d、Cas12e、Cas12g、Cas12h、Cas12i或Cas12j/CasΦ的对应氨基酸处。在其他实施方案中,催化性结构域插入在BhCas12b的氨基酸P153与S154之间。在其他实施方案中,催化性结构域插入在BhCas12b的氨基酸K255与E256之间。在其他实施方案中,催化性结构

域插入在BhCas12b的氨基酸D980与G981之间。在其他实施方案中,催化性结构域插入在BhCas12b的氨基酸K1019与L1020之间。在其他实施方案中,催化性结构域插入在BhCas12b的氨基酸F534与P535之间。在其他实施方案中,催化性结构域插入在BhCas12b的氨基酸K604与G605之间。在其他实施方案中,催化性结构域插入在BhCas12b的氨基酸H344与F345之间。在其他实施方案中,催化性结构域插入在BvCas12b的氨基酸位置147与148、248与249、299与300、991与992、或1031与1032之间,或Cas12a、Cas12c、Cas12d、Cas12e、Cas12g、Cas12h、Cas12i或Cas12j/CasΦ的对应氨基酸处。在其他实施方案中,催化性结构域插入在BvCas12b的氨基酸P147与D148之间。在其他实施方案中,催化性结构域插入在BvCas12b的氨基酸G248与G249之间。在其他实施方案中,催化性结构域插入在BvCas12b的氨基酸P299与E300之间。在其他实施方案中,催化性结构域插入在BvCas12b的氨基酸G991与E992之间。在其他实施方案中,催化性结构域插入在BvCas12b的氨基酸K1031与M1032之间。在其他实施方案中,催化性结构域插入在AaCas12b的氨基酸位置157与158、258与259、310与311、1008与1009、或1044与1045之间,或Cas12a、Cas12c、Cas12d、Cas12e、Cas12g、Cas12h、Cas12i或Cas12j/CasΦ的对应氨基酸处。在其他实施方案中,催化性结构域插入在AaCas12b的氨基酸P157与G158之间。在其他实施方案中,催化性结构域插入在AaCas12b的氨基酸V258与G259之间。在其他实施方案中,催化性结构域插入在AaCas12b的氨基酸D310与P311之间。在其他实施方案中,催化性结构域插入在AaCas12b的氨基酸G1008与E1009之间。在其他实施方案中,催化性结构域插入在AaCas12b的氨基酸G1044与K1045之间。

[1148] 在其他实施方案中,融合蛋白含有核定位信号(例如,二分体核定位信号)。在其他实施方案中,核定位信号的氨基酸序列是MAPKKKRKVGIHGVPAA。在上述方面的其他实施方案中,核定位信号由以下序列编码:

[1149] ATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCC。在其他实施方案中,Cassette2b多肽含有沉默RuvC结构域催化活性的突变。在其他实施方案中,Cas12b多肽含有D574A、D829A和/或D952A突变。在其他实施方案中,融合蛋白进一步含有标签(例如,流感血细胞凝集素标签)。

[1150] 在一些实施方案中,融合蛋白包含具有内容融合的和碱基编辑结构域(例如,脱氨酶结构域,例如腺苷脱氨酶结构域的全部或部分)的napDNAbp结构域(例如,源自Cas12的结构域)。在一些实施方案中,napDNAbp是Cas12b。在一些实施方案中,碱基编辑器包含具有BhCas12b结构域,该结构域具有插入在下表10中提供的基因座处的内部融合的TadA*8结构域。

[1151] 表10:Cas12b蛋白中的插入基因座

[1152]

BhCas12b	插入位点	插入在aa之间
位置1	153	PS
位置2	255	KE
位置3	306	DE
位置4	980	DG
位置5	1019	KL
位置6	534	FP
位置7	604	KG

位置8	344	HF
BvCas12b	插入位点	插入在aa之间
位置1	147	PD
位置2	248	GG
位置3	299	PE
位置4	991	GE
位置5	1031	KM
AaCas12b	插入位点	插入在aa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

[1153] 作为非限制性示例,腺苷脱氨酶(例如,TadA*8.13)可以插入到BhCas12b中以产生有效编辑核酸序列的融合蛋白(例如,TadA*8.13-BhCas12b)。

[1154] 在一些实施方案中,本文所述的碱基编辑系统是ABE,该ABE具有插入到Cas9中的TadA。提供了具有插入到Cas9中的TadA的相关ABE的序列。

[1155] 101 Cas9 TadAins 1015

[1156] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVL
NNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGA
AGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDYDVRKMIAKSEQEIGKATAKYFF
YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI

IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[1157] 102 Cas9 TadAins 1022

[1158] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVP
VGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGV
RNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDAKSEQEIGKATAKYFF
YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[1159] 103 Cas9 TadAins 1029

[1160] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRA

RDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRI
GRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGKATAKYFF
YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[1161] 103 Cas9 TadAins 1040

[1162] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA
RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVD
STDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSR
RLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNL
SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNEL
TKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMG
RHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAE
RGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSGSSGSETPGTSESATP
ESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNY
RLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFF
RMPRQVFNAQKKAQSSTDNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKT
EVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSF
EKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPE
DNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKY
FDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD105 Cas9 TadAins 1068

[1163] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE

MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGEGS
SGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEI
MALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGIL
ADECAALLCYFFRMPRQVFNAQKKAQSSTDTGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[1164] 106 Cas9 TadAins 1247

[1165] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLV
LNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTG
AAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[1166] 107 Cas9 TadAins 1054

[1167] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK

PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGSSGSETPGTSESATP
ESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYR
LIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRM
PRQVFNAQKKAQSSTDGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[1168] 108 Cas9 TadAins 1026

[1169] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDE
REVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRV
VFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDQEIGKATAKYFF
YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[1170] 109 Cas9 TadAins 768

[1171] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLH
DPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHR
VEITEGILADECAALLCYFFRMPRTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRD
MYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFD
NLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITL
ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK
KYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFEL
ENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILA
DANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI
DLSQLGGD

[1172] 110.1 Cas9 TadAins 1250

[1173] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAV

LVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAK
TGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPREDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILA
DANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI
DLSQLGGD

[1174] 110.2 Cas9 TadAins 1250

[1175] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPGSSGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPV
GAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVR
NAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPREDNEQKQLFVEQHKHYLDEIIEQISEFSKRV
ILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYE
TRIDLSQLGGD

[1176] 110.3 Cas9 TadAins 1250

[1177] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF

IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPGSSGSSGSETPGTSESATPESGSSSGSEVEFSHEYWMRHALTLAKRARDERE
VPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVF
GVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPREDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITG
LYETRIDLSQLGGD

[1178] 110.4 Cas9 TadAins 1250

[1179] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPGSSGSSGSETPGTSESATPESGSSSGSEVEFSHEYWMRHALTLAKRARDERE
VPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVF
GVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMRREDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITG
LYETRIDLSQLGGD

[1180] 110.5 Cas9 TadAins 1249

[1181] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT

EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSGSSGSSGSETPGTSESATPESGSSSGSEVEFSHEYWMRHALTLAKRARDEREV
PVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFG
VRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMRRPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITG
LYETRIDLSQLGGD

[1182] 110.5 Cas9 TadAinsδ59-66 1250

[1183] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPGSSGSSGSETPGTSESATPESGSSGSEVEFSHEYWMRHALTLAKRARDEREV
PVGAVLVLNNRVIGEGWNRAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGA
AGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDEDNEQKQLFVEQHKHYLDEIIEQ
ISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH
QSITGLYETRIDLSQLGGD

[1184] 110.6 Cas9 TadAins 1251

[1185] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST

DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPEGSSGSSGSETPGTSESATPESGSSSGSEVEFSHEYWMRHALTLAKRARDER
EVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVV
FGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMRRDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITG
LYETRIDLSQLGGD

[1186] 110.7 Cas9 TadAins 1252

[1187] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPEDGSSGSSGSETPGTSESATPESGSSSGSEVEFSHEYWMRHALTLAKRARDE
REVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRV
VFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMRRNEQKQLFVEQHKHYLDEIIEQISEFS

KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITG
LYETRIDLSQLGGD

[1188] 110.8 Cas9 TadAinsδ59-66C截短1250

[1189] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAV
LVLNNRVIGEGWNRAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLM
DVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKV
LSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGG
D

[1190] 111.1 Cas9 TadAins 997

[1191] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDAT

LYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVF
NAQKKAQSSTDGSSGSETPGTSESATPESSGIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIM
NFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLI
ARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIK
LPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQIS
EFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQS
ITGLYETRIDLSQLGGD

[1192] 111.2 Cas9 TadAins 997

[1193] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDAT
LYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVF
NAQKKAQSSTDGSSGSSGSETPGTSESATPESSGGSSIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYF
FYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVK
KDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDE
IIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDA
TLIHQSITGLYETRIDLSQLGGD

[1194] 112δHNH TadA

[1195] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS

DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRA
IGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPG
MNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMN
TKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYD
VRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNI
VKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMER
SSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGS
PEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK
YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1196] 113N端单TadA螺旋截短165末端

[1197] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
YFFRMPRSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTD
RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYN
QLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDT
YDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILR
RQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKN
LPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSV
EISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRR
YTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNE
KLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQL
LNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLV
SDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS
NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSD
KLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDL
IIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIE
QISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLI
HQSITGLYETRIDLSQLGGD

[1198] 114 N端单TadA螺旋截短165末端δ59-65

[1199] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRTAHAEIMALRQGGLVMQNYRLI
DATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPR
SGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKN
LIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIV
DEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN

LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ
SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYP
FLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVL
PKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRL
SRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQ
TVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYL
QNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLIT
QRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDF
QFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFK
TEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKK
DWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKY
SLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSK
RVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGL
YETRIDLSQLGGD

[1200] 115.1 Cas9 TadAins1004

[1201] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRA
IGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPG
MNHRVEITEGILADECAALLCYFFRMPRQLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP
KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVIL
ADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGD

[1202] 115.2 Cas9 TadAins1005

[1203] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR

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

[1204] 115.3 Cas9 TadAins1006

[1205] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLEGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWN
RAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHY
PGMNHRVEITEGILADECAALLCYFFRMPRQSEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP
KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE

LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVIL
ADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGD

[1206] 115.4Cas9 TadAins1007

[1207] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGW
NRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLH
YPGMNHRVEITEGILADECAALLCYFFRMPRQEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP
KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVIL
ADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGD

[1208] 116.1Cas9 TadAins C端截短2 792

[1209] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREV
PVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFG
VRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQSQILKEHPVENTQLQNEKLYLYYLQNGR

DMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYK
VREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP
KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVIL
ADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGD

[1210] 116.2 Cas9 TadAins C端截短2 791

[1211] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVP
VGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGV
RNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQGSQILKEHPVENTQLQNEKLYLYYLQNGR
DMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYK
VREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP
KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVIL
ADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGD

[1212] 116.3 Cas9 TadAins C端截短2 790

[1213] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD

NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKEGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPV
GAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVR
NAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQLGSQILKEHPVENTQLQNEKLYLYYLQNGR
DMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYK
VREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP
KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVIL
ADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGD

[1214] 117 Cas9δ1017-1069

[1215] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYSSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLH
DPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHR
VEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKE
SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAK
GYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQH
KHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTST
KEVLDATLIHQSITGLYETRIDLSQLGGD

[1216] 118 Cas9 TadA-CP116ins 1067

[1217] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNMNHR
VEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAK
RARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHS
RIGRVVFGVRNAKTGAAGSLMDVLHYPGGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[1218] 119 Cas9 TadAins 701

[1219] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIG
EGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMD
VLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDLTFKEDIQKAQVSGQGDSLHEHIANLAGS
PAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ
NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWR
QLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSK
LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFF
YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[1220] 120 Cas9 TadACP136ins 1248

[1221] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTS
ESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[1222] 121 Cas9 TadACP136ins 1052

[1223] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAMNHRVEITEGILADECA
ALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAV
LVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAK

TGAAGSLMDVLHYPGNGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[1224] 122 Cas9 TadACP136ins 1041

[1225] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSMNHRVEITEGILADECAALLCYFFRMPRQV
FNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWN
RAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHY
PGNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[1226] 123 Cas9 TadACP139ins 1299

[1227] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD

HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRMNH
RVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLA
KRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIH
SRIGRVVFGVRNAKTGAAGSLMDVLHYPGDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[1228] 124 Cas9δ792-872 TadAins

[1229] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRA
IGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPG
MNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHA
HDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRP
LIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT
VAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLA
SAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLS
AYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1230] 125 Cas9 delta 792-906 TadAins

[1231] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA
RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVD
STDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSR
RLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNL
SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNEL
TKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMG
RHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLN
NRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGA
AGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGLSELDKAGFIKRQLVETRQIT
KHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKL
ESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF
ATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKK
LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKY
VNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAEN
IIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD126 TadA CP65ins
1003

[1232] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVL
HYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWM
RHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFF
YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[1233] 127 TadA CP65ins 1016

[1234] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK

PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNA
KTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESS
GSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPYDVRKMIAKSEQEIGKATAKYFF
YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[1235] 128 TadA CP65ins 1022

[1236] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMITAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRV
VFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSE
SATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPAKSEQEIGKATAKYFF
YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[1237] 129 TadA CP65 ins 1029

[1238] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEITAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMI
HSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGS
ETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPGKATAKYFF
YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[1239] 130 TadA CP65ins 1041

[1240] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSTAHAEIMALRQGGLVMQNYRLIDATLYVTF
EPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKK
AQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLH
DPNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN

SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[1241] 131 TadA CP65ins 1054

[1242] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANTAHAEIMALRQGGLVM
QNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCY
FFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNN
RVIGEGWNRAIGLHDPGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[1243] 132 TadA CP65ins 1246

[1244] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF

IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRN
AKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPES
SGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD

[1245] 在一些实施方案中,生成腺苷脱氨酶碱基编辑器以将TadA或其变体在经鉴别的位置插入Cas9多肽内。

[1246] 示例性但非限制性的融合蛋白在国际PCT申请号PCT/US2020/016285和美国专利临时申请号62/852,228和62/852,224中有所描述,其内容通过引用以其整体并入本文。

[1247] A到G编辑

[1248] 在一些实施方案中,本文所述的碱基编辑器包含融合蛋白,该融合蛋白包含腺苷脱氨酶结构域。碱基编辑器的这种腺苷脱氨酶结构域可促进通过将腺嘌呤A)脱氨基以形成肌苷(I)(其表现出鸟嘌呤(G)的碱基配对性质)而将A核碱基编辑为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基(即,去除氨基团)。

[1249] 在一些实施方案中,本文所提供的核碱基编辑器可通过下述制备:将一个或多个蛋白质结构域融合在一起,从而生成融合蛋白。在某些实施方案中,本文所提供的融合蛋白包含一个或多个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含具有降低的核酸酶活性的Cas9结构域。在一些实施方案中,本文所提供的融合蛋白可具有没有核酸酶活性的Cas9结构域(dCas9),或切断双螺旋DNA分子的一条链的Cas9结构域,称为Cas9切口酶(nCas9)。不欲受缚于任何特定理论,催化性残基(例如,H840)的存在维持Cas9切割含有与所靶向的A相对的T的非靶向(例如,非脱氨基)链的活性。Cas9的催化性残基的突变(例如,D10到A10)防止切割含有所靶向的A残基的被编辑链。此类Cas9变体能够在基于gRNA定义的靶标序列的特异性位置处生成单链DNA断裂(切口),导致非编辑链的修复,最终导致非编辑链上的T到C的改变。在一些实施方案中,A到G碱基编辑器进一步包含肌苷碱基切除修复的抑制剂,例如,尿嘧啶糖基化酶抑制剂(UGI)结构域或无催化活性的肌苷特异性核酸酶。不欲受缚于任何特定理论,UGI结构域或无催化活性的肌苷特异性核酸酶可抑制或防止被脱氨基的腺苷残基(例如,肌苷)的碱基切除修复,这可以改善碱基编辑器的活性或效率。

[1250] 包含腺苷脱氨酶的甲基编辑器可作用于任何多核苷酸,包括DNA、RNA和DNA-RNA杂交物。在某些实施方案中,包含腺苷脱氨酶的碱基编辑器可将包含RNA的多核苷酸的靶标A脱氨基。例如,碱基编辑器可包含能够将RNA多核苷酸和/或DNA-RNA杂交物多核苷酸的靶标A脱氨基的腺苷脱氨酶结构域。在一种实施方案中,并入碱基编辑器中的腺苷脱氨酶包含作用于RNA的腺苷脱氨酶(ADAR,例如,ADAR1或ADAR2)的全部或一部分。在另一实施方案中,并入碱基编辑器中的腺苷脱氨酶包含作用于tRNA的腺苷脱氨酶(ADAT)的全部或一部分。包含腺苷脱氨酶结构域的碱基编辑器也可以能够将DNA多核苷酸的A核碱基脱氨基。在一种实施

方案中,碱基编辑器的腺苷脱氨酶结构域包含ADAT的全部或部分,该ADAT包含允许该ADAT将DNA中的靶标A脱氨基的一个或多个突变。例如,碱基编辑器可包含来自大肠杆菌的ADAT(EcTadA)的全部或部分,该ADAT包含以下突变中的一者或多者:D108N、A106V、D147Y、E155V、L84F、H123Y、I156F,或另一腺苷脱氨酶中的对应突变。

[1251] 腺苷脱氨酶可以源自任何合适的生物体(例如,大肠杆菌)。在一些实施方案中,腺苷脱氨酶来自原核生物。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(Salmonella typhi)、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。在一些实施方案中腺苷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对应残基可以通过例如序列比对并确定同源残基来鉴定。可据此生成任何天然存在的腺苷脱氨酶(例如,与ecTadA具有同源性)中的对应于本文所述任何突变(例如,在ecTadA中鉴定的任何突变)的突变。

[1252] 腺苷脱氨酶

[1253] 在一些实施方案中,本发明的融合蛋白包含一个或多个腺苷脱氨酶结构域。在一些实施方案中,本文提供的腺苷脱氨酶能够将腺嘌呤脱氨基。在一些实施方案中,本文提供的腺苷脱氨酶能够将DNA的脱氧腺苷中的腺嘌呤脱氨基。腺苷脱氨酶可以源自任何合适的生物体(例如,大肠杆菌(E.coli))。在一些实施方案中腺苷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。本领域技术人员将能够鉴定任何同源蛋白质中的对应残基,通过序列比对并确定同源残基。据此,本领域技术人员将能够生成任何天然存在的腺苷脱氨酶(例如,与ecTadA具有同源性)中的对应于本文所述任何突变(例如,在ecTadA中鉴定的任何突变)的突变。在一些实施方案中,腺苷脱氨酶来自原核生物。在一些实施方案中,腺苷脱氨酶来自细菌。在一些实施方案中,腺苷脱氨酶来自大肠杆菌、金黄色葡萄球菌、伤寒沙门氏菌(Salmonella typhi)、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。

[1254] 本发明提供腺苷脱氨酶变体,这些变体具有增加的效率(>50-60%)和特异性。特别地,本文所述的腺苷脱氨酶变体更可能编辑多核苷酸内的所希望的碱基,并且更不可能编辑不希望被改变的碱基(即,“旁观者”)。

[1255] 在一些施方案中,腺苷脱氨酶是TadA脱氨酶。在特定实施方案中,TadA是PCT/US2017/045381(WO 2018/027078)中描述的TadA中的任一种,该专利通过引用并入本文。

[1256] 野生型TadA(wt)腺苷脱氨酶具有以下序列(也称为TadA参考序列):

[1257] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD

[1258] 在一些实施方案中,腺苷脱氨酶是全长大肠杆菌TadA脱氨酶。例如,在某些实施方案中,腺苷脱氨酶包含氨基酸序列:

[1259] MRRAFITGVFFLSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEIT

EGILADECAALLSDFFRMRRQEIKAQKKAQSSTD。

[1260] 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(Escherichia coli(E.coli))、金黄色葡萄球菌(Staphylococcus aureus(S.aureus))、鼠伤寒沙门氏菌(Salmonella typhimurium(S.typhimurium))、腐败希瓦氏菌(Shewanella putrefaciens(S.putrefaciens))、流感嗜血杆菌(Haemophilus influenzae(H.influenzae))、新月柄杆菌(Caulobacter crescentus(C.crescentus))、硫还原地杆菌(Geobacter sulfurreducens(G.sulfurreducens))或枯草芽孢杆菌(Bacillus subtilis)。在一些实施方案中,腺苷脱氨酶来自大肠杆菌。

[1261] 但应了解,其他可用于本申请的腺苷脱氨酶对于本领域技术人员而言将会是显而易见的并且出于本公开的范畴内。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。非限制地,示例性AD AT同源物的氨基酸序列包括下列:

[1262] 金黄色葡萄球菌(Staphylococcus aureus(S.aureus))TadA:

[1263] MGSHMTNDIYFMTLAIEEAKKAAQLGEVPIGAIITKDDEVIARAHNLRETLQQPTAHAEHIAIERAAKVLGSWRLEGCTLYVTLEPCVMCAGTIVMSRIPRVVYGADDPKGGCSGSLMNLLQQSNFNHRAIVDKGVLKEACSTLLTTFFKNLRANKKSTN

[1264] 枯草芽孢杆菌(Bacillus subtilis(B.subtilis))TadA:

[1265] MTQDELYMKEAIKEAKKAEEKGEVPIGAVLVINGEIIARAHNLRETEQRSIAHAEMLVIDEACKALGTWRLEGATLYVTLEPCPMCAGAVVLSRVEKVVFGAFDPKGGCSGTLMNLLQEERFNHQAEVVSGVLEEECGGMLSAFFRELRKKKKAARKNLSE

[1266] 鼠伤寒沙门氏菌(Salmonella typhimurium (S.typhimurium))TadA:

[1267] MPPAFITGVTSLSDVELDHEYWMRHALTLAKRAWDEREVPVGAVLVHNHRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVLQNYRLLDTTLYVTLEPCVMCAGAMVHSRIGRVVFGARDAKTGAAGSLIDVLHHPGMNHRVEIIEGVLRDECATLLSDFFRMRRQEIKALKKADRAEGAGPAV

[1268] 腐败希瓦氏菌(Shewanella putrefaciens(S.putrefaciens))TadA:

[1269] MDEYWMQVAMQMAEKAEAAGEVPVGAVLVKDGQQIATGYNLSISQHDPTAHAEILCLRSAGKKLENYRLLDATLYITLEPCAMCAGAMVHSRIARVVYGARDEKTGAAGTVVNLLQHPAFNHQVEVTSGVLAEACSAQLSRFFKRRRDEKKALKLAQRAQQGIE

[1270] 流感嗜血杆菌(Haemophilus influenzae F3031(H.influenzae))TadA:

[1271] MDAAKVRSEFDEKMMRYALELADKAEALGEIPVGAVLVDDARNIIGEGWNLSIVQSDPTAHAEIIALRNGAKNIQNYRLLNSTLYVTLEPCTMCAGAILHSRIKRLVFGASDYKTGAIGSRFHFFDDYKMNHTLEITSGVLAEECSQKLSTFFQKRREEKKIEKALLKSLSDK

[1272] 新月柄杆菌(Caulobacter crescentus(C.crescentus))TadA:

[1273] MRTDESEDQDHRMMRLALDAARAAAEAGETPVGAVILDPSTGEVIATAGNGPIAAHDPTAHAEIAAMRAAAAKLGNYRLTDLTLVVTLEPCAMCAGAISHARIGRVVFGADDPKGGAVVHGPKFFAQPTCHWRPEVTGGVLADESADLLRGFFRARRKAKI

[1274] 硫还原地杆菌(Geobacter sulfurreducens(G.sulfurreducens))TadA:

[1275] MSSLKKTPIRDDAYWMGKAIREAAKAAARDEVPIGAVIVRDGAVIGRGHNLREGSNDPSAHAEMIAIRQAARRSANWRLTGATLYVTLEPCLMCMGAIILARLERVVFGCYDPKGGAAGSLYDLSADPRLNHQVRLSPGVCQEECG

TMLSDFFRDLRRRKKAKATPALFIDERKVPPEP

[1276] 大肠杆菌TadA(ecTadA)的一个实施方案包括下列：

[1277] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC YFFRMPRQVFNAQKKAQSSTD

[1278] 在一些实施方案中,腺苷脱氨酶包含氨基酸序列,该氨基酸序列与本文所提供的任何腺苷脱氨酶中详述的氨基酸序列中的任一者至少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。应了解,本文所提供的腺苷脱氨酶可包括一个或多个突变(例如,本文所提供的任何突变)。本公开提供任何具有某一百分比的一致性加上任何本文所述突变或其组合的脱氨酶结构域。在一些实施方案中,腺苷脱氨酶包含氨基酸序列,与参考序列或本文所提供的任何腺苷脱氨酶相比,该氨基酸序列具有1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、21、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50个或更多个突变。在一些实施方案中,腺苷脱氨酶包含氨基酸序列,与本领域已知或本文所述的氨基酸序列中的任一者相比,该氨基酸序列具有至少5、至少10、至少15、至少20、至少25、至少30、至少35、至少40、至少45、至少50、至少60、至少70、至少80、至少90、至少100、至少110、至少120、至少130、至少140、至少150、至少160、或至少170个相同的连续氨基酸残基。

[1279] 应了解,本文所提供的任何突变(例如,基于TadA参考序列)可引入其他腺苷脱氨酶中,诸如大肠杆菌TadA(ecTadA)、金黄色葡萄球菌TadA(saTadA),或其他腺苷脱氨酶(例如,细菌腺苷脱氨酶)。对本领域技术人员将会显而易见的是,可类似地比对其他脱氨酶以鉴定可以如本文所提供者进行突变的同源氨基酸残基。因此,TadA参考序列中鉴定的任何突变可以在具有同源氨基酸残基的其他腺苷脱氨酶(例如,ecTada)中作出。应了解,本文所提供的任何突变可以独立地作出或与TadA参考序列或另一腺苷脱氨酶中的突变以任意组合作出。

[1280] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的D108X突变,或另一腺苷脱氨酶中的对应突变,其中X指示任何除野生型腺苷脱氨酶中对应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的D108G、D108N、D108V、D108A或D108Y突变,或另一腺苷脱氨酶中的对应突变。但应了解,可类似地比对其他脱氨酶以鉴定可以如本文所提供者进行突变的同源氨基酸残基。

[1281] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的A106X突变,或另一腺苷脱氨酶中的对应突变,其中X指示任何除野生型腺苷脱氨酶中对应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的A106V突变,或另一腺苷脱氨酶(例如,ecTadA)中的对应突变。

[1282] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的E155X突变,或另一腺苷脱氨酶中的对应突变,其中X的存在指示任何除野生型腺苷脱氨酶中对应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的E155D、E155G或E155V突变,或另一腺苷脱氨酶(例如,ecTadA)中的对应突变。

[1283] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的D147X突变,或另一腺苷脱

氨酶中的对应突变,其中X的存在指示任何除野生型腺苷脱氨酶中对应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的D147Y突变,或另一腺苷脱氨酶(例如,ecTadA)中的对应突变。

[1284] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的A106X、E155X或D147X突变,或另一腺苷脱氨酶(例如,ecTadA)中的对应突变,其中X指示任何除野生型腺苷脱氨酶中对应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含E155D、E155G或E155V突变。在一些实施方案中,腺苷脱氨酶包含D147Y。

[1285] 应了解,本文所提供的任何突变(例如,基于TadA参考序列的ecTadA氨基酸序列)可引入其他腺苷脱氨酶中,诸如金黄色葡萄球菌TadA(saTadA),或其他腺苷脱氨酶(例如,细菌腺苷脱氨酶)。ecTadA中突变残基的同源性如何对本领域技术人员将会显而易见的。因此,ecTadA中鉴定的任何突变可以在具有同源氨基酸残基的其他腺苷脱氨酶中作出。应了解,本文所提供的任何突变可以独立地作出或与edTadA或另一腺苷脱氨酶中的突变以任意组合作出。

[1286] 例如,腺苷脱氨酶可含有TadA参考序列中的D108N、A106V、E155V和/或D147Y突变,或另一腺苷脱氨酶(例如,ecTadA)中的对应突变。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的下列突变组(突变组以“;”分隔),或另一腺苷脱氨酶中的对应突变:D108N和A106V;D108N和E155V;D108N和D147Y;A106V和E155V;A106V和D147Y;E155V和D147Y;D108N、A106V和E155V;D108N、A106V和D147Y;D108N、E155V和D147Y;A106V、E155V和D147Y;以及D108N、A106V、E155V和D147Y。但应了解,本文所提供的对应突变的任何组合可在腺苷脱氨酶(例如,ecTadA)中作出。

[1287] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H8X、T17X、L18X、W23X、L34X、W45X、R51X、A56X、E59X、E85X、M94X、I95X、V102X、F104X、A106X、R107X、D108X、K110X、M118X、N127X、A138X、F149X、M151X、R153X、Q154X、I156X和/或K157X突变中的一者或多者,或另一腺苷脱氨酶中的一个或多个对应突变,其中X的存在指示除野生型腺苷脱氨酶中的对应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H8Y、T17S、L18E、W23L、L34S、W45L、R51H、A56E,或A56S、E59G、E85K,或E85G、M94L、I95L、V102A、F104L、A106V、R107C,或R107H,或R107P、D108G,或D108N,或D108V,或D108A,或D108Y、K110I、M118K、N127S、A138V、F149Y、M151V、R153C、Q154L、I156D和/或K157R突变中的一者或多者,或另一腺苷脱氨酶中的一个或多个对应突变。

[1288] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H8X、D108X和/或N127X突变中的一者或多者,或另一腺苷脱氨酶中的一个或多个对应突变,其中X指示任何氨基酸的存在。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H8Y、D108N和/或N127S突变中的一者或多者,或另一腺苷脱氨酶(中的一个或多个对应突变。

[1289] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H8X、R26X、M61X、L68X、M70X、A106X、D108X、A109X、N127X、D147X、R152X、Q154X、E155X、K161X、Q163X和/或T166X突变中的一者或多者,或另一腺苷脱氨酶中的一个或多个对应突变,其中X指示除野生型腺苷脱氨酶中的对应氨基酸以外的任何氨基酸的存在。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H8Y、R26W、M61I、L68Q、M70V、A106T、D108N、A109T、N127S、D147Y、R152C、Q154H或Q154R、E155G或E155V或E155D、K161Q、Q163H、和/或T166P突变中的一者或多者,或

另一腺苷脱氨酶中的一个或多个对应突变。

[1290] 在一些实施方案中,腺苷脱氨酶包含选自由TadA参考序列中的H8X、D108X、N127X、D147X、R152X和Q154X所组成的群组的一个、两个、三个、四个、五个或六个突变,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个对应突变,其中X指示除野生型腺苷脱氨酶中的对应氨基酸以外的任何氨基酸的存在。在一些实施方案中,腺苷脱氨酶包含选自由TadA参考序列中的H8X、M61X、M70X、D108X、N127X、Q154X、E155X和Q163X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,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个对应突变,其中X指示除野生型腺苷脱氨酶中的对应氨基酸以外的任何氨基酸的存在。在一些实施方案中,腺苷脱氨酶包含选自由TadA参考序列中的H8X、D108X、N127X、E155X和T166X所组成的群组的一个、两个、三个、四个或五个突变,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个对应突变,其中X指示除野生型腺苷脱氨酶中的对应氨基酸以外的任何氨基酸的存在。

[1291] 在一些实施方案中,腺苷脱氨酶包含选自由H8X、A106X和D108X所组成的群组的一个、两个、三个、四个、五个或六个突变,或另一腺苷脱氨酶中的一个或多个突变,其中X指示除野生型腺苷脱氨酶中的对应氨基酸以外的任何氨基酸的存在。在一些实施方案中,腺苷脱氨酶包含选自由TadA参考序列中的H8X、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,或另一腺苷脱氨酶中的一个或多个对应突变,其中X指示除野生型腺苷脱氨酶中的对应氨基酸以外的任何氨基酸的存在。

[1292] 在一些实施方案中,腺苷脱氨酶包含选自由TadA参考序列中的H8X、R126X、L68X、D108X、N127X、D147X和E155X所组成的群组的一个、两个、三个、四个、五个、六个或七个突变,或另一腺苷脱氨酶中的一个或多个对应突变,其中X指示除野生型腺苷脱氨酶中的对应氨基酸以外的任何氨基酸的存在。在一些实施方案中,腺苷脱氨酶包含选自由TadA参考序列中的H8X、D108X、A109X、N127X和E155X所组成的群组的一个、两个、三个、四个或五个突变,或另一腺苷脱氨酶中的一个或多个对应突变,其中X指示除野生型腺苷脱氨酶中的对应氨基酸以外的任何氨基酸的存在。

[1293] 在一些实施方案中,腺苷脱氨酶包含选自由TadA参考序列中的H8Y、D108N、N127S、D147Y、R152C和Q154H所组成的群组的一个、两个、三个、四个、五个或六个突变,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个对应突变。在一些实施方案中,腺苷脱氨酶包含选自由TadA参考序列中的H8Y、M61I、M70V、D108N、N127S、Q154R、E155G和Q163H所组成的群组的一个、两个、三个、四个、五个、六个、七个或八个突变,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个对应突变。在一些实施方案中,腺苷脱氨酶包含选自由TadA参考序列中的H8Y、D108N、N127S、E155V和T166P所组成的群组的一个、两个、三个、四个或五个突变,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个对应突变。在一些实施方案中,腺苷脱氨酶包含选自由TadA参考序列中的H8Y、A106T、D108N、N127S、E155D和K161Q所组成的群组的一个、两个、三个、四个、五个或六个突变,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个对应突变。在一些实施方案中,腺苷脱氨酶包含选自由TadA参考序列中的H8Y、R26W、L68Q、D108N、N127S、D147Y和E155V所组成的群组的一个、两个、三个、四个、五个、六个、七个或八个突变,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个对应突变。在一些实施方案中,

腺苷脱氨酶包含选自由TadA参考序列中的H8Y、D108N、A109T、N127S和E155G所组成的群组的一个、两个、三个、四个或五个突变,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个对应突变。

[1294] 在一些实施方案中,腺苷脱氨酶包含另一腺苷脱氨酶中的一个或多个突变中的一者或多者。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的D108N、D108G或D108V突变,或另一腺苷脱氨酶中的对应突变。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的A106V和D108N突变,或另一腺苷脱氨酶中的对应突变。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的R107C和D108N突变,或另一腺苷脱氨酶中的对应突变。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H8Y、D108N、N127S、D147Y和Q154H突变,或另一腺苷脱氨酶中的对应突变。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H8Y、D108N、N127S、D147Y和E155V突变,或另一腺苷脱氨酶中的对应突变。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的D108N、D147Y和E155V突变,或另一腺苷脱氨酶中的对应突变。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H8Y、D108N和N127S突变,或另一腺苷脱氨酶中的对应突变。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的A106V、D108N、D147Y和E155V突变,或另一腺苷脱氨酶(例如,ecTadA)中的对应突变。

[1295] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的S2X、H8X、I49X、L84X、H123X、N127X、I156X和/或K160X突变中的一者或多者,或另一腺苷脱氨酶中的一个或多个对应突变,其中X的存在指示除野生型腺苷脱氨酶中的对应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的S2A、H8Y、I49F、L84F、H123Y、N127S、I156F和/或K160S突变中的一者或多者,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个对应突变。

[1296] 在一些实施方案中,腺苷脱氨酶包含L84X突变腺苷脱氨酶,其中指示任何除野生型腺苷脱氨酶中对应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的L84F突变,或另一腺苷脱氨酶(例如,ecTadA)中的对应突变。

[1297] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H123X突变,或另一腺苷脱氨酶中的对应突变,其中X指示任何除野生型腺苷脱氨酶中对应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H123Y突变,或另一腺苷脱氨酶中的对应突变。

[1298] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的I156X突变,或另一腺苷脱氨酶中的对应突变,其中X指示任何除野生型腺苷脱氨酶中对应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的I156F突变,或另一腺苷脱氨酶中的对应突变。

[1299] 在一些实施方案中,腺苷脱氨酶包含选自由TadA参考序列中的L84X、A106X、D108X、H123X、D147X、E155X和I156X所组成的群组的一个、两个、三个、四个、五个、六个或七个突变,或另一腺苷脱氨酶中的一个或多个对应突变,其中X指示除野生型腺苷脱氨酶中的对应氨基酸以外的任何氨基酸的存在。在一些实施方案中,腺苷脱氨酶包含选自由TadA参考序列中的S2X、I49X、A106X、D108X、D147X和E155X所组成的群组的一个、两个、三个、四个、五个或六个突变,或另一腺苷脱氨酶中的一个或多个对应突变,其中X指示除野生型腺苷脱氨酶中的对应氨基酸以外的任何氨基酸的存在。在一些实施方案中,腺苷脱氨酶包含选自

由TadA参考序列中的H8X、A106X、D108X、N127X和K160X所组成的群组的一个、两个、三个、四个或五个突变,或另一腺苷脱氨酶中的一个或多个对应突变,其中X指示除野生型腺苷脱氨酶中的对应氨基酸以外的任何氨基酸的存在。

[1300] 在一些实施方案中,腺苷脱氨酶包含选自由TadA参考序列中的L84F、A106V、D108N、H123Y、D147Y、E155V和I156F所组成的群组的一个、两个、三个、四个、五个、六个或七个突变,或另一腺苷脱氨酶中的一个或多个对应突变。在一些实施方案中,腺苷脱氨酶包含选自由TadA参考序列中的S2A、I49F、A106V、D108N、D147Y和E155V所组成的群组的一个、两个、三个、四个、五个或六个突变。

[1301] 在一些实施方案中,腺苷脱氨酶包含选自由TadA参考序列中的H8Y、A106T、D108N、N127S和K160S所组成的群组的一个、两个、三个、四个或五个突变,或另一腺苷脱氨酶中的一个或多个对应突变。

[1302] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的E25X、R26X、R107X、A142X和/或A143X突变中的一者或多者,或另一腺苷脱氨酶中的一个或多个对应突变,其中X的存在指示除野生型腺苷脱氨酶中的对应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的E25M、E25D、E25A、E25R、E25V、E25S、E25Y、R26G、R26N、R26Q、R26C、R26L、R26K、R107P、R107K、R107A、R107N、R107W、R107H、R107S、A142N、A142D、A142G、A143D、A143G、A143E、A143L、A143W、A143M、A143S、A143Q和/或A143R突变中的一者或多者,或另一腺苷脱氨酶中的一个或多个对应突变。在一些实施方案中,腺苷脱氨酶包含对应于TadA参考序列中的本文所述突变中的一者或多者,或另一腺苷脱氨酶中的一个或多个对应突变。

[1303] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的E25X突变,或另一腺苷脱氨酶中的对应突变,其中X指示任何除野生型腺苷脱氨酶中对应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的E25M、E25D、E25A、E25R、E25V、E25S或E25Y突变,或另一腺苷脱氨酶(例如,ecTadA)中的对应突变。

[1304] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的R26X突变,或另一腺苷脱氨酶中的对应突变,其中X指示任何除野生型腺苷脱氨酶中对应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的R26G、R26N、R26Q、R26C、R26L或R26K突变,或另一腺苷脱氨酶(例如,ecTadA)中的对应突变。

[1305] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的R107X突变,或另一腺苷脱氨酶中的对应突变,其中X指示任何除野生型腺苷脱氨酶中对应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的R107P、R107K、R107A、R107N、R107W、R107H或R107S突变,或另一腺苷脱氨酶(例如,ecTadA)中的对应突变。

[1306] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的A142X突变,或另一腺苷脱氨酶中的对应突变,其中X指示任何除野生型腺苷脱氨酶中对应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的A142N、A142D、A142G突变,或另一腺苷脱氨酶(例如,ecTadA)中的对应突变。

[1307] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的A143X突变,或另一腺苷脱氨酶中的对应突变,其中X指示任何除野生型腺苷脱氨酶中对应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的A143D、A143G、A143E、A143L、A143W、

A143M、A143S、A143Q和/或A143R突变,或另一腺苷脱氨酶(例如,ecTadA)中的对应突变。

[1308] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H36X、N37X、P48X、I49X、R51X、M70X、N72X、D77X、E134X、S S146X、Q154X、K157X和/或K161X突变中的一者或多者,或另一腺苷脱氨酶中的一个或多个对应突变,其中X的存在指示除野生型腺苷脱氨酶中的对应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H36L、N37T、N37S、P48T、P48L、I49V、R51H、R51L、M70L、N72S、D77G、E134G、S146R、S146C、Q154H、K157N和/或K161T突变中的一者或多者,或另一腺苷脱氨酶(例如,ecTadA)中的一个或多个对应突变。

[1309] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H36X突变,或另一腺苷脱氨酶中的对应突变,其中X指示任何除野生型腺苷脱氨酶中对应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H36L突变,或另一腺苷脱氨酶中的对应突变。

[1310] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的N37X突变,或另一腺苷脱氨酶中的对应突变,其中X指示任何除野生型腺苷脱氨酶中对应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的N37T或N37S突变,或另一腺苷脱氨酶中的对应突变。

[1311] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的P48X突变,或另一腺苷脱氨酶中的对应突变,其中X指示任何除野生型腺苷脱氨酶中对应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的P48T或P48L突变,或另一腺苷脱氨酶中的对应突变。

[1312] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的P51X突变,或另一腺苷脱氨酶中的对应突变,其中X指示任何除野生型腺苷脱氨酶中对应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的R51H或R51L突变,或另一腺苷脱氨酶中的对应突变。

[1313] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的S146X突变,或另一腺苷脱氨酶中的对应突变,其中X指示任何除野生型腺苷脱氨酶中对应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的S146R或S146C突变,或另一腺苷脱氨酶中的对应突变。

[1314] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的K157X突变,或另一腺苷脱氨酶中的对应突变,其中X指示任何除野生型腺苷脱氨酶中对应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的K157N突变,或另一腺苷脱氨酶中的对应突变。

[1315] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的P48X突变,或另一腺苷脱氨酶中的对应突变,其中X指示任何除野生型腺苷脱氨酶中对应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的P48S、P48T或P48A突变,或另一腺苷脱氨酶中的对应突变。

[1316] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的A142X突变,或另一腺苷脱氨酶中的对应突变,其中X指示任何除野生型腺苷脱氨酶中对应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的A142N突变,或另一腺苷脱氨酶中的

对应突变。

[1317] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的W23X突变,或另一腺苷脱氨酶中的对应突变,其中X指示任何除野生型腺苷脱氨酶中对应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的W23R或W23L突变,或另一腺苷脱氨酶中的对应突变。

[1318] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的R152X突变,或另一腺苷脱氨酶中的对应突变,其中X指示任何除野生型腺苷脱氨酶中对应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的R152P或R52H突变,或另一腺苷脱氨酶中的对应突变。

[1319] 在一种实施方案中,腺苷脱氨酶可包含突变H36L、R51L、L84F、A106V、D108N、H123Y、S146C、D147Y、E155V、I156F和K157N。在一些实施方案中,腺苷脱氨酶包含以下相对于TadA参考序列的突变组合,其中组合的每个突变以“_”分隔并且突变的每个组合在括号内:(A106V_D108N)、(R107C_D108N)、(H8Y_D108N_N127S_D147Y_Q154H)、(H8Y_D108N_N127S_D147Y_E155V)、(D108N_D147Y_E155V)、(H8Y_D108N_N127S)、(H8Y_D108N_N127S_D147Y_Q154H)、(A106V_D108N_D147Y_E155V)、(D108Q_D147Y_E155V)、(D108M_D147Y_E155V)、(D108L_D147Y_E155V)、(D108K_D147Y_E155V)、(D108I_D147Y_E155V)、(D108F_D147Y_E155V)、(A106V_D108N_D147Y)、(A106V_D108M_D147Y_E155V)、(E59A_A106V_D108N_D147Y_E155V)、(E59A催化死亡的_A106V_D108N_D147Y_E155V)、(L84F_A106V_D108N_H123Y_D147Y_E155V_I156Y)、(L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)、(D103A_D104N)、(G22P_D103A_D104N)、(D103A_D104N_S138A)、(R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F)、(E25G_R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F)、(E25D_R26G_L84F_A106V_R107K_D108N_H123Y_A142N_A143G_D147Y_E155V_I156F)、(R26Q_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F)、(E25M_R26G_L84F_A106V_R107P_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F)、(R26C_L84F_A106V_R107H_D108N_H123Y_A142N_D147Y_E155V_I156F)、(L84F_A106V_D108N_H123Y_A142N_A143L_D147Y_E155V_I156F)、(R26G_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F)、(E25A_R26G_L84F_A106V_R107N_D108N_H123Y_A142N_A143E_D147Y_E155V_I156F)、(R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F)、(A106V_D108N_A142N_D147Y_E155V)、(R26G_A106V_D108N_A142N_D147Y_E155V)、(E25D_R26G_A106V_R107K_D108N_A142N_A143G_D147Y_E155V)、(R26G_A106V_D108N_R107H_A142N_A143D_D147Y_E155V)、(E25D_R26G_A106V_D108N_A142N_D147Y_E155V)、(A106V_R107K_D108N_A142N_D147Y_E155V)、(A106V_D108N_A142N_A143G_D147Y_E155V)、(A106V_D108N_A142N_A143L_D147Y_E155V)、(H36L_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N)、(N37T_P48T_M70L_L84F_A106V_D108N_H123Y_D147Y_I49V_E155V_I156F)、(N37S_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K161T)、(H36L_L84F_A106V_D108N_H123Y_D147Y_Q154H_E155V_I156F)、(N72S_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F)、(H36L_P48L_L84F_A106V_D108N_H123Y_E134G_D147Y_E155V_I156F)、(H36L_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K157N)、(H36L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F)、(L84F_A106V_

D108N_H123Y_S146R_D147Y_E155V_I156F_K161T)、(N37S_R51H_D77G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)、(R51L_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K157N)、(D24G_Q71R_L84F_H96L_A106V_D108N_H123Y_D147Y_E155V_I156F_K160E)、(H36L_G67V_L84F_A106V_D108N_H123Y_S146T_D147Y_E155V_I156F)、(Q71L_L84F_A106V_D108N_H123Y_L137M_A143E_D147Y_E155V_I156F)、(E25G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_Q159L)、(L84F_A91T_F104I_A106V_D108N_H123Y_D147Y_E155V_I156F)、(N72D_L84F_A106V_D108N_H123Y_G125A_D147Y_E155V_I156F)、(P48S_L84F_S97C_A106V_D108N_H123Y_D147Y_E155V_I156F)、(W23G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)、(D24G_P48L_Q71R_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_Q159L)、(L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F)、(H36L_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F_K157N)、(N37S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_K161T)、(L84F_A106V_D108N_D147Y_E155V_I156F)、(R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K161T)、(L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K161T)、(L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K160E_K161T)、(L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K160E)、(R74Q_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)、(R74A_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)、(L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)、(R74Q_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)、(L84F_R98Q_A106V_D108N_H123Y_D147Y_E155V_I156F)、(L84F_A106V_D108N_H123Y_R129Q_D147Y_E155V_I156F)、(P48S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F)、(P48S_A142N)、(P48T_I49V_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_L157N)、(P48T_I49V_A142N)、(H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N)、(H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S146C_A142N_D147Y_E155V_I156)F、(H36L_P48T_I49V_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N)、(H36L_P48T_I49V_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F_K157N)、(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N)、(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F_K157N)、(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_A142N_D147Y_E155V_I156F_K157N)、(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N)、(W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N)、(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F_K161T)、(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152H_E155V_I156F_K157N)、(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F_K157N)、(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F_K157N)、(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142A_S146C_D147Y_E155V_I156F_K157N)、(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142A_S146C_D147Y_R152P_E155V_I156F_K157N)、(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F_K161T)、(W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F_K157N)、(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_

A142N_S146C_D147Y_R152P_E155V_I156F_K157N)。

[1320] 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*7.10。在特定实施方案中,融合蛋白包含单一TadA*7.10(例如,作为单体提供)。在其他实施方案中,融合蛋白包含TadA*7.10和TadA(wt),它们能够形成异二聚体。在一种实施方案中,本发明的融合蛋白包含链接至TadA*7.10的野生型TadA,TadA7.10链接至Cas9切口酶。

[1321] 在一些实施方案中,TadA*7.10包含至少一个改变。在一些实施方案中,腺苷脱氨酶包含在以下序列中的改变:

[1322] TadA*7.10

[1323] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

[1324] 在一些实施方案中,TadA*7.10包含氨基酸82和/或166处的改变。在特定实施方案中,TadA*7.10包含以下改变中的一者或多者:Y147T、Y147R、Q154S、Y123H、V82S、T166R和/或Q154R。在其他实施方案中,TadA*7.10的变体包含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;和I76Y+V82S+Y123H+Y147R+Q154R。

[1325] 在一些实施方案中,腺苷脱氨酶变体(例如,TadA*8)包含缺少。在一些实施方案中,腺苷脱氨酶变体包含C端的缺失。在特定实施方案中,腺苷脱氨酶变体包含在残基149、150、151、152、153、154、155、156和157处开始的C端缺失,相对于TadA*7.10(TadA参考序列),或另一TadA中的相对应突变。

[1326] 在其他实施方案中,腺苷脱氨酶变体(例如,TadA*8)是单体,其包含以下改变中的一者或多者:Y147T、Y147R、Q154S、Y123H、V82S、T166R和/或Q154R,相对于TadA*7.10(TadA参考序列),或另一TadA中的对应突变。在其他实施方案中,腺苷脱氨酶变体(TadA*8)是单体,其包含选自以下组的改变的组合:Y147T+Q154R;Y147T+Q154S;Y147R+Q154S;V82S+Q154S;V82S+Y147R;V82S+Q154R;V82S+Y123H;I76Y+V82S;V82S+Y123H+Y147T;V82S+Y123H+Y147R;V82S+Y123H+Q154R;Y147R+Q154R+Y123H;Y147R+Q154R+I76Y;Y147R+Q154R+T166R;Y123H+Y147R+Q154R+I76Y;V82S+Y123H+Y147R+Q154R;和I76Y+V82S+Y123H+Y147R+Q154R,相对于TadA*7.10(TadA参考序列),或另一TadA中的对应突变。

[1327] 在其他实施方案中,腺苷脱氨酶变体是均二聚体,其包含两个腺苷脱氨酶结构域(例如,TadA*8),各自具有以下突变中的一者或多者:Y147T、Y147R、Q154S、Y123H、V82S、T166R和/或Q154R,相对于TadA*7.10(TadA参考序列),或另一TadA中的对应突变。在其他实施方案中,腺苷脱氨酶变体是均二聚体,其包含两个腺苷脱氨酶结构域(例如,TadA*8),各自具有选自以下的组的改变的组合:Y147T+Q154R;Y147T+Q154S;Y147R+Q154S;V82S+Q154S;V82S+Y147R;V82S+Q154R;V82S+Y123H;I76Y+V82S;V82S+Y123H+Y147T;V82S+Y123H+Y147R;V82S+Y123H+Q154R;Y147R+Q154R+Y123H;Y147R+Q154R+I76Y;Y147R+Q154R+T166R;Y123H+Y147R+Q154R+I76Y;V82S+Y123H+Y147R+Q154R;和I76Y+V82S+Y123H+Y147R+Q154R,

相对于TadA*7.10(TadA参考序列),或另一TadA中的对应突变。

[1328] 在其他实施方案中,腺苷脱氨酶变体是野生型腺苷脱氨酶结构域和腺苷脱氨酶变体(例如,TadA*8)的异二聚体,该腺苷脱氨酶变体包含以下改变中的一者或多者:Y147T、Y147R、Q154S、Y123H、V82S、T166R和/或Q154R,相对于TadA*7.10(TadA参考序列),或另一TadA中的对应突变。在其他实施方案中,腺苷脱氨酶变体是野生型腺苷脱氨酶结构域和腺苷脱氨酶变体结构域(例如,TadA*8)的异二聚体,该腺苷脱氨酶变体结构域包含选自以下组的改变的组合:Y147T+Q154R;Y147T+Q154S;Y147R+Q154S;V82S+Q154S;V82S+Y147R;V82S+Q154R;V82S+Y123H;I76Y+V82S;V82S+Y123H+Y147T;V82S+Y123H+Y147R;V82S+Y123H+Q154R;Y147R+Q154R+Y123H;Y147R+Q154R+I76Y;Y147R+Q154R+T166R;Y123H+Y147R+Q154R+I76Y;V82S+Y123H+Y147R+Q154R;和I76Y+V82S+Y123H+Y147R+Q154R,相对于TadA*7.10(TadA参考序列),或另一TadA中的对应突变。

[1329] 在其他实施方案中,腺苷脱氨酶变体是TadA*7.10结构域和腺苷脱氨酶变体(例如,TadA*8)的异二聚体,该腺苷脱氨酶变体包含以下改变中的一者或多者:Y147T、Y147R、Q154S、Y123H、V82S、T166R和/或Q154R,相对于TadA*7.10(TadA参考序列),或另一TadA中的对应突变。在其他实施方案中,腺苷脱氨酶变体是TadA*7.10结构域和腺苷脱氨酶变体结构域(例如,TadA*8)的异二聚体,该腺苷脱氨酶变体结构域包含选自以下组的改变的组合:Y147T+Q154R;Y147T+Q154S;Y147R+Q154S;V82S+Q154S;V82S+Y147R;V82S+Q154R;V82S+Y123H;I76Y+V82S;V82S+Y123H+Y147T;V82S+Y123H+Y147R;V82S+Y123H+Q154R;Y147R+Q154R+Y123H;Y147R+Q154R+I76Y;Y147R+Q154R+T166R;Y123H+Y147R+Q154R+I76Y;V82S+Y123H+Y147R+Q154R;和I76Y+V82S+Y123H+Y147R+Q154R,相对于TadA*7.10(TadA参考序列),或另一TadA中的对应突变。

[1330] 在特定实施方案中,腺苷脱氨酶异二聚体包含TadA*8结构域和腺苷脱氨酶结构域,该腺苷脱氨酶结构域选自金哈UN各色葡萄球菌TadA、枯草芽孢杆菌TadA、鼠伤寒沙门氏菌TadA、腐败希瓦氏菌TadA、流感嗜血杆菌F3031 TadA、新月柄杆菌TadA、硫还原地杆菌TadA或TadA*7.10。

[1331] 在一些实施方案中,腺苷脱氨酶是TadA*8。在一种实施方案中,腺苷脱氨酶包含TadA*8,其包含以下序列或其具有腺苷脱氨酶活性的片段或基本上由其组成:

[1332] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD

[1333] 一些实施方案中,TadA*8被截短。在一些实施方案中,相对于全长TadA*8,截短的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N端氨基酸残基。在一些实施方案中,相对于全长TadA*8,截短的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,腺苷脱氨酶是全长TadA*8。

[1334] 在一些实施方案中,TadA*8是TadA*8.1、TadA*8.2、TadA*8.3、TadA*8.4、TadA*8.5、TadA*8.6、TadA*8.7、TadA*8.8、TadA*8.9、TadA*8.10、TadA*8.11、TadA*8.12、TadA*8.13、TadA*8.14、TadA*8.15、TadA*8.16、TadA*8.17、TadA*8.18、TadA*8.19、TadA*8.20、TadA*8.21、TadA*8.22、TadA*8.23或TadA*8.24。

[1335] 在其他实施方案中,本公开的碱基编辑器包含腺苷脱氨酶变体(例如,TadA*8)单体,该单体包含以下改变中的一者或多者:R26C、V88A、A109S、T111R、D119N、H122N、Y147D、F149Y、T166I and/or D167N,相对于TadA*7.10(TadA参考序列),或另一TadA中的对应突变。在其他实施方案中,腺苷脱氨酶变体(TadA*8)单体包含选自以下组的改变的组合:R26C+A109S+T111R+D119N+H122N+Y147D+F149Y+T166I+D167N;V88A+A109S+T111R+D119N+H122N+F149Y+T166I+D167N;R26C+A109S+T111R+D119N+H122N+F149Y+T166I+D167N;V88A+T111R+D119N+F149Y;and A109S+T111R+D119N+H122N+Y147D+F149Y+T166I+D167N,相对于TadA*7.10(TadA参考序列),或另一TadA中的对应突变。

[1336] 在其他实施方案中,碱基编辑器包含野生型腺苷脱氨酶结构域和腺苷脱氨酶变体结构域(例如,TadA*8)的异二聚体,该腺苷脱氨酶变体包含以下改变中的一者或多者:R26C、V88A、A109S、T111R、D119N、H122N、Y147D、F149Y、T166I和/或D167N,相对于TadA*7.10(TadA参考序列),或另一TadA中的对应突变。在其他实施方案中,碱基编辑器包含野生型腺苷脱氨酶结构域和腺苷脱氨酶变体结构域(例如,TadA*8)的异二聚体,该腺苷脱氨酶变体结构域包含选自以下组的改变的组合:R26C+A109S+T111R+D119N+H122N+Y147D+F149Y+T166I+D167N;V88A+A109S+T111R+D119N+H122N+F149Y+T166I+D167N;R26C+A109S+T111R+D119N+H122N+F149Y+T166I+D167N;V88A+T111R+D119N+F149Y;and A109S+T111R+D119N+H122N+Y147D+F149Y+T166I+D167N,相对于TadA*7.10(TadA参考序列),或另一TadA中的对应突变。

[1337] 在其他实施方案中,碱基编辑器包含TadA*7.10结构域和腺苷脱氨酶变体结构域(例如,TadA*8)的异二聚体,该腺苷脱氨酶变体包含以下改变中的一者或多者:R26C、V88A、A109S、T111R、D119N、H122N、Y147D、F149Y、T166I和/或D167N,相对于TadA*7.10(TadA参考序列),或另一TadA中的对应突变。在其他实施方案中,碱基编辑器包含TadA*7.10结构域和腺苷脱氨酶变体结构域(例如,TadA*8)的异二聚体,该腺苷脱氨酶变体结构域包含选自以下组的改变的组合:R26C+A109S+T111R+D119N+H122N+Y147D+F149Y+T166I+D167N;V88A+A109S+T111R+D119N+H122N+F149Y+T166I+D167N;R26C+A109S+T111R+D119N+H122N+F149Y+T166I+D167N;V88A+T111R+D119N+F149Y;and A109S+T111R+D119N+H122N+Y147D+F149Y+T166I+D167N,相对于TadA*7.10(TadA参考序列),或另一TadA中的对应突变。

[1338] 在一些实施方案中,TadA*8是如表11中所示的变体。表11显示TadA氨基酸序列中的某些氨基酸位置编号,并且存在于TadA-7.10腺苷脱氨酶中的那些位置处的氨基酸。表11也显示在噬菌体辅助非连续进化(PANCE)和噬菌体辅助连续进化(PACE)后,TadA变体中相对于TadA-7.10的氨基酸变化,如M.Richter等人,2020,Nature Biotechnology,doi.org/10.1038/s41587-020-0453-z中所述,其整体内容通过引用并入本文。在一些实施方案中,TadA*8是TadA*8a、TadA*8b、TadA*8c、TadA*8d或TadA*8e。在一些实施方案中,TadA*8是TadA*8e。

[1339] 表11.其他TadA*8变体

[1340]

		TadA氨基酸编号									
	TadA	26	88	109	111	119	122	147	149	166	167
	TadA-7.10	R	V	A	T	D	H	Y	F	T	D
PANCE 1					R						
PANCE 2				S/T	R						
	TadA-8a	C		S	R	N	N	D	Y	I	N
	TadA-8b		A	S	R	N	N		Y	I	N
PACE	TadA-8c	C		S	R	N	N		Y	I	N
	TadA-8d		A		R	N			Y		
	TadA-8e			S	R	N	N	D	Y	I	N

[1341] 在一种实施方案中,本发明的融合蛋白包含链接至Cas9切口酶的本文所述的腺苷脱氨酶变体(例如,TadA*8)。在特定实施方案中,融合蛋白包含单一TadA*8(例如,作为单体提供)。在其他实施方案中,融合蛋白包含TadA*8和TadA(wt),它们能够形成异二聚体。

[1342] 在一些实施方案中,腺苷脱氨酶包含氨基酸序列,该氨基酸序列与本文所提供的任何腺苷脱氨酶中详述的氨基酸序列中的任一者至少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。应了解,本文所提供的腺苷脱氨酶可包括一个或多个突变(例如,本文所提供的任何突变)。本公开提供任何具有某一百分比的一致性加上任何本文所述突变或其组合的脱氨酶结构域。在一些实施方案中,腺苷脱氨酶包含氨基酸序列,与参考序列或本文所提供的任何腺苷脱氨酶相比,该氨基酸序列具有1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、21、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50个或更多个突变。在一些实施方案中,腺苷脱氨酶包含氨基酸序列,与本领域已知或本文所述的氨基酸序列中的任一者相比,该氨基酸序列具有至少5、至少10、至少15、至少20、至少25、至少30、至少35、至少40、至少45、至少50、至少60、至少70、至少80、至少90、至少100、至少110、至少120、至少130、至少140、至少150、至少160、或至少170个相同的连续氨基酸残基。

[1343] 在特定实施方案中,TadA*8包含在以下以黑体显示的任何位置处的一个或多个突变。在其他实施方案中,TadA*8包含在以下划线显示的任何位置处的一个或多个突变。

[1344]

MSEVEFSHEY WMRHALTLAK RARDEREVPV GAVLVLNNRV IGEGWNRAIG 50

LHDPTAHAEI MALRQGGLVM QNYRLIDATL YVTFEPCVMC AGAMIHSRIG 100

RVVFGVRNAK TGAAGSLMDV LHYPGMNHRV EITEGILADE CAALLCYFFR 150

MPRQVFNAQK KAQSSTD

[1345] 例如,TadA*8包含在氨基酸位置82和/或166处的改变(例如,V82S,T166R),该改变单独存在或与以下中的一者或多者组合:Y147T、Y147R、Q154S、Y123H和/或Q154R,相对于TadA*7.10(TadA参考序列),或另一TadA的相对应突变。在特定实施方案中,改变的组合选自由以下的组: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;和I76Y+V82S+Y123H+Y147R+Q154R,相对于TadA*7.10(TadA参考序列),或另一TadA中的对应突变。

[1346] 在一些实施方案中,腺苷脱氨酶包含TadA*8,其包含以下序列或其具有腺苷脱氨酶活性的片段或基本上由其组成:

[1347] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD

[1348] 一些实施方案中,TadA*8被截短。在一些实施方案中,相对于全长TadA*8,截短的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N端氨基酸残基。在一些实施方案中,相对于全长TadA*8,截短的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,腺苷脱氨酶是全长TadA*8。

[1349] 在一种实施方案中,本发明的融合蛋白包含链接至Cas9切口酶的本文所述的腺苷脱氨酶变体(例如,TadA*8)。在特定实施方案中,融合蛋白包含单一TadA*8(例如,作为单体提供)。在其他实施方案中,碱基编辑器包含TadA*8和TadA(wt),它们能够形成异二聚体。

[1350] 在特定实施方案中,融合蛋白包含单一(例如,作为单体提供)TadA*8。在一些实施方案中,TadA*8连接至Cas9切口酶。在一些实施方案中,本发明的融合蛋白包含野生型TadA(TadA(wt))链接至TadA*8的异二聚体。在其他实施方案中,本发明的融合蛋白包含野生型TadA*7.10链接至TadA*8的异二聚体。在一些实施方案中,碱基编辑器是包含TadA*8变体单体的ABE8。在一些实施方案中,碱基编辑器是包含TadA*8与TadA(wt)异二聚体的ABE8。在一些实施方案中,碱基编辑器是包含TadA*8与TadA*7.10的异二聚体的ABE8。在一些实施方案中,碱基编辑器是包含TadA*8的异二聚体的ABE8。在一些实施方案中,TadA*8选自表11、13或14。在一些实施方案中,ABE8选自表13、14或16。

[1351] 在一些实施方案中,腺苷脱氨酶是TadA*9变体。在一些实施方案中,腺苷脱氨酶是选自下述变体的TadA*9变体且参考以下序列(称为TadA*7.10):

[1352]

```
        10        20        30        40        50
MSEVEFSHEY WMRHALTLAK RARDEREVPV GAVLVLNNRV IGEGWNRAIG
        60        70        80        90       100
LHDPTAHAEI MALRQGGLVM QNYRLIDATL YVTFEPCVMC AGAMIHSRIG
       110       120       130       140       150
RVVFGVRNAK TGAAGSLMDV LHYPGMNHRV EITEGILADE CAALLCYFFR
       160
MPRQVFNAQK KAQSSTD
```

[1353] 在一些实施方案中,腺苷脱氨酶包含以下改变中的一者或多者:R21N、R23H、E25F、N38G、L51W、P54C、M70V、Q71M、N72K、Y73S、V82T、M94V、P124W、T133K、D139L、D139M、C146R和A158K。一种或多种改变在上述序列中以加下划线和加粗字体显示。

[1354] 在一些实施方案中,腺苷脱氨酶包含以下改变的组合中的一者或多者:V82S+Q154R+Y147R;V82S+Q154R+Y123H;V82S+Q154R+Y147R+Y123H;Q154R+Y147R+Y123H+I76Y+V82S;V82S+I76Y;V82S+Y147R;V82S+Y147R+Y123H;V82S+Q154R+Y123H;Q154R+Y147R+Y123H+I76Y;V82S+Y147R;V82S+Y147R+Y123H;V82S+Q154R+Y123H;V82S+Q154R+Y147R;V82S+Q154R+Y147R;Q154R+Y147R+Y123H+I76Y;Q154R+Y147R+Y123H+I76Y+V82S;I76Y_V82S_

Y123H_Y147R_Q154R;Y147R+Q154R+H123H;和V82S+Q154R。

[1355] 在一些实施方案中,腺苷脱氨酶包含以下改变的组合中的一者或多者:E25F+V82S+Y123H,T133K+Y147R+Q154R;E25F+V82S+Y123H+Y147R+Q154R;L51W+V82S+Y123H+C146R+Y147R+Q154R;Y73S+V82S+Y123H+Y147R+Q154R;P54C+V82S+Y123H+Y147R+Q154R;N38G+V82T+Y123H+Y147R+Q154R;N72K+V82S+Y123H+D139L+Y147R+Q154R;E25F+V82S+Y123H+D139M+Y147R+Q154R;Q71M+V82S+Y123H+Y147R+Q154R;E25F+V82S+Y123H+T133K+Y147R+Q154R;E25F+V82S+Y123H+Y147R+Q154R;V82S+Y123H+P124W+Y147R+Q154R;L51W+V82S+Y123H+C146R+Y147R+Q154R;P54C+V82S+Y123H+Y147R+Q154R;Y73S+V82S+Y123H+Y147R+Q154R;N38G+V82T+Y123H+Y147R+Q154R;R23H+V82S+Y123H+Y147R+Q154R;R21N+V82S+Y123H+Y147R+Q154R;V82S+Y123H+Y147R+Q154R+A158K;N72K+V82S+Y123H+D139L+Y147R+Q154R;E25F+V82S+Y123H+D139M+Y147R+Q154R;和M70V+V82S+M94V+Y123H+Y147R+Q154R

[1356] 在一些实施方案中,腺苷脱氨酶包含以下改变的组合中的一者或多者:Q71M+V82S+Y123H+Y147R+Q154R;E25F+I76Y+V82S+Y123H+Y147R+Q154R;I76Y+V82T+Y123H+Y147R+Q154R;N38G+I76Y+V82S+Y123H+Y147R+Q154R;R23H+I76Y+V82S+Y123H+Y147R+Q154R;P54C+I76Y+V82S+Y123H+Y147R+Q154R;R21N+I76Y+V82S+Y123H+Y147R+Q154R;I76Y+V82S+Y123H+D139M+Y147R+Q154R;Y73S+I76Y+V82S+Y123H+Y147R+Q154R;E25F+I76Y+V82S+Y123H+Y147R+Q154R;I76Y+V82T+Y123H+Y147R+Q154R;N38G+I76Y+V82S+Y123H+Y147R+Q154R;R23H+I76Y+V82S+Y123H+Y147R+Q154R;P54C+I76Y+V82S+Y123H+Y147R+Q154R;R21N+I76Y+V82S+Y123H+Y147R+Q154R;I76Y+V82S+Y123H+D139M+Y147R+Q154R;Y73S+I76Y+V82S+Y123H+Y147R+Q154R;and V82S+Q154R;N72K_V82S+Y123H+Y147R+Q154R;Q71M_V82S+Y123H+Y147R+Q154R;V82S+Y123H+T133K+Y147R+Q154R;V82S+Y123H+T133K+Y147R+Q154R+A158K;M70V+Q71M+N72K+V82S+Y123H+Y147R+Q154R;N72K_V82S+Y123H+Y147R+Q154R;Q71M_V82S+Y123H+Y147R+Q154R;M70V+V82S+M94V+Y123H+Y147R+Q154R;V82S+Y123H+T133K+Y147R+Q154R;V82S+Y123H+T133K+Y147R+Q154R+A158K;和M70V+Q71M+N72K+V82S+Y123H+Y147R+Q154R。在一些实施方案中,腺苷脱氨酶作为单体表达。在其他实施方案中,腺苷脱氨酶作为异二聚体表达。在一些实施方案中,脱氨酶或其他多肽序列缺少甲硫氨酸,例如,当作为融合蛋白的组分而被包括时。这可能改变位置的编号。但是,技术人员将会理解,此类相对应突变是指相同突变,例如,Y73S与Y72S以及D139M与D138M。

[1357] 在一些实施方案中,TadA*9变体包含表17中所述的改变,如本文所述。在一些实施方案中,TadA*9变体是单体。在一些实施方案中,TadA*9变体是具有野生型TadA腺苷脱氨酶的异二聚体。在一些实施方案中,TadA*9变体是具有另一TadA变体(例如,TadA*8、TadA*9)的异二聚体。TadA*9腺苷脱氨酶的其他细节在国际PCT申请号PCT/2020/049975中描述,其通过引用以其整体并入本文。

[1358] 本文所提供的任何突变和任何额外突变(例如,基于ecTadA氨基酸序列)可引入任何其他腺苷脱氨酶内。本文所提供的任何突变可以独立地作出或与TadA参考序列或另一腺苷脱氨酶(例如,ecTadA)中的突变以任意组合作出。

[1359] A到G核碱基编辑蛋白的细节描述于国际PCT申请号PCT/2017/045381(WO2018/027078)和Gaudelli,N.M.,等人,"Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage"Nature,551,464-471(2017)中,其整体内容通过引

用并入本文。

[1360] C到T编辑

[1361] 在一些实施方案中,本文所公开的碱基编辑器包含融合蛋白,该融合蛋白包含能够将多核苷酸的靶标胞苷(C)碱基脱氨基以产生尿苷(U)(其具有胸腺嘧啶的碱基配对性质)的胞苷脱氨酶。在一些实施方案中,例如其中多核苷酸是双链的(例如,DNA),尿苷碱基随后可以被置换为胸苷碱基(例如,通过细胞修复机构)以引起C:G到T:A的转换。在其他实施方案中,通过碱基编辑器将核酸中的C脱氨基为U不能伴随U到T的置换。

[1362] 将多核苷酸中的靶标C氨基以给出U是一种类型的碱基编辑的非限制性示例,该碱基编辑可以通过本文所述的碱基编辑器执行。在另一示例中,包含胞苷脱氨酶结构域的碱基编辑器可介导胞嘧啶(C)碱基到鸟嘌呤(G)碱基的转变。例如,通过碱基编辑器的胞苷脱氨酶结构域使胞苷脱氨基而产生的多核苷酸的U可以通过碱基切除修复机制(例如,通过尿嘧啶DNA糖基化酶(UDG)结构域)从多核苷酸中切除,从而产生无碱基位点。然后,可以通过例如跨损伤聚合酶将与该无碱基位点相对的核碱基置换为(例如,通过甲基修复机构)另一碱基,诸如C。尽管与无碱基位点相对的核碱基通常被替换为C,但也可能发生其他置换(例如,A、G或T)。

[1363] 据此,在一些实施方案中,本文所述的碱基编辑器包含脱氨酶结构域(例如,胞苷脱氨酶结构域),该脱氨酶结构域能够将多核苷酸中的靶标C脱氨基为U。再者,如下文所述,碱基编辑器可包含其他结构域,在一些实施方案中,该其他结构域促进脱氨基所致的U转换为T或G。例如,包含胞苷脱氨酶结构域的碱基编辑器可进一步包含尿嘧啶糖基化酶抑制剂(UGI)结构域以介导U被T置换,从而完成C到T的碱基编辑事件。在另一示例中,碱基编辑器可合并有跨损伤聚合酶以改善C到G碱基编辑的效率,因为跨损伤聚合酶可促进并入与无碱基位点相对的C(即,导致G并入在无碱基位点处,从而完成C到G的碱基编辑事件)。

[1364] 包含胞苷脱氨酶作为结构域的碱基编辑器可将任何多核苷酸(包括DNA、RNA和DNA-RNA杂交物)中的靶标C脱氨基。通常,胞苷脱氨酶催化在多核苷酸的单链部分的情境中定位的C核碱基。在一些实施方案中,包含靶标C的整体多核苷酸可以是单链的。例如,并入碱基编辑器内的胞苷脱氨酶可将单链RNA多核苷酸中的靶标C脱氨基。在其他实施方案中,包含胞苷脱氨酶的碱基编辑器可作用于双链多核苷酸,但靶标C可以定位在多核苷酸的在脱氨基反应期间处于单链状态的部分中。例如,在其中NAGPB结构域包含Cas9结构域的实施方案中,几个核苷酸在Cas9-gRNA-靶标DNA复合物的形成期间可以保持未配对,胞质形成Cas9“R环复合物”。这些未配对的核苷酸可以形成单链DNA鼓泡,该鼓泡可以充当单链特异性核苷酸脱氨酶(例如,胞苷脱氨酶)的底物。

[1365] 在一些实施方案中,碱基编辑器的胞苷脱氨酶可包含载脂蛋白BmRNA编辑复合物(APOBEC)家族脱氨酶的全部或部分。APOBEC是一个进化上保守的胞苷脱氨酶家族。这一家族的成员是C到U编辑酶。APOBEC样蛋白的N端结构域是催化性结构域,而C端结构域是假催化性结构域。更具体地,催化性结构域是锌依赖性胞苷脱氨酶结构域,且对于胞苷脱氨基而言非常重要。APOBEC家族成员包括APOBEC1、APOBEC2、APOBEC3A、APOBEC3B、APOBEC3C、APOBEC3D(现在称为“APOBEC3E”)、APOBEC3F、APOBEC3G、APOBEC3H、APOBEC4和激活诱导型(胞苷)脱氨酶。在一些实施方案中,并入碱基编辑器中的脱氨酶包含APOBEC1脱氨酶的全部或一部分。在一些实施方案中,并入碱基编辑器中的脱氨酶包含APOBEC2脱氨酶的全部或一

部分。在一些实施方案中,并入碱基编辑器中的脱氨酶包含APOBEC3脱氨酶的全部或一部分。在一些实施方案中,并入碱基编辑器中的脱氨酶包含APOBEC3A脱氨酶的全部或一部分。在一些实施方案中,并入碱基编辑器中的脱氨酶包含APOBEC3B脱氨酶的全部或一部分。在一些实施方案中,并入碱基编辑器中的脱氨酶包含APOBEC3C脱氨酶的全部或一部分。在一些实施方案中,并入碱基编辑器中的脱氨酶包含APOBEC3D脱氨酶的全部或一部分。在一些实施方案中,并入碱基编辑器中的脱氨酶包含APOBEC3E脱氨酶的全部或一部分。在一些实施方案中,并入碱基编辑器中的脱氨酶包含APOBEC3F脱氨酶的全部或一部分。在一些实施方案中,并入碱基编辑器中的脱氨酶包含APOBEC3G脱氨酶的全部或一部分。在一些实施方案中,并入碱基编辑器中的脱氨酶包含APOBEC3H脱氨酶的全部或一部分。在一些实施方案中,并入碱基编辑器中的脱氨酶包含APOBEC4脱氨酶的全部或一部分。在一些实施方案中,并入碱基编辑器中的脱氨酶包含激活诱导脱氨酶(AID)的全部或一部分。在一些实施方案中,并入碱基编辑器中的脱氨酶包含胞苷脱氨酶1(CDA1)的全部或一部分。应了解,碱基编辑器可包含来自任何适合生物体(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来源自大鼠(例如,大鼠APOBEC1)。在一些实施方案中,碱基编辑器的脱氨酶结构域是人类APOBEC1。在一些实施方案中,碱基编辑器的脱氨酶结构域是pmCDA1。

[1366] pmCDA1的氨基酸和核酸序列显示在下文中。

[1367] >tr|A5H718|A5H718_PETMA胞嘧啶脱氨酶OS=海七鳃鳗(Petromyzon marinus)OX=7757PE=2SV=1氨基酸序列:

[1368] MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQSGTERGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLKIWACKLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKRRSELSIMIQVKILHTTKSPAV

[1369] 核酸序列:>EF094822.1海七鳃鳗分离PmCDA.21胞嘧啶脱氨酶mRNA,完全编码序列:

[1370] TGACACGACACAGCCGTGTATATGAGGAAGGGTAGCTGGATGGGGGGGGGGGGAATACGTTCAGAGAGGACATTAGCGAGCGTCTTGTTGGTGGCCTTGAGTCTAGACACCTGCAGACATGACCGACGCTGAGTACGTGAGAATCCATGAGAAGTTGGACATCTACACGTTTAAGAAACAGTTTTTCAACAACAAAAAATCCGTGTCGCATAGATGCTACGTTCTCTTTGAATTAAAACGACGGGGTGAACGTAGAGCGTGTTTTTGGGGCTATGCTGTGAATAAACCACAGAGCGGGACAGAACGTGGAATTCACGCCGAAATCTTTAGCATTAGAAAAGTCGAAGAATACCTGCGCGACAACCCCGGACAATTCACGATAAATTGGTACTCATCCTGGAGTCCTTGTGCAGATTGCGCTGAAAAGATCTTAGAATGGTATAACCAGGAGCTGCGGGGGAACGGCCACACTTTGAAAATCTGGGCTTGCAAACTCTATTACGAGAAAAATGCGAGGAATCAAATTGGGCTGTGGAACCTCAGAGATAACGGGGTTGGGTTGAATGTAATGGTAAGTGAACACTACCAATGTTGCAGGAAAATATTCATCCAATCGTCGCACAATCAATTGAATGAGAATAGATGGCTTGAGAAGACTTTGAAGCGAGCTGAAAAACGACGGAGCGAGTTGTCCATTATGATTCAGGTAAAAATACTCCACACCACTAAGAGTCCTGCTGTTTAAGAGGCTATGCGGATGGTTTTC

[1371] 人类激活诱导型胞苷脱氨酶(AID)的编码序列(CDS)的氨基酸和核酸序列显示于下。

[1372] >tr|Q6QJ80|Q6QJ80_人类激活诱导型胞苷脱氨酶OS=智人OX=9606GN=AICDA

PE＝2SV＝1氨基酸序列：

[1373] MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELLFLRYISDWD
LDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKAPV

[1374] 人类激活诱导型胞苷脱氨酶(AID)的编码序列(CDS)的氨基酸和核酸序列显示于下。>tr|Q6QJ80|Q6QJ80_人类激活诱导型胞苷脱氨酶OS＝智人OX＝9606GN＝AICDA PE＝2SV＝1氨基酸序列：

[1375] MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELLFLRYISDWD
LDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKAPV

[1376] 核酸序列:>NG_011588.1:5001-15681智人激活诱导型胞苷脱氨酶(AICDA),RefSeqGene(LRG_17),在12号染色体上：

[1377] AGAGAACCATCATTAATTGAAGTGAGATTTTTCTGGCCTGAGACTTGCAGGGAGGCAAGAAGACACTCT
GGACACCACTATGGACAGGTAAAGAGGCAGTCTTCTCGTGGGTGATTGCACTGGCCTTCCTCTCAGAGCAAATCTGA
GTAATGAGACTGGTAGCTATCCCTTTCTCTCATGTAACTGTCTGACTGATAAGATCAGCTTGATCAATATGCATATA
TATTTTTTGATCTGTCTCCTTTTCTTCTATTCAGATCTTATACGCTGTCAGCCCAATTCTTTCTGTTTCAGACTTCT
CTTGATTTCCCTCTTTTTCATGTGGCAAAAGAAGTAGTGCGTACAATGTACTGATTCGTCCTGAGATTTGTACCATG
GTTGAAACTAATTTATGGTAATAATATTAACATAGCAAATCTTTAGAGACTCAAATCATGAAAAGGTAATAGCAGTA
CTGTACTAAAAACGGTAGTGCTAATTTTCGTAATAATTTTGTAAATATTCAACAGTAAAACAACTTGAAGACACACT
TTCCTAGGGAGGCGTTACTGAAATAATTTAGCTATAGTAAGAAAATTTGTAATTTTAGAAATGCCAAGCATTCTAAA
TTAATTGCTTGAAAGTCACTATGATTGTGTCCATTATAAGGAGACAAATTCATTCAAGCAAGTTATTTAATGTTAAA
GGCCCAATTGTTAGGCAGTTAATGGCACTTTTACTATTAACTAATCTTTCCATTTGTTCAGACGTAGCTTAACTTAC
CTCTTAGGTGTGAATTTGGTTAAGGTCCTCATAATGTCTTTATGTGCAGTTTTTGATAGGTTATTGTCATAGAACTT
ATTCTATTCCTACATTTATGATTACTATGGATGTATGAGAATAACACCTAATCCTTATACTTTACCTCAATTTAACT
CCTTTATAAAGAACTTACATTACAGAATAAAGATTTTTTAAAAATATATTTTTTTGTAGAGACAGGGTCTTAGCCCA
GCCGAGGCTGGTCTCTAAGTCCTGGCCCAAGCGATCCTCCTGCCTGGGCCTCCTAAAGTGCTGGAATTATAGACATG
AGCCATCACATCCAATATACAGAATAAAGATTTTTAATGGAGGATTTAATGTTCTTCAGAAAATTTTCTTGAGGTCA
GACAATGTCAAATGTCTCCTCAGTTTACACTGAGATTTTGAAAACAAGTCTGAGCTATAGGTCCTTGTGAAGGGTCC
ATTGGAAATACTTGTTCAAAGTAAAATGGAAAGCAAAGGTAAAATCAGCAGTTGAAATTCAGAGAAAGACAGAAAAG
GAGAAAAGATGAAATTCAACAGGACAGAAGGGAAATATATTATCATTAAGGAGGACAGTATCTGTAGAGCTCATTAG
TGATGGCAAAATGACTTGGTCAGGATTATTTTTAACCCGCTTGTTTCTGGTTTGCACGGCTGGGGATGCAGCTAGGG
TTCTGCCTCAGGGAGCACAGCTGTCCAGAGCAGCTGTCAGCCTGCAAGCCTGAAACACTCCCTCGGTAAAGTCCTTC
CTACTCAGGACAGAAATGACGAGAACAGGGAGCTGGAAACAGGCCCCTAACCAGAGAAGGGAAGTAATGGATCAACA
AAGTTAACTAGCAGGTCAGGATCACGCAATTCATTTCACTCTGACTGGTAACATGTGACAGAAACAGTGTAGGCTTA
TTGTATTTTCATGTAGAGTAGGACCCAAAAATCCACCCAAAGTCCTTTATCTATGCCACATCCTTCTTATCTATACT
TCCAGGACACTTTTTCTTCCTTATGATAAGGCTCTCTCTCTCTCCACACACACACACACACACACACACACACACAC
ACACACACACACAAACACACACCCCGCCAACCAAGGTGCATGTAAAAAGATGTAGATTCCTCTGCCTTTCTCATCTA
CACAGCCCAGGAGGGTAAGTTAATATAAGAGGGATTTATTGGTAAGAGATGATGCTTAATCTGTTTAACACTGGGCC
TCAAAGAGAGAATTTCTTTTCTTCTGTACTTATTAAGCACCTATTATGTGTTGAGCTTATATATACAAAGGGTTATT
ATATGCTAATATAGTAATAGTAATGGTGGTTGGTACTATGGTAATTACCATAAAAATTATTATCCTTTTAAAATAAA
GCTAATTATTATTGGATCTTTTTTAGTATTCATTTTATGTTTTTTATGTTTTTGATTTTTTAAAAGACAATCTCACC

CTGTTACCCAGGCTGGAGTGCAGTGGTGCAATCATAGCTTTCTGCAGTCTTGAACTCCTGGGCTCAAGCAATCCTCC
TGCCTTGGCCTCCCAAAGTGTTGGGATACAGTCATGAGCCACTGCATCTGGCCTAGGATCCATTTAGATTAAAATAT
GCATTTTAAATTTTAAAATAATATGGCTAATTTTTACCTTATGTAATGTGTATACTGGCAATAAATCTAGTTTGCTG
CCTAAAGTTTAAAGTGCTTTCCAGTAAGCTTCATGTACGTGAGGGGAGACATTTAAAGTGAAACAGACAGCCAGGTG
TGGTGGCTCACGCCTGTAATCCCAGCACTCTGGGAGGCTGAGGTGGGTGGATCGCTTGAGCCCTGGAGTTCAAGACC
AGCCTGAGCAACATGGCAAAACGCTGTTTCTATAACAAAAATTAGCCGGGCATGGTGGCATGTGCCTGTGGTCCCAG
CTACTAGGGGGCTGAGGCAGGAGAATCGTTGGAGCCCAGGAGGTCAAGGCTGCACTGAGCAGTGCTTGCGCCACTGC
ACTCCAGCCTGGGTGACAGGACCAGACCTTGCCTCAAAAAAATAAGAAGAAAAATTAAAAATAAATGGAAACAACTA
CAAAGAGCTGTTGTCCTAGATGAGCTACTTAGTTAGGCTGATATTTTGGTATTTAACTTTTAAAGTCAGGGTCTGTC
ACCTGCACTACATTATTAAAATATCAATTCTCAATGTATATCCACACAAAGACTGGTACGTGAATGTTCATAGTACC
TTTATTCACAAAACCCCAAAGTAGAGACTATCCAAATATCCATCAACAAGTGAACAAATAAACAAAATGTGCTATAT
CCATGCAATGGAATACCACCCTGCAGTACAAAGAAGCTACTTGGGGATGAATCCCAAAGTCATGACGCTAAATGAAA
GAGTCAGACATGAAGGAGGAGATAATGTATGCCATACGAAATTCTAGAAAATGAAAGTAACTTATAGTTACAGAAAG
CAAATCAGGGCAGGCATAGAGGCTCACACCTGTAATCCCAGCACTTTGAGAGGCCACGTGGGAAGATTGCTAGAACT
CAGGAGTTCAAGACCAGCCTGGGCAACACAGTGAAACTCCATTCTCCACAAAAATGGGAAAAAAAGAAAGCAAATCA
GTGGTTGTCCTGTGGGGAGGGGAAGGACTGCAAAGAGGGAAGAAGCTCTGGTGGGGTGAGGGTGGTGATTCAGGTTC
TGTATCCTGACTGTGGTAGCAGTTTGGGGTGTTTACATCCAAAAATATTCGTAGAATTATGCATCTTAAATGGGTGG
AGTTTACTGTATGTAAATTATACCTCAATGTAAGAAAAAATAATGTGTAAGAAAACTTTCAATTCTCTTGCCAGCAA
ACGTTATTCAAATTCCTGAGCCCTTTACTTCGCAAATTCTCTGCACTTCTGCCCCGTACCATTAGGTGACAGCACTA
GCTCCACAAATTGGATAAATGCATTTCTGGAAAAGACTAGGGACAAAATCCAGGCATCACTTGTGCTTTCATATCAA
CCATGCTGTACAGCTTGTGTTGCTGTCTGCAGCTGCAATGGGGACTCTTGATTTCTTTAAGGAAACTTGGGTTACCA
GAGTATTTCCACAAATGCTATTCAAATTAGTGCTTATGATATGCAAGACACTGTGCTAGGAGCCAGAAAACAAAGAG
GAGGAGAAATCAGTCATTATGTGGGAACAACATAGCAAGATATTTAGATCATTTTGACTAGTTAAAAAAGCAGCAGA
GTACAAAATCACACATGCAATCAGTATAATCCAAATCATGTAAATATGTGCCTGTAGAAAGACTAGAGGAATAAACA
CAAGAATCTTAACAGTCATTGTCATTAGACACTAAGTCTAATTATTATTATTAGACACTATGATATTTGAGATTTAA
AAAATCTTTAATATTTTAAAATTTAGAGCTCTTCTATTTTTCCATAGTATTCAAGTTTGACAATGATCAAGTATTAC
TCTTTCTTTTTTTTTTTTTTTTTTTTTTTTTGAGATGGAGTTTTGGTCTTGTTGCCCATGCTGGAGTGGAATGGCAT
GACCATAGCTCACTGCAACCTCCACCTCCTGGGTTCAAGCAAAGCTGTCGCCTCAGCCTCCCGGGTAGATGGGATTA
CAGGCGCCCACCACCACACTCGGCTAATGTTTGTATTTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGT
CTCAAACTCCTGACCTCAGAGGATCCACCTGCCTCAGCCTCCCAAAGTGCTGGGATTACAGATGTAGGCCACTGCGC
CCGGCCAAGTATTGCTCTTATACATTAAAAAACAGGTGTGAGCCACTGCGCCCAGCCAGGTATTGCTCTTATACATT
AAAAAATAGGCCGGTGCAGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAAGCCAAGGCGGGCAGAACACCCGAGG
TCAGGAGTCCAAGGCCAGCCTGGCCAAGATGGTGAAACCCCGTCTCTATTAAAAATACAAACATTACCTGGGCATGA
TGGTGGGCGCCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGGATCCGCGGAGCCTGGCAGATCTGCCTGAGC
CTGGGAGGTTGAGGCTACAGTAAGCCAAGATCATGCCAGTATACTTCAGCCTGGGCGACAAAGTGAGACCGTAACAA
AAAAAAAAAAATTTAAAAAAAGAAATTTAGATCAAGATCCAACTGTAAAAAGTGGCCTAAACACCACATTAAAGAGT
TTGGAGTTTATTCTGCAGGCAGAAGAGAACCATCAGGGGGTCTTCAGCATGGGAATGGCATGGTGCACCTGGTTTTT
GTGAGATCATGGTGGTGACAGTGTGGGGAATGTTATTTTGGAGGGACTGGAGGCAGACAGACCGGTTAAAAGGCCAG
CACAACAGATAAGGAGGAAGAAGATGAGGGCTTGGACCGAAGCAGAGAAGAGCAAACAGGGAAGGTACAAATTCAAG

AAATATTGGGGGGTTTGAATCAACACATTTAGATGATTAATTAAATATGAGGACTGAGGAATAAGAAATGAGTCAAG
GATGGTTCCAGGCTGCTAGGCTGCTTACCTGAGGTGGCAAAGTCGGGAGGAGTGGCAGTTTAGGACAGGGGGCAGTT
GAGGAATATTGTTTTGATCATTTTGAGTTTGAGGTACAAGTTGGACACTTAGGTAAAGACTGGAGGGGAAATCTGAA
TATACAATTATGGGACTGAGGAACAAGTTTATTTTATTTTTTGTTTCGTTTTCTTGTTGAAGAACAAATTTAATTGT
AATCCCAAGTCATCAGCATCTAGAAGACAGTGGCAGGAGGTGACTGTCTTGTGGGTAAGGGTTTGGGGTCCTTGATG
AGTATCTCTCAATTGGCCTTAAATATAAGCAGGAAAAGGAGTTTATGATGGATTCCAGGCTCAGCAGGGCTCAGGAG
GGCTCAGGCAGCCAGCAGAGGAAGTCAGAGCATCTTCTTTGGTTTAGCCCAAGTAATGACTTCCTTAAAAAGCTGAA
GGAAAATCCAGAGTGACCAGATTATAAACTGTACTCTTGCATTTTCTCTCCCTCCTCTCACCCACAGCCTCTTGATG
AACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCCGCTGGGCTAAGGGTCGGCGTGAGACCTACCTGTGCTACGT
AGTGAAGAGGCGTGACAGTGCTACATCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGGTATCAATTAAAGTCG
GCTTTGCAAGCAGTTTAATGGTCAACTGTGAGTGCTTTTAGAGCCACCTGCTGATGGTATTACTTCCATCCTTTTTT
GGCATTTGTGTCTCTATCACATTCCTCAAATCCTTTTTTTTATTTCTTTTTCCATGTCCATGCACCCATATTAGACA
TGGCCCAAAATATGTGATTTAATTCCTCCCCAGTAATGCTGGGCACCCTAATACCACTCCTTCCTTCAGTGCCAAGA
ACAACTGCTCCCAAACTGTTTACCAGCTTTCCTCAGCATCTGAATTGCCTTTGAGATTAATTAAGCTAAAAGCATTT
TTATATGGGAGAATATTATCAGCTTGTCCAAGCAAAAATTTTAAATGTGAAAAACAAATTGTGTCTTAAGCATTTTT
GAAAATTAAGGAAGAAGAATTTGGGAAAAAATTAACGGTGGCTCAATTCTGTCTTCCAAATGATTTCTTTTCCCTCC
TACTCACATGGGTCGTAGGCCAGTGAATACATTCAACATGGTGATCCCCAGAAAACTCAGAGAAGCCTCGGCTGATG
ATTAATTAAATTGATCTTTCGGCTACCCGAGAGAATTACATTTCCAAGAGACTTCTTCACCAAAATCCAGATGGGTT
TACATAAACTTCTGCCCACGGGTATCTCCTCTCTCCTAACACGCTGTGACGTCTGGGCTTGGTGGAATCTCAGGGAA
GCATCCGTGGGGTGGAAGGTCATCGTCTGGCTCGTTGTTTGATGGTTATATTACCATGCAATTTTCTTTGCCTACAT
TTGTATTGAATACATCCCAATCTCCTTCCTATTCGGTGACATGACACATTCTATTTCAGAAGGCTTTGATTTTATCA
AGCACTTTCATTTACTTCTCATGGCAGTGCCTATTACTTCTCTTACAATACCCATCTGTCTGCTTTACCAAAATCTA
TTTCCCCTTTTCAGATCCTCCCAAATGGTCCTCATAAACTGTCCTGCCTCCACCTAGTGGTCCAGGTATATTTCCAC
AATGTTACATCAACAGGCACTTCTAGCCATTTTCCTTCTCAAAAGGTGCAAAAAGCAACTTCATAAACACAAATTAA
ATCTTCGGTGAGGTAGTGTGATGCTGCTTCCTCCCAACTCAGCGCACTTCGTCTTCCTCATTCCACAAAAACCCATA
GCCTTCCTTCACTCTGCAGGACTAGTGCTGCCAAGGGTTCAGCTCTACCTACTGGTGTGCTCTTTTGAGCAAGTTGC
TTAGCCTCTCTGTAACACAAGGACAATAGCTGCAAGCATCCCCAAAGATCATTGCAGGAGACAATGACTAAGGCTAC
CAGAGCCGCAATAAAAGTCAGTGAATTTTAGCGTGGTCCTCTCTGTCTCTCCAGAACGGCTGCCACGTGGAATTGCT
CTTCCTCCGCTACATCTCGGACTGGGACCTAGACCCTGGCCGCTGCTACCGCGTCACCTGGTTCACCTCCTGGAGCC
CCTGCTACGACTGTGCCCGACATGTGGCCGACTTTCTGCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCG
CGCCTCTACTTCTGTGAGGACCGCAAGGCTGAGCCCGAGGGGCTGCGGCGGCTGCACCGCGCCGGGGTGCAAATAGC
CATCATGACCTTCAAAGGTGCGAAAGGGCCTTCCGCGCAGGCGCAGTGCAGCAGCCCGCATTCGGGATTGCGATGCG
GAATGAATGAGTTAGTGGGGAAGCTCGAGGGGAAGAAGTGGGCGGGGATTCTGGTTCACCTCTGGAGCCGAAATTAA
AGATTAGAAGCAGAGAAAAGAGTGAATGGCTCAGAGACAAGGCCCCGAGGAAATGAGAAAATGGGGCCAGGGTTGCT
TCTTTCCCCTCGATTTGGAACCTGAACTGTCTTCTACCCCCATATCCCCGCCTTTTTTTCCTTTTTTTTTTTTTGAA
GATTATTTTTACTGCTGGAATACTTTTGTAGAAAACCACGAAAGAACTTTCAAAGCCTGGGAAGGGCTGCATGAAAA
TTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCCTTTTGGTAAGGGGCTTCCTCGCTTTTTAAATTTTCTTTCTTT
CTCTACAGTCTTTTTTGGAGTTTCGTATATTTCTTATATTTTCTTATTGTTCAATCACTCTCAGTTTTCATCTGATG
AAAACTTTATTTCTCCTCCACATCAGCTTTTTCTTCTGCTGTTTCACCATTCAGAGCCCTCTGCTAAGGTTCCTTTT

CCCTCCCTTTTCTTTCTTTTGTTGTTTCACATCTTTAAATTTCTGTCTCTCCCCAGGGTTGCGTTTCCTTCCTGGTC
AGAATTCTTTTCTCCTTTTTTTTTTTTTTTTTTTTTTTTTTTAAACAAACAAACAAAAAACCCAAAAAAACTCTTTC
CCAATTTACTTTCTTCCAACATGTTACAAAGCCATCCACTCAGTTTAGAAGACTCTCCGGCCCCACCGACCCCCAAC
CTCGTTTTGAAGCCATTCACTCAATTTGCTTCTCTCTTTCTCTACAGCCCCTGTATGAGGTTGATGACTTACGAGAC
GCATTTCGTACTTTGGGACTTTGATAGCAACTTCCAGGAATGTCACACACGATGAAATATCTCTGCTGAAGACAGTG
GATAAAAAACAGTCCTTCAAGTCTTCTCTGTTTTTATTCTTCAACTCTCACTTTCTTAGAGTTTACAGAAAAAATAT
TTATATACGACTCTTTAAAAAGATCTATGTCTTGAAAATAGAGAAGGAACACAGGTCTGGCCAGGGACGTGCTGCAA
TTGGTGCAGTTTTGAATGCAACATTGTCCCCTACTGGGAATAACAGAACTGCAGGACCTGGGAGCATCCTAAAGTGT
CAACGTTTTTCTATGACTTTTAGGTAGGATGAGAGCAGAAGGTAGATCCTAAAAAGCATGGTGAGAGGATCAAATGT
TTTTATATCAACATCCTTTATTATTTGATTCATTTGAGTTAACAGTGGTGTTAGTGATAGATTTTTCTATTCTTTTC
CCTTGACGTTTACTTTCAAGTAACACAAACTCTTCCATCAGGCCATGATCTATAGGACCTCCTAATGAGAGTATCTG
GGTGATTGTGACCCCAAACCATCTCTCCAAAGCATTAATATCCAATCATGCGCTGTATGTTTTAATCAGCAGAAGCA
TGTTTTTATGTTTGTACAAAAGAAGATTGTTATGGGTGGGGATGGAGGTATAGACCATGCATGGTCACCTTCAAGCT
ACTTTAATAAAGGATCTTAAAATGGGCAGGAGGACTGTGAACAAGACACCCTAATAATGGGTTGATGTCTGAAGTAG
CAAATCTTCTGGAAACGCAAACTCTTTTAAGGAAGTCCCTAATTTAGAAACACCCACAAACTTCACATATCATAATT
AGCAAACAATTGGAAGGAAGTTGCTTGAATGTTGGGGAGAGGAAAATCTATTGGCTCTCGTGGGTCTCTTCATCTCA
GAAATGCCAATCAGGTCAAGGTTTGCTACATTTTGTATGTGTGTGATGCTTCTCCCAAAGGTATATTAACTATATAA
GAGAGTTGTGACAAAACAGAATGATAAAGCTGCGAACCGTGGCACACGCTCATAGTTCTAGCTGCTTGGGAGGTTGA
GGAGGGAGGATGGCTTGAACACAGGTGTTCAAGGCCAGCCTGGGCAACATAACAAGATCCTGTCTCTCAAAAAAAAA
AAAAAAAAAAAGAAAGAGAGAGGGCCGGGCGTGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGCCGG
GCGGATCACCTGTGGTCAGGAGTTTGAGACCAGCCTGGCCAACATGGCAAAACCCCGTCTGTACTCAAAATGCAAAA
ATTAGCCAGGCGTGGTAGCAGGCACCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCAG
GAGGTGGAGGTTGCAGTAAGCTGAGATCGTGCCGTTGCACTCCAGCCTGGGCGACAAGAGCAAGACTCTGTCTCAGA
AAAAAAAAAAAAAAAGAGAGAGAGAGAGAAAGAGAACAATATTTGGGAGAGAAGGATGGGGAAGCATTGCAAGGAAA
TTGTGCTTTATCCAACAAAATGTAAGGAGCCAATAAGGGATCCCTATTTGTCTCTTTTGGTGTCTATTTGTCCCTAA
CAACTGTCTTTGACAGTGAGAAAAATATTCAGAATAACCATATCCCTGTGCCGTTATTACCTAGCAACCCTTGCAAT
GAAGATGAGCAGATCCACAGGAAAACTTGAATGCACAACTGTCTTATTTTAATCTTATTGTACATAAGTTTGTAAAA
GAGTTAAAAATTGTTACTTCATGTATTCATTTATATTTTATATTATTTTGCGTCTAATGATTTTTTATTAACATGAT
TTCCTTTTCTGATATATTGAAATGGAGTCTCAAAGCTTCATAAATTTATAACTTTAGAAATGATTCTAATAACAACG
TATGTAATTGTAACATTGCAGTAATGGTGCTACGAAGCCATTTCTCTTGATTTTTAGTAAACTTTTATGACAGCAAA
TTTGCTTCTGGCTCACTTTCAATCAGTTAAATAAATGATAAATAATTTTGGAAGCTGTGAAGATAAAATACCAAATA
AAATAATATAAAAGTGATTTATATGAAGTTAAAATAAAAAATCAGTATGATGGAATAAACTTG

[1378] 可根据本公开的方面融合至Cas9的其他示例性脱氨酶提供于下。在实施方案中，脱氨酶是激活诱导型脱氨酶(AID)。应理解，在一些实施方案中，可使用相应序列的活性结构域，例如灭有定位信号的结构域(核定位序列，没有核输出信号、细胞质定位信号)。

[1379] 人类AID：

[1380] MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL

[1381] (下划线:核定位序列;双下划线:核输出信号)

[1382] 小鼠AID:

[1383] MDSLLMKQKKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSCSLDFGHLRNKSGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVAEFLRWNPNLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIGIMTFKDYFYCWNTFVENRERTFKAWEGLHENSVRLTRQLRRILLPLYEVDDLRDAFRMLGF

[1384] (下划线:核定位序列;双下划线:核输出信号)

[1385] 犬AID:

[1386] MDSLLMKQRKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLDFGHLRNKSGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFAARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENREKTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL

[1387] (下划线:核定位序列;双下划线:核输出信号)

[1388] 牛AID:

[1389] MDSLLKKQRQFLYQFKNVRWAKGRHETYLCYVVKRRDSPTSFSLDFGHLRNKAGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFTARLYFCDKERKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL

[1390] (下划线:核定位序列;双下划线:核输出信号)

[1391] 大鼠AID:

[1392] MAVGSKPKAALVGPHWERERIWCFLCSTGLGTQQTGQTSRWLRPAATQDPVSPPRSLLMKQRKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLDFGYLRNKSGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLTGWGALPAGLMSPARPSDYFYCWNTFVENHERTFKAWEGLHENSVRLSRRLRRILLPLYEVDDLRDAFRTLGL

[1393] (下划线:核定位序列;双下划线:核输出信号)

[1394] clAID(家犬亚种(Canis lupus familiaris)):

[1395] MDSLLMKQRKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLDFGHLRNKSGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFAARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENREKTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL

[1396] btAID(牛(Bos Taurus)):

[1397] MDSLLKKQRQFLYQFKNVRWAKGRHETYLCYVVKRRDSPTSFSLDFGHLRNKAGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFTARLYFCDKERKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL

[1398] mAID(小家鼠(Mus musculus)):

[1399] MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELLFLRYISDWD LDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFY CWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL

[1400] rAPOBEC-1(褐家鼠(Rattus norvegicus)):

[1401] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEK FTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTE QESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHIL WATGLK

[1402] maAPOBEC-1(金仓鼠(Mesocricetus auratus)):

[1403] MSSETGPVVVDPTLRRRIEPHEFDAFFDQGELRKETCLLYEIRWGGRHNIWRHTGQNTSRHVEINFIEK FTSERYFYPSTRCSIVWFLSWSPCGECSKAITEFLSGHPNVTLFIYAARLYHHTDQRNRQGLRDLISRGVTIRIMTE QEYCYCWRNFVNYPPSNEVYWPRYPNLWMRLYALELYCIHLGLPPCLKIKRRHQYPLTFFRLNLQSCHYQRIPPHIL WATGFI

[1404] ppAPOBEC-1(猩猩(Pongo pygmaeus)):

[1405] MTSEKGPSTGDPTLRRRIESWEFDVFYDPRELRKETCLLYEIKWGMSRKIWRSSGKNTTNHVEVNFIKK FTSERRFHSSISCSITWFLSWSPCWECSQAIREFLSQHPGVTLVIYVARLFWHMDQRNRQGLRDLVNSGVTIQIMRA SEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKISRRWQNHLAFFRLHLQNCHYQTIPPHIL LATGLIHPSVTWR

[1406] ocAPOBEC1(穴兔(Oryctolagus cuniculus)):

[1407] MASEKGPSNKDYTLRRRIEPWEFEVFFDPQELRKEACLLYEIKWGASSKTWRSSGKNTTNHVEV

[1408] NFLEKLTSEGRLGPSTCCSITWFLSWSPCWECSMAIREFLSQHPGVTLIIFVARLFQHMDRRNRQGLKD LVTSGVTVRVMSVSEYCYCWENFVNYPPGKAAQWPRYPPRWMLMYALELYCIILGLPPCLKISRRHQKQLTFFSLTP QYCHYKMIPPYILLATGLLQPSVPWR

[1409] mdAPOBEC-1(短尾负鼠(Monodelphis domestica):

[1410] MNSKTGPSVGDATLRRRIKPWEFVAFFNPQELRKETCLLYEIKWGNQNIWRHSNQNTSQHAEINFMEKF TAERHFNSSVRCSITWFLSWSPCWECSKAIRKFLDHYPNVTLAIFISRLYWHMDQQHRQGLKELVHSGVTIQIMSYS EYHYCWRNFVDYPQGEEDYWPKYPYLWIMLYVLELHCIILGLPPCLKISGSHSNQLALFSLDLQDCHYQKIPYNVLV ATGLVQPFVTWR

[1411] ppAPOBEC-2(猩猩(Pongo pygmaeus):

[1412] MAQKEEAAAATEAASQNGEDLENLDDPEKLKELIELPPFEIVTGERLPANFFKFQFRNVEYSSGRNKTF LCYVVEAQGKGGQVQASRGYLEDEHAAAHAEEAFFNTILPAFDPALRYNVTWYVSSSPCAACADRIIKTLSKTKNLR LLILVGRLFMWEELEIQDALKKLKEAGCKLRIMKPQDFEYVWQNFVEQEEGESKAFQPWEDIQENFLYYEEKLADIL K

[1413] btAPOBEC-2(牛(Bos Taurus):

[1414] MAQKEEAAAAAEPASQNGEEVENLEDPEKLKELIELPPFEIVTGERLPAHYFKFQFRNVEYSSGRNKTF LCYVVEAQSKGGQVQASRGYLEDEHATNHAEEAFFNSIMPTFDPALRYMVTWYVSSSPCAACADRIVKTLNKTKNLR LLILVGRLFMWEEPEIQAALRKLKEAGCRLRIMKPQDFEYIWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADIL K

[1415] mAPOBEC-3-(1)(小家鼠(Mus musculus):

[1416] MQPQRLGPRAGMGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLHHGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQIVRFLATHHNLSLDIFSSRLYNVQDPETQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWKRLLTNFRYQDSKLQEILRPCYISVPSSSSSTLSNICLTKGLPETRFWVEGRRMDPLSEEEFYSQFYNQRVKHLCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKGKQHAEILFLDKIRSMELSQVTITCYLTWSPCPNCAWQLAAFKRDRPDLILHIYTSRLYFHWKRPFQKGLCSLWQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLRRIKESWGLQDLVNDFGNLQLGPPMS

[1417] 小鼠APOBEC-3-(2):

[1418] MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLHHGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQIVRFLATHHNLSLDIFSSRLYNVQDPETQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWKRLLTNFRYQDSKLQEILRPCYIPVPSSSSSTLSNICLTKGLPETRFCVEGRRMDPLSEEEFYSQFYNQRVKHLCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKGKQHAEILFLDKIRSMELSQVTITCYLTWSPCPNCAWQLAAFKRDRPDLILHIYTSRLYFHWKRPFQKGLCSLWQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLRRIKESWGLQDLVNDFGNLQLGPPMS

[1419] (斜体:核酸编辑结构域)

[1420] 大鼠APOBEC-3:

[1421] MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNRLRYAIDRKDTFLCYEVTRKDCDSPVSLHHGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQVLRFLATHHNLSLDIFSSRLYNIRDPENQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWKKLLTNFRYQDSKLQEILRPCYIPVPSSSSSTLSNICLTKGLPETRFCVERRRVHLLSEEEFYSQFYNQRVKHLCYYHGVKPYLCYQLEQFNGQAPLKGCLLSEKGKQHAEILFLDKIRSMELSQVIITCYLTWSPCPNCAWQLAAFKRDRPDLILHIYTSRLYFHWKRPFQKGLCSLWQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLHRIKESWGLQDLVNDFGNLQLGPPMS

[1422] (斜体:核酸编辑结构域)

[1423] hAPOBEC-3A(智人(Homo sapiens)):

[1424] MEASPASGPRHLMDPHIFTSNFNNGIGRHKTYLCYEVERLDNGTSVKMDQHRGFLHNQAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGEVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFKHCWDTFVDHQGCPFQPWDGLDEHSQALSGRLRAILQNQGN(

[1425] hAPOBEC-3F(智人(Homo sapiens)):

[1426] MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPRLDAKIFRGQVYSQPEHHAEMCFLSWFCGNQLPAYKCFQITWFVSWTPCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERDYRRALCRLSQAGARVKIMDDEEFAYCWENFVYSEGQPFMPWYKFDDNYAFLHRTLKEILRNPMEAMYPHIFYFHFKNLRKAYGRNESWLCFTMEVVKHHSPVSWKRGVFRNQVDPETHCHAERCFLSWFCDDILSPNTNYEVTWYTSWSPCPECAGEVAEFLARHSNVNLTIFTARLYYFWDTDYQEGLRSLSQEGASVEIMGYKDFKYCWENFVYNDDEPFKPWKGLKYNFLFLDSKLQEILE

[1427] 恒河猴APOBEC-3G:

[1428] MVEPMDPRTFVSNFNNRPILSGLNTVWLCCEVKTKDPSGPPLDAKIFQGKVYSKAKYHPEMRFLRWFHKWRQLHHDQEYKVTWYVSWSPCTRCANSVATFLAKDPKVTLTIFVARLYYFWKPDYQQALRILCQKRGGPHATMKIMNYNEFQDCWNKFVDGRGKPFKPRNNLPKHYTLLQATLGELLRHLMDPGTFTSNFNNKPWVSGQHETYLCYKVERLHNDTWVPLNQHRGFLRNQAPNIHGFPKGRHAELCFLDLIPFWKLDGQQYRVTCFTSWSPCFSCAQEMAKFISNNEHVSLCIFAARIYDDQGRYQEGLRALHRDGAKIAMMNYSEFEYCWDTFVDRQGRPFQPWDGLDEHSQALSGRLRAI

[1429] (斜体:核酸编辑结构域;下划线:细胞质定位信号)

[1430] 黑猩猩APOBEC-3G：

[1431] MKPHFRNPVERMYQDTFSDNFYNRPILSHRNTVWLCYEVKTKGPSRPPLDAKIFRGQVYSKLKYHPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDVATFLAEDPKVTLTIFVARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEPWNNLPKYYILLHIMLGEILRHSMDPPTFTSNFNNELWVRGRHETYLCYEVERLHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLHQDYRVTCFTSWSPCFSCAQEMAKFISNNKHVSLCIFAARIYDDQGRCQEGLRTLAKAGAKISIMTYSEFKHCWDTFVDHQGCPFQPWDGLEEHSQALSGRLRAILQNQGN

[1432] (斜体:核酸编辑结构域;下划线:细胞质定位信号)

[1433] 绿猴APOBEC-3G：

[1434] MNPQIRNMVEQMEPDIFVYYFNNRPILSGRNTVWLCYEVKTKDPSGPPLDANIFQGKLYPEAKDHPEMKFLHWFRKWRQLHRDQEYEVTWYVSWSPCTRCANSVATFLAEDPKVTLTIFVARLYYFWKPDYQQALRILCQERGGPHATMKIMNYNEFQHCWNEFVDGQGKPFKPRKNLPKHYTLLHATLGELLRHVMDPGTFTSNFNNKPWVSGQRETYLCYKVERSHNDTWVLLNQHRGFLRNQAPDRHGFPKGRHAELCFLDLIPFWKLDDQQYRVTCFTSWSPCFSCAQKMAKFISNNKHVSLCIFAARIYDDQGRCQEGLRTLHRDGAKIAVMNYSEFEYCWDTFVDRQGRPFQPWDGLDEHSQALSGRLRAI

[1435] (斜体:核酸编辑结构域;下划线:细胞质定位信号)

[1436] 人类APOBEC-3G：

[1437] MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQVYSELKYHPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPKVTLTIFVARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEPWNNLPKYYILLHIMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQNQEN

[1438] (斜体:核酸编辑结构域;下划线:细胞质定位信号)

[1439] 人类APOBEC-3F：

[1440] MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPRLDAKIFRGQVYSQPEHHAEMCFLSWFCGNQLPAYKCFQITWFVSWTPCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERDYRRALCRLSQAGARVKIMDDEEFAYCWENFVYSEGQPFMPWYKFDDNYAFLHRTLKEILRNPMEAMYPHIFYFHFKNLRKAYGRNESWLCFTMEVVKHHSPVSWKRGVFRNQVDPETHCHAERCFLSWFCDDILSPNTNYEVTWYTSWSPCPECAGEVAEFLARHSNVNLTIFTARLYYFWDTDYQEGLRSLSQEGASVEIMGYKDFKYCWENFVYNDDEPFKPWKGLKYNFLFLDSKLQEILE

[1441] (斜体:核酸编辑结构域)

[1442] 人类APOBEC-3B：

[1443] MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGQVYFKPQYHAEMCFLSWFCGNQLPAYKCFQITWFVSWTPCPDCVAKLAEFLSEHPNVTLTISAARLYYYWERDYRRALCRLSQAGARVTIMDYEEFAYCWENFVYNEGQQFMPWYKFDENYAFLHRTLKEILRYLMDPDTFTFNFNNDPLVLRRRQTYLCYEVERLDNGTWVLMDQHMGFLCNEAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGEVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVYRQGCPFQPWDGLEEHSQALSGRLRAILQNQGN

[1444] (斜体:核酸编辑结构域)

[1445] 大鼠APOBEC-3B：

[1446] MQPQGLGPNAGMGPVCLGCSHRRPYSPIRNPLKKLYQQTFYFHFKNVRYAWGRKNNFLCYEVNGMDCAL PVPLRQGVFRKQGHIHAELCFIYWFHDKVLRVLSPMEEFKVTWYMSWSPCSKCAEQVARFLAAHRNLSLAIFSSRLY YYLRNPNYQQKLCRLIQEGVHVAAMDLPEFKKCWNKFVDNDGQPFRPWMRLRINFSFYDCKLQEIFSRMNLLREDVF YLQFNNSHRVKPVQNRYYRRKSYLCYQLERANGQEPLKGYLLYKKGEQHVEILFLEKMRSMELSQVRITCYLTWSPC PNCARQLAAFKKDHPDLILRIYTSRLYFWRKKFQKGLCTLWRSGIHVDVMDLPQFADCWTNFVNPQRPFRPWNELEK NSWRIQRRLRRIKESWGL

[1447] 牛APOBEC-3B:

[1448] DGWEVAFRSGTVLKAGVLGVSMTEGWAGSGHPGQGACVWTPGTRNTMNLLREVLFKQQFGNQPRVPAPY YRRKTYLCYQLKQRNDLTLDRGCFRNKKQRHAERFIDKINSLDLNPSQSYKIICYITWSPCPNCANELVNFITRNNH LKLEIFASRLYFHWIKSFKMGLQDLQNAGISVAVMTHTEFEDCWEQFVDNQSRPFQPWDKLEQYSASIRRRLQRILT API

[1449] 黑猩猩APOBEC-3B:

[1450] MNPQIRNPMEWMYQRTFYYNFENEPILYGRSYTWLCYEVKIRRGHSNLLWDTGVFRGQMYSQPEHHAEM CFLSWFCGNQLSAYKCFQITWFVSWTPCPDCVAKLAKFLAEHPNVTLTISAARLYYYWERDYRRALCRLSQAGARVK IMDDEEFAYCWENFVYNEGQPFMPWYKFDDNYAFLHRTLKEIIRHLMDPDTFTFNFNNDPLVLRRHQTYLCYEVERL DNGTWVLMDQHMGFLCNEAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGQVRAFLQEN THVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVYRQGCPFQPWDGLEEHSQALSGRLRAIL QVRASSLCMVPHRPPPPPQSPGPCLPLCSEPPLGSLLPTGRPAPSLPFLLTASFSFPPPASLPPLPSLSLSPGHLPV PSFHSLTSCSIQPPCSSRIRETEGWASVSKEGRDLG

[1451] 人类APOBEC-3C:

[1452] MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFRNQVDSETHCHAER CFLSWFCDDILSPNTKYQVTWYTSWSPCPDCAGEVAEFLARHSNVNLTIFTARLYYFQYPCYQEGLRSLSQEGVAVE IMDYEDFKYCWENFVYNDNEPFKPWKGLKTNFRLLKRRLRESLQ

[1453] (斜体:核酸编辑结构域)

[1454] 大猩猩APOBEC-3C

[1455] MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFRNQVDSETHCHAER CFLSWECDDILSPNTNYQVTWYTSWSPCPECAGEVAEFLARHSNVNLTIFTARLYYFQDTDYQEGLRSLSQEGVAVK IMDYKDFKYCWENFVYNDDEPFKPWKGLKYNFRFLKRRLQEILE

[1456] (斜体:核酸编辑结构域)

[1457] 人类APOBEC-3A:

[1458] MEASPASGPRHLMDPHIFTSNFNNGIGRHKTYLCYEVERLDNGTSVKMDQHRGFLHNQAKNLLCGFYGR HAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGEVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDA GAQVSIMTYDEFKHCWDTFVDHQGCPFQPWDGLDEHSQALSGRLRAILQNQGN

[1459] (斜体:核酸编辑结构域)

[1460] 恒河猴APOBEC-3A:

[1461] MDGSPASRPRHLMDPNTFTFNFNNDLSVRGRHQTYLCYEVERLDNGTWVPMDERRGFLCNKAKNVPCGD YGCHVELRFLCEVPSWQLDPAQTYRVTWFISWSPCFRRGCAGQVRVFLQENKHVRLRIFAARIYDYDPLYQEALRTL RDAGAQVSIMTYEEFKHCWDTFVDRQGRPFQPWDGLDEHSQALSGRLRAILQNQGN

[1462] (斜体:核酸编辑结构域)

[1463] 牛APOBEC-3A：

[1464] MDEYTFTENFNNQGWPSKTYLCYEMERLDGDATIPLDEYKGFVRNKGLDQPEKPCHAELYFLGKIHSWNLDRNQHYRLTCFISWSPCYDCAQKLTTFLKENHHISLHILASRIYTHNRFGCHQSGLCELQAAGARITIMTFEDFKHCWETFVDHKGKPFQPWEGLNVKSQALCTELQAILKTQQN

[1465] (斜体:核酸编辑结构域)

[1466] 人类APOBEC-3H：

[1467] MALLTAETFRLQFNNKRRLRRPYYPRKALLCYQLTPQNGSTPTRGYFENKKKCHAEICFINEIKSMGLDETQCYQVTCYLTWSPCSSCAWELVDFIKAHDHLNLGIFASRLYYHWCKPQQKGLRLLCGSQVPVEVMGFPKFADCWENFVDHEKPLSFNPYKMLEELDKNSRAIKRRLERIKIPGVRAQGRYMDILCDAEV(斜体:核酸编辑结构域)

[1468] 恒河猴APOBEC-3H：

[1469] MALLTAKTFSLQFNNKRRVNKPYYPRKALLCYQLTPQNGSTPTRGHLKNKKKDHAEIRFINKIKSMGLDETQCYQVTCYLTWSPCPSCAGELVDFIKAHRHLNLRIFASRLYYHWRPNYQEGLLLLCGSQVPVEVMGLPEFTDCWENFVDHKEPPSFNPSEKLEELDKNSQAIKRRLERIKSRSVDVLENGLRSLQLGPVTPSSSIRNSR

[1470] 人类APOBEC-3D：

[1471] MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGPVLPKRQSNHRQEVYFRFENHAEMCFLSWFCGNRLPANRRFQITWFVSWNPCLPCVVKVTKFLAEHPNVTLTISAARLYYYRDRDWRWVLLRLHKAGARVKIMDYEDFAYCWENFVCNEGQPFMPWYKFDDNYASLHRTLKEILRNPMEAMYPHIFYFHFKNLLKACGRNESWLCFTMEVTKHHSAVFRKRGVFRNQVDPETHCHAERCFLSWFCDDILSPNTNYEVTWYTSWSPCPECAGEVAEFLARHSNVNLTIFTARLCYFWDTDYQEGLCSLSQEGASVKIMGYKDFVSCWKNFVYSDDEPFKPWKGLQTNFRLLKRRLREILQ

[1472] (斜体:核酸编辑结构域)

[1473] 人类APOBEC-1：

[1474] MTSEKGPSTGDPTLRRRIEPWEFDVFYDPRELRKEACLLYEIKWGMSRKIWRSSGKNTTNHVEVNFIKKFTSERDFHPSMSCSITWFLSWSPCWECSQAIREFLSRHPGVTLVIYVARLFWHMDQQNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKISRRWQNHLTFFRLHLQNCHYQTIPPHILLATGLIHPSVAWR

[1475] 小鼠APOBEC-1：

[1476] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSVWRHTSQNTSNHVEVNFLEKFTTERYFRPNTRCSITWFLSWSPCGECSRAITEFLSRHPYVTLFIYIARLYHHTDQRNRQGLRDLISSGVTIQIMTEQEYCYCWRNFVNYPPSNEAYWPRYPHLWVKLYVLELYCIILGLPPCLKILRRKQPQLTFFTITLQTCHYQRIPPHLLWATGLK

[1477] 大鼠APOBEC-1：

[1478] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK

[1479] 人类APOBEC-2：

[1480] MAQKEEAAVATEAASQNGEDLENLDDPEKLKELIELPPFEIVTGERLPANFFKFQFRNVEYSSGRNKTFLCYVVEAQGKGGQVQASRGYLEDEHAAAHAEEAFFNTILPAFDPALRYNVTWYVSSSPCAACADRIIKTLSKTKNLR

LLILVGRLFMWEEPEIQAALKKLKEAGCKLRIMKPQDFEYVWQNFVEQEEGESKAFQPWEDIQENFLYYEEKLADIL
K

[1481] 小鼠APOBEC-2:

[1482] MAQKEEAAEAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFRNVEYSSGRNKTF
LCYVVEVQSKGGQAQATQGYLEDEHAGAHAEEAFFNTILPAFDPALKYNVTWYVSSSPCAACADRILKTLSKTKNLR
LLILVSRLFMWEEPEVQAALKKLKEAGCKLRIMKPQDFEYIWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADIL
K

[1483] 大鼠APOBEC-2:

[1484] MAQKEEAAEAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFRNVEYSSGRNKTF
LCYVVEAQSKGGQVQATQGYLEDEHAGAHAEEAFFNTILPAFDPALKYNVTWYVSSSPCAACADRILKTLSKTKNLR
LLILVSRLFMWEEPEVQAALKKLKEAGCKLRIMKPQDFEYLWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADIL
K

[1485] 牛APOBEC-2:

[1486] MAQKEEAAAAAEPASQNGEEVENLEDPEKLKELIELPPFEIVTGERLPAHYFKFQFRNVEYSSGRNKTF
LCYVVEAQSKGGQVQASRGYLEDEHATNHAEEAFFNSIMPTFDPALRYMVTWYVSSSPCAACADRIVKTLNKTKNLR
LLILVGRLFMWEEPEIQAALRKLKEAGCRLRIMKPQDFEYIWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADIL
K

[1487] 海七鳃鳗(Petromyzon marinus)CDA1(pmCDA1):

[1488] MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQSGTERGIHAEI
FSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLKIWACKLYYEKNARNQIGLWNLRDNGV
GLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKRRSELSFMIQVKILHTTKSPAV

[1489] 人类APOBEC3G D316R D317R:

[1490] MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQVYSELKYHPEMR
FFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPKVTLTIFVARLYYFWDPDYQEALRSLCQKRDGPR
ATMKFNYDEFQHCWSKFVYSQRELFEPWNNLPKYYILLHFMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYLCYEV
ERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISK
KHVSLCIFTARIYRRQGRCQEGLRTLAEAGAKISFTYSEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQ
NQEN

[1491] 人类APOBEC3G A链:

[1492] MDPPTFTFNFNNEPWWGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVI
PFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISFTYSEFK
HCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQ

[1493] 人类APOBEC3G A链D120R D121R:

[1494] MDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDV
IPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIYRRQGRCQEGLRTLAEAGAKISFMTYSE
FKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQ

[1495] hAPOBEC-4(智人(Homo sapiens)):

[1496] MEPIYEEYLANHGTIVKPYYWLSFSLDCSNCPYHIRTGEEARVSLTEFCQIFGFPYGTTFPQTKHLTFY
ELKTSSGSLVQKGHASSCTGNYIHPESMLFEMNGYLDSAIYNNDSIRHIILYSNNSPCNEANHCCISKMYNFLITYP

GITLSIYFSQLYHTEMDFPASAWNREALRSLASLWPRVVLSPISGGIWHSVLHSFISGVSGSHVFQPILTGRALADR
HNAYEINAITGVKPYFTDVLLQTKRNPNTKAQEALESYPLNNAFPGQFFQMPSGQLQPNLPPDLRAPVVFVLVPLRD
LPPMHMGQNPNKPRNIVRHLNMPQMSFQETKDLGRLPTGRSVEIVEITEQFASSKEADEKKKKKGKK

[1497] mAPOBEC-4(小家鼠(Mus musculus)):

[1498] MDSLLMKQKKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSCSLDFGHLRNKSGCHVELLFLRYISDWD
LDPGRCYRVTWFTSWSPCYDCARHVAEFLRWNPNLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIGIMTFKDYFY
CWNTFVENRERTFKAWEGLHENSVRLTRQLRRILLPLYEVDDLRDAFRMLGF

[1499] rAPOBEC-4(褐家鼠(Rattus norvegicus)):

[1500] MEPLYEEYLTHSGTIVKPYYWLSVSLNCTNCPYHIRTGEEARVPYTEFHQTFGFPWSTYPQTKHLTFYE
LRSSSGNLIQKGLASNCTGSHTHPESMLFERDGYLDSLIFHDSNIRHIILYSNNSPCDEANHCCISKMYNFLMNYPE
VTLSVFFSQLYHTENQFPTSAWNREALRGLASLWPQVTLSAISGGIWQSILETFVSGISEGLTAVRPFTAGRTLTDR
YNAYEINCITEVKPYFTDALHSWQKENQDQKVWAASENQPLHNTTPAQWQPDMSQDCRTPAVFMLVPYRDLPPIHVN
PSPQKPRTVVRHLNTLQLSASKVKALRKSPSGRPVKKEEARKGSTRSQEANETNKSKWKKQTLFIKSNICHLLEREQ
KKIGILSSWSV

[1501] mfAPOBEC-4(食蟹猕猴(Macaca fascicularis)):

[1502] MEPTYEEYLANHGTIVKPYYWLSFSLDCSNCPYHIRTGEEARVSLTEFCQIFGFPYGTTYPQTKHLTFY
ELKTSSGSLVQKGHASSCTGNYIHPESMLFEMNGYLDSAIYNNDSIRHIILYCNNSPCNEANHCCISKVYNFLITYP
GITLSIYFSQLYHTEMDFPASAWNREALRSLASLWPRVVLSPISGGIWHSVLHSFVSGVSGSHVFQPILTGRALTDR
YNAYEINAITGVKPFFTDVLLHTKRNPNTKAQMALESYPLNNAFPGQSFQMTSGIPPDLRAPVVFVLLPLRDLPPMH
MGQDPNKPRNIIRHLNMPQMSFQETKDLERLPTRRSVETVEITERFASSKQAEEKTKKKKGKK

[1503] pmCDA-1(海七鳃鳗(Petromyzon marinus)):

[1504] MAGYECVRVSEKLDFDTFEFQFENLHYATERHRTYVIFDVKPQSAGGRSRRLWGYIINNPNVCHAELIL
MSMIDRHLESNPGVYAMTWYMSWSPCANCSSKLNPWLKNLLEEQGHTLTMHFSRIYDRDREGDHRGLRGLKHVSNSF
RMGVVGRAEVKECLAEYVEASRRTLTWLDTTESMAAKMRRKLFCILVRCAGMRESGIPLHLFTLQTPLLSGRVVWWR
V

[1505] pmCDA-2(海七鳃鳗(Petromyzon marinus)):

[1506] MELREVVDCALASCVRHEPLSRVAFLRCFAAPSQKPRGTVILFYVEGAGRGVTGGHAVNYNKQGTSIHA
EVLLLSAVRAALLRRRRCEDGEEATRGCTLHCYSTYSPCRDCVEYIQEFGASTGVRVVIHCCRLYELDVNRRRSEAE
GVLRSLSRLGRDFRLMGPRDAIALLLGGRLANTADGESGASGNAWVTETNVVEPLVDMTGFGDEDLHAQVQRNKQIR
EAYANYASAVSLMLGELHVDPDKFPFLAEFLAQTSVEPSGTPRETRGRPRGASSRGPEIGRQRPADFERALGAYGLF
LHPRIVSREADREEIKRDLIVVMRKHNYQGP

[1507] pmCDA-5(海七鳃鳗(Petromyzon marinus)):

[1508] MAGDENVRVSEKLDFDTFEFQFENLHYATERHRTYVIFDVKPQSAGGRSRRLWGYIINNPNVCHAELIL
MSMIDRHLESNPGVYAMTWYMSWSPCANCSSKLNPWLKNLLEEQGHTLMMHFSRIYDRDREGDHRGLRGLKHVSNSF
RMGVVGRAEVKECLAEYVEASRRTLTWLDTTESMAAKMRRKLFCILVRCAGMRESGMPLHLFT

[1509] yCD(酿酒酵母(Saccharomyces cerevisiae)):

[1510] MVTGGMASKWDQKGMDIAYEEAALGYKEGGVPIGGCLINNKDGSVLGRGHNMRFQKGSATLHGEISTLE
NCGRLEGKVYKDTTLYTTLSPCDMCTGAIIMYGIPRCVVGENVNFKSKGEKYLQTRGHEVVVVDDERCKKIMKQFID
ERPQDWFEDIGE

[1511] rAPOBEC-1(δ177-186):

[1512] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEK FTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTE QESGYCWRNFVNYSPSNEAHWPRYPHLWVRGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK

[1513] rAPOBEC-1(δ202-213):

[1514] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEK FTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTE QESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQHYQRLPPHILWATGLK

[1515] 小鼠APOBEC-3:

[1516] MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLHHGVFKNKDN IHAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQIVRFLATHHNLSLDIFSSRLYNVQDPETQQNLCRL VQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWKRLLTNFRYQDSKLQEILRPCYIPVPSSSSSTLSNICLTKGLPE TRFCVEGRRMDPLSEEEFYSQFYNQRVKHLCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKGKQHAEILFLDKIRSM ELSQVTITCYLTWSPCPNCAWQLAAFKRDRPDLILHIYTSRLYFHWKRPFQKGLCSLWQSGILVDVMDLPQFTDCWT NFVNPKRPFWPWKGLEIISRRTQRRLRRIKESWGLQDLVNDFGNLQLGPPMS

[1517] (斜体:核酸编辑结构域)

[1518] 本公开的一些方面基于以下认知:例如通过在脱氨酶结构域中作出点突变来调节本文所述的任何融合蛋白的脱氨酶结构域催化活性,将会影响该融合蛋白(例如,碱基编辑器)的持续合成能力。例如,降低但不消除碱基编辑融合蛋白内的脱氨酶结构域的催化活性的突变可以使得该脱氨酶结构域不太可能催化与靶标残基相邻的残基的脱氨基反应,从而缩窄脱氨基窗口。缩窄脱氨基窗口的能力可以防止不希望的与特异性靶标残基相邻的残基的脱氨基反应,这可以降低或预防脱靶效应。

[1519] 例如,在一些实施方案中,并入碱基编辑器内的APOBEC脱氨酶可包含选自由以下所组成的群组的一个或多个突变:rAPOBEC1的H121X、H122X、R126X、R126X、R118X、W90X、W90X和R132X,或另一APOBEC脱氨酶中的一个或多个对应突变,其中X是任何氨基酸。在一些实施方案中,并入碱基编辑器内的APOBEC脱氨酶可包含选自由以下所组成的群组的一个或多个突变:rAPOBEC1的H121R、H122R、R126A、R126E、R118A、W90A、W90Y和R132E,或另一APOBEC脱氨酶中的一个或多个对应突变。

[1520] 在一些实施方案中,并入碱基编辑器内的APOBEC脱氨酶可包含选自由以下所组成的群组的一个或多个突变:hAPOBEC3G的D316X、D317X、R320X、R320X、R313X、W285X、W285X、R326X,或另一APOBEC脱氨酶中的一个或多个对应突变,其中X是任何氨基酸。在一些实施方案中,本文提供的任何融合蛋白包含APOBEC脱氨酶,该脱氨酶包含选自由以下所组成的群组的一个或多个突变:hAPOBEC3G的D316R、D317R、R320A、R320E、R313A、W285A、W285Y、R326E,或另一APOBEC脱氨酶的一个或多个对应突变。

[1521] 在一些实施方案中,并入碱基编辑器的APOBEC脱氨酶可包含rAPOBEC1的H121R和H122R突变,或另一APOBEC脱氨酶中的一个或多个对应突变。在一些实施方案中,并入碱基编辑器的APOBEC脱氨酶可包含APOBEC脱氨酶,该脱氨酶包含rAPOBEC1的R126A突变,或另一APOBEC脱氨酶中的一个或多个对应突变。在一些实施方案中,并入碱基编辑器的APOBEC脱氨酶可包含APOBEC脱氨酶,该脱氨酶包含rAPOBEC1的R126E突变,或另一APOBEC脱氨酶中的

一个或多个对应突变。在一些实施方案中,并入碱基编辑器的APOBEC脱氨酶可包含APOBEC脱氨酶,该脱氨酶包含rAPOBEC1的R118A突变,或另一APOBEC脱氨酶中的一个或多个对应突变。在一些实施方案中,并入碱基编辑器的APOBEC脱氨酶可包含APOBEC脱氨酶,该脱氨酶包含rAPOBEC1的W90A突变,或另一APOBEC脱氨酶中的一个或多个对应突变。在一些实施方案中,并入碱基编辑器的APOBEC脱氨酶可包含APOBEC脱氨酶,该脱氨酶包含rAPOBEC1的W90Y突变,或另一APOBEC脱氨酶中的一个或多个对应突变。在一些实施方案中,并入碱基编辑器的APOBEC脱氨酶可包含APOBEC脱氨酶,该脱氨酶包含rAPOBEC1的R132E突变,或另一APOBEC脱氨酶中的一个或多个对应突变。在一些实施方案中,并入碱基编辑器的APOBEC脱氨酶可包含APOBEC脱氨酶,该脱氨酶包含rAPOBEC1的W90Y和R126E突变,或另一APOBEC脱氨酶中的一个或多个对应突变。在一些实施方案中,并入碱基编辑器的APOBEC脱氨酶可包含APOBEC脱氨酶,该脱氨酶包含rAPOBEC1的R126E和R132E突变,或另一APOBEC脱氨酶中的一个或多个对应突变。在一些实施方案中,并入碱基编辑器的APOBEC脱氨酶可包含APOBEC脱氨酶,该脱氨酶包含rAPOBEC1的W90Y和R132E突变,或另一APOBEC脱氨酶中的一个或多个对应突变。在一些实施方案中,并入碱基编辑器的APOBEC脱氨酶可包含APOBEC脱氨酶,该脱氨酶包含rAPOBEC1的W90Y、R126E和R132E突变,或另一APOBEC脱氨酶中的一个或多个对应突变。

[1522] 在一些实施方案中,并入碱基编辑器的APOBEC脱氨酶可包含APOBEC脱氨酶,该脱氨酶包含hAPOBEC3G的D316R和D317R突变,或另一APOBEC脱氨酶中的一个或多个对应突变。在一些实施方案中,本文所提供的任何融合蛋白包含APOBEC脱氨酶,该脱氨酶包含hAPOBEC3G的R320A突变,或另一APOBEC脱氨酶中的一个或多个对应突变。在一些实施方案中,并入碱基编辑器的APOBEC脱氨酶可包含APOBEC脱氨酶,该脱氨酶包含hAPOBEC3G的R320E突变,或另一APOBEC脱氨酶中的一个或多个对应突变。在一些实施方案中,并入碱基编辑器的APOBEC脱氨酶可包含APOBEC脱氨酶,该脱氨酶包含hAPOBEC3G的R313A突变,或另一APOBEC脱氨酶中的一个或多个对应突变。在一些实施方案中,并入碱基编辑器的APOBEC脱氨酶可包含APOBEC脱氨酶,该脱氨酶包含hAPOBEC3G的W285A突变,或另一APOBEC脱氨酶中的一个或多个对应突变。在一些实施方案中,并入碱基编辑器的APOBEC脱氨酶可包含APOBEC脱氨酶,该脱氨酶包含hAPOBEC3G的W285Y突变,或另一APOBEC脱氨酶中的一个或多个对应突变。在一些实施方案中,并入碱基编辑器的APOBEC脱氨酶可包含APOBEC脱氨酶,该脱氨酶包含hAPOBEC3G的R326E突变,或另一APOBEC脱氨酶中的一个或多个对应突变。在一些实施方案中,并入碱基编辑器的APOBEC脱氨酶可包含APOBEC脱氨酶,该脱氨酶包含hAPOBEC3G的W2865Y和R320E突变,或另一APOBEC脱氨酶中的一个或多个对应突变。在一些实施方案中,并入碱基编辑器的APOBEC脱氨酶可包含APOBEC脱氨酶,该脱氨酶包含hAPOBEC3G的R320E和R326E突变,或另一APOBEC脱氨酶中的一个或多个对应突变。在一些实施方案中,并入碱基编辑器的APOBEC脱氨酶可包含APOBEC脱氨酶,该脱氨酶包含hAPOBEC3G的W285Y和R326E突变,或另一APOBEC脱氨酶中的一个或多个对应突变。在一些实施方案中,并入碱基编辑器的APOBEC脱氨酶可包含APOBEC脱氨酶,该脱氨酶包含hAPOBEC3G的W285Y、R320E和R326E突变,或另一APOBEC脱氨酶中的一个或多个对应突变。

[1523] 大量经修饰的胞苷脱氨酶是可商购的,包括但不限于,SaBE3、SaKKH-BE3、VQR-BE3、EQR-BE3、VRER-BE3、YE1-BE3、EE-BE3、YE2-BE3和YEE-BE3,它们可以从Addgene获得(质粒85169、85170、85171、85172、85173、85174、85175、85176、85177)。在一些实施方案中,并

入碱基编辑器中的脱氨酶包含APOBEC1脱氨酶的全部或一部分。

[1524] C到T核碱基编辑蛋白的细节描述于国际PCT申请号PCT/US2016/058344(WO2017/070632)和Komor,A.C.,等人,“Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage”Nature 533,420-424(2016)中,其整体内容通过引用并入本文。

[1525] 胞苷脱氨酶

[1526] 在一些实施方案中,本发明的融合蛋白包含一个或多个胞苷脱氨酶结构域。在一些实施方案中,本文提供的胞苷脱氨酶能够将胞嘧啶或5-甲基胞嘧啶脱氨基为尿嘧啶或胸腺嘧啶。在一些实施方案中,本文提供的胞苷脱氨酶能够将DNA中的胞嘧啶脱氨基。胞苷脱氨酶可以源自任何合适的生物体。在一些实施方案中胞苷脱氨酶是天然存在的胞苷脱氨酶,其包括对应于本文所提供的任何突变的一个或多个突变。本领域技术人员将能够鉴定任何同源蛋白质中的对应残基,通过序列比对并确定同源残基。据此,本领域技术人员将能够生成任何天然存在的胞苷脱氨酶中的对应于本文所述任何突变的突变。在一些实施方案中,胞苷脱氨酶来自原核生物。在一些实施方案中,胞苷脱氨酶来自细菌。在一些实施方案中,胞苷脱氨酶来自哺乳动物(例如,人类)。

[1527] 在一些实施方案中,胞苷脱氨酶包含氨基酸序列,该氨基酸序列与本文详述的胞苷脱氨酶氨基酸序列中的任一者至少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。应了解,本文所提供的胞苷脱氨酶可包括一个或多个突变(例如,本文所提供的任何突变)。一些实施方案提供编码任何前述方面或如本文所述的胞苷脱氨酶核碱基编辑器的多核苷酸分子。在一些实施方案中,该多核苷酸经密码子优化。

[1528] 本公开提供任何具有某一百分比的一致性加上任何本文所述突变或其组合的脱氨酶结构域。在一些实施方案中,胞苷脱氨酶包含氨基酸序列,与参考序列或本文所提供的任何胞苷脱氨酶相比,该氨基酸序列具有1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、21、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50个或更多个突变。在一些实施方案中,胞苷脱氨酶包含氨基酸序列,与本领域已知或本文所述的氨基酸序列中的任一者相比,该氨基酸序列具有至少5、至少10、至少15、至少20、至少25、至少30、至少35、至少40、至少45、至少50、至少60、至少70、至少80、至少90、至少100、至少110、至少120、至少130、至少140、至少150、至少160、或至少170个相同的连续氨基酸残基。

[1529] 本发明的融合蛋白第二蛋白包含两个或更多个核酸编辑结构域。在一些实施方案中,核酸编辑结构域可以催化C到U的碱基变化。在一些实施方案中,核酸编辑结构域是脱氨酶结构域。在一些实施方案中,脱氨酶是胞苷脱氨酶。在一些实施方案中,脱氨酶是载脂蛋白BmRNA编辑复合物(APOBEC)家族脱氨酶。在一些实施方案中,脱氨酶是APOBEC1脱氨酶。在一些实施方案中,脱氨酶是APOBEC2脱氨酶。在一些实施方案中,脱氨酶是APOBEC3脱氨酶。在一些实施方案中,脱氨酶是APOBEC3A脱氨酶。在一些实施方案中,脱氨酶是APOBEC3B脱氨酶。在一些实施方案中,脱氨酶是APOBEC3C脱氨酶。在一些实施方案中,脱氨酶是APOBEC3D脱氨酶。在一些实施方案中,脱氨酶是APOBEC3E脱氨酶。在一些实施方案中,脱氨酶是APOBEC3F脱氨酶。在一些实施方案中,脱氨酶是APOBEC3G脱氨酶。在一些实施方案中,脱氨

酶是APOBEC3H脱氨酶。在一些实施方案中,脱氨酶是APOBEC4脱氨酶。在一些实施方案中,脱氨酶是激活诱导型脱氨酶(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,脱氨酶结构域是人类、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,脱氨酶是海七鳃鳗(Petromyzon marinus)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,脱氨酶是包含D316R D317R突变的人类APOBEC3G变体。在一些实施方案中,脱氨酶是人类APOBEC3G的片段并且包含对应于D316R D317R突变的突变。在一些实施方案中,核酸编辑结构域与本文所述的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。

[1530] 在某些实施方案中,本文所提供的融合蛋白包含一个或多个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包含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没有核酸酶活性的Cas9结构域(dCas9),或切断双螺旋DNA分子的一条链的Cas9结构域,称为Cas9切口酶(nCas9)。

[1531] 高保真度Cas9结构域

[1532] 本公开的一些方面提供高保真度Cas9结构域。在一些实施方案中,高保真度Cas9结构域是包含一个或多个突变的经工程改造的Cas9结构域,与对应的野生型Cas9结构域相比,该突变降低Cas9结构域与DNA的糖-磷酸酯主链之间的静电相互作用。不受缚于任何特定理论,与DNA的糖-磷酸酯主链的静电相互作用降低的高保真度Cas9结构域可具有较低的脱靶效应。在一些实施方案中,Cas9结构域(例如,野生型Cas9结构域)包含一个或多个突变,该突变降低Cas9结构域与DNA的糖-磷酸酯主链之间的缔合。在一些实施方案中,Cas9结构域包含一个或多个突变,该突变将Cas9结构域与DNA的糖-磷酸酯主链之间的缔合降低至少1%、至少2%、至少3%、至少4%、至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、或至少70%。

[1533] 在一些实施方案中,本文所提供的任何Cas9融合蛋白包含N497X、R661X、Q695X和/或Q926X突变中的一者或多者,或本文所提供的任何氨基酸序列中的对应突变,其中X是任何氨基酸。在一些实施方案中,本文所提供的任何Cas9融合蛋白包含N497A、R661A、Q695A和/或Q926A突变中的一者或多者,或本文所提供的任何氨基酸序列中的对应突变。在一些实施方案中,Cas9结构域包含D10A突变,或本文所提供的任何氨基酸序列中的对应突变。具有高保真度的Cas9结构域是本领域中已知的,并且对于熟练技术人员将会是显而易见的。例如,具有高保真度的Cas9结构域已经描述于Kleinstiver,B.P.,等人"High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects." Nature 529,490-495(2016)和Slaymaker,I.M.,等人"Rationally engineered Cas9 nucleases with improved specificity."Science 351,84-88(2015)中,各自的整体内容通过引用并入本文。

[1534] 在一些实施方案中,经修饰的Cas9是高保真度Cas9酶。在一些实施方案中,高保真度Cas9酶是SpCas9(K855A)、eSpCas9(1.1)、SpCas9-HF1或超精确Cas9变体(HypaCas9)。经

修饰的Cas9eSpCas9(1.1)含有丙氨酸置换,该丙氨酸置换削弱HNH/RuvC槽与非靶标DNA链之间的相互作用,防止链分离并且在脱靶位点切断。同样,SpCas9-HF1通过丙氨酸置换减少脱靶标记,该丙氨酸置换中断Cas9与DNA磷酸酯主链的相互作用。HypaCas9在REC3结构域中含有增加Cas9校对和靶标识别的突变(SpCas9 N692A/M694A/Q695A/H698A)。全部三种高保真度酶均产生低于野生型Cas9的脱靶编辑。

[1535] 示例性高保真度Cas9提供如下。相对于Cas9的高保真度Cas9结构域以粗体和下划线显示

[1536]
DKKYSIGL**A**IGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF
GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN
SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGL
FGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNL
SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS
KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQI
HLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETI
TPWNFEEVVDKGASAQSFIERMT**A**FDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASL
GTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWG**A**LSRKLINGIRDKQSGKTILDFLKSDGFANRNFM**A**LIHDDSLTFKEDIQ
KAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQ
TTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQE
LDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQL
LNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR**A**ITKHVAQILDSRMNTKYDE
NDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKL
ESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLI
ETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIAR
KKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDF
LEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS
HYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRD
KPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYET
RIDLSQLGGD

[1538] 包含napDNAbp以及胞苷脱氨酶和/或腺苷脱氨酶的融合蛋白

[1539] 本公开的一些方面提供融合蛋白,该融合蛋白包含Cas9结构域或其他核酸可编程DNA结合蛋白(例如,Cas12)以及一个或多个胞苷脱氨酶或腺苷脱氨酶结构域和/或DNA糖基化酶结构域。应了解,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/或腺苷脱氨酶融合。本文所公开的碱基编辑器的机构与可以以任何次序排列。

[1540] 在一些实施方案中,融合蛋白包含结构域A-C、A-D或A-E:

[1541] NH_2-[A-B-C]-COOH;

[1542] NH_2-[A-B-C-D]-COOH;或

[1543] NH_2-[A-B-C-D-E]-COOH;

[1544] 其中A和C或者A、C和E各自包含以下一者或多者:

[1545] 腺苷脱氨酶结构域或其活性片段,

[1546] 胞苷脱氨酶结构域或其活性片段,

[1547] DNA糖基化酶结构域或其活性片段;并且

[1548] 其中B或者B和D各自包含具有核酸序列特异性结合活性的一个或多个结构域。

[1549] 在一些实施方案中,融合蛋白包含以下结构:

[1550] NH_2-[A_n-B_o-C_n]-COOH;

[1551] NH_2-[A_n-B_o-C_n-D_o]-COOH;或

[1552] NH_2-[A_n-B_o-C_p-D_o-E_q]-COOH;

[1553] 其中A和C或者A、C和E各自包含以下一者或多者:

[1554] 腺苷脱氨酶结构域或其活性片段,

[1555] 胞苷脱氨酶结构域或其活性片段,

[1556] DNA糖基化酶结构域或其活性片段;并且

[1557] 其中n是整数:1、2、3、4或5,其中p是整数:1、2、3、4或5,其中q是整数:1、2、3、4或5;并且其中B或者B和D各自包含具有核酸序列特异性结合活性的结构域;并且其中o是整数:1、2、3、4或5。

[1558] 例如但不限于,在一些实施方案中,融合蛋白包含结构:

[1559] NH_2-[腺苷脱氨酶]-[Cas9结构域]-COOH;

[1560] NH_2-[Cas9结构域]-[腺苷脱氨酶]-COOH;

[1561] NH_2-[胞苷脱氨酶]-[Cas9结构域]-COOH;

[1562] NH_2-[Cas9结构域]-[胞苷脱氨酶]-COOH;

[1563] NH_2-[胞苷脱氨酶]-[Cas9结构域]-[腺苷脱氨酶]-COOH;

[1564] NH_2-[腺苷脱氨酶]-[Cas9结构域]-[胞苷脱氨酶]-COOH;

[1565] NH_2-[腺苷脱氨酶]-[胞苷脱氨酶]-[Cas9结构域]-COOH;

[1566] NH_2-[胞苷脱氨酶]-[腺苷脱氨酶]-[Cas9结构域]-COOH;

[1567] NH_2-[Cas9结构域]-[腺苷脱氨酶]-[胞苷脱氨酶]-COOH;或

[1568] NH_2-[Cas9结构域]-[胞苷脱氨酶]-[腺苷脱氨酶]-COOH。

[1569] 在一些实施方案中,本文提供的任何Cas12结构域或Cas12蛋白可以与本文提供的任何胞苷或腺苷脱氨酶融合。例如但不限于,在一些实施方案中,融合蛋白包含结构:

[1570] NH_2-[腺苷脱氨酶]-[Cas12结构域]-COOH;

[1571] NH_2-[Cas12结构域]-[腺苷脱氨酶]-COOH;

[1572] NH_2-[胞苷脱氨酶]-[Cas12结构域]-COOH;

[1573] NH_2-[Cas12结构域]-[胞苷脱氨酶]-COOH;

[1574] NH_2-[胞苷脱氨酶]-[Cas12结构域]-[腺苷脱氨酶]-COOH;

[1575] NH_2-[腺苷脱氨酶]-[Cas12结构域]-[胞苷脱氨酶]-COOH;

[1576] NH_2-[腺苷脱氨酶]-[胞苷脱氨酶]-[Cas12结构域]-COOH;

[1577] NH_2-[胞苷脱氨酶]-[腺苷脱氨酶]-[Cas12结构域]-COOH;

[1578] NH_2-[Cas12结构域]-[腺苷脱氨酶]-[胞苷脱氨酶]-COOH;或

[1579] NH_2-[Cas12结构域]-[胞苷脱氨酶]-[腺苷脱氨酶]-COOH。

[1580] 在一些实施方案中,腺苷脱氨酶是TadA*8。示例性融合蛋白结构包括下列:

[1581] NH_2-[TadA*8]-[Cas9结构域]-COOH;

[1582] NH_2-[Cas9结构域]-[TadA*8]-COOH;

[1583] NH_2-[TadA*8]-[Cas12结构域]-COOH;或

[1584] NH_2-[Cas12结构域]-[TadA*8]-COOH。

[1585] 在一些实施方案中,融合蛋白的腺苷脱氨酶包含TadA*8和胞苷脱氨酶和/或腺苷脱氨酶。在一些实施方案中,TadA*8是TadA*8.1、TadA*8.2、TadA*8.3、TadA*8.4、TadA*8.5、TadA*8.6、TadA*8.7、TadA*8.8、TadA*8.9、TadA*8.10、TadA*8.11、TadA*8.12、TadA*8.13、TadA*8.14、TadA*8.15、TadA*8.16、TadA*8.17、TadA*8.18、TadA*8.19、TadA*8.20、TadA*8.21、TadA*8.22、TadA*8.23或TadA*8.24。

[1586] 示例性融合蛋白结构包括下列:

[1587] NH_2-[TadA*8]-[Cas9/Cas12]-[腺苷脱氨酶]-COOH;

[1588] NH_2-[腺苷脱氨酶]-[Cas9/Cas12]-[TadA*8]-COOH;

[1589] NH_2-[TadA*8]-[Cas9/Cas12]-[胞苷脱氨酶]-COOH;或

[1590] NH_2-[胞苷脱氨酶]-[Cas9/Cas12]-[TadA*8]-COOH。

[1591] 在一些实施方案中,融合蛋白的腺苷脱氨酶包含TadA*9和胞苷脱氨酶和/或腺苷脱氨酶。示例性融合蛋白结构包括下列:

[1592] NH_2-[TadA*9]-[Cas9/Cas12]-[腺苷脱氨酶]-COOH;

[1593] NH_2-[腺苷脱氨酶]-[Cas9/Cas12]-[TadA*9]-COOH;

[1594] NH_2-[TadA*9]-[Cas9/Cas12]-[胞苷脱氨酶]-COOH;或

[1595] NH_2-[胞苷脱氨酶]-[Cas9/Cas12]-[TadA*9]-COOH。

[1596] 在一些实施方案中,融合蛋白可包含脱氨酶,该脱氨酶的侧翼是Cas9或Cas12多肽的N端片段和C端片段。在一些实施方案中,融合蛋白包含胞苷脱氨酶,该脱氨酶的侧翼是Cas9或Cas12多肽的N端片段和C端片段。在一些实施方案中,融合蛋白包含腺苷脱氨酶,该脱氨酶的侧翼是Cas9或Cas12多肽的N端片段和C端片段。

[1597] 在一些实施方案中,包含胞苷脱氨酶或腺苷脱氨酶和napDNAbp(例如,Cas9或Cas12结构域)的融合蛋白不包括连接子序列。在一些实施方案中,连接子存在于胞苷或腺苷脱氨酶与napDNAbp之间。在一些实施方案中上述通用架构中所用的“-”表示任选的连接子的存在。在一些实施方案中,胞苷或腺苷脱氨酶与napDNAbp经由本文所提供的任何连接子融合。例如,在一些实施方案中,胞苷或腺苷脱氨酶与napDNAbp经由本文所提供的任何连接子融合。

[1598] 应了解,本公开的融合蛋白可包含一个或多个额外特征。例如,在一些实施方案中,融合蛋白可包含抑制剂、细胞质定位序列、输出序列诸如核输出序列、或其他定位序列,

以及可用于融合蛋白的增溶、纯化或检测的序列标签。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、血细胞凝集素(HA)-标签、聚组氨酸标签(也称为组氨酸标签或His-标签)、麦芽糖结合蛋白(MBP)-标签、nus-标签、谷胱甘肽-S-转移酶(GST)-标签、绿色荧光蛋白(GFP)-标签、硫氧还蛋白-标签、S-标签、Softag(例如,Softag 1、Softag 3)、strep-标签、生物素连接酶标签、FlAsH标签、V5标签和SBP-标签。其他合适的序列对于本领域技术人员将是显而易见的。在一些实施方案中,融合蛋白包含一个或多个His标签。

[1599] 示例性但非限制性的融合蛋白在国际PCT申请CT/2017/044935、PCT/US2019/044935和PCT/US2020/016288中描述,其各自通过引用而以其整体并入本文。

[1600] 包含核定位序列F(NLS)的融合蛋白

[1601] 在一些实施方案中,本文所提供的融合蛋白进一步包含一个或多个(例如,2、3、4、5个)核靶向序列,例如,核定位序列(NLS)。在一种实施方案中,使用二分体NLS。在一些实施方案中,NLS包含氨基酸序列,该氨基酸序列促进包含NLS的蛋白质进入细胞核内(例如,通过核转运)。在一些实施方案中,本文所提供的任何融合蛋白进一步包含核定位序列(NLS)。在一些实施方案中,NLS融合至融合蛋白的N端。在一些实施方案中,NLS融合至融合蛋白的C端。在一些实施方案中,NLS融合至Cas9结构域的N端。在一些实施方案中,NLS融合至Cas9结构域的C端。在一些实施方案中,NLS融合至nCas9结构域或dCas9结构域的C端。在一些实施方案中,NLS融合至Cas12结构域的N端。在一些实施方案中,NLS融合至Cas12结构域的C端。在一些实施方案中,NLS融合至胞苷或腺苷脱氨酶的N端。在一些实施方案中,NLS融合至胞苷或腺苷脱氨酶的C端。在一些实施方案中,NLS经由一个或多个连接子融合至融合蛋白。在一些实施方案中,NLS融合至融合蛋白而无需连接子。在一些实施方案中,NLS包含本文所提供或引用的NLS序列中的任一者的氨基酸序列。其他核定位序列是本领域中已知的,并且对于熟练技术人员将会是显而易见的。例如,NLS序列在Plank等人的PCT/EP2000/011690中有所描述,其内容通过引用针对其对示例性核定位序列的公开内容并入本文。在一些实施方案中,NLS包含氨基酸序列PKKKRKVEGADKRTADGSEFESPKKKRKV、KRTADGSEFESPKKKRKV、KRPAATKKAGQAKKKK、KKTELQTTNAENKTKKL、KRGINDRNFWRGENGRKTR、RKSGKIAAIVVKRPRKPKKKRKV或MDSLLMNRRKFLYQFKNVRWAKGRRETYLC。

[1602] 在一些实施方案中,包含胞苷或腺苷脱氨酶、Cas9结构域和NLS的融合蛋白不包含连接子序列。在一些实施方案中,在结构域或蛋白质(例如,胞苷或腺苷脱氨酶、Cas9结构域或NLS)中的一者或多者之间存在连接子序列。在一些实施方案中,连接子存在于胞苷脱氨酶与腺苷脱氨酶结构域与napDNAbp之间。在一些实施方案中下述通用架构中所用的“-”表示任选的连接子的存在。在一些实施方案中,胞苷脱氨酶与腺苷脱氨酶与napDNAbp经由本文所提供的任何连接子融合。例如,在一些实施方案中,胞苷脱氨酶和腺苷脱氨酶与napDNAbp经由本文所提供的任何连接子融合。

[1603] 在一些实施方案中,具有胞苷或腺苷脱氨酶以及napDNAbp(例如,Cas9或Cas12)的示例性napDNAbp(例如,Cas9或Cas12)融合蛋白的通用架构包含以下结构中的任一者,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,且COOH是融合蛋白的C端。

[1604] NH_2-NLS-[胞苷脱氨酶]-[napDNAbp结构域]-COOH;

[1605] NH_2-NLS[napDNAbp结构域]-[胞苷脱氨酶]-COOH;

[1606] NH_2-[胞苷脱氨酶]-[napDNAbp结构域]-NLS-COOH;

[1607] NH_2-[napDNAbp结构域]-[胞苷脱氨酶]-NLS-COOH;

[1608] NH_2-NLS-[腺苷脱氨酶]-[napDNAbp结构域]-COOH;

[1609] NH_2-NLS[napDNAbp结构域]-[腺苷脱氨酶]-COOH;

[1610] NH_2-[腺苷脱氨酶]-[napDNAbp结构域]-NLS-COOH;

[1611] NH_2-[napDNAbp结构域]-[腺苷脱氨酶]-NLS-COOH;

[1612] NH_2-NLS-[胞苷脱氨酶]-[napDNAbp结构域]-[腺苷脱氨酶]-COOH;

[1613] NH_2-NLS-[腺苷脱氨酶]-[napDNAbp结构域]-[胞苷脱氨酶]-COOH;

[1614] NH_2-NLS-[腺苷脱氨酶][胞苷脱氨酶]-[napDNAbp结构域]-COOH;

[1615] NH_2-NLS-[胞苷脱氨酶]-[腺苷脱氨酶]-[napDNAbp结构域]-COOH;

[1616] NH_2-NLS-[napDNAbp结构域]-[腺苷脱氨酶]-[胞苷脱氨酶]-COOH;

[1617] NH_2-NLS-[napDNAbp结构域]-[胞苷脱氨酶]-[腺苷脱氨酶]-COOH;

[1618] NH_2-[胞苷脱氨酶]-[napDNAbp结构域]-[腺苷脱氨酶]-NLS-COOH;

[1619] NH_2-[腺苷脱氨酶]-[napDNAbp结构域]-[胞苷脱氨酶]-NLS-COOH;

[1620] NH_2-[腺苷脱氨酶][胞苷脱氨酶]-[napDNAbp结构域]-NLS-COOH;

[1621] NH_2-[胞苷脱氨酶]-[腺苷脱氨酶]-[napDNAbp结构域]-NLS-COOH;

[1622] NH_2-[napDNAbp结构域]-[腺苷脱氨酶]-[胞苷脱氨酶]-NLS-COOH;或

[1623] NH_2-[napDNAbp结构域]-[胞苷脱氨酶]-[腺苷脱氨酶]-NLS-COOH。在一些实施方案中,NLS存在于连接子内,或NLS位于连接子侧翼,例如,如本文所述。二分体NLS包含两个碱性氨基酸簇,它们通过相对短的间隔序列分隔(因此二分体-2个部分,而单分体NLS则不然)。核质蛋白的NLS,KR[PAATKKAGQA]KKKK,是无处不在的二分体信号的原型:两个碱性氨基酸的簇,通过约10个氨基酸的间隔序列分隔。示例性二分体NLS的序列如下:

[1624] PKKKRKVEGADKRTADGSEFESPKKKRKV

[1625] 可使用编码包含一个或多个核定位序列(NLS)的CRISPR酶的载体。例如,可使用(约)1、2、3、4、5、6、7、8、9、10个NLS。CRISPR酶可包含在氨基端处或氨基酸附近的NLS,在羧基端处或羧基端附件的约或超过约1、2、3、4、5、6、7、8、9、10个NLS,或其任意组合(例如,在氨基酸的一个或多个NLS以及在羧基端的一个或多个NLS)。当存在超过一个NLS时,各自可彼此独立地选择,使得单一NLS可存在于超过一个拷贝内且/或与一个或多个拷贝中的一个或多个其他NLS组合。

[1626] 该方法中使用的CRISPR酶可包含约6个NLS。当最接近NLS的氨基酸沿着从N端到C端在多肽链的约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,则该NLS视为接近N端或C端。

[1627] 额外结构域

[1628] 本文所述的碱基编辑器可包括任何有助于促进核碱基编辑、多核苷酸的核碱基修饰或改变的结构域。在一些实施方案中,碱基编辑器包含多核苷酸可编程核苷酸结合结构域(例如,Cas9)、核碱基编辑结构域(例如,脱氨酶结构域)和一个或多个额外结构域。在一些实施方案中,额外结构域可促进碱基编辑器的酶促和催化功能、碱基编辑器的结合功能,或可以是将会干扰所希望的碱基编辑结果的细胞机构(例如,酶)的抑制剂。在一些实施方

案中,碱基编辑器可包含核酸酶I、切口酶、重组酶、脱氨酶、甲基转移酶、甲基化酶、乙酰化酶、乙酰转移酶、转录激活因子或转录抑制因子结构域。

[1629] 在一些实施方案中,碱基编辑器可包含尿嘧啶糖基化酶抑制剂(UGI)结构域。在一些实施方案中,响应U:G异源双链DNA的存在的细胞DNA修复可能对细胞中核碱基编辑效率的降低负责。在此类实施方案中,尿嘧啶DNA糖基化酶(UDG)可催化U从细胞中的DNA移除,这可能启动碱基切除修复(BER),主要导致U:G对逆转为C:G对。在此类实施方案中,BER可以在包含一个或多个结构域的碱基编辑器中被抑制,这些结构域结合单链,阻断被编辑的碱基,抑制UGI,抑制BER,包含被编辑的碱基,和/或促进对未被编辑的链的修复。因此,本公开设想包含UGI结构域的碱基编辑器融合蛋白。

[1630] 在一些实施方案中,碱基编辑器包含双链断裂(DSB)结合蛋白的全部或一部分作为结构域。例如,DSB结合蛋白可包括细菌噬菌体μ的Gam蛋白,其结合至DSB的末端并且可保护它们免于降解。参见Komor,A.C.,等人,"Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity"Science Advances3:eaao4774(2017),其整体内容通过引用并入本文。

[1631] 此外,在一些实施方案中,Gam蛋白可以融合至碱基编辑器的N端。在一些实施方案中,Gam蛋白可以融合至碱基编辑器的C端。细菌噬菌体μ的Gam蛋白可以结合至双链断裂(DSB)的末端并保护它们免于降解。在一些实施方案中,使用Gam结合DSB的游离末端可以减少碱基编辑器期间的插入缺失的形成。在一些实施方案中,将174个残基的Gam蛋白融合至碱基编辑器的N端。参见Komor,A.C.,等人,"Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity"Science Advances 3:eaao4774(2017)。在一些实施方案中,一个或多个突变可相对于野生型结构域改变碱基编辑器结构域的长度。例如,在至少一个结构域中至少一个氨基酸的缺失可以减少碱基编辑器的长度。在另一种情况下,一个或多个突变相对于野生型结构域不改变结构域的长度。例如,在任何结构域中的置换不改变碱基编辑器的长度。

[1632] 其中全部结构域的长度均与野生型结构域相同的此类碱基编辑器的非限制性示例包括:

[1633] NH2-[核碱基编辑结构域]-连接子1-[APOBEC1]-连接子2-[核碱基编辑结构域]-COOH;

[1634] NH2-[核碱基编辑结构域]-连接子1-[APOBEC1]-[核碱基编辑结构域]-COOH;

[1635] NH2-[核碱基编辑结构域]-[APOBEC1]-连接子2-[核碱基编辑结构域]-COOH;

[1636] NH2-[核碱基编辑结构域]-[APOBEC1]-[核碱基编辑结构域]-COOH;

[1637] NH2-[核碱基编辑结构域]-连接子1-[APOBEC1]-连接子2-[核碱基编辑结构域]-[UGI]-COOH;

[1638] NH2-[核碱基编辑结构域]-连接子1-[APOBEC1]-[核碱基编辑结构域]-[UGI]-COOH;

[1639] NH2-[核碱基编辑结构域]-[APOBEC1]-连接子2-[核碱基编辑结构域]-[UGI]-COOH;

[1640] NH2-[核碱基编辑结构域]-[APOBEC1]-[核碱基编辑结构域]-[UGI]-COOH;

[1641] NH2-[UGI]-[核碱基编辑结构域]-连接子1-[APOBEC1]-连接子2-[核碱基编辑结构域]-COOH;

[1642] NH2-[UGI]-[核碱基编辑结构域]-连接子1-[APOBEC1]-[核碱基编辑结构域]-COOH;

[1643] NH2-[UGI]-[核碱基编辑结构域]-[APOBEC1]-连接子2-[核碱基编辑结构域]-COOH;或

[1644] NH2-[UGI]-[核碱基编辑结构域]-[APOBEC1]-[核碱基编辑结构域]-COOH。

[1645] 碱基编辑器系统

[1646] 本文提供使用碱基编辑器系统来编辑核碱基的系统、组合物和方法。在一些实施方案中,碱基电击器系统包含碱基编辑器(BE),该BE包含(1)多核苷酸可编程核苷酸结合结构域和用于编辑核碱基的核碱基编辑结构域(例如,脱氨酶结构域),和(2)与该多核苷酸可编程核苷酸结合结构域协同作用的向导多核苷酸(例如,向导RNA)。在一些实施方案中,碱基编辑器系统是胞苷碱基编辑器(CBE)。在一些实施方案中,碱基编辑器系统是腺苷碱基编辑器(ABE)。在一些实施方案中,多核苷酸可编程核苷酸结合结构域是多核苷酸可编程DNA结合结构域。在一些实施方案中,多核苷酸可编程核苷酸结合结构域是多核苷酸可编程RNA结合结构域。在一些实施方案中,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胞苷脱氨酶或胞嘧啶脱氨酶。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,腺苷碱基编辑器可以将DNA中的腺嘌呤脱氨基。在一些实施方案中,碱基编辑器能够将DNA内的胞苷脱氨基。

[1647] 在一些实施方案中本文所提供的碱基编辑系统提供新的基因组编辑途径,其使用含有催化缺陷型酿脓链球菌Cas9、脱氨酶(例如,胞苷或腺苷脱氨酶)、碱基切除修复抑制剂的融合蛋白来诱导DNA中的可编程单核苷酸(C→T或A→G)变化而不生成双链DNA断裂,不修复供体DNA模板,并且不诱导过量的随机插入和缺失。

[1648] 核碱基编辑蛋白的细节在国际PCT申请号PCT/2017/045381(WO2018/027078)和PCT/US2016/058344(WO2017/070632)中描述,其各自通过引用以其整体并入本文。也参见Komor,A.C.,等人,"Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage"Nature 533,420-424(2016);Gaudelli,N.M.,等人,"Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage"Nature 551,464-471(2017);和Komor,A.C.,等人,"Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:Abase editors with higher efficiency and product purity"Science Advances 3:eaao4774(2017),其整体内容通过引用并入本文。

[1649] 本文所提供的碱基编辑器系统的使用包括以下步骤:(a)使受试者的多核苷酸的靶标核苷酸序列(例如,双链或单链DNA或RNA)与包含核碱基编辑器(例如,腺苷碱基编辑器或胞苷碱基编辑器)和向导多核酸(例如,gRNA)的碱基编辑器系统接触,其中靶标核苷酸序列包含被靶向的核碱基对;(b)诱导所述靶标区的链分离;(c)将所述靶标区的单链中的靶标核碱基对的第一核碱基转变为第二核碱基;以及(d)切断所述靶标区的不超过一条链,其中与第一核碱基互补的第三核碱基被替换为与第二核碱基的第四核碱基。应了解,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个或多个基因中的多个核碱基对。在一些实施方案中,本文所提供的碱基编辑器系统能够对一个或多个基因中的多个核碱基对进行多重编辑。在一些实施方案中,该多个核碱基对位于同一基因内。在一些实施方案中,该多个核碱基对位于一个或多个基因内,其中至少一个基因位于不同的基因座中。

[1650] 在一些实施方案中,被切断的单链(切口链)杂交至向导核酸。在一些实施方案中,被切断的单链与包含第一核碱基的链相对。在一些实施方案中,碱基编辑器包含Cas9结构域。在一些实施方案中,第一碱基是腺嘌呤,且第二碱基不是G、C、A或T。在一些实施方案中,第二碱基是肌苷。

[1651] 在一些实施方案中,可利用单向导多核苷酸将脱氨酶靶向至靶标核酸序列。在一些实施方案中,可利用一对向导多核苷酸将不同的脱氨酶靶向至靶标核酸序列。

[1652] 碱基编辑器系统核碱基组分和多核苷酸可编程核苷酸结合组分可以共价地或非共价地彼此缔合。例如,在一些实施方案中,脱氨酶结构域可以通过多核苷酸可编程核苷酸结合结构域被靶向至靶标核苷酸序列。在一些实施方案中,多核苷酸可编程核苷酸结合结构域可融合或链接至脱氨酶结构域。在一些实施方案中,多核苷酸可编程核苷酸结合结构域可通过与脱氨酶结构域非共价地相互作用或缔合而将脱氨酶结构域靶向至靶标核苷酸序列。例如,在一些实施方案中,核碱基编辑组分,例如,脱氨酶组分可包含额外的异源部分或结构域,该异源部分或结构域能够与作为多核苷酸可编程核苷酸结合结构域的一部分的额外异源部分或结构域相互作用、缔合或形成复合物。在一些实施方案中,额外的异源部分可能能够与多肽结合、相互作用、缔合或形成复合物。在一些实施方案中,额外的异源部分可能能够与多核苷酸结合、相互作用、缔合或形成复合物。在一些实施方案中,额外的异源部分可能能够结合至向导多核苷酸。在一些实施方案中,额外的异源部分可能能够结合至多肽连接子。在一些实施方案中,额外的异源部分可能能够结合至多核苷酸连接子。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(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。

[1653] 碱基编辑器系统可进一步包含向导多核苷酸组分。应了解,碱基编辑器系统的组分可以经由共价键、非共价相互作用、或其缔合和相互作用的任意组合彼此缔合。在一些实施方案中,脱氨结构域可以通过向导多核苷酸被靶向至靶标核苷酸序列。例如,在一些实施方案中,碱基编辑器系统的核碱基编辑组分,例如脱氨酶组分,可包含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),该额外的异源部分或结构域能够与向导多核苷酸的部分或链段(例如,多核苷酸基序)相互作用、缔合或形成复合物。在一些实施方案中,额外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)可融合或链接至脱氨酶结构域。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、缔合或形成复合物。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、缔合或形成复合物。在一些实施方案中,额外的异源部分可能能够结合至向导多核苷酸。在一些实施方案中,额外的异源部分可能能够结合至多肽连接子。在一些实施方案中,额外的异源部分可能能够结合至多核苷酸连接子。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(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。

[1654] 在一些实施方案中,碱基编辑器系统可进一步包含碱基切除修复(BER)的抑制剂组分。应了解,碱基编辑器系统的组分可以经由共价键、非共价相互作用、或其缔合和相互作用的任意组合彼此缔合。BER的抑制剂组分可包含碱基切除修复抑制剂。在一些实施方案中,碱基切除修复的抑制剂可以是尿嘧啶DNA糖基化酶抑制剂(UGI)。在一些实施方案中,碱基切除修复的抑制剂可以是肌苷碱基切除修复抑制剂。在一些实施方案中,碱基切除修复的抑制剂可以通过多核苷酸可编程核苷酸结合结构域被靶向至靶标核苷酸序列。在一些实施方案中,多核苷酸可编程核苷酸结合结构域可融合或链接至碱基切除修复的抑制剂。在一些实施方案中,多核苷酸可编程核苷酸结合结构域可融合或链接至脱氨酶结构域和碱基切除修复的抑制剂。在一些实施方案中,多核苷酸可编程核苷酸结合结构域可通过与碱基切除修复的抑制剂非共价地相互作用或缔合而将碱基切除修复的抑制剂靶向至靶标核苷酸序列。例如,在一些实施方案中,碱基切除修复的抑制剂组分可包含额外的异源部分或结构域,该异源部分或结构域能够与作为多核苷酸可编程核苷酸结合结构域的一部分的额外异源部分或结构域相互作用、缔合或形成复合物。在一些实施方案中,碱基切除修复的抑制剂可以通过向导多核苷酸被靶向至靶标核苷酸序列。例如,在一些实施方案中,碱基编辑器系统的碱基切除修复的抑制剂组分,可包含额外的异源部分或结构域(例如,多核苷酸结合结构域诸如RNA或DNA结合蛋白),该额外的异源部分或结构域能够与向导多核苷酸的部分或链段(例如,多核苷酸基序)相互作用、缔合或形成复合物。在一些实施方案中,向导多核苷酸的额外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)可融合或链接至碱基切除修复的抑制剂。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、缔合或形成复合物。在一些实施方案中,额外的异源部分可能能够结合至向导多核苷酸。在一些实施方案中,额外的异源部分可能能够结合至多肽连接子。在一些实施方案中,额外的异源部分可能能够结合至多核苷酸连接子。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(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。

[1655] 在一些实施方案中,碱基编辑器抑制被编辑链的碱基切除修复(BER)。在一些实施方案中,碱基编辑器保护或结合未被编辑的链。在一些实施方案中,碱基编辑器包含UGI活性。在一些实施方案中,碱基编辑器包含无催化活性的肌苷特异性核酸酶。在一些实施方案中,碱基编辑器包含切口酶活性。在一些实施方案中,预期编辑的碱基对在PAM位点上游。在一些实施方案中,预期编辑的碱基对在PAM位点上游1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个核苷酸。在一些实施方案中,预期编辑的碱基对在PAM位点下游。在一些实施方案中,预期编辑的碱基对在PAM位点下游1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个核苷酸。

[1656] 在一些实施方案中,该方法不需要标准(例如,NGG)PAM位点。在一些实施方案中,核碱基编辑器包含连接子或间隔序列。在一些实施方案中,连接子或间隔序列是1至25个氨基酸的长度。在一些实施方案中,连接子或间隔序列是5至20个氨基酸的长度。在一些实施方案中,连接子或间隔序列是10、11、12、13、14、15、16、17、18、19或20个氨基酸的长度。

[1657] 在一些实施方案中,本文所提供的碱基编辑融合蛋白需要定位在精确位置处,例如,所定义区(例如,“脱氨基窗口”)内的靶标碱基被放置之处。在一些实施方案中,靶标可以在4个碱基区内。在一些实施方案中,这种所定义的靶标区可以是PAM上游约15个碱基。参见Komor,A.C.,等人,“Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage”Nature 533,420-424(2016);Gaudelli,N.M.,等人,“Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage”Nature 551,464-471(2017);和Komor,A.C.,等人,“Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity”Science Advances 3:eaao4774(2017),其整体内容通过引用并入本文。

[1658] 在一些实施方案中,靶标区包含靶标窗口,其中靶标窗口包含靶标核碱基对。在一些实施方案中,靶标窗口包含1至10个核苷酸。在一些实施方案中,靶标窗口是1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个核苷酸的长度。在一些实施方案中,预期编辑的碱基对在靶标窗口内。在一些实施方案中,靶标窗口包含预期编辑的碱基对。在一些实施方案中,该方法使用本文所提供的任何碱基编辑器执行。在一些实施方案中,靶标窗口是脱氨基窗口。脱氨基窗口可以是定义区,其中碱基编辑器作用于靶标核苷酸并将靶标核苷酸脱氨基。在一些实施方案中,脱氨基窗口在2、3、4、5、6、7、8、9或10个碱基区内。在一些实施方案中,脱氨基窗口是PAM上游5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24或25个碱基。

[1659] 本公开的碱基编辑器可包含任何促进对靶标多核苷酸序列进行编辑的结构域、特征或氨基酸序列。例如,在一些实施方案中,碱基编辑器包含核定位序列(NLS)。在一些实施方案中,碱基编辑器的NLS定位在脱氨酶结构域与多核苷酸可编程核苷酸结合结构域之间。在一些实施方案中,碱基编辑器的NLS定位在多核苷酸可编程核苷酸结合结构域的C端。

[1660] 可存在于本文所公开的碱基编辑器中的其他示例性特征是定位序列诸如细胞质定位序列、输出序列诸如核输出序列、或其他定位序列,以及可用于融合蛋白的增溶、纯化或检测的序列标签。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、血细胞凝集素(HA)-标签、聚组氨酸标签(也称为组氨酸标签或His-标签)、麦芽糖结合蛋白(MBP)-标签、nus-标签、谷胱甘肽-S-转移酶(GST)-标签、绿色荧光蛋白(GFP)-标签、硫氧还蛋白-标签、S-标签、Softag(例如,Softag 1、Softag3)、strep-标签、生物素连接酶标签、FlAsH标签、V5标签和SBP-标签。其他合适的序列对于本领域技术人员将是显而易见的。在一些实施方案中,融合蛋白包含一个或多个His标签。

[1661] 在一些实施方案中,非限制性的示例性胞苷碱基编辑器(CBE)包括BE1(APOBEC1-XTEN-dCas9)、BE2(APOBEC1-XTEN-dCas9-UGI)、BE3(APOBEC1-XTEN-dCas9(A840H)-UGI)、BE3-Gam、saBE3、saBE4-Gam、BE4、BE4-Gam、saBE4或saB4E-Gam。BE4将APOBEC1-Cas9n(D10A)连接子延长至32个氨基酸并将Cas9n-UGI连接子延长至9个氨基酸,并且将第二个UGI拷贝追加至该构建体的C端,其中另一个9个氨基酸的连接子并入单一碱基编辑器构建体内。碱基编辑器saBE3和saBE4中,酿脓链球菌Cas9n(D10A)被替换为较小的金黄色葡萄球菌Cas9n(D10A)。BE3-Gam、saBE3-Gam、BE4-Gam和saBE4-Gam中,Gam蛋白的174个残基经由该16个氨

基酸的XTEN连接子融合至N-terminus of BE3、saBE3、BE4和saBE4的N端。

[1662] 在一些实施方案中,腺苷碱基编辑器(ABE)可以将DNA中的腺嘌呤脱氨基。在一些实施方案中,通过将BE3的APOBEC1组分替换为天然的或工程化的大肠杆菌TadA、人类ADAR2、小鼠ADA或人类ADAT2,生成了ABE。在一些实施方案中,ABE包含进化的adA变体。在一些实施方案中,ABE是ABE 1.2(TadA*-XTEN-nCas9-NLS)。在一些实施方案中,TadA*包含A106V和D108N突变。

[1663] 在一些实施方案中,ABE是第二代ABE。在一些实施方案中,ABE是ABE2.1,其包含TadA*中的额外突变D147Y和E155V(TadA*2.1)。在一些实施方案中,ABE是ABE2.2,其中ABE2.1融合至无催化活性形式的人类烷基腺苷DNA糖基化酶(具有E125Q突变的AAG)。在一些实施方案中,ABE是ABE2.3,其中ABE2.1融合至无催化活性形式的大肠杆菌Endo V(通过D35A突变失活)。在一些实施方案中,ABE是ABE2.6,其具有长度为ABE2.1中的连接子两倍的连接子(32个氨基酸,$(SGGS)_2$-XTEN-$(SGGS)_2$)。在一些实施方案中,ABE是ABE2.7,它是与额外的野生型TadA单体系结的ABE2.1。在一些实施方案中,ABE是ABE2.8,它是与额外的TadA*2.1单体系结的ABE2.1。在一些实施方案中,ABE是ABE2.9,它是进化的TadA(TadA*2.1)至ABE2.1的N端的直接融合物。在一些实施方案中,ABE是ABE2.10,它是野生型TadA至ABE2.1的N端的直接融合物。在一些实施方案中,ABE是ABE2.11,它是在TadA*单体的N端处具有失活E59A突变的ABE2.9。在一些实施方案中,ABE是ABE2.12,它是在内部TadA*单体中具有失活E59A突变的ABE2.9。

[1664] 在一些实施方案中,ABE是第三代ABE。在一些实施方案中,ABE是ABE3.1,它是具有三个额外的TadA突变(L84F、H123Y和I156F)的ABE2.3。

[1665] 在一些实施方案中,ABE是第四代ABE。在一些实施方案中,ABE是ABE4.3,它是具有额外的TadA突变A142N(TadA*4.3)的ABE3.1。

[1666] 在一些实施方案中,ABE是第五代ABE。在一些实施方案中,ABE是ABE5.1,它通过将来自存活克隆体的共有突变组(H36L、R51L、S146C和K157N)导入ABE3.1内而生成的。在一些实施方案中,ABE是ABE5.3,它具有异二聚构建体,该构建体含有融合自内部进化的TadA*的野生型大肠杆菌TadA。在一些实施方案中,ABE还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12中所示。在一些实施方案中,ABE是第六代ABE。在一些实施方案中,ABE是ABE6.1、ABE6.2、ABE6.3、ABE6.4、ABE6.5或ABE6.6,如下表12中所示。在一些实施方案中,ABE是第七代ABE。在一些实施方案中,ABE是ABE7.1、ABE7.2、ABE7.3、ABE7.4、ABE7.5、ABE7.6、ABE7.7、ABE7.8、ABE 7.9或ABE7.10,如下表12中所示。

[1667] 表12.ABE的基因型

[1668]

	23	26	36	37	48	49	51	72	84	87	106	108	123	125	142	146	147	152	155	156	157	161
ABE0.1	W	R	H	N	P		R	N	L	S	A	D	H	G	A	S	D	R	E	I	K	K
ABE0.2	W	R	H	N	P		R	N	L	S	A	D	H	G	A	S	D	R	E	I	K	K
ABE1.1	W	R	H	N	P		R	N	L	S	A	N	H	G	A	S	D	R	E	I	K	K
ABE1.2	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	D	R	E	I	K	K
ABE2.1	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.2	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.3	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K

[1669]

	23	26	36	37	48	49	51	72	84	87	106	108	123	125	142	146	147	152	155	156	157	161
ABE2.4	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.5	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.6	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.7	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.8	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.9	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.10	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.11	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.12	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE3.1	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.2	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.3	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.4	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.5	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.6	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.7	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.8	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE4.1	W	R	H	N	P		R	N	L	S	V	N	H	G	N	S	Y	R	V	I	K	K
ABE4.2	W	G	H	N	P		R	N	L	S	V	N	H	G	N	S	Y	R	V	I	K	K
ABE4.3	W	R	H	N	P		R	N	F	S	V	N	Y	G	N	S	Y	R	V	F	K	K
ABE5.1	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.2	W	R	H	S	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	T
ABE5.3	W	R	L	N	P		L	N	I	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.4	W	R	H	S	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	T
ABE5.5	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.6	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.7	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.8	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.9	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.10	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.11	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.12	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.13	W	R	H	N	P		L	D	F	S	V	N	Y	A	A	S	Y	R	V	F	K	K
ABE5.14	W	R	H	N	S		L	N	F	C	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE6.1	W	R	H	N	S		L	N	F	S	V	N	Y	G	N	S	Y	R	V	F	K	K
ABE6.2	W	R	H	N	T	V	L	N	F	S	V	N	Y	G	N	S	Y	R	V	F	N	K
ABE6.3	W	R	L	N	S		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE6.4	W	R	L	N	S		L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
ABE6.5	W	R	L	N	T	V	L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE6.6	W	R	L	N	T	V	L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
ABE7.1	W	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE7.2	W	R	L	N	A		L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
ABE7.3	L	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE7.4	R	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE7.5	W	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	H	V	F	N	K
ABE7.6	W	R	L	N	A		L	N	I	S	V	N	Y	G	A	C	Y	P	V	F	N	K
ABE7.7	L	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	P	V	F	N	K
ABE7.8	L	R	L	N	A		L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
ABE7.9	L	R	L	N	A		L	N	F	S	V	N	Y	G	N	C	Y	P	V	F	N	K

[1670]

	23	26	36	37	48	49	51	72	84	87	106	108	123	125	142	146	147	152	155	156	157	161
ABE7.10	R	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	P	V	F	N	K

[1671] 在一些实施方案中,碱基编辑器是第八代ABE(ABE8)。在一些实施方案中,ABE8含有TadA*8变体。在一些实施方案中,ABE8具有含有TadA*8变体的单体构建体(“ABE8.x-m”)。在一些实施方案中,ABE8是ABE8.1-m,其具有含有具有Y147T突变的TadA*7.10(TadA*8.1)的单体构建体。在一些实施方案中,ABE8是ABE8.2-m,其具有含有具有Y147R突变的TadA*7.10(TadA*8.2)的单体构建体。在一些实施方案中,ABE8是ABE8.3-m,其具有含有具有Q154S突变的TadA*7.10(TadA*8.3)的单体构建体。在一些实施方案中,ABE8是ABE8.4-m,其具有含有具有Y123H突变的TadA*7.10(TadA*8.4)的单体构建体。在一些实施方案中,ABE8是ABE8.5-m,其具有含有具有V82S突变的TadA*7.10(TadA*8.5)的单体构建体。在一些实施方案中,ABE8是ABE8.6-m,其具有含有具有T166R突变的TadA*7.10(TadA*8.6)的单体构建体。在一些实施方案中,ABE8是ABE8.7-m,其具有含有具有Q154R突变的TadA*7.10(TadA*8.7)的单体构建体。在一些实施方案中,ABE8是ABE8.8-m,其具有含有具有Y147R、Q154R和Y123H突变的TadA*7.10(TadA*8.8)的单体构建体。在一些实施方案中,ABE8是ABE8.9-m,其具有含有具有Y147R、Q154R和I76Y突变的TadA*7.10(TadA*8.9)的单体构建体。在一些实施方案中,ABE8是ABE8.10-m,其具有含有具有Y147R、Q154R和T166R突变的TadA*7.10(TadA*8.10)的单体构建体。在一些实施方案中,ABE8是ABE8.11-m,其具有含有具有Y147T和Q154R突变的TadA*7.10(TadA*8.11)的单体构建体。在一些实施方案中,ABE8是ABE8.12-m,其具有含有具有Y147T和Q154S突变的TadA*7.10(TadA*8.12)的单体构建体。

[1672] 在一些实施方案中,ABE8是ABE8.13-m,其具有含有具有Y123H(Y123H从H123Y恢复而来)、Y147R、Q154R和I76Y突变的TadA*7.10(TadA*8.13)的单体构建体。在一些实施方案中,ABE8是ABE8.14-m,其具有含有具有I76Y和V82S突变的TadA*7.10(TadA*8.14)的单体构建体。在一些实施方案中,ABE8是ABE8.15-m,其具有含有具有V82S和Y147R突变的TadA*7.10(TadA*8.15)的单体构建体。在一些实施方案中,ABE8是ABE8.16-m,其具有含有具有V82S、Y123H(Y123H从H123Y恢复而来)和Y147R突变的TadA*7.10(TadA*8.16)的单体构建体。在一些实施方案中,ABE8是ABE8.17-m,其具有含有具有V82S和Q154R突变的TadA*7.10(TadA*8.17)的单体构建体。在一些实施方案中,ABE8是ABE8.18-m,其具有含有具有V82S、Y123H(Y123H从H123Y恢复而来)和Q154R突变的TadA*7.10(TadA*8.18)的单体构建体。在一些实施方案中,ABE8是ABE8.19-m,其具有含有具有V82S、Y123H(Y123H从H123Y恢复而来)、Y147R和Q154R突变的TadA*7.10(TadA*8.19)的单体构建体。在一些实施方案中,ABE8是ABE8.20-m,其具有含有具有I76Y、V82S、Y123H(Y123H从H123Y恢复而来)、Y147R和Q154R突变的TadA*7.10(TadA*8.20)的单体构建体。在一些实施方案中,ABE8是ABE8.21-m,其具有含有具有Y147R和Q154S突变的TadA*7.10(TadA*8.21)的单体构建体。在一些实施方案中,ABE8是ABE8.22-m,其具有含有具有V82S和Q154S突变的TadA*7.10(TadA*8.22)的单体构建体。在一些实施方案中,ABE8是ABE8.23-m,其具有含有具有V82S和Y123H(Y123H从H123Y恢复而来)突变的TadA*7.10(TadA*8.23)的单体构建体。在一些实施方案中,ABE8是ABE8.24-m,其具有含有具有V82S、Y123H(Y123H从H123Y恢复而来)和Y147T突变的TadA*7.10(TadA*8.24)的单体构建体。

[1673] 在一些实施方案中,ABE8具有含有野生型大肠杆菌TadA与TadA*8变体(“ABE8.x-d”)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.1-d,其具有含有野生型大肠杆菌TadA与具有Y147T突变的TadA*7.10(TadA*8.1)融合的异二聚构建体。在一些实施方案

中,ABE8是ABE8.2-d,其具有含有野生型大肠杆菌TadA与具有Y147R突变的TadA*7.10(TadA*8.2)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.3-d,其具有含有野生型大肠杆菌TadA与具有Q154S突变的TadA*7.10(TadA*8.3)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.4-d,其具有含有野生型大肠杆菌TadA与具有Y123H突变的TadA*7.10(TadA*8.4)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.5-d,其具有含有野生型大肠杆菌TadA与具有V82S突变的TadA*7.10(TadA*8.5)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.6-d,其具有含有野生型大肠杆菌TadA与具有T166R突变的TadA*7.10(TadA*8.6)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.7-d,其具有含有野生型大肠杆菌TadA与具有Q154R突变的TadA*7.10(TadA*8.7)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.8-d,其具有含有野生型大肠杆菌TadA与具有Y147R、Q154R和Y123H突变的TadA*7.10(TadA*8.8)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.9-d,其具有含有野生型大肠杆菌TadA与具有Y147R、Q154R和I76Y突变的TadA*7.10(TadA*8.9)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.10-d,其具有含有野生型大肠杆菌TadA与具有Y147R、Q154R和T166R突变的TadA*7.10(TadA*8.10)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.11-d,其具有含有野生型大肠杆菌TadA与具有Y147T和Q154R突变的TadA*7.10(TadA*8.11)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.12-d,其具有含有野生型大肠杆菌TadA与具有Y147T和Q154S突变的TadA*7.10(TadA*8.12)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.13-d,其具有含有野生型大肠杆菌TadA与具有Y123H(Y123H从H123Y恢复而来)、Y147R、Q154R和I76Y突变的TadA*7.10(TadA*8.13)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.14-d,其具有含有野生型大肠杆菌TadA与具有I76Y和V82S突变的TadA*7.10(TadA*8.14)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.15-d,其具有含有野生型大肠杆菌TadA与具有V82S和Y147R突变的TadA*7.10(TadA*8.15)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.16-d,其具有含有野生型大肠杆菌TadA与具有V82S、Y123H(Y123H从H123Y恢复而来)和Y147R突变的TadA*7.10(TadA*8.16)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.17-d,其具有含有野生型大肠杆菌TadA与具有V82S和Q154R突变的TadA*7.10(TadA*8.17)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.18-d,其具有含有野生型大肠杆菌TadA与具有V82S、Y123H(Y123H从H123Y恢复而来)和Q154R突变的TadA*7.10(TadA*8.18)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.19-d,其具有含有野生型大肠杆菌TadA与具有V82S、Y123H(Y123H从H123Y恢复而来)、Y147R和Q154R突变的TadA*7.10(TadA*8.19)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.20-d,其具有含有野生型大肠杆菌TadA与具有I76Y、V82S、Y123H(Y123H从H123Y恢复而来)、Y147R和Q154R突变的TadA*7.10(TadA*8.20)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.21-d,其具有含有野生型大肠杆菌TadA与具有Y147R和Q154S突变的TadA*7.10(TadA*8.21)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.22-d,其具有含有野生型大肠杆菌TadA与具有V82S和Q154S突变的TadA*7.10(TadA*8.22)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.23-d,其具有含有野生型大肠杆菌TadA与具有V82S和Y123H(Y123H从H123Y恢复而来)突变的TadA*7.10(TadA*8.23)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.24-d,其具有含有野生型大肠杆菌TadA与具有V82S、Y123H(Y123H从H123Y恢复而来)和Y147T突变的TadA*

7.10(TadA*8.23)融合的异二聚构建体。

[1674] 在一些实施方案中,ABE8具有含有TadA*7.10与TadA*8变体融合的异二聚构建体(“ABE8.x-7”)。在一些实施方案中,ABE8是ABE8.1-7,其具有含有具有Y147T突变的TadA*7.10(TadA*8.1)的异二聚构建体。在一些实施方案中,ABE8是ABE8.2-7,其具有含有具有Y147R突变的TadA*7.10(TadA*8.2)的异二聚构建体。在一些实施方案中,ABE8是ABE8.3-7,其具有含有具有Q154S突变的TadA*7.10(TadA*8.3)的异二聚构建体。在一些实施方案中,ABE8是ABE8.4-7,其具有含有具有Y123H突变的TadA*7.10(TadA*8.4)的异二聚构建体。在一些实施方案中,ABE8是ABE8.5-7,其具有含有具有V82S突变的TadA*7.10(TadA*8.5)的异二聚构建体。在一些实施方案中,ABE8是ABE8.6-7,其具有含有具有T166R突变的TadA*7.10(TadA*8.6)的异二聚构建体。在一些实施方案中,ABE8是ABE8.7-7,其具有含有具有Q154R突变的TadA*7.10(TadA*8.7)的异二聚构建体。在一些实施方案中,ABE8是ABE8.8-7,其具有含有TadA*7.10与具有Y147R、Q154R和Y123H突变的TadA*7.10(TadA*8.8)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.9-7,其具有含有TadA*7.10与具有Y147R、Q154R和I76Y突变的TadA*7.10(TadA*8.9)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.10-7,其具有含有TadA*7.10与具有Y147R、Q154R和T166R突变的TadA*7.10(TadA*8.10)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.11-7,其具有含有TadA*7.10与具有Y147T和Q154R突变的TadA*7.10(TadA*8.11)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.12-7,其具有含有TadA*7.10与具有Y147T和Q154S突变的TadA*7.10(TadA*8.12)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.13-7,其具有含有TadA*7.10与具有Y123H(Y123H从H123Y恢复而来)、Y147R、Q154R和I76Y突变的TadA*7.10(TadA*8.13)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.14-7,其具有含有TadA*7.10与具有I76Y和V82S突变的TadA*7.10(TadA*8.14)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.15-7,其具有含有TadA*7.10与具有V82S和Y147R突变的TadA*7.10(TadA*8.15)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.16-7,其具有含有TadA*7.10与具有V82S、Y123H(Y123H从H123Y恢复而来)和Y147R突变的TadA*7.10(TadA*8.16)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.17-7,其具有含有TadA*7.10与具有V82S和Q154R突变的TadA*7.10(TadA*8.17)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.18-7,其具有含有TadA*7.10与具有V82S、Y123H(Y123H从H123Y恢复而来)和Q154R突变的TadA*7.10(TadA*8.18)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.19-7,其具有含有TadA*7.10与具有V82S、Y123H(Y123H从H123Y恢复而来)、Y147R和Q154R突变的TadA*7.10(TadA*8.19)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.20-7,其具有含有TadA*7.10与具有I76Y、V82S、Y123H(Y123H从H123Y恢复而来)、Y147R和Q154R突变的TadA*7.10(TadA*8.20)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.21-7,其具有含有TadA*7.10与具有Y147R和Q154S突变的TadA*7.10(TadA*8.21)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.22-7,其具有含有TadA*7.10与具有V82S和Q154S突变的TadA*7.10(TadA*8.22)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.23-7,其具有含有TadA*7.10与具有V82S和Y123H(Y123H从H123Y恢复而来)突变的TadA*7.10(TadA*8.23)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8.24-7,其具有含有TadA*7.10与具有V82S、Y123H(Y123H从H123Y恢复而来)和y147T突变的TadA*7.10(TadA*8.24)融合的异二聚

构建体。

[1675] 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13中所示。

[1676] 表13:腺苷脱氨酶碱基编辑器8(ABE8)变体

[1677]

ABE8	腺苷脱氨酶	腺苷脱氨酶说明
ABE8.1-m	TadA*8.1	单体_TadA*7.10 + Y147T
ABE8.2-m	TadA*8.2	单体_TadA*7.10 + Y147R
ABE8.3-m	TadA*8.3	单体_TadA*7.10 + Q154S
ABE8.4-m	TadA*8.4	单体_TadA*7.10 + Y123H
ABE8.5-m	TadA*8.5	单体_TadA*7.10 + V82S
ABE8.6-m	TadA*8.6	单体_TadA*7.10 + T166R
ABE8.7-m	TadA*8.7	单体_TadA*7.10 + Q154R
ABE8.8-m	TadA*8.8	单体_TadA*7.10 + Y147R_Q154R_Y123H
ABE8.9-m	TadA*8.9	单体_TadA*7.10 + Y147R_Q154R_I76Y
ABE8.10-m	TadA*8.10	单体_TadA*7.10 + Y147R_Q154R_T166R
ABE8.11-m	TadA*8.11	单体_TadA*7.10 + Y147T_Q154R
ABE8.12-m	TadA*8.12	单体_TadA*7.10 + Y147T_Q154S
ABE8.13-m	TadA*8.13	单体_TadA*7.10 + Y123H_Y147R_Q154R_I76Y
ABE8.14-m	TadA*8.14	单体_TadA*7.10 + I76Y_V82S
ABE8.15-m	TadA*8.15	单体_TadA*7.10 + V82S_Y147R
ABE8.16-m	TadA*8.16	单体_TadA*7.10 + V82S_Y123H_Y147R
ABE8.17-m	TadA*8.17	单体_TadA*7.10 + V82S_Q154R
ABE8.18-m	TadA*8.18	单体_TadA*7.10 + V82S_Y123H_Q154R
ABE8.19-m	TadA*8.19	单体_TadA*7.10 + V82S_Y123H_Y147R_Q154R
ABE8.20-m	TadA*8.20	单体_TadA*7.10 + I76Y_V82S_Y123H_Y147R_Q154R
ABE8.21-m	TadA*8.21	单体_TadA*7.10 + Y147R_Q154S
ABE8.22-m	TadA*8.22	单体_TadA*7.10 + V82S_Q154S

[1678]

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	异二聚体_(WT) + (TadA*7.10 + Y147T)
ABE8.2-d	TadA*8.2	异二聚体_(WT) + (TadA*7.10 + Y147R)
ABE8.3-d	TadA*8.3	异二聚体_(WT) + (TadA*7.10 + Q154S)
ABE8.4-d	TadA*8.4	异二聚体_(WT) + (TadA*7.10 + Y123H)
ABE8.5-d	TadA*8.5	异二聚体_(WT) + (TadA*7.10 + V82S)
ABE8.6-d	TadA*8.6	异二聚体_(WT) + (TadA*7.10 + T166R)
ABE8.7-d	TadA*8.7	异二聚体_(WT) + (TadA*7.10 + Q154R)
ABE8.8-d	TadA*8.8	异二聚体_(WT) + (TadA*7.10 + Y147R_Q154R_Y123H)
ABE8.9-d	TadA*8.9	异二聚体_(WT) + (TadA*7.10 + Y147R_Q154R_I76Y)
ABE8.10-d	TadA*8.10	异二聚体_(WT) + (TadA*7.10 + Y147R_Q154R_T166R)
ABE8.11-d	TadA*8.11	异二聚体_(WT) + (TadA*7.10 + Y147T_Q154R)
ABE8.12-d	TadA*8.12	异二聚体_(WT) + (TadA*7.10 + Y147T_Q154S)
ABE8.13-d	TadA*8.13	异二聚体_(WT) + (TadA*7.10 + Y123H_Y147T_Q154R_I76Y)
ABE8.14-d	TadA*8.14	异二聚体_(WT) + (TadA*7.10 + I76Y_V82S)
ABE8.15-d	TadA*8.15	异二聚体_(WT) + (TadA*7.10 + V82S_ Y147R)
ABE8.16-d	TadA*8.16	异二聚体_(WT) + (TadA*7.10 + V82S_Y123H_Y147R)
ABE8.17-d	TadA*8.17	异二聚体_(WT) + (TadA*7.10 + V82S_Q154R)
ABE8.18-d	TadA*8.18	异二聚体_(WT) + (TadA*7.10 + V82S_Y123H_Q154R)
ABE8.19-d	TadA*8.19	异二聚体_(WT) + (TadA*7.10 + V82S_Y123H_Y147R_Q154R)
ABE8.20-d	TadA*8.20	异二聚体_(WT) + (TadA*7.10 + I76Y_V82S_Y123H_Y147R_Q154R)
ABE8.21-d	TadA*8.21	异二聚体_(WT) + (TadA*7.10 + Y147R_Q154S)
ABE8.22-d	TadA*8.22	异二聚体_(WT) + (TadA*7.10 + V82S_Q154S)
ABE8.23-d	TadA*8.23	异二聚体_(WT) + (TadA*7.10 + V82S_Y123H)
ABE8.24-d	TadA*8.24	异二聚体_(WT) + (TadA*7.10 + V82S_Y123H_Y147T)

[1679] 在一些实施方案中,ABE8是ABE8a-m,其具有含有具有R26C、A109S、T111R、D119N、H122N、Y147D、F149Y、T166I和D167N突变的TadA*7.10(TadA*8a)的单体构建体。在一些实施方案中,ABE8是ABE8b-m,其具有含有具有V88A、A109S、T111R、D119N、H122N、F149Y、T166I和D167N突变的TadA*7.10(TadA*8b)的单体构建体。在一些实施方案中,ABE8是ABE8c-m,其具有含有具有R26C、A109S、T111R、D119N、H122N、F149Y、T166I和D167N突变的TadA*7.10(TadA*8c)的单体构建体。在一些实施方案中,ABE8是ABE8d-m,其具有含有具有V88A、T111R、D119N和F149Y突变的TadA*7.10(TadA*8d)的单体构建体。在一些实施方案中,ABE8是ABE8e-m,其具有含有具有A109S、T111R、D119N、H122N、Y147D、F149Y、T166I和D167N突变的TadA*7.10(TadA*8e)的单体构建体。

[1680] 在一些实施方案中,ABE8是ABE8a-d,其具有含有野生型大肠杆菌TadA与具有R26C、A109S、T111R、D119、H122N、Y147D、F149Y、T166I和D167N突变的TadA*7.10(TadA*8a)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8b-d,其具有含有野生型大肠杆菌TadA与具有V88A、A109S、T111R、D119N、H122N、F149Y、T166I和D167N突变的TadA*7.10(TadA*8b)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8c-d,其具有含有野生型大肠杆菌TadA与具有R26C、A109S、T111R、D119N、H122N、F149Y、T166I和D167N突变的TadA*7.10(TadA*8c)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8d-d,其具有含有野生型大肠杆菌TadA与具有V88A、T111R、D119N和F149Y突变的TadA*7.10(TadA*8d)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8e-d,其具有含有野生型大肠杆菌TadA与具有

A109S、T111R、D119N、H122N、Y147D、F149Y、T166I和D167N突变的TadA*7.10(TadA*8e)融合的异二聚构建体。

[1681] 在一些实施方案中,ABE8是ABE8a-7,其具有含有TadA*7.10与具有R26C、A109S、T111R、D119、H122N、Y147D、F149Y、T166I和D167N突变的TadA*7.10(TadA*8a)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8b-7,其具有含有TadA*7.10与具有V88A、A109S、T111R、D119N、H122N、F149Y、T166I和D167N突变的TadA*7.10(TadA*8b)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8c-7,其具有含有TadA*7.10与具有R26C、A109S、T111R、D119N、H122N、F149Y、T166I和D167N突变的TadA*7.10(TadA*8c)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8d-7,其具有含有TadA*7.10与具有V88A、T111R、D119N和F149Y突变的TadA*7.10(TadA*8d)融合的异二聚构建体。在一些实施方案中,ABE8是ABE8e-7,其具有含有TadA*7.10与具有A109S、T111R、D119N、H122N、Y147D、F149Y、T166I和D167N突变的TadA*7.10(TadA*8e)融合的异二聚构建体。

[1682] 在一些实施方案中,ABE是ABE8a-m、ABE8b-m、ABE8c-m、ABE8d-m、ABE8e-m、ABE8a-d、ABE8b-d、ABE8c-d、ABE8d-d或ABE8e-d,如下表14中所示。在一些实施方案中,ABE是ABE8e-m或ABE8e-d。在与除SpCas9以外的Cas同源物例如SaCas9、SaCas9-KKH、Cas12a同源物(例如,LbCas12a、enAs-Cas12a、SpCas9-NG)和环状排列的CP1028-SpCas9和CP1041-SpCas9合用时,ABE8e显示有效的腺嘌呤碱基编辑活性和低的插入缺失形成。除了表14中针对ABE8e显示的突变以外,通过将V106W置换引入TadA结构域内,降低了脱靶RNA和DNA编辑(如M.Richter等人,2020,Nature Biotechnology,doi.org/10.1038/s41587-020-0453-z中所述,其整体内容通过引用并入本文)。

[1683] 表14;其他腺苷脱氨酶碱基编辑器8变体

[1684]

ABE8 碱基编辑器	腺苷脱氨酶	腺苷脱氨酶说明
ABE8a-m	TadA*8a	单体_TadA*7.10 + R26C + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N
ABE8b-m	TadA*8b	单体_TadA*7.10 + V88A + A109S + T111R + D119N + H122N + F149Y + T166I + D167N
ABE8c-m	TadA*8c	单体_TadA*7.10 + R26C + A109S + T111R + D119N + H122N + F149Y + T166I + D167N
ABE8d-m	TadA*8d	单体_TadA*7.10 + V88A + T111R + D119N + F149Y
ABE8e-m	TadA*8e	单体_TadA*7.10 + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N
ABE8a-d	TadA*8a	异二聚体_(WT) + (TadA*7.10 + R26C + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N)
ABE8b-d	TadA*8b	异二聚体_(WT) + (TadA*7.10 + V88A + A109S + T111R + D119N + H122N + F149Y + T166I + D167N)
ABE8c-d	TadA*8c	异二聚体_(WT) + (TadA*7.10 + R26C + A109S + T111R + D119N + H122N + F149Y + T166I + D167N)
ABE8d-d	TadA*8d	异二聚体_(WT) + (TadA*7.10 + V88A + T111R + D119N + F149Y)
ABE8e-d	TadA*8e	异二聚体_(WT) + (TadA*7.10 + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N)

[1685] 在一些实施方案中,通过将腺苷脱氨酶变体(例如,TadA*8)克隆到支架中来生成碱基编辑器(例如,ABE8),该支架包括环状排列的Cas9(例如,CP5或CP6)和二分体核定位序列。在一些实施方案中,碱基编辑器(例如,ABE7.9、ABE7.10或ABE8)是NGC PAM CP5变体(酿

脓链球菌Cas9或spVRQR Cas9)。在一些实施方案中,碱基编辑器(例如,ABE7.9、ABE7.10或ABE8)是AGA PAM CP5变体(酿脓链球菌Cas9或spVRQR Cas9)。在一些实施方案中,碱基编辑器(例如,ABE7.9、ABE7.10或ABE8)是NGC PAM CP6变体(酿脓链球菌Cas9或spVRQR Cas9)。在一些实施方案中,碱基编辑器(例如,ABE7.9、ABE7.10或ABE8)是AGA PAM CP6变体(酿脓链球菌Cas9或spVRQR Cas9)。

[1686] 在一些实施方案中,ABE具有如下表15中所示的基因型。

[1687] 表15.ABE的基因型

[1688]

	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

[1689] 如下表16中所示,描述了40种ABE8的基因型。指出了ABE的在进化的大肠杆菌TadA部分中的残基位置。当不同于ABE7.10突变时,显示了ABE8中的突变改变。在一些实施方案中,ABE具有如下表16中所示ABE中的一个的基因型。

[1690] 表16.进化的TadA中的残基身份

[1691]

	23	36	48	51	76	82	84	106	108	123	146	147	152	154	155	156	157	166
ABE7.10	R	L	A	L	I	V	F	V	N	Y	C	Y	P	Q	V	F	N	T
ABE8.1-m												T						
ABE8.2-m												R						
ABE8.3-m														S				
ABE8.4-m										H								
ABE8.5-m						S												
ABE8.6-m																		R
ABE8.7-m														R				
ABE8.8-m										H		R		R				
ABE8.9-m					Y							R		R				
ABE8.10-m												R		R				R
ABE8.11-m												T		R				
ABE8.12-m												T		S				
ABE8.13-m					Y					H		R		R				
ABE8.14-m					Y	S												
ABE8.15-m						S						R						
ABE8.16-m						S				H		R						
ABE8.17-m						S								R				
ABE8.18-m						S				H				R				
ABE8.19-m						S				H		R		R				
ABE8.20-m					Y	S				H		R		R				
ABE8.21-m												R		S				
ABE8.22-m						S								S				
ABE8.23-m						S				H								
ABE8.24-m						S				H		T						
ABE8.1-d												T						
ABE8.2-d												R						
ABE8.3-d														S				
ABE8.4-d										H								
ABE8.5-d						S												
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				

[1692]

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						

[1693] 在一些实施方案中,碱基编辑器是ABE8.1,其包含以下序列或其具有腺苷脱氨酶活性的片段或基本上由其组成:

[1694] ABE8.1_Y147T_CP5_NGC PAM_单体

[1695] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAAL
LCTFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSEIGKATAKYFFYSNIMNFFKT
EITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKK
DWDPKKYGGFMQPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAKFLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF
SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYRSTKEVLDATLIHQSI
TGLYETRIDLSQLGGDGGSGGSGGSGGSGGSGGSGGSGGMDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDR
HSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSK
DTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLP
EKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKI
ECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVM
KQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHI
ANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVV
KKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE
VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQE
GADKRTADGSEFESPKKKRKV*

[1696] 在上述序列中,纯文本表示腺苷脱氨酶结构域,粗体序列指示源自Cas9的序列,斜体序列表示连接子序列,且加下划线序列表示二分体核定位序列。

[1697] 在一些实施方案中,碱基编辑器是ABE8.1,其包含以下序列或其具有腺苷脱氨酶活性的片段或基本上由其组成:

[1698] pNMG-B335 ABE8.1_Y147T_CP5_NGC PAM_单体

[1699] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAAL
LCTFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSEIGKATAKYFFYSNIMNFFKT
EITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKK
DWDPKKYGGFMQPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAKFLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF
SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYRSTKEVLDATLIHQSI
TGLYETRIDLSQLGGDGGSGGSGGSGGSGGSGGSGGMDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDR
HSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSK
DTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLP
EKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKI
ECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVM
KQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHI
ANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVV
KKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE
VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQE
GADKRTADGSEFESPKKKRKV*

[1700] 在上述序列中,纯文本表示腺苷脱氨酶结构域,粗体序列指示源自Cas9的序列,斜体序列表示连接子序列,且加下划线序列表示二分体核定位序列。

[1701] 在一些实施方案中,碱基编辑器是ABE8.14,其包含以下序列或其具有腺苷脱氨酶活性的片段或基本上由其组成:

[1702] 具有NGC PAM CP5的pNMG-357_ABE8.14

[1703] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAAL
LSDFFRMRRQEIKAQKKAQSSTDGGSSGGSSGSETPGTSESATPESSGGSSGGSMSEVEFSHEYWMRHALTLAKRA
RDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSR
IGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTDSGGSSGG
SSGSETPGTSESATPESSGGSSGGSEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGR
DFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPTVAYSVLVVAKVEKGKS
KKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAKFLQKGNELALPS
KYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQA
ENIIHLFTLTNLGAPRAFKYFDTTIARKEYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGSGGSGGSGGSGG
SGGSGGMDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR
RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST

DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL
ENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSD
AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYK
FIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTK
VKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIK
DKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKT
ILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRH
KPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI
NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG
GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYH
HAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEGADKRTADGSEFESPKKKRKV*

[1704] 在上述序列中,纯文本表示腺苷脱氨酶结构域,粗体序列指示源自Cas9的序列,斜体序列表示连接子序列,且加下划线序列表示二分体核定位序列。

[1705] 在一些实施方案中,碱基编辑器是ABE8.8-m,其包含以下序列或其具有腺苷脱氨酶活性的片段或基本上由其组成:

[1706] ABE8.8-m

[1707]
MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAE
IMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLM
DVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSG*
*SETPGTSESATPESSGGSSGGS*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNT
DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSF
FHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYL
ALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSAR
LSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDD
LDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVK
LNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYY
VGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPK
HSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYF
KKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFED
REMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG
FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDE
LVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQ
LQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNR

GKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVE
TRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYH
HAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSN
IMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV

[1708] QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKS
VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG
ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRY
TSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[1709] 在上述序列中,纯文本表示腺苷脱氨酶结构域,粗体序列指示源自Cas9的序列,斜体序列表示连接子序列,加下划线序列表示二分体核定位序列,且加双下划线序列指示突变。

[1710] 在一些实施方案中,碱基编辑器是ABE8.8-d,其包含以下序列或其具有腺苷脱氨酶活性的片段或基本上由其组成:

[1711] ABE8.8-d

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAE
IMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLM
DVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD*SGGSSGGSSG*
*SETPGTSESATPESSGGSSGGS*SEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNN
RVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHS
RIGRVVFGVRNAKTGAAGSLMDVLH<u>H</u>PGMNHRVEITEGILADECAALLC<u>R</u>FFRMPR<u>R</u>VF
NAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGS*DKKYSIGL<u>A</u>IGTNSVG
WAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRK
NRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI

[1712] YHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQ
LFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFK
SNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEI
TKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEE
FYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFY
PFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAI
VDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFL
DNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKL
INGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA

[1713]

NLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRI
EEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVP
QSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTK
AERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSK
LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK
MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRD
FATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT
VAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIK
LPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQ
LFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTL
TNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKR
TADGSEFESPKKKRKV*

[1714] 在上述序列中,纯文本表示腺苷脱氨酶结构域,粗体序列指示源自Cas9的序列,斜体序列表示连接子序列,加下划线序列表示二分体核定位序列,且加双下划线序列指示突变。

[1715] 在一些实施方案中,碱基编辑器是ABE8.13-m,其包含以下序列或其具有腺苷脱氨酶活性的片段或基本上由其组成:

[1716] ABE8.13-m

[1717]

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAE
IMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLM
DVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSG*
*SETPGTSESATPESSGGSSGGS*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNT
DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSF
FHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYL
ALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSAR
LSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDD
LDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVK
LNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYY
VGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPK
HSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYF
KKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFED
REMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG
FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDE

LVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQ

LQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNR

GKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVE

TRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYH

HAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSN

[1718] IMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV

QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKS

VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG

ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS

KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRY

TSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[1719] 在上述序列中,纯文本表示腺苷脱氨酶结构域,粗体序列指示源自Cas9的序列,斜体序列表示连接子序列,加下划线序列表示二分体核定位序列,且加双下划线序列指示突变。

[1720] 在一些实施方案中,碱基编辑器是ABE8.13-d,其包含以下序列或其具有腺苷脱氨酶活性的片段或基本上由其组成:

[1721] ABE8.13-d

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAE

IMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLM

DVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD*SGGSSGGSSG*

*SETPGTSESATPESSGGSSGGS*SEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNN

RVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHS

RIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVF

NAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGS*DKKYSIGLAIGTNSVG

WAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRK

[1722] NRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI

YHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQ

LFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFK

SNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEI

TKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEE

FYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFY

PFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ

SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAI

VDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFL

DNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKL
INGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA
NLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRI
EEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVP
QSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTK
AERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSK
LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK
[1723] MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRD
FATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT
VAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIK
LPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQ
LFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTL
TNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKR
TADGSEFESPKKKRKV*

[1724] 在上述序列中,纯文本表示腺苷脱氨酶结构域,粗体序列指示源自Cas9的序列,斜体序列表示连接子序列,加下划线序列表示二分体核定位序列,且加双下划线序列指示突变。

[1725] 在一些实施方案中,碱基编辑器是ABE8.17-m,其包含以下序列或其具有腺苷脱氨酶活性的片段或基本上由其组成:

[1726] ABE8.17-m

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAE
IMALRQGGLVMQNYRLIDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLM
DVLHYPGMNHRVEITEGILADECAALLCYFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSG*
*SETPGTSESATPESSGGSSGGS*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNT
DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSF
FHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYL
ALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSAR
[1727] LSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDD
LDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVK
LNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYY
VGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPK
HSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYF
KKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFED

REMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG
FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDE
LVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQ
LQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNR
GKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVE
TRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYH

[1728] HAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSN
IMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKS
VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG
ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRY
TSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[1729] 在上述序列中,纯文本表示腺苷脱氨酶结构域,粗体序列指示源自Cas9的序列,斜体序列表示连接子序列,加下划线序列表示二分体核定位序列,且加双下划线序列指示突变。

[1730] 在一些实施方案中,碱基编辑器是ABE8.17-d,其包含以下序列或其具有腺苷脱氨酶活性的片段或基本上由其组成:

[1731] ABE8.17-d

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAE
IMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLM
DVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD*SGGSSGGSSG*
*SETPGTSESATPESSGGSSGGS*SEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNN
RVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYSTFEPCVMCAGAMIHS
RIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRRVF
NAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGS*DKKYSIGLAIGTNSVG

[1732] WAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRK
NRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI
YHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQ
LFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFK
SNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEI
TKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEE
FYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFY
PFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ

SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAI
VDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFL
DNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKL
INGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA
NLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRI
EEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVP
QSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTK
AERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSK
[1733]
LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK
MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRD
FATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT
VAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIK
LPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQ
LFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTL
TNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKR
TADGSEFESPKKKRKV*

[1734] 在上述序列中,纯文本表示腺苷脱氨酶结构域,粗体序列指示源自Cas9的序列,斜体序列表示连接子序列,加下划线序列表示二分体核定位序列,且加双下划线序列指示突变。

[1735] 在一些实施方案中,碱基编辑器是ABE8.20-m,其包含以下序列或其具有腺苷脱氨酶活性的片段或基本上由其组成:

[1736] ABE8.20-m

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAE
IMALRQGGLVMQNYRLYDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLM
DVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSG*
*SETPGTSESATPESSGGSSGGS*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNT
DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSF
FHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYL
[1737]
ALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSAR
LSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDD
LDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVK
LNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYY
VGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPK

HSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYF
KKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFED
REMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG
FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDE
LVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQ
LQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNR
GKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVE

[1738] TRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYH
HAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSN
IMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKS
VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG
ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRY
TSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[1739] 在上述序列中,纯文本表示腺苷脱氨酶结构域,粗体序列指示源自Cas9的序列,斜体序列表示连接子序列,加下划线序列表示二分体核定位序列,且加双下划线序列指示突变。

[1740] 在一些实施方案中,碱基编辑器是ABE8.20-d,其包含以下序列或其具有腺苷脱氨酶活性的片段或基本上由其组成:

[1741] ABE8.20-d

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALR
QGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVE
ITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGG*
*S*SEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG
LVMQNYRLYDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITE
GILADECAALLCRFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGS*DK
KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR
RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH
LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

[1742] GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY
DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL
VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET
ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF
LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF
LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK
QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT
VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ

[1743] NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT
KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF
VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD
KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYS
VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG
RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF
SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL
DATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[1744] 在上述序列中,纯文本表示腺苷脱氨酶结构域,粗体序列指示源自Cas9的序列,斜体序列表示连接子序列,加下划线序列表示二分体核定位序列,且加双下划线序列指示突变。

[1745] 在一些实施方案中,本发明的ABE8选自以下序列:01.單ABE8.1_bpNLS+Y147T

[1746] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
TFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD
SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[1747] 02.單ABE8.1_bpNLS+Y147R

[1748] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
RFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS

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

[1749] 03.單ABE8.1_bpNLS+Q154S

[1750] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
YFFRMPRSVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD
SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[1751] 04.單ABE8.1_bpNLS+Y123H

[1752] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLC
YFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD
SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[1753] 05.單ABE8.1_bpNLS+V82S

[1754] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
YFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD
SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG

GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[1755] 06.單ABE8.1_bpNLS+T166R

[1756] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
YFFRMPRQVFNAQKKAQSSRDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD
SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[1757] 07.單ABE8.1_bpNLS+Q154R

[1758] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
YFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD

SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[1759] 08.單ABE8.1_bpNLS+Y147R_Q154R_Y123H

[1760] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLC
RFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD
SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[1761] 09.單ABE8.1_bpNLS+Y147R_Q154R_I76Y

[1762] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
RFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT

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

[1763] 10.單ABE8.1_bpNLS+Y147R_Q154R_T166R

[1764] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
RFFRMPRRVFNAQKKAQSSRDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD
SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[1765] 11.單ABE8.1_bpNLS+Y147T_Q154R

[1766] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC

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

[1767] 12.單ABE8.1_bpNLS+Y147T_Q154S

[1768] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
TFFRMPRSVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD
SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL

FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[1769] 13.單ABE8.1_bpNLS+H123Y123H_Y147R_Q154R_I76Y

[1770] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLC
RFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD
SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[1771] 14.單ABE8.1_bpNLS+V82S+Q154R

[1772] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLC
YFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD
SLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS

EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID
FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
QYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[1773] 在一些实施方案中,碱基编辑器是第九代ABE(ABE9)。在一些实施方案中,ABE9含有TadA*9变体。ABE9碱基编辑器包括腺苷脱氨酶变体,该腺苷脱氨酶变体包含氨基酸序列,该氨基酸序列含有相对于ABE 7*10参考序列的改变,如本文所述。如表19中所用,术语“单体”是指单体形式的TadA*7.10,其包含如表19中所述的改变。如表19中所用,术语“异二聚体”是指所指定野生型大肠杆菌TadA腺苷脱氨酶融合至包含表19中所述改变的TadA*7.10,如本文所述。ABE9碱基编辑器的细节在国际PCT申请号PCT/2020/049975中描述,其通过引用以其整体并入本文。

[1774] 表17.腺苷脱氨酶碱基编辑器9(ABE9)变体

[1775]

ABE9 说明	改变
ABE9.1_单体	E25F、V82S、Y123H、T133K、Y147R、Q154R
ABE9.2_单体	E25F、V82S、Y123H、Y147R、Q154R
ABE9.3 _单体	V82S、Y123H、P124W、 Y147R、Q154R
ABE9.4_单体	L51W、V82S、Y123H、C146R、Y147R、Q154R
ABE9.5_单体	P54C、V82S、Y123H、Y147R、Q154R
ABE9.6_单体	Y73S、V82S、Y123H、Y147R、Q154R
ABE9.7_单体	N38G、V82T、Y123H、Y147R、Q154R
ABE9.8_单体	R23H、V82S、Y123H、Y147R、Q154R
ABE9.9_单体	R21N、V82S、Y123H、Y147R、Q154R
ABE9.10_单体	V82S、Y123H、Y147R、Q154R、A158K
ABE9.11_单体	N72K、V82S、Y123H、D139L、Y147R、Q154R,
ABE9.12_单体	E25F、V82S、Y123H、D139M、Y147R、Q154R
ABE9.13_单体	M70V、V82S、M94V、Y123H、Y147R、Q154R
ABE9.14_单体	Q71M、 V82S、Y123H、Y147R、Q154R
ABE9.15_异二聚体	E25F、V82S、Y123H、T133K、Y147R、Q154R
ABE9.16_异二聚体	E25F、V82S、Y123H、Y147R、Q154R
ABE9.17_异二聚体	V82S、Y123H、P124W、 Y147R、Q154R
ABE9.18_异二聚体	L51W、V82S、Y123H、C146R、Y147R、Q154R
ABE9.19_异二聚体	P54C、V82S、Y123H、Y147R、Q154R
ABE9.2 _异二聚体	Y73S、V82S、Y123H、Y147R、Q154R
ABE9.21_异二聚体	N38G、V82T、Y123H、Y147R、Q154R

ABE9.22_异二聚体	R23H、V82S、Y123H、Y147R、Q154R
ABE9.23_异二聚体	R21N、V82S、Y123H、Y147R、Q154R
ABE9.24_异二聚体	V82S、Y123H、Y147R、Q154R、A158K
ABE9.25_异二聚体	N72K、V82S、Y123H、D139L、Y147R、Q154R,
ABE9.26_异二聚体	E25F、V82S、Y123H、D139M、Y147R、Q154R
ABE9.27 _异二聚体	M70V、V82S、M94V、Y123H、Y147R、Q154R
ABE9.28_异二聚体	Q71M、 V82S、Y123H、Y147R、Q154R
ABE9.29_单体	E25F_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.30_单体	I76Y_V82T_Y123H_Y147R_Q154R
ABE9.31_单体	N38G_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.32_单体	N38G_I76Y_V82T_Y123H_Y147R_Q154R
ABE9.33_单体	R23H_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.34_单体	P54C_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.35_单体	R21N_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.36_单体	I76Y_V82S_Y123H_D138M_Y147R_Q154R
ABE9.37_单体	Y72S_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.38_异二聚体	E25F_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.39_异二聚体	I76Y_V82T_Y123H_Y147R_Q154R
ABE9.40_异二聚体	N38G_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.41_异二聚体	N38G_I76Y_V82T_Y123H_Y147R_Q154R
ABE9.42_异二聚体	R23H_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.43_异二聚体	P54C_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.44_异二聚体	R21N_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.45_异二聚体	I76Y_V82S_Y123H_D138M_Y147R_Q154R
ABE9.46_异二聚体	Y72S_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.47_单体	N72K_V82S、Y123H、Y147R、Q154R
ABE9.48_单体	Q71M_V82S、Y123H、Y147R、Q154R
ABE9.49_单体	M70V,V82S、M94V、Y123H、Y147R、Q154R
ABE9.50_单体	V82S、Y123H、T133K、Y147R、Q154R
ABE9.51_单体	V82S、Y123H、T133K、Y147R、Q154R、A158K
ABE9.52_单体	M70V,Q71M,N72K,V82S、Y123H、Y147R、Q154R
ABE9.53_异二聚体	N72K_V82S、Y123H、Y147R、Q154R
ABE9.54_异二聚体	Q71M_V82S、Y123H、Y147R、Q154R
ABE9.55_异二聚体	M70V,V82S、M94V、Y123H、Y147R、Q154R
ABE9.56_异二聚体	V82S、Y123H、T133K、Y147R、Q154R
ABE9.57_异二聚体	V82S、Y123H、T133K、Y147R、Q154R、A158K
ABE9.58_异二聚体	M70V、Q71M、N72K、V82S、Y123H、Y147R、Q154R

[1776]

[1777] 在一些实施方案中,碱基编辑器是融合蛋白,该融合蛋白包含多核苷酸可编程核苷酸结合结构域(例如,源自Cas9的结构域)融合至核碱基编辑结构域(例如,脱氨酶结构域的全部或一部分)。在某些实施方案中,本文所提供的融合蛋白包含一个或多个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包含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没有核酸酶活性的Cas9结构域(dCas9),或切断双螺旋DNA分子的一条链的Cas9结构域,称为Cas9切口酶(nCas9)。

[1778] 在一些实施方案中,碱基编辑器进一步包含一个结构域,该结构域包含尿嘧啶糖基化酶抑制剂(UGI)的全部或一部分。在一些实施方案中,碱基编辑器包含结构域,该结构域包含尿嘧啶结合蛋白(UBP)诸如尿嘧啶DNA糖基化酶(UDG)的全部或一部分。在一些实施方案中,碱基编辑器包含结构域,该结构域包含核酸聚合酶的全部或一部分。在一些实施方

案中,碱基编辑器可包含核酸聚合酶(NAP)的全部或一部分作为结构域。例如,碱基编辑器可包含真核NAP的全部或一部分。在一些实施方案中,并入碱基编辑器内的NAP或其部分是DNA聚合酶。在一些实施方案中,并入碱基编辑器内的NAP或其部分具有跨损伤聚合酶活性。在一些实施方案中,并入碱基编辑器内的NAP或其部分是跨损伤DNA聚合酶。在一些实施方案中,并入碱基编辑器内的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,并入碱基编辑器内的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,并入碱基编辑器内的NAP或其部分包含氨基酸序列,该氨基酸序列与核酸聚合酶(例如,跨损伤DNA聚合酶)至少75%、80%、85%、90%、95%、96%、97%、98%、99%或99.5%相同。在一些实施方案中,并入碱基编辑器内的核酸聚合酶或其部分是跨损伤DNA聚合酶。

[1779] 在一些实施方案中,碱基编辑器的结构域可包含多个结构域。例如,包含源自Cas9的多核苷酸可编程核苷酸结合结构域的碱基编辑器可包含对应于野生型或天然Cas9的REC叶和NUC叶的REC叶和NUC叶。在另一示例中,碱基编辑器可包含RuvCI结构域、BH结构域、REC1结构域、REC2结构域、RuvCII结构域、L1结构域、HNH结构域、L2结构域、RuvCIII结构域、WED结构域、TOPO结构域或CTD结构域中的一者或多者。在一些实施方案中,碱基编辑器的一个或多个结构域包含相对于包含该结构域的野生型版本的多核苷酸的突变(例如,置换、插入、缺失)。例如,多核苷酸可编程DNA结合结构域的HNH结构域可包含H840A置换。在另一示例中,多核苷酸可编程DNA结合结构域的RuvCI结构域可包含D10A置换。

[1780] 本文所公开的碱基编辑器的不同结构域(例如,相邻结构域)可以使用或不适用一个或多个连接子结构域(例如,XTEN连接子结构域)彼此连结。在一些实施方案中,连接子结构域可以是键(例如,共价键)、化学基团或分子,其连接两个分子或部分,例如,融合蛋白的两个结构域,例如,第一结构域(例如,源自Cas9的结构域)和第二结构域(例如,腺苷脱氨酶结构域或胞苷脱氨酶结构域)。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,连接子是酰胺链接的碳氮键。在某些实施方案中,连接子是环状的或非环状的、取代的或未取代的、分支的或不分支的脂肪族或杂脂族连接子。在某些实施方案中,连接子是聚合性的(例如,聚乙烯、聚乙二醇、聚酰胺、聚酯等)。在某些实施方案中,连接子包含氨基烷酸的单体、二聚体、或聚合物。在一些实施方案中,连接子包含氨基烷酸(例如,甘氨酸、乙酸、丙氨酸、β-丙氨酸、3-氨基丙酸、4-氨基丁酸、5-戊酸等)。在一些实施方案中,连接子包含氨基己酸(Ahx)的单体、二聚体或聚合物。在某些实施方案中,连接子基于碳环部分(例如,环戊烷、环己烷)。在其他实施方案中,连接子包含聚乙二醇部分(PEG)。在某些实施方案中,连接子包含芳基或杂芳基部分。在某些实施方案中,连接子基于苯基环。连接子可包括经功能化的部分以促使来自肽的亲核基团(例如,巯基、氨基)附接至连接子。任何亲电基团可用作连接子的一部分。示例性亲电基团包括但不限于,活化的酯、Michael受体、烷基卤化物、芳基卤化物和异硫氰酸酯。在一些实施方案中,连接子接合RNA可编程核酸酶的gRNA接合结构域(包括Cas9核酸酶结构域)与核酸编辑蛋白的催化性结构域。在一些实施方案中,连接子接合dCas9与第二结构域(例如,UGI等)。

[1781] 连接子

[1782] 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。连接子可以简单到共价键,或者可以是很多个原子长度的聚合性连接子。在某些实施方案中,连接子是

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,连接子是酰胺链接的碳氮键。在某些实施方案中,连接子是环状的或非环状的、取代的或未取代的、分支的或不分支的脂肪族或杂脂族连接子。在某些实施方案中,连接子是聚合性的(例如,聚乙烯、聚乙二醇、聚酰胺、聚酯等)。在某些实施方案中,连接子包含氨基烷酸的单体、二聚体、或聚合物。在某些实施方案中,连接子包含氨基烷酸(例如,甘氨酸、乙酸、丙氨酸、β-丙氨酸、3-氨基丙酸、4-氨基丁酸、5-戊酸等)。在某些实施方案中,连接子包含氨基己酸(Ahx)的单体、二聚体、或聚合物。在某些实施方案中,连接子基于碳环部分(例如,环戊烷、环己烷)。在其他实施方案中,连接子包含聚乙二醇部分(PEG)。在其他实施方案中,连接子包含氨基酸。在某些实施方案中,连接子包含肽。在某些实施方案中,连接子包含芳基或杂芳基部分。在某些实施方案中,连接子基于苯基环。连接子可包括经功能化的部分以促使来自肽的亲核基团(例如,巯基、氨基)附接至连接子。任何亲电基团可用作连接子的一部分。示例性亲电基团包括但不限于,活化的酯、Michael受体、烷基卤化物、芳基卤化物和异硫氰酸酯。

[1783] 典型地,连接子位于两个基团、分子或其他部分之间或连接子侧翼具有两个基团、分子或其他部分,并且彼此经由共价键联接,从而将两者联接。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,连接子是2至100个氨基酸的长度,例如,2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、30、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,连接子是约3至约104(例如,5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、55、60、65、70、75、80、85、90、95或100)个氨基酸的长度。也考虑更长或更短的连接子。

[1784] 在一些实施方案中,本文所提供的任何融合蛋白包含胞苷或腺苷结构域和Cas9结构域,这些结构域经由连接子彼此融合。在胞苷或腺苷脱氨酶与Cas9结构域之间可采用任何连接子长度和柔性(例如,范围从非常柔性的$(GGGS)_n$、$(GGGGS)_n$和$(G)_n$形式的连接子到非常刚性的$(EAAAK)_n$、$(SGGS)_n$、SGSETPGTSESATPES形式的连接子(参见,例如,Guilinger JP,等人Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification.Nat.Biotechnol.2014;32(6):577-82;整体内容通过引用并入本文)和$(XP)_n$),以便实现用于胞苷或腺苷脱氨酶核碱基编辑器活性的最优长度。在一些实施方案中,n是1、2、3、4、5、6、7、8、9、10、11、12、13、14或15。在一些实施方案中,连接子包含$(GGS)_n$基序,其中n是1、3或7。在一些实施方案中,本文所提供的任何融合蛋白的胞苷脱氨酶或腺苷脱氨酶和Cas9结构域经由包含氨基酸序列SGSETPGTSESATPES的连接子融合,该连接子也称为XTEN连接子。在一些实施方案中,连接子包含多个脯氨酸残基并且是5至21、5至14、5至9、5至7个氨基酸的长度,例如,PAPAP、PAPAPA、PAPAPAP、PAPAPAPA、$P(AP)_4$、$P(AP)_7$、$P(AP)_{10}$(参见,例如,Tan J,Zhang F,Karcher D,Bock R.Engineering of high-precision base editors for site-specific single nucleotide replacement.Nat Commun.2019Jan 25;10(1):439;其整体内容通过引用并入本文)。此类富含脯氨酸的连接子也称为“刚性”连接子。

[1785] 在另一实施方案中,碱基编辑器系统包含一种组分(蛋白质),其与脱氨酶(DNA脱氨酶)例如腺苷或胞苷脱氨酶非共价地相互作用并且将腺苷或胞苷脱氨酶短暂地附着至靶标多核苷酸序列中的靶标核碱基以进行特异性编辑,且旁观者或靶标邻近效应在最低限度或降低。这种包括脱氨酶相互作用蛋白质的非共价系统和方法发挥作用以将DNA脱氨酶附着至特定基因组靶标核碱基并将上靶编辑与靶标邻近编辑事件解耦,因此增强对更精确的单碱基置换突变的实现。在一种实施方案中,脱氨酶相互作用蛋白结合至脱氨酶(例如,腺苷脱氨酶或胞苷脱氨酶)而不阻断或干扰脱氨酶的活性(催化)位点与靶标核碱基(例如,分别为腺苷或胞苷)衔接。诸如名为"MagnEdit"的系统包括系结至Cas9和gRNA复合物的相互作用蛋白并且可以附着经共表达的腺苷或胞苷脱氨酶(外源性的或内源性的)以编辑特异性基因组靶标位点,并且描述在McCann,J.等人,2020,"MagnEdit-interacting factors that recruit DNA-editing enzymes to single base targets,"Life-Science-Alliance,Vol.3,No.4(e201900606),(doi 10.26508/lsa.201900606)中,其内容通过引用以其整体并入本文。在一种实施方案中,DNA脱氨酶是本文所述的腺苷脱氨酶变体(例如,TadA*8)。

[1786] 在另一实施方案中,被称为"Suntag"的系统包括非共价相互作用组分,该组分用于将碱基编辑器的蛋白质(例如腺苷脱氨酶或胞苷脱氨酶)组分或其多个拷贝招募至多核苷酸靶标位点,以实现在该位点处的碱基编辑且邻近靶标编辑减少,例如,如Tanenbaum,M.E.等人,"A protein tagging system for signal amplification in gene expression and fluorescence imaging,"Cell.2014October 23;159(3):635-646.doi:10.1016/j.cell.2014.09.039中和Huang,Y.-H.等人,2017,"DNA epigenome editing using CRISPR-Cas SunTag-directed DNMT3A,"Genome Biol 18:176.doi:10.1186/s13059-017-1306-z中所述,其各自的内容通过引用以其整体并入本文。在一种实施方案中,DNA脱氨酶是本文所述的腺苷脱氨酶变体(例如,TadA*8)。

[1787] 核酸可编程DNA结合蛋白与向导RNA

[1788] 本文提供用于在宿主细胞例如免疫细胞(例如,T细胞或NK细胞)中进行碱基编辑的组合物和方法。本文进一步提供组合物,其包含向导多核酸序列,例如,向导RNA序列,或本文所提供的2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20或更多个RNA的组合。在一些实施方案中,本文所提供的用于碱基编辑组合物进一步包含编码碱基编辑器例如C碱基编辑器或A碱基编辑器的多核苷酸。例如,用于碱基编辑的组合物可包含编码BE、BE4、ABE的mRNA以及所提供的一种或多种向导RNA的组合。用于碱基编辑的组合物可包含碱基编辑器多肽以及本文所提供的一种或多种本文所提供的任何向导RNA的组合。该组合物可用来通过不同递送途径例如电穿孔、核转染、病毒转导或转染在免疫细胞中实现碱基编辑。在一些实施方案中,用于碱基编辑的组合物包含编码碱基编辑器的mRNA以及本文提供的一种或多种向导RNA序列的组合用于电穿孔。

[1789] 本公开的一些方面提供复合物,其包含本文所提供的任何融合蛋白以及结合至融合蛋白的核酸可编程DNA结合蛋白(napDNAbp)结构域(例如,Cas9(例如,dCas9、核酸酶活性Cas9、或Cas9切口酶)或Cas12)的向导RNA。这些复合物也称为核糖核蛋白(RNP)。在一些实施方案中,向导核酸(例如,向导RNA)具有15至100个核苷酸的长度且包含与靶标序列互补的至少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包含与靶标序列互补的15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39或40个连续核苷酸的序列。在一些实施方案中,靶标序列是DNA序列。在一些实施方案中,靶标序列是RNA序列。在一些实施方案中,靶标序列是细菌、酵母、真菌、昆虫、植物或动物基因组中的序列。在一些实施方案中,靶标序列是人类基因组中的序列。在一些实施方案中,靶标序列的3'末端紧邻标准PAM序列(NGG)。在一些实施方案中,靶标序列的3'末端紧邻非标准PAM序列(例如,表5中所列的序列或5'-NAA-3')。在一些实施方案中,向导核酸(例如,向导RNA)与感兴趣的基因(例如,与疾病或疾患相关的基因)中的序列互补。

[1790] 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含以下的方法:使DNA分子与本文所提供的任何融合蛋白接触并且与至少一种向导RNA接触,其中向导RNA具有约15至100个核苷酸的长度并且包含与靶标序列互补的至少10个连续核苷酸的序列。在一些实施方案中,靶标序列的3'末端紧邻AGC、GAG、TTT、GTG或CAA序列。在一些实施方案中,靶标序列的3'末端紧邻NGA、NGCG、NGN、NNGRRT、NNNRRT、NGCG、NGCN、NGTN、NGTN、NGTN或5'(TTTV)序列。在一些实施方案中,靶标序列的3'末端紧邻例如TTN、DTTN、GTTN、ATTN、ATTC、DTTNT、WTTN、HATY、TTTN、TTTV、TTTC、TG、RTR或YTN PAM位点。

[1791] 应理解,相应序列中的具体位置或残基的编号取决于特定蛋白质和所使用的编号方案。编号可以不同,例如,在成熟蛋白质的前体和成熟蛋白质本身中,并且来自物种与物种之间的序列差异可能影响编号。本领域技术人员将能够通过本领域周知的方法鉴定任何同源蛋白质中和相应编码核酸中的相应残基,例如,通过序列比对并确定同源残基。

[1792] 对本领域技术人员将是显而易见的是,为了将本文所公开的任何融合蛋白靶向至靶标位点,例如,包含待编辑的突变的位点,通常必要的是将该融合蛋白与向导RNA共表达。如在本文其他处更详细阐述的,向导RNA通常包含允许napDNAbp(例如,Cas9或Cas12)与向导序列结合的tracrRNA框架,并且该向导序列赋予napDNAbp:核酸编辑酶/结构域融合蛋白以特异性。或者,向导RNA和tracrRNA可以作为两个核酸分子独立地提供。在一些实施方案中,向导RNA包含结构,其中向导序列包含与靶标序列互补的序列。向导序列通常是20个核苷酸的长度。基于本公开,适用于将napDNAbp:核酸编辑酶/结构域融合蛋白百香之特异性基因组位点的向导RNA的序列对于本领域技术人员将是显而易见的。此类合适的向导RNA序列通常包含与待编辑的靶标核苷酸上游或下游50个核苷酸内的核酸序列互补的向导序列。一些适用于将所提供的任何融合蛋白靶向至特异性靶标序列的示例性向导RNA提供在本文中。

[1793] sgRNA的不同部分预计会形成与Cas9(例如,SpyCas9)和/或DNA靶标相互作用的各种特征。已经在天然的native crRNA:tracrRNA双螺旋和单向导RNA(sgRNA)中鉴定出六个引导Cas9和us安内切酶活性的保守模块(参见,Briner等人,Guide RNA Functional Modules Direct Cas9 Activity and Orthogonality Mol Cell.20140ct23;56(2):333-339)。六个模块包括负责DNA靶向的间隔序列、上部茎、凸起、由CRISPR重复序列:tracrRNA双螺旋形成的下部茎、结点、和来自tracrRNA的3'末端的发夹圈。上部茎和下部茎主要通过与磷酸酯主链的序列依赖性相互作用与Cas9相互作用。在一些实施方案中,上部茎是可有

可无的。在一些实施方案中,在下部茎的基部的保守尿嘧啶核苷酸序列是可有可无的。凸起参与到与Cas9的Rec1结构域的特异性侧链相互作用中。核碱基U44与Tyr 325和His 328的侧链相互作用,而G43与Tyr 329相互作用。结点形成sgRNA:Cas9相互作用的核心,并且位于sgRNA与Cas9和靶标DNA两者之间的插入处。核碱基A51和A52与Phe 1105的侧链相互作用;U56与Arg 457和Asn 459相互作用;核碱基U59插入到由Arg 74、Asn 77、Pro 475、Leu 455、Phe 446和Ile 448的侧链定义的疏水袋中;C60与Leu 455、Ala 456和Asn 459相互作用,且C61与Arg 70的侧链相互通,继而与C15相互作用。在一些实施方案中,这些突变中的一个或多个在用于Cas9(例如,spyCas9)的sgRNA的凸起和/或结点中作出,以优化sgRNA:Cas9相互作用。

[1794] 此外,tracrRNA结点和发夹圈对于Cas9配对至关重要并且可交换以跨越分离不同Cas9蛋白的正交屏障,这有利于进一步利用正交Cas9蛋白。在一些实施方案中,结点和发夹圈经交换以靶向正交Cas9蛋白。在一些实施方案中,sgRNA没有上部茎、发夹圈1、和/或下部茎的柔软性,以设计更紧凑且构象稳定的向导RNA。在一些实施方案中,模块经修饰以优化使用具有各种嵌合向导的单Cas9进行的或通过同时使用具有不同嵌合sgRNA的组合的正交系统进行的多重编辑。关于向导功能模块及其方法的细节描述于例如Briner等人,Guide RNA Functional Modules Direct Cas9 Activity and Orthogonality Mol Cell.20140ct 23;56(2):333-339中,其内容通过引用以其整体并入本文。

[1795] 本文所公开的碱基编辑器的机构与可以以任何次序排列。包含融合蛋白(包含例如多核苷酸可编程核苷酸结合结构域(例如,Cas9和Cas12)和脱氨酶结构域(例如,胞苷或腺苷脱氨酶))的碱基编辑器的非限制性示例可排列如下:

[1796] NH_2-[核碱基编辑结构域]-连接子1-[核碱基编辑结构域]-COOH;

[1797] NH_2-[脱氨酶]-连接子1-[核碱基编辑结构域]-COOH;

[1798] NH_2-[脱氨酶]-连接子1-[核碱基编辑结构域]-连接子2-[UGI]-COOH;

[1799] NH_2-[脱氨酶]-连接子1-[核碱基编辑结构域]-COOH;

[1800] Nh2-[腺苷$_2$-[腺苷脱氨酶]-连接子1-[核碱基编辑结构域]-COOH;

[1801] NH_2-[核碱基编辑结构域]-[脱氨酶]-COOH;

[1802] NH_2-[脱氨酶]-[核碱基编辑结构域]-[肌苷BER抑制剂]-COOH;

[1803] NH_2-[脱氨酶]-[肌苷BER抑制剂]-[核碱基编辑结构域]-COOH;

[1804] NH_2-[肌苷BER抑制剂]-[脱氨酶]-[核碱基编辑结构域]-COOH;

[1805] NH_2-[核碱基编辑结构域]-[脱氨酶]-[肌苷BER抑制剂]-COOH;

[1806] NH_2-[核碱基编辑结构域]-[肌苷BER抑制剂]-[脱氨酶]-COOH;

[1807] NH_2-[肌苷BER抑制剂]-[核碱基编辑结构域]-[脱氨酶]-COOH;

[1808] NH2-[核碱基编辑结构域]-连接子1-[脱氨酶]-连接子2-[核碱基编辑结构域]-COOH;

[1809] NH2-[核碱基编辑结构域]-连接子1-[脱氨酶]-[核碱基编辑结构域]-COOH;

[1810] NH2-[核碱基编辑结构域]-[脱氨酶]-连接子2-[核碱基编辑结构域]-COOH;

[1811] NH2-[核碱基编辑结构域]-[脱氨酶]-[核碱基编辑结构域]-COOH;

[1812] NH2-[核碱基编辑结构域]-连接子1-[脱氨酶]-连接子2-[核碱基编辑结构域]-[肌苷BER抑制剂]-COOH;

[1813] NH2-[核碱基编辑结构域]-连接子1-[脱氨酶]-[核碱基编辑结构域]-[肌苷BER抑制剂]-COOH;

[1814] NH2-[核碱基编辑结构域]-[脱氨酶]-连接子2-[核碱基编辑结构域]-[肌苷BER抑制剂]-COOH;

[1815] NH2-[核碱基编辑结构域]-[脱氨酶]-[核碱基编辑结构域]-[肌苷BER抑制剂]-COOH;

[1816] NH2-[肌苷BER抑制剂]-[核碱基编辑结构域]-连接子1-[脱氨酶]-连接子2-[核碱基编辑结构域]-COOH;

[1817] NH2-[肌苷BER抑制剂]-[核碱基编辑结构域]-连接子1-[脱氨酶]-[核碱基编辑结构域]-COOH;

[1818] NH2-[肌苷BER抑制剂]-[核碱基编辑结构域]-[脱氨酶]-连接子2-[核碱基编辑结构域]-COOH;或

[1819] NH2-[肌苷BER抑制剂]NH2-[核碱基编辑结构域]-[脱氨酶]-[核碱基编辑结构域]-COOH。

[1820] 在一些实施方案中,本文所提供的碱基编辑融合蛋白需要定位在精确位置处,例如,所定义区(例如,“脱氨基窗口”)内的靶标碱基被放置之处。在一些实施方案中,靶标可以在4个碱基区内。在一些实施方案中,这种所定义的靶标区可以是PAM上游约15个碱基。参见Komor,A.C.,等人,“Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage”Nature 533,420-424(2016);Gaudelli,N.M.,等人,“Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage”Nature 551,464-471(2017);和Komor,A.C.,等人,“Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity”Science Advances 3:eaao4774(2017),其整体内容通过引用并入本文。

[1821] 定义的靶标区可以是脱氨基窗口。脱氨基窗口可以是定义区,其中碱基编辑器作用于靶标核苷酸并将靶标核苷酸脱氨基。在一些实施方案中,脱氨基窗口在2、3、4、5、6、7、8、9或10个碱基区内。在一些实施方案中,脱氨基窗口是PAM上游5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24或25个碱基。

[1822] 本公开的碱基编辑器可包含任何促进对靶标多核苷酸序列进行编辑的结构域、特征或氨基酸序列。例如,在一些实施方案中,碱基编辑器包含核定位序列(NLS)。在一些实施方案中,碱基编辑器的NLS定位在脱氨酶结构域与napDNAbp之间。在一些实施方案中,碱基编辑器的NLS定位在napDNAbp的C端。

[1823] 可包括在融合蛋白中的蛋白质结构域的非限制性示例包括脱氨酶结构域(例如,腺苷脱氨酶或胞苷脱氨酶)、尿嘧啶糖基化酶抑制剂(UGI)结构域、表位标签、受体基因序列和/或具有以下一种或多种活性的蛋白质结构域:甲基化酶活性、去甲基酶活性、转录活化活性、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裂解活性和核酸结合活性。另外的结构域恶意是异源功能性结构域。此类异源功能性结构域可赋予功能活性,诸如与靶标DNA缔合的靶标多肽(例如,组蛋白、DNA结合蛋白等)的修饰,导致例如组蛋白甲基化、组蛋白乙酰化、组蛋白泛素化等。所赋

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、磺酰化活性、去磺酰化活性或上述的任意组合。

[1824] 所赋予的其他功能可包括甲基转移酶活性、去甲基酶活性、脱氨基活性、歧化酶活性、烷基化活性、脱嘌呤活性、氧化活性、嘧啶二聚体形成活、整合酶活性、转座酶活性、重组酶活性、聚合酶活性、连接酶活性、解旋酶活性、光解酶活性或糖基化酶活性、乙酰转移酶活性、脱乙酰酶活性、激酶活性、磷酸酶活性、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、磺酰化活性、去磺酰化活性、核糖基化活性、脱核糖基化活性、肉豆蔻酰化活性、重塑活性、蛋白酶活性、氧化还原酶活性、转移酶活性、水解酶活性、裂解酶活性、异构酶活性、合酶活性、合成酶活性和脱肉豆蔻酰化活性,或其任意组合。

[1825] 结构域可以经检测或使用表位标签、受体蛋白、其他结合结构域标记。表位标签的非限制性示例包括组氨酸(His)标签、V5标签、FLAG标签、流感血凝素(HA)标签、Myc标签、VSV-G标签和硫氧还蛋白(Trx)标签。受体基因的示例包括但不限于,谷胱甘肽-5-转移酶(GST)、辣根过氧化物酶(HRP)、氯霉素乙酰转移酶(CAT)、β-半乳糖苷酶、β-葡萄糖醛酸酶、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)、和包括蓝色荧光蛋白(BFP)在内的自发荧光蛋白。另外的蛋白质序列可包括氨基酸序列,其结合DNA分子或结合其他细胞分子,包括但不限于,麦芽糖结合蛋白(MBP)、S-tag、Lex A DNA结合结构域(DBD)融合物、GAL4 DNA结合结构域融合物、和单纯疱疹病毒(HSV)BP16蛋白融合物。

[1826] 使用包含胞苷或腺苷脱氨酶和Cas9结构域的融合蛋白的方法

[1827] 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含以下的方法:使DNA分子与本文所提供的任何融合蛋白接触并且与至少一种向导RNA接触,其中向导RNA具有约15至100个核苷酸的长度并且包含与靶标序列互补的至少10个连续核苷酸的序列。在一些实施方案中,靶标序列的3'末端紧邻标准PAM序列(NGG)。在一些实施方案中,靶标序列的3'末端不紧邻标准PAM序列(NGG)。在一些实施方案中,靶标序列的3'末端紧邻AGC、GAG、TTT、GTG或CAA序列。在一些实施方案中,靶标序列的3'末端紧邻NGA、NGCG、NGN、NNGRRT、NNNRRT、NGCG、NGCN、NGTN、NGTN、NGTN或5'(TTTV)序列。

[1828] 在一些实施方案中,本发明的融合蛋白用于诱变感兴趣的靶标。特别地,本文所述的胞苷脱氨酶或腺苷脱氨酶核碱基编辑器能够在靶标序列内作出多个突变。这些突变可能影响靶标的功能。例如,当使用胞苷脱氨酶或腺苷脱氨酶核碱基编辑器来靶向调节区时,该调节区的功能被改变并且下游蛋白的表达被减少或消除。

[1829] 应理解,相应序列中的具体位置或残基的编号取决于特定蛋白质和所使用的编号方案。编号可以是不同的,例如,在成熟蛋白质的前体和成熟蛋白质本身中,并且来自物种与物种之间的序列差异可能影响编号。本领域技术人员将能够通过本领域周知的方法鉴定任何同源蛋白质中和相应编码核酸中的相应残基,例如,通过序列比对并确定同源残基。

[1830] 对本领域技术人员将是显而易见的是,为了将如本文所公开的任何包含Cas9结构域和胞苷或腺苷脱氨酶融合蛋白靶向至靶标位点,例如,包含待编辑的突变的位点,通常必要的是将该融合蛋白与向导RNA例如sgRNA共表达。如在本文其他处更详细阐述的,向导RNA通常包含允许Cas9与向导序列结合的tracrRNA框架,并且该向导序列赋予Cas9:核酸编辑

酶/结构域融合蛋白以特异性。或者,向导RNA和tracrRNA可以作为两个核酸分子独立地提供。在一些实施方案中,向导RNA包含结构,其中向导序列包含与靶标序列互补的序列。向导序列通常是20个核苷酸的长度。基于本公开,适用于将Cas9:核酸编辑酶/结构域融合蛋白百香之特异性基因组位点的向导RNA的序列对于本领域技术人员将是显而易见的。此类合适的向导RNA序列通常包含与待编辑的靶标核苷酸上游或下游50个核苷酸内的核酸序列互补的向导序列。一些适用于将所提供的任何融合蛋白靶向至特异性靶标序列的示例性向导RNA提供在本文中。

[1831] 碱基编辑器效率

[1832] 在一些实施方案中,本文所提供的方法的目的是经由基因标记中断基因产物的正常功能。本文所提供的核碱基编辑蛋白质可以在体外用于基于基因编辑的人类疗法。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,包含多核苷酸可编程核苷酸结合结构域(例如,Cas9)和核碱基编辑结构域(例如,腺苷脱氨酶结构域和胞苷脱氨酶结构域)的融合蛋白可用来进行A到G或C到T的核苷酸编辑。

[1833] CRISPR-Cas9核酸酶已经广泛用于介导靶向基因组编辑。在大多数基因组编辑应用中,Cas9与向导多核苷酸(例如,单向导(sgRNA))形成复合物并在该sgRNA序列特异性的靶标位点处诱导双链DNA断裂(DSB)。细胞主要通过非同源性末端接合(NHEJ)修复途径对这一DSB作出应答,这导致随机插入或缺失(插入缺失),可造成移码突变,从而中断该基因。在具有与DSB侧翼序列具有高度同源性的供体DNA模板的存在下,可通过称为同源定向修复(HDR)的替代性途径实现基因校正。不幸的是,在大多数非微扰条件下,HDR是无效的,取决于细胞状态和细胞类型,且更高频率的插入缺失占据优势。由于大多数已知的与人类疾病相关的基因变异是点突变,需要可能更有效且干净地作出精确点突变的方法。本文所提供的碱基编辑系统提供一种新的方式来提供基因组编辑而不生成双链DNA断裂,无需供体DNA目标,并且不会诱导过量的随机插入和缺失。

[1834] 在一些实施方案中,本公开提供碱基编辑器,该碱基编辑器有效地在核酸(例如,受试者基因组内的核酸)中生成“期望突变”诸如STOP密码子,而不生成大量的非期望突变诸如非期望的点突变。在一些实施方案中,期望突变是通过结合至向导多核苷酸(例如,gRNA)的经专门设计以生成期望突变的特异性碱基编辑器(例如,腺苷碱基编辑器或胞苷碱基编辑器)生成的突变。在一些实施方案中,预期突变是在与疾病或疾患(例如,T细胞或NK细胞恶性病变)相关靶标抗原缔合的基因内。在一些实施方案中,预期突变是在与疾病或疾患(例如,T细胞或NK细胞恶性病变)相关靶标抗原缔合的基因内的腺嘌呤(A)到鸟嘌呤(G)点突变(例如,SNP)。在一些实施方案中,预期突变是基因的编码区或非编码区(例如,调节区或元件)内的腺嘌呤(A)到鸟嘌呤(G)点突变。在一些实施方案中,预期突变是在与疾病或疾患(例如,T细胞或NK细胞恶性病变)相关靶标抗原缔合的基因内的胞嘧啶(C)到胸腺嘧啶(T)点突变(例如,SNP)。在一些实施方案中,预期突变是基因的编码区或非编码区(例如,调节区或元件)内的胞嘧啶(C)到胸腺嘧啶(T)点突变。在一些实施方案中,预期突变是点突变,其生成STOP密码子,例如,基因的编码区内的提前STOP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。

[1835] 本发明的碱基编辑器有利地修饰特异性核苷酸碱基编码蛋白而不生成显著比例的插入缺失。如本文所用,“插入缺失”是指核苷酸在核酸内的插入或缺失。此类插入或缺失

可导致基因的编码区内的框移突变。在一些实施方案中,希望生成碱基编辑器,其有效地修饰(例如,突变)核酸内的特异性核苷酸而不在该核酸内生成大量的插入或缺失(即,插入缺失)。在一些实施方案中,希望生成碱基编辑器,其有效地修饰(例如,突变或甲基化)核酸内的特异性核苷酸而不在该核酸内生成大量的插入或缺失(即,插入缺失)。在某些实施方案中,本文所提供的任何碱基编辑器均可生成与插入缺失相比更高比例的预期修饰(例如,甲基化)。在某些实施方案中,本文所提供的任何碱基编辑器均可生成与插入缺失相比更高比例的预期修饰(例如,突变)。

[1836] 在一些实施方案中,本文所提供的碱基编辑器能够生成大于1:1的预期突变与插入缺失比率(即,预期点突变:非预期点突变)。在一些实施方案中,本文所提供的碱基编辑器能够生成预期突变与插入缺失的比率,该比率为至少1.5:1、至少2:1、至少2.5:1、至少3:1、至少3.5:1、至少4:1、至少4.5:1、至少5:1、至少5.5:1、至少6:1、至少6.5:1、至少7:1、至少7.5:1、至少8:1、至少10:1、至少12:1、至少15:1、至少20:1、至少25:1、至少30:1、至少40:1、至少50:1、至少100:1、至少200:1、至少300:1、至少400:1、至少500:1、至少600:1、至少700:1、至少800:1、至少900:1、或至少1000:1或更高。可使用任何合适的方法确定预期突变和插入缺失的数量。

[1837] 在一些实施方案中,本文所提供的碱基编辑器可限制插入缺失在核酸区内的形成。在一些实施方案中,该区在被碱基编辑器编辑的核苷酸处或是与被碱基编辑器编辑的核苷酸相距2、3、4、5、6、7、8、9或10和核苷酸内的区。在一些实施方案中,本文所提供的任何碱基编辑器可将核酸区处的插入缺失形成限制为少于1%、少于1.5%、少于2%、少于2.5%、少于3%、少于3.5%、少于4%、少于4.5%、少于5%、少于6%、少于7%、少于8%、少于9%、少于10%、少于12%、少于15%、或少于20%。在核酸区处形成的插入缺失数量可以取决于核酸(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,插入缺失的数量或比例在将核酸(例如,细胞基因组内的核酸)暴露于碱基编辑器中至少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天后确定。

[1838] 本公开的一些方面是基于以下认知: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)中生成预期突变而不生成可观数量的非预期突变(例如,谬误的脱靶编辑或旁观者编辑)。在一些实施方案中,预期突变是通过结合至gRNA的特异性碱基编辑器产生的,该碱基编辑器经专门设计以生成预期突变。在一些实施方案中,预期突变是生成终止密码子的突变,例如,基因的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。在一些实施方案中,预期突变是改变基因剪接的突变。在一些实施方案中,预期突变是改变基因调节序列(例如,基因启动子或基因受体)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生成大于1:1的预期突变与非预期比率(例如,预期突变:非预期突变)。在一些实施方案中,本文所提供的任何碱基编辑器能够生成预期突变与非预期突变的比率,该比率为至少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。应了解,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。

[1839] 碱基编辑通常称为“修饰”,诸如基因的修饰、基因修饰和核酸序列的修饰,并且基于修饰是碱基编辑修饰的语境而可以清楚地理解。因此,碱基编辑修饰是核苷酸碱基水平上的修饰,例如,作为本公开通篇所讨论的脱氨酶活性的结果,其随后导致基因序列的改变,并且可以影响基因产物。因此,本质上,本文所述的基因编辑修饰可导致结构上和/或功能上的接引修饰,其中基因产物的表达可以经修饰,例如,该基因的表达被敲除;或反而被增强,或者在一些情况下,基因功能或活性可以经修饰。使用本文所公开的方法,碱基编辑效率可以确定为在其中进行碱基编辑的基因的敲低效率,其中碱基编辑旨在敲低该基因的表达。敲低水平可以通过下列定量验证:通过任何检测测定法确定表达水平,诸如用于例如通过流式细胞术测定蛋白质表达水平的测定法;用于检测RNA表达的测定法,诸如定量RT-PCR、北方印迹分析,或任何其他合适的测定法,诸如焦磷酸测序;并且可以通过核苷酸测序反应来定量验证。

[1840] 在一些实施方案中,修饰,例如,单碱基编辑导致所靶向的基因的表达降低至少10%。在一些实施方案中,碱基编辑效率可导致所靶向的基因的表达降低至少10%。在一些实施方案中,碱基编辑效率可导致所靶向的基因的表达降低至少20%。在一些实施方案中,碱基编辑效率可导致所靶向的基因的表达降低至少30%。在一些实施方案中,碱基编辑效率可导致所靶向的基因的表达降低至少40%。在一些实施方案中,碱基编辑效率可导致所靶向的基因的表达降低至少50%。在一些实施方案中,碱基编辑效率可导致所靶向的基因的表达降低至少60%。在一些实施方案中,碱基编辑效率可导致所靶向的基因的表达降低至少70%。在一些实施方案中,碱基编辑效率可导致所靶向的基因的表达降低至少80%。在一些实施方案中,碱基编辑效率可导致所靶向的基因的表达降低至少90%。在一些实施方案中,碱基编辑效率可导致所靶向的基因的表达降低至少91%。在一些实施方案中,碱基编辑效率可导致所靶向的基因的表达降低至少92%。在一些实施方案中,碱基编辑效率可导致所靶向的基因的表达降低至少93%。在一些实施方案中,碱基编辑效率可导致所靶向的基因的表达降低至少94%。在一些实施方案中,碱基编辑效率可导致所靶向的基因的表达降低至少95%。在一些实施方案中,碱基编辑效率可导致所靶向的基因的表达降低至少96%。在一些实施方案中,碱基编辑效率可导致所靶向的基因的表达降低至少97%。在一些实施方案中,碱基编辑效率可导致所靶向的基因的表达降低至少98%。在一些实施方案中,碱基编辑效率可导致所靶向的基因的表达降低至少99%。在一些实施方案中,碱基编辑效率可导致被靶向的基因的敲除(基因表达的100%敲低)。

[1841] 在一些实施方案中,本文所提供的任何碱基编辑器系统在靶标多核苷酸序列中导致少于50%、少于40%、少于30%、少于20%、少于19%、少于18%、少于17%、少于16%、少于15%、少于14%、少于13%、少于12%、少于11%、少于10%、少于9%、少于8%、少于7%、少于6%、少于5%、少于4%、少于3%、少于2%、少于1%、少于0.9%、少于0.8%、少于0.7%、少于0.6%、少于0.5%、少于0.4%、少于0.3%、少于0.2%、少于0.1%、少于0.09%、少于0.08%、少于0.07%、少于0.06%、少于0.05%、少于0.04%、少于0.03%、少于0.02%、或少于0.01%的插入缺失形成。

[1842] 在一些实施方案中,使用靶向修饰例如单碱基编辑同时将用于碱基编辑的至少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种不同的内源性序列与不同的向导RNA一起靶向。在一些实施方案中,使用靶向修饰例如单碱基编辑依次将用于碱基编辑的至少4、5、6、7、8、9、10、11、12 13、14、15、16、17,18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50或更多种不同的内源性基因序列与不同的向导RNA一起靶向。

[1843] 本公开的一些方面是基于以下认知: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)中生成预期突变诸如点突变而不生成显著数量的非预期突变诸如非预期点突变(即,旁观者的突变)。在一些实施方案中,本文所提供的任何碱基编辑器能够生成至少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生成至少0.01%、1%、2%、3%、4%、5%、10%、15%、20%、25%、30%、40%、45%、50%、60%、70%、80%、90%、95%或99%的插入缺失突变。

[1844] 在一些实施方案中,本文所述的任何包含ABE8碱基编辑器变体的碱基编辑器系统在靶标多核苷酸序列中导致少于50%、少于40%、少于30%、少于20%、少于19%、少于18%、少于17%、少于16%、少于15%、少于14%、少于13%、少于12%、少于11%、少于10%、少于9%、少于8%、少于7%、少于6%、少于5%、少于4%、少于3%、少于2%、少于1%、少于0.9%、少于0.8%、少于0.7%、少于0.6%、少于0.5%、少于0.4%、少于0.3%、少于0.2%、少于0.1%、少于0.09%、少于0.08%、少于0.07%、少于0.06%、少于0.05%、少于0.04%、少于0.03%、少于0.02%、或少于0.01%的插入缺失形成。在一些实施方案中,包含本文所述ABE8碱基编辑器变体中的一个的任何碱基编辑器系统在靶标多核苷酸序列中导致少于0.8%的插入缺失形成。在一些实施方案中,包含本文所述ABE8碱基编辑器变体中的一个的任何碱基编辑器系统在靶标多核苷酸序列中导致至多0.8%的插入缺失形成。在一些实施方案中,包含本文所述ABE8碱基编辑器变体中的一个的任何碱基编辑器系统在靶标多核苷酸序列中导致少于0.3%的插入缺失形成。在一些实施方案中,在靶标多核苷酸序列中导致更低的插入缺失形成。在一些实施方案中,与包含ABE7.10的碱基编辑器系统相比,包含本文所述ABE8碱基编辑器变体中的一个的任何碱基编辑器系统在靶标多核苷酸序列中导致更低的插入缺失形成。

[1845] 在一些实施方案中,与包含ABE7碱基编辑器中的一个的碱基编辑器系统相比,包含所述ABE8碱基编辑器变体中的一个的任何碱基编辑器系统具有降低的插入缺失频率。在一些实施方案中,与包含ABE7碱基编辑器中的一个的碱基编辑器系统相比,包含所述ABE8碱基编辑器变体中的一个的任何碱基编辑器系统具有降低至少0.01%、至少1%、至少2%、至少3%、至少4%、至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、或至少95%的插入缺失频率。在一些实施方案中,与包含ABE7.10的碱基编辑器系统相比,包含所述ABE8碱基编辑器变体中的一个的碱基编辑器系统具有降低至少0.01%、至少1%、至少2%、至少3%、至少4%、至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、或至少95%的插入缺失频率。

[1846] 本发明提供腺苷脱氨酶变体(例如,ABE8变体),其具有增加的效率和特异性。特别地,本文所述的腺苷脱氨酶变体更可能编辑多核苷酸内的所希望的碱基,并且更不可能编辑不希望被改变的碱基(例如,“旁观者”)。

[1847] 在一些实施方案中,包含本文所述ABE碱基编辑器变体中的一个的任何碱基编辑系统具有降低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,在靶标核苷酸序列的靶标窗内的非预期或非靶标位置中的靶标碱基(例如,A或C)的碱基编辑。在一些实施方案中,与包含ABE7碱基编辑器例如ABE7.10的碱基编辑器系统相比,包含本文所述ABE碱基编辑器变体中的一个的任何碱基编辑系统具有降低的旁观者编辑或突变。在一些实施方案中,与包含ABE7碱基编辑器例如ABE7.10的碱基编辑器系统相比,包含本文所述ABE碱基编辑器变体中的一个的任何碱基编辑系统具有降低至少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%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、或至少99%的旁观者编辑或突变。在一些实施方案中,与包含ABE7碱基编辑器例如ABE7.10的碱基编辑器系统相比,包含本文所述ABE碱基编辑器变体中的一个的任何碱基编辑系统具有降低至少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、或至少3.0倍的旁观者编辑或突变。

[1848] 在一些实施方案中,包含本文所述ABE碱基编辑器变体中的一个的任何碱基编辑系统具有降低的谬误编辑。在一些实施方案中,非预期编辑或突变是谬误突变或谬误编辑,例如,在基因组的非预期或非靶标区内的靶标碱基(例如,A或C)的非特异性编辑或不依赖于向导的编辑。在一些实施方案中,与包含ABE7碱基编辑器例如ABE7.10的碱基编辑器系统相比,包含本文所述ABE碱基编辑器变体中的一个的任何碱基编辑系统具有降低的谬误编辑。在一些实施方案中,与包含ABE7碱基编辑器例如ABE7.10的碱基编辑器系统相比,包含本文所述ABE碱基编辑器变体中的一个的任何碱基编辑系统具有降低至少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%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、或至少99%的谬误编辑。在一些实施方案中,与包含ABE7碱基编辑器例如ABE7.10的碱基编辑器系统相比,包含本文所述ABE碱基编辑器变体中的一个的任何碱基编辑系统具有降低至少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、或至少3.0倍的谬误编辑。

[1849] 在一些实施方案中,本文所述的任何ABE8碱基编辑器变体的碱基编辑效率为至少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%。在一些实施方案中,通过计算细胞群中经编辑的核碱基的百分比来测量碱基编辑效率。在一些实施方案中,本文所述的任何ABE8碱基编辑器变体的碱基编辑效率为至少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%,如通过细胞群中经编辑的核碱基所测量。

[1850] 在一些实施方案中,与ABE7碱基编辑器相比,本文所述的任何ABE8碱基编辑器变体的碱基编辑效率更高。在一些实施方案中,与ABE7碱基编辑器例如ABE7.10相比,本文所述的任何ABE8碱基编辑器变体的碱基编辑效率高出至少1%、至少2%、至少3%、至少4%、至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少99%、至少100%、至少105%、至少110%、至少115%、至少120%、至少125%、至少130%、至少135%、至少140%、至少145%、至少150%、至少155%、至少160%、至少165%、至少170%、至少175%、至少180%、至少185%、至少190%、至少195%、至少200%、至少210%、至少220%、至少230%、至少240%、至少250%、至少260%、至少270%、至少280%、至少290%、至少300%、至少310%、至少320%、至少330%、至少340%、至少350%、至少360%、至少370%、至少380%、至少390%、至少400%、至少450%、或至少500%。

[1851] 在一些实施方案中,与ABE7碱基编辑器例如ABE7.10相比,本文所述的任何ABE8碱基编辑器变体的碱基编辑效率高出至少1.1倍、至少1.2倍、至少1.3倍、至少1.4倍、至少1.5倍、至少1.6倍、至少1.7倍、至少1.8倍、至少1.9倍、至少2.0倍、至少2.1倍、至少2.2倍、至少2.3倍、至少2.4倍、至少2.5倍、至少2.6倍、至少2.7倍、至少2.8倍、至少2.9倍、至少3.0倍、至少3.1倍、至少3.2、至少3.3倍、至少3.4倍、至少3.5倍、至少3.6倍、至少3.7倍、至少3.8倍、至少3.9倍、至少4.0倍、至少4.1倍、至少4.2倍、至少4.3倍、至少4.4倍、至少4.5倍、至少4.6倍、至少4.7倍、至少4.8倍、至少4.9倍、或至少5.0倍。

[1852] 在一些实施方案中,本文所述的任何ABE8碱基编辑器变体的上靶碱基编辑效率为至少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%。在一些实施方案中,本文所述的任何ABE8碱基编辑器变体的上靶碱基编辑效率为至少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%,如通过细胞群中经编辑的靶标核碱基所测量。

[1853] 在一些实施方案中,与ABE7碱基编辑器相比,本文所述的任何ABE8碱基编辑器变体的上靶碱基编辑效率更高。在一些实施方案中,与ABE7碱基编辑器例如ABE7.10相比,本文所述的任何ABE8碱基编辑器变体的上靶碱基编辑效率高出至少1%、至少2%、至少3%、至少4%、至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少99%、至少100%、至少105%、至少110%、至少115%、至少120%、至少125%、至少130%、至少135%、至少140%、至少145%、至少150%、至少

155％、至少160％、至少165％、至少170％、至少175％、至少180％、至少185％、至少190％、至少195％、至少200％、至少210％、至少220％、至少230％、至少240％、至少250％、至少260％、至少270％、至少280％、至少290％、至少300％、至少310％、至少320％、至少330％、至少340％、至少350％、至少360％、至少370％、至少380％、至少390％、至少400％、至少450％、或至少500％。

[1854] 在一些实施方案中，与ABE7碱基编辑器例如ABE7.10相比，本文所述的任何ABE8碱基编辑器变体的上靶碱基编辑效率高出至少1.1倍、至少1.2倍、至少1.3倍、至少1.4倍、至少1.5倍、至少1.6倍、至少1.7倍、至少1.8倍、至少1.9倍、至少2.0倍、至少2.1倍、至少2.2倍、至少2.3倍、至少2.4倍、至少2.5倍、至少2.6倍、至少2.7倍、至少2.8倍、至少2.9倍、至少3.0倍、至少3.1倍、至少3.2倍、至少3.3倍、至少3.4倍、至少3.5倍、至少3.6倍、至少3.7倍、至少3.8倍、至少3.9倍、至少4.0倍、至少4.1倍、至少4.2倍、至少4.3倍、至少4.4倍、至少4.5倍、至少4.6倍、至少4.7倍、至少4.8倍、至少4.9倍、或至少5.0倍。

[1855] 本文所述的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，本文所述的任何ABE8碱基编辑器变体作为mRNA递送至宿主细胞。.在一些实施方案中，经由基于核酸的递送系统例如mRNA递送的ABE8碱基编辑器具有至少1％、至少2％、至少3％、至少4％、至少5％、至少10％、至少15％、至少20％、至少25％、至少30％、至少35％、至少40％、至少45％、至少50％、至少55％、至少60％、至少65％、至少70％、至少75％、至少80％、至少85％、至少90％、至少95％、或至少99％的上靶编辑效率，如通过经编辑的核碱基所测量。在一些实施方案中，与通过质粒或载体系统递送的ABE8碱基编辑器相比，通过mRNA系统递送的ABE8碱基编辑器具有更高的甲基编辑效率。在一些实施方案中，与通过质粒或载体系统递送相比，本文所述的任何ABE8碱基编辑器变体在通过mRNA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％高、至少310％、至少320％、至少330％、至少340％、至少350％、至少360％、至少370％、至少380％、至少390％、至少400％、至少450％、或至少500％的上靶编辑效率。在一些实施方案中，与通过质粒或载体系统递送相比，本文所述的任何ABE8碱基编辑器变体在通过mRNA递送时具有至少高出1.1倍、至少1.2倍、至少1.3倍、至少1.4倍、至少1.5倍、至少1.6倍、至少1.7倍、至少1.8倍、至少1.9倍、至少2.0倍、至少2.1倍、至少2.2倍、至少2.3倍、至少2.4倍、至少2.5倍、至少2.6倍、至少2.7倍、至少2.8倍、至少2.9倍、至少3.0倍、至少3.1倍、至少3.2倍、至少3.3倍、至少3.4倍、至少3.5倍、至少3.6倍、至少3.7倍、至少3.8倍、至少3.9倍、至少4.0倍、至少4.1倍、至少4.2倍、至少4.3倍、至少4.4倍、至少4.5倍、至少4.6倍、至少4.7倍、至少4.8倍、至少4.9倍、或至少5.0倍的上靶编辑效率。

[1856] 在一些实施方案中，本文所述的任何包含ABE8碱基编辑器变体的碱基编辑器系统在靶标多核苷酸序列中导致少于50％、少于40％、少于30％、少于20％、少于19％、少于

18%、少于17%、少于16%、少于15%、少于14%、少于13%、少于12%、少于11%、少于10%、少于9%、少于8%、少于7%、少于6%、少于5%、少于4%、少于3%、少于2%、少于1%、少于0.9%、少于0.8%、少于0.7%、少于0.6%、少于0.5%、少于0.4%、少于0.3%、少于0.2%、少于0.1%、少于0.09%、少于0.08%、少于0.07%、少于0.06%、少于0.05%、少于0.04%、少于0.03%、少于0.02%、或少于0.01%的脱靶编辑。

[1857] 在一些实施方案中,与通过质粒或载体系统递送时相比,本文所述的任何ABE8碱基编辑器变体当通过mRNA系统递送时具有较低的向导依赖性脱靶编辑效率。在一些实施方案中,与通过质粒或载体系统递送时相比,本文所述的任何ABE8碱基编辑器变体当通过mRNA系统递送时具有降低至少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%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、或至少99%的向导依赖性脱靶编辑效率。在一些实施方案中,与通过质粒或载体系统递送时相比,本文所述的任何ABE8碱基编辑器变体当通过mRNA系统递送时具有降低至少1.1倍、至少1.2倍、至少1.3倍、至少1.4倍、至少1.5倍、至少1.6倍、至少1.7倍、至少1.8倍、至少1.9倍、至少2.0倍、至少2.1倍、至少2.2倍、至少2.3倍、至少2.4倍、至少2.5倍、至少2.6倍、至少2.7倍、至少2.8倍、至少2.9倍、或至少3.0倍的向导依赖性脱靶编辑效率。在一些实施方案中,与通过质粒或载体系统递送时相比,本文所述的任何ABE8碱基编辑器变体当通过mRNA系统递送时的向导依赖性脱靶编辑效率下降至少约2.2倍。

[1858] 在一些实施方案中,与通过质粒或载体系统递送时相比,本文所述的任何ABE8碱基编辑器变体当通过mRNA系统递送时具有较低的不依赖于向导的脱靶编辑效率。在一些实施方案中,与通过质粒或载体系统递送时相比,本文所述的任何ABE8碱基编辑器变体当通过mRNA系统递送时具有降低至少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%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、或至少99%的不依赖于向导的脱靶编辑效率。在一些实施方案中,与通过质粒或载体系统递送相比,本文所述的任何ABE8碱基编辑器变体在通过mRNA递送时具有降低至少1.1倍、至少1.2倍、至少1.3倍、至少1.4倍、至少1.5倍、至少1.6倍、至少1.7倍、至少1.8倍、至少1.9倍、至少2.0倍、至少2.1倍、至少2.2倍、至少2.3倍、至少2.4倍、至少2.5倍、至少2.6倍、至少2.7倍、至少2.8倍、至少2.9倍、至少3.0倍、至少5.0倍、至少10.0倍、至少20.0倍、至少50.0倍、至少70.0倍、至少100.0倍、至少120.0倍、至少130.0倍、或至少150.0倍的不依赖于向导的脱靶编辑效率。在一些实施方案中,与通过质粒或载体系统递送相比,本文所述的ABE8碱基编辑器变体在通过mRNA递送时具有降低134.0倍的不依赖于向导的脱靶编辑效率(例如,谬误RNA脱氨基)。在一些实施方案中,本文所述的ABE8碱基编辑器变体不跨基因组增加不依赖于向导的突变率。

[1859] 在一些实施方案中,单个基因递送事件(例如,通过转导、转染、电穿孔或任何其他方法)可用于细胞基因组内5个序列的靶标碱基编辑。在一些实施方案中,单个基因递送事件可用于细胞基因组内6个序列的靶标碱基编辑。在一些实施方案中,单个基因递送事件可用于细胞基因组内7个序列的靶标碱基编辑。在一些实施方案中,单个电穿孔事件可用于细胞基因组内8个序列的靶标碱基编辑。在一些实施方案中,单个基因递送事件可用于细胞基

因组内9个序列的靶标碱基编辑。在一些实施方案中,单个基因递送事件可用于细胞基因组内10个序列的靶标碱基编辑。在一些实施方案中,单个基因递送事件可用于细胞基因组内20个序列的靶标碱基编辑。在一些实施方案中,单个基因递送事件可用于细胞基因组内30个序列的靶标碱基编辑。在一些实施方案中,单个基因递送事件可用于细胞基因组内40个序列的靶标碱基编辑。在一些实施方案中,单个基因递送事件可用于细胞基因组内50个序列的靶标碱基编辑。

[1860] 在一些实施方案中,本文所述的方法,例如,碱基编辑方法具有极低到没有的脱靶效应。

[1861] 在一些实施方案中,本文所述的碱基编辑方法导致细胞群的至少50%已经被成功地编辑(即,细胞已经成功地工程改造)。在一些实施方案中,本文所述的碱基编辑方法导致细胞群的至少55%已经被成功地编辑。在一些实施方案中,本文所述的碱基编辑方法导致细胞群的至少60%已经被成功地编辑。在一些实施方案中,本文所述的碱基编辑方法导致细胞群的至少65%已经被成功地编辑。在一些实施方案中,本文所述的碱基编辑方法导致细胞群的至少70%已经被成功地编辑。在一些实施方案中,本文所述的碱基编辑方法导致细胞群的至少75%已经被成功地编辑。在一些实施方案中,本文所述的碱基编辑方法导致细胞群的至少80%已经被成功地编辑。在一些实施方案中,本文所述的碱基编辑方法导致细胞群的至少85%已经被成功地编辑。在一些实施方案中,本文所述的碱基编辑方法导致细胞群的至少90%已经被成功地编辑。在一些实施方案中,本文所述的碱基编辑方法导致细胞群的至少95%已经被成功地编辑。在一些实施方案中,本文所述的碱基编辑方法导致细胞群的约91%、92%、93%、94%、95%、96%、97%、98%、99%或100%已经被成功地编辑。

[1862] 在一些实施方案中,在碱基编辑干预后的肝细胞恢复率是在碱基编辑事件开始时细胞群的超过至少60%、70%、80%、90%。在一些实施方案中,上述肝细胞恢复率是约70%。在一些实施方案中,上述肝细胞恢复率是约75%。在一些实施方案中,上述肝细胞恢复率是约80%。在一些实施方案中,上述肝细胞恢复率是约85%。在一些实施方案中,上述肝细胞恢复率是在碱基编辑事件的时间该细胞群中约90%、或约91%、92%、93%、94%95%、96%、97%、98%、或99%、或100%的细胞。

[1863] 在一些实施方案中,经工程改造的细胞群可进一步在体外扩张约2倍、约3倍、约4倍、约5倍、约6倍、约7倍、约8倍、约9倍、约10倍、约15倍、约20倍、约25倍、约30倍、约35倍、约40倍、约45倍、约50倍、或约100倍。

[1864] 预期突变和插入缺失的数量可使用任何合适方法确定,例如,如国际PCT申请号PCT/2017/045381(WO2018/027078)和PCT/US2016/058344(WO2017/070632);Komor,A.C.,等人,"Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage"Nature 533,420-424(2016);Gaudelli,N.M.,等人,"Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage"Nature 551,464-471(2017);和Komor,A.C.,等人,"Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:Abase editors with higher efficiency and product purity"Science Advances 3:eaao4774(2017)中所述,其整体内容通过引用并入本文。

[1865] 在一些实施方案中,为了计算插入缺失频率,扫描了测序读数,以将其确切匹配至位于可能出现插入缺失的窗口两侧的两个10-bp序列。如果没有定位到确切匹配,则将该读数从分析中排除。如果这一插入缺失窗口的长度与参考序列确切匹配,则将该读数归类为不含插入缺失。如果插入缺失窗口比参考序列长或短两个或更多个碱基,则将该测序读数分别归类为插入或缺失。在一些实施方案中,本文所提供的碱基编辑器可限制插入缺失在核酸区内的形成。在一些实施方案中,该区在被碱基编辑器编辑的核苷酸处或是与被碱基编辑器编辑的核苷酸相距2、3、4、5、6、7、8、9或10和核苷酸内的区。

[1866] 在靶标核苷酸区处形成的插入缺失数量可以取决于核酸(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,插入缺失的数量或比例在将靶标核苷酸序列(例如,细胞基因组内的核酸)暴露于碱基编辑器中至少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天后确定。应了解,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。

[1867] 碱基编辑器效率的细节在国际PCT申请号PCT/2017/045381(WO 2018/027078)和PCT/US2016/058344(WO 2017/070632)中描述,其各自通过引用以其整体并入本文。也参见Komor,A.C.,等人,"Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage"Nature 533,420-424(2016);Gaudelli,N.M.,等人,"Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage"Nature 551,464-471(2017);和Komor,A.C.,等人,"Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:Abase editors with higher efficiency and product purity"Science Advances 3:eaao4774(2017),其整体内容通过引用并入本文。在一些实施方案中,使用本文所提供的方法在一个或多个基因中编辑多个核碱基对导致至少一个预期突变的形成。在一些实施方案中,所述至少一个预期突变的所述形成导致基因正常功能的中断。在一些实施方案中,所述至少一个预期突变的所述形成导致基因编码的蛋白质的表达下降或消除。应了解,多重编辑可使用本文所提供的任何方法或方法组合实施。

[1868] 多重编辑

[1869] 在一些实施方案中,本文所提供的碱基编辑器系统能够对一个或多个基因中的多个核碱基对进行多重编辑。在一些实施方案中,该多个核碱基对位于同一基因内。在一些实施方案中,该多个核碱基对位于一个或多个基因内,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,多重编辑可包含一个或多个向导多核苷酸。在一些实施方案中,多重编辑可包含一个或多个碱基编辑器系统。在一些实施方案中,多重编辑可包含一个或多个碱基编辑器系统和一个或多个向导多核苷酸。在一些实施方案中,多重编辑可包含一个或多个碱基编辑器系统和多个向导多核苷酸。在一些实施方案中,多重编辑可包含一个或多个向导多核苷酸和单个碱基编辑器系统。在一些实施方案中,多重编辑可包含至少一个不需要PAM序列来靶向结合至靶标多核苷酸序列的向导多核苷酸。在一些实施方案中,多重编辑可包含至少一个需要PAM序列来靶向结合至靶标多核苷酸序列的向导多核苷酸。在一些实施方案中,多重编辑可包含至少一个不需要PAM序列来靶向结合至靶标多核苷酸序列的向导多核苷酸与至少一个需要PAM序列来靶向结合至靶标多核苷酸序列的向导多核苷酸

的混合物。应了解,使用本文所述的任何碱基编辑器进行的多重编辑特征可应用于使用本文所提供的任何碱基编辑器的方法的任意组合。也应了解,使用本文所述的任何碱基编辑器进行的多重编辑可包括对多个核碱基对的依次编辑。

[1870] 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,该多个核碱基对在同一基因内。在一些实施方案中,一个或多个基因中的至少一个基因定位在不同基因座中。

[1871] 在一些实施方案中,编辑是对至少一个蛋白质编码区内的多个核碱基对的编辑。在一些实施方案中,编辑是对至少一个非蛋白质编码区内的多个核碱基对的编辑。在一些实施方案中,编辑是对至少一个蛋白质编码区和至少一个非蛋白质编码区内的多个核碱基对的编辑。

[1872] 在一些实施方案中,编辑与一个或多个向导多核苷酸协同作用。在一些实施方案中,该碱基编辑器系统可包含一个或多个碱基编辑器系统。在一些实施方案中,该碱基编辑器系统可包含一个或多个碱基编辑器系统与一个或多个向导多核苷酸系统作用。在一些实施方案中,该碱基编辑器系统可包含一个或多个碱基编辑器系统与多个向导多核苷酸系统作用。在一些实施方案中,该编辑与一个或多个向导多核苷酸和单个碱基编辑器系统协同作用。在一些实施方案中,该编辑与至少一个不需要PAM序列来靶向结合至靶标多核苷酸序列的向导多核苷酸协同作用。在一些实施方案中,该编辑与至少一个需要PAM序列来靶向结合至靶标多核苷酸序列的向导多核苷酸协同作用。在一些实施方案中,该编辑与至少一个不需要PAM序列来靶向结合至靶标多核苷酸序列的向导多核苷酸与至少一个需要PAM序列来靶向结合至靶标多核苷酸序列的向导多核苷酸的混合物协同作用。应了解,使用本文所述的任何碱基编辑器进行的多重编辑特征可应用于使用本文所提供的任何碱基编辑器的方法的任意组合。也应了解,编辑可包括对多个核碱基对的依次编辑。

[1873] 在一些实施方案中,能够对一个或多个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含ABE9碱基编辑器中的一个。在一些实施方案中,能够对一个或多个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含本文所述的ABE8碱基编辑器变体中的一个。在一些实施方案中,能够对一个或多个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含ABE7碱基编辑器中的一个。在一些实施方案中,与能够进行多重编辑的包含ABE7碱基编辑器中的一个的碱基编辑器系统相比,能够进行多重编辑的包含本文所述ABE8碱基编辑器变体中的一个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,与能够进行多重编辑的包含ABE7碱基编辑器中的一个的碱基编辑器系统相比,能够进行多重编辑的包含本文所述ABE8碱基编辑器变体中的一个的碱基编辑器系统具有高出至少1%、至少2%、至少3%、至少4%、至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少99%、至少100%、至少105%、至少110%、至少115%、至少120%、至少125%、至少130%、至少135%、至少140%、至少145%、至少150%、至少155%、至少160%、至少165%、至少170%、至少175%、至少180%、至少185%、至少190%、至少195%、至少200%、至少210%、至少220%、至少230%、至少240%、至少250%、至少260%、至少270%、至少280%、至少290%,at least 300% higher、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％的多重编辑效率。在一些实施方案中，与能够进行多重编辑的包含ABE7碱基编辑器中的一个的碱基编辑器系统相比，能够进行多重编辑的包含本文所述ABE8碱基编辑器变体中的一个的碱基编辑器系统具有高出至少1.1倍、至少1.2倍、至少1.3倍、至少1.4倍、至少1.5倍、至少1.6倍、至少1.7倍、至少1.8倍、至少1.9倍、至少2.0倍、至少2.1倍、至少2.2倍、至少2.3倍、至少2.4倍、至少2.5倍、至少2.6倍、至少2.7倍、至少2.8倍、至少2.9倍、至少3.0倍、至少3.1倍、至少3.2倍、至少3.3倍、至少3.4倍、至少3.5倍、至少4.0倍、至少4.5倍、至少5.0倍、至少5.5倍、或至少6.0的多重编辑效率。

[1874] 递送系统

[1875] 如本文所述，评定了核碱基编辑器靶向基因(例如，CD2)中的一个或多个核苷酸的适用性。在一种实施方案中，将感兴趣的单细胞用编码本文所述的碱基编辑系统的一个或多个核酸分子与少量的编码受体(例如，GFP)的载体一起转染、转导或以其他方式修饰。这些细胞可以是本领域已知的任何细胞，包括免疫细胞(例如，T细胞或NK细胞)，或永生化的人类细胞系诸如293T、K562或U20S。或者，可使用原代细胞(例如，人类)。细胞也可获自受试者或个体，诸如来自组织活检样品、手术、血液、徐诶请或其他生物学流体。此类细胞可以与最终的细胞靶标相关。

[1876] 可使用病毒载体执行递送。在一个实施方案中，可使用液体转染(诸如Lipofectamine或Fugene)或通过电穿孔执行转染。转染后，可通过荧光显微镜或通过流式细胞术确定受体(例如，GFP)的表达，以证实一致且高水平的转染。这些初步转染可以由不同的碱基编辑组成，以确定哪些编辑组合具有最大的活性。系统可包含一种或多种不同的载体。在一种实施方案中，针对所希望的细胞类型的表达对碱基编辑器进行密码子优化，该细胞类型优先是真核细胞，优选哺乳动物细胞或人类细胞。

[1877] 如本文所述评估和碱基编辑器的活性，即，通过将细胞基因组测序以监测靶标序列中的改变。对于Sanger测序，将经纯化的PCR扩增子克隆到质粒主链中，转化，小规模制备并使用单引物测序。测序可以使用下一代测序(NGS)技术执行。当使用下一代测序时，扩增子可以是300-500bp，具有不对称设置的预期切断位点。在PCR后，可以将下一代测序衔接子和条码(例如，Illumina多重衔接子和指标物)添加到扩增子的末端，例如，用于高通量测序(例如，在Illumina MiSeq上)。可以选择在初始测试中诱导最高水平的靶标特异性改变的融合蛋白进行进一步评定。

[1878] 在特定实施方案中，使用核碱基编辑器来靶向感兴趣的多核苷酸。在一种实施方案中，本发明的核碱基编辑器被递送至细胞(例如，免疫细胞(例如，T细胞或NK细胞))，与用于靶向细胞基因组内一个或多个感兴趣的核酸序列的一个或多个向导RNA协同使用，从而改变基因(例如，CD2)。在一些实施方案中，通过一个或多个向导RNA将碱基编辑器靶向至一个或多个感兴趣的基因(例如，CD2、TRAC、B2M、CIITA、TRBC1、TRBC2、PD-1、CD52)。在一些实施方案中，对一种或多种感兴趣基因的序列的一个或多个编辑降低或消除由该基因编码的蛋白质在宿主细胞(例如，免疫细胞(例如，T细胞或NK细胞))内的表达。在一些实施方案中，在宿主细胞(例如，免疫细胞(例如，T细胞或NK细胞))内，由一种或多种感兴趣基因(例如，CD2)编码的一种或多种蛋白质的表达被完全消除或消除。

[1879] 在一些实施方案中，宿主细胞是哺乳动物细胞。在一些实施方案中，宿主细胞是人

类细胞。在一些实施方案中,宿主细胞是免疫细胞。在一些实施方案中,免疫细胞是T细胞。在一些实施方案中,免疫细胞是NK细胞。在一些实施方案中,将一个或多个编辑引入选自CD2、CD3、CD5、CD7、CD52、B2M、CIITA、TRBC1、TRBC2、TRAC和PD-1或其组合的一种或多种基因中。在一些实施方案中,将一个或多个编辑引入CD2基因中。在一些实施方案中,将一个或多个编辑引入CD5基因中。在一些实施方案中,将一个或多个编辑引入CD7基因中。在一些实施方案中,将一个或多个编辑引入CD2、CD5、TRAC和PD_1基因中。在一些实施方案中,将一个或多个编辑引入CD5、CD5、TRAC和PD_1基因中。在一些实施方案中,将一个或多个编辑引入CD7、CD3、CD52和PD_1基因中。

[1880] 核碱基编辑器和gRNA的基于核酸的递送

[1881] 可以通过本领域已知方法或如本文所述将根据本公开的编码胞苷或腺苷脱氨酶的核酸给药至受试者或在体外或体内递送至细胞内。例如,可以通过例如载体(例如,病毒或非病毒载体)、基于非载体的方法(例如,使用裸DNA、DNA复合物、脂质纳米颗粒)或其组合递送胞苷或腺苷脱氨酶。

[1882] 编码胞苷或腺苷脱氨酶核碱基编辑器的核酸可以作为裸DNA或RNA直接递送至细胞(例如,免疫细胞),例如通过转染或电穿孔的手段,或可以缀合至促进被靶标细胞吸收的分子(例如,N-乙酰基半乳糖胺)。也可使用核酸载体,诸如本文所述的载体。在特定实施方案中,多核苷酸,例如,编码碱基编辑器或其功能性组分的mRNA,可以与本文所述多个向导RNA的组合共电穿孔。

[1883] 核酸载体可包含一个或多个编码本文所述融合蛋白的结构域的序列。载体也可包含编码单个肽(例如,用于核定位、核仁定位或线粒体定位)的序列,与编码蛋白的序列缔合(例如,插入至该序列内或与该序列融合)。作为一个示例,核酸载体可包括包括一个或多个核定位序列(例如,来自SV40的核定位序列)的Cas9编码序列,和一个或多个脱氨酶。

[1884] 核酸载体也可包括任何合适数量的调节/控制元件,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)。这些元件是本领域中周知的。

[1885] 根据本公开的核酸载体包括重组病毒载体。示例性病毒载体在上文详述。也可使用本领域已知的其他病毒载体。此外,病毒颗粒可用来递送核酸和/或肽形式的基因组编辑系统组分。例如,“空”病毒颗粒可以经组装以含有任何合适的载舱。病毒载体和病毒颗粒也可经工程改造以并入靶向配体,从而改变靶标组织特异性。

[1886] 除了病毒载体外,非病毒载体也可用来递送编码根据本公开的基因组编辑系统的核酸。一个重要的非病毒核酸载体类别是纳米颗粒,其可以是有机的或无机的。纳米颗粒是本领域中周知的。任何合适的纳米粒子设计可以用来递送基因组编辑系统组分或编码此类组分的核酸。例如,在本公开的某些实施方案中,有机(例如,脂质和/或聚合物)纳米颗粒可适用于用作递送媒介物。用于纳米颗粒调配物和/或基因转移中的示例性脂质显示在表18(下)中。

[1887] 表18

[1888]

用于基因转移的脂质		
脂质	缩写	特征
1,2-二油酰-sn-甘油-3-磷脂酰胆碱	DOPC	辅助
1,2-二油酰-sn-甘油-3-磷脂酰乙醇胺	DOPE	辅助
胆固醇		辅助
N-[1-(2,3-二油酰氧基)丙基]N,N,N-三甲基氯化铵	DOTMA	阳离子性
1,2-二油酰氧基-3-三甲基铵-丙烷	DOTAP	阳离子性
二-十八烷基酰胺基甘氨酰基精胺	DOGS	阳离子性
N-(3-氨基丙基)-N,N-二甲基-2,3-双(十二烷氧基)-1-丙烷溴化铵	GAP-DLRIE	阳离子性
鲸腊基三甲基溴化铵	CTAB	阳离子性
6-月桂氧基己基鸟氨酸酯	LHON	阳离子性
1-(2,3-二油酰氧基丙基)-2,4,6-三甲基吡啶鎓	2Oc	阳离子性
2,3-二油酰氧基-N-[2(精胺甲酰胺基-乙基]-N,N-二甲基-1-丙烷三氟乙酸铵	DOSPA	阳离子性
1,2-二油烯基-3-三甲基铵-丙烷	DOPA	阳离子性
N-(2-羟基乙基)-N,N-二甲基-2,3-双(十四烷氧基)-1-丙烷溴化铵	MDRIE	阳离子性
二肉豆蔻酰氧基丙基甲基羟乙基溴化铵	DMRI	阳离子性
3β-[N-(N',N'-二甲基氨基乙烷)-氨基甲酰基]胆固醇	DC-Chol	阳离子性
双胍-二乙三胺-胆固醇	BGTC	阳离子性
1,3-二油酰氧基-2-(6-羧基-精胺基)-丙酰胺	DOSPER	阳离子性
二甲基十八烷基溴化铵	DDAB	阳离子性

[1889]

用于基因转移的脂质		
脂质	缩写	特征
二-十八烷基酰胺基甘氨酰基亚精胺	DSL	阳离子性
外消旋-[(2,3-二-十八烷氧基丙基)(2-羟基乙基)]-二甲基氯化铵	CLIP-1	阳离子性
外消旋-[2(2,3-二-十六烷氧基丙基-氧甲氧基)乙基]三甲基溴化铵	CLIP-6	阳离子性
乙基二肉豆蔻酰基磷脂酰胆碱	EDMPC	阳离子性
1,2-二硬脂氧基-N,N-二甲基-3-氨基丙烷	DSDMA	阳离子性
1,2-二肉豆蔻酰基-三甲基铵丙烷	DMTAP	阳离子性
O,O'-二肉豆蔻酰基-N-赖氨酰基天冬氨酸酯	DMKE	阳离子性
1,2-硬脂酰氧基-sn-甘油-3-乙基磷酸胆碱	DSEPC	阳离子性
N-棕榈酰基 D-赤型-鞘氨苷基氨基甲酰基-精胺	CCS	阳离子性
N-叔丁基-NO-十四烷基-3-十四烷基氨基丙脒	diC14-amidine	阳离子性
十八烯酰氧基[乙基-2-十七烯基-3-羟乙基]氯化咪唑鎓	DOTIM	阳离子性
N1-胆甾烯氧羰基-3,7-二氮杂壬烷-1,9-二胺	CDAN	阳离子性
2-(3-[双(3-氨基-丙基)-氨基]丙基氨基)-N-二-十四氨基氨基甲酰基甲基-乙酰胺	RPR209120	阳离子性
1,2-二亚油氧基-3-二甲基氨基丙烷	DLinDMA	阳离子性
2,2-二亚油基-4-二甲基氨基乙基-[1,3]-二氧杂环戊烷	DLin-KC2-DMA	阳离子性
二亚油基-甲基-4-甲基氨基丁酸酯	DLin-MC3-DMA	阳离子性

[1890] 表19列出了用于基因转移和/或纳米颗粒调配物的示例性聚合物。

[1891] 表19

[1892]

用于基因转移的聚合物	
聚合物	缩写
聚乙二醇	PEG
聚乙烯亚胺	PEI
二硫代双(琥珀酰亚胺基丙酸酯)	DSP
二甲基-3,3'-二硫代双丙亚胺酸酯	DTBP
聚(乙烯亚胺)双氨基甲酸酯	PEIC
聚(L-赖氨酸)	PLL
经组氨酸修饰的 PLL	
聚(N-乙烯基吡咯烷酮)	PVP
聚(丙烯亚胺)	PPI
聚(酰胺基胺)	PAMAM
聚(酰胺基乙烯亚胺)	SS-PAEI
三亚乙基四胺	TETA
聚(β-氨基酯)	
聚(4-羟基-L-脯氨酸酯)	PHP
聚(烯丙基胺)	
聚(α-[4-氨基丁基]-L-乙醇酸)	PAGA
聚(D,L-乳酸-共-乙醇酸)	PLGA
聚(N-乙基-4-乙烯基溴化吡啶鎓)	
聚(磷腈)	PPZ
聚(磷酯)	PPE
聚(磷酰胺酯)	PPA
聚(N-2-羟基丙基甲基丙烯酰胺)	pHPMA
聚(甲基丙烯酸 2-(二甲基氨基)乙酯)	pDMAEMA
聚(2-氨基乙基亚丙基磷酸酯)	PPE-EA
脱乙酰壳多糖	
半乳糖基化的脱乙酰壳多糖	
N-脱乙酰化的脱乙酰壳多糖	
组蛋白	
胶原	
葡聚糖-精胺	D-SPM

[1893] 表20总结了用于编码本文所述融合蛋白的多核苷酸的的递送方法。

[1894] 表20

递送	载体/模式	递送至非分裂细胞	表达的持续时间	基因组整合	递送的分子类型
物理	（例如，电穿孔、基因枪、磷酸钙转染	是	瞬时	否	核酸和蛋白质
病毒	逆转录病毒	否	稳定	是	RNA
	慢病毒	是	稳定	是/否具有修饰	RNA
	腺病毒	是	瞬时	否	DNA
	腺相关病毒(AAV)	是	稳定	否	DNA
	疫苗病毒	是	非常短暂	否	DNA
	单纯疱疹病毒	是	稳定	否	DNA
非病毒	阳离子脂质体	是	瞬时	取决于递送的是什么	核酸和蛋白质
	聚合物纳米颗粒	是	瞬时	取决于递送的是什么	核酸和蛋白质
生物学非病毒	减毒细菌	是	瞬时	否	核酸
递送媒介物	工程改造的噬菌体	是	瞬时	否	核酸
	哺乳动物的病毒样颗粒	是	瞬时	否	核酸
	生物脂质体:红细胞血影和外泌体	是	瞬时	否	核酸

[1895]

[1896] 在另一方面,基因组编辑系统组分或编码此类组分的核酸(例如,核酸结合蛋白诸如,举例而言,Cas9或其变体)以及靶向感兴趣的基因组核酸序列的gRNA的递送,可以通过递送核糖核蛋白(RNP)至细胞而得以实现。RNP包含核酸结合蛋白,例如,Cas9,其与靶向gRNA复合。RNP可以使用已知方法递送至细胞,诸如电穿孔、核转染或阳离子脂质介导的方法,例如,如Zuris,J.A.等人,2015,Nat.Biotechnology,33(1):73-80所报导。RNP有利于在CRISPR碱基编辑系统中使用,特别是用于难以转染的细胞诸如原代细胞。此外,RNP也可以缓解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,CMV或EF1A,其可以用于CRISPR之中)没有被良好表达时。有利的是,RNP的使用不需要将外来DNA递送至细胞内。此外,因为包含核酸结合蛋白和gRNA复合物的RNP随时间推移被降解,RNP的使用具有限制脱靶效应的潜力。以类似于用于基于质粒的技术的方式,RNP可用来递送结合蛋白(例如,Cas9变体)和引导同源定向修复(HDR)。

[1897] 用于驱动碱基编辑器编码核酸分子表达的启动子可包括AAV ITR。这可能有利于消除对另外的启动子元件的需求,该元件占据载体内的空间。释放出来的另外空间可用于驱动另外元素的表达,诸如向导核酸或可选标志物。ITR活性相对较弱,所以它可以用来降低所选核酸酶过表达带来的潜在毒性。

[1898] 任何合适的启动子可用来驱动碱基编辑器(并且若适用,向导核酸)的表达。对于普遍表达,可使用的启动子包括CMV、CAG、CBh、PGK、SV40、铁蛋白重链或轻链等。对于大脑或其他CNS细胞表达,合适的启动子可包括:适用于所有神经元的SynapsinI,适用于兴奋性神经元的Camkiα,适用于GABA能神经元的GAD67或GAD65或VGAT等。对于肝细胞表达,合适的启动子包括白蛋白启动子。对于肺细胞表达,合适的启动子可包括SP-B。对于内皮细胞,合适的启动子可包括ICAM。对于造血细胞,合适的启动子可包括IFNβ或CD45。对于成骨细胞,合

适的启动子可包括OG-2。

[1899] 在一些实施方案中,本公开的碱基编辑器的尺寸足够小,以至于允许独立的启动子驱动在同一核酸分子内的碱基编辑器和相容向导核酸的表达。例如,载体或病毒载体可以包含可操作地链接至编码碱基编辑器的核酸的第一启动子以及可操作地链接至向导核酸的第二启动子。

[1900] 用来驱动向导核酸的表达的启动子可包括:Pol III启动子,诸如U6或H1。使用Pol II启动子和内含子匣来表达gRNA腺相关病毒(AAV)。

[1901] 在特定实施方案中,本发明的融合蛋白由病毒载体(例如,腺相关病毒(AAV)、AAV3、AAV3b、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAVrh8、AAV10及其变体)中存在的多核苷酸或任何病毒载体的合适衣壳蛋白编码。因此,在一些方面,本公开涉及融合蛋白的病毒递送。病毒载体的示例包括逆转录病毒载体(例如,Maloney鼠白血病病毒,MML-V)、腺病毒载体(例如,AD100)、慢病毒载体(基于HIV和FIV的载体)、疱疹病毒载体(例如,HSV-2)。

[1902] 在一些方面,本文所述的用于编辑免疫细胞中特异性基因的方法可用于在基因上修饰CAR-T细胞。此类CAR-T细胞和生产此类CAR-T细胞的方法描述在国际申请号PCT/US2016/060736、PCT/US2016/060734、PCT/US2016/034873、PCT/US2015/040660、PCT/EP2016/055332、PCT/IB2015/058650、PCT/EP2015/067441、PCT/EP2014/078876、PCT/EP2014/059662、PCT/IB2014/061409、PCT/US2016/019192、PCT/US2015/059106、PCT/US2016/052260、PCT/US2015/020606、PCT/US2015/055764、PCT/CN2014/094393、PCT/US2017/059989、PCT/US2017/027606和PCT/US2015/064269中,各自的内容以其整体并入本文。

[1903] 病毒载体

[1904] 本文所述的碱基编辑器因此可使用病毒载体递送。在一些实施方案中,本文所公开的碱基编辑器可编码在病毒载体中所包含的核酸上。在一些实施方案中,碱基编辑器系统的一个或多个组分可以被编码在一种或多种病毒载体上。例如,碱基编辑器和向导核酸可以编码在单个病毒载体上。在其他实施方案中,碱基编辑器和向导核酸编码在不同病毒载体上。在每种情况下,碱基编辑器和向导核酸可以各自可操作地链接至启动子和终止子。被编码在病毒载体上的组分的组合可以通过所选病毒载体的载荷尺寸约束条件确定。

[1905] 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,该过程将病毒靶向至培养物或宿主中的特异性细胞,并将病毒有效载荷运输到细胞核或宿主细胞基因组。病毒载体可以直接给药至培养物中的细胞、患者(体内),或它们可以处理细胞(体外),并且经修饰的细胞可任选地给药至患者(间接体内)。传统的基于病毒的系统将会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。可以使用逆转录病毒、慢病毒和腺相关病毒基因转移方法整合到宿主基因组中,经常导致被插入的转基因的长期表达。此外,已经在很多不同的细胞类型和靶标组织中观察到了高转导效率。

[1906] 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,Maloney鼠白血病病毒,MML-V)、疱疹病毒(例如,HSV-2)和腺相关病毒(AAV),或者其他质粒或病毒载体类型,特别是使用来自美国专利号8,454,972(适用于腺病毒的调配物、剂量)、美国专利号8,404,658(适用于AAV的调配物、剂量)和美国专利号5,846,946(适用于DNA质粒的调配物、剂量)和来自临床试验以及关于前述慢病毒AAV和腺病毒的临床

试验的出版物的调配物和剂量。例如,对于AAV,给药途径、调配物和剂量可以如美国专利号8,454,972和牵涉AAV的临床试验中所用。对于腺病毒,给药途径、调配物和剂量可以如美国专利号8,404,658和牵涉腺病毒的临床试验中所用。对于质粒递送,给药途径、调配物和剂量可以如美国专利号5,846,946和牵涉质粒的临床研究中所用。剂量可以基于平均70kg的个体(例如,成年男性人类)或自其外推,并且可针对不同体重和物种的患者、受试者、哺乳动物调整。给药频率处于医疗或兽医从业者(例如,医师、兽医)的能力范围内,取决于常规因素(包括患者或受试者的年龄、性别、一般健康状况、其他情况)和正在处理的特定病症或症状。病毒载体可以注射至感兴趣的组织内。对于细胞类型特异性碱基编辑,碱基编辑器和任选使用的向导核酸的表达可以由细胞类型特异性启动子驱动。

[1907] 逆转录病毒的向性可以通过加入外来包膜蛋白来改变,从而扩大靶标细胞的潜在靶标细胞群。慢病毒载体是能够转导或感染非分裂细胞并典型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会取决于靶标组织。逆转录病毒载体由顺式作用的长末端重复序列构成,该重复序列对外来序列的封装能力为至多6-10kb。最小的顺式作用LTR足以复制并封装载体,该载体然后用于将治疗基因整合到靶标细胞中,以提供永久性转基因表达。广泛使用的逆转录病毒载体包括基于鼠白血病病毒(MuLV)、长臂猿白血病病毒(GaLV)、猴免疫缺陷病毒(SIV)、人类免疫缺陷病毒(HIV)的那些及其组合(参见,例如,Buchscher等人,J.Virol.66:2731-2739(1992);Johann等人,J.Virol.66:1635-1640(1992);Sommnerfelt等人,Virol.176:58-59(1990);Wilson等人,J.Virol.63:2374-2378(1989);Miller等人,J.Virol.65:2220-2224(1991);PCT/US94/05700)。

[1908] 逆转录病毒载体,尤其是慢病毒载体,可能需要多核苷酸序列小于用于有效整合到靶标细胞内的给定长度。例如,与使用较小尺寸的那些相比,长度大于9kb的逆转录病毒载体可导致低病毒滴度。在一些方面,本公开的碱基编辑器的尺寸足以能有效封装并经由逆转录病毒载体递送至靶标细胞。在一些实施方案中,即便当与向导核酸和/或可靶向的核酸酶系统的其他组分一起被表达时,碱基编辑器的尺寸也允许进行有效封装和递送。

[1909] 封装细胞典型永安里形成能够感染宿主细胞的病毒颗粒。此类细胞包括封装慢病毒的293细胞,以及封装逆转录病毒的ψ2细胞或PA317细胞。用于基因疗法中的病毒载体经常通过生产将核酸载体封装到病毒颗粒内的细胞系来生成。载体典型含有封装和后续整合到宿主内所需的最小病毒序列,其他病毒序列由用于待表达的多核苷酸的表达匣替换。缺失的病毒功能典型通过封装细胞系反式供给。例如,用于基因疗法中的腺相关病毒(“AAV”)典型仅具备来自AAV基因组的ITR序列,它们是封装和整合到宿主基因组内所需要的。病毒DNA可以封装在细胞系中,该细胞系含有编码其他AAV基因(即rep和cap,但缺乏ITR序列)的辅助质粒。细胞系也可以用腺病毒作为辅助进行转染。辅助病毒可促进AAV载体的复制和来自辅助质粒的AAV基因的表达。在一些情况下,辅助质粒由于缺乏TR序列而不以显著量被封装。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。

[1910] 在优选瞬时表达的应用中,可使用基于腺病毒的系统。基于腺病毒的载体能够在很多细胞类型中实现非常高的转导效率且不需要细胞分裂。已经使用此类载体获得了高滴度和高水平的表达。该载体可在相对简单的系统中以大数量生产。腺相关病毒(“AAV”)载体也可用来施用靶标核酸转导细胞,例如,在核酸和肽的体外生产中,以及用于体内和间接体内基因疗法过程(参见,例如,West等人,Virology160:38-47(1987);美国专利号4,797,

368;WO 93/24641;Kotin,Human Gene Therapy 5:793-801(1994);Muzyczka,J.Clin.Invest.94:1351(1994))。重组AAV载体的构建在大量出版物中描述,包括美国专利号5,173,414;Tratschin等人,Mol.Cell.Biol.5:3251-3260(1985);Tratschin,等人,Mol.Cell.Biol.4:2072-2081(1984);Hermonat&Muzyczka,PNAS 81:6466-6470(1984);和Samulski等人,J.Virol.63:03822-3828(1989)。

[1911] AAV是小的、单链DNA依赖性病毒,属于细小病毒科。4.7kb野生型(wt)AAV基因组由两个基因构成,这两个基因分别编码四个复制蛋白质和三个质粒蛋白质,并且每一侧各具有145-bp反向末端重复序列(ITR)。病毒体由三个质粒蛋白Vp1、Vp2和Vp3构成,它们以1:1:10的比率从同一开放阅读框产生但来自差异剪接位点(Vp1)和替代性翻译起始位点(分别为Vp2和Vp3)。Vp3是病毒体中丰度最高的亚基并且参与到定义该病毒向性的细胞表面处的受体识别中。已经在Vp1的独特的N端鉴定出了在病毒感染性中发挥作用的磷脂酶结构域。

[1912] 类似于wt AAV,重组AAV(rAAV)利用顺式作用145-bp ITR填充病毒转基因匣的侧翼,提供至多4.5kb用于封装外来DNA。转染后,rAAV可表达本发明的融合蛋白并且通过以环状头到尾连环体作为游离基因存在而保持不整合到宿主基因组内。尽管存在大量在体外和体内成功使用该系统的rAAV示例,但当基因的编码序列的长度在尺寸上大于或等于wt AAV基因组时,其有限的封装能力仍限制了AAV介导的基因疗法的使用。

[1913] 可基于应用来选择病毒载体。例如,对于体内基因递送,AAV可能比其他病毒载体有利。在一些实施方案中,AAV导致低毒性,者可能是由于纯化方法不需要对细胞颗粒进行可能激活免疫应答的超离心。在一些实施方案中,AAV不整合到宿主基因组中,因此导致插入突变的可能性较低。腺病毒是常用的疫苗,因为它们诱导强免疫应答。病毒载体的封装能力可能限制可以被封装在该载体内的碱基编辑器的尺寸。

[1914] AAV的封装能力为约4.5Kb或4.75Kb,包括两个145碱基的反向末端重复序列(ITR)。这意味着所公开的碱基编辑器以及启动子和转录终止子可以配合到单个病毒载体中。大于4.5或4.75Kb的构建体可导致显著降低的病毒产生。例如,SpCas9相当大,该基因自身超过4.1Kb,这使得它难以封装到AAV中。因此,本公开的实施方案包括利用在长度上比传统碱基编辑器短的所公开的碱基编辑器T。在一些示例中,碱基编辑器小于4kb。所公开的碱基编辑器可能小于4.5kb、4.4kb、4.3kb、4.2kb、4.1kb、4kb、3.9kb、3.8kb、3.7kb、3.6kb、3.5kb、3.4kb、3.3kb、3.2kb、3.1kb、3kb、2.9kb、2.8kb、2.7kb、2.6kb、2.5kb、2kb或1.5kb。在一些实施方案中,所公开的碱基编辑器是4.5kb或更小的长度。

[1915] AAV可以是AAV1、AAV2、AAV5或其任意组合。可以根据目标细胞选择AAV的类型;例如,针对靶向脑或神经细胞,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;针对靶向心脏组织,可以选择AAV4。AAV8可用于递送至肝脏。关于这些细胞的某些AAV血清型的列表可见于Grimm,D.et al,J.Virol.82:5887-5911(2008)中。

[1916] 在一些实施方案中,使用慢病毒载体来转导经嵌合抗原受体修饰的免疫细胞(例如,T细胞或NK细胞)。慢病毒是复合逆转录病毒,其具有感染并在有丝分裂细胞和有丝分裂后细胞两者中表达其基因的能力。最常见的慢病毒是人类免疫缺陷病毒(HIV),其使用其他病毒的包膜糖蛋白来靶向广泛范围的细胞类型。

[1917] 慢病毒可如下制备。在克隆pCasES10(其含有慢病毒转移质粒主链)后,在用10%胎牛血清且不使用抗生素在DMEM转染前一天,将低传代(p=5)的HEK293FT接种到T-75烧瓶

中直到50％融合度。20小时后,将培养基更换为OptiMEM(无血清)培养基,并在4小时后进行转染。细胞用10μg的慢病毒转移质粒(pCasES10)和以下封装质粒转染:5μg的pMD2.G(VSV-g甲型)和7.5μg的psPAX2(gag/pol/rev/tat)。转染可以在4mL OptiMEM中使用阳离子脂质递送剂(50μl Lipofectamine2000和100μl Plus试剂)进行。6小时后,将培养基更换为具有10％胎牛血清的无抗生素DMEM。这些方法在细胞培养期间使用血清,但优选无血清方法。

[1918] 慢病毒可如下纯化。在48小时后收获病毒上清液。上清液首先进行碎片澄清,并通过0.45μm低蛋白质结合(PVDF)过滤器过滤。然后将它们在超离心机中以24,000rpm旋转2小时。在4℃将病毒沉淀物重新悬浮在50μl的DMEM中过夜。然后将它们分为等分小样并立即在-80℃冷冻。

[1919] 在另一实施方案中,也预期基于马传染性贫血病毒(EIAV)的最小非灵长类慢病毒载体。在另一实施方案中,预期通过视网膜下注射递送RetinoStat®,一种基于马传染性贫血病毒的慢病毒基因治疗载体,其表达血管抑制蛋白内皮抑素和血管抑素。在另一实施方案中,预期使用自灭活慢病毒载体。

[1920] 该系统的任何RNA,例如向导RNA或碱基编辑器编码mRNA,可以RNA的形式递送。碱基编辑器编码mRNA可使用体外转录来生成。例如,核酸酶mRNA可使用含有以下元件的PCR匣来合成:T7启动子、任选的kozak序列(GCCACC)、核酸酶序列、和3'UTR诸如来自β球蛋白-polyA尾部的3'UTR。该匣可用于通过T7核酸酶进行的转录。向导多核苷酸(例如,gRNA)也可使用体外转录从含有T7启动子以及后面的序列"GG"和向导多核苷酸序列的匣转录。

[1921] 为了增强表达并降低可能的毒性,碱基编辑器编码序列和/或向导核酸可经修饰以包括一个或多个经修饰的核苷酸,例如,使用假-U或5-甲基-C。

[1922] AAV载体的封装能力小使得大量超出这一尺寸的基因的递送和/或大生理调节元件的使用具有挑战性。这些挑战例如可以通过将待递送的蛋白分为两个或更多个片段解决,其中N端片段融合至分体内含肽-N且C端片段融合至分体内含肽-C。这些片段随后被封装到两个或更多个AAV载体内。如本文所用,"内含肽"是指自剪接蛋白质内含子(例如,肽),其连接侧翼N端和C端外显肽(例如,待接合的片段)。使用某些内含肽来接合异源蛋白片段在例如Wood等人,J.Biol.Chem.289(21);14512-9(2014)中描述。例如,当融合至独立的蛋白质片段时,内含肽IntN和IntC彼此识别,将其自身剪出并同时连接它们所融合至的蛋白质片段的侧翼N端和C端外显肽,从而从两个蛋白质片段重构全长蛋白质。其他合适的内含肽对于本领域技术人员将是显而易见的。

[1923] 本发明的融合蛋白的片段的长度可变。在一些实施方案中,蛋白质片段的长度在2个氨基酸到约1000个氨基酸的范围。在一些实施方案中,蛋白质片段的长度在约5个氨基酸到约500个氨基酸的范围。在一些实施方案中,蛋白质片段的长度在约20个氨基酸到约200个氨基酸的范围。在一些实施方案中,蛋白质片段的长度在约10个氨基酸到约100个氨基酸的范围。合适的其他长度的蛋白质片段对于本领域技术人员将是显而易见的。

[1924] 在一种实施方案中,通过将大的转基因盒剪接为独立的两半(5'和3'末端,或头部和尾部)而产生双AAV载体,其中该盒的每一半都被封装到单个AAV载体(尺寸<5kb)中。全长转基因表达匣的再组装随后在通过双重AAV载体对相同细胞进行共转染然后进行以下项时实现:(1)5'与3'基因组(双重AAV重叠载体)之间的同源重组(HR);(2)5'和3'基因组(双重AAV反式剪接载体)的ITR介导的尾到头连环化;或(3)这两种机制(双重AAV杂交载体)的组

合。在体内使用双重AAV载体导致全长蛋白质的表达。使用双重AAV载体平台代表用于尺寸>4.7kb的转基因的有效且可用的基因转移策略。

[1925] 内含肽

[1926] 内含肽(干预蛋白质)是见于多种不同生物体中的自加工结构域,其进行倍称为蛋白质剪接的过程。蛋白质剪接是由肽键的切割和形成所构成的多步生化反应。尽管蛋白质剪接的内源性底物是见于含内含肽生物体中的蛋白质,但内含肽也可用来在化学上操纵几乎任何多肽主链。

[1927] 在蛋白质剪接中,内含肽通过切割两个肽键将其自身从前体多肽切除,从而经由形成新肽键来连接侧翼外显肽(外部蛋白质)。这一重排发生在翻译后(或可能在翻译时)。内含肽介导的蛋白质剪接自发地发生,仅需要内含肽结构域的折叠。

[1928] 约5%的内含肽是分体内含肽,其被转录并翻译为两个独立的多肽,N-内含肽和C-内含肽,各自融合至一个外显肽。翻译时,内含肽片段自发地且非共价地组装为经典内含肽结构,以进行反式蛋白质剪接。蛋白质剪接的机制牵涉一系列酰基转移反应,这些翻译导致两个肽键在内含肽-外显肽连接处切割,并在N-外显肽和C-外显肽之间形成一个新的肽键。这一过程通过接合N-外显肽与内含肽N端的肽键的激活来启动。几乎全部内含肽均在其N端具有半胱氨酸或丝氨酸,该氨基酸附接至C端N-外显肽残基的羰基碳。这一N到O/S的酰基位移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一起促成,导致线性(硫代)酯中间体的形成。接着,这一中间体经历通过第一个C-外显肽残基(+1)的亲核性攻击进行的反式-(硫代)酯化,该残基是半胱氨酸、丝氨酸或苏氨酸。所得支链(硫代)酯中间体通过独特的转化分解,该转化是内含肽的高度保守的C端天冬酰胺的环化。这一过程由组氨酸(见于高度保守的HNF基序中)和倒数第二个组氨酸促成,并且也可可能牵涉天冬氨酸盐。这一琥珀酰亚胺形成反应将内含肽从反应性复合物切除并留下通过非肽链接附接的外显肽。这一结构快速重排为不依赖于内含肽模式的稳定肽键T。

[1929] 在一些实施方案中,核酸酶(例如,Cas9)的一部分或片段融合至内含肽。核酸酶可以融合至内含肽的N端或C端。在一些实施方案中,融合蛋白的一部分或片段融合至内含肽并且融合至AAV衣壳蛋白。内含肽、核酸酶和衣壳蛋白可以任何排列融合在一起(例如,核酸酶-内含肽-衣壳、内含肽-核酸酶-衣壳、衣壳-内含肽-核酸酶等)。在一些实施方案中,碱基编辑器(例如,ABE、CBE)的N端片段融合至分体内含肽-N,且C端片段融合至分体内含肽-C。这些片段随后被封装到两个或更多个AAV载体中。在一些实施方案中,内含肽的N端融合之融合蛋白的C端,且内含肽的C端融合至AAV衣壳蛋白的N端。

[1930] 在一种实施方案中,利用内含肽来接合接枝到AAV衣壳蛋白上的胞苷或腺苷脱氨酶碱基编辑器蛋白的片段或部分。使用某些内含肽来接合异源蛋白片段在例如Wood等人,J.Biol.Chem.289(21);14512-9(2014)中描述。例如,当融合至独立的蛋白质片段时,内含肽IntN和IntC彼此识别,将其自身剪出并同时连接它们所融合至的蛋白质片段的侧翼N端和C端外显肽,从而从两个蛋白质片段重构全长蛋白质。其他合适的内含肽对于本领域技术人员将是显而易见的。

[1931] 在一些实施方案中,ABE在SpCas9的所选区内的Ala、Ser、Thr或Cys残基处分体为N端片段和C端片段。这些区对应于通过Cas9晶体结构分析鉴定的环区。每个片段的N端融合至内含肽-N,且每个片段的C端融合至内含肽-C,融合位点是氨基酸位置S303、T310、T313、

S355、A456、S460、A463、T466、S469、T472、T474、C574、S577、A589和S590，这些位点在下述序列中以加粗大写字母表示。

[1932]

```
   1 mdkkysigld igtnsvgwav itdeykvpsk kfkvlgntdr hsikknliga llfdsgetae
  61 atrlkrtarr rytrrknric ylqeifsnem akvddsffhr leesflveed kkherhpifg
 121 nivdevayhe kyptiyhlrk klvdstdkad lrliylalah mikfrghfli egdlnpdnsd
 181 vdklfiqlvq tynqlfeenp inasgvdaka ilsarlsksr rlenliaqlp gekknglfgn
 241 lialslgltp nfksnfdlae daklqlskdt ydddldnlla qigdqyadlf laaknlsdai
 301 llSdilrvnT eiTkaplsas mikrydehhq dltllkalvr qqlpekykei ffdqSkngya
 361 gyidggasqe efykfikpil ekmdgteell vklnredllr kqrtfdngsi phqihlgelh
 421 ailrrqedfy pflkdnreki ekiltfripy yvgplArgnS rfAwmTrkSe eTiTpwnfee
 481 vvdkgasaqs fiermtnfdk nlpnekvlpk hsllyeyftv yneltkvkyv tegmrkpafl
 541 sgeqkkaivd llfktnrkvt vkqlkedyfk kieCfdSvei sgvedrfnAS lgtyhdllki
 601 ikdkdfldne enedilediv ltltlfedre mieerlktya hlfddkvmkq lkrrrytgwg

 661 rlsrklingi rdkqsgktil dflksdgfan rnfmqlihdd sltfkediqk aqvsgqgdsl
 721 hehianlags paikkgilqt vkvvdelvkv mgrhkpeniv iemarenqtt qkgqknsrer
 781 mkrieegike lgsqilkehp ventqlqnek lylyylqngr dmyvdqeldi nrlsdydvdh
 841 ivpqsflkdd sidnkvltrs dknrgksdnv pseevvkkmk nywrqllnak litqrkfdnl
 901 tkaergglse ldkagfikrq lvetrqitkh vaqildsrmn tkydendkli revkvitlks
 961 klvsdfrkdf qfykvreinn yhhahdayln avvgtalikk ypklesefvy gdykvydvrk
1021 miakseqeig katakyffys nimnffktei tlangeirkr plietngetg eivwdkgrdf
1081 atvrkvlsmp qvnivkktev qtggfskesi lpkrnsdkli arkkdwdpkk yggfdsptva
1141 ysvlvvakve kgkskklksv kellgitime rssfeknpid fleakgykev kkdliiklpk
1201 yslfelengr krmlasagel qkgnelalps kyvnflylas hyeklkgspe dneqkqlfve
1261 qhkhyldeii eqisefskrv iladanldkv lsaynkhrdk pireqaenii hlftltnlga
 1321 paafkyfdtt idrkrytstk evldatlihq sitglyetri dlsqlggd
```

[1933]

[1934] 核碱基编辑器的使用

[1935] 与肿瘤细胞相关的靶标抗原也可以在健康免疫细胞上表达。因此，使用经修饰的免疫细胞(例如，CAR-T细胞)来靶向肿瘤细胞上的抗原也可导致杀死也表达该抗原的健康免疫细胞，这称为同类相杀或自杀。使用经修饰的免疫细胞与可降低或消除靶标抗原(例如，CD2)在经修饰的免疫细胞上的表达的碱基编辑器组合来提供同类相杀抗性，为使用在治疗性和基础研究中的应用进行基因校正提供了新的策略。

[1936] 本公开提供用于生产抗同类相杀的经修饰的免疫细胞的方法。在一些实施方案中，提供一种方法，其包括使用核碱基编辑器(例如，腺苷脱氨酶碱基编辑器或胞苷脱氨酶碱基编辑器)和一种或多种向导RNA编辑获自健康受试者或供体的免疫细胞(例如，T细胞或NK细胞)以降低或消除靶标抗原(例如，CD2)在该免疫细胞(例如，T细胞或NK细胞)上的表达。然后，用被引导至靶标抗原(例如，CD2)的嵌合抗原受体转导免疫细胞。所得经修饰的免疫细胞(例如，CAR-T)是抗同类相杀的，并且能够靶向肿瘤细胞而不靶向健康免疫细胞。在一些实施方案中，抗同类相杀的经修饰的免疫细胞如实施例2中提供的生产。在一些实施方案中，提供一种生产抗同类相杀的CD2 CAR-T细胞的方法。

[1937] 本公开也提供使用本文所提供的免疫细胞治疗被诊断为患有疾病(例如，T细胞或NK细胞恶性病变)的受试者的方法。在一些实施方案中，提供一种方法，其包括向患有或易患疾病(例如，T细胞或NK细胞恶性病变)的受试者给药有效量的如本文所提供生产的经修饰的免疫细胞。本公开提供用于治疗与可通过脱氨酶介导的基因编辑来校正的靶标抗原相关的T细胞或NK细胞恶性病变的方法。基于本公开，可使用本文所提供的策略和经修饰的免疫细胞治疗的合适疾病对于本领域技术人员将是显而易见的。

[1938] 本文提供使用碱基编辑器或碱基编辑器系统来编辑靶标核苷酸序列中的核碱基。在一些实施方案中，碱基编辑器系统(例如，包含脱氨酶、Cas9结构域和一个或多个向导RNA)的活性导致剪接位点的中断或导致提前STOP密码子被引入靶标核苷酸序列内。

[1939] 在一些实施方案中,本文提供的脱氨酶(例如,腺苷脱氨酶或胞苷脱氨酶)能够将DNA的脱氧腺苷残基脱氨基。本公开的其他方面提供融合蛋白,该融合蛋白包含脱氨酶(例如,腺苷脱氨酶或胞苷脱氨酶)和能够结合至特异性核苷酸序列的结构域(例如,Cas9或Cpf1蛋白)。例如,腺苷可以被转变为肌苷残基,而肌苷残基通常与胞苷残基进行碱基配对。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DNA的体外靶向编辑,例如,用于生成突变型细胞或动物;用于引入靶向突变,例如,用于细胞中的基因缺陷的间接体内校正,例如,在获自受试者且后续被重新引入同一受试者或另一受试者体内的细胞中;以及用于靶向突变的体内引入,例如,可以使用本文所提供的核碱基编辑器来进行基因缺陷的校正或将引入的疾病相关基因中的去活性突变处理为G到A或T到C突变。本公开提供经修饰的免疫细胞、融合蛋白、核酸、载体、组合物、方法、试剂盒、系统等,它们利用脱氨酶和核碱基编辑器。

[1940] 用于编辑核酸的方法

[1941] 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,该方法是用于编辑核酸的核碱基(例如,双链DNA序列的碱基对)的方法。在一些实施方案中,该方法包括以下步骤:a)使核酸(例如,双链DNA序列)的靶标区与包含碱基编辑器(例如,Cas9结构域融合至胞苷或腺苷脱氨酶)和向导核酸(例如,gRNA)的复合物接触,其中靶标区包含被靶向的核碱基对,b)诱导所述靶标区的链分离;c)将所述靶标区的单链中的靶标核碱基对的第一核碱基转变为第二核碱基,以及d)切断所述靶标区的不超过一条链,其中与第一核碱基互补的第三核碱基被替换为与第二核碱基的第四核碱基。在一些实施方案中,该方法导致在核酸中形成少于20%的插入缺失。应了解,在一些实施方案中,省略了步骤b。在一些实施方案中,该方法导致少于19%、18%、16%、14%、12%、10%、8%、6%、4%、2%、1%、0.5%、0.2%或少于0.1%的插入缺失。在一些实施方案中,该方法进一步包括将第二核碱基替换为与第四核碱基互补的第五核碱基,从而生成预期的经编辑的碱基对(例如,C·G到T·A)。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。

[1942] 在一些实施方案中,靶标核苷酸中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,被切断的单链(切口链)杂交至向导核酸。在一些实施方案中,被切断的单链与包含第一核碱基的链相对。在一些实施方案中,碱基编辑器包含Cas9结构域。在一些实施方案中,碱基编辑器保护或结合未被编辑的链。在一些实施方案中,碱基编辑器包含切口酶活性。在一些实施方案中,预期经编辑的碱基对在PAM位点上游。在一些实施方案中,预期经编辑的碱基对在PAM位点上游1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个核苷酸。在一些实施方案中,预期经编辑的碱基对在PAM位点下游。在一些实施方案中,预期经编辑的碱基对在PAM位点下游1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个核苷酸。在一些实施方案中,该方法不需要标准(例如,NGG)PAM位点。在一些实施方案中,核碱基编辑器包含连接子。在一些实施方案中,连接子是1至25个氨基酸的长度。在一些实施方案中,连接子是5至20个氨基酸的长度。在一些实施方案中,连接子是10、11、12、13、14、15、16、17、18、19或20个氨基酸的长度。在一种实施方案中,连接子是32个氨基酸

的长度。在另一实施方案中,“长连接子”是至少60个氨基酸的长度。在其他实施方案中,连接子是约3至100个氨基酸之间的长度。在一些实施方案中,靶标区包含靶标窗口,其中靶标窗口包含靶标核碱基对。在一些实施方案中,靶标窗口包含1至10个核苷酸。在一些实施方案中,靶标窗口是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个核苷酸的长度。在一些实施方案中,靶标窗口是1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个核苷酸的长度。在一些实施方案中,预期经编辑的碱基对在靶标窗口内。在一些实施方案中,靶标窗口包含预期经编辑的碱基对。在一些实施方案中,该方法使用本文所提供的任何碱基编辑器执行。在一些实施方案中,靶标窗口是甲基化窗口。

[1943] 在一些实施方案中,本公开提供用于编辑核苷酸的方法。在一些实施方案中,本发明提供用于编辑双链DNA序列的核碱基对的方法。在一些实施方案中,该方法包括:a)使双链DNA序列的靶标区与包含碱基编辑器和向导核酸(例如,gRNA)的复合物接触,其中靶标区包含靶标核碱基对,b)诱导所述靶标区的链分离,c)将靶标区的单链中的所述靶标核碱基对的第一核碱基转变为第二核碱基,d)切断所述靶标区的不超过一条链,其中与第一核碱基互补的第三核碱基被替换为与第二核碱基互补的第四核碱基,并且第二核碱基被替换为与第四核碱基互补的第五核碱基,从而生成预期经编辑的碱基对,其中生成预期经编辑的碱基对的效率是至少5%。应了解,在一些实施方案中,省略了步骤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,通过本文所述方法进行的碱基编辑在任何特定基因位点可具有至少10%的碱基转变效率。在一些实施方案中,通过本文所述方法进行的碱基编辑在任何特定基因位点可具有至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%或至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%或至少95%、96%、97%、98%或至少99%的碱基转变效率。在一些实施方案中,通过本文所述方法进行的碱基编辑在任何特定基因位点可具有至少70%的碱基转变效率。在一些实施方案中,通过本文所述方法进行的碱基编辑在任何特定基因位点可具有至少80%的碱基转变效率。在一些实施方案中,通过本文所述方法进行的碱基编辑在任何特定基因位点可具有至少90%的碱基转变效率。

[1944] 在一些实施方案中,该方法造成少于19%、18%、16%、14%、12%、10%、8%、6%、4%、2%、1%、0.5%、0.2%或少于0.1%的插入缺失。在一些实施方案中,在靶标核苷酸处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,被切断的单链杂交至向导核酸。在一些实施方案中,被切断的单链与包含第一核碱基的链相对。在一些实施方案中,核碱基编辑器包含切口酶活性。在一些实施方案中,预期经编辑的碱基对在PAM位点上游。在一些实施方案中,预期经编辑的碱基对在PAM位点上游1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个核苷酸。在一些实施方案中,预期经编辑的碱基对在PAM位点下游。在一些实施方案中,预期经编辑的碱基对在PAM位点下游1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个核苷酸。在一些实施方案中,该方法不需要标准(例如,NGG)PAM位点。在一些实施方案中,核碱基编辑器包含连接子。在一些实施方案中,连接子是1至25个氨基酸的长度。在一些实施方案中,连接子是5至20个氨基酸的长度。在一些

实施方案中,连接子是10、11、12、13、14、15、16、17、18、19或20个氨基酸的长度。例如,在一些实施方案中,靶标区包含靶标窗口,其中靶标窗口包含靶标核碱基对。在一些实施方案中,靶标窗口包含1至10个核苷酸。在一些实施方案中,靶标窗口是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个核苷酸的长度。在一些实施方案中,靶标窗口是1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个核苷酸的长度。在一些实施方案中,预期经编辑的碱基对出现在靶标窗口内。在一些实施方案中,靶标窗口包含预期经编辑的碱基对。在一些实施方案中,核碱基编辑器是本文所提供的任一种碱基编辑器。

[1945] 药物组合物

[1946] 在一些方面,本发明提供一种药物组合物,其包含本文所述的经基因修饰的免疫细胞、碱基编辑器、融合蛋白或融合蛋白-向导多核苷酸复合物中的任一者。更具体地,本文提供药物组合物,其包含表达嵌合抗原受体(CAR)的经基因修饰的免疫细胞或此类免疫细胞群,其中所述经修饰的免疫细胞或其群具有至少一个经编辑的基因以提供同类相杀抗性,增强经修饰的免疫细胞的功能,或降低经修饰的免疫细胞的免疫阻遏或抑制,其中经编辑的基因的表达被敲除或敲低。在一些实施方案中,该至少一个经编辑的基因是CD3、CD5、CD7、CD33、CD123、TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M、PDCD1/PD1、CBLB、TGFBR2、ZAP70、NFATc1、TET2或其组合。在一些实施方案中,CD3、CD5、CD7、CD33或CD123被编辑。在一些实施方案中,CD3、CD5、CD7、CD33或CD123与选自TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和PDCD1/PD1中的一种或多种基因组合被编辑。

[1947] 在一些实施方案中,药物组合物包含经修饰的免疫效应细胞(例如,CAR-T细胞)或其群,其缺乏或具有降低水平的CD5并且表达CD5嵌合抗原受体。在一些实施方案中,药物组合物包含经修饰的免疫效应细胞(例如,CAR-T细胞)或其群,其缺乏或具有降低水平的CD7并且表达CD7嵌合抗原受体。在一些实施方案中,药物组合物包含经修饰的免疫效应细胞(例如,CAR-T细胞)或其群,其缺乏或具有降低水平的CD33并且表达CD33嵌合抗原受体。在一些实施方案中,药物组合物包含经修饰的免疫效应细胞(例如,CAR-T细胞)或其群,其缺乏或具有降低水平的CD3并且表达CD3嵌合抗原受体。在一些实施方案中,药物组合物包含经修饰的免疫效应细胞(例如,CAR-T细胞)或其群,其缺乏或具有降低水平的CD123并且表达CD123嵌合抗原受体。

[1948] 在一些实施方案中,药物组合物包含经修饰的免疫效应细胞(例如,CAR-T细胞)或其群,其缺乏或具有降低水平的CD5和选自TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和PDCD1/PD1的至少一种经编辑的基因并且表达CD5嵌合抗原受体。在一些实施方案中,药物组合物包含经修饰的免疫效应细胞(例如,CAR-T细胞)或其群,其缺乏或具有降低水平的CD7和选自TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M的至少一种经编辑的基因并且表达CD7嵌合抗原受体。在一些实施方案中,药物组合物包含经修饰的免疫效应细胞(例如,CAR-T细胞)或其群,其缺乏或具有降低水平的CD33和选自TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M的至少一种经编辑的基因并且表达CD33嵌合抗原受体。在一些实施方案中,药物组合物包含经修饰的免疫效应细胞(例如,CAR-T细胞)或其群,其缺乏或具有降低水平的CD3和选自TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M的至少一种经编辑的基因并且表达CD3嵌合抗原受体。在一些实施方案中,药物组合物包含经修饰的免疫效应细胞(例如,CAR-T细胞)或其群,其缺乏或具有降低水平的CD123和选自TRAC、LAG-3、FAS、CIITA、

TRBC1、TRBC2、CD52、B2M的至少一种经编辑的基因并且表达CD123嵌合抗原受体。

[1949] 本发明的药物组合物可以根据已知技术制备。参见,例如,《雷明顿药物科学(第二十一版)》(Remington,The Science And Practice of Pharmacy(21st ed.2005))。通常,在给药或储存之前将免疫细胞或其群与合适的载体混合,并且在一些实施方案中,药物组合物进一步包含药学上可接受的载体。合适的药学上可接受的载体通常包含惰性物质,其辅助将药物组合物给药至受试者,辅助将药物组合物加工为可递送的制剂,或辅助在给药前储存药物组合物。药学上可接受的载体可包括可以稳定化、优化或以其他方式改变调配物的形式、粘稠度、粘性、pH、药代动力学、溶解度的试剂。此类试剂包括缓冲剂、润湿剂、乳化剂、稀释剂、成胶囊剂和皮肤渗透增强剂。例如,载体可包括但不限于,盐水、缓冲盐水、葡萄糖、精氨酸、蔗糖、水、甘油、乙醇、山梨醇、葡聚糖、羧甲基纤维素钠及其组合。

[1950] 可用作药学上可接受的载剂的材料的非限制性示例包括:(1)糖类,诸如乳糖、葡萄糖和蔗糖;(2)淀粉类,诸如玉米淀粉和马铃薯淀粉;(3)纤维素及其衍生物,诸如羧甲基纤维素钠、甲基纤维素、乙基纤维素、微晶纤维素和醋酸纤维素;(4)粉末化黄芪胶;(5)麦芽;(6)明胶;(7)润滑剂,诸如硬脂酸镁、月桂基硫酸钠和滑石;(8)赋形剂,诸如可可脂和栓蜡;(9)油类,诸如花生油、棉籽油、红花油、芝麻油、橄榄油、玉米油和豆油;(10)二醇类,诸如丙二醇;(11)多元醇类,诸如甘油、山梨醇、甘露醇和聚乙二醇(PEG);(12)酯类,诸如油酸乙酯和月桂酸乙酯;(13)瓜尔胶;(14)缓冲剂,诸如氢氧化镁和氢氧化铝;(15)藻酸;(16)无热原水;(17)等渗生理盐水;(18)林格溶液;(19)乙醇;(20)pH缓冲溶液;(21)聚酯类、聚碳酸酯类和/或聚酐类;(22)填充剂,诸如多肽和氨基酸;(23)血清醇类,诸如乙醇;和(23)其他用于药物调配物中的无毒相容性物质。润湿剂、着色剂、脱模剂、包衣剂、甜味剂、风味剂、芳香剂、防腐剂和抗氧化剂也可以存在于调配物中。

[1951] 药物组合物可包含一种或多种pH缓冲化合物以将调配物的pH维持在预设水平,该预设水平反映生理pH,诸如在约5.0至约8.0的范围。用于水性液体调配物中的pH缓冲化合物可以是氨基酸或氨基酸的混合物,诸如组氨酸,或者氨基酸诸如组氨酸与甘氨酸的混合物。或者,pH缓冲化合物优选是将调配物的pH维持在预设水平(诸如约5.0至约8.0的范围)并且不螯合钙离子的试剂。此类pH缓冲化合物的示例性示例包括但不限于,咪唑和醋酸根离子。pH缓冲化合物可以任何适合将调配物的pH维持在预设水平的量存在。

[1952] 药物组合物也可含有一种或多种渗透调控剂,例如,将调配物的渗透性质(例如,紧张性、渗透度和/或渗透压)调控在接纳者个体的血流和血细胞可接受水平的试剂。渗透调控剂可以是不螯合钙离子的试剂。渗透调控剂可以是本领域技术人员已知或可获得的调控调配物的渗透性质的任何化合物。本领域技术人员可以经验性地确定给定渗透调控剂用于本发明调配物中的适用性。合适类型的渗透调控剂的示例性示例包括但不限于:盐类,诸如氯化钠和醋酸钠;糖类,诸如蔗糖、葡萄糖和甘露醇;氨基酸类,诸如甘氨酸;以及这些试剂和/或这些类型的试剂中一种或多种的混合物。渗透调控剂可以以足以调控调配物渗透性质的任何浓度存在。

[1953] 本发明的药物组合物除了包括经修饰的免疫细胞或其群以及载体外,还可包括至少一种可用于治疗疾病的另外治疗剂。例如,本文所述药物组合物的一些实施方案进一步包含化疗剂。在一些实施方案中,药物组合物进一步包含细胞因子肽或编码细胞因子肽的核酸序列。在一些实施方案中,包含经修饰的免疫细胞或其群的药物组合物可以独立于另

外治疗剂给药。

[1954] 在一些实施方案中,该至少一种另外治疗剂是一种或多种另外的经修饰的免疫细胞或其经修饰的免疫细胞的一个或多个群。在一些实施方案中,该一种或多种另外的经修饰的免疫效应细胞包含至少一个经编辑的基因以敲除或敲低该经编辑的基因的表达。在一些实施方案中,该至少一个经编辑的基因是CD3、CD5、CD7、CD33、CD123、TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M、PD1或其组合。在一些实施方案中,CD3、CD5、CD7、CD33或CD123被编辑。在一些实施方案中,CD3、CD5、CD7、CD33或CD123与选自TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和/或PD1中的一种或多种基因组合被编辑。

[1955] 在一些实施方案中,药物组合物包含其经修饰的免疫细胞的两个或更多个群。在一些实施方案中,药物组合物包含1)经修饰的免疫效应细胞(例如,CAR-T细胞)或其群,其缺乏或具有降低水平的CD5以及选自TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M的至少一种经编辑的基因,并且表达CD5嵌合抗原受体;和2)经修饰的免疫效应细胞(例如,CAR-T细胞)或其群,其缺乏或具有降低水平的CD7以及选自TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M的至少一种经编辑的基因,并且表达CD7嵌合抗原受体。

[1956] 在一些实施方案中,药物组合物包含1)经修饰的免疫效应细胞(例如,CAR-T细胞)或其群,其缺乏或具有降低水平的CD55以及选自TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M的至少一种经编辑的基因,并且表达CD55嵌合抗原受体;和2)经修饰的免疫效应细胞(例如,CAR-T细胞)或其群,其缺乏或具有降低水平的CD123以及选自TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M的至少一种经编辑的基因,并且表达CD123嵌合抗原受体。

[1957] 本发明的药物组合物可用来治疗任何对于自体或同种异体免疫细胞免疫疗法有应答的疾病或病症。例如,在一些实施方案中,药物组合物可用于治疗瘤形成。在一些实施方案中,瘤形成是实体肿瘤。在其他实施方案中,瘤形成是液态肿瘤。在一些实施方案中,瘤形成是血液学癌症。在一些实施方案中,血液学癌症是B细胞癌症,且在一些实施方案中,B细胞癌症是多发性骨髓瘤。在一些实施方案中,B细胞癌症是复发性/难治性多发性骨髓瘤。在一些实施方案中,血液学癌症是白血病、骨髓瘤和/或淋巴瘤。在一些情况下,白血病是急性白血病。急性白血病包括,例如,急性骨髓性白血病(AML)。急性白血病也包括,例如,急性淋巴性白血病或急性淋巴细胞性白血病(ALL);ALL包括B谱系ALL、T谱系ALL和T细胞急性淋巴细胞性白血病(T-ALL)。

[1958] 瘤形成的非限制性示例包括T细胞急性成淋巴细胞性白血病(T-ALL)、蕈样肉芽肿(MF)、Sézary综合征(SS)、外周T/NK细胞淋巴瘤、间变性大细胞淋巴瘤ALK^+、原发性皮肤T细胞淋巴瘤、T细胞大颗粒淋巴细胞型白血病、血管免疫母细胞性T/NK细胞淋巴瘤、肝脾T细胞淋巴瘤、原发性皮肤$CD30^+$淋巴增生性疾病、结外NK/T细胞淋巴瘤、成人T细胞白血病/淋巴瘤、T细胞幼淋巴细胞白血病、皮下脂膜炎样T细胞淋巴瘤、原发性皮肤γδT细胞淋巴瘤、侵袭性NK细胞白血病和肠病相关T细胞淋巴瘤。在一些实施方案中,瘤形成是T细胞急性成淋巴细胞性白血病(T-ALL)。在一些实施方案中,瘤形成是T细胞急性骨髓性白血病(AML)。

[1959] 在一些实施方案中,向患有或易患瘤形成(例如,白血病)的受试者给药有效量以下一项或多项:1)经修饰的免疫效应细胞(例如,CAR-T细胞)或其群,其缺乏或具有降低水平的CD5以及选自TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M的至少一种经编辑的基因,并且表达CD5嵌合抗原受体;2)经修饰的免疫效应细胞(例如,CAR-T细胞)或其群,其缺

乏或具有降低水平的CD7以及选自TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M的至少一种经编辑的基因,并且表达CD7嵌合抗原受体;3)经修饰的免疫效应细胞(例如,CAR-T细胞)或其群,其缺乏或具有降低水平的CD33以及选自TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M的至少一种经编辑的基因,并且表达CD33嵌合抗原受体;和4)经修饰的免疫效应细胞(例如,CAR-T细胞)或其群,其缺乏或具有降低水平的CD123以及选自TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M的至少一种经编辑的基因,并且表达CD123嵌合抗原受体。

[1960] 一种关于本发明的基因上经修饰的免疫细胞的治疗性用途的考量是实现最优或令人满意的效果所必需的细胞数量。待给药的细胞数量针对待治疗的受试者可变。在一种实施方案中,向人类受试者给药介于10^4到10^{10}之间、介于10^5到10^9之间、或介于10^6到10^8之间的本发明的基因上经修饰的免疫应答细胞。在一些实施方案中,向人类受试者给药至少约1x 10^8、2x 10^8、3x 10^8、4x 10^8和5x 10^8的本发明的基因上经修饰的免疫细胞。可以基于多种因素,包括其尺寸、年龄、性别、体重和病症,针对每个个体受试者确定精确的有效剂量。本领域技术人员可以从本发明和本领域知识容易地确定剂量。

[1961] 本领域技术人员可以容易地确定将在本发明方法中给药的组合物中的细胞数量和任选使用的添加剂、媒介物和/或载体的量。典型地,添加剂(除了活性免疫细胞之外)以0.001%至50%(重量)溶液的量存在于磷酸盐缓冲盐水中,并且活性成分以微克到毫克的数量级存在,诸如约0.0001至约5wt%,优选约0.0001至约1wt%,在再更优选约0.0001至约0.05wt%或约0.001至约20wt%,优选约0.01至约10wt%,且在更优选约0.05至约5wt%。当然,对于待给药至动物或人类的任何组合物,以及对于任何特定给药方法,优选因此确定:毒性,诸如通过在合适动物模型(例如,啮齿动物诸如小时)中确定致死剂量(LD)和LD50;以及引起合适应答的组合物的剂量、其中组分的浓度和给药组合物的时机。此类确定不需要根据本领域技术人员的知识、本公开和本文引用的文件进行不必要的实验。而且,不需要进行不必要的实验即可确定依次给药的时间。

[1962] 在一种实施方案中,本文所述方法和组合物可用于生成经工程改造的T细胞,其表达CAR并且可具有一个或多个经碱基编辑的修饰,使得经工程改造的T细胞可表现针对靶标的免疫应答。可以将CAR特异性地引导向抗原靶标,该抗原可以由宿主体内的细胞呈递。在一些实施方案中,免疫应答涵盖细胞毒性。在一些实施方案中,经工程改造的T细胞具有增强的针对其靶标的细胞毒性应答。在一些实施方案中,与未经工程改造的T细胞相比,经工程改造的T细胞诱导增强的细胞毒性应答。在一些实施方案中,与未经工程改造的细胞相比,经工程改造的T细胞表现处增强至少1.1倍、1.5倍、2倍、3倍、4倍、5倍、6倍、7倍、8倍、9倍、10倍、11倍、12倍、13倍、14倍、15倍、16倍、17倍、18倍、19倍、20倍或更多的细胞毒性应答。在一些实施方案中,经工程改造的T细胞可以比未经工程改造的细胞多杀死至少10%、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200%、至少500%或至少1000%的靶标细胞。在一些实施方案中,T细胞可以诱导较高的记忆应答。在一些实施方案中,T细胞可能诱导比未经工程改造的细胞更低水平的炎症细胞因子,换言之,经工程改造的细胞不造成细胞因子风暴应答。

[1963] 在一些实施方案中,将经工程改造的T细胞给药至同种异体宿主,其中经工程改造的T细胞不被该宿主排异。在一些实施方案中,同种异体T细胞诱导可忽略的或最低限度的宿主排异。在一些实施方案中,经工程改造T细胞具有同类相杀抗性。

[1964] 在一些实施方案中,将药物组合物配置为用于递送至受试者。给药本文所述的药物组合物的合适途径包括但不限于:外用、皮下、透皮、皮内、病灶内、关节内、腹腔内、膀胱内、跨粘膜、牙龈、牙内、耳蜗内、经鼓室、器官内、硬膜外、鞘内、肌肉内、静脉内、血管内、骨内、眼周、肿瘤内、脑内和脑室内给药。

[1965] 在一些实施方案中,将本文所述的药物组合物局部给药至患病位点(例如,肿瘤位点)。在一些实施方案中,通过注射、通过导管、通过栓剂或通过植入物将本文所述的药物组合物给药至受试者,该植入物是多孔的、无孔的或胶冻样材料,包括膜如硅橡胶膜或纤维。

[1966] 在其他实施方案中,本文所述的药物组合物在控释系统中递送。在一种实施方案中,可以使用泵(参见,例如,Langer,1990,Science249:1527-1533;Sefton,1989,CRC Crit.Ref.Biomed.Eng.14:201;Buchwald等人,1980,Surgery 88:507;Saudek et al,1989,N.Engl.J.Med.321:574)。在另一实施方案中,可使用聚合物材料。(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。也参见Levy等人,1985,Science 228:190;During等人,1989,Ann.Neurol.25:351;Howard等人,1989,J.Neurosurg.71:105。)其他控释系统例如在上文的Langer中讨论。

[1967] 在一些实施方案中,根据常规过程将药物组合物配置为适用于静脉内或皮下给药至受试者例如人类的组合物。在一些实施方案中,用于通过注射给药的药物组合物是用作增溶剂和局部麻醉剂的无菌等渗溶液,诸如利多卡因,用于缓解注射部位的疼痛。通常,成分或单独提供或混合在一起以单位剂型提供,例如作为干燥的冻干粉或无水浓缩物提供在指示活性剂数量的气密密封的容器诸如安瓿或小袋内。若药物组合物待通过输液给药,它可以用含有无菌药用级水或盐水的输液瓶分配。若药物组合物通过注射给药,可提供无菌注射用水或盐水的安瓿,使得成分可在给药之前混合。

[1968] 用于全身给药的药物组合物可以是液态,例如无菌盐水、乳酸化林格溶液或汉克溶液。此外,药物组合物可以是固体形式,并且在重新溶解或悬浮后立即使用。也设想了冻干的形式。药物组合物可以包含在液体颗粒或媒介物诸如脂质体或微晶中,且也适用于肠胃外给药。颗粒可以是任何合适的结构,诸如单层或多层,只要组合物包含在其中即可。化合物可以入陷在“稳定化质粒-脂质颗粒(SPLP)”中,该颗粒含有融合性脂质二油酰基磷脂酰乙醇胺(DOPE)、低水平(5-10mol%)的阳离子脂质,并且通过聚乙二醇(PEG)涂层稳定化(Zhang Y.P.等人,Gene Ther.1999,6:1438-47)。带正电的脂质诸如N-[1-(2,3-二油酰氧基)丙基]-N,N,N-三甲基-甲基硫酸铵或“DOTAP”对于此类颗粒和囊泡是特佳的。此类脂质颗粒的制备是周知的。参见,例如,美国专利号4,880,635、4,906,477、4,911,928、4,917,951、4,920,016和4,921,757,其各自通过引用并入本文。

[1969] 本文所述的药物组合物可以例如作为单位剂量给药或包装。当术语“单位剂量”关于本公开的药物组合物使用时,是指适合作何用于受试者的单一剂量的物理上离散的单元,每个单元含有经计算以产生所希望治疗效果的预设数量的活性材料联合所需的稀释剂;即,载体或媒介物。

[1970] 再者,药物组合物可提供为药物试剂盒,其包含(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,用于重构或稀释冻干的本发明化合物的无菌稀释剂)的第二容器。任选地与此类容器相关联的可以是政府机构预设形式的关于制造、使用或销售药品或生物产品的注意事项,该注意事项反映了该机构对制造、使用或销售用于人体给药的批准。

[1971] 另一方面,包括了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,制品包含容器和标记。合适的容器包括,例如,瓶、小瓶、注射器和试管。容器可以由多种材料诸如玻璃或塑料形成。在一些实施方案中,容器容纳本文所述还可有效治疗疾病的组合物并且可具有无菌出入口。例如,容器可以是具有可被皮下注射针刺穿的塞子的静脉内溶液袋或小瓶。组合物中的活性剂是本发明的化合物。在一些实施方案中,容器上或与容器关联的标记指示组合物用于治疗所选择的疾病。制品可进一步包含第二容器,该第二容器包含药学上可接受的缓冲液,诸如磷酸盐缓冲盐水、林格溶液或葡萄糖溶液。它可以进一步包括来自商用和使用者观点的其他所希望的材料,包括其他缓冲剂、稀释剂、过滤器、针、注射器和具有使用说明的包装插页。

[1972] 在一些实施方案中,提供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。在一些实施方案中,药物组合物包含本文所提供的任何融合蛋白。在一些实施方案中,药物组合物包含本文所提供的任何复合物。在一些实施方案中,药物组合物包含一种核糖核蛋白复合物,该复合物包含RNA引导的核酸酶(例如,Cas9),其与gRNA和阳离子脂质形成复合物。在一些实施方案中,药物组合物包含gRNA、核酸可编程DNA结合蛋白、阳离子脂质和药学上可接收的赋形剂。药物组合物可任选地包含一种或多种另外的治疗活性剂。

[1973] 在一些实施方案中,将本文所提供的组合物给药至受试者例如人类受试者,以便在受试者内实现靶向基因组修饰。在一些实施方案中,细胞获自受试者并且与本文所提供的任何药物组合物接触。在一些实施方案中,从受试者去除并在间接体内与药物组合物接触的细胞被重新引入该受试者体内,任选地在已经在细胞中实现或检测到所希望的基因组修饰之后。递送包含核酸酶的药物组合物的方法是已知的,并且在例如美国专利号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中描述,其全部的公开内容通过引用以其整体并入本文。尽管本文所提供的药物组合物的描述主要针对适于给药至人类的药物组合物,但本领域技术人员将理解,此类组合物通常适于给药至动物或各种生物体,例如用于兽医用途。

[1974] 对适用于给药至人类的药物组合物进行修饰以便使该组合物适用于给药至各种动物很好理解,并且兽医药理学领域一般技术人员可以使用极常规(如果有)的实验来设计并/或执行修饰。药物组合物所给药至的受试者预期包括但不限于,人类和/或其他灵长动物;哺乳动物、驯养动物、宠物和商用相关哺乳动物,诸如牛、猪、马、绵羊、狗、小鼠和/或大鼠;和/或鸟类,包括商用相关鸟类诸如鸡、鸭、鹅和/或火鸡。

[1975] 本文所述药物组合物的调配物可以通过任何已知方法或此后制药领域开发的方法制备。通常,此类制备方法包括将活性成分带至与赋形剂和/或一种或多种其他辅助成分联合的步骤,且随后,若必要及/或希望,将产品成形且/或包装o为所希望的单或多剂量单位。药物调配物可另外包含药学上可接受的赋形剂,如本文所用,该赋形剂包括任何和全部溶剂、分散介质、稀释剂或其他液体媒介物、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,如适合所希望的剂型。《雷明顿药物科学与实践(第二十一版)》(Remington's The Science and Practice of Pharmacy,21st Edition,A.R.Gennaro(Lippincott,Williams&Wilkins,Baltimore,MD,2006);通过引用以其整体并入本文)公开了用于配置药物组合物的各种赋形剂和用于其制备的已知技术。关于用于生产包含核酸酶的药物组合物的另外的合适方法、试剂、赋形剂和溶剂,也参见PCT申请PCT/US2010/055131(公开号WO2011/053982A8,2010年11月2日提交),通过引用以其整体并入本文。

[1976] 除非任何常规赋形剂介质与物质或其衍生物不相容,诸如通过产生任何不良生物效应或以其他有害方式与药物组合物的任何其他组分相互作用,否则其使用预期在本发明的范围内。

[1977] 上文所述的组合物以有效量给药。有效量将取决于给药模式、待治疗的特定病症和所希望的结果。它也可能取决于病症的阶段、受试者的年龄和身体状况、并行疗法的属性(如果有)和医疗从业人员周知的类似因素。对于治疗性应用,它是足以实现医学上所希望的结果的量。

[1978] 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、疾患和/或病症中的任一者。

[1979] 治疗方法

[1980] 本发明的一些方面提供治疗有此需要的受试者的方法,该方法包括向有此需要的受试者给药有效治疗量的本文所述药物组合物。更具体地,治疗方法包括向有此需要的受试者给药一种或多种药物组合物,该药物组合物包含表达嵌合抗原受体(CAR)且具有至少一个经编辑的基因的经修饰的免疫细胞群,其中该至少一个经编辑的基因提供同类相杀抗性,增强经修饰的免疫细胞的功能或降低其免疫阻遏或抑制,并且其中该至少一种经编辑的基因的表达被敲除或敲低。在一些实施方案中,该治疗方法是自体免疫细胞疗法。在其他实施方案中,该治疗方法是同种异体免疫细胞疗法。

[1981] 在某些实施方案中,通过在基因上修饰免疫细胞以表达本文预期的嵌合抗原受体,免疫细胞的特异性被重新引导至子啊受试者的患病细胞或被改变的细胞表面上表达的标志物。在一些实施方案中,治疗方法包括向受试者给药本文所述的免疫细胞,其中免疫细胞已经在基因上经修饰以将其特异性重新引导至在肿瘤细胞上表达的标志物。在一些实施方案中,瘤形成是B细胞癌症;例如,B细胞癌症诸如淋巴瘤、白血病或骨髓瘤,例如,多发性骨髓瘤。在一些实施方案中,瘤形成是实体肿瘤。在其他实施方案中,瘤形成是液态肿瘤。在一些实施方案中,瘤形成是血液学癌症。在一些实施方案中,血液学癌症是白血病、骨髓瘤和/或淋巴瘤。在一些情况下,白血病包括白血病前期。在一些情况下,白血病是急性白血病。急性白血病包括,例如,急性骨髓性白血病(AML)。急性白血病也包括,例如,急性淋巴性白血病或急性淋巴细胞性白血病(ALL);ALL包括B谱系ALL、T谱系ALL和T细胞急性淋巴细胞性白血病(T-ALL)。

[1982] 瘤形成的非限制性示例包括T细胞急性成淋巴细胞性白血病(T-ALL)、蕈样肉芽肿(MF)、Sézary综合征(SS)、外周T/NK细胞淋巴瘤、间变性大细胞淋巴瘤ALK^+、原发性皮肤T细胞淋巴瘤、T细胞大颗粒淋巴细胞型白血病、血管免疫母细胞性T/NK细胞淋巴瘤、肝脾T细胞淋巴瘤、原发性皮肤$CD30^+$淋巴增生性疾病、结外NK/T细胞淋巴瘤、成人T细胞白血病/淋巴

瘤、T细胞幼淋巴细胞白血病、皮下脂膜炎样T细胞淋巴瘤、原发性皮肤γδT细胞淋巴瘤、侵袭性NK细胞白血病和肠病相关T细胞淋巴瘤。在一些实施方案中,瘤形成是T细胞急性成淋巴细胞性白血病(T-ALL)。在一些实施方案中,瘤形成是T细胞急性骨髓性白血病(AML)。因此,本公开的一些方面提供治疗受试者的瘤形成的方法。

[1983] 在一些实施方案中,治疗方法包括向患有瘤形成(例如,白血病)或具有发展出瘤形成倾向的受试者给药有效量的经修饰的免疫细胞(例如,CAR-T细胞),该免疫细胞缺乏或具有降低水平的功能性T细胞受体α恒定区(TRAC)、分化簇33(CD33)、分化簇3(CD3)、分化簇5(CD5)、分化簇7(CD7)、分化簇52(CD52)、分化簇123(CD123)、程序性细胞死亡1(PD-1)、β-2微球蛋白(B2M)、Fas细胞表面死亡受体(FAS)、淋巴细胞激活基因3(LAG-3)、II类主要组织相容性复合物、反式激活因子(CIITA)、T细胞受体β恒定区1(TRBC1)、T细胞受体β恒定区2(TRBC2)、或其组合。在一些方案中,向患有瘤形成(例如,T或NK细胞恶性病变)或具有发展出瘤形成潜力的受试者给药有效量的经修饰的免疫效应细胞(例如,CAR-T细胞),该细胞缺乏或具有降低水平的CD123并且表达CD123嵌合抗原受体。在一些方案中,向患有瘤形成(例如,T或NK细胞恶性病变)或具有发展出瘤形成潜力的受试者给药有效量的经修饰的免疫效应细胞(例如,CAR-T细胞),该细胞缺乏或具有降低水平的CD5并且表达CD5嵌合抗原受体。在一些方案中,向患有瘤形成(例如,T或NK细胞恶性病变)或具有发展出瘤形成潜力的受试者给药有效量的经修饰的免疫效应细胞(例如,CAR-T细胞),该细胞缺乏或具有降低水平的CD7并且表达CD7嵌合抗原受体。在一些方案中,向患有瘤形成(例如,T或NK细胞恶性病变)或具有发展出瘤形成潜力的受试者给药有效量的经修饰的免疫效应细胞(例如,CAR-T细胞),该细胞缺乏或具有降低水平的CD33并且表达CD33嵌合抗原受体。在一些方案中,向患有瘤形成(例如,T或NK细胞恶性病变)或具有发展出瘤形成潜力的受试者给药有效量的经修饰的免疫效应细胞(例如,CAR-T细胞),该细胞缺乏或具有降低水平的CD3并且表达CD3嵌合抗原受体。

[1984] 在一些方案中,向患有瘤形成(例如,T或NK细胞恶性病变)或具有发展出瘤形成潜力的受试者给药有效量的经修饰的免疫效应细胞(例如,CAR-T细胞),该细胞缺乏或具有降低水平的功能性TRAC。在一些方案中,向患有瘤形成(例如,T或NK细胞恶性病变)或具有发展出瘤形成潜力的受试者给药有效量的经修饰的免疫效应细胞(例如,CAR-T细胞),该细胞缺乏或具有降低水平的功能性B2M。在一些方案中,向患有瘤形成(例如,T或NK细胞恶性病变)或具有发展出瘤形成潜力的受试者给药有效量的经修饰的免疫效应细胞(例如,CAR-T细胞),该细胞缺乏或具有降低水平的功能性FAS。在一些方案中,向患有瘤形成(例如,T或NK细胞恶性病变)或具有发展出瘤形成潜力的受试者给药有效量的经修饰的免疫效应细胞(例如,CAR-T细胞),该细胞缺乏或具有降低水平的功能性LAG-3。在一些方案中,向患有瘤形成(例如,T或NK细胞恶性病变)或具有发展出瘤形成潜力的受试者给药有效量的经修饰的免疫效应细胞(例如,CAR-T细胞),该细胞缺乏或具有降低水平的功能性TRBC1和/或TRBC2。在一些方案中,向患有瘤形成(例如,T或NK细胞恶性病变)或具有发展出瘤形成潜力的受试者给药有效量的经修饰的免疫效应细胞(例如,CAR-T细胞),该细胞缺乏或具有降低水平的功能性CIITA。在一些方案中,向患有瘤形成(例如,T或NK细胞恶性病变)或具有发展出瘤形成潜力的受试者给药有效量的经修饰的免疫效应细胞(例如,CAR-T细胞),该细胞缺乏或具有降低水平的CD52。在一些方案中,向患有瘤形成(例如,T或NK细胞恶性病变)或具

有发展出瘤形成潜力的受试者给药有效量的经修饰的免疫效应细胞(例如,CAR-T细胞),该细胞缺乏或具有降低水平的PD-1。

[1985] 在一些实施方案中,向患有瘤形成(例如,T或NK细胞恶性病变)或具有发展出瘤形成潜力的受试者给药有效量的经修饰的免疫效应细胞(例如,CAR-T细胞)或其群,该细胞缺乏或具有降低水平的CD123并且缺乏或具有降低水平的TRAC、LAG-3、FAS、CD52、B2M、CIITA、TRBC1、TRBC2或/或PD-1或其组合,并且表达CD123嵌合抗原受体。在一些实施方案中,向患有瘤形成(例如,T或NK细胞恶性病变)或具有发展出瘤形成潜力的受试者给药有效量的经修饰的免疫效应细胞(例如,CAR-T细胞),该细胞缺乏或具有降低水平的CD5并且缺乏或具有降低水平的TRAC、LAG-3、FAS、CD52、B2M、CIITA、TRBC1、TRBC2或/或PD-1或其组合,并且表达CD5嵌合抗原受体。在一些实施方案中,向患有瘤形成(例如,T或NK细胞恶性病变)或具有发展出瘤形成潜力的受试者给药有效量的经修饰的免疫效应细胞(例如,CAR-T细胞),该细胞缺乏或具有降低水平的CD7并且缺乏或具有降低水平的TRAC、LAG-3、FAS、CD52、B2M、CIITA、TRBC1、TRBC2或/或PD-1或其组合,并且表达CD7嵌合抗原受体。在一些实施方案中,向患有瘤形成(例如,T或NK细胞恶性病变)或具有发展出瘤形成潜力的受试者给药有效量的经修饰的免疫效应细胞(例如,CAR-T细胞),该细胞缺乏或具有降低水平的CD33并且缺乏或具有降低水平的TRAC、LAG-3、FAS、CD52、B2M、CIITA、TRBC1、TRBC2或/或PD-1或其组合,并且表达CD33嵌合抗原受体。在一些实施方案中,向患有瘤形成(例如,T或NK细胞恶性病变)或具有发展出瘤形成潜力的受试者给药有效量的经修饰的免疫效应细胞(例如,CAR-T细胞),该细胞缺乏或具有降低水平的CD3并且缺乏或具有降低水平的TRAC、LAG-3、FAS、CD52、B2M、CIITA、TRBC1、TRBC2或/或PD-1或其组合,并且表达CD3嵌合抗原受体。

[1986] 在一些实施方案中,治疗受试者的瘤形成的方法包括向受试者给药本文所述的免疫细胞和一种或多种另外治疗剂。例如,本发明的免疫细胞可以与一种或多种另外的经修饰的免疫细胞或其群共同给药。在一些实施方案中,向患有瘤形成(例如,T或NK细胞恶性病变)或具有发展出瘤形成潜力的受试者共同给药:有效量的经修饰的免疫效应细胞(例如,CAR-T细胞),其缺乏或具有降低水平的CD123并且表达CD123嵌合抗原受体;以及有效量的经修饰的免疫效应细胞(例如,CAR-T细胞),其缺乏或具有降低水平的CD33并且表达CD33嵌合抗原受体。在一些实施方案中,向患有瘤形成(例如,T或NK细胞恶性病变)或具有发展出瘤形成潜力的受试者共同给药:有效量的经修饰的免疫效应细胞(例如,CAR-T细胞),其缺乏或具有降低水平的CD5并且表达CD5嵌合抗原受体;以及有效量的经修饰的免疫效应细胞(例如,CAR-T细胞),其缺乏或具有降低水平的CD7并且表达CD7嵌合抗原受体。在一些实施方案中,任何免疫细胞进一步包含在CD3、CD5、CD7、CD33、CD123、TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和/或PD1或其组合中的一者或多者中的突变。在一些实施方案中,向患有瘤形成(例如,T或NK细胞恶性病变)或具有发展出瘤形成潜力的受试者共同给药:有效量的经修饰的免疫效应细胞(例如,CAR-T细胞),其缺乏或具有降低水平的CD3并且表达CD3嵌合抗原受体;以及有效量的经修饰的免疫效应细胞(例如,CAR-T细胞),其缺乏或具有降低水平的CD7并且表达CD7嵌合抗原受体。在一些实施方案中,向患有瘤形成(例如,T或NK细胞恶性病变)或具有发展出瘤形成潜力的受试者共同给药:有效量的经修饰的免疫效应细胞(例如,CAR-T细胞),其缺乏或具有降低水平的CD3并且表达CD3嵌合抗原受体;以及有效量的经修饰的免疫效应细胞(例如,CAR-T细胞),其缺乏或具有降低水平的CD123并且表

达CD123嵌合抗原受体。

[1987] 在一些实施方案中,本发明的免疫细胞可以与细胞因子共同给药。在一些实施方案中,细胞因子是IL-2、IFN-α、IFN-γ或其组合。在一些实施方案中,免疫细胞与化疗剂共同给药。化疗剂可以是环磷酰胺、阿霉素、长春新碱、强的松或利妥昔单抗或其组合。其他化疗剂包括奥妥珠单抗、苯达莫司汀、苯丁酸氮芥、环磷酰胺、依鲁替尼、氨甲蝶呤、阿糖胞苷、地塞米松、顺铂、硼替佐米、氟达拉滨、艾代拉里斯(idelalisib)、阿卡替尼(acalabrutinib)、来那度胺、维奈托克、换新酰胺、异环磷酰胺、依托泊苷、喷司他丁、美法仑、卡非佐米、伊沙佐米、帕比司他、达雷木单抗、埃罗妥珠单抗、沙利度胺、来那度胺或泊马度胺或其组合。"共同给药"是指在疗程期间给药两种或更多种治疗剂或药物组合物。该共同给药可以是同时给药或依次给药。较晚给药的治疗剂或药物组合物的依次给药可能在给药第一种药物组合物或治疗剂后在疗程期间的任何时间发生。

[1988] 在本发明的一些实施方案中,所给药的免疫细胞在体内增殖并且可在受试者体内长期持续存在。在一些实施方案中,本发明的免疫细胞可以成熟为记忆免疫细胞并且保留在受试者的循环中,从而生成能对表达被嵌合抗原受体识别的标志物的患病或被改变细胞的重现作出主动应答的细胞群。

[1989] 本文提供使用经修饰的免疫细胞(例如,获自健康受试者的免疫细胞,或者T细胞诸如细胞毒T细胞、调节T细胞或T辅助细胞)治疗瘤形成或液态癌症的方法,该经修饰的免疫细胞通过以下方法产生:在免疫细胞中表达核碱基编辑器多肽或将核碱基编辑器多肽引入免疫细胞中,并且使该细胞与能够靶向编码至少一个多肽的核酸分子的两个或更多个向导RNA接触,其中第一多肽是Cbl原癌基因B(CBLB)且第二多肽选自由T细胞受体α恒定区(TRAC)、β-2微球蛋白(B2M)和程序性细胞死亡蛋白1(PD1)多肽所组成的群组,其中核碱基编辑器多肽包含核酸可编程DNA结合蛋白(napDNAbp)和胞苷脱氨酶。在该方法的一个实例中,三种向导RNA被表达在该细胞中或与该细胞接触,各自靶向B2M、TRAC、PDCD1或CBLB多核苷酸。在该方法的另一实例中,四种向导RNA被表达在该细胞中或与该细胞接触,各自靶向B2M、TRAC、PDCD1和CBLB多核苷酸中的一个。在一个实例中,两种向导RNA被表达在该细胞中或与该细胞接触,各自靶向B2M或TRAC多核苷酸。在另一实例中,三种向导RNA被表达在该细胞中或与该细胞接触,各自靶向B2M、TRAC、PDCD1或CBLB多核苷酸。在另一实例中,四种向导RNA被表达在该细胞中或与该细胞接触,各自靶向B2M、TRAC、PDCD1和CBLB多核苷酸中的一个。在另一实例中,三种向导RNA被表达在该细胞中或与该细胞接触,各自靶向B2M、TRAC和PDCD1多核苷酸。在另一实例中,向导RNA靶向TRAC外显子4剪接受体位点、B2M外显子1剪接供体位点、或PDCD1外显子1剪接供体位点。在另一实施方案中,生成经修饰的免疫细胞的方法进一步包括在该细胞中表面向导RNA,该向导RNA靶向细胞毒性T淋巴细胞相关蛋白4(CTLA4)多核苷酸。

[1990] 本文提供使用经修饰的免疫细胞(例如,获自健康受试者的免疫细胞,或者T细胞诸如细胞毒T细胞、调节T细胞或T辅助细胞)治疗瘤形成或液态癌症的方法,该经修饰的免疫细胞通过以下方法产生:在免疫细胞中表达核碱基编辑器多肽或将核碱基编辑器多肽引入免疫细胞中,并且使该细胞与能够靶向编码至少一个多肽的核酸分子的两个或更多个向导RNA接触,该多肽选自由Cbl原癌基因B(CBLB)、T细胞受体α恒定区(TRAC)、β-2微球蛋白(B2M)和程序性细胞死亡蛋白1(PD1)多肽所组成的群组,其中核碱基编辑器多肽包含核酸

可编程DNA结合蛋白(napDNAbp)和腺苷脱氨酶。在一个实例中,两种向导RNA被表达在该细胞中或与该细胞接触,各自靶向B2M或TRAC多核苷酸。在另一实例中,三种向导RNA被表达在该细胞中或与该细胞接触,各自靶向B2M、TRAC、PDCD1或CBLB多核苷酸。在另一实例中,四种向导RNA被表达在该细胞中或与该细胞接触,各自靶向B2M、TRAC、PDCD1和CBLB多核苷酸中的一个。在另一实例中,三种向导RNA被表达在该细胞中或与该细胞接触,各自靶向B2M、TRAC和PDCD1多核苷酸。在另一实例中,向导RNA靶向TRAC外显子4剪接受体位点、B2M外显子1剪接供体位点、或PDCD1外显子1剪接供体位点。在另一实施方案中,生成经修饰的免疫细胞的方法进一步包括在该细胞中表面向导RNA,该向导RNA靶向细胞毒性T淋巴细胞相关蛋白4(CTLA4)多核苷酸。

[1991] 本文提供使用经修饰的免疫细胞(例如,获自健康受试者的免疫细胞,或者T细胞诸如细胞毒性T细胞、调节T细胞或T辅助细胞)治疗瘤形成或液态癌症的方法,该经修饰的免疫细胞通过以下方法生产:在免疫细胞中表达核碱基编辑器多肽或将核碱基编辑器多肽引入免疫细胞中,并且使该细胞与能够靶向编码至少一种多肽的核酸分子的两个或更多个向导RNA接触;该至少一个多肽选自由以下所组成的群组:V-Set免疫调节受体(VISTA)、T细胞免疫球蛋白粘蛋白3(Tim-3)、具有Ig和ITIM结构域的T细胞免疫受体(TIGIT)、转化生长因子β受体II(TGFbRII)、含有调节因子X相关锚蛋白的蛋白质(RFXANK)、含PVR相关免疫球蛋白结构域(PVRIG)、淋巴细胞激活基因3(Lag3)、细胞毒性T淋巴细胞相关蛋白4(CTLA-4)、几丁质酶3样1(Chi3l1)、分化簇96(CD96)、B和T淋巴细胞相关(BTLA)、Tet甲基胞苷双加氧酶2(TET2)、Sprouty RTK信号传导拮抗剂1(Spry1)、Sprouty RTK信号传导拮抗剂2(Spry2)、II类主要组织相容性复合物反式激活因子(CIITA)、分化簇3(CD3)、分化簇7(CD7)、分化簇33(CD33)、分化簇52(CD52)、分化簇123(CD123)、T细胞受体β恒定区1(TRBC1)、T细胞受体β恒定区2(TRBC2)、细胞因子诱导型含SH2蛋白质(CISH)、乙酰辅酶A乙酰转移酶1(ACAT1)、细胞色素P450家族11亚家族A成员1(Cyp11a1)、GATA结合蛋白3(GATA3)、核受体亚家族4A组成员1(NR4A1)、核受体亚家族4A组成员2(NR4A2)、核受体亚家族4A组成员3(NR4A3)、甲基化控制的J蛋白(MCJ)、Fas细胞表面死亡受体(FAS)、和选择蛋白P配体体/P-选择蛋白糖蛋白配体-1(SELPG/PSGL1)多肽;其中核碱基编辑器多肽包含核酸可编程DNA结合蛋白(napDNAbp)和腺苷或胞苷脱氨酶。一种实施方案中,生产细胞的方法进一步包括使细胞与能够靶向编码核酸分子的两个或更多个向导RNA接触,该核酸分子选自由T细胞受体α恒定区(TRAC)、β-2微球蛋白(B2M)和程序性细胞死亡蛋白1(PD1)多肽所组成的群组。在另一实施方案中,生成经修饰的免疫细胞的方法进一步包括在该细胞中表面向导RNA,该向导RNA靶向细胞毒性T淋巴细胞相关蛋白4(CTLA4)多核苷酸。

[1992] 本文提供一种如本文所述使用经修饰的免疫细胞(例如,获自健康受试者的免疫细胞,或者T细胞诸如细胞毒性T细胞、调节T细胞或T辅助细胞)治疗瘤形成或液态癌症的方法,该经修饰的免疫细胞具有抗瘤形成活性,通过包含在从患有瘤形成的受试者分离的免疫细胞内表达一个或多个向导RNA和和碱基编辑器多肽的方法产生,该向导RNA靶向编码选自由CBLB、PD1、CTLA4、TGFBR2、TIGIT或最小组织相容性多肽所组成的群组的多肽的多核苷酸,且该核碱基编辑器多肽包含核酸可编程DNA结合蛋白(napDNAbp)和胞苷脱氨酶或腺苷脱氨酶。在一种情况下,向导RNA各自靶向CBLB和PDCD1多核苷酸。在一种情况下,向导RNA各自靶向PDCD1和CTLA4多核苷酸。在一种情况下,向导RNA各自靶向CBLB、PDCD1和CTLA4多核

苷酸。在一种情况下,向导RNA各自靶向CBLB、PD1和CTLA4多核苷酸。

[1993] 在上述任何产生经修饰的免疫细胞的方法中,在一种情况下,向导RNA靶向靶标多核苷酸中的剪接受体位点或剪接供体位点。在上述任何产生经修饰的免疫细胞的方法中,在一种情况下,核碱基编辑器多肽在靶标多核苷酸中生成终止密码子。在上述任何产生经修饰的免疫细胞的方法中,在一种情况下,核碱基编辑器多肽在PDCD1外显子2中生成终止密码子。在上述任何产生经修饰的免疫细胞的方法中,在一种情况下,核碱基编辑器多肽进一步包含表达一种或多种核定位信号(NLS)诸如二分体NLS、N端NLS、C端NLS、或一种或多种尿嘧啶糖基化酶抑制剂。

[1994] 在上述任何产生经修饰的免疫细胞的方法中,在一种情况下,napDNAbp选自由Cas9、CasX、CasY、Cpf1、Cas12b/C2c1、Cas12c/C2c3,、Cas12j/CasΦ和BvCas12b或其活性片段所组成的群组。在上述任何产生经修饰的免疫细胞的方法中,在一种情况下,napDNAbp结构域包含核酸酶死亡的Cas9(dCas9)、Cas9切口酶(nCas9)或核酸酶活性Cas9。在上述任何产生经修饰的免疫细胞的方法中,在一种情况下,胞苷脱氨酶是海七鳃鳗胞嘧啶脱氨酶1(pmCDA1)或激活诱导型胞苷脱氨酶(AICDA),并且腺苷脱氨酶是TadA。

[1995] 在上述任何产生经修饰的免疫细胞的方法中,在一种情况下,该方法进一步包括在免疫细胞中表达一个或多个靶向TIGIT多核苷酸、TGFBR2多核苷酸、ZAP70多核苷酸、NFATc1多核苷酸和TET2多核苷酸的向导RNA。

[1996] 应理解,在一些情况下,待给药至受试者的经修饰的免疫细胞不包含可检测的易位。

[1997] 本文提供一种使用具有降低的免疫原性和增加的抗瘤形成活性的经修饰的免疫细胞(例如,获自健康受试者的免疫细胞,或者T细胞诸如细胞毒性T细胞、调节T细胞或T辅助细胞)治疗瘤形成或液态癌症的方法,该经修饰的免疫细胞包含在TRAC、B2M、CBLB或PD1多核苷酸或其组合中的突变。在一种情况下,经修饰的免疫细胞包含在B2M和TRAC多核苷酸中的突变。在另一种情况下,经修饰的免疫细胞包含在B2M、TRAC和PDCD1多核苷酸中的突变。在另一种情况下,经修饰的免疫细胞包含在B2M、TRAC和CBLB多核苷酸中的突变。在另一种情况下,经修饰的免疫细胞包含在B2M、TRAC、CBLB和PD1多核苷酸中的突变。在另一种情况下,经修饰的免疫细胞包含在一种或多种编码TIGIT、TGFBR2、ZAP70、NFATc1或TET2的多核苷酸中的突变。在另一种情况下,经修饰的免疫细胞包含在编码以下项的一种或多种多核苷酸中的突变:V-Set免疫调节受体(VISTA)、T细胞免疫球蛋白粘蛋白3(Tim-3)、具有Ig和ITIM结构域的T细胞免疫受体(TIGIT)、转化生长因子β受体II(TGFbRII)、含有调节因子X相关锚蛋白的蛋白质(RFXANK)、含PVR相关免疫球蛋白结构域(PVRIG)、淋巴细胞激活基因3(Lag3)、细胞毒性T淋巴细胞相关蛋白4(CTLA-4)、几丁质酶3样1(Chi311)、分化簇96(CD96)、B和T淋巴细胞相关(BTLA)、Tet甲基胞苷双加氧酶2(TET2)、,Sprouty RTK信号传导拮抗剂1(Spry1)、Sprouty RTK信号传导拮抗剂2(Spry2)、II类主要组织相容性复合物反式激活因子(CIITA)、分化簇3(CD3)、分化簇7(CD7)、分化簇33(CD33)、分化簇52(CD52)、分化簇123(CD123)、T细胞受体β恒定区1(TRBC1)、T细胞受体β恒定区2(TRBC2)、细胞因子诱导型含SH2蛋白质(CISH)、乙酰辅酶A乙酰转移酶1(ACAT1)、细胞色素P450家族11亚家族A成员1(Cyp11a1)、GATA结合蛋白3(GATA3)、核受体亚家族4A组成员1(NR4A1)、核受体亚家族4A组成员2(NR4A2)、核受体亚家族4A组成员3(NR4A3)、甲基化控制的J蛋白(MCJ)、Fas细胞表面

死亡受体(FAS)、或选择蛋白P配体体/P-选择蛋白糖蛋白配体-1(SELPG/PSGL1)。在另一种情况下,经修饰的免疫细胞表达嵌合抗原受体,该嵌合抗原受体包含,例如,具有对于与瘤形成相关的标志物(例如,B细胞成熟抗原(BCMA))的亲和性的细胞外结构域。

[1998] 本文提供一种使用具有降低的免疫原性和增加的抗瘤形成活性的经修饰的免疫细胞(例如,获自健康受试者的免疫细胞,或者T细胞诸如细胞毒性T细胞、调节T细胞或T辅助细胞)治疗瘤形成或液态癌症的方法,该经修饰的免疫细胞包含在CD33、PD-1、Cd2、TRAC多核苷酸或其组合中的一者或多者中的突变。在各种实施方案中,经修饰的免疫细胞包含在CD33、PD-1、CD52、TRAC或其组合中的两者或更多者中的突变。在各种实施方案中,经修饰的免疫细胞包含在CD33、PD-1、CD52、TRAC或其组合中的三者或更多者中的突变。在一种情况下,经修饰的免疫细胞包含在CD33、PD-1、CD52和TRAC中的突变。

[1999] 本文提供一种使用具有降低的免疫原性和增加的抗瘤形成活性的经修饰的免疫细胞(例如,获自健康受试者的免疫细胞,或者T细胞诸如细胞毒性T细胞、调节T细胞或T辅助细胞)治疗瘤形成或液态癌症的方法,该经修饰的免疫细胞包含在CD3、PD-1、Cd2、TRAC、CD7多核苷酸或其组合中的一者或多者中的突变。在各种实施方案中,经修饰的免疫细胞包含在CD3、PD-1、CD52、TRAC、CD7或其组合中的两者或更多者中的突变。在一种情况下,经修饰的免疫细胞包含在CD3和CD7中的突变。在各种实施方案中,经修饰的免疫细胞包含在CD3、PD-1、CD52、TRAC、CD7或其组合中的三者或更多者中的突变。在一种情况下,经修饰的免疫细胞包含在CD3、PD-1、CD52、TRAC和CD7中的突变。

[2000] 本文提供使用经修饰的免疫细胞(例如,获自健康受试者的免疫细胞,或者T细胞诸如细胞毒T细胞、调节T细胞或T辅助细胞)治疗瘤形成或液态癌症的方法,该经修饰的免疫细胞通过在免疫细胞中表达以下项或将以下项引入免疫细胞中的方法产生:核碱基编辑器多肽;两个或更多个向导RNA能够靶向编码选自由β-2微球蛋白(B2M)和程序性细胞死亡蛋白1(PD1)多肽所组成的群组的至少一个多肽的核酸分子;Cas12多肽;能够靶向编码TRAC的核酸分子的向导RNA;和编码嵌合抗原受体(CAR)或T细胞受体的DNA供体模板,其中中核碱基编辑器多肽包含核酸可编程DNA结合蛋白(napDNAbp)和胞苷脱氨酶。

[2001] 本文提供一种使用具有降低的免疫原性和增加的抗瘤形成活性的经修饰的免疫细胞(例如,获自健康受试者的免疫细胞,或者T细胞诸如细胞毒性T细胞、调节T细胞或T辅助细胞)治疗瘤形成或液态癌症的方法,该经修饰的免疫细胞包含在B2M或PD-1中的突变、编码嵌合抗原受体(CAR)的核酸在TRAC基因座处的插入以及在免疫细胞表面上表达CAR。

[2002] 在本发明的一些实施方案中,本文提供一种治疗有此需要的受试者的液态癌症,包括向有此需要的受试者给药有效治疗量的本文所述药物组合物。待通过此类方法治疗的液态癌症包括但不限于,淋巴瘤、白血病、骨髓瘤或其组合。

[2003] 在一种情况下,本文提供一种治疗有此需要的受试者的液态癌症,包括向有此需要的受试者给药有效治疗量的本文所述药物组合物,其中液态癌症包含淋巴瘤。待使用本文所述方法治疗的液态淋巴瘤包括但不限于,恶性浆细胞肿瘤,霍奇金淋巴瘤;结节性淋巴细胞为主的霍奇金淋巴瘤;真性红细胞增多症,霍奇金淋巴瘤;小淋巴细胞性淋巴瘤(SLL);弥散性大B细胞淋巴瘤;免疫母细胞性大细胞淋巴瘤;前体B成淋巴细胞性淋巴瘤;套细胞淋巴瘤;边缘带淋巴瘤;蕈样肉芽肿病;间变性大细胞淋巴瘤;Sezary综合征;前体T成淋巴细胞性淋巴瘤;和具有弥散性大B细胞淋巴瘤与经典霍奇金淋巴瘤之间的中间特征的未归类B

细胞淋巴瘤;B细胞非霍奇金淋巴瘤(NHL);滤泡性淋巴瘤;伯基特淋巴瘤;伯基特样淋巴瘤;边缘带淋巴瘤;讨套细胞淋巴瘤;弥漫性大B细胞淋巴瘤(DLBCL、原发性纵膈B细胞淋巴瘤;大B细胞淋巴瘤;前体B成淋巴细胞性淋巴瘤;具有弥散性大B细胞淋巴瘤与伯基特淋巴瘤之间的中间特征的未归类B细胞淋巴瘤;边缘带淋巴瘤;粘膜相关淋巴组织淋巴瘤;小细胞淋巴细胞性淋巴瘤;套细胞淋巴瘤;原发性纵隔(胸腺)大B细胞淋巴瘤;淋巴浆细胞性淋巴瘤(诸如华氏巨球蛋白血症);结节边缘区B细胞淋巴瘤;脾边缘带淋巴瘤;血管内大B细胞淋巴瘤;原发性渗出性淋巴瘤;淋巴瘤样肉芽肿病;富含T细胞/组织细胞的大B细胞淋巴瘤;原发性中枢神经系统淋巴瘤;原发性皮肤弥漫性大B细胞淋巴瘤(腿型);老年性EBV阳性弥漫性大B细胞淋巴瘤;与炎症相关的弥漫性大B细胞淋巴瘤;血管内大B细胞淋巴瘤;ALK阳性大B细胞淋巴瘤;浆母细胞性淋巴瘤;在HHV8相关的多中心Castleman病中发生的大B细胞淋巴瘤;纵膈(胸腺)大B细胞淋巴瘤;血管内大B细胞淋巴瘤;原发性渗出性淋巴瘤;鼻型、肠病型T细胞淋巴瘤;肝脾T细胞淋巴瘤;母细胞NK细胞淋巴瘤;原发皮肤间变性大细胞淋巴瘤;淋巴瘤样丘疹病;血管免疫母细胞性t细胞淋巴瘤;外周T细胞淋巴瘤;未指明的间变性大细胞淋巴瘤;经典霍奇金淋巴瘤(结节性硬化症,混合细胞结构,淋巴细胞丰富的,淋巴细胞耗尽或未耗尽);结节性淋巴细胞为主的霍奇金淋巴瘤;成人T细胞淋巴瘤;B细胞淋巴瘤-DLBCL、B细胞淋巴瘤-DLBCL-生发中心样;B细胞淋巴瘤-DLBCL-激活的B细胞样;非霍奇金淋巴瘤;原发性中枢神经系统(CNS)淋巴瘤;T细胞淋巴瘤;外周T细胞淋巴瘤,皮肤T细胞淋巴瘤(CTCL);间变性大细胞淋巴瘤(ALCL);AIDS相关淋巴瘤;肠病型T细胞淋巴瘤;肝脾T细胞淋巴瘤;母细胞NK细胞淋巴瘤;原发皮肤间变性大细胞淋巴瘤;淋巴瘤样丘疹病,血管免疫母细胞性T细胞淋巴瘤;外周T细胞淋巴瘤;未指明的间变性大细胞淋巴瘤;经典霍奇金淋巴瘤(结节性硬化症,混合细胞结构,淋巴细胞丰富的,淋巴细胞耗尽或未耗尽);和结节性淋巴细胞为主的霍奇金淋巴瘤;其他淋巴瘤;及此类淋巴瘤的组合。

[2004] 在一种情况下,本文提供一种治疗有此需要的受试者的液态癌症,包括向有此需要的受试者给药有效治疗量的本文所述药物组合物,其中液态癌症包含白血病。在一些情况下,白血病包括白血病前期。白血病前期包括,例如,骨髓增生异常综合征(MDS)。在一些情况下,白血病是急性白血病。急性白血病包括,例如,急性骨髓性白血病(AML)。急性白血病也包括,例如,急性淋巴性白血病或急性淋巴细胞性白血病(ALL);ALL包括B谱系ALL、T谱系ALL和T细胞急性淋巴细胞性白血病(T-ALL)。其他待使用本文所述方法治疗的白血病包括但不限于,毛细胞白血病(HCL);慢性粒细胞白血病(CIVIL);慢性粒单核细胞白血病(CMML);慢性淋巴细胞白血病(CLL);B细胞幼淋巴细胞白血病,粒细胞白血病;伯基特白血病;T细胞幼淋巴细胞白血病;T细胞大大颗粒淋巴细胞白血病;侵略性自然杀手(NK)细胞白血病;成人T细胞白血病(ATL);急性早幼粒细胞白血病(APML);幼淋巴细胞白血病(PLL);成红细胞白血病;急性巨核细胞白血病;大颗粒淋巴细胞白血病(LGF);慢性粒细胞白血病(CML);其他白血病;及此类白血病的组合。

[2005] 在一种情况下,本文提供一种治疗有此需要的受试者的液态癌症,包括向有此需要的受试者给药有效治疗量的本文所述药物组合物,其中液态癌症包含骨髓瘤。但使用本文所述方法治疗的骨髓瘤包括但不限于多发性骨髓瘤;阴燃多发性骨髓瘤;浆细胞骨髓瘤;非分泌性骨髓瘤,IgD骨髓瘤;骨硬化性骨髓瘤;其他骨髓瘤;以及此类骨髓瘤的组合。

[2006] 在一种情况下,本文提供一种治疗有此需要的受试者的液态癌症的方法,包括向

有此需要的受试者给药有效治疗量的本文所述药物组合物,其中液态癌症包含卡勒氏病;骨髓性白血病;浆细胞瘤;骨孤立性浆细胞瘤;髓外浆细胞瘤霉菌病真菌病/Sezary综合征;原发性皮肤CD30阳性T细胞淋巴增殖性疾病;嗜酸性粒细胞增多症;慢性嗜酸性粒细胞增多症;和其他造血细胞相关癌症。

[2007] 在一种实施方案中,向受试者给药至少0.1×10^5细胞、至少0.5×10^5细胞、至少1×10^5细胞、至少5×10^5细胞、至少1×10^6细胞、至少0.5×10^7细胞、至少1×10^7细胞、至少0.5×10^8细胞、至少1×10^8细胞、至少0.5×10^9细胞、至少1×10^9细胞、至少2×10^9细胞、至少3×10^9细胞、至少4×10^9细胞、至少5×10^9细胞、或至少1×10^{10}细胞。在特定实施方案中,向受试者给药约1×10^7细胞至约1×10^9细胞、约2×10^7细胞至约0.9×10^9细胞、约3×10^7细胞至约0.8×10^9细胞、约4×10^7细胞至约0.7×10^9细胞、约5×10^7细胞至约0.6×10^9细胞、或约5×10^7细胞至约0.5×10^9细胞。

[2008] 在一些实施方案中,向受试者给药至少0.1×10^4细胞/kg体重、至少0.5×10^4细胞/kg体重、至少1×10^4细胞/kg体重、至少5×10^4细胞/kg体重、至少1×10^5细胞/kg体重、至少0.5×10^6细胞/kg体重、至少1×10^6细胞/kg体重、至少0.5×10^7细胞/kg体重、至少1×10^7细胞/kg体重、至少0.5×10^8细胞/kg体重、至少1×10^8细胞/kg体重、至少2×10^8细胞/kg体重、至少3×10^8细胞/kg体重、至少4×10^8细胞/kg体重、至少5×10^8细胞/kg体重、或至少1×10^9细胞/kg体重。在特定实施方案中,向受试者给药约1×10^6细胞/kg体重至约1×10^8细胞/kg体重、约2×10^6T细胞/kg体重至约0.9×10^8细胞/kg体重、约3×10^6细胞/kg体重至约0.8×10^8细胞/kg体重、约4×10^6细胞/kg体重至约0.7×10^8细胞/kg体重、约5×10^6细胞/kg体重至约0.6×10^8细胞/kg体重、或约5×10^6细胞/kg体重至约0.5×10^8细胞/kg体重。

[2009] 本领域技术人员将会了解,在特定实施方案中可能需要设想的多次给药t来实现所希望的治疗。例如,组合物可以在历时1周、2周、3周、1个月、2个月、3个月、4个月、5个月、6个月、1年、2年、5年、10年或更久的时间跨度内向受试者给药1、2、3、4、5、6、7、8、9或10次或更多次。在任何此类方法中,该方法可包括向受试者给药有效量的经修饰的免疫细胞,并且该方法增加或降低免疫应答。在任何此类方法中,该方法可包括每天给药一个或多个剂量的有效量的经修饰的免疫细胞。在任何此类方法中,该方法可包括每天给药两个或更多个剂量的有效量的经修饰的免疫细胞。在任何此类方法中,该方法可包括每天给药三个或更多个剂量的有效量的经修饰的免疫细胞。在任何此类方法中,该方法可包括每周给药一个或多个剂量的有效量的经修饰的免疫细胞。在任何此类方法中,该方法可包括每周给药两个或更多个剂量的有效量的经修饰的免疫细胞。在任何此类方法中,该方法可包括每周给药三个或更多个剂量的有效量的经修饰的免疫细胞。在任何此类方法中,该方法可包括每个月给药一个或多个剂量的有效量的经修饰的免疫细胞。在任何此类方法中,该方法可包括每个月给药两个或更多个剂量的有效量的经修饰的免疫细胞。在任何此类方法中,该方法可包括每个月给药三个或更多个剂量的有效量的经修饰的免疫细胞。

[2010] 在某些实施方案中,可能希望向受试者给药激活的药物组合物,然后重新抽血(或已经执行血液成分单采),激活来自其的T细胞,并将这些激活且扩张的T细胞回输给患者。这一过程每数周可进行多次。在某些实施方案中,T细胞可从来自10cc至400cc的抽取血液激活。在某些实施方案中,T细胞可从来自20cc、30cc、40cc、50cc、60cc、70cc、80cc、90cc、100cc、150cc、200cc、250cc、300cc、350cc、或400cc或更多的抽取血液激活。

[2011] 本文设想的药物组合物给药可以使用传统技术进行,包括但不限于输液、输血或肠胃外地。在一些实施方案中,肠胃外给药包括血管内、静脉内、肌肉内、动漫、鞘内、肿瘤内、皮内、腹膜内、透过器官、皮下、表皮下、关节内、包膜下、蛛网膜下和胸骨内输液或注射。

[2012] 个体受试者的应答可以表征为完全缓解、部分缓解、病情稳定和病情进展。在一些实施方案中,应答是完全缓解(CR)。完全缓解定义为全部循环肿瘤细胞(CTC)或单核血细胞(MNBC)的消失,即,任何病理淋巴结(无论靶标或非靶标)必须具有减少至<10mm的短轴。在一些示例(例如,AML)中,完全缓解可以定义如下:骨髓原始细胞<5%;不存在具有奥氏小体棒的原始细胞;不存在髓外疾病;绝对中性粒细胞计数>1.0x 10^9/L(1000/μL);血小板计数>100x 10^9/L(100 000/μL);和红细胞输血的非依赖性。在某些实施方案中,应答是伴有不完全恢复的CR(CRi)。在一些示例(例如,AML)中,伴有不完全恢复的CR可以定义为包括全部CR评判标准,但残余中性粒细胞减少(<1.0x 10^9/L[1000/μL])或血小板减少症(<100x 10^9/L[100 000/μL])除外。在某些实施方案中,应答是形态学无白血病状态。在一些示例(例如,AML)中,形态学无白血病状态可以定义为包括骨髓原始细胞<5%;不存在具有奥氏小体棒的原始细胞;不存在髓外疾病;以及不需要血液学恢复。在某些实施方案中,应答是部分缓解(PR)。部分缓解可以定义为血清中循环肿瘤细胞(CTC)或单核血细胞(MNBC)的直径总和平均减少至少30%,以基线直径总和作为参考。在一些示例(例如,AML)中,部分缓解可以定义为包括CR的全部血液学评判标准;骨髓原始细胞百分比降低到5%至25%;以及治疗前骨髓原始细胞百分比减少至少50%。在某些实施方案中,应答是形态学无白血病状态。在一些示例(例如,AML)中,形态学无白血病状态可以定义为包括骨髓原始细胞<5%;不存在具有奥氏小体棒的原始细胞;不存在髓外疾病;以及不需要血液学恢复。在某些实施方案中,应答是复发。

[2013] 在一些示例(例如,AML)中,复发可以定义为包括骨髓原始细胞<5%;不存在具有奥氏小体棒的原始细胞;不存在髓外疾病;以及不需要血液学恢复。在一些实施方案中,应答是病情进展(PD)。病情进展可以定义为循环肿瘤细胞(CTC)或单核血细胞(MNBC)的数量增加至少20%,以研究期间的最小数量为参考(这包括基线数量,如果基线数量是最小的);以及循环肿瘤细胞(CTC)或单核血细胞(MNBC)的绝对数量增加至少1、至少2、至少3、至少4、至少5、至少6、至少7、至少8、至少9、至少10、至少11、至少12、至少13、至少14、至少15、至少16、至少17、至少18、至少19、至少20、至少30、至少40、至少50、或至少100个或更多个。

[2014] 一个或多个新病灶的出现也视为进展。在一些实施方案中,疾病可以是病情稳定(SD)。病情稳定的特征可以在于,液态癌症细胞数量既没有足够降低到符合PR标准也没有足够增加到符合PD标准,以CTC和/或MNBC在研究期间的最小数量为参考。在某些实施方案中,应答是病理完全缓解。病理完全缓解,例如,如由病理学家在检查在手术或活检时间去除的组织后所确定,一般是指手术标本中不存在侵袭性和/或非侵袭性液态癌症细胞的组织学证据。

[2015] 在一些实施方案中,药物组合物每周给药两次或三次。在另一实施方案中,药物组合物每周给药两次。在另一实施方案中,药物组合物每2或3周给药一次。在另一实施方案中,药物组合物每1或2周给药一次。在另一实施方案中,药物组合物在21天周期的第1、4、8和11天给药。在另一实施方案中,药物组合物在28天周期的第1、8和15天给药。

[2016] 在一些实施方案中,所给药化合物的量是约0.5-30mg每千克人类受试者体重。在

另一实施方案中,所给药化合物的量是约0.5-20mg每千克人类受试者体重。在另一实施方案中,所给药化合物的量是约0.5-10mg每千克人类受试者体重。在另一实施方案中,所给药化合物的量是约0.04mg、约0.08mg、约0.16mg、约0.32mg、约0.64mg、约1.25mg、约1.28mg、约1.92mg、约2.5mg、约3.56mg、约3.75mg、约5.0mg、约7.12mg、约7.5mg、约10mg、约14.24mg、约15mg、约20mg、或约30mg每千克人类受试者体重。在另一实施方案中,所给药化合物的量是约1.92mg、约3.75mg、约7.5mg、约15.0mg、或约30.0mg每千克人类受试者体重,且该药物每周给药两次。在另一实施方案中,所给药化合物的量是约1.28mg、约2.56mg、约5.0mg、约10mg、或约20mg每千克人类受试者体重,且该药物每周给药两次。在另一实施方案中,所给药化合物的量是约1.92mg、约3.75mg、约7.5mg、约15.0mg、或约30.0mg每千克人类受试者体重,且该药物每周给药一次。在另一实施方案中,所给药化合物的量是约1.28mg、约2.56mg、约5.0mg、约10mg、或约20mg每千克人类受试者体重,且该药物每周给药一次。在另一实施方案中,所给药化合物的量是约1.92mg、约3.75mg、约7.5mg、约15.0mg、或约30.0mg每千克人类受试者体重,且该药物每天给药一次,在每个七天时间内执行三次、五次或七次。在另一实施方案中,化合物经静脉内每天给药一次,在七天时间内执行七次。在另一实施方案中,所给药化合物的量是约1.28mg、约2.56mg、约5.0mg、约10mg、或约20mg每千克人类受试者体重,且该药物每天给药一次,在每个七天时间内执行三次、五次或七次。在另一实施方案中,化合物经静脉内每天给药一次,在七天时间内执行七次。

[2017] 在一些实施方案中,化合物历经0.25h、0.5h、1h、2h、3h、4h、5h、6h、7h、8h、9h、10h、11h或12h的时间给药。在另一实施方案中,化合物历经0.25至2h的时间给药。在另一实施方案中,化合物历经1h的时间逐步给药。在另一实施方案中,化合物历经2h的时间逐步给药。

[2018] 在一些实施方案中,化合物的体内循环半衰期是约1h、2h、3h、4h、5h、6h、7h、8h、9h、10h或12h。在另一实施方案中,化合物的体内循环半衰期是约4h。在另一实施方案中,化合物的体内循环半衰期是约6h。在一些实施方案中,化合物的生物组织半衰期是约1h、2h、3h、4h、5h、6h、7h、8h、9h、10h或12h。在另一实施方案中,化合物的生物组织半衰期是约10h。

[2019] 在一些实施方案中,自治疗开始后1个月的时间内,治疗导致液态癌症细胞的数量降低约95%、约90%、约85%、约80%、约75%、约70%、约65%、约60%、约55%、约50%、约45%、约40%、约35%、约30%、约25%、约20%、约15%、约10%、或约5%。在另一实施方案中,自治疗开始后1个月的时间内,治疗导致液态癌症细胞的数量降低至少约60%、至少约65%、至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、或至少约95%。在另一实施方案中,自治疗开始后1年的时间内,治疗导致液态癌症细胞的数量降低约95%、约90%、约85%、约80%、约75%、约70%、约65%、约60%、约55%、约50%、约45%、约40%、约35%、约30%、约25%、约20%、约15%、约10%、或约5%。在另一实施方案中,自治疗开始后1年的时间内,治疗导致液态癌症细胞的数量降低至少约60%、至少约65%、至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、或至少约95%。在另一实施方案中,自治疗开始后6个月的时间内,治疗导致液态癌症细胞的数量降低约95%、约90%、约85%、约80%、约75%、约70%、约65%、约60%、约55%、约50%、约45%、约40%、约35%、约30%、约25%、约20%、约15%、约10%、或约5%。在另一实施方案中,自治疗开始后6个月的时间内,治疗导致液态癌症细胞的数量降低至少约60%、至少约65%、至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、或至少约95%。在另一实施方案中,自治疗开始后3个

月的时间内,治疗导致液态癌症细胞的数量降低约95%、约90%、约85%、约80%、约75%、约70%、约65%、约60%、约55%、约50%、约45%、约40%、约35%、约30%、约25%、约20%、约15%、约10%、或约5%。在另一实施方案中,自治疗开始后3个月的时间内,治疗导致液态癌症细胞的数量降低至少约60%、至少约65%、至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、或至少约95%。

[2020] 在一些实施方案中,治疗导致人类受试者的存活时间比该人类受试者未使用该化合物治疗时该人类受试者的预期存活时间增加。在另一实施方案中,人类受试者的存活时间增加至少30天。在另一实施方案中,人类受试者的存活时间增加至少3个月。在另一实施方案中,人类受试者的存活时间增加至少6个月。在另一实施方案中,人类受试者的存活时间增加至少1年。

[2021] 在一些实施方案中,人类受试者是难治性的和/或对于一种或多种液体癌症的治疗不耐受。在另一实施方案中,人类受试者已经进行至少一种不成功的液态癌症既往治疗和/或疗法。

[2022] 在一些实施方案中,待使用所述方法治疗的人类受试者是儿童(例如,年龄为0至18岁)。在其他实施方案中,待使用所述方法治疗的人类受试者是成人(例如,年龄为18+岁)。

[2023] 试剂盒

[2024] 本发明提供用于治疗受试者的瘤形成的试剂盒。在一些实施方案中,试剂盒用于治疗实体肿瘤。在其他实施方案中,试剂盒用于治疗液态肿瘤。在一些实施方案中,试剂盒用于治疗血液学癌症。在一些实施方案中,血液学癌症是B细胞癌症。在一些实施方案中,血液学癌症是白血病、骨髓瘤和/或淋巴瘤。在一些实施方案中,试剂盒用于治疗选自由以下所组成的群组的瘤形成:T细胞急性成淋巴细胞性白血病(T-ALL)、蕈样肉芽肿(MF)、Sézary综合征(SS)、外周T/NK细胞淋巴瘤、间变性大细胞淋巴瘤ALK+、原发性皮肤T细胞淋巴瘤、T细胞大颗粒淋巴细胞型白血病、血管免疫母细胞性T/NK细胞淋巴瘤、肝脾T细胞淋巴瘤、原发性皮肤CD30+淋巴增生性疾病、结外NK/T细胞淋巴瘤、成人T细胞白血病/淋巴瘤、T细胞幼淋巴细胞白血病、皮下脂膜炎样T细胞淋巴瘤、原发性皮肤γδT细胞淋巴瘤、侵袭性NK细胞白血病和肠病相关T细胞淋巴瘤。在一些实施方案中,试剂盒用于治疗T细胞急性成淋巴细胞性白血病(T-ALL)。在一些实施方案中,试剂盒用于治疗急性骨髓性白血病(AML)。在一些实施方案中,试剂盒用于治疗人类受试者。

[2025] 在一些实施方案中,试剂盒包含本文所提供的任何嵌合抗原受体。在一些实施方案中,试剂盒包含编码本文所提供的任何嵌合抗原受体的核酸。在一些实施方案中,试剂盒包含本文所提供的经修饰的免疫细胞。在一些实施方案中,试剂盒包括表达CD3 CAR、CD5 CAR、CD7 CAR、CD33 CAR和/或CD123 CAR的经修饰的免疫细胞。在一些实施方案中,任何免疫细胞进一步包含在CD3、CD5、CD7、CD33、CD123、TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和/或PD1或其组合中的一者或多者中的突变。

[2026] 在一些实施方案中,试剂盒包括本文所提供的经修饰的免疫细胞群。在一些实施方案中,试剂盒包含本文所提供的药物组合物。在一些实施方案中,试剂盒包括经CD123修饰的免疫细胞群或者药物组合物,该药物组合物包含经CD123修饰的免疫细胞或经修饰的免疫细胞群。在一些实施方案中,试剂盒包括经CD5修饰的免疫细胞群或者药物组合物,该

药物组合物包含经CD5修饰的免疫细胞或经修饰的免疫细胞群。在一些实施方案中,试剂盒包括经CD7修饰的免疫细胞群或者药物组合物,该药物组合物包含经CD7修饰的免疫细胞或经修饰的免疫细胞群。在一些实施方案中,试剂盒包括经CD33修饰的免疫细胞群或者药物组合物,该药物组合物包含经CD33修饰的免疫细胞或经修饰的免疫细胞群。在一些实施方案中,试剂盒包括经CD3修饰的免疫细胞群或者药物组合物,该药物组合物包含经CD3修饰的免疫细胞或经修饰的免疫细胞群。

[2027] 在一些实施方案中,试剂盒进一步包括碱基编辑器多肽或编码碱基编辑器多肽的多核苷酸,其中碱基编辑器多肽包含核酸可编程DNA结合蛋白(napDNAbp)和脱氨酶。在一些实施方案中,napDNAbp是Cas9或Cas12。在一些实施方案中,编码碱基编辑器的多核苷酸是mRNA序列。在一些实施方案中,脱氨酶是胞苷脱氨酶或腺苷脱氨酶。

[2028] 本发明提供包含核酸构建体的试剂盒,该核酸构建体包含编码胞苷或腺苷脱氨酶核碱基编辑器的核苷酸序列和至少一个向导RNA,每个向导RNA具有核酸序列,该核酸序列与编码TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M、PD1、CBLB和/或CTLA4的基因的核酸序列至少85%互补。在一些实施方案中,编码胞苷或腺苷脱氨酶的核苷酸序列包含异源启动子,该异源启动子驱动胞苷或腺苷脱氨酶核碱基编辑器的表达。

[2029] 本公开的一些方面提供包含核酸构建体的试剂盒,该核酸构建体包含编码和碱基编辑器的核苷酸和向导RNA。在一些实施方案中,核酸构建体包含异源启动子,该异源启动子驱动核碱基编辑器的表达。在一些实施方案中,本公开提供包含核酸构建体的试剂盒,该核酸构建体包含(a)编码本文所提供的融合至胞苷或腺苷脱氨酶的(a)Cas9结构域的核苷酸序列;和(b)异源启动子,其驱动(a)的序列的表达。在一些实施方案中,Cas9结构域融合至胞苷脱氨酶。在一些实施方案中,Cas9结构域融合至腺苷脱氨酶。

[2030] 在一些实施方案中,试剂盒包含胞苷脱氨酶核碱基编辑器和向导RNA。在一些实施方案中,试剂盒包含腺苷脱氨酶核碱基编辑器和向导RNA。在一些实施方案中,试剂盒进一步包含一个或多个向导核酸序列。在一些实施方案中,该一个或多个向导核酸序列靶向CD3、CD5、CD7、CD33或CD123。在一些实施方案中,该一个或多个向导核酸序列各自靶向CD3、CD5、CD7、CD33、CD123、TRAC、LAG-3、FAS、CIITA、TRBC1、TRBC2、CD52、B2M和/或PD1中的一者。在一些实施方案中,向导RNA具有核酸序列,该核酸序列与编码CD5的基因的核酸序列至少85%互补。在一些实施方案中,试剂盒包含核碱基编辑器和CD5向导RNA。在一些实施方案中,向导RNA具有核酸序列,该核酸序列与编码CD7的基因的核酸序列至少85%互补。在一些实施方案中,试剂盒包含核碱基编辑器和CD7向导RNA。在一些实施方案中,向导RNA具有核酸序列,该核酸序列与编码CD33的基因的核酸序列至少85%互补。在一些实施方案中,试剂盒包含核碱基编辑器和CD33向导RNA。在一些实施方案中,向导RNA具有核酸序列,该核酸序列与编码CD3的基因的核酸序列至少85%互补。在一些实施方案中,试剂盒包含核碱基编辑器和CD3向导RNA。在一些实施方案中,向导RNA具有核酸序列,该核酸序列与编码CD123的基因的核酸序列至少85%互补。在一些实施方案中,试剂盒包含胞苷脱氨酶核碱基编辑器和CD123向导RNA。在一些实施方案中,试剂盒可进一步包括一个或多个另外的向导RNA,每个向导RNA具有核酸序列,该核酸序列与编码TRAC、LAG-3、FAS、PD1、B2M、CIITA、TRBC1、TRBC2和/或CD52的基因的核酸序列至少85%互补。

[2031] 瘤形成治疗试剂盒可进一步包括关于使用经修饰的免疫细胞治疗瘤形成的书面

使用说明。在其他实施方案中,使用说明包括以下至少一项:注意事项;警告;临床研究;和/或参考资料。使用说明可以直接印刷在容器(当存在时)上,或作为标签施加到容器上,或作为独立的纸张、小册子、卡片或折页提供在容器内或随容器提供。在进一步的实施方案中,试剂盒可包括标签或独立插页(包装插页)形式的针对合适的操作参数的使用说明。在又一实施方案中,试剂盒可包含一个或多个具有一种或多种适当阳性和阴性对照样品的容器,这些对照样品待用作标品用于检测、校正或归一化。试剂盒可进一步包含第二容器,该第二容器包含药学上可接受的缓冲液,诸如(无菌)磷酸盐缓冲盐水、林格溶液或葡萄糖溶液。它可以进一步包括来自商用和使用者观点的其他所希望的材料,包括其他缓冲剂、稀释剂、过滤器、针、注射器和具有使用说明的包装插页。

[2032] 除非另外指定,否则本发明的实践采用分子生物学(包括重组技术)、微生物学、细胞生物学、生物化学和免疫学的传统技术,这些技术完全在该领域技术人员的权限之内。此类技术在文献中有完整阐述,例如,《分子克隆:实验室手册(第二版)》(Molecular Cloning:ALaboratory Manual,second edition(Sambrook,1989));《寡核苷酸合成》(Oligonucleotide Synthesis(Gait,1984));《动物细胞培养》(Animal Cell Culture(Freshney,1987));《酶学方法:实验免疫学手册》(Methods in Enzymology Handbook of Experimental Immunology(Weir,1996));《用于哺乳动物细胞的转基因载体》(Gene Transfer Vectors for Mammalian Cells(Miller and Calos,1987));《分子生物学现代方法》(Current Protocols in Molecular Biology(Ausubel,1987));《PCR:聚合酶链反应》(PCR:The Polymerase Chain Reaction,(Mullis,1994));《免疫学现代方法》(Current Protocols in Immunology(Coligan,1991))。这些技术可用于生产本发明的多核苷酸和多肽,并因此可视为作出并实践本发明。尤其可用于具体实施方案的技术将在以下章节中讨论。

[2033] 详述以下实施例以向该领域技术人员提供对如何制备和使用本发明的检验、筛选和治疗性方法的完全公开和说明,并且以下实施例不试图限制发明人所认为的本发明的范畴。

[2034] [实施例]

[2035] 实施例1:在免疫细胞中表达的基因中剪接位点的中断和终止密码子的引入

[2036] 使用核碱基编辑器BE4来中断剪接位点并将终止密码子插入到在免疫细胞中表面的基因子集中。质粒构建体pCMV_BE4max编码BE4,其包含具有胞苷脱氨酶活性的APOBEC-1胞苷脱氨酶结构域、包含D10A突变且具有切口酶活性的Cas9结构域、和两个尿嘧啶DNA糖基化酶抑制剂(UGI)结构域。UGI是一种83个氨基酸残基的蛋白质,其源自枯草芽孢杆菌噬菌体PBS1,有效地阻断以编辑在免疫细胞中表达的某些基因的剪接位点。BE4进一步包含N端和C端核定位序列(NLS)。

[2037] >pCMV_BE4max

[2038] ATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTAGAGATCCGCGGCCGCT

AATACGACTCACTATAGGGAGAGCCGCCACCATGAAACGGACAGCCGACGGAAGCGAGTTCGAGTCACCAAAGAAGA
AGCGGAAAGTCTCCTCAGAGACTGGGCCTGTCGCCGTCGATCCAACCCTGCGCCGCCGGATTGAACCTCACGAGTTT
GAAGTGTTCTTTGACCCCCGGGAGCTGAGAAAGGAGACATGCCTGCTGTACGAGATCAACTGGGGAGGCAGGCACTC
CATCTGGAGGCACACCTCTCAGAACACAAATAAGCACGTGGAGGTGAACTTCATCGAGAAGTTTACCACAGAGCGGT
ACTTCTGCCCCAATACCAGATGTAGCATCACATGGTTTCTGAGCTGGTCCCCTTGCGGAGAGTGTAGCAGGGCCATC
ACCGAGTTCCTGTCCAGATATCCACACGTGACACTGTTTATCTACATCGCCAGGCTGTATCACCACGCAGACCCAAG
GAATAGGCAGGGCCTGCGCGATCTGATCAGCTCCGGCGTGACCATCCAGATCATGACAGAGCAGGAGTCCGGCTACT
GCTGGCGGAACTTCGTGAATTATTCTCCTAGCAACGAGGCCCACTGGCCTAGGTACCCACACCTGTGGGTGCGCCTG
TACGTGCTGGAGCTGTATTGCATCATCCTGGGCCTGCCCCCTTGTCTGAATATCCTGCGGAGAAAGCAGCCCCAGCT
GACCTTCTTTACAATCGCCCTGCAGTCTTGTCACTATCAGAGGCTGCCACCCCACATCCTGTGGGCCACAGGCCTGA
AGTCTGGAGGATCTAGCGGAGGATCCTCTGGCAGCGAGACACCAGGAACAAGCGAGTCAGCAACACCAGAGAGCAGT
GGCGGCAGCAGCGGCGGCAGCGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCGT
GATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGA
ACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGA
TACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTT
CTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCG
TGGACGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAG
GCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCT
GAACCCCGACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACC
CCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTG
ATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAA
CTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACA
ACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTG
AGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCA
CCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACC
AGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATC
CTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTT
CGACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACC
CATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCC
AGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGT
GGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGC
TGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGA
ATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGT
GACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAG
ATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAG
GAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCT
GAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGC
TGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGC
TTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGT

GTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGC
AGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCC
AGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGA
GCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACC
TGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATC
GTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAG
CGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTA
CCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAG
AGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGA
CGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATT
TCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACC
GCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGAT
GATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCA
AGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATC
GTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGAC
CGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGA
AGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTG
GAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGA
GAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGT
ACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTG
GCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAA
TGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCA
AGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGA
GAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACAC
CACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCC
TGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGTGACAGCGGCGGGAGCGGCGGGAGCGGGGGGAGCACTAAT
CTGAGCGACATCATTGAGAAGGAGACTGGGAAACAGCTGGTCATTCAGGAGTCCATCCTGATGCTGCCTGAGGAGGT
GGAGGAAGTGATCGGCAACAAGCCAGAGTCTGACATCCTGGTGCACACCGCCTACGACGAGTCCACAGATGAGAATG
TGATGCTGCTGACCTCTGACGCCCCCGAGTATAAGCCTTGGGCCCTGGTCATCCAGGATTCTAACGGCGAGAATAAG
ATCAAGATGCTGAGCGGAGGATCCGGAGGATCTGGAGGCAGCACCAACCTGTCTGACATCATCGAGAAGGAGACAGG
CAAGCAGCTGGTCATCCAGGAGAGCATCCTGATGCTGCCCGAAGAAGTCGAAGAAGTGATCGGAAACAAGCCTGAGA
GCGATATCCTGGTCCATACCGCCTACGACGAGAGTACCGACGAAAATGTGATGCTGCTGACATCCGACGCCCCAGAG
TATAAGCCCTGGGCTCTGGTCATCCAGGATTCCAACGGAGAGAACAAAATCAAAATGCTGTCTGGCGGCTCAAAAAG
AACCGCCGACGGCAGCGAATTCGAGCCCAAGAAGAAGAGGAAAGTCTAACCGGTCATCATCACCATCACCATTGAGT
TTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCT
TGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGT
CATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGA
TGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCGATACCGTCGACCTCTAGCTAGAGCTTGG
CGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGC

ATAAAGTGTAAAGCCTAGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCA
GTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCT
CTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCG
GTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAA
CCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAA
GTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCT
GTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACG
CTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACC
GCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACT
GGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACAC
TAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCG
GCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAA
GAAGATCCTTTGATCTTTTCTACGGGGTCTGACACTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAG
ATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGT
AAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATA
GTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACC
GCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTC
CTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGT
TTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGG
TTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCG
TTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCA
TCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTG
CTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTT
CTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGA
TCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAAT
AAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTC
TCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTG
CCACCTGACGTCGACGGATCGGGAGATCGATCTCCCGATCCCCTAGGGTCGACTCTCAGTACAATCTGCTCTGATGC
CGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTA
CAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTA
CGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAG
CCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCAT
TGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTA
CGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATC

[2039] 为了查明BE4在敲低或敲除免疫细胞中的蛋白质表达中的有效性,将第一个免疫细胞群用编码BE4的mRNA和sgRNA共转染,该sgRNA靶向与TRAC外显子2、TRAC外显子4或B2M外显子1的供体或受体剪接位点的G碱基互补的C碱基,取决于特异性靶标位点。mRNA通过体外转录产生(TriLin Biotechnologies)。简而言之,将4g的BE4 mRNA和2g的合成gRNA电穿孔到1M CD3+ T细胞内(Nucleofector™平台,Lonza Bioscience)。然后将细胞培养3天以进

行足够时间的碱基编辑。为了进行比较,将第二个免疫细胞群用编码Cas9核酸酶和靶向B2M外显子1的供体剪接位点的G碱基的sgRNA共转染。未观察到BE4编辑与Cas9编辑之间的明显差异,每个经编辑的基因的敲低率均大于90%,而未经电穿孔的对照细胞没有明显的敲低(图2)。

[2040] 假设如果用编码BE4的mRNA(其催化单链缺口)或用Cas9核酸酶(其催化双链断裂)转染细胞,则造成观察到的靶向基因敲低的基因修饰将会有所不同。为了测试这一假设,用编码BE4的2g BE4/1g sgRNA(中)或4g BE4/2g sgRNA(高)RNA和靶向B2M外显子1的供体剪接位点的G碱基的sgRNA共转染免疫细胞。在孵育3、5和7天后,收集DNA并测序。参考图3,大多数碱基编辑表明了仅中断剪接位点并且以预期方式进行(即,并入了反义链中的C到T转换,导致有义链中的G到A转换)。这些结果与从用Cas9核酸酶转染的细胞中获得的结果形成对比,后者显示,在用Cas9转染的细胞中,大多数编辑是插入缺失(图3)。

[2041] 剪接位点的中断和终止密码子的插入可能在敲低靶标基因的表达中有效。BE4介导的对TRAC外显子4中剪接受体以及B2M外显子1和PDCD1外显子1中剪接供体的编辑导致全长蛋白质的表达减少(图4和5)。在剪接位点中观察到的BE4介导的改变是C到T转换,但也观察到了插入缺失和C到G转换。将赭色终止密码子插入到PDCD1基因的外显子2中(其中该外显子中的连续胞苷残基被靶向并编辑为胸苷残基),也导致基因表达的显著敲低,尽管敲低程度低于关于TRAC和B2M基因所见(图4)。这些结果进一步表明,BE4介导的对免疫细胞中表达的基因的单个或连续胞苷碱基的编辑导致基因表达的有效减少。

[2042] 实施例2:剪接位点中断和终止密码子插入的计算机模拟分析

[2043] 为了确定所设计的gRNA是否会结合至非位点靶标,使用CAS-OFFinder分析gRNA的核酸序列。参考图6,“X”凸起类型指示gRNA与基因DNA对齐并且任何矛盾之处均为误配。随着误配数从一增加到四,潜在的非位点结合增加。例如,关于TRAC外显子4剪接受体的结果显示,当存在三个误配时,存在26种非位点结合可能性,而存在四个误配时则存在164种可能性。

[2044] 如果gRNA具有凸起,其中gRNA具有二十个碱基对但与基因组DNA的十九个碱基对对齐,则导致凸起。再参考图6,当TRAC外显子4剪接受体gRNA具有一个碱基对的凸起时,非位点结合可能性的数量随着误配的增加而增加;然而,当不存在凸起(即,当凸起尺寸为零时)时,该可能性的数量显著降低。

[2045] 实施例3:免疫细胞内的多重碱基编辑

[2046] 为了确定BE4是否将会介导多个基因的碱基编辑以生成多敲低细胞,用编码BE4碱基编辑器的mRNA连同靶向B2M、TRAC、PD1或其组合中的靶标特异性位点的sgRNA一起共转染免疫细胞。参考图7,BE4系统引起有效的敲低,如通过流式细胞术所测量,以鉴定在单、双、三基因编辑中蛋白质生产降低的细胞百分比。将细胞基于B2M和CD3表达进行选通,其中CD3表达充当TRAC表达的指标。因为PD1染色效率低下,并未执行对表达这一蛋白质的细胞的直接测量。在具有单、双和三基因编辑的细胞群之间未观察到差异,且经修饰以敲低B2M、TRAC和PD1(三基因编辑)的表达的免疫细胞可检测地不同于未经修饰的免疫细胞(图8)。

[2047] 造成蛋白质表达降低的基因修饰总结在图9中。具体地且类似于实施例1中所述的导致在单基因修饰中表达降低的机制,C到T转换构成了在经修饰的B2M单修饰基因细胞群中和B2M+PD1、B2M+TRAC和B2M+TRAC+PD1多修饰基因细胞群中观察到的大量编辑。插入缺失

和转换构成了在经编辑的基因中的基因改变的不显著的很小一部分。

[2048] 因此,通过碱基编辑对三个基因座进行同步修饰产生了高效敲除而没有可检测的易位事件,如通过单向靶向测序(UDiTaS;Giannoukos等人,BMC Genomics.2018Mar 21;19(1):212.doi:10.1186/s12864-018-4561-9)所评估。此外,在经BE4编辑的基因中未检测到易位。采用微滴式数字聚合酶链反应(ddPCR)策略(图10)来检测经BE4编辑的B2M、TRAC和PD1基因之间的易位。使用QX200微滴式数字仪器(Bio-Rad),用下一代测序(NGS)来分析从经BE4或Cas9编辑以生成B2M+TRAC+PD1编辑的细胞提取的DNA,以确定BE4和Cas9编辑的确切序列。如图11的左图所示,在大多数细胞中,B2M、TRAC和PD1基因被修饰。ddPCR分析显示,在经BE4编辑的细胞中不存在易位,但在大约1.7%的经Cas9编辑的细胞中观察到了易位(图11,右图)。表21进一步示例性说明在经BE4编辑的细胞中没有观察到易位。

[2049] 表21

[2050]

碱基编辑器	易位	对照扩增子微滴	实验扩增子微滴
Cas9核酸酶	B2M-TRAC	61,206	585
	B2M-PDCD1	55,970	291
	PDCD1-TRAC	59,600	112
BE4	B2M-TRAC	90,717	0
	B2M-PDCD1	89,028	0
	PDCD1-TRAC	83,501	0

[2051] 实施例4:BE4介导的对Cbl原癌基因B(CBLB)的编辑

[2052] Cbl-b是一种T细胞受体(TCR)信号传导蛋白,其负向调节TCR复合物信号传导(图12)。因为T细胞在Cbl-b信号传导被抑制时具有较低的活化阈值,敲除或敲低这一基因将会显著改善T细胞或表达CAR的T细胞的有效性。为了确定Cbl-b基因是否对胞苷脱氨基介导的修饰敏感,用编码BE4的mRNA和sgRNA(其靶向外显子8和16的剪接位点受体、外显子8、10、11和12的剪接位点供体,或者将会促进STOP密码子插入到外显子1、4和8中)转染细胞。所得细胞使用流式细胞术分析。

[2053] 参考图13,中断外显子12的剪接位点供体和外显子8的剪接位点受体导致Cbl-b表达的最大减少(分别为67.2%和57.4%)。在用外显子8剪接位点受体和外显子12剪接位点供体sgRNA转染的细胞中,略微超过60%的细胞被成功地编辑(图13,柱状图)。

[2054] 实施例5:免疫细胞中的Cas12b核酸酶表征

[2055] Cas12b/c2c1位点特异性地靶向并切割双链核酸分子的两条链。通过确定两种不同的Cas12b/c2c1蛋白(BhCas12b和BvCas12b)介导靶标核酸分子中的插入缺失的倾向性,表征了这两种酶。将编码Cas12b/c2c1蛋白的mRNA连同向导RNA电穿孔到T细胞中,该向导RNA对于GRIN2B基因中的基因座和DNMT1基因中的基因座具有特异性。将细胞培养3至5天,然后分离细胞DNA。通过下一代测序来确定插入缺失率。参考图14,关于所具有的在GRIN2B基因中的插入缺失百分比,与分离自经BvCas12b蛋白处理的细胞的DNA(大约20%)相比,分离自经BhCas12b蛋白处理的细胞的DNA中要高得多(大约75%)。也观察到,关于DNMT1基因中的插入缺失,与分离自经BvCas12b处理的细胞的DNA(大约0%)相比,分离自经BhCas12b处理的细胞的DNA中要高得多(大约20%)。

[2056] 使用BhCas12b(V4)蛋白中断TRAC基因。经由使用编码BhCas12b(V4)蛋白的mRNA连

同对GRIN2B、DNMT1和TRAC基因中的基因座具有特异性的向导RNA进行电穿孔来转导T细胞。在电穿孔后96小时，使用荧光辅助细胞分选(FACS)分析来评估细胞，其中将细胞针对CD3(TRAC的指标)选通。参考图15，用编码GFP的质粒或者用BhCas12b(V4)和对GRIN2B或DNMT1具有特异性的RNA转导的T细胞中的大约95％是CD3+。那些经转导以表达BhCas12b(V4)和对TRAC基因中的基因座具有特异性的向导RNA的细胞较不可能是CD3+(大约2％到大约50％，取决于所使用的向导RNA)。所测试的十一种TRAC向导RNA中有三种导致大约100％的BhCas12b(V4)介导的插入缺失。

[2057] 实施例6:CAR-P2A-mCherry慢病毒表达表征

[2058] 使用CAR-P2A-mCherry慢病毒将细胞转导以表达嵌合抗原受体(CAR)，并使用荧光辅助细胞分选(FACS)进行针对CAR表达的分析。将细胞脱色，与缀合至R-藻红蛋白(PE)或异硫氰酸荧光素酯(FITC)的BCMA蛋白一起孵育。因为BCMA是CAR的靶标抗原，表达CAR的细胞将会结合染料标记的BCMA。参考图16，对于未染色的细胞，FACS分析仅检测经转导的样品中mCherry的存在，其中一些溢出到PE通道内。BCMA-PE通道显示超出在溢出效应中所见的高度阳性信号，并且这些结果在与BCMA-FITC一起孵育的细胞中被证实。染料标记的BCMA蛋白检测结果表明了几乎与针对mCherry所见相同的CAR表达。参考图17，经由FACS分析在经聚(1,8-辛二醇柠檬酸酯)(POC)慢病毒载体转导的细胞中检测到85％ CAR表达。

[2059] 实施例7:BE4产生有效、持久的基因敲除且具有高产物纯度

[2060] BE4介导多个基因的碱基编辑以生成多敲低细胞。用编码BE4碱基编辑器的mRNA连同靶向B2M、TRAC、PD1或其组合中的靶标特异性位点的sgRNA一起共转染免疫细胞。如测序资料所示，碱基编辑在修饰细胞中有效并且持续长达至少7天(图18)。观察到了高产物纯度，因为C到T转换构成了大量的所观察到的编辑。插入缺失以及C到G和C到A转换构成了在经编辑的基因中的基因改变的不显著的很小一部分。在生成所希望的修饰方面，碱基编辑与spCas9核酸酶一样有效。

[2061] BE4系统引起了有效敲低，如通过流式细胞术所测量，这鉴定了具有降低的表面表达的细胞百分比(图19A)。针对B2M表达选通的细胞展示出细胞表面上B2M蛋白的丧失。如通过流式细胞术所测量，在生成B2M蛋白敲除方面，碱基编辑也与spCas9核酸酶一样有效。

[2062] 实施例8:正交易位检测测定不能检测BE4介导的经三重编辑的T细胞中的重排

[2063] 用编码BE4碱基编辑器的mRNA连同靶向B2M、TRAC和PD1中的靶标特异性位点的sgRNA一起共转染免疫细胞。使用易位检测测定来评定经三重编辑的T细胞，该测定能够检测B2M、TRAC和PD1靶标基因之间的不希望的特异性易位(图20)。注意，这些特异性易位无一在任何经BE4编辑的基因中检出(表22)。相比之下，经Cas9处理的细胞展示出低但可检测水平的易位。因此，使用BE4碱基编辑器对T细胞进行的多重编辑并未生成易位，与使用Cas9核酸酶的多重编辑形成对比。

[2064] 表22

[2065]

类型	虚假(%)	经 BE4 处理的(%)	经 Cas9 处理的(%)
上靶修饰 (B2M/TRAC/PDCD1)	0	89.9/97.9/89.1	53.0/77.2/55.2
B2M-A / TRAC-A	0	**0**	**0.925**
B2M-A / TRAC-B	0	**0**	**0.353**
B2M-A / PDCD1-A	0	**0**	**1.647**
B2M-A / PDCD1-B	0	**0**	**0.508**
B2M-B / TRAC-A*	0	**0**	**0.505**

[2066] $LLoD_{BE4}$=0.1%

[2067] *如果易位包括在B2M基因座处的局部重排,则在该实验中仅测量B2M-B

[2068] 实施例9:多重碱基编辑不显著破坏细胞表达

[2069] 使用BE4和spCas9两种sgRNA执行了跨B2M、TRAC和PD1基因的广泛向导筛选。基于单定位点测试,选择高边际效率和扩张的向导。最终细胞产率在使用BE4和spCas9进行的1、2和3重编辑之间进行比较并归一化至仅电穿孔对照。经BE4编辑的具有所希望编辑的细胞当作出高达3重编辑时展示高产率(图21)。相比之下,经spCas9编辑细胞当作出递增数量的多重编辑时显示降低的产率。因此,经多重碱基编辑的细胞即使当作出高达3重编辑时仍未知高细胞扩张。因此,BE4生成经多重编辑的T细胞,与经spCas9处理的样品相比,该细胞没有可检测的基因组重排,同时也维持高细胞扩张。

[2070] 实施例10:BE4以与spCas9类似的上靶编辑效率和细胞表型生成经三重编辑的T细胞

[2071] 用编码BE4碱基编辑器的mRNA连同靶向B2M、TRAC和PD1中的靶标特异性位点的sgRNA一起共转染T细胞。如测序资料所示,碱基编辑在全部三个位点处修饰细胞(图22)。通过碱基编辑进行的基因修饰类似于使用spCas9核酸酶进行的基因修饰。流式细胞术也显示B2M和CD3的表面表达降低(图23,上图)。与仅电穿孔对照细胞相比,经BE4和Cas9多重编辑的细胞展示细胞表面上的B2M和CD3蛋白显著减少(>95% $CD3^-/B2M^-$)。尽管PD1染色效率不高,但与仅电穿孔对照细胞相比,仍在经BE4和Cas9多重编辑的细胞中的观察到了PD1的显著减少(~90%)(图23,下图)。

[2072] 实施例11:BE4编辑不改变CAR表达或抗原依赖性细胞杀死

[2073] 用编码BE4碱基编辑器的mRNA连同靶向B2M、TRAC和PD1中的靶标特异性位点的sgRNA一起共转染T细胞。通过整合编码抗BCMACAR的慢病毒载体,引入靶向BCMA的嵌合抗原受体(CAR)。通过流式细胞术在经BE4和Cas9编辑的细胞中观察到了CAR表达(图24),与未接受慢病毒载体的未经处理的细胞比较。通过对表达BCMA的细胞进行核染色并检测核染色的丧失(指示细胞死亡),评定CAR-T细胞的细胞杀死。在经载体转导并表达CAR的细胞中,包括经BE4和Cas9编辑的T细胞中,观察到了抗原依赖性细胞杀死(图25)。相比之下,未使用载体

转导的未经处理的细胞不展示细胞杀死活性。因此,与它们的仅核酸酶对应物相比,经BE4生成的CAR-T细胞表明可比较的基因中断、细胞表型和抗原依赖性细胞杀死。

[2074] 实施例12:Cas12b和BE4可经配对以在T细胞中进行高效的多重编辑

[2075] 仅使用BE4或者使用BE4和Cas12b生成$CD3^-B2M^-$T细胞。对于仅使用BE4生成的T细胞,用编码BE4碱基编辑器的mRNA连同靶向B2M和TRAC中的靶标特异性位点的sgRNA一起共转染T细胞。对于使用BE4和Cas12b生成的T细胞,用编码BE4碱基编辑器的mRNA以及靶向B2M中特异性位点的sgRNA、编码BhCas12b(V4)的mRNA、和靶向TRAC基因外显子4的Cas12b sgRNA(其用来中断TRAC基因)共转染T细胞。使用荧光辅助细胞分选(FACS)分析来评估所得的T细胞,以检测B2M和CD3细胞表面表达。仅使用BE4的敲除展示出与使用BE4和Cas12b的那些类似的情况。特别地,高该百分比的T细胞是$CD3^-B2M^-$:86%(仅BE4)和88%(BE4+Cas12b),而其他可能表型$CD3^-B2M^+$;$CD3^+B2M^+$T细胞;和$CD3^+B2M^-$在细胞群中占比较少(图26)。相比之下,仅电穿孔对照显示群体具有高百分比(97.8%)的$CD3^+B2M^+$细胞和非常低百分比的$CD3^-B2M^-$细胞。

[2076] 使用Cas12b生成$CD3^-CAR^+$T细胞。用编码BhCas12b(V4)的mRNA,靶向TRAC基因外显子4的Cas12b sgRNA和编码抗BCMACAR的双链DNA(dsDNA)供体模板共转染T细胞。使用荧光辅助细胞分选(FACS)分析来评估T细胞,以检测CD3和BCMA细胞表面表达。当在sgRNA存在下将递增量的Cas12b引入细胞内时,CD3表达降低,如通过细胞群向$CD3^-$象限的转移所见(图27)。当在相同条件下将递增量的供体模板引入细胞内时,观察到细胞群向$CD3^-CAR^+$象限的转移。

[2077] 因此,Cas12b可以与BE4配对以生成经多重编辑的T细胞,将由多个双链断裂造成的基因组重排降至最低。

[2078] 实施例13:在动物模型中的体内功效

[2079] 为了确定本文所述化合物在体内的抗致癌活性,将化合物例如单独给予(IP、IV、PO、通过吸入或鼻途径)或与次优剂量的相关化疗剂(例如,环磷酰胺、阿霉素、依托泊苷)组合。在一个示例中,在NOD-SCID小鼠已经经历全身辐射后3小时(hrs),将$5x\ 10^6$个稳定地表达荧光素酶的RS4;11细胞(从具有急性成淋巴细胞性白血病的受试者的骨髓建立)通过尾静脉注射到该小鼠体内。如果不做处理,在这一模型中,这种形式的白血病在3周内致命。例如,通过向小鼠注射D-荧光素(60mg/kg)并对经麻醉的动物成像(例如,Xenogen体外成像系统,Caliper Life Sciences,Hopkinton,MA),容易地监测白血病。通过活体成像软件(Caliper Life Sciences,Hopkinton,MA)对光子通量(光子/秒)积分,将全身生物发光定量。将本文所述化合物单独地或与次优剂量的相关化疗剂例如通过尾静脉或IP途径以0.1mg/kg至50mg/kg的剂量给药至白血病小时(注射后10天/实验的第1天,在14-16的生物发光范围内),持续7至21天。任选地,在整个实验过程中每两天对小鼠成像,并在实验持续期间每天监测存活。在实验结束时,可选择对死去的小鼠进行尸检。另一动物模型是将DoHH2(一种源自人类滤泡性白血病的细胞系,其稳定地表达荧光素酶的)移植到NOD-SCID小鼠体内。这些体内测试任选地生成初步药代动力学、药效学和毒理学数据。

[2080] 实施例14:临床试验

[2081] 为了确定本文所述化合物用于治疗人类的适用性,执行了临床试验。例如,可以选择被诊断为患有液态癌症并且需要治疗的受试者并分为治疗组和一个或多个对照组,其中

向治疗组给药本文所述化合物,同时对照组接受安慰剂或已知的抗癌药物。因此,可通过对受试者组执行关于多种因素诸如存活和生命品质的比较,评定本文所述化合物的治疗安全性和功效。本实施例中,与使用安慰剂治疗的受试者对照组相比,用本文所述化合物治疗的受试者组可显示改善的长期存活。

[2082] 研究设计

[2083] 本研究包括一项剂量递增、2臂研究,经设计为评定在表达WT p53蛋白的晚期液态癌症(例如,白血病、骨髓瘤或液态淋巴瘤)受试者中,采用28天或21天周期的2种不同给药方案,通过IV输液给药的化合物的安全性、耐受性、PK、PD和抗液态癌症细胞效果。例如,本文所述化合物可用于患有复发性/难治性急性髓性白血病(AML)和/或急性淋巴性白血病(ALL)的受试者。受试者接受本文所述化合物,对于28天周期,每周一次,持续连续的三周;或者对于21天周期,每周两次,持续连续的两周。很多患有液态淋巴瘤的受试者在外周血中呈现循环肿瘤细胞(CTC),这些细胞可以使用流式细胞术检测并分析。因此,在这些细胞中检测研究药物-特异性靶标衔接是可能的。

[2084] 研究由剂量递增期(DEP)和扩展期(EXP)组成。DEP是一种“3+3”剂量递增设计,以建立本文所述化合物的MTD或OBD。EXP纳入至多2个不同的在MTD或OBD具有特异性液态癌症的受试者组,以进一步探索该剂量水平的临床安全性和潜在功效水平。对进行EXP的受试者的选择是基于DEP结果以及来自另外的非临床药理学研究数据确定的。后者包括对多个液态癌症细胞系(例如,白血病、液体淋巴瘤、骨髓瘤等)的研究,这些细胞系有助于在不同类型的液态癌症中比较细胞系对本文所述化合物的敏感性。

[2085] 在研究的剂量递增和剂量扩展阶段,受试者的治疗将继续进行,直到记录疾病进展、不可接受的毒性,或受试者或医生决定停止治疗。

[2086] 在MTD或OBD建立后,另外的受试者可以被纳入最多两个单独的扩展队列(每个扩展队列大约30名受试者),以获得在该剂量水平以及在特定受试者或液态癌症细胞类型中的进一步经验。对受试者或液态癌症细胞类型的选择部分地基于在该研究的剂量递增期作出的观察结果来确定。

[2087] 基于负面事件(AE)的发生率、严重程度、持续时间、因果性、严重性和类型以及受试者身体检查、生命体征和临床实验室结果的改变来评定安全性。研究人员使用NCI CTCAE 4.0版来评估AE的严重程度。

[2088] 因为这一研究的主要目标是基于安全性和PK,统计分析本质上是描述性的并说明全部研究剂量和全部观察到的应答,包括根据IWG(2014)标准获得完全缓解(CR)或部分缓解(PR)或保持病情稳定(SD)的受试者。接受至少一个剂量的本文所述化合物的受试者构成了安全性群体并且包括在全部安全性分析中。完成本文所述化合物的至少一个周期并经历治疗后目标疾病评估的受试者构成了功效可评定受试者群体。

[2089] 实施例15:鉴定并沉默基因的示例性方法

[2090] 在本实施例中,提供用于对CAR-T的靶标进行基因沉默的方法。图28A提供待靶向的基因的示意图以及关于该基因的注意事项。靶标包括,例如,CD4、CD5、CD7、CS-1、TRAC、CAR、FAS、LAG-3、PD-1、β2M、CD52和TCR中的一者或多者。

[2091] 有两种使用碱基编辑器进行沉默的策略。第一种策略访问靶标蛋白并随后改变终止密码子的谷氨酰胺、精氨酸或色氨酸中的一者或多者,以沉默蛋白表达(参见,图28B的项

目1)。第二种策略包括用CBE中断剪接(参见,图28B的项目2)以沉默蛋白表达。

[2092] 实施例16:使用多个CAR-T群来解决AML或T-ALL中的疾病克隆性

[2093] 据推测,白血病由于肿瘤微环境中的选择压力,通过克隆扩张和选择的反复过程而进化,在某些治疗中会产生耐药性或复发。作为示例,AML癌细胞的不同克隆体可以表现出各种抗原,诸如CD33+CD123-;CD33+CD123+或CD33-CD123+。靶向CD33+癌细胞的治疗剂不会访问缺乏这一特定抗原诸如CD33-CD123+的细胞。因此,预期本文所述的多个CAR-T细胞群将会解决疾病克隆性。

[2094] 在示例性方法中,如图29中所述生成CD33和CD123 CAR-T细胞。将CD33和CD123 CAR-T细胞给药至具有CD33+CD123+AML标志物的AML患者。CD33和CD123 CAR-T细胞消除了患者体内的表现为CD33+CD123-、CD33+CD123+和CD33-CD123+的AML癌细胞。

[2095] 在另一示例性方法中,通过本文所述的方法,如图30中所述,生成CD3和CD7 CAR-T细胞。将CD3和CD7 CAR-T细胞给药至具有CD3+CD7+T-ALL标志物的T-ALL患者。CD3和CD7 CAR-T细胞消除了表现为CD3+CD7-、CD3+CD7+和CD3-CD7+的T-ALL癌细胞。

[2096] 实施例17:修饰多个CAR-T群以消除同类相杀

[2097] 传统T细胞和CAR-T疗法的主要限制是,激活的细胞可能在细胞制造、储存期间或当输注到患者体内时经历自杀或同类相杀。为了避免同类相杀,本文所述的CAR-T细胞经修饰以在编码该细胞的CAR对其具有特异性的抗原的多核苷酸中包含突变。作为示例,CD7 CAR-T细胞经修饰以在CD7多核苷酸中包含突变,从而阻止同类相杀。作为领难以示例,CD3 CAR-T细胞经修饰以在CD3多核苷酸中包含突变,从而阻止同类相杀。在多个CAR-T群中,执行另外的修饰以阻止交叉同类相杀(即,在两个CAR-T群中的CD3和CD7两种多核苷酸中的突变)。另外的修饰包括对基因的突变,该基因可能牵涉到免疫抑制、移植物抗宿主、宿主抗移植物等中。

[2098] 在示例性方法中,如图29中所述生成用于治疗AML的CD33和CD123 CAR-T细胞。CD33和CD123 CAR-T细胞两者均经修饰以在CD33中具有突变,使得细胞不靶向它们自身。在PD-1、CD52和TRAC中的突变分别降低或阻止免疫抑制、宿主抗移植物和移植物抗宿主。

[2099] 在示例性方法中,如图30中所述生成用于治疗T-ALL的CD3和CD7CAR-T细胞。CD3和CD7 CAR-T细胞两者均经修饰以在CD3(TRAC)和CD7中具有突变,使得细胞不靶向它们自身或其他群。在PD-1、CD52和TRAC中的突变分别降低或阻止免疫抑制、宿主抗移植物和移植物抗宿主。

[2100] 实施例18:与核酸酶相比,对于T-ALL CAR细胞的四重同时碱基编辑不影响产率

[2101] 在四重T-ALL编辑情况下的理论产率在测试一下组分的测定中确定。对照是无电穿孔(无EP)、仅EP、CBE变体1、CBE变体2、和Cas9。结果显示在图31A中,并且对应的活力结果显示在图31B中。确定了与核酸酶相比,对于T-ALL CAR细胞的四重同时碱基编辑不影响产率。

[2102] 实施例19:扩大Lonza电穿孔诱导T-ALL CAR-T细胞程序的高效、多重T细胞碱基编辑。

[2103] 将T细胞在激活后培养3天。用10μg用于编辑器(CBE 1、CBE 2或spCas9)的mRNA和5μg用于T-ALL向导(TRAC、CD52、CD7、PD1)的sgRNA将5×10^6个细胞电穿孔。合适的向导提供在下表23中:

[2104] 表23.用于T-ALL CAR-T的向导RNA

[2105]

向导 RNA (gRNA)	向导核酸序列
TRAC gRNA	ususcsGUAUCUGUAAAACCAAGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAU AAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUsususu
CD52 gRNA	ususasCCUGUACCAUAACCAGGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAU AAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUsususu
CD52 gRNA	csuscsUUACCUGUACCAUAACCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAU AAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUsususu
CD7 gRNA	usgscsACCUCUGGGGAGGACCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAU AAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUsususu
CD7 gRNA	cscsusACCUGUCACCAGGACCAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAU AAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUsususu
PD1 gRNA	csascsCUACCUAAGAACCAUCCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAU AAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUsususu

[2106] 2'-O-甲基类似物和3'硫代磷酸酯核苷酸间链接用在前三个5'和3'末端RNA残基处,在每个末端处的化学修饰以后接该核苷酸的“s”指示。

[2107] CBE 1是指大鼠-APOBEC-BE4,且CBE2是指红毛猩猩“BE4”(ppAPOBEC)。无EP和仅EP是对照,其要么没有受到攻击(zap)要么受到攻击但不使用编辑器/向导。

[2108] 理论产率测量如下:每天计数细胞,并使用AO/DAPI染色进行活力测量(活力读数也使用计数-AML数字和EP后数小时活力读数执行)。每天将细胞分割回0.5×10^6细胞/mL(理论上因为分割,事实上并没有扩张全部细胞)。随后进行计算以确定,如果培养物没有被分割,理论产率将会是多少。确定了基因修饰的百分比(图32)。

[2109] 通过NGS获取细胞并确定理论细胞产率(图33)。

[2110] 细胞杀死通过模型肿瘤细胞中的荧光标志物的丧失表明(图34)。

[2111] 所测试的全部CAR均包含结合剂克隆物,要么是重到轻(H2L)取向要么是轻到重(L2H)取向,后面是CD8a铰链、CD8跨膜结构域、和CD28z信号传导结构域。

[2112] 实施例20:示例性CAR-T细胞工序流程和定制

[2113] 获得来自健康供体的冷冻单采样品(apheresis)leukopak并根据制造商的使用说明解冻。根据制造商的使用说明,使用例如CliniMACS Prodigy分离CD4和CD8细胞。然后,根据制造商的使用说明,使用TransAct激活T细胞。根据制造商的使用说明,使用Lonza 4DNucleofactor进行mRNA/RNP碱基编辑器的电穿孔递送,以敲除CD3、CD7、CD52和/或PD1。使用本领域公认的技术执行TCRα/β的耗竭。扩张T细胞,并进行CAR和7CAR的慢病毒转导。用程序冷冻仪(CRF)将CAR-T细胞低温保存。

[2114] 实施例21:用CAR-T细胞治疗T-ALL

[2115] T细胞急性成淋巴细胞性白血病(T-ALL)是一种侵袭性的骨髓恶性肿瘤。约12-15%的ALL病例在儿童中诊出。复发后5年存活率低于25%。治疗标准是使用化疗来诱导第二次缓解,然后进行同种异体造血干细胞移植(alloHSCT)。然而,很多患者是化疗难治性的或具有高肿瘤负荷并且不能诱导作为alloHSCT桥梁的深度缓解。此外,先前进行深度治疗

的患者一般不适合进行自体CAR-T治疗。因此,需要替代性治疗选项来治疗T-ALL。

[2116] 为了检查经编辑的CAR-T细胞是否可在患者体内诱导深度缓解以便架设alloHSCT治疗的桥梁,使用多重编辑癌创建抗同类相杀的CAR-T细胞。如图35中所示,表达CD7嵌合抗原受体的T细胞经编辑以减少或消除CD7、TRAC、CD52和PD-1的表达,用于靶向$CD3^{-}CD7^{+}$或$CD3^{+}CD7^{+}$ T-ALL肿瘤细胞。如图36中所示,在第-7天,使用Cy/Flu/Campath在20至30名T-ALL患者中执行淋巴细胞删除。在第0天,向患者输注CD7 CAR-T细胞。然后,在第65天,使用Cy/Flu/TBI/ATG对患者进行预调理。在第70天,患者接受alloHSCT治疗。预期使用CD7 CAR-T细胞进行的治疗诱导T细胞发育不全并允许T-ALL患者在没有桥梁治疗的情况下用alloHSCT治疗。

[2117] 实施例22:产生CD7 CAR-T细胞

[2118] 为了产生如图37A中所示的TALL017 CD7 CAR-T细胞,接收了来自健康供体的冷冻单采样品。然后用Plasmatherm(Barkey GmbH&Co.KG)解冻单采样品。使用CliniMACS Prodigy(Miltenyi Biotec)从解冻的单采样品分离CD4和CD8细胞。在第2天,使用Lonza SD Nucleofector通过电穿孔将mRNA碱基编辑器和向导RNA递送至T细胞内。碱基编辑后,经由慢病毒转导(7CAR8)将CD7嵌合抗原受体(CAR)递送至T细胞内。然后在第2-8天,将T细胞在培养物中扩张。扩张后,随后将收获的CAR-T细胞(>1B)低温保存。当解冻时,TALL017过程产生异源细胞群。

[2119] 在解冻后24小时,将未经转导的(UTD)与经转导的TALL017细胞进行比较。通过将细胞针对CD2+和CD4+细胞进行选通,使用流式细胞术验证了CAR激活(图37C)。如图37C中所述,由抗CD7 CAR诱导的中间过程同类相杀导致7CAR T在TALL017冷冻保存时仍被激活。在冷冻保存时中间激活的7CAR-T解冻极差。

[2120] 为了产生如图37B中所示的TALL038 CD7 CAR-T细胞,接收了来自健康供体的冷冻单采样品。然后用Plasmatherm(Barkey GmbH&Co.KG)解冻单采样品。使用CliniMACS Prodigy(Miltenyi Biotec)从解冻的单采样品分离CD4和CD8细胞。在第2天,使用Lonza SD Nucleofector通过电穿孔将mRNA碱基编辑器和向导RNA递送至T细胞内。碱基编辑后,经由慢病毒转导(7CAR8)将CD7嵌合抗原受体(CAR)递送至T细胞内。然后在第2-10天,将T细胞在培养物中扩张。扩张后,将CRα/β+细胞耗竭。然后在第10-12天,将T细胞在培养物中进一步扩张。随后在第12天将收获的CAR-T细胞(>2B)低温保存。当解冻时,TALL038过程产生同源单分散细胞群。

[2121] 实施例23:TALL017和TALL038 CD7 CAR-T细胞的评定

[2122] 使用下一代测序(NGS)来比较在TALL017和TALL038 CD7 CAR-T细胞中的编辑效率。如图38中所示,在两种类型的CD7 CAR-T细胞中,CD52、CD7、TRAC和PD-1的编辑均在80-99%效率范围内。使用未转导的(UTD)细胞作为对照。

[2123] 在蛋白融化后24小时,检查TALL017和TALL038 CD7 CAR-T细胞中的TCRα/β、CD52和CD7的表达(图39A和39B)。TALL038细胞表明了<0.4%的TCRα/β+细胞。对于TALL017和TALL038 CD7 CAR-T两者,蛋白质表达数据均与冷冻前指标和融化后NGS数据匹配。使用未转导的(UTD)细胞作为对照。

[2124] 为了评定CD7 CAR-T细胞群体中的细胞身份,使用了如图40中所示的选通方案。细胞按以下秩序选通:FSC/SSC,单细胞,活/死,CD45(全部白血球),CD19-(B细胞排除),

CD56-/CD14-(NK/单核细胞排除),CD2+(T细胞),和CD2+的CD4:CD8。如图41中所示,TALL017和TALL038 CD7 CAR-T细胞群体的身份是CD2+/-和CD56+/-细胞的混合物。没有可检测的单核细胞或B细胞。

[2125] 最后,评定了CAR表达。如图42中所示,TALL017和TALL038 CD7CAR-T细胞群体两者均维持CAR+细胞表达,直至融化后48小时。使用未转导的(UTD)细胞作为对照。

[2126] 也在融化后24小时和48小时评定了TALL017和TALL038 CD7CAR-T细胞中的CD25和CD69表达。如图43中所示,与TALL017细胞相比,TALL038细胞中的融化后的CD25表达较低。CD25是中期T细胞激活的标志物。不同于大多数CAR,抗CD7 CAR诱导中间培养物同类相杀,这进一步激活7CART。冷冻激活状态的TALL017细胞对细胞健康有害。与TALL017相比,TALL038细胞具有剧烈降低的CD25表达,这可能暗示了更高的抗冷冻性。

[2127] 实施例24:TALL038 CD7 CAR-T细胞的体外表征

[2128] 为了在体外表征TALL038 CD7 CAR-T细胞,开发了基于珠的效价法(图44)。在第0天,将CD7与Dynabeads偶联,并将CD7 CAR-T细胞和未转导的(UTD)细胞融化。在第1天,将CD7珠纯化并与CD7CAR-T细胞共培养,铺板,并在37℃孵育。在第2天,收集上清液并使用酶联凝集素测定(ELLA)获得读数。

[2129] 使用该测定,评定了TALL038 CD7 CAR-T细胞的IFNγ、TNFα、IL-10和IL-2释放。对CD7抗原作出应答,TALL038 CD7 CAR-T细胞释放了IFNγ、TNFα和IL-2(图45A和45B)。使用未转导的(UTD)细胞作为对照。

[2130] 为了进一步在体外表征TALL038 CD7 CAR-T细胞,开发了共培养方法来测量靶标肿瘤细胞的CAR引导的T细胞杀死。图46提供了该测定的概述。首先,接种肿瘤细胞。然后将CAR-T细胞添加到接种的肿瘤细胞中。然后使用IncuCyte®扫描系统分析细胞。最后,使用IncuCyte®软件获得读数。

[2131] 将TALL038 CD7 CAR-T细胞添加到接种的CCRF-CEM GFP+肿瘤细胞中。以4:1CAR:CCRF、2:1CAR:CCRF或1:1CAR:CCRF的E:T比率测试表24中的T细胞批次:

[2132] 表24.T细胞批次

[2133]

批次	%CAR+
TALL017UTD	0%CAR+
TALL017 7CAR8	~50%CAR+
TALL038UTD	0%CAR+
TALL038 7CAR8	~50%CAR+

[2134] 如图47A和47B中所示,与TALL017 CAR T相比,TALL038 CD7 CAR-T细胞在再次挑战时表现出增加的CCRF。

[2135] 实施例25:TALL038 CD7 CAR-T细胞的体内表征

[2136] 先前研究已经表明,CCRF模型显示对CD7引导的CAR-T疗法产生应答(参见,例如,Gomes-Silva D.,等人,CD7-edited T cells expressing a CD7-specific CAR for the therapy of T-cell malignancies.Blood.2017;130:285-296)。

[2137] 为了在体内表征TALL038 CD7 CAR-T细胞,开发了一种研究来评估TALL038药物产品在T-ALL的CCRF模型中对于体内生长的功效。表25总结了研究策略。

[2138] 表25.CCRF模型体内研究

[2139]

组	肿瘤	每只小鼠的CCRF-GFP+细胞	每组的小鼠数	治疗	每只小鼠的（T细胞）细胞*	途径	组织收集Tissue Collection
1	CCRF GFP+	1e5	10	DPBS	无	IV	血液
2	CCRF GFP+	1e5	10	UTD CAR	1.00E+07	IV	血液
3	CCRF GFP+	1e5	10	7CAR8	1.00E+07	IV	血液
4	CCRF GFP+	1e5	10	7CAR8	3.00E+06	IV	血液
5	CCRF GFP+	1e5	10	7CAR8	1.00E+06	IV	血液
6	CCRF GFP+	1e5	10	在归一化期间作为异常值排除			

[2140] 十只6至7周龄雌性NSG小鼠(Jackson Labs:货号005557)经静脉内(IV)移植CCRF-GFP+细胞(1e5)。在第11天,小鼠经IV移植250μL总TALL038 1e7细胞/小鼠。在注射后,每周3次进行IVIS成像测量,持续2至3周,以监测肿瘤负荷。每周1次采集全血,持续2至3周。在移植后进行每周3次的体重称重和临床观察。

[2141] 图48显示在CCRF移植后第10天/CD7 CAR-T治疗第-1天的生物发光辐射数据(均值,SEM)。在该时间点,小鼠不呈现CD7 CAR-T治疗所需的可测量肿瘤负荷。图49A和49B显示在CCRF移植后第19天/7CAR8治疗后第8天的生物发光辐射数据(均值,SEM)。在该时间点,观察到应答CD7 CAR-T治疗的剂量依赖性肿瘤生长抑制。小鼠应答CD7CAR-T治疗而不损失体重。图50A和50B显示在CCRF移植后第38天/CD7CAR-T治疗后第27天的生物发光辐射数据(均值,SEM)。在每组中第一只小鼠到达仁慈终点(通量>1e10)的日期后,不再绘制均值和标准误差。在该时间点,在复发之前,在每组中均观察到持续的剂量依赖性抑制。通过使用图50B中的对数标度,将CD7 CAR-T组的1e7复发可视化。单个小鼠在CCRF移植后第38天/在CD7 CAR-T治疗后第27天的生物发光辐射数据显示在图51A、51B、52A和52B中。对于上述实验,UTD和媒介物用作对照。

[2142] 实施例26:靶向CD5的CAR-T用于T-ALL治疗和淋巴瘤单一疗法

[2143] 先前研究已经认识到了对于通过靶向多种癌症抗原来创建改进的CAR-T产品的需求(参见,例如,Xinj ie Xu,等人,Mechanisms of Relapse After CD19 CAR T-Cell Therapy for Acute Lymphoblastic Leukemia and Its Prevention and Treatment Strategies.Front Immunol.2019;10:2664,Nov.12,2019;E.Mejstríková,等人,CD19-negative relapse of pediatric B-cell precursor acute lymphoblastic leukemia following blinatumomab treatment.Blood Cancer Journal,7:659(2017);Hanren Dai,等人,Bispecific CAR-T cells targeting both CD19 and CD22 for therapy of

adults with relapsed or refractory B cell acute lymphoblastic leukemia.J.Hematology&Oncology.13:30(2020);Robbie G.Majzner and Crystal L.Mackall,Tumor Antigen Escape from CAR T-cell Therapy.Cancer Discov.2018Oct;8(10):1219-1226)。

[2144] 已经在T和NK细胞恶性肿瘤中鉴别出了CD5表达,这些恶性肿瘤包括但不限于T-ALL、MF/SS、原发性皮肤T细胞淋巴瘤(NOS)、T细胞大颗粒淋巴细胞性白血病、血管免疫母细胞性T/NK细胞淋巴瘤、肝脾T细胞淋巴瘤、原发性皮肤CD30+LPD、成人T细胞白血病/淋巴瘤和T细胞前淋巴细胞性白血病。将CD5与CD7组合将解锁液态癌症疗法用于某些癌症诸如T-ALL、MF/SS、原发性皮肤T细胞淋巴瘤(NOS)、血管免疫母细胞性T/NK细胞淋巴瘤和成人T细胞白血病/淋巴瘤的更大潜力。

[2145] 为了减少或消除T细胞内的CD5表达,生成了如下表26中提供的42种胞苷碱基编辑器单向导RNA(sgRNA)(Agilent)。

[2146] 表26.CD5向导RNA

[2147]

向导名	向导分组	向导序列
RCM sgRNA 102	Ex1 SD(Pos 6)	ACUCACCCAGCAUCCCCAGC
RCM sgRNA 103	Ex2 SA(Pos 6)	AGCGACUGCAGAAAGAAGAG
RCM sgRNA 104	Ex2 STOP(pos 5/6)	CAUACCAGCUGAGCCGUCCG
RCM sgRNA 105	Ex3 SA(Pos 8)	UGGAAAUCUGGGGGUCAGAA
RCM sgRNA 106	Ex3 SD(Pos 9)	GUUACCCACCUAAGCAGGUC
RCM sgRNA 107	Ex3 STOP(pos 5)	CUGCCAGCGGCUGAACUGUG
RCM sgRNA 108	Ex3 STOP(Pos 5/6)	CCUCCCACUGCUUGGAGCUC
RCM sgRNA 109	Ex3 STOP(pos 6)	UCUGCCAGCGGCUGAACUGU
RCM sgRNA 110	Ex3 STOP(pos 7)	GUCUGCCAGCGGCUGAACUG
RCM sgRNA 111	Ex3 STOP(Pos 8)	GAAGUGCCAGGGCCAGCUGG
RCM sgRNA 112	Ex3 STOP(Pos 8/9)	CCAUGUGCCAUCCGUCCUUG
RCM sgRNA 113	Ex3 STOP(Pos 9)	UUUGCAGCCAGAGCUGGGGC
RCM sgRNA 114	Ex4 SA(Pos 5)	GGUUCUGCAAUGAGACACUC
RCM sgRNA 115	Ex5 SA(Pos 5)	GAGCUAGGAGAGGAGAGAGC
RCM sgRNA 116	Ex5 SD(Pos 8)	UCACUUACCUGAGCAAAGGA
RCM sgRNA 117	Ex5 SD(Pos 9)	CUCACUUACCUGAGCAAAGG
RCM sgRNA 118	Ex5 STOP(Pos 4)	CUGCAGCUGGUGGCACAGUC
RCM sgRNA 119	Ex5 STOP(Pos 5)	CGGCCAGCACUGUGCCGGCG
RCM sgRNA 120	Ex5 STOP(Pos 6/7)	AUCUUCCAUUGGAUUGGCAA
RCM sgRNA 121	Ex5 STOP(Pos 7/8)	GAUCUUCCAUUGGAUUGGCA
RCM sgRNA 122	Ex6 SA(Pos 4)	AACCUGAGAGGGGAAGCAAU
RCM sgRNA 123	Ex6 SA(Pos 5)	AAACCUGAGAGGGGAAGCAA
RCM sgRNA 124	Ex6 STOP(Pos 4)	GUGCAGAGCCGUCUGGUGGG
RCM sgRNA 125	Ex6 STOP(Pos 4/5)	CUCCCACCGCAGCGAGCUCC
RCM sgRNA 126	Ex6 STOP(Pos 4/5)	CUCCCAAAGUUCGUGGCACU

RCM sgRNA 127	Ex6 STOP(Pos 5)	AGGUGCAGAGCCGUCUGGUG
RCM sgRNA 128	Ex6 STOP(Pos 5)	GGUGCAGAGCCGUCUGGUGG
RCM sgRNA 129	Ex6 STOP(Pos 5/6)	UCUCCCAAAGUUCGUGGCAC
RCM sgRNA 130	Ex6 STOP(Pos 6)	UGUCCCAGUGCCACGAACUU
RCM sgRNA 131	Ex6 STOP(Pos 7)	AAGGUGCAGAGCCGUCUGGU
RCM sgRNA 132	Ex6 STOP(Pos 7)	UCCUAUCGAGUGCUGGACGC
RCM sgRNA 133	Ex6 STOP(Pos 8)	CAAGGUGCAGAGCCGUCUGG
RCM sgRNA 134	Ex6 STOP(Pos 8)	GGUGCGCCAGGGGGCUCAGU
RCM sgRNA 135	Ex6 STOP(Pos 8/9)	GGGCUGCCCACUGAGCCCCC
RCM sgRNA 136	Ex6 STOP(Pos 9)	AGGUGCGCCAGGGGGCUCAG
RCM sgRNA 137	Ex7 STOP(Pos 4)	GGCCAGGAUCCAAACCCCGC
RCM sgRNA 138	Ex8 STOP(Pos 4)	CGCCAGUGGAUUGGCCCAAC
RCM sgRNA 139	Ex8 STOP(Pos 7)	AAGAAGCAGCGCCAGUGGAU
RCM sgRNA 140	Ex9 SA(Pos 8)	AAAGACACUGGGCAGAUGGU
RCM sgRNA 141	Ex9 SD(Pos 6)	GCUUACCUGGAUAAGCUGAC
RCM sgRNA 142	Ex10 SA(Pos 9)	UUCCAGAGCUGGGGAAAGAA
RCM sgRNA 143	Ex10 STOP(Pos 9)	AUGGGGCUCAGAGGCUGUAA

[2148] 通过在电穿孔72小时后,使用流式细胞术、下一代测序(NGS)、RNA和西方印迹分析,筛选针对最少n＝3个供体和最少n＝5个抗体克隆物的向导,验证CD5编辑(图53)。使用仅电穿孔(EP)作为阴性对照。将胞苷碱基编辑器与CD5向导RNA结合使用有效地减少或消除T细胞中的CD5表达。

[2149] 选择了两个候选向导sgRNA 103(剪接中断)和sgRNA 104(STOP密码子),用于经由NGS进一步测试当与BE4结合使用时的碱基编辑效率。sgRNA 103和sgRNA 104两者均表现出约85％的编辑效率。下表27提供使用sgRNA 103和sgRNA 104执行的计算机模拟OT分析的结果。

[2150] 表27.计算机模拟OT分析

[2151]

		凸起_类型	无	无	无	无	无	无	DNA	DNA	RNA	RNA
		凸起_长度	0	0	0	0	0	0	1	1	1	1
		num_mm	0	2	3	4	5	6	1	2	1	2
靶标_位点_名称	pam	ot_类别										
sgRCM103	NAA	外显子	0	1	5	46	323	1902	0	3	0	2
sgRCM103	NAA	无	0	3	52	647	5918	40065	0	8	0	80
sgRCM103	NAA	TS 外显子	0	0	0	3	27	94	0	0	0	0
sgRCM103	NAG	外显子	0	2	4	41	340	1883	0	0	0	2
sgRCM103	NAG	无	0	2	34	444	3633	24615	0	5	1	37
sgRCM103	NAG	TS 外显子	0	0	1	2	19	104	0	0	0	2
sgRCM103	NGA	外显子	0	0	0	25	200	1214	0	2	0	4
sgRCM103	NGA	无	0	0	23	381	3357	22500	2	5	0	29
sgRCM103	NGA	TS 外显子	0	0	0	0	18	60	0	1	0	0
sgRCM103	NGG	外显子	1	1	2	29	202	1202	0	1	0	2
sgRCM103	NGG	无	0	0	29	294	2710	18612	1	4	0	28
sgRCM103	NGG	TS 外显子	0	0	0	0	10	65	0	0	0	0
sgRCM104	NAA	外显子	0	0	0	4	26	256	0	0	0	1
sgRCM104	NAA	无	0	0	2	26	386	3483	0	0	0	3
sgRCM104	NAA	TS 外显子	0	0	0	0	1	7	0	0	0	0
sgRCM104	NAG	外显子	0	0	0	5	44	502	0	0	0	0
sgRCM104	NAG	无	0	0	2	36	461	5459	0	0	0	4
sgRCM104	NAG	TS 外显子	0	0	0	0	0	27	0	0	0	0
sgRCM104	NGA	外显子	0	0	0	2	38	364	0	0	0	1
sgRCM104	NGA	无	0	0	0	51	1101	9541	0	0	1	0
sgRCM104	NGA	TS 外显子	0	0	0	0	1	27	0	0	0	0
sgRCM104	NGG	外显子	1	0	0	2	70	506	0	0	0	0
sgRCM104	NGG	无	0	0	1	28	461	4267	0	2	0	1
sgRCM104	NGG	TS 外显子	0	0	0	0	2	34	0	0	0	0

[2152] 实施例27:产生靶向CD5的CAR-T

[2153] 为了产生靶向CD5的CAR-T细胞,将慢病毒载体(LVV)工程改造为编码CD5 CAR(图55)。用该CD5 CAR LVV转导患者T细胞。对CAR-T细胞执行7至14天的体外扩张,再回输给患者。图55显示了示例性CD5 CAR构建体设计。示例性CD5 CAR构建体设计包括抗CD5 scFv、铰链/间隔序列、跨膜结构域、信号2共刺激结构域和CD3ζ共刺激结构域。铰链/间隔序列区可包括CH3、CD8α或CD28。跨膜结构域可包括CD8α或CD28。共刺激结构域可包括41bb或CD28。

[2154] 生成了三十种LVV。下表28提供各种LVV CD5 CAR构建体。

[2155] 表28.CD5 CAR LVV构建体

[2156]

ID	序列
pCAR_BTx063	MND-5CAR-CH3-CD28TM-CD28-CD3Z-WPRE密码子opt 1
pCAR_BTx064	MND-5CAR-CH3-CD28TM-CD28-CD3Z-WPRE密码子opt 2
pCAR_BTx065	MND-5CAR-CH3-CD28TM-CD28-CD3Z-WPRE密码子opt 3
pCAR_BTx066	MND-5CAR-CD8aH-CD28TM-CD28-CD3Z-WPRE密码子opt 1
pCAR_BTx067	MND-5CAR-CD8aH-CD28TM-CD28-CD3Z-WPRE密码子opt 2
pCAR_BTx068	MND-5CAR-CD8aH-CD28TM-CD28-CD3Z-WPRE密码子opt 3

pCAR_BTx069	MND-5CAR-CD28H-CD28TM-CD28-CD3Z-WPRE密码子opt 1
pCAR_BTx070	MND-5CAR-CD28H-CD28TM-CD28-CD3Z-WPRE密码子opt 2
pCAR_BTx071	MND-5CAR-CD28H-CD28TM-CD28-CD3Z-WPRE密码子opt 3
pCAR_BTx072	MND-5CAR-CH3-CD8aTM-41BB-CD3Z-WPRE密码子opt 1
pCAR_BTx073	MND-5CAR-CH3-CD8aTM-41BB-CD3Z-WPRE密码子opt 2
pCAR_BTx074	MND-5CAR-CH3-CD8aTM-41BB-CD3Z-WPRE密码子opt 3
pCAR_BTx075	MND-5CAR-CD8aH-CD8aTM-41BB-CD3Z-WPRE密码子opt 1
pCAR_BTx076	MND-5CAR-CD8aH-CD8aTM-41BB-CD3Z-WPRE密码子opt 2
pCAR_BTx077	MND-5CAR-CD8aH-CD8aTM-41BB-CD3Z-WPRE密码子opt 3
pCAR_BTx078	PGK-5CAR-CH3-CD28TM-CD28-CD3Z-WPRE密码子opt 1
pCAR_BTx079	PGK-5CAR-CH3-CD28TM-CD28-CD3Z-WPRE密码子opt 2
pCAR_BTx080	PGK-5CAR-CH3-CD28TM-CD28-CD3Z-WPRE密码子opt 3
pCAR_BTx081	PGK-5CAR-CD8aH-CD28TM-CD28-CD3Z-WPRE密码子opt 1
pCAR_BTx082	PGK-5CAR-CD8aH-CD28TM-CD28-CD3Z-WPRE密码子opt 2
pCAR_BTx083	PGK-5CAR-CD8aH-CD28TM-CD28-CD3Z-WPRE密码子opt 3
pCAR_BTx084	PGK-5CAR-CD28H-CD28TM-CD28-CD3Z-WPRE密码子opt 1
pCAR_BTx085	PGK-5CAR-CD28H-CD28TM-CD28-CD3Z-WPRE密码子opt 2
pCAR_BTx086	PGK-5CAR-CD28H-CD28TM-CD28-CD3Z-WPRE密码子opt 3
pCAR_BTx087	PGK-5CAR-CH3-CD8aTM-41BB-CD3Z-WPRE密码子opt 1
pCAR_BTx088	PGK-5CAR-CH3-CD8aTM-41BB-CD3Z-WPRE密码子opt 2
pCAR_BTx089	PGK-5CAR-CH3-CD8aTM-41BB-CD3Z-WPRE密码子opt 3
pCAR_BTx090	PGK-5CAR-CD8aH-CD8aTM-41BB-CD3Z-WPRE密码子opt 1
pCAR_BTx091	PGK-5CAR-CD8aH-CD8aTM-41BB-CD3Z-WPRE密码子opt 2
pCAR_BTx092	PGK-5CAR-CD28H-CD28TM-CD28-CD3Z-WPRE密码子opt 1

[2157] 使用sgRNA 103和sgRNA 104筛选来自表27的具有MND启动子或PGK启动子的每一种CD5 CAR LVV构建体,以检查在原代人类T细胞中的CAR表达。使用未转导的(UTD)细胞作为对照。如图56中所示,CD5CAR LVV构建体导致具有0%至82%的CAR表达。在10天培养期结束时,二十四种构建体显示可检测的在T细胞表面上的CAR表达。

[2158] 为了确定铰链/间隔序列CAR构建体序列是否影响CAR表达,检查了表29中的含有不同铰链/间隔序列(以加粗字体表示)的CAR构建体与sgRNA 103和sgRNA 104的组合。使用未转导的(UTD)细胞作为对照。如图57中所示,铰链/间隔序列确实影响CAR表达。

[2159] 表29.CD5 CAR构建体

[2160]

ID	序列
UTD	未经转导(无慢病毒)
pCAR_BTx063	MND-5CAR-CH3-CD28TM-CD28-CD3Z-WPRE
pCAR_BTx067	MND-5CAR-CD8aH-CD28TM-CD28-CD3Z-WPRE
pCAR_BTx071	MND-5CAR-CD28H-CD28TM-CD28-CD3Z-WPRE

[2161] 为了进一步在体外表征CD5 LVV CAR-T细胞,执行了活体成像细胞杀死测定。将CD5 LVV CAR-T细胞(LV66、LV67、LV68和LV69)添加到CCRF肿瘤细胞中。如图58A-58D中所

示,经编辑和未经编辑的构建体对于CD5+CCRF细胞都是细胞毒性的。

[2162] 使用sgRNA 103和sgRNA 104在单独的T细胞中或在CCRF细胞中在体外评定CD5 CAR构建体(LV63至LV69)的IFNγ产生(图59A和59B)。在CD5+CCRF-CEM白血病细胞系存在下,全部CD5 CAR构建体均释放IFNγ。

[2163] 实施例28:使用多修饰的免疫效应细胞的组合疗法

[2164] 为了检查使用经多重基因修饰的CAR-T细胞是否可增强对具有T细胞或NK细胞恶性病变诸如液态癌症的患者的治疗,向患者给药一种或多种基于其免疫表型的经修饰的CAR-T细胞。使用流式细胞术和/或序列分析,对患者样品中的免疫细胞进行针对CD5、CD7、CD33和CD123存在或不存在的免疫分型。通过电穿孔将mRNA碱基编辑器和向导RNA递送至免疫细胞内,以编辑CD5、CD7、CD33或CD123与TRAC、LAG-3、FAS、CD52、B2M、CIITA、TRBC1、TRBC2和PD-1中的一者或多者组合,从而减少或消除这些基因在免疫细胞中的表达。在碱基编辑后,经由慢病毒转导将CD5 CAR、CD7 CAR、CD33 CAR或CD123 CAR递送至经修饰的免疫细胞内,以分别创建CD5、CD7、CD33和CD123 CAR-T细胞。

[2165] 经免疫分型为$CD5^+CD7^-$的患者用CD5 CAR-T细胞或其群治疗。经免疫分型为$CD5^-CD7^+$的患者用CD7 CAR-T细胞或其群治疗。经免疫分型为$CD5^+CD7^+$的患者用CD5 CAR-T细胞或其群和/或CD7 CAR细胞或其群治疗。经免疫分型为CD33-CD123+的患者用CD123 CAR-T细胞或其群治疗。经免疫分型为CD33+CD123-的患者用CD33 CAR-T细胞或其群治疗。经免疫分型为CD33+CD123+的患者用CD3 CAR-T细胞或其群和/或CD123 CAR细胞或其群治疗。经免疫分型为$CD5^+CD7^+CD33^+$CD123+的患者用CD33 CAR-T细胞或其群、CD5 CAR-T细胞或其群、CD7 CAR-T细胞或其群和/或CD123 CAR-T细胞或其群的任意组合治疗。

[2166] 其他等效物

[2167] 从前述说明书明显可见,可对本文所述的发明作出变更和修饰以使其适用于各种用途和条件。此类实施方案也处于权利要求书的范畴内。

[2168] 本文中的任何变量定义中对一系列元件的叙述包括该变量作为任何单一元件或所列元件的组合(亚组合)的定义。本文中对于实施方案的叙述包括该实施方案作为任何单一实施方案或与其他实施方案或其部分组合。

[2169] 本说明书中提及的全部出版物、专利和专利申请通过引用以与各独立出版物、专利或专利申请被具体且独立地指明为通过引用并入相同的程度并入本文。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而以其整体并入本文。

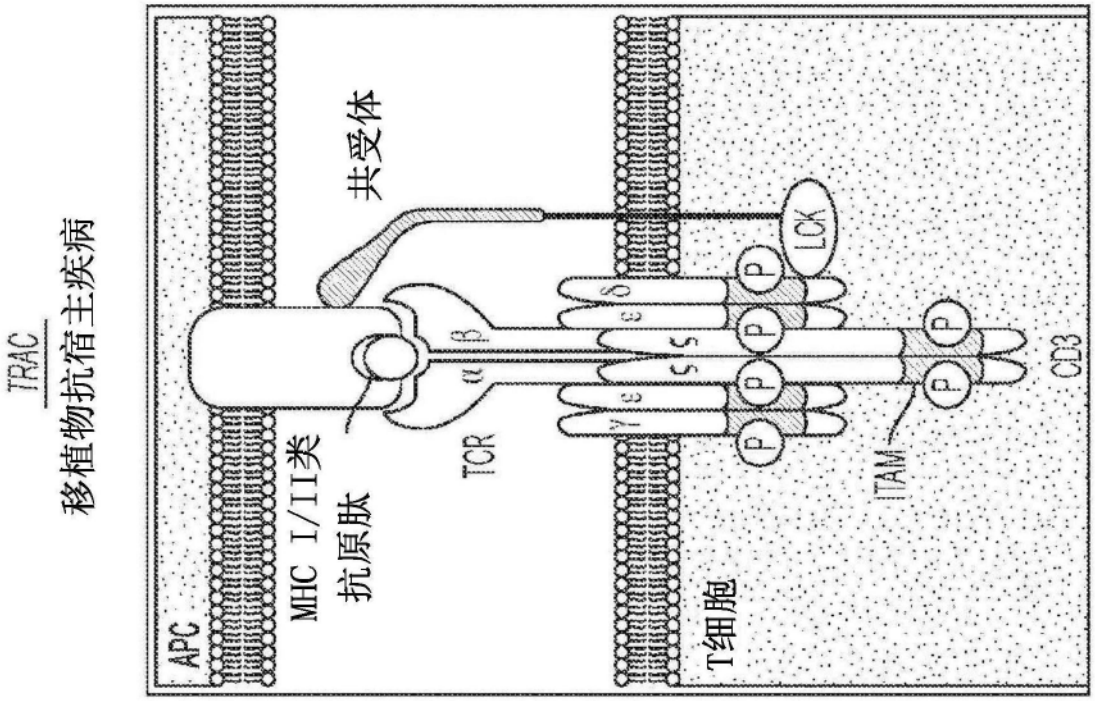

TRAC
移植物抗宿主疾病
APC
MHC I/II类
抗原肽
共受体
TCR
T细胞
ITAM
LCK
CD3

图1A

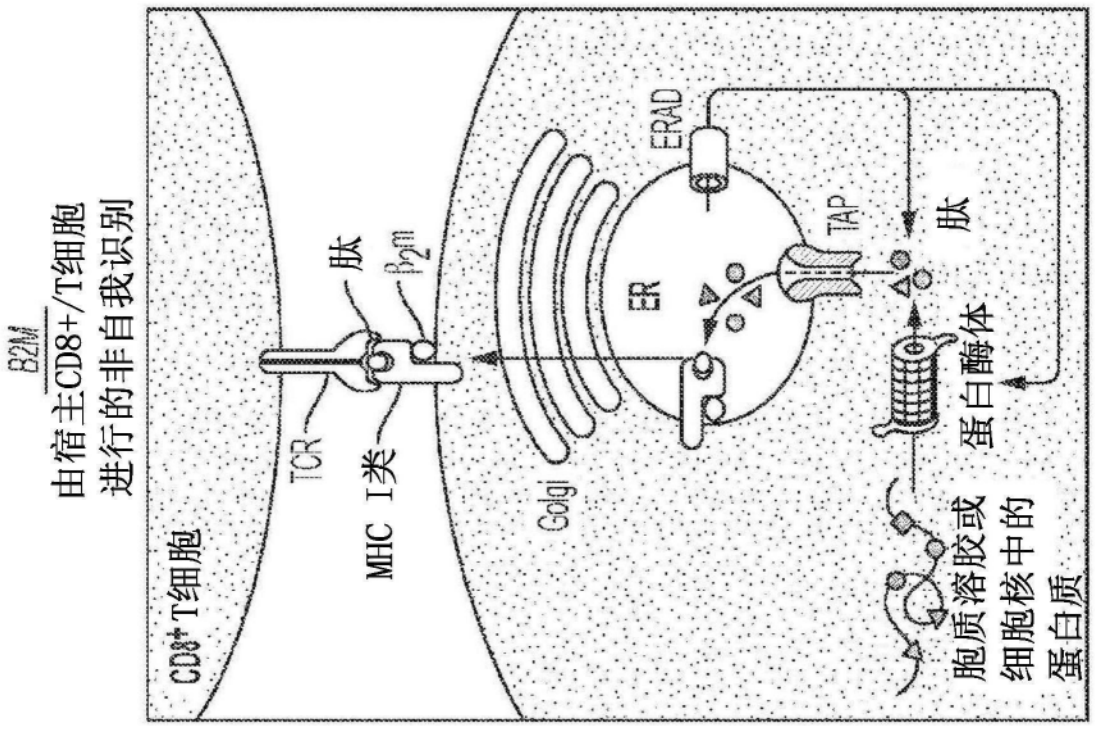

B2M
由宿主CD8+/T细胞
进行的非自我识别
CD8+T细胞
TCR
肽
MHC I类
β2m
Golgi
ER
ERAD
TAP
肽
蛋白酶体
胞质溶胶或
细胞胞核中的
蛋白质

图1B

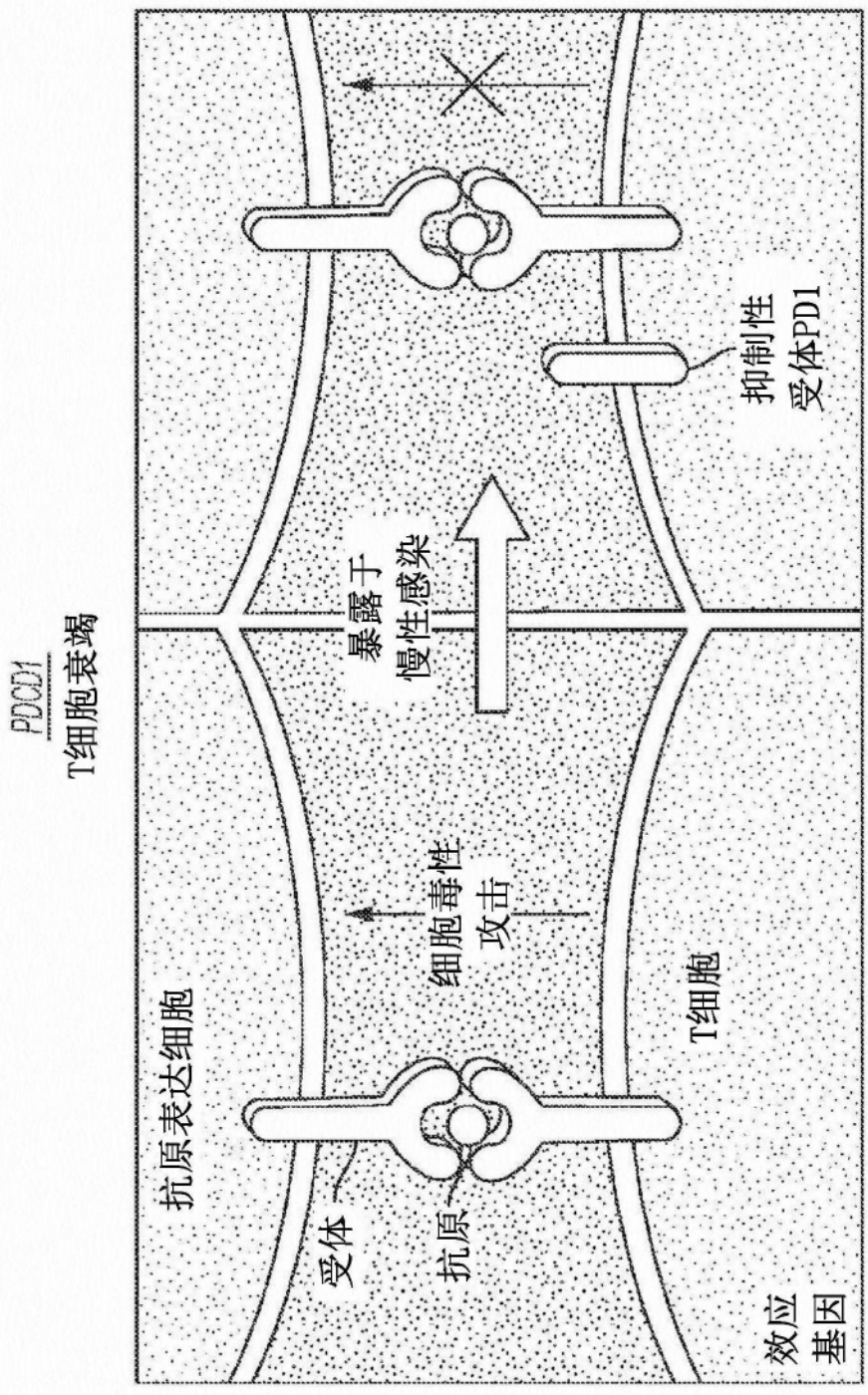


图1C

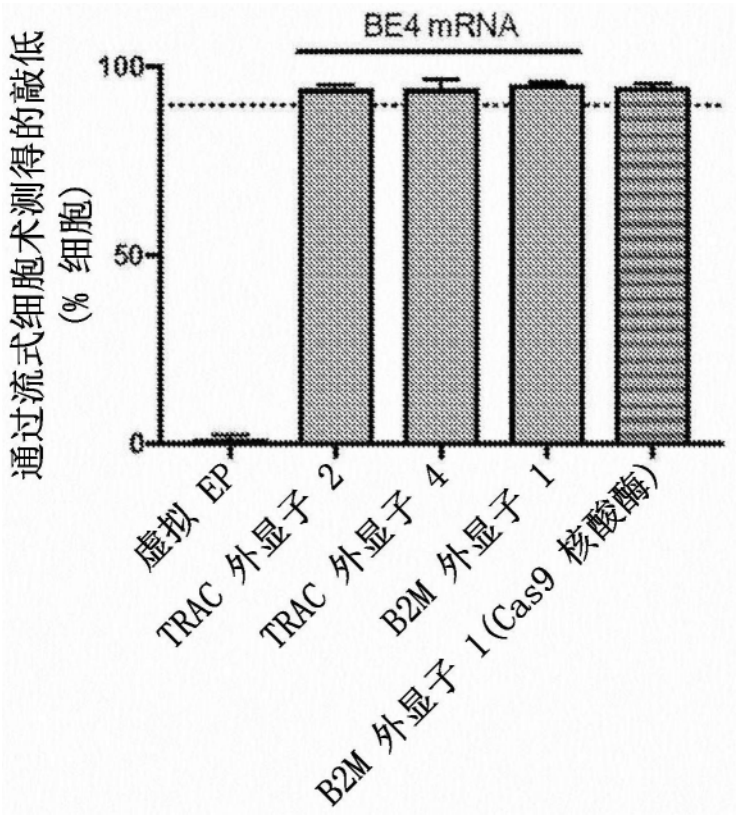

BE4 mRNA
通过流式细胞术测得的敲低
(% 细胞)
100
50
0
虚拟 EP
TRAC 外显子 2
TRAC 外显子 4
B2M 外显子 1
B2M 外显子 1(Cas9 核酸酶)

图2

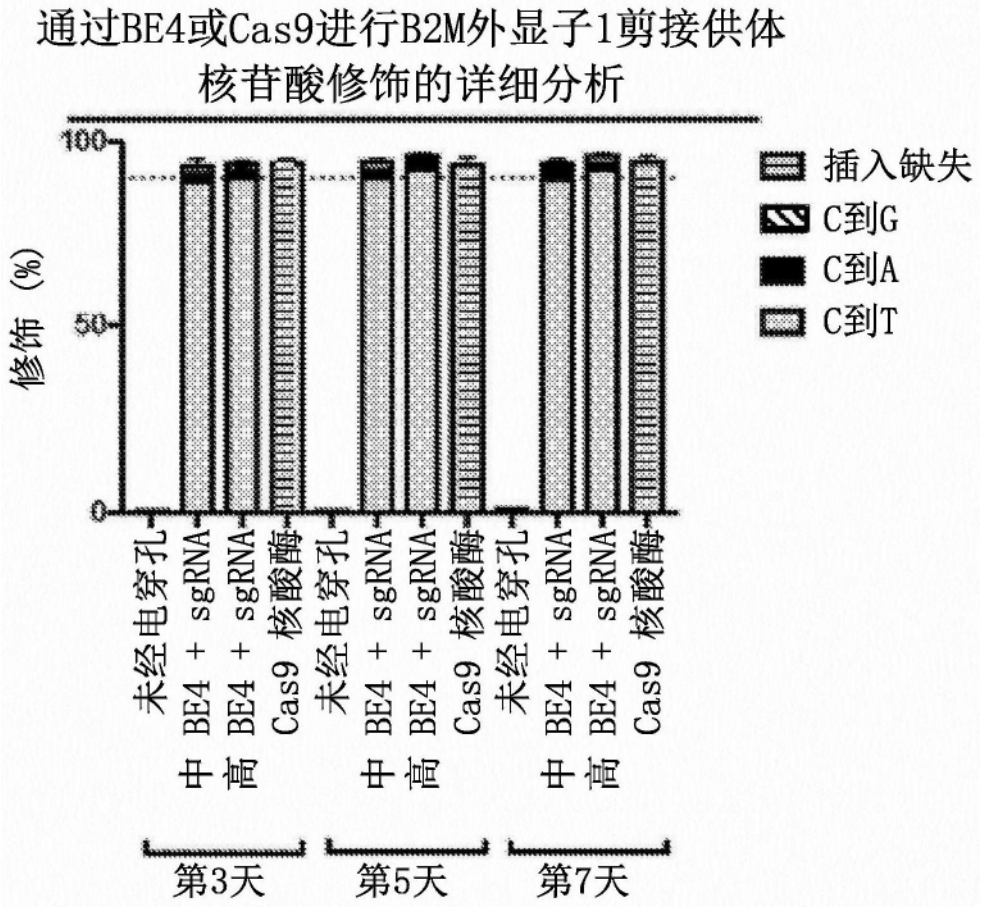

通过BE4或Cas9进行B2M外显子1剪接供体
核苷酸修饰的详细分析
修饰 (%)
100
50
0
插入缺失
C到G
C到A
C到T
未经电穿孔
中 BE4 + sgRNA
高 BE4 + sgRNA
Cas9 核酸酶
第3天
第5天
第7天

图3

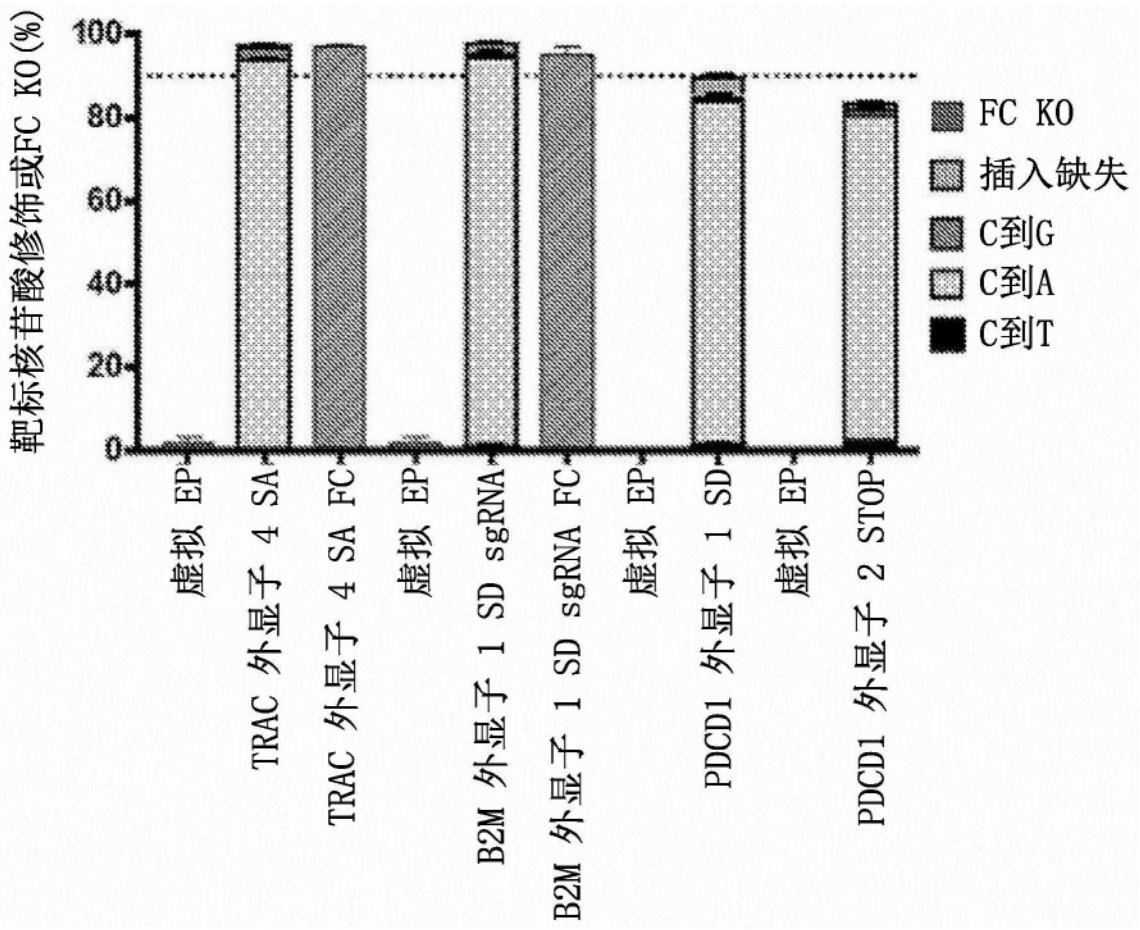

靶标核苷酸修饰或FC KO(%)
100
80
60
40
20
0
虚拟 EP
TRAC 外显子 4 SA
TRAC 外显子 4 SA FC
虚拟 EP
B2M 外显子 1 SD sgRNA
B2M 外显子 1 SD sgRNA FC
虚拟 EP
PDCD1 外显子 1 SD
虚拟 EP
PDCD1 外显子 2 STOP
FC KO
插入缺失
C到G
C到A
C到T

图4

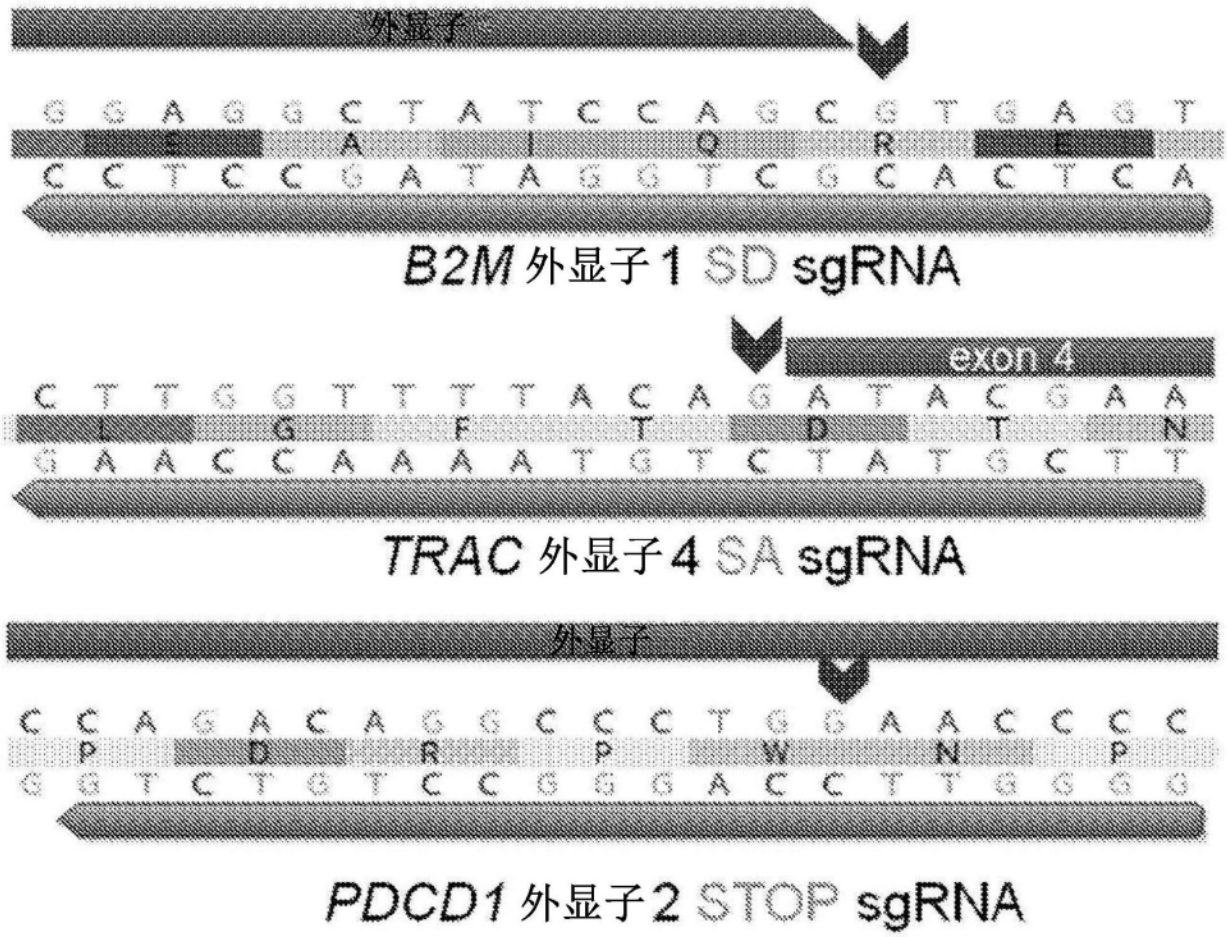
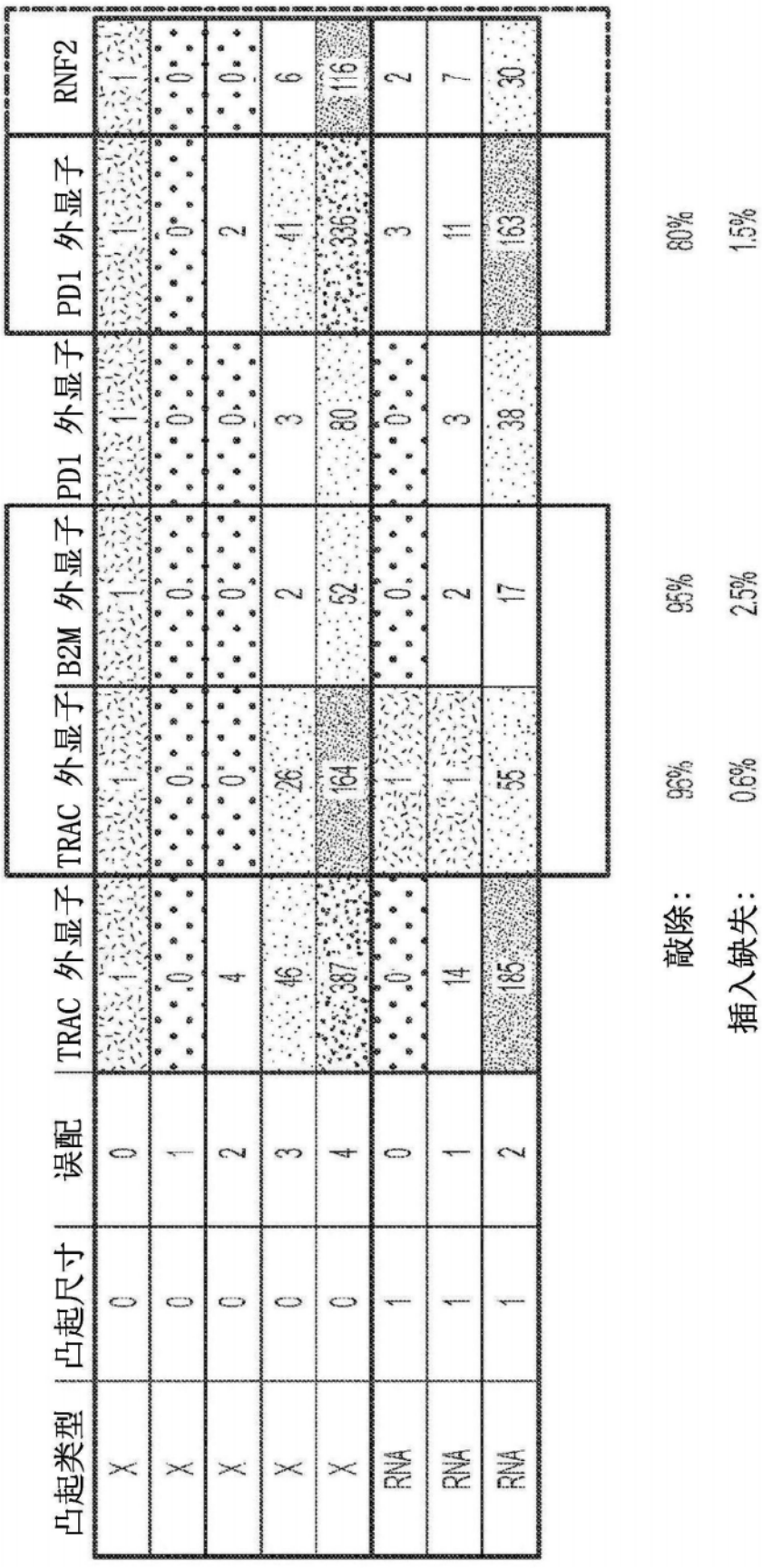

外显子
B2M 外显子1 SD sgRNA
exon 4
TRAC 外显子4 SA sgRNA
外显子
PDCD1 外显子2 STOP sgRNA

图5



凸起类型	凸起尺寸	误配	TRAC 外显子	TRAC 外显子	B2M 外显子	PD1 外显子	PD1 外显子	RNF2
X	0	0	1	1	1	1	1	1
X	0	1	0	0	0	0	0	0
X	0	2	4	0	0	0	2	0
X	0	3	46	26	2	3	41	6
X	0	4	387	164	52	80	336	116
RNA	1	0	0	1	0	0	3	2
RNA	1	1	14	1	2	3	11	7
RNA	1	2	185	55	17	38	163	30

敲除： 96% 95% 80%

插入缺失： 0.6% 2.5% 1.5%

图6

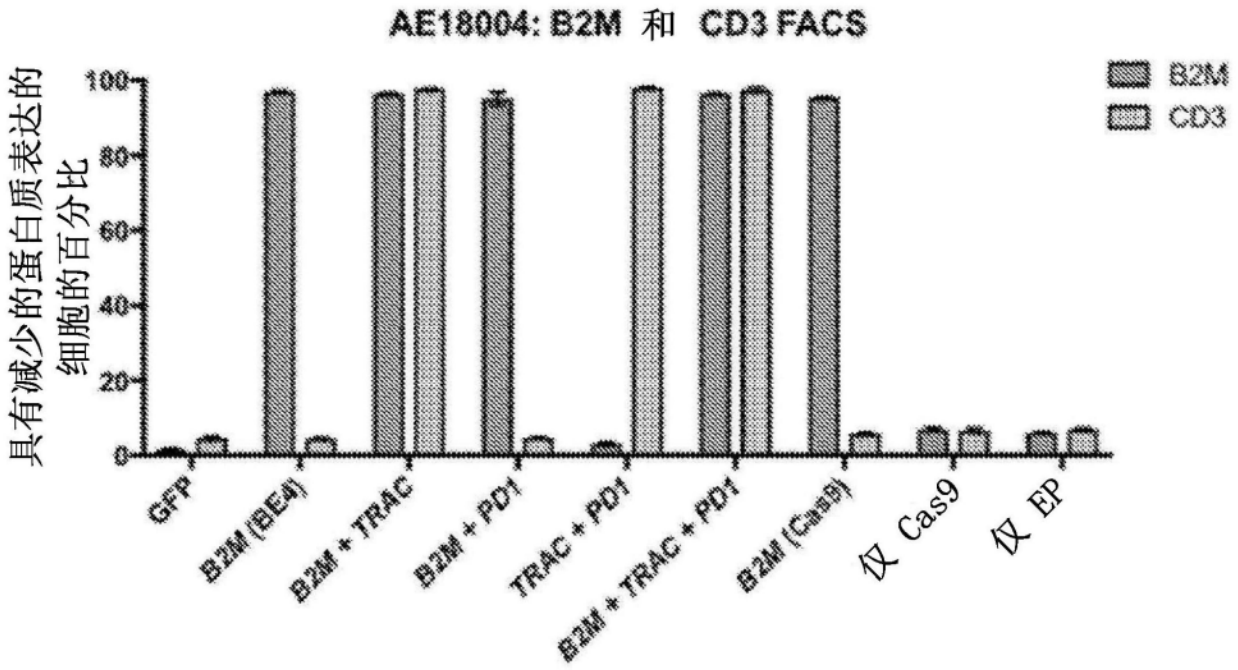

AE18004: B2M 和 CD3 FACS
具有减少的蛋白质表达的细胞的百分比
B2M
CD3
100
80
60
40
20
0
GFP
B2M (BE4)
B2M + TRAC
B2M + PD1
TRAC + PD1
B2M + TRAC + PD1
B2M (Cas9)
仅 Cas9
仅 EP

图7

对照

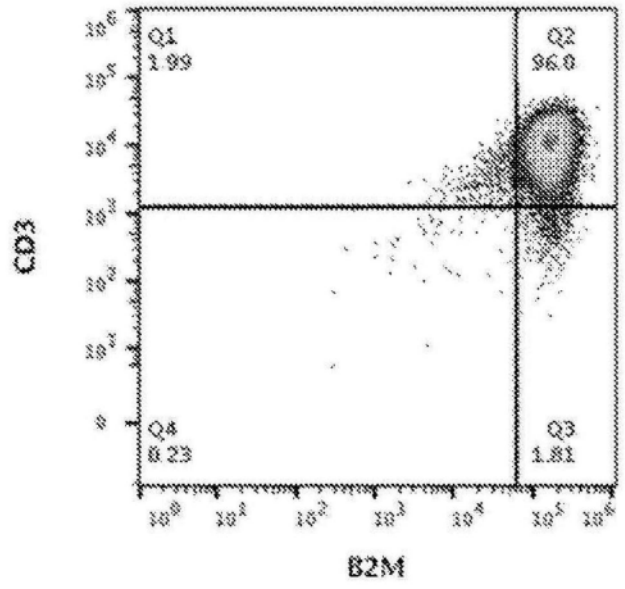

Q1
1.99
Q2
96.0
Q4
0.23
Q3
1.81
CD3
B2M

图8A

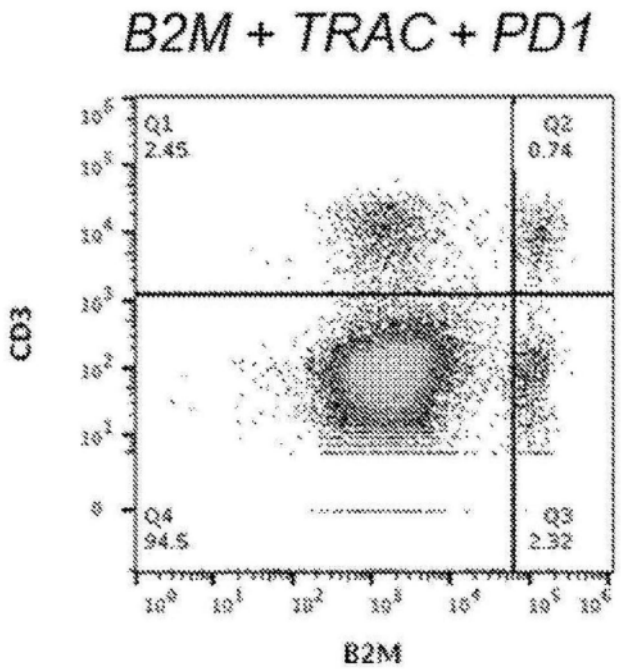

B2M + TRAC + PD1
Q1
2.45
Q2
0.74
Q4
94.5
Q3
2.32
CD3
B2M

图8B

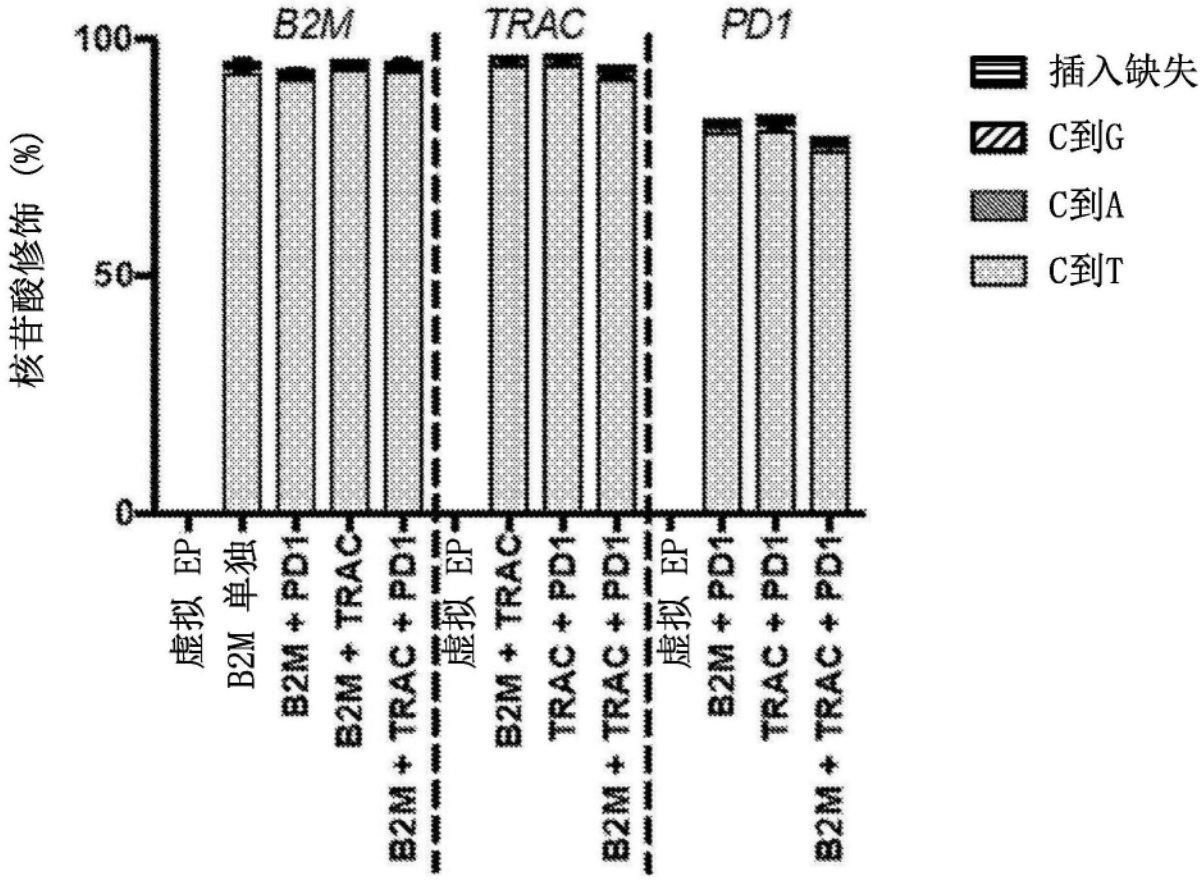

B2M
TRAC
PD1
100
50
0
核苷酸修饰 (%)
插入缺失
C到G
C到A
C到T
虚拟 EP
B2M 单独
B2M + PD1
B2M + TRAC
B2M + TRAC + PD1
虚拟 EP
B2M + TRAC
TRAC + PD1
B2M + TRAC + PD1
虚拟 EP
B2M + PD1
TRAC + PD1
B2M + TRAC + PD1

图9

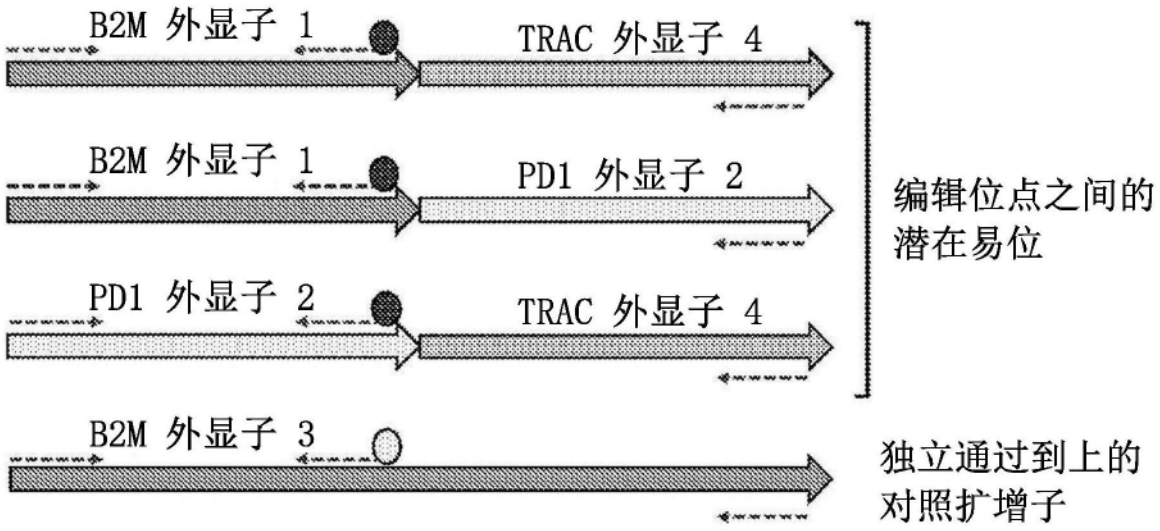

B2M 外显子 1
TRAC 外显子 4
B2M 外显子 1
PD1 外显子 2
编辑位点之间的
潜在易位
PD1 外显子 2
TRAC 外显子 4
B2M 外显子 3
独立通过到上的
对照扩增子

图10

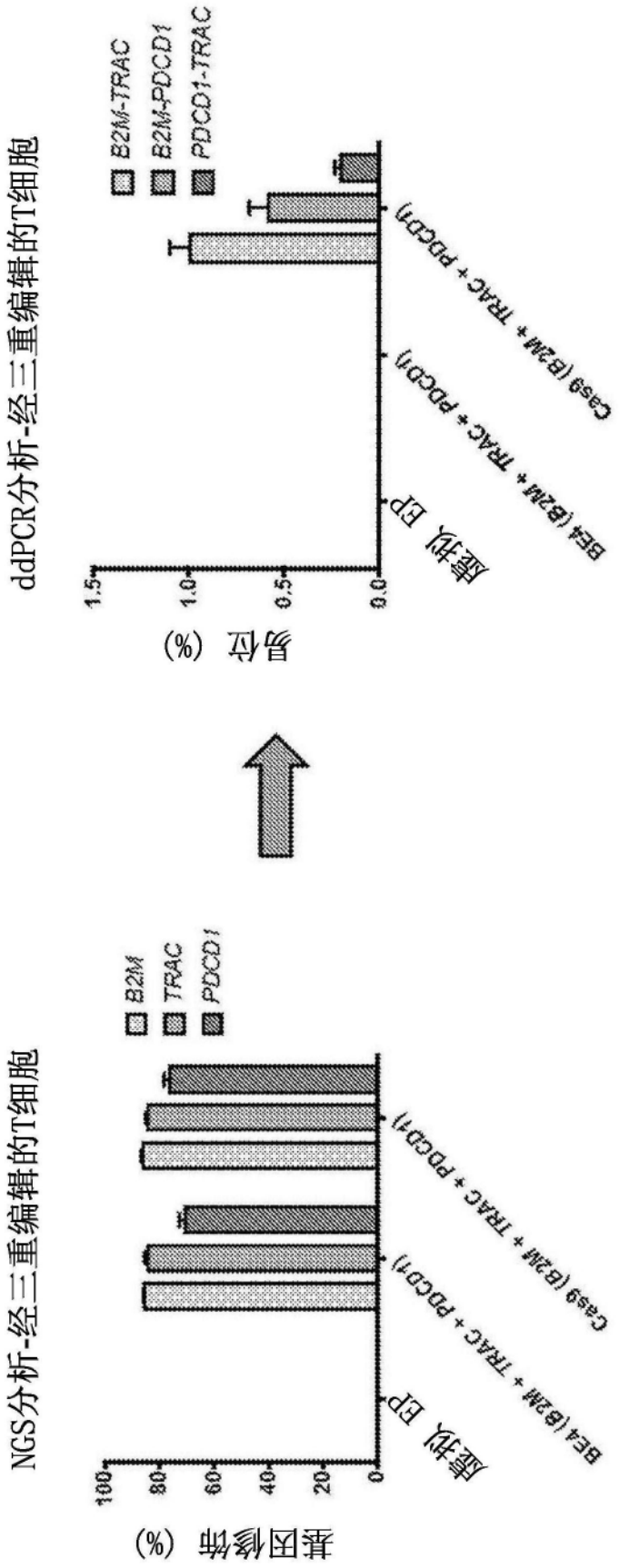

ddPCR分析-经三重编辑的T细胞
B2M-TRAC
B2M-PDCD1
PDCD1-TRAC
易位 (%)
1.5
1.0
0.5
0.0
假拟 EP
BE4 (B2M + TRAC + PDCD1)
Cas9 (B2M + TRAC + PDCD1)
NGS分析-经三重编辑的T细胞
B2M
TRAC
PDCD1
基因修饰 (%)
100
80
60
40
20
0
假拟 EP
BE4 (B2M + TRAC + PDCD1)
Cas9 (B2M + TRAC + PDCD1)

图11

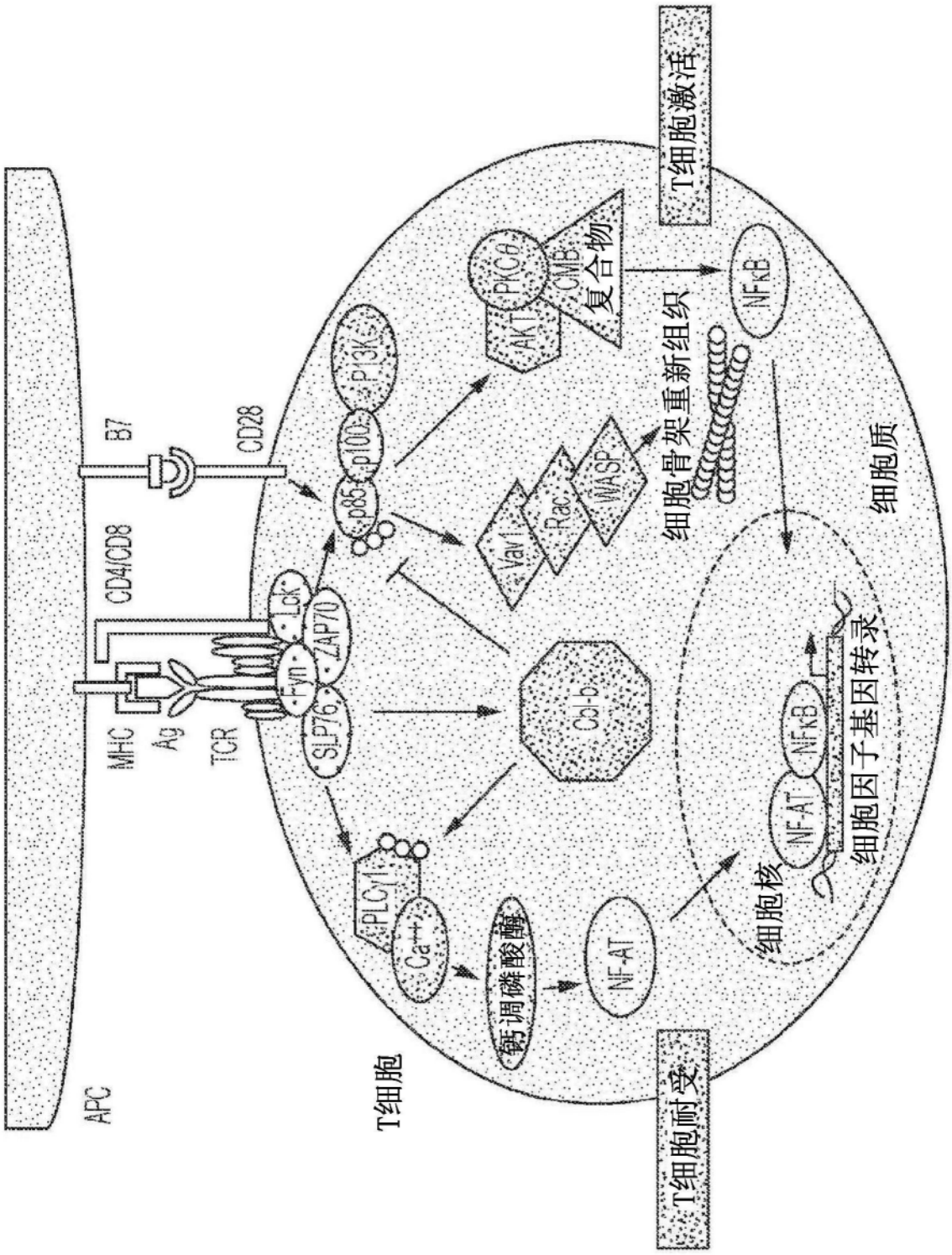

APC
MHC
Ag
TCR
CD4/CD8
B7
CD28
T细胞
T细胞激活
T细胞耐受
细胞质
细胞核
细胞骨架重新组织
细胞因子基因转录
Lck
Fyn
ZAP70
SLP76
PLCγ1
Ca++
钙调磷酸酶
NF-AT
p85
p110
PI3K
AKT
PKCθ
CMB
复合物
Vav1
Rac
WASP
Cbl-b
NFκB

图12

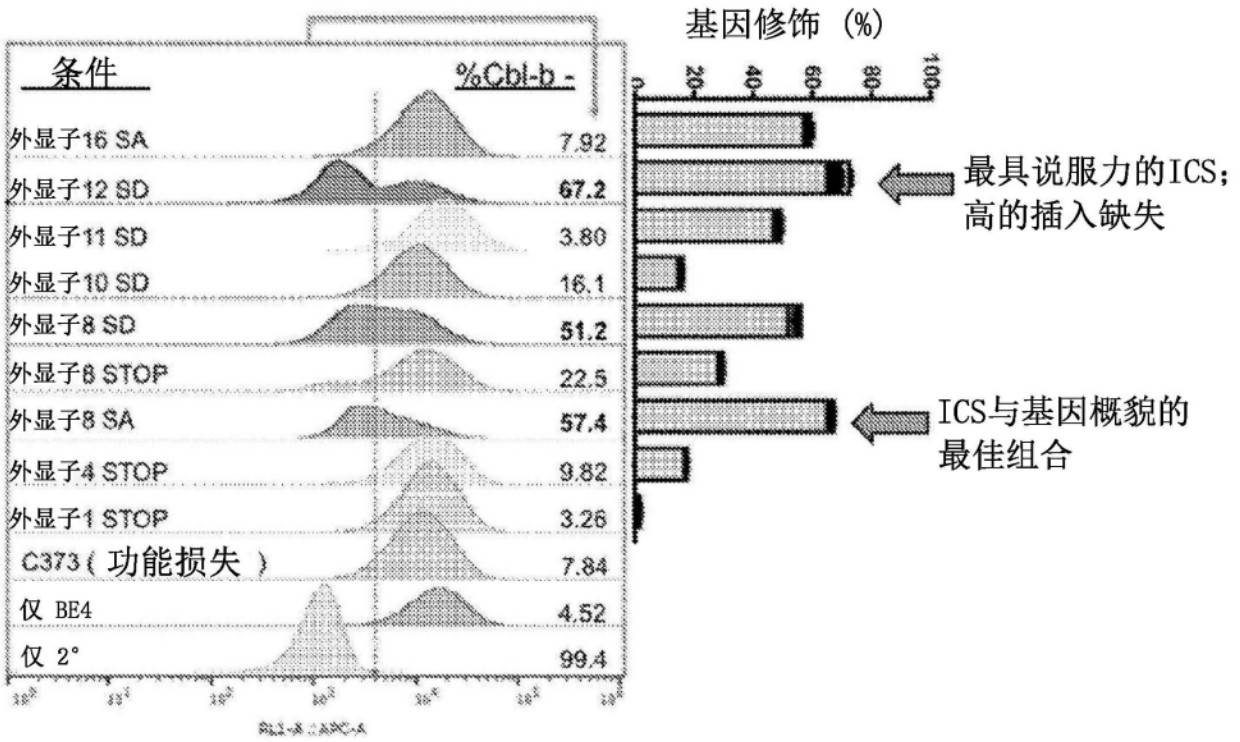

条件
%Cbl-b -
外显子16 SA
7.92
外显子12 SD
67.2
外显子11 SD
3.80
外显子10 SD
16.1
外显子8 SD
51.2
外显子8 STOP
22.5
外显子8 SA
57.4
外显子4 STOP
9.82
外显子1 STOP
3.26
C373（功能损失）
7.84
仅 BE4
4.52
仅 2°
99.4
基因修饰（%）
0
20
40
60
80
100
最具说服力的ICS；
高的插入缺失
ICS与基因概貌的
最佳组合

图13

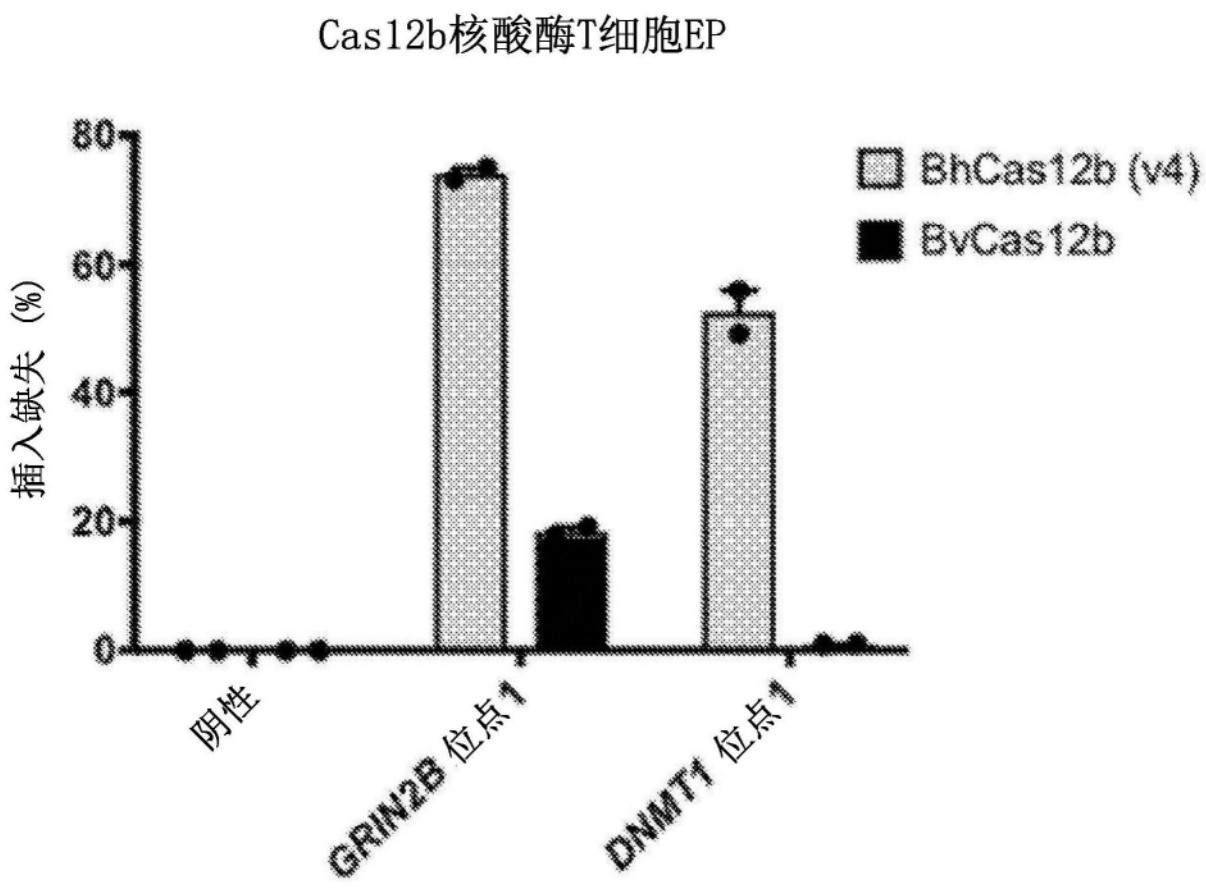

Cas12b核酸酶T细胞EP
插入缺失（%）
80
60
40
20
0
BhCas12b (v4)
BvCas12b
阴性
GRIN2B 位点1
DNMT1 位点1

图14

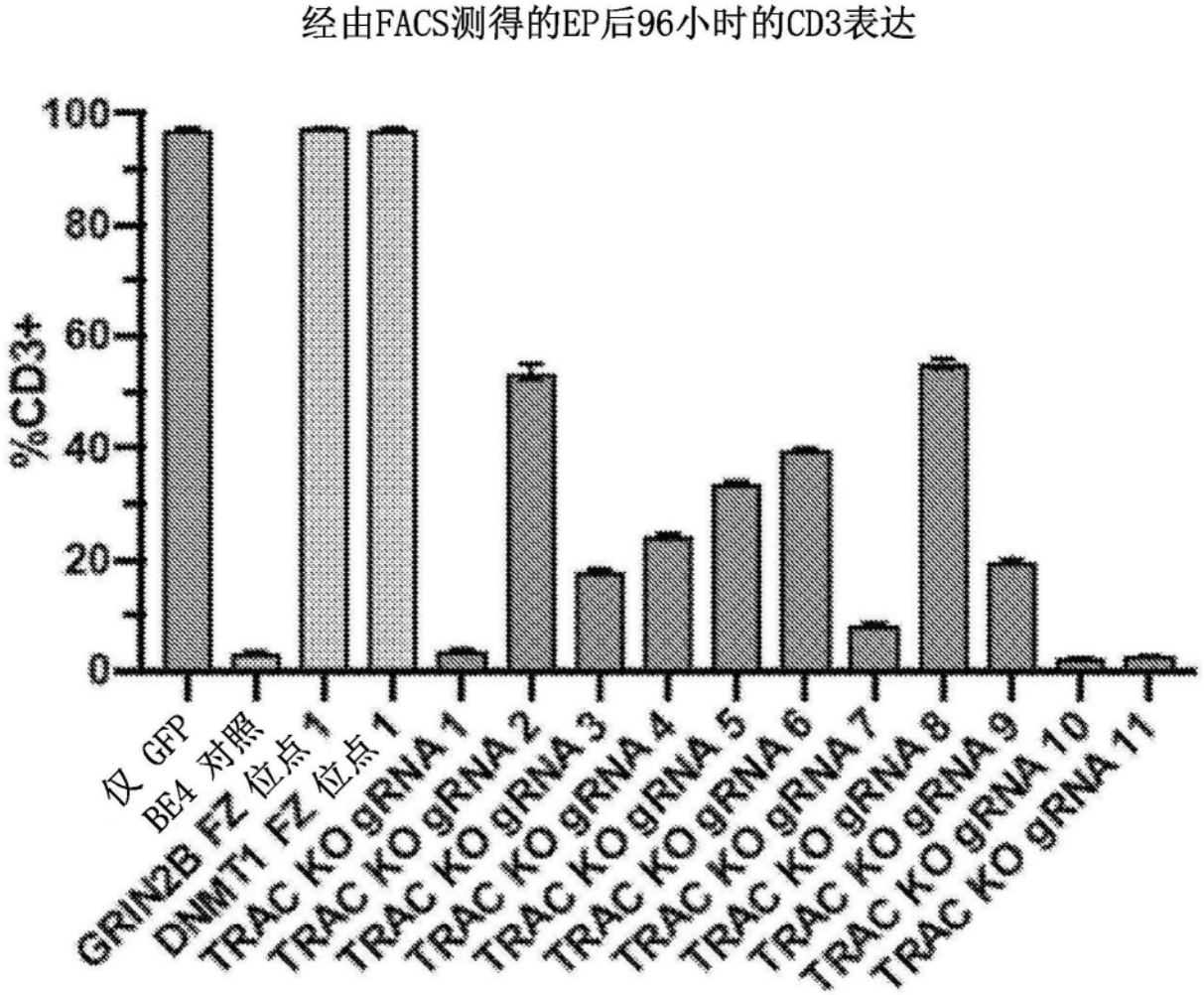

经由FACS测得的EP后96小时的CD3表达
%CD3+
100
80
60
40
20
0
仅 GFP
BE4 对照
GRIN2B FZ 位点 1
DNMT1 FZ 位点 1
TRAC KO gRNA 1
TRAC KO gRNA 2
TRAC KO gRNA 3
TRAC KO gRNA 4
TRAC KO gRNA 5
TRAC KO gRNA 6
TRAC KO gRNA 7
TRAC KO gRNA 8
TRAC KO gRNA 9
TRAC KO gRNA 10
TRAC KO gRNA 11

图15

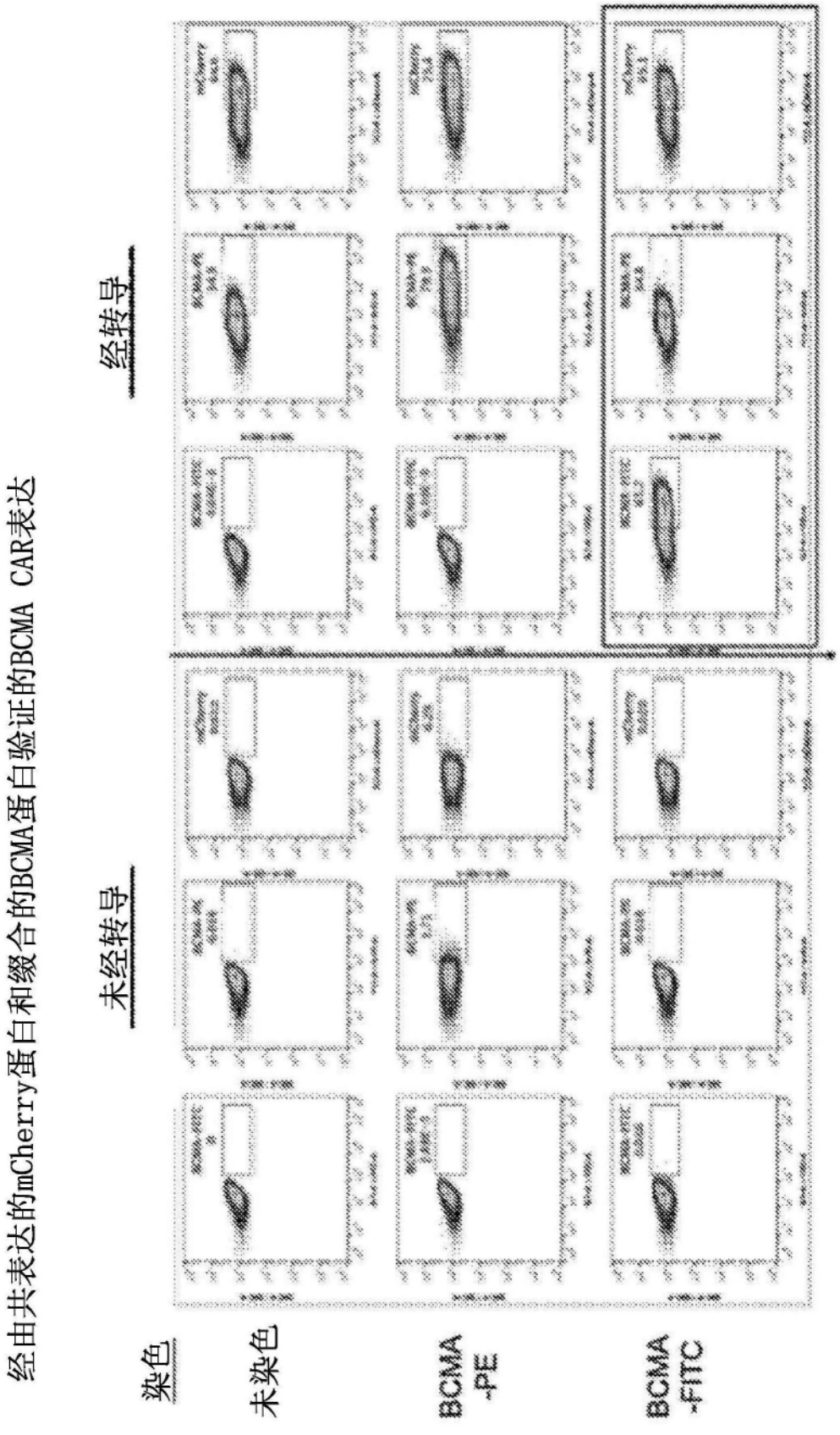

经由共表达的mCherry蛋白和缀合的BCMA蛋白验证的BCMA CAR表达
未经转导
经转导
染色
未染色
BCMA -PE
BCMA -FITC

图16

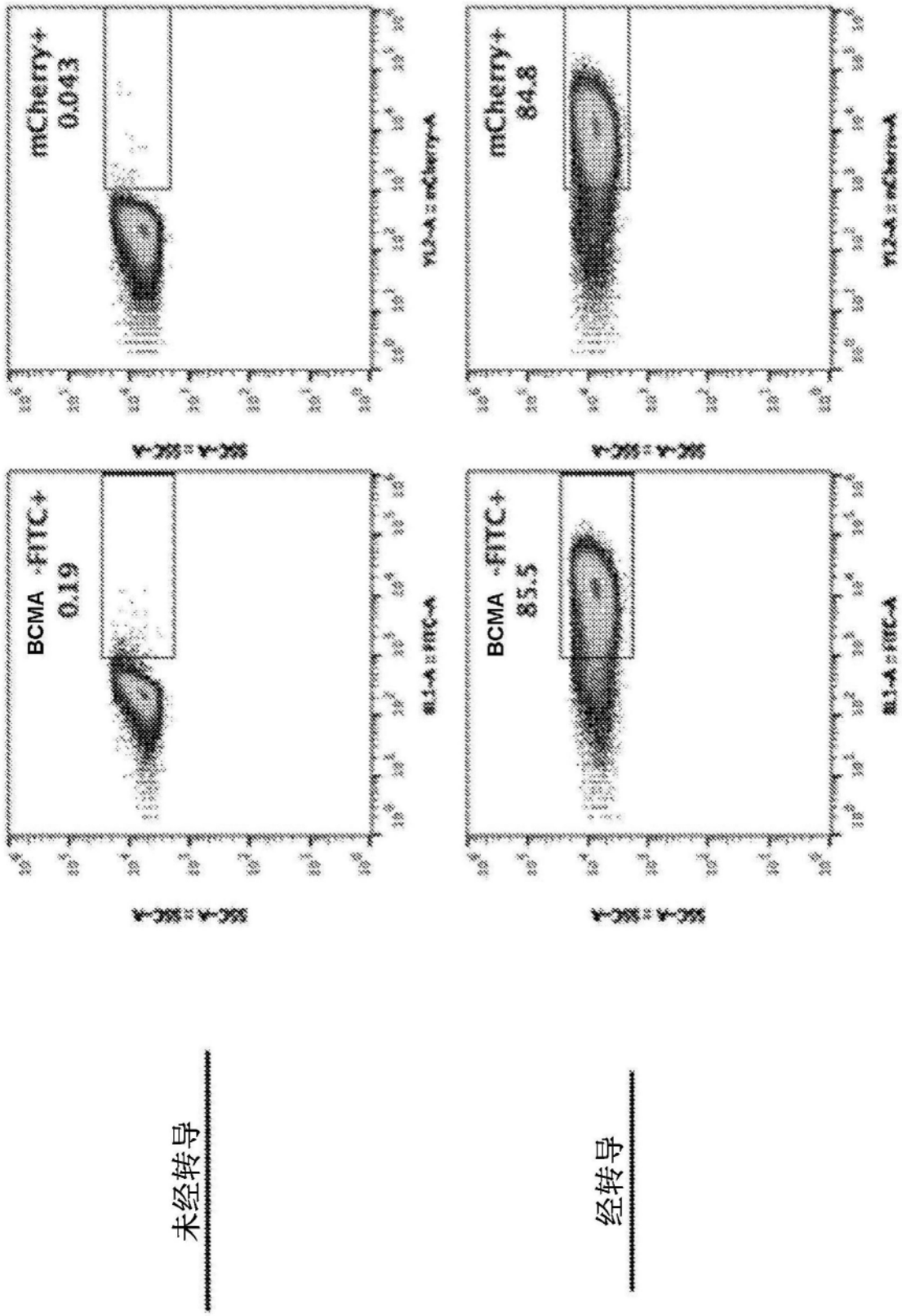

mCherry+
0.043
BCMA -FITC+
0.19
mCherry+
84.8
BCMA -FITC+
85.5
未经转导
经转导

图17

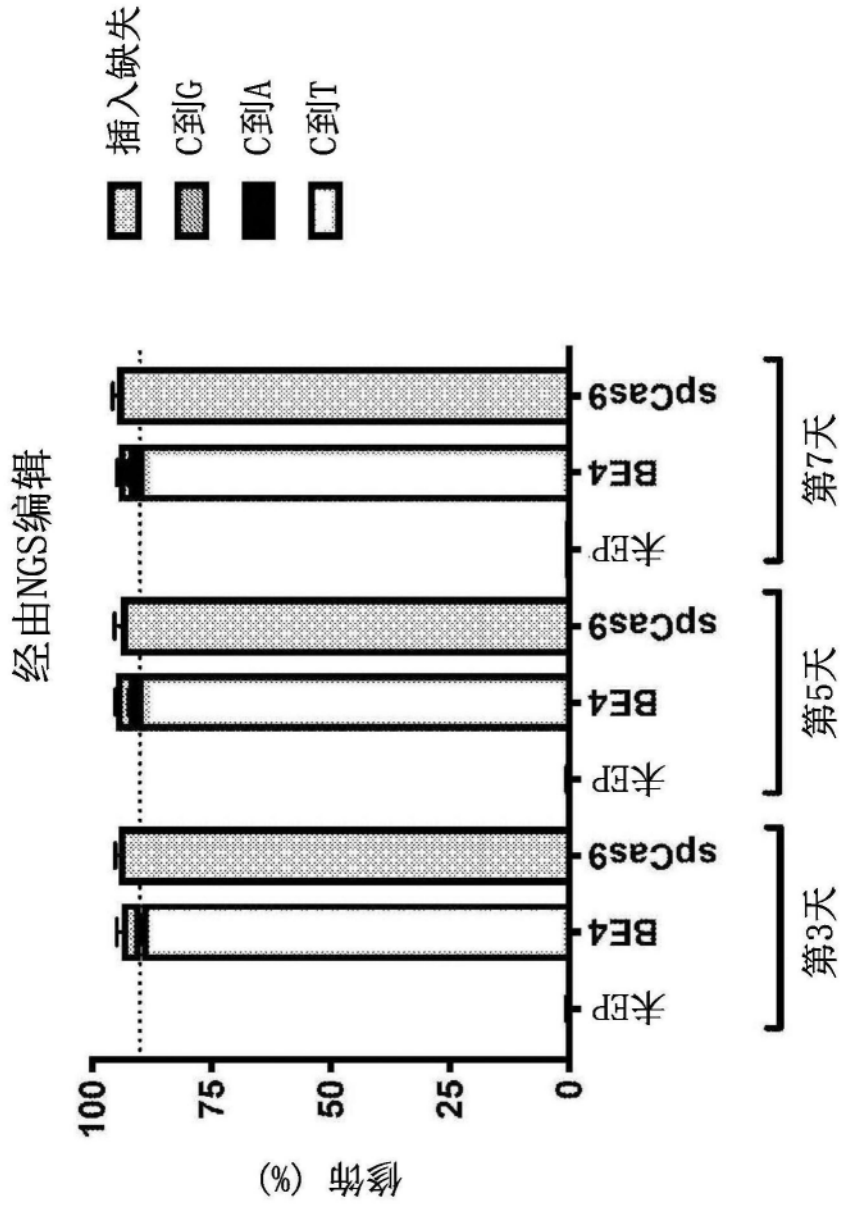

插入缺失
C到G
C到A
C到T
经由NGS编辑
修饰 (%)
100
75
50
25
0
未EP
BE4
spCas9
第3天
第5天
第7天

图18

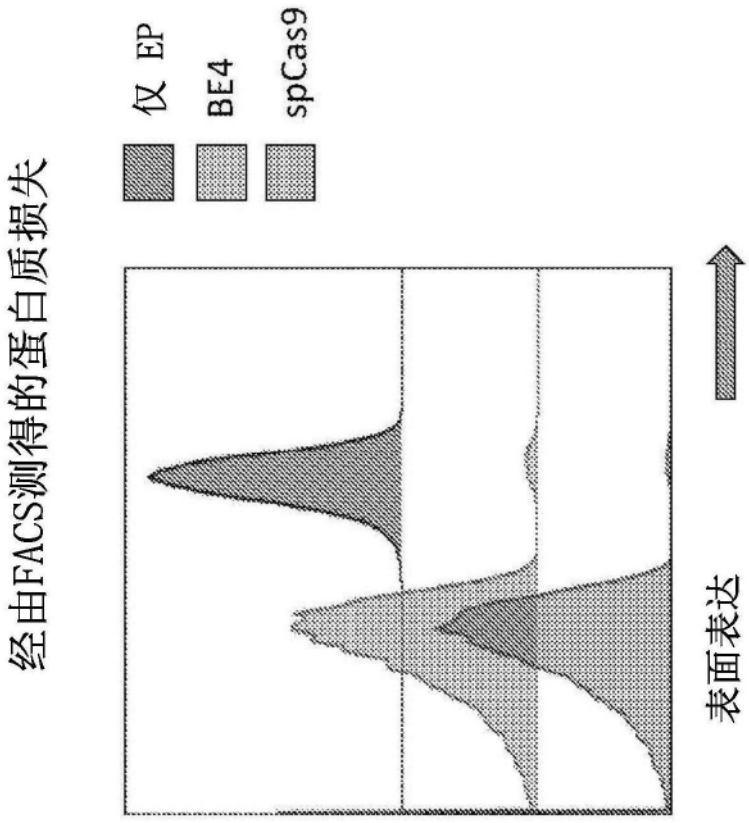


图19A

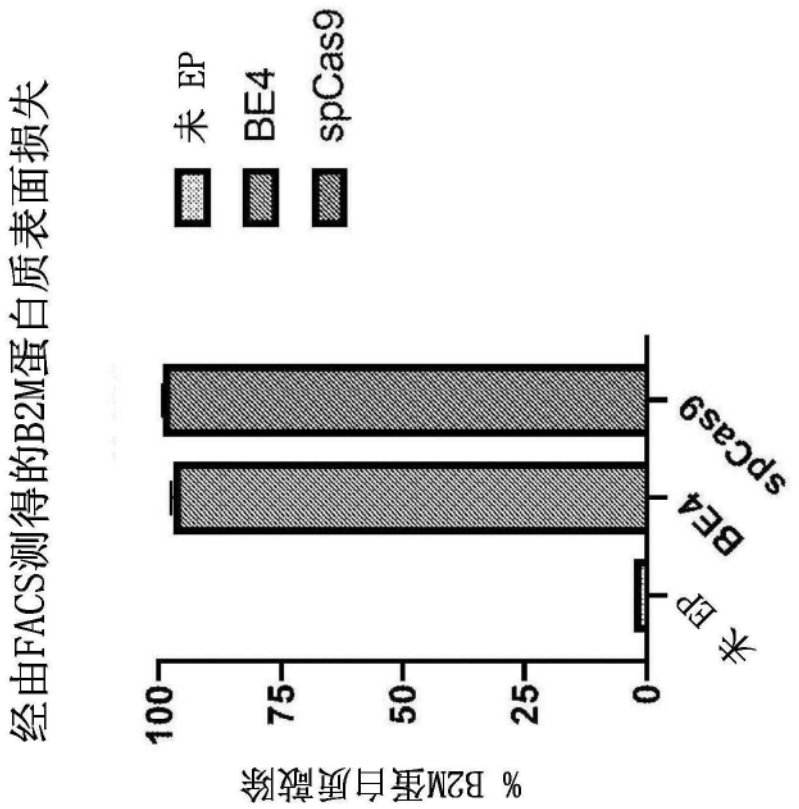

经由FACS测得的B2M蛋白质表面损失
未 EP
BE4
spCas9
% B2M蛋白质敲除
100
75
50
25
0
未 EP
BE4
spCas9

图19B

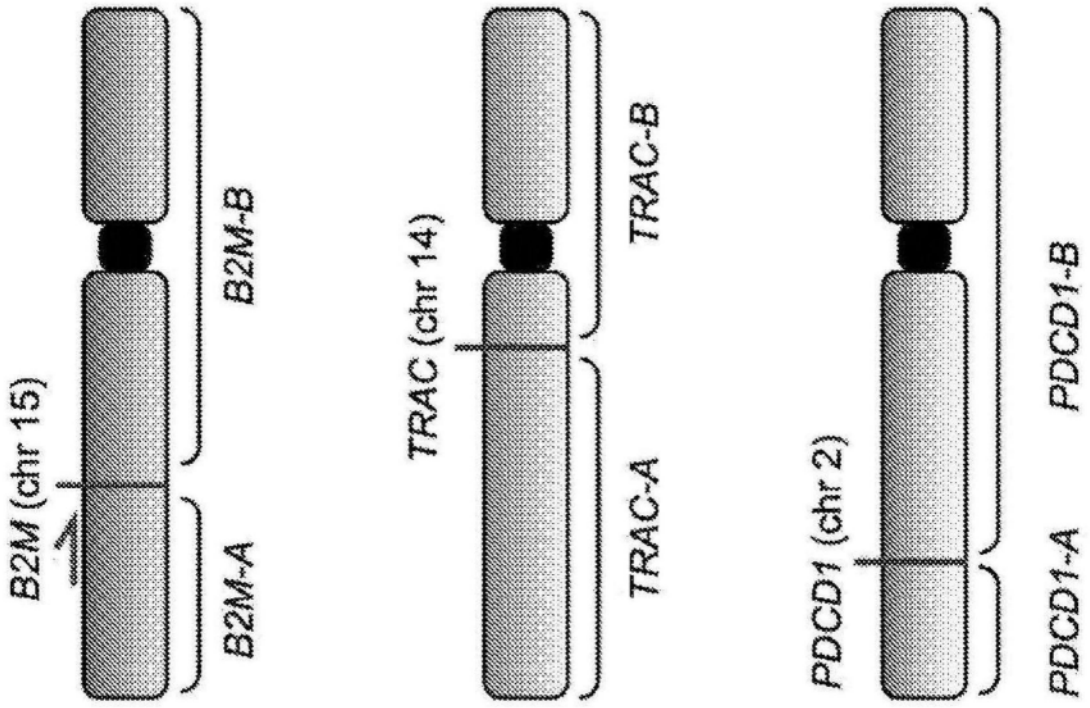

B2M (chr 15)
B2M-A
B2M-B
TRAC (chr 14)
TRAC-A
TRAC-B
PDCD1 (chr 2)
PDCD1-A
PDCD1-B

图20

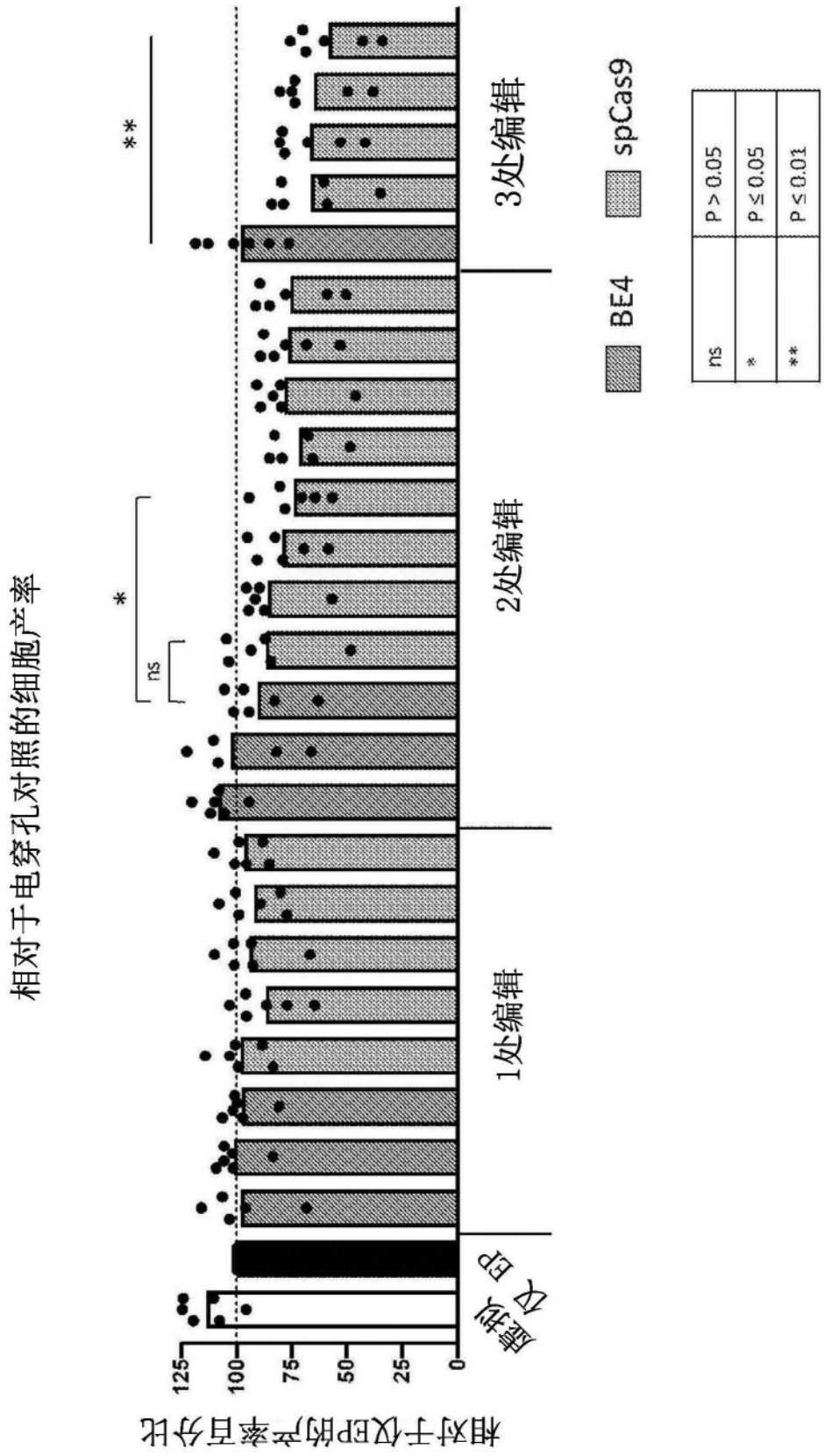

相对于电穿孔对照的细胞产率
相对于仅EP的产率百分比
125
100
75
50
25
0
仅EP
1处编辑
2处编辑
3处编辑
ns
*
**
BE4
spCas9
ns
P > 0.05
*
P ≤ 0.05
**
P ≤ 0.01

图21

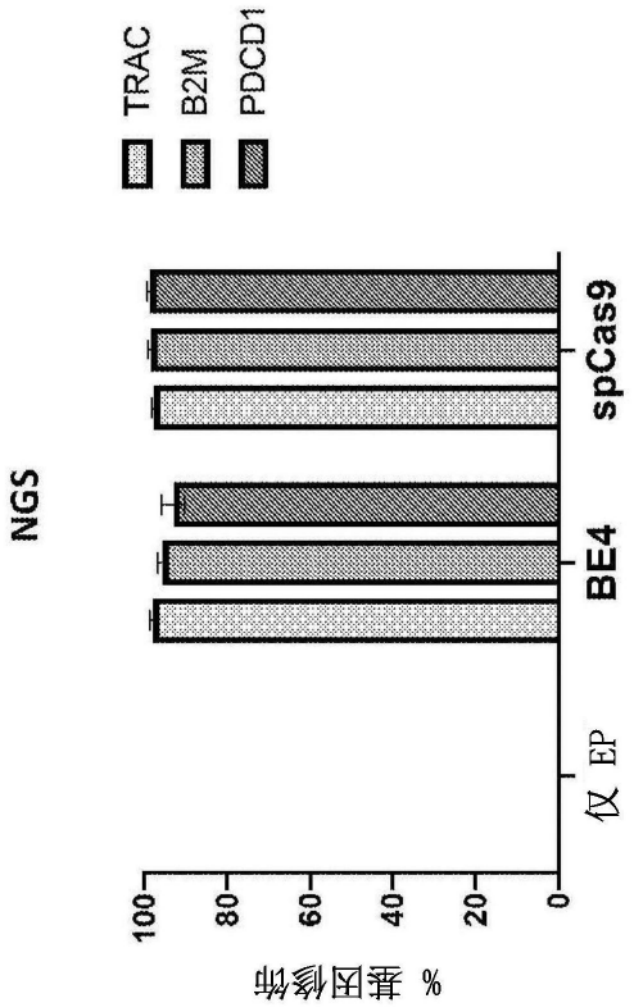


图22

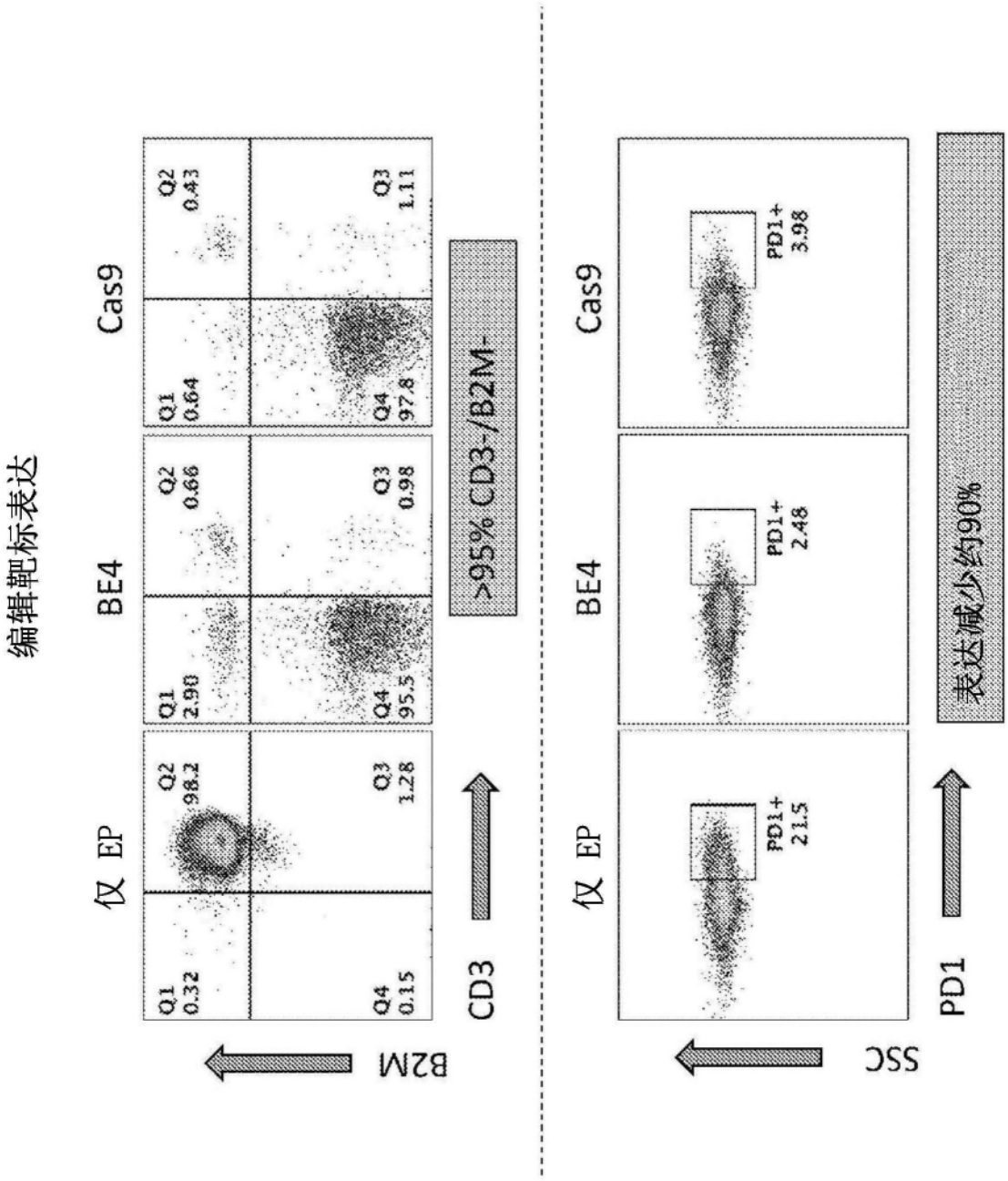

编辑靶标表达
仅 EP
BE4
Cas9
Q1 0.32
Q2 98.2
Q4 0.15
Q3 1.28
Q1 2.90
Q2 0.66
Q4 95.5
Q3 0.98
Q1 0.64
Q2 0.43
Q4 97.8
Q3 1.11
B2M
CD3
>95% CD3-/B2M-
PD1+ 21.5
PD1+ 2.48
PD1+ 3.98
表达减少约90%
SSC
PD1

图23

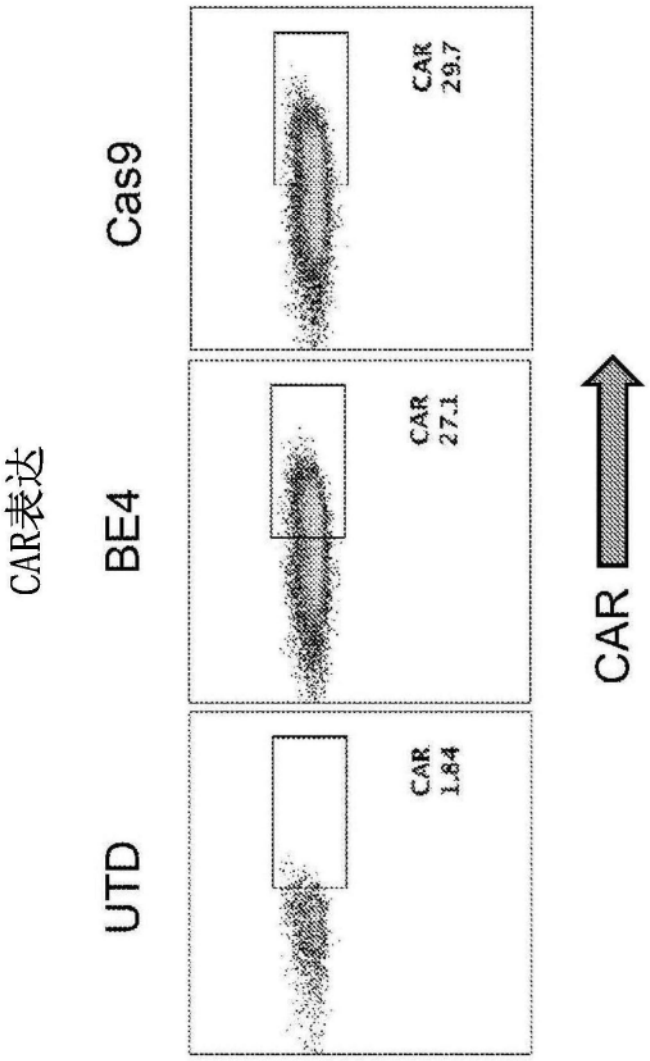


图24

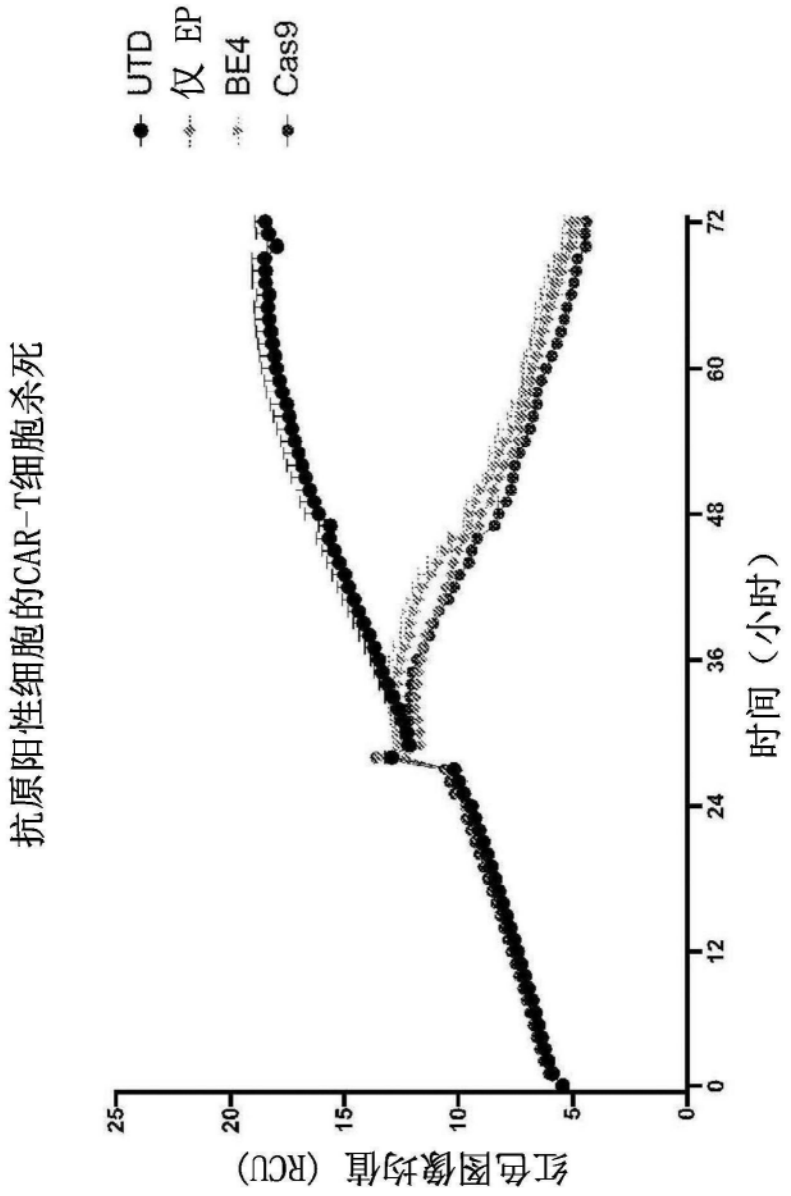

抗原阳性细胞的CAR-T细胞杀死
UTD
仅 EP
BE4
Cas9
红色图像均值 (RCU)
25
20
15
10
5
0
0
12
24
36
48
60
72
时间（小时）

图25

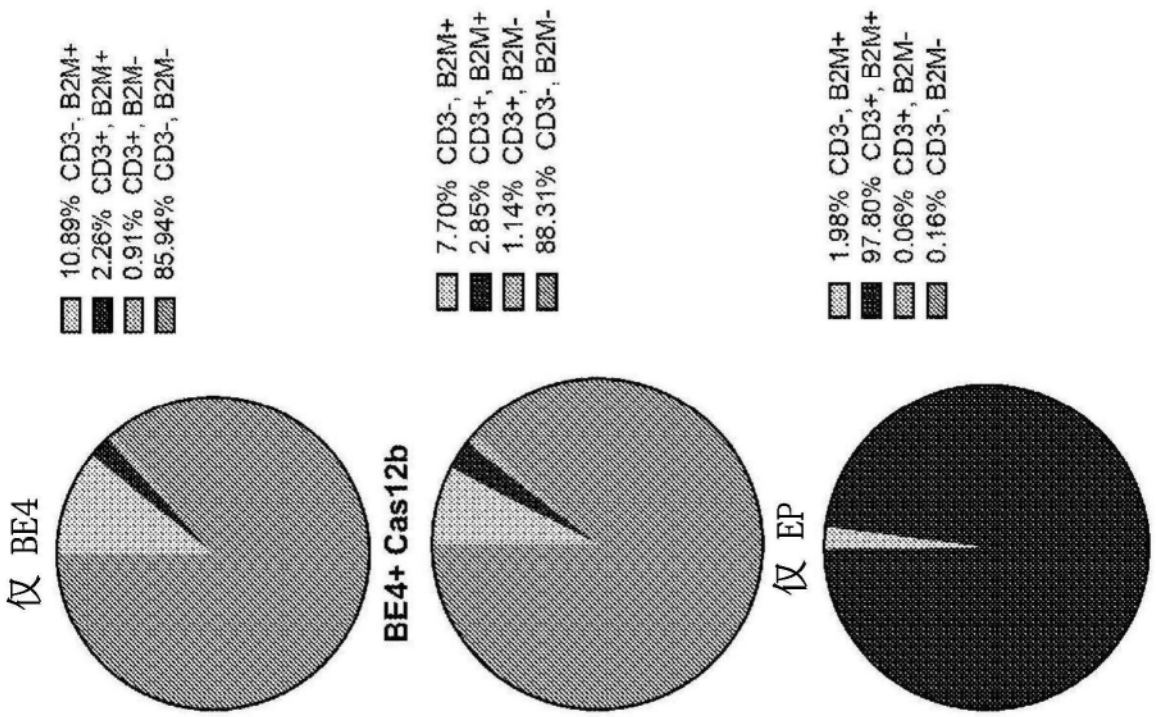

仅 BE4
10.89% CD3-, B2M+
2.26% CD3+, B2M+
0.91% CD3+, B2M-
85.94% CD3-, B2M-
BE4+ Cas12b
7.70% CD3-, B2M+
2.85% CD3+, B2M+
1.14% CD3+, B2M-
88.31% CD3-, B2M-
仅 EP
1.98% CD3-, B2M+
97.80% CD3+, B2M+
0.06% CD3+, B2M-
0.16% CD3-, B2M-

图26

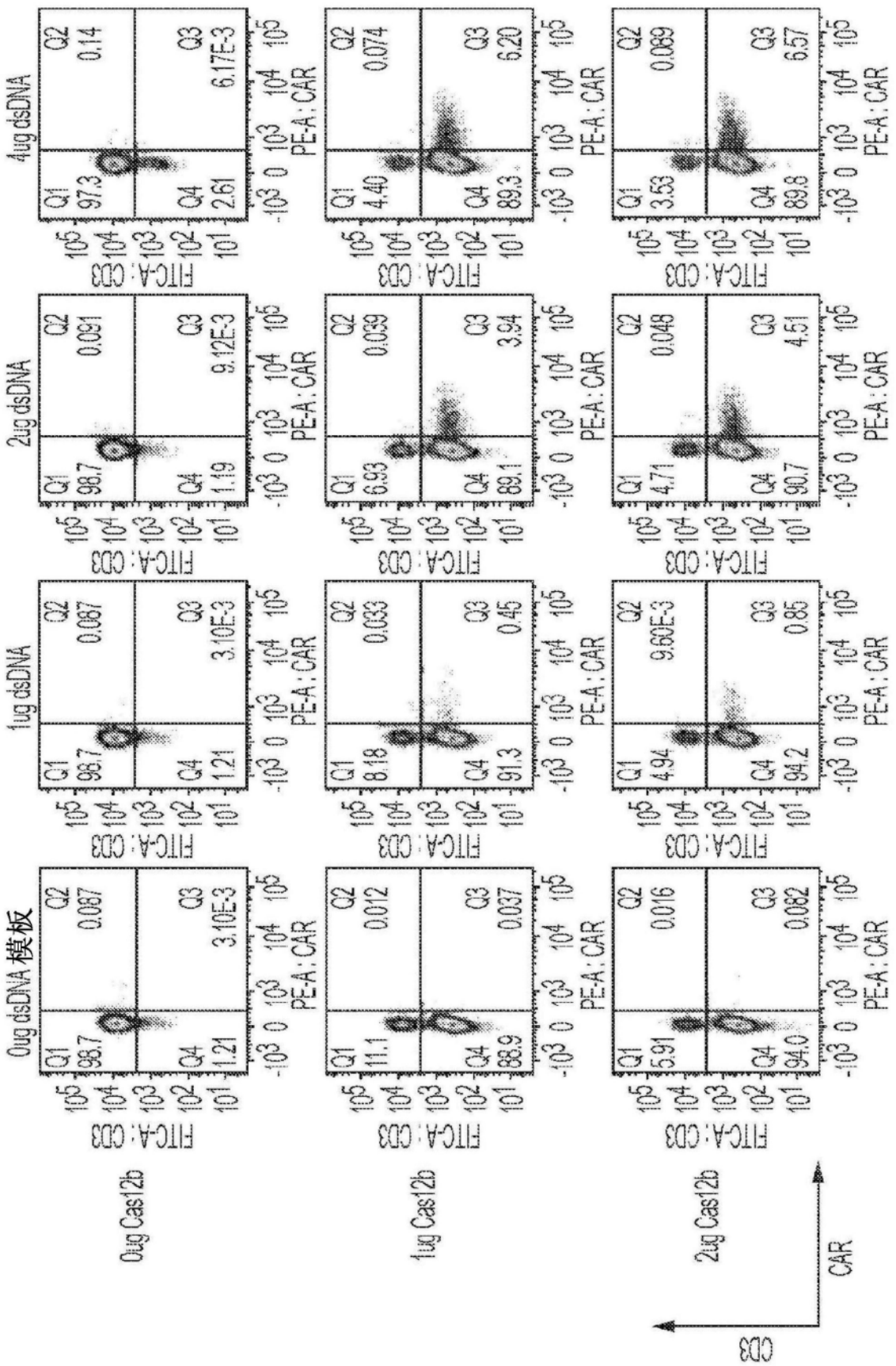

0ug dsDNA 模板
1ug dsDNA
2ug dsDNA
4ug dsDNA
0ug Cas12b
1ug Cas12b
2ug Cas12b
CD3
CAR
FITC-A: CD3
PE-A: CAR
Q1 98.7
Q2 0.087
Q3 3.10E-3
Q4 1.21
Q1 11.1
Q2 0.012
Q3 0.037
Q4 88.9
Q1 5.91
Q2 0.016
Q3 0.082
Q4 94.0
Q1 98.7
Q2 0.091
Q3 9.12E-3
Q4 1.19
Q1 6.93
Q2 0.039
Q3 3.94
Q4 89.1
Q1 4.71
Q2 0.048
Q3 4.51
Q4 90.7
Q1 97.3
Q2 0.14
Q3 6.17E-3
Q4 2.61
Q1 4.40
Q2 0.074
Q3 6.20
Q4 89.3
Q1 3.53
Q2 0.089
Q3 6.57
Q4 89.8
Q1 96.7
Q2 0.087
Q3 3.10E-3
Q4 1.21
Q1 8.18
Q2 0.033
Q3 0.45
Q4 91.3
Q1 4.94
Q2 9.60E-3
Q3 0.85
Q4 94.2

图27

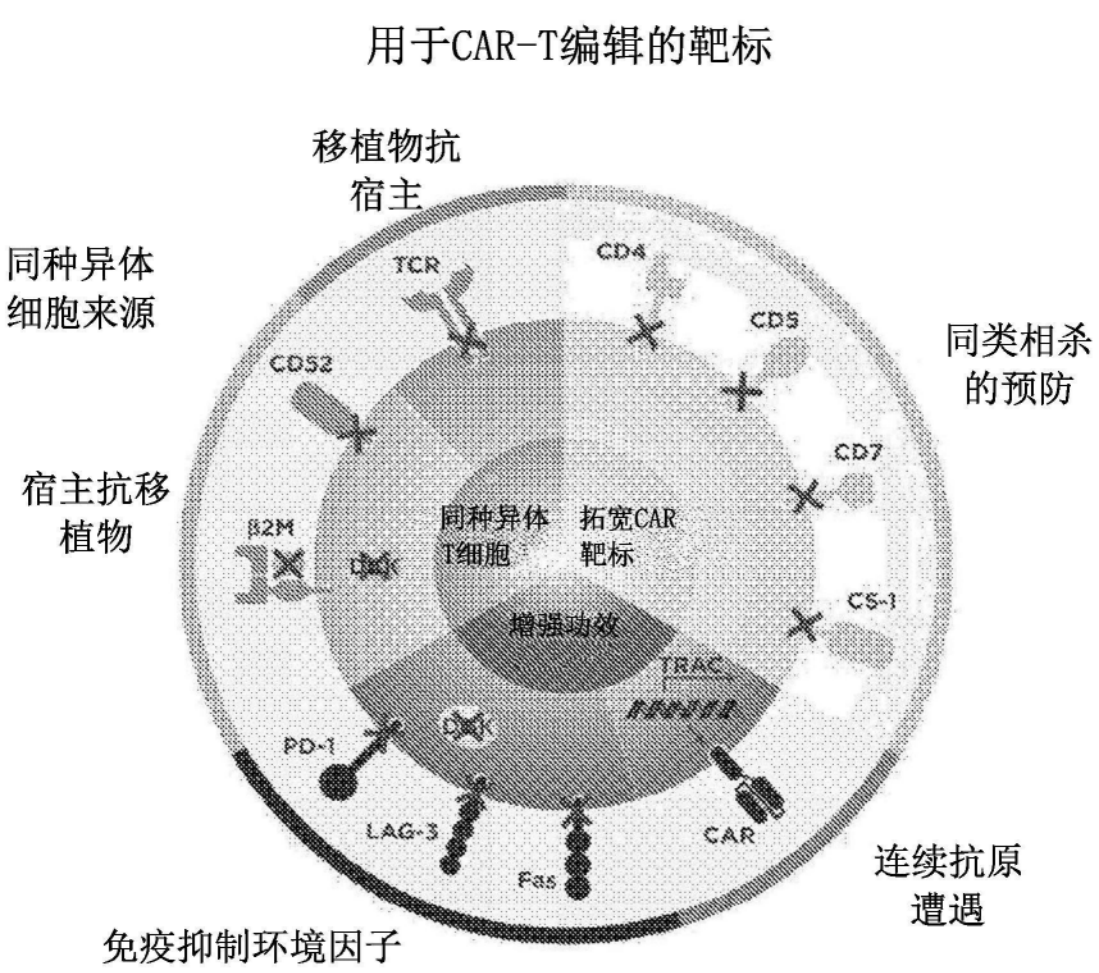

用于CAR-T编辑的靶标
移植物抗
宿主
同种异体
细胞来源
同类相杀
的预防
宿主抗移
植物
连续抗原
遭遇
免疫抑制环境因子
TCR
CD4
CD5
CD52
CD7
CS-1
β2M
同种异体
T细胞
拓宽CAR
靶标
增强功效
TRAC
PD-1
LAG-3
Fas
CAR

图28A

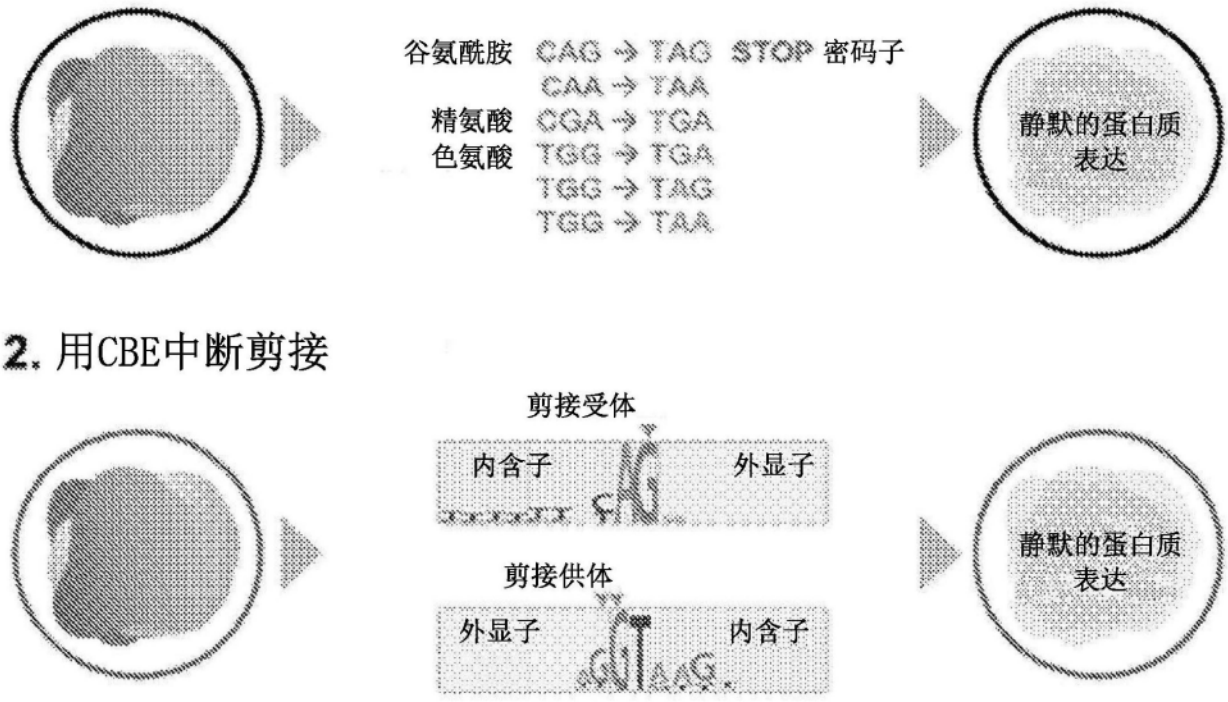

使用碱基编辑器进行静默的两个策略
1. 用CBE创建终止密码子
谷氨酰胺 CAG → TAG STOP 密码子
CAA → TAA
精氨酸 CGA → TGA
色氨酸 TGG → TGA
TGG → TAG
TGG → TAA
静默的蛋白质表达
2. 用CBE中断剪接
剪接受体
内含子
外显子
剪接供体
外显子
内含子
静默的蛋白质表达

图28B

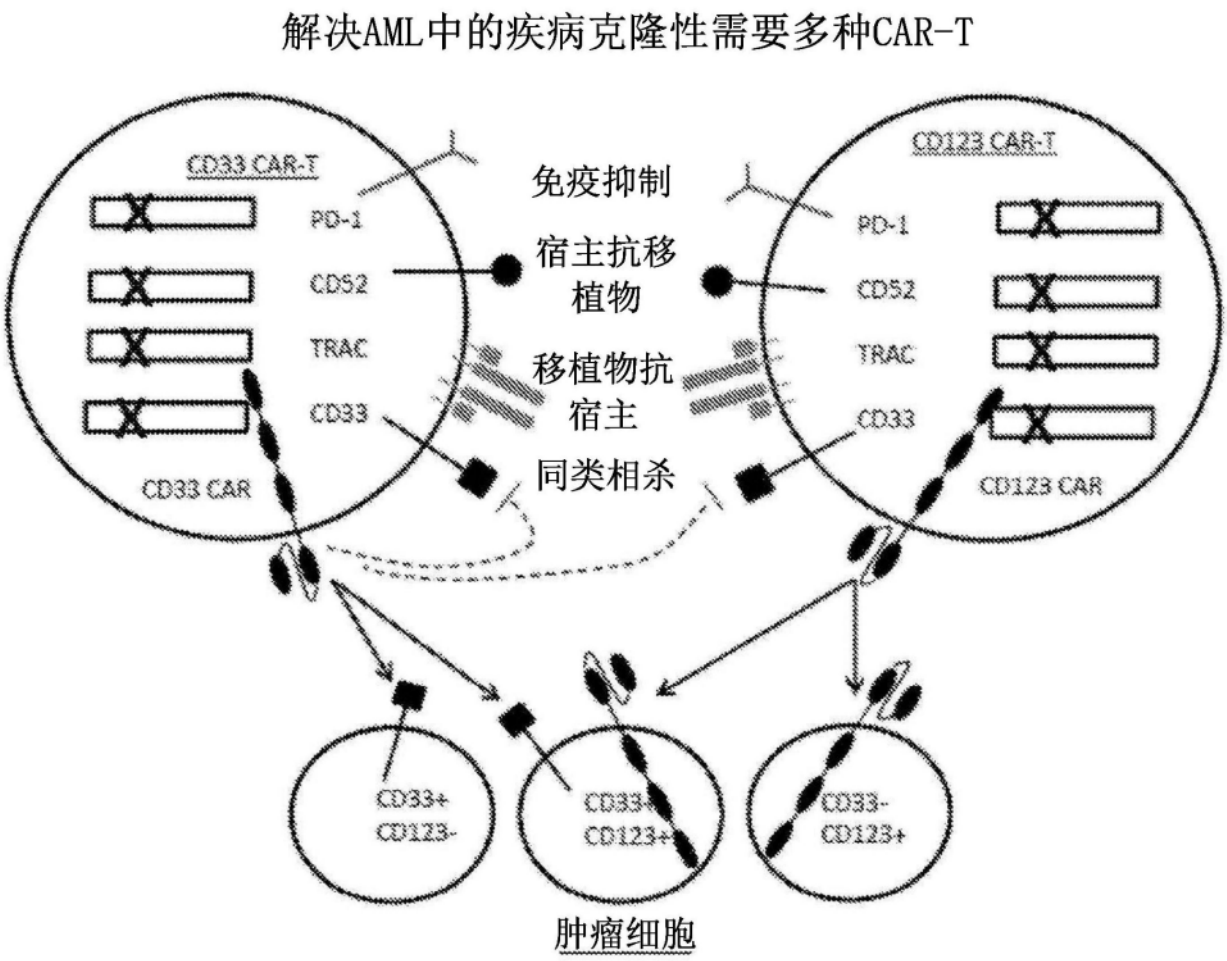

解决AML中的疾病克隆性需要多种CAR-T
CD33 CAR-T
PD-1
CD52
TRAC
CD33
CD33 CAR
免疫抑制
宿主抗移植物
移植物抗宿主
同类相杀
CD123 CAR-T
PD-1
CD52
TRAC
CD33
CD123 CAR
CD33+ CD123-
CD33+ CD123+
CD33- CD123+
肿瘤细胞

图29

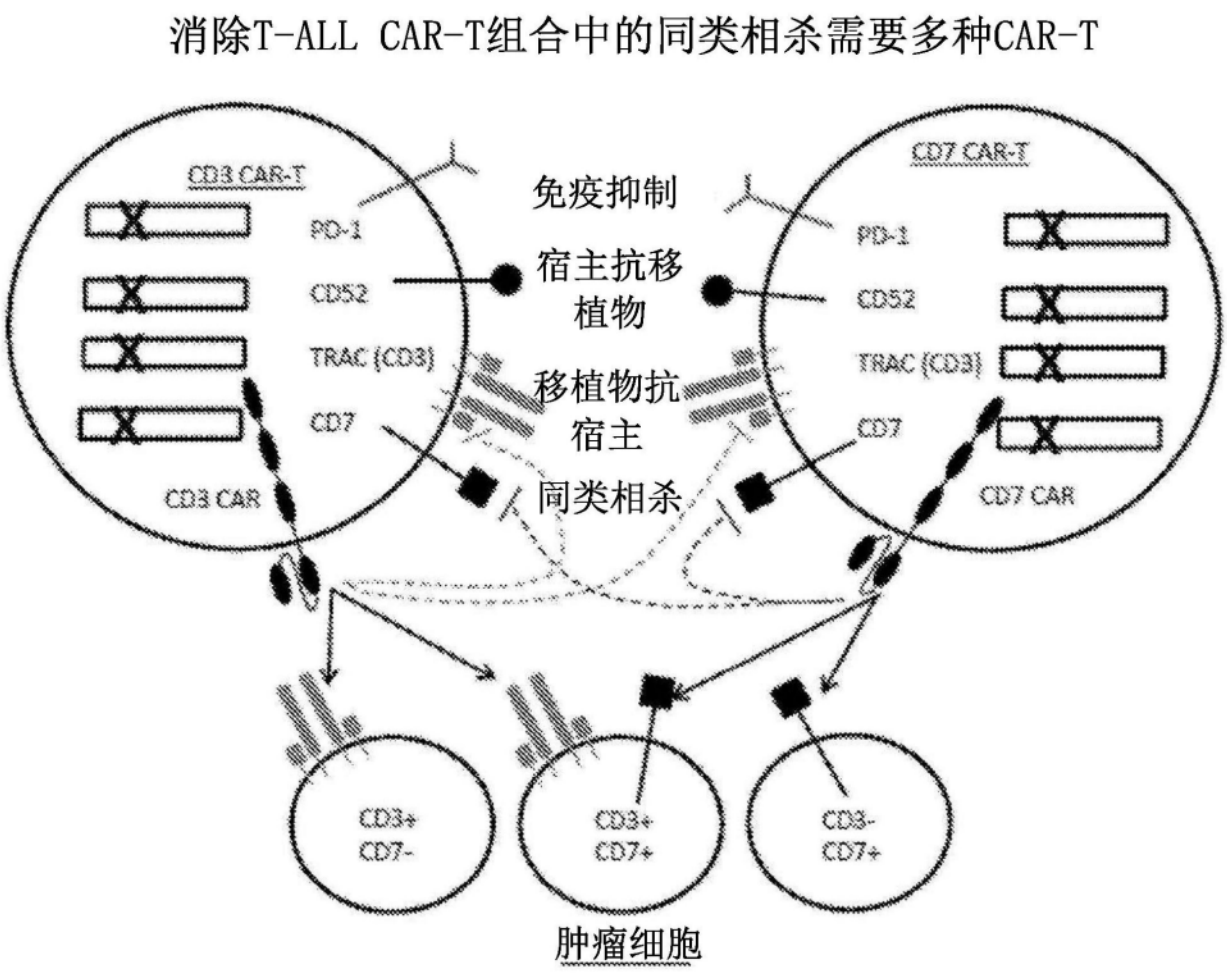

消除T-ALL CAR-T组合中的同类相杀需要多种CAR-T
CD3 CAR-T
PD-1
CD52
TRAC (CD3)
CD7
CD3 CAR
免疫抑制
宿主抗移植物
移植物抗宿主
同类相杀
CD7 CAR-T
PD-1
CD52
TRAC (CD3)
CD7
CD7 CAR
CD3+ CD7-
CD3+ CD7+
CD3- CD7+
肿瘤细胞

图30

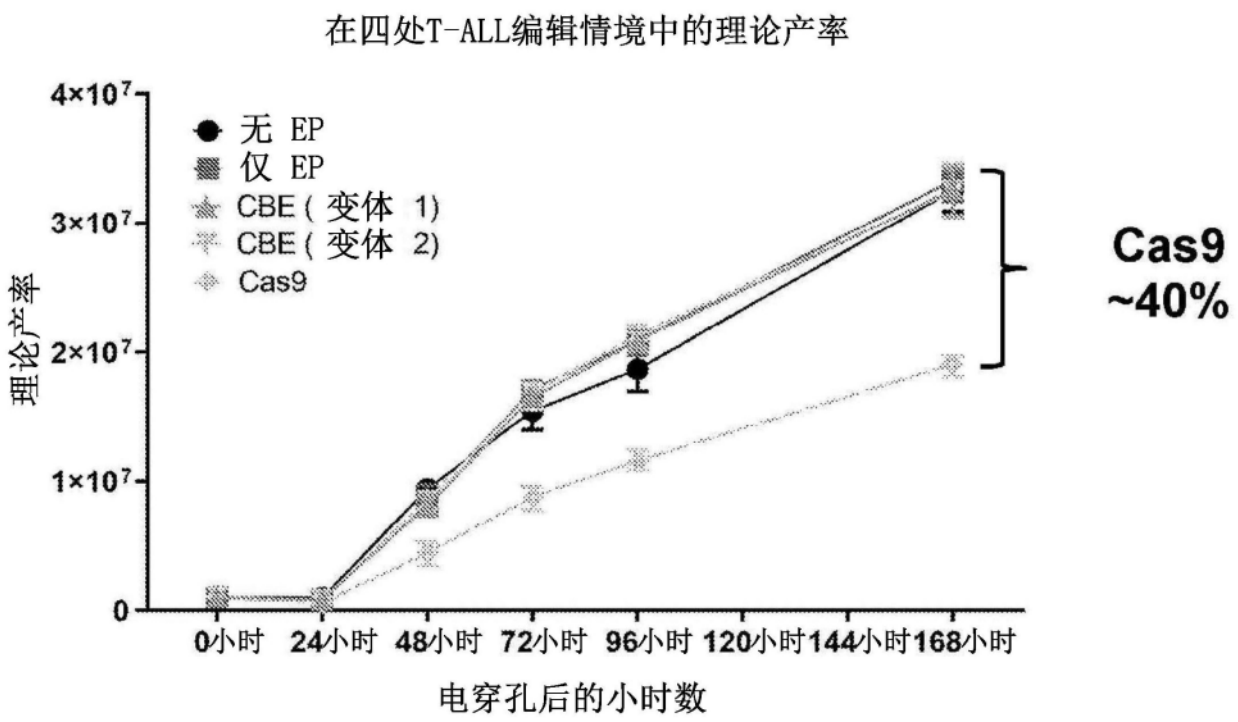

在四处T-ALL编辑情境中的理论产率
理论产率
4×10⁷
3×10⁷
2×10⁷
1×10⁷
0
无 EP
仅 EP
CBE (变体 1)
CBE (变体 2)
Cas9
Cas9
~40%
0小时
24小时
48小时
72小时
96小时
120小时
144小时
168小时
电穿孔后的小时数

图31A

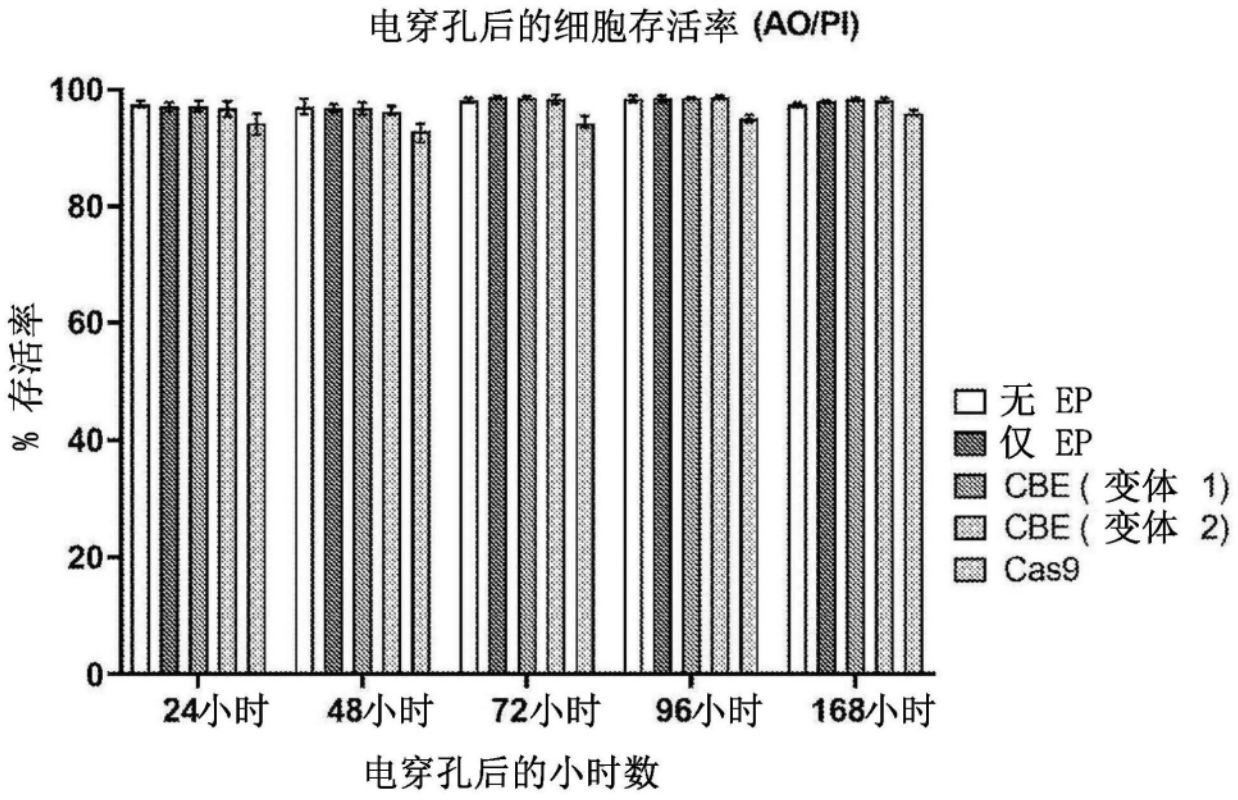

电穿孔后的细胞存活率 (AO/PI)
% 存活率
100
80
60
40
20
0
无 EP
仅 EP
CBE (变体 1)
CBE (变体 2)
Cas9
24小时
48小时
72小时
96小时
168小时
电穿孔后的小时数

图31B

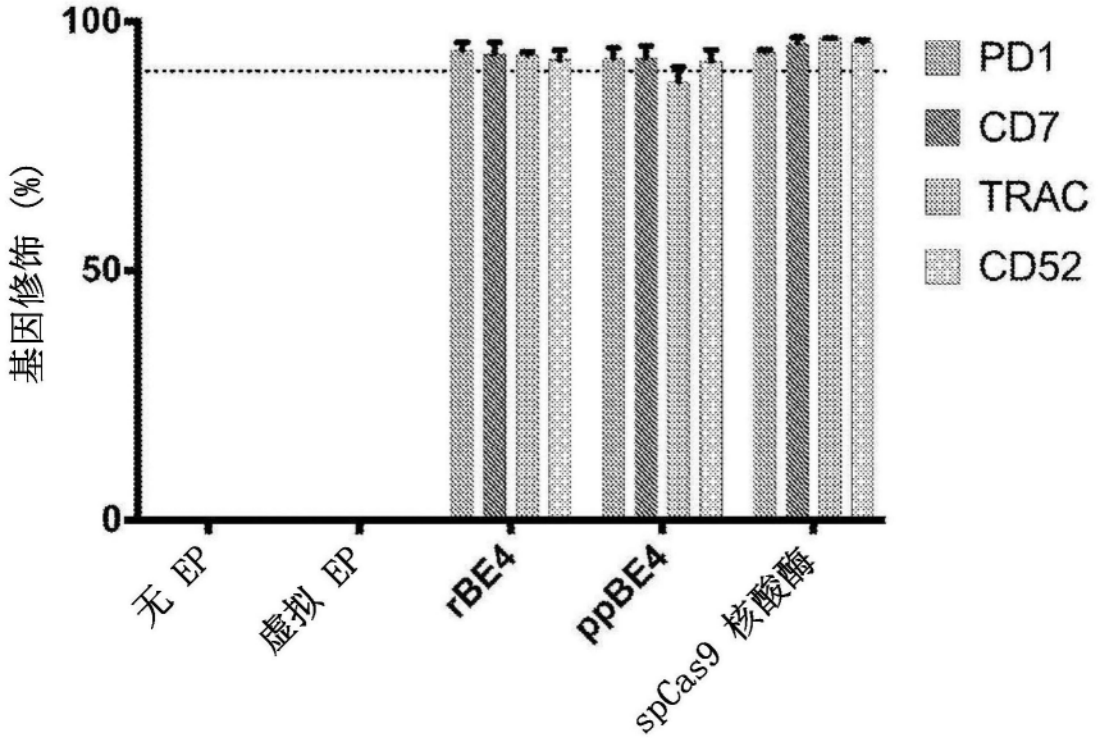

100
50
0
基因修饰（%）
PD1
CD7
TRAC
CD52
无 EP
虚拟 EP
rBE4
ppBE4
spCas9 核酸酶

图32

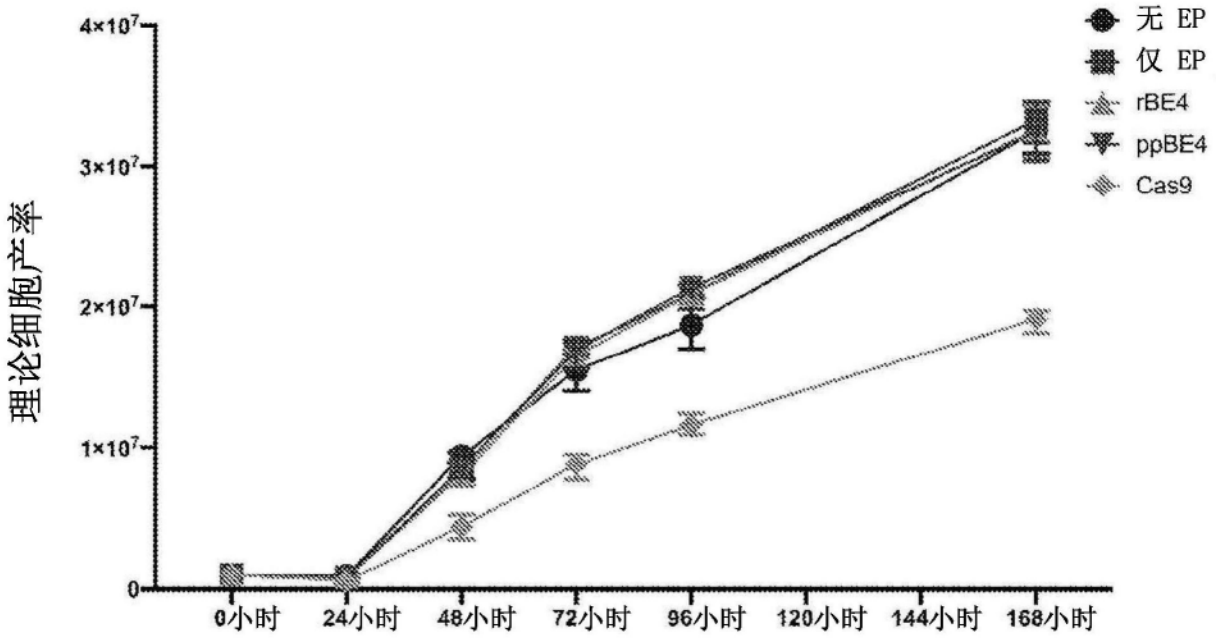

4×10⁷
3×10⁷
2×10⁷
1×10⁷
0
理论细胞产率
无 EP
仅 EP
rBE4
ppBE4
Cas9
0小时
24小时
48小时
72小时
96小时
120小时
144小时
168小时

图33

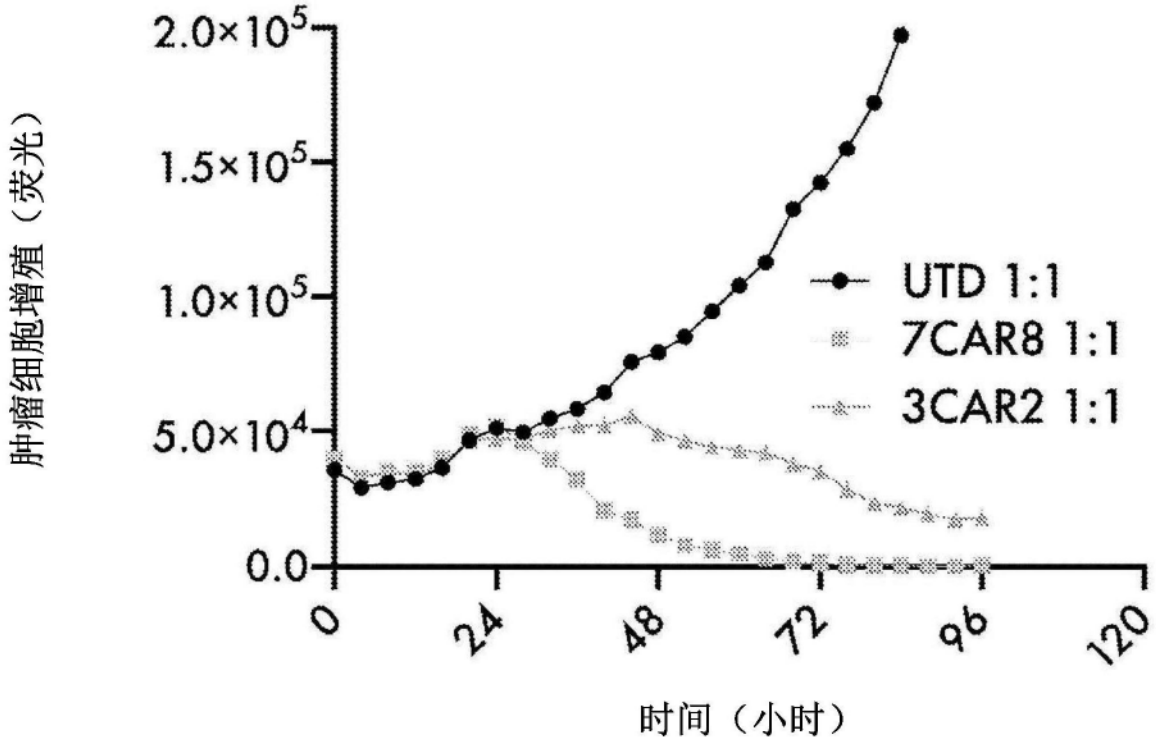

肿瘤细胞增殖（荧光）
2.0×10^5
1.5×10^5
1.0×10^5
5.0×10^4
0.0
0
24
48
72
96
120
时间（小时）
UTD 1:1
7CAR8 1:1
3CAR2 1:1

图34

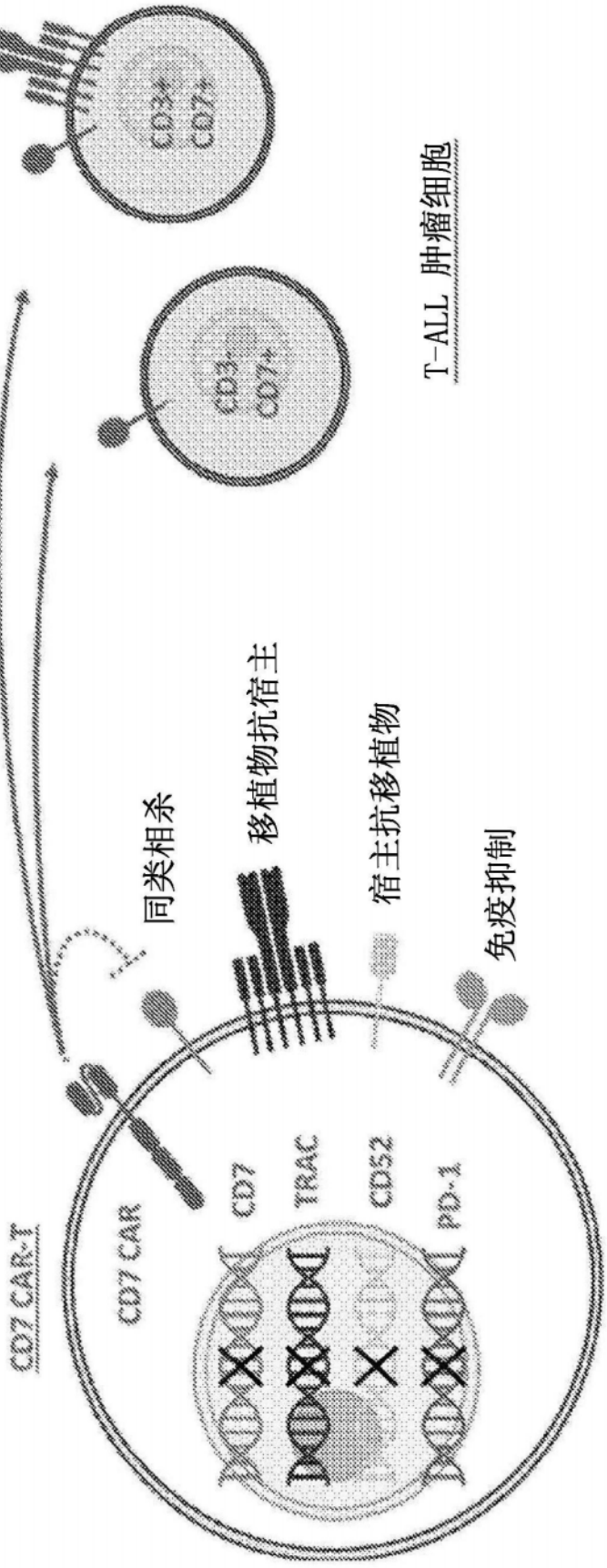

CD3+
CD7+
CD3-
CD7+
T-ALL 肿瘤细胞
同类相杀
移植物抗宿主
宿主抗移植物
免疫抑制
CD7 CAR-T
CD7 CAR
CD7
TRAC
CD52
PD-1

图35

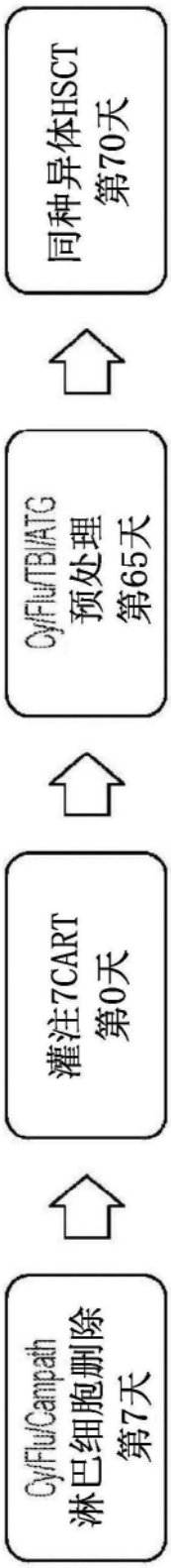

Cy/Flu/Campath
淋巴细胞删除
第7天
灌注7CART
第0天
Cy/Flu/TBI/ATG
预处理
第65天
同种异体HSCT
第70天

图36

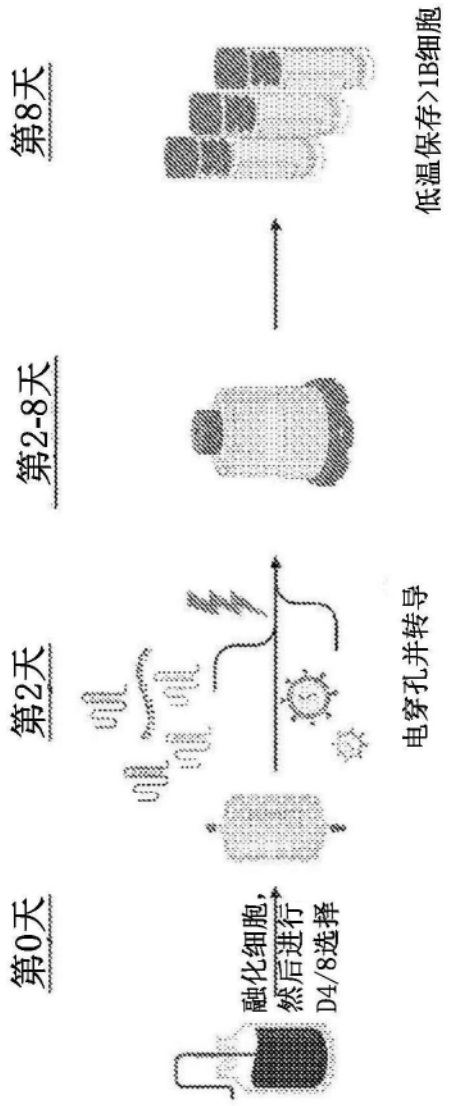


图37A

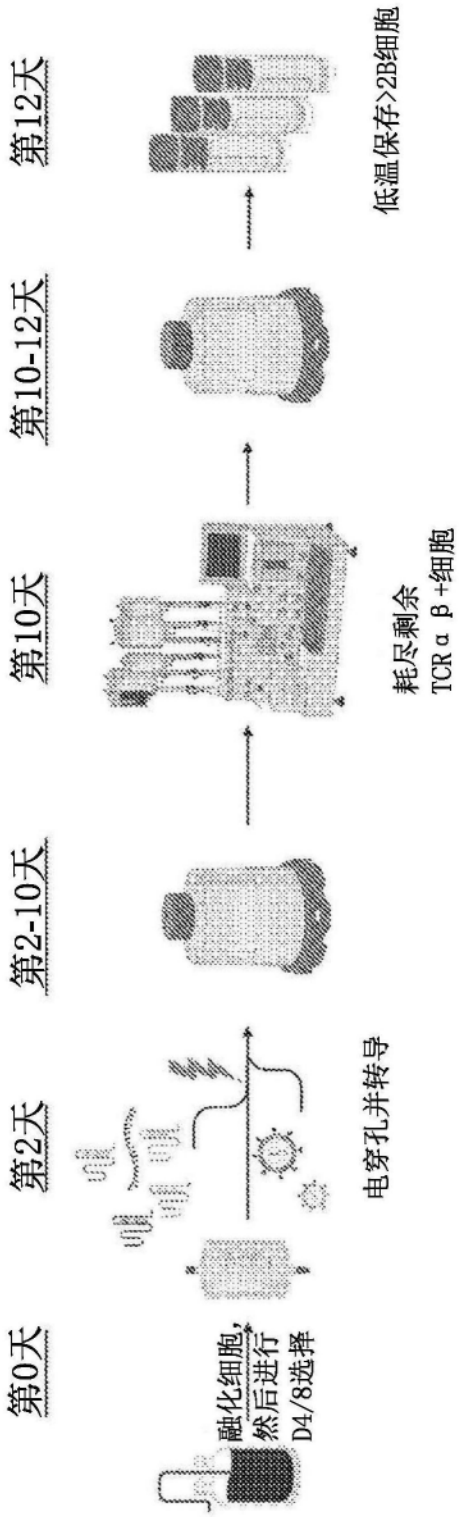

第0天
第2天
第2-10天
第10天
第10-12天
第12天
融化细胞，然后进行D4/8选择
电穿孔并转导
耗尽剩余TCRα β+细胞
低温保存>2B细胞

图37B

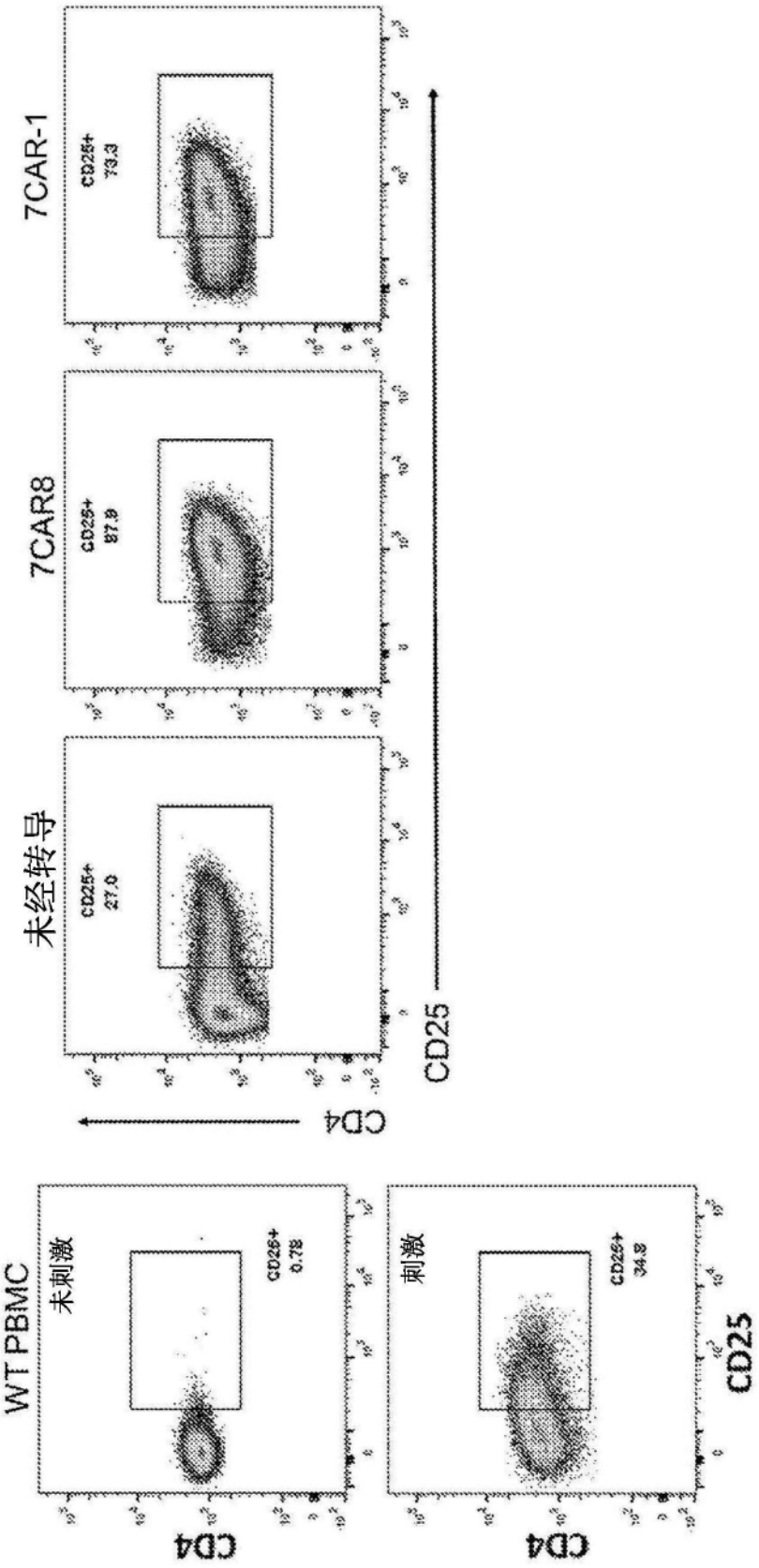

WT PBMC
未刺激
CD25+
0.78
刺激
CD25+
34.8
CD4
CD25
未经转导
CD25+
27.0
7CAR8
CD25+
87.9
7CAR-1
CD25+
73.3
CD4
CD25

图37C

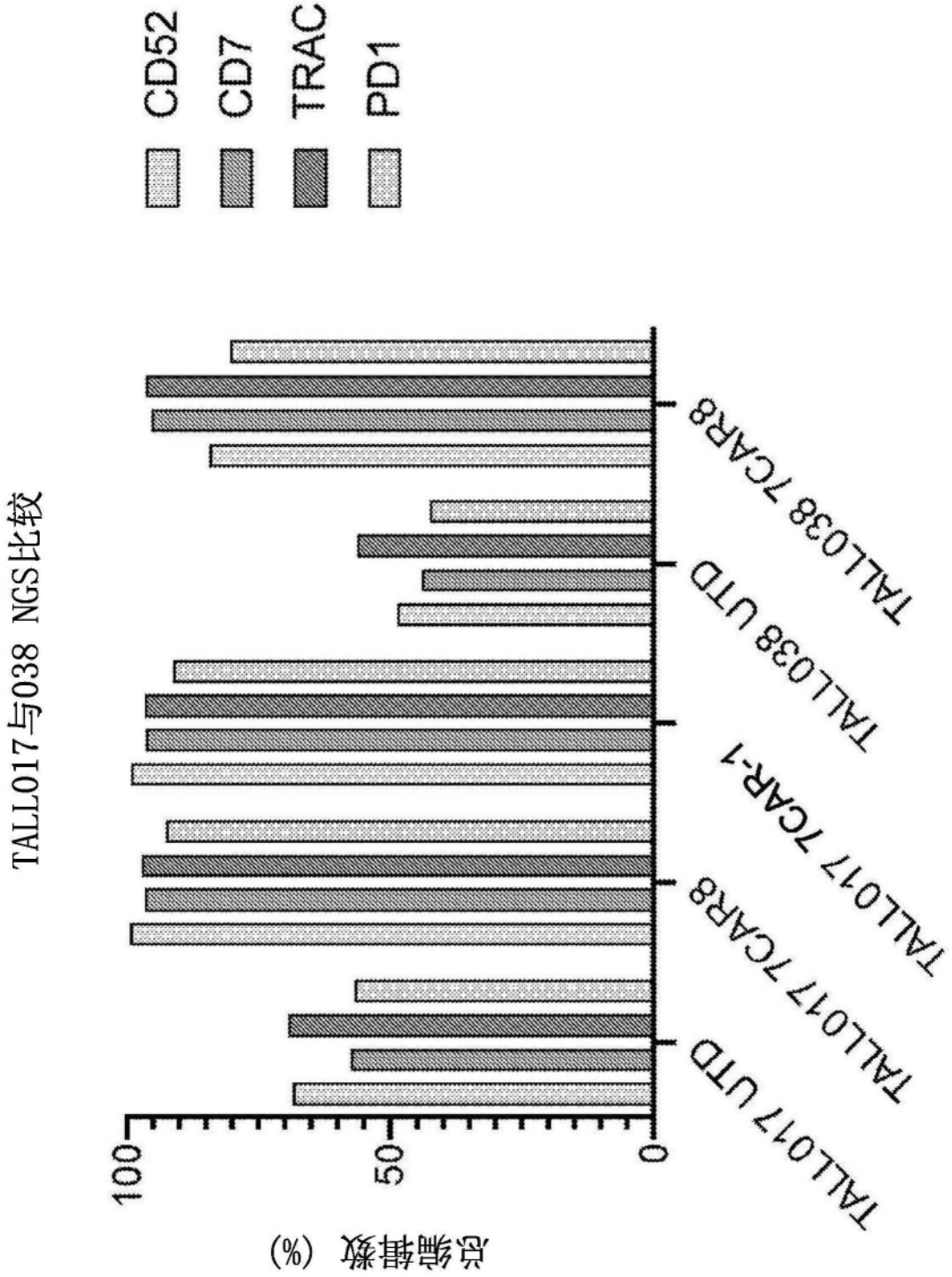

CD52
CD7
TRAC
PD1
TALL017与038 NGS比较
总编辑数 (%)
100
50
0
TALL017 UTD
TALL017 7CAR8
TALL017 7CAR-1
TALL038 UTD
TALL038 7CAR8

图38

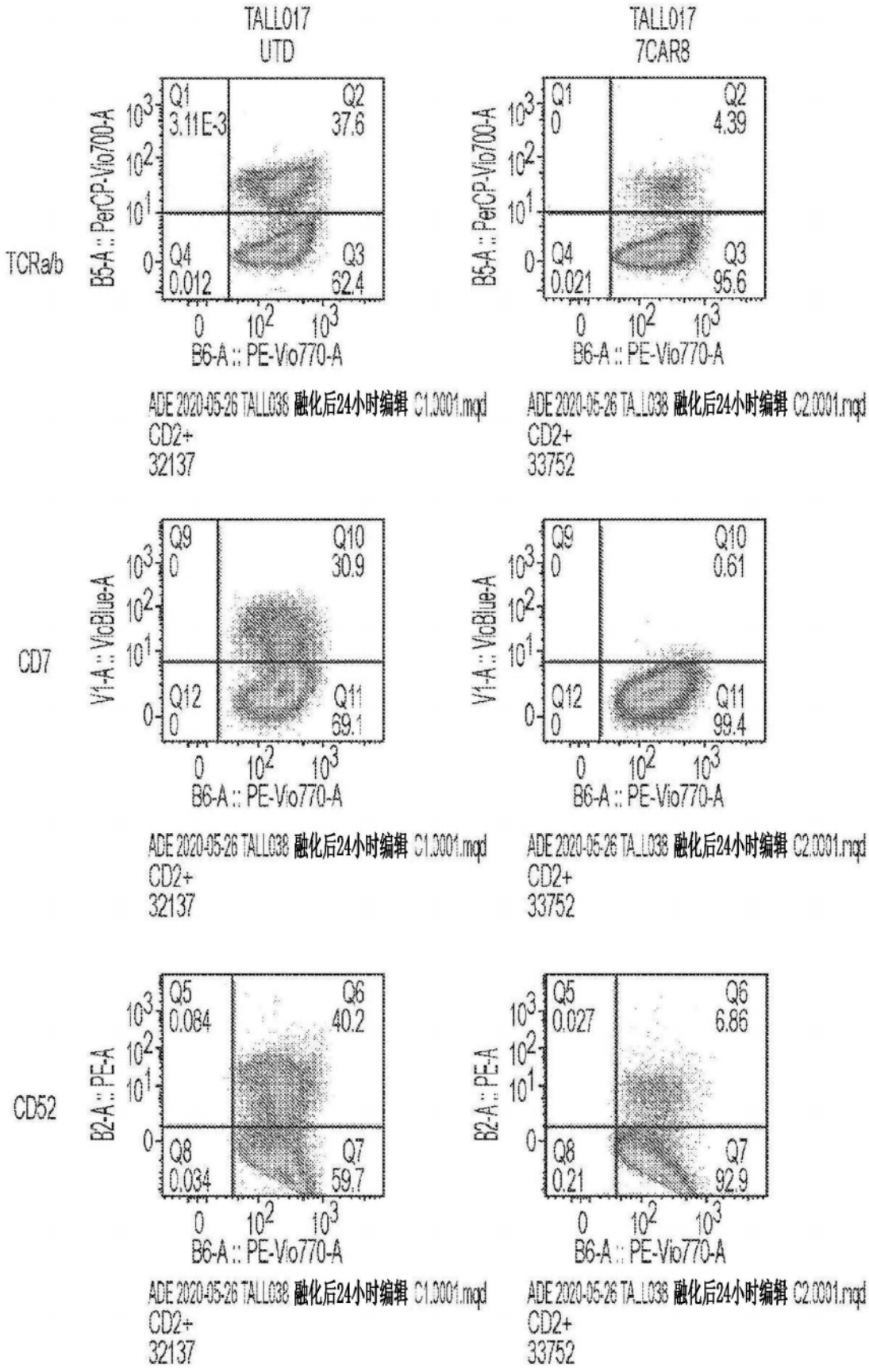

TALL017
UTD
TALL017
7CAR8
TCRa/b
CD7
CD52
B5-A :: PerCP-Vio700-A
B6-A :: PE-Vio770-A
V1-A :: VioBlue-A
B2-A :: PE-A
Q1 3.11E-3
Q2 37.6
Q4 0.012
Q3 62.4
Q1 0
Q2 4.39
Q4 0.021
Q3 95.6
Q9 0
Q10 30.9
Q12 0
Q11 69.1
Q9 0
Q10 0.61
Q12 0
Q11 99.4
Q5 0.084
Q6 40.2
Q8 0.034
Q7 59.7
Q5 0.027
Q6 6.86
Q8 0.21
Q7 92.9
ADE 2020-05-26 TALL038 融化后24小时编辑 C1.0001.mqd
CD2+
32137
ADE 2020-05-26 TA_L038 融化后24小时编辑 C2.0001.mqd
CD2+
33752

图39A

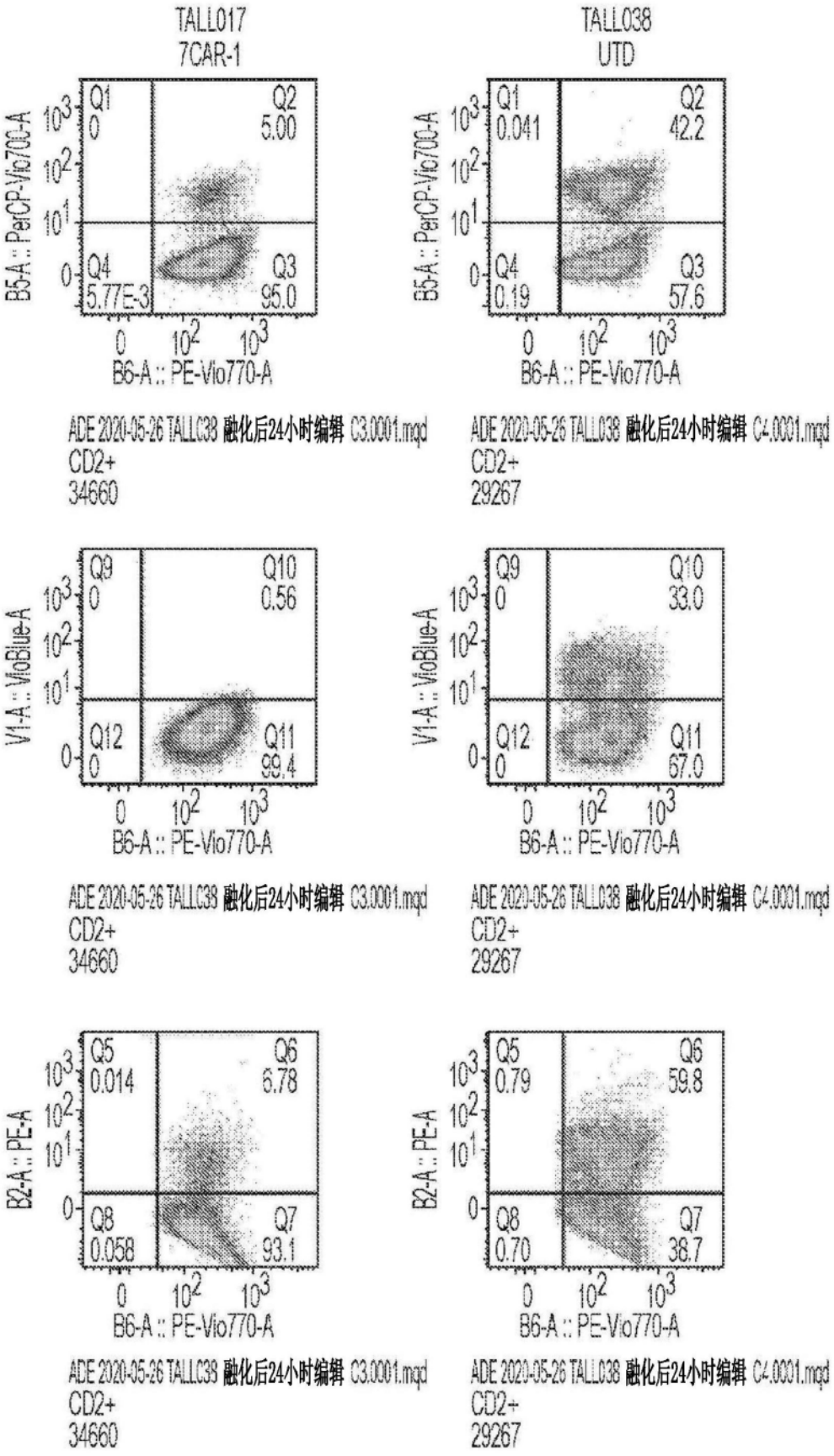

TALL017
7CAR-1
TALL038
UTD
B5-A :: PerCP-Vio700-A
B6-A :: PE-Vio770-A
Q1 0
Q2 5.00
Q4 5.77E-3
Q3 95.0
Q1 0.041
Q2 42.2
Q4 0.19
Q3 57.6
ADE 2020-05-26 TALL038 融化后24小时编辑 C3.0001.mqd
CD2+
34660
ADE 2020-05-26 TALL038 融化后24小时编辑 C4.0001.mqd
CD2+
29267
V1-A :: VioBlue-A
Q9 0
Q10 0.56
Q12 0
Q11 99.4
Q9 0
Q10 33.0
Q12 0
Q11 67.0
B2-A :: PE-A
Q5 0.014
Q6 6.78
Q8 0.058
Q7 93.1
Q5 0.79
Q6 59.8
Q8 0.70
Q7 38.7

图39A(续)

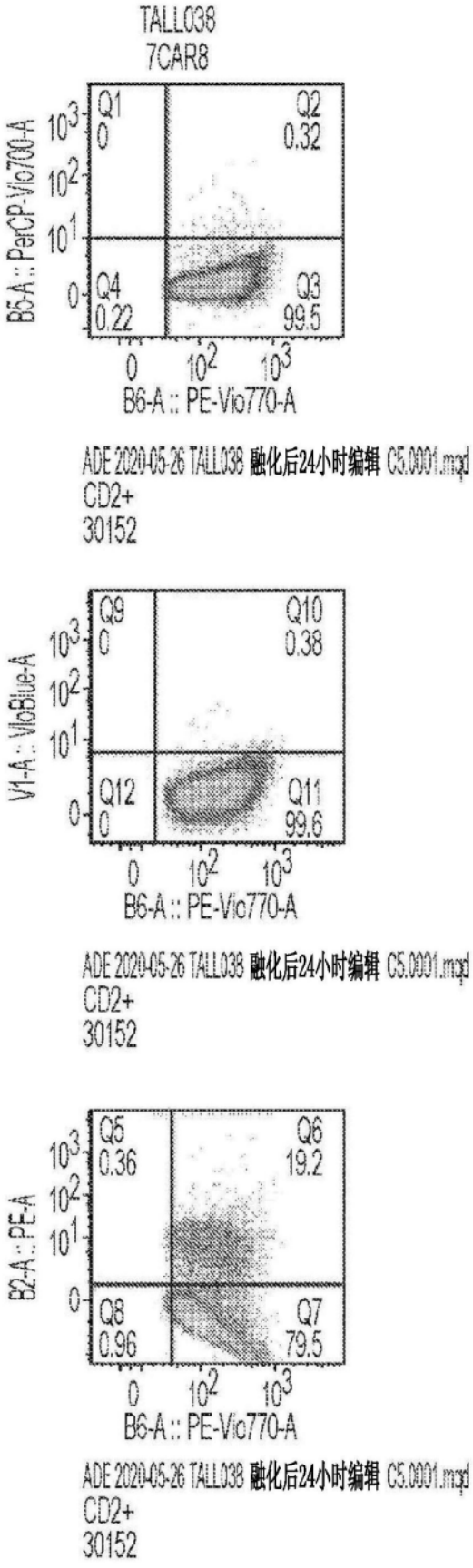

TALL038
7CAR8
Q1
0
Q2
0.32
Q4
0.22
Q3
99.5
B5-A :: PerCP-Vio700-A
B6-A :: PE-Vio770-A
ADE 2020-05-26 TALL038 融化后24小时编辑 C5.0001.mopl
CD2+
30152
Q9
0
Q10
0.38
Q12
0
Q11
99.6
V1-A :: VioBlue-A
B6-A :: PE-Vio770-A
ADE 2020-05-26 TALL038 融化后24小时编辑 C5.0001.mopl
CD2+
30152
Q5
0.36
Q6
19.2
Q8
0.96
Q7
79.5
B2-A :: PE-A
B6-A :: PE-Vio770-A
ADE 2020-05-26 TALL038 融化后24小时编辑 C5.0001.mopl
CD2+
30152

图39A(续)

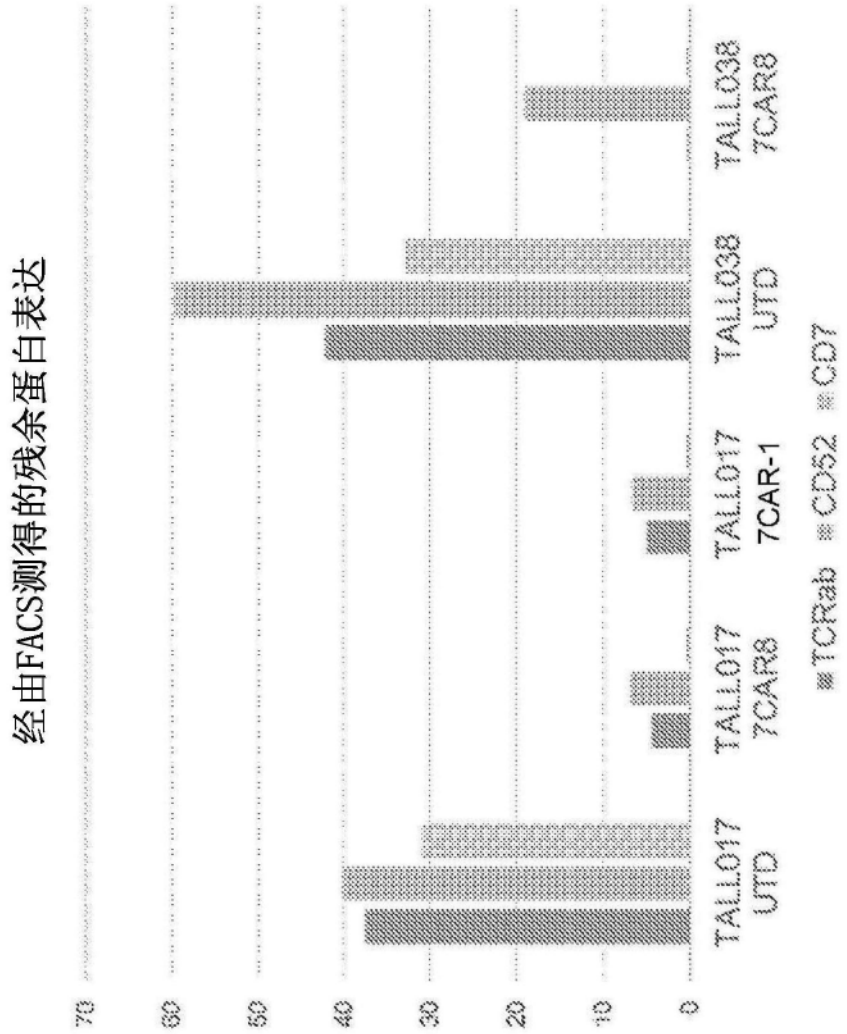

经由FACS测得的残余蛋白表达
70
60
50
40
30
20
10
0
TALL017 UTD
TALL017 7CAR8
TALL017 7CAR-1
TALL038 UTD
TALL038 7CAR8
TCRab
CD52
CD7

图39B

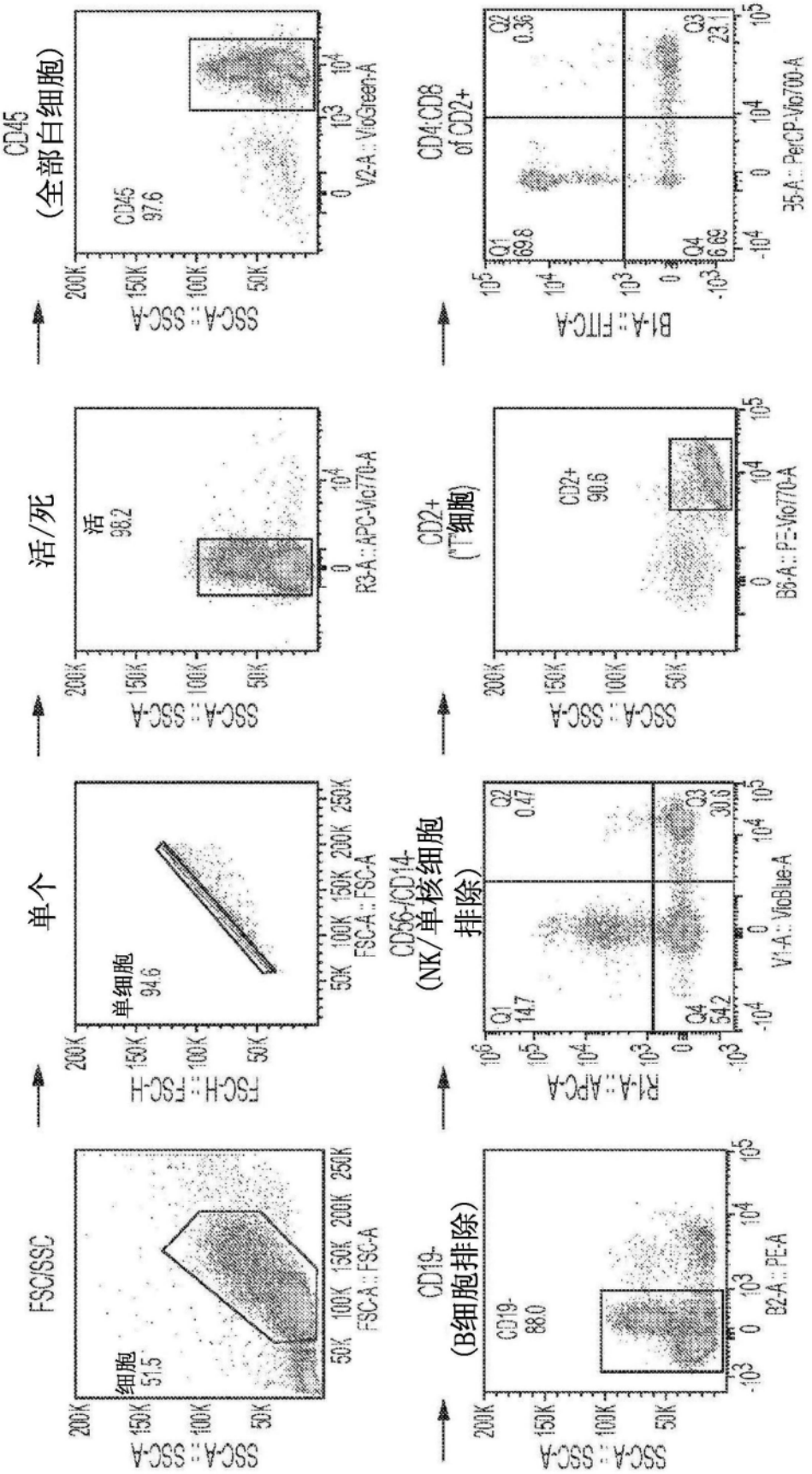

FSC/SSC
细胞
51.5
SSC-A :: SSC-A
FSC-A :: FSC-A
单个
单细胞
94.6
FSC-H :: FSC-A
FSC-A :: FSC-A
活/死
活
98.2
SSC-A :: SSC-A
R3-A :: APC-Vio770-A
CD45
(全部白细胞)
CD45
97.6
SSC-A :: SSC-A
V2-A :: VioGreen-A
CD19-
(B细胞排除)
CD19-
88.0
SSC-A :: SSC-A
B2-A :: PE-A
CD56-/CD14-
(NK/单核细胞
排除)
Q1
14.7
Q2
0.47
Q4
54.2
Q3
30.6
R1-A :: APC-A
V1-A :: VioBlue-A
CD2+
("T"细胞)
CD2+
90.6
SSC-A :: SSC-A
B6-A :: PE-Vio770-A
CD4:CD8
of CD2+
Q1
69.8
Q2
0.36
Q4
6.69
Q3
23.1
B1-A :: FITC-A
35-A :: PerCP-Vio700-A

图40

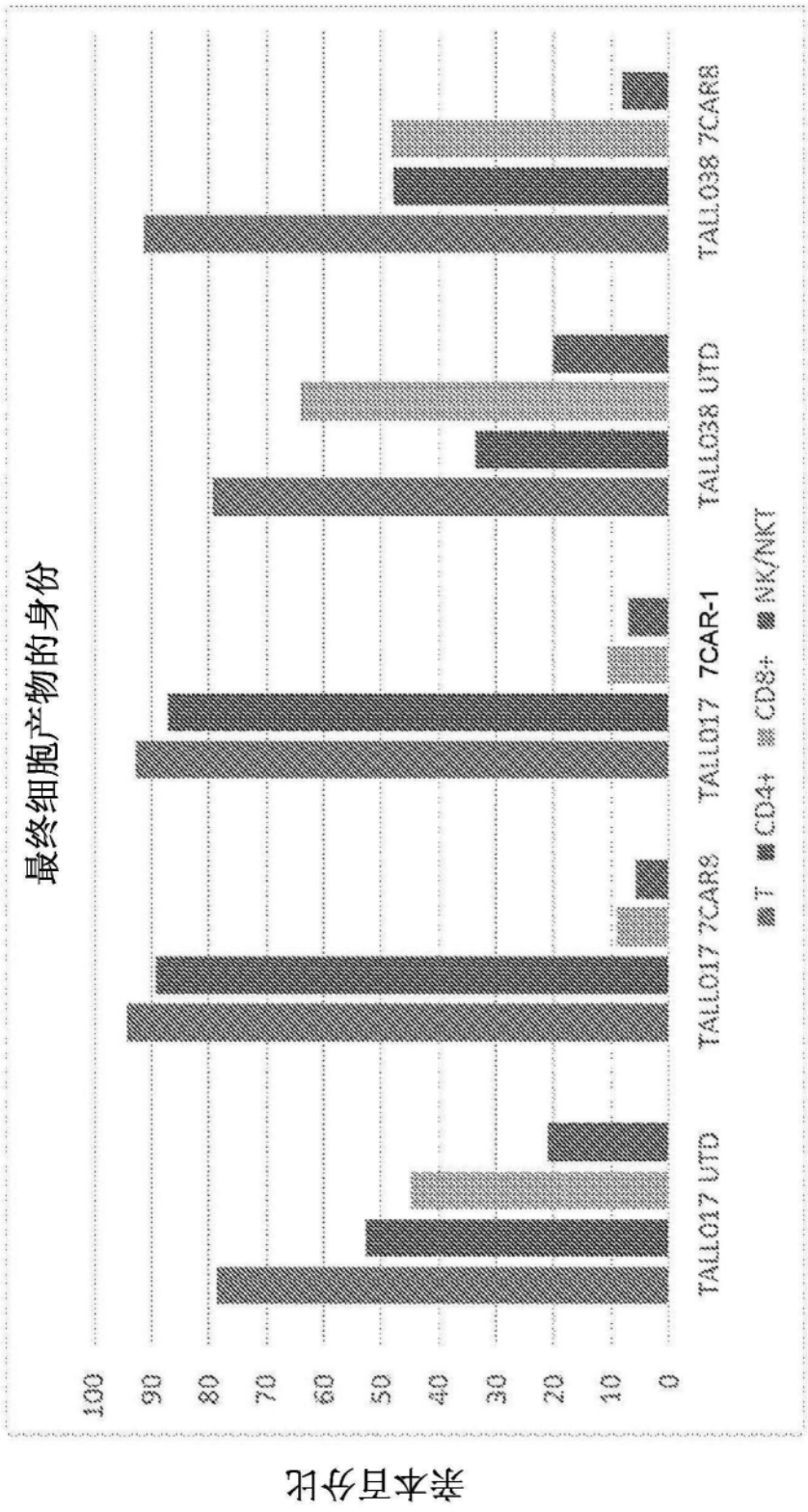

最终细胞产物的身份
样本百分比
100
90
80
70
60
50
40
30
20
10
0
TALL017 UTD
TALL017 7CAR8
TALL017 7CAR-1
TALL038 UTD
TALL038 7CAR8
T
CD4+
CD8+
NK/NKT

图41

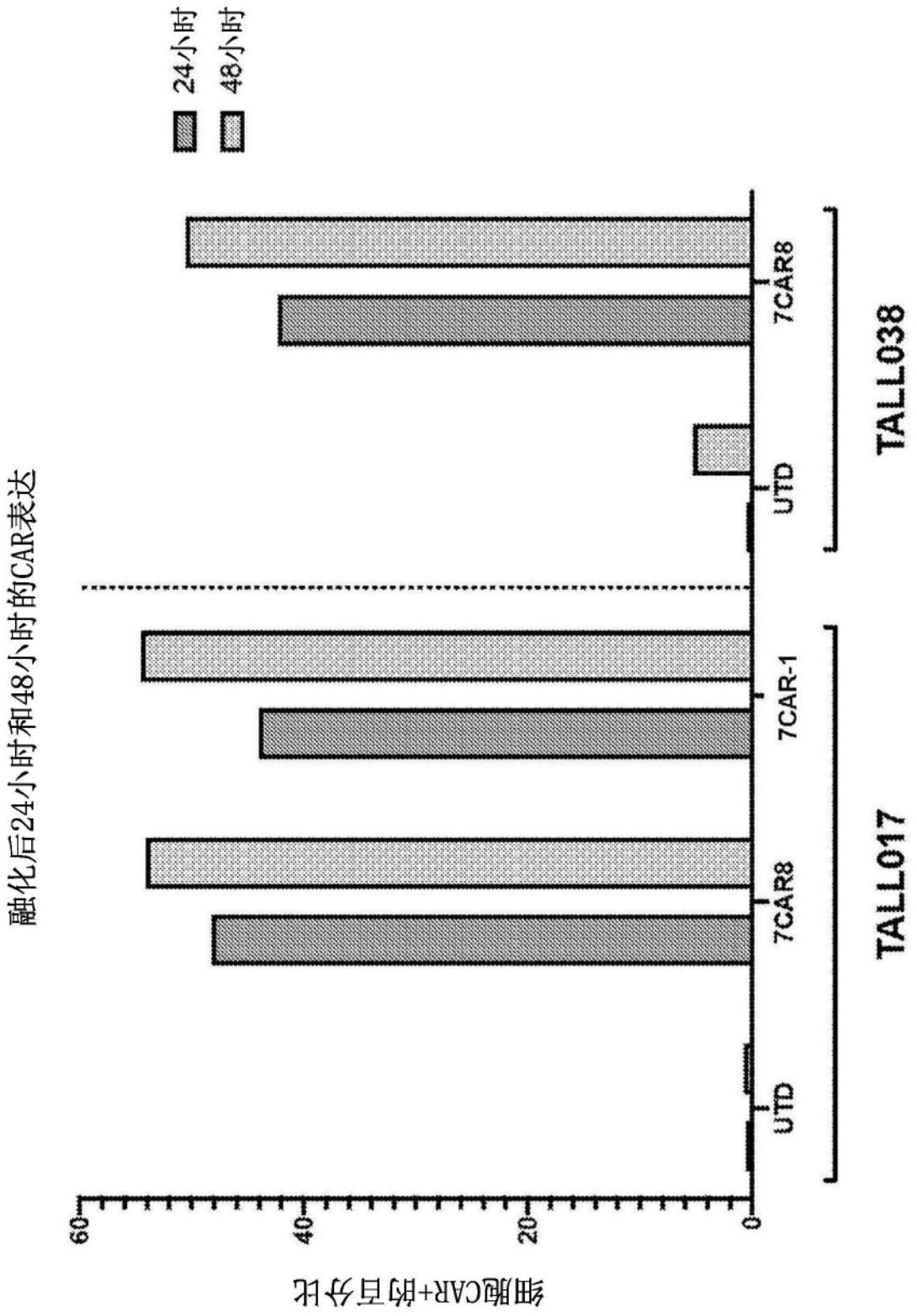

融化后24小时和48小时的CAR表达
24小时
48小时
细胞CAR+的百分比
0
20
40
60
UTD
7CAR8
7CAR-1
7CAR8
UTD
TALL017
TALL038

图42

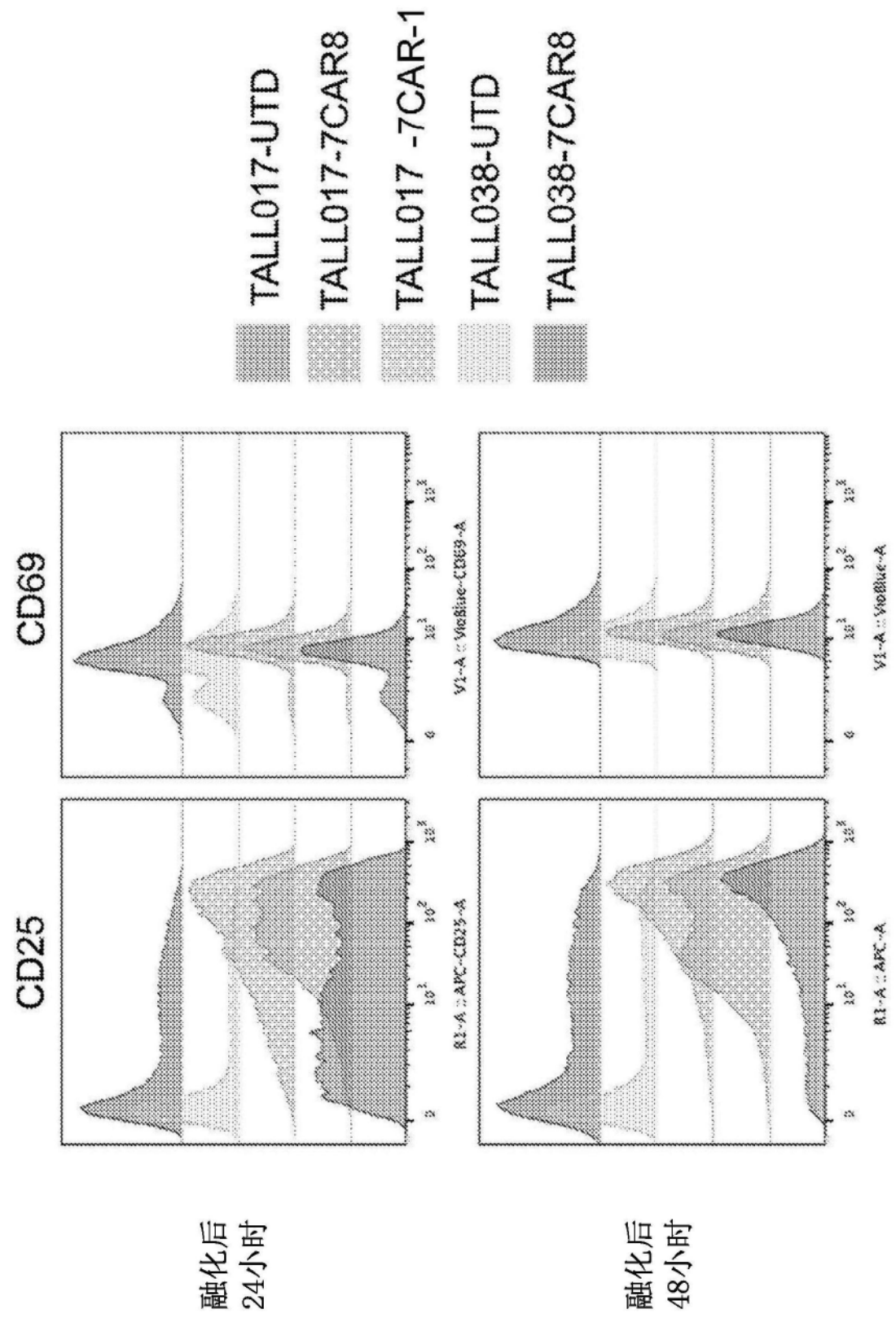

TALL017-UTD
TALL017-7CAR8
TALL017 -7CAR-1
TALL038-UTD
TALL038-7CAR8
CD69
CD25
融化后
24小时
融化后
48小时

图43

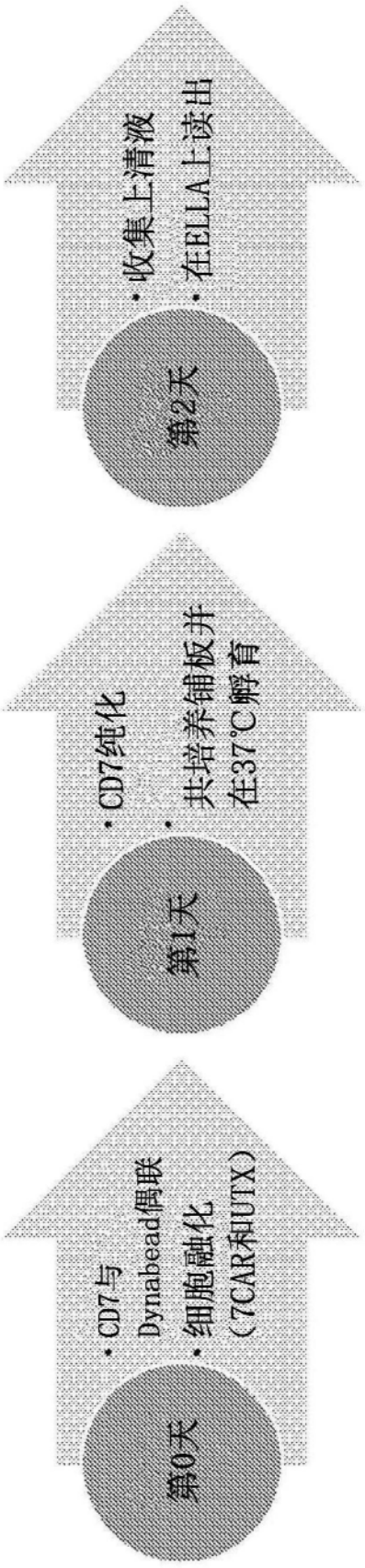

第0天
·CD7与Dynabead偶联
·细胞融化（7CAR和UTX）
第1天
·CD7纯化
·共培养铺板并在37℃孵育
第2天
·收集上清液
·在ELLA上读出

图44

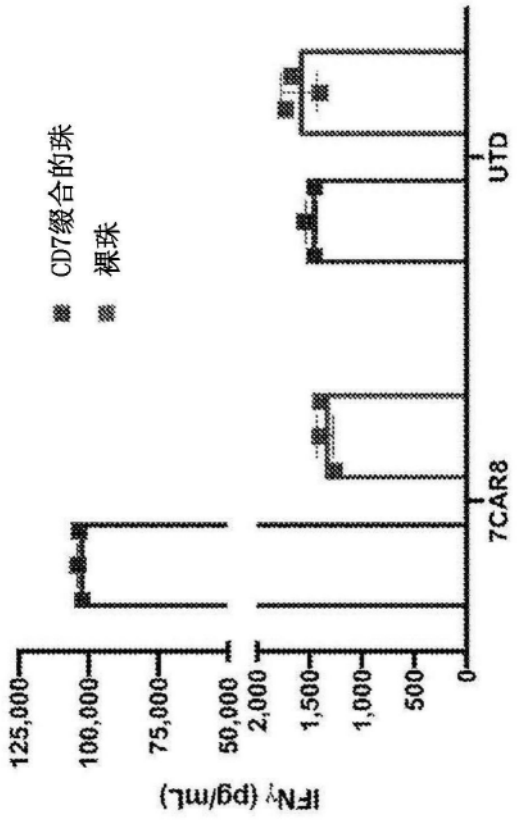

CD7缀合的珠
裸珠
IFNγ (pg/mL)
125,000
100,000
75,000
50,000
2,000
1,500
1,000
500
0
7CAR8
UTD

图45A

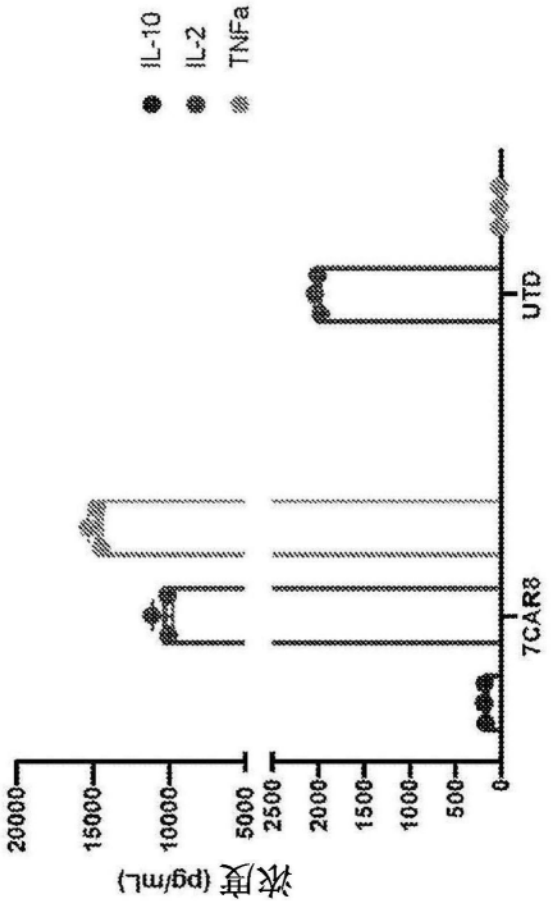

IL-10
IL-2
TNFa
UTD
7CAR8
浓度 (pg/mL)
20000
15000
10000
5000
2500
2000
1500
1000
500
0

图45B

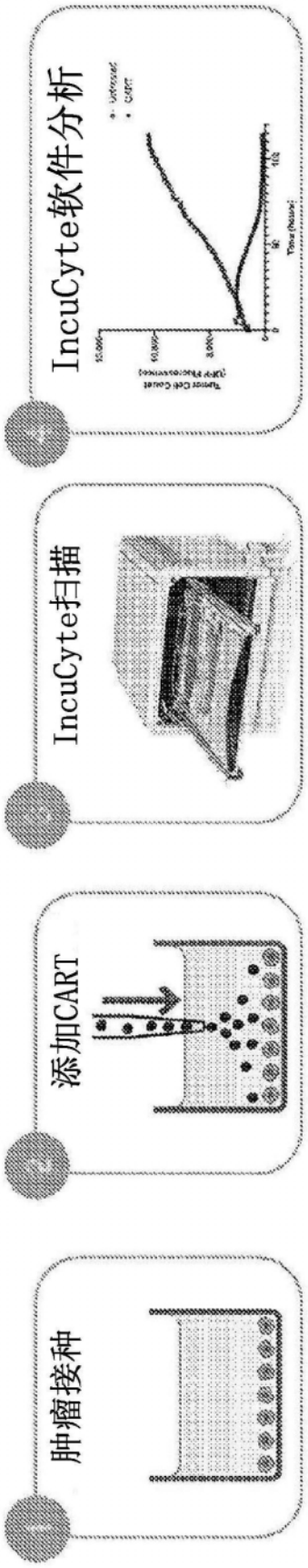


图46

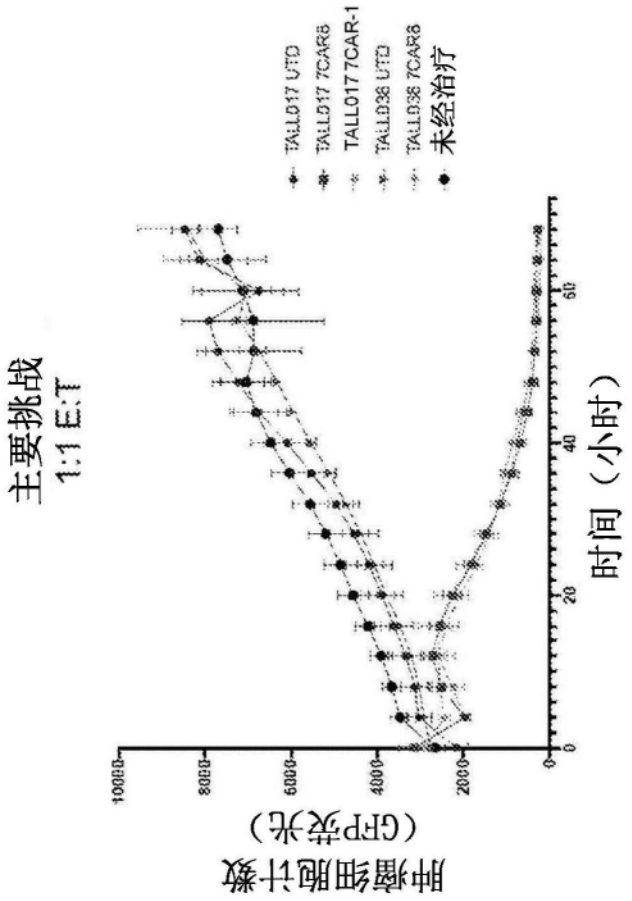

主要挑战
1:1 E:T
肿瘤细胞计数
(GFP荧光)
时间（小时）
TALL017 UTD
TALL017 7CAR8
TALL017 7CAR-1
TALL038 UTD
TALL038 7CAR8
未经治疗

图47A

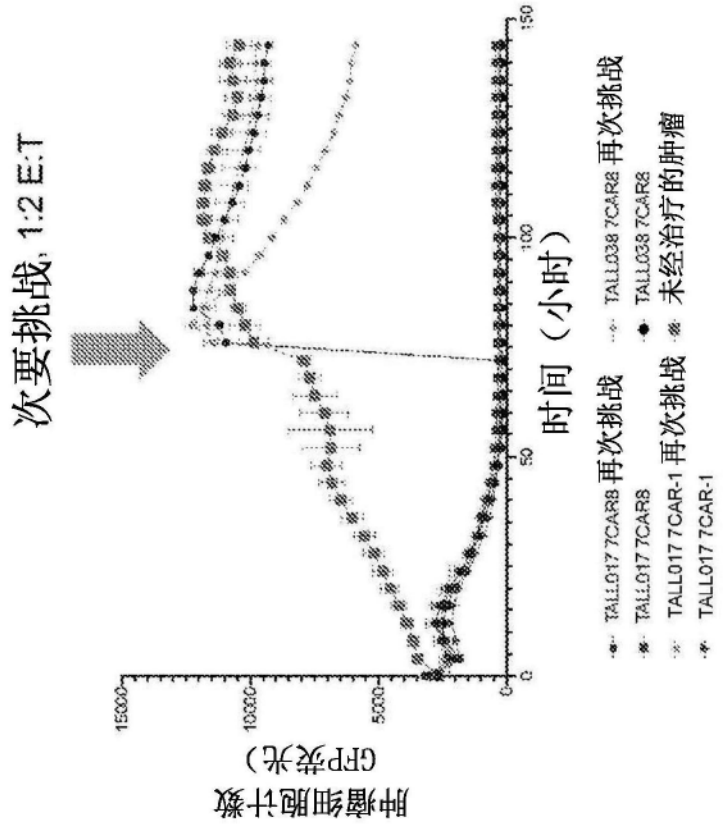

次要挑战, 1:2 E:T
肿瘤细胞计数
(GFP荧光)
15000
10000
5000
0
0
50
100
150
时间（小时）
TALL017 7CAR8 再次挑战
TALL017 7CAR8
TALL017 7CAR-1 再次挑战
TALL017 7CAR-1
TALL038 7CAR8 再次挑战
TALL038 7CAR8
未经治疗的肿瘤

图47B

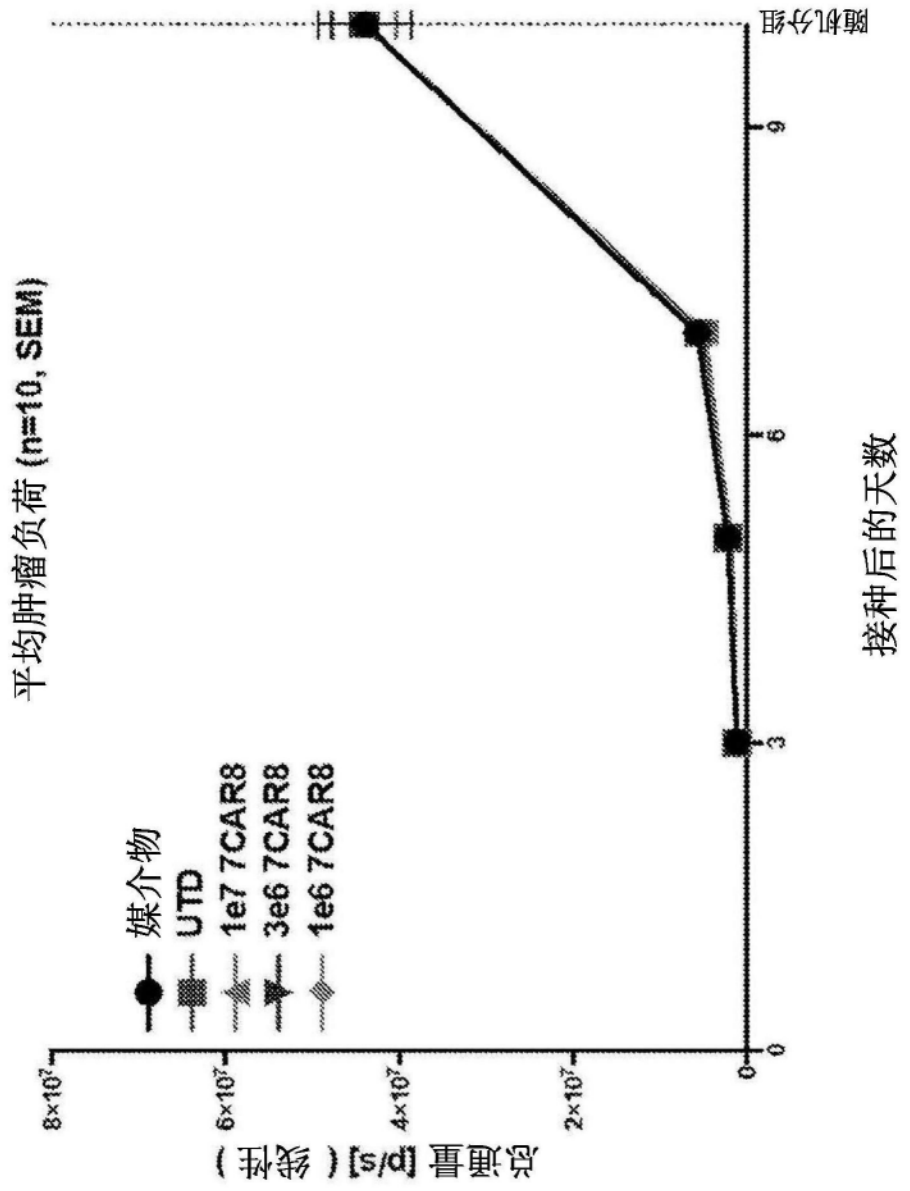

平均肿瘤负荷 (n=10, SEM)
媒介物
UTD
1e7 7CAR8
3e6 7CAR8
1e6 7CAR8
总通量 [p/s] (线性)
8×10⁷
6×10⁷
4×10⁷
2×10⁷
0
0
3
6
9
随机分组
接种后的天数

图48

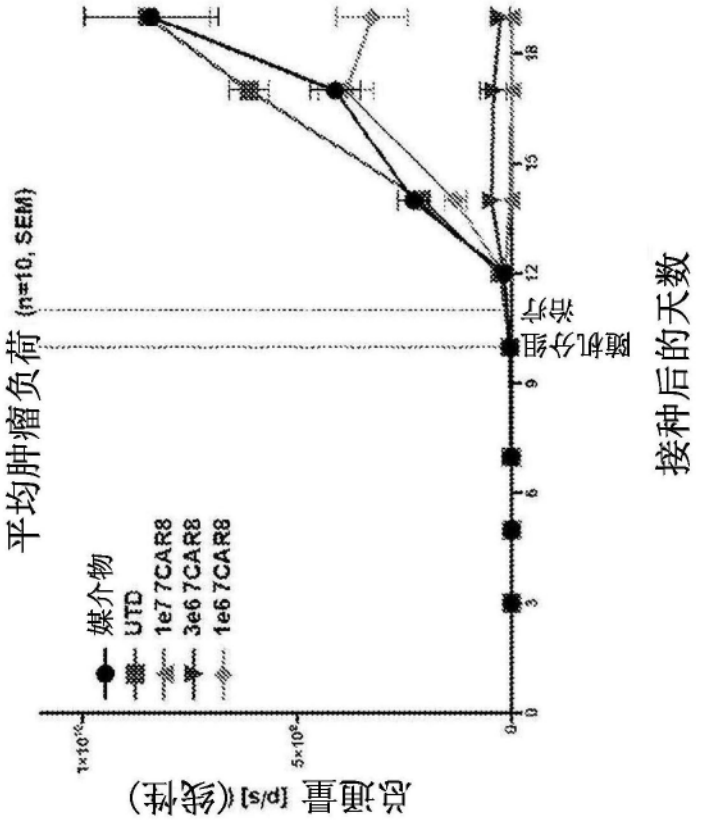

平均肿瘤负荷 (n=10, SEM)
媒介物
UTD
1e7 7CAR8
3e6 7CAR8
1e6 7CAR8
随机分组
治疗
接种后的天数
总通量 [p/s] (线性)

图49A

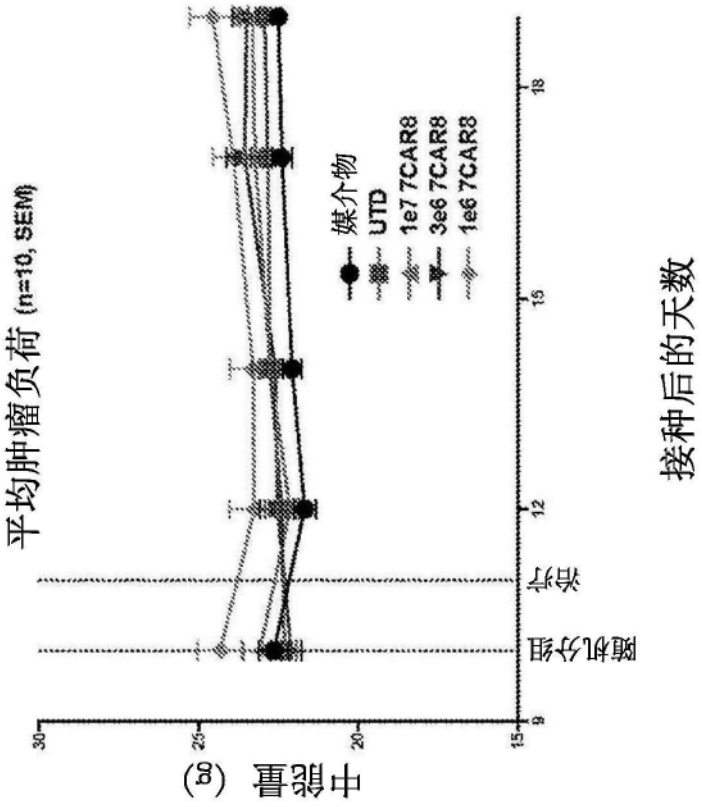

平均肿瘤负荷 (n=10, SEM)
中能量 (g)
接种后的天数
随机分组
治疗
媒介物
UTD
1e7 7CAR8
3e6 7CAR8
1e6 7CAR8
30
25
20
15
9
12
15
18

图49B

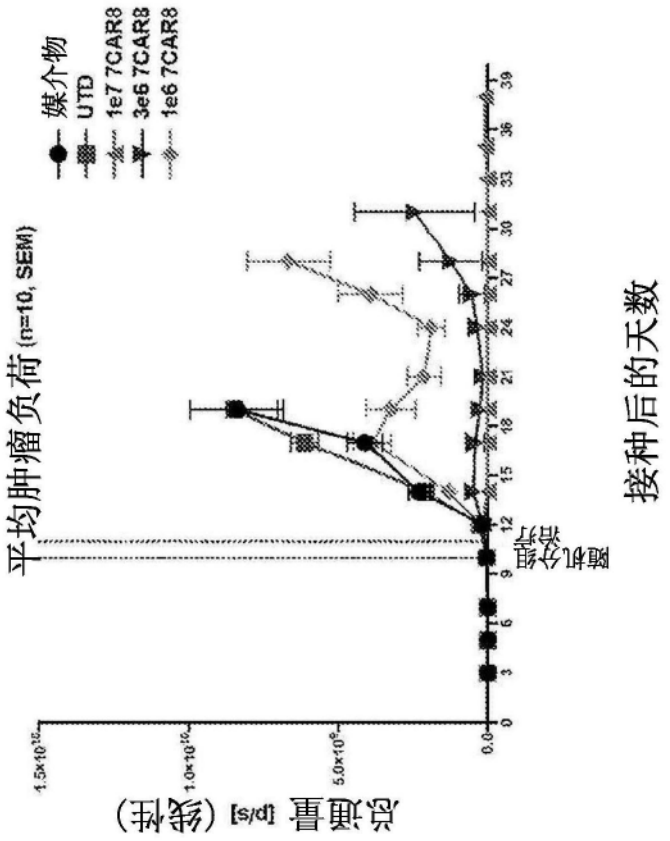

媒介物
UTD
1e7 7CAR8
3e6 7CAR8
1e6 7CAR8
平均肿瘤负荷 (n=10, SEM)
总通量 [p/s] (线性)
1.5x10^10
1.0x10^10
5.0x10^9
0.0
随机分组
治疗
0
3
6
9
12
15
18
21
24
27
30
33
36
39
接种后的天数

图50A

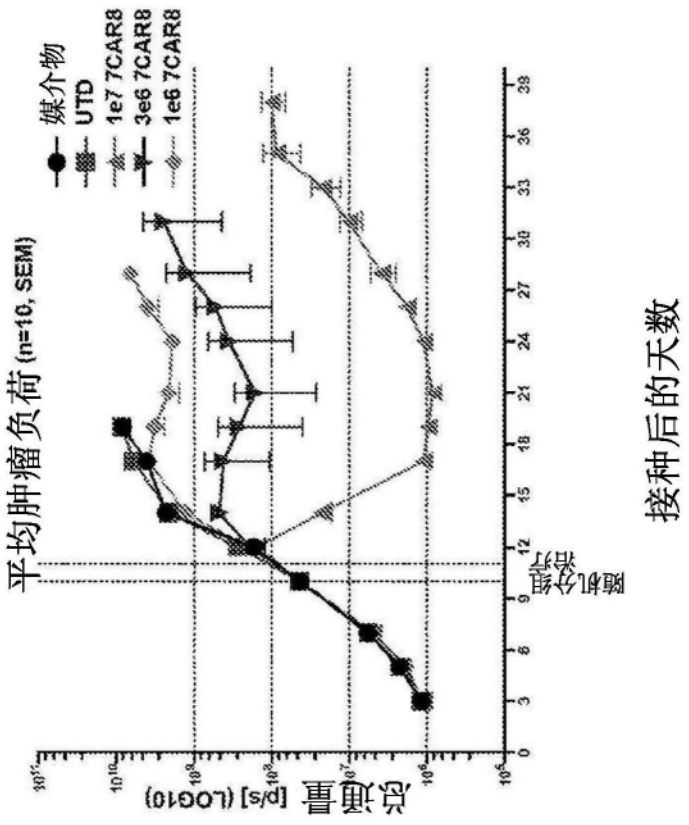

媒介物
UTD
1e7 7CAR8
3e6 7CAR8
1e6 7CAR8
平均肿瘤负荷 (n=10, SEM)
总通量 [p/s] (LOG10)
随机分组
治疗
接种后的天数

图50B

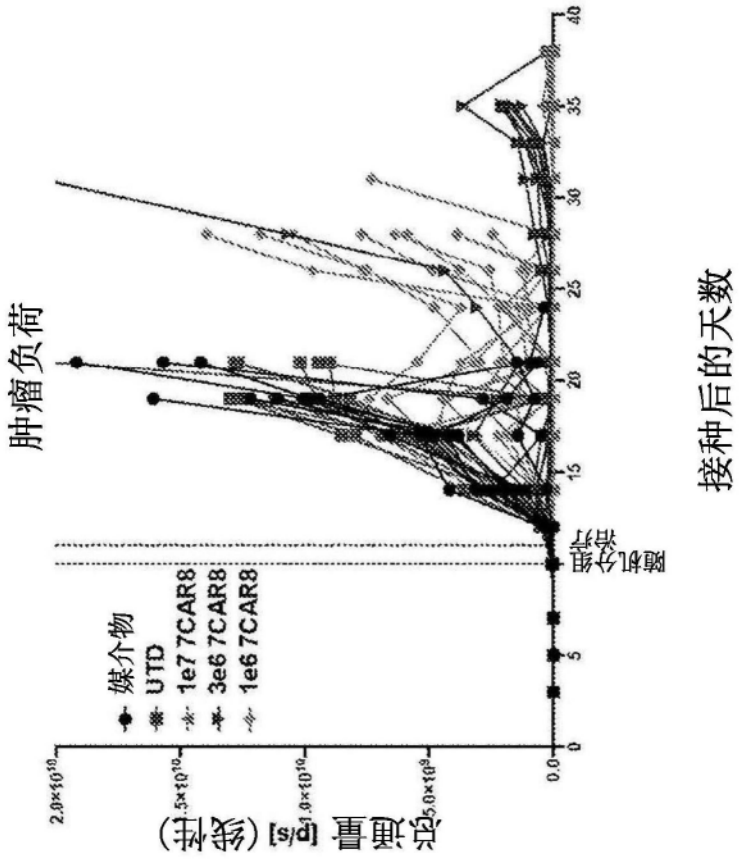

肿瘤负荷
接种后的天数
总通量 [p/s] (线性)
媒介物
UTD
1e7 7CAR8
3e6 7CAR8
1e6 7CAR8
随机分组
治疗
0
5
10
15
20
25
30
35
40
0.0
5.0×10⁹
3.0×10¹⁰
1.5×10¹⁰
2.0×10¹⁰

图51A

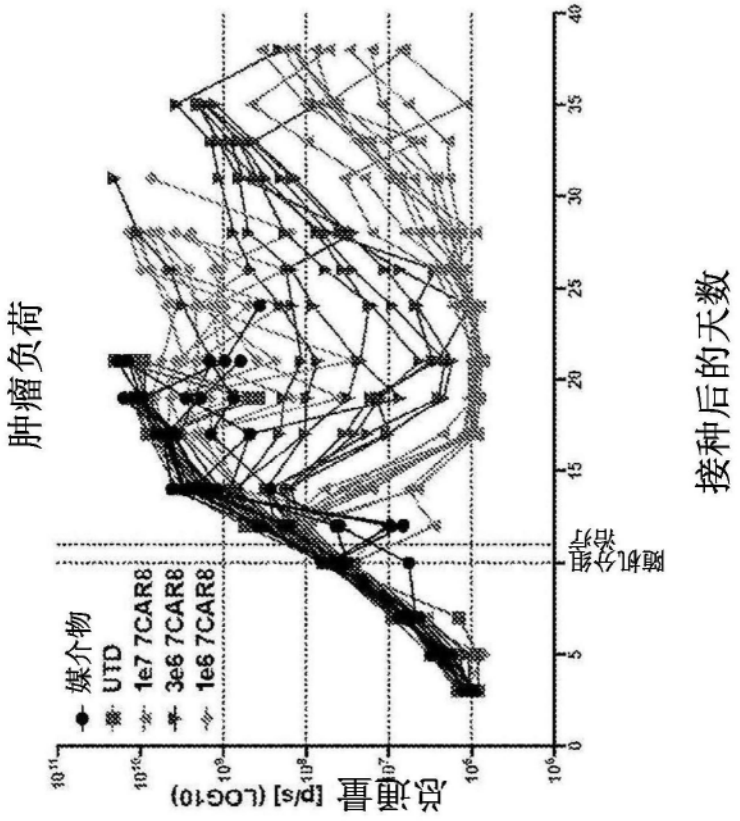

肿瘤负荷
总通量 [p/s] (LOG10)
接种后的天数
随机分组
治疗
媒介物
UTD
1e7 7CAR8
3e6 7CAR8
1e6 7CAR8

图51B

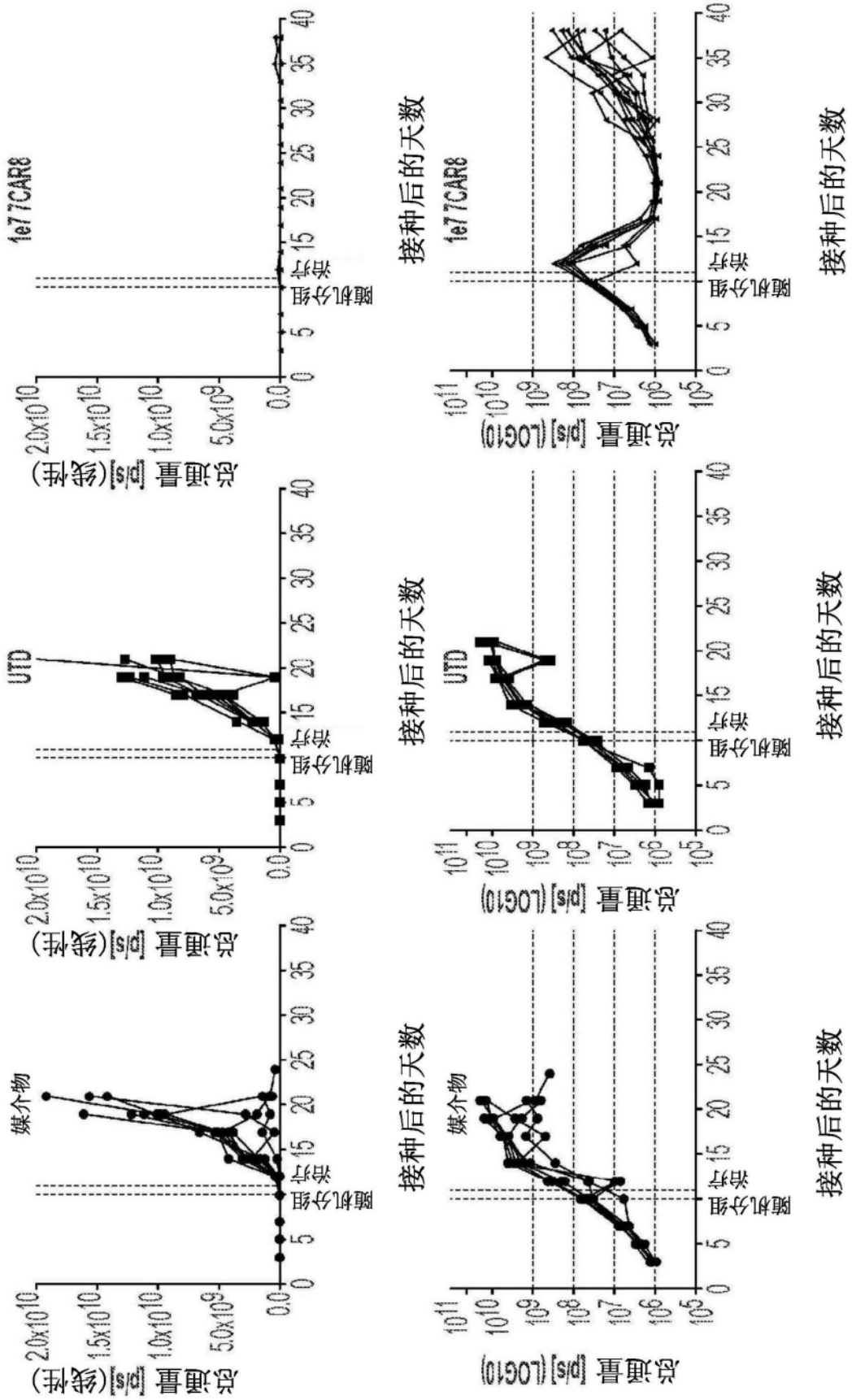

媒介物
UTD
1e7 7CAR8
随机分组
治疗
接种后的天数
总通量 [p/s] (线性)
总通量 [p/s] (LOG10)
0
5
15
20
25
30
35
40
0.0
5.0x10^9
1.0x10^10
1.5x10^10
2.0x10^10
10^5
10^6
10^7
10^8
10^9
10^10
10^11

图52A

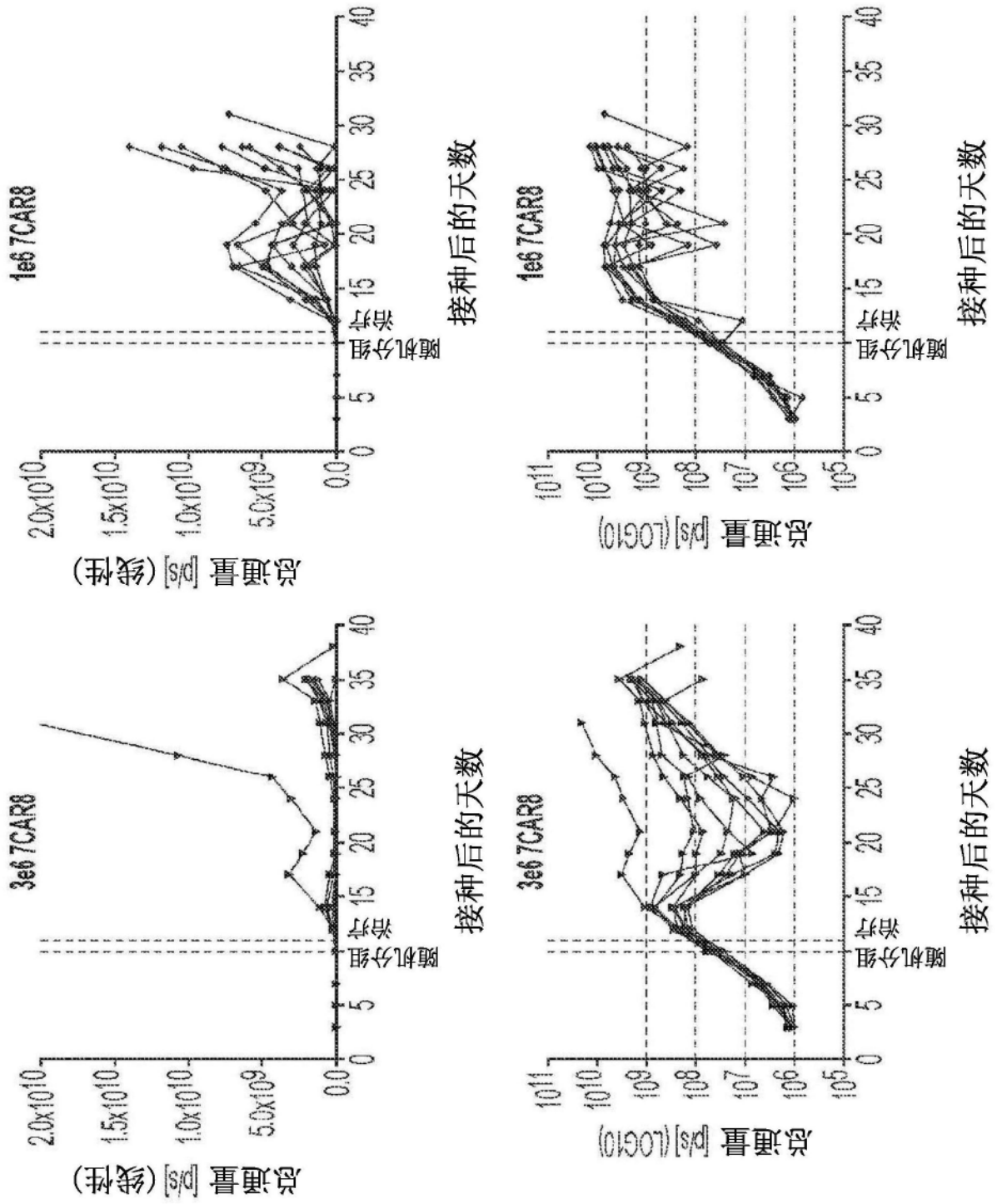

1e6 7CAR8
3e6 7CAR8
随机分组
治疗
接种后的天数
总通量 [p/s] (线性)
总通量 [p/s] (LOG10)
2.0x10^10
1.5x10^10
1.0x10^10
5.0x10^9
0.0
10^11
10^10
10^9
10^8
10^7
10^6
10^5
0
5
15
20
25
30
35
40

图52B

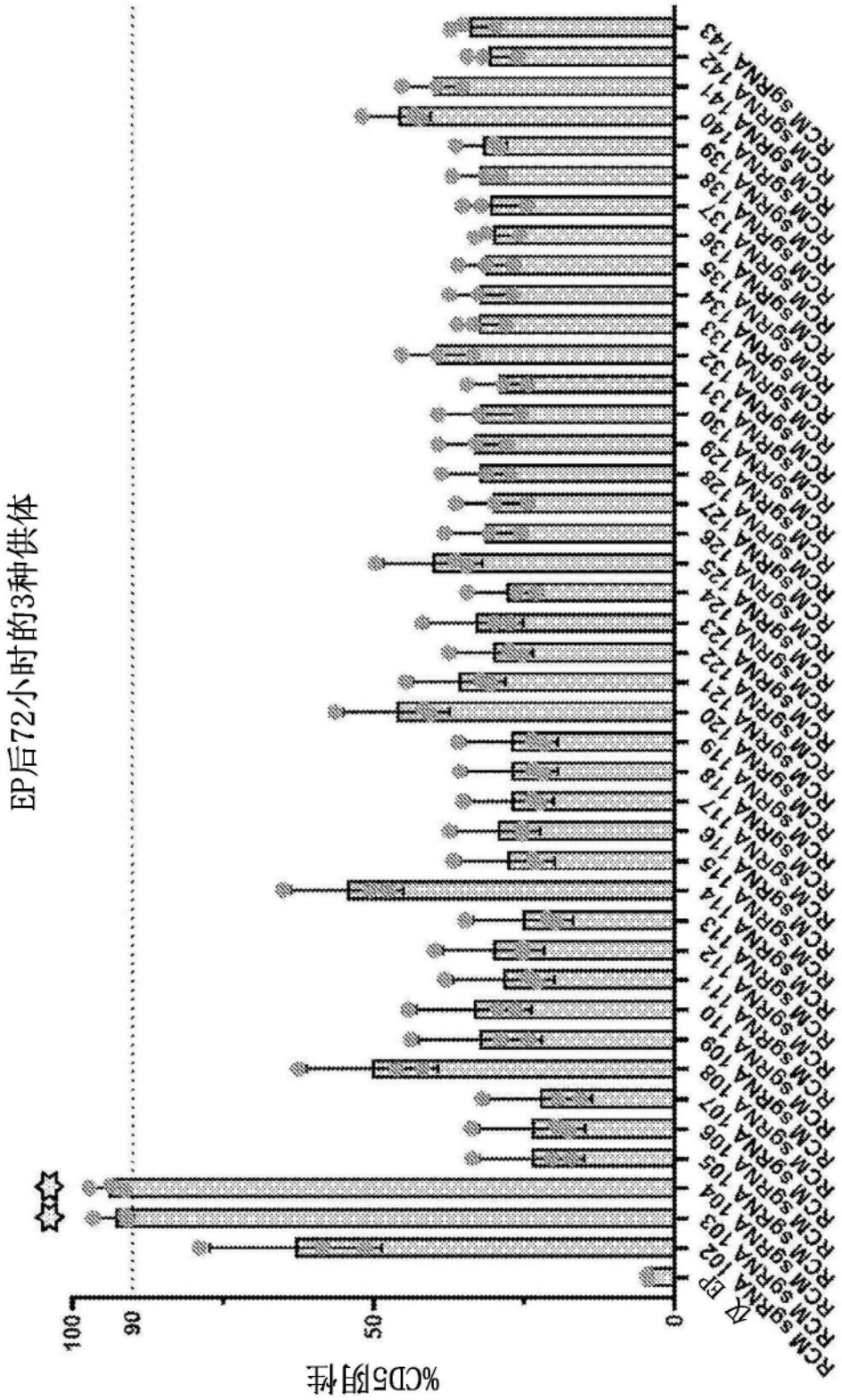

EP后72小时的3种供体
%CD5阴性
100
90
50
0
EP
RCM sgRNA 102
RCM sgRNA 103
RCM sgRNA 104
RCM sgRNA 105
RCM sgRNA 106
RCM sgRNA 107
RCM sgRNA 108
RCM sgRNA 109
RCM sgRNA 110
RCM sgRNA 111
RCM sgRNA 112
RCM sgRNA 113
RCM sgRNA 114
RCM sgRNA 115
RCM sgRNA 116
RCM sgRNA 117
RCM sgRNA 118
RCM sgRNA 119
RCM sgRNA 120
RCM sgRNA 121
RCM sgRNA 122
RCM sgRNA 123
RCM sgRNA 124
RCM sgRNA 125
RCM sgRNA 126
RCM sgRNA 127
RCM sgRNA 128
RCM sgRNA 129
RCM sgRNA 130
RCM sgRNA 131
RCM sgRNA 132
RCM sgRNA 133
RCM sgRNA 134
RCM sgRNA 135
RCM sgRNA 136
RCM sgRNA 137
RCM sgRNA 138
RCM sgRNA 139
RCM sgRNA 140
RCM sgRNA 141
RCM sgRNA 142
RCM sgRNA 143

图53

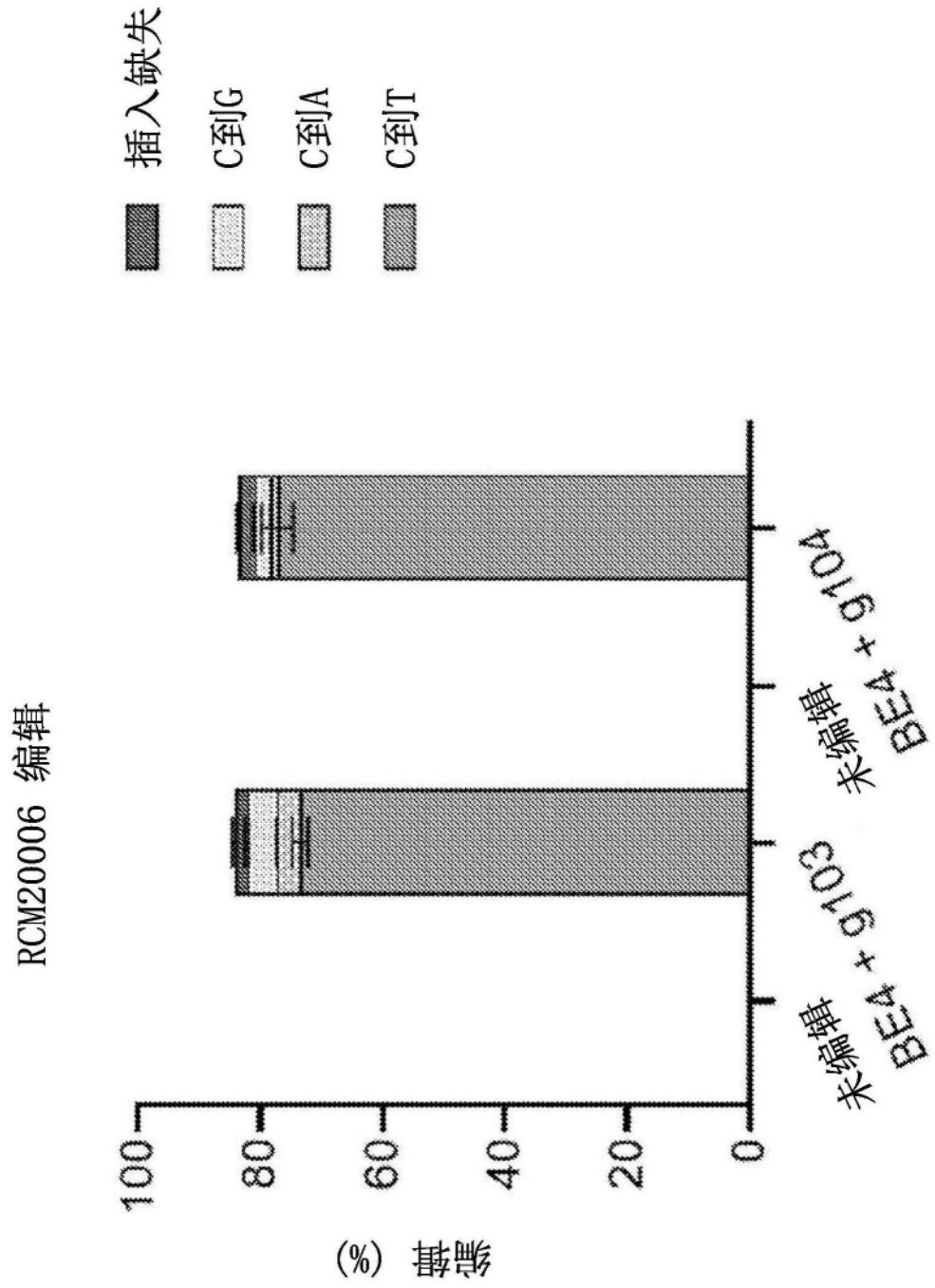

插入缺失
C到G
C到A
C到T
RCM20006 编辑
编辑 (%)
100
80
60
40
20
0
未编辑
BE4 + 9103
未编辑
BE4 + 9104

图54

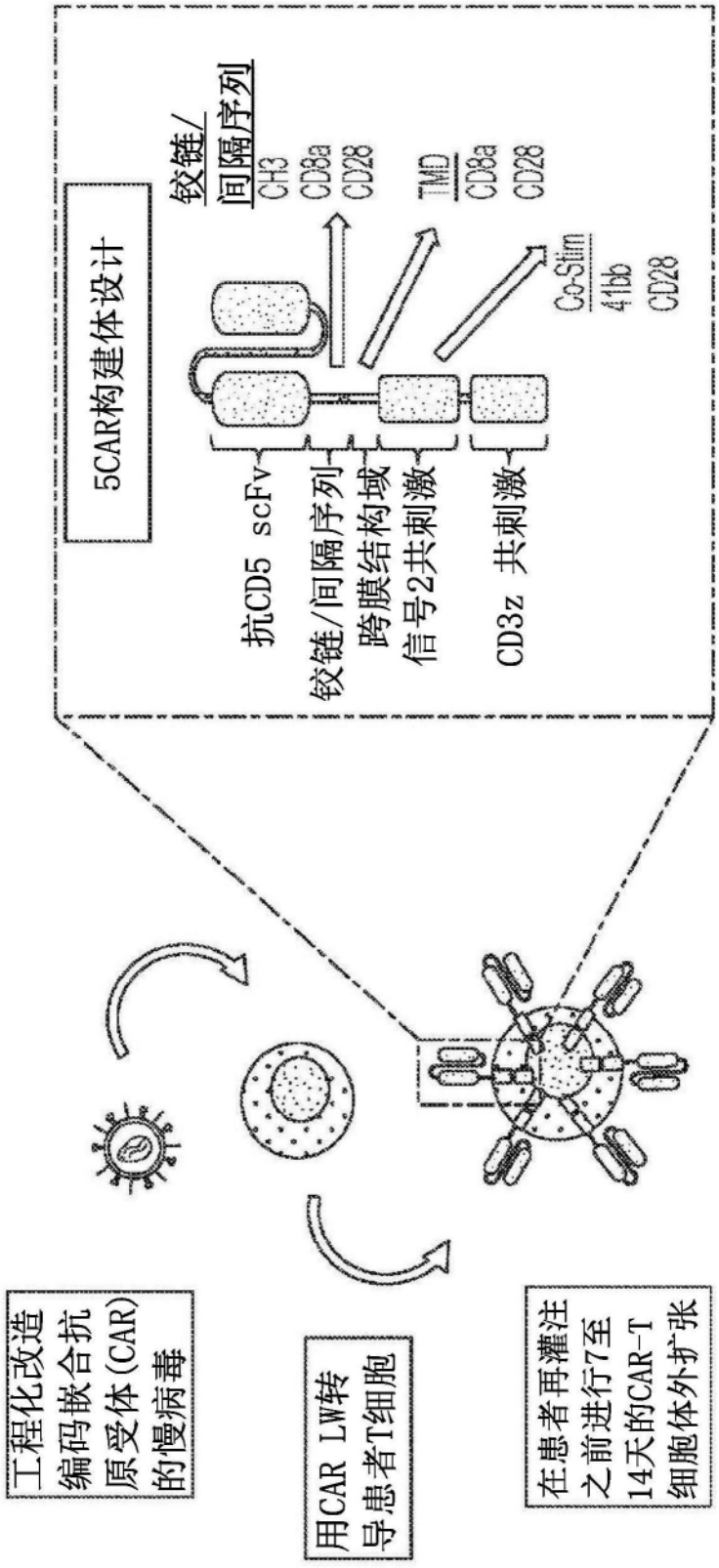

5CAR构建体设计
铰链/
间隔序列
CH3
CD8a
CD28
TMD
CD8a
CD28
Co-Stim
41bb
CD28
抗CD5 scFv
铰链/间隔序列
跨膜结构域
信号2共刺激
CD3z 共刺激
工程化改造
编码嵌合抗
原受体(CAR)
的慢病毒
用CAR LW转
导患者T细胞
在患者再灌注
之前进行7至
14天的CAR-T
细胞体外扩张

图55

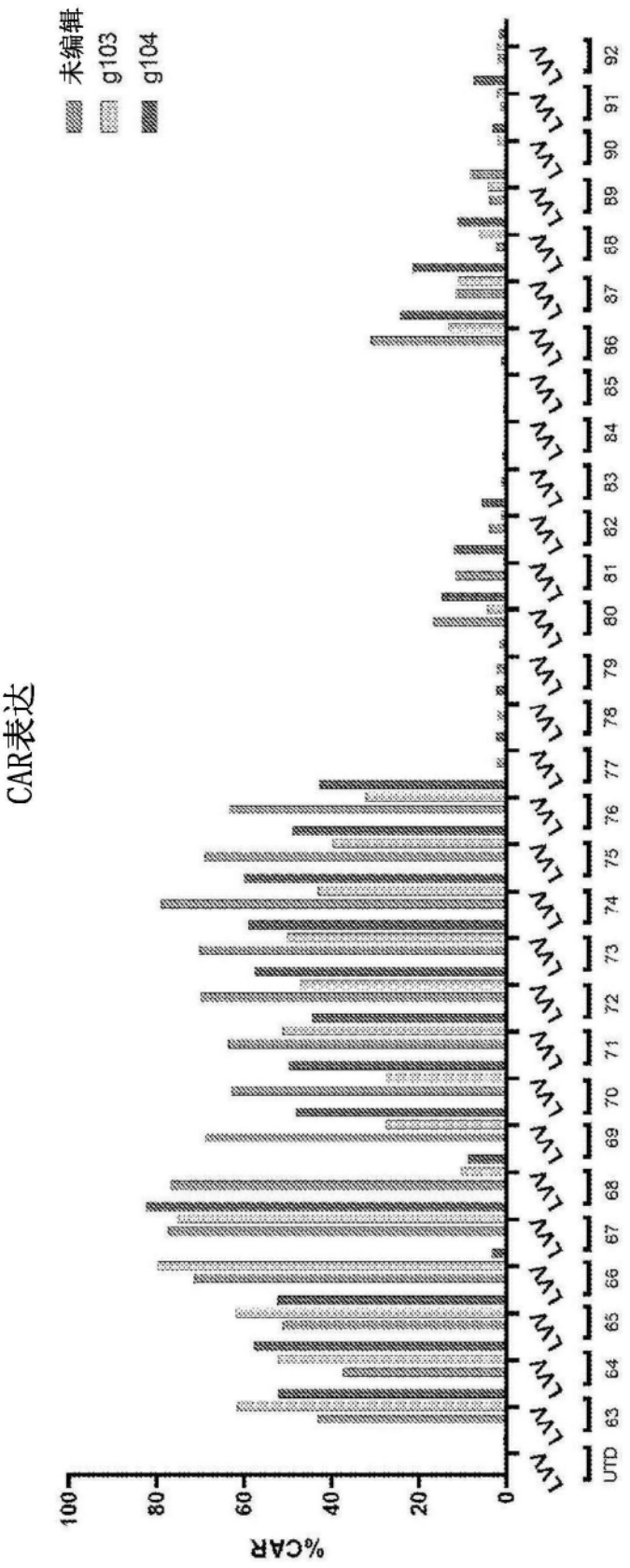

未编辑
g103
g104
CAR表达
%CAR
0
20
40
60
80
100
UTD
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92

图56

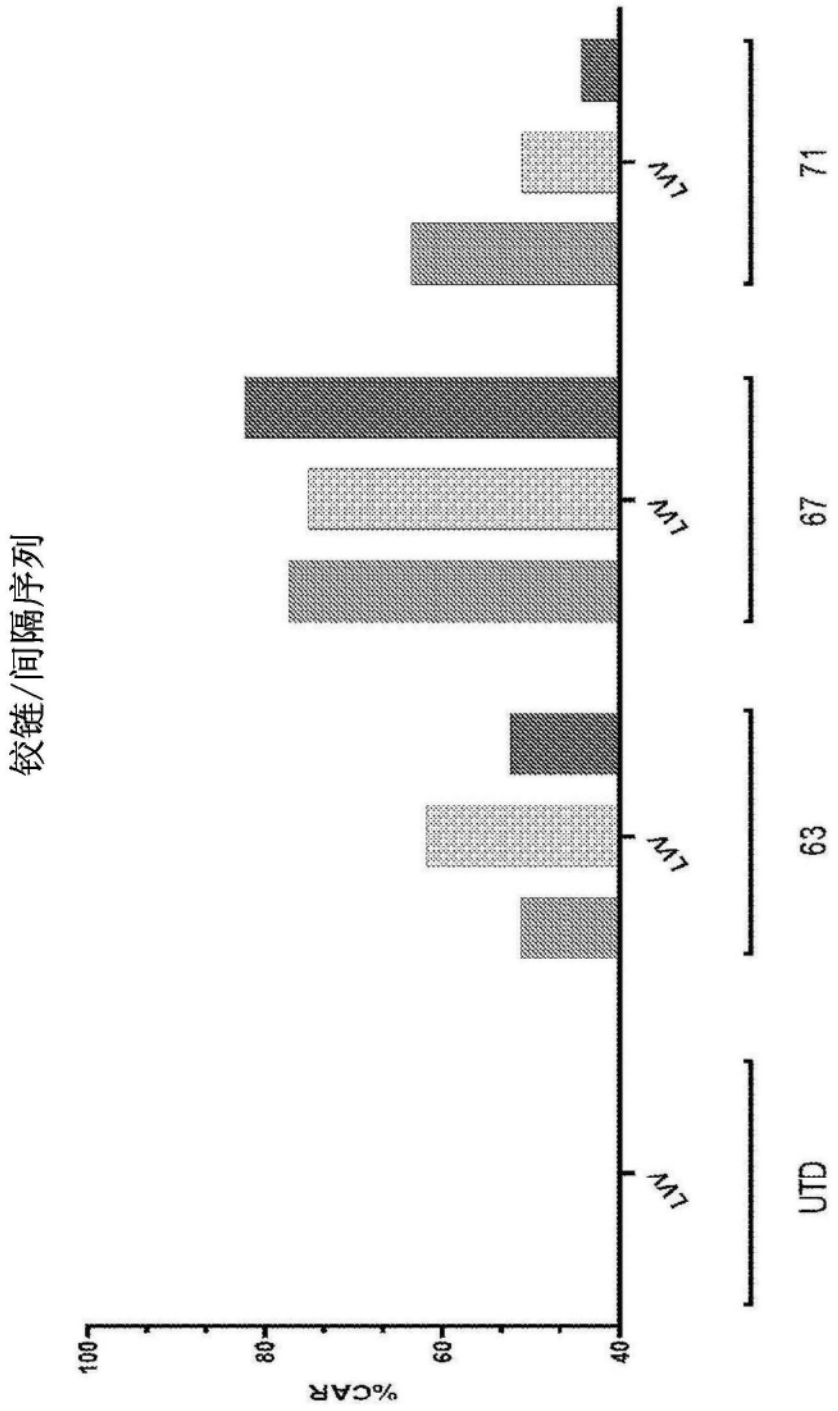


图57

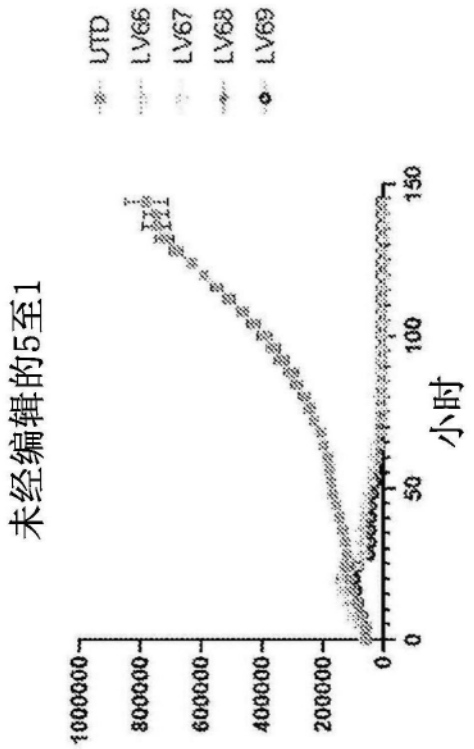

UTD
LV66
LV67
LV68
LV69
未经编辑的5至1
1000000
800000
600000
400000
200000
0
50
100
150
小时

图58A

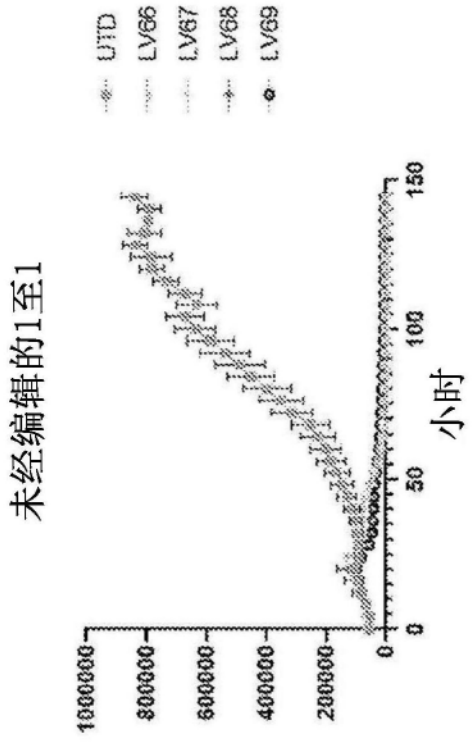

UTD
LV66
LV67
LV68
LV69
未经编辑的1至1
1000000
800000
600000
400000
200000
0
50
100
150
小时

图58B

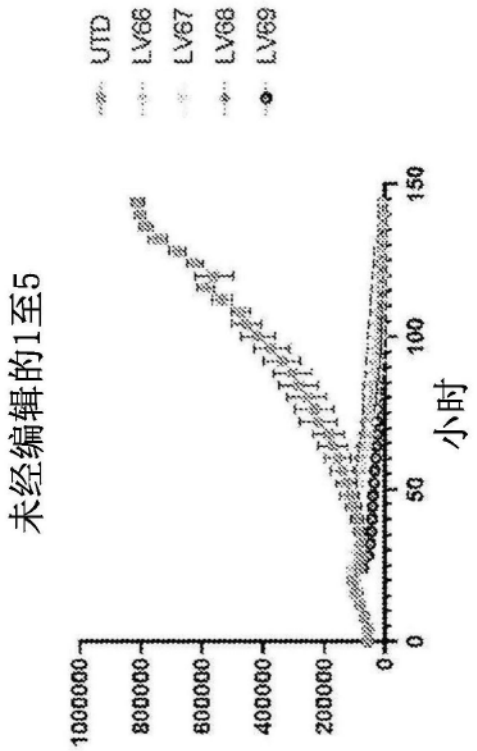

UTD
LV66
LV67
LV68
LV69
未经编辑的1至5
小时
1000000
800000
600000
400000
200000
0
50
100
150

图58C

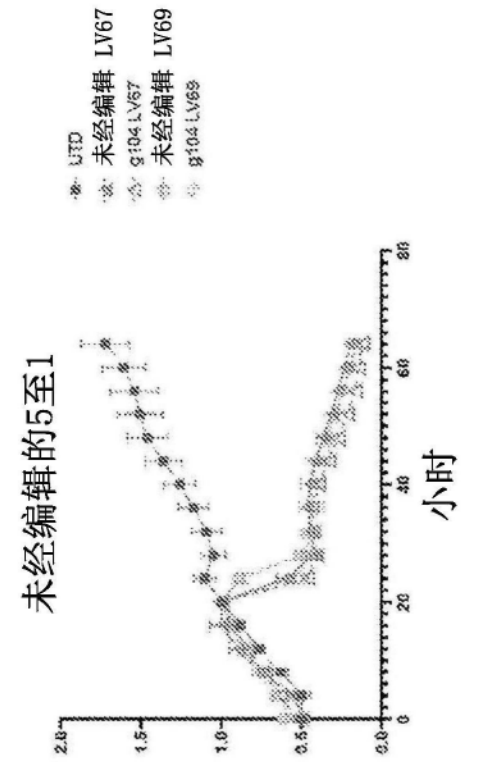

UTD
未经编辑 LV67
g104 LV67
未经编辑 LV69
g104 LV69
未经编辑的5至1
小时
2.0
1.5
1.0
0.5
0.0
0
20
40
60
80

图58D

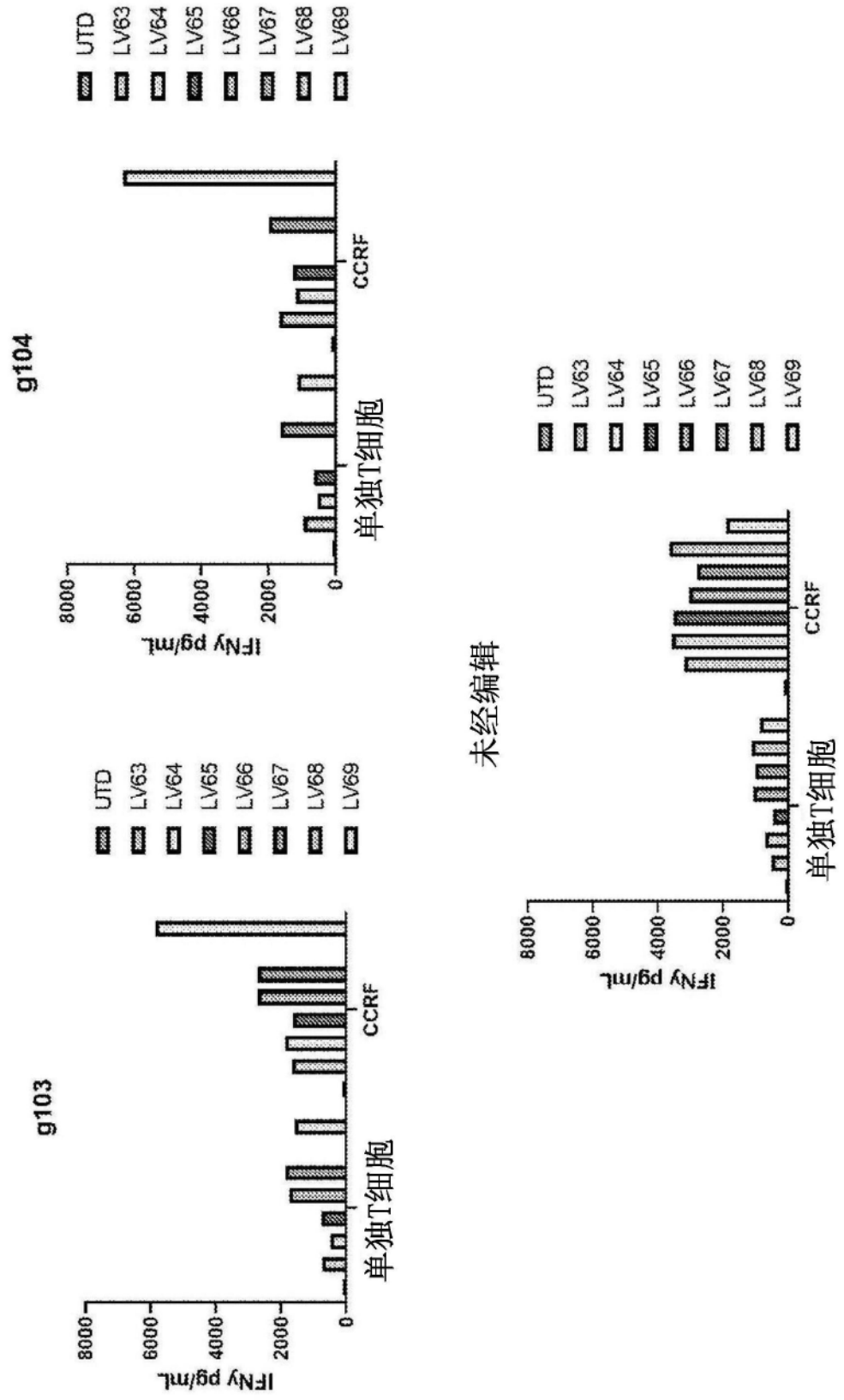

g104
UTD
LV63
LV64
LV65
LV66
LV67
LV68
LV69
IFNγ pg/mL
8000
6000
4000
2000
0
单独T细胞
CCRF
g103
未经编辑

图59A

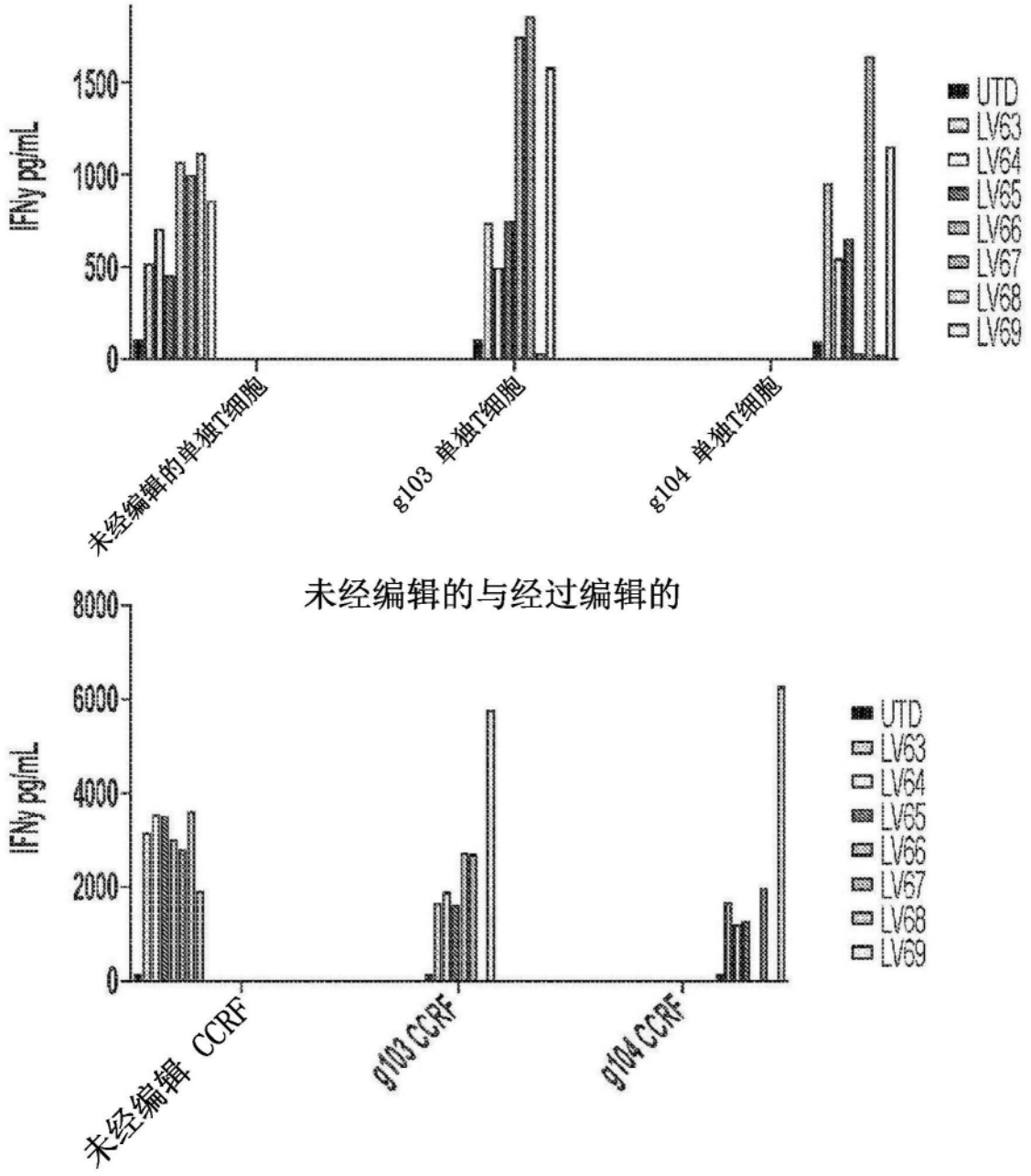

IFNγ pg/mL
1500
1000
500
0
UTD
LV63
LV64
LV65
LV66
LV67
LV68
LV69
未经编辑的单独T细胞
g103 单独T细胞
g104 单独T细胞
未经编辑的与经过编辑的
IFNγ pg/mL
8000
6000
4000
2000
0
UTD
LV63
LV64
LV65
LV66
LV67
LV68
LV69
未经编辑 CCRF
g103 CCRF
g104 CCRF

图59B