



(12) 发明专利

(10) 授权公告号 CN 112469446 B

(45) 授权公告日 2024. 12. 17

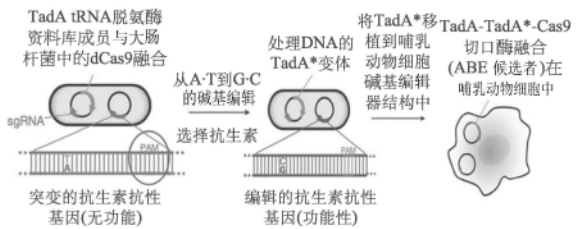
(21) 申请号 201980046479.3	(73) 专利权人 比姆医疗股份有限公司
(22) 申请日 2019.05.11	地址 美国马萨诸塞州
(65) 同一申请的已公布的文献号 申请公布号 CN 112469446 A	(72) 发明人 D·布赖森 J·埃文斯 M·帕克 J·M·格尔克 N·彼得罗相
(43) 申请公布日 2021.03.09	(74) 专利代理机构 北京戈程知识产权代理有限公司 11314
(30) 优先权数据 62/670,588 2018.05.11 US 62/780,838 2018.12.17 US 62/817,986 2019.03.13 US	专利代理师 程伟
(85) PCT国际申请进入国家阶段日 2021.01.11	(51) Int.Cl. A61K 48/00 (2006.01)
(86) PCT国际申请的申请数据 PCT/US2019/031898 2019.05.11	(56) 对比文件 W0 2017165862 A1,2017.09.28 W0 2018027078 A1,2018.02.08
(87) PCT国际申请的公布数据 W02019/217943 EN 2019.11.14	审查员 郭晓烨
	权利要求书4页 说明书193页 序列表(电子公布) 附图11页

(54) 发明名称

利用可编程碱基编辑器系统编辑单核苷酸多态性的方法

(57) 摘要

本发明提供一种组合物和使用碱基编辑器的方法,所述碱基编辑器包含多核苷酸可编程核苷酸结合结构域以及结合指导多核苷酸的核碱基编辑结构域。本发明还提供一种用于编辑靶核苷酸序列的核碱基的碱基编辑器系统。



1. 一种与一种或多种指导多核苷酸复合的碱基编辑器于制备用于编辑包含与 α -1抗胰蛋白酶缺乏症(A1AD)相关的单核苷酸多态性(SNP)的SERPINA1多核苷酸的药物的用途,所述药物与所述SERPINA1多核苷酸接触,其中所述碱基编辑器包含对具有核酸序列5'-NGC-3'的原间隔物相邻基序(PAM)具有特异性的脓性链球菌Cas9内切酶的多核苷酸可编程DNA结合结构域、由MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD氨基酸序列所组成的野生型TadA腺苷脱氨酶结构域和由SEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD氨基酸序列所组成的TadA*7.10腺苷脱氨酶结构域,其中一种或多种所述指导多核苷酸靶向所述碱基编辑器,以造成SERPINA1多核苷酸的A·T至G·C改变,其中所述一种或多种指导多核苷酸的间隔物由选自以下所组成的组的核苷酸序列所组成:

ACCAUCGACAAGAAAGGGACUGA;

CAUCGACAAGAAAGGGACUGA;

AUCGACAAGAAAGGGACUGA;

UCGACAAGAAAGGGACUGA;和

CGACAAGAAAGGGACUGA;

其中所述与A1AD相关的SNP导致在氨基酸位置342具有赖氨酸的 α -1抗胰蛋白酶(A1AT)多肽的表达,以及其中所述A·T至G·C改变导致在氨基酸位置341包含甘氨酸的A1AT多肽的表达。

2. 根据权利要求1所述的用途,其中所述接触是在细胞中。

3. 根据权利要求2所述的用途,其中所述细胞是体内或离体的。

4. 根据权利要求1所述的用途,其中所述一种或多种指导多核苷酸的支架由以下核苷酸序列所组成:5'-GUUUUAGAGCUAGAAAUAGCAAGUAAAAUAAGGCUAGUCCG UUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU-3'。

5. 一种通过引入以下(i)和(ii)至细胞或其祖细胞而产生的细胞:

(i) 碱基编辑器或编码所述碱基编辑器的多核苷酸,其中所述碱基编辑器包含对具有核酸序列5'-NGC-3'的原间隔物相邻基序(PAM)具有特异性的脓性链球菌Cas9内切酶的多核苷酸可编程DNA结合结构域、由MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD氨基酸序列所组成的野生型TadA腺苷脱氨酶结构域和由SEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD氨基酸序列所组成的TadA*7.10腺苷脱氨酶结构域;以及

(ii) 一种或多种指导多核苷酸,其靶向碱基编辑器,以造成包含与A1AD相关的SNP的SERPINA1多核苷酸的A·T至G·C改变,其中所述一种或多种指导多核苷酸的间隔物由选自以下所组成的组的核苷酸序列所组成:

ACCAUCGACAAGAAAGGGACUGA;

CAUCGACAAGAAAGGGACUGA;

AUCGACAAGAAAGGGACUGA;
UCGACAAGAAAGGGACUGA;和
CGACAAGAAAGGGACUGA;

其中所述与A1AD相关的SNP导致在氨基酸位置342具有赖氨酸的 α -1抗胰蛋白酶(A1AT)多肽的表达,以及其中所述A•T至G•C改变导致在氨基酸位置341包含甘氨酸的A1AT多肽的表达。

6. 根据权利要求5所述的细胞,其中所述产生的细胞是肝细胞。
7. 根据权利要求5或权利要求6所述的细胞,其中所述细胞是肝细胞。
8. 根据权利要求5或权利要求6所述的细胞,其中所述细胞的祖细胞是诱导性多能干细胞。
9. 根据权利要求6所述的细胞,其中所述肝细胞表达A1AT多肽。
10. 根据权利要求5或权利要求6所述的细胞,其中所述细胞来自患有A1AD的受试者。
11. 根据权利要求5或权利要求6所述的细胞,其中所述细胞是哺乳动物细胞。
12. 根据权利要求5或权利要求6所述的细胞,其中所述细胞被选择用于所述与A1AD相关的SNP的所述A•T至G•C改变。

13. 根据权利要求5所述的细胞,其中所述一种或多种指导多核苷酸的支架由以下核苷酸序列所组成:5'-GUUUUAGAGCUAGAAAUAGCAAGUAAAAUAAGGCUAGUCCG UUAUCAACUUGAAAAAG UGGCACCGAGUCGGUGCUUUU-3'。

14. 根据权利要求5或权利要求6所述的细胞,其中所述碱基编辑器和所述一种或多种指导多核苷酸在所述细胞中形成复合物。

15. 一种根据权利要求5至14中任一项所述的细胞于制备用于对此需要的受试者治疗A1AD的药物的用途,其中所述受试者表达在氨基酸位置342具有赖氨酸突变的A1AT多肽。

16. 根据权利要求15所述的用途,其中所述细胞对于所述受试者是自体的,同种异体的或异种的。

17. 一种分离细胞或细胞群,其从权利要求5至14中任一项所述的细胞繁殖或扩增。

18. 一种碱基编辑器或编码所述碱基编辑器的多核苷酸以及一种或多种靶向碱基编辑器的指导多核苷酸于制备用于对受试者治疗A1AD的药物的用途,其中所述受试者表达在氨基酸位置342具有赖氨酸突变的A1AT多肽,

其中所述碱基编辑器包含对具有核酸序列5'-NGC-3'的原间隔物相邻基序(PAM)具有特异性的脓性链球菌Cas9内切酶的多核苷酸可编程DNA结合结构域、由MSEVEFSHEYWMRHAL TLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAG AMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRQEIKAKKQSSSTD氨基酸序列所组成的野生型TadA腺苷脱氨酶结构域和由SEVEFSHEYWMRHAL TLAKRARDEREVPVG AVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVR NAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAKKQSSSTD氨基酸序列所组成的TadA*7.10腺苷脱氨酶结构域;以及

其中所述指导多核苷酸靶向碱基编辑器,以造成包含与A1AD相关的SNP的SERPINA1多核苷酸的A•T至G•C改变,其中所述一种或多种指导多核苷酸的间隔物由选自由以下所组成的组的核苷酸序列所组成:

ACCAUCGACAAGAAAGGGACUGA;
CAUCGACAAGAAAGGGACUGA;
AUCGACAAGAAAGGGACUGA;
UCGACAAGAAAGGGACUGA;和
CGACAAGAAAGGGACUGA;

其中所述与A1AD相关的SNP导致在氨基酸位置342具有赖氨酸的 α -1抗胰蛋白酶(A1AT)多肽的表达,以及其中所述A•T至G•C改变导致在氨基酸位置341包含甘氨酸的A1AT多肽的表达。

19.根据权利要求18所述的用途,其中所述受试者是哺乳动物。

20.根据权利要求18或权利要求19所述的用途,其中所述药物递送至所述受试者的细胞。

21.根据权利要求20所述的用途,其中所述细胞是肝细胞。

22.根据权利要求20所述的用途,其中所述细胞是肝细胞的祖细胞。

23.根据权利要求18所述的用途,其中所述一种或多种指导多核苷酸的支架由以下核苷酸序列所组成:5'-GUUUUAGAGCUAGAAAUAGCAAGUAAAAUAAGGCUAGUCCG UUAUCAACUUGAAAA AGUGGCACCGAGUCGGUGCUUUU-3'。

24.一种产生肝细胞的方法,所述方法包括:

(a)将包含SERPINA1多核苷酸的诱导性多能干细胞或肝细胞祖细胞引入以下(i)和(ii),所述SERPINA1多核苷酸包含与A1AD相关的SNP,

(i)碱基编辑器或编码所述碱基编辑器的多核苷酸,其中所述碱基编辑器包含对具有核苷酸序列5'-NGC-3'的原间隔物相邻基序(PAM)具有特异性的脓性链球菌Cas9内切酶的多核苷酸可编程核苷酸结合结构域、由MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGW NRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVL HHPGMNHRVEITEGILADECAALLSDFFRMRQEIKAKKAQSSTD氨基酸序列所组成的野生型TadA腺苷脱氨酶结构域和由SEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD氨基酸序列所组成的TadA*7.10腺苷脱氨酶结构域;以及

(ii)一种或多种指导多核苷酸,其中所述一种或多种指导多核苷酸靶向碱基编辑器,以造成所述SERPINA1多核苷酸的A•T至G•C改变,其中所述一种或多种指导多核苷酸的间隔物由选自以下所组成的组的核苷酸序列所组成:

ACCAUCGACAAGAAAGGGACUGA;
CAUCGACAAGAAAGGGACUGA;
AUCGACAAGAAAGGGACUGA;
UCGACAAGAAAGGGACUGA;和
CGACAAGAAAGGGACUGA;

其中所述与A1AD相关的SNP导致在氨基酸位置342具有赖氨酸的 α -1抗胰蛋白酶(A1AT)多肽的表达,以及其中所述A•T至G•C改变导致在氨基酸位置341包含甘氨酸的A1AT多肽的表达;以及

(b) 将诱导性多能干细胞或肝细胞祖细胞分化为肝细胞。

25. 根据权利要求24所述的方法, 其中所述肝细胞祖细胞获自患有A1AD的受试者。

26. 根据权利要求24或权利要求25所述的方法, 其中所述肝细胞或肝细胞祖细胞是哺乳动物细胞或人类细胞。

27. 根据权利要求24或权利要求25所述的方法, 其中所述细胞被选择用于与A1AD相关的SNP的A • T至G • C改变。

28. 根据权利要求24所述的方法, 其中所述一种或多种指导多核苷酸的支架由以下核苷酸序列所组成: 5' -GUUUUAGAGCUAGAAAUAGCAAGUAAAAUAAGGCUAGUCCG UUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU-3'。

29. 根据权利要求24或权利要求25所述的方法, 其中所述碱基编辑器和所述一种或多种指导多核苷酸在所述细胞中形成复合物。

利用可编程碱基编辑器系统编辑单核苷酸多态性的方法

[0001] 相关申请

[0002] 本申请主张2018年5月11日提交的美国临时申请号62/670,588,2018年12月17日提交的美国临时申请号62/780,838和2019年3月13日提交的美国临时申请62/817,986的优先权和利益,其整体内容通过引用并入本文。

背景技术

[0003] 对于大多数已知的遗传疾病,研究或解决疾病的根本原因,需要校正目标基因座中的点突变,而不是基因的随机破坏。当前的基因组编辑技术利用规律间隔成簇短回文重复序列(clustered regularly interspaced short palindromic repeat,CRISPR)系统在目标基因座处引入双链DNA断裂,作为基因校正的第一步。为响应双链DNA断裂,细胞DNA修复过程通常会通过非同源末端连接(non-homologous end joining,NHEJ)导致DNA切割位点的随机插入或缺失(插入缺失,Indels)。尽管大多数遗传疾病是由点突变引起的,但当前的点突变校正方法效率低下,通常会在靶基因座上引起大量的随机插入和缺失(Indels),这是由于细胞对dsDNA断裂的反应所致。因此,需要一种改进形式的基因组编辑,其更有效且具有更少的副产物,例如随机插入或缺失(Indels)或易位。

[0004] α -1抗胰蛋白酶缺乏症(A1AD)是一种遗传疾病,其中编码 α -1抗胰蛋白酶(A1AT)蛋白的SERPINA1基因中的致病性突变导致患有所述疾病的个体减少蛋白质产生。A1AT是嗜中性粒细胞弹性蛋白酶特别好的抑制剂,可保护肺等组织器官免受弹性蛋白降解。因此,具有A1AD的患者的肺中的弹性蛋白更容易被嗜中性粒细胞弹性蛋白酶降解,且随着时间的流逝,肺弹性的流失发展成慢性阻塞性肺疾病(COPD)。在健康个体中,A1AT是由肝脏中的肝细胞生成,并分泌到全身循环中,在所述循环中所述蛋白质起蛋白酶抑制剂的作用。

[0005] 最常见的致病性A1AT变异是SERPINA1基因中的鸟嘌呤至腺嘌呤(G $\rightarrow$ A)突变,导致A1AT蛋白第342位氨基酸被谷氨酸取代为赖氨酸。此种取代导致蛋白质在肝细胞内错误折叠并聚合,最终,有毒聚集体可导致肝损伤和肝硬化。尽管可能通过基因敲除(CRISPR/ZFN/TALEN)或基因敲除(siRNA)解决肝脏毒性,但这些方法均未解决肺部病理学问题。尽管肺部病理可以用蛋白质替代疗法解决,但此种疗法无法解决肝脏毒性。基因治疗也不足以解决A1AT基因缺陷。由于A1AD患者的肝脏已经处于由内源性A1AT聚集引起的严重疾病负担中,因此增加肝脏中A1AT的基因治疗会适得其反。因此,需要一种治疗患有A1AD的患者的方法,其同时解决所述疾病伴随的肺病理学和肝毒性。

[0006] 参考文献援引

[0007] 本申请说明书中提到的所有出版、专利和专利申请都以相同的程度通过引用并入本文,如同明确地和独立地指出个别独立的出版物、专利或专利申请通过引用并入。除非另有说明,否则本说明书中提及的出版物、专利和专利申请通过引用整体并入本文。

发明内容

[0008] 如本文所述,本发明提供一种使用可编程核碱基编辑器精确校正与疾病或病症相

关的蛋白质中的病原性氨基酸的组合物和方法。在一个特定方面,所述组合物和方法可用于治疗 α -1抗胰蛋白酶缺乏症(A1AD)。在一个实施方案中,所述用于治疗A1AD的组合物和方法是利用腺苷(A)碱基编辑器(ABE-(NGC变体))来精确地校正内源SERPINA1基因中的有害的单核苷酸多态性(SNP)。在一个实施方案中,所述组合物和方法校正有害突变E342K,其影响编码的 α -1抗胰蛋白酶(A1AT)蛋白的活性和功能。此种校正同时消除肝脏中的致病蛋白负担,并恢复了肺部功能蛋白。

[0009] 在一方面,本发明提供一种编辑包含与 α 1抗胰蛋白酶缺乏症(A1AD)相关的单核苷酸多态性(SNP)的SERPINA1多核苷酸的方法,其中所述方法包括使SERPINA1多核苷酸与碱基编辑器接触,所述碱基编辑器与一种或多种指导多核苷酸复合,其中碱基编辑器包含一个多核苷酸可编程的DNA结合结构域和一个腺苷脱氨酶结构域,且其中一种或多种指导多核苷酸靶向所述碱基编辑器,从而使与A1AD相关的SERPINA1基因中的SNP的A·T到G·C改变。在一个实施方案中,所述方法包括接触细胞,例如真核细胞、哺乳动物细胞或人类细胞。在另一个实施方案中,细胞是体内或离体的。

[0010] 在另一方面,本发明的特征在于一种细胞,所述细胞是通过将碱基编辑器、编码所述碱基编辑器的多核苷酸引入细胞或其祖细胞中而产生的,其中所述碱基编辑器包含多核苷酸可编程DNA结合域和腺苷脱氨酶域。以及靶向所述碱基编辑器的一种或多种指导多核苷酸,以造成与A1AD相关的基因(例如SERPINA1基因)中SNP的A·T到G·C改变。在一实施方案中,产生的细胞是肝细胞。在另一个实施方案中,所述细胞或其祖细胞是诱导性多能干细胞或肝细胞。在另一个实施方案中,肝细胞表达A1AT多肽。在另一个实施方案中,细胞来自患有A1AD的受试者。在另一个实施方案中,细胞是哺乳动物细胞或人细胞。

[0011] 在另一方面,本发明的特征在于一种在受试者中治疗A1AD的方法,所述方法包括对此需要的受试者施用如以上描述的方面和实施方案中所述的细胞。在一实施方案中,细胞对于受试者是自体的或同种异体的或异种的。

[0012] 在另一方面,本发明的特征在于从任何上述方面的细胞繁殖或扩增的分离的细胞或细胞群。

[0013] 在另一方面,本发明的特征在于一种对受试者治疗A1AD的方法,其中所述方法包括对此需要的受试者施用碱基编辑器或编码所述碱基编辑器的多核苷酸,其中所述碱基编辑器包含多核苷酸可编程的DNA结合域和腺苷脱氨酶域;以及靶向所述碱基编辑器的一种或多种指导多核苷酸,以造成与A1AD相关的基因(例如SERPINA1基因)中SNP的A·T到G·C改变。在一个实施方案中,受试者是哺乳动物或人类。在另一个实施方案中,所述方法涉及将碱基编辑器或编码碱基编辑器的多核苷酸以及一种或多种指导多核苷酸递送至受试者的细胞。在另一个实施方案中,细胞是肝细胞。在另一个实施方案中,细胞是肝细胞的祖细胞。在另一个实施方案中,肝细胞表达一含有突变的A1AT多肽。

[0014] 在另一方面,本发明的特征在于一种生产肝细胞或其祖细胞的方法,其中所述方法涉及(a)引入与A1AD相关的基因(例如SERPINA1基因)中含有SNP的诱导型多能干细胞或肝细胞祖细胞、碱基编辑器或编码碱基编辑器的多核苷酸,其中所述碱基编辑器包含多核苷酸可编程核苷酸结合结构域和腺苷脱氨酶结构域;以及一种或多种指导多核苷酸,其中所述一种或多种指导多核苷酸靶向碱基编辑器以造成与A1AD相关的SNP的A·T至G·C改变; (b) 将诱导性多能干细胞或肝细胞祖细胞分化为肝细胞。在一实施方案中,所述方法涉

及将诱导性多能干细胞分化为肝细胞或其祖细胞。在另一个实施方案中,诱导性多能干细胞包含E342K突变。在另一个实施方案中,肝细胞祖细胞获自患有A1AD的受试者。在另一个实施方案中,肝细胞或肝细胞祖细胞是哺乳动物细胞或人细胞。

[0015] 在另一个方面,碱基编辑器(BE)中所描述的组合物和方法中使用包括包含氨基酸序列的多肽:(i):

[0016] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRREQEIKAQKKAQSSTDSSGGSSGGSSGSETPGTSESATPESSGGSSGSSSEVEFSHEYWMRHALTLAKRADEREVPVGAVLVNNRVIGEGWNRRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSSGGSSGSSGSETPGTSESATPESSGGSSGSSGSDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSIDKADRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEEPNINASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDITYDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPLKDNREKIEKILTFRIPYYVGPLARGNSRFAMWTRKSEETITPWNFEVVDKGASQSFIERMTNFDKNLPNEKVLPKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKIKDKDFLDNEENEDILEDIVLTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAICKGILQTVKVVDELVKVMGRHKPENIVEMARENQTTQKGQKNSRERMKRIEEGKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNLTKAERGGELSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVVREINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPPKYGGMqPTVAYSVLVAKVEKGKSKLKS VKELLGITIMERSSEFKNPIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLASakfLQKGNELALPSKYVNFLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIIEQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPrAFKYFDTTIaRKeYrSTKEVLDATLIHQSIITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRK;以及(ii)腺苷脱氨酶结构域。

[0017] 在另一方面,本发明的特征在于一种指导RNA(gRNA),其包含以下核酸序列:

[0018] 5'-GUUUUAGAGCUAGAAAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU-3';

[0019] 5'-ACCAUCGACAAGAAAGGGACUGAGUUUUUAGAGAGUAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU-3';

[0020] 5'-CCAUCGACAAGAAAGGGACUGAGUUUUUAGAGCUAGAAAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU-3';

[0021] 5'-CAUCGACAAGAAAGGGACUGAGUUUUUAGAGGAGAGUAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU-3';

[0022] 5'-AUCGACAAGAAAGGGACUGAGUUUUUAGAGCUAGAAAUAGCAAGUAAAAUAAGGCUAGUCCGUUA

UCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU-3' ;

[0023] 5'-UCGACAAGAAAGGGACUGAGUUUUAGAGGUGAGAAAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU-3' ;以及

[0024] 5'-CGACAAGAAAGGGACUGAGUUUUAGAGGUGAGAAAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU-3' 。

[0025] 在另一方面,本发明的特征在于蛋白质核酸复合物,其包含本文所述的前述方面或实施方案中任一项的碱基编辑器和指导RNA。

[0026] 在本文所述的本发明的上述方面或任何其他方面的各个实施方案中,于所述与A1AD相关的SNP处的A • T至G • C改变将A1AT多肽中的赖氨酸改变为谷氨酸。在以上描述的方面或本文描述的本发明的任何其他方面的各个实施方案中,与A1AD相关的SNP导致表达在氨基酸位置342处具有赖氨酸的A1AT多肽。在另一个实施方案中,碱基编辑器的校正是以谷氨酸替换与A1AD相关的A1AT多肽的位置342处的赖氨酸。在以上描述的本发明的上述方面或本发明的任何其他方面的各种实施方案中,多核苷酸可编程DNA结合结构域是修饰的化脓链球菌Cas9 (SpCas9) 或其变体。

[0027] 在以上描述的本发明的上述方面或本发明的任何其他方面的各种实施方案中,多核苷酸可编程DNA结合结构域包含具有改变的原间隔物相邻基序 (PAM) 特异性的修饰的SpCas9。在上述方面或本文所述的本发明的任何其他方面的各种实施方案中,修饰的SpCas9对核酸序列5' -AGC-3' 具有特异性。在本文所述的本发明的上述方面或任何其他方面的各个实施方案中,修饰的SpCas9包含氨基酸取代D1332A和D1135M、S1136Q、G1218K、E1219F、D1332A、R1335E和T1337R中的一个或多个,或相应的氨基酸其替代。在以上描述的本发明的上述方面或本发明的任何其他方面的各种实施方案中,多核苷酸可编程DNA结合结构域包含SpCas9的变体,其具有改变的原间隔物相邻基序 (PAM) 特异性。在本文描述的上述方面或本发明的任何其他方面的各种实施方案中,SpCas9的变体对核酸序列5' -NGC-3' 具有特异性。在本文所述的本发明的上述方面或任何其他方面的各种实施方案中,修饰的SpCas9包含氨基酸取代D1135M、S1136Q、G1218K、E1219F、A1322R、D1332A、R1335E和T1337R,或其相应的氨基酸取代。在本文描述的上述方面或本发明的任何其他方面的各种实施方案中,多核苷酸可编程DNA结合结构域是核酸酶失活或切口酶变体。在以上描述的本发明的上述方面或本发明的任何其他方面的各个实施方案中,切口酶变体包含氨基酸取代D10A或其相应的氨基酸取代。在以上描述的本发明的上述方面或本发明的任何其他方面的各种实施方案中,腺苷脱氨酶结构域能够使脱氧核糖核酸 (DNA) 中的腺苷脱氨。在本文所述的本发明的上述方面或任何其他方面的各个实施方案中,腺苷脱氨酶是自然界中不存在的修饰的腺苷脱氨酶。在本文所述的本发明的上述方面或任何其他方面的各种实施方案中,腺苷脱氨酶是TadA脱氨酶 (例如,TadA*7.10)。在本文所述的本发明的上述方面或任何其他方面的各个实施方案中,一种或多种指导RNA包含CRISPR RNA (crRNA) 和反式编码的小RNA (tracrRNA),其中crRNA包含与SERPINA1核酸序列互补的核酸序列,其中SERPINA1核酸序列包含与A1AD相关的SNP。在上述方面的各个实施方案中,碱基编辑器与单链指导RNA (sgRNA) 复合,所述单链指导RNA (sgRNA) 包含与SERPINA1核酸序列互补的核酸序列,其中SERPINA1核酸序列包含与A1AD相关的SNP。

[0028] 在另一方面,本文提供一种用于校正基因中的致病性单核苷酸多态性 (SNP) 的碱

基编辑器系统,其中所述碱基编辑器系统包括(a)碱基编辑器,其包括:(i)多核苷酸可编程DNA结合域,以及(ii)能够使病原性SNP或其互补核碱基脱氨的脱氨酶域;(b)与多核苷酸可编程的DNA结合结构域结合的指导多核苷酸,其中所述指导多核苷酸将碱基编辑器靶向至靶多核苷酸序列,所述靶多核苷酸序列的至少一部分位于所述基因或其反向互补序列中;其中使病原体SNP或其互补核碱基脱氨基导致病原体SNP转化为其野生型等位基因,从而校正病原性突变,例如本文表3A和3B中所列。

[0029] 在另一方面,本文提供一种用于校正基因中的致病性单核苷酸多态性(SNP)的方法,其中所述方法包括:与靶核苷酸序列接触,所述靶核苷酸序列的至少一部分位于所述基因或其反向互补序列中,其碱基编辑器包含:(i)多核苷酸可编程的DNA结合结构域,与将碱基编辑器靶向目标多核苷酸序列的指导多核苷酸结合,其至少一部分位于基因或其反向互补序列中(ii)能够使病原性SNP或其互补核碱基脱氨的脱氨酶域;以及通过在将碱基编辑器靶向靶核苷酸序列后使病原性SNP或其互补核碱基脱氨来编辑病原性SNP,其中使病原性SNP或其互补核碱基脱氨导致病原性SNP转化为其野生型等位基因,从而校正病原性突变,例如本文表3A或表3B中所列。

[0030] 在另一方面,本文提供一种通过校正基因中的致病性单核苷酸多态性(SNP)来治疗受试者的遗传疾病的方法,其中所述方法包括将碱基编辑器或编码所述碱基编辑器的多核苷酸施用于有此需要的受试者,其中所述碱基编辑器包含:(i)多核苷酸可编程的DNA结合结构域,和(ii)能够使病原性SNP或其补体核碱基脱氨基的脱氨酶结构域;给予所述对象的指导多核苷酸,其中所述指导多核苷酸将所述碱基编辑器靶向靶核苷酸序列,所述靶核苷酸序列的至少一部分位于所述基因或其反向互补序列中;以及通过在将碱基编辑器靶向靶核苷酸序列后使病原性SNP或其互补核碱基脱氨来编辑病原性SNP,其中使病原性SNP或其互补核碱基脱氨导致病原性SNP转化为其野生型等位基因,从而校正表3A或3B中列出的致病突变,并治疗遗传性疾病。

[0031] 本文提供一种通过校正细胞、组织或器官的基因中的致病性单核苷酸多态性(SNP)来产生用于治疗有此需要的受试者的遗传性疾病的细胞、组织或器官的方法,其中所述方法包括:使细胞、组织或器官与碱基编辑器接触,其中所述碱基编辑器包括:(i)多核苷酸可编程的DNA结合结构域,和(ii)能够使病原性SNP或其互补核碱基脱氨的脱氨酶结构域;以及使细胞、组织或器官与指导多核苷酸接触,其中所述指导多核苷酸将碱基编辑器靶向至靶核苷酸序列,所述靶核苷酸序列的至少一部分位于所述基因或其反向互补序列中;以及通过在将碱基编辑器靶向靶核苷酸序列后使病原性SNP或其互补核碱基脱氨来编辑病原性SNP,其中使病原性SNP或其互补核碱基脱氨导致病原性SNP转化为其野生型等位基因,从而校正诸如表3A或3B所列的致病突变,并产生用于治疗遗传疾病的细胞、组织或器官。在一些实施方案中,所述方法进一步包括将细胞、组织或器官施用给受试者。在一些实施方案中,细胞、组织或器官对于受试者是自体的。在一些实施方案中,细胞、组织或器官对受试者是同种异体的。在一些实施方案中,细胞、组织或器官对受试者是异种的

[0032] 在一些实施方案中,致病性SNP与斯特格病变(Stargardt disease)相关;任选地,致病性SNP在ABCA4基因中;以及,任选地,病原性突变包括A1038V、L541P、G1961E或其组合。在一些实施方案中,致病性SNP与弹性假性黄瘤相关。任选地,致病性SNP在ABCC6基因中;以及,任选地,病原突变包括R1141*。在一些实施方案中,致病性SNP与中链酰基辅酶A脱氢酶

缺乏症相关;任选地,致病性SNP在ACADM基因中;以及,任选地,病原性突变包括K329E。在一些实施方案中,致病性SNP与严重的联合免疫缺陷相关;任选地,所述致病性SNP在ADA基因中;以及,任选地,病原性突变包括G216R、Q3*或其组合。

[0033] 在一些实施方案中,致病性SNP与原发性低血尿症有关。任选地,所述致病性SNP在AGXT基因中;以及,任选地,病原性突变包括G170R。在一些实施方案中,致病性SNP与常染色体隐性遗传性高胆固醇血症有关;任选地,所述致病性SNP在ARH基因中;以及,任选地,致病突变包括Q136*。在一些实施方案中,致病性SNP与异染色质的白细胞营养不良相关。任选地,所述病原性SNP在ARSA基因中;以及,任选地,致病性突变包含P426L、c.459+1G>A或其组合。在一些实施方案中,致病性SNP与马特鲁兹-拉米综合征(Marteauz-Lamy Syndrome, MSPVI)相关;任选地,所述致病性SNP在ARSB基因中;以及,任选地,致病突变包括Y210C。在一些实施方案中,致病性SNP与I型瓜氨酸血症相关;例如,I型瓜氨酸血症。任选地,所述致病性SNP在ASS基因中;以及,任选地,致病突变包括G390R。在一些实施方案中,致病性SNP与Darier氏病有关。任选地,所述致病性SNP在ATP2A2基因中;以及,任选地,致病突变包括N767S。

[0034] 在一些实施方案中,致病性SNP与经典高半胱氨酸尿症相关;任选地,所述致病性SNP在CBS基因中;任选地,致病突变包括G307S、T191M或其组合。在一些实施方案中,致病性SNP与囊性纤维化有关;任选地,所述致病性SNP在CFTR基因中;任选地,致病突变包括G551D、W1282*、R553*、R117H或其组合。在一些实施方案中,致病性SNP与脉络膜增生症相关;任选地,所述致病性SNP在CHM基因中;以及,任选地,致病突变包括R293*、R270*、A117A或其组合。在一些实施方案中,致病性SNP与神经元类脂褐质沉积病(Neuronal ceroid lipofuscinosis, NCL)相关。任选地,所述致病性SNP在CLN2基因中;以及,任选地,致病突变包括R208*。在一些实施方案中,致病性SNP与常染色体显性遗传性聋相关;任选地,所述病原性SNP在COCH基因中;以及,任选地,致病突变包括G88E。在一些实施方案中,致病性SNP与肉碱棕榈酰转移酶II缺乏症相关;任选地,所述致病性SNP在CPT2基因中;以及,任选地,致病突变包括S113L。

[0035] 在一些实施方案中,致病性SNP与胱氨酸病有关。任选地,所述致病性SNP在CTNS基因中;以及,任选地,致病突变包括W138*。在一些实施方案中,病原性SNP与常染色体隐性遗传性耳聋有关;任选地,所述致病性SNP在CX30基因中;以及,任选地,致病突变包括T5M。在一些实施方案中,病原性SNP与常染色体隐性遗传性耳聋有关;任选地,所述病原性SNP在DFNB59基因中;以及,任选地,病原性突变包括R183W。在一些实施方案中,致病性SNP与分离的丙种球蛋白血症相关;任选地,致病性SNP在E47基因中;病原性突变包括E555K。在一些实施方案中,致病性SNP与先天性因子XI缺乏症相关;任选地,所述致病性SNP在F11基因中;以及,任选地,病原性突变包括E117*、F283L或其组合。在一些实施方案中,致病性SNP与先天性因子V缺乏症相关;任选地,病原性SNP在F5基因中;以及,任选地,病原性突变包括R506Q、R534Q或其组合。在一些实施方案中,致病性SNP与先天性因子VII缺乏症相关;任选地,致病性SNP在F7基因中;以及,任选地,病原性突变包括A294V、C310F、R304Q、Q100R或其组合。

[0036] 在一些实施方案中,致病性SNP与A型血友病A相关;任选地,致病性SNP在F8基因中;以及,任选地,病原性突变包括R2169H、R1985Q、R2178C、R550C或其组合。在一些实施方案中,病原性SNP与B型血友病B有关;任选地,病原性SNP在F9基因中;以及,任选地,病原性

突变包括T342M、R294Q、R43Q、R191H、G106S、A279T、R75*、R294*、R379Q或其组合。在一些实施方案中,致病性SNP与1型酪氨酸血症相关;任选地,所述致病性SNP在FAH基因中;以及,任选地,病原性突变包括P261L。在一些实施方案中,致病性SNP与常染色体显性低磷酸盐血症性病相关;任选地,所述病原性SNP在FGF23基因中;以及,任选地,病原性突变包括R176Q。

[0037] 在一些实施方案中,致病性SNP与冯吉尔克病(von Gierke disease)有关;任选地,所述致病性SNP在G6PC基因中;以及,任选地,病原突变包括Q347*。在一些实施方案中,致病性SNP与地中海G6PD缺乏有关;任选地,致病性SNP在G6PD基因中;以及,任选地,病原性突变包括S188D。在一些实施方案中,病原性SNP与莫尔基奥综合症(MPSIVA)相关;任选地,所述致病性SNP在GALNS基因中;以及,任选地,病原性突变包括R386C。在一些实施方案中,致病性SNP与经典半乳糖血症相关;任选地,致病性SNP在GALT基因中;以及,任选地,病原性突变包括Q188R。

[0038] 在一些实施方案中,致病性SNP与高雪氏(Gaucher)症相关;任选地,所述病原性SNP在GBA基因中;以及,任选地,病原性突变包括N370S、L444P或其组合。在一些实施方案中,致病性SNP与戊二酰辅酶A脱氢酶缺乏有关;任选地,所述致病性SNP在GCDH基因中;以及,任选地,病原性突变包括R138G、M263V、R402W或其组合。在一些实施方案中,致病性SNP与甘氨酸脑病有关;任选地,所述病原性SNP在GLDC基因中;以及,任选地,病原性突变包括A389V、G771R、T269M或其组合。在一些实施方案中,致病性SNP与锥状-杆状细胞营养不良有关;任选地,致病性SNP在GUCY2D基因中;以及,任选地,病原性突变包括R838C。在一些实施方案中,致病性SNP与Sly综合征(MPSVII)相关;可选地,所述致病性SNP在GUSB基因中;以及,任选地,病原性突变包括L175F。

[0039] 在一些实施方案中,致病性SNP与镰状细胞病有关;任选地,所述病原性SNP在HBB基因中;以及,任选地,病原突变包括E26K;E7K;c.-138C>T;IVS2,654C>T;或其组合。在一些实施方案中,致病性SNP与间歇性紫质症相关;任选地,所述致病性SNP在HMBS基因中;以及,任选地,病原性突变包括R173W。在一些实施方案中,致病性SNP与莱希-尼亨症候群(Lesch-Nyhan syndrome)相关;任选地,所述致病性SNP在HPRT1基因中;以及,任选地,病原性突变包括R51*、R170*或其组合。在一些实施方案中,致病性SNP与亨特综合征相关;任选地,病原性SNP在IDS基因中;以及,任选地,病原性突变包括R88C、G374G或其组合。在一些实施方案中,致病性SNP与贺勒氏症(Hurler)(MPS1)相关。任选地,病原性SNP在IDUA基因中;以及,任选地,病原突变包括Q70*。

[0040] 在一些实施方案中,致病性SNP与视网膜色素病变相关;任选地,所述致病性SNP在IMPDH1基因中;以及,任选地,病原性突变包括D226N。在某些实施方案中,致病性SNP与安德森氏症候群(Andersen-Tawil syndrome)有关;任选地,所述病原性SNP在KCNJ2基因中;以及,任选地,病原性突变包括R218W。在一些实施方案中,致病性SNP与Meesmann上皮角膜营养不良有关;因此,SNP与Meesmann上皮角膜营养不良有关。任选地,所述病原性SNP在KRT12基因中;以及,任选地,病原性突变包括L132P。在一些实施方案中,致病性SNP与帕金森氏症(Parkinson's disease)相关;任选地,所述致病性SNP在LRRK2基因中;以及,任选地,病原性突变包括G2109S。在一些实施方案中,致病性SNP与雷特氏症(Rett syndrome)有关;任选地,所述致病性SNP在MECP2基因中;以及,任选地,病原性突变包括R106W、R133C、R306C、R168*、R255*或其组合。在一些实施方案中,致病性SNP与圣菲利柏氏症B型(Sanfilippo

Syndrome B, MPSIIIB) 有关; 任选地, 病原性SNP在NAGLU基因中; 以及, 任选地, 病原性突变包括R297*、Y140C或其组合。

[0041] 在一些实施方案中, 病原性SNP与CADASIL综合征相关; 任选地, 病原性SNP在NOTCH3基因中; 以及, 任选地, 病原性突变包括R90C、R141C或其组合。在一些实施方案中, 致病性SNP与蓝锥单色性相关; 任选地, 所述致病性SNP在OPN1LW基因中; 以及, 任选地, 病原性突变包括C203R。在一些实施方案中, 致病性SNP与苯丙酮尿症有关; 任选地, 所述病原性SNP在PAH基因中; 以及, 任选地, 病原性突变包括R408W、I65T、R261Q、IVS10-11G>A或其组合。在一些实施方案中, 致病性SNP与尤塞氏综合症1F型 (Usher syndrome type 1F) 相关; 任选地, 致病性SNP在PCDH15基因中; 以及, 任选地, 病原性突变包括R245*。在一些实施方案中, 致病性SNP与视网膜色素病变相关; 任选地, 病原性SNP在PDE6A基因中; 以及, 任选地, 病原性突变包括V685M、D670G或其组合。

[0042] 在一些实施方案中, 致病性SNP与Pendred氏综合症相关; 任选地, 所述病原性SNP在PDS基因中; 以及, 任选地, 病原性突变包括L236P; c. 1001+1G>A; IVS8, +1G>A或其组合。在一些实施方案中, 病原性SNP与异位型紫质症 (Variegate porphyria) 相关; 任选地, 所述致病性SNP在PPOX基因中; 以及, 任选地, 病原突变包括R59W。在一些实施方案中, 病原性SNP与1型神经元类脂褐质沉积病有关; 任选地, 病原性SNP在PPT1基因中; 以及, 任选地, 病原性突变包括R151*。在一些实施方案中, 致病性SNP与库贾氏病 (Creutzfeldt-Jakob Disease, CJD) 有关; 任选地, 所述致病性SNP在PRNP基因中; 以及, 任选地, 病原性突变包括M129V、P102L、D178N或其组合。在一些实施方案中, 致病性SNP与视网膜色素病变相关; 任选地, 所述病原性SNP在PRPF3基因中; 以及, 任选地, 病原性突变包括T494M。在一些实施方案中, 致病性SNP与视网膜色素病变相关; 任选地, 致病性SNP在PRPF8基因中; 以及, 任选地, 病原性突变包括H2309R。

[0043] 在一些实施方案中, 致病性SNP与遗传性慢性胰腺炎相关; 任选地, 病原性SNP在PRSS1基因中; 以及, 任选地, 病原突变包括R122H。在一些实施方案中, 致病性SNP与视网膜色素病变相关; 任选地, 所述致病性SNP在RHO基因中; 以及, 任选地, 病原性突变包括P347L、D190N或其组合。在一些实施方案中, 致病性SNP与视网膜色素病变相关; 任选地, 所述致病性SNP在RP1基因中; 以及, 任选地, 病原性突变包括R667*。在一些实施方案中, 致病性SNP与2型莱伯氏先天性黑蒙症有关; 任选地, 致病性SNP在RPE65基因中; 以及, 任选地, 病原性突变包括R44*; IVS1, G-A, +5; 或其组合。在一些实施方案中, 致病性SNP与戴布氏贫血 (Diamond-Blackfan anaemia) 有关; 任选地, 所述致病性SNP在RPS19基因中; 以及, 任选地, 病原突变包括R62Q。

[0044] 在一些实施方案中, 致病性SNP与X-连接的视网膜骨分裂症相关; 任选地, 所述病原性SNP在视黄醛毒素 (RS1) 基因中; 以及, 任选地, 病原性突变包括R102W、R141C或其组合。在一些实施方案中, 致病性SNP与A1AD相关; 任选地, 致病性SNP在SERPINA1基因中; 以及, 任选地, 病原性突变包括E342K、R48C (R79C) 或其组合。在一些实施方案中, 致病性SNP与圣菲利柏氏症A型 (Sanfilippo Syndrome A, MPSIIIA) 相关; 任选地, 所述致病性SNP在SGSH基因中; 以及, 任选地, 病原性突变包括R74C。在一些实施方案中, 致病性SNP与A型尼曼匹克症 (Niemann-Pick disease type A) 相关; 任选地, 所述致病性SNP在SMPD1基因中; 以及, 任选地, 病原性突变包括L302P。

[0045] 在一些实施方案中,致病性SNP与常染色体显性帕金森氏病有关;任选地,病原性SNP在SNCA基因中;以及,任选地,病原突变包括A53T。在一些实施方案中,致病性SNP与家族性肌萎缩性侧索硬化症(ALS)有关。任选地,所述病原性SNP在超氧化物歧化酶1(SOD1)基因中;以及,任选地,病原性突变包括A4V、H46R、G37R或其组合。在一些实施方案中,致病性SNP与常染色体显性遗传性聋相关;任选地,致病性SNP在TECTA基因中;以及,任选地,病原性突变包括Y1870C。在一些实施方案中,病原性SNP与常染色体隐性遗传性耳聋有关;任选地,所述致病性SNP在TMC1基因中;病原性突变包括Y182C。在一些实施方案中,致病性SNP与ATTR淀粉样变性有关;任选地,所述病原性SNP在TTR基因中;以及,任选地,病原性突变包括V50M/V30M。在一些实施方案中,病原性SNP与视网膜色素病变/尤塞氏综合症1C型相关;任选地,所述致病性SNP在USH1C基因中;以及,任选地,病原性突变包括V72V。

[0046] 在一些实施方案中,致病性SNP与视网膜色素病变相关;任选地,致病性SNP在USH2a基因中;以及,任选地,病原性突变包括C759F。在一些实施方案中,致病性SNP与肌小管肌病有关;任选地,所述致病性SNP在MTM1基因中;以及,任选地,病原性突变包括c.1261-10A>G。

[0047] 在一些实施方案中,本文提供的任何碱基编辑器系统或方法可以进一步包含用于编辑另外的核碱基的第二指导多核苷酸。在一些实施方案中,另外的核碱基不在基因中。在一些实施方案中,另外的核碱基位于基因中。在一些实施方案中,另外的核碱基位于蛋白质编码区。在一些实施方案中,另外的核碱基位于蛋白质非编码区。在一些实施方案中,蛋白质非编码区是基因调节组件。在一些实施方案中,脱氨酶结构域是胞苷脱氨酶结构域或腺苷脱氨酶结构域。在一些实施方案中,脱氨酶结构域是胞苷脱氨酶结构域。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,腺苷脱氨酶结构域能够使脱氧核糖核酸(DNA)中的腺苷脱氨。在一些实施方案中,指导多核苷酸包含核糖核酸(RNA)或脱氧核糖核酸(DNA)。在一些实施方案中,指导多核苷酸包含CRISPR RNA(crRNA)序列、反式激活的CRISPR RNA(tracrRNA)序列或其组合。

[0048] 在一些实施方案中,本文提供的任何碱基编辑器系统或方法可进一步包含第二指导多核苷酸。在一些实施方案中,第二指导多核苷酸包含核糖核酸(RNA)或脱氧核糖核酸(DNA)。在一些实施方案中,第二指导多核苷酸包含CRISPR RNA(crRNA)序列、反式激活的CRISPR RNA(tracrRNA)序列或其组合。在一些实施方案中,第二指导多核苷酸将碱基编辑器靶向第二靶核苷酸序列。

[0049] 在一些实施方案中,在本文提供的任何碱基编辑器系统或方法中,多核苷酸可编程DNA结合结构域包含Cas9结构域、Cpf1结构域、CasX结构域、CasY结构域、Cas12b/C2c1结构域或Cas12c/C2c3域。在一些实施方案中,多核苷酸可编程DNA结合结构域是核酸酶死亡的。在一些实施方案中,多核苷酸可编程的DNA结合结构域是切口酶。在一些实施方案中,多核苷酸可编程DNA结合结构域包含Cas9结构域。在一些实施方案中,Cas9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。在一些实施方案中,Cas9结构域包含Cas9切口酶。在一些实施方案中,多核苷酸可编程的DNA结合结构域是设计或修饰的多核苷酸可编程的DNA结合结构域。

[0050] 在一些实施例中,本文提供的任何碱基编辑器系统或方法可以进一步包括第二碱基编辑器。在一些实施例中,第二碱基编辑器包括与碱基编辑器不同的脱氨酶域。

[0051] 在一些实施方案中,在本文提供的任何方法中,编辑导致少于20%的插入缺失形成。在一些实施例中,编辑导致少于15%的插入缺失形成。在一些实施例中,编辑导致少于10%的插入缺失形成。在一些实施例中,编辑导致少于5%的插入缺失形成。在一些实施例中,编辑导致少于4%的插入缺失形成。在一些实施例中,编辑导致少于3%的插入缺失形成。在一些实施例中,编辑导致少于2%的插入缺失形成。在一些实施例中,编辑导致少于1%的插入缺失形成。在一些实施例中,编辑导致少于0.5%的插入缺失形成。在一些实施例中,编辑导致少于0.1%的插入缺失形成。在一些实施例中,编辑不导致易位。

[0052] 一方面,本发明提供了一种编辑G6PC多核苷酸的方法,所述G6PC多核苷酸包含与糖原贮积病1a型(GSD1a)相关的单核苷酸多态性(SNP),所述方法包括使G6PC多核苷酸与碱基编辑器接触,所述碱基编辑器与一种或多种指导多核苷酸复合,其中碱基编辑器包含多核苷酸可编程DNA结合结构域和腺苷脱氨酶结构域,且其中一种或多种指导多核苷酸靶向碱基编辑器,以造成与GSD1a相关的SNP的A·T到G·C改变。在一个实施方案中,与糖原贮积病1a型(GSD1a)相关的SNP处的A·T至G·C改变将谷氨酰胺(Q)改变为非谷氨酰胺(X)氨基酸。在本文描述的上述方面或本发明的任何其他方面的各个实施方案中,与糖原贮积病1a型(GSD1a)相关的SNP处的A·T至G·C改变将在G6PC多肽中精氨酸(R)改变为非精氨酸(X)。

[0053] 在上述方面或本文描述的本发明的任何其他方面的各个实施方案中,与GSD1a相关的SNP导致在位置347处具有非谷氨酰胺(X)氨基酸或在位置83处的非精氨酸(X)氨基酸的G6PC多肽的表达。在一个实施方案中,碱基编辑器校正是以非谷氨酰胺氨基酸(X)取代在位置347处的谷氨酰胺。在另一个实施方案中,碱基编辑器的校正是用非精氨酸氨基酸(X)代替位置83处的精氨酸。

[0054] 在以上描述的本发明的上述方面或本发明的任何其他方面的各种实施方案中,多核苷酸可编程DNA结合结构域是修饰的化脓链球菌Cas9(SpCas9)或其变体。在以上描述的本发明的上述方面或本发明的任何其他方面的各种实施方案中,多核苷酸可编程DNA结合域包含具有改变的原间隔物相邻基序(PAM)特异性的修饰的SpCas9。在一个实施方案中,修饰的SpCas9对核酸序列5'-NGA-3'或5'-NGG-3'具有特异性。在本文所述的本发明的上述方面或任何其他方面的各种实施方案中,腺苷脱氨酶为ABE7.10。

[0055] 一方面,通过向细胞或其祖细胞中引入碱基编辑器,编码所述碱基编辑器的多核苷酸到细胞中来生产细胞,其中所述碱基编辑器包含多核苷酸可编程DNA结合域和腺苷脱氨酶域;以及一种或多种靶向碱基编辑器的指导多核苷酸,以造成与糖原贮积症1a型(GSD1a)相关的SNP的A·T到G·C改变。在本文所述的本发明的上述方面或任何其他方面的各个实施方案中,所述细胞是肝细胞,肝细胞前体或iPSc衍生的肝细胞。

[0056] 在以上描述的本发明的上述方面或本发明的任何其他方面的各种实施方案中,细胞来自患有GSD1a的受试者。在以上描述的本发明的上述方面或本发明的任何其他方面的各种实施方案中,细胞具有Q347X突变。在本文描述的上述方面或本发明的任何其他方面的各种实施方案中,与GSD1a相关的SNP处的A·T至G·C改变将谷氨酰胺改变为非谷氨酰胺(X)氨基酸。在本文描述的上述方面或本发明的任何其他方面的各种实施方案中,与GSD1a相关的SNP处的A·T至G·C将G6PC多肽中的精氨酸改变为非精氨酸(X)氨基酸。在以上描述的本发明的上述方面或本发明的任何其他方面的各个实施方案中,与GSD1a相关的SNP导致在位置347处具有非谷氨酰胺(X)氨基酸或在位置83处的非精氨酸(X)氨基酸的G6PC多肽的

表达。

[0057] 在一个方面,本发明提供一种对此需要的受试者中治疗1a型糖原贮积症(GSD1a)或冯吉尔克病的方法,所述方法包括对受试者施用本文描述的本发明的上述方面或任何其他方面的各种实施方案的细胞。

[0058] 在另一方面,本发明提供一种产生肝细胞或其祖细胞的方法,所述方法包括:(a) 引入包含与GSD1a相关的SNP的诱导性多能干细胞或肝细胞祖细胞、碱基编辑器或编码所述碱基编辑器的多核苷酸,其中所述碱基编辑器包含多核苷酸可编程核苷酸结合结构域和腺苷脱氨酶结构域;以及一种或多种指导多核苷酸,其中所述一种或多种指导多核苷酸靶向所述碱基编辑器,以造成与GSD1a相关的SNP的A·T至G·C改变;(b) 将诱导性多能干细胞或肝细胞祖细胞分化为肝细胞。在另一方面,所述方法包括将诱导性多能干细胞分化为肝细胞或其祖细胞。在各种实施方案中,步骤(a)的诱导性多能干细胞包含Q347X突变。

[0059] 在本文所述的本发明的上述方面或任何其他方面的各种实施方案中,所述肝细胞祖细胞获自患有GSD1a的受试者。在各种实施方案中,肝细胞或肝细胞祖细胞是哺乳动物细胞或人细胞。在以上描述的本发明的上述方面或本发明的任何其他方面的各种实施方案中,与GSD1a相关的SNP处的A·T至G·C改变将谷氨酰胺改变为非谷氨酰胺(X)氨基酸或将精氨酸改变为G6PC多肽中的非精氨酸(X)氨基酸。在各种实施方案中,与GSD1a相关的SNP导致在位置347具有非谷氨酰胺(X)氨基酸的G6PC多肽的表达。在各种实施方案中,与GSD1a相关的SNP导致在位置83处的具有精氨酸(X)氨基酸的G6PC多肽的表达。在各种实施方案中,与GSD1a相关的SNP用非谷氨酰胺(X)氨基酸取代谷氨酰胺。在各种实施方案中,与GSD1a相关的SNP用非精氨酸(X)氨基酸替代精氨酸。在上述方面或本文描述的本发明的任何其他方面的各种实施方案中,所述细胞被选择用于与GSD1a相关的SNP的A·T至G·C改变。

[0060] 一方面,本发明提供一种编辑IDUA多核苷酸的方法,所述IDUA多核苷酸包含与1型粘多糖贮积病(MPS1)相关的单核苷酸多态性(SNP),所述方法包括使IDUA多核苷酸与碱基编辑器接触,且所述碱基编辑器一种或多种指导多核苷酸复合,其中碱基编辑器包含多核苷酸可编程DNA结合结构域和腺苷脱氨酶结构域,且其中一种或多种指导多核苷酸靶向碱基编辑器,以造成与MPS1相关的SNP的A·T至G·C改变。在一个实施方案中,多核苷酸可编程DNA结合结构域是修饰的化脓性链球菌Cas9(SpCas9)或其变体。在另一个实施方案中,多核苷酸可编程DNA结合结构域包含具有改变的原间隔物相邻基序(PAM)特异性的修饰的SpCas9。在以上描述的本发明的上述方面或本发明的任何其他方面的各种实施方案中,修饰的SpCas9对核酸序列5'-NGG-3'具有特异性。在本文所述的本发明的上述方面或任何其他方面的各种实施方案中,腺苷脱氨酶为ABE7.10。在各种实施方案中,指导多核苷酸包含人核酸序列ACTCTaGGCAGAGGTCTCAA AGG。在各种实施方案中,指导多核苷酸包含小鼠核酸序列GCTCTaGGCCGAAGTGTCGC AGG。

[0061] 一方面,通过向细胞或其祖细胞中引入碱基编辑器,编码所述碱基编辑器的多核苷酸到细胞中来生产细胞,其中所述碱基编辑器包含多核苷酸可编程DNA结合结构域和腺苷脱氨酶结构域;以及一种或多种靶向碱基编辑器的指导多核苷酸,以造成与1型粘多糖贮积病(MPS1)相关的SNP的A·T到G·C改变。在各种实施方案中,细胞是干细胞、干细胞前体或诱导性多能干细胞(iPSC)。在以上描述的本发明的上述方面或本发明的任何其他方面的各种实施方案中,细胞来自患有MPS1的受试者。

[0062] 另一方面,本发明提供一种对此需要的受试者治疗MPS1的方法,所述方法包括对受试者施用上述方面或本文所述的本发明的任何其他方面的细胞。

附图说明

[0063] 在所附权利要求中具体阐述本申请的特征。通过参考下面的详细描述,可以更好地理解本发明的特征和优点,所述详细描述阐述说明性实施例,在实施例中应用本申请的原理及其附图:

[0064] 图1是比较健康受试者和抗胰蛋白酶缺乏症(A1AD)患者的示意图。在健康受试者中, α -1抗胰蛋白酶(A1AT)保护肺部免受蛋白酶损害,肝脏将 α -1抗胰蛋白酶释放到血液中。在患有A1AD的患者中,缺乏正常功能的A1AT蛋白会导致肺组织损伤。此外,肝细胞中异常A1AT的积累会导致肝硬化。

[0065] 图2显示不同基因型(正常(MM); α -1抗胰蛋白酶缺乏症的杂合子携带者(MZ、SZ);以及纯合子缺乏症(SS、ZZ))的血清 α -1抗胰蛋白酶(A1AT)的典型水平范围。所述血清 α -1抗胰蛋白酶(AAT)浓度在左侧“y”轴以 μ M表示,这在文献中很常见。右侧的“y”轴显示临床实验室和其他测量技术(浊度法或放射免疫扩散法)常见报告的血清AAT浓度到mg/dL单位的近似换算。

[0066] 图3A至3C显示碱基编辑靶序列,以及与精确校正编码A1AT蛋白的SERPINA1基因中的致病突变有关的图3A显示针对编码A1AT的SERPINA1基因中的突变的精确的校正碱基编辑策略。A7(“目标A”)可被编辑以恢复野生型(WT)表型。在某些情况下,“A”核碱基A5/A7可被编辑以将氨基酸D341G引入A1AT蛋白中。在某些情况下,A7/A8可被编辑以将氨基酸E342G引入A1AT蛋白中。图3B提供一核酸序列,其显示了SERPINA1基因内靶标A核碱基的位置和编码的氨基酸,以及显示从表达野生型(WT)的HEK293T细胞分泌的A1AT水平(ng/ml)的图,或包含E342K、D341G或E342G的A1AT变体。图3C是显示野生型(WT)A1AT蛋白中的弹性蛋白酶活性与含有E342K或D341G的A1AT变体中的弹性蛋白酶活性的图。

[0067] 图4显示从TadA tRNA脱氨酶开始发展DNA脱氧腺苷脱氨酶的策略的示意图。显示的是大肠杆菌资料库,其中包含与dCas9融合的突变ecTadA(TadA*)基因的质粒资料库,以及一个需要靶向A•T到G•C突变以修复抗生素抗性基因的选择质粒。将来自存活的TadA*变异的突变导入到ABE体系结构中,用于在人类细胞中进行碱基编辑。

[0068] 图5提供一种显示SERPINA1基因内的靶“A”核碱基的位置和编码的氨基酸的核酸序列,以及显示SERPINA1基因中的位置A5或A7处的编辑百分比作为指导RNA长度的函数的图。

[0069] 图6A和6B描绘SpCas9突变体的资料库,SpCas9突变体产生以丰富Cas9的PAM相互作用(PI)结构域内的突变。可由所述资料库筛选以了解具有改变的PAM特异性的SpCas9。

[0070] 图7是显示出使用碱基编辑器的Q347X突变的校正策略的示意图,以利用NGG和NGA PAM识别序列在靶位点处转换A>G(突出显示)。精确的校正产生TAG>CAG(终止密码子>谷氨酰胺)转换。

[0071] 图8A和8B提供基于GSD1a iPSc衍生的肝细胞的成熟周期的转染时间表。图8A提供转染时间表的时间表,其显示了接种、转染和细胞收获的代表性时间点。图8B显示在第5天和第7天成熟的GSD1a iPSc衍生的肝细胞的图像。

[0072] 图9A和9B提供数据,且所述数据示出针对GSD1a的G6PC Q347X的碱基编辑精确校正。图9A提供核酸序列,其显示了G6PC基因内靶标和旁观者“A”核碱基的位置以及相应的NGG和NGA PAM序列,以及显示HEK293T细胞中G6PC Q347X对ABE-的碱基编辑效率百分比的图。使用NGA PAM或NGG PAM定位目标上的ABE-旁观者,插入/缺失和核酸酶-插入/缺失。图9B提供了图表,其显示了在患者iPSc衍生的肝细胞中,使用NGA PAM或NGG PAM在G6PC Q347X中对ABE-靶标上、ABE-旁观者、插入缺失和核酸酶插入缺失的碱基编辑效率。使用实线表示实验的平均值。

[0073] 图10提供核酸序列,其显示了G6PC基因中靶标和旁观者“A”核碱基的位置以及相应的GGA PAM序列,以及显示用于ABE-On靶标、ABE-Bystander和使用mRNA变体的插入的源自iPSc的肝细胞的患者G6PC Q347X的A>G碱基编辑百分比的图,。

[0074] 图11提供显示使用ABE7.10碱基编辑器在小鼠和人类IDUA基因中碱基编辑效率的百分比的图。

具体实施方式

[0075] 此处的描述和示例详细示出本申请的实施例。应理解的是,本申请不限于本文描述的特定实施例,因此可以变化。本领域技术人员将认识到,有许多变化和修改包含在本申请范围内。

[0076] 所有术语为如本领域技术人员将理解的。除非另有定义,否则本文所用的所有技术和科学术语具有与本领域技术人员本申请通常所理解相同含义。

[0077] 除非另有说明,否则本文公开的一些实施方案的实施采用免疫学、生物化学、化学、分子生物学、微生物学、细胞生物学、基因组学和重组DNA领域技术范围内的常规技术。参见例如Sambrook and Green, *Molecular Cloning: A Laboratory Manual*, 4th Edition (2012); the series *Current Protocols in Molecular Biology* (F.M. Ausubel, et al. eds.); the series *Methods In Enzymology* (Academic Press, Inc.), PCR 2: A Practical Approach (M.J. MacPherson, B.D. Hames and G.R. Taylor eds. (1995)), Harlow and Lane, eds. (1988); *Antibodies, A Laboratory Manual*, and *Culture of Animal Cells: A Manual of Basic Technique and Specialized Applications*, 6th Edition (R.I. Freshney, ed. (2010))。

[0078] 在此使用的章节标题仅出于组织目的,并且不应被解释为限制所描述的主题。

[0079] 尽管可以在单个实施例的上下文中描述本申请的各种特征,但也可以独立地或以任何合适的组合来提供特征。相反,为了清楚起见,本文中可以在单独的实施例的上下文中描述本申请,但本申请也可在单个实施例中实现。

[0080] 定义

[0081] 以下定义是本领域技术的补充,并且是针对本申请的,而不应归因于任何相关或不相关的情况,例如,任何共同拥有的专利或申请。尽管本文描述的那些方法或材料类似或等同的任何方法和材料都可在实践中用于测试本申请,本文仍描述优选的材料和方法。因此,本文所使用的术语仅出于描述特定实施例的目的,而无意于进行限制。

[0082] 除非另有定义,否则本文所用的所有技术和科学术语具有本领域技术人员通常理解的含义。下列参考文献为本领域技术人员提供本发明中使用的许多术语的一般定义:

Singleton等人,Dictionary of Microbiology and Molecular Biology(2nd ed.1994);The Cambridge Dictionary of Science and Technology(Walker ed.,1988);The Glossary of Genetics,5th Ed.,R.Rieger等人(eds.),Springer Verlag(1991);以及Hale&Marham,The Harper Collins Dictionary of Biology(1991)。

[0083] 在本申请中,单数的使用包括复数,除非另有明确说明。必须注意,在说明书中使用的单数形式“一”、“一个”和“所述”包括复数对象,除非上下文另有明确规定。在本申请中,除非另外说明,否则“或”的使用表示“和/或”。此外,术语“包括”以及其他形式(例如“包括(include)”、“包含(includes)”和“包含(included)”)的使用不是限制性的。

[0084] 如本说明书和权利要求书中所使用的,词语“包括”(以及包括的任何形式,例如“包括(comprise)”和“包括(comprises)”),“具有”(以及具有的任何形式,例如“具有(have)”和“具有(has)”),“包含”(以及任何形式的包含,例如“包含(includes)”和“包含(include)”)或“含”(以及任何形式的包含,例如“含有(contains)”和“包含(contain)”)是包含性的或开放式的,并且不排除其他未引用的元素或方法步骤。可预期地,可相对于本申请的任何方法或组成来实现本说明书中讨论的任何实施例,反之亦然。此外,本申请的组合物可用于实现本申请的方法。

[0085] 术语“大约(about)”或“大约(approximately)”是指本领域技术人员确定的特定值在可接受的误差范围内,这将部分取决于如何测量或确定所述值,即测量系统的局限性。例如,根据本领域的实践,“约”可以表示在1个或大于1个标准偏差之内。可替代地,“约”可以表示给定值的最高20%、最高10%、最高5%或最高1%的范围。可替代地,特别是关于生物系统或过程,所述术语可以意指在数值的一个数量级内,优选在5倍以内,更优选在2倍以内。在本申请和权利要求书中描述了特定值的情况下,除非另有说明,否则应假设术语“约”表示所述特定值在可接受的误差范围内。

[0086] 在说明书中对“一些实施例”、“一个实施例”、“一个实施例”或“其他实施例”的引用是指结合这些实施例描述的特定特征,结构或特性包括在至少一些实施例中,但不一定是本申请的所有实施例。

[0087] “腺苷脱氨酶”是指能够催化腺嘌呤或腺苷的水解脱氨的多肽或其片段。在一些实施方案中,脱氨酶或脱氨酶结构域是催化腺苷水解为肌苷或脱氧腺苷水解为脱氧肌苷的腺苷脱氨酶。在一些实施方案中,腺苷脱氨酶催化脱氧核糖核酸(DNA)中腺嘌呤或腺苷的水解脱氨。本文提供的腺苷脱氨酶(例如工程化腺苷脱氨酶,进化的腺苷脱氨酶)可以来自任何生物,例如细菌。

[0088] “试剂”是指任何小分子化合物、抗体、核酸分子或多肽或其片段。

[0089] “改善”是指减少、抑制、减弱、减少、阻止或稳定疾病的发展或进程。

[0090] “改变”是指基因或多肽的表达水平或活性的改变(增加或减少),如通过标准领域已知的方法(例如本文所述的方法)检测到的。如本文所用,改变包括表达水平的10%变化,优选25%的变化,更优选40%的变化,最优选50%或更大的表达水平的变化。

[0091] “类似物”是指不相同但具有相似功能或结构特征的分子。举例来说,多肽类似物保留了相应的天然存在的多肽的生物活性,同时具有某些生物化学修饰,相对于天然存在的多肽,其可以增强类似物的功能。这种生化修饰可增加类似物的蛋白酶抗性、膜通透性或半衰期,而不会改变例如配体结合。类似物可包括非天然氨基酸。

[0092] 本文中的“给药”是指对患者或受试者提供本文所述的一种或多种组合物。举例但不限于,可以通过静脉内(iv)注射、皮下(sc)注射、皮内(id)注射、腹膜内(ip)注射或肌肉内(im)注射进行组合物施用,可以采用一种或多种这样的途径。例如,肠胃外给药可以通过推注或随时间逐渐灌注。可替代地或同时地,通过口服途径施用。

[0093] “ α -1抗胰蛋白酶(A1AT)蛋白”是指与UniProt登录号P01009的氨基酸序列具有至少约95%的氨基酸序列同一性的多肽或其片段。在特定的实施方案中,A1AT蛋白相对于以下参考序列包含一种或多种改变。在一个特定的实施方案中,与A1AD相关的A1AT蛋白包含E342K突变。以下提供示例性A1AT氨基酸序列(>sp|P01009|A1AT_人类 α -1-抗胰蛋白酶OS=智人OX=9606GN=SERPINA1 PE=1SV=3):

[0094] MPSSVSWGILLLAGLCLVPVSLAEDPQGDAQKTDTSHHDDQDHPTFNKITPNLAFAFS

[0095] LYRQLAHQSNTNIFSPVSIATAFAMLSLGTKADTHDEILEGLNFNLTEIPEAQIHEGF

[0096] QELLRTLNPDSQLQLTTGNGLFLSEGLKLVDFLEDVKKLYHSEFTVNFGDTEEAQKQ

[0097] INDYVEKGTQGKIVDLVKELDRDITVFALVNYIFFKKGWERPFVEKDTTEEDFHVDQVTTV

[0098] KVPMMKRLGMFNIQHCKKLSSWVLLMKYLGNATAIFFLPDEGKLQHLENELTHDIITKFL

[0099] ENEDRRSASLHLPKLSITGTDLKSVLGQLGITKVFSGADLSGVTEEAPLKLSKAVHKA

[0100] VLTIDEKGTEAAGAMFLEAIPMSIPPEVKFNKPFVFLMIEQNTKSPLFMGKVVNPTQK。

[0101] “碱基编辑器(BE)”或“核碱基编辑器(NBE)”是所述指结合多核苷酸并具有核碱基修饰活性的试剂。在各种实施方案中,所述碱基编辑器包含核碱基修饰多肽(例如脱氨酶)和与指导多核苷酸(例如,指导RNA)结合的多核苷酸可编程核苷酸结合结构域。在各种实施方案中,所述试剂是生物分子复合物,其包含具有碱基编辑活性的蛋白质结构域,即能够修饰核酸分子(例如,A、T、C、G或U)中的碱基(例如,脱氧核糖核酸)。在一些实施方案中,所述多核苷酸可编程DNA结合结构域与脱氨酶结构域融合或连接。在一个实施方案中,所述试剂是融合蛋白,其包含具有碱基编辑活性的结构域。在另一个实施方案中,具有碱基编辑活性的蛋白结构域与指导RNA连接(例如,通过所述指导RNA上的RNA结合基序,以及与脱氨酶融合的RNA结合域)。在一些实施方案中,具有碱基编辑活性的结构域能够使核酸分子内的碱基脱氨。在一些实施方案中,所述碱基编辑器能够使DNA分子内的碱基脱氨。在一些实施方案中,所述碱基编辑器能够使DNA内的胞嘧啶(C)或腺苷(A)脱氨。在一些实施方案中,所述碱基编辑器是胞苷碱基编辑器(CBE)。在一些实施例中,所述碱基编辑器是腺苷碱基编辑器(ABE)。在一些实施方案中,所述腺苷脱氨酶从TadA进化而来。在一些实施方案中,所述多核苷酸可编程DNA结合结构域是CRISPR相关的(例如,Cas或Cpf1)酶。在一些实施方案中,所述碱基编辑器是与脱氨酶结构域融合的催化死亡的Cas9(dCas9)。在一些实施方案中,所述碱基编辑器是一与脱氨酶结构域融合的Cas9切口酶(ncas9)。在一些实施方案中,所述碱基编辑器是与碱基切除修复(BER)抑制剂融合。在一些实施方案中,所述碱基切除修复抑制剂是尿嘧啶DNA糖基化酶抑制剂(UGI)。在一些实施方案中,所述碱基切除修复抑制剂是肌苷碱基切除修复抑制剂。碱基编辑器的详细信息在国际PCT第PCT/2017/045381号专利申请(WO第2018/027078号国际专利申请)和国际第PCT/US2016/058344号专利申请案(WO第2017/070632号专利申请)中进行了描述,在此将其全部内容通过引用并入本文。另请参见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:A base editors with higher efficiency and product purity”Science Advances 3:eaa04774 (2017) 以及 Rees, H.A. 等人 “Base editing: precision chemistry on the genome and transcriptome of living cells.”Nat Rev Genet. 2018 Dec; 19(12): 770-788. doi:10.1038/s41576-018-0059-1, 其全部内容通过引用合并于此。

[0102] 例如, 在此描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具有以下核酸序列(8877个碱基对), (Addgene, Watertown, MA.; Komor AC, et al., 2017, Sci Adv., 30; 3(8): eaa04774. doi:10.1126/sciadv.aao4774), 如下所示。还包括有具与BE4核酸序列至少95%或更高同一性的多核苷酸序列。

```
1 atatgccaag tacgccccct attgacgtca atgacggtaa atggcccgcc tggcattatg
61 cccagtacat gaccttatgg gactttccta ctggcagta catctacgta ttagtcatcg
121 ctattaccat ggtgatgcgg ttgtggcagt acatcaatgg gcgtggatag cggtttgact
181 cacgggggatt tccaagtctc caccacattg acgtcaatgg gagtttgttt tggcaccaaa
241 atcaacggga ctttccaaaa tgcgtgaaca actccgcccc attgacgcaa atgggcggtg
301 ggcgtgtacg gtgggaggtc tatataagca gagctggttt agtgaaccgt cagatccgct
361 agagatccgc ggccgctaatac gactcact ataggagag cgcaccat gagtcagag
421 actggcccag tggctgtgga cccacattg agacggcgga tcgagccca tgagtttgag
481 gtattcttcg atccgagaga gctccgcaag gagacctgcc tgctttacga aattaattgg
541 gggggccggc actccatttg gcgacataca tcacagaaca ctaacaagca cgtcgaagtc
[0103] 601 aacttcatcg agaagttcac gacagaaaga tatttctgtc cgaacacaag gtgcagcatt
661 acctggtttc tcagctggag cccatgcggc gaatgtagta gggccatcac tgaattctg
721 tcaaggtatc cccacgtcac tctgtttatt tacatcgcaa ggctgtacca ccacgtgac
781 ccccgaatc gacaaggcct gcgggatttg atctctcag gtgtgactat ccaaattatg
841 actgagcagg agtcaggata ctgctggaga aactttgtga attatagccc gagtaatgaa
901 gcccactggc ctaggtatcc ccatctgtgg gtacgactgt acgttctga actgtactgc
961 atcatactgg gcctgcctcc ttgtctcaac attctgagaa ggaagcagcc acagctgaca
1021 ttctttacca tcgtcttca gtctgtcat taccagcgac tgccccaca cattctctgg
1081 gccaccgggt tgaaatctgg tggttcttct ggtggttcta gcggcagcga gactccggg
1141 acctcagagt cggccacacc cgaaagtct ggtggttct ctggtggttc tgataaaaag
1201 tattctattg gtttagccat cggcactaat tccgttggat gggctgtcat aaccgatgaa
```

[0104]

1261 taaaaagtac cttcaaagaa atttaaggtg ttggggaaca cagaccgtca ttcgattaaa
1321 aagaatctta tcggtgccct cctattcgat agtggcgaaa cggcagaggc gactcgcctg
1381 aaacgaaccg ctcggaagag gtatacacgt cgcaagaacc gaatatgtta cttacaagaa
1441 atttttagca atgagatggc caaagttgac gattcttct ttcaccgttt ggaagagtcc
1501 ttcttgtcg aagaggacaa gaaacatgaa cggcacccca tctttgaaa catagtagat
1561 gaggtggcat atcatgaaaa gtaccaacg attatcacc tcagaaaaaa gctagttagc
1621 tcaactgata aagcggacct gaggttaac tacttggctc ttgccatat gataaagttc
1681 cgtgggcact ttctattga gggtagtcta aatccggaca actcggatgt cgacaaactg
1741 ttcatccagt tagtacaac ctataatcag ttgtttgaag agaaccctat aaatgcaagt
1801 ggcgtggatg cgaaggctat tcttagcgcc cgcctctcta aatcccgcg gctagaaaaac
1861 ctgatgcac aattaccgg agagaagaaa aatgggtgt tcgtaacct tatagcgtc
1921 tcactaggcc tgacacaaa tttaagtcg aacttcgact tagctgaaga tgccaaattg
1981 cagcttagta aggacacgta cgaatgacgt ctgcacaac tactggcaca aattggagat
2041 cagtatgcgg acttatttt ggctgcaaaa aaccttagcg atgcaatcct cctatctgac
2101 atactgagag ttaatactga gattaccaag gcgccgttat ccgcttcaat gatcaaaagg
2161 tacgatgaac atcaccaaga cttgacact ctcaaggccc tagtccgtca gcaactgcct
2221 gagaaatata aggaaatatt ctttgatcag tcgaaaaacg ggtacgcagg ttatattgac
2281 ggcggagcga gtcaagagga attctacaag ttatcaaac ccatattaga gaagatggat
2341 gggacggaag agttgctgt aaaactcaat cgcgaagatc tactgcgaaa gcagcggact
2401 ttcgacaacg gtagcattcc acatcaaat cacttaggcg aattgcatgc tatacttaga
2461 aggaggagg attttatcc gtctctaaa gacaatcgtg aaaagattga gaaaatccta
2521 acctttcgca taccttacta tgtgggaccc ctggcccgag ggaactctcg gttcgcatgg
2581 atgacaagaa agtccgaaga aacgattact ccatggaatt ttgaggaagt tgtcgataaa
2641 ggtgcgtcag ctcaatcgt catcgagagg atgaccaact ttgacaagaa ttaccgaac
2701 gaaaaagtat tgcctaagca cagtttact tacgagtatt tcacagtga caatgaactc
2761 acgaaagtta agtatgtcac tgaggcatg cgtaaacccg ctttctaag cggagaacag
2821 aagaaagcaa tagtagatct gttattcaag accaaccgca aagtgcagt taagcaattg
2881 aaagaggact actttaagaa aattgaatgc ttcgattctg tcgagatctc cggggtagaa
2941 gatcgattta atgcgtcact tggtagctat catgacctcc taaagataat taaagataag
3001 gacttctgg ataacgaaga gaatgaagat atcttagaag atatagtgtt gactcttacc
3061 ctctttgaag atcgggaaat gattgaggaa agactaaaaa catacgtca cctgttcgac
3121 gataaggtta tgaacagtt aaagaggcgt cgctatcgg gctggggacg attgtcgcgg
3181 aaacttatca acgggataag agacaagcaa agtggtaaaa ctattctga ttttctaaag
3241 agcgacggct tcgccaatag gaactttatg cagctgatcc atgatgactc ttaaccttc

[0105]

3301 aaagaggata tacaaaaggc acaggtttcc ggacaagggg actcattgca cgaacatatt
3361 gcgaatcttg ctggttcgcc agccatcaaa aagggcatac tccagacagt caaagtagtg
3421 gatgagctag ttaaggtcat gggacgtcac aaaccggaaa acattgtaat cgagatggca
3481 cgcgaaaatc aaacgactca gaagggggcaa aaaaacagtc gagagcggat gaagagaata
3541 gaagagggtg ttaaagaact gggcagccag atcttaagg agcatcctgt ggaaaatacc
3601 caattgcaga acgagaaact ttaccttat tactacaaa atggaaggga catgtatgtt
3661 gatcaggaac tggacataaa ccgtttatct gattacgacg tcatcacat tgtaccccaa
3721 tcctttttga aggacgattc aatcgacaat aaagtgttta cagctcggga taagaaccga
3781 gggaaaagtg acaatgttcc aagcgaggaa gtcgtaaaga aaatgaaga ctattggcgg
3841 cagctcctaa atcgaaact gataacgcaa agaaagttcg ataactaac taaagctgag
3901 aggggtggct tgtctgaact tgacaaggcc ggatttatta aacgtcagct cgtggaaacc
3961 cgccaaatca caaagcatgt tgcacagata ctagattccc gaatgaatac gaaatcacg
4021 gagaacgata agctgattcg ggaagtcaaa gtaatcactt taaagtcaaa attggtgtcg
4081 gacttcagaa aggattttca attctataaa gttagggaga taaataacta ccaccatgcg
4141 cagcagctt atcttaatgc cgtcgtagg accgcactca ttaagaaata cccgaagcta
4201 gaaagtgagt ttgtgtatgg tgattacaaa gtttatgacg tccgtaagat gatcgcgaaa
4261 agcgaacagg agataggcaa ggctacagcc aaatacttct ttattctaa cattatgaat
4321 ttctttaaga cggaaatcac tctggcaaac ggagagatag gcaaacgacc ttaattgaa
4381 accaatgggg agacaggtga aatcgtatgg gataagggcc gggacttcgc gacggtgaga
4441 aaagtttgt ccatgcccc agtcaacata gtaaagaaaa ctgaggtgca gaccggaggg
4501 ttttcaaagg aatcgattct tccaaaaagg aatagtata agctcatcgc tcgtaaaaag
4561 gactgggacc cgaaaaagta cgggtggctc gatagcccta cagttgccta ttctgtccta
4621 gtatgtgcaa aagttgagaa gggaaaatcc aagaaactga agtcagtcaa agaattattg
4681 gggataacga ttatggagcg ctctctttt gaaaagaacc ccatcgactt ccttgaggcg
4741 aaaggttaca aggaagtaaa aaaggatctc ataattaaac taccaaagta tagtctgttt
4801 gagttagaaa atggccgaaa acggatgttg gctagcgccg gagagcttca aaaggggaac
4861 gaactcgcac taccgtctaa atacgtgaat ttctgtatt tagcgtccca ttacgagaag
4921 ttgaaagggt cacctgaaga taacgaacag aagcaacttt ttgttgagca gcacaaacat
4981 tatctcgacg aaatcataga gcaaatttcg gaattcagta agagagtcac cctagctgat
5041 gccaatctgg acaaagtatt aagcgcatac aacaagcaca gggataaacc catacgtgag
5101 caggcggaaa atattatcca ttgtttact ctaccaacc tcggcgtcc agccgcattc
5161 aagtattttg acacaacgat agatcgcaaa cgatacactt ctaccaagga ggtgctagac
5221 gcgacactga ttaccaatc catcacggga ttatatgaaa ctcgataga ttgtcacag
5281 cttgggggtg actctggtgg ttctggagga tctggtggtt ctactaatct gtcagatatt

[0106]

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 gtcattgctt tgactagcga cgcccctgaa tacaagcctt gggctctggt catacaggat
5821 agcaacggtg agaacaagat taagatgctc tctggtggtt ctccaagaa gaagaggaaa
5881 gtctaaccgg tcatacacc catcaccatt gagtttaaac ccgctgatca gcctcgactg
5941 tgccttctag ttgccagcca tctgttgtt gccctcccc cgtgccttcc ttgacctgg
6001 aaggtgccac tccactgtc ctctcctaat aaaatgagga aattgcatcg cattgtctga
6061 gtaggtgtca ttctattctg gggggtgggg tggggcagga cagcaagggg gaggattggg
6121 aagacaatag caggcatgct ggggatgagg tgggctctat ggcttctgag gcggaaagaa
6181 ccagctgggg ctcgataccg tcgacctcta gctagagctt ggcgtaata tggtcatagc
6241 tgtttctgt gtgaaattgt tatccgtca caattccaca caacatacga gccggaagca
6301 taaagtgtaa agcctagggt gcctaagag tgagctaact cacattaatt gcgttgcgt
6361 cactgcccgc ttccagtcg ggaaacctgt cgtgccagct gcattaatga atcggccaac
6421 gcgcggggag aggcggttg cgtattgggc gctcttccgc ttctcgtc actgactgc
6481 tgcgctcgt cgttcggctg cggcgagcgg tatcagctca ctcaaaggcg gtaatacgtt
6541 tatccacaga atcaggggat aacgcaggaa agaactgtg agcaaaaggc cagcaaaagg
6601 ccaggaaccg taaaaaggcc gcgttgcctg cgttttcca taggctccgc cccctgacg
6661 agcatcaca aaatcgacgc tcaagtcaga ggtggcgaaa cccgacagga ctataaagat
6721 accaggcgtt tccccctgga agctccctcg tgcgctcctc tgttccgacc ctgccgtta
6781 ccgataacct gtccgcttt ctccctcgg gaagcgtggc gctttctcat agtcacgt
6841 gtaggtatct cagttcgtg taggtcgtc gctccaagct gggctgtgtg cacgaacccc
6901 ccgttcagcc cgaccgtgc gccttatccg gtaactatcg tcttgagtc aaccggtaa
6961 gacacgact atcgcactg gcagcagcca ctggtaacag gattagcaga gcgaggtatg
7021 taggcgtgc tacagagttc ttgaagtgt ggcctaacta cggctacact agaagaacag
7081 tatttggtat ctgcgtctg ctgaagccag ttacctcgg aaaaagagtt ggtagctctt
7141 gatccggcaa acaaaccacc gctgtagcg gtggttttt tgttgcaag cagcagatta
7201 cgcgcagaaa aaaaggatct caagaagatc cttgatctt ttctacgggg tctgacgctc
7261 agtggaaaca aaactcacgt taagggtatt tggatcatg attatcaaaa aggatcttca
7321 cctagatcct tttaaattaa aatgaagtt ttaaataat ctaaagtata tatgagtaa

7381 cttggtctga cagttaccaa tgcttaatca gtgaggcacc tatctcagcg atctgtctat
 7441 ttcgttcac catagttgcc tgactccccg tcgtgtagat aactacgata cgggagggct
 7501 taccatctgg cccagtgct gcaatgatac cgcgagaccc acgtcaccg gctccagatt
 7561 tatcagcaat aaaccagcca gccggaaggg ccgagcgcag aagtggctct gcaactttat
 7621 ccgcctccat ccagtctatt aattgttgcc gggaagctag agtaagtagt tcgccagtta
 7681 atagtttgcg caacgttgtt gccattgcta caggcatcgt ggtgtcacgc tcgtcgtttg
 7741 gtagggcttc attcagctcc ggttcccaac gatcaaggcg agttacatga tccccatgt
 7801 tgtgcaaaaa agcggtagc tcctcggtc ctccgatcgt tgcagaagt aagtggccg
 7861 cagtgttacc actcatgggt atggcagcac tgcataatc tcttactgtc atgcatccg
 7921 taagatgctt ttctgtgact ggtgagtact caaccaagtc attctgagaa tagtgtatgc
 7981 ggcgaccgag ttgctcttgc ccggcgtaaa tacgggataa taccgcgcca catagcagaa
 8041 ctttaaaagt gctcatcatt ggaaaacgtt ctccggggcg aaaactctca aggatcttac
 [0107] 8101 cgctgttgag atccagttcg atgtaacca ctctgcacc caactgatct tcagcatctt
 8161 ttactttcac cagcgtttct gggtgagcaa aaacaggaag gcaaaatgcc gcaaaaaagg
 8221 gaataagggc gacacggaaa tgtlgaatac tcatactctt ctttttcaa tattattgaa
 8281 gcatttatca gggttattgt ctcatgagcg gatacatatt tgaatgtatt tagaaaaata
 8341 aacaaatagg ggtccgcgc acattcccc gaaaagtgcc acctgacgtc gacggatcgg
 8401 gagatcgatc tcccgatccc ctagggtcga ctctcagtac aatctgctct gatccgcat
 8461 agttaagcca gtatctgtc cctgcttggt tgttgagggt cgctgagtag tgcgcgagca
 8521 aaatttaagc tacaacaagg caaggcttga ccgacaattg catgaagaat ctgcttaggg
 8581 ttaggcggtt tgcgtgctt cgcgatgtac gggccagata tacgcgttga cattgattat
 8641 tgactagtta ttaatagtaa tcaattacgg ggtcattagt tcatagccca tatatggagt
 8701 tccgcgttac ataactacg gtaaatggcc cgctggctg accgcccac gacccccgc
 8761 cattgacgtc aataatgacg tatgttccca tagtaacgcc aatagggact ttccattgac
 8821 gtcaatgggt ggagtattta cggtaaactg cccacttggc agtacatcaa gtgtatc

[0108] 在一些实施方案中, BE4核酸序列选自以下之一者:

[0109] 原始BE4

[0110] ATGagctcagagactggcccagtggtgtggacccacattgagacggcgatcgagcccatgagtt
 tgaggtattcttcgatccgagagagctccgcaaggagacctgcctgctttacgaaattaattgggggggcccggcac
 tccatttggcgacatacatcacagaacactaacaagcacgtcgaagtcaacttcacgacagaaa
 gatattttctgtccgaacacaaggtgcagcattacctggtttctcagctggagcccatgcccgaatgtagtagggc
 catcactgaattcctgtcaaggtatccccacgtcactctgtttatttacatcgcaaggctgtaccaccacgtgac
 ccccgcaatcgacaaggcctgcgggatttgatctcttcaggtgtgactatccaaattatgactgagcaggagtcag
 gatactgctggagaaactttgtgaattatagcccagtaataagcccactggcctaggtatccccatctgtgggt
 acgactgtacgttcttgaactgtactgcatcatactgggcctgcctccttgtctcaacattctgagaaggaagcag

ccacagctgacattctttaccatcgctcttcagtccttgctcattaccagcgactgccccacacattctctgggcca
ccgggttgaaatctgggtggttcttctgggtggttcttagcggcagcgagactcccgggacctcagagtccgccacacc
cgaaagttctgggtggttcttctgggtggttctgataaaaagtattctattgggttagccatcggcactaattccggtt
ggatgggctgtcataaccgatgaatacaaagtaccttcaaagaaatttaaggtgttggggaacacagaccgtcatt
cgattaaaaagaatcttatcggtgccctcctattcgatagtggcgaaacggcagaggcgactcgcctgaaacgaac
cgctcggagaaggtatacacgtcgcaagaaccgaatatgttacttacaagaaatttttagcaatgagatggccaaa
gttgacgattctttctttcaccgtttggagagtccttcttgcgaagaggacaagaaacatgaacggcacccca
tctttggaaacatagtagatgaggtggcatatcatgaaaagtacccaacgatttatcacctcagaaaaagctagt
tgactcaactgataaagcggacctgaggttaatctacttggctcttggccatatagataaagttccgtgggcacttt
ctcattgagggatgatctaaatccggacaactcggatgtcgacaaactgttcatccagttagtacaacctataatc
agttgtttgaagagaacctataaatgcaagtggcgtggatgcgaaggctatttcttagcgcccgcctctctaaatc
ccgacggctagaaaacctgatcgacaaattacccggagagaagaaaaatgggtgttcggttaaccttatagcgctc
tcaactaggcctgacaccaaattttaagtcgaacttcgacttagctgaagatgccaaattgcagcttagtaaggaca
cgtaacgatgacgatctcgacaatctactggcacaattggagatcagtatgcggacttatttttggctgccaaaaa
ccttagcgatgcaatcctcctatctgacatactgagagttaatactgagattaccaaggcgccgttatccgcttca
atgatcaaaaggtacgatgaacatcaccaagacttgacacttctcaaggccctagtccgtcagcaactgcctgaga
aatataaggaaatattctttgatcagtcgaaaaacgggtacgcaggttatattgacggcggagcgagtcaagagga
attctacaagtttatcaaaccatattagagaagatggatgggacggaagagttgcttgtaaaactcaatcgcgaa
gatctactgcgaaagcagcggactttcgacaacggtagcattccacatcaaattccacttaggcgaattgcatgcta
tacttagaaggcaggaggatttttatccgttccctcaaagacaatcgtgaaaagattgagaaaatcctaacctttcg
cataccttactatgtgggacccctggcccgagggaactctcggttcgcatggatgacaagaaagtccgaagaaacg
attactccatggaattttgaggaaagttgtcgataaaggtgcgtcagctcaatcgttcatcgagaggatgaccaact
ttgacaagaatttaccgaacgaaaaagtattgcctaagcacagtttactttacgagtatttcacagtgtacaatga
actcacgaaagttaagtatgtcactgagggcatgcgtaaacccgcctttctaagcggagaacagaagaaagcaata
gtagatctgttattcaagaccaaccgcaaagtgcagtttaagcaattgaaagaggactactttaagaaaattgaat
gcttcgattctgtcgagatctccggggtagaagatcgatttaatgcgtcacttgggtacgtatcatgacctcctaaa
gataattaaagataaggacttcttgataacgaagagaatgaagatatcttagaagatatagtgttgactcttacc
ctctttgaagatcgggaaatgattgaggaaagactaaaaacatacgcctcacctgttcgacgataagggttatgaaac
agttaaagaggcgctcgctatacgggctggggacgattgtcgcggaaacttatcaacgggataagagacaagcaaag
tggtaaaactattctcgattttctaaagagcgacggcttcgccaataggaactttatgcagctgatccatgatgac
tctttaaccttcaaagaggatatacaaaaggcacaggtttccggacaaggggactcattgcacgaacatattgcga
atcttgcgtggttcgccagccatcaaaaagggcatactccagacagctcaaagtagtgatgagctagtttaaggtcat
gggacgtcacaaaccgaaaaacattgtaatcgagatggcacgcgaaaatcaaacgactcagaaggggcaaaaaaac
agtcgagagcggatgaagagaatagaagagggtattaaagaactgggcagccagatcttaaaggagcatcctgtgg
aaaatacccaattgcagaacgagaaactttacctctattacctacaaaatggaaggacatgtatgttgatcagga
actggacataaaccgtttatctgattacgacgtcgatcacattgtaccccaatcctttttgaaggacgattcaatc
gacaataaagtgttacacgctcggataagaaccgagggaaaagtgacaatgttccaagcgaggaagtcgtaaaga
aatgaagaactattggcggcagctcctaaatgcgaaactgataacgcaaagaaagttcgataacttaactaaagc
tgagaggggtggcttgtctgaacttgacaaggccgattttattaaacgtcagctcgtggaacccgccaaatcaca

27

CTATAAAAAAAAAACCTGATCGGCGCCCTTTTGTGTTGACAGTGGTGAGACTGCTGAAGCGACTCGCCTGAAGCGAAC
TGCCAGGAGGCGGTATACGAGGCGAAAAACCGAATTTGTACCTCCAGGAGATTTTCTCAAATGAAATGGCCAAG
GTAGATGATAGTTTTTTTCCACCGCTTGAAGAAAGTTTTCTCGTTGAGGAGGACAAAAAGCACGAGAGGCACCCAA
TCTTTGGCAACATAGTCGATGAGGTCGCATACCATGAGAAATATCCTACGATCTATCATCTCCGCAAGAAGCTGGT
CGATAGCACGGATAAAGCTGACCTCCGGCTGATCTACCTTGCTCTTGCTCACATGATTAAATTCAGGGGCCATTTC
CTGATAGAAGGAGACCTCAATCCCGACAATTCTGATGTCGACAAACTGTTTATTCAGCTCGTTCAGACCTATAATC
AACTCTTTGAGGAGAACCCCATCAATGCTTCAGGGGTGGACGCAAAGGCCATTTTGTCCGCGCGCTTGAGTAAATC
ACGACGCCTCGAGAATTTGATAGCTCAACTGCCGGGTGAGAAGAAAAACGGGTGTTTGGGAATCTCATAGCGTTG
AGTTTGGGACTTACGCCAACTTTAAGTCTAACTTTGATTTGGCCGAAGATGCCAAATTGCAGCTGTCCAAAGATA
CCTATGATGACGACTTGGATAACCTTCTTGCGCAGATTGGTGACCAATACGCGGATCTGTTTCTTGCCGCAAAAAA
TCTGTCCGACGCCATACTCTTGCCGATATACTGCGCGTCAATACTGAGATAACTAAGGCTCCCTCAGCGCGTCC
ATGATTAAGATACGATGAGCACCACCAAGATCTCACTCTGTTGAAAGCCCTGGTTCGCCAGCAGCTTCCAGAGA
AGTATAAGGAGATATTTTCGACCAATCTAAAAACGGCTATGCGGGTTACATTGACGGTGGCGCTCTCAAGAAGA
ATTCTACAAGTTTATAAAGCCGATACTTGAGAAAATGGACGGTACAGAGGAATTGTTGGTTAAGCTCAATCGCGAG
GACTTGTTGAGAAAGCAGCGCACATTTGACAATGGTAGTATTCCACACCAGATTCATCTGGGCGAGTTGCATGCCA
TTCTTAGAAGACAAGAAGATTTTATCCGTTTCTGAAAGATAACAGAGAAAAGATTGAAAAGATACTTACCTTTCG
CATACCGTATTATGTAGGTCCCCTGGCTAGAGGGAACAGTCGCTTCGCTTGATGACTCGAAAATCAGAAGAAACA
ATAACCCCTGGAATTTTGAAGAAGTGGTAGATAAAGGTGCGAGTGCCCAATCTTTTATTGAGCGGATGACAAATT
TTGACAAGAATCTGCCTAACGAAAAGGTGCTTCCCAAGCATTCCCTTTTGTATGAATACTTTACAGTATATAATGA
ACTGACTAAAGTGAAGTACGTTACCGAGGGGATGCGAAAGCCAGCTTTTCTCAGTGGCGAGCAGAAAAAGCAATA
GTTGACCTGCTGTTCAAGACGAATAGGAAGTTACCGTCAAACAGCTCAAAGAAGATTACTTTAAAAAGATCGAAT
GTTTTGATTGAGTTGAGATAAGCGGAGTAGAGGATAGATTTAACGCAAGTCTTGAACTTATCATGACCTTTTGAA
GATCATCAAGGATAAAGATTTTGGACAACGAGGAGAATGAAGATATCCTGGAAGATATAGTACTTACCTTGACG
CTTTTGAAGATCGAGAGATGATCGAGGAGCGACTTAAGACGTACGCACATCTCTTGACGATAAGGTTATGAAAC
AATTGAAACGCCGGGTATACTGGCTGGGGCAGGCTTCTCGAAAAGCTGATTAATGGTATCCGCGATAAGCAGTC
TGGAAGACAATCCTTGACTTTCTGAAAAGTGATGGATTTGCAAATAGAACTTTATGCAGCTTATACATGATGAC
TCTTTGACGTTCAAGGAAGACATCCAGAAGGCACAGGTATCCGGCCAAGGGGATAGCCTCCATGAACACATAGCCA
ACCTGGCCGGCTCACCAGCTATTAAAAAGGGAATATTGCAAACCGTTAAGGTTGTTGACGAACTCGTTAAGGTTAT
GGGCCGACACAAACCAGAGAATATCGTGATTGAGATGGCTAGGGAGAATCAGACCACTCAAAAAGGTCAGAAAAAT
TCTCGCGAAAGGATGAAGCGAATTGAAGAGGGAATCAAAGAAGTTGGCTCTCAAATTTTGAAGAGCACCCGGTAG
AAAACACTCAGCTGCAGAATGAAAAGCTGTATCTGTATTATCTGCAGAATGGTCGAGATATGTACGTTGATCAGGA
GCTGGATATCAATAGGCTCAGTGACTACGATGTCGACCACATCGTTCCTCAATCTTTCCTGAAAGATGACTCTATC
GACAACAAAGTGTTGACGCGATCAGATAAGAACCGGGGAAAAATCCGACAATGTACCCTCAGAAGAAGTTGTCAAGA
AGATGAAAACTATTGGAGACAATTGCTGAACGCCAAGCTCATAACACAACGCAAGTTCGATAACTTGACGAAAGC
CGAAAGAGGTGGGTTGTCAGAATTGGACAAAGCTGGCTTTATTAAGCGCCAATTGGTGGAGACCCGGCAGATTACG
AAACACGTAGCACAAATTTTGGATTACGAATGAATACCAATACGACGAAAACGACAAATTGATACGCGAGGTGA
AAGTGATTACGCTTAAGAGTAAGTTGGTTTCCGATTTAGGAAGGATTTTCAGTTTACAAAGTAAGAGAAATAAA
CAACTACCACCACGCCCATGATGCTTACCTCAACGCGGTAGTTGGCACAGCTCTTATCAAAAAATATCCAAAGCTG
GAAAGCGAGTTCGTTTACGGTGACTATAAAGTATACGACGTTCGGAAGATGATAGCCAAATCAGAGCAGGAAATTG

GGAAGGCAACCGCAAAATACTTCTTCTATTCAAACATCATGAACTTCTTTAAGACGGAGATTACGCTCGCGAACGG
CGAAATACGCAAGAGGCCCTCATAGAGACTAACGGCGAAACCGGGGAGATCGTATGGGACAAAGGACGGGACTTT
GCGACCGTTAGAAAAGTACTTTCAATGCCACAAGTGAATATTGTTAAAAAGACAGAAGTACAAACAGGGGGTTCA
GTAAGGAATCCATTTTGCCCAAGCGGAACAGTGATAAATTGATAGCAAGGAAAAAAGATTGGGACCCTAAGAAGTA
CGGTGGTTTTCGACTCTCCTACCGTTGCATATTCAGTCCTTGTAGTTGCGAAAGTGGAAGGGGAAAAGTAAGAAG
CTTAAGAGTGTTAAAGAGCTTCTGGGCATAACCATAATGGAACGGTCTAGCTTCGAGAAAAATCCAATTGACTTTC
TCGAGGCTAAAGGTTACAAGGAGGTAAAAAAGGACCTGATAATTAACTCCCAAAGTACAGTCTCTTCGAGTTGGA
GAATGGGAGGAAGAGAATGTTGGCATCTGCAGGGGAGCTCCAAAAGGGGAACGAGCTGGCTCTGCCTTCAAATAC
GTGAACTTTCTGTACCTGGCCAGCCACTACGAGAACTCAAGGGTTCTCCTGAGGATAACGAGCAGAAACAGCTGT
TTGTAGAGCAGCACAAGCATTACCTGGACGAGATAATTGAGCAAATTAGTGAGTTCTCAAAAAGAGTAATCCTTGC
AGACGCGAATCTGGATAAAGTTCTTTCCGCCTATAATAAGCACCGGGACAAGCCTATACGAGAACAAGCCGAGAAC
ATCATTACCTCTTTACCCTTACTAATCTGGGCGCGCCGGCCGCCTTCAAATACTTCGACACCACGATAGACAGGA
AAAGGTATACGAGTACCAAAGAAGTACTTGACGCCACTCTCATCCACCAGTCTATAACAGGGTTGTACGAAACGAG
GATAGATTTGTCCAGCTCGGCGGCGACTCAGGAGGGTCAGGCGGCTCCGGTGGATCAACGAATCTTTCCGACATA
ATCGAGAAAGAAACCGGCAAACAGTTGGTGATCCAAGAATCAATCCTGATGCTGCCTGAAGAAGTAGAAGAGGTGA
TTGGCAACAAACCTGAGTCTGACATTCTTGTCCACACCGCGTATGACGAGAGCACGGACGAGAACGTTATGCTTCT
CACTAGCGACGCCCCCTGAGTATAAACCATGGGCGCTGGTCATCCAAGATTCCAATGGGGAAAACAAGATTAAGATG
CTTAGTGGTGGGTCTGGAGGGAGCGGTGGGTCCACGAACCTCAGCGACATTATTGAAAAAGAGACTGGTAAACAAC
TTGTAATACAAGAGTCTATTCTGATGTTGCCTGAAGAGGTGGAGGAGGTGATTGGGAACAAACCGGAGTCTGATAT
ACTTGTTTCATACCGCCTATGACGAATCTACTGATGAGAATGTGATGCTTTTaaACGTCAGACGCTCCCGAGTACAAA
CCCTGGGCTCTGGTGATTACAGGACAGCAATGGTGAGAATAAGATTAAAATGTTGAGTGGGGGCTCAAAGCGCACGG
CTGACGGTAGCGAATTTGAGAGCCCCAAAAAAGGTCGAAtaa

[0113] BE4密码子优化2

[0114] ATGAGCAGCGAGACAGGCCCTGTGGCTGTGGATCCTACACTGCGGAGAAGAATCGAGCCCCACGAGTT
CGAGGTGTTCTTCGACCCCAGAGAGCTGCGGAAAGAGACATGCCTGCTGTACGAGATCAACTGGGGCGGCAGACAC
TCTATCTGGCGGCACACAAGCCAGAACACCAACAAGCACGTGGAAGTGAACTTTATCGAGAAGTTTACGACCGAGC
GGTACTTCTGCCCCAACCCAGATGCAGCATCACCTGGTTTCTGAGCTGGTCCCCTTGCGGCGAGTGCAGCAGAGC
CATCACCGAGTTTCTGTCCAGATATCCCCACGTGACCCTGTTTCATCTATATCGCCCGGCTGTACCACCACGCCGAT
CCTAGAAATAGACAGGGACTGCGCGACCTGATCAGCAGCGGAGTGACCATCCAGATCATGACCGAGCAAGAGAGCG
GCTACTGCTGGCGGAACTTCGTGAACCTACAGCCCCAGCAACGAAGCCCACTGGCCTAGATATCCTCACCTGTGGGT
CCGACTGTACGTGCTGGAACCTGTACTGCATCATCCTGGGCCTGCCTCCATGCCTGAACATCCTGAGAAGAAAGCAG
CCTCAGCTGACCTTCTTCACAATCGCCCTGCAGAGCTGCCACTACCAGAGACTGCCTCCACACATCCTGTGGGCCA
CCGGACTTAAGAGCGGAGGATCTAGCGGCGGCTCTAGCGGATCTGAGACACCTGGCACAAGCGAGTCTGCCACACC
TGAGAGTAGCGGCGGATCTTCTGGCGGCTCCGACAAGAAGTACTCTATCGGACTGGCCATCGGCACCAACTCTGTT
GGATGGGCCGTGATCACCGACGAGTACAAGGTGCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACA
GCATCAAGAAGAATCTGATCGGCGCCCTGCTGTTTCGACTCTGGCGAAACAGCCGAAGCCACCAGACTGAAGAGAAC
CGCCAGGCGGAGATACACCCGGCGGAAGAACCGGATCTGCTACCTGCAAGAGATCTTCAGCAACGAGATGGCCAAG
GTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCTGGTGGAAAGAGGACAAGAAGCACGAGCGGCACCCCA
TCTTCGGCAACATCGTGGATGAGGTGGCCTACCACGAGAAGTACCCACCATCTACCACCTGAGAAAGAACTGGT

GGACAGCACCGACAAGGCCGACCTGAGACTGATCTACCTGGCTCTGGCCACATGATCAAGTTCGGGGCCACTTT
CTGATCGAGGGCGATCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTTCATCCAGCTGGTGCAGACCTACAACC
AGCTGTTTCGAGGAAAAACCCCATCAACGCCTCTGGCGTGGACGCCAAGGCTATCCTGTCTGCCAGACTGAGCAAGAG
CAGAAGGCTGGAAAACTGATCGCCCAGCTGCCTGGCGAGAAGAAGAATGGCCTGTTTCGGCAACCTGATTGCCCTG
AGCCTGGGACTGACCCCTAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAACTGCAGCTGAGCAAGGACA
CCTACGACGACGACCTGGACAATCTGCTGGCCAGATCGGCGATCAGTACGCCGACTTGTTTTCTGGCCGCCAAGAA
CCTGTCCGACGCCATCCTGCTGAGCGATATCCTGAGAGTGAACACCGAGATCACAAAGGCCCTCTGAGCGCCTCT
ATGATCAAGAGATACGACGAGCACCACCAGGATCTGACCCTGCTGAAGGCCCTCGTTAGACAGCAGCTGCCAGAGA
AGTACAAAGAGATTTTCTTCGATCAGTCCAAGAACGGCTACGCCGGCTACATTGATGGCGGAGCCAGCCAAGAGGA
ATTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTGGTCAAGCTGAACAGAGAG
GACCTGCTGCGGAAGCAGCGGACCTTCGACAATGGCTCTATCCCTCACCAGATCCACCTGGGAGAGCTGCACGCCA
TTCTGCGGAGACAAGAGGACTTTTACCCATTCTGAAGGACAACCGGGAAAAAGATCGAGAAGATCCTGACCTTCAG
GATCCCTACTACGTGGGACCACTGGCCAGAGGCAATAGCAGATTTCGCTGGATGACCAGAAAGAGCGAGGAAACC
ATCACACCCTGGAACCTTCGAGGAAGTGGTGGACAAGGGCGCCAGCGCTCAGTCCTTCATCGAGCGGATGACCAACT
TCGATAAGAACCTGCCTAACGAGAAGGTGCTGCCAAGCACTCCCTGCTGTATGAGTACTTCACCGTGTACAACGA
GCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTTCTGAGCGGCGAGCAGAAAAAGGCCATT
GTGGATCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGT
GCTTCGACAGCGTGGAATCAGCGGCGTGGAAGATCGGTTCAATGCCAGCCTGGGCACATACCACGACCTGCTGAA
AATTATCAAGGACAAGGACTTCCTGGACAACGAAGAGAACGAGGACATTCTCGAGGACATCGTGCTGACCCTGACA
CTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACATACGCCACCTGTTTCGACGACAAAGTGATGAAGC
AACTGAAGCGGAGGCGGTACACAGGCTGGGGCAGACTGTCTCGGAAGCTGATCAACGGCATCCGGGATAAGCAGTC
CGGCAAGACAATCCTGGATTTCTGAAGTCCGACGGCTTCGCCAACAGAACTTCATGCAGCTGATCCACGACGAC
AGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAAGGCGATTCTCTGCACGAGCACATTGCCA
ACCTGGCCGGATCTCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTTGTGAAAGTGAT
GGGCAGACACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACACAGAAGGGCCAGAAGAAC
AGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACCCCCGTGG
AAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGACGGGATATGTACGTGGACCAAGA
GCTGGACATCAACCGGTGAGCGACTACGATGTGGACCATATCGTGCCCCAGAGCTTTCTGAAGGACGACTCCATC
GATAACAAGGTCTTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGATAACGTGCCCTCCGAAGAGGTGGTCAAGA
AGATGAAGAACTACTGGCGACAGCTGCTGAACGCCAAGCTGATTACCCAGCGGAAGTTCGATAACCTGACCAAGGC
CGAGAGAGGCGGCTGAGCGAACTTGATAAGGCCGGCTTCATTAAGCGGCAGCTGGTGGAAACCCGGCAGATCACC
AAACACGTGGCACAGATTCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTGA
AAGTCATCACCTGAAGTCTAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTCTACAAAGTGCGGGAAATCAA
CAACTACCATCACGCCACGACGCCTACCTGAATGCCGTTGTTGGAACAGCCCTGATCAAGAAGTATCCCAAGCTG
GAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAACAAGAGATCG
GCAAGGCTACCGCCAAGTACTTTTTCTACAGCAACATCATGAACTTTTTCAAGACAGAGATCACCCTGGCCAACGG
CGAGATCCGGAAAAGACCCCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCAGAGATTTT
GCCACAGTGCAGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAGAAAACCGAGGTGCAGACAGGCGGCTTCA
GCAAAGAGTCTATCCTGCCTAAGCGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTA

CGGCGGCTTCGATAGCCCTACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAGGGCAAGTCCAAAAAG
CTCAAGAGCGTGAAAGAGCTGCTGGGGATCACCATCATGGAAGAAGCAGCTTTGAGAAGAACCCGATCGACTTTC
TGGAAGCCAAGGGCTACAAAGAAGTCAAGAAGGACCTCATCATCAAGTCCCCAAGTACAGCCTGTTTCGAGCTGGA
AAATGGCCGGAAGCGGATGCTGGCCTCAGCAGGCGAACTGCAGAAAGGCAATGAACTGGCCCTGCCTAGCAAATAC
GTCAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCAGCCCCGAGGACAATGAGCAAAAGCAGCTGT
TTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGC
CGACGCTAACCTGGATAAGGTGCTGTCTGCCTATAACAAGCACCGGGACAAGCCTATCAGAGAGCAGGCCGAGAAT
ATCATCCACCTGTTTACCCTGACCAACCTGGGAGCCCCTGCCGCCTTCAAGTACTTCGACACCACCATCGACCGGA
AGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACACTGATCCACCAGTCTATCACCGGCCTGTACGAAACCCG
GATCGACCTGTCTCAGCTCGGCGGCGATTCTGGTGGTTCTGGCGGAAGTGGCGGATCCACCAATCTGAGCGACATC
ATCGAAAAAGAGACAGGCAAGCAGCTCGTGATCCAAGAATCCATCCTGATGCTGCCTGAAGAGTTGAGGAAGTGA
TCGGCAACAAGCCTGAGTCCGACATCCTGGTGCACACCGCCTACGATGAGAGCACCGATGAGAACGTCATGCTGCT
GACAAGCGACGCCCTGAGTACAAGCCTTGGGCTCTCGTGATTGAGGACAGCAATGGGGAGAACAAGATCAAGATG
CTGAGCGGAGGTAGCGGAGGCAGTGGCGGAAGCACAAACCTGTCTGATATCATTGAAAAAGAAACCGGAAGCAAC
TGGTGATTCAAGAGTCCATTCTCATGCTCCCGGAAGAAGTCGAGGAAGTCATTGAAACAAACCCGAGAGCGATAT
TCTGGTCCACACAGCCTATGACGAGTCTACAGACGAAAACGTGATGCTCCTGACCTCTGACGCTCCCGAGTATAAG
CCCTGGGCACCTTGTTATCCAGGACTCTAACGGGGAAAACAAAATCAAAATGTTGTCCGGCGGCAGCAAGCGGACAG
CCGATGGATCTGAGTTCGAGAGCCCCAAGAAGAAACGGAAGGTgGAGtaa

[0115] “碱基编辑活性”是指用于化学改变多核苷酸内的碱基。在一个实施例中,第一碱基被转换为第二碱基。在一个实施方案中,碱基编辑活性是胞苷脱氨酶活性,例如,将靶C • G转化为T • A。在另一个实施方案中,碱基编辑活性是腺苷或腺嘌呤脱氨酶活性,例如将A • T转化为G • C。

[0116] 术语“碱基编辑器系统”是指用于编辑靶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(BE)系统包含(1)用于使所述核碱基脱氨的多核苷酸可编程核苷酸结合结构域和脱氨酶结构域;(2)指导多核苷酸(例如,指导RNA)与多核苷酸可编程核苷酸结合结构域结合。在一些实施方案中,碱基编辑器系统包括(1)碱基编辑器(BE),其包含多核苷酸可编程DNA结合结构域和用于使所述核碱基脱氨的脱氨酶结构域;(2)指导RNA与多核苷酸可编程DNA结合结构域结合。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)。

[0117] 术语“保守氨基酸取代”或“保守突变”是指一个氨基酸被具有共同性质的另一氨基酸替代。定义单个氨基酸之间共有特性的一种功能方法是分析同源生物体相应蛋白质之间氨基酸变化的标准化频率(Schulz,G.E.和Schirmer,R.H.,Principles of Protein Structure, Springer-Verlag, 纽约(1979))。根据这样的分析,可以定义氨基酸组,其中组内的氨基酸优先彼此交换,因此在它们对整体蛋白质结构的影响上最相似(Schulz,G.E.和Schirmer,R.H.,同上)。保守突变的非限制性实例包括氨基酸的氨基酸取代,例如赖氨酸取代精氨酸且反之亦然,故可以保持正电荷;谷氨酸取代天冬氨酸且反之亦然,从而可保持负电荷;苏氨酸取代丝氨酸,以维持游离的-OH;以及氨酰胺取代天冬酰胺,从而维持游离的-NH₂。

[0118] 如本文可互换使用的,术语“编码序列”或“蛋白质编码序列”是指编码蛋白质的多核苷酸的区段。所述区域或序列的起始密码子在5'末端附近,终止密码子在3'末端附近。编码序列也可称为开放阅读框(opening reading frames)。

[0119] “胞苷脱氨酶”是指能够催化将氨基转化为羰基的脱氨反应的多肽或其片段。在一实施方案中,胞苷脱氨酶将胞嘧啶转化为尿嘧啶或将5-甲基胞嘧啶转化为胸腺嘧啶。PmCDA1,衍生自海洋七鳃鳗(*Petromyzon marinus*) (海洋七鳃鳗胞嘧啶脱氨酶,Petromyzon marinus cytosine deaminase 1,“PmCDA1”)、AID(活化诱导的胞苷脱氨酶;AICDA),其衍生自哺乳动物(例如人类、猪、牛、马、猴等)以及作为示例性的胞苷脱氨酶的APOBEC。

[0120] 如本文所用,术语“脱氨酶”或“脱氨酶结构域”是指催化脱氨反应的蛋白质或酶。在一些实施方案中,脱氨酶或脱氨酶结构域是胞苷脱氨酶,其分别催化胞苷或脱氧胞苷水解为尿苷或脱氧尿苷。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,其催化胞嘧啶水解脱氨为尿嘧啶。在一些实施方案中,脱氨酶是腺苷脱氨酶,其催化腺嘌呤水解为次黄嘌呤。在一些实施方案中,脱氨酶是腺苷脱氨酶,其催化腺苷或腺嘌呤(A)水解为肌苷(I)。在一些实施方案中,脱氨酶或脱氨酶结构域是腺苷脱氨酶,其分别催化腺苷或脱氧腺苷水解为肌苷或脱氧肌苷。在一些实施方案中,腺苷脱氨酶催化脱氧核糖核酸(DNA)中腺苷的水解脱氨。本文提供的腺苷脱氨酶(例如工程化腺苷脱氨酶,进化的腺苷脱氨酶)可以来自任何生物,例如细菌。在一些实施方案中,腺苷脱氨酶来自细菌,例如大肠杆菌、金黄色葡萄球菌、伤寒链球菌、腐烂链球菌、流感嗜血杆菌或新月形梭菌。在一些实施方案中,腺苷脱氨酶是TadA脱氨酶。在一些实施方案中,脱氨酶或脱氨酶结构域是来自生物体例如人类、黑猩猩、大猩猩、猴子、牛、狗、大鼠或小鼠的天然存在的脱氨酶的变体。在一些实施方案中,脱氨酶或脱氨酶结构域本质上不存在。例如,在一些实施方案中,脱氨酶或脱氨酶结构域为至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%、至少99.1%、至少99.2%、至少99.3%、至少99.4%、至少99.5%、至少99.6%、至少99.7%、至少99.8%或至少99.9%的同一性自然发生的脱氨酶。举例来说,脱氨酶结构域描述于国际第PCT/2017/045381号专利申请(WO第2018/027078号国际专利申请)和国际第PCT/US2016/058344号专利申请(WO第2017/070632号国际专利申请)中,其各自通过引用整体并入本文。另请参阅Komor,A.C.等人,“Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage”*Nature* 533,420-424(2016); Gaudelli,N.M.等人,“Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage”*Nature* 551,464-471(2017);Komor,A.C.等人,“Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity”*Science Advances* 3: eaao4774(2017)以及Rees,H.A.等人,“Base editing:precision chemistry on the genome and transcriptome of living cells.”*Nat Rev Genet.*2018Dec;19(12):770-788.doi:10.1038/s41576-018-0059-1,其全部内容通过引用合并于本文中。

[0121] “可检测标记”是指一种组合物,当其与感兴趣的分子连接时,使其能够通过光谱、光化学、生物化学、免疫化学或化学手段进行检测。例如,有用的标记包括放射性同位素、磁珠、金属珠、胶体颗粒、荧光染料、电子致密试剂、酶(例如,如在ELISA中通常使用的)、生物

素、过氧羟基洋地黄毒苷(digoxigenin)或半抗原。

[0122] “疾病”是指损害或干扰细胞、组织或器官正常功能的任何状况或障碍。疾病的例子包括视网膜色素病变、尤塞氏综合症、镰状细胞病、 β -地中海贫血、 α -1抗胰蛋白酶缺乏症(A1AD)、肝紫质症、中链酰基辅酶A脱氢酶(MCAD)缺乏症、溶酶体酸性脂肪酶(LAL)缺乏症、苯丙酮尿症、血色素沉着症、冯吉尔克病、庞贝氏(Pompe)症、高雪氏症、贺勒氏症、囊性纤维化或慢性疼痛。在一个特定的实施方案中,所述疾病是A1AD。

[0123] “有效量”是指相对于未经治疗的患者或没有疾病的个体,缓解本文所述的试剂或活性化合物的量,例如本文所述的碱基编辑器。相对于未经治疗的患者或没有疾病的个体,即健康个体,改善疾病症状所需要的“治疗”。用于实施本发明以治疗疾病的活性化合物的有效量,根据给药方式、年龄、体重和受试者的总体健康状况而变化。最终,主治医师或兽医将决定适当的量和剂量方案。所述剂量称为“有效”量。在一个实施方案中,有效量是足以在细胞(例如体外或体内细胞)中引入目的基因改变的本发明的碱基编辑器的量。在一个实施方案中,有效量是达到治疗效果(例如,减少或控制视网膜色素病变、尤塞氏综合症、镰状细胞病、 β -地中海贫血、 α -1抗胰蛋白酶缺乏症(A1AD)、肝紫质症、中链酰基辅酶A脱氢酶(MCAD)缺乏症、溶酶体酸性脂肪酶(LAL)缺乏症、苯丙酮尿症、血色素沉着症、冯吉尔克病、庞贝氏症、高雪氏症、贺勒氏症、囊性纤维化或慢性疼痛或症状或其条件)。这种治疗效果不足以改变受试者、组织或器官的所有细胞中的致病基因,而仅足以改变致病基因的约1%、5%、10%、25%、50%、75%或受试者、组织或器官中存在更多的细胞。在一个实施方案中,有效量足以缓解一种或多种疾病的症状(例如,视网膜色素病变、尤塞氏综合症、镰状细胞病、 β -地中海贫血、 α -1抗胰蛋白酶缺乏症(A1AD)、肝紫质症、中链酰基辅酶A脱氢酶(MCAD)缺乏症、溶酶体酸性脂肪酶(LAL)缺乏症、苯丙酮尿症、血色素沉着症、冯吉尔克病、庞贝氏症、高雪氏症、贺勒氏症、囊性纤维化或慢性疼痛)。

[0124] “片段”是指多肽或核酸分子的一部分。所述部分优选包含参考核酸分子或多肽的全长的至少10%、20%、30%、40%、50%、60%、70%、80%或90%。片段可以包含10、20、30、40、50、60、70、80、90或100、200、300、400、500、600、700、800、900或1000个核苷酸或氨基酸。

[0125] “杂交”是指互补核碱基之间的氢键,其可以是Watson-Crick、Hoogsteen或反向Hoogsteen氢键。举例来说,腺嘌呤和胸腺嘧啶是通过形成氢键而配对的互补核碱基。

[0126] 术语“碱基修复抑制剂”、“碱基修复抑制剂”、“IBR”或其语法等同物是指能够抑制核酸修复酶例如碱基切除修复酶活性的蛋白质。在一些实施方案中,IBR是肌苷碱基切除修复的抑制剂。碱基修复的示例性抑制剂包括APE1、Endo III、Endo IV、Endo V、Endo VIII、Fpg、hOGG1、hNEIL1、T7 Endo1、T4PDG、UDG、hSMUG1和hAAG的抑制剂。在一些实施方案中,碱基修复抑制剂是Endo V或hAAG的抑制剂。在一些实施方案中,IBR是Endo V或hAAG的抑制剂。在一些实施方案中,IBR是催化失活的EndoV或催化失活的hAAG。在一些实施方案中,碱基修复抑制剂是催化失活的EndoV或催化失活的hAAG。在一些实施方案中,碱基修复抑制剂是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,本文提供的UGI蛋白包括UGI的片段和与UGI或UGI片段同源的蛋白。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“催化失活的肌苷特异性核酸酶”或“死亡肌苷特异性核酸酶”。不希望受到任何特定理论的束缚,催

化失活的肌苷糖基化酶(例如烷基腺嘌呤糖基化酶(AAG))可以结合肌苷,但不能产生脱碱基位点或去除肌苷,从而在空间上阻止新形成的肌苷部分受到DNA损伤/修复机制。在一些实施方案中,催化失活的肌苷特异性核酸酶能够结合核酸中的肌苷,但不切割核酸。非限制性示例性的催化失活的肌苷特异性核酸酶包括例如来自人的催化失活的烷基腺苷糖基化酶(AAG核酸酶)和例如来自大肠杆菌的催化失活的核酸内切酶V(EndoV核酸酶)。在一些实施方案中,催化失活的AAG核酸酶在另一个AAG核酸酶中包含E125Q突变或相应的突变。

[0127] 术语“分离的”、“纯化的”或“生物学上纯净的”是指不含各种程度不同组分材料,且所述各组分通常在其材料的天然状态下随其伴随。“隔离”表示与原始来源或周围环境的隔离程度。“纯化”表示高于隔离的分离程度。“纯化的”或“生物纯净的”蛋白质充分不含其他物质,以使任何杂质均不会实质性影响所述蛋白质的生物学特性或引起其他不利后果。即,如果本发明的核酸或肽在通过重组DNA技术生产时基本不含细胞材料、病毒材料或培养基,或者在化学合成时基本不含化学前体或其他化学品,则将其纯化。纯度和均质性通常使用分析化学技术确定,例如聚丙烯酰胺凝胶电泳或高效液相色谱法。术语“纯化的”可以表示一核酸或一蛋白质在电泳凝胶中本质上形成一条带。对于可进行修饰的蛋白质,例如磷酸化或糖基化,不同的修饰可以产生不同的分离的蛋白质,可以将其分别纯化。

[0128] “分离的多核苷酸”是指不含位于基因侧翼的基因的核酸(例如DNA),所述基因在衍生本发明的核酸分子的生物的天然存在的基因组中。因此,所述术语包括,例如,掺入载体中的重组DNA;进入自主复制的质粒或病毒;或进入原核生物或真核生物的基因组DNA;或以独立于其他序列的独立分子(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)等形式存在。另外,所述术语包括从DNA分子转录的RNA分子,以及作为编码附加多肽序列的杂合基因的一部分的重组DNA。

[0129] “分离的多肽”是指已经与天然伴随其的组分分离的本发明的多肽。通常,当多肽按重量计至少60%不含与之天然结合的蛋白质和天然存在的有机分子时,将其分离。优选地,所述制剂是按重量计至少75%,更优选地至少90%,并且最优选地至少99%的本发明的多肽。本发明的分离的多肽可通过例如从天然来源提取、通过表达编码这种多肽的重组核酸而获得、或通过化学合成蛋白质,且可以通过任何合适的方法,例如,管柱色谱法、聚丙烯酰胺凝胶电泳或通过HPLC分析来测量纯度。

[0130] 如本文所用,术语“接头”可指共价接头(例如,共价键)、非共价接头、化学基团或连接两个分子或部分,例如,蛋白质复合物或核糖核酸复合物的两个成分,或融合蛋白的两个结构域,例如多核苷酸可编程DNA结合结构域(例如,dCas9)和脱氨酶结构域(例如,腺苷脱氨酶或胞苷脱氨酶)。接头可以连接碱基编辑器系统的不同组件或组件的不同部分。例如,在一些实施方案中,接头可以连接多核苷酸可编程核苷酸结合结构域的指导多核苷酸结合结构域和脱氨酶的催化结构域。在一些实施方案中,接头可以连接CRISPR多肽和脱氨酶。在一些实施方案中,接头可以连接Cas9和脱氨酶。在一些实施方案中,接头可以连接dCas9和脱氨酶。在一些实施方案中,接头可以连接nCas9和脱氨酶。在一些实施方案中,接头可以连接指导多核苷酸和脱氨酶。在一些实施方案中,接头可以连接碱基编辑器系统的脱氨基组分和多核苷酸可编程核苷酸结合组分。在一些实施方案中,接头可以连接脱氨基成分的RNA结合部分和碱基编辑器系统的多核苷酸可编程核苷酸结合成分。在一些实施方案中,接头可连接脱碱基组分的RNA结合部分和碱基编辑器系统的多核苷酸可编程核苷酸

结合组分的RNA结合部分。接头可位于两个基团,分子或其他部分之间或在其两侧,并通过共价键或非共价相互作用连接至每个,从而将两者连接。在一些实施方案中,接头可以是有机分子、基团、聚合物或化学部分。在一些实施方案中,接头可以是多核苷酸。在一些实施方案中,接头可以是DNA接头。在一些实施方案中,接头可以是RNA接头。在一些实施方案中,接头可包含能够结合配体的适体。在一些实施方案中,配体可以是碳水化合物、肽、蛋白质或核酸。在一些实施方案中,接头可包含可从核糖开关衍生的适体。衍生适体的核糖开关可选自茶碱核糖开关、硫胺焦磷酸 (TPP) 核糖开关、腺苷钴胺素 (AdoCbl) 核糖开关、S-腺苷甲硫氨酸 (SAM) 核糖开关、SAH核糖开关、黄素单核苷酸 (FMN) 核糖开关、四氢叶酸核糖开关、赖氨酸核糖开关、甘氨酸核糖开关、嘌呤核糖开关、GlmS核糖开关或pre-queosine1 (PreQ1) 核糖开关。在一些实施方案中,接头可包含结合至多肽或蛋白质结构域例如多肽配体的适体。在一些实施方案中,多肽配体可以是K同源性 (KH) 结构域、MS2外壳蛋白结构域、PP7外壳蛋白结构域、SfMu Com外壳蛋白结构域、无菌 α 基序、端粒酶Ku结合基序和Ku蛋白、端粒酶Sm7结合基序和Sm7蛋白或RNA识别基序。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,核碱基编辑成分可包含脱氨酶结构域和RNA识别基序。

[0131] 在一些实施方案中,接头可以是氨基酸或多个氨基酸(例如,肽或蛋白质)。在一些实施方案中,接头的长度可以是约5至100个氨基酸,例如约5、6、7、8、9、10、11、12、13、14、15、16、17、18、19。长度为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。在一些实施方案中,接头的长度可以是约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。也可以考虑更长或更短的接头。

[0132] 在一些实施方案中,接头连接RNA可编程核酸酶的gRNA结合结构域,包括Cas9核酸酶结构域,和核酸编辑蛋白的催化结构域(例如胞苷或腺苷脱氨酶)。在一些实施方案中,接头连接dCas9和核酸编辑蛋白。例如,接头位于两个基团、分子或其他部分之间或在两个基团、分子或其他部分的侧面,并通过共价键连接至每个,从而将两者连接。在一些实施方案中,接头是氨基酸或多个氨基酸(例如,肽或蛋白质)。在一些实施方案中,接头是有机分子、基团、聚合物或化学部分。在一些实施方案中,接头的长度为5至200个氨基酸,例如5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、35、45、50、55、60、65、70、75、80、85、90、95、100、101、102、103、104、105、110、120、130,长度为140、150、160、175、180、190或200个氨基酸。也可以考虑更长或更短的接头。在一些实施方案中,接头包含氨基酸序列SGSETPGTSESATPES,其也可以称为XTEN接头。在一些实施方案中,接头包含氨基酸序列SGGS。在一些实施方案中,接头包含 $(SGGS)_n$ 、 $(GGGS)_n$ 、 $(GGGGS)_n$ 、 $(G)_n$ 、 $(EAAAK)_n$ 、 $(GGG)_n$ 、SGSETPGTSESATPES或 $(XP)_n$ 基序,或以下各项的组合其中任何一个,其中n独立地是1到30之间的整数,并且X是任何氨基酸。在一些实施方案中,n为1、2、3、4、5、6、7、8、9、10、11、12、13、14或15。在一些实施方案中,连接基包含多个脯氨酸残基和长度为5至21、5至14、5至9、5至7个氨基酸,例如PAPAP、PAPAPA、PAPAPAP、PAPAPAPA、 $P(AP)_4$ 、 $P(AP)_7$ 、 $P(AP)_{10}$ 。这种富含脯氨酸的接头也被称为“刚性”接头。

[0133] 在一些实施方案中,碱基编辑器的结构域通过包含以下氨基酸序列的接头融合:

[0134] SGGSSGSETPGTSESATPESGGGS、SGGSSGGSSGSETPGTSESATPESGGSSGGGS或GGSGGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTE

PSEGSAPGTSTEPSEGSAPGTSESATP。在一些实施方案中,碱基编辑器的结构域通过包含氨基酸序列SGSETPGTSESATPES的接头被融合,所述接头也被称为XTEN接头。在一些实施方案中,接头的长度为24个氨基酸。在一些实施方案中,接头包含氨基酸序列SGGSSGGSSGSETPGTSESATPES。在一些实施方案中,接头的长度为40个氨基酸。在一些实施方案中,接头包含氨基酸序列SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGS。在一些实施方案中,接头的长度为64个氨基酸。在一些实施方案中,接头包含氨基酸序列SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGSPGSETPGTSESATPESSGGS SGGS。在一些实施方案中,接头的长度为92个氨基酸。在一些实施方案中,接头包含氨基酸序列PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAP GTSTEPSEGSAPGTSESATPESGPGSEPATs。

[0135] 如本文所用,术语“突变”是指例如核酸或氨基酸序列的序列中的残基被另一残基取代,或序列中一个或多个残基的缺失或插入。突变通常在本文中通过鉴定原始残基,且所述原始残基随后在序列中残基的位置及新取代的残基的身份来描述。本文提供的各种用于进行氨基酸取代(突变)的方法是本领域众所周知的,并且例如由Green和Sambrook, *Molecular Cloning: A Laboratory Manual* (第4版, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 纽约(2012))所提供。在一些实施方案中,当前公开的碱基编辑器可以在核酸(例如,受试者的基因组内的核酸)中有效地产生“预期突变”,例如点突变,而不会产生大量的意外突变,例如意外的点突变。在一些实施方案中,预期的突变是由与专门设计用于产生预期的突变的指导多核苷酸(例如, gRNA) 结合的特定碱基编辑器(例如, 胞苷碱基编辑器或腺苷碱基编辑器)产生的突变。

[0136] 通常,相对于参考(或野生型)序列,即不包含突变的序列,对序列(例如,本文所述的氨基酸序列)中产生或鉴定的突变进行编号。本领域技术人员将容易理解如何确定氨基酸和核酸序列中的突变相对于参考序列的位置。

[0137] 术语“非保守突变”涉及不同基团之间的氨基酸取代,例如赖氨酸代替色氨酸或苯丙氨酸代替丝氨酸等。在这种情况下,非保守氨基酸取代优选不被取代、干扰功能性变体或抑制其功能。非保守氨基酸取代可增强功能变体的生物学活性,从而与野生型蛋白质相比,功能变体的生物学活性增加。

[0138] 术语“核定位序列”、“核定位信号”或“NLS”是指促进蛋白质导入细胞核的氨基酸序列。核定位序列是本领域已知的,并且例如描述于Plank等人于2000年11月23日提交的国际第PCT/EP2000/011690号专利申请,且于2001年5月31日公开的W0第2001/038547号国际专利申请,所述专利的内容因其示例性核定位序列的公开内容通过引用并入本文。在其他实施方案中,NLS是例如Koblan等人, *Nature Biotech*, 2018, doi:10.1038/nbt.4172所述的优化NLS。在一些实施方案中,NLS包含氨基酸序列KRTADGSEFESPKKKRKV、KRPAATKKAGQAKKKK、KKTELQTTNAENKTKKL、KRGINDRNFWRGENGRKTR、RKSGKIAAIVVKRPRK、PKKKRKV或MDSLLMNRKFLYQFKNVRWAKGRRETYLC。

[0139] 在本文中可互换使用的术语“核碱基”、“氮碱基”或“碱基”是指形成核苷的含氮生物化合物,其又是核苷酸的组分。核碱基形成碱基对并彼此堆叠的能力直接导致长链螺旋结构,例如核糖核酸(RNA)和脱氧核糖核酸(DNA)。五个核碱基-腺嘌呤(A)、胞嘧啶(C)、鸟嘌呤(G)、胸腺嘧啶(T)和尿嘧啶(U)-被称为初级或典范。腺嘌呤和鸟嘌呤衍生自嘌呤、胞嘧啶、尿嘧啶和胸腺嘧啶衍生自嘧啶。DNA和RNA也可以包含其他修饰的(非主要)碱基。非限制

性示例性修饰核碱基可包括次黄嘌呤、黄嘌呤、7-甲基鸟嘌呤、5,6-二氢尿嘧啶、5-甲基胞嘧啶 (m5C) 和5-氢甲基胞嘧啶。次黄嘌呤和黄嘌呤可以通过诱变剂的存在而产生,它们都可以通过脱氨基作用(用羰基取代胺基)。次黄嘌呤可以由腺嘌呤修饰。黄嘌呤可以由鸟嘌呤修饰。尿嘧啶可以由胞嘧啶脱氨基产生。“核苷”由核碱基和五碳糖(核糖或脱氧核糖)组成。核苷的实例包括腺苷、鸟苷、尿苷、胞苷、5-甲基尿苷 (m5U)、脱氧腺苷、脱氧鸟苷、胸苷、脱氧尿苷和脱氧胞苷。具有修饰的核碱基的核苷的实例包括肌苷 (I)、黄嘌呤 (X)、7-甲基鸟苷 (m7G)、二氢尿苷 (D)、5-甲基胞苷 (m5C) 和假尿苷 (Ψ)。“核苷酸”由核碱基,五碳糖(核糖或脱氧核糖)和至少一个磷酸基团组成。

[0140] 如本文所用,术语“核酸”和“核酸分子”是指包含核碱基和酸性部分(例如,核苷、核苷酸或核苷酸的聚合物)的化合物。通常,聚合核酸,例如包含三个或更多个核苷酸的核酸分子是线性分子,其中相邻的核苷酸通过磷酸二酯键彼此连接。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”是指包含三个或更多个个体核苷酸残基的寡核苷酸链。如本文所用,术语“寡核苷酸”、“多核苷酸”和“多核酸”可互换使用,以指代核苷酸的聚合物(例如,至少三个核苷酸的字符串)。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,例如在基因组、转录物、mRNA、tRNA、rRNA、siRNA、snRNA、质粒、粘粒、染色体、染色单体或其他天然存在的核酸分子的背景下。另一方面,核酸分子可以是非天然存在的分子,例如重组DNA或RNA、人工染色体、工程基因组或其片段,或合成DNA、RNA、DNA/RNA杂种,或包括非天然存在的核苷酸或核苷。此外,术语“核酸”、“DNA”、“RNA”和/或类似术语包括核酸类似物,例如具有除磷酸二酯主链以外的类似物。核酸可以从天然来源中纯化,使用重组表达系统产生,并任选地纯化、化学合成等。在适当的情况下,例如,在化学合成分子的情况下,核酸可以包含核苷类似物,例如具有化学修饰的碱基或糖和骨架修饰。除非另有说明,否则核酸序列以5'至3'方向显示。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);或核苷类似物(例如2-氨基腺苷、2-硫代胸苷、肌苷、吡咯并嘧啶、3-甲基腺苷、5-甲基胞苷、2-氨基腺苷、C5-溴尿苷、C5-氟尿苷、C5-碘尿苷、C5-丙炔基尿苷、C5-丙炔基胞苷、C5-甲基胞嘧啶、2-氨基腺苷、7-脱氮杂腺苷、7-脱氮鸟苷、8-氧代腺苷、8-氧代鸟苷、06-甲基鸟嘌呤和2-硫代胞苷);化学修饰的碱;生物修饰的碱(例如甲基化的碱);插层基;修饰的糖(例如2'-氟核糖、核糖、2'-脱氧核糖、阿拉伯糖和己糖);和/或修饰的磷酸基团(例如,硫代磷酸酯和5'-N-亚磷酰胺键)。

[0141] 术语“核酸可编程DNA结合蛋白”或“napDNAbp”可与“多核苷酸可编程核苷酸结合域”互换使用,以指代与核酸缔合的蛋白质(例如DNA或RNA),例如引导核酸,将napDNAbp引导至特定核酸序列。例如,Cas9蛋白可以与指导RNA结合,所述指导RNA将Cas9蛋白引导至与指导RNA互补的特定DNA序列。在一些实施方案中,napDNAbp是Cas9结构域,例如核酸酶活性Cas9、Cas9切口酶 (nCAs9) 或核酸酶非活性Cas9 (dCas9)。核酸可编程DNA结合蛋白的实例包括但不限于Cas9(例如dCas9和nCAs9)、Cas12a/Cpf1、Cas12b/C2c1、Cas12c/C2c3、Cas12d/CasY、Cas12e/CasX、Cas12g、Cas12h和Cas12i。其他核酸可编程的DNA结合蛋白也在本申请的范围内,尽管它们可能未在本申请中具体列出。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*.2019Jan4;363(6422):88-91.doi:10.1126/science.aav7271,每个文献的全部内容通过引用并入本文。

[0142] 如本文所用,术语“核碱基编辑结构域”或“核碱基编辑蛋白”是指可催化RNA或DNA中的核碱基修饰的蛋白质或酶,例如胞嘧啶(或胞苷)变为尿嘧啶(或尿苷)或胸腺嘧啶(或胸腺嘧啶核苷)、以及腺嘌呤(或腺苷)对次黄嘌呤(或肌苷)的脱氨作用,以及非模板核苷酸的添加和插入。在一些实施方案中,核碱基编辑结构域是脱氨酶结构域(例如胞苷脱氨酶、胞嘧啶脱氨酶、腺嘌呤脱氨酶或腺苷脱氨酶)。在一些实施方案中,核碱基编辑结构域可以是天然存在的核碱基编辑结构域。在一些实施方案中,核碱基编辑结构域可以是来自天然存在的核碱基编辑结构域的工程化或进化的核碱基编辑结构域。核碱基编辑结构域可以来自任何生物,例如细菌、人类、黑猩猩、大猩猩、猴子、牛、狗、大鼠或小鼠。例如,核碱基编辑蛋白在国际第PCT/2017/045381号专利申请(WO第2018/027078号国际专利申请)和国际第PCT/US2016/058344号专利申请(WO第2017/070632号国际专利申请)中描述,其各自通过引用整体并入本文。另请参阅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);以及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),其内容通过引用合并于此。

[0143] 如本文中所使用的,“获得”与“获得试剂”中的包括合成、购买或以其他方式获得试剂。

[0144] 如本文所用,“患者”或“受试者”是指被诊断患有、具有患上或发育中、或怀疑患有或发展成疾病或病症的哺乳动物受试者或个体。在一些实施方案中,术语“患者”是指患有疾病或病症的可能性高于平均可能性的哺乳动物受试者。示例性患者可以是人类、非人类灵长类动物、猫、狗、猪、牛、猫、马、骆驼、美洲驼、山羊、绵羊、啮齿动物(例如小鼠、兔子、大鼠或豚鼠)以及其他可以受益于本文公开的疗法。示例性的人类患者可以是男性和/或女性。

[0145] “有此需要的患者”或“有此需要的受试者”在本文中是指被诊断为或怀疑患有疾病或病症的患者,例如但不限于 α -1抗胰蛋白酶缺乏症(A1AD)。

[0146] 术语“致病性突变”、“致病性变异”、“疾病外壳突变”、“引起疾病的变异”、“有害突变”或“易感突变”是指遗传变异或突变,会增加个体的易感性或易患某种疾病或病症。在一些实施方案中,致病性突变包含由基因编码的蛋白质中的至少一种致病性氨基酸取代的至少一种野生型氨基酸。

[0147] 术语“蛋白质”、“肽”、“多肽”及其语法等同物在本文可互换使用,是指通过肽(酰胺)键连接在一起的氨基酸残基的聚合物。所述术语是指任何大小、结构或功能的蛋白质、肽或多肽。通常,蛋白质、肽或多肽的长度至少为三个氨基酸。蛋白质、肽或多肽可以指单个蛋白质或蛋白质集合。蛋白质、肽或多肽中的一个或多个氨基酸可以被修饰,例如,通过添加化学实体,例如碳水化合物基团、羟基、磷酸基团、法尼基、异法尼基、脂肪酸基团,用于缀合、功能化或其他修饰的接头等。蛋白质、肽或多肽也可以是单分子或可以是多分子复合

物。蛋白质、肽或多肽可以只是天然存在的蛋白质或肽的片段。蛋白质、肽或多肽可以是天然存在的、重组的或合成的,或其任何组合。如本文所用,术语“融合蛋白”是指包含来自至少两种不同蛋白的蛋白结构域的杂合多肽。一种蛋白质可以位于融合蛋白的氨基末端(N-末端)部分或羧基末端(C-末端)蛋白质上,从而分别形成氨基末端融合蛋白或羧基末端融合蛋白。蛋白质可以包含不同的结构域,例如核酸结合结构域(例如,指导蛋白质与靶位点结合的Cas9的gRNA结合结构域)和核酸切割结构域或核酸的催化结构域酸编辑蛋白。在一些实施方案中,蛋白质包含蛋白质部分,例如构成核酸结合结构域的氨基酸序列,和有机化合物,例如可以充当核酸切割剂的化合物。在一些实施方案中,蛋白质与核酸例如RNA或DNA复合或相关。本文提供的任何蛋白质可以通过本领域已知的任何方法产生。例如,本文提供的蛋白可通过重组蛋白表达和纯化产生,其特别适合于包含肽接头的融合蛋白。重组蛋白表达和纯化的方法是众所周知的,包括Green和Sambrook, *Molecular Cloning: A Laboratory Manual* (第4版, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 纽约(2012))中描述的方法,全部内容通过引用将其合并于此。

[0148] 本文公开的多肽和蛋白质(包括其功能部分和功能变体)可以包含合成氨基酸代替一种或多种天然存在的氨基酸。这样的合成氨基酸是本领域已知的,并且包括例如氨基环己烷羧酸、正亮氨酸、 α -氨基正癸酸、高丝氨酸、S-乙酰氨基甲基-半胱氨酸、反式3-和反式4-羟基脯氨酸、4-氨基苯丙氨酸、4-硝基苯丙氨酸、4-氯苯丙氨酸、4-羧基苯丙氨酸、 β -苯基丝氨酸、 β -羟基苯丙氨酸、苯基甘氨酸、 α -萘丙氨酸、环己基丙氨酸、环己基甘氨酸、二氢吡啶-2-羧酸、1,2,3,4-四氢异喹啉-3-羧酸、氨基丙二酸、氨基丙二酸单酰胺、N'-苄基-N'-甲基赖氨酸、N', N'-二苄基赖氨酸、6-羟基赖氨酸、鸟氨酸、 α -氨基环戊烷羧酸、 α -氨基环己烷羧酸、 α -氨基环庚烷羧酸、 α -(2-氨基-2-降冰片烷)-羧酸、 α , γ -二氨基丁酸、 α , β -二氨基丙酸、高苯丙氨酸和 α -叔丁基甘氨酸。多肽和蛋白质可以与多肽构建体的一个或多个氨基酸的翻译后修饰相关。翻译后修饰的非限制性实例包括磷酸化、酰化(包括乙酰化和甲酰化)、糖基化(包括N-连接和O-连接)、酰胺化、羟基化、烷基化(包括甲基化和乙基化)、泛素化、吡咯烷酮羧酸的加成、二硫键的形成、硫酸化、肉豆蔻酰化、棕榈酰化、异戊二烯化(isoprenylation)、法尼基化、香叶酰化(geranylation)、糖基磷脂酰肌醇化(glypiation)、硫辛酰化(lipoylation)和碘化。

[0149] 术语“多核苷酸可编程核苷酸结合结构域”是指与核酸(例如,DNA或RNA)缔合的蛋白质,例如指导多核苷酸可编程DNA结合域的指导多核苷酸(例如,指导RNA),其将多核苷酸可编程DNA结合结构域引导至特定核酸序列。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蛋白可以与指导RNA结合,所述指导RNA将Cas9蛋白引导至与指导RNA互补的特定DNA序列。在一些实施方案中,多核苷酸可编程核苷酸结合结构域是Cas9结构域,例如核酸酶活性Cas9、Cas9切口酶(nCas9)或核酸酶失活Cas9(dCas9)。核酸可编程DNA结合蛋白的非限制性实例包括Cas9(例如,dCas9和nCas9)、Cas12a/Cpf1、Cas12b/C2c1和Cas12c/C2c3、Cas12d/CasY、Cas12e/CasX、Cas12g、Cas12h和Cas12i。Cas酶的非限制性实例包括Cas1、Cas1B、Cas2、Cas3、Cas4、Cas5、Cas5d、Cas5t、Cas5h、Cas5a、Cas6、Cas7、Cas8、Cas8a、Cas8b、Cas8c、Cas9(也称为Csn1或Csx12)、Cas10、Cas10d、Cas12a/Cpf1、Cas12b/

C2c1、Cas12c/C2c3、Cas12d/CasY、Cas12e/CasX、Cas12g、Cas12h、Cas12i、Csy1、Csy2、Csy3、Csy4、Cse1、Cse2、Cse3、Cse4、Cse5e、Csc 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效应子蛋白、CARF、DinG、其同源物或其修饰或设计化形式。其他核酸可编程的DNA结合蛋白也在本申请的范围内,尽管它们未在本申请中具体列出。

[0150] 本文在蛋白质或核酸的上下文中使用的术语“重组”是指自然界中不存在而是人类工程化的产物的蛋白质或核酸。例如,在一些实施方案中,重组蛋白或核酸分子包含氨基酸或核苷酸序列,所述氨基酸或核苷酸序列包含至少一个、至少两个、至少三个、至少四个、至少五个、至少六个或至少一个与任何自然发生的序列相比有七个突变。

[0151] “减少”是指至少10%、25%、50%、75%或100%的负变化。

[0152] “参考”是指标准或控制条件。在一个实施方案中,参照是野生型或健康细胞。

[0153] “参考序列”是定义的序列,用作序列比较的基础。参考序列可以是指定序列的子集或全部。例如,全长cDNA或基因序列的片段,或完整的cDNA或基因序列。对于多肽,参考多肽序列的长度通常为至少约16个氨基酸,优选至少约20个氨基酸,更优选至少约25个氨基酸,甚至更优选约35个氨基酸、约50个氨基酸、或约100个氨基酸。对于核酸,参考核酸序列的长度通常为至少约50个核苷酸,优选至少约60个核苷酸,更优选至少约75个核苷酸,甚至更优选约100个核苷酸或约300个核苷酸或其附近或之间的任何整数。

[0154] 术语“RNA可编程核酸酶”和“RNA指导的核酸酶”与不是切割靶标的一种或多种RNA一起使用(例如,结合或相关)。在一些实施方案中,当RNA可编程的核酸酶与RNA复合时,可被称为核酸酶:RNA复合物。通常,结合的RNA被称为指导RNA(gRNA)。gRNA可以以两个或多个RNA的复合体形式存在,也可以以单个RNA分子形式存在。以单链RNA分子形式存在的gRNA可以称为单链指导RNA(sgRNA),尽管“gRNA”可以互换使用,是指以单个分子或两个或多个分子的复合物形式存在的指导RNA。通常,以单链RNA种类存在的gRNA包含两个域:(1)与靶核酸具有同源性的域(例如,并指导Cas9复合物与靶标结合);(2)结合Cas9蛋白的结构域。在一些实施方案中,结构域(2)对应于称为tracrRNA的序列,并包含茎环结构。例如,在一些实施方案中,结构域(2)是与Jinek等人在Science337:816-821(2012)中所提供的tracrRNA相同或同源,其全部内容通过引用并入本文。gRNA的其他实例(例如,包括结构域2的所述些)可以于2013年9月6日提交的名称为“Switchable Cas9 Nucleases and Uses”的美国U.S.S.N.第61/874,682号专利临时申请以及于2013年9月6日提交名称为“Delivery System For Functional Nucleases”的美国专利申请U.S.S.N.第61/874,746号专利临时申请的全部内容通过引用整体并入本文。在一些实施方案中,gRNA包含结构域(1)和(2)中的两个或更多个,并且可以被称为“延伸的gRNA”。例如,如本文所述,延伸的gRNA将例如结合两个或更多个Cas9蛋白且在两个或更多个不同的区域结合靶核酸。gRNA包含与靶位点互补的核苷酸序列,其介导核酸酶/RNA复合物与所述靶位点的结合,提供核酸酶:RNA复合物的序列特异性。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,来自化脓链球菌的Cas9(Csn1)(参见,例如,“Complete genome sequence of an M1 strain of Streptococcus pyogenes.”,Ferretti JJ.,McShan W.M.,Ajdic D.J.,

Savic D.J.、Savic G.、Lyon K.、Primeaux C.、Sezate S.、Suvorov A.N.、Kenton S.、Lai H.S.、Lin S.P.、Qian Y.、Jia H.G.、Najar F.Z.、Ren Q.、Zhu H.、Song L.、White J.、Yuan X.、Clifton S.W.、Roe B.A.、McLaughlin R.E.,*Proc.Natl.Acad.Sci.USA* 98:4658-4663 (2001);“CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.”, Deltcheva E.、Chylinski K.、Sharma C.M.、Gonzales K.、Chao Y.、Pirzada Z.A.、Eckert M.R.、Vogel J.、Charpentier E.,*Nature* 471:602 -607(2011)。

[0155] 术语“单核苷酸多态性(SNP)”是发生在基因组中特定位置的单链核苷酸中的变异,其中每个变异在群体中以一定程度(例如> 1%)存在。例如,在人类基因组的特定碱基位置,C核苷酸可以出现在大多数个体中,但在少数个体中,所述位置被A占据。这意味着在所述特定位置存在一个SNP,并且这两个可能的核苷酸变异C或A被认为是所述位置的等位基因。SNP是对疾病易感性的差异基础。疾病的严重程度以及人体对治疗的反应方式也是遗传变异的表现。SNP可以落入基因的编码区域,基因的非编码区域或基因间区域(基因之间的区域)内。在一些实施方案中,由于遗传密码的简并性,编码序列内的SNP不一定改变所产生的蛋白质的氨基酸序列。编码区中的SNP有两种类型:同义和非同义SNP。同义SNP不会影响蛋白质序列,而非同义SNP会改变蛋白质的氨基酸序列。非同义SNP有两种类型:错义和无义。不在蛋白质编码区的SNP仍会影响基因剪接,转录因子结合,信使RNA降解或非编码RNA的序列。受这种SNP类型影响的基因表达称为eSNP(表达SNP),可以在基因的上游或下游。单核苷酸变异体(SNV)是单链核苷酸的变异体,没有频率限制,可以在体细胞中产生。体细胞单核苷酸变异(例如,由癌症引起)也可以称为单核苷酸变异。

[0156] “SERPINA1多核苷酸”是指编码A1AT蛋白或其片段的核酸分子。示例性SERPINA1多核苷酸的序列,其可见于NCBI登录号第NM_000295号,并提供如下:

```
1 acaatgactc ctttcggtaa gtgcagtga agctgtacac tgcccaggca aagcgtccgg
61 gcagcgtagg cgggcgactc agatcccagc cagtggactt agcccctgtt tgctctccg
121 ataactgggg tgaccttggg taatattcac cagcagctc ccccgttgcc cctctggatc
181 cactgcttaa atacggacga ggacagggcc ctgtctctc agcttcagge accaccactg
[0157] 241 acctgggaca gtgaatcgac aatgccgtct tctgtctcgt ggggcacctc cctgtctggca
301 ggcctgtgct gcctgggtccc tgtctccctg gctgaggatc ccaggggaga tgctgccag
361 aagacagata catccacca tgatcaggat cacccaacct tcaacaagat caccaccaac
421 ctggctgagt tcgccttcag cctataccgc cagctggcac accagtccaa cagcaccaat
481 atcttcttct cccagtgag catcgctaca gcctttgcaa tgctctccct ggggaccaag
```

541 gctgacactc acgatgaaat cctggagggc ctgaattca acctcacgga gattccggag
601 gctcagatcc atgaaggctt ccaggaactc ctccgtaccc tcaaccagcc agacagccag
661 ctccagctga ccaccggcaa tggcctgttc ctacgcgagg gcctgaagct agtggataag
721 ttttggagg atgttaaaaa gtgtaccac tcagaagcct tcactgtcaa ctccggggac
781 accgaagagg ccaagaaaca gatcaacgat tacgtggaga aggggtactca agggaaaatt
841 gtggatttgg tcaaggagct tgacagagac acagttttg ctctggtgaa ttacatcttc
901 tttaaaggca aatgggagag accctttgaa gtcaaggaca ccgagggaaga ggacttccac
961 gtggaccagg tgaccaccgt gaagggtcct atgatgaagc gtttaggcat gtttaacatc
1021 cagcactgta agaagctgtc cagctgggtg ctgctgatga aatactggg caatgccacc
1081 gccatcttct tcctgcctga tgaggggaaa ctacagcacc tggaaaatga actcaccac
1141 gataatcata ccaagttcct ggaaaatgaa gacagaaggt ctgccagctt acatttacc
1201 aaactgtcca ttactggaac ctatgatctg aagagcgtcc tgggtcaact gggcatcact
1261 aaggcttcca gcaatggggc tgacctctcc ggggtcacag aggaggcacc cctgaagctc
1321 tccaaggccg tgcataaggc tgtgctgacc atcgacgaga aagggactga agctgctggg
1381 gccatgtttt tagaggccat acctatgtct atccccccg aggtcaagtt caacaaacce
1441 ttgtcttct taatgattga acaaaatacc aagtcctccc tcttcatggg aaaagtgggtg
1501 aatcccaccc aaaataaact gcctctcgtc cctcaacccc tccccctcat ccttgcccc
1561 ctccctggat gacattaaag aagggttgag ctggtccctg cctgcatgtg actgtaaate
1621 cctcccatgt ttctctgag tctcccttg cctgctgagg ctgtatgtgg gctccaggta
1681 acagtgtgt ctccgggccc cctgaactgt gticattggag catctggctg ggtaggcaca
[0158] 1741 tgctgggctt gaatccaggg gggactgaat cctcagctta cggacctggg cccatctgtt
1801 tctggagggc tccagtcttc ctgtcctgt ctggagtc ccaagaagga atcacagggg
1861 aggaaccaga taccagccat gaccccagc tccaccaagc atcttcatgt cccccctc
1921 atccccact cccccccacc cagagtgtct catctgcca gggctggctg tgcccccc
1981 aaggctgccc tctgggggc ccagaactg cctgacgtg ccgtggcca gttttgtgg
2041 atctgcagca acacaagaga gaggacaatg tctcctctt gacccgtgt cacctaacca
2101 gactcgggcc ctgcacctc caggcactc tggaaaatga ctgaggcaga ttcttctga
2161 agccattct ccatggggca acaaggacac ctattctgtc ctgtccttc catcgtgcc
2221 ccagaaagcc tcacatact cgttttagaa tcaggctcct tctcccaga tgaaggagg
2281 ggtctctgt ttgtttctc tatctctcc tcagactga ccaggcccag caggcccag
2341 aagaccatta cctatatcc ctctctcc ctatgcacat ggccataggc ctgctgatgg
2401 ctcaggaagg ccattgcaag gactctcag ctatgggaga ggaagccat caccattga
2461 ccccgcaac cctccctt cctctctga gtcccgactg gggccacatg cagcctgact
2521 tctttgtgcc tgttctgtc cctgcagtct tcagagggcc accgcagctc cagtgccacg
2581 gcaggaggt gtctctgaat agcccctgt glaagggcca ggagagtcct tcatcctcc
2641 aaggccctgc taaaggacac agcagccagg aagtcctctg ggcccctagc tgaaggacag
2701 cctgctcct cgtctctac caggaatggc ctgtcctat ggaaggcact gccccatccc
2761 aaactaatct aggaatcact gtctaaccac tcactgtcat gaatgtgtac ttaaaggatg
2821 aggttgagtc ataccaata gtgatttga tagttcaaaa tgggtgaaatt agcaattcta
2881 catgattcag tctaataat ggataccgac tgtttccac acaagtctcc tgttcttta
2941 agcttactca ctgacagcct ttactctcc acaatacat taaagatatg gccatcacca

3001 agccccctag gatgacacca gacctgagag tctgaagacc tggatccaag ttctgacttt
[0159] 3061 tccccctgac agctgtgtga ccttcgtgaa gtcgccaac ctctctgagc cccagtcatt
3121 gctagtaaga cctgcctttg agttggtagt atgttcaagt tagataacaa aatgtttata
3181 cccattagaa cagagaataa atagaactac atttcttgca

[0160] PAM序列以斜体和双底线显示,并显示腺苷碱基编辑后的正确序列。

[0161] “特异性结合”是指识别并结合多肽和/或核酸的核酸分子、多肽或其复合物(例如,核酸可编程DNA结合域和指导核酸)、化合物或分子。本发明的酸性分子,但是基本上不识别并结合样品,例如生物样品中的其他分子。

[0162] 可用于本发明方法的核酸分子包括编码本发明多肽或其片段的任何核酸分子。这样的核酸分子不必与内源核酸序列100%相同,但是通常将显示出实质同一性。与内源序列具有“基本同一性”的多核苷酸通常能够与双链核酸分子的至少一条链杂交。可用于本发明方法的核酸分子包括编码本发明多肽或其片段的任何核酸分子。这样的核酸分子不必与内源核酸序列100%相同,但是通常将显示出实质同一性。与内源序列具有“基本同一性”的多核苷酸通常能够与双链核酸分子的至少一条链杂交。“杂交”是指在各种严格条件下在互补多核苷酸序列(例如,本文描述的基因)或其部分之间形成双链分子的对。(参见例如Wahl, G.M.和S.L.Berger (1987) *Methods Enzymol.* 152:399; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507)。

[0163] 例如,严格的盐浓度通常小于约750 mM的NaCl和75 mM的柠檬酸三钠,优选小于约500 mM的NaCl和50 mM的柠檬酸三钠,更优选小于约250 mM的NaCl和25 mM的柠檬酸三钠盐。在不存在有机溶剂例如甲酰胺的情况下可获得低严格度杂交,而在至少约35%的甲酰胺,更优选至少约50%的甲酰胺的存在下可以获得高严谨度杂交。严格的温度条件通常将包括至少约30℃,更优选至少约37℃,最优选至少约42℃的温度。改变其他参数,例如杂交时间、去污剂的浓度,例如,十二烷基硫酸钠(SDS)以及包含或排除载体DNA是本领域技术人员众所周知的。通过根据需要组合这些各种条件来实现各种严谨度。在一个优选的实施方案中,杂交将在30℃在750mM的NaCl、75mM的柠檬酸三钠和1% SDS中发生。在一个更优选的实施方案中,杂交将在37℃下在500mM的NaCl、50mM的柠檬酸三钠、1%的SDS、35%的甲酰胺和100μg/ml的变性鲑鱼精DNA(ssDNA)中进行。在最优选的实施方案中,杂交将在42℃在250mM的NaCl、25mM的柠檬酸三钠、1%的SDS、50%的甲酰胺和200μg/ml的ssDNA中发生。在这些条件下有用的变化对于本领域技术人员是显而易见的。

[0164] 对于大多数应用,杂交后的洗涤步骤的严谨度也会有所不同。洗涤严谨度条件可以通过盐浓度和温度来定义。如上所述,可以通过降低盐浓度或通过提高温度来提高洗涤严谨度。例如,洗涤步骤的严谨的盐浓度将优选小于约30mM的NaCl和3mM的柠檬酸三钠,最优选小于约15mM的NaCl和1.5mM的柠檬酸三钠。洗涤步骤的严格温度条件通常包括至少约25℃,更优选至少约42℃,甚至更优选至少约68℃的温度。在优选的实施方案中,洗涤步骤将为在25℃的30mM的NaCl、3mM的柠檬酸三钠和0.1%的SDS中发生。在更优选的实施方案中,洗涤步骤将在42℃下在15mM的NaCl、1.5mM的柠檬酸三钠和0.1%的SDS中进行。在更优选的实施方案中,洗涤步骤将在68℃下在15mM的NaCl、1.5mM的柠檬酸三钠和0.1%的SDS中进行。这些条件的其他变化对于本领域技术人员是显而易见的。杂交技术是本领域技术人员众所周知的,并且描述于例如Benton和Davis (*Science* 196:180, 1977); Grunstein和

Hogness (Proc. Natl. Acad. Sci., USA 72:3961, 1975); Ausubel 等人 (Current Protocols in Molecular Biology, Wiley Interscience, 纽约, 2001); Berger 和 Kimmel (Guide to Molecular Cloning Techniques, 1987 年, Academic Press, 纽约); 和 Sambrook 等人, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, 纽约。

[0165] “基本上相同”是指与参考氨基酸序列 (例如, 本文所述的任何一种氨基酸序列) 或核酸序列 (例如, 本文所述的任何一种核酸序列)。优选地, 这种序列在氨基酸水平或核酸水平上与用于比较的序列至少 60%, 更优选 80% 或 85%, 并且更优选 90%、95% 或甚至 99% 相同。

[0166] 序列同一性通常使用序列分析软件 (例如, 威斯康星大学 (1710 University Avenue, Madison, Wis. 53705) 生物技术中心遗传学计算机组的序列分析软件包、BLAST、BESTFIT、GAP 或 PILEUP/PRETTYBOX 程序进行测量)。此类软件通过将同源性程度分配给各种替换、缺失和/或其他修饰来匹配相同或相似的序列。保守取代通常包括以下组内的取代: 甘氨酸、丙氨酸; 缬氨酸、异亮氨酸、亮氨酸; 天冬氨酸、谷氨酸、天冬酰胺、谷氨酰胺; 丝氨酸、苏氨酸、赖氨酸、精氨酸; 和苯丙氨酸、酪氨酸。在确定同一性程度的示例性方法中, 可以使用 BLAST 程序, 其中 e^{-3} 和 e^{-100} 之间的概率分数表示密切相关的序列。

[0167] 例如, COBALT 用于以下参数:

[0168] a) 对齐参数: 空位罚分 -11、-1 和空位罚分 5、-1,

[0169] b) CDD 参数: 使用 RPS BLAST; Blast E 值 0.003; 搜寻保守列并重新计算, 然后

[0170] c) 查询聚类参数: 使用查询聚类; 字元长度 4; 最大丛集距离 0.8; 常规字母。

[0171] EMBOSS Needle 例如使用以下参数:

[0172] a) 矩阵: BLOSUM62;

[0173] b) 开放空位: 10;

[0174] c) 扩展空位: 0.5;

[0175] d) 输出格式: 成对;

[0176] e) 最终空位罚分: 假;

[0177] f) 最终空位开启: 10; 以及

[0178] g) 最终空位扩展: 0.5。

[0179] “受试者”是指哺乳动物, 包括但不限于人类或非人类哺乳动物, 例如牛、马、犬、绵羊或猫。

[0180] 术语“靶位点”是指核酸分子内被核碱基编辑器修饰的序列。在一个实施方案中, 靶位点被脱氨酶或包含脱氨酶的融合蛋白 (例如胞苷或腺嘌呤脱氨酶) 脱氨。

[0181] 因 RNA 可编程核酸酶 (例如, Cas9) 使用 RNA:DNA 杂交来靶向 DNA 切割位点, 所以这些蛋白质原则上能够靶向由指导 RNA 指定的任何序列。使用 RNA 可编程核酸酶 (例如 Cas9) 进行位点特异性切割 (例如, 修饰基因组) 的方法是本领域已知的 (例如, 参见 Cong, L. 等人, Multiplex genome engineering using CRISPR/Cas systems., Science 339, 819-823 (2013); Mali, P. 等人, RNA-guided human genome engineering via Cas9., Science 339, 823-826 (2013); Hwang, W.Y. 等人, Efficient genome editing in zebrafish using a CRISPR-Cas system., Nature biotechnology 31, 227-229 (2013); Jinek, M. 等人, RNA-

programmed genome editing in human cells.eLife 2,e00471(2013);Dicarlo,J.E.等人,Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems.,Nucleic acids research(2013);Jiang,W.等人,RNA-guided editing of bacterial genomes using CRISPR-Cas systems.,Nature biotechnology 31,233-239(2013);每个的全部内容通过引用并入本文)。

[0182] 如本文所用,术语“治疗”是指减轻或改善与之相关的病症和/或症状或获得所需的药理和/或生理作用。应当理解,尽管没有排除,但治疗疾病或病症并不需要完全消除与之相关的疾病、病症或症状。在一些实施方案中,所述作用是治疗性的,即,但不限于,所述作用部分或完全减轻(reduces)、减少(diminishes)、消除(abrogates)、减轻(abates)、减轻(alleviates)、减轻(decreases)疾病的强度或治愈(cures)疾病和/或归因于所述疾病的不良症状。在一些实施方案中,所述作用是预防性的,即所述作用保护或预防疾病或病症的发生或再发生。为此,本发明公开的方法包括给予治疗有效量的本文所述的组合物。

[0183] “尿嘧啶糖基化酶抑制剂”是指抑制尿嘧啶切除修复系统的药剂。在一个实施方案中,所述试剂是结合宿主尿嘧啶-DNA糖基化酶并防止从DNA去除尿嘧啶残基的蛋白质或其片段。

[0184] 本文提供的范围应理解为所述范围内所有值的简写。例如,1到50的范围应理解为包括任何数字,数字的组合或选自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的子范围。

[0185] 在本文对变量的任何定义中列举化学基团的列举包括将所述变量定义为所列基团的任何单个基团或组合。本文中对变量或方面的实施例的叙述包括所述实施例作为任何单个实施例或与任何其他实施例或其部分组合。

[0186] 本文提供的任何组合物或方法可以与本文提供的任何其他组合物或方法中的一种或多种结合。

[0187] DNA编辑已成为通过在遗传水平上校正致病突变来改变疾病状态的可行方法。直到最近,所有的DNA编辑平台都通过在指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的结果来起作用,从而导致复杂的遗传产物种群。尽管可以通过同源引导修复(homology directed repair,HDR)途径实现精确的用户定义的修复结果,但许多挑战已阻止了在治疗相关细胞类型中使用HDR进行高效修复。在实践中,相对于竞争性的,易于出错的非同源末端连接途径的效率低下。此外,HDR严格限制在细胞周期的G1期和S期,从而防止有丝分裂后细胞中DSB的精确修复。结果,已经证明难以或不可能以用户定义的可编程方式在这些人群中高效地改变基因组序列。

[0188] 核碱基编辑器

[0189] 本文公开了用于编辑、修饰或改变多核苷酸的靶核苷酸序列的碱基编辑器或核碱基编辑器。本文描述的是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的核碱基编辑器或碱基编辑器。当与结合的指导多核苷酸(例如,gRNA)结合时,多核苷酸可编程核苷酸结合结构域可以特异性地结合至靶多核苷酸序列(即,通过结合的引导核酸的碱基与靶多核苷酸序列的碱基之间的互补碱基配对),从而将碱基编辑器定位到希望被编辑的靶核酸序列。在一些实施方案中,靶多核苷酸序列包含单链DNA或双链DNA。在一些实施方案

中,靶多核苷酸序列包含RNA。在一些实施方案中,靶多核苷酸序列包含DNA-RNA杂合体。

[0190] 多核苷酸可编程核苷酸结合域

[0191] 术语“多核苷酸可编程核苷酸结合结构域”是指与核酸(例如,DNA或RNA)相关的蛋白质,例如指导多核苷酸(例如,指导RNA),其将多核苷酸可编程核苷酸结合结构域引导至特定核酸序列。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。在一些实施方案中,多核苷酸可编程核苷酸结合结构域是Cpf1蛋白。

[0192] 应当理解,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。例如,多核苷酸可编程核苷酸结合结构域可以与将多核苷酸可编程核苷酸结合结构域引导至RNA的核酸结合。其他核酸可编程的DNA结合蛋白也在本申请的范围内,尽管它们未在本申请中具体列出。

[0193] 碱基编辑器的多核苷酸可编程核苷酸结合结构域本身可以包含一个或多个结构域。例如,多核苷酸可编程核苷酸结合结构域可包含一个或多个核酸酶结构域。在一些实施方案中,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。在本文中,术语“核酸外切酶”是指能够从自由端消化核酸(例如,RNA或DNA)的蛋白质或多肽,并且术语“核酸内切酶”是指能够催化(例如,切割)内部区域的蛋白质或多肽。在核酸(例如DNA或RNA)中。在一些实施方案中,核酸内切酶可以切割双链核酸的单链。在一些实施方案中,核酸内切酶可以切割双链核酸分子的两条链。在一些实施方案中,多核苷酸可编程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。

[0194] 在一些实施方案中,多核苷酸可编程核苷酸结合结构域的核酸酶结构域可以切割靶多核苷酸的零、一或两条链。在一些情况下,多核苷酸可编程核苷酸结合结构域可包含切口酶结构域。本文中,术语“切口酶”是指包含核酸酶结构域的多核苷酸可编程核苷酸结合结构域,所述核酸酶结构域仅能够切割双链核酸分子(例如DNA)中两条链中的一条链。在一些实施方案中,可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域中来从多核苷酸可编程核苷酸结合域的完全催化活性(例如天然)形式衍生切口酶。例如,在多核苷酸可编程核苷酸结合结构域包含衍生自Cas9的切口酶结构域的情况下,源自Cas9的切口酶结构域可以在位置840处包含D10A突变和组氨酸。在这种情况下,残基H840保留催化活性并因此可以切割核酸双链体的单链。在另一个实例中,Cas9衍生的切口酶结构域可包含H840A突变,而位置10的氨基酸残基仍为D。在一些实施方案中,切口酶可源自多核苷酸的完全催化活性(例如天然)形式。通过去除切口酶活性不需要的全部或部分核酸酶结构域来实现可程序设计的核苷酸结合结构域。例如,在多核苷酸可编程核苷酸结合结构域包含衍生自Cas9的切口酶结构域的情况下,源自Cas9的切口酶结构域可以包含RuvC结构域或HNH结构域的全部或部分的缺失。

[0195] 示例性的催化活性Cas9的氨基酸序列如下:

[0196] MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSVDKLFILQVQTYNQLFEENPINASGVDAKAILSARLSKSR

RLENLIAQLPGEKKNGLFGNLIASLGLTPNFKSNFDLAEDAKLQLSKDQYDDDLNLLAQIGDQYADLFLAAKNL
SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPKEYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRI
PYVVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASQSFIERMTNFDKNLPNEKVLPHSLLYEYFTVYNEL
TKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKI
IKDKDFLDNEENEDILEDIVLTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWRLSRKLINGIRDKQSG
KTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQDLSLHEHIANLAGSPAICKGILQTVKVVDELVKVMG
RHKPENIVIAMARENQTTQKGQKNSRERMKRIEELGSGILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKEA
RGGLSELKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKREINN
YHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVDVRKMIKSEQEI GKATAYFFYSNIMNFFKTEITLANGE
IRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYG
GFDSPTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSEFKNPIDFLEAKGYKEVKKDLIIKLPKYSLELEN
GRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIEQISEFSKRVLAD
ANLDKVL SAYNKHDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRI
DLSQLGGD。

[0197] 因此,包含多核苷酸可编程核苷酸结合结构域的碱基编辑器包含核苷酸酶结构域,所述碱基编辑器能够在特定的多核苷酸靶序列(例如,由结合的指导核酸的互补序列确定)上产生单链DNA断裂(切口)。在一些实施方案中,被包含切口酶结构域(例如,Cas9衍生的切口酶结构域)的碱基编辑器切割的核酸双链体靶多核苷酸序列的链是未被碱基编辑器编辑的链(即,由碱基编辑器切割的片段与包含要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,Cas9衍生的切口酶结构域)的碱基编辑器可以切割靶向于编辑的DNA分子的链。在这种情况下,非靶向链不会被切割。

[0198] 本文还提供了包含多核苷酸可编程核苷酸结合结构域的碱基编辑器,所述多核苷酸可编程核苷酸结合结构域被催化性死亡(即,不能切割靶多核苷酸序列)。在本文中,术语“催化死的”和“核酸酶死的”可互换使用,是指具有一个或多个导致其不能切割核酸链的突变和/或缺失的多核苷酸可编程核苷酸结合结构域。在一些实施方案中,由于一个或多个核酸酶结构域中的特定点突变,催化死亡的多核苷酸可编程核苷酸结合结构域碱基编辑器可缺乏核酸酶活性。例如,在碱基编辑器包含Cas9结构域的情况下,Cas9可同时包含D10A突变和H840A突变。这样的突变使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,催化死亡的多核苷酸可编程核苷酸结合结构域可包含催化结构域(例如RuvC1和/或HNH结构域)的全部或一部分的一个或多个缺失。在进一步的实施方案中,催化性死亡的多核苷酸可编程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部分缺失。

[0199] 本文还考虑能够从多核苷酸可编程核苷酸结合结构域的先前功能形式产生催化死亡的多核苷酸可编程核苷酸结合结构域的突变。例如,在催化死亡的Cas9(“dCas9”)的情况下,提供了具有D10A和H840A以外的突变的变体,其导致核酸酶失活的Cas9。举例来说,此类突变包括在D10和H840处的其他氨基酸取代,或在Cas9的核酸酶结构域内的其他取代(例如,在HNH核酸酶子结构域和/或RuvC1子结构域中的取代)。基于本申请和本领域的知识,其

他合适的核酸酶失活的dCas9结构域对本领域技术人员而言是显而易见的,并且在本申请的范围内。这样的另外的示例性的适合的核酸酶失活的Cas9结构域包括但不限于D10A/H840A,D10A/D839A/H840A和D10A/D839A/H840A/N863A突变域(参见,例如,Prashant等人,CAS9transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering.,Nature Biotechnology.,2013;31(9):833-838,其全部内容通过引用并入本文)。

[0200] 可掺入碱基编辑器中的多核苷酸可编程核苷酸结合结构域的非限制性实例包括CRISPR蛋白衍生的结构域、限制性核酸酶、大范围核酸酶、TAL核酸酶(TALEN)和锌指核酸酶(ZFN)。在一些情况下,碱基编辑器包含多核苷酸可编程核苷酸结合结构域,其包含天然或修饰的蛋白质或其部分,其通过结合的指导核酸能够在CRISPR过程中结合核酸序列(即,成簇的规则间隔的短回文重复序列)介导的核酸修饰。这种蛋白质在本文中被称为“CRISPR蛋白质”。因此,本文公开的是包含多核苷酸可编程核苷酸结合结构域的碱基编辑器,所述多核苷酸可编程核苷酸结合结构域包含CRISPR蛋白的全部或部分(即,包含CRISPR蛋白的全部或部分作为域的碱基编辑器,也称为碱基编辑器的“CRISPR蛋白衍生域”)。与野生型或天然版本的CRISPR蛋白质相比,掺入碱基编辑器中的CRISPR蛋白质的域可以进行修饰。例如,如下所述,相对于野生型或天然形式的CRISPR蛋白,源自CRISPR蛋白的结构域可包含一个或多个突变、插入、缺失、重排和/或重组。

[0201] CRISPR是一种适应性免疫系统,可针对移动遗传组件(病毒,转座因子和结合质粒)提供保护。CRISPR簇包含间隔子,与先前的移动组件互补的序列以及靶向入侵的核酸。CRISPR簇被转录并加工成CRISPR RNA(crRNA)。在II型CRISPR系统中,正确处理pre-crRNA需要反编码小RNA(tracrRNA),内源性核糖核酸酶3(rnc)和Cas9蛋白。tracrRNA可作为核糖核酸酶3辅助pre-crRNA加工的指南。随后,Cas9/crRNA/tracrRNA进行核酸内切切割与间隔物互补的线性或环状dsDNA靶标。不与crRNA互补的靶链首先被核酸内切地切割,然后进行3'-5'核酸外切。实际上,DNA结合和切割通常需要蛋白质和两个RNA。但是,可以对单链指导RNA(“sgRNA”或简称为“gRNA”)进行改造,以便将crRNA和tracrRNA的各个方面都整合到单个RNA物种中。参见例如Jinek M.、Chylinski K.、Fonfara I.、Hauer M.、Doudna J.A.、Charpentier E.Science 337:816-821(2012),其全部内容通过引用结合于此。Cas9识别CRISPR重复序列中的短基序(PAM或原间隔子相邻基序),以帮助区分自身和非自身。

[0202] 在一些实施方案中,本文描述的方法可以利用工程改造的Cas蛋白。指导RNA(gRNA)是一种简短的合成RNA,由Cas结合所需的支架序列和用户定义的~20个核苷酸间隔区组成,所述间隔区定义了要修饰的基因组靶标。因此,本领域技术人员可以改变Cas蛋白的基因组靶标,部分地由与基因组其余部分相比gRNA靶向序列对基因组靶标的特异性来部分确定。

[0203] 在一些实施方案中,gRNA支架序列如下:GUUUUAGAGC UAGAAAUAGC AAGUUA AAAU AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUUUU。

[0204] 在一些实施方案中,掺入碱基编辑器中的源自CRISPR蛋白的域是当与结合的指导核酸结合时能够结合靶多核苷酸的核酸内切酶(例如,脱氧核糖核酸酶或核糖核酸酶)。在一些实施方案中,掺入碱基编辑器中的源自CRISPR蛋白质的域是当与结合的指导核酸结合时能够结合靶多核苷酸的切口酶。在一些实施方案中,掺入碱基编辑器中的源自CRISPR蛋

白的域是与结合的指导核酸结合时能够结合靶多核苷酸的催化性死亡域。在一些实施方案中,与碱基编辑器的CRISPR蛋白衍生的域结合的靶多核苷酸是DNA。在一些实施方案中,与碱基编辑器的CRISPR蛋白来源的结构域结合的靶多核苷酸是RNA。

[0205] 可以在本文中使用的Cas蛋白包括1类和2类。Cas蛋白的非限制性实例包括Cas1、Cas1B、Cas2、Cas3、Cas4、Cas5、Cas5d、Cas5t、Cas5h、Cas5a、Cas6、Cas7、Cas8、Cas9(也称为Csn1或Csx12)、Cas10、Csy1、Csy2、Csy3、Csy4、Cse1、Cse2、Cse3、Cse4、Cse5e、Csc1、Csc2、Csa5、Csn1、Csn2、Csm1、Csm2、Csm3、Csm3、Csm4、Csm5、Csm6、Cmr1、Cmr3、Cmr4、Cmr5、Cmr6、Csb1、Csb2、Csb3、Csx17、Csx14、Csx10、Csx16、CsaX、Csx3、Csx1、Csx1S、Csf1、Csf2、Cs0、Csf4、Csd1、Csd2、Cst1、Cst2、Csh1、Csh2、Csa1、Csa2、Csa3、Csa4、Csa5、Cas12a/Cpf1、Cas12b/C2c1、Cas12c/C2c3、Cas12d/CasY、Cas12e/CasX、Cas12g、Cas12h和Cas12i CARF、DinG、其同源物或已修饰者。未经修饰的CRISPR酶可以具有DNA切割活性,例如Cas9,它具有两个功能性核酸内切酶结构域:RuvC和HNH。CRISPR酶可以指导在靶序列上例如在靶序列内和/或在靶序列的互补序列内切割一条或两条链。例如,CRISPR酶可以指导一条或两条链在大约1、2、3、4、5、6、7、8、9、10、15、20、25、50、100、200、500或自靶序列的第一个或最后一个核苷酸的多个碱基对。

[0206] 可使用编码相对于相应的野生型酶突变的CRISPR酶的载体,使突变的CRISPR酶缺乏切割含有靶序列的靶多核苷酸的一条或两条链的能力。Cas9可以指与野生型示例性Cas9多肽(例如,化脓性链球菌的Cas9)具有至少或至少约50%、60%、70%、80%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%序列同一性和/或序列同源性的多肽。Cas9可以指与野生型示例性Cas9多肽(例如,来自化脓性链球菌)具有最多或最多约50%、60%、70%、80%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%的序列同一性和/或序列同源性的多肽。Cas9可指Cas9蛋白的野生型或修饰形式,其可包含氨基酸变化,例如缺失、插入、取代、变体、突变、融合、嵌合体或上述之任何组合。

[0207] 在一些实施方案中,碱基编辑器的CRISPR蛋白来源的结构域可包括来自溃疡棒状杆菌(*Corynebacterium ulcerans*) (NCBI Refs:NC_015683.1,NC_017317.1);白喉棒状杆菌(*Corynebacterium diphtheria*) (NCBI Ref:NC_016782.1,NC_016786.1);栖蚜蝇螺原体(*Spiroplasma syrphidicola*) (NCBI Ref:NC_021284.1);中间型普氏杆菌(*Prevotella intermedia*) (NCBI Ref:NC_017861.1);中国台湾螺旋体(*Spiroplasma taiwanense*) (NCBI Ref:NC_021846.1);海豚链球菌(*Streptococcus iniae*) (NCBI Ref:NC_021314.1);波罗的海贝尔氏菌(*Belliella baltica*) (NCBI Ref:NC_018010.1);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);无害李斯特菌(*Listeria innocua*) (NCBI Ref:NP_472073.1);空肠弯曲菌(*Campylobacter jejuni*) (NCBI Ref:YP_002344900.1);脑膜炎双球菌(*Neisseria meningitidis*) (NCBI Ref:YP_002342100.1),化脓性链球菌(*Streptococcus pyogenes*)或金黄色葡萄球菌的Cas9的全部或部分。

[0208] 核碱基编辑器的Cas9域

[0209] 术语“Cas9”或“Cas9结构域”是指包含Cas9蛋白或其片段(例如,包含Cas9的活性、非活性或部分活性的DNA切割域的蛋白、和/或Cas9的gRNA结合结构域)。Cas9核酸酶有时也称为casn1核酸酶或CRISPR(规律间隔成簇短回文重复序列)相关的核酸酶。示例性的Cas9

是化脓性链球菌Cas9(spCas9),其氨基酸序列如下。

[0210] MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFGSGETAETRLKRTARRRYTRRKNRICYLQEFSNEMAKVDDSFHRLEESF
LVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLADSTDKADLRILIYALA
HMIKFRGHFLIEGDLNPDNSDVKLFQILVQIYNQLFEENPINASRVDAKAILSA
RLSKSRLENLIAQLPGEKRNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKD
TYDDDLNLLAQIGDQYADFLAAKNLSDAILLSDILRVNSEITKAPLSASMIKR
YDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPL
KDNREKIEKILTRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS
AQSFIERMNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL
SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAY
HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRGMIEERLKTIAHLFDDKVMK
QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLT
FKEDIQKAQVSGQGHSLEQIANLAGSPAIAKKGILQTVKIVDELVKVMGHKPEN
IVIMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLY
LYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRG
KSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQF
YKVBREINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIK
SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRD
FATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYG
GFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSSFENPIDFLEAKG
YKEVKKDLIIKLPKYSLEFENGRKRMLASAGELQKGNELALPSKYVNFLYLAS
HYEKLKGSPEQNEQQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAY
NKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIH
QSITGLYETRIDLSQLGGD (底线: HNH 域; 双底线: RuvC 域)

[0211] Cas9核酸酶序列和结构是本领域技术人员众所周知的(例如,参见“Complete genome sequence of an M1 strain of Streptococcus pyogenes.”,Ferretti等人, J.J.、McShan W.M.、Ajdic D.J.、Savic D.J.、Savic G.、Lyon K.、Primeaux C.、Sezate S.、Suvorov A.N.、Kenton S.、Lai H.S.、Lin S.P.、Qian Y.、Jia H.G.、Najar F.Z.、Ren Q.、Zhu H.、Song L.、White J.、Yuan X.、Clifton S.W.、Roe B.A.、McLaughlin R.E.、Proc.Natl.Acad.Sci.U.S.A.98:4658-4663(2001);“CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.”Deltcheva E.、Chylinski K.、Sharma C.M.、Gonzales K.、Chao Y.、Pirzada Z.A.、Eckert M.R.、Vogel J.、Charpentier E.、Nature471:602-607(2011);and“A programmabledual-RNA-guided DNA endonuclease in adaptive bacterial immunity.”,Jinek M.、Chylinski K.、Fonfara I.、Hauer M.、Doudna J.A.、Charpentier E.Science337:816-821(2012),每个都通过引用并入本文)。已经在各种物种中描述Cas9直向同源物,包括但不限于化脓性链球菌和嗜热链球菌。基于所述公开,其他合适的Cas9核酸酶和序列对本领域技术人员是显而易见的,并且此类Cas9核

酸酶和序列包括来自Chylinski、Rhun及Charpentier,“The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems”(2013)RNA Biology 10:5,726-737;其全部内容通过引用并入本文。

[0212] 在一些方面,核酸可编程DNA结合蛋白(napDNAbp)是Cas9结构域。本文提供非限制性的示例性Cas9结构域。Cas9结构域可以是核酸酶活性Cas9结构域、核酸酶非活性Cas9结构域或Cas9切口酶。在一些实施方案中,Cas9结构域是核酸酶活性结构域。例如,Cas9结构域可以是切割双链核酸的两条链(例如,双链DNA分子的两条链)的Cas9结构域。在一些实施方案中,Cas9结构域包含如本文所述的任何氨基酸序列。在一些实施方案中,Cas9结构域包含的氨基酸序列与本文列出的任何氨基酸序列至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相同。在一些实施方案中,与本文列出的任何氨基酸序列相比,Cas9结构域包含具有1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50个或更多个突变。在一些实施方案中,与本文列出的任何氨基酸序列相比,Cas9结构域包含具有至少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。

[0213] 在一些实施方案中,提供了包含Cas9的片段的蛋白质。例如,在一些实施方案中,蛋白质包含两个Cas9结构域之一:(1) Cas9的gRNA结合结构域;或(2) Cas9的DNA切割结构域。在一些实施方案中,包含Cas9或其片段的蛋白质被称为“Cas9变体”。Cas9变体与Cas9或其片段具有同源性。例如,Cas9变体在以下条件下至少约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。与野生型Cas9具有至少约98%的同源性、至少约99%的同源性、至少约99.5%的同源性或至少约99.9%的同源性。在一些实施方案中,与野生型Cas9相比,Cas9变体可以具有1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、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或更多氨基酸变化的氨基酸。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),使得所述片段与野生型Cas9的相应片段至少约70%相同、至少约80%相同、至少约90%相同、至少约95%相同、至少约96%相同、至少约97%相同、至少约98%相同、至少约99%相同、至少约99.5%相同或至少约99.9%相同。在一些实施方案中,所述片段至少是相应野生型Cas9氨基酸长度的至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%、至少99.5%。在一些实施方案中,所述片段的长度为至少100个氨基酸。在一些实施方案中,所述片段的长度为至少100、150、200、250、300、350、400、450、500、550、600、650、700、750、800、850、900、950、1000、1050、1100、1150、1200、1250或1300个氨基酸。

[0214] 在一些实施方案中,本文提供的Cas9融合蛋白包含Cas9蛋白的全长氨基酸序列,例如本文提供的Cas9序列之一。然而,在其他实施方案中,本文提供的融合蛋白不包含全长Cas9序列,而仅包含其一个或多个片段。本文提供了合适的Cas9结构域和Cas9片段的示例性氨基酸序列,并且Cas9结构域和片段的其他合适的序列对本领域技术人员而言是显而易

见的。

[0215] Cas9蛋白可以与指导RNA缔合,指导RNA将Cas9蛋白引导至与指导RNA具有互补性的特定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Cas9、Cas9切口酶 (nCas9) 或核酸酶失活Cas9 (dCas9)。核酸可编程DNA结合蛋白的实例包括但不限于Cas9 (例如,dCas9和nCas9)、CasX、CasY、Cpf1、Cas12b/C2C1和Cas12c/C2C3。

[0216] 在一些实施方案中,野生型Cas9对应于化脓性链球菌的Cas9 (NCBI参考序列:NC_017053.1,核苷酸和氨基酸序列如下所示)。ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACA AATAGCGTCGGATGGGCGGTGATCACTGATGATTATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAG ACCGCCACAGTATCAAAAAAATCTTATAGGGGCTCTTTTATTTGGCAGTGGAGAGACAGCGGAAGCGACTCGTCT CAAACGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAG ATGGCGAAAGTAGATGATAGTTTCTTTTCATCGACTTGAAGAGTCTTTTTTGGTGGAAGAAGACAAGAAGCATGAAC GTCATCTATTTTTTGAAATATAGTAGATGAAGTTGCTTATCATGAGAAATATCCAACATCTATCATCTGCGAAA AAAATTGGCAGATTCTACTGATAAAGCGGATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGT GGTCATTTTTTGATTGAGGGAGATTTAAATCCTGATAATAGTGATGTGGACAAACTATTTATCCAGTTGGTACAAA TCTACAATCAATTATTTGAAGAAAACCTATTAACGCAAGTAGAGTAGATGCTAAAGCGATTCTTTCTGCACGATT GAGTAAATCAAGACGATTAGAAAATCTCATTGCTCAGCTCCCCGGTGAGAAGAGAAATGGCTTGTTTGGGAATCTC ATTGCTTTGTCATTGGGATTGACCCCTAATTTTAAATCAAATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTT CAAAAGATACTTACGATGATGATTTAGATAATTTATTGGCGCAAATTTGGAGATCAATATGCTGATTTGTTTTTGGC AGCTAAGAATTTATCAGATGCTATTTTACTTTTCAGATATCCTAAGAGTAAATAGTGAAATAACTAAGGCTCCCCTA TCAGCTTCAATGATTAAGCGCTACGATGAACATCATCAAGACTTGACTCTTTTAAAAGCTTTAGTTCGACAACAAC TTCCAGAAAAGTATAAAGAAATCTTTTTTGATCAATCAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAG CCAAGAAGAATTTTATAAATTTATCAAACCAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTGGTGAAACTA AATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACAACGGCTCTATTCCCCATCAAATTCATTGGGTGAGC TGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTTTAAAAGACAATCGTGAGAAGATTGAAAAAATCTT GACTTTTTCGAATTCCTTATTATGTTGGTCCATTGGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCT GAAGAAACAATTACCCCATGGAATTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTATTGAACGCA TGACAAACTTTGATAAAAAATCTTCCAAATGAAAAAGTACTACCAAAACATAGTTTGCTTTATGAGTATTTTACGGT TTATAACGAATTGACAAAGGTCAAATATGTTACTGAGGGAATGCGAAAACCAGCATTCTTTTACAGGTGAACAGAAG AAAGCCATTGTTGATTTACTCTTCAAAACAAATCGAAAAGTAACCGTTAAGCAATTAAAAGAAGATTATTTCAAAA AAATAGAATGTTTTGATAGTGTGAAATTTTCAAGAGTTGAAGATAGATTTAATGCTTCATTAGGCGCCTACCATGA TTTTGCTAAAAATTATTAAGATAAAGATTTTTTGGATAATGAAGAAAATGAAGATATCTTAGAGGATATTGTTTTA ACATTGACCTTATTTGAAGATAGGGGGATGATTGAGGAAAGACTT AAAACATATGCTCACCTCTTTGATGATAAG GTGATGAAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTTTGTCTCGAAAATTGATTAATGGTATTAGGGA TAAGCAATCTGGCAAAACAATATTAGATTTTTTGAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCC ATGATGATAGTTTGACATTTAAAGAAGATATTCAAAAAGCACAGGTGTCTGGACAAGGCCATAGTTTACATGAACAG ATTGCTAACTTAGCTGGCAGTCTTGCTATTAATAAAGGTATTTTACAGACTGTAAAAATTGTTGATGAACTGGTCAA AGTAATGGGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACGTGAAAATCAGACAACCTAAAAGGGCCAGAAAA ATTCGCGAGAGCGTATGAAACGAATCGAAGAAGGTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTT

GAAAATACTCAATTGCAAAATGAAAAGCTCTATCTCTATTATCTACAAAATGGAAGAGACATGTATGTGGACCAAGA
ATTAGATATTAATCGTTTAAAGTGATTATGATGTCGATCACATTGTTCCACAAAGTTTCATTAAAGACGATTCAATAG
ACAATAAGGTACTAACGCGTTCTGATAAAAAATCGTGGTAAATCGGATAACGTTCCAAGTGAAGAAGTAGTCAAAAAG
ATGAAAACTATTGGAGACAACCTTCTAAACGCCAAGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGA
ACGTGGAGGTTTGAGTGAACCTTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAAATCACTAAGC
ATGTGGCACAAATTTTGGATAGTCGCATGAATACTAAATACGATGAAAATGATAAACTTATTCGAGAGGTTAAAGTG
ATTACCTTAAAATCTAAATTAGTTTCTGACTTCGAAAAGATTTCCAATTCTATAAAGTACGTGAGATTAACAATTA
CCATCATGCCCATGATGCGTATCTAAATGCCGTCGTTGGAAGTCTTTGATTAAGAAATATCCAAAAGTTGAATCGG
AGTTTGTCTATGGTGATTATAAAGTTTATGATGTTTCGTAATAATGATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCA
ACCGCAAAATATTTCTTTTACTCTAATATCATGAACTTCTTCAAAACAGAAATTACACTTGCAAATGGAGAGATTCTG
CAAACGCCCTCTAATCGAACTAATGGGGAACTGGAGAAATGTCTGGGATAAAGGGCGAGATTTTGCCACAGTGC
GCAAAGTATTGTCCATGCCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCA
ATTTTACCAAAAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGATCCAAAAAATATGGTGGTTTTGA
TAGTCCAACGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGAAATCGAAGAAGTTAAATCCGTTA
AAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAATCCGATTGACTTTTTAGAAAGCTAAAGGA
TATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACCTAAATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACG
GATGCTGGCTAGTGCCGGAGAATTACAAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATT
TAGCTAGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTTGTGGAGCAGCATAAG
CATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCGTGTTATTTTAGCAGATGCCAATTTAGATAA
AGTTCTTAGTGATATAACAAACATAGAGACAAACCAATACGTGAACAAGCAGAAAATATTATTCATTTATTTACGT
TGACGAATCTTGGAGCTCCCGCTGCTTTTAAATATTTTGATACAACAATTGATCGTAAACGATATACGTCTACAAAA
GAAGTTTTAGATGCCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAGCTAGG
AGGTGACTGA

[0217]

MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFG
SGETAETRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVE
EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLADSTDKADLRILIYLAHMI
KFRGHFLIEGDLNPDNSDVKLFIQLVQIYNQLFEENPINASRVDAKAILSARLS
KSRRLLENLIAQLPGEKRNGLFGNLIASLGLTPNFKSNFDLAEDAKLQLSKDTY
DDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNSEITKAPLSASMIKRYD
EHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILE
KMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPFLKD
NREKIEKILTRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGE
QKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDL
LKIIKDKDFLDNEENEDILEDIVLTLTLFEDRGMIEERLKTIAHLFDDKVMKQLK
RRRYTGWGRLSRKLINGIRDQKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKE
DIQKAQVSGQGHSLEQIANLAGSPAIIKKGILQTVKIVDELVKVMGHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYY
LQNGRDMYVDQELDINRLSDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL
VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQ
EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFAT
VRKVLSPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD
SPTVAYSVLVVAKEKGKSKKLKSVKELLGITIMERSSSFENPIDFLEAKGYKE
VKKDLIIKLPKYSLELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYE
KLKGSPEDEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAITLIHQ
SITGLYETRIDLSQLGGD

[0218] (底线:HNH域;双底线:RuvC域)

[0219] 在一些实施方案中,野生型Cas9对应于或包含以下核苷酸和/或氨基酸序列:

[0220] ATGGATAAAAGTATTCTATTGGTTTAGACATCGGCACTAATCCGTTGGATGGGCTGTCATAACCGAT
GAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACCGTCATTTCGATTAAAAAGAATCTTATCGG
TGCCCTCCTATTTCGATAGTGGCGAAACGGCAGAGGCGACTCGCCTGAAACGAACCGCTCGGAGAAGGTATACACGTC
GCAAGAACCGAATATGTTACTTACAAGAAATTTTAGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTTACCGT
TTGGAAGAGTCCTTCCTTGTGAAGAGGACAAGAAACATGAACGGCACCCCATCTTTGGAAACATAGTAGATGAGGT
GGCATATCATGAAAAGTACCAACGATTTATCACCTCAGAAAAAGCTAGTTGACTCAACTGATAAAGCGGACCTGA
GGTTAATCTACTTGGCTCTTGCCCATATGATAAAGTTCCGTGGGCACCTTTCTCATTGAGGGTGATCTAAATCCGGAC
AACTCGGATGTCGACAACTGTTTCATCCAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGAACCCTATAAATGC
AAGTGGCGTGGATGCGAAGGCTATTCTTAGCGCCCGCCTCTCTAAATCCCGACGGCTAGAAAACCTGATCGCACAAAT
TACCCGGAGAGAAGAAAAATGGGTTGTTTCGGTAACCTTATAGCGCTCTCACTAGGCCTGACACCAAATTTTAAGTCG
AACTTCGACTTAGCTGAAGATGCCAAATTGCAGCTTAGTAAGGACACGTACGATGACGATCTCGACAATCTACTGGC
ACAAATTGGAGATCAGTATGCGGACTTATTTTGGCTGCCAAAAACCTTAGCGATGCAATCCTCCTATCTGACATAC

TGAGAGTTAATACTGAGATTACCAAGGCGCCGTTATCCGCTTCAATGATCAAAAGGTACGATGAACATCACCAAGAC
TTGACACTTCTCAAGGCCCTAGTCCGTCAGCAACTGCCTGAGAAATATAAGGAAATATTCTTTGATCAGTCGAAAAA
CGGGTACGCAGGTTATATTGACGGCGGAGCGAGTCAAGAGGAATTCTACAAGTTTATCAAACCCATATTAGAGAAGA
TGGATGGGACGGAAGAGTTGCTTGTA AAAACTCAATCGCGAAGATCTACTGCGAAAGCAGCGGACTTTTCGACAACGGT
AGCATTCCACATCAAATCCACTTAGGCGAATTGCATGCTATACTTAGAAGGCAGGAGGATTTTTATCCGTTCTCTCAA
AGACAATCGTGAAAAGATTGAGAAAAATCCTAACCTTTCGCATACCTTACTATGTGGGACCCCTGGCCCCGAGGGAAC
CTCGGTTTCGCATGGATGACAAGAAAAGTCCGAAGAAACGATTACTCCATGGAATTTTGAGGAAGTTGTCGATAAAGGT
GCGTCAGCTCAATCGTTCATCGAGAGGATGACCAACTTTGACAAGAATTTACCGAACGAAAAAGTATTGCCTAAGCA
CAGTTTACTTTACGAGTATTTACAGTGTACAATGAACTCACGAAAGTTAAGTATGTCACTGAGGGCATGCGTAAAC
CCGCCTTTCTAAGCGGAGAACAGAAGAAAAGCAATAGTAGATCTGTTATTCAAGACCAACCGCAAAGTGACAGTTAAG
CAATTGAAAGAGGACTACTTTAAGAAAATTGAATGCTTCGATTCTGTGCGAGATCTCCGGGGTAGAAGATCGATTTAA
TGCGTCACTTGGTACGTATCATGACCTCCTAAAGATAATTAAGATAAGGACTTCCTGGATAACGAAGAGAATGAAG
ATATCTTAGAAGATATAGTGTGACTCTTACCCTCTTTGAAGATCGGGAAATGATTGAGGAAAGACTAAAAACATAC
GCTCACCTGTTTCGACGATAAGGTTATGAAACAGTTAAAGAGGCGTCGCTATACGGGCTGGGGACGATTGTGCGGGAA
ACTTATCAACGGGATAAGAGACAAGCAAAGTGGTAAAACTATTCTCGATTTTCTAAAGAGCGACGGCTTCGCCAATA
GGAACTTTATGCAGCTGATCCATGATGACTCTTTAACCTTCAAAGAGGATATACAAAAGGCACAGGTTTCCGGACAA
GGGGACTCATTGCACGAACATATTGCGAATCTTGCTGGTTCGCCAGCCATCAAAAAGGGCATACTCCAGACAGTCAA
AGTAGTGGATGAGCTAGTTAAGGTCATGGGACGTCACAAACCGGAAAACATTGTAATCGAGATGGCACGCGAAAATC
AAACGACTCAGAAGGGGCAAAAAAACAGTCGAGAGCGGATGAAGAGAATAGAAGAGGGTATTAAAGAACTGGGCAGC
CAGATCTTAAAGGAGCATCCTGTGGA AAAATACCCAATTGCAGAACGAGAACTTTACCTCTATTACCTACAAAATGG
AAGGGACATGTATGTTGATCAGGAACTGGACATAAACCGTTTATCTGATTACGACGTCGATCACATTGTACCCCAAT
CCTTTTTGAAGGACGATTCAATCGACAATAAAGTGCTTACACGCTCGGATAAGAACCGAGGGAAAAGTGACAATGTT
CCAAGCGAGGAAGTCGTAAAGAAAATGAAGAACTATTGGCGGCAGCTCCTAAATGCGAAACTGATAACGCAAAGAAA
GTTTCGATAACTTAACTAAAGCTGAGAGGGGTGGCTTGCTGAACTTGACAAGGCCGATTTATTAAACGTCAGCTCG
TGGA AACCCGCCAAATCACAAAGCATGTTGCACAGATACTAGATTCCCGAATGAATACGAAATACGACGAGAACGAT
AAGCTGATTCGGGAAGTCAAAGTAATCACTTTAAAGTCAAATTTGGTGTGCGACTTCAGAAAGGATTTTCAATTCTA
TAAAGTTAGGGAGATAAAATAACTACCACCATGCGCACGACGCTTATCTTAATGCCGTCGTAGGGACCGCACTCATT
AGAAATACCCGAAGCTAGAAAGTGAGTTTGTGTATGGTGATTACAAAGTTTATGACGTCCGTAAGATGATCGCGAAA
AGCGAACAGGAGATAGGCAAGGCTACAGCCAAATACTTCTTTTATTCTAACATTATGAATTTCTTTAAGACGGAAAT
CACTCTGGCAAACGGAGAGATACGCAAACGACCTTTAATTGAAACCAATGGGGAGACAGGTGAAATCGTATGGGATA
AGGGCCGGGACTTCGCGACGGTGAGAAAAAGTTTGTCCATGCCCCAAGTCAACATAGTAAAGAAAAGTGGAGTGCAG
ACCGGAGGGTTTTTCAAAGGAATCGATTCTTCAAAAAGGAATAGTGATAAGCTCATCGCTCGTAAAAAGGACTGGGA
CCCGAAAAAGTACGGTGGCTTCGATAGCCCTACAGTTGCCTATTCTGTCTAGTAGTGGCAAAAGTTGAGAAGGGAA
AATCCAAGAACTGAAGTCAGTCAAAGAATTATTGGGGATAACGATTATGGAGCGCTCGTCTTTTGAAAAGAACCCC
ATCGACTTCCTTGAGGCGAAAGGTTACAAGGAAGTAAAAAAGGATCTCATAATTAACTACCAAAGTATAGTCTGTT
TGAGTTAGAAAATGGCCGAAAACGGATGTTGGCTAGCGCCGAGAGCTTCAAAGGGGAACGAACTCGCACTACCGT
CTAAATACGTGAATTTCTGTATTTAGCGTCCCATACGAGAAGTTGAAAGGTTACCTGAAGATAACGAACAGAAG
CAACTTTTTGTTGAGCAGCACAAACATTATCTCGACGAAATCATAGAGCAAATTTCCGAATTCAGTAAGAGAGTCAT
CCTAGCTGATGCCAATCTGGACAAAGTATTAAGCGCATACAACAAGCACAGGGATAAACCCATACGTGAGCAGGCGG

AAAATATTATCCATTTGTTTACTCTTACCAACCTCGGCGCTCCAGCCGCATTCAAGTATTTTGACACAACGATAGAT
CGCAAACGATACACTTCTACCAAGGAGGTGCTAGACGCGACACTGATTCACCAATCCATCACGGGATTATATGAAAC
TCGGATAGATTTGTACAGCTTGGGGGTGACGGATCCCCAAGAAGAAGAGGAAAGTCTCGAGCGACTACAAAGACC
ATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGGCTGCAGGA

[0221] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFD
SGETAETRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVE
EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYLAHMI
KFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLS
KSRRLENLIAQLPGEKKNGLFGNLIASLGLTPNFKSNFDLAEDAKLQLSKDTY
DDDLNLLAQIGDQYADFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYD
EHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILE
KMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPFLKD
NREKIEKILTRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGE
QKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLK
RRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKE
DIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDDELVKVMGRHKPENI
[0222] VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYL
YYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGK
SDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGELSELDKAGFIKR
QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFY
KVREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIAS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDE
ATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD PKKYGG
FDSPTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSEFKNPIDFLEAKGY
KEVKKDLIIKLPKYSLFEL ENGRKRMLASAGELQKGNELALPSKYVNFLYLASH
YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYN
KHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ
SITGLYETRIDLSQLGGD

[0223] (单底线:HNH域;双底线:RuvC域)。

[0224] 在一些实施方案中,野生型Cas9对应于化脓性链球菌的Cas9 (NCBI参考序列:NC_002737.2 (核苷酸序列如下);以及Uniprot参考序列:Q99ZW2 (氨基酸序列如下):

[0225] ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGATCACTGAT
GAATATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTATCAAAAAAATCTTATAGG
GGCTCTTTTATTTGACAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTC
GGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGA
CTTGAAGAGTCTTTTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAAGT
TGCTTATCATGAGAAATATCCAATCTATCATCTGCGAAAAAATTTGGTAGATTCTACTGATAAAGCGGATTTGC
GCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAGATTTAAATCCTGAT

AATAGTGATGTGGACAAACTATTTATCCAGTTGGTACAAACCTACAATCAATTATTTGAAGAAAACCCTATTAACGC
AAGTGGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGCTCAGC
TCCCCGGTGAGAAGAAAAATGGCTTATTTGGGAATCTCATTGCTTTGTCATTGGGTTTGACCCCTAATTTTAAATCA
AATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTAGATAATTTATTGGC
GCAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGATGCTATTTTACTTTTCAGATATCC
TAAGAGTAAATACTGAAATAACTAAGGCTCCCTATCAGCTTCAATGATTAAACGCTACGATGAACATCATCAAGAC
TTGACTCTTTTAAAGCTTTAGTTTCGACAACAACCTCCAGAAAAGTATAAAGAAATCTTTTTTGATCAATCAAAAAA
CGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAAATTTATCAAACCAATTTTAGAAAAA
TGGATGGTACTGAGGAATTATTGGTGAAACTAAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACAACGGC
TCTATTCCCCATCAAATTCATTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTAAA
AGACAATCGTGAGAAGATTGAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTCCATTGGCGCGTGGCAATA
GTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCATGGAATTTTGAAGAAGTTGTCGATAAAGGT
GCTTCAGCTCAATCATTTATTGAACGCATGACAACTTTGATAAAAAATCTTCAAATGAAAAAGTACTACCAAAACA
TAGTTTGCTTTATGAGTATTTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTGAAGGAATGCGAAAAC
CAGCATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAACAAATCGAAAAGTAACCGTTAAG
CAATTAAGAAGATTATTTCAAAAAATAGAATGTTTTGATAGTGTGAAATTTCAGGAGTTGAAGATAGATTTAA
TGCTTCATTAGGTACCTACCATGATTTGCTAAAAATTTATAAAGATAAAGATTTTTTGGATAATGAAGAAAATGAAG
ATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAAGATAGGGAGATGATTGAGGAAAGACTTAAACATAT
GCTCACCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTTTGTCTCGAAA
ATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTAGATTTTTTGAAATCAGATGGTTTTGCCAATC
GCAATTTTATGCAGCTGATCCATGATGATAGTTTGACATTTAAAGAAGACATTCAAAAAGCACAAGTGTCTGGACAA
GGCGATAGTTTACATGAACATATTGCAAATTTAGCTGGTAGCCCTGCTATTA AAAAAGGTATTTTACAGACTGTAAA
AGTTGTTGATGAATTGGTCAAAGTAATGGGGCGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACGTGAAAATC
AGACAACCTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAATCGAAGAAGGTATCAAAGAATTAGGAAGT
CAGATTCTTAAAGAGCATCCTGTTGAAAAATACTCAATTGCAAAATGAAAAGCTCTATCTCTATTATCTCCAAATGG
AAGAGACATGTATGTGGACCAAGAATTAGATATTAATCGTTTAAAGTATTATGATGTCGATCACATTGTTCCACAAA
GTTTCCTTAAAGACGATTCAATAGACAATAAGGTCTTAACGCGTTCTGATAAAAAATCGTGGTAAATCGGATAACGTT
CCAAGTGAAGAAGTAGTCAAAAAGATGAAAACTATTGGAGACAACCTCTAAACGCCAAGTTAATCACTCAACGTAA
GTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTGAACCTTGATAAAGCTGGTTTTATCAAACGCCAATTGG
TTGAAACTCGCCAAATCACTAAGCATGTGGCACAATTTTGGATAGTCGCATGAATACTAAATACGATGAAAATGAT
AACTTATTCGAGAGGTTAAAGTGATTACCTTAAATCTAAATTAGTTTCTGACTTCCGAAAAGATTTCCAATTCTA
TAAAGTACGTGAGATTAACAATTACCATCATGCCATGATGCGTATCTAAATGCCGTCGTTGGAAGTCTTTTGATTA
AGAAATATCCAAAACCTGAATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTTCGTAAAATGATTGCTAAG
TCTGAGCAAGAAATAGGCAAGCAACCGCAAAATATTTCTTTTACTCTAATATCATGAACTTCTTCAAACAGAAAT
TAACTTGCAAATGGAGAGATTTCGAAACGCCCTCTAATCGAACTAATGGGGAACTGGAGAAATTGTCTGGGATA
AAGGGCGAGATTTTGCCACAGTGCAGCAAGTATTGTCCATGCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAG
ACAGGCGGATTCTCAAGGAGTCAATTTTACCAAAAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGA
TCCAAAAAATATGGTGGTTTTGATAGTCCAACGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGA
AATCGAAGAAGTTAAATCCGTTAAAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAATCCG

ATTGACTTTTTAGAAAGCTAAAGGATATAAGGAAGTTAAAAAGACTTAATCATTAACCTACCTAAATATAGTCTTTT
TGAGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAATTACAAAAAGGAAATGAGCTGGCTCTGCCAA
GCAATATGTGAATTTTTTATATTTAGCTAGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAA
CAATTGTTTGTGGAGCAGCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCGTGTTAT
TTTAGCAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAGACAAACCAATACGTGAACAAGCAG
AAAATATTATTCATTTATTTACGTTGACGAATCTTGGAGCTCCCGCTGCTTTTAAATATTTTGATACAACAATTGAT
CGTAAACGATATACGTCTACAAAAGAAGTTTATAGTGCCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAAC
ACGCATTGATTTGAGTCAGCTAGGAGGTGACTGA

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFD
SGETAETRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSSFFHRLEESFLVE
EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADRLIYLALAHMI
KFRGHFLIEGDLNPDNSDVDKLFQILVQTYNQLFEENPINASGVDAKAILSARLS
KSRRLENLIAQLPGEKKNGLFGNLIASLGLTPNFKSNFDLAEDAKLQLSKDTY
DDDLNLLAQIGDQYADLFLAAKNLSAILLSILRVNTEITKAPLSASMIKRYD
EHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILE
KMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKD
[0226] NREKIEKILTRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGE
QKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLK
RRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKE
DIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYL
YYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGK
SDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR
QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFY
KVREINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIAS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF
ATVRKVLSPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGG
[0227] FDSPTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSEFKNPIDFLEAKGY
KEVKKDLIIKLPKYSLELENGRKRMLASAGELQKGNELALPSKYVNFYLYLASH
YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVL SAYN
KHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ
SITGLYETRIDLSQLGGD (底线: HNH 域; 双底线: RuvC 域)

[0228] 在一些实施方案中, Cas9是指来自溃疡棒状杆菌 (NCBI Refs: NC_015683.1, NC_017317.1); 白喉棒状杆菌 (NCBI Ref: NC_016782.1, NC_016786.1); 栖蝇蝇螺原体 (NCBI Ref: NC_021284.1); 中间型普氏杆菌 (NCBI Ref: NC_017861.1); 中国台湾螺旋体 (NCBI Ref: NC_021846.1); 海豚链球菌 (NCBI Ref: NC_021314.1); 波罗的海贝尔氏菌 (NCBI Ref: NC_018010.1); 扭曲冷弯曲菌 (NCBI Ref: NC_018721.1); 嗜热链球菌 (NCBI Ref: YP_820832.1), 无害李斯特菌 (NCBI Ref: NP_472073.1), 空肠弯曲杆菌 (NCBI Ref: YP_

002344900.1) 或脑膜炎双球菌 (NCBI Ref:YP_002342100.1) 的Cas9,或来自其他任何生物的Cas9。

[0229] 应了解,包括其变体和同源物在内的其他Cas9蛋白(例如,核酸酶死亡的Cas9(dCas9)、Cas9切口酶(nCas9)或核酸酶活性Cas9),包括其变体和同源物。示例性的Cas9蛋白包括但不限于以下提供的那些。在一些实施方案中,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。

[0230] 在一些实施方案中,Cas9结构域是无核酸酶的Cas9结构域(dCas9)。例如,dCas9结构域可以结合双链核酸分子(例如,通过gRNA分子),而不切割双链核酸分子的任一链。在一些实施方案中,无核酸酶的dCas9结构域包含本文列出的氨基酸序列的D10X突变和H840X突变,或本文提供的任何氨基酸序列中的相应突变,其中X是任何氨基酸改变。在一些实施方案中,无核酸酶的dCas9结构域包含本文列出的氨基酸序列的D10A突变和H840A突变,或本文提供的任何氨基酸序列中的相应突变。作为一个实例,无核酸酶的Cas9结构域包含克隆载体pPlatTET-gRNA2(登录号BAV54124)中列出的氨基酸序列。

[0231] 示例性的催化失活Cas9(dCas9)的氨基酸序列如下:

[0232] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRNRIICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEEINPINSAGVDAKAILSARLSKSRRL ENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAI LLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLP EKYKEIFFDQSKNGYAGYIDGGASQEEFYKF I K PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPYYVGP LARGNSRFAWMTRKSEETITPWNFEVVDKGASQSFIERMTNFDKNLPNEKVLPHSLLYEFYTVYNELTKVKYVT EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKKI IKDKDFLD NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGSLSHEHIANLAGSPAIIKKGILQTVKVVDLVKVMGRHKPENIVIE MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVD AIVPQSFLKDDSIDNKVLTRSDKNRGKSDNPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG EIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN ELALPSKYVNFYLLASHYEKLGKSPEDNEQKQLFVEQHKHYLDEIIIEQISEFSKRVILADANLDKVL SAYNKHDKP IREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDLSQLGGD

[0233] (例如,参见Qi等人,“Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression.”Cell.,2013;152(5):1173-83,其全部内容通过引用并入本文)。

[0234] 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,即,所述Cas9是切口酶,称为“nCas9”蛋白(作为“切口酶”Cas9)。核酸酶灭活的Cas9蛋白可以互换地称为“dCas9”蛋白(对于核酸酶“死亡”Cas9)或催化失活的Cas9。产生具有失活的DNA切割

结构域的Cas9蛋白(或其片段)的方法是已知的(例如,参见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,其每一个的全部内容通过引用并入本文。例如,已知Cas9的DNA切割结构域包括两个亚结构域,即HNH核酸酶亚结构域和RuvC1亚结构域。HNH子域切割与gRNA互补的链,而RuvC1子域切割非互补链。这些亚结构域内的突变可以使Cas9的核酸酶活性沉默。例如,突变D10A和H840A完全灭活了化脓性链球菌Cas9的核酸酶活性(Jinek等人,Science.337:816-821(2012);Qi等人,Cell.,28;152(5):1173-83(2013))。

[0235] 在一些实施方案中,dCas9结构域包含与本文提供的任何dCas9结构域至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相同的氨基酸序列。在一些实施方案中,与本文列出的任何氨基酸序列相比,Cas9结构域包含具有1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50个或更多个或更多突变。在一些实施方案中,与本文列出的任何氨基酸序列相比,Cas9结构域包含具有至少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、至少具有900、至少1000、至少1100或至少1200个相同的连续氨基酸残基。

[0236] 在一些实施方案中,dCas9对应于或部分或全部包含具有使Cas9核酸酶活性失活的一个或多个突变的Cas9氨基酸序列。例如,在一些实施方案中,dCas9结构域包含D10A和H840A突变或另一Cas9中的相应突变。

[0237] 在一些实施方案中,dCas9包含dCas9的氨基酸序列(D10A和H840A):

[0238] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFD

SGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVE
EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADRLIYLALAHMI
KFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLS
KSRRLLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKS NFDLAEDAKLQLSKDTY
DDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYD
EHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILE
KMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPFLKD
NREKIEKILTRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEVVDKGASAQ
SFIERMTNFDKNLPNEKVLPHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGE
QKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIERLKTYAHLFDDKVMKQLK
RRRYTGWGRLSRKLLINGIRDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKE
 [0239] DIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYL
YYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGK
SDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR
QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFY
KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF
ATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGG
FDSPTVAYSVLVVAKEKGKSKKLKSVKELLGITIMERS SFEKNPIDFLEAKGY
KEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH
YEKLKGSPEDNEQKQLFVEQHKKHYLDEIIEQISEF SKRVILADANLDKVL SAYN
KHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTTIDRKRYTSTKEVL DATLIHQ
SITGLYETRIDLSQLGGD (单底线: HNH 域; 双底线: RuvC 域)。

[0240] 在一些实施方案中, Cas9结构域包含D10A突变, 而位置840上的残基在以上提供的氨基酸序列中或在本文提供的任何氨基酸序列中的相应位置处仍为组氨酸。

[0241] 在其他实施方案中, 提供了具有除D10A和H840A以外的突变的dCas9变体, 其例如导致核酸酶灭活的Cas9 (dCas9)。举例来说, 此类突变包括在D10和H840处的其他氨基酸取代, 或在Cas9的核酸酶结构域内的其他取代(例如, 在HNH核酸酶子结构域和/或RuvC1子结构域中的取代)。在一些实施方案中, 提供了dCas9的变体或同源物, 其至少约70%相同、至少约80%相同、至少约90%相同、至少约95%相同、至少约98%相同、至少约99%相同、至少约99.5%相同或至少约99.9%相同。在一些实施方案中, 提供的dCas9变体具有以下氨基酸序列, 所述氨基酸序列较短或更长约5个氨基酸、约10个氨基酸、约15个氨基酸、约20个氨基酸、约25个氨基酸、约30个氨基酸、约40个氨基酸、约50个氨基酸、约75个氨基酸、约100个氨基酸或更多。

[0242] 在一些实施方案中, Cas9结构域是Cas9切口酶。Cas9切口酶可以是Cas9蛋白, 其能够仅切割双链核酸分子(例如, 双链DNA分子)的一条链。在一些实施方案中, Cas9切口酶切割双链核酸分子的靶链, 这意味着Cas9切口酶切割与结合至Cas9的gRNA(例如, sgRNA)碱基配对(互补)的链。在一些实施方案中, Cas9切口酶包括D10A突变并且在840位具有组氨酸。

在一些实施方案中,Cas9切口酶切割双链核酸分子的非靶标、非碱基编辑的链,这意味着Cas9切口酶切割未与结合至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相应的突变。在一些实施方案中,Cas9切口酶包含的氨基酸序列至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%与本文提供的任何Cas9切口酶相同。基于本申请和本领域的知识,其他合适的Cas9切口酶对于本领域技术人员是显而易见的,并且在本申请的范围内。

[0243] 示例性的催化Cas9切口酶(nCas9)的氨基酸序列如下:MDKKYSIGLAIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLNLIQPLGEEKNGFLGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLEPKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKI IKDKDFLDNEENEDI EDIVLTLTLFEDREMIEERLKT YAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAICKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVFREINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGSKKLKSVKELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLGKSPEDNEQKQLFVEQHKHYLDEIIIEQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDLSQLGGD

[0244] 在一些实施方案中,Cas9指来自古细菌(例如,纳米古细菌)的Cas9,其构成单细胞原核微生物的域和界。在一些实施方案中,可编程核苷酸结合蛋白可以是CasX或CasY蛋白,其已经描述于例如Burstein等人,“New CRISPR-Cas systems from uncultivated microbes.”,Cell Res,2017年2月21日,doi:10.1038/cr.2017.21,其全部内容通过引用合并于本文。使用基因组分辨的宏基因组学(metagenomic),鉴定了许多CRISPR-Cas系统,包括生命古细菌域中首次报导的Cas9。这种差异化的Cas9蛋白是在鲜为研究的nanoarchaea中发现的,是主动CRISPR-Cas系统的一部分。在细菌中,发现了两个以前未知的系统,即CRISPR-CasX和CRISPR-CasY,它们是迄今为止发现的最紧凑的系统。在一些实施例,在本文所述的碱基编辑器系统中,Cas9被CasX或CasX的变体替代。在一些实施例,在本文描述的碱基编辑器系统中,Cas9被CasY或CasY的变体替代。应当理解,其他RNA引导的DNA结合蛋白可以用作核酸可编程DNA结合蛋白(napDNAbp),并且在本申请的范围内。

[0245] 在一些实施方案中,本文提供的任何融合蛋白的核酸可编程DNA结合蛋白(napDNAbp)可以是CasX或CasY蛋白。在一些实施方案中,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包含的氨基酸序列至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%与天然存在的CasX或CasY蛋白相同。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,可编程核苷酸结合蛋白包含与本文所述的任何CasX或CasY蛋白具有至少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%的同一性的氨基酸序列。应当理解,根据本申请,也可以使用来自其他细菌物种的CasX和CasY。

[0246] 示例性的CasX((uniprot.org/uniprot/F0NN87;uniprot.org/uniprot/F0NH53)tr|F0NN87|F0NN87_SULIHCRISPR相关的Casx蛋白OS=冰岛硫化叶菌(菌株HVE10/4)GN=SiH_0402PE=4SV=1)氨基酸序列如下:

[0247] MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVNLAYKIAKNNEDAAAERRGKAKKKKG
EEGETTTSNIIPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKECEEVSAPSFVKPEFYEFGRSPGMVERT
RRVKLEVEPHYLIIAAGWVLTRLGKAKVSEG DYVGVNVFTPTRGILYSLIQNVNGIVPGIKPETAFLWIARKVV
SSVTNPVSVVRIYITISDAVGQNPTTINGGFSIDLTKLLEKRYLLSERLEAIARNALSISSNMRERYIVLANIYE
YLTG SKRLEDLLYFANRDLIMNLSDDGKVRDLKLISAYVNGELIRGEG。

[0248] 示例性的CasX(>tr|F0NH53|F0NH53_SULIR CRISPR相关蛋白,Casx OS=冰岛硫化叶菌(菌株REY15A)GN=SiRe_0771 PE=4 SV=1)氨基酸序列如下:

[0249] MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVNLAYKIAKNNEDAAAERRGKAKKKKG
EEGETTTSNIIPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKECEEVSAPSFVKPEFYKFGRSPGMVERT
RRVKLEVEPHYLIIMAAAGWVLTRLGKAKVSEG DYVGVNVFTPTRGILYSLIQNVNGIVPGIKPETAFLWIARKVV
SSVTNPVSVVSIYITISDAVGQNPTTINGGFSIDLTKLLEKRDLLSERLEAIARNALSISSNMRERYIVLANIYE
YLTGSKRLEDLLYFANRDLIMNLSDDGKVRDLKLISAYVNGELIRGEG。

[0250] δ -变形菌CasX

[0251] MEKRINKIRKKLSADNATKPVSRSGPMKTLVVRVMTDDLKKRLEKRRKKPEVMPQVISNNAANNLRML
LDDYTKMKEAILQVYWQEFKDDHVGLMCKFAQPASKKIDQNKLPKPEMDEKGNLT TAGFACSQCQGPLFVYKLEQVS
EKGKAYTNYFGRCNVAEHEKLILLAQLKPKVKSDEAVTYSLGKFGQRALDFYSIHVTKESTHPVKPLAQIAGNRYA
SGPVGKALS DACMGTIASFLSKYQDIIIEHQVVKGNQKRLESLRELAKENLEYPSTLPPQPHTKEGVDfAYNE
VIARVRMWNVNLNLWQKLKLSRDDAKPLRLKGFPSFPVVERRENEVDWWNTINEVKKLIDAKRDMGRVFWSGVTAEK
RNTILEGYNYLPNENDHKKREGSLENPKKPAKRQFGDLLLYLEKKYAGDWGKVFDEAWERIDKKIAGLTSHIEREEA
RNAEDAQSKAVLTDWLRKASFVLERLKEMDEKEFYACEIQLQKWYGDLRGNPFAVEAENRVVDISGFSIGSDGHSI
QYRNLLAWKYLENGKREFYLLMNYGKKGRIRFTDGTDIKSGKWQGLLYGGGKAKVIDLTFDPDDEQLIILPLAFGT
RQGREFIWNDLLSLETGLIKLANGRVIEKTIYNKKIGRDEPALFVALTFERREVDPSPNIKPVNLIGVARGENIPAV
IALTDPEGCPLPEFKDSSGGPTDILRIGEGYKEQRAIQAAKEVEQRRAGGYSRKFASKSRNLADDMVRNSARDLFY
HAVTHDAVLVFANLSRGFGRQGKRTFMTERQYTKMEDWLTAKLAYEGLTSKTYLSKTLAQYTSKTC SNCGFTITYAD
MDVMLVRLKKTSDGWATTLNNKELKAEQITYYNNRYKRQTVEKELSAELDRLSEESGNNDISKWTKGRDEALFLK
KRFSHRPVQE QFVCLDCGHEVHAAEQ AALNIARSWLFLNSNSTEFKSYKSGKQPFVGAWQAFYKRRLKEVWKPNA

[0252] 示例性的CasY((ncbi.nlm.nih.gov/蛋白质/APG80656.1)>APG80656.1 CRISPR相关蛋白CasY[未培养的俭菌总门群细菌])氨基酸序列如下:

[0253] MSKRHPRISGVKGYRLHAQRLEYTGKSGAMRTIKYPLYSSPSGGRTVPREIVSAINDDYVGLYGLSNF
DDLNAEKRNEEKVYSVLDWFYDCVQYGAVFSYAPGLLNVAEVRGGSYELTKTLKGSHLYDELQIDKVIKFLNK
KEISRANGSLDKLKKDIIDCFKAEYRERHKDQCNKLADDIKNAKKDAGASLGERQKKLFRDFFGISEQSENDKPSF
TNPLNLTCCLLPFDTVNNNRNRGEVLFNKLKEYAQKLDKNEGSLEMWEYIGIGNSGTAFSNFLGEGFLGRLRENKI
TELKKAMMDITDAWRGQEQEEELEKRLRILAALTIKLREPKFDNHWGGRSDINGKLSSWLQNYINQTVKIKEDLK
GHKKDLKKAKEMINRFGESDTKEEAVVSSLLESIEKIVPDDSDADEKPDIPAIAIYRRFLSDGRLTLNRFVQREDV
QEALIKERLEAEKKKKPKRKKKSDAEDEKETIDFKELFPHLAKPLKLVNPFYGDSKRELYKKYNAAIYTDALWK
AVEKIYKSAFSSSLKNSFFDITDFDKDFFIKRLQKIFSVYRRFNTDKWKPIVKNSFAPYCDIVSLAENEVLYKPKQS
RSRKSAAIDKNRVRPLSTENIAKAGIALARELSVAGFDWKDLLKKEEHEEYIDLIELHKTALALLAVTETQLDIS
ALDFVENGTVKDFMKTRDGNLVLEGRFLEMFSQSIVFSELRLAGLMSRKEFITRSAIQTMNGKQAELLYIPHEFQ
SAKITTPKEMSRAFLDLAPAEFATSLEPESLSEKSLKLKQMRYYPHYFGYELTRTGQIDGGVAENALRLEKSPV
KKREIKCKQYKTLGRGQNKIVLYVRSSYYQTQFLEWFLHRPKNVQTDVAVSGSFLIDEKKVKTRWNYDALTVALEP
VSGSERVFVSQPFTIFPEKSAEEEGQRYLGIDIGEYGIAYTALEITGDSAKILDQNFISDPQLKTLREEVKGLKLD
QRRGTFAMPSTKIARIRESLVHSLRNRHHLALKHKAKIVYELEVS RFEEGKQKIKKVYATLKKADVSEIDADKN
LQTTVWGKLAVASEISASYTSQFCGACKKLWRAEMQVDETITTTQELIGTVRVIKGGTLIDAIKDFMRPPIFDENDT
PFPKYRDFCDKHHISKMRGNSCLFICPFCRANADADIQASQTIALLRVYVKEEKKVEDYFERFRKLKNIKVLGQMK
KI。

[0254] 在一些实施方案中,核酸可编程DNA结合蛋白(napDNAbp)是微生物CRISPR-Cas系统的单个效应子。微生物CRISPR-Cas系统的单个效应子包括但不限于Cas9、Cpf1、Cas12b/C2c1和Cas12c/C2c3。通常,微生物CRISPR-Cas系统分为1类和2类系统。1类系统具有多亚基效应子复合物,而2类系统具有单个蛋白质效应子。例如,Cas9和Cpf1是2类效应器。除Cas9和Cpf1外,Shmakov等人,“Discovery and Functional Characterization of Diverse Class 2CRISPR Cas Systems”,Mol.Cell,2015年11月5日;60(3):385-397还描述了三个不同的2类CRISPR-Cas系统,其全部内容通过引用合并于此。两个系统的效应子Cas12b/C2c1和Cas12c/C2c3包含与Cpf1相关的RuvC样核酸内切酶结构域。第三系统包含具有两个谓词HEPN RNase结构域的效应子。与Cas12b/C2c1产生的CRISPR RNA不同,成熟的CRISPR RNA的产生不依赖tracrRNA。Cas12b/C2c1取决于CRISPR RNA和tracrRNA的DNA切割。

[0255] 已经报导了酸环脂双歧杆菌Cas12b/C2c1(AacC2c1)的晶体结构与嵌合单分子指导RNA(sgRNA)复合。例如,参见Liu等人,“C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism”,Mol.Cell,2017年1月19日;65(2):310-322,其全部内容通过引用合并于此。还已经报导了酸热脂环酸杆菌C2c1以三元复合物结合到靶DNA上的晶体结构。例如,参见Yang等人,“PAM-dependent Target DNA Recognition and Cleavage by C2C1CRISPR-Cas endonuclease”,Cell,2016年12月15日;167(7):1814-1828,其全部内容通过引用结合于此。具有目标和非目标DNA链的AacC2c1具有催化能力的构象已被独立捕获,放置在单个RuvC催化口袋中,Cas12b/C2c1介导的切割导致目标DNA错开七个核苷酸。Cas12b/C2c1三元复合物与先前鉴定的Cas9和Cpf1对应物之间的结构比较证明了CRISPR-Cas9系统使用的机制的多样性。

[0256] 在一些实施方案中,本文提供的任何融合蛋白的核酸可编程DNA结合蛋白(napDNAbp)可以是Cas12b/C2c1或Cas12c/C2c3蛋白。在一些实施方案中,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%相同的氨基酸序列。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包含与本文提供的任何napDNAbp序列至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%的氨基酸序列相同。应当理解,根据本申请,也可以使用来自其他细菌物种的Cas12b/C2c1或Cas12c/C2c3。

[0257] 一个Cas12b/C2c1((uniprot.org/uniprot/T0D7A2#2) sp|T0D7A2|C2C1_ALIAG CRISPR相关的核酸内切酶C2c1 OS=嗜酸酸脂环芽孢杆菌(菌株ATCC 49025/DSM 3922/CIP 106132/NCIMB 13)GN=c2c1 PE=1 SV=1)氨基酸序列如下:

[0258] MAVKSIKVKLRLLDDMPEIRAGLWKLHKEVNAGVRYYTEWLSLLRQENLYRRSPNGDGEQECDKTAEEC
KAELLERLRARQVENGHRGPAGSDELLQLARQLYELLVPAIGAKGDAQIARKFLSPLADKDAVGGLGIKAGN
KPRWVRMREAGEPGWEEKEKAETRKSADRTADVLRALADFGKPLMRVYTDSEMSSVEWKPLRKGQAVRTWDRDM
FQQAIERMMSWESWNQRVGQYAKLVEQKNRFEQKNFVGQEHLVHLVNQLQQDMKEASPGLESKEQTAHYVTGRAL
RGSDKVFEEKWGLAPDAPFDLYDAEIKNVQRRNTRRFGSHDLFAKLAEPEYQALWREDASFLTRYAVYNSILRKLN
HAKMFATFTLPDATAHPIWTRFDKLGGLNHQYTFLEFNEFGERRHAIRFHKLLKVENGVAEVDVTVPIISMSEQLD
NLLPRDPNEPIALYFRDYGAEQHFTGEFGGAKIQCRRDQLAHMHRRRGARDVYLVNSVRVQSQSEARGERRPPYAA
VFRLVGDNHRAVFVHFDKLSDYLAHPDDGKLGSEGLSGLRVMSVDLGLRTSASISVFRVARKDELKPNKGRVPF
FFPIKGNLNLVAVHERSQLLKLPGETESKDLRAIREERQRTLRLTLQLAYLRLLVRCGSEDVGRRERSWAKLIEQ
PVDAANHMTDPDWREAFENELQKLKSLHGICSDKEWMDAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYAKDV
VGGNSIEQIEYLERQYKFLKSWFFGKVSQVIRAEKGSRFATLREHIDHAKEDRLKKLADRIIMEALGYVYALD
ERGGKQWVAKYPPCQLILLEELSEYQFNDRPPSENNQLMQWSHRGVFQELINQAQVHDLVGTMYAAFSSRFDA
TGAPGIRCRRVPARCTQEHNPFPWWLNKFVVEHTLDACPLRADDLIPTGEGEIFVSPFSAEEGDFHQIHADLNA
AQNLQQLWSDFDISQIRLRCDWGEVDGELVLIPRLTGKRTADSYSNKVFYTNVGTYYERERGGKRRKVFQAEKL
SEEEAELLVEADEAREKSVVLMRDPGIIINRGWTRQKEFWSMV NQRIEGLYLVKQIRSRVPLQDSACENTGDI。

[0259] BhCas12b(组氨酸芽孢杆菌)NCBI参考序列:WP_095142515MAPKKRKGVIHGVPAAAT
RSFILKIEPNNEEVKKGLWKTHEVLNHGIAYYMNLKLRQEAIEYEHHEQDPKNPKKVSKEIQAEWDFVLKMQKCN
SFTHEVDKDEVFNILRELYEELVPSSVEKKGEANQLSNKFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEE
EKKKWEEDKKKDLAKILGKLAEYGLIPLFIPYTDSENPVKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWESWN
LKVKEEYKVEKEYKTLERIKEDIQALKALEQYEKERQEQLLRDTLNTNEYRLSKRGLRGWREIIQKWLKMDENEP
SEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHF IWRNHPEYPYLYATFCEIDKKKKDAKQATFTLADPINHPLW
VRFEERSGSNLNKYRILTEQLHTEKLKKKLTQQLDRLIYPTESGGWEEKGKVDIVLLPSRQFYNQIFLDIEEKGKHA
FTYKDESIKFLKGTGGARVQFDRDHLRRYPHKVESGNVGRIFYNMTVNIPTESPVSKSLKIHRDDFPKVVNFKP
KELTEWIKDSKGKKLKSGIESLEIGLRVMSIDLGGQRAAAASIFEVDQKPDIEGKLFPIKGTLEYAVHRASFNIK
LPGETLVKSREVLARKAREDNLKLNMQKLNFLRNVLHFQQFEDITEREKRVTKWISRQENSVDPLVYQDELIQIRELM
YKPYKDWVAFKQLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDEIDRTRKFLLRWSLRPTEPGEVRRLEPG
QRFAIDQLNHLNALKEDRLKKMANTIIMHALGYCYDVRKKKWKAKNPACQIILFEDLSNYPYEERSRFENSKLMKW
SRREIPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKLQDNRFKLNQREGRLTLDKIAVLKEGD

LYPDKGGEKFISLSKDRKCVTTHADINAAQNLQKRFWTRTHGFIYKYCKAYQVDGQTVYIPESKDQKQKIIIEFGEG
YFILKDGVEWVNAGKLKIKKGSSKQSSSELVSDILKDSFDLASELKGEKMLYRDPSPGNVFPSPDKWMAAGVFFGK
LERILISKLTNQYSISTIEDSSKQSMKRPAATKKAGQAKKKK

[0260] 在一些实施方案中,Cas12b是BvCas12B,其是BhCas12b的变体,并且相对于BhCas12B包含以下变化:S893R,K846R和E837G。BvCas12b(芽孢杆菌V3-13)NCBI参考序列:WP_101661451.1MAIRSIKLMKTNSTDSIYLRKALWRTHQLINEGIAYYMNLLTYRQEAIGDKTKEAYQAE
LINIIRNQQRNNGSSEEHGSDQEILALLRQLYELIIPSSIGESGDANQLGNKFLYPLVDPNSQSGKGTSNAGRKPRW
KRLKEEGNPDWLEKKKDEERKAKDPTVKIFDNLNKYGLLPLFPLFTNIQKDIEWLPLGKRQSVRKWDKDMFIQAIE
RLLSWESWNRRADEYKQLKEKTESYYKEHLTGGEWIEKIRKFEKERNMELEKNAFAPNDGYFITSRQIRGWDRVY
EKWSKLPESASPEELWKVVAEQNKMSSEFGDPKVFSLANRENNDIWRGHSEIRYHIAAYNGLQKKLSRTKEQATF
TLPDAIEHPLWIRYESPGGTNLNLFKLEEKQKKNYYVTLSKIIWPSEEKWKIEKENIEIPLAPSIQFNRIKQKQHVK
GKQEISFSDYSSRISLDGVLGGSRIQFNRYIKNHKELLGEGDIGPVFFNLVVDVAPLQETRNGRLQSPIGKALKVI
SSDFSVIDYKPKELMDWMNTGSASNSFGVASLLEGMVRMSIDMGQRTSASVSIFEVVKELPKDQEQLFYSINDTE
LFAIHKRSFLLNLPGEVVTKNNKQQRQERRKKRQFVRSQIRMLANVLRLETKKTPDERKKAHKLMEIVQSYDSWTA
SQKEVWEKELNLLTNMAAFNDEIWKESLVELHHRIEPYVGQIVSKWRKGLSEGRKNLAGISMWNIDELEDTRLLIS
WSKRSRTPGEANRIETDEPFGSSLLQHIQNVKDDRLKQMANLIIMTALGFKYDKEEKDRYKRWKETYPACQIILFEN
LNRYLFNLDRSRRENSRLMKWAHRSIPRTVSMQGMFGLQVGDRSEYSSRFHAKTGAPGIRCHALTEEDLKAGSNT
LKRLIEDGFINESELAYLKKGDIIPSQGGELFVTLISKRYKKDSNNELTVIHADINAAQNLQKRFWQQNSEVYRVPC
QLARMGEDKLYIPKSTETIKKYFGKGSFVKNNTEQEVYKWEKSEKMKIKTDTTDFDLQDLGDFEDISKTIELAQQQ
KKYLTMFRDPSGYFFNNETWRPQKEYWSIVNNIIKSLCKKKILSNKVEL

[0261] Cas9核酸酶具有两个功能性核酸内切酶结构域:RuvC和HNH。Cas9在靶标结合后经历第二个构象变化,所述变化使核酸酶结构域定位,以切割靶标DNA的相对链。Cas9介导的DNA切割的最终结果是靶DNA(PAM序列上游约3-4个核苷酸)内的双链断裂(DSB)。然后,通过两种一般的修复途径之一修复所得的DSB:(1)有效但容易出错的非同源末端连接(NHEJ)途径;或(2)效率较低但高保真同源引导修复(HDR)途径。

[0262] 非同源末端连接(NHEJ)和/或同源引导修复(HDR)的“效率”可以通过任何方便的方法来计算。例如,在某些情况下,效率可以用成功HDR的百分比表示。例如,可以使用测量员核酸酶测定法(surveyor nuclease assay)来产生切割产物,并且可以用产物与受质的比例来计算百分比。例如,作为成功的HDR的结果,可以使用直接切割含有新整合的限制性序列的DNA的测量员核酸酶。切割的受质越多,表明HDR百分比越高(HDR效率越高)。作为说明性实例,可以使用以下等式[(切割产物)/(受质加上切割产物)]计算HDR的分数(百分比)(例如,(b+c)/(a+b+c),其中“a”是DNA受质的条带强度,“b”和“c”是切割产物。

[0263] 在某些情况下,效率可以用NHEJ成功率来表示。例如,T7核酸内切酶I测定法可用于产生切割产物,产物与受质的比例可用于计算NHEJ的百分比。T7核酸内切酶可切割由野生型和突变型DNA链杂交引起的错配异源双链DNA(NHEJ在原始断裂位点产生小的随机插入或缺失(Indels))。切割程度越高,表示NHEJ的百分比越高(NHEJ的效率越高)。作为说明性示例,可以使用以下等式计算NHEJ的分数(百分比):(1-(1-(b(c+c)/(a+b+c))^{1/2}))×100,其中“a”是DNA受质的条带强度,“b”和“c”是切割产物(Ran等人,Ce11,2013年9月12日;154(6):1380-9;Ran等人,Nat Protoc.,2013年11月;8(11):2281-2308)。

[0264] NHEJ修复途径是最活跃的修复机制,它经常在DSB位点引起小的核苷酸插入或缺失(Indels)。NHEJ介导的DSB修复的随机性具有重要的实际意义,因为表达Cas9和gRNA或指导多核苷酸的细胞群会导致各种各样的突变。在大多数情况下,NHEJ在靶标DNA中产生小的插入,导致氨基酸缺失、插入或移码突变,从而导致靶标基因的开放阅读框(ORF)内的提前终止密码子。理想的最终结果是靶基因内的功能丧失突变。

[0265] 虽然NHEJ介导的DSB修复经常破坏基因的开放阅读框,但是同源引导修复(HDR)可用于产生特定的核苷酸变化,范围从单核苷酸变化到大插入,如添加荧光团或标签。

[0266] 为了利用HDR进行基因编辑,可以将含有所需序列的DNA修复模板与gRNA和Cas9或Cas9切口酶一起传递到感兴趣的细胞类型中。修复模板可以包含期望的编辑以及紧接在靶标的上游和下游的其他同源序列(称为左和右同源臂)。每个同源臂的长度可以取决于引入的改变的大小,其中较大的插入需要更长的同源臂。修复模板可以是单链寡核苷酸、双链寡核苷酸或双链DNA质粒。即使在表达Cas9,gRNA和外源性修复模板的细胞中,HDR的效率通常也较低(<修饰等位基因的10%)。由于HDR发生在单元周期的S和G2阶段,因此可以通过同步单元来提高HDR的效率。NHEJ中涉及的化学或遗传抑制基因也可以增加HDR频率。

[0267] 在一些实施方案中,Cas9是修饰的Cas9。给定的gRNA靶向序列在整个基因组中可能存在部分同源的其他位点。这些位点称为脱靶位点,在设计gRNA时需要加以考虑。除了优化gRNA设计外,还可以通过修饰Cas9来提高CRISPR特异性。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切口酶是SpCas9的D10A突变体,保留一个核酸酶结构域并产生DNA切口而不是DSB。切口酶系统也可以与HDR介导的基因编辑结合起来进行特定的基因编辑。

[0268] 在某些情况下,Cas9是Cas9蛋白的变体。当与野生型Cas9蛋白的氨基酸序列相比时,变体Cas9多肽具有的氨基酸序列相差一个氨基酸(例如,具有缺失、插入、取代、融合)。在一些情况下,变体Cas9多肽具有降低Cas9多肽的核酸酶活性的氨基酸变化(例如,缺失、插入或取代)。例如,在某些情况下,变体Cas9多肽具有小于50%、小于40%、小于30%、小于20%、小于10%、小于5%或小于1%的核酸酶。相应的野生型Cas9蛋白的活性。在某些情况下,变体Cas9蛋白没有实质的核酸酶活性。当主题Cas9蛋白是没有实质核酸酶活性的变体Cas9蛋白时,其可以被称为“dCas9”。

[0269] 在一些情况下,变体Cas9蛋白具有降低的核酸酶活性。例如,变体Cas9蛋白展现出核酸内切酶活性的小于约20%、小于约15%、小于约10%、小于约5%、小于约1%或小于约0.1%。野生型Cas9蛋白,例如野生型Cas9蛋白。

[0270] 在一些情况下,变体Cas9蛋白可以切割指导靶序列的互补链,但是具有降低的切割双链指导靶序列的非互补链的能力。例如,变体Cas9蛋白可具有降低RuvC结构域功能的突变(氨基酸取代)。作为非限制性实例,在一些实施方案中,变体Cas9蛋白具有D10A(在氨基酸位置10处为谷氨酸的天冬氨酸),因此可以切割双链导向靶序列的互补链,但切割双链导向靶序列的非互补链的能力降低(因此,当变体Cas9蛋白切割双链靶核酸时,导致单链断裂(SSB)而不是双链断裂(DSB))(例如,参见Jinek等人,Science.2012年八月17日;337(6096):816-21)。

[0271] 在一些情况下,变体Cas9蛋白可以切割双链指导靶序列的非互补链,但是具有降低的切割指导靶序列的互补链的能力。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(氨基酸取代)。作为非限制性实例,在一些实施方案中,变体Cas9蛋白具有H840A(在氨基酸位置840处的组氨酸至丙氨酸)突变,并因此可以切割指导靶序列的非互补链,但是具有降低的切割靶序列的能力。指导目标序列的互补链(因此,当变体Cas9蛋白切割双链指导目标序列时,将产生SSB而不是DSB)。这样的Cas9蛋白切割指导靶序列(例如单链指导靶序列)的能力降低,但是保留了结合指导靶序列(例如单链指导靶序列)的能力。

[0272] 在一些情况下,变体Cas9蛋白切割双链靶DNA的互补链和非互补链的能力降低。作为非限制性实例,在一些情况下,变体Cas9蛋白既具有D10A突变又具有H840A突变,使得所述多肽具有降低的切割双链靶DNA的互补链和非互补链的能力。这样的Cas9蛋白切割靶DNA(例如单链靶DNA)的能力降低,但是保留了结合靶DNA(例如单链靶DNA)的能力。

[0273] 作为另一个非限制性实例,在一些情况下,变体Cas9蛋白具有W476A和W1126A突变,使得多肽具有降低的切割靶DNA的能力。这样的Cas9蛋白切割靶DNA(例如单链靶DNA)的能力降低,但是保留了结合靶DNA(例如单链靶DNA)的能力。

[0274] 作为另一个非限制性实例,在一些情况下,变体Cas9蛋白具有P475A、W476A、N477A、D1125A、W1126A和D1127A突变,使得所述多肽具有降低的切割靶DNA的能力。这样的Cas9蛋白切割靶DNA(例如单链靶DNA)的能力降低,但是保留了结合靶DNA(例如单链靶DNA)的能力。

[0275] 作为另一个非限制性实例,在一些情况下,变体Cas9蛋白具有H840A、W476A和W1126A突变,使得多肽具有降低的切割靶DNA的能力。这样的Cas9蛋白切割靶DNA(例如单链靶DNA)的能力降低,但是保留了结合靶DNA(例如单链靶DNA)的能力。作为另一个非限制性实例,在某些情况下,变体Cas9蛋白带有H840A、D10A、W476A和W1126A突变,使得所述多肽切割靶DNA的能力降低。这样的Cas9蛋白切割靶DNA(例如单链靶DNA)的能力降低,但是保留了结合靶DNA(例如单链靶DNA)的能力。在一些实施方案中,变体Cas9已经在Cas9 HNH结构域(A840H)中的位置840处恢复了催化的His残基。

[0276] 作为另一个非限制性实例,在一些情况下,变体Cas9蛋白带有H840A、P475A、W476A、N477A、D1125A、W1126A和D1127A突变,使得所述多肽切割靶DNA的能力降低。这样的Cas9蛋白切割靶DNA(例如单链靶DNA)的能力降低,但是保留了结合靶DNA(例如单链靶DNA)的能力。作为另一个非限制性实例,在一些情况下,变体Cas9蛋白具有D10A、H840A、P475A、W476A、N477A、D1125A、W1126A和D1127A突变,使得所述多肽具有降低的切割靶DNA的能力。这样的Cas9蛋白切割靶DNA(例如单链靶DNA)的能力降低,但是保留了结合靶DNA(例如单链靶DNA)的能力。在某些情况下,当变异的Cas9蛋白带有W476A和W1126A突变,或者变异的Cas9蛋白带有P475A、W476A、N477A、D1125A、W1126A和D1127A突变时,变异的Cas9蛋白不能有效地与PAM序列结合。因此,在一些这样的情况下,当在结合方法中使用这种变体Cas9蛋白时,所述方法不需要PAM序列。换句话说,在某些情况下,当将这种Cas9变异蛋白用于结合方法时,所述方法可包含指导RNA,但所述方法可在不存在PAM序列的情况下进行(因此,结合的特异性由指导RNA的靶向片段提供)。可以使其他残基突变以获得上述效果(即,使一个或另一个核酸酶部分失活)。作为非限制性实例,残基D10、G12、G17、E762、H840、N854、N863、H982、H983、A984、D986和/或A987可以被改变(即被取代)。同样,除丙氨酸取代以外的突变也是合适的。

[0277] 在一些实施方案中,具有降低的催化活性的变体Cas9蛋白(例如,当Cas9蛋白具有D10、G12、G17、E762、H840、N854、N863、H982、H983、A984、D986和/或如果是A987突变,例如D10A、G12A、G17A、E762A、H840A、N854A、N863A、H982A、H983A、A984A和/或D986A),则Cas9变体蛋白仍可以以位点特异性方式与靶DNA结合(因为只要它保留与指导RNA相互作用的能力,它仍然可以被指导RNA引导至目标DNA序列)。

[0278] 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。

[0279] 化脓性链球菌Cas9的替代品可以包括Cpf1家族的RNA指导的内切核酸酶,在哺乳动物细胞中显示出切割活性。来自普氏杆菌(*Prevotella*)和弗朗西斯菌1(*Francisella* 1)(CRISPR/Cpf1)的CRISPR是一种类似于CRISPR/Cas9系统的DNA编辑技术。Cpf1是II类CRISPR/Cas系统的RNA指导的核酸内切酶。在普氏杆菌和弗朗西斯菌中发现了这种获得性免疫机制。Cpf1基因与CRISPR基因座相关联,编码一种内切核酸酶,所述核酸内切酶使用指导RNA来发现和切割病毒DNA。Cpf1是一种比Cas9小且简单的核酸内切酶,克服了CRISPR/Cas9系统的某些局限性。与Cas9核酸酶不同,Cpf1介导的DNA切割的结果是带有短3'突出端的双链断裂。Cpf1的交错切割模式可以打开方向性基因转移的可能性,类似于传统的限制性内切酶克隆,可以提高基因编辑的效率。像上述的Cas9变体和直向同源物一样,Cpf1还可以将CRISPR靶向的AT位富集区域或AT基因组缺少SpCas9支持的NGG PAM位点的位点数目扩展。Cpf1基因座包含一个混合的 α/β 域、一后跟一个螺旋区域、一个RuvC-II和一个锌指状域的RuvC-I。Cpf1蛋白具有类似于Cas9的RuvC结构域的RuvC样核酸内切酶结构域。此外,Cpf1不具有HNH核酸内切酶结构域,并且Cpf1的N端不具有Cas9的 α -螺旋识别叶。Cpf1 CRISPR-Cas域结构显示Cpf1在功能上是独特的,被归类为2类,V型CRISPR系统。Cpf1基因座编码的Cas1、Cas2和Cas4蛋白与I型和III型相比更类似于II型系统。功能性Cpf1不需要反式激活CRISPR RNA(tracrRNA),因此只需要CRISPR(crRNA)。这有利于基因组编辑,因为Cpf1不仅比Cas9小,而且具有更小的sgRNA分子(约为Cas9的一半)。与Cas9靶向的富含G的PAM相比,Cpf1-crRNA复合物可通过识别与5'-YTN-3'相邻的原间隔子来切割目标DNA或RNA。鉴定出PAM后,Cpf1引入了一个4或5个核苷酸突出端的粘性末端样DNA双链断裂。

[0280] 本申请内容的一些方面提供了融合蛋白,其包含充当核酸可编程DNA结合蛋白的结构域,其可用于将蛋白质(例如碱基编辑器)引导至特定核酸(例如DNA或RNA)序列。在特定实施方案中,融合蛋白包含核酸可编程DNA结合蛋白结构域和脱氨酶结构域。DNA结合蛋白包括但不限于Cas9(例如dCas9和nCas9)、Cas12a/Cpf1、Cas12b/C2c1、Cas12c/C2c3、Cas12d/CasY、Cas12e/CasX、Cas12g、Cas12h和Cas12i。具有不同于Cas9的PAM特异性的可编程多核苷酸结合蛋白的一个例子是成簇的来自普氏杆菌和弗朗西斯菌1(Cpf1)的规则间隔的短回文重复序列。与Cas9相似,Cpf1也是2类CRISPR效应子。已经显示Cpf1介导具有不同于Cas9的特征的强大的DNA干扰。Cpf1是一种缺少tracrRNA的单链RNA指导的内切核酸酶,其利用了一个富含T的邻近原间隔子的基序(TTN,TTN或YTN)。此外,Cpf1通过交错的DNA双链断裂切割DNA。在16种Cpf1家族蛋白中,来自酸性氨基球菌和毛螺菌的两种酶显示在人细胞中具有有效的基因组编辑活性。Cpf1蛋白是本领域已知的,并且先前已经进行了描述,例如Yamano等人,“Crystal structure of Cpf1 in complex with guide RNA and target DNA.”*Cell*. (165)2016,p.949-962;其全部内容通过引用结合于此。

[0281] 在本发明的组合物和方法中还有用的是无核酸酶的Cpf1 (dCpf1) 变体,其可用作指导核苷酸序列可编程的多核苷酸结合蛋白结构域。Cpf1蛋白具有类似于Cas9的RuvC结构域的RuvC样核酸内切酶结构域,但不具有HNH核酸内切酶结构域,并且Cpf1的N端不具有Cas9的 α -螺旋识别叶。在Zetsche等人,Cell.,163,759-771,2015(通过引用并入本文)中显示,Cpf1的RuvC-样结构域负责切割DNA链和RuvC-样结构域的失活。使Cpf1核酸酶活性失活。例如,Novicida novicida Cpf1中对应于D917A、E1006A或D1255A的突变会失活Cpf1核酸酶活性。在一些实施方案中,本申请的dCpf1包含对应于D917A、E1006A、D1255A、D917A/E1006A、D917A/D1255A、E1006A/D1255A或D917A/E1006A/D1255A的突变。应当理解,根据本发明,可以使用任何突变,例如使Cpf1的RuvC结构域失活的取代突变、缺失或插入。在一些实施方案中,本文提供的任何融合蛋白的核酸可编程核苷酸结合蛋白可以是Cpf1蛋白。在一些实施方案中,Cpf1蛋白是Cpf1切口酶(nCpf1)。在一些实施方案中,Cpf1蛋白是核酸酶失活的Cpf1(dCpf1)。在一些实施方案中,Cpf1、nCpf1或dCpf1包含与本文公开的Cpf1序列具有至少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%同一性氨基酸序列。在一些实施方案中,dCpf1包含与本文公开的Cpf1序列具有至少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%相同性的氨基酸序列,并且包含对应于D917A、E1006A、D1255A、D917A/E1006A、D917A/D1255A、E1006A/D1255A或D917A/E1006A/D1255A。应当理解,根据本申请,也可以使用来自其他细菌物种的Cpf1。

[0282] 野生型新弗朗西斯菌(Francisella novicida)Cpf1的氨基酸序列如下。D917、E1006和D1255用粗体加底线。

[0283]

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIID
KYHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSSDDNLQKDFKSAKDTIKKQIS
EYIKDSEKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEA
LEIHSFKGWTTYFKGFHENRKNVYSSNDIPTSIHYRIVDDNLPKFLENKAKYESL
KDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQS
GITKFNTIIGGKFFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSD
TESKSFVIDKLEDDSDVVTMMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKL
DLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELI
AKKTEKAKYLSLETIKLALAEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNK
DNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNLLHKLKIFHISQSED
KANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTL
ANGWDKNKEPDNTAILFIKDDKYLLGVMNKKNNKIFDDKAIKENKGEQYKKI
VYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEEF
NIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFE
NISESYIDSVVNQGKLYLFQIYNKDFSAYS~~K~~GRPNLHTLYWKALFDERNLQDV
VYKLNGEAE~~L~~FYRKQSIPKKITHPAKEAIA~~N~~KNKDNPKKESVFEYDLIKDKRFT
EDKFFFHC~~P~~ITINFKSSGANKFNDEINLLLKEKANDVHILSIDRGERHLAYYTLV
DGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKE
GYLSQVVHEIAKL~~V~~IEYNAIVVFEDLNFGFKRGRFKVEKQVYQKLEKMLIEKLN
YLVFKDNEFDKTTGGVLRAYQLTAPFETFKKM~~G~~KQTGIIYYVPAGFTSKICPVTG
FVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTI
ASFGSRLINFRNSDKNHNWDTREVVPTKELEKLLKDYSIEYGHGECIKAAICGE
SDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQ
DADANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN.

[0284] 以下为新弗朗西斯菌Cpf1 D917A的氨基酸序列。(A917、E1006和D1255用粗体和底线标出)。

[0285]

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIID
KYHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSSDDNLQKDFKSAKDTIKKQIS
EYIKDSEKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEA
LEIHSFKGWTTYFKGFHENRKNVYSSNDIPTSIHYRIVDDNLPKFLENKAKYESL
KDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQS
GITKFNTIIGGKFFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSD
TESKSFVIDKLEDDSDVVTMMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKL
DLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELI
AKKTEKAKYLSLETIKLALAEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNK
DNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNLLHKLKIFHISQSED
KANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTL
ANGWDKNKEPDNTAILFIKDDKYLLGVMNKKNNKIFDDKAIKENKGEQYKKI
VYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEEF
NIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFE

[0286] NISESYIDSVVNQGKLYLFQIYNKDFSAYS SKGRPNLHTLYWKALFDERNLQDV
VYKLNGEAE L F Y R K Q S I P K K I T H P A K E A I A N K N K D N P K K E S V F E Y D L I K D K R F T
E D K F F F H C P I T I N F K S S G A N K F N D E I N L L L K E K A N D V H I L S I A R G E R H L A Y Y T L V
D G K G N I I K Q D T F N I I G N D R M K T N Y H D K L A A I E K D R D S A R K D W K K I N N I K E M K E
G Y L S Q V V H E I A K L V I E Y N A I V V F E D L N F G F K R G R F K V E K Q V Y Q K L E K M L I E K L N
Y L V F K D N E F D K T G G V L R A Y Q L T A P F E T F K K M G K Q T G I I Y Y V P A G F T S K I C P V T G
F V N Q L Y P K Y E S V S K S Q E F F S K F D K I C Y N L D K G Y F E F S F D Y K N F G D K A A K G K W T I
A S F G S R L I N F R N S D K N H N W D T R E V Y P T K E L E K L L K D Y S I E Y G H G E C I K A A I C G E
S D K K F F A K L T S V L N T I L Q M R N S K T G T E L D Y L I S P V A D V N G N F F D S R Q A P K N M P Q
D A D A N G A Y H I G L K G L M L L G R I K N N Q E G K K L N L V I K N E E Y F E F V Q N R N N .

[0287] 以下为新弗朗西斯菌Cpf1 E1006A的氨基酸序列。(D917, A1006和D1255用粗体和
底线标出)。

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIID
KYHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQIS
EYIKDSEKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEA
LEIISFKGWTTYFKGFHENRKNVYSSNDIPTSIYRIVDDNLPKFLENKAKYESL
KDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQS
GITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSD
TESKSFVIDKLEDDSDVVTMMSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKL
DLSKIYFKNDKSLTDLSSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELI
AKKTEKAKYLSLETIKLAL EEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNK
DNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNLLHKLKIFHISQSED
KANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTL
ANGWDKNKEPDNTAILFIKDDKYLLGVMNKKNNKIFDDKAIKENKGEYKKI
[0288] VYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEF
NIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFE
NISESYIDSVVNQGKLYLFQIYNKDFSAYS SKGRPNLHTLYWKALFDERNLQDV
VYKLNGEAE L F Y R K Q S I P K K I T H P A K E A I A N K N K D N P K K E S V F E Y D L I K D K R F T
E D K F F F H C P I T I N F K S S G A N K F N D E I N L L L K E K A N D V H I L S I D R G E R H L A Y Y T L V
D G K G N I I K Q D T F N I I G N D R M K T N Y H D K L A A I E K D R D S A R K D W K K I N N I K E M K E
G Y L S Q V V H E I A K L V I E Y N A I V V F A D L N F G F K R G R F K V E K Q V Y Q K L E K M L I E K L
N Y L V F K D N E F D K T G G V L R A Y Q L T A P F E T F K K M G K Q T G I I Y Y V P A G F T S K I C P V T
G F V N Q L Y P K Y E S V S K S Q E F F S K F D K I C Y N L D K G Y F E F S F D Y K N F G D K A A K G K W
T I A S F G S R L I N F R N S D K N H N W D T R E V Y P T K E L E K L L K D Y S I E Y G H G E C I K A A I C G
E S D K K F F A K L T S V L N T I L Q M R N S K T G T E L D Y L I S P V A D V N G N F F D S R Q A P K N M P
Q D A D A N G A Y H I G L K G L M L L G R I K N N Q E G K K L N L V I K N E E Y F E F V Q N R N N .

[0289] 以下为新弗朗西斯菌Cpf1 D1255A的氨基酸序列。(D917、E1006和A1255突变位置
以粗体显示并带有底线)。

[0290] MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIID

KYHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQIS
EYIKDSEKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEA
LEIHSFKGWTTYFKGFHENRKNVYSSNDIPTSHYRIVDDNLPKFLENKAKYESL
KDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQS
GITKFNTIIGGKFFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSD
TESKSFVIDKLEDDSDVVTMMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKL
DLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELI
AKKTEKAKYLSLETIKLALAEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNK
DNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNLLHKLKIFHISQSED
KANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTL
ANGWDKNKEPDNTAILFIKDDKYLLGVMNKKNNKIFDDKAIKENKGEGYKKI
[0291] VYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKF
NIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFE
NISESYIDSVVNQGKLYLFQIYNKDFSAYS~~K~~GRPNLHTLYWKALFDERNLQDV
VYKLNGEAE~~L~~FYRKQSIPKKITHPAKEAIANKNKNKDNPKKESVF~~E~~YDLIKDKRFT
EDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIDRGERHLAYYTLV
DGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKE
GYLSQVVHEIAKL~~V~~IEYNAIVVFEDLNF~~G~~FKRGRFKVEKQVYQKLEKMLIEKLN
YLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTG
FVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTI
ASFGSRLINFRNSDKNHNWDTRE~~V~~YPTKELEKLLKDYSIEYGHGECIKAAICGE
SDKKFFAKLTSVLNTILQMRNSKTGT~~E~~LDYLISP~~V~~ADVNGNFFDSRQAPKNMPQ
DAANGAYHIGLKGLMLLGRIKNNQEGK~~L~~NLVIKNEEYFEFVQNRNN

[0292] 新弗朗西斯菌Cpf1 D917A/E1006A的氨基酸序列如下。(A917、A1006和D1255用粗体和底线标出)。

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIID
KYHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQIS
EYIKDSEKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEA
LEIHSFKGWTTYFKGFHENRKNVYSSNDIPTSHYRIVDDNLPKFLENKAKYESL
KDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQS
GITKFNTIIGGKFFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSD
TESKSFVIDKLEDDSDVVTMMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKL
[0293] DLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELI
AKKTEKAKYLSLETIKLALAEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNK
DNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNLLHKLKIFHISQSED
KANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTL
ANGWDKNKEPDNTAILFIKDDKYLLGVMNKKNNKIFDDKAIKENKGEGYKKI
VYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKF
NIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFE
NISESYIDSVVNQGKLYLFQIYNKDFSAYS~~K~~GRPNLHTLYWKALFDERNLQDV

VYKLNGEAE~~L~~FYRKQSIPKKITHPAKEAIA~~N~~KNKDNPKKESVFEYDLIKDKRFT
EDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIARGERHLAYYTLV
DGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKE
GYLSQVVHEIAKL~~V~~IEYNAIVVFADLNFGFKRGRFKVEKQVYQKLEKMLIEKL
[0294] NYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVT
GFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKW
TIASFGSRLINFRNSDKNHNWDTREVVYPTKELEKLLKDYSIEYGHGECIKAAICG
ESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMP
QDADANGAYHIGLKGLMMLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN.

[0295] 以下为新弗朗西斯菌Cpf1 D917A/D1255A的氨基酸序列。(A917、E1006和A1255用粗体和底线标出)。

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIID
KYHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQIS
EYIKDSEKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEA
LEIHSFKGWTTYFKGFHENRKNVYSSNDIPTSIYRIVDDNLPKFLENKAKYESL
KDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQS
GITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSD
TESKSFVIDKLEDDSDVVTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKL
DLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELI
AKKTEKAKYLSLETIKLAL~~E~~EFNKH~~R~~DIDKQCRFEEILANFAAIPMIFDEIAQNK
DNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNLLHKLKIFHISQSED
KANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTL
ANGWDKNKEPDNTAILFIKDDKY~~L~~GVMMNKKNNKIFDDKAIKENKGE~~G~~YKKI
[0296] VYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEF
NIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFE
NISESYIDSVVNQGKLYLFQIYNKDFSAYS~~K~~GRPNLHTLYWKALFDERNLQDV
VYKLNGEAE~~L~~FYRKQSIPKKITHPAKEAIA~~N~~KNKDNPKKESVFEYDLIKDKRFT
EDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIARGERHLAYYTLV
DGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKE
GYLSQVVHEIAKL~~V~~IEYNAIVVFEDLNFGFKRGRFKVEKQVYQKLEKMLIEKL
YLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTG
FVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTI
ASFGSRLINFRNSDKNHNWDTREVVYPTKELEKLLKDYSIEYGHGECIKAAICGE
SDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQ
DAAANGAYHIGLKGLMMLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN.

[0297] 新弗朗西斯菌Cpf1 E1006A/D1255A的氨基酸序列如下。(D917、A1006和A1255用粗体和底线标出)。

[0298] MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIID
KYHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQIS

EYIKDSEKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEA
LEIHSFKGWTTYFKGFHENRKNVYSSNDIPTSIHYRIVDDNLPKFLENKAKYESL
KDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNYYLNQS
GITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSD
TESKSFVIDKLEDDSDVVTMMSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKL
DLSKIYFKNDKSLTDLSSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELI
AKKTEKAKYLSLETIKLALAEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNK
DNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNLLHKLKIFHISQSED
KANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTL
ANGWDKNKEPDNTAILFIKDDKYLLGVMNKKNNKIFDDKAIKENKGEYKKI
[0299] VYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEF
NIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFE
NISESYIDSVVNQGKLYLFQIYNKDFSAYS~~K~~GRPNLHTLYWKALFDERNLQDV
VYKLNGEAE~~L~~FYRKQSIPKKITHPAKEAIANKNDNP~~K~~KESVFEYDLIKDKRFT
EDKFFFHCPITINFKSSGANKFNDEINLLLEKANDVHILSIDRGERHLAYYTLV
DGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKE
GYLSQVVHEIAKL~~V~~IEYNAIVVFADLNF~~G~~FKRGRFKVEKQVYQKLEKMLIEKL
NYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVT
GFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKW
TIASFGSRLINFRNSDKNHNWDTREVVYPTKELEKLLKDYSIEYGHGECIKAAICG
ESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMP
QDAANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN.

[0300] 以下为新弗朗西斯菌Cpf1 D917A/E1006A/D1255A的氨基酸序列。(A917、A1006和A1255用粗体和底线标出)。

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIID
KYHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSSDDNLLQKDFKSAKDTIKKQIS
EYIKDSEKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEA
LEIHSFKGWTTYFKGFHENRKNVYSSNDIPTSIHYRIVDDNLPKFLENKAKYESL
KDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNYYLNQS
GITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSD
TESKSFVIDKLEDDSDVVTMMSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKL
[0301] DLSKIYFKNDKSLTDLSSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELI
AKKTEKAKYLSLETIKLALAEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNK
DNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNLLHKLKIFHISQSED
KANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTL
ANGWDKNKEPDNTAILFIKDDKYLLGVMNKKNNKIFDDKAIKENKGEYKKI
VYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEF
NIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFE
NISESYIDSVVNQGKLYLFQIYNKDFSAYS~~K~~GRPNLHTLYWKALFDERNLQDV
VYKLNGEAE~~L~~FYRKQSIPKKITHPAKEAIANKNDNP~~K~~KESVFEYDLIKDKRFT

[0302]

EDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIARGERHLAYYTLV
DGKGNIHKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKE
GYLSQVVHEIAKLVEYNAIVVFADLNFGFKRGRFKVEKQVYQKLEKMLIEKL
NYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVT
GFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKW
TIASFGSRLINFRNSDKNHNWDTREVVPTKELEKLLKDYSIEYGHGECIKAAICG
ESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMP
QDAAANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN.

[0303] 在一些实施方案中,融合蛋白中存在的Cas9结构域之一可以被不需要PAM序列的指导核苷酸序列可编程DNA结合蛋白结构域替代。

[0304] 在一些实施方案中,Cas结构域是来自金黄色葡萄球菌(SaCas9)的Cas9结构域。在一些实施方案中,SaCas9结构域是核酸酶活性SaCas9、核酸酶非活性SaCas9(SaCas9d)或SaCas9切口酶(SaCas9n)。在一些实施方案中,SaCas9结构域在本文提供的任何氨基酸序列中包含N579A突变或相应的突变。

[0305] 在一些实施方案中,SaCas9结构域、SaCas9d结构域或SaCas9n结构域可以结合具有非典型PAM的核酸序列。在一些实施方案中,SaCas9结构域、SaCas9d结构域或SaCas9n结构域可以结合具有NGRRT或NGRRT PAM序列的核酸序列。在一些实施方案中,SaCas9结构域在本文提供的任何氨基酸序列中包含E781X、N967X和R1014X突变或相应突变中的一个或多个,其中X是任何氨基酸。在一些实施方案中,SaCas9结构域包含本文提供的任何氨基酸序列中的一个或多个E781K、N967K和R1014H突变,或一个或多个相应突变。在一些实施方案中,SaCas9结构域在本文提供的任何氨基酸序列中包含E781K、N967K或R1014H突变或相应的突变。

[0306] 示例性的SaCas9的氨基酸序列如下:

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRL
KRRRRHRIQRVKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALL
HLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGE
VRGSINRFKTSDYVKEAKQLLKVKQAYHQLDQSFIDTYIDLLETRRTYYEGPGE
GSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITR
DENEKLEYEYKQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFT
NLKVYHDIKDITARKEIENAELLDQIAKILTIYQSSEDIQEELTNLNSLTQEEIE
QISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVKKVDLSQQKEI
PTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINE

[0307]

[0308] MQQRNRQTNERIEEIIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLN
NPFNYEVDHIIPRSVSFDNSFNKVLVKQEENSKKGNRTPFQYLSSSDSKISYET
FKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLM
NLLRSYFRVNNLDVKVKSINGGFTSFLRRKWFKKERNKGYKHAEDALIIAN
ADFIFKEWKKLDKAKKVMENQMFEKQAESMPEIETEQEYKEIFITPHQIKHIK
DFKDYKYSHRVDKKNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKL
KKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYEEETGNYLTKYS
KKDNGPVIKKIKYYGNKLNAHLDTDDYPNSRNKVVKL~~SLK~~PYRFDVYLDNGV
YKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKING
ELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIKTIASKTQSIKKYST
DILGNLYEVKSKKHPQIIKKG。

[0309] 在所述序列中,带底线和粗体的残基N579可以被突变(例如,突变为A579)以产生SaCas9切口酶。

[0310] 示例性的SaCas9n的氨基酸序列如下:

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRL
KRRRRHRIQRVKLLFDYNLLTDHSELGINPYEARVKGLSQKLSEEEFSAALL
HLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGE
VRGSINRFKTS~~DY~~VKEAKQLLKVKAYHQLDQSFIDTYIDLLETRRTYYEGPGE
GSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITR
DENEKLEYEYEFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFT
NLKVYHDIKDITARKEIENAELLDQIAKILTIYQSSEDIQEELTNLNSLTQEEIE
QISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEI
PTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIELAREKNSKDAQKMINE
MQQRNRQTNERIEEIIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLN
[0311] NPFNYEVDHIIPRSVSFDNSFNKVLVKQEEASKKGNRTPFQYLSSSDSKISYET
FKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLM
NLLRSYFRVNNLDVKVKSINGGFTSFLRRKWFKKERNKGYKHAEDALIIAN
ADFIFKEWKKLDKAKKVMENQMFEKQAESMPEIETEQEYKEIFITPHQIKHIK
DFKDYKYSHRVDKKNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKL
KKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYEEETGNYLTKYS
KKDNGPVIKKIKYYGNKLNAHLDTDDYPNSRNKVVKL~~SLK~~PYRFDVYLDNGV
YKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKING
ELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIKTIASKTQSIKKYST
DILGNLYEVKSKKHPQIIKKG。

[0312] 在所述序列中,可以从N579突变以产生SaCas9切口酶的残基A579用底线和粗体标出。

[0313] 示例性的SaKKH Cas9的氨基酸序列如下:

[0314] KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRL
KRRRRHRIQRVKLLFDYNLLTDHSELGINPYEARVKGLSQKLSEEEFSAALL

HLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGE
 VRGSINRFKTSYVKEAKQLLKVKQAYHQLDQSFIDTYIDLLETRRTYYEGPGE
 GSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITR
 DENEKLEYEYKFKQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFT
 NLKVYHDIKDITARKEIENAELLDQIAKILTIYQSSEDIQEELTNLNSLTQEEIE
 QISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEI
 PTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINE
 MQKRNRQTNERIEEIIHRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLN
 NPFNYEVDHIIPRSVSFDNSFNKVLVKQEEASKKGNRTPFQYLSSSDSKISYET
 [0315] FKKHILNLAGKGRISKTKKEYLLEERDINRFSVQKDFINRLVDTRYATRGLM
 NLLRSYFRVNNLDVKVKSINGGFTSFLRRKWFKKERNKGYKHAEDALIIAN
 ADFIFKEWKKLDKAKKVMENQMFEKQAESMPEIETEQEYKEIFITPHQIKHIK
 DFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKL
 KKLINKSPEKLLMYHHPQTYQKLKLIMEQYGDEKNPLYKYEEETGNYLTKYS
 KKDNGPVIKKIKYYGNKLNAHLDDITDDYPNSRNKVVKLKSLKPYRFDVYLDNGV
 YKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYKNDLIKING
 ELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPHIIKTIASKTQSIKKYST
 DILGNLYEVKSKKHPQIIKKG。

[0316] 上面的残基A579用底线和粗体标出,可以从N579突变以产生SaCas9切口酶。可以从E781、N967和R1014突变以产生SaKKH Cas9的上述残基K781、K967和H1014用底线和斜体标出。

[0317] 高保真Cas9域

[0318] 本申请的一些方面提供了高保真度的Cas9结构域。在一些实施方案中,高保真Cas9结构域是经工程改造的Cas9结构域,其包含相对于相应的野生型Cas9结构域降低Cas9结构域与DNA的糖-磷酸骨架之间的静电相互作用的一个或多个突变。与DNA的糖-磷酸主链的静电相互作用降低的高保真Cas9结构域可具有较少的脱靶效应。在一些实施方案中,Cas9结构域(例如,野生型Cas9结构域)包含一种或多种降低Cas9结构域与DNA的糖-磷酸主链之间的缔合的突变。在一些实施方案中,Cas9结构域包含一种或多种突变,其使Cas9结构域与DNA的糖-磷酸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%。

[0319] 在一些实施方案中,本文提供的任何Cas9融合蛋白在本文提供的任何氨基酸序列中包含N497X、R661X、Q695X和/或Q926X突变或相应突变中的一个或多个,其中X是任何氨基酸。在一些实施方案中,本文提供的任何Cas9融合蛋白在本文提供的任何氨基酸序列中包含N497A、R661A、Q695A和/或Q926A突变或相应突变中的一个或多个。在一些实施方案中,Cas9结构域在本文提供的任何氨基酸序列中包含D10A突变或相应的突变。具有高保真度的Cas9结构域是本领域已知的,并且对本领域技术人员而言是显而易见的。例如,Kleinstiver,B.P.等人,“High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects.”,Nature 529,490-495(2016);and Slaymaker,I.M.等人,“Rationally engineered Cas9 nucleases with improved specificity.”,

Science 351,84-88(2015);每个的全部内容通过引用并入本文。

[0320] 在一些实施方案中,修饰的Cas9是高保真Cas9酶。在一些实施方案中,高保真Cas9酶是SpCas9(K855A)、eSpCas9(1.1)、SpCas9-HF1或超精确Cas9变体(HypaCas9)。修饰后的Cas9 eSpCas9(1.1)包含丙氨酸取代,可削弱HNH/RuvC凹槽与非目标DNA链之间的相互作用,从而防止链分离和在脱靶位点切割。类似地,SpCas9-HF1通过丙氨酸取代降低了脱靶编辑,丙氨酸取代破坏了Cas9与DNA磷酸骨架之间的相互作用。HypaCas9在REC3域中包含突变(SpCas9N692A/M694A/Q695A/H698A),可增加Cas9校对和目标识别力。与野生型Cas9相比,所有三种高保真酶产生的脱靶编辑均更少。

[0321] 下面提供示例性的高保真Cas9。

[0322] 相对于Cas9的高保真Cas9域突变以粗体和底线显示

MDKKYSIGL**A**IGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFD
SGETAETRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVE
EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADRLIYLALAHMI
KFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLS
KSRRLENLIAQLPGEKKNGLFGNLIASLGLTPNFKSNFDLAEDAKLQLSKDTY
DDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYD
EHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILE
KMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKD
NREKIEKILTRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
SFIERMT**A**FDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGE
QKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
LKIHKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTIAHLFDDKVMKQLK
RRRYTGWG**A**LSRKLINGIRDKQSGKTILDFLKSDGFANRNF**M**LIHDDSLTFKE
DIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYL
YYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGK
SDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR
QLVETRA**I**ITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFY
KVREINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF
ATVRKVLSPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGG
FDSPTVAYSVLVVAKEKGKSKKLKSVKELLGITIMERSSEFKNPIDFLEAKGY
KEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH
YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVL SAYN
KHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ
SITGLYETRIDLSQLGGD

[0324] 指导多核苷酸

[0325] 如本文所用,术语“(一个或多个)指导多核苷酸”是指可以对靶序列特异并且可以与多核苷酸可编程核苷酸结合域蛋白(例如,Cas9或Cpf1)形成复合物的多核苷酸。在一个实施方案中,指导多核苷酸是指导RNA。如本文所用,术语“指导RNA(gRNA)”及其语法等同物

可以指可以对靶DNA具有特异性并且可以与Cas蛋白形成复合物的RNA。RNA/Cas复合物可以帮助将Cas蛋白“引导”至目标DNA。Cas9/crRNA/tracrRNA核酸内切切割与间隔物互补的线性或环状dsDNA靶标。不与crRNA互补的靶链首先在核酸内切,然后在核酸外切3'-5'。实际上,DNA结合和切割通常需要蛋白质和两个RNA。但是,可以对单链指导RNA(“sgRNA”或简称为“gRNA”)进行改造,以便将crRNA和tracrRNA的各个方面都整合到单个RNA物种中。例如,参见Jinek M.等人,Science 337:816-821(2012),其全部内容通过引用并入本文。Cas9识别CRISPR重复序列中的短基序(PAM或原间隔子相邻基序),以帮助区分自身和非自身。Cas9核酸酶序列和结构是本领域技术人员众所周知的(参见例如“Complete genome sequence of an M1 strain of Streptococcus pyogenes.”Ferretti, JJ等人,Natl. Acad. Sci. USA 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);以及Jinek M.等人“Programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.”,Science 337:816-821(2012),其每一个的全部内容通过引用并入本文。已经在各种物种中描述了Cas9直向同源物,包括但不限于化脓性链球菌和嗜热链球菌。基于所述公开,其他合适的Cas9核酸酶和序列对于本领域技术人员而言是显而易见的,并且此类Cas9核酸酶和序列包括来自Chylinski, Rhun和Charpentier,“The tracrRNA and Cas9 families of type IICRISPR-Cas immunity systems”(2013) RNA Biology 10:5,726-737;其全部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,失活的)DNA切割结构域,即,Cas9是切口酶。

[0326] 在一些实施方案中,指导多核苷酸是至少一个单一指导RNA(“sgRNA”或“gRNA”)。在一些实施方案中,指导多核苷酸是至少一种tracrRNA。在一些实施方案中,指导多核苷酸不需要PAM序列来将多核苷酸可编程DNA结合结构域(例如,Cas9或Cpf1)引导至靶核苷酸序列。

[0327] 本文公开的碱基编辑器的多核苷酸可编程核苷酸结合结构域(例如,CRISPR衍生的结构域)可以通过与指导多核苷酸结合而识别靶多核苷酸序列。指导多核苷酸(例如,gRNA)通常是单链的,并且可以被编程为与多核苷酸的靶序列位点特异性结合(即,通过互补碱基配对),从而指导与指导结合的碱基编辑器靶序列的核酸。指导多核苷酸可以是DNA。指导多核苷酸可以是RNA。在某些情况下,指导多核苷酸包含天然核苷酸(例如,腺苷)。在一些情况下,指导多核苷酸包含非天然(或非天然)核苷酸(例如,肽核酸或核苷酸类似物)。在一些情况下,指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个核苷酸在长度上。指导核酸的靶向区域的长度可以在10至30个核苷酸之间,或在15至25个核苷酸之间,或在15至20个核苷酸之间。

[0328] 在一些实施方案中,指导多核苷酸包含两个或更多个单独的多核苷酸,其可以通过例如互补碱基配对(例如双链指导多核苷酸)彼此相互作用。例如,指导多核苷酸可以包含CRISPR RNA(crRNA)和反式激活CRISPR RNA(tracrRNA)。例如,指导多核苷酸可以包含一种或多种反式激活CRISPR RNA(tracrRNA)。

[0329] 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序列的序列的第一RNA分子(crRNA)和第二RNA分子(trRNA)之间进行互补碱基配对,所述序列识别靶序列和包含重复序列的第二个RNA分子(trRNA),所述重复序列形成稳定指导

RNA-CRISPR蛋白复合物的支架区域。这样的双链指导RNA系统可以用作指导多核苷酸,以将本文公开的碱基编辑器引导至靶多核苷酸序列。

[0330] 在一些实施方案中,本文提供的碱基编辑器利用单链指导多核苷酸(例如,gRNA)。在一些实施方案中,本文提供的碱基编辑器利用双链指导多核苷酸(例如双链gRNA)。在一些实施方案中,本文提供的碱基编辑器利用一种或多种指导多核苷酸(例如,多种gRNA)。在一些实施方案中,单链指导多核苷酸用于本文所述的不同碱基编辑器。例如,单链指导多核苷酸可用于胞苷碱基编辑器和腺苷碱基编辑器。

[0331] 在其他实施方案中,指导多核苷酸可以在单个分子(即,单分子指导核酸)中既包含核酸的多核苷酸靶向部分又包含核酸的支架部分。例如,单分子指导多核苷酸可以是单链指导RNA(sgRNA或gRNA)。在本文中,术语“指导多核苷酸序列”涵盖能够与靶标多核苷酸序列相互作用并将其引导至靶标多核苷酸序列的任何单分子、双分子或多分子核酸。

[0332] 通常,指导多核苷酸(例如crRNA/trRNA复合物或gRNA)包含“多核苷酸靶向片段”和“蛋白质结合片段”,所述片段包含能够识别并结合靶多核苷酸序列的序列。在碱基编辑器的多核苷酸可编程核苷酸结合结构域组分内稳定指导多核苷酸。在一些实施方案中,指导多核苷酸的多核苷酸靶向区段识别并结合至DNA多核苷酸,从而促进DNA中碱基的编辑。在其他情况下,指导多核苷酸的多核苷酸靶向片段识别并结合至RNA多核苷酸,从而促进RNA中碱基的编辑。在本文中,“片段”是指分子的部分或区域,例如在指导多核苷酸中核苷酸的连续延伸;片段也可以指复合物的区域/部分,使得片段可以包含大于或等于1的区域。例如,在指导多核苷酸包含多个核酸分子的情况下,其蛋白质结合区段可以包括例如沿着互补性区域杂交的多个单独分子的全部或一部分。包含两个单独分子的靶向DNA的RNA的结合片段可包含(i)第一个RNA分子的40至75个碱基对,长度为100个碱基对;和(ii)第二个RNA分子的10至25个碱基对长度为50个碱基对。“段”的定义,除非在特定上下文中另外明确定义,不限于特定数量的总碱基对,也不限于来自以下任何特定数量的碱基对。对于给定的RNA分子,不限于复合物中特定数量的单独分子,并且可以包括具有任何总长度的RNA分子区域,并且可以包括与其他分子互补的区域。

[0333] 指导RNA或指导多核苷酸可包含两个或更多个RNA,例如,CRISPR RNA(crRNA)和反式激活crRNA(tracrRNA)。指导RNA或指导多核苷酸有时可以包含单链RNA,或通过crRNA和tracrRNA的一部分(例如功能部分)融合形成的单指导RNA(sgRNA)。指导RNA或指导多核苷酸也可以是包含crRNA和tracrRNA的双链RNA。此外,crRNA可以与靶DNA杂交。

[0334] 如上所述,指导RNA或指导多核苷酸可以是表达产物。例如,编码指导RNA的DNA可以是包含编码指导RNA的序列的载体。可以通过用分离的指导RNA或包含编码指导RNA和启动子的序列的质粒DNA转染细胞来将指导RNA或指导多核苷酸转移到细胞中。指导RNA或指导多核苷酸也可以以其他方式转移到细胞中,例如使用病毒介导的基因递送。

[0335] 可以分离出指导RNA或指导多核苷酸。例如,指导RNA可以以分离的RNA的形式转染到细胞或生物体中。可以使用本领域已知的任何体外转录系统通过体外转录制备指导RNA。指导RNA可以以分离的RNA的形式而不是包含指导RNA的编码序列的质粒的形式转移至细胞。

[0336] 指导RNA或指导多核苷酸可包含三个区域:在5'端的第一区域,其可与染色体序列中的靶位点互补;第二内部区域,其可形成茎环结构;和可以是单链的第三个3'区域。每个

指导RNA的第一区域也可以是不同的,使得每个指导RNA将融合蛋白引导至特定的靶位点。此外,每个指导RNA的第二和第三区域在所有指导RNA中可以相同。

[0337] 指导RNA或指导多核苷酸的第一区域可以与染色体序列中靶位点的序列互补,使得指导RNA的第一区域可以与靶位点碱基配对。在一些情况下,指导RNA的第一区域可包含约10个核苷酸至25个核苷酸或约10个核苷酸至25个核苷酸(即,10个核苷酸至核苷酸;或约10个核苷酸至约25个核苷酸;或10个核苷酸至约25个核苷酸;或10个核苷酸至约25个核苷酸;或大约10个核苷酸至25个核苷酸)或更多。例如,在指导RNA的第一区域和染色体序列中的靶位点之间的碱基配对区域可以是或可以是大约10、11、12、13、14、15、16、17、18、19、20、22、23、24、25个或更多个核苷酸长度。有时,指导RNA的第一区域的长度可以是或可以是大约19、20或21个核苷酸。

[0338] 指导RNA或指导多核苷酸也可以包含形成二级结构的第二区域。例如,由指导RNA形成的二级结构可包含茎(或发夹)和环。环和茎的长度可以变化。例如,环的长度可以在约3至10个核苷酸的范围内或约3至10个核苷酸,并且茎的长度可以在约6至20个碱基对的范围或约6至20个碱基对。茎可包含一个或多个1至10个或大约10个核苷酸的突出。第二区域的总长度的长度可以在约16至60个核苷酸的范围内或16至60个核苷酸。例如,环的长度可以是或可以是约4个核苷酸,茎的长度可以是或可以是约12个碱基对。

[0339] 指导RNA或指导多核苷酸还可以在3'端包含基本上可以是单链的第三区域。例如,第三区域有时与目的细胞中的任何染色体序列都不互补,并且有时与其余的指导RNA不互补。此外,第三区域的长度可以变化。第三区域的长度可以大于或大于约4个核苷酸。例如,第三区域的长度可以在约5至60个核苷酸的长度范围内。

[0340] 指导RNA或指导多核苷酸可以靶向基因靶标的任何外显子或内含子。在某些情况下,指导可以靶向基因的第1外显子或第2外显子;指导可以靶向基因的外显子3或4。组合物可以包含全部靶向相同外显子的多个指导RNA,或者在某些情况下,可以包含可以靶向不同外显子的多个指导RNA。基因的外显子和内含子可以被靶向。

[0341] 指导RNA或指导多核苷酸可以靶向20个核苷酸或约20个核苷酸的核酸序列。靶核酸可以小于或小于约20个核苷酸。靶核酸的长度可以是至少或至少约5、10、15、16、17、18、19、20、21、22、23、24、25、30,或长度在1至100个核苷酸之间。靶核酸可以是至多或至多约5、10、15、16、17、18、19、20、21、22、23、24、25、30、40、50或1至100之间的任意值的核苷酸长度。靶核酸序列可以是或可以是PAM第一个核苷酸5'的20个碱基。指导RNA可以靶向核酸序列。靶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。

[0342] 指导多核苷酸,例如指导RNA,可以指可以与另一种核酸杂交的核酸,例如细胞基因组中的靶核酸或原间隔子。指导多核苷酸可以是RNA。指导多核苷酸可以是DNA。指导多核苷酸可以被编程或设计成与位点特异性地结合核酸序列。指导多核苷酸可以包含多核苷酸链,并且可以被称为单链指导多核苷酸。指导多核苷酸可以包含两条多核苷酸链,并且可以被称为双链指导多核苷酸。指导RNA可以作为RNA分子引入细胞中。例如,RNA分子可以在体外转录和/或可以化学合成。可以从合成DNA分子(例如gBlocks®基因片段)转录RNA。然后将指导RNA作为RNA分子引入细胞中。指导RNA也可以以非RNA核酸分子例如DNA分子的形式引入细胞。例如,可以将编码指导RNA的DNA可操作地连接至启动子控制序列,以在目标

细胞中表达指导RNA。RNA编码序列可以可操作地连接至被RNA聚合酶III (Pol III) 识别的启动子序列。可用于表达指导RNA的质粒载体包括但不限于px330载体和px333载体。在某些情况下,质粒载体(例如px333载体)可包含至少两个编码指导RNA的DNA序列。

[0343] 选择、设计和验证指导多核苷酸的方法,例如指导RNA和靶向序列在本文中描述并且是本领域技术人员已知的。例如,为了使脱碱基酶结构域(例如,AID结构域)中的脱氨酶结构域的潜在受质混杂的影响最小化,可以无意地靶向脱氨的残基数量(例如,脱靶的C残基可能可以最小化在靶核酸基因座内的ssDNA上的残基)。另外,可以使用软件工具来优化对应于靶核酸序列的gRNA,例如以最小化整个基因组的脱靶活性。例如,对于使用化脓性链球菌Cas9进行的每个可能的靶向结构域选择,可以在整个基因组中鉴定出所有脱靶序列(在选择PAM之前,例如NAG或NGG),其中包含多达一定数量(例如1、2、3、4、5、6、7、8、9或10)个不匹配的碱基对的。可以鉴定出与靶位点互补的gRNA的第一区域,并且可以根据其总的预测脱靶得分对所有第一区域(例如crRNA)进行排名;排名最高的定位域表示目标活动最多且目标最少的活动。可以通过使用本领域已知的和/或本文阐述的方法在功能上评估候选靶向gRNA。

[0344] 作为非限制性实例,可以使用DNA序列搜索算法来鉴定与Cas9一起使用的指导RNA的crRNA中的靶DNA杂交序列。可以使用基于公共工具cas-offer的定制gRNA设计软件来进行gRNA设计,如Bae S., Park J., 和Kim J.-S中所述。Cas-OFFinder:一种快速且通用的算法,可搜索Cas9 RNA引导的核酸内切酶的潜在脱靶位点, *Bioinformatics* 30,1473-1475 (2014)。所述软件会在计算指南全基因组偏离目标的倾向后对指南进行评分。通常,对于长度从17到24不等的指南,会考虑从完全匹配到7个不匹配的匹配。一旦非目标位点通过计算确定后,便会为每个指南计算出总得分,并使用网络版汇总为表格输出接口。除了识别与PAM序列相邻的潜在靶位点外,所述软件还识别与所选靶位点相差1、2、3或3个以上核苷酸的所有PAM相邻序列。靶核酸序列的基因组DNA序列,例如可以获得靶基因,并可以使用公开可用的工具(例如RepeatMasker程序)筛选重复组件。RepeatMasker在输入的DNA序列中搜索重复元素和低复杂度的区域。输出是给定查询序列中存在的重复的详细注释。

[0345] 在鉴定之后,指导RNA的第一区域,例如,crRNA可以根据它们与靶位点的距离,其正交性和与相关PAM序列(例如,基于鉴定包含以下内容的人类基因组中的紧密匹配的5' G)的紧密匹配而存在的5' 核苷酸进行分类:相关的PAM,例如化脓性链球菌的NGG PAM,金黄色葡萄球菌的NNGRRT或NNGRRV PAM)。如本文所用,正交性是指人类基因组中包含与靶序列的最小错配的最小数目的序列的数目。“高水平的正交性”或“良好的正交性”可以例如是指20聚体靶向结构域,其在人类基因组中除了预期的靶标之外没有相同的序列,也没有在靶标中包含一个或两个错配的任何序列。可以选择具有良好正交性的靶向结构域,以最大程度地减少脱靶DNA的切割。

[0346] 在一些实施方案中,报告系统可用于检测碱基编辑活性并测试候选指导多核苷酸。在一些实施方案中,报告系统可以包括基于报导基因的测定,其中碱基编辑活性导致报导基因的表达。例如,报告系统可以包括报告基因,所述报告基因包含失活的起始密码子,例如模板链上从3' -TAC-5' 到3' -CAC-5' 的突变。靶C成功脱氨后,相应的mRNA将被转录为5' -AUG-3' 而不是5' -GUG-3',从而能够翻译报告基因。合适的报导基因对于本领域技术人员是显而易见的。报告基因的非限制性实例包括编码绿色荧光蛋白(GFP)、红色荧光蛋白

(RFP)、荧光素酶、分泌性碱性磷酸酶 (SEAP) 的基因,或对本领域技术人员显而易见的表达的任何其他基因。报告系统可用于测试许多不同的gRNA,例如,以便确定相应脱氨酶将靶向靶DNA序列的哪个残基。还可以测试靶向非模板链的sgRNA,以评估特定碱基编辑蛋白(例如,Cas9脱氨酶融合蛋白。在一些实施方案中,可以设计这样的gRNA,使得突变的起始密码子不会与gRNA碱基配对。指导多核苷酸可包含标准核糖核苷酸,修饰的核糖核苷酸(例如假尿苷),核糖核苷酸异构体和/或核糖核苷酸类似物。在一些实施方案中,指导多核苷酸可以包含至少一种可检测标记。可检测标记可以是荧光团(例如,FAM、TMR、Cy3、Cy5、Texas Red、Oregon Green、Alexa Fluors、Halo标签或合适的荧光染料)、检测标签(例如生物素,异羟基洋地黄毒苷等)、量子点或金粒子。

[0347] 所述指导多核苷酸可以化学合成、酶合成或其组合。例如,可以使用基于标准亚磷酸酰胺的固相合成方法合成指导RNA。或者,可通过将编码指导RNA的DNA与噬菌体RNA聚合酶识别的启动子控制序列可操作地连接,在体外合成指导RNA。合适的噬菌体启动子序列的实例包括T7、T3、SP6启动子序列或其变体。在其中指导RNA包含两个分开的分子(例如,crRNA和tracrRNA)的实施方案中,crRNA可以被化学合成并且tracrRNA可以被酶合成。

[0348] 在一些实施方案中,碱基编辑器系统可以包含多个指导多核苷酸,例如,gRNA。例如,gRNA可以靶向一个或多个靶基因座(例如,至少1gRNA、至少2gRNA、至少5gRNA、至少10gRNA、至少20gRNA、至少30gRNA、至少50g gRNA)包含在碱基编辑器系统中。所述多个gRNA序列可以串联排列,并且优选被直接重复隔开。

[0349] 编码指导RNA或指导多核苷酸的DNA序列也可以是载体的一部分。此外,载体可包含其他表达控制序列(例如增强子序列、Kozak序列、聚腺苷酸化序列、转录终止序列等)、选择性标记序列(例如GFP或抗生素抗性基因,例如嘌呤霉素)、复制起点等等。编码指导RNA的DNA分子也可以是线性的。编码指导RNA或指导多核苷酸的DNA分子也可以是环状的。

[0350] 在一些实施方案中,碱基编辑器系统的一个或多个组件可以由DNA序列编码。可以将此类DNA序列引入表达系统,例如cDNA。一个单元、一起或分开。例如,可以将编码多核苷酸可编程核苷酸结合结构域和指导RNA的DNA序列引入细胞,每个DNA序列可以是单独分子的一部分(例如,一个包含多核苷酸可编程核苷酸结合结构域编码序列的载体和另一个包含指导RNA编码序列的载体或两者可以是同一分子的一部分(例如,一个包含多核苷酸可编程核苷酸结合域和指导RNA的编码(和调节)序列的载体)。

[0351] 指导多核苷酸可以包含一个或多个修饰,以提供具有新的或增强的特征的核酸。指导多核苷酸可以包含核酸亲和标签。指导多核苷酸可以包含合成核苷酸、合成核苷酸类似物、核苷酸衍生物和/或修饰的核苷酸。

[0352] 在某些情况下,gRNA或指导多核苷酸可包含修饰。可以在gRNA或指导多核苷酸的任何位置进行修饰。可以对单链gRNA或指导多核苷酸进行一个以上的修饰。修饰后,gRNA或指导多核苷酸可以进行质量控制。在某些情况下,质量控制可以包括PAGE、HPLC、MS或其任意组合。

[0353] gRNA或指导多核苷酸的修饰可以是取代、插入、缺失、化学修饰、物理修饰、稳定化、纯化或其任何组合。

[0354] 也可以通过5' 腺苷酸、5' 鸟苷三磷酸帽、5' N7-甲基鸟苷三磷酸帽、5' 三磷酸帽、3' 磷酸酯、3' 硫代磷酸酯、5' 磷酸酯、5' 硫代磷酸酯、顺式胸腺嘧啶二聚体、三聚体、C12间隔

基、C3间隔基、C6间隔基、d间隔基、PC间隔基、r间隔基、18间隔基9,3'-3' 修饰、5'-5' 修饰、无碱基、吡啶、偶氮苯、生物素、生物素BB、生物素TEG、胆甾醇TEG、脱硫生物素TEG、DNP TEG、DNP-X、DOTA、dT-生物素、双重生物素、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、羧基接头、硫醇接头、2'-脱氧核糖核苷类似物嘌呤、2'-脱氧核糖核苷类似物嘧啶、核糖核苷类似物、2'-O-甲基核糖核苷类似物、糖修饰的类似物、摆动/通用碱基、荧光染料标记、2'-氟RNA、2'-O-甲基RNA、磷酸甲酯、磷酸二酯脱氧核糖核酸、磷酸二酯RNA、硫代磷酸酯DNA、硫代磷酸酯RNA、UNA、伪尿苷5'-三磷酸酯、5'-甲基胞苷5'-三磷酸酯或其任意组合修饰gRNA或指导多核苷酸。

[0355] 在某些情况下,修饰是永久性的。在其他情况下,修饰是暂时的。在一些情况下,对gRNA或指导多核苷酸进行了多种修饰。gRNA或指导多核苷酸修饰可以改变核苷酸的物理化学性质,例如其构象、极性、疏水性、化学反应性、碱基配对相互作用或其任何组合。

[0356] 修饰也可以是硫代磷酸酯替代物。在某些情况下,天然磷酸二酯键可能易于被细胞核酸酶快速降解,并且;使用硫代磷酸酯(PS)键替代物修饰核苷酸间键对通过细胞降解的水解更稳定。修饰可以增加gRNA或指导多核苷酸的稳定性。修饰还可以增强生物活性。在某些情况下,硫代磷酸酯增强的RNA gRNA可抑制RNase A、RNase T1、小牛血清核酸酶或它们的任何组合。这些特性可允许将PS-RNA gRNA用于在体内或体外极有可能暴露于核酸酶的应用中。例如,可以在gRNA的5' 或"端的最后3-5个核苷酸之间引入硫代磷酸酯(PS)键,这可以抑制核酸外切酶降解。在某些情况下,可以在整个gRNA中添加硫代磷酸酯键,以减少核酸内切酶的攻击。

[0357] 原间隔子相邻主题

[0358] 术语“间隔子邻近基序(PAM)”或PAM样基序是指在CRISPR细菌适应性免疫系统中紧随Cas9核酸酶靶向的DNA序列之后的2-6个碱基对的DNA序列。在一些实施例中,PAM可以是5' PAM(即,位于原间隔物的5' 端的上游)。在其他实施例中,PAM可以是3' PAM(即位于原间隔物的5' 端的下游)。

[0359] 原间隔子相邻基序(PAM)或PAM样基序是指在CRISPR细菌适应性免疫系统中紧随Cas9核酸酶靶向的DNA序列后的2-6个碱基对的DNA序列。在一些实施例中,PAM可以是5' PAM(即,位于原间隔物的5' 端的上游)。在其他实施例中,PAM可以是3' PAM(即,位于原间隔物的5' 端的下游)。PAM序列对于靶标结合至关重要,但是确切的序列取决于Cas蛋白的类型。

[0360] 本文提供的碱基编辑器可以包含CRISPR蛋白衍生的结构域,其能够结合包含典型的或非典型的原间隔子相邻基序(PAM)序列的核苷酸序列。PAM位点是接近靶多核苷酸序列的核苷酸序列。本申请的一些方面提供了包含全部或部分具有不同PAM特异性的CRISPR蛋白质的碱基编辑器。例如,典型的Cas9蛋白,例如化脓性链球菌的Cas9(spCas9),需要典型的NGG PAM序列结合特定的核酸区域,其中“NGG”中的“N”是腺嘌呤(A)、胸腺嘧啶(T)、鸟嘌呤(G)或胞嘧啶(C),且G为鸟嘌呤。PAM可以是CRISPR蛋白质特异性的,并且在包含不同CRISPR蛋白质衍生域的不同碱基编辑器之间可以不同。PAM可以是目标序列的5'或3'。PAM可以在靶序列的上游或下游。PAM的长度可以是1、2、3、4、5、6、7、8、9、10个或更多个核苷酸。通常,PAM的长度在2-6个核苷酸之间。

[0361] 在一些实施方案中,Cas9结构域是化脓性链球菌(SpCas9)的Cas9结构域。在一些

实施方案中,SpCas9结构域是核酸酶活性SpCas9,核酸酶非活性SpCas9 (SpCas9d) 或SpCas9切口酶 (SpCas9n)。在一些实施方案中,SpCas9在本文提供的任何氨基酸序列中包含D9X突变或相应的突变,其中X是除D外的任何氨基酸。在一些实施方案中,SpCas9包括D9A突变或相应的突变。在本文提供的任何氨基酸序列中。在一些实施方案中,SpCas9结构域、SpCas9d结构域或SpCas9n结构域可以结合具有非典型PAM的核酸序列。在一些实施方案中,SpCas9结构域、SpCas9d结构域或SpCas9n结构域可以结合具有NGG、NGA或NGCG PAM序列的核酸序列。在一些实施方案中,SpCas9结构域在本文提供的任何氨基酸序列中包含D1135X、R1335X和T1336X突变或相应突变中的一个或多个,其中X是任何氨基酸。在一些实施方案中,SpCas9结构域在本文提供的任何氨基酸序列中包含D1135E、R1335Q和T1336R突变或相应突变中的一个或多个。在一些实施方案中,SpCas9结构域包含D1135E、R1335Q和T1336R突变,或本文提供的任何氨基酸序列中的相应突变。在一些实施方案中,SpCas9结构域在本文提供的任何氨基酸序列中包含D1135X、R1335X和T1336X突变或相应突变中的一个或多个,其中X是任何氨基酸。在一些实施方案中,SpCas9结构域在本文提供的任何氨基酸序列中包含D1135V、R1335Q和T1336R突变或相应突变中的一个或多个。在一些实施方案中,SpCas9结构域包含D1135V、R1335Q和T1336R突变,或本文提供的任何氨基酸序列中的相应突变。在一些实施方案中,SpCas9结构域在本文提供的任何氨基酸序列中包含D1135X、G1217X、R1335X和T1336X突变或相应突变中的一个或多个,其中X是任何氨基酸。在一些实施方案中,SpCas9结构域在本文提供的任何氨基酸序列中包含D1135V、G1217R、R1335Q和T1336R突变或相应突变中的一个或多个。在一些实施方案中,SpCas9结构域在本文提供的任何氨基酸序列中包含D1135V、G1217R、R1335Q和T1336R突变或相应的突变。

[0362] 在一些实施方案中,本文提供的任何融合蛋白的Cas9结构域包含至少60%、至少65%、至少70%、至少75%、至少80%的氨基酸序列,与本文所述的Cas9多肽具有至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%的同一性。在一些实施方案中,本文提供的任何融合蛋白的Cas9结构域包含本文描述的任何Cas9多肽的氨基酸序列。在一些实施方案中,本文提供的任何融合蛋白的Cas9结构域由本文描述的任何Cas9多肽的氨基酸序列组成。

[0363] 示例性的结合PAM的SpCas9的氨基酸序列如下:

[0364] MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFILVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLEPKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPIYYVGPLARGNSRFAMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIMARENQTTQKGQKNSRERMKRIEIGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDHDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAE

RGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKREINN
YHHAHDAYLNAVVGTALEKYPKLESEFVYGDYKVDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGE
IRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYG
GFDSPVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSEFKNPIDFLEAKGYKEVKKDLIIKLPKYSLELEN
GRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIEQISEFSKRVLAD
ANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIIHQSIITGLYETRI
DLSQLGGD。

[0365] 示例性的结合PAM的SpCas9n的氨基酸序列如下：

[0366] MDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA
RRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVD
STDKADRLRIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFILQVQTYNQLFEENPINASGVDAKAILSARLSKSR
RLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADFLAAKNL
SDAILLSDILRVNTEITKAPLSAMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRI
PYVVGPLARGNSRFAWMTRKSEETITPWNFEVVDKGASQSFIERMTNFDKNLPNEKVLPHKSLLYEYFTVYNEL
TKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKI
IKDKDFLDNEENEDILEDIVLTTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
KTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGSLEHIANLAGSPAIIKKGILQTVKVVDELVKVMG
RHKPENIVIEMARENQTTQKGQKNSRERMKRIEIEGKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQEL
DINRLSDYDHDHVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAE
RGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKREINN
YHHAHDAYLNAVVGTALEKYPKLESEFVYGDYKVDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGE
IRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYG
GFDSPVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSEFKNPIDFLEAKGYKEVKKDLIIKLPKYSLELEN
GRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIEQISEFSKRVLAD
ANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIIHQSIITGLYETRI
DLSQLGGD。

[0367] 示例性的PAM结合SpEQR Cas9的氨基酸序列如下：

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFD
SGETAETRLKRTARRRYTRRKNRICYLQEFSNEMAKVDDSSFFHRLEESFVEE
DKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYALAHMIK
FRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSK
SRRLENLIAQLPGEKKNGLFGNLIASLGLTPNFKSNFDLAEDAKLQLSKDTYD
DDLNDLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDE
HHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILE
KMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKD
NREKIEKILTRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
SFIERMTNFDKNLPNEKVLPHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGE
QKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTIAHLFDDKVMKQLK
RRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKE
[0368] DIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGKELGSQILKEHPVENTQLQNEKLYL
YYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGK
SDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR
QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFY
KVREINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF
ATVRKVLSPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGG
FESPTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSEKPNIDFLEAKGY
KEVKKDLIIKLPKYSLELENGRKRMLASAGELQKGNELALPSKYVNFYLYLASH
YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVL SAYN
KHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL DATLIHQ
SITGLYETRIDLSQLGGD。

[0369] 在此序列中, 可以从D1135、R1335和T1337突变以产生SpEQR Cas9的残基E1135、Q1335和R1337用底线和粗体标出。

[0370] 示例性的结合PAM的SpVQR Cas9的氨基酸序列如下:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFD
SGETAETRLKRTARRRYTRRKNRICYLQEFSNEMAKVDDSSFFHRLEESFLVE
EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYALAHMI
KFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLS
KSRRLENLIAQLPGEKKNGLFGNLIASLGLTPNFKSNFDLAEDAKLQLSKDTY
[0371] DDDLNDLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYD
EHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILE
KMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKD
NREKIEKILTRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
SFIERMTNFDKNLPNEKVLPHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGE
QKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL

LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLK
RRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKE
DIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYL
YYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGK
SDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR
[0372] QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFY
KVREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF
ATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD PKKYGG
FVSPTVAYSVLVVAKEVEKGKSKKLKSVKELLGITIMERS SFEKNPIDFLEAKGY
KEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH
YEKLKGSPEDNEQKQLFVEQHKHYLDEIIIEQISEFSKRVLADANLDKVL SAYN
KHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK **QYR**STKEVL DATLIHQ
SITGLYETRIDLSQLGGD。

[0373] 在此序列中, 可以从D1135、R1335和T1336突变以产生SpVQR Cas9的残基V1135、Q1335和R1336以底线和粗体标出。

[0374] 示例性的结合PAM的SpVRER Cas9的氨基酸序列如下:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFD
SGETAETRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDD SFFHRLEESFLVE
EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADRLIYLALAHMI
KFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLS
KSRRLENLIAQLPGEKKNGLFGNLIASLGLTPNFKSNFDLAEDAKLQLSKDTY
DDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYD
EHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILE
KMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPFLKD
NREKIEKILTFRIPIYVGPLARGNSRFAWMTRKSEETITPWNFEVVDKGASAQ
SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGE
[0375] QKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLK
RRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKE
DIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYL
YYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGK
SDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR
QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFY
KVREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF
ATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD PKKYGG
FVSPTVAYSVLVVAKEVEKGKSKKLKSVKELLGITIMERS SFEKNPIDFLEAKGY

[0376] KEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFYLYLASH
YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYN
KHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKEYRSTKEVL DATLIHQ
SITGLYETRIDLSQLGGD.

[0377] 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域是核酸酶活性的SpyMacCas9、核酸酶非活性的SpyMacCas9 (SpyMacCas9d) 或SpyMacCas9切口酶(SpyMacCas9n)。在一些实施方案中,SaCas9结构域,SaCas9d结构域或SaCas9n结构域可以结合具有非典型PAM的核酸序列。在一些实施方案中,SpyMacCas9结构域、SpCas9d结构域或SpCas9n结构域可以结合具有NAA PAM序列的核酸序列。

[0378] 示例性SpyMacCas9

[0379] MDKKYSIGLDIGTNSVGWAVITDDYKVP SKKFKVLGNTDRHS IKKNLIGALLFGSGETAETRLKRTA
RRRYTRRKNRICYLQEIFSNEMAKVDDSFHRL EESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLAD
STDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQIYNQLFEENPINASRVDAKAILSARLSKSR
RLENLIAQLPGEKRNGLFGNLIALSLGLTPNFKS NFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNL
SDAILLS DILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGS IPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRI
PYVVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNEL
TKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVT VKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDLLKI
IKDKDFLDNEENEDILEDIVLTTLTFEDRGMIEERL KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
KTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGHS LHEQIANLAGSPAIKKGILQTVKIVDELVKVMG
HKPENIV IEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELD
INRLSDYDV DHIVPQSF IKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAER
GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFYK VREINNY
HHAHDAYLNAVVG TALIKKYPKLESEFVYG DYKVYDVRKMI AKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI
RKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEIQTVGQNGGLFDDNPKSPLEVTPSKLVPLKKELN
PKKYGGYQKPTTAYPVLLITDTKQLIPISVMNKKQFEQNPVKFLRDRGYQQVGKNDFIKLPKYTLVDIGDGIKRLW
ASSKEIHKGNQLVVS KKSQILLYHAHHLDS DLSNDYLQNHNNQQFDVLFNEIISFSKKCKLGKEHIQKIENVYSNKK
NSASIEELAESFIKLLGFTQLGATSPFNFLGVKLNQKQYKGKKDYILPCTEGTLIRQSITGLYETRVDLSKIGED.

[0380] 在一些情况下,变体Cas9蛋白带有H840A、P475A、W476A、N477A、D1125A、W1126A和D1218A突变,使得所述多肽切割靶DNA或RNA的能力降低。这样的Cas9蛋白切割靶DNA(例如单链靶DNA)的能力降低,但是保留了结合靶DNA(例如单链靶DNA)的能力。作为另一个非限制性实例,在一些情况下,变体Cas9蛋白带有D10A、H840A、P475A、W476A、N477A、D1125A、W1126A和D1218A突变,使得所述多肽切割靶DNA的能力降低。这样的Cas9蛋白切割靶DNA(例如单链靶DNA)的能力降低,但是保留了结合靶DNA(例如单链靶DNA)的能力。在某些情况下,当变异的Cas9蛋白带有W476A和W1126A突变,或者变异的Cas9蛋白带有P475A、W476A、N477A、D1125A、W1126A和D1218A突变时,变异的Cas9蛋白不能有效地与PAM序列结合。因此,在一些这样的情况下,当在结合方法中使用这种变体Cas9蛋白时,所述方法不需要PAM序列。换句话说,在某些情况下,当将这种Cas9变异蛋白用于结合方法时,所述方法可包含指

导RNA,但所述方法可在不存在PAM序列的情况下进行(结合的特异性为因此由指导RNA的靶向片段提供)。可以使其他残基突变以获得上述效果(即,使一个或另一个核酸酶部分失活)。作为非限制性实例,残基D10、G12、G17、E762、H840、N854、N863、H982、H983、A984、D986和/或A987可以被改变(即被取代)。同样,除丙氨酸取代以外的突变也是合适的。

[0381] 在一些实施方案中,碱基编辑器的CRISPR蛋白来源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Cas9衍生的结构域可以采用非典型的PAM序列。这样的序列已在本领域中描述,并且对本领域技术人员而言是显而易见的。例如,在Kleinstiver,B.P.等人,“Engineered CRISPR-Cas9 nucleases with altered PAM specificities”,*Nature* 523,481-485(2015)中已经描述了结合非典型PAM序列的Cas9结构域。Kleinstiver,B.P.等人,“Broadening the targeting range of *Staphylococcus aureus* CRISPR-Cas9 by modifying PAM recognition”*Nature Biotechnology* 33,1293-1298(2015);每个的全部内容通过引用合并于此。

[0382] 在一些实例中,可以将本文公开的碱基编辑器的CRISPR蛋白衍生的结构域识别的PAM提供给编码碱基编辑器的插入物(例如AAV插入物)的单独寡核苷酸上的细胞。在这种情况下,在单独的寡核苷酸上提供PAM可以切割原本无法被切割的靶序列,因为在与靶序列相同的多核苷酸上不存在相邻的PAM。

[0383] 在一个实施方案中,化脓性链球菌Cas9(SpCas9)可作用于基因组工程的CRISPR核酸内切酶。但是,可以使用其他方法。在某些情况下,可以使用不同的核酸内切酶靶向某些基因组靶标。在某些情况下,可以使用具有非NGG PAM序列的合成的SpCas9衍生变体。另外,已经鉴定了来自各种物种的其他Cas9直系同源物,并且这些“非SpCas9s”可以结合多种PAM序列,这也可用于本申请。例如,相对较大的SpCas9大小(大约4kb编码序列)会导致携带SpCas9 cDNA的质粒无法在细胞中有效表达。相反,金黄色葡萄球菌Cas9(SaCas9)的编码序列比SpCas9短约1千个碱基,可能使其在细胞中有效表达。与SpCas9相似,SaCas9核酸内切酶能够在体外和体内修饰哺乳动物细胞中的靶基因。在某些情况下,Cas蛋白可以靶向不同的PAM序列。在某些情况下,靶基因可能与Cas9 PAM(例如5'-NGG)相邻。在其他情况下,其他Cas9直系同源物可能具有不同的PAM要求。例如,其他PAM,例如嗜热链球菌(对于CRISPR1为5'-NNAGAA,对于CRISPR3为5'-NGGNG)和脑膜炎双球菌(5'-NNNGATT)。

[0384] 在一些实施方案中,对于化脓性链球菌系统,靶基因序列可以在5'-NGG PAM之前(即5'至),并且20-nt指导RNA序列可以与相反链碱基配对介以与PAM相邻的Cas9切割。在某些情况下,相邻的片段可以是或可以是PAM上游的约3个碱基对。在某些情况下,相邻的片段可以是或可以是PAM上游的约10个碱基对。在一些情况下,相邻的片段可以是PAM上游或可以是PAM上游的约0-20个碱基对。例如,相邻的切割可以紧邻,1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20PAM上游的21、22、23、24、25、26、27、28、29或30个碱基对。相邻的片段也可以在PAM的下游1到30个碱基对。

[0385] 包含核定位序列(nuclear localization sequence,NLS)的融合蛋白

[0386] 在一些实施方案中,本文提供的融合蛋白进一步包含一个或多个(例如2、3、4、5)核靶向序列,例如核定位序列(NLS)。在一实施例中,使用二分式NLS。在一些实施方案中,NLS包含有助于将包含NLS的蛋白质导入细胞核的氨基酸序列(例如,通过核运输)。在一些实施方案中,本文提供的任何融合蛋白还包含核定位序列(NLS)。在一些实施方案中,将NLS

融合至融合蛋白的N端。在一些实施方案中,将NLS融合至融合蛋白的C端。在一些实施方案中,NLS与Cas9结构域的N-末端融合。在一些实施方案中,将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与脱氨酶的N-末端融合。在一些实施方案中,NLS与脱氨酶的C-末端融合。在一些实施方案中,NLS通过一个或多个接头与融合蛋白融合。在一些实施方案中,将NLS与没有接头的融合蛋白融合。在一些实施方案中,NLS包含本文提供或引用的任何NLS序列的氨基酸序列。另外的核定位序列在本领域中是已知的,并且对本领域技术人员而言是显而易见的。例如,NLS序列在Plank等人,PCT/EP2000/011690中描述,其内容通过引用并入本文作为示例性核定位序列的参考。在一些实施方案中,NLS包含氨基酸序列PKKKRKVEGADKRTADGSEFESPKKKRKV、KRTADGSEFESPKKKRKV、KRPAATKKAGQAKKKK、KKTELQTTNAENKTKKL、KRGINDRNFWRGENGKTR、RKSGKIAAIVVKRPRKPKKKRKVK或KMLKNRLCFYQLY。在一些实施方案中,NLS存在于接头中或NLS侧接于接头,例如本文所述的接头。在一些实施方案中,N末端或C末端NLS是二分NLS。二分体NLS包含两个碱性氨基酸簇,它们由相对较短的间隔序列隔开(因此,二分体-2个部分,而单分体NLS没有)。核纤溶酶的NLS KR[PAATKKAGQA]KKKK是普遍存在的二分信号的原型:两个碱性氨基酸簇,由大约10个氨基酸的间隔区隔开。示例性两方NLS的序列如下:PKKKRKVEGADKRTADGSEFES PKKKRKV。

[0387] 在一些实施方案中,本发明的融合蛋白不包含接头序列。在一些实施方案中,存在一个或多个结构域或蛋白质之间的接头序列。

[0388] 应了解,本发明的融合蛋白可包含一个或多个附加特征。例如,在一些实施方案中,融合蛋白可包含抑制剂、胞质定位序列、输出序列(例如核输出序列或其他定位序列),以及可用于增溶、纯化或检测融合蛋白的序列标签。本文提供的合适的蛋白质标签包括但不限于生物素羧化酶载体蛋白(BCCP)标签、myc标签、钙调蛋白标签、FLAG标签、血凝素(HA)标签、多组氨酸标签(也称为组氨酸标签或His-标签)、麦芽糖结合蛋白(MBP)标签、nus标签、谷胱甘肽S-转移酶(GST)标签、绿色荧光蛋白(GFP)标签、硫氧还蛋白标签、S标签、Softags(例如,Softag 1、Softag 3)、链球菌标签、生物素连接酶标签、FlAsH标签、V5标签和SBP标签。对于本领域技术人员而言,其他合适的序列是显而易见的。在一些实施方案中,融合蛋白包含一个或多个His标签。

[0389] 可以使用编码包含一个或多个核定位序列(NLS)的CRISPR酶的载体。例如,可以使用或大约使用1、2、3、4、5、6、7、8、9、10个NLS。CRISPR酶可以在弹药末端或附近包含NLS,在羧基末端或附近包含约或大于约1、2、3、4、5、6、7、8、9、10个NLS这些的组合(例如,在氨基末端的一个或多个NLS和在羧基末端的一个或多个NLS)。当存在一个以上的NLS时,可以彼此独立地选择一个,这样单个NLS可以一个以上的副本存在和/或与一个或多个其他的NLS一起存在一个或多个副本中。

[0390] 所述方法中使用的CRISPR酶可包含约6个NLS。当最接近NLS的氨基酸在距N端或C端的多肽链中约50个氨基酸以内时,例如在1、2、3、4、5、10、15、20、25、30、40或50个氨基酸内,则认为NLS接近N端或C端。

[0391] 在一些实施方案中,NLS包含氨基酸序列PKKKRKVEGADKRTADGSEFES PKKKRKV、KRTADGSEFESPKKKRKV、KRPAATKKAGQAKKKK、KKTELQTTNAENKTKKL、KRGINDRNFWRGENGKTR、RKSGKIAAIVVKRPRKPKKKRKV或MDSLLMNRKFLYQFKNVRWAKGRRETYLC。

[0392] 在一些实施方案中,NLS存在于接头中或NLS侧接于接头,例如本文所述的接头。在

一些实施方案中,N末端或C末端NLS是二分NLS。二分体NLS包含两个碱性氨基酸簇,它们被相对较短的间隔序列隔开(因此,二分体-2个部分,而单分体NLS没有)。核纤溶酶的NLS KR [PAATKKAGQA]KKKK是普遍存在的二分信号的原型:两个碱性氨基酸簇,由大约10个氨基酸的间隔区隔开。示例性两方NLS的序列如下:PKKKRKVEGADKRTADGSEFES PKKKRKV。

[0393] 在一些实施方案中,本发明的融合蛋白不包含接头序列。在一些实施方案中,存在一个或多个结构域或蛋白质之间的接头序列。

[0394] PAM序列可以是本领域已知的任何PAM序列。合适的PAM序列包括但不限于NGG、NGA、NGC、NGN、NGT、NGCG、NGAG、NGAN、NGNG、NGCN、NGCG、NGTN、NNGRRT、NNNRRT、NGRRR (N)、TTTV、TYCV、TYCV、TATV、NNGTAT、NNAGAAW或NAAAAC。Y是嘧啶;N是任何核苷酸碱基;W是A或T。

[0395] 核碱基编辑结构域

[0396] 本文描述了包含融合蛋白的碱基编辑器,所述融合蛋白包括多核苷酸可编程核苷酸结合结构域和核苷碱基编辑结构域(例如,脱氨酶结构域)。通过与能够识别靶序列的指导多核苷酸相互作用,可以对碱基编辑器进行编程以编辑靶多核苷酸序列中的一个或多个碱基。一旦识别出靶序列,就将碱基编辑器锚定在将要进行编辑的多核苷酸上,然后碱基编辑器的脱氨酶结构域组分可以编辑靶碱基。

[0397] 在一些实施方案中,核碱基编辑结构域是脱氨酶结构域。在某些情况下,脱氨酶结构域可以是胞嘧啶脱氨酶或胞苷脱氨酶。在一些实施方案中,术语“胞嘧啶脱氨酶”和“胞苷脱氨酶”可以互换使用。在某些情况下,脱氨酶结构域可以是腺嘌呤脱氨酶或腺苷脱氨酶。在一些实施方案中,术语“腺嘌呤脱氨酶”和“腺苷脱氨酶”可以互换使用。核碱基编辑蛋白的细节在国际PCT第PCT/2017/045381号专利申请(WO第2018/027078号国际专利申请)和PCT/US2016/058344 (WO2017/070632)中进行了描述,将其全部内容通过引用并入本文。另请参阅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);以及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),全文其内容通过引用合并于此。

[0398] C到T编辑

[0399] 在一些实施方案中,本文公开的碱基编辑器包括包含胞苷脱氨酶的融合蛋白,所述胞嘧啶脱氨酶能够使多核苷酸的靶胞苷(C)碱基脱氨以产生尿苷(U),所述尿苷具有胸腺嘧啶的碱基配对特性。在一些实施方案中,例如在多核苷酸是双链的(例如DNA)的情况下,然后尿苷碱基可以被胸苷碱基取代(例如通过细胞修复机制)以产生C:G至T:A的转变。在其他实施方案中,碱基编辑器将核酸中的C脱氨为U时,不能将U取代为T。

[0400] 多核苷酸中靶C的脱氨基以产生U是可以由本文所述的碱基编辑器执行的碱基编辑类型的非限制性实例。在另一个实例中,包含胞苷脱氨酶结构域的碱基编辑器可以介导胞嘧啶(C)碱基向鸟嘌呤(G)碱基的转化。例如,可以通过碱基切除修复机制(例如,通过尿嘧啶DNA糖基化酶(UDG)结构域),从碱基编辑器的胞嘧啶脱氨酶结构域的胞嘧啶脱氨基所

产生的多核苷酸的U,从多核苷酸中切除,产生一个无基础的网站。然后通过例如转移损伤聚合酶将与无碱基位点相对的核碱基(例如通过碱基修复机制)替换为另一碱基,例如C。尽管通常将与无碱基位点相反的核碱基用C取代,但也可以发生其他取代(例如A,G或T)。

[0401] 因此,在一些实施方案中,本文所述的碱基编辑器包含能够将多核苷酸中的靶C脱氨到U的脱氨结构域(例如胞苷脱氨酶结构域)。此外,如下所述,碱基编辑器可包含促进由脱氨基产生的U转化为T或G的其他结构域。例如,包含胞苷脱氨酶域的碱基编辑器还可包含尿嘧啶糖基化酶抑制剂(UGI)域来介导U被T取代,从而完成C到T碱基编辑事件。在另一个实例中,碱基编辑器可以掺入跨病变聚合酶以提高C-to-G碱基编辑的效率,因为跨病变聚合酶可以促进与无碱基位点相对的C的掺入(即,导致在无碱基位点掺入G,完成C-to-G碱基编辑活动)。

[0402] 包含胞苷脱氨酶作为结构域的碱基编辑器可以使任何多核苷酸,包括DNA、RNA和DNA-RNA杂种中的靶C脱氨基。通常,胞苷脱氨酶催化位于多核苷酸单链部分的上下文中的C核碱基。在一些实施方案中,包含靶C的整个多核苷酸可以是单链的。例如,并入碱基编辑器中的胞苷脱氨酶可以使单链RNA多核苷酸中的靶C脱氨基。在其他实施方案中,包含胞苷脱氨酶结构域的碱基编辑器可以作用于双链多核苷酸,但是靶标C可以位于在脱氨反应时处于单链状态的多核苷酸的一部分中。例如,在其中NAGPB结构域包含Cas9结构域的实施方案中,在形成Cas9-gRNA-靶DNA复合物期间,几个核苷酸可以不成对,导致形成Cas9“R-环复合物”。这些未配对的核苷酸可以形成单链DNA的气泡,所述单链DNA可以用作单链特异性核苷酸脱氨酶(例如胞苷脱氨酶)的受质。

[0403] 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B mRNA编辑复合体(APOBEC)家族脱氨酶的全部或一部分。APOBEC是进化上保守的胞苷脱氨酶家族。所述家族的成员是C到U(C-to-U)编辑酶。像蛋白质一样,APOBEC的N末端结构域是催化结构域,而C末端结构域是伪催化结构域。更具体地说,催化结构域是锌依赖性胞苷脱氨酶结构域,并且对于胞苷脱氨是重要的。APOBEC家族成员包括APOBEC1、APOBEC2、APOBEC3A、APOBEC3B、APOBEC3C、APOBEC3D(现在称为“APOBEC3E”)、APOBEC3F、APOBEC3G、APOBEC3H、APOBEC4和激活诱导的(胞苷)脱氨酶。在一些实施方案中,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全部或部分活化诱导的脱氨酶(AID)。在一些实施方案中,并入碱基编辑器中的脱氨酶包含全部或部分胞苷脱氨酶1(CDA1)。应当理解,碱基编

辑器可以包含来自任何合适生物体(例如人或大鼠)的脱氨酶。在一些实施方案中,碱基编辑器的脱氨酶结构域来自人类、黑猩猩、大猩猩、猴子、牛、狗、大鼠或小鼠。在一些实施方案中,碱基编辑器的脱氨酶结构域衍生自大鼠(例如大鼠APOBEC1)。在一些实施方案中,碱基编辑器的脱氨酶结构域是人类APOBEC1。在一些实施方案中,碱基编辑器的脱氨酶结构域是pmCDA1。

[0404] PmCDA1的氨基酸和核酸序列在下文显示。

[0405] >tr|A5H718|A5H718_PETMA胞嘧啶脱氨酶OS=海七鳃鳗OX=7757 PE=2 SV=1氨基酸序列:

[0406] MTDAEYVRIHEKLDIYTFKKQFFNNKKS VSHRCYVLFELKRRGERRACFWGYAVNKPQSG

[0407] TERGIHAEIFSIRKVEEYLRDNPQGFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLK

[0408] IWACKLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKT

[0409] LKRAEKRRSELSIMI QVKILHTTKSPAV

[0410] 核酸序列:>EF094822.1滨海双嗜菌分离物PmCDA.21胞嘧啶脱氨酶mRNA,完整cds:

[0411] TGACACGACACAGCCGTGTATATGAGGAAGGGTAGCTGGATGGGGGGGGGGGAATACGTT CAGAGAGG
ACATTAGCGAGCGTCTTGTGGTGGCCTTGAGTCTAGACACCTGCAGACATGACCGACGCTGAGTACGTGAGAATCC
ATGAGAAGTTGGACATCTACACGTTTAAGAAACAGTTTTTCAACAACAAAAATCCGTGTCGCATAGATGCTACGTT
CTCTTTGAATTAACGACGGGGTGAACGTAGAGCGTGT TTTTGGGGCTATGCTGTGAATAAACACAGAGCGGGAC
AGAACGTGGAATTCACGCCGAAATCTTTAGCATTAGAAAAGTCGAAGAATACCTGCGCGACAACCCCGGACAATTCA
CGATAAATTGGTACTCATCCTGGAGTCCTTGTGCAGATTGCGCTGAAAAGATCTTAGAATGGTATAACCAGGAGCTG
CGGGGGAACGGCCACACTTTGAAAATCTGGGCTTGCAA CTCTATTACGAGAAAAATGCGAGGAATCAAATTGGGCT
GTGGAACCTCAGAGATAACGGGGTTGGGTTGAATGTAATGGTAAGTGAACACTACCAATGTTGCAGGAAAATATTCA
TCCAATCGTCGCACAATCAATTGAATGAGAATAGATGGCTTGAGAAGACTTTGAAGCGAGCTGAAAAACGACGGAGC
GAGTTGTCCATTATGATT CAGGTAAAAATACTCCACACCACTAAGAGTCCTGCTGTTTAAGAGGCTATGCGGATGGT
TTTC

[0412] 人激活诱导的胞苷脱氨酶(AID)的编码序列(CDS)的氨基酸和核酸序列如下所示。

[0413] >tr|Q6QJ80|Q6QJ80_HUMAN激活诱导的胞苷脱氨酶OS=智人OX=9606 GN=AICDA
PE=2 SV=1氨基酸序列:

[0414] MDSLLMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNKGCHVELLFLRYISDWD
LDPGRCYRVTWFTSWSPCYDCARHVADFLRGPNLSLRIFTARLYFCEDRKAEPGLRRLHRAGVQIAIMTFKAPV

[0415] 人激活诱导的胞苷脱氨酶(AID)的编码序列(CDS)的氨基酸和核酸序列如下所示。

[0416] >tr|Q6QJ80|Q6QJ80_HUMAN激活诱导的胞苷脱氨酶OS=智人OX=9606 GN=AICDA
PE=2SV=1氨基酸序列:

[0417] MDSLLMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNKGCHVELLFLRYISDWD
LDPGRCYRVTWFTSWSPCYDCARHVADFLRGPNLSLRIFTARLYFCEDRKAEPGLRRLHRAGVQIAIMTFKAPV

[0418] 核酸序列:>NG_011588.1:5001-15681智人激活诱导的胞苷脱氨酶(AICDA),
RefSeqGene(LRG_17)在12号染色体上:

[0419] AGAGAACCATCATTAATTGAAGTGAGATTTTCTGGCCTGAGACTTGCAGGGAGGCAAGAAGACACTCT
GGACACCACTATGGACAGGTAAAGAGGCAGTCTTCTCGTGGGTGATTGCACTGGCCTTCCTCTCAGAGCAAATCTGA
GTAATGAGACTGGTAGCTATCCCTTTCTCTCATGTAAGTGTCTGACTGATAAGATCAGCTTGATCAATATGCATATA

TATTTTTTGATCTGTCTCCTTTTCTTCTATTTCAGATCTTATACGCTGTCAGCCCAATTCTTTCTGTTTCAGACTTCT
CTTGATTTCCCTCTTTTTTCATGTGGCAAAAGAAGTAGTGCGTACAATGTACTGATTCGTCCTGAGATTTGTACCATG
GTTGAAACTAATTTATGGTAATAATATTAACATAGCAAATCTTTAGAGACTCAAATCATGAAAAGGTAATAGCAGTA
CTGTACTAAAAACGGTAGTGCTAATTTTCGTAATAATTTTGTAATATTCAACAGTAAACAACCTTGAAGACACACT
TTCCTAGGGAGGCGTTACTGAAATAATTTAGCTATAGTAAGAAAATTTGTAATTTTAGAAATGCCAAGCATTCTAAA
TTAATTGCTTGAAAGTCACTATGATTGTGTCCATTATAAGGAGACAAATTCATTCAAGCAAGTTATTTAATGTTAAA
GGCCCAATTGTTAGGCAGTTAATGGCACTTTTACTATTAACCTAATCTTTCCATTTGTTTCAGACGTAGCTTAACCTTAC
CTCTTAGGTGTGAATTTGGTTAAGGTCCTCATAATGTCTTTATGTGCAGTTTTTGATAGGTTATTGTCATAGAAGTT
ATTCTATTCTACATTTATGATTACTATGGATGTATGAGAATAACACCTAATCCTTATACTTTACCTCAATTTAACT
CCTTTATAAAGAACTTACATTACAGAATAAAGATTTTTTAAAAATATATTTTTTTGTAGAGACAGGGTCTTAGCCCA
GCCGAGGCTGGTCTCTAAGTCCTGGCCCAAGCGATCCTCCTGCCTGGGCCCTCTAAAGTGCTGGAATTATAGACATG
AGCCATCACATCCAATATACAGAATAAAGATTTTTAATGGAGGATTTAATGTTCTTCAGAAAATTTTCTTGAGGTCA
GACAATGTCAAATGTCTCCTCAGTTTACACTGAGATTTTGAAAACAAGTCTGAGCTATAGGTCCTTGTGAAGGGTCC
ATTGGAATACTTGTTCAAAGTAAAATGGAAGCAAAGGTAAAATCAGCAGTTGAAATTCAGAGAAAGACAGAAAAG
GAGAAAAGATGAAATTCAACAGGACAGAAGGGAAATATATTATCATTAAAGGAGGACAGTATCTGTAGAGCTCATTAG
TGATGGCAAAATGACTTGGTCAGGATTATTTTAACCCGCTTGTTTCTGGTTTGACGGCTGGGGATGCAGCTAGGG
TTCTGCCTCAGGGAGCACAGCTGTCCAGAGCAGCTGTCAGCCTGCAAGCCTGAAACACTCCCTCGGTAAAGTCCTTC
CTACTCAGGACAGAAATGACGAGAACAGGGAGCTGGAAACAGGCCCTAACCAGAGAAGGGAAGTAATGGATCAACA
AAGTTAACTAGCAGGTCAAGATCACGCAATTCATTTCACTCTGACTGGTAACATGTGACAGAAACAGTGTAGGCTTA
TTGTATTTTCATGTAGAGTAGGACCCAAAAATCCACCCAAAGTCCTTTATCTATGCCACATCCTTCTTATCTATACT
TCCAGGACACTTTTTCTTCCTTATGATAAGGCTCTCTCTCTCCACACACACACACACACACACACACACACAC
ACACACACACACAAACACACACCCCGCCAACCAAGGTGCATGTAAAAAGATGTAGATTCCTCTGCCTTTCTCATCTA
CACAGCCAGGAGGGTAAGTTAATATAAGAGGGATTTATTGGTAAGAGATGATGCTTAATCTGTTTAACACTGGGCC
TCAAAGAGAGAATTTCTTTTCTTCTGTACTTATTAAGCACCTATTATGTGTTGAGCTTATATATACAAAGGGTTATT
ATATGCTAATATAGTAATAGTAATGGTGGTTGGTACTATGGTAATTACCATAAAAATTATTATCCTTTTAAAAATAAA
GCTAATTATTATTGGATCTTTTTTAGTATTCATTTTATGTTTTTTATGTTTTTGATTTTTTAAAAGACAATCTCACC
CTGTTACCCAGGCTGGAGTGCAGTGGTGCATCATAGCTTTCTGCAGTCTTGAACCTCTGGGCTCAAGCAATCCTCC
TGCCTTGGCCTCCCAAAGTGTTGGGATACAGTCATGAGCCACTGCATCTGGCCTAGGATCCATTTAGATTAATAATAT
GCATTTTAAATTTTAAAAATAATATGGCTAATTTTTACCTTATGTAATGTGTATACTGGCAATAAATCTAGTTTGCTG
CCTAAAGTTTAAAGTGCTTTCCAGTAAGCTTCATGTACGTGAGGGGAGACATTTAAAGTGAAACAGACAGCCAGGTG
TGGTGGCTCACGCCTGTAATCCCAGCACTCTGGGAGGCTGAGGTGGGTGGATCGCTTGAGCCCTGGAGTTCAAGACC
AGCCTGAGCAACATGGCAAAACGCTGTTTCTATAACAAAAATTAGCCGGGCATGGTGGCATGTGCCTGTGGTCCCAG
CTACTAGGGGGCTGAGGCAGGAGAATCGTTGGAGCCCAGGAGGTCAAGGCTGCACTGAGCAGTGCTTGCGCCACTGC
ACTCCAGCCTGGGTGACAGGACCAGACCTTGCTCAAAAAATAAGAAGAAAAATTAAAAATAAATGGAACAACATA
CAAAGAGCTGTTGTCCTAGATGAGCTACTTAGTTAGGCTGATATTTTGGTATTTAACTTTTAAAGTCAGGGTCTGTC
ACCTGCACTACATTATTAATAATATCAATTCTCAATGTATATCCACACAAAGACTGGTACGTGAATGTTCATAGTACC
TTTATTACAAAACCCCAAAGTAGAGACTATCCAAATATCCATCAACAAGTGAACAAATAAACAAAATGTGCTATAT
CCATGCAATGGAATACCACCCTGCAGTACAAAGAAGCTACTTGGGGATGAATCCCAAAGTCATGACGCTAAATGAAA
GAGTCAGACATGAAGGAGGAGATAATGTATGCCATACGAAATCTAGAAAATGAAAGTAACTTATAGTTACAGAAAG

CAAATCAGGGCAGGCATAGAGGCTCACACCTGTAATCCCAGCACTTTGAGAGGCCACGTGGGAAGATTGCTAGA
ACTCAGGAGTTCAAGACCAGCCTGGGCAACACAGTGAACTCCATTCTCCACAAAAATGGGAAAAAAGAAAGCAAATCA
GTGGTTGTCCTGTGGGGAGGGGAAGGACTGCAAAGAGGGGAAGAAGCTCTGGTGGGGTGAGGGTGGTGATT
CAGGTTC TGTATCCTGACTGTGGTAGCAGTTTGGGGTGTTTACATCCAAAAATATTCGTAGAATTATGCATCTTAAATGGGTGG
AGTTTACTGTATGTAAATTATACCTCAATGTAAGAAAAAATAATGTGTAAGAAAACCTTTCAATTCTCTTGCCAGCAA
ACGTTATTCAAATTCCTGAGCCCTTTACTTCGCAAATTCTCTGCACTTCTGCCCCGTACCATTAGGTGACAGCACTA
GCTCCACAAATTGGATAAATGCATTTCTGGAAAAAGACTAGGGACAAAATCCAGGCATCACTTGTGCTTTCATATCAA
CCATGCTGTACAGCTTGTGTTGCTGTCTGCAGCTGCAATGGGGACTCTTGATTTCTTTAAGGAACTTGGGTTACCA
GAGTATTTCCACAAATGCTATTCAAATTAGTGCTTATGATATGCAAGACACTGTGCTAGGAGCCAGAAAACAAAGAG
GAGGAGAAATCAGTCATTATGTGGGAACAACATAGCAAGATATTTAGATCATTGACTAGTTAAAAAAGCAGCAGA
GTACAAAATCACACATGCAATCAGTATAATCCAAATCATGTAAATATGTGCCTGTAGAAAGACTAGAGGAATAAACA
CAAGAATCTTAACAGTCATTGTCATTAGACACTAAGTCTAATTATTATTATTAGACACTATGATATTTGAGATTTAA
AAAATCTTTAATATTTTTAAATTTAGAGCTCTTCTATTTTTCCATAGTATTCAAGTTTGACAATGATCAAGTATTAC
TCTTTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTGTAGATGGAGTTTGGTCTTGTGCCCATGCTGGAGTGGAATGGCAT
GACCATAGCTCACTGCAACCTCCACCTCCTGGGTTCAGGCAAGCTGTGCGCTCAGCCTCCCGGGTAGATGGGATTA
CAGGCGCCACCACCACACTCGGCTAATGTTTGTATTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGT
CTCAAACCTGACCTCAGAGGATCCACCTGCCTCAGCCTCCCAAAGTGCTGGGATTACAGATGTAGGCCACTGCGC
CCGGCCAAGTATTGCTCTTATACATTAAAAAACAGGTGTGAGCCACTGCGCCAGCCAGGTATTGCTCTTATACATT
AAAAATAGGCCGGTGCAGTGCTCAGCCTGTAATCCCAGCACTTTGGGAAGCCAAGGCGGGCAGAACACCCGAGG
TCAGGAGTCCAAGGCCAGCCTGGCCAAGATGGTGAAACCCCGTCTCTATTTAAAAATACAAACATTACCTGGGCATGA
TGGTGGGCGCCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGGATCCGCGGAGCCTGGCAGATCTGCCTGAGC
CTGGGAGGTTGAGGCTACAGTAAGCCAAGATCATGCCAGTATACTTCAGCCTGGGCGACAAAGTGAGACCGTAACAA
AAAAAAAAAAATTTAAAAAAGAAATTTAGATCAAGATCCAACTGTAAAAAGTGCCCTAAACACCACATTAAAGAGT
TTGGAGTTTATTCTGCAGGCAGAAGAGAACCATCAGGGGGTCTTCAGCATGGGAATGGCATGGTGCACCTGGTTTTT
GTGAGATCATGGTGGTGACAGTGTGGGGAATGTTATTTTGGAGGGACTGGAGGCAGACAGACCGGTTAAAAGGCCAG
CACAACAGATAAGGAGGAAGAAGATGAGGGCTTGGACCGAAGCAGAGAAGAGCAAACAGGGAAGGTACAAATTCAAG
AAATATTGGGGGGTTTGAATCAACACATTTAGATGATTAAATTAATATGAGGACTGAGGAATAAGAAATGAGTCAAG
GATGGTTCCAGGCTGCTAGGCTGCTTACCTGAGGTGGCAAAGTCGGGAGGAGTGGCAGTTTAGGACAGGGGGCAGTT
GAGGAATATTGTTTTGATCATTTTGAGTTTGAGGTACAAGTTGGACACTTAGGTAAAGACTGGAGGGGAAATCTGAA
TATACAATTATGGGACTGAGGAACAAGTTTATTTTATTTTTGTTTCGTTTTCTTGTTGAAGAACAAATTTAATTGT
AATCCCAAGTCATCAGCATCTAGAAGACAGTGGCAGGAGGTGACTGTCTTGTTGGGTAAGGGTTTGGGGTCCTTGATG
AGTATCTCTCAATTGGCCTTAAATATAAGCAGGAAAAAGGAGTTTATGATGGATTCCAGGCTCAGCAGGGCTCAGGAG
GGCTCAGGCAGCCAGCAGAGGAAGTCAGAGCATCTTCTTTGGTTTAGCCCAAGTAATGACTTCCTTAAAAAGCTGAA
GGAAAATCCAGAGTGACCAGATTATAAACTGTACTCTTGCATTTTCTCTCCCTCCTCTACCCACAGCCTCTTGATG
AACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCCGCTGGGCTAAGGGTCGGCGTGAGACCTACCTGTGCTACGT
AGTGAAGAGGCGTGACAGTGCTACATCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGGTATCAATTAAAGTCG
GCTTTGCAAGCAGTTTAATGGTCAACTGTGAGTGCTTTAGAGCCACCTGCTGATGGTATTACTTCCATCCTTTTTT
GGCATTGTGTCTCTATCACATTCCTCAAATCCTTTTTTTTATTTCTTTTTCCATGTCCATGCACCCATATTAGACA
TGGCCCAAATATGTGATTTAATTCCTCCCCAGTAATGCTGGGCACCCTAATACCACTCCTTCCTTCAGTGCCAAGA

ACAACCTGCTCCCAAACTGTTTACCAGCTTTCCTCAGCATCTGAATTGCCTTTGAGATTAATTAAGCTAAAAGCATTT
TTATATGGGAGAATATTATCAGCTTGTCCAAGCAAAAATTTTAAATGTGAAAAACAAATTGTGTCTTAAGCATTTTT
GAAAATTAAGGAAGAAGAATTTGGGAAAAAATTAACGGTGGCTCAATTCTGTCTTCCAAATGATTTCTTTTCCCTCC
TACTCACATGGGTCGTAGGCCAGTGAATACATTCAACATGGTGATCCCCAGAAAACCTCAGAGAAGCCTCGGCTGATG
ATTAATTAATGATCTTTTCGGCTACCCGAGAGAATTACATTTCCAAGAGACTTCTTCACCAAAATCCAGATGGGTT
TACATAAACTTCTGCCCACGGGTATCTCCTCTCTCCTAACACGCTGTGACGTCTGGGCTTGGTGGAATCTCAGGGAA
GCATCCGTGGGGTGAAGGTCATCGTCTGGCTCGTTGTTTGATGGTTATATTACCATGCAATTTTCTTTGCCTACAT
TTGTATTGAATACATCCCAATCTCCTTCCTATTCGGTGACATGACACATTCTATTTTCTCAGAAGGCTTTGATTTTATCA
AGCACTTTTCATTTACTTCTCATGGCAGTGCCTATTACTTCTCTTACAATACCCATCTGTCTGCTTTACCAAAATCTA
TTTCCCTTTTTCAGATCCTCCCAATGGTCTCATAACTGTCTGCCTCCACCTAGTGGTCCAGGTATATTTCCAC
AATGTTACATCAACAGGCATTCTAGCCATTTTCTTCTCAAAAGGTGCAAAAAGCAACTTCATAAACACAAATTA
ATCTTCGGTGAGGTAGTGTGATGCTGCTTCTCCTCCCAACTCAGCGCACTTCGTCTTCTCATTCCACAAAAACCCATA
GCCTTCTCTTCACTCTGCAGGACTAGTGCTGCCAAGGGTTCAGCTCTACCTACTGGTGTGCTCTTTTGAGCAAGTTGC
TTAGCCTCTCTGTAACACAAGGACAATAGCTGCAAGCATCCCCAAAGATCATTGCAGGAGACAATGACTAAGGCTAC
CAGAGCCGCAATAAAAGTCAGTGAATTTTAGCGTGGTCTCTCTGTCTCTCCAGAACGGCTGCCACGTGGAATTGCT
CTTCTCCGCTACATCTCGGACTGGGACCTAGACCCTGGCCGCTGCTACCGCGTCACCTGGTTCACCTCCTGGAGCC
CCTGCTACGACTGTGCCCAGCATGTGGCCGACTTTCTGCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCG
CGCCTCTACTTCTGTGAGGACCGCAAGGCTGAGCCCGAGGGGCTGCGGCGCTGCACCGCGCCGGGGTGCAAATAGC
CATCATGACCTTCAAAGGTGCGAAAGGGCCTTCCGCGCAGGCGCAGTGCAGCAGCCGCATTTCGGGATTGCGATGCG
GAATGAATGAGTTAGTGGGGAAGCTCGAGGGGAAGAAGTGGGCGGGGATTCTGGTTCACCTCTGGAGCCGAAATTA
AGATTAGAAGCAGAGAAAAGAGTGAATGGCTCAGAGACAAGGCCCGAGGAAATGAGAAAATGGGGCCAGGGTTGCT
TCTTTCCCTCGATTTGGAACCTGAACTGTCTTCTACCCCCATATCCCCGCCTTTTTTTTCTTTTTTTTTTTTGA
GATTATTTTTTACTGCTGGAATACTTTTGTAGAAAACACGAAAGAACTTTCAAAGCCTGGGAAGGGCTGCATGAAAA
TTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCCTTTTGGTAAGGGGCTTCTCGCTTTTAAATTTTCTTTCTTT
CTCTACAGTCTTTTTTGGAGTTTCGTATATTTCTTATATTTTCTTATTGTTCAATCACTCTCAGTTTTCATCTGATG
AAAACCTTTATTTCTCCTCCACATCAGCTTTTTCTTCTGCTGTTTCACCATTCAGAGCCCTCTGCTAAGGTTCTTTT
CCCTCCCTTTTCTTTCTTTTGTGTTTCACATCTTTAAATTTCTGTCTCTCCCCAGGGTTGCGTTTCTTCTGCTC
AGAATTCTTTTCTCCTTTTTTTTTTTTTTTTTTTTTTTTTTTTAAACAAACAAACAAAAACCCAAAAAACTCTTTC
CCAATTTACTTTCTTCCAACATGTTACAAAGCCATCCACTCAGTTTAGAAGACTCTCCGGCCCCACCGACCCCAAC
CTCGTTTTGAAGCCATTCACTCAATTTGCTTCTCTCTTTCTCTACAGCCCCTGTATGAGGTTGATGACTTACGAGAC
GCATTTCTGACTTTGGGACTTTGATAGCAACTTCCAGGAATGTCACACACGATGAAATATCTCTGCTGAAGACAGTG
GATAAAAAACAGTCTTCAAGTCTTCTCTGTTTTTATTCTTCAACTCTCACTTTCTTAGAGTTTACAGAAAAAATAT
TTATATACGACTCTTTAAAAAGATCTATGTCTTGAAAAATAGAGAAGGAACACAGGTCTGGCCAGGGACGTGCTGCAA
TTGGTGACGTTTTGAATGCAACATTGTCCCTACTGGGAATAACAGAACTGCAGGACCTGGGAGCATCCTAAAGTGT
CAACGTTTTTCTATGACTTTTAGGTAGGATGAGAGCAGAAGGTAGATCCTAAAAAGCATGGTGAGAGGATCAAATGT
TTTTATATCAACATCCTTTATTATTTGATTCATTTGAGTTAACAGTGGTGTTAGTGATAGATTTTTTCTATTCTTTTC
CCTTGACGTTTACTTTCAAGTAACACAACTCTTCCATCAGGCCATGATCTATAGGACCTCCTAATGAGAGTATCTG
GGTGATTGTGACCCCAACCATCTCTCCAAAGCATTAATATCCAATCATGCGCTGTATGTTTTAATCAGCAGAAGCA
TGTTTTTATGTTTGTACAAAAGAAGATTGTTATGGGTGGGGATGGAGGTATAGACCATGCATGGTCACCTTCAAGCT

ACTTTAATAAAGGATCTTAAAATGGGCAGGAGGACTGTGAACAAGACACCCTAATAATGGGTTGATGTCTGAAGTAG
CAAATCTTCTGGAAACGCAAACTCTTTTAAGGAAGTCCCTAATTTAGAAACACCCACAACTTCACATATCATAATT
AGCAAACAATTGGAAGGAAGTTGCTTGAATGTTGGGGAGAGGAAAATCTATTGGCTCTCGTGGGTCTCTTCATCTCA
GAAATGCCAATCAGGTCAAGGTTTGCTACATTTTGTATGTGTGTGATGCTTCTCCCAAAGGTATATTAACATATAA
GAGAGTTGTGACAAAACAGAATGATAAAGCTGCGAACCGTGGCACACGCTCATAGTTCTAGCTGCTTGGGAGGTTGA
GGAGGGAGGATGGCTTGAACACAGGTGTTCAAGGCCAGCCTGGGCAACATAACAAGATCCTGTCTCTCAAAAAAAAA
AAAAAAAAAAGAAAGAGAGAGAGGGCCGGGCGTGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGCCGG
GCGGATCACCTGTGGTCAGGAGTTTGAGACCAGCCTGGCCAACATGGCAAAACCCCGTCTGTACTCAAAATGCAAAA
ATTAGCCAGGCGTGGTAGCAGGCACCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCAG
GAGGTGGAGGTTGCAGTAAGCTGAGATCGTGCCGTTGCACTCCAGCCTGGGCGACAAGAGCAAGACTCTGTCTCAGA
AAAAAAAAAAAAAAAAAGAGAGAGAGAGAGAAAAGAGAACAATATTTGGGAGAGAAGGATGGGGAAGCATTGCAAGGAAA
TTGTGCTTTATCCAACAAAATGTAAGGAGCCAATAAGGGATCCCTATTTGTCTCTTTTGGTGTCTATTTGTCCCTAA
CAACTGTCTTTGACAGTGAGAAAAATATTCAGAATAACCATATCCCTGTGCCGTTATTACCTAGCAACCCCTTGAAT
GAAGATGAGCAGATCCACAGGAAAACCTGAATGCACAACTGTCTTATTTTAATCTTATTGTACATAAGTTTGTAAAA
GAGTTAAAAATTGTTACTTCATGTATTCAATTTATATTTTATATTTTTCGCTCTAATGATTTTTTTATTAACATGAT
TTCCTTTTCTGATATATTGAAATGGAGTCTCAAAGCTTCATAAATTTATAACTTTAGAAATGATTCTAATAACAACG
TATGTAATTGTAACATTGCAGTAATGGTGCTACGAAGCCATTCTCTTGATTTTTTAGTAACTTTTATGACAGCAAA
TTTGCTTCTGGCTCACTTTCAATCAGTTAAATAAATGATAAATAATTTTGAAGCTGTGAAGATAAAATACCAAATA
AAATAATATAAAAGTGATTTATATGAAGTTAAAATAAAAAATCAGTATGATGGAATAAACTTG

[0420] 以下提供了根据本申请的方面可以与Cas9融合的其他示例性脱氨酶。应当理解，
在一些实施方案中，可以使用相应序列的活性结构域，例如没有定位信号的结构域（核定位
序列、无核输出信号、胞质定位信号）。

[0421] 人类AID：

MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATS
FSLDFGYLRNKNKGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSP
[0422] CYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKA EPEGLRRLHRA
GVQIAIMTFKDYFYCWNTFVENHERTFKAW EGLHENS VRLSRQLR
RILLPLYEVDDLRLDAFRTLGL

[0423] (底线：核定位序列；双底线：核输出信号)

[0424] 小鼠AID：

MDSLLMKQKKFLYHFKNVRWAKGRHETLYCYVVKRRDSATS
CSLDFGHLRNKSGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSP
[0425] CYDCARHVAEFLRWPNPNLSLRIFTARLYFCEDRKA EPEGLRRLHRA
GVQIGIMTFKDYFYCWNTFVENRERTFKAW EGLHENS VRLTRQLR
RILLPLYEVDDLRLDAFRMLGF

[0426] (底线：核定位序列；双底线：核输出信号)

[0427] 犬类AID：

[0428] MDSLLMKQRKFLYHFKNVRWAKGRHETYLCYVVKRRDSATS
FSLDFGHLRNKSGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPC
YDCARHVADFLRGYPNLSLRIFAARLYFCEDRKAEPEGLRRLHRAG
VQIAIMTFKDYFYCWNTFVENREKTFKAWEGLHENSVRLSRQLRRI
LLPLYEVDDLRLDAFRTLGL

[0429] (底线:核定位序列;双底线:核输出信号)

[0430] 牛类AID:

MDSLLKKQRQFLYQFKNVRWAKGRHETYLCYVVKRRDSPTS
FSLDFGHLRNKAGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSP
[0431] CYDCARHVADFLRGYPNLSLRIFTARLYFCDKERKAEPEGLRRLHR
AGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQL
RRILLPLYEVDDLRLDAFRTLGL

[0432] (底线:核定位序列;双底线:核输出信号)

[0433] 大鼠AID:

MAVGSKPKAALVGPHWERERIWCFLCSTGLGTQQTGQTSRW
LRPAATQDPVSPPRSLLMKQRKFLYHFKNVRWAKGRHETYLCYVV
KRRDSATSFSLDFGYLRNKSGCHVELLFLRYISDWDLDPGRCYRVT
[0434] WFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLTGWGALPAGLM
SPARPSDYFYCWNTFVENHERTFKAWEGLHENSVRLSRRLRRILLP
LYEVDDLRLDAFRTLGL

[0435] (底线:核定位序列;双底线:核输出信号)

[0436] 小鼠APOBEC-3

MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLGYAKGRKDTF
LCYEVTRKDCDSPVSLHHGVFKNKDNIHAEICFLYWFHDKVLKVLSP
REEFKITWYMSWSPCFECAEQIVRFLATHHNLSLDIFSSRLYNVQDPE
TQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWKR
LLTNFRYQDSKLQEILRPCYIPVPSSSSSTLSNICLTKGLPETRFCVEG
[0437] RRMDPLSEEFYSQFYNQVRVKHLCYYHRMKPYLCYQLEQFNGQAP
LKGCLLSEKGKQHAEILFLDKIRSMELSQVTITCYLTWSPCPNCAWQL
AAFKRDRPDLILHIYTSRLYFHWKRPFQKGLCSLWQSGILVDVMDL
PQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRRLRIKESWGLQDL
VNDFGNLQLGPPMS

[0438] (斜体:核酸编辑域)

[0439] 大鼠APOBEC-3:

- [0440] MGPFC LGCSHRKCYSPIRNLISQETFKFHFKNRLRYAIDRKDTF
LCYEVTRKDCDSPVSLHHGVFKNKDNIHAEICFLYWFHDKVLKVLSP
REEFKITWYMSWSPCFECAEQVLRFLATHHNLSLDIFSSRLYNIRDPE
NQNLCLRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWKK
LLTNFRYQDSKLQEILRPCYIPVPSSSSSTLSNICLTKGLPETRFCVER
RRVHLLSEEEFY SQFY NQRVKHLCYYHGVKPYLCYQLEQFNGQAP
LKGCLLSEK GKQHAEILFLDKIRSMELSQVIITCYLTWSPCPNCAWQL
AAFKRDRPDLILHIYTSRLYFHWKRPFQKGLCSLWQSGILVDVMDL
PQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRR LHRIKESWGLQDL
VNDFGNLQLGPPMS
- [0441] (斜体:核酸编辑域)
- [0442] 猕猴APOBEC-3G:
- [0443] MVEPMDPRTFVS NFNNRPILSGLNTVWLCCEVKTKDPSGPPLD
AKIFQGKVYSKAKYHPEMRFLRWFHKWRQLHHDQEYKVTWYVSW
PCTRCANSVATFLAKDPKVTLTIFVARLYYFWKPDYQQALRILCQK
RGGPHATMKIMNYNEFQDCWNKFVDGRGKPFKPRNNLPKH YTL
QATLGELLRHLMDPGTFTSNFNKPPWVSGQHET YLCYKVERLHND
[0444] TWVPLNQHRGFLRNQAPNIHGFPKGRHAELCFLDLIPFWKLDGQQY
RVTCFTSWSPCFSCAQEMAKFISNNEHVS LCIFAARIYDDQGRYQEG
LRALHRDGAKIAMMNYSEFEYCWDTFVDRQGRPFQPWDGLDEHS
QALSGRLRAI
- [0445] (斜体:核酸编辑域;底线:胞质定位信号)
- [0446] 黑猩猩APOBEC-3G:
- [0447] MKPHFRNPVERMYQDTFSDNFYNRPILSHRNTVWLCYEVKTK
GPSRPPLDAKIFRGQVYSKLKYHPEMRFFHWFSKWRKLHRDQEYEV
TWYISWSPCTKCTR DVATFLAEDPKVTLTIFVARLYYFWDPDYQEAL
RSLCQKRDGP RATMKIMNYDEFQHCWSKFVYSQRELFEPWNNLPK
[0447] YYILLHIMLGEILRHSM DPPTFTSNFNNELWVRGRHET YLCYEVER
LHN DTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKL
DLHQDYRVTCFTSWSPCFSCAQEMAKFISNNKHVS LCIFAARIYDDQ
GRCQEGLRTLAKAGAKISIMTYSEFKHCWDTFVDHQGCPFPWDG
LEEHSQALSGRLRAILQNQGN
- [0448] (斜体:核酸编辑域;底线:胞质定位信号)
- [0449] 绿猴APOBEC-3G:

MNPQIRNMVEQMEPDIFVYYFNNRPILSGRNTVWLCYEVKTK
DPSGPPLDANIFQGKLYPEAKDHP*EMKFLHWFRKWRQLHRDQEYEV*
TWYVSWSPCTRCANSVATFLAEDPKVTLTIFVARLYYFWKPDYQQA
LRILCQERGGPHATMKIMNYNEFQHCWNEFVDGQGKPFKPRKNLP
[0450] KHYTLLHATLGELLRHVMDPGTFTSNFNNKPWVSGQRETYLCYKV
ERSHNDTWVLLNQHRGFLRNQAPDRHGFPGRHAELCFLDLIPFWK
LDDQQYRVTCFTSWSPCFSCAQKMAKFISNNKHVSLCIFAARIYDDQ
GRCQEGLRTLHRDGAKIAVMNYSEFEYCWDTFVDRQGRPFQPWD
GLDEHSQALSGRLRAI

[0451] (斜体:核酸编辑域;底线:胞质定位信号)

[0452] 人类APOBEC-3G:

MKPHFRNTVERMYRDTFSYNFYNNRPILSRRNTVWLCYEVKTK
[0453] GPSRPPLDAKIFRGQVYSELKYHPEMRFFHWFSKWRKLHRDQEYEV
*TWYISWSPCTKCTRD*MATFLAEDPKVTLTIFVARLYYFWDPDYQEA
LRSLCQKRDGPRAATMKIMNYDEFQHCWSKFVYSQRELFEPWNNLP
KYYILLHIMLGEILRHSM DPPTFTFNFNNEPWVRGRHETLYCYEVE
[0454] RMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWK
LDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNNKHVSLCIFTARIYDD
QGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVDHQGC PFQPWD
GLDEHSQDLSGRLRAILQNQEN

[0455] (斜体:核酸编辑域;底线:胞质定位信号)

[0456] 人类APOBEC-3F:

MKPHFRNTVERMYRDTFSYNFYNNRPILSRRNTVWLCYEVKTK
GPSRPRLDAKIFRGQVYSQPEHHAEMCFLSWFCGNQLPAYKCFQITW
FVSWTPCPDCVAKLAEFLAHPNVTLTISAARLYYYWERDYRRALC
RLSQAGARVKIMDDEEFAYCWENFVYSEGQPFMPWYKFDDNYAF
[0457] LHRTLKEILRNPMEAMYPHIFYFHFKNLRKAYGRNESWLCFTMEV
VKHHSPVSWKRGVFRNQVDPETHCHAERCFLSWFCDDILSPNTNYE
VTWYTSWSPCPECAGEVAEF
LARHSNVNLTIFTARLYYFWDDTDYQEGLRSLSQEGASVEIMGYKDF
KYCWENFVYNDDEPFKPWKGLKYNFLFLDSKLQEILE

[0458] (斜体:核酸编辑域)

[0459] 人类APOBEC-3B:

MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIK
RGRSNLLWDTGVFRGQVYFKPQYHAEMCFLSWFCGNQLPAYKCFQI
TWFVSWTPCPDCVAKLAEFLSEHPNVTLTISAARLYYYWERDYRRA
LCRLSQAGARVTIMDYEEFAYCWENFVYNEGQQFMPWYKFDENY
[0460] AFLHRTLKEILRYLMDPDTFTFNFNNDPLVLRRTQTYLCYEVERLD
NGTWVLMQHMGLFCNEAKNLLCGFYGRHAELRFLDLVPSLQLDP
AQIYRVTFISWSPCFSWGCAGEVRAFLQENTHVRLRIFAARIYDYD
PLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVYRQGCPFPQWD
GLEEHSQALSGRLRAILQNQGN

[0461] (斜体:核酸编辑域)

[0462] 大鼠APOBEC-3B:

MQPQGLGPNAGMGPVCLGCSHRRPYSPIRNPLKKLYQQTFYF
[0463] HFKNVRYAWGRKNNFLCYEVNGMDCALPVPLRQGVFRKQGHIHA
ELCFIYWFHDKVLRVLSPMEEFKVTWYMSWSPCSKCAEQVARFLA
AHRNLSLAIFSSRLYYYLRNPNYQQKLCRLIQEGVHVAAMDLPFEK
KCWNKFVDNDGQPFRPWMRLRINFSDYDCKLQEIFSRMNLLREDV
FYLQFNNSHRVKPVQNRYRRKSYLCYQLERANGQEPLKGYLLYK
[0464] KGEQHVEILFLEKMRSMELSQVRITCYLTWSPCPNCARQLAAFKKD
HPDLILRIYTSRLYFWRKKFQKGLCTLWRSGIHVDVMDLPQFADC
WTNFVNPQRPFPPWNELEKNSWRIQRRLRRIKESWGL

[0465] 牛类APOBEC-3B:

[0466] DGWEVAFRSGTVLKAGVLGVSMTEGWAGSGHPGQGACVWTPGTRNTMNLREVLFKQQFGNQPRVPAPY
YRRKTYLCYQLKQRNDLTLDRCFRNKKQRHAERFIDKINSLDLNPSQSYKIIICYITWSPCPNCANELVNFITRNH
LKLEIFASRLYFHWIKSFKMGLQDLQNAGISVAVMTHTEFEDCWEQFVDNQSRPFQPWDKLEQYSASIRRLQRILT
API

[0467] 黑猩猩APOBEC-3B:

[0468] MNPQIRNPMEMYQRTFYNFENEPILYGRSYTWLCYEVKIRRGHSNLLWDTGVFRGQMYSQPEHHAEM
CFLSWFCGNQLSAYKCFQITWFVSWTPCPDCVAKLAKFLAEHPNVTLTISAARLYYYWERDYRRALCRLSQAGARVK
IMDDEEFAYCWENFVYNEGQPFMPWYKFDNYAFLHRTLKEIIRHLMDPDTFTFNFNNDPLVLRHQTYLCYEVERL
DNGTWVLMQHMGLFCNEAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTFISWSPCFSWGCAGQVRAFLQEN
THVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVYRQGCPFPQWDGLEEHSQALSGRLRAIL
QVRASSLCMVHRPPPPQSPGPCLPLCSEPPLGSLPTGRPAPSLPFLLTASFSPPPASLPPLPSLSLSPGHLPV
PSFHSLTSCSIQPPCSSRIRETGWASVSKEGRDLG

[0469] 人类APOBEC-3C:

- MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGI
KRRSVVSWKTGVFRNQVDSETH*CHAERCFLSWFCDDILSPNTKYQV*
[0470] *TWYTSWSPCPDC*AGEVAEFLARHSNVNLTIFTARLYYFQYPCYQEG
LRSLSQEGVAVEIMDYEDFKYCWENFVYNDNEPFPKPKWKGLKTNFR
LLKRRLRESLQ
- [0471] (斜体:核酸编辑域)
- [0472] 大猩猩APOBEC-3C
- [0473] MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGI
KRRSVVSWKTGVFRNQVDSETH*CHAERCFLSWECDILSPNTNYQV*
[0474] *TWYTSWSPCPEC*AGEVAEFLARHSNVNLTIFTARLYYFQDQTDYQEG
LRSLSQEGVAVKIMDYKDFKYCWENFVYNDDEPFPKPKWKGLKYNF
RFLKRRLQEILE
- [0475] 人类APOBEC-3A:
MEASPASGPRHLMDPHIFTSNFNNGIGRHKTYLCYEVEERLDNG
TSVKMDQHRGFLHNQAKNLLCGFYGR*HAELRFLDLVPSLQLDPAQI*
[0476] *YRVTFWISWSPCF*SWGCAGEVRAFLQENTHVRLRIFAARIYDYDPLY
KEALQMLRDAGAQVSIMTYDEFKHCWDTFVDHQGCPFPQWDGLD
EHSQALSGRLRAILQNQGN
- [0477] (斜体:核酸编辑域)
- [0478] 猕猴APOBEC-3A:
MDGSPASRPRHLMDPNTFTFNFNNDLSVRGRHQTYLCYEVEER
LDNGTWVPMDERRGFLCNKAKNVPCGDYGC*HVELRFLCEVPSWQL*
[0479] *DPAQTYRVTFWISWSPCF*RRCAGQVRVFLQENKHVRLRIFAARIYD
YDPLYQEALRTL RDAGAQVSIMTYEEFKHCWDTFVD RQGRPFQPW
DGLDEHSQALSGRLRAILQNQGN
- [0480] (斜体:核酸编辑域)
- [0481] 牛类APOBEC-3A:
MDEYTFTENFNNQGWPSKTYLCYEMERLDGDATIPLDEYKGF
VRNKGLDQPEKPC*HAELYFLGKIHSWNLDNRNQHYRLTCFISWSPCYD*
[0482] *CAQKLTTFLKENHHISLHILASRIYTHNRFGCHQSGLC*ELQAAGARI
TIMTFEDFKHCWETFVDHKGKPFQPWEGLNVKSQALCTELQAILKT
QQN
- [0483] (斜体:核酸编辑域)
- [0484] 人类APOBEC-3H:

MALLTAETFRLQFNNKRRRLRRPYYPKALLCYQLTPQNGSTPT
RGYFENKKKCHAEICFINEIKSMGLDETQCYQVTCYLTWSPCSSCAWE
[0485] LVDFIKAHDHLNLGIFASRLYYHWCKPQQKGLRLLCGSQVPVEVM
GFPKFADCWENFVDHEKPLSFNPYKMLEELDKNSRAIKRRLERIKIP
GVRAQGRYMDILCDAEV

[0486] (斜体:核酸编辑域)

[0487] 猕猴APOBEC-3H:

[0488] MALLTAKTFSLQFNNKRRVKNPYYPKALLCYQLTPQNGSTPTRGHLKNKKKDHAEIFINKIKSMGLD
ETQCYQVTCYLTWSPCPCAGELVDFIKAHRHLNLRIFASRLYYHWRPNYQEGLLLLCGSQVPVEVMGLPEFTDCWE
NFVDHKEPPSFNPSEKLEELDKNSQAIKRRLERIKSRSDVLENGLRSLQLGPVTPSSSIRNSR

[0489] 人类APOBEC-3D:

MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIK
RGRSNLLWDTGVFRGPVLPKRQSNHRQEVYFRFENHAEMCFLSWF
CGNRLPANRRFQITWFWVSWNPCLPCVVKVTKFLAEHPNVTLTISAAR
LYYYRDRDWRWVLLRLHKAGARVKIMDYEDFAYCWENFVCNEG
[0490] QPFMPWYKFDDNYASLHRTLKEILRNPMEAMYPHIFYFHFKNLLK
ACGRNESWLCFTMEVTKHHSVFRKRGVFRNQVDPETHCHAERCF
LSWFCDDILSPNTNYEVTWYTSWSPCPECAGEVAEFLARHSNVNLTIF
TARLCYFWDTDYQEGLCSLSQEGASVKIMGYKDFVSCWKNFVYS
DDEPFKPWKGLQTNFRLLKRRRLREILQ

[0491] (斜体:核酸编辑域)

[0492] 人类APOBEC-1:

[0493] MTSEKGPSTGDPTLRRRIEPWEFDVFDPRELRKEACLLYEIKWGMSRKIWRSSGKNTTNHVEVNFIEK
FTSERDFHPSMSCSITWFLSWSPCWECSQAIREFLSRHPGVTLVIIYVARLFWHMDQQNRQGLRDLVNSGVTIQIMRA
SEYYHCWRNFVNYPPGDEAHPQYPPLWMLYALELHCIIISLPPCLKISRWNHLTFFRLHLQNCHYQTIPPHIL
LATGLIHPSVAWR

[0494] 小鼠APOBEC-1:

[0495] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKRETCCLYEINWGGRHSVWRHTSQNTSNHVEVNFLEK
FTTERYFRPNTRCSITWFLSWSPCGECSRAITEFLSRHPVTLFIYIARLYHHTDQRNRQGLRDLISSGVTIQIMTE
QEYCYCWRNFVNYPPSNEAYWPRYPHLWVKLYVLELYCIIILGLPPCLKILRRKQPQLTFFTITLQTCHYQRIPPHLL
WATGLK

[0496] 大鼠APOBEC-1:

[0497] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKRETCCLYEINWGGRHSIWRHTSQNTNKHVEVNFIEK
FTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTE
QESGYCWRNFVNYSPSNEAHPRYPHLWVRLYVLELYCIIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHIL
WATGLK

[0498] 人类APOBEC-2:

[0499] MAQKEEAAVATEAASQNGEDLENLDDPEKLKELIELPPFEIVTGERLPANFFKFQFRNVEYSSGRNKTF
LCYVVEAQSGGGQVQASRGYLEDEHAAAHAEEAFFNTILPAFDPALRYNVTWYVSSSPCAACADRIIKTLSKTKNLR
LLILVGRLFMWEEPEIQAALKKLKEAGCKLRIMKPQDFEYVWQNFVEQEEGESKAFQPWEDIQENFLYYEEKLADIL
K

[0500] 小鼠APOBEC-2:

[0501] MAQKEEAAEAAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFRNVEYSSGRNK
FLCYVVEVQSKGGQVQATQGYLEDEHAGAHAAEEAFFNTILPAFDPALKYNVTWYVSSSPCAACADRILKTLSTKN
LRLILVSRILFMWEEPEVQAALKKLKEAGCKLRIMKPQDFEYIWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLA
DILK大鼠APOBEC-2:

[0502] MAQKEEAAEAAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFRNVEYSSGRNK
FLCYVVEAQSKGGQVQATQGYLEDEHAGAHAAEEAFFNTILPAFDPALKYNVTWYVSSSPCAACADRILKTLSTKN
LRLILVSRILFMWEEPEVQAALKKLKEAGCKLRIMKPQDFEYLWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLA
DILK牛类APOBEC-2:

[0503] MAQKEEAAAAAEPASQNGEEVENLEDPEKLKELIELPPFEIVTGERLPAHYFKFQFRNVEYSSGRNK
FLCYVVEAQSKGGQVQASRGYLEDEHATNHAEEAFFNSIMPTFDPALRYMVTWYVSSSPCAACADRIVKTLNKTKN
LRLILVGRILFMWEEPEIQAALRKLKEAGCRLRIMKPQDFEYIWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLA
DILK海七鳃鳗CDA1 (pmCDA1):

[0504] MTDAEYVRIHEKLDIYTFKKQFFNNKSVSHRCYVLFELKRRGERRACFWGYAVNK

[0505] PQSGTERGIHAEIFSIRKVEEYLRDNPQGFTINWYSSWSPCADCAEKILEWYNQELRG

[0506] NGHTLKIWACKLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQ

[0507] LNENRWLEKTLKRAEKRRSELSFMIQVKILHTTKSPAV

[0508] 人类APOBEC3G D316R D317R:

[0509] MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQ

[0510] VYSELKYHPEMRFFHWFWSKWRKLHRDQEYEVWYISWSPCTKCTRDMATFLAEDP

[0511] KVTLTIFVARLYYFWDPDYQALRSLCQKRDGPRATMKFNYDEFQHCWSKFVYSQ

[0512] RELFEPWNNLPKYIILLHFMLEILRHSMPTFTFNFNNEPWVRGRHETILCYEVER

[0513] MHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTC

[0514] FTSWSPCFSCAQEMAKFISKHVSLEIFTARIYRRQGRCQEGRLTLAEAGAKISFTYSEFKHCWDTFV
DHQGCPFPWDGLDEHSQDLSGRLRAILQNQEN人类APOBEC3G链A:

[0515] MDPPTFTFNFNNEPWVRGRHETILCYEVERMHNDTWVLLNQRRGFLCNQAPHKHG

[0516] FLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCI

[0517] FTARIYDDQGRCQEGRLTLAEAGAKISFTYSEFKHCWDTFVDHGGCPFPWDGLD
EHSQDLSGRLRAILQ

[0518] 人APOBEC3G链A D120R D121R:

[0519] MDPPTFTFNFNNEPWVRGRHETILCYEVERMHNDTWVLLNQRRGFLCNQAPHKHG

[0520] FLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCI

[0521] FTARIYRRQGRCQEGRLTLAEAGAKISFMTYSEFKHCWDTFVDHGGCPFPWDGLD
HSQDLSGRLRAILQ

[0522] 本申请的一些方面基于以下认识:调节本文所述的任何融合蛋白的脱氨酶结构域

催化活性,例如通过在脱氨酶结构域中进行点突变来影响融合蛋白的合成能力(例如,碱基编辑器)。例如,减少但不消除碱基编辑融合蛋白中脱氨酶结构域催化活性的突变可使脱氨酶结构域催化邻近靶残基的残基脱氨的可能性降低,从而缩小了脱氨窗口。缩小脱氨窗口的能力可以防止与特定目标残基相邻的残基发生不希望的脱氨基,这可以减少或防止脱靶效应。

[0523] 例如,在一些实施方案中,并入碱基编辑器中的APOBEC脱氨酶可包含选自rAPOBEC1的H121X、H122X、R126X、R126X、R118X、W90X、W90X和R132X的一个或多个突变,或另一个APOBEC脱氨酶中的一个或多个相应突变,其中X是任何氨基酸。在一些实施方案中,并入碱基编辑器中的APOBEC脱氨酶可包含一种或多种选自rAPOBEC1的H121R、H122R、R126A、R126E、R118A、W90A、W90Y和R132E的突变,或一种或多种相应的突变。另一个APOBEC脱氨酶。

[0524] 在一些实施方案中,并入碱基编辑器中的APOBEC脱氨酶可以包含一种或多种选自hAPOBEC3G的D316X、D317X、R320X、R320X、R313X、W285X、W285X、R326X或一个或多个对应的突变。另一个APOBEC脱氨酶中的突变,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,其包含选自hAPOBEC3G的D316R、D317R、R320A、R320E、R313A、W285A、W285Y、R326E的一个或多个突变,或一个或多个相应突变。在另一个APOBEC脱氨酶中。

[0525] 在一些实施方案中,并入碱基编辑器中的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,并入碱基编辑器中的APOBEC脱氨酶可包含APOBEC脱氨酶,其包含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,并入碱基编辑器中的APOBEC脱氨酶可包含APOBEC脱氨酶,其包含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,并入碱基编辑器中的APOBEC脱氨酶可包含APOBEC脱氨酶,其包含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,并入碱基编辑器中的APOBEC脱氨酶可包含APOBEC脱氨酶,其包含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,并入碱基编辑器中的APOBEC脱氨酶可包含APOBEC脱氨酶,其包含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,并入碱基编辑器中的APOBEC脱氨酶可包含APOBEC脱氨酶,其包含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,并入碱基编辑器中的APOBEC脱氨酶可包含APOBEC脱氨酶,其包含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,并入碱基编辑器中的APOBEC脱氨酶可以包含APOBEC脱氨酶,其包含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,并入碱基编辑器中的APOBEC脱氨酶可包含APOBEC脱氨酶,其包含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,并入碱基编辑器中的APOBEC脱氨酶可包含APOBEC脱氨酶,其包含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。

[0526] 在一些实施方案中,并入碱基编辑器中的APOBEC脱氨酶可包含APOBEC脱氨酶,其包含hAPOBEC3G的D316R和D317R突变,或另一APOBEC脱氨酶中的一个或多个相应突变。在一些实施方案中,本文提供的任何融合蛋白包含APOBEC脱氨酶,其包含hAPOBEC3G的R320A突

变,或另一APOBEC脱氨酶中的一个或多个相应突变。在一些实施方案中,并入碱基编辑器中的APOBEC脱氨酶可包含APOBEC脱氨酶,其包含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,并入碱基编辑器中的APOBEC脱氨酶可包含APOBEC脱氨酶,其包含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,并入碱基编辑器中的APOBEC脱氨酶可包含APOBEC脱氨酶,其包含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,并入碱基编辑器中的APOBEC脱氨酶可包含APOBEC脱氨酶,其包含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,并入碱基编辑器中的APOBEC脱氨酶可包含APOBEC脱氨酶,其包含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,并入碱基编辑器中的APOBEC脱氨酶可包含APOBEC脱氨酶,其包含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,并入碱基编辑器中的APOBEC脱氨酶可包含APOBEC脱氨酶,其包含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,并入碱基编辑器中的APOBEC脱氨酶可包含APOBEC脱氨酶,其包含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。

[0527] 许多修饰的胞苷脱氨酶可商购获得,包括但不限于SaBE3、SaKKH-BE3、VQR-BE3、EQR-BE3、VRER-BE3、YE1-BE3、EE-BE3、YE2-BE3、和YEE-BE3、可得自Addgene(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。

[0528] C到T核碱基编辑蛋白的详细信息在国际PCT第PCT/US2016/058344号专利申请(WO第2017/070632号国际专利申请)和Komor, A.C.等人中进行了描述,“Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage” Nature 533,420-424(2016),其全部内容通过引用结合于此。

[0529] A到G编辑

[0530] 在一些实施方案中,本文所述的碱基编辑器可包含脱氨酶结构域,其包括腺苷脱氨酶。碱基编辑器的这种腺苷脱氨酶结构域可以通过使A脱氨基形成肌苷(I)来促进腺嘌呤(A)核苷碱基向鸟嘌呤(G)核苷碱基的编辑,表现出G的碱基配对特性。腺苷脱氨酶能够脱氧核糖核酸(DNA)中的脱氧腺苷残基的腺嘌呤脱氨(即除去胺基)的方法。

[0531] 在一些实施方案中,可通过将一个或多个蛋白结构域融合在一起,从而产生融合蛋白来制备本文提供的核碱基编辑器。在某些实施方案中,本文提供的融合蛋白包含改善融合蛋白的碱基编辑活性(例如,效率、选择性和特异性)的一种或多种特征。例如,本文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,本文提供的融合蛋白可具有不具有核酸酶活性的Cas9结构域(dCas9)或切割双链DNA分子的一条链的Cas9结构域,称为Cas9切口酶(nCas9)。不希望受到任何特定理论的束缚,催化残基(例如,H840)的存在维持了Cas9切割含有与靶标A相对的T的未编辑(例如,未脱氨基)链的活性。Cas9的催化残基(例如,D10至A10)的“残基”阻止了含有目标A残基的编辑链的切割。此类Cas9变体能够基于gRNA定义的靶序列在特定位置产生单链DNA断裂(缺口),从而导致未

编辑链的修复,最终导致非编辑的子线。在一些实施方案中,A至G碱基编辑器还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)结构域或催化失活的肌苷特异性核酸酶。不希望受任何特定理论的束缚,UGI结构域或催化失活的肌苷特异性核酸酶可以抑制或阻止脱氨基腺苷残基(例如肌苷)的碱基切除修复,这可以提高碱基编辑器的活性或效率。

[0532] 包含腺苷脱氨酶的碱基编辑器可以作用于任何多核苷酸,包括DNA、RNA和DNA-RNA杂交体。在某些实施方案中,包含腺苷脱氨酶的碱基编辑器可以使包含RNA的多核苷酸的靶标A脱氨基。例如,碱基编辑器可以包含能够使RNA多核苷酸和/或DNA-RNA杂合多核苷酸的靶A脱氨基的腺苷脱氨酶结构域。在一个实施方案中,并入碱基编辑器中的腺苷脱氨酶包含作用于RNA(ADAR,例如,ADAR1或ADAR2)的全部或部分腺苷脱氨酶。在另一个实施方案中,掺入碱基编辑器中的腺苷脱氨酶包含作用于tRNA(ADAT)的全部或部分腺苷脱氨酶。包含腺苷脱氨酶结构域的碱基编辑器也能够使DNA多核苷酸的A核碱基脱氨基。在一个实施方案中,碱基编辑器的腺苷脱氨酶结构域包含ADAT的全部或部分,其包含一个或多个突变,所述突变允许ADAT将DNA中的靶标A脱氨基。例如,碱基编辑器可以包含来自大肠杆菌(EcTadA)的ADAT的全部或部分,其包含以下一个或多个突变:D108N、A106V、D147Y、E155V、L84F、H123Y、I157F或另一个突变腺苷脱氨酶。

[0533] 腺苷脱氨酶可以衍生自任何合适的生物(例如大肠杆菌)。在一些实施方案中,腺嘌呤脱氨酶是天然存在的腺苷脱氨酶,其包括与本文提供的任何突变相对应的一个或多个突变(例如,ecTadA中的突变)。任何同源蛋白质中的相应残基可通过例如序列比对和同源残基的测定来鉴定。可以相应地产生与本文所述的任何突变(例如,在ecTadA中鉴定的任何突变)相对应的任何天然存在的腺苷脱氨酶中的突变(例如,与ecTadA具有同源性)。

[0534] TadA

[0535] 在特定的实施方案中,TadA是PCT/US2017/045381(WO2018/027078)中描述的TadA中的任何一种,其通过引用整体并入本文。

[0536] 在一个实施方案中,本发明的融合蛋白包含与TadA7.10连接的野生型TadA,其与Cas9切口酶连接。在特定实施方案中,融合蛋白包含单个TadA7.10结构域(例如,以单体形式提供)。在其他实施方式中,ABE7.10编辑器包括能够形成异二聚体的TadA7.10和TadA(wt)。相关序列如下:

[0537] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRREQEIKAKKAQSSTD(称为“Ta”)。

[0538] TadA7.10:

[0539] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNRRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKAQSSTD

[0540] 在一些实施方案中,腺苷脱氨酶包含与本文提供的任何腺苷脱氨酶中列出的任何氨基酸序列至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相同的氨基酸序列。应当理解,本文提供的腺苷脱氨酶可包括一个或多个突变(例如,本文提供的任何突变)。本申请内

容提供了具有一定百分比同一性的任何脱氨酶结构域,加上本文所述的任何突变或其组合。在一些实施方案中,与参考序列或本文提供的任何腺苷脱氨酶相比,腺苷脱氨酶包含具有1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50或更多个突变的氨基酸序列。在一些实施方案中,与本领域已知或本文所述的任何氨基酸序列相比,腺苷脱氨酶包含具有至少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。

[0541] 在一些实施方案中,TadA脱氨酶是全长大肠杆菌TadA脱氨酶。例如,在某些实施方案中,腺苷脱氨酶包含氨基酸序列:

[0542] MRRAFITGVFFLSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAKKQASSTD。

[0543] 然而,应当理解,对本领域技术人员而言可用于本申请的另外的腺苷脱氨酶是显而易见的,并且在本申请的范围内。例如,腺苷脱氨酶可以是作用于tRNA (ADAT) 的腺苷脱氨酶的同源物。非限制性地,示例性AD AT同源物的氨基酸序列包括以下:

[0544] 金黄色葡萄球菌TadA:

[0545] MGSHTNDIYFMTLAIEEAKKAAQLGEVPIGAIITKDDEVIAHAHNLRETLQQPTAHAEHIAIERAAKVLGSRWLEGCTLYVTLEPCVMCAGTIVMSRIPRVVYGADDPKGGCSGLMNLQQSNFNHRAIVDKGVLKEACSTLLTTFKKNLRANKKSTN

[0546] 枯草芽孢杆菌 (*Bacillus subtilis*) TadA:

[0547] MTQDELYMKEAIKEAKKAEKGEVPIGAVLVINGEIIARAHNLRETEQRSIAHAEMLVIDEACKALGTWRLEGATLYVTLEPCPM CAGAVVLSRVEKVVFGAFDPKGGC SGTLMNLLQEERFNHQAEEVSGVLEECCGMLSAFFRELRRKKKAARKNLSE

[0548] 鼠伤寒沙门氏菌 (*Salmonella typhimurium*, *S.typhimurium*) TadA:

[0549] MPPAFITGVTSLSDELVDHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVLQNYRLDDTTLYVTLEPCVMCAGAMVHSRIGRVVFGARDAKTGAAGSLIDVLHHPGMNHRVEIIEGVL RDECATLLSDFFRMRRQEIKALKKADRAEGAGPAV

[0550] 腐败希瓦氏菌 (*Shewanella putrefaciens*, *S.putrefaciens*) TadA: MDEYWMQVAMQMAEKAEAAAGEVPVGAVLVKDGQQIATGYNLSISQHDPT

[0551] AHAEILCLRSAGKKLENYRLDDATLYITILEPCAMCAGAMVHSRIARVVYGARDEKTGAAGTVVNLLQHPAFNHQVEVTSGVLAEACSAQLSRFFKRRRDEKKALKLAQRAQQGIE

[0552] 流感嗜血杆菌F3031 (*Haemophilus influenzae*, *H.influenzae*) TadA:

[0553] MDAKVRSEFDEKMMRYALELADKAEALGEIPVGAVLVDDARNIIGEGWNLSIVQSDPTA HAEIIALRNGAKNIQNYRLNSTLYVTLEPCTMCAGAILHSRIKRLVFG

[0554] ASDYKTGAIGSRFHFFDDYKMNHTLEITSGVLAEESQKLSTFFQKRREEKKIEKALLKSLSDK

[0555] 新月形杆菌 (*Caulobacter crescentus*, *C.crescentus*) TadA:

[0556] MRTDESEDQDHMMRLALDAARAAAEEAGETPVGAVILDPSTGEVIATAGNGPIAAHDPTAHAEIAAMRAAAAKLGNRYRLDTLTLVVTLEPCAMCAGAI SHARIGRVVFGADD

[0557] PKGGAVVHGPKFFAQPTCHWRPEVTGGVLADESADLLRGFFRARRKAKI

[0558] 硫还原泥土杆菌(*Geobacter sulfurreducens*, *G. sulphurreducens*) TadA:

[0559] MSSLKKTPIRDDAYWMGKAIREAAKAAARDEVPIGAVIVRDGAVIGRHNLRGNSDPSAHAEMIAIRQ
AARRSANWRLTGATLYVTLEPCLMCMGAIILARLERVVFVGCYDPKGGGAAGSLYDLSADPRLNHQVRLSPGVCQEECG
TMLSDFRDLRRRKAKATPALFIDERKVPPEP

[0560] 大肠杆菌TadA(ecTadA)的实施例包括以下内容:

[0561] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNRHVEITEGILADECAALLC
YFFRMPRQVFNAQKKAQSSTD

[0562] 在一些实施方案中,腺苷脱氨酶来自原核生物。在一些实施方案中,腺苷脱氨酶来自细菌。在一些实施方案中,腺苷脱氨酶来自大肠杆菌、金黄色葡萄球菌、伤寒沙门氏菌、腐烂希瓦氏菌、流感嗜血杆菌、新月形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。

[0563] 在一个实施方案中,本发明的融合蛋白包含与TadA7.10连接的野生型TadA,其与Cas9切口酶连接。在特定实施方案中,融合蛋白包含单个TadA7.10结构域(例如,以单体形式提供)。在其他实施方式中,ABE7.10编辑器包括能够形成异二聚体的TadA7.10和TadA(wt)。

[0564] 在一些实施方案中,腺苷脱氨酶包含与本文提供的任何腺苷脱氨酶中列出的任何氨基酸序列至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相同的氨基酸序列。应当理解,本文提供的腺苷脱氨酶可包括一个或多个突变(例如,本文提供的任何突变)。本申请内容提供了具有一定百分比同一性的任何脱氨酶结构域,加上本文所述的任何突变或其组合。在一些实施方案中,腺苷脱氨酶包含与参考序列或本文提供的任何腺苷脱氨酶相比具有1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50或更多个突变的氨基酸序列。在一些实施方案中,与本领域已知或本文所述的任何氨基酸序列相比,所述腺苷脱氨酶包含具有至少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个相同连续氨基酸残基的氨基酸序列。

[0565] 应当理解,可以将本文提供的任何突变(例如,基于TadA参考序列)引入其他腺苷脱氨酶(如大肠杆菌TadA(ecTadA))、金黄色葡萄球菌TadA(saTadA)或其他腺苷脱氨酶(例如细菌腺苷脱氨酶)。对于本领域技术人员显而易见的是,可以类似地比对额外的脱氨酶以鉴定可以如本文所提供的突变的同源氨基酸残基。因此,可以在具有同源氨基酸残基的其他腺苷脱氨酶(例如ecTadA)中进行在TadA参考序列中鉴定的任何突变。还应当理解,本文提供的任何突变可以在TadA参考序列或另一种腺苷脱氨酶中单独或以任何组合进行。

[0566] 在一些实施方案中,腺苷脱氨酶在TadA参考序列中包含D108X突变,或在另一个腺苷脱氨酶(例如ecTadA)中包含相应的突变,其中X表示野生型中相应氨基酸以外的任何氨基酸。腺苷脱氨酶。在一些实施方案中,腺苷脱氨酶包含D108G、D108N、D108V、D108A或D108Y突变,或另一腺苷脱氨酶中的相应突变。

[0567] 在一些实施方案中,腺苷脱氨酶在TadA参考序列中包含A106X突变,或在另一个腺苷脱氨酶(例如ecTadA)中的相应突变,其中X表示野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶在TadA参考序列中包含A106V突变,或在另一腺苷脱氨酶(例如野生型TadA或ecTadA)中包含相应的突变。

[0568] 在一些实施方案中,腺苷脱氨酶在TadA参考序列中包含E155X突变,或在另一个腺苷脱氨酶(例如ecTadA)中的相应突变,其中X的存在表示野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶在TadA参考序列中包含E155D、E155G或E155V突变,或在另一腺苷脱氨酶(例如ecTadA)中的相应突变。

[0569] 在一些实施方案中,腺苷脱氨酶在TadA参考序列中包含D147X突变,或在另一个腺苷脱氨酶(例如ecTadA)中的相应突变,其中X的存在指示野生物中相应氨基酸以外的任何氨基酸。型腺苷脱氨酶。在一些实施方案中,腺苷脱氨酶包含D147Y、TadA参考序列中的突变或另一种腺苷脱氨酶(例如ecTadA)中的相应突变。

[0570] 在一些实施方案中,腺苷脱氨酶在TadA参考序列中包含A106X、E155X或D147X突变,或在另一个腺苷脱氨酶中的相应突变(例如ecTadA),其中X表示野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含E155D、E155G或E155V突变。在一些实施方案中,腺苷脱氨酶包含D147Y。

[0571] 举例来说,腺苷脱氨酶可在TadA参考序列中包含D108N、A106V、E155V和/或D147Y突变,或在另一腺苷脱氨酶(例如ecTadA)中的相应突变。在一些实施方案中,腺苷脱氨酶在TadA参考序列中包含以下突变组(突变组由“;”分隔),或在另一腺苷脱氨酶(例如ecTadA)中的相应突变:D108N和A106V;D108N和E155V;D108N和D147Y;A106V和E155V;A106V和D147Y;E155V和D147Y;D108N,A106V和E155V;D108N,A106V和D147Y;D108N,E155V和D147Y;A106V,E155V和D147Y;和D108N,A106V,E155V和D147Y。然而,应当理解,可以在腺苷脱氨酶(例如ecTadA)中进行本文提供的相应突变的任何组合。

[0572] 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,或另一种腺苷脱氨酶(例如ecTadA)中的一个或多个相应突变,其中X的存在表示任何野生型腺苷脱氨酶中相应氨基酸以外的其他氨基酸。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的H8Y、T17S、L18E、W23L、L34S、W45L、R51H、A56E或A56S、E59G、E85K或E85G、M94L、I95L、V102A、F104L、A106V、R107C、R107H、R107P、D108G或D108N、D108V或D108A或D108Y、K110I、M118K、N127S、A138V、F149Y、M151V、R153C、Q154L、I156D和/或K157R突变另一个腺苷脱氨酶中的一个或多个相应突变(例如ecTadA)。

[0573] 在一些实施方案中,腺苷脱氨酶在TadA参考序列中包含一个或多个H8X、D108X和/或N127X突变,或在另一个腺苷脱氨酶中的一个或多个相应突变(例如ecTadA),其中X表示存在任何氨基酸。在一些实施方案中,腺苷脱氨酶在TadA参考序列中包含H8Y、D108N和/或N127S突变中的一种或多种,或在另一种腺苷脱氨酶(例如ecTadA)中一种或多种相应的突变。

[0574] 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

变中的一种或多种,或另一个腺苷脱氨酶(例如ecTadA)中的一个或多个相应突变,其中X表示野生型腺苷脱氨酶中存在除相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包括TadA参考序列中的H8Y、R26W、M61I、L68Q、M70V、A106T、D108N、A109T、N127S、D147Y、R152C、Q154H或Q154R、E155G或E155V或E155D、K161Q、Q163H和/或T166P突变中的一个或多个,或另一个腺苷脱氨酶(例如ecTadA)中的一个或多个相应突变。

[0575] 在一些实施方案中,腺苷脱氨酶包含1、2、3、4、5或6个突变,其选自TadA参考序列中的H8X、D108X、N127X、D147X、R152X和Q154X、或相应的另一个腺苷脱氨酶(例如ecTadA)中的一个或多个突变,其中X表示野生型腺苷脱氨酶中存在除相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶在包含1、2、3、4、5、6、7或8个突变,所述突变选自TadA参考序列中的H8X、M61X、M70X、D108X、N127X、Q154X、E155X和Q163X组成的组中或相应突变,或另一种腺苷脱氨酶(例如ecTadA)中的突变,其中X表示野生型腺苷脱氨酶中存在除相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含1、2、3、4或5个突变,所述突变选自TadA参考序列中的H8X、D108X、N127X、E155X和T166X,或另一腺苷中的相应突变。脱氨酶(例如ecTadA),其中X表示野生型腺苷脱氨酶中存在除相应氨基酸以外的任何氨基酸。

[0576] 在一些实施方案中,腺苷脱氨酶包含1、2、3、4、5或6个突变,所述突变选自H8X、A106X、D108X,另一个腺苷脱氨酶中的一个或多个突变,其中X表示存在野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含1、2、3、4、5、6、7或8个选自以下的突变:H8X、R126X、L68X、D108X、N127X、D147X和E155X,或相应的突变或另一种腺苷脱氨酶的突变,其中X表示野生型腺苷脱氨酶中存在除相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含1、2、3、4或5个突变,所述突变选自TadA参考序列中的H8X、D108X、A109X、N127X和E155X,或另一个腺苷脱氨酶(例如ecTadA)中的相应突变,其中X表示野生型腺苷脱氨酶中存在除相应氨基酸以外的任何氨基酸。

[0577] 在一些实施方案中,腺苷脱氨酶包含1、2、3、4、5或6个的突变,所述突变选自TadA参考序列中的H8Y、D108N、N127S、D147Y、R152C和Q154H组成的组,或相应的一个或多个其他腺苷脱氨酶(例如ecTadA)的突变。在一些实施方案中,腺苷脱氨酶在TadA参考序列中包含1、2、3、4、5、6、7或8个突变,所述突变选自H8Y、M61I、M70V、D108N、N127S、Q154R、E155G和Q163H组成的组,或另一个腺苷脱氨酶的相应突变(例如ecTadA)。在一些实施方案中,腺苷脱氨酶包含1、2、3、4或5个突变,所述突变选自TadA参考序列中的H8Y、D108N、N127S、E155V和T166P组成的组,或另一种腺苷脱氨酶(例如ecTadA)中的相应突变。在一些实施方案中,腺苷脱氨酶包含1、2、3、4、5或6个选自TadA参考序列中的H8Y、A106T、D108N、N127S、E155D和K161Q组成的组,或相应的一个或多个突变,或另一种腺苷脱氨酶(例如ecTadA)中的突变。在一些实施方案中,腺苷脱氨酶包含1、2、3、4、5、6、7或8个突变,所述突变选自在TadA参考序列中H8Y、R126W、L68Q、D108N、N127S、D147Y和E155V,或另一个腺苷脱氨酶(例如ecTadA)的相应突变。在一些实施方案中,腺苷脱氨酶包含1、2、3、4或5个突变,所述突变选自TadA参考序列中的H8Y、D108N、A109T、N127S和E155G组成的组,或另一个腺苷脱氨酶(例如ecTadA)中的相应突变。

[0578] 可以将本文提供的任何突变和任何其他突变(例如,基于ecTadA氨基酸序列)引入任何其他腺苷脱氨酶中。本文提供的任何突变均可在TadA参考序列或另一种腺苷脱氨酶(例如ecTadA)中单独或以任何组合进行。

[0579] A至G核碱基编辑蛋白的详细信息在国际第PCT/2017/045381号专利申请(WO第2018/027078号国际专利申请)和Gaudelli,N.M.等人发表的“Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage”,Nature,551,464-471(2017)进行了描述,其全部内容通过引用并入本文。

[0580] 在一些实施方案中,腺苷脱氨酶在另一种腺苷脱氨酶(例如ecTadA)中包含一个或多个相应的突变。在一些实施方案中,腺苷脱氨酶在TadA参考序列中包含D108N、D108G或D108V突变,或在另一个腺苷脱氨酶(例如ecTadA)中的相应突变。在一些实施方案中,腺苷脱氨酶在TadA参考序列中包含A106V和D108N突变,或在另一腺苷脱氨酶(例如ecTadA)中相应的突变。在一些实施方案中,腺苷脱氨酶在TadA参考序列中包含R107C和D108N突变,或在另一个腺苷脱氨酶(例如ecTadA)中的相应突变。在一些实施方案中,腺苷脱氨酶在TadA参考序列中包含H8Y、D108N、N127S、D147Y和Q154H突变,或在另一腺苷脱氨酶(例如ecTadA)中的相应突变。在一些实施方案中,腺苷脱氨酶在TadA参考序列中包含H8Y、R24W、D108N、N127S、D147Y和E155V突变,或在另一腺苷脱氨酶(例如ecTadA)中的相应突变。在一些实施方案中,腺苷脱氨酶在TadA参考序列中包含D108N、D147Y和E155V突变,或在另一腺苷脱氨酶(例如,ecTadA)中相应的突变。在一些实施方案中,腺苷脱氨酶在TadA参考序列中包含H8Y、D108N和N127S突变,或在另一腺苷脱氨酶(例如,ecTadA)中相应的突变。在一些实施方案中,腺苷脱氨酶在TadA参考序列中包含A106V、D108N、D147Y和E155V突变,或在另一个腺苷脱氨酶(例如,ecTadA)中的相应突变。

[0581] 在一些实施方案中,腺苷脱氨酶在TadA参考序列中包含a、S2X、H8X、I49X、L84X、H123X、N127X、I156X和/或K160X突变中的一种或多种,或另一种腺苷脱氨酶中的一种或多种相应突变,其中X的存在表示野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的S2A、H8Y、I49F、L84F、H123Y、N127S、I156F和/或K160S突变中的一种或多种,或另一种腺苷脱氨酶(例如ecTadA)中的一种或多种相应的突变。

[0582] 在一些实施方案中,腺苷脱氨酶包含L84X突变腺苷脱氨酶,其中X表示野生型腺苷脱氨酶中除相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶在TadA参考序列中包含L84F突变,或在另一腺苷脱氨酶(例如,ecTadA)中包含相应的突变。

[0583] 在一些实施方案中,腺苷脱氨酶在TadA参考序列中包含H123X突变,或在另一腺苷脱氨酶(例如ecTadA)中包含相应的突变,其中X表示野生型腺苷脱氨酶中除相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶在TadA参考序列中包含H123Y突变,或在另一腺苷脱氨酶(例如,ecTadA)中包含相应的突变。

[0584] 在一些实施方案中,腺苷脱氨酶在TadA参考序列中包含I157X突变,或在另一腺苷脱氨酶(例如ecTadA)中包含相应的突变,其中X表示野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶在TadA参考序列中包含I157F突变,或在另一腺苷脱氨酶(例如,ecTadA)中包含相应的突变。

[0585] 在一些实施方案中,腺苷脱氨酶在TadA参考序列中包含1、2、3、4、5、6或7个突变,所述突变选自L84X、A106X、D108X、H123X、D147X、E155X和I156X组成的组,或另一种腺苷脱氨酶(例如ecTadA)中的相应突变,其中X表示野生型腺苷脱氨酶中存在除相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含1、2、3、4、5或6个突变,所述突变选自

TadA参考序列中的S2X、I49X、A106X、D108X、D147X和E155X组成的组,或在另一种腺苷脱氨酶(例如ecTadA)中相应的一个或多个突变,其中X表示野生型腺苷脱氨酶中存在除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1、2、3、4或5个突变,所述突变选自TadA参考序列中的H8X、A106X、D108X、N127X和K160X组成的组,或另一腺苷脱氨酶(例如ecTadA)中的相应突变,其中X表示野生型腺苷脱氨酶中存在相应氨基酸以外的任何氨基酸。

[0586] 在一些实施方案中,腺苷脱氨酶在TadA参考序列中包含1、2、3、4、5、6或7个突变,所述突变选自L84F、A106V、D108N、H123Y、D147Y、E155V和I156F组成的组,或另一个腺苷脱氨酶(例如ecTadA)的相应突变。在一些实施方案中,腺苷脱氨酶包含1、2、3、4、5或6个突变,所述突变选自在TadA参考序列中S2A、I49F、A106V、D108N、D147Y和E155V组成的组。

[0587] 在一些实施方案中,腺苷脱氨酶包含1、2、3、4或5个突变,所述突变选自TadA参考序列中的H8Y、A106T、D108N、N127S和K160S,或在另一种腺苷脱氨酶(例如ecTadA)中相应的一个或多个突变。

[0588] 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突变中的一种或多种,或另一种腺苷脱氨酶(例如ecTadA)中的一种或多种相应突变,其中X的存在表示野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的E25M、E25D、E25A、E25R、E25V、E25S、E25Y、R26G、R26N、R26Q、R26C、R26L、R26K、R107P、R07K、R107A、R107N、R107W、R107H、R107S、A142N、A142D、A142G、A143D、A143G、A143E、A143L、A143W、A143M、A143S、A143Q和/或A143R突变,或另一种腺苷脱氨酶(例如ecTadA)中的一个或多个相应突变。在一些实施方案中,腺苷脱氨酶包含一种或多种本文所述的对应于TadA参考序列的突变,或另一种腺苷脱氨酶(例如ecTadA)中的一种或多种相应的突变。

[0589] 在一些实施方案中,腺苷脱氨酶在TadA参考序列中包含E25X突变,或在另一个腺苷脱氨酶(例如ecTadA)中相应的突变,其中X表示野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含在TadA参考序列中的E25M、E25D、E25A、E25R、E25V、E25S或E25Y突变,或在另一腺苷脱氨酶(例如ecTadA)中的相应突变。

[0590] 在一些实施方案中,腺苷脱氨酶包含在TadA参考序列中的R26X突变,或在另一个腺苷脱氨酶(例如ecTadA)中包含相应的突变,其中X表示野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含在TadA参考序列中的R26G、R26N、R26Q、R26C、R26L或R26K突变,或在另一腺苷脱氨酶(例如ecTadA)中的相应突变。

[0591] 在一些实施方案中,腺苷脱氨酶包含在TadA参考序列中的R107X突变,或在另一腺苷脱氨酶(例如ecTadA)中包含相应的突变,其中X表示野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶在TadA参考序列中包含R107P、R07K、R107A、R107N、R107W、R107H或R107S突变,或在另一个腺苷脱氨酶(例如ecTadA)中的相应突变。

[0592] 在一些实施方案中,腺苷脱氨酶包含在TadA参考序列中的A142X突变,或在另一个腺苷脱氨酶(例如ecTadA)中包含相应的突变,其中X表示野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的A142N、A142D、A142G突变或另一种腺苷脱氨酶(例如,ecTadA)中的相应突变。

[0593] 在一些实施方案中,腺苷脱氨酶包含在TadA参考序列中的A143X突变,或在另一个腺苷脱氨酶(例如ecTadA)中相应的突变,其中X表示野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含在TadA参考序列中的A143D、A143G、A143E、A143L、A143W、A143M、A143S、A143Q和/或A143R突变,或在另一个腺苷脱氨酶(例如ecTadA)中的相应突变。

[0594] 在一些实施方案中,腺苷脱氨酶包含在TadA参考序列中的H36X、N37X、P48X、I49X、R51X、M70X、N72X、D77X、E134X、S146X、Q154X、K157X和/或K161X突变中的一种或多种,或另一个腺苷脱氨酶(例如ecTadA)中的一个或多个相应突变,其中X的存在表示野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H36L、N37T、N37S、P48T、P48L、I49V、R51H、R51L、M70L、N72S、D77G、E134G、S146R、S146C、Q154H、K157N和/或K161T突变中的一种或多种,或另一种腺苷脱氨酶(例如ecTadA)中的一个或多个相应突变。

[0595] 在一些实施方案中,腺苷脱氨酶在TadA参考序列中包含H36X突变,或在另一个腺苷脱氨酶(例如ecTadA)中包含相应的突变,其中X表示野生型腺苷脱氨酶中除相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶在TadA参考序列中包含H36L突变,或在另一个腺苷脱氨酶(例如,ecTadA)中相应的突变。

[0596] 在一些实施方案中,腺苷脱氨酶包含在TadA参考序列中N37X突变,或在另一个腺苷脱氨酶(例如ecTadA)中包含相应的突变,其中X表示野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含在TadA参考序列中N37T或N37S突变,或在另一腺苷脱氨酶(例如ecTadA)中相应的突变。

[0597] 在一些实施方案中,腺苷脱氨酶在TadA参考序列中包含P48X突变,或在另一个腺苷脱氨酶(例如ecTadA)中相应突变,其中X表示野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶在TadA参考序列中包含P48T或P48L突变,或在另一腺苷脱氨酶(例如ecTadA)中相应的突变。

[0598] 在一些实施方案中,腺苷脱氨酶在TadA参考序列中包含R51X突变,或在另一个腺苷脱氨酶中包含相应的突变,其中X表示野生型腺苷脱氨酶中除相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶在TadA参考序列中包含R51H或R51L突变,或在另一腺苷脱氨酶(例如ecTadA)中相应的突变。

[0599] 在一些实施方案中,腺苷脱氨酶在TadA参考序列中包含S146X突变,或在另一个腺苷脱氨酶(例如ecTadA)中的相应突变,其中X表示野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶在TadA参考序列中包含S146R或S146C突变,或在另一腺苷脱氨酶(例如ecTadA)中相应的突变。

[0600] 在一些实施方案中,腺苷脱氨酶在TadA参考序列中包含K157X突变,或在另一个腺苷脱氨酶(例如ecTadA)中的相应突变,其中X表示除野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶在TadA参考序列中包含K157N突变,或在另一腺苷脱氨酶(例如ecTadA)中相应的突变。

[0601] 在一些实施方案中,腺苷脱氨酶包含在TadA参考序列中的P48X突变,或在另一个腺苷脱氨酶(例如ecTadA)中包含相应的突变,其中X表示野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶在TadA参考序列中包含P48S、P48T或

P48A突变,或在另一腺苷脱氨酶(例如ecTadA)中相应的突变。

[0602] 在一些实施方案中,腺苷脱氨酶包含在TadA参考序列中A142X突变,或在另一个腺苷脱氨酶(例如ecTadA)中的相应突变,其中X表示野生型腺苷中相应氨基酸以外的任何氨基酸。脱氨酶。在一些实施方案中,腺苷脱氨酶包含在TadA参考序列中A142N突变,或在另一腺苷脱氨酶(例如ecTadA)中相应的突变。

[0603] 在一些实施方案中,腺苷脱氨酶包含在TadA参考序列中W23X突变,或在另一个腺苷脱氨酶(例如ecTadA)中包含相应的突变,其中X表示野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含在TadA参考序列中W23R或W23L突变,或在另一个腺苷脱氨酶(例如ecTadA)中相应的突变。

[0604] 在一些实施方案中,腺苷脱氨酶包含在TadA参考序列中R152X突变,或在另一个腺苷脱氨酶(例如ecTadA)中包含相应的突变,其中X表示野生型腺苷脱氨酶中相应氨基酸以外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含在TadA参考序列中R152P或R52H突变,或在另一腺苷脱氨酶(例如ecTadA)中相应的突变。

[0605] 在一个实施方案中,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用“_”分隔,并且每个突变的组合在括号之间:

- [0606] (A106V_D108N) ,
- [0607] (R107C_D108N) ,
- [0608] (H8Y_D108N_N127S_D147Y_Q154H) ,
- [0609] (H8Y_R24W_D108N_N127S_D147Y_E155V) ,
- [0610] (D108N_D147Y_E155V) ,
- [0611] (H8Y_D108N_N127S) ,
- [0612] (H8Y_D108N_N127S_D147Y_Q154H) ,
- [0613] (A106V_D108N_D147Y_E155V) ,
- [0614] (D108Q_D147Y_E155V) ,
- [0615] (D108M_D147Y_E155V) ,
- [0616] (D108L_D147Y_E155V) ,
- [0617] (D108K_D147Y_E155V) ,
- [0618] (D108I_D147Y_E155V) ,
- [0619] (D108F_D147Y_E155V) ,
- [0620] (A106V_D108N_D147Y) ,
- [0621] (A106V_D108M_D147Y_E155V) ,
- [0622] (E59A_A106V_D108N_D147Y_E155V) ,
- [0623] (E59A_cat_dead_A106V_D108N_D147Y_E155V) ,
- [0624] (L84F_A106V_D108N_H123Y_D147Y_E155V_I156Y) ,
- [0625] (L84F_A106V_D108N_H123Y_D147Y_E155V_I156F) ,
- [0626] (D103A_D104N) ,
- [0627] (G22P_D103A_D104N) ,

[0628] (G22P_D103A_D104N_S138 A) ,
[0629] (D103A_D104N_S138A) ,
[0630] (R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F) ,
[0631] (E25G_R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F) ,
[0632] (E25D_R26G_L84F_A106V_R107K_D108N_H123Y_A142N_A143G_D147Y_E155V_I156F) ,
[0633] (R26Q_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F) ,
[0634] (E25M_R26G_L84F_A106V_R107P_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F) ,
[0635] (R26C_L84F_A106V_R107H_D108N_H123Y_A142N_D147Y_E155V_I156F) , (L84F_A106V_D108N_H123Y_A142N_A143L_D147Y_E155V_I156F) ,
[0636] (R26G_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F) ,
[0637] (E25A_R26G_L84F_A106V_R107N_D108N_H123Y_A142N_A143E_D147Y_E155V_I156F) ,
[0638] (R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F) ,
[0639] (A106V_D108N_A142N_D147Y_E155V) ,
[0640] (R26G_A106V_D108N_A142N_D147Y_E155V) ,
[0641] (E25D_R26G_A106V_R107K_D108N_A142N_A143G_D147Y_E155V) ,
[0642] (R26G_A106V_D108N_R107H_A142N_A143D_D147Y_E155V) ,
[0643] (E25D_R26G_A106V_D108N_A142N_D147Y_E155V) ,
[0644] (A106V_R107K_D108N_A142N_D147Y_E155V) ,
[0645] (A106V_D108N_A142N_A143G_D147Y_E155V) ,
[0646] (A106V_D108N_A142N_A143L_D147Y_E155V) ,
[0647] (H36L_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N) ,
[0648] (N37T_P48T_M70L_L84F_A106V_D108N_H123Y_D147Y_I49V_E155V_I156F) ,
[0649] (N37S_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K161T) ,
[0650] (H36L_L84F_A106V_D108N_H123Y_D147Y_Q154H_E155V_I156F) ,
[0651] (N72S_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F) ,
[0652] (H36L_P48L_L84F_A106V_D108N_H123Y_E134G_D147Y_E155V_I156F) ,
[0653] (H36L_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K157N) ,
[0654] (H36L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F) ,
[0655] (L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F_K161T) ,
[0656] (N37S_R51H_D77G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F) ,
[0657] (R51L_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K157N) ,
[0658] (D24G_Q71R_L84F_H96L_A106V_D108N_H123Y_D147Y_E155V_I156F_K160E) ,
[0659] (H36L_G67V_L84F_A106V_D108N_H123Y_S146T_D147Y_E155V_I156F) ,
[0660] (Q71L_L84F_A106V_D108N_H123Y_L137M_A143E_D147Y_E155V_I156F) ,
[0661] (E25G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_Q159L) ,

[0662] (L84F_A91T_F104I_A106V_D108N_H123Y_D147Y_E155V_I156F) ,
[0663] (N72D_L84F_A106V_D108N_H123Y_G125A_D147Y_E155V_I156F) ,
[0664] (P48S_L84F_S97C_A106V_D108N_H123Y_D147Y_E155V_I156F) ,
[0665] (W23G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F) ,
[0666] (D24G_P48L_Q71R_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_Q159L) ,
[0667] (L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F) ,
[0668] (H36L_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F_K157N) ,
[0669] (N37S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_K161T) ,
[0670] (L84F_A106V_D108N_D147Y_E155V_I156F) ,
[0671] (R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K161T) ,
[0672] (L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K161T) ,
[0673] (L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K160E_K161T) ,
[0674] (L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K160E) ,
[0675] (R74Q_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F) ,
[0676] (R74A_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F) ,
[0677] (L84F_A106V_D108N_H123Y_D147Y_E155V_I156F) ,
[0678] (R74Q_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F) ,
[0679] (L84F_R98Q_A106V_D108N_H123Y_D147Y_E155V_I156F) ,
[0680] (L84F_A106V_D108N_H123Y_R129Q_D147Y_E155V_I156F) ,
[0681] (P48S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F) ,
[0682] (P48S_A142N) ,
[0683] (P48T_I49V_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_L157N) ,
[0684] (P48T_I49V_A142N) ,
[0685] (H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N) ,
[0686] (H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S146C_A142N_D147Y_E155V_I156F
[0687] (H36L_P48T_I49V_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N) ,
[0688] (H36L_P48T_I49V_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F_K157N) ,
[0689] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N) ,
[0690] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F_K157N) ,
[0691] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_A142N_D147Y_E155V_I156F_K157N) ,
[0692] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N) ,

[0693] (W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N) ,

[0694] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F_K161T) ,

[0695] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152H_E155V_I156F_K157N) ,

[0696] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F_K157N) ,

[0697] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F_K157N) ,

[0698] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142A_S146C_D147Y_E155V_I156F_K157N) ,

[0699] _I156F_K157N) ,

[0700] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142A_S146C_D147Y_R152P_E155V_I156F_K157N) ,

[0701] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F_K161T) ,

[0702] (W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F_K157N) ,

[0703] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_R152P_E155V_I156F_K157N) ,

[0704] _I156F_K157N) 。

[0705] 在某些实施方案中,本文提供的融合蛋白包含改善融合蛋白的碱基编辑活性的一种或多种特征。举例来说,本文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,本文提供的任何融合蛋白可具有不具有核酸酶活性的Cas9结构域(dCas9),或切割双链DNA分子的一条链的Cas9结构域,称为Cas9切口酶(nCas9)。

[0706] 胞苷脱氨酶

[0707] 在一个实施方案中,本发明的融合蛋白包含胞苷脱氨酶。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-甲基胞嘧啶脱氨为尿嘧啶或胸腺嘧啶。在一些实施方案中,本文提供的胞嘧啶脱氨酶能够使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。在一些实施方案中,胞苷脱氨酶是天然存在的胞苷脱氨酶,其包含与本文提供的任何突变相对应的一个或多个突变。本领域技术人员将能够例如通过序列比对和同源残基的确定来鉴定任何同源蛋白中的相应残基。因此,本领域技术人员将能够在与本文所述的任何突变相对应的任何天然存在的胞苷脱氨酶中产生突变。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)。

[0708] 在一些实施方案中,胞苷脱氨酶包含的氨基酸序列与本文所示的任何一个胞苷脱氨酶氨基酸序列至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相同。应当理解,本文提供的胞苷脱氨酶可包括一个或多个突变(例如,本文提供的任何突变)。本申请内容提供具有一定百分比同一性的任何脱氨酶结构域,加上本文所述的任何突变或其组合。在一些实施

方案中,胞苷脱氨酶与参考序列或本文提供的任何胞苷脱氨酶相比,具有1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50个或更多个突变的氨基酸序列。在一些实施方案中,胞苷脱氨酶与本领域已知或本文所述的任何氨基酸序列相比,包含具有至少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。

[0709] 本发明的融合蛋白包含核酸编辑结构域。在一些实施方案中,核酸编辑域可以催化C到U碱基的改变。在一些实施方案中,核酸编辑结构域是脱氨酶结构域。在一些实施方案中,脱氨酶是胞苷脱氨酶或腺苷脱氨酶。在一些实施方案中,脱氨酶是载脂蛋白B mRNA编辑复合体(APOBEC)家族脱氨酶。在一些实施方案中,脱氨酶是APOBEC1脱氨酶。在一些实施方案中,脱氨酶是APOBEC2脱氨酶。在一些实施方案中,脱氨酶是APOBEC3脱氨酶。在一些实施方案中,脱氨酶是APOBEC3A脱氨酶。在一些实施方案中,脱氨酶是APOBEC3B脱氨酶。在一些实施方案中,脱氨酶是APOBEC3C脱氨酶。在一些实施方案中,脱氨酶是APOBEC3D脱氨酶。在一些实施方案中,脱氨酶是APOBEC3E脱氨酶。在一些实施方案中,脱氨酶是APOBEC3F脱氨酶。在一些实施方案中,脱氨酶是APOBEC3G脱氨酶。在一些实施方案中,脱氨酶是APOBEC3H脱氨酶。在一些实施方案中,脱氨酶是APOBEC4脱氨酶。在一些实施方案中,脱氨酶是激活诱导的脱氨酶(AID)。在一些实施方案中,脱氨酶是脊椎动物脱氨酶。在一些实施方案中,脱氨酶是无脊椎动物脱氨酶。在一些实施方案中,脱氨酶是人类、黑猩猩、大猩猩、猴子、牛、狗、大鼠或小鼠的脱氨酶。在一些实施方案中,脱氨酶是人脱氨酶。在一些实施方案中,脱氨酶是大鼠脱氨酶,例如rAPOBEC1。在一些实施方案中,脱氨酶是马油双胞胎胞嘧啶核苷脱氨酶1(pmCDA1)。在一些实施方案中,脱氨酶是人APOBEC3G。在一些实施方案中,脱氨酶是人APOBEC3G的片段。在一些实施方案中,脱氨酶是包含D316R D317R突变的人APOBEC3G变体。在一些实施方案中,脱氨酶是人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具有至少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%的同一性。

[0710] 排他性降低的Cas9域

[0711] 通常,Cas9蛋白(如化脓性链球菌的Cas9(spCas9))需要典型的NGG PAM序列结合特定的核酸区域,其中“NGG”中的“N”是腺苷(A)、胸苷(T)或胞嘧啶(C),而G为鸟苷。这可能会限制编辑基因组中所需碱基的能力。在一些实施方案中,本文提供的碱基编辑融合蛋白可能需要放置在精确的位置,例如包含在PAM上游的靶碱基的区域。例如,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),其全部内容通过引用并入本文。因此,在一些实施方案中,本文提供的任何融合蛋白可包含Cas9结构域,其能够结合不包含典型(例如NGG)PAM序列的核苷酸序列。结合非典型PAM序列的Cas9结构域已经在本领域中描述,并且对本领域技术人员而言是显而易见的。例如,在Kleinstiver,B.P.等人发表的“具有改变的PAM特异性的工程化CRISPR-Cas9核酸酶”,*Nature* 523,481-485(2015)中已经描述了结合非典型PAM序列的Cas9结构域。Kleinstiver,B.P.等人发表的“Broadening the

targeting range of *Staphylococcus aureus* CRISPR-Cas9 by modifying PAM recognition”*Nature Biotechnology* 33,1293-1298(2015);Nishimasu,H.等人发表的“Engineered CRISPR-Cas9nuclease with expanded targeting space”*Science*,2018Sep 21;361(6408):1259-1262,Chatterjee,P.等人发表的“Minimal PAM specificity of a highly similar SpCas9 ortholog”,*Sci Adv.*2018 Oct 24;4(10):eaau0766.doi:10.1126/sciadv.aau0766,其各自的全部内容通过引用合并于此。下表1中描述了几种PAM变体。下表1中描述了PAM变体的几个非限制性示例:

[0712] 表1.Cas9蛋白和相应的PAM序列

变体	PAM
spCas9	NGG
spCas9-VRQR	NGA
spCas9-VRER	NGCG
xCas9 (sp)	NGN
saCas9	NNGRRT
saCas9-KKH	NNNRRT
spCas9-MQKSER	NGCG
spCas9-MQKSER	NGCN
spCas9-LRKIQK	NGTN
spCas9-LRVSQK	NGTN
spCas9-LRVSQL	NGTN
SpyMacCas9	NAA

[0714]

Cpf1	5' (TTTV)
------	-----------

[0715] 具有指导RNA的Cas9复合物

[0716] 本申请的一些方面提供了复合物,其包含本文提供的任何融合蛋白和指导RNA(例如靶向SERPINA1的指导)。在一些实施方案中,指导核酸(例如,指导RNA)长为15-100个核苷

酸,并且包含与靶序列互补的至少10个连续核苷酸的序列。在一些实施方案中,指导RNA是15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49或50个核苷酸长。在一些实施方案中,指导RNA包含与靶序列互补的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个的连续核苷酸的序列。在一些实施方案中,靶序列是DNA序列。在一些实施方案中,靶序列是细菌、酵母、真菌、昆虫、植物或动物的基因组中的序列。在一些实施方案中,靶序列是人基因组中的序列。在一些实施方案中,靶序列的3'末端紧邻典型PAM序列(NGG)。在一些实施方案中,靶序列的3'末端紧邻非典型的PAM序列(例如,表1或5'-NAA-3'中列出的序列)。在一些实施方案中,指导核酸(例如,指导RNA)与目的基因(例如, SERPINA1)中的序列互补。

[0717] 本申请的一些方面提供了使用本文提供的融合蛋白或复合物的方法。例如,本申请的一些方面提供了包括使DNA分子与本文提供的任何融合蛋白以及至少一个指导RNA接触的方法,其中所述指导RNA长约15-100个核苷酸,并且包含至少一个序列。与靶序列互补的10个连续核苷酸。在一些实施方案中,靶序列的3'末端紧邻AGC、GAG、TTT、GTG或CAA序列。在一些实施方案中,靶序列的3'末端紧邻NGA、NGCG、NGN、NNGRRT、NNNRRT、NGCG、NGCN、NGTN、NGTN、NGTN或5' (TTTV) 序列。

[0718] 应所述理解的是,各个序列中特定位置或残基的编号取决于所使用的特定蛋白质和编号方案。编号可能不同,例如在成熟蛋白的前体和成熟蛋白本身中,编号之间的差异可能会影响编号。本领域技术人员将能够通过本领域公知的方法,例如通过序列比对和同源残基的确定,来鉴定任何同源蛋白和各自编码核酸中的各自残基。

[0719] 对于本领域技术人员而言显而易见的是,为了将本文公开的任何融合蛋白靶向至靶位点,例如包含待编辑的突变的位点,通常必须共同与指导RNA一起表达融合蛋白。如本文其他地方更详细地解释的,指导RNA通常包含允许Cas9结合的tracrRNA框架和指导序列,其赋予Cas9:核酸编辑酶/结构域融合蛋白序列特异性。备选地,指导RNA和tracrRNA可以作为两个核酸分子分开提供。在一些实施方案中,指导RNA包含结构,其中指导序列包含与靶序列互补的序列。指导序列通常为20个核苷酸长。基于本申请内容,用于将Cas9:核酸编辑酶/结构域融合蛋白靶向特定基因组靶位点的合适的指导RNA的序列对于本领域技术人员是显而易见的。此类合适的指导RNA序列通常包含与待编辑靶核苷酸上游或下游50个核苷酸内的核酸序列互补的指导序列。本文提供了一些适用于将任何提供的融合蛋白靶向特定靶序列的示例性指导RNA序列。

[0720] 其他结构域

[0721] 本文所述的碱基编辑器可包括任何有助于促进多核苷酸的核碱基的编辑、修饰或改变的域。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)、核苷碱基编辑结构域(例如,脱氨酶结构域)和一个或多个其他结构域。在某些情况下,额外的结构域可以促进碱基编辑器的酶促或催化功能、碱基编辑器的结合功能,或者是可能干扰所需碱基编辑结果的细胞机器抑制剂(例如酶)。在一些实施方案中,碱基编辑器可以包含核酸酶、切口酶、重组酶、脱氨酶、甲基转移酶、甲基化酶、乙酰基酶、乙酰基转移酶、转录激活剂或转录阻遏物结构域。

[0722] 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。UGI

结构域可以例如通过抑制由C的脱氨基形成的U向C核碱基的转化来提高包含胞苷脱氨酶结构域的碱基编辑器的效率。在某些情况下,对U:G异源双链DNA的存在的细胞DNA修复反应可能导致细胞中核碱基编辑效率的降低。在这种情况下,尿嘧啶DNA糖基酶(UDG)可以催化从细胞DNA中去除U,这可以启动碱基切除修复(base excision repair, BER),主要导致U:G对还原为C:G对。在这种情况下,可以在包含一个或多个结合单链、阻断编辑的碱基、抑制UGI、抑制BER、保护编辑的碱基和/或促进未编辑的链的一个或多个域的碱基编辑器中抑制BER。因此,本发明涵盖了包含UGI结构域的碱基编辑器融合蛋白。

[0723] 在一些实施方案中,碱基编辑器包含全部或部分双链断裂(DSB)结合蛋白作为结构域。例如,DSB结合蛋白可以包括噬菌体Mu的Gam蛋白,其可以结合到DSB的末端并可以保护它们免于降解。参见Komor, A.C.等人发表的“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),全文其内容通过引用合并于此。

[0724] 在一些实施方案中,碱基编辑器可包含全部或一部分核酸聚合酶(NAP)作为域。例如,碱基编辑器可以包含真核NAP的全部或一部分。在一些实施方案中,掺入碱基编辑器中的NAP或其部分是DNA聚合酶。在一些实施方案中,并入碱基编辑器中的NAP或其部分具有跨病变聚合酶活性。在某些情况下,掺入碱基编辑器中的NAP或其部分是转病DNA聚合酶。在一些实施方案中,并入碱基编辑器中的NAP或其部分是Rev7、Rev1复合物、聚合酶 ι 、聚合酶 κ 或聚合酶 η 。在一些实施方案中,并入碱基编辑器中的NAP或其部分是真核聚合酶 α 、 β 、 γ 、 δ 、 ϵ 、 γ 、 η 、 ι 、 κ 、 λ 、 μ 或 ν 组分。在一些实施方案中,掺入碱基编辑器中的NAP或其部分包含与核酸聚合酶(例如,转损DNA聚合酶)至少75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%相同的氨基酸序列。

[0725] 碱基编辑器系统

[0726] 本文提供的碱基编辑器系统包括以下步骤:(a)使受试者的多核苷酸(例如,双链DNA或RNA、单链DNA或RNA)的靶核苷酸序列与碱基接触。编辑器系统,其包含核碱基编辑器(例如,腺苷碱基编辑器或胞苷碱基编辑器)和指导多核酸(例如,gRNA),其中所述靶核苷酸序列包含靶核碱基对;(b)诱导靶区域的链分离;(c)将靶区域的单链中的靶核碱基对的第一核碱基转化为第二核碱基;(d)切割不超过靶区域的一条链,其中与第一核碱基互补的第三核碱基被与第二核碱基互补的第四核碱基替代。应当理解,在一些实施例中,步骤(b)被省略。在一些实施方案中,靶向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,多个核碱基对位于同一基因中。在一些实施方案中,多个核碱基对位于一个或多个基因中,其中至少一个基因位于不同的基因座中。

[0727] 在一些实施方案中,切割的单链(带切口的链)与指导核酸杂交。在一些实施方案中,切割的单链与包含第一核碱基的链相反。在一些实施例中,碱基编辑器包括Cas9域。在一些实施方式中,第一碱基是腺嘌呤,而第二碱基不是G、C、A或T。在一些实施方式中,第二碱基是肌苷。

[0728] 如本文所提供的碱基编辑系统提供了一种新的基因组编辑方法,其使用融合蛋白,所述融合蛋白包含催化缺陷的化脓性链球菌Cas9、胞苷脱氨酶和碱基切除修复抑制剂

以诱导可编程的单核苷酸(C→T或A→G) DNA的变化而不会产生双链DNA断裂,不需要供体DNA模板,也不会引起过多的随机插入和缺失。

[0729] 本文提供用于使用碱基编辑器系统编辑核碱基的系统、组合物和方法。在一些实施方案中,碱基编辑器系统包括(1)碱基编辑器(BE),其包含多核苷酸可编程核苷酸结合域和用于编辑核碱基的核碱基编辑结构域(例如,脱氨酶域);(2)指导多核苷酸(例如,指导RNA)与多核苷酸可编程核苷酸结合结构域结合。在一些实施方案中,碱基编辑器系统包括胞嘧啶碱基编辑器(CBE)。在一些实施例中,碱基编辑器系统包括腺苷碱基编辑器(ABE)。在一些实施方案中,多核苷酸可编程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。在某些情况下,脱氨酶结构域可以是胞嘧啶脱氨酶或胞苷脱氨酶。在一些实施方案中,术语“胞嘧啶脱氨酶”和“胞苷脱氨酶”可以互换使用。在某些情况下,脱氨酶结构域可以是腺嘌呤脱氨酶或腺苷脱氨酶。在一些实施方案中,术语“腺嘌呤脱氨酶”和“腺苷脱氨酶”可以互换使用。核碱基编辑蛋白的细节在国际第PCT/2017/045381号专利申请(WO第2018/027078号国际专利申请)和国际第PCT/US2016/058344号专利申请(WO第2017/070632号国际专利申请)中进行了描述,将其全部内容通过引用并入本文。另请参阅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);以及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),全文其内容通过引用合并于此。

[0730] 在一些实施方案中,核碱基编辑器系统可以包括一个以上的碱基编辑组件。例如,核碱基编辑器系统可以包括一个以上的脱氨酶。在一些实施方案中,核酸酶碱基编辑器系统可包括一种或多种胞苷脱氨酶和/或一种或多种腺苷脱氨酶。在一些实施方案中,可以使用单链指导多核苷酸将不同的脱氨酶靶向靶核酸序列。在一些实施方案中,可以利用一对指导多核苷酸将不同的脱氨酶靶向靶核酸序列。

[0731] 碱基编辑器系统的核碱基成分和多核苷酸可编程核苷酸结合成分可以彼此共价或非共价结合。例如,在一些实施方案中,可以通过多核苷酸可编程核苷酸结合结构域将脱氨酶结构域靶向靶核苷酸序列。在一些实施方案中,可将多核苷酸可编程核苷酸结合结构域融合或连接至脱氨酶结构域。在一些实施方案中,多核苷酸可编程核苷酸结合结构域可通过与脱氨酶结构域非共价相互作用或与其缔合而将脱氨酶结构域靶向靶核苷酸序列。例如,在一些实施方案中,核碱基编辑组件,例如,核糖核酸编辑组件。脱氨酶组分可包含能够与作为多核苷酸可编程核苷酸结合域的一部分的其他异源部分或结构域相互作用,与其缔合或与其形成复合物的其他异源部分或结构域。在一些实施方案中,其他异源部分可能能够与多肽结合、与多肽相互作用、与多肽缔合或与其形成复合物。在一些实施方案中,其他异源部分可能能够与多核苷酸结合、与多核苷酸相互作用、与其缔合或与其形成复合物。在一些实施方案中,所述额外的异源部分可能能够结合至指导多核苷酸。在一些实施方案中,其他异源部分可能能够结合多肽接头。在一些实施方案中,其他异源部分可能能够结合多

核苷酸接头。额外的异源部分可以是蛋白质结构域。在一些实施方案中,所述额外的异源部分可以是K同源性(KH)结构域、MS2外壳蛋白结构域、PP7外壳蛋白结构域、SfMu Com外壳蛋白结构域、无体 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。

[0732] 碱基编辑器系统可以进一步包含指导多核苷酸组分。应当理解,碱基编辑器系统的组件可以通过共价键、非共价相互作用,或它们的缔合和相互作用的任何组合彼此关联。在一些实施方案中,脱氨酶结构域可以被指导多核苷酸靶向靶核苷酸序列。例如,在一些实施方案中,碱基编辑器系统的核碱基编辑组件(例如,脱氨酶成分)可包含其他异源部分或域(例如,多核苷酸结合域,例如RNA或DNA结合蛋白),所述其他异源部分或域能够与指导多核苷酸的一部分或区段(例如,多核苷酸基序)相互作用、缔合或与之形成复合物。在一些实施方案中,可以将其他异源部分或域(例如,多核苷酸结合结构域,例如RNA或DNA结合蛋白)融合或连接至脱氨酶结构域。在一些实施方案中,其他异源部分能与多肽结合,与多肽相互作用、与多肽缔合或与其形成复合物。在一些实施方案中,其他异源部分能与多核苷酸结合、与多核苷酸相互作用、与其缔合或与其形成复合物。在一些实施方案中,所述其他异源部分能结合至指导多核苷酸。在一些实施方案中,其他异源部分能结合多肽接头。在一些实施方案中,其他异源部分可能能够结合多核苷酸接头。其他异源部分可以是蛋白质结构域。在一些实施方案中,其他异源部分可以是K同源性(KH)结构域、MS2外壳蛋白结构域、PP7外壳蛋白结构域、SfMu Com外壳蛋白结构域、无体 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。

[0733] 在一些实施例中,碱基编辑器系统可以进一步包括碱基切除修复(BER)成分的抑制剂。应当理解,碱基编辑器系统的组件可以通过共价键、非共价相互作用、或它们的缔合和相互作用的任何组合彼此关联。BER组分的抑制剂可以包括碱基切除修复抑制剂。在一些实施方案中,碱基切除修复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靶向靶核苷酸序列。在一些实施方案中,可将多核苷酸可编程核苷酸结合结构域融合或连接至碱基切除修复的抑制剂。在一些实施方案中,可将多核苷酸可编程核苷酸结合结构域融合或连接至脱氨酶结构域和碱基切除修复的抑制剂。在一些实施方案中,多核苷酸可编程核苷酸结合结构域可通过与碱基切除修复抑制剂非共价相互作用或与其缔合而将碱基切除修复抑制剂靶向靶核苷酸序列。例如,在一些实施方案中,碱基切除修复组分的抑制剂可包含其他异源部分或域,所述其他异源部分或域能与作为多核苷酸可编程核苷酸结合域的一部分的其他异源部分或域相互作用、缔合或能够形成复合物。在一些实施方案中,碱基切除修复的抑制剂可以被指导多核苷酸靶向靶核苷酸序列。例如,在一些实施方案中,碱基切除修复的抑制剂可包含其他异源部分或结构域(例如多核苷酸结合结构域,例如RNA或DNA结合蛋白),所述其他异源部分或域能与指导多核苷酸的一部分或区段(例如,多核苷酸基序)相互作用、缔合或能够与之形成复合物。在一些实施方案中,指导多核苷酸(例如,多核苷酸结合结构域,例如RNA或DNA结合蛋白)的其他异源部分或域可与碱基切除修复的抑制剂融合或连接。在一些实施方案中,其他异源部分能与多核苷酸结合,与多核苷酸相互作用、与其缔合或与其形成复合物。在一些实施方案中,所述其他的异源部分可能能够结合至指导多核苷酸。在一些实施方

案中,其他异源部分可能能够结合多肽接头。在一些实施方案中,其他异源部分可能能够结合多核苷酸接头。其他异源部分可以是蛋白质结构域。在一些实施方案中,其他异源部分可以是K同源性(KH)结构域、MS2外壳蛋白结构域、PP7外壳蛋白结构域、SfMu Com外壳蛋白结构域、无菌 α 基序、端粒酶Ku结合基序和Ku蛋白、端粒酶Sm7结合基序和Sm7蛋白或RNA识别基序。

[0734] 在一些实施方案中,碱基编辑器禁止编辑的链的碱基切除修复。在一些实施方案中,碱基编辑器保护或结合未编辑的链。在一些实施例中,碱基编辑器包括UGI活动。在一些实施方案中,碱基编辑器包含催化失活的肌苷特异性核酸酶。在一些实施例中,碱基编辑器包括切口酶活性。在一些实施方案中,碱基对的预期编辑在PAM位点的上游。在一些实施方案中,碱基对的预期编辑是PAM位点上游的1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个核苷酸。在一些实施方案中,碱基对的预期编辑在PAM位点的下游。在一些实施例中,预期的经编辑的碱基对是PAM位点下游的1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20核苷酸。

[0735] 在一些实施例中,所述方法不需要典型的(例如,NGG)PAM位点。在一些实施方案中,核碱基编辑器包含接头或间隔子。在一些实施方案中,接头或间隔区的长度为1至25个氨基酸。在一些实施方案中,接头或间隔区的长度是5至20个氨基酸。在一些实施方案中,接头或间隔区的长度为10、11、12、13、14、15、16、17、18、19或20个氨基酸。

[0736] 在一些实施方案中,靶标区域包括靶标窗口,其中所述靶标窗口包括靶标核碱基对。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个核苷酸长度。在一些实施例中,碱基对的预期编辑在目标窗口内。在一些实施例中,目标窗口包括碱基对的预期编辑。在一些实施例中,使用本文提供的任何碱基编辑器来执行所述方法。在一些实施例中,目标窗口是脱氨窗口。

[0737] 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,非限制性的示例性CBE是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)接头延伸至32个氨基酸,将Cas9n-UGI接头延伸至9个氨基酸,并将UGI的第二个副本附加至所述构建体的C末端,并将另外9个氨基酸接头附加到一个碱基编辑器中构造。碱基编辑器saBE3和saBE4将化脓性链球菌Cas9n(D10A)替换为较小的金黄色葡萄球菌Cas9n(D10A)。BE3-Gam、saBE3-Gam、BE4-Gam和saBE4-Gam具有174个残基,通过16个氨基酸XTEN接头与BE3、saBE3、BE4和saBE4的N端融合。

[0738] 在一些实施例中,碱基编辑器是腺苷碱基编辑器(ABE)。在一些实施方案中,腺苷碱基编辑器可以使DNA中的腺嘌呤脱氨基。在一些实施方案中,通过用天然或工程改造的大肠杆菌TadA、人类ADAR2、小鼠ADA或人类ADAT2代替BE3的APOBEC1组分来产生ABE。在一些实施方案中,ABE包含进化的TadA变体。在一些实施方案中,ABE是ABE 1.2(TadA*-XTEN-nCas9-NLS)。在一些实施方案中,TadA*包含A106V和D108N突变。

[0739] 在一些实施例中,ABE是第二代ABE。在一些实施方案中,ABE是ABE2.1,其在TadA*(TadA*2.1)中包含另外的突变D147Y和E155V。在一些实施方案中,ABE是与人类烷基腺嘌呤DNA糖基化酶(具有E125Q突变的AAG)的催化失活形式融合的ABE2.2、ABE2.1。在一些实施方

案中,所述ABE是与催化灭活的大肠杆菌Endo V(被D35A突变灭活)的ABE2.3、ABE2.1融合。在一些实施方案中,ABE是ABE2.6,其具有比ABE2.1中的接头长两倍的接头(32个氨基酸,(SGGS)₂-XTEN-(SGGS)₂)。在一些实施方案中,ABE是ABE2.7,其是与另外的野生型TadA单体束缚的ABE2.1。在一些实施方案中,ABE是ABE2.8,其是与另外的TadA*2.1单体束缚的ABE2.1。在一些实施方案中,ABE是ABE2.9,其是进化的TadA(TadA*2.1)与ABE2.1的N末端的直接融合。在一些实施方案中,ABE是ABE2.10,其是野生型TadA与ABE2.1的N-末端的直接融合。在一些实施方案中,ABE是ABE2.11,其是在TadA*单体的N末端具有失活的E59A突变的ABE2.9。在一些实施方案中,ABE是ABE2.12,其是在内部TadA*单体中具有失活的E59A突变的ABE2.9。

[0740] 在一些实施例,ABE是第三代ABE。在一些实施方案中,ABE是ABE3.1,其是具有三个另外的TadA突变(L84F、H123Y和I157F)的ABE2.3。

[0741] 在一些实施例,ABE是第四代ABE。在一些实施方案中,ABE是ABE4.3,其是具有额外的TadA突变A142N(TadA*4.3)的ABE3.1。

[0742] 在一些实施例,ABE是第五代ABE。在一些实施方案中,ABE是ABE5.1,其通过将来自存活克隆(H36L、R51L、S146C和K157N)的共有突变集合导入ABE3.1中而产生。在一些实施方案中,ABE是ABE5.3,其具有包含与内部进化的TadA*融合的野生型大肠杆菌TadA的异二聚体构建体。在一些实施方案中,ABE是ABE5.2、ABE5.4、ABE5.5、ABE5.6、ABE5.7、ABE5.8、ABE5.9、ABE5.10、ABE5.11、ABE5.12、ABE5.13或ABE5.14,如下表2所示。在某些实施例中,ABE是第六代ABE。在一些实施例中,ABE是ABE6.1、ABE6.2、ABE6.3、ABE6.4、ABE6.5或ABE6.6,如下表2所示。在一些实施例中,ABE是第七代ABE。在一些实施方案中,ABE为ABE7.1、ABE7.2、ABE7.3、ABE7.4、ABE7.5、ABE7.6、ABE7.7、ABE7.8、ABE7.9或ABE7.10在下面的表2。

[0743] 表2.ABE的基因型

[0744]

	23	26	36	37	48	49	51	72	84	87	105	108	123	125	142	145	147	152	155	156	157	16
ABE0.1	W	R	H	N	P		R	N	L	S	A	D	H	G	A	S	D	R	E	I	K	K
ABE0.2	W	R	H	N	P		R	N	L	S	A	D	H	G	A	S	D	R	E	I	K	K
ABE1.1	W	R	H	N	P		R	N	L	S	A	N	H	G	A	S	D	R	E	I	K	K
ABE1.2	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	D	R	E	I	K	K
ABE2.1	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.2	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.3	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.4	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.5	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.6	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.7	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.8	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.9	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.10	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.11	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.12	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K

[0745]

	23	26	36	37	48	49	51	72	84	87	105	108	123	125	142	145	147	152	155	156	157	16
ABE3.1	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.2	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.3	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.4	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.5	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.6	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.7	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.8	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE4.1	W	R	H	N	P		R	N	L	S	V	N	H	G	N	S	Y	R	V	I	K	K
ABE4.2	W	G	H	N	P		R	N	L	S	V	N	H	G	N	S	Y	R	V	I	K	K
ABE4.3	W	R	H	N	P		R	N	F	S	V	N	Y	G	N	S	Y	R	V	F	K	K
ABE5.1	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.2	W	R	H	S	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	T
ABE5.3	W	R	L	N	P		L	N	I	S	V	N	Y	G	A	C	Y	R	V	I	N	K
ABE5.4	W	R	H	S	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	T
ABE5.5	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.6	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.7	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.8	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.9	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.10	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.11	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.12	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.13	W	R	H	N	P		L	D	F	S	V	N	Y	A	A	S	Y	R	V	F	K	K
ABE5.14	W	R	H	N	S		L	N	F	C	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE6.1	W	R	H	N	S		L	N	F	S	V	N	Y	G	N	S	Y	R	V	F	K	K
ABE6.2	W	R	H	N	T	V	L	N	F	S	V	N	Y	G	N	S	Y	R	V	F	N	K
ABE6.3	W	R	L	N	S		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE6.4	W	R	L	N	S		L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
ABE6.5	W	R	L	N	I	V	L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE6.6	W	R	L	N	T	V	L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
ABE7.1	W	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE7.2	W	R	L	N	A		L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K

	23	26	36	37	48	49	51	72	84	87	105	108	123	125	142	145	147	152	155	156	157	16
ABE7.3	I	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE7.4	R	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE7.5	W	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	H	V	F	N	K
[0746] ABE7.6	W	R	L	N	A		L	N	I	S	V	N	Y	G	A	C	Y	P	V	I	N	K
ABE7.7	L	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	P	V	F	N	K
ABE7.8	I	R	L	N	A		L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
ABE7.9	L	R	L	N	A		L	N	F	S	V	N	Y	G	N	C	Y	P	V	F	N	K
ABE7.10	R	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	P	V	F	N	K

[0747] 在一些实施方案中,碱基编辑器是融合蛋白,其包含与核碱基编辑结构域(例如,脱氨酶域的全部或一部分)融合的多核苷酸可编程核苷酸结合域(例如,Cas9衍生的域)。在一些实施方案中,碱基编辑器还包含包含全部或部分尿嘧啶糖基化酶抑制剂(UGI)的结构域。在一些实施方案中,碱基编辑器包含包含尿嘧啶结合蛋白(UBP)例如尿嘧啶DNA糖基化酶(UDG)的全部或一部分的结构域。在一些实施方案中,碱基编辑器包含包含全部或部分核酸聚合酶的结构域。在一些实施方案中,掺入碱基编辑器中的核酸聚合酶或其部分是绕道(translesion)DNA聚合酶。

[0748] 在一些实施例中,碱基编辑器的域可以包括多个域。例如,包含衍生自Cas9的多核苷酸可编程核苷酸结合结构域的碱基编辑器可以包含REC叶和对应于野生型或天然Cas9的REC叶和NUC叶的REC叶和NUC叶。在另一个示例中,碱基编辑器可以包括RuvCI域、BH域、REC1域、REC2域、RuvCII域、L1域、HNH域、L2域、RuvCIII域、WED域、TOPO域或CTD域中的一个或多个。在一些实施方案中,碱基编辑器的一个或多个结构域包含相对于包含所述结构域的多肽的野生型版本的突变(例如,取代、插入、缺失)。举例来说,多核苷酸可编程DNA结合结构域的HNH结构域可包含H840A取代。在另一个实例中,多核苷酸可编程DNA结合结构域的RuvCI结构域可包含D10A取代。

[0749] 本文公开的碱基编辑器的不同域(例如,相邻域)可以在使用或不使用一个或多个接头域(例如,XTEN接头域)的情况下彼此连接。在一些情况下,接头结构域可以是键(例如,共价键)、化学基团或连接两个分子或部分的分子,例如融合蛋白的两个结构域,例如第一结构域(例如,Cas9衍生的结构域)和第二结构域(例如胞苷脱氨酶结构域或腺苷脱氨酶结构域)。在一些实施方案中,连接基为共价键(例如,碳-碳键、二硫键、碳-杂原子键等)。在某些实施方案中,连接基是酰胺键的碳氮键。在某些实施方案中,接头是环状或无环的、取代或未取代的、支链或直链的脂族或杂脂族接头。在某些实施方案中,接头是聚合物的(例如,聚乙烯、聚乙二醇、聚酰胺、聚酯等)。在某些实施方案中,连接基包含氨基链烷酸的单体、二聚体或聚合物。在一些实施方案中,连接基包含氨基链烷酸(例如,甘氨酸、乙酸、丙氨酸、 β -丙氨酸、3-氨基丙酸、4-氨基丁酸、5-戊酸等)。在一些实施方案中,接头包含氨基己酸(Ahx)的单体、二聚体或聚合物。在某些实施方案中,连接基基于碳环部分(例如环戊烷、环己烷)。在其他实施方案中,接头包含聚乙二醇部分(PEG)。在某些实施方案中,连接基包含芳基或杂芳基部分。在某些实施方案中,连接基基于苯环。接头可以包括官能化的部分,以促进亲

核试剂(例如硫醇、氨基)从肽到接头的连接。任何亲电子试剂都可以用作接头的一部分。示例性的亲电试剂包括但不限于活化的酯、活化的酰胺、迈克尔受体、烷基卤化物、芳基卤化物、酰基卤化物和异硫氰酸酯。在一些实施方案中,连接子连接RNA可编程核酸酶的gRNA结合结构域,包括Cas9核酸酶结构域,以及核酸编辑蛋白的催化结构域。在一些实施方案中,接头连接dCas9和第二结构域(例如胞苷脱氨酶、UGI等)。

[0750] 通常,连接子位于两个基团、分子或其他部分之间或在两个基团、分子或其他部分的侧面,并通过共价键连接至每个,从而将两者连接。在一些实施方案中,接头是氨基酸或多个氨基酸(例如,肽或蛋白质)。在一些实施方案中,连接基是有机分子、基团、聚合物或化学部分。在一些实施方案中,接头的长度为2至100个氨基酸,例如长度为2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、30至35、35至40、40至45、45至50、50至60、60至70、70至80、80至90、90至100、100至150或150至200个氨基酸。在一些实施方案中,接头长度为约3至约104个氨基酸(例如长度为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)。也可以考虑更长或更短的接头。在一些实施方案中,接头结构域包含氨基酸序列SGSETPGTSESATPES,其也可以称为XTEN接头。可以使用任何用于连接融合蛋白结构域的方法(例如,从形式非常灵活的接头(SGGS)_n、(GGGS)_n、(GGGGS)_n和(G)_n到形式更刚性的接头(EAAAK)_n、(GGG)_n、SGSETPGTSESATPES(例如参见Guilinger JP、Thompson DB、Liu DR发表的“Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification”,*Nat. Biotechnol.* 2014;32(6):577-82;全部内容通过引用并入本文)或(XP)_n基序,或这些的任何组合,以实现核碱基编辑器活性的最佳长度,其中n独立为整数在一些实施方案中,n为1、2、3、4、5、6、7、8、9、10、11、12、13、14或15,其中X为任何氨基酸。在一些实施方案中,接头包含(GGS)_n基序,其中n为1、3或7。在一些实施方案中,本文提供的融合蛋白的Cas9结构域通过含氨基酸序列SGSETPGTSESATPES的接头进行融合。在一些实施方案中,接头包含多个脯氨酸残基,并且长度为5至21、5至14、5至9、5至7个氨基酸,例如PAPAP、PAPAPA、PAPAPAP、PAPAPAPA、P(AP)₄、P(AP)₇、P(AP)₁₀(例如,参见Tan J、Zhang F、Karcher D、Bock R.在2019年1月25日发表的“Engineering of high-precision base editors for site-specific single nucleotide replacement.”,*Nat Commun.*;10(1):439;全部内容通过引用并入本文。这种富含脯氨酸的接头也被称为“刚性”接头。

[0751] 本文公开的碱基编辑器的域可以任何序列排列。包含如下融合蛋白的碱基编辑器的非限制性实例可以如下排列:所述融合蛋白包含例如多核苷酸可编程核苷酸结合结构域和脱氨酶结构域:

[0752] NH₂-[核碱基编辑域]-接头1-[例如,Cas9衍生域]-COOH;

[0753] NH₂-[例如,胞苷脱氨酶]-接头1-[例如,Cas9衍生域]-COOH;

[0754] NH₂-[例如,胞苷脱氨酶]-接头1-[例如,Cas9衍生域]-接头2-[UGI]-COOH;

[0755] NH₂-[例如,APOBEC]-接头1-[例如,Cas9衍生域]-COOH;

[0756] NH₂-[例如,胞苷脱氨酶]-接头1-[例如,Cas9衍生域]-COOH;

[0757] NH₂-[例如,APOBEC]-接头1-[例如,Cas9衍生域]-COOH;

- [0758] NH_2 -[例如,APOBEC]-接头1-[例如,Cas9衍生域]-接头2-[UGI]-COOH
- [0759] NH_2 -[例如,肌苷BER抑制剂]-[例如,Cas9衍生域]-COOH;
- [0760] NH_2 -[例如,Cas9衍生域]-[例如,肌苷BER抑制剂]-COOH;
- [0761] NH_2 -[例如,肌苷BER抑制剂]-[例如,Cas9衍生域]-[肌苷BER抑制剂]-COOH;
- [0762] NH_2 -[例如,肌苷BER抑制剂]-[肌苷BER抑制剂]-[例如,Cas9衍生域]-COOH;
- [0763] NH_2 -[肌苷BER抑制剂]-[例如,肌苷BER抑制剂]-[例如,Cas9衍生域]-COOH;
- [0764] NH_2 -[例如,Cas9衍生域]-[例如,肌苷BER抑制剂]-[肌苷BER抑制剂]-COOH;
- [0765] NH_2 -[例如,Cas9衍生域]-[肌苷BER抑制剂]-[例如,肌苷BER抑制剂]-COOH;or
- [0766] NH_2 -[肌苷BER抑制剂]-[例如,Cas9衍生域]-[例如,肌苷BER抑制剂]-COOH。
- [0767] 另外,在某些情况下,Gam蛋白可以与碱基编辑器的N末端融合。在某些情况下,Gam蛋白可以融合到碱基编辑器的C末端。噬菌体Mu的Gam蛋白可以结合到双链断裂(DSB)的末端,并保护其免于降解。在一些实施方案中,使用Gam结合DSB的自由端可以减少碱基编辑过程中插入缺失的形成。在一些实施方案中,将174个残基的Gam蛋白融合至碱基编辑器的N末端。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)。在某些情况下,一个或多个突变可以改变碱基编辑域相对于野生型域的长度。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,一个或多个突变不改变相对于野生型结构域的结构域的长度。例如,任何域中的替换都不会更改碱基编辑器的长度。这些碱基编辑器的非限制性示例(其中所有域的长度均与野生型域相同)可以包括:
- [0768] NH_2 -[APOBEC1]-接头1-[Cas9(D10A)]-接头2-[UGI]-COOH;
- [0769] NH_2 -[CDA1]-接头1-[Cas9(D10A)]-接头2-[UGI]-COOH;
- [0770] NH_2 -[AID]-接头1-[Cas9(D10A)]-接头2-[UGI]-COOH;
- [0771] NH_2 -[APOBEC1]-接头1-[Cas9(D10A)]-接头2-[SSB]-COOH;
- [0772] NH_2 -[UGI]-接头1-[APOBEC1]-接头2-[Cas9(D10A)]-COOH;
- [0773] NH_2 -[APOBEC1]-接头1-[Cas9(D10A)]-接头2-[UGI]-接头3-[UGI]-COOH;
- [0774] NH_2 -[Cas9(D10A)]-接头1-[CDA1]-接头2-[UGI]-COOH;
- [0775] NH_2 -[Gam]-接头1-[APOBEC1]-接头2-[Cas9(D10A)]-接头3-[UGI]-COOH;
- [0776] NH_2 -[Gam]-接头1-[APOBEC1]-接头2-[Cas9(D10A)]-接头3-[UGI]-接头4-[UGI]-COOH;
- [0777] NH_2 -[APOBEC1]-接头1-[dCas9(D10A,H840A)]-接头2-[UGI]-COOH;or
- [0778] NH_2 -[APOBEC1]-接头1-[dCas9(D10A,H840A)]-COOH。
- [0779] 在一些实施方案中,本文提供的碱基编辑融合蛋白需要定位在精确的位置,例如,靶碱基被放置在限定的区域内(例如,“脱氨窗口”)。在某些情况下,目标可以在4个碱基区域内。在某些情况下,这样定义的目标区域可以在PAM上游约15个碱基处。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);以及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),全文其内容通过引用合并于此。

[0780] 定义的目标区域可以是脱氨窗口。脱氨基窗口可以是碱基编辑器作用于靶核苷酸并使其脱氨基的定义区域。在一些实施例中,脱氨窗在2、3、4、5、6、7、8、9或10个碱基区域内。在一些实施例中,脱氨窗为在PAM的上游5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24或25碱基。

[0781] 本申请的碱基编辑器可包含任何结构域、特征或氨基酸序列,其有助于靶多核苷酸序列的编辑。例如,在一些实施例中,碱基编辑器包括核定位序列(NLS)。在一些实施方案中,碱基编辑器的NLS位于脱氨酶结构域和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位于C端至多核苷酸可编程核苷酸结合结构域。

[0782] 如本文公开的,可存在于碱基编辑器中的其他示例性特征是定位序列,例如细胞质定位序列、输出序列,例如核输出序列或其他定位序列,以及可用于融合蛋白的增溶、纯化或检测。本文提供的合适的蛋白质标签包括但不限于生物素羧化酶载体蛋白(BCCP)标签、myc标签、钙调蛋白标签、FLAG标签、血凝素(HA)标签、多组氨酸标签(也称为组氨酸标签或His标签)、麦芽糖结合蛋白(MBP)标签、nus标签、谷胱甘肽S-转移酶(GST)标签、绿色荧光蛋白(GFP)标签、硫氧还蛋白标签、S标签、软标签(例如,软标签1、软标签3)、链球菌标签、生物素连接酶标签、FlAsH标签、V5标签和SBP标签。对于本领域技术人员而言,其他合适的序列是显而易见的。在一些实施方案中,融合蛋白包含一个或多个His标签。

[0783] 可包含在融合蛋白中的蛋白质结构域的非限制性实例包括脱氨酶结构域(例如胞苷脱氨酶和/或腺苷脱氨酶)、尿嘧啶糖基化酶抑制剂(UGI)结构域、表位标签、报告基因序列、和/或具有以下一种或多种活性的蛋白质结构域:甲基酶活性、脱甲基酶活性、转录激活活性、转录抑制活性、转录释放因子活性、组蛋白修饰活性、RNA裂解活性和核酸结合活性。其他域可以是异源功能域。这样的异源功能域可以赋予功能活性,例如DNA甲基化、DNA损伤、DNA修复、与靶DNA相关的靶多肽(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。

[0784] 赋予的其他功能可以包括甲基转移酶活性、脱甲基酶活性、脱氨基活性、歧化酶活性、烷基化活性、脱嘌呤活性、氧化活性、嘧啶二聚体形成活性、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、乙酰基转移酶活性、脱乙酰基酶活性、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、去烯基化活性、SUMO化化活性、脱SUMO化化活性、核糖基化活性、去核糖基化活性、肉豆蔻酰基化活性、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、合成酶活性和脱豆基甲酰化活性或其任何组合。

[0785] 表位标签的非限制性实例包括组氨酸(His)标签、V5标签、FLAG标签、流感血凝素(HA)标签、Myc标签、VSV-G标签和硫氧还蛋白(Trx)标签。报告基因的例子包括但不限于谷胱甘肽-S-转移酶(GST)、辣根过氧化物酶(HRP)、氯霉素乙酰基转移酶(CAT) β -半乳糖苷酶、 β -葡萄糖醛酸糖苷酶、荧光素酶、绿色荧光蛋白(GFP)、HcRed、DsRed、青色荧光蛋白(CFP)、黄色荧光蛋白(YFP)和包括蓝色荧光蛋白(BFP)的自发荧光蛋白。其他蛋白质序列可包括结合

DNA分子或结合其他细胞分子的氨基酸序列,包括但不限于麦芽糖结合蛋白(MBP)、S标签、Lex A DNA结合域(DBD)融合物、GAL4 DNA结合域融合物、和单纯疱疹病毒(HSV)BP16蛋白融合体。

[0786] 碱基编辑效率

[0787] CRISPR-Cas9核酸酶已被广泛用于介导靶向基因组编辑。在大多数基因组编辑应用中,Cas9与指导多核苷酸(例如单指导RNA(sgRNA))形成复合物,并在由sgRNA序列指定的靶位点诱导双链DNA断裂(DSB)。细胞主要通过非同源末端连接(NHEJ)修复途径对DSB作出反应,这会导致随机插入或缺失(Indels),从而导致移码突变破坏基因。在存在与DSB侧翼具有高度同源性的供体DNA模板的情况下,可以通过称为同源引导修复(HDR)的另一种途径实现基因校正。不幸的是,在大多数非摄动条件下,HDR效率低下,取决于细胞状态和细胞类型,并以更大频率的插入/缺失为主导。由于与人类疾病相关的大多数已知遗传变异都是点突变,因此需要能够更有效,更干净地进行精确点突变的方法。本文提供的碱基编辑系统提供了一种新的编辑基因组编辑的方式,而不会产生双链DNA断裂,不需要供体DNA模板,并且不会引起过多的随机插入和缺失。

[0788] 本文提供的碱基编辑器能够修饰特定的核苷酸碱基而不会产生显着比例的插入缺失。如本文所用,术语“插入/缺失”是指核酸内核苷酸碱基的插入或缺失。此类插入或缺失可导致基因编码区内的移码突变。在一些实施方案中,期望产生有效地修饰(例如,突变或脱氨基)核酸内的特定核苷酸而不在靶核苷酸序列中产生大量插入或缺失(即,插入缺失)的碱基编辑器。在某些实施方案中,与插入缺失相比,本文提供的任何碱基编辑器能够产生更大比例的预期修饰(例如,点突变或脱氨基)。

[0789] 在一些实施例中,本文提供的任何碱基编辑器系统导致在靶多核苷酸序列中小于50%、小于40%、小于30%、小于20%、小于19%、小于18%、小于17%、小于16%、小于15%、小于14%、小于13%、小于12%、小于11%、小于10%、小于9%、小于8%、小于7%、小于6%、小于5%、小于4%、小于3%、小于2%、小于1%、小于0.9%、小于0.8%、小于0.7%、小于0.6%、小于0.5%、小于0.4%、小于0.3%、小于0.2%、小于0.1%、小于0.09%、小于0.08%、小于0.07%、小于0.06%、小于0.05%、小于0.04%、小于0.03%、小于0.02%或小于0.01%的插入缺失形成。

[0790] 本申请的一些方面是基于这样的认识,即本文提供的任何碱基编辑器均能够在核酸(例如DNA的基因组内的核酸)中有效产生预期的突变(例如点突变),而不会产生大量的意外突变(例如意外的点突变)。

[0791] 在一些实施方案中,本文提供的任何碱基编辑器均能够产生至少0.01%的预期突变(即,至少0.01%的碱基编辑效率)。在一些实施例中,本文提供的任何碱基编辑器均能够产生至少0.01%、1%、2%、3%、4%、5%、10%、15%、20%、25%、30%、40%、45%、50%、60%、70%、80%、90%、95%或99%的预期突变。

[0792] 在一些实施方案中,本文提供的碱基编辑器能够产生预期的点突变与插入缺失的比率大于1:1。在一些实施方案中,本文提供的碱基编辑器能够产生预期的点突变与插入缺失的比率为至少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8.5:1、至少9:1、至少10:1、至少11:1、至少12:1、至少13:1、至少14:1、至少15:1、至少20:1、至少25:1、至

少30:1、至少40:1、至少50:1、至少100:1、至少200:1、至少300 1:1、至少400:1、至少500:1、至少600:1、至少700:1、至少800:1、至少900:1或至少1000:1或更多。

[0793] 预期的突变和插入/缺失的数目可以使用任何合适的方法来确定,例如,如国际第PCT/2017/045381号专利申请(WO第2018/027078号国际专利申请)和第PCT/US2016/058344号专利申请(WO第2017/070632号国际专利申请)中所述;KKomor,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);以及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:eaa04774(2017);其全部内容通过引用结合于此。

[0794] 在一些实施方案中,为了计算插入缺失频率、扫描测序读数以寻找与两个10-bp序列的精确匹配,所述两个10-bp序列位于可能发生插入缺失的窗口的两侧。如果未找到完全匹配的内容,则将其从分析中排除。如果此插入缺失窗口的长度与参考序列完全匹配,则将读数分类为不包含插入缺失。如果插入缺失窗口比参考序列长或短两个或多个碱基,则测序读数分别分类为插入或缺失。在一些实施方案中,本文提供的碱基编辑器可以限制核酸区域中插入缺失的形成。在一些实施方案中,所述区域位于碱基编辑器靶向的核苷酸处或碱基编辑器靶向的核苷酸的2、3、4、5、6、7、8、9或10个核苷酸内的区域。

[0795] 在靶核苷酸区域形成的插入缺失的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,在将目标核苷酸序列(例如,细胞基因组内的核酸)暴露于碱基编辑器至少1小时、至少2小时、至少6小时、至少12小时、至少24小时、至少36小时,至少48小时,至少3天,至少4天,至少5天,至少7天,至少10天或至少14天后,确定插入缺失的数量或比例。应当理解,如本文所述的碱基编辑器的特征可以应用于任何融合蛋白,或使用本文提供的融合蛋白的方法。

[0796] 多重编辑

[0797] 在一些实施方案中,本文提供的碱基编辑器系统能够多重编辑一个或多个基因中的多个核碱基对。在一些实施方案中,多个核碱基对位于同一基因中。在一些实施方案中,多个核碱基对位于一个或多个基因中,其中至少一个基因位于不同的基因座中。在一些实施方案中,多重编辑可包含一个或多个指导多核苷酸。在一些实施例,多重编辑可以包括一个或多个碱基编辑器系统。在一些实施方案中,多重编辑可包括具有单链指导多核苷酸的一个或多个碱基编辑器系统。在一些实施方案中,多重编辑可以包括具有多个指导多核苷酸的一个或多个碱基编辑器系统。在一些实施方案中,多重编辑可包含具有单个碱基编辑器系统的一个或多个指导多核苷酸。在一些实施方案中,多重编辑可包含至少一种不需要PAM序列靶向结合靶多核苷酸序列的指导多核苷酸。在一些实施方案中,多重编辑可包含至少一种需要PAM序列以靶向结合靶多核苷酸序列的指导多核苷酸。在一些实施方案中,多重编辑可以包含至少一种不需要PAM序列靶向结合靶多核苷酸序列的指导多核苷酸和至少一种需要PAM序列靶向结合靶多核苷酸的指导多核苷酸的混合物序列。应当理解,使用本文所述的任何碱基编辑器的多路编辑的特征可以应用于使用本文提供的任何碱基编辑器的方法的任意组合。还应当理解,使用本文所述的任何碱基编辑器进行的多重编辑可包括多

个核碱基对的序列编辑。

[0798] 本文提供的方法包括以下步骤：(a) 使受试者的多核苷酸的靶核苷酸序列（例如，双链DNA序列）与包含核碱基编辑器（例如，腺苷）的碱基编辑器系统接触。碱基编辑器或胞苷碱基编辑器）和引导多核酸（例如，gRNA），其中靶核苷酸序列包含靶核碱基对；(b) 诱导靶区域的链分离；(c) 将靶区域的单链中的靶核碱基对的第一核碱基编辑为第二核碱基；(d) 切割不超过靶区域的一条链，其中与第一核碱基互补的第三核碱基被与第二核碱基互补的第四核碱基替代。

[0799] 在一些实施方案中，多个核碱基对在一个以上的基因中。在一些实施方案中，多个核碱基对在同一基因中。在一些实施方案中，一个或多个基因中的至少一个基因位于不同的基因座中。

[0800] 在一些实施方案中，所述编辑是在至少一个蛋白质编码区中的多个核碱基对的编辑。在一些实施方案中，所述编辑是对至少一个蛋白质非编码区中的多个核碱基对的编辑。在一些实施方案中，所述编辑是编辑至少一个蛋白质编码区和至少一个蛋白质非编码区中的多个核碱基对。

[0801] 在一些实施方案中，所述编辑与一种或多种指导多核苷酸结合。在一些实施例中，碱基编辑器系统可以包括一个或多个碱基编辑器系统。在一些实施方案中，碱基编辑器系统可包括与单链指导多核苷酸结合的一个或多个碱基编辑器系统。在一些实施方案中，碱基编辑器系统可包含一个或多个碱基编辑器系统以及多个指导多核苷酸。在一些实施方案中，使用单个碱基编辑器系统将编辑与一个或多个指导多核苷酸结合。在一些实施方案中，编辑与至少一种不需要PAM序列靶向结合靶多核苷酸序列的指导多核苷酸结合。在一些实施方案中，所述编辑与至少一种需要PAM序列以靶向结合靶多核苷酸序列的指导多核苷酸结合。在一些实施方案中，所述编辑与至少一种不需要PAM序列靶向结合靶多核苷酸序列的指导多核苷酸和至少一种需要PAM序列靶向结合靶标的指导多核苷酸的混合物结合。多核苷酸序列。应当理解，使用本文所述的任何碱基编辑器的多路编辑的特性可以应用于使用本文提供的任何碱基编辑器的方法的任意组合。还应当理解，编辑可以包括多个核碱基对的序列编辑。

[0802] 碱基编辑器的使用方法

[0803] 疾病相关基因和等位基因中点突变的校正为治疗和基础研究中的应用开辟了基因校正的新策略。

[0804] 本申请提供了用于治疗被诊断患有与点突变相关的疾病或由点突变引起的疾病的对象的方法，所述疾病可以通过本文提供的碱基编辑器系统来校正。例如，在一些实施方案中，提供了一种方法，其包括对患有这种疾病（例如，由遗传突变引起的疾病）的受试者施用有效量的核碱基编辑器（例如，腺苷脱氨酶碱基编辑器或胞苷脱氨酶碱基编辑器），可校正疾病相关基因中的点突变。

[0805] 在一些实施方案中，所述疾病是增生性疾病。在一些实施方案中，所述疾病是遗传性疾病。在一些实施方案中，所述疾病是赘生性疾病。在一些实施方案中，所述疾病是代谢性疾病。在一些实施方案中，所述疾病是溶酶体贮积病。示例性的合适疾病和病症包括但不限于视网膜色素病变（例如，adRP-PRPF3、adRP-RHO）、尤塞氏综合症1F型、镰状细胞病、 α -1抗胰蛋白酶缺乏症（A1AD）、肝紫质症、MCAD缺乏症、LAL缺乏症、苯丙酮尿症（PKU）、血色素沉着

症、冯吉尔克病 (GSD1a)、庞贝氏症 (GSDII)、高雪氏症、贺勒氏症 (MPS1)、囊性纤维化、高半胱氨酸尿症 (HCU) 或慢性疼痛。可以通过校正点突变或将失活突变引入疾病相关基因来治疗的其他疾病是本领域技术人员已知的,并且本申请不限于此。提供了用于治疗其他疾病或病症的方法,所述其他疾病或病症例如与可以通过脱氨酶介导的基因编辑来校正的点突变相关或由其引起的疾病或病症。本文描述了这样的疾病,并且基于本申请,可以用本文提供的策略和融合蛋白治疗的其他合适的疾病对于本领域技术人员是显而易见的。

[0806] 在某些方面,提供了用于治疗A1AD的方法,其与点突变相关或由其引起(例如,在编码A1AT蛋白的SERPINA1基因中)并且可以通过脱氨酶介导的基因编辑来校正。

[0807] 应当理解,相应序列中特定位置或残基的编号,例如疾病相关基因或其编码的蛋白质的多核苷酸或氨基酸序列,分别取决于所用的特定蛋白质和编号方案。编号可能不同,例如,在成熟蛋白的前体和成熟蛋白本身中,编号之间的差异会影响编号。本领域技术人员将能够通过本领域公知的方法,例如通过序列比对和同源残基的确定,来鉴定任何同源蛋白和各自编码核酸中的各自残基。

[0808] 本文提供了使用碱基编辑器或碱基编辑器系统来编辑与疾病或病症相关的靶核苷酸序列中的核碱基的方法。在一些实施方案中,碱基编辑器的活性(例如,包含腺苷脱氨酶和Cas9结构域)导致点突变的校正。在一些实施方案中,靶DNA序列包含与疾病或病症相关的G→A点突变,并且其中突变体A碱基的脱氨基导致不与疾病或病症相关的序列。在一些实施方案中,靶DNA序列包含与疾病或病症相关的T→C点突变,并且其中突变体C碱基的脱氨基导致不与疾病或病症相关的序列。

[0809] 在一些实施方案中,靶DNA序列编码蛋白质,并且点突变是密码子,与野生型密码子相比,突变密码子编码的氨基酸发生改变。在一些实施方案中,突变体A的脱氨基导致突变体密码子编码的氨基酸的改变。在一些实施方案中,突变体A的脱氨基产生编码野生型氨基酸的密码子。在一些实施方案中,突变体C的脱氨基导致突变体密码子编码的氨基酸的改变。在一些实施方案中,突变体C的脱氨基导致编码野生型氨基酸的密码子。在一些实施方案中,受试者已经或已经被诊断出疾病或病症。

[0810] 在一些实施方案中,本文提供的腺苷脱氨酶能够使DNA的脱氧腺苷残基的腺嘌呤脱氨基。本申请内容的其他方面提供了融合蛋白,其包含能够结合至特定核苷酸序列的腺苷脱氨酶(例如,如本文所述将DNA中的脱氧腺苷脱氨的腺苷脱氨酶)和结构域(例如,Cas9或Cpf1蛋白)。例如,腺苷可被转化为肌苷残基,其通常与胞嘧啶残基碱基对。这样的融合蛋白尤其可用于核酸序列的靶向编辑。这样的融合蛋白可用于体外DNA的靶向编辑,例如用于突变细胞或动物的产生。用于引入靶向突变,例如,用于校正离体细胞(例如,从受试者获得的细胞中的遗传缺陷),然后将其重新引入相同或另一个受试者中;对于体内定向突变的引入,例如,遗传缺陷的校正或疾病相关基因在G到A或T到C到突变的疾病相关基因的引入可以使用本文提供的核碱基编辑器进行处理。本申请提供了利用所述脱氨酶和核碱基编辑器的脱氨酶,融合蛋白,核酸,载体,细胞,组合物,方法,试剂盒,系统等。

[0811] 使用核碱基编辑器靶向SERPINA1基因中的核苷酸

[0812] 如本文所述评估靶向SERPINA1基因中核苷酸的核碱基编辑器的适用性。在一个实施方案中,用一种或多种编码本文所述的核碱基编辑器的核酸分子和少量编码报导分子(例如GFP)的载体转染,转导或以其他方式修饰目的细胞。这些细胞可以是永生化的细胞

胞系,例如293T、K562或U20S。或者,可以使用原代人细胞。细胞也可以获自受试者或个体,例如获自组织活检、手术、血液、血浆、血清或其他生物流体。这些细胞可能与最终的细胞靶标有关,

[0813] 可以使用病毒载体进行递送,如下文进一步描述。在一个实施方案中,可以使用脂质转染(例如脂转染胺或Fugene)或通过电穿孔进行转染。转染后,可以通过荧光显微镜或流式细胞术确定GFP的表达,以确认一致和高水平的转染。这些初步转染可以包含不同的核碱基编辑器,以确定哪些编辑器组合提供最大的活性。

[0814] 如本文所述评估核碱基编辑器的活性,即通过对靶基因进行测序以检测靶序列的改变。对于Sanger测序,将纯化的PCR扩增子克隆到质粒主链中,进行转化,微量制备并用单个引物测序。测序也可以使用下一代测序技术进行。使用下一代测序时,扩增子可能为300-500bp,预期的切割位点不对称放置。PCR之后,可以将下一代测序适配器和条形码(例如Illumina多重适配器和索引)添加到扩增子的末端,例如用于高通量测序(例如在Illumina MiSeq上)。

[0815] 可以选择在初始测试中诱导最大水平的靶标特异性改变的融合蛋白进行进一步评估。

[0816] 在特定的实施方案中,核碱基编辑器用于靶向目的多核苷酸。在一个实施方案中,将本发明的核碱基编辑器与用于靶向核酸序列的指导RNA一起递送至细胞(例如,肝细胞),所述核酸序列例如具有AIAD相关突变的SERPINA1多核苷酸,从而改变靶标基因,即SERPINA1。

[0817] 在一些实施方案中,指导RNA靶向碱基编辑器以将一个或多个编辑引入目的基因的序列。在一些实施方案中,引入SERPINA1或SERPINC1基因的一个或多个改变如下表3A和3B所示。

[0818] 生成预期的突变

[0819] 在一些实施方案中,本文提供的方法的目的是通过基因编辑来恢复功能异常的基因的功能。在一些实施方案中,功能异常的基因的功能通过引入预期的突变来恢复。可以例如通过校正人细胞培养物中与疾病相关的突变来验证本文提供的核碱基编辑蛋白用于体外基于基因编辑的人治疗剂。本领域技术人员将理解,本文提供的核碱基编辑蛋白,例如,包含多核苷酸可编程核苷酸结合结构域(例如,Cas9)和核碱基编辑结构域(例如,腺苷脱氨酶结构域或胞苷脱氨酶结构域)的融合蛋白。可用于校正从A到G或C到T的任何单点突变。在第一种情况下,将突变体A脱氨基至I校正突变,在后一种情况下,将与突变体T碱基配对的A脱氨基,然后进行一轮复制,以校正突变。

[0820] 在一些实施方案中,本申请提供了碱基编辑器,其可以有效地在核酸(例如,受试者基因组内的核酸)中产生预期的突变,例如点突变,而无需产生大量的突变。意外突变,例如意外点突变。在一些实施方案中,预期的突变是由与专门设计用于产生预期的突变的指导多核苷酸(例如,gRNA)结合的特定碱基编辑器(例如,胞苷碱基编辑器或腺苷碱基编辑器)产生的突变。在一些实施方案中,预期的突变是与疾病或病症相关的突变。在一些实施方案中,预期的突变是与疾病或病症相关的腺嘌呤(A)至鸟嘌呤(G)点突变。在一些实施方案中,预期的突变是与疾病或病症相关的胞嘧啶(C)至胸腺嘧啶(T)点突变。在一些实施方案中,预期的突变是基因的编码区或非编码区内的腺嘌呤(A)至鸟嘌呤(G)点突变。在一些

实施方案中,预期的突变是基因的编码区或非编码区内的胞嘧啶(C)至胸腺嘧啶(T)点突变。在一些实施方案中,预期的突变是点突变,其在基因的编码区域内产生终止密码子,例如过早的终止密码子。在一些实施方案中,预期的突变是消除终止密码子的突变。

[0821] 在一些实施方案中,本文提供的任何碱基编辑器均能够产生大于1:1的预期突变与非预期突变(例如,预期点突变:非预期点突变)的比率。本文提供的碱基编辑器中的至少一个能够产生预期突变与非预期突变(例如,预期点突变:非预期点突变)的比率至少为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或者更多。

[0822] 碱基编辑器效率的细节在国际PCT申请第PCT/2017/045381号专利申请(WO第201/027078号国际专利申请)和第PCT/US2016/058344号专利申请案(WO第2017/070632号国际专利申请)中进行了描述,将其全部内容通过引用合并在此。另请参阅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);以及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:eaa04774(2017),全文其内容通过引用合并于此。

[0823] 在一些实施方案中,对一个或多个基因中多个核碱基对的编辑导致至少一种预期突变的形成。在一些实施方案中,至少一个预期突变的形成导致引起突变的疾病的精确校正。应当理解,如本文所述的碱基编辑器的多重编辑的特征可以应用于使用本文提供的碱基编辑器的方法的组合中的任何一种。

[0824] 致病突变的精确校正

[0825] 在一些实施方案中,预期的突变是与疾病或病理学相关的基因中的致病突变或致病突变的精确校正。病原性突变可以是病原性单核苷酸多态性(SNP),也可以由SNP引起。例如,致病突变可以是基因编码的蛋白质中的氨基酸变化。在另一个实例中,致病突变可以是基因中的致病SNP。精确的校正可以使病原体突变恢复到其野生型状态。在一些实施方案中,病原性突变是与疾病或病症相关的G→A点突变,其中用A-to-G碱基编辑器(ABE)对突变体A碱基的脱氨基产生与以下目的无关的序列:疾病或失调。在一些实施方案中,致病性突变是C→T点突变。可以例如通过将A-to-G碱基编辑器(ABE)靶向相反链并编辑病原性T核碱基的补体A来校正C→T点突变。在一些实施方案中,病原性突变是与疾病或病症相关的T→C点突变,并且其中用C-to-T碱基编辑器(BE或CBE)对突变体C碱基的脱氨基产生的序列是与疾病或病症无关。在一些实施方案中,致病突变是A→G点突变。可以通过例如将CBE靶向相反链并编辑病原性G核碱基的补体C来校正A→G点突变。非限制性示例性致病突变或致病突变在本文表3A和3B中列出,以及可用于通过将致病突变编辑回其野生型状态来校正突变的碱基编辑器。指示的碱基编辑器可以靶向致病性SNP或致病性SNP的互补序列,描述于Den Dunnen,J.T.和S.E.Antonarakis发表的“突变命名法扩展和描述复杂突变的建议:讨论”。

Human Mutation 15:712(2000),其全部内容通过引用合并于此。

[0826] 表3A:SERPINA1或SERPINC1基因中致病性突变的精确校正

基因	致病突变	碱基编辑器	gRNA 靶向序列	PAM
<i>SERPINA1</i>	E342K	ABE	GACAAGAAAGGGACUGAAGC	NGC
<i>SERPINA1</i>	E342K	ABE	AUCGACAAGAAAGGGACUGA	NGC
SERPINC1	R48C (R79C)	ABE	ACACACCGGUUGGUGGCCUC	NGG

[0828] 表3B:疾病相关基因中致病突变的精确校正

基因	致病突变	碱基编辑器	gRNA 靶向序列	PAM
ABCA4	A1038V	ABE	CUCCAGCUGGACCUCUCCU	GGG
ABCA4	A1038V	ABE	UCCCAGGAGGAGGUCCAGCU	NGA
ABCA4	L541P	CBE	CUCUCCACUGGAGGAAAACA	NGT
ABCA4	G1961E	ABE	CUGUGUGUCGAAGUUCGCCC	TGG
ABCA4	G1961E	ABE	UGUGUGUCGAAGUUCGCCC	GGAG
ABCA4	G1961E	ABE	GUCGAAGUUCGCCCUGGAGA	NGT
ABCA4	G1961E	ABE	UGUCGAAGUUCGCCCUGGAG	NGG
ABCC6	R1141*	ABE	GUUCAGAAUGCCCGGACCAC	NGT
ACADM	K329E	CBE	UCAACUCCAUGCCAUAUUC	NGC
ACADM	K329E	CBE	CUCCAUAUGCCAUAUUCAGCC	NGC
ADA	G216R	ABE	GCCAGGGAGGUGGGCUCGGC	NGA
ADA	G216R	ABE	CCACGCCAGGGAGGUGGGCU	NGG
ADA	Q3*	ABE	UCUAGGCCAUGGUGCCCUCG	NGCG
AGXT	G170R	ABE	GCUUCAGGGAACUCUGCCAC	NGG
ARH	Q136*	ABE	UGCUGAGCUCUGGGCGAUGUA	NGC
ARSA	P426L	ABE	GCAGGGGCUCAUGAGCAGUC	NGA
ARSB	Y210C	CBE	GAACACAUAUUUUUAUAUCC	NGT
ASS	G390R	ABE	AUGCCACCAGGUUCAUAAC	NNNRRT
ATP2A2	N767S	CBE	CGCUGGACGAGAUGAGGUAG	NGG
ATP2A2	N767S	CBE	CCCCGACGCUGGACGAGAUG	NGG
ATP2A2	N767S	CBE	ACGCUGGACGAGAUGAGGUA	NNGRRT

[0830]

基因	致病突变	碱基 编辑 器	gRNA 靶向序列	PAM
CBS	T191M	ABE	GUGGGCAUCCUCACAAUCUC	NGC
CFTR	G551D	ABE	CUGAGUGGAGAUCAACGAGC	NNGRRT
CFTR	W1282*	ABE	CAACAGUGAAGGAAAGCCUU	NGG
CFTR	R553*	ABE	GCUCAUUGACCUCCACUCAG	NNNRRT
CFTR	R117H	ABE	CACUCUAUCGCGAUUUAUCU	NGG
CHM	R270*	ABE	CUCAUCCUUCUCGAAAUGCA	NGA
CHM	A117A	CBE	CUGCAGCGCACCAGCUUCUU	NGA
CLN2	R208*	ABE	UAUCACUUACGGAUCACAGA	NGG
COCH	G88E	ABE	GGGAACCUGUACGAGUCUAU	NGC
CPT2	S113L	ABE	UACCCAAAUGUAGCUUGUA	NGT
CX30	T5M	ABE	UGCAGCAUCCCCCAAUCCAU	NGCG
DFNB59	R183W	ABE	UUACCCACAUUGC U U C C C C U	NGA
E47	E555K	ABE	GCGGAAGCGGGUGCGCGUGC	NGG
F11	E117*	ABE	UUGGCAUUAUUGAGCACUCU	NGG
F11	F283L	CBE	UGAUCUCUUGGGAGAAGAAC	NGG
F5	R506Q	ABE	GGCAAGGAAUACAGGUUUU	NGT
F5	R534Q	CBE	UCCUCGCCUGUCCAGGGAUC	NGC
F7	A294V	ABE	GAGCUCCAGGACCGUGGCGC	NNNRRT
F7	C310F	ABE	GCAGGAAGUCCUGGGUCAUC	NGC
F7	R304Q	ABE	CGUGCCCCAGCUGAUGACCC	NGG
F7	Q100R	CBE	GCAGUACCGCUCACAGCCGC	NGT
F8	R2178C	ABE	GUGCAA AUGCUAUAAUGAGU	NGG

[0831]

基因	致病突变	碱基 编辑 器	gRNA 靶向序列	PAM
F8	R550C	ABE	UAAUAGCAGGUCAGGCACCG	NGG
F9	T342M	ABE	AUGUUCAUGUAUUCUUGUC	NNNRRT
F9	R294Q	ABE	AGCAAAAGCAAAUGUGAUU	NGA
F9	G106S	ABE	AAUGGCAGCAGUUGCAAGGA	NGA
F9	A279T	ABE	GUUGUCACAGGUAAAUACAC	NGA
F9	R294*	ABE	CAUUUCACUUUUGCUCUGUA	NGT
F9	R379Q	ABE	UUGACCAAGCCACAUGUCUU	NGA
FAH	P261L	ABE	CACCACCCACAGAGAGACAG	NGG
FGF23	R176Q	ABE	CGGCAGCACACCCGGAGCGC	NGA
G6PC	Q347*	ABE	GACCUAGGCGAGGCAGUAGG	NGA
G6PC	Q347*	ABE	GGACCUAGGCGAGGCAGUAG	NGG
G6PC	Q347*	ABE	AGGACCUAGGCGAGGCAGUA	NNGRRT
G6PC	R83C	ABE	CAGUAUGGACACUGUCCAAA	NNGRRT
G6PD	S188F	ABE	GGAGAAGAUGUGGUUGGACA	NNNRRT
GALNS	R386C	ABE	UCGCCACAGUAAUAGAAGAU	NGG
GALT	Q188R	CBE	UUACCCGGCAGUGGGGGUGG	NGG
GBA	N370S	CBE	UACAGGAGGCUCUAGGGUAA	NGA
GBA	N370S	CBE	AGGCUCUAGGGUAAGGACAA	NGG
GBA	L444P	CBE	AACGACCCGGACGCAGUGGC	NNNRRT
GBA	L444P	CBE	CGACCCGGACGCAGUGGCAC	NGA
GCDH	M263V	ABE	GAUGAUCACGCCUGUGGCUG	NGG
GCDH	R402W	ABE	GUGCCAGAUCACGUGAUACU	NGT

[0832]

基因	致病突变	碱基 编辑 器	gRNA 靶向序列	PAM
GLDC	A389V	ABE	AUUCACCAAGAGGGCCUAAA	NGA
GLDC	G771R	ABE	CCCAUCAGAGUGUAAGUUCU	NGG
GLDC	T269M	ABE	CCCCUCCAUGUCUGGGUACU	NGA
GUCY2D	R838C	ABE	GCACACGGAGGAGCUGGAGC	NGG
GUSB	L175F	ABE	GUGAAUGUGUUGUUGAUGGC	NNNRRT
HBB	E26K	ABE	UUGGUGGUAAGGCCCUUGGC	NGG
HBB	E26K	ABE	UGGUAAGGCCCUUGGCAGGU	NGG
HBB	E7K	ABE	ACUCCUAAGGAGAAGUCUGC	NGT
HMBS	R173W	ABE	GCCAGGUGUUGAGGUUUC	NGC
HPRT1	R51*	ABE	AUCACAUCUCAAGCAAGACG	NNNRRT
HPRT1	R170*	ABE	UUCAUGGGGUCCUUUUCACC	NGC
IDS	G374G	ABE	UUCUCACCUGCCUCCGGAAG	NGA
IDUA	Q70*	ABE	CUGCUAGUCCCAGCUGAGGA	NGT
IMPDH1	D226N	ABE	ACCAACCUGAAGAAGAACCG	NGA
KCNJ2	R218W	ABE	UUUCCAAAGAUUGCCCACU	NGC
KRT12	L132P	CBE	CAAAAUCCUAAUGAUAGAUU	NGC
LRRK2	G2019S	ABE	ACUACAGCAUUGCUCAGUAC	NGC
MECP2	R106W	ABE		
MECP2	R133C	ABE		
MECP2	R306C	ABE		
MECP2	R168*	ABE		
MECP2	R255*	ABE		

[0833]

基因	致病突变	碱基 编辑 器	gRNA 靶向序列	PAM
NAGLU	R297*	ABE	CUCUCACAGGAAGAGGCUCC	NGA
NAGLU	Y140C	CBE	CACGAAAGAGCAGCUUUGCG	NGC
OPN1LW	C203R	CBE	GACUUCACGCGGCCAGACG	NGT
PAH	R408W	ABE	AAGGGCCAAGGUAUUGUGGC	NGC
PAH	I65T	CBE	ACACUGAAUCUAGACCUUCU	NGT
PAH	R261Q	ABE	CUUCCAAGUCUUCCACUGCA	NNNRRT
PCDH15	R245*	ABE	UGGUGGUUCACCUCUCAUUC	AGAT
PCDH15	R245*	ABE	UGGUGGUGGUUCACCUCUCAU	TCAGAT
PCDH15	R245*	ABE	UUCACCUCUCAUUCAGAUUU	NGG
PDE6A	V685M	ABE	AAAGAUCAUGGAUCAGUCUA	NGA
PDS	L236P	CBE	CAGCCAAAGAUUGUCCUCAA	NGT
PPOX	R59W	ABE	UCCCCAAGGUCCAAGCUCAA	NGA
PRNP	E200K	ABE	CACCAAGACCGACGUUAAGA	NGA
PRNP	M129V	CBE	UUCCCAGCACGUAGCCGCCA	NGG
PRNP	P102L	ABE	CUCAGCUUGUUCCACUGACU	NNGRRT
PRNP	D178N	CBE	GUGCACAACUGCGUCAUAU	NNNRRT
PRPF3	T494M	ABE	ACCUUCAUGGGGUCUUGAAC	NGC
PRPF8	H2309R	CBE	GGGCCUGCGCACCUCGUGGU	NGA
RHO	P347L	ABE	GUCUUAGGCCAGGGCCACCU	NGC
RHO	P347L	ABE	UAGGCCAGGGCCACCUGGCU	NGT
RHO	D190N	ABE	AAUCAACUACUACACGCUCA	NGC
RP1	R667*	ABE	UCAAGAUUUUUUCUUCUUUU	NGC

[0834]

基因	致病突变	碱基编辑器	gRNA 靶向序列	PAM
RPE65	R44*	ABE	ACAUCAAGGAGACUGCCGG	NGA
RPS19	R62Q	ABE	GCGCAGCACCUGUACCUCCG	NGG
RS1	R102W	ABE	CUGUUGAGCCAGGCCUUGUU	NNNRRT
RS1	R141C	ABE	AUGUCACAGCACCCUGGGU	NNGRRT
<i>SERPINA1</i>	E342K	ABE	GACAAGAAAGGGACUGAAGC	NGC
<i>SERPINA1</i>	E342K	ABE	AUCGACAAGAAAGGGACUGA	NGC
SERPINC1	R48C (R79C)	ABE	ACACACCGGUUGGUGGCCUC	NGG
SGSH	R74C	ABE	GGCGCAGCUGGGAGAGCAGC	NGC
SMPD1	L302P	CBE	CACCUGUGAGGAAGUUCCUG	NGG
SNCA	A53T	ABE	UGACAACAGGUAAGCUCCAU	NGT
SOD1	A4V	ABE	CGACCUUCGUCGCCAUAAACU	NGC
SOD1	H46R	CBE	CAUGAACACGGAAUCCAUGC	NGG
SOD1	G37R	ABE	UAAAAGACUGACUGAAGGCC	NGC
TECTA	Y1870C	ABE	UCAUGUAUAAAAACACACUC	NGG
TTR	V50M/V30M	ABE	GGCCAUGCAUGUGUUCAGAA	NGG
USH1C	V72V	ABE	CCAGGUAGAAUAUGAUCAGC	NGA
USH2a	C759F	ABE	GGAUUGAAGAAUUGUUCAC	NGA
MTM1	c.1261-10A>G	CBE	AACUGAUGAAGAUAAUUGU	NNNRRT
PAH	IVS10-11G>A	ABE	UCACUUAGGGCCUACAGUAC	NGC
PDS	IVS8, +1 G>A	ABE	GGGAUGAGUGUGGUGUCCU	NNNRRT
ARSA	c.459+1G>A	ABE	CGACCAGAUAGGAACCAACC	NGG

[0835] *是终止密码子。

[0836] 在一些实施方案中,所述疾病或病症是 α -1抗胰蛋白酶缺乏症(A1AD)。在一些实施方案中,致病性突变在编码A1AT蛋白的SERPINA1基因中。在一些实施方案中,SERPINA1编码的A1AT蛋白的突变是E342K(PiZ等位基因)(图3A)。在一些实施方案中,将SERPINA1等位基

因的位置7的核碱基“A”编辑为“G”,以将PiZ等位基因恢复为野生型等位基因。(图3B和3C)。

[0837] 输送系统

[0838] 本文公开的碱基编辑器可以被编码在病毒载体中包含的核酸上。病毒载体可包括慢病毒、腺病毒、逆转录病毒和腺相关病毒(AAV)。可以根据应用选择病毒载体。例如,AAV由于其温和的免疫原性而通常用于体内基因递送。腺病毒由于其诱导的强烈的免疫原性应答而通常用作疫苗。病毒载体的包装能力会限制可包装到载体中的碱基编辑器的大小。例如,AAV的包装容量约为4.5kb,包括两个145个碱基的反向末端重复序列(ITR)。

[0839] AAV是属于细小病毒家族的小型单链DNA依赖病毒。4.7kb野生型(wt)AAV基因组由分别编码四个复制蛋白和三个衣壳蛋白的两个基因组成,并且在两侧均带有145-bp反向末端重复序列(ITR)。病毒体由三种衣壳蛋白Vp1、Vp2和Vp3组成,它们以1:1:10的比例由同一开放阅读框产生,但由差异剪接(Vp1)和其他翻译起始位点(分别为Vp2和Vp3)产生。Vp3是病毒体中最丰富的亚基,在细胞表面参与受体识别,从而确定病毒的向性。在Vp1的唯一N末端已鉴定出在病毒感染性中起作用的磷脂酶结构域。

[0840] 与野生型AAV类似,重组AAV(rAAV)利用顺式作用145bp的ITR修饰载体转基因盒的侧翼,可提供高达4.5kb的外源DNA包装。感染后,rAAV可以表达本发明的融合蛋白,并且通过游离形式存在于圆环头尾接头中而持续存在而不整合到宿主基因组中。尽管在体外和体内使用所述系统成功进行rAAV成功的例子很多,但当基因编码序列的长度等于或大于wt AAV基因组时,有限的包装能力就限制了AAV介导的基因传递的使用。

[0841] AAV载体的小包装能力使超过所述大小的许多基因的传递和/或使用大的生理调节组件具有挑战性。这些挑战可以通过例如将待递送的蛋白质分成两个或更多个片段来解决,其中N-末端片段与分裂的内含子-N融合,C-末端片段与分裂的内含子-C融合。然后将这些片段包装成两个或多个AAV载体。如本文所用,“内含子”是指连接N-末端和C-末端蛋白质(例如,待连接的片段)侧翼的自剪接蛋白质内含子(例如,肽)。某些内含子用于连接异源蛋白片段的用途描述于例如Wood等人,J.Biol.Chem.289(21):14512-9(2014)。例如,当融合至分离的蛋白质片段时,内含子IntN和IntC相互识别,将自身剪接在一起,并同时连接与它们融合的蛋白质片段的侧翼N和C末端蛋白,从而重建全长来自两个蛋白质片段的蛋白质。其他合适的内含子对于本领域技术人员是显而易见的。

[0842] 本发明的融合蛋白的片段的长度可以变化。在一些实施方案中,蛋白质片段的长度范围为2个氨基酸至约1000个氨基酸。在一些实施方案中,蛋白质片段的长度范围为约5个氨基酸至约500个氨基酸。在一些实施方案中,蛋白质片段的长度范围为约20个氨基酸至约200个氨基酸。在一些实施方案中,蛋白质片段的长度范围为约10个氨基酸至约100个氨基酸。其他长度的合适的蛋白质片段对本领域技术人员是显而易见的。

[0843] 在一些实施方案中,核酸酶(例如,Cas9)的一部分或片段与内含子融合。核酸酶可以融合到内含子的N末端或C末端。在一些实施方案中,融合蛋白的一部分或片段与内含子融合并且与AAV衣壳蛋白融合。内含子,核酸酶和衣壳蛋白可以以任何排列融合在一起(例如,核酸酶-内含子-衣壳,内含子-核酸酶-衣壳,衣壳-内含子-核酸酶等)。在一些实施方案中,将内含子的N端融合至融合蛋白的C端,并且将内含子的C端融合至AAV衣壳蛋白的N端。

[0844] 在一个实施方案中,通过将大的转基因表达盒分成两个分开的两半(5'和3'末端,或头和尾)来产生双重AAV载体,其中盒的每一半被包装在单个AAV载体中(<5kb)。然后,通

过两个双重AAV载体共同感染同一细胞,然后进行以下操作:完成全长转基因表达盒的重组:(1)5'和3'基因组之间的同源重组(HR)(双重AAV重叠向量);(2)ITR介导的5'和3'基因组的尾对头串联连接(双重AAV转拼载体);或(3)这两种机制的组合(双重AAV混合载体)。体内使用双重AAV载体可导致全长蛋白质的表达。双重AAV载体平台的使用代表了大小大于4.7kb的转基因的有效且可行的基因转移策略。

[0845] 所公开的用于设计碱基编辑器的策略可用于产生能够被包装到病毒载体中的碱基编辑器。基于RNA或DNA病毒的系统用于传递碱基编辑器的方法利用了高度进化的过程,所述过程可将病毒靶向培养物中或宿主中的特定细胞,并将病毒有效载荷转运至细胞核或宿主细胞基因组。病毒载体可直接施用于培养中的细胞(患者)(体内),或它们可用于体外处理细胞,并且修饰的细胞可任选地施用于患者(离体)。基于病毒的常规系统可以包括逆转录病毒,慢病毒,腺病毒,腺相关病毒和单纯疱疹病毒载体,用于基因转移。用逆转录病毒,慢病毒和腺相关病毒基因转移方法可以将其整合到宿主基因组中,这通常会导致插入的转基因长期表达。另外,已经在许多不同的细胞类型和靶组织中观察到高转导效率。

[0846] 逆转录病毒的嗜性可以通过掺入外来包膜蛋白,扩大靶细胞的潜在靶群体来改变。慢病毒载体是能够转导或感染非分裂细胞并通常产生高病毒滴度的逆转录病毒载体。因此,逆转录病毒基因转移系统的选择将取决于靶组织。逆转录病毒载体由顺式作用的长末端重复序列组成,其包装能力可容纳6-10kb的外源序列。最小的顺式作用LTR足以复制和包装载体,然后将其用于将治疗性基因整合到靶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)、猿猴免疫缺陷病毒(SIV)、人免疫缺陷病毒(HIV)及其组合的载体(例如,参见Buchscher等人, J.Virol.66:2731-2739(1992);Johann等人, J.Virol.66:1635-1640(1992);Johann等人, J.Virol.66:1635-1640(1992);Sommnerfelt等, Virol.176:58-59(1990);Wilson等人, J.Virol.63:2374-2378(1989);Miller等人, J.Virol.65:2220-2224(1991);国际第PCT/US94/05700号专利申请)。

[0847] 逆转录病毒载体,特别是慢病毒载体,可能需要小于给定长度的多核苷酸序列以有效整合入靶细胞。例如,与较小的逆转录病毒载体相比,长度大于9kb的逆转录病毒载体可导致较低的病毒滴度。在一些方面,本申请的碱基编辑器具有足够的大小,以使得能够有效地包装和经由逆转录病毒载体递送至靶细胞。在一些情况下,碱基编辑器的大小应使得即使与指导核酸和/或可靶向核酸酶系统的其他组分一起表达时也可以有效包装和递送。

[0848] 在优选瞬时表达的应用中,可以使用基于腺病毒的系统。基于腺病毒的载体在许多细胞类型中都具有很高的转导效率,并且不需要细胞分裂。使用这样的载体,已经获得了高滴度和表达水平。所述载体可以在相对简单的系统中大量产生。腺相关病毒("AAV")载体还可用于在靶标核酸和肽的体外生产中,以及在体内和离体基因治疗程序中,用靶核酸转导细胞(例如,参见West等人, Virology 160:38-47(1987);美国第4,797,368号专利;W0第93/24641号专利申请;Kotin, Human Gene Therapy 5:793-801(1994);Muzyczka, J.Clin. Invest.94:1351(1994)。在许多出版物中描述了重组AAV载体的构建,包括美国第5,173,414号专利;Tratschin等,分子细胞生物学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)。

[0849] 因此,本文所述的碱基编辑器可以与病毒载体一起递送。碱基编辑器系统的一个或多个组件可以被编码在一个或多个病毒载体上。例如,碱基编辑器和指导核酸可以被编码在单个病毒载体上。在其他情况下,碱基编辑器和指导核酸被编码在不同的病毒载体上。在任一种情况下,碱基编辑器和指导核酸均可以可操作地连接至启动子和终止子。

[0850] 编码在病毒载体上的成分的组合可以通过所选择的病毒载体的货物大小限制来确定。

[0851] 碱基编辑器的非病毒式输送

[0852] 还提供了针对碱基编辑者的非病毒传递方法。一类重要的非病毒核酸载体是纳米颗粒,其可以是有机或无机的。纳米颗粒是本领域众所周知的。任何合适的纳米颗粒设计可用于递送基因组编辑系统组件或编码此类组件的核酸。例如,有机(例如脂质和/或聚合物)纳米颗粒可以适合用作本申请的某些实施方案中的递送载体。用于纳米颗粒制剂和/或基因转移的示例性脂质示于表4中(以下)。

[0853] 表4

用于基因转移的脂质

脂质	缩写	特征
1,2-二油酰基-sn-甘油-3-磷脂酰胆碱	DOPC	辅助者
1,2-二油酰基-sn-甘油-3-磷脂酰乙醇胺	DOPE	辅助者
胆固醇		辅助者
N-[1-(2,3-二油氧基)预防 N,N,N-三甲基氯化铵	DOTMA	阳离子型
1,2-二油酰氧基-3-三甲基铵丙烷	DOTAP	阳离子型
二十八碳酰胺基糖精	DOGS	阳离子型
N-(3-氨基丙基)-N,N-二甲基-2,3-双(十二烷氧基)-1-溴化丙烷	GAP-DLRIE	阳离子型
十六烷基三甲基溴化铵	CTAB	阳离子型
6-月桂酸己酯	LHON	阳离子型
1-(2,3-二油酰氧基丙基)-2,4,6-三甲基吡啶鎓	2Oc	阳离子型
2,3-二油基氧基-N-[2(精胺羧酰胺基-乙基)-N,N-二甲基-1-丙三氟乙酸铵	DOSPA	阳离子型
1,2-二油基-3-三甲基铵丙烷	DOPA	阳离子型
N-(2-羟乙基)-N,N-二甲基-2,3-双(十四烷氧基)-1-溴化丙烷	MDRIE	阳离子型
二肉豆蔻氧基丙基二甲基羟乙基溴化铵	DMRI	阳离子型
3β-[N-(N',N'-二甲基氨基乙烷)-氨基甲酰基]胆固醇	DC-Chol	阳离子型
(N,N_双胍乙基氨基乙烷)氨基甲酰-胆固醇	BGTC	阳离子型
1,3-二脱氧-2-(6-羧基-亚丙基)-丙基酰胺	DOSPER	阳离子型
二甲基十八烷基溴化铵	DDAB	阳离子型
二十八烷基酰胺基水杨酸亚精胺	DSL	阳离子型
rac-[(2,3-二十八烷基氧丙基)(2-羟乙基)]二甲基氯化铵	CLIP-1	阳离子型
rac-[2(2,3-二十六烷基氧基丙基氧基甲基氧基)乙基]三甲基溴化铵	CLIP-6	阳离子型
乙基二肉豆蔻酰基磷脂酰胆碱	EDMPC	阳离子型
1,2-二硬脂基氧基-N,N-二甲基-3-氨基丙烷	DSDMA	阳离子型
1,2-二肉豆蔻基三甲基铵丙烷	DMTAP	阳离子型
O, O'-二肉豆蔻基-N-赖氨酰天冬氨酸	DMKE	阳离子型
1,2-二硬脂酰基-sn-甘油-3-乙基磷酸胆碱	DSEPC	阳离子型
N-棕榈酰基 D-赤型鞘氨醇氨基甲酰基-亚精胺	CCS	阳离子型
N-叔丁基-N0-十四烷基-3-十四烷基氨基丙酸脒	diC14-amidine	阳离子型

[0854]

用于基因转移的脂质

脂质	缩写	特征
十八碳烯氧基[乙基-2-十七碳烯基-3 羟乙基]咪唑啉鎓氯化物	DOTIM	阳离子型
N1-胆固醇氧羰基-3,7-二氮杂壬烷-1,9-二胺	CDAN	阳离子型
[0855] 2-(3-[双(3-氨基-丙基)-氨基]丙基氨基)-N	RPR209120	阳离子型
二十四烷基氨基甲酰基甲乙乙酰胺		
1,2-二亚油基氧基-3-二甲基氨基丙烷	DLinDMA	阳离子型
2,2-二亚油基-4-二甲基氨基乙基-[1,3]-二氧戊环	DLin-KC2-DMA	阳离子型
二亚油酰基-甲基-4-二甲基氨基丁酸酯	DLin-MC3-DMA	阳离子型

[0856] 表5列出了用于基因转移和/或纳米颗粒制剂的示例性聚合物。

[0857] 表5

用于基因转移的聚合物

聚合物	缩写
聚乙二醇	PEG
聚乙烯亚胺	PEI
二硫代双(琥珀酰亚胺丙酸酯)	DSP
3,3'-二硫代双丙二酸二甲酯	DTBP
聚(亚乙基亚胺)双氨基甲酸酯	PEIC
聚(L-赖氨酸)	PLL
[0858] 组氨酸修饰的 PLL	
聚(N-乙烯基吡咯烷酮)	PVP
聚丙二胺	PPI
聚酰胺基胺	PAMAM
聚(氨基乙亚胺)	SS-PAEI
三亚乙基四胺	TETA
聚(β -氨基酯)	
聚(4-羟基-L-脯氨酸酯)	PHP
聚烯丙胺	
聚(α -[4-氨基丁基]-L-乙醇酸)	PAGA
聚(D,L-乳酸-乙醇酸共聚物)	PLGA

用于基因转移的聚合物

	聚合物	缩写
[0859]	聚(N-乙基-4-乙烯基溴化吡啶鎓)	
	聚磷腈	PPZ
	聚磷酸酯	PPE
	聚氨基磷酸酯	PPA
	聚(N-2-羟丙基甲基丙烯酰胺)	pHPMA
	聚(甲基丙烯酸 2-(二甲氨基)乙酯)	pDMAEMA
	聚(2-氨基乙基丙烯磷酸酯)	PPE-EA
	壳聚糖	
	半乳糖基化壳聚糖	
	N-二酰化壳聚糖	
	组蛋白	
	胶原	
	葡聚糖-精胺	D-SPM

[0860] 表6总结了编码本文所述融合蛋白的多核苷酸的递送方法。

[0861] 表6

	递送	载体/模式	递送至非分		基因组整合	输送的分子类型
			裂细胞	表达时间		
[0862]	物理	(例如电穿孔、粒子枪、磷酸钙转染)	是	短暂的	否	核酸和蛋白质
	病毒式	逆转录病毒	否	稳定	是	RNA
		慢病毒	是	稳定	是/否, 修饰后	RNA
		腺病毒	是	短暂的	否	DNA
		腺相关病毒 (AAV)	是	稳定	否	DNA
		牛痘病毒	是	非常短暂	否	DNA
		单纯疱疹病毒	是	稳定	否	DNA
	非病毒式	阳离子脂质体	是	短暂的	取决于交付的物品	核酸和蛋白质
		聚合纳米粒子	是	短暂的	取决于交付的物品	核酸和蛋白质

	递送	载体/模式	递送至非分		基因组整合	输送的分子类型
			裂细胞	表达时间		
[0863]	生物非病毒式	减毒细菌	是	短暂的	否	核酸
	递送载体	工程噬菌体	是	短暂的	否	核酸
		哺乳动物病毒样颗粒	是	短暂的	否	核酸
		生物脂质体: 红细胞幽灵和外来体	是	短暂的	否	核酸

[0864] 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,例如Cas9或其变体,以及靶向RNA的基因组核酸序列的gRNA的递送可以通过将核糖核蛋白(RNP)传递到细胞来实现这一目标。RNP包含与靶向gRNA复合的核酸结合蛋白,例如Cas9。可以使用已知方法将RNPs递送到细胞中,例如电穿孔、核转染或阳离子脂质介导的方法,例如Zuris, J.A.等人,2015, Nat. Biotechnology, 33(1):73-80。RNP有利于用于CRISPR碱基编辑系统,特别是对于难以转染的细胞,例如原代细胞。另外,RNP还可以减轻细胞中蛋白质表达可能发生的困难,特别是当不能很好地表达在CRISPR质粒中使用的真核启动子例如CMV或EF1A时。有利地,使用RNP不需要将外源DNA递送到细胞中。而且,由于包含核酸结合蛋白和gRNA复合物的RNP随时间降解,因此使用RNP具有限制脱靶作用的潜力。以类似于基于质粒的技术的方式,RNP可用于递送结合蛋白(例如,Cas9变体)并指导同源引导修复(HDR)。

[0865] 用于驱动碱基编辑器编码核酸分子表达的启动子可以包括AAV ITR。这对于消除对额外的启动子组件的需求可能是有利的,所述额外的启动子组件可以占用载体中的空间。释放的额外空间可用于驱动其他元素的表达,例如指导核酸或选择标记。ITR活性相对较弱,因此由于所选核酸酶的过表达,它可用于减少潜在的毒性。

[0866] 任何合适的启动子都可用于驱动碱基编辑器以及在适当的情况下指导核酸的表达。对于普遍表达,可以使用的启动子包括CMV、CAG、CBh、PGK、SV40、铁蛋白重链或轻链等。对于大脑或其他CNS细胞表达,合适的启动子可以包括:用于所有神经元的Synapsin I,用于兴奋性神经元的CaMKII α ,用于GABA能神经元的GAD67或GAD65或VGAT等。对于肝细胞表达,合适的启动子包括白蛋白启动子。对于肺细胞表达,合适的启动子可以包括SP-B。对于内皮细胞,合适的启动子可以包括ICAM。对于造血细胞,合适的启动子可以包括IFN β 或CD45。对于成骨细胞,合适的启动子可以包括OG-2。

[0867] 在一些情况下,本申请内容的碱基编辑器的大小足够小以允许单独的启动子驱动碱基编辑器和同一核酸分子内的兼容性指导核酸的表达。例如,载体或病毒载体可包含可操作地连接至编码碱基编辑器的核酸的第一启动子和可操作地连接至指导核酸的第二启动子。

[0868] 用于驱动指导核酸表达的启动子可包括:Pol III启动子,例如Pol II启动子和内含子盒的U6或H1使用,以表达gRNA腺伴随病毒(AAV)。

[0869] 可以使用腺相关病毒(AAV)、慢病毒、腺病毒或其他质粒或病毒载体类型,尤其是使用例如来自美国第8,454,972号专利(腺病毒的制剂,剂量)、美国专利号第8,404,658号专利(AAV的制剂,剂量)和美国专利号第5,846,946号专利(DNA质粒的制剂,剂量)以及来自

涉及涉及的临床试验的临床试验和出版物慢病毒、AAV和腺病毒。例如,对于AAV、给药途径、制剂和剂量可以如美国专利号第8,454,972号专利中以及涉及AAV的临床试验中所描述。对于腺病毒、给药途径、制剂和剂量可以如美国专利号第8,404,658号专利中以及涉及腺病毒的临床试验中所述。对于质粒递送、给药途径、制剂和剂量可以如美国专利号第5,846,946号专利中以及涉及质粒的临床研究中所述。剂量可以基于或推断为平均70公斤的个体(例如成年男性),并且可以针对患者,受试者,不同体重和物种的哺乳动物进行调整。给药频率在医学或兽医(例如医师,兽医)的能力范围内,这取决于通常的因素,包括年龄、性别、总体健康状况,患者或受试者的其他状况以及要解决的特定状况或症状。可以将病毒载体注射到目标组织中。对于细胞类型特异性碱基编辑,可以由细胞类型特异性启动子驱动碱基编辑器和任选的指导核酸的表达。

[0870] 对于体内递送,AAV可能优于其他病毒载体。在某些情况下,AAV毒性低,这可能是由于纯化方法不需要超离心可激活免疫反应的细胞颗粒所致。在某些情况下,AAV不会引起宿主诱变,因为它没有整合到宿主基因组中。

[0871] AAV的包装极限为4.5或4.75Kb。这意味着公开的碱基编辑器以及启动子和转录终止子可以适合单个病毒载体。大于4.5或4.75Kb的构建体可能导致病毒产生明显减少。例如,SpCas9很大,基因本身超过4.1Kb,这使其很难包装到AAV中。因此,本申请的实施例包括利用公开的碱基编辑器,所述碱基编辑器的长度比常规碱基编辑器的长度短。在某些示例中,碱基编辑器小于4kb。公开的碱基编辑器可以小于4.5kb、4.4kb、4.3kb、4.2kb、4.1kb、4kb、3.9kb、3.8kb、3.7kb、3.6kb、3.5kb、3.4kb、3.3kb、3.2kb、3.1kb、3kb、2.9kb、2.8kb、2.7kb、2.6kb、2.5kb、2kb或1.5kb。在某些情况下,公开的碱基编辑器的长度为4.5kb或更短。

[0872] AAV可以是AAV1、AAV2、AAV5或其任何组合。可以针对要靶向的细胞选择AAV的类型;例如,可以选择AAV血清型1、2、5或混合衣壳AAV1、AAV2、AAV5或其任何组合以靶向脑或神经元细胞;并且可以选择AAV4靶向心脏组织。AAV8可用于递送至肝脏。关于这些细胞的某些AAV血清型的列表可以在Grimm,D.等人,J.Virol.82:5887-5911(2008))中找到。

[0873] 慢病毒是具有在有丝分裂和有丝分裂后细胞中感染并表达其基因的能力的复杂逆转录病毒。最常见的慢病毒是人类免疫缺陷病毒(HIV),它使用其他病毒的包膜糖蛋白靶向广泛的细胞类型。

[0874] 慢病毒可以如下制备。在pCasES10(包含慢病毒转移质粒主链)克隆化后,将低传代率($p=5$)的HEK293FT接种到T-75烧瓶中,直到转染前一天在含10%胎牛血清且无抗生素的DMEM中接种至50%融合。20小时后,将培养基更改为OptiMEM(无血清)培养基,并在4小时后进行转染。用10 μ g慢病毒转移质粒(pCasES10)和以下包装质粒转染细胞:5 μ g pMD2.G(VSV-g假型)和7.5 μ g psPAX2(gag/pol/rev/tat)。可以在带有阳离子脂质输送剂(50 μ l Lipofectamine 2000和100 μ l Plus试剂)的4mL OptiMEM中进行转染。6小时后,将培养基更换为含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,但是无血清方法是首选。

[0875] 慢病毒可以如下纯化。48小时后收集病毒上清液。首先清除上清液中的碎屑,然后通过0.45 μ m低蛋白结合(PVDF)过滤器进行过滤。然后将它们在超速离心机中以24,000rpm旋转2小时。将病毒沉淀在4 $^{\circ}$ C下重悬于50 μ l DMEM中过夜,然后等分并立即冷冻在-80 $^{\circ}$ C下。

[0876] 在另一个实施方案中,还考虑了基于马传染性贫血病毒(EIAV)的最少的非灵长类慢病毒载体。在另一个实施方案中,RetinoStatRTM是一种基于马传染性贫血病毒的慢病毒基因治疗载体,其表达血管抑制蛋白内皮抑素和血管抑素,预期通过视网膜下注射来递送。在另一个实施方案中,考虑使用自我灭活的慢病毒载体。

[0877] 系统的任何RNA,例如指导RNA或编码碱基编辑器的mRNA,都可以RNA的形式递送。可以使用体外转录产生编码碱基编辑器的mRNA。例如,可以使用包含以下组件的PCR盒合成核酸酶mRNA:T7启动子,可选的kozak序列(GCCACC),核酸酶序列和3' UTR,例如来自 β 珠蛋白-polyA尾巴的3' UTR。所述盒可用于通过T7聚合酶转录。指导多核苷酸(例如,gRNA)也可以使用体外转录从含T7启动子的盒中转录,随后是序列“GG”和指导多核苷酸序列。

[0878] 为了增强表达并减少可能的毒性,可以将碱基编辑器编码序列和/或指导核酸进行修饰,以包括一种或多种经修饰的核苷,例如核苷酸。

[0879] 使用伪U或5-甲基C

[0880] 在一些实施方案中,本申请内容包括修饰细胞或生物的方法。所述细胞可以是原核细胞或真核细胞。所述细胞可以是哺乳动物细胞。哺乳动物细胞可以是非人类的灵长类、牛、猪、啮齿动物或小鼠细胞。由本申请内容的碱基编辑器,组合物和方法引入细胞的修饰可以是使得改变细胞和细胞的后代以改善生物产物如抗体、淀粉、酒精或其他所需细胞输出的生产。通过本申请的方法引入细胞的修饰可以是使得细胞和细胞的后代包括改变产生的生物产物的改变。

[0881] 系统可以包括一个或多个不同的载体。一方面,对碱基编辑器进行密码子优化以表达所需的细胞类型,优选为真核细胞,优选为哺乳动物细胞或人细胞。

[0882] 通常,密码子优化是指修饰核酸序列以在目的宿主细胞中增强表达的过程,方法是用所述宿主细胞的基因中更频繁或最常用的密码子替换天然序列的至少一个密码子(例如约或大于约1、2、3、4、5、10、15、20、25、50或更多密码子),同时保持天然氨基酸序列。各种物种对特定氨基酸的某些密码子表现出特定的偏见。密码子偏倚(生物之间密码子使用的差异)通常与信使RNA(mRNA)的翻译效率相关,而信使RNA(mRNA)的翻译效率又被认为尤其取决于翻译的密码子的性质和特定密码子的可用性。转移RNA(tRNA)分子。所选tRNA在细胞中的优势通常反映了肽合成中最常使用的密码子。因此,可以基于密码子优化来定制基因以在给定生物中最佳基因表达。密码子使用表很容易获得,例如,可在www.kazusa.or.jp/codon/上的“Codon Usage Database”(2002年7月9日访问)上找到,并且这些表可以通过多种方式进行修饰。参见Nakamura,Y.等人发表的“Codon usage tabulated from the international DNA sequence databases:status for the year 2000”,Nucl.Acids Res.28:292(2000)。也可获得用于优化特定序列以在特定宿主细胞中表达的密码子的计算机算法,例如Gene Forge(Aptagen;Jacobuss,Pa.)。在一些实施方案中,编码工程化核酸酶的序列中的一个或多个密码子(例如1、2、3、4、5、10、15、20、25、50或更多或所有密码子)对应于最频繁使用的特定氨基酸的密码子。

[0883] 包装细胞通常用于形成能够感染宿主细胞的病毒颗粒。这样的细胞包括包装腺病毒的293细胞和包装逆转录病毒的psi.2细胞或PA317细胞。基因疗法中使用的病毒载体通常是通过产生将核酸载体包装成病毒颗粒的细胞系而产生的。载体通常包含包装和随后整合入宿主所需的最小病毒序列,其他病毒序列被表达盒替换以表达多核苷酸。缺失的病毒

功能通常由包装细胞系反式提供。例如,用于基因治疗的AAV载体通常仅具有来自AAV基因组的ITR序列,其是包装和整合入宿主基因组所需的。病毒DNA可以包装在细胞系中,所述细胞系包含编码其他AAV基因(即rep和cap)但缺少ITR序列的辅助质粒。所述细胞系也可以被腺病毒作为辅助感染。辅助病毒可以促进AAV载体的复制和辅助质粒中AAV基因的表达。在某些情况下,由于缺乏ITR序列,辅助质粒的包装量不大。可以通过例如腺病毒比AAV更敏感的热处理来减少腺病毒的污染。

[0884] 药物成分

[0885] 本申请的其他方面涉及药物组合物,其包含本文所述的任何碱基编辑器,融合蛋白或融合蛋白-指导多核苷酸复合物。如本文所用,术语“药物组合物”是指配制用于药物用途的组合物。在一些实施方案中,药物组合物还包含药学上可接受的载体。在一些实施方案中,药物组合物包含另外的试剂(例如,用于特异性递送,增加的半衰期或其他治疗化合物)。

[0886] 如本文所用,术语“药学上可接受的载体”是指药学上可接受的材料,组合物或媒介物,例如液体或固体填充剂、稀释剂、赋形剂、制造助剂(例如润滑剂、滑石镁、钙或硬脂酸锌(或硬脂酸锌)或溶剂封装材料,涉及将化合物从身体的一个部位(例如,输送部位)携带或运输到另一部位(例如,器官、组织或身体的一部分)。在与制剂的其他成分兼容并且对受试者的组织无害的意义上(例如,生理兼容性、无菌、生理pH等),药学上可接受的载体是“可接受的”。

[0887] 可以用作药学上可接受的载体的材料的一些非限制性实例包括:(1)糖,例如乳糖、葡萄糖和蔗糖;(2)淀粉,例如玉米淀粉和马铃薯淀粉;(3)纤维素及其衍生物,例如羧甲基纤维素钠、甲基纤维素、乙基纤维素、微晶纤维素和乙酸纤维素;(4)黄芪粉;(5)麦芽;(6)明胶;(7)润滑剂,如硬脂酸镁,十二烷基硫酸钠和滑石粉;(8)赋形剂,例如可可脂和栓剂蜡;(9)花生油、棉籽油、红花油、芝麻油、橄榄油、玉米油、大豆油等油;(10)二醇,例如丙二醇;(11)多元醇,例如甘油、山梨糖醇、甘露糖醇和聚乙二醇(PEG);(12)酯,例如油酸乙酯和月桂酸乙酯;(13)琼脂;(14)缓冲剂,例如氢氧化镁和氢氧化铝;(15)海藻酸;(16)无热原水;(17)等渗盐水;(18)林格的解答;(19)乙醇;(20)pH缓冲溶液;(21)聚酯,聚碳酸酯和/或聚酞;(22)填充剂,例如多肽和氨基酸;(23)血清醇,例如乙醇;(23)药物制剂中使用的其他无毒兼容性物质。制剂中还可存在润湿剂、着色剂、脱模剂、包衣剂、甜味剂、调味剂、加香剂、防腐剂和抗氧化剂。诸如“赋形剂”、“载体”、“药学上可接受的载体”、“载体”等的术语在本文中可互换使用。

[0888] 药物组合物可包含一种或多种pH缓冲化合物,以将制剂的pH维持在反映生理pH的预定水平,例如在约5.0至约8.0的范围内。用于水性液体制剂的pH缓冲化合物可以是氨基酸或氨基酸混合物,例如组氨酸、或氨基酸混合物,例如组氨酸和甘氨酸。或者,pH缓冲化合物优选是将制剂的pH保持在预定水平,例如在约5.0至约8.0的范围内并且不螯合钙离子的试剂。这种pH缓冲化合物的说明性实例包括但不限于咪唑和乙酸根离子。pH缓冲化合物可以以适合将制剂的pH维持在预定水平的任何量存在。

[0889] 药物组合物还可包含一种或多种渗透调节剂,即将制剂的渗透特性(例如张度,重量克分子渗透压浓度和/或渗透压)调节至血流可接受的水平的化合物。和接受者的血细胞。渗透调节剂可以是不螯合钙离子的试剂。渗透调节剂可以是本领域技术人员已知或可

获得的调节剂的渗透特性的任何化合物。本领域技术人员可以凭经验确定给定的渗透调节剂用于本发明制剂的适用性。合适类型的渗透调节剂的说明性实例包括但不限于：盐，例如氯化钠和乙酸钠；和盐。糖，如蔗糖，葡萄糖和甘露醇；氨基酸，例如甘氨酸；以及这些试剂和/或试剂类型中的一种或多种的混合物。渗透调节剂可以以足以调节剂的渗透特性的任何浓度存在。

[0890] 在一些实施方案中，配制药物组合物以递送至受试者，例如，用于基因编辑。施用本文所述药物组合物的合适途径包括但不限于：局部、皮下、经皮、皮内、病变内、关节内、腹膜内、膀胱内、经粘膜、牙龈、齿内、耳蜗内、鼓膜内、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。

[0891] 在一些实施方案中，本文所述的药物组合物局部施用于患病部位（例如肿瘤部位）。在一些实施方案中，本文所述的药物组合物通过注射、通过导管、通过栓剂或通过植入物施用于受试者，所述植入物为多孔、无孔或凝胶状材料。包括诸如唾液弹性膜的膜或纤维。

[0892] 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。在一个实施方案中，可以使用泵（例如，参见Langer, 1990, Science 249:1527-1533; Sefton, 1989, CRC Crit. Ref. Biomed. Eng. 14:201; Buchwald等, 1980, Surgery 88:507; Saudek等人, 1989, N. Engl. J. Med. 321:574）。在另一个实施方案中，可以使用聚合物材料。（例如，参见”Medical Applications of Controlled Release”（Langer和Wise编辑, CRC Press, Boca Raton, Fla., 1974）; Controlled Drug Bioavailability, Drug Product Design and Performance（Smolen和Ball编辑, Wiley, New York, 1984; Ranger和Peppas, 1983, Macromol. Sci. Rev. Macromol. Chem. 23:61。另见Levy等人, 1985, Science 228:190; During等人, 1989, Ann. Neurol. 25:351; Howard等人, 1989, J. Neurosurg. 71:105。）其他控释系统例如在上文的Langer中讨论。

[0893] 在一些实施方案中，按照常规程序将药物组合物配制为适于静脉内或皮下施用给受试者例如人的组合物。在一些实施方案中，用于注射给药的药物组合物是无菌等渗用途的溶液作为增溶剂和局部麻醉剂如利多卡因以减轻注射部位的疼痛。通常，将成分分别或以单位剂型混合在一起提供，例如，作为干燥的冻干粉或无水浓缩物在指示活性剂的量的密闭容器中，例如安瓿或小袋中。如果要通过输液方式给药，可以将其分配到装有无菌药物级水或盐水的输液瓶中。在通过注射施用药物组合物的情况下，可以提供安瓿注射用无菌水或盐水，以便可以在施用之前将成分混合。

[0894] 用于全身给药的药物组合物可以是液体，例如无菌盐水、乳酸林格氏液或汉克氏溶液。另外，药物组合物可以是固体形式，并在使用前立即重新溶解或悬浮。也可以考虑冻干形式。药物组合物可以包含在脂质颗粒或囊泡中，例如脂质体或微晶，其也适合于肠胃外给药。颗粒可以具有任何合适的结构，例如单层或多层，只要其中包含组合物即可。可以将化合物截留在“稳定的质粒-脂质颗粒”（SPLP）中，所述颗粒包含融合脂质二油酰磷脂酰乙醇胺（dioleoylphosphatidylethanolamine, DOPE），低含量（5-10mol%）的阳离子脂质，并通过聚乙二醇（PEG）涂层使其稳定（张永鹏等，基因治疗（Gene Ther.）1999, 6:1438-47）。带正电荷的脂质，例如N-[1-(2,3-二羟乙基)丙基]-N,N,N-三甲基-铵甲基硫酸盐或“DOTAP”，对于此类颗粒和囊泡是特别优选的。这种脂质颗粒的制备是众所周知的。例如，参见美国第

4,880,635号专利;第4,906,477号专利;第4,911,928号专利;第4,917,951号专利;第4,920,016号专利;第4,921,757号专利;每一个均通过引用并入本文。

[0895] 例如,本文所述的药物组合物可以作为单位剂量施用或包装。当用于本申请的药物组合物时,术语“单位剂量”是指适合作为对象的单位剂量的物理上离散的单位,每个单位包含预定量的活性物质,所述预定量的活性物质经计算可产生期望的治疗效果并与之结合。所需的稀释剂;即载体或车辆。

[0896] 此外,可以将药物组合物作为药物试剂盒提供,其包括(a)包含冻干形式的本发明化合物的容器,和(b)包含药学上可接受的稀释剂(例如,用于重构或灭菌的无菌容器)的第二容器。本发明的冻干化合物的稀释液,可以任选地与这种容器结合,可以是由政府机构规定的规范药品或生物制品的生产,使用或销售的通知,所述通知反映了所述机构的批准。制造,使用或销售以进行人为管理。

[0897] 另一方面,包括包含用于治疗上述疾病的材料的制品。在一些实施例中,制品包括容器和标签。合适的容器包括例如瓶、小瓶、注射器和试管。容器可以由多种材料形成,例如玻璃或塑料。在一些实施方案中,容器容纳有效治疗本文所述疾病的组合物,并且可以具有无菌入口。例如,容器可以是静脉注射溶液袋或具有可被皮下注射针刺穿的塞子的小瓶。组合物中的活性剂是本发明的化合物。在一些实施方案中,在容器上或与容器相关的标签表明所述组合物用于治疗选择的疾病。所述制品可以进一步包括第二容器,所述第二容器包含药学上可接受的缓冲剂,例如磷酸盐缓冲盐水、林格氏溶液或右旋糖溶液。从商业和用户的角度来看,它还可包括其他材料,包括其他缓冲液、稀释剂、过滤器、针头、注射器和带有使用说明书的包装说明书。

[0898] 在一些实施方案中,本文描述的任何融合蛋白、gRNA和/或复合物作为药物组合物的一部分提供。在一些实施方案中,药物组合物包含本文提供的任何融合蛋白。在一些实施方案中,药物组合物包含本文提供的任何复合物。在一些实施方案中,药物组合物包含核糖核蛋白复合物,其包含与gRNA和阳离子脂质形成复合物的RNA引导的核酸酶(例如,Cas9)。在一些实施方案中,药物组合物包含gRNA,核酸可编程DNA结合蛋白、阳离子脂质和药学上可接受的赋形剂。药物组合物可以任选地包含一种或多种其他治疗活性物质。

[0899] 治疗A1AD的方法

[0900] 还提供了治疗A1AD和/或引起A1AD的SERPINA1中的基因突变的方法,其包括向受试者(例如,哺乳动物,例如人)施用治疗有效量的包含多核苷酸的药物组合物编码此处描述的碱基编辑器系统(例如,碱基编辑器和gRNA)。在一些实施方案中,碱基编辑器是融合蛋白,其包含多核苷酸可编程DNA结合结构域和腺苷脱氨酶结构域或胞苷脱氨酶结构域。用碱基编辑器和靶向所述碱基编辑器的一个或多个指导多核苷酸转导受试者的细胞以实现A • T到G • C的改变(如果所述细胞被腺苷脱氨酶结构域转导)或C • G含有SERPINA1基因突变的核酸序列的U • A改变(如果用胞苷脱氨酶结构域转导细胞)。

[0901] 本文的方法包括对受试者(包括被鉴定为需要这种治疗的受试者,或怀疑有患疾病的风险并且需要这种治疗的受试者)给予有效量的本文所述的组合物。识别需要这种治疗的受试者可以在受试者或卫生保健专业人员的判断中,并且可以是主观的(例如意见)或客观的(例如可以通过测试或诊断方法测量)。

[0902] 治疗方法通常包括给予治疗有效量的药物组合物,所述药物组合物包括例如编码

碱基编辑器的载体和靶向受试者(例如人类患者)的SERPINA1基因的gRNA。需要它。这样的治疗将适当地施用于患有,患有A1AD,易感或有A1AD风险的受试者,特别是人类受试者。本文的组合物也可以用于治疗可能涉及A1AD的任何其他疾病。

[0903] 在一个实施例中,提供了一种监视治疗进度的方法。所述方法包括以下步骤:确定患有或易患与A1AD相关的疾病或其症状的受试者中的诊断标记物(Marker)(例如,与A1AD相关的SNP)或诊断测量值(例如,筛选,测定)的水平。已向受试者施用足以治疗所述疾病或其症状的治疗量的本文组合物。所述方法确定的标志物水平可以与健康正常对照者或其他患病患者的已知标志物水平进行比较,以确定受试者的疾病状况。在优选的实施方案中,在确定第一水平之后的时间点确定受试者中第二水平的标志物,并且比较这两个水平以监测疾病的进程或治疗的功效。在某些优选的实施方案中,在根据本发明开始治疗之前确定受试者中标志物的治疗前水平。然后将所述治疗前的标志物水平与治疗开始后受试者中标志物的水平进行比较,以确定治疗的有效性。

[0904] 在一些实施方案中,将本文提供的组合物施用于受试者,例如人类受试者,以在受试者内实现靶向的基因组修饰。在一些实施方案中,细胞获自受试者,并与本文提供的任何药物组合物接触。在一些实施方案中,任选地在细胞中已经实现或检测到所需的基因组修饰之后,将从受试者中取出并与药物组合物离体接触的细胞重新引入受试者中。递送包含核酸酶的药物组合物的方法是已知的,并且描述于例如美国第6,453,242号专利;第6,503,717号专利;第6,534,261号专利;第6,599,692号专利;第6,607,882号专利;第6,689,558号专利;第6,824,978号专利;第6,933,113号专利;第6,979,539号专利;第7,013,219号专利;和第7,163,824号专利,其全部公开内容通过引用整体并入本文。尽管本文提供的药物组合物的描述主要针对适合于向人给药的药物组合物,但是本领域技术人员应理解,此类组合物通常适合于向各种动物或生物体给药,例如,兽医用途。

[0905] 为使组合物适于向各种动物给药而对适于向人给药的药物组合物的修饰是众所周知的,并且普通熟练的兽医药理学家可以仅通过普通的实验(如果有的话)来设计和/或进行这种修饰。预期给予药物组合物的受试者包括但不限于人和/或其他灵长类;哺乳动物、家养动物、宠物和与商业有关的哺乳动物,例如牛、猪、马、绵羊、猫、狗、小鼠和/或大鼠;和/或鸟类,包括与商业相关的鸟类,例如鸡、鸭、鹅和/或火鸡。

[0906] 本文描述的药物组合物的制剂可以通过药理学领域中已知的或以后开发的任何方法来制备。通常,这样的制备方法包括以下步骤:使一种或多种活性成分与赋形剂和/或一种或多种其他辅助成分结合,然后,如果必要和/或期望的话,将产品成形和/或包装成粉末。所需的单剂量或多剂量单位。药物制剂可以另外包含药学上可接受的赋形剂,如本文所用,其包括任何和所有溶剂、分散介质、稀释剂或其他液体媒介物、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。适用于所需特定剂型的固体粘合剂、润滑剂等。Remington的The Science and Practice of Pharmacy,第21版,A.R.Gennaro (Lippincott,Williams&Wilkins,Baltimore,MD,2006,通过引用整体并入),公开了用于配制药物组合物的各种赋形剂及其制备已知技术。关于用于生产包含核酸酶的药物组合物的其他合适的方法、试剂、赋形剂和溶剂,还参见2010年11月2日提交的国际第PCT/US2010/055131号专利申请(WO第2011/053982A8号国际专利申请)。

[0907] 除非任何常规的赋形剂介质与物质或其衍生物不兼容,例如通过产生任何不希望

的生物学作用或以有害的方式与药物组合物的任何其他组分相互作用,否则预期其用途在本申请的范围内。

[0908] 如上所述的组合物可以有效量施用。有效量将取决于给药方式,所治疗的特定病症和期望的结果。它还可能取决于状况的阶段,受试者的年龄和身体状况,同时治疗的性质(如果有)以及执业医师众所周知的类似因素。对于治疗应用,所述量足以达到医学上理想的结果。

[0909] 在一些实施方案中,根据本发明的组合物可以用于治疗多种疾病、病症和/或病状中的任何一种,包括但不限于以下一种或多种:自身免疫症病(例如,糖尿病、狼疮、多发性硬化症、牛皮癣、类风湿关节炎);炎性疾病(例如关节炎、盆腔炎);传染病(例如病毒感染(例如HIV、HCV、RSV)、细菌感染、真菌感染、败血症);神经系统疾病(例如阿尔茨海默氏病、亨廷顿氏病、自闭症、杜兴氏肌肉营养不良症);心血管疾病(例如动脉粥样硬化、高胆固醇血症、血栓形成,凝血障碍,如黄斑变性的血管生成性疾病);增生性疾病(例如,癌症,良性肿瘤);呼吸系统疾病(例如慢性阻塞性肺疾病);消化系统疾病(例如炎症性肠病,溃疡);肌肉骨骼疾病(例如,纤维肌痛、关节炎);内分泌、代谢和营养失调(例如糖尿病、骨质疏松);泌尿系统疾病(例如肾脏疾病);心理障碍(例如抑郁症、精神分裂症);皮肤疾病(例如伤口、湿疹);血液和淋巴系统疾病(例如贫血、血友病);等等

[0910] 套件

[0911] 本申请的各个方面提供了包括碱基编辑器系统的试剂盒。在一个实施方案中,试剂盒包含核酸构建体,所述核酸构建体包含编码核碱基编辑器融合蛋白的核苷酸序列。融合蛋白包含脱氨酶(例如胞苷脱氨酶或腺嘌呤脱氨酶)和核酸可编程DNA结合蛋白(napDNAbp)。在一些实施方案中,试剂盒包含至少一种能够靶向感兴趣的核酸分子例如A1AD相关突变的指导RNA。在一些实施方案中,试剂盒包含核酸构建体、所述核酸构建体包含编码至少一个指导RNA的核苷酸序列。

[0912] 在一些实施方案中,试剂盒提供了使用所述试剂盒编辑一个或多个与A1AD相关的突变的说明。说明书通常将包括有关使用试剂盒编辑核酸分子的信息。在其他实施例中,所述指令包括以下至少之一:注意事项;警告;临床研究;和/或参考。这些说明可以直接印在容器上(如果有),也可以贴在容器上的标签,也可以印在容器内或随容器提供的单独的纸页、小册子、卡片或文件夹中。在另一个实施方案中,试剂盒可以包括标签或单独的插入物(包装插入物)形式的用于合适的操作参数的说明书。在又一个实施方案中,试剂盒可包含一个或多个带有适当的阳性和阴性对照或对照样品的容器,以用作检测、校准或归一化的标准。所述试剂盒可以进一步包括第二容器,所述第二容器包含药学上可接受的缓冲液,例如(无菌的)磷酸盐缓冲盐水、林格氏溶液或右旋糖溶液。从商业和用户的角度来看,它还可包括其他材料,包括其他缓冲液、稀释剂、过滤器、针头、注射器和带有使用说明的包装说明书。

[0913] 在某些实施方案中,所述试剂盒可用于治疗患有 α -1抗胰蛋白酶缺乏症(A1AD)的受试者。

[0914] 本文设想了涵盖碱基编辑器系统和用途的方法和组成的以下编号的附加实施例:

[0915] 1.一种对有此需要的受试者中治疗 α -1抗胰蛋白酶缺乏症(A1AD)的方法,包括向所述受试者施用碱基编辑器系统,所述系统包括

- [0916] 指导多核苷酸或编码所述指导多核苷酸的核酸；
- [0917] 多核苷酸可编程DNA结合结构域或编码多核苷酸可编程DNA结合结构域的核酸，以及
- [0918] 腺苷脱氨酶结构域或编码腺苷脱氨酶结构域的核酸，
- [0919] 其中所述指导多核苷酸能够靶向所述碱基编辑器系统，以实现所述受试者细胞的SERPINA1多核苷酸中的A • T到G • C单核苷酸多态性 (SNP) 的改变，从而治疗A1AD；
- [0920] 其中SNP是A1AD的病因。
- [0921] 2. 一种对此需要的受试者中治疗 α -1抗胰蛋白酶缺乏症 (A1AD) 的方法，包括
- [0922] (a) 将碱基编辑器系统引入单元格，所述系统包括
- [0923] 指导多核苷酸或编码所述指导多核苷酸的核酸；
- [0924] 多核苷酸可编程DNA结合结构域或编码多核苷酸可编程DNA结合结构域的核酸，和
- [0925] 腺苷脱氨酶结构域或编码腺苷脱氨酶结构域的核酸，以及
- [0926] (b) 将细胞施用于受试者，
- [0927] 其中所述指导多核苷酸能够靶向所述碱基编辑器系统，以实现所述细胞中SERPINA1多核苷酸中单核苷酸多态性 (SNP) 的A • T至G • C改变，从而治疗A1AD；
- [0928] 其中SNP是A1AD的病因。
- [0929] 3. 根据实施方案2的方法，其中所述细胞是肝细胞或其祖细胞。
- [0930] 4. 根据实施方案2或3中任一项所述的方法，其中所述细胞对于所述受试者是自体的、同种异体的或异种的。
- [0931] 5. 一种校正SERPINA1多核苷酸中引起 α -1抗胰蛋白酶缺乏 (A1AD) 的单核苷酸多态性 (SNP) 的方法，所述方法包括使SERPINA1多核苷酸与碱基编辑器系统接触，所述碱基编辑器系统包括：
- [0932] 指导多核苷酸；
- [0933] 多核苷酸可编程DNA结合域，以及
- [0934] 腺苷脱氨酶结构域，
- [0935] 其中所述指导多核苷酸能够靶向所述碱基编辑器系统以实现SERPINA1多核苷酸中SNP的A • T至G • C，从而校正所述SNP。
- [0936] 6. 一种生产用于治疗 α -1抗胰蛋白酶缺乏症 (A1AD) 的修饰细胞的方法，包括将碱基编辑器系统引入细胞中，所述系统包括
- [0937] 指导多核苷酸或编码所述指导多核苷酸的核酸；
- [0938] 多核苷酸可编程DNA结合结构域或编码多核苷酸可编程DNA结合结构域的核酸，以及
- [0939] 腺苷脱氨酶结构域或编码腺苷脱氨酶结构域的核酸，
- [0940] 其中所述指导多核苷酸能够靶向所述碱基编辑器系统，以引起细胞中SERPINA1多核苷酸中引起A1AD的单核苷酸多态性 (SNP) 的A • T到G • C改变。
- [0941] 7. 根据实施方案6的方法，其中所述引入是体内的。
- [0942] 8. 根据实施方案6的方法，其中所述引入是离体的。
- [0943] 9. 根据实施方案6至8中任一项的方法，其中所述细胞是其祖细胞的肝细胞。
- [0944] 10. 根据实施方案6至9中任一项的方法，其中所述细胞获自患有A1AD的受试者。

[0945] 11. 前述实施方案中任一项的方法, 其中SERPINA1多核苷酸编码由SNP产生的在位置342处包含赖氨酸的A1AT蛋白。

[0946] 12. 实施方案11的方法, 其中所述A • T至G • C改变用野生型氨基酸替代赖氨酸。

[0947] 13. 一种在有此需要的受试者中治疗 α -1抗胰蛋白酶缺乏症 (A1AD) 的方法, 包括向所述受试者施用包括以下内容的碱基编辑器系统:

[0948] 指导多核苷酸或编码所述指导多核苷酸的核酸;

[0949] 多核苷酸可编程DNA结合结构域或编码多核苷酸可编程DNA结合结构域的核酸, 以及

[0950] 腺苷脱氨酶结构域或编码腺苷脱氨酶结构域的核酸,

[0951] 其中所述指导多核苷酸能够靶向所述碱基编辑器系统以引起所述对象细胞的SERPINA1多核苷酸中导致A1AD的单核苷酸多态性 (SNP) 的A • T至G • C改变,

[0952] 其中SERPINA1多核苷酸编码由SNP产生的在位置342处包含赖氨酸的A1AT蛋白,

[0953] 其中A • T到G • C的改变用野生型氨基酸取代赖氨酸, 从而治疗A1AD。

[0954] 14. 一种在有此需要的受试者中治疗 α -1抗胰蛋白酶缺乏症 (A1AD) 的方法, 包括

[0955] (a) 使单元与碱基编辑器系统联系, 所述碱基编辑器系统包括指导多核苷酸或编码所述指导多核苷酸的核酸;

[0956] 多核苷酸可编程DNA结合结构域或编码多核苷酸可编程DNA结合结构域的核酸, 以及

[0957] 腺苷脱氨酶结构域或编码腺苷脱氨酶结构域的核酸,

[0958] (b) 将细胞施用于受试者,

[0959] 其中所述指导多核苷酸能够靶向所述碱基编辑系统, 以引起所述细胞SERPINA1多核苷酸中导致A1AD的单核苷酸多态性 (SNP) 从A • T到G • C的改变,

[0960] 其中SERPINA1多核苷酸编码由SNP产生的在位置342处包含赖氨酸的A1AT蛋白,

[0961] 其中A • T到G • C的改变用野生型氨基酸取代赖氨酸, 从而治疗A1AD。

[0962] 15. 根据实施方案14的方法, 其中所述细胞是肝细胞或其祖细胞。

[0963] 16. 根据实施方案14或15的方法, 其中所述细胞对于所述受试者是自体的、同种异体的或异种的。

[0964] 17. 一种校正SERPINA1多核苷酸中引起 α -1抗胰蛋白酶缺乏 (A1AD) 的单核苷酸多态性 (SNP) 的方法, 所述方法包括使SERPINA1多核苷酸与包含以下内容的碱基编辑器系统接触:

[0965] 指导多核苷酸;

[0966] 多核苷酸可编程DNA结合域, 以及

[0967] 腺苷脱氨酶结构域,

[0968] 其中所述指导多核苷酸能够靶向所述碱基编辑系统以实现所述SNP的A • T至G • C,

[0969] 其中SERPINA1多核苷酸编码由SNP产生的在位置342处包含赖氨酸的A1AT蛋白,

[0970] 其中A • T到G • C的改变用野生型氨基酸取代赖氨酸, 从而校正SNP。

[0971] 18. 一种生产用于治疗A1AD的修饰细胞的方法, 其包括将碱基编辑器系统引入细胞中, 所述碱基编辑器系统包括:

- [0972] 指导多核苷酸或编码所述指导多核苷酸的核酸；
- [0973] 多核苷酸可编程DNA结合结构域或编码多核苷酸可编程DNA结合结构域的核酸，以及
- [0974] 腺苷脱氨酶结构域或编码腺苷脱氨酶结构域的核酸，
- [0975] 其中所述指导多核苷酸能够靶向所述碱基编辑系统，以引起所述细胞SERPINA1多核苷酸中导致A1AD的单核苷酸多态性 (SNP) 从A • T到G • C的改变，
- [0976] 其中SERPINA1多核苷酸编码由SNP产生的在位置342处包含赖氨酸的A1AT蛋白，
- [0977] 其中A • T到G • C的改变用野生型氨基酸取代赖氨酸。
- [0978] 19. 根据实施方案18的方法，其中所述引入是体内的。
- [0979] 20. 根据实施方案18的方法，其中所述引入是离体的。
- [0980] 21. 根据实施方案18至20中任一项的方法，其中所述细胞是肝细胞或其祖细胞。
- [0981] 22. 根据实施方案18至21中任一项的方法，其中所述细胞获自患有A1AD的受试者。
- [0982] 23. 根据实施方案12至22中任一项的方法，其中所述野生型氨基酸是谷氨酸。
- [0983] 24. 根据前述实施方案中任一项的方法，其中所述多核苷酸可编程DNA结合结构域是Cas9结构域。
- [0984] 25. 根据实施方案24的方法，其中所述Cas9结构域是核酸酶失活的Cas9结构域。
- [0985] 26. 根据实施方案24的方法，其中所述Cas9结构域是Cas9切口酶结构域。
- [0986] 27. 根据实施方案24至26中任一项的方法，其中所述Cas9结构域包含SpCas9结构域。
- [0987] 28. 根据实施方案27的方法，其中所述SpCas9结构域包含D10A和/或H840A氨基酸取代或其相应的氨基酸取代。
- [0988] 29. 根据实施方案27或28的方法，其中所述SpCas9结构域对NGG PAM具有特异性。
- [0989] 30. 根据实施方案27至29中任一项的方法，其中所述SpCas9结构域对NGA PAM, NGT PAM或NGC PAM具有特异性。
- [0990] 31. 根据实施方案27至30中任一项的方法，其中所述SpCas9结构域包含氨基酸取代L1111R、D1135V、G1218R、E1219F、A1322R、R1335V、T1337R和L1111、D1135L、S1136R、G1218S、E1219V、D1332A、R1335Q、T1337I、T1337V、T1337F和T1337M中的一个或多个，或其相应的氨基酸取代。
- [0991] 32. 根据实施方案27至31中任一项的方法，其中所述SpCas9结构域包含氨基酸取代L1111R、D1135V、G1218R、E1219F、A1322R、R1335V、T1337R以及L1111、D1135L、S1136R、G1218S、E1219V、D1332A、D1332S、D1332T、D1332V、D1332L、D1332K、D1332R、R1335Q、T1337I、T1337V、T1337F、T1337S、T1337N、T1337K、T1337R、T1337H、T1337Q和T1337M中的一个或多个，或相应的氨基酸取代。
- [0992] 33. 根据实施方案27至32中任一项的方法，其中所述SpCas9结构域包含氨基酸取代D1135L、S1136R、G1218S、E1219V、A1322R、R1335Q、T1337和A1322R，以及L1111、D1135L、S1136R、G1218S、E1219V、D1332A、D1332S、D1332T、D1332V、D1332L、D1332K、D1332R、R1335Q、T1337I、T1337V、T1337F、T1337S、T1337N、T1337K、T1337R、T1337H、T1337Q或T1337M中的一个或多个，或其相应的氨基酸取代。
- [0993] 34. 根据实施方案27至33中任一项的方法，其中所述SpCas9结构域包含氨基酸取

代D1135M、S1136Q、G1218K、E1219F、A1322R、D1332A、R1335E和T1337R,或其相应的氨基酸取代。

[0994] 35. 根据实施方案27至34中任一项的方法,其中所述SpCas9结构域对NG PAM、NGG PAM、GAA PAM、GAT PAM或CAA PAM具有特异性。

[0995] 36. 根据实施方案35的方法,其中所述SpCas9结构域包含氨基酸取代E480K、E543K和E1219V或其相应的氨基酸取代。

[0996] 37. 根据实施方案27至29中任一项的方法,其中所述Cas9结构域包含SaCas9结构域。

[0997] 38. 根据实施例27所述的方法,其中所述SaCas9结构域对NNRRT PAM具有特异性。

[0998] 39. 根据实施例38所述的方法,其中所述SaCas9结构域对NGRRT PAM具有特异性。

[0999] 40. 根据实施方案37至39中任一项的方法,其中所述SaCas9结构域包含氨基酸取代N579A或其相应的氨基酸取代。

[1000] 41. 根据实施方案37至40中任一项的方法,其中所述SaCas9结构域包含氨基酸取代E782K、N968K和R1015H或其相应的氨基酸取代。

[1001] 42. 根据实施方案27至29中任一项的方法,其中所述Cas9结构域包括St1Cas9结构域:

[1002] 43. 根据实施例40所述的方法,其中所述St1Cas9结构域对NNACCA PAM具有特异性。

[1003] 44. 根据前述实施方案中任一项的方法,其中所述腺苷脱氨酶结构域是自然界中不存在的修饰的腺苷脱氨酶结构域。

[1004] 45. 根据实施方案44的方法,其中所述腺苷脱氨酶结构域包含TadA结构域。

[1005] 46. 根据实施方案45的方法,其中TadA结构域包含TadA 7.10的氨基酸序列。

[1006] 47. 根据前述实施例中任一项的方法,其中所述碱基编辑器系统进一步包括锌指结构域。

[1007] 48. 根据实施方案47的方法,其中所述锌指结构域包含识别螺旋序列RNEHLEV, QSTTLKR和RTEHLAR或识别螺旋序列RGEHLRQ, QSGTLKR和RNDKLVP。

[1008] 49. 根据实施方案47或48的方法,其中所述锌指结构域是zf1ra或zf1rb。

[1009] 50. 根据前述实施例中任一项所述的方法,其中,所述碱基编辑器系统还包括核定位信号(NLS)。

[1010] 51. 根据前述实施例中任一项所述的方法,其中,所述碱基编辑器系统还包括一个或多个链接器。

[1011] 52. 根据实施方案51的方法,其中所述多核苷酸可编程DNA结合结构域、腺苷脱氨酶结构域,锌指结构域和NLS中的两个或更多个经由接头连接。

[1012] 53. 根据实施方案52的方法,其中所述接头是肽接头,从而形成碱基编辑融合蛋白。

[1013] 54. 根据实施方案53的方法,其中所述肽接头包含选自以下的氨基酸序列:

[1014] SGGSSGSETPGTSESATPESSGGS,

[1015] SGGSSGGSSGSETPGTSESATPESSGGSSGGS,

[1016] GGGSGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPG

TSTEPSEGSAPGTSESATPESGPGSEPATSGGSGGS,

[1017] SGGSSGGSSGSETPGTSESATPES,

[1018] SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGS,

[1019] SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGSSGSETPGTSESATPESSGGSSGGS,

[1020] PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATS, (SGGS)_n, (GGGS)_n, (GGGGS)_n, (G)_n, (EAAAK)_n, (GGS)_n, SGSETPGTSESATPES, and (XP)_n.

[1021] 55. 根据实施方案53或54的方法, 其中所述碱基编辑融合蛋白包含选自以下的氨基酸序列:

[1022] MPKKKRKVSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRQEIKAQKKAQSSTDSSGGSSGGSSGSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVHNNRVIGEGWNRRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSSGGSSGGSSGSETPGTSESATPESDLVLGLAIGIGSVGVGILNKVTGEIHKNSRIFPAAQAENNLVRRTNRQGRRLARRKKHRRVRLNRLFEEGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLLDASDDGNSVGDYAIQIVKENSQKLETKTPGQIQLERYQTYGQLRGDFTVEKDGGKKHRLINVFPTSAYRSEALRILQTQQEFNPQITDEFINRYLEILTGKRKYHGPNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEFRAAKASYTAQEFNLLNDLNNLTVPPTETKKLSKEQKNQIINYVKNKAMGPAKLFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETLDKLAYVLTNTEREGIQEALHEFADGSFSQKQVDELVQFRKANSSIFGKGWHNFSVKLMME LIPELYETSEEQMTILTRLGKQKTSSSNKTKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKANKDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTYPYQALDSMDDAWSFRELKAFVRESKTLNKKKE YLLTEEDISKFDVRKKFIERNLVDTLYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTHHH AVDALIIAASSQLNLWKKQKNTLVSYSEDQLLDIETGELISDDEYKESVFKAPYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKADETYVLGKIKDIYTDQGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPIL ENYPNKQINDKGKEVPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLGNHIDITPKDSNNKVVLQSVSPWR ADVYFNKTTGKYEILGLKYADLQFDKGTGYKISQEKYNDIKKKEGVSDSEFKFTLYKNDLLL VKDTETKEQQLF RFLSRTMPKQKHVELKPYDKQKFEGGEALIKVLGNVANSGQCKKGLGKSNISYKVRTDVLGNQHIKNEGDKPK LDFPKKKRKVEGADKRTADGSEFESPKKKRKV, MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGE GWNRPPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMD VLHHPGMNHRVEITEGILADECAALLSDFFRMRQEIKAQKKAQSSTDSSGGSSGGSSGSETPGTSESATPESSGGS SGGSSSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVHNNRVIGEGWNRRAIGLHDPTAHAEIMALRQGGLVMQNY RLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFF RMPRQVFNAQKKAQSSTDSSGGSSGGSSGSETPGTSESATPESSGGSSGGSKRNYILGLAIGITSVGYGIIDYETRD VIDAGVRLFKEANVENNEGRRSKRGRARRLRRRRRHRIQRVKLLFDYNLLTDHSELGINPYEARVKGLSQKLSSEE EFS AALLHLAKRRGVHNVNEVEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTS DYVKE AKQLLKVKQAYHQLDQSFIDTYIDLLETRRTYYEGPGEPSFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADL YNALNDLNNLVITRDENEKLEYEKFQI IENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHD

IKDITARKEI IENAELLDQIAKILTIYQSSEDIQEELTNLNSLTQEEIEQISNLKGYTGTHNLSLKAINLILDEL
WHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAI IKKYGLPNDII IELAREKN
SKDAQKMINEMQKRNRQTNERIEEI IRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNPNFYEV DHI I
PRSVSFDNSFNKVLVKQEENSKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGGRISKTKKEYLLEERDINRF
SVQKDFINRNLVDTRYATRGLMNLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHAEDALI IANA
DFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETE QEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDT
LYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHPDQTYQKLKLIMEQYGDEKNPLYKYEEETGN
YLTKYSKKDNGPVIKKIKYYGNKLN AHL DITDDYPNSRNKVVKLSLKPYPFDVYLDNGVYKFVTVKNLDVIKKENY
YEVNSKCYEEAKKLKKSINQAEFIASFYKNDLIKINGEL YRVIGVNNDLLNRIEVMIDITYREYLENMNDKRPPH
IIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKGEGADKRTADGSEFESPKKKRKV,

[1023] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAAL
LSDFFRMRRQEIKAQKKAQSSTDSSGSSGGSSGSETPGTSESATPESSGGSSGSSSEVEFSHEYWMRHALTLAKRA
RDEREVPVGAVLVHNNRVIGEGWNRRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSR
IGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSSGSSGG
SSGSETPGTSESATPESSGGSSGSSSKRNYILGLAIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRG
ARRLKRRRRHRIQRVKLLFDYNLLTDHSELSGINPYEARVKLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDT
GNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTS DYVKEAKQLLKVQKAYHQLDQSFIDTYIDL
LETRRTYYEGPGEESPFGWKDIKEWYMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYEYK
FQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEI IENAELLDQIAKILT
IYQSSEDIQEELTNLNSLTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQ
QKEIPTTLVDDFILSPVVKRSFIQSIKVINAI IKKYGLPNDII IELAREKNSKDAQKMINEMQKRNRQTNERIEEI
IRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNPNFYEV DHI IPRSVSFDNSFNKVLVKQEENSKKG
NRTPFQYLSSSDSKISYETFKKHILNLAKGGRISKTKKEYLLEERDINRF SVQKDFINRNLVDTRYATRGLMNL
RSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHAEDALI IANADFIKKEWKKLDKAKKVMENQMFEEK
QAESMPEIETE QEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIVNNLNGLYDKD
NDKLKKLINKSPEKLLMYHHPDQTYQKLKLIMEQYGDEKNPLYKYEEETGNLYLTKYSKKDNGPVIKKIKYYGNKLN
AHL DITDDYPNSRNKVVKLSLKPYPFDVYLDNGVYKFVTVKNLDVIKKENY YEVNSKCYEEAKKLKKSINQAEFIA
SFYKNDLIKINGEL YRVIGVNNDLLNRIEVMIDITYREYLENMNDKRPPHIIKTIASKTQSIKKYSTDILGNLYE
VSKSKHPQIIKKGEGADKRTADGSEFESPKKKRKVSSGNSNANSRGPSFSSGLVPLSLRGS SRPGERPFQCRICM
RNFSRNEHLEVHTRTHTGEKPFQCRICMRNFSQSTTLKRHLRTHTGEKPFQCRICMRNFSRTEHLARHLKTHLRGS
SAQ, or

[1024] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAAL
LSDFFRMRRQEIKAQKKAQSSTDSSGSSGGSSGSETPGTSESATPESSGGSSGSSSEVEFSHEYWMRHALTLAKRA
RDEREVPVGAVLVHNNRVIGEGWNRRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSR
IGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSSGSSGG
SSGSETPGTSESATPESSGGSSGSSSKRNYILGLAIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRG
ARRLKRRRRHRIQRVKLLFDYNLLTDHSELSGINPYEARVKLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDT

GNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSYVKEAKQLLKVQKAYHQLDQSFIDTYIDL
LETRRTYYEGPGEGSPFGWKDIKEWYMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYEYK
FQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILT
IYQSSEDIQEELTNLNSLTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQ
QKEIPTTLVDDFILSPVVKRSFIQSIKVINAIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNQRTNERIEEI
IRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNKVLVKQEENSKKG
NRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNL
RSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHAEDALIIANADFIKKEWKLDKAKVMENQMFEEK
QAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKSHRVDKKPNRKLINDTLYSTRKDDKGNTLIVNNLNGLYDKD
NDKLKLINKSPEKLLMYHHPQTYQKLKLIMEQYGDEKNPLYKYEEETGNYLTKYSKKDNGPVIKKIKYYGNKLN
AHLDTDDYPNSRNKVVKLSLKPYPFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCY EEAKKKKISNQAEFI
ASFYKNDLIKINGELYRVIGVNNDLLNRIEVMIDITYREYLENMNDKRPPHIKTIASKTQSIKKYSTDILGNLY
EVKSKKHPQIIKKGEGADKRTADGSEFESPKKKRKVSSGNSNANSRGPSSFSSGLVPLSLRGSHSRPGERPFQCRIC
MRNFSRGEHLRQHTRTHTGEKPFQCRICMRNFSQSGTLKRHLRTHTEKPFQCRICMRNFSRNDKLVPHLKTHLRG
SSAQ.

[1025] 56. 根据前述实施方案中任一项的方法, 其中所述指导多核苷酸包含两个单独的多核苷酸, 其中所述两个单独的多核苷酸是两个DNA、两个RNA或DNA和RNA。

[1026] 57. 根据前述实施方案中任一项的方法, 其中所述指导多核苷酸包含crRNA和tracrRNA, 其中所述crRNA包含与SERPINA1多核苷酸中的靶序列互补的核酸序列。

[1027] 58. 根据实施方案57的方法, 其中靶序列包括选自GACAAGAAAGGGACTGAAGC, ATCAGACAAGAAAGGGACTGA和ACACACCGTTGGTGGGGCCTC或其互补序列的序列。

[1028] 59. 根据实施方案57或58的方法, 其中所述碱基编辑器系统包括单链指导RNA (sgRNA)。

[1029] 60. 根据实施方案59的方法, 其中所述sgRNA包含选自ACTCTaGGCAGAGGTCTCAAAAAA和GCTCTaGGCCGAAGTGTCGAGG的序列。

[1030] 61. 根据前述实施方案中任一项的方法, 其中所述碱基编辑器系统包含载体, 所述载体包含指导多核苷酸、多核苷酸可编程DNA结合结构域和脱氨酶结构域中的一个或多个。

[1031] 62. 根据实施方案61的方法, 其中所述载体是腺病毒载体、AAV载体、慢病毒载体或逆转录病毒载体。

[1032] 63. 一种包括碱基编辑器系统的修饰单元, 所述碱基编辑器系统包括

[1033] 指导多核苷酸或编码所述指导多核苷酸的核酸;

[1034] 多核苷酸可编程DNA结合结构域或编码多核苷酸可编程DNA结合结构域的核酸, 以及

[1035] 腺苷脱氨酶结构域或编码腺苷脱氨酶结构域的核酸,

[1036] 其中所述指导多核苷酸能够靶向所述碱基编辑器系统, 以引起所述细胞中SERPINA1多核苷酸中引起 α -1抗胰蛋白酶缺乏 (A1AD) 的单核苷酸多态性 (SNP) 的A • T至G • C改变。

[1037] 64. 一种包括碱基编辑器系统的修饰单元, 所述碱基编辑器系统包括

[1038] 指导多核苷酸或编码所述指导多核苷酸的核酸;

- [1039] 多核苷酸可编程DNA结合结构域或编码多核苷酸可编程DNA结合结构域的核酸,以及
- [1040] 腺苷脱氨酶结构域或编码腺苷脱氨酶结构域的核酸,
- [1041] 其中所述指导多核苷酸能够靶向所述碱基编辑器系统,以引起所述细胞中SERPINA1多核苷酸中引起 α -1抗胰蛋白酶缺乏(A1AD)的单核苷酸多态性(SNP)的A·T到G·C改变,
- [1042] 其中SERPINA1多核苷酸编码由SNP产生的在位置324处包含赖氨酸的A1AT蛋白,
- [1043] 其中A·T到G·C的改变用野生型氨基酸取代赖氨酸。
- [1044] 65.根据实施方案63的修饰的细胞,其中SERPINA1多核苷酸编码由SNP产生的在位置342处包含赖氨酸的A1AT蛋白。
- [1045] 66.根据实施方案65的修饰的细胞,其中所述A·T至G·C改变用野生型氨基酸替代赖氨酸。
- [1046] 67.根据实施方案63至66中任一项的修饰的细胞,其中所述细胞是肝细胞或其祖细胞。
- [1047] 68.根据实施方案67的经修饰的细胞,其中所述细胞获自患有A1AD的受试者。
- [1048] 69.根据实施方案66至68中任一项的修饰的细胞,其中所述野生型氨基酸是谷氨酸。
- [1049] 70.根据实施方案63至66中任一项的修饰的细胞,其中所述多核苷酸可编程DNA结合结构域是Cas9结构域。
- [1050] 71.实施方案70的修饰的细胞,其中所述Cas9结构域是核酸酶失活的Cas9结构域。
- [1051] 72.根据实施方案71的修饰的细胞,其中所述Cas9结构域是Cas9切口酶结构域。
- [1052] 73.根据实施方案70至72中任一项的修饰的细胞,其中所述Cas9结构域包含SpCas9结构域。
- [1053] 74.根据实施方案73的修饰的细胞,其中所述SpCas9结构域包含D10A和/或H840A氨基酸取代或其相应的氨基酸取代。
- [1054] 75.根据实施方案73或74的修饰的细胞,其中所述SpCas9结构域对NGG PAM具有特异性。
- [1055] 76.根据实施方案73至75中任一项的修饰的细胞,其中所述SpCas9结构域对NGA PAM、NGT PAM或NGC PAM具有特异性。
- [1056] 77.根据实施方案73至76中任一项的修饰的细胞,其中所述SpCas9结构域包含氨基酸取代L1111R、D1135V、G1218R、E1219F、A1322R、R1335V、T1337R和一个或多个L1111、D1135L、S1136R、G1218S、E1219V、D1332A、R1335Q、T1337I、T1337V、T1337F和T1337M或其相应的氨基酸取代。
- [1057] 78.根据实施方案73至76中任一项的修饰的细胞,其中所述SpCas9结构域包含氨基酸取代L1111R、D1135V、G1218R、E1219F、A1322R、R1335V、T1337R和L1111、D1135L、S1136R、G1218S、E1219V、D1332A、D1332S、D1332T、D1332V、D1332L、D1332K、D1332R、R1335Q、T1337I、T1337V、T1337F、T1337S、T1337N、T1337K、T1337R、T1337H、T1337Q和T1337M或相应的氨基酸取代。
- [1058] 79.根据实施方案73至76任一项的修饰的细胞,其中所述SpCas9结构域包含氨基

酸取代D1135L、S1136R、G1218S、E1219V、A1322R、R1335Q、T1337和A1322R、以及L1111、D1135L、S1136R、G1218S、E1219V、D1332A、D1332S、D1332T、D1332V、D1332L、D1332K、D1332R、R1335Q、T1337I、T1337V、T1337F、T1337S、T1337N、T1337K、T1337R、T1337H、T1337Q和其一T1337M中的一个或多个,或其相应的氨基酸取代。

[1059] 80. 实施方案73至76中任一项的修饰的细胞,其中所述SpCas9结构域包含氨基酸取代D1135M、S1136Q、G1218K、E1219F、A1322R、D1332A、R1335E和T1337R,或其相应的氨基酸取代。

[1060] 81. 根据实施方案73至75中任一项的修饰的细胞,其中所述SpCas9结构域对NG PAM、NGG PAM、GAA PAM、GAT PAM或CAA PAM具有特异性。

[1061] 82. 根据实施方案81的修饰的细胞,其中所述SpCas9结构域包含氨基酸取代E480K、E543K和E1219V或其相应的氨基酸取代。

[1062] 83. 根据实施方案70-72中任一项的修饰的细胞,其中所述Cas9结构域包含SaCas9结构域。

[1063] 84. 根据实施方案83的修饰的细胞,其中所述SaCas9结构域对NNNRRT PAM具有特异性。

[1064] 85. 根据实施方案84的修饰的细胞,其中所述SaCas9结构域对NNGRRT PAM具有特异性。

[1065] 86. 根据实施方式83至85中任一项的修饰的细胞,其中所述SaCas9结构域包含氨基酸取代N579A或其相应的氨基酸取代。

[1066] 87. 根据实施方案83-86中任一项的修饰的细胞,其中所述SaCas9结构域包含氨基酸取代E782K、N968K和R1015H或其相应的氨基酸取代。

[1067] 88. 根据实施方案70至72中任一项的修饰的细胞,其中所述Cas9结构域包含St1Cas9结构域:

[1068] 89. 根据实施方案88的修饰的细胞,其中所述St1Cas9结构域对NNACCA PAM具有特异性。

[1069] 90. 根据前述实施方案中任一项的修饰的细胞,其中所述腺苷脱氨酶结构域是自然界中不存在的修饰的腺苷脱氨酶结构域。

[1070] 91. 根据实施方案90的修饰的细胞,其中所述腺苷脱氨酶结构域包含TadA结构域。

[1071] 92. 根据实施方案91的修饰的细胞,其中TadA结构域包含TadA7.10的氨基酸序列。

[1072] 93. 根据实施方案63至92中任一项的修饰的单元,其中所述碱基编辑器系统进一步包含锌指结构域。

[1073] 94. 根据实施方案93的修饰的细胞,其中所述锌指结构域包含识别螺旋序列RNEHLEV、QSTTLKR和RTEHLAR或识别螺旋序列RGEHLRQ、QSGTLKR和RNDKLVP。

[1074] 95. 根据实施方案93或94的修饰的细胞,其中所述锌指结构域是zf1ra或zf1rb。

[1075] 96. 根据前述实施例中任一项所述的修饰单元,其中,所述碱基编辑器系统还包括核定位信号(NLS)。

[1076] 97. 根据实施例63-96中任一实施例所述的经修饰的单元格,其中,所述碱基编辑器系统还包括一个或多个链接器。

[1077] 98. 根据实施方案97的修饰的细胞,其中所述多核苷酸可编程DNA结合结构域,腺

昔脱氨酶结构域, 锌指结构域和NLS中的两个或更多个经由接头连接。

[1078] 99. 根据实施方案98的修饰的细胞, 其中所述接头是肽接头, 从而形成碱基编辑融合蛋白。

[1079] 100. 根据实施方案99的修饰的细胞, 其中所述肽接头包含选自以下的氨基酸序列:

[1080] SGGSSGSETPGTSESATPESSGGS,

[1081] SGGSSGGSSGSETPGTSESATPESSGGSSGGS,

[1082] GSGGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSGGS,

[1083] SGGSSGGSSGSETPGTSESATPES,

[1084] SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGS,

[1085] SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGSSGSETPGTSESATPESSGGSSGGS,

[1086] PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATS, (SGGS)_n, (GGGS)_n, (GGGS)_n, (G)_n, (EAAAK)_n, (GGS)_n, SGSETPGTSESATPES, and (XP)_n.

[1087] 101. 根据实施方案99或100的修饰的细胞, 其中所述碱基编辑融合蛋白包含选自以下的氨基酸序列:

[1088] MPKKKRKVSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIM
ALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGIL
ADECAALLSDFFRMRRQEIKAKKAQSSTDGSGSSGGSSGSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHA
LTLAKRARDEREVPVGAVLVHNNRVIGEGWNRRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCA
GAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAKKAQSSTD
SGGSSGGSSGSETPGTSESATPESDLVLGLAIGIGSVGVGILNKVTGEIHKNSRIFPAAQAENNLVRRTNRQGR
LARRKKHRRVRLNRLFEEGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLLDASDDGNS
SVGDYAIQIVKENSQKLETKTPGQIQRLERYQTYGQLRGDFTVEKDGGKHRLINVFPTSAYRSEALRILQTQQEFNPQ
ITDEFINRYLEILTGKRKYHGPNGEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEFRAAKASYTAQEFNLLN
DLNNLTVPPTETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMT
LETLDIEQMDRETLDKLAYVLTNTEREGIQEALHEFADGSFSQKQVDELVQFRKANSSIFGKGWHNFSVKLMME
LIPELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLLTEEIYNPVAKSVRQAIKIVNAAIKEYGDFDNIVIE
ARETNEDDEKKAIQKIQKANKDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTGTISIH
DLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQALDSMDDAWSFRELKAFVRESKTLNKKKE
YLLTEEDISKFDVRKKFIERNLVDTLYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTHHH
AVDALIIAASSQLNLWKKQKNTLVSYSEDQLLDIETGELISDDEYKESVFKAPYQHFVDTLKSKEFEDSILFSYQV
DSKFNRKISDATIYATRQAKVGKDKADETYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPIL
ENYPNKQINDKGKEVPCNPFLKYKEEHGYIRKYSKKGNPEIKSLKYYDSKLGHNHIDITPKDSNNKVVLSVSPWR
ADVYFNKTTGKYEILGLKYADLQFDKGTGYKISQEKYNDIKKKEGVDSDSEFKFTLYKNDLLL VKDTETKEQQLF
RFLSRTMPKQKHVELKPYDKQKFEGGEALIKVLGNVANSGQCKKGLGKSNISYKVRTDVLGNQHI IKNEGDKPK
LDFPKKKRKVEGADKRTADGSEFESPKKKRKV, MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGE
GWNRP IGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMD

VLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGS
SGGSSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNY
RLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFF
RMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGSSGGSKRNYILGLAIGITSVGYGIIDYETRD
VIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELGINPYEARVKGLSQKLSEE
EFS AALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSYVKE
AKQLLKVKQAYHQLDQSFIDTYIDLLETRRTYYEGPGEPSFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADL
YNALNDLNNLVITRDENEKLEYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHD
IKDITARKEI IENAELLDQIAKILTIYQSSEDIQEELTNLSEL TQEEIEQISNLKGYTGTHNLSLKAINLILDEL
WHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAI IKKYGLPNDII IELAREKN
SKDAQKMINEMQKRNRQTNERIEE IIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHII
PRSVSFDNSFNKVLVKQEENSKKGNRTPFQYLSSSDSKISYETFKKHILNLA KGGRISKTKKEYLLEERDINRF
SVQKDFINRNLVDTRYATRGLMNLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHAEDALIIANA
DFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETE QEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDT
LYSTRKDDKGNTLIVNNLNGLYDKDNDKLLKLINKSPEKLLMYHHPDQTYQKLKLIMEQYGDEKNPLYKYEEETGN
YLTYSKKDNGPVIKKIKYYGNKLN AHLDITDDYPNSRNKVVKLSLKPYPFDVYLDNGVYKFVTVKNLDV IKKENY
YEVNSKCYEEAKKLLKISNQAEFIASFYKNDLIKINGEL YRVIGVNNDLLNRIEVMIDITYREYLENMNDKRPPH
I IKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKGEGADKRTADGSEFESPKKKRKV,

[1089] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAAL
LSDFFRMRRQEIKAAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGSSGGSSSEVEFSHEYWMRHALTLAKRA
RDEREVPVGAVLVNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSR
IGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSGGSSGG
SSGSETPGTSESATPESSGSSGGSKRNYILGLAIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRG
ARRLKRRRRHRIQRVKKLLFDYNLLTDHSELGINPYEARVKGLSQKLSEEEFS AALLHLAKRRGVHNVNEVEEDT
GNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSYVKEAKQLLKVKQAYHQLDQSFIDTYIDL
LETRRTYYEGPGEPSFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYEKF
FQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEI IENAELLDQIAKILT
IYQSSEDIQEELTNLSEL TQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQ
QKEIPTTLVDDFILSPVVKRSFIQSIKVINAI IKKYGLPNDII IELAREKNSKDAQKMINEMQKRNRQTNERIEE I
IRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHII PRSVSFDNSFNKVLVKQEENSKKG
NRTPFQYLSSSDSKISYETFKKHILNLA KGGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNL
RSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHAEDALIIANADDFIFKEWKKLDKAKKVMENQMFEEK
QAESMPEIETE QEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIVNNLNGLYDKD
NDKLLKLINKSPEKLLMYHHPDQTYQKLKLIMEQYGDEKNPLYKYEEETGN YLTYSKKDNGPVIKKIKYYGNKLN
AHLDITDDYPNSRNKVVKLSLKPYPFDVYLDNGVYKFVTVKNLDV IKKENY YEVNSKCYEEAKKLLKISNQAEFI A
SFYKNDLIKINGEL YRVIGVNNDLLNRIEVMIDITYREYLENMNDKRPPHI IKTIASKTQSIKKYSTDILGNLYE
VSKKHPQIIKKGEGADKRTADGSEFESPKKKRKVSSGNSNANSRGPSFSSGLVPLSLRGS HRPGERPFQCRICM
RNFSRNEHLEVHTRTHTGEKPFQCRICMRNFSQSTTLKRHLRTHTGEKPFQCRICMRNFSRTEHLARHLKTHLRGS

SAQ, or

[1090] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTDSSGSSGGSSGSETPGTSESATPESSGGSSGSSSEVEFSHEYWMRHALTLAKRADREVEVPVGAVLVLNRRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSSGSSGSSGSETPGTSESATPESSGGSSGSSGSKRNYILGLAIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSRGARRLKRRRRHRIQRVKLLFDYNLLTDHSELGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRSKALEEKYVAELQLERLKKDGEVRGSINRFKTSYVKEAKQLLKVKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYEYEQI IENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEI IENAELLDQIAKILT IYQSSEDIQEELTNLSEL TQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQ QKEIPTTLVDDFILSPVVKRSFIQS I KVINAI I KKYGLPNDI I IELAREKNSKDAQKMINEMQKRNRQTNERIEEI IRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNKVLVKQEENSKKG NRTPFQYLSSSDSKISYETFKKHILNLA KGGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNL RSYFRVNNLDVKVKSINGGFTSFLRRKWFKKERNKGYKHAEDALI IANADFIKWKKLDKAKKVMENQMFEEK QAESMPEIETE QEYKEIFITPHQIKHIKDFKDYKSHRVDKKPNRKLINDTLYSTRKDDKGNTLIVNNLNGLYDKD NDKLKKLINKSPEKLLMYHHPQTYQKLKLIMEQYGDEKNPLYKYEEETGNYLTKYSKKDNGPVIKKIKYYGNKLN AHLDITDDYPNSRNKVVKLSLKPYPYFDVYLDNGVYKFVTVKNLDVIKKENYEVNSKCYEEAKKKKISNQAEFIA SFYKNDLIKINGELYRVIGVNDLLNRIEVMIDITYREYLENMNDKRPPHI IKT IASKTQSIKKYSTDILGNLYE VKSKKHPI I KKGEGADKRTADGSEFESPKKKRKVSSGNSNANSRGPSFSSGLVPLSLRGSHSRPGERPFQCRICMRNFSRGEHLRQHTRTHTGEKPFQCRICMRNFSQSGTLKRHLRTHTGEKPFQCRICMRNFSRNDKLVPHLKTHLRGS SAQ.

[1091] 102. 根据实施方案63至101中任一项的修饰的细胞,其中所述指导多核苷酸包含两个单独的多核苷酸,其中所述两个单独的多核苷酸是两个DNA、两个RNA或DNA和RNA。

[1092] 103. 根据前述实施方案中任一项的修饰的细胞,其中指导多核苷酸包含crRNA和tracrRNA,其中crRNA包含与SERPINA1多核苷酸中的靶序列互补的核酸序列。

[1093] 104. 根据实施方案103的修饰的细胞,其中靶序列包含选自GACAAGAAAGGGACTGAAGC、ATCAGACAAGAAAGGGACTGA和ACACACCGGTTGGTGGGGCCTC或其互补序列的序列。

[1094] 105. 根据实施方案102或103的修饰的细胞,其中所述碱基编辑器系统包含单链指导RNA (sgRNA)。

[1095] 106. 根据实施方案105的修饰细胞,其中所述sgRNA包含选自ACTCTaGGCAGAGGTCTCAAAGG和GCTCTaGGCCGAAGTGTGCGAGG的序列。

[1096] 107. 根据实施方案63-106中任一项的修饰的细胞,其中所述碱基编辑器系统包含载体,所述载体包含指导多核苷酸、多核苷酸可编程DNA结合结构域和脱氨酶结构域中的一个或多个。

[1097] 108. 根据实施方案107的修饰的细胞,其中所述载体是腺病毒载体、AAV载体、慢病毒载体或逆转录病毒载体。

- [1098] 109.一种基本编辑器系统,包括
- [1099] 指导多核苷酸或编码所述指导多核苷酸的核酸;
- [1100] 多核苷酸可编程DNA结合结构域或编码多核苷酸可编程DNA结合结构域的核酸,以及
- [1101] 腺苷脱氨酶结构域或编码腺苷脱氨酶结构域的核酸,
- [1102] 其中所述指导多核苷酸能够靶向所述碱基编辑器系统以引起SERPINA1多核苷酸中引起 α -1抗胰蛋白酶缺乏 (A1AD) 的单核苷酸多态性 (SNP) 的A • T至G • C改变。
- [1103] 110.一种碱基编辑器系统,包括
- [1104] 指导多核苷酸或编码所述指导多核苷酸的核酸;
- [1105] 多核苷酸可编程DNA结合结构域或编码多核苷酸可编程DNA结合结构域的核酸,以及
- [1106] 腺苷脱氨酶结构域或编码腺苷脱氨酶结构域的核酸,
- [1107] 其中所述指导多核苷酸能够靶向所述碱基编辑器系统以引起SERPINA1多核苷酸中引起 α -1抗胰蛋白酶缺乏 (A1AD) 的单核苷酸多态性 (SNP) 的A • T到G • C改变,
- [1108] 其中SERPINA1多核苷酸编码由SNP产生的在位置324处包含赖氨酸的A1AT蛋白,
- [1109] 其中A • T到G • C的改变用野生型氨基酸取代赖氨酸。
- [1110] 111.根据实施方案109的碱基编辑器系统,其中SERPINA1多核苷酸编码由SNP产生的在位置342处包含赖氨酸的A1AT蛋白。
- [1111] 112.根据实施方案111的碱基编辑器系统,其中所述A • T至G • C改变用野生型氨基酸替代赖氨酸。
- [1112] 113.根据实施方案110或112的碱基编辑器系统,其中所述野生型氨基酸是谷氨酸。
- [1113] 114.根据实施方案109至113中任一项的碱基编辑器系统,其中所述多核苷酸可编程DNA结合结构域是Cas9结构域。
- [1114] 115.根据根据实施例114所述的碱基编辑器系统,其中所述Cas9结构域是核酸酶失活的Cas9结构域。
- [1115] 116.根据根据实施例114所述的碱基编辑器系统,其中,所述Cas9域是Cas9切口酶域。
- [1116] 117.根据实施例114至116中任一实施例所述的碱基编辑器系统,其中,所述Cas9域包括SpCas9域。
- [1117] 118.根据实施方案117的碱基编辑器系统,其中所述SpCas9结构域包含D10A和/或H840A氨基酸取代或其相应的氨基酸取代。
- [1118] 119.根据实施例117或118所述的碱基编辑器系统,其中所述SpCas9结构域对NGG PAM具有特异性。
- [1119] 120.根据实施例117至119中任一实施例所述的碱基编辑器系统,其中,所述SpCas9域对NGA PAM,NGT PAM或NGC PAM具有特异性。
- [1120] 121.根据实施方案117至119中任一项的碱基编辑器系统,其中所述SpCas9结构域包含氨基酸取代L1111R、D1135V、G1218R、E1219F、A1322R、R1335V、T1337R和L1111、D1135L、S1136R、G1218S、E1219V、D1332A、R1335Q、T1337I、T1337V、T1337F和T1337M中的一个或多个

个,或其相应的氨基酸取代。

[1121] 122.根据实施方案117至119中任一项的碱基编辑器系统,其中所述SpCas9结构域包含氨基酸取代L1111R、D1135V、G1218R、E1219F、A1322R、R1335V、T1337R和L1111、D1135L、S1136R、G1218S、E1219V、D1332A、D1332S、D1332T、D1332V、D1332L、D1332K、D1332R、R1335Q、T1337I、T1337V、T1337F、T1337S、T1337N、T1337K、T1337R、T1337H、T1337Q和T1337M中的一个或多个,或相应的氨基酸取代。

[1122] 123.根据实施方式117至119中任一项的碱基编辑器系统,其中所述SpCas9结构域包含氨基酸取代D1135L、S1136R、G1218S、E1219V、A1322R、R1335Q、T1337和A1322R、以及L1111、D1135L、S1136R、G1218S、E1219V、D1332A、D1332S、D1332T、D1332V、D1332L、D1332K、D1332R、R1335Q、T1337I、T1337V、T1337F、T1337S、T1337N、T1337K、T1337R、T1337H、T1337M中的一个或多个,或相应的氨基酸取代。

[1123] 124.根据实施方案117至119中任一项的碱基编辑器系统,其中所述SpCas9结构域包含氨基酸取代D1135M、S1136Q、G1218K、E1219F、A1322R、D1332A、R1335E和T1337R,或其相应的氨基酸取代。

[1124] 125.根据实施例117至119中任一实施例所述的碱基编辑器系统,其中,所述SpCas9结构域具有对NG PAM、NGN PAM、GAA PAM、GAT PAM或CAA PAM的特异性。

[1125] 126.根据实施方案125的碱基编辑器系统,其中所述SpCas9结构域包含氨基酸取代E480K、E543K和E1219V或其相应的氨基酸取代。

[1126] 127.根据实施例114至116中任一实施例所述的碱基编辑器系统,其中,所述Cas9域包括SaCas9域。

[1127] 128.根据实施例127所述的碱基编辑器系统,其中,所述SaCas9结构域对NNNRRT PAM具有特异性。

[1128] 129.根据实施例128所述的碱基编辑器系统,其中所述SaCas9域对NNGRRT PAM具有特异性。

[1129] 130.根据实施方案127至129中任一项的碱基编辑器系统,其中所述SaCas9结构域包含氨基酸取代N579A或其对应的氨基酸取代。

[1130] 131.根据实施方案127至130中任一项的碱基编辑器系统,其中所述SaCas9结构域包含氨基酸取代E782K、N968K和R1015H或其相应的氨基酸取代。

[1131] 132.根据实施例117至119中任一项所述的碱基编辑器系统,其中,所述Cas9域包括St1Cas9域:

[1132] 133.根据实施例132所述的碱基编辑器系统,其中,所述St1Cas9结构域对NNACCA PAM具有特异性。

[1133] 134.根据实施方案中任一项的碱基编辑器系统,其中所述腺苷脱氨酶结构域是自然界中不存在的修饰的腺苷脱氨酶结构域。

[1134] 135.根据实施方案90的碱基编辑器系统,其中所述腺苷脱氨酶结构域包含TadA结构域。

[1135] 136.根据实施方案91的碱基编辑器系统,其中TadA结构域包含TadA 7.10的氨基酸序列。

[1136] 137.根据实施例109至136中任一实施例所述的碱基编辑器系统,其中,所述碱基

编辑器系统还包括锌指结构域。

[1137] 138. 根据实施方案137的碱基编辑器系统, 其中所述锌指结构域包含识别螺旋序列RNEHLEV、QSTTLKR和RTEHLAR或识别螺旋序列RGEHLRQ、QSGTLKR和RNDKLVP。

[1138] 139. 根据实施方案136或137的碱基编辑器系统, 其中所述锌指结构域是zf1ra或zf1rb。

[1139] 140. 根据前述实施例中任一项所述的碱基编辑器系统, 其中, 所述碱基编辑器系统还包括核定位信号(NLS)。

[1140] 141. 根据实施例109至140中任一实施例所述的碱基编辑器系统, 其中, 所述碱基编辑器系统还包括一个或多个链接器。

[1141] 142. 根据实施方案141的碱基编辑器系统, 其中所述多核苷酸可编程DNA结合结构域、腺苷脱氨酶结构域、锌指结构域和NLS中的两个或更多个经由接头连接。

[1142] 143. 根据实施方案142的碱基编辑器系统, 其中所述接头是肽接头, 从而形成碱基编辑融合蛋白。

[1143] 144. 根据实施方案143的碱基编辑器系统, 其中所述肽接头包含选自以下的氨基酸序列:

[1144] SGGSSGSETPGTSESATPESSGGS,

[1145] SGGSSGGSSGSETPGTSESATPESSGGSSGGS,

[1146] GSGGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSSGGS,

[1147] SGGSSGGSSGSETPGTSESATPES,

[1148] SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGS,

[1149] SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGSSGSETPGTSESATPESSGGSSGGS,

[1150] PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATS, (SGGS)_n, (GGGS)_n, (GGGS)_n, (G)_n, (EAAAK)_n, (GGS)_n, SGSETPGTSESATPES, and (XP)_n.

[1151] 145. 根据实施方案143或144的碱基编辑器系统, 其中所述碱基编辑融合蛋白包含选自以下的氨基酸序列:

[1152] MPKKRKRVSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRQEIKAQKKAQSSTDSSGGSSGGSSGSETPGTSESATPESSGGSSGSSSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVHNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSSGGSSGGSSGSETPGTSESATPESDLVLGLAIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGRRLARRKKHRRVRLNRLFEEGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLLDASDDGNSVGDYAIQIVKENSQKLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQITDEFINRYLEILTGKRKYHGPNGEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEFRAAKASYTAQEFNLLNLDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETLDKLAYVLTNTEREGIQEALHEFADGSFSQKQVDELVQFRKANSSIFGKGWHNFSVKLMME LIPELYETSEEQMTILTRLGKQKTSSSNKTKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVIM

ARETNEDDEKKAIQKIQKANKDEKDAAMLKAANQYNGKAELPHSVFHGKQLATKIRLWHQQGERCLYTGKTISIH
DLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTYPYQALDSMDDAWSFRELKAFVRESKTL SNKKKE
YLLTEEDISKFDVRKKFIERNLVDTLYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTHHH
AVDALIIAASSQLNLWKKQKNTLVSYSEDQLLDIETGELISDDEYKESVFKAPYQHFDVTLKSKEFEDSILFSYQV
DSKFNRKISDATIYATRQAKVGKDKADETYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPIL
ENYPNKQINDKGKEVPCNPFLKYKEEHGYIRKYSKKGNPEIKSLKYYDSKLGNHIDITPKDSNNKVVLQSVSPWR
ADVYFNKTTGKYEILGLKYADLQFDKGTGYKISQEKYNDIKKKEGVDSDSEFKFTLYKNDLLLVKDTETKEQQLF
RFLSRTMPKQKHVELKPYDKQKFEGGEALIKVLGNVANSGQCKKGLGKSNISYKVRTDVLGNQHI IKNEGDKPK
LDFPKKKRKVEGADKRTADGSEFESPKKKRKV , MSEVEFSHEYWMRHAL TLAKRAWDEREVPVGAVLVHNNRVI
GWNRPGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMD
VLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTDSSGSSGGSSGSETPGTSESATPESGSS
SGGSSEVEFSHEYWMRHAL TLAKRARDEREVPVGAVLVLNNRVI GEGWNRAIGLHDPTAHAEIMALRQGGLVMQNY
RLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFF
RMPRQVFNAQKKAQSSTDSSGSSGGSSGSETPGTSESATPESGSSGGSSGSKRNYILGLAIGITSVGYGIIDYETRD
VIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELGINPYEARVKGLSQKLSEE
EFSALLHLAKRRGVHNVNEVEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSYVKE
AKQLLKVKQAYHQLDQSFIDTYIDLLETRRTYYEGPGEPSFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADL
YNALNDLNNLVITRDENEKLEYEKFQI IENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHD
IKDITARKEI IENAELLDQIAKILTIYQSSEDIQEELTNLNSLTQEEIEQISNLKGYTGTHNLSLKAINLILDEL
WHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDF ILSPVVKRSFIQSIKVINAI IKKYGLPNDI I IELAREKN
SKDAQKMINEMQKRNRQTNERIEE I IRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNPNFYEV DHI I
PRSVSFDNSFNKVLVKQEENSKGNRTPFQYLSSSDSKISYETFKKHILNLAKGGRISKTKKEYLLEERDINRF
SVQKDFINRNLVDTRYATRGLMNLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHAEDAL I IANA
DFIFKEWKKLDKAKKVMENQMFEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDT
LYSTRKDDKGNLTIVNNLGLYDKDNDKLKKL INKSPEKLLMYHHPQTYQKLKLIMEQYGDEKNPLYKYEEETGN
YLTKYSKKDNGPVIKKIKYYGNKLNALHDITDDYPNSRNKVVKLSLKPYPFDVYLDNGVYKFVTVKNLDVIKKENY
YEVNSKCYEEAKKLKISNQAEFIASFYKNDLIKINGELYRVIGVNNDLLNRIEVMIDITYREYLENMNDKRPPH
I IKTIASKTQSIKKYSTDILGNLYEVKSKKHPQI IKKGEGADKRTADGSEFESPKKKRKV ,

[1153] MSEVEFSHEYWMRHAL TLAKRAWDEREVPVGAVLVHNNRVI GEGWNRPGRHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAAL
LSDFFRMRRQEIKAQKKAQSSTDSSGSSGGSSGSETPGTSESATPESGSSGGSSGSEVEFSHEYWMRHAL TLAKRA
RDEREVPVGAVLVLNNRVI GEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSR
IGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSSGSSGG
SSGSETPGTSESATPESGSSGGSSGSKRNYILGLAIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRG
ARRLKRRRRHRIQRVKKLLFDYNLLTDHSELGINPYEARVKGLSQKLSEEEFSALLHLAKRRGVHNVNEVEDT
GNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSYVKEAKQLLKVKQAYHQLDQSFIDTYIDL
LETRRTYYEGPGEPSFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYEKF
FQI IENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEI IENAELLDQIAKILT
IYQSSEDIQEELTNLNSLTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQ

QKEIPTTLVDDFILSPVVKRSFIQSIKVINAI IKKYGLPNDII IELAREKNSKDAQKMINEMQKRNRQTNERIEEI
IRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHI IPRSVSFDNSFNKVLVKQEENSKKG
NRTPFQYLSSSDSKISYETFKKHILNLAKGGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNL
RSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHAEDALI IANADFIKKEWKLDKAKKVMENQMFECK
QAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSRVDKKNRKLINDTLYSTRKDDKGNTLIVNNLNGLYDKD
NDKLKLINKSPEKLLMYHHPQTYQKLKLIMEQYGDEKNPLYKYEEETGNYLTKYSKKDNGPVIKKIKYYGNKLN
AHLDTDDYPNSRNKVVKLSLKPYPFDVYLDNGVYKFVTVKNLDVIKKENYEVNSKCYEEAKKKKISNQAEFIA
SFYKNDLIKINGELYRVIGVNNDLLNRIEVMIDITYREYLENMNDKRPPHI IKTIASTQSIKKYSTDILGNLYE
VKSKKHPQIIKKGEGADKRTADGSEFESPKKKRKVSSGNSNANSRGPSFSSGLVPLSLRGSHSRPGERPFQCRICM
RNFSRNEHLEVHTRTHTGEKPFQCRICMRNFSQSTTLKRHLRTHHTGEKPFQCRICMRNFSRTEHLARHLKTHLRGS
SAQ, or

[1154] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAAL
LSDFFRMRREQEIKAQKKAQSSTDSSGSSGGSSGSETPGTSESATPESSGGSSGSSSEVEFSHEYWMRHALTLAKRA
RDEREVPVGAVLVHNNRVIGEGWNRIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSR
IGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSSGSSGG
SSGSETPGTSESATPESSGGSSGSSGSKRNYILGLAIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRG
ARRLKRRRRHRIQRVKKLLFDYNLLTDHSELGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDT
GNELSTKEQISRSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDL
LETRRTYYEGPGEKSPFGWKDIKEWYMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYEYK
FQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEI IENAELLDQIAKILT
IYQSSEDIQEELTNLSELQTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVKKVDLSQ
QKEIPTTLVDDFILSPVVKRSFIQSIKVINAI IKKYGLPNDII IELAREKNSKDAQKMINEMQKRNRQTNERIEEI
IRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHI IPRSVSFDNSFNKVLVKQEENSKKG
NRTPFQYLSSSDSKISYETFKKHILNLAKGGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNL
RSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHAEDALI IANADFIKKEWKLDKAKKVMENQMFECK
QAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSRVDKKNRKLINDTLYSTRKDDKGNTLIVNNLNGLYDKD
NDKLKLINKSPEKLLMYHHPQTYQKLKLIMEQYGDEKNPLYKYEEETGNYLTKYSKKDNGPVIKKIKYYGNKLN
AHLDTDDYPNSRNKVVKLSLKPYPFDVYLDNGVYKFVTVKNLDVIKKENYEVNSKCYEEAKKKKISNQAEFIA
SFYKNDLIKINGELYRVIGVNNDLLNRIEVMIDITYREYLENMNDKRPPHI IKTIASTQSIKKYSTDILGNLYE
VKSKKHPQIIKKGEGADKRTADGSEFESPKKKRKVSSGNSNANSRGPSFSSGLVPLSLRGSHSRPGERPFQCRICM
RNFSRGEHLRQHTRTHTGEKPFQCRICMRNFSQSGTLKRHLRTHHTGEKPFQCRICMRNFSRNDKLVPHLKTHLRGS
SAQ.

[1155] 146. 根据前述实施方案中任一项的碱基编辑器系统, 其中所述指导多核苷酸包含两个单独的多核苷酸, 其中所述两个单独的多核苷酸是两个DNA、两个RNA或DNA和RNA。

[1156] 147. 根据实施方案109-146中任一项的碱基编辑器系统, 其中所述指导多核苷酸包含crRNA和tracrRNA, 其中所述crRNA包含与SERPINA1多核苷酸中的靶序列互补的核酸序列。

[1157] 148. 根据实施方案147的碱基编辑器系统, 其中所述靶序列包含选自以下的序列:

GACAAGAAAGGGACUGAAGC、AUCGACAAGAAAGGGACUGA和ACACACCGGUUGGUGGCCUC或它们的互补物。

[1158] 149. 根据实施方案147或148的碱基编辑器系统,其中所述碱基编辑器系统包括单链指导RNA (sgRNA)。

[1159] 150. 根据实施方案149的碱基编辑器系统,其中所述sgRNA包含选自ACTCTaGGCAGAGGTCTCAAAGG和GCTCTaGGCCGAAGTGTGCGCAGG的序列。

[1160] 151. 根据实施方案109至150中任一项的碱基编辑器系统,其中所述碱基编辑器系统包含载体,所述载体包含指导多核苷酸、多核苷酸可编程DNA结合结构域和脱氨酶结构域中的一个或多个。

[1161] 152. 根据实施方案151的碱基编辑器系统,其中所述载体是腺病毒载体、AAV载体、慢病毒载体或逆转录病毒载体。

[1162] 153. 一种在有需要的受试者中治疗疾病的方法,包括对受试者施用碱基编辑器系统,所述系统包括

[1163] 指导多核苷酸或编码所述指导多核苷酸的核酸;

[1164] 多核苷酸可编程DNA结合结构域或编码多核苷酸可编程DNA结合结构域的核酸,以及

[1165] 脱氨酶结构域或编码腺苷脱氨酶结构域的核酸,

[1166] 其中所述指导多核苷酸能够靶向所述碱基编辑器系统以实现所述受试者细胞的靶多核苷酸中的病原性单核苷酸多态性 (SNP) 的脱氨基,

[1167] 其中病原性SNP是表3A或表3B中病原性氨基酸突变的原因,其中病原性SNP的脱氨基导致病原性SNP转化为其野生型等位基因,从而治疗所述疾病。

[1168] 154. 一种在有需要的受试者中治疗疾病的方法,包括

[1169] (a) 将碱基编辑器系统引入单元格,所述系统包括

[1170] 指导多核苷酸或编码所述指导多核苷酸的核酸;

[1171] 多核苷酸可编程DNA结合结构域或编码多核苷酸可编程DNA结合结构域的核酸,和

[1172] 脱氨酶结构域或编码脱氨酶结构域的核酸,以及

[1173] (b) 将细胞施用于受试者,

[1174] 其中所述指导多核苷酸能够靶向所述碱基编辑器系统以实现所述受试者细胞的靶多核苷酸中的病原性单核苷酸多态性 (SNP) 的脱氨基,

[1175] 其中病原性SNP是表3A或表3B中病原性氨基酸突变的原因,其中病原性SNP的脱氨基导致病原性SNP转化为其野生型等位基因,从而治疗所述疾病。

[1176] 155. 一种校正靶多核苷酸中引起疾病的SNP病因的方法,其包括使所述靶多核苷酸与碱基编辑器系统接触,所述碱基编辑器系统包括

[1177] 指导多核苷酸;

[1178] 多核苷酸可编程DNA结合域,和

[1179] 脱氨酶域,

[1180] 其中所述指导多核苷酸能够靶向所述碱基编辑器系统以实现所述受试者细胞的靶多核苷酸中的病原性单核苷酸多态性 (SNP) 的脱氨基,

[1181] 其中,病原体SNP是表3A或表3B中的病原体氨基酸突变的病因,其中,病原体SNP的

脱氨基导致病原体SNP向其野生型等位基因的转化,从而校正了靶多核苷酸中的病原体SNP。

[1182] 156.一种产生用于治疗疾病的修饰细胞的方法,其包括将碱基编辑器系统引入细胞中,所述碱基编辑器系统包括:

[1183] 指导多核苷酸或编码所述指导多核苷酸的核酸;

[1184] 多核苷酸可编程DNA结合结构域或编码多核苷酸可编程DNA结合结构域的核酸,以及

[1185] 脱氨酶结构域或编码脱氨酶结构域的核酸,

[1186] 其中所述指导多核苷酸能够靶向所述碱基编辑器系统以实现所述受试者细胞的靶多核苷酸中的病原性单核苷酸多态性(SNP)的脱氨基,

[1187] 其中病原性SNP是表3A或表3B中的病原性氨基酸突变的原因,其中SNP的脱氨基导致病原性SNP转化为其野生型等位基因,从而校正靶多核苷酸中的病原性SNP。

[1188] 157.根据实施方案156的方法,其中所述引入是体内或离体的。

[1189] 158.根据实施方案156或157的方法,其中所述细胞获自患有所述疾病的受试者。

[1190] 159.根据实施方案156至158中任一项的方法,其中所述疾病选自斯特格病变、弹性假黄瘤、中链酰基辅酶A脱氢酶缺乏症、严重的联合免疫缺陷、原发性低血尿症、常染色体隐性遗传性高胆固醇血症、变色白细胞营养不良、马特鲁斯-拉米综合征(MSPVI)、I型瓜氨酸血症、经典的高半胱氨酸尿症的Darier氏病、囊性纤维化、脉络膜炎、神经元类脂褐质尿毒症、常染色体显性耳聋肉碱棕榈酰转移酶II缺乏症、胱氨酸症、常染色体隐性遗传性耳聋、无配子体血友病性耳聋V因子缺乏症、先天性VII因子缺乏症、A型血友病、B型血友病、酪氨酸血症1型、常染色体显性遗传性低磷酸盐血症性病、冯吉尔克病、地中海G6PD缺乏症候群、莫尔奎(Morquio)综合征、MPSIVA经典半乳糖血症、高雪氏症、戊二酰-CoA脱氢酶缺乏症、甘氨酸脑病、锥状-杆状细胞营养不良、Sly综合征(MPSVII)、镰状细胞病、间歇性紫质症、莱希-尼亨症候群、韩特氏(Hunter)症、贺勒氏症(MSPII)、视网膜色素病变Andersen-Tawil综合征、Meesmann上皮角膜营养不良、帕金森氏病、Sanfilippo综合征B(MPSIIB)、CADASIL综合征蓝锥性单色症苯丙酮尿症候群杂色斑状紫质症、1型神经元类脂褐质沉积病、1型库贾氏病(CJD)、遗传性慢性胰腺炎、莱柏氏先天黑蒙症2、戴布氏贫血、SanmannippoA综合征(A型)ATTR淀粉样变性病、视网膜色素病变/尤塞氏综合征1C型和肌小管肌病。

[1191] 160.根据实施方案156至159中任一项的方法,其中靶多核苷酸包含表3A或表3B中的基因。

[1192] 161.根据实施方案156至160中任一项的方法,其中所述病原性氨基酸突变包括表3A或表3B中的病原性突变。

[1193] 162.根据实施方案156至161中任一项的方法,其中所述多核苷酸可编程DNA结合结构域是Cas9结构域。

[1194] 163.根据实施方案162的方法,其中所述Cas9结构域是核酸酶失活的Cas9结构域或Cas9切口酶结构域。

[1195] 164.根据实施方案162或163的方法,其中所述Cas9结构域包括SpCas9结构域。

[1196] 165.根据实施方案164的方法,其中所述SpCas9结构域包含D10A和/或H840A氨基

酸取代或其相应的氨基酸取代。

[1197] 166. 根据实施方案164或165的方法, 其中所述SpCas9结构域对NGN PAM具有特异性。

[1198] 167. 根据实施方案164至166中任一项的方法, 其中所述Cas9结构域包含氨基酸取代D1135M、S1136Q、G1218K、E1219F、A1322R、D1332A、R1335E和T1337R, 或其相应的氨基酸取代。

[1199] 168. 根据实施方案164或165的方法, 其中所述SpCas9结构域对NG PAM、NGG PAM、GAA PAM、GAT PAM或CAA PAM具有特异性。

[1200] 169. 根据实施方案168的方法, 其中所述SpCas9结构域包含氨基酸取代E480K、E543K和E1219V或其相应的氨基酸取代。

[1201] 170. 根据实施方式162或163的方法, 其中所述Cas9结构域包含SaCas9。

[1202] 171. 根据实施方案170的方法, 其中所述SaCas9结构域对NNRRRT PAM具有特异性。

[1203] 172. 根据实施方案170或171的方法, 其中所述SaCas9结构域包含氨基酸取代N579A或其相应的氨基酸取代。

[1204] 173. 根据实施方案170至172中任一项的方法, 其中所述SaCas9结构域包含氨基酸取代E782K、N968K和R1015H或其相应的氨基酸取代。

[1205] 174. 根据实施方案162或163的方法, 其中所述Cas9结构域包括St1Cas9结构域。

[1206] 175. 根据实施方案174的方法, 其中所述St1Cas9结构域对NNACCA PAM具有特异性。

[1207] 176. 根据实施方案156-175中任一项的方法, 其中所述脱氨酶结构域包含胞苷脱氨酶结构域。

[1208] 177. 根据实施方案176的方法, 其中所述胞苷脱氨酶结构域包含APOBEC1结构域。

[1209] 178. 根据实施方案156至175中任一项的方法, 其中所述脱氨酶结构域包含腺苷脱氨酶结构域。

[1210] 179. 根据实施方案178的方法, 其中所述腺苷脱氨酶结构域包含TadA 7.10结构域。

[1211] 180. 根据实施例156至179中任一实施例所述的方法, 其中所述碱基编辑器系统还包括UGI域。

[1212] 181. 根据实施方案156至180中任一项的方法, 其中所述碱基编辑器系统进一步包含锌指结构域。

[1213] 182. 根据实施例156至181中任一实施例所述的方法, 其中所述碱基编辑器系统还包括一个或多个链接器。

[1214] 183. 根据实施方案182的方法, 其中所述多核苷酸可编程DNA结合结构域, 脱氨酶结构域, UGI结构域和锌指结构域中的两个或更多个经由接头连接,

[1215] 184. 根据实施方案183的方法, 其中所述接头是肽接头, 从而形成碱基编辑融合蛋白。

[1216] 185. 根据实施方案184的方法, 其中所述碱基编辑融合蛋白包含BE4的氨基酸序列。

[1217] 186. 根据实施方案184的方法, 其中所述碱基编辑融合蛋白包含以下氨基酸序列:

[1218] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAAQKKAQSSTDSSGSSGGSSGSETPGTSESATPESSGGSSGSSSEVEFSHEYWMRHALTLAKRADREVEVPVGAVLVNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSSGSSGGSSGSETPGTSESATPESSGGSSGSSGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKI IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQDLSLHEHIANLAGSPAIKKGILQTVKVVDDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVFREINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPPKYYGGFmqPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSEFKNPIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLASakfLQKGNELALPSKYVNFLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIIEQISEFSKRVI LADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPrAFKYFDTTIaRKeYrSTKEVLDATLIHQSI TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV.

[1219] 187. 根据实施方案156至186中任一项的方法,其中所述脱氨基导致少于10%的插入缺失形成。

[1220] 188. 一种碱基编辑器系统,包括

[1221] 指导多核苷酸或编码所述指导多核苷酸的核酸;

[1222] 多核苷酸可编程DNA结合结构域或编码多核苷酸可编程DNA结合结构域的核酸,和

[1223] 脱氨酶结构域或编码腺苷脱氨酶结构域的核酸,

[1224] 其中所述指导多核苷酸能够靶向所述碱基编辑器系统以实现靶多核苷酸中的病原性单核苷酸多态性(SNP)的脱氨基,

[1225] 其中,病原体SNP是表3A或表3B中的病原体氨基酸突变的原因,其中,病原体SNP的脱氨基导致病原体SNP向其野生型等位基因的转化,其中靶标多核苷酸在其中包含靶向序列表3A或表3B。

[1226] 实施例

[1227] 提供以下实施例仅用于说明目的,并不旨在限制本文提供的权利要求的范围。

[1228] 实施例1. 碱基编辑器中的PAM变体验证

[1229] 新颖的CRISPR系统和PAM变体使碱基编辑器可以对目标SNP进行精确校正。几种新

颖的PAM变体已得到评估和验证。PAM评估和碱基编辑器的详细信息在例如国际第PCT/2017/045381号专利申请(WO第2018/027078号国际专利申请)和国际第PCT/US2016/058344号专利申请(WO第2017/070632号国际专利申请)中进行了描述,它们各自通过引用并入本文完整。另请参阅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);以及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),全文通过引用将其每一个的内容合并于此。

[1230] 实施例2.基因编辑以校正 α -1抗胰蛋白酶缺乏症(A1AD)

[1231] α -1抗胰蛋白酶(A1A或A1AT)是由14号染色体上的SERPINA1基因编码的蛋白酶抑制剂。这种糖蛋白主要在肝脏中合成,并分泌到血液中,血清浓度为1.5-3.0g/L。(20-52 μ mol/L)在健康成年人中(图1)。A1AT扩散到肺间质和肺泡衬里液中,在那里使嗜中性粒细胞弹性蛋白酶失活,从而保护肺组织免受蛋白酶介导的损害。 α -1抗胰蛋白酶缺乏症(A1AD)以常染色体显性方式遗传。

[1232] 已经描述了SERPINA1基因的100多个遗传变异,但并非所有变异都与疾病有关。这些变体的字母名称基于其在凝胶电泳中的迁移速度。最常见的变异是M(中等迁移率)等位基因,两个最常见的缺陷等位基因是PiS和PiZ(后者迁移速度最慢)。已经描述了几种不产生可测血清蛋白的突变;这些被称为“无效”等位基因。最常见的基因型是MM,可产生正常的血清 α -1抗胰蛋白酶水平。大多数严重缺乏症的人都是Z等位基因(ZZ)的纯合子。Z蛋白在肝细胞内质网中产生时会折叠并聚合。这些异常的聚合物被困在肝脏中,大大降低了 α -1抗胰蛋白酶的血清水平。 α -1抗胰蛋白酶缺乏症患者所见的肝病是由肝细胞中异常的 α -1抗胰蛋白酶蛋白的积累以及随之而来的细胞反应(包括自噬,内质网应激反应和凋亡)引起的。图2显示了最常见的基因型(MM、MZ、SS、SZ和ZZ)以及 α -1抗胰蛋白酶的各自血清水平。 α -1抗胰蛋白酶的循环水平降低导致肺中性粒细胞弹性蛋白酶活性增加;蛋白酶和抗蛋白酶活性的这种不平衡导致与这种状况有关的肺部疾病(图1)。

[1233] α -1抗胰蛋白酶缺乏症(A1AD)在高加索人中最常见,并且最常影响到肺部和肝脏。在肺部,最常见的表现是早发(30多岁和40多岁的患者)全肺部肺气肿,最常见于肺部。但是,支气管扩张也可能发生弥漫性或上叶肺气肿。最常见的症状包括呼吸困难,喘息和咳嗽。患病者的肺功能检查显示与COPD一致;但是,可能会观察到支气管扩张药反应性,并可能被误诊为哮喘。

[1234] 由ZZ基因型引起的肝病以各种方式表现出来。患病的婴儿可能在新生儿期出现胆汁淤积性黄疸,有时表现为无胆色粪(苍白或粘土色)和肝肿大。血液中的结合胆红素、转氨酶和 γ -谷氨酰转移酶水平升高。年龄较大的儿童和成人的肝脏疾病可能偶发性转氨酶升高或肝硬化的迹象,包括静脉曲张出血或腹水。 α -1抗胰蛋白酶缺乏症也使患者易患肝细胞癌。尽管纯合的ZZ基因型对于肝病的发展是必不可少的,但杂合的Z突变可以通过赋予更大的更严重的肝病风险(例如在丙型肝炎感染和囊性纤维化肝病中)作为其他疾病的遗传修饰子。

[1235] A1AD的两种最常见的临床变异是E264V (PiS) 和E342K (PiZ) 等位基因。超过一半的A1AD患者携带至少一个E342K突变拷贝。通过同源引导修复 (HDR) 进行的核酸酶基因组编辑效率低下,丰富的插入将降低循环水平并恶化肺部症状。由于额外的错误折叠蛋白,通过AAV进行肝脏的基因治疗使肝脏病理恶化。编码野生型A1AT和可击倒E342K A1AT的siRNA的AAV有望解决这两种疾病。

[1236] 图3A显示了SERPINA1基因突变的精确的校正碱基编辑策略。可以编辑序列 (A7) 中位置7的“A”核碱基,“靶标A”以恢复野生型。图3B显示了根据DNA编辑器 (E342K、D341G、E342G) 产生的替代等位基因的A1AT蛋白分泌的表征。用编码A1AT变体的载体转染HEK293T,并通过ELISA评估上清液的A1AT含量。所述测定法表征了相对于WT A1AT,含E342K的A1AT的分泌减少。结果表明,PAM选项为AGCT,预计将导致编辑SERPINA1序列第5位和/或第7位 (A5和/或A7) 的“A”。当A7编辑产生野生型蛋白时,发现A1AT功能得以恢复;当A5和A7编辑导致在位置342和D341G处的谷氨酸 (E) 时;当A7和A8转换为WT和E342G时。评估的重组突变体A1AT的表型 (活性/功能) 包括细胞分泌蛋白质和抑制中性粒细胞弹性蛋白酶。A1AT变体的功能活性在图3C (抑制弹性蛋白酶) 中示出。有趣的是,D341G突变具有显着的弹性蛋白酶活性,这证实了A5和A7编辑的恢复结果。图4显示了产生多肽变体 (例如,腺苷脱氨酶变体) 的策略。

[1237] 实施例3. HEK298T细胞的碱基编辑

[1238] 在HEK298T细胞中测试了不同基因中各种SNP的碱基编辑效率 (表7)。对于质粒转染,使用250ng的gRNA质粒和750ng的碱基编辑质粒,以3 μ l:1 μ g的比例,以3 μ l:1 μ g的比例瞬时转染了针对HEK293细胞优化的高效,低毒性的DNA转染试剂Mirus TransIT293,将HEK293T细胞瞬时转染。对于mRNA转染,使用Neon系统在1150V电压下用两个20ms脉冲将HEK293T细胞用3 μ g总RNA电穿孔。质粒转染四天和RNA电穿孔两天后,用含有0.05% SDS, 25 μ g/ml蛋白酶K, 10mM Tris pH 8.0的简单裂解缓冲液从细胞中提取基因组DNA,然后在85 $^{\circ}$ C加热灭活。PCR扩增基因组位点并在MiSeq上测序。如前所述,分析了每个位置的基频和插入缺失百分比的结果。例如,插入缺失计算的细节在国际第PCT/2017/045381号国际专利申请 (W0第2018/027078号专利申请) 和国际第PCT/US2016/058344号专利申请 (W0第2017/070632号国际专利申请) 中进行了描述,其全部内容通过引用整体并入本文。另请参见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); 以及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), 全文其内容通过引用合并于此。

[1239] 表7. SNP校正的肝靶标编辑

[1240]

疾病	靶标	美国患者估算	碱基编辑器	原间隔子	PAM	精确校正百分比
粘多糖贮积病 1	IDUA W402*	600	ABE	GCTCTAGGC CGAAGTGTC GC	NGG	12.53
同型半胱氨酸尿	CBS I278T	900	CBE	TCACTGGGG TGGATCCCG AA	NGG	26.1
同型半胱氨酸尿	CBS T191M	unknown	ABE	GTGGGCATC CTCACAATCT C	NGC	1.11
糖原贮积症 1a	G6PC Q347*	500	ABE	GGACCTAGG CGAGGCAGT AG	NGG	37.22
糖原贮积症 1a	G6PC Q347*	500	ABE	GACCTAGGC GAGGCAGTA GG	NGA	43.89
糖原贮积症 1a	G6PC R83C	900	ABE	CAGTATGGA CACTGTCCA AA	NNG RRT	1.41
α -1 抗胰蛋白酶缺乏症 (A1AD)	<i>SERPINA</i> I E342K	30000	ABE	ATCGACAAG AAAGGGACT GA	NGC	0.39

[1241] 实施例4.通过突变改组产生的改进的NGC-PAM ABE

[1242] 使用表达Lipo2000的表达质粒转染包含表达含有E342K的A1AT变体的PiZ ORF的整合慢病毒盒的HEK293T细胞 (HEK293T-E342K)。转染后,表征A1AT中的碱基编辑。所述方法确定了许多改良的ABE,包括Var-3,其具有以下氨基酸序列:

[1243] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRREQEIKAQKKAQSSTDSSGGSSGGSSGSETPGTSESATPESSGGSSGSSSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVHNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSSGGSSGSSGSETPGTSESATPESSGGSSGSSGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVSDTKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFILVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFP

LKDNREKIEKILTFRIPYYVGPLARGNSRFAMTRKSEETITPWNFEVVVDKGASAQSFIERMTNFDKNLPNEKVL
PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVE
DRFNASLGTYHDLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKG
ILQTVKVVDDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLY
LYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLN
AKLITQRKFDNLTAKAERGGLSELDKAGFIKRLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVS
DFRKDFQFYKVVREINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYS
NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPPKYGGMqPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSEKPNIDFLEAKGYKEVKK
DLIIKLPKYSLEFELNGRKRMLASakfLQKGNELALPSKYVNFLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDE
IIIEQISEFSKRVIADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPrAFKYFDTTIIaRKeYrSTKEVLD
ATLIHQSIITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[1244] 使用2.5 μ gVar-3 ABE mRNA和1000ng gRNA 191长度20nt的Neon电穿孔法转染HEK293T-E342K。

[1245] 为spCas9碱基编辑器所提供作为sgRNA的gRNA骨架如下：

[1246] 5'-GUUUUAGAGCUAGAAAUAGCAAGUUAUUAAUAAGGCUAGUCCGUUAUCAAC
UUGAAAAAGUGGCACCGAGUCGGUGCUUUU-3'

[1247] 可用于本发明方法的gRNA包括：

[1248] 5'-ACCAUCGACAAGAAAGGGACUGAGUUUUAGAGCUAGAAAUAGCAAGUUAUUAAUAAGGCUAGUCCG
UUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU-3'；

[1249] 5'-CCAUCGACAAGAAAGGGACUGAGUUUUAGAGCUAGAAAUAGCAAGUUAUUAAUAAGGCUAGUCCG
UAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU-3'；

[1250] 5'-CAUCGACAAGAAAGGGACUGAGUUUUAGAGCUAGAAAUAGCAAGUUAUUAAUAAGGCUAGUCCGU
UAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU-3'；

[1251] 5'-AUCGACAAGAAAGGGACUGAGUUUUAGAGCUAGAAAUAGCAAGUUAUUAAUAAGGCUAGUCCGU
UCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU-3'；

[1252] 5'-UCGACAAGAAAGGGACUGAGUUUUAGAGCUAGAAAUAGCAAGUUAUUAAUAAGGCUAGUCCGU
CAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU-3'；and

[1253] 5'-CGACAAGAAAGGGACUGAGUUUUAGAGCUAGAAAUAGCAAGUUAUUAAUAAGGCUAGUCCGU
AACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU-3'。

[1254] 用mRNA和sgRNA进行的编辑主要在A5和A7处是未连锁的。Var-3 ABE的性能大大优于其他ABE,并且总有益校正率超过35% (仅A7=WT)+(A5+A7=WT+D341G-分泌无变化)。

[1255] 实施例5. 优化gRNA长度以校正HEK293T细胞中的E342K

[1256] 使用Neon系统,一式两份将ABE Var-3的Synthego gRNA和mRNA转染到HEK293T-E342K中。与A5位置相比,目标互补长度为18和19个核苷酸的gRNA在所需的A7位置产生更好的编辑率(图5)。所述模式与截短的gRNA缩短R环可及区域并因此降低靠近gRNA末端位置的编辑的机制一致。

[1257] 测试了18和19个核苷酸的gRNA和ABE Var-3在诱导的多能干细胞来源的E342肝细

胞和PiZZ小鼠模型中校正E342K突变的能力。使用NGC PAM进化来产生具有改变的PAM结合特异性的SpCas9突变体(图6A、6B)。选择这些SpCas9突变体可丰富Cas9的PAM相互作用(PI)结构域内的突变。筛选所述库中具有改变的PAM特异性的SpCas9。

[1258] 实施例6. 基因编辑以校正1a型糖原贮积障碍(冯吉尔克病)

[1259] 1型糖原贮积障碍(也称为GSD1或冯吉尔克病)是一种遗传性疾病,导致糖原分解和糖异生不足,组织中糖原积累,导致严重的低血糖症和乳酸性酸中毒,可能对CNS造成损害。在美国,大约每10万名新生儿中就有1名患有GSD1。

[1260] GSD1有两种类型,分别是1a型(GSD1a)和1b型(GSD1b),它们是由不同的基因突变引起的。GSD1a是由葡萄糖6磷酸酶(G6PC)的突变引起的,影响了约80%的GSD1患者。约25%的白种人患者携带隐性突变Q347*。当前的治疗方案包括在两餐之间定期或连续喂食玉米淀粉(淀粉酶将淀粉直接转化为葡萄糖)。

[1261] 校正Q347X突变的碱基编辑策略

[1262] 直接校正Q347*以恢复G6PC的表达并使葡萄糖代谢正常化。通过在目标位置有效转换A>G,可以使用碱基编辑器来校正Q347X。代表性的目标部位(突出显示)在图7中示出。在此位点进行精确校正将产生以下转换:TAG>CAG(终止密码子>谷氨酰胺)。碱基编辑器可以使用NGG PAM识别序列或NGA PAM识别序列。组织和递送策略可包括肝脂质纳米颗粒(LNP)递送。

[1263] 带有Q347X突变的iPSc衍生肝细胞的体外转染

[1264] 使用体外转染方法在iPSc衍生的肝细胞(Definigen,批次00419F002)中测试了碱基编辑。GSD1a iPSc衍生的肝细胞是复合杂合体(Q347X/G222R),具有Q347X突变。将GSD1a细胞铺板并使其成熟。

[1265] 如图8A和8B中所示,基于GSD1a iPSc衍生的肝细胞的已知成熟周期选择转染时间表。图8A提供了转染时间表的时间表,其显示了接种,转染和细胞收获的代表性时间点。图8B显示了在第5天和第7天成熟的GSD1a iPSc来源的肝细胞的代表性图像。成熟后(例如第12天),用碱基编辑器ABE7.10 VRQR/gRNA 272转染GSD1a细胞。转染后48-72小时(例如第14天),收获GSD1a转染的细胞用于gDNA。

[1266] 在图9A和9B中示出了用于GSD1a的G6PC Q347X的代表性碱基编辑精确校正数据。如图9A所示,使用NGA PAM或NGG PAM序列检查了HEK293T细胞中G6PC Q347X对ABE-0n靶标,ABE-Bystander,插入/缺失和核酸酶-插入/缺失的碱基编辑效率。使用NGG PAM的G6PC的靶向/插入序列如下所示:

[1267] **gga cct agg cga ggc agt ag ggg**

[1268] 上面的靶标/插入序列包含两个“a”核碱基,分别对应于旁观者(斜体和底线显示)和靶标(粗体和底线显示)。

[1269] 使用NGA PAM的G6PC的靶向/插入序列如下所示:

[1270] **ga cct agg cga ggc agt ag gga**

[1271] 上面的靶标/插入序列含有两个“a”核碱基,分别对应于旁观者(斜体和底线所示)和靶标(粗体和底线显示)。

[1272] NGA PAM gRNA 272在低插入缺失且没有可检测的旁观者V384A的HEK293T细胞中

对Q347X产生了>40%的精确校正(图9A)。因此,使用HEK293细胞获得了高的碱基编辑效率。

[1273] 如图9B所示,使用NGA PAM或NGG PAM序列检查了患者iPSc衍生的肝细胞中G6PC Q347X对ABE-On靶标,ABE-Bystander,插入和核酸酶-插入s的碱基编辑效率。使用NGA (n=4) 和NGG (n=2),旁观者V384A可以忽略不计的插入,观察到患者iPSc衍生的肝细胞中G6PC Q347X的相似编辑。如图9B所示,在杂合患者iPS衍生的Q347X肝细胞中的精确校正导致约8%-15%的A>G转化效率。虽然低于HEK293细胞,但与旁观者相比,Q347X iPSc肝细胞中的碱基编辑产生了更清晰的结果。

[1274] 患者iPSc衍生肝细胞中ABE变体的A到G转换效率

[1275] 测试了G6PC Q347X患者iPSc衍生的肝细胞中ABE变体的A-G转化效率。在培养的第12天,产生mRNA变体(TriLink pBxt464、MSP464、MSP465、MSP471),并转染到GSD1a患者iPSc来源的肝细胞中。在第14天,收获细胞用于gDNA分离并进行PAS染色。

[1276] 使用GGA PAM的G6PC的靶向/插入序列如下所示:

[1277] **gacctaggcgaggcagtagg**

[1278] 上面的靶标/插入序列包含两个“a”核碱基,分别对应于旁观者(斜体和底线显示)和靶标上(粗体和底线显示)。

[1279] Q347X的碱基编辑校正效率的百分比在mRNA变体之间相似,对于杂合序列而言,具有约10%的A至G转化效率(图10)。

[1280] 实施例7. 基因编辑以校正1型粘多糖贮积病(贺勒氏症)

[1281] 1型粘多糖贮积病(也称为MPS1或贺勒氏症)是一种罕见的常染色体隐性溶酶体贮积病,大约每200,000例出生中发生一次。MPS1的特征是骨骼异常,认知障碍,心脏病,呼吸系统疾病,肝脏和脾脏肿大以及预期寿命缩短。MPS1是由 α -L-艾杜糖醛酸酶基因(α -L-iduronidase, IDUA)的突变引起的,导致 α -L-艾杜糖醛酸酶的缺乏,这对于溶酶体糖胺聚糖的分解至关重要。

[1282] 以下提供了在NCBI参考序列号NP_032351.2下找到的代表性小鼠 α -L-艾杜糖醛酸酶(IDUA)蛋白的氨基酸序列:

1 MRPPRPSSAM LTFFAAFLAA PLALAESPYL VRVDAARPLR PLLPFWRSTG FCPPLPHDQA
61 DQYDLSWDQQ LNLAYIGAVP HSGIEQVRIH WLLDLITARK SPGQGLMYNF THLDAFLDLL
121 MENQLLPGE LMGSPSGYFT DFDDKQQVFE WKDLVSLAR RYIGRYGLTH VSKWNFETWN
181 EPDHHDFDNV SMTTQGLFNY YDACSEGLRI ASPTLKLGGP GDSFHPLPRS PMCWSLLGHC
241 ANGTNFFTGE VGVRLDYISL HKKGAGSSIA ILEQEMAVVE QVQQLFPEFK DTPIYNDEAD
[1283] 301 PLVGWSLPQP WRADVITYAAL VVKVIAQHQN LLFANSSSSM RYVLLSNDNA FLSYHPYPFS
361 QRTLTARFQV NNTHPPHVQL LRKPVLTVMG LMALLDGEQL WAEVSKAGAV LDSNHTVGVL
421 ASTHHPEGSA AAWSTTVLIY TSDDTHAHPN HSIPVTLRLR GVPPGLDLVY IVLYLDNQLS
481 SPYSAWQHMG QPVFPSAEQF RRMVMVEDPV AEAPRPFPAR GRLTLHRKLP VPSLLLVHVC
541 TRPLKPPGQV SRLRALPLTH GQLILVWSDE RVGSKCLWTY EIQFSQKGEE YAPINRRPST
601 FNLVVFSPDT AVVSGSYRVR ALDYWARPGP FSDPVTYLDV PAS

[1284] 以下提供了在NCBI参考序列号NM_008325.4下找到的代表性小鼠 α -L-艾杜糖醛酸酶(IDUA)核酸序列:

1 ctctgtgcc acccactgcc aagagggaca ggtctcaaag gtcagggcag tgtcccggga
61 aggagggcat cggctcctgg gagcggcctt aggacgcggg gtggactctc accatgcac
121 aggaagccag ccagtcccca gatgaagtcc gagcagaggt ggcagaagag cacctacagg
181 cctccagcga gaccgagaca gccgcaagaa taatggccgc tctgagacac ccaagcactg
241 ctaatgttgg ttccattttt ggagcgctg ggacgcagcg gaactcgcca gcacggggcg
301 gcgcgtgact gggttccttt ttgtccggc ctggcgagag gtcacgtggg gcgttacga
361 gaggcggaac actgcgaccg ccgcctaaaa agcttgctgt ttggggcac ctggatatcc
421 caaccatgcg acccccgcgt cctcctcag ctatgtgac gtttttct gcgttcttg
[1285] 481 ccgcgcctt ggcgtggct gactcaccgt acctggtgct tgtggacga gcccgccgc
541 tgaggcctct gttgcccttc tggaggagca ccggcttctg cccccactg cctcacgac
601 aggtgacca gtacgacctt agttgggacc agcaactgaa cctgcctac ataggtgccg
661 tacctcacag tggcattgag caggtccgga tacactggct gctggatctc atcacagca
721 ggaagtcacc tgggcaggga cttatgtaca acttcacca ctggatgca ttcttgacc
781 ttctcatgga gaaccagctt ctccctggat ttgagctcat gggcagtcct tctgggtact
841 tcacggactt tgatgacaag cagcaggtgt ttgaatggaa ggacctggtt tctctcttg
901 ccaggagata cattggtagg tatgggctga cacacgttc caagtggaac ttgagactt
961 ggaatgaacc agaccacat gactttgaca acgtgtccat gaccacacaa ggcttcctga
1021 attactatga tgcctgctct gaggggctgc gcattgccag cccactttg aagtgggtg

1081 gtctctgggga ttcttccac cccctgccaa ggtcaccaat gtgctggagc ctctctgggtc
1141 actgtgccaa tggaaccaac ttcttactg gcgaggtggg cgtgcgtctg gattacatct
1201 ccctgcacaa gaagggtgca ggtagctcca tcgccatcct ggagcaggag atggcagttg
1261 tggagcaggt ccagcagctc ttccctgagt tcaaggatac cctatttac aatgacgagg
1321 cagaccctct ggtagggctgg tccctgccac aacctggag agctgatgtg acttatgcgg
1381 ccttggtggt gaaggtcatt gcacagcacc agaacctgct gtttgccaac agcagttcct
1441 ccatgcgcta tgtgtctctc agcaatgaca atgccttctt gagctaccac ccgtaccctt
1501 tctcccagcg cacacttact gctcgattcc aggtcaacaa tactcaccca ccccacgtgc
1561 agttgctgcg aaagccagta ctacagtca tggggctcat ggccctgttg gatggagaac
1621 aactctgggc agaggttca aaggctgggg ctgtgttgga cagcaatcat acagtgggtg
1681 tcttggccag caccatcac cctgaaggt cgcagcggc ctggagtacc acagtctca
1741 tetacactag tgatgacacc cagcacacc ccaaccacag tatccctgtg actcttcgcc
1801 tgcgtggggg acctctggc ttggatctt tctacatagt acttactta gacaatcaac
1861 tcagcagccc ctacagtgcg tggcagcaca tgggccagcc agtcttccc tctgcagagc
1921 agttccgacg tatgcgatg gtggaggacc ccgtggctga ggcaccacgc ccttttctg
1981 ctagggggcg cctgacctc caccggaagc ttccggtgcc atcactctg ctggtgcatg
2041 tatgcacag cccctgaag ccacctgggc aggtcagccg gctccgtgca ctgccctga
2101 cacatggaca gctgatttg gtctgtcag atgagcgtgt gggctccaag tgcctgtgga
[1286] 2161 catatgagat ccagtttcc cagaaagggt aagagtatgc cccaatcaac aggaggccgt
2221 ctacttttaa cctctttgt ttacgccag acacagctgt ggtctctggc tctaccgag
2281 ttcgagcatt ggattactgg gcccgccag gcccttctc cgacctgtg acttacctgg
2341 atgtccctgc ctcatgagag ccactggctc ctagtactt gtgaatctgt gctgactggt
2401 gaatggagtc aaccagtatg agctagactg ccattagcta ggcagctgac tgcagcttc
2461 tattgttctt cccctattt cctttaaagt gtcttctct acctcagact tagggtcagt
2521 ctttgtggct aagcacttta taggcccagt tggagtgacc ttgcccacc ttctcccca
2581 tgcccagctg ttcaaaaagt ttaaatgtgg gactggaaag atggctcagt agataaagt
2641 ctgtctgtgc aggcccagg acttgtgtt agatactag cactcatgta taggctgggc
2701 atggcaatat atgcctattg tctagtgtt ggtggaagg acagagacag gccagggttc
2761 cctggccttc cagtctacct gaaactgcaa gctccaggtt cagtaagaaa cctgtttta
2821 gaaaaatcaa gtagagtgt tggtagacac acacacacac acacacagag
2881 tctaaattta gtttcttgag ctctgtaat atcaaaaatg aagttatgta ctctgaaat
2941 acaatactgc acagagtaag catcttcatt ccaacaggaa aaagaaatga cagggaagga
3001 tttagtga acaagaccaa agcatagcaa gacaaacgtt aaatcctgca gctccattct
3061 cagcatcggg gcccatgac ctgtgatgt ctggacagtc tgtgtctgtg gtgttgcct
3121 ttccagccgc catgacctt ctctaggct ggtgtcttg gcttctatt gattcgataa
3181 accgcgatga atgagcagaa gcactgggt aggggaagct tgacttact tgtattccta
3241 cattacagtc tatcatcgaa ggcagtcagg caggcacctg gaggcaggaa gtcattggaga

3301 ggccatggag ggggtgctgt tactcagact atgatcttac acatcccggg atcaccagcc
3361 aaagggtggg cccccacca caacggtctg gacctccca catcaatcac tagtttaaga
3421 aaacaggctt atctataggt caatcttgtt ggggcatttt tctcaattga ggttccttt
3481 tcccaaatga ctctagcttg taataaactg aaataaagcc accccaatct tgccacacat
3541 tacctggcct cccaagtctt ccttgaaat ctgggtggaa gccaacataa cctgtcact
3601 gtaactctga tattctacct tctaagccag catcctgtgg atcacagcct actatcagct
3661 tgagtgttag ttgaggactc ctgggtcctc catggctaca ataagcatga agtgcctgag
3721 gcttagtccc atgcatttgt ggagaatatt atgatgatga tatctagtag aggggggagg
3781 ctgttcacct caaaggaac aggaagtaga gtggggatag aattaagcta aatactctt
[1287] 3841 cactgaccgg ctttgtttta ctacgccctg catcctaaag ttctagaat ctccccaaac
3901 agacctatta cctggaaact accttaagg tgaagcctg ggctgatctt aataggctga
3961 tgttctacc ttggttctgg ccacgggaag aacctctct ctttagtac aaacctctgg
4021 tgtgccccac cagagagcct gttggacaca cgtgtcctta attcattctg cacattttc
4081 ttcttccat cagcacacag aagtcagca tacagaggtt tgtgctgaaa tgtaggcgta
4141 ctcccaagct ctcccagag tactatcacc tactgtccag gaaatgagtc tgagtgcagt
4201 gctcatatac actgcatggc tacatccaaa gtcagagttc ctctgccctc atgcctcttg
4261 aagtgaacga aatgtgatga cttctgcag ggtgttttt agtcctctgt ggaccttagg
4321 ctggccttgg catcttggct cacctgtccc agagtacta ctattaagat tacaggtgtg
4381 taccaccatg cctgccaatt acctctact ttaataaaaa tatgacattt

[1288] 以下提供了在NCBI参考序列号NP_000194.2下找到的代表性人 α -L-艾杜糖醛酸酶(IDUA)蛋白质的氨基酸序列:

1 MRPLRPRAAL LALLASLLAA PPVAPAEAPH LVHVDAARAL WPLRRFWRST GFCPLPHSQ
61 ADQYVLSWDQ QLNLAYVGAV PHRGIKQVRT HWLLELVTR GSTGRGLSYN FTHLDGYLDL
121 LRENQLLPGF ELMGSASGHF TDFEDKQQVF EWKDLVSSLA RRYIGRYGLA HVSKWNFETW
181 NEPDHHDFFD VSMTMQGFLN YYDACSEGLR AASPALRLGG PGDSFHTPPR SPLSWGLLRH
241 CHDGTNFFTG EAGVRLDYIS LHRKGARSSI SILEQEKVVA QQIRQLFPKF ADTPIYNDEA
301 DPLVGWSLPQ PWRADVTYAA MVVKVIAHQH NLLANTTSA FPYALLSNDN AFLSYHPHPF
[1289] 361 AQRILTARFQ VNNTRPPHVQ LLRKPVLTAM GLLALLDEEQ LWAEVSQAGT
VLDSNHTVG
421 LASAHRPQGP ADAWRAAVLI YASDDTRAHP NRSVAVTLRL RGVP PGPLV YVTRYLDNGL
481 CSPDGEWRRL GRPVFPTAEQ FRRMRAAEDP VAAAPRPLPA GGRLTLR PAL RLPSLLLHVH
541 CARPEKPPGQ VTRLRALPLT QGQLVLVWSD EHVGSKCLWT YEIQFSQDGK AYTPVSRKPS
601 TFNLFVFS PD TGAVSGSYRV RALDYWARPG PFSDPV PYLE VPVPRGPPSP GNP

[1290] 以下提供了在NCBI参考序列号NM_000203下找到的代表性人 α -L-艾杜糖醛酸酶(IDUA)核酸序列:

1 agtgcagccc gaagccccgc agtccccgag cacgcgtggc catgcgtccc ctgcgcccc
 [1291] 61 gcgcgcgcgt gctggcgctc ctggcctcgc tctggccgc gccccgggtg gccccggccg
 121 aggccccgca cctggtgcat gtggacgcgg ccgcgcgcgt gtggccctg cggcgttct
 181 ggaggagcac aggccttctgc ccccgcctgc cacacagcca ggctgaccag tacgtctca
 241 gctgggacca gcagctcaac ctgcctatg tgggcgcctt ccctaccgc ggcataaagc
 301 aggtccggac cactggctg ctggagcttg tcaccaccag ggggtccact ggacggggcc
 361 tgagctacaa cttaccacac ctggacgggt acctggacct tctcaggag aaccagctcc
 421 tcccagggtt tgagctgatg ggcagcgctt cgggccactt cactgacttt gaggacaagc
 481 agcaggtgtt tgagtgaag gacttggctt ccagcctggc caggagatac atcggtaggt
 541 acggactggc gcatgtttcc aagtggaaact tcgagacgtg gaatgagcca gaccaccag
 601 actttgacaa cgtctccatg accatgcaag gcttctgaa ctactacgat gcctgctcg
 661 agggcttgcg cgcgcgcagc ccgccttgc ggctgggagg ccccgccgac tcttccaca
 721 cccaccgcg atccccgtg agctggggcc tctgcgcca ctgccacgac ggtaccaact
 781 tcttactgg ggaggcgggc gtgcggctgg actacatct cctccacagg aagggtgcgc
 841 gcagctccat ctccatctg gagcaggaga aggtcgtcgc gcagcagatc cggcagctt
 901 tcccaagt cgcggacacc ccaattaca acgacgaggc ggaccgcgtg gtgggtggt
 961 ccttgcaca gccgtggagg gcggacgtga cctacgcggc catggtgtg aaggtcatc
 1021 cgcagcatca gaacctgcta ctggccaaca ccacctcgc ctccctac gcgtcctga
 1081 gcaacgacaa tgccttctg agctaccacc cgcacctt cgcgcagcgc acgtcaccg
 1141 cgcgttcca ggtcaacaac acccgcccgc cgcacgtgca gctgtgcgc aagccggtgc
 [1292] 1201 tcacggccat ggggtgctg gcgtgctgg atgaggagca gctctggcc gaagtgtgc
 1261 aggccgggac cgtcctggac agcaaccaca cgggtggcgt cctggccagc gccaccgcc
 1321 cccaggggcc ggccgacgcc tggcgcgcgc cgggtgctgat ctacgcgagc gacgacacc
 1381 gcgcccacc caaccgcagc gtgcgggtga cctgcggct gcgcggggtg cccccggcc
 1441 cgggcctggt ctacgtcacg cgtacctgg acaacgggct ctgcagcccc gacggcgagt
 1501 ggcggcgcct gggccggccc gtcttccca cggcagagca gttccggcgc atgcgcgcg
 1561 ctgaggacce gttggccgcg gcgccccgc cttaccgc cggcgggcgc ctgacctgc
 1621 gccccgcgt gcggctgccg tcgttttgc tgggtcacgt gtgtgcgcgc cccgagaagc
 1681 cggcgggga ggtcacgcgc ctccgcgc tgccttgac ccaagggcag ctggttctgg
 1741 tctgtcgga tgaacacgt ggctccaagt gcctgtggac atacgagatc cagtctctc
 1801 aggacggtaa ggcgtacacc ccggtcagca ggaagccatc gaccttaac ctcttctgt
 1861 tcagcccaga cacaggtgct gtcttggct cctaccgagt tcgagccctg gactactggg
 1921 cccgaccagg ccccttctgc gacctgtgc cgtacctgga ggtccctgtg ccaagagggc
 1981 cccatcccc gggaatcca tgacctgtg ctgagccca gtgggttgca cctccaccg
 2041 cagtcagcga gctggggctg cactgtccc atgtgccct cccatcccc ctttgcatt
 2101 atattttat attttattt ttcttttat atcttggtac caacgcccc tttaaagcgg
 2161 cttgcacag gtca

[1293] 在HEK293T细胞系中用小鼠和人类IDUA靶序列测试了碱基编辑(图11)。小鼠IDUA

的靶向/插入序列如下所示,并对应于上文所示的小鼠IDUA基因序列的核酸1612-1824。

[1294] **atggagaaca actctaggca gaggtctcaa aggctggggc tgtgttgac agcaatcata**
cagtgggtgt cctggccagc acccatcacc ctgaaggctc cgcagcggcc tggagtacca
[1295] **cagtctcat ctacactagt gatgacaccc acgcacaccc caaccacagt atccctgtga**
ctcttcgcct gcgtggggta cctcctggct tgg

[1296] 以上小鼠靶标/插入序列包含“a”核碱基(以粗体和底线显示),而小鼠IDUA基因序列在IDUA序列的位置1627处包含“g”核碱基。

[1297] 人类IDUA多核苷酸序列中的靶向/插入序列如下所示,并且对应于以上所示人类IDUA基因序列的核酸1231-1324-。

[1298] **atgaggagca gctctaggcc gaagtgtcgc aggccgggac cgtcctggac agcaaccaca**
cggtgggcgt cctggccagc gccaccgcc ccca

[1299] 上述人类靶标/插入序列包含“a”核碱基(以粗体和底线显示),而人类IDUA基因序列在IDUA序列的位置1246处包含“g”核碱基。

[1300] 对于质粒转染,将HEK293T细胞用250ng的gRNA质粒和750ng的碱基编辑质粒以30,000个细胞/孔的48孔板转染。使用具有NGG PAM序列的碱基编辑器SpCas9-ABE7.10。小鼠gRNA序列如下所示,并包含一个“a”核碱基(以粗体和底线显示):

[1301] **gctctaggccgaagtgtcgc agg**

[1302] 人类gRNA序列如下所示,并包含一个“a”核碱基(以粗体和底线显示):

[1303] **actctaggcagaggtctcaa agg**

[1304] 实施例8.材料和方法

[1305] 使用以下材料和方法获得本文所述实施例中提供的结果。

[1306] 克隆化

[1307] 本文描述了靶多核苷酸和gRNA的DNA序列以及所用的引物。对于gRNA,显示了以下支架序列:GUUUUAGAGC UAGAAAUAGC AAGUAAAAU AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUUUU。所述支架用于本文表中所示的PAM,例如,NGG、NGA、NGC、NGT PAM;等等。gRNA涵盖本文提供的或根据本领域技术人员的通常知识确定的疾病相关基因(例如表3A和3B)的支架序列和间隔子序列(靶序列)。在艺术上。(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);Gaudelli,N.M.等人“Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage”Nature 551,464-471(2017);Komor,A.C.等人“Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity”Science Advances 3:eaao4774(2017)以及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)。

[1308] 使用VeraSeq Ultra DNA聚合酶(Enzymatics)或Q5 Hot Start高保真DNA聚合酶

(New England Biolabs) 进行PCR。使用USER克隆(New England Biolabs) 构建碱基编辑器(BE) 质粒。脱氨酶基因合成为gBlocks基因片段(Integrated DNA Technologies)。下面列出了使用的Cas9基因。从先前报导的质粒获得Cas9基因。将脱氨酶和融合基因克隆到pCMV(哺乳动物密码子优化) 或pET28b(大肠杆菌密码子优化) 骨架中。使用定点诱变构建sgRNA表达质粒。

[1309] 简而言之,根据制造商的说明,使用T4多核苷酸激酶(New England Biolabs) 将上文列出的引物5' 磷酸化。接下来,根据制造商的说明,使用Q5热启动高保真聚合酶(New England Biolabs),以磷酸化引物和编码目的基因的质粒为模板,进行PCR。将PCR产物与DpnI(20U,New England Biolabs) 在37℃下孵育1小时,在QIAprep旋转柱(Qiagen) 上纯化,并根据制造商的说明使用QuickLigase(New England Biolabs) 连接。使用Mach1感受态细胞(ThermoFisher Scientific) 进行DNA载体扩增。

[1310] ssDNA的体外脱氨酶测定

[1311] 所有ssDNA受质的序列在下面提供。所有Cy3标记的受质均获自Integrated DNA Technologies(IDT)。使用TNT T7快速偶联转录/翻译试剂盒(Promega) 根据制造商的说明,在体外使用1μg质粒表达脱氨酶。蛋白表达后,将5μl裂解物与35μlssDNA(1.8μM) 和USER酶(1个单位) 在CutSmart缓冲液(New England Biolabs) (50mM乙酸钾,29mM Tris-乙酸盐,10mM乙酸镁) 中合并,100μgml-1BSA,pH 7.9) 于37℃孵育2小时。在10%TBE-脲凝胶(Bio-Rad) 上,将未切割的含U受质从全长未修饰的受质中分离出来。

[1312] His6-rAPOBEC1-接头-dCas9融合蛋白的表达和纯化。

[1313] 用质粒(例如,编码pET28b-His6-rAPOBEC1-接头-dCas9的质粒) 转化大肠杆菌BL21 STAR(DE3) 感受态细胞(ThermoFisher Scientific)。所得表达菌株在含有100μgml-1卡那霉素的Luria-Bertani(LB) 肉汤中于37℃过夜生长。将细胞以1:100的比例稀释到相同的生长培养基中,并在37℃下生长至OD600 ≈ 0.6。在2小时内将培养物冷却至4℃,并以0.5mM的浓度加入异丙基-β-D-1-硫代半乳糖吡喃糖苷(IPTG) 诱导蛋白质表达。约16小时后,以4,000g离心力收集细胞,并将其重悬于裂解缓冲液(50mM三(羟甲基)-氨基甲烷(Tris)-HCl(pH 7.5),1M NaCl,20%甘油,10mM tris(2-羧乙基) 膦(TCEP,Soltec Ventures))。通过超声(20s脉冲打开,20s脉冲关闭,在6W输出下总共8分钟) 进行超声处理裂解细胞,并在25,000g离心15分钟后分离出裂解物上清液。将裂解物与His-Pur的镍-氨乙酸(nickel-NTA) 树脂(ThermoFisher Scientific) 在4℃下孵育1小时,以捕获带His标签的融合蛋白。将树脂转移到柱上,并用40ml裂解缓冲液洗涤。将带有His标签的融合蛋白在补充了285mM咪唑的裂解缓冲液中洗脱,并通过超滤(Amicon-Millipore,截断分子量为100kDa) 浓缩至1ml总体积。将蛋白质在低盐纯化缓冲液中稀释至20ml,所述缓冲液包含50mM三(羟甲基)-氨基甲烷(Tris)-HCl(pH 7.0)、0.1M NaCl、20%甘油、10mM TCEP,并加载到SP Sepharose Fast Flow中树脂(GE生命科学)。用40ml这种低盐缓冲液洗涤树脂,并用5ml的含有50mM三(羟甲基)-氨基甲烷(Tris)-HCl(pH 7.0)、0.5M NaCl、20%甘油的活性缓冲液洗脱蛋白质、10毫米TCEP。洗脱的蛋白质通过SDS-PAGE定量。

[1314] sgRNA的体外转录

[1315] 根据制造商的说明,使用TranscriptAid T7高产量转录试剂盒(ThermoFisher Scientific) 在体外转录包含T7启动子和20bp sgRNA目标序列的线性DNA片段。sgRNA产品

使用MEGAclear试剂盒(ThermoFisher Scientific)根据制造商的说明进行纯化,并通过UV吸收进行定量。

[1316] Cy3偶联的dsDNA受质的制备

[1317] 通常,将未标记的序列链(例如80-nt未标记链的序列)订购为来自IDT的经PAGE纯化的寡核苷酸。订购了与每个80-nt受质物3'末端互补的25-nt Cy3标记的引物,作为IDT的HPLC纯化寡核苷酸。为了生成Cy3标记的dsDNA受质,将80nt的链(5 μ l的100 μ M溶液)与Cy3标记的引物(5 μ l的100 μ M溶液)在NEBuffer 2(38.25 μ l的50mM)中合并并将NaCl、10mM Tris-HCl、10mM MgCl₂、1mM DTT、pH 7.9溶液、New England Biolabs)与dNTPs(0.75 μ l的100mM溶液)加热至95 $^{\circ}$ C 5分钟,然后逐步冷却以每秒0.1 $^{\circ}$ C的速率升至45 $^{\circ}$ C。在此退火期之后,加入Klenow exo- (5U, New England Biolabs),并将反应液在37 $^{\circ}$ C孵育1小时。溶液用缓冲液PB(250 μ l, Qiagen)和异丙醇(50 μ l)稀释,并在QIAprep旋转柱(Qiagen)上纯化,用50 μ l Tris缓冲液洗脱。dsDNA上的脱氨酶测定。将纯化的融合蛋白(20 μ l 1.9 μ M的活性缓冲液)与1当量的适当sgRNA合并,并在环境温度下孵育5分钟。加入Cy3标记的dsDNA受质至终浓度125nM,并将所得溶液在37 $^{\circ}$ C下孵育2小时。通过添加缓冲液PB(100 μ l, Qiagen)和异丙醇(25 μ l)从融合中分离dsDNA,并在EconoSpin微旋转柱(EPOCH Life Science)上纯化,用20 μ l CutSmart缓冲液(New England Biolabs)洗脱。将USER酶(1U, New England Biolabs)添加到纯化的,编辑的dsDNA中,并在37 $^{\circ}$ C下孵育1h。通过将5 μ l反应溶液与15 μ l基于DMSO的上样缓冲液(5mM Tris、0.5mM EDTA、12.5%甘油、0.02%溴酚蓝、0.02%二甲苯)合并,将Cy3标记的链从其补体中完全变性。青色,80%DMSO)。在10%TBE-脲凝胶(Bio-Rad)上,将全长的含C受质与任何切割的含U的编辑受质分离,并在GE Amersham Typhoon成像仪上成像。

[1318] 制备用于高通量测序的体外编辑的dsDNA

[1319] 寡核苷酸获自IDT。在Tris缓冲液中合并互补序列(5 μ l的100 μ M溶液),并通过加热至95 $^{\circ}$ C持续5分钟进行退火,然后以0.1 $^{\circ}$ C/s的速率逐渐冷却至45 $^{\circ}$ C,以生成60-bp dsDNA受质。将纯化的融合蛋白(20 μ l 1.9 μ M活性缓冲液)与1当量的适当sgRNA合并,并在环境温度下孵育5分钟。加入60-mer dsDNA受质至终浓度125nM,并将所得溶液在37 $^{\circ}$ C下孵育2h。通过添加缓冲液PB(100 μ l, Qiagen)和异丙醇(25 μ l)从融合物中分离dsDNA,并在EconoSpin微旋转柱(EPOCH Life Science)上纯化,用20 μ l Tris缓冲液洗脱。通过高通量测序引物对和VeraSeq Ultra(Enzymatics)按照制造商的说明,通过PCR扩增得到的编辑后的DNA(以1 μ l为模板),并进行13个扩增循环。使用RapidTips(Diffinity Genomics)纯化PCR反应产物,并使用含有测序接头的引物通过PCR扩增纯化的DNA,进行纯化,并如前所述在MiSeq高通量DNA测序仪(Illumina)上测序。

[1320] 细胞培养

[1321] HEK293T(ATCC CRL-3216)和U2OS(ATCC HTB-96)在37 $^{\circ}$ C的Dulbecco's Modified Eagle's Medium加GlutaMax(ThermoFisher)中补充了10%(v/v)胎牛血清(FBS)、含5%的二氧化碳。将HCC1954细胞(ATCC CRL-2338)保持在如上所述补充的RPMI-1640培养基(ThermoFisher Scientific)中。含有目的基因的永生细胞(例如SERPINA1、G6PC、IDUA等)(Taconic Biosciences)在补充有10%(v/v)胎牛血清(FBS)的Dulbecco改良Eagle培养基加GlutaMax(ThermoFisher Scientific)中培养。和200 μ g/ml-1遗传霉素(ThermoFisher Scientific)。

[1322] 转染

[1323] 将HEK293T细胞接种在48孔胶原包覆的BioCoat平板(Corning)上,并以约85%汇合度转染。简而言之,根据制造商的协议,每孔使用1.5 μ l Lipofectamine 2000 (ThermoFisher Scientific)转染750ng BE和250ng sgRNA表达质粒。根据制造商的说明,使用适当的Amaya Nucleofector II程序转染HEK293T细胞(对于HEK293T细胞,使用程序Q-001的V试剂盒)。

[1324] 基因组DNA样品的高通量DNA测序

[1325] 3天后收获转染的细胞,并根据制造商的说明使用Agencourt DNAdvance基因组DNA分离试剂盒(贝克曼库尔特公司)分离基因组DNA。通过侧翼高通量测序引物对,通过PCR扩增了目标上和目标外的基因组区域。按照制造商的说明,使用Phusion高保真DNA聚合酶(ThermoFisher)进行PCR扩增,使用5ng基因组DNA作为模板。分别确定每个引物对的循环数,以确保反应在线性扩增范围内停止。PCR产物使用RapidTips (Diffinity Genomics)纯化。用含有测序接头的引物通过PCR扩增纯化的DNA。使用Quant-iT PicoGreen dsDNA分析试剂盒(ThermoFisher)和KAPA Library Quantification Kit-Illumina (KAPA Biosystems)对产品进行凝胶纯化和定量。如前所述(Pattanayak, Nature Biotechnol. 31, 839-843 (2013)),在Illumina MiSeq上对样品进行测序。

[1326] 数据分析

[1327] 测序读数使用MiSeq Reporter(Illumina)自动解复用,并使用自定义的Matlab分析单个FASTQ文件。使用Smith-Waterman算法,将每个读数成对对齐至适当的参考序列。Q得分低于31的碱基被Ns取代,因此在计算核苷酸频率时被排除在外。此处理产生的预期MiSeq碱基调用错误率约为1,000的1。将比对序列,其中阅读和参考序列不包含缺口被存储在比对表中,从所述表中可以列出每个基因座的基本频率。使用先前描述的标准(Zuris等人, Nature Biotechnol. 33, 73-80 (2015)),使用定制的Matlab脚本对插入频率进行定量(Zuris等, Nature Biotechnol. 33, 73-80 (2015))。如果没有精确匹配,则将其从分析中排除;如果此插入窗口的长度与参考序列完全匹配,则将其归类为不包含插入;如果插入窗口为两个如果一个或多个碱基比参考序列长或短,则将测序读数分别分类为插入或缺失。

[1328] 序列

[1329] 下表8列出了野生型和变体(E342K) SERPINA1多核苷酸编码的野生型和变体(E342K) SERPINA1氨基酸序列和开放阅读框(ORF)核酸序列的代表性列表。所描述的实施例。

[1330] 表8. 示例性序列

[1331]

	SEQ ID NO	序列
<i>SERPINA1</i> 氨基酸		MPSSVSWGILLLAGLCCLVPVSLAEDPQGDAQAQKTDTSHH QDHPTFNKITPNLAFAFSLYRQLAHQSNSTNIFFSPVSIATAF AMLSLGTKADTHDEILEGLNFNLTEIPEAQIHEGFQELLRTL QPDSQLQLTTGNGLFLSEGLKLVDKFLEDVKKLYHSEAFTVN FGDTEEAKKQINDYVEKGTQGKIVDLVKELDRDTVFALVNYI FFKGKWERPFVVKDTEEEEDFHVDQVTTVKVPMMLKRLGMFN IQHCKKLSSWVLLMKYLGNAIAIFFLPDEGKLQHLENELTHD IITKFLNEDRRSASLHLPKLSITGTDLKSVLGLGITKVFNS GADLSGVTEEAPLKLSKAVHKAVLTIDEKGTAAAGAMFLEAI PMSIPPEVKFNKPFVFLMIEQNTKSPLFMGKVVNPTQK
<i>SERPINA1</i> ORF		ATGCCGTCTTCTGTCTCGTGGGGCATCCTCCTGCTGGCAGG CCTGTGCTGCCTGGTCCCTGTCTCCCTGGCTGAGGATCCCC AGGGAGATGCTGCCCAGAAGACAGATACATCCCACCATG ATCAGGATCACCCAACCTTCAACAAGATCAACCCCAACCT GGCTGAGTTTCGCCTTCAGCCTATACCGCCAGCTGGCACAC CAGTCCAACAGCACCAATATCTTCTTCTCCCCAGTGAGCA TCGCTACAGCCTTTGCAATGCTCTCCCTGGGGACCAAGGC TGACACTCACGATGAAATCCTGGAGGGCCTGAATTTCAAC CTCACGGAGATTCCGGAGGCTCAGATCCATGAAGGCTTCC AGGAACTCCTCCGTACCCTCAACCAGCCAGACAGCCAGCT CCAGCTGACCACCGCAATGGCCTGTTCTCAGCGAGGGC CTGAAGCTAGTGGATAAGTTTTTGGAGGATGTTAAAAAGT TGTACCACTCAGAAGCCTTCACTGTCAACTTCGGGGACAC CGAAGAGGCCAAGAAACAGATCAACGATTACGTGGAGAA GGGTACTCAAGGGAAAATTGTGGATTTGGTCAAGGAGCTT GACAGAGACACAGTTTTTGTCTCTGGTGAATTACATCTTCTT TAAAGGCAAATGGGAGAGACCCTTTGAAGTCAAGGACAC CGAGGAAGAGGACTTCCACGTGGACCAGGTGACCACCGT GAAGGTGCCTATGATGAAGCGTTTAGGCATGTTTAACATC CAGCACTGTAAGAAGCTGTCCAGCTGGGTGCTGCTGATGA AATACCTGGGCAATGCCACCGCCATCTTCTCCTGCCTGAT

[1332]

		GAGGGGAAACTACAGCACCTGGAAAATGAACTCACCCAC GATATCATCACCAAGTTCCTGGAAAATGAAGACAGAAGGT CTGCCAGCTTACATTTACCCAAACTGTCCATTACTGGAACC TATGATCTGAAGAGCGTCCTGGGTCAACTGGGCATCACTA AGGTCTTCAGCAATGGGGCTGACCTCTCCGGGGTCACAGA GGAGGCACCCCTGAAGCTCTCCAAGGCCGTGCATAAGGCT GTGCTGACCATCGACGAGAAAGGGACTGAAGCTGCTGGG GCCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCGA GGTCAAGTTCAACAAACCCTTTGTCTTCTTAATGATTGAAC AAAATACCAAGTCTCCCCTCTTCATGGGAAAAGTGGTGAA TCCCACCCAAAAA
<i>SERPINA1</i> E342K 氨基酸		MPSSVSWGILLLAGLCCLVPVSLAEDPQGDAAQKTDTSHH QDHPTFNKITPNLAFAFSLYRQLAHQSNSTNIFFSPVSIATF AMLSLGTKADTHDEILEGLNFNLTEIPEAQIHEGFQELLRTL NPDSQLQLTTGNGLFLSEGLKLVDKFLEDVKKLYHSEAFV NFGDTEEAKKQINDYVEKGTQGKIVDLVKELDRDTVFALV NYIFFKGKWERPFVKDTEEEEDFHVDQVTTVKVPMMKRL GMFNIQHCKKLSSWVLLMKYLGNAIAIFLPDEGKLQHLE NELTHDIITKFLNEDRRSASLHLPKLSITGTYDLKSVL GQLGITKVFSGADLSGVTEEAPLKLSKAVHKAVLTIDK KGTEAAGAMFLEAIPMSIPPEVKFNKPFVFLMIEQNTK SPLFMGKVVNPTQK
<i>SERPINA1</i> E342K ORF		ATGCCGTCTTCTGTCTCGTGGGGCATCCTCCTGCTGGCAGG CCTGTGCTGCCTGGTCCCTGTCTCCCTGGCTGAGGATCCCC AGGGAGATGCTGCCCAGAAGACAGATACATCCCACCATG ATCAGGATCACCAACCTTCAACAAGATACCCCCAACCT GGCTGAGTTCGCCTTCAGCCTATACCGCCAGCTGGCACAC CAGTCCAACAGCACCAATATCTTCTTCTCCCCAGTGAGCA TCGCTACAGCCTTTGCAATGCTCTCCCTGGGGACCAAGGC TGACACTCACGATGAAATCCTGGAGGGCCTGAATTTCAAC CTCACGGAGATTCCGGAGGCTCAGATCCATGAAGGCTTCC AGGAACTCCTCCGTACCCTCAACCAGCCAGACAGCCAGCT CCAGCTGACCACCGGCAATGGCCTGTTCTCAGCGAGGGC CTGAAGCTAGTGGATAAGTTTTTGGAGGATGTAAAAAGT TGTACCACTCAGAAGCCTTCACTGTCAACTTCGGGGACAC CGAAGAGGCCAAGAAACAGATCAACGATTACGTGGAGAA GGGTACTCAAGGGAAAATTGTGGATTGTTCAAGGAGCTT GACAGAGACACAGTTTTTGTCTGCTGGTGAATTACATCTTCT TAAAGGCAAATGGGAGAGACCCTTTGAAGTCAAGGACAC CGAGGAAGAGGACTTCCACGTGGACCAGGTGACCACCGT GAAGGTGCCTATGATGAAGCGTTTAGGCATGTTTAACATC

[1333]

CAGCACTGTAAGAAGCTGTCCAGCTGGGTGCTGCTGATGA AATACCTGGGCAATGCCACCGCCATCTTCTTCCTGCCTGAT GAGGGGAAACTACAGCACCTGGAAAATGAACTCACCCAC GATATCATCACCAAGTTCCTGGAAAATGAAGACAGAAGGT CTGCCAGCTTACATTTACCCAACTGTCCATTACTGGAACC TATGATCTGAAGAGCGTCCTGGGTCAACTGGGCATCACTA AGGTCTTCAGCAATGGGGCTGACCTCTCCGGGGTCACAGA GGAGGCACCCCTGAAGCTCTCCAAGGCCGTGCATAAGGCT GTGCTGACCATCGACaAGAAAGGGACTGAAGCTGCTGGGG CCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCCAG GTCAAGTTCAACAAACCCTTTGTCTTCTTAATGATTGAACA AAATACCAAGTCTCCCCTCTTCATGGGAAAAGTGGTGAAT CCCACCCAAAAA

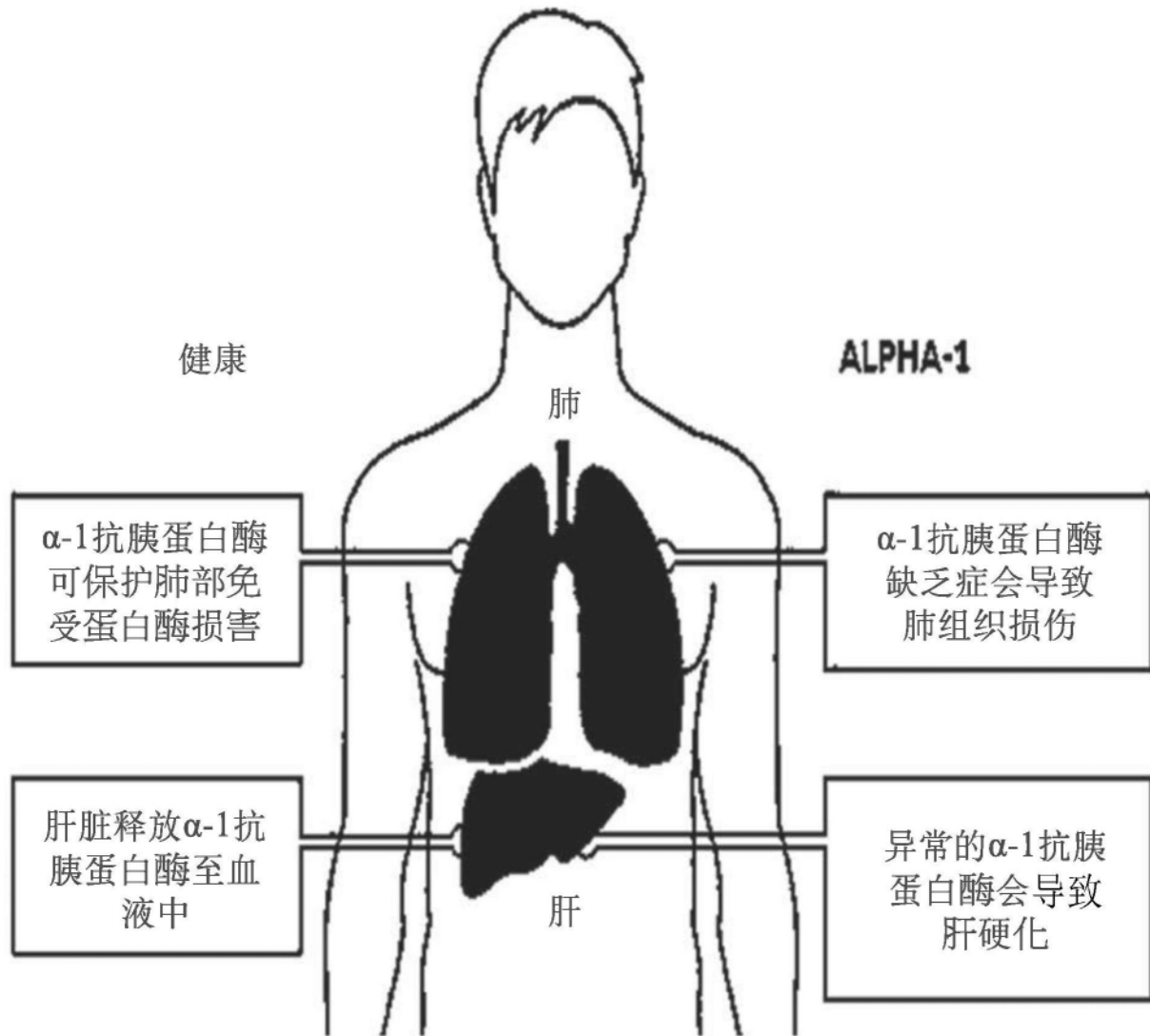


图1

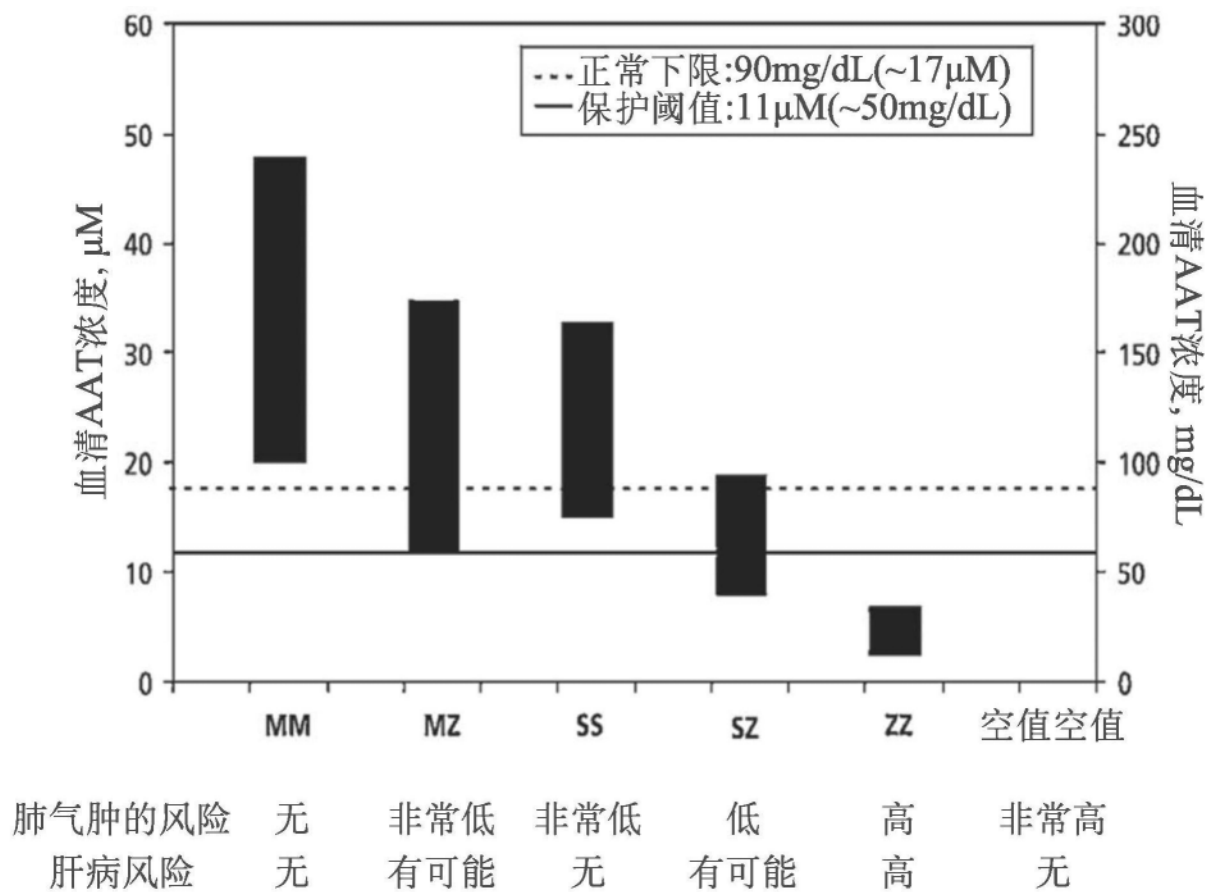


图2

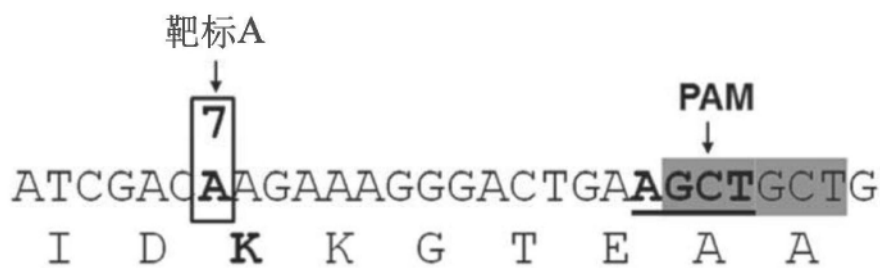


图3A

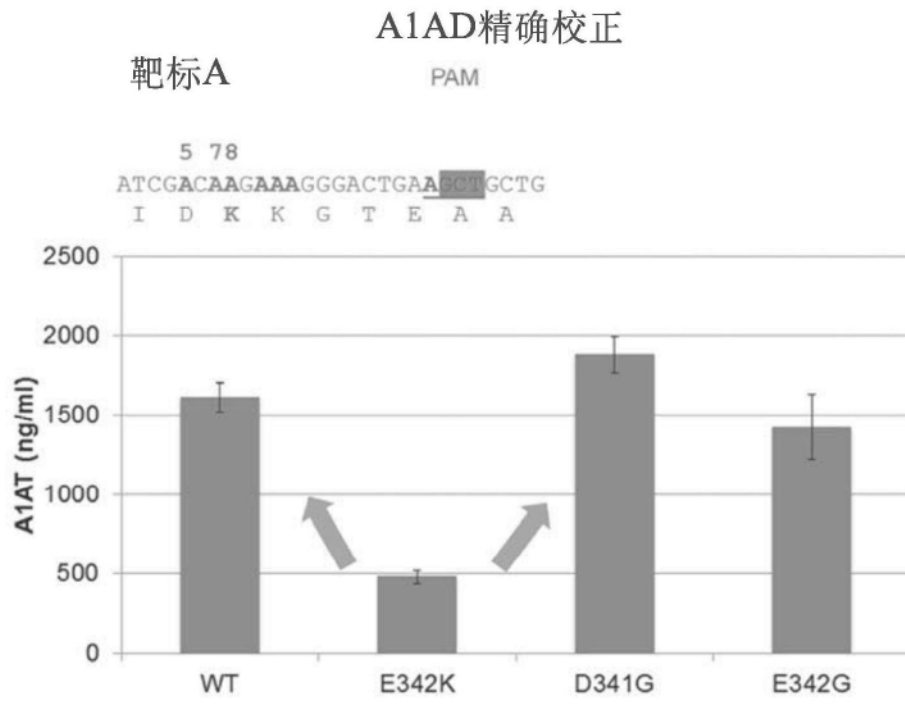


图3B

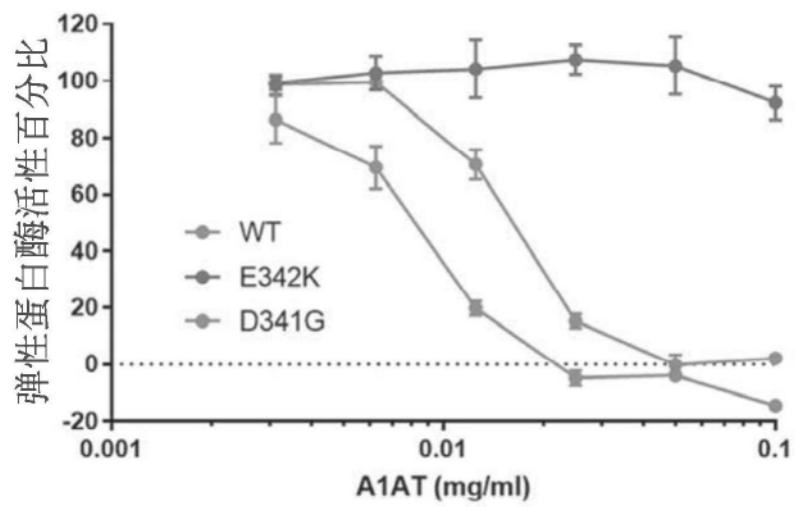


图3C

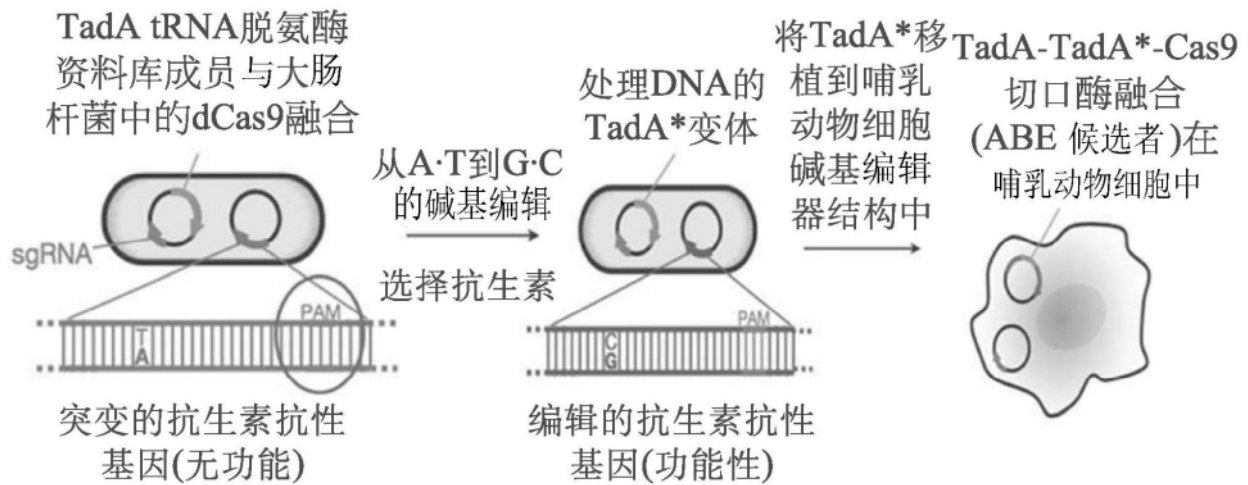


图4

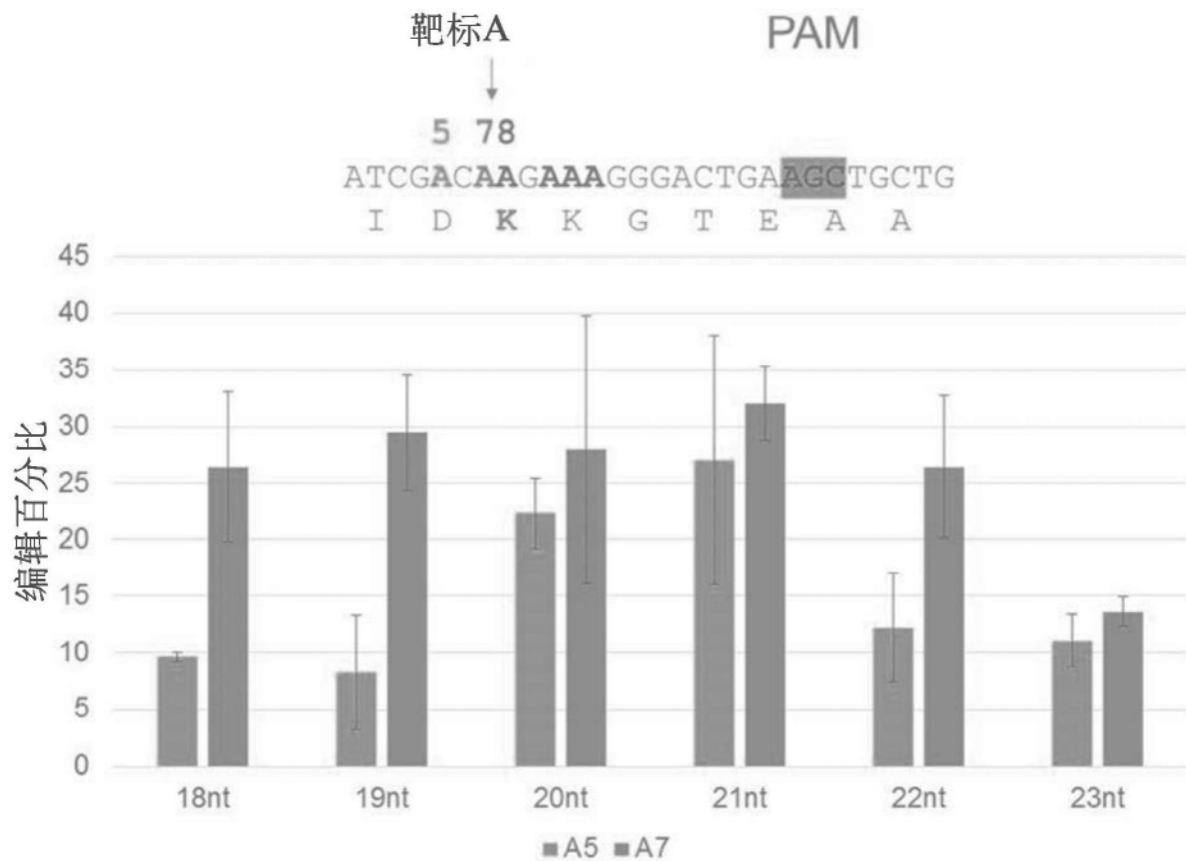


图5

NGC PAM演变
自资料库665第2b轮的富集变异

A	0	0	0	0	0	0	0	0	0	12.7	0	0	0	0	35.9	0	0	0	0
C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16	0	0
D	0	0	0	0	0	0	0	9.39	0	0	0	87.3	0	0	0	0	0	0	0.55
E	0	0	0	0	0	0	0	86.7	87.3	1.1	0	0	49.2	0	64.1	0	0	0	0
F	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G	0	0	0	0	0	0	0	12.7	0	0	12.7	0	0	0	0	0	0	0	1.66
H	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4.97	0	0
I	0	0	0	0	0	0	0	0	0	0	0	0	49.7	0	0	0	0	0	0
K	84.5	0	0	59.7	90.1	85.1	0	0.55	0	65.2	0	0	50.8	0	0	0	0	0	0
L	0	74	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	33.7	0	0	0	0	0	0	0	0	18.2
N	0	0	0	0	0	0	90.6	0	0	0	0	49.2	0	0	0	1.66	0	0	0
P	0	26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.1	0	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R	15.5	0	0	40.3	9.94	14.9	0	0	0	0	0	0	64.1	35.9	0	63	0	0	0
S	0	0	92.8	0	0	0	0	0	0	0	0	1.1	0	0	0	0	0	0	0
T	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	35.9	0	0
V	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	79.6	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Y	0	0	7.18	0	0	0	0	0	0	0	0	0	0	0	0	77.3	0	0	0
B	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
X	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Z	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
J	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
O	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

图6A

[illegible]

图6B

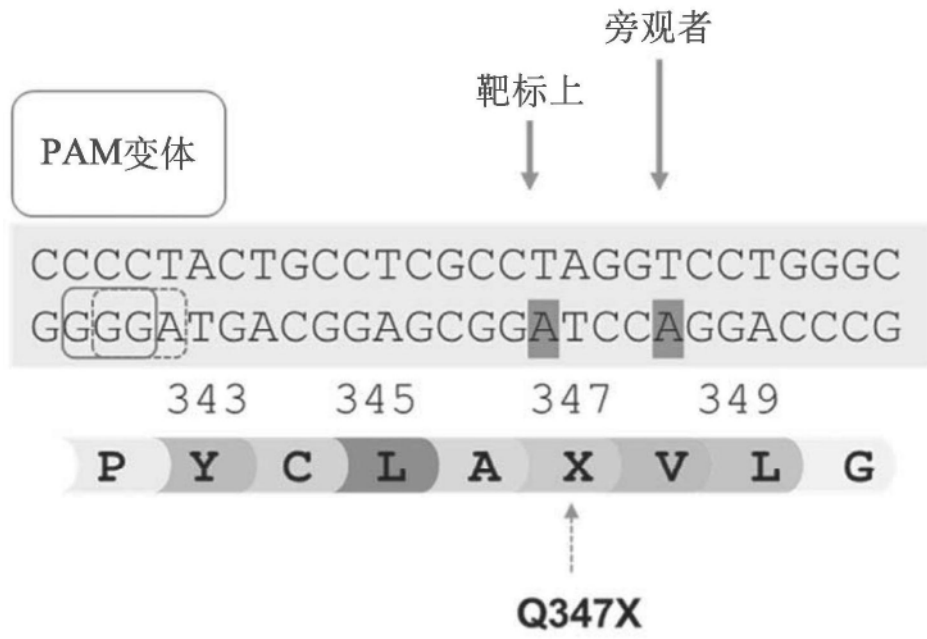


图7



图8A

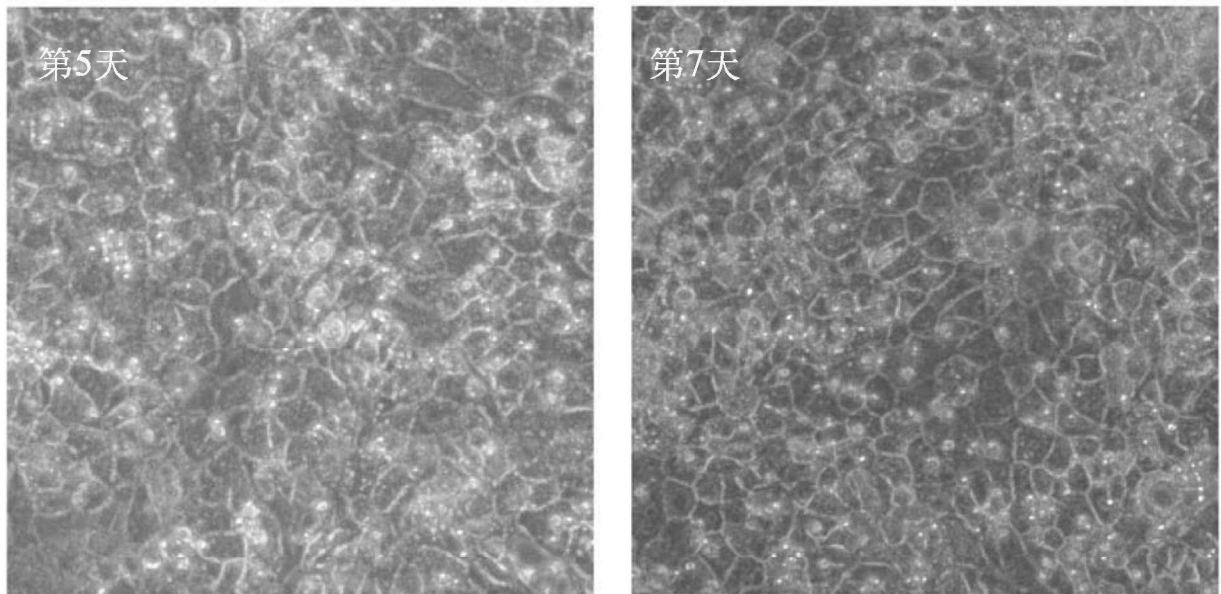


图8B

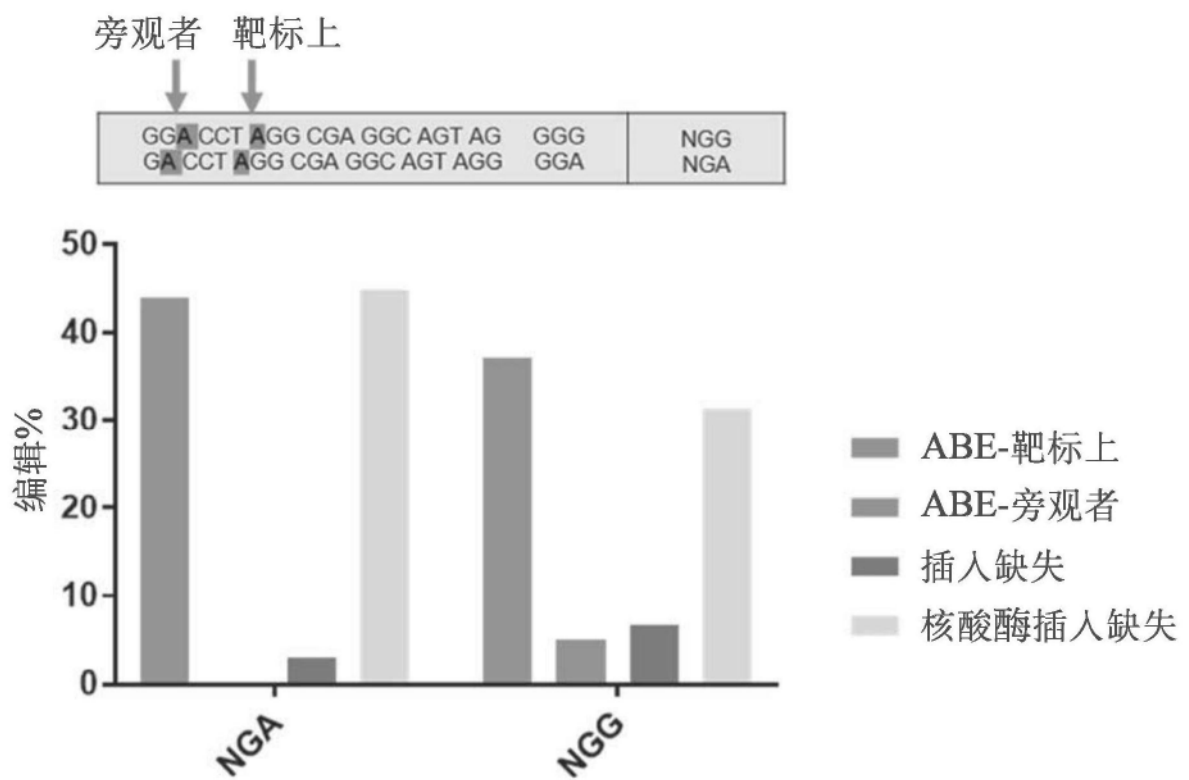


图9A

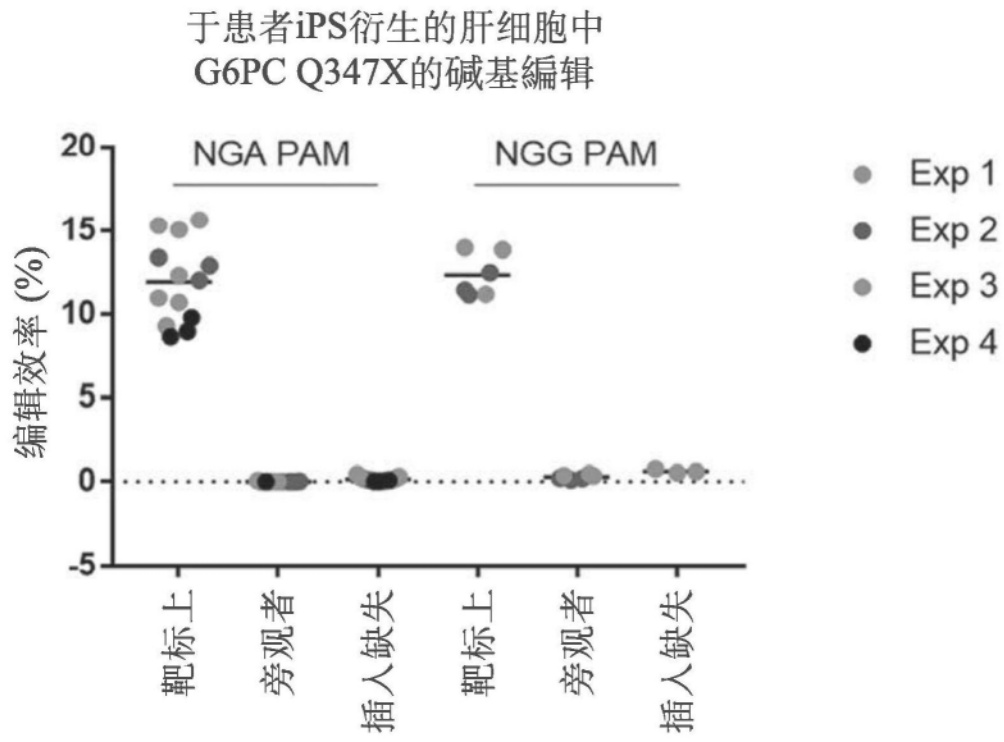


图9B

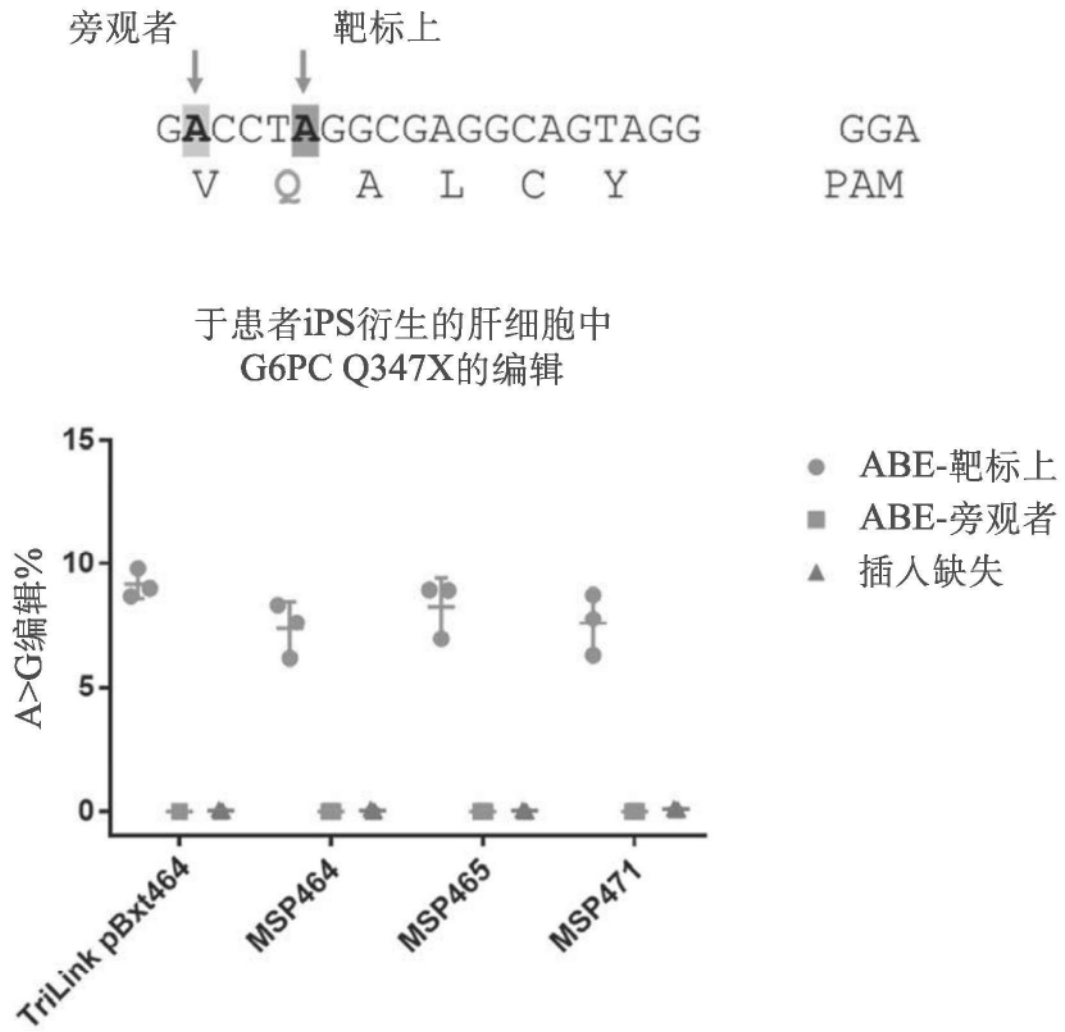


图10

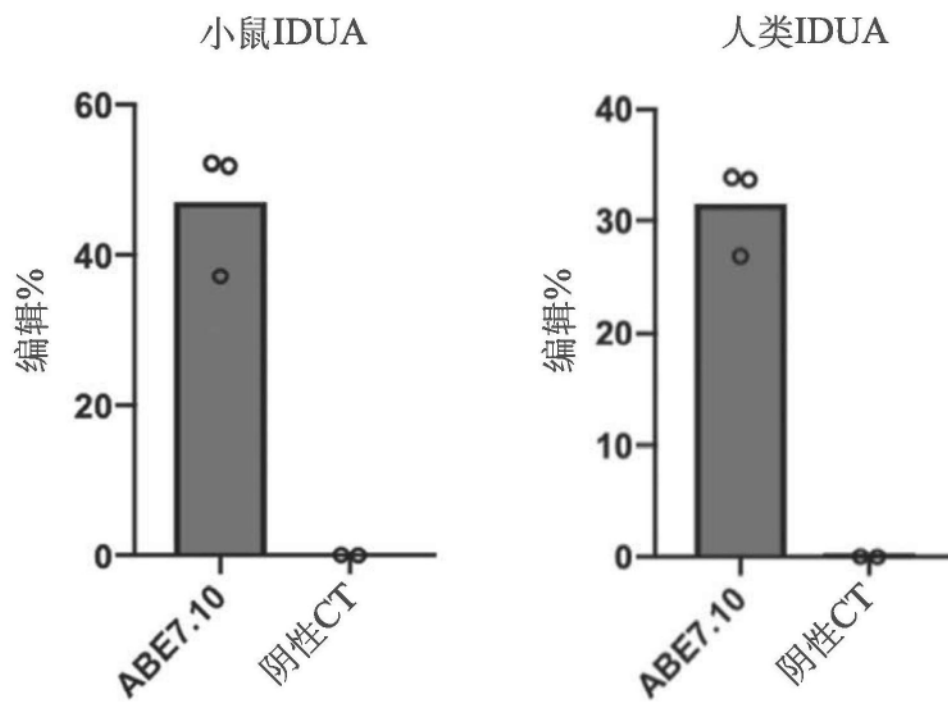


图11