(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 110799525 A
(43)申请公布日 2020.02.14

(21)申请号 201880042346.4

(22)申请日 2018.04.23

(30)优先权数据
62/488,426 2017.04.21 US
62/616,066 2018.01.11 US

(85)PCT国际申请进入国家阶段日
2019.12.19

(86)PCT国际申请的申请数据
PCT/US2018/028919 2018.04.23

(87)PCT国际申请的公布数据
WO2018/195545 EN 2018.10.25

(71)申请人 通用医疗公司
地址 美国马萨诸塞州

(72)发明人 J·K·乔昂格 B·克莱因斯蒂弗 A·苏萨

(74)专利代理机构 北京林达刘知识产权代理事务所(普通合伙) 11277
代理人 刘新宇 李茂家

(51)Int.Cl.
C07K 14/195(2006.01)
C12N 9/16(2006.01)
C12N 9/22(2006.01)
C12N 15/11(2006.01)
C12N 15/62(2006.01)
C12N 15/63(2006.01)
C12N 15/79(2006.01)
C12N 15/90(2006.01)

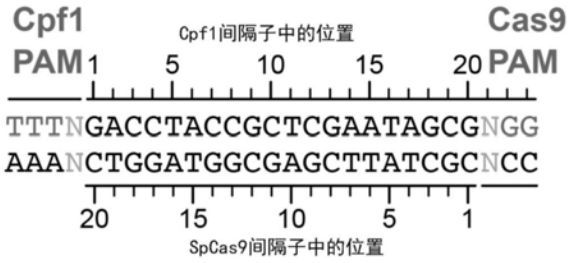
权利要求书4页 说明书108页 附图103页

(54)发明名称

具有改变的PAM特异性的CPF1(CAS12a)的变体

(57)摘要

具有改善的靶向范围和增强的中靶活性的工程化的普氏菌属和弗朗西斯氏菌属源的CRISPR 1(Cpf1)核酸酶,以及它们在基因组工程化、表观基因组工程化、碱基编辑、基因组靶向、基因组编辑和体外诊断中的用途。



CN 110799525 A

1.一种分离的、来自氨基酸球菌种(Acidaminococcus sp.)BV3L6的、普氏菌属和弗朗西斯氏菌属源的CRISPR 1(CRISPR from Prevotella and Francisella 1)(Cpf1)蛋白(AsCpf1),其包含与SEQ ID NO:2的氨基酸序列至少80%相同并在以下位置中的一处或多处具有突变的序列:SEQ ID NO:2的E174、S170、K548、N551、T167、T539、N552、M604和/或K607。

2.根据权利要求1所述的分离的Cpf1蛋白,其进一步包含S542处的突变。

3.根据权利要求2所述的分离的蛋白,其包含突变S542Q、S542K或S542R。

4.根据权利要求1所述的分离的蛋白,其包含与SEQ ID NO:2的氨基酸序列至少95%相同的序列。

5.根据权利要求1所述的分离的蛋白,其包含一个或多个以下突变:E174R、S170R、K548R、S170K、E174K、T167A、T539K、T539R、K548V、N551R、N552R、M604A、K607Q、K607R、K607S和/或K607H。

6.根据权利要求2所述的分离的蛋白,其包含以下突变:SEQ ID NO:2的S170R/E174R、E174R/K548R、S170R/K548R、E174R/S542R、S170R/S542R、E174R/S542R/K548R、E174R/N551R、S170R/N551R、S542R/K548R、S542R/N551R、S542R/N552R、K548R/N551R、S170R/S542R/K607R、E174R/S542R/K607R、E174R/S542R/K607H、E174R/S542R/K548R/N551R、S170R/S542R/K548V/N552R、E174R/S542R/K548V/N552R、S170R/S542R/K607R或E174R/S542R/K607R。

7.根据权利要求1所述的分离的蛋白,其进一步包含降低核酸酶活性的一个或多个突变,所述突变优选选自由D908、E993、R1226、D1235和/或D1263处的突变组成的组,优选选自由D908A、E993A、R1226A、D1235A和/或D1263A组成的组。

8.根据权利要求6所述的分离的蛋白,其进一步包含降低核酸酶活性的一个或多个突变,所述突变优选选自由D908、E993、R1226、D1235和/或D1263处的突变组成的组,优选选自由D908A、E993A、R1226A、D1235A和/或D1263A组成的组。

9.根据权利要求1-8任一项所述的分离的蛋白,其进一步包含N282、N178、S186、N278、R301、T315、S376、N515、K523、K524、K603、K965、Q1013、Q1014和/或K1054中的一处或多处,优选N282、T315、N515或N278处的突变,优选地其中所述突变提高所述蛋白的特异性。

10.根据权利要求9所述的分离的蛋白,其包含选自由N282A、T315A、N515A或N278A组成的组的突变。

11.一种分离的、来自毛螺旋菌科细菌ND2006的、普氏菌属和弗朗西斯氏菌属源的CRISPR 1(Cpf1)蛋白(LbCpf1),其包含与SEQ ID NO:11至少80%相同的序列,在以下位置中的一处或多处具有:SEQ ID NO:11的T152、D156、G532和/或K538。

12.根据权利要求11所述的分离的蛋白,其包含与SEQ ID NO:11的氨基酸序列至少95%相同的序列。

13.根据权利要求11所述的分离的蛋白,其包含一个或多个以下突变T152R、T152K、D156R、D156K、G532R和/或K538R。

14.根据权利要求11所述的分离的蛋白,其包含以下突变:D156R/G532R/K538R。

15.根据权利要求11所述的分离的蛋白,其进一步包含降低核酸酶活性的一个或多个突变,所述突变例如选自由D832、E925、R1138、D1148和/或D1180处的突变组成的组,优选选

自由D832A、E925A、R1138A、D1148A和/或D1180A组成的组。

16.根据权利要求14所述的分离的蛋白，其进一步包含降低核酸酶活性的一个或多个突变，所述突变例如选自由D832、E925、R1138、D1148和/或D1180处的突变组成的组，优选选自由D832A、E925A、R1138A、D1148A和/或D1180A组成的组。

17.根据权利要求11-16任一项所述的分离的蛋白，其进一步包含S202、N274、N278、K290、K367、K532、K609、K915、Q962、K963、K966、K1002和/或S1003的一处或多处的突变，优选地其中所述突变提高所述蛋白的特异性。

18.根据权利要求17所述的分离的蛋白，其包含一个或多个以下突变：S202A、N274A、N278A、K290A、K367A、K532A、K609A、K915A、Q962A、K963A、K966A、K1002A和/或S1003A。

19.一种分离的、来自土拉弗朗西斯氏菌的、普氏菌属和弗朗西斯氏菌属源的CRISPR 1(Cpf1)蛋白(FnCpf1)，其包含与SEQ ID NO:4至少80％相同并在以下位置中的一处或多处具有突变的序列：SEQ ID NO:4的K180、E184、N607、K613、D616、N617和/或K671。

20.根据权利要求19所述的分离的蛋白，其包含与SEQ ID NO:4的氨基酸序列至少95％相同的序列。

21.根据权利要求19所述的分离的蛋白，其包含一个或多个以下突变：K180R、E184R、N607R、K613R、K613V、D616R、N617R、K671H和K671R。

22.根据权利要求19所述的分离的蛋白，其包含以下突变：N607R/K613R，N607R/K613V，N607R/K613V/D616R，或N607R/K613R/D616R。

23.根据权利要求19所述的分离的蛋白，其进一步包含降低核酸酶活性的一个或多个突变，所述突变优选地选自由D917、E1006、R1218、D1227和/或D1255处的突变组成的组，优选选自由D917A、E1006A、R1218A、D1227A和/或D1255A组成的组。

24.根据权利要求20所述的分离的蛋白，其进一步包含降低核酸酶活性的一个或多个突变，所述突变优选地选自由D917、E1006、R1218、D1227和/或D1255处的突变组成的组，优选选自由D917A、E1006A、R1218A、D1227A和/或D1255A组成的组。

25.一种融合蛋白，其包含经任选的间插接头与异源功能结构域融合的根据权利要求1-24任一项所述的分离的蛋白，其中所述接头不干扰所述融合蛋白的活性。

26.根据权利要求25所述的融合蛋白，其中所述异源功能结构域是转录激活结构域。

27.根据权利要求26所述的融合蛋白，其中所述转录激活结构域是VP16、VP64、Rta、NF-κB p65或VPR融合物。

28.根据权利要求25所述的融合蛋白，其中所述异源功能结构域是转录沉默子或转录阻遏结构域。

29.根据权利要求28所述的融合蛋白，其中所述转录阻遏结构域是克鲁贝尔相关盒(KRAB)结构域、ERF阻遏蛋白结构域(ERD)或mSin3A相互作用结构域(SID)。

30.根据权利要求28所述的融合蛋白，其中所述转录沉默子是异染色质蛋白1(HP1)。

31.根据权利要求25所述的融合蛋白，其中所述异源功能结构域是修饰DNA甲基化状态的酶。

32.根据权利要求31所述的融合蛋白，其中所述修饰DNA甲基化状态的酶是DNA甲基转移酶(DNMT)或TET蛋白。

33.根据权利要求32所述的融合蛋白，其中所述TET蛋白是TET1。

34.根据权利要求25所述的融合蛋白,其中所述异源功能结构域是修饰组蛋白亚单位的酶。

35.根据权利要求34所述的融合蛋白,其中所述修饰组蛋白亚单位的酶是组蛋白乙酰基转移酶(HAT)、组蛋白脱乙酰酶(HDAC)、组蛋白甲基转移酶(HMT)、或组蛋白脱甲基酶。

36.根据权利要求35所述的融合蛋白,其中所述异源功能结构域是生物系链。

37.根据权利要求36所述的融合蛋白,其中所述生物系链是MS2、Csy4或λN蛋白。

38.根据权利要求25所述的融合蛋白,其中所述异源功能结构域是FokI。

39.根据权利要求25所述的融合蛋白,其中所述异源功能结构域是脱氨酶。

40.根据权利要求39所述的融合蛋白,其中所述异源功能结构域是胞苷脱氨酶,优选地选自由APOBEC1、APOBEC2、APOBEC3A、APOBEC3B、APOBEC3C、APOBEC3D/E、APOBEC3F、APOBEC3G、APOBEC3H、APOBEC4、活化诱导的胞苷脱氨酶(AID)、胞嘧啶脱氨酶1(CDA1)、CDA2和作用于tRNA的胞嘧啶脱氨酶(CDAT)组成的组。

41.根据权利要求39所述的融合蛋白,其中所述异源功能结构域是腺苷脱氨酶,优选地选自由腺苷脱氨酶1(ADA1)、ADA2;作用于RNA的腺苷脱氨酶1(ADAR1)、ADAR2、ADAR3;作用于tRNA的腺苷脱氨酶1(ADAT1)、ADAT2、ADAT3;和天然发生的或工程化的tRNA特异性腺苷脱氨酶(TadA)组成的组。

42.根据权利要求25所述的融合蛋白,其中所述异源功能结构域是抑制或增强内源DNA修复或碱基切除修复(BER)途径的酶、结构域或肽。

43.根据权利要求42所述的融合蛋白,其中所述异源功能结构域是抑制尿嘧啶DNA糖基化酶(UDG,也称为尿嘧啶N-糖基化酶,或UNG)的尿嘧啶DNA糖基化酶抑制剂(UGI);或来自噬菌体Mu的Gam。

44.一种分离的核酸,其编码根据权利要求1-43任一项所述的蛋白或融合蛋白。

45.一种载体,其包含根据权利要求44所述的分离的核酸。

46.根据权利要求45所述的载体,其中所述分离的核酸编码根据权利要求1所述的分离的蛋白,并且可操作地连接至一个或多个调节结构域,用于表达分离的、来自氨基酸球菌种BV3L6的、普氏菌属和弗朗西斯氏菌属源的CRISPR 1(Cpf1)蛋白(AsCpf1),所述蛋白在以下位置中的一处或多处具有突变:SEQ ID NO:2的E174、S170、K548、N551、T167、T539、N552、M604和/或K607。

47.根据权利要求45所述的载体,其中所述分离的核酸编码根据权利要求9所述的分离的蛋白,并且可操作地连接至一个或多个调节结构域,用于表达分离的、来自毛螺旋菌科细菌ND2006的、普氏菌属和弗朗西斯氏菌属源的CRISPR 1(Cpf1)蛋白(LbCpf1),所述蛋白在以下位置中的一处或多处具有突变:SEQ ID NO:11的T152、D156、G532和/或K538。

48.根据权利要求45所述的载体,其中所述分离的核酸编码根据权利要求15所述的分离的蛋白,并且可操作地连接至一个或多个调节结构域,用于表达分离的、来自土拉弗朗西斯氏菌的、普氏菌属和弗朗西斯氏菌属源的CRISPR 1(Cpf1)蛋白(FnCpf1),所述蛋白包含与SEQ ID NO:4至少80%相同并在以下位置中的一处或多处具有突变的序列:SEQ ID NO:4的K180、E184、N607、K613、D616、N617和/或K671。

49.一种宿主细胞,优选哺乳动物宿主细胞,其包含根据权利要求44所述的核酸,并且任选地表达根据权利要求1-43任一项所述的蛋白或融合蛋白。

50.一种改变细胞的基因组的方法,所述方法包括在所述细胞中表达、或使所述细胞接触:根据权利要求1-43任一项所述的分离的蛋白或融合蛋白,以及至少一种指导RNA,所述指导RNA具有与所述细胞的基因组的选定部分互补的区域。

51.根据权利要求50所述的方法,其中所述分离的蛋白或融合蛋白包含核定位序列、细胞穿透肽序列和/或亲和标签的一种或多种。

52.根据权利要求50所述的方法,其中所述细胞是干细胞。

53.根据权利要求50所述的方法,其中所述细胞是胚胎干细胞、间充质干细胞、或诱导多能干细胞;是在活的动物中的;或是在胚胎中的。

54.一种改变双链DNA(dsDNA)分子的方法,所述方法包括使所述dsDNA分子接触根据权利要求1-43任一项所述的分离的蛋白或融合蛋白,以及指导RNA,所述指导RNA具有与所述dsDNA分子的选定部分互补的区域。

55.根据权利要求54的方法,其中所述dsDNA分子是在体外的。

56.一种体外检测样品中的目标ssDNA或dsDNA的方法,所述方法包括:

使所述样品接触(I)根据权利要求1-43任一项所述的分离的蛋白或融合蛋白,(ii)具有与所述目标ssDNA或dsDNA分子的选定部分互补的区域的指导RNA,和(iii)标记的检测子DNA,其中所述分离的蛋白或融合蛋白在结合所述目标ssDNA或dsDNA时切割所述检测子DNA;和

测量由所述标记的检测子DNA的切割产生的可检测信号,从而检测所述目标ssDNA或dsDNA。

57.根据权利要求56所述的方法,其中测量所述标记的检测子DNA包括检测金纳米颗粒、荧光团、荧光偏振、胶体相转变/分散、电化学信号和基于半导体的信号中的一种或多种。

58.根据权利要求56所述的方法,其中所述标记的检测子DNA在被切割之前产生一定量的可检测信号,当所述标记的检测子DNA被切割时所述可检测信号的量降低。

59.根据权利要求56所述的方法,其中所述标记的检测子DNA在被切割之前产生第一可检测信号,当所述标记的检测子DNA被切割时产生第二可检测信号。

60.根据权利要求56所述的方法,其中所述标记的检测子DNA包含淬灭剂/fluor对或FRET对。

具有改变的PAM特异性的CPF1(CAS12a)的变体

[0001] 优先权声明

[0002] 本申请要求2017年4月21日提交的美国专利申请系列号No.62/488,426,以及2018年1月11日提交的美国专利申请系列号No.62/616,066的权益。上述的全部内容通过引用并入本文。

[0003] 联邦政府赞助的研究或开发

[0004] 本发明是在美国国立卫生研究院授予的拨款号No.GM105378和GM118158的政府支持下完成的。美国政府在本发明中享有一定的权利。

技术领域

[0005] 本发明至少部分地涉及具有改变和改进的目标特异性的、工程化的普氏菌属(Prevotella)和弗朗西斯氏菌属(Francisella)CRISPR 1(CRISPR from Prevotella and Francisella 1,Cpf1)核酸酶,以及它们在基因组工程、表观基因组工程、基因组靶向、基因组编辑和体外诊断中的用途。

背景技术

[0006] CRISPR-Cas Cpf1核酸酶(也称为Cas12a核酸酶)最近被描述为基因组编辑应用中作为对Cas9核酸酶的替代(Zetsche等人.Cell 163,759-771(2015);Shmakov等人,Mol Cell.2015 Nov 5;60(3):385-97;Kleinstiver等人,Nat Biotechnol.2016 Aug;34(8):869-74;Kim等人,Nat Biotechnol.2016 Aug;34(8):863-8)。Cpf1核酸酶拥有许多潜在有益的性质,包括但不限于:识别富T前间区序列邻近基序(T-rich protospacer-adjacent motif,PAM)序列,与野生型化脓性链球菌(Streptococcus pyogenes)Cas9(SpCas9)相比在人类细胞中相对更大的基因组广度的特异性,简化了多个位点的同时靶向(多重化)的加工pre-crRNA的核糖核酸内切酶活性,产生5'DNA突出端的DNA核酸内切酶活性(而不是SpCas9观察到的平端的双链断裂),以及在距离PAM的最远端切割前间区序列的DNA序列(与用SpCas9和SaCas9观察到的在前间区序列的PAM近端切割相比)。迄今为止,来自氨基酸球菌种(Acidaminococcus sp.)BV3L6、毛螺旋菌科细菌(Lachnospiraceae bacterium)ND2006和土拉弗朗西斯氏菌新杀手亚种(Francisella tularensis subsp.novicida)U112的Cpf1直系同源基因(分别为AsCpf1、LbCpf1和FnCpf1)是已被描述为在人类细胞中强健地起作用的仅有的直系同源基因。虽然有这些能力,与SpCas9相比,Cpf1核酸酶没有那么快地用于基因组编辑。一种可能的原因是需要更长的PAM,与SpCas9的每8bp一次相比,其将靶向限制到大约每43bp的随机DNA序列中一次。

[0007] 在此我们通过利用结构导向的工程化方法来产生不仅具有大幅扩展的靶向范围、还具有实质上改进的中靶活性的AsCpf1变体解决了这种靶向范围限制。除了作为用于基因组编辑的核酸酶的改进的效力和多面性之外,我们证实了这些变体还可以用于其他应用,包括多重核酸酶靶向、表观基因组编辑,C-T碱基编辑和Cpf1介导的DNA检测,都是野生型AsCpf1以前不可能达到的水平。

发明内容

[0008] 本文描述的是一系列AsCpf1、FnCpf1和LbCpf1变体，与它们的野生型对应物相比识别更广范围的PAM，从而提高了可以通过这类RNA-指导核酸酶靶向的位点的范围。此外，在识别和修饰带有典型的TTTN PAM的目标位点方面，这些变体比野生型Cpf1核酸酶表现更好。本文描述的变体的增强的活性改善了AsCpf1用于基因组编辑、表观基因组编辑、碱基编辑和体外DNA检测的活性。

[0009] 因而，本文提供的是分离的、来自氨基酸球菌种BV3L6的、普氏菌属和弗朗西斯氏菌属源的CRISPR 1(Cpf1)蛋白(AsCpf1)，其包含与SEQ ID NO:2的氨基酸序列至少80％相同并在以下位置中的一处或多处具有突变的序列：SEQ ID NO:2的E174、S170、K548、N551、T167、T539、N552、M604和/或K607。

[0010] 在某些实施方案中，所述分离的Cpf1蛋白包含E174R处的突变，任选地在S170R、K548、N551、T167、T539、S542、N552、M604和/或K607处具有一个或多个额外的突变。在某些实施方案中，所述分离的Cpf1蛋白包括S170R的突变，任选地在E174R、K548、N551、T167、T539、S542、N552、M604和/或K607处具有一个或多个额外的突变。

[0011] 在某些实施方案中，所述分离的Cpf1蛋白包括S542处的突变。

[0012] 在某些实施方案中，所述分离的Cpf1蛋白包括突变S542Q、S542K或S542R。

[0013] 在某些实施方案中，所述分离的Cpf1蛋白包括与SEQ ID NO:2的氨基酸序列至少95％相同的序列。

[0014] 在某些实施方案中，所述分离的Cpf1蛋白包括一个或多个以下突变：E174R、S170R、K548R、S170K、E174K、T167A、T539K、T539R、K548V、N551R、N552R、M604A、K607Q、K607R、K607S和/或K607H。

[0015] 在某些实施方案中，所述分离的Cpf1蛋白包括以下突变：SEQ ID NO:2的S170R/E174R、E174R/K548R、S170R/K548R、E174R/S542R、S170R/S542R、E174R/S542R/K548R、E174R/N551R、S170R/N551R、S542R/K548R、S542R/N551R、S542R/N552R、K548R/N551R、S170R/S542R/K607R、E174R/S542R/K607R、E174R/S542R/K607H、E174R/S542R/K548R/N551R、S170R/S542R/K548V/N552R、E174R/S542R/K548V/N552R、S170R/S542R/K607R或E174R/S542R/K607R。

[0016] 在某些实施方案中，所述分离的Cpf1蛋白包含降低核酸酶活性的一个或多个突变，例如选自由D908、E993、R1226、D1235和/或D1263处的突变组成的组，优选选自由D908A、E993A、R1226A、D1235A和/或D1263A组成的组。

[0017] 在某些实施方案中，所述分离的Cpf1蛋白包含在N282、N178、S186、N278、R301、T315、S376、N515、K523、K524、K603、K965、Q1013、Q1014和/或K1054中的一处或多处，优选N282、T315、N515或N278处的突变，优选地其中所述突变提高所述蛋白的特异性。在某些实施方案中，所述分离的Cpf1蛋白包括选自由N282A、T315A、N515A或N278A组成的组的突变。

[0018] 本文还提供的是分离的、来自毛螺旋菌科细菌ND2006的、普氏菌属和弗朗西斯氏菌属源的CRISPR 1(Cpf1)蛋白(LbCpf1)，其包含与SEQ ID NO:11至少80％相同的序列，在以下位置中的一处或多处具有：SEQ ID NO:11的T152、D156、G532和/或K538。

[0019] 在某些实施方案中，所述分离的Cpf1蛋白包括与SEQ ID NO:11的氨基酸序列至少95％相同的序列。

[0020] 在某些实施方案中，所述分离的Cpf1蛋白包括一个或多个以下突变：T152R、T152K、D156R、D156K、G532R和/或K538R。

[0021] 在某些实施方案中，所述分离的Cpf1蛋白包括以下突变：D156R/G532R/K538R。

[0022] 在某些实施方案中，所述分离的Cpf1蛋白包含降低核酸酶活性的一个或多个突变，所述突变例如选自由D832、E925、R1138、D1148和/或D1180处的突变组成的组，优选地选自由D832A、E925A、R1138A、D1148A和/或D1180A组成的组。

[0023] 在某些实施方案中，所述分离的Cpf1蛋白包含S202、N274、N278、K290、K367、K532、K609、K915、Q962、K963、K966、K1002和/或S1003中的一处或多处的突变，优选地其中所述突变提高所述蛋白的特异性。在某些实施方案中，所述分离的Cpf1蛋白包括一个或多个以下突变：S202A、N274A、N278A、K290A、K367A、K532A、K609A、K915A、Q962A、K963A、K966A、K1002A和/或S1003A。

[0024] 本文还提供的是分离的、来自土拉弗朗西斯氏菌的、普氏菌属和弗朗西斯氏菌属源的CRISPR 1(Cpf1)蛋白(FnCpf1)，其包含与SEQ ID NO:4至少80％相同并在以下位置中的一处或多处具有突变的序列：SEQ ID NO:4的K180、E184、N607、K613、D616、N617和/或K671。

[0025] 在某些实施方案中，所述分离的Cpf1蛋白包括与SEQ ID NO:4的氨基酸序列至少95％相同的序列。

[0026] 在某些实施方案中，所述分离的Cpf1蛋白包括一个或多个以下突变：K180R、E184R、N607R、K613R、K613V、D616R、N617R、K671H和K671R。

[0027] 在某些实施方案中，所述分离的Cpf1蛋白包括以下突变：N607R/K613R、N607R/K613V、N607R/K613V/D616R或N607R/K613R/D616R。

[0028] 在某些实施方案中，所述分离的Cpf1蛋白包含降低核酸酶活性的一个或多个突变，所述突变例如选自由D917、E1006、R1218、D1227和/或D1255处的突变组成的组，优选地选自由D917A、E1006A、R1218A、D1227A和/或D1255A组成的组。

[0029] 本文还提供的融合蛋白，其包含经任选的间插接头与异源功能结构域融合的本文描述的Cpf1蛋白，其中所述接头不干扰所述融合蛋白的活性。

[0030] 在某些实施方案中，所述异源功能结构域是转录激活结构域，例如，称为VP64的四聚VP16融合物、Rta、NF-κB p65或VPR(一种VP64、p65、Rta融合蛋白)。

[0031] 在某些实施方案中，所述异源功能结构域是转录沉默子或转录阻遏结构域。在某些实施方案中，所述转录阻遏结构域是克鲁贝尔相关盒(Krueppel-associated box,KRAB)结构域、ERF阻遏蛋白结构域(ERD)或mSin3A相互作用结构域(SID)。在某些实施方案中，所述转录沉默子是异染色质蛋白1(Heterochromatin Protein 1,HP1)。

[0032] 在某些实施方案中，所述异源功能结构域是修饰DNA甲基化状态的酶，例如，DNA甲基转移酶(DNMT)或TET蛋白，例如，TET1。

[0033] 在某些实施方案中，所述异源功能结构域是修饰组蛋白亚单位的酶，例如，组蛋白乙酰基转移酶(HAT)、组蛋白脱乙酰酶(HDAC)、组蛋白甲基转移酶(HMT)、或组蛋白脱甲基酶。

[0034] 在某些实施方案中，所述异源功能结构域是修饰胞嘧啶DNA碱基的脱氨酶，例如，来自载脂蛋白B mRNA编辑酶的胞苷脱氨酶，脱氨酶的催化多肽样(APOBEC)家族，包括

APOBEC1、APOBEC2、APOBEC3A、APOBEC3B、APOBEC3C、APOBEC3D/E、APOBEC3F、APOBEC3G、APOBEC3H、APOBEC4，活化诱导的胞苷脱氨酶(AID)、胞嘧啶脱氨酶1(CDA1)和CDA2，以及作用于tRNA的胞嘧啶脱氨酶(CDAT)。

[0035] 在某些实施方案中，所述异源功能结构域是修饰腺苷DNA碱基的脱氨酶，例如，所述脱氨酶是腺苷脱氨酶1(ADA1)、ADA2；作用于RNA的腺苷脱氨酶1(ADAR1)、ADAR2、ADAR3；作用于tRNA的腺苷脱氨酶1(ADAT1)、ADAT2、ADAT3；和天然发生的或工程化的tRNA特异性腺苷脱氨酶(TadA)。

[0036] 在某些实施方案中，所述异源功能结构域是抑制或增强内源DNA修复或碱基切除修复(BER)途径的酶、结构域或肽，例如，尿嘧啶DNA糖基化酶抑制剂(UGI)，其抑制尿嘧啶DNA糖基化酶(UDG，也称为尿嘧啶N-糖基化酶，或UNG)介导的尿嘧啶切除以启动BER；或DNA末端结合蛋白，例如，来自噬菌体Mu的Gam。

[0037] 在某些实施方案中，所述异源功能结构域是生物系链(biological tether)，例如，MS2、Csy4或λN蛋白。

[0038] 在某些实施方案中，所述异源功能结构域是FokI。

[0039] 本文还提供的是编码本文描述的Cpf1变体蛋白和融合蛋白的分离的核酸。

[0040] 此外，本文提供的是包含所述核酸的载体。在某些实施方案中，所述分离的核酸编码分离的、来自氨基酸球菌种BV3L6的、普氏菌属和弗朗西斯氏菌属源的CRISPR 1(Cpf1)蛋白(AsCpf1)，所述蛋白在以下位置中的一处或多处具有突变：SEQ ID NO:2的T167、S170、E174、T539、K548、N551、N552、M604和/或K607，并且与一个或多个调节结构域可操作地连接，用于表达分离的、来自氨基酸球菌种BV3L6的、普氏菌属和弗朗西斯氏菌属源的CRISPR1(Cpf1)蛋白(AsCpf1)，所述蛋白在以下位置中的一处或多处具有突变：SEQ ID NO:2的T167、S170、E174、T539、K548、N551、N552、M604和/或K607。

[0041] 在某些实施方案中，所述分离的核酸编码分离的、来自毛螺旋菌科细菌ND2006的、普氏菌属和弗朗西斯氏菌属源的CRISPR 1(Cpf1)蛋白，所述蛋白在以下位置中的一处或多处具有突变：SEQ ID NO:11的T152、D156、G532和/或K538，并且与一个或多个调节结构域可操作地连接，用于表达分离的、来自毛螺旋菌科细菌ND2006的、普氏菌属和弗朗西斯氏菌属源的CRISPR 1(Cpf1)蛋白(LbCpf1)，所述蛋白在以下位置中的一处或多处具有突变：SEQ ID NO:11的T152、D156、G532和/或K538。

[0042] 在某些实施方案中，所述分离的核酸编码分离的、来自土拉弗朗西斯氏菌的、普氏菌属和弗朗西斯氏菌属源的CRISPR 1(Cpf1)蛋白(FnCpf1)，所述蛋白包含与SEQ ID NO:4至少80％相同并在以下位置中的一处或多处具有突变的序列：SEQ ID NO:4的K180、E184、N607、K613、D616、N617和/或K671，并且与一个或多个调节结构域可操作地连接，用于表达分离的、来自土拉弗朗西斯氏菌的、普氏菌属和弗朗西斯氏菌属源的CRISPR 1(Cpf1)蛋白(FnCpf1)，所述蛋白包含与SEQ ID NO:4至少80％相同并在以下位置中的一处或多处具有突变的序列：SEQ ID NO:4的K180、E184、N607、K613、D616、N617和/或K671。

[0043] 本文还提供的是宿主细胞，优选哺乳动物宿主细胞，其包含本文描述的所述核酸，并且任选地表达本文描述的蛋白或融合蛋白。

[0044] 此外，本文提供的是改变细胞的基因组的方法，所述方法包括在所述细胞中表达、或使所述细胞接触：本文描述的分离的蛋白或融合蛋白以及至少一种指导RNA，所述指导

RNA具有与所述细胞的基因组的选定部分互补的区域,即,其中所述互补的区域邻近于与所述蛋白或融合蛋白相应的PAM序列,例如,如表B中所示。在某些实施方案中,所述分离的蛋白或融合蛋白包含一个或多个核定位序列、细胞穿透肽序列和/或亲和标签。在某些实施方案中,所述细胞是干细胞。在某些实施方案中,所述细胞是胚胎干细胞、间充质干细胞、或诱导多能干细胞;是在活的动物中的;或是在胚胎中的。

[0045] 还提供的是改变双链DNA(dsDNA)分子的方法,所述方法包括使所述dsDNA分子接触本文描述的分离的蛋白或融合蛋白以及指导RNA,所述指导RNA具有与所述dsDNA分子的选定部分互补的区域,即,其中所述互补的区域邻近于与所述蛋白或融合蛋白相应的PAM序列,例如,如表B中所示。在某些实施方案中,所述dsDNA分子是在体外的。

[0046] 还提供的是在体外检测样品中的目标ssDNA或dsDNA的方法。所述方法包括使所述样品接触(i)权利要求1-43的分离的蛋白或融合蛋白,(ii)具有与所述目标ssDNA或dsDNA分子的选定部分互补的区域的指导RNA;和(iii)标记的检测子DNA,其中所述分离的蛋白或融合蛋白在结合所述目标ssDNA或dsDNA时切割所述检测子DNA;并且测量由所述标记的检测子DNA的切割产生的可检测信号,从而检测所述目标ssDNA或dsDNA。

[0047] 在某些实施方案中,测量所述标记的检测子DNA包括检测金纳米颗粒、荧光团、荧光偏振、胶体相转变/分散、电化学信号和基于半导体的信号中的一种或多种。

[0048] 在某些实施方案中,所述标记的检测子DNA在被切割之前产生一定量的可检测信号,当所述标记的检测子DNA被切割时所述可检测信号的量降低。在某些实施方案中,所述标记的检测子DNA在被切割之前产生第一可检测信号,当所述标记的检测子DNA被切割时产生第二可检测信号。

[0049] 在某些实施方案中,所述标记的检测子DNA包含淬灭剂/fluor对(pair)或FRET对。

[0050] 除非另外定义,本文使用的所有技术和科学术语具有本发明所属技术领域的普通技术人员通常所理解的相同含义。本文描述用于本发明的方法和材料;也可以使用本领域已知的其他适合的方法和材料。所述材料、方法和实施例仅是说明性的,而不意图是限制性的。本文提及的所有公开物、专利申请、专利、序列、数据库条目和其他参考文献以其全部内容并入以作参考。在冲突的情况下,以本说明书、包括定义为准。

[0051] 根据以下的详细说明和附图以及根据权利要求,本发明的其他特征和益处将是明显的。

附图说明

[0052] 图1:Cpf1和SpCas9核酸酶的代表性的目标位点。AsCpf1和LbCpf1识别目标位点间隔子的5'末端的TTTN PAM序列,而SpCas9识别间隔子的3'末端的NGG PAM。有义序列:SEQ ID NO:1。

[0053] 图2:野生型AsCpf1和LbCpf1核酸酶的PAM识别分布。使用靶向具有典型TTTN PAM的位点、或具有携带单碱基差异的非典型PAM的位点的crRNA,在基于人类细胞的EGFP破坏测定中测定了野生型AsCpf1和LbCpf1识别目标位点的能力。由于缺乏EGFP报告基因中的目标位点,除了ATTN和TTAN之外,检查了每种非典型PAM的三个位点。

[0054] 图3:野生型AsCpf1对于具有典型PAM和携带两个和三个碱基差异的非典型PAM的位点的PAM识别分布。使用靶向具有指定PAM的位点的crRNA,在基于人类细胞的EGFP破坏测

定中测定了AsCpf1识别和破坏目标位点的能力。由于缺乏EGFP报告基因中的目标位点,除了ATTN和TTAN之外,在可能的情况下,检查了每种非典型PAM的三个位点。误差条表示s.e.m.,n=3,否则n=1。

[0055] 图4:AsCpf1和LbCpf1对典型和非典型(单碱基差异)PAM位点的PAM识别轮廓。使用靶向具有典型TTTN PAM或具有携带单碱基差异的非典型PAM的位点的crRNA,在U2OS人类细胞中测定了野生型AsCpf1和LbCpf1识别和破坏内源目标位点的能力。在可能的情况下,检查了每种非典型PAM的三个位点。误差条表示s.e.m.,n=2或3,否则n=1。

[0056] 图5A-5H:通过物理上靠近PAM DNA碱基的残基的合理突变(rational mutation),改变AsCpf1 PAM识别特异性。使用人类细胞EGFP破坏测定,或通过定量人类U2OS细胞中内源基因位点处诱导的插入或删除突变(indel)频率,使用靶向携带典型TTTN或非典型PAM位点的位点的crRNA,评估了野生型AsCpf1和携带单突变或双突变的变体的活性。野生型AsCpf1与以下比较:(A)T167/T539变体,使用靶向EGFP中的位点的crRNA;(B)S170和E174变体,使用靶向EGFP中的位点的crRNA;(C和D)S542变体,使用靶向EGFP中的位点(子图C)或内源人类基因位点(子图D)的crRNA;(E)N551和N552变体,使用靶向EGFP中的位点的crRNA;(F)K607变体,使用靶向EGFP中的位点的crRNA;(G和H)S542/K607变体,使用靶向EGFP中的位点(子图G)和内源人类基因位点(子图H)的crRNA。

[0057] 图6A-6B:野生型AsCpf1和AsCpf1-S542R变体的PAM识别分布。使用靶向携带典型TTTN PAM、或具有单碱基差异(子图A)、双碱基或三碱基差异(子图B)的非典型PAM的位点的crRNA,在基于人类细胞的EGFP破坏测定中测定了这些核酸酶的活性。由于EGFP报告基因中缺乏足够目标位点,除了ATTN和TTAN之外,在可能的情况下,检查了每种非典型PAM的三个位点。n=1。

[0058] 图7A-7B:野生型AsCpf1和AsCpf1-S542R变体对内源人类基因目标位点的PAM识别分布。使用靶向携带典型TTTN PAM、或具有单碱基差异(子图A)、双碱基或三碱基差异(子图B)的非典型PAM的位点的crRNA,在人类U2OS细胞中评估了核酸酶诱变内源基因目标位点的能力。在可能的情况下,检查了每种非典型PAM的三个位点。

[0059] 图8A-8C:在位置靠近PAM DNA碱基的残基处携带额外突变的、合理设计的AsCpf1变体的PAM识别分布。在子图A、B和C中显示的独立实验中,使用基于人类细胞的EGFP报告子分析试验了单氨基酸取代、以及双氨基酸取代或三氨基酸取代的组合,来评估它们识别携带典型TTTN或非典型PAM位点的目标位点的能力。

[0060] 图9A-9B:野生型AsCpf1和AsCpf1变体的PAM。(A)野生型AsCpf1的活性与携带单个S542R取代或E174R、S542R和K548R取代组合的变体的活性相比较。使用靶向携带典型TTTN PAM或具有单碱基差异的PAM的位点的crRNA,使用基于人类细胞的EGFP破坏测定试验了这些核酸酶的活性。(B)还使用靶向具有典型TTTN PAM或带有双碱基或三碱基差异的PAM的位点的crRNA,使用基于人类细胞的EGFP报告子分析,将野生型AsCpf1的活性与E174R/S542R/K548R变体进行比较。在可能的情况下,检查了每种非典型PAM的三个位点。误差条表示s.e.m.,n=2或3,否则n=1。

[0061] 图10A-10B:野生型LbCpf1和AsCpf1变体的PAM识别分布的比较。野生型LbCpf1修饰内源人类基因目标位点的能力与AsCpf1 E174R/S542R/K548R变体进行比较。这个实验使用的crRNA靶向携带典型TTTN PAM和具有单碱基差异的PAM(子图A)、或具有双碱基或三碱

基差异的PAM(子图B)的位点。

[0062] 图11A-11C:野生型AsCpf1和各种AsCpf1变体的PAM识别分布。使用靶向携带典型TTTN PAM、或具有单碱基、双碱基或三碱基差异的PAM的位点的crRNA,将野生型AsCpf1的活性与AsCpf1 PAM变体的活性进行比较。(A)使用基于人类细胞的EGFP破坏测定、或(B)通过评估内源人类基因目标位点处的突变频率(通过T7EI核酸内切酶分析来判断)进行比较。(C)使用靶向携带典型TTTN PAM或携带单碱基、双碱基或三碱基差异的PAM的内源人类基因位点的crRNA,比较了两种AsCpf1 PAM变体的突变活性。

[0063] 图12A-12B:野生型LbCpf1和工程化的LbCpf1变体的PAM识别分布。(A)使用靶向携带典型TTTN PAM或具有单碱基差异的PAM的位点的crRNA,使用基于人类细胞的EGFP破坏测定,野生型LbCpf1和携带各种单氨基酸取代的变体的活性。(B)还使用靶向携带典型TTTN PAM或具有单碱基差异的PAM的位点的crRNA,使用基于人类细胞的EGFP破坏测定,将野生型LbCpf1的活性与LbCpf1-D156R/G532R/K538R变体进行比较。n=1。

[0064] 图13A-13C:野生型FnCpf1和工程化的LbCpf1变体的PAM识别分布。(A)使用靶向携带典型TTN PAM或具有单碱基差异的PAM的位点的crRNA,使用基于人类细胞的EGFP破坏测定,野生型FnCpf1的活性。(B)使用靶向携带TTTN PAM的位点的crRNA,野生型FnCpf1、AsCpf1和LbCpf1针对内源人类细胞目标位点的活性。(C)使用靶向携带典型TTTN PAM或具有单碱基差异的PAM的位点的crRNA,使用基于人类细胞的EGFP破坏测定,比较野生型FnCpf1与工程化的FnCpf1 PAM变体的活性。n=1。

[0065] 图14A-14C:本申请中描述的AsCpf1变体与其他工作中公开的不同的AsCpf1变体的PAM识别分布的比较。(A,B)使用靶向携带典型TTTN PAM或具有单碱基或双碱基差异的PAM的位点的crRNA,使用基于人类细胞的EGFP破坏测定,将许多我们的工程化AsCpf1 PAM识别变体与S542R/K548V/N552R(子图A)和S542R/K607R(子图B)PAM识别变体的活性进行比较。(C)使用靶向具有典型TTTN PAM、或有单碱基、双碱基或三碱基差异的PAM的位点的crRNA,通过检测这些核酸酶诱变内源人类基因位点的能力,进行了我们的AsCpf1变体与S542R/K607R变体的额外比较。

[0066] 图15A-15G:具有扩展的目标范围的AsCas12a变体的工程化和表征。(A),携带氨基酸取代的AsCas12a变体对人类细胞中内源位点的修饰。通过T7E1分析评估活性;显示平均值、s.e.m.和单独的数据点,n≥3。(B),通过PAM测定分析(PAMDA)评估的,野生型AsCas12a和E174R/S542R/K548R变体的PAM偏好分布。$\log_{10}$速率常数(k)是四个重复的平均值,每两个针对两种不同的间隔子序列(参见图21D)。(C),E174R/S542R和E174R/S542R/K548R AsCas12a对非典型PAM的平均活性图,其中黑色线条代表每种PAM分类的12～20个位点(点)的平均值(也参见图23A、23B和23D)。(D),野生型、E174R/S542R和E174R/S542R/K548R AsCas12a对于20个编码非典型PAM的位点(针对VTTN、TTCN和TATN分类的各PAM)的活性概述(也参见图19A、23A和23B;所有位点编号为'1')。(E),AsCas12a、E174R/S542R变体和eAsCas12a对TTTN PAM的平均活性图,其中黑色线条代表每种PAM分类的5～8个位点(点)的平均值(参见图23G)。(F),eAsCas12a(E174R/S542R/K548R)的各种可靶向和不可靶向的PAM的人类细胞活性和PAMDA速率常数(k)的概述的叠加图。显示了通过T7E1分析确定的人类细胞活性的箱须图。层级(tier)1PAM展现了人类细胞中大于20%的平均靶向并且PAMDA k大于0.01,而满足细胞中大于10%平均靶向的最低阈值、PAMDA k大于0.005的PAM被认为是层

级2PAM。(G),与野生型AsCas12a相比,AsCas12a变体允许的靶向范围方面的改善的计算,通过在指定的序列特征内对完整PAM序列进行枚举并对元素大小进行归一化来确定。TSS,转录起始位点;每个AsCas12a变体可靶向的PAM序列是:野生型,TTTV;eAsCas12a,参见针对PAM层级的子图F和扩展数据图23I;RVR,TATV;RR,TYCV。

[0067] 图16A-16E:AsCas12a变体的增强的活性。(A),Cas12a直系同源体和变体对编码PAMDA位点1靶点的线性化质粒底物的时程体外切割反应的定量,在37℃、32℃和25℃下进行(分别为左、中和右图)。使用一相指数衰减方程拟合曲线;误差条代表s.e.m,n=3。(B-D),野生型和变体AsCas12a核酸酶对于编码TTTN PAM(子图B)、TATN PAM(子图C)和TYCN PAM(子图D)的位点的活性概述(也分别参见图24A-C)。(E),每种NNNN PAM的PAMDA测定的速率常数的散点图,以比较AsCas12a变体的PAM偏好(RVR对比eRVR,左图;RR对比eRR,右图)。编码以下取代的变体:eAsCas12a,E174R/S542R/K548R;RVR,S542R/K548V/N552R;eRVR,E174R/S542R/K548V/N552R;RR,S542R/K607R;eRR,E174R/S542R/K607R。

[0068] 图17A-17H:eAsCas12a特异性的表征和改善。(A),AsCas12a、eAsCas12a和eAsCas12a-HF1的GUIDE-seq基因组广度特异性分布,每一个与靶向具有TTTV PAM的位点的crRNA配对。脱靶位点中的错配位置是高亮的;还在序列的右侧显示了GUIDE-Seq读取结果计数;黄色圆圈表示仅被不对称的GUIDE-Seq读取结果支持的脱靶位点;绿色圆圈表示早先对LbCas12a鉴定的脱靶位点(Kleinstiver等人,Nat Biotechnol.,2016,34:869-74);平均PAMDA ks>0.005的eAsCas12a的非典型PAM中可选的核苷酸没有作为错配进行标色或高亮。SEQ ID NO:449-477为按照出现顺序。(B),子图A中检查的位点中,AsCas12a变体的GUIDE-seq检测的脱靶位点数量的直方图。(C),每个NNNN PAM的PAMDA测定的速率常数的散点图,以比较eAsCas12a和eAsCas12a-HF1的PAM偏好。(D),eAsCas12a和eAsCas12a-HF1对于靶向具有非典型PAM的位点的crRNA的GUIDE-Seq基因组广度特异性分布。对子图a描述了图例;eAsCas12a-HF1没有对CTTA-1、CTTC-2或TATC-1进行评估。SEQ ID NO:478-530为按照出现顺序。(E),在子图a中检测的位点中,eAsCas12a和eAsCas12a-HF1的GUIDE-seq检测的脱靶位点数量的直方图;na,未评估。(F),通过针对中靶位点观察到的计数将脱靶GUIDE-Seq读取结果计数归一化来计算的脱靶效率比。(G,H),野生型、eAsCas12a和eAsCas12a-HF1对于编码TTTN PAM(子图G)或非典型PAM(子图H)的位点的中靶活性概述(也分别参见图25I和25J)。

[0069] 图18A-18K:eAsCas12a对于多重靶向、基因激活和碱基编辑的应用。(A-C),AsCas12a、eAsCas12a和LbCas12a的多重中靶修饰效率的比较,当使用TTTV PAM靶向的crRNA阵列进行编程时,所述crRNA阵列编码3个独立的crRNA,从聚合酶III启动子(U6,子图A和B)或聚合酶II启动子(CAG,子图C)表达。使用相同的基因组DNA样品,通过T7E1分析评估了三个独立的基因座处的活性;显示平均值、s.e.m.和单独的数据点,n=3。(D),当使用集中的crRNA质粒、或表达靶向附近的(~100bp)基因组的位点(genomic loci)的两个crRNA的多重crRNA阵列时,编辑效率的评估。通过T7E1分析评估活性;显示平均值、s.e.m.和单独的数据点,n=4。(E-G),使用靶向典型PAM位点(子图E)和非典型PAM位点(子图F和G)的三种crRNA的库,使用dCas12a-VPR(1.1)融合物对内源人类基因的激活(参见图26A)。通过RT-qPCR评估的活性和RNA的倍数变化相对于HPRT1水平进行归一化;显示了三个生物学重复的三个技术重复的平均值、s.e.m.和单独的数据点(n=9)。(H),具有不同的NLS和接头组成的

dCas12a碱基编辑(BE)构建体的示意图。(I),通过靶向深度测序评估的,对于八个不同目标位点的dCas12a-BEs所指导的胞嘧啶到胸腺嘧啶(C到T)转换效率。在-5～+25窗口内检查了三个生物学重复的C到T编辑的平均百分比;对于每个目标位点,这个窗口中所有的C用绿色突出显示;目标位点内C的位置在热图底部标出。SEQ ID NO.531-538为按照出现的顺序。(J),使用dCas12a-BEs,对于所有八个目标位点在20nt目标位点间隔子序列内C到T编辑效率的集合概述。(K),对于每个dCas12a-BE实验(八个目标位点和三个重复),含有插入或删除突变(indel)的测序读取结果的百分比倍数变化的概述,相对于对照样品(LbBE1.4和空的U6质粒)中观察到的插入或删除突变百分比进行归一化。VPR,合成的VP64-p65-Rta激活结构域(Chavez等人,Nat Methods.,2015,12:326-8);NLS(sv),SV40核定位信号;NLS(nuc),核质蛋白的核定位信号;rAPO1,大鼠APOBEC1;gs,甘氨酸-丝氨酸肽接头;UGI,尿嘧啶糖基化酶抑制剂。

[0070] 图19A-19B:人类细胞中Cas12a直系同源体的活性。(A),靶向人类细胞中携带TTTN或VTTN PAM的内源位点的Cas12a直系同源体的活性。修饰百分比通过T7E1分析评估;显示平均值、s.e.m.和单独的数据点,n＝3。(B),Cas12a直系同源体针对具有NTTN PAM序列的24个位点的活性的概述(显示了来自子图a的数据的平均活性)。

[0071] 图20A-20D:AsCas12a变体的工程化和表征。(A),与crRNA配对的、与编码原型TTTV PAM的推定的目标位点相互作用的Cas12a示意图和结构示意图。在结构示意图中,靠近PAM DNA碱基的氨基酸残基以绿色突出显示;图像产自PDBID:5B43(Yamano等人,Cell.2016May 5;165(4):949-62)),在PyMOL(v 1.8.6.0)中进行可视化。(B,C),当针对人类细胞中携带典型(子图B)或非典型(子图C)PAM的内源位点试验时,携带单个氨基酸取代的AsCas12a变体的活性。修饰百分比通过T7E1分析评估;显示平均值、s.e.m.和单独的数据点,n＝3。(D),对于携带典型和非典型PAM的位点,与野生型AsCas12a相比AsCas12a变体的平均活性的倍数变化。与野生型AsCas12a的活性相比较的倍数变化根据图15A中的百分比修饰数据计算。

[0072] 图21A-21H:体外PAM表征分析的优化。(A),纯化的Cas12a直系同源体和AsCas12a变体的代表性的SDS-PAGE凝胶图像;s.m,大小标准(size marker)以kDa计。(B),用作体外切割反应的底物、携带PAM和间隔子的组合的线性化质粒的示意图。SEQ ID NOs.539-540.(C),野生型AsCas12a(左图)和E174R/S542R/K548R变体(右图)对子图b中说明的底物的时程体外切割反应分布。使用一相指数衰减方程拟合曲线;误差条代表s.e.m,n＝3。(D),PAM测定分析(PAMDA)的示意图。带有代替PAM的8个随机化核苷酸的线性化质粒文库使用Cas12a核糖核蛋白(RNP)复合物进行体外切割反应。在各个时间点停止等分试样的反应,随后用作PCR的模板。带有不完全可靶向PAM的底物进行扩增和测序,以允许定量从起始文库开始随时间的PAM消耗率。(E),野生型AsCas12a(左图)和E174R/S542R/K548R变体(右图)的重复中PAMDA速率常数(k)之间的相关性。(F),两个间隔子序列的来自平均PAMDA值的速率常数之间的相关性。(G),野生型和E174R/S542R/K548R AsCas12a的PAMDA速率常数的直方图。(H),编码标明的PAM序列的底物随时间的消耗分布。使用一相指数衰减方程拟合曲线;误差条代表s.e.m,n＝4。

[0073] 图22A-22C:eAsCas12a衍生变体的PAM特异性的解卷积(deconvolution)。(A),野生型AsCas12a、E174R/S542R/K548R变体和所有中间的单取代和双取代变体的PAM偏好分布,通过PAMDA评估。$\log_{10}$速率常数(k)是四个重复的平均值,每两个针对两种不同的间隔子

序列(参见图21A-21H)。(B),E174R/S542R和E174R/S542R/K548R变体对于所有128种NNYN PAM的PAM偏好分布的比较。(C),Cas12a直系同源体的比对,其具有在这项研究中使用红色边界突出的、对于改变PAM偏好重要的残基。(SEQ ID NO:541-442)。

[0074] 图23A-23I:eAsCas12a的改善的靶向范围的评估。(A,B),E174R/S542R和E174R/S542R/K548R AsCas12a对人类细胞中携带非典型VTTN和TTCN PAM(子图a)或TATN PAM(子图b)的内源位点的活性比较。(C),野生型AsCas12a对于具有TTCN或TATN PAM的位点的活性。(D,E),E174R/S542R/K548R变体对于具有TGTV PAM的位点(子图D)或其他具有各种非典型PAM的位点(子图E)的活性。(F),人类细胞中子图A-E中试验的PAM的PAMDA速率常数和平均修饰之间的相关性。灰色阴影的方框表示0.005(或$10^{-2.25}$)的任意的PAMDA速率常数阈值,其大略地预示人类细胞中的活性。(G),野生型、E174R/S542R和E174R/S542R/K548R AsCas12a对于具有TTTN PAM的位点的活性比较。(H),AsCas12a、E174R/S542R变体和eAsCas12a对于编码TTTN PAM的26个位点的平均活性的概述(也参见图23G)。(I),eAsCas12a可靶向的PAM的概述。PAM的层级:1,高置信度PAM(平均k>0.01,平均修饰百分比>13%);2,中等置信度PAM(平均k>0.005,平均修饰百分比>10%);3,低活性或不一致的PAM(平均修饰百分比<10%,或平均k与修饰百分比之间不一致)。对于所有的子图,修饰百分比通过T7E1分析进行评估;显示平均值、s.e.m.和单独的数据点,n≥3)。

[0075] 图24A-24E:增强的Cas12a PAM变体的活性。(A-C),AsCas12a变体对具有TTTN PAM(子图A)、TATN PAM(子图B)和TYCN PAM(子图C)的内源位点修饰活性的比较。修饰百分比通过T7E1分析评估;显示平均值、s.e.m.和单独的数据点,n=3。(D),通过PAMDA评估的,原始的和增强的RVR和RR AsCas12a变体的PAM偏好分布。$\log_{10}$速率常数是四个重复的平均值,每两个针对两种不同的间隔子序列(参见图21A-21H)。(E),RVR/eRVR(上图)和RR/eRR(底图)变体对于所有128种NNYN PAM的PAM偏好分布的比较。AsCas12a变体编码以下取代:eAsCas12a,E174R/S542R/K548R;RVR,S542R/K548V/N552R;eRVR,E174R/S542R/K548V/N552R;RR,S542R/K607R;eRR,E174R/S542R/K607R。

[0076] 图25A-25K:AsCas12a和eAsCas12a的特异性的评估和改善。(A),GUIDE-Seq方法的示意图。(B,C),AsCas12a核酸酶对于GUIDE-Seq样品的中靶诱变(子图b)和GUIDE-Seq dsODN标签整合(子图c)活性的比较。修饰百分比和标签整合分别通过T7E1和RFLP分析进行评估;显示平均值、s.e.m.和单独的数据点,n=3。(D),AsCas12a核酸酶的GUIDE-Seq dsODN标签整合与总体诱变的比率;数据来自子图b和c。(E),当使用完全匹配中靶位点、或编码单个核苷酸错配的crRNA时,野生型AsCas12a或携带单取代的变体的活性。修饰百分比通过T7E1分析评估;显示平均值、s.e.m.和单独的数据点,n=3。SEQ ID NOs.553-554。(F),如子图e中评估的、携带单个氨基酸取代的eAsCas12a变体的活性。修饰百分比通过T7E1分析评估;显示平均值、s.e.m.和单独的数据点,n=3。(G),通过PAMDA评估的eAsCas12a和eAsCas12a-HF1的PAM偏好分布。$\log_{10}$速率常数是四个重复的平均值,每两个针对两种不同的间隔子序列(参见图21A-21H)。(H),eAsCas12a和eAsCas12a-HF1对于所有128种NNYN PAM的PAM偏好分布的比较。(I,J),AsCas12a、eAsCas12a和eAsCas12a-HF1对于带有TTTN PAM(子图i)或非典型的VTTV、TATV和TTCVPAM(子图j)的中靶活性评估。修饰百分比通过T7E1分析评估;显示平均值、s.e.m.和单独的数据点,n=3。(K),Cas12a直系同源体和变体对PAMDA位点1底物的时程体外切割反应,在37℃、32℃和25℃下进行(分别为左、中和右图)。使用一

相指数衰减方程拟合曲线;误差条代表n＝3的s.e.m。AsCas12a变体编码以下取代:eAsCas12a,E174R/S542R/K548R;eAsCas12a-HF1,E174R/N282A/S542R/K548R。

[0077] 图26A-26F:使用Cas12a融合物的基因激活。(A),VRP激活结构域融合至DNase失活的Cas12a(dCas12a)直系同源体和变体的示意图。(B),大略地涵盖VEGFA转录起始位点(TSS)上游700bp的序列窗口的例示,标明了SpCas9和Cas12a的目标位点。(C,D),dCas12a-VPR和dSpCas9-VPR架构(分别使用crRNA或sgRNA对)的活性比较;crRNA靶向VEGFA启动子中具有TTTV PAM(子图C)或TTCV PAM(子图D)的位点。与含有deAs-VPR(1.3)的对照转染和模拟crRNA质粒相比,通过VEGFA产生方面的改变来评估的活性;显示平均值、s.e.m.和单独的数据点,n＝4。(E,F),使用三个或两个(分别为子图e和f)crRNA或sgRNA的库,对于一系列具有dCas12a-VPR融合物的典型和非典型PAM的位点,通过dCas12a-VPR(1.1)或dSpCas9-VPR融合蛋白的VEGFA激活;显示平均值、s.e.m.和单独的数据点,n≥3。VPR,合成的VP64-p65-Rta激活结构域(Chavez等人,Nat Methods.,2015,12:326-8);NLS(sv),SV40核定位信号;NLS(nuc),核质蛋白的核定位信号;HA,人类流感血细胞凝集素标签;gs,甘氨酸-丝氨酸肽接头。

[0078] 图27A-D:使用Cas12a的碱基编辑。(A),在8个目标位点的20nt间隔子中的所有C之中,与未处理的对照相比在C到T编辑方面的倍数变化。(B),前面的(5')碱基的身份对胞嘧啶到胸腺嘧啶(C到T)的转换的影响。对于涵盖每个目标位点的-14～+30区域的窗口(4nt PAM的上游另外10nt,以及20nt间隔子序列下游的10nt)中的所有的C,对八个目标位点的C到T编辑效率(参见图18I)进行标绘。(C),五个目标位点中六个选定的胞嘧啶处的编辑纯度的分析。相对于每个BE构建体的位置上所有的非C出现的总和,标绘每个非C身份的分数。(D),通过确定-14～+30窗口内编码插入或删除突变的等位基因的百分比,不计算仅有取代的等位基因,对每个BE/crRNA对计算了Cas12a-BE的插入或删除突变(indel)活性。

[0079] 图28A-28B:使用AsCas12和eAsCas12a的DNA检测。(A),通过DNase诱导的报告分子荧光,使用野生型AsCas12a的时程DNA检测。评估了用不同的活性和失活底物进行编程时的活性。(B),eAsCas12a和eAsCas12a-HF1(分别为顶部和底部子图)随时间的DNA-检测活性。评估了用携带典型PAM、非典型PAM和非可靶向序列的底物编程时的活性。使用λ_{ex}＝485nm和λ_{em}＝528nm,每60秒采集荧光的测量值,持续三小时。

具体实施方式

[0080] 迄今为止表征的Cpf1酶识别位于间隔子序列的5'的富T PAM(图1)。AsCpf1和LbCpf1都被报道识别形式TTTN的PAM,但是强烈地偏向TTTV(其中,V＝A、C或G)。预计TTTV PAM序列在随机DNA的每43个碱基中大约出现一次,潜在地限制了AsCpf1和LbCpf1用于基因组编辑的靶向范围(和有用性)。工程化核酸酶的靶向范围对于需要DNA双链断裂(DSB)的精确靶向或放置的应用是特别重要的,包括但不限于:1)在小的遗传元件例如短的开放阅读框(ORF)、转录因子结合位点、微肽、miRNA等中产生插入或删除突变(indel);2)同源介导修复(HDR),其中DSB与所需序列改变的接近程度可以显著地影响修复效率,3)通过将SNP变异置于前间区序列或PAM之内而实现的等位基因特异性编辑;4)通过引入成对的DSB来产生限定长度的基因组删除或转位;5)进行基因或基因调节元件的饱和诱变;和6)使用与DNA修饰酶融合的工程化的RNA指导核酸酶或切口酶进行碱基编辑。考虑到提高的靶向范围的这些

清晰的优点，我们试图改变或松弛AsCpf1和LbCpf1的PAM特异性，以改善它们识别更多样范围的DNA序列的能力。

[0081] 在此我们证明了，在PAM邻近氨基酸残基之处或附近的取代可以改变AsCpf1和LbCpf1两者的PAM偏好，产生可以识别非相关PAM序列的这些核酸酶的变体，从而提高这种平台的靶向范围。这些工程化的CRISPR-Cas12a变体具有显著改善的性质，同时展现了扩大的靶向范围和增强的靶向活性。就我们所知，这是可以改进CRISPR核酸酶的中靶活性的氨基酸取代的首次描述。eAsCas12a的增强的性质提供了相对于当前可用的Cas12a直系同源体和变体的主要优点，展现了靶向范围方面大于8倍的改进，同时还允许更强力的多重编辑、基因激活、DNA检测和碱基编辑应用，效率是野生型AsCas12a早先不能实现的。eAsCas12a碱基编辑技术的发展扩展了基因组中可靶向碱基的范围，并且这样做几乎没有附带插入或删除突变的证据。重要地，eAsCas12a的靶向范围与早先描述的工程化的SpCas9核酸酶是可比较的，为需要扩大的PAM识别的Cas12a应用(例如，在限定的或小的基因组窗口内的靶向，多重基因组或表观基因组编辑，聚焦的编码或非编码crRNA-铺瓦(tiling)筛选，或设想复杂的组合文库筛选时)提供了更大的目标位点密度。本文描述的变体包括eAsCas12a的改进的性质使得它们成为迄今为止描述的最广泛可靶向的和活性的Cas12a酶之一。

[0082] Cpf1

[0083] 成簇规律间隔的短回文重复(CRISPR)系统编码细菌适应性免疫所必需的、RNA指导的内切核酸酶(Wright等人，Cell 164,29-44(2016))。CRISPR相关(Cas)核酸酶可以容易地进行编程来切割目标DNA序列，用于各种生物体中的基因组编辑[2-5]。这些核酸酶的一个类别，称为Cas9蛋白，与两个短的RNA复合：crRNA和反式激活crRNA(tracrRNA)[7,8]。最常用的Cas9直系同源体，SpCas9，使用在其5'末端具有与目标DNA位点的“前间区序列”区域互补的20个核苷酸(nt)的crRNA。高效的切割还需要SpCas9识别前间区序列邻近基序(PAM)。crRNA和tracrRNA通常组合成单个～100-nt的指导RNA(gRNA)[7,9-11]，其指导SpCas9的DNA切割活性。已经使用许多不同的方法表征了与不同的gRNA配对的SpCas9核酸酶的基因组广度的特异性[12-15]。还已经工程化了具有实质上改善的基因组广度特异性的SpCas9变体[16,17]。

[0084] 近来，已经鉴定出称为Cpf1的Cas蛋白，其也可以进行编程来切割目标DNA序列[1,18-20]。不同于SpCas9，Cpf1仅需要单个42-nt crRNA，其在它的3'末端具有与目标DNA序列的前间区序列互补的23nt[1]。此外，虽然SpCas9识别前间区序列3'的NGG PAM序列，AsCpf1和LbCp1识别位于前间区序列5'的TTTN PAM [1]。AsCpf1和LbCpf1的早期实验显示，这些核酸酶可以被编程来编辑人类细胞中的目标位点[1]，但是仅在少量位点上对其进行了试验。近期的研究展现了，在人类细胞中AsCpf1和LbCpf1都具有稳健的中靶活性和高基因组广度的特异性(参见，例如，Kleinstiver&Tsai等人，Nature Biotechnology 2016；和Kim等人，Nat Biotechnol.2016)。也参见US20160208243。

[0085] 本发现为工程化的AsCpf1和LbCpf1变体提供支持，其在本文中一起称为“变体”或“所述变体”。

[0086] 本文描述的所有变体可以快速地掺入现有的和广泛使用的载体，例如，通过简单的定点诱变。

[0087] 因而，本文提供的是AsCpf1变体。AsCpf1野生型蛋白质序列如下：

[0088] AsCpf1-V型CRISPR相关蛋白Cpf1[氨基酸球菌种BV3L6],NCBI参考序列:WP_021736722.1

[0089]

```
   1 MTQFEGFTNL YQVSKTLRFE LIPQGKTLKH IQEQGFIEED KARNDHYKEL KPIIDRIYKT
  61 YADQCLQLVQ LDWENLSAAI DSYRKEKTEE TRNALIEEQA TYRNAIHDYF IGRTDNLTDA
 121 INKRHAEIYK GLFKAELFNG KVLKQLGTVT TTEHENALLR SFDKFTTYFS GFYENRKNVF
 181 SAEDISTAIP HRIVQDNFPK FKENCHIFTR LITAVPSLRE HFENVKKAIG IFVSTSIEEV
 241 FSFPFYNQLL TQTQIDLYNQ LLGGISREAG TEKIKGLNEV LNLAIQKNDE TAHIIASLPH
 301 RFIPLFKQIL SDRNTLSFIL EEFKSDEEVI QSFCKYKTLL RNENVLETAE ALFNELNSID
 361 LTHIFISHKK LETISSALCD HWDTLRNALY ERRISELTGK ITKSAKEKVQ RSLKHEDINL
 421 QEIISAAGKE LSEAFKQKTS EILSHAHAAL DQPLPTTLKK QEEKEILKSQ LDSLLGLYHL
 481 LDWFAVDESN EVDPEFSARL TGIKLEMEPS LSFYNKARNY ATKKPYSVEK FKLNFQMPTL
 541 ASGWDVNKEK NNGAILFVKN GLYYLGIMPK QKGRYKALSF EPTEKTSEGF DKMYYDYFPD
 601 AAKMIPKCST QLKAVTAHFQ THTTPILLSN NFIEPLEITK EIYDLNNPEK EPKKFQTAYA
 661 KKTGDQKGYR EALCKWIDFT RDFLSKYTKT TSIDLSSLRP SSQYKDLGEY YAELNPLLYH
 721 ISFQRIAEKE IMDAVETGKL YLFQIYNKDF AKGHHGKPNL HTLYWTGLFS PENLAKTSIK
 781 LNGQAELFYR PKSRMKRMAH RLGEKMLNKK LKDQKTPIPD TLYQELYDYV NHRLSHDLSD
 841 EARALLPNVI TKEVSHEIIK DRRFTSDKFF FHVPITLNYQ AANSPSKFNQ RVNAYLKEHP
 901 ETPIIGIDRG ERNLIYITVI DSTGKILEQR SLNTIQQFDY QKKLDNREKE RVAARQAWSV
 961 VGTIKDLKQG YLSQVIHEIV DLMIHYQAVV VLENLNFGFK SKRTGIAEKA VYQQFEKMLI

1021 DKLNCLVLKD YPAEKVGGVL NPYQLTDQFT SFAKMGTQSG FLFYVPAPYT SKIDPLTGFV
1081 DPFVWKTIKN HESRKHFLEG FDFLHYDVKT GDFILHFKMN RNLSFQRGLP GFMPAWDIVF
1141 EKNETQFDAK GTPFIAGKRI VPVIENHRFT GRYRDLYPAN ELIALLEEKG IVFRDGSNIL
1201 PKLLENDDSH AIDTMVALIR SVLQMRNSNA ATGEDYINSP VRDLNGVCFD SRFQNPEWPM
1261 DADANGAYHI ALKGQLLLNH LKESKDLKLQ NGISNQDWLA YIQELRN  (SEQ ID NO:2)
```

[0090] 本文描述的AsCpf1变体可以包括SEQ ID NO:2的氨基酸序列,例如,至少包含SEQ ID NO:2的氨基酸1-1307,在表1中的一个或多个位置处具有突变(即,用不同的氨基酸例如,丙氨酸、甘氨酸或丝氨酸(除非其中所述天然氨基酸是丝氨酸)替换天然氨基酸),例如,在以下位置:SEQ ID NO:2的T167、S170、E174、T539、K548、N551、N552、M604和/或K607(或例如SEQ ID NO:9的与之类似的位置)。在某些实施方案中,所述AsCpf1变体与SEQ ID NO:2的氨基酸序列至少80%,例如至少85%、90%或95%相同,例如,在被替换的SEQ ID NO:2的残基的高达5%、10%、15%或20%处有差异,例如除了本文描述的突变之外,具有保守性突变。在优选的实施方案中,所述变体保持了亲本(parent)的期望的活性,例如核酸酶活性(除非其中所述亲本是切口酶或死的Cpf1),和/或与指导RNA和目标DNA相互作用的能力。

[0091] 本文还提供的是LbCpf1变体。LbCpf1野生型蛋白质序列如下:

[0092] LbCpf1-V型CRISPR相关蛋白Cpf1[毛螺旋菌科细菌ND2006],GenBank Acc No.WP_051666128.1

[0093]

```
   1 MLKNVGIDRL DVEKGRKNMS KLEKFTNCYS LSKTLRFKAI PVGKTQENID NKRLLVEDEK
  61 RAEDYKGVKK LLDRYYLSFI NDVLHSIKLK NLNNYISLFR KKTRTEKENK ELENLEINLR
 121 KEIAKAFKGN EGYKSLFKKD IIETILPEFL DDKDEIALVN SFNGFTTAFT GFFDNRENMF
 181 SEEAKSTSIA FRCINENLTR YISNMDIFEK VDAIFDKHEV QEIKEKILNS DYDVEDFFEG
 241 EFFNFVLTQE GIDVYNAIIG GFVTESGEKI KGLNEYINLY NQKTKQKLPK FKPLYKQVLS
 301 DRESLSFYGE GYTSDEEVLE VFRNTLNKNS EIFSSIKKLE KLFKNFDEYS SAGIFVKNGP
 361 AISTISKDIF GEWNVIRDKW NAEYDDIHLK KKAVVTEKYE DDRRKSFKKI GSFSLEQLQE
 421 YADADLSVVE KLKEIIIQKV DEIYKVYGSS EKLFDADFVL EKSLKKNDAV VAIMKDLLDS
 481 VKSFENYIKA FFGEGKETNR DESFYGDFVL AYDILLKVDH IYDAIRNYVT QKPYSKDKFK
 541 LYFQNPQFMG GWDKDKETDY RATILRYGSK YYLAIMDKKY AKCLQKIDKD DVNGNYEKIN
 601 YKLLPGPNKM LPKVFFSKKW MAYYNPSEDI QKIYKNGTFK KGDMFNLNDC HKLIDFFKDS
 661 ISRYPKWSNA YDFNFSETEK YKDIAGFYRE VEEQGYKVSF ESASKKEVDK LVEEGKLYMF
 721 QIYNKDFSDK SHGTPNLHTM YFKLLFDENN HGQIRLSGGA ELFMRRASLK KEELVVHPAN
 781 SPIANKNPDN PKKTTTLSYD VYKDKRFSED QYELHIPIAI NKCPKNIFKI NTEVRVLLKH
 841 DDNPYVIGID RGERNLLYIV VVDGKGNIVE QYSLNEIINN FNGIRIKTDY HSLLDKKEKE
 901 RFEARQNWTS IENIKELKAG YISQVVHKIC ELVEKYDAVI ALEDLNSGFK NSRVKVEKQV
 961 YQKFEKMLID KLNYMVDKKS NPCATGGALK GYQITNKFES FKSMSTQNGF IFYIPAWLTS
1021 KIDPSTGFVN LLKTKYTSIA DSKKFISSFD RIMYVPEEDL FEFALDYKNF SRTDADYIKK
1081 WKLYSYGNRI RIFRNPKKNN VFDWEEVCLT SAYKELFNKY GINYQQGDIR ALLCEQSDKA
1141 FYSSFMALMS LMLQMRNSIT GRTDVDFLIS PVKNSDGIFY DSRNYEAQEN AILPKNADAN
1201 GAYNIARKVL WAIGQFKKAE DEKLDKVKIA ISNKEWLEYA QTSVKH  (SEQ ID NO:3)
```

[0094] 本文描述的LbCpf1变体可以包括SEQ ID NO:3的氨基酸序列,例如,至少包含SEQ ID NO:3的氨基酸23-1246,在以下位置中的一处或多处具有突变(即,用不同的氨基酸例如,丙氨酸、甘氨酸或丝氨酸替换天然氨基酸):SEQ ID NO:11的T152、D156、G532和/或K538(或与之类似的位置,例如,SEQ ID NO:3的T170、D174、G550和/或K556);SEQ ID NO:3的氨基酸19-1246与SEQ ID NO:11的1-1228相同(SEQ ID NO:11的氨基酸1-1228在本文中也称为LbCPF1(-18))。在某些实施方案中,所述LbCpf1变体与SEQ ID NO:3的氨基酸序列至少80%,例如至少85%、90%或95%相同,例如,在被替换的SEQ ID NO:3的残基的高达5%、10%、15%或20%处有差异,例如除了本文描述的突变之外,具有保守性突变。在优选的实施方案中,所述变体保持了亲本的期望的活性,例如核酸酶活性(除非其中所述亲本是切口酶或死的Cpf1),和/或与指导RNA和目标DNA相互作用的能力。本工作实施例中使用的LbCpf1的形式起始于MSKLEK基序,省略了上文方框中的前18个氨基酸,如Zetsche等人Cell 163,759-771(2015)中所描述的。

[0095] 本文还提供的是FnCpf1变体。FnCpf1野生型蛋白质序列如下:

[0096] FnCpf1-V型CRISPR相关蛋白Cpf1[土拉弗朗西斯氏菌],GenBank Acc No.WP_003040289.1

[0097]

```
   1 MSIYQEFVNK YSLSKTLRFE LIPQGKTLEN IKARGLILDD EKRAKDYKKA KQIIDKYHQF
  61 FIEEILSSVC ISEDLLQNYS DVYFKLKKSD DDNLQKDFKS AKDTIKKQIS EYIKDSEKFK
 121 NLFNQNLIDA KKGQESDLIL WLKQSKDNGI ELFKANSDIT DIDEALEIIK SFKGWTTYFK
 181 GFHENRKNVY SSNDIPTSII YRIVDDNLPK FLENKAKYES LKDKAPEAIN YEQIKKDLAE
 241 ELTFDIDYKT SEVNQRVFSL DEVFEIANFN NYLNQSGITK FNTIIGGKFV NGENTKRKGI
 301 NEYINLYSQQ INDKTLKKYK MSVLFKQILS DTESKSFVID KLEDDSDVVT TMQSFYEQIA
 361 AFKTVEEKSI KETLSLLFDD LKAQKLDLSK IYFKNDKSLT DLSQQVFDDY SVIGTAVLEY
 421 ITQQIAPKNL DNPSKKEQEL IAKKTEKAKY LSLETIKLAL EEFNKHRDID KQCRFEEILA
 481 NFAAIPMIFD EIAQNKDNLA QISIKYQNQG KKDLLQASAE DDVKAIKDLL DQTNNLLHKL
 541 KIFHISQSED KANILDKDEH FYLVFEECYF ELANIVPLYN KIRNYITQKP YSDEKFKLNF
 601 ENSTLANGWD KNKEPDNTAI LFIKDDKYYL GVMNKKNNKI FDDKAIKENK GEGYKKIVYK
 661 LLPGANKMLP KVFFSAKSIK FYNPSEDILR IRNHSTHTKN GSPQKGYEKF EFNIEDCRKF
 721 IDFYKQSISK HPEWKDFGFR FSDTQRYNSI DEFYREVENQ GYKLTFENIS ESYIDSVVNQ
 781 GKLYLFQIYN KDFSAYSKGR PNLHTLYWKA LFDERNLQDV VYKLNGEAEL FYRKQSIPKK
 841 ITHPAKEAIA NKNKDNPKKE SVFEYDLIKD KRFTEDKFFF HCPITINFKS SGANKFNDEI
 901 NLLLKEKAND VHILSIDRGE RHLAYYTLVD GKGNIIKQDT FNIIGNDRMK TNYHDKLAAI
 961 EKDRDSARKD WKKINNIKEM KEGYLSQVVH EIAKLVIEYN AIVVFEDLNF GFKRGRFKVE
1021 KQVYQKLEKM LIEKLNYLVF KDNEFDKTGG VLRAYQLTAP FETFKKMGKQ TGIIYYVPAG
1081 FTSKICPVTG FVNQLYPKYE SVSKSQEFFS KFDKICYNLD KGYFEFSFDY KNFGDKAAKG
1141 KWTIASFGSR LINFRNSDKN HNWDTREVYP TKELEKLLKD YSIEYGHGEC IKAAICGESD
1201 KKFFAKLTSV LNTILQMRNS KTGTELDYLI SPVADVNGNF FDSRQAPKNM PQDADANGAY
1261 HIGLKGLMLL GRIKNNQEGK KLNLVIKNEE YFEFVQNRNN    (SEQ ID NO:4)
```

[0098] 本文描述的FnCpf1变体可以包括SEQ ID NO:4的氨基酸序列,在以下位置中的一处或多处具有突变(即,用不同的氨基酸例如,丙氨酸、甘氨酸或丝氨酸替换天然的氨基酸):SEQ ID NO:4的K180、E184、N607、K613、D616、N617和/或K671。在某些实施方案中,所述FnCpf1变体与SEQ ID NO:4的氨基酸序列至少80%,例如至少85%、90%或95%相同,例如,在被替换的SEQ ID NO:4的残基的高达5%、10%、15%或20%处有差异,例如除了本文描述的突变之外,具有保守性突变。在优选的实施方案中,所述变体保持了亲本的期望的活性,例如核酸酶活性(除非其中所述亲本是切口酶或死的Cpf1),和/或与指导RNA和目标DNA相互作用的能力。

[0099] 为了确定两个核酸序列的同一性百分比,为了最佳比较的目的将序列比对(例如,可以在第一和第二氨基酸或核酸序列之一或两者中引入空位用于最佳的比对,并且出于比较的目的可以忽视非同源的序列)。出于比较目的而比对的参考序列的长度是参考序列长度的至少80%,在某些实施方案中是至少90%或100%。然后比较相应的氨基酸位置或核苷酸位置处的核苷酸。当第一序列中的位置被与第二序列中相应位置相同的核苷酸占据时,则该分子在该位置是同一的(即,如本文使用的,核酸“同一性”等同于核酸“同源性”)。两个序列之间的同一性百分比是序列共有的同一的位置的数量的函数,考虑了为了两个序列的最佳比对需要引入的空位的数量、每个空位的长度。按照本领域技术技术范围内的各种方式确定两个多肽或核酸序列之间的同一性百分比,例如,使用公众可获得的计算机软件,例如,Smith Waterman Alignment(Smith,T.F.和M.S.Waterman(1981)J Mol Biol 147:195-7);已合并到GeneMatcher Plus™中的“BestFit”(Smith和Waterman,Advances in Applied Mathematics,482-489(1981)),Schwarz和Dayhof(1979)Atlas of Protein Sequence and Structure,Dayhof,M.O.,编著,pp 353-358;BLAST程序(Basic Local Alignment Search Tool;(Altschul,S.F.,W.Gish,等人(1990)J Mol Biol215:403-10),BLAST-2、BLAST-P、BLAST-N、BLAST-X、WU-BLAST-2、ALIGN、ALIGN-2、CLUSTAL或Megalign(DNASTAR)软件。此外,本领域技术人员可以确定用于衡量比对的合适参数,包括在被比较的序列全长上实现最大比对所需的任何算法。一般地,对于蛋白质或核酸,比较的长度可以是任何长度,达到或包

括全长(例如,5%、10%、20%、30%、40%、50%、60%、70%、80%、90%、95%或100%)。出于本组合物和方法的目的,序列全长的至少80%进行比对。

[0100] 为了本发明的目的,序列的比较和两个序列之间同一性百分比的确定可以使用Blossum 62评分矩阵、使用空位罚分12、空位延伸罚分4以及移框缺口罚分5来完成。

[0101] 保守性取代一般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。

[0102] 在某些实施方案中,所述突变体具有代替野生型氨基酸的丙氨酸。在某些实施方案中,所述突变体具有精氨酸或赖氨酸(或天然氨基酸)以外的任何氨基酸。

[0103] 在某些实施方案中,所述Cpf1变体还包括表A中列出的以下突变之一,其降低或破坏Cpf1的核酸酶活性:

[0104] 表A

[0105]

涉及DNA和RNA催化的残基				
	AsCpf1	LbCpf1(+18)	*LbCpf1*	FnCpf1
DNA靶向	D908	D850	*D832*	D917
	E911	E853	*E835*	E920
	N913	N855	*N837*	H922
	Y916	Y858	*Y840*	Y925
	E993	E943	*E925*	E1006
	R1226	R1156	*R1138*	R1218
	S1228	S1158	*S1140*	S1220
	D1235	D1166	*D1148*	D1227
	D1263	D1198	*D1180*	D1255
RNA加工	H800	H777	*H759*	H843
	K809	K786	*K768*	K852
	K860	K803	*K785*	K869
	F864	F807	*F789*	F873
将Cpf1转变为切口酶的突变				
	R1226A	R1156A	*R1138A*	R1218A

[0106] 参见,例如,Yamano等人,Cell.2016 May 5;165(4):949-62;Fonfara等人,Nature.2016 Apr 28;532(7600):517-21;Dong等人,Nature.2016 Apr 28;532(7600):522-6;和Zetsche等人,Cell.2015 Oct 22;163(3):759-71。注意,“LbCpf1(+18)”是指SEQ ID NO:3的全部氨基酸序列1-1246,而LbCpf1是指Zetsche等人中LbCpf1的序列,本文也显示为SEQ ID NO:11的氨基酸1-1228和SEQ ID NO:3的氨基酸19-1246。

[0107] 因而,在某些实施方案中,对于AsCpf1,催化活性破坏突变在D908和E993处进行,例如D908A和E993A;对于LbCpf1,催化活性破坏突变在D832和E925处进行,例如D832A和E925A。

[0108] 在某些实施方案中,所述Cpf1变体还包括提高特异性的突变(即,诱导实质上更少的脱靶效应),例如,WO2018/022634中描述的。例如,LbCpf1变体蛋白可以包括一个、两个、三个、四个、五个、六个或全部七个以下位置处的一个或多个突变:S202、N274、N278、K290、

K367、K532、K609、K915、Q962、K963、K966、K1002和/或S1003，例如，S202A、N274A、N278A、K290A、K367A、K532A、K609A、K915A、Q962A、K963A、K966A、K1002A和/或S1003A。AsCpf1变体蛋白可以包括一个、两个、三个、四个、五个或六个以下位置处的一个或多个突变：N178、N278、N282、R301、T315、S376、N515、K523、K524、K603、K965、Q1013和/或K1054，例如，N178A、N278A、N282A、R301A、T315A、S376A、N515A、K523A、K524A、K603A、K965A、Q1013A和/或K1054A。在某些实施方案中，所述AsCpf1变体包括N282A、T315A、N515A或N278A的突变。

[0109] 本文还提供的是编码所述Cpf1变体的分离的核酸；载体，其包含任选地与一个或多个调节结构域可操作连接的分离的核酸，以用于表达所述变体蛋白；以及宿主细胞，例如哺乳动物宿主细胞，其包含所述核酸并任选地表达所述变体蛋白。

[0110] 本文描述的变体可以用于改变细胞的基因组；所述方法一般包括在所述细胞中表达所述变体蛋白，以及具有与所述细胞的基因组的选定部分互补的区域的指导RNA。选择性地改变细胞的基因组的方法是本领域已知的，参见，例如，US 8,993,233;US 20140186958;US 9,023,649;WO/2014/099744;WO 2014/089290;WO2014/144592;WO144288;WO2014/204578;WO2014/152432;WO2115/099850;US8,697,359;US20160024529;US20160024524;US20160024523;US20160024510;US20160017366;US20160017301;US20150376652;US20150356239;US20150315576;US20150291965;US20150252358;US20150247150;US20150232883;US20150232882;US20150203872;US20150191744;US20150184139;US20150176064;US20150167000;US20150166969;US20150159175;US20150159174;US20150093473;US20150079681;US20150067922;US20150056629;US20150044772;US20150024500;US20150024499;US20150020223;;US20140356867;US20140295557;US20140273235;US20140273226;US20140273037;US20140189896;US20140113376;US20140093941;US20130330778;US20130288251;US20120088676;US20110300538;US20110236530;US20110217739;US20110002889;US20100076057;US20110189776;US20110223638;US20130130248;US20150050699;US20150071899;US20150045546;US20150031134;US20150024500;US20140377868;US20140357530;US20140349400;US20140335620;US20140335063;US20140315985;US20140310830;US20140310828;US20140309487;US20140304853;US20140298547;US20140295556;US20140294773;US20140287938;US20140273234;US20140273232;US20140273231;US20140273230;US20140271987;US20140256046;US20140248702;US20140242702;US20140242700;US20140242699;US20140242664;US20140234972;US20140227787;US20140212869;US20140201857;US20140199767;US20140189896;US20140186958;US20140186919;US20140186843;US20140179770;US20140179006;US20140170753;WO/2008/108989;WO/2010/054108;WO/2012/164565;WO/2013/098244;WO/2013/176772;Makarova等人，"Evolution and classification of the CRISPR-Cas systems"9(6)Nature Reviews Microbiology 467-477(1-23)(Jun.2011);Wiedenheft等人，"RNA-guided genetic silencing systems in bacteria and archaea"482Nature 331-338(Feb.16,2012);Gasiunas等人，"Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria"109(39)Proceedings of theNational Academy of Sciences USA E2579-E2586(Sep.4,2012);Jinek等人，"A Programmable

Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity"337Science 816-821(Aug.17,2012);Carroll,"ACRISPR Approach to Gene Targeting"20(9) Molecular Therapy 1658-1660(Sep.2012);2012年5月25日提交的美国申请No.61/652,086;Al-Attar等人,Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs):The Hallmark of an Ingenious Antiviral Defense Mechanism in Prokaryotes,Biol Chem.(2011)vol.392,Issue 4,pp.277-289;Hale等人,Essential Features and Rational Design of CRISPR RNAs That Function With the Cas RAMP Module Complex to Cleave RNAs,Molecular Cell,(2012)vol.45,Issue 3,292-302。

[0111] 本文描述的变体蛋白可以用于代替或补充上述参考文献中描述的任何Cas9或Cpf1蛋白,或与其中描述的类似突变体组合,与适合于选定Cpf1的指导RNA一起使用,即,与靶向不同于野生型PAM的目标序列的指导RNA一起使用,例如,具有根据以下表B的PAM序列的指导RNA。

[0112]

表B		
变体蛋白	**较强的PAM**	**较弱的PAM**
AsCpf1 S170R	TTTN, CTTN, GTTN	TATN, TCTN, TTAN, TTCN, TTGN
AsCpf1 E174R	TTTN, CTTN, GTTN, TTCN	TATN, TCTN, TTAN, TTGN
AsCpf1 S542K	TTTN, GTTN, TTCN	CTTN, TCTN, TTAN
AsCpf1 S542Q	TTTN	TTCN
AsCpf1 S542R	TTTN, ATTN, CTTN, GTTN, TCTN, TTCN, ATCN, CCCN, CCTN, GCTN, GGTN, TCCN	TGTN, TATN, CTCN, TGCN
AsCpf1 N551R	TTTN	GTTN
AsCpf1 N552R		TTTN
AsCpf1 T167A/T539K	TTTN, GTTN	

[0113]

表B		
变体蛋白	**较强的PAM**	**较弱的PAM**
AsCpf1 T167A/T539R	TTTN, GTTN	
AsCpf1 E174R/S542R	TTTN, CTTN, TTCN	
AsCpf1 S542R/K548R	TTTN, CTTN, GTTN, TTCN	TATN, TGTN, TTAN
AsCpf1 S542R/N551R	TTTN, GTTN, TTCN	CTTN
AsCpf1 S542R/N552R	TTTN	GTTN
AsCpf1 K548R/N551R	TTTN, GTTN	
AsCpf1 K548R/N552R		TTTN
AsCpf1 S542R/M604A	TTTN	TTCN
AsCpf1 S542R/K607H	TTTN, TTCN	
AsCpf1 E174R/S542R/K548R	TTTN, CTTN, GTTN, ATTN, TATN, TGTN, TTAN, TTCN, AGTN, CATN, CCCN, CGTN, CTCN, GATN, GCTN, GGTN, GTCN, TACN, TCCN, TGCN, ACCN	TCTN, TTGN, AAAN, ACTN, ATCN, CCTN, CTAN
AsCpf1 S542R/K548R/N551R	TTTN, GTTN, TTCN	CTTN, TATN, TGTN, TTAN
AsCpf1 S170R/S542R/K607R	TTTN, TTCN, TCCN, TCTN, ACCN	GTTN, TTAN
AsCpf1 E174R/S542R/K607H	CTTN, TCTN, TTCN, TCCN, TTTN	CCCN, ACCN
AsCpf1 E174R/S542R/K607R	TTTN, TTCN, TCCN, CCCN, ACCN, GCCN	CTTN, GTTN, TCTN, TTAN, TTGN
AsCpf1 E174R/S542R/K548R/N551R	TTTN, CTTN, GTTN, TTCN, TCCN, CCCN, ACCN	TATN
AsCpf1 E174R/S542R/K548R/N552R	TTTN, CTTN, GTTN	TATN
AsCpf1 E174R/S542R/K548V/N552R	TTTN, CTTN, GTTN, TATN	TCTN, TGTN, TTCN, TCCN
AsCpf1 S170R/S542R/K548V/N552R	TTTN, GTTN, TATN,	CTTN, TGTN, TTCN
LbCpf1 T152R	TTTN, TTCN	
LbCpf1 T152K	TTTN, TTCN	
LbCpf1 D156R	TTTN, TTCN	
LbCpf1 D156K	TTTN, TTCN	
LbCpf1 G532R	TTTN, TTCN	
LbCpf1 K538R	TTTN	TTCN
LbCpf1 D156R/G532R/K538R	TTTN, CTTN, GTTN, TTAN, TTCN, TTGN, TCCN	TATN, TCTN
FnCpf1 K180R	TTTN, CTTN, GTTN, NTTN, TCTN	TTAN, TTCN

[0114]

表B		
变体蛋白	较强的PAM	较弱的PAM
FnCpf1 N607R	TTTN, CTTN, GTTN, NTTN, TCTN, TTCN	TTAN
FnCpf1 K613R	TTTN, CTTN, GTTN, NTTN,	TTCN, TGTN
FnCpf1 K613V	TTTN, CTTN, GTTN, NTTN,	TATN, TGTN, TTCN
FnCpf1 D616R	TTTN, CTTN, GTTN, NTTN, TCTN, TTCN	TTAN
FnCpf1 N617R	TTTN, CTTN, GTTN, NTTN, TCTN	TTCN
FnCpf1 K671R	TTTN, TCTN	CTTN, GTTN, NTTN, TTCN
FnCpf1 K671H	TTTN, CTTN, GTTN, TCTN, NTTN	TTCN
FnCpf1 K607R/K613V	TTTN, CTTN, GTTN, NTTN, TGTN, TTCN	TATN, TCTN, TTAN
FnCpf1 K607R/K613V/D616R	TTTN, GTTN, GTTN, NTTN, TATN, TCTN, TGTN, TTAN, TTCN	
FnCpf1 K607R/K613R/D616R	TTTN, GTTN, GTTN, NTTN, TGTN, TTAN, TTCN	TCTN,

[0115] 本文描述的变体还可以用于体外检测样品中的目标ssDNA或dsDNA的方法，例如US20170362644;East-Seletsky等人,Nature.2016 Oct 13;538(7624):270-273;Gootenberg等人,Science.2017 Apr 28;356(6336):438-442;Gootenberg等人,Science 10.1126/science.aaq0179(2018);Chen等人,Science.2018 Feb 15.pii:eaar6245;Science.2018 Feb 15.pii:eaaq0179;和WO2017219027A1中描述的。在这些方法中，所述变体与它的靶点的结合诱导针对其他靶点的非特异性DNase活性。所述方法包括使已知或怀疑包括目标ssDNA或dsDNA的样品接触融合蛋白(或多个融合蛋白)、与该融合蛋白一起工作的相关的指导RNA、以及标记的检测子DNA(例如，3-30nt、3-20、5-20、5-15或其他适合的长度的报告子ssDNA)。当融合蛋白结合它的靶点时，非特异性DNase活性切割检测子DNA，产生信号。测量来自标记的检测子DNA的信号的方法是本领域已知的，可以包括，例如，检测金纳米颗粒、荧光团、荧光偏振、胶体相转变/分散、电化学信号和基于半导体的信号中的一种或多种。在某些实施方案中，标记的检测子DNA在被切割之前产生一定量的可检测信号，当标记的检测子DNA被切割时可检测信号的量降低。可选地，标记的检测子DNA在被切割之前可以产生第一可检测信号，当标记的检测子DNA被切割时产生第二可检测信号。在某些实施方案中，标记的检测子DNA包含淬灭剂/fluor对。在某些实施方案中，还包括Csm6，一种辅助CRISPR相关酶。

[0116] 此外，本文描述的变体可以用于本领域已知的融合蛋白中代替野生型Cas9或其他Cas9突变体(例如，dCas9或Cas9切口酶)，例如US 8,993,233;US 20140186958;US 9,023,649;WO/2014/099744;WO 2014/089290;WO2014/144592;WO144288;WO2014/204578;WO2014/152432;WO2115/099850;US8,697,359;US2010/0076057;US2011/0189776;US2011/0223638;US2013/0130248;WO/2008/108989;WO/2010/054108;WO/2012/164565;WO/2013/098244;WO/2013/176772;US20150050699;US 20150071899和WO 2014/124284中描述的具

有异源功能结构域的融合蛋白。例如，所述变体，优选地包含一个或多个核酸酶-降低或杀灭突变，可以在Cpf1的N或C末端融合至转录激活结构域(例如，来自单纯性疱疹病毒的VP16结构域(Sadowski等人，1988，Nature，335:563-564)或VP64的转录激活结构域；来自细胞转录因子NF-κB的p65结构域(Ruben等人，1991，Science，251:1490-93)；或融合于dCas9的三联效应子，其由串联连接的激活子VP64、p65和Rta(VPR)组成，Chavez等人，Nat Methods.2015Apr;12(4):326-8)或其他异源功能结构域(例如，转录阻遏蛋白(例如，KRAB、ERD、SID和其它的蛋白，例如，ets2阻遏因子(ERF)阻遏蛋白结构域(ERD)的氨基酸473-530，KOX1的KRAB结构域的氨基酸1-97，或Mad mSIN3相互作用结构域(SID)的氨基酸1-36；参见Beerli等人，PNAS USA 95:14628-14633(1998))或沉默子，例如异染色质蛋白1(HP1，也称为swi6)，例如，HP1α或HP1β；可以募集长的非编码RNA(lncRNA)的蛋白或肽，其与固定的RNA结合序列融合，例如，被MS2衣壳蛋白、核糖核酸内切酶Csy4或λN蛋白结合的那些；碱基编辑子(修饰DNA甲基化状态的酶(例如，DNA甲基转移酶(DNMT)或TET蛋白)；或者也可以使用本领域已知的修饰组蛋白亚单位的酶(例如，组蛋白乙酰基转移酶(HAT)，组蛋白脱乙酰基酶(HDAC)，组蛋白甲基转移酶(例如，用于赖氨酸或精氨酸残基的甲基化)，或组蛋白脱甲基酶(例如，用于赖氨酸或精氨酸残基的脱甲基化)。这样的结构域的许多序列是本领域已知的，例如，催化DNA中甲基化的胞嘧啶的羟基化的结构域。示范性的蛋白包括十-十一-转位(TET)1-3家族，是将DNA中的5-甲基胞嘧啶(5-mC)转换成5-羟基甲基胞嘧啶(5-hmC)的酶。

[0117] 人类TET1-3的序列是本领域已知的，在下表中显示：

[0118]

	GenBank登录号	
基因	氨基酸	核酸
TET1	NP_085128.2	NM_030625.2
TET2*	NP_001120680.1(变体1) NP_060098.3(变体2)	NM_001127208.2 NM_017628.4
TET3	NP_659430.1	NM_144993.1

[0119] *变体(1)代表更长的转录产物，并且编码更长的同种型(a)。与变体1相比，变体(2)在5'UTR和3'UTR以及编码序列上不同。与同种型a相比，产生的同种型(b)是较短的，并具有不同的C-末端。

[0120] 在某些实施方案中，可以包括所述催化结构域的全长序列的全部或部分，例如，包含富半胱氨酸延伸的催化模块以及由7个高度保守的外显子编码的2OGFeDO结构域，例如，包含氨基酸1580-2052的Tet1催化结构域，包含氨基酸1290-1905的Tet2，和包含氨基酸966-1678的Tet3。参见，例如，Iyer等人，Cell Cycle.2009 Jun 1;8(11):1698-710.Epub 2009 Jun 27的图1，其中的比对说明了所有三种Tet蛋白的关键催化残基，以及全长序列的其他补充材料(可从ftp站点ftp.ncbi.nih.gov/pub/aravind/DONS/supplementary_material_DONS.html获得)(参见，例如，seq 2c)；在某些实施方案中，所述序列包括Tet1的氨基酸1418-2136或Tet2/3中的相应区域。

[0121] 其他催化模块可以来自Iyer等人，2009中鉴定的蛋白。

[0122] 在某些实施方案中，所述异源功能结构域是碱基编辑子，例如，修饰胞嘧啶DNA碱基的脱氨酶，例如，来自载脂蛋白B mRNA编辑酶的胞苷脱氨酶，脱氨酶的催化性多肽样(APOBEC)家族，包括APOBEC1、APOBEC2、APOBEC3A、APOBEC3B、APOBEC3C、APOBEC3D/E、APOBEC3F、APOBEC3G、APOBEC3H、APOBEC4(参见，例如，Yang等人，J Genet Genomics.2017

Sep 20;44(9):423-437);活化诱导的胞苷脱氨酶(AID),例如,活化诱导的胞苷脱氨酶(AICDA),胞嘧啶脱氨酶1(CDA1)和CDA2,以及作用于tRNA的胞嘧啶脱氨酶(CDAT)。下表提供了示范性的序列;也可以使用其他序列。

[0123]

	GenBank登录号	
脱氨酶	核酸	氨基酸
hAID/AICDA	NM_020661.3同种型1 NM_020661.3同种型2	NP_065712.1变体1 NP_065712.1变体2
APOBEC1	NM_001644.4同种型a NM_005889.3同种型b	NP_001635.2变体1 NP_005880.2变体3
APOBEC2	NM_006789.3	NP_006780.1
APOBEC3A	NM_145699.3同种型a NM_001270406.1同种型b	NP_663745.1变体1 NP_001257335.1变体2
APOBEC3B	NM_004900.4同种型a NM_001270411.1同种型b	NP_004891.4变体1 NP_001257340.1变体2
APOBEC3C	NM_014508.2	NP_055323.2
APOBEC3D/E	NM_152426.3	NP_689639.2
APOBEC3F	NM_145298.5同种型a NM_001006666.1同种型b	NP_660341.2 变体1 NP_001006667.1变体2
APOBEC3G	NM_021822.3(同种型a)	NP_068594.1(变体1)
APOBEC3H	NM_001166003.2	NP_001159475.2(变体SV-200)
APOBEC4	NM_203454.2	NP_982279.1
CDA1*	NM_127515.4	NP_179547.1

[0124] *来自酿酒酵母(Saccharomyces cerevisiae)S288C

[0125] 在某些实施方案中,所述异源功能结构域是修饰腺苷DNA碱基的脱氨酶,例如,所述脱氨酶是腺苷脱氨酶1(ADA1)、ADA2;作用于RNA的腺苷脱氨酶1(ADAR1)、ADAR2、ADAR3(参见,例如,Savva等人,Genome Biol.2012 Dec 28;13(12):252);作用于tRNA的腺苷脱氨酶1(ADAT1)、ADAT2、ADAT3(参见Keegan等人,RNA.2017Sep;23(9):1317-1328以及Schaub和Keller,Biochimie.2002Aug;84(8):791-803);以及天然发生的或工程化的tRNA特异性腺苷脱氨酶(TadA)(参见,例如,Gaudelli等人,Nature.2017 Nov 23;551(7681):464-471)(NP_417054.2(大肠杆菌(Escherichia coli)str.K-12substr.MG1655);参见,例如Wolf等人,EMBO J.2002Jul 15;21(14):3841-51)。下表提供了示范性的序列;也可以使用其他序列。

[0126]

	GenBank登录号	
脱氨酶	核酸	氨基酸
ADA(ADA1)	NM_000022.3 变体1	NP_000013.2 同种型1
ADA2	NM_001282225.1	NP_001269154.1
ADAR	NM_001111.4	NP_001102.2
ADAR2(ADARB1)	NM_001112.3 变体1	NP_001103.1 同种型1
ADAR3(ADARB2)	NM_018702.3	NP_061172.1
ADAT1	NM_012091.4 变体1	NP_036223.2 同种型1
ADAT2	NM_182503.2 变体1	NP_872309.2 同种型1
ADAT3	NM_138422.3 变体1	NP_612431.2 同种型1

[0127] 在某些实施方案中,所述异源功能结构域是抑制或增强内源DNA修复或碱基切除修复(BER)途径的酶、结构域或肽,例如,胸腺嘧啶DNA糖基化酶(TDG;GenBank Acc Nos.NM_003211.4(核酸)和NP_003202.3(蛋白))或尿嘧啶DNA糖基化酶(UDG,也称为尿嘧啶N-糖基化酶,或UNG;GenBank Acc Nos.NM_003362.3(核酸)和NP_003353.1(蛋白),或抑制UNG介导的尿嘧啶切除以启动BER的尿嘧啶DNA糖基化酶抑制剂(UGI)(参见,例如Mol等人,Cell 82,701-708(1995);Komor等人,Nature.2016 May 19;533(7603)),或DNA末端结合蛋白,例如Gam,其是一种来自噬菌体的蛋白质,结合游离的DNA末端、抑制DNA修复酶、并引起更精确的编辑(更少的非预期碱基编辑;Komor等人,Sci Adv.2017 Aug 30;3(8):eaao4774)。

[0128] 在某些实施方案中,可以使用所述蛋白的全部或部分,例如,至少是保持了酶的预期功能的催化结构域。

[0129] 在某些实施方案中,所述异源功能结构域是生物系链,包含MS2衣壳蛋白、核糖核酸内切酶Csy4或λN蛋白的全部或部分(例如,来自它们的DNA结合结构域)。这些蛋白可以用于将含有特定茎-环结构的RNA分子募集到dCpf1gRNA靶向序列指定的位置。例如,与MS2衣壳蛋白、核糖核酸内切酶Csy4或λN融合的dCpf1变体可以用于募集长的非编码RNA(lncRNA)例如XIST或HOTAIR;参见,例如,Keryer-Bibens等人,Biol.Cell 100:125-138(2008),其与所述Csy4、MS2或λN结合序列连接。可选地,Csy4、MS2或λN蛋白结合序列可以连接到另一蛋白,例如上文的Keryer-Bibens等人中描述的,所述蛋白可以使用本文描述的方法和组合物靶向dCpf1变体结合位点。在某些实施方案中,Csy4是无催化活性的。在某些实施方案中,Cpf1变体,优选地dCpf1变体,融合至如US 8,993,233;US 20140186958;US 9,023,649;WO/2014/099744;WO 2014/089290;WO2014/144592;WO144288;WO2014/204578;WO2014/152432;WO2115/099850;US8,697,359;US2010/0076057;US2011/0189776;US2011/0223638;US2013/0130248;WO/2008/108989;WO/2010/054108;WO/2012/164565;WO/2013/098244;WO/2013/176772;US20150050699;US 20150071899和WO 2014/204578中描述的FokI。

[0130] 在某些实施方案中,所述融合蛋白包括所述Cpf1变体和所述异源功能结构域之间

的接头。可以用于这些融合蛋白(或串联结构的融合蛋白之间)的接头可以包括不干扰所述融合蛋白的功能的任何序列。在优选的实施方案中,所述接头是短的,例如,2-20个氨基酸,并且一般是柔性的(即,包含高自由度的氨基酸,例如,甘氨酸、丙氨酸和丝氨酸)。在某些实施方案中,所述接头包含由GGGS(SEQ ID NO:12)或GGGGS(SEQ ID NO:13)组成的一个或多个单元,例如,GGGS(SEQ ID NO:12)或GGGGS(SEQ ID NO:13)单元的两个、三个、四个或多个重复。也可以使用其他接头序列。

[0131] 在某些实施方案中,所述变体蛋白包括促进向细胞内空间递送的细胞穿透肽序列,例如,HIV衍生的TAT肽、穿透素(penetratin)、转运蛋白(transportan)或hCT衍生的细胞穿透性肽,参见,例如Caron等人,(2001)Mol Ther.3(3):310-8;Langel,Cell-Penetrating Peptides:Processes and Applications(CRC Press,Boca Raton FL 2002);El-Andaloussi等人,(2005)Curr Pharm Des.11(28):3597-611;和Deshayes等人,(2005)Cell Mol Life Sci.62(16):1839-49。

[0132] 细胞穿透肽(CPP)是短肽,其促进各种各样的生物分子跨越细胞膜移动进入细胞质或其他细胞器,例如线粒体和核。可以通过CPP递送的分子的实例包括治疗药物、质粒DNA、寡核苷酸、siRNA、肽-核酸(PNA)、蛋白、肽、纳米颗粒和脂质体。CPP一般是30个氨基酸或更少,来源于天然或非天然发生的蛋白或嵌合序列,含有高相对丰度的正电荷氨基酸,例如赖氨酸或精氨酸,或者交替模式的极性和非极性氨基酸。本领域中常用的CPP包括Tat(Frankel等人,(1988)Cell.55:1189-1193,Vives等人,(1997)J.Biol.Chem.272:16010-16017)、穿透素(Derossi等人,(1994)J.Biol.Chem.269:10444-10450)、聚精氨酸肽序列(Wender等人,(2000)Proc.Natl.Acad.Sci.USA 97:13003-13008,Futaki等人,(2001)J.Biol.Chem.276:5836-5840)和转运蛋白(Pooga等人,(1998)Nat.Biotechnol.16:857-861)。

[0133] CPP可以通过共价或非共价策略与它们的货物连接。共价连接CPP和它的货物的方法是本领域已知的,例如,化学交联(Stetsenko等人,(2000)J.Org.Chem.65:4900-4909,Gait等人(2003)Cell.Mol.Life.Sci.60:844-853)或克隆融合蛋白(Nagahara等人,(1998)Nat.Med.4:1449-1453)。货物与包含极性和非极性结构域的短的两亲性CPP之间的非共价偶联通过静电和疏水性相互作用来建立。

[0134] 本领域中已经利用CPP来向细胞递送潜在治疗性的生物分子。实例包括用于免疫抑制的连接到聚精氨酸的环孢霉素(Rothbard等人,(2000)Nature Medicine 6(11):1253-1257)、用于抑制肿瘤发生的、与CPP连接的针对周期蛋白B1的siRNA,称为MPG(Crombez等人,(2007)Biochem Soc.Trans.35:44-46),与CPP连接的肿瘤抑制物p53肽以降低癌细胞生长(Takenobu等人,(2002)Mol.Cancer Ther.1(12):1043-1049,Snyder等人,(2004)PLoS Biol.2:E36),以及与Tat融合的Ras的显性阴性形式或磷酸肌醇3激酶(PI3K)以治疗哮喘(Myou等人,(2003)J.Immunol.171:4399-4405)。

[0135] 本领域已经利用CPP将造影剂转运到细胞内用于成像和生物传感应用。例如,附着到Tat的绿色荧光蛋白(GFP)已经用于标记癌细胞(Shokolenko等人,(2005)DNA Repair 4(4):511-518)。缀合至量子点的Tat已经用于成功地跨越血脑屏障,用于大鼠脑的可视化(Santra等人,(2005)Chem.Commun.3144-3146)。CPP还与磁共振成象技术组合用于细胞成像(Liu等人,(2006)Biochem.and Biophys.Res.Comm.347(1):133-140)。还参见Ramsey和

Flynn,Pharmacol Ther.2015 Jul 22.pii:S0163-7258(15)00141-2。

[0136] 在某些实施方案中,可选地或另外地,所述变体蛋白可以包括核定位序列,例如,SV40大T抗原NLS(PKKKRRV(SEQ ID NO:16))和核质蛋白NLS(KRPAATKKAGQAKKKK(SEQ ID NO:7))。其他NLS是本领域已知的,参见,例如,Cokol等人,EMBO Rep.2000 Nov 15;1(5):411-415;Freitas和Cunha,Curr Genomics.2009 Dec;10(8):550-557。

[0137] 在某些实施方案中,所述变体包括对配体例如GST、FLAG或六组氨酸序列具有高亲和性的部分。这样的亲和标签可以便于重组变体蛋白的纯化。

[0138] 对于其中将变体蛋白递送给细胞的方法,所述蛋白可以使用本领域已知的任何方法产生,例如,通过体外翻译,或在合适的宿主细胞中从编码所述变体蛋白的核酸表达;生产蛋白的许多方法是本领域已知的。例如,所述蛋白可以在酵母、大肠杆菌(E.coli)、昆虫细胞系、植物、转基因动物或培养的哺乳动物细胞中产生并从中纯化;参见,例如,Palomares等人,"Production of Recombinant Proteins:Challenges and Solutions," Methods Mol Biol.2004;267:15-52。此外,所述变体蛋白可以连接到便于转移入细胞的部分,例如,脂质纳米颗粒,任选地使用一旦蛋白处于细胞之内就被切割的接头。参见,例如LaFountaine等人,Int J Pharm.2015 Aug 13;494(1):180-194。

[0139] 表达系统

[0140] 为了使用本文描述的Cpf1变体,可期望从编码它们的核酸中表达它们。这可以以多种方式进行。例如,编码Cpf1变体的核酸可以克隆到中间载体中,用于转化入原核或真核细胞用于复制和/或表达。中间载体一般是原核生物载体,例如,质粒,或穿梭载体、或昆虫载体,用于编码所述Cpf1变体的核酸的保存或操作,用于生产Cpf1变体。编码Cpf1变体的核酸还可以克隆到表达载体中,用于向植物细胞、动物细胞,优选哺乳动物细胞或人类细胞、真菌细胞、细菌细胞或原生动物细胞施用。

[0141] 为了获得表达,一般将编码Cpf1变体的序列亚克隆到含有指导转录的启动子的表达载体中。适合的细菌和真核启动子是本领域公知的,例如,在Sambrook等人,Molecular Cloning,A Laboratory Manual(第3版,2001);Kriegler,Gene Transfer and Expression:A Laboratory Manual(1990);和Current Protocols in Molecular Biology(Ausubel等人编辑,2010)中描述。用于表达工程化的蛋白的细菌表达系统可以获自例如大肠杆菌、芽胞杆菌属物种(Bacillus sp.)和沙门氏菌属(Salmonella)(Palva等人,1983,Gene 22:229-235)。用于这类表达系统的试剂盒是商购可得的。用于哺乳动物细胞、酵母和昆虫细胞的真核表达系统是本领域公知的,也是商购可得的。

[0142] 用于指导核酸表达的启动子取决于具体的应用。例如,强组成型启动子一般用于融合蛋白的表达和纯化。相比之下,当Cpf1变体将被体内施用而用于基因调节时,可以使用组成型或诱导型启动子,取决于Cpf1变体的具体运用。此外,用于Cpf1变体施用的优选的启动子可以是弱启动子,例如,HSVTK或具有类似活性的启动子。启动子还可以包括响应于反式激活的元件,例如,低氧响应元件、Gal4响应元件、lac阻遏物响应元件和小分子控制系统,例如四环素调节系统和RU-486系统(参见,例如,Gossen&Bujard,1992,Proc.Natl.Acad.Sci.USA,89:5547;Oligino等人,1998,Gene Ther.,5:491-496;Wang等人,1997,Gene Ther.,4:432-441;Neering等人,1996,Blood,88:1147-55;和Rendahl等人,1998,Nat.Biotechnol.,16:757-761)。

[0143] 除了启动子之外，表达载体一般地含有转录单位或表达盒，其含有核酸在原核或真核宿主细胞中表达所需的所有其他元件。因而典型的表达盒含有启动子，其可操作地连接至，例如，编码所述Cpf1变体的核酸序列，以及例如转录产物的高效多聚腺苷酸化、转录终止、核糖体结合位点或翻译终止所需的任何信号。所述盒的其他元件可以包括，例如，增强子和异源剪接的内含子信号。

[0144] 针对所述Cpf1变体的预期用途选择用于将遗传信息转运入细胞的具体表达载体，例如，在植物、动物、细菌、真菌、原生动物等中表达。标准的细菌表达载体包括质粒，例如基于pBR322的质粒、pSKF、pET23D和商购可得的标签-融合物表达系统例如GST和LacZ。

[0145] 含有来自真核生物病毒的调节元件的表达载体常常用于真核表达载体中，例如，SV40载体、乳头状瘤病毒载体和来自Epstein-Barr病毒的载体。其他示范性的真核载体包括pMSG、pAVO09/A+、pMTO10/A+、pMAMneo-5、杆状病毒pDSVE，和任何其他载体，所述载体容许蛋白质在SV40早期启动子、SV40晚期启动子、金属硫蛋白启动子、鼠类乳腺肿瘤病毒启动子、劳氏肉瘤病毒启动子、多角体蛋白(polyhedrin)启动子或其他显示了真核细胞中的表达有效性的启动子的指导下表达。

[0146] 用于表达所述Cpf1变体的载体可以包括RNA Pol III启动子以驱动指导RNA的表达，例如，H1、U6或7SK启动子。这些人类启动子容许Cpf1变体在质粒转染之后在哺乳动物细胞中表达。

[0147] 某些表达系统具有用于选择稳定转染的细胞系的标志物，例如，胸苷激酶、潮霉素B磷酸转移酶和二氢叶酸还原酶。高产量表达系统也是适合的，例如，在昆虫细胞中使用杆状病毒载体，以及在多角体蛋白启动子或其他强杆状病毒启动子的指导下的gRNA编码序列。

[0148] 一般包括在表达载体中的元件还包括在大肠杆菌中起作用的复制子，编码抗生素抗性的基因以允许选择带有重组质粒的细菌，处在质粒的非关键区域中的独特的限制性位点，以容许重组序列的插入。

[0149] 使用标准的转染方法来产生表达大量蛋白的细菌、哺乳动物、酵母或昆虫细胞系，然后使用标准技术进行纯化(参见，例如，Colley等人，1989，J.Biol.Chem.，264:17619-22；Guide to Protein Purification，in Methods in Enzymology，vol.182(Deutscher编辑，1990))。根据标准技术进行真核和原核细胞的转化(参见，例如，Morrison，1977，J.Bacteriol.132:349-351；Clark-Curtiss&Curtiss，Methods in Enzymology 101:347-362(Wu等人，编辑，1983)。

[0150] 可以使用向宿主细胞中引入外源核苷酸序列的任何已知的操作。这些包括利用磷酸钙转染、聚凝胺、原生质体融合、电穿孔、核染、脂质体、显微注射、裸DNA、质粒载体、病毒载体，游离的和整合的，以及任何其他公知的方法将克隆的基因组DNA、cDNA、合成的DNA或其他外源遗传材料导入宿主细胞中(参见，例如，上文的Sambrook等人)。仅仅必需的是，所使用的具体的遗传工程操作能够向能表达所述Cpf1变体的宿主细胞中成功地导入至少一种基因。

[0151] 本发明还包括载体和包含载体的细胞。

[0152] 本文还提供包含本文描述的变体的组合物和试剂盒。在某些实施方案中，所述试剂盒包括融合蛋白和相关的指导RNA(即，结合所述蛋白并将它导向适合该蛋白的目标序列

的指导RNA)。在某些实施方案中,所述试剂盒还包括标记的检测子DNA,例如,用于检测目标ssDNA或dsDNA的方法中。标记的检测子DNA是本领域已知的,例如,在US20170362644;East-Seletsky等人,Nature.2016 Oct 13;538(7624):270-273;Gootenberg等人,Science.2017 Apr 28;356(6336):438-442和WO2017219027A1中描述,可以包括包含荧光共振能量转移(FRET)对或淬灭剂/fluor对或这两者的标记的检测子DNA。所述试剂盒还可以包括一种或多种另外的试剂,例如,另外的酶(例如,RNA聚合酶)和缓冲液,例如,用于本文描述的方法。

[0153] 实施例

[0154] 在以下实施例中进一步描述本发明,其不限制本发明的范围,本发明的范围在权利要求中描述。

[0155] 方法

[0156] 除非另有说明,以下的材料和方法用于下文的实施例。

[0157] 质粒和寡核苷酸

[0158] 实施例1B、4B和5-8中使用的crRNA的目标位点序列和寡核苷酸序列可以分别从表2A-2D和3A-3D获得。野生型AsCas12a、LbCas12a、FnCas12a和MbCas12a的人类表达质粒(分别为SQT1659、SQT1665、AAS1472、AAS2134)通过将来自质粒pY010、pY016、pY004和pY014(Addgene质粒69982、69988、69976和69986;Feng Zhang赠送)的核酸酶开放阅读框亚克隆至pCAG-CFP(Addgene质粒11179;Connie Cepko赠送)的NotI和AgeI位点中来产生。通过将As、Fn和MbCas12a的人类密码子优化的开放阅读框、以及细菌密码子优化的LbCas12a开放阅读框(来自Addgene质粒79008;Jin Soo Kim赠送)分别克隆到pET28b-Cas9(Addgene质粒47327;Alex Schier赠送)NcoI和FseI位点中来产生BPK3541、RTW656、RTW660和RTW645。所有的Cas12a变体、激活子构建体和碱基编辑子融合物通过标准的分子克隆和等温组装来产生。通过将与间隔子序列双螺旋相应的寡核苷酸退火并连接至BsmBI消化的BPK3079、BPK3082、BPK4446和BPK4449中,分别产生用于Cas12a crRNA的人类细胞表达质粒,用于As、Lb、Fn和MbCas12a crRNA的U6启动子驱动的转录。用于体外切割反应的底物质粒通过将目标位点克隆到pUC19的NheI和HindIII位点中产生。通过将与间隔子序列双螺旋相应的寡核苷酸退火并连接到BsaI消化的MSP3491、MSP3495、RTW763和RTW767中,分别产生用于Cas12a crRNA的体外转录的质粒,用于As、Lb、Fn和MbCas12a crRNA的T7启动子驱动的转录。

[0159] 表2A.用于核酸酶实验的Cas12a crRNA

[0160]

crRNA ID	基因名称	4nt PAM	20nt 间隔子	SEQ ID NO:
AACC-1	*FANCF*	AACC	AGTGGAGGCAAGAGGGCGGC	26.
AACC-2	*RUNX1*	AACC	AAGACAGGTCACTGTTTCAG	27.
AACC-3	*EMX1*	AACC	ACACCTTCACCTGGGCCAGG	28.
AACC-4	*EMX1*	AACC	GGTGGCGCATTGCCACGAAG	29.

[0161]

crRNA ID	基因名称	4nt PAM	20nt 间隔子	SEQ ID NO:
AATA-1	*FANCF*	AATA	GCATTGCAGAGAGGCGTATC	30.
AATA-2	*RUNX1*	AATA	TGCTGTCTGAAGCCATCGCT	31.
AATA-3	*DNMT1*	AATA	AGTGGCAGAGTGCTAAGGGA	32.
AATA-4	*EMX1* (扩增子 2)	AATA	TGGAGCCTGCTCCAGGTGGG	33.
AATC-1	*FANCF*	AATC	AGTACGCAGAGAGTCGCCGT	34.
AATC-2	*CFTR*	AATC	CTAACTGAGACCTTACACCG	35.
AATG-1	*EMX1*	AATG	CGCCACCGGTTGATGTGATG	36.
ACCC-1	*VEGFA*	ACCC	CGGCTCTGGCTAAAGAGGGA	37.
ACCC-2	*VEGFA*	ACCC	CCTATTTCTGACCTCCCAAA	38.
ACCC-3	*DNMT1*	ACCC	AGAGGCTCAAGTGAGCAGCT	39.
ACCC-4	*EMX1*	ACCC	TAGTCATTGGAGGTGACATC	40.
ACCC-5	*EMX1*	ACCC	ACGAGGGCAGAGTGCTGCTT	41.
ACCC-6	*DNMT1*	ACCC	AATAAGTGGCAGAGTGCTAA	42.
AGCC-1	*FANCF*	AGCC	GCCCTCTTGCCTCCACTGGT	43.
AGCC-2	*RUNX1*	AGCC	ATCGCTTCCTCCTGAAAATG	44.
AGCC-3	*RUNX1*	AGCC	TCACCCCTCTAGCCCTACAT	45.
AGCC-4	*EMX1* (扩增子 2)	AGCC	TGCTCCAGGTGGGGAATAAG	46.
AGTA-1	*DNMT1*	AGTA	ACAGACATGGACCATCAGGA	47.
AGTA-2	*CFTR*	AGTA	CCAGATTCTGAGCAGGGAGA	48.
AGTC-1	*DNMT1*	AGTC	TCCGTGAACGTTCCCTTAGC	49.
AGTC-2	*CFTR*	AGTC	TGTCCTGAACCTGATGACAC	50.
ATCA-1	*DNMT1*	ATCA	GGAAACATTAACGTACTGAT	51.
ATCA-2	*CFTR*	ATCA	GAATCCTCTTCGATGCCATT	52.
ATCC-1	*DNMT1*	ATCC	TCACAGCAGCCCCTTGAGAA	53.
ATCC-2	*CFTR*	ATCC	AATCAACTCTATACGAAAAT	54.
ATCC-3	*DNMT1*	ATCC	CCAACATGCACTGATGTTGT	55.
ATCC-4	*FANCF*	ATCC	ATCGGCGCTTTGGTCGGCAT	56.
ATTA-1	*DNMT1*	ATTA	ACGTACTGATGTTAACAGCT	57.
ATTA-2	*EMX1* (扩增子 2)	ATTA	ACATTAACAAGAAGCATTTG	58.
ATTA-3	*EMX1* (扩增子 2)	ATTA	TTCAAGTGGCGCAGATCTAG	59.
ATTA-4	*CFTR*	ATTA	GAAGGAGATGCTCCTGTCTC	60.
ATTC-1	*DNMT1*	ATTC	ACCGAGCAGGAGTGAGGGAA	61.
ATTC-2	*EMX1* (扩增子 2)	ATTC	CCCACCTGGAGCAGGCTCCA	62.
ATTC-3	*CFTR*	ATTC	TGATGAGCCTTTAGAGAGAA	63.
ATTC-4	*VEGFA*	ATTC	CCTCTTTAGCCAGAGCCGGG	64.
ATTC-5	*FANCF*	ATTC	GCACGGCTCTGGAGCGGCGG	65.
ATTG-1	*DNMT1*	ATTG	GGTCAGCTGTTAACATCAGT	66.
ATTG-2	*EMX1* (扩增子 2)	ATTG	TTATGAACCTGGGTGAAGTC	67.
ATTG-3	*VEGFA*	ATTG	GAATCCTGGAGTGACCCCTG	68.
ATTG-4	*CFTR*	ATTG	GATTGAGAATAGAATTCTTC	69.
ATTG-5	*FANCF*	ATTG	GAACATCCGCGAAATGATAC	70.
ATTT-1	*DNMT1*	ATTT	GGCTCAGCAGGCACCTGCCT	71.
ATTT-2	*EMX1* (扩增子 2)	ATTT	GCTTTCCACCCACCTTTCCC	72.
ATTT-3	*VEGFA*	ATTT	CTGACCTCCCAAACAGCTAC	73.

[0162]

crRNA ID	基因名称	4nt PAM	20nt 间隔子	SEQ ID NO:
ATTT-4	*CFTR*	ATTT	CTTCTTTCTGCACTAAATTG	74.
ATTT-5	*FANCF*	ATTT	CGCGGATGTTCCAATCAGTA	75.
CACC-1	*FANCF*	CACC	GTGCGCCGGGCCTTGCAGTG	76.
CACC-2	*RUNX1*	CACC	GAGGCATCTCTGCACCGAGG	77.
CCCC-1	*FANCF*	CCCC	GCCCAAAGCCGCCCTCTTGC	78.
CCCC-2	*RUNX1*	CCCC	GCCTTCAGAAGAGGGTGCAT	79.
CCCC-3	*DNMT1*	CCCC	AGAGGGTTCTAGACCCAGAG	80.
CCCC-4	*DNMT1*	CCCC	AGGGCCAGCCCAGCAGCCAA	81.
CGCA-1	*FANCF*	CGCA	CGGCTCTGGAGCGGCGGCTG	82.
CGCA-2	*EMX1*	CGCA	TTGCCACGAAGCAGGCCAAT	83.
CGCC-1	*FANCF*	CGCC	GCTCCAGAGCCGTGCGAATG	84.
CGCC-2	*EMX1*	CGCC	ACCGGTTGATGTGATGGGAG	85.
CGCC-3	*FANCF*	CGCC	ACATCCATCGGCGCTTTGGT	86.
CGCC-4	*FANCF*	CGCC	GATGGATGTGGCGCAGGTAG	87.
CGTC-1	*FANCF*	CGTC	AGCACCTGGGACCCCGCCAC	88.
CGTC-2	*FANCF*	CGTC	TCCAAGGTGAAAGCGGAAGT	89.
CTCA-1	*DNMT1*	CTCA	AACGGTCCCCAGAGGGTTCT	90.
CTCA-2	*CFTR*	CTCA	AAACTCATGGGATGTGATTC	91.
CTCC-1	*DNMT1*	CTCC	GTGAACGTTCCCTTAGCACT	92.
CTCC-2	*CFTR*	CTCC	TTCTAATGAGAAACGGTGTA	93.
CTCC-3	*FANCF*	CTCC	ACTGGTTGTGCAGCCGCCGC	94.
CTCC-4	*FANCF*	CTCC	AGAGCCGTGCGAATGGGGCC	95.
CTCT-1	*DNMT1*	CTCT	GGGGAACACGCCCGGTGTCA	96.
CTTA-1	*DNMT1*	CTTA	TTGGGTCAGCTGTTAACATC	97.
CTTA-2	*EMX1* (扩增子 2)	CTTA	TTCCCCACCTGGAGCAGGCT	98.
CTTA-3	*RUNX1*	CTTA	CTAATCAGATGGAAGCTCTT	99.
CTTA-4	*CFTR*	CTTA	CACCGTTTCTCATTAGAAGG	100.
CTTC-1	*FANCF*	CTTC	CGCTTTCACCTTGGAGACGG	101.
CTTC-2	*EMX1* (扩增子 2)	CTTC	ACCCAGGTTCATAACAATGT	102.
CTTC-3	*VEGFA*	CTTC	TCCCCGCTCCAACGCCCTCA	103.
CTTC-4	*CFTR*	CTTC	TAATGAGAAACGGTGTAAGG	104.
CTTC-5	*FANCF*	CTTC	GCGCACCTCATGGAATCCCT	105.
CTTG-1	*DNMT1*	CTTG	ACAGGCGAGTAACAGACATG	106.
CTTG-2	*EMX1* (扩增子 2)	CTTG	TTAATGTTAATAACTTGCTT	107.
CTTG-3	*CFTR*	CTTG	GTTAACTGAGTGTGTCATCA	108.
CTTG-4	*RUNX1*	CTTG	GGGAGTCCCAGAGGTATCCA	109.
CTTT-1	*DNMT1*	CTTT	GGTCAGGTTGGCTGCTGGGC	110.
CTTT-2	*EMX1* (扩增子 2)	CTTT	CCCTGGCCTACCTCACTGGC	111.
CTTT-3	*VEGFA*	CTTT	AGCCAGAGCCGGGGTGTGCA	112.
CTTT-4	*CFTR*	CTTT	AGAGAGAAGGCTGTCCTTAG	113.
CTTT-5	*FANCF*	CTTT	GGTCGGCATGGCCCCATTCG	114.
GCCC-1	*DNMT1*	GCCC	GGTGTCACGCCACTTGACAG	115.
GCCC-2	*CFTR*	GCCC	CACGCTTCAGGCACGAAGGA	116.
GTCA-1	*DNMT1*	GTCA	CGCCACTTGACAGGCGAGTA	117.

[0163]

crRNA ID	基因名称	4nt PAM	20nt 间隔子	SEQ ID NO:
GTCA-2	*CFTR*	GTCA	TCAGGTTCAGGACAGACTGC	118.
GTCC-1	*DNMT1*	GTCC	CCAGAGGGTTCTAGACCCAG	119.
GTCC-2	*CFTR*	GTCC	AGGAGACAGGAGCATCTCCT	120.
GTCC-3	*FANCF*	GTCC	CAGGTGCTGACGTAGGTAGT	121.
GTCC-4	*EMX1*	GTCC	TCCCCATTGGCCTGCTTCGT	122.
GTTA-1	*DNMT1*	GTTA	CTCGCCTGTCAAGTGGCGTG	123.
GTTA-2	*EMX1*（扩增子 2）	GTTA	TGAACCTGGGTGAAGTCCCA	124.
GTTA-3	*EMX1*（扩增子 2）	GTTA	ATGTTAATAACTTGCTTCAA	125.
GTTA-4	*CFTR*	GTTA	ACCAAGGTCAGAACATTCAC	126.
GTTC-1	*DNMT1*	GTTC	CCTTAGCACTCTGCCACTTA	127.
GTTC-2	*EMX1*（扩增子 2）	GTTC	ATTTGTCCAGAGGAAACCAC	128.
GTTC-3	*RUNX1*	GTTC	CCTGTCTTGTTTGTGAGAGG	129.
GTTC-4	*CFTR*	GTTC	AGGACAGACTGCCTCCTTCG	130.
GTTG-1	*DNMT1*	GTTG	GGGATTCCTGGTGCCAGAAA	131.
GTTG-2	*EMX1*（扩增子 2）	GTTG	GGACTTCACCCAGGTTCATA	132.
GTTG-3	*VEGFA*	GTTG	AGGGCGTTGGAGCGGGGAGA	133.
GTTG-4	*CFTR*	GTTG	ATTGGATTGAGAATAGAATT	134.
GTTG-5	*FANCF*	GTTG	TGCAGCCGCCGCTCCAGAGC	135.
GTTT-1	*DNMT1*	GTTT	CCTGATGGTCCATGTCTGTT	136.
GTTT-2	*EMX1*（扩增子 2）	GTTT	GACTTGGGATAGTGGAATAG	137.
GTTT-3	*VEGFA*	GTTT	GGGAGGTCAGAAATAGGGGG	138.
GTTT-4	*CFTR*	GTTT	CTCATTAGAAGGAGATGCTC	139.
GTTT-5	*RUNX1*	GTTT	CACCTCGGTGCAGAGATGCC	140.
TACA-1	*RUNX1*	TACA	TCTCTCTTTCTTCTCCCCTC	141.
TACA-2	*RUNX1*	TACA	GGCAAAGCTGAGCAAAAGTA	142.
TACA-3	*EMX1*	TACA	AACGGCAGAAGCTGGAGGAG	143.
TACA-4	*RUNX1*	TACA	AGACCAGCATGTACTCACCT	144.
TACC-1	*DNMT1*	TACC	CACGTTCGTGGCCCCATCTT	145.
TACC-2	*CFTR*	TACC	AGATTCTGAGCAGGGAGAGG	146.
TACC-3	*EMX1*（扩增子 2）	TACC	TCACTGGCCCCACCCCAGAG	147.
TACC-4	*FANCF*	TACC	TGCGCCACATCCATCGGCGC	148.
TATA-1	*CFTR*	TATA	GAGTTGATTGGATTGAGAAT	149.
TATA-2	*CFTR*	TATA	TTCAAGAAGGTTATCTCAAG	150.
TATA-3	*CFTR*	TATA	TATTCAAGAAGGTTATCTCA	151.
TATA-4	*VEGFA*（扩增子 2）	TATA	GACATGTCCCATTTGTGGGA	152.
TATC-1	*CFTR*	TATC	GCCTCTCCCTGCTCAGAATC	153.
TATC-2	*CFTR*	TATC	TCAAGAAACTGGCTTGGAAA	154.
TATC-3	*CFTR*	TATC	CAGTTCAGTCAAGTTTGCCT	155.
TATC-4	*EMX1*（扩增子 2）	TATC	CCAAGTCAAACTTCTCTTCA	156.
TATG-1	*VEGFA*（扩增子 2）	TATG	TTCGGGTGCTGTGAACTTCC	157.
TATG-2	*EMX1*（扩增子 2）	TATG	AACCTGGGTGAAGTCCCAAC	158.
TATG-3	*VEGFA*	TATG	TAGCTGTTTGGGAGGTCAGA	159.
TATG-4	*CFTR*	TATG	GGACATTTTCAGAACTCCAA	160.
TATT-1	*DNMT1*	TATT	GGGTCAGCTGTTAACATCAG	161.

[0164]

crRNA ID	基因名称	4nt PAM	20nt 间隔子	SEQ ID NO:
TATT-2	*VEGFA*	TATT	TCTGACCTCCCAAACAGCTA	162.
TATT-3	*CFTR*	TATT	CTCAATCCAATCAACTCTAT	163.
TATT-4	*FANCF*	TATT	GGTCGAAATGCATGTCAATC	164.
TCCA-1	*DNMT1*	TCCA	TGTCTGTTACTCGCCTGTCA	165.
TCCA-2	*CFTR*	TCCA	GGAGACAGGAGCATCTCCTT	166.
TCCA-3	*VEGFA*	TCCA	GTCCCAAATATGTAGCTGTT	167.
TCCC-1	*DNMT1*	TCCC	CAGAGGGTTCTAGACCCAGA	168.
TCCC-2	*CFTR*	TCCC	CAAACTCTCCAGTCTGTTTA	169.
TCCC-3	*DNMT1*	TCCC	GTCACCCCTGTTTCTGGCAC	170.
TCCC-4	*FANCF*	TCCC	AGGTGCTGACGTAGGTAGTG	171.
TCCC-5	*VEGFA*	TCCC	TCTTTAGCCAGAGCCGGGGT	172.
TCCG-1	*DNMT1*	TCCG	TGAACGTTCCCTTAGCACTC	173.
TCCG-2	*FANCF*	TCCG	AGCTTCTGGCGGTCTCAAGC	174.
TCCG-3	*VEGFA*	TCCG	CACGTAACCTCACTTTCCTG	175.
TCCT-1	*DNMT1*	TCCT	GATGGTCCATGTCTGTTACT	176.
TGCA-1	*DNMT1*	TGCA	CACAGCAGGCCTTTGGTCAG	177.
TGCA-2	*CFTR*	TGCA	GAAAGAAGAAATTCAATCCT	178.
TGCC-1	*DNMT1*	TGCC	ACTTATTGGGTCAGCTGTTA	179.
TGCC-2	*CFTR*	TGCC	TCGCATCAGCGTGATCAGCA	180.
TGCC-3	*FANCF*	TGCC	TCCACTGGTTGTGCAGCCGC	181.
TGCC-4	*FANCF*	TGCC	GACCAAAGCGCCGATGGATG	182.
TGTA-1	*RUNX1*	TGTA	ATGAAATGGCAGCTTGTTTC	183.
TGTA-2	*EMX1*	TGTA	CTTTGTCCTCCGGTTCTGGA	184.
TGTA-3	匹配的位点 5	TGTA	CCTCACCACTGACATTAATT	185.
TGTA-4	匹配的位点 5	TGTA	ACCACAGTCAAGTAGTTAAT	186.
TGTA-5	*CFTR*	TGTA	AGGTCTCAGTTAGGATTGAA	187.
TGTC-1	*FANCF*	TGTC	AATCTCCCAGCGTCTTTATC	188.
TGTC-2	*RUNX1*	TGTC	TTGTTTGTGAGAGGAATTCA	189.
TGTC-3	*EMX1* (扩增子 2)	TGTC	CAGAGGAAACCACTGTTGGG	190.
TGTC-4	*EMX1* (扩增子 2)	TGTC	TATTCCACTATCCCAAGTCA	191.
TGTC-5	*EMX1* (扩增子 2)	TGTC	CCTCACCCATCTCCCTGTGA	192.
TGTG-1	*FANCF*	TGTG	CAGCCGCCGCTCCAGAGCCG	193.
TGTG-2	*RUNX1*	TGTG	AGAGGAATTCAAACTGAGGC	194.
TGTG-3	*FANCF*	TGTG	GCGCAGGTAGCGCGCCCACT	195.
TGTG-4	*EMX1*	TGTG	GTTCCAGAACCGGAGGACAA	196.
TGTG-5	*EMX1*	TGTG	ATGGGAGCCCTTCTTCTTCT	197.
TTAC-2	*RUNX1*	TTAC	AGGCAAAGCTGAGCAAAAGT	198.
TTAC-3	*EMX1* (扩增子 2)	TTAC	TCATCTCTGCCAGACCACCT	199.
TTAC-4	匹配的位点 5	TTAC	TGATTCTGGGGTCAACATCT	200.
TTAC-5	*CFTR*	TTAC	AAATGAATGGCATCGAAGAG	201.
TTAC-6	*CFTR*	TTAC	ACCGTTTCTCATTAGAAGGA	202.
TTAC-7	匹配的位点 5	TTAC	TAGGGCAATAAGCAACACCT	203.
TTCA-1	*DNMT1*	TTCA	GTCTCCGTGAACGTTCCCTT	204.
TTCA-2	*EMX1* (扩增子 2)	TTCA	CCCAGGTTCATAACAATGTT	205.

[0165]

crRNA ID	基因名称	4nt PAM	20nt 间隔子	SEQ ID NO:
TTCA-3	*VEGFA*	TTCA	CCCAGCTTCCCTGTGGTGGC	206.
TTCA-4	*CFTR*	TTCA	ATCCTAACTGAGACCTTACA	207.
TTCA-5	*FANCF*	TTCA	CCTTGGAGACGGCGACTCTC	208.
TTCC-1	*DNMT1*	TTCC	TGATGGTCCATGTCTGTTAC	209.
TTCC-2	*EMX1* (扩增子 2)	TTCC	CTGGCCTACCTCACTGGCCC	210.
TTCC-3	*VEGFA*	TTCC	AAAGCCCATTCCCTCTTTAG	211.
TTCC-4	*CFTR*	TTCC	ATTGTGCAAAAGACTCCCTT	212.
TTCC-5	*FANCF*	TTCC	GAGCTTCTGGCGGTCTCAAG	213.
TTCG-1	*DNMT1*	TTCG	TGGCCCCATCTTTCTCAAGG	214.
TTCG-2	*VEGFA*	TTCG	AGAGTGAGGACGTGTGTGTC	215.
TTCG-3	*RUNX1*	TTCG	GAGCGAAAACCAAGACAGGT	216.
TTCG-4	*CFTR*	TTCG	ACCAATTTAGTGCAGAAAGA	217.
TTCG-5	*FANCF*	TTCG	CACGGCTCTGGAGCGGCGGC	218.
TTCT-1	*DNMT1*	TTCT	GCCCTCCCGTCACCCCTGTT	219.
TTCT-2	*EMX1* (扩增子 2)	TTCT	GCCCTTTACTCATCTCTGCC	220.
TTCT-3	*VEGFA*	TTCT	GACCTCCCAAACAGCTACAT	221.
TTCT-4	*CFTR*	TTCT	TTCGACCAATTTAGTGCAGA	222.
TTCT-5	*FANCF*	TTCT	GGCGGTCTCAAGCACTACCT	223.
TTTA-1	*DNMT1*	TTTA	TTTCCCTTCAGCTAAAATAA	224.
TTTA-2	*DNMT1*	TTTA	TTTTAGCTGAAGGGAAATAA	225.
TTTA-3	*FANCF*	TTTA	TCCGTGTTCCTTGACTCTGG	226.
TTTA-4	*RUNX1*	TTTA	CCTTCGGAGCGAAAACCAAG	227.
TTTA-5	匹配的位点 5	TTTA	GGATGCCACTAAAAGGGAAA	228.
TTTA-6	匹配的位点 1	TTTA	GATTGAAGGAAAAGTTACAA	229.
TTTC-1	*DNMT1*	TTTC	CCTCACTCCTGCTCGGTGAA	230.
TTTC-2	*DNMT1*	TTTC	CTGATGGTCCATGTCTGTTA	231.
TTTC-3	*EMX1*	TTTC	TCATCTGTGCCCCTCCCTCC	232.
TTTC-4	*FANCF*	TTTC	ACCTTGGAGACGGCGACTCT	233.
TTTC-5	*RUNX1*	TTTC	GCTCCGAAGGTAAAAGAAAT	234.
TTTC-6	*RUNX1*	TTTC	AGCCTCACCCCTCTAGCCCT	235.
TTTC-7	*RUNX1*	TTTC	TTCTCCCCTCTGCTGGATAC	236.
TTTC-8	*FANCF*	TTTC	CGAGCTTCTGGCGGTCTCAA	237.
TTTG-1	*DNMT1*	TTTG	AGGAGTGTTCAGTCTCCGTG	238.
TTTG-2	*DNMT1*	TTTG	GCTCAGCAGGCACCTGCCTC	239.
TTTG-3	*EMX1*	TTTG	TCCTCCGGTTCTGGAACCAC	240.
TTTG-4	*EMX1*	TTTG	TGGTTGCCCACCCTAGTCAT	241.
TTTG-5	*EMX1*	TTTG	TACTTTGTCCTCCGGTTCTG	242.
TTTG-6	*FANCF*	TTTG	GGCGGGGTCCAGTTCCGGGA	243.
TTTG-7	*FANCF*	TTTG	GTCGGCATGGCCCCATTCGC	244.
TTTT-1	*DNMT1*	TTTT	ATTTCCCTTCAGCTAAAATA	245.
TTTT-2	*RUNX1*	TTTT	CAGGAGGAAGCGATGGCTTC	246.
TTTT-3	*FANCF*	TTTT	CCGAGCTTCTGGCGGTCTCA	247.
TTTT-4	*CFTR*	TTTT	CGTATAGAGTTGATTGGATT	248.
TTTT-5	*CFTR*	TTTT	GAGCTAAAGTCTGGCTGTAG	249.

[0166] 表2B.用于基因激活实验的Cas12a crRNA

crRNA ID	基因名称	4nt PAM	20nt 间隔子	SEQ ID NO:	与 TSS 的距离(bp)‡	链
AR-TTTV-a-1	*AR*	TTTG	AGAGTCTGGATGAGAAATGC	250.	639	C
AR-TTTV-a-2	*AR*	TTTC	TACCCTCTTCTCTGCCTTTC	251.	588	T
AR-TTTV-a-3	*AR*	TTTG	CTCTAGGAACCCTCAGCCCC	252.	550	T
AR-TTTV-b-1	*AR*	TTTC	TCCAAAGCCACTAGGCAGGC	253.	141	C
AR-TTTV-b-2	*AR*	TTTA	GGAAAGCAGGAGCTATTCAG	254.	231	C
AR-TTTV-b-3	*AR*	TTTG	GAACCAAATTTGGTGAGTGC	255.	296	C
AR-ATTV-1	*AR*	ATTC	AGGAAGCAGGGGTCCTCCAG	256.	142	C
AR-ATTV-2	*AR*	ATTG	GGCTTTGGAACCAAATTTGG	257.	303	C
AR-ATTV-3	*AR*	ATTC	CGTCATAGGGATAGATCGGG	258.	508	T
AR-CTTV-1	*AR*	CTTG	TTTCTCCAAAGCCACTAGGC	259.	145	C
AR-CTTV-2	*AR*	CTTC	CTGAATAGCTCCTGCTTTCC	260.	227	T
AR-CTTV-3	*AR*	CTTA	TCAGTCCTGAAAAGAACCCC	261.	398	C
AR-GTTV-1	*AR*	GTTG	CATTTGCTCTCCACCTCCCA	262.	9	C
AR-GTTV-2	*AR*	GTTA	GCGCGCGGTGAGGGGAGGGG	263.	117	C
AR-GTTV-3	*AR*	GTTC	CAAAGCCCAATCTAAAAAAC	264.	312	T
AR-TTCV-1	*AR*	TTCA	GGAAGCAGGGGTCCTCCAGG	265.	212	C
AR-TTCV-2	*AR*	TTCC	TGGAGGCCAGCACTCACCAA	266.	283	T
AR-TTCV-3	*AR*	TTCA	GGACTGATAAGAGCGCGCAG	267.	407	T
[0167] AR-CTCC-1	*AR*	CTCC	AAAGCCACTAGGCAGGCGTT	268.	138	C
AR-CTCC-2	*AR*	CTCC	AGGAAATCTGGAGCCCTGGC	269.	268	C
AR-CTCC-3	*AR*	CTCC	CTCCCTCGCCTCCACCCTGT	270.	338	C
AR-TCCC-1	*AR*	TCCC	GCCCCCACCGGGCCGGCCTC	271.	48	T
AR-TCCC-2	*AR*	TCCC	CTCACCGCGCGCTAACGCCT	272.	121	T
AR-TCCC-3	*AR*	TCCC	TCGCCTCCACCCTGTTGGTT	273.	333	C
HBB-TTTV-1	*HBB*	TTTG	TACTGATGGTATGGGGCCAA	274.	203	C
HBB-TTTV-2	*HBB*	TTTG	AAGTCCAACTCCTAAGCCAG	275.	150	C
HBB-TTTV-3	*HBB*	TTTG	CAAGTGTATTTACGTAATAT	276.	248	T
HBB-ATTV-1	*HBB*	ATTG	GCCAACCCTAGGGTGTGGCT	277.	71	T
HBB-ATTV-2	*HBB*	ATTG	CTACTAAAAACATCCTCCTT	278.	226	T
HBB-ATTV-3	*HBB*	ATTG	GGAAAACGATCTTCAATATG	279.	293	T
HBB-CTTV-1	*HBB*	CTTA	GACCTCACCCTGTGGAGCCA	280.	90	C
HBB-CTTV-2	*HBB*	CTTA	GGAGTTGGACTTCAAACCCT	281.	154	T
HBB-CTTV-3	*HBB*	CTTA	CCAAGCTGTGATTCCAAATA	282.	269	C
HBB-TATV-1	*HBB*	TATG	CCCAGCCCTGGCTCCTGCCC	283.	28	T
HBB-TATV-2	*HBB*	TATC	TCTTGGCCCCATACCATCAG	284.	197	T
HBB-TATV-3	*HBB*	TATC	CCAAAGCTGAATTATGGTAG	285.	369	C
HBB-TGTV-1	*HBB*	TGTC	ATCACTTAGACCTCACCCTG	286.	98	C
HBB-TGTV-2	*HBB*	TGTA	CTGATGGTATGGGGCCAAGA	287.	203	C
HBB-TGTV-4	*HBB*	TGTA	GATGGATCTCTTCCTGCGTC	288.	393	T
HBB-TTCV-1	*HBB*	TTCA	AACCCTCAGCCCTCCCTCTA	289.	167	T

[0168]

crRNA ID	基因名称	4nt PAM	20nt 间隔子	SEQ ID NO:	与 TSS 的距离 (bp)‡	链
HBB-TTCV-2	*HBB*	TTCC	AAATATTACGTAAATACACT	290.	254	C
HBB-TTCV-3	*HBB*	TTCA	GCTTTGGGATATGTAGATGG	291.	378	T
HBB-CTCC-1	*HBB*	CTCC	CTGCTCCTGGGAGTAGATTG	292.	51	T
HBB-CTCC-2	*HBB*	CTCC	CTCTAAGATATATCTCTTGG	293.	183	T
HBB-CTCC-3	*HBB*	CTCC	AGAATATGCAAAATACTTAC	294.	417	T
HBB-TACC-1	*HBB*	TACC	TGTCCTTGGCTCTTCTGGCA	295.	126	T
HBB-TACC-2	*HBB*	TACC	ATCAGTACAAATTGCTACTA	296.	212	T
HBB-TACC-3	*HBB*	TACC	ATAATTCAGCTTTGGGATAT	297.	370	T
NPY1R-TTTV-1	*NPY1R*	TTTC	AAGCCTCGGGAAACTGCCCT	298.	256	C
NPY1R-TTTV-2	*NPY1R*	TTTC	TTTGTTTGCAGGTCAGTGCC	299.	299	T
NPY1R-TTTV-3	*NPY1R*	TTTG	GGCTGGCGCTCGAGCTCTCC	300.	350	C
NPY1R-ATTV-1	*NPY1R*	ATTC	CTGGTTTGGGCTGGCGCTCG	301.	382	C
NPY1R-ATTV-2	*NPY1R*	ATTA	GTGCCATTATTGTGGCGAAT	302.	407	C
NPY1R-ATTV-3	*NPY1R*	ATTC	TCGGCACTGGCGTGAGAGTT	303.	464	C
NPY1R-CTTV-1	*NPY1R*	CTTC	CCCGGAGTCGAGGACTGTGG	304.	230	C
NPY1R-CTTV-2	*NPY1R*	CTTC	GGCCACAAGATGGCACTGAC	305.	314	C
NPY1R-CTTV-3	*NPY1R*	CTTA	TAAAGTGAGGAAAACAAATT	306.	485	C
NPY1R-TTCV-1	*NPY1R*	TTCC	CCGGAGTCGAGGACTGTGGG	307.	229	C
NPY1R-TTCV-2	*NPY1R*	TTCG	GCCACAAGATGGCACTGACC	308.	313	C
NPY1R-TTCV-3	*NPY1R*	TTCC	CAGCGAGCCCTTTGATTCCT	309.	376	C
NPY1R-CTCC-1	*NPY1R*	CTCC	GGGGAAGGCAGGGCAGTTTC	310.	243	T
NPY1R-CTCC-2	*NPY1R*	CTCC	AGCCGGGTATGACTTCGGCC	311.	330	C
NPY1R-CTCC-3	*NPY1R*	CTCC	TTTCTTTGGCCCACTGAGAA	312.	554	T
VEGFA-TTTV-1	*VEGFA*	TTTC	AGGCTGTGAACCTTGGTGGG	313.	200	C
VEGFA-TTTV-2	*VEGFA*	TTTC	CTGCTCCCTCCTCGCCAATG	314.	274	C
VEGFA-TTTV-3	*VEGFA*	TTTG	CTAGGAATATTGAAGGGGGC	315.	338	T
VEGFA-ATTV-1	*VEGFA*	ATTG	CGGCGGGCTGCGGGCCAGGC	316.	159	C
VEGFA-ATTV-2	*VEGFA*	ATTA	CCCATCCGCCCCCGGAAACT	317.	274	T
VEGFA-ATTV-3	*VEGFA*	ATTC	CTAGCAAAGAGGGAACGGCT	318.	326	C
VEGFA-CTTV-1	*VEGFA*	CTTC	CCCTTCATTGCGGCGGGCTG	319.	114	T
VEGFA-CTTV-2	*VEGFA*	CTTC	CCCTTCATTGCGGCGGGCTG	320.	169	C
VEGFA-CTTV-3	*VEGFA*	CTTC	CCCTGCCCCCTTCAATATTC	321.	346	C
VEGFA-GTTV-1	*VEGFA*	GTTC	ACAGCCTGAAAATTACCCAT	322.	209	T
VEGFA-GTTV-2	*VEGFA*	GTTA	CGTGCGGACAGGGCCTGAGA	323.	303	T
VEGFA-GTTV-3	*VEGFA*	GTTG	GAGCGGGGAGAAGGCCAGGG	324.	435	C
VEGFA-TTCV-1	*VEGFA*	TTCC	ACACGCGGCTCGGGCCCGGG	325.	115	T
VEGFA-TTCV-2	*VEGFA*	TTCA	GGCTGTGAACCTTGGTGGGG	326.	199	C
VEGFA-TTCV-3	*VEGFA*	TTCC	TGCTCCCTCCTCGCCAATGC	327.	213	C
VEGFA-TTCV-4	*VEGFA*	TTCC	CCTTCATTGCGGCGGGCTGC	328.	185	C
VEGFA-TTCV-5	*VEGFA*	TTCC	CCTGCCCCCTTCAATATTCC	329.	362	C
VEGFA-TCCC-1	*VEGFA*	TCCC	CTTCATTGCGGCGGGCTGCG	330.	167	C
VEGFA-TCCC-2	*VEGFA*	TCCC	TCCTCGCCAATGCCCCGCGG	331.	266	C

[0169]

crRNA ID	基因名称	4nt PAM	20nt 间隔子	SEQ ID NO:	与 TSS 的距离(bp)‡	链
VEGFA-TCCC-3	*VEGFA*	TCCC	CTGCCCCCTTCAATATTCCT	332.	344	C
VEGFA-CTCC-1	*VEGFA*	CTCC	TCGCCAATGCCCCGCGGGCG	333.	263	C
VEGFA-CTCC-2	*VEGFA*	CTCC	CTCCTCGCCAATGCCCCGCG	334.	267	C
VEGFA-CTCC-3	*VEGFA*	CTCC	AGGATTCCAATAGATCTGTG	335.	407	C

[0170] C,编码;T,模板;‡,对于模板链指导,从TSS到PAM的-4位置测量,或对于编码链指导,到间隔子的第20个nt

[0171] 表2C:用于VEGFA基因激活实验的SpCas9 sgRNA

[0172]

指导 ID	3nt PAM	20nt 间隔子	SEQ ID NO:	与 TSS 的距离(bp)‡	链
VEGFA-NGG-a-1	AGG	GTGTGCAGACGGCAGTCACT	336.	571	编码
VEGFA-NGG-a-2	AGG	GAGCAGCGTCTTCGAGAGTG	337.	509	编码
VEGFA-NGG-a-3	TGG	GGTGAGTGAGTGTGTGCGTG	338.	469	编码
VEGFA-NGG-b-4	AGG	GGGGCGGATGGGTAATTTTC	339.	217	编码
VEGFA-NGG-b-5	AGG	GGCATTGGCGAGGAGGGAGC	340.	272	模板
VEGFA-NGG-b-6	AGG	GCAAAGAGGGAACGGCTCTC	341.	320	编码

[0173] ‡,对于编码链指导,从TSS到PAM的-3位置测量,或对于模板链到间隔子的第20个nt

[0174] 表2D:用于碱基编辑子实验的Cas12a crRNA

[0175]

crRNA ID	基因名称	4nt PAM	20nt 间隔子	SEQ ID NO:
TTTA-3	*FANCF*	TTTA	TCCGTGTTCCTTGACTCTGG	342.
TTTC-1	*DNMT1*	TTTC	CCTCACTCCTGCTCGGTGAA	343.
TTTC-3	*EMX1*	TTTC	TCATCTGTGCCCCTCCCTCC	344.
TTTC-6	*RUNX1*	TTTC	AGCCTCACCCCTCTAGCCCT	345.
TTTC-7	*RUNX1*	TTTC	TTCTCCCCTCTGCTGGATAC	346.
TTTC-8	*FANCF*	TTTC	CGAGCTTCTGGCGGTCTCAA	347.
TTTG-4	*EMX1*	TTTG	TGGTTGCCCACCCTAGTCAT	348.
TTTG-7	*FANCF*	TTTG	GTCGGCATGGCCCCATTCGC	349.

[0176] 表3A.本研究中使用的寡核苷酸-用于T7E1和RFLP实验

[0177]

描述	序列	SEQ ID NO:
扩增人类细胞中 *DNMT1* 基因座的正向 PCR 引物	CCAGAATGCACAAAGTACTGCAC	350.
扩增人类细胞中 *DNMT1* 基因座的反向 PCR 引物	GCCAAAGCCCGAGAGAGTGCC	351.
扩增人类细胞中 *CFTR* 基因座的正向 PCR 引物	GCTGTGTCTGTAAACTGATGGCTAACA	352.
扩增人类细胞中 *CFTR* 基因座的反向 PCR 引物	TTGCATTCTACTCAATTGCATTCTGTGGG	353.
扩增人类细胞中 *EMX1* 基因座的正向 PCR 引物	GGAGCAGCTGGTCAGAGGGG	354.
扩增人类细胞中 *EMX1* 基因座的反向 PCR 引物	CCATAGGGAAGGGGGACACTGG	355.
扩增人类细胞中 *EMX1*(扩增子 2)基因座的正向 PCR 引物	CTGCCTCCTATTCATACACACTTACGGG	356.

描述	序列	SEQ ID NO:
扩增人类细胞中 *EMX1*(扩增子 2)基因座的反向 PCR 引物	CTCTGTTGGTGGAAACTCCCTGACC	357.
扩增人类细胞中 *FANCF* 基因座的正向 PCR 引物	GGGCCGGGAAAGAGTTGCTG	358.
扩增人类细胞中 *FANCF* 基因座的反向 PCR 引物	GCCCTACATCTGCTCTCCCTCC	359.
扩增人类细胞中 *RUNX1* 基因座的正向 PCR 引物	CCAGCACAACTTACTCGCACTTGAC	360.
扩增人类细胞中 *RUNX1* 基因座的反向 PCR 引物	CATCACCAACCCACAGCCAAGG	361.
扩增人类细胞中 *VEGFA* 基因座的正向 PCR 引物	CAGCTCCACAAACTTGGTGCCAAATTC	362.
扩增人类细胞中 *VEGFA* 基因座的反向 PCR 引物	CCGCAATGAAGGGGAAGCTCGAC	363.
[0178] 扩增人类细胞中 *VEGFA*(扩增子 2)基因座的正向 PCR 引物	CGCTGTTCAGGTCTCTGCTAGAAGTAGG	364.
扩增人类细胞中 *VEGFA*(扩增子 2)基因座的反向 PCR 引物	CCAGACCAGAGACCACTGGGAAG	365.
扩增人类细胞中匹配的位点 1 基因座的正向 PCR 引物	GACAAATGTATCATGCTATTATAAGATGTTGAC	366.
扩增人类细胞中匹配的位点 1 基因座的反向 PCR 引物	CCATTTACTGAGAGTAATTATAATTGTGC	367.
扩增人类细胞中匹配的位点 5 基因座的正向 PCR 引物	CCAAGGACAGGAATATCTTATACCCTCTGT	368.
扩增人类细胞中匹配的位点 5 基因座的反向 PCR 引物	TGTCATTGTCCTTGTCCTTTAGCTACCG	369.

[0179] 表3B.本研究中使用的寡核苷酸-用于PAM测定分析(PAMDA)和其他体外切割实验

描述	序列	SEQ ID NO:
扩增随机化 PAM 基因座的反向 PCR 引物	CAAAACAGCCAAGCTTGCATGC	370.
扩增随机化 PAM 基因座的正向 PCR 引物，添加 CCAT 条码	AGCTGCCATCGGTATTTCACACCGCATACGTAC	371.
扩增随机化 PAM 基因座的正向 PCR 引物，添加 GCAA 条码	AGCTGGCAACGGTATTTCACACCGCATACGTAC	372.
扩增随机化 PAM 基因座的正向 PCR 引物，添加 ATGC 条码	AGCTGATGCCGGTATTTCACACCGCATACGTAC	373.
扩增随机化 PAM 基因座的正向 PCR 引物，添加 GATG 条码	AGCTGGATGCGGTATTTCACACCGCATACGTAC	374.
[0180] 扩增随机化 PAM 基因座的正向 PCR 引物，添加 CGAT 条码	AGCTGCGATCGGTATTTCACACCGCATACGTAC	375.
将要克隆入 p11-lacY-wtx1 的 EcoRI/SphI 的、NNNNNNNN PAM 消耗文库间隔子 1 的顶链(top strand)寡聚物	AGACCGGAATTCNNNGTNNNNNNNNNNGGAATCCCTTCTGCAGCACCTGGGCGCAGGTCACGAGGCATG	376.
将要克隆入 p11-lacY-wtx1 的 EcoRI/SphI 的、NNNNNNNN PAM 消耗文库间隔子 2 的顶链寡聚物	AGACCGGAATTCNNNGTNNNNNNNNNNCTGATGGTCCATGTCTGTTACTCGCGCAGGTCACGAGGCATG	377.
填入文库寡聚物的反向引物	/5Phos/CCTCGTGACCTGCGC	378.
将要克隆入 p11-lacY-wtx1 的 EcoRI/SphI 的、	AATTCTTTAGGAATCCCTTCTGCAGCACCT	379.

[0181]

描述	序列	SEQ ID NO:
具有 TTTA PAM 靶点的间隔子 1 的顶链	GGGCATG	
将要克隆入 p11-lacY-wtx1 的 EcoRI/SphI 的、具有 TTTA PAM 靶点的间隔子 1 的底链	CCCAGGTGCTGCAGAAGGGATTCCTAAAG	380.
将要克隆入 p11-lacY-wtx1 的 EcoRI/SphI 的、具有 CTTA PAM 靶点的间隔子 1 的顶链	AATTCCTTAGGAATCCCTTCTGCAGCACCTGGGCATG	381.
将要克隆入 p11-lacY-wtx1 的 EcoRI/SphI 的、具有 CTTA PAM 靶点的间隔子 1 的底链	CCCAGGTGCTGCAGAAGGGATTCCTAAGG	382.
将要克隆入 p11-lacY-wtx1 的 EcoRI/SphI 的、具有 ACCT PAM 靶点的间隔子 1 的顶链	AATTCACCTGGAATCCCTTCTGCAGCACCTGGGCATG	383.
将要克隆入 p11-lacY-wtx1 的 EcoRI/SphI 的、具有 ACCT PAM 靶点的间隔子 1 的底链	CCCAGGTGCTGCAGAAGGGATTCCAGGTG	384.
将要克隆入 p11-lacY-wtx1 的 EcoRI/SphI 的、具有 TTTA PAM 靶点的间隔子 2 的顶链	AATTCTTTACTGATGGTCCATGTCTGTTACTCGCATG	385.
将要克隆入 p11-lacY-wtx1 的 EcoRI/SphI 的、具有 TTTA PAM 靶点的间隔子 2 的底链	CGAGTAACAGACATGGACCATCAGTAAAG	386.
将要克隆入 p11-lacY-wtx1 的 EcoRI/SphI 的、具有 CTTA PAM 靶点的间隔子 2 的顶链	AATTCCTTACTGATGGTCCATGTCTGTTACTCGCATG	387.
将要克隆入 p11-lacY-wtx1 的 EcoRI/SphI 的、具有 CTTA PAM 靶点的间隔子 2 的底链	CGAGTAACAGACATGGACCATCAGTAAGG	388.
将要克隆入 p11-lacY-wtx1 的 EcoRI/SphI 的、具有 ACCT PAM 靶点的间隔子 2 的顶链	AATTCACCTCTGATGGTCCATGTCTGTTACTCGCATG	389.
将要克隆入 p11-lacY-wtx1 的 EcoRI/SphI 的、具有 ACCT PAM 靶点的间隔子 2 的底链	CGAGTAACAGACATGGACCATCAGAGGTG	390.

[0182] 表3C.本研究中使用的寡核苷酸—用于激活子RT-qPCR实验

[0183]

描述	序列	SEQ ID NO:
人类 *NPY1R* 基因的正向 RT-qPCR 引物	ATGGTGAGCAGAGTGCCCTATC	391.
人类 *NPY1R* 基因的反向 RT-qPCR 引物	ATGGTCCCTGGCAGTCTCCAAA	392.
人类 *AR* 基因的正向 RT-qPCR 引物	CCATCGGACTCTCATAGGTTGTC	393.
人类 *AR* 基因的反向 RT-qPCR 引物	GACCTGTACTTATTGTCTCTCATC	394.
人类 *HBB* 基因的正向 RT-qPCR 引物	GCACGTGGATCCTGAGAACT	395.
人类 *HBB* 基因的反向 RT-qPCR 引物	ATTGGACAGCAAGAAAGCGAG	396.
人类 *HPRT1* 基因的正向 RT-qPCR 引物	CATTATGCTGAGGATTTGGAAAGG	397.
人类 *HPRT1* 基因的反向 RT-qPCR 引物	CTTGAGCACACAGAGGGCTACA	398.

[0184] 表3D.本研究中使用的寡核苷酸—用于碱基编辑子深度测序实验

[0185]

描述	序列	SEQ ID NO:
扩增人类细胞中 TTTA PAM 位点 3 的正向 PCR 引物	GCTCCAGAGCCGTGCGAATGG	399.
扩增人类细胞中 TTTA PAM 位点 3 的反向 PCR 引物	GCCCTACATCTGCTCTCCCTCC	400.
扩增人类细胞中 TTTC PAM 位点 1 的正向 PCR 引	CAGCTGACCCAATAAGTGGCAGAGTG	401.

[0186]

描述	序列	SEQ ID NO:
物		
扩增人类细胞中 TTTC PAM 位点 1 的反向 PCR 引物	TCAGGTTGGCTGCTGGGCTGG	402.
扩增人类细胞中 TTTC PAM 位点 3 的正向 PCR 引物	CCCCAGTGGCTGCTCTGGG	403.
扩增人类细胞中 TTTC PAM 位点 3 的反向 PCR 引物	CATCGATGTCCTCCCCATTGGC	404.
扩增人类细胞中 TTTC PAM 位点 6 的正向 PCR 引物	GCTGTCTGAAGCCATCGCTTCC	405.
扩增人类细胞中 TTTC PAM 位点 6 的反向 PCR 引物	CAGAGGTATCCAGCAGAGGGGAG	406.
扩增人类细胞中 TTTC PAM 位点 7 的正向 PCR 引物	CCTTCGGAGCGAAAACCAAGACAG	407.
扩增人类细胞中 TTTC PAM 位点 7 的反向 PCR 引物	CAGGCAGGACGAATCACACTGAATG	408.
扩增人类细胞中 TTTC PAM 位点 8 的正向 PCR 引物	GCTCCAGAGCCGTGCGAATGG	409.
扩增人类细胞中 TTTC PAM 位点 8 的反向 PCR 引物	GCACCTCATGGAATCCCTTCTGC	410.
扩增人类细胞中 TTTG PAM 位点 4 的正向 PCR 引物	GAAGCTGGAGGAGGAAGGGC	411.
扩增人类细胞中 TTTG PAM 位点 4 的反向 PCR 引物	CAGCAGCAAGCAGCACTCTGC	412.
扩增人类细胞中 TTTG PAM 位点 7 的正向 PCR 引物	GCCCTCTTGCCTCCACTGGTTG	413.
扩增人类细胞中 TTTG PAM 位点 7 的反向 PCR 引物	CCAATAGCATTGCAGAGAGGCGT	414.

[0187] 细胞培养条件

[0188] 人类U2OS(来自Toni Cathomen,Freiburg)和HEK293细胞(Invitrogen)分别在补充有10％热活灭FBS、1％青霉素和链霉素以及2mM GlutaMax的高级杜氏改良伊格尔培养基(A-DMEM)和DMEM中培养(例外是用于RT-qPCR分析实验而培养的HEK293细胞使用含有0.1％青霉素和链霉素、缺乏GlutaMax的培养基)。所有细胞培养试剂购自Life Technologies,细胞在5％CO_2中在37℃下生长。每两周分析培养基上清液中支原体的存在,细胞系身份通过STR分布来确认(ATCC)。除非另有陈述,阴性对照转染包括Cas12a表达和U6-null质粒。

[0189] 通过T7E1或深度测序的基因和碱基编辑评估

[0190] 对于核酸酶和碱基编辑子实验,Cas12a和crRNA表达质粒(分别为500ng和250ng)使用4D-Nucleofector核转染仪(Lonza)用SE细胞系核转染(Cell Line Nucleofector)试剂盒通过DN-100程序电穿孔到大约2×10^5个U2OS细胞中。使用Agencourt DNAdvance核酸分离试剂盒(Beckman Coulter),或通过常规切割和顺磁性珠子提取,在核染后大约72或120小时提取基因组DNA(gDNA)(分别用于核酸酶或碱基编辑实验)。顺磁性珠子按照早先描述的类似地制备(Rohland等人,Genome Res.,2012,22:939-46)(GE Healthcare Sera-Mag

SpeedBeads(Fisher Scientific),在0.1×TE中洗涤,悬浮在20%PEG-8000(w/v)、1.5M NaCl、10mM Tris-HCl pH 8、1mM EDTA pH 8和0.05%Tween20中)。对于细胞切割,除去培养基上清液,进行500μL PBS洗涤,细胞在55℃下用200μL裂解缓冲液(100mM Tris HCl pH 8.0,200mM NaCl,5mM EDTA,0.05%SDS,1.4mg/ml蛋白酶K(NEB)和12.5mM DTT)处理12-20小时。为了提取gDNA,溶胞产物与165μL顺磁性珠子组合,彻底混合,孵育5分钟,在磁性平板上分离,用70%EtOH洗涤3次,容许干燥5分钟,在65μL洗脱缓冲液(1.2mM Tris-HCl pH 8.0)中洗脱。基因组基因座使用100ng gDNA作为模板和表3中列出的引物,用Phusion Hot Start Flex DNA聚合酶(New England Biolabs;NEB)通过PCR扩增。在QIAxcel毛细管电泳仪(Qiagen)上分析之后,使用顺磁性珠子纯化PCR产物。

[0191] 对于核酸酶实验,内源人类目标位点的修饰百分比通过T7核酸内切酶(T7EI)分析进行测定,与早先描述的类似(Reyon等人,Nat Biotechnol.,2012,30:460-5)。简要地说,将200ng纯化的PCR产物变性、退火,并在37℃下用10U T7EI(NEB)消化25分钟。使用顺磁性珠子纯化消化物,使用QIAxcel进行分析来估计目标位点修饰。

[0192] 对于碱基编辑实验,基本上如早先描述的(Kleinstiver等人,Nature,2016,529:490-5)进行靶向深度测序。使用KAPA HTP文库制备试剂盒(KAPA BioSystems)从纯化并集中的PCR产物中产生双重索引的Tru-seq文库,在Illumina MiSeq测序仪上测序。样品测序的平均读取结果计数为55,000,最小8,500个读取结果。使用CRISPResso的修改版本(Pinello等人,Nat Biotechnol.,2016,34:695-7)分析核苷酸取代以及插入或删除突变(indel),进行另外的常规分析来检查在涵盖每个目标位点的-14～+30区域的44nt窗口(4nt PAM上游的另外10nt,以及20nt间隔子序列下游的10nt)中的插入或删除突变百分比(定义为[修饰的读取结果-仅取代的读取结果]/总读取结果*100)。

[0193] GUIDE-seq

[0194] 如早先描述的(Tsai等人,Nat Biotechnol.,2015,33:187-197)进行GUIDE-Seq实验。简要地说,如上文描述的对U2OS细胞进行电穿孔,但是包括100pmol的双链寡聚脱氧核苷酸(dsODN)GUIDE-Seq标签。分别进行限制片段长度多态性(RFLP)分析(如早先描述的进行,Kleinstiver等人,Nature,2015,523:481-5)和T7E1分析(如上所述)来评估GUIDE-Seq标签整合和中靶修饰百分比。使用Illumina MiSeq测序仪对GUIDE-seq文库进行测序,在下游数据处理之前使用guideseq(Tsai等人,Nat Biotechnol.,2016,34:483)v1.1以75bp窗口并允许达到9个错配对数据进行分析。使用SAMTools鉴定了高置信度、细胞类型特异性的单核苷酸多态性(SNP)。

[0195] 基因激活实验

[0196] 对于使用靶向VEGFA启动子的crRNA或sgRNA的实验,每孔1.6×10^5个HEK293细胞接种到24孔平板中,大约24小时之后用编码Cas12a或Cas9激活子的质粒以及crRNA或sgRNA的库(分别750ng和250ng)、1.5μL TransIT-LT1(Mirus)和Opti-MEM达到总体积50μL进行转染。转染后22小时改变细胞培养基,在转染后44小时采集培养基上清液的等分试样,以使用人类VEGFQuantikine ELISA试剂盒(R&D Systems)确定VEGFA浓度。

[0197] 对于使用靶向AR、HBB或NPY1R启动子的crRNA的实验,每孔8.6×10^4个HEK293细胞接种到12孔平板中,大约24小时之后用750ng Cas12a激活子表达质粒、250ng crRNA质粒库、3μL TransIT-LT1(Mirus)和100μL Opti-MEM转染。转染后72小时使用NucleoSpin RNA

Plus试剂盒(Clontech)从转染的细胞提取总RNA。使用250ng纯化的RNA使用High-Capacity RNA-to-cDNA试剂盒(ThermoFisher)进行cDNA合成,3μL 1:20稀释的cDNA使用Fast SYBR Green Master Mix(ThermoFisher)和表3中列出的引物通过定量逆转录PCR(RT-qPCR)进行扩增。在LightCycler480(Roche)上使用以下循环程序进行RT-qPCR反应:95℃初始变性20秒,随后是95℃3秒和60℃30秒的45个循环。如果样品扩增在35个循环之后没有达到检测阈值,由于极低水平表达的转录产物典型的Ct(阈值循环)波动,Ct值被认为是35。将相对于阴性对照实验(Cas12a激活子和空的crRNA质粒)的基因表达水平归一化为HPRT1的表达。

[0198] Cas12a蛋白的表达和纯化

[0199] 编码Cas12a-SV40NLS-6xHis融合蛋白的质粒转化入Rosetta 2(DE3)大肠杆菌,将单个菌落接种到含有50mg/L卡那霉素和25mg/L氯霉素(Kan/Cm)的25mL LB培养基中,之后在25℃下生长16小时。然后将起始培养物1:100稀释到含有Kan/Cm的150mL LB培养基中,在37℃下生长直到OD_{600}达到0.4。培养物然后用0.2mM异丙基β-D-硫代半乳糖苷进行诱导,之后在18℃下摇动23小时。通过1200g下离心15分钟从50mL的培养物中收获细胞团粒,悬浮在含有20mM Hepes(羟乙基哌嗪乙硫磺酸)pH 7.5、100mM KCl、5mM $MgCl_2$、5%甘油、1mM DTT、Sigmafast蛋白酶抑制物(Sigma-Aldrich)和0.1%Triton X-100的1mL裂解缓冲液中。细胞悬浮液装入1mL AFA纤维milliTUBE(Covaris),根据以下条件使用E220evolution聚焦超声发生器(Covaris)进行裂解:峰值强度功率150W,每次猝发200个循环,占空比为10%,在5℃下处理20分钟。细胞溶胞产物在21,000g和4℃下离心20分钟,上清液与等体积的结合缓冲液(裂解缓冲液+10mM咪唑)混合,添加到在结合缓冲液中预平衡的400μL HisPur Ni-NTA树脂(Thermo Fisher Scientific),在4℃下摇动8小时。结合了蛋白的树脂用1mL洗涤缓冲液(20mM Hepes pH 7.5,500mM KCl,5mM $MgCl_2$,5%甘油,25mM咪唑和0.1%Triton X-100)洗涤三次,用1mL结合缓冲液洗涤一次,然后使用500μL洗脱缓冲液(20mM Hepes pH 7.5,100mM KCl,5mM $MgCl_2$,10%甘油和500mM咪唑)进行三次连续洗脱。将选择洗脱物集中,使用Spectra/Por 4标准级纤维素透析袋(Standard Cellulose Dialysis Tubing)(Spectrum Chemical Manufacturing Corp)在三个连续的1:500缓冲液交换中进行透析,前两次在透析缓冲液(300mM NaCl,10mM Tris-HCl pH 7.4,0.1mM EDTA和1mM DTT)中,最后一次在含有20%甘油的透析缓冲液中。然后用Amicon Ultra-0.5mL离心式过滤器(Centrifugal Filter Units)(Millipore Sigma)浓缩蛋白,用具有80%甘油的等体积的透析缓冲液稀释,在-20℃下保存。

[0200] 体外切割反应

[0201] 37℃下使用T7 RiboMAX快速大量RNA制备试剂盒(Express Large Scale RNA Production kit)(Promega)从大约1μg HindIII线性化的crRNA转录质粒中体外转录Cas12a crRNA16小时。通过添加1μL RQ1 DNase在37℃下降解DNA模板15分钟,随后用miRNeasy Mini试剂盒(Qiagen)纯化RNA。体外切割反应由裂解缓冲液(10mM Hepes pH 7.5,150mM NaCl和5mM $MgCl_2$)中25nM PvuI线性化的底物质粒、300nM crRNA和200nM纯化的Cas12a蛋白组成,在37℃下进行,除非另有陈述。温度耐受性分析的质粒底物编码具有TTTA PAM的PAMDA位点2间隔子。制备切割反应主混合物,然后等分为每时间点5μL体积,在热循环仪中孵育,通过添加10μL终止缓冲液(0.5%SDS,50mM EDTA)来停止。使用顺磁性珠子纯化停止的等分试样,通过QIAxcel ScreenGel软件(v1.4)定量切割百分比。

[0202] PAM测定分析

[0203] 编码具有随机序列的目标位点的质粒文库使用Klenoq(-exo)(NEB)进行克隆,以填入两个独立的寡聚物的底链,所述寡聚物带有两个不同间隔子序列5'的10nt随机序列(表3)。双链产物使用EcoRI消化,连接到EcoRI和SphI消化的p11-lacY-wtx1中(Addgene质粒69056;Huimin Zhao赠送)。连接物转化入电感受态XL1 Blue大肠杆菌中,在9mL SOC中在37℃下回收1小时,然后在具有100mg/L羧苄西林的150mL LB培养基中生长16小时。基于观察到的转化体的数量,每个文库的复杂度估计为大于10^6。

[0204] 如上所述进行随机PAM质粒文库的切割反应,等分试样在3、6、12、24和48分钟时停止。使用磁性珠子纯化反应,大约1-5ng用作模板,用于未切割的分子的PCR扩增,使用Phusion Hot Start Flex DNA聚合酶(NEB)15个循环。在PCR反应期间,将4nt唯一的分子索引(UMI)添加到PAM的上游,以允许时间点样品的去多重化,还从未消化的质粒产生产物,以确定文库中的初始PAM表示。使用QuantiFluor dsDNA系统(Promega)定量纯化的PCR产物,归一化,集中,用于使用KAPA HTP PCR-free文库制备试剂盒(KAPA BioSystems)用Illumina双重索引的衔接子进行文库制备。使用Universal KAPA Illumina Library qPCR定量试剂盒(KAPA Biosystems)对文库进行定量,使用300-cycle v2试剂盒(Illumina)在Illumina MiSeq测序仪上测序。

[0205] 使用常规的Python脚本分析测序读取结果,以估计给定蛋白对每种PAM的切割速率。通过Phred评分(≥Q30)过滤配对的末端读取结果,然后以时间点UMI、PAM和间隔子序列完美匹配的要求进行合并。对于所有时间点、蛋白和间隔子,对每4nt和5nt PAM产生计数。然后对于测序深度中的样品间差异(inter-sample difference)校正PAM计数,转换成原始质粒文库中该PAM的初始表示(通过未消化的对照确定的)的分数,然后归一化,以说明由于切割的底物的消耗而造成的随时间的未切割底物提高的分数表示(通过在所有时间点中选择具有最高平均计数的5个PAM来代表不可切割的底物的分布)。然后通过线性最小二乘回归将每种PAM随时间的消耗拟合于指数衰减模型(y(t)=Aeˆ(-kt),其中y(t)是归一化的PAM计数,t是时间(分钟),k是速率常数,A是常数)。

[0206] 靶向范围计算

[0207] 使用GENCODE's Release 27GTF文件,在各种注释的基因组元件上评估野生型和变体AsCas12a核酸酶的靶向范围。在涵盖所有转录起始位点(TSS)上游1kb的区域内、所有基因的第一个外显子内、以及所有注释的miRNA内,对可靶向的4nt PAM的完全出现进行枚举。GTF文件中每个元件的参数值是:Exon1,feature-type exon,exon_number 1,gene_type protein_coding;TSS,feature-type transcript,gene_type protein_codingor miRNA;miRNA,feature-type gene,gene_type miRNA。对于每个元件,通过长度对PAM计数进行归一化,通过boxplot可视化。PAM鉴定和枚举脚本将根据要求提供。Cas12a核酸酶可靶向的PAM包括:TTTV,野生型AsCas12a;TTYN、RTTC、CTTV、TATM、CTCC、TCCC、TACA(层级1)和RTTS、TATA、TGTV、ANCC、CVCC、TGCC、GTCC、TTAC(层级2)PAM,eAsCas12a(参见图1g和扩展数据图5h);TATV,AsCas12a-RVR;和TYCV,AsCas12a-RR。

[0208] DNA检测分析

[0209] 通过在4℃下孵育500nM纯化的AsCas12a蛋白和750nM化学合成的crRNA(IDT)5分钟形成Cas12a-crRNA RNP复合物。所有反应在10mM Hepes pH 7.5、150mM NaCl和5mM

$MgCl_2$中进行。接下来，RNP稀释到100nM并与250nM线性化的活化质粒DNA在15μL的总体积中混合，所述质粒DNA带有与典型(TTTA)或非典型(CTTA或ACCT)PAM匹配的目标位点或不相关的目标位点(阴性对照)。容许反应在37℃下进行30分钟，之后在4℃下孵育。对于荧光检测，然后将10μL的RNP/目标-DNA反应与等体积的100nM常规荧光淬灭报告子(IDT)在小体积384孔黑色平板(Corning)中混合。检测反应在37℃下进行3小时，使用Synergy HTX微量培养板读取器(BTX)用λ_{ex}＝485nm和λ_{em}＝528nm每60秒进行测量。

[0210] 序列

[0211] 以下的实施例中使用以下构建体。

[0212] BPK3079:U6-AsCas12a-crRNA-BsmBI盒(cassette)

[0213] U6启动子为黑色，AsCas12a crRNA为斜体，U6转录所需的鸟嘌呤为粗体，间隔子进入盒是小写字母，BsmBI位点是双下划线的，U6终止子(**TTTTTTT**)是**粗体双下划线的**。

[0214] TGTACAAAAAAGCAGGCTTTAAAGGAACCAATTCAGTCGACTGGATCCGGTACCAAGGTCGGGCAGG
AAGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTAG
AATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGG
GTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTC
GATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACC**G***TAATTTCTACTCTTGTAGAT*ggagacgattaat
gcgtctcc**TTTTTTT**(SEQ ID NO: 415)

[0215] BPK3082:U6-LbCas12a-crRNA-BsmBI盒

[0216] U6启动子为绿色，LbCas12a crRNA为斜体有色，U6转录所需的鸟嘌呤为粗体，间隔子进入盒是小写字母，BsmBI位点是双下划线的，U6终止子(**TTTTTTT**)是**粗体双下划线的**。

[0217] TGTACAAAAAAGCAGGCTTTAAAGGAACCAATTCAGTCGACTGGATCCGGTACCAAGGTCGGGCAGG
AAGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTAG
AATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGG
GTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTC
GATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACC**G***AATTTCTACTAAGTGTAGAT*ggagacgattaat
gcgtctcc**TTTTTTT**(SEQ ID NO: 416)

[0218] BPK4446:U6-FnCas12a-crRNA-BsmBI盒

[0219] U6启动子为黑色，FnCas12a crRNA为斜体有色，U6转录所需的鸟嘌呤为粗体，间隔子进入盒是小写字母，BsmBI位点是双下划线的，U6终止子(**TTTTTTT**)是**粗体双下划线的**。

[0220] TGTACAAAAAAGCAGGCTTTAAAGGAACCAATTCAGTCGACTGGATCCGGTACCAAGGTCGGGCAGG
AAGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTAG
AATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGG
GTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTC
GATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACC**G***AATTTCTACTGTTGTAGAT*ggagacgattaatg
cgtctcc**TTTTTTT**(SEQ ID NO: 417)

[0221] BPK4449:U6-MbCas12a-crRNA-BsmBI盒

[0222] U6启动子为黑色，MbCas12a crRNA为斜体有色，U6转录所需的鸟嘌呤为粗体，间隔子进入盒是小写字母，BsmBI位点是双下划线的，U6终止子(**TTTTTTT**)是**粗体双下划线的**。

[0223] TGTACAAAAAAGCAGGCTTTAAAGGAACCAATTCAGTCGACTGGATCCGGTACCAAGGTCGGGCAGG AAGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTAG AATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGG GTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTC GATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACC**G***AATTTCTACTGTTTGTAGAT*ggagacgattaat gcgtctcc**TTTTTTT**(SEQ ID NO:418)

[0224] MSP3491:T7-AsCas12a-crRNA-BsaIcassette

[0225] T7启动子为黑色,T7转录所需的鸟嘌呤为粗体,AsCas12a crRNA为斜体,间隔子进入盒是小写字母,BsaI位点是双下划线的,用于线性化的DraI的限制位点(tttaaa)是粗体小写字母,HindIII(**AAGCTT**)是**粗体双下 划线的**

[0226] TAATACGACTCACTATA**G***TAATTTCTACTCTTGTAGAT*ggagacccatgccatagcgttgttcggaatatgaatttttgaacagattcacca acacctagtggtctcc**tttaaa****AAGCTT**(SEQ ID NO: 419)

[0227] MSP3495:T7-LbCas12a-crRNA-BsaI盒

[0228] T7启动子为黑色,T7转录所需的鸟嘌呤为粗体,LbCas12acrRNA为斜体,间隔子进入盒是小写字母,BsaI位点是双下划线的,用于线性化的DraI的限制位点(tttaaa)是粗体小写字母,HindIII(**AAGCTT**)是**粗体双下 划线的**:

[0229] TAATACGACTCACTATA**G***AATTTCTACTAAGTGTAGAT*ggagacccatgccatagcgttgttcggaatatgaatttttgaacagattcacca acacctagtggtctcc**tttaaa****AAGCTT**(SEQ ID NO:420)

[0230] pCAG-人类AsCpf1-NLS-3xHA的核苷酸序列

[0231] 人类密码子优化的AsCpf1为普通字体(NTs 1-3921),NLS为小写字母(aaaaggccg gcggccacgaaaaaggccggccaggcaaaaaagaaaaag,SEQ ID NO:21),3xHA标签(TACCCATACGATG TTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCC,SEQ ID NO:5)为粗体

[0232]

ATGACACAGTTCGAGGGCTTTACCAACCTGTATCAGGTGAGCAAGACACTGCGGTTTGAGCTGA
TCCCACAGGGCAAGACCCTGAAGCACATCCAGGAGCAGGGCTTCATCGAGGAGGACAAGGCCCGCAA
TGATCACTACAAGGAGCTGAAGCCCATCATCGATCGGATCTACAAGACCTATGCCGACCAGTGCCTGCA
GCTGGTGCAGCTGGATTGGGAGAACCTGAGCGCCGCCATCGACTCCTATAGAAAGGAGAAAACCGAGG
AGACAAGGAACGCCCTGATCGAGGAGCAGGCCACATATCGCAATGCCATCCACGACTACTTCATCGGCC
GGACAGACAACCTGACCGATGCCATCAATAAGAGACACGCCGAGATCTACAAGGGCCTGTTCAAGGCC
GAGCTGTTTAATGGCAAGGTGCTGAAGCAGCTGGGCACCGTGACCACAACCGAGCACGAGAACGCCC
TGCTGCGGAGCTTCGACAAGTTTACAACCTACTTCTCCGGCTTTTATGAGAACAGGAAGAACGTGTTCA
GCGCCGAGGATATCAGCACAGCCATCCCACACCGCATCGTGCAGGACAACTTCCCCAAGTTTAAGGAG
AATTGTCACATCTTCACACGCCTGATCACCGCCGTGCCCAGCCTGCGGGAGCACTTTGAGAACGTGAA
GAAGGCCATCGGCATCTTCGTGAGCACCTCCATCGAGGAGGTGTTTTCCTTCCCTTTTTATAACCAGCTG
CTGACACAGACCCAGATCGACCTGTATAACCAGCTGCTGGGAGGAATCTCTCGGGAGGCAGGCACCGA
GAAGATCAAGGGCCTGAACGAGGTGCTGAATCTGGCCATCCAGAAGAATGATGAGACAGCCCACATCA
TCGCCTCCCTGCCACACAGATTCATCCCCCTGTTTAAGCAGATCCTGTCCGATAGGAACACCCTGTCTTT
CATCCTGGAGGAGTTTAAGAGCGACGAGGAAGTGATCCAGTCCTTCTGCAAGTACAAGACACTGCTGA
GAAACGAGAACGTGCTGGAGACAGCCGAGGCCCTGTTTAACGAGCTGAACAGCATCGACCTGACACA
CATCTTCATCAGCCACAAGAAGCTGGAGACAATCAGCAGCGCCCTGTGCGACCACTGGGATACACTGA
GGAATGCCCTGTATGAGCGGAGAATCTCCGAGCTGACAGGCAAGATCACCAAGTCTGCCAAGGAGAAG
GTGCAGCGCAGCCTGAAGCACGAGGATATCAACCTGCAGGAGATCATCTCTGCCGCAGGCAAGGAGCT
GAGCGAGGCCTTCAAGCAGAAAACCAGCGAGATCCTGTCCCACGCACACGCCGCCCTGGATCAGCCA
CTGCCTACAACCCTGAAGAAGCAGGAGGAGAAGGAGATCCTGAAGTCTCAGCTGGACAGCCTGCTGG
GCCTGTACCACCTGCTGGACTGGTTTGCCGTGGATGAGTCCAACGAGGTGGACCCCGAGTTCTCTGCC
CGGCTGACCGGCATCAAGCTGGAGATGGAGCCTTCTCTGAGCTTCTACAACAAGGCCAGAAATTATGCC
ACCAAGAAGCCCTACTCCGTGGAGAAGTTCAAGCTGAACTTTCAGATGCCTACACTGGCCTCTGGCTG
GGACGTGAATAAGGAGAAGAACAATGGCGCCATCCTGTTTGTGAAGAACGGCCTGTACTATCTGGGCA
TCATGCCAAAGCAGAAGGGCAGGTATAAGGCCCTGAGCTTCGAGCCCACAGAGAAAACCAGCGAGGG
CTTTGATAAGATGTACTATGACTACTTCCCTGATGCCGCCAAGATGATCCCAAAGTGCAGCACCCAGCTG
AAGGCCGTGACAGCCCACTTTCAGACCCACACAACCCCCATCCTGCTGTCCAACAATTTCATCGAGCCT
CTGGAGATCACAAAGGAGATCTACGACCTGAACAATCCTGAGAAGGAGCCAAAGAAGTTTCAGACAG
CCTACGCCAAGAAAACCGGCGACCAGAAGGGCTACAGAGAGGCCCTGTGCAAGTGGATCGACTTCAC
AAGGGATTTTCTGTCCAAGTATACCAAGACAACCTCTATCGATCTGTCTAGCCTGCGGCCATCCTCTCAG
TATAAGGACCTGGGCGAGTACTATGCCGAGCTGAATCCCCTGCTGTACCACATCAGCTTCCAGAGAATC
GCCGAGAAGGAGATCATGGATGCCGTGGAGACAGGCAAGCTGTACCTGTTCCAGATCTATAACAAGGA
CTTTGCCAAGGGCCACCACGGCAAGCCTAATCTGCACACACTGTATTGGACCGGCCTGTTTTCTCCAGA
GAACCTGGCCAAGACAAGCATCAAGCTGAATGGCCAGGCCGAGCTGTTCTACCGCCCTAAGTCCAGGA
TGAAGAGGATGGCACACCGGCTGGGAGAGAAGATGCTGAACAAGAAGCTGAAGGATCAGAAAACCCC
AATCCCCGACACCCTGTACCAGGAGCTGTACGACTATGTGAATCACAGACTGTCCCACGACCTGTCTGA
TGAGGCCAGGGCCCTGCTGCCCAACGTGATCACCAAGGAGGTGTCTCACGAGATCATCAAGGATAGGC
GCTTTACCAGCGACAAGTTCTTTTTCCACGTGCCTATCACACTGAACTATCAGGCCGCCAATTCCCCATC
TAAGTTCAACCAGAGGGTGAATGCCTACCTGAAGGAGCACCCCGAGACACCTATCATCGGCATCGATCG
GGGCGAGAGAAACCTGATCTATATCACAGTGATCGACTCCACCGGCAAGATCCTGGAGCAGCGGAGCC
TGAACACCATCCAGCAGTTTGATTACCAGAAGAAGCTGGACAACAGGGAGAAGGAGAGGGTGGCAGC

[0233] AAGGCAGGCCTGGTCTGTGGTGGGCACAATCAAGGATCTGAAGCAGGGCTATCTGAGCCAGGTCATCC
ACGAGATCGTGGACCTGATGATCCACTACCAGGCCGTGGTGGTGCTGGAGAACCTGAATTTCGGCTTTA
AGAGCAAGAGGACCGGCATCGCCGAGAAGGCCGTGTACCAGCAGTTCGAGAAGATGCTGATCGATAA
GCTGAATTGCCTGGTGCTGAAGGACTATCCAGCAGAGAAAGTGGGAGGCGTGCTGAACCCATACCAGC
TGACAGACCAGTTCACCTCCTTTGCCAAGATGGGCACCCAGTCTGGCTTCCTGTTTTACGTGCCTGCCC
CATATACATCTAAGATCGATCCCCTGACCGGCTTCGTGGACCCCTTCGTGTGGAAAACCATCAAGAATCA
CGAGAGCCGCAAGCACTTCCTGGAGGGCTTCGACTTTCTGCACTACGACGTGAAAACCGGCGACTTCA
TCCTGCACTTTAAGATGAACAGAAATCTGTCCTTCCAGAGGGGCCTGCCCGGCTTTATGCCTGCATGGG
ATATCGTGTTCGAGAAGAACGAGACACAGTTTGACGCCAAGGGCACCCCTTTCATCGCCGGCAAGAGA
ATCGTGCCAGTGATCGAGAATCACAGATTCACCGGCAGATACCGGGACCTGTATCCTGCCAACGAGCTG
ATCGCCCTGCTGGAGGAGAAGGGCATCGTGTTCAGGGATGGCTCCAACATCCTGCCAAAGCTGCTGGA
GAATGACGATTCTCACGCCATCGACACCATGGTGGCCCTGATCCGCAGCGTGCTGCAGATGCGGAACTC
CAATGCCGCCACAGGCGAGGACTATATCAACAGCCCCGTGCGCGATCTGAATGGCGTGTGCTTCGACTC
CCGGTTTCAGAACCCAGAGTGGCCCATGGACGCCGATGCCAATGGCGCCTACCACATCGCCCTGAAGG
GCCAGCTGCTGCTGAATCACCTGAAGGAGAGCAAGGATCTGAAGCTGCAGAACGGCATCTCCAATCAG
GACTGGCTGGCCTACATCCAGGAGCTGCGCAACaaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaagGG
ATCC**TACCCATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATAT**
GATGTCCCCGACTATGCCTAA (SEQ ID NO:6)

[0234] AsCpf1-NLS-3xHA的氨基酸序列

[0235] AsCpf1为普通字体(AAs 1-1306),NLS(krpaatkkagqakkkk,SEQ ID NO:7)为小写字母,3xHA标签(YPYDVPDYAYPYDVPDYAYPYDVPDYA,SEQ ID NO:8)为粗体

[0236] MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTYADQCLQL
VQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKAELFNG
KVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITA
VPSLREHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKN
DETAHIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSIDLTHI
FISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEIISAAGKELSEAFKQK
TSEILSHAHAALDQPLPTTLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEPSL
SFYNKARNYATKKPYSVEKFKLNFQMPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFE
PTEKTSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKK
FQTAYAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIA
EKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKR
MAHRLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDK
FFFHVPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKK
LDNREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFKSKRTGIAEKAVYQQ
FEKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVW
KTIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIA
GKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIRSVLQMRNSN
AATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDW
LAYIQELRNkrpaatkkagqakkkkGS**YPYDVPDYAYPYDVPDYAYPYDVPDYA** (SEQ ID NO:9)

[0237] SQT1659:pCAG-hAsCas12a-NLS-3xHA

[0238] 人类密码子优化的氨基酸球菌种BV3L6 Cas12a(AsCas12a)为黑色,核质蛋白NLS(aaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaag,SEQ ID NO:21)为小写字母,接头序列为斜体,3xHA标签(TACCCATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCA TACCCATATGATGTCCCCGACTATGCC,SEQ ID NO:5)为粗体

[0239]

ATGACACAGTTCGAGGGCTTTACCAACCTGTATCAGGTGAGCAAGACACTGCGGTTTGAGCTG
ATCCCACAGGGCAAGACCCTGAAGCACATCCAGGAGCAGGGCTTCATCGAGGAGGACAAGGCCCGC
AATGATCACTACAAGGAGCTGAAGCCCATCATCGATCGGATCTACAAGACCTATGCCGACCAGTGCCT
GCAGCTGGTGCAGCTGGATTGGGAGAACCTGAGCGCCGCCATCGACTCCTATAGAAAGGAGAAAACC
GAGGAGACAAGGAACGCCCTGATCGAGGAGCAGGCCACATATCGCAATGCCATCCACGACTACTTCA
TCGGCCGGACAGACAACCTGACCGATGCCATCAATAAGAGACACGCCGAGATCTACAAGGGCCTGTT
CAAGGCCGAGCTGTTTAATGGCAAGGTGCTGAAGCAGCTGGGCACCGTGACCACAACCGAGCACGA
GAACGCCCTGCTGCGGAGCTTCGACAAGTTTACAACCTACTTCTCCGGCTTTTATGAGAACAGGAAG
AACGTGTTCAGCGCCGAGGATATCAGCACAGCCATCCCACACCGCATCGTGCAGGACAACTTCCCCA
AGTTTAAGGAGAATTGTCACATCTTCACACGCCTGATCACCGCCGTGCCCAGCCTGCGGGAGCACTTT
GAGAACGTGAAGAAGGCCATCGGCATCTTCGTGAGCACCTCCATCGAGGAGGTGTTTTCCTTCCCTTT
TTATAACCAGCTGCTGACACAGACCCAGATCGACCTGTATAACCAGCTGCTGGGAGGAATCTCTCGGG
AGGCAGGCACCGAGAAGATCAAGGGCCTGAACGAGGTGCTGAATCTGGCCATCCAGAAGAATGATG
AGACAGCCCACATCATCGCCTCCCTGCCACACAGATTCATCCCCCTGTTTAAGCAGATCCTGTCCGATA
GGAACACCCTGTCTTTCATCCTGGAGGAGTTTAAGAGCGACGAGGAAGTGATCCAGTCCTTCTGCAA
GTACAAGACACTGCTGAGAAACGAGAACGTGCTGGAGACAGCCGAGGCCCTGTTTAACGAGCTGAA
CAGCATCGACCTGACACACATCTTCATCAGCCACAAGAAGCTGGAGACAATCAGCAGCGCCCTGTGC
GACCACTGGGATACACTGAGGAATGCCCTGTATGAGCGGAGAATCTCCGAGCTGACAGGCAAGATCA
CCAAGTCTGCCAAGGAGAAGGTGCAGCGCAGCCTGAAGCACGAGGATATCAACCTGCAGGAGATCA
TCTCTGCCGCAGGCAAGGAGCTGAGCGAGGCCTTCAAGCAGAAAACCAGCGAGATCCTGTCCCACG
CACACGCCGCCCTGGATCAGCCACTGCCTACAACCCTGAAGAAGCAGGAGGAGAAGGAGATCCTGA
AGTCTCAGCTGGACAGCCTGCTGGGCCTGTACCACCTGCTGGACTGGTTTGCCGTGGATGAGTCCAA
CGAGGTGGACCCCGAGTTCTCTGCCCGGCTGACCGGCATCAAGCTGGAGATGGAGCCTTCTCTGAGC
TTCTACAACAAGGCCAGAAATTATGCCACCAAGAAGCCCTACTCCGTGGAGAAGTTCAAGCTGAACT
TTCAGATGCCTACACTGGCCTCTGGCTGGGACGTGAATAAGGAGAAGAACAATGGCGCCATCCTGTTT
GTGAAGAACGGCCTGTACTATCTGGGCATCATGCCAAAGCAGAAGGGCAGGTATAAGGCCCTGAGCT
TCGAGCCCACAGAGAAAACCAGCGAGGGCTTTGATAAGATGTACTATGACTACTTCCCTGATGCCGCC
AAGATGATCCCAAAGTGCAGCACCCAGCTGAAGGCCGTGACAGCCCACTTTCAGACCCACACAACCC
CCATCCTGCTGTCCAACAATTTCATCGAGCCTCTGGAGATCACAAAGGAGATCTACGACCTGAACAAT
CCTGAGAAGGAGCCAAAGAAGTTTCAGACAGCCTACGCCAAGAAAACCGGCGACCAGAAGGGCTAC
AGAGAGGCCCTGTGCAAGTGGATCGACTTCACAAGGGATTTTCTGTCCAAGTATACCAAGACAACCT
CTATCGATCTGTCTAGCCTGCGGCCATCCTCTCAGTATAAGGACCTGGGCGAGTACTATGCCGAGCTGA
ATCCCCTGCTGTACCACATCAGCTTCCAGAGAATCGCCGAGAAGGAGATCATGGATGCCGTGGAGAC
AGGCAAGCTGTACCTGTTCCAGATCTATAACAAGGACTTTGCCAAGGGCCACCACGGCAAGCCTAATC
TGCACACACTGTATTGGACCGGCCTGTTTTCTCCAGAGAACCTGGCCAAGACAAGCATCAAGCTGAA
TGGCCAGGCCGAGCTGTTCTACCGCCCTAAGTCCAGGATGAAGAGGATGGCACACCGGCTGGGAGAG
AAGATGCTGAACAAGAAGCTGAAGGATCAGAAAACCCCAATCCCCGACACCCTGTACCAGGAGCTGT
ACGACTATGTGAATCACAGACTGTCCCACGACCTGTCTGATGAGGCCAGGGCCCTGCTGCCCAACGT
GATCACCAAGGAGGTGTCTCACGAGATCATCAAGGATAGGCGCTTTACCAGCGACAAGTTCTTTTTCC
ACGTGCCTATCACACTGAACTATCAGGCCGCCAATTCCCCATCTAAGTTCAACCAGAGGGTGAATGCC
TACCTGAAGGAGCACCCCGAGACACCTATCATCGGCATCGATCGGGGCGAGAGAAACCTGATCTATAT
CACAGTGATCGACTCCACCGGCAAGATCCTGGAGCAGCGGAGCCTGAACACCATCCAGCAGTTTGAT

TACCAGAAGAAGCTGGACAACAGGGAGAAGGAGAGGGTGGCAGCAAGGCAGGCCTGGTCTGTGGT
GGGCACAATCAAGGATCTGAAGCAGGGCTATCTGAGCCAGGTCATCCACGAGATCGTGGACCTGATG
ATCCACTACCAGGCCGTGGTGGTGCTGGAGAACCTGAATTTCGGCTTTAAGAGCAAGAGGACCGGCA
TCGCCGAGAAGGCCGTGTACCAGCAGTTCGAGAAGATGCTGATCGATAAGCTGAATTGCCTGGTGCT
GAAGGACTATCCAGCAGAGAAAGTGGGAGGCGTGCTGAACCCATACCAGCTGACAGACCAGTTCAC
CTCCTTTGCCAAGATGGGCACCCAGTCTGGCTTCCTGTTTTACGTGCCTGCCCCATATACATCTAAGAT
CGATCCCCTGACCGGCTTCGTGGACCCCTTCGTGTGGAAAACCATCAAGAATCACGAGAGCCGCAAG
CACTTCCTGGAGGGCTTCGACTTTCTGCACTACGACGTGAAAACCGGCGACTTCATCCTGCACTTTAA
GATGAACAGAAATCTGTCCTTCCAGAGGGGCCTGCCCGGCTTTATGCCTGCATGGGATATCGTGTTCG

[0240] AGAAGAACGAGACACAGTTTGACGCCAAGGGCACCCCTTTCATCGCCGGCAAGAGAATCGTGCCAG
TGATCGAGAATCACAGATTCACCGGCAGATACCGGGACCTGTATCCTGCCAACGAGCTGATCGCCCTG
CTGGAGGAGAAGGGCATCGTGTTCAGGGATGGCTCCAACATCCTGCCAAAGCTGCTGGAGAATGACG
ATTCTCACGCCATCGACACCATGGTGGCCCTGATCCGCAGCGTGCTGCAGATGCGGAACTCCAATGCC
GCCACAGGCGAGGACTATATCAACAGCCCCGTGCGCGATCTGAATGGCGTGTGCTTCGACTCCCGGTT
TCAGAACCCAGAGTGGCCCATGGACGCCGATGCCAATGGCGCCTACCACATCGCCCTGAAGGGCCAG
CTGCTGCTGAATCACCTGAAGGAGAGCAAGGATCTGAAGCTGCAGAACGGCATCTCCAATCAGGACT
GGCTGGCCTACATCCAGGAGCTGCGCAACCaaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaag*GGATCC*
TACCCATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGAT
GTCCCCGACTATGCCTAA (SEQ ID NO: 421)

[0241] AAS826:pCAG-hAsCas12a(E174R/S542R)-NLS-3xHA

[0242] 人类密码子优化的氨基酸球菌种BV3L6 Cas12a(AsCas12a)为黑色,修饰的密码子(E174R/S542R)为双下划线的,核质蛋白NLS(aaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaag,SEQ ID NO:21)为小写字母,接头序列为斜体,3xHA标签(TACCCATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCC,SEQ ID NO:5)为粗体

ATGACACAGTTCGAGGGCTTTACCAACCTGTATCAGGTGAGCAAGACACTGCGGTTTGAGCTGATCCC
ACAGGGCAAGACCCTGAAGCACATCCAGGAGCAGGGCTTCATCGAGGAGGACAAGGCCCGCAATGA
TCACTACAAGGAGCTGAAGCCCATCATCGATCGGATCTACAAGACCTATGCCGACCAGTGCCTGCAGC
TGGTGCAGCTGGATTGGGAGAACCTGAGCGCCGCCATCGACTCCTATAGAAAGGAGAAAACCGAGG
AGACAAGGAACGCCCTGATCGAGGAGCAGGCCACATATCGCAATGCCATCCACGACTACTTCATCGG
CCGGACAGACAACCTGACCGATGCCATCAATAAGAGACACGCCGAGATCTACAAGGGCCTGTTCAAG
GCCGAGCTGTTTAATGGCAAGGTGCTGAAGCAGCTGGGCACCGTGACCACAACCGAGCACGAGAAC
GCCCTGCTGCGGAGCTTCGACAAGTTTACAACCTACTTCTCCGGCTTTTATAGAAACAGGAAGAACGT
GTTCAGCGCCGAGGATATCAGCACAGCCATCCCACACCGCATCGTGCAGGACAACTTCCCCAAGTTTA

[0243] AGGAGAATTGTCACATCTTCACACGCCTGATCACCGCCGTGCCCAGCCTGCGGGAGCACTTTGAGAA
CGTGAAGAAGGCCATCGGCATCTTCGTGAGCACCTCCATCGAGGAGGTGTTTTCCTTCCCTTTTTATAA
CCAGCTGCTGACACAGACCCAGATCGACCTGTATAACCAGCTGCTGGGAGGAATCTCTCGGGAGGCA
GGCACCGAGAAGATCAAGGGCCTGAACGAGGTGCTGAATCTGGCCATCCAGAAGAATGATGAGACA
GCCCACATCATCGCCTCCCTGCCACACAGATTCATCCCCCTGTTTAAGCAGATCCTGTCCGATAGGAAC
ACCCTGTCTTTCATCCTGGAGGAGTTTAAGAGCGACGAGGAAGTGATCCAGTCCTTCTGCAAGTACA
AGACACTGCTGAGAAACGAGAACGTGCTGGAGACAGCCGAGGCCCTGTTTAACGAGCTGAACAGCA
TCGACCTGACACACATCTTCATCAGCCACAAGAAGCTGGAGACAATCAGCAGCGCCCTGTGCGACCA
CTGGGATACACTGAGGAATGCCCTGTATGAGCGGAGAATCTCCGAGCTGACAGGCAAGATCACCAAG
TCTGCCAAGGAGAAGGTGCAGCGCAGCCTGAAGCACGAGGATATCAACCTGCAGGAGATCATCTCTG

[0244] CCGCAGGCAAGGAGCTGAGCGAGGCCTTCAAGCAGAAAACCAGCGAGATCCTGTCCCACGCACACG
CCGCCCTGGATCAGCCACTGCCTACAACCCTGAAGAAGCAGGAGGAGAAGGAGATCCTGAAGTCTC
AGCTGGACAGCCTGCTGGGCCTGTACCACCTGCTGGACTGGTTTGCCGTGGATGAGTCCAACGAGGT
GGACCCCGAGTTCTCTGCCCGGCTGACCGGCATCAAGCTGGAGATGGAGCCTTCTCTGAGCTTCTAC
AACAAGGCCAGAAATTATGCCACCAAGAAGCCCTACTCCGTGGAGAAGTTCAAGCTGAACTTTCAGA
TGCCTACACTGGCCAGAGGCTGGGACGTGAATAAGGAGAAGAACAATGGCGCCATCCTGTTTGTGAA
GAACGGCCTGTACTATCTGGGCATCATGCCAAAGCAGAAGGGCAGGTATAAGGCCCTGAGCTTCGAG
CCCACAGAGAAAACCAGCGAGGGCTTTGATAAGATGTACTATGACTACTTCCCTGATGCCGCCAAGAT
GATCCCAAAGTGCAGCACCCAGCTGAAGGCCGTGACAGCCCACTTTCAGACCCACACAACCCCCATC
CTGCTGTCCAACAATTTCATCGAGCCTCTGGAGATCACAAAGGAGATCTACGACCTGAACAATCCTGA
GAAGGAGCCAAAGAAGTTTCAGACAGCCTACGCCAAGAAAACCGGCGACCAGAAGGGCTACAGAG
AGGCCCTGTGCAAGTGGATCGACTTCACAAGGGATTTTCTGTCCAAGTATACCAAGACAACCTCTATC
GATCTGTCTAGCCTGCGGCCATCCTCTCAGTATAAGGACCTGGGCGAGTACTATGCCGAGCTGAATCC
CCTGCTGTACCACATCAGCTTCCAGAGAATCGCCGAGAAGGAGATCATGGATGCCGTGGAGACAGGC
AAGCTGTACCTGTTCCAGATCTATAACAAGGACTTTGCCAAGGGCCACCACGGCAAGCCTAATCTGCA
CACACTGTATTGGACCGGCCTGTTTTCTCCAGAGAACCTGGCCAAGACAAGCATCAAGCTGAATGGC
CAGGCCGAGCTGTTCTACCGCCCTAAGTCCAGGATGAAGAGGATGGCACACCGGCTGGGAGAGAAG
ATGCTGAACAAGAAGCTGAAGGATCAGAAAACCCCAATCCCCGACACCCTGTACCAGGAGCTGTACG
ACTATGTGAATCACAGACTGTCCCACGACCTGTCTGATGAGGCCAGGGCCCTGCTGCCCAACGTGATC
ACCAAGGAGGTGTCTCACGAGATCATCAAGGATAGGCGCTTTACCAGCGACAAGTTCTTTTTCCACGT
GCCTATCACACTGAACTATCAGGCCGCCAATTCCCCATCTAAGTTCAACCAGAGGGTGAATGCCTACC
TGAAGGAGCACCCCGAGACACCTATCATCGGCATCGATCGGGGCGAGAGAAACCTGATCTATATCACA
GTGATCGACTCCACCGGCAAGATCCTGGAGCAGCGGAGCCTGAACACCATCCAGCAGTTTGATTACC
AGAAGAAGCTGGACAACAGGGAGAAGGAGAGGGTGGCAGCAAGGCAGGCCTGGTCTGTGGTGGGC
ACAATCAAGGATCTGAAGCAGGGCTATCTGAGCCAGGTCATCCACGAGATCGTGGACCTGATGATCCA
CTACCAGGCCGTGGTGGTGCTGGAGAACCTGAATTTCGGCTTTAAGAGCAAGAGGACCGGCATCGCC
GAGAAGGCCGTGTACCAGCAGTTCGAGAAGATGCTGATCGATAAGCTGAATTGCCTGGTGCTGAAGG
ACTATCCAGCAGAGAAAGTGGGAGGCGTGCTGAACCCATACCAGCTGACAGACCAGTTCACCTCCTT
TGCCAAGATGGGCACCCAGTCTGGCTTCCTGTTTTACGTGCCTGCCCCATATACATCTAAGATCGATCC
CCTGACCGGCTTCGTGGACCCCTTCGTGTGGAAAACCATCAAGAATCACGAGAGCCGCAAGCACTTC
CTGGAGGGCTTCGACTTTCTGCACTACGACGTGAAAACCGGCGACTTCATCCTGCACTTTAAGATGAA
CAGAAATCTGTCCTTCCAGAGGGGCCTGCCCGGCTTTATGCCTGCATGGGATATCGTGTTCGAGAAGA
ACGAGACACAGTTTGACGCCAAGGGCACCCCTTTCATCGCCGGCAAGAGAATCGTGCCAGTGATCGA
GAATCACAGATTCACCGGCAGATACCGGGACCTGTATCCTGCCAACGAGCTGATCGCCCTGCTGGAGG
AGAAGGGCATCGTGTTCAGGGATGGCTCCAACATCCTGCCAAAGCTGCTGGAGAATGACGATTCTCA
CGCCATCGACACCATGGTGGCCCTGATCCGCAGCGTGCTGCAGATGCGGAACTCCAATGCCGCCACA
GGCGAGGACTATATCAACAGCCCCGTGCGCGATCTGAATGGCGTGTGCTTCGACTCCCGGTTTCAGAA
CCCAGAGTGGCCCATGGACGCCGATGCCAATGGCGCCTACCACATCGCCCTGAAGGGCCAGCTGCTG
CTGAATCACCTGAAGGAGAGCAAGGATCTGAAGCTGCAGAACGGCATCTCCAATCAGGACTGGCTGG
CCTACATCCAGGAGCTGCGCAACCaaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaagGGATCCTACCC
ATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCC
CGACTATGCCTAA (SEQ ID NO: 422)

[0245] AAS848:pCAG-heAsCas12a(E174R/S542R/K548R)-NLS-3xHA

[0246] 人类密码子优化的氨基酸球菌种BV3L6 Cas12a(AsCas12a)为黑色,eAsCas12a的修饰的密码子(E174R/S542R/K548R)为双下划线的,核质蛋白NLS(aaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaag,SEQ ID NO:21)为小写字母,接头序列为斜体,3xHA标签(TACCCATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCA

TACCCATATGATGTCCCCGACTATGCC,SEQ ID NO:5)为粗体

[0247]

ATGACACAGTTCGAGGGCTTTACCAACCTGTATCAGGTGAGCAAGACACTGCGGTTTGAGCTGATCCC
ACAGGGCAAGACCCTGAAGCACATCCAGGAGCAGGGCTTCATCGAGGAGGACAAGGCCCGCAATGA
TCACTACAAGGAGCTGAAGCCCATCATCGATCGGATCTACAAGACCTATGCCGACCAGTGCCTGCAGC
TGGTGCAGCTGGATTGGGAGAACCTGAGCGCCGCCATCGACTCCTATAGAAAGGAGAAAACCGAGG
AGACAAGGAACGCCCTGATCGAGGAGCAGGCCACATATCGCAATGCCATCCACGACTACTTCATCGG
CCGGACAGACAACCTGACCGATGCCATCAATAAGAGACACGCCGAGATCTACAAGGGCCTGTTCAAG
GCCGAGCTGTTTAATGGCAAGGTGCTGAAGCAGCTGGGCACCGTGACCACAACCGAGCACGAGAAC
GCCCTGCTGCGGAGCTTCGACAAGTTTACAACCTACTTCTCCGGCTTTTATAGAAACAGGAAGAACGT
GTTCAGCGCCGAGGATATCAGCACAGCCATCCCACACCGCATCGTGCAGGACAACTTCCCCAAGTTTA
AGGAGAATTGTCACATCTTCACACGCCTGATCACCGCCGTGCCCAGCCTGCGGGAGCACTTTGAGAA
CGTGAAGAAGGCCATCGGCATCTTCGTGAGCACCTCCATCGAGGAGGTGTTTTCCTTCCCTTTTTATAA
CCAGCTGCTGACACAGACCCAGATCGACCTGTATAACCAGCTGCTGGGAGGAATCTCTCGGGAGGCA
GGCACCGAGAAGATCAAGGGCCTGAACGAGGTGCTGAATCTGGCCATCCAGAAGAATGATGAGACA
GCCCACATCATCGCCTCCCTGCCACACAGATTCATCCCCCTGTTTAAGCAGATCCTGTCCGATAGGAAC
ACCCTGTCTTTCATCCTGGAGGAGTTTAAGAGCGACGAGGAAGTGATCCAGTCCTTCTGCAAGTACA
AGACACTGCTGAGAAACGAGAACGTGCTGGAGACAGCCGAGGCCCTGTTTAACGAGCTGAACAGCA
TCGACCTGACACACATCTTCATCAGCCACAAGAAGCTGGAGACAATCAGCAGCGCCCTGTGCGACCA
CTGGGATACACTGAGGAATGCCCTGTATGAGCGGAGAATCTCCGAGCTGACAGGCAAGATCACCAAG
TCTGCCAAGGAGAAGGTGCAGCGCAGCCTGAAGCACGAGGATATCAACCTGCAGGAGATCATCTCTG
CCGCAGGCAAGGAGCTGAGCGAGGCCTTCAAGCAGAAAACCAGCGAGATCCTGTCCCACGCACACG
CCGCCCTGGATCAGCCACTGCCTACAACCCTGAAGAAGCAGGAGGAGAAGGAGATCCTGAAGTCTC
AGCTGGACAGCCTGCTGGGCCTGTACCACCTGCTGGACTGGTTTGCCGTGGATGAGTCCAACGAGGT
GGACCCCGAGTTCTCTGCCCGGCTGACCGGCATCAAGCTGGAGATGGAGCCTTCTCTGAGCTTCTAC
AACAAGGCCAGAAATTATGCCACCAAGAAGCCCTACTCCGTGGAGAAGTTCAAGCTGAACTTTCAGA
TGCCTACACTGGCCAGAGGCTGGGACGTGAATAGAGAGAAGAACAATGGCGCCATCCTGTTTGTGAA
GAACGGCCTGTACTATCTGGGCATCATGCCAAAGCAGAAGGGCAGGTATAAGGCCCTGAGCTTCGAG
CCCACAGAGAAAACCAGCGAGGGCTTTGATAAGATGTACTATGACTACTTCCCTGATGCCGCCAAGAT
GATCCCAAAGTGCAGCACCCAGCTGAAGGCCGTGACAGCCCACTTTCAGACCCACACAACCCCCATC
CTGCTGTCCAACAATTTCATCGAGCCTCTGGAGATCACAAAGGAGATCTACGACCTGAACAATCCTGA
GAAGGAGCCAAAGAAGTTTCAGACAGCCTACGCCAAGAAAACCGGCGACCAGAAGGGCTACAGAG
AGGCCCTGTGCAAGTGGATCGACTTCACAAGGGATTTTCTGTCCAAGTATACCAAGACAACCTCTATC
GATCTGTCTAGCCTGCGGCCATCCTCTCAGTATAAGGACCTGGGCGAGTACTATGCCGAGCTGAATCC
CCTGCTGTACCACATCAGCTTCCAGAGAATCGCCGAGAAGGAGATCATGGATGCCGTGGAGACAGGC
AAGCTGTACCTGTTCCAGATCTATAACAAGGACTTTGCCAAGGGCCACCACGGCAAGCCTAATCTGCA
CACACTGTATTGGACCGGCCTGTTTTCTCCAGAGAACCTGGCCAAGACAAGCATCAAGCTGAATGGC
CAGGCCGAGCTGTTCTACCGCCCTAAGTCCAGGATGAAGAGGATGGCACACCGGCTGGGAGAGAAG
ATGCTGAACAAGAAGCTGAAGGATCAGAAAACCCCAATCCCCGACACCCTGTACCAGGAGCTGTACG
ACTATGTGAATCACAGACTGTCCCACGACCTGTCTGATGAGGCCAGGGCCCTGCTGCCCAACGTGATC
ACCAAGGAGGTGTCTCACGAGATCATCAAGGATAGGCGCTTTACCAGCGACAAGTTCTTTTTCCACGT
GCCTATCACACTGAACTATCAGGCCGCCAATTCCCCATCTAAGTTCAACCAGAGGGTGAATGCCTACC

TGAAGGAGCACCCCGAGACACCTATCATCGGCATCGATCGGGGCGAGAGAAACCTGATCTATATCACA
GTGATCGACTCCACCGGCAAGATCCTGGAGCAGCGGAGCCTGAACACCATCCAGCAGTTTGATTACC
AGAAGAAGCTGGACAACAGGGAGAAGGAGAGGGTGGCAGCAAGGCAGGCCTGGTCTGTGGTGGGC
ACAATCAAGGATCTGAAGCAGGGCTATCTGAGCCAGGTCATCCACGAGATCGTGGACCTGATGATCCA
CTACCAGGCCGTGGTGGTGCTGGAGAACCTGAATTTCGGCTTTAAGAGCAAGAGGACCGGCATCGCC
GAGAAGGCCGTGTACCAGCAGTTCGAGAAGATGCTGATCGATAAGCTGAATTGCCTGGTGCTGAAGG
ACTATCCAGCAGAGAAAGTGGGAGGCGTGCTGAACCCATACCAGCTGACAGACCAGTTCACCTCCTT
TGCCAAGATGGGCACCCAGTCTGGCTTCCTGTTTTACGTGCCTGCCCCATATACATCTAAGATCGATCC
CCTGACCGGCTTCGTGGACCCCTTCGTGTGGAAAACCATCAAGAATCACGAGAGCCGCAAGCACTTC
CTGGAGGGCTTCGACTTTCTGCACTACGACGTGAAAACCGGCGACTTCATCCTGCACTTTAAGATGAA
[0248] CAGAAATCTGTCCTTCCAGAGGGGCCTGCCCGGCTTTATGCCTGCATGGGATATCGTGTTCGAGAAGA
ACGAGACACAGTTTGACGCCAAGGGCACCCCTTTCATCGCCGGCAAGAGAATCGTGCCAGTGATCGA
GAATCACAGATTCACCGGCAGATACCGGGACCTGTATCCTGCCAACGAGCTGATCGCCCTGCTGGAGG
AGAAGGGCATCGTGTTCAGGGATGGCTCCAACATCCTGCCAAAGCTGCTGGAGAATGACGATTCTCA
CGCCATCGACACCATGGTGGCCCTGATCCGCAGCGTGCTGCAGATGCGGAACTCCAATGCCGCCACA
GGCGAGGACTATATCAACAGCCCCGTGCGCGATCTGAATGGCGTGTGCTTCGACTCCCGGTTTCAGAA
CCCAGAGTGGCCCATGGACGCCGATGCCAATGGCGCCTACCACATCGCCCTGAAGGGCCAGCTGCTG
CTGAATCACCTGAAGGAGAGCAAGGATCTGAAGCTGCAGAACGGCATCTCCAATCAGGACTGGCTGG
CCTACATCCAGGAGCTGCGCAACaaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaag*GGATCC***TACCC**
ATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCC
CGACTATGCCTAA (SEQ ID NO:423)

[0249] AAS1815:pCAG-heAsCas12a-HF1(E174R/N282A/S542R/K548R)-NLS-3xHA

[0250] 人类密码子优化的氨基酸球菌种BV3L6 Cas12a(AsCas12a)为黑色,修饰的密码子(E174R/N282A/S542R/K548R)为双下划线的,核质蛋白NLS(aaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaag,SEQ ID NO:21)为小写字母,接头序列为斜体,3xHA标签(TACCCATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCC,SEQ ID NO:5)为粗体

ATGACACAGTTCGAGGGCTTTACCAACCTGTATCAGGTGAGCAAGACACTGCGGTTTGAGCTGATCCC
ACAGGGCAAGACCCTGAAGCACATCCAGGAGCAGGGCTTCATCGAGGAGGACAAGGCCCGCAATGA
TCACTACAAGGAGCTGAAGCCCATCATCGATCGGATCTACAAGACCTATGCCGACCAGTGCCTGCAGC
TGGTGCAGCTGGATTGGGAGAACCTGAGCGCCGCCATCGACTCCTATAGAAAGGAGAAAACCGAGG
AGACAAGGAACGCCCTGATCGAGGAGCAGGCCACATATCGCAATGCCATCCACGACTACTTCATCGG
CCGGACAGACAACCTGACCGATGCCATCAATAAGAGACACGCCGAGATCTACAAGGGCCTGTTCAAG
GCCGAGCTGTTTAATGGCAAGGTGCTGAAGCAGCTGGGCACCGTGACCACAACCGAGCACGAGAAC
[0251] GCCCTGCTGCGGAGCTTCGACAAGTTTACAACCTACTTCTCCGGCTTTTATAGAAACAGGAAGAACGT
GTTCAGCGCCGAGGATATCAGCACAGCCATCCCACACCGCATCGTGCAGGACAACTTCCCCAAGTTTA
AGGAGAATTGTCACATCTTCACACGCCTGATCACCGCCGTGCCCAGCCTGCGGGAGCACTTTGAGAA
CGTGAAGAAGGCCATCGGCATCTTCGTGAGCACCTCCATCGAGGAGGTGTTTTCCTTCCCTTTTTATAA
CCAGCTGCTGACACAGACCCAGATCGACCTGTATAACCAGCTGCTGGGAGGAATCTCTCGGGAGGCA
GGCACCGAGAAGATCAAGGGCCTGAACGAGGTGCTGGCCCTGGCCATCCAGAAGAATGATGAGACA
GCCCACATCATCGCCTCCCTGCCACACAGATTCATCCCCCTGTTTAAGCAGATCCTGTCCGATAGGAAC
ACCCTGTCTTTCATCCTGGAGGAGTTTAAGAGCGACGAGGAAGTGATCCAGTCCTTCTGCAAGTACA
AGACACTGCTGAGAAACGAGAACGTGCTGGAGACAGCCGAGGCCCTGTTTAACGAGCTGAACAGCA

TCGACCTGACACACATCTTCATCAGCCACAAGAAGCTGGAGACAATCAGCAGCGCCCTGTGCGACCA
CTGGGATACACTGAGGAATGCCCTGTATGAGCGGAGAATCTCCGAGCTGACAGGCAAGATCACCAAG
TCTGCCAAGGAGAAGGTGCAGCGCAGCCTGAAGCACGAGGATATCAACCTGCAGGAGATCATCTCTG
CCGCAGGCAAGGAGCTGAGCGAGGCCTTCAAGCAGAAAACCAGCGAGATCCTGTCCCACGCACACG
CCGCCCTGGATCAGCCACTGCCTACAACCCTGAAGAAGCAGGAGGAGAAGGAGATCCTGAAGTCTC
AGCTGGACAGCCTGCTGGGCCTGTACCACCTGCTGGACTGGTTTGCCGTGGATGAGTCCAACGAGGT
GGACCCCGAGTTCTCTGCCCGGCTGACCGGCATCAAGCTGGAGATGGAGCCTTCTCTGAGCTTCTAC
AACAAGGCCAGAAATTATGCCACCAAGAAGCCCTACTCCGTGGAGAAGTTCAAGCTGAACTTTCAGA
TGCCTACACTGGCCAGAGGCTGGGACGTGAATAGAGAGAAGAACAATGGCGCCATCCTGTTTGTGAA
GAACGGCCTGTACTATCTGGGCATCATGCCAAAGCAGAAGGGCAGGTATAAGGCCCTGAGCTTCGAG
CCCACAGAGAAAACCAGCGAGGGCTTTGATAAGATGTACTATGACTACTTCCCTGATGCCGCCAAGAT
GATCCCAAAGTGCAGCACCCAGCTGAAGGCCGTGACAGCCCACTTTCAGACCCACACAACCCCCATC
CTGCTGTCCAACAATTTCATCGAGCCTCTGGAGATCACAAAGGAGATCTACGACCTGAACAATCCTGA
GAAGGAGCCAAAGAAGTTTCAGACAGCCTACGCCAAGAAAACCGGCGACCAGAAGGGCTACAGAG
AGGCCCTGTGCAAGTGGATCGACTTCACAAGGGATTTTCTGTCCAAGTATACCAAGACAACCTCTATC
GATCTGTCTAGCCTGCGGCCATCCTCTCAGTATAAGGACCTGGGCGAGTACTATGCCGAGCTGAATCC
CCTGCTGTACCACATCAGCTTCCAGAGAATCGCCGAGAAGGAGATCATGGATGCCGTGGAGACAGGC
AAGCTGTACCTGTTCCAGATCTATAACAAGGACTTTGCCAAGGGCCACCACGGCAAGCCTAATCTGCA
CACACTGTATTGGACCGGCCTGTTTTCTCCAGAGAACCTGGCCAAGACAAGCATCAAGCTGAATGGC
CAGGCCGAGCTGTTCTACCGCCCTAAGTCCAGGATGAAGAGGATGGCACACCGGCTGGGAGAGAAG
ATGCTGAACAAGAAGCTGAAGGATCAGAAAACCCCAATCCCCGACACCCTGTACCAGGAGCTGTACG
ACTATGTGAATCACAGACTGTCCCACGACCTGTCTGATGAGGCCAGGGCCCTGCTGCCCAACGTGATC

[0252] ACCAAGGAGGTGTCTCACGAGATCATCAAGGATAGGCGCTTTACCAGCGACAAGTTCTTTTTCCACGT
GCCTATCACACTGAACTATCAGGCCGCCAATTCCCCATCTAAGTTCAACCAGAGGGTGAATGCCTACC
TGAAGGAGCACCCCGAGACACCTATCATCGGCATCGATCGGGGCGAGAGAAACCTGATCTATATCACA
GTGATCGACTCCACCGGCAAGATCCTGGAGCAGCGGAGCCTGAACACCATCCAGCAGTTTGATTACC
AGAAGAAGCTGGACAACAGGGAGAAGGAGAGGGTGGCAGCAAGGCAGGCCTGGTCTGTGGTGGGC
ACAATCAAGGATCTGAAGCAGGGCTATCTGAGCCAGGTCATCCACGAGATCGTGGACCTGATGATCCA
CTACCAGGCCGTGGTGGTGCTGGAGAACCTGAATTTCGGCTTTAAGAGCAAGAGGACCGGCATCGCC
GAGAAGGCCGTGTACCAGCAGTTCGAGAAGATGCTGATCGATAAGCTGAATTGCCTGGTGCTGAAGG
ACTATCCAGCAGAGAAAGTGGGAGGCGTGCTGAACCCATACCAGCTGACAGACCAGTTCACCTCCTT
TGCCAAGATGGGCACCCAGTCTGGCTTCCTGTTTTACGTGCCTGCCCCATATACATCTAAGATCGATCC
CCTGACCGGCTTCGTGGACCCCTTCGTGTGGAAAACCATCAAGAATCACGAGAGCCGCAAGCACTTC
CTGGAGGGCTTCGACTTTCTGCACTACGACGTGAAAACCGGCGACTTCATCCTGCACTTTAAGATGAA
CAGAAATCTGTCCTTCCAGAGGGGCCTGCCCGGCTTTATGCCTGCATGGGATATCGTGTTCGAGAAGA
ACGAGACACAGTTTGACGCCAAGGGCACCCCTTTCATCGCCGGCAAGAGAATCGTGCCAGTGATCGA
GAATCACAGATTCACCGGCAGATACCGGGACCTGTATCCTGCCAACGAGCTGATCGCCCTGCTGGAGG
AGAAGGGCATCGTGTTCAGGGATGGCTCCAACATCCTGCCAAAGCTGCTGGAGAATGACGATTCTCA
CGCCATCGACACCATGGTGGCCCTGATCCGCAGCGTGCTGCAGATGCGGAACTCCAATGCCGCCACA
GGCGAGGACTATATCAACAGCCCCGTGCGCGATCTGAATGGCGTGTGCTTCGACTCCCGGTTTCAGAA
CCCAGAGTGGCCCATGGACGCCGATGCCAATGGCGCCTACCACATCGCCCTGAAGGGCCAGCTGCTG
CTGAATCACCTGAAGGAGAGCAAGGATCTGAAGCTGCAGAACGGCATCTCCAATCAGGACTGGCTGG
CCTACATCCAGGAGCTGCGCAACCaaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaag*GGATCC***TACC**
CATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGT
CCCCGACTATGCCTAA (SEQ ID NO: 424)

[0253] BPK3541:pET-28b-hAsCas12a-NLS-6xHis

[0254] 人类密码子优化的氨基酸球菌种BV3L6 Cas12a(AsCas12a)为黑色,具有沉默突变以移除NcoI位点的密码子为双下划线的,插入的甘氨酸是点式下划线的,核质蛋

白NLS(aaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaag,SEQ ID NO:21)为小写字母,接头序列为斜体,6xHis为粗体

[0255] ATGGGGACACAGTTCGAGGGCTTTACCAACCTGTATCAGGTGAGCAAGACACTGCGGTTTGAGCTGA
TCCCACAGGGCAAGACCCTGAAGCACATCCAGGAGCAGGGCTTCATCGAGGAGGACAAGGCCCGCA
ATGATCACTACAAGGAGCTGAAGCCCATCATCGATCGGATCTACAAGACCTATGCCGACCAGTGCCTG
CAGCTGGTGCAGCTGGATTGGGAGAACCTGAGCGCCGCCATCGACTCCTATAGAAAGGAGAAAACCG
AGGAGACAAGGAACGCCCTGATCGAGGAGCAGGCCACATATCGCAATGCCATCCACGACTACTTCAT
CGGCCGGACAGACAACCTGACCGATGCCATCAATAAGAGACACGCCGAGATCTACAAGGGCCTGTTC
AAGGCCGAGCTGTTTAATGGCAAGGTGCTGAAGCAGCTGGGCACCGTGACCACAACCGAGCACGAG
AACGCCCTGCTGCGGAGCTTCGACAAGTTTACAACCTACTTCTCCGGCTTTTATGAGAACAGGAAGA
ACGTGTTCAGCGCCGAGGATATCAGCACAGCCATCCCACACCGCATCGTGCAGGACAACTTCCCCAA
GTTTAAGGAGAATTGTCACATCTTCACACGCCTGATCACCGCCGTGCCCAGCCTGCGGGAGCACTTTG
AGAACGTGAAGAAGGCCATCGGCATCTTCGTGAGCACCTCCATCGAGGAGGTGTTTTCCTTCCCTTTT
TATAACCAGCTGCTGACACAGACCCAGATCGACCTGTATAACCAGCTGCTGGGAGGAATCTCTCGGGA
GGCAGGCACCGAGAAGATCAAGGGCCTGAACGAGGTGCTGAATCTGGCCATCCAGAAGAATGATGA
GACAGCCCACATCATCGCCTCCCTGCCACACAGATTCATCCCCCTGTTTAAGCAGATCCTGTCCGATAG
GAACACCCTGTCTTTCATCCTGGAGGAGTTTAAGAGCGACGAGGAAGTGATCCAGTCCTTCTGCAAG
TACAAGACACTGCTGAGAAACGAGAACGTGCTGGAGACAGCCGAGGCCCTGTTTAACGAGCTGAAC
AGCATCGACCTGACACACATCTTCATCAGCCACAAGAAGCTGGAGACAATCAGCAGCGCCCTGTGCG
ACCACTGGGATACACTGAGGAATGCCCTGTATGAGCGGAGAATCTCCGAGCTGACAGGCAAGATCAC
CAAGTCTGCCAAGGAGAAGGTGCAGCGCAGCCTGAAGCACGAGGATATCAACCTGCAGGAGATCAT
CTCTGCCGCAGGCAAGGAGCTGAGCGAGGCCTTCAAGCAGAAAACCAGCGAGATCCTGTCCCACGC
ACACGCCGCCCTGGATCAGCCACTGCCTACAACCCTGAAGAAGCAGGAGGAGAAGGAGATCCTGAA
GTCTCAGCTGGACAGCCTGCTGGGCCTGTACCACCTGCTGGACTGGTTTGCCGTGGATGAGTCCAAC
GAGGTGGACCCCGAGTTCTCTGCCCGGCTGACCGGCATCAAGCTGGAGATGGAGCCTTCTCTGAGCT
TCTACAACAAGGCCAGAAATTATGCCACCAAGAAGCCCTACTCCGTGGAGAAGTTCAAGCTGAACTT
TCAGATGCCTACACTGGCCTCTGGCTGGGACGTGAATAAGGAGAAGAACAATGGCGCCATCCTGTTT
GTGAAGAACGGCCTGTACTATCTGGGCATCATGCCAAAGCAGAAGGGCAGGTATAAGGCCCTGAGCT
TCGAGCCCACAGAGAAAACCAGCGAGGGCTTTGATAAGATGTACTATGACTACTTCCCTGATGCCGCC
AAGATGATCCCAAAGTGCAGCACCCAGCTGAAGGCCGTGACAGCCCACTTTCAGACCCACACAACCC
CCATCCTGCTGTCCAACAATTTCATCGAGCCTCTGGAGATCACAAAGGAGATCTACGACCTGAACAAT
CCTGAGAAGGAGCCAAAGAAGTTTCAGACAGCCTACGCCAAGAAAACCGGCGACCAGAAGGGCTAC
AGAGAGGCCCTGTGCAAGTGGATCGACTTCACAAGGGATTTTCTGTCCAAGTATACCAAGACAACCT
CTATCGATCTGTCTAGCCTGCGGCCATCCTCTCAGTATAAGGACCTGGGCGAGTACTATGCCGAGCTGA
ATCCCCTGCTGTACCACATCAGCTTCCAGAGAATCGCCGAGAAGGAGATCATGGATGCCGTGGAGAC
AGGCAAGCTGTACCTGTTCCAGATCTATAACAAGGACTTTGCCAAGGGCCACCACGGCAAGCCTAATC
TGCACACACTGTATTGGACCGGCCTGTTTTCTCCAGAGAACCTGGCCAAGACAAGCATCAAGCTGAA
TGGCCAGGCCGAGCTGTTCTACCGCCCTAAGTCCAGGATGAAGAGGATGGCACACCGGCTGGGAGAG
AAGATGCTGAACAAGAAGCTGAAGGATCAGAAAACCCCAATCCCCGACACCCTGTACCAGGAGCTGT
ACGACTATGTGAATCACAGACTGTCCCACGACCTGTCTGATGAGGCCAGGGCCCTGCTGCCCAACGT
GATCACCAAGGAGGTGTCTCACGAGATCATCAAGGATAGGCGCTTTACCAGCGACAAGTTCTTTTTCC

ACGTGCCTATCACACTGAACTATCAGGCCGCCAATTCCCCATCTAAGTTCAACCAGAGGGTGAATGCC
TACCTGAAGGAGCACCCCGAGACACCTATCATCGGCATCGATCGGGGCGAGAGAAACCTGATCTATAT
CACAGTGATCGACTCCACCGGCAAGATCCTGGAGCAGCGGAGCCTGAACACCATCCAGCAGTTTGAT
TACCAGAAGAAGCTGGACAACAGGGAGAAGGAGAGGGTGGCAGCAAGGCAGGCCTGGTCTGTGGT
GGGCACAATCAAGGATCTGAAGCAGGGCTATCTGAGCCAGGTCATCCACGAGATCGTGGACCTGATG
ATCCACTACCAGGCCGTGGTGGTGCTGGAGAACCTGAATTTCGGCTTTAAGAGCAAGAGGACCGGCA
TCGCCGAGAAGGCCGTGTACCAGCAGTTCGAGAAGATGCTGATCGATAAGCTGAATTGCCTGGTGCT
GAAGGACTATCCAGCAGAGAAAGTGGGAGGCGTGCTGAACCCATACCAGCTGACAGACCAGTTCAC
CTCCTTTGCCAAGATGGGCACCCAGTCTGGCTTCCTGTTTTACGTGCCTGCCCCATATACATCTAAGAT
CGATCCCCTGACCGGCTTCGTGGACCCCTTCGTGTGGAAAACCATCAAGAATCACGAGAGCCGCAAG

[0256] CACTTCCTGGAGGGCTTCGACTTTCTGCACTACGACGTGAAAACCGGCGACTTCATCCTGCACTTTAA
GATGAACAGAAATCTGTCCTTCCAGAGGGGCCTGCCCGGCTTTATGCCTGCATGGGATATCGTGTTCG
AGAAGAACGAGACACAGTTTGACGCCAAGGGCACCCCTTTCATCGCCGGCAAGAGAATCGTGCCAG
TGATCGAGAATCACAGATTCACCGGCAGATACCGGGACCTGTATCCTGCCAACGAGCTGATCGCCCTG
CTGGAGGAGAAGGGCATCGTGTTCAGGGATGGCTCCAACATCCTGCCAAAGCTGCTGGAGAATGACG
ATTCTCACGCCATCGAC**ACG**ATGGTGGCCCTGATCCGCAGCGTGCTGCAGATGCGGAACTCCAATGCC
GCCACAGGCGAGGACTATATCAACAGCCCCGTGCGCGATCTGAATGGCGTGTGCTTCGACTCCCGGTT
TCAGAACCCAGAGTGG**CCA**ATGGACGCCGATGCCAATGGCGCCTACCACATCGCCCTGAAGGGCCAG
CTGCTGCTGAATCACCTGAAGGAGAGCAAGGATCTGAAGCTGCAGAACGGCATCTCCAATCAGGACT
GGCTGGCCTACATCCAGGAGCTGCGCAAC*aaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaag**GGAGC*
*GGCCGCACTCGAG***CACCACCACCACCACCAC**TGA (SEQ ID NO:425)

[0257] RTW645:pET-28b-bLbCas12a-NLS-6xHis

[0258] 细菌密码子优化的毛螺旋菌科细菌ND2006 Cas12a(LbCas12a)为黑色,插入的甘氨酸为点式下划线的,核质蛋白NLS(aaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaag,SEQ ID NO:21)为小写字母,接头序列为斜体,6xHis为粗体

ATGGGGAGCAAACTGGAAAAATTTACGAATTGTTATAGCCTGTCCAAGACCCTGCGTTTCAAAGCCAT
CCCCGTTGGCAAAACCCAGGAGAATATTGATAATAAACGTCTGCTGGTTGAGGATGAAAAAAGAGCA
GAAGACTATAAGGGAGTCAAAAAACTGCTGGATCGGTACTACCTGAGCTTTATAAATGACGTGCTGCA
TAGCATTAAACTGAAAAATCTGAATAACTATATTAGTCTGTTCCGCAAGAAAACCCGAACAGAGAAAG
AAAATAAAGAGCTGGAAAACCTGGAGATCAATCTGCGTAAAGAGATCGCAAAAGCTTTTAAAGGAAA
TGAAGGTTATAAAAGCCTGTTCAAAAAAGACATTATTGAAACCATCCTGCCGGAATTTCTGGATGATA
AAGACGAGATAGCGCTCGTGAACAGCTTCAACGGGTTCACGACCGCCTTCACGGGCTTTTTCGATAA
CAGGGAAAATATGTTTTCAGAGGAAGCCAAAAGCACCTCGATAGCGTTCCGTTGCATTAATGAAAATT
TGACAAGATATATCAGCAACATGGATATTTTCGAGAAAGTTGATGCGATCTTTGACAAACATGAAGTG

[0259] CAGGAGATTAAGGAAAAAATTCTGAACAGCGATTATGATGTTGAGGATTTTTTCGAGGGGGAATTTTT
TAACTTTGTACTGACACAGGAAGGTATAGATGTGTATAATGCTATTATCGGCGGGTTCGTTACCGAATC
CGGCGAGAAAATTAAGGGTCTGAATGAGTACATCAATCTGTATAACCAAAAGACCAAACAGAAACTG
CCAAAATTCAAACCGCTGTACAAGCAAGTCCTGAGCGATCGGGAAAGCTTGAGCTTTTACGGTGAAG
GTTATACCAGCGACGAGGAGGTACTGGAGGTCTTTCGCAATACCCTGAACAAGAACAGCGAAATTTT
CAGCTCCATTAAAAAGCTGGAGAAACTGTTTAAGAATTTTGACGAGTACAGCAGCGCAGGTATTTTTG
TGAAGAACGGACCTGCCATAAGCACCATTAGCAAGGATATTTTTGGAGAGTGGAATGTTATCCGTGAT
AAATGGAACGCGGAATATGATGACATACACCTGAAAAAGAAGGCTGTGGTAACTGAGAAATATGAAG
ACGATCGCCGCAAAAGCTTTAAAAAAATCGGCAGCTTTAGCCTGGAGCAGCTGCAGGAATATGCGGA
CGCCGACCTGAGCGTGGTCGAGAAACTGAAGGAAATTATTATCCAAAAAGTGGATGAGATTTACAAG

GTATATGGTAGCAGCGAAAAACTGTTTGATGCGGACTTCGTTCTGGAAAAAAGCCTGAAAAAAAATG
ATGCTGTTGTTGCGATCATGAAAGACCTGCTCGATAGCGTTAAGAGCTTTGAAAATTACATTAAAGCAT
TCTTTGGCGAGGGCAAAGAAACAAACAGAGACGAAAGCTTTTATGGCGACTTCGTCCTGGCTTATGA
CATCCTGTTGAAGGTAGATCATATATATGATGCAATTCGTAATTACGTAACCCAAAAGCCGTACAGCAA
AGATAAGTTCAAACTGTATTTCCAGAACCCGCAGTTTATGGGTGGCTGGGACAAAGACAAGGAGACA
GACTATCGCGCCACTATTCTGCGTTACGGCAGCAAGTACTATCTCGCCATCATGGACAAAAAATATGCA
AAGTGTCTGCAGAAAATCGATAAAGACGACGTGAACGGAAATTACGAAAAGATTAATTATAAGCTGCT
GCCAGGGCCCAACAAGATGTTACCGAAAGTATTTTTTTCCAAAAAATGGATGGCATACTATAACCCGA
GCGAGGATATACAGAAGATTTACAAAAATGGGACCTTCAAAAAGGGGGATATGTTCAATCTGAATGAC
TGCCACAAACTGATCGATTTTTTTAAAGATAGCATCAGCCGTTATCCTAAATGGTCAAACGCGTATGAT
TTTAATTTCTCCGAAACGGAGAAATATAAAGACATTGCTGGTTTCTATCGCGAAGTCGAAGAACAGGG
TTATAAAGTTAGCTTTGAATCGGCCAGCAAGAAAGAGGTTGATAAACTGGTGGAGGAGGGTAAGCTG
TATATGTTTCAGATTTATAACAAAGACTTTAGCGACAAAAGCCACGGTACTCCTAATCTGCATACGATG
TACTTTAAACTGCTGTTTGATGAGAATAACCACGGCCAAATCCGTCTCTCCGGTGGAGCAGAACTTTT
TATGCGGCGTGCGAGCCTAAAAAAGGAAGAACTGGTGGTGCATCCCGCCAACAGCCCGATTGCTAAC
AAAAATCCAGATAATCCTAAGAAGACCACCACACTGTCGTACGATGTCTATAAGGATAAACGTTTCTC
GGAAGACCAGTATGAATTGCATATACCGATAGCAATTAATAAATGCCCAAAAAACATTTTCAAAATCAA
CACTGAAGTTCGTGTGCTGCTGAAACATGATGATAATCCGTATGTGATCGGAATTGACCGTGGGGAGA
[0260] GAAATCTGCTGTATATTGTAGTCGTTGATGGCAAGGGCAACATCGTTGAGCAGTATAGCCTGAATGAAA
TAATTAATAATTTTAACGGTATACGTATTAAAACCGACTATCATAGCCTGCTGGATAAAAAGGAGAAAG
AGCGTTTTGAGGCACGCCAAAATTGGACGAGCATCGAAAACATCAAGGAACTGAAGGCAGGATATAT
CAGCCAAGTAGTCCATAAAATCTGTGAACTGGTGGAGAAGTACGACGCTGTCATTGCCCTGGAAGAC
CTCAATAGCGGCTTTAAAAACAGCCGGGTGAAGGTGGAGAAACAGGTATACCAAAAGTTTGAAAAGA
TGCTCATTGATAAGCTGAACTATATGGTTGATAAAAAGAGCAACCCGTGCGCCACTGGCGGTGCACTG
AAAGGGTACCAAATTACCAATAAATTTGAAAGCTTTAAAAGCATGAGCACGCAGAATGGGTTTATTTT
TTATATACCAGCATGGCTGACGAGCAAGATTGACCCCAGCACTGGTTTTGTCAATCTGCTGAAAACCA
AATACACAAGCATTGCGGATAGCAAAAAATTTATTTCGAGCTTCGACCGTATTATGTATGTTCCGGAGG
AAGATCTGTTTGAATTTGCCCTGGATTATAAAAACTTCAGCCGCACCGATGCAGATTATATCAAAAAAT
GGAAGCTGTACAGTTATGGTAATCGTATACGTATCTTCCGTAATCCGAAGAAAAACAATGTGTTCGATT
GGGAAGAGGTCTGTCTGACCAGCGCGTATAAAGAACTGTTCAACAAGTACGGAATAAATTATCAGCA
AGGTGACATTCGCGCACTGCTGTGTGAACAGTCAGATAAAGCATTTTATAGCAGCTTTATGGCGCTGAT
GAGCCTGATGCTCCAGATGCGCAACAGCATAACCGGTCGCACAGATGTTGACTTTCTGATCAGCCCTG
TGAAGAATAGCGACGGCATCTTCTACGATTCCAGGAACTATGAAGCACAGGAAAACGCTATTCTGCCT
AAAAATGCCGATGCCAACGGCGCCTATAATATTGCACGGAAGGTTCTGTGGGCGATTGGACAGTTCAA
GAAAGCGGAAGATGAGAAGCTGGATAAGGTAAAAATTGCTATTAGCAATAAGGAATGGCTGGAGTAC
GCACAGACATCGGTTAAACAC*GGTAGT*aaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaag*GGAGCGGC*
*CGCACTCGAG***CACCACCACCACCACCAC**TGA (SEQ ID NO:426)

[0261] AAS1885:pET-28b-heAsCas12a(E174R/S542R/K548R)-NLS-6xHis

[0262] 人类密码子优化的氨基酸球菌种BV3L6 Cas12a(AsCas12a)为黑色,eAsCas12a的修饰的密码子(E174R/S542R/K548R)为双下划线的小写字母,具有沉默突变以移除NcoI位点的密码子为双下划线的 大 写 字 母 ,插入的甘氨酸为点 式 下 划 线 的 ,核质蛋白NLS(aaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaag,SEQ ID NO:21)为小写字母,接头序列为斜体,6xHis为粗体

ATGGGGACACAGTTCGAGGGCTTTACCAACCTGTATCAGGTGAGCAAGACACTGCGGTTTGAGCTGA
TCCCACAGGGCAAGACCCTGAAGCACATCCAGGAGCAGGGCTTCATCGAGGAGGACAAGGCCCGCA
ATGATCACTACAAGGAGCTGAAGCCCATCATCGATCGGATCTACAAGACCTATGCCGACCAGTGCCTG
CAGCTGGTGCAGCTGGATTGGGAGAACCTGAGCGCCGCCATCGACTCCTATAGAAAGGAGAAAACCG
AGGAGACAAGGAACGCCCTGATCGAGGAGCAGGCCACATATCGCAATGCCATCCACGACTACTTCAT
CGGCCGGACAGACAACCTGACCGATGCCATCAATAAGAGACACGCCGAGATCTACAAGGGCCTGTTC
AAGGCCGAGCTGTTTAATGGCAAGGTGCTGAAGCAGCTGGGCACCGTGACCACAACCGAGCACGAG
AACGCCCTGCTGCGGAGCTTCGACAAGTTTACAACCTACTTCTCCGGCTTTTATagaAACAGGAAGAAC
GTGTTCAGCGCCGAGGATATCAGCACAGCCATCCCACACCGCATCGTGCAGGACAACTTCCCCAAGT
TTAAGGAGAATTGTCACATCTTCACACGCCTGATCACCGCCGTGCCCAGCCTGCGGGAGCACTTTGAG
AACGTGAAGAAGGCCATCGGCATCTTCGTGAGCACCTCCATCGAGGAGGTGTTTTCCTTCCCTTTTTAT
AACCAGCTGCTGACACAGACCCAGATCGACCTGTATAACCAGCTGCTGGGAGGAATCTCTCGGGAGG
CAGGCACCGAGAAGATCAAGGGCCTGAACGAGGTGCTGAATCTGGCCATCCAGAAGAATGATGAGA
CAGCCCACATCATCGCCTCCCTGCCACACAGATTCATCCCCCTGTTTAAGCAGATCCTGTCCGATAGGA
ACACCCTGTCTTTCATCCTGGAGGAGTTTAAGAGCGACGAGGAAGTGATCCAGTCCTTCTGCAAGTAC
AAGACACTGCTGAGAAACGAGAACGTGCTGGAGACAGCCGAGGCCCTGTTTAACGAGCTGAACAGC
ATCGACCTGACACACATCTTCATCAGCCACAAGAAGCTGGAGACAATCAGCAGCGCCCTGTGCGACC
ACTGGGATACACTGAGGAATGCCCTGTATGAGCGGAGAATCTCCGAGCTGACAGGCAAGATCACCAA
GTCTGCCAAGGAGAAGGTGCAGCGCAGCCTGAAGCACGAGGATATCAACCTGCAGGAGATCATCTCT
GCCGCAGGCAAGGAGCTGAGCGAGGCCTTCAAGCAGAAAACCAGCGAGATCCTGTCCCACGCACAC
GCCGCCCTGGATCAGCCACTGCCTACAACCCTGAAGAAGCAGGAGGAGAAGGAGATCCTGAAGTCT
CAGCTGGACAGCCTGCTGGGCCTGTACCACCTGCTGGACTGGTTTGCCGTGGATGAGTCCAACGAGG
[0263] TGGACCCCGAGTTCTCTGCCCGGCTGACCGGCATCAAGCTGGAGATGGAGCCTTCTCTGAGCTTCTAC
AACAAGGCCAGAAATTATGCCACCAAGAAGCCCTACTCCGTGGAGAAGTTCAAGCTGAACTTTCAGA
TGCCTACACTGGCCagaGGCTGGGACGTGAATagaGAGAAGAACAATGGCGCCATCCTGTTTGTGAAGA
ACGGCCTGTACTATCTGGGCATCATGCCAAAGCAGAAGGGCAGGTATAAGGCCCTGAGCTTCGAGCC
CACAGAGAAAACCAGCGAGGGCTTTGATAAGATGTACTATGACTACTTCCCTGATGCCGCCAAGATGA
TCCCAAAGTGCAGCACCCAGCTGAAGGCCGTGACAGCCCACTTTCAGACCCACACAACCCCCATCCT
GCTGTCCAACAATTTCATCGAGCCTCTGGAGATCACAAAGGAGATCTACGACCTGAACAATCCTGAG
AAGGAGCCAAAGAAGTTTCAGACAGCCTACGCCAAGAAAACCGGCGACCAGAAGGGCTACAGAGA
GGCCCTGTGCAAGTGGATCGACTTCACAAGGGATTTTCTGTCCAAGTATACCAAGACAACCTCTATCG
ATCTGTCTAGCCTGCGGCCATCCTCTCAGTATAAGGACCTGGGCGAGTACTATGCCGAGCTGAATCCCC
TGCTGTACCACATCAGCTTCCAGAGAATCGCCGAGAAGGAGATCATGGATGCCGTGGAGACAGGCAA
GCTGTACCTGTTCCAGATCTATAACAAGGACTTTGCCAAGGGCCACCACGGCAAGCCTAATCTGCACA
CACTGTATTGGACCGGCCTGTTTTCTCCAGAGAACCTGGCCAAGACAAGCATCAAGCTGAATGGCCA
GGCCGAGCTGTTCTACCGCCCTAAGTCCAGGATGAAGAGGATGGCACACCGGCTGGGAGAGAAGATG
CTGAACAAGAAGCTGAAGGATCAGAAAACCCCAATCCCCGACACCCTGTACCAGGAGCTGTACGACT
ATGTGAATCACAGACTGTCCCACGACCTGTCTGATGAGGCCAGGGCCCTGCTGCCCAACGTGATCAC
CAAGGAGGTGTCTCACGAGATCATCAAGGATAGGCGCTTTACCAGCGACAAGTTCTTTTTCCACGTGC
CTATCACACTGAACTATCAGGCCGCCAATTCCCCATCTAAGTTCAACCAGAGGGTGAATGCCTACCTG
AAGGAGCACCCCGAGACACCTATCATCGGCATCGATCGGGGCGAGAGAAACCTGATCTATATCACAGT
GATCGACTCCACCGGCAAGATCCTGGAGCAGCGGAGCCTGAACACCATCCAGCAGTTTGATTACCAG
AAGAAGCTGGACAACAGGGAGAAGGAGAGGGTGGCAGCAAGGCAGGCCTGGTCTGTGGTGGGCAC
AATCAAGGATCTGAAGCAGGGCTATCTGAGCCAGGTCATCCACGAGATCGTGGACCTGATGATCCACT
ACCAGGCCGTGGTGGTGCTGGAGAACCTGAATTTCGGCTTTAAGAGCAAGAGGACCGGCATCGCCGA

GAAGGCCGTGTACCAGCAGTTCGAGAAGATGCTGATCGATAAGCTGAATTGCCTGGTGCTGAAGGAC
TATCCAGCAGAGAAAGTGGGAGGCGTGCTGAACCCATACCAGCTGACAGACCAGTTCACCTCCTTTG
CCAAGATGGGCACCCAGTCTGGCTTCCTGTTTTACGTGCCTGCCCCATATACATCTAAGATCGATCCCC
TGACCGGCTTCGTGGACCCCTTCGTGTGGAAAACCATCAAGAATCACGAGAGCCGCAAGCACTTCCT
GGAGGGCTTCGACTTTCTGCACTACGACGTGAAAACCGGCGACTTCATCCTGCACTTTAAGATGAAC
AGAAATCTGTCCTTCCAGAGGGGCCTGCCCGGCTTTATGCCTGCATGGGATATCGTGTTCGAGAAGAA
CGAGACACAGTTTGACGCCAAGGGCACCCCTTTCATCGCCGGCAAGAGAATCGTGCCAGTGATCGAG
[0264] AATCACAGATTCACCGGCAGATACCGGGACCTGTATCCTGCCAACGAGCTGATCGCCCTGCTGGAGGA
GAAGGGCATCGTGTTCAGGGATGGCTCCAACATCCTGCCAAAGCTGCTGGAGAATGACGATTCTCAC
GCCATCGAC**ACG**ATGGTGGCCCTGATCCGCAGCGTGCTGCAGATGCGGAACTCCAATGCCGCCACAG
GCGAGGACTATATCAACAGCCCCGTGCGCGATCTGAATGGCGTGTGCTTCGACTCCCGGTTTCAGAAC
CCAGAGTGG**CCA**ATGGACGCCGATGCCAATGGCGCCTACCACATCGCCCTGAAGGGCCAGCTGCTGC
TGAATCACCTGAAGGAGAGCAAGGATCTGAAGCTGCAGAACGGCATCTCCAATCAGGACTGGCTGGC
CTACATCCAGGAGCTGCGCAACaaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaag*GGAGCGGCCGCA*
*CTCGAG***CACCACCACCACCACCAC**TGA SEQ ID NO:427)

[0265] AAS1880:pET-28b-hAsCas12a(E174R/S542R)-NLS-6xHis

[0266] 人类密码子优化的氨基酸球菌种BV3L6 Cas12a(AsCas12a)为黑色,修饰的密码子(E174R/S542R)为双下 划线的小写字母,具有沉默突变以移除NcoI位点的密码子为双下划线的大写字母,插入的甘氨酸为点 式下划线的,核质蛋白NLS(aaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaag,SEQ ID NO:21)为小写字母,接头序列为斜体,6xHis为粗体

ATGGGGACACAGTTCGAGGGCTTTACCAACCTGTATCAGGTGAGCAAGACACTGCGGTTTGAG
CTGATCCCACAGGGCAAGACCCTGAAGCACATCCAGGAGCAGGGCTTCATCGAGGAGGACAAGGCC
CGCAATGATCACTACAAGGAGCTGAAGCCCATCATCGATCGGATCTACAAGACCTATGCCGACCAGTG
CCTGCAGCTGGTGCAGCTGGATTGGGAGAACCTGAGCGCCGCCATCGACTCCTATAGAAAGGAGAAA
ACCGAGGAGACAAGGAACGCCCTGATCGAGGAGCAGGCCACATATCGCAATGCCATCCACGACTACT
TCATCGGCCGGACAGACAACCTGACCGATGCCATCAATAAGAGACACGCCGAGATCTACAAGGGCCT
GTTCAAGGCCGAGCTGTTTAATGGCAAGGTGCTGAAGCAGCTGGGCACCGTGACCACAACCGAGCA
CGAGAACGCCCTGCTGCGGAGCTTCGACAAGTTTACAACCTACTTCTCCGGCTTTTATagaAACAGGAA
GAACGTGTTCAGCGCCGAGGATATCAGCACAGCCATCCCACACCGCATCGTGCAGGACAACTTCCCC
AAGTTTAAGGAGAATTGTCACATCTTCACACGCCTGATCACCGCCGTGCCCAGCCTGCGGGAGCACTT
TGAGAACGTGAAGAAGGCCATCGGCATCTTCGTGAGCACCTCCATCGAGGAGGTGTTTTCCTTCCCTT
[0267] TTTATAACCAGCTGCTGACACAGACCCAGATCGACCTGTATAACCAGCTGCTGGGAGGAATCTCTCGG
GAGGCAGGCACCGAGAAGATCAAGGGCCTGAACGAGGTGCTGAATCTGGCCATCCAGAAGAATGAT
GAGACAGCCCACATCATCGCCTCCCTGCCACACAGATTCATCCCCCTGTTTAAGCAGATCCTGTCCGAT
AGGAACACCCTGTCTTTCATCCTGGAGGAGTTTAAGAGCGACGAGGAAGTGATCCAGTCCTTCTGCA
AGTACAAGACACTGCTGAGAAACGAGAACGTGCTGGAGACAGCCGAGGCCCTGTTTAACGAGCTGA
ACAGCATCGACCTGACACACATCTTCATCAGCCACAAGAAGCTGGAGACAATCAGCAGCGCCCTGTG
CGACCACTGGGATACACTGAGGAATGCCCTGTATGAGCGGAGAATCTCCGAGCTGACAGGCAAGATC
ACCAAGTCTGCCAAGGAGAAGGTGCAGCGCAGCCTGAAGCACGAGGATATCAACCTGCAGGAGATC
ATCTCTGCCGCAGGCAAGGAGCTGAGCGAGGCCTTCAAGCAGAAAACCAGCGAGATCCTGTCCCAC
GCACACGCCGCCCTGGATCAGCCACTGCCTACAACCCTGAAGAAGCAGGAGGAGAAGGAGATCCTG
AAGTCTCAGCTGGACAGCCTGCTGGGCCTGTACCACCTGCTGGACTGGTTTGCCGTGGATGAGTCCA
ACGAGGTGGACCCCGAGTTCTCTGCCCGGCTGACCGGCATCAAGCTGGAGATGGAGCCTTCTCTGAG

CTTCTACAACAAGGCCAGAAATTATGCCACCAAGAAGCCCTACTCCGTGGAGAAGTTCAAGCTGAAC
TTTCAGATGCCTACACTGGCCagaGGCTGGGACGTGAATAAGGAGAAGAACAATGGCGCCATCCTGTT
TGTGAAGAACGGCCTGTACTATCTGGGCATCATGCCAAAGCAGAAGGGCAGGTATAAGGCCCTGAGC
TTCGAGCCCACAGAGAAAACCAGCGAGGGCTTTGATAAGATGTACTATGACTACTTCCCTGATGCCGC
CAAGATGATCCCAAAGTGCAGCACCCAGCTGAAGGCCGTGACAGCCCACTTTCAGACCCACACAACC
CCCATCCTGCTGTCCAACAATTTCATCGAGCCTCTGGAGATCACAAAGGAGATCTACGACCTGAACAA
TCCTGAGAAGGAGCCAAAGAAGTTTCAGACAGCCTACGCCAAGAAAACCGGCGACCAGAAGGGCTA
CAGAGAGGCCCTGTGCAAGTGGATCGACTTCACAAGGGATTTTCTGTCCAAGTATACCAAGACAACC
TCTATCGATCTGTCTAGCCTGCGGCCATCCTCTCAGTATAAGGACCTGGGCGAGTACTATGCCGAGCTG
AATCCCCTGCTGTACCACATCAGCTTCCAGAGAATCGCCGAGAAGGAGATCATGGATGCCGTGGAGA
CAGGCAAGCTGTACCTGTTCCAGATCTATAACAAGGACTTTGCCAAGGGCCACCACGGCAAGCCTAAT
CTGCACACACTGTATTGGACCGGCCTGTTTTCTCCAGAGAACCTGGCCAAGACAAGCATCAAGCTGA
ATGGCCAGGCCGAGCTGTTCTACCGCCCTAAGTCCAGGATGAAGAGGATGGCACACCGGCTGGGAGA
GAAGATGCTGAACAAGAAGCTGAAGGATCAGAAAACCCCAATCCCCGACACCCTGTACCAGGAGCT
GTACGACTATGTGAATCACAGACTGTCCCACGACCTGTCTGATGAGGCCAGGGCCCTGCTGCCCAAC
GTGATCACCAAGGAGGTGTCTCACGAGATCATCAAGGATAGGCGCTTTACCAGCGACAAGTTCTTTTT
CCACGTGCCTATCACACTGAACTATCAGGCCGCCAATTCCCCATCTAAGTTCAACCAGAGGGTGAATG
CCTACCTGAAGGAGCACCCCGAGACACCTATCATCGGCATCGATCGGGGCGAGAGAAACCTGATCTAT
[0268] ATCACAGTGATCGACTCCACCGGCAAGATCCTGGAGCAGCGGAGCCTGAACACCATCCAGCAGTTTG
ATTACCAGAAGAAGCTGGACAACAGGGAGAAGGAGAGGGTGGCAGCAAGGCAGGCCTGGTCTGTG
GTGGGCACAATCAAGGATCTGAAGCAGGGCTATCTGAGCCAGGTCATCCACGAGATCGTGGACCTGA
TGATCCACTACCAGGCCGTGGTGGTGCTGGAGAACCTGAATTTCGGCTTTAAGAGCAAGAGGACCGG
CATCGCCGAGAAGGCCGTGTACCAGCAGTTCGAGAAGATGCTGATCGATAAGCTGAATTGCCTGGTG
CTGAAGGACTATCCAGCAGAGAAAGTGGGAGGCGTGCTGAACCCATACCAGCTGACAGACCAGTTC
ACCTCCTTTGCCAAGATGGGCACCCAGTCTGGCTTCCTGTTTTACGTGCCTGCCCCATATACATCTAAG
ATCGATCCCCTGACCGGCTTCGTGGACCCCTTCGTGTGGAAAACCATCAAGAATCACGAGAGCCGCA
AGCACTTCCTGGAGGGCTTCGACTTTCTGCACTACGACGTGAAAACCGGCGACTTCATCCTGCACTTT
AAGATGAACAGAAATCTGTCCTTCCAGAGGGGCCTGCCCGGCTTTATGCCTGCATGGGATATCGTGTT
CGAGAAGAACGAGACACAGTTTGACGCCAAGGGCACCCCTTTCATCGCCGGCAAGAGAATCGTGCC
AGTGATCGAGAATCACAGATTCACCGGCAGATACCGGGACCTGTATCCTGCCAACGAGCTGATCGCCC
TGCTGGAGGAGAAGGGCATCGTGTTCAGGGATGGCTCCAACATCCTGCCAAAGCTGCTGGAGAATGA
CGATTCTCACGCCATCGAC**ACG**ATGGTGGCCCTGATCCGCAGCGTGCTGCAGATGCGGAACTCCAATG
CCGCCACAGGCGAGGACTATATCAACAGCCCCGTGCGCGATCTGAATGGCGTGTGCTTCGACTCCCGG
TTTCAGAACCCAGAGTGG**CCA**ATGGACGCCGATGCCAATGGCGCCTACCACATCGCCCTGAAGGGCC
AGCTGCTGCTGAATCACCTGAAGGAGAGCAAGGATCTGAAGCTGCAGAACGGCATCTCCAATCAGGA
CTGGCTGGCCTACATCCAGGAGCTGCGCAACaaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaagGGA
GCGGCCGCACTCGAG**CACCACCACCACCACCAC**TGA SEQ ID NO:428)

[0269] AAS1935:pET-28b-heAsCas12a-HF1(E174R/N282A/S542R/K548R)-NLS-6xHis

[0270] 人类密码子优化的氨基酸球菌种BV3L6 Cas12a(AsCas12a)为黑色,eAsCas12a-HF1的修饰的密码子(E174R/N282A/S542R/K548R)为双下划线的小写字母,具有沉默突变以移除NcoI位点的密码子为双下划线的大写字母,插入的甘氨酸为点式下划线的,核质蛋白NLS(aaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaag,SEQ ID NO:21)为小写字母,接头序列为斜体,6xHis为粗体

[0271] ATGGGGACACAGTTCGAGGGCTTTACCAACCTGTATCAGGTGAGCAAGACACTGCGGTTTGAG

CTGATCCCACAGGGCAAGACCCTGAAGCACATCCAGGAGCAGGGCTTCATCGAGGAGGACAAGGCC
CGCAATGATCACTACAAGGAGCTGAAGCCCATCATCGATCGGATCTACAAGACCTATGCCGACCAGTG
CCTGCAGCTGGTGCAGCTGGATTGGGAGAACCTGAGCGCCGCCATCGACTCCTATAGAAAGGAGAAA
ACCGAGGAGACAAGGAACGCCCTGATCGAGGAGCAGGCCACATATCGCAATGCCATCCACGACTACT
TCATCGGCCGGACAGACAACCTGACCGATGCCATCAATAAGAGACACGCCGAGATCTACAAGGGCCT
GTTCAAGGCCGAGCTGTTTAATGGCAAGGTGCTGAAGCAGCTGGGCACCGTGACCACAACCGAGCA
CGAGAACGCCCTGCTGCGGAGCTTCGACAAGTTTACAACCTACTTCTCCGGCTTTTATagaAACAGGAA
GAACGTGTTCAGCGCCGAGGATATCAGCACAGCCATCCCACACCGCATCGTGCAGGACAACTTCCCC
AAGTTTAAGGAGAATTGTCACATCTTCACACGCCTGATCACCGCCGTGCCCAGCCTGCGGGAGCACTT
TGAGAACGTGAAGAAGGCCATCGGCATCTTCGTGAGCACCTCCATCGAGGAGGTGTTTTCCTTCCCTT
TTTATAACCAGCTGCTGACACAGACCCAGATCGACCTGTATAACCAGCTGCTGGGAGGAATCTCTCGG
GAGGCAGGCACCGAGAAGATCAAGGGCCTGAACGAGGTGCTGgccCTGGCCATCCAGAAGAATGATG
AGACAGCCCACATCATCGCCTCCCTGCCACACAGATTCATCCCCCTGTTTAAGCAGATCCTGTCCGATA
GGAACACCCTGTCTTTCATCCTGGAGGAGTTTAAGAGCGACGAGGAAGTGATCCAGTCCTTCTGCAA
GTACAAGACACTGCTGAGAAACGAGAACGTGCTGGAGACAGCCGAGGCCCTGTTTAACGAGCTGAA
CAGCATCGACCTGACACACATCTTCATCAGCCACAAGAAGCTGGAGACAATCAGCAGCGCCCTGTGC
GACCACTGGGATACACTGAGGAATGCCCTGTATGAGCGGAGAATCTCCGAGCTGACAGGCAAGATCA
CCAAGTCTGCCAAGGAGAAGGTGCAGCGCAGCCTGAAGCACGAGGATATCAACCTGCAGGAGATCA
TCTCTGCCGCAGGCAAGGAGCTGAGCGAGGCCTTCAAGCAGAAAACCAGCGAGATCCTGTCCCACG
CACACGCCGCCCTGGATCAGCCACTGCCTACAACCCTGAAGAAGCAGGAGGAGAAGGAGATCCTGA
AGTCTCAGCTGGACAGCCTGCTGGGCCTGTACCACCTGCTGGACTGGTTTGCCGTGGATGAGTCCAA
CGAGGTGGACCCCGAGTTCTCTGCCCGGCTGACCGGCATCAAGCTGGAGATGGAGCCTTCTCTGAGC
[0272] TTCTACAACAAGGCCAGAAATTATGCCACCAAGAAGCCCTACTCCGTGGAGAAGTTCAAGCTGAACT
TTCAGATGCCTACACTGGCCagaGGCTGGGACGTGAATagaGAGAAGAACAATGGCGCCATCCTGTTTG
TGAAGAACGGCCTGTACTATCTGGGCATCATGCCAAAGCAGAAGGGCAGGTATAAGGCCCTGAGCTT
CGAGCCCACAGAGAAAACCAGCGAGGGCTTTGATAAGATGTACTATGACTACTTCCCTGATGCCGCCA
AGATGATCCCAAAGTGCAGCACCCAGCTGAAGGCCGTGACAGCCCACTTTCAGACCCACACAACCCC
CATCCTGCTGTCCAACAATTTCATCGAGCCTCTGGAGATCACAAAGGAGATCTACGACCTGAACAATC
CTGAGAAGGAGCCAAAGAAGTTTCAGACAGCCTACGCCAAGAAAACCGGCGACCAGAAGGGCTACA
GAGAGGCCCTGTGCAAGTGGATCGACTTCACAAGGGATTTTCTGTCCAAGTATACCAAGACAACCTCT
ATCGATCTGTCTAGCCTGCGGCCATCCTCTCAGTATAAGGACCTGGGCGAGTACTATGCCGAGCTGAAT
CCCCTGCTGTACCACATCAGCTTCCAGAGAATCGCCGAGAAGGAGATCATGGATGCCGTGGAGACAG
GCAAGCTGTACCTGTTCCAGATCTATAACAAGGACTTTGCCAAGGGCCACCACGGCAAGCCTAATCTG
CACACACTGTATTGGACCGGCCTGTTTTCTCCAGAGAACCTGGCCAAGACAAGCATCAAGCTGAATG
GCCAGGCCGAGCTGTTCTACCGCCCTAAGTCCAGGATGAAGAGGATGGCACACCGGCTGGGAGAGA
AGATGCTGAACAAGAAGCTGAAGGATCAGAAAACCCCAATCCCCGACACCCTGTACCAGGAGCTGTA
CGACTATGTGAATCACAGACTGTCCCACGACCTGTCTGATGAGGCCAGGGCCCTGCTGCCCAACGTG
ATCACCAAGGAGGTGTCTCACGAGATCATCAAGGATAGGCGCTTTACCAGCGACAAGTTCTTTTTCCA
CGTGCCTATCACACTGAACTATCAGGCCGCCAATTCCCCATCTAAGTTCAACCAGAGGGTGAATGCCT
ACCTGAAGGAGCACCCCGAGACACCTATCATCGGCATCGATCGGGGCGAGAGAAACCTGATCTATATC
ACAGTGATCGACTCCACCGGCAAGATCCTGGAGCAGCGGAGCCTGAACACCATCCAGCAGTTTGATT
ACCAGAAGAAGCTGGACAACAGGGAGAAGGAGAGGGTGGCAGCAAGGCAGGCCTGGTCTGTGGTG
GGCACAATCAAGGATCTGAAGCAGGGCTATCTGAGCCAGGTCATCCACGAGATCGTGGACCTGATGA
TCCACTACCAGGCCGTGGTGGTGCTGGAGAACCTGAATTTCGGCTTTAAGAGCAAGAGGACCGGCAT
CGCCGAGAAGGCCGTGTACCAGCAGTTCGAGAAGATGCTGATCGATAAGCTGAATTGCCTGGTGCTG

[0273] AAGGACTATCCAGCAGAGAAAGTGGGAGGCGTGCTGAACCCATACCAGCTGACAGACCAGTTCACCT
CCTTTGCCAAGATGGGCACCCAGTCTGGCTTCCTGTTTTACGTGCCTGCCCCATATACATCTAAGATCG
ATCCCCTGACCGGCTTCGTGGACCCCTTCGTGTGGAAAACCATCAAGAATCACGAGAGCCGCAAGCA
CTTCCTGGAGGGCTTCGACTTTCTGCACTACGACGTGAAAACCGGCGACTTCATCCTGCACTTTAAGA
TGAACAGAAATCTGTCCTTCCAGAGGGGCCTGCCCGGCTTTATGCCTGCATGGGATATCGTGTTCGAG
AAGAACGAGACACAGTTTGACGCCAAGGGCACCCCTTTCATCGCCGGCAAGAGAATCGTGCCAGTGA
TCGAGAATCACAGATTCACCGGCAGATACCGGGACCTGTATCCTGCCAACGAGCTGATCGCCCTGCTG
GAGGAGAAGGGCATCGTGTTCAGGGATGGCTCCAACATCCTGCCAAAGCTGCTGGAGAATGACGATT
CTCACGCCATCGAC**ACG**ATGGTGGCCCTGATCCGCAGCGTGCTGCAGATGCGGAACTCCAATGCCGC
CACAGGCGAGGACTATATCAACAGCCCCGTGCGCGATCTGAATGGCGTGTGCTTCGACTCCCGGTTTC
AGAACCCAGAGTGG**CCA**ATGGACGCCGATGCCAATGGCGCCTACCACATCGCCCTGAAGGGCCAGCT
GCTGCTGAATCACCTGAAGGAGAGCAAGGATCTGAAGCTGCAGAACGGCATCTCCAATCAGGACTGG
CTGGCCTACATCCAGGAGCTGCGCAAC*aaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaag*GGAGCGG
*CCGCACTCGAG***CACCACCACCACCACCAC**TGA SEQ ID NO: 429)

[0274] SQT1665 pCAG-人类LbCpf1-NLS-3xHA的核苷酸序列

[0275] 人类密码子优化的LbCpf1为正常自体,(nts 1-3684),NLS(aaaaggccggcggccacg aaaaaggccggccaggcaaaaaagaaaaag,SEQ ID NO:21)为小写字母,3xHA标签(TACCCATACGAT GTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCC,SEQ ID NO:5)为粗体,接头序列为斜体

[0276] ATGAGCAAGCTGGAGAAGTTTACAAACTGCTACTCCCTGTCTAAGACCCTGAGGTTCAAGGCCA
TCCCTGTGGGCAAGACCCAGGAGAACATCGACAATAAGCGGCTGCTGGTGGAGGACGAGAAGAGAGC
CGAGGATTATAAGGGCGTGAAGAAGCTGCTGGATCGCTACTATCTGTCTTTTATCAACGACGTGCTGCA
CAGCATCAAGCTGAAGAATCTGAACAATTACATCAGCCTGTTCCGGAAGAAAACCAGAACCGAGAAGG
AGAATAAGGAGCTGGAGAACCTGGAGATCAATCTGCGGAAGGAGATCGCCAAGGCCTTCAAGGGCAA
CGAGGGCTACAAGTCCCTGTTTAAGAAGGATATCATCGAGACAATCCTGCCAGAGTTCCTGGACGATAA
GGACGAGATCGCCCTGGTGAACAGCTTCAATGGCTTTACCACAGCCTTCACCGGCTTCTTTGATAACAG
AGAGAATATGTTTTCCGAGGAGGCCAAGAGCACATCCATCGCCTTCAGGTGTATCAACGAGAATCTGAC
CCGCTACATCTCTAATATGGACATCTTCGAGAAGGTGGACGCCATCTTTGATAAGCACGAGGTGCAGGA
GATCAAGGAGAAGATCCTGAACAGCGACTATGATGTGGAGGATTTCTTTGAGGGCGAGTTCTTTAACTT
TGTGCTGACACAGGAGGGCATCGACGTGTATAACGCCATCATCGGCGGCTTCGTGACCGAGAGCGGCG
AGAAGATCAAGGGCCTGAACGAGTACATCAACCTGTATAATCAGAAAACCAAGCAGAAGCTGCCTAAG
TTTAAGCCACTGTATAAGCAGGTGCTGAGCGATCGGGAGTCTCTGAGCTTCTACGGCGAGGGCTATACA
TCCGATGAGGAGGTGCTGGAGGTGTTTAGAAACACCCTGAACAAGAACAGCGAGATCTTCAGCTCCAT
CAAGAAGCTGGAGAAGCTGTTCAAGAATTTTGACGAGTACTCTAGCGCCGGCATCTTTGTGAAGAACG
GCCCCGCCATCAGCACAATCTCCAAGGATATCTTCGGCGAGTGGAACGTGATCCGGGACAAGTGGAAT
GCCGAGTATGACGATATCCACCTGAAGAAGAAGGCCGTGGTGACCGAGAAGTACGAGGACGATCGGA
GAAAGTCCTTCAAGAAGATCGGCTCCTTTTCTCTGGAGCAGCTGCAGGAGTACGCCGACGCCGATCTG
TCTGTGGTGGAGAAGCTGAAGGAGATCATCATCCAGAAGGTGGATGAGATCTACAAGGTGTATGGCTCC
TCTGAGAAGCTGTTCGACGCCGATTTTGTGCTGGAGAAGAGCCTGAAGAAGAACGACGCCGTGGTGG
CCATCATGAAGGACCTGCTGGATTCTGTGAAGAGCTTCGAGAATTACATCAAGGCCTTCTTTGGCGAGG
GCAAGGAGACAAACAGGGACGAGTCCTTCTATGGCGATTTTGTGCTGGCCTACGACATCCTGCTGAAG
GTGGACCACATCTACGATGCCATCCGCAATTATGTGACCCAGAAGCCCTACTCTAAGGATAAGTTCAAG
CTGTATTTTCAGAACCCTCAGTTCATGGGCGGCTGGGACAAGGATAAGGAGACAGACTATCGGGCCAC
CATCCTGAGATACGGCTCCAAGTACTATCTGGCCATCATGGATAAGAAGTACGCCAAGTGCCTGCAGAA

GATCGACAAGGACGATGTGAACGGCAATTACGAGAAGATCAACTATAAGCTGCTGCCCGGCCCTAATAA
GATGCTGCCAAAGGTGTTCTTTTCTAAGAAGTGGATGGCCTACTATAACCCCAGCGAGGACATCCAGAA
GATCTACAAGAATGGCACATTCAAGAAGGGCGATATGTTTAACCTGAATGACTGTCACAAGCTGATCGA
CTTCTTTAAGGATAGCATCTCCCGGTATCCAAAGTGGTCCAATGCCTACGATTTCAACTTTTCTGAGACA
GAGAAGTATAAGGACATCGCCGGCTTTTACAGAGAGGTGGAGGAGCAGGGCTATAAGGTGAGCTTCGA
GTCTGCCAGCAAGAAGGAGGTGGATAAGCTGGTGGAGGAGGGCAAGCTGTATATGTTCCAGATCTATA
ACAAGGACTTTTCCGATAAGTCTCACGGCACACCCAATCTGCACACCATGTACTTCAAGCTGCTGTTTG
ACGAGAACAATCACGGACAGATCAGGCTGAGCGGAGGAGCAGAGCTGTTCATGAGGCGCGCCTCCCT
GAAGAAGGAGGAGCTGGTGGTGCACCCAGCCAACTCCCCTATCGCCAACAAGAATCCAGATAATCCCA
AGAAAACCACAACCCTGTCCTACGACGTGTATAAGGATAAGAGGTTTTCTGAGGACCAGTACGAGCTG
CACATCCCAATCGCCATCAATAAGTGCCCCAAGAACATCTTCAAGATCAATACAGAGGTGCGCGTGCTG
CTGAAGCACGACGATAACCCCTATGTGATCGGCATCGATAGGGGCGAGCGCAATCTGCTGTATATCGTG
GTGGTGGACGGCAAGGGCAACATCGTGGAGCAGTATTCCCTGAACGAGATCATCAACAACTTCAACGG
CATCAGGATCAAGACAGATTACCACTCTCTGCTGGACAAGAAGGAGAAGGAGAGGTTCGAGGCCCGC
CAGAACTGGACCTCCATCGAGAATATCAAGGAGCTGAAGGCCGGCTATATCTCTCAGGTGGTGCACAA
[0277] GATCTGCGAGCTGGTGGAGAAGTACGATGCCGTGATCGCCCTGGAGGACCTGAACTCTGGCTTTAAGA
ATAGCCGCGTGAAGGTGGAGAAGCAGGTGTATCAGAAGTTCGAGAAGATGCTGATCGATAAGCTGAAC
TACATGGTGGACAAGAAGTCTAATCCTTGTGCAACAGGCGGCGCCCTGAAGGGCTATCAGATCACCAAT
AAGTTCGAGAGCTTTAAGTCCATGTCTACCCAGAACGGCTTCATCTTTTACATCCCTGCCTGGCTGACAT
CCAAGATCGATCCATCTACCGGCTTTGTGAACCTGCTGAAAACCAAGTATACCAGCATCGCCGATTCCA
AGAAGTTCATCAGCTCCTTTGACAGGATCATGTACGTGCCCGAGGAGGATCTGTTCGAGTTTGCCCTGG
ACTATAAGAACTTCTCTCGCACAGACGCCGATTACATCAAGAAGTGGAAGCTGTACTCCTACGGCAACC
GGATCAGAATCTTCCGGAATCCTAAGAAGAACAACGTGTTCGACTGGGAGGAGGTGTGCCTGACCAGC
GCCTATAAGGAGCTGTTCAACAAGTACGGCATCAATTATCAGCAGGGCGATATCAGAGCCCTGCTGTGC
GAGCAGTCCGACAAGGCCTTCTACTCTAGCTTTATGGCCCTGATGAGCCTGATGCTGCAGATGCGGAAC
AGCATCACAGGCCGCACCGACGTGGATTTTCTGATCAGCCCTGTGAAGAACTCCGACGGCATCTTCTAC
GATAGCCGGAACTATGAGGCCCAGGAGAATGCCATCCTGCCAAAGAACGCCGACGCCAATGGCGCCTA
TAACATCGCCAGAAAGGTGCTGTGGGCCATCGGCCAGTTCAAGAAGGCCGAGGACGAGAAGCTGGAT
AAGGTGAAGATCGCCATCTCTAACAAGGAGTGGCTGGAGTACGCCCAGACCAGCGTGAAGCACaaaaggc
cggcggccacgaaaaaggccggccaggcaaaaaagaaaaag*GGATCC***TACCCATACGATGTTCCAGATTACGCTTATCCCT**
ACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCCTAA (SEQ ID NO:10)

[0278] LbCpf1-NLS-3xHA的氨基酸序列

[0279] LbCpf1为正常文本(AAs 1-1228),NLS(krpaatkkagqakkkk,SEQ ID NO:7)为小写字母,3xHA标签(YPYDVPDYAYPYDVPDYAYPYDVPDYA,SEQ ID NO:8)为粗体

MSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRAEDYKGVKKLLDRYYLSFINDVLH
SIKLKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAFKGNEGYKSLFKKDIIETILPEFLDDKDEIALV
NSFNGFTTAFTGFFDNRENMFSEEAKSTSIAFRCINENLTRYISNMDIFEKVDAIFDKHEVQEIKEKILNSDYD
VEDFFEGEFFNFVLTQEGIDVYNAIIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDRESLS
FYGEGYTSDEEVLEVFRNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDK
[0280] WNAEYDDIHLKKKAVVTEKYEDDRRKSFKKIGSFSLEQLQEYADADLSVVEKLKEIIIQKVDEIYKVYGSSE
KLFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEGKETNRDESFYGDFVLAYDILLKVDHIY
DAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAKCLQKIDKDD
VNGNYEKINYKLLPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNLNDCHKLIDFFKDSI
SRYPKWSNAYDFNFSETEKYKDIAGFYREVEEQGYKVSFESASKKEVDKLVEEGKLYMFQIYNKDFSDKSH

[0281]

GTPNLHTMYFKLLFDENNHGQIRLSGGAELFMRRASLKKEELVVHPANSPIANKNPDNPKKTTTLSYDVYK
DKRFSEDQYELHIPIAINKCPKNIFKINTEVRVLLKHDDNPYVIGIDRGERNLLYIVVVDGKGNIVEQYSLNEI
INNFNGIRIKTDYHSLLDKKEKERFEARQNWTSIENIKELKAGYISQVVHKICELVEKYDAVIALEDLNSGFK
NSRVKVEKQVYQKFEKMLIDKLNYMVDKKSNPCATGGALKGYQITNKFESFKSMSTQNGFIFYIPAWLTSK
IDPSTGFVNLLKTKYTSIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRN
PKKNNVFDWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDV
DFLISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIGQFKKAEDEKLDKVKIAISNKEW
LEYAQTSVKHkrpaatkkagqakkkkGS**YPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO:11)**

[0282] AAS1472 pCAG-人类FnCpf1-NLS-3xHA的核苷酸序列

[0283] 人类密码子优化的FnCpf1为正常字体,nts 1-3900),NLS(aaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaag,SEQ ID NO:21)为小写字母,3xHA标签(TACCCATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCC,SEQ ID NO:5)为粗体

[0284]

ATGAGCATCTACCAGGAGTTCGTCAACAAGTATTCACTGAGTAAGACACTGCGGTTCGAGCTGAT
CCCACAGGGCAAGACACTGGAGAACATCAAGGCCCGAGGCCTGATTCTGGACGATGAGAAGCGGGCA
AAAGACTATAAGAAAGCCAAGCAGATCATTGATAAATACCACCAGTTCTTTATCGAGGAAATTCTGAGC
TCCGTGTGCATCAGTGAGGATCTGCTGCAGAATTACTCAGACGTGTACTTCAAGCTGAAGAAGAGCGA
CGATGACAACCTGCAGAAGGACTTCAAGTCCGCCAAGGACACCATCAAGAAACAGATTAGCGAGTACA
TCAAGGACTCCGAAAAGTTTAAAAATCTGTTCAACCAGAATCTGATCGATGCTAAGAAAGGCCAGGAG
TCCGACCTGATCCTGTGGCTGAAACAGTCTAAGGACAATGGGATTGAACTGTTCAAGGCTAACTCCGAT
ATCACTGATATTGACGAGGCACTGGAAATCATCAAGAGCTTCAAGGGATGGACCACATACTTTAAAGGC
TTCCACGAGAACCGCAAGAACGTGTACTCCAGCAACGACATTCCTACCTCCATCATCTACCGAATCGTC
GATGACAATCTGCCAAAGTTCCTGGAGAACAAGGCCAAATATGAATCTCTGAAGGACAAAGCTCCCGA
GGCAATTAATTACGAACAGATCAAGAAAGATCTGGCTGAGGAACTGACATTCGATATCGACTATAAGAC
TAGCGAGGTGAACCAGAGGGTCTTTTCCCTGGACGAGGTGTTTGAAATCGCCAATTTCAACAATTACCT
GAACCAGTCCGGCATTACTAAATTCAATACCATCATTGGCGGGAAGTTTGTGAACGGGGAGAATACCAA
GCGCAAGGGAATTAACGAATACATCAATCTGTATAGCCAGCAGATCAACGACAAAACTCTGAAGAAATA
CAAGATGTCTGTGCTGTTCAAACAGATCCTGAGTGATACCGAGTCCAAGTCTTTTGTCATTGATAAACTG
GAAGATGACTCAGACGTGGTCACTACCATGCAGAGCTTTTATGAGCAGATCGCCGCTTTCAAGACAGTG
GAGGAAAAATCTATTAAGGAAACTCTGAGTCTGCTGTTCGATGACCTGAAAGCCCAGAAGCTGGACCT
GAGTAAGATCTACTTCAAAAACGATAAGAGTCTGACAGACCTGTCACAGCAGGTGTTTGATGACTATTC
CGTGATTGGGACCGCCGTCCTGGAGTACATTACACAGCAGATCGCTCCAAAGAACCTGGATAATCCCTC
TAAGAAAGAGCAGGAACTGATCGCTAAGAAAACCGAGAAGGCAAAATATCTGAGTCTGGAAACAATTA
AGCTGGCACTGGAGGAGTTCAACAAGCACAGGGATATTGACAAACAGTGCCGCTTTGAGGAAATCCTG
GCCAACTTCGCAGCCATCCCCATGATTTTTGATGAGATCGCCCAGAACAAAGACAATCTGGCTCAGATC
AGTATTAAGTACCAGAACCAGGGCAAGAAAGACCTGCTGCAGGCTTCAGCAGAAGATGACGTGAAAG
CCATCAAGGATCTGCTGGACCAGACCAACAATCTGCTGCACAAGCTGAAAATCTTCCATATTAGTCAGT
CAGAGGATAAGGCTAATATCCTGGATAAAGACGAACACTTCTACCTGGTGTTCGAGGAATGTTACTTCG
AGCTGGCAAACATTGTCCCCCTGTATAACAAGATTAGGAACTACATCACACAGAAGCCTTACTCTGACG
AGAAGTTTAAACTGAACTTCGAAAATAGTACCCTGGCCAACGGGTGGGATAAGAACAAGGAGCCTGAC
AACACAGCTATCCTGTTCATCAAGGATGACAAGTACTATCTGGGAGTGATGAATAAGAAAAACAATAAG
ATCTTCGATGACAAAGCCATTAAGGAGAACAAAGGGGAAGGATACAAGAAAATCGTGTATAAGCTGCT
GCCCGGCGCAAATAAGATGCTGCCTAAGGTGTTCTTCAGCGCCAAGAGTATCAAATTCTACAACCCATC
CGAGGACATCCTGCGGATTAGAAATCACTCAACACATACTAAGAACGGGAGCCCCCAGAAGGGATATG

[0285] AGAAATTTGAGTTCAACATCGAGGATTGCAGGAAGTTTATTGACTTCTACAAGCAGAGCATCTCCAAAC
ACCCTGAATGGAAGGATTTTGGCTTCCGGTTTTCCGACACACAGAGATATAACTCTATCGACGAGTTCTA
CCGCGAGGTGGAAAATCAGGGGTATAAGCTGACTTTTGAGAACATTTCTGAAAGTTACATCGACAGCGT
GGTCAATCAGGGAAAGCTGTACCTGTTCCAGATCTATAACAAAGATTTTTCAGCATACAGCAAGGGCAG
ACCAAACCTGCATACACTGTACTGGAAGGCCCTGTTCGATGAGAGGAATCTGCAGGACGTGGTCTATAA
ACTGAACGGAGAGGCCGAACTGTTTTACCGGAAGCAGTCTATTCCTAAGAAAATCACTCACCCAGCTA
AGGAGGCCATCGCTAACAAGAACAAGGACAATCCTAAGAAAGAGAGCGTGTTCGAATACGATCTGATT
AAGGACAAGCGGTTCACCGAAGATAAGTTCTTTTTCCATTGTCCAATCACCATTAACTTCAAGTCAAGC
GGCGCTAACAAGTTCAACGACGAGATCAATCTGCTGCTGAAGGAAAAAGCAAACGATGTGCACATCCT
GAGCATTGACCGAGGAGAGCGGCATCTGGCCTACTATACCCTGGTGGATGGCAAAGGGAATATCATTAA
GCAGGATACATTCAACATCATTGGCAATGACCGGATGAAAACCAACTACCACGATAAACTGGCTGCAAT
CGAGAAGGATAGAGACTCAGCTAGGAAGGACTGGAAGAAAATCAACAACATTAAGGAGATGAAGGAA
GGCTATCTGAGCCAGGTGGTCCATGAGATTGCAAAGCTGGTCATCGAATACAATGCCATTGTGGTGTTC
GAGGATCTGAACTTCGGCTTTAAGAGGGGGCGCTTTAAGGTGGAAAAACAGGTCTATCAGAAGCTGGA
GAAAATGCTGATCGAAAAGCTGAATTACCTGGTGTTTAAAGATAACGAGTTCGACAAGACCGGAGGCG
TCCTGAGAGCCTACCAGCTGACAGCTCCCTTTGAAACTTTCAAGAAAATGGGAAAACAGACAGGCATC
ATCTACTATGTGCCAGCCGGATTCACTTCCAAGATCTGCCCCGTGACCGGCTTTGTCAACCAGCTGTACC
CTAAATATGAGTCAGTGAGCAAGTCCCAGGAATTTTTCAGCAAGTTCGATAAGATCTGTTATAATCTGGA
CAAGGGGTACTTCGAGTTTTCCTTCGATTACAAGAACTTCGGCGACAAGGCCGCTAAGGGGAAATGGA
CCATTGCCTCCTTCGGATCTCGCCTGATCAACTTTCGAAATTCCGATAAAAACCACAATTGGGACACTAG
GGAGGTGTACCCAACCAAGGAGCTGGAAAAGCTGCTGAAAGACTACTCTATCGAGTATGGACATGGCG
AATGCATCAAGGCAGCCATCTGTGGCGAGAGTGATAAGAAATTTTTCGCCAAGCTGACCTCAGTGCTGA
ATACAATCCTGCAGATGCGGAACTCAAAGACCGGGACAGAACTGGACTATCTGATTAGCCCCGTGGCTG
ATGTCAACGGAAACTTCTTCGACAGCAGACAGGCACCCAAAAATATGCCTCAGGATGCAGACGCCAAC
GGGGCCTACCACATCGGGCTGAAGGGACTGATGCTGCTGGGCCGGATCAAGAACAATCAGGAGGGGA
AGAAGCTGAACCTGGTCATTAAGAACGAGGAATACTTCGAGTTTGTCCAGAATAGAAATAACaaaaggccgg
cggccacgaaaaaggccggccaggcaaaaaagaaaaagGGATCCTACCCATACGATGTTCCAGATTACGCTTATCCCTAC
GACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCCTAA (SEQ ID NO:17)

[0286] FnCpf1-NLS-3xHA的氨基酸序列

[0287] FnCpf1为正常文本(AAs 1-1300),NLS(krpaatkkagqakkkk,SEQ ID NO:7)为小写字母,3xHA标签(YPYDVPDYAYPYDVPDYAYPYDVPDYA,SEQ ID NO:8)为粗体

[0288] MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSS
VCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESDLI
LWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFL
ENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTI
IGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSF
YEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAP
KNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNL
AQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYF
ELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKI
FDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFE
FNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKL
YLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKN
KDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIDRGERHLAY

[0289] YTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAK
LVIEYNAIVVFEDLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETF
KKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGD
KAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFA
KLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDADANGAYHIGLKGLMLLGRIK
NNQEGKKLNLVIKNEEYFEFVQNRNNkrpaatkkagqakkkkGS**YPYDVPDYAYPYDVPDYAYPYDVPDYA**
(SEQ ID NO:18)

[0290] AAS2134:pCAG-hMbCas12a-NLS-3xHA

[0291] 人类密码子优化的牛眼莫拉氏菌(Moraxella bovoculi)237Cas12a(MbCas12a)为黑色,核质蛋白NLS(aaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaag,SEQ ID NO:21)为小写字母,接头序列为斜体,3xHA标签(TACCCATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCC,SEQ ID NO:5)为粗体

[0292] ATGCTGTTCCAGGACTTTACCCACCTGTATCCACTGTCCAAGACAGTGAGATTTGAGCTGAAGC
CCATCGATAGGACCCTGGAGCACATCCACGCCAAGAACTTCCTGTCTCAGGACGAGACAATGGCCGA
TATGCACCAGAAGGTGAAAGTGATCCTGGACGATTACCACCGCGACTTCATCGCCGATATGATGGGCG
AGGTGAAGCTGACCAAGCTGGCCGAGTTCTATGACGTGTACCTGAAGTTTCGGAAGAACCCAAAGGA
CGATGAGCTGCAGAAGCAGCTGAAGGATCTGCAGGCCGTGCTGAGAAAGGAGATCGTGAAGCCCAT
CGGCAATGGCGGCAAGTATAAGGCCGGCTACGACAGGCTGTTCGGCGCCAAGCTGTTTAAGGACGGC
AAGGAGCTGGGCGATCTGGCCAAGTTCGTGATCGCACAGGAGGGAGAGAGCTCCCCAAAGCTGGCC
CACCTGGCCCACTTCGAGAAGTTTTCCACCTATTTCACAGGCTTTCACGATAACCGGAAGAATATGTAT
TCTGACGAGGATAAGCACACCGCCATCGCCTACCGCCTGATCCACGAGAACCTGCCCCGGTTTATCGA
CAATCTGCAGATCCTGACCACAATCAAGCAGAAGCACTCTGCCCTGTACGATCAGATCATCAACGAGC
TGACCGCCAGCGGCCTGGACGTGTCTCTGGCCAGCCACCTGGATGGCTATCACAAGCTGCTGACACA
GGAGGGCATCACCGCCTACAATACACTGCTGGGAGGAATCTCCGGAGAGGCAGGCTCTCCTAAGATC
CAGGGCATCAACGAGCTGATCAATTCTCACCACAACCAGCACTGCCACAAGAGCGAGAGAATCGCCA
AGCTGAGGCCACTGCACAAGCAGATCCTGTCCGACGGCATGAGCGTGTCCTTCCTGCCCTCTAAGTTT
GCCGACGATAGCGAGATGTGCCAGGCCGTGAACGAGTTCTATCGCCACTACGCCGACGTGTTCGCCA
AGGTGCAGAGCCTGTTCGACGGCTTTGACGATCACCAGAAGGATGGCATCTACGTGGAGCACAAGAA
CCTGAATGAGCTGTCCAAGCAGGCCTTCGGCGACTTTGCACTGCTGGGACGCGTGCTGGACGGATAC
TATGTGGATGTGGTGAATCCAGAGTTCAACGAGCGGTTTGCCAAGGCCAAGACCGACAATGCCAAGG
CCAAGCTGACAAAGGAGAAGGATAAGTTCATCAAGGGCGTGCACTCCCTGGCCTCTCTGGAGCAGGC
CATCGAGCACTATACCGCAAGGCACGACGATGAGAGCGTGCAGGCAGGCAAGCTGGGACAGTACTTC
AAGCACGGCCTGGCCGGAGTGGACAACCCCATCCAGAAGATCCACAACAATCACAGCACCATCAAG
GGCTTTCTGGAGAGGGAGCGCCCTGCAGGAGAGAGAGCCCTGCCAAAGATCAAGTCCGGCAAGAAT
CCTGAGATGACACAGCTGAGGCAGCTGAAGGAGCTGCTGGATAACGCCCTGAATGTGGCCCACTTCG
CCAAGCTGCTGACCACAAAGACCACACTGGACAATCAGGATGGCAACTTCTATGGCGAGTTTGGCGT
GCTGTACGACGAGCTGGCCAAGATCCCCACCCTGTATAACAAGGTGAGAGATTACCTGAGCCAGAAG
CCTTTCTCCACCGAGAAGTACAAGCTGAACTTTGGCAATCCAACACTGCTGAATGGCTGGGACCTGA
ACAAGGAGAAGGATAATTTCGGCGTGATCCTGCAGAAGGACGGCTGCTACTATCTGGCCCTGCTGGA
CAAGGCCCACAAGAAGGTGTTTGATAACGCCCCTAATACAGGCAAGAGCATCTATCAGAAGATGATCT
ATAAGTACCTGGAGGTGAGGAAGCAGTTCCCCAAGGTGTTCTTTTCCAAGGAGGCCATCGCCATCAA
CTACCACCCTTCTAAGGAGCTGGTGGAGATCAAGGACAAGGGCCGGCAGAGATCCGACGATGAGCGC
CTGAAGCTGTATCGGTTTATCCTGGAGTGTCTGAAGATCCACCCTAAGTACGATAAGAAGTTCGAGGG

CGCCATCGGCGACATCCAGCTGTTTAAGAAGGATAAGAAGGGCAGAGAGGTGCCAATCAGCGAGAA
GGACCTGTTCGATAAGATCAACGGCATCTTTTCTAGCAAGCCTAAGCTGGAGATGGAGGACTTCTTTA
TCGGCGAGTTCAAGAGGTATAACCCAAGCCAGGACCTGGTGGATCAGTATAATATCTACAAGAAGATC
GACTCCAACGATAATCGCAAGAAGGAGAATTTCTACAACAATCACCCCAAGTTTAAGAAGGATCTGG
TGCGGTACTATTACGAGTCTATGTGCAAGCACGAGGAGTGGGAGGAGAGCTTCGAGTTTTCCAAGAA
GCTGCAGGACATCGGCTGTTACGTGGATGTGAACGAGCTGTTTACCGAGATCGAGACACGGAGACTG
AATTATAAGATCTCCTTCTGCAACATCAATGCCGACTACATCGATGAGCTGGTGGAGCAGGGCCAGCT
GTATCTGTTCCAGATCTACAACAAGGACTTTTCCCCAAAGGCCCACGGCAAGCCCAATCTGCACACCC
TGTACTTCAAGGCCCTGTTTTCTGAGGACAACCTGGCCGATCCTATCTATAAGCTGAATGGCGAGGCC
CAGATCTTCTACAGAAAGGCCTCCCTGGACATGAACGAGACAACAATCCACAGGGCCGGCGAGGTGC
TGGAGAACAAGAATCCCGATAATCCTAAGAAGAGACAGTTCGTGTACGACATCATCAAGGATAAGAG
GTACACACAGGACAAGTTCATGCTGCACGTGCCAATCACCATGAACTTTGGCGTGCAGGGCATGACA
ATCAAGGAGTTCAATAAGAAGGTGAACCAGTCTATCCAGCAGTATGACGAGGTGAACGTGATCGGCA
TCGATCGGGGCGAGAGACACCTGCTGTACCTGACCGTGATCAATAGCAAGGGCGAGATCCTGGAGCA
GTGTTCCCTGAACGACATCACCACAGCCTCTGCCAATGGCACACAGATGACCACACCTTACCACAAG
ATCCTGGATAAGAGGGAGATCGAGCGCCTGAACGCCCGGGTGGGATGGGGCGAGATCGAGACAATCA

[0293] AGGAGCTGAAGTCTGGCTATCTGAGCCACGTGGTGCACCAGATCAGCCAGCTGATGCTGAAGTACAA
CGCCATCGTGGTGCTGGAGGACCTGAATTTCGGCTTTAAGAGGGGCCGCTTTAAGGTGGAGAAGCAG
ATCTATCAGAACTTCGAGAATGCCCTGATCAAGAAGCTGAACCACCTGGTGCTGAAGGACAAGGCCG
ACGATGAGATCGGCTCTTACAAGAATGCCCTGCAGCTGACCAACAATTTCACAGATCTGAAGAGCATC
GGCAAGCAGACCGGCTTCCTGTTTTATGTGCCCGCCTGGAACACCTCTAAGATCGACCCTGAGACAG
GCTTTGTGGATCTGCTGAAGCCAAGATACGAGAACATCGCCCAGAGCCAGGCCTTCTTTGGCAAGTT
CGACAAGATCTGCTATAATGCCGACAAGGATTACTTCGAGTTTCACATCGACTACGCCAAGTTTACCGA
TAAGGCCAAGAATAGCCGCCAGATCTGGACAATCTGTTCCCACGGCGACAAGCGGTACGTGTACGAT
AAGACAGCCAACCAGAATAAGGGCGCCGCCAAGGGCATCAACGTGAATGATGAGCTGAAGTCCCTG
TTCGCCCGCCACCACATCAACGAGAAGCAGCCCAACCTGGTCATGGACATCTGCCAGAACAATGATA
AGGAGTTTCACAAGTCTCTGATGTACCTGCTGAAAACCCTGCTGGCCCTGCGGTACAGCAACGCCTC
CTCTGACGAGGATTTCATCCTGTCCCCCGTGGCAAACGACGAGGGCGTGTTCTTTAATAGCGCCCTGG
CCGACGATACACAGCCTCAGAATGCCGATGCCAACGGCGCCTACCACATCGCCCTGAAGGGCCTGTG
GCTGCTGAATGAGCTGAAGAACTCCGACGATCTGAACAAGGTGAAGCTGGCCATCGACAATCAGACC
TGGCTGAATTTCGCCCAGAACAGGaaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaag*GGATCC***TAC**
CCATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATG
TCCCCGACTATGCCTAA (SEQ ID NO: 430)

[0294] (RTW876)pCAG-人类-dAsCpf1(D908A)-NLS(核质蛋白)-3xHA-VPR的核苷酸序列

[0295] 人类密码子优化的dAsCpf1(D908A)为正常字体(NTs 1-3921),核质蛋白NLS为小写字母(aaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaag,SEQ ID NO:21),3xHA标签(TACCCATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCG ACTATGCC,SEQ ID NO:5)为粗体,VPR为双下划线的

ATGACACAGTTCGAGGGCTTTACCAACCTGTATCAGGTGAGCAAGACACTGCGGTTTGAGCTG
ATCCCACAGGGCAAGACCCTGAAGCACATCCAGGAGCAGGGCTTCATCGAGGAGGACAAGGCCCGC
AATGATCACTACAAGGAGCTGAAGCCCATCATCGATCGGATCTACAAGACCTATGCCGACCAGTGCCT
GCAGCTGGTGCAGCTGGATTGGGAGAACCTGAGCGCCGCCATCGACTCCTATAGAAAGGAGAAAACC
GAGGAGACAAGGAACGCCCTGATCGAGGAGCAGGCCACATATCGCAATGCCATCCACGACTACTTCA

TCGGCCGGACAGACAACCTGACCGATGCCATCAATAAGAGACACGCCGAGATCTACAAGGGCCTGTT
CAAGGCCGAGCTGTTTAATGGCAAGGTGCTGAAGCAGCTGGGCACCGTGACCACAACCGAGCACGA
GAACGCCCTGCTGCGGAGCTTCGACAAGTTTACAACCTACTTCTCCGGCTTTTATGAGAACAGGAAG
AACGTGTTCAGCGCCGAGGATATCAGCACAGCCATCCCACACCGCATCGTGCAGGACAACTTCCCCA
AGTTTAAGGAGAATTGTCACATCTTCACACGCCTGATCACCGCCGTGCCCAGCCTGCGGGAGCACTTT
GAGAACGTGAAGAAGGCCATCGGCATCTTCGTGAGCACCTCCATCGAGGAGGTGTTTTCCTTCCCTTT
TTATAACCAGCTGCTGACACAGACCCAGATCGACCTGTATAACCAGCTGCTGGGAGGAATCTCTCGGG
AGGCAGGCACCGAGAAGATCAAGGGCCTGAACGAGGTGCTGAATCTGGCCATCCAGAAGAATGATG
AGACAGCCCACATCATCGCCTCCCTGCCACACAGATTCATCCCCCTGTTTAAGCAGATCCTGTCCGATA
GGAACACCCTGTCTTTCATCCTGGAGGAGTTTAAGAGCGACGAGGAAGTGATCCAGTCCTTCTGCAA
GTACAAGACACTGCTGAGAAACGAGAACGTGCTGGAGACAGCCGAGGCCCTGTTTAACGAGCTGAA
CAGCATCGACCTGACACACATCTTCATCAGCCACAAGAAGCTGGAGACAATCAGCAGCGCCCTGTGC
GACCACTGGGATACACTGAGGAATGCCCTGTATGAGCGGAGAATCTCCGAGCTGACAGGCAAGATCA
CCAAGTCTGCCAAGGAGAAGGTGCAGCGCAGCCTGAAGCACGAGGATATCAACCTGCAGGAGATCA
TCTCTGCCGCAGGCAAGGAGCTGAGCGAGGCCTTCAAGCAGAAAACCAGCGAGATCCTGTCCCACG
CACACGCCGCCCTGGATCAGCCACTGCCTACAACCCTGAAGAAGCAGGAGGAGAAGGAGATCCTGA
AGTCTCAGCTGGACAGCCTGCTGGGCCTGTACCACCTGCTGGACTGGTTTGCCGTGGATGAGTCCAA
CGAGGTGGACCCCGAGTTCTCTGCCCGGCTGACCGGCATCAAGCTGGAGATGGAGCCTTCTCTGAGC
TTCTACAACAAGGCCAGAAATTATGCCACCAAGAAGCCCTACTCCGTGGAGAAGTTCAAGCTGAACT
TTCAGATGCCTACACTGGCCTCTGGCTGGGACGTGAATAAGGAGAAGAACAATGGCGCCATCCTGTTT
GTGAAGAACGGCCTGTACTATCTGGGCATCATGCCAAAGCAGAAGGGCAGGTATAAGGCCCTGAGCT
TCGAGCCCACAGAGAAAACCAGCGAGGGCTTTGATAAGATGTACTATGACTACTTCCCTGATGCCGCC
[0296] AAGATGATCCCAAAGTGCAGCACCCAGCTGAAGGCCGTGACAGCCCACTTTCAGACCCACACAACCC
CCATCCTGCTGTCCAACAATTTCATCGAGCCTCTGGAGATCACAAAGGAGATCTACGACCTGAACAAT
CCTGAGAAGGAGCCAAAGAAGTTTCAGACAGCCTACGCCAAGAAAACCGGCGACCAGAAGGGCTAC
AGAGAGGCCCTGTGCAAGTGGATCGACTTCACAAGGGATTTTCTGTCCAAGTATACCAAGACAACCT
CTATCGATCTGTCTAGCCTGCGGCCATCCTCTCAGTATAAGGACCTGGGCGAGTACTATGCCGAGCTGA
ATCCCCTGCTGTACCACATCAGCTTCCAGAGAATCGCCGAGAAGGAGATCATGGATGCCGTGGAGAC
AGGCAAGCTGTACCTGTTCCAGATCTATAACAAGGACTTTGCCAAGGGCCACCACGGCAAGCCTAATC
TGCACACACTGTATTGGACCGGCCTGTTTTCTCCAGAGAACCTGGCCAAGACAAGCATCAAGCTGAA
TGGCCAGGCCGAGCTGTTCTACCGCCCTAAGTCCAGGATGAAGAGGATGGCACACCGGCTGGGAGAG
AAGATGCTGAACAAGAAGCTGAAGGATCAGAAAACCCCAATCCCCGACACCCTGTACCAGGAGCTGT
ACGACTATGTGAATCACAGACTGTCCCACGACCTGTCTGATGAGGCCAGGGCCCTGCTGCCCAACGT
GATCACCAAGGAGGTGTCTCACGAGATCATCAAGGATAGGCGCTTTACCAGCGACAAGTTCTTTTTCC
ACGTGCCTATCACACTGAACTATCAGGCCGCCAATTCCCCATCTAAGTTCAACCAGAGGGTGAATGCC
TACCTGAAGGAGCACCCCGAGACACCTATCATCGGCATCGCCCGGGGCGAGAGAAACCTGATCTATAT
CACAGTGATCGACTCCACCGGCAAGATCCTGGAGCAGCGGAGCCTGAACACCATCCAGCAGTTTGAT
TACCAGAAGAAGCTGGACAACAGGGAGAAGGAGAGGGTGGCAGCAAGGCAGGCCTGGTCTGTGGT
GGGCACAATCAAGGATCTGAAGCAGGGCTATCTGAGCCAGGTCATCCACGAGATCGTGGACCTGATG
ATCCACTACCAGGCCGTGGTGGTGCTGGAGAACCTGAATTTCGGCTTTAAGAGCAAGAGGACCGGCA
TCGCCGAGAAGGCCGTGTACCAGCAGTTCGAGAAGATGCTGATCGATAAGCTGAATTGCCTGGTGCT
GAAGGACTATCCAGCAGAGAAAGTGGGAGGCGTGCTGAACCCATACCAGCTGACAGACCAGTTCAC
CTCCTTTGCCAAGATGGGCACCCAGTCTGGCTTCCTGTTTTACGTGCCTGCCCCATATACATCTAAGAT
CGATCCCCTGACCGGCTTCGTGGACCCCTTCGTGTGGAAAACCATCAAGAATCACGAGAGCCGCAAG
CACTTCCTGGAGGGCTTCGACTTTCTGCACTACGACGTGAAAACCGGCGACTTCATCCTGCACTTTAA

[0297] GATGAACAGAAATCTGTCCTTCCAGAGGGGCCTGCCCGGCTTTATGCCTGCATGGGATATCGTGTTCG
AGAAGAACGAGACACAGTTTGACGCCAAGGGCACCCCTTTCATCGCCGGCAAGAGAATCGTGCCAG
TGATCGAGAATCACAGATTCACCGGCAGATACCGGGACCTGTATCCTGCCAACGAGCTGATCGCCCTG
CTGGAGGAGAAGGGCATCGTGTTCAGGGATGGCTCCAACATCCTGCCAAAGCTGCTGGAGAATGACG
ATTCTCACGCCATCGACACCATGGTGGCCCTGATCCGCAGCGTGCTGCAGATGCGGAACTCCAATGCC
GCCACAGGCGAGGACTATATCAACAGCCCCGTGCGCGATCTGAATGGCGTGTGCTTCGACTCCCGGTT
TCAGAACCCAGAGTGGCCCATGGACGCCGATGCCAATGGCGCCTACCACATCGCCCTGAAGGGCCAG
CTGCTGCTGAATCACCTGAAGGAGAGCAAGGATCTGAAGCTGCAGAACGGCATCTCCAATCAGGACT
GGCTGGCCTACATCCAGGAGCTGCGCAACaaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaagGGATCC
TACCCATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGAT
GTCCCCGACTATGCCGGAAGCGAGGCCAGCGGTTCCGGACGGGCTGACGCATTGGACGATTTTGAT
CTGGATATGCTGGGAAGTGACGCCCTCGATGATTTTGACCTTGACATGCTTGGTTCGGATGCCCTTGAT
GACTTTGACCTCGACATGCTCGGCAGTGACGCCCTTGATGATTTCGACCTGGACATGCTGATTAACTCT
AGAAGTTCCGGATCTCCGAAAAAGAAACGCAAAGTTGGTAGCCAGTACCTGCCCGACACCGACGAC
CGGCACCGGATCGAGGAAAAGCGGAAGCGGACCTACGAGACATTCAAGAGCATCATGAAGAAGTCC
CCCTTCAGCGGCCCCACCGACCCTAGACCTCCACCTAGAAGAATCGCCGTGCCCAGCAGATCCAGCG
CCAGCGTGCCAAAACCTGCCCCCCAGCCTTACCCCTTCACCAGCAGCCTGAGCACCATCAACTACGA
CGAGTTCCCTACCATGGTGTTCCCCAGCGGCCAGATCTCTCAGGCCTCTGCTCTGGCTCCAGCCCCTC
CTCAGGTGCTGCCTCAGGCTCCTGCTCCTGCACCAGCTCCAGCCATGGTGTCTGCACTGGCTCAGGCA
CCAGCACCCGTGCCTGTGCTGGCTCCTGGACCTCCACAGGCTGTGGCTCCACCAGCCCCTAAACCTA
CACAGGCCGGCGAGGGCACACTGTCTGAAGCTCTGCTGCAGCTGCAGTTCGACGACGAGGATCTGG
GAGCCCTGCTGGGAAACAGCACCGATCCTGCCGTGTTCACCGACCTGGCCAGCGTGGACAACAGCG
AGTTCCAGCAGCTGCTGAACCAGGGCATCCCTGTGGCCCCTCACACCACCGAGCCCATGCTGATGGA
ATACCCCGAGGCCATCACCCGGCTCGTGACAGGCGCTCAGAGGCCTCCTGATCCAGCTCCTGCCCCTC
TGGGAGCACCAGGCCTGCCTAATGGACTGCTGTCTGGCGACGAGGACTTCAGCTCTATCGCCGATATG
GATTTCTCAGCCTTGCTGGGCTCTGGCAGCGGCAGCCGGGATTCCAGGGAAGGGATGTTTTTGCCGA
AGCCTGAGGCCGGCTCCGCTATTAGTGACGTGTTTGAGGGCCGCGAGGTGTGCCAGCCAAAACGAAT
CCGGCCATTTCATCCTCCAGGAAGTCCATGGGCCAACCGCCCACTCCCCGCCAGCCTCGCACCAACA
CCAACCGGTCCAGTACATGAGCCAGTCGGGTCACTGACCCCGGCACCAGTCCCTCAGCCACTGGATC
CAGCGCCCGCAGTGACTCCCGAGGCCAGTCACCTGTTGGAGGATCCCGATGAAGAGACGAGCCAGG
CTGTCAAAGCCCTTCGGGAGATGGCCGATACTGTGATTCCCCAGAAGGAAGAGGCTGCAATCTGTGG
CCAAATGGACCTTTCCCATCCGCCCCCAAGGGGCCATCTGGATGAGCTGACAACCACACTTGAGTCCA
TGACCGAGGATCTGAACCTGGACTCACCCCTGACCCCGGAATTGAACGAGATTCTGGATACCTTCCTG
AACGACGAGTGCCTCTTGCATGCCATGCATATCAGCACAGGACTGTCCATCTTCGACACATCTCTGTTT
TAA SEQ ID NO: 431)

[0298] dAsCpf1(D908A)-NLS(核质蛋白)-3xHA-VPR的氨基酸序列

[0299] AsCpf1为正常字体(AAs 1-1306),NLS(核质蛋白)(krpaatkkagqakkkk,SEQ ID NO:7)为小写字母,3xHA标签(YPYDVPDYAYPYDVPDYAYPYDVPDYA,SEQ ID NO:8)为粗体,VPR为双下划线的

[0300] MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTYADQCLQ
LVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKAELF
NGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTR
LITAVPSLREHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLA
IQKNDETAHIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSI

[0301] DLTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEIISAAGKELS EAFKQKTSEILSHAHAALDQPLPTTLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPEFSARLTGI KLEMEPSLSFYNKARNYATKKPYSVEKFKLNFQMPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQ KGRYKALSFEPTEKTSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIY DLNNPEKEPKKFQTAYAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAE LNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQ AELFYRPKSRMKRMAHRLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKE VSHEIIKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIARGERNLIYITVIDSTGKIL EQRSLNTIQQFDYQKKLDNREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLN FGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFLFYVP APYTSKIDPLTGFVDPFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAW DIVFEKNETQFDAKGTPFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDS HAIDTMVALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLL NHLKESKDLKLQNGISNQDWLAYIQELRNkrpaatkkagqakkkkGS**YPYDVPDYAYPYDVPDYAYPYDVPDY A**GSEASGSGRADALDDFDLDMLGSDALDDFDLDMLGSDALDDFDLDMLGSDALDDFDLDMLINSRSSG SPKKKRKVGSQYLPDTDDRHRIEEKRKRTYETFKSIMKKSPFSGPTDPRPPPRRIAVPSRSSASVPKPAPQP YPFTSSLSTINYDEFPTMVFPSGQISQASALAPAPPQVLPQAPAPAPAPAMVSALAQAPAPVPVLAPGPPQA VAPPAPKPTQAGEGTLSEALLQLQFDDEDLGALLGNSTDPAVFTDLASVDNSEFQQLLNQGIPVAPHTTEP MLMEYPEAITRLVTGAQRPPDPAPAPLGAPGLPNGLLSGDEDFSSIADMDFSALLGSGSGSRDSREGMFLP KPEAGSAISDVFEGREVCQPKRIRPFHPPGSPWANRPLPASLAPTPTGPVHEPVGSLTPAPVPQPLDPAPAVT PEASHLLEDPDEETSQAVKALREMADTVIPQKEEAAICGQMDLSHPPPRGHLDELTTTLESMTEDLNLDSP LTPELNEILDTFLNDECLLHAMHISTGLSIFDTSLF SEQ ID NO: 432)

[0302] (RTW776)pCAG-人类-dAsCpf1(D908A)三重变体(E174R/S542R/K548R)-NLS(核质蛋白)-3xHA-VPR的核苷酸序列

[0303] 人类密码子优化的dAsCpf1(D908A)为正常字体(NTs 1-3921),核质蛋白NLS为小写字母(aaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaag,SEQ ID NO:21),3xHA标签(TACCCATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCG ACTATGCC,SEQ ID NO:5)为粗体,VPR为双下划线的

[0304] ATGACACAGTTCGAGGGCTTTACCAACCTGTATCAGGTGAGCAAGACACTGCGGTTTGAGCTG ATCCCACAGGGCAAGACCCTGAAGCACATCCAGGAGCAGGGCTTCATCGAGGAGGACAAGGCCCGC AATGATCACTACAAGGAGCTGAAGCCCATCATCGATCGGATCTACAAGACCTATGCCGACCAGTGCCT GCAGCTGGTGCAGCTGGATTGGGAGAACCTGAGCGCCGCCATCGACTCCTATAGAAAGGAGAAAACC GAGGAGACAAGGAACGCCCTGATCGAGGAGCAGGCCACATATCGCAATGCCATCCACGACTACTTCA TCGGCCGGACAGACAACCTGACCGATGCCATCAATAAGAGACACGCCGAGATCTACAAGGGCCTGTT CAAGGCCGAGCTGTTTAATGGCAAGGTGCTGAAGCAGCTGGGCACCGTGACCACAACCGAGCACGA GAACGCCCTGCTGCGGAGCTTCGACAAGTTTACAACCTACTTCTCCGGCTTTTATAGAAACAGGAAGA ACGTGTTCAGCGCCGAGGATATCAGCACAGCCATCCCACACCGCATCGTGCAGGACAACTTCCCCAA GTTTAAGGAGAATTGTCACATCTTCACACGCCTGATCACCGCCGTGCCCAGCCTGCGGGAGCACTTTG AGAACGTGAAGAAGGCCATCGGCATCTTCGTGAGCACCTCCATCGAGGAGGTGTTTTCCTTCCCTTTT TATAACCAGCTGCTGACACAGACCCAGATCGACCTGTATAACCAGCTGCTGGGAGGAATCTCTCGGGA GGCAGGCACCGAGAAGATCAAGGGCCTGAACGAGGTGCTGAATCTGGCCATCCAGAAGAATGATGA GACAGCCCACATCATCGCCTCCCTGCCACACAGATTCATCCCCCTGTTTAAGCAGATCCTGTCCGATAG GAACACCCTGTCTTTCATCCTGGAGGAGTTTAAGAGCGACGAGGAAGTGATCCAGTCCTTCTGCAAG

TACAAGACACTGCTGAGAAACGAGAACGTGCTGGAGACAGCCGAGGCCCTGTTTAACGAGCTGAAC
AGCATCGACCTGACACACATCTTCATCAGCCACAAGAAGCTGGAGACAATCAGCAGCGCCCTGTGCG
ACCACTGGGATACACTGAGGAATGCCCTGTATGAGCGGAGAATCTCCGAGCTGACAGGCAAGATCAC
CAAGTCTGCCAAGGAGAAGGTGCAGCGCAGCCTGAAGCACGAGGATATCAACCTGCAGGAGATCAT
CTCTGCCGCAGGCAAGGAGCTGAGCGAGGCCTTCAAGCAGAAAACCAGCGAGATCCTGTCCCACGC
ACACGCCGCCCTGGATCAGCCACTGCCTACAACCCTGAAGAAGCAGGAGGAGAAGGAGATCCTGAA
GTCTCAGCTGGACAGCCTGCTGGGCCTGTACCACCTGCTGGACTGGTTTGCCGTGGATGAGTCCAAC
GAGGTGGACCCCGAGTTCTCTGCCCGGCTGACCGGCATCAAGCTGGAGATGGAGCCTTCTCTGAGCT
TCTACAACAAGGCCAGAAATTATGCCACCAAGAAGCCCTACTCCGTGGAGAAGTTCAAGCTGAACTT
TCAGATGCCTACACTGGCCAGAGGCTGGGACGTGAATAGAGAGAAGAACAATGGCGCCATCCTGTTT
GTGAAGAACGGCCTGTACTATCTGGGCATCATGCCAAAGCAGAAGGGCAGGTATAAGGCCCTGAGCT
TCGAGCCCACAGAGAAAACCAGCGAGGGCTTTGATAAGATGTACTATGACTACTTCCCTGATGCCGCC
AAGATGATCCCAAAGTGCAGCACCCAGCTGAAGGCCGTGACAGCCCACTTTCAGACCCACACAACCC
CCATCCTGCTGTCCAACAATTTCATCGAGCCTCTGGAGATCACAAAGGAGATCTACGACCTGAACAAT
CCTGAGAAGGAGCCAAAGAAGTTTCAGACAGCCTACGCCAAGAAAACCGGCGACCAGAAGGGCTAC
AGAGAGGCCCTGTGCAAGTGGATCGACTTCACAAGGGATTTTCTGTCCAAGTATACCAAGACAACCT
CTATCGATCTGTCTAGCCTGCGGCCATCCTCTCAGTATAAGGACCTGGGCGAGTACTATGCCGAGCTGA
ATCCCCTGCTGTACCACATCAGCTTCCAGAGAATCGCCGAGAAGGAGATCATGGATGCCGTGGAGAC
AGGCAAGCTGTACCTGTTCCAGATCTATAACAAGGACTTTGCCAAGGGCCACCACGGCAAGCCTAATC
TGCACACACTGTATTGGACCGGCCTGTTTTCTCCAGAGAACCTGGCCAAGACAAGCATCAAGCTGAA
TGGCCAGGCCGAGCTGTTCTACCGCCCTAAGTCCAGGATGAAGAGGATGGCACACCGGCTGGGAGAG
AAGATGCTGAACAAGAAGCTGAAGGATCAGAAAACCCCAATCCCCGACACCCTGTACCAGGAGCTGT
[0305] ACGACTATGTGAATCACAGACTGTCCCACGACCTGTCTGATGAGGCCAGGGCCCTGCTGCCCAACGT
GATCACCAAGGAGGTGTCTCACGAGATCATCAAGGATAGGCGCTTTACCAGCGACAAGTTCTTTTTCC
ACGTGCCTATCACACTGAACTATCAGGCCGCCAATTCCCCATCTAAGTTCAACCAGAGGGTGAATGCC
TACCTGAAGGAGCACCCCGAGACACCTATCATCGGCATCGCCCGGGGCGAGAGAAACCTGATCTATAT
CACAGTGATCGACTCCACCGGCAAGATCCTGGAGCAGCGGAGCCTGAACACCATCCAGCAGTTTGAT
TACCAGAAGAAGCTGGACAACAGGGAGAAGGAGAGGGTGGCAGCAAGGCAGGCCTGGTCTGTGGT
GGGCACAATCAAGGATCTGAAGCAGGGCTATCTGAGCCAGGTCATCCACGAGATCGTGGACCTGATG
ATCCACTACCAGGCCGTGGTGGTGCTGGAGAACCTGAATTTCGGCTTTAAGAGCAAGAGGACCGGCA
TCGCCGAGAAGGCCGTGTACCAGCAGTTCGAGAAGATGCTGATCGATAAGCTGAATTGCCTGGTGCT
GAAGGACTATCCAGCAGAGAAAGTGGGAGGCGTGCTGAACCCATACCAGCTGACAGACCAGTTCAC
CTCCTTTGCCAAGATGGGCACCCAGTCTGGCTTCCTGTTTTACGTGCCTGCCCCATATACATCTAAGAT
CGATCCCCTGACCGGCTTCGTGGACCCCTTCGTGTGGAAAACCATCAAGAATCACGAGAGCCGCAAG
CACTTCCTGGAGGGCTTCGACTTTCTGCACTACGACGTGAAAACCGGCGACTTCATCCTGCACTTTAA
GATGAACAGAAATCTGTCCTTCCAGAGGGGCCTGCCCGGCTTTATGCCTGCATGGGATATCGTGTTCG
AGAAGAACGAGACACAGTTTGACGCCAAGGGCACCCCTTTCATCGCCGGCAAGAGAATCGTGCCAG
TGATCGAGAATCACAGATTCACCGGCAGATACCGGGACCTGTATCCTGCCAACGAGCTGATCGCCCTG
CTGGAGGAGAAGGGCATCGTGTTCAGGGATGGCTCCAACATCCTGCCAAAGCTGCTGGAGAATGACG
ATTCTCACGCCATCGACACCATGGTGGCCCTGATCCGCAGCGTGCTGCAGATGCGGAACTCCAATGCC
GCCACAGGCGAGGACTATATCAACAGCCCCGTGCGCGATCTGAATGGCGTGTGCTTCGACTCCCGGTT
TCAGAACCCAGAGTGGCCCATGGACGCCGATGCCAATGGCGCCTACCACATCGCCCTGAAGGGCCAG
CTGCTGCTGAATCACCTGAAGGAGAGCAAGGATCTGAAGCTGCAGAACGGCATCTCCAATCAGGACT
GGCTGGCCTACATCCAGGAGCTGCGCAACaaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaagGGATCC
TACCCATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGAT

[0306] **GTCCCCGACTATGCC**GGAAGCGAGGCCAGCGGTTCCGGACGGGCTGACGCATTGGACGATTTTGAT
CTGGATATGCTGGGAAGTGACGCCCTCGATGATTTTGACCTTGACATGCTTGGTTCGGATGCCCTTGAT
GACTTTGACCTCGACATGCTCGGCAGTGACGCCCTTGATGATTTCGACCTGGACATGCTGATTAACTCT
AGAAGTTCCGGATCTCCGAAAAAGAAACGCAAAGTTGGTAGCCAGTACCTGCCCGACACCGACGAC
CGGCACCGGATCGAGGAAAAGCGGAAGCGGACCTACGAGACATTCAAGAGCATCATGAAGAAGTCC
CCCTTCAGCGGCCCCACCGACCCTAGACCTCCACCTAGAAGAATCGCCGTGCCCAGCAGATCCAGCG
CCAGCGTGCCAAAACCTGCCCCCCAGCCTTACCCCTTCACCAGCAGCCTGAGCACCATCAACTACGA
CGAGTTCCCTACCATGGTGTTCCCCAGCGGCCAGATCTCTCAGGCCTCTGCTCTGGCTCCAGCCCCTC
CTCAGGTGCTGCCTCAGGCTCCTGCTCCTGCACCAGCTCCAGCCATGGTGTCTGCACTGGCTCAGGCA
CCAGCACCCGTGCCTGTGCTGGCTCCTGGACCTCCACAGGCTGTGGCTCCACCAGCCCCTAAACCTA
CACAGGCCGGCGAGGGCACACTGTCTGAAGCTCTGCTGCAGCTGCAGTTCGACGACGAGGATCTGG
GAGCCCTGCTGGGAAACAGCACCGATCCTGCCGTGTTCACCGACCTGGCCAGCGTGGACAACAGCG
AGTTCCAGCAGCTGCTGAACCAGGGCATCCCTGTGGCCCCTCACACCACCGAGCCCATGCTGATGGA
ATACCCCGAGGCCATCACCCGGCTCGTGACAGGCGCTCAGAGGCCTCCTGATCCAGCTCCTGCCCCTC
TGGGAGCACCAGGCCTGCCTAATGGACTGCTGTCTGGCGACGAGGACTTCAGCTCTATCGCCGATATG
GATTTCTCAGCCTTGCTGGGCTCTGGCAGCGGCAGCCGGGATTCCAGGGAAGGGATGTTTTTGCCGA
AGCCTGAGGCCGGCTCCGCTATTAGTGACGTGTTTGAGGGCCGCGAGGTGTGCCAGCCAAAACGAAT
CCGGCCATTTCATCCTCCAGGAAGTCCATGGGCCAACCGCCCACTCCCCGCCAGCCTCGCACCAACA
CCAACCGGTCCAGTACATGAGCCAGTCGGGTCACTGACCCCGGCACCAGTCCCTCAGCCACTGGATC
CAGCGCCCGCAGTGACTCCCGAGGCCAGTCACCTGTTGGAGGATCCCGATGAAGAGACGAGCCAGG
CTGTCAAAGCCCTTCGGGAGATGGCCGATACTGTGATTCCCCAGAAGGAAGAGGCTGCAATCTGTGG
CCAAATGGACCTTTCCCATCCGCCCCCAAGGGGCCATCTGGATGAGCTGACAACCACACTTGAGTCCA
TGACCGAGGATCTGAACCTGGACTCACCCCTGACCCCGGAATTGAACGAGATTCTGGATACCTTCCTG
AACGACGAGTGCCTCTTGCATGCCATGCATATCAGCACAGGACTGTCCATCTTCGACACATCTCTGTTT
SEQ ID NO: 433)

[0307] dAsCpf1(D908A)三重变体(E174R/S542R/K548R)-NLS(核质蛋白)-3xHA-VPR的氨基酸序列

[0308] AsCpf1为正常字体(AAs 1-1307),NLS(核质蛋白)(krpaatkkagqakkkk,SEQ ID NO:7)为小写字母,3xHA标签(YPYDVPDYAYPYDVPDYAYPYDVPDYA,SEQ ID NO:8)为粗体,VPR为双下划线的

MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTYADQCLQ
LVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKAELF
NGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYRNRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFT
RLITAVPSLREHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNL
AIQKNDETAHIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELN
SIDLTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEIISAAGKEL
SEAFKQKTSEILSHAHAALDQPLPTTLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPEFSARLTGI
KLEMEPSLSFYNKARNYATKKPYSVEKFKLNFQMPTLARGWDVNKEKNNGAILFVKNGLYYLGIMPKQ
KGRYKALSFEPTEKTSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIY
DLNNPEKEPKKFQTAYAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAE
LNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQ
AELFYRPKSRMKRMAHRLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKE
VSHEIIKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIARGERNLIYITVIDSTGKIL
EQRSLNTIQQFDYQKKLDNREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLN
FGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFLFYVP

[0309] APYTSKIDPLTGFVDPFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAW
DIVFEKNETQFDAKGTPFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDS
HAIDTMVALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLL
NHLKESKDLKLQNGISNQDWLAYIQELRNkrpaatkkagqakkkkGS**YPYDVPDYAYPYDVPDYAYPYDVPDY**
AGSEASGSGRADALDDFDLDMLGSDALDDFDLDMLGSDALDDFDLDMLGSDALDDFDLDMLINSRSSG
SPKKKRKVGSQYLPDTDDRHRIEEKRKRTYETFKSIMKKSPFSGPTDPRPPPRRIAVPSRSSASVPKPAPQP
YPFTSSLSTINYDEFPTMVFPSGQISQASALAPAPPQVLPQAPAPAPAPAMVSALAQAPAPVPVLAPGPPQA
VAPPAPKPTQAGEGTLSEALLQLQFDDEDLGALLGNSTDPAVFTDLASVDNSEFQQLLNQGIPVAPHTTEP
MLMEYPEAITRLVTGAQRPPDPAPAPLGAPGLPNGLLSGDEDFSSIADMDFSALLGSGSGSRDSREGMFLP
KPEAGSAISDVFEGREVCQPKRIRPFHPPGSPWANRPLPASLAPTPTGPVHEPVGSLTPAPVPQPLDPAPAVT
PEASHLLEDPDEETSQAVKALREMADTVIPQKEEAAICGQMDLSHPPPRGHLDELTTTLESMTEDLNLDSP
LTPELNEILDTFLNDECLLHAMHISTGLSIFDTSLF SEQ ID NO:434)

[0310] RTW1017:pCAG-2xNLS-hdeAsCas12a(E174R/S542R/K548R/D908A)-NLS-gs-3xHA-gs-VPR(deAs-VPR(1.2))人类密码子优化的氨基酸球菌种BV3L6 Cas12a(AsCas12a)为黑色,DNase失活用修饰的密码子(D908A)eAsCas12a(E174R/S542R/K548R)为双下划线的小写字母,具有沉默突变以移除NcoI位点的密码子为双下划线的大写字母,插入的甘氨酸为点式下划线的,核质蛋白NLS(aaaaggccggcggccacgaaaaaggccggccaggcaaaaaaagaaaaag,SEQ ID NO:21)为小写字母,接头序列为斜体,3xHA标签(YPYDVPDYAYPYDVPDYAYPYDVPDYA,SEQ ID NO:8)为粗体,SV40 NLS为斜体小写字母,VP64-p65-RTA(VPR)为*双下划线的斜体*

[0311] ATGGGC*ccaaagaaaaagaggaaagtc**GGCAGTGGA**cctaaaaagaaacgaaaggtt**GGGTCAGGT***ACACAGTTCGAGGG
CTTTACCAACCTGTATCAGGTGAGCAAGACACTGCGGTTTGAGCTGATCCCACAGGGCAAGACCCTG
AAGCACATCCAGGAGCAGGGCTTCATCGAGGAGGACAAGGCCCGCAATGATCACTACAAGGAGCTG
AAGCCCATCATCGATCGGATCTACAAGACCTATGCCGACCAGTGCCTGCAGCTGGTGCAGCTGGATTG
GGAGAACCTGAGCGCCGCCATCGACTCCTATAGAAAGGAGAAAACCGAGGAGACAAGGAACGCCCT
GATCGAGGAGCAGGCCACATATCGCAATGCCATCCACGACTACTTCATCGGCCGGACAGACAACCTG
ACCGATGCCATCAATAAGAGACACGCCGAGATCTACAAGGGCCTGTTCAAGGCCGAGCTGTTTAATG
GCAAGGTGCTGAAGCAGCTGGGCACCGTGACCACAACCGAGCACGAGAACGCCCTGCTGCGGAGCT
TCGACAAGTTTACAACCTACTTCTCCGGCTTTTATagaAACAGGAAGAACGTGTTCAGCGCCGAGGATA
TCAGCACAGCCATCCCACACCGCATCGTGCAGGACAACTTCCCCAAGTTTAAGGAGAATTGTCACATC
TTCACACGCCTGATCACCGCCGTGCCCAGCCTGCGGGAGCACTTTGAGAACGTGAAGAAGGCCATCG
GCATCTTCGTGAGCACCTCCATCGAGGAGGTGTTTTCCTTCCCTTTTTATAACCAGCTGCTGACACAGA
CCCAGATCGACCTGTATAACCAGCTGCTGGGAGGAATCTCTCGGGAGGCAGGCACCGAGAAGATCAA
GGGCCTGAACGAGGTGCTGAATCTGGCCATCCAGAAGAATGATGAGACAGCCCACATCATCGCCTCC
CTGCCACACAGATTCATCCCCCTGTTTAAGCAGATCCTGTCCGATAGGAACACCCTGTCTTTCATCCTG
GAGGAGTTTAAGAGCGACGAGGAAGTGATCCAGTCCTTCTGCAAGTACAAGACACTGCTGAGAAAC
GAGAACGTGCTGGAGACAGCCGAGGCCCTGTTTAACGAGCTGAACAGCATCGACCTGACACACATCT
TCATCAGCCACAAGAAGCTGGAGACAATCAGCAGCGCCCTGTGCGACCACTGGGATACACTGAGGAA
TGCCCTGTATGAGCGGAGAATCTCCGAGCTGACAGGCAAGATCACCAAGTCTGCCAAGGAGAAGGTG
CAGCGCAGCCTGAAGCACGAGGATATCAACCTGCAGGAGATCATCTCTGCCGCAGGCAAGGAGCTGA
GCGAGGCCTTCAAGCAGAAAACCAGCGAGATCCTGTCCCACGCACACGCCGCCCTGGATCAGCCACT
GCCTACAACCCTGAAGAAGCAGGAGGAGAAGGAGATCCTGAAGTCTCAGCTGGACAGCCTGCTGGG
CCTGTACCACCTGCTGGACTGGTTTGCCGTGGATGAGTCCAACGAGGTGGACCCCGAGTTCTCTGCC
CGGCTGACCGGCATCAAGCTGGAGATGGAGCCTTCTCTGAGCTTCTACAACAAGGCCAGAAATTATG

CCACCAAGAAGCCCTACTCCGTGGAGAAGTTCAAGCTGAACTTTCAGATGCCTACACTGGCCagaGGC
TGGGACGTGAATagaGAGAAGAACAATGGCGCCATCCTGTTTGTGAAGAACGGCCTGTACTATCTGGG
CATCATGCCAAAGCAGAAGGGCAGGTATAAGGCCCTGAGCTTCGAGCCCACAGAGAAAACCAGCGA
GGGCTTTGATAAGATGTACTATGACTACTTCCCTGATGCCGCCAAGATGATCCCAAAGTGCAGCACCC
AGCTGAAGGCCGTGACAGCCCACTTTCAGACCCACACAACCCCCATCCTGCTGTCCAACAATTTCATC
GAGCCTCTGGAGATCACAAAGGAGATCTACGACCTGAACAATCCTGAGAAGGAGCCAAAGAAGTTT
CAGACAGCCTACGCCAAGAAAACCGGCGACCAGAAGGGCTACAGAGAGGCCCTGTGCAAGTGGATC
GACTTCACAAGGGATTTTCTGTCCAAGTATACCAAGACAACCTCTATCGATCTGTCTAGCCTGCGGCCA
TCCTCTCAGTATAAGGACCTGGGCGAGTACTATGCCGAGCTGAATCCCCTGCTGTACCACATCAGCTTC
CAGAGAATCGCCGAGAAGGAGATCATGGATGCCGTGGAGACAGGCAAGCTGTACCTGTTCCAGATCT
ATAACAAGGACTTTGCCAAGGGCCACCACGGCAAGCCTAATCTGCACACACTGTATTGGACCGGCCT
GTTTTCTCCAGAGAACCTGGCCAAGACAAGCATCAAGCTGAATGGCCAGGCCGAGCTGTTCTACCGC
CCTAAGTCCAGGATGAAGAGGATGGCACACCGGCTGGGAGAGAAGATGCTGAACAAGAAGCTGAAG
GATCAGAAAACCCCAATCCCCGACACCCTGTACCAGGAGCTGTACGACTATGTGAATCACAGACTGTC
CCACGACCTGTCTGATGAGGCCAGGGCCCTGCTGCCCAACGTGATCACCAAGGAGGTGTCTCACGAG
ATCATCAAGGATAGGCGCTTTACCAGCGACAAGTTCTTTTTCCACGTGCCTATCACACTGAACTATCAG
GCCGCCAATTCCCCATCTAAGTTCAACCAGAGGGTGAATGCCTACCTGAAGGAGCACCCCGAGACAC
CTATCATCGGCATCgccCGGGGCGAGAGAAACCTGATCTATATCACAGTGATCGACTCCACCGGCAAGA
TCCTGGAGCAGCGGAGCCTGAACACCATCCAGCAGTTTGATTACCAGAAGAAGCTGGACAACAGGG
AGAAGGAGAGGGTGGCAGCAAGGCAGGCCTGGTCTGTGGTGGGCACAATCAAGGATCTGAAGCAGG
GCTATCTGAGCCAGGTCATCCACGAGATCGTGGACCTGATGATCCACTACCAGGCCGTGGTGGTGCTG
GAGAACCTGAATTTCGGCTTTAAGAGCAAGAGGACCGGCATCGCCGAGAAGGCCGTGTACCAGCAGT
[0312] TCGAGAAGATGCTGATCGATAAGCTGAATTGCCTGGTGCTGAAGGACTATCCAGCAGAGAAAGTGGG
AGGCGTGCTGAACCCATACCAGCTGACAGACCAGTTCACCTCCTTTGCCAAGATGGGCACCCAGTCT
GGCTTCCTGTTTTACGTGCCTGCCCCATATACATCTAAGATCGATCCCCTGACCGGCTTCGTGGACCCC
TTCGTGTGGAAAACCATCAAGAATCACGAGAGCCGCAAGCACTTCCTGGAGGGCTTCGACTTTCTGC
ACTACGACGTGAAAACCGGCGACTTCATCCTGCACTTTAAGATGAACAGAAATCTGTCCTTCCAGAG
GGGCCTGCCCGGCTTTATGCCTGCATGGGATATCGTGTTCGAGAAGAACGAGACACAGTTTGACGCCA
AGGGCACCCCTTTCATCGCCGGCAAGAGAATCGTGCCAGTGATCGAGAATCACAGATTCACCGGCAG
ATACCGGGACCTGTATCCTGCCAACGAGCTGATCGCCCTGCTGGAGGAGAAGGGCATCGTGTTCAGG
GATGGCTCCAACATCCTGCCAAAGCTGCTGGAGAATGACGATTCTCACGCCATCGAC**ACG**ATGGTGG
CCCTGATCCGCAGCGTGCTGCAGATGCGGAACTCCAATGCCGCCACAGGCGAGGACTATATCAACAG
CCCCGTGCGCGATCTGAATGGCGTGTGCTTCGACTCCCGGTTTCAGAACCCAGAGTGG**CCA**ATGGAC
GCCGATGCCAATGGCGCCTACCACATCGCCCTGAAGGGCCAGCTGCTGCTGAATCACCTGAAGGAGA
GCAAGGATCTGAAGCTGCAGAACGGCATCTCCAATCAGGACTGGCTGGCCTACATCCAGGAGCTGCG
CAACaaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaag*GGATCC***TACCCATACGATGTTCCAGATTAC**
GCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCC*GGAAGCGAG*
GCCAGCGGTTCCGGACGGGCTGACGCATTGGACGATTTTGATCTGGATATGCTGGGAAGTGACGCCCTCG
ATGATTTTGACCTTGACATGCTTGGTTCGGATGCCCTTGATGACTTTGACCTCGACATGCTCGGCAGTGACG
CCCTTGATGATTTCGACCTGGACATGCTGATTAACTCTAGAAGTTCCGGATCTCCGAAAAAGAAACGCAAAGT
TGGTAGCCAGTACCTGCCCGACACCGACGACCGGCACCGGATCGAGGAAAAGCGGAAGCGGACCTACGA
GACATTCAAGAGCATCATGAAGAAGTCCCCCTTCAGCGGCCCCACCGACCCTAGACCTCCACCTAGAAGAA
TCGCCGTGCCCAGCAGATCCAGCGCCAGCGTGCCAAAACCTGCCCCCCAGCCTTACCCCTTCACCAGCAG
CCTGAGCACCATCAACTACGACGAGTTCCCTACCATGGTGTTCCCCAGCGGCCAGATCTCTCAGGCCTCTG
CTCTGGCTCCAGCCCCTCCTCAGGTGCTGCCTCAGGCTCCTGCTCCTGCACCAGCTCCAGCCATGGTGTC

[0313] *TGCACTGGCTCAGGCACCAGCACCCGTGCCTGTGCTGGCTCCTGGACCTCCACAGGCTGTGGCTCCACCA GCCCCTAAACCTACACAGGCCGGCGAGGGCACACTGTCTGAAGCTCTGCTGCAGCTGCAGTTCGACGACG AGGATCTGGGAGCCCTGCTGGGAAACAGCACCGATCCTGCCGTGTTCACCGACCTGGCCAGCGTGGACAA CAGCGAGTTCCAGCAGCTGCTGAACCAGGGCATCCCTGTGGCCCCTCACACCACCGAGCCCATGCTGATG GAATACCCCGAGGCCATCACCCGGCTCGTGACAGGCGCTCAGAGGCCTCCTGATCCAGCTCCTGCCCCTC TGGGAGCACCAGGCCTGCCTAATGGACTGCTGTCTGGCGACGAGGACTTCAGCTCTATCGCCGATATGGAT TTCTCAGCCTTGCTGGGCTCTGGCAGCGGCAGCCGGGATTCCAGGGAAGGGATGTTTTTGCCGAAGCCTG AGGCCGGCTCCGCTATTAGTGACGTGTTTGAGGGCCGCGAGGTGTGCCAGCCAAAACGAATCCGGCCATTT CATCCTCCAGGAAGTCCATGGGCCAACCGCCCACTCCCCGCCAGCCTCGCACCAACACCAACCGGTCCAG TACATGAGCCAGTCGGGTCACTGACCCCGGCACCAGTCCCTCAGCCACTGGATCCAGCGCCCGCAGTGAC TCCCGAGGCCAGTCACCTGTTGGAGGATCCCGATGAAGAGACGAGCCAGGCTGTCAAAGCCCTTCGGGAG ATGGCCGATACTGTGATTCCCCAGAAGGAAGAGGCTGCAATCTGTGGCCAAATGGACCTTTCCCATCCGCC CCCAAGGGGCCATCTGGATGAGCTGACAACCACACTTGAGTCCATGACCGAGGATCTGAACCTGGACTCAC CCCTGACCCCGGAATTGAACGAGATTCTGGATACCTTCCTGAACGACGAGTGCCTCTTGCATGCCATGCATA TCAGCACAGGACTGTCCATCTTCGACACATCTCTGTTT*TAA (SEQ ID NO: 435)

[0314] RTW1130:pCAG-hdeAsCas12a(E174R/S542R/K548R/D908A)-gs-NLS-gs-VPR(deAs-VPR(1.3))人类密码子优化的氨基酸球菌种BV3L6 Cas12a(AsCas12a)为黑色,DNase失活用修饰的密码子(D908A)eAsCas12a(E174R/S542R/K548R)为双下划线的小写字母,具有沉默突变以移除NcoI位点的密码子为双下划线的大 写字母,接头序列为斜体,SV40NLS为斜体小写字母,VP64-p65-RTA(VPR)为*双下划线的斜体*

[0315] ATGACACAGTTCGAGGGCTTTACCAACCTGTATCAGGTGAGCAAGACACTGCGGTTTGAGCTGATCCC ACAGGGCAAGACCCTGAAGCACATCCAGGAGCAGGGCTTCATCGAGGAGGACAAGGCCCGCAATGA TCACTACAAGGAGCTGAAGCCCATCATCGATCGGATCTACAAGACCTATGCCGACCAGTGCCTGCAGC TGGTGCAGCTGGATTGGGAGAACCTGAGCGCCGCCATCGACTCCTATAGAAAGGAGAAAACCGAGG AGACAAGGAACGCCCTGATCGAGGAGCAGGCCACATATCGCAATGCCATCCACGACTACTTCATCGG CCGGACAGACAACCTGACCGATGCCATCAATAAGAGACACGCCGAGATCTACAAGGGCCTGTTCAAG GCCGAGCTGTTTAATGGCAAGGTGCTGAAGCAGCTGGGCACCGTGACCACAACCGAGCACGAGAAC GCCCTGCTGCGGAGCTTCGACAAGTTTACAACCTACTTCTCCGGCTTTTATagaAACAGGAAGAACGTG TTCAGCGCCGAGGATATCAGCACAGCCATCCCACACCGCATCGTGCAGGACAACTTCCCCAAGTTTAA GGAGAATTGTCACATCTTCACACGCCTGATCACCGCCGTGCCCAGCCTGCGGGAGCACTTTGAGAAC GTGAAGAAGGCCATCGGCATCTTCGTGAGCACCTCCATCGAGGAGGTGTTTTCCTTCCCTTTTTATAAC CAGCTGCTGACACAGACCCAGATCGACCTGTATAACCAGCTGCTGGGAGGAATCTCTCGGGAGGCAG GCACCGAGAAGATCAAGGGCCTGAACGAGGTGCTGAATCTGGCCATCCAGAAGAATGATGAGACAG CCCACATCATCGCCTCCCTGCCACACAGATTCATCCCCCTGTTTAAGCAGATCCTGTCCGATAGGAACA CCCTGTCTTTCATCCTGGAGGAGTTTAAGAGCGACGAGGAAGTGATCCAGTCCTTCTGCAAGTACAA GACACTGCTGAGAAACGAGAACGTGCTGGAGACAGCCGAGGCCCTGTTTAACGAGCTGAACAGCAT CGACCTGACACACATCTTCATCAGCCACAAGAAGCTGGAGACAATCAGCAGCGCCCTGTGCGACCAC TGGGATACACTGAGGAATGCCCTGTATGAGCGGAGAATCTCCGAGCTGACAGGCAAGATCACCAAGT CTGCCAAGGAGAAGGTGCAGCGCAGCCTGAAGCACGAGGATATCAACCTGCAGGAGATCATCTCTGC CGCAGGCAAGGAGCTGAGCGAGGCCTTCAAGCAGAAAACCAGCGAGATCCTGTCCCACGCACACGC CGCCCTGGATCAGCCACTGCCTACAACCCTGAAGAAGCAGGAGGAGAAGGAGATCCTGAAGTCTCA GCTGGACAGCCTGCTGGGCCTGTACCACCTGCTGGACTGGTTTGCCGTGGATGAGTCCAACGAGGTG GACCCCGAGTTCTCTGCCCGGCTGACCGGCATCAAGCTGGAGATGGAGCCTTCTCTGAGCTTCTACA ACAAGGCCAGAAATTATGCCACCAAGAAGCCCTACTCCGTGGAGAAGTTCAAGCTGAACTTTCAGAT GCCTACACTGGCCagaGGCTGGGACGTGAATagaGAGAAGAACAATGGCGCCATCCTGTTTGTGAAGAA

CGGCCTGTACTATCTGGGCATCATGCCAAAGCAGAAGGGCAGGTATAAGGCCCTGAGCTTCGAGCCC
ACAGAGAAAACCAGCGAGGGCTTTGATAAGATGTACTATGACTACTTCCCTGATGCCGCCAAGATGAT
CCCAAAGTGCAGCACCCAGCTGAAGGCCGTGACAGCCCACTTTCAGACCCACACAACCCCCATCCTG
CTGTCCAACAATTTCATCGAGCCTCTGGAGATCACAAAGGAGATCTACGACCTGAACAATCCTGAGA
AGGAGCCAAAGAAGTTTCAGACAGCCTACGCCAAGAAAACCGGCGACCAGAAGGGCTACAGAGAG
GCCCTGTGCAAGTGGATCGACTTCACAAGGGATTTTCTGTCCAAGTATACCAAGACAACCTCTATCGA
TCTGTCTAGCCTGCGGCCATCCTCTCAGTATAAGGACCTGGGCGAGTACTATGCCGAGCTGAATCCCCT
GCTGTACCACATCAGCTTCCAGAGAATCGCCGAGAAGGAGATCATGGATGCCGTGGAGACAGGCAAG
CTGTACCTGTTCCAGATCTATAACAAGGACTTTGCCAAGGGCCACCACGGCAAGCCTAATCTGCACAC
ACTGTATTGGACCGGCCTGTTTTCTCCAGAGAACCTGGCCAAGACAAGCATCAAGCTGAATGGCCAG
GCCGAGCTGTTCTACCGCCCTAAGTCCAGGATGAAGAGGATGGCACACCGGCTGGGAGAGAAGATGC
TGAACAAGAAGCTGAAGGATCAGAAAACCCCAATCCCCGACACCCTGTACCAGGAGCTGTACGACTA
TGTGAATCACAGACTGTCCCACGACCTGTCTGATGAGGCCAGGGCCCTGCTGCCCAACGTGATCACC
AAGGAGGTGTCTCACGAGATCATCAAGGATAGGCGCTTTACCAGCGACAAGTTCTTTTTCCACGTGCC
TATCACACTGAACTATCAGGCCGCCAATTCCCCATCTAAGTTCAACCAGAGGGTGAATGCCTACCTGA
AGGAGCACCCCGAGACACCTATCATCGGCATCgccCGGGGCGAGAGAAACCTGATCTATATCACAGTGA
TCGACTCCACCGGCAAGATCCTGGAGCAGCGGAGCCTGAACACCATCCAGCAGTTTGATTACCAGAA
GAAGCTGGACAACAGGGAGAAGGAGAGGGTGGCAGCAAGGCAGGCCTGGTCTGTGGTGGGCACAA
TCAAGGATCTGAAGCAGGGCTATCTGAGCCAGGTCATCCACGAGATCGTGGACCTGATGATCCACTAC
CAGGCCGTGGTGGTGCTGGAGAACCTGAATTTCGGCTTTAAGAGCAAGAGGACCGGCATCGCCGAGA
AGGCCGTGTACCAGCAGTTCGAGAAGATGCTGATCGATAAGCTGAATTGCCTGGTGCTGAAGGACTAT
CCAGCAGAGAAAGTGGGAGGCGTGCTGAACCCATACCAGCTGACAGACCAGTTCACCTCCTTTGCCA
[0316] AGATGGGCACCCAGTCTGGCTTCCTGTTTTACGTGCCTGCCCCATATACATCTAAGATCGATCCCCTGA
CCGGCTTCGTGGACCCCTTCGTGTGGAAAACCATCAAGAATCACGAGAGCCGCAAGCACTTCCTGGA
GGGCTTCGACTTTCTGCACTACGACGTGAAAACCGGCGACTTCATCCTGCACTTTAAGATGAACAGA
AATCTGTCCTTCCAGAGGGGCCTGCCCGGCTTTATGCCTGCATGGGATATCGTGTTCGAGAAGAACGA
GACACAGTTTGACGCCAAGGGCACCCCTTTCATCGCCGGCAAGAGAATCGTGCCAGTGATCGAGAAT
CACAGATTCACCGGCAGATACCGGGACCTGTATCCTGCCAACGAGCTGATCGCCCTGCTGGAGGAGA
AGGGCATCGTGTTCAGGGATGGCTCCAACATCCTGCCAAAGCTGCTGGAGAATGACGATTCTCACGC
CATCGACACGATGGTGGCCCTGATCCGCAGCGTGCTGCAGATGCGGAACTCCAATGCCGCCACAGGC
GAGGACTATATCAACAGCCCCGTGCGCGATCTGAATGGCGTGTGCTTCGACTCCCGGTTTCAGAACCC
AGAGTGGCCAATGGACGCCGATGCCAATGGCGCCTACCACATCGCCCTGAAGGGCCAGCTGCTGCTG
AATCACCTGAAGGAGAGCAAGGATCTGAAGCTGCAGAACGGCATCTCCAATCAGGACTGGCTGGCCT
ACATCCAGGAGCTGCGCAAC*GGTGGAAGCGGAGGGAGTcccaagaagaagaggaaagtcGGGGGTTCCGGAGG*
AAGCGAGGCCAGCGGTTCCGGACGGGCTGACGCATTGGACGATTTTGATCTGGATATGCTGGGAAGTGAC
GCCCTCGATGATTTTGACCTTGACATGCTTGGTTCGGATGCCCTTGATGACTTTGACCTCGACATGCTCGGC
AGTGACGCCCTTGATGATTTCGACCTGGACATGCTGATTAACTCTAGAAGTTCCGGATCTCCGAAAAAGAAA
CGCAAAGTTGGTAGCCAGTACCTGCCCGACACCGACGACCGGCACCGGATCGAGGAAAAGCGGAAGCGG
ACCTACGAGACATTCAAGAGCATCATGAAGAAGTCCCCCTTCAGCGGCCCCACCGACCCTAGACCTCCACC
TAGAAGAATCGCCGTGCCCAGCAGATCCAGCGCCAGCGTGCCAAAACCTGCCCCCCAGCCTTACCCCTTCA
CCAGCAGCCTGAGCACCATCAACTACGACGAGTTCCCTACCATGGTGTTCCCCAGCGGCCAGATCTCTCAG
GCCTCTGCTCTGGCTCCAGCCCCTCCTCAGGTGCTGCCTCAGGCTCCTGCTCCTGCACCAGCTCCAGCCA
TGGTGTCTGCACTGGCTCAGGCACCAGCACCCGTGCCTGTGCTGGCTCCTGGACCTCCACAGGCTGTGGC
TCCACCAGCCCCTAAACCTACACAGGCCGGCGAGGGCACACTGTCTGAAGCTCTGCTGCAGCTGCAGTTC
GACGACGAGGATCTGGGAGCCCTGCTGGGAAACAGCACCGATCCTGCCGTGTTCACCGACCTGGCCAGC

[0317] *GTGGACAACAGCGAGTTCCAGCAGCTGCTGAACCAGGGCATCCCTGTGGCCCCTCACACCACCGAGCCCA TGCTGATGGAATACCCCGAGGCCATCACCCGGCTCGTGACAGGCGCTCAGAGGCCTCCTGATCCAGCTCC TGCCCCTCTGGGAGCACCAGGCCTGCCTAATGGACTGCTGTCTGGCGACGAGGACTTCAGCTCTATCGCC GATATGGATTTCTCAGCCTTGCTGGGCTCTGGCAGCGGCAGCCGGGATTCCAGGGAAGGGATGTTTTTGCC GAAGCCTGAGGCCGGCTCCGCTATTAGTGACGTGTTTGAGGGCCGCGAGGTGTGCCAGCCAAAACGAATC CGGCCATTTCATCCTCCAGGAAGTCCATGGGCCAACCGCCCACTCCCCGCCAGCCTCGCACCAACACCAA CCGGTCCAGTACATGAGCCAGTCGGGTCACTGACCCCGGCACCAGTCCCTCAGCCACTGGATCCAGCGCC CGCAGTGACTCCCGAGGCCAGTCACCTGTTGGAGGATCCCGATGAAGAGACGAGCCAGGCTGTCAAAGCC CTTCGGGAGATGGCCGATACTGTGATTCCCCAGAAGGAAGAGGCTGCAATCTGTGGCCAAATGGACCTTTC CCATCCGCCCCCAAGGGGCCATCTGGATGAGCTGACAACCACACTTGAGTCCATGACCGAGGATCTGAACC TGGACTCACCCCTGACCCCGGAATTGAACGAGATTCTGGATACCTTCCTGAACGACGAGTGCCTCTTGCAT GCCATGCATATCAGCACAGGACTGTCCATCTTCGACACATCTCTGTTT*TAA (SEQ ID NO: 436)

[0318] RTW1319:pCAG-2xNLS-hdeAsCas12a(E174R/S542R/K548R/D908A)-gs-NLS-gs-VPR(deAs-VPR(1.4))

[0319] 人类密码子优化的氨基酸球菌种BV3L6 Cas12a(AsCas12a)为黑色,DNase失活用修饰的密码子(D908A)eAsCas12a(E174R/S542R/K548R)为双下划线的小写字母,具有沉默突变以移除NcoI位点的密码子为双 下划线的大写字母,插入的甘氨酸为点式下划线的,接头序列为斜体,3xHA标签(YPYDVPDYAYPYDVPDYAYPYDVPDYA,SEQ ID NO:8)为粗体,SV40 NLS为斜体小写字母,VP64-p65-RTA(VPR)为*双下划线的斜体*

[0320] ATGGGC*ccaaagaaaaagaggaaagtcGGCAGTGGAcctaaaaagaaacgaaaggttGGGTCAGGT*ACACAGTTCGAGGG
CTTTACCAACCTGTATCAGGTGAGCAAGACACTGCGGTTTGAGCTGATCCCACAGGGCAAGACCCTG
AAGCACATCCAGGAGCAGGGCTTCATCGAGGAGGACAAGGCCCGCAATGATCACTACAAGGAGCTG
AAGCCCATCATCGATCGGATCTACAAGACCTATGCCGACCAGTGCCTGCAGCTGGTGCAGCTGGATTG
GGAGAACCTGAGCGCCGCCATCGACTCCTATAGAAAGGAGAAAACCGAGGAGACAAGGAACGCCCT
GATCGAGGAGCAGGCCACATATCGCAATGCCATCCACGACTACTTCATCGGCCGGACAGACAACCTG
ACCGATGCCATCAATAAGAGACACGCCGAGATCTACAAGGGCCTGTTCAAGGCCGAGCTGTTTAATG
GCAAGGTGCTGAAGCAGCTGGGCACCGTGACCACAACCGAGCACGAGAACGCCCTGCTGCGGAGCT
TCGACAAGTTTACAACCTACTTCTCCGGCTTTTATagaAACAGGAAGAACGTGTTCAGCGCCGAGGATA
TCAGCACAGCCATCCCACACCGCATCGTGCAGGACAACTTCCCCAAGTTTAAGGAGAATTGTCACATC
TTCACACGCCTGATCACCGCCGTGCCCAGCCTGCGGGAGCACTTTGAGAACGTGAAGAAGGCCATCG
GCATCTTCGTGAGCACCTCCATCGAGGAGGTGTTTTCCTTCCCTTTTTATAACCAGCTGCTGACACAGA
CCCAGATCGACCTGTATAACCAGCTGCTGGGAGGAATCTCTCGGGAGGCAGGCACCGAGAAGATCAA
GGGCCTGAACGAGGTGCTGAATCTGGCCATCCAGAAGAATGATGAGACAGCCCACATCATCGCCTCC
CTGCCACACAGATTCATCCCCCTGTTTAAGCAGATCCTGTCCGATAGGAACACCCTGTCTTTCATCCTG
GAGGAGTTTAAGAGCGACGAGGAAGTGATCCAGTCCTTCTGCAAGTACAAGACACTGCTGAGAAAC
GAGAACGTGCTGGAGACAGCCGAGGCCCTGTTTAACGAGCTGAACAGCATCGACCTGACACACATCT
TCATCAGCCACAAGAAGCTGGAGACAATCAGCAGCGCCCTGTGCGACCACTGGGATACACTGAGGAA
TGCCCTGTATGAGCGGAGAATCTCCGAGCTGACAGGCAAGATCACCAAGTCTGCCAAGGAGAAGGTG
CAGCGCAGCCTGAAGCACGAGGATATCAACCTGCAGGAGATCATCTCTGCCGCAGGCAAGGAGCTGA
GCGAGGCCTTCAAGCAGAAAACCAGCGAGATCCTGTCCCACGCACACGCCGCCCTGGATCAGCCACT
GCCTACAACCCTGAAGAAGCAGGAGGAGAAGGAGATCCTGAAGTCTCAGCTGGACAGCCTGCTGGG
CCTGTACCACCTGCTGGACTGGTTTGCCGTGGATGAGTCCAACGAGGTGGACCCCGAGTTCTCTGCC
CGGCTGACCGGCATCAAGCTGGAGATGGAGCCTTCTCTGAGCTTCTACAACAAGGCCAGAAATTATG
CCACCAAGAAGCCCTACTCCGTGGAGAAGTTCAAGCTGAACTTTCAGATGCCTACACTGGCCagaGGC
TGGGACGTGAATAGAGAGAAGAACAATGGCGCCATCCTGTTTGTGAAGAACGGCCTGTACTATCTGG

GCATCATGCCAAAGCAGAAGGGCAGGTATAAGGCCCTGAGCTTCGAGCCCACAGAGAAAACCAGCG
AGGGCTTTGATAAGATGTACTATGACTACTTCCCTGATGCCGCCAAGATGATCCCAAAGTGCAGCACC
CAGCTGAAGGCCGTGACAGCCCACTTTCAGACCCACACAACCCCCATCCTGCTGTCCAACAATTTCAT
CGAGCCTCTGGAGATCACAAAGGAGATCTACGACCTGAACAATCCTGAGAAGGAGCCAAAGAAGTT
TCAGACAGCCTACGCCAAGAAAACCGGCGACCAGAAGGGCTACAGAGAGGCCCTGTGCAAGTGGAT
CGACTTCACAAGGGATTTTCTGTCCAAGTATACCAAGACAACCTCTATCGATCTGTCTAGCCTGCGGCC
ATCCTCTCAGTATAAGGACCTGGGCGAGTACTATGCCGAGCTGAATCCCCTGCTGTACCACATCAGCTT
CCAGAGAATCGCCGAGAAGGAGATCATGGATGCCGTGGAGACAGGCAAGCTGTACCTGTTCCAGATC
TATAACAAGGACTTTGCCAAGGGCCACCACGGCAAGCCTAATCTGCACACACTGTATTGGACCGGCCT
GTTTTCTCCAGAGAACCTGGCCAAGACAAGCATCAAGCTGAATGGCCAGGCCGAGCTGTTCTACCGC
CCTAAGTCCAGGATGAAGAGGATGGCACACCGGCTGGGAGAGAAGATGCTGAACAAGAAGCTGAAG
GATCAGAAAACCCCAATCCCCGACACCCTGTACCAGGAGCTGTACGACTATGTGAATCACAGACTGTC
CCACGACCTGTCTGATGAGGCCAGGGCCCTGCTGCCCAACGTGATCACCAAGGAGGTGTCTCACGAG
ATCATCAAGGATAGGCGCTTTACCAGCGACAAGTTCTTTTTCCACGTGCCTATCACACTGAACTATCAG
GCCGCCAATTCCCCATCTAAGTTCAACCAGAGGGTGAATGCCTACCTGAAGGAGCACCCCGAGACAC
CTATCATCGGCATCgccCGGGGCGAGAGAAACCTGATCTATATCACAGTGATCGACTCCACCGGCAAGA
TCCTGGAGCAGCGGAGCCTGAACACCATCCAGCAGTTTGATTACCAGAAGAAGCTGGACAACAGGG
AGAAGGAGAGGGTGGCAGCAAGGCAGGCCTGGTCTGTGGTGGGCACAATCAAGGATCTGAAGCAGG
GCTATCTGAGCCAGGTCATCCACGAGATCGTGGACCTGATGATCCACTACCAGGCCGTGGTGGTGCTG
GAGAACCTGAATTTCGGCTTTAAGAGCAAGAGGACCGGCATCGCCGAGAAGGCCGTGTACCAGCAGT
TCGAGAAGATGCTGATCGATAAGCTGAATTGCCTGGTGCTGAAGGACTATCCAGCAGAGAAAGTGGG
AGGCGTGCTGAACCCATACCAGCTGACAGACCAGTTCACCTCCTTTGCCAAGATGGGCACCCAGTCT
[0321] GGCTTCCTGTTTTACGTGCCTGCCCCATATACATCTAAGATCGATCCCCTGACCGGCTTCGTGGACCCC
TTCGTGTGGAAAACCATCAAGAATCACGAGAGCCGCAAGCACTTCCTGGAGGGCTTCGACTTTCTGC
ACTACGACGTGAAAACCGGCGACTTCATCCTGCACTTTAAGATGAACAGAAATCTGTCCTTCCAGAG
GGGCCTGCCCGGCTTTATGCCTGCATGGGATATCGTGTTCGAGAAGAACGAGACACAGTTTGACGCCA
AGGGCACCCCTTTCATCGCCGGCAAGAGAATCGTGCCAGTGATCGAGAATCACAGATTCACCGGCAG
ATACCGGGACCTGTATCCTGCCAACGAGCTGATCGCCCTGCTGGAGGAGAAGGGCATCGTGTTCAGG
GATGGCTCCAACATCCTGCCAAAGCTGCTGGAGAATGACGATTCTCACGCCATCGACACGATGGTGGC
CCTGATCCGCAGCGTGCTGCAGATGCGGAACTCCAATGCCGCCACAGGCGAGGACTATATCAACAGC
CCCGTGCGCGATCTGAATGGCGTGTGCTTCGACTCCCGGTTTCAGAACCCAGAGTGGCCAATGGACG
CCGATGCCAATGGCGCCTACCACATCGCCCTGAAGGGCCAGCTGCTGCTGAATCACCTGAAGGAGAG
CAAGGATCTGAAGCTGCAGAACGGCATCTCCAATCAGGACTGGCTGGCCTACATCCAGGAGCTGCGC
AAC*GGTGGAAGCGGAGGGAGTcccaagaagaagaggaaagtcGGGGGTTCCGGAGGAAGC*GAGGCCAGCGGTTC
CGGACGGGCTGACGCATTGGACGATTTTGATCTGGATATGCTGGGAAGTGACGCCCTCGATGATTTTGACC
TTGACATGCTTGGTTCGGATGCCCTTGATGACTTTGACCTCGACATGCTCGGCAGTGACGCCCTTGATGATT
TCGACCTGGACATGCTGATTAACTCTAGAAGTTCCGGATCTCCGAAAAAGAAACGCAAAGTTGGTAGCCAGT
ACCTGCCCGACACCGACGACCGGCACCGGATCGAGGAAAAGCGGAAGCGGACCTACGAGACATTCAAGAG
CATCATGAAGAAGTCCCCCTTCAGCGGCCCCACCGACCCTAGACCTCCACCTAGAAGAATCGCCGTGCCCA
GCAGATCCAGCGCCAGCGTGCCAAAACCTGCCCCCCAGCCTTACCCCTTCACCAGCAGCCTGAGCACCAT
CAACTACGACGAGTTCCCTACCATGGTGTTCCCCAGCGGCCAGATCTCTCAGGCCTCTGCTCTGGCTCCAG
CCCCTCCTCAGGTGCTGCCTCAGGCTCCTGCTCCTGCACCAGCTCCAGCCATGGTGTCTGCACTGGCTCA
GGCACCAGCACCCGTGCCTGTGCTGGCTCCTGGACCTCCACAGGCTGTGGCTCCACCAGCCCCTAAACCT
ACACAGGCCGGCGAGGGCACACTGTCTGAAGCTCTGCTGCAGCTGCAGTTCGACGACGAGGATCTGGGA
GCCCTGCTGGGAAACAGCACCGATCCTGCCGTGTTCACCGACCTGGCCAGCGTGGACAACAGCGAGTTCC

[0322] *AGCAGCTGCTGAACCAGGGCATCCCTGTGGCCCCTCACACCACCGAGCCCATGCTGATGGAATACCCCGA GGCCATCACCCGGCTCGTGACAGGCGCTCAGAGGCCTCCTGATCCAGCTCCTGCCCCTCTGGGAGCACCA GGCCTGCCTAATGGACTGCTGTCTGGCGACGAGGACTTCAGCTCTATCGCCGATATGGATTTCTCAGCCTTG CTGGGCTCTGGCAGCGGCAGCCGGGATTCCAGGGAAGGGATGTTTTTGCCGAAGCCTGAGGCCGGCTCC GCTATTAGTGACGTGTTTGAGGGCCGCGAGGTGTGCCAGCCAAAACGAATCCGGCCATTTCATCCTCCAGG AAGTCCATGGGCCAACCGCCCACTCCCCGCCAGCCTCGCACCAACACCAACCGGTCCAGTACATGAGCCA GTCGGGTCACTGACCCCGGCACCAGTCCCTCAGCCACTGGATCCAGCGCCCGCAGTGACTCCCGAGGCC AGTCACCTGTTGGAGGATCCCGATGAAGAGACGAGCCAGGCTGTCAAAGCCCTTCGGGAGATGGCCGATA CTGTGATTCCCCAGAAGGAAGAGGCTGCAATCTGTGGCCAAATGGACCTTTCCCATCCGCCCCCAAGGGGC CATCTGGATGAGCTGACAACCACACTTGAGTCCATGACCGAGGATCTGAACCTGGACTCACCCCTGACCCC GGAATTGAACGAGATTCTGGATACCTTCCTGAACGACGAGTGCCTCTTGCATGCCATGCATATCAGCACAGG ACTGTCCATCTTCGACACATCTCTGTTT*TAA (SEQ ID NO:437)

[0323] RTW1351:pCAG-rAPOBEC1-gs-XTEN-gs-hdAsCas12a(D908A)-NLS-gs-UGI-NLS (AsBE1.1)

[0324] 大鼠APOBEC1(rAPOBEC1)为粗体大写字母,插入的甘氨酸为点式下划线的,甘氨酸/丝氨酸接头为斜体,XTEN接头为斜体小写字母,人类密码子优化的氨基酸球菌种BV3L6 Cas12a(AsCas12a)为黑色,DNase失活用修饰的密码子(D908A)AsCas12a为双下划线的小写字母,具有沉默突变以移除NcoI位点的密码子为双下划线的大写字母,核质蛋白NLS为粗体小写字母,UGI为**虚下划线的粗体大写字母**,SV40NLS为*双下划线的斜体*

[0325] **ATGGGCAGCTCAGAGACTGGCCCAGTGGCTGTGGACCCCACATTGAGACGGCGGATCGAGCCC CATGAGTTTGAGGTATTCTTCGATCCGAGAGAGCTCCGCAAGGAGACCTGCCTGCTTTACGAAA TTAATTGGGGGGGCCGGCACTCCATTTGGCGACATACATCACAGAACACTAACAAGCACGTCG AAGTCAACTTCATCGAGAAGTTCACGACAGAAAGATATTTCTGTCCGAACACAAGGTGCAGCAT TACCTGGTTTCTCAGCTGGAGCCCATGCGGCGAATGTAGTAGGGCCATCACTGAATTCCTGTCA AGGTATCCCCACGTCACTCTGTTTATTTACATCGCAAGGCTGTACCACCACGCTGACCCCCGCA ATCGACAAGGCCTGCGGGATTTGATCTCTTCAGGTGTGACTATCCAAATTATGACTGAGCAGGA GTCAGGATACTGCTGGAGAAACTTTGTGAATTATAGCCCGAGTAATGAAGCCCACTGGCCTAGG TATCCCCATCTGTGGGTACGACTGTACGTTCTTGAACTGTACTGCATCATACTGGGCCTGCCTC CTTGTCTCAACATTCTGAGAAGGAAGCAGCCACAGCTGACATTCTTTACCATCGCTCTTCAGTC TTGTCATTACCAGCGACTGCCCCCACACATTCTCTGGGCCACCGGGTTGAAA*TCTGGTGGTTCTT CTGGTGGTTCTagcggcagcgagactcccgggacctcagagtccgccacacccgaaagtTCCGGAGGGAGTAGCGGCGGGTCTA* CACAGTTCGAGGGCTTTACCAACCTGTATCAGGTGAGCAAGACACTGCGGTTTGAGCTGATCCCACA GGGCAAGACCCTGAAGCACATCCAGGAGCAGGGCTTCATCGAGGAGGACAAGGCCCGCAATGATCA CTACAAGGAGCTGAAGCCCATCATCGATCGGATCTACAAGACCTATGCCGACCAGTGCCTGCAGCTGG TGCAGCTGGATTGGGAGAACCTGAGCGCCGCCATCGACTCCTATAGAAAGGAGAAAACCGAGGAGA CAAGGAACGCCCTGATCGAGGAGCAGGCCACATATCGCAATGCCATCCACGACTACTTCATCGGCCGG ACAGACAACCTGACCGATGCCATCAATAAGAGACACGCCGAGATCTACAAGGGCCTGTTCAAGGCCG AGCTGTTTAATGGCAAGGTGCTGAAGCAGCTGGGCACCGTGACCACAACCGAGCACGAGAACGCCC TGCTGCGGAGCTTCGACAAGTTTACAACCTACTTCTCCGGCTTTTATGAGAACAGGAAGAACGTGTTC AGCGCCGAGGATATCAGCACAGCCATCCCACACCGCATCGTGCAGGACAACTTCCCCAAGTTTAAGG AGAATTGTCACATCTTCACACGCCTGATCACCGCCGTGCCCAGCCTGCGGGAGCACTTTGAGAACGT GAAGAAGGCCATCGGCATCTTCGTGAGCACCTCCATCGAGGAGGTGTTTTCCTTCCCTTTTTATAACC AGCTGCTGACACAGACCCAGATCGACCTGTATAACCAGCTGCTGGGAGGAATCTCTCGGGAGGCAGG CACCGAGAAGATCAAGGGCCTGAACGAGGTGCTGAATCTGGCCATCCAGAAGAATGATGAGACAGC

CCACATCATCGCCTCCCTGCCACACAGATTCATCCCCCTGTTTAAGCAGATCCTGTCCGATAGGAACAC
CCTGTCTTTCATCCTGGAGGAGTTTAAGAGCGACGAGGAAGTGATCCAGTCCTTCTGCAAGTACAAG
ACACTGCTGAGAAACGAGAACGTGCTGGAGACAGCCGAGGCCCTGTTTAACGAGCTGAACAGCATC
GACCTGACACACATCTTCATCAGCCACAAGAAGCTGGAGACAATCAGCAGCGCCCTGTGCGACCACT
GGGATACACTGAGGAATGCCCTGTATGAGCGGAGAATCTCCGAGCTGACAGGCAAGATCACCAAGTC
TGCCAAGGAGAAGGTGCAGCGCAGCCTGAAGCACGAGGATATCAACCTGCAGGAGATCATCTCTGCC
GCAGGCAAGGAGCTGAGCGAGGCCTTCAAGCAGAAAACCAGCGAGATCCTGTCCCACGCACACGCC
GCCCTGGATCAGCCACTGCCTACAACCCTGAAGAAGCAGGAGGAGAAGGAGATCCTGAAGTCTCAG
CTGGACAGCCTGCTGGGCCTGTACCACCTGCTGGACTGGTTTGCCGTGGATGAGTCCAACGAGGTGG
ACCCCGAGTTCTCTGCCCGGCTGACCGGCATCAAGCTGGAGATGGAGCCTTCTCTGAGCTTCTACAAC
AAGGCCAGAAATTATGCCACCAAGAAGCCCTACTCCGTGGAGAAGTTCAAGCTGAACTTTCAGATGC
CTACACTGGCCTCTGGCTGGGACGTGAATAAGGAGAAGAACAATGGCGCCATCCTGTTTGTGAAGAA
CGGCCTGTACTATCTGGGCATCATGCCAAAGCAGAAGGGCAGGTATAAGGCCCTGAGCTTCGAGCCC
ACAGAGAAAACCAGCGAGGGCTTTGATAAGATGTACTATGACTACTTCCCTGATGCCGCCAAGATGAT
CCCAAAGTGCAGCACCCAGCTGAAGGCCGTGACAGCCCACTTTCAGACCCACACAACCCCCATCCTG
CTGTCCAACAATTTCATCGAGCCTCTGGAGATCACAAAGGAGATCTACGACCTGAACAATCCTGAGA
AGGAGCCAAAGAAGTTTCAGACAGCCTACGCCAAGAAAACCGGCGACCAGAAGGGCTACAGAGAG
GCCCTGTGCAAGTGGATCGACTTCACAAGGGATTTTCTGTCCAAGTATACCAAGACAACCTCTATCGA
TCTGTCTAGCCTGCGGCCATCCTCTCAGTATAAGGACCTGGGCGAGTACTATGCCGAGCTGAATCCCCT
GCTGTACCACATCAGCTTCCAGAGAATCGCCGAGAAGGAGATCATGGATGCCGTGGAGACAGGCAAG
CTGTACCTGTTCCAGATCTATAACAAGGACTTTGCCAAGGGCCACCACGGCAAGCCTAATCTGCACAC
ACTGTATTGGACCGGCCTGTTTTCTCCAGAGAACCTGGCCAAGACAAGCATCAAGCTGAATGGCCAG
[0326] GCCGAGCTGTTCTACCGCCCTAAGTCCAGGATGAAGAGGATGGCACACCGGCTGGGAGAGAAGATGC
TGAACAAGAAGCTGAAGGATCAGAAAACCCCAATCCCCGACACCCTGTACCAGGAGCTGTACGACTA
TGTGAATCACAGACTGTCCCACGACCTGTCTGATGAGGCCAGGGCCCTGCTGCCCAACGTGATCACC
AAGGAGGTGTCTCACGAGATCATCAAGGATAGGCGCTTTACCAGCGACAAGTTCTTTTTCCACGTGCC
TATCACACTGAACTATCAGGCCGCCAATTCCCCATCTAAGTTCAACCAGAGGGTGAATGCCTACCTGA
AGGAGCACCCCGAGACACCTATCATCGGCATCgccCGGGGCGAGAGAAACCTGATCTATATCACAGTGA
TCGACTCCACCGGCAAGATCCTGGAGCAGCGGAGCCTGAACACCATCCAGCAGTTTGATTACCAGAA
GAAGCTGGACAACAGGGAGAAGGAGAGGGTGGCAGCAAGGCAGGCCTGGTCTGTGGTGGGCACAA
TCAAGGATCTGAAGCAGGGCTATCTGAGCCAGGTCATCCACGAGATCGTGGACCTGATGATCCACTAC
CAGGCCGTGGTGGTGCTGGAGAACCTGAATTTCGGCTTTAAGAGCAAGAGGACCGGCATCGCCGAGA
AGGCCGTGTACCAGCAGTTCGAGAAGATGCTGATCGATAAGCTGAATTGCCTGGTGCTGAAGGACTAT
CCAGCAGAGAAAGTGGGAGGCGTGCTGAACCCATACCAGCTGACAGACCAGTTCACCTCCTTTGCCA
AGATGGGCACCCAGTCTGGCTTCCTGTTTTACGTGCCTGCCCCATATACATCTAAGATCGATCCCCTGA
CCGGCTTCGTGGACCCCTTCGTGTGGAAAACCATCAAGAATCACGAGAGCCGCAAGCACTTCCTGGA
GGGCTTCGACTTTCTGCACTACGACGTGAAAACCGGCGACTTCATCCTGCACTTTAAGATGAACAGA
AATCTGTCCTTCCAGAGGGGCCTGCCCGGCTTTATGCCTGCATGGGATATCGTGTTCGAGAAGAACGA
GACACAGTTTGACGCCAAGGGCACCCCTTTCATCGCCGGCAAGAGAATCGTGCCAGTGATCGAGAAT
CACAGATTCACCGGCAGATACCGGGACCTGTATCCTGCCAACGAGCTGATCGCCCTGCTGGAGGAGA
AGGGCATCGTGTTCAGGGATGGCTCCAACATCCTGCCAAAGCTGCTGGAGAATGACGATTCTCACGC
CATCGACACGATGGTGGCCCTGATCCGCAGCGTGCTGCAGATGCGGAACTCCAATGCCGCCACAGGC
GAGGACTATATCAACAGCCCCGTGCGCGATCTGAATGGCGTGTGCTTCGACTCCCGGTTTCAGAACCC
AGAGTGGCCAATGGACGCCGATGCCAATGGCGCCTACCACATCGCCCTGAAGGGCCAGCTGCTGCTG
AATCACCTGAAGGAGAGCAAGGATCTGAAGCTGCAGAACGGCATCTCCAATCAGGACTGGCTGGCCT

[0327] ACATCCAGGAGCTGCGCAACaaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaagGGATCCTCTGGTG GTTCTGGAGGATCTGGTGGTTCTACTAATCTGTCAGATATTATTGAAAAGGAGACCGGTAAGCAACT GGTTATCCAGGAATCCATCCTCATGCTCCCAGAGGAGGTGGAAGAAGTCATTGGGAACAAGCC GGAAAGCGATATACTCGTGCACACCGCCTACGACGAGAGCACCGACGAGAATGTCATGCTTCT GACTAGCGACGCCCCTGAATACAAGCCTTGGGCTCTGGTCATACAGGATAGCAACGGTGAGAA CAAGATTAAGATGCTCTCTGGTGGTTCTCCCAAGAAGAAGAGGAAAGTCTAA (SEQ ID NO:438)

[0328] RTW1295:pCAG-rAPOBEC1-gs-XTEN-gs-hdLbCas12a(D832A)-NLS-gs-UGI-NLS (LbBE1.1)

[0329] 大鼠APOBEC1(rAPOBEC1)为粗体大写字母,插入的甘氨酸为点式下划线的,甘氨酸/丝氨酸接头为斜体,XTEN接头为小写字母斜体,人类密码子优化的毛螺旋菌科细菌ND2006 Cas12a(LbCas12a)为黑色,用于DNase失活用修饰的密码子(D832A)LbCas12a为双下划线的小写字母,具有沉默突变以移除NcoI位点的密码子为双下划线的大写字母,核质蛋白NLS为粗体小写字母,UGI为**虚下划线的粗体大写字母**,SV40 NLS为*双下划线的斜体*

[0330] ATGGGCAGCTCAGAGACTGGCCCAGTGGCTGTGGACCCCACATTGAGACGGCGGATCGAGCCC CATGAGTTTGAGGTATTCTTCGATCCGAGAGAGCTCCGCAAGGAGACCTGCCTGCTTTACGAAA TTAATTGGGGGGGCCGGCACTCCATTTGGCGACATACATCACAGAACACTAACAAGCACGTCG AAGTCAACTTCATCGAGAAGTTCACGACAGAAAGATATTTCTGTCCGAACACAAGGTGCAGCAT TACCTGGTTTCTCAGCTGGAGCCCATGCGGCGAATGTAGTAGGGCCATCACTGAATTCCTGTCA AGGTATCCCCACGTCACTCTGTTTATTTACATCGCAAGGCTGTACCACCACGCTGACCCCCGCA ATCGACAAGGCCTGCGGGATTTGATCTCTTCAGGTGTGACTATCCAAATTATGACTGAGCAGGA GTCAGGATACTGCTGGAGAAACTTTGTGAATTATAGCCCGAGTAATGAAGCCCACTGGCCTAGG TATCCCCATCTGTGGGTACGACTGTACGTTCTTGAACTGTACTGCATCATACTGGGCCTGCCTC CTTGTCTCAACATTCTGAGAAGGAAGCAGCCACAGCTGACATTCTTTACCATCGCTCTTCAGTC TTGTCATTACCAGCGACTGCCCCCACACATTCTCTGGGCCACCGGGTTGAAA*TCTGGTGGTTCTT CTGGTGGTTCTagcggcagcgagactcccgggacctcagagtccgccacacccgaaagtTCCGGAGGGAGTAGCGGCGGGTCTA* GCAAGCTGGAGAAGTTTACAAACTGCTACTCCCTGTCTAAGACCCTGAGGTTCAAGGCCATCCCTGT GGGCAAGACCCAGGAGAACATCGACAATAAGCGGCTGCTGGTGGAGGACGAGAAGAGAGCCGAGG ATTATAAGGGCGTGAAGAAGCTGCTGGATCGCTACTATCTGTCTTTTATCAACGACGTGCTGCACAGCA TCAAGCTGAAGAATCTGAACAATTACATCAGCCTGTTCCGGAAGAAAACCAGAACCGAGAAGGAGA ATAAGGAGCTGGAGAACCTGGAGATCAATCTGCGGAAGGAGATCGCCAAGGCCTTCAAGGGCAACG AGGGCTACAAGTCCCTGTTTAAGAAGGATATCATCGAGACAATCCTGCCAGAGTTCCTGGACGATAAG GACGAGATCGCCCTGGTGAACAGCTTCAATGGCTTTACCACAGCCTTCACCGGCTTCTTTGATAACAG AGAGAATATGTTTTCCGAGGAGGCCAAGAGCACATCCATCGCCTTCAGGTGTATCAACGAGAATCTGA CCCGCTACATCTCTAATATGGACATCTTCGAGAAGGTGGACGCCATCTTTGATAAGCACGAGGTGCAG GAGATCAAGGAGAAGATCCTGAACAGCGACTATGATGTGGAGGATTTCTTTGAGGGCGAGTTCTTTAA CTTTGTGCTGACACAGGAGGGCATCGACGTGTATAACGCCATCATCGGCGGCTTCGTGACCGAGAGC GGCGAGAAGATCAAGGGCCTGAACGAGTACATCAACCTGTATAATCAGAAAACCAAGCAGAAGCTGC CTAAGTTTAAGCCACTGTATAAGCAGGTGCTGAGCGATCGGGAGTCTCTGAGCTTCTACGGCGAGGGC TATACATCCGATGAGGAGGTGCTGGAGGTGTTTAGAAACACCCTGAACAAGAACAGCGAGATCTTCA GCTCCATCAAGAAGCTGGAGAAGCTGTTCAAGAATTTTGACGAGTACTCTAGCGCCGGCATCTTTGTG AAGAACGGCCCCCGCCATCAGCACAATCTCCAAGGATATCTTCGGCGAGTGGAACGTGATCCGGGACA AGTGGAATGCCGAGTATGACGATATCCACCTGAAGAAGAAGGCCGTGGTGACCGAGAAGTACGAGG ACGATCGGAGAAAGTCCTTCAAGAAGATCGGCTCCTTTTCTCTGGAGCAGCTGCAGGAGTACGCCGA CGCCGATCTGTCTGTGGTGGAGAAGCTGAAGGAGATCATCATCCAGAAGGTGGATGAGATCTACAAG

[0331] GTGTATGGCTCCTCTGAGAAGCTGTTCGACGCCGATTTTGTGCTGGAGAAGAGCCTGAAGAAGAACG
ACGCCGTGGTGGCCATCATGAAGGACCTGCTGGATTCTGTGAAGAGCTTCGAGAATTACATCAAGGC
CTTCTTTGGCGAGGGCAAGGAGACAAACAGGGACGAGTCCTTCTATGGCGATTTTGTGCTGGCCTAC
GACATCCTGCTGAAGGTGGACCACATCTACGATGCCATCCGCAATTATGTGACCCAGAAGCCCTACTC
TAAGGATAAGTTCAAGCTGTATTTTCAGAACCCTCAGTTCATGGGCGGCTGGGACAAGGATAAGGAG
ACAGACTATCGGGCCACCATCCTGAGATACGGCTCCAAGTACTATCTGGCCATCATGGATAAGAAGTA
CGCCAAGTGCCTGCAGAAGATCGACAAGGACGATGTGAACGGCAATTACGAGAAGATCAACTATAAG
CTGCTGCCCGGCCCTAATAAGATGCTGCCAAAGGTGTTCTTTTCTAAGAAGTGGATGGCCTACTATAAC
CCCAGCGAGGACATCCAGAAGATCTACAAGAATGGCACATTCAAGAAGGGCGATATGTTTAACCTGA
ATGACTGTCACAAGCTGATCGACTTCTTTAAGGATAGCATCTCCCGGTATCCAAAGTGGTCCAATGCCT
ACGATTTCAACTTTTCTGAGACAGAGAAGTATAAGGACATCGCCGGCTTTTACAGAGAGGTGGAGGA
GCAGGGCTATAAGGTGAGCTTCGAGTCTGCCAGCAAGAAGGAGGTGGATAAGCTGGTGGAGGAGGG
CAAGCTGTATATGTTCCAGATCTATAACAAGGACTTTTCCGATAAGTCTCACGGCACACCCAATCTGCA
CACCATGTACTTCAAGCTGCTGTTTGACGAGAACAATCACGGACAGATCAGGCTGAGCGGAGGAGCA
GAGCTGTTCATGAGGCGCGCCTCCCTGAAGAAGGAGGAGCTGGTGGTGCACCCAGCCAACTCCCCTA
TCGCCAACAAGAATCCAGATAATCCCAAGAAAACCACAACCCTGTCCTACGACGTGTATAAGGATAAG
AGGTTTTCTGAGGACCAGTACGAGCTGCACATCCCAATCGCCATCAATAAGTGCCCCAAGAACATCTT
CAAGATCAATACAGAGGTGCGCGTGCTGCTGAAGCACGACGATAACCCCTATGTGATCGGCATCgccAG
GGGCGAGCGCAATCTGCTGTATATCGTGGTGGTGGACGGCAAGGGCAACATCGTGGAGCAGTATTCC
CTGAACGAGATCATCAACAACTTCAACGGCATCAGGATCAAGACAGATTACCACTCTCTGCTGGACA
AGAAGGAGAAGGAGAGGTTCGAGGCCCGCCAGAACTGGACCTCCATCGAGAATATCAAGGAGCTGA
AGGCCGGCTATATCTCTCAGGTGGTGCACAAGATCTGCGAGCTGGTGGAGAAGTACGATGCCGTGATC
GCCCTGGAGGACCTGAACTCTGGCTTTAAGAATAGCCGCGTGAAGGTGGAGAAGCAGGTGTATCAGA
AGTTCGAGAAGATGCTGATCGATAAGCTGAACTACATGGTGGACAAGAAGTCTAATCCTTGTGCAACA
GGCGGCGCCCTGAAGGGCTATCAGATCACCAATAAGTTCGAGAGCTTTAAGTCCATGTCTACCCAGAA
CGGCTTCATCTTTTACATCCCTGCCTGGCTGACATCCAAGATCGATCCATCTACCGGCTTTGTGAACCT
GCTGAAAACCAAGTATACCAGCATCGCCGATTCCAAGAAGTTCATCAGCTCCTTTGACAGGATCATGT
ACGTGCCCGAGGAGGATCTGTTCGAGTTTGCCCTGGACTATAAGAACTTCTCTCGCACAGACGCCGAT
TACATCAAGAAGTGGAAGCTGTACTCCTACGGCAACCGGATCAGAATCTTCCGGAATCCTAAGAAGA
ACAACGTGTTCGACTGGGAGGAGGTGTGCCTGACCAGCGCCTATAAGGAGCTGTTCAACAAGTACGG
CATCAATTATCAGCAGGGCGATATCAGAGCCCTGCTGTGCGAGCAGTCCGACAAGGCCTTCTACTCTA
GCTTTATGGCCCTGATGAGCCTGATGCTGCAGATGCGGAACAGCATCACAGGCCGCACCGACGTGGAT
TTTCTGATCAGCCCTGTGAAGAACTCCGACGGCATCTTCTACGATAGCCGGAACTATGAGGCCCAGGA
GAATGCCATCCTGCCAAAGAACGCCGACGCCAATGGCGCCTATAACATCGCCAGAAAGGTGCTGTGG
GCCATCGGCCAGTTCAAGAAGGCCGAGGACGAGAAGCTGGATAAGGTGAAGATCGCCATCTCTAACA
AGGAGTGGCTGGAGTACGCCCAGACCAGCGTGAAGCAC**aaaaggccggcggccacgaaaaaggccggccaggcaaaaaa**
agaaaaag*GGATCCTCTGGTGGTTCTGGAGGATCTGGTGGTTCT***ACTAATCTGTCAGATATTATTGAAAAG**
GAGACCGGTAAGCAACTGGTTATCCAGGAATCCATCCTCATGCTCCCAGAGGAGGTGGAAGAA
GTCATTGGGAACAAGCCGGAAAGCGATATACTCGTGCACACCGCCTACGACGAGAGCACCGAC
GAGAATGTCATGCTTCTGACTAGCGACGCCCCTGAATACAAGCCTTGGGCTCTGGTCATACAGG
ATAGCAACGGTGAGAACAAGATTAAGATGCTC*TCTGGTGGTTCTCCCAAGAAGAAGAGGAAAGTC*TA
A (SEQ ID NO:439)

[0332] RTW1352:

[0333] pCAG-rAPOBEC1-gs-XTEN-gs-hdeAsCas12a(E174R/S542R/K548R/D908A)-NLS-gs-UGI-NLS(eAsBE1.1)大鼠APOBEC1(rAPOBEC1)为粗体大写字母,插入的甘氨酸为点式下划线的,甘氨酸/丝氨酸接头为斜体,XTEN接头为小写字母斜体,人类密码子优化的氨基酸球菌种BV3L6 Cas12a(AsCas12a)为黑色,DNase失活用修饰的密码子(D908A)

eAsCas12a(E174R/S542R/K548R)为双下划线的小写字母,具有沉默突变以移除NcoI位点的密码子为双下划线的大写字母,核质蛋白NLS为粗体小写字母,UGI为**虚下划线的粗 体大写字母**,SV40NLS为*双下划线的斜体*

[0334]

ATGGGCAGCTCAGAGACTGGCCCAGTGGCTGTGGACCCCACATTGAGACGGCGGATCGAGCCC
CATGAGTTTGAGGTATTCTTCGATCCGAGAGAGCTCCGCAAGGAGACCTGCCTGCTTTACGAAA
TTAATTGGGGGGGCCGGCACTCCATTTGGCGACATACATCACAGAACACTAACAAGCACGTCG
AAGTCAACTTCATCGAGAAGTTCACGACAGAAAGATATTTCTGTCCGAACACAAGGTGCAGCAT
TACCTGGTTTCTCAGCTGGAGCCCATGCGGCGAATGTAGTAGGGCCATCACTGAATTCCTGTCA
AGGTATCCCCACGTCACTCTGTTTATTTACATCGCAAGGCTGTACCACCACGCTGACCCCCGCA
ATCGACAAGGCCTGCGGGATTTGATCTCTTCAGGTGTGACTATCCAAATTATGACTGAGCAGGA
GTCAGGATACTGCTGGAGAAACTTTGTGAATTATAGCCCGAGTAATGAAGCCCACTGGCCTAGG
TATCCCCATCTGTGGGTACGACTGTACGTTCTTGAACTGTACTGCATCATACTGGGCCTGCCTC
CTTGTCTCAACATTCTGAGAAGGAAGCAGCCACAGCTGACATTCTTTACCATCGCTCTTCAGTC
TTGTCATTACCAGCGACTGCCCCCACACATTCTCTGGGCCACCGGGTTGAAA*TCTGGTGGTTCTT*
CTGGTGGTTCTagcggcagcgagactcccgggacctcagagtccgccacacccgaaagtTCCGGAGGGAGTAGCGGCGGGTCTA
CACAGTTCGAGGGCTTTACCAACCTGTATCAGGTGAGCAAGACACTGCGGTTTGAGCTGATCCCACA
GGGCAAGACCCTGAAGCACATCCAGGAGCAGGGCTTCATCGAGGAGGACAAGGCCCGCAATGATCA
CTACAAGGAGCTGAAGCCCATCATCGATCGGATCTACAAGACCTATGCCGACCAGTGCCTGCAGCTGG
TGCAGCTGGATTGGGAGAACCTGAGCGCCGCCATCGACTCCTATAGAAAGGAGAAAACCGAGGAGA
CAAGGAACGCCCTGATCGAGGAGCAGGCCACATATCGCAATGCCATCCACGACTACTTCATCGGCCGG
ACAGACAACCTGACCGATGCCATCAATAAGAGACACGCCGAGATCTACAAGGGCCTGTTCAAGGCCG
AGCTGTTTAATGGCAAGGTGCTGAAGCAGCTGGGCACCGTGACCACAACCGAGCACGAGAACGCCC
TGCTGCGGAGCTTCGACAAGTTTACAACCTACTTCTCCGGCTTTTATagaAACAGGAAGAACGTGTTCA
GCGCCGAGGATATCAGCACAGCCATCCCACACCGCATCGTGCAGGACAACTTCCCCAAGTTTAAGGA
GAATTGTCACATCTTCACACGCCTGATCACCGCCGTGCCCAGCCTGCGGGAGCACTTTGAGAACGTG
AAGAAGGCCATCGGCATCTTCGTGAGCACCTCCATCGAGGAGGTGTTTTCCTTCCCTTTTTATAACCA
GCTGCTGACACAGACCCAGATCGACCTGTATAACCAGCTGCTGGGAGGAATCTCTCGGGAGGCAGGC
ACCGAGAAGATCAAGGGCCTGAACGAGGTGCTGAATCTGGCCATCCAGAAGAATGATGAGACAGCC
CACATCATCGCCTCCCTGCCACACAGATTCATCCCCCTGTTTAAGCAGATCCTGTCCGATAGGAACACC
CTGTCTTTCATCCTGGAGGAGTTTAAGAGCGACGAGGAAGTGATCCAGTCCTTCTGCAAGTACAAGA
CACTGCTGAGAAACGAGAACGTGCTGGAGACAGCCGAGGCCCTGTTTAACGAGCTGAACAGCATCG
ACCTGACACACATCTTCATCAGCCACAAGAAGCTGGAGACAATCAGCAGCGCCCTGTGCGACCACTG
GGATACACTGAGGAATGCCCTGTATGAGCGGAGAATCTCCGAGCTGACAGGCAAGATCACCAAGTCT
GCCAAGGAGAAGGTGCAGCGCAGCCTGAAGCACGAGGATATCAACCTGCAGGAGATCATCTCTGCCG
CAGGCAAGGAGCTGAGCGAGGCCTTCAAGCAGAAAACCAGCGAGATCCTGTCCCACGCACACGCCG
CCCTGGATCAGCCACTGCCTACAACCCTGAAGAAGCAGGAGGAGAAGGAGATCCTGAAGTCTCAGC
TGGACAGCCTGCTGGGCCTGTACCACCTGCTGGACTGGTTTGCCGTGGATGAGTCCAACGAGGTGGA
CCCCGAGTTCTCTGCCCGGCTGACCGGCATCAAGCTGGAGATGGAGCCTTCTCTGAGCTTCTACAACA
AGGCCAGAAATTATGCCACCAAGAAGCCCTACTCCGTGGAGAAGTTCAAGCTGAACTTTCAGATGCC
TACACTGGCCagaGGCTGGGACGTGAATagaGAGAAGAACAATGGCGCCATCCTGTTTGTGAAGAACGG
CCTGTACTATCTGGGCATCATGCCAAAGCAGAAGGGCAGGTATAAGGCCCTGAGCTTCGAGCCCACA
GAGAAAACCAGCGAGGGCTTTGATAAGATGTACTATGACTACTTCCCTGATGCCGCCAAGATGATCCC

AAAGTGCAGCACCCAGCTGAAGGCCGTGACAGCCCACTTTCAGACCCACACAACCCCCATCCTGCTG
TCCAACAATTTCATCGAGCCTCTGGAGATCACAAAGGAGATCTACGACCTGAACAATCCTGAGAAGG
AGCCAAAGAAGTTTCAGACAGCCTACGCCAAGAAAACCGGCGACCAGAAGGGCTACAGAGAGGCCC
TGTGCAAGTGGATCGACTTCACAAGGGATTTTCTGTCCAAGTATACCAAGACAACCTCTATCGATCTG
TCTAGCCTGCGGCCATCCTCTCAGTATAAGGACCTGGGCGAGTACTATGCCGAGCTGAATCCCCTGCT
GTACCACATCAGCTTCCAGAGAATCGCCGAGAAGGAGATCATGGATGCCGTGGAGACAGGCAAGCTG
TACCTGTTCCAGATCTATAACAAGGACTTTGCCAAGGGCCACCACGGCAAGCCTAATCTGCACACACT
GTATTGGACCGGCCTGTTTTCTCCAGAGAACCTGGCCAAGACAAGCATCAAGCTGAATGGCCAGGCC
GAGCTGTTCTACCGCCCTAAGTCCAGGATGAAGAGGATGGCACACCGGCTGGGAGAGAAGATGCTGA
ACAAGAAGCTGAAGGATCAGAAAACCCCAATCCCCGACACCCTGTACCAGGAGCTGTACGACTATGT
GAATCACAGACTGTCCCACGACCTGTCTGATGAGGCCAGGGCCCTGCTGCCCAACGTGATCACCAAG
GAGGTGTCTCACGAGATCATCAAGGATAGGCGCTTTACCAGCGACAAGTTCTTTTTCCACGTGCCTAT
CACACTGAACTATCAGGCCGCCAATTCCCCATCTAAGTTCAACCAGAGGGTGAATGCCTACCTGAAGG
AGCACCCCGAGACACCTATCATCGGCATCgccCGGGGCGAGAGAAACCTGATCTATATCACAGTGATCG
ACTCCACCGGCAAGATCCTGGAGCAGCGGAGCCTGAACACCATCCAGCAGTTTGATTACCAGAAGAA
GCTGGACAACAGGGAGAAGGAGAGGGTGGCAGCAAGGCAGGCCTGGTCTGTGGTGGGCACAATCA
AGGATCTGAAGCAGGGCTATCTGAGCCAGGTCATCCACGAGATCGTGGACCTGATGATCCACTACCAG
GCCGTGGTGGTGCTGGAGAACCTGAATTTCGGCTTTAAGAGCAAGAGGACCGGCATCGCCGAGAAG
[0335] GCCGTGTACCAGCAGTTCGAGAAGATGCTGATCGATAAGCTGAATTGCCTGGTGCTGAAGGACTATCC
AGCAGAGAAAGTGGGAGGCGTGCTGAACCCATACCAGCTGACAGACCAGTTCACCTCCTTTGCCAA
GATGGGCACCCAGTCTGGCTTCCTGTTTTACGTGCCTGCCCCATATACATCTAAGATCGATCCCCTGAC
CGGCTTCGTGGACCCCTTCGTGTGGAAAACCATCAAGAATCACGAGAGCCGCAAGCACTTCCTGGAG
GGCTTCGACTTTCTGCACTACGACGTGAAAACCGGCGACTTCATCCTGCACTTTAAGATGAACAGAAA
TCTGTCCTTCCAGAGGGGCCTGCCCGGCTTTATGCCTGCATGGGATATCGTGTTCGAGAAGAACGAGA
CACAGTTTGACGCCAAGGGCACCCCTTTCATCGCCGGCAAGAGAATCGTGCCAGTGATCGAGAATCA
CAGATTCACCGGCAGATACCGGGACCTGTATCCTGCCAACGAGCTGATCGCCCTGCTGGAGGAGAAG
GGCATCGTGTTCAGGGATGGCTCCAACATCCTGCCAAAGCTGCTGGAGAATGACGATTCTCACGCCAT
CGACACGATGGTGGCCCTGATCCGCAGCGTGCTGCAGATGCGGAACTCCAATGCCGCCACAGGCGAG
GACTATATCAACAGCCCCGTGCGCGATCTGAATGGCGTGTGCTTCGACTCCCGGTTTCAGAACCCAGA
GTGGCCAATGGACGCCGATGCCAATGGCGCCTACCACATCGCCCTGAAGGGCCAGCTGCTGCTGAAT
CACCTGAAGGAGAGCAAGGATCTGAAGCTGCAGAACGGCATCTCCAATCAGGACTGGCTGGCCTACA
TCCAGGAGCTGCGCAAC**aaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaag***GGATCCTCTGGTGGTT*
*CTGGAGGATCTGGTGGTTCT***ACTAATCTGTCAGATATTATTGAAAAGGAGACCGGTAAGCAACTGG**
TTATCCAGGAATCCATCCTCATGCTCCCAGAGGAGGTGGAAGAAGTCATTGGGAACAAGCCGG
AAAGCGATATACTCGTGCACACCGCCTACGACGAGAGCACCGACGAGAATGTCATGCTTCTGA
CTAGCGACGCCCCTGAATACAAGCCTTGGGCTCTGGTCATACAGGATAGCAACGGTGAGAACA
AGATTAAGATGCTC*TCTGGTGGTTCTCCCAAGAAGAAGAGGAAAGTC*TAA (SEQ ID NO:440)

[0336] RTW1348:

[0337] pCAG-2xNLS-rAPOBEC1-gs-XTEN-gs-hdeAsCas12a(E174R/S542R/K548R/D908A)-NLS-gs-UGI-NLS(eAsBE1.2)

[0338] 大鼠APOBEC1(rAPOBEC1)为粗体大写字母,插入的甘氨酸为点式下划线的,SV40NLS为*双下划线斜体 的*,甘氨酸/丝氨酸接头为斜体,XTEN接头为斜体小写字母,人类密码子优化的氨基酸球菌种BV3L6Cas12a(AsCas12a)为黑色,失活用修饰的密码子(D908A)eAsCas12a(E174R/S542R/K548R)为双下划线 的小写字母,具有沉默突变以移除NcoI位点的密码子为双下划线的大写字母,核质蛋白NLS为粗体小写字母,UGI为**虚下划线的粗体大写字母**,

ATGGGC*CCAAAGAAAAAGAGGAAAGTC*GGCAGTGGA*CCTAAAAAGAAACGAAAGGTT*GGGTCAGGT**AGCT**
CAGAGACTGGCCCAGTGGCTGTGGACCCCACATTGAGACGGCGGATCGAGCCCCATGAGTTTG
AGGTATTCTTCGATCCGAGAGAGCTCCGCAAGGAGACCTGCCTGCTTTACGAAATTAATTGGGG
GGGCCGGCACTCCATTTGGCGACATACATCACAGAACACTAACAAGCACGTCGAAGTCAACTT
CATCGAGAAGTTCACGACAGAAAGATATTTCTGTCCGAACACAAGGTGCAGCATTACCTGGTTT
CTCAGCTGGAGCCCATGCGGCGAATGTAGTAGGGCCATCACTGAATTCCTGTCAAGGTATCCCC
ACGTCACTCTGTTTATTTACATCGCAAGGCTGTACCACCACGCTGACCCCCGCAATCGACAAGG
CCTGCGGGATTTGATCTCTTCAGGTGTGACTATCCAAATTATGACTGAGCAGGAGTCAGGATAC
TGCTGGAGAAACTTTGTGAATTATAGCCCGAGTAATGAAGCCCACTGGCCTAGGTATCCCCATC
TGTGGGTACGACTGTACGTTCTTGAACTGTACTGCATCATACTGGGCCTGCCTCCTTGTCTCAA
CATTCTGAGAAGGAAGCAGCCACAGCTGACATTCTTTACCATCGCTCTTCAGTCTTGTCATTAC
CAGCGACTGCCCCCACACATTCTCTGGGCCACCGGGTTGAAA*TCTGGTGGTTCTTCTGGTGGTTCT*
*agcggcagcgagactcccgggacctcagagtccgccacacccgaaagt**TCCGGAGGGAGTAGCGGCGGGTCT***ACACAGTTCGA
GGGCTTTACCAACCTGTATCAGGTGAGCAAGACACTGCGGTTTGAGCTGATCCCACAGGGCAAGACC
CTGAAGCACATCCAGGAGCAGGGCTTCATCGAGGAGGACAAGGCCCGCAATGATCACTACAAGGAG
CTGAAGCCCATCATCGATCGGATCTACAAGACCTATGCCGACCAGTGCCTGCAGCTGGTGCAGCTGGA
TTGGGAGAACCTGAGCGCCGCCATCGACTCCTATAGAAAGGAGAAAACCGAGGAGACAAGGAACGC
CCTGATCGAGGAGCAGGCCACATATCGCAATGCCATCCACGACTACTTCATCGGCCGGACAGACAACC
TGACCGATGCCATCAATAAGAGACACGCCGAGATCTACAAGGGCCTGTTCAAGGCCGAGCTGTTTAAT
GGCAAGGTGCTGAAGCAGCTGGGCACCGTGACCACAACCGAGCACGAGAACGCCCTGCTGCGGAGC
TTCGACAAGTTTACAACCTACTTCTCCGGCTTTTATAGAAACAGGAAGAACGTGTTCAGCGCCGAGGA
[0339] TATCAGCACAGCCATCCCACACCGCATCGTGCAGGACAACTTCCCCAAGTTTAAGGAGAATTGTCACA
TCTTCACACGCCTGATCACCGCCGTGCCCAGCCTGCGGGAGCACTTTGAGAACGTGAAGAAGGCCAT
CGGCATCTTCGTGAGCACCTCCATCGAGGAGGTGTTTTCCTTCCCTTTTTATAACCAGCTGCTGACACA
GACCCAGATCGACCTGTATAACCAGCTGCTGGGAGGAATCTCTCGGGAGGCAGGCACCGAGAAGATC
AAGGGCCTGAACGAGGTGCTGAATCTGGCCATCCAGAAGAATGATGAGACAGCCCACATCATCGCCT
CCCTGCCACACAGATTCATCCCCCTGTTTAAGCAGATCCTGTCCGATAGGAACACCCTGTCTTTCATCC
TGGAGGAGTTTAAGAGCGACGAGGAAGTGATCCAGTCCTTCTGCAAGTACAAGACACTGCTGAGAA
ACGAGAACGTGCTGGAGACAGCCGAGGCCCTGTTTAACGAGCTGAACAGCATCGACCTGACACACA
TCTTCATCAGCCACAAGAAGCTGGAGACAATCAGCAGCGCCCTGTGCGACCACTGGGATACACTGAG
GAATGCCCTGTATGAGCGGAGAATCTCCGAGCTGACAGGCAAGATCACCAAGTCTGCCAAGGAGAAG
GTGCAGCGCAGCCTGAAGCACGAGGATATCAACCTGCAGGAGATCATCTCTGCCGCAGGCAAGGAGC
TGAGCGAGGCCTTCAAGCAGAAAACCAGCGAGATCCTGTCCCACGCACACGCCGCCCTGGATCAGCC
ACTGCCTACAACCCTGAAGAAGCAGGAGGAGAAGGAGATCCTGAAGTCTCAGCTGGACAGCCTGCT
GGGCCTGTACCACCTGCTGGACTGGTTTGCCGTGGATGAGTCCAACGAGGTGGACCCCGAGTTCTCT
GCCCGGCTGACCGGCATCAAGCTGGAGATGGAGCCTTCTCTGAGCTTCTACAACAAGGCCAGAAATT
ATGCCACCAAGAAGCCCTACTCCGTGGAGAAGTTCAAGCTGAACTTTCAGATGCCTACACTGGCC**aga**
GGCTGGGACGTGAAT**aga**GAGAAGAACAATGGCGCCATCCTGTTTGTGAAGAACGGCCTGTACTATCT
GGGCATCATGCCAAAGCAGAAGGGCAGGTATAAGGCCCTGAGCTTCGAGCCCACAGAGAAAACCAG
CGAGGGCTTTGATAAGATGTACTATGACTACTTCCCTGATGCCGCCAAGATGATCCCAAAGTGCAGCA
CCCAGCTGAAGGCCGTGACAGCCCACTTTCAGACCCACACAACCCCCATCCTGCTGTCCAACAATTT
CATCGAGCCTCTGGAGATCACAAAGGAGATCTACGACCTGAACAATCCTGAGAAGGAGCCAAAGAA
GTTTCAGACAGCCTACGCCAAGAAAACCGGCGACCAGAAGGGCTACAGAGAGGCCCTGTGCAAGTG

[0340] GATCGACTTCACAAGGGATTTTCTGTCCAAGTATACCAAGACAACCTCTATCGATCTGTCTAGCCTGCG
GCCATCCTCTCAGTATAAGGACCTGGGCGAGTACTATGCCGAGCTGAATCCCCTGCTGTACCACATCA
GCTTCCAGAGAATCGCCGAGAAGGAGATCATGGATGCCGTGGAGACAGGCAAGCTGTACCTGTTCCA
GATCTATAACAAGGACTTTGCCAAGGGCCACCACGGCAAGCCTAATCTGCACACACTGTATTGGACCG
GCCTGTTTTCTCCAGAGAACCTGGCCAAGACAAGCATCAAGCTGAATGGCCAGGCCGAGCTGTTCTA
CCGCCCTAAGTCCAGGATGAAGAGGATGGCACACCGGCTGGGAGAGAAGATGCTGAACAAGAAGCT
GAAGGATCAGAAAACCCCAATCCCCGACACCCTGTACCAGGAGCTGTACGACTATGTGAATCACAGA
CTGTCCCACGACCTGTCTGATGAGGCCAGGGCCCTGCTGCCCAACGTGATCACCAAGGAGGTGTCTC
ACGAGATCATCAAGGATAGGCGCTTTACCAGCGACAAGTTCTTTTTCCACGTGCCTATCACACTGAAC
TATCAGGCCGCCAATTCCCCATCTAAGTTCAACCAGAGGGTGAATGCCTACCTGAAGGAGCACCCCGA
GACACCTATCATCGGCATCgccCGGGGCGAGAGAAACCTGATCTATATCACAGTGATCGACTCCACCGG
CAAGATCCTGGAGCAGCGGAGCCTGAACACCATCCAGCAGTTTGATTACCAGAAGAAGCTGGACAAC
AGGGAGAAGGAGAGGGTGGCAGCAAGGCAGGCCTGGTCTGTGGTGGGCACAATCAAGGATCTGAAG
CAGGGCTATCTGAGCCAGGTCATCCACGAGATCGTGGACCTGATGATCCACTACCAGGCCGTGGTGGT
GCTGGAGAACCTGAATTTCGGCTTTAAGAGCAAGAGGACCGGCATCGCCGAGAAGGCCGTGTACCAG
CAGTTCGAGAAGATGCTGATCGATAAGCTGAATTGCCTGGTGCTGAAGGACTATCCAGCAGAGAAAG
TGGGAGGCGTGCTGAACCCATACCAGCTGACAGACCAGTTCACCTCCTTTGCCAAGATGGGCACCCA
GTCTGGCTTCCTGTTTTACGTGCCTGCCCCATATACATCTAAGATCGATCCCCTGACCGGCTTCGTGGA
CCCCTTCGTGTGGAAAACCATCAAGAATCACGAGAGCCGCAAGCACTTCCTGGAGGGCTTCGACTTT
CTGCACTACGACGTGAAAACCGGCGACTTCATCCTGCACTTTAAGATGAACAGAAATCTGTCCTTCCA
GAGGGGCCTGCCCGGCTTTATGCCTGCATGGGATATCGTGTTCGAGAAGAACGAGACACAGTTTGAC
GCCAAGGGCACCCCTTTCATCGCCGGCAAGAGAATCGTGCCAGTGATCGAGAATCACAGATTCACCG
GCAGATACCGGGACCTGTATCCTGCCAACGAGCTGATCGCCCTGCTGGAGGAGAAGGGCATCGTGTT
CAGGGATGGCTCCAACATCCTGCCAAAGCTGCTGGAGAATGACGATTCTCACGCCATCGACACGATG
GTGGCCCTGATCCGCAGCGTGCTGCAGATGCGGAACTCCAATGCCGCCACAGGCGAGGACTATATCA
ACAGCCCCGTGCGCGATCTGAATGGCGTGTGCTTCGACTCCCGGTTTCAGAACCCAGAGTGGCCAAT
GGACGCCGATGCCAATGGCGCCTACCACATCGCCCTGAAGGGCCAGCTGCTGCTGAATCACCTGAAG
GAGAGCAAGGATCTGAAGCTGCAGAACGGCATCTCCAATCAGGACTGGCTGGCCTACATCCAGGAGC
TGCGCAAC**aaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaag***GGATCCTCTGGTGGTTCTGGAGGATC*
*TGGTGGTTCT***ACTAATCTGTCAGATATTATTGAAAAGGAGACCGGTAAGCAACTGGTTATCCAGG**
AATCCATCCTCATGCTCCCAGAGGAGGTGGAAGAAGTCATTGGGAACAAGCCGGAAAGCGATA
TACTCGTGCACACCGCCTACGACGAGAGCACCGACGAGAATGTCATGCTTCTGACTAGCGACG
CCCCTGAATACAAGCCTTGGGCTCTGGTCATACAGGATAGCAACGGTGAGAACAAGATTAAGAT
GCTC*TCTGGTGGTTCTCCCAAGAAGAAGAGGAAAGTC*TAA (SEQ ID NO:441)

[0341] RTW1296:

[0342] pCAG-rAPOBEC1-gs-XTEN-gs-hdeAsCas12a(E174R/S542R/K548R/D908A)-gs-UGI-NLS(eAsBE1.3)

[0343] Rat APOBEC1(rAPOBEC1)为粗体大写字母,插入的甘氨酸为点式下划线的,甘氨酸/丝氨酸接头为斜体,XTEN接头为斜体小写字母,人类密码子优化的氨基酸球菌种BV3L6 Cas12a(AsCas12a)为黑色,DNase失活用修饰的密码子(D908A)eAsCas12a(E174R/S542R/K548R)为双下划线的小写字母,具有沉默突变以移除NcoI位点的密码子为粗体下划线黑色的,UGI为**虚下划线的粗体大写字母**,SV40NLS为*双下划 线的斜体*

[0344] **ATGGGCAGCTCAGAGACTGGCCCAGTGGCTGTGGACCCCACATTGAGACGGCGGATCGAGCCC**
CATGAGTTTGAGGTATTCTTCGATCCGAGAGAGCTCCGCAAGGAGACCTGCCTGCTTTACGAAA

TTAATTGGGGGGGCCGGCACTCCATTTGGCGACATACATCACAGAACACTAACAAGCACGTCG
AAGTCAACTTCATCGAGAAGTTCACGACAGAAAGATATTTCTGTCCGAACACAAGGTGCAGCAT
TACCTGGTTTCTCAGCTGGAGCCCATGCGGCGAATGTAGTAGGGCCATCACTGAATTCCTGTCA
AGGTATCCCCACGTCACTCTGTTTATTTACATCGCAAGGCTGTACCACCACGCTGACCCCCGCA
ATCGACAAGGCCTGCGGGATTTGATCTCTTCAGGTGTGACTATCCAAATTATGACTGAGCAGGA
GTCAGGATACTGCTGGAGAAACTTTGTGAATTATAGCCCGAGTAATGAAGCCCACTGGCCTAGG
TATCCCCATCTGTGGGTACGACTGTACGTTCTTGAACTGTACTGCATCATACTGGGCCTGCCTC
CTTGTCTCAACATTCTGAGAAGGAAGCAGCCACAGCTGACATTCTTTACCATCGCTCTTCAGTC
TTGTCATTACCAGCGACTGCCCCCACACATTCTCTGGGCCACCGGGTTGAAA*TCTGGTGGTTCTT*
*CTGGTGGTTCT*AGCGGCAGCGAGACTCCCGGGACCTCAGAGTCCGCCACACCCGAAAGT*TCCGGAGG*
*GAGTAGCGGCGGGTCT*ACACAGTTCGAGGGCTTTACCAACCTGTATCAGGTGAGCAAGACACTGCGG
TTTGAGCTGATCCCACAGGGCAAGACCCTGAAGCACATCCAGGAGCAGGGCTTCATCGAGGAGGAC
AAGGCCCGCAATGATCACTACAAGGAGCTGAAGCCCATCATCGATCGGATCTACAAGACCTATGCCGA
CCAGTGCCTGCAGCTGGTGCAGCTGGATTGGGAGAACCTGAGCGCCGCCATCGACTCCTATAGAAAG
GAGAAAACCGAGGAGACAAGGAACGCCCTGATCGAGGAGCAGGCCACATATCGCAATGCCATCCAC
GACTACTTCATCGGCCGGACAGACAACCTGACCGATGCCATCAATAAGAGACACGCCGAGATCTACA
AGGGCCTGTTCAAGGCCGAGCTGTTTAATGGCAAGGTGCTGAAGCAGCTGGGCACCGTGACCACAA
CCGAGCACGAGAACGCCCTGCTGCGGAGCTTCGACAAGTTTACAACCTACTTCTCCGGCTTTTATagaA
ACAGGAAGAACGTGTTCAGCGCCGAGGATATCAGCACAGCCATCCCACACCGCATCGTGCAGGACAA
CTTCCCCAAGTTTAAGGAGAATTGTCACATCTTCACACGCCTGATCACCGCCGTGCCCAGCCTGCGGG
AGCACTTTGAGAACGTGAAGAAGGCCATCGGCATCTTCGTGAGCACCTCCATCGAGGAGGTGTTTTC
CTTCCCTTTTTATAACCAGCTGCTGACACAGACCCAGATCGACCTGTATAACCAGCTGCTGGGAGGAA
[0345] TCTCTCGGGAGGCAGGCACCGAGAAGATCAAGGGCCTGAACGAGGTGCTGAATCTGGCCATCCAGA
AGAATGATGAGACAGCCCACATCATCGCCTCCCTGCCACACAGATTCATCCCCCTGTTTAAGCAGATC
CTGTCCGATAGGAACACCCTGTCTTTCATCCTGGAGGAGTTTAAGAGCGACGAGGAAGTGATCCAGT
CCTTCTGCAAGTACAAGACACTGCTGAGAAACGAGAACGTGCTGGAGACAGCCGAGGCCCTGTTTA
ACGAGCTGAACAGCATCGACCTGACACACATCTTCATCAGCCACAAGAAGCTGGAGACAATCAGCAG
CGCCCTGTGCGACCACTGGGATACACTGAGGAATGCCCTGTATGAGCGGAGAATCTCCGAGCTGACA
GGCAAGATCACCAAGTCTGCCAAGGAGAAGGTGCAGCGCAGCCTGAAGCACGAGGATATCAACCTG
CAGGAGATCATCTCTGCCGCAGGCAAGGAGCTGAGCGAGGCCTTCAAGCAGAAAACCAGCGAGATC
CTGTCCCACGCACACGCCGCCCTGGATCAGCCACTGCCTACAACCCTGAAGAAGCAGGAGGAGAAG
GAGATCCTGAAGTCTCAGCTGGACAGCCTGCTGGGCCTGTACCACCTGCTGGACTGGTTTGCCGTGG
ATGAGTCCAACGAGGTGGACCCCGAGTTCTCTGCCCGGCTGACCGGCATCAAGCTGGAGATGGAGCC
TTCTCTGAGCTTCTACAACAAGGCCAGAAATTATGCCACCAAGAAGCCCTACTCCGTGGAGAAGTTC
AAGCTGAACTTTCAGATGCCTACACTGGCCagaGGCTGGGACGTGAATagaGAGAAGAACAATGGCGCC
ATCCTGTTTGTGAAGAACGGCCTGTACTATCTGGGCATCATGCCAAAGCAGAAGGGCAGGTATAAGGC
CCTGAGCTTCGAGCCCACAGAGAAAACCAGCGAGGGCTTTGATAAGATGTACTATGACTACTTCCCTG
ATGCCGCCAAGATGATCCCAAAGTGCAGCACCCAGCTGAAGGCCGTGACAGCCCACTTTCAGACCCA
CACAACCCCCATCCTGCTGTCCAACAATTTCATCGAGCCTCTGGAGATCACAAAGGAGATCTACGACC
TGAACAATCCTGAGAAGGAGCCAAAGAAGTTTCAGACAGCCTACGCCAAGAAAACCGGCGACCAGA
AGGGCTACAGAGAGGCCCTGTGCAAGTGGATCGACTTCACAAGGGATTTTCTGTCCAAGTATACCAA
GACAACCTCTATCGATCTGTCTAGCCTGCGGCCATCCTCTCAGTATAAGGACCTGGGCGAGTACTATGC
CGAGCTGAATCCCCTGCTGTACCACATCAGCTTCCAGAGAATCGCCGAGAAGGAGATCATGGATGCC
GTGGAGACAGGCAAGCTGTACCTGTTCCAGATCTATAACAAGGACTTTGCCAAGGGCCACCACGGCA
AGCCTAATCTGCACACACTGTATTGGACCGGCCTGTTTTCTCCAGAGAACCTGGCCAAGACAAGCATC

AAGCTGAATGGCCAGGCCGAGCTGTTCTACCGCCCTAAGTCCAGGATGAAGAGGATGGCACACCGGC
TGGGAGAGAAGATGCTGAACAAGAAGCTGAAGGATCAGAAAACCCCAATCCCCGACACCCTGTACC
AGGAGCTGTACGACTATGTGAATCACAGACTGTCCCACGACCTGTCTGATGAGGCCAGGGCCCTGCT
GCCCAACGTGATCACCAAGGAGGTGTCTCACGAGATCATCAAGGATAGGCGCTTTACCAGCGACAAG
TTCTTTTTCCACGTGCCTATCACACTGAACTATCAGGCCGCCAATTCCCCATCTAAGTTCAACCAGAGG
GTGAATGCCTACCTGAAGGAGCACCCCGAGACACCTATCATCGGCATCgccCGGGGCGAGAGAAACCT
GATCTATATCACAGTGATCGACTCCACCGGCAAGATCCTGGAGCAGCGGAGCCTGAACACCATCCAGC
AGTTTGATTACCAGAAGAAGCTGGACAACAGGGAGAAGGAGAGGGTGGCAGCAAGGCAGGCCTGGT
CTGTGGTGGGCACAATCAAGGATCTGAAGCAGGGCTATCTGAGCCAGGTCATCCACGAGATCGTGGA
CCTGATGATCCACTACCAGGCCGTGGTGGTGCTGGAGAACCTGAATTTCGGCTTTAAGAGCAAGAGG
ACCGGCATCGCCGAGAAGGCCGTGTACCAGCAGTTCGAGAAGATGCTGATCGATAAGCTGAATTGCC
TGGTGCTGAAGGACTATCCAGCAGAGAAAGTGGGAGGCGTGCTGAACCCATACCAGCTGACAGACC
AGTTCACCTCCTTTGCCAAGATGGGCACCCAGTCTGGCTTCCTGTTTTACGTGCCTGCCCCATATACAT
CTAAGATCGATCCCCTGACCGGCTTCGTGGACCCCTTCGTGTGGAAAACCATCAAGAATCACGAGAG
[0346] CCGCAAGCACTTCCTGGAGGGCTTCGACTTTCTGCACTACGACGTGAAAACCGGCGACTTCATCCTG
CACTTTAAGATGAACAGAAATCTGTCCTTCCAGAGGGGCCTGCCCGGCTTTATGCCTGCATGGGATAT
CGTGTTCGAGAAGAACGAGACACAGTTTGACGCCAAGGGCACCCCTTTCATCGCCGGCAAGAGAATC
GTGCCAGTGATCGAGAATCACAGATTCACCGGCAGATACCGGGACCTGTATCCTGCCAACGAGCTGAT
CGCCCTGCTGGAGGAGAAGGGCATCGTGTTCAGGGATGGCTCCAACATCCTGCCAAAGCTGCTGGAG
AATGACGATTCTCACGCCATCGAC**ACG**ATGGTGGCCCTGATCCGCAGCGTGCTGCAGATGCGGAACT
CCAATGCCGCCACAGGCGAGGACTATATCAACAGCCCCGTGCGCGATCTGAATGGCGTGTGCTTCGAC
TCCCGGTTTCAGAACCCAGAGTGG**CCA**ATGGACGCCGATGCCAATGGCGCCTACCACATCGCCCTGA
AGGGCCAGCTGCTGCTGAATCACCTGAAGGAGAGCAAGGATCTGAAGCTGCAGAACGGCATCTCCAA
TCAGGACTGGCTGGCCTACATCCAGGAGCTGCGCAAC*TCTGGTGGTTCTGGAGGATCTGGTGGTTCT***AC**
TAATCTGTCAGATATTATTGAAAAGGAGACCGGTAAGCAACTGGTTATCCAGGAATCCATCCTCA
TGCTCCCAGAGGAGGTGGAAGAAGTCATTGGGAACAAGCCGGAAAGCGATATACTCGTGCACA
CCGCCTACGACGAGAGCACCGACGAGAATGTCATGCTTCTGACTAGCGACGCCCCTGAATACA
AGCCTTGGGCTCTGGTCATACAGGATAGCAACGGTGAGAACAAGATTAAGATGCTC*TCTGGTGG*
*TTCT*CCCAAGAAGAAGAGGAAAGTCTAA (SEQ ID NO:442)

[0347] (JG1211)pCAG-人-dLbCpf1(D832A)-NLS(核质蛋白)-3xHA-VPR的核苷酸序列

[0348] 人类密码子优化的dLbCpf1(D832A)为正常字体(NTs 1-3921),核质蛋白NLS为小写字母(aaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaag,SEQ ID NO:21),3xHA标签(TACCCATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCG ACTATGCC,SEQ ID NO:5)为粗体,VPR为双下划线的

ATGAGCAAGCTGGAGAAGTTTACAAACTGCTACTCCCTGTCTAAGACCCTGAGGTTCAAGGCC
ATCCCTGTGGGCAAGACCCAGGAGAACATCGACAATAAGCGGCTGCTGGTGGAGGACGAGAAGAGA
GCCGAGGATTATAAGGGCGTGAAGAAGCTGCTGGATCGCTACTATCTGTCTTTTATCAACGACGTGCT
GCACAGCATCAAGCTGAAGAATCTGAACAATTACATCAGCCTGTTCCGGAAGAAAACCAGAACCGAG
[0349] AAGGAGAATAAGGAGCTGGAGAACCTGGAGATCAATCTGCGGAAGGAGATCGCCAAGGCCTTCAAG
GGCAACGAGGGCTACAAGTCCCTGTTTAAGAAGGATATCATCGAGACAATCCTGCCAGAGTTCCTGG
ACGATAAGGACGAGATCGCCCTGGTGAACAGCTTCAATGGCTTTACCACAGCCTTCACCGGCTTCTTT
GATAACAGAGAGAATATGTTTTCCGAGGAGGCCAAGAGCACATCCATCGCCTTCAGGTGTATCAACGA
GAATCTGACCCGCTACATCTCTAATATGGACATCTTCGAGAAGGTGGACGCCATCTTTGATAAGCACGA

GGTGCAGGAGATCAAGGAGAAGATCCTGAACAGCGACTATGATGTGGAGGATTTCTTTGAGGGCGAG
TTCTTTAACTTTGTGCTGACACAGGAGGGCATCGACGTGTATAACGCCATCATCGGCGGCTTCGTGAC
CGAGAGCGGCGAGAAGATCAAGGGCCTGAACGAGTACATCAACCTGTATAATCAGAAAACCAAGCA
GAAGCTGCCTAAGTTTAAGCCACTGTATAAGCAGGTGCTGAGCGATCGGGAGTCTCTGAGCTTCTACG
GCGAGGGCTATACATCCGATGAGGAGGTGCTGGAGGTGTTTAGAAACACCCTGAACAAGAACAGCGA
GATCTTCAGCTCCATCAAGAAGCTGGAGAAGCTGTTCAAGAATTTTGACGAGTACTCTAGCGCCGGCA
TCTTTGTGAAGAACGGCCCCGCCATCAGCACAATCTCCAAGGATATCTTCGGCGAGTGGAACGTGATC
CGGGACAAGTGGAATGCCGAGTATGACGATATCCACCTGAAGAAGAAGGCCGTGGTGACCGAGAAGT
ACGAGGACGATCGGAGAAAGTCCTTCAAGAAGATCGGCTCCTTTTCTCTGGAGCAGCTGCAGGAGTA
CGCCGACGCCGATCTGTCTGTGGTGGAGAAGCTGAAGGAGATCATCATCCAGAAGGTGGATGAGATC
TACAAGGTGTATGGCTCCTCTGAGAAGCTGTTCGACGCCGATTTTGTGCTGGAGAAGAGCCTGAAGA
AGAACGACGCCGTGGTGGCCATCATGAAGGACCTGCTGGATTCTGTGAAGAGCTTCGAGAATTACAT
CAAGGCCTTCTTTGGCGAGGGCAAGGAGACAAACAGGGACGAGTCCTTCTATGGCGATTTTGTGCTG
GCCTACGACATCCTGCTGAAGGTGGACCACATCTACGATGCCATCCGCAATTATGTGACCCAGAAGCC
CTACTCTAAGGATAAGTTCAAGCTGTATTTTCAGAACCCTCAGTTCATGGGCGGCTGGGACAAGGATA
AGGAGACAGACTATCGGGCCACCATCCTGAGATACGGCTCCAAGTACTATCTGGCCATCATGGATAAG
AAGTACGCCAAGTGCCTGCAGAAGATCGACAAGGACGATGTGAACGGCAATTACGAGAAGATCAACT
ATAAGCTGCTGCCCGGCCCTAATAAGATGCTGCCAAAGGTGTTCTTTTCTAAGAAGTGGATGGCCTACT
ATAACCCCAGCGAGGACATCCAGAAGATCTACAAGAATGGCACATTCAAGAAGGGCGATATGTTTAAC
CTGAATGACTGTCACAAGCTGATCGACTTCTTTAAGGATAGCATCTCCCGGTATCCAAAGTGGTCCAAT
GCCTACGATTTCAACTTTTCTGAGACAGAGAAGTATAAGGACATCGCCGGCTTTTACAGAGAGGTGGA
GGAGCAGGGCTATAAGGTGAGCTTCGAGTCTGCCAGCAAGAAGGAGGTGGATAAGCTGGTGGAGGA

[0350] GGGCAAGCTGTATATGTTCCAGATCTATAACAAGGACTTTTCCGATAAGTCTCACGGCACACCCAATCT
GCACACCATGTACTTCAAGCTGCTGTTTGACGAGAACAATCACGGACAGATCAGGCTGAGCGGAGGA
GCAGAGCTGTTCATGAGGCGCGCCTCCCTGAAGAAGGAGGAGCTGGTGGTGCACCCAGCCAACTCC
CCTATCGCCAACAAGAATCCAGATAATCCCAAGAAAACCACAACCCTGTCCTACGACGTGTATAAGGA
TAAGAGGTTTTCTGAGGACCAGTACGAGCTGCACATCCCAATCGCCATCAATAAGTGCCCCAAGAACA
TCTTCAAGATCAATACAGAGGTGCGCGTGCTGCTGAAGCACGACGATAACCCCTATGTGATCGGCATC
GCCAGGGGCGAGCGCAATCTGCTGTATATCGTGGTGGTGGACGGCAAGGGCAACATCGTGGAGCAGT
ATTCCCTGAACGAGATCATCAACAACTTCAACGGCATCAGGATCAAGACAGATTACCACTCTCTGCTG
GACAAGAAGGAGAAGGAGAGGTTCGAGGCCCGCCAGAACTGGACCTCCATCGAGAATATCAAGGAG
CTGAAGGCCGGCTATATCTCTCAGGTGGTGCACAAGATCTGCGAGCTGGTGGAGAAGTACGATGCCG
TGATCGCCCTGGAGGACCTGAACTCTGGCTTTAAGAATAGCCGCGTGAAGGTGGAGAAGCAGGTGTA
TCAGAAGTTCGAGAAGATGCTGATCGATAAGCTGAACTACATGGTGGACAAGAAGTCTAATCCTTGTG
CAACAGGCGGCGCCCTGAAGGGCTATCAGATCACCAATAAGTTCGAGAGCTTTAAGTCCATGTCTACC
CAGAACGGCTTCATCTTTTACATCCCTGCCTGGCTGACATCCAAGATCGATCCATCTACCGGCTTTGTG
AACCTGCTGAAAACCAAGTATACCAGCATCGCCGATTCCAAGAAGTTCATCAGCTCCTTTGACAGGAT
CATGTACGTGCCCGAGGAGGATCTGTTCGAGTTTGCCCTGGACTATAAGAACTTCTCTCGCACAGACG
CCGATTACATCAAGAAGTGGAAGCTGTACTCCTACGGCAACCGGATCAGAATCTTCCGGAATCCTAAG
AAGAACAACGTGTTCGACTGGGAGGAGGTGTGCCTGACCAGCGCCTATAAGGAGCTGTTCAACAAGT
ACGGCATCAATTATCAGCAGGGCGATATCAGAGCCCTGCTGTGCGAGCAGTCCGACAAGGCCTTCTAC
TCTAGCTTTATGGCCCTGATGAGCCTGATGCTGCAGATGCGGAACAGCATCACAGGCCGCACCGACGT
GGATTTTCTGATCAGCCCTGTGAAGAACTCCGACGGCATCTTCTACGATAGCCGGAACTATGAGGCCC
AGGAGAATGCCATCCTGCCAAAGAACGCCGACGCCAATGGCGCCTATAACATCGCCAGAAAGGTGCT
GTGGGCCATCGGCCAGTTCAAGAAGGCCGAGGACGAGAAGCTGGATAAGGTGAAGATCGCCATCTCT

AACAAGGAGTGGCTGGAGTACGCCCAGACCAGCGTGAAGCACaaaaggccggcggccacgaaaaaggccggccaggca
aaaaagaaaaagGGATCC**TACCCATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATG**
CATACCCATATGATGTCCCCGACTATGCCGGAAGCGAGGCCAGCGGTTCCGGACGGGCTGACGCAT
TGGACGATTTTGATCTGGATATGCTGGGAAGTGACGCCCTCGATGATTTTGACCTTGACATGCTTGGTT
CGGATGCCCTTGATGACTTTGACCTCGACATGCTCGGCAGTGACGCCCTTGATGATTTCGACCTGGAC
ATGCTGATTAACTCTAGAAGTTCCGGATCTCCGAAAAAGAAACGCAAAGTTGGTAGCCAGTACCTGCC
CGACACCGACGACCGGCACCGGATCGAGGAAAAGCGGAAGCGGACCTACGAGACATTCAAGAGCAT
CATGAAGAAGTCCCCCTTCAGCGGCCCCACCGACCCTAGACCTCCACCTAGAAGAATCGCCGTGCCC
AGCAGATCCAGCGCCAGCGTGCCAAAACCTGCCCCCCAGCCTTACCCCTTCACCAGCAGCCTGAGCA
CCATCAACTACGACGAGTTCCCTACCATGGTGTTCCCCAGCGGCCAGATCTCTCAGGCCTCTGCTCTG
GCTCCAGCCCCTCCTCAGGTGCTGCCTCAGGCTCCTGCTCCTGCACCAGCTCCAGCCATGGTGTCTGC
ACTGGCTCAGGCACCAGCACCCGTGCCTGTGCTGGCTCCTGGACCTCCACAGGCTGTGGCTCCACCA
GCCCCTAAACCTACACAGGCCGGCGAGGGCACACTGTCTGAAGCTCTGCTGCAGCTGCAGTTCGACG
[0351] ACGAGGATCTGGGAGCCCTGCTGGGAAACAGCACCGATCCTGCCGTGTTCACCGACCTGGCCAGCGT
GGACAACAGCGAGTTCCAGCAGCTGCTGAACCAGGGCATCCCTGTGGCCCCTCACACCACCGAGCCC
ATGCTGATGGAATACCCCGAGGCCATCACCCGGCTCGTGACAGGCGCTCAGAGGCCTCCTGATCCAG
CTCCTGCCCCTCTGGGAGCACCAGGCCTGCCTAATGGACTGCTGTCTGGCGACGAGGACTTCAGCTCT
ATCGCCGATATGGATTTCTCAGCCTTGCTGGGCTCTGGCAGCGGCAGCCGGGATTCCAGGGAAGGGAT
GTTTTTGCCGAAGCCTGAGGCCGGCTCCGCTATTAGTGACGTGTTTGAGGGCCGCGAGGTGTGCCAG
CCAAAACGAATCCGGCCATTTCATCCTCCAGGAAGTCCATGGGCCAACCGCCCACTCCCCGCCAGCCT
CGCACCAACACCAACCGGTCCAGTACATGAGCCAGTCGGGTCACTGACCCCGGCACCAGTCCCTCAG
CCACTGGATCCAGCGCCCGCAGTGACTCCCGAGGCCAGTCACCTGTTGGAGGATCCCGATGAAGAGA
CGAGCCAGGCTGTCAAAGCCCTTCGGGAGATGGCCGATACTGTGATTCCCCAGAAGGAAGAGGCTGC
AATCTGTGGCCAAATGGACCTTTCCCATCCGCCCCCAAGGGGCCATCTGGATGAGCTGACAACCACAC
TTGAGTCCATGACCGAGGATCTGAACCTGGACTCACCCCTGACCCCGGAATTGAACGAGATTCTGGAT
ACCTTCCTGAACGACGAGTGCCTCTTGCATGCCATGCATATCAGCACAGGACTGTCCATCTTCGACAC
ATCTCTGTTTTAA SEQ ID NO:443)

[0352] dLbCpf1(D832A)-NLS(核质蛋白)-3xHA-VPR的氨基酸序列

[0353] LbCpf1为正常字体(AAs 1-1228),NLS(核质蛋白)(krpaatkkagqakkkk,SEQ ID NO:7)为小写字母,3xHA标签(YPYDVPDYAYPYDVPDYAYPYDVPDYA,SEQ ID NO:8)为粗体,VPR为双下划线的

MSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRAEDYKGVKKLLDRYYLSFINDVL
HSIKLKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAFKGNEGYKSLFKKDIIETILPEFLDDKDEIA
LVNSFNGFTTAFTGFFDNRENMFSEEAKSTSIAFRCINENLTRYISNMDIFEKVDAIFDKHEVQEIKEKILNS
DYDVEDFFEGEFFNFVLTQEGIDVYNAIIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSD
RESLSFYGEGYTSDEEVLEVFRNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWN
VIRDKWNAEYDDIHLKKKAVVTEKYEDDRRKSFKKIGSFSLEQLQEYADADLSVVEKLKEIIIQKVDEIYK
[0354] VYGSSEKLFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEGKETNRDESFYGDFVLAYDIL
LKVDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAKC
LQKIDKDDVNGNYEKINYKLLPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNLNDCH
KLIDFFKDSISRYPKWSNAYDFNFSETEKYKDIAGFYREVEEQGYKVSFESASKKEVDKLVEEGKLYMFQI
YNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAELFMRRASLKKEELVVHPANSPIANKNPDNP
KKTTTLSYDVYKDKRFSEDQYELHIPIAINKCPKNIFKINTEVRVLLKHDDNPYVIGIARGERNLLYIVVVD
GKGNIVEQYSLNEIINNFNGIRIKTDYHSLLDKKEKERFEARQNWTSIENIKELKAGYISQVVHKICELVEK

[0355] YDAVIALEDLNSGFKNSRVKVEKQVYQKFEKMLIDKLNYMVDKKSNPCATGGALKGYQITNKFESFKSM
STQNGFIFYIPAWLTSKIDPSTGFVNLLKTKYTSIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYI
KKWKLYSYGNRIRIFRNPKKNNVFDWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMA
LMSLMLQMRNSITGRTDVDFLISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIGQFK
KAEDEKLDKVKIAISNKEWLEYAQTSVKHkrpaatkkagqakkkkGS**YPYDVPDYAYPYDVPDYAYPYDVPDY**
AGSEASGSGRADALDDFDLDMLGSDALDDFDLDMLGSDALDDFDLDMLGSDALDDFDLDMLINSRSSG
SPKKKRKVGSQYLPDTDDRHRIEEKRKRTYETFKSIMKKSPFSGPTDPRPPPRRIAVPSRSSASVPKPAPQP
YPFTSSLSTINYDEFPTMVFPSGQISQASALAPAPPQVLPQAPAPAPAPAMVSALAQAPAPVPVLAPGPPQA
VAPPAPKPTQAGEGTLSEALLQLQFDDEDLGALLGNSTDPAVFTDLASVDNSEFQQLLNQGIPVAPHTTEP
MLMEYPEAITRLVTGAQRPPDPAPAPLGAPGLPNGLLSGDEDFSSIADMDFSALLGSGSGSRDSREGMFLP
KPEAGSAISDVFEGREVCQPKRIRPFHPPGSPWANRPLPASLAPTPTGPVHEPVGSLTPAPVPQPLDPAPAVT
PEASHLLEDPDEETSQAVKALREMADTVIPQKEEAAICGQMDLSHPPPRGHLDELTTTLESMTEDLNLDSP
LTPELNEILDTFLNDECLLHAMHISTGLSIFDTSLF SEQ ID NO:444)

[0356] (RTW1008)pCAG-NLS(SV40)x2-rAPOBEC1-gsXTENgs接头-人类-dAsCpf1(D908A)-NLS(核质蛋白)-GS接头-UGI-NLS(SV40)的核苷酸序列

[0357] 人类密码子优化的dAsCpf1(D908A)为正常字体(NTs 844-4764),rAPOBEC1为粗体(NTs 67-750),核质蛋白NLS为小写字母(aaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaaag,SEQ ID NO:21),SV40 NLS为小写字母(ccaaagaaaaagaggaaagtc,cctaaaaagaaacgaaaggtt,或cccaagaagaagaggaaagtc,分别为SEQ ID NOs:19、20或22),gsXTENgs接头(tctggtggttcttctggtggttctagcggcagcgagactcccgggacctcagagtccgccacacccgaaagttccggagggagtagcggcggg,SEQ ID NO:23)为小写字母,UGI为双下划线的

[0358] ATGGGCccaaagaaaaagaggaaagtcGGCAGTGGAcctaaaaagaaacgaaaggttGGGTCAGGT**AGCTCAGAGA**
CTGGCCCAGTGGCTGTGGACCCCACATTGAGACGGCGGATCGAGCCCCATGAGTTTGAGGTAT
TCTTCGATCCGAGAGAGCTCCGCAAGGAGACCTGCCTGCTTTACGAAATTAATTGGGGGGGCC
GGCACTCCATTTGGCGACATACATCACAGAACACTAACAAGCACGTCGAAGTCAACTTCATCGA
GAAGTTCACGACAGAAAGATATTTCTGTCCGAACACAAGGTGCAGCATTACCTGGTTTCTCAGC
TGGAGCCCATGCGGCGAATGTAGTAGGGCCATCACTGAATTCCTGTCAAGGTATCCCCACGTCA
CTCTGTTTATTTACATCGCAAGGCTGTACCACCACGCTGACCCCCGCAATCGACAAGGCCTGCG
GGATTTGATCTCTTCAGGTGTGACTATCCAAATTATGACTGAGCAGGAGTCAGGATACTGCTGG
AGAAACTTTGTGAATTATAGCCCGAGTAATGAAGCCCACTGGCCTAGGTATCCCCATCTGTGGG
TACGACTGTACGTTCTTGAACTGTACTGCATCATACTGGGCCTGCCTCCTTGTCTCAACATTCTG
AGAAGGAAGCAGCCACAGCTGACATTCTTTACCATCGCTCTTCAGTCTTGTCATTACCAGCGAC
TGCCCCCACACATTCTCTGGGCCACCGGGTTGAAAtctggtggttcttctggtggttctagcggcagcgagactcccgggacct
cagagtccgccacacccgaaagttccggagggagtagcggcgggTCTACACAGTTCGAGGGCTTTACCAACCTGTATCAGGTG
AGCAAGACACTGCGGTTTGAGCTGATCCCACAGGGCAAGACCCTGAAGCACATCCAGGAGCAGGGC
TTCATCGAGGAGGACAAGGCCCGCAATGATCACTACAAGGAGCTGAAGCCCATCATCGATCGGATCTA
CAAGACCTATGCCGACCAGTGCCTGCAGCTGGTGCAGCTGGATTGGGAGAACCTGAGCGCCGCCATC
GACTCCTATAGAAAGGAGAAAACCGAGGAGACAAGGAACGCCCTGATCGAGGAGCAGGCCACATAT
CGCAATGCCATCCACGACTACTTCATCGGCCGGACAGACAACCTGACCGATGCCATCAATAAGAGACA
CGCCGAGATCTACAAGGGCCTGTTCAAGGCCGAGCTGTTTAATGGCAAGGTGCTGAAGCAGCTGGGC
ACCGTGACCACAACCGAGCACGAGAACGCCCTGCTGCGGAGCTTCGACAAGTTTACAACCTACTTCT
CCGGCTTTTATGAGAACAGGAAGAACGTGTTCAGCGCCGAGGATATCAGCACAGCCATCCCACACCG
CATCGTGCAGGACAACTTCCCCAAGTTTAAGGAGAATTGTCACATCTTCACACGCCTGATCACCGCCG

TGCCCAGCCTGCGGGAGCACTTTGAGAACGTGAAGAAGGCCATCGGCATCTTCGTGAGCACCTCCAT
CGAGGAGGTGTTTTCCTTCCCTTTTTATAACCAGCTGCTGACACAGACCCAGATCGACCTGTATAACC
AGCTGCTGGGAGGAATCTCTCGGGAGGCAGGCACCGAGAAGATCAAGGGCCTGAACGAGGTGCTGA
ATCTGGCCATCCAGAAGAATGATGAGACAGCCCACATCATCGCCTCCCTGCCACACAGATTCATCCCC
CTGTTTAAGCAGATCCTGTCCGATAGGAACACCCTGTCTTTCATCCTGGAGGAGTTTAAGAGCGACGA
GGAAGTGATCCAGTCCTTCTGCAAGTACAAGACACTGCTGAGAAACGAGAACGTGCTGGAGACAGC
CGAGGCCCTGTTTAACGAGCTGAACAGCATCGACCTGACACACATCTTCATCAGCCACAAGAAGCTG
GAGACAATCAGCAGCGCCCTGTGCGACCACTGGGATACACTGAGGAATGCCCTGTATGAGCGGAGAA
TCTCCGAGCTGACAGGCAAGATCACCAAGTCTGCCAAGGAGAAGGTGCAGCGCAGCCTGAAGCACG
AGGATATCAACCTGCAGGAGATCATCTCTGCCGCAGGCAAGGAGCTGAGCGAGGCCTTCAAGCAGAA
AACCAGCGAGATCCTGTCCCACGCACACGCCGCCCTGGATCAGCCACTGCCTACAACCCTGAAGAAG
CAGGAGGAGAAGGAGATCCTGAAGTCTCAGCTGGACAGCCTGCTGGGCCTGTACCACCTGCTGGACT
GGTTTGCCGTGGATGAGTCCAACGAGGTGGACCCCGAGTTCTCTGCCCGGCTGACCGGCATCAAGCT
GGAGATGGAGCCTTCTCTGAGCTTCTACAACAAGGCCAGAAATTATGCCACCAAGAAGCCCTACTCC
GTGGAGAAGTTCAAGCTGAACTTTCAGATGCCTACACTGGCCTCTGGCTGGGACGTGAATAAGGAGA
AGAACAATGGCGCCATCCTGTTTGTGAAGAACGGCCTGTACTATCTGGGCATCATGCCAAAGCAGAA
GGGCAGGTATAAGGCCCTGAGCTTCGAGCCCACAGAGAAAACCAGCGAGGGCTTTGATAAGATGTAC
TATGACTACTTCCCTGATGCCGCCAAGATGATCCCAAAGTGCAGCACCCAGCTGAAGGCCGTGACAG
CCCACTTTCAGACCCACACAACCCCCATCCTGCTGTCCAACAATTTCATCGAGCCTCTGGAGATCACA
AAGGAGATCTACGACCTGAACAATCCTGAGAAGGAGCCAAAGAAGTTTCAGACAGCCTACGCCAAG
AAAACCGGCGACCAGAAGGGCTACAGAGAGGCCCTGTGCAAGTGGATCGACTTCACAAGGGATTTT
CTGTCCAAGTATACCAAGACAACCTCTATCGATCTGTCTAGCCTGCGGCCATCCTCTCAGTATAAGGAC
[0359] CTGGGCGAGTACTATGCCGAGCTGAATCCCCTGCTGTACCACATCAGCTTCCAGAGAATCGCCGAGAA
GGAGATCATGGATGCCGTGGAGACAGGCAAGCTGTACCTGTTCCAGATCTATAACAAGGACTTTGCCA
AGGGCCACCACGGCAAGCCTAATCTGCACACACTGTATTGGACCGGCCTGTTTTCTCCAGAGAACCT
GGCCAAGACAAGCATCAAGCTGAATGGCCAGGCCGAGCTGTTCTACCGCCCTAAGTCCAGGATGAAG
AGGATGGCACACCGGCTGGGAGAGAAGATGCTGAACAAGAAGCTGAAGGATCAGAAAACCCCAATC
CCCGACACCCTGTACCAGGAGCTGTACGACTATGTGAATCACAGACTGTCCCACGACCTGTCTGATGA
GGCCAGGGCCCTGCTGCCCAACGTGATCACCAAGGAGGTGTCTCACGAGATCATCAAGGATAGGCGC
TTTACCAGCGACAAGTTCTTTTTCCACGTGCCTATCACACTGAACTATCAGGCCGCCAATTCCCCATCT
AAGTTCAACCAGAGGGTGAATGCCTACCTGAAGGAGCACCCCGAGACACCTATCATCGGCATCGCCC
GGGGCGAGAGAAACCTGATCTATATCACAGTGATCGACTCCACCGGCAAGATCCTGGAGCAGCGGAG
CCTGAACACCATCCAGCAGTTTGATTACCAGAAGAAGCTGGACAACAGGGAGAAGGAGAGGGTGGC
AGCAAGGCAGGCCTGGTCTGTGGTGGGCACAATCAAGGATCTGAAGCAGGGCTATCTGAGCCAGGTC
ATCCACGAGATCGTGGACCTGATGATCCACTACCAGGCCGTGGTGGTGCTGGAGAACCTGAATTTCGG
CTTTAAGAGCAAGAGGACCGGCATCGCCGAGAAGGCCGTGTACCAGCAGTTCGAGAAGATGCTGATC
GATAAGCTGAATTGCCTGGTGCTGAAGGACTATCCAGCAGAGAAAGTGGGAGGCGTGCTGAACCCAT
ACCAGCTGACAGACCAGTTCACCTCCTTTGCCAAGATGGGCACCCAGTCTGGCTTCCTGTTTTACGTG
CCTGCCCCATATACATCTAAGATCGATCCCCTGACCGGCTTCGTGGACCCCTTCGTGTGGAAAACCATC
AAGAATCACGAGAGCCGCAAGCACTTCCTGGAGGGCTTCGACTTTCTGCACTACGACGTGAAAACCG
GCGACTTCATCCTGCACTTTAAGATGAACAGAAATCTGTCCTTCCAGAGGGGCCTGCCCGGCTTTATG
CCTGCATGGGATATCGTGTTCGAGAAGAACGAGACACAGTTTGACGCCAAGGGCACCCCTTTCATCG
CCGGCAAGAGAATCGTGCCAGTGATCGAGAATCACAGATTCACCGGCAGATACCGGGACCTGTATCCT
GCCAACGAGCTGATCGCCCTGCTGGAGGAGAAGGGCATCGTGTTCAGGGATGGCTCCAACATCCTGC
CAAAGCTGCTGGAGAATGACGATTCTCACGCCATCGACACGATGGTGGCCCTGATCCGCAGCGTGCT

[0360] GCAGATGCGGAACTCCAATGCCGCCACAGGCGAGGACTATATCAACAGCCCCGTGCGCGATCTGAAT
GGCGTGTGCTTCGACTCCCGGTTTCAGAACCCAGAGTGGCCAATGGACGCCGATGCCAATGGCGCCT
ACCACATCGCCCTGAAGGGCCAGCTGCTGCTGAATCACCTGAAGGAGAGCAAGGATCTGAAGCTGCA
GAACGGCATCTCCAATCAGGACTGGCTGGCCTACATCCAGGAGCTGCGCAACaaaaggccggcggccacgaaaaa
ggccggccaggcaaaaaagaaaaagGGATCCTCTGGTGGTTCTGGAGGATCTGGTGGTTCTACTAATCTGTCAGATA
TTATTGAAAAGGAGACCGGTAAGCAACTGGTTATCCAGGAATCCATCCTCATGCTCCCAGAGGAGGTG
GAAGAAGTCATTGGGAACAAGCCGGAAAGCGATATACTCGTGCACACCGCCTACGACGAGAGCACC
GACGAGAATGTCATGCTTCTGACTAGCGACGCCCCTGAATACAAGCCTTGGGCTCTGGTCATACAGGA
TAGCAACGGTGAGAACAAGATTAAGATGCTCTCTGGTGGTTCTcccaagaagaagaggaaagtc (SEQ ID NO:445)

[0361] NLS(SV40)x2-rAPOBEC1-gsXTENgs接头-人类-dAsCpf1(D908A)-NLS(核质蛋白)-GS接头-UGI-NLS(SV40)的氨基酸序列

[0362] AsCpf1为正常字体(AAs x-xxxx),rAPOBEC1为粗体(AAs 23-250),NLS(核质蛋白)(krpaatkkagqakkkk,SEQ ID NO:7)为小写字母,SV40 NLS为小写字母(pkkkrkv,SEQ ID NO:24),gsXTENgs接头(sggssggssgsetpgtsesatpessggssgg,SEQ ID NO:25)为小写字母,UGI为双下划线的

[0363] MGpkkkrkvGSGpkkkrkvGSG**SSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGG
RHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFI
YIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVL
ELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK**sggssggssgsetpgtsesatpessggssgg
STQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTYADQCLQLVQLD
WENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKAELFNGKVL
KQLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVP
SLREHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKND
ETAHIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSIDLTHI
FISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEIISAAGKELSEAFKQ
KTSEILSHAHAALDQPLPTTLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEP
SLSFYNKARNYATKKPYSVEKFKLNFQMPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKA
LSFEPTEKTSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPE
KEPKKFQTAYAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLY
HISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYR
PKSRMKRMAHRLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIK
DRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIARGERNLIYITVIDSTGKILEQRSLNT
IQQFDYQKKLDNREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFKSKR
TGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKI
DPLTGFVDPFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKN
ETQFDAKGTPFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTM
VALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHLKES
KDLKLQNGISNQDWLAYIQELRNkrpaatkkagqakkkkGSSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLP
EEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSpkkkrkv
(SEQ ID NO:446)

[0364] (RTW1028)pCAG-NLS(SV40)x2-rAPOBEC1-gsXTENgs接头-人类-dAsCpf1(D908A)三重变体(E174R/S542R/K548R)-NLS(核质蛋白)-GS接头-UGI-NLS(SV40)的核苷酸序列

[0365] 人类密码子优化的dAsCpf1(D908A)为正常字体(NTs 844-4764),rAPOBEC1为粗体(NTs 67-750),核质蛋白NLS为小写字母(aaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaaag,SEQ ID NO:21),SV40 NLS为小写字母(ccaaagaaaaagaggaaagtc,

cctaaaaagaaacgaaaggtt,或cccaagaagaagaggaaagtc,分别为SEQ ID NOs:19、20或22),gsXTENgs接头(tctggtggttcttctggtggttctagcggcagcgagactcccgggacctcagagtccgccacacccgaaagttccggagggagtagcggcggg,SEQ ID NO:23)为小写字母,UGI为双下划线的

[0366]

ATGGGCccaaagaaaaagaggaaagtcGGCAGTGGAcctaaaaagaaacgaaaggttGGGTCAGGT**AGCTCAGAGA**
CTGGCCCAGTGGCTGTGGACCCCACATTGAGACGGCGGATCGAGCCCCATGAGTTTGAGGTAT
TCTTCGATCCGAGAGAGCTCCGCAAGGAGACCTGCCTGCTTTACGAAATTAATTGGGGGGGCC
GGCACTCCATTTGGCGACATACATCACAGAACACTAACAAGCACGTCGAAGTCAACTTCATCGA
GAAGTTCACGACAGAAAGATATTTCTGTCCGAACACAAGGTGCAGCATTACCTGGTTTCTCAGC
TGGAGCCCATGCGGCGAATGTAGTAGGGCCATCACTGAATTCCTGTCAAGGTATCCCCACGTCA
CTCTGTTTATTTACATCGCAAGGCTGTACCACCACGCTGACCCCCGCAATCGACAAGGCCTGCG
GGATTTGATCTCTTCAGGTGTGACTATCCAAATTATGACTGAGCAGGAGTCAGGATACTGCTGG
AGAAACTTTGTGAATTATAGCCCGAGTAATGAAGCCCACTGGCCTAGGTATCCCCATCTGTGGG
TACGACTGTACGTTCTTGAACTGTACTGCATCATACTGGGCCTGCCTCCTTGTCTCAACATTCTG
AGAAGGAAGCAGCCACAGCTGACATTCTTTACCATCGCTCTTCAGTCTTGTCATTACCAGCGAC
TGCCCCCACACATTCTCTGGGCCACCGGGTTGAAAtctggtggttcttctggtggttctagcggcagcgagactcccgggacct
cagagtccgccacacccgaaagttccggagggagtagcggcgggTCTACACAGTTCGAGGGCTTTACCAACCTGTATCAGGTG
AGCAAGACACTGCGGTTTGAGCTGATCCCACAGGGCAAGACCCTGAAGCACATCCAGGAGCAGGGC
TTCATCGAGGAGGACAAGGCCCGCAATGATCACTACAAGGAGCTGAAGCCCATCATCGATCGGATCTA
CAAGACCTATGCCGACCAGTGCCTGCAGCTGGTGCAGCTGGATTGGGAGAACCTGAGCGCCGCCATC
GACTCCTATAGAAAGGAGAAAACCGAGGAGACAAGGAACGCCCTGATCGAGGAGCAGGCCACATAT
CGCAATGCCATCCACGACTACTTCATCGGCCGGACAGACAACCTGACCGATGCCATCAATAAGAGACA
CGCCGAGATCTACAAGGGCCTGTTCAAGGCCGAGCTGTTTAATGGCAAGGTGCTGAAGCAGCTGGGC
ACCGTGACCACAACCGAGCACGAGAACGCCCTGCTGCGGAGCTTCGACAAGTTTACAACCTACTTCT
CCGGCTTTTATGCCAACAGGAAGAACGTGTTCAGCGCCGAGGATATCAGCACAGCCATCCCACACCG
CATCGTGCAGGACAACTTCCCCAAGTTTAAGGAGAATTGTCACATCTTCACACGCCTGATCACCGCCG
TGCCCAGCCTGCGGGAGCACTTTGAGAACGTGAAGAAGGCCATCGGCATCTTCGTGAGCACCTCCAT
CGAGGAGGTGTTTTCCTTCCCTTTTTATAACCAGCTGCTGACACAGACCCAGATCGACCTGTATAACC
AGCTGCTGGGAGGAATCTCTCGGGAGGCAGGCACCGAGAAGATCAAGGGCCTGAACGAGGTGCTGA
ATCTGGCCATCCAGAAGAATGATGAGACAGCCCACATCATCGCCTCCCTGCCACACAGATTCATCCCC
CTGTTTAAGCAGATCCTGTCCGATAGGAACACCCTGTCTTTCATCCTGGAGGAGTTTAAGAGCGACGA
GGAAGTGATCCAGTCCTTCTGCAAGTACAAGACACTGCTGAGAAACGAGAACGTGCTGGAGACAGC
CGAGGCCCTGTTTAACGAGCTGAACAGCATCGACCTGACACACATCTTCATCAGCCACAAGAAGCTG
GAGACAATCAGCAGCGCCCTGTGCGACCACTGGGATACACTGAGGAATGCCCTGTATGAGCGGAGAA
TCTCCGAGCTGACAGGCAAGATCACCAAGTCTGCCAAGGAGAAGGTGCAGCGCAGCCTGAAGCACG
AGGATATCAACCTGCAGGAGATCATCTCTGCCGCAGGCAAGGAGCTGAGCGAGGCCTTCAAGCAGAA
AACCAGCGAGATCCTGTCCCACGCACACGCCGCCCTGGATCAGCCACTGCCTACAACCCTGAAGAAG
CAGGAGGAGAAGGAGATCCTGAAGTCTCAGCTGGACAGCCTGCTGGGCCTGTACCACCTGCTGGACT
GGTTTGCCGTGGATGAGTCCAACGAGGTGGACCCCGAGTTCTCTGCCCGGCTGACCGGCATCAAGCT
GGAGATGGAGCCTTCTCTGAGCTTCTACAACAAGGCCAGAAATTATGCCACCAAGAAGCCCTACTCC

GTGGAGAAGTTCAAGCTGAACTTTCAGATGCCTACACTGGCCGCCGGCTGGGACGTGAATAAGGCCA
AGAACAATGGCGCCATCCTGTTTGTGAAGAACGGCCTGTACTATCTGGGCATCATGCCAAAGCAGAA
GGGCAGGTATAAGGCCCTGAGCTTCGAGCCCACAGAGAAAACCAGCGAGGGCTTTGATAAGATGTAC
TATGACTACTTCCCTGATGCCGCCAAGATGATCCCAAAGTGCAGCACCCAGCTGAAGGCCGTGACAG
CCCACTTTCAGACCCACACAACCCCCATCCTGCTGTCCAACAATTTCATCGAGCCTCTGGAGATCACA
AAGGAGATCTACGACCTGAACAATCCTGAGAAGGAGCCAAAGAAGTTTCAGACAGCCTACGCCAAG
AAAACCGGCGACCAGAAGGGCTACAGAGAGGCCCTGTGCAAGTGGATCGACTTCACAAGGGATTTT
CTGTCCAAGTATACCAAGACAACCTCTATCGATCTGTCTAGCCTGCGGCCATCCTCTCAGTATAAGGAC
CTGGGCGAGTACTATGCCGAGCTGAATCCCCTGCTGTACCACATCAGCTTCCAGAGAATCGCCGAGAA
GGAGATCATGGATGCCGTGGAGACAGGCAAGCTGTACCTGTTCCAGATCTATAACAAGGACTTTGCCA
AGGGCCACCACGGCAAGCCTAATCTGCACACACTGTATTGGACCGGCCTGTTTTCTCCAGAGAACCT
GGCCAAGACAAGCATCAAGCTGAATGGCCAGGCCGAGCTGTTCTACCGCCCTAAGTCCAGGATGAAG
AGGATGGCACACCGGCTGGGAGAGAAGATGCTGAACAAGAAGCTGAAGGATCAGAAAACCCCAATC
CCCGACACCCTGTACCAGGAGCTGTACGACTATGTGAATCACAGACTGTCCCACGACCTGTCTGATGA
GGCCAGGGCCCTGCTGCCCAACGTGATCACCAAGGAGGTGTCTCACGAGATCATCAAGGATAGGCGC
TTTACCAGCGACAAGTTCTTTTTCCACGTGCCTATCACACTGAACTATCAGGCCGCCAATTCCCCATCT
AAGTTCAACCAGAGGGTGAATGCCTACCTGAAGGAGCACCCCGAGACACCTATCATCGGCATCGCCC
GGGGCGAGAGAAACCTGATCTATATCACAGTGATCGACTCCACCGGCAAGATCCTGGAGCAGCGGAG
CCTGAACACCATCCAGCAGTTTGATTACCAGAAGAAGCTGGACAACAGGGAGAAGGAGAGGGTGGC
AGCAAGGCAGGCCTGGTCTGTGGTGGGCACAATCAAGGATCTGAAGCAGGGCTATCTGAGCCAGGTC
[0367] ATCCACGAGATCGTGGACCTGATGATCCACTACCAGGCCGTGGTGGTGCTGGAGAACCTGAATTTCGG
CTTTAAGAGCAAGAGGACCGGCATCGCCGAGAAGGCCGTGTACCAGCAGTTCGAGAAGATGCTGATC
GATAAGCTGAATTGCCTGGTGCTGAAGGACTATCCAGCAGAGAAAGTGGGAGGCGTGCTGAACCCAT
ACCAGCTGACAGACCAGTTCACCTCCTTTGCCAAGATGGGCACCCAGTCTGGCTTCCTGTTTTACGTG
CCTGCCCCATATACATCTAAGATCGATCCCCTGACCGGCTTCGTGGACCCCTTCGTGTGGAAAACCATC
AAGAATCACGAGAGCCGCAAGCACTTCCTGGAGGGCTTCGACTTTCTGCACTACGACGTGAAAACCG
GCGACTTCATCCTGCACTTTAAGATGAACAGAAATCTGTCCTTCCAGAGGGGCCTGCCCGGCTTTATG
CCTGCATGGGATATCGTGTTCGAGAAGAACGAGACACAGTTTGACGCCAAGGGCACCCCTTTCATCG
CCGGCAAGAGAATCGTGCCAGTGATCGAGAATCACAGATTCACCGGCAGATACCGGGACCTGTATCCT
GCCAACGAGCTGATCGCCCTGCTGGAGGAGAAGGGCATCGTGTTCAGGGATGGCTCCAACATCCTGC
CAAAGCTGCTGGAGAATGACGATTCTCACGCCATCGACACGATGGTGGCCCTGATCCGCAGCGTGCT
GCAGATGCGGAACTCCAATGCCGCCACAGGCGAGGACTATATCAACAGCCCCGTGCGCGATCTGAAT
GGCGTGTGCTTCGACTCCCGGTTTCAGAACCCAGAGTGGCCAATGGACGCCGATGCCAATGGCGCCT
ACCACATCGCCCTGAAGGGCCAGCTGCTGCTGAATCACCTGAAGGAGAGCAAGGATCTGAAGCTGCA
GAACGGCATCTCCAATCAGGACTGGCTGGCCTACATCCAGGAGCTGCGCAACaaaaggccggcggccacgaaaaa
ggccggccaggcaaaaaagaaaaagGGATCCTCTGGTGGTTCTGGAGGATCTGGTGGTTCTACTAATCTGTCAGATA
TTATTGAAAAGGAGACCGGTAAGCAACTGGTTATCCAGGAATCCATCCTCATGCTCCCAGAGGAGGTG
GAAGAAGTCATTGGGAACAAGCCGGAAAGCGATATACTCGTGCACACCGCCTACGACGAGAGCACC
GACGAGAATGTCATGCTTCTGACTAGCGACGCCCCTGAATACAAGCCTTGGGCTCTGGTCATACAGGA
TAGCAACGGTGAGAACAAGATTAAGATGCTCTCTGGTGGTTCTcccaagaagaagaggaaagtc (SEQ ID
NO:447)

[0368] NLS(SV40)x2-rAPOBEC1-gsXTENgs接头-人类-dAsCpf1(D908A)三重变体(E174R/S542R/K548R)-NLS(核质蛋白)-GS接头-UGI-NLS(SV40)的氨基酸序列

[0369] AsCpf1为正常字体(AAs x-xxxx),rAPOBEC1为粗体(AAs 23-250),NLS(核质蛋白)(krpaatkkagqakkkk,SEQ ID NO:7)为小写字母,SV40 NLS为小写字母(pkkkrkv,SEQ ID NO:24),gsXTENgs接头(sggssggssgsetpgtsesatpessggssgg,SEQ ID NO:25)为小写字母,UGI为双下划线的

MGpkkkrkvGSGpkkkrkvGSG**SSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGG**
RHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFI
YIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVL
ELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKsggssggssgsetpgtsesatpessggssgg
STQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTYADQCLQLVQLD
WENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKAELFNGKVL
KQLGTVTTTEHENALLRSFDKFTTYFSGFYRNRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVP
SLREHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKND
ETAHIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSIDLTHI
FISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEIISAAGKELSEAFKQ
KTSEILSHAHAALDQPLPTTLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEP
SLSFYNKARNYATKKPYSVEKFKLNFQMPTLARGWDVNREKNNGAILFVKNGLYYLGIMPKQKGRYKA
[0370] LSFEPTEKTSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPE
KEPKKFQTAYAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLY
HISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYR
PKSRMKRMAHRLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIK
DRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIARGERNLIYITVIDSTGKILEQRSLNT
IQQFDYQKKLDNREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFKSKR
TGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKI
DPLTGFVDPFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKN
ETQFDAKGTPFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTM
VALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHLKES
KDLKLQNGISNQDWLAYIQELRNkrpaatkkagqakkkkGSSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLP
EEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSpkkkrkv
(SEQ ID NO:448)

[0371] 实施例1.具有改变的PAM特异性的AsCpf1的变体

[0372] 为了试图改变Cpf1核酸酶的靶向范围,我们首先检查了AsCpf1和LbCpf1的现有的晶体结构(Dong,Nature 2016;Yamano,Cell 2016)。在其他观察结果中,这些结构证明PAM特异性由静电相互作用与间接碱基读数的组合所介导。因此我们猜测,在空间上接近PAM碱基的DNA碱基的残基处某些氨基酸取代的组合可能产生具有改变的或松弛的PAM识别偏好的变体。为了试验这一点,我们检查了PAM附近跨越残基G131-L137、S161-S181、N534-I555、Y595-T616、L628-F632和S685-I693的AsCpf1的区域(表1)。我们聚焦于其三维位置满足以下至少一个指标的参考AsCpf1序列中的氨基酸:1)空间上接近PAM DNA碱基(目标链或非目标链上),2)位置在DNA大沟或小沟之内,和/或3)残基的位置使得用带正电的替代物例如精氨酸、赖氨酸或组氨酸取代现有氨基酸预期可以提高侧链与磷酸二酯骨架的接近性(以及推测的相互作用)。由于含有crRNA的晶体结构和含有PAM的DNA仅可用于AsCpf1,因此根据三种同源基因之间的序列比对鉴定了LbCpf1和FnCpf1中的同源位置(表1)。

[0373] 表1:AsCpf1和LbCpf1中突变以产生改变的PAM识别特异性变体的候选残基的比较。使用或不使用FnCpf1进行比对。

[0374]

	AsCpf1	LbCpf1	LbCpf1(+18)	FnCpf1	比对参数
G131-L137	G131	**S117***	**S135***	**G133****	* 是指LbCpf1残基，仅来自与AsCpf1的比对 ** 是指FnCpf1残基，仅来自与AsCpf1的比对 黑色**文字**是指来自与所有三种Cpf1直系同源体的比对的残基(AsCpf1作为参考)
	L132	**L118***	**L136***	**L142****	
	F133	**F119***	**F137***	**K143****	
	K134	**K120***	**K138***	**Q144****	
	A135	**K121***	**K139***	**S145****	
	E136	**D122***	**D140***	**K146****	
	L137	**I123***	**I141***	**D147****	
S161-S181	S161	S143	S161	s171	黑色**文字**是指来自与所有三种Cpf1直系同源体的比对的残基
	F162	F144	F162	F172	
	D163	N145	N163	K173	
	K164	G146	G164	G174	
	F165	F147	F165	W175	
	T166	T148	T166	T176	
	T167	T149	T167	T177	
	Y168	A150	A168	Y178	
	F169	F151	F169	F179	
	S170	T152	T170	K180	
	G171	G153	G171	G181	
	F172	F154	F172	F182	
	Y173	F155	F173	H183	
	E174	D156	D174	E184	
	N175	N157	N175	N185	
	R176	R158	R176	R186	
	K177	E159	E177	K187	
	N178	N160	N178	N188	

[0375]

	V179	M161	M179	V189	
	F180	F162	F180	Y190	
	S181	S163	S181	S191	
N534-I555	N534	Y524	Y542	N599	黑色**文字**是指来自与所有三种Cpf1直系同源体的比对的残基
	F535	F525	F543	F600	
	Q536	Q526	Q544	E601	
	M537	N527	N545	N602	
	P538	P528	P546	S603	
	T539	Q529	Q547	T604	
	L540	F530	F548	L605	
	A541	M531	M549	A606	
	S542	G532	G550	N607	
	G543	G533	G551	G608	
	W544	W534	W552	W609	
	D545	D535	D553	D610	
	V546	K536	K554	K611	
	N547	D537	D555	N612	
	K548	K538	K556	K613	
	E549	E539	E557	E614	
	K550	T540	T558	P615	
	N551	D541	D559	D616	
	N552	Y542	Y560	N617	
	G553	R543	R561	T618	
	A554	A544	A562	A619	
	I555	T545	T563	I620	
Y595-T616	Y595	Y583	Y601	Y659	黑色**文字**是指来自与所有三种Cpf1直系同源体的比对的残基
	D596	K584	K602	K660	
	Y597	L585	L603	L661	
	F598	L586	L604	L662	
	P599	P587	P605	P663	
	D600	G588	G606	G664	
	A601	P589	P607	A665	
	A602	N590	N608	N666	
	K603	K591	K609	K667	
	M604	M592	M610	M668	
	I605	L593	L611	L669	
	P606	P594	P612	P670	
	K607	K595	K613	K671	
	C608	V596	V614	V672	
	S609	F597	F615	F673	
	T610	F598	F616	F674	
	Q611	S599	S617	S675	
	L612	K600	K618	A676	

[0376]

	K613	K601	K619	K677	
	A614	W602	W620	S678	
	V615	M603	M621	I679	
	T616	A604	A622	K680	
L628-F632	L628	**F598***	**F616***	**R690****	* 是指LbCpf1残基，仅来自与AsCpf1的比对
	S629	**S599***	**S617***	**I691****	** 是指FnCpf1残基，仅来自与AsCpf1的比对
	N630	**K600***	**K618***	**R692****	如果与AsCpf1、LbCpf1和FnCpf1进行比对，在这个区域内LbCpf1和FnCpf1都不与AsCpf1比对
	N631	**K601***	**K619***	**N693****	
	F632	**W602***	**W620***	**H694****	
S685-I693	S685	S644	S662	S729	黑色**文字**是指来自与所有三种Cpf1直系同源体的比对的残基
	K686	R645	R663	K730	** 是指FnCpf1残基，仅来自与AsCpf1的比对
	Y687	Y646	Y664	H731	
	T688	P647	P665	P732	
	K689	K648	K666	E733	
	T690	W649	W667	W734	
	T691	S650	S668	K735	
	S692	N651	N669	D736	
	I693	A652	A670	**F737****	

[0377] 在初始实验中，通过试验这些核酸酶在人类细胞中针对可选择的PAM目标位点的活性，所述目标位点在这些酶的初始表征中限定的TTTN基序内具有碱基差异(Zetsche, Cell 2015)，我们首先试图仔细定义野生型AsCpf1和LbCpf1的PAM偏好。使用我们的充分确立的基于人类细胞的EGFP破坏测定，我们首先试验了AsCpf1和LbCpf1诱变带有典型TTTN或非典型PAM的各种目标位点的能力，所述非典型PAM携带TTT基序内一系列不同的单碱基错配。我们确定了，虽然AsCpf1和LbCpf1都可以耐受PAM中的非典型碱基，但LbCpf1的识别是更加混杂的(图2)。在我们的使用可选择的CTTN、GTTN和TTCN PAM的EGFP破坏测定中，两种Cpf1酶都有效地和一致地靶向位点。(EGFP破坏测定测量EGFP表达的损失，作为基因组编辑核酸酶在EGFP序列内位点靶向的替代(Reyon,Nature Biotechnology 2012)。我们还检查了AsCpf1识别EGFP中含有更加趋异的PAM序列的位点的能力，所述PAM序列在TTTNPAM的TTT序列中具有两个或三个碱基差异。除了针对含有CCCN、CCTN、GATN、GCTN和TCCN PAM的位点的某些轻微的和可变的活性之外，我们发现野生型AsCpf1核酸酶没有有效地靶向任何这些可选择的PAM(图3)。

[0378] 然而，由于EGFP破坏测定中观察到的活性可能代表核酸酶介导的基因破坏与DNA结合介导的转录阻遏(没有切割)的组合，我们试验了AsCpf1和LbCpf1对内源人类基因目标位点的活性，所述内源人类基因目标位点含有相对于典型PAM具有单碱基差异的可选择的PAM序列，因为这项分析的读出结果真实地定量基因破坏事件(图4)。我们发现，虽然在我们的EGFP破坏测定中对于带有特定非典型PAM位点的位点用AsCpf1和LbCpf1观察到适度的活性，我们在带有这些可选择的PAM序列的内源基因位点上没有观察到插入或删除突变的可

比较的诱导。这个结果表明，虽然野生型AsCpf1和LbCpf1核酸酶可以有效地结合具有单碱基差异的可选择的PAM序列，在某些情况下它们不能有效地切割这些位点。

[0379] 考虑到野生型Cpf1核酸酶切割非TTTN PAM的有限能力，我们试图对可以靶向并破坏带有这些可选择的PAM的位点的Cpf1变体进行工程化。在初始实验中，我们试图工程化具有松弛的PAM识别特异性的Cpf1核酸酶。我们使用EGFP破坏测定来快速地鉴定和筛选AsCpf1中可以改变PAM识别的单氨基酸取代，原因是考虑到上文描述的基于EGFP的测定的限制，我们最终将需要验证我们鉴定的任意变体在人类细胞中切割实际的内源基因中的位点的能力。为了试验AsCpf1的理论性取代是否可以改变PAM特异性的假说，我们专注于试验以下位置处的氨基酸取代的作用：T167、S170、E174、T539、S542、K548、N551、N552、M604和K607。

[0380] 注意的是，随后的PAM编号基于以$T_4T_3T_2N_1$编号的TTTN PAM，每个AsCpf1变体的初始试验仅检查了所有可能的PAM的小的子集，对于变体的组合随后进行更为扩展的试验。

[0381] T167/T539：基于AsCpf1的共晶结构中T539对PAM的T_4的接近性，我们预想T539K/T167A或T539R/T167A的并发取代可以通过以下来允许PAM的第四位置处G的碱基特异性识别：1)T539取代为K或R允许鸟嘌呤的碱基特异性读出；和/或2)T167A取代减轻了T539残基改变所诱导的其他干扰或不利的接触。相对于野生型AsCpf1，T539K/T167A和T539R/T167A变体都在EGFP破坏测定中对于带有GTTN PAM的位点显示了改善的活性，对于带有典型TTTN PAM的位点的活性仅有最小的或仅仅适度的降低。

[0382] S170/E174：两个残基都位于DNA大沟之内，S170靠近PAM的T_2，E174位置靠近PAM的T_2或N_1位置，并靠近目标链DNA骨架。我们预想，这些位置处的精氨酸取代可以松弛PAM特异性，允许创造对DNA骨架的非特异性接触，或潜在地建立TTGN或TTTG PAM的碱基特异性识别。在EGFP破坏测定中S170R和E174R变体都提高了典型TTTN PAM处的活性，同时还提高了对携带GTTN PAM的位点的活性(图5B)。我们还显示了，S170R和E174R变体都赋予了靶向CTTN和TTCN PAM位点的提高的能力(图8C)。

[0383] S542：该残基位于大沟中非常接近PAM的T_3和T_2碱基。因此，我们猜测S542K或S542R突变可以通过以下来起到松弛PAM特异性的作用：1)在PAM结合界面中提供额外的非特异性能量以容纳非典型碱基，和/或2)产生可以潜在地识别非目标链上的G_3或G_2的碱基特异性接触，或可能是目标链上的G_3或G_2，其将是PAM的非目标链上的C_3或C_2。我们的假说将预测，在S542上携带取代的变体可以预计允许识别带有TGTN、TTGN、TCTN或TTCN PAM的位点。使用EGFP报告子分析，我们发现，具有S542K或S542R取代(而不是具有S542Q取代)的AsCpf1变体展现了对具有非典型TTCN PAM的目标位点的提高的活性(图5C)。感兴趣的是，当试验它们切割和诱变内源人类基因靶点的能力时，携带S542Q、S542K或S542R突变的AsCpf1变体都显示了对携带TTTN PAM的目标位点诱导插入或删除突变的提高的能力(图5D)，但是仅S542K和S542R变体显示了对携带非典型TTCN PAM的目标位点的提高的活性(图5D)。还在图7中进一步检查了这些变体对其他非典型PAM的识别(见下文)。

[0384] K548：该残基位于非PAM DNA链的A_4和A_3附近，并且靠近目标链DNA的骨架。因此我们猜测该位置处的取代可能潜在地提高针对具有非典型CTTN、TCTN或CCTN PAM的目标位点的活性。我们发现，引入K548R突变看起来自身没有对PAM特异性赋予实质上的改变，但是确实在位置S542、N551和N552的其他取代的环境下有助于松弛PAM识别(参见下文图9)。

[0385] N551/N552:残基N551和N552都位于目标和非目标DNA链骨架之间的大沟中,但是N552还非常靠近非PAM DNA链的A_3,并且靠近目标链DNA骨架。而N551R取代看起来没有不利影响,或在一种情况下可能稍微改善了对具有非典型GTTN PAM序列的目标位点的AsCpf1活性(不影响具有典型TTTN PAM的位点的识别),N552R取代看起来取消了对具有TTTN或GTTNPAM的目标位点的活性(图5E)。我们还研究了N551R和N552R取代与S542R突变的组合以及其他突变组合(参见下文的图8B和图11)。

[0386] M604/K607:残基M604位于靠近PAM的N1位置以及非PAM目标DNA链的A2核苷酸的DNA小沟中。残基K607也位于小沟中,与(PAM的)T_3和T_2以及非PAM DNA链的A_3形成网络。单独的K607处的多种不同取代看起来负面影响AsCpf1的活性(图5f),但是将K607H取代与S542R突变组合产生了具有针对带有典型TTTN或非典型TTCN PAM的位点的提高的活性(在图5g和5h中,分别通过EGFP破坏测定或通过内源人类基因目标位点诱变来判断)。类似地,当使用EGFP破坏测定进行分析时,M604A取代与S542R取代组合改善了针对带有典型TTTN或非典型TTCN PAM的目标位点的活性(图5g)。

[0387] 为了进一步试验由引入S542R突变产生的松弛的PAM特异性表型,我们比较了野生型AsCpf1和AsCpf1-S542R针对带有相对于典型序列具有单碱基差异的PAM的EGFP中的目标位点的EGFP破坏活性(包括我们推测可被S542R变体识别的四种非典型PAM)(图6A)。在这些实验中,我们观察到AsCpf1-S542R变体针对携带多种非典型PAM位点的目标位点的提高的活性(包括我们的碱基特异性接触模型所推测的四种之外的PAM(表明PAM结合亲和力的一般性改善和相关的PAM特异性的松弛)。为了进一步检查S542突变体扩展AsCpf1靶向范围的潜力,我们在一系列具有PAM的EGFP目标位点上比较了野生型AsCpf1和S542R变体,所述PAM带有相对于典型位点的2个或3个碱基差异(图6B)。在所有位点上S542R突变体维持了与野生型AsCpf1观察到的至少相同的活性水平,并显著地改善了针对许多带有两个或三个取代的PAM的位点的活性(在某些情况下,数倍)(图6B)。我们的EGFP破坏测定的结果表明,AsCpf1-S542R变体可以识别带有以下PAM的位点:TTTN、CTTN、GTTN、TCTN、TGTN、TTAN、TTCN、ATCN、CCCN、CCTN、GCTN、GGTN、TCCN和TGCN。我们接下来在携带具有一个(图7A)或者2个或3个碱基取代(图7B)的PAM的内源人类基因目标位点上试验了S542R变体。这些实验再一次揭示了,S542R变体可以切割广泛的错配PAM基序,但是这些位点的谱没有我们在EGFP破坏测定中观察到的那么广。基于我们使用这些内源人类基因目标位点的实验的结果,我们断定,AsCpf1-S542R变体可以切割带有以下PAM的位点:TTTN、ATTN、CTTN、GTTN、TCTN、TTCN、CCCN和TCCN。

[0388] S542处的单取代可以扩展AsCpf1的PAM识别特异性的观察结果表明,通过向该变体添加单个或多个突变有可能进一步提高靶向范围。如图5中所显示的和上文详细描述的,我们发现,S170、E174、K548、N551和K607(单独或组合)处的氨基酸取代产生了稍微改变的PAM识别特异性。因而,我们设法研究这些位置与S542R突变一起的其他取代组合是否可以进一步改善AsCpf1的靶向范围。首先,我们确定了,相对于单独的S542R取代,包括S542R/K548R、S542R/N551R和K548R/N551R的取代组合在大多数情况下可以改善对带有典型TTTN或非典型GTTN PAM的目标位点的活性(图8A)。接下来,使用EGFP破坏测定,我们确定对于具有更多样化的PAM序列的更大数量的目标位点:1)带有单个的S170R或E174R取代的变体对于许多目标位点可以超越S542R取代,2)对于一系列不同的目标位点,E174R/S542R、S542R/

K548R和S542R/N551R变体表现与单独的S542R一样好或更好，和3)针对带有典型TTTN和非典型CTTN、GTTN、TATN、TCTN、TGTN、TTAN、TTCN和TTGN PAM位点的一系列目标位点，在我们试验的一大批AsCpf1变体之中，E174R/S542R/K548R三取代变体表现了最高的活性水平(图8B和8C)。

[0389] 在携带具有单碱基差异的PAM的EGFP目标位点上AsCpf1E174R/S542R/K548R变体与野生型AsCpf1和AsCpf1-S542R变体的进一步比较(图9A)揭示了，E174R/S542R/K548R变体对所试验的几乎所有具有各种PAM位点的目标位点具有比野生型AsCpf1和S542更高的EGFP破坏活性。此外，对于具有携带两个或三个错配的PAM的一系列EGFP位点，当E174R/S542R/K548R变体与野生型AsCpf1进行比较时，观察到对于许多这些目标位点在EGFP破坏方面实质上的提高(图9B)。E174R/S542R/K548R三取代变体还显示了与S542R变体相比总体上更高的活性和针对更广范围的变体PAM(比较图9B和图6B)，包括带有以下PAM的位点的高效识别：TTTN、CTTN、GTTN、TATN、TCTN、TGTN、TTAN、TTCN、TTGN、AGTN、ATCN、CATN、CCCN、CCTN、CGTN、CTAN、CTCN、GATN、GCTN、GGTN、GTCN、TACN、TCCN和TGCN。

[0390] 由于上文提到的EGFP破坏测定对于评估Cpf1核酸酶活性的限制，我们接下来评估了我们的AsCpf1 E174R/S542R/K548R变体针对带有具有单碱基差异(图10A)或两个或三个碱基差异(图10B)的PAM的一系列内源人类基因目标位点的活性。这些结果证明，AsCpf1 E174R/S542R/K548R可以有效地切割携带以下PAM的位点：TTTN、ATTN、CTTN、GTTN、TATN、TCTN、TGTN、TTCN、ATCN、CCCN、CCTN、CTCN、GCTN、GGTN、TCCN和TGCN。重要的是注意到，AsCpf1-E174R/S542R/K548R没有对具有所有可能的PAM的目标位点进行试验。显然，携带大部分这些可选PAM的目标位点甚至不能被野生型LbCpf1切割，而野生型LbCpf1具有比野生型AsCpf1更为松弛的PAM偏好(图2和4)。此外，对于可以被野生型LbCpf1切割的变体PAM位点，根据插入或删除突变诱导效率来判断，我们观察到AsCpf1-E174R/S542R/K548R变体一致地超越野生型LbCpf1。感兴趣的是，AsCpf1-E174R/S542R/K548R变体也展现了针对典型TTTN PAM位点的改善的活性，甚至展现了对用AsCpf1或LbCpf1早先不能靶向的TTTN-6位点的实质上的活性(图10A和10B)。TTTN-6位点携带PAM的第一个位置上的T(TTTTPAM)，表明三取代AsCpf1变体可以改善针对携带PAM的第一位置中的T的位点的活性。因而，相对于野生型AsCpf1和LbCpf1核酸酶，AsCpf1-E174R/S542R/K548R变体实质上改善了Cpf1平台对于具有非典型PAM的位点的靶向范围，并且总体上也显示了对具有典型PAM的位点的改善的活性。

[0391] 接下来，为了尝试进一步松弛我们的AsCpf1 PAM变体的PAM特异性和/或改善在任何给定的PAM下的活性量级，我们对E174R/S542R/K548R变体添加了更多的氨基酸取代。首先，我们添加了N551R或N552R取代，从而分别产生四取代变体E174R/S542R/K548R/N551R和E174R/S542R/K548R/N552R。这两种四取代变体与野生型AsCpf1和E174R/S542R/K548R变体的比较揭示了，通过EGFP破坏测定或通过它们诱导内源人类基因目标位点中的插入或删除突变的能力来判断，E174R/S542R/K548R/N551R变体可以改善带有携带单个、两个或三个差异的各种非典型PAM的位点上的基因破坏活性(分别为图11A和11B)。相反，E174R/S542R/K548R/N552R四取代变体在这些相同的实验中不显示改善的活性，在很多情况下实际上取消活性(图11A和11B)。为了进一步比较E174R/S542R/K548R和E174R/S542R/K548R/N551R变体的PAM识别特异性，我们检查了它们对含有典型或非典型(单碱基差异)PAM的扩展的更大

数量的内源人类基因目标位点的活性。我们观察到两种变体对大部分位点的可比较的活性,少数情况下一种或另一种变体展现了稍微改善的活性(图11C)。

[0392] 实施例1B.具有改变的PAM特异性和改善的中靶活性的AsCas12a变体的进一步表征

[0393] 先前在人类细胞中表征Cas12a直系同源体揭示了,对于具有TTTV PAM的位点,As和LbCas12a一致地是更有效的核酸酶(Kim等人,Nat Biotechnol.,2016,34:863-8),Fn和MbCas12a可能拥有NTTN的松弛的PAM偏好(Zetsche等人,Cell,2015,163:759-71)。为更加彻底地评估每种直系同源体的活性和PAM偏好,在人类细胞中使用靶向带有TTTN或VTTN PAM的位点的两组十二个crRNA检查了它们的基因组编辑活性(图19A)。我们观察到对于TTTN PAM位点在四种直系同源体之间类似的基因破坏,然而观察到靶点特异性差异。此外,与As和LbCas12a相比时,Fn和Mb可以更有效地靶向VTTN PAM,但是与先前的报道一致的是,它们对VTTN位点的平均活性过低而不能将这些PAM表征为真实的靶点(图19A和19B)。这些结果支持了早先的观察,即针对带有TTTV PAM的位点Cas12a核酸酶大都是有效的(Kim等人,NatBiotechnol.,2016,34:863-8),迄今为止表征的天然发生的Cas12a直系同源体都没有显示克服了人类细胞中这种限制性的PAM需求。

[0394] 为了扩展Cas12a核酸酶的靶向范围并扩大有用性,我们利用AsCas12a三元复合物的结构研究(Yamano等人,Cell.2016May 5;165(4):949-62)来工程化能够识别典型和非典型PAM两者的单一变体。鉴定了空间上紧密接近PAMDNA碱基的残基(图20A),我们猜测这些残基的取代可以通过产生新的碱基相互作用或非特异性相互作用来改变或松弛PAM识别。我们首先在人类细胞中针对编码典型和非典型PAM的位点试验了在这些位置上编码单取代的变体的活性。与野生型AsCas12a相比,四种单取代变体(S170R、E174R、S542R、K548R)展现了对于典型的TTTA或TTTC PAM位点的优越的活性,同时还允许对具有非典型CTTA或TTCC PAM的位点的更有效的靶向(分别为图20B和20C)。

[0395] 人类细胞中这些取代的组合试验揭示了与野生型AsCas12a相比对四种携带非典型PAM(ATTC、CTTA、GTTC和TTCC)的其他位点的活性的实质上提高,并重现了对典型TTTG PAM位点的改善的活性的观察结果(图15A)。当组合E174R和S542R取代时观察到一些活性方面最突出的提高和靶向范围的扩大,与AsCas12a相比,E174R/S542R和E174R/S542R/K548R变体展现了对非典型PAM位点的4倍至32倍之间的改善的活性,以及对典型PAM位点的几乎2倍增强的活性(图20D)。因而,我们选择这两者变体用于进一步的表征。

[0396] 为了全面地剖析我们的AsCas12a变体的扩展的PAM偏好,我们优化了无偏差体外高通量PAM测定分析(PAMDA;图21A-21H)。我们首先纯化并分析了野生型和E174R/S542R/K548R AsCas12a(图21A)。在编码具有各种可靶向和非可靶向PAM的两个不同间隔子序列的质粒底物上验证了这些Cas12a核酸酶的体外切割活性,证实了早先在人类细胞分析中观察到的活性改善(图21B和21C)。我们然后通过构建编码这两个间隔子序列、但是带有随机的8-mer序列来代替PAM的两个独立的质粒文库,使该工作流程进行PAMDA(图21D)。使用AsCas12a/crRNA核糖核蛋白(RNP)复合物对这两种线性化的质粒文库进行时程切割反应,随后PCR扩增并测序未切割的底物,来计算可靶向PAM的消耗速率(图21D)。在独立的PAMDA实验之间,对于重复和间隔子之间的大多数间隔子近4nt PAM序列,观察到PAM比速率常数(k;群体中PAM随时间的消耗)之间的强相关性(分别为图21E和21F)。对于野生型和E174R/

S542R/K548R AsCas12a，可能的256种4nt PAM的每一个的$log_{10}k$值计数的分箱(binning)提示了处在-2.25$log_{10}k$范围内的真实PAM识别和靶向的近似阈值(图21G)。对来自相同PAM/间隔子组合的PAMDA数据的消耗曲线的分析显示了一致的切割分布，所述PAM/间隔子组合用于优化使用静态PAM底物的体外分析(图21H)。

[0397] 为了进行PAMDA，纯化的Cas12a核酸酶与crRNA复合，来查询带有随机8nt序列代替PAM的质粒文库，允许计算所述群体中可靶向PAM的消耗的体外速率常数(k)。如所预计的，野生型AsCas12a对所有可能的4nt PAM序列的平均$log_{10}k$值的曲线揭示了，靶向仅对具有TTTV PAM的位点是有效的(图15B)。反之，E174R/S542R/K548R变体展现了显著扩大的PAM类别靶向，包括TTTN和TTCN(TTYN)；ATTV、CTTV和GTTV(VTTV)；TATV和TGTV(TRTV)；以及许多其他PAM(图15B)。重要的是，该分析也支持了我们的观察结果，即，变体维持了典型TTTV PAM的有力的识别。

[0398] 接下来，为了对每种取代赋予的靶向范围改善获得更完整的了解，我们对产生E174R/S542R/K548R所必需的单取代和双取代中间变体进行PAMDA(图22A)。与我们的人类细胞分析数据(图15A)一致，该分析揭示了，E174R/S542R变体也展现了对广大范围的PAM的改善的活性。比较的E174R/S542R和E174R/S542R/K548R对于NNYN PAM的平均$log_{10}k$ PAMDA值证明两种变体都具有扩展的靶向范围(图22B)，表明E174R和S542R取代是造成大部分扩大的靶向范围的原因。感兴趣的是，这些残基的身份不是Cas12a直系同源体间共有的，但是存在于侧翼氨基酸严格保守的区域中(图22C)。

[0399] 为了进一步研究人类细胞中E174R/S542R和E174R/S542R/K548R变体展现的靶向范围改善，我们表征了它们对于PAMDA鉴定为可靶向或非可靶向的位点的活性，包括75个VTTN和TTCN位点，其带有现在每种变体都可接近的PAM(VTTT作为阴性对照；图23A)，以及17个具有TATN PAM的位点，其中TATV位点应当仅被E174R/S542R/K548R有效地靶向(图23B)。通过所述变体的PAMDA结果所预测的，我们观察到E174R/S542R和E174R/S542R/K548R对具有VTTV和TTCN PAM的位点的一致且稳健的靶向，其对VTTT位点的修饰无效，并观察到E174R/S542R/K548R变体仅对TATV位点的有效靶向(图15C、23A和23B)。重要地，与野生型AsCas12a相比，两种变体在靶向这些非典型PAM位点方面都是更有效得多的(图15D、19A和23C)。由于E174R/S542R/K548R变体的PAMDA结果表明它还可以潜在地识别超出我们已试验的那些的扩展的PAM范围，我们检查了人类细胞中15个带有TGTVPAM的位点以及83个其他位点，其携带处在-2.25的平均log_{10}(k)PAMDA阈值之处或附近的可选择的PAM(分别为图23D和23E)。我们观察到带有这些另外的非典型PAM的许多位点的稳健的修饰，以及大多数PAM的平均人类细胞活性和PAMDA ks之间的强相关性(图23F)。

[0400] 来自PAMDA的另一个观察结果是，E174R/S542R和E174R/S542R/K548R变体现在可以靶向野生型AsCas12a早先不能接近的TTTT PAM(图22A)。为测定这些变体是否可以有效地靶向具有非典型TTTT PAM的位点，同时还维持对典型TTTV PAM的活性，我们比较了它们在人类细胞中对25种另外的TTTN位点的活性(图23G)。与我们较早的发现(图15A)一致，我们观察到在修饰携带每种TTTV PAM的位点方面大约2倍的提高，以及对编码TTTT PAM的位点的大大改善的靶向(图15E)。这些结果表明，携带E174R和S542R的组合的变体不仅显著地改善靶向范围，还可令人惊讶地增强了对具有TTTNPAM的位点的靶向(图23H)。

[0401] 总体上，E174R/S542R/K548R变体，此后称为增强的AsCas12a(eAsCas12A)，允许靶

向范围的显著扩展和中靶活性的改善。现在eAsCas12a可接近的PAM可以基于PAMDA和人类细胞数据之间的一致性分类到置信度层级中(图15F和23I)。我们观察到人类细胞中的平均修饰百分比与体外测定的平均PAMDA ks之间的强相关性(图23F),表明PAMDA是人类细胞中可靶向和非可靶向PAM的合理预示。值得考虑的是,所述PAMDA数据产自编码两种独立的间隔子序列的文库,因而可能的是,从这些文库观察到的PAM偏好分布不代表所有间隔子序列的PAM排序(虽然我们确实观察到检查的两种间隔子文库之间良好的相关性,图21F)。

[0402] 因而,我们将对于所有检查的位点在人类细胞中满足大于20%平均靶向的严格阈值并且PAMDAk大于0.01的PAM分类为"层级1"PAM(TTYN、CTTV、RTTC、TATM、CTCC、TCCC和TACA),将满足细胞中大于10%的平均靶向的中度靶向阈值并且PAMDAk大于0.005的PAM分类为"层级2"PAM(RTTS、TATA、TGTV、ANCC、CVCC、TGCC、GTCC、TTAC)(图23I)。不一致的PAM(人类细胞数据与PAMDA数据之间的弱相关性)以及在人类细胞中的平均修饰小于10%的被分类为"层级3"PAM,考虑到我们当前的数据,不推荐用于大多数基因组编辑应用。对于靶向范围不是限制性的并且效率是主要目标的应用,我们建议根据它们的PAMDA以及人类细胞排序优先考虑层级1或2中的PAM(图23I)。

[0403] 总之,eAsCas12a和其他AsCas12a变体将靶向改善了超过8倍,允许基因组的编码区和非编码区的更高分辨率的靶向(图15G)。

[0404] eAsCas12a的改善的中靶活性

[0405] 在靶向范围之外,基因组编辑核酸酶的另一个关键性质是强力的中靶活性。因此我们设法更好地了解哪些取代有助于我们用eAsCas12a观察到的增强的靶向效率,就我们所知,还没有描述过提高CRISPR核酸酶的编辑效率的氨基酸取代。因而,我们首先确定了eAsCas12a或其衍生变体是否可以在如早先对野生型AsCas12a描述的更低的温度下还原DNA切割缺陷(Moreno-Mateos等人,Nat Commun.,2017,8:2024)。在37℃、32℃和25℃下的比较性体外切割反应揭示了,eAsCas12a几乎消除了在AsCas12a和LbCas12a之间观察到的温度依赖性切割差异,表型恢复很大程度上归因于E174R和S542R取代(图16A)。

[0406] Cas12a核酸酶的独特性质是它们从聚-crRNA转录产物中加工单个crRNA的能力(Fonfara等人,Nature,2016,532:517-21),简化了细胞中的多重靶向(Zetsche等人,Nat Biotechnol.,2017,35:31-34;Tak等人,Nat Methods,2017,14:1163-1166)。为了评估eAsCas12a的活性增强是否可以延及多重靶向,我们比较了用聚-crRNA阵列编程时As、eAs和LbCas12a的活性,所述聚-crRNA阵列各自编码靶向人类细胞中独立基因的三个crRNA(图18A-18C)。大多数情况下,我们观察到当聚-crRNA阵列从RNA聚合酶-III启动子表达时eAsCas12a的优越靶向性,推测是由于eAsCas12a对具有典型PAM的位点的增强的活性(图18A和18B)。当聚-crRNA从RNA聚合酶II启动子表达时也观察到多重靶向的这种改善,扩展了多重编辑应用的范围(图18C)。我们还设计了编码两组邻近靶向的crRNA的多重阵列以产生小的基因组删除。从聚-crRNA转录产物表达crRNA对,或者将单个crRNA质粒的库转染到细胞中,我们再次观察到eAsCas12a的改善的多重靶向性(图18D)。

[0407] 实施例4提供了额外的证据来支持E174R取代增强中靶活性的观察结果。

[0408] 实施例2.具有改变的PAM特异性的LbCpf1的变体

[0409] 由于AsCpf1和LbCpf1在围绕PAM的蛋白质-DNA接触附近的残基之间享有高度同源性(基于三维晶体结构和初步序列比对,参见表1),我们构造了LbCpf1 PAM变体,其将带有

与我们在AsCpf1中突变的位置相应残基处的残基。当与它们对应的AsCpf1变体(图5)相比较时，LbCpf1中位置T152、D156、G532和K538处的单取代(其相应于AsCpf1中的残基S170、E174、S542和K548)仅揭示了针对具有非典型PAM的位点在EGFP破坏活性方面的中度提高(图12A)。然而，在EGFP破坏测定中与野生型LbCpf1相比，三取代LbCpf1-D156R/G532R/K538R变体(类似于AsCpf1-E174R/S542R/K548R)展现了靶向具有非典型PAM的位点方面稍微更大幅度的提高(图12B)。

[0410] 实施例3.具有改变的PAM特异性的FnCpf1的变体

[0411] 之前的报道已经暗示FnCpf1在人类细胞中不工作或具有不良的活性(Zetsche,Cell 2015;Kim,Nature Biotechnology 2016)。由于AsCpf1和FnCpf1享有高度的同源性，我们首先设法试验野生型FnCpf1是否实际上在人类细胞中起作用，然后我们是否可以松弛早先报道的TTN的PAM特异性(Zetsche,Cell 2015)。

[0412] 为了检查FnCpf1在人类细胞中的活性，我们在我们的人类细胞EGFR破坏测定中试验了其针对含有NTTN、TNTN和TTNN形式的PAM的目标位点的活性(图13A)。我们的结果揭示了，野生型FnCpf1实际上可以在人类细胞中介导针对NTTN位点的稳健的EGFP破坏，针对TCTN和TTCN位点具有一定的可检测活性(图13A)。接下来，我们在携带TTTN PAM的10个不同的内源目标位点上将FnCpf1的内源基因破坏活性与AsCpf1和LbCpf1进行比较。在很多情况下，我们观察到FnCpf1与AsCpf1和LbCpf1可比较的活性，证明FnCpf1实际上在人类细胞中强健地起作用(图13B)。

[0413] 由于FnCpf1在人类细胞中起作用，我们设法确定，通过在导致PAM特异性改变的AsCpf1的同源位置的残基处产生携带取代的FnCpf1变体，是否可以产生FnCpf1 PAM变体(表1)。在我们检查的取代之中，K180R、N607R和D616R的单取代在TTTN、TNTN和NTTN PAM位点处产生相对于野生型AsCpf1的活性提高(图13C)。另外，K671H突变可以提高针对TCTN PAM位点的活性。我们还观察到，携带取代组合的变体，包括N607R/K613R、N607R/K613V、N607R/K613V/D616R或N607R/K613R/D616R，在TTTN、CTTN、GTTN、TATN、TCTN、TCTN、TTAN或TTCN形式的某些PAM处相对于野生型FnCpf1的活性改善(图13C)。

[0414] 实施例4.具有改变的PAM特异性的其他AsCpf1变体

[0415] Gao等人近来公开了具有改变的PAM特异性的其他Cpf1变体(Gao等人，“Engineered Cpf1 Enzymes with Altered PAM Specificities,”bioRxiv 091611;doi:https://doi.org/10.1101/091611)。这些变体以及它们声称的对典型和/或非典型PAM的活性如下：

[0416] 1)AsCpf1-S542R/K548V/N552R-针对TATV PAM位点起作用

[0417] 2)AsCpf1-S542R/K607R-针对TYCV PAM位点最佳地起作用，但是展现了针对典型的TTTV位点的活性损失。

[0418] 为了针对S542R/K548V/N552R变体评估本文描述的三取代或四取代AsCpf1变体(分别为E174R/S542R/K548R和E174R/S542R/K548R/N551R)，我们对携带典型的TTTN、TATN(据报道被S542R/K548V/N552R变体识别)和具有单碱基或双碱基差异的PAM的目标位点使用EGFP破坏测定比较这三种变体的活性(图14A)。对于所有试验的位点，我们观察到在TTTV、TATN和其他非典型的CTTN、GTTN、TCTN、TGTN、TTAN、TTCN、TTGN和TCCN PAM处我们的三取代和四取代变体优于S542R/K548V/N552R变体(图14A)。接下来，基于我们早先的观察结

果,即S170R或E174R取代当与其他取代组合时可以提高AsCpf1变体的活性,我们研究了向S542R/K548V/N552R变体添加这些取代之一是否也可以改善其活性。在将S170R/S542R/K548V/N552R和E174R/S542R/K548V/N552R四取代变体与亲本S542R/K548V/N552R比较时,我们观察到添加S170R或E174R取代实质上改善了活性(E174R的作用大于S170R,而添加S170R也产生改善;图14A)。

[0419] 对于携带典型TTTN PAM序列、或具有单碱基差异的PAM的EGFP中的许多目标位点,我们还比较了我们的三取代和四取代AsCpf1变体(分别为E174R/S542R/K548R和E174R/S542R/K548R/N551R)相对于S542R/K607R变体的活性。对于所有这些位点,当与S542R/607R变体比较时,我们的三取代和四取代变体(E174R/S542R/K548R和E174R/S542R/K548R/N551R)具有大约相等的或更高水平的EGFP破坏活性(图14B)。S542R/K607R仅在具有TCCNPAM的目标位点上优于我们的三重和四重变体。因此,我们向S542R/K607R变体添加S170R或E174R取代以产生三取代的S170R/S542R/K607R和E174R/S542R/K607R变体。在所有试验的位点上,这些额外的三取代变体与S542R/K607R变体相比表现同样好或者更好,值得注意的是在具有TCCNPAM的位点上也表现良好(图14B)。最后,我们在携带非典型PAM(PAM中有1、2或3个碱基差异)的U2OS细胞中的一系列内源人类基因目标位点处比较了我们的三取代变体之一(E174R/S542R/K548R)与S542R/K607R变体。除了携带NCCC PAM的位点之外在所有试验的位点上,我们的E174R/S542R/K548R变体与S542R/K607R变体相比表现同样好或更好(图14C)。对于S542R/K607R变体展示了更高的基因破坏活性的NCCC PAM位点,我们比较了该变体与我们的S170R/S542R/K607R、E174R/S542R/K607R和E174R/S542R/K607H变体,并观察到在所有情况下,我们的三取代变体之一优于S542R/K607H变体(图14C)。

[0420] 实施例4B.改善AsCas12a PAM变体的中靶活性

[0421] 由于我们的结果表明,E174R和S542R产生人类细胞中eAsCas12a的活性增强,我们猜测在早先描述的已经编码S542R的AsCas12a变体中包含E174R也可以改善它们的活性。因而,E174R取代与RVR(S542R/K548V/N552R)和RR(S542R/K607R)变体组合来创造这些核酸酶的增强形式(分别为eRVR和eRR)。比较eAsCas12a、RVR、eRVR、RR和eRR变体在人类细胞中针对具有TTTN PAM的11个位点的活性(图24A),揭示了虽然早先公开的RVR和RR变体具有与野生型AsCas12a相比类似的或更弱的活性,添加E174R来创造eRVR和eRR变体引起了它们的活性方面大于2倍的提高(虽然仍然低于eAsCas12a;图16B)。这些结果强化了该观察结果,即携带E174R和S542R的组合的变体可以改善中靶活性。

[0422] 接下来,由于我们对eAsCas12a的PAMDA评估揭示了识别早先报道为RVR和RR变体可接近的初始PAM(分别为TATV和TYCV PAM;图15B),在人类细胞中这些位点处,我们比较了eAsCas12a与这些核酸酶的公开的和增强的形式。对于12个TATN位点(图24B),我们注意到eAsCas12a展示了与RVR变体大约相等的活性(图16C)。感兴趣的是,向RVR添加E174R引起活性的2倍改善,表明对于靶向TATN位点是主要目标的应用,eRVR是最佳的变体(图16C)。然后我们在人类细胞中29个携带TYCN PAM的位点上评估了eAsCas12a、RR和eRR(图24C)。对于18个TTCN位点eAsCas12a展现了与RR相比更高的修饰,而含有E174R的eRR变体具有与eAsCas12a可比较的活性(图16D)。在11个TCCN位点上进一步比较这些变体揭示了,虽然在具有TCCNPAM的位点上RR变体是相比eAsCas12a更有效的核酸酶,含有E174R的eRR变体再一次产生了对于所有TCCN位点最稳健的修饰(图16D)。

[0423] 为了确定通过添加E174R是否已经改变了增强的eRVR和eRR变体的靶向范围，我们将PAMDA应用于RVR、RR和它们的增强变体(图24D)。与我们的人类细胞数据一致，我们观察到，eRVR和eRR核酸酶具有与它们的亲本RVR和RR变体类似的靶向范围，但是通过添加E174R取代改善了它们的中靶效力(图16E和24E)。总之，这些结果证明，E174R和S542R取代不仅改善靶向范围，它们还可以改善AsCas12a核酸酶的中靶活性。

[0424] 实施例5.增强eAsCas12a的基因组广度特异性

[0425] 考虑到相比野生型AsCas12a，eAsCas12a展现了增强的活性和松弛的PAM识别，我们设法比较这些核酸酶的特异性，因为它们辨别中靶和脱靶位点的能力对于研究和治疗应用都是关键的。就此来说，我们和其他人早先已经显示，野生型Cas12a核酸酶拥有稳健的基因组广度特异性，并且相对不耐受错配的脱靶位点，所述脱靶位点在间隔子的紧邻PAM近端、中间和PAM远端区域中带有单个或两个错配(Kleinstiver等人,Nat Biotechnol.,2016,34:869-74;Kim等人,Nat Biotechnol.,2016,34:863-8;WO2018/022634)。因此，我们使用了通过测序(GUIDE-seq)方法而可行的基因组广度的DSB无偏鉴定(Tsai等人,Nat Biotechnol.,2015,33:187-197)，来比较As和eAsCas12a在四个具有TTTV PAM的位点上的基因组广度特异性(图25A-25D)。野生型AsCas12a通过GUIDE-Seq检测到少数脱靶，对于eAsCas12a我们观察到脱靶数量的提高(图17A和17B)。对eAsCas12a观察到的许多脱靶是早先使用LbCas12a的GUIDE-seq实验中已经鉴定出的(Kleinstiver等人,Nat Biotechnol.,2016,34:869-74)，在已知耐受核苷酸取代的位置中含有错配(Kleinstiver等人,Nat Biotechnol.,2016,34:869-74;Kim等人,Nat Biotechnol.,2016,34:863-8)，或编码现在可靶向的非典型PAM(图17A)。

[0426] 为了研究定向进化方法是否可以用于改善eAsCas12a的保真度，我们检查了AsCas12a中预计造成与DNA的非特异性接触的残基处氨基酸取代的影响。我们分析了九种不同的推测的高保真度(HF)取代的单错配耐受性(包括早先描述的K949A变体)，观察到虽然某些取代改善了AsCas12a对于两个位点的单错配耐受性分布，许多取代也降低了使用匹配的crRNA时的活性(图25E)。我们将最有前景的取代与eAsCas12a组合，观察到eAsCas12a的N282A形式(命名为eAsCas12a-HF1)产生了在单错配不耐受性方面最希望的改善，并且维持了中靶活性(图25F)。使用PAMDA评估eAsCas12a和eAsCas12a-HF1揭示了几乎相同的PAM偏好分布(图25G和25H)，表明N282A HF突变不改变PAM识别或靶向范围(图17C)。

[0427] 接下来，为了确定eAsCas12a-HF1是否可以改善基因组广度特异性，我们使用同样的四个早先检查的TTTV PAM靶向的crRNA进行GUIDE-Seq。与eAsCas12a相比，我们观察到对于4种crRNA中的3种，在用eAsCas12a-HF1检测到脱靶的位置观察到数量和频率的降低(图17A和17B)，其中它们的特异性分布现在更加接近于野生型AsCas12a所观察到的。进行了其他GUIDE-Seq实验，对于具有非典型PAM的位点比较eAsCas12a和eAsCas12a-HF1(图17D和25B-25D)，我们再次观察到与eAsCas12a相比使用eAsCas12a-HF1时脱靶的数量和频率降低(分别为图17E和17F)

[0428] 我们然后比较了对于典型和非典型PAM位点eAsCas12a和eAsCas12a-HF1的中靶活性(分别为图25I和25J)，来检查N282A取代是否影响靶向效率。我们观察到对于8个TTTN PAM位点的类似的基因修饰(再一次相对于野生型AsCas12a提高了近3倍的效率;图17G)，以及在15个携带非典型PAM的位点上eAsCas12a与eAsCas12a-HF1之间可比较的活性(图17H)。

此外，评估温度耐受性的体外切割分析揭示了在37℃、32℃和25℃下eAsCas12a、eAsCas12a-HF1和LbCas12a之间相似的切割分布(图25K)。总之，这些结果证明，eAsCas12a-HF1可以改善基因组广度特异性，同时维持重要的靶向范围、提高的活性以及温度耐受性性质。

[0429] 实施例6.利用eAsCas12a的增强的性质用于基因激活和表观基因组编辑应用

[0430] CRISPR-Cas12a的另一个卓越的适配是用于表观基因组编辑，其中已经显示DNase失活的Cas12a(dCas12a)与异源效应子的融合物调节基因表达。我们早先证明了，与人类细胞中等价的dAsCas12a融合物相比，与合成的VPR反式-激活结构域(VP64、p65和Rta的组合；Chavez等人，Nat Methods.，2015，12:326-8)的dLbCas12a融合物介导更强力的基因激活(Tak等人，Nat Methods，2017，14:1163-1166)。为了研究与LbCas12a效应子相比eAsCas12a是否可以改善表观基因组编辑，我们首先设计并试验了不同结构的与VPR的dAs、deAs和dLbCas12a融合物(图26A)。比较它们对VEGFA启动子近端的典型TTTV和非典型TTCV PAM的活性，建立了最佳的dCas12a-VPR融合物结构(版本1.1；图26B-26D)，还揭示了在使用靶向典型或非典型位点的crRNA时，相对于dAs与dLbCas12a融合物，deAs-VPR效应子促成了更大的VEGFA生产(图26E)。在比较deAs-VPR与原型dSpCas9-VPR融合物的实验中(靶向独立但附近的位点)，我们用deAs-VPR观察到可比较的或更好的基因激活(图26B-C和26E-F)。

[0431] 使用靶向三种其他内源基因(NPY1R、HBB和AR)的启动子中的位点的dAs、deAs和dLb-VPR融合物的额外实验再一次揭示了在使用典型PAM(图18E)或非典型PAM靶向的crRNA(图18F和18G)时deAs-VPR的最强力的基因激活。deAs-VPR融合物实现了10至10,000倍之间的基因激活，经常地优于dAs或dLbCas12a-VPR超过10-100倍。

[0432] 总的来说，当靶向典型的TTTV PAM位点时，与公开的dLbCas12a-VPR融合物相比，与VPR的deAsCas12a融合物可以以相等的或更高的效率介导稳健的基因激活，并且通过靶向仅可用这种eAsCas12a变体接近的非典型PAM位点，还提供了新的能力来激活基因。这些结果重现了eAsCas12a的增强的活性和改善的靶向范围，并提供了强力的和广泛可靶向的基因激活技术，其也可以适用于其他表观基因组编辑应用。

[0433] 实施例7.用于碱基编辑应用的AsCas12a和LbCas12a的变体

[0434] 近年来已经使用工程化的SpCas9碱基编辑子(BE)构建体展现了进行精确的单碱基编辑事件的能力(参见，例如，Komor等人，Nature.2016 May 19;533(7603):420-4；Nishida等人，Science.2016 Sep 16;353(6305)；Kim等人，Nat Biotechnol.2017 Apr;35(4):371-376；Komor等人，Sci Adv.2017 Aug 30;3(8):eaao4774；和Gaudelli等人，Nature.2017 Nov 23;551(7681):464-471)，其利用了SpCas9-gRNA形成的R-环形成，其导致非目标DNA链的ssDNA可接近性。因此异源的胞苷或腺嘌呤脱氨酶的酶结构域与SpCas9的融合可以作用于暴露的ssDNA链，分别引起C到T或A到G的有效导入。由于细胞碱基切除修复(BER)采用尿嘧啶DNA糖基化酶(UDG；也称为尿嘧啶N-糖基化酶，或UNG)来切下尿嘧啶碱基，由于胞苷的脱氨基化产生尿嘧啶中间物，这种内源的过程可以有效地逆转胞苷BE产生的编辑。因此，为了改善胞苷BE的效率，异源效应子结构域例如尿嘧啶糖基化酶抑制剂(UGI)也可以融合到SpCas9以抑制UDG，破坏BER的起始并提高胞苷BE的有效性。

[0435] 由于我们早先的观察表明，eAsCas12a(E174R/S542R/K548R)拥有增强的活性，因此我们想知道eAsCas12a是否可以允许开发推测的AsCas12a碱基-编辑子(AsBEs)。为了试

验这种假说，我们克隆了四种不同的DNase失活的eAsBE结构(BE-1.1-1.4;图18H)，其包括rAPOBEC1的N-末端融合物、D908A取代以取消核酸酶介导的DNA水解活性、以及UGI的C-末端融合物，并使用八种不同的crRNA将它们的活性与野生型AsBE1.1和1.4进行比较。对于8个位点的7个中的所有的C，我们用AsBEs观察到最小的(<1％)C到T编辑(图18I)。感兴趣的是，对于同样的八个位点，eAsBE融合物展现了远大于绝对水平的C到T转换(2-34％的编辑;图18I和18J)，相对于AsBEs显著改善编辑(图27A)。两种相当的LbBE结构的评估(范围在2-19％的C到T编辑)揭示了相对于eAsBEs可比较水平的C到T编辑(图18I和18J)。对于所有构建体，编辑效率根据目标位点和BE结构而不同(图18I)，类似于使用SpCas9BEs的观察结果，C的5'存在G看起来阻抑C到T编辑(图27B)。对于高效率编辑的位置，使用Cas12a-BEs观察到希望的编辑纯度(优势的C到T，作为主要产物)(图27C)。Cas12a-BEs观察到低水平的插入或删除突变，推测是由于它们的DNase活性的失活(图18K和图27D)。总之，这些结果证明，eAsCas12a的活性增强允许以早先用AsBEs不可实现的水平、以及以与LbBEs可比较的效力进行C到T编辑，并且扩展了CRISPR碱基编辑试剂的潜力。

[0436] 实施例8.用于DNA检测的AsCas12a变体

[0437] 近年来描述的CRISPR-Cas12a核酸酶的其他应用基于该观察结果，即Cas12a分子展现了靶点编程的非特异性DNase活性(Chen等人，Science，2018，doi:10.1126/science.aar6245)，这种性质已经用于溶液中特定DNA分子的敏感检测(Chen等人，Science，2018，doi:10.1126/science.aar6245;Gootenberg等人，Science，2018，doi:10.1126/science.aaq0179)。当Cas12a-crRNA复合物结合目标DNA时，催化性的RuvC DNase活性位点采取超活性构象，其不加区别地消化附近的DNA。合成的淬灭荧光团DNA-报告分子可以添加到溶液中，方便Cas12a-DNase活性的定量，其将释放荧光报告子(East-Seletsky，Nature，2016，538:270-273)。因而，eAsCas12a的扩展的靶向范围和改善的活性，通过增强敏感性、便于检测具有非典型PAM的DNA分子、或通过允许检测用于诊断目的的变体等位基因，可潜在地改善DNA检测方法。

[0438] 因此，我们设法评价野生型AsCas12a和eAsCas12a的附带的反式-DNase活性，来评估我们的工程化变体与DNA检测工作流程的相容性。我们在体外组装了Cas12-crRNA复合物，用活化性(匹配报告分子)或非活化性(对照)DNA底物对它们编程，之后添加报告分子。我们还改变了活化性DNA底物上编码的PAM，以确定eAsCas12a的扩展的靶向范围是否在这种体外分析中再现。在用野生型AsCas12a进行的实验中，在存在编码典型TTTA PAM位点的匹配底物的情况下，我们观察到稳健的检测，以及在用携带非典型ACCT PAM的底物编程时，观察到大大降低的活性(图28A)。接下来，eAsCas12a和eAsCas12-HF1都展现了在TTTA PAM底物上与野生型AsCas12a相比可比较水平的检测(图28B)，但是也可以稳健地检测非典型的CTTA PAM底物(考虑到eAsCas12a的扩展的PAM偏好分布所预计的)(图15B)。令人惊讶地，eAsCas12a酶也能检测携带非可靶向ACCT PAM的DNA底物(图28B)，暗示在对于原型目标DNA切割的PAM要求或者非特异性反式-DNase活性方面的潜在差异。这些结果证明，eAsCas12a和eAsCas12a-HF1都是用于DNA检测的强力的工程化核酸酶，其提供了相对于野生型AsCas12a的靶向范围和潜在特异性的优点。

[0439] 其他实施方案

[0440] 要理解的是虽然已经连同其详细说明一起描述了本发明，以上的描述意图说明而

不是限制本发明的范围，本发明的范围由附随的权利要求的范围所定义。其他方面、优点和改变在所附权利要求的范围之内。

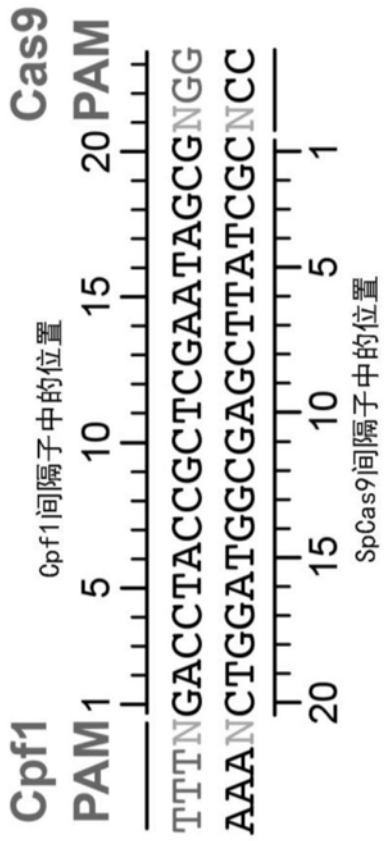

Cpf1
PAM
1
5
10
15
20
Cpf1间隔子中的位置
TTTNGACCTACCGCTCGAATAGCGGNGG
AAANCTGGATGGCGAGCTTATCGCCNCC
20
15
10
5
1
SpCas9间隔子中的位置
Cas9
PAM

图1

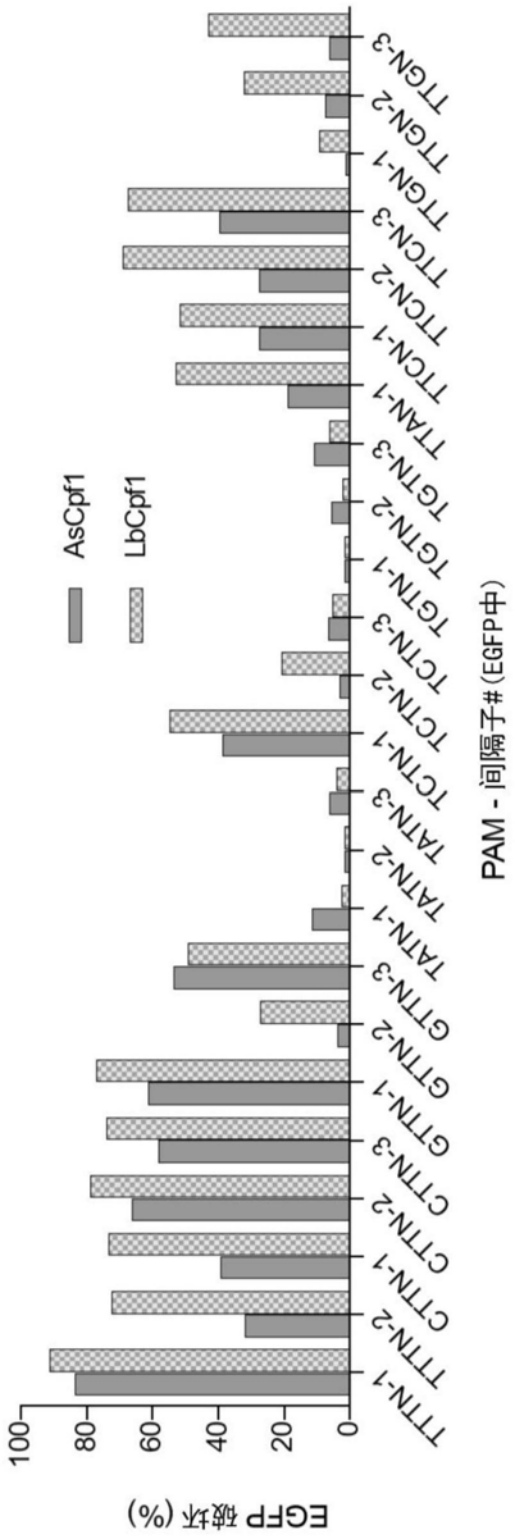

EGFP 破坏 (%)
100
80
60
40
20
0
AsCpf1
LbCpf1
PAM - 间隔子#(EGFP中)
TTTN-1
TTTN-2
CTTN-1
CTTN-2
CTTN-3
GTTN-1
GTTN-2
GTTN-3
TATN-1
TATN-2
TATN-3
TCTN-1
TCTN-2
TCTN-3
TGTN-1
TGTN-2
TGTN-3
TTAN-1
TTCN-1
TTCN-2
TTCN-3
TTGN-1
TTGN-2
TTGN-3

图2

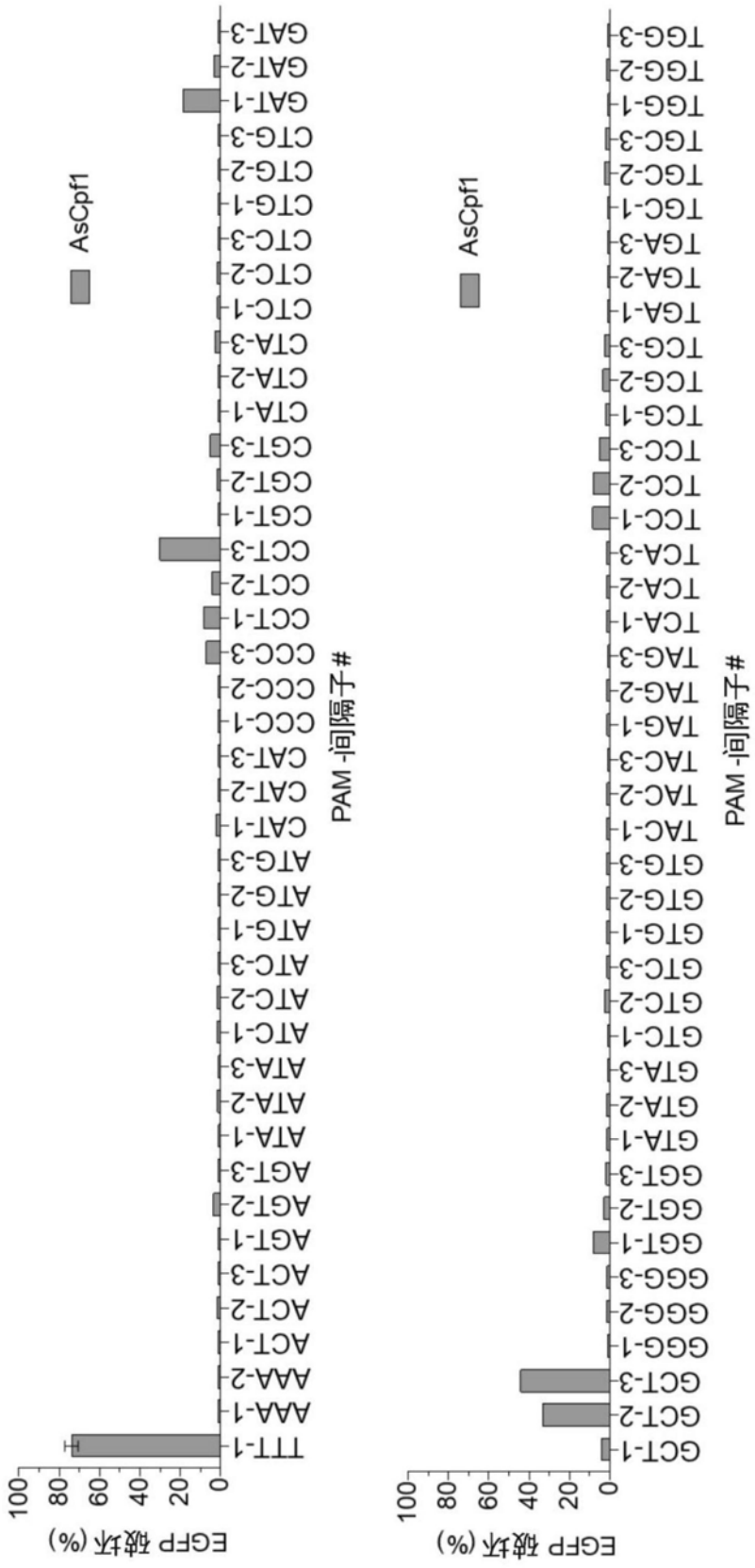

AsCpf1
EGFP 破坏(%)
PAM -间隔子#
TTT-1
AAA-1
AAA-2
ACT-1
ACT-2
ACT-3
AGT-1
AGT-2
AGT-3
ATA-1
ATA-2
ATA-3
ATC-1
ATC-2
ATC-3
ATG-1
ATG-2
ATG-3
CAT-1
CAT-2
CAT-3
CCC-1
CCC-2
CCC-3
CCT-1
CCT-2
CCT-3
CGT-1
CGT-2
CGT-3
CTA-1
CTA-2
CTA-3
CTC-1
CTC-2
CTC-3
CTG-1
CTG-2
CTG-3
GAT-1
GAT-2
GAT-3
AsCpf1
EGFP 破坏(%)
PAM -间隔子#
GCT-1
GCT-2
GCT-3
GGG-1
GGG-2
GGG-3
GGT-1
GGT-2
GGT-3
GTA-1
GTA-2
GTA-3
GTC-1
GTC-2
GTC-3
GTG-1
GTG-2
GTG-3
TAC-1
TAC-2
TAC-3
TAG-1
TAG-2
TAG-3
TCA-1
TCA-2
TCA-3
TCC-1
TCC-2
TCC-3
TCG-1
TCG-2
TCG-3
TGA-1
TGA-2
TGA-3
TGC-1
TGC-2
TGC-3
TGG-1
TGG-2
TGG-3

图3

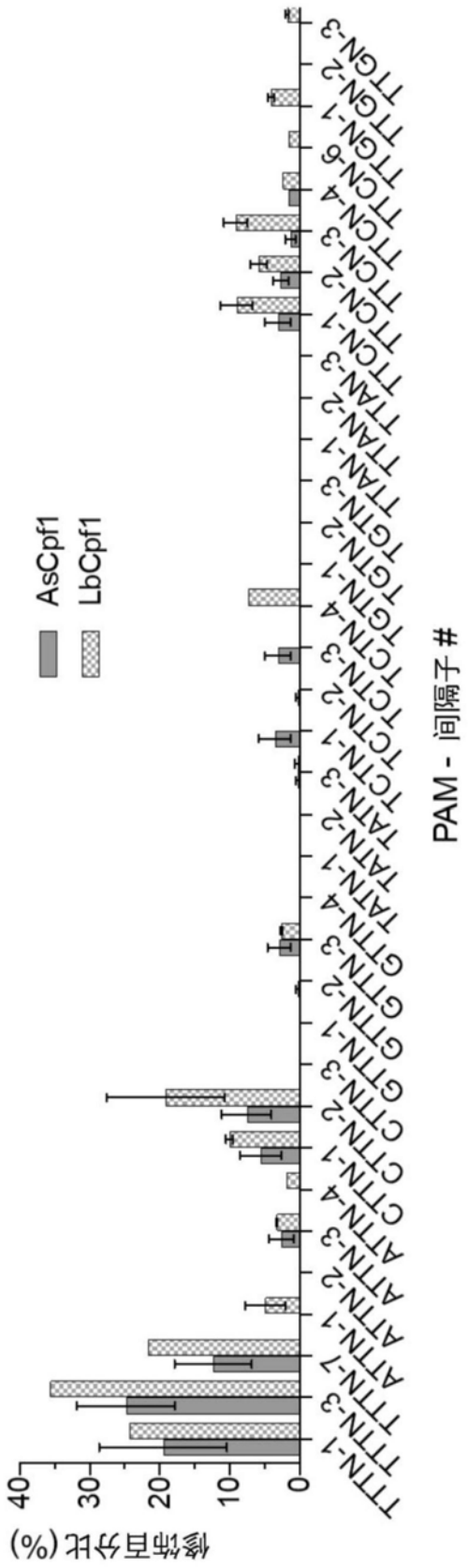

AsCpf1
LbCpf1
修饰百分比(%)
0
10
20
30
40
PAM - 间隔子 #
TTTN-1
TTTN-3
ATTN-1
ATTN-2
ATTN-3
ATTN-4
CTTN-1
CTTN-2
CTTN-3
GTTN-1
GTTN-2
GTTN-3
GTTN-4
TATN-1
TATN-2
TATN-3
TCTN-1
TCTN-2
TCTN-3
TCTN-4
TGTN-1
TGTN-2
TGTN-3
TTAN-1
TTAN-2
TTAN-3
TTCN-1
TTCN-2
TTCN-3
TTCN-4
TTCN-6
TTGN-1
TTGN-2
TTGN-3

图4

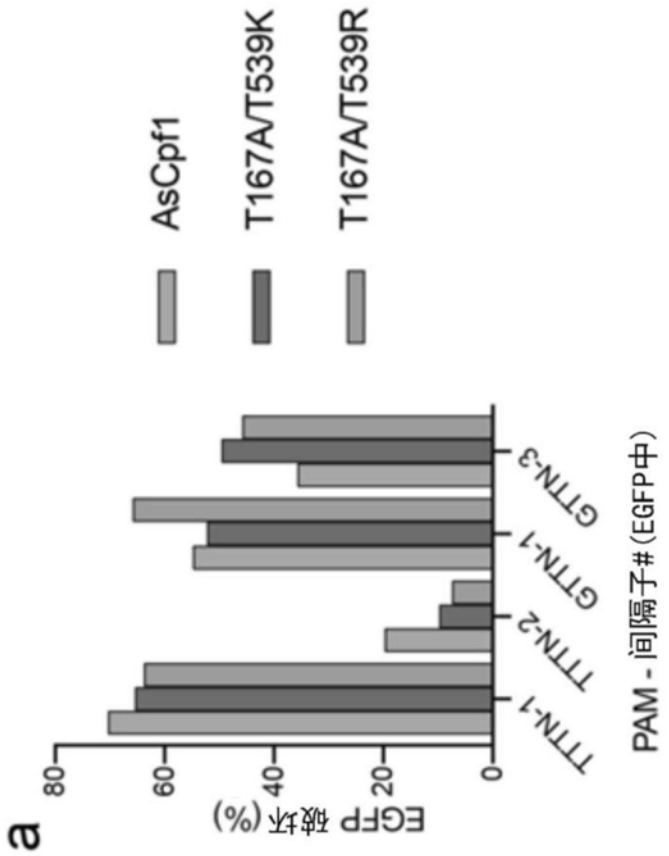

a
AsCpf1
T167A/T539K
T167A/T539R
EGFP 破坏 (%)
0
20
40
60
80
TTTN-1
TTTN-2
GTTN-1
GTTN-3
PAM - 间隔子# (EGFP中)

图5A

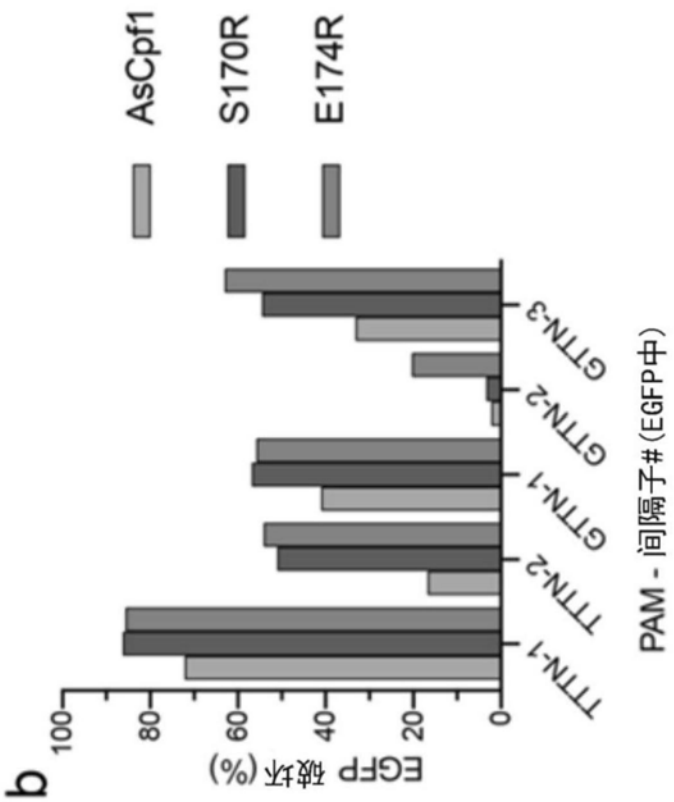

b
AsCpf1
S170R
E174R
EGFP 破坏(%)
0
20
40
60
80
100
TTTN-1
TTTN-2
GTTN-1
GTTN-2
GTTN-3
PAM - 间隔子#(EGFP中)

图5B

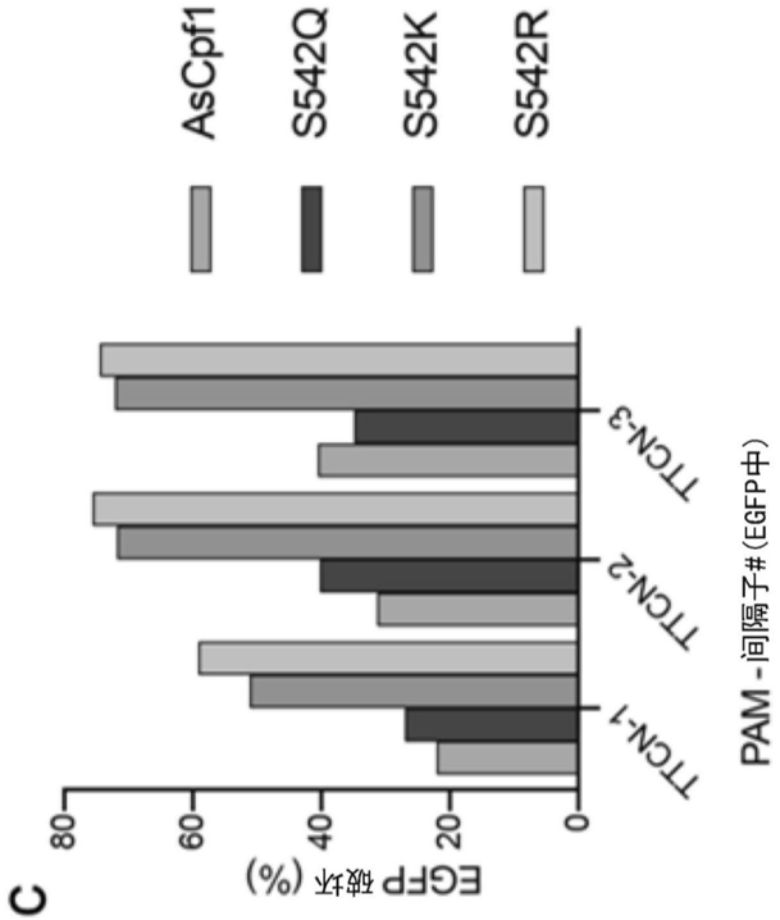

C
AsCpf1
S542Q
S542K
S542R
EGFP 破坏 (%)
0
20
40
60
80
TTCN-1
TTCN-2
TTCN-3
PAM - 间隔子#(EGFP中)

图5C

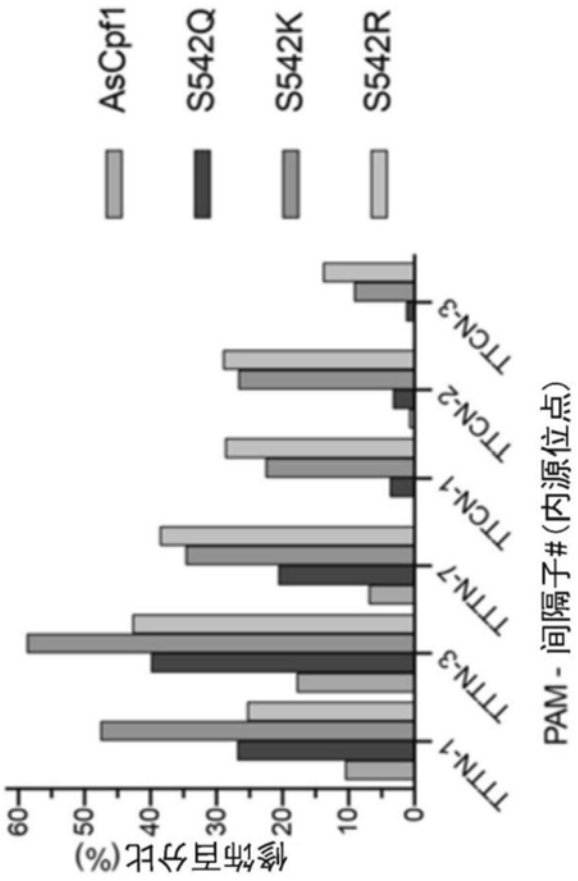

AsCpf1
S542Q
S542K
S542R
修饰百分比(%)
0
10
20
30
40
50
60
TTTN-1
TTTN-3
TTTN-7
TTCN-1
TTCN-2
TTCN-3
PAM - 间隔子#(内源位点)

图5D

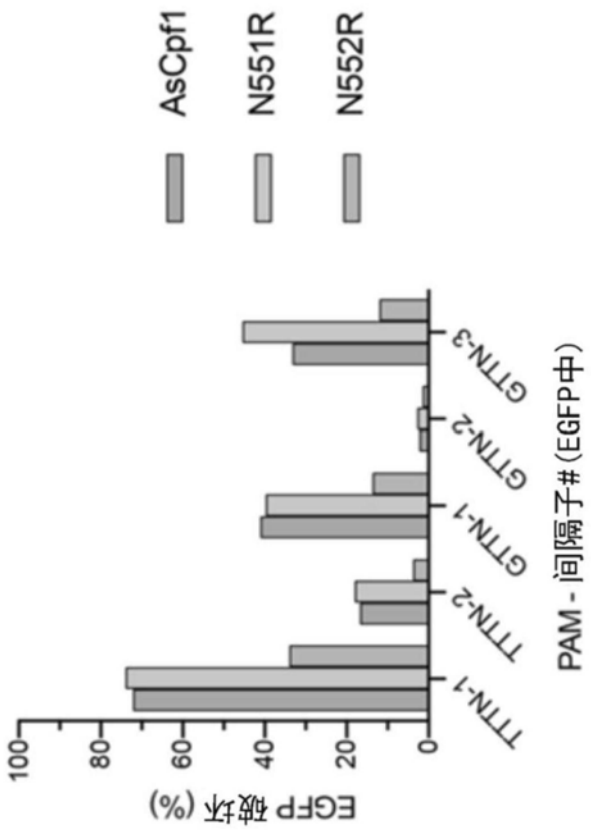

AsCpf1
N551R
N552R
EGFP 破坏 (%)
0
20
40
60
80
100
TTTN-1
TTTN-2
GTTN-1
GTTN-2
GTTN-3
PAM - 间隔子#(EGFP中)

图5E

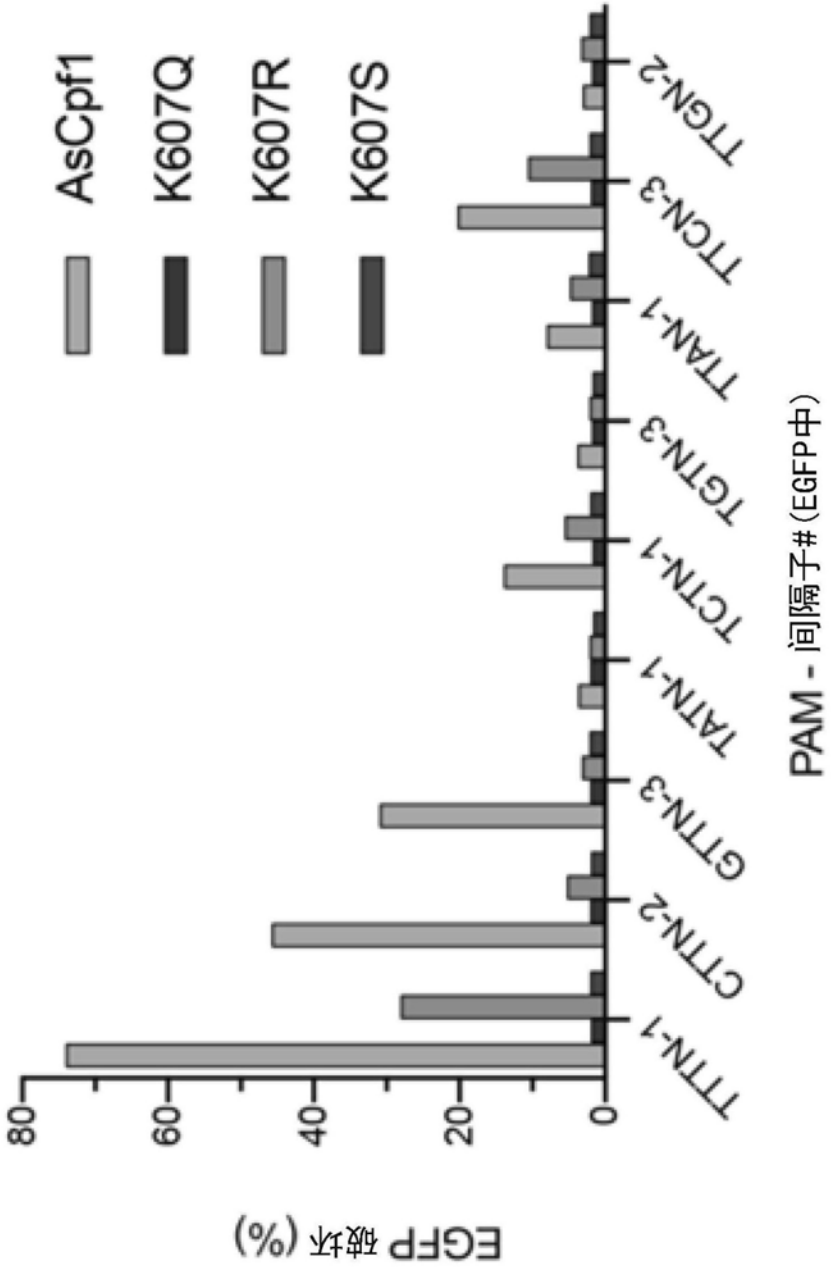

AsCpf1
K607Q
K607R
K607S
EGFP 破坏 (%)
0
20
40
60
80
TTTN-1
CTTN-2
GTTN-3
TATN-1
TCTN-1
TGTN-3
TTAN-1
TTCN-3
TTGN-2
PAM - 间隔子#(EGFP中)

图5F

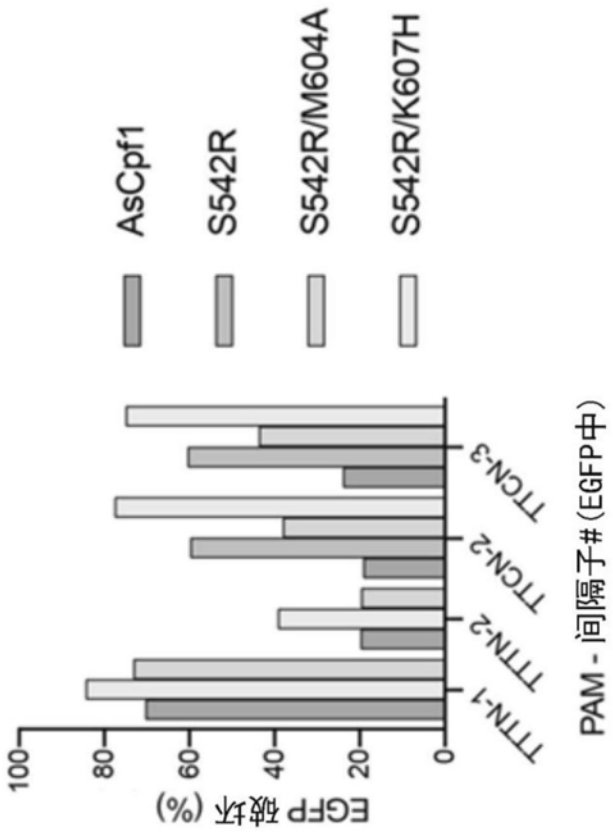

AsCpf1
S542R
S542R/M604A
S542R/K607H
EGFP 破坏 (%)
100
80
60
40
20
0
TTTN-1
TTTN-2
TTCN-2
TTCN-3
PAM - 间隔子#(EGFP中)

图5G

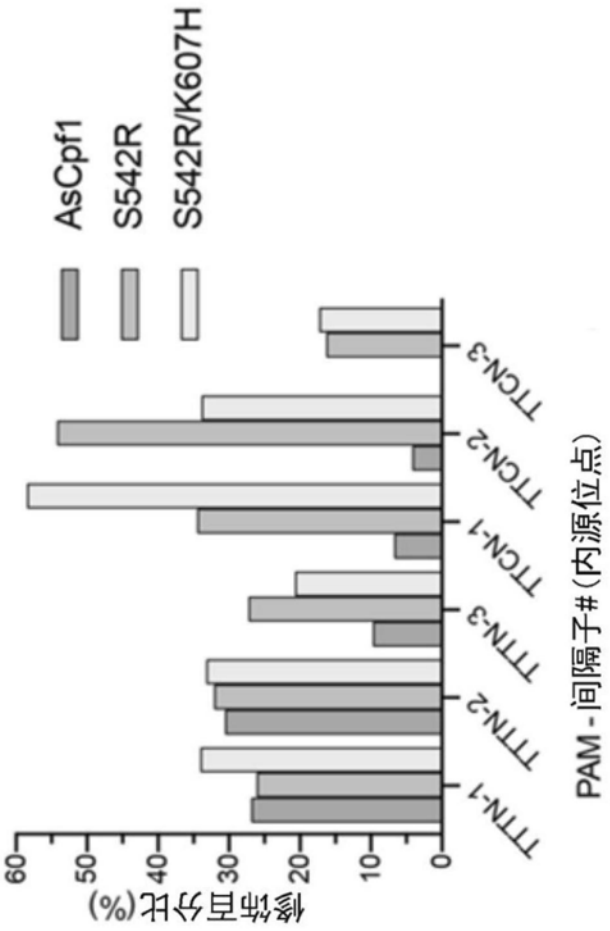

AsCpf1
S542R
S542R/K607H
修饰百分比(%)
0
10
20
30
40
50
60
TTTN-1
TTTN-2
TTTN-3
TTCN-1
TTCN-2
TTCN-3
PAM - 间隔子#(内源位点)

图5H

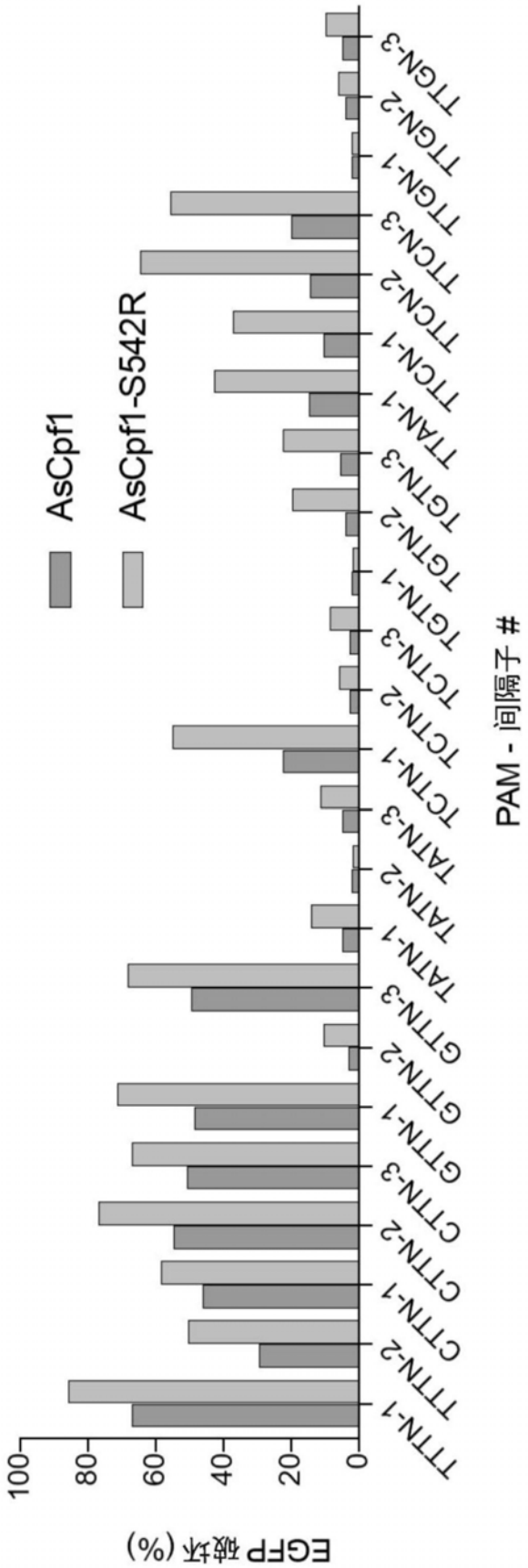

AsCpf1
AsCpf1-S542R
EGFP 破坏 (%)
0
20
40
60
80
100
PAM - 间隔子 #
TTTN-1
TTTN-2
TTTN-3
CTTN-1
CTTN-2
CTTN-3
GTTN-1
GTTN-2
GTTN-3
TATN-1
TATN-2
TATN-3
TCTN-1
TCTN-2
TCTN-3
TGTN-1
TGTN-2
TGTN-3
TTAN-1
TTCN-1
TTCN-2
TTCN-3
TTGN-1
TTGN-2
TTGN-3

图6A

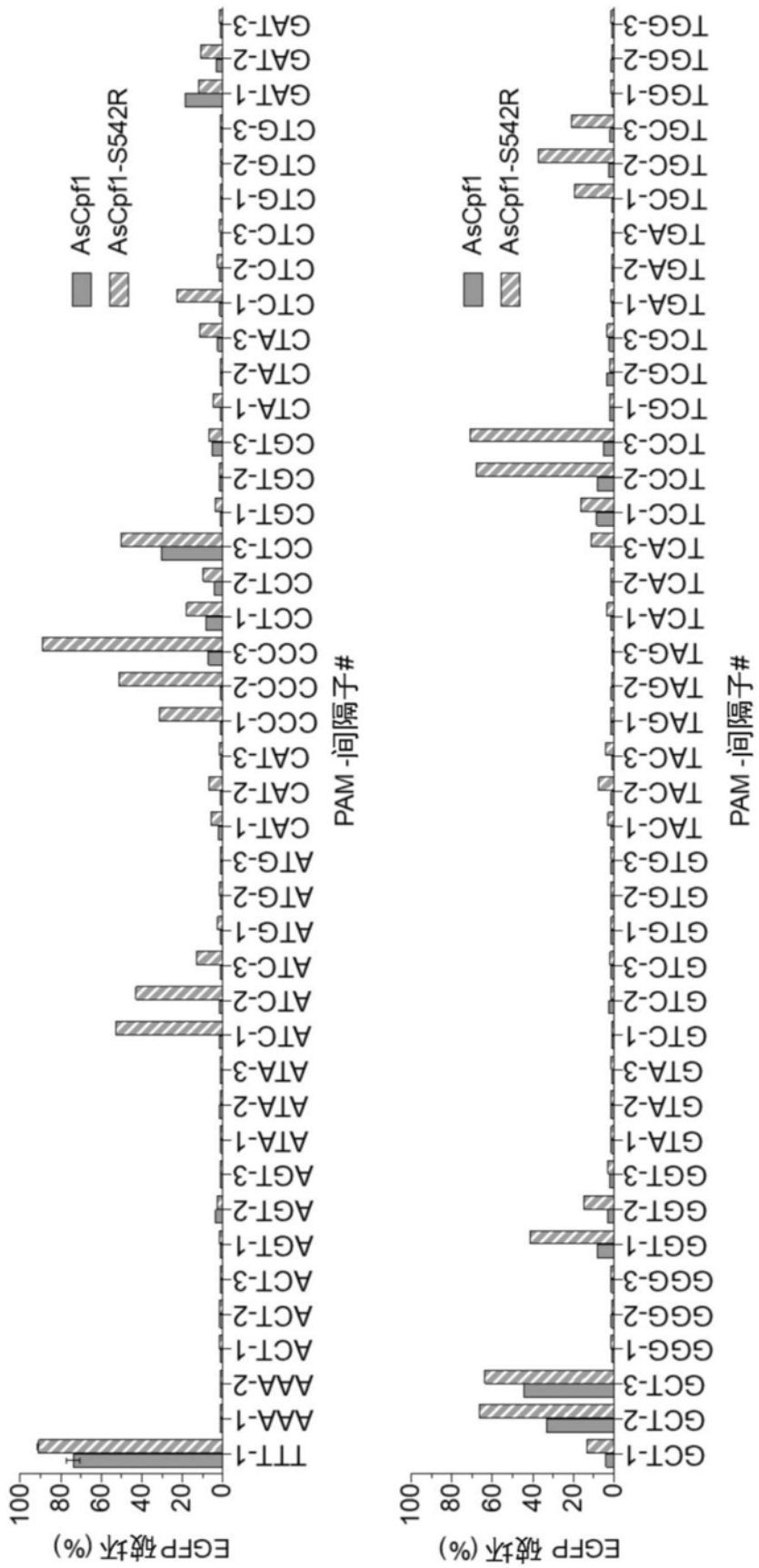

AsCpf1
AsCpf1-S542R
EGFP破坏(%)
PAM-间隔子#
TTT-1
AAA-1
AAA-2
ACT-1
ACT-2
ACT-3
AGT-1
AGT-2
AGT-3
ATA-1
ATA-2
ATA-3
ATC-1
ATC-2
ATC-3
ATG-1
ATG-2
ATG-3
CAT-1
CAT-2
CAT-3
CCC-1
CCC-2
CCC-3
CCT-1
CCT-2
CCT-3
CGT-1
CGT-2
CGT-3
CTA-1
CTA-2
CTA-3
CTC-1
CTC-2
CTC-3
CTG-1
CTG-2
CTG-3
GAT-1
GAT-2
GAT-3
GCT-1
GCT-2
GCT-3
GGG-1
GGG-2
GGG-3
GGT-1
GGT-2
GGT-3
GTA-1
GTA-2
GTA-3
GTC-1
GTC-2
GTC-3
GTG-1
GTG-2
GTG-3
TAC-1
TAC-2
TAC-3
TAG-1
TAG-2
TAG-3
TCA-1
TCA-2
TCA-3
TCC-1
TCC-2
TCC-3
TCG-1
TCG-2
TCG-3
TGA-1
TGA-2
TGA-3
TGC-1
TGC-2
TGC-3
TGG-1
TGG-2
TGG-3
0
20
40
60
80
100

图6B

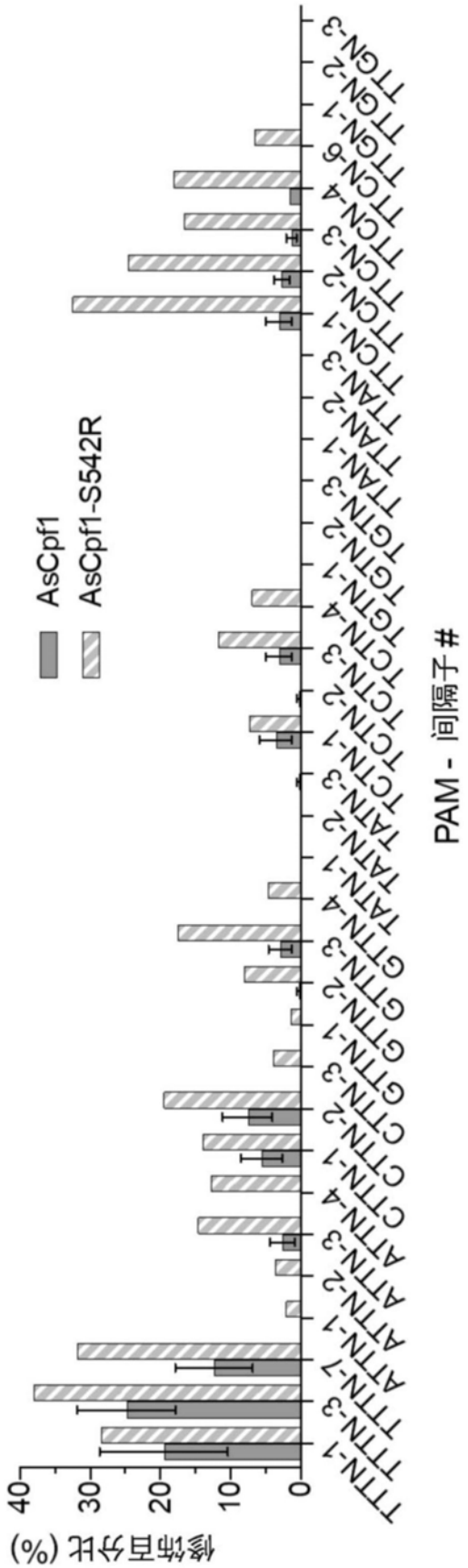

修饰百分比 (%)
0
10
20
30
40
AsCpf1
AsCpf1-S542R
PAM - 间隔子 #
TTTN-1
TTTN-3
TTTN-7
ATTN-1
ATTN-2
ATTN-3
ATTN-4
CTTN-1
CTTN-2
CTTN-3
GTTN-1
GTTN-2
GTTN-3
GTTN-4
TATN-1
TATN-2
TCTN-1
TCTN-2
TCTN-3
TCTN-4
TGTN-1
TGTN-2
TGTN-3
TTAN-1
TTAN-2
TTAN-3
TTCN-1
TTCN-2
TTCN-3
TTCN-4
TTCN-6
TTGN-1
TTGN-2
TTGN-3

图7A

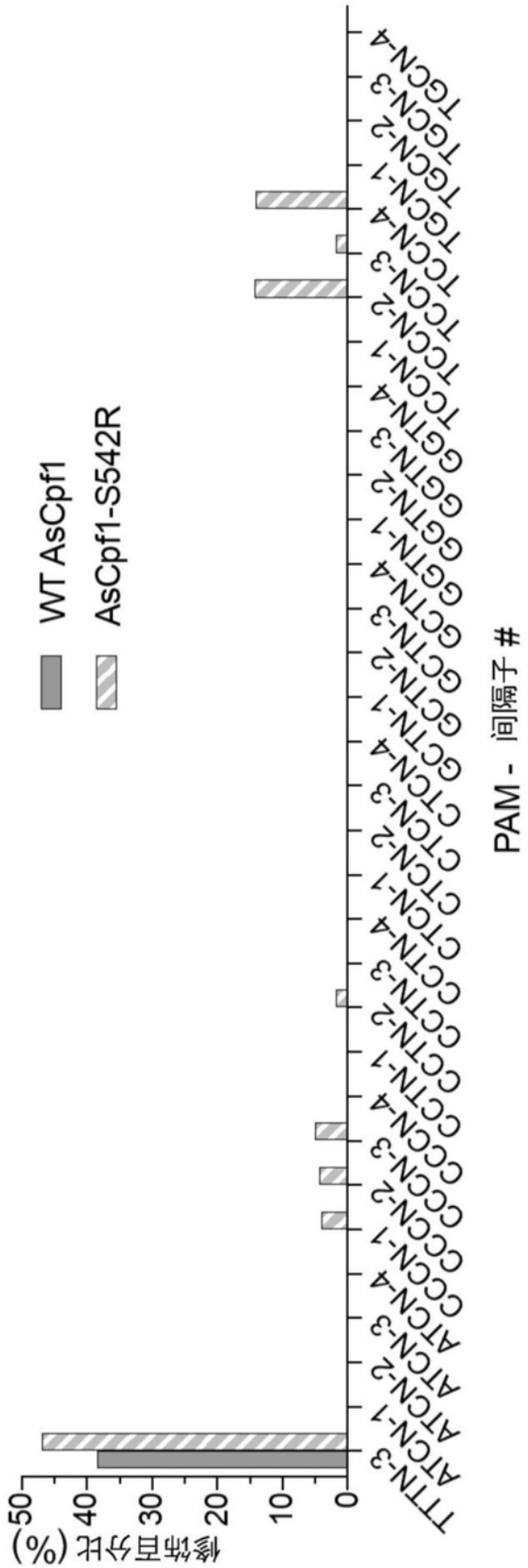

WT AsCpf1
AsCpf1-S542R
修饰百分比 (%)
0
10
20
30
40
50
PAM - 间隔子 #
TTTN-3
ATCN-1
ATCN-2
ATCN-3
ATCN-4
CCCN-1
CCCN-2
CCCN-3
CCCN-4
CCTN-1
CCTN-2
CCTN-3
CCTN-4
CTCN-1
CTCN-2
CTCN-3
CTCN-4
GCTN-1
GCTN-2
GCTN-3
GCTN-4
GGTN-1
GGTN-2
GGTN-3
GGTN-4
TCCN-1
TCCN-2
TCCN-3
TCCN-4
TGCN-1
TGCN-2
TGCN-3
TGCN-4

图7B

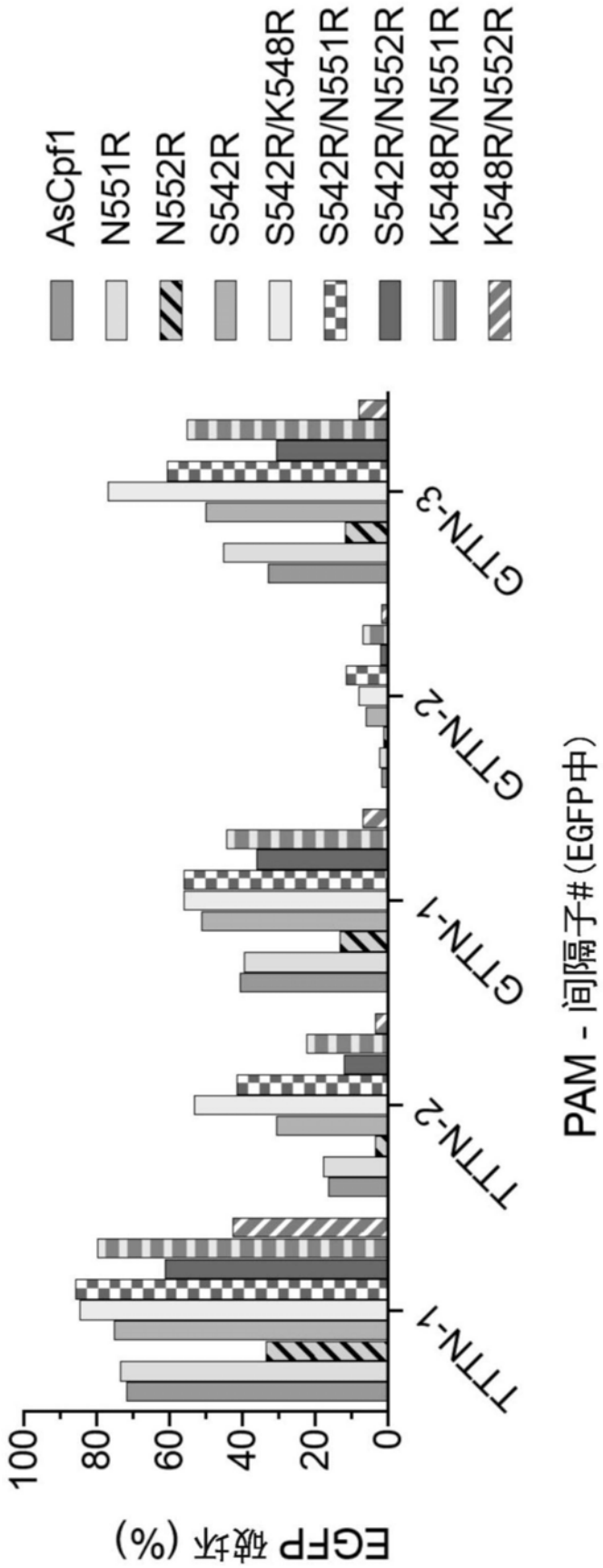

AsCpf1
N551R
N552R
S542R
S542R/K548R
S542R/N551R
S542R/N552R
K548R/N551R
K548R/N552R
EGFP 破坏 (%)
100
80
60
40
20
0
TTTN-1
TTTN-2
GTTN-1
GTTN-2
GTTN-3
PAM - 间隔子#(EGFP中)

图8A

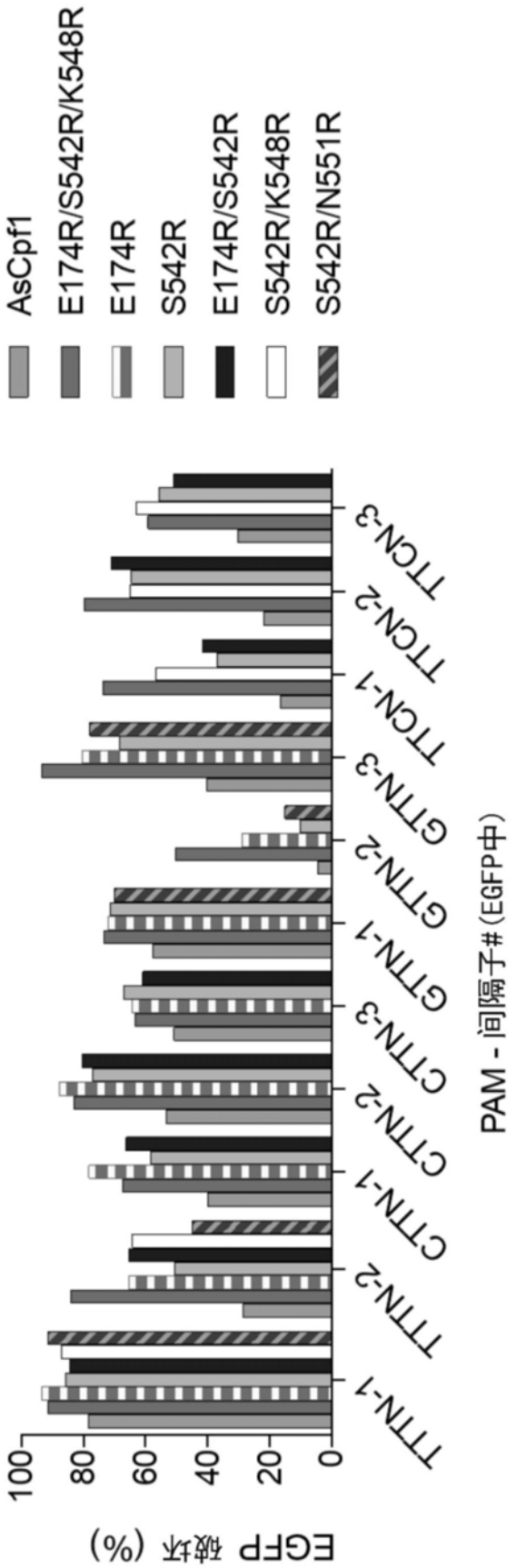

AsCpf1
E174R/S542R/K548R
E174R
S542R
E174R/S542R
S542R/K548R
S542R/N551R
EGFP 破坏 (%)
100
80
60
40
20
0
TTTN-1
TTTN-2
CTTN-1
CTTN-2
CTTN-3
GTTN-1
GTTN-2
GTTN-3
TTCN-1
TTCN-2
TTCN-3
PAM - 间隔子#(EGFP中)

图8B

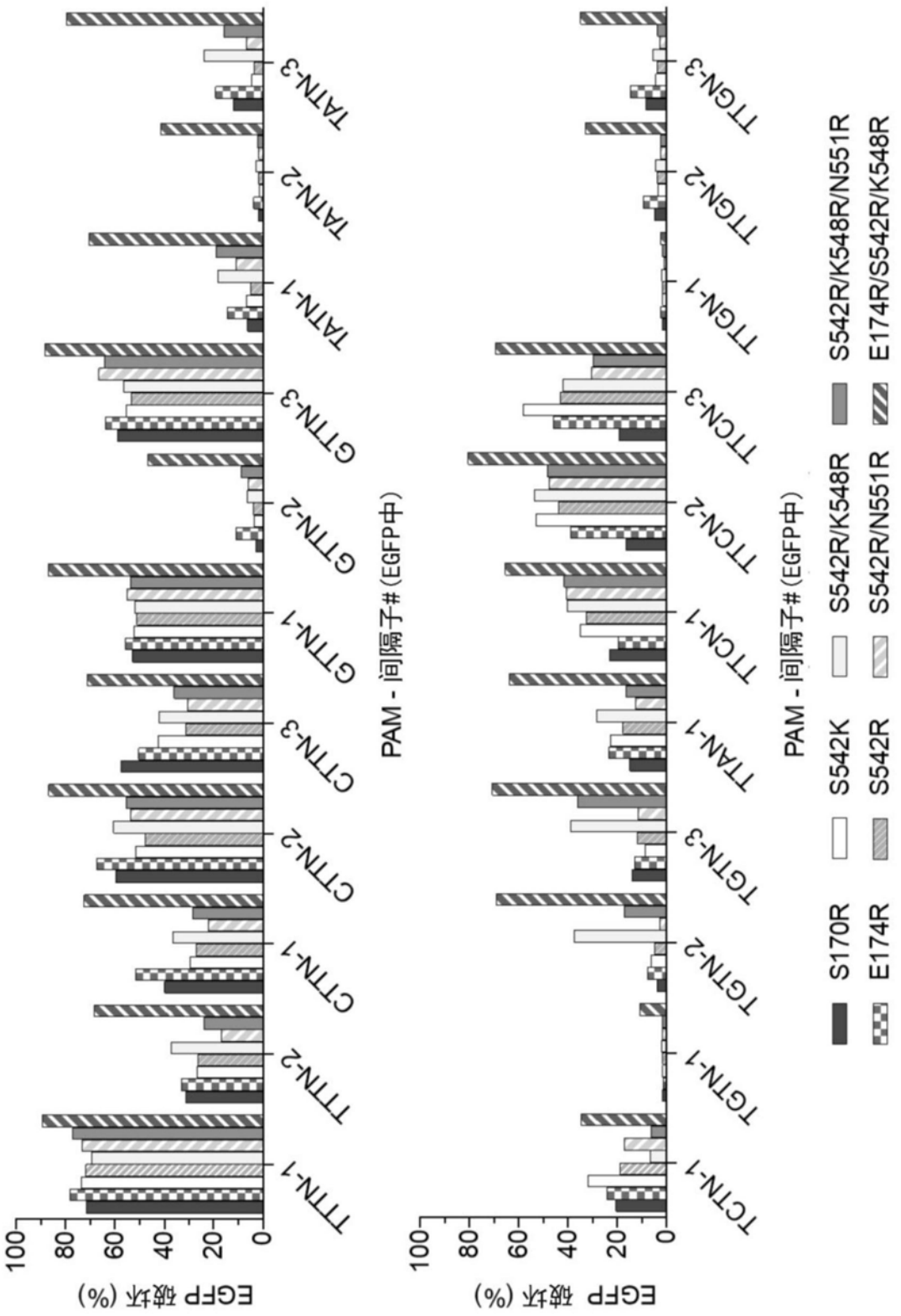

EGFP 破坏 (%)
100
80
60
40
20
0
TTTN-1
TTTN-2
CTTN-1
CTTN-2
CTTN-3
GTTN-1
GTTN-2
GTTN-3
TATN-1
TATN-2
TATN-3
PAM - 间隔子#(EGFP中)
EGFP 破坏 (%)
100
80
60
40
20
0
TCTN-1
TGTN-1
TGTN-2
TGTN-3
TTAN-1
TTCN-1
TTCN-2
TTCN-3
TTGN-1
TTGN-2
TTGN-3
PAM - 间隔子#(EGFP中)
S170R
E174R
S542K
S542R
S542R/K548R
S542R/N551R
S542R/K548R/N551R
E174R/S542R/K548R

图8C

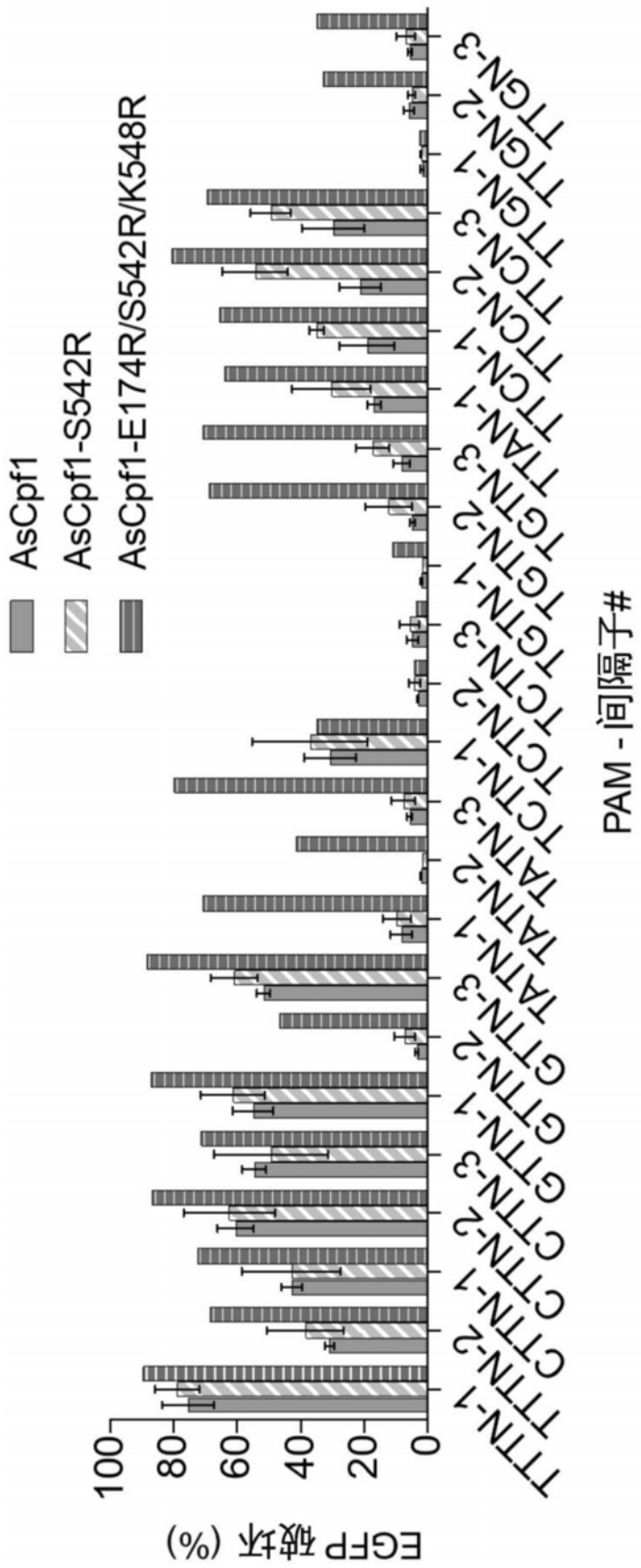

AsCpf1
AsCpf1-S542R
AsCpf1-E174R/S542R/K548R
EGFP 破坏 (%)
0
20
40
60
80
100
TTTN-1
TTTN-2
TTTN-3
CTTN-1
CTTN-2
CTTN-3
GTTN-1
GTTN-2
GTTN-3
TATN-1
TATN-2
TATN-3
TCTN-1
TCTN-2
TCTN-3
TGTN-1
TGTN-2
TGTN-3
TTAN-1
TTCN-1
TTCN-2
TTCN-3
TTGN-1
TTGN-2
TTGN-3
PAM - 间隔子#

图9A

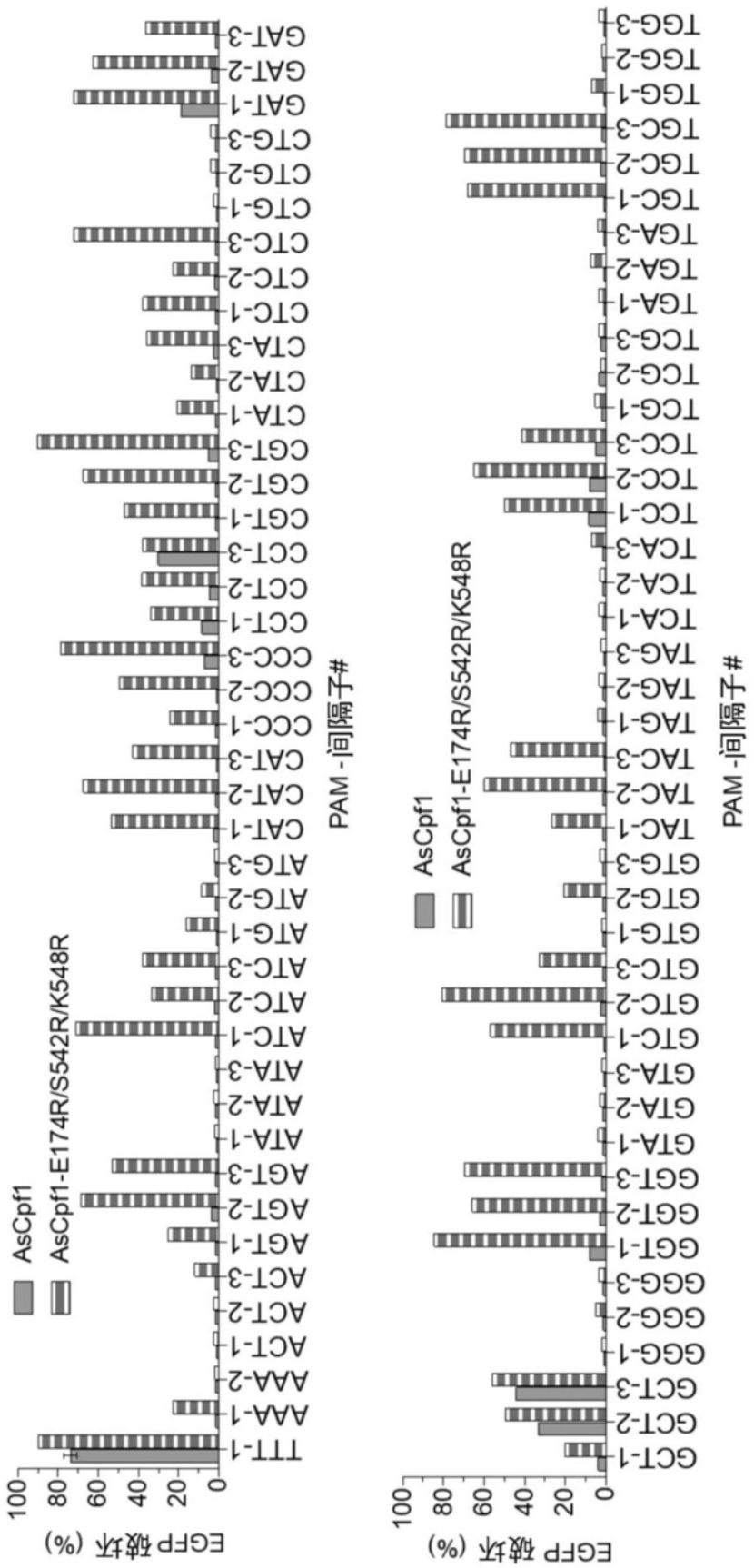

EGFP 破坏 (%)
100
80
60
40
20
0
AsCpf1
AsCpf1-E174R/S542R/K548R
PAM -间隔子#
TTT-1
AAA-1
AAA-2
ACT-1
ACT-2
ACT-3
AGT-1
AGT-2
AGT-3
ATA-1
ATA-2
ATA-3
ATC-1
ATC-2
ATC-3
ATG-1
ATG-2
ATG-3
CAT-1
CAT-2
CAT-3
CCC-1
CCC-2
CCC-3
CCT-1
CCT-2
CCT-3
CGT-1
CGT-2
CGT-3
CTA-1
CTA-2
CTA-3
CTC-1
CTC-2
CTC-3
CTG-1
CTG-2
CTG-3
GAT-1
GAT-2
GAT-3
GCT-1
GCT-2
GCT-3
GGG-1
GGG-2
GGG-3
GGT-1
GGT-2
GGT-3
GTA-1
GTA-2
GTA-3
GTC-1
GTC-2
GTC-3
GTG-1
GTG-2
GTG-3
TAC-1
TAC-2
TAC-3
TAG-1
TAG-2
TAG-3
TCA-1
TCA-2
TCA-3
TCC-1
TCC-2
TCC-3
TCG-1
TCG-2
TCG-3
TGA-1
TGA-2
TGA-3
TGC-1
TGC-2
TGC-3
TGG-1
TGG-2
TGG-3

图9B

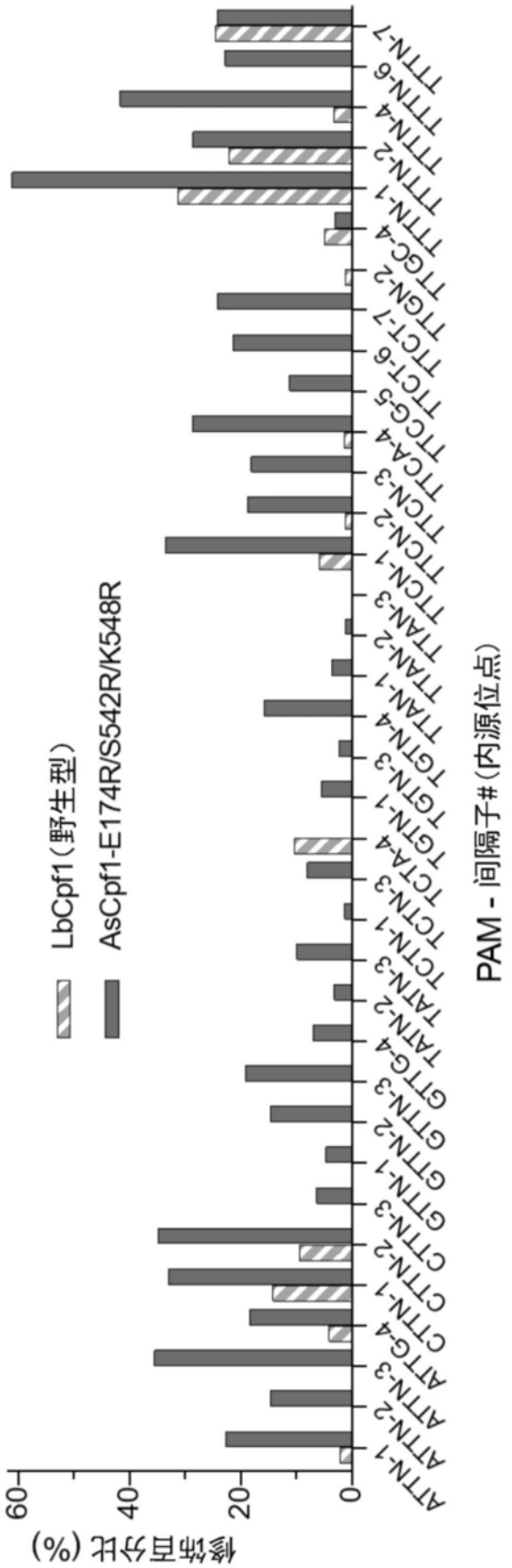

修饰百分比 (%)
0
20
40
60
LbCpf1(野生型)
AsCpf1-E174R/S542R/K548R
PAM - 间隔子#(内源位点)
ATTN-1
ATTN-2
ATTN-3
ATTG-4
CTTN-1
CTTN-2
CTTN-3
GTTN-1
GTTN-2
GTTN-3
GTTG-4
TATN-2
TATN-3
TCTN-1
TCTN-3
TCTA-4
TGTN-1
TGTN-3
TGTN-4
TTAN-1
TTAN-2
TTAN-3
TTCN-1
TTCN-2
TTCN-3
TTCA-4
TTCG-5
TTCT-6
TTCT-7
TTGN-2
TTGC-4
TTTN-1
TTTN-2
TTTN-4
TTTN-6
TTTN-7

图10A

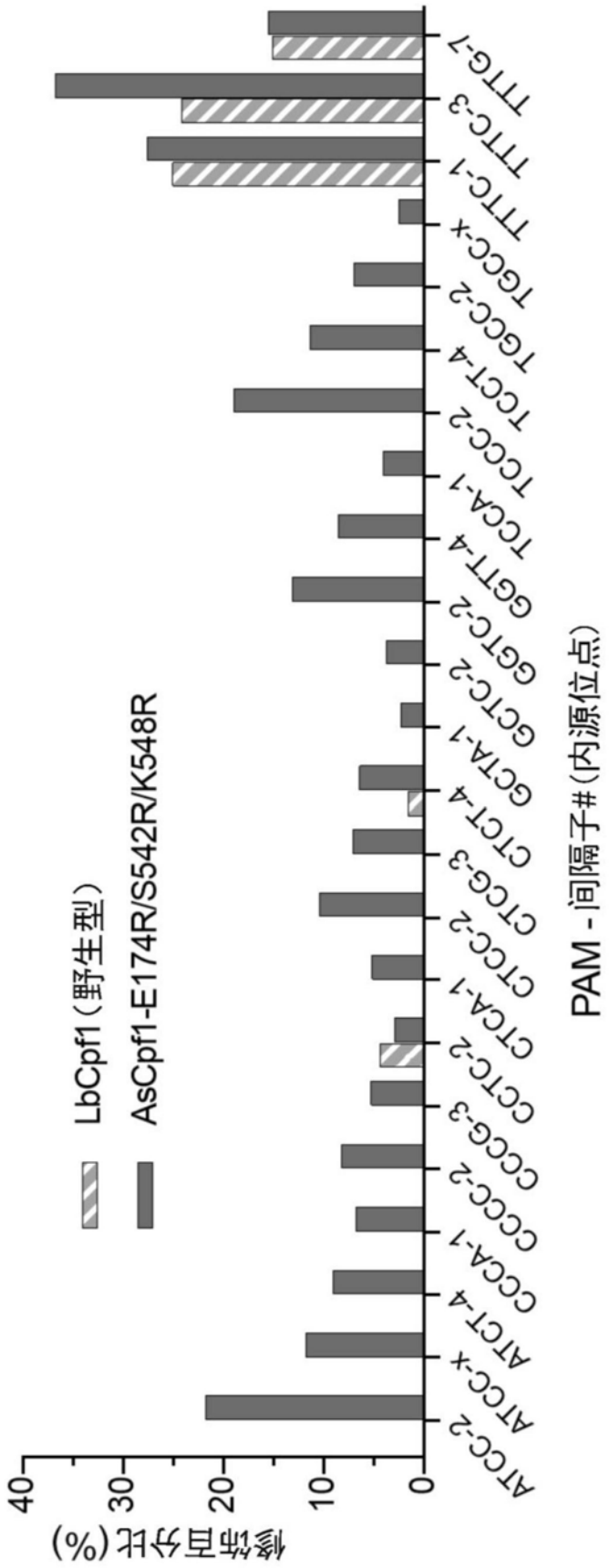

修饰百分比(%)
40
30
20
10
0
LbCpf1(野生型)
AsCpf1-E174R/S542R/K548R
PAM - 间隔子#(内源位点)
ATCC-2
ATCC-x
ATCT-4
CCCA-1
CCCC-2
CCCG-3
CCTC-2
CTCA-1
CTCC-2
CTCG-3
CTCT-4
GCTA-1
GCTC-2
GGTC-2
GGTT-4
TCCA-1
TCCC-2
TCCT-4
TGCC-2
TTTC-x
TTTC-1
TTTC-3
TTTG-7

图10B

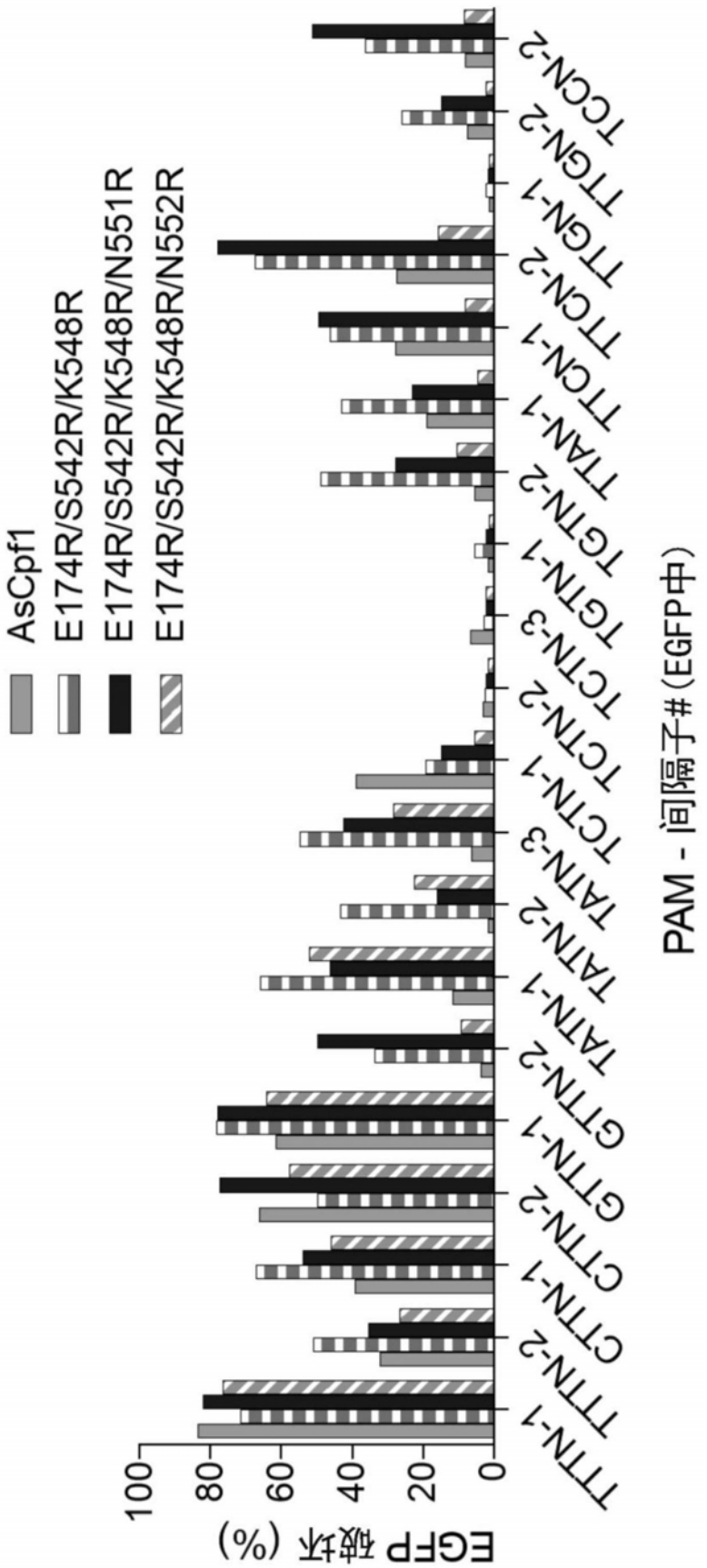

AsCpf1
E174R/S542R/K548R
E174R/S542R/K548R/N551R
E174R/S542R/K548R/N552R
EGFP 破坏 (%)
100
80
60
40
20
0
PAM - 间隔子#(EGFP中)
TTTN-1
TTTN-2
CTTN-1
CTTN-2
GTTN-1
GTTN-2
TATN-1
TATN-2
TATN-3
TCTN-1
TCTN-2
TCTN-3
TGTN-1
TGTN-2
TTAN-1
TTCN-1
TTCN-2
TTGN-1
TTGN-2
TCCN-2

图11A

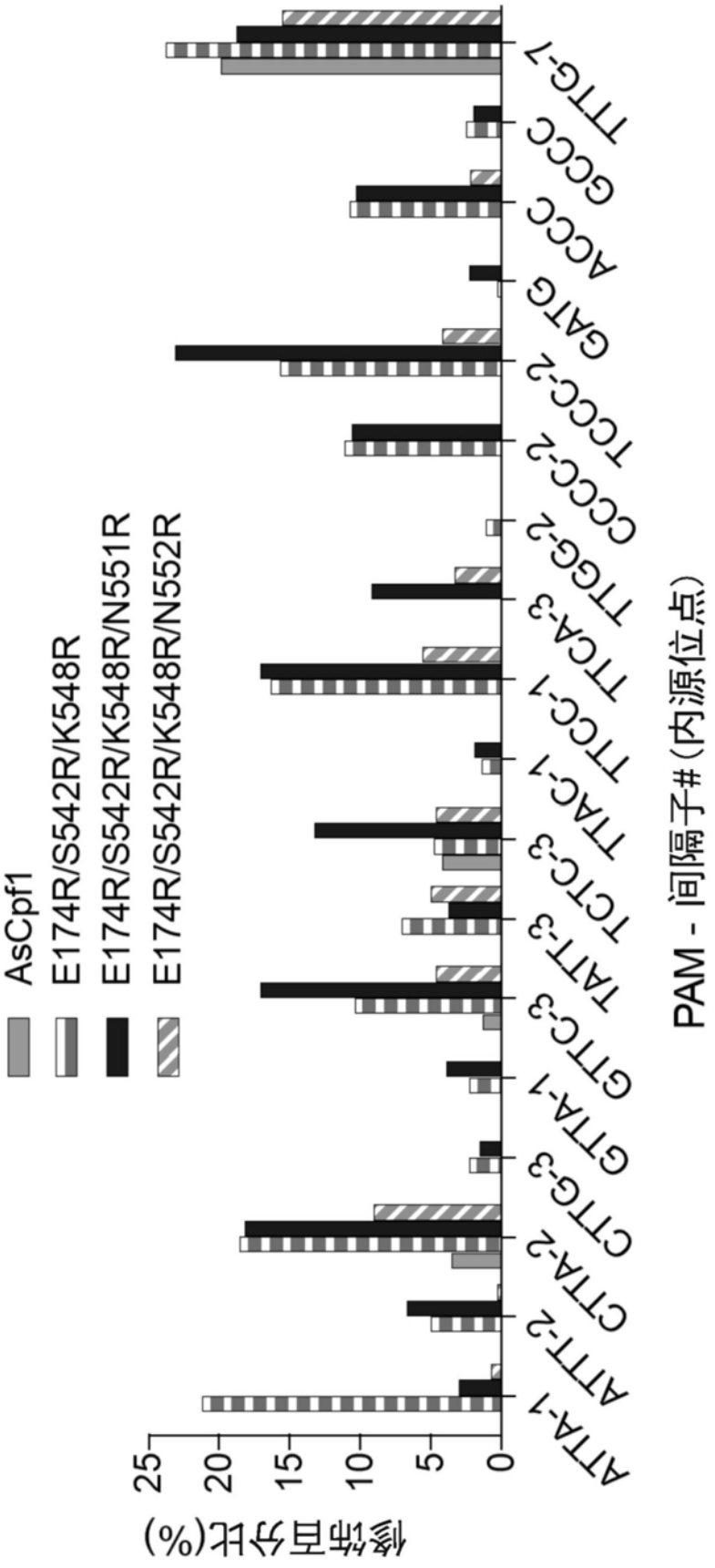

修饰百分比(%)
25
20
15
10
5
0
ATTA-1
ATTT-2
CTTA-2
CTTG-3
GTTA-1
GTTC-3
TATT-3
TCTC-3
TTAC-1
TTCC-1
TTCA-3
TTGG-2
CCCC-2
TCCC-2
GATG
ACCC
GCCC
TTTG-7
PAM - 间隔子#(内源位点)
AsCpf1
E174R/S542R/K548R
E174R/S542R/K548R/N551R
E174R/S542R/K548R/N552R

图11B

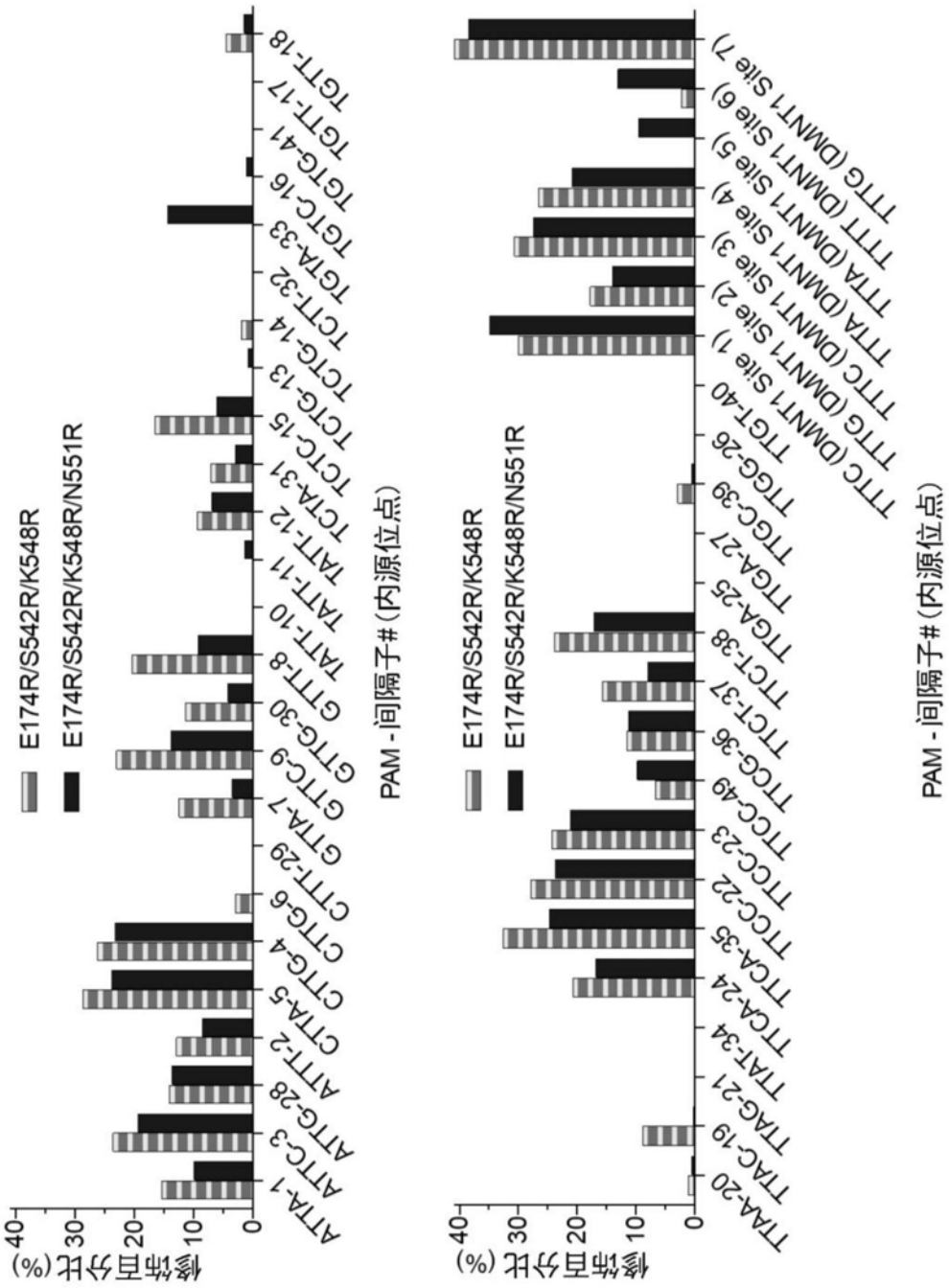

E174R/S542R/K548R
E174R/S542R/K548R/N551R
修饰百分比(%)
0
10
20
30
40
PAM - 间隔子#(内源位点)
ATTA-1
ATTC-3
ATTG-28
ATTT-2
CTTA-5
CTTG-4
CTTT-6
CTTT-29
GTTA-7
GTTC-9
GTTG-30
GTTT-8
TATT-10
TATT-11
TATT-12
TCTA-31
TCTC-15
TCTG-13
TCTG-14
TCTT-32
TGTA-33
TGTC-16
TGTG-41
TGTT-17
TGTT-18
TTAA-20
TTAC-19
TTAG-21
TTAT-34
TTCA-24
TTCA-35
TTCC-22
TTCC-23
TTCC-49
TTCG-36
TTCT-37
TTCT-38
TTGA-25
TTGA-27
TTGC-39
TTGG-26
TTGT-40
TTTC (DMNT1 Site 1)
TTTG (DMNT1 Site 2)
TTTC (DMNT1 Site 3)
TTTA (DMNT1 Site 4)
TTTT (DMNT1 Site 5)
TTTA (DMNT1 Site 6)
TTTG (DMNT1 Site 7)

图11C

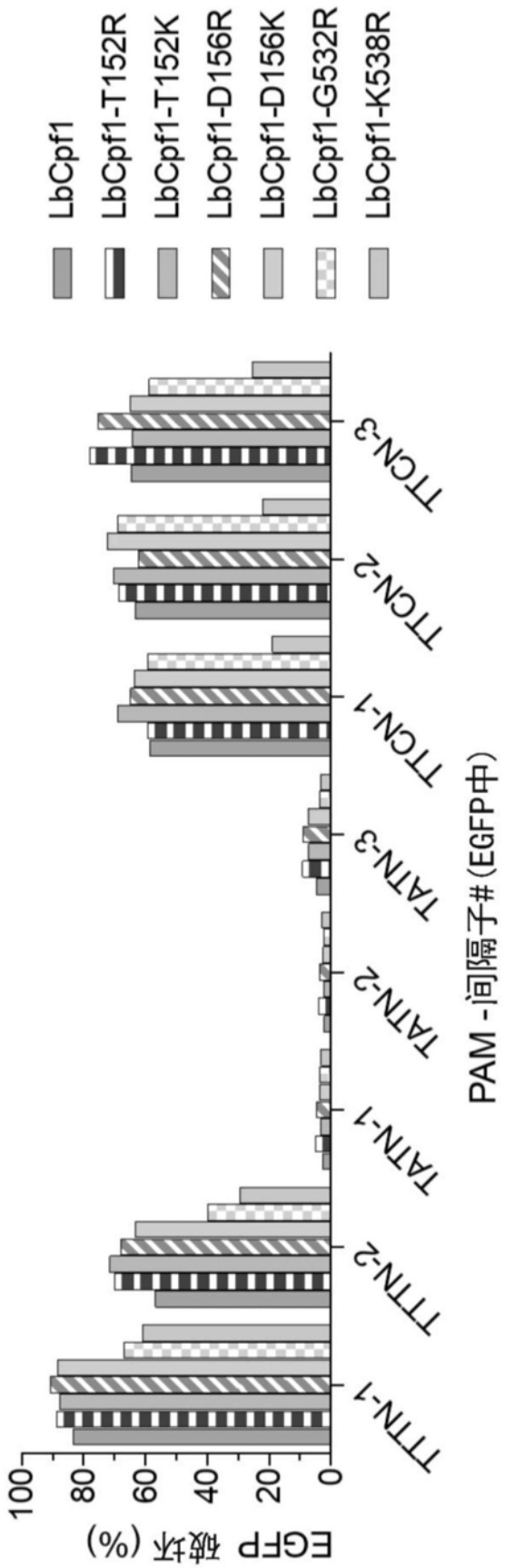

LbCpf1
LbCpf1-T152R
LbCpf1-T152K
LbCpf1-D156R
LbCpf1-D156K
LbCpf1-G532R
LbCpf1-K538R
EGFP 破坏 (%)
0
20
40
60
80
100
TTTN-1
TTTN-2
TATN-1
TATN-2
TATN-3
TTCN-1
TTCN-2
TTCN-3
PAM -间隔子#(EGFP中)

图12A

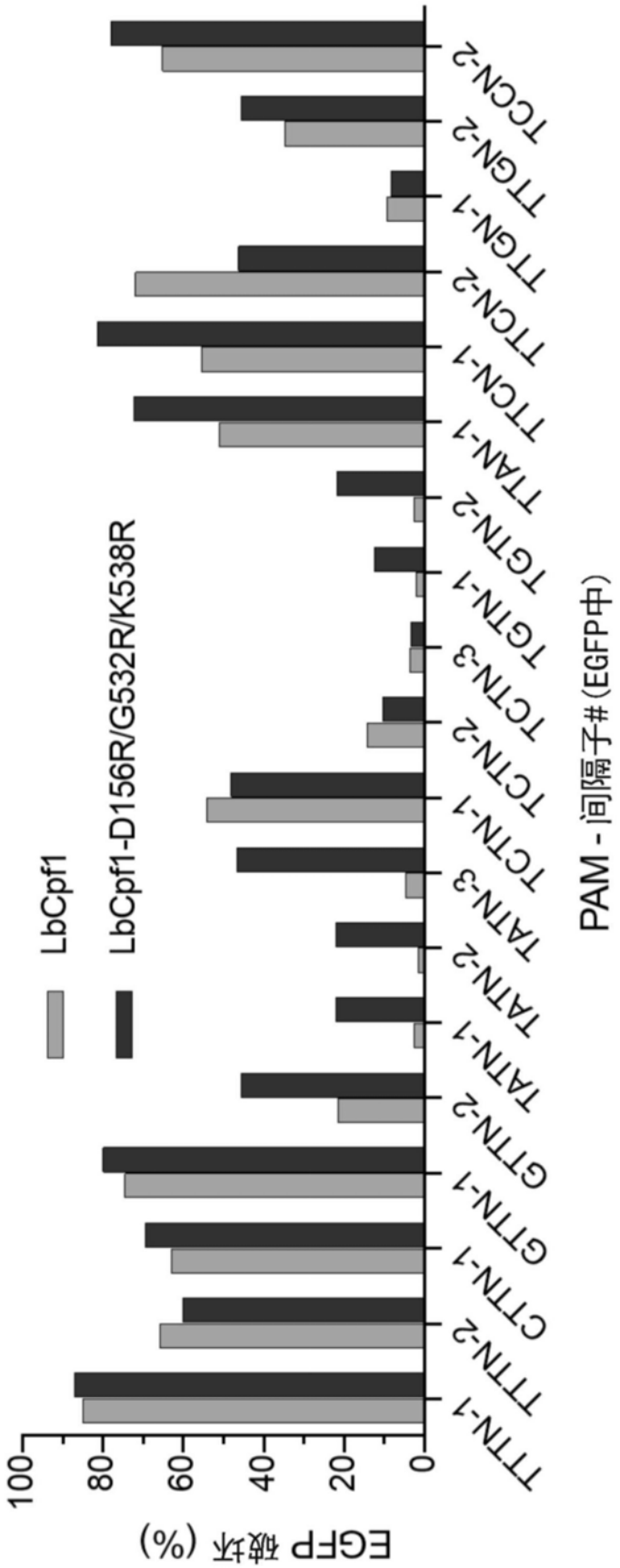

EGFP 破坏 (%)
100
80
60
40
20
0
LbCpf1
LbCpf1-D156R/G532R/K538R
PAM - 间隔子#(EGFP中)
TTTN-1
TTTN-2
CTTN-1
CTTN-2
GTTN-1
GTTN-2
TATN-1
TATN-2
TATN-3
TCTN-1
TCTN-2
TCTN-3
TGTN-1
TGTN-2
TTAN-1
TTCN-1
TTCN-2
TTGN-1
TTGN-2
TCCN-2

图12B

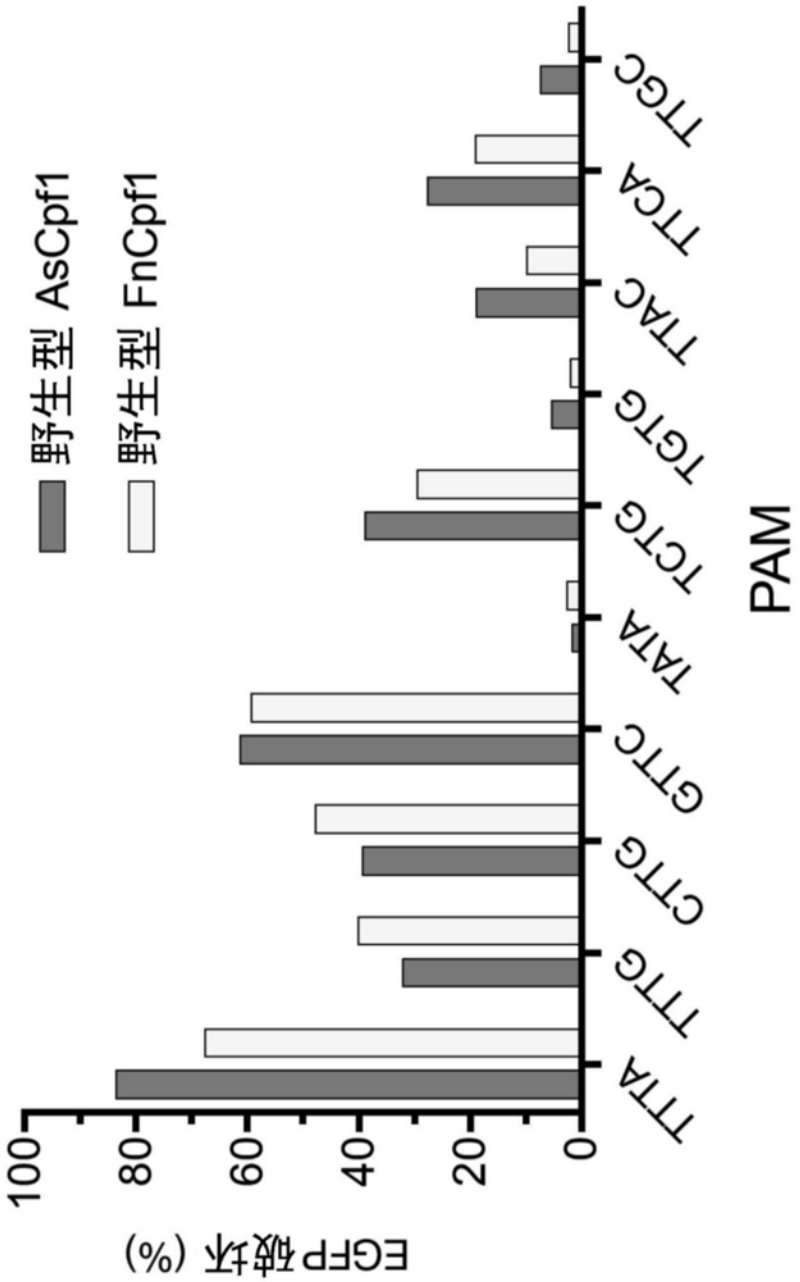

野生型 AsCpf1
野生型 FnCpf1
EGFP破坏 (%)
100
80
60
40
20
0
TTTA
TTTG
CTTG
GTTC
TATA
TCTG
TGTG
TTAC
TTCA
TTGC
PAM

图13A

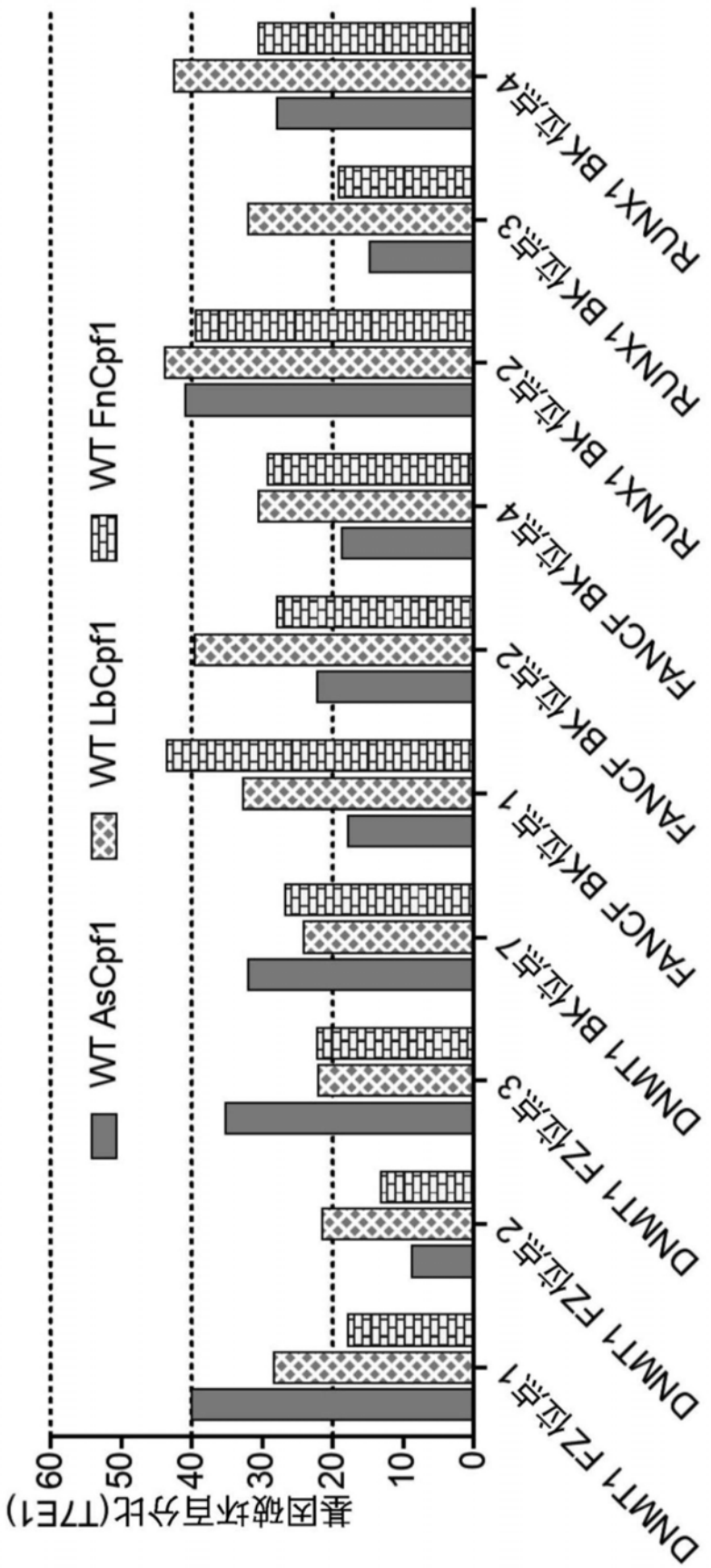

基因破坏百分比(T7E1)
0
10
20
30
40
50
60
WT AsCpf1
WT LbCpf1
WT FnCpf1
DNMT1 FZ位点1
DNMT1 FZ位点2
DNMT1 FZ位点3
DNMT1 BK位点7
FANCF BK位点1
FANCF BK位点2
FANCF BK位点4
RUNX1 BK位点2
RUNX1 BK位点3
RUNX1 BK位点4

图13B

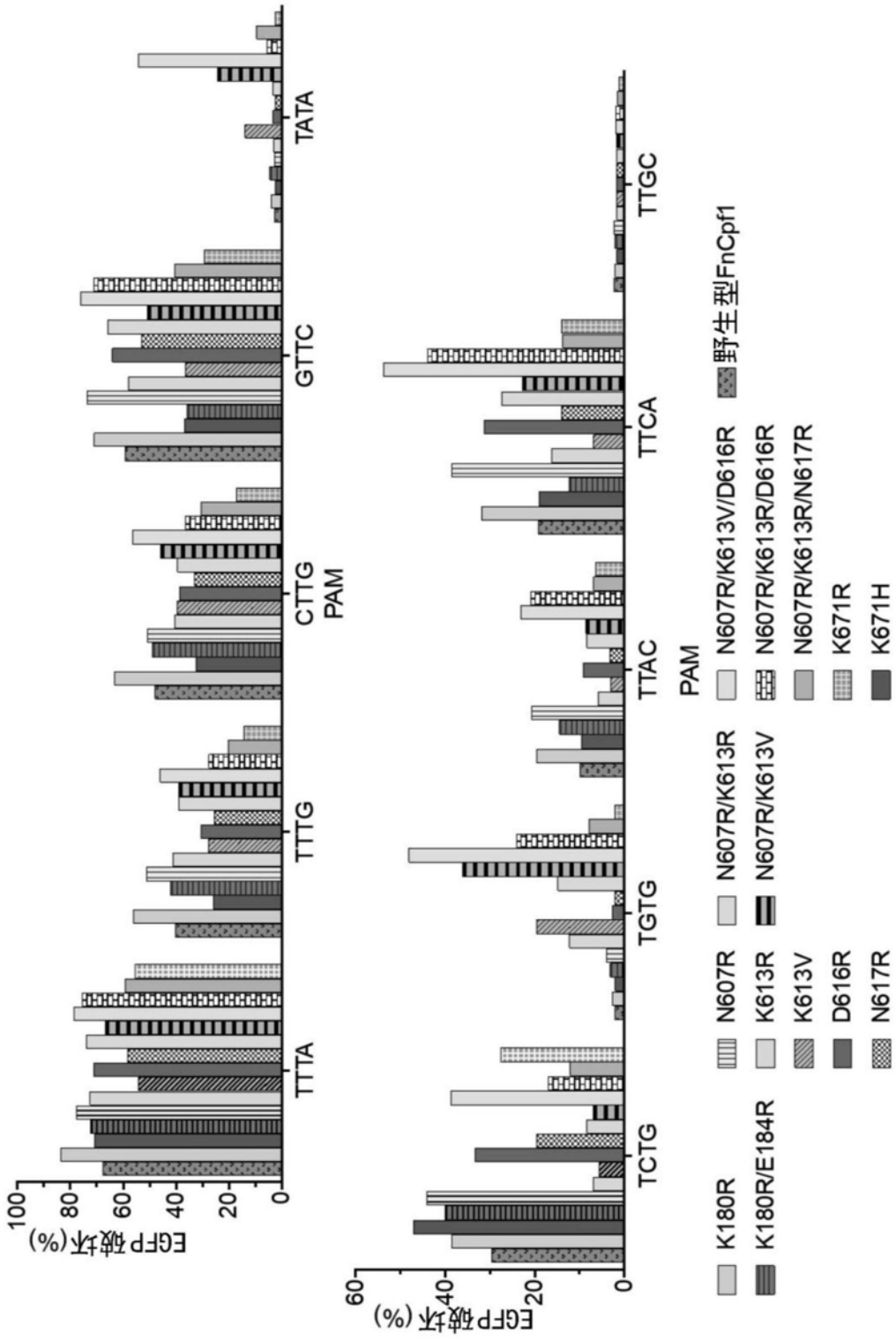

EGFP破坏(%)
0
20
40
60
80
100
TTTA
TTTG
CTTG
GTTC
TATA
PAM
EGFP破坏(%)
0
20
40
60
TCTG
TGTG
TTAC
TTCA
TTGC
PAM
K180R
K180R/E184R
N607R
K613R
K613V
D616R
N617R
N607R/K613R
N607R/K613V
N607R/K613V/D616R
N607R/K613R/D616R
N607R/K613R/N617R
K671R
K671H
野生型FnCpf1

图13C

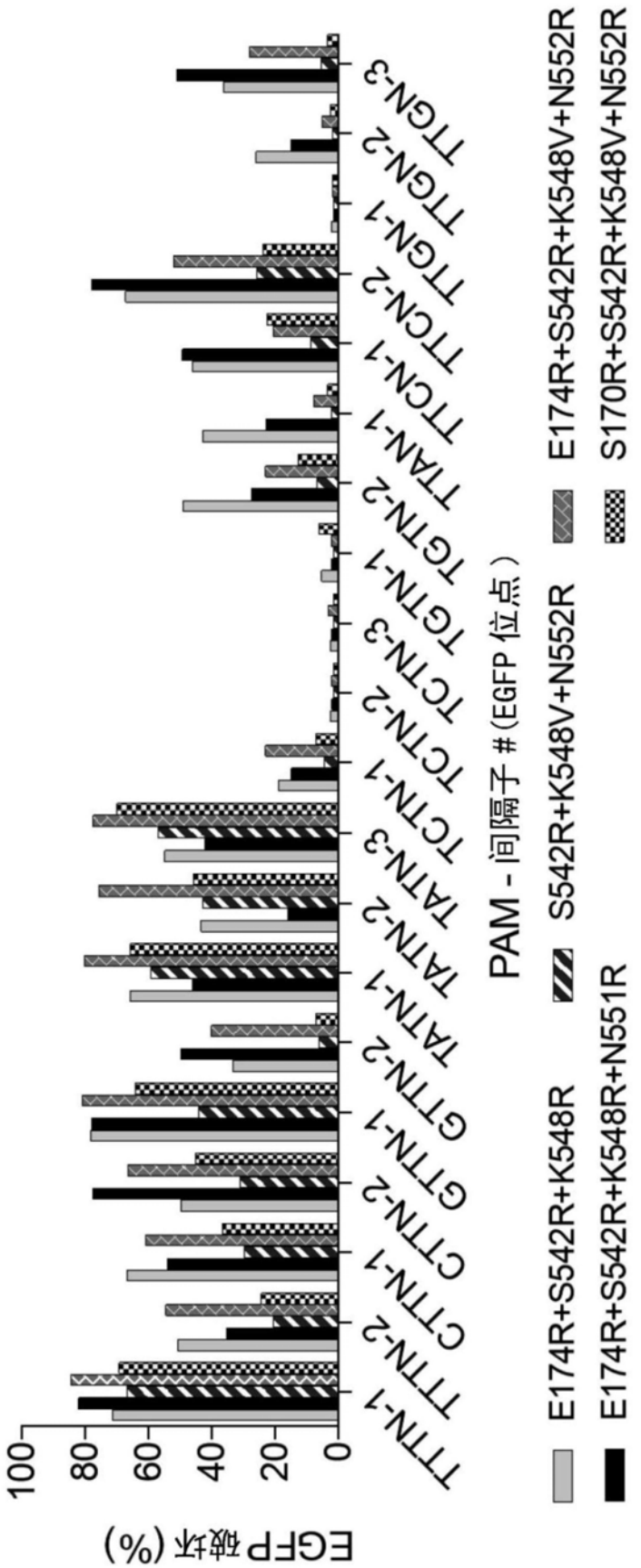

EGFP破坏(%)
100
80
60
40
20
0
TTTN-1
TTTN-2
CTTN-1
CTTN-2
GTTN-1
GTTN-2
TATN-1
TATN-2
TATN-3
TCTN-1
TCTN-2
TCTN-3
TGTN-1
TGTN-2
TTAN-1
TTCN-1
TTCN-2
TTGN-1
TTGN-2
TTGN-3
PAM - 间隔子 #(EGFP 位点)
E174R+S542R+K548R
S542R+K548V+N552R
E174R+S542R+K548V+N552R
E174R+S542R+K548R+N551R
S170R+S542R+K548V+N552R

图14A

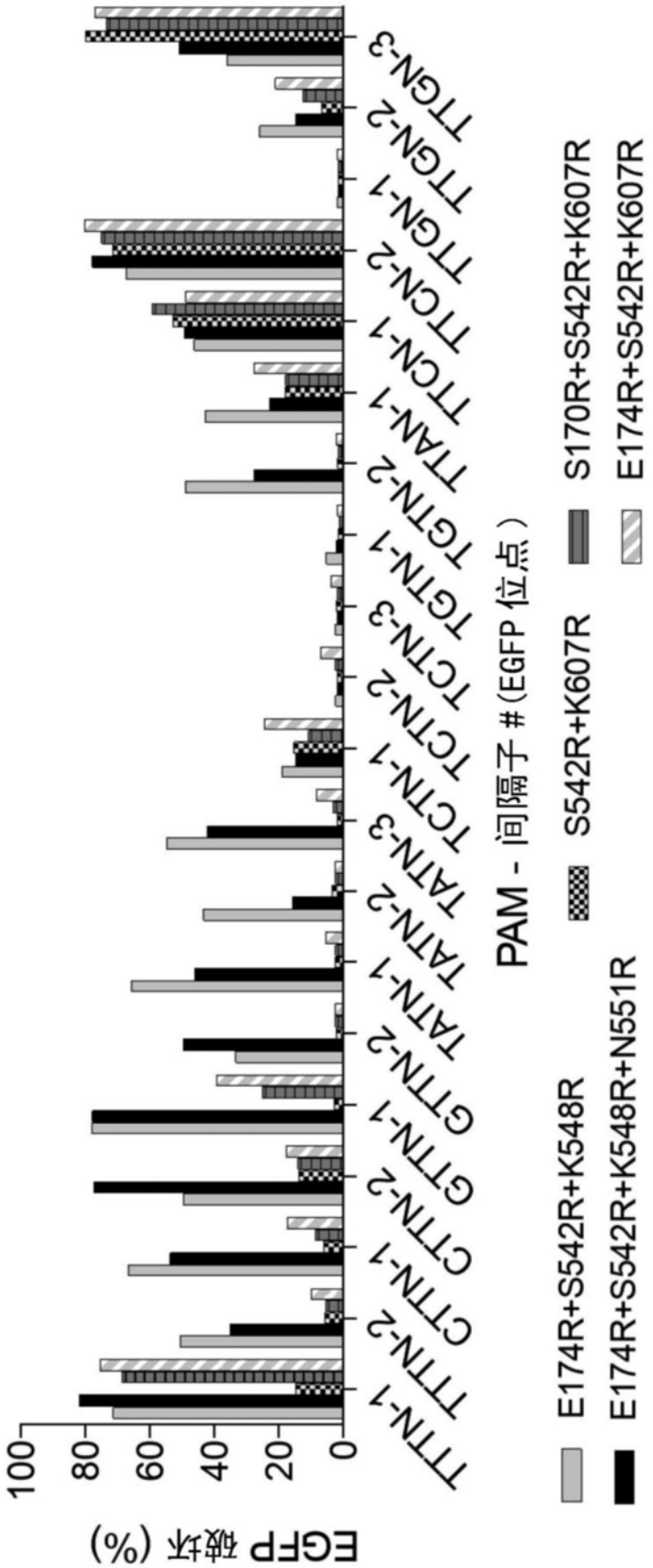

EGFP 破坏 (%)
100
80
60
40
20
0
PAM - 间隔子 #(EGFP 位点)
TTTN-1
TTTN-2
CTTN-1
CTTN-2
GTTN-1
GTTN-2
TATN-1
TATN-2
TCTN-3
TCTN-2
TCTN-1
TGTN-3
TGTN-2
TGTN-1
TTAN-1
TTCN-1
TTCN-2
TTGN-1
TTGN-2
TTGN-3
E174R+S542R+K548R
S542R+K607R
S170R+S542R+K607R
E174R+S542R+K548R+N551R
E174R+S542R+K607R

图14B

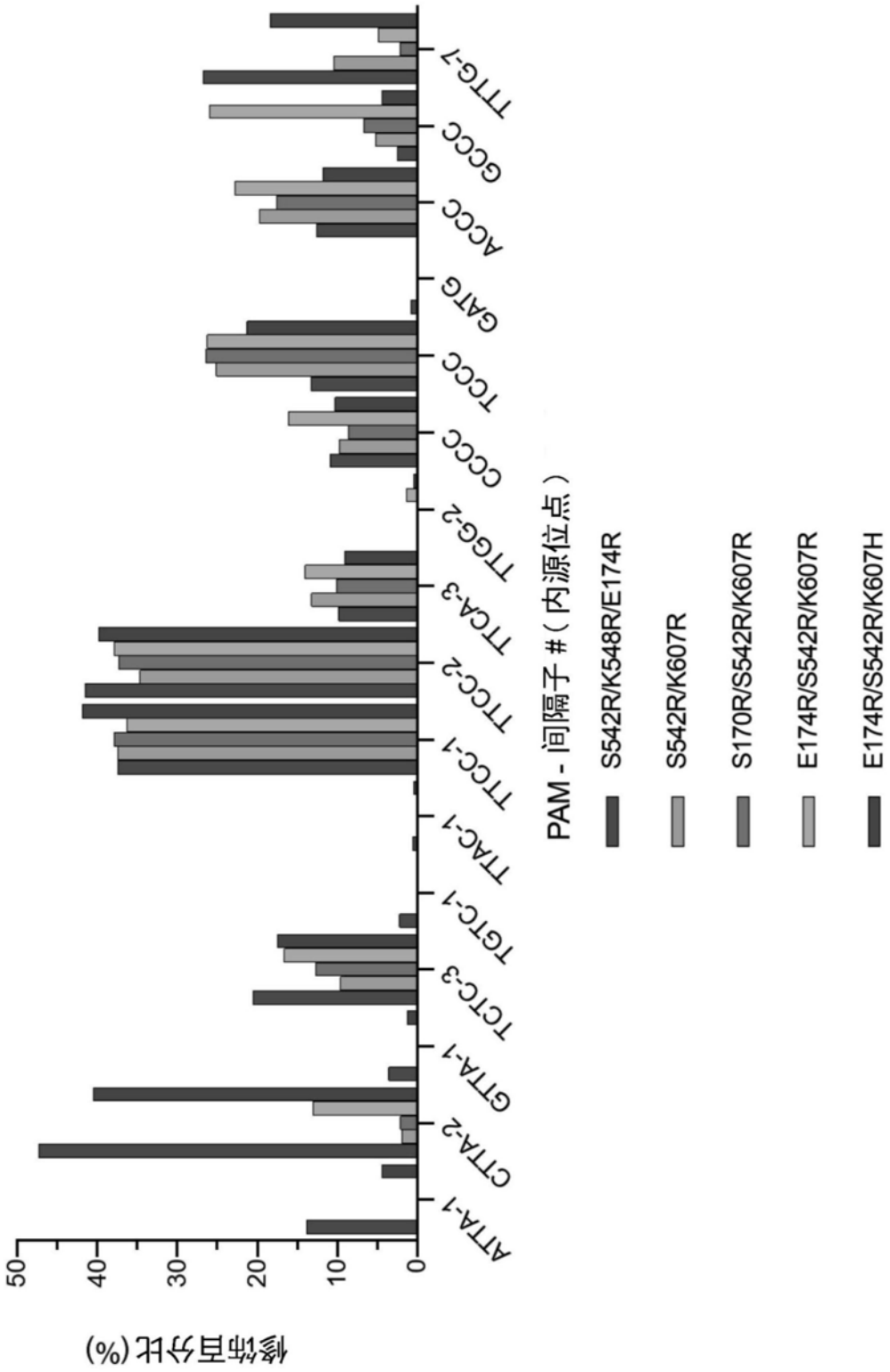

修饰百分比(%)
0
10
20
30
40
50
ATTA-1
CTTA-2
GTTA-1
TCTC-3
TGTC-1
TTAC-1
TTCC-1
TTCC-2
TTCA-3
TTGG-2
CCCC
TCCC
GATG
ACCC
GCCC
TTTG-7
PAM - 间隔子 #（内源位点）
S542R/K548R/E174R
S542R/K607R
S170R/S542R/K607R
E174R/S542R/K607R
E174R/S542R/K607H

图14C

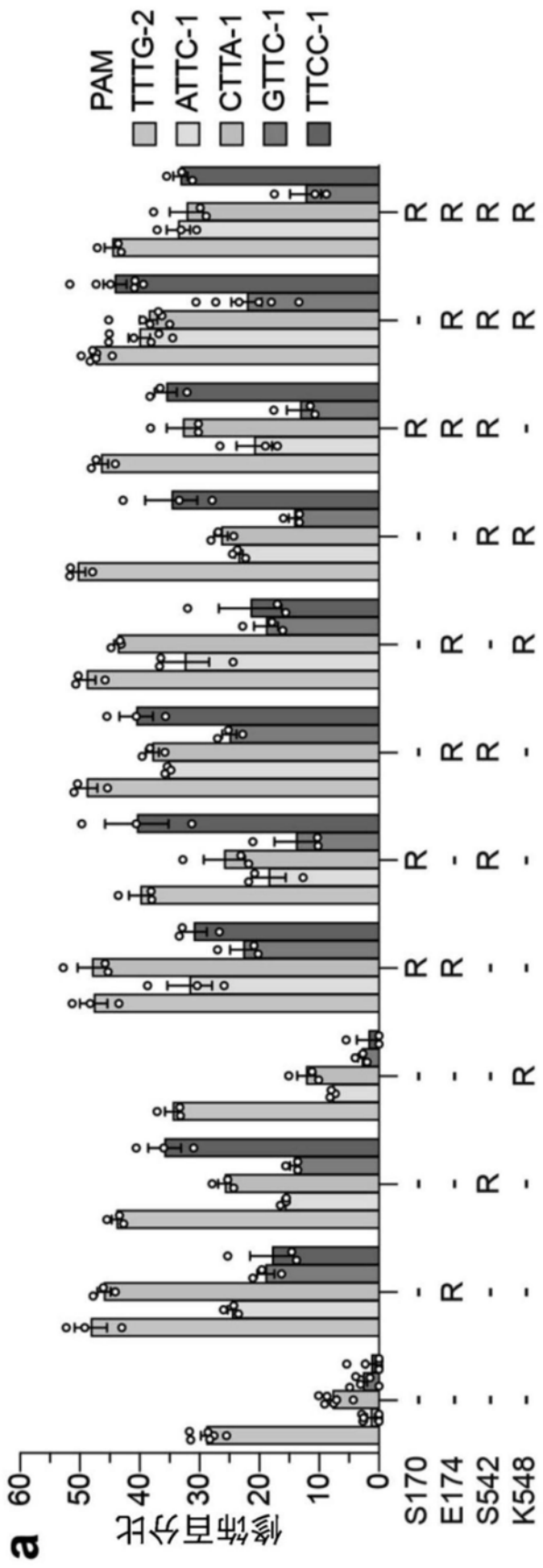

PAM
TTTG-2
ATTC-1
CTTA-1
GTTC-1
TTCC-1
a
修饰百分比
60
50
40
30
20
10
0
S170
E174
S542
K548

图15A

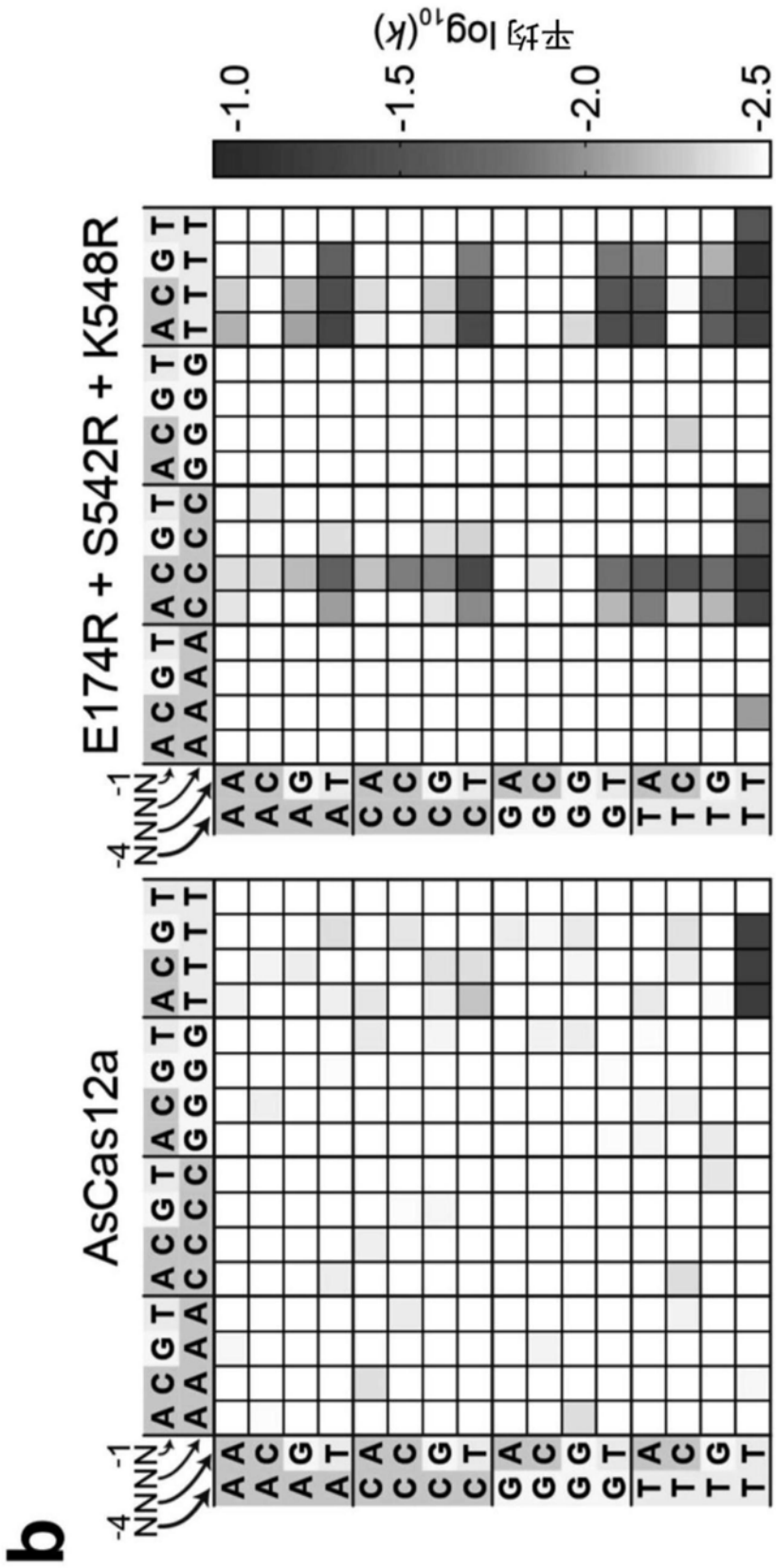


图15B

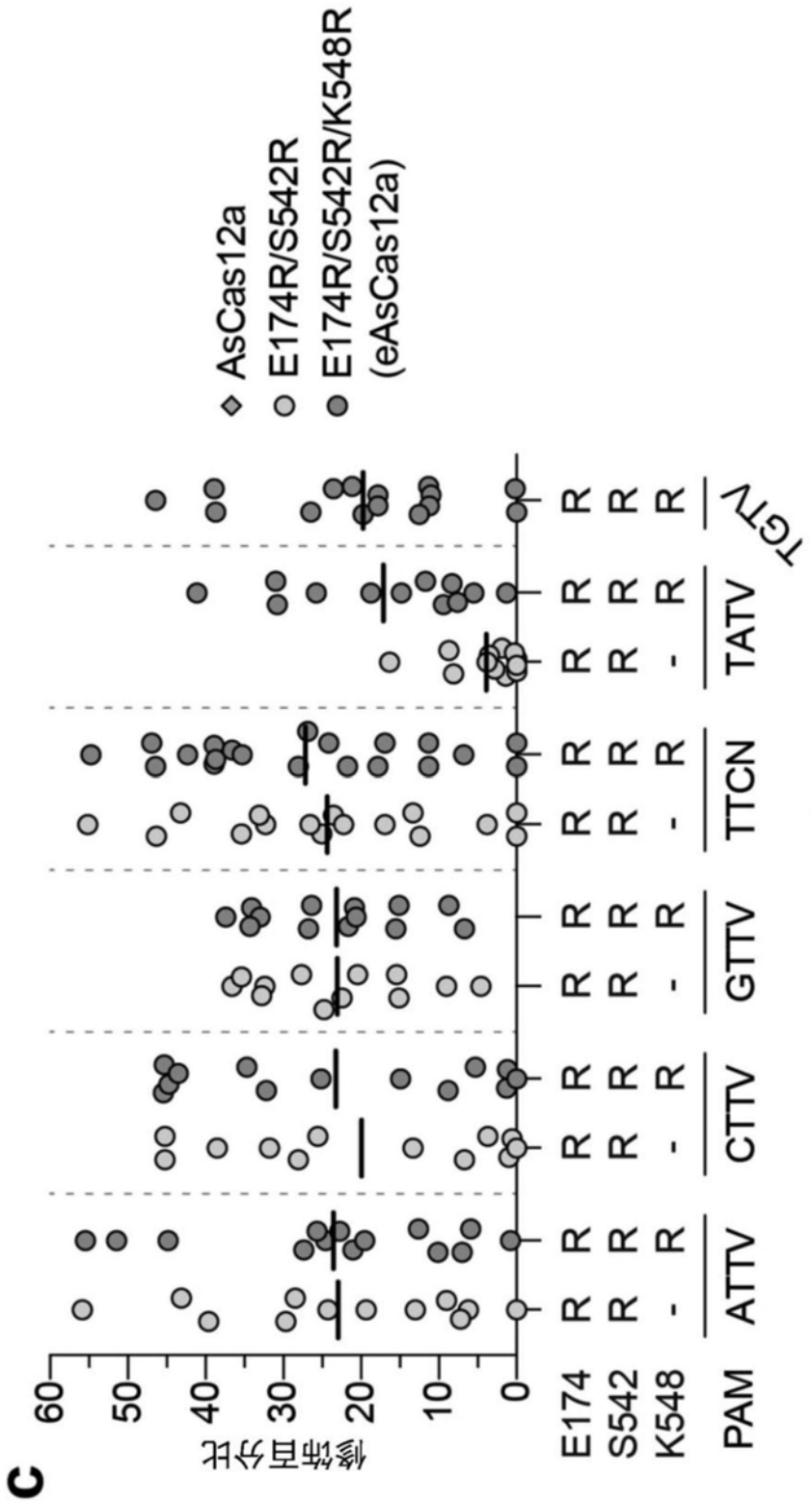

C
修饰百分比
60
50
40
30
20
10
0
AsCas12a
E174R/S542R
E174R/S542R/K548R
(eAsCas12a)
E174 R R R R R R R R
S542 R R R R R R R R
K548 - R - R - R - R
PAM ATTV CTTV GTTV TTCN TATV TGTV

图15C

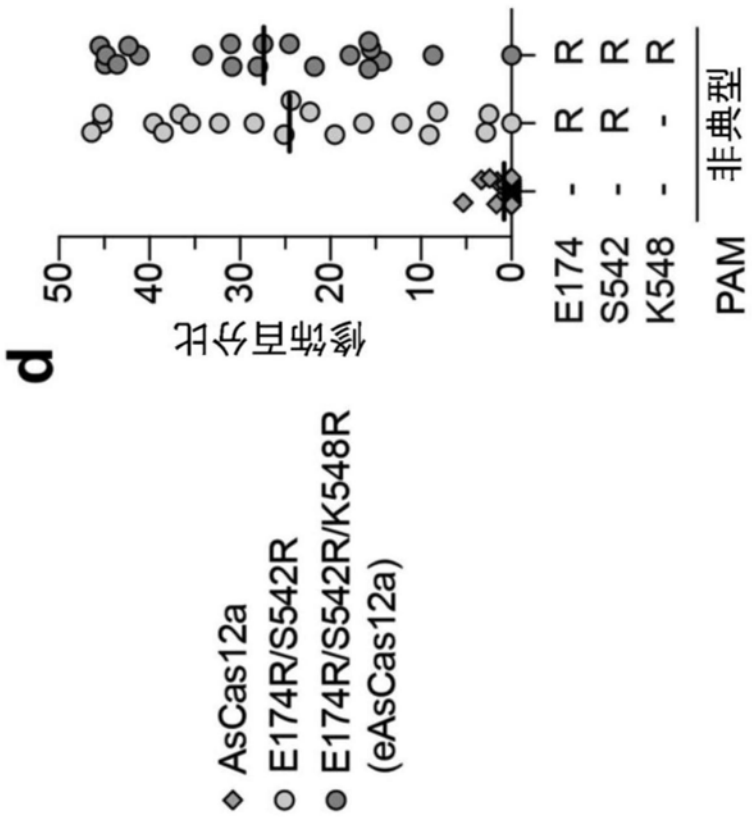

d
修饰百分比
50
40
30
20
10
0
E174 - R R
S542 - R R
K548 - - R
PAM 非典型
AsCas12a
E174R/S542R
E174R/S542R/K548R
(eAsCas12a)

图15D

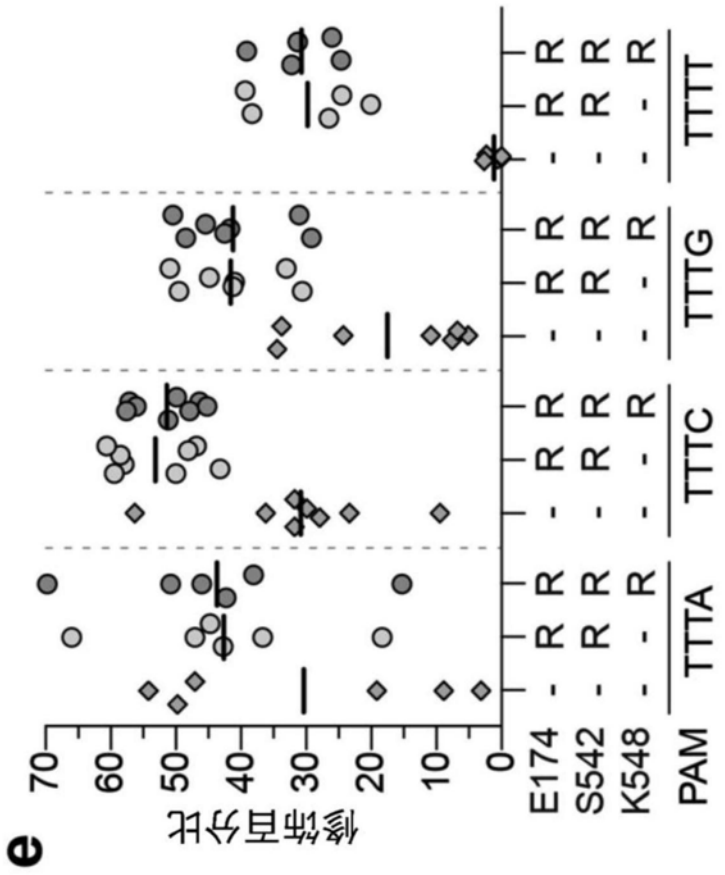

e
修饰百分比
0
10
20
30
40
50
60
70
PAM
K548
S542
E174
TTTA
TTTC
TTTG
TTTT
- R R
- R R
- - R

图15E

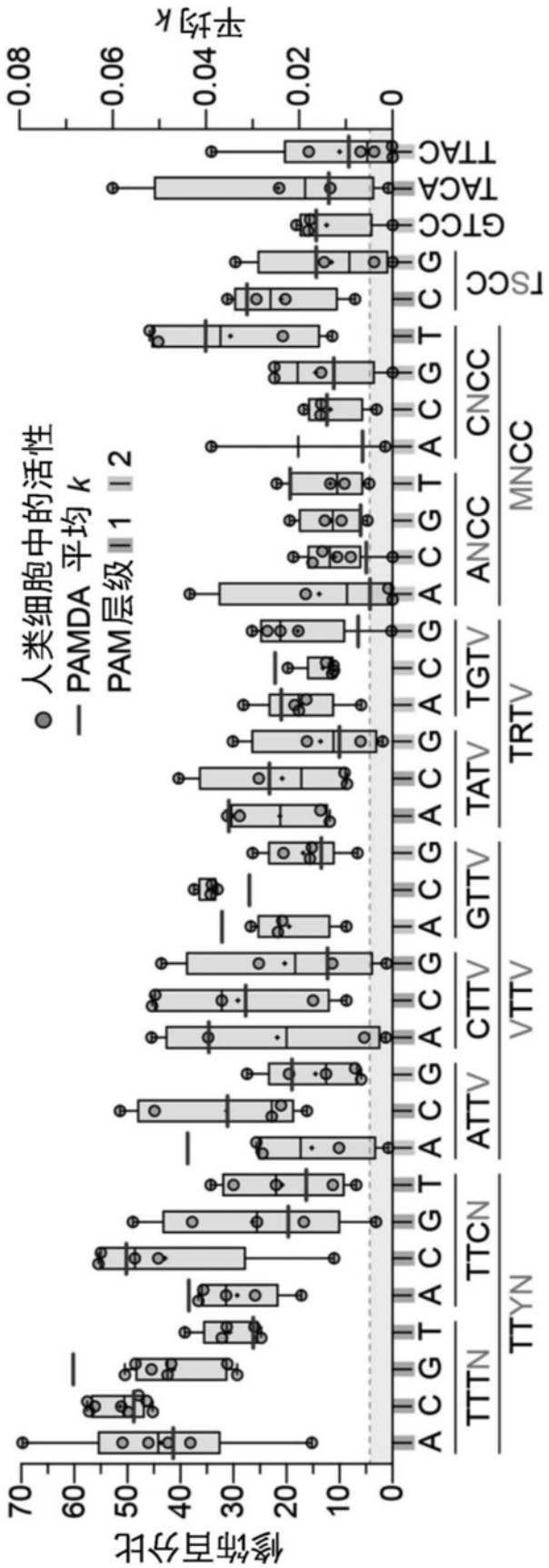

平均k
0.08
0.06
0.04
0.02
0
人类细胞中的活性
PAMDA 平均 k
PAM层级 1 2
修饰百分比
70
60
50
40
30
20
10
0
TTTN
TTCN
ATTV
CTTV
GTTV
TATV
TGTV
ANCC
CNCC
CCST
TTYN
VTTV
TRTV
MNCC
A C G T A C G T A C G A C G A C G A C G A C G A C G T A C G T C G

图15F

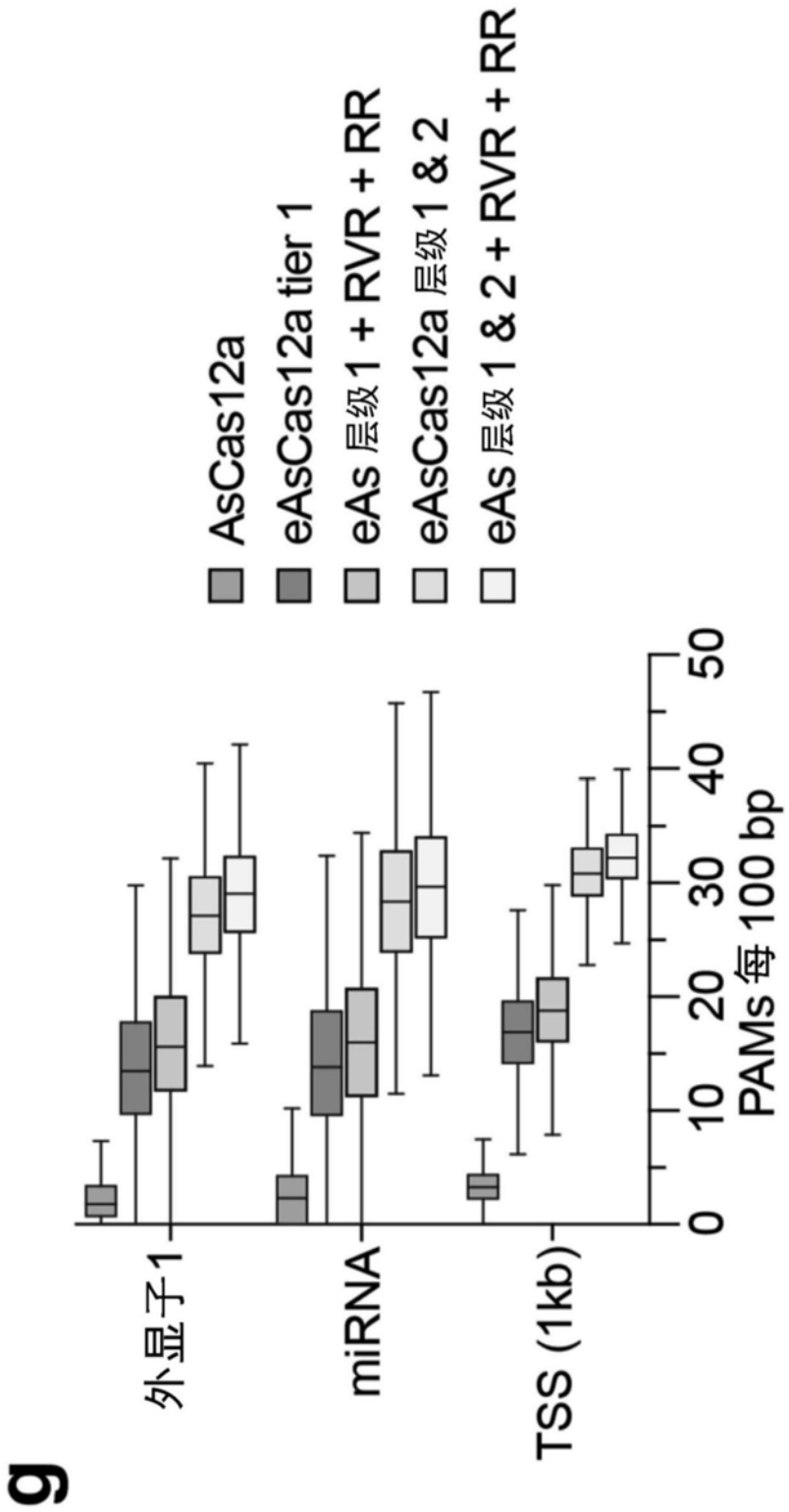

g
AsCas12a
eAsCas12a tier 1
eAs 层级1 + RVR + RR
eAsCas12a 层级1 & 2
eAs 层级1 & 2 + RVR + RR
外显子1
miRNA
TSS (1kb)
0
10
20
30
40
50
PAMs 每 100 bp

图15G

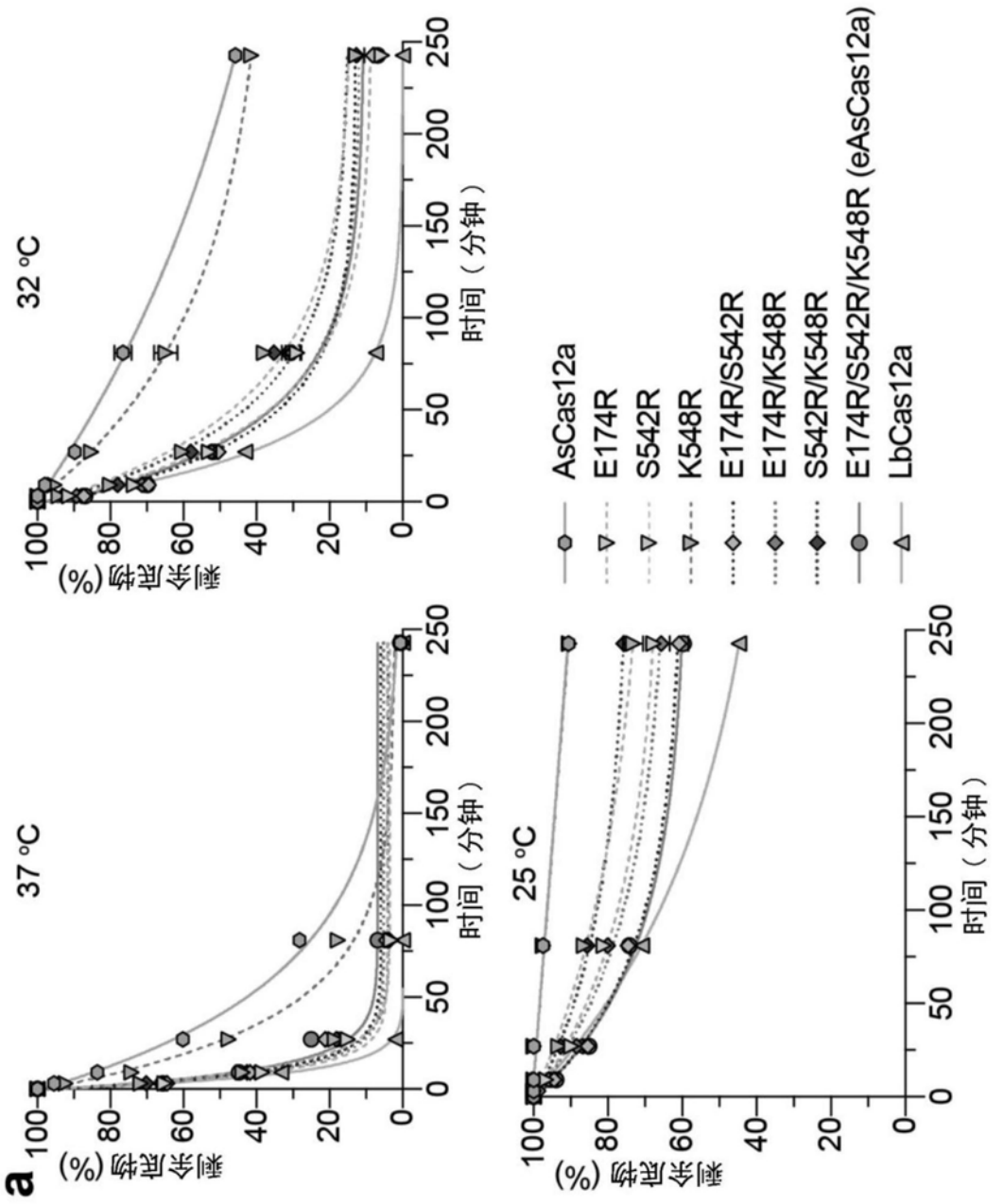

a
37 °C
25 °C
32 °C
剩余底物 (%)
时间（分钟）
AsCas12a
E174R
S542R
K548R
E174R/S542R
E174R/K548R
S542R/K548R
E174R/S542R/K548R (eAsCas12a)
LbCas12a

图16A

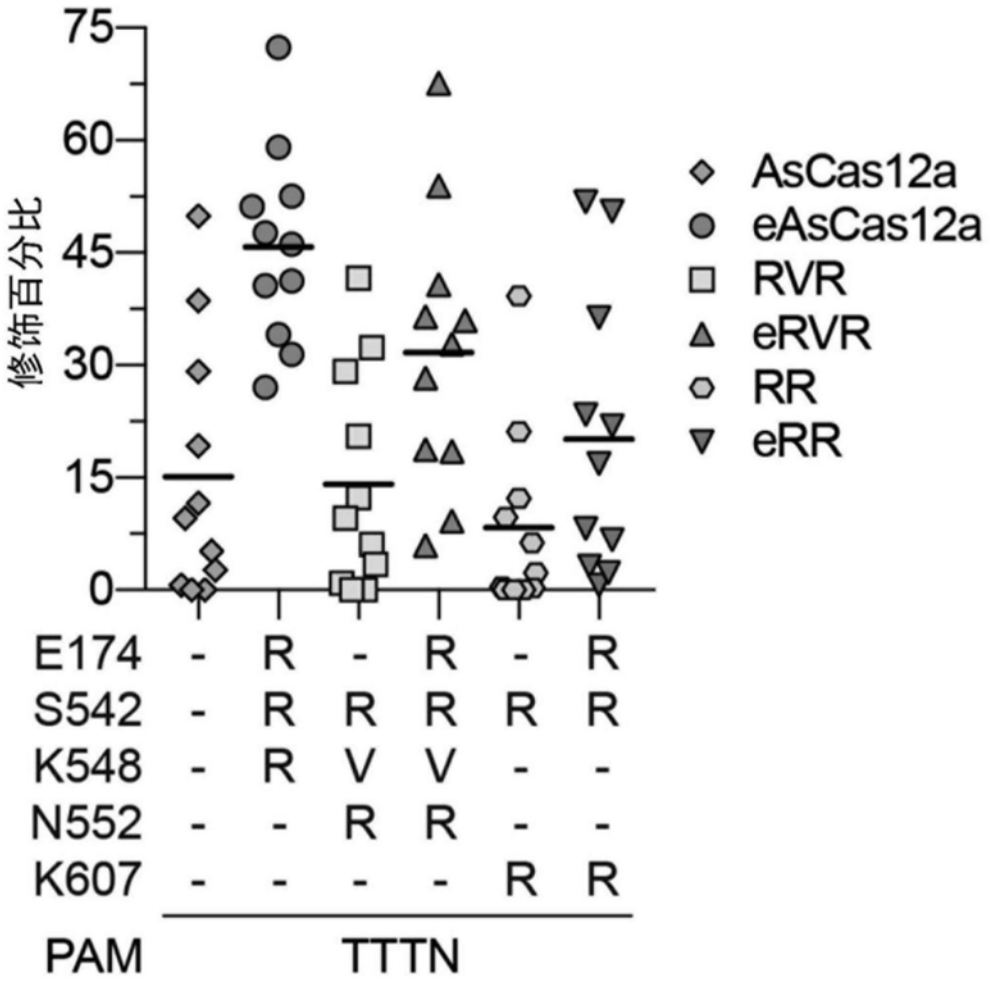

75
60
45
30
15
0
修饰百分比
AsCas12a
eAsCas12a
RVR
eRVR
RR
eRR
E174 - R - R - R
S542 - R R R R R
K548 - R V V - -
N552 - - R R - -
K607 - - - - R R
PAM TTTN

图16B

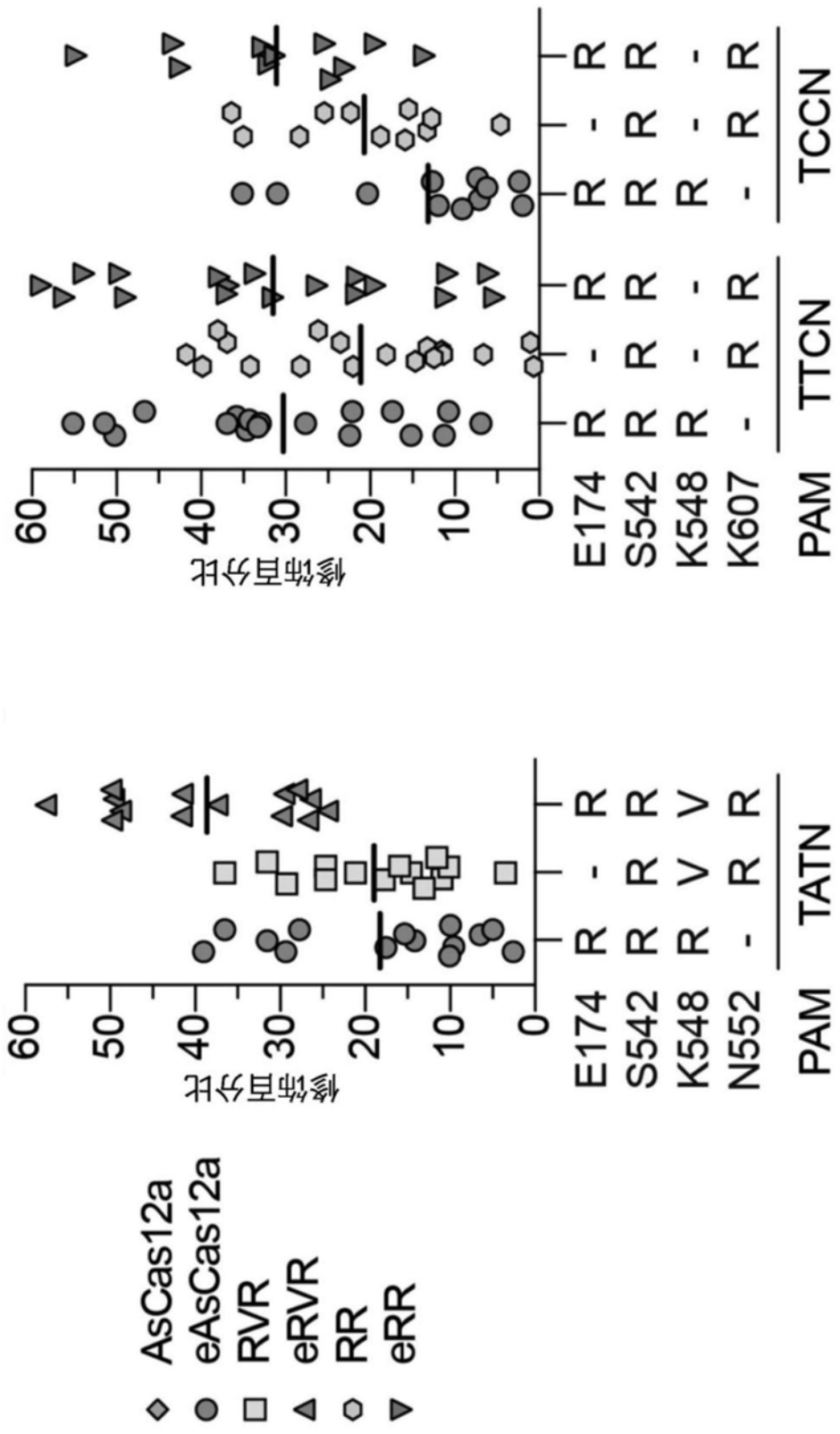

修饰百分比
60
50
40
30
20
10
0
E174 R - R
S542 R R R
K548 R - - R - -
K607 - R R - R R
PAM TTCN TCCN
修饰百分比
60
50
40
30
20
10
0
E174 R - R
S542 R R R
K548 R V V
N552 - R R
PAM TATN
AsCas12a
eAsCas12a
RVR
eRVR
RR
eRR

图16C-D

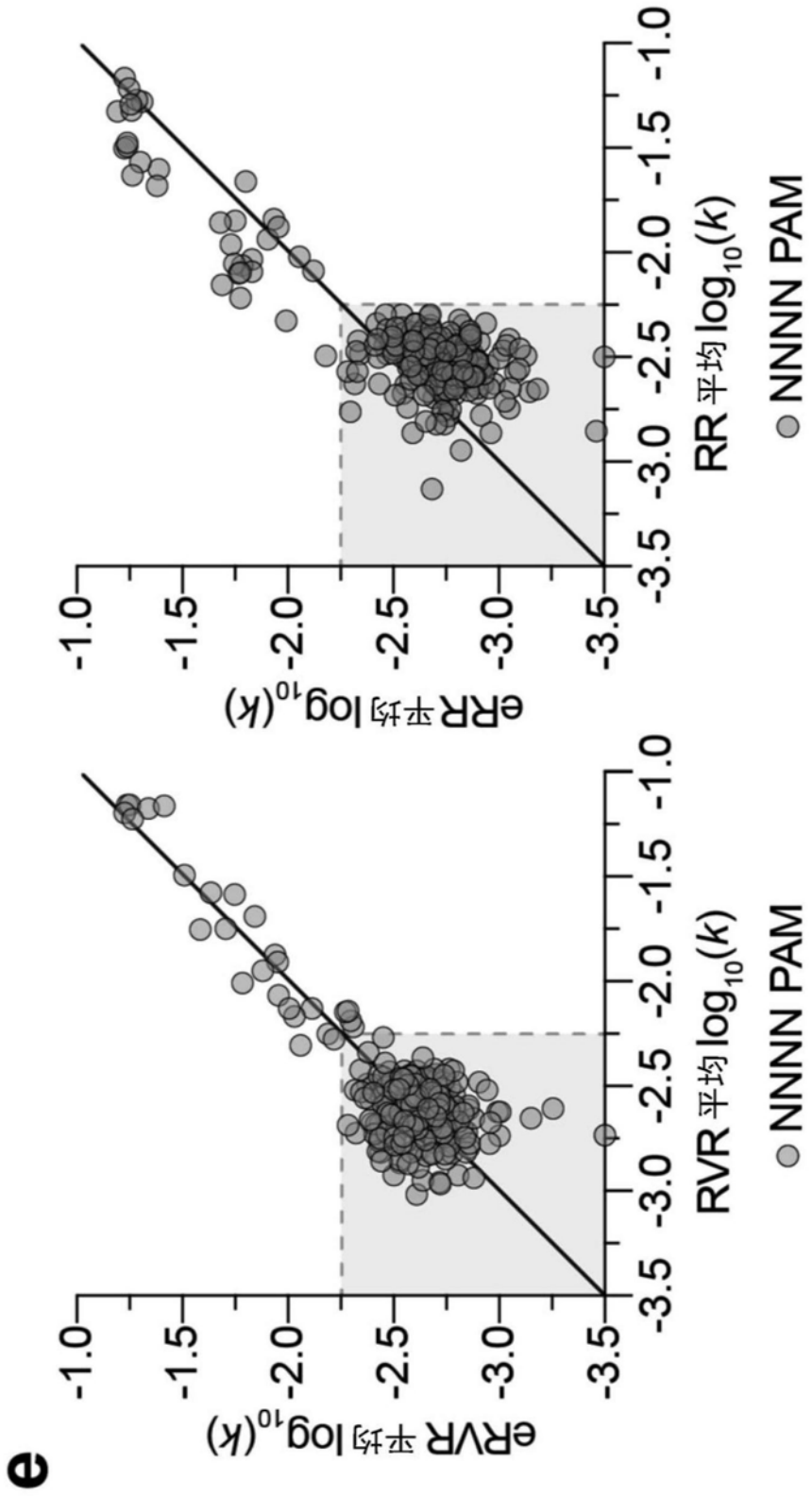

e
eRVR 平均 $\log_{10}(k)$
RVR 平均 $\log_{10}(k)$
NNNN PAM
eRR 平均 $\log_{10}(k)$
RR 平均 $\log_{10}(k)$
NNNN PAM
-1.0
-1.5
-2.0
-2.5
-3.0
-3.5

图16E

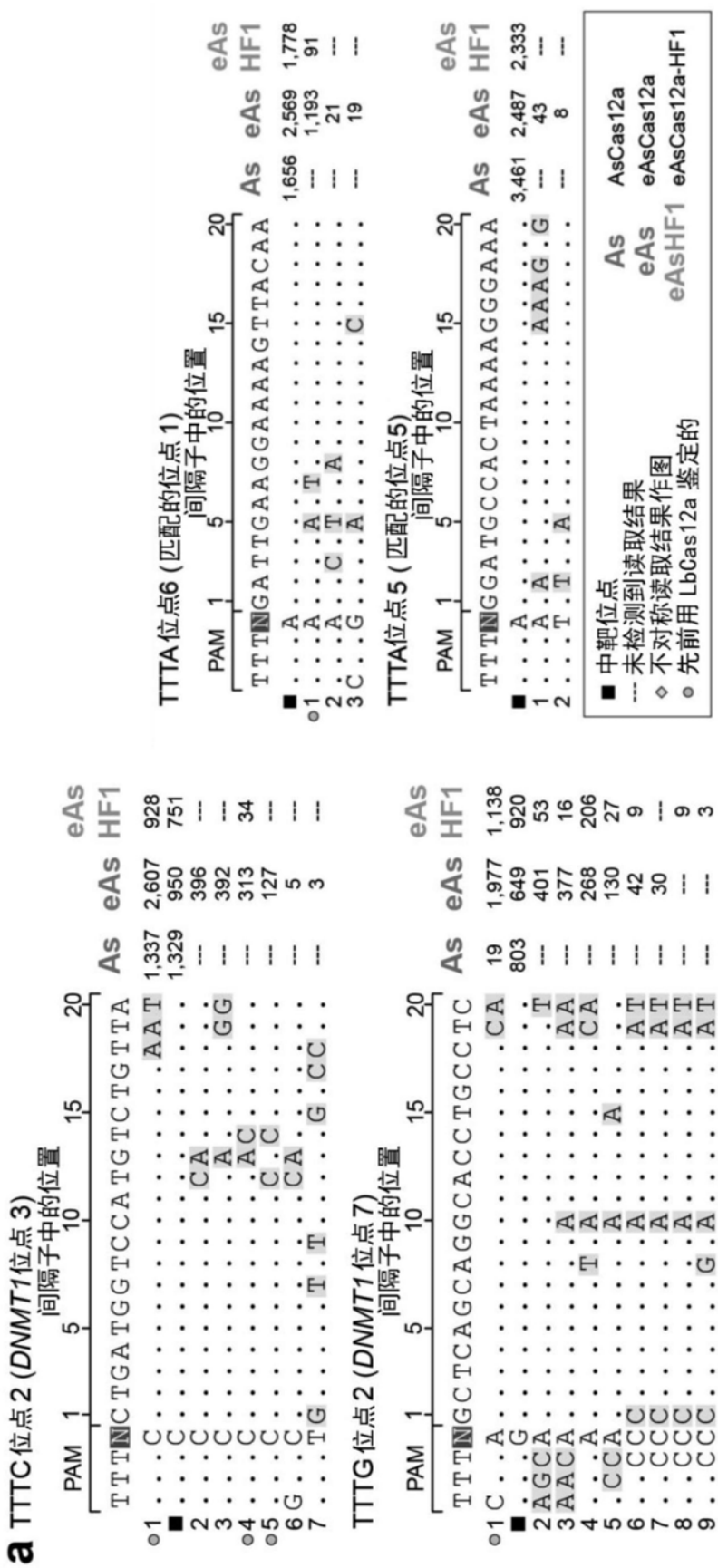

a
TTTC位点2 (DNMT1位点3)
间隔子中的位置
PAM
TTTNCTGATGGTCCATGTCTGTTA
As
eAs
eAs HF1
TTTG位点2 (DNMT1位点7)
间隔子中的位置
PAM
TTTNGCTCAGCAGGCACCTGCCTC
As
eAs
eAs HF1
TTTA位点6 (匹配的位点1)
间隔子中的位置
PAM
TTTNGATTGAAGGAAAAGTTACAA
As
eAs
eAs HF1
TTTA位点5 (匹配的位点5)
间隔子中的位置
PAM
TTTNGGATGCCACTAAAAGGGAAA
As
eAs
eAs HF1
■ 中靶位点
--- 未检测到读取结果
◇ 不对称读取结果作图
● 先前用 LbCas12a 鉴定的
As AsCas12a
eAs eAsCas12a
eAsHF1 eAsCas12a-HF1

图17A

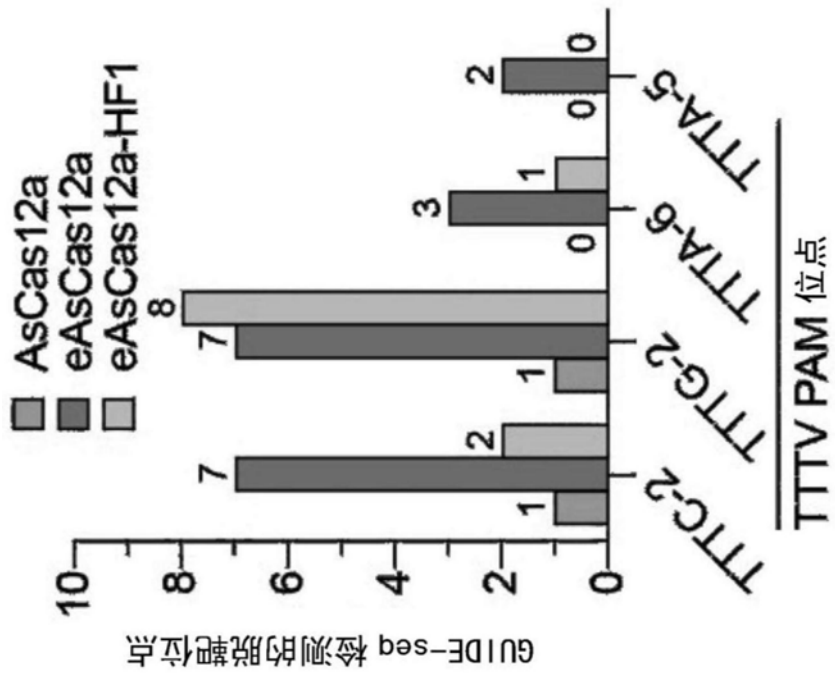

AsCas12a
eAsCas12a
eAsCas12a-HF1
GUIDE-seq 检测的脱靶位点
TTTV PAM 位点
TTTC-2
TTTG-2
TTTA-6
TTTA-5

图17B

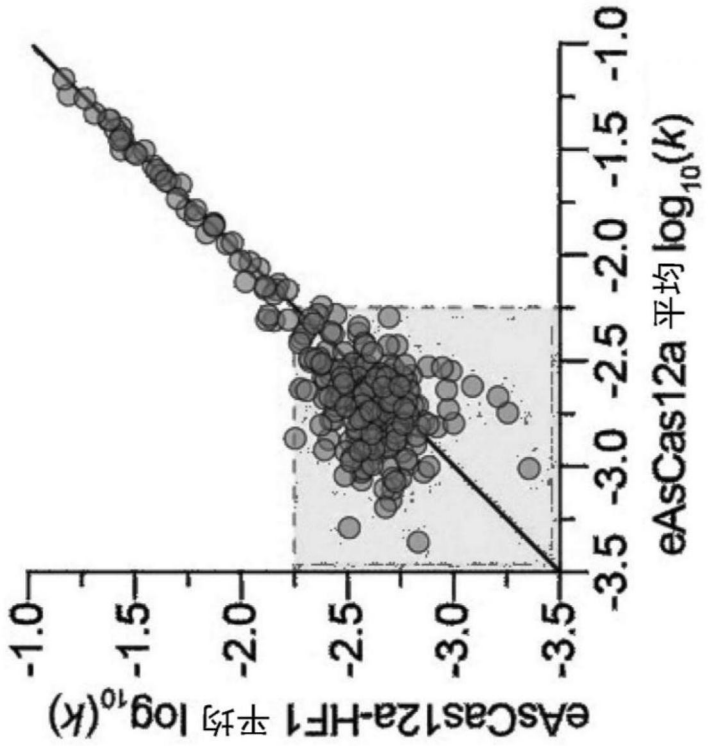

eAsCas12a-HF1 平均 log10(k)
eAsCas12a 平均 log10(k)

图17C

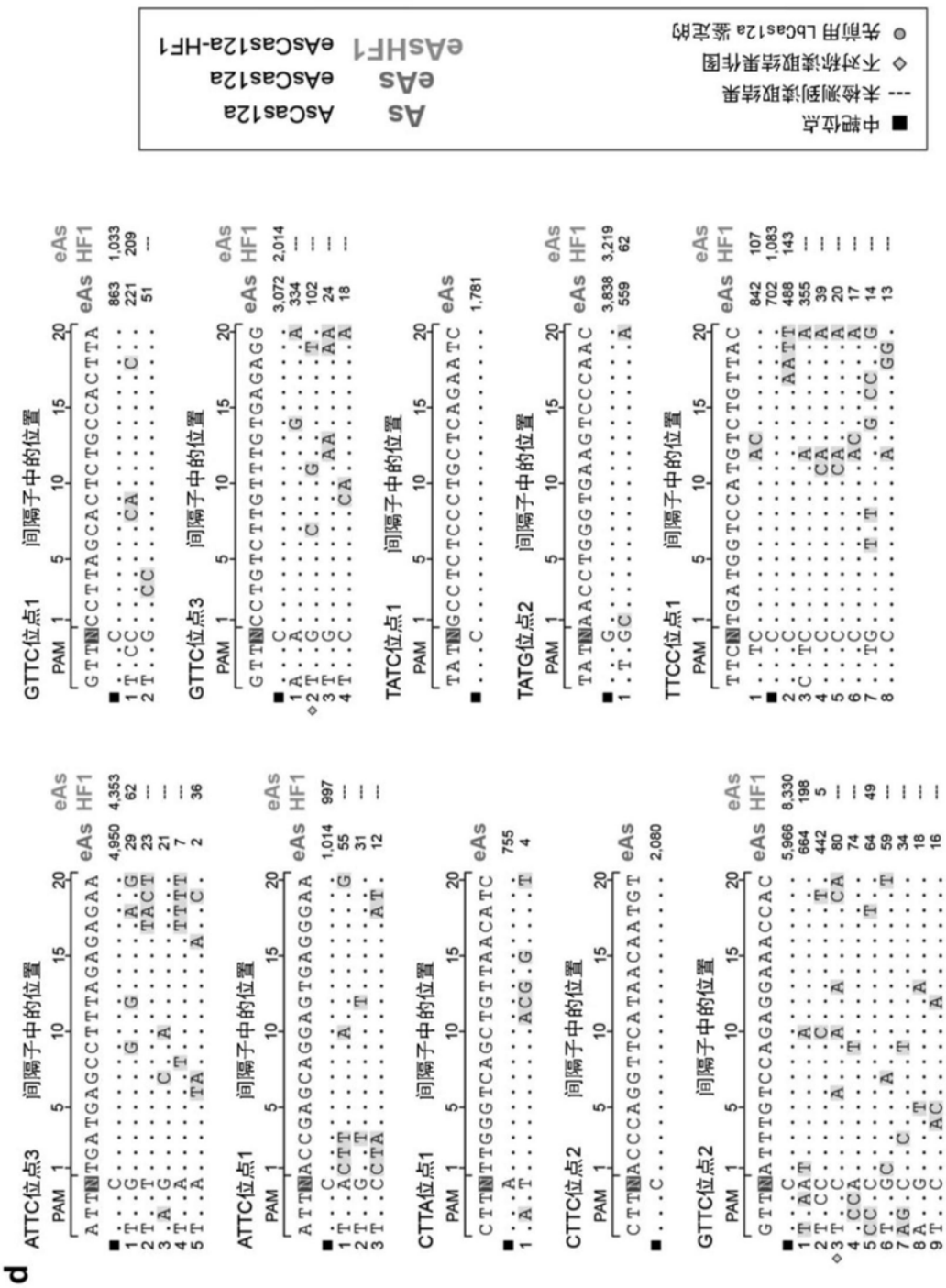

d
ATTC位点3 间隔子中的位置
GTTC位点1 间隔子中的位置
GTTC位点3 间隔子中的位置
TATC位点1 间隔子中的位置
TATG位点2 间隔子中的位置
TTCC位点1 间隔子中的位置
ATTC位点1 间隔子中的位置
CTTA位点1 间隔子中的位置
CTTC位点2 间隔子中的位置
GTTC位点2 间隔子中的位置
PAM
eAs
HF1
AS
AsCas12a
eAs
eAsCas12a
eASHF1
eAsCas12a-HF1
中靶位点
未检测到读取结果
不对称读取结果作图
先前用 LbCas12a 鉴定的

图17D

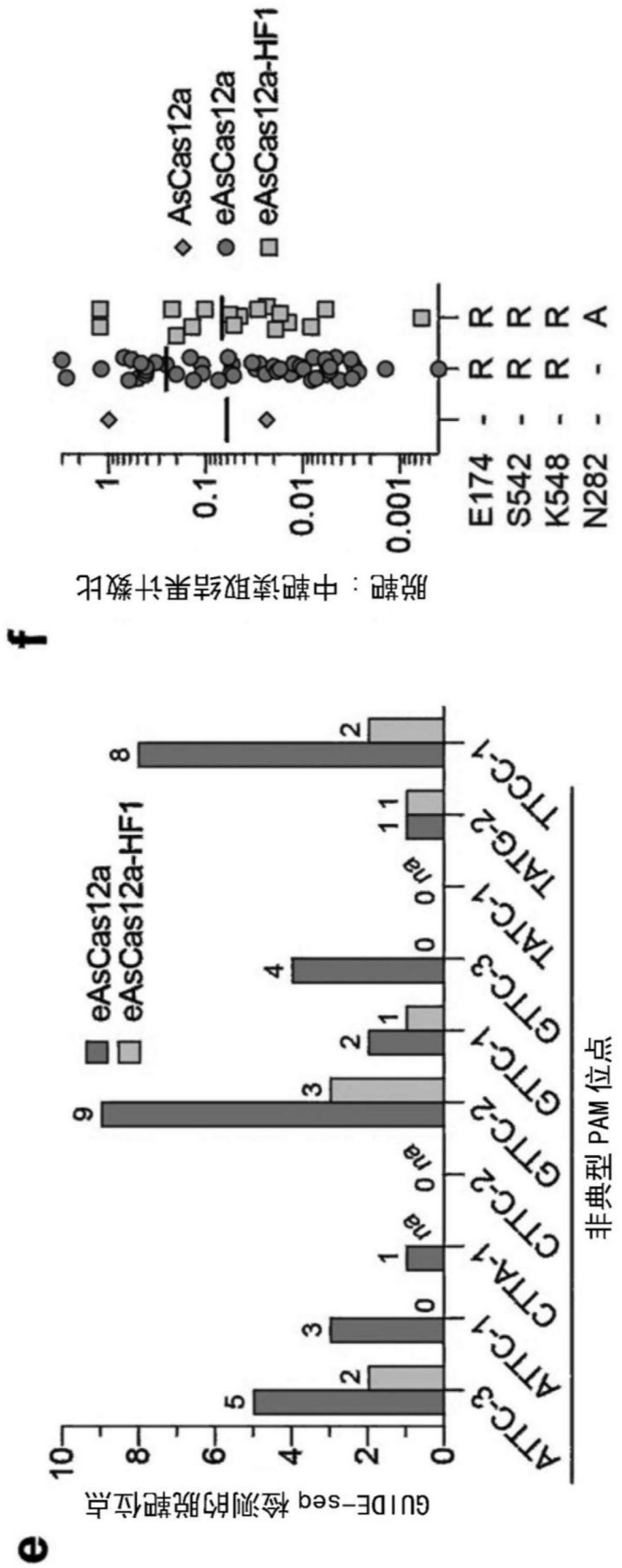

f
脱靶：中靶读取结果计数比
AsCas12a
eAsCas12a
eAsCas12a-HF1
1
0.1
0.01
0.001
E174 - R R
S542 - R R
K548 - R R
N282 - - A
e
GUIDE-seq 检测的脱靶位点
eAsCas12a
eAsCas12a-HF1
ATTC-3
ATTC-1
CTTA-1
CTTC-2
GTTC-2
GTTC-1
GTTC-3
TATC-1
TATG-2
TTCC-1
非典型 PAM 位点

图17E-F

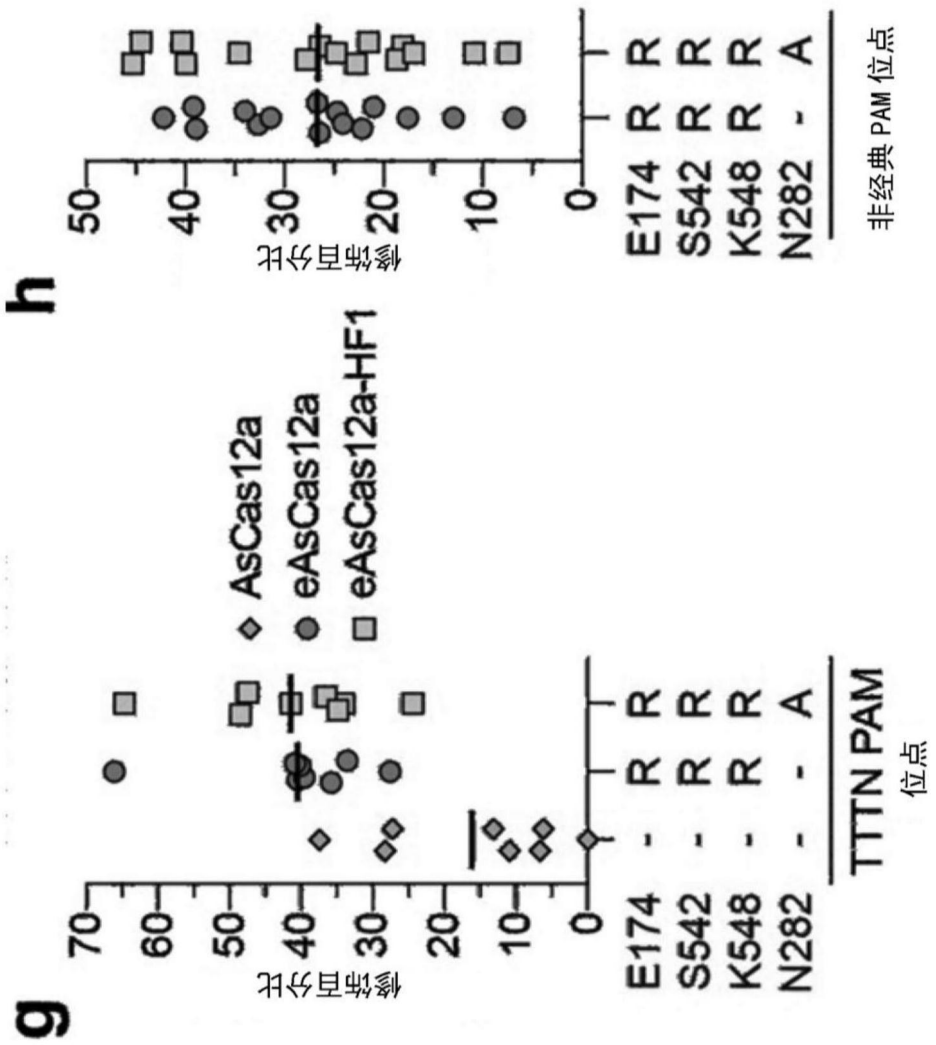

g
h
修饰百分比
AsCas12a
eAsCas12a
eAsCas12a-HF1
E174
S542
K548
N282
TTTN PAM
位点
非经典 PAM 位点

图17G-H

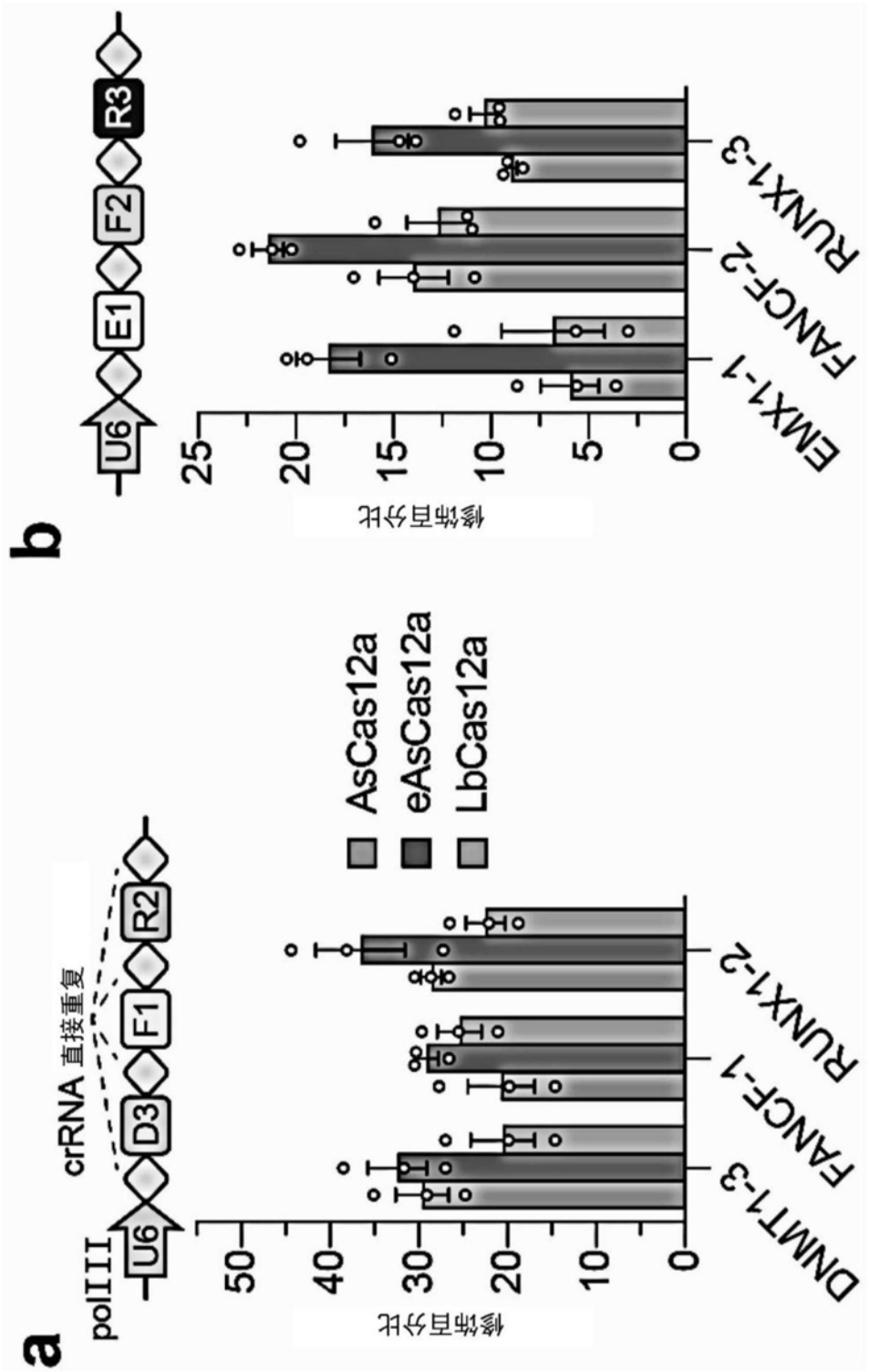

a
polIII
U6
D3
F1
R2
crRNA 直接重复
AsCas12a
eAsCas12a
LbCas12a
修饰百分比
0
10
20
30
40
50
DNMT1-3
FANCF-1
RUNX1-2
b
U6
E1
F2
R3
修饰百分比
0
5
10
15
20
25
EMX1-1
FANCF-2
RUNX1-3

图18A-B

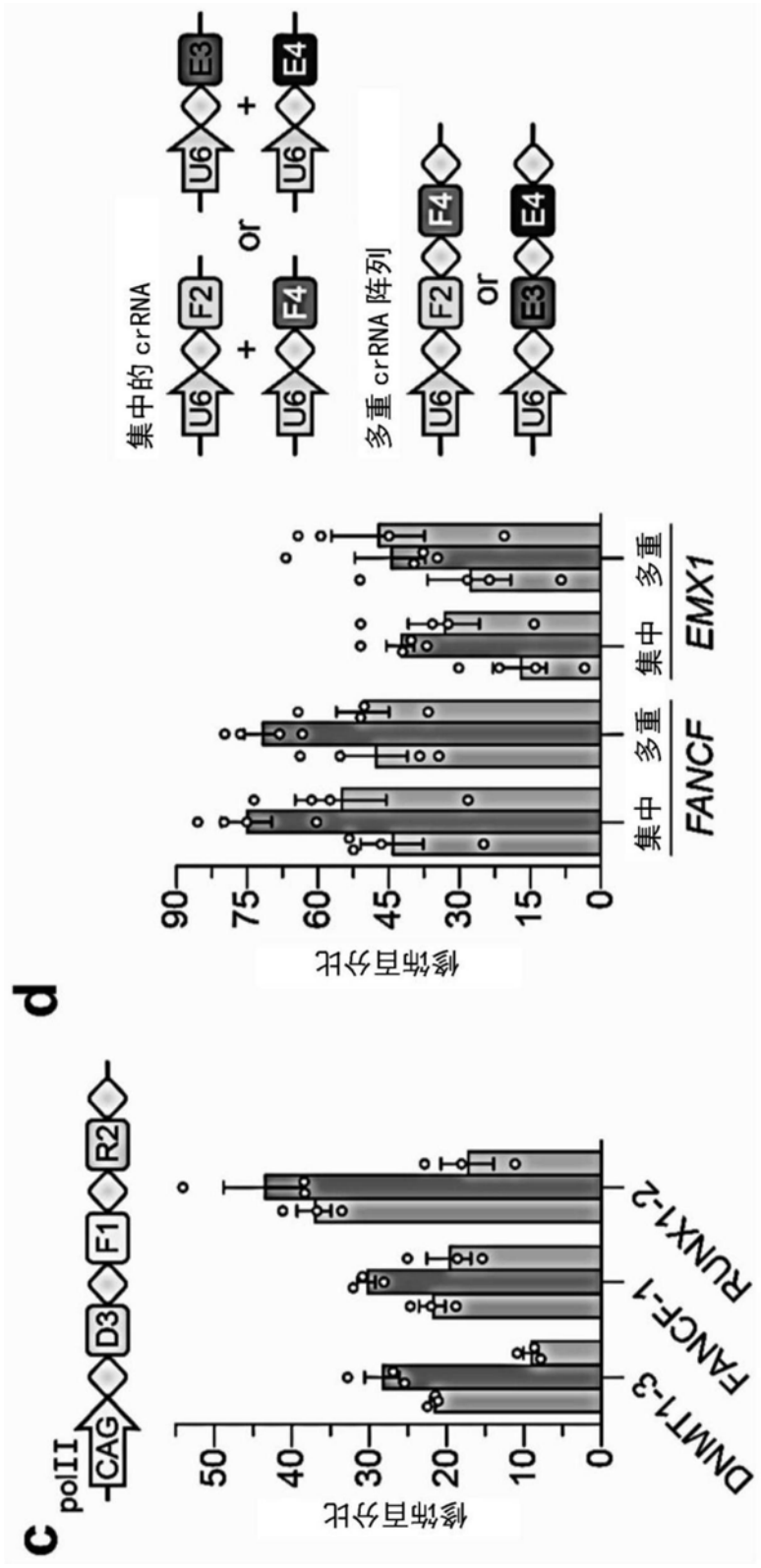

c
polII
CAG
D3
F1
R2
修饰百分比
0
10
20
30
40
50
DNMT1-3
FANCF-1
RUNX1-2
d
集中的 crRNA
U6
F2
F4
or
U6
E3
U6
E4
多重 crRNA 阵列
U6
F2
F4
or
U6
E3
E4
修饰百分比
0
15
30
45
60
75
90
集中
多重
FANCF
集中
多重
EMX1

图18C-D

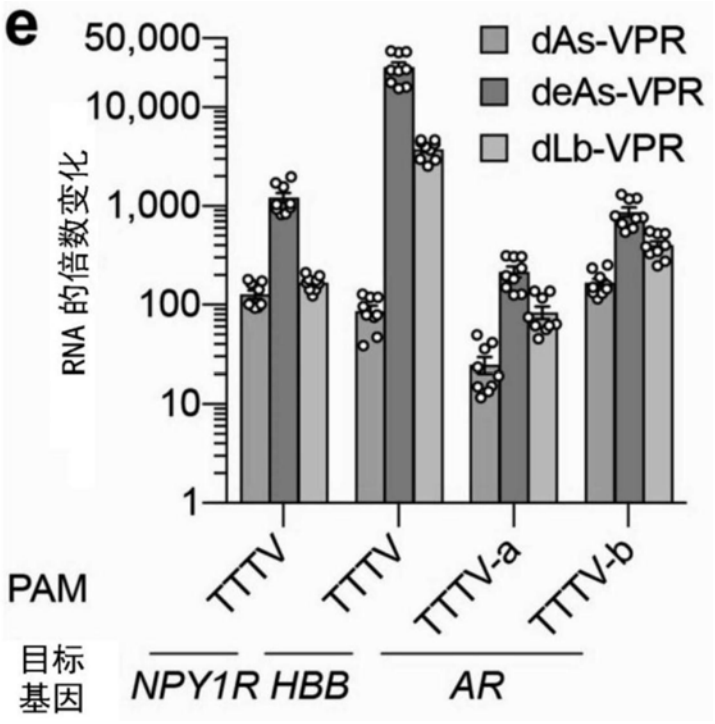

e
50,000
10,000
1,000
100
10
1
RNA 的倍数变化
dAs-VPR
deAs-VPR
dLb-VPR
PAM
TTTV
TTTV
TTTV-a
TTTV-b
目标基因
NPY1R
HBB
AR

图18E

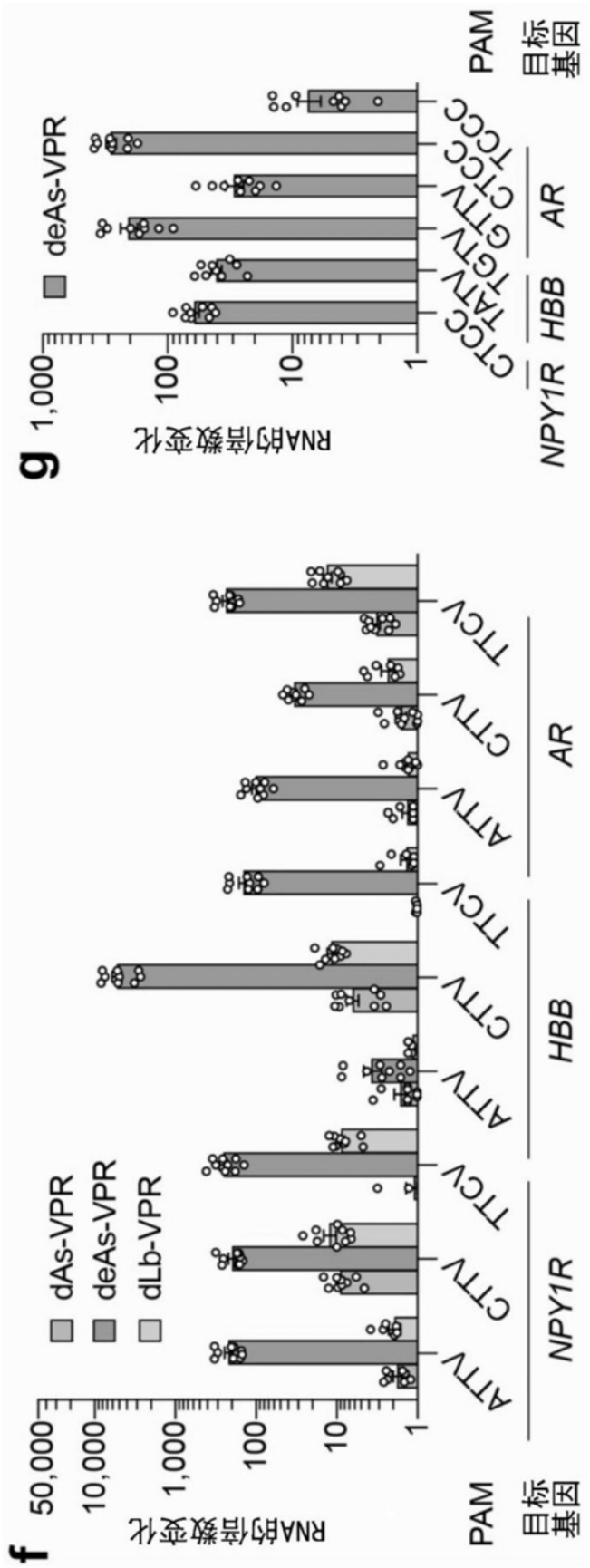

g
deAs-VPR
RNA的倍数变化
1,000
100
10
1
PAM
目标基因
CTCC
ATTV
CTTV
TTCV
NPY1R
HBB
AR
f
dAs-VPR
deAs-VPR
dLb-VPR
RNA的倍数变化
50,000
10,000
1,000
100
10
1
PAM
目标基因
ATTV
CTTV
TTCV
NPY1R
HBB
AR

图18F-G

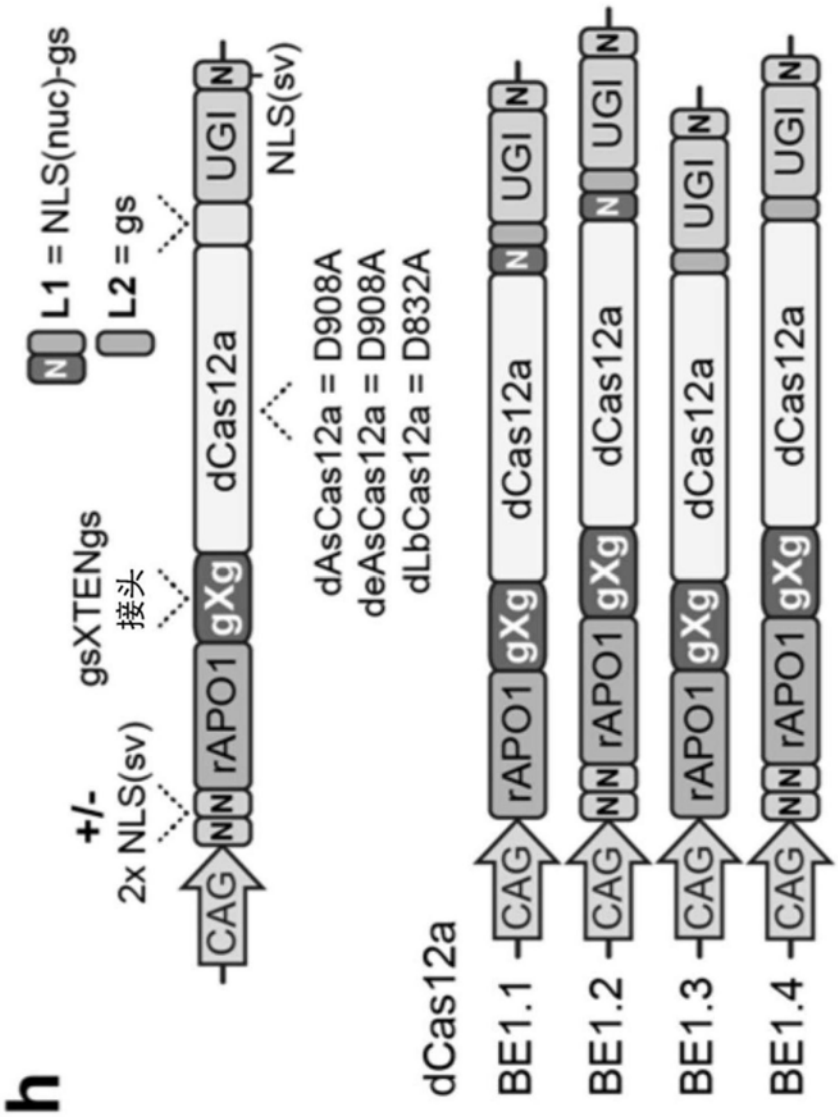

h
dCas12a
BE1.1
BE1.2
BE1.3
BE1.4
CAG
rAPO1
gXg
dCas12a
UGI
NLS(sv)
2x NLS(sv)
+/-
gsXTENgs
接头
L1 = NLS(nuc)-gs
L2 = gs
dAsCas12a = D908A
deAsCas12a = D908A
dLbCas12a = D832A

图18H

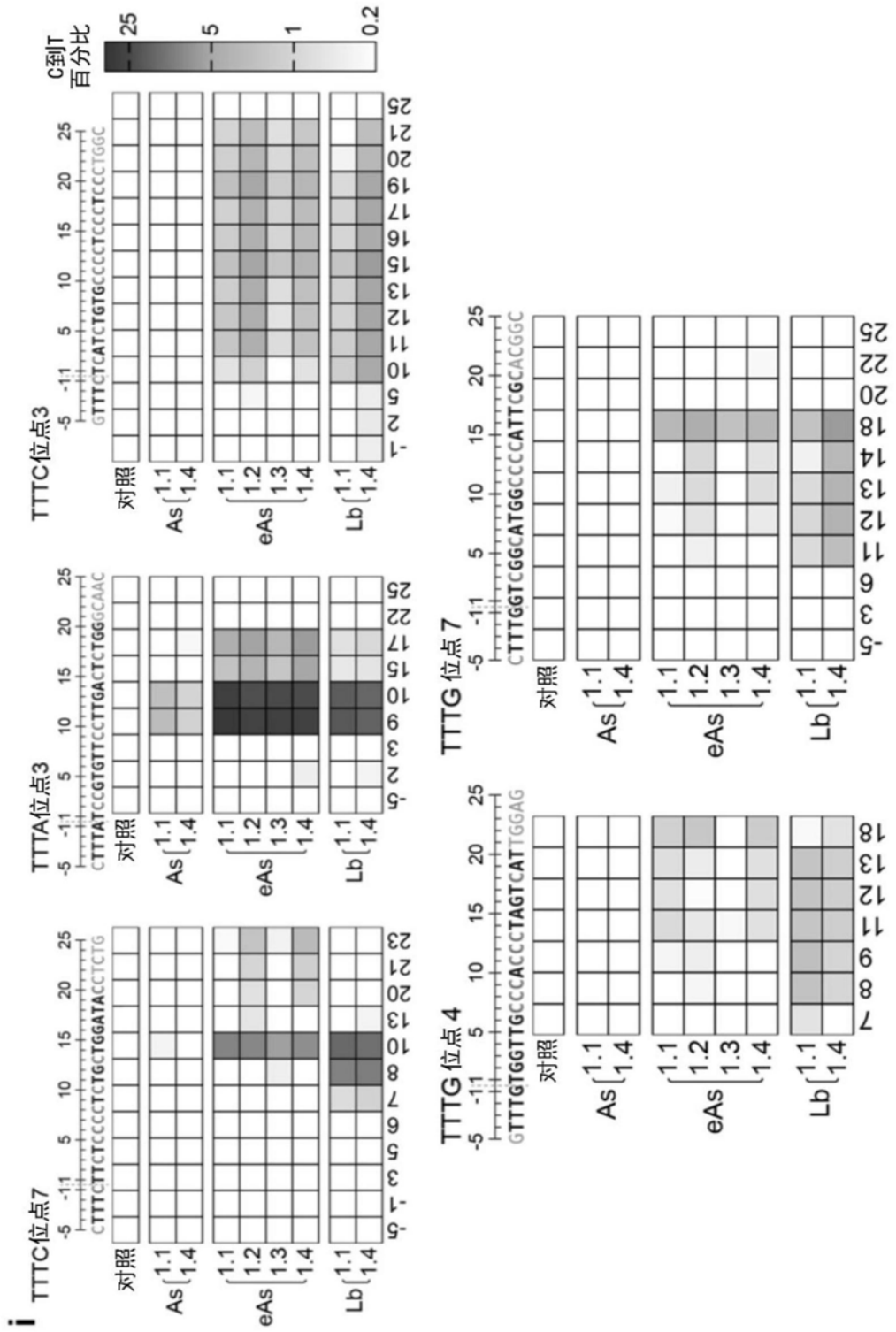

i
TTTC位点7
TTTA位点3
TTTC位点3
TTTG位点4
TTTG位点7
C到T百分比
25
5
1
0.2
对照
As
eAs
Lb

图18I

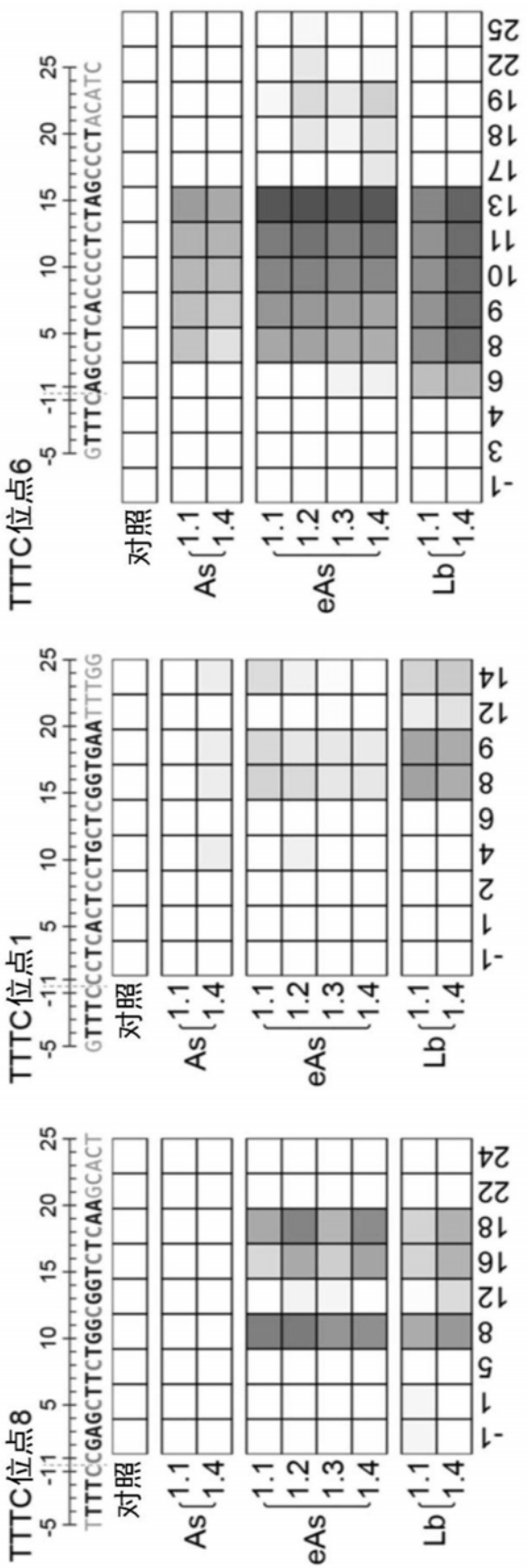

TTTC位点6
GTTTCAGCCTCACCCCCTCTAGCCCTACATC
-5 -1 1 5 10 15 20 25
对照
As
1.1
1.4
eAs
1.1
1.2
1.3
1.4
Lb
1.1
1.4
-1 3 4 6 8 9 10 11 13 17 18 19 22 25
TTTC位点1
GTTTCCCTCACTCCTGCTCGGTGAATTTGG
-5 -1 1 5 10 15 20 25
对照
As
1.1
1.4
eAs
1.1
1.2
1.3
1.4
Lb
1.1
1.4
-1 1 2 4 6 8 9 12 14
TTTC位点8
TTTTCCGAGCTTCTGGCGGTCTCAAGCACT
-5 -1 1 5 10 15 20 25
对照
As
1.1
1.4
eAs
1.1
1.2
1.3
1.4
Lb
1.1
1.4
-1 1 5 8 12 16 18 22 24

图18I(续)

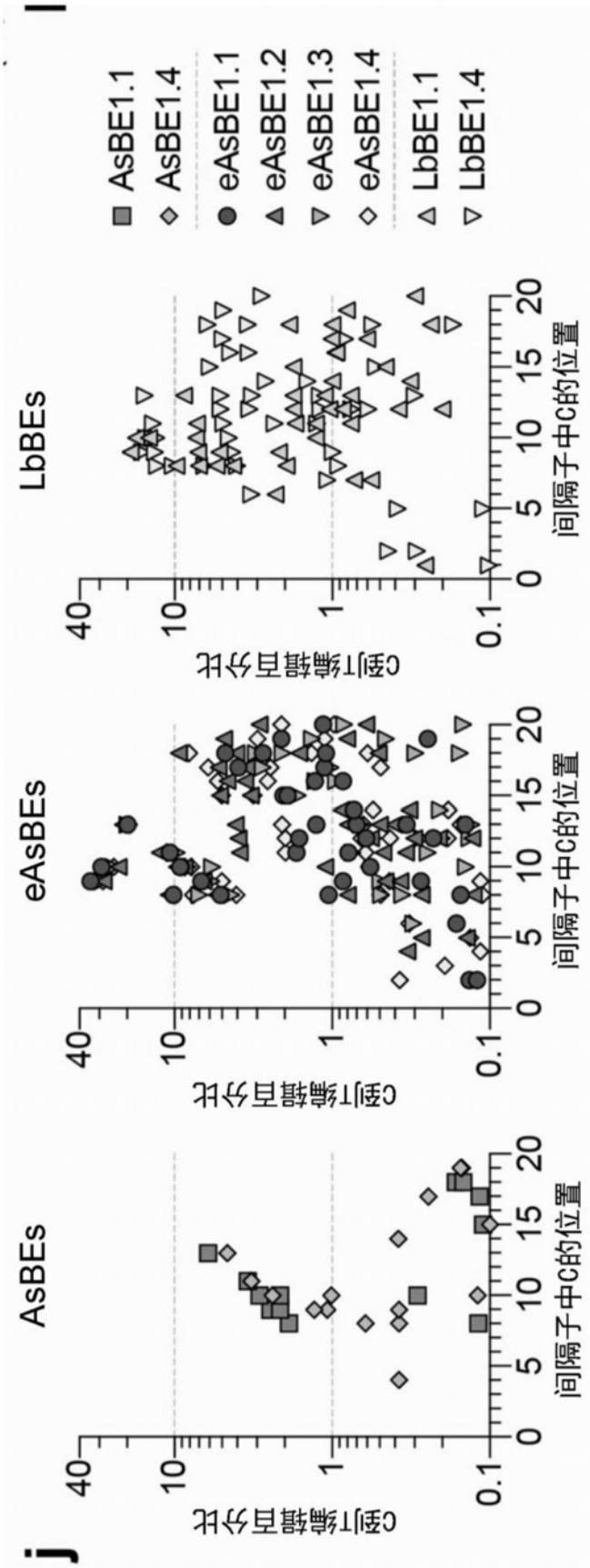

I
AsBE1.1
AsBE1.4
eAsBE1.1
eAsBE1.2
eAsBE1.3
eAsBE1.4
LbBE1.1
LbBE1.4
LbBEs
eAsBEs
AsBEs
C到T编辑百分比
间隔子中C的位置
40
10
1
0.1
0
5
10
15
20
j

图18J

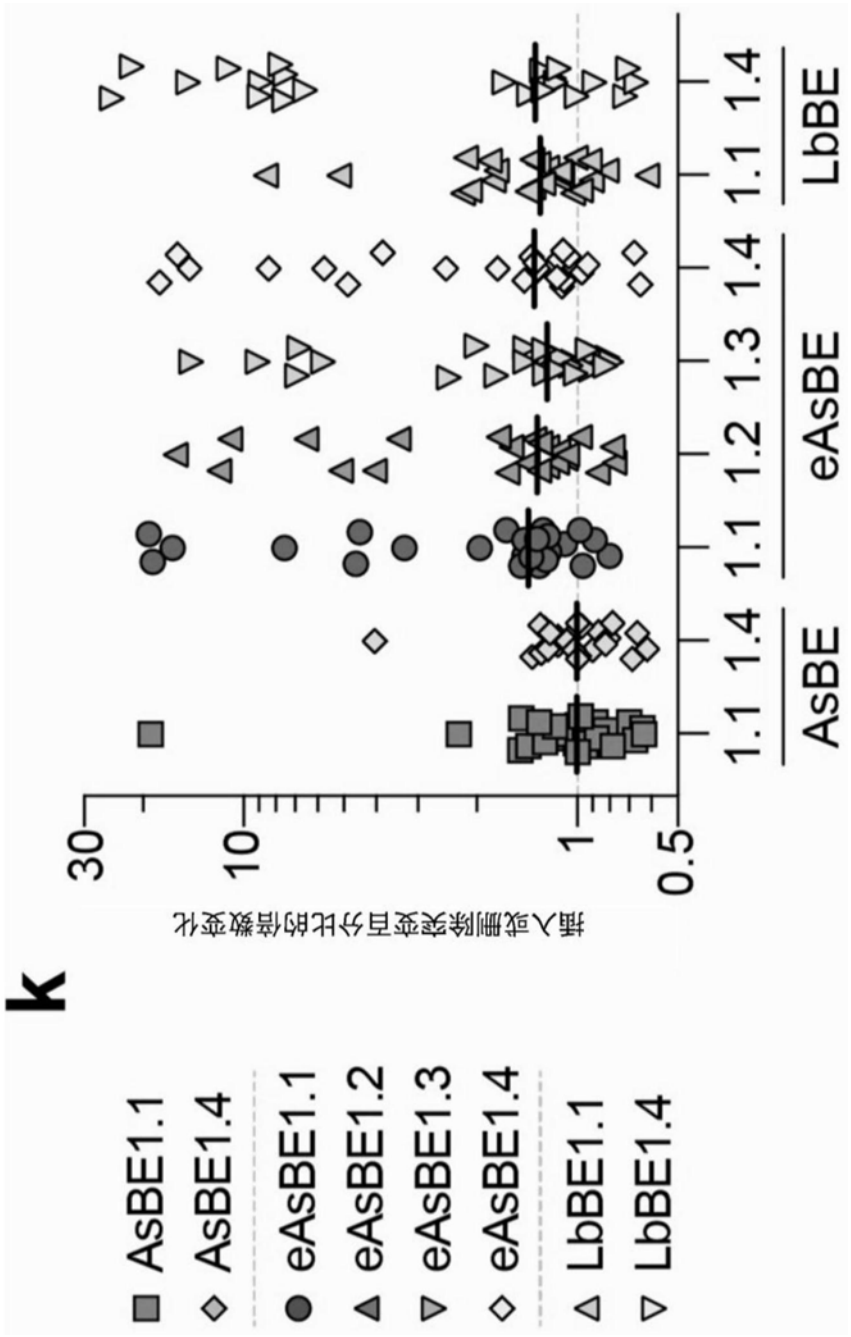

k
插入或删除突变百分比的倍数变化
30
10
1
0.5
1.1 1.4 1.1 1.2 1.3 1.4 1.1 1.4
AsBE
eAsBE
LbBE
AsBE1.1
AsBE1.4
eAsBE1.1
eAsBE1.2
eAsBE1.3
eAsBE1.4
LbBE1.1
LbBE1.4

图18K

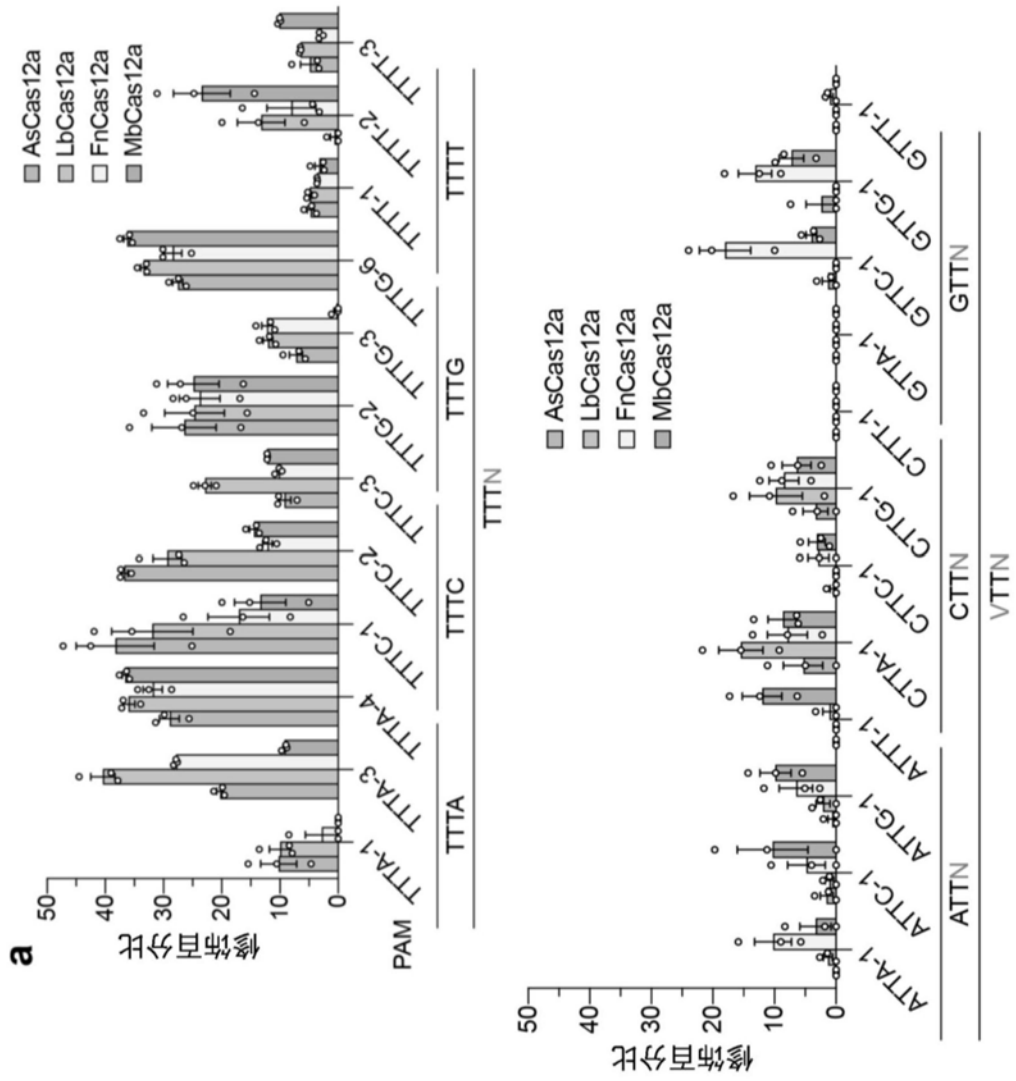

a
AsCas12a
LbCas12a
FnCas12a
MbCas12a
修饰百分比
0
10
20
30
40
50
PAM
TTTA
TTTC
TTTG
TTTT
TTTN
TTTA-1
TTTA-3
TTTA-4
TTTC-1
TTTC-2
TTTC-3
TTTG-2
TTTG-3
TTTG-6
TTTT-1
TTTT-2
TTTT-3
ATTA-1
ATTC-1
ATTG-1
ATTT-1
CTTA-1
CTTC-1
CTTG-1
CTTT-1
GTTA-1
GTTC-1
GTTG-1
GTTT-1
ATTN
CTTN
GTTN
VTTN

图19A

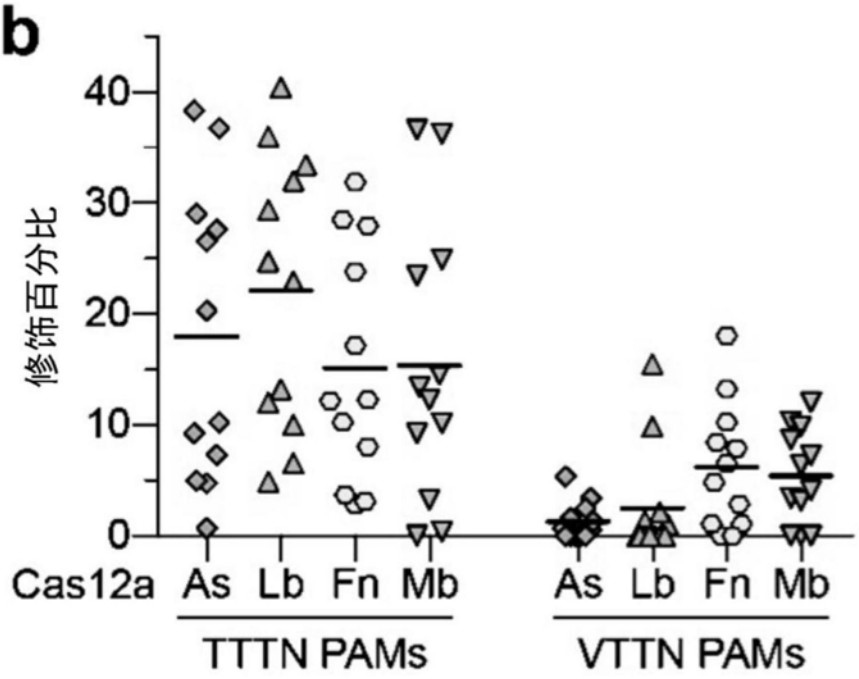

b
40
30
20
10
0
修饰百分比
Cas12a
As
Lb
Fn
Mb
As
Lb
Fn
Mb
TTTN PAMs
VTTN PAMs

图19B

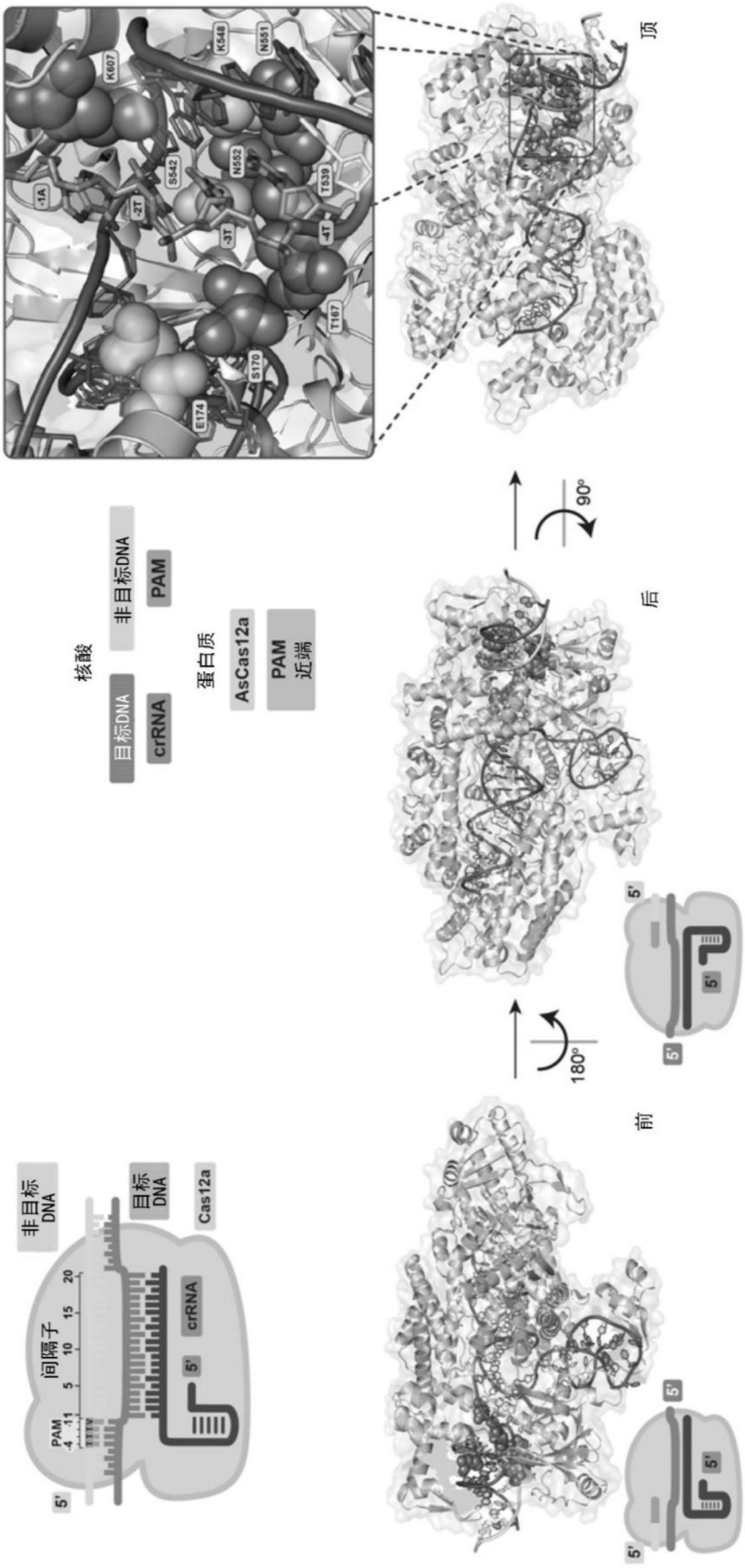

顶
后
前
90°
180°
核酸
非目标DNA
目标DNA
crRNA
PAM
蛋白质
AsCas12a
PAM
近端
非目标
DNA
目标
DNA
Cas12a
crRNA
间隔子
PAM
5'
K607
K548
N551
N552
S542
T539
T167
S170
E174
-1A
-2T
-3T
-4T

图20A

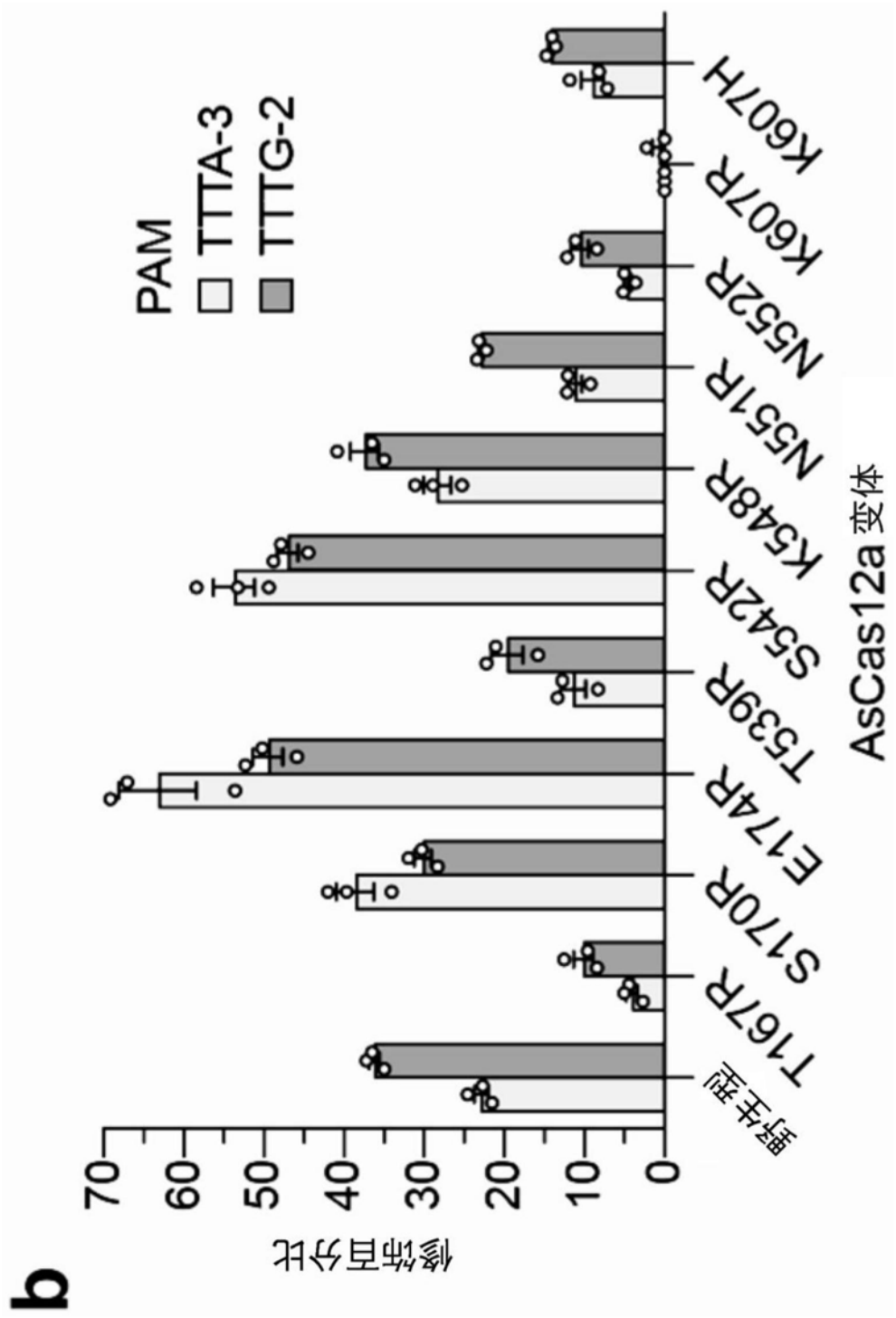

b
修饰百分比
70
60
50
40
30
20
10
0
野生型
T167R
S170R
E174R
T539R
S542R
K548R
N551R
N552R
K607R
K607H
AsCas12a 变体
PAM
TTTA-3
TTTG-2

图20B

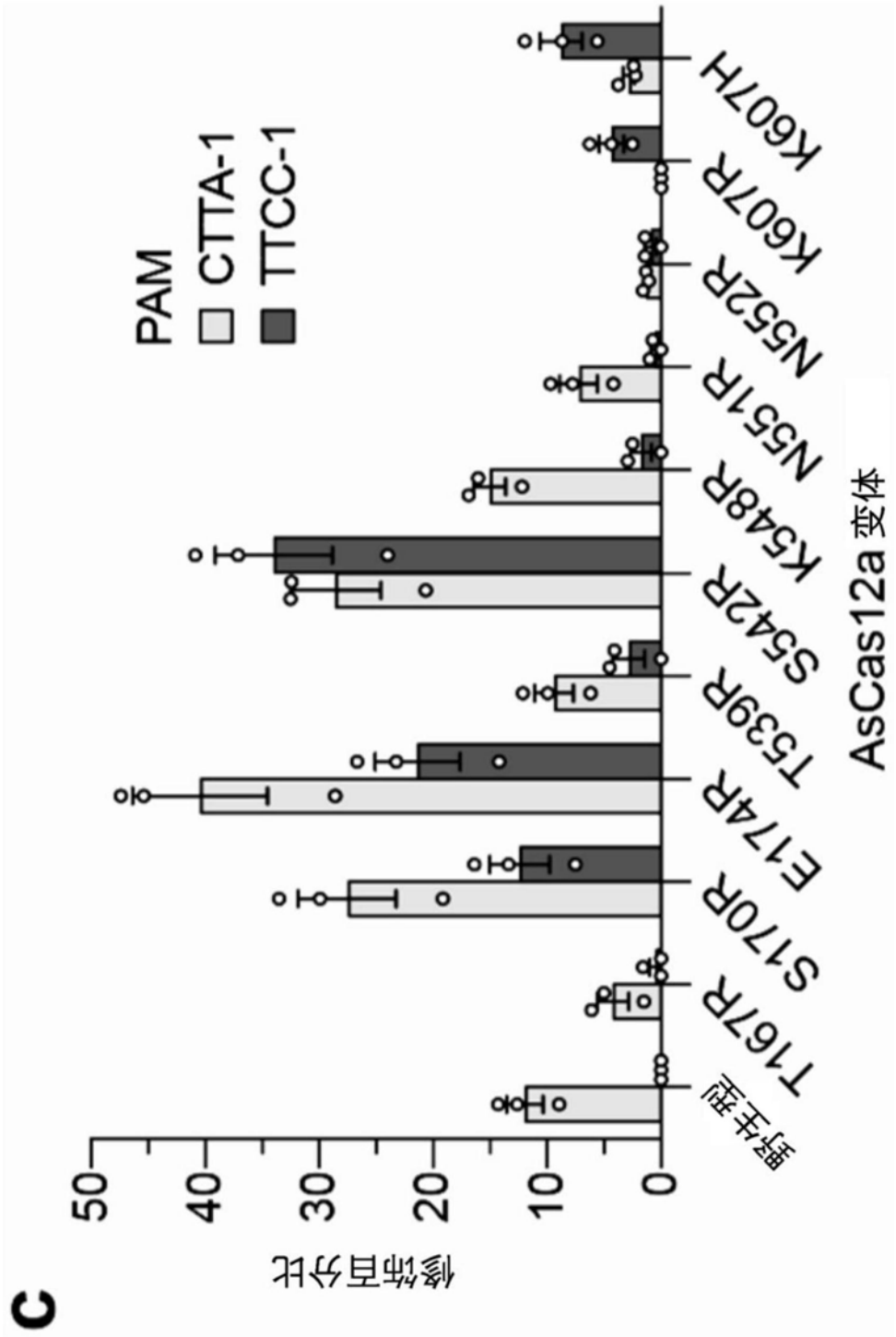

C
修饰百分比
50
40
30
20
10
0
野生型
T167R
S170R
E174R
T539R
S542R
K548R
N551R
N552R
K607R
K607H
AsCas12a 变体
PAM
CTTA-1
TTCC-1

图20C

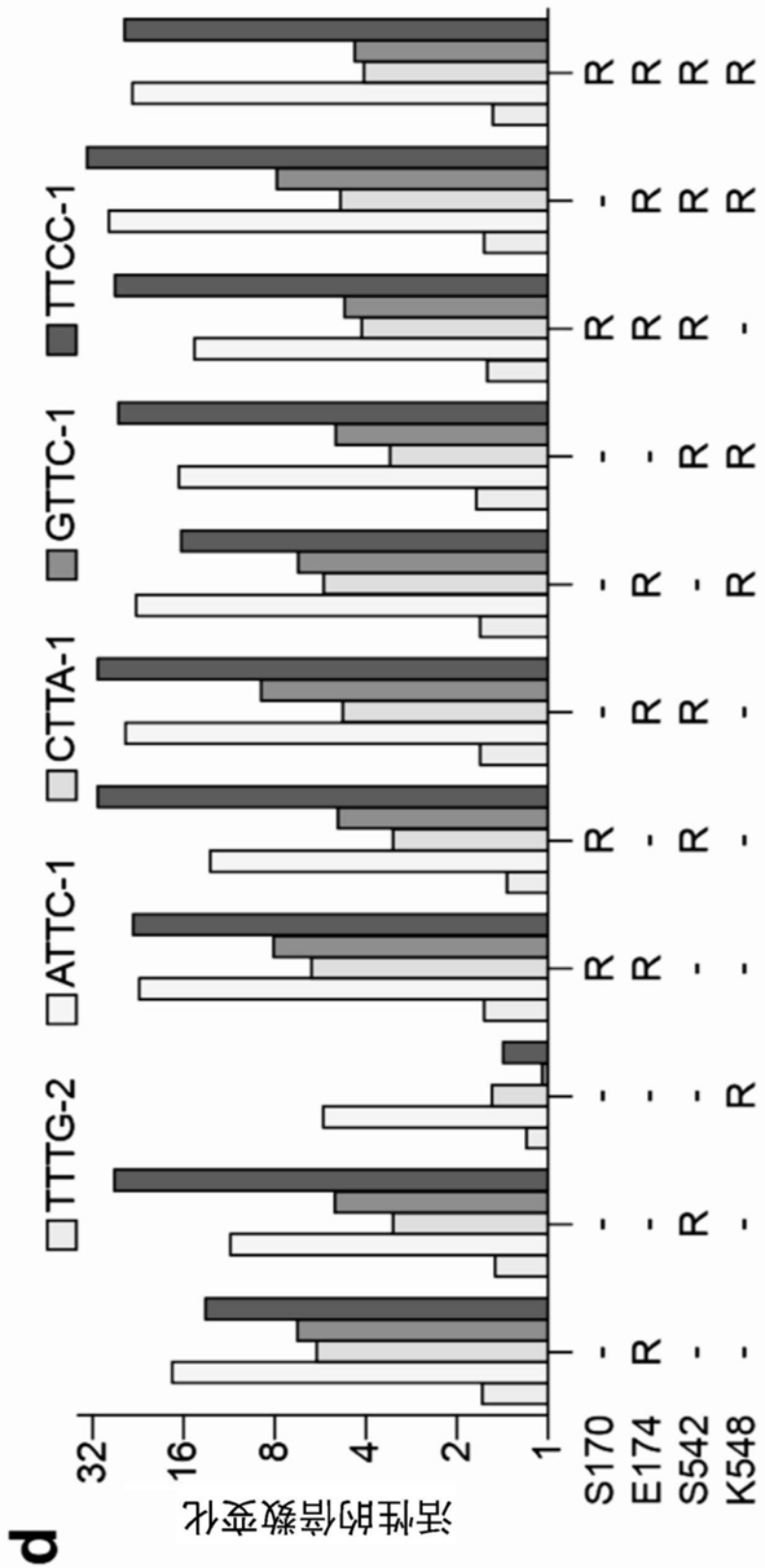

d
活性的倍数变化
32
16
8
4
2
1
S170
E174
S542
K548
TTTG-2
ATTC-1
CTTA-1
GTTC-1
TTCC-1

图20D

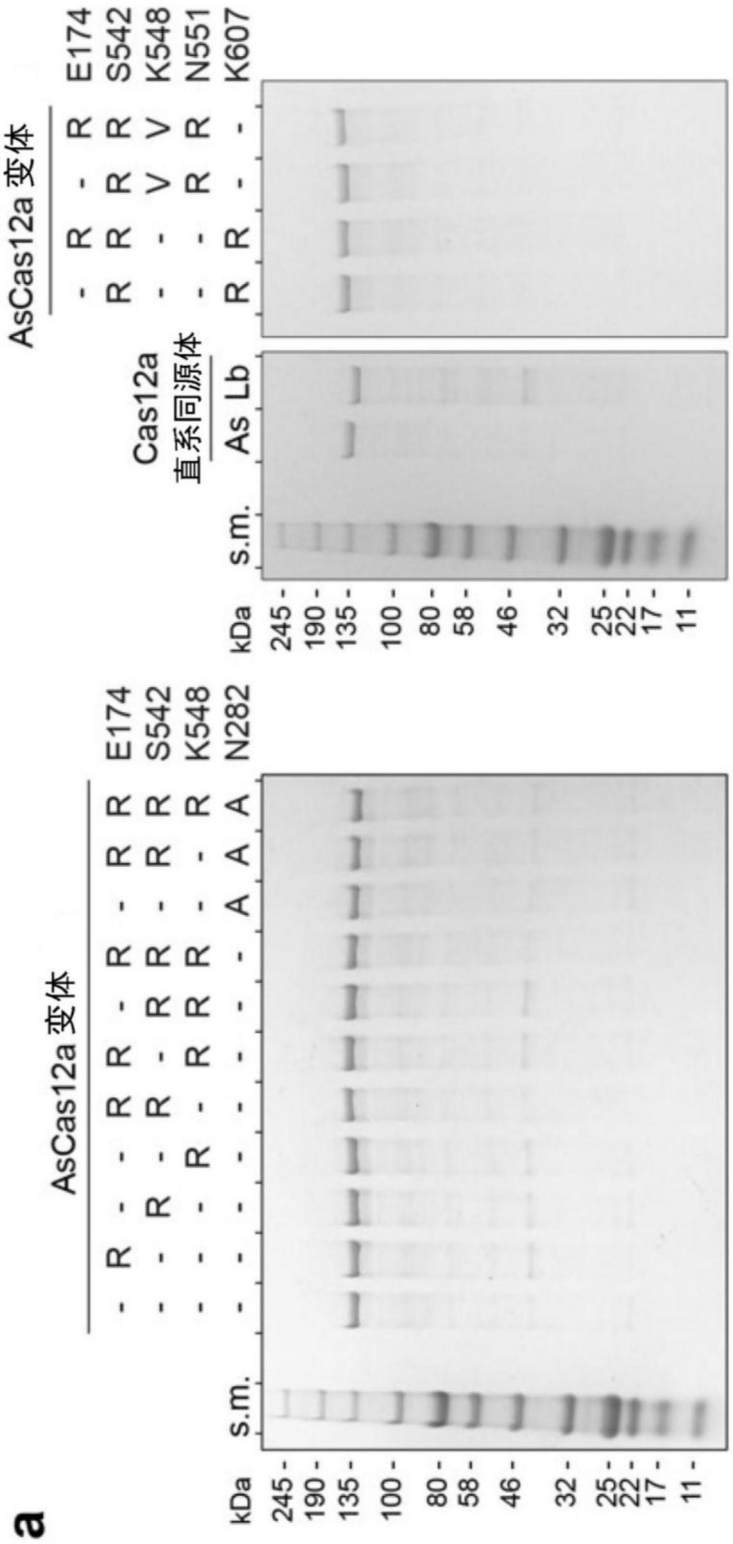

a
AsCas12a 变体
- R - - R R - R - R R
- - - R - R - R R - R
- - - - R - R R R - -
- - - - - - - - A A A
s.m.
kDa
245
190
135
100
80
58
46
32
25
22
17
11
E174
S542
K548
N282
Cas12a
直系同源体
As Lb
s.m.
kDa
245
190
135
100
80
58
46
32
25
22
17
11
AsCas12a 变体
- R - - R R - R - R R
R R - R - R R - R - -
- - R - R - - V V - -
R R - - R R - - - -
E174
S542
K548
N551
K607

图21A

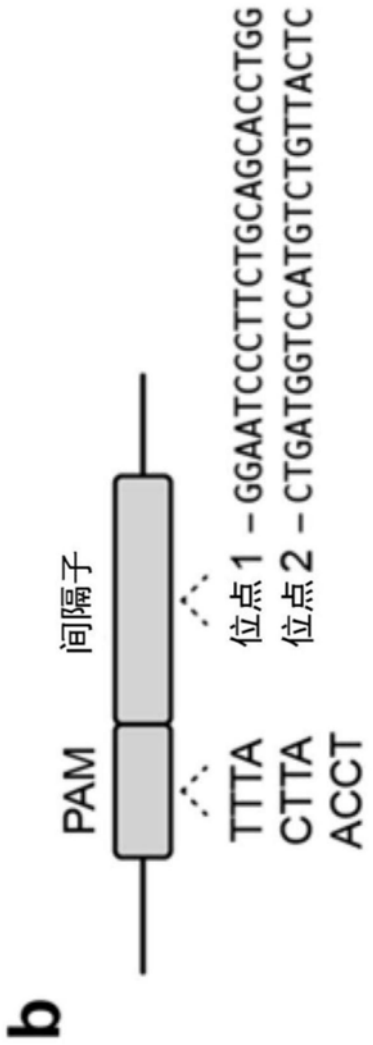

b
PAM
间隔子
TTTA
CTTA
ACCT
位点1 – GGAATCCCTTCTGCAGCACCTGG
位点2 – CTGATGGTCCATGTCTGTTACTC

图21B

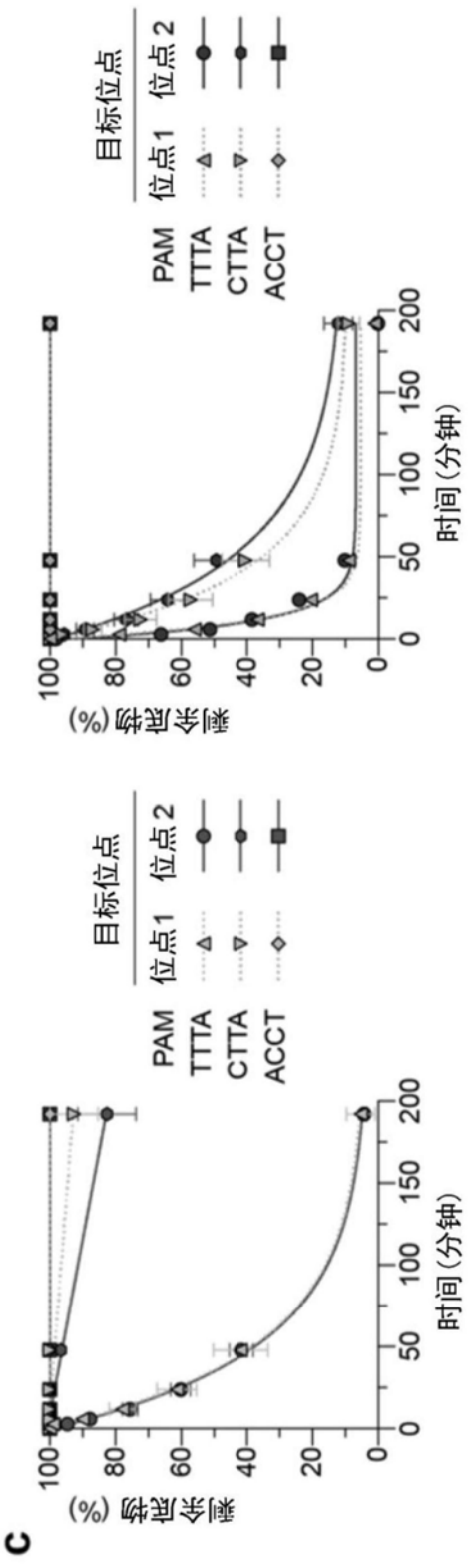

C
剩余底物 (%)
时间 (分钟)
0
50
100
150
200
20
40
60
80
100
目标位点
位点1
位点2
PAM
TTTA
CTTA
ACCT

图21C

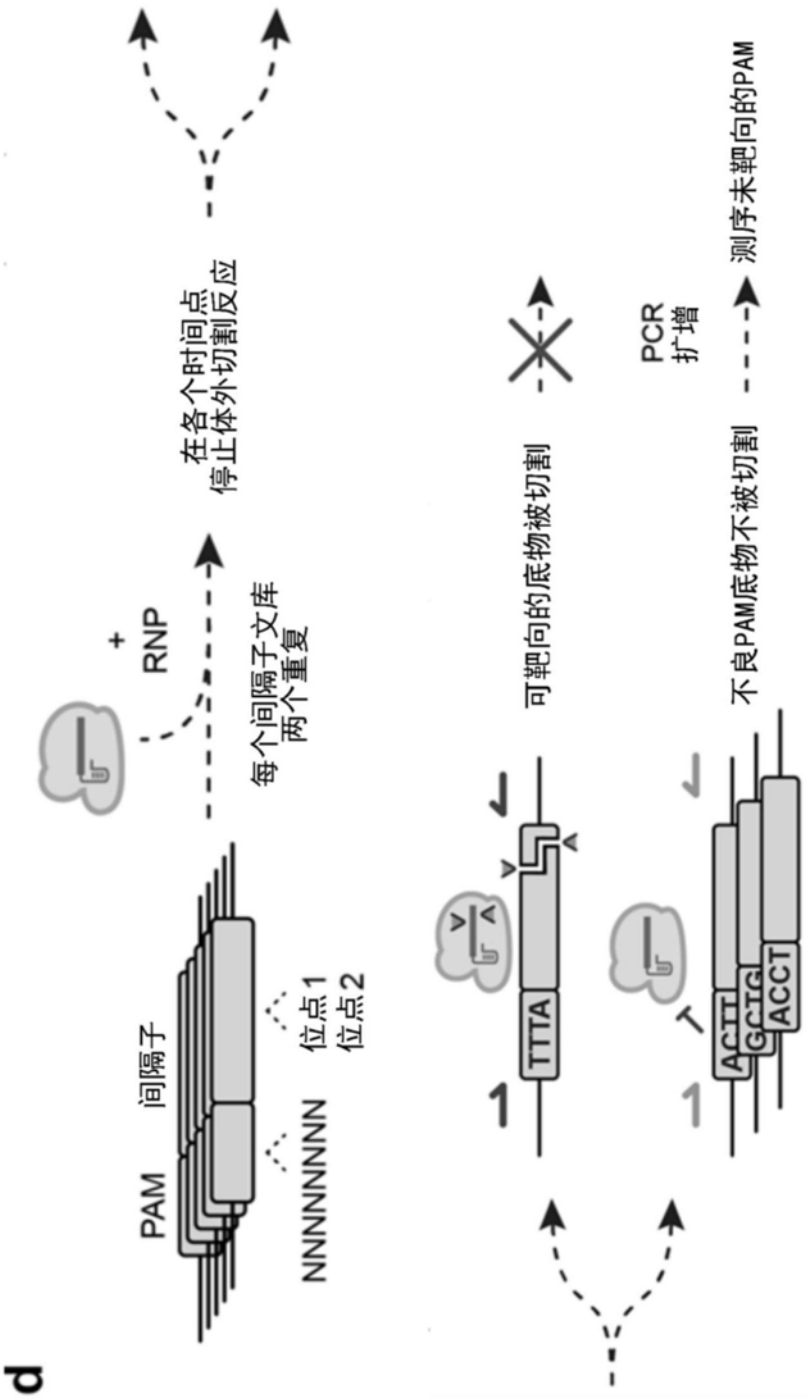

d
PAM
间隔子
NNNNNNNN
位点1
位点2
每个间隔子文库
两个重复
+
RNP
在各个时间点
停止体外切割反应
可靶向的底物被切割
TTTA
不良PAM底物不被切割
ACTT
GCTG
ACCT
PCR
扩增
测序未靶向的PAM

图21D

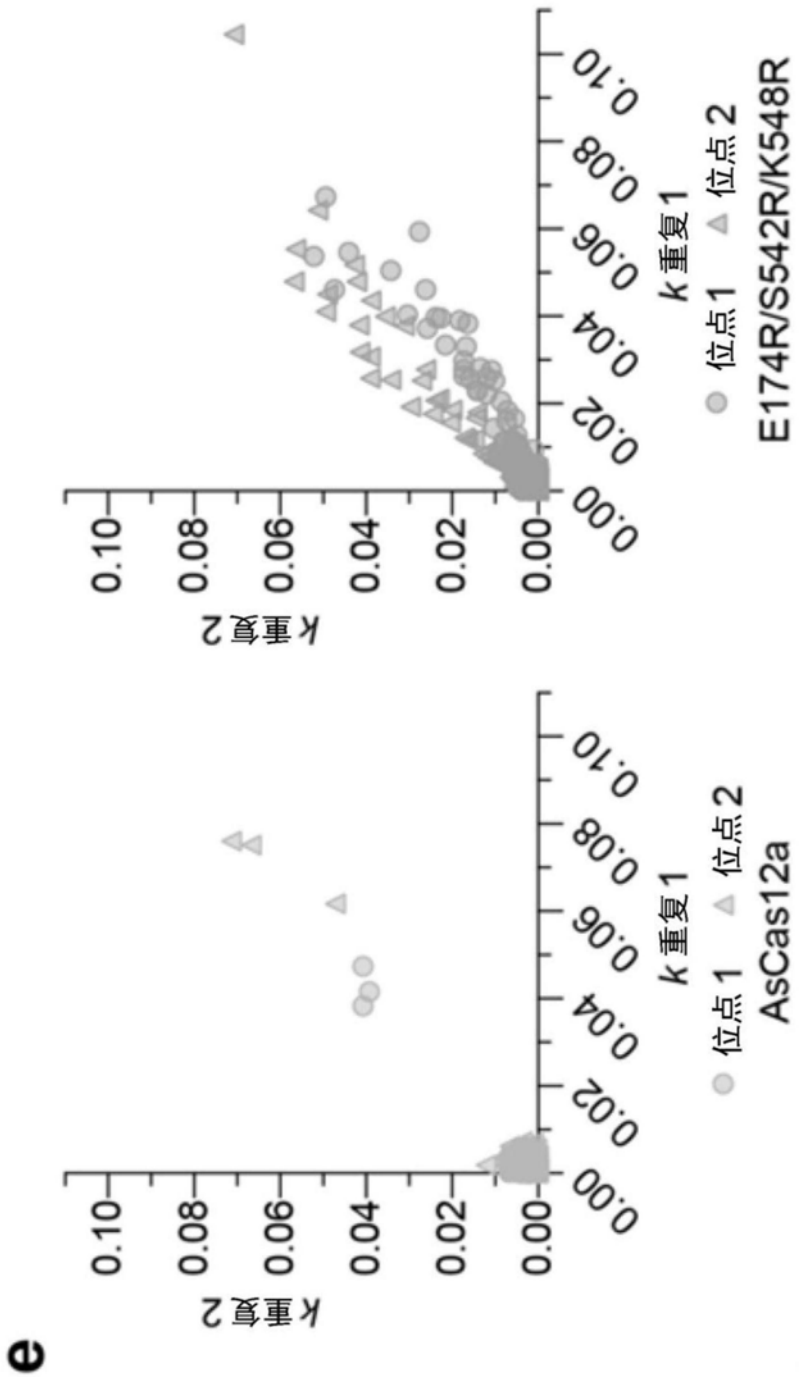

e
k重复2
k重复1
位点1
位点2
AsCas12a
E174R/S542R/K548R
0.00
0.02
0.04
0.06
0.08
0.10

图21E

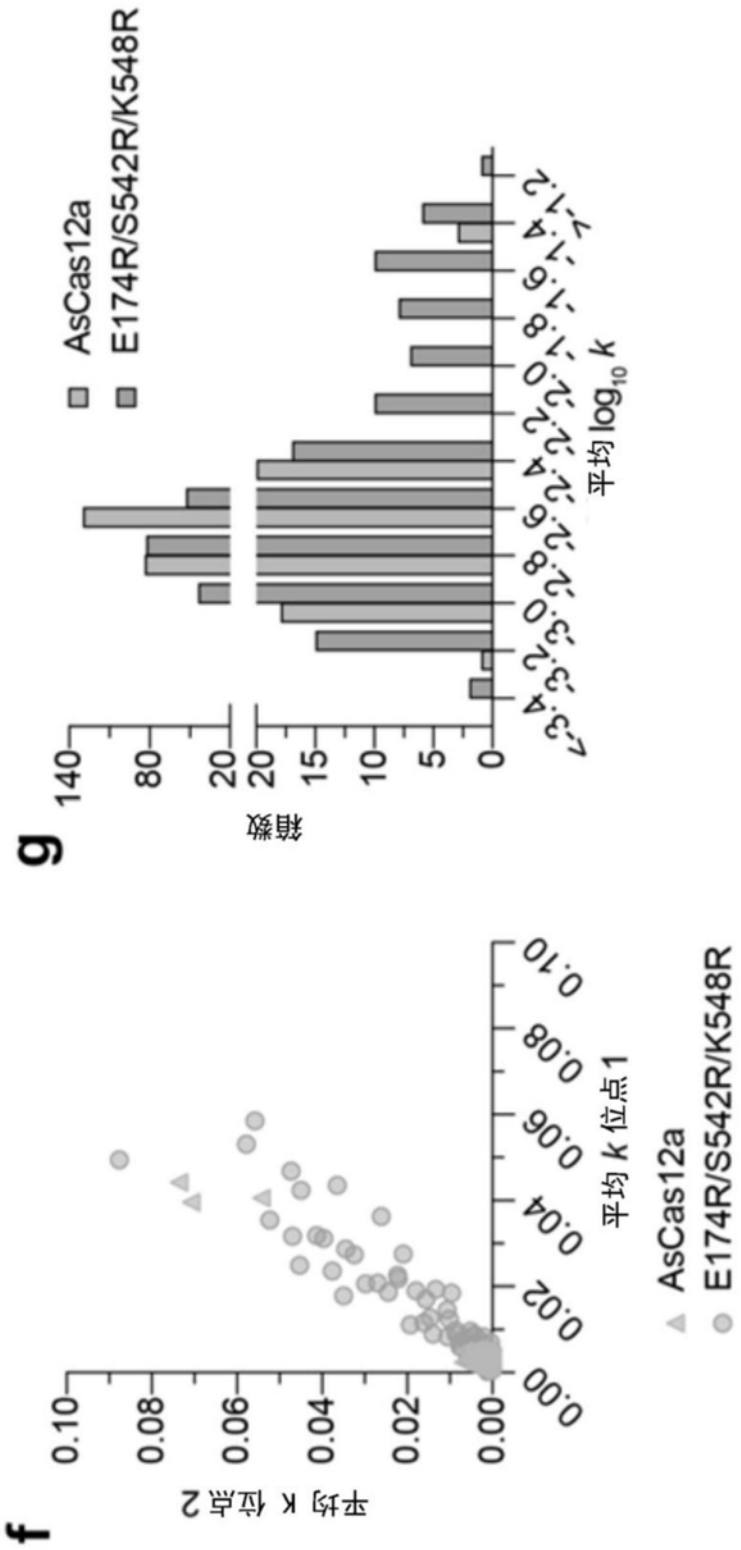

f
平均 k 位点 2
平均 k 位点 1
AsCas12a
E174R/S542R/K548R
g
箱数
平均 log10 k
AsCas12a
E174R/S542R/K548R

图21F-G

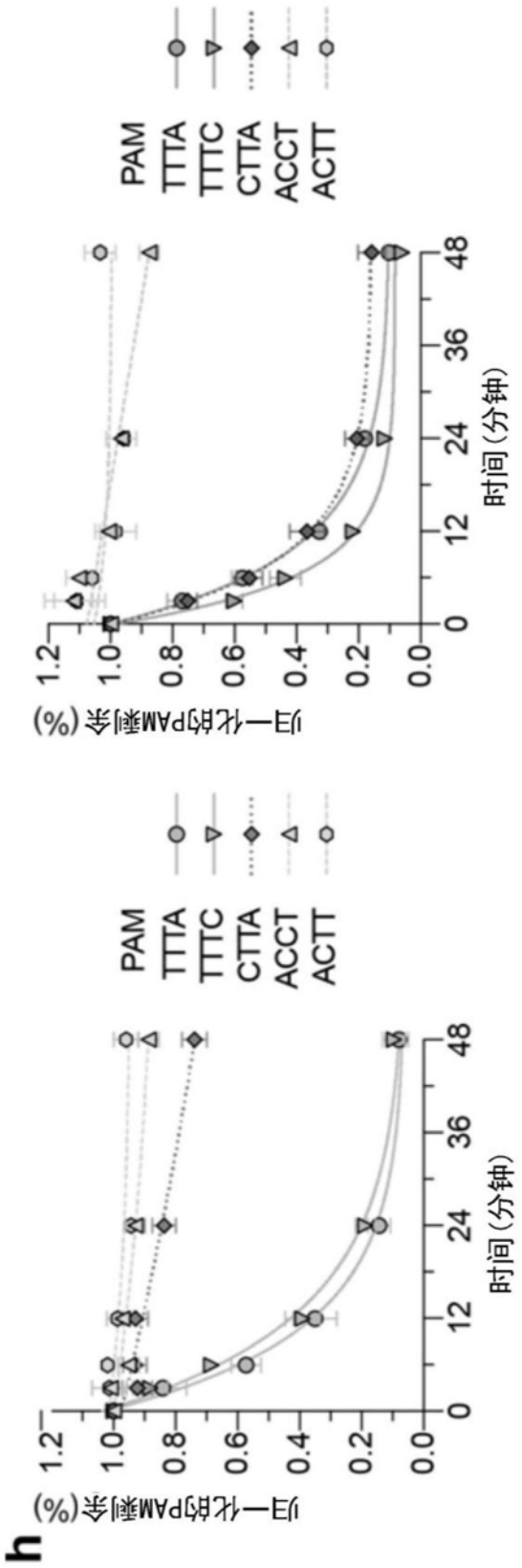

h
归一化的PAM剩余(%)
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
12
24
36
48
时间(分钟)
PAM
TTTA
TTTC
CTTA
ACCT
ACTT

图21H

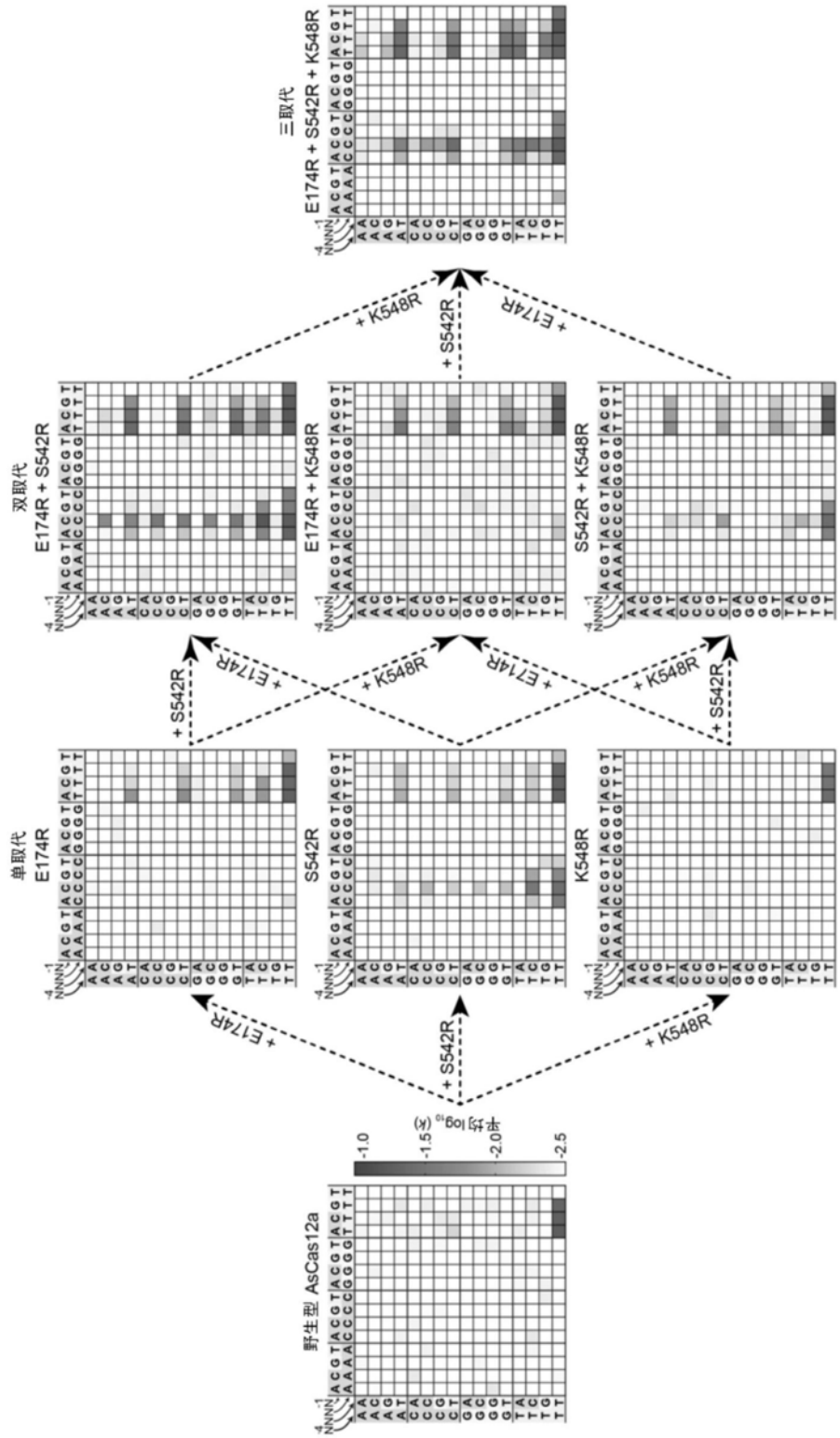

三取代
E174R + S542R + K548R
双取代
E174R + S542R
E174R + K548R
S542R + K548R
单取代
E174R
S542R
K548R
野生型 AsCas12a
+ K548R
+ S542R
+ E174R
+ S542R
+ E174R
+ K548R
+ E174R
+ K548R
+ S542R
+ S542R
+ K548R
平均 log10 (k)
-1.0
-1.5
-2.0
-2.5

图22A

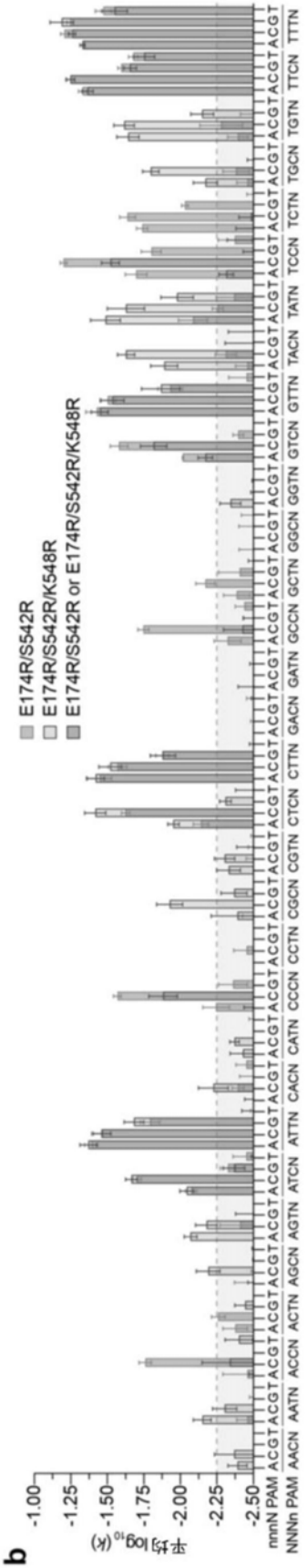

b
平均log10(k)
-1.00
-1.25
-1.50
-1.75
-2.00
-2.25
-2.50
E174R/S542R
E174R/S542R/K548R
E174R/S542R or E174R/S542R/K548R
nnnN PAM ACGTACGTACGTACGTACGTACGTACGTACGTACGTACGTACGTACGTACGTACGTACGTACGTACGTACGTACGTACGT
NNNn PAM AACN AATN ACCN ACTN AGCN AGTN ATCN ATTN CACN CATN CCCN CCTN CGCN CGTN CTCN CTTN GACN GATN GCCN GCTN GGCN GGTN GTCN GTTN TACN TATN TCCN TCTN TGCN TGTN TTCN TTTN

图22B

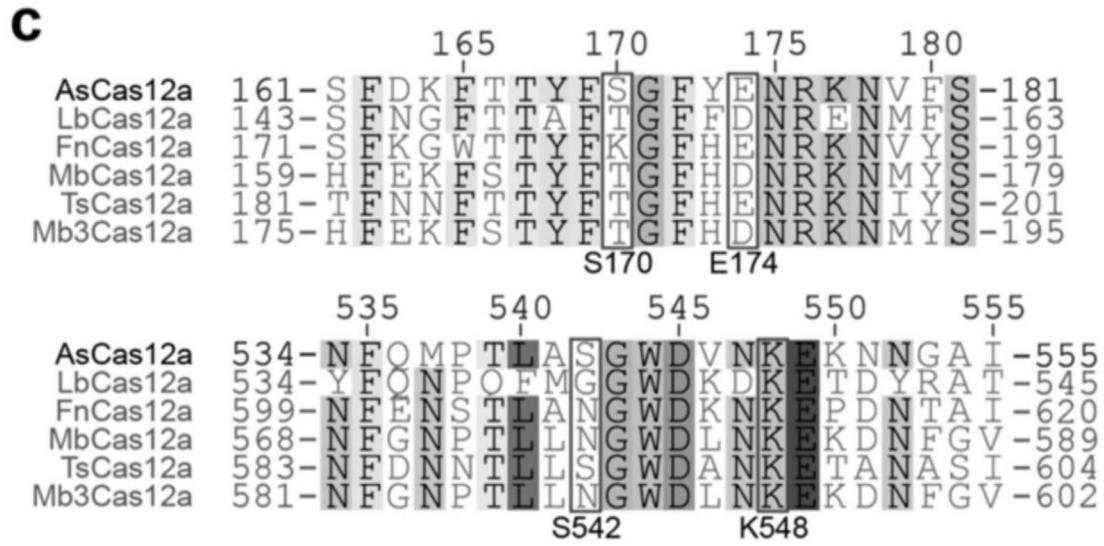

C
165
170
175
180
AsCas12a 161-SFDKFTTYFSGFYENRKNVFS-181
LbCas12a 143-SFNGFTTAFTGFFDNRENMFS-163
FnCas12a 171-SFKGWTTYFKGFHENRKNVYS-191
MbCas12a 159-HFEKFSTYFTGFHDNRKNMYS-179
TsCas12a 181-TFNNFTTYFTGFHENRKNIYS-201
Mb3Cas12a 175-HFEKFSTYFTGFHDNRKNMYS-195
S170
E174
535
540
545
550
555
AsCas12a 534-NFQMPTLASGWDVNKEKNNGAI-555
LbCas12a 534-YFQNPQFMGGWDKDKETDYRAT-545
FnCas12a 599-NFENSTLANGWDKNKEPDNTAI-620
MbCas12a 568-NFGNPTLLNGWDLNKEKDNFGV-589
TsCas12a 583-NFDNNTLLSGWDANKETANASI-604
Mb3Cas12a 581-NFGNPTLLNGWDLNKEKDNFGV-602
S542
K548

图22C

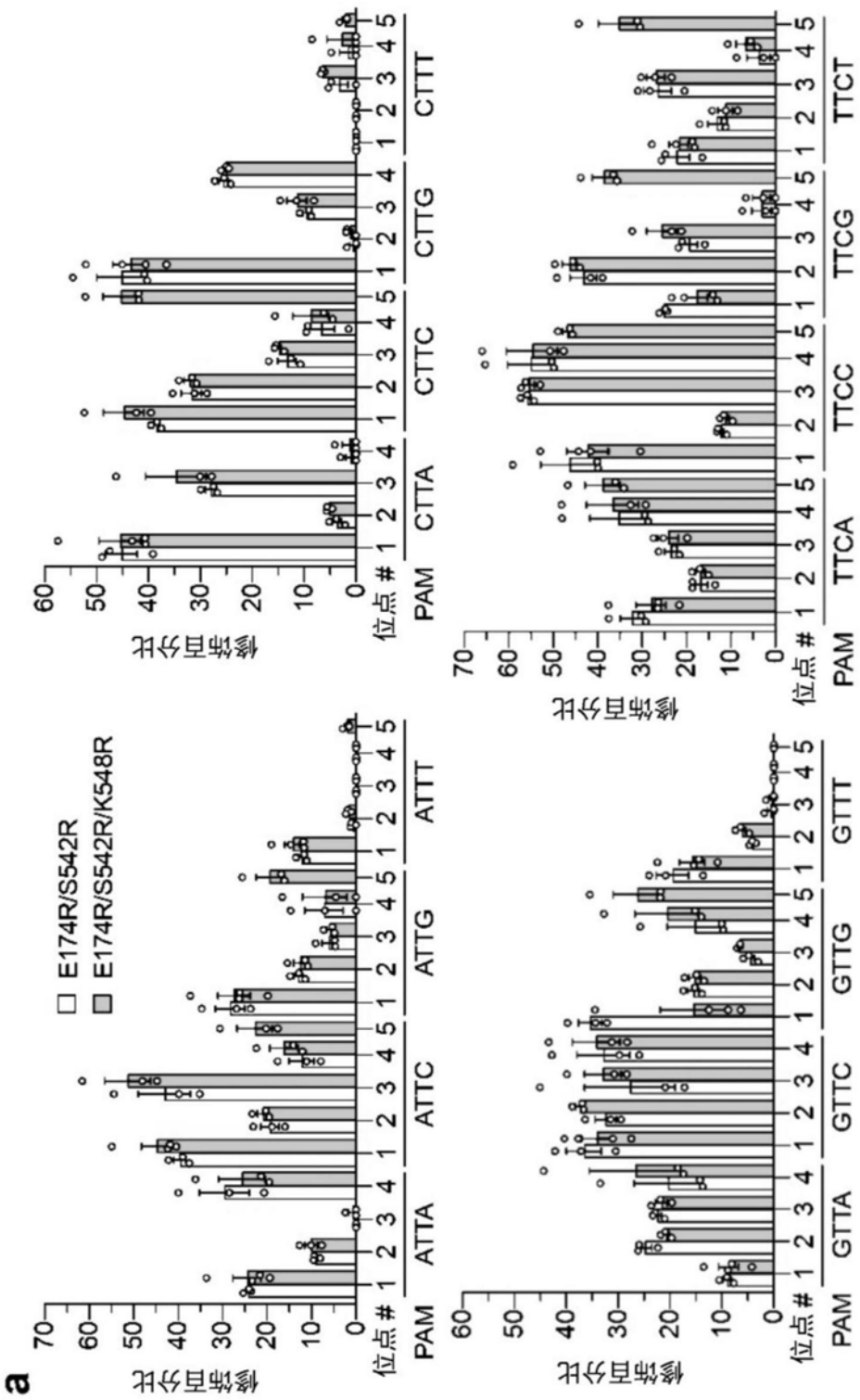

a
E174R/S542R
E174R/S542R/K548R
修饰百分比
位点 #
PAM
ATTA
ATTC
ATTG
ATTT
CTTA
CTTC
CTTG
CTTT
GTTA
GTTC
GTTG
GTTT
TTCA
TTCC
TTCG
TTCT

图23A

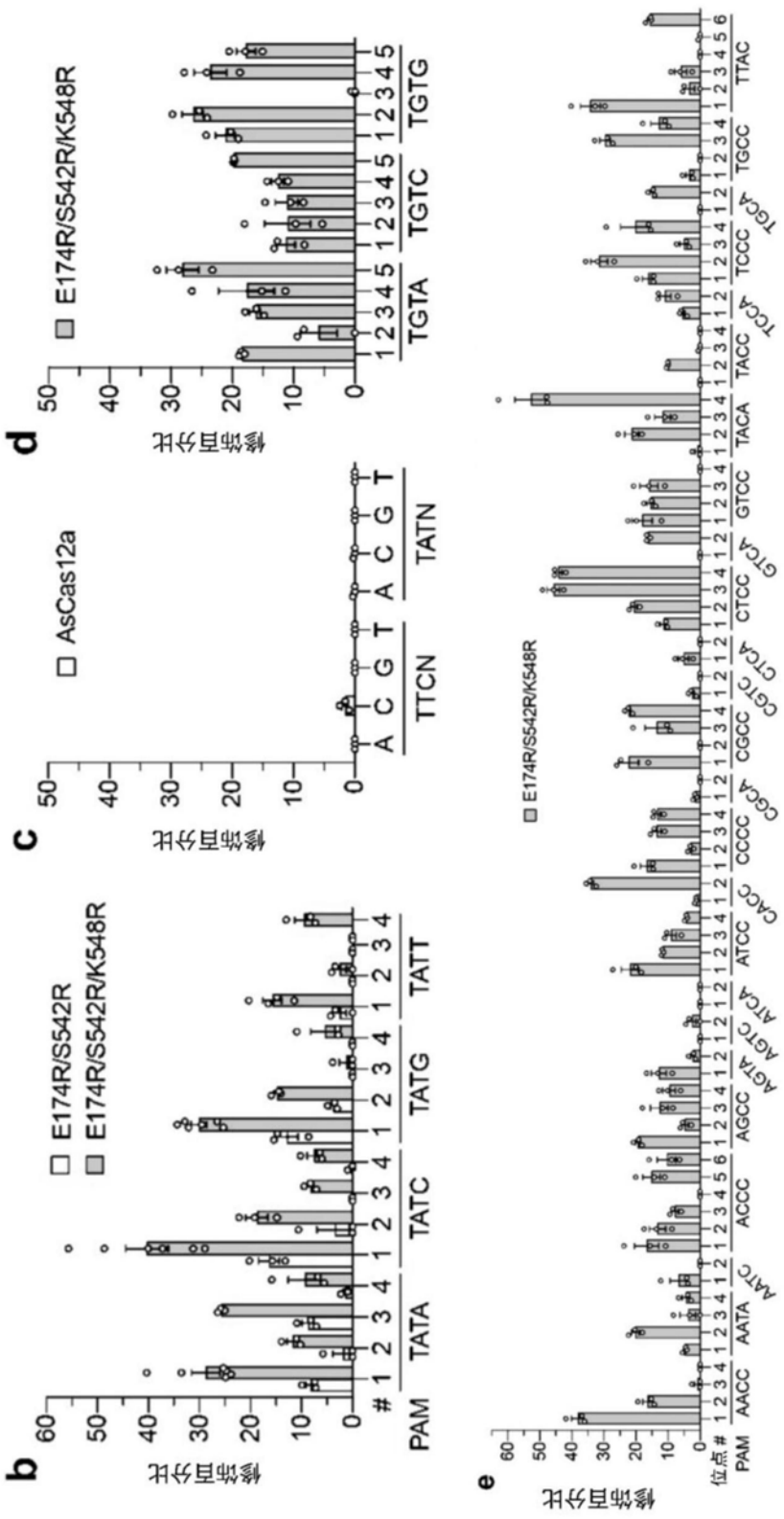

b
E174R/S542R
E174R/S542R/K548R
修饰百分比
PAM
TATA
TATC
TATG
TATT
c
AsCas12a
修饰百分比
TTCN
TATN
d
E174R/S542R/K548R
修饰百分比
TGTA
TGTC
TGTG
e
E174R/S542R/K548R
修饰百分比
位点#
PAM
AACC
AATA
AATC
ACCC
AGCC
AGTA
AGTC
ATCA
ATCC
CACC
CCCC
CGCA
CGCC
CGTC
CTCA
CTCC
GTCA
GTCC
TACA
TACC
TCCA
TCCC
TGCA
TGCC
TTAC

图23B-E

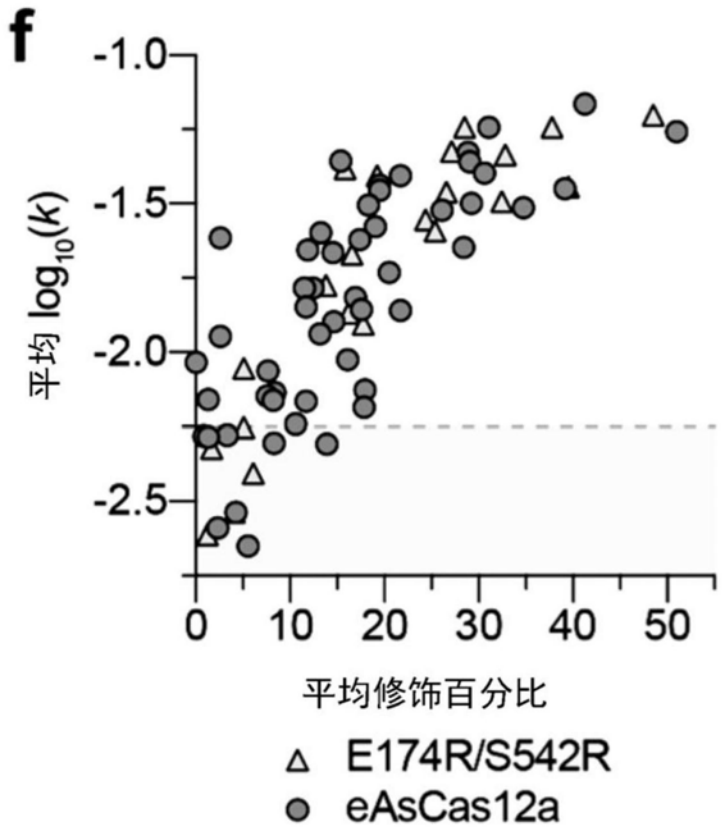

f
-1.0
-1.5
-2.0
-2.5
平均 log10(k)
0
10
20
30
40
50
平均修饰百分比
E174R/S542R
eAsCas12a

图23F

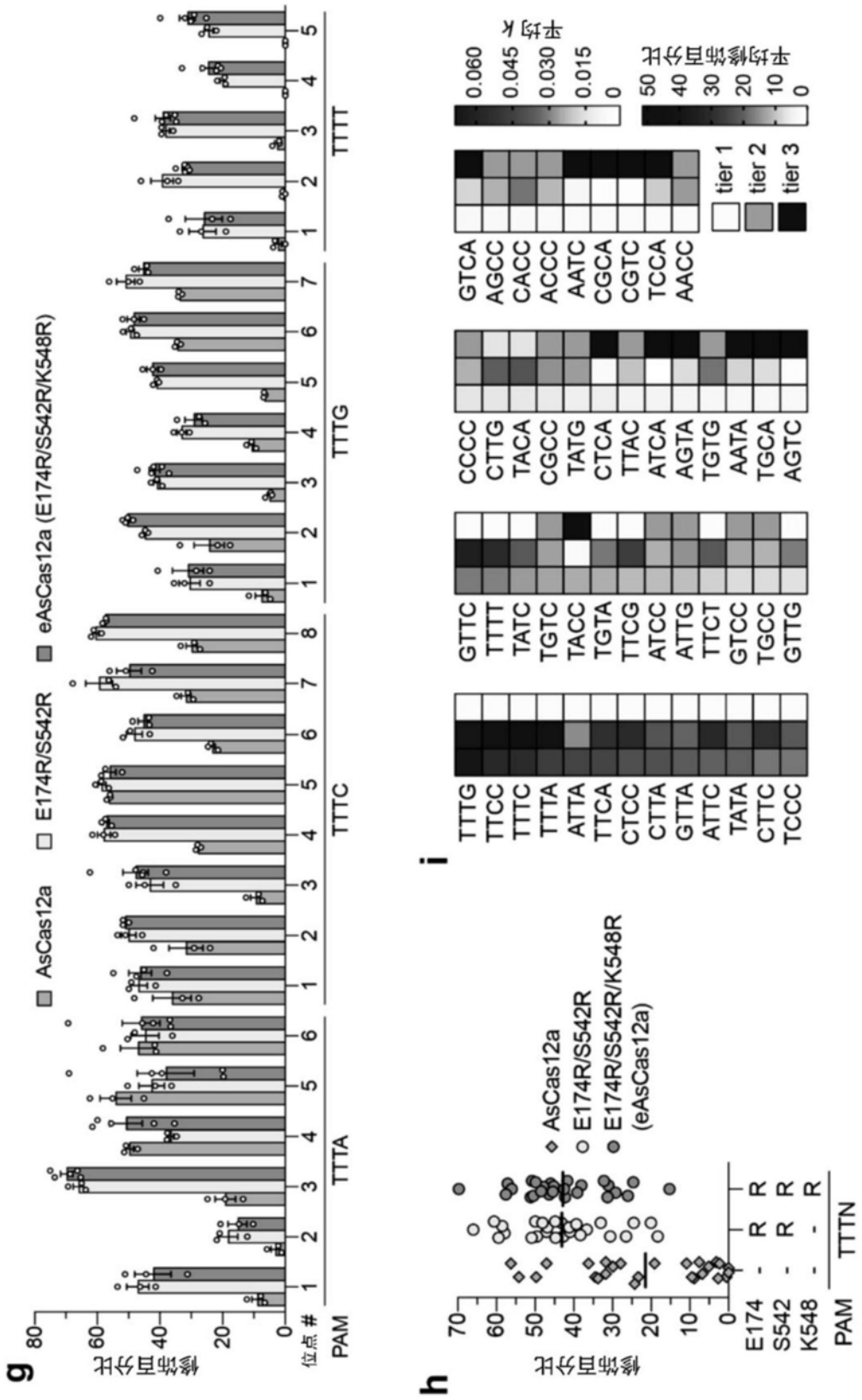

g
AsCas12a
E174R/S542R
eAsCas12a (E174R/S542R/K548R)
修饰百分比
位点 #
PAM
TTTA
TTTC
TTTG
TTTT
i
平均修饰百分比
平均 k
tier 1
tier 2
tier 3
h
AsCas12a
E174R/S542R
E174R/S542R/K548R (eAsCas12a)
修饰百分比
E174
S542
K548
PAM
TTTN

图23G-I

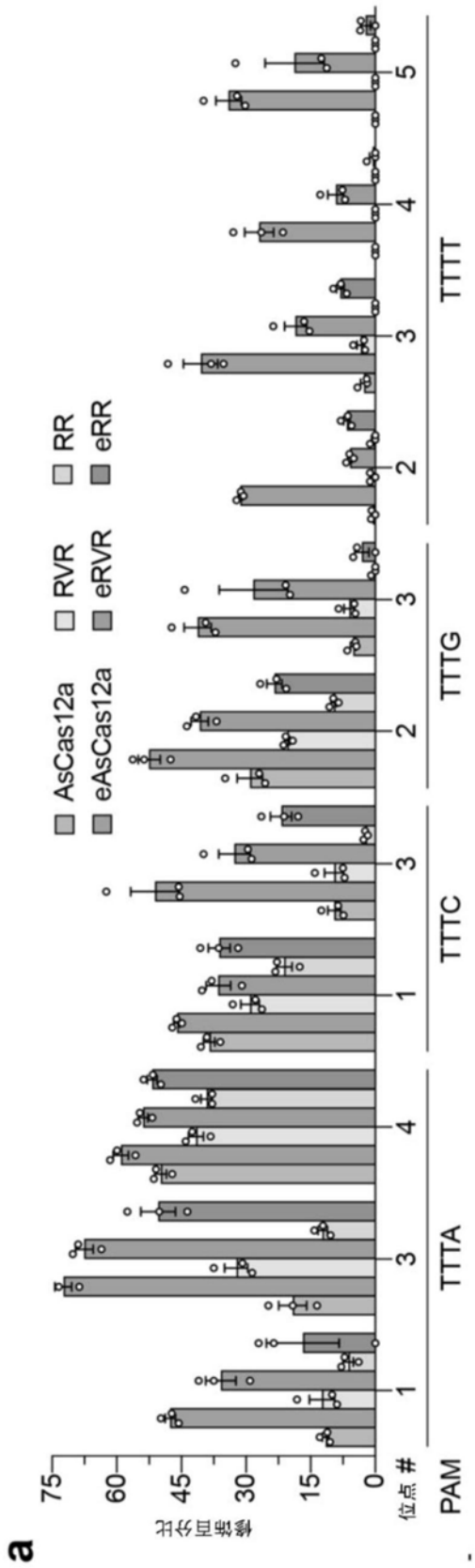

a
修饰百分比
75
60
45
30
15
0
位点 #
PAM
TTTA
TTTC
TTTG
TTTT
1
2
3
4
5
AsCas12a
eAsCas12a
RVR
eRVR
RR
eRR

图24A

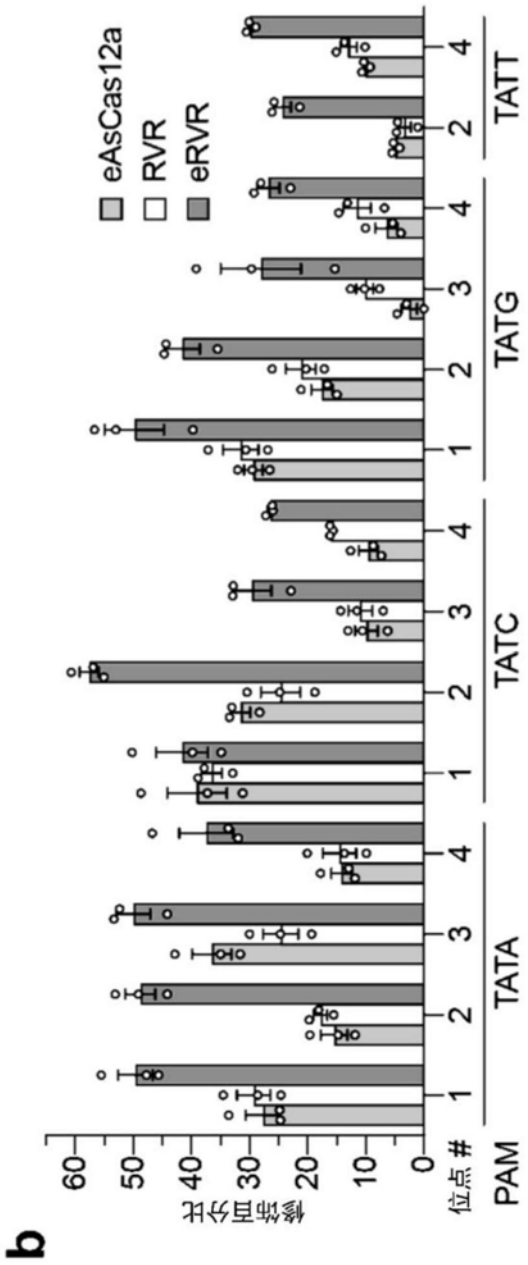

b
eAsCas12a
RVR
eRVR
修饰百分比
60
50
40
30
20
10
0
位点#
PAM
TATA
TATC
TATG
TATT
1
2
3
4

图24B

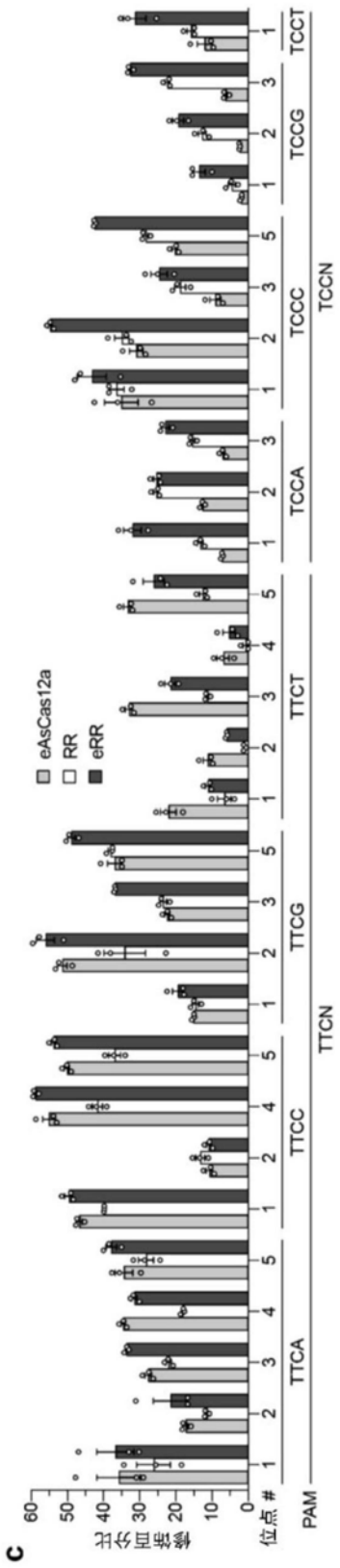

C
修饰百分比
60
50
40
30
20
10
0
eAsCas12a
RR
eRR
位点 #
PAM
TTCA
TTCC
TTCG
TTCT
TTCN
TCCA
TCCC
TCCG
TCCT
TCCN

图24C

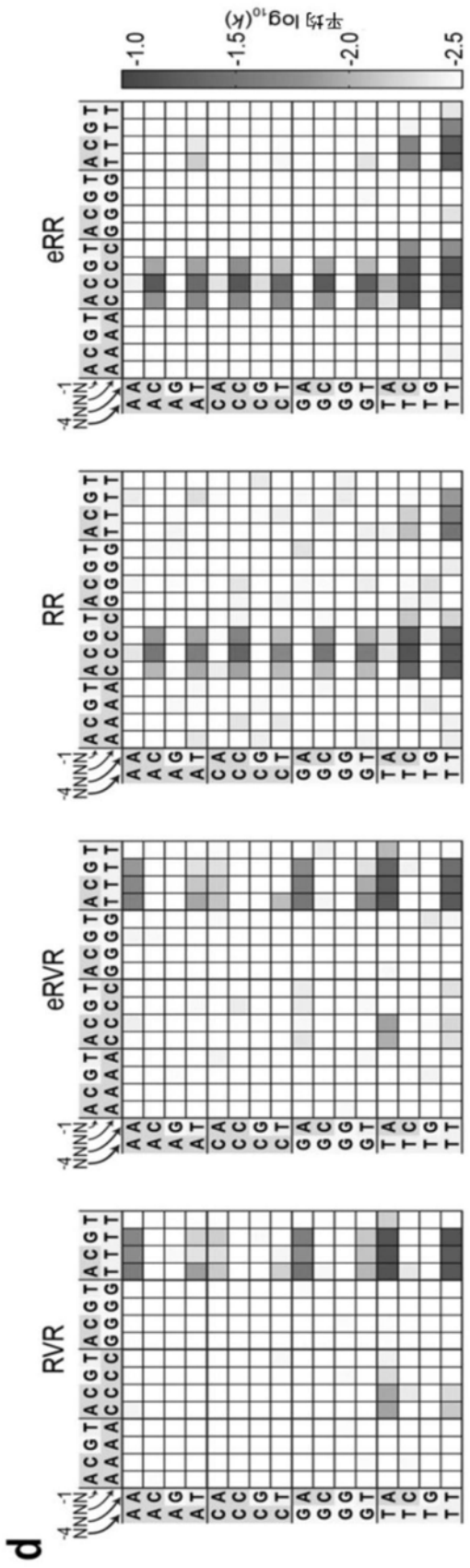


图24D

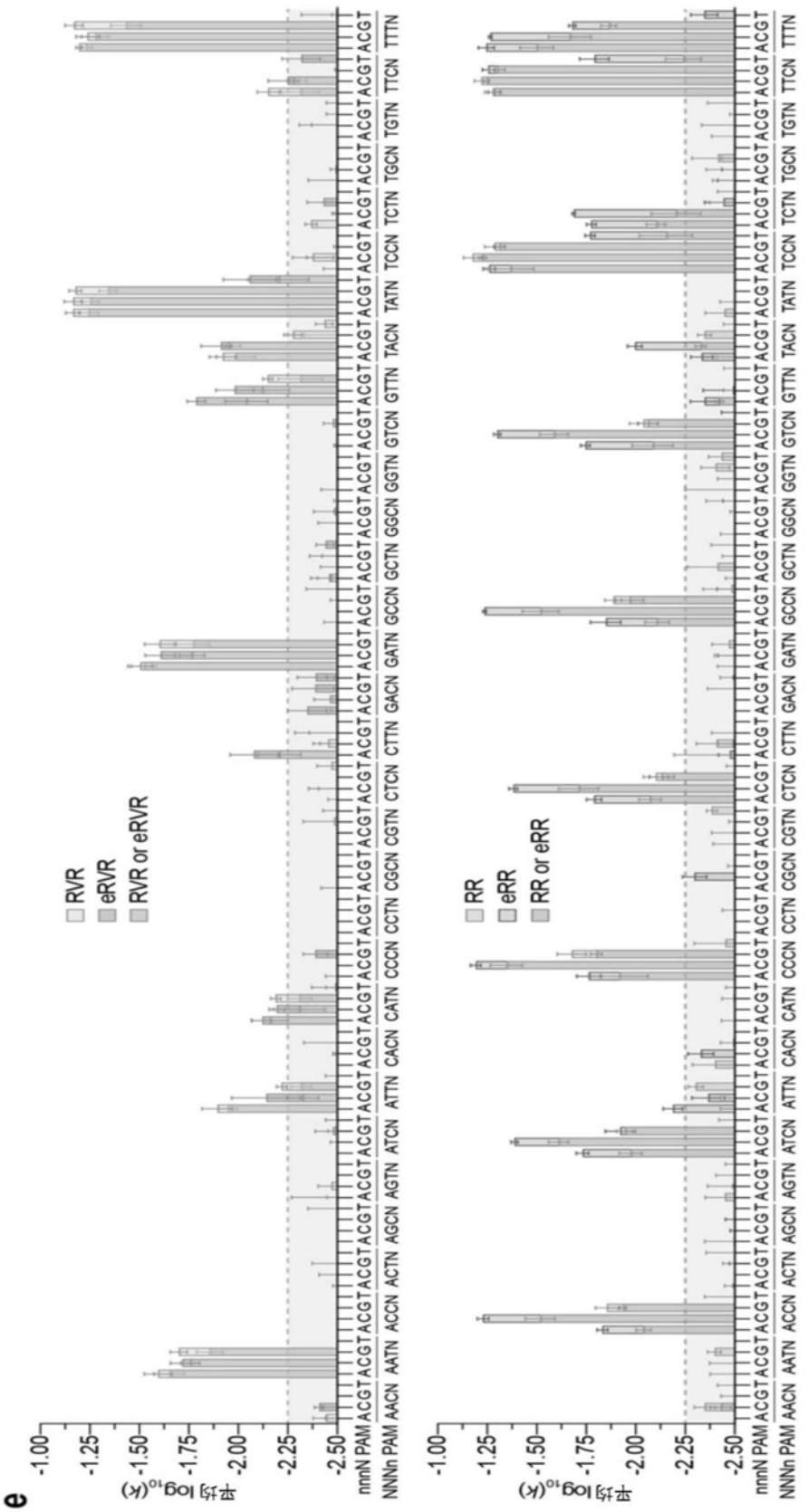

e
RVR
eRVR
RVR or eRVR
RR
eRR
RR or eRR
平均 log10(K)
-1.00
-1.25
-1.50
-1.75
-2.00
-2.25
-2.50
nnnN PAM ACGTACGTACGTACGTACGTACGTACGTACGTACGTACGTACGTACGTACGTACGTACGTACGTACGTACGTACGTACGTACGTACGTACGT
NNNn PAM AACN AATN ACCN ACTN AGCN AGTN ATCN ATTN CACN CATN CCCN CCTN CGCN CGTN CTCN CTTN GACN GATN GCCN GCTN GGCN GGTN GTCN GTTN TACN TATN TCCN TCTN TGCN TGTN TTCN TTTN

图24E

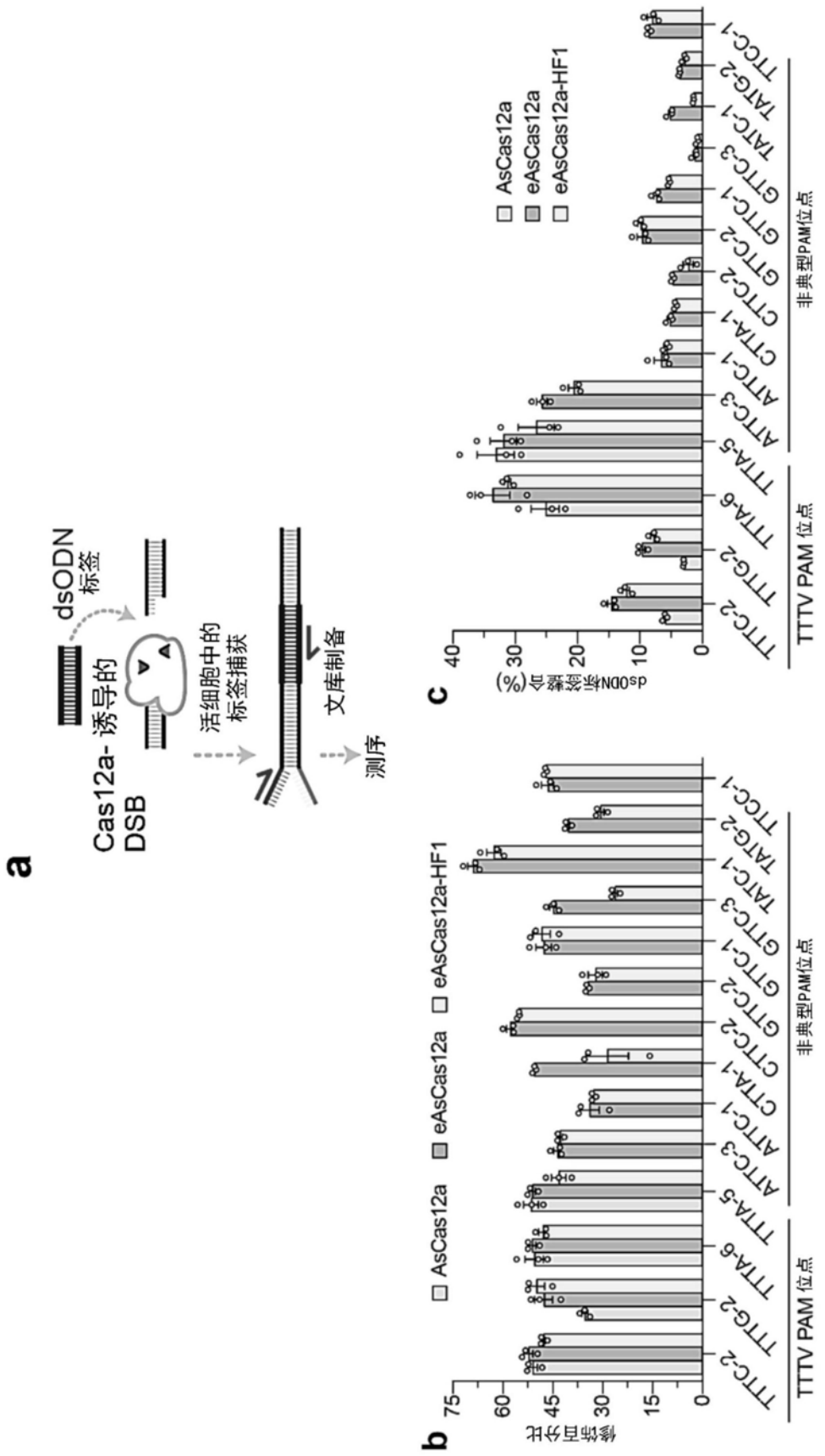

a
dsODN
标签
Cas12a-
诱导的
DSB
活细胞中的
标签捕获
文库制备
测序
b
AsCas12a
eAsCas12a
eAsCas12a-HF1
修饰百分比
TTTV PAM 位点
非典型PAM位点
TTTC-2
TTTG-2
TTTA-6
TTTA-5
ATTC-3
ATTC-1
CTTA-1
CTTC-2
GTTC-2
GTTC-1
GTTC-3
TATC-1
TATG-2
TTCC-1
c
AsCas12a
eAsCas12a
eAsCas12a-HF1
dsODN标签整合(%)
TTTV PAM 位点
非典型PAM位点

图25A-C

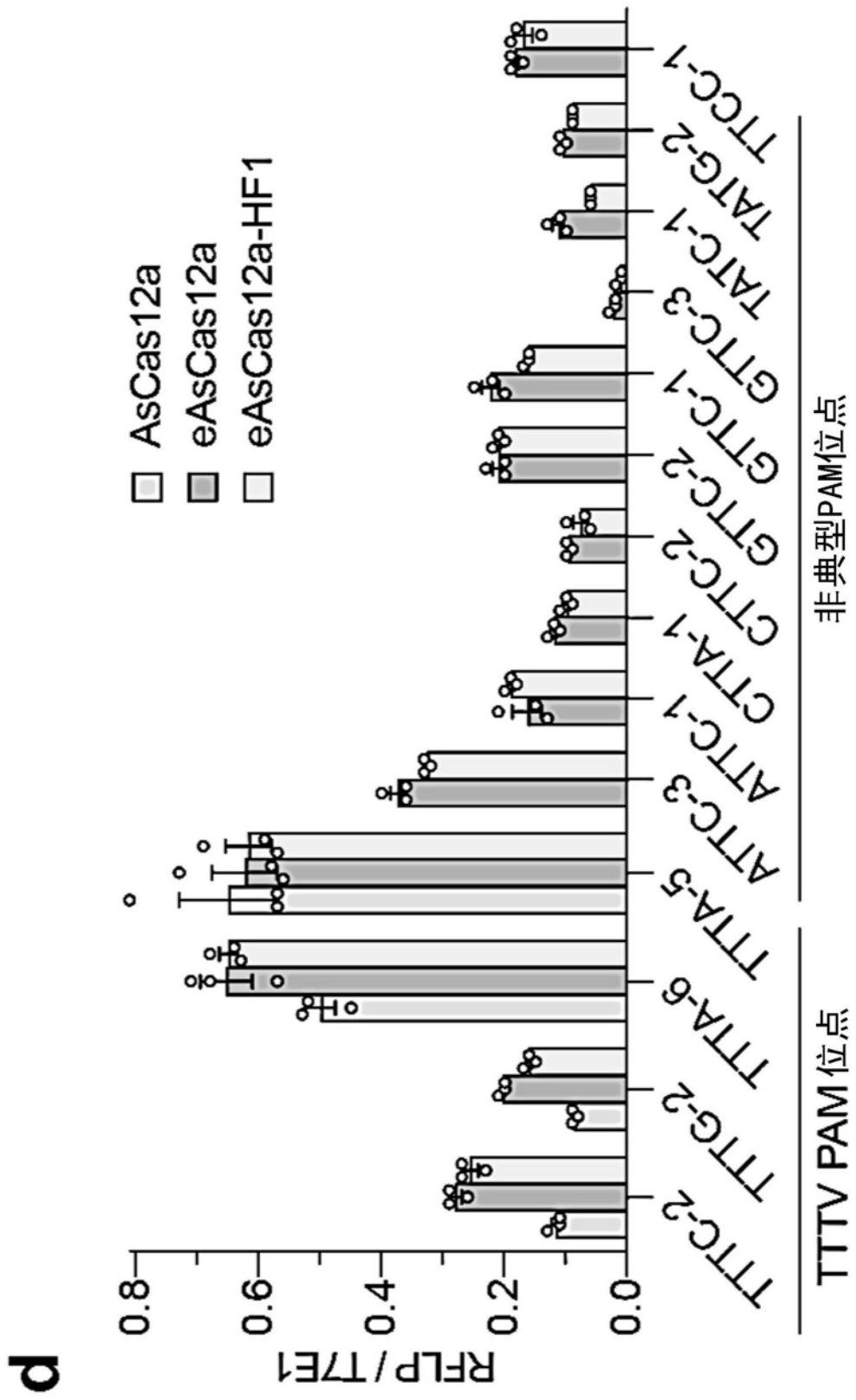

d
RFLP / T7E1
0.0
0.2
0.4
0.6
0.8
AsCas12a
eAsCas12a
eAsCas12a-HF1
TTTC-2
TTTG-2
TTTA-6
TTTA-5
ATTC-3
ATTC-1
CTTA-1
CTTC-2
GTTC-2
GTTC-1
GTTC-3
TATC-1
TATG-2
TTCC-1
TTTV PAM位点
非典型PAM位点

图25D

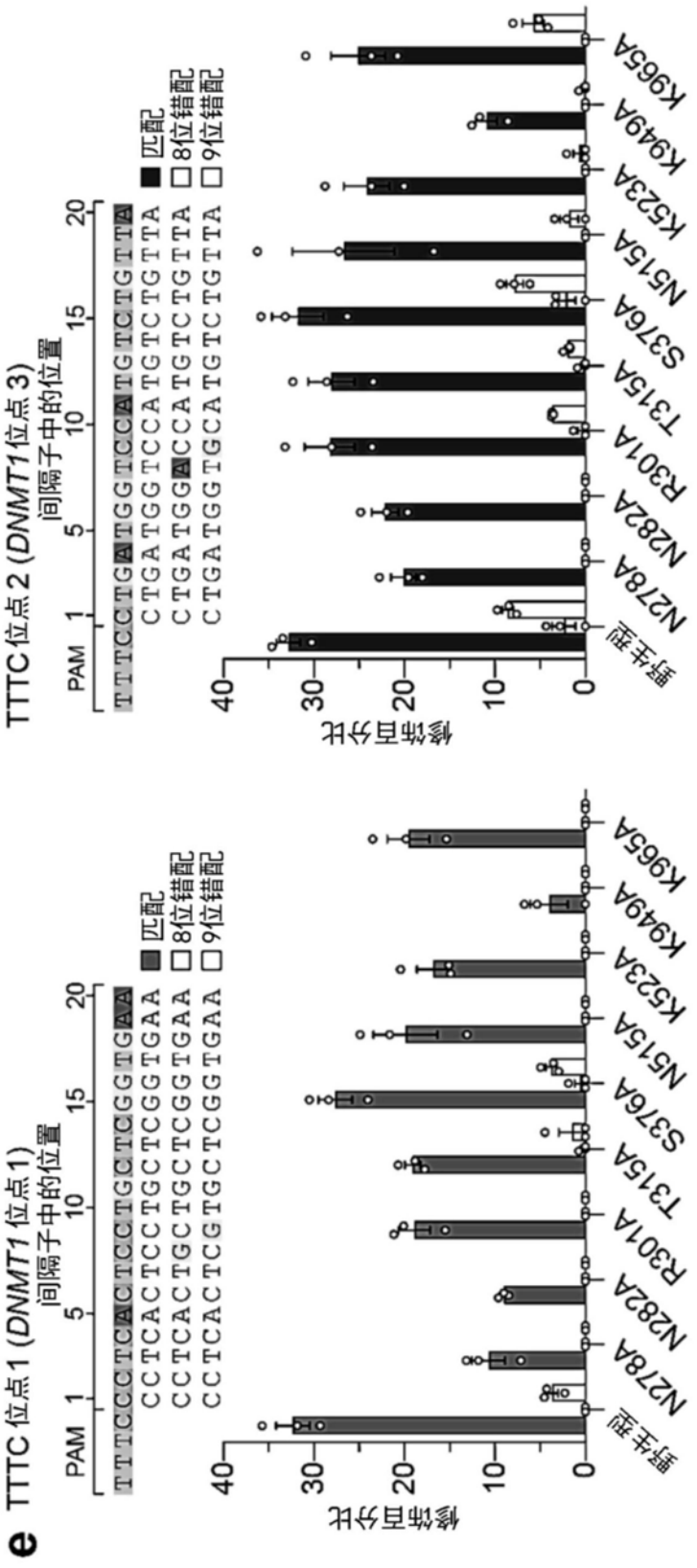

e
TTTC位点1 (DNMT1位点1)
间隔子中的位置
PAM
匹配
8位错配
9位错配
修饰百分比
野生型
N278A
N282A
R301A
T315A
S376A
N515A
K523A
K949A
K965A
TTTC位点2 (DNMT1位点3)

图25E

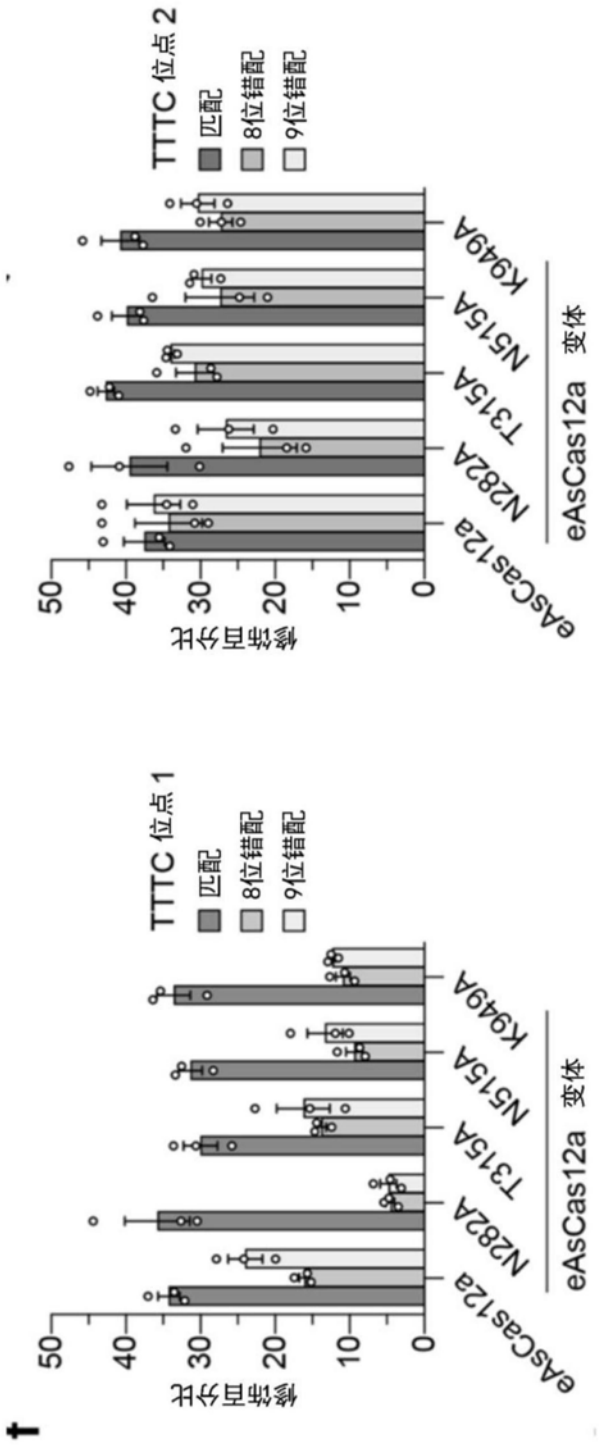

TTTC 位点 2
匹配
8位错配
9位错配
修饰百分比
50
40
30
20
10
0
eAsCas12a
N282A
T315A
N515A
K949A
eAsCas12a 变体
TTTC 位点 1
匹配
8位错配
9位错配
修饰百分比
50
40
30
20
10
0
eAsCas12a
N282A
T315A
N515A
K949A
eAsCas12a 变体

图25F

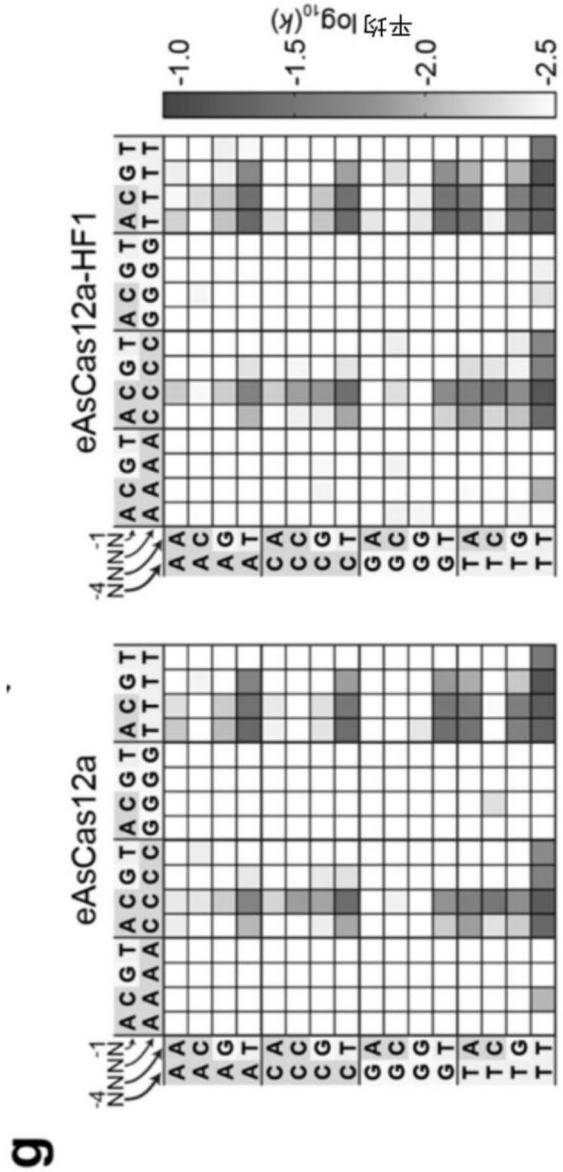


图25G

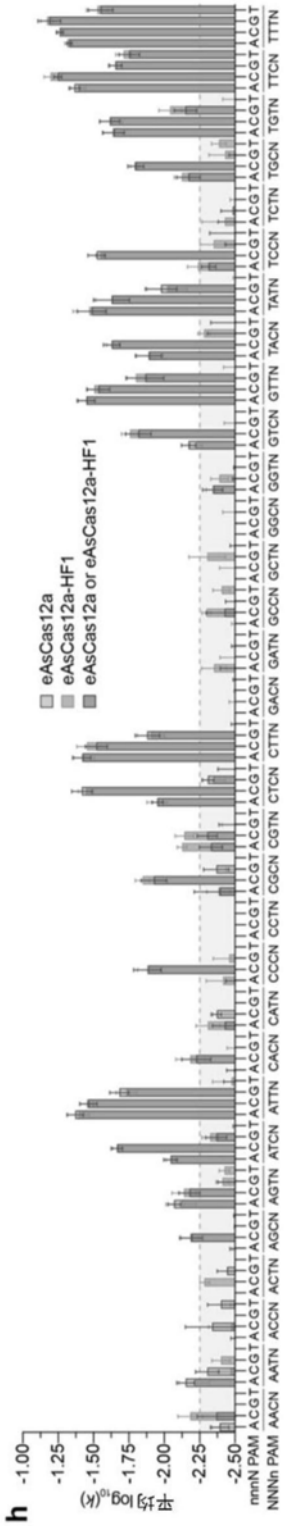

h
平均log10(K)
-1.00
-1.25
-1.50
-1.75
-2.00
-2.25
-2.50
eAsCas12a
eAsCas12a-HF1
eAsCas12a or eAsCas12a-HF1
nnnN PAM
NNNn PAM
AACN AATN ACCN ACTN AGCN AGTN ATCN ATTN CACN CATN CCCN CCTN CGCN CGTN CTCN CTTN GACN GATN GCCN GCTN GGCN GGTN GTCN GTTN TACN TATN TCCN TCTN TGCN TGTN TTCN TTTN

图25H

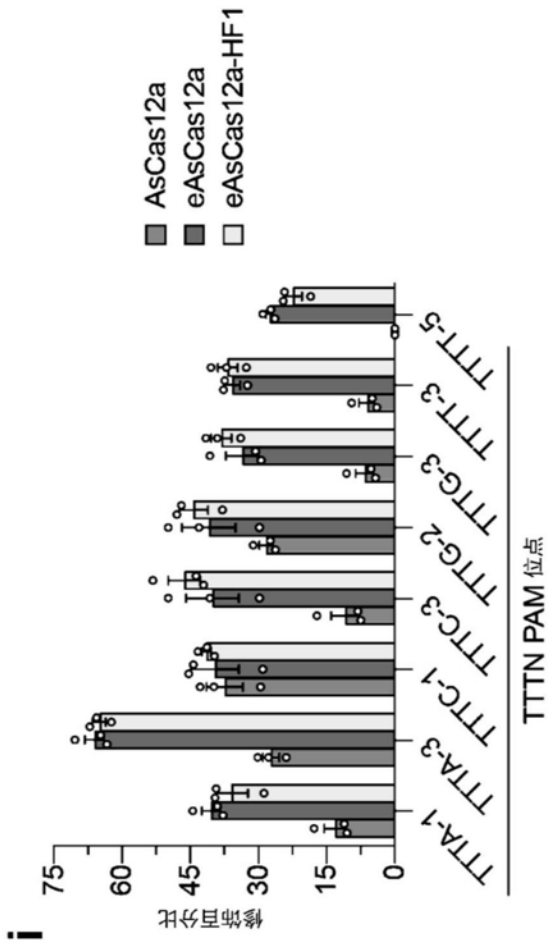

AsCas12a
eAsCas12a
eAsCas12a-HF1
i
修饰百分比
75
60
45
30
15
0
TTTA-1
TTTA-3
TTTC-1
TTTC-3
TTTG-2
TTTG-3
TTTT-3
TTTT-5
TTTN PAM 位点

图25I

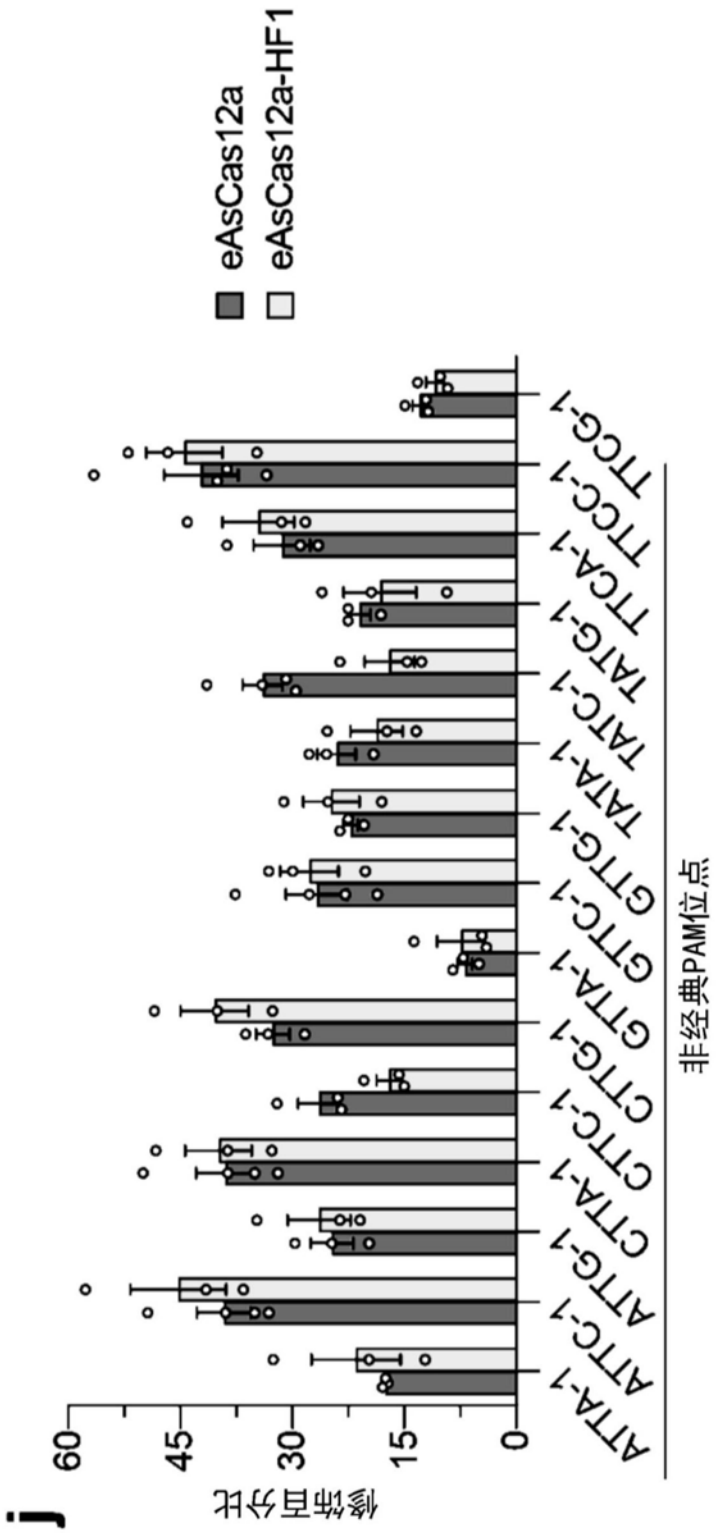

eAsCas12a
eAsCas12a-HF1
修饰百分比
60
45
30
15
0
ATTA-1
ATTC-1
ATTG-1
CTTA-1
CTTC-1
CTTG-1
GTTA-1
GTTC-1
GTTG-1
TATA-1
TATC-1
TATG-1
TTCA-1
TTCC-1
TTCG-1
非经典PAM位点
j

图25J

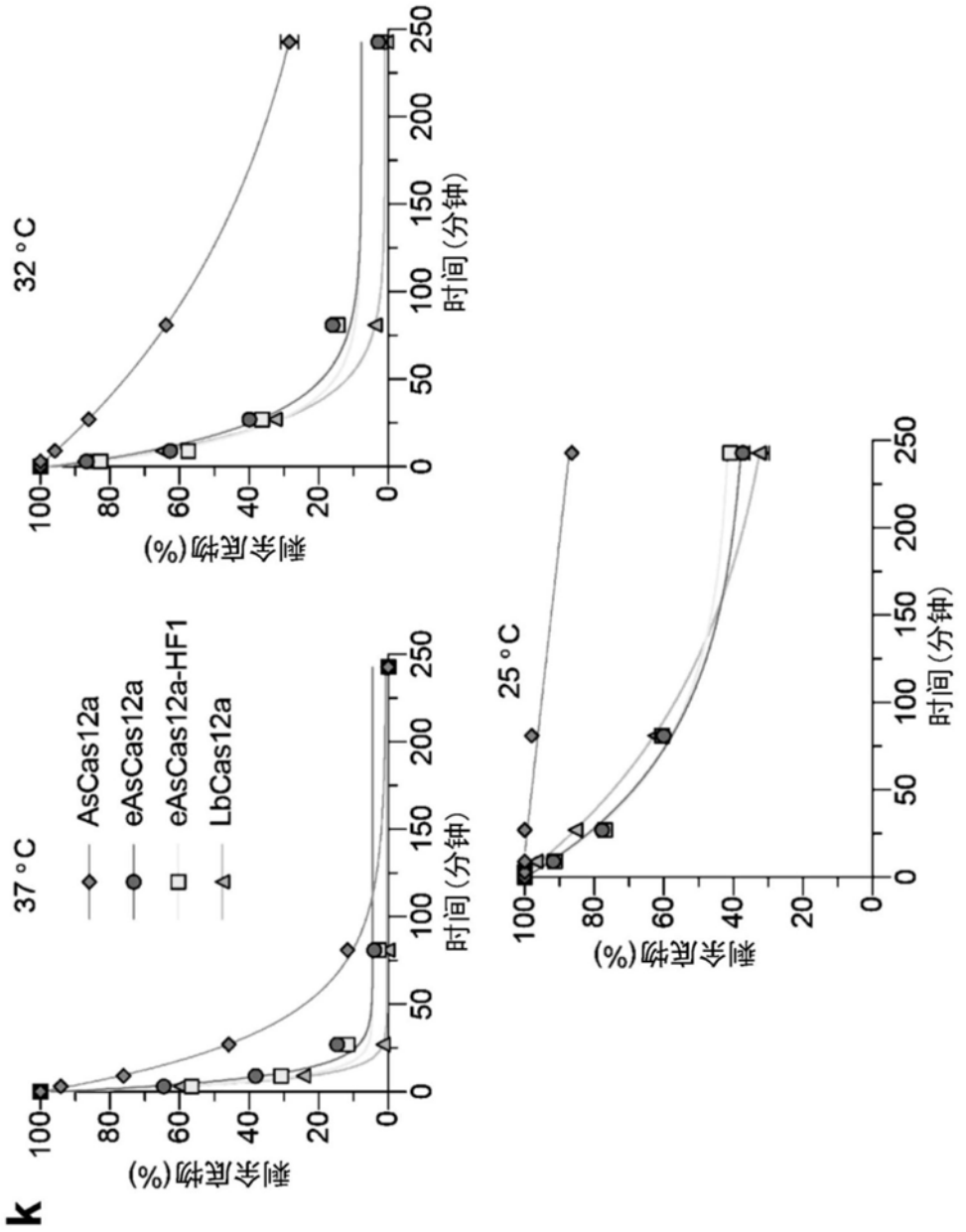

k
37 °C
32 °C
25 °C
AsCas12a
eAsCas12a
eAsCas12a-HF1
LbCas12a
剩余底物(%)
时间(分钟)
0
20
40
60
80
100
0
50
100
150
200
250

图25K

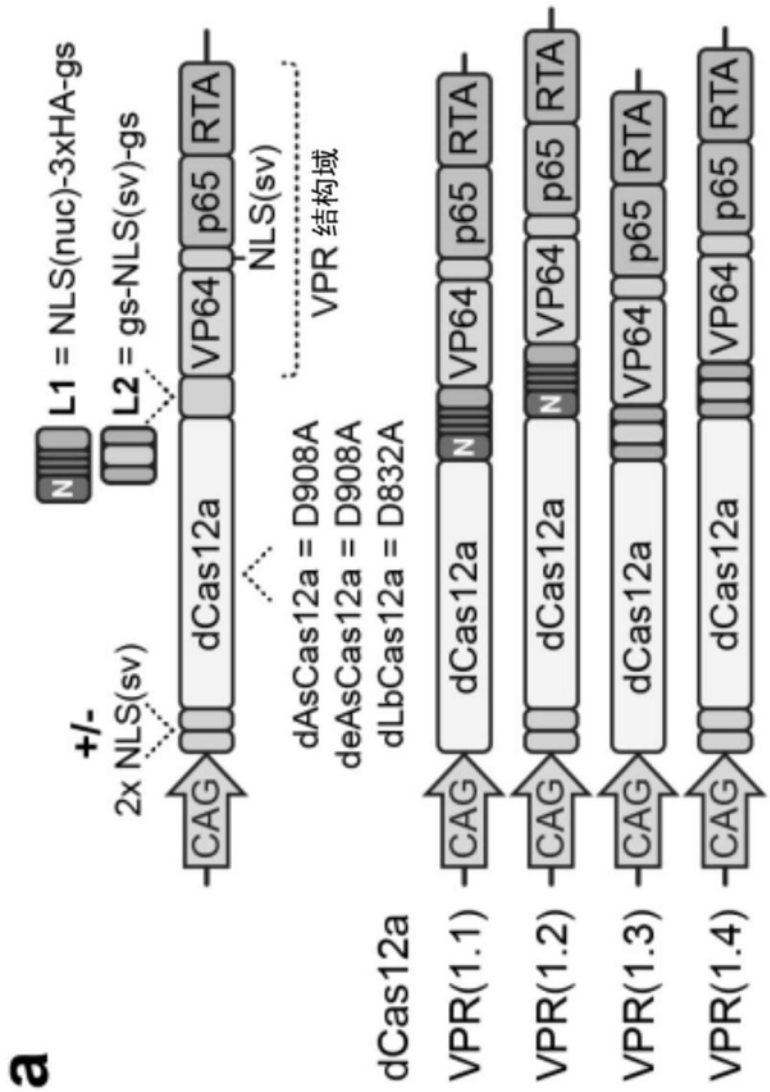

a
dCas12a
VPR(1.1)
VPR(1.2)
VPR(1.3)
VPR(1.4)
CAG
+/-
2x NLS(sv)
dCas12a
N
L1 = NLS(nuc)-3xHA-gs
L2 = gs-NLS(sv)-gs
VP64
p65
RTA
NLS(sv)
VPR 结构域
dAsCas12a = D908A
deAsCas12a = D908A
dLbCas12a = D832A

图26A

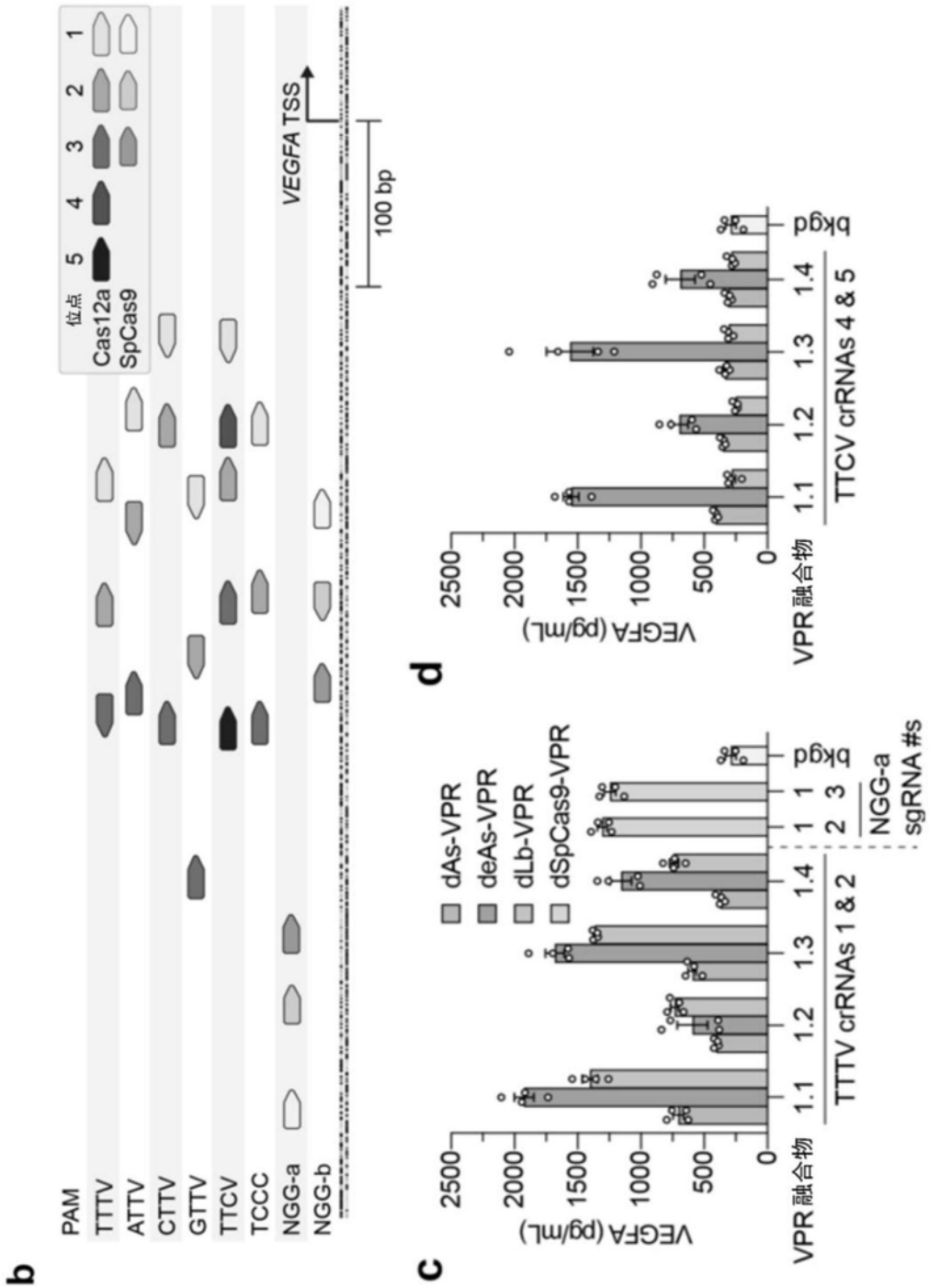

b
PAM
TTTV
ATTV
CTTV
GTTV
TTCV
TCCC
NGG-a
NGG-b
位点 5 4 3 2 1
Cas12a
SpCas9
VEGFA TSS
100 bp
c
dAs-VPR
deAs-VPR
dLb-VPR
dSpCas9-VPR
VEGFA (pg/mL)
2500
2000
1500
1000
500
0
VPR 融合物
1.1 1.2 1.3 1.4
TTTV crRNAs 1 & 2
1 2 3
NGG-a
sgRNA #s
bkgd
d
VEGFA (pg/mL)
VPR 融合物
1.1 1.2 1.3 1.4
TTCV crRNAs 4 & 5
bkgd

图26B-D

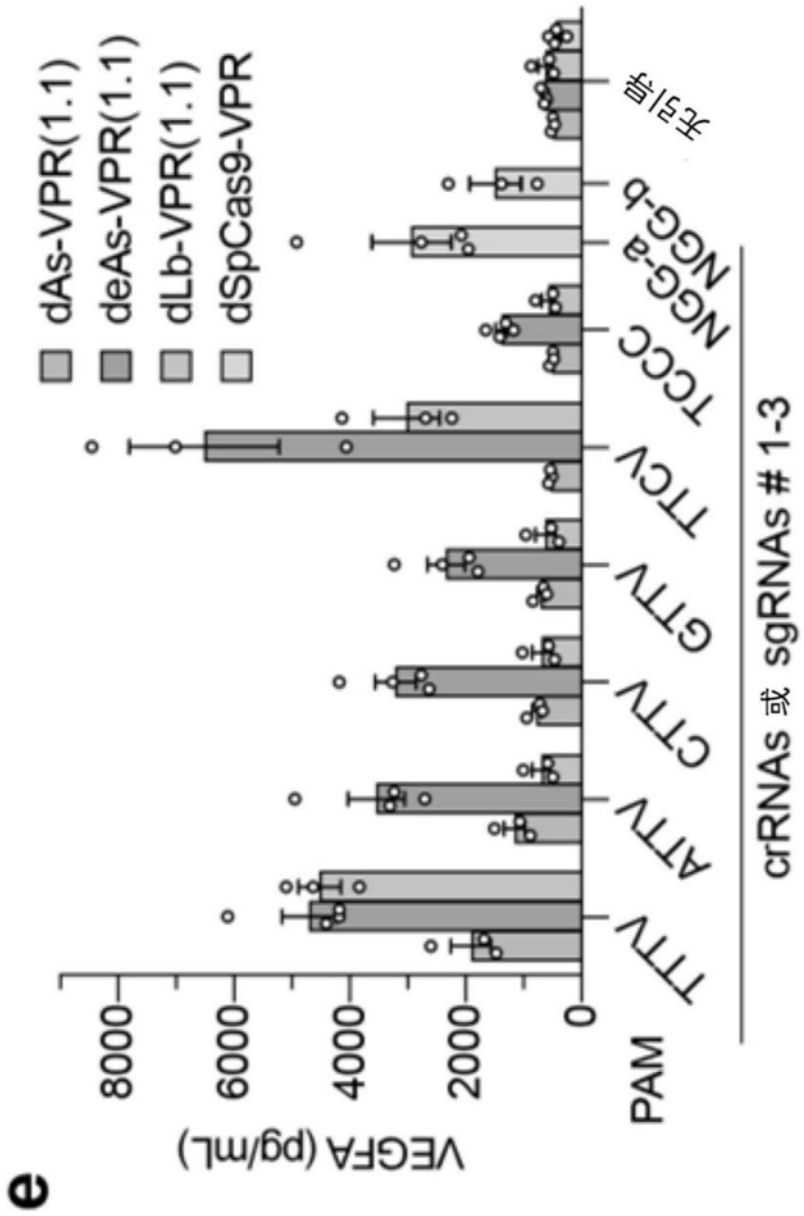

e
VEGFA (pg/mL)
8000
6000
4000
2000
0
dAs-VPR(1.1)
deAs-VPR(1.1)
dLb-VPR(1.1)
dSpCas9-VPR
PAM
TTTV
ATTV
CTTV
GTTV
TTCV
TCCC
NGG-a
NGG-b
无引导
crRNAs 或 sgRNAs # 1-3

图26E

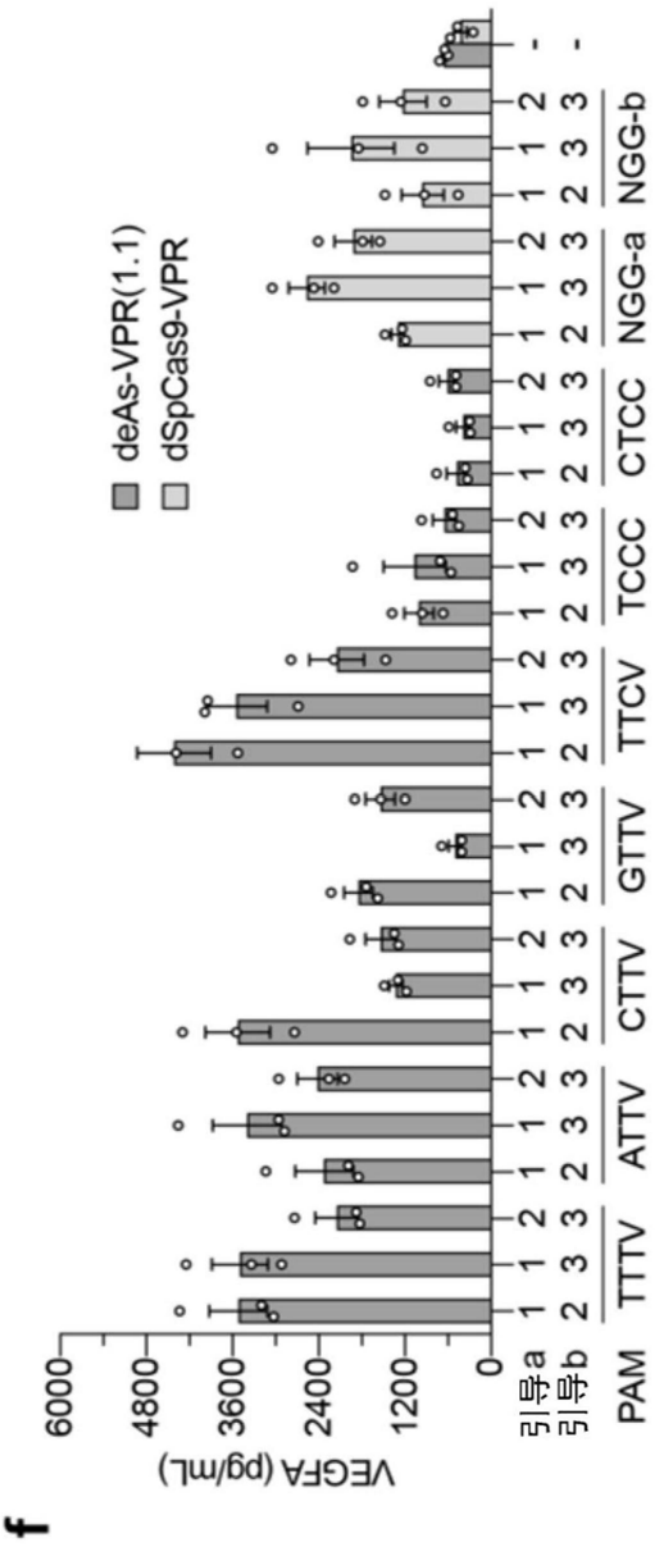

f
VEGFA (pg/mL)
6000
4800
3600
2400
1200
0
deAs-VPR(1.1)
dSpCas9-VPR
PAM
引导b
引导a
TTTV
ATTV
CTTV
GTTV
TTCV
TCCC
CTCC
NGG-a
NGG-b
1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 - -

图26F

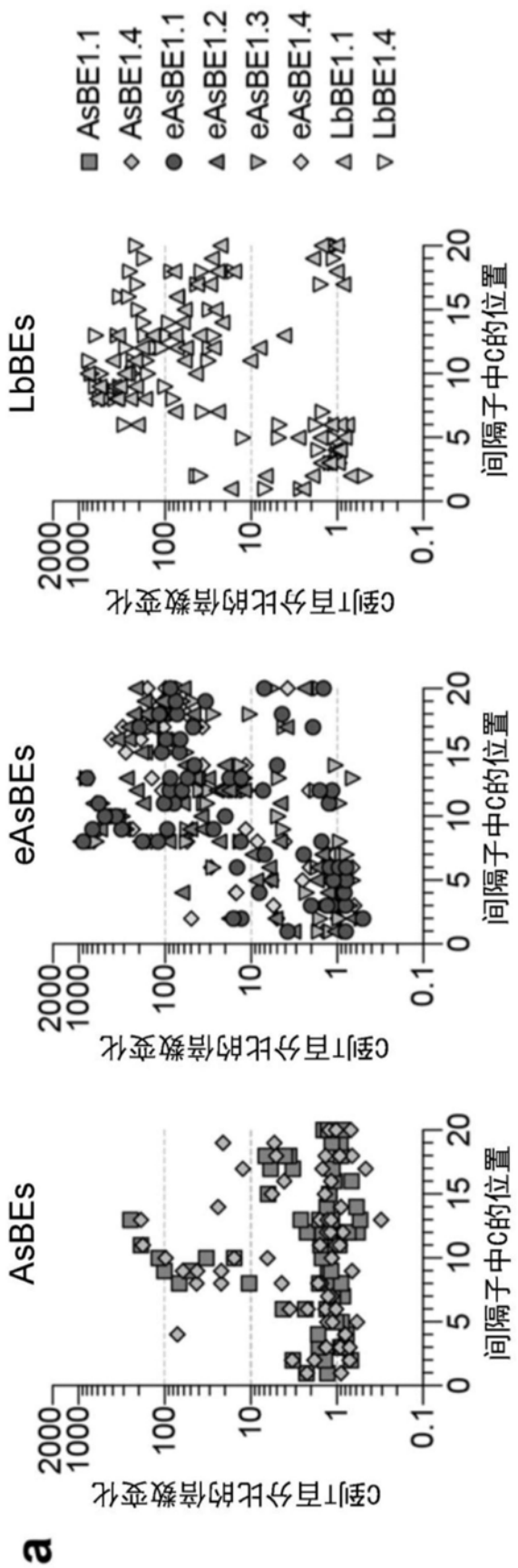

a
AsBEs
eAsBEs
LbBEs
AsBE1.1
AsBE1.4
eAsBE1.1
eAsBE1.2
eAsBE1.3
eAsBE1.4
LbBE1.1
LbBE1.4
C到T百分比的倍数变化
间隔子中C的位置
2000
1000
100
10
1
0.1
0
5
10
15
20

图27A

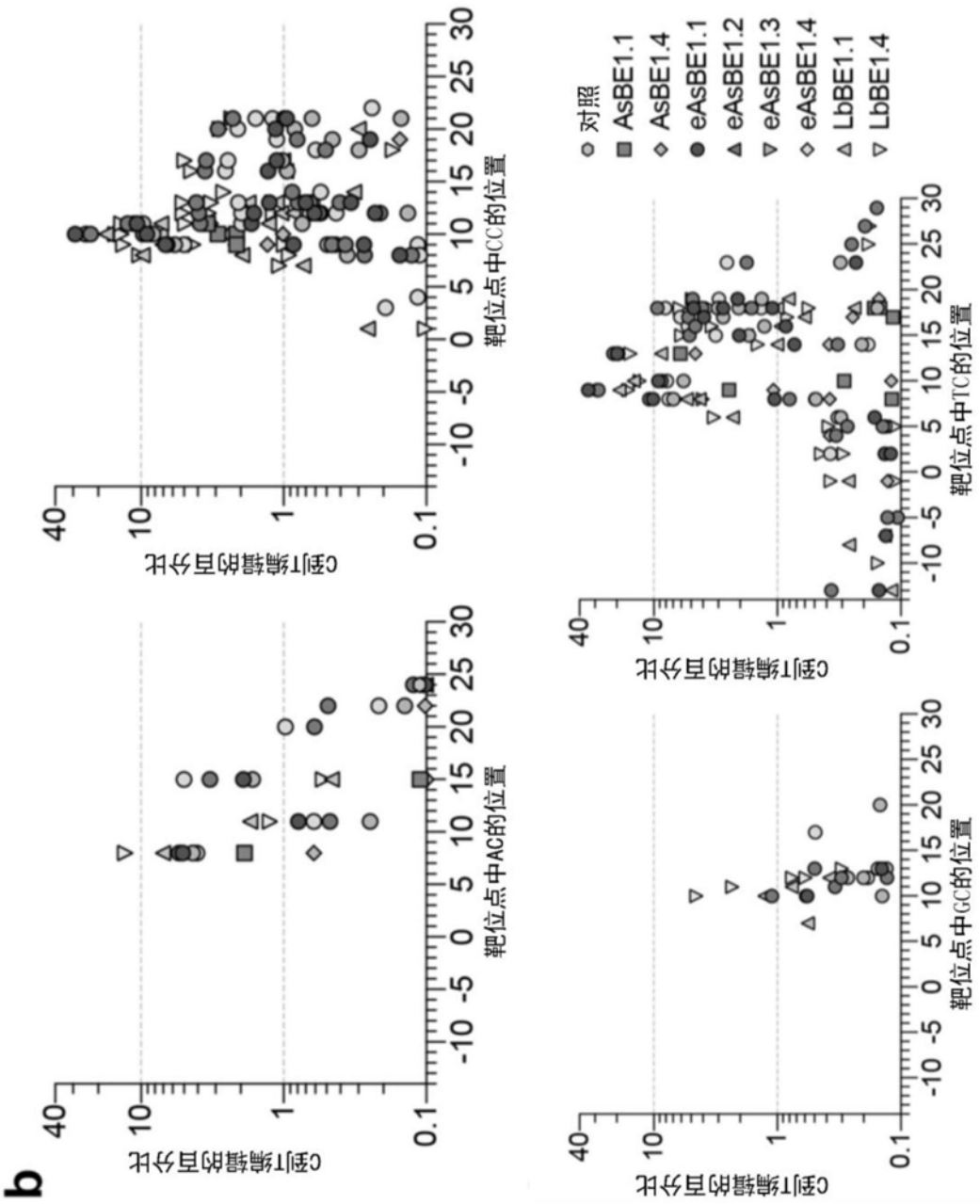

b
C到T编辑的百分比
靶位点中AC的位置
靶位点中CC的位置
靶位点中GC的位置
靶位点中TC的位置
对照
AsBE1.1
AsBE1.4
eAsBE1.1
eAsBE1.2
eAsBE1.3
eAsBE1.4
LbBE1.1
LbBE1.4

图27B

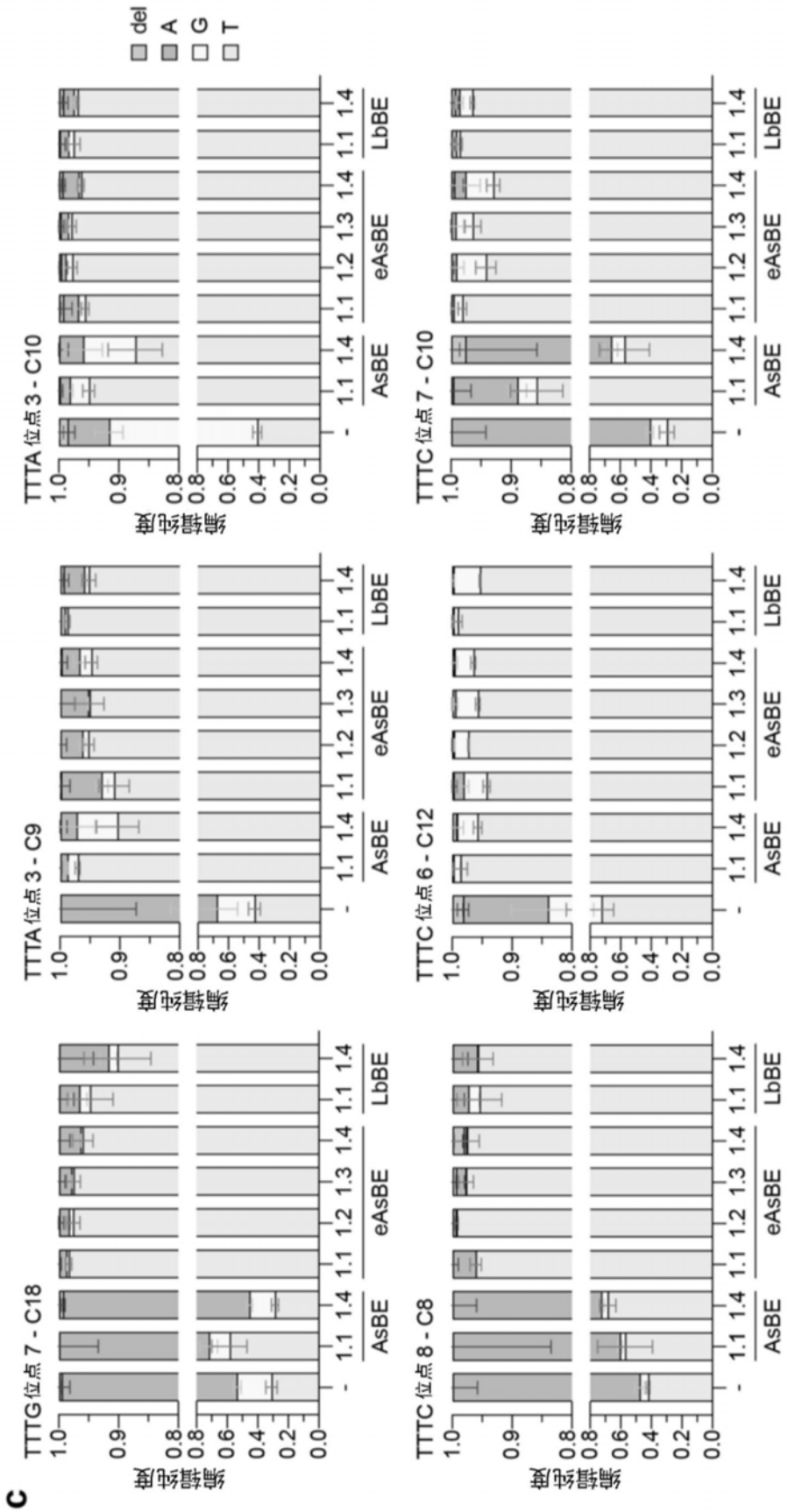

C
del
A
G
T
TTTG位点7 - C18
TTTA位点3 - C9
TTTC位点6 - C12
TTTA位点3 - C10
TTTC位点7 - C10
TTTC位点8 - C8
编辑纯度
1.0
0.9
0.8
0.6
0.4
0.2
0.0
-
1.1
1.4
1.1
1.2
1.3
1.4
1.1
1.4
AsBE
eAsBE
LbBE

图27C

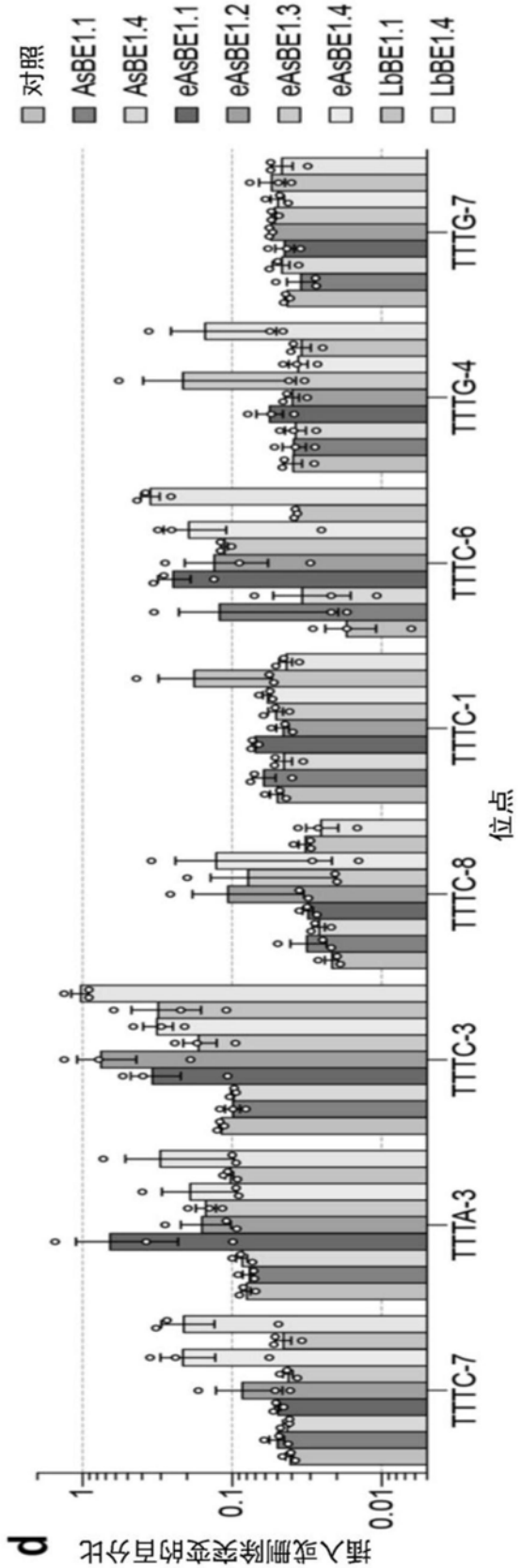

对照
AsBE1.1
AsBE1.4
eAsBE1.1
eAsBE1.2
eAsBE1.3
eAsBE1.4
LbBE1.1
LbBE1.4
TTTG-7
TTTG-4
TTTC-6
TTTC-1
TTTC-8
TTTC-3
TTTA-3
TTTC-7
位点
1
0.1
0.01
插入或删除突变的百分比
d

图27D

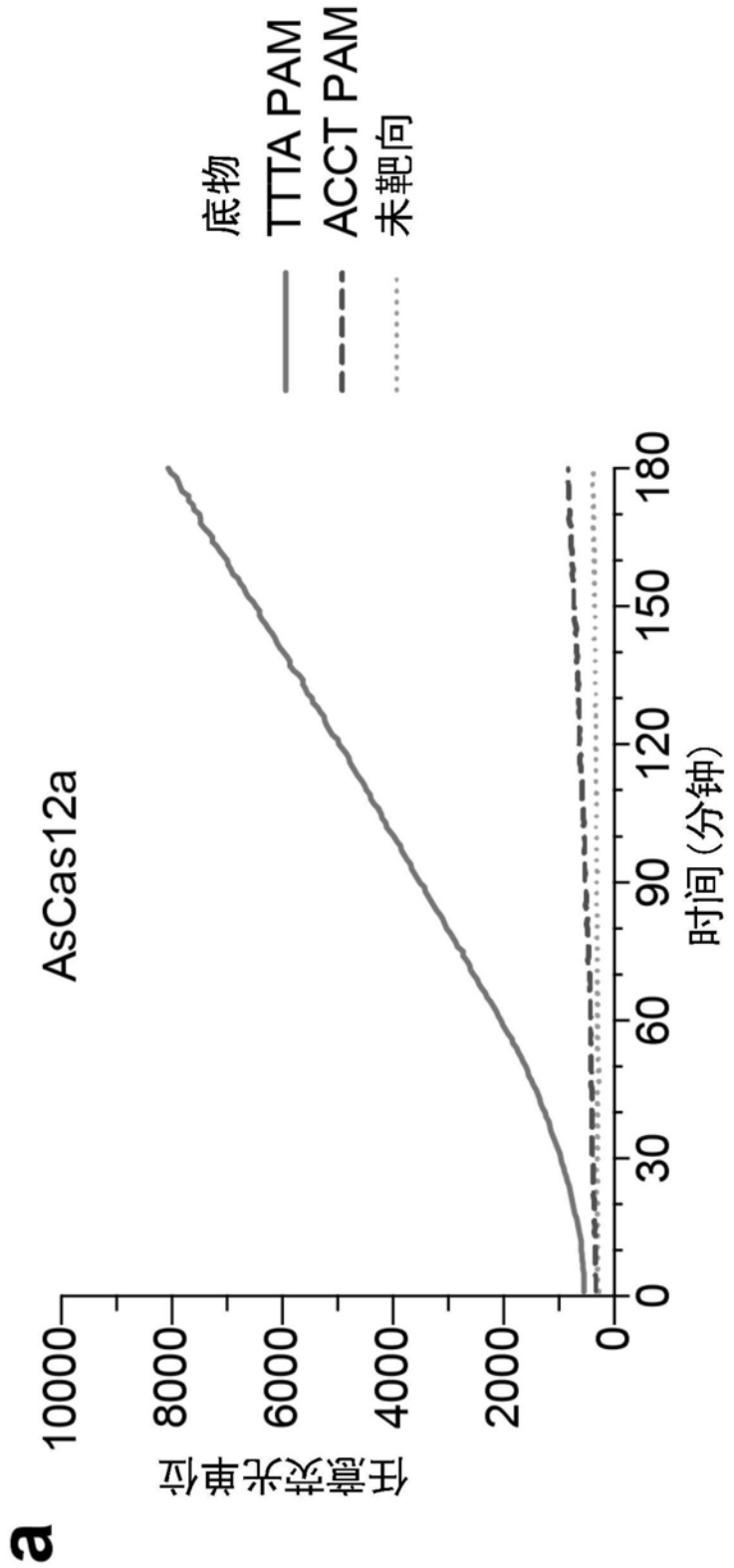

a
AsCas12a
任意荧光单位
10000
8000
6000
4000
2000
0
0
30
60
90
120
150
180
时间(分钟)
底物
TTTA PAM
ACCT PAM
未靶向

图28A

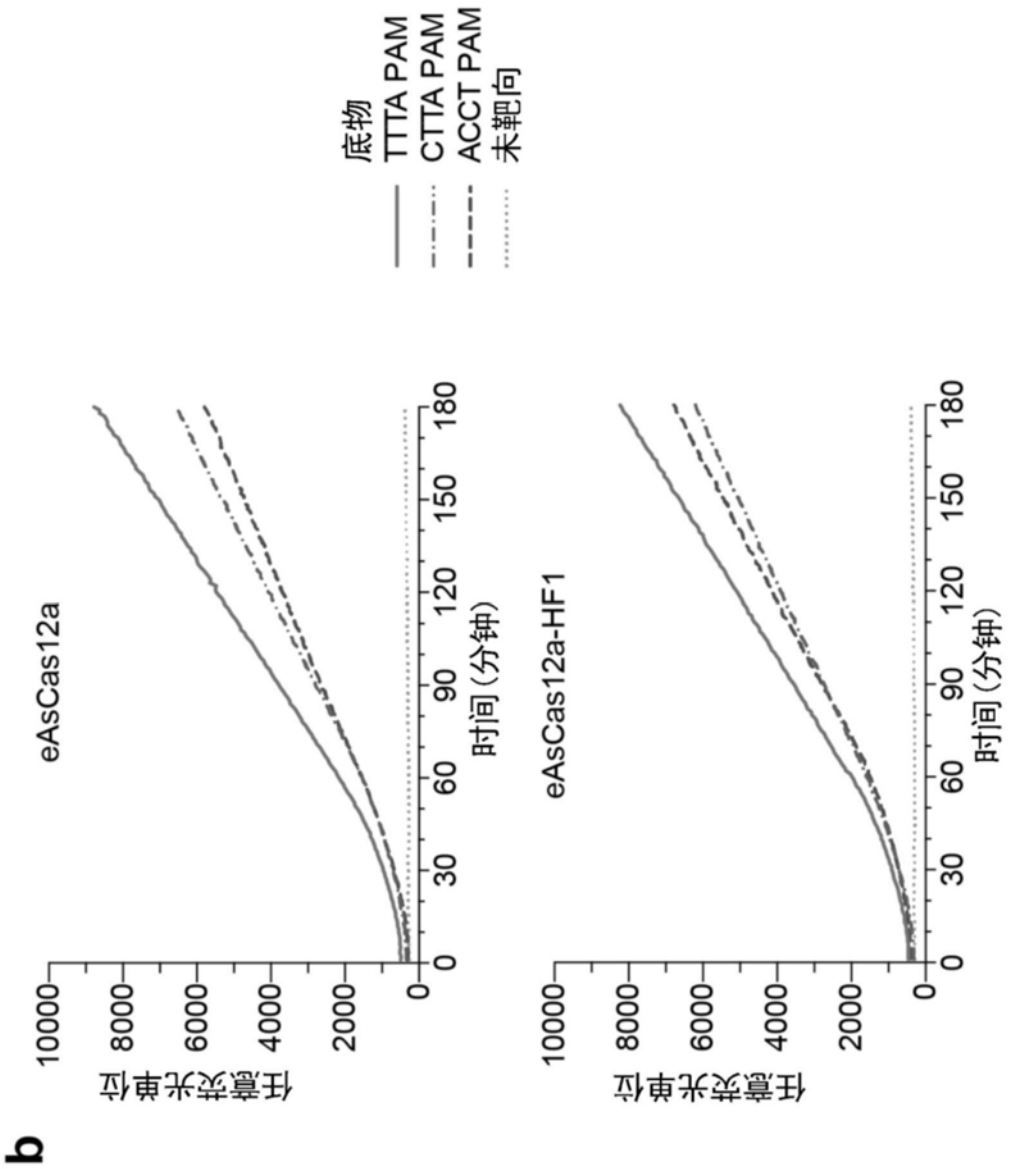

b
底物
TTTA PAM
CTTA PAM
ACCT PAM
未靶向
eAsCas12a
eAsCas12a-HF1
任意荧光单位
时间(分钟)
0
2000
4000
6000
8000
10000
0
30
60
90
120
150
180

图28B