(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 110914426 A
(43)申请公布日 2020.03.24

(21)申请号 201880033446.0

(22)申请日 2018.03.23

(30)优先权数据
62/475,830 2017.03.23 US
62/490,587 2017.04.26 US
62/511,934 2017.05.26 US
62/551,951 2017.08.30 US

(85)PCT国际申请进入国家阶段日
2019.11.20

(86)PCT国际申请的申请数据
PCT/US2018/024208 2018.03.23

(87)PCT国际申请的公布数据
WO2018/176009 EN 2018.09.27

(71)申请人 哈佛大学的校长及成员们
地址 美国马萨诸塞州

(72)发明人 D.R.刘 A.C.科摩 L.陈 H.A.里斯

(74)专利代理机构 北京市柳沈律师事务所 11105
代理人 余晓文

(51)Int.Cl.
C12N 15/11(2006.01)
C12N 15/10(2006.01)

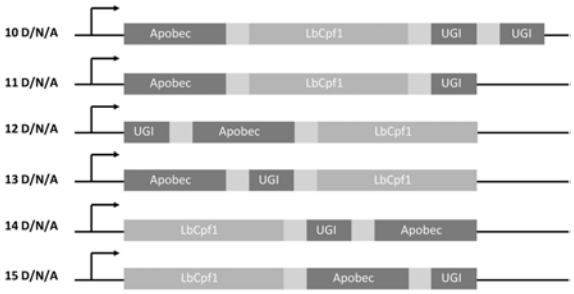
权利要求书11页 说明书222页
序列表（电子公布） 附图215页

(54)发明名称

包含核酸可编程DNA结合蛋白的核碱基编辑器

(57)摘要

本公开内容的一些方面提供了可用于核酸的靶向编辑，包括编辑细胞或受试者的基因组内，例如人基因组内的单一位点的策略、系统、试剂、方法和试剂盒。在一些实施方案中，提供了核酸可编程DNA结合蛋白(napDNAbp)例如Cpf1或其变体与核酸编辑蛋白或蛋白质域，例如脱氨酶域的融合蛋白。在一些实施方案中，提供了用于靶向核酸编辑的方法。在一些实施方案中，提供了用于产生靶向核酸编辑蛋白，例如napDNAbp(例如，CasX，CasY，Cpf1，C2c1，C2c2，C2C3和Argonaute)与核酸编辑蛋白或域的融合蛋白的试剂和试剂盒。



CN 110914426 A

1.融合蛋白,其包含:(i)核酸可编程DNA结合蛋白(napDNAbp);(ii)胞苷脱氨酶域;和(iii)尿嘧啶糖基化酶抑制剂(UGI)域,其中所述napDNAbp是CasX、CasY、Cpf1、C2c1、C2c2、C2c3或Argonaute蛋白。

2.权利要求1的融合蛋白,其中所述融合蛋白包含2、3、4或5个UGI域。

3.权利要求1或2的融合蛋白,其中所述核酸可编程DNA结合蛋白(napDNAbp)是CasX蛋白。

4.权利要求3的融合蛋白,其中所述CasX蛋白包含

(i)与SEQ ID NO:29或30至少90%相同的氨基酸序列;或

(ii)SEQ ID NO:29或30的氨基酸序列。

5.权利要求1或2的融合蛋白,其中所述核酸可编程DNA结合蛋白(napDNAbp)是CasY蛋白。

6.权利要求5的融合蛋白,其中所述CasY蛋白包含与SEQ ID NO:31至少90%相同的氨基酸序列。

7.权利要求5或6的融合蛋白,其中所述CasY蛋白包含SEQ ID NO:31的氨基酸序列。

8.权利要求1或2的融合蛋白,其中所述核酸可编程DNA结合蛋白(napDNAbp)是Cpf1或Cpf1突变体蛋白。

9.权利要求8的融合蛋白,其中所述Cpf1或Cpf1突变体蛋白包含与SEQ ID NO:9-16中的任一项至少90%相同的氨基酸序列。

10.权利要求8或9的融合蛋白,其中所述Cpf1或Cpf1突变体蛋白包含SEQ ID NO:9-16中任一项的氨基酸序列。

11.权利要求1或2的融合蛋白,其中所述核酸可编程DNA结合蛋白(napDNAbp)是C2c1蛋白。

12.权利要求11的融合蛋白,其中所述C2c1蛋白包含与SEQ ID NO:26至少90%相同的氨基酸序列。

13.权利要求11或12的融合蛋白,其中所述C2c1蛋白包含SEQ ID NO:26的氨基酸序列。

14.权利要求1或2的融合蛋白,其中所述核酸可编程DNA结合蛋白(napDNAbp)是C2c2蛋白。

15.权利要求14的融合蛋白,其中所述C2c2蛋白包含与SEQ ID NO:27至少90%相同的氨基酸序列。

16.权利要求14或15的融合蛋白,其中所述C2c2蛋白包含SEQ ID NO:27的氨基酸序列。

17.权利要求1或2的融合蛋白,其中所述核酸可编程DNA结合蛋白(napDNAbp)是C2c3蛋白。

18.权利要求17的融合蛋白,其中所述C2c3蛋白包含与SEQ ID NO:28至少90%相同的氨基酸序列。

19.权利要求17或18的融合蛋白,其中所述C2c3蛋白包含SEQ ID NO:28的氨基酸序列。

20.权利要求1或2的融合蛋白,其中所述核酸可编程DNA结合蛋白(napDNAbp)是Argonaute蛋白。

21.权利要求20的融合蛋白,其中所述Argonaute蛋白包含与SEQ ID NO:25至少90%相同的氨基酸序列。

22.权利要求20或21的融合蛋白,其中所述Argonaute蛋白包含SEQ ID NO:25的氨基酸序列。

23.权利要求1-22中任一项的融合蛋白,其中所述胞苷脱氨酶域是来自载脂蛋白B mRNA编辑复合物(APOBEC)家族脱氨酶的脱氨酶。

24.权利要求23的融合蛋白,其中所述APOBEC家族脱氨酶选自下组:APOBEC1脱氨酶、APOBEC2脱氨酶、APOBEC3A脱氨酶、APOBEC3B脱氨酶、APOBEC3C脱氨酶、APOBEC3D脱氨酶、APOBEC3F脱氨酶、APOBH3G脱氨酶和APOBEC3H脱氨酶。

25.权利要求1-24中任一项的融合蛋白,其中所述胞苷脱氨酶域包含与SEQ ID NO:49-84的氨基酸序列至少85%相同的氨基酸序列。

26.权利要求1至25中任一项的融合蛋白,其中所述胞苷脱氨酶域包含SEQ ID NO:49-84的氨基酸序列。

27.权利要求1-22中任一项的融合蛋白,其中所述胞苷脱氨酶域是大鼠APOBEC1(rAPOBEC1)脱氨酶,其包含一个或多个选自下组的突变:SEQ ID NO:76的W90Y、R126E和R132E,或另一种APOBEC脱氨酶中的一个或多个相应突变。

28.权利要求1-22中任一项的融合蛋白,其中所述胞苷脱氨酶域是人APOBEC1(hAPOBEC1)脱氨酶,其包含一个或多个选自下组的突变:SEQ ID NO:74的W90Y、Q126E和R132E,或另一种APOBEC脱氨酶中的一个或多个相应突变。

29.权利要求1-22中任一项的融合蛋白,其中所述胞苷脱氨酶域是人APOBEC3G(hAPOBEC3G)脱氨酶,其包含一个或多个选自下组的突变:SEQ ID NO:60的W285Y、R320E和R326E,或另一种APOBEC脱氨酶中的一个或多个相应突变。

30.权利要求1-22中任一项的融合蛋白,其中所述胞苷脱氨酶域是激活诱导的脱氨酶(AID)。

31.权利要求1-22中任一项的融合蛋白,其中所述胞苷脱氨酶域是来自海七鳃鳗(Petromyzon marinus)的胞苷脱氨酶1(pmCDA1)。

32.权利要求1-31中任一项的融合蛋白,其中所述UGI域包含能够抑制UDG活性的域。

32.权利要求1-32中任一项的融合蛋白,其中所述UGI域包含与SEQ ID NO:134至少85%相同的氨基酸序列。

33.权利要求1-32中任一项的融合蛋白,其中所述UGI域包含以SEQ ID NO:134所示的氨基酸序列。

34.权利要求1-33中任一项的融合蛋白,其中所述融合蛋白包含以下结构:

NH_2-[胞苷脱氨酶域]-[napDNAbp]-[UGI域]-COOH;

NH_2-[胞苷脱氨酶域]-[napDNAbp]-[UGI]-[UGI]-COOH;

NH_2-[胞苷脱氨酶域]-[napDNAbp]-[UGI]-COOH;

NH_2-[UGI]-[Apobec]-[napDNAbp]-COOH;

NH_2-[胞苷脱氨酶域]-[UGI]-[napDNAbp]-COOH;

NH_2-[napDNAbp]-[UGI]-[胞苷脱氨酶域]-COOH;或

NH_2-[napDNAbp]-[胞苷脱氨酶域]-[UGI]-COOH;

其中“-”的每个情况包括任选的接头。

35.权利要求34中任一项的融合蛋白,其中(ii)的胞苷脱氨酶域和(i)的napDNAbp域经

由接头连接，所述接头包含以下基序或其组合：氨基酸序列(GGGS)$_n$(SEQ ID NO:613)、(GGGGS)$_n$(SEQ ID NO:607)、(G)$_n$(SEQ ID NO:608)，(EAAAK)$_n$(SEQ ID NO:609)、(GGS)$_n$(SEQ ID NO:610)、(SGGS)$_n$(SEQ ID NO:606)、SGSETPGTSESATPES(SEQ ID NO:604)或(XP)$_n$(SEQ ID NO:611)，其中n独立地是1和30之间的整数，包括1和30，且其中X是任意氨基酸。

36.权利要求1-35中任一项的融合蛋白，其中(ii)的胞苷脱氨酶域和(i)的napDNAbp经由包含氨基酸序列SGSETPGTSESATPES(SEQ ID NO:604)的接头连接。

37.权利要求1-36中任一项的融合蛋白，其还包含核定位序列(NLS)。

38.权利要求37的融合蛋白，其中所述NLS包含氨基酸序列PKKKRKV(SEQ ID NO:614)、MDSLLMNRRKFLYQFKNVRWAKGRRETYLC(SEQ ID NO:615)或KRTADGSEFEPKKKRKV(SEQ ID NO:740)。

39.权利要求37或38的融合蛋白，其中所述融合蛋白包含以下结构：

NH_2-[胞苷脱氨酶域]-[napDNAbp]-[UGI域]-[NLS]-COOH，且其中“-”的每个情况包含任选的接头。

40.权利要求37-39中任一项的融合蛋白，其中所述UGI域和所述NLS经由包含氨基酸序列SGGS(SEQ ID NO:606)的接头连接，或其中所述napDNAbp和所述UGI域经由包含氨基酸序列SGGS(SEQ ID NO:606)的接头连接。

41.复合物，其包含权利要求1-40中任一项的融合蛋白和与所述融合蛋白的napDNAbp结合的引导RNA。

42.方法，其包括使核酸分子与权利要求1-40中任一项的融合蛋白和引导RNA接触，其中所述引导RNA包含至少10个连续核苷酸的序列，该序列与生物体的基因组中的靶序列互补，并且包含靶碱基对。

43.权利要求42的方法，其中所述靶碱基对包含与疾病或病症相关的T至C点突变，并且其中所述突变体C碱基的脱氨基导致与疾病或病症无关的序列。

44.权利要求42或43的方法，其中所述接触导致在碱基编辑后小于20％的插入/缺失形成。

45.权利要求42-44中任一项的方法，其中所述接触导致碱基编辑后至少2:1的意图与非意图产物。

46.复合物，其包含权利要求1-40中任一项的融合蛋白和与所述napDNAbp结合的RNA。

47.权利要求46的复合物，其中所述RNA是引导RNA(gRNA)。

48.权利要求46的复合物，其中所述RNA是非编码RNA，和/或其中所述RNA包含SEQ ID NO:741的氨基酸序列，其中SEQ ID NO:741的每个T是尿嘧啶(U)。

49.权利要求46的复合物，其中所述RNA是微小RNA(miRNA)。

50.权利要求46的复合物，其中所述RNA是小干扰RNA(siRNA)。

51.权利要求46的复合物，其中所述RNA是Piwi相互作用RNA(piRNA)。

52.权利要求46-51中任一项的复合物，其中所述RNA长为10-100个核苷酸，并且包含与靶序列互补的至少10个连续核苷酸的序列。

53.权利要求52的复合物，其中所述RNA长为10、11、12、13、14,15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49或50个核苷酸。

54.权利要求46-53中任一项的复合物，其中所述RNA包含与靶序列互补的15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39或40个连续核苷酸的序列。

55.权利要求46-54中任一项的复合物，其中所述靶序列是DNA序列。

56.权利要求46-54中任一项的复合物，其中所述靶序列是RNA序列。

57.权利要求55的复合物，其中所述靶序列在生物体的基因组中。

58.权利要求56的复合物，其中所述靶序列是从生物体的基因组转录的。

59.权利要求57或58的复合物，其中所述生物体是原核生物。

60.权利要求59的复合物，其中所述原核生物是细菌。

61.权利要求57或58的复合物，其中所述生物体是真核生物。

62.权利要求61的复合物，其中所述生物体是植物。

63.权利要求61的复合物，其中所述生物体是脊椎动物。

64.权利要求63的复合物，其中所述脊椎动物是哺乳动物。

65.权利要求64的复合物，其中所述哺乳动物是小鼠或大鼠。

66.权利要求64的复合物，其中所述哺乳动物是人。

67.方法，其包括使核酸分子与权利要求46-66中任一项的复合物接触。

68.权利要求67的方法，其中所述核酸是DNA。

69.权利要求68的方法，其中所述核酸是双链DNA。

70.权利要求67的方法，其中所述核酸是RNA。

71.权利要求70的方法，其中所述RNA是mRNA。

72.权利要求67-71中任一项的方法，其中所述核酸包含与疾病或病症相关的靶序列。

73.权利要求72的方法，其中所述靶序列包括与疾病或病症相关的点突变。

74.权利要求73的方法，其中所述融合蛋白或所述复合物的活性导致所述点突变的校正。

75.权利要求73-74中任一项的方法，其中所述靶序列包含与疾病或病症相关的T至C点突变，并且其中所述突变体C碱基的脱氨基导致与疾病或病症无关的序列。

76.权利要求75的方法，其中所述靶序列编码蛋白质，并且其中所述点突变在密码子中，并且与野生型密码子相比导致由所述突变体密码子编码的氨基酸的变化。

77.权利要求75或76的方法，其中所述突变体C的脱氨基导致所述突变体密码子编码的氨基酸的变化。

78.权利要求75或76的方法，其中所述突变体C的脱氨基导致编码野生型氨基酸的密码子。

79.权利要求67-78中任一项的方法，其中所述接触在受试者的体内进行。

80.权利要求67-78中任一项的方法，其中所述接触在体外进行。

81.权利要求79的方法，其中所述受试者已经被诊断出患有疾病或病症。

82.权利要求72-82中任一项的方法，其中所述疾病或病症是囊性纤维化(cystic fibrosis)、苯丙酮尿症(phenylketonuria)、表皮松解角化过度症(epidermolytic hyperkeratosis)(EHK)、夏科-马里-图思病(Charcot-Marie-Toot disease)4J型、神经母细胞瘤(neuroblastoma)(NB)、血管性血友病(von Willebrand disease)(vWD)、先天性肌

强直(myotonia congenital)、遗传性肾淀粉样变性(hereditary renal amyloidosis)、扩张型心肌病(dilated cardiomyopathy)(DCM)、遗传性淋巴水肿(hereditary lymphedema)、家族性阿尔茨海默氏病、HIV、朊病毒病(Prion disease)、慢性婴儿神经系统皮肤关节综合征(chronic infantile neurologic cutaneous articular syndrome)(CINCA)、结蛋白相关性肌病(desmin-related myopathy)(DRM)、与PI3KCA突变体蛋白、CTNNB1突变体蛋白、HRAS突变体蛋白或p53突变体蛋白相关的肿瘤性疾病(neoplastic disease)。

83.包含核酸构建体的试剂盒,所述核酸构建体包含

(a)编码权利要求1-40中任一项的融合蛋白的核酸序列;和

(b)驱动(a)的序列表达的异源启动子。

84.权利要求83的试剂盒,其进一步包含编码引导RNA主链的表达构建体,其中所述构建体包含克隆位点,所述克隆位点定位为允许将与靶序列相同或互补的核酸序列克隆到所述引导RNA主链中。

85.编码权利要求1-40中任一项的融合蛋白的多核苷酸。

86.载体,其包含权利要求85的多核苷酸。

87.权利要求86的载体,其中所述载体包含驱动所述多核苷酸表达的异源启动子。

88.细胞,其包含权利要求1-40中任一项的融合蛋白。

89.细胞,其包含权利要求46-66中任一项的复合物。

90.细胞,其包含编码权利要求1-40中任一项的融合蛋白的核酸分子。

91.用于产生核糖核蛋白(RNP)复合物的方法,所述方法包括:

(i)在水性溶液中将碱基编辑器蛋白与RNA复合,从而在所述水性溶液中形成包含所述碱基编辑器和所述RNA的复合物;并且

(ii)使(i)的复合物与阳离子脂质接触。

92.权利要求91的方法,其中(i)的所述碱基编辑器蛋白和所述RNA以1:1至1:1.5的摩尔比复合。

93.权利要求91或92的方法,其中(i)的所述碱基编辑器蛋白和所述RNA以1:1.05至1:1.2的摩尔比复合。

94.权利要求91-93中任一项的方法,其中(i)的所述碱编辑蛋白和所述RNA以约1:1.1的摩尔比复合。

95.权利要求91-94中任一项的方法,其中(i)的所述碱编辑蛋白和所述RNA以1:1.1的摩尔比复合。

96.权利要求91-95中任一项的方法,其中所述碱基编辑器蛋白以10μM至100μM的浓度存在于所述水性溶液中。

97.权利要求91-96中任一项的方法,其中所述碱基编辑器蛋白以20μM至50μM的浓度存在于所述水性溶液中。

98.权利要求91-97中任一项的方法,其中所述碱基编辑器蛋白以30μM至40μM的浓度存在于所述水性溶液中。

99.权利要求91-98中任一项的方法,其中所述碱基编辑器蛋白以11μM至110μM的浓度存在于所述水性溶液中。

100.权利要求91-99中任一项的方法，其中所述碱基编辑器蛋白以22μM至55μM的浓度存在于所述水性溶液中。

101.权利要求91至100中任一项的方法，其中所述碱基编辑器蛋白以33μM至44μM的浓度存在于所述水性溶液中。

102.权利要求91-101中任一项的方法，其中所述RNA是sgRNA。

103.权利要求91-102中任一项的方法，其中使(i)的所述水性溶液中的所述复合物与(ii)的所述阳离子脂质以1:2至2:1的体积比接触。

104.权利要求91-103中任一项的方法，其中使(i)的所述水性溶液中的所述复合物与(ii)的所述阳离子脂质以1:1.5至1.5:1的体积比接触。

105.权利要求91-104中任一项的方法，其中使(i)的所述水性溶液中的所述复合物与(ii)的所述阳离子脂质以1:1.2至1.2:1的体积比接触。

106.权利要求91-105中任一项的方法，其中使(i)的所述水性溶液中的所述复合物与(ii)的所述阳离子脂质以1:1.1至1.1:1的体积比接触。

107.权利要求91-106中任一项的方法，其中使(i)的所述水性溶液中的所述复合物与(ii)的所述阳离子脂质以约1:1的体积比接触。

108.权利要求91-107中任一项的方法，其中使(i)的所述水性溶液中的所述复合物与(ii)的所述阳离子脂质以1:1的体积比接触。

109.权利要求91-108中任一项的方法，其中所述阳离子脂质为Lipofectamine®。

110.权利要求109的方法，其中所述Lipofectamine®选自下组：Lipofectamine® 2000、Lipofectamine®3000、Lipofectamine® MessengerMAX、Lipofectamine® LTX和Lipofectamine®RNAiMAX。

111.权利要求91-110中任一项的方法，其中所述碱基编辑器是权利要求1-40中任一项的融合蛋白。

112.权利要求91-111中任一项的方法，其中(i)的所述复合物是权利要求46-66中任一项的复合物。

113.通过权利要求91-112中任一项的方法产生的药物组合物。

114.药物组合物，其包含权利要求1-40中任一项的融合蛋白。

115.药物组合物，其包含权利要求41的复合物。

116.药物组合物，其包含权利要求46-66中任一项的复合物。

117.权利要求113-116中任一项的药物组合物，其进一步包含药学上可接受的赋形剂。

118.用于纯化碱基编辑器蛋白的方法，所述方法包括：

(i)在细胞中表达所述碱基编辑器蛋白，其中所述碱基编辑器蛋白包含亲和标签；

(ii)裂解(i)的所述细胞，从而产生裂解物；并且

(iii)对所述裂解物进行亲和层析，从而产生包含纯化的碱基编辑器蛋白的洗脱液。

119.权利要求118的方法，其中所述细胞是细菌细胞。

120.权利要求119的方法，其中所述细菌细胞是大肠杆菌细胞。

121.权利要求120的方法，其中所述大肠杆菌细胞是BL21 DE3感受态细胞。

122.权利要求118-121中任一项的方法，其中所述亲和标签是聚组氨酸标签。

123.权利要求118-122中任一项的方法，其中步骤(iii)包括使(ii)的所述裂解物与镍-NTA树脂接触，其中所述碱基编辑器蛋白与所述镍-NTA树脂结合。

124.权利要求123的方法，其中所述方法进一步包括：

(iv)从所述镍-NTA树脂洗脱所述碱基编辑器蛋白，从而产生包含所述碱基编辑器蛋白的洗脱液。

125.权利要求118-124中任一项的方法，其中将包含所述碱基编辑器蛋白的洗脱液进行阳离子交换层析。

126.权利要求125的方法，其中所述阳离子交换层析包括使包含所述碱基编辑器蛋白的洗脱液与阳离子交换柱接触。

127.方法，其包括将权利要求1-40中任一项的融合蛋白、权利要求41的复合物、权利要求46-66中的任一项的复合物或权利要求113-117中的任一项的药物组合物递送至受试者的内耳。

128.权利要求127的方法，其包括将权利要求1-40中任一项的融合蛋白、权利要求41的复合物、权利要求46-66中任一项的复合物或权利要求113-117中任一项的药物组合物递送至所述受试者的耳蜗。

129.权利要求127或128的方法，其中将权利要求1-40中任一项的融合蛋白、权利要求41的复合物、权利要求46-66中任一项的复合物或权利要求113-117中任一项的药物组合物注射到所述受试者的内耳中。

130.权利要求127-129中任一项的方法，其中将权利要求1-40中任一项的融合蛋白、权利要求41的复合物、权利要求46-66中任一项的复合物或权利要求113-117中任一项的药物组合物注射到所述受试者的耳蜗中。

131.权利要求127-130中任一项的方法，其进一步包括对所述受试者的耳产生耳廓后切口。

132.方法，其包括将权利要求1-40中任一项的融合蛋白、权利要求41的复合物、权利要求46-66中任一项的复合物或权利要求113-117中的任一项的药物组合物递送至斑马鱼胚胎。

133.权利要求132的方法，其中将权利要求1-40中任一项的融合蛋白、权利要求41的复合物、权利要求46-66中任一项的复合物或权利要求113-117中的任一项的药物组合物注射到单细胞阶段的斑马鱼胚胎中。

134.权利要求132或133的方法，其中将权利要求1-40中任一项的融合蛋白、权利要求41的复合物、权利要求46-66中任一项的复合物或权利要求113-117中任一项的药物组合物注射到两细胞、四细胞、八细胞或十六细胞阶段的斑马鱼胚胎中。

135.权利要求132-134中任一项的方法，其中将0.5nmol和50nmol之间的与gRNA复合的碱基编辑器蛋白注射到所述斑马鱼胚胎中。

136.权利要求132-135中任一项的方法，其中将2nmol和30nmol之间的与gRNA复合的碱基编辑器蛋白注射到所述斑马鱼胚胎中。

137.融合蛋白，其包含：(i)核酸可编程DNA结合蛋白(napDNAbp)；(ii)胞苷脱氨酶域；(iii)第一尿嘧啶糖基化酶抑制剂(UGI)域；和(iv)第二尿嘧啶糖基化酶抑制剂(UGI)域。

138.权利要求137的融合蛋白，其中所述核酸可编程DNA结合蛋白(napDNAbp)是CasX、

CasY、Cpf1、Cpf1切口酶、dCpf1、C2c1、C2c2、C2c3、Cas9、dCas9、Cas9切口酶或Argonaute蛋白。

139.权利要求138的融合蛋白，其中所述dCas9蛋白包含与SEQ ID NO:6或7至少90％相同的氨基酸序列。

140.权利要求137或138的融合蛋白，其中所述dCas9蛋白包含SEQ ID NO:7的氨基酸序列。

141.权利要求138的融合蛋白，其中所述dCas9蛋白是无核酸酶活性的金黄色葡萄球菌(Staphylococcus aureus)Cas9(SaCas9d)蛋白。

142.权利要求141的融合蛋白，其中所述SaCas9d蛋白包含与SEQ ID NO:33-36至少85％、90％、95％或98％相同的氨基酸序列。

143.权利要求141或142的融合蛋白，其中所述SaCas9d蛋白包含SEQ ID NO:34的氨基酸序列。

144.权利要求137-143中任一项的融合蛋白，其中所述胞苷脱氨酶域是来自载脂蛋白B mRNA编辑复合物(APOBEC)家族脱氨酶的脱氨酶。

145.权利要求144的融合蛋白，其中所述APOBEC家族脱氨酶选自下组：APOBEC1脱氨酶、APOBEC2脱氨酶、APOBEC3A脱氨酶、APOBEC3B脱氨酶、APOBEC3C脱氨酶、APOBEC3D脱氨酶、APOBEC3F脱氨酶、APOBEC3G脱氨酶、和APOBEC3H脱氨酶。

146.权利要求137-145中任一项的融合蛋白，其中所述胞苷脱氨酶域包含与SEQ ID NO:49-84的氨基酸序列至少85％、90％、95％或98％相同的氨基酸序列。

147.权利要求137-146中任一项的融合蛋白，其中所述胞苷脱氨酶域包含SEQ ID NO:49-84的氨基酸序列。

148.权利要求137-143中任一项的融合蛋白，其中所述胞苷脱氨酶域是激活诱导的脱氨酶(AID)。

149.权利要求137-143中任一项的融合蛋白，其中所述胞苷脱氨酶域是胞苷脱氨酶1(CDA1)。

150.权利要求137-149中任一项的融合蛋白，其中所述UGI域包含能够抑制UDG活性的域。

151.权利要求137-150中任一项的融合蛋白，其中所述UGI域包含与SEQ ID NO:134至少85％、90％、95％或98％相同的氨基酸序列。

152.权利要求137-150中任一项的融合蛋白，其中所述UGI域包含以SEQ ID NO:134所示的氨基酸序列。

153.权利要求137-152中任一项的融合蛋白，其中所述融合蛋白包含以下结构：

NH_2-[胞苷脱氨酶域]-[napDNAbp]-[第一UGI域]-[第二UGI域]-COOH；

NH_2-[第一UGI域]-[第二UGI域]-[胞苷脱氨酶域]-[napDNAbp]-COOH；

NH_2-[napDNAbp]-[胞苷脱氨酶域]-[第一UGI域]-[第二UGI域]-COOH;或

NH_2-[第一UGI域]-[第二UGI域]-[napDNAbp]-[胞苷脱氨酶域]-COOH；

其中“-”的每个情况包括任选的接头。

154.权利要求153中任一项的融合蛋白，其中所述胞苷脱氨酶域和所述napDNAbp域经由接头连接，所述接头包含以下基序或其组合：氨基酸序列$(GGGS)_n$(SEQ ID NO:613)、

$(GGGGS)_n$(SEQ ID NO:607)、$(G)_n$(SEQ ID NO:608)、$(EAAAK)_n$(SEQ ID NO:609)、$(GGS)_n$(SEQ ID NO:610)、$(SGGS)_n$(SEQ ID NO:606)、SGSETPGTSESATPES(SEQ ID NO:604)、SGGS$(GGS)_n$(SEQ ID NO:612)、SGGSSGGSSGSETPGTSESATPESSGGSSGGS(SEQ ID NO:605)或$(XP)_n$(SEQ ID NO:611),其中n独立地是1至30之间的整数,包括1和30,并且X是任意氨基酸。

155.权利要求137-154中任一项的融合蛋白,其中所述胞苷脱氨酶域和所述napDNAbp经由包含氨基酸序列SGGSSGGSSGSETPGTSESATPESSGGSSGGS(SEQ ID NO:605)的接头连接。

156.权利要求137-154中任一项的融合蛋白,其中所述napDNAbp和所述第一UGI域经由包含氨基酸序列SGGS$(GGS)_n$(SEQ ID NO:612)的接头连接,其中n是2。

157.权利要求137-154中任一项的融合蛋白,其中所述第一UGI域和所述第二UGI域经由包含氨基酸序列SGGS$(GGS)_n$(SEQ ID NO:612)的接头连接,其中n是2。

158.权利要求137-157中任一项的融合蛋白,其进一步包含核定位序列(NLS)。

159.权利要求158的融合蛋白,其中所述NLS包含氨基酸序列PKKKRKV(SEQ ID NO:614)或KRTADGSEFEPKKKRKV(SEQ ID NO:740)。

160.权利要求158或159的融合蛋白,其中所述融合蛋白包含以下结构:

$[NH_2]$-[胞苷脱氨酶域]-[napDNAbp]-[第一UGI域]-[第二UGI域]-[NLS]-[COOH],并且“-”的每个情况包含任选的接头。

161.权利要求158-160中任一项的融合蛋白,其中所述融合蛋白包含以下结构:

[胞苷脱氨酶域]-[dCas9]-[第一UGI域]-[第二UGI域]-[NLS],并且“-”的每个情况包含任选的接头。

162.权利要求160-161中任一项的融合蛋白,其中所述第二UGI域和所述NLS经由包含氨基酸序列SGGS(SEQ ID NO:606)的接头连接。

163.复合物,其包含权利要求137-162中任一项的融合蛋白和与所述融合蛋白的napDNAbp结合的引导RNA。

164.方法,其包括使核酸分子与权利要求137-162中任一项的融合蛋白和引导RNA接触,其中所述引导RNA包含至少10个连续核苷酸的序列,其与生物体的基因组中的靶序列互补,并且包含靶碱基对。

165.权利要求164的方法,其中所述靶碱基对包含与疾病或病症相关的T至C点突变,并且其中所述突变体C碱基的脱氨基导致与疾病或病症无关的序列。

166.权利要求164或165的方法,其中所述接触导致碱基编辑后小于20%的插入/缺失形成。

167.权利要求164或165的方法,其中所述靶序列包含与疾病或病症相关的点突变。

168.权利要求164-167中任一项的方法,其中所述融合蛋白或所述复合物的活性导致所述点突变的校正。

169.权利要求164-165中任一项的方法,其中所述靶序列包含与疾病或病症相关的T至C点突变,并且其中所述突变体C碱基的脱氨基导致与所述疾病或病症无关的序列。

170.权利要求169的方法,其中所述靶序列编码蛋白质,并且其中所述点突变是在密码子中,并且与野生型密码子相比导致由所述突变体密码子编码的氨基酸的变化。

171.权利要求169或170的方法,其中所述突变体C的脱氨基导致由所述突变体密码子编码的氨基酸的变化。

172.权利要求169或170的方法，其中所述突变体C的脱氨基导致编码野生型氨基酸的密码子。

173.权利要求164-172中任一项的方法，其中所述接触在受试者体内进行。

174.权利要求164-172中任一项的方法，其中所述接触在体外进行。

175.权利要求173的方法，其中所述受试者已被诊断出患有疾病或病症。

176.用于编辑双链DNA序列的核碱基对的方法，所述方法包括：

使所述双链DNA序列的靶区域与包含核碱基编辑器和引导核酸的复合物接触，其中所述靶区域包含靶核碱基对；

诱导所述靶区域的链分离；

将所述靶区域的单链中所述靶核碱基对的第一核碱基转化为第二核碱基；并且

切割所述靶区域的不超过一条链；

其中与所述第一核碱基互补的第三核碱基用与所述第二核碱基互补的第四核碱基替换；

其中所述方法导致所述双链DNA序列中小于20％的插入/缺失形成；且

其中所述核碱基编辑器包含CasX、CasY、Cpf1、C2c1、C2c2、C2c3或Argonaute。

177.权利要求176的方法，其中所述方法引起小于19％，18％，17％，16％，15％，14％，13％，12％，11％，10％，9％，8％，7％，6％，5％，4％，3％，2％或1％的插入/缺失形成。

178.权利要求176的方法，其进一步包括用与所述第四核碱基互补的第五核碱基替换所述第二核碱基，从而产生意图编辑的碱基对。

179.权利要求178的方法，其中产生所述意图编辑的碱基对的效率为至少5％。

180.权利要求179的方法，其中所述效率为至少10％、20％、30％、40％或50％。

181.权利要求178的方法，其中所述意图编辑的碱基对与非意图编辑的碱基对的比率为2:1、3:1、4:1、5:1、6:1、7:1或8:1。

182.权利要求178的方法，其中意图编辑的碱基对与插入/缺失形成的比率为2:1至1000:1。

183.权利要求176的方法，其中所述切割的单链与所述引导核酸杂交。

184.权利要求176的方法，其中所述切割的单链与包含所述第一核碱基的链相对。

185.权利要求176的方法，其中所述第一碱基是胞嘧啶。

186.权利要求176的方法，其中所述第二核碱基是非标准核碱基。

187.权利要求176的方法，其中所述第二核碱基不同于G、C、A或T。

188.权利要求176的方法，其中所述第二碱基是尿嘧啶。

189.权利要求176的方法，其中所述核碱基编辑器包含UGI活性。

190.权利要求176的方法，其中所述核碱基编辑器包含UGI域。

191.权利要求176的方法，其中所述核碱基编辑器包含两个UGI域。

192.权利要求176的方法，其中所述核碱基编辑器包括切口酶活性。

193.权利要求178的方法，其中所述意图编辑的碱基对在PAM位点的上游。

194.权利要求193的方法，其中所述意图编辑的碱基对在PAM位点上游的1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个核苷酸。

195.权利要求178的方法，其中所述意图编辑的碱基对在PAM位点的下游。

196.权利要求195的方法，其中所述意图编辑的碱基对PAM位点下游的1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个核苷酸。

197.权利要求176的方法，其中所述方法不需要规范的PAM位点。

198.权利要求176的方法，其中所述核碱基编辑器包括接头。

199.权利要求198的方法，其中所述接头的长度为1-25个氨基酸。

200.权利要求198的方法，其中所述接头的长度为5-20个氨基酸。

201.权利要求198的方法，其中所述接头的长度为10、11、12、13、14、15、16、17、18、19或20个氨基酸。

202.权利要求178的方法，其中所述靶区域包括靶窗，其中所述靶窗包含靶核碱基对。

203.权利要求202的方法，其中所述靶窗包含1-10个核苷酸。

204.权利要求202的方法，其中所述靶窗的长度为1-9、1-8、1-7、1-6、1-5、1-4、1-3、1-2或1个核苷酸。

205.权利要求202的方法，其中所述靶窗的长度是1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个核苷酸。

206.权利要求202-205的方法，其中所述意图编辑的碱基对在所述靶窗内发生。

207.权利要求202-206的方法，其中所述靶窗包含所述意图编辑的碱基对。

包含核酸可编程DNA结合蛋白的核碱基编辑器

[0001] 发明背景

[0002] 核酸序列的靶向编辑，例如基因组DNA的靶向切割或特定修饰对基因组DNA的靶向引入是研究基因功能的非常有希望的方法，并且还具有提供针对人遗传病的新疗法的潜力[1]。理想的核酸编辑技术拥有三个特征：(1)安装期望的修饰的高效率；(2)最小的脱靶活性；和(3)能够被编程以精确编辑给定核酸中任何位点，例如人基因组内的任何位点的能力[2]。当前的基因组工程工具，包括工程化的锌指核酸酶(ZFN)[3]、转录激活物样效应物核酸酶(TALEN)[4]，和最近地RNA引导的DNA内切核酸酶Cas9[5]实现基因组中序列特异性的DNA切割。此可编程的切割可通过非同源末端连接(NHEJ)导致切割位点处的DNA突变，或通过同源性指导的修复(HDR)导致切割位点周围的DNA替换[6,7]。

[0003] 当前技术的一个缺点是NHEJ和HDR都是随机过程，其通常导致适度的基因编辑效率以及可与期望的改变竞争的不想要的基因改变[8]。由于原则上可以通过实现基因组中特定位置处的特定核苷酸变化(例如，与疾病相关的基因的特定密码子中的C至T变化)治疗许多遗传疾病[9]，实现此类精确基因编辑的可编程方式的开发将代表强大的新研究工具以及得到基于基因编辑的人治疗剂的潜在新方法两者。

[0004] 发明概述

[0005] 核酸可编程DNA结合蛋白(napDNAbp)，例如成簇规则间隔短回文重复序列(CRISPR)系统，是最近发现的原核适应性免疫系统[10]，其已经经过修饰以实现多种生物体中的稳健且通用的基因组工程[11]。CRISPR-Cas(与CRISPR相关的)系统是蛋白质-RNA复合物，其使用RNA分子(sgRNA)作为引导物以通过碱基配对将复合物定位于靶DNA序列[12]。在自然系统中，Cas蛋白然后作用为内切核酸酶以切割靶定的DNA序列[13]。靶DNA序列必须既与sgRNA互补，又在互补区3’端还包含“原间隔物邻近基序”(PAM)以使系统发挥功能[14]。

[0006] 在已知的Cas蛋白中，酿脓链球菌Cas9已广泛用作基因组工程的工具[15]。此Cas9蛋白是一种大的多域蛋白，其含有两个不同的核酸酶域。可以将点突变引入Cas9中以消除核酸酶活性，从而导致死的Cas9(dCas9)，其仍然保留其以sgRNA编程方式结合DNA的能力[16]。原则上，当与另一种蛋白质或域融合时，dCas9仅仅通过与适当的sgRNA共表达就可以将该蛋白质或域靶向到几乎任何DNA序列。

[0007] dCas9复合物用于基因组工程目的潜力是巨大的。在理论上，它将蛋白质带到由sgRNA编程的基因组中特定位点的独特能力可以开发成核酸酶以外的多种位点特异性基因组工程工具，包括脱氨酶(例如胞苷脱氨酶)、转录激活物、转录阻遏物、组蛋白修饰蛋白、整合酶和重组酶[11]。这些潜在的应用中的一些最近已通过与转录激活物的dCas9融合来实现，以提供RNA引导的转录激活物[17,18]，转录阻遏物[16,19,20]和染色质修饰酶[21]。这些融合物与多种sgRNA的简单共表达导致靶基因的特异性表达。这些有潜力的研究已经为设计和构建容易可编程的序列特异性效应物以精确操作基因组铺平了道路。

[0008] 本公开内容的一些方面是基于以下认识：由接头融合的核酸可编程DNA结合蛋白(napDNAbp)，例如CasX，CasY，Cpf1，C2c1，C2c2，C2c3或Argonaute蛋白和胞苷脱氨酶域的某些构造可用于有效使靶胞苷残基脱氨基。本公开内容的其他方面涉及以下认识：具有通过

接头与napDNAbp的N末端融合的胞苷脱氨酶域的核碱基编辑融合蛋白能够有效地使双链DNA靶分子中的靶核酸脱氨基。参见例如下面的实施例3和4,其证明了融合蛋白(在本文中也称为碱基编辑器)比其他碱基编辑器(例如没有UGI域的碱基编辑器)产生更少的插入/缺失并更有效地使靶核酸脱氨基。本公开内容的其他方面涉及以下认识:当融合蛋白由超过一个UGI域构成时,具有通过接头与napDNAbp的N末端融合的胞苷脱氨酶域的核碱基编辑融合蛋白以更高的效率进行碱基编辑并且大大改善产物纯度。参见例如实施例17,其证明包含两个UGI域的融合蛋白(例如,碱基编辑器)与其他碱基编辑器(例如包含一个UGI域的那些)相比,产生更少的插入/缺失并更有效地使靶核酸脱氨基。

[0009] 在一些实施方案中,融合蛋白包含:(i)核酸可编程DNA结合蛋白(napDNAbp);(ii)胞苷脱氨酶域;和(iii)尿嘧啶糖基化酶抑制剂(UGI)域,其中napDNAbp是CasX,CasY,Cpf1,C2c1,C2c2,C2c3或Argonaute蛋白。在一些实施方案中,核酸可编程DNA结合蛋白(napDNAbp)是CasX蛋白。在一些实施方案中,CasX蛋白包含与SEQ ID NO:29或30至少90%相同的氨基酸序列。在一些实施方案中,CasX蛋白包含SEQ ID NO:29或30的氨基酸序列。

[0010] 在一些实施方案中,融合蛋白包含:(i)核酸可编程DNA结合蛋白(napDNAbp);(ii)胞苷脱氨酶域;(iii)第一尿嘧啶糖基化酶抑制剂(UGI)域;和(iv)第二尿嘧啶糖基化酶抑制剂(UGI)域,其中napDNAbp是Cas9,dCas9或Cas9切口酶蛋白。在一些实施方案中,napDNAbp是dCas9蛋白。在一些实施方案中,napDNAbp是CasX,CasY,Cpf1,C2c1,C2c2,C2c3或Argonaute蛋白。在一些实施方案中,dCas9蛋白是酿脓链球菌dCas9(SpCas9d)。在一些实施方案中,dCas9蛋白是含有D10A突变的酿脓链球菌dCas9。在一些实施方案中,dCas9蛋白包含与SEQ ID NO:6或7至少90%相同的氨基酸序列。在一些实施方案中,dCas9蛋白包含SEQ ID NO:6或7的氨基酸序列。在一些实施方案中,dCas9蛋白是金黄色葡萄球菌dCas9(SaCas9d)。在一些实施方案中,dCas9蛋白是含有D10A突变的金黄色葡萄球菌dCas9。在一些实施方案中,dCas9蛋白包含与SEQ ID NO:33-36至少90%相同的氨基酸序列。在一些实施方案中,dCas9蛋白包含SEQ ID NO:33-36的氨基酸序列。

[0011] 在一些实施方案中,核酸可编程DNA结合蛋白(napDNAbp)是CasY蛋白。在一些实施方案中,CasY蛋白包含与SEQ ID NO:31至少90%相同的氨基酸序列。在一些实施方案中,CasY蛋白包含SEQ ID NO:31的氨基酸序列。

[0012] 在一些实施方案中,核酸可编程DNA结合蛋白(napDNAbp)是Cpf1或Cpf1突变体蛋白。在一些实施方案中,Cpf1或Cpf1突变体蛋白包含与SEQ ID NO:9-24中的任一项至少90%相同的氨基酸序列。在一些实施方案中,Cpf1或Cpf1突变体蛋白包含SEQ ID NO:9-24中任一项的氨基酸序列。

[0013] 在一些实施方案中,核酸可编程DNA结合蛋白(napDNAbp)是C2c1蛋白。在一些实施方案中,C2c1蛋白包含与SEQ ID NO:26至少90%相同的氨基酸序列。在一些实施方案中,C2c1蛋白包含SEQ ID NO:26的氨基酸序列。

[0014] 在一些实施方案中,核酸可编程DNA结合蛋白(napDNAbp)是C2c2蛋白。在一些实施方案中,C2c2蛋白包含与SEQ ID NO:27至少90%相同的氨基酸序列。在一些实施方案中,C2c2蛋白包含SEQ ID NO:27的氨基酸序列。

[0015] 在一些实施方案中,核酸可编程DNA结合蛋白(napDNAbp)是C2c3蛋白。在一些实施方案中,C2c3蛋白包含与SEQ ID NO:28至少90%相同的氨基酸序列。在一些实施方案中,

C2c3蛋白包含SEQ ID NO:28的氨基酸序列。

[0016] 在一些实施方案中,核酸可编程DNA结合蛋白(napDNAbp)是Argonaute蛋白。在一些实施方案中,Argonaute蛋白包含与SEQ ID NO:25至少90%相同的氨基酸序列。在一些实施方案中,Argonaute蛋白包含SEQ ID NO:25的氨基酸序列。

[0017] 本公开内容的一些方面是基于以下认识:本文提供的融合蛋白能够产生一个或多个突变(例如,C至T突变)而不产生大比例的插入/缺失。在一些实施方案中,本文提供的任何融合蛋白(例如,碱基编辑器蛋白)产生小于10%的插入/缺失。在一些实施方案中,本文提供的任何融合蛋白(例如,碱基编辑器蛋白)产生小于10%,9%,8%,7%,6%,5.5%,5%,4.5%,4%,3.5%,3%,2.5%,2%,1.5%,1%,0.5%或0.1%插入/缺失。

[0018] 在一些实施方案中,融合蛋白包含napDNAbp和载脂蛋白B mRNA编辑复合物1(APOBEC1)脱氨酶域,其中脱氨酶域通过包含氨基酸序列SGSETPGTSESATPES(SEQ ID NO:604)的接头与napDNAbp域的N末端融合。在一些实施方案中,napDNAbp包含本文提供的任何napDNAbp的氨基酸序列。在一些实施方案中,脱氨酶是大鼠APOBEC1(SEQ ID NO:76)。在一些实施方案中,脱氨酶是人APOBEC1(SEQ ID NO:74)。在一些实施方案中,脱氨酶是pmCDA1(SEQ ID NO:81)。在一些实施方案中,脱氨酶是人APOBEC3G(SEQ ID NO:60)。在一些实施方案中,脱氨酶是(SEQ ID NO:82-84)中任一项的人APOBEC3G变体。在一些实施方案中,融合蛋白包含napDNAbp和载脂蛋白B mRNA编辑复合物1催化多肽样3G(APOBEC3G)脱氨酶域,其中所述脱氨酶域通过任何长度或组成(例如氨基酸序列,肽,聚合物或键)的接头与napDNAbp域的N末端融合。在一些实施方案中,接头包含氨基酸序列SGSETPGTSESATPES(SEQ ID NO:604)。在一些实施方案中,接头包含氨基酸序列SGGSSGGSSGSETPGTSESATPESSGGSSGGS(SEQ ID NO:605)。

[0019] 在一些实施方案中,融合蛋白包含napDNAbp和胞苷脱氨酶1(CDA1)脱氨酶域,其中所述脱氨酶域通过包含氨基酸序列SGSETPGTSESATPES(SEQ ID NO:604)的接头与napDNAbp域的N末端融合。在一些实施方案中,接头包含氨基酸序列SGGSSGGSSGSETPGTSESATPESSGGSSGGS(SEQ ID NO:605)。在一些实施方案中,napDNAbp包含本文提供的任何napDNAbps的氨基酸序列。

[0020] 在一些实施方案中,融合蛋白包含napDNAbp和活化诱导的胞苷脱氨酶(AID)脱氨酶域,其中脱氨酶域通过包含氨基酸序列SGSETPGTSESATPES(SEQ ID NO:604)的接头与napDNAbp域的N末端融合。在一些实施方案中,接头包含氨基酸序列SGGSSGGSSGSETPGTSESATPESSGGSSGGS(SEQ ID NO:605)。在一些实施方案中,napDNAbp包含本文提供的任何napDNAbps的氨基酸序列。

[0021] 本公开内容的一些方面是基于以下认识:由接头融合的napDNAbp和胞苷脱氨酶域的某些构造可用于有效使靶胞苷残基脱氨基。本公开内容的其他方面涉及以下认识:具有通过包含氨基酸序列SGSETPGTSESATPES(SEQ ID NO:604)的接头与napDNAbp的N末端融合的载脂蛋白B mRNA编辑复合物1(APOBEC1)脱氨酶域的核碱基编辑融合蛋白能够有效使双链DNA靶分子中的靶核酸脱氨基。在一些实施方案中,融合蛋白包含napDNAbp域和载脂蛋白B mRNA编辑复合物1(APOBEC1)脱氨酶域,其中脱氨酶域通过包含氨基酸序列SGSETPGTSESATPES(SEQ ID NO:604)的接头与napDNAbp的N末端融合。

[0022] 本公开内容的一些方面提供了可用于靶向编辑核酸的策略、系统、试剂、方法和试

剂盒，包括编辑受试者的基因组，例如人的基因组内的单个位点。在一些实施方案中，提供了napDNAbp(例如，CasX，CasY，Cpf1，C2c1，C2c2，C2c3或Argonaute蛋白)和脱氨酶或脱氨酶域的融合蛋白。在一些实施方案中，提供了用于靶向核酸编辑的方法。在一些实施方案中，提供了用于产生靶向核酸编辑蛋白，例如napDNAbp和脱氨酶或脱氨酶域的融合蛋白的试剂和试剂盒。

[0023] 本公开内容的一些方面提供了包含如本文提供的napDNAbp的融合蛋白，所述napDNAbp与第二蛋白(例如，酶促域如胞苷脱氨酶域)融合，从而形成融合蛋白。在一些实施方案中，第二蛋白质包含酶促域或结合域。在一些实施方案中，酶促域是核酸酶、切口酶、重组酶、脱氨酶、甲基转移酶、甲基化酶、乙酰基酯酶、乙酰基转移酶、转录激活物或转录阻遏物域。在一些实施方案中，酶促域是核酸编辑域。在一些实施方案中，核酸编辑域是脱氨酶域。在一些实施方案中，脱氨酶是胞嘧啶脱氨酶或胞苷脱氨酶。在一些实施方案中，脱氨酶是载脂蛋白B mRNA编辑复合物(APOBEC)家族脱氨酶。在一些实施方案中，脱氨酶是APOBEC1脱氨酶。在一些实施方案中，脱氨酶是APOBEC2脱氨酶。在一些实施方案中，脱氨酶是APOBEC3脱氨酶。在一些实施方案中，脱氨酶是APOBEC3A脱氨酶。在一些实施方案中，脱氨酶是APOBEC3B脱氨酶。在一些实施方案中，脱氨酶是APOBEC3C脱氨酶。在一些实施方案中，脱氨酶是APOBEC3D脱氨酶。在一些实施方案中，脱氨酶是APOBEC3E脱氨酶。在一些实施方案中，脱氨酶是APOBEC3F脱氨酶。在一些实施方案中，脱氨酶是APOBEC3G脱氨酶。在一些实施方案中，脱氨酶是APOBEC3H脱氨酶。在一些实施方案中，脱氨酶是APOBEC4脱氨酶。在一些实施方案中，脱氨酶是活化诱导的脱氨酶(AID)。应当理解，脱氨酶可以来自任何合适的生物体(例如，人或大鼠)。在一些实施方案中，脱氨酶来自人，黑猩猩，大猩猩，猴，牛，狗，大鼠或小鼠。在一些实施方案中，脱氨酶是大鼠APOBEC1(SEQ ID NO:76)。在一些实施方案中，脱氨酶是人APOBEC1(SEQ ID NO:74)。在一些实施方案中，脱氨酶是pmCDA1。

[0024] 本公开内容的一些方面提供了融合蛋白，其包含：(i)CasX，CasY，Cpf1，C2c1，C2c2，C2c3或Argonaute蛋白域，其包含SEQ ID NO:32的氨基酸序列；(ii)载脂蛋白B mRNA编辑复合物1(APOBEC1)脱氨酶域，其中所述脱氨酶域通过包含SGSETPGTSESATPES(SEQ ID NO:604)的氨基酸序列的接头与napDNAbp的N末端融合。在一些实施方案中，脱氨酶是大鼠APOBEC1(SEQ ID NO:76)。在一些实施方案中，脱氨酶是人APOBEC1(SEQ ID NO:74)。在一些实施方案中，融合蛋白包含SEQ ID NO:591的氨基酸序列。在一些实施方案中，融合蛋白包含SEQ ID NO:5737的氨基酸序列。在一些实施方案中，脱氨酶是pmCDA1(SEQ ID NO:81)。在一些实施方案中，脱氨酶是人APOBEC3G(SEQ ID NO:60)。在一些实施方案中，脱氨酶是SEQ ID NO:82-84中任一项的人APOBEC3G变体。

[0025] 本公开内容的其他方面涉及以下认识：包含脱氨酶域、napDNAbp域和尿嘧啶糖基化酶抑制剂(UGI)域的融合蛋白表现出改善的使核酸分子中靶核苷酸脱氨基的效率。不希望受到任何特定理论的束缚，对U:G异双链体DNA的存在的细胞DNA修复应答可以导致细胞中核碱基编辑效率的降低。尿嘧啶DNA糖基化酶(UDG)催化从细胞中的DNA中除去U，这可以启动碱基切除修复，其中最常见的结果是将U:G对转化为C:G对。如本文中证明，尿嘧啶DNA糖基化酶抑制剂(UGI)可抑制人UDG活性。不希望受任何特定理论的束缚，碱基切除修复可以被结合单链，阻断经编辑的碱基，抑制UGI，抑制碱基切除修复，保护经编辑的碱基和/或促进非编辑链的“固定”的分子抑制。因此，本公开内容涵盖了包含与UGI域融合的

napDNAbp-胞苷脱氨酶域的融合蛋白。

[0026] 本公开内容的其他方面涉及如下的认识:包含脱氨酶域、napDNAbp域和超过一个尿嘧啶糖基化酶抑制剂(UGI)域(例如,一个,两个,三个,四个,五个或更多个UGI域)的融合蛋白表现出改善的使核酸分子中的靶核苷酸脱氨基的效率和/或改善的核酸产物纯度。不希望受到任何特定理论的束缚,第二UGI域的添加可以实质上减少UDG对G:U碱基编辑中间体的接近,从而改善碱基编辑的效率。

[0027] 本公开内容的一些方面是基于以下认识:本文提供的任何碱基编辑器能够修饰特定的核苷酸碱基而不产生相当大比例的插入/缺失。如本文所用“,插入/缺失”是指核酸内核苷酸碱基的插入或缺失。此类插入或缺失可导致基因编码区内的移码突变。在一些实施方案中,期望产生有效地修饰(例如突变或脱氨基)核酸内的特定核苷酸而不在核酸中产生插入或缺失(即插入/缺失)的碱基编辑器。在某些实施方案中,本文提供的任何碱基编辑器能够产生更大比例的意图修饰(例如,点突变或脱氨基)与插入/缺失。

[0028] 在某些实施方案中,本文提供的任何碱基编辑器能够产生一定百分比的期望突变。在一些实施方案中,期望的突变是C至T突变。在一些实施方案中,期望的突变是C至A突变。在一些实施方案中,期望的突变是C至G突变。在一些实施方案中,本文提供的任何碱基编辑器能够产生至少1%的期望突变。在一些实施方案中,本文提供的任何碱基编辑器能够产生至少1%,2%,3%,4%,5%,10%,15%,20%,25%,30%,40%,45%,50%,60%,70%,80%,90%,95%或99%期望突变。

[0029] 本公开内容的一些方面是基于以下的认识,即本文提供的任何碱基编辑器能够在核酸(例如受试者的基因组内的核酸)中有效产生意图的突变,例如点突变,而不产生大量的非意图突变,例如非意图的点突变。

[0030] 在一些实施方案中,融合蛋白的脱氨酶域与napDNAbp域的N末端融合。在一些实施方案中,UGI域与napDNAbp域的C末端融合。在一些实施方案中,napDNAbp和核酸编辑域通过接头融合。在一些实施方案中,napDNAbp域和UGI域通过接头融合。在一些实施方案中,第二UGI域与第一UGI域的C末端融合。在一些实施方案中,第一UGI域和第二UGI域通过接头融合。

[0031] 在某些实施方案中,接头可用于连接本发明的任何肽或肽域。接头可以仅是共价键,或者它可以是许多原子长度的聚合接头。在某些实施方案中,接头是多肽或基于氨基酸。在其他实施方案中,接头不是肽样的。在某些实施方案中,接头是共价键(例如,碳-碳键、二硫键、碳-杂原子键等)。在某些实施方案中,接头是酰胺连接的碳-氮键。在某些实施方案中,接头是环状或无环的,取代或未取代的,分支或非分支的脂族或杂脂族接头。在某些实施方案中,接头是聚合的(例如,聚乙烯,聚乙二醇,聚酰胺,聚酯等)。在某些实施方案中,接头包含氨基链烷酸的单体,二聚体或多聚体。在某些实施方案中,接头包含氨基链烷酸(例如,甘氨酸,乙酸,丙氨酸,β-丙氨酸,3-氨基丙酸,4-氨基丁酸,5-戊酸等)。在某些实施方案中,接头包含氨基己酸(Ahx)的单体,二聚体或多聚体。在某些实施方案中,接头基于碳环部分(例如环戊烷,环己烷)。在其他实施方案中,接头包含聚乙二醇部分(PEG)。在其他实施方案中,接头包含氨基酸。在某些实施方案中,接头包含肽。在某些实施方案中,接头包含芳基或杂芳基部分。在某些实施方案中,接头基于苯环。接头可以包括官能化部分,以促进亲核试剂(例如硫醇,氨基)从肽至接头的连接。任何亲电体可以用作接头的一部分。示例

性亲电体包括但不限于活化的酯、活化的酰胺、迈克尔受体、烷基卤化物、芳基卤化物、酰基卤化物和异硫氰酸酯。

[0032] 在一些实施方案中，接头包含氨基酸序列(GGGGS)$_n$(SEQ ID NO:607)，(G)$_n$(SEQ ID NO:608)，(EAAAK)$_n$(SEQ ID NO:609)，(GGS)$_n$(SEQ ID NO:610)，(SGGS)$_n$(SEQ ID NO:606)，SGSETPGTSESATPES(SEQ ID NO:604)，(XP)$_n$(SEQ ID NO:611)，SGGS(GGS)$_n$(SEQ ID NO:612)，SGGSSGGSSGSETPGTSESATPESSGGSSGGS(SEQ ID NO:605)或其任意组合，其中n独立地是1和30之间的整数，且X是任意氨基酸。在一些实施方案中，接头包含氨基酸序列(GGS)$_n$(SEQ ID NO:610)，其中n是1、3或7。在一些实施方案中，接头包含氨基酸序列SGGS(GGS)$_n$(SEQ ID NO:612)，其中n为2。在一些实施方案中，接头包含氨基酸序列SGSETPGTSESATPES(SEQ ID NO:604)。在一些实施方案中，接头包含氨基酸序列SGGSSGGSSGSETPGTSESATPESSGGSSGGS(SEQ ID NO:605)。

[0033] 在一些实施方案中，融合蛋白包含结构[核酸编辑域]-[任选的接头序列]-[napDNAbp]-[任选的接头序列]-[UGI]。在一些实施方案中，融合蛋白包含结构[核酸编辑域]-[任选的接头序列]-[UGI]-[任选的接头序列]-[napDNAbp]；[UGI]-[任选的接头序列]-[核酸编辑域]-[任选的接头序列]-[napDNAbp]；[UGI]-[任选的接头序列]-[napDNAbp]-[任选的接头序列]-[核酸编辑域]；[napDNAbp]-[任选的接头序列]-[UGI]-[任选的接头序列]-[核酸编辑域]；[napDNAbp]-[任选的接头序列]-[核酸编辑域]-[任选的接头序列]-[UGI]；或[核酸编辑域]-[任选的接头序列]-[napDNAbp]-[任选的接头序列]-[第一UGI]-[任选的接头序列]-[第二UGI]。

[0034] 在一些实施方案中，核酸编辑域包含脱氨酶。在一些实施方案中，核酸编辑域包含脱氨酶。在一些实施方案中，脱氨酶是胞苷脱氨酶。在一些实施方案中，脱氨酶是载脂蛋白B mRNA编辑复合物(APOBEC)家族脱氨酶。在一些实施方案中，脱氨酶是APOBEC1脱氨酶、APOBEC2脱氨酶、APOBEC3A脱氨酶、APOBEC3B脱氨酶、APOBEC3C脱氨酶、APOBEC3D脱氨酶、APOBEC3F脱氨酶、APOBEC3G脱氨酶、APOBEC3H脱氨酶、或APOBEC4脱氨酶。在一些实施方案中，脱氨酶是活化诱导的脱氨酶(AID)。在一些实施方案中，脱氨酶是胞苷脱氨酶1(CDA1)。在一些实施方案中，脱氨酶是八目鳗CDA1(pmCDA1)脱氨酶。

[0035] 在一些实施方案中，脱氨酶来自人，黑猩猩，大猩猩，猴，牛，狗，大鼠或小鼠。在一些实施方案中，脱氨酶来自人。在一些实施方案中，脱氨酶来自大鼠。在一些实施方案中，脱氨酶是包含(SEQ ID NO:76)所示的氨基酸序列的大鼠APOBEC1脱氨酶。在一些实施方案中，脱氨酶是包含(SEQ ID NO:74)所示的氨基酸序列的人APOBEC1脱氨酶。在一些实施方案中，脱氨酶是pmCDA1(SEQ ID NO:81)。在一些实施方案中，脱氨酶是人APOBEC3G(SEQ ID NO:60)。在一些实施方案中，脱氨酶是(SEQ ID NO:82-84)中任一项的人APOBEC3G变体。在一些实施方案中，脱氨酶与SEQ ID NO:49-84所示的氨基酸序列中的任一项至少80％，至少85％，至少90％，至少92％，至少95％，至少96％，至少97％，至少98％，至少99％或至少99.5％相同。

[0036] 在一些实施方案中，UGI域包含与SEQ ID NO:134至少80％，至少85％，至少90％，至少92％，至少95％，至少96％，至少97％，至少98％，至少99％或至少99.5％相同的氨基酸序列。在一些实施方案中，UGI域包含SEQ ID NO:134所示的氨基酸序列。

[0037] 本公开内容的一些方面提供了复合物，其包含如本文提供的napDNAbp融合蛋白和

与napDNAbp结合的引导RNA。

[0038] 本公开内容的一些方面提供了使用本文提供的napDNAbp,融合蛋白或复合物的方法。例如,本公开内容的一些方面提供了方法,其包括使DNA分子(a)与如本文提供的napDNAbp或融合蛋白以及与引导RNA接触,其中,引导RNA长约15-100个核苷酸,并包含与靶序列互补的至少10个连续核苷酸的序列;或者(b)与如本文提供的napDNAbp,napDNAbp融合蛋白或napDNAbp或与gRNA的napDNAbp复合物接触。

[0039] 本公开内容的一些方面提供了包含核酸构建体的试剂盒,所述核酸构建体包含(a)编码如本文提供的napDNAbp或napDNAbp融合蛋白的核苷酸序列;和(b)驱动(a)序列表达的异源启动子。在一些实施方案中,试剂盒进一步包含编码引导RNA主链的表达构建体,其中构建体包含克隆位点,其定位成允许将与靶序列相同或互补的核酸序列克隆到引导RNA主链中。

[0040] 本公开内容的一些方面提供了编码如本文提供的融合蛋白的napDNAbp的多核苷酸。本公开内容的一些方面提供了包含此类多核苷酸的载体。在一些实施方案中,载体包含驱动多核苷酸表达的异源启动子。

[0041] 本公开内容的一些方面提供了细胞,其包含如本文提供的napDNAbp蛋白,融合蛋白,核酸分子和/或载体。

[0042] 应当理解,包含Cas9域(例如Cas9,nCas9或dCas9)的本文提供的任何融合蛋白可以用本文提供的任何napDNAbp,例如CasX,CasY,Cpf1,C2c1,C2c2,C2c3或Argonaute蛋白替换。

[0043] 上述报告物系统的示例性实施方案的描述仅出于说明目的而提供,并不意味着限制。本公开内容还包括其他报告物系统,例如,上文详细描述的示例性系统的变型。

[0044] 以上概述旨在以非限制性方式显示本文所公开的技术的一些实施方案,优点,特征和用途。根据详细描述,附图,实施例和权利要求书,本文公开的技术的其他实施方案,优点,特征和用途将是明显的。

[0045] 附图简述

[0046] 图1显示了脱氨酶在单链DNA底物上的脱氨酶活性。使用随机化PAM序列(NNN PAM)的单链DNA底物用作阴性对照。使用的规范PAM序列包括(NGG PAM)。

[0047] 图2显示了Cas9:脱氨酶融合蛋白在单链DNA底物上的活性。

[0048] 图3显示了Cas9:脱氨酶:sgRNA复合物的双链DNA底物结合。

[0049] 图4显示了双链DNA脱氨基测定法。

[0050] 图5显示了Cas9融合物可以靶向双链DNA靶序列的位置3-11(根据图5中的示意图编号)。上凝胶:1μM rAPOBEC1-GGS-dCas9,125nM dsDNA,1当量sgRNA。中间凝胶:1μM rAPOBEC1-(GGS)3(SEQ ID NO:610)-dCas9,125nM dsDNA,1当量sgRNA。下凝胶:1.85μM rAPOBEC1-XTEN-dCas9,125nM dsDNA,1当量sgRNA。

[0051] 图6证明了脱氨酶活性需要正确的引导RNA,例如正确的sgRNA。

[0052] 图7显示了脱氨酶-dCas9:sgRNA复合物对体内靶序列的靶DNA结合的机制。

[0053] 图8显示了示例性疾病相关靶序列的成功脱氨基。

[0054] 图9显示了使用His6-rAPOBEC1-XTEN-dCas9的体外C→T编辑效率。

[0055] 图10显示了通过与UGI融合大大增强HEK293T细胞中的C→T编辑效率。

[0056] 图11A至11C显示了NBE1在体外介导特异性的引导RNA程序的C至U转化。图11A:核碱基编辑策略。在引导RNA规定的基因座处具有靶C的DNA被dCas9结合,所述dCas9介导DNA底物的局部变性。通过栓系的APOBEC1酶进行的胞嘧啶脱氨基将靶C转化为U。在DNA复制或修复之后,所得的G:U异双链体可以永久转化为A:T碱基对。若U在模板DNA链中,则它还会在转录后产生含有G至A突变的RNA转录物。图11B:脱氨基测定法,显示约五个核苷酸的活性窗。将NBE1-sgRNA复合物与dsDNA底物在37℃温育2h后,5'荧光团标记的DNA分离,并与USER酶(尿嘧啶DNA糖基化酶和内切核酸酶VIII)在37℃下温育1h以诱导任何尿嘧啶位点处的DNA切割。将得到的DNA在变性聚丙烯酰胺凝胶上解析,并显现任何荧光团连接的链。每个泳道根据原间隔物中的靶C位置标记,或者若不存在靶C,则用"-"标记,将PAM远端的碱基计数为位置1。图11C:脱氨酶测定法,显示序列特异性和NBE1的sgRNA依赖性。将在第7位处具有靶C的DNA底物与如图11B中的NBE1及正确的sgRNA、错配的sgRNA或无sgRNA一起温育。错配的sgRNA或无sgRNA均未观察到C至U编辑。阳性对照样品含有具有在第7位上合成掺入的U的DNA序列。

[0057] 图12A至12B显示了序列背景和靶C位置对体外核碱基编辑效率的影响。图12A:改变靶C周围的序列对体外编辑效率的影响。将原间隔物序列5′-TTATTTCGTGGATTTATTTA-3′(SEQ ID NO:591)中C_7的80%靶定链(来自两条链的总测序读段的40%)的脱氨基产率定义为1.0,并且显示了在位置1-6和8-13处含有所有可能的单碱基突变的底物的相对脱氨基效率。值和误差棒反映在不同天数进行的两个或更多个独立生物学重复的平均值和标准差。图12B:每个NC基序对体外编辑效率的位置效应。如右侧所示的序列中指示,每个NC靶基序在原间隔物内的位置1至8变化(以红色显示PAM,也显示了原间隔物加上原间隔物5'的1个碱基)。与NBE1温育后,在每个编号的靶C位置处含有T的总序列读段的百分比显示在图中。注意体外最大可能的脱氨基产率为总测序读段的50%(靶定链的100%)。值和误差棒反映在不同天数进行的两个或三个独立生物学重复的平均值和标准差。图12B从上至下分别描绘了SEQ ID NO:619至626。

[0058] 图13A至13C显示了人细胞中的核碱基编辑。图13A:核碱基编辑器靶向的六个哺乳动物细胞基因组基因座的原间隔物和PAM序列。靶C用与它们在原间隔物内的位置相对应的下标数字指示。图13A从上至下分别描绘了SEQ ID NO:127至132。图13B:用表达NBE1、NBE2或NBE3和合适的sgRNA的质粒转染HEK293T细胞。转染后三天,提取基因组DNA并通过六个基因座处的高通量DNA测序进行分析。对于所有六个基因组基因座处的NBE1、NBE2和NBE3,以及在6个位点中的3处(EMX1,HEK293位点3和HEK293位点4)具有供体HDR模板的wt Cas9显示了细胞C至T转化百分比,定义为在指定的靶位置处具有T的总DNA测序读段的百分比。值和误差棒反映了在不同天数进行的三个独立生物学重复的平均值和标准差。图13C:在对于所有六个基因组基因座用NBE2和NBE3,或者在6个位点的3处(EMX1,HEK293位点3和HEK293位点4)用wt Cas9和用于HDR的单链DNA模板处理HEK293T细胞后显示如方法中所述的那样计算的插入/缺失形成的频率。值反映了在不同天数进行的至少三个独立生物学重复的平均值。

[0059] 图14A至14C显示了NBE2和NBE3介导的哺乳动物细胞中三个与疾病相关的突变的校正。对于每个位点,原间隔物的序列在突变名称的右边指示,PAM和造成突变的碱基以粗体指示,下标数字对应于其在原间隔物中的位置。显示了每个与疾病相关的等位基因上方

的氨基酸序列，以及红色的核碱基编辑后的校正氨基酸序列。每个序列下方是具有相应碱基的总测序读段的百分比。用编码NBE2或NBE3和合适的sgRNA的质粒对细胞进行核转染。核转染后两天，提取基因组DNA并通过HTS分析以评估致病性突变校正。图14A：在总读段的11％(44％的经核转染的星形胶质细胞)中，NBE3在小鼠星形胶质细胞中将与阿尔茨海默氏病相关的APOE4等位基因转换为APOE3’。附近的两个C也被转化为T，但是所得蛋白的预测序列没有变化(SEQ ID NO:627)。图14B在11％的经处理的对突变杂合的人淋巴瘤细胞(经核转染的细胞的12％)中，NBE2校正与癌症相关的p53 N239D突变(SEQ ID NO:628)。图14C在7.6％的核转染的人乳腺癌细胞中NBE3校正p53 Y163C突变(SEQ ID NO:629)。

[0060] 图15A至15D显示脱氨酶-dCas9接头长度和组成对核碱基编辑的影响。基于凝胶的脱氨酶测定法，显示具有GGS(图15A)，$(GGS)_3$(SEQ ID NO:610)(图15B)，XTEN(图15C)或$(GGS)_7$(SEQ ID NO:610)的脱氨酶-Cas9接头的核碱基编辑器的脱氨基窗(图15D)。将1.85μM编辑器-sgRNA复合物与125nMdsDNA底物在37℃温育2小时后，将缀合有染料的DNA分离并与USER酶(尿嘧啶DNA糖基化酶和内切核酸酶VIII)在37℃再温育1小时以在任何尿嘧啶位点处切割DNA主链。将得到的DNA在变性聚丙烯酰胺凝胶上解析，并对缀合有染料的链成像。每个泳道根据靶C在原间隔物内的位置进行编号，若没有靶C，则使用-编号。8U是具有第8位处合成掺入的U的阳性对照序列。

[0061] 图16A至16B显示了NBE1能够在体外校正疾病相关的突变。图16A：7个疾病相关突变的原间隔物和PAM序列。在每种情况下疾病相关的靶C用带下标的数字指示，该数字反映其在原间隔物内的位置。对于除两个APOE4SNP以外的所有突变，靶C驻留于模板(非编码)链中。图16A从上至下分别描绘了SEQ ID NO:631至636。图16B：脱氨酶测定法，显示在与NBE1温育，DNA分离以及与USER酶温育以在含有U的位置处切割DNA之前(-)和(+)之后的每个dsDNA寡核苷酸。用指示U位置的相应数字显示来自在原间隔物内各个位置处含有U的合成寡核苷酸与USER酶温育的阳性对照泳道。

[0062] 图17显示了NBE1的持续合成能力。顶部显示了包含8个连续C的60聚体DNA寡核苷酸的原间隔物和PAM。将寡核苷酸(125nM)与NBE1(2μM)在37℃温育2小时。分离DNA并通过高通量测序进行分析。显示的是观察到的最频繁的九个序列的总读段的百分比。绝大多数编辑链(>93％)具有超过一个转化为T的C。此图描绘了SEQ ID NO:309。

[0063] 图18A至18H显示了将UGI与NBE1融合以生成NBE2的效果。图18A：用核碱基编辑器靶向的六个哺乳动物细胞基因组基因座的原间隔物和PAM序列。可编辑C用与它们在原间隔物内的位置相对应的标记指示。图18A从上至下分别描绘了SEQ ID NO:127至132。图18B至18G：HEK293T细胞用表达NBE1，NBE2或NBE1和UGI以及合适的sgRNA的质粒转染。转染后三天，提取基因组DNA并通过六个基因座处的高通量DNA测序进行分析。对于所有六个基因组基因座处的NBE1、NBE1和UGI、和NBE2显示了细胞C至T转化百分比，定义为在指示的靶位置处具有T的总DNA测序读段的百分比。图18H：显示了NBE1、不同质粒上的NBE1加UGI、NBE2和未处理的细胞的感兴趣的原间隔物周围的510C处的C至T突变率。数据显示来自1.5x106个细胞的3,000,000个DNA测序读段的结果。值反映了在不同天数进行的至少两次生物学实验的平均值。

[0064] 图19显示了U2OS和HEK293T细胞中NBE2的核碱基编辑效率。对于HEK293T细胞和U2OS细胞中的六个靶定的基因组基因座中的每个，显示了通过NBE2得到的细胞C至T转化百

分比。使用lipofectamine 2000转染HEK293T细胞，并对U2OS细胞进行核转染。U2OS核转染效率为74％。质粒递送后三天，提取基因组DNA，并通过HTS分析六个基因组基因座处的核碱基编辑。值和误差棒反映了在不同天数完成的至少两次生物学实验的平均值和标准差。

[0065] 图20显示了核碱基编辑在多个细胞分裂中持续存在。在对细胞进行传代之前和之后，HEK293T细胞中两个基因组基因座处显示了通过NBE2得到的细胞C至T转化百分比。使用Lipofectamine 2000转染HEK293T细胞。转染后三天，收获细胞并分成两半。一半进行HTS分析，并且允许另一半增殖约5次细胞分裂，然后收获并进行HTS分析。

[0066] 图21显示了来自ClinVar的遗传变体，其原则上可以通过核碱基编辑来校正。对人遗传变异及其对应表型的NCBI ClinVar数据库[68]搜索可通过当前的核碱基编辑技术校正的遗传疾病。通过施加左侧列出的连续限制对结果进行过滤。x轴以对数标度显示满足该限制和所有以上限制的出现数目。

[0067] 图22显示了六个基因组基因座中可编辑C的体外鉴定。将具有与六个不同基因组位点匹配的序列的合成80-聚体与NBE1温育，然后通过HTS分析核碱基编辑。对于每个位点，原间隔物的序列在位点名称的右侧指示，PAM用红色突出显示。每个序列下方是具有相应碱基的总DNA测序读段的百分比。若体外转化效率>10％，则靶C视为“可编辑”。注意，由于非靶定链不是核碱基编辑的底物，因此最大产率为DNA测序总读段的50％。该图从上至下分别描绘了SEQ ID NO:127至132。

[0068] 图23显示了在EMX1脱靶处NBE1，NBE2和NBE3的活性。使用Lipofectamine 2000，用表达NBE1，NBE2或NBE3和与EMX1序列匹配的sgRNA的质粒转染HEK293T细胞。转染后三天，提取基因组DNA，通过PCR扩增，并通过在如先前使用GUIDE-seq方法[55]确定的EMX1 sgRNA的中靶基因座，加上前10个已知Cas9脱靶基因座处的高通量DNA测序进行分析。EMX1脱靶5位点未扩增并且未显示。显示了中靶和脱靶的原间隔物和原间隔物邻近基序(PAM)的序列。对于NBE1，NBE2和NBE3，显示了细胞C至T的转化百分比，定义为在原间隔物内原始C的每个位置处具有T的总DNA测序读段的百分比。最右边显示每个序列报告的测序读段总数。该图从上至下分别描绘了SEQ ID NO:127和637至645。

[0069] 图24显示了在FANCF脱靶处NBE1，NBE2和NBE3的活性。使用Lipofectamine 2000，用表达NBE1，NBE2或NBE3和与FANCF序列匹配的sgRNA的质粒转染HEK293T细胞。转染后三天，提取基因组DNA，通过PCR扩增，并通过在如先前使用GUIDE-seq方法[55]确定的FANCF sgRNA的中靶基因座，加上所有已知Cas9脱靶基因座处的高通量DNA测序进行分析。显示了中靶和脱靶原间隔物和原间隔物邻近基序(PAM)的序列。对于NBE1，NBE2和NBE3，显示了细胞C至T的转化百分比，定义为在原间隔物内原始C的每个位置处具有T的总DNA测序读段的百分比。最右边显示每个序列报告的测序读段总数。该图从上至下分别描绘了SEQ ID NO:128和646至653。

[0070] 图25显示了HEB293位点2脱靶处的NBE1，NBE2和NBE3的活性。使用Lipofectamine 2000，用表达NBE1，NBE2或NBE3和与HEK293位点2序列匹配的sgRNA的质粒转染HEK293T细胞。转染后三天，提取基因组DNA，通过PCR扩增，并通过在如先前使用GUIDE-seq方法[55]确定的HEK293位点2sgRNA的中靶基因座，加上所有已知Cas9脱靶基因座处的高通量DNA测序进行分析。显示了中靶和脱靶原间隔物和原间隔物邻近基序(PAM)的序列。对于NBE1，NBE2和NBE3，显示了细胞C至T的转化百分比，定义为在原间隔物内原始C的每个位置处具有T的总

DNA测序读段的百分比。最右边显示每个序列报告的测序读段总数。该图从上至下分别描绘了SEQ ID NO:129、654和655。

[0071] 图26显示了HEB293位点3脱靶处的NBE1,NBE2和NBE3的活性。使用Lipofectamine 2000,用表达NBE1,NBE2或NBE3和与HEK293位点3序列匹配的sgRNA的质粒转染HEK293T细胞。转染后三天,提取基因组DNA,通过PCR扩增,并通过在如先前使用GUIDE-seq方法[55]确定的HEK293位点3sgRNA的中靶基因座,加上所有已知Cas9脱靶基因座处的高通量DNA测序进行分析。显示了中靶和脱靶原间隔物和原间隔物邻近基序(PAM)的序列。对于NBE1,NBE2和NBE3,显示了细胞C至T的转化百分比,定义为在原间隔物内原始C的每个位置处具有T的总DNA测序读段的百分比。最右边显示每个序列报告的测序读段总数。该图从上至下分别描绘了SEQ ID NO:130和656至660。

[0072] 图27显示了HEB293位点4脱靶处的NBE1,NBE2和NBE3的活性。使用Lipofectamine 2000,用表达NBE1,NBE2或NBE3和与HEK293位点4序列匹配的sgRNA的质粒转染HEK293T细胞。转染后三天,提取基因组DNA,通过PCR扩增,并通过在如先前使用GUIDE-seq方法[55]确定的HEK293位点4sgRNA的中靶基因座,加上前10种已知Cas9脱靶基因座处的高通量DNA测序进行分析。显示了中靶和脱靶原间隔物和原间隔物邻近基序(PAM)的序列。对于NBE1,NBE2和NBE3,显示了细胞C至T的转化百分比,定义为在原间隔物内原始C的每个位置处具有T的总DNA测序读段的百分比。最右边显示每个序列报告的测序读段总数。该图从上至下分别描绘了SEQ ID NO:131和661至670。

[0073] 图28显示了非靶C突变率。此处显示的是在测试的6个中靶和34个脱靶基因座周围的2,500个不同胞嘧啶处的C至T突变率,代表从约1.8x106个细胞衍生的总计14,700,000个序列读段。

[0074] 图29A至29C显示了人细胞中的碱基编辑。图29A显示了哺乳动物细胞中可能的碱基编辑结果。初始编辑导致U:G错配。尿嘧啶DNA糖基化酶(UDG)对U的识别和切除引发碱基切除修复(BER),其导致回复至C:G起始状态。BER被BE2和BE3阻碍,这抑制了UDG。U:G错配也通过错配修复(MMR)处理,所述错配修复优先修复错配的切口链。BE3对未编辑的含有G的链产生切口,有利于U:G错配至期望的U:A或T:A结果的解析。图29B显示了如以下实施例的材料和方法中所述处理的HEK293T细胞。在指示的靶位置处具有T的总DNA测序读取的百分比显示用BE1,BE2或BE3的处理,或用具有供体HDR模板的wt Cas9的处理。图29C显示了图29B中的处理之后插入/缺失形成的频率。在图34中列出值。对于图29B和29C,值和误差棒反映在不同天数进行的三个独立的生物学重复的平均值和S.D.。

[0075] 图30A至30B显示了哺乳动物细胞中两种与疾病相关的突变的BE3介导的校正。突变的右侧显示了原间隔物的序列,PAM和靶碱基为红色,带有下标的数字指示其在原间隔物中的位置。每个序列下方是具有相应碱基的总测序读段的百分比。如材料和方法中所述处理细胞。图30A显示在总读段的74.9%中,BE3在小鼠星形胶质细胞中将与阿尔茨海默氏病相关的APOE4等位基因转换为APOE3r。两个附近的C也被转换为T,但所得蛋白的预测序列没有变化。用wt Cas9和供体ssDNA相同地处理这些细胞导致仅0.3%的校正,具有26.1%的插入/缺失形成。该图描绘了SEQ ID NO:671和627。图30B显示了在经核转染的人乳腺癌细胞的7.6%中,BE3在0.7%插入/缺失形成的情况下校正与癌症相关的p53 Y163C突变。用wt Cas9和供体ssDNA相同处理这些细胞导致无突变校正,具有6.1%插入/缺失形成。该图描绘

了SEQ ID NO:672和629。

[0076] 图31显示了在HEK293位点2脱靶处BE1,BE2和BE3的活性。使用Lipofectamine 2000,用表达BE1,BE2或BE3和与HEK293位点2序列匹配的sgRNA的质粒转染HEK293T细胞。转染后三天,提取基因组DNA,通过PCR扩增,并通过如Joung和同事先前使用GUIDE-seq方法(63)以及Adli和同事使用染色质免疫沉淀高通量测序(ChIP-seq)实验(18)确定的HEK293位点2sgRNA的中靶基因座,加上所有已知Cas9和dCas9脱靶基因座处的高通量DNA测序进行分析。显示了中靶和脱靶的原间隔物和原间隔物邻近基序(PAM)的序列。对于BE1,BE2和BE3,显示了细胞C至T的转化百分比,定义为原间隔物中原始C的每个位置处具有T的总DNA测序读段的百分比。最右边显示报告的测序读段的总数,以及对每个序列报告的ChIP-seq信号强度。该图从上至下分别描绘了SEQ ID NO:129、654、655和673至677。

[0077] 图32显示了在HEK293位点3脱靶处BE1,BE2和BE3的活性。使用Lipofectamine 2000,用表达BE1,BE2或BE3和与HEK293位点3序列匹配的sgRNA的质粒转染HEK293T细胞。转染后三天,提取基因组DNA,通过PCR扩增,并通过如Joung和同事使用GUIDE-seq方法[54],以及使用染色质免疫沉淀高通量测序(ChIP-seq)实验[61]先前确定的HEK293位点3sgRNA的中靶基因座,加上所有已知Cas9脱靶基因座和前5种已知dCas9脱靶基因座处的高通量DNA测序进行分析。显示了中靶和脱靶的原间隔物和原间隔物邻近基序(PAM)的序列。对于BE1,BE2和BE3,显示了细胞C至T的转化百分比,定义为原间隔物中原始C的每个位置处具有T的总DNA测序读段的百分比。最右边显示报告的测序读段的总数,以及对每个序列报告的ChIP-seq信号强度。该图从上至下分别描绘了SEQ ID NO:130、656至660和678-682。

[0078] 图33显示了显示了在HEK293位点4脱靶处BE1,BE2和BE3的活性。使用Lipofectamine 2000,用表达BE1,BE2或BE3和与HEK293位点4序列匹配的sgRNA的质粒转染HEK293T细胞。转染后三天,提取基因组DNA,通过PCR扩增,并如使用GUIDE-seq方法[54],以及使用染色质免疫沉淀高通量测序(ChIP-seq)实验[61]先前确定的HEK293位点4sgRNA的中靶基因座,加上前10种已知Cas9脱靶基因座和前5种已知dCas9脱靶基因座处的高通量DNA测序进行分析。显示了中靶和脱靶原间隔物和原间隔物邻近基序(PAM)的序列。对于BE1,BE2和BE3,显示了细胞C至T的转化百分比,定义为原间隔物中原始C的每个位置处具有T的总DNA测序读段的百分比。最右边显示报告的测序读段的总数,以及对每个序列报告的ChIP-seq信号强度。该图从上至下分别描绘了SEQ ID NO:131、661至670、683和684。

[0079] 图34显示了在小鼠星形胶质细胞中BE3介导的阿尔茨海默氏病相关的APOE4等位基因至APOE3r的校正后非原间隔物碱基的突变率。以每个碱基相对于原间隔物的位置显示了来自图30A和图34B的原间隔物任一侧的50个碱基的DNA序列。在PAM的情况中,PAM远侧的原间隔物侧用正数字指明,而包含PAM的侧用负数字指明。在每个序列下方是未处理的细胞,用BE3和靶向APOE4 C158R突变的sgRNA处理的细胞,或用BE3和靶向VEGFA基因座的sgRNA处理的细胞的具有相应碱基的总DNA测序读段的百分比。两种经BE3处理的样品均未导致突变率高于未处理对照的突变率。该图从上至下分别描绘了SEQ ID NO:685至688。

[0080] 图35显示了在HCC1954人细胞中与癌症相关的p53 Y163C突变的BE3介导的校正后非原间隔物碱基的突变率。以每个碱基相对于原间隔物的位置显示了来自图30B和图39B的原间隔物任一侧50个碱基的DNA序列。PAM远侧的原间隔物侧用正数字指明,而包含PAM的侧用负数字与PAM指明。在每个序列下方是未处理的细胞,用BE3和靶向TP53 Y163C突变的

sgRNA处理的细胞，或用BE3和靶向VEGFA基因座的sgRNA处理的细胞的具有相应碱基的总测序读段的百分比。两种经BE3处理的样品均未导致突变率高于未处理对照的突变率。该图从上至下分别描绘了SEQ ID NO:689至692。

[0081] 图36A至36F显示了脱氨酶、接头长度和接头组成对碱基编辑的影响。图36A显示了基于凝胶的脱氨酶测定法，显示了rAPOBEC1,pmCDA1,hAID,hAPOBEC3G,rAPOBEC1-GGS-dCas9,rAPOBEC1-$(GGS)_3$(SEQ ID NO:610)-dCas9和dCas9-$(GGS)_3$(SEQ ID NO:610)-rAPOBEC1对ssDNA的活性。将酶在哺乳动物细胞裂解物衍生的体外转录-翻译系统中表达，并与1.8μM缀合有染料的ssDNA和USER酶(尿嘧啶DNA糖基化酶和内切核酸酶VIII)在37℃温育2小时。将得到的DNA在变性聚丙烯酰胺凝胶上分离并成像。阳性对照是具有与靶C在同一位置处合成掺入的U的序列。图36B显示了图36C至36F中使用的表达和纯化的蛋白质的考马斯染色变性PAGE凝胶。图36C至36F显示了基于凝胶的脱氨酶测定法，显示了具有GGS(图36C)，$(GGS)_3$(SEQ IDNO:610)(图36D)，XTEN(图36E)，或$(GGS)_7$(SEQ ID NO:610)(图36F)的脱氨酶-Cas9接头的碱基编辑器的脱氨基窗。将与sgRNA复合的1.85μM脱氨酶-dCas9融合蛋白与125nM dsDNA底物在37℃温育2小时后，分离缀合有染料的DNA，并在37℃与USER酶一起温育1小时以在任何尿嘧啶位点处切割DNA主链。将得到的DNA在变性聚丙烯酰胺凝胶上解析，并对缀合有染料的链成像。每个泳道根据靶C在原间隔物内的位置进行编号，若没有靶C，则使用-编号。8U是具有第8位处合成掺入的U的阳性对照序列。

[0082] 图37A至37C显示了在哺乳动物细胞中，BE1碱基的编辑效率显著降低。图37A碱基编辑器靶向的六个哺乳动物细胞基因组基因座的原间隔物和PAM序列。靶C用红色指示，带有下标的数字对应于它们在原间隔物内的位置。图37B显示将具有与六个不同基因组位点匹配的序列的合成80聚体与BE1一起温育，然后通过HTS分析碱基编辑。对于每个位点，原间隔物的序列在位点名称的右侧与PAM一起指示。每个序列下方是具有相应碱基的总DNA测序读段的百分比。若体外转化效率>10％，则我们认为靶C是“可编辑的”。注意，由于非靶定链不受BE1影响，因此最大产率为DNA测序总读段的50％。显示来自单个实验的值。图37C显示用表达BE1和合适的sgRNA的质粒转染HEK293T细胞。转染后三天，提取基因组DNA并通过六个基因座处的高通量DNA测序进行分析。在所有六个基因组基因座处对于BE1显示了细胞C至T转化百分比，定义为在指示的靶位置处具有T的总DNA测序读段的百分比。来自HEK293T细胞的所有数据的值和误差棒反映了在不同天数进行的三个独立生物学重复的平均值和标准差。图37A从上至下分别描绘了SEQ ID NO:127至132。图37B从上至下分别描绘了SEQ ID NO:127至132。

[0083] 图38显示了碱基编辑在多个细胞分裂中持续存在。对于在对细胞进行传代之前和之后的HEK293T细胞中HEK293位点3和4显示了BE2和BE3得到的细胞C至T转化百分比。将HEK293T细胞用表达BE2或BE3和靶向HEK293位点3或4的sgRNA的质粒进行核转染。核转染后三天，收获细胞并分成两半。一半进行HTS分析，并且允许另一半增殖约5次细胞分裂，然后收获并进行HTS分析。值和误差棒反映了至少两个生物学实验的平均值和标准差。

[0084] 图39A至39C显示了非靶C/G突变率。这里显示的是在测试的6个中靶和34个脱靶基因座周围的2,500个独特的胞嘧啶和鸟嘌呤处的C至T和G至A突变率，代表从约$1.8x10^6$个细胞衍生的总共14,700,000个序列读段。图39A和39B显示了通过BE1、BE2和BE3得到的细胞非靶C至T和G至A转化百分比针对它们相对于所有2,500个胞嘧啶/鸟嘌呤的原间隔物的位置

个别绘图。PAM远端的原间隔物侧用正数字指明，而包含PAM的侧用负数字指明。图39C显示了由BE1，BE2和BE3得到的平均非靶细胞C至T和G至A的转化百分比，以及最高和最低的个别转化百分比得到显示。

[0085] 图40A至40B显示了哺乳动物细胞中两个与疾病相关的突变的BE3介导的校正的其他数据集。对于每个位点，原间隔物的序列在突变名称的右边指示，PAM和造成突变的碱基以红色粗体指示，带下标的数字对应于其在原间隔物中的位置。显示了每个与疾病相关的等位基因上方的氨基酸序列，以及碱基编辑后的校正氨基酸序列。每个序列下方是具有相应碱基的总测序读段的百分比。用编码BE3和合适的sgRNA的质粒对细胞进行核转染。核转染后两天，从经核转染的细胞中提取基因组DNA，并通过HTS分析以评估致病性突变校正。图40A显示仅在用正确的sgRNA处理时在总读段的58.3％中BE3在小鼠星形胶质细胞中将与阿尔茨海默氏病相关的APOE4等位基因转换为APOE3r。两个附近的C也转换为T，但不改变所得蛋白质的预测序列。用wtCas9和供体ssDNA相同处理这些细胞产生0.2％的校正，具有26.7％的插入/缺失形成。图40B显示了仅当用正确的sgRNA处理时在经核转染的人乳腺癌细胞的3.3％中BE3校正与癌症相关的p53 Y163C突变。用wt Cas9和供体ssDNA相同处理这些细胞以8.0％插入/缺失形成导致检测不到的突变校正。图40A至40B描绘了SEQ ID NO:671、627、672和629。

[0086] 图41显示了示例性的基于USER(尿嘧啶特异性切除试剂)酶的测定法的示意图，该测定法可用于测试各种脱氨基酶对单链DNA(ssDNA)底物的活性。

[0087] 图42是pmCDA-nCas9-UGI-NLS构建体及其在HeK-3位点处相对于碱基编辑器(rAPOBEC1)和阴性对照(未处理)的活性的示意图。该图描绘了SEQ ID NO:693。

[0088] 图43是pmCDA1-XTEN-nCas9-UGI-NLS构建体及其在HeK-3位点处相对于碱基编辑器(rAPOBEC1)和阴性对照(未处理)的活性的示意图。该图描述了SEQ ID NO:694。

[0089] 图44显示了使用胞苷脱氨酶(CDA)或APOBEC得到的具有转化为T的靶C的总测序读段的百分比。

[0090] 图45显示了使用脱氨酶(CDA)或APOBEC得到的具有转化为A的靶C的总测序读段的百分比。

[0091] 图46显示了使用脱氨酶(CDA)或APOBEC得到的具有转化为G的靶C的总测序读段的百分比。

[0092] 图47是huAPOBEC3G-XTEN-nCas9-UGI-NLS构建体及其在HeK-2位点处相对于突变形式(huAPOBEC3G*(D316R_D317R)-XTEN-nCas9-UGI-NLS，碱基编辑器(rAPOBEC1)和阴性对照(未处理)的活性，该图描述了SEQ ID NO:695。

[0093] 图48显示了在实施例7的选择测定法中使用的LacZ构建体的示意图。

[0094] 图49显示了来自不同质粒和构建体的回复数据。

[0095] 图50显示了lacZ回复的验证和回复克隆的纯化。

[0096] 图51是描述实施例7中使用的脱氨基选择质粒的示意图。

[0097] 图52显示了氯霉素回复测定法(pmCDA1融合)的结果。

[0098] 图53A至53B证明了两种构建体的DNA校正诱导。

[0099] 图54显示了氯霉素回复测定法(huAPOBEC3G融合)的结果。

[0100] 图55显示了在EMX1脱靶处BE3和HF-BE3的活性。从上到下的序列对应于SEQ ID

NO:127和637-645。

[0101] 图56显示了BE3和HF-BE3的中靶碱基编辑效率。

[0102] 图57是表明突变以不同程度影响胞苷脱氨基的图。各自略微损害催化的突变组合允许一个位置处相对于其他位置的选择性脱氨基。FANCF位点是GGAATC$_6$C$_7$C$_8$TTC$_{11}$TGCAGCACCTGG(SEQ ID NO:128)。

[0103] 图58是描绘下一代碱基编辑器的示意图。

[0104] 图59是显示从Cas9变体制备的新碱基编辑器的示意图。

[0105] 图60显示了不同NGA PAM位点的碱基编辑百分比。

[0106] 图61显示了使用NGCG PAM EMX(VRER BE3)和C$_1$TC$_3$C$_4$C$_5$ATC$_8$AC$_{10}$ATCAACCGGT(SEQ ID NO:696)间隔物的胞苷的碱基编辑百分比。

[0107] 图62显示了源自不同NNGRRT PAM位点的碱基编辑百分比。

[0108] 图63显示了源自不同NNHRRT PAM位点的碱基编辑百分比。

[0109] 图64A至64C显示了使用Cpf1 BE2源自不同TTTN PAM位点的碱基编辑百分比。使用的间隔物为:TTTCCTC$_3$C$_4$C$_5$C$_6$C$_7$C$_8$C$_9$AC$_{11}$AGGTAGAACAT(图64A,SEQ ID NO:697),TTTCC$_1$C$_2$TC$_4$TGTC$_8$C$_9$AC$_{11}$ACCCTCATCCTG(Figure 64B,SEQ ID NO:698)和TTTCC$_1$C$_2$C$_3$AGTC$_7$C$_8$TC$_{10}$C$_{11}$AC$_{13}$AC$_{15}$C$_{16}$C$_{17}$TGAAAC(图64C,SEQ ID NO:699)。

[0110] 图65是描绘如通过胞苷脱氨酶点诱变的动力学调节实现的选择性脱氨基的示意图。

[0111] 图66是显示各种突变对在间隔物中具有多个胞苷的细胞培养中探测的脱氨基窗的影响的图。使用的间隔物是:TGC$_3$C$_4$C$_5$C$_6$TC$_8$C$_9$C$_{10}$TC$_{12}$C$_{13}$C$_{14}$TGGCCC(SEQ ID NO:700)。

[0112] 图67是显示各种突变对在间隔物中具有多个胞苷的细胞培养中探测的脱氨基窗的影响的图。使用的间隔物是:AGAGC$_5$C$_6$C$_7$C$_8$C$_9$C$_{10}$C$_{11}$TC$_{13}$AAAGAGA(SEQ ID NO:701)。

[0113] 图68是显示各种突变对具有有限数目的胞苷的FANCF位点的影响的图。使用的间隔物是:GGAATC$_6$C$_7$C$_8$TTC$_{11}$TGCAGCACCTGG(SEQ ID NO:128)。注意,三重突变体(W90Y,R126E,R132E)优先编辑第六位的胞苷。

[0114] 图69是显示各种突变对具有有限数目的胞苷的HEK3位点的影响的图。使用的间隔物是:GGCC$_4$C$_5$AGACTGAGCACGTGATGG(SEQ ID NO:702)。注意,双重突变体和三重突变体相对于第四位的胞苷优先编辑在第五位的胞苷。

[0115] 图70是显示各种突变对具有有限数目的胞苷的EMX1位点的影响的图。使用的间隔物是:GAGTC$_5$C$_6$GAGCAGAAGAAGAAGGG(SEQ ID NO:703)。注意,三重突变体仅编辑在第五位而不是第六位的胞苷。

[0116] 图71是显示各种突变对具有有限数目的胞苷的HEK2位点的影响的图。使用的间隔物是:GAAC$_4$AC$_6$AAAGCATAGACTGCGGG(SEQ ID NO:704)。

[0117] 图72显示了永生化星形胶质细胞中BE3和包含突变W90Y R132E的BE3的中靶碱基编辑效率。

[0118] 图73描绘了三种Cpf1融合构建体的示意图。

[0119] 图74显示了BE3和HF-BE3(EMX1,FANCF和RNF2)的质粒递送的比较。

[0120] 图75显示了BE3和HF-BE3(HEK3和HEK 4)的质粒递送的比较。

[0121] 图76显示了所有10个位点处的EMX-1脱靶编辑。该图描述了SEQ ID NO:127和637-

645。

[0122] 图77显示了使用GAGTCCGAGCAGAAGAAGAAG(SEQ ID NO:705)间隔物对HEK细胞的脱氨酶蛋白脂转染。检查了EMX-1中靶和EMX-1脱靶位点2。

[0123] 图78显示了使用GGAATCCCTTCTGCAGCACCTGG(SEQ ID NO:706)间隔物对HEK细胞的脱氨酶蛋白脂转染。检查了FANCF中靶和FANCF脱靶位点1。

[0124] 图79显示了使用GGCCCAGACTGAGCACGTGA(SEQ ID NO:707)间隔物对HEK细胞的脱氨酶蛋白脂转染。检查了HEK-3中靶位点。

[0125] 图80显示了使用GGCACTGCGGCTGGAGGTGGGGG(SEQ ID NO:708)间隔物对HEK细胞的脱氨酶蛋白脂转染。HEK-4中靶、脱靶位点1、位点3和位点4。

[0126] 图81显示了sgHR_13(GTCAGGTCGAGGGTTCTGTC(SEQ ID NO:709)间隔物;C8靶物:G51至终止),sgHR_14(GGGCCGCAGTATCCTCACTC(SEQ ID NO:710)间隔物;C7靶物;C7靶物:Q68至终止)和sgHR_15(CCGCCAGTCCCAGTACGGGA(SEQ ID NO:711)间隔物;C10和C11是靶物:W239或W237至终止)的sgRNA活性的体外测定法结果。

[0127] 图82显示了sgHR_17(CAACCACTGCTCAAAGATGC(SEQ ID NO:712)间隔物;C4和C5是靶物:W410至终止)和sgHR_16(CTTCCAGGATGAGAACACAG(SEQ ID NO:713)间隔物,C4和C5是靶物:W273至终止)的体外测定法结果。

[0128] 图83显示了斑马鱼胚胎中与sgHR_13复合的BE3蛋白的直接注射。

[0129] 图84显示了斑马鱼胚胎中与sgHR_16复合的BE3蛋白的直接注射。

[0130] 图85显示了斑马鱼胚胎中与sgHR_17复合的BE3蛋白的直接注射。

[0131] 图86显示了可以使用碱基编辑器产生的示例性核酸变化,所述碱基编辑器能够产生胞嘧啶至胸腺嘧啶变化。

[0132] 图87显示了载脂蛋白E(APOE)同种型的示例,证明了可以如何使用碱基编辑器(例如,BE3)将一个APOE同种型(例如,APOE4)编辑为与阿尔茨海默氏病风险降低相关的另一种APOE同种型(例如,APOE3r)。

[0133] 图88显示了小鼠星形胶质细胞中APOE4至APOE3r的碱基编辑。该图描绘了SEQ ID No:671和627。

[0134] 图89显示了PRNP的碱基编辑,以引起精氨酸残基37处的蛋白质早期截短。该图描绘了SEQ ID No:577和714。

[0135] 图90显示了敲除UDG(UGI抑制的)显著改善C至T碱基编辑的效率的清洁度。

[0136] 图91显示了使用具有切口酶但没有UGI的碱基编辑器导致结果的混合,具有很高的插入/缺失率。

[0137] 图92A至92G显示了SaBE3,SaKKH-BE3,VQR-BE3,EQR-BE3和VRER-BE3介导人细胞中含有非NGG PAM的靶位点处的有效碱基编辑。图92A显示了使用酿脓链球菌和金黄色葡萄球菌Cas9的碱基编辑器构造。图92B显示了具有交替或宽松的PAM要求的最近表征的Cas9变体。图92C和92D显示了用如实施例12中所述的碱基编辑器变体处理的HEK293T细胞。显示了在指定的靶位置处具有转化为T的C的总DNA测序读段的百分比(不富集转染细胞)。X轴下方显示了每个测试靶物的PAM序列。图显示了具有NNGRRT PAM的基因组基因座处SaBE3和SaKKH-BE3的结果(图92C),具有NNNRRT PAM的基因组基因座处SaBE3和SaKKH-BE3的结果(图92D),具有NGAG PAM(图92E),以及具有NGAH PAM(图92F)的基因组基因座处的VQR-BE3

和EQR-BE3 PAM的结果，和在具有NGCG PAM的基因组基因座处VRER-BE3的结果(图92G)。值和误差棒反映至少两个生物学重复的平均值和标准差。

[0138] 图93A至93C证明在胞苷脱氨酶域中具有突变的碱基编辑器显示出窄的编辑窗。图93A至93C显示了用表达突变体碱基编辑器和合适的sgRNA的质粒转染的HEK293T细胞。转染后三天，提取基因组DNA并通过指定基因座处的高通量DNA测序进行分析。对于EMX1位点(SEQ ID NO:721)，HEK293位点3(SEQ ID NO:719)，FANCF位点(SEQ ID NO:722)，HEK293位点2(SEQ ID NO:720)，位点A(SEQ ID NO:715)和位点B(SEQ ID NO:718)基因座，显示了在指示的靶位置处具有变为T的C的总DNA测序读段的百分比(不富集转染细胞)。图93A显示了某些胞苷脱氨酶突变，其使碱基编辑窗变窄。关于其他突变的表征，参见图98。图93B显示了影响基因组基因座上的编辑窗宽度的胞苷脱氨酶突变的影响。组合有益突变对缩窄编辑窗具有累加效应。图93C显示了YE1-BE3，YE2-BE3，EE-BE3和YEE-BE3实现碱基编辑的产物分布，与BE3相比主要产生单修饰产物。值和误差棒反映至少两个生物学重复的平均值和标准差。

[0139] 图94A和94B显示了来自ClinVar的遗传变体，其原则上可以由这项工作中开发的碱基编辑器进行校正。对人遗传变异及其对应表型的NCBI ClinVar数据库搜索在理论上可以通过碱基编辑校正的遗传疾病。图94A证明了通过用具有改变的PAM特异性的碱基编辑器在ClinVar数据库中所有致病性T→C突变中碱基编辑靶向范围的改善。白色分数表示基于BE3或BE3以及在此工作中开发的五种经修饰的PAM碱基编辑器的PAM要求，可接近的致病性T→C突变的比例。图94B显示了通过使用具有变窄的活性窗的碱基编辑器，ClinVar数据库中所有致病性T→C突变中碱基编辑靶向范围的改善。如图93A至93C所示，假设BE3以相当的效率编辑位置4-8中的C。假设YEE-BE3在其活性窗内以C5>C6>C7>其他偏爱进行编辑。白色分数表示可以在不对其他C进行相当编辑的情况下由BE3编辑(左)或者在不对其他C进行相当编辑的情况下可以由BE3或YEE-BE3编辑(右)的致病性T→C突变比例。

[0140] 图95A至图95B显示了截短的引导RNA对碱基编辑窗宽度的影响。用表达BE3和不同5’截短长度的sgRNA的质粒转染HEK293T细胞。如实施例中所述分析经处理的细胞。图95A显示了原间隔物和PAM序列(顶部，SEQ ID NO:715)和细胞C至T的转化百分比，其定义为在EMX1基因组基因座内的位点处在指示的靶位置处具有T的总DNA测序读段的百分比。在此位点处，通过使用17nt截短的gRNA改变碱基编辑窗。图95B显示了原间隔物和PAM序列(顶部，SEQ ID NOs:715和716)以及细胞C至T的转化百分比，定义为在HEK位置3和位点4基因组基因座内的位点处在指示的靶位置处具有T的总DNA测序读段的百分比。在这些位点处，未观察到碱基编辑窗的变化，但是注意到随着sgRNA被截断，所有底物碱基的编辑效率呈线性下降。

[0141] 图96显示了APOBEC1-Cas9接头长度对碱基编辑窗宽度的影响。用表达具有XTEN，GGS，$(GGS)_3$(SEQ ID NO:610)，$(GGS)_5$(SEQ ID NO:610)，或$(GGS)_7$(SEQ ID NO:610)的rAPOBEC1-Cas9接头的碱基编辑器和sgRNA的质粒转染HEK293T细胞。如实施例中所述分析经处理的细胞。对于具有不同接头的各种碱基编辑器，显示了细胞C至T的转化百分比，定义为在指示的靶位置处具有T的总DNA测序读段的百分比。

[0142] 图97A至图97C显示了rAPOBEC突变对碱基编辑窗宽度的影响。图97C显示用表达靶向位点A或位点B的sgRNA和指示的BE3点突变体的质粒转染的HEK293T细胞。如实施例中所

述分析经处理的细胞。显示原间隔物中以及原间隔物的三个碱基对内的所有C，并显示细胞C至T的转化百分比。对于所有测试的突变体，显示“编辑窗宽度”，定义为在其内部编辑效率超出半最大值的计算的核苷酸数目。

[0143] 图98显示了APOBEC1突变对哺乳动物细胞中碱基编辑的产物分布的影响。用表达BE3或其突变体和合适的sgRNA的质粒转染HEK293T细胞。如实施例中所述分析经处理的细胞。显示了(左)细胞C至T的转化百分比，定义为在指示的靶位置处具有T的总DNA测序读段的百分比。右侧显示了含有C至T转化的总测序读段的百分比。BE3点突变体没有显著影响HEK位点4(仅一个靶胞苷的位点)处的碱基编辑效率。

[0144] 图99显示了BE3和HF-BE3中的中靶编辑血浆递送的比较。

[0145] 图100显示了BE3的蛋白质和血浆递送中的中靶编辑的比较。

[0146] 图101显示了HF-BE3的蛋白质和血浆中的中靶编辑的比较。

[0147] 图102显示脂转染和安装HF突变两者减少脱靶脱氨基事件。菱形指示未检测到脱靶，并且特异性比设置为100。

[0148] 图103显示了具有原间隔物(NNNNTC$_2$TC$_4$TC$_6$TC$_8$TC$_{10}$TC$_{12}$TC$_{14}$TC$_{16}$TC$_{18}$TC$_{20}$NGG，SEQ ID NO:723)中的偶数位置处放置的C的合成底物上的体外C至T编辑。

[0149] 图104显示了具有原间隔物(NNNNTC$_2$TC$_4$TC$_6$TC$_8$TC$_{10}$TC$_{12}$TC$_{14}$TC$_{16}$TC$_{18}$TC$_{20}$NGG，SEQ ID NO:723)中的奇数位置处放置的C的合成底物上的体外C至T编辑。

[0150] 图105包括两个图，描绘了用质粒对蛋白质递送进行碱基编辑的特异性比。

[0151] 图106A至106B显示了在非NGG PAM位点上的BE3活性。用表达BE3和合适的sgRNA的质粒转染HEK293T细胞。如实施例中所述分析经处理的细胞。图106A显示了SaBE3或SaKKH-BE3可以有效地靶向BE3活性。BE3在NAG PAM上显示出低但明显的活性。此图描绘了SEQ ID NO:728和729。图106B显示了与VQR-BE3或VRER-BE3相比，BE3在具有NGA或NGCG PAM的位点处具有显著减少的编辑。该图描绘了SEQ ID NO:730和731。

[0152] 图107A至107B显示了APOBEC1突变对VQR-BE3和SaKKH-BE3的影响。用表达VQR-BE3，SaKKH-BE3或其突变体和合适的sgRNA的质粒转染HEK293T细胞。如以下实施例所述分析经处理的细胞。显示了细胞C至T的转化百分比，定义为在指示的靶位置处具有T的总DNA测序读段的百分比。图107A显示了可以将窗调节突变应用于VQR-BE3，以在NGA PAM可靶向的位点处进行选择性碱基编辑。此图描绘了SEQ ID NO:732和733。图107B显示了当应用于SaKKH-BE3时，突变引起碱基编辑效率的整体下降，而不在靶窗内赋予碱基选择性。此图描绘了SEQ ID NO:728和734。

[0153] 图108显示了核苷酸编辑的示意图。使用以下缩写：(MMR)-错配修复，(BE3切口酶)-指碱基编辑器3，其包含Cas9切口酶域，(UGI)-尿嘧啶糖基化酶抑制剂，(UDG)-尿嘧啶DNA糖基化酶，(APOBEC)-指APOBEC胞苷脱氨酶。

[0154] 图109显示了示例性碱基编辑构建体的示意图。显示了BE3，BE4-pmCDA1，BE4-hAID，BE4-3G，BE4-N，BE4-SSB，BE4-(GGS)$_3$，BE4-XTEN，BE4-32aa，BE4-2xUGI和BE4的碱基编辑构建体的结构排列。接头显示为灰色(XTEN，SGGS(SEQ ID NO:606)，(GGS)$_3$(SEQ ID NO:610)和32aa)。显示了脱氨酶(rAPOBEC1，pmCDA1，hAID和hAPOBEC3G)。显示了尿嘧啶DNA糖基化酶抑制剂(UGI)。单链DNA结合蛋白(SSB)以紫色显示。Cas9切口酶dCas9(A840H)以红色显示。图109还显示了以下靶序列：EMX1，FANCF，HEK2，HEK3，HEK4和RNF2。氨基酸序列从上至下

在SEQ ID NO:127-132中指示。PAM序列是最后三个核苷酸。靶胞嘧啶(C)编号并以红色指示。

[0155] 图110显示了指示的碱基编辑构建体(BE3,pmCDA1 hAID,hAPOBEC3G,BE4-N,BE4-SSB,BE4-$(GGS)_3$,BE-XTEN,BE4-32aa和BE4-2xUGI)对EMX1序列GAGTC$_5$CGAGCAGAAGAAGAAGGG(SEQ ID NO:127)的靶定胞嘧啶(C_5)的碱基编辑结果。对于每个碱基编辑构建体,在"C_5"下指示了突变的靶定胞嘧啶(C_5)的总百分比。对于每个碱基编辑构建体,在"插入/缺失"下指示插入/缺失的总百分比。对于饼图中的每个碱基编辑构建体,指示突变为腺嘌呤(A),鸟嘌呤(G)或胸腺嘧啶(T)的突变胞嘧啶的比例。

[0156] 图111显示了指示的碱基编辑构建体(BE3,pmCDA1 hAID,hAPOBEC3G,BE4-N,BE4-SSB,BE4-$(GGS)_3$,BE-XTEN,BE4-32aa和BE4-2xUGI)对FANCF序列GGAATCCC$_8$TTCTGCAGCACCTGG(SEQ ID NO:128)的靶定胞嘧啶(C_8)的碱基编辑结果。对于每个碱基编辑构建体,在"C_8"下指示了突变的靶定胞嘧啶(C_8)的总百分比。对于每个碱基编辑构建体,在"插入/缺失"下指示插入/缺失的总百分比。对于饼图中的每个碱基编辑构建体,指示突变为腺嘌呤(A),鸟嘌呤(G)或胸腺嘧啶(T)的突变胞嘧啶的比例。

[0157] 图112显示了指示的碱基编辑构建体(BE3,pmCDA1 hAID,hAPOBEC3G,BE4-N,BE4-SSB,BE4-$(GGS)_3$,BE-XTEN,BE4-32aa,和BE4-2xUGI)对HEK2序列GAACAC$_6$AAAGCATAGACTGCGGG(SEQ ID NO:129)的靶定胞嘧啶(C_6)的碱基编辑结果。对于每个碱基编辑构建体,在"C_6"下指示了突变的靶定胞嘧啶(C_6)的总百分比。对于每个碱基编辑构建体,在"插入/缺失"下指示插入/缺失的总百分比。对于饼图中的每个碱基编辑构建体,指示突变为腺嘌呤(A),鸟嘌呤(G)或胸腺嘧啶(T)的突变胞嘧啶的比例。

[0158] 图113显示了指示的碱基编辑构建体(BE3,pmCDA1 hAID,hAPOBEC3G,BE4-N,BE4-SSB,BE4-$(GGS)_3$,BE-XTEN,BE4-32aa,和BE4-2xUGI)对HEK3序列GGCCC$_5$AGACTGAGCACGTGATGG(SEQ ID NO:130)的靶定胞嘧啶(C_5)的碱基编辑结果。对于每个碱基编辑构建体,在"C_5"下指示了突变的靶定胞嘧啶(C_5)的总百分比。对于每个碱基编辑构建体,在"插入/缺失"下指示插入/缺失的总百分比。对于饼图中的每个碱基编辑构建体,指示突变为腺嘌呤(A),鸟嘌呤(G)或胸腺嘧啶(T)的突变胞嘧啶的比例。

[0159] 图114显示了指示的碱基编辑构建体(BE3,pmCDA1 hAID,hAPOBEC3G,BE4-N,BE4-SSB,BE4-$(GGS)_3$,BE-XTEN,BE4-32aa,和BE4-2xUGI)对HEK4序列GGCAC$_5$TGCGGCTGGAGGTCCGGG(SEQ ID NO:131)的靶定胞嘧啶(C_5)的碱基编辑结果。对于每个碱基编辑构建体,在"C_5"下指示了突变的靶定胞嘧啶(C_5)的总百分比。对于每个碱基编辑构建体,在"插入/缺失"下指示插入/缺失的总百分比。对于饼图中的每个碱基编辑构建体,指示突变为腺嘌呤(A),鸟嘌呤(G)或胸腺嘧啶(T)的突变胞嘧啶的比例。

[0160] 图115显示了指示的碱基编辑构建体(BE3,pmCDA1 hAID,hAPOBEC3G,BE4-N,BE4-SSB,BE4-$(GGS)_3$,BE-XTEN,BE4-32aa,和BE4-2xUGI)对RNF2序列GTCATC$_6$TTAGTCATTACCTGAGG(SEQ ID NO:132)的靶定胞嘧啶(C_6)的碱基编辑结果。对于每个碱基编辑构建体,在"C_6"下指示了突变的靶定胞嘧啶(C_5)的总百分比。对于每个碱基编辑构建体,在"插入/缺失"下指示插入/缺失的总百分比。对于饼图中的每个碱基编辑构建体,指示突变为腺嘌呤(A),鸟嘌呤(G)或胸腺嘧啶(T)的突变胞嘧啶的比例。

[0161] 图116显示了示例性的荧光标记的(Cy3标记的)DNA构建体,其用于测试对靶链产

生切口的Cpf1突变体。在1的DNA构建体中，非靶链(顶部链)和靶链(底部链)两者是荧光标记的。在2的DNA构建体中，非靶链(顶部链)是荧光标记的，而靶链(底部链)不是荧光标记的。在3的DNA构建体中，非靶链(顶部链)不是荧光标记的，而靶链(底部链)是荧光标记的。

[0162] 图117显示了数据，其证明了各种Cpf1构建体(例如，R836A，R1138A，野生型)在30分钟(30min)或大于2小时(2h+)的反应时间内切割图116中所示的DNA构建体的靶链和非靶链的能力。

[0163] 图118显示了数据，其证明了具有结构APOBEC-AsCpf1(R912A)-UGI的碱基编辑器能够编辑C残基(例如，靶序列FANCF1，FANCF2，HEK3-3和HEK3-4的)，具有从靶序列的第7个碱基至第11个碱基的窗。BG指示背景突变水平(未处理)。AsCpf1指示仅AsCpf1处理的(对照)，APOBEC-AsCpf1(R912A)-UGI指示包含优先剪切靶链的Cpf1的碱基编辑器，并且APOBEC-AsCpf1(R1225A)-UGI指示包含切割非靶链的Cpf1的自我挫败(self-defeating)碱基编辑器。FANCF1，FANCF2，HEK3-3和HEK3-4的靶序列如下：

[0164] FANCF1 GCGGATGTTCCAATCAGTACGCA(SEQ ID NO:724)

[0165] FANCF2 CGAGCTTCTGGCGGTCTCAAGCA(SEQ ID NO:725)

[0166] HEK3-3 TGCTTCTCCAGCCCTGGCCTGG(SEQ ID NO:726)

[0167] HEK3-4 AGACTGAGCACGTGATGGCAGAG(SEQ ID NO:727)

[0168] 图119显示了包含Cpf1蛋白(例如，AsCpf1或LbCpf1)的碱基编辑器的示意图。对不同接头序列(例如XTEN，GGS，$(GGS)_3$(SEQ ID NO:610)，$(GGS)_5$(SEQ ID NO:610)，和$(GGS)_7$(SEQ ID NO:610))测试了标记为“接头”的部分，其结果在图120中显示。

[0169] 图120显示了数据，其证明了图119中所示的构建体编辑HEK3位点TGCTTCTC_8AGCCCTGGCCTGG(SEQ ID NO:592)的C_8残基的能力。测试了将APOBEC域连接到Cpf1域(例如LbCpf1(R836A)或AsCpf1(R912A))的不同接头序列。测试的示例性接头包括XTEN，GGS，$(GGS)_3$(SEQ ID NO:610)，$(GGS)_5$(SEQ ID NO:610)，和$(GGS)_7$(SEQ ID NO:610)。

[0170] 图121显示了数据，其证明了图119所示的具有LbCpf1域的构建体编辑HEK3 TGCTTCTC_8C_9AGCCCTGGCCTGG(SEQ ID NO:592)的C_8和C_9残基的能力。测试了来自整合生物信息学中心VU(Centre of Integrative Bioinformatics VU)维护的数据库的不同接头序列，它们将APOBEC域连接到LbCpf1域。测试的示例性接头包括1au7，1c1k，1c20，1ee8，1flz，1ign，1jmc，1sfe，2ezx和2reb。

[0171] 图122显示了AsCpf1的结构的示意图，其中指示了N和C末端。

[0172] 图123显示了SpCas9的结构的示意图，其中指示了N和C末端。

[0173] 图124显示了AsCpf1的示意图，其中红色圆圈指示编辑窗所在的预测区域。正方形指示可能阻碍APOBEC活性的螺旋区域。

[0174] 图125A和125B显示了高保真性碱基编辑器(HF-BE3)的工程化和体外表征。图125A显示了HF-BE3的示意图。显示了引入BE3中以产生HF-BE3的点突变。表示法使用PDB结构4UN3(Cas9)，4ROV(胞苷脱氨酶)和1UGI(尿嘧啶DNA糖基化酶抑制剂)。图125B显示了含有“TC”重复原间隔物的合成底物的体外脱氨基。值和误差棒反映在不同天数进行的两次独立重复的平均值和范围。

[0175] 图126A至126C显示了碱基编辑器蛋白的纯化。图126A显示了用于碱基编辑器表达的最佳大肠杆菌菌株的选择。在18℃下IPTG诱导的蛋白质表达16小时后，分析粗细胞裂解

液的蛋白质含量。BL21 Star(DE3)(Thermo Fisher)细胞显示出BE3和HF-BE3两者的最有希望的表达后水平,并用于表达碱基编辑器。图126B显示了表达的碱基编辑器蛋白的纯化。将His6标签放置在碱基编辑器的C末端导致生成BE3和HF-BE3两者的截短产物(泳道1和2)。出乎意料的是,通过将His6标签放置在蛋白质的N末端除去此种截短产物(泳道3-6)。在OD600=0.7(泳道4-5)的细胞密度(晚于Cas9表达的最佳者(OD600=0.4)1)下诱导碱基编辑器的表达改善碱基编辑器蛋白的产率。使用手动HisPur树脂柱,然后进行阳离子交换FPLC(Akta)进行纯化。图126C显示了纯化的BE3和HF-BE3。使用加热和LDS使不同浓度的纯化BE3和HF-BE3变性,然后加载到聚丙烯酰胺凝胶上。蛋白质样品代表本研究中使用的蛋白质。图126A至126C中的凝胶是BOLT Bis-Tris Plus 4-12%聚丙烯酰胺(Thermo Fisher)。如方法中所述进行电泳和染色。

[0176] 图127A至127D显示了在人细胞中的高保真性碱基编辑器(HF-BE3)活性。图127A至127C显示了使用基因组DNA的高通量测序测定与BE3和HF-BE3质粒转染相关的中靶和脱靶编辑,所述基因组DNA来自用靶向非重复基因组基因座EMX1(图127A),FANCF(图127B)和HEK293位点3(图127C)的sgRNA处理的HEK293T细胞。与每个sgRNA相关的中靶和脱靶基因座由垂直线分开。图127D显示了与靶向VEGFA位点2的高度重复sgRNA相关的中靶和脱靶编辑。值和误差棒反映在不同天数进行的三个独立的生物学重复的平均值±S.D.。对于图127A至127C,星号指示基于经处理的样品与未处理的对照之间的比较的显著编辑。*p≤0.05,**p≤0.01和***p≤0.001(Student两尾t检验)。对于图127D,由于相对于对照,所有经处理的样品显示出显著的编辑,未显示星号。表16中列出了个别p值。

[0177] 图128A至128C显示了BE3蛋白或质粒的剂量对培养的人细胞中的中靶和脱靶碱基编辑效率的影响。图128A显示了四个基因组基因座中的每处的中靶编辑效率在每个sgRNA的活性窗中的所有经编辑的胞嘧啶间取平均。值和误差棒反映了在不同天数进行的三个独立生物学重复的平均值±S.E.M。图128B和128C显示了HEK293T细胞中源自BE3质粒滴定(图128B)或BE3蛋白滴定(图128C)的EMX1位点处的中靶和脱靶编辑。值和误差棒反映了在不同天数进行的三个独立的生物学重复的平均值±S.D.。

[0178] 图129A至129B显示了BE3和HF-BE3的质粒和蛋白质递送的中靶:脱靶碱基编辑频率比。通过将活性窗中特定胞嘧啶处的中靶编辑百分比除以所指示的脱靶基因座的相应胞嘧啶处的脱靶编辑百分比计算碱基编辑中靶:脱靶特异性比(参见方法)。当脱靶编辑低于检测阈值(测序读段的0.025%)时,我们将脱靶编辑设置为检测限(0.025%),然后将中靶编辑百分比除以此上限。在用◆表示的这些情况下,显示的特异性比代表下限。对非重复sgRNAFANCF,HEK 293位点3和FANCF(图129A)和高度重复sgRNA VEGFA位点2(图129B)显示了特异性比。值和误差棒反映在不同天数进行的三个独立的生物学重复的平均值±S.D.。

[0179] 图130A至图130D显示了碱基编辑器向培养的人细胞中的蛋白质递送。图130A至130D显示了与碱基编辑器的RNP传递相关的中靶和脱靶编辑,所述碱基编辑器与靶向EMX1(图130A),FANCF(图130B),HEK293位点3(图130C)和VEGFA位点2(图130D)的sgRNA复合。非重复sgRNA的所有测序基因座处检测不到脱靶碱基编辑。值和误差棒反映在不同天数进行的三个独立的生物学重复的平均值±S.D.。星号表示基于经处理的样品与未处理的对照之间的比较进行的明显编辑。*p≤0.05,**p≤0.01和***p≤0.001(Student两尾t检验)。

[0180] 图131A至131C显示了与基因组基因座处的碱基编辑相关的插入/缺失形成。图

131A显示了VEGFA位点2,EMX1,FANCF和HEK293位点3sgRNA在中靶位点处的插入/缺失频率。图131B显示了碱基编辑:插入/缺失形成的比率。菱形(◆)指示未检测到插入/缺失(在经处理的样品中以及未处理的对照中插入/缺失频率无显著差异)。图131C显示了在与图131A中询问的中靶位点相关的脱靶位点处观察到的插入/缺失。值和误差棒反映在不同天数进行的三个独立的生物学重复的平均值±S.D.。

[0181] 图132A至132D显示了使用BE3的RNP递送在斑马鱼胚胎中和活小鼠的内耳中的无DNA的体内碱基编辑。图132A显示了注射与指示的sgRNA复合的BE3后4天收获的斑马鱼中的中靶基因组编辑。值和误差棒反映三只注射斑马鱼和三只对照斑马鱼的平均值±s.d.。给对照注射与无关的sgRNA复合的BE3。图132B显示了示意图,其显示了包囊到阳离子脂质纳米颗粒中的BE3:sgRNA复合物的体内注射。图132C显示了在VEGFA位点2基因组基因座处的碱基编辑器窗中的胞嘧啶残基的碱基编辑。图132D显示了VEGFA位点2靶基因座的碱基编辑窗中的每个胞嘧啶处的中靶编辑。图132D(图132C和132D)显示了值和误差棒反映注射靶向VEGFA位点2的sgRNA的三只小鼠,未注射的三只小鼠和注射无关的sgRNA的一只小鼠的平均值±S.E.M.。

[0182] 图133A至133E显示了鼠NIH/3T3细胞中的中靶和脱靶碱基编辑。图133A显示了与“VEGFA位点2”sgRNA相关的中靶碱基编辑(对于序列,参见图132E)。阴性对照对应于用编码BE3但无sgRNA的质粒处理的细胞。值和误差棒反映在不同天数进行的三个独立的生物学重复的平均值±S.D.。图133B至图133E显示了使用此sgRNA的前四个预测的脱靶基因座(图132E中所示的序列)使用高通量DNA测序测量与该位点相关的脱靶编辑。图133B显示了脱靶2,图133C显示了脱靶1,图133D显示了脱靶3,图133E显示了脱靶4。值和误差棒反映了在不同天数进行的三个独立的生物学重复的平均值±S.D.。

[0183] 图134A至134B显示了来自体内编辑的鼠组织的脱靶碱基编辑和中靶插入/缺失分析。图134A显示了在与VEGFA位点2相关的脱靶基因座的碱基编辑窗中每个胞嘧啶绘制的编辑。图134B显示了在中靶碱基编辑器基因座处的插入/缺失率。值和误差棒反映了三只注射的小鼠和三只对照小鼠的平均值±S.E.M。

[0184] 图135A至135C显示了敲除UNG对碱基编辑产物纯度的影响。图135A显示了如实施例17的材料和方法中所述用BE3处理的HAP1(UNG^+)和HAP1UNG^-细胞。显示了经编辑的DNA测序读段(靶C得到突变的读段)中的产物分布。图135B显示了测试的基因组基因座的原间隔物和PAM序列,图135A中分析的靶C显示为红色。图135C显示了在HAP1细胞或HAP1 UNG^-细胞中用BE3处理后插入/缺失形成的频率。值和误差条反映在不同天数进行三个独立的生物学重复的平均值±S.D.。

[0185] 图136A至136D显示了多C碱基编辑对产物纯度的影响。图136A显示了未处理的、经BE3处理的和经AID-BE3处理的人HEK293T细胞的代表性高通量测序数据。原间隔物的序列显示在顶部,PAM和靶C为红色,带下标的数字指示它们在原间隔物内的位置。每个序列下方是具有相应碱基的总测序读段的百分比。清洁编辑为T而不是非T碱基的靶C的相对百分比,对于经AID-BE3处理的细胞(其编辑此基因座处的3个C)比对于经BE3处理的细胞(其仅编辑1个C)高得多。图136B显示了如实施例17的材料和方法中所述用BE3,CDA1-BE3和AID-BE3处理的HEK293T细胞。显示了在经编辑的DNA测序读段(靶C得到突变的读段)中的产物分布。图136C显示了研究的基因组基因座的原间隔物和PAM序列,在图136B中分析的靶C用红色显

示。图136D显示了在图136A中所示的处理之后插入/缺失形成的频率。值和误差条反映在不同天数进行三个独立的生物学重复的平均值±S.D.。

[0186] 图137A至137C显示了改变BE3的结构对C至T编辑效率和产物纯度的影响。图137A显示了研究的基因组基因座的原间隔物和PAM序列，图137C中的靶C以紫色和红色显示，并且图137B中的靶C以红色显示。图137B显示了如实施例17的材料和方法中所述用BE3，SSB-BE3，N-UGI-BE3和BE3-2xUGI处理的HEK293T细胞。对于BE3、N-UGI-BE3和BE3-2xUGI显示了经编辑的DNA测序读段(靶C得到突变的读段)中的产物分布。图137C显示了C至T的碱基编辑效率。值和误差棒反映在不同天数进行的三个独立的生物学重复的平均值±S.D.。

[0187] 图138A至图138D显示了BE3中的接头长度变化对C至T编辑效率和产物纯度的影响。图138A显示了BE3，BE3C，BE3D和BE3E的结构。图138B显示了所研究的基因组基因座的原间隔物和PAM序列，图138C中的靶C用紫色和红色显示，图138D中的靶C用红色显示。图138C显示了如实施例17的材料和方法中所述用BE3，BE3C，BE3D或BE3E处理的HEK293T细胞。显示了C至T碱基编辑效率。图138D显示了BE3，BE3C，BE3D和BE3E的经编辑的DNA测序读段(靶C得到突变的读段)中的产物分布。值和误差条反映在不同天数进行的三个独立的生物学重复的平均值±S.D.。

[0188] 图139A至139D显示了与BE3相比BE4增加碱基编辑效率和产物纯度。图139A显示了BE3，BE4和Target-AID的结构。图139B显示了研究的基因组基因座的原间隔物和PAM序列，图139C中的靶C用紫色和红色显示，图139D中的靶C用红色显示。图139C显示了如实施例17的材料和方法中所述用BE3，BE4或Target-AID处理的HEK293T细胞。显示了C至T碱基的编辑效率。图139D显示了BE3和BE4的经编辑的DNA测序读段(靶C得到突变的读段)中的产物分布。值和误差条反映在不同天数进行的三个独立的生物学重复的平均值±S.D.。

[0189] 图140A至图140C显示了CDA1-BE3和AID-BE3比BE3更有效编辑靶G后的C。图140A显示了研究的基因组基因座的原间隔物和PAM序列，BE3，CDA1-BE3和AID-BE3编辑的靶C用红色显示，仅CDA1-BE3和AID-BE3编辑的靶C(在G后)用紫色显示。图140B显示了如实施例17的材料和方法中所述用BE3，CDA1-BE3，AID-BE3或APOBEC3G-BE3处理的HEK293T细胞。显示了C至T碱基的编辑效率。图140C显示了来自HEK293T细胞的个别DNA测序读段，所述HEK293T细胞用靶向HEK2基因座的BE3，CDA1-BE3或AID-BE3处理并根据原间隔物的序列进行分箱并进行分析，揭示了具有CDA1-BE3和AID-BE3得到的清洁C至T编辑物的测序读段的>85％具有两个编辑为T的C(图140C)。

[0190] 图141A至图141C显示了在具有多个可编辑的C的位点中不均匀的编辑导致较低的产物纯度。图141A显示了研究的基因组基因座的原间隔物和PAM序列，图141C中的靶C以紫色和红色显示，并且图141B中的靶C以红色显示。图141B和141C显示了如实施例17的材料和方法中所述用BE3处理的HEK293T细胞。显示了在经编辑的DNA测序读段(靶C得到突变的读段)中的产物分布。当相同基因座内两个C的编辑效率不相等时，C至非T的编辑更为频繁。值和误差棒反映在不同天数进行的三个独立的生物学重复的平均值±S.D.。

[0191] 图142A至图142D显示了多个C的碱基编辑导致较高的碱基编辑产物纯度。图142A显示了研究的基因组基因座的原间隔物和PAM序列，图142B中研究的靶C用红色显示。图142B显示了如实施例17的材料和方法中所述用BE3或BE3B(其缺少UGI)处理的HEK293T细胞。显示了经编辑的DNA测序读段(靶C得到突变的读段)中的产物分布。图142C显示了根据

第6位的主要靶C的身份对来自用靶向HEK2基因座的BE3或BE3B(其缺乏UGI)处理的HEK293T细胞的HTS读段进行分箱。然后,针对第4位的次要靶C处的碱基身份分析所得的读段。当在所述读取中仅有单一编辑事件时C_6更可能被不正确编辑为非T。图142D显示了用靶向HEK4基因座(仅有单一可编辑C的位点)的BE3或BE3B处理的细胞中C_5处具有A,G和T的经编辑的读段的分布,例示了通过UNG启动的碱基切除修复处理单一G:U错配以给出产物的混合物。值和误差条反映在不同天数进行的三个独立的生物学重复的平均值±S.D.。

[0192] 图143显示了多个C的碱基编辑导致HEK3和RNF2基因座处的较高的碱基编辑产物纯度。根据主要靶C位置(红色)处碱基的身份分开来自用靶向HEK3和RNF2基因座的BE3或BE3B(无UGI)处理的HEK293T细胞的DNA测序读段。然后,对四组测序读段询问次要靶C位置(紫色)处的碱基身份。对于BE3,当主要靶C(红色)被不正确编辑为G时,次要靶C更可能保留C。相反,当主要靶C(红色)转化为T时,次要靶C在相同测序读段中更可能也编辑为T。值和误差棒反映在不同天数进行的三个独立的生物学重复的平均值±S.D.。

[0193] 图144A至144C显示BE4比BE3诱导更低的插入/缺失频率,并且Target-AID表现出与CDA1-BE3相似的产物纯度。图144A显示了如实施例17的材料和方法中所述用BE3,BE4或Target-AID处理的HEK293T细胞。显示了插入/缺失形成的频率(参见实施例17的材料和方法)。图144B显示了如实施例17的材料和方法中所述用CDA1-BE3或Target-AID处理的HEK293T细胞。显示了在经编辑的DNA测序读段(靶C得到突变的读段)中的产物分布。图144C显示了所研究的基因组基因座的原间隔物和PAM序列,图144B中所研究的靶C以红色显示。值和误差棒反映在不同天数进行的三个独立的生物学重复的平均值±S.D.。

[0194] 图145A至145C显示了与SaBE3相比,SaBE4表现出增加的碱基编辑产率和产物纯度。图145A显示了如实施例17的材料和方法中所述用SaBE3和SaBE4处理的HEK293T细胞。显示了在指示的靶位置处具有T的总DNA测序读段的百分比。图145B显示了所研究的基因组基因座的原间隔物和PAM序列,图145A中的靶C以紫色和红色显示,图145C中所研究的靶C以红色显示。图145C显示了经编辑的DNA测序读段(靶C得到突变的读段)中的产物分布。值和误差条反映在不同天数进行的三个独立的生物学重复的平均值±S.D.。

[0195] 图146显示了来自在EMX1基因座处用BE3,CDA1-BE3,AID-BE3或APOBEC3G-BE3的处理的碱基编辑结果。原间隔物的序列显示在顶部,PAM和靶碱基为红色,带有下标的数字指示它们在原间隔物中的位置。序列下方是具有相应碱基的总测序读段的百分比。如实施例17的材料和方法中所述处理细胞。显示的值来自一个代表性实验。

[0196] 图147显示了来自在FANCF基因座处用BE3,CDA1-BE3,AID-BE3或APOBEC3G-BE3的处理的碱基编辑结果。原间隔物的序列显示在顶部,PAM和靶碱基为红色,带下标的数字指示它们在原间隔物中的位置。序列下方是具有相应碱基的总测序读段的百分比。如实施例17的材料和方法中所述处理细胞。显示的值来自一个代表性实验。

[0197] 图148显示了来自在HEK2基因座处用BE3,CDA1-BE3,AID-BE3或APOBEC3G-BE3的处理的碱基编辑结果。原间隔物的序列显示在顶部,PAM和靶碱基为红色,带下标的数字指示它们在原间隔物中的位置。序列下方是具有相应碱基的总测序读段的百分比。如实施例17的材料和方法中所述处理细胞。显示的值来自一个代表性实验。

[0198] 图149显示了来自在HEK3基因座处用BE3,CDA1-BE3,AID-BE3或APOBEC3G-BE3的处理的碱基编辑结果。原间隔物的序列显示在顶部,PAM和靶碱基为红色,带下标的数字指示

它们在原间隔物中的位置。序列下方是具有相应碱基的总测序读段的百分比。如实施例17的材料和方法中所述处理细胞。显示的值来自一个代表性实验。

[0199] 图150显示了来自在HEK4基因座处用BE3,CDA1-BE3,AID-BE3或APOBEC3G-BE3的处理的碱基编辑结果。原间隔物的序列显示在顶部,PAM和靶碱基为红色,带下标的数字指示它们在原间隔物中的位置。序列下方是具有相应碱基的总测序读段的百分比。如实施例17的材料和方法中所述处理细胞。显示的值来自一个代表性实验。

[0200] 图151显示了来自在RNF2基因座处用BE3,CDA1-BE3,AID-BE3或APOBEC3G-BE3的处理的碱基编辑结果。原间隔物的序列显示在顶部,PAM和靶碱基为红色,带下标的数字指示它们在原间隔物中的位置。序列下方是具有相应碱基的总测序读段的百分比。如实施例17的材料和方法中所述处理细胞。显示的值来自一个代表性实验。

[0201] 图152显示了LBCpf1融合构建体的示意图。构建体10具有[Apobec]-[LbCpf1]-[UGI]-[UGI]的域排列;构建体11具有[Apobec]-[LbCpf1]-[UGI]的域排列;构建体12具有[UGI]-[Apobec]-[LbCpf1]的域排列;构造13具有[Apobec]-[UGI]-[LbCpf1]的域排列;构建体14具有[LbCpf1]-[UGI]-[Apobec]的域排列;构建体15具有[LbCpf1]-[Apobec]-[UGI]的域排列。对于每个构建体,使用了三种不同的LbCpf1蛋白(D/N/A,其指核酸酶死亡的LbCpf1(D);LbCpf1切口酶(N)和核酸酶活性LbCpf1(A))。

[0202] 图153显示了使用长度为19个核苷酸的引导,即EMX19:TACTTTGTCCTCCGGTTCT(SEQ ID NO:744),在EMX靶物TTTGTAC$_3$TTTGTC$_9$C$_{10}$TC12C$_{13}$GGTTC$_{18}$TG(SEQ ID NO:738)中6个C残基的C至T编辑的百分比。对图152中所示的构建体中的几种测试编辑。

[0203] 图154显示了使用长度为18个核苷酸的引导,即EMX18:TACTTTGTCCTCCGGTTC(SEQ ID NO:745),在EMX靶物TTTGTAC$_3$TTTGTC$_9$C$_{10}$TC12C$_{13}$GGTTC$_{18}$TG(SEQ ID NO:738)中6个C残基的C至T编辑的百分比。对图152中所示的构建体中的几种测试编辑。

[0204] 图155显示了使用长度为17个核苷酸的引导,即EMX17:TACTTTGTCCTCCGGTT(SEQ ID NO:746),在EMX靶物TTTGTAC$_3$TTTGTC$_9$C$_{10}$TC12C$_{13}$GGTTC$_{18}$TG(SEQ ID NO:738)中6个C残基的C至T编辑的百分比。对图152中所示的构建体中的几种测试编辑。

[0205] 图156显示了使用长度为23个核苷酸的引导,即Hek2_23:CAGCCCGCTGGCCCTGTAAAGGA(SEQ ID NO:747),在HEK2靶物TTTCC$_1$AGC$_4$C$_5$C$_6$GC$_8$TGGC$_{12}$C$_{13}$C$_{14}$TGTAAA(SEQ ID NO:739)中8个C残基的C至T编辑的百分比。对图152中所示的构建体中的几种测试编辑。

[0206] 图157显示了使用长度为20个核苷酸的引导,即Hek2_20:CAGCCCGCTGGCCCTGTAAA(SEQ ID NO:748),在HEK2靶物TTTCC$_1$AGC$_4$C$_5$C$_6$GC$_8$TGGC$_{12}$C$_{13}$C$_{14}$TGTAAA(SEQ ID NO:739)中8个C残基的C至T编辑的百分比。对图152中所示的构建体中的几种测试编辑。

[0207] 图158显示了使用长度为19个核苷酸的引导,即Hek2_19:CAGCCCGCTGGCCCTGTAA(SEQ ID NO:749),在HEK2靶物TTTCC$_1$AGC$_4$C$_5$C$_6$GC$_8$TGGC$_{12}$C$_{13}$C$_{14}$TGTAAA(SEQ ID NO:739)中8个C残基的C至T编辑的百分比。对图152中所示的构建体中的几种测试编辑。

[0208] 图159显示了使用长度为18个核苷酸的引导,即Hek2_18:CAGCCCGCTGGCCCTGTA(SEQ ID NO:750),在HEK2靶物TTTCC$_1$AGC$_4$C$_5$C$_6$GC$_8$TGGC$_{12}$C$_{13}$C$_{14}$TGTAAA(SEQ ID NO:739)中8个C残基的C至T编辑百分比。对图152中所示的构建体中的几种测试编辑。

[0209] 图160显示了图153中描绘的实验的编辑百分比值(基于插入/缺失计数进行调整

后)和插入/缺失百分比。

[0210] 图161显示了图154中描绘的实验的编辑百分比值(基于插入/缺失计数进行调整后)和插入/缺失百分比。

[0211] 图162显示了图155中描绘的实验的编辑百分比值(基于插入/缺失计数进行调整后)和插入/缺失百分比。

[0212] 图163显示了图156中描绘的实验的编辑百分比值(基于插入/缺失计数进行调整后)和插入/缺失百分比。

[0213] 图164显示了图157中描绘的实验的编辑百分比值(基于插入/缺失计数进行调整后)和插入/缺失百分比。

[0214] 图165显示了图158中描绘的实验的编辑百分比值(基于插入/缺失计数进行调整后)和插入/缺失百分比。

[0215] 图166显示了图159中描绘的实验的编辑百分比值(基于插入/缺失计数进行调整后)和插入/缺失百分比。

[0216] 定义

[0217] 如本文和权利要求书中所用,除非上下文另外明确指出,否则单数形式“一种”、“一个”和“该/所述”包括单数和复数提及。因此,例如,提及“试剂”包括单一试剂和多个此类试剂。

[0218] 术语“核酸可编程DNA结合蛋白”或“napDNAbp”是指与核酸(例如,DNA或RNA),例如引导核酸(例如,gRNA)缔合的蛋白质,所述核酸例如通过与靶核酸序列杂交将napDNAbp引导至特定核酸序列。例如,Cas9蛋白可以与引导RNA缔合,该引导RNA将Cas9蛋白引导至与引导RNA互补的特定DNA序列。在一些实施方案中,napDNAbp是2类微生物CRISPR-Cas效应物。在一些实施方案中,napDNAbp是Cas9域,例如,核酸酶活性Cas9,Cas9切口酶(nCas9)或无核酸酶活性Cas9(dCas9)。核酸可编程DNA结合蛋白的实例包括但不限于Cas9(例如,dCas9和nCas9),CasX,CasY,Cpf1,C2c1,C2c2,C2C3和Argonaute。然而,应当理解,核酸可编程DNA结合蛋白也包括结合RNA的核酸可编程蛋白。例如,napDNAbp可以与将napDNAbp引导至RNA的核酸缔合。其他核酸可编程DNA结合蛋白也在本公开内容的范围内,尽管它们可能未在本公开内容中具体描述。

[0219] 在一些实施方案中,napDNAby是“RNA可编程核酸酶”或“RNA引导的核酸酶”。该术语在本文中可互换使用,并且是指与一种或多种不作为切割靶物的RNA形成复合物(例如与之结合或缔合)的核酸酶。在一些实施方案中,当与RNA复合物时,RNA可编程核酸酶可以称为核酸酶:RNA复合物。通常,结合的RNA称为引导RNA(gRNA)。gRNA可以以两个或更多个RNA的复合物形式存在,或者以单一RNA分子形式存在。以单一RNA分子形式存在的gRNA可以称为单引导RNA(sgRNA),尽管“gRNA”也用于指以单个分子或两个或更多个分子的复合物形式存在的引导RNA。通常,以单一RNA种类存在的gRNA包含两个域:(1)与靶核酸共享同源性的域(即,指导Cas9复合物与靶物的结合);和(2)结合Cas9蛋白的域。在一些实施方案中,域(2)对应于称为tracrRNA的序列,并包含茎-环结构。在一些实施方案中,域(2)与如Jinek et al.,Science 337:816-821(2012)(其全部内容通过引用并入本文)中提供的tracrRNA相同或同源。gRNA的其他实例(例如,包括域2的那些)可以参见2013年9月6日提交的美国临时专利申请U.S.S.N.61/874,682,题为“Switchable Cas9 Nucleases And Uses Thereof”

和于2013年9月6日提交的美国临时专利申请U.S.S.N.61/874,746,题为“Delivery System For Functional Nucleases”,每篇的全部内容通过引用完整并入本文。在一些实施方案中,gRNA包含域(1)和(2)中的两个或更多个,并且可以称为“延伸的gRNA”。例如,延伸的gRNA将结合两个或更多个Cas9蛋白,并且在两个或更多个独特区域处结合靶核酸,如本文所述。gRNA包含与靶位点互补的核苷酸序列,其介导核酸酶/RNA复合物与所述靶位点的结合,提供核酸酶:RNA复合物的序列特异性。在一些实施方案中,RNA可编程核酸酶是(CRISPR相关系统)Cas9内切核酸酶,例如来自酿脓链球菌的Cas9(Csn1)(参见例如“Complete genome sequence of an M1 strain of Streptococcus pyogenes.”Ferretti J.J.,McShan W.M.,Ajdic D.J.,Savic D.J.,Savic G.,Lyon K.,Primeaux C.,Sezate S.,Suvorov A.N.,Kenton S.,Lai H.S.,Lin S.P.,Qian Y.,Jia H.G.,Najar F.Z.,Ren Q.,Zhu H.,Song L.,White J.,Yuan X.,Clifton S.W.,Roe B.A.,McLaughlin R.E.,Proc.Natl.Acad.Sci.U.S.A.98:4658-4663(2001);“CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.”Deltcheva E.,Chylinski K.,Sharma C.M.,Gonzales K.,Chao Y.,Pirzada Z.A.,Eckert M.R.,Vogel J.,Charpentier E.,Nature 471:602-607(2011);和“A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.”Jinek M.,Chylinski K.,Fonfara I.,Hauer M.,Doudna J.A.,Charpentier E.Science 337:816-821(2012),每篇的全部内容通过引用并入本文。

[0220] 在一些实施方案中,本文提供的任何sgRNA包含与napDNAbp结合的序列,例如sgRNA主链序列。例如,sgRNA已经记载于Jinek M,Chylinski K,Fonfara I,Hauer M,Doudna JA,and Charpentier E(2012)A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.Science,337,816-812;Mali P,Esvelt KM,Church GM(2013)Cas9 as a versatile tool for engineering biology.Nature Methods,10,957-963;Li JF,Norville JE,Aach J,McCromack M,Zhang D,Bush J,Church GM,and Sheen J(2013)Multiplex and homologous recombination-mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9.Nature Biotech,31,688-691;Hwang WY,Fu Y,Reyon D,Maeder ML,Tsai SQ,Sander JD,Peterson RT,YehJRJ,Joung JK(2013)Efficient in vivo genome editing using RNA-guided nucleases.Nat Biotechnol,31,227-229;Cong L,Ran FA,Cox D,Lin S,Barretto R,Habib N,Hsu PD,Wu X,Jiang W,Marraffini LA,Zhang F(2013)Multiplex genome engineering using CRIPSR/Cas systems.Science,339,819-823;Cho SW,Kim S,Kim JM,Kim JS(2013)Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease.Nat Biotechnol,31,230-232;Jinek MJ,East A,Cheng A,Lin S,Ma E,Doudna J(2013)RNA-programmed genome editing in human cells.eLIFE,2:e00471;DiCarlo JE,Norville JE,Mali P,Rios,Aach J,Church GM(2013)Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems.Nucl Acids Res,41,4336-4343;Briner AE,Donohoue PD,Gomaa AA,Selle K,Slorach EM,Nye CH,Haurwitz RE,Beisel CL,May AP,and Barrangou R(2014)Guide RNA functional modules direct Cas9 activity and orthogonality.Mol Cell,56,333-339;每篇的内容

通过引用并入本文。在一些实施方案中，本文提供的任何gRNA(例如sgRNA)包含GTAATTTCTACTAAGTGTAGAT(SEQ ID NO:741)的核酸序列，其中SEQ ID NO:741的每个T是尿嘧啶(U)，即GUAAUUUCUACUAAGUGUAGAU，或序列GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUU-3'(SEQ ID NO:618)。

[0221] 由于RNA可编程核酸酶(例如，Cas9)使用RNA:DNA杂交来靶向DNA切割位点，这些蛋白质原则上可以靶向由引导RNA规定的任何序列。使用RNA可编程核酸酶(例如Cas9)进行位点特异性切割(例如以修饰基因组)的方法是本领域已知的(参见例如Cong,L.et al.,Multiplex genome engineering using CRISPR/Cas systems.Science339,819-823(2013);Mali,P.et al.,RNA-guided human genome engineering via Cas9.Science339,823-826(2013);Hwang,W.Y.et al.,Efficient genome editing in zebrafish using a CRISPR-Cas system.Nature biotechnology31,227-229(2013);Jinek,M.et al.RNA-programmed genome editing in human cells.eLife2,e00471(2013);Dicarlo,J.E.et al.,Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems.Nucleic Acids Research(2013);Jiang,W.et al.,RNA-guided editing of bacterial genomes using CRISPR-Cas systems.Nature Biotechnology31,233-239(2013);每篇的全部内容其通过引用并入本文)。

[0222] 术语“Cas9”或“Cas9域”是指包含Cas9蛋白或其片段(例如，包含Cas9的活性、无活性或部分活性的DNA切割域，和/或Cas9的gRNA结合域的蛋白质)的RNA引导的核酸酶。Cas9核酸酶有时也称为casn1核酸酶或CRISPR(聚簇规则间隔短回文重复)相关核酸酶。CRISPR是适应性免疫系统，其提供针对移动遗传元件(病毒、可转座元件和接合质粒)的保护。CRISPR簇含有间隔物，与先前的移动元件互补的序列，并靶向侵入核酸。CRISPR簇得以转录并加工成CRISPR RNA(crRNA)。在II型CRISPR系统中，对pre-crRNA的正确加工需要反式编码的小RNA(tracrRNA)、内源性核糖核酸酶3(rnc)和Cas9蛋白。tracrRNA充当用于pre-crRNA的核糖核酸酶3辅助加工的引导。随后，Cas9/crRNA/tracrRNA以内切核水解方式切割与间隔物互补的线性或环状dsDNA靶物。首先以内切核水解方式切割不与crRNA互补的靶链，然后以3'-5'外切核水解方式修剪(trim)。在自然界中，DNA结合和切割通常需要蛋白质和这两种RNA。然而，单一引导RNA(“sgRNA”或简称“gNRA”)可以经工程化以将crRNA和tracrRNA两者的方面并入单一RNA种类中。参见例如Jinek M.,Chylinski K.,Fonfara I.,Hauer M.,Doudna J.A.,Charpentier E.Science337:816-821(2012)，其全部内容通过引用并入本文。Cas9识别CRISPR重复序列中的短基序(PAM或原间隔物邻近基序)，以帮助区分自我与非自我。Cas9核酸酶序列和结构是本领域技术人员公知的(参见例如“Complete genome sequence of an M1 strain of Streptococcus pyogenes.”Ferretti et al.,J.J.,McShan W.M.,Ajdic D.J.,Savic D.J.,Savic G.,Lyon K.,Primeaux C.,Sezate S.,Suvorov A.N.,Kenton S.,Lai H.S.,Lin S.P.,Qian Y.,Jia H.G.,Najar F.Z.,Ren Q.,Zhu H.,Song L.,White J.,Yuan X.,Clifton S.W.,Roe B.A.,McLaughlin R.E.,Proc.Natl.Acad.Sci.U.S.A.98:4658-4663(2001);“CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.”Deltcheva E.,Chylinski K.,Sharma C.M.,Gonzales K.,Chao Y.,Pirzada Z.A.,Eckert M.R.,Vogel J.,Charpentier E.,Nature 471:602-607(2011);和“A programmable dual-RNA-guided DNA endonuclease

in adaptive bacterial immunity."Jinek M.,Chylinski K.,Fonfara I.,Hauer M.,Doudna J.A.,Charpentier E.Science337:816-821(2012),其各自的全部内容通过引用并入本文)。已经在各种物种中描述了Cas9直系同源物,包括但不限于酿脓链球菌(S.pyogenes)和嗜热链球菌(S.thermophilus)。基于本公开,其他合适的Cas9核酸酶和序列对于本领域技术人员将是明显的,并且此类Cas9核酸酶和序列包括来自Chylinski,Rhun,and Charpentier,"The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems"(2013)RNA Biology 10:5,726-737中公开的生物体和基因座的Cas9序列;其全部内容通过引用并入本文。在一些实施方案中,Cas9核酸酶具有无活性的(例如失活的)DNA切割域,也就是说,Cas9是切口酶。

[0223] 核酸酶失活的Cas9蛋白可以互换地称为"dCas9"蛋白(代表核酸酶-"死亡的"Cas9)。用于生成具有无活性的DNA切割域的Cas9蛋白(或其片段)的方法是已知的(参见例如Jinek et al.,Science.337:816-821(2012);Qi et al.,"Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression"(2013)Cell.28;152(5):1173-83,其各自的全部内容通过引用并入本文)。例如,已知Cas9的DNA切割域包括两个亚域,即HNH核酸酶亚域和RuvC1亚域。HNH亚域切割与gRNA互补的链,而RuvC1亚域切割非互补链。这些亚域内的突变可以沉默Cas9的核酸酶活性。例如,突变D10A和H840A完全使酿脓链球菌Cas9的核酸酶活性失活(Jinek et al.,Science.337:816-821(2012);Qi et al.,Cell.28;152(5):1173-83(2013))。在一些实施方案中,提供了包含Cas9的片段的蛋白质。例如,在一些实施方案中,蛋白质包含两个Cas9域的一个:(1)Cas9的gRNA结合域;或(2)Cas9的DNA切割域。在一些实施方案中,包含Cas9或其片段的蛋白质称为"Cas9变体"。Cas9变体与Cas9或其片段共享同源性。例如,Cas9变体与野生型Cas9至少约70%相同、至少约80%相同、至少约90%相同、至少约95%相同、至少约96%相同、至少约97%相同、至少约98%相同、至少约99%相同、至少约99.5%相同或至少约99.9%相同。在一些实施方案中,Cas9变体与野生型Cas9相比,可以具有1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、21、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50个或更多个氨基酸变化。在一些实施方案中,Cas9变体包含Cas9的片段(例如,gRNA结合域或DNA切割域),使得该片段与野生型Cas9的相应的片段至少约70%相同、至少约80%相同、至少约90%相同、至少约95%相同、至少约96%相同、至少约97%相同、至少约98%相同、至少约99%相同、至少约99.5%相同或至少约99.9%相同。在一些实施方案中,片段是相应的野生型Cas9的氨基酸长度的至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%相同、至少96%、至少97%、至少98%、至少99%或至少99.5%。

[0224] 在一些实施方案中,片段的长度为至少100个氨基酸。在一些实施方案中,片段的长度为至少100、150、200、250、300、350、400、450、500、550、600、650、700、750、800、850、900、950、1000、1050、1100、1150、1200、1250或至少1300个氨基酸。在一些实施方案中,野生型Cas9对应于来自酿脓链球菌(Streptococcus pyogenes)的Cas9(NCBI参考序列:NC_017053.1,SEQ ID NO:1(核苷酸);SEQ ID NO:2(氨基酸))。

[0225] ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGATCACTGA

TGATTATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTATCAAAAAAAATCTTATA
GGGGCTCTTTTATTTGGCAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACAC
GTCGGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCA
TCGACTTGAAGAGTCTTTTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGAT
GAAGTTGCTTATCATGAGAAATATCCAACTATCTATCATCTGCGAAAAAAATTGGCAGATTCTACTGATAAAGCGG
ATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAGATTTAAA
TCCTGATAATAGTGATGTGGACAAACTATTTATCCAGTTGGTACAAATCTACAATCAATTATTTGAAGAAAACCCT
ATTAACGCAAGTAGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCA
TTGCTCAGCTCCCCGGTGAGAAGAGAAATGGCTTGTTTGGGAATCTCATTGCTTTGTCATTGGGATTGACCCCTAA
TTTTAAATCAAATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTAGAT
AATTTATTGGCGCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGATGCTATTTTAC
TTTCAGATATCCTAAGAGTAAATAGTGAAATAACTAAGGCTCCCCTATCAGCTTCAATGATTAAGCGCTACGATGA
ACATCATCAAGACTTGACTCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTATAAAGAAATCTTTTTT
GATCAATCAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAAATTTATCAAAC
CAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTGGTGAAACTAAATCGTGAAGATTTGCTGCGCAAGCAACG
GACCTTTGACAACGGCTCTATTCCCCATCAAATTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGAC
TTTTATCCATTTTTAAAAGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTC
CATTGGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCATGGAATTTTGA
AGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACAAACTTTGATAAAAATCTTCCAAAT
GAAAAAGTACTACCAAAACATAGTTTGCTTTATGAGTATTTTACGGTTTATAACGAATTGACAAAGGTCAAATATG
TTACTGAGGGAATGCGAAAACCAGCATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAAC
AAATCGAAAAGTAACCGTTAAGCAATTAAAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGAAATT
TCAGGAGTTGAAGATAGATTTAATGCTTCATTAGGCGCCTACCATGATTTGCTAAAAATTATTAAAGATAAAGATT
TTTTGGATAATGAAGAAAATGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAAGATAGGGGGAT
GATTGAGGAAAGACTTAAAACATATGCTCACCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTAT
ACTGGTTGGGGACGTTTGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTAGATT
TTTTGAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGACATTTAAAGAAGA
TATTCAAAAAGCACAGGTGTCTGGACAAGGCCATAGTTTACATGAACAGATTGCTAACTTAGCTGGCAGTCCTGCT
ATTAAAAAAGGTATTTTACAGACTGTAAAAATTGTTGATGAACTGGTCAAAGTAATGGGGCATAAGCCAGAAAATA
TCGTTATTGAAATGGCACGTGAAAATCAGACAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAAT
CGAAGAAGGTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAATACTCAATTGCAAAATGAA
AAGCTCTATCTCTATTATCTACAAAATGGAAGAGACATGTATGTGGACCAAGAATTAGATATTAATCGTTTAAGTG
ATTATGATGTCGATCACATTGTTCCACAAAGTTTCATTAAAGACGATTCAATAGACAATAAGGTACTAACGCGTTC
TGATAAAAATCGTGGTAAATCGGATAACGTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAA
CTTCTAAACGCCAAGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTGAAC
TTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAAATCACTAAGCATGTGGCACAAATTTTGGA
TAGTCGCATGAATACTAAATACGATGAAAATGATAAACTTATTCGAGAGGTTAAAGTGATTACCTTAAAATCTAAA
TTAGTTTCTGACTTCCGAAAAGATTTCCAATTCTATAAAGTACGTGAGATTAACAATTACCATCATGCCCATGATG
CGTATCTAAATGCCGTCGTTGGAACTGCTTTGATTAAGAAATATCCAAAACTTGAATCGGAGTTTGTCTATGGTGA

TTATAAAGTTTATGATGTTCGTAAAATGATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAAAATATTTC
TTTTACTCTAATATCATGAACTTCTTCAAAACAGAAATTACACTTGCAAATGGAGAGATTCGCAAACGCCCTCTAA
TCGAAACTAATGGGGAAACTGGAGAAATTGTCTGGGATAAAGGGCGAGATTTTGCCACAGTGCGCAAAGTATTGTC
CATGCCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAATTTTACCAAAA
AGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGATCCAAAAAAATATGGTGGTTTTGATAGTCCAACGG
TAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGAAATCGAAGAAGTTAAAATCCGTTAAAGAGTTACT
AGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAAATCCGATTGACTTTTTAGAAGCTAAAGGATATAAGGAA
GTTAAAAAAGACTTAATCATTAAACTACCTAAATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGG
CTAGTGCCGGAGAATTACAAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATTTAGCTAG
TCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTTGTGGAGCAGCATAAGCATTAT
TTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCGTGTTATTTTAGCAGATGCCAATTTAGATAAAGTTC
TTAGTGCATATAACAAACATAGAGACAAACCAATACGTGAACAAGCAGAAAATATTATTCATTTATTTACGTTGAC
GAATCTTGGAGCTCCCGCTGCTTTTAAATATTTTGATACAACAATTGATCGTAAACGATATACGTCTACAAAAGAA
GTTTTAGATGCCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAGCTAGGAG
GTGACTGA(SEQ ID NO:1)

[0226] MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETAEATRLKRT
ARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI
YHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQIYNQLFEENPINA
SRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY
DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKAL
VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP
WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL
SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDLLKIIKDKDFL
DNEENEDILEDIVLTLTLFEDRGMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQ
SGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGHSLHEQIANLAGSPAIKKGILQTVK
IVDELVKVMGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEK
LYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVV
KKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT
KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESE
FVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD
KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAY
SVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG
RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF
SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
ATLIHQSITGLYETRIDLSQLGGD(SEQ ID NO:2)

[0227] (单下划线:HNH域;双下划线:RuvC域)

[0228] 在一些实施方案中,野生型Cas9对应于或包含SEQ ID NO:3(核苷酸)和/或SEQ ID NO:4(氨基酸):

[0229] ATGGATAAAAAGTATTCTATTGGTTTAGACATCGGCACTAATTCCGTTGGATGGGCTGTCATAACCGA
TGAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACCGTCATTCGATTAAAAAGAATCTTATC
GGTGCCCTCCTATTCGATAGTGGCGAAACGGCAGAGGCGACTCGCCTGAAACGAACCGCTCGGAGAAGGTATACAC
GTCGCAAGAACCGAATATGTTACTTACAAGAAATTTTTAGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTTCA
CCGTTTGGAAGAGTCCTTCCTTGTCGAAGAGGACAAGAAACATGAACGGCACCCCATCTTTGGAAACATAGTAGAT
GAGGTGGCATATCATGAAAAGTACCCAACGATTTATCACCTCAGAAAAAAGCTAGTTGACTCAACTGATAAAGCGG

ACCTGAGGTTAATCTACTTGGCTCTTGCCCATATGATAAAGTTCCGTGGGCACTTTCTCATTGAGGGTGATCTAAA
TCCGGACAACTCGGATGTCGACAAACTGTTCATCCAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGAACCCT
ATAAATGCAAGTGGCGTGGATGCGAAGGCTATTCTTAGCGCCCGCCTCTCTAAATCCCGACGGCTAGAAAACCTGA
TCGCACAATTACCCGGAGAGAAGAAAAATGGGTTGTTCGGTAACCTTATAGCGCTCTCACTAGGCCTGACACCAAA
TTTTAAGTCGAACTTCGACTTAGCTGAAGATGCCAAATTGCAGCTTAGTAAGGACACGTACGATGACGATCTCGAC
AATCTACTGGCACAAATTGGAGATCAGTATGCGGACTTATTTTTGGCTGCCAAAAACCTTAGCGATGCAATCCTCC
TATCTGACATACTGAGAGTTAATACTGAGATTACCAAGGCGCCGTTATCCGCTTCAATGATCAAAAGGTACGATGA
ACATCACCAAGACTTGACACTTCTCAAGGCCCTAGTCCGTCAGCAACTGCCTGAGAAATATAAGGAAATATTCTTT
GATCAGTCGAAAAACGGGTACGCAGGTTATATTGACGGCGGAGCGAGTCAAGAGGAATTCTACAAGTTTATCAAAC
CCATATTAGAGAAGATGGATGGGACGGAAGAGTTGCTTGTAAAACTCAATCGCGAAGATCTACTGCGAAAGCAGCG
GACTTTCGACAACGGTAGCATTCCACATCAAATCCACTTAGGCGAATTGCATGCTATACTTAGAAGGCAGGAGGAT
TTTTATCCGTTCCTCAAAGACAATCGTGAAAAGATTGAGAAAATCCTAACCTTTCGCATACCTTACTATGTGGGAC
CCCTGGCCCGAGGGAACTCTCGGTTCGCATGGATGACAAGAAAGTCCGAAGAAACGATTACTCCATGGAATTTTGA
GGAAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTTCATCGAGAGGATGACCAACTTTGACAAGAATTTACCGAAC
GAAAAAGTATTGCCTAAGCACAGTTTACTTTACGAGTATTTCACAGTGTACAATGAACTCACGAAAGTTAAGTATG
TCACTGAGGGCATGCGTAAACCCGCCTTTCTAAGCGGAGAACAGAAGAAAGCAATAGTAGATCTGTTATTCAAGAC
CAACCGCAAAGTGACAGTTAAGCAATTGAAAGAGGACTACTTTAAGAAAATTGAATGCTTCGATTCTGTCGAGATC
TCCGGGGTAGAAGATCGATTTAATGCGTCACTTGGTACGTATCATGACCTCCTAAAGATAATTAAAGATAAGGACT
TCCTGGATAACGAAGAGAATGAAGATATCTTAGAAGATATAGTGTTGACTCTTACCCTCTTTGAAGATCGGGAAAT
GATTGAGGAAAGACTAAAAACATACGCTCACCTGTTCGACGATAAGGTTATGAAACAGTTAAAGAGGCGTCGCTAT
ACGGGCTGGGGACGATTGTCGCGGAAACTTATCAACGGGATAAGAGACAAGCAAAGTGGTAAAACTATTCTCGATT
TTCTAAAGAGCGACGGCTTCGCCAATAGGAACTTTATGCAGCTGATCCATGATGACTCTTTAACCTTCAAAGAGGA
TATACAAAAGGCACAGGTTTCCGGACAAGGGGACTCATTGCACGAACATATTGCGAATCTTGCTGGTTCGCCAGCC
ATCAAAAAGGGCATACTCCAGACAGTCAAAGTAGTGGATGAGCTAGTTAAGGTCATGGGACGTCACAAACCGGAAA
ACATTGTAATCGAGATGGCACGCGAAAATCAAACGACTCAGAAGGGGCAAAAAAACAGTCGAGAGCGGATGAAGAG
AATAGAAGAGGGTATTAAAGAACTGGGCAGCCAGATCTTAAAGGAGCATCCTGTGGAAAATACCCAATTGCAGAAC
GAGAAACTTTACCTCTATTACCTACAAAATGGAAGGGACATGTATGTTGATCAGGAACTGGACATAAACCGTTTAT
CTGATTACGACGTCGATCACATTGTACCCCAATCCTTTTTGAAGGACGATTCAATCGACAATAAAGTGCTTACACG
CTCGGATAAGAACCGAGGGAAAAGTGACAATGTTCCAAGCGAGGAAGTCGTAAAGAAAATGAAGAACTATTGGCGG
CAGCTCCTAAATGCGAAACTGATAACGCAAAGAAAGTTCGATAACTTAACTAAAGCTGAGAGGGGTGGCTTGTCTG
AACTTGACAAGGCCGGATTTATTAAACGTCAGCTCGTGGAAACCCGCCAAATCACAAAGCATGTTGCACAGATACT
AGATTCCCGAATGAATACGAAATACGACGAGAACGATAAGCTGATTCGGGAAGTCAAAGTAATCACTTTAAAGTCA
AAATTGGTGTCGGACTTCAGAAAGGATTTTCAATTCTATAAAGTTAGGGAGATAAATAACTACCACCATGCGCACG
ACGCTTATCTTAATGCCGTCGTAGGGACCGCACTCATTAAGAAATACCCGAAGCTAGAAAGTGAGTTTGTGTATGG
TGATTACAAAGTTTATGACGTCCGTAAGATGATCGCGAAAAGCGAACAGGAGATAGGCAAGGCTACAGCCAAATAC
TTCTTTTATTCTAACATTATGAATTTCTTTAAGACGGAAATCACTCTGGCAAACGGAGAGATACGCAAACGACCTT
TAATTGAAACCAATGGGGAGACAGGTGAAATCGTATGGGATAAGGGCCGGGACTTCGCGACGGTGAGAAAAGTTTT
GTCCATGCCCCAAGTCAACATAGTAAAGAAAACTGAGGTGCAGACCGGAGGGTTTTCAAAGGAATCGATTCTTCCA
AAAAGGAATAGTGATAAGCTCATCGCTCGTAAAAAGGACTGGGACCCGAAAAAGTACGGTGGCTTCGATAGCCCTA

CAGTTGCCTATTCTGTCCTAGTAGTGGCAAAAGTTGAGAAGGGAAAATCCAAGAAACTGAAGTCAGTCAAAGAATT
ATTGGGGATAACGATTATGGAGCGCTCGTCTTTTGAAAAGAACCCCATCGACTTCCTTGAGGCGAAAGGTTACAAG
GAAGTAAAAAAGGATCTCATAATTAAACTACCAAAGTATAGTCTGTTTGAGTTAGAAAATGGCCGAAAACGGATGT
TGGCTAGCGCCGGAGAGCTTCAAAAGGGGAACGAACTCGCACTACCGTCTAAATACGTGAATTTCCTGTATTTAGC
GTCCCATTACGAGAAGTTGAAAGGTTCACCTGAAGATAACGAACAGAAGCAACTTTTTGTTGAGCAGCACAAACAT
TATCTCGACGAAATCATAGAGCAAATTTCGGAATTCAGTAAGAGAGTCATCCTAGCTGATGCCAATCTGGACAAAG
TATTAAGCGCATACAACAAGCACAGGGATAAACCCATACGTGAGCAGGCGGAAAATATTATCCATTTGTTTACTCT
TACCAACCTCGGCGCTCCAGCCGCATTCAAGTATTTTGACACAACGATAGATCGCAAACGATACACTTCTACCAAG
GAGGTGCTAGACGCGACACTGATTCACCAATCCATCACGGGATTATATGAAACTCGGATAGATTTGTCACAGCTTG
GGGGTGACGGATCCCCCAAGAAGAAGAGGAAAGTCTCGAGCGACTACAAAGACCATGACGGTGATTATAAAGATCA
TGACATCGATTACAAGGATGACGATGACAAGGCTGCAGGA(SEQ ID NO:3)

[0230] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRT
ARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI
YHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPIN
ASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDT
YDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKA
LVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRT
FDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETIT
PWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF
LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFL
DNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQ
SGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVK
VVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQN
EKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRM
NTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLE
SEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIV
WDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT
VAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFEL
ENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQ
ISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKE
VLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO:4)

[0231] (单下划线:HNH域;双下划线:RuvC域)

[0232] 在一些实施方案中,野生型Cas9对应于来自酿脓链球菌的Cas9(NCBI参考序列:NC_002737.2,SEQ ID NO:5(核苷酸);和Uniport参考序列:Q99ZW2,SEQ ID NO:6(氨基酸)。

[0233] ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGATCACTGA
TGAATATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTATCAAAAAAAATCTTATA
GGGGCTCTTTTATTTGACAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACAC
GTCGGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCA
TCGACTTGAAGAGTCTTTTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGAT
GAAGTTGCTTATCATGAGAAATATCCAACTATCTATCATCTGCGAAAAAAATTGGTAGATTCTACTGATAAAGCGG
ATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAGATTTAAA
TCCTGATAATAGTGATGTGGACAAACTATTTATCCAGTTGGTACAAACCTACAATCAATTATTTGAAGAAAACCCT
ATTAACGCAAGTGGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCA
TTGCTCAGCTCCCCGGTGAGAAGAAAAATGGCTTATTTGGGAATCTCATTGCTTTGTCATTGGGTTTGACCCCTAA

TTTTAAATCAAATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTAGAT
AATTTATTGGCGCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGATGCTATTTTAC
TTTCAGATATCCTAAGAGTAAATACTGAAATAACTAAGGCTCCCCTATCAGCTTCAATGATTAAACGCTACGATGA
ACATCATCAAGACTTGACTCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTATAAAGAAATCTTTTTT
GATCAATCAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAAATTTATCAAAC
CAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTGGTGAAACTAAATCGTGAAGATTTGCTGCGCAAGCAACG
GACCTTTGACAACGGCTCTATTCCCCATCAAATTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGAC
TTTTATCCATTTTTAAAAGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTC
CATTGGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCATGGAATTTTGA
AGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACAAACTTTGATAAAAATCTTCCAAAT
GAAAAAGTACTACCAAAACATAGTTTGCTTTATGAGTATTTTACGGTTTATAACGAATTGACAAAGGTCAAATATG
TTACTGAAGGAATGCGAAAACCAGCATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAAC
AAATCGAAAAGTAACCGTTAAGCAATTAAAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGAAATT
TCAGGAGTTGAAGATAGATTTAATGCTTCATTAGGTACCTACCATGATTTGCTAAAAATTATTAAAGATAAAGATT
TTTTGGATAATGAAGAAAATGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAAGATAGGGAGAT
GATTGAGGAAAGACTTAAAACATATGCTCACCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTAT
ACTGGTTGGGGACGTTTGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTAGATT
TTTTGAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGACATTTAAAGAAGA
CATTCAAAAAGCACAAGTGTCTGGACAAGGCGATAGTTTACATGAACATATTGCAAATTTAGCTGGTAGCCCTGCT
ATTAAAAAAGGTATTTTACAGACTGTAAAAGTTGTTGATGAATTGGTCAAAGTAATGGGGCGGCATAAGCCAGAAA
ATATCGTTATTGAAATGGCACGTGAAAATCAGACAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACG
AATCGAAGAAGGTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAATACTCAATTGCAAAAT
GAAAAGCTCTATCTCTATTATCTCCAAAATGGAAGAGACATGTATGTGGACCAAGAATTAGATATTAATCGTTTAA
GTGATTATGATGTCGATCACATTGTTCCACAAAGTTTCCTTAAAGACGATTCAATAGACAATAAGGTCTTAACGCG
TTCTGATAAAAATCGTGGTAAATCGGATAACGTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGA
CAACTTCTAAACGCCAAGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTG
AACTTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAAATCACTAAGCATGTGGCACAAATTTT
GGATAGTCGCATGAATACTAAATACGATGAAAATGATAAACTTATTCGAGAGGTTAAAGTGATTACCTTAAAATCT
AAATTAGTTTCTGACTTCCGAAAAGATTTCCAATTCTATAAAGTACGTGAGATTAACAATTACCATCATGCCCATG
ATGCGTATCTAAATGCCGTCGTTGGAACTGCTTTGATTAAGAAATATCCAAAACTTGAATCGGAGTTTGTCTATGG
TGATTATAAAGTTTATGATGTTCGTAAAATGATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAAAATAT
TTCTTTTACTCTAATATCATGAACTTCTTCAAAACAGAAATTACACTTGCAAATGGAGAGATTCGCAAACGCCCTC
TAATCGAAACTAATGGGGAAACTGGAGAAATTGTCTGGGATAAAGGGCGAGATTTTGCCACAGTGCGCAAAGTATT
GTCCATGCCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAATTTTACCA
AAAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGATCCAAAAAAATATGGTGGTTTTGATAGTCCAA
CGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGAAATCGAAGAAGTTAAAATCCGTTAAAGAGTT
ACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAAATCCGATTGACTTTTTAGAAGCTAAAGGATATAAG
GAAGTTAAAAAAGACTTAATCATTAAACTACCTAAATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGC
TGGCTAGTGCCGGAGAATTACAAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATTTAGC

TAGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTTGTGGAGCAGCATAAGCAT
TATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCGTGTTATTTTAGCAGATGCCAATTTAGATAAAG
TTCTTAGTGCATATAACAAACATAGAGACAAACCAATACGTGAACAAGCAGAAAATATTATTCATTTATTTACGTT
GACGAATCTTGGAGCTCCCGCTGCTTTTAAATATTTTGATACAACAATTGATCGTAAACGATATACGTCTACAAAA
GAAGTTTTAGATGCCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAGCTAG
GAGGTGACTGA(SEQ ID NO:5)

[0234] MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRT
ARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI
YHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPIN
ASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDT
YDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKA
LVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRT
FDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETIT
PWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF
LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFL
DNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQ
SGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVK
VVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQN
EKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRM
NTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLE
SEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIV
WDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT
VAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFEL
ENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQ
ISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKE
VLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 6) (单下划线:HNH域;双下划线:RuvC域)

[0235] 在一些实施方案中,Cas9是指来自以下的Cas9:溃疡棒杆菌(Corynebacterium ulcerans)(NCBI Ref:NC_015683.1、NC_017317.1);白喉棒杆菌(Corynebacterium diphtheria)(NCBI Ref:NC_016782.1、NC_016786.1);Spiroplasma syrphidicola(NCBI Ref:NC_021284.1);间型普雷沃氏菌(Prevotella intermedia)(NCBI Ref:NC_017861.1);台湾螺原体(Spiroplasma taiwanense)(NCBI Ref:NC_021846.1);海豚链球菌(Streptococcus iniae)(NCBI Ref:NC_021314.1);波罗的海贝尔氏菌(Belliella baltica)(NCBI Ref:NC_018010.1);Psychroflexus torquisI(NCBI Ref:NC_018721.1);嗜热链球菌(Streptococcus thermophilus)(NCBI Ref:YP_820832.1)、无害李斯特氏菌(Listeria innocua)(NCBI Ref:NP_472073.1)、空肠弯曲杆菌(Campylobacter jejuni)(NCBI Ref:YP_002344900.1)或脑膜炎奈瑟氏球菌(Neisseria.meningitides)(NCBI Ref:YP_002342100.1)或者是指来自实施例5中列出的任何生物体的Cas9。

[0236] 在一些实施方案中,dCas9对应于或包含部分或全部的Cas9氨基酸序列,所述Cas9氨基酸序列具有一个或多个使Cas9核酸酶活性失活的突变。例如,在一些实施方案中,dCas9域包含D10A和H840A突变。

[0237] dCas9(D10A和H840A):

[0238] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL
KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEK
YPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE
NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQL
SKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS
EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMR
KPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKD
KDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
RDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGIL
QTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVE
NTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNR
GKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI
TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRK
RPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDW
DPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFV
EQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFD
TTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 7) (单下划线：HNH域；双下划线：RuvC域).

[0239] 在一些实施方案中，Cas9域包含D10A突变，而位置840处的残基在SEQ ID NO:6中提供的氨基酸序列中或在另一个Cas9域，例如本文提供的任何Cas9蛋白中提供的任何氨基酸序列中的相应位置处仍为组氨酸。不希望受任何特定理论的束缚，催化残基H840的存在恢复Cas9切割含有与靶定C相对的G的非编辑(例如非脱氨基)链的活性。H840的恢复(例如从A840)不导致含有C的靶链的切割。此类Cas9变体能够基于gRNA限定的靶序列在特定位置处产生单链DNA断裂(缺口)，从而导致非编辑链的修复，最终导致非编辑链中的G至A变化。此过程的示意图示于图108中。简言之，C-G碱基对的C可以通过脱氨酶，例如APOBEC脱氨酶被脱氨基成U。对具有G的非编辑链产生缺口有助于通过错配修复机制除去G。UGI抑制UDG，这防止U除去。

[0240] 在其他实施方案中，提供了具有除D10A和H840A之外的突变的dCas9变体，其例如导致核酸酶失活的Cas9(dCas9)。举例来说，此类突变包括D10和H820处的其他氨基酸取代，或Cas9的核酸酶域内的其他取代(例如，HNH核酸酶亚域和/或RuvC1亚域中的取代)。在一些实施方案中，提供了dCas9的变体或同源物(例如SEQ ID NO:6的变体)，其与SEQ ID NO:6至少约70％相同、至少约80％相同、至少约90％相同、至少约95％相同、至少约98％相同、至少约99％相同、至少约99.5％相同或至少约99.9％相同。在一些实施方案中，提供了dCas9的变体(例如SEQ ID NO:6的变体)，其具有比SEQ ID NO:6短或长约5个氨基酸、约10个氨基酸、约15个氨基酸、约20个氨基酸、约25个氨基酸、约30个氨基酸、约40个氨基酸、约50个氨基酸、约75个氨基酸、约100个氨基酸或更多的氨基酸序列。

[0241] 在一些实施方案中，如本文提供的Cas9融合蛋白包含Cas9蛋白的全长氨基酸序列，例如本文提供的Cas9序列之一。然而，在其他实施方案中，如本文提供的融合蛋白不包含全长Cas9序列，而仅包含其片段。例如，在一些实施方案中，本文提供的Cas9融合蛋白包含Cas9片段，其中所述片段结合crRNA和tracrRNA或sgRNA，但不包含功能性核酸酶域，例如

其中其仅包含截短形式的核酸酶域或根本没有核酸酶域。本文提供了合适的Cas9域和Cas9片段的示例性氨基酸序列，并且对于本领域技术人员来说，Cas9域和片段的另外合适的序列将是明显的。

[0242] 在一些实施方案中，Cas9是指来自以下的Cas9：溃疡棒杆菌(Corynebacterium ulcerans)(NCBI Ref:NC_015683.1、NC_017317.1)；白喉棒杆菌(Corynebacterium diphtheria)(NCBI Ref:NC_016782.1、NC_016786.1)；Spiroplasma syrphidicola(NCBI Ref:NC_021284.1)；间型普雷沃氏菌(Prevotella intermedia)(NCBI Ref:NC_017861.1)；台湾螺原体(Spiroplasma taiwanense)(NCBI Ref:NC_021846.1)；海豚链球菌(Streptococcus iniae)(NCBI Ref:NC_021314.1)；波罗的海贝尔氏菌(Belliella baltica)(NCBI Ref:NC_018010.1)；Psychroflexus torquisI(NCBI Ref:NC_018721.1)；嗜热链球菌(Streptococcus thermophilus)(NCBI Ref:YP_820832.1)；无害李斯特氏菌(Listeria innocua)(NCBI Ref:NP_472073.1)；空肠弯曲杆菌(Campylobacter jejuni)(NCBI Ref:YP_002344900.1)；或脑膜炎奈瑟氏球菌(Neisseria.meningitides)(NCBI Ref:YP_002342100.1)。

[0243] 如本文所用，术语“脱氨酶”或“脱氨酶域”是指催化脱氨基反应的蛋白质或酶。在一些实施方案中，脱氨酶或脱氨酶域是胞苷脱氨酶，其分别催化胞苷或脱氧胞苷水解为尿苷或脱氧尿苷。在一些实施方案中，脱氨酶或脱氨酶域是胞苷脱氨酶域，其催化胞嘧啶水解脱氨基为尿嘧啶。在一些实施方案中，脱氨酶或脱氨酶域是来自生物体，例如人，黑猩猩，大猩猩，猴，牛，狗，大鼠或小鼠的天然存在的脱氨酶。在一些实施方案中，脱氨酶或脱氨酶域是自然界中不存在的来自生物体的天然存在的脱氨酶的变体。例如，在一些实施方案中，脱氨酶或脱氨酶域与来自生物体的天然存在的脱氨酶为至少50％，至少55％，至少60％，至少65％，至少70％，至少75％，至少80％，至少85％，至少90％，至少95％，至少96％，至少97％，至少98％，至少99％或至少99.5％相同。

[0244] 如本文所用，术语“有效量”是指足以引起期望的生物应答的生物活性剂的量。例如，在一些实施方案中，核酸酶的有效量可以指足以诱导被核酸酶特异性结合和切割的靶位点的切割的核酸酶的量。在一些实施方案中，本文提供的融合蛋白的有效量，例如包含无核酸酶活性的Cas9域和核酸编辑域(例如，脱氨酶域)的融合蛋白的量可以指的是足以诱导对由融合蛋白特异性结合和编辑的靶位点的编辑。如熟练技术人员将理解的，试剂，例如融合蛋白，核酸酶，脱氨酶，重组酶，杂合蛋白，蛋白质二聚体，蛋白质(或蛋白质二聚体)和多核苷酸的复合物或多核苷酸的有效量可以随多种因素而变化，如例如期望的生物学应答，例如，特定的等位基因，基因组或待编辑的靶位点，所靶向的细胞或组织以及使用的试剂。

[0245] 如本文所用，术语“接头”是指连接两个分子或部分，例如融合蛋白的两个域，诸如例如无核酸酶活性的Cas9域和核酸编辑域(例如，脱氨酶域)的化学基团或分子。接头可以是例如任何长度和组成的氨基酸序列，肽或聚合物。在一些实施方案中，接头连接RNA可编程核酸酶的gRNA结合域，包括Cas9核酸酶域，和核酸编辑蛋白的催化域。在一些实施方案中，接头连接dCas9和核酸编辑蛋白。通常，接头位于两个基团，分子或其他部分之间或侧翼有两个基团，分子或其他部分，并通过共价键彼此连接，从而连接两者。在一些实施方案中，接头是氨基酸或多个氨基酸(例如，肽或蛋白质)。在一些实施方案中，接头是有机分子，基团，聚合物或化学部分。在一些实施方案中，接头的长度为1-100个氨基酸，例如长度为1,2,

3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,30-35,35-40,40-45,45-50,50-60,60-70,70-80,80-90,90-100,100-150,或150-200个氨基酸。也可以考虑更长或更短的接头。

[0246] 如本文所用,术语“突变”是指序列,例如核酸或氨基酸序列内的残基用另一个残基取代,或序列内的一个或多个残基的缺失或插入。突变通常在本文中通过鉴定原始残基,随后在序列中残基的位置以及新取代的残基的身份来描述。本文提供的用于进行氨基酸取代(突变)的各种方法是本领域公知的,并且例如由Green and Sambrook,Molecular Cloning:A Laboratory Manual(第4版,Cold Spring Harbor Laboratory Press,Cold Spring Harbor,N.Y.(2012))提供。

[0247] 如本文所用,术语“核酸”和“核酸分子”是指包含核碱基和酸性部分(例如核苷、核苷酸或核苷酸的聚合物)的化合物。通常,聚合核酸,例如包含三个或更多个核苷酸的核酸分子是线性分子,其中相邻的核苷酸经由磷酸二酯连接彼此连接。在一些实施方案中,“核酸”是指个别的核酸残基(例如核苷酸和/或核苷)。在一些实施方案中,“核酸”是指包含三个或更多个个别核苷酸残基的寡核苷酸链。如本文所用,术语“寡核苷酸”和“多核苷酸”可以可互换地使用以指核苷酸的聚合物(例如,至少三个核苷酸的串)。在一些实施方案中,“核酸”涵盖RNA以及单链和/或双链DNA。核酸可以是天然存在的,例如在基因组、转录物、mRNA、tRNA、rRNA、siRNA、snRNA、质粒、粘粒、染色体、染色单体或其他天然存在的核酸分子的背景下。另一方面,核酸分子可以是非天然存在的分子,例如重组DNA或RNA、人工染色体、工程化的基因组或其片段,或合成DNA、RNA、DNA/RNA杂交体,或包括非天然存在的核苷酸或核苷。此外,术语“核酸”、“DNA”、“RNA”和/或类似术语包括核酸类似物,例如具有除磷酸二酯主链之外的类似物。核酸可以从天然来源纯化,使用重组表达系统产生并任选地纯化、化学合成等。在适当的情况下,例如在化学合成分子的情况下,核酸可以包含核苷类似物,例如具有化学修饰的碱基或糖、和主链修饰的类似物。除非另有说明,核酸序列以5’至3’方向呈现。在一些实施方案中,核酸是或包含天然核苷(例如腺苷、胸苷、鸟苷、胞苷、尿苷、脱氧腺苷、脱氧胸苷、脱氧鸟苷和脱氧胞苷);核苷类似物(例如2-氨基腺苷、2-硫代胸苷、肌苷、吡咯并嘧啶、3-甲基腺苷、5-甲基胞苷、2-氨基腺苷、C5-溴尿苷、C5-氟尿苷、C5-碘尿苷、C5-丙炔基-尿苷、C5-丙炔基-胞苷、C5-甲基胞苷、2-氨基腺苷、7-脱氮腺苷、7-脱氮鸟苷、8-氧代腺苷、8-氧代鸟苷、O(6)-甲基鸟嘌呤和2-硫代胞苷);化学修饰的碱基;生物修饰的碱基(例如甲基化碱基);插入的碱基;修饰的糖(例如2’-氟核糖、核糖、2’-脱氧核糖、阿拉伯糖和己糖);和/或修饰的磷酸基团(例如硫代磷酸酯和5’-N-亚磷酰胺连接)。

[0248] 如本文所用,术语“核酸编辑域”是指能够对核酸(例如DNA或RNA)进行一个或多个修饰(例如胞苷残基的脱氨基)的蛋白质或酶。示例性核酸编辑域包括但不限于脱氨酶,核酸酶,切口酶,重组酶,甲基转移酶,甲基化酶,乙酰基酯酶(acetylase),乙酰基转移酶,转录激活物或转录阻遏物域。在一些实施方案中,核酸编辑域是脱氨酶(例如胞苷脱氨酶,例如APOBEC或AID脱氨酶)。

[0249] 如本文所用,术语“增殖性疾病”是指其中细胞或组织稳态受到干扰,使得细胞或细胞群表现出异常升高的增殖速率的任何疾病。增殖性疾病包括过度增殖性疾病,如肿瘤前期增生性状况和肿瘤性疾病。新生性疾病的特征是细胞的异常增殖,并包括良性和恶性新生物两者。恶性新生物也称为癌症。

[0250] 术语“蛋白质”、“肽”和“多肽”在本文中可互换使用，并且是指通过肽(酰胺)键连接在一起的氨基酸残基的聚合物。该术语是指具有任何大小、结构或功能的蛋白质、肽或多肽。通常，蛋白质、肽或多肽将是至少三个氨基酸长。蛋白质、肽或多肽可以指个别的蛋白质或蛋白质的集合。蛋白质、肽或多肽中的一个或多个氨基酸可以被修饰，例如通过添加化学实体如碳水化合物基团、羟基、磷酸基团、法呢基、异法呢基、脂肪酸基团，用于缀合、官能化或其他修饰的接头等。蛋白质、肽或多肽也可以是单个分子或者可以是多分子复合物。蛋白质、肽或多肽可以仅仅是天然存在的蛋白质或肽的片段。蛋白质、肽或多肽可以是天然存在的、重组的或合成的，或其任何组合。如本文所用的术语“融合蛋白”是指包含来自至少两种不同蛋白质的蛋白质域的杂合多肽。一种蛋白质可以位于融合蛋白的氨基端(N端)部分或羧基端(C端)蛋白质，从而分别形成“氨基端融合蛋白”或“羧基端融合蛋白”。蛋白质可以包含不同的域，例如核酸结合域(例如指导蛋白质与靶位点结合的Cas9的gRNA结合域)和核酸编辑蛋白的核酸切割域或催化域。在一些实施方案中，蛋白质包含蛋白质性部分，例如构成核酸结合域的氨基酸序列，和有机化合物，例如可以起核酸切割试剂作用的化合物。在一些实施方案中，蛋白质与核酸例如RNA复合或缔合。本文提供的任何蛋白质可以通过本领域已知的任何方法产生。例如，本文提供的蛋白质可以经由重组蛋白质表达和纯化产生，其特别适用于包含肽接头的融合蛋白。用于重组蛋白质表达和纯化的方法是熟知的，并且包括Green and Sambrook,Molecular Cloning:A Laboratory Manual(第4版,Cold Spring Harbor Laboratory Press,Cold Spring Harbor,N.Y.(2012))描述的那些，其全部内容通过引用并入本文。

[0251] 如本文所用，术语“受试者”是指个体生物体，例如个体哺乳动物。在一些实施方案中，受试者是人。在一些实施方案中，受试者是非人哺乳动物。在一些实施方案中，受试者是非人灵长类动物。在一些实施方案中，受试者是啮齿动物。在一些实施方案中，受试者是绵羊、山羊、牛、猫或狗。在一些实施方案中，受试者是脊椎动物、两栖动物、爬行动物、鱼、昆虫、苍蝇或线虫。在一些实施方案中，受试者是研究动物。在一些实施方案中，受试者是经遗传工程化的，例如基因遗传化的非人受试者。受试者可以是任何一个性别和处于任何发展阶段的。

[0252] 术语“靶位点”是指由脱氨酶或包含脱氨酶的融合蛋白(例如本文提供的dCas9-脱氨酶融合蛋白)脱氨基的核酸分子内的序列。

[0253] 术语“治疗/处理”是指如本文所述旨在逆转、缓解疾病或病症或其一种或多种症状、延迟疾病或病症或其一种或多种症状的发作或抑制疾病或病症或其一种或多种症状进展的临床干预。如本文所用，术语“治疗/处理”是指如本文所述旨在逆转、缓解疾病或病症或其一种或多种症状、延迟疾病或病症或其一种或多种症状的发作或抑制疾病或病症或其一种或多种症状进展的临床干预。在一些实施方案中，可以在一种或多种症状已经得以形成之后和/或疾病已经得到诊断之后施用治疗。在其他实施方案中，可以在没有症状的情况下施用治疗，例如用于预防或延迟症状的发作或抑制疾病的发作或进展。例如，可以在症状发作之前(例如，鉴于症状的历史和/或鉴于遗传或其他易感性因素)施用治疗于易感个体。治疗也可以在症状消退后继续进行，例如以预防或延迟其复发。

[0254] 如本文中在蛋白质或核酸的背景中使用，术语“重组”是指自然界中不存在，但是作为人工程化的产物的蛋白质或核酸。例如，在一些实施方案中，重组蛋白质或核酸分子包

含氨基酸或核苷酸序列，其相比于任何天然存在的序列包含至少一个、至少两个、至少三个、至少四个、至少五个、至少六个或至少七个突变。

[0255] 如本文所用，术语“药物组合物”是指可以在疾病或病症的治疗中向受试者施用的组合物。在一些实施方案中，药物组合物包含活性成分，例如核酸酶或编码核酸酶的核酸，以及药学上可接受的赋形剂。

[0256] 如本文所用，术语“碱基编辑器(BE)”或“核碱基编辑器(NBE)”是指包含能够对核酸序列(例如DNA或RNA)内的碱基(例如，A，T，C，G或U)进行修饰的多肽的试剂。在一些实施方案中，碱基编辑器能够使核酸内的碱基脱氨基。在一些实施方案中，碱基编辑器能够使DNA分子内的碱基脱氨基。在一些实施方案中，碱基编辑器能够使DNA中的胞嘧啶(C)脱氨基。在一些实施方案中，碱基编辑器是融合蛋白，其包含与胞苷脱氨酶域融合的核酸可编程DNA结合蛋白(napDNAbp)。在一些实施方案中，碱基编辑器包含与胞苷脱氨酶融合的Cas9(例如，dCas9和nCas9)，CasX，CasY，Cpf1，C2c1，C2c2，C2c3或Argonaute蛋白。在一些实施方案中，碱基编辑器包含与胞苷脱氨酶融合的Cas9切口酶(nCas9)。在一些实施方案中，碱基编辑器包含与胞苷脱氨酶融合的无核酸酶活性的Cas9(dCas9)。在一些实施方案中，碱基编辑器与碱基切除修复的抑制剂例如UGI域融合。在一些实施方案中，碱基编辑器包含与胞苷脱氨酶融合的CasX蛋白。在一些实施方案中，碱基编辑器包含与胞苷脱氨酶融合的CasY蛋白。在一些实施方案中，碱基编辑器包含与胞苷脱氨酶融合的Cpf1蛋白。在一些实施方案中，碱基编辑器包含与胞苷脱氨酶融合的C2c1蛋白。在一些实施方案中，碱基编辑器包含与胞苷脱氨酶融合的C2c2蛋白。在一些实施方案中，碱基编辑器包含与胞苷脱氨酶融合的C2c3蛋白。在一些实施方案中，碱基编辑器包含与胞苷脱氨酶融合的Argonaute蛋白。

[0257] 如本文所用，术语“尿嘧啶糖基化酶抑制剂”或“UGI”是指能够抑制尿嘧啶DNA糖基化酶碱基切除修复酶的蛋白质。

[0258] 如本文所用，术语“Cas9切口酶”是指能够裂解仅双链体的核酸分子(例如，双链体DNA分子)的一条链的Cas9蛋白。在一些实施方案中，Cas9切口酶包含D10A突变并且具有SEQ ID NO:6的H840位置处的组氨酸，或在另一个Cas9域，例如本文提供的任何Cas9蛋白中的相应突变。例如，Cas9切口酶可以包含如SEQ ID NO:8所示的氨基酸序列。此类Cas9切口酶具有活性的HNH核酸酶域并且能够切割DNA的非靶定链，即由gRNA结合的链。此外，此类Cas9切口酶具有无活性的RuvC核酸酶域，并且不能切割DNA的靶定链，即期望碱基编辑的链。

[0259] 示例性的Cas9切口酶(克隆载体pPlatTET-gRNA2；登录号BAV54124)。

[0260] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMG

RHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAE
RGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGE
IRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYG
GFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELEN
GRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD
ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI
DLSQLGGD(SEQ ID NO:8)

[0261] 发明详述

[0262] 本公开内容的一些方面提供融合蛋白,其包含能够结合核苷酸序列的域(例如,Cas9或Cpf1蛋白)和酶域,例如DNA编辑域,诸如例如脱氨酶域。脱氨酶对核碱基的脱氨基可以导致相应残基处的点突变,其在本文中称为核酸编辑。因此,包含Cas9变体或域和DNA编辑域的融合蛋白可用于核酸序列的靶向编辑。此类融合蛋白可用于体外DNA的靶向编辑,例如用于突变体细胞或动物的产生;用于引入靶向突变,例如用于校正离体细胞,例如从受试者获得的细胞中的遗传缺陷,所述细胞然后重新引入相同或另一个受试者中;以及用于引入靶向突变,例如在受试者中校正遗传缺陷或在疾病相关基因中引入失活突变。通常,本文所述融合蛋白的Cas9域没有任何核酸酶活性,而是Cas9片段或dCas9蛋白或域。本发明的其他方面提供了融合蛋白,其包含(i)能够结合核酸序列的域(例如,Cas9或Cpf1蛋白);(ii)酶域,例如,DNA编辑域(例如,脱氨酶域);和(iii)一个或多个尿嘧啶糖基化酶抑制剂(UGI)域。与没有UGI域的融合蛋白相比,至少一个UGI域的存在提高了碱基编辑效率。与具有一个UGI域或没有UGI域的融合蛋白相比,包含两个UGI域的融合蛋白进一步提高了碱基编辑效率和产物纯度。还提供了使用本文所述的Cas9融合蛋白的方法。

[0263] 核酸可编程DNA结合蛋白

[0264] 本公开内容的一些方面提供了核酸可编程DNA结合蛋白,其可以用于将蛋白质(例如碱基编辑器)引导至特定的核酸(例如,DNA或RNA)序列。应当理解,本文提供的任何融合蛋白(例如,碱基编辑器)可以包含任何核酸可编程DNA结合蛋白(napDNAbp)。例如,包含Cas9域的本文所述的任何融合蛋白可以使用另一个napDNAbp,例如CasX,CasY,Cpf1,C2c1,C2c2,C2c3和Argonaute替换Cas9域。核酸可编程DNA结合蛋白包括但不限于Cas9(例如,dCas9和nCas9),CasX,CasY,Cpf1,C2c1,C2c2,C2C3和Argonaute。与Cas9具有不同PAM特异性的核酸可编程DNA结合蛋白的一个例子是来自Prevotella和弗朗西丝菌属(Francisella)的成簇规则间隔短回文重复序列1(Cpf1)。与Cas9相似,Cpf1也是2类CRISPR效应物。已经显示Cpf1介导具有不同于Cas9的特征的稳健的DNA干扰。Cpf1是一种缺少tracrRNA的单一RNA引导的内切核酸酶,并且它利用富含T的原间隔物邻近基序(TTN,TTTN或YTN)。此外,Cpf1通过交错的DNA双链断裂切割DNA。在16种Cpf1家族蛋白中,显示来自氨基酸球菌属(Acidaminococcus)和毛螺菌科(Lachnospiraceae)的两种酶在人细胞中具有有效的基因组编辑活性。Cpf1蛋白在本领域中是已知的,并且先前已经记载于Yamano et al.,“Crystal structure of Cpf1 in complex with guide RNA and target DNA.”Cell(165)2016,p.949-962;其全部内容通过引用并入本文。

[0265] 在本组合物和方法中也有用的是无核酸酶活性的Cpf1(dCpf1)变体,其可用作引导核苷酸序列可编程DNA结合蛋白域。Cpf1蛋白具有类似于Cas9的RuvC域的RuvC样内切核酸酶域,但不具有HNH内切核酸酶域,并且Cpf1的N末端不具有Cas9的α-螺旋识别叶。Zetsche et al.,Cell,163,759-771,2015(通过引用并入本文)显示了Cpf1的RuvC样域负责切割两条DNA链并且RuvC样域的失活使Cpf1核酸酶活性失活。例如,与新凶手弗朗西斯菌Cpf1(SEQ ID NO:15)中的D917A,E1006A或D1255A相对应的突变使Cpf1核酸酶活性失活。在一些实施方案中,死亡Cpf1(dCpf1)包含对应于SEQ ID NO:9中的D917A,E1006A,D1255A,D917A/E1006A,D917A/D1255A,E1006A/D1255A,或D917A/E1006A/D1255A的突变。应当理解,根据本公开内容,可以使用使Cpf1的RuvC域失活的任何突变,例如取代突变,缺失或插入。

[0266] 在一些实施方案中,本文提供的任何融合蛋白的核酸可编程DNA结合蛋白(napDNAbp)可以是Cpf1蛋白。在一些实施方案中,Cpf1蛋白是Cpf1切口酶(nCpf1)。在一些实施方案中,Cpf1蛋白是无核酸酶活性的Cpf1(dCpf1)。在一些实施方案中,Cpf1,nCpf1或dCpf1包含与SEQ ID NO:9-24中的任一项至少85%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%或至少99.5%相同的氨基酸序列。在一些实施方案中,dCpf1包含与SEQ ID NO:9-16中的任一项至少85%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%或至少99.5%相同的氨基酸序列,并且包含对应于SEQ ID NO:9中的D917A,E1006A,D1255A,D917A/E1006A,D917A/D1255A,E1006A/D1255A,或D917A/E1006A/D1255A的突变。在一些实施方案中,dCpf1包含SEQ ID NO:9-16中任一项的氨基酸序列。应当理解,根据本公开内容,也可以使用来自其他物种的Cpf1。

[0267] 野生型新凶手弗朗西斯菌Cpf1(SEQ ID NO:9)(D917,E1006和D1255是粗体且加下划线的)

[0268]
MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSSVC
ISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESD
LILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDN
LPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQ
SGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLED
DSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYS
VIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILA
NFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQS
EDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEP
DNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKF
YNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSI
DEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNL
QDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITI
NFKSSGANKFNDEINLLLKEKANDVHILSIDRGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNY
HDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVFEDLNFGFKRGRFKV
EKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKI
CPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLI
NFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRN
SKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDADANGAYHIGLKGLMLLGRIKNNQEGKKLNL
VIKNEEYFEFVQNRNN (SEQ ID NO:9)

[0269] 新凶手弗朗西斯菌Cpf1 D917A(SEQ ID NO:10)(A917,E1006和D1255是粗体且加下划线的)

[0270] MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSSVC
ISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESD
LILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDN
LPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQ
SGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLED
DSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYS
VIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILA
NFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQS
EDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEP
DNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKF
YNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSI
DEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNL
QDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITI
NFKSSGANKFNDEINLLLKEKANDVHILSIARGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNY
HDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVFEDLNFGFKRGRFKV
EKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKI
CPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLI
NFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRN
SKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDADANGAYHIGLKGLMLLGRIKNNQEGKKLNL
VIKNEEYFEFVQNRNN (SEQ ID NO: 10)

[0271] 新凶手弗朗西斯菌Cpf1 E1006A(SEQ ID NO:11)(D917,A1006和D1255是粗体且加下划线的)

[0272] MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSSVC
ISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESD
LILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDN
LPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQ
SGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLED
DSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYS
VIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILA
NFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQS
EDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEP
DNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKF
YNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSI
DEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNL
QDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITI
NFKSSGANKFNDEINLLLKEKANDVHILSIDRGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNY
HDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVFADLNFGFKRGRFKV
EKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKI
CPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLI
NFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRN
SKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDADANGAYHIGLKGLMLLGRIKNNQEGKKLNL
VIKNEEYFEFVQNRNN (SEQ ID NO: 11)

[0273] 新凶手弗朗西斯菌Cpf1 D1255A(SEQ ID NO:12)(D917,E1006和A1255是粗体且加下划线的)

[0274] MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSSVC
ISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESD
LILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDN

[0275] LPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQ
SGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLED
DSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYS
VIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILA
NFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQS
EDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEP
DNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKF
YNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSI
DEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNL
QDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITI
NFKSSGANKFNDEINLLLKEKANDVHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNY
HDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**E**DLNFGFKRGRFKV
EKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKI
CPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLI
NFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRN
SKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**A**ANGAYHIGLKGLMLLGRIKNNQEGKKLNL
VIKNEEYFEFVQNRNN (SEQ ID NO: 12)

[0276] 新凶手弗朗西斯菌Cpf1 D917A/E1006A(SEQ ID NO:13)(A917,A1006,和D1255是粗体且加下划线的)

[0277] MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSSVC
ISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESD
LILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDN
LPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQ
SGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLED
DSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYS
VIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILA
NFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQS
EDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEP
DNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKF
YNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSI
DEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNL
QDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITI
NFKSSGANKFNDEINLLLKEKANDVHILSI**A**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNY
HDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**A**DLNFGFKRGRFKV
EKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKI
CPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLI
NFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRN
SKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**D**ANGAYHIGLKGLMLLGRIKNNQEGKKLNL
VIKNEEYFEFVQNRNN (SEQ ID NO: 13)

[0278] 新凶手弗朗西斯菌Cpf1 D917A/D1255A(SEQ ID NO:14)(A917,E1006,和A1255是粗体且加下划线的)

[0279] MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSSVC
ISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESD
LILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDN
LPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQ
SGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLED
DSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYS
VIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILA
NFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQS

[0280] EDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEP
DNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKF
YNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSI
DEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNL
QDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITI
NFKSSGANKFNDEINLLLKEKANDVHILSIARGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNY
HDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVFEDLNFGFKRGRFKV
EKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKI
CPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLI
NFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRN
SKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDAAANGAYHIGLKGLMLLGRIKNNQEGKKLNL
VIKNEEYFEFVQNRNN (SEQ ID NO: 14)

[0281] 新凶手弗朗西斯菌Cpf1 E1006A/D1255A(SEQ ID NO:15)(D917,A1006,和A1255是粗体且加下划线的)

[0282] MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSSVC
ISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESD
LILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDN
LPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQ
SGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLED
DSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYS
VIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILA
NFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQS
EDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEP
DNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKF
YNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSI
DEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNL
QDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITI
NFKSSGANKFNDEINLLLKEKANDVHILSIDRGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNY
HDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVFADLNFGFKRGRFKV
EKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKI
CPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLI
NFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRN
SKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDAAANGAYHIGLKGLMLLGRIKNNQEGKKLNL
VIKNEEYFEFVQNRNN (SEQ ID NO: 15)

[0283] 新凶手弗朗西斯菌Cpf1 D917A/E1006A/D1255A(SEQ ID NO:16)(A917,A1006和A1255是粗体且加下划线的)

[0284] MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSSVC
ISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESD
LILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDN
LPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQ
SGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLED
DSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYS
VIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILA
NFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQS
EDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEP
DNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKF
YNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSI
DEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNL
QDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITI

[0285] NFKSSGANKFNDEINLLLKEKANDVHILSIARGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNY
HDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVFADLNFGFKRGRFKV
EKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKI
CPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLI
NFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRN
SKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDAAANGAYHIGLKGLMLLGRIKNNQEGKKLNL
VIKNEEYFEFVQNRNN (SEQ ID NO: 16)

[0286] 在一些实施方案中，核酸可编程DNA结合蛋白是来自氨基酸球菌属物种的Cpf1蛋白(AsCpf1)。来自自氨基酸球菌属物种的Cpf1蛋白先前已经得到描述并且对于熟练技术人员将是明显的。示例性的氨基酸球菌属Cpf1蛋白(AsCpf1)包括但不限于本文中提供的任何AsCpf1蛋白。

[0287] 野生型AsCpf1-残基R912以粗体加下划线指示并且残基661-667以斜体且加下划线指示。

[0288] TQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTYADQCLQLV
QLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKAE
LFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVFSAEDISTAIPHRIVQDNFPKFKEN
CHIFTRLITAVPSLREHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIK
GLNEVLNLAIQKNDETAHIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVL
ETAEALFNELNSIDLTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKH
EDINLQEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTMLKKQEEKEILKSQLDSLLGLYHLLD
WFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNFQMPTLASGWDVN
KEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFDKMYYDYFPDAAKMIPKCSTQLK
AVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTAYAKKTGDQKGYREALCKWIDFTR
DFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDF
AKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKMLNKKLKDQ
KTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFFFHVPITLNYQAAN
SPSKFNQRVNAYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDNREKERVA
ARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFKSKRTGIAEKAVYQQFEKM
LIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFV
WKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAK
GTPFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIR
SVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHLKESK
DLKLQNGISNQDWLAYIQELRN (SEQ ID NO: 17)

[0289] AsCpf1(R912A)-残基A912以粗体加下划线指示并且残基661-667以斜体加下划线指示。

[0290] TQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTYADQCLQLV
QLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKAE
LFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVFSAEDISTAIPHRIVQDNFPKFKEN
CHIFTRLITAVPSLREHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIK
GLNEVLNLAIQKNDETAHIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVL
ETAEALFNELNSIDLTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKH
EDINLQEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTMLKKQEEKEILKSQLDSLLGLYHLLD
WFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNFQMPTLASGWDVN
KEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFDKMYYDYFPDAAKMIPKCSTQLK
AVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTAYAKKTGDQKGYREALCKWIDFTR
DFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDF

[0291] AKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKMLNKKLKDQ
KTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFFFHVPITLNYQAAN
SPSKFNQRVNAYLKEHPETPIIGIDRGEANLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDNREKERVA
ARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFKSKRTGIAEKAVYQQFEKM
LIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFV
WKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAK
GTPFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIR
SVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHLKESK
DLKLQNGISNQDWLAYIQELRN (SEQ ID NO: 19)

[0292] 在一些实施方案中,核酸可编程DNA结合蛋白是来自毛螺菌科物种的Cpf1蛋白(LbCpf1)。先前已经描述了来自毛螺菌科物种的Cpf1蛋白,并且对熟练技术人员是明显的。示例性毛螺菌科Cpf1蛋白(LbCpf1)包括但不限于本文提供的任何LbCpf1蛋白。

[0293] 在一些实施方案中,LbCpf1是切口酶。在一些实施方案中,LbCpf1切口酶包含相对于SEQ ID NO:18的R836X突变体,其中X是除R之外的任何氨基酸。在一些实施方案中,LbCpf1切口酶包含相对于SEQ ID NO:18的R836A突变体。在一些实施方案中,LbCpf1是无核酸酶活性的LbCpf1(dLbCpf1)。在一些实施方案中,dLbCpf1包含相对于SEQ ID NO:18的D832X突变体,其中X是除D外的任何氨基酸。在一些实施方案中,dLbCpf1包含相对于SEQ ID NO:18的D832A突变体。其它dCpf1蛋白已经在本领域中记载于例如Li et al.“Base editing with a Cpf1-cytidine deaminase fusion”Nature Biotechnology;March 2018DOI:10.1038/nbt.4102;其全部内容通过引用并入本文。在一些实施方案中,dCpf1包含Li et al中描述的Cpf1的点突变D832A,E1006A,D1125A中的1、2或3个。

[0294] 野生型LbCpf1-残基R836和R1138以粗体加下划线指示。

[0295] MSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRAEDYKGVKKLLDRYYLSFINDVL
HSIKLKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAFKGNEGYKSLFKKDIIETILPEFLDDKD
EIALVNSFNGFTTAFTGFFDNRENMFSEEAKSTSIAFRCINENLTRYISNMDIFEKVDAIFDKHEVQEIK
EKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNAIIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKP
LYKQVLSDRESLSFYGEGYTSDEEVLEVFRNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAI
STISKDIFGEWNVIRDKWNAEYDDIHLKKKAVVTEKYEDDRRKSFKKIGSFSLEQLQEYADADLSV
VEKLKEIIIQKVDEIYKVYGSSEKLFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEGK
ETNRDESFYGDFVLAYDILLKVDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYR
ATILRYGSKYYLAIMDKKYAKCLQKIDKDDVNGNYEKINYKLLPGPNKMLPKVFFSKKWMAYYN
PSEDIQKIYKNGTFKKGDMFNLNDCHKLIDFFKDSISRYPKWSNAYDFNFSETEKYKDIAGFYREVE
EQGYKVSFESASKKEVDKLVEEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSG
GAELFMRRASLKKEELVVHPANSPIANKNPDNPKKTTTLSYDVYKDKRFSEDQYELHIPIAINKCPK
NIFKINTEVRVLLKHDDNPYVIGIDRGERNLLYIVVVDGKGNIVEQYSLNEIINNFNGIRIKTDYHSLL
DKKEKERFEARQNWTSIENIKELKAGYISQVVHKICELVEKYDAVIALEDLNSGFKNSRVKVEKQV
YQKFEKMLIDKLNYMVDKKSNPCATGGALKGYQITNKFESFKSMSTQNGFIFYIPAWLTSKIDPSTG
FVNLLKTKYTSIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNP
KKNNVFDWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITG
RTDVDFLISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIGQFKKAEDEKLDKV
KIAISNKEWLEYAQTSVKH (SEQ ID NO: 18)

[0296]

[0297] LbCpf1(R836A)-残基A836以粗体加下划线指示。

[0298] MSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRAEDYKGVKKLLDRYYLSFINDVL
HSIKLKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAFKGNEGYKSLFKKDIIETILPEFLDDKD
EIALVNSFNGFTTAFTGFFDNRENMFSEEAKSTSIAFRCINENLTRYISNMDIFEKVDAIFDKHEVQEIK
EKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNAIIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKP
LYKQVLSDRESLSFYGEGYTSDEEVLEVFRNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAI
STISKDIFGEWNVIRDKWNAEYDDIHLKKKAVVTEKYEDDRRKSFKKIGSFSLEQLQEYADADLSV
VEKLKEIIIQKVDEIYKVYGSSEKLFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEGK
ETNRDESFYGDFVLAYDILLKVDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYR
ATILRYGSKYYLAIMDKKYAKCLQKIDKDDVNGNYEKINYKLLPGPNKMLPKVFFSKKWMAYYN
PSEDIQKIYKNGTFKKGDMFNLNDCHKLIDFFKDSISRYPKWSNAYDFNFSETEKYKDIAGFYREVE
EQGYKVSFESASKKEVDKLVEEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSG
GAELFMRRASLKKEELVVHPANSPIANKNPDNPKKTTTLSYDVYKDKRFSEDQYELHIPIAINKCPK
NIFKINTEVRVLLKHDDNPYVIGIDRGE**A**NLLYIVVVDGKGNIVEQYSLNEIINNFNGIRIKTDYHSLL
DKKEKERFEARQNWTSIENIKELKAGYISQVVHKICELVEKYDAVIALEDLNSGFKNSRVKVEKQV
YQKFEKMLIDKLNYMVDKKSNPCATGGALKGYQITNKFESFKSMSTQNGFIFYIPAWLTSKIDPSTG
FVNLLKTKYTSIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNP
KKNNVFDWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITG
RTDVDFLISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIGQFKKAEDEKLDKV
KIAISNKEWLEYAQTSVKH (SEQ ID NO: 20)

[0299] LbCpf1(R1138A)-残基A1138以粗体加下划线指示。

[0300] MSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRAEDYKGVKKLLDRYYLSFINDVL
HSIKLKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAFKGNEGYKSLFKKDIIETILPEFLDDKD
EIALVNSFNGFTTAFTGFFDNRENMFSEEAKSTSIAFRCINENLTRYISNMDIFEKVDAIFDKHEVQEIK
EKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNAIIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKP
LYKQVLSDRESLSFYGEGYTSDEEVLEVFRNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAI
STISKDIFGEWNVIRDKWNAEYDDIHLKKKAVVTEKYEDDRRKSFKKIGSFSLEQLQEYADADLSV
VEKLKEIIIQKVDEIYKVYGSSEKLFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEGK
ETNRDESFYGDFVLAYDILLKVDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYR
ATILRYGSKYYLAIMDKKYAKCLQKIDKDDVNGNYEKINYKLLPGPNKMLPKVFFSKKWMAYYN
PSEDIQKIYKNGTFKKGDMFNLNDCHKLIDFFKDSISRYPKWSNAYDFNFSETEKYKDIAGFYREVE
EQGYKVSFESASKKEVDKLVEEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSG
GAELFMRRASLKKEELVVHPANSPIANKNPDNPKKTTTLSYDVYKDKRFSEDQYELHIPIAINKCPK
NIFKINTEVRVLLKHDDNPYVIGIDRGERNLLYIVVVDGKGNIVEQYSLNEIINNFNGIRIKTDYHSLL
DKKEKERFEARQNWTSIENIKELKAGYISQVVHKICELVEKYDAVIALEDLNSGFKNSRVKVEKQV
YQKFEKMLIDKLNYMVDKKSNPCATGGALKGYQITNKFESFKSMSTQNGFIFYIPAWLTSKIDPSTG
FVNLLKTKYTSIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNP
KKNNVFDWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQM**A**NSITG
RTDVDFLISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIGQFKKAEDEKLDKV
KIAISNKEWLEYAQTSVKH (SEQ ID NO: 21)

[0301] 在一些实施方案中,Cpf1蛋白是残缺的Cpf1蛋白。如本文所用,“残缺的Cpf1”蛋白是与野生型Cpf1蛋白相比核酸酶活性降低的Cpf1蛋白。在一些实施方案中,残缺的Cpf1蛋白比非靶链优先更有效地切割靶链。例如,Cpf1蛋白优先切割待编辑的核苷酸驻留的双链体核酸分子的链。在一些实施方案中,残缺的Cpf1蛋白比靶链优先更有效地切割非靶链。例如,Cpf1蛋白优先切割待编辑的核苷酸不驻留的双链体核酸分子的链。在一些实施方案中,残缺的Cpf1蛋白比它切割非靶链有效至少5%优先切割靶链。在一些实施方案中,残缺的Cpf1蛋白比它切割非靶链有效至少5%,10%,15%,20%,25%,30%,35%,40%,50%,60%,70%,80%,90%或至少100%优先切割靶链

[0302] 在一些实施方案中,残缺的Cpf1蛋白是非天然存在的Cpf1蛋白。在一些实施方案中,残缺的Cpf1蛋白相对于野生型Cpf1蛋白包含一个或多个突变。在一些实施方案中,残缺的Cpf1蛋白相对于野生型Cpf1蛋白包含1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、

18、19或20个突变。在一些实施方案中，残缺的Cpf1蛋白包含如SEQ ID NO:18，或在另一个Cpf1蛋白的相应氨基酸中所示的R836A突变。应当理解，也可以将包含与SEQ ID NO:18的R836A的同源残基(例如，对应的氨基酸)的Cpf1突变以实现相似的结果。在一些实施方案中，残缺的Cpf1蛋白包含如SEQ ID NO:18，或在另一个Cpf1蛋白的相应氨基酸中所示的R1138A突变。在一些实施方案中，残缺的Cpf1蛋白包含SEQ ID NO:17，或在另一个Cpf1蛋白的相应氨基酸中所示的R912A突变。不希望受到任何特定理论的束缚，SEQ ID NO:18(LbCpf1)的残基R836和SEQ ID NO:17(AsCpf1)的残基R912是相应(例如同源)残基的实例。例如，SEQ ID NO:17和18之间的比对的部分显示R912和R836是相应的残基。

[0303]
```
AsCpf1      YQAANSPSKFNQRVNAYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQ--
LbCpf1      KCPKN-IFKINTEVRVLLKHDDNPYVIGIDRGERNLLYIVVVDGKGNIVEQYSLNEIINN
              *   *:* .*.. **.. :  :**********:**.*:*..*:*:** *** *
```

[0304] 在一些实施方案中，本文提供的任何Cpf1蛋白包含一个或多个氨基酸缺失。在一些实施方案中，本文提供的任何Cpf1蛋白包含1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个氨基酸缺失。不希望受任何特定理论的束缚，Cpf1中存在有螺旋区，该螺旋区包含AsCpf1(SEQ ID NO:17)的残基661-667，其可以阻碍与Cpf1融合的脱氨酶(例如APOBEC)的功能。此区域包含氨基酸序列KKTGDQK。因此，本公开内容的方面提供了包含破坏Cpf1中的此螺旋区的突变(例如，缺失)的Cpf1蛋白。在一些实施方案中，Cpf1蛋白包含SEQ ID NO:17中以下残基的一个或多个缺失，或在另一个Cpf1蛋白中的一个或多个相应缺失：K661,K662,T663,G664,D665,Q666和K667。在一些实施方案中，Cpf1蛋白包含SEQ ID NO:17中的T663和D665缺失，或在另一个Cpf1蛋白中相应的缺失。在一些实施方案中，Cpf1蛋白包含在SEQ ID NO:17中的K662,T663,D665和Q666缺失，或在另一个Cpf1蛋白中的相应的缺失。在一些实施方案中，Cpf1蛋白包含SEQ ID NO:17中的K661,K662,T663,D665,Q666和K667缺失，或在另一个Cpf1蛋白中的相应的缺失。

[0305] AsCpf1(缺失的T663和D665)

[0306] TQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTYADQCLQLV
QLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKAELFNGKVLKQ
LGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREH
FENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIAS
LPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSIDLTHIFISHKKLET
ISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAA
LDQPLPTTMLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKK
PYSVEKFKLNFQMPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFDKMYYDYFPD
AAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTAYAKKGQKGYREALCKWID
FTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGH
HGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKMLNKKLKDQKTPIPDTLYQELYDYV
NHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHPETPIIG
IDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDNREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDL
MIHYQAVVVLENLNFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQ
SGFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMP
AWDIVFEKNETQFDAKGTPFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSH

AIDTMVALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHLKESKD
LKLQNGISNQDWLAYIQELRN(SEQ ID NO:22)

[0307] AsCpf1(缺失的K662,T663,D665和Q666)

[0308] TQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTYADQCLQLV
QLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKAELFNGKVLKQ
LGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREH
FENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIAS
LPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSIDLTHIFISHKKLET
ISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAA
LDQPLPTTMLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKK
PYSVEKFKLNFQMPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFDKMYYDYFPD
AAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTAYAKGKGYREALCKWIDFT
RDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHG
KPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKMLNKKLKDQKTPIPDTLYQELYDYVNH
RLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGID
RGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDNREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMI
HYQAVVVLENLNFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQSG
FLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAW
DIVFEKNETQFDAKGTPFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAI
DTMVALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHLKESKDLK
LQNGISNQDWLAYIQELRN(SEQ ID NO:23)

[0309] AsCpf1(缺失的K661,K662,T663,D665,Q666和K667)

[0310] TQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTYADQCLQLV
QLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKAELFNGKVLKQ
LGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREH
FENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIAS
LPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSIDLTHIFISHKKLET
ISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAA
LDQPLPTTMLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKK
PYSVEKFKLNFQMPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFDKMYYDYFPD
AAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTAYAGGYREALCKWIDFTRD
FLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKP
NLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRL
SHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIDRG
ERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDNREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHY
QAVVVLENLNFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFL
FYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDI
VFEKNETQFDAKGTPFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDT
MVALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHLKESKDLKLQ

NGISNQDWLAYIQELRN(SEQ ID NO:24)

[0311] 在一些实施方案中,核酸可编程DNA结合蛋白(napDNAbp)是不需要靶序列中的规范(NGG)PAM序列的核酸可编程DNA结合蛋白。在一些实施方案中,napDNAbp是Argonaute蛋白。此类核酸可编程DNA结合蛋白的一个实例是来自格氏嗜盐碱杆菌(Natronobacterium gregoryi)的Argonaute蛋白(NgAgo)。NgAgo是ssDNA引导的内切核酸酶。NgAgo结合长度约24个核苷酸的5′磷酸化ssDNA(gDNA)以将其引导至靶位点,并在gDNA位点处产生DNA双链断裂。与Cas9相比,NgAgo-gDNA系统不需要原间隔物相邻基序(PAM)。使用无核酸酶活性的NgAgo(dNgAgo)可以极大地扩展可以靶向的碱基。NgAgo的表征和使用已经描述于Gao et al.,Nat.Biotechnol.,2016Jul;34(7):768-73.PubMed PMID:27136078;Swarts et al.,Nature 507(7491)(2014):258-61;和Swarts et al.,Nucleic Acids Res.43(10)(2015):5120-9,每篇通过引用并入本文。格氏嗜盐碱杆菌Argonaute的序列提供于SEQ ID NO:25中。

[0312] 在一些实施方案中,napDNAbp是Argonaute蛋白。在一些实施方案中,napDNAbp包含与天然存在的Argonaute蛋白至少85%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%或至少99.5%相同的氨基酸序列。在一些实施方案中,napDNAbp是天然存在的Argonaute蛋白。在一些实施方案中,napDNAbp包含与SEQ ID NO:25的任一项至少85%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%或至少99.5%相同的氨基酸序列。在一些实施方案中,napDNAbp包含SEQ ID NO:25的任一项的氨基酸序列。

[0313] 野生型格氏嗜盐碱杆菌Argonaute(SEQ ID NO:25)

[0314] MTVIDLDSTTTADELTSGHTYDISVTLTGVYDNTDEQHPRMSLAFEQDNGERRYITLWKNTTPKDVFTYDYATGSTYIFTNIDYEVKDGYENLTATYQTTVENATAQEVGTTDEDETFAGGEPLDHHLDDALNETPDDAETESDSGHVMTSFASRDQLPEWTLHTYTLTATDGAKTDTEYARRTLAYTVRQELYTDHDAAPVATDGLMLLTPEPLGETPLDLDCGVRVEADETRTLDYTTAKDRLLARELVEEGLKRSLWDDYLVRGIDEVLSKEPVLTCDEFDLHERYDLSVEVGHSGRAYLHINFRHRFVPKLTLADIDDDNIYPGLRVKTTYRPRRGHIVWGLRDECATDSLNTLGNQSVVAYHRNNQTPINTDLLDAIEAADRRVVETRRQGHGDDAVSFPQELLAVEPNTHQIKQFASDGFHQQARSKTRLSASRCSEKAQAFAERLDPVRLNGSTVEFSSEFFTGNNEQQLRLLYENGESVLTFRDGARGAHPDETFSKGIVNPPESFEVAVVLPEQQADTCKAQWDTMADLLNQAGAPPTRSETVQYDAFSSPESISLNVAGAIDPSEVDAAFVVLPPDQEGFADLASPTETYDELKKALANMGIYSQMAYFDRFRDAKIFYTRNVALGLLAAAGGVAFTTEHAMPGDADMFIGIDVSRSYPEDGASGQINIAATATAVYKDGTILGHSSTRPQLGEKLQSTDVRDIMKNAILGYQQVTGESPTHIVIHRDGFMNEDLDPATEFLNEQGVEYDIVEIRKQPQTRLLAVSDVQYDTPVKSIAAINQNEPRATVATFGAPEYLATRDGGGLPRPIQIERVAGETDIETLTRQVYLLSQSHIQVHNSTARLPITTAYADQASTHATKGYLVQTGAFESNVGFL(SEQ ID NO:25)

[0315] 在一些实施方案中,napDNAbp是Argonaute蛋白的原核同源物。Argonaute蛋白的原核同源物是已知的,并且已经记载于例如Makarova K.,et al.,“Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements”,Biol.Direct.2009Aug 25;4:29.doi:10.1186/1745-6150-4-29,其通过引用并入本文。在一些实施方案中,napDNAbp是Marinitoga piezophila Argunaute(MpAgo)蛋白。CRISPR相关的Marinitoga piezophila Argunaute(MpAgo)蛋白使用5’磷酸化的引导切割单链靶序列。所有已知的Argonautes都使

用5'引导。MpAgo-RNA复合物的晶体结构显示了引导链结合位点,该位点包含阻断5'磷酸相互作用的残基。此数据表明对5'-羟基化引导具有非规范特异性的Argonaute亚类的进化。参见例如Kaya et al.,"A bacterial Argonaute with noncanonical guide RNA specificity",Proc Natl Acad Sci U S A.2016Apr 12;113(15):4057-62,其全部内容通过引用并入本文。应当理解,可以在本文中描述的任何融合蛋白(例如碱基编辑器)中使用其他Argonaute蛋白以引导脱氨酶(胞苷脱氨酶)至靶核酸(例如ssRNA)。

[0316] 在一些实施方案中,核酸可编程DNA结合蛋白(napDNAbp)是微生物CRISPR-Cas系统的单一效应物。微生物CRISPR-Cas系统的单一效应物包括但不限于Cas9、Cpf1、C2c1、C2c2和C2c3。通常,微生物CRISPR-Cas系统分为1类和2类系统。1类系统具有多亚基效应物复合物,而2类系统具有单一蛋白质效应物。例如,Cas9和Cpf1是2类效应物。除了Cas9和Cpf1之外,Shmakov et al.,"Discovery and Functional Characterization of Diverse Class 2CRISPR Cas Systems",Mol.Cell,2015Nov 5;60(3):385-397(其全部内容在此通过引用并入)已经描述了三种不同的2类CRISPR-Cas系统(C2c1、C2c2和C2c3)。系统中的两种(C2c1和C2c3)的效应物含有与Cpf1相关的RuvC样内切核酸酶域。第三种系统C2c2含有具有两个预测的HEPN RNA酶域的效应物。与C2c1产生CRISPR RNA不同,成熟CRISPR RNA的产生是不依赖tracrRNA的。C2c1依赖于CRISPR RNA和tracrRNA两者用于DNA切割。已显示细菌性C2c2对于CRISPR RNA成熟具有独特的RNA酶活性,不同于其RNA激活的单链RNA降解活性。这些RNA酶功能彼此不同,并且与Cpf1的CRISPR RNA加工行为不同。参见例如East-Seletsky,et al.,"Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection",Nature,2016Oct 13;538(7624):270-273,其全部内容在此通过引用并入。Leptotrichia shahii中C2c2的体外生化分析已显示,C2c2由单一CRISPR RNA引导,并且可以编程以切割携带互补原间隔物的ssRNA靶物。两个保守的HEPN域中的催化残基介导切割。催化残基中的突变产生无催化活性的RNA结合蛋白。参见例如Abudayyeh et al.,"C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector,"Science,2016Aug 5;353(6299),其全部内容在此通过引用并入。

[0317] 已经报告了与嵌合单分子引导RNA(sgRNA)复合的酸土脂环酸芽孢杆菌(Alicyclobaccillus acidoterrastris)C2c1(AacC2c1)的晶体结构。参见例如Liu et al.,"C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism",Mol.Cell,2017Jan 19;65(2):310-322,其全部内容通过引用并入本文。还已经报告了作为三元复合物与靶DNA结合的酸土脂环酸芽孢杆菌C2c1的晶体结构。参见例如Yang et al.,"PAM-dependent Target DNA Recognition and Cleavage by C2C1 CRISPR-Cas endonuclease",Cell,2016Dec 15;167(7):1814-1828,其全部内容在此通过引用并入。具有靶DNA链和非靶DNA链两者的AacC2c1的催化能力构象已被独立地捕获,定位在单一RuvC催化口袋内,C2c1介导的切割导致靶DNA的交错的七核苷酸断裂。C2c1三元复合物与先前鉴定的Cas9和Cpf1对应物之间的结构比较证明了CRISPR-Cas9系统使用的机制的多样性。

[0318] 在一些实施方案中,本文提供的任何融合蛋白的核酸可编程DNA结合蛋白(napDNAbp)可以是C2c1、C2c2或C2c3蛋白。在一些实施方案中,napDNAbp是C2c1蛋白。在一些实施方案中,napDNAbp是C2c2蛋白。在一些实施方案中,napDNAbp是C2c3蛋白。在一些实

施方案中，napDNAbp包含与天然存在的C2c1、C2c2或C2c3蛋白至少85％、至少90％、至少91％、至少92％、至少93％、至少94％、至少95％、至少96％、至少97％、至少98％、至少99％或至少99.5％相同的氨基酸序列。在一些实施方案中，napDNAbp是天然存在的C2c1、C2c2或C2c3蛋白。在一些实施方案中，napDNAbp包含与SEQ ID NO:26-28的任一项至少85％、至少90％、至少91％、至少92％、至少93％、至少94％、至少95％、至少96％、至少97％、至少98％、至少99％或至少99.5％相同的氨基酸序列。在一些实施方案中，napDNAbp包含SEQ ID 26-28中任一项的氨基酸序列。应当理解，根据本公开也可以使用来自其他细菌物种的C2c1、C2c2或C2c3。

[0319] C2c1(uniprot.org/uniprot/T0D7A2#)

[0320] sp|T0D7A2|C2C1_ALIAG CRISPR相关的内切核酸酶C2c1 OS＝酸土脂环酸芽孢杆菌(菌株ATCC 49025/DSM 3922/CIP 106132/NCIMB 13137/GD3B)GN＝c2c1 PE＝1SV＝1

[0321] MAVKSIKVKLRLDDMPEIRAGLWKLHKEVNAGVRYYTEWLSLLRQENLYRRSPNGDGEQECDKTAEECKAELLERLRARQVENGHRGPAGSDDELLQLARQLYELLVPQAIGAKGDAQQIARKFLSPLADKDAVGGLGIAKAGNKPRWVRMREAGEPGWEEEKEKAETRKSADRTADVLRALADFGLKPLMRVYTDSEMSSVEWKPLRKGQAVRTWDRDMFQQAIERMMSWESWNQRVGQEYAKLVEQKNRFEQKNFVGQEHLVHLVNQLQQDMKEASPGLESKEQTAHYVTGRALRGSDKVFEKWGKLAPDAPFDLYDAEIKNVQRRNTRRFGSHDLFAKLAEPEYQALWREDASFLTRYAVYNSILRKLNHAKMFATFTLPDATAHPIWTRFDKLGGNLHQYTFLFNEFGERRHAIRFHKLLKVENGVAREVDDVTVPISMSEQLDNLLPRDPNEPIALYFRDYGAEQHFTGEFGGAKIQCRRDQLAHMHRRRGARDVYLNVSVRVQSQSEARGERRPPYAAVFRLVGDNHRAFVHFDKLSDYLAEHPDDGKLGSEGLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNSKGRVPFFFPIKGNDNLVAVHERSQLLKLPGETESKDLRAIREERQRTLRQLRTQLAYLRLLVRCGSEDVGRRERSWAKLIEQPVDAANHMTPDWREAFENELQKLKSLHGICSDKEWMDAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYAKDVVGGNSIEQIEYLERQYKFLKSWSFFGKVSGQVIRAEKGSRFAITLREHIDHAKEDRLKKLADRIIMEALGYVYALDERGKGKWVAKYPPCQLILLEELSEYQFNNDRPPSENNQLMQWSHRGVFQELINQAQVHDLLVGTMYAAFSSRFDARTGAPGIRCRRVPARCTQEHNPEPFPWWLNKFVVEHTLDACPLRADDLIPTGEGEIFVSPFSAEEGDFHQIHADLNAAQNLQQRLWSDFDISQIRLRCDWGEVDGELVLIPRLTGKRTADSYSNKVFYTNTGVTYYERERGKKRRKVFAQEKLSEEEAELLVEADEAREKSVVLMRDPSGIINRGNWTRQKEFWSMVNQRIEGYLVKQIRSRVPLQDSACENTGDI(SEQ ID NO:26)

[0322] C2c2(uniprot.org/uniprot/P0D0C6)

[0323] >sp|P0D0C6|C2C2_LEPSD CRISPR相关的内切核酸酶C2c2 OS＝Leptotrichia shahii(菌株DSM 19757/CCUG 47503/CIP 107916/JCM 16776/LB37)GN＝c2c2PE＝1SV＝1

[0324] MGNLFGHKRWYEVRDKKDFKIKRKVKVKRNYDGNKYILNINENNNKEKIDNNKFIRKYINYKKNDNILKEFTRKFHAGNILFKLKGKEGIIRIENNDDFLETEEVVLYIEAYGKSEKLKALGITKKKIIDEAIRQGITKDDKKIEIKRQENEEEIEIDIRDEYTNKTLNDCSIILRIIENDELETKKSIYEIFKNINMSLYKIIEKIIENETEKVFENRYYEEHLREKLLKDDKIDVILTNFMEIREKIKSNLEILGFVKFYLNVGGDKKKSKNKKMLVEKILNINVDLTVEDIADFVIKELEFWNITKRIEKVKKVNNEFLEKRRNRTYIKSYVLLDKHEKFKIERENKKDKIVKFFVENIKNNSIKEKIEKILAEFKIDELIKKLEKELKKGNCDTEIFGIFKKHYKVNFDSKKFSKKSDEEKELYKIIYRYLKGRIEKILVNEQKVRLKKMEKIEIEKILNESILSEKILKRVKQYTLEHIMYLGKLRHNDIDMTTVNTDDFSRLHAKEELDLELITFFASTNMELNKIFSRENINNDENIDFFGGDREKNYVLDKKILNSKIKIIRDLDFIDNKNNITNNFIRKFTKIGTNERNRILHAISKERDLQGTQDDYNKVINIIQNLKISDEEVSKALNLDVVFKDKKNIITKINDIKISEENNNDIKYLPSFSKV

LPEILNLYRNNPKNEPFDTIETEKIVLNALIYVNKELYKKLILEDDLEENESKNIFLQELKKTLGNIDEIDENIIE
NYYKNAQISASKGNNKAIKKYQKKVIECYIGYLRKNYEELFDFSDFKMNIQEIKKQIKDINDNKTYERITVKTSDK
TIVINDDFEYIISIFALLNSNAVINKIRNRFFATSVWLNTSEYQNIIDILDEIMQLNTLRNECITENWNLNLEEFI
QKMKEIEKDFDDFKIQTKKEIFNNYYEDIKNNILTEFKDDINGCDVLEKKLEKIVIFDDETKFEIDKKSNILQDEQ
RKLSNINKKDLKKKVDQYIKDKDQEIKSKILCRIIFNSDFLKKYKKEIDNLIEDMESENENKFQEIYYPKERKNEL
YIYKKNLFLNIGNPNFDKIYGLISNDIKMADAKFLFNIDGKNIRKNKISEIDAILKNLNDKLNGYSKEYKEKYIKK
LKENDDFFAKNIQNKNYKSFEKDYNRVSEYKKIRDLVEFNYLNKIESYLIDINWKLAIQMARFERDMHYIVNGLRE
LGIIKLSGYNTGISRAYPKRNGSDGFYTTTAYYKFFDEESYKKFEKICYGFGIDLSENSEINKPENESIRNYISHF
YIVRNPFADYSIAEQIDRVSNLLSYSTRYNNSTYASVFEVFKKDVNLDYDELKKKFKLIGNNDILERLMKPKKVSV
LELESYNSDYIKNLIIELLTKIENTNDTL(SEQ ID NO:27)

[0325] C2c3,翻译自>CEPX01008730.1海洋宏基因组基因组装配TARA_037_MES_0.1-0.22,重叠群TARA_037_MES_0.1-0.22_scaffold22115_1,全基因组鸟枪序列。

[0326] MRSNYHGGRNARQWRKQISGLARRTKETVFTYKFPLETDAAEIDFDKAVQTYGIAEGVGHGSLIGLVC
AFHLSGFRLFSKAGEAMAFRNRSRYPTDAFAEKLSAIMGIQLPTLSPEGLDLIFQSPPRSRDGIAPVWSENEVRNR
LYTNWTGRGPANKPDEHLLEIAGEIAKQVFPKFGGWDDLASDPDKALAAADKYFQSQGDFPSIASLPAAIMLSPAN
STVDFEGDYIAIDPAAETLLHQAVSRCAARLGRERPDLDQNKGPFVSSLQDALVSSQNNGLSWLFGVGFQHWKEKS
PKELIDEYKVPADQHGAVTQVKSFVDAIPLNPLFDTTHYGEFRASVAGKVRSWVANYWKRLLDLKSLLATTEFTLP
ESISDPKAVSLFSGLLVDPQGLKKVADSLPARLVSAEEAIDRLMGVGIPTAADIAQVERVADEIGAFIGQVQQFNN
QVKQKLENLQDADDEEFLKGLKIELPSGDKEPPAINRISGGAPDAAAEISELEEKLQRLLDARSEHFQTISEWAEE
NAVTLDPIAAMVELERLRLAERGATGDPEEYALRLLLQRIGRLANRVSPVSAGSIRELLKPVFMEEREFNLFFHNR
LGSLYRSPYSTSRHQPFSIDVGKAKAIDWIAGLDQISSDIEKALSGAGEALGDQLRDWINLAGFAISQRLRGLPDT
VPNALAQVRCPDDVRIPPLLAMLLEEDDIARDVCLKAFNLYVSAINGCLFGALREGFIVRTRFQRIGTDQIHYVPK
DKAWEYPDRLNTAKGPINAAVSSDWIEKDGAVIKPVETVRNLSSTGFAGAGVSEYLVQAPHDWYTPLDLRDVAHLV
TGLPVEKNITKLKRLTNRTAFRMVGASSFKTHLDSVLLSDKIKLGDFTIIIDQHYRQSVTYGGKVKISYEPERLQV
EAAVPVVDTRDRTVPEPDTLFDHIVAIDLGERSVGFAVFDIKSCLRTGEVKPIHDNNGNPVVGTVAVPSIRRLMKA
VRSHRRRRQPNQKVNQTYSTALQNYRENVIGDVCNRIDTLMERYNAFPVLEFQIKNFQAGAKQLEIVYGS(SEQ ID
NO:28)

[0327] 在一些实施方案中,本文提供的任何融合蛋白的核酸可编程DNA结合蛋白(napDNAbp)是来自构成单细胞原核微生物的域和界的古生菌(arehaea)(例如纳古生菌(nanoarchaea))的Cas9。在一些实施方案中,napDNAbp是CasX或CasY,其已经描述于例如Burstein et al.,“New CRISPR-Cas systems from uncultivated microbes.”Cell Res.2017Feb 21.doi:10.1038/cr.2017.21,其通过引用并入本文。使用基因组分辨的宏基因组学,鉴定了许多CRISPR-Cas系统,包括在生命的古生菌域中首次报告的Cas9。这种趋异的Cas9蛋白在纳古生菌中作为活性CRISPR-Cas系统的一部分发现。在细菌中,发现了两个以前未知的系统,CRISPR-CasX和CRISPR-CasY,它们是迄今发现的最紧凑的系统之一。在一些实施方案中,Cas9是指CasX或CasX的变体。在一些实施方案中,Cas9是指CasY或CasY的变体。应当理解,其他RNA引导的DNA结合蛋白可以用作核酸可编程DNA结合蛋白(napDNAbp),并且在本公开的范围内。

[0328] 在一些实施方案中,本文提供的任何融合蛋白的核酸可编程DNA结合蛋白

(napDNAbp)是CasX或CasY蛋白。在一些实施方案中,napDNAbp是CasX蛋白。在一些实施方案中,napDNAbp是CasY蛋白。在一些实施方案中,napDNAbp包含与天然存在的CasX或CasY蛋白至少85%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%或至少99.5%相同的氨基酸序列。在一些实施方案中,napDNAbp是天然存在的CasX或CasY蛋白。在一些实施方案中,napDNAbp包含与SEQ ID NO:29-31中的任一项至少85%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少具有96%,至少97%,至少98%,至少99%或至少99.5%相同的氨基酸序列。在一些实施方案中,napDNAbp包含SEQ ID NO:29-31中任一项的氨基酸序列。应当理解,根据本公开内容,也可以使用来自其他细菌物种的CasX和CasY。

[0329] CasX(uniprot.org/uniprot/F0NN87;uniprot.org/uniprot/F0NH53)

[0330] >tr|F0NN87|F0NN87_SULIH CRISPR相关的Casx蛋白OS＝冰岛硫化叶菌(Sulfolobus islandicus)(菌株HVE10/4)GN＝SiH_0402PE＝4SV＝1

[0331] MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAERRGKAKKKKGEEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKECEEVSAPSFVKPEFYEFGRSPGMVERTRRVKLEVEPHYLIIAAAGWVLTRLGKAKVSEGDYVGVNVFTPTRGILYSLIQNVNGIVPGIKPETAFGLWIARKVVSSVTNPNVSVVRIYTISDAVGQNPTTINGGFSIDLTKLLEKRYLLSERLEAIARNALSISSNMRERYIVLANYIYEYLTGSKRLEDLLYFANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG(SEQ ID NO:29)

[0332] >tr|F0NH53|F0NH53_SULIR CRISPR相关蛋白,Casx OS＝冰岛硫化叶菌(菌株REY15A)GN＝SiRe_0771PE＝4SV＝1

[0333] MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAERRGKAKKKKGEEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKECEEVSAPSFVKPEFYKFGRSPGMVERTRRVKLEVEPHYLIMAAAGWVLTRLGKAKVSEGDYVGVNVFTPTRGILYSLIQNVNGIVPGIKPETAFGLWIARKVVSSVTNPNVSVVSIYTISDAVGQNPTTINGGFSIDLTKLLEKRDLLSERLEAIARNALSISSNMRERYIVLANYIYEYLTGSKRLEDLLYFANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG(SEQ ID NO:30)

[0334] CasY(ncbi.nlm.nih.gov/protein/APG80656.1)

[0335] >APG80656.1 CRISPR相关蛋白CasY[未培养的Parcubacteria组细菌]

[0336] MSKRHPRISGVKGYRLHAQRLEYTGKSGAMRTIKYPLYSSPSGGRTVPREIVSAINDDYVGLYGLSNFDDLYNAEKRNEEKVYSVLDFWYDCVQYGAVFSYTAPGLLKNVAEVRGGSYELTKTLKGSHLYDELQIDKVIKFLNKKEISRANGSLDKLKKDIIDCFKAEYRERHKDQCNKLADDIKNAKKDAGASLGERQKKLFRDFFGISEQSENDKPSFTNPLNLTCCLLPFDTVNNNRNRGEVLFNKLKEYAQKLDKNEGSLEMWEYIGIGNSGTAFSNFLGEGFLGRLRENKITELKKAMMDITDAWRGQEQEEELEKRLRILAALTIKLREPKFDNHWGGYRSDINGKLSSWLQNYINQTVKIKEDLKGHKKDLKKAKEMINRFGESDTKEEAVVSSLLESIEKIVPDDSADDEKPDIPAIAIYRRFLSDGRLTLNRFVQREDVQEALIKERLEAEKKKKPKKRKKKSDAEDEKETIDFKELFPHLAKPLKLVPNFYGDSKRELYKKYKNAAIYTDALWKAVEKIYKSAFSSSLKNSFFDTDFDKDFFIKRLQKIFSVYRRFNTDKWKPIVKNSFAPYCDIVSLAENEVLYKPKQSRSRKSAAIDKNRVRLPSTENIAKAGIALARELSVAGFDWKDLLKKEEHEEYIDLIELHKTALALLLAVTETQLDISALDFVENGTVKDFMKTRDGNLVLEGRFLEMFSQSIVFSELRGLAGLMSRKEFITRSAIQTMNGKQAELLYIPHEFQSAKITTPKEMSRAFLDLAPAEFATSLEPESLSEKSLLKLKQMRYYPHYFGYELTRTGQGIDGGVAENALRLEKSPVKKREIKCKQYKTLGRGQNKIVLYVRSSYYQTQFLEWFLHRPKNVQTDVAVSGSFLIDEKKVKTRWNYDALTVALEPVSGSERVFVSQPFTIFPEKSAEEEGQRYLGIDIGEYGIAYTALEITGDSAKILDQNFISDPQLKTLREEVKGLKLD

QRRGTFAMPSTKIARIRESLVHSLRNRIHHLALKHKAKIVYELEVSRFEEGKQKIKKVYATLKKADVYSEIDADKN
LQTTVWGKLAVASEISASYTSQFCGACKKLWRAEMQVDETITTQELIGTVRVIKGGTLIDAIKDFMRPPIFDENDT
PFPKYRDFCDKHHISKKMRGNSCLFICPFCRANADADIQASQTIALLRYVKEEKKVEDYFERFRKLKNIKVLGQMK
KI(SEQ ID NO:31)

[0337] 核碱基编辑器的Cas9域

[0338] 本文提供了非限制性的示例性Cas9域。Cas9域可以是核酸酶活性Cas9域,无核酸酶活性Cas9域或Cas9切口酶。在一些实施方案中,Cas9域是核酸酶活性域。例如,Cas9域可以是切割双链体核酸的两条链(例如,双链体DNA分子的两条链)的Cas9域。在一些实施方案中,Cas9域包含任何Cas9蛋白,例如本文提供的任何Cas9氨基酸序列。在一些实施方案中,Cas9域包含与任何Cas9蛋白,诸如如本文提供的任一种Cas9氨基酸序列至少60%,至少65%,至少70%,至少75%,至少80%,至少85%,至少90%,至少95%,至少96%,至少97%,至少98%,至少99%或至少99.5%相同的氨基酸序列。在一些实施方案中,Cas9域包含与任何Cas9蛋白,诸如如本文提供的任一种Cas9氨基酸序列相比具有1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,21,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50或更多个突变的氨基酸序列。在一些实施方案中,Cas9域包含与任何Cas9蛋白质,诸如如本文提供的任一种Cas9氨基酸序列相比具有至少10,至少15,至少20,至少30,至少40,至少50,至少60,至少70,至少80,至少90,至少100,至少150,至少200,至少250,至少300,至少350,至少400,至少500,至少600,至少700,至少800,至少900,至少1000,至少1100或至少1200个相同的连续氨基酸残基的氨基酸序列。

[0339] 在一些实施方案中,Cas9域是无核酸酶活性的Cas9域(dCas9)。例如,dCas9域可以结合双链体核酸分子(例如,通过gRNA分子),而不切割双链体核酸分子的任一链。在一些实施方案中,无核酸酶活性的dCas9域包含以SEQ ID NO:6所示的氨基酸序列的D10X突变和H840X突变,或任何Cas9蛋白,诸如如本文提供的任一种Cas9氨基酸中的相应突变,其中X是任意氨基酸的变化。在一些实施方案中,无核酸酶活性的dCas9域包含以SEQ ID NO:6所示的氨基酸序列的D10A突变和H840A突变,或任何Cas9蛋白,诸如如本文提供的任一种Cas9氨基酸中的相应突变。作为一个实例,无核酸酶活性的Cas9域包含以SEQ ID NO:32所示的氨基酸序列(克隆载体pPlatTET-gRNA2,登录号BAV54124)。

[0340] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA
RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVD
STDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSR
RLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNL
SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNEL
TKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMG
RHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL

DINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAE
RGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGE
IRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYG
GFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELEN
GRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD
ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI
DLSQLGGD(SEQ ID NO:32;参见例如Qi et al.,Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression.Cell.2013;152(5):1173-83,其全部内容通过引用并入本文)。

[0341] 基于本公开内容和本领域的知识,其他合适的无核酸酶活性的dCas9域对本领域技术人员将是明显的,并且在本公开内容的范围内。此类另外的示例性的适合的无核酸酶活性的Cas9域包括但不限于D10A/H840A,D10A/D839A/H840A,和D10A/D839A/H840A/N863A突变体域(参见例如Prashant et al.,CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering.Nature Biotechnology.2013;31(9):833-838,其全部内容通过引用并入本文)。在一些实施方案中,dCas9域包含与本文提供的任一种dCas9域至少60%,至少65%,至少70%,至少75%,至少80%,至少85%,至少90%,至少95%,至少96%,至少97%,至少98%,至少99%或至少99.5%相同的氨基酸序列。在一些实施方案中,Cas9域包含与任何Cas9蛋白质,诸如如本文提供的任一种Cas9氨基酸序列相比具有1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,21,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50个或更多的突变。在一些实施方案中,Cas9域包含与任何Cas9蛋白质,诸如如本文提供的任一种Cas9氨基酸序列相比具有至少10,至少15,至少20,至少30,至少40,至少50,至少60,至少70,至少80,至少90,至少100,至少150,至少200,至少250,至少300,至少350,至少400,至少500,至少600,至少700,至少800,至少900,至少1000,至少1100或至少1200个相同的连续氨基酸残基的氨基酸序列。

[0342] 在一些实施方案中,Cas9域是Cas9切口酶。Cas9切口酶可以是仅能够切割双链体核酸分子(例如双链体DNA分子)的一条链的Cas9蛋白。在一些实施方案中,Cas9切口酶切割双链体核酸分子的靶链,这意味着Cas9切口酶切割与结合至Cas9的gRNA(例如,sgRNA)碱基配对(互补)的链。在一些实施方案中,Cas9切口酶包含D10A突变并且具有SEQ ID NO:6的位置840处的组氨酸,或者在任何Cas9蛋白,诸如如本文提供的任一种Cas9氨基酸序列中的突变。例如,Cas9切口酶可以包含如SEQ ID NO:8所示的氨基酸序列。在一些实施方案中,Cas9切口酶切割双链体核酸分子的非靶标非碱基编辑的链,这意味着Cas9切口酶切割不与结合至Cas9的gRNA(例如sgRNA)碱基配对的链。在一些实施方案中,Cas9切口酶包含H840A突变并且具有SEQ ID NO:6的位置10处的天冬氨酸残基,或在任何Cas9蛋白,诸如如本文提供的任一种Cas9氨基酸序列中的突变。在一些实施方案中,Cas9切口酶包含与本文提供的任何Cas9切口酶至少60%,至少65%,至少70%,至少75%,至少80%,至少85%,至少90%,至少95%,至少96%,至少97%,至少98%,至少99%或至少99.5%相同的氨基酸序列。基于本公开内容和本领域的知识,其他合适的Cas9切口酶对于本领域技术人员将是明显的,并且在

本公开内容的范围内。

[0343] 具有降低的PAM排他性的Cas9域

[0344] 本公开内容的一些方面提供了具有不同PAM特异性的Cas9域。通常,Cas9蛋白,例如来自酿脓链球菌的Cas9(spCas9),需要规范的NGG PAM序列来结合特定的核酸区域。这可以限制在基因组内编辑期望的碱基的能力。在一些实施方案中,本文提供的碱基编辑融合蛋白需要定位于精确的位置处,例如,其中靶碱基在4碱基区域(例如"脱氨基作用窗口")内,其在PAM的上游的约15个碱基。参见Komor,A.C.,et al.,"Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage"Nature 533,420-424(2016),其全部内容在此通过引用并入。因此,在一些实施方案中,本文提供的任何融合蛋白可以含有能够结合不含规范的(例如,NGG)PAM序列的核苷酸序列的Cas9域。本领域已经描述了结合非规范PAM序列的Cas9域,并且其对于熟练技术人员而言是明显的。例如,结合非规范PAM序列的Cas9域已经描述于Kleinstiver,B.P.,et al.,"Engineered CRISPR-Cas9 nucleases with altered PAM specificities"Nature 523,481-485(2015);和Kleinstiver,B.P.,et al.,"Broadening the targeting range ofStaphylococcus aureus CRISPR-Cas9 by modifying PAM recognition"Nature Biotechnology 33,1293-1298(2015);每篇的全部内容在此通过引用并入。

[0345] 在一些实施方案中,Cas9域是来自金黄色葡萄球菌的Cas9域(SaCas9)。在一些实施方案中,SaCas9域是核酸酶活性的SaCas9、无核酸酶活性的SaCas9(SaCas9d)或SaCas9切口酶(SaCas9n)。在一些实施方案中,SaCas9包含氨基酸序列SEQ ID NO:33。在一些实施方案中,SaCas9包含SEQ ID NO:33的N579X突变或任何Cas9蛋白,诸如如本文中提供的任一种氨基酸序列中的相应的突变,其中X是除N之外的任何氨基酸。在一些实施方案中,SaCas9包含SEQ ID NO:33的N579A突变或任何Cas9蛋白,诸如如本文中提供的任一种氨基酸序列中的相应的突变。在一些实施方案中,SaCas9域、SaCas9d域或SaCas9n域可以结合具有非规范PAM的核酸序列。在一些实施方案中,SaCas9域、SaCas9d域或SaCas9n域可以结合具有NNGRRT PAM序列的核酸序列。在一些实施方案中,SaCas9域包含SEQ ID NO:33的E781X、N967X和R1014X突变中的一个或多个或任何Cas9蛋白,诸如如本文中提供的任一种氨基酸序列中的相应的突变,其中X是任意氨基酸。在一些实施方案中,SaCas9域包含SEQ ID NO:33的E781K、N967K和R1014H突变中的一个或多个或任何Cas9蛋白,诸如如本文中提供的任一种氨基酸序列中的相应的突变。在一些实施方案中,SaCas9域包含SEQ ID NO:33的E781K、N967K或R1014H突变或任何Cas9蛋白,诸如如本文中提供的任一种氨基酸序列中的相应的突变。

[0346] 在一些实施方案中,本文提供的任何融合蛋白的Cas9域包含与SEQ ID NO:33-36中的任一个至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相同的氨基酸序列。在一些实施方案中,本文提供的任何融合蛋白的Cas9域包含SEQ ID NO:33-36中任一个的氨基酸序列。在一些实施方案中,本文提供的任何融合蛋白的Cas9域由SEQ ID NO:33-36中任一个的氨基酸序列组成。

[0347] 示例性SaCas9序列

[0348] KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVK
KLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTK
EQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYI
DLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVI
TRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDIT
ARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDEL
WHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIEL
AREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLED
LLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISYETFKKHILNLAKG
KGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFT
SFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIET
EQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDN

[0349] DKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIK
KIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVN
SKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDK
RPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG (SEQ ID NO: 33)

[0350] SEQ ID NO:33的残基N579(其是加下划线且粗体的)可以突变(例如突变为A579)以产生SaCas9切口酶。

[0351] 示例性SaCas9d序列

[0352] KRNYILGLAIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVK
KLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTK
EQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYI
DLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVI
TRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDIT
ARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDEL
WHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIEL
AREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLED
LLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISYETFKKHILNLAKG
KGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFT
SFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIET
EQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDN
DKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIK
KIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVN
SKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDK
RPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG (SEQ ID NO: 34)

[0353] SEQ ID NO:34的残基D10(其是加下划线且粗体的)可以突变(例如突变为A10)以产生无核酸酶活性的SaCas9d。

[0354] 示例性SaCas9n序列

[0355] KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVK
KLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTK
EQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYI
DLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVI
TRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDIT
ARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDEL
WHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIEL
AREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLED
LLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEE<u>A</u>SKKGNRTPFQYLSSSDSKISYETFKKHILNLAKG
KGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFT
SFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIET
EQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDN
DKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIK
KIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVN
SKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDK
RPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG (SEQ ID NO: 35)。

[0356] SEQ ID NO:35的残基A579(其可以从SEQ ID NO:33的N579突变以产生SaCas9切口酶)是加下划线且粗体的。

[0357] 示例性SaKKH Cas9

[0358] KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVK
KLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTK
EQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYI
DLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVI
TRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDIT
ARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDEL
WHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIEL
AREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLED
LLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEE<u>A</u>SKKGNRTPFQYLSSSDSKISYETFKKHILNLAKG
KGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFT
SFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIET
EQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNR<u>*K*</u>LINDTLYSTRKDDKGNTLIVNNLNGLYDKDN
DKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIK
KIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVN
SKCYEEAKKLKKISNQAEFIASFY<u>*K*</u>NDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDK
RPP<u>*H*</u>IIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG(SEQ ID NO: 36)。

[0359] SEQ ID NO:36的残基A579(其可以从SEQ ID NO:36的N579突变以产生SaCas9切口酶)带下划线并为粗体。SEQ ID NO:36的残基K781,K967和H1014(其可以从SEQ ID NO:36的E781,N967和R1014突变以产生SaKKHCas9)是加下划线且且斜体的。

[0360] 在一些实施方案中,Cas9域是来自酿脓链球菌(SpCas9)的Cas9域。在一些实施方案中,SpCas9域是核酸酶活性SpCas9,无核酸酶活性SpCas9(SpCas9d)或SpCas9切口酶(SpCas9n)。在一些实施方案中,SpCas9包含氨基酸序列SEQ ID NO:37。在一些实施方案中,SpCas9包含SEQ ID NO:37的D9X突变,或任何Cas9蛋白,例如如本文中提供的任一种Cas9中的相应突变,其中X是除D以外的任何氨基酸。在一些实施方案中,SpCas9包含SEQ ID NO:37的D9A突变,或任何Cas9蛋白,例如如本文中提供的任一种Cas9中的相应突变。在一些实施方案中,SpCas9域,SpCas9d域或SpCas9n域可以结合具有非规范PAM的核酸序列。在一些实施方案中,SpCas9域,SpCas9d域或SpCas9n域可以结合具有NGG,NGA或NGCG PAM序列的核酸序列。在一些实施方案中,SpCas9域包含SEQ ID NO:37的D1134X,R1334X和T1336X突变中的一个或多个,或任何Cas9蛋白,例如如本文中提供的任一种Cas9中的相应突变,X是任意氨

基酸。在一些实施方案中，SpCas9域包含SEQ ID NO:37的D1134E，R1334Q和T1336R突变中的一个或多个，或任何Cas9蛋白，例如如本文中提供的任一种Cas9中的相应突变。在一些实施方案中，SpCas9域包含SEQ ID NO:37的D1134E，R1334Q和T1336R突变，或任何Cas9蛋白，例如如本文中提供的任一种Cas9中的相应突变。在一些实施方案中，SpCas9域包含SEQ ID NO:37的D1134X，R1334X和T1336X突变中的一个或多个，或任何Cas9蛋白，例如如本文中提供的任一种Cas9中的相应突变，X是任意氨基酸。在一些实施方案中，SpCas9域包含SEQ ID NO:37的D1134V，R1334Q和T1336R突变中的一个或多个，或任何Cas9蛋白，例如如本文中提供的任一种Cas9中的相应突变。在一些实施方案中，SpCas9域包含SEQ ID NO:37的D1134V，R1334Q和T1336R突变，或任何Cas9蛋白，例如如本文中提供的任一种Cas9中的相应突变。在一些实施方案中，SpCas9域包含SEQ ID NO:37的D1134X，G1217X，R1334X和T1336X突变中的一个或多个，或任何Cas9蛋白，例如如本文中提供的任一种Cas9中的相应突变，其中X是任意氨基酸。在一些实施方案中，SpCas9域包含SEQ ID NO:37的D1134V，G1217R，R1334Q和T1336R突变中的一个或多个，或任何Cas9蛋白，例如如本文中提供的任一种Cas9中的相应突变。在一些实施方案中，SpCas9域包含SEQ ID NO:37的D1134V，G1217R，R1334Q和T1336R突变，或任何Cas9蛋白，例如如本文中提供的任一种Cas9中的相应突变。

[0361] 在一些实施方案中，本文提供的任何融合蛋白的Cas9域包含与SEQ ID NO:37-41中任一项至少60％，至少65％，至少70％，至少75％，至少80％，至少85％，至少90％，至少95％，至少96％，至少97％，至少98％，至少99％或至少99.5％相同的氨基酸序列。在一些实施方案中，本文提供的任何融合蛋白的Cas9域包含SEQ ID NO:37-41中任一项的氨基酸序列。在一些实施方案中，本文提供的任何融合蛋白的Cas9域由SEQ ID NO:37-41中任一项的氨基酸序列组成。

[0362] 示例性SpCas9

[0363] DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDS
TDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRR
LENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY
KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIP
YYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELT
KVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKII
KDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGK
TILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGR
HKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELD
INRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAER
GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNY
HHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI
RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGG
FDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG
RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA

NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRID LSQLGGD(SEQ ID NO:37)

[0364] 示例性SpCas9n

[0365] DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDS TDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRR LENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIP YYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELT KVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKII KDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGK TILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGR HKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELD INRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAER GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNY HHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGG FDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRID LSQLGGD(SEQ ID NO:38)

[0366] 示例性SpEQR Cas9

[0367] DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIY HLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNE KLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEV VKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMN TKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLES EFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVW DKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF**E**SPTVA YSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELEN GRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISE

[0368] FSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK**Q**Y**R**STKEVL DATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 39)

[0369] SEQ ID NO:39的残基E1134,Q1334和R1336(其可以从SEQ ID NO:39的D1134,R1334和T1336突变以产生SpEQR Cas9)是加下划线且粗体的。

[0370] 示例性SpVQR Cas9

[0371] DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA
RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIY
HLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS
GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY
DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL
VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP
WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL
SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV
VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNE
KLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEV
VKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMN
TKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLES
EFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVW
DKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF**V**SPTVA
YSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELEN
GRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISE
FSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK**Q**Y**R**STKEVL
DATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 40)

[0372] SEQ ID NO:40的残基V1134,Q1334和R1336(其可以从SEQ ID NO:40的D1134,R1334和T1336突变以产生SpVQR Cas9)是加下划线且粗体的。

[0373] 示例性SpVRER Cas9

[0374] DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA
RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIY
HLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS
GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY
DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL
VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP
WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL
SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV
VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNE
KLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEV
VKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMN
TKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLES
EFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVW
DKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF**V**SPTVA
YSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELEN
GRKRMLASA**R**ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISE
FSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK**E**Y**R**STKEVL
DATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 41)

[0375]

[0376] SEQ ID NO:41的残基V1134,R1217,Q1334和R1336(其可以从SEQ ID NO:41的D1134,G1217,R1334和T1336突变以产生SpVRER Cas9)是加下划线且粗体的。

[0377] 以下是示例性的融合蛋白(例如碱基编辑蛋白),其能够结合具有非规范(例如非NGG)PAM序列的核酸序列:

[0378] 示例性SaBE3(rAPOBEC1-XTEN-SaCas9n-UGI-NLS)

[0379] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIE KFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSG VTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFF TIALQSCHYQRLPPHILWATGLKSGSETPGTSESATPESKRNYILGLDIGITSVGYGIIDYETRDVIDAG VRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQ KLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEV RGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWY EMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTL KQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEE LTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEI PTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEE IIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVK QEEASKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRN LVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIAN ADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDK KPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLI MEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLS LKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKIN GELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSK KHPQIIKKGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLL TSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV (SEQ ID NO: 42)

[0380] 示例性SaKKH-BE3(rAPOBEC1-XTEN-SaCas9n-UGI-NLS)

[0381] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIE KFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIM TEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPP HILWATGLKSGSETPGTSESATPESKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRG ARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDT GNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDL LETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEK FQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILT IYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQ QKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEI IRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEEASKKG NRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLL RSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEK QAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIVNNLNGLYDKD NDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLN AHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIA SFYKNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPHIIKTIASKTQSIKKYSTDILGNLYE VKSKKHPQIIKKGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTS DAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV(SEQ ID NO:43)

[0382] 示例性EQR-BE3(rAPOBEC1-XTEN-Cas9n-UGI-NLS)MSSETGPVAVDPTLRRRIEPHEFEVF

FDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITE
FLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLY
VLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPGTSESATPESDKKYSIGLA
IGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIF
SNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMI
KFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF
GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITK
APLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELL
VKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMT
RKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSG
EQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILED
IVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFM
QLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTT
QKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF
LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLV
ETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALI
KKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVW
DKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFESPTVAYSVLVVAKVE
KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNEL
ALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDI
IEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKM
LSGGSPKKKRKV(SEQ ID NO:44)

[0383] VQR-BE3(rAPOBEC1-XTEN-Cas9n-UGI-NLS)MSSETGPVAVDPTLRRRIEPHEFEVFFDPREL
RKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYP
HVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYC
IILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSV
GWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAK
VDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHF
LIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIAL
SLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS
MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE
DLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET
ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAI
VDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLT
LFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKN
SRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSI
DNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQIT

KHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKL
ESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF
ATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKK
LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKY
VNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAEN
IIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETG
KQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSP
KKKRKV(SEQ ID NO:45)

[0384] VRER-BE3

[0385] (rAPOBEC1-XTEN-Cas9n-UGI-NLS)MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLL
YEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIY
IARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPP
CLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITD
EYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFH
RLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPN
FKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDE
HHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQR
TFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE
EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKT
NRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREM
IEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKED
IQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKR
IEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTR
SDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQIL
DSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYG
DYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKEL
LGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLA
SHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTL
TNLGAPAAFKYFDTTIDRKEYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQE
SILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV
(SEQ ID NO:46)

[0386] 高保真碱基编辑

[0387] 本公开内容的一些方面提供了包含具有高保真性的Cas9域的Cas9融合蛋白(例如,本文提供的任何融合蛋白)。本公开内容的其他方面提供了Cas9融合蛋白(例如,本文提供的任何融合蛋白),其包含与野生型Cas9域相比具有降低的Cas9域和DNA的糖-磷酸主链之间的相互作用静电相互作用的Cas9域。在一些实施方案中,Cas9域(例如,野生型Cas9域)包含一个或多个降低Cas9域和DNA的糖磷酸主链之间的缔合的突变。在一些实施方案中,本

文提供的任何Cas9融合蛋白包含SEQ ID NO:6中提供的氨基酸序列的N497X,R661X,Q695X和/或Q926X突变中的一个或多个,或任何Cas9蛋白,诸如如本文提供的任一种Cas9氨基酸序列中的相应突变,其中X是任意氨基酸。在一些实施方案中,本文提供的任何Cas9融合蛋白包含SEQ ID NO:6中提供的氨基酸序列的N497A,R661A,Q695A和/或Q926A突变中的一个或多个,或任何Cas9蛋白,诸如如本文提供的任一种Cas9氨基酸序列中的相应突变。在一些实施方案中,Cas9域包含SEQ ID NO:6中提供的氨基酸序列的D10A突变,或或任何Cas9蛋白,诸如如本文提供的任一种Cas9氨基酸序列中的相应突变。在一些实施方案中,Cas9域(例如,本文提供的任何融合蛋白的Cas9域)包含如SEQ ID NO:47所示的氨基酸序列。在一些实施方案中,融合蛋白包含如SEQ ID NO:48所示的氨基酸序列。具有高保真性的Cas9域是本领域已知的,并且对熟练技术人员而言是明显的。例如,具有高保真性的Cas9域已经记载于Kleinstiver,B.P.,et al.“High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects.”Nature 529,490-495(2016);和Slaymaker,I.M.,et al.“Rationally engineered Cas9 nucleases with improved specificity.”Science 351,84-88(2015);每篇的全部内容通过引用并入本文。

[0388] 应当理解,本文提供的碱基编辑器,例如碱基编辑器2(BE2)或碱基编辑器3(BE3)可以通过如本文所述修饰Cas9域以生成高保真性碱基编辑器,例如,高保真性碱基编辑器2(HF-BE2)或高保真性碱基编辑器3(HF-BE3)而转化为保真性碱基编辑器。在一些实施方案中,碱基编辑器2(BE2)包括脱氨酶域,dCas9和UGI域。在一些实施方案中,碱基编辑器3(BE3)包含脱氨酶域,Cas9域和UGI域。

[0389] Cas9域,其中相对于SEQ ID NO:6的Cas9的突变以粗体和下划线显示

[0390] DKKYSIGL**A**IGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF
GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD

NSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG
LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAK
NLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ
SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQ
IHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETI
TPWNFEEVVDKGASAQSFIERMT**A**FDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYV
TEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFN
ASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKV
MKQLKRRRYTGWG**A**LSRKLINGIRDKQSGKTILDFLKSDGFANRNFM**A**LIHDDSLTFK
EDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEM
[0391] ARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRD
MYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR**A**ITKHVAQIL
DSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESI
LPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITI
MERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNEL
ALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA
NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
ATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 47)

[0392] HF-BE3

[0393] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIE

KFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIM
TEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPP
HILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALL
FDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS
GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLA
QIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSK
NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPF
LKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTAFDKNLPNEKVL
PKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVE
DRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
ALSRKLINGIRDKQSGKTILDFLKSDGFANRNFMALIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKG
ILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLY
LYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLN
AKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRAITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVS
DFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS
NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDE
IIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
ATLIHQSITGLYETRIDLSQLGGD(SEQ ID NO:48)

[0394] Cas9融合蛋白

[0395] 本文公开的任何Cas9域(例如,核酸酶活性Cas9蛋白,无核酸酶活性的dCas9蛋白或Cas9切口酶蛋白)可以与第二蛋白融合,因此本文提供的融合蛋白包含如本文提供的Cas9域和第二蛋白质或"融合配偶体"。在一些实施方案中,第二蛋白质与Cas9域的N末端融合。然而,在其他实施方案中,第二蛋白质与Cas9域的C末端融合。在一些实施方案中,与Cas9域融合的第二蛋白质是核酸编辑域。在一些实施方案中,Cas9域和核酸编辑域通过接头融合,而在其他实施方案中,Cas9域和核酸编辑域彼此直接融合。在一些实施方案中,Cas9域和核酸编辑域通过任何长度或组成的接头融合。例如,接头可以是键,一个或多个氨基酸,任何长度和组成的聚合物或肽。在一些实施方案中,接头包含$(GGGS)_n$(SEQ ID NO:613),$(GGGGS)_n$(SEQ ID NO:607),$(G)_n$(SEQ ID NO:608),$(EAAAK)_n$(SEQ ID NO:609),$(GGS)_n$(SEQ ID NO:610),$(SGGS)_n$(SEQ ID NO:606),SGSETPGTSESATPES(SEQ ID NO:604),SGGS$(GGS)_n$(SEQ ID NO:612),SGGSSGGSSGSETPGTSESATPESSGGSSGGS(SEQ ID NO:605),或$(XP)_n$(SEQ ID NO:611)基序,或这些中的任意组合,其中n独立地是1和30之间的整数,并且X是任意氨基酸。在一些实施方案中,接头包含$(GGS)_n$基序,其中n为1、3或7。在一些实施方案中,接头包含$(GGS)_n$(SEQ ID NO:610)基序,其中n为1,2,3,4,5,6,7,8,9,10,11,12,13,14或15。在一些实施方案中,接头包含氨基酸序列SGGS$(GGS)_n$(SEQ ID NO:612),其中n为1、2、3、4、5、6、7、8、9或10.在一些实施方案中,接头包含氨基酸序列SGGS$(GGS)_n$(SEQ ID NO:612),其中n为2。在一些实施方案中,接头包含SGSETPGTSESATPES的氨基酸序列(SEQ ID

NO:604)，在实施例中也称为XTEN接头。在一些实施方案中，接头包含氨基酸序列SGGSSGGSSGSETPGTSESATPESSGGSSGGS(SEQ ID NO:605)，在实施例中也称为32个氨基酸的接头。接头的长度可以影响待编辑的碱基，如实施例中显示。例如，长3个氨基酸的接头(例如$(GGS)_1$)相对于PAM序列可以给出2-5、2-4、2-3、3-4个碱基的编辑窗，而9个氨基酸的接头(例如$(GGS)_3$(SEQ ID NO:610))可以给出相对于PAM序列的2-6,2-5,2-4,2-3,3-6,3-5,3-4,4-6,4-5,5-6个碱基的编辑窗。16个氨基酸的接头(例如，XTEN接头)可以给出相对于具有异常强活性的PAM序列的2-7,2-6,2-5,2-4,2-3,3-7,3-6,3-5,3-4,4-7,4-6,4-5,5-7,5-6,6-7个碱基的窗，并且21个氨基酸的接头(例如，$(GGS)_7$(SEQ ID NO:610))可以给出相对于PAM序列的3-8,3-7,3-6,3-5,3-4,4-8,4-7,4-6,4-5,5-8,5-7,5-6,6-8,6-7,7-8个碱基的编辑窗。改变接头长度可以使本公开内容的dCas9融合蛋白编辑距PAM序列不同距离的核碱基的新发现提供了重要的临床意义，因为PAM序列可以与基因中要校正的致病突变具有不同距离。应当理解，本文作为实例描述的接头长度并不意味着是限制性的。

[0396] 在一些实施方案中，第二蛋白质包含酶域。在一些实施方案中，酶域是核酸编辑域。此类核酸编辑域可以是但不限于核酸酶，切口酶，重组酶，脱氨酶，甲基转移酶，甲基化酶，乙酰基酶或乙酰基转移酶。可以根据本公开内容使用的非限制性示例性结合域包含转录活化物域和转录阻遏物域。

[0397] 脱氨酶域

[0398] 在一些实施方案中，第二蛋白质包含核酸编辑域。在一些实施方案中，核酸编辑域可以催化从C到U的碱基变化。在一些实施方案中，核酸编辑域是脱氨酶域。在一些实施方案中，脱氨酶是胞苷脱氨酶或胞苷脱氨酶。在一些实施方案中，脱氨酶是载脂蛋白B mRNA编辑复合物(APOBEC)家族脱氨酶。在一些实施方案中，脱氨酶是APOBEC1脱氨酶。在一些实施方案中，脱氨酶是APOBEC2脱氨酶。在一些实施方案中，脱氨酶是APOBEC3脱氨酶。在一些实施方案中，脱氨酶是APOBEC3A脱氨酶。在一些实施方案中，脱氨酶是APOBEC3B脱氨酶。在一些实施方案中，脱氨酶是APOBEC3C脱氨酶。在一些实施方案中，脱氨酶是APOBEC3D脱氨酶。在一些实施方案中，脱氨酶是APOBEC3E脱氨酶。在一些实施方案中，脱氨酶是APOBEC3F脱氨酶。在一些实施方案中，脱氨酶是APOBEC3G脱氨酶。在一些实施方案中，脱氨酶是APOBEC3H脱氨酶。在一些实施方案中，脱氨酶是APOBEC4脱氨酶。在一些实施方案中，脱氨酶是活化诱导的脱氨酶(AID)。在一些实施方案中，脱氨酶是脊椎动物脱氨酶。在一些实施方案中，脱氨酶是无脊椎动物脱氨酶。在一些实施方案中，脱氨酶是人，黑猩猩，大猩猩，猴，牛，狗，大鼠或小鼠脱氨酶。在一些实施方案中，脱氨酶是人脱氨酶。在一些实施方案中，脱氨酶是大鼠脱氨酶，例如rAPOBEC1。在一些实施方案中，脱氨酶是活化诱导的胞苷脱氨酶(AID)。在一些实施方案中，脱氨酶是胞苷脱氨酶1(CDA1)。在一些实施方案中，脱氨酶是海七鳃鳗(Petromyzon marinus)胞苷脱氨酶1(pmCDA1)。在一些实施方案中，脱氨酶是人APOBEC3G(SEQ ID NO:60)。在一些实施方案中，脱氨酶是人APOBEC3G的片段(SEQ ID NO:83)。在一些实施方案中，脱氨酶是包含D316R_D317R突变的人APOBEC3G变体(SEQ ID NO:82)。在一些实施方案中，脱氨酶是人APOBEC3G的，并且包含对应于SEQ ID NO:60中的D316R_D317R突变的突变的片段(SEQ ID NO:84)。

[0399] 在一些实施方案中，核酸编辑域与SEQ ID NO:49-84中任一项的脱氨酶域是至少80％，至少85％，至少90％，至少92％，至少95％，至少96％，至少97％，至少98％，至少99％

或至少99.5％相同的。在一些实施方案中，核酸编辑域包含SEQ ID NO:49-84中任一项的氨基酸序列。

[0400] 调节碱基编辑器的编辑窗的脱氨酶域

[0401] 本公开内容的一些方面是基于以下认识：例如通过在脱氨酶域中进行点突变来调节本文提供的任何融合蛋白的脱氨酶域催化活性影响融合蛋白(例如，碱基编辑器)的持续合成能力。例如，减少但不消除碱基编辑融合蛋白内脱氨酶域催化活性的突变可以使得脱氨酶域将催化靶残基附近的残基脱氨基变得不太可能的，从而缩窄脱氨基窗。缩窄脱氨基窗的能力可以防止特定靶残基附近的残基的不希望的脱氨基，这可以减少或防止脱靶效应。

[0402] 在一些实施方案中，本文提供的任何融合蛋白包含具有降低的催化脱氨酶活性的脱氨酶域(例如，胞苷脱氨酶域)。在一些实施方案中，本文提供的任何融合蛋白包含脱氨酶域(例如，胞苷脱氨酶域)，其与合适的对照相比具有降低的催化脱氨酶活性。例如，合适的对照可以是在将一个或多个突变引入脱氨酶之前的脱氨酶的脱氨酶活性。在其他实施方案中，合适的对照可以是野生型脱氨酶。在一些实施方案中，合适的对照是野生型载脂蛋白BmRNA编辑复合物(APOBEC)家族脱氨酶。在一些实施方案中，合适的对照是APOBEC1脱氨酶，APOBEC2脱氨酶，APOBEC3A脱氨酶，APOBEC3B脱氨酶，APOBEC3C脱氨酶，APOBEC3D脱氨酶，APOBEC3F脱氨酶，APOBEC3G脱氨酶或APOBEC3H脱氨酶。在一些实施方案中，合适的对照是活化诱导的脱氨酶(AID)。在一些实施方案中，合适的对照是来自海七鳃鳗的胞苷脱氨酶1(pmCDA1)。在一些实施方案中，脱氨酶域可以是具有与合适的对照相比小至少1％，至少5％，至少15％，至少20％，至少25％，至少30％，至少40％，至少50％，至少60％，至少70％，至少80％，至少90％或至少95％的催化脱氨酶活性。

[0403] 在一些实施方案中，本文提供的任何融合蛋白包含APOBEC脱氨酶，其包含一个或多个选自下组的突变：rAPOBEC1(SEQ ID NO:76)的H121X，H122X，R126X，R126X，R118X，W90X，W90X和R132X的突变，或另一种APOBEC脱氨酶中的一个或多个相应突变，其中X是任意氨基酸。在一些实施方案中，本文提供的任何融合蛋白包含APOBEC脱氨酶，其包含一个或多个选自下组的突变：rAPOBEC1(SEQ ID NO:76)的H121R，H122R，R126A，R126E，R118A，W90A，W90Y和R132E，或另一种APOBEC脱氨酶中的一个或多个相应突变。

[0404] 在一些实施方案中，本文提供的任何融合蛋白包含APOBEC脱氨酶，其包含一个或多个选自下组的突变：hAPOBEC3G(SEQ ID NO:60)的D316X，D317X，R320X，R320X，R313X，W285X，W285X，R326X，或另一种APOBEC脱氨酶中的一个或多个相应突变，其中X是任意氨基酸。在一些实施方案中，本文提供的任何融合蛋白包含APOBEC脱氨酶，其包含一个或多个选自下组的突变：hAPOBEC3G(SEQ ID NO:60)的D316R，D317R，R320A，R320E，R313A，W285A，W285Y，R326E，或另一种APOBEC脱氨酶中的一个或多个相应突变。

[0405] 在一些实施方案中，本文提供的任何融合蛋白包含APOBEC脱氨酶，其包含rAPOBEC1(SEQ ID NO:76)的H121R和H122R突变，或另一种APOBEC脱氨酶中的一个或多个相应突变。在一些实施方案中，本文提供的任何融合蛋白包含APOBEC脱氨酶，其包含rAPOBEC1(SEQ ID NO:76)的R126A突变，或另一种APOBEC脱氨酶中的一个或多个相应突变。在一些实施方案中，本文提供的任何融合蛋白包含APOBEC脱氨酶，其包含rAPOBEC1(SEQ ID NO:76)的R126E突变，或另一种APOBEC脱氨酶中的一个或多个相应突变。在一些实施方案中，本文

提供的任何融合蛋白包含APOBEC脱氨酶,其包含rAPOBEC1(SEQ ID NO:76)的R118A突变,或另一种APOBEC脱氨酶中的一个或多个相应突变。在一些实施方案中,本文提供的任何融合蛋白包含APOBEC脱氨酶,其包含rAPOBEC1(SEQ ID NO:76)的W90A突变,或另一种APOBEC脱氨酶中的一个或多个相应突变。在一些实施方案中,本文提供的任何融合蛋白包含APOBEC脱氨酶,其包含rAPOBEC1(SEQ ID NO:76)的W90Y突变,或另一种APOBEC脱氨酶中的一个或多个相应突变。在一些实施方案中,本文提供的任何融合蛋白包含APOBEC脱氨酶,其包含rAPOBEC1(SEQ ID NO:76)的R132E突变,或另一种APOBEC脱氨酶中的一个或多个相应突变。在一些实施方案中,本文提供的任何融合蛋白包含APOBEC脱氨酶,其包含rAPOBEC1(SEQ ID NO:76)的W90Y和R126E突变,或另一种APOBEC脱氨酶中的一个或多个相应突变。在一些实施方案中,本文提供的任何融合蛋白包含APOBEC脱氨酶,其包含rAPOBEC1的R126E和R132E突变(SEQ ID NO:76),或另一种APOBEC脱氨酶中的一个或多个相应突变。本文提供的融合蛋白包含APOBEC脱氨酶,其包含rAPOBEC1(SEQ ID NO:76)的W90Y和R132E突变,或另一种APOBEC脱氨酶中的一个或多个相应突变。在一些实施方案中,本文提供的任何融合蛋白包含APOBEC脱氨酶,其包含rAPOBEC1(SEQ ID NO:76)的W90Y,R126E和R132E突变,或另一种APOBEC脱氨酶中的一个或多个相应突变。

[0406] 在一些实施方案中,本文提供的任何融合蛋白包含APOBEC脱氨酶,其包含hAPOBEC3G(SEQ ID NO:60)的D316R和D317R突变,或另一种APOBEC脱氨酶中的一个或多个相应突变。在一些实施方案中,本文提供的任何融合蛋白包含APOBEC脱氨酶,其包含hAPOBEC3G(SEQ ID NO:60)的R320A突变,或另一种APOBEC脱氨酶中的一个或多个相应突变。在一些实施方案中,本文提供的任何融合蛋白包含APOBEC脱氨酶,其包含hAPOBEC3G(SEQ ID NO:60)的R320E突变,或另一种APOBEC脱氨酶中的一个或多个相应突变。在一些实施方案中,本文提供的任何融合蛋白包含APOBEC脱氨酶,其包含hAPOBEC3G(SEQ ID NO:60)的R313A突变,或另一种APOBEC脱氨酶中的一个或多个相应突变。在一些实施方案中,本文提供的任何融合蛋白包含APOBEC脱氨酶,其包含hAPOBEC3G(SEQ ID NO:60)的W285A突变,或另一种APOBEC脱氨酶中的一个或多个相应突变。在一些实施方案中,本文提供的任何融合蛋白包含包含hAPOBEC3G(SEQ ID NO:60)的W285Y突变或另一种APOBEC脱氨酶中的一个或多个相应突变。在一些实施方案中,本文提供的任何融合蛋白包含APOBEC脱氨酶,其包含hAPOBEC3G(SEQ ID NO:60)的R326E突变,或另一种APOBEC脱氨酶中的一个或多个相应突变。在一些实施方案中,本文提供的任何融合蛋白均包含APOBEC脱氨酶,其包含hAPOBEC3G(SEQ ID NO:60)的W285Y和R320E突变,或另一种APOBEC脱氨酶中的一个或多个相应突变。在一些实施方案中,本文提供的任何融合蛋白包含APOBEC脱氨酶,其包含hAPOBEC3G(SEQ ID NO:60)的R320E和R326E突变,或另一种APOBEC脱氨酶中的一个或多个相应突变。本文提供的任何融合蛋白包含APOBEC脱氨酶,其包含hAPOBEC3G(SEQ ID NO:60)的W285Y和R326E突变,或另一种APOBEC脱氨酶中的一个或多个相应突变。在一些实施方案中,本文提供的任何融合蛋白包含APOBEC脱氨酶,其包含hAPOBEC3G(SEQ ID NO:60)的W285Y,R320E和R326E突变,或另一种APOBEC脱氨酶中的一个或多个相应突变。

[0407] 本公开内容的一些方面提供了融合蛋白,其包含(i)无核酸酶活性的Cas9域;和(ii)核酸编辑域。在一些实施方案中,无核酸酶活性的Cas9域(dCas9)包含与如由任何Cas9蛋白提供的Cas9氨基酸序列,诸如如本文提供的任一种Cas9氨基酸序列至少80%,至少

85％，至少90％，至少92％，至少95％，至少96％，至少97％，至少98％，至少99％或至少99.5％相同的氨基酸序列，并且包含使Cas9的核酸酶活性失活的突变。使Cas9的核酸酶域变为无活性的突变是本领域中公知的。例如，已知Cas9的DNA切割域包括两个亚域，即HNH核酸酶亚域和RuvC1亚域。HNH亚域切割与gRNA互补的链，而RuvC1亚域切割非互补链。这些亚域内的突变可以使Cas9的核酸酶活性沉默。例如，突变D10A和H840A使酿脓链球菌Cas9的核酸酶活性完全失活(Jinek et al.,Science.337:816-821(2012);Qi et al.,Cell.28;152(5):1173-83(2013))。在一些实施方案中，本公开内容的dCas9包含SEQ ID NO:6中提供的氨基酸序列的D10A突变，或任何Cas9蛋白，诸如如本文提供的任一种Cas9氨基酸序列中的相应突变。在一些实施方案中，本公开内容的dCas9包含SEQ ID NO:6中提供的氨基酸序列的H840A突变，或任何Cas9蛋白，诸如如本文提供的任一种Cas9氨基酸序列中的相应突变。在一些实施方案中，本公开内容的dCas9包含SEQ ID NO:6中提供的氨基酸序列的D10A和H840A突变两者，或任何Cas9蛋白，诸如如本文提供的任一种Cas9氨基酸序列中的相应突变。在一些实施方案中，Cas9还包含在SEQ ID NO:6中提供的氨基酸序列的位置840处的组氨酸残基，或任何Cas9蛋白，诸如如本文提供的任一种Cas9氨基酸序列中的相应突变。催化残基H840的存在恢复Cas9切割含有与靶定C相对的G的非编辑链的能力。H840的恢复不导致含有C的靶链的切割。dCas9包含SEQ IDNO:32的氨基酸序列。应当理解，使Cas9的核酸酶域失活的其他突变也可以包括在本公开内容的dCas9中。

[0408] 包含本文公开的突变的Cas9或dCas9域可以是全长Cas9或其片段。在一些实施方案中，包含Cas9或其片段的蛋白质称为“Cas9变体”。Cas9变体与Cas9或其片段共享同源性。例如，Cas9变体与野生型Cas9是至少约70％相同，至少约80％相同，至少约90％相同，至少约95％相同，至少约96％相同，至少约97％相同，至少约98％相同，至少约99％相同，至少约99.5％相同或至少约99.9％。在一些实施方案中，Cas9变体包含Cas9的片段(例如，gRNA结合域或DNA切割域)，使得该片段与野生型Cas9的相应片段，例如包含SEQ ID NO:6的氨基酸序列的Cas9至少约70％相同，至少约80％相同，至少约90％相同，至少约95％相同，至少约96％相同，至少约97％相同，至少约98％相同，至少约99％相同，至少约99.5％相同或至少约99.9％相同。

[0409] 本公开内容的任何Cas9融合蛋白可进一步包含核酸编辑域(例如，能够修饰核酸的酶，例如脱氨酶)。在一些实施方案中，核酸编辑域是DNA编辑域。在一些实施方案中，核酸编辑域具有脱氨酶活性。在一些实施方案中，核酸编辑域包含脱氨酶或脱氨酶域或由脱氨酶或脱氨酶域组成。在一些实施方案中，脱氨酶是胞苷脱氨酶。在一些实施方案中，脱氨酶是载脂蛋白BmRNA编辑复合物(APOBEC)家族脱氨酶。在一些实施方案中，脱氨酶是APOBEC1家族脱氨酶。在一些实施方案中，脱氨酶是活化诱导的胞苷脱氨酶(AID)。本文详细描述了一些核酸编辑域以及包括此类域的Cas9融合蛋白。基于本公开内容和本领域的知识，其他合适的核酸编辑域对熟练技术人员将是明显的。

[0410] 本公开内容的一些方面提供了融合蛋白，其包含与核酸编辑域融合的Cas9域，其中核酸编辑域与Cas9域的N末端融合。在一些实施方案中，Cas9域和核酸编辑-编辑域通过接头融合。在一些实施方案中，接头包含$(GGGS)_n$(SEQ ID NO:613)，$(GGGGS)_n$(SEQ ID NO:607)，$(G)_n$(SEQ ID NO:608)，an$(EAAAK)_n$(SEQ ID NO:609)，$(GGS)_n$(SEQ ID NO:610)，$(SGGS)_n$(SEQ ID NO:606)，SGSETPGTSESATPES(SEQ ID NO:604)基序(参见例如Guilinger

JP,Thompson DB,Liu DR.Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification.Nat.Biotechnol.2014;32(6):577-82;全部内容通过引用并入本文),或(XP)$_n$(SEQ ID NO:611)基序或这些的任意组合,其中n独立地是1和30之间的整数。在一些实施方案中,n独立地是1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29或30,或者,若存在超过一个接头或超过一个接头基序,则其任意组合。在一些实施方案中,接头包含(GGS)$_n$(SEQ ID NO:610)基序,其中n是1,2,3,4,5,6,7,8,9,10,11,12,13,14或15。在一些实施方案中,接头包含(GGS)$_n$(SEQ ID NO:610)基序,其中n为1、3或7。在一些实施方案中,接头包含氨基酸序列SGSETPGTSESATPES(SEQ ID NO:604)。其他合适的接头基序和接头构造对本领域技术人员将是明显的。在一些实施方案中,合适的接头基序和构造包括Chen et al.,Fusion protein linkers:property,design and functionality.Adv Drug Deliv Rev.2013;65(10):1357-69(其全部内容通过引用并入本文)中描述的那些。基于本公开内容,其他合适的接头序列对本领域技术人员将是明显的。在一些实施方案中,本文提供的示例性Cas9融合蛋白的一般结构包括以下结构:

[0411] [NH_2]-[核酸编辑域]-[Cas9]-[COOH]或

[0412] [NH_2]-[核酸编辑域]-[接头]-[Cas9]-[COOH],

[0413] 其中NH_2是融合蛋白的N末端,而COOH是融合蛋白的C末端。

[0414] 本公开内容的融合蛋白可以包含一个或多个额外的特征。例如,在一些实施方案中,融合蛋白包含核定位序列(NLS)。在一些实施方案中,融合蛋白的NLS位于核酸编辑域和Cas9域之间。在一些实施方案中,融合蛋白的NLS位于Cas9域的C末端。

[0415] 可能存在的其他示例性特征是定位序列,例如胞质定位序列,输出序列,例如核输出序列或其他定位序列,以及可用于增溶,纯化或检测其的序列标签。本文提供的合适的蛋白质标签包括但不限于生物素羧化酶载体蛋白(BCCP)标签,myc标签,钙调蛋白标签,FLAG标签,血凝素(HA)标签,聚组氨酸标签,也称为作为组氨酸标签或His标签,麦芽糖结合蛋白(MBP)标签,nus标签,谷胱甘肽S-转移酶(GST)标签,绿色荧光蛋白(GFP)标签,硫氧还蛋白标签,S标签,Softags(例如,Softag 1,Softag 3),strep标签,生物素连接酶标签,FlAsH标签,V5标签和SBP标签。对于本领域技术人员而言,其他合适的序列将是明显的。在一些实施方案中,融合蛋白包含一个或多个His标签。

[0416] 在一些实施方案中,核酸编辑域是脱氨酶。例如,在一些实施方案中,具有脱氨酶域的示例性Cas9融合蛋白的一般结构包括以下结构:

[0417] [NH_2]-[NLS]-[脱氨酶]-[Cas9]-[COOH],

[0418] [NH_2]-[Cas9]-[脱氨酶]-[COOH],

[0419] [NH_2]-[脱氨酶]-[Cas9]-[COOH],或

[0420] [NH_2]-[脱氨酶]-[Cas9]-[NLS]-[COOH];

[0421] 其中,NLS是核定位序列,NH_2是融合蛋白的N末端,并且COOH是融合蛋白的C末端。核定位序列是本领域已知的,并且对熟练技术人员而言是明显的。例如,NLS序列记载于Plank et al.,PCT/EP2000/011690,对于其对示例性核定位序列的公开,其内容通过引用并入本文。在一些实施方案中,NLS包含氨基酸序列PKKKRKV(SEQ ID NO:614)或MDSLLMNRRKFLYQFKNVRWAKGRRETYLC(SEQ ID NO:615)。在一些实施方案中,在Cas9和脱氨酶

之间插入接头。在一些实施方案中,NLS位于Cas9域的C末端。在一些实施方案中,NLS位于Cas9域的N末端。在一些实施方案中,NLS位于脱氨酶和Cas9域之间。在一些实施方案中,NLS位于脱氨酶域的N末端。在一些实施方案中,NLS位于脱氨酶域的C-末端。

[0422] 一种示例性的合适的核酸编辑域类型是例如APOBEC家族的胞苷脱氨酶。胞苷脱氨酶的载脂蛋白B mRNA编辑复合物(APOBEC)家族涵盖11种蛋白质,所述蛋白质以受控且有益的方式启动诱变[29]。一种家族成员,活化诱导的胞苷脱氨酶(AID)通过以转录依赖性链偏爱方式将ssDNA中的胞嘧啶转化为尿嘧啶负责抗体的成熟[30]。载脂蛋白B编辑复合物3(APOBEC3)酶通过逆转录病毒ssDNA中的胞嘧啶脱氨基针对某种HIV-1毒株对人细胞提供保护[31]。这些蛋白都需要Zn^{2+}配位基序(His-X-Glu-X_{23-26}-Pro-Cys-X_{2-4}-Cys;SEQ ID NO:616)和结合水分子以发挥催化活性。Glu残基的作用是将水分子活化为氢氧化锌,从而在脱氨基反应中发生亲核攻击。每个家族成员优先在其自身的特定"热点"处脱氨基,从用于hAID的WRC(W是A或T,R是A或G)变化至用于hAPOBEC3F的TTC[32]。APOBEC3G催化域的最新晶体结构揭示了由侧翼有六个α-螺旋的五股β-片层核心构成的二级结构,认为其在整个家族中是保守的[33]。已经显示了活性中心环负责ssDNA结合和确定"热点"身份[34]。这些酶的过表达与基因组不稳定性和癌症有关,因此突出显示了序列特异性靶向的重要性[35]。

[0423] 本公开内容的一些方面涉及以下认识:胞苷脱氨酶的活性如APOBEC酶可以导向基因组DNA中的特定位点。不希望受到任何特定理论的束缚,使用Cas9作为识别剂的优点包括:(1)可以通过简单地改变sgRNA序列容易地改变Cas9的序列特异性;并且(2)Cas9结合其靶序列,通过使dsDNA变性,从而形成单链的,因此作为脱氨酶的可行底物的DNA段。应当理解,其他催化域或来自其他脱氨酶的催化域也可以用于产生具有Cas9的融合蛋白,并且本公开内容不限于此。

[0424] 本公开内容的一些方面是基于以下认识:Cas9:脱氨酶融合蛋白可以根据图3中的编号方案有效地使位置3-11处的核苷酸脱氨基。鉴于关于可以被Cas9:脱氨酶融合蛋白靶向的核苷酸的本文提供的结果,本领域技术人员将能够设计合适的引导RNA,以将融合蛋白靶向到包含要脱氨基的核苷酸的靶序列。

[0425] 在一些实施方案中,脱氨酶域和Cas9域通过接头彼此融合。可以采用脱氨酶域(例如,AID)和Cas9域之间的各种接头长度和灵活性(例如,范围为形式$(GGGGS)_n$(SEQ ID NO:607),$(GGS)_n$(SEQ ID NO:610)和$(G)_n$(SEQ ID NO:608)的非常柔性接头至形式$(EAAAK)_n$(SEQ ID NO:609),$(SGGS)_n$(SEQ ID NO:606),SGSETPGTSESATPES(SEQ ID NO:604)(参见例如Guilinger JP,Thompson DB,Liu DR.Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification.Nat.Biotechnol.2014;32(6):577-82;全部内容通过引用并入本文)和$(XP)_n$(SEQ ID NO:611)的更刚性接头)[36]以实现脱氨酶活性对于特定应用的最佳长度。在一些实施方案中,接头包含$(GGS)_n$(SEQ ID NO:610)基序,其中n是1、3或7。在一些实施方案中,接头包含SGSETPGTSESATPES(SEQ ID NO:604)基序。

[0426] 下面提供了一些示例性的合适的核酸编辑域,例如脱氨酶和脱氨酶域,其可以根据本公开内容的方面与Cas9域融合。应当理解,在一些实施方案中,可以使用相应序列的活性域,例如没有定位信号的域(核定位序列,没有核输出信号,胞质定位信号)。

[0427] 人AID:

[0428] MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL (SEQ ID NO: 49)

[0429] (下划线:核定位序列;双下划线:核输出序列;)

[0430] 小鼠AID:

[0431] MDSLLMKQKKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSCSLDFGHLRNKSGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVAEFLRWNPNLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIGIMTFKDYFYCWNTFVENRERTFKAWEGLHENSVRLTRQLRRILLPLYEVDDLRDAFRMLGF (SEQ ID NO: 51)

[0432] (下划线:核定位序列;双下划线:核输出序列;)

[0433] 狗AID:

[0434] MDSLLMKQRKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLDFGHLRNKSGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFAARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENREKTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL (SEQ ID NO: 52)

[0435] (下划线:核定位序列;双下划线:核输出序列;)

[0436] 牛AID:

[0437] MDSLLKKQRQFLYQFKNVRWAKGRHETYLCYVVKRRDSPTSFSLDFGHLRNKAGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFTARLYFCDKERKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL (SEQ ID NO: 53)

[0438] (下划线:核定位序列;双下划线:核输出序列;)

[0439] 大鼠:AID:

[0440] MAVGSKPKAALVGPHWERERIWCFLCSTGLGTQQTGQTSRWLRPAATQDPVSPPRSLLMKQRKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLDFGYLRNKSGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLTGWGALPAGLMSPARPSDYFYCWNTFVENHERTFKAWEGLHENSVRLSRRLRRILLPLYEVDDLRDAFRTLGL (SEQ ID NO: 54)

[0441] (下划线:核定位序列;双下划线:核输出序列;)

[0442] 小鼠APOBEC-3:

[0443] MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLGYAKGRKDTFLCYEVTR KDCDSPVSLHHGVFKNKDNI*HAEICFLYWFHDKVLKVLSPREEFKITWYMS WSPCFEC*AEQIVRFLATHHNLSLDIFSSRLYNVQDPETQQNLCRLVQEGAQ VAAMDLYEFKKCWKKFVDNGGRRFRPWKRLLTNFRYQDSKLQEILRPCY IPVPSSSSSTLSNICLTKGLPETRFCVEGRRMDPLSEEEFYSQFYNQRVKHL CYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKGKQ*HAEILFLDKIRSMELS QVTITCYLTWSPCPNC*AWQLAAFKRDRPDLILHIYTSRLYFHWKRPFQKGL CSLWQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLR RIKESWGLQDLVNDFGNLQLGPPMS (SEQ ID NO: 55)

[0444] (斜体字:核酸编辑域)

[0445] 大鼠APOBEC-3:

[0446] MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLRYAIDRKDTFLCYEVTRK DCDSPVSLHHGVFKNKDNI*HAEICFLYWFHDKVLKVLSPREEFKITWYMSW SPCFEC*AEQVLRFLATHHNLSLDIFSSRLYNIRDPENQQNLCRLVQEGAQV AAMDLYEFKKCWKKFVDNGGRRFRPWKKLLTNFRYQDSKLQEILRPCYI PVPSSSSSTLSNICLTKGLPETRFCVERRRVHLLSEEEFYSQFYNQRVKHLC YYHGVKPYLCYQLEQFNGQAPLKGCLLSEKGKQ*HAEILFLDKIRSMELSQ VIITCYLTWSPCPNC*AWQLAAFKRDRPDLILHIYTSRLYFHWKRPFQKGLC SLWQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLHR IKESWGLQDLVNDFGNLQLGPPMS (SEQ ID NO: 56)

[0447] (斜体字:核酸编辑域)

[0448] 猕猴APOBEC-3G:

[0449] MVEPMDPRTFVSNFNNRPILSGLNTVWLCCEVKTKDPSGPPLDAKIFQGK VYSKAKY*HPEM**RFLRWFHKWRQLHHDQEYKVTWYVSWSPCTRC*ANSVATF LAKDPKVTLTIFVARLYYFWKPDYQQALRILCQKRGGPHATMKIMNYNE FQDCWNKFVDGRGKPFKPRNNLPKHYTLLQATLGELLRHLMDPGTFTSN FNNKPWVSGQHETYLCYKVERLHNDTWVPLNQHRGFLRNQAPNIHGFP KGR*HAELCFLDLIPFWKLDGQQYRVTCFTSWSPCFSC*AQEMAKFISNNEHV SLCIFAARIYDDQGRYQEGLRALHRDGAKIAMMNYSEFEYCWDTFVDRQ GRPFQPWDGLDEHSQALSGRLRAI (SEQ ID NO: 57)

[0450] (斜体字:核酸编辑域;下划线:胞质定位信号)

[0451] 黑猩猩APOBEC-3G:

[0452] MKPHFRNPVERMYQDTFSDNFYNRPILSHRNTVWLCYEVKTKGPSRPPL DAKIFRGQVYSKLKY*HPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTK C*TRDVATFLAEDPKVTLTIFVARLYYFWDPDYQEALRSLCQKRDGPRATM KIMNYDEFQHCWSKFVYSQRELFEPWNNLPKYYILLHIMLGEILRHSMDP PTFTSNFNNELWVRGRHETYLCYEVERLHNDTWVLLNQRRGFLCNQAP HKHGFLEGR*HAELCFLDVIPFWKLDLHQDYRVTCFTSWSPCFSC*AQEMAKF ISNNKHVSLCIFAARIYDDQGRCQEGLRTLAKAGAKISIMTYSEFKHCWD TFVDHQGCPFQPWDGLEEHSQALSGRLRAILQNQGN (SEQ ID NO: 58)

[0453] (斜体字:核酸编辑域;下划线:胞质定位信号)

[0454] 绿猴APOBEC-3G:

[0455] MNPQIRNMVEQMEPDIFVYYFNNRPILSGRNTVWLCYEVKTKDPSGPPL DANIFQGKLYPEAKD*HPEMKFLHWFRKWRQLHRDQEYEVTWYVSWSPCTR CANSVATFLAEDPKVTLTIFVARLYYFWKPDYQQALRILCQERGGPHATM KIMNYNEFQHCWNEFVDGQGKPFKPRKNLPKHYTLLHATLGELLRHVM DPGTFTSNFNNKPWVSGQRETYLCYKVERSHNDTWVLLNQHRGFLRNQ APDRHGFPKGR*HAELCFLDLIPFWKLDDQQYRVTCFTSWSPCFSC*AQKMA KFISNNKHVSLCIFAARIYDDQGRCQEGLRTLHRDGAKIAVMNYSEFEYC WDTFVDRQGRPFQPWDGLDEHSQALSGRLRAI (SEQ ID NO: 59)

[0456] (斜体字:核酸编辑域;下划线:胞质定位信号)

[0457] 人APOBEC-3G:

[0458] MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPL DAKIFRGQVYSELKY*HPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTK C*TRDMATFLAEDPKVTLTIFVARLYYFWDPDYQEALRSLCQKRDGPRAT MKIMNYDEFQHCWSKFVYSQRELFEPWNNLPKYYILLHIMLGEILRHSM DPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQ APHKHGFLEGR*HAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSC*AQEMA KFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTYSEFKHCW DTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQNQEN (SEQ ID NO: 60)

[0459] (斜体字:核酸编辑域;下划线:胞质定位信号)

[0460] 人APOBEC-3F:

[0461] MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPRL DAKIFRGQVYSQPEH*HAEMCFLSWFCGNQLPAYKCFQITWFVSWTPCPDC*V AKLAEFLAEHPNVTLTISAARLYYYWERDYRRALCRLSQAGARVKIMDD

[0462] EEFAYCWENFVYSEGQPFMPWYKFDDNYAFLHRTLKEILRNPMEAMYPH IFYFHFKNLRKAYGRNESWLCFTMEVVKHHSPVSWKRGVFRNQVDPET HC*HAERCFLSWFCDDILSPNTNYEVTWYTSWSPCPEC*AGEVAEFLARHSNV NLTIFTARLYYFWDTDYQEGLRSLSQEGASVEIMGYKDFKYCWENFVYN DDEPFKPWKGLKYNFLFLDSKLQEILE (SEQ ID NO: 61)

[0463] (斜体字:核酸编辑域)

[0464] 人APOBEC-3B:

[0465] MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLL WDTGVFRGQVYFKPQY*HAEMCFLSWFCGNQLPAYKCFQITWFVSWTPCPD C*VAKLAEFLSEHPNVTLTISAARLYYYWERDYRRALCRLSQAGARVTIMD YEEFAYCWENFVYNEGQQFMPWYKFDENYAFLHRTLKEILRYLMDPDTF TFNFNNDPLVLRRRQTYLCYEVERLDNGTWVLMDQHMGFLCNEAKNLL CGFY*GRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGC*AGEVRAFLQ ENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFEYCWDT FVYRQGCPFQPWDGLEEHSQALSGRLRAILQNQGN (SEQ ID NO: 62)

[0466] (斜体字:核酸编辑域)

[0467] 大鼠APOBEC-3B:

[0468] MQPQGLGPNAGMGPVCLGCSHRRPYSPIRNPLKKLYQQTFYFHFKNVRYAWGRKNNFLCYEVNGMDCA LPVPLRQGVFRKQGHIHAELCFIYWFHDKVLRVLSPMEEFKVTWYMSWSPCSKCAEQVARFLAAHRNLSLAIFSSR LYYYLRNPNYQQKLCRLIQEGVHVAAMDLPEFKKCWNKFVDNDGQPFRPWMRLRINFSFYDCKLQEIFSRMNLLRE

DVFYLQFNNSHRVKPVQNRYYRRKSYLCYQLERANGQEPLKGYLLYKKGEQHVEILFLEKMRSMELSQVRITCYLT WSPCPNCARQLAAFKKDHPDLILRIYTSRLYFYWRKKFQKGLCTLWRSGIHVDVMDLPQFADCWTNFVNPQRPFRP WNELEKNSWRIQRRLRRIKESWGL(SEQ ID NO:63)

[0469] 牛APOBEC-3B:

[0470] DGWEVAFRSGTVLKAGVLGVSMTEGWAGSGHPGQGACVWTPGTRNTMNLLREVLFKQQFGNQPRVPAP YYRRKTYLCYQLKQRNDLTLDRGCFRNKKQRHAEIRFIDKINSLDLNPSQSYKIICYITWSPCPNCANELVNFITR NNHLKLEIFASRLYFHWIKSFKMGLQDLQNAGISVAVMTHTEFEDCWEQFVDNQSRPFQPWDKLEQYSASIRRRLQ RILTAPI(SEQ ID NO:64)

[0471] 黑猩猩APOBEC-3B:

[0472] MNPQIRNPMEWMYQRTFYYNFENEPILYGRSYTWLCYEVKIRRGHSNLLWDTGVFRGQMYSQPEHHAE MCFLSWFCGNQLSAYKCFQITWFVSWTPCPDCVAKLAKFLAEHPNVTLTISAARLYYYWERDYRRALCRLSQAGAR VKIMDDEEFAYCWENFVYNEGQPFMPWYKFDDNYAFLHRTLKEIIRHLMDPDTFTFNFNNDPLVLRRHQTYLCYEV ERLDNGTWVLMDQHMGFLCNEAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGQVRAF LQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVYRQGCPFQPWDGLEEHSQALSGR LRAILQVRASSLCMVPHRPPPPPQSPGPCLPLCSEPPLGSLLPTGRPAPSLPFLLTASFSFPPPASLPPLPSLSLS PGHLPVPSFHSLTSCSIQPPCSSRIRETEGWASVSKEGRDLG(SEQ ID NO:65)

[0473] 人APOBEC-3C:

[0474] MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVV SWKTGVFRNQVDSETH*CHAERCFLSWFCDDILSPNTKYQVTWYTSWSPCPD CAGEVAEFLARHSNVNLTIFTARLYYFQYPCYQEGLRSLSQEGVAVEIMDY

[0475] EDFKYCWENFVYNDNEPFKPWKGLKTNFRLLKRRLRESLQ (SEQ ID NO: 66)

[0476] (斜体字:核酸编辑域)

[0477] 大猩猩APOBEC3C:

[0478] MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFR NQVDSETH*CHAERCFLSWFCDDILSPNTNYQVTWYTSWSPCPEC*AGEVAEFLARHSNVN LTIFTARLYYFQDTDYQEGLRSLSQEGVAVKIMDYKDFKYCWENFVYNDDEPFKPWK GLKYNFRFLKRRLQEILE (SEQ ID NO: 67)

[0479] (斜体字:核酸编辑域)

[0480] 人APOBEC-3A:

[0481] MEASPASGPRHLMDPHIFTSNFNNGIGRHKTYLCYEVERLDNGTSVKMD QHRGFLHNQAKNLLCGFYGR*HAELRFLDLVPSLQLDPAQIYRVTWFISWSPC FSWGC*AGEVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQV SIMTYDEFKHCWDTFVDHQGCPFQPWDGLDEHSQALSGRLRAILQNQGN (SEQ ID NO: 68)

[0482] (斜体字:核酸编辑域)

[0483] 猕猴APOBEC-3A:

[0484] MDGSPASRPRHLMDPNTFTFNFNNDLSVRGRHQTYLCYEVERLDNGTWVPMDERRG FLCNKAKNVPCGDYGC*HVELRFLCEVPSWQLDPAQTYRVTWFISWSPC*FRRGCAGQVR VFLQENKHVRLRIFAARIYDYDPLYQEALRTLRDAGAQVSIMTYEEFKHCWDTFVDR QGRPFQPWDGLDEHSQALSGRLRAILQNQGN (SEQ ID NO: 69)

[0485] (斜体字:核酸编辑域)

[0486] 牛APOBEC-3A:

[0487] MDEYTFTENFNNQGWPSKTYLCYEMERLDGDATIPLDEYKGFVRNKGLDQPEKPC*HAELYFLGKIHSWNLDRNQHYRLTCFISWSPC*YDCAQKLTTFLKENHHISLHILASRIYTHNRFGCHQSGLCELQAAGARITIMTFEDFKHCWETFVDHKGKPFQPWEGLNVKSQALCTELQAILKTQQN (SEQ ID NO: 70)

[0488] (斜体字:核酸编辑域)

[0489] 人APOBEC-3H:

[0490] MALLTAETFRLQFNNKRRLRRPYYPRKALLCYQLTPQNGSTPTRGYFENKKKC*HAEICFINEIKSMGLDETQCYQVTCYLTWSPCSSC*AWELVDFIKAHDHLNLGIFASRLYYHWCKPQQKGLRLLCGSQVPVEVMGFPKFADCWENFVDHEKPLSFNPYKMLEELDKNSRAIKRRLERIKIPGVRAQGRYMDILCDAEV (SEQ ID NO: 71)

[0491] (斜体字:核酸编辑域)

[0492] 猕猴APOBEC-3H:

[0493] MALLTAKTFSLQFNNKRRVNKPYYPRKALLCYQLTPQNGSTPTRGHLKNKKKDHAEIRFINKIKSMGLDETQCYQVTCYLTWSPCPSCAGELVDFIKAHRHLNLRIFASRLYYHWRPNYQEGLLLLCGSQVPVEVMGLPEFTDCWENFVDHKEPPSFNPSEKLEELDKNSQAIKRRLERIKSRSVDVLENGLRSLQLGPVTPSSSIRNSR(SEQ ID NO: 72)

[0494] 人APOBEC-3D:

[0495] MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGPVLPKRQSNHRQEVYFRFEN*HAEMCFLSWFCGNRLPANRRFQITWFVSWNPCLPC*VVKVTKFLAEHPNVTLTISAARLYYYRDRDWRWVLLRLHKAGARVKIMDYEDFAYCWENFVCNEGQPFMPWYKFDDNYASLHRTLKEILRNPMEAMYPHIFYFHFKNLLKACGRNE

[0496] SWLCFTMEVTKHHSAVFRKRGVFRNQVDPETHC*HAERCFLSWFCDDILSPNTNYEVTW YTSWSPCPEC*AGEVAEFLARHSNVNLTIFTARLCYFWDTDYQEGLCSLSQEGASVKIMGYKDFVSCWKNFVYSDDEPFKPWKGLQTNFRLLKRRLREILQ (SEQ ID NO: 73)

[0497] (斜体字:核酸编辑域)

[0498] 人APOBEC-1:

[0499] MTSEKGPSTGDPTLRRRIEPWEFDVFYDPRELRKEACLLYEIKWGMSRKIWRSSGKNTTNHVEVNFIKKFTSERDFHPSMSCSITWFLSWSPCWECSQAIREFLSRHPGVTLVIYVARLFWHMDQQNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKISRRWQNHLTFFRLHLQNCHYQTIPPHILLATGLIHPSVAWR(SEQ ID NO:74)

[0500] 小鼠APOBEC-1:

[0501] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSVWRHTSQNTSNHVEVNFLEKFTTERYFRPNTRCSITWFLSWSPCGECSRAITEFLSRHPYVTLFIYIARLYHHTDQRNRQGLRDLISSGVTIQIMTEQEYCYCWRNFVNYPPSNEAYWPRYPHLWVKLYVLELYCIILGLPPCLKILRRKQPQLTFFTITLQTCHYQRIPPHLLWATGLK(SEQ ID NO:75)

[0502] 大鼠APOBEC-1:

[0503] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPP

HILWATGLK(SEQ ID NO:76)

[0504] 人APOBEC-2:

[0505] MAQKEEAAVATEAASQNGEDLENLDDPEKLKELIELPPFEIVTGERLPANFFKFQFRNVEYSSGRNKT FLCYVVEAQGKGGQVQASRGYLEDEHAAAHAEEAFFNTILPAFDPALRYNVTWYVSSSPCAACADRIIKTLSKTKN LRLLILVGRLFMWEEPEIQAALKKLKEAGCKLRIMKPQDFEYVWQNFVEQEEGESKAFQPWEDIQENFLYYEEKLA DILK(SEQ ID NO:77)

[0506] 小鼠APOBEC-2:

[0507] MAQKEEAAEAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFRNVEYSSGRNKT FLCYVVEVQSKGGQAQATQGYLEDEHAGAHAEEAFFNTILPAFDPALKYNVTWYVSSSPCAACADRILKTLSKTKN LRLLILVSRLFMWEEPEVQAALKKLKEAGCKLRIMKPQDFEYIWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLA DILK(SEQ ID NO:78)

[0508] 大鼠APOBEC-2:

[0509] MAQKEEAAEAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFRNVEYSSGRNKT FLCYVVEAQSKGGQVQATQGYLEDEHAGAHAEEAFFNTILPAFDPALKYNVTWYVSSSPCAACADRILKTLSKTKN LRLLILVSRLFMWEEPEVQAALKKLKEAGCKLRIMKPQDFEYLWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLA DILK(SEQ ID NO:79)

[0510] 牛APOBEC-2:

[0511] MAQKEEAAAAAEPASQNGEEVENLEDPEKLKELIELPPFEIVTGERLPAHYFKFQFRNVEYSSGRNKT FLCYVVEAQSKGGQVQASRGYLEDEHATNHAEEAFFNSIMPTFDPALRYMVTWYVSSSPCAACADRIVKTLNKTKN LRLLILVGRLFMWEEPEIQAALRKLKEAGCRLRIMKPQDFEYIWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLA DILK(SEQ ID NO:80)

[0512] 海七鳃鳗CDA1(pmCDA1)

[0513] MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQSGTERGIHAE IFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLKIWACKLYYEKNARNQIGLWNLRDN GVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKRRSELSIMIQVKILHTTKSPAV(SEQ ID NO: 81)

[0514] 人APOBEC3G D316R_D317RMKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSR PPLDAKIFRGQVYSELKYHPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPKVTLTIFVAR LYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEPWNNLPKYYILLHIMLGEILRHSMD PPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLD QDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIYRRQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTF VDHQGCPFQPWDGLDEHSQDLSGRLRAILQNQEN(SEQ ID NO:82)

[0515] 人APOBEC3G链AMDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHG FLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRT LAEAGAKISIMTYSEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQ(SEQ ID NO:83)

[0516] 人APOBEC3G链A D120R_D121RMDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRR GFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIY RRQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQ(SEQ ID NO: 84)

[0517] 在一些实施方案中，如本文提供的融合蛋白包含核酸编辑酶的全长氨基酸，例如以上提供的序列之一。然而，在其他实施方案中，如本文提供的融合蛋白不包含核酸编辑酶的全长序列，而仅包含其片段。例如，在一些实施方案中，本文提供的融合蛋白包含Cas9域和核酸编辑酶的片段，例如，其中片段包含核酸编辑域。核酸编辑域的示例性氨基酸序列在上面的序列中以斜体字母显示，并且此类域的其他合适的序列对本领域技术人员而言是明显的。

[0518] 其他合适的核酸编辑酶序列，例如脱氨酶和域序列(它们可以根据本发明的方面使用，例如可以与无核酸酶活性的Cas9域融合)基于本公开内容对于本领域技术人员而言将是明显的。在一些实施方案中，此类另外的酶序列包含脱氨酶或脱氨酶域序列，其与本文提供的序列至少70％，至少75％，至少80％，至少85％，至少90％，至少95％，至少96％，至少97％，至少98％或至少99％相似。其他合适的Cas9域，变体和序列对本领域技术人员而言也是明显的。此类其他合适的Cas9域的实例包括但不限于D10A，D10A/D839A/H840A，和D10A/D839A/H840A/N863A突变域(参见例如Prashant et al.，CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering.Nature Biotechnology.2013;31(9):833-838，其全部内容通过引用并入本文)。在一些实施方案中，Cas9包含SEQ ID NO:6中提供的氨基酸序列的位置840处的组氨酸残基，或任何Cas9蛋白，诸如如本文提供的任一种Cas9氨基酸序列中的相应突变。催化残基H840的存在恢复Cas9切割含有与靶定C相对的G的非编辑链的能力。H840的恢复不导致含有C的靶链的切割。

[0519] 结合本领域的常识，基于本公开内容，用于产生包含Cas9域和脱氨酶域的融合蛋白的其他合适的策略对于本领域技术人员将是明显的。考虑到本公开内容和本领域的知识，使用接头或不使用接头产生根据本公开内容的方面的融合蛋白的合适策略对于本领域技术人员也是明显的。例如，Gilbert et al.，CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes.Cell.2013;154(2):442-51显示了使用2个NLS作为接头(SPKKKRKVEAS，SEQ ID NO:617)的Cas9与VP64的C端融合物可用于转录激活。Mali et al.，CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering.Nat Biotechnol.2013;31(9):833-8报告了无接头的与VP64的C端融合物可用于转录激活。并且，Maeder et al.，CRISPR RNA-guided activation of endogenous human genes.Nat Methods.2013;10:977-979报告了使用Gly_4Ser(SEQ ID NO:613)接头与VP64的C端融合物可用作转录激活剂。最近，已成功产生dCas9-FokI核酸酶融合物，并与亲本Cas9酶相比显示出改善的酶特异性(于Guilinger JP，Thompson DB，Liu DR.Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification.Nat.Biotechnol.2014;32(6):577-82，和Tsai SQ，Wyvekens N，Khayter C，Foden JA，Thapar V，Reyon D，Goodwin MJ，Aryee MJ，Joung JK.Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing.Nat Biotechnol.2014;32(6):569-76.PMID:24770325；分别在FokI-dCas9融合蛋白中使用SGSETPGTSESATPES(SEQ ID NO:604)或GGGGS(SEQ ID NO:607)接头)。

[0520] 本公开内容的一些方面提供了融合蛋白，其包含(i)Cas9酶或域(例如，第一蛋

白);和(ii)核酸编辑酶或域(例如第二种蛋白质)。在一些方面,本文提供的融合蛋白还包括(iii)可编程的DNA结合蛋白,例如锌指域,TALE或第二Cas9蛋白(例如第三蛋白)。不希望受任何特定理论的束缚,将可编程的DNA结合蛋白(例如,第二Cas9蛋白)与包含(i)Cas9酶或域(例如,第一蛋白);和(ii)核酸编辑酶或域(例如第二蛋白)的融合蛋白融合可以用于改善融合蛋白对靶核酸序列的特异性,或用于改善融合蛋白结合不含规范PAM(NGG)序列的核酸序列的特异性或结合亲和力。在一些实施方案中,第三蛋白是Cas9蛋白质(例如,第二Cas9蛋白)。在一些实施方案中,第三蛋白是本文提供的任何Cas9蛋白。在一些实施方案中,第三蛋白与Cas9蛋白(例如,第一蛋白)N末端的融合蛋白融合。在一些实施方案中,第三蛋白与Cas9蛋白质(例如第一蛋白)C末端的融合蛋白融合。在一些实施方案中,通过接头(例如,第二接头)融合Cas9域(例如,第一蛋白)和第三蛋白(例如,第二Cas9蛋白)。在一些实施方案中,接头包含$(GGGGS)_n$(SEQ ID NO:607),$(G)_n$(SEQ ID NO:608),$(EAAAK)_n$(SEQ ID NO:609),$(GGS)_n$(SEQ ID NO:610),$(SGGS)_n$(SEQ ID NO:606),SGSETPGTSESATPES(SEQ ID NO:604),SGGS$(GGS)_n$(SEQ ID NO:612),SGGSSGGSSGSETPGTSESATPESSGGSSGGS(SEQ ID NO:605)或$(XP)_n$(SEQ ID NO:611)基序或这些的任意组合,其中n独立地是1和30之间的整数。在一些实施方案中,本文提供的示例性napDNAbp融合蛋白的一般结构包含以下结构:

[0521] [NH2]-[核酸编辑酶或域]-[napDNAbp]-[第三蛋白]-[COOH];

[0522] [NH2]-[第三蛋白]-[napDNAbp]-[核酸编辑酶或域]-[COOH];

[0523] [NH2]-[napDNAbp]-[核酸编辑酶或域]-[第三蛋白]-[COOH];

[0524] [NH2]-[第三蛋白]-[核酸编辑酶或域]-[napDNAbp]-[COOH];

[0525] [NH2]-[UGI]-[核酸编辑酶或域]-napDNAbp]-[第三蛋白]-[COOH];

[0526] [NH2]-[UGI]-[第三蛋白]-[napDNAbp]-[核酸编辑酶或域]-[COOH];

[0527] [NH2]-[UGI]-[napDNAbp]-[核酸编辑酶或域]-[第三蛋白]-[COOH];

[0528] [NH2]-[UGI]-[第三蛋白]-[核酸编辑酶或域]-[napDNAbp]-[COOH];

[0529] [NH2]-[核酸编辑酶或域]-[napDNAbp]-[第三蛋白]-[UGI]-[COOH];

[0530] [NH2]-[第三蛋白]-[napDNAbp]-[核酸编辑酶或域]-[UGI]-[COOH];

[0531] [NH2]-[NapDNAbp]-[核酸编辑酶或域]-[第三蛋白]-[UGI]-[COOH];

[0532] [NH2]-[第三蛋白]-[核酸编辑酶或域]-[NapDNAbp]-[UGI]-[COOH];或

[0533] [NH2]-[核酸编辑酶或域]-[NapDNAbp]-[第一UGI域]-[第二UGI域]-[COOH]

[0534] 其中NH2是融合蛋白的N末端,而COOH是融合蛋白的C末端。在一些实施方案中,以上一般结构中使用的"]-["表示存在任选的接头序列。在其他实例中,本文提供的示例性NapDNAbp融合蛋白的一般结构包括以下结构:

[0535] [NH2]-[核酸编辑酶或域]-[NapDNAbp]-[第二NapDNAbp蛋白]-[COOH];

[0536] [NH2]-[第二NapDNAbp蛋白]-[NapDNAbp]-[核酸编辑酶或域]-[COOH];

[0537] [NH2]-[NapDNAbp]-[核酸编辑酶或域]-[第二NapDNAbp蛋白]-[COOH];

[0538] [NH2]-[第二NapDNAbp蛋白]-[核酸编辑酶或域]-[NapDNAbp]-[COOH];

[0539] [NH2]-[UGI]-[核酸编辑酶或域]-[NapDNAbp]-[第二NapDNAbp蛋白]-[COOH],

[0540] [NH2]-[UGI]-[第二NapDNAbp蛋白]-[NapDNAbp]-[核酸编辑酶或域]-[COOH];

[0541] [NH2]-[UGI]-[NapDNAbp]-[核酸编辑酶或域]-[第二NapDNAbp蛋白]-[COOH];

[0542] [NH2]-[UGI]-[第二NapDNAbp蛋白]-[核酸编辑酶或域]-[NapDNAbp]-[COOH];

[0543] [NH2]-[核酸编辑酶或域]-[NapDNAbp]-[第二NapDNAbp蛋白]-[UGI]-[COOH]；

[0544] [NH2]-[第二NapDNAbp蛋白]-[NapDNAbp]-[核酸编辑酶或域]-[UGI]-[COOH]；

[0545] [NH2]-[NapDNAbp]-[核酸编辑酶或域]-[第二NapDNAbp蛋白]-[UGI]-[COOH]；或

[0546] [NH2]-[第二NapDNAbp蛋白]-[核酸编辑酶或域]-[NapDNAbp]-[UGI]-[COOH]；

[0547] 其中NH_2是融合蛋白的N末端，而COOH是融合蛋白的C末端。在一些实施方案中，以上一般结构中使用的"]-["表示存在任选的接头序列。在一些实施方案中，第二NapDNAbp是dCas9蛋白。在一些实例中，本文提供的示例性Cas9融合蛋白的一般结构包括如图3所示的结构。应当理解，示例性Cas9融合蛋白的任何一般结构中提供的任何蛋白可以通过本文提供的一个或多个接头连接。在一些实施方案中，接头是相同的。在一些实施方案中，接头是不同的。在一些实施方案中，示例性Cas9融合蛋白的任何一般结构中提供的一个或多个蛋白质不经由接头融合。在一些实施方案中，融合蛋白还包含核靶向序列，例如核定位序列。在一些实施方案中，本文提供的融合蛋白还包含核定位序列(NLS)。在一些实施方案中，将NLS与融合蛋白的N末端融合。在一些实施方案中，将NLS与融合蛋白的C末端融合。在一些实施方案中，将NLS与第三蛋白质的N末端融合。在一些实施方案中，将NLS与第三蛋白的C末端融合。在一些实施方案中，将NLS与Cas9蛋白的N末端融合。在一些实施方案中，将NLS与Cas9蛋白的C末端融合。在一些实施方案中，将NLS与核酸编辑酶或域的N末端融合。在一些实施方案中，将NLS与核酸编辑酶或域的C末端融合。在一些实施方案中，将NLS与UGI蛋白的N末端融合。在一些实施方案中，将NLS与UGI蛋白的C末端融合。在一些实施方案中，通过一个或多个接头将NLS与融合蛋白融合。在一些实施方案中，不用接头将NLS与融合蛋白融合。

[0548] 尿嘧啶糖基化酶抑制剂融合蛋白

[0549] 本公开内容的一些方面涉及包含尿嘧啶糖基化酶抑制剂(UGI)域的融合蛋白。在一些实施方案中，本文提供的包含Cas9域(例如，核酸酶活性Cas9域，无核酸酶活性的dCas9域或Cas9切口酶)的任何融合蛋白可以进一步直接或经由接头与UGI域融合。本公开内容的一些方面提供了具有增加的核碱基编辑效率的脱氨酶-dCas9融合蛋白，脱氨酶-核酸酶活性Cas9融合蛋白和脱氨酶-Cas9切口酶融合蛋白。不希望受到任何特定理论的束缚，对U:G异双链体DNA的存在的细胞DNA修复应答可以造成细胞中核碱基编辑效率的降低。例如，尿嘧啶DNA糖基化酶(UDG)催化从细胞中的DNA除去U，这可以启动碱基切除修复，其中最常见的结果是将U:G对回复为C:G对。如以下实施例中证明，尿嘧啶DNA糖基化酶抑制剂(UGI)可抑制人UDG活性。因此，本公开内容涵盖了包含进一步与UGI域融合的dCas9-核酸编辑域的融合蛋白。本公开内容还考虑了包含进一步与UGI域融合的Cas9切口酶-核酸编辑域的融合蛋白。应当理解，使用UGI域可以提高能够催化C至U变化的核酸编辑域的编辑效率。例如，包含UGI域的融合蛋白可以更有效地使C残基脱氨基。在一些实施方案中，融合蛋白包含以下结构：

[0550] [脱氨酶]-[任选的接头序列]-[dCas9]-[任选的接头序列]-[UGI]；

[0551] [脱氨酶]-[任选的接头序列]-[UGI]-[任选的接头序列]-[dCas9]；

[0552] [UGI]-[任选的接头序列]-[脱氨酶]-[任选的接头序列]-[dCas9]；

[0553] [UGI]-[任选的接头序列]-[dCas9]-[任选的接头序列]-[脱氨酶]；

[0554] [dCas9]-[任选的接头序列]-[脱氨酶]-[任选的接头序列]-[UGI]；

[0555] [dCas9]-[任选的接头序列]-[UGI]-[任选的接头序列]-[脱氨酶]；

[0556] [脱氨酶]-[任选的接头序列]-[dCas9]-[任选的接头序列]-[第一UGI]-[任选的接头序列]-[第二UGI]；

[0557] [脱氨酶]-[任选的接头序列]-[第一UGI]-[任选的接头序列]-[第二UGI]-[任选的接头序列]-[dCas9]；

[0558] [第一UGI]-[任选的接头序列]-[第二UGI]-[任选的接头序列]-[脱氨酶]-[任选的接头序列]-[dCas9]；

[0559] [第一UGI]-[任选的接头序列]-[第二UGi]-[任选的接头序列]-[dCas9]-[任选的接头序列]-[脱氨酶]；

[0560] [dCas9]-[任选的接头序列]-[脱氨酶]-[任选的接头序列]-[第一UGI]-[任选的接头序列]-[第二UGI]；或

[0561] [dCas9]-[任选的接头序列]-[第一UGI]-[任选的接头序列]-[第二UGI]-[任选的接头序列]-[脱氨酶]。

[0562] 在其他实施方案中，融合蛋白包含以下结构：

[0563] [脱氨酶]-[任选的接头序列]-[Cas9切口酶]-[任选的接头序列]-[UGI]；

[0564] [脱氨酶]-[任选的接头序列]-[UGI]-[任选的接头序列]-[Cas9切口酶]；

[0565] [UGI]-[任选的接头序列]-[脱氨酶]-[任选的接头序列]-[Cas9切口酶]；

[0566] [UGI]-[任选的接头序列]-[Cas9切口酶]-[任选的接头序列]-[脱氨酶]；

[0567] [Cas9切口酶]-[任选的接头序列]-[脱氨酶]-[任选的接头序列]-[UGI]

[0568] [Cas9切口酶]-[任选的接头序列]-[UGI]-[任选的接头序列]-[脱氨酶]

[0569] [脱氨酶]-[任选的接头序列]-[Cas9切口酶]-[任选的接头序列]-[第一UGI]-[任选的接头序列]-[第二UGI]；

[0570] [脱氨酶]-[任选的接头序列]-[第一UGI]-[任选的接头序列]-[第二UGI]-[任选的接头序列]-[Cas9切口酶]；

[0571] [第一UGI]-[任选的接头序列]-[第二UGI]-[任选的接头序列]-[脱氨酶]-[任选的接头序列]-[Cas9切口酶]；

[0572] [第一UGI]-[任选的接头序列]-[第二UGi]-[任选的接头序列]-[Cas9切口酶]-[任选的接头序列]-[脱氨酶]；

[0573] [Cas9切口酶]-[任选的接头序列]-[脱氨酶]-[任选的接头序列]-[第一UGI]-[任选的接头序列]-[第二UGI]；或

[0574] [Cas9切口酶]-[任选的接头序列]-[第一UGI]-[任选的接头序列]-[第二UGI]-[任选的接头序列]-[脱氨酶]。

[0575] 应当理解，上文描述的任何融合蛋白可以由以下构成：(i)核酸可编程DNA结合蛋白(napDNAbp)；(ii)胞苷脱氨酶域；和(iii)两个或更多个UGI域，其中两个或更多个UGI域可以在构建体中彼此相邻(例如，[第一UGI]-[第二UGI]，其中“-”为任选的接头)，或两个或更多个UGI域可以与(i)的napDNAbp和/或(ii)的胞苷脱氨酶域分开(例如[[第一UGI]-[脱氨酶]-[第二UGI]，[第一UGI]-[napDNAbp]-[第二UGI]，[第一UGI]-[脱氨酶]-[napDNAbp]-[第二UGI]等，其中“-”是任选的接头)。

[0576] 另一方面，融合蛋白包含：(i)Cas9酶或域；(ii)核酸编辑酶或域(例如第二蛋白)(例如胞苷脱氨酶域)；(iii)第一尿嘧啶糖基化酶抑制剂域(UGI)(例如，第三蛋白)；和(iv)

第二尿嘧啶糖基化酶抑制剂域(UGI)(例如,第四蛋白)。第一和第二尿嘧啶糖基化酶抑制剂域(UGI)可以是相同或不同的。在一些实施方案中,Cas9域(例如,第一蛋白)和脱氨酶(例如,第二蛋白)通过接头融合。在一些实施方案中,Cas9域与脱氨酶的C末端融合。在一些实施方案中,Cas9蛋白(例如,第一蛋白)和第一UGI域(例如,第三蛋白)经由接头(例如,第二接头)融合。在一些实施方案中,第一UGI域与Cas9蛋白的C-末端融合。在一些实施方案中,第一UGI域(例如,第三蛋白)和第二UGI域(例如,第四蛋白)通过接头(例如,第三接头)融合。在一些实施方案中,第二UGI域与第一UGI域的C末端融合。在一些实施方案中,接头包含$(GGGGS)_n$(SEQ ID NO:607),$(G)_n$(SEQ ID NO:608),$(EAAAK)_n$(SEQ ID NO:609),$(GGS)_n$(SEQ ID NO:610),$(SGGS)_n$(SEQ ID NO:606),SGSETPGTSESATPES(SEQ ID NO:604),SGGS$(GGS)_n$(SEQ ID NO:612),SGGSSGGSSGSETPGTSESATPESSGGSSGGS(SEQ ID NO:605)或$(XP)_n$(SEQ ID NO:611)基序或这些的任意组合,其中n独立地是1和30之间的整数。在一些实施方案中,第一接头包含氨基酸1-50个氨基酸的序列。在一些实施方案中,第一接头包含1-40个氨基酸的氨基酸序列。在一些实施方案中,第一接头包含1-35个氨基酸的氨基酸序列。在一些实施方案中,第一接头包含1-30个氨基酸的氨基酸序列。在一些实施方案中,第一接头包含1-20个氨基酸的氨基酸序列。在一些实施方案中,第一接头包含10-20个氨基酸的氨基酸序列。在一些实施方案中,第一接头包含30-40个氨基酸的氨基酸序列。在一些实施方案中,第一接头包含14、16或18个氨基酸的氨基酸序列。在一些实施方案中,第一接头包含16个氨基酸的氨基酸序列。在一些实施方案中,第一接头包含30、32或34个氨基酸的氨基酸序列。在一些实施方案中,第一接头包含32个氨基酸的氨基酸序列。在一些实施方案中,第一接头包含SGSETPGTSESATPES(SEQ ID NO:604)基序。在一些实施方案中,第一接头包含SGGSSGGSSGSETPGTSESATPESSGGSSGGS(SEQ ID NO:605)基序。在一些实施方案中,第二接头包含1-50个氨基酸的氨基酸序列。在一些实施方案中,第二接头包含1-40个氨基酸的氨基酸序列。在一些实施方案中,第二接头包含1-35个氨基酸的氨基酸序列。在一些实施方案中,第二接头包含1-30个氨基酸的氨基酸序列。在一些实施方案中,第二接头包含1-20个氨基酸的氨基酸序列。在一些实施方案中,第二接头包含2-20个氨基酸的氨基酸序列。在一些实施方案中,第二接头包含2-10个氨基酸的氨基酸序列。在一些实施方案中,第二接头包含10-20个氨基酸的氨基酸序列。在一些实施方案中,第二接头包含2、4或6个氨基酸的氨基酸序列。在一些实施方案中,第二接头包含7、9或11个氨基酸的氨基酸序列。在一些实施方案中,第二接头包含14、16或18个氨基酸的氨基酸序列。在一些实施方案中,第二接头包含4个氨基酸的氨基酸序列。在一些实施方案中,第二接头包含9个氨基酸的氨基酸序列。在一些实施方案中,第二接头包含16个氨基酸的氨基酸序列。在一些实施方案中,第二接头包含$(SGGS)_n$(SEQ ID NO:606)基序,其中n是1和30之间的整数。在一些实施方案中,第二接头包含$(SGGS)_n$(SEQ ID NO:606)基序,其中n为1。在一些实施方案中,第二接头包含SGGS$(GGS)_n$(SEQ ID NO:612)基序,其中n是1和30之间的整数。在一些实施方案中,第二接头包含SGGS$(GGS)_n$(SEQ ID NO:612)基序,其中n为2。在一些实施方案中,第三接头包含1-50个氨基酸的氨基酸序列。在一些实施方案中,第三接头包含1-40个氨基酸的氨基酸序列。在一些实施方案中,第三接头包含1-35个氨基酸的氨基酸序列。在一些实施方案中,第三接头包含1-30个氨基酸的氨基酸序列。在一些实施方案中,第三接头包含1-20个氨基酸的氨基酸序列。在一些实施方案中,第三接头包含2-20个氨基酸的氨基酸序列。在一些实施方案中,第三接头包含

2-10个氨基酸的氨基酸序列。在一些实施方案中，第三接头包含10-20个氨基酸的氨基酸序列。在一些实施方案中，第三接头包含2、4或6个氨基酸的氨基酸序列。在一些实施方案中，第三接头包含7、9或11个氨基酸的氨基酸序列。在一些实施方案中，第三接头包含14、16或18个氨基酸的氨基酸序列。在一些实施方案中，第三接头包含4个氨基酸的氨基酸序列。在一些实施方案中，第三接头包含9个氨基酸的氨基酸序列。在一些实施方案中，第三接头包含16个氨基酸的氨基酸序列。在一些实施方案中，第三接头包含$(SGGS)_n$(SEQ ID NO:606)基序，其中n是1和30之间的整数。在一些实施方案中，第三接头包含$(SGGS)_n$(SEQ ID NO:606)基序，其中n为1。在一些实施方案中，第三接头包含SGGS$(GGS)_n$(SEQ ID NO:612)基序，其中n是1和30之间的整数。在一些实施方案中，第三接头包含SGGS$(GGS)_n$(SEQ ID NO:612)基序，其中n为2。

[0577] 在一些实施方案中，融合蛋白包含以下结构：

[0578] [脱氨酶]-[任选的接头序列]-[dCas9]-[任选的接头序列]-[第一UGI]-[任选的接头序列]-[第二UGI]；

[0579] [脱氨酶]-[任选的接头序列]-[Cas9切口酶]-[任选的接头序列]-[第一UGI]-[任选的接头序列]-[第二UGI]；或

[0580] [脱氨酶]-[任选的接头序列]-[Cas9]-[任选的接头序列]-[第一UGI]-[任选的接头序列]-[第二UGI]。

[0581] 另一方面，融合蛋白包含：(i)Cas9酶或域；(ii)核酸编辑酶或域(例如第二蛋白)(例如胞苷脱氨酶域)；(iii)超过两个尿嘧啶糖基化酶抑制剂(UGI)域。

[0582] 在一些实施方案中，本文提供的融合蛋白不包含接头序列。在一些实施方案中，存在任选的接头序列中的一个或两个。在一些实施方案中，存在任选的接头序列中的一个、两个或三个。

[0583] 在一些实施方案中，以上一般结构中使用的“-”表示存在任选的接头序列。在一些实施方案中，包含UGI的融合蛋白还包含核靶向序列，例如核定位序列。在一些实施方案中，本文提供的融合蛋白还包含核定位序列(NLS)。在一些实施方案中，将NLS与融合蛋白的N末端融合。在一些实施方案中，将NLS与融合蛋白的C末端融合。在一些实施方案中，将NLS与UGI蛋白的N末端融合。在一些实施方案中，将NLS与UGI蛋白的C末端融合。在一些实施方案中，将NLS与Cas9蛋白的N末端融合。在一些实施方案中，将NLS与Cas9蛋白的C末端融合。在一些实施方案中，将NLS与脱氨酶的N末端融合。在一些实施方案中，将NLS融合至脱氨酶的C末端。在一些实施方案中，将NLS与第二Cas9的N末端融合。在一些实施方案中，将NLS与第二Cas9的C末端融合。在一些实施方案中，通过一个或多个接头将NLS与融合蛋白融合。在一些实施方案中，不用接头将NLS与融合蛋白融合。在一些实施方案中，NLS包含本文提供或引用的任一种NLS序列的氨基酸序列。在一些实施方案中，NLS包含如SEQ ID NO:614或SEQ ID NO:615所示的氨基酸序列。

[0584] 在一些实施方案中，UGI域包含野生型UGI或以SEQ ID NO:134所示的UGI。在一些实施方案中，本文提供的UGI蛋白包含UGI的片段和与UGI或UGI片段同源的蛋白质。例如，在一些实施方案中，UGI域包含SEQ ID NO:134所示的氨基酸序列的片段。在一些实施方案中，UGI片段包含氨基酸序列，该氨基酸序列包含如SEQ ID NO:134所示的氨基酸序列的至少60％，至少65％，至少70％，至少75％，至少80％，至少85％，至少90％，至少95％，至少96％，

至少97％，至少98％，至少99％，或至少99.5％。在一些实施方案中，UGI包含与SEQ ID NO:134所示氨基酸序列同源的氨基酸序列或与SEQ ID NO:134所示氨基酸序列的片段同源的序列。在一些实施方案中，包含UGI或UGI的片段或UGI或UGI片段的同源物的蛋白质称为“UGI变体”。UGI变体与UGI或其片段共享同源性。例如，UGI变体与野生型UGI或如SEQ ID NO:134所示的UGI是至少70％相同，至少75％相同，至少80％相同，至少85％相同，至少90％相同，至少95％相同，至少96％相同，至少97％相同，至少98％相同，至少99％相同，至少99.5％相同或至少99.9％相同。在一些实施方案中，UGI变体包含UGI的片段，使得片段与野生型UGI或如SEQ ID NO:134所示的UGI的相应片段至少70％相同，至少80％相同，至少90％相同，至少95％相同，至少96％相同，至少97％相同，具有至少98％相同，至少99％相同，至少99.5％相同或至少99.9％。UGI包含以下氨基酸序列：

[0585] >sp|P14739|UNGI_BPPB2尿嘧啶DNA糖基化酶抑制剂MTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML(SEQ ID NO:134)

[0586] 本文提供了合适的UGI蛋白和核苷酸序列，并且其他合适的UGI序列是本领域技术人员已知的，并且包括例如以下发表的那些：Wang et al.,Uracil-DNA glycosylase inhibitor gene of bacteriophage PBS2 encodes a binding protein specific for uracil-DNA glycosylase.J.Biol.Chem.264:1163-1171(1989)；Lundquist et al.,Site-directed mutagenesis and characterization of uracil-DNA glycosylase inhibitor protein.Role of specific carboxylic amino acids in complex formation with Escherichia coli uracil-DNA glycosylase.J.Biol.Chem.272:21408-21419(1997)；Ravishankar et al.,X-ray analysis of a complex of Escherichia coli uracil DNA glycosylase(EcUDG)with a proteinaceous inhibitor.The structure elucidation of a prokaryotic UDG.Nucleic Acids Res.26:4880-4887(1998)；和Putnam et al.,Protein mimicry of DNA from crystal structures of the uracil-DNA glycosylase inhibitor protein and its complex with Escherichia coli uracil-DNA glycosylase.J.Mol.Biol.287:331-346(1999)，每篇的全部内容通过引用并入本文。

[0587] 应当理解，其他蛋白质可以是尿嘧啶糖基化酶抑制剂。例如，能够抑制(例如，空间阻断)尿嘧啶-DNA糖基化酶碱基切除修复酶的其他蛋白质在本公开内容的范围内。另外，任何阻断或抑制碱基切除修复的蛋白质也在本公开内容的范围内。在一些实施方案中，本文描述的融合蛋白包含一个UGI域。在一些实施方案中，本文的融合蛋白包含超过一个UGI域。在一些实施方案中，本文描述的融合蛋白包含两个UGI域。在一些实施方案中，本文描述的融合蛋白包含超过两个UGI域。在一些实施方案中，使用结合DNA的蛋白质。在另一个实施方案中，使用UGI的替代物。在一些实施方案中，尿嘧啶糖基化酶抑制剂是结合单链DNA的蛋白质。例如，尿嘧啶糖基化酶抑制剂可以是塔斯曼尼亚欧文氏菌(Erwinia tasmaniensis)单链结合蛋白。在一些实施方案中，单链结合蛋白包含氨基酸序列(SEQ ID NO:135)。在一些实施方案中，尿嘧啶糖基化酶抑制剂是结合尿嘧啶的蛋白质。在一些实施方案中，尿嘧啶糖基化酶抑制剂是结合DNA中的尿嘧啶的蛋白质。在一些实施方案中，尿嘧啶糖基化酶抑制剂是无催化活性的尿嘧啶DNA-糖基化酶蛋白。在一些实施方案中，尿嘧啶糖基化酶抑制剂是不从DNA中切除尿嘧啶的无催化活性的尿嘧啶DNA-糖基化酶蛋白。例如，尿嘧啶糖基化酶抑制剂是UdgX。在一些实施方案中，UdgX包含氨基酸序列(SEQ ID NO:136)。作为另一个例子，

尿嘧啶糖基化酶抑制剂是无催化活性的UDG。在一些实施方案中，无催化活性的UDG包含氨基酸序列(SEQ ID NO:137)。应当理解，其他尿嘧啶糖基化酶抑制剂对于熟练技术人员将是明显的，并且在本公开内容的范围内。在一些实施方案中，尿嘧啶糖基化酶抑制剂是与SEQ ID NO:135-137或143-148中的任一项同源的蛋白质。在一些实施方案中，尿嘧啶糖基化酶抑制剂与SEQ ID NO:135-137或143-148中的任一项是至少50％相同，至少55％相同，至少60％相同，至少65％相同，至少70％相同，至少75％相同，至少50％相同的蛋白质。与SEQ中的任一序列具有80％相同，至少85％相同，至少90％相同，至少95％相同，至少96％相同，至少98％相同，至少99％相同或至少99.5％相同的。塔斯曼尼亚欧文氏菌SSB(热稳定性单链DNA结合蛋白)

[0588] MASRGVNKVILVGNLGQDPEVRYMPNGGAVANITLATSESWRDKQTGETKEKTEWHRVVLFGKLAEVAGEYLRKGSQVYIEGALQTRKWTDQAGVEKYTTEVVVNVGGTMQMLGGRSQGGGASAGGQNGGSNNGWGQPQQPQGGNQFSGGAQQQARPQQQPQQNNAPANNEPPIDFDDDIP(SEQ ID NO:135)

[0589] UdgX(结合DNA中的尿嘧啶但不切除)

[0590] MAGAQDFVPHTADLAELAAAAGECRGCGLYRDATQAVFGAGGRSARIMMIGEQPGDKEDLAGLPFVGPAGRLLDRALEAADIDRDALYVTNAVKHFKFTRAAGGKRRIHKTPSRTEVVACRPWLIAEMTSVEPDVVVLLGATAAKALLGNDFRVTQHRGEVLHVDDVPGDPALVATVHPSSLLRGPKEERESAFAGLVDDLRVAADVRP(SEQ ID NO:136)

[0591] UDG(无催化活性的人UDG结合DNA中的尿嘧啶但不切除)

[0592] MIGQKTLYSFFSPSPARKRHAPSPEPAVQGTGVAGVPEESGDAAAIPAKKAPAGQEEPGTPPSSPLSAEQLDRIQRNKAAALLRLAARNVPVGFGESWKKHLSGEFGKPYFIKLMGFVAEERKHYTVYPPPHQVFTWTQMCDIKDVKVVILGQEPYHGPNQAHGLCFSVQRPVPPPPSLENIYKELSTDIEDFVHPGHGDLSGWAKQGVLLLNAVLTVRAHQANSHKERGWEQFTDAVVSWLNQNSNGLVFLLWGSYAQKKGSAIDRKRHHVLQTAHPSPLSVYRGFFGCRHFSKTNELLQKSGKKPIDWKEL(SEQ ID NO:137)

[0593] 下文显示了可以用作UGI的另外的单链DNA结合蛋白。应当理解，其他单链结合蛋白可以用作UGI，例如以下描述的那些:Dickey TH,Altschuler SE,Wuttke DS.Single-stranded DNA-binding proteins:multiple domains for multiple functions.Structure.2013Jul 2;21(7):1074-84.doi:10.1016/j.str.2013.05.013.Review.;Marceau AH.Functions of single-strand DNA-binding proteins in DNAreplication,recombination,and repair.Methods Mol Biol.2012;922:1-21.doi:10.1007/978-1-62703-032-8_1.;Mijakovic,Ivan,et al.;Bacterial single-stranded DNA-binding proteins are phosphorylated on tyrosine.Nucleic Acids Res 2006;34(5):1588-1596.doi:10.1093/nar/gkj514;Mumtsidu E,Makhov AM,Konarev PV,Svergun DI,Griffith JD,Tucker PA.Structural features of the single-stranded DNA-binding protein of Epstein-Barrvirus.J Struct Biol.2008Feb;161(2):172-87.Epub 2007Nov 1;Nowak M,Olszewski M,Śpibida M,Kur J.Characterization of single-strandedDNA-binding proteins from the psychrophilic bacteria Desulfotalea psychrophila,Flavobacterium psychrophilum,Psychrobacter arcticus,Psychrobactercryohalolentis,Psychromonas ingrahamii,Psychroflexus torquis,and

[0594] Photobacterium profundum.BMC Microbiol.2014Apr 14;14:91.doi:10.1186/1471-2180-14-91;Tone T,Takeuchi A,Makino O.Single-stranded DNA binding protein Gp5 ofBacillus subtilis phageΦ29is required for viral DNA replication ingrowth-temperature dependent fashion.Biosci Biotechnol Biochem.2012;76(12):2351-3.Epub 2012Dec 7;Wold.REPLICATION PROTEIN A:A Heterotrimeric,Single-Stranded DNA-Binding Protein Required for Eukaryotic DNA Metabolism.Annual Review of Biochem.1997;66:61-92.doi:10.1146/annurev.biochem.66.1.61;Wu Y,Lu J,Kang T.Human single-stranded DNA binding proteins:guardians ofgenome stability.ActaBiochim Biophys Sin(Shanghai).2016Jul;48(7):671-7.doi:10.1093/abbs/gmw044.Epub 2016May 23.Review;每篇的全部内容在此通过引用并入。

[0595] mtSSB-SSBP1单链DNA结合域1[智人(人)](UniProtKB:Q04837;NP_001243439.1)

[0596] MFRRPVLQVLRQFVRHESETTTSLVLERSLNRVHLLGRVGQDPVLRQVEGKNPVTIFSLATNEMWRSGDSEVYQLGDVSQKTTWHRISVFRPGLRDVAYQYVKKGSRIYLEGKIDYGEYMDKNNVRRQATTIIADNIIFLSDQTKEKE(SEQ ID NO:138)

[0597] 单链DNA结合蛋白3同种型A[小家鼠](UniProtKB-Q9D032-1;NCBI Ref:NP_076161.2)

[0598] MFAKGKGSAVPSDGQAREKLALYVYEYLLHVGAQKSAQTFLSEIRWEKNITLGEPPGFLHSWWCVFWDLYCAAPERRDTCEHSSEAKAFHDYSAAAAPSPVLGNIPPNDGMPGGPIPPGFFQGPPGSQPSPHAQPPPHNPSSMMGPHSQPFMSPRYAGGPRPPIRMGNQPPGGVPGTQPLLPNSMDPTRQQGHPNMGGSMQRMNPPRGMGPMGPGPQNYGSGMRPPPNSLGPAMPGINMGPGAGRPWPNPNSANSIPYSSSSPGTYVGPPGGGGPPGTPIMPSPADSTNSSDNIYTMINPVPPGGSRSNFPMGPGSDGPMGGMGGMEPHHMNGSLGSGDIDGLPKNSPNNISGISNPPGTPRDDGELGGNFLHSFQNDNYSPSMTMSV(SEQ ID NO:139)

[0599] RPA 1-复制蛋白A 70kDa DNA-结合亚基(UniProtKB:P27694;NCBI Ref:NM_002945.3)

[0600] MVGQLSEGAIAAIMQKGDTNIKPILQVINIRPITTGNSPPRYRLLMSDGLNTLSSFMLATQLNPLVEEEQLSSNCVCQIHRFIVNTLKDGRRVVILMELEVLKSAEAVGVKIGNPVPYNEGLGQPQVAPPAPAASPAASSRPQPQNGSSGMGSTVSKAYGASKTFGKAAGPSLSHTSGGTQSKVVPIASLTPYQSKWTICARVTNKSQIRTWSNSRGEGKLFSLELVDESGEIRATAFNEQVDKFFPLIEVNKVYYFSKGTLKIANKQFTAVKNDYEMTFNNETSVMPCEDDHHLPTVQFDFTGIDDLENKSKDSLVDIIGICKSYEDATKITVRSNNREVAKRNIYLMDTSGKVVTATLWGEDADKFDGSRQPVLAIKGARVSDFGGRSLSVLSSSTIIANPDIPEAYKLRGWFDAEGQALDGVSISDLKSGGVGGSNTNWKTLYEVKSENLGQGDKPDYFSSVATVVYLRKENCMYQACPTQDCNKKVIDQQNGLYRCEKCDTEFPNFKYRMILSVNIADFQENQWVTCFQESAEAILGQNAAYLGELKDKNEQAFEEVFQNANFRSFIFRVRVKVETYNDESRIKATVMDVKPVDYREYGRRLVMSIRRSALM(SEQ ID NO:140)

[0601] RPA 2-复制蛋白A 32kDa亚基(UniProtKB:P15927;NCBI Ref:NM_002946)

[0602] MWNSGFESYGSSSYGGAGGYTQSPGGFGSPAPSQAEKKSRARAQHIVPCTISQLLSATLVDEVFRIGNVEISQVTIVGIIRHAEKAPTNIVYKIDDMTAAPMDVRQWVDTDDTSSENTVVPPETYVKVAGHLRSFQNKKSLVAFKIMPLEDMNEFTTHILEVINAHMVLSKANSQPSAGRAPISNPGMSEAGNFGGNSFMPANGLTVAQNQVLNLIKACPRPEGLNFQDLKNQLKHMSVSSIKQAVDFLSNEGHIYSTVDDDHFKSTDAE(SEQ ID NO:141)

[0603] RPA 3-复制蛋白A14kDa亚基(UniProtKB:P35244;NCBI Ref:NM_002947.4)

[0604] MVDMMDLPRSRINAGMLAQFIDKPVCFVGRLEKIHPTGKMFILSDGEGKNGTIELMEPLDEEISGIVEVVGRVTAKATILCTSYVQFKEDSHPFDLGLYNEAVKIIHDFPQFYPLGIVQHD(SEQ ID NO:142)

[0605] 细菌单链DNA结合蛋白：

[0606] ssbA-单链DNA结合蛋白[枯草芽孢杆菌枯草亚种168株](UniProtKB:P37455;NCBIRef:)

[0607] MLNRVVLVGR LTKDPELRYT PNGAAVATFT LAVNRTFTNQ SGEREADFIN CVTWRRQAEN VANFLKKGSL AGVDGRLQTR NYENQQGQRV FVTEVQAESV QFLEPKNGGG SGSGGYNEGN SGGGQYFGGG QNDNPFGGNQ NNQRRNQGNS FNDDPFANDG KPIDISDDDLPF(SEQ ID NO:143)

[0608] 单链DNA结合蛋白2[天蓝色链霉菌(Streptomyces coelicolor)A3(2)](UniProtKB:Q9X8U3;NCBI Ref:NP_628093.1)

[0609] MAGETVITVVGNLVDDPELRFTPSGAAVAKFRVASTPRTFDRQTNEWKDGESLFLTCSVWRQAAENVAESLQRGMRVIVQGRLKQRSYEDREGVKRTVYELDVDEVGASLRSATAKVTKTSGQGRGGQGGYGGGGGGQGGGGWGGGPGGGQQGGGAPADDPWATGGAPAGGQQGGGGQGGGGWGGGSGGGGGYSDEPPF(SEQ ID NO:144)

[0610] 单链DNA结合蛋白[肺炎链球菌(Streptococcus pneumonia)R6](UniProtKB:P66855;NCBI Ref:NP_358988.1)

[0611] MINNVVLVGRMTRDAELRYTPSNVAVATFTLAVNRTFKSQNGEREADFINVVMWRQQAENLANWAKKGSLIGVTGRIQTRSYDNQQGQRVYVTEVVAENFQMLESRSVREGHTGGAYSAPTANYSAPTNSVPDFSRNENPFGATNPLDISDDDLPF(SEQ ID NO:145)病毒单链DNA结合蛋白：

[0612] 单链DNA结合蛋白[人alpha疱疹病毒1](UniProtKB:P04296;NCBI Ref:YP_009137104.1)

[0613] METKPKTATTIKVPPGPLGYVYARACPSEGIELLALLSARSGDSDVAVAPLVVGLTVESGFEANVAVVVGSRTTGLGGTAVSLKLTPSHYSSSVYVFHGGRHLDPSTQAPNLTRLCERARRHFGFSDYTPRPGDLKHETTGEALCERLGLDPDRALLYLVVTEGFKEAVCINNTFLHLGGSDKVTIGGAEVHRIPVYPLQLFMPDFSRVIAEPFNANHRSIGENFTYPLPFFNRPLNRLLFEAVVGPAAVALRCRNVDAVARAAAHLAFDENHEGAALPADITFTAFEASQGKTPRGGRDGGGKGPAGGFEQRLASVMAGDAALALESIVSMAVFDEPPTDISAWPLFEGQDTAAARANAVGAYLARAAGLVGAMVFSTNSALHLTEVDDAGPADPKDHSKPSFYRFFLVPGTHVAANPQVDREGHVVPGFEGRPTAPLVGGTQEFAGEHLAMLCGFSPALLAKMLFYLERCDGGVIVGRQEMDVFRYVADSNQTDVPCNLCTFDTRHACVHTTLMRLRARHPKFASAARGAIGVFGTMNSMYSDCDVLGNYAAFSALKRADGSETARTIMQETYRAATERVMAELETLQYVDQAVPTAMGRLETIITNREALHTVVNNVRQVVDREVEQLMRNLVEGRNFKFRDGLGEANHAMSLTLDPYACGPCPLLQLLGRRSNLAVYQDLALSQCHGVFAGQSVEGRNFRNQFQPVLRRRVMDMFNNGFLSAKTLTVALSEGAAICAPSLTAGQTAPAESSFEGDVARVTLGFPKELRVKSRVLFAGASANASEAAKARVASLQSAYQKPDKRVDILLGPLGFLLKQFHAAIFPNGKPPGSNQPNPQWFWTALQRNQLPARLLSREDIETIAFIKKFSLDYGAINFINLAPNNVSELAMYYMANQILRYCDHSTYFINTLTAIIAGSRRPPSVQAAAAAWSAQGGAGLEAGARALMDAVDAHPGAWTSMFASCNLLRPVMAARPMVVLGLSISKYYGMAGNDRVFQAGNWASLMGGKNACPLLIFDRTRKFVLACPRAGFVCAASSLGGGAHESSLCEQLRGIISEGGAAVASSVFVATVKSLGPRTQQLQIEDWLALLEDEYLSEEMMELTARALERGNGEWSTDAALEVAHEAEALVSQLGNAGEVFNFGDFGCEDDNATPFGGPGAPGPAFAGRKRAFHGDDPFGEGPPDKKGDLTLDML(SEQ ID NO:146)

[0614] 来自芽孢杆菌病毒phi29的单链DNA结合蛋白(UniProtKB:Q38504.1;NCBI Ref:

YP_002004532.1)

[0615] MENTNIVKATFDTETLEGQIKIFNAQTGGGQSFKNLPDGTIIEANAIAQYKQVSDTYGDAKEETVTTIFAADGSLYSAISKTVAEAASDLIDLVTRHKLETFKVKVVQGTSSKGNVFFSLQLSL(SEQ ID NO:147)

[0616] 单链DNA结合蛋白[伯霍尔德杆菌病毒DC1](UniProtKB:I6NRL7;NCBI Ref:YP_006589943.1)

[0617] MASVNKVILVGNLGADPETRYLPSGDAISNIRLATTDRYKDKASGEMKESTEWHRVSFFGRLAEIVDEYLRKGAPVYIEGRIRTRKWQDNAGQDRYTTEIVAEKMQMLGDRRDGGERQQRAPQQQQQRTQRNGYADATGRAQPSQRPAAGGGFDEMDDDIPF(SEQ ID NO:148)

[0618] 在一些实施方案中,核酸编辑域是脱氨酶域。在一些实施方案中,脱氨酶是胞嘧啶脱氨酶或胞苷脱氨酶。在一些实施方案中,脱氨酶是载脂蛋白BmRNA编辑复合物(APOBEC)家族脱氨酶。在一些实施方案中,脱氨酶是APOBEC1脱氨酶。在一些实施方案中,脱氨酶是APOBEC2脱氨酶。在一些实施方案中,脱氨酶是APOBEC3脱氨酶。在一些实施方案中,脱氨酶是APOBEC3A脱氨酶。在一些实施方案中,脱氨酶是APOBEC3B脱氨酶。在一些实施方案中,脱氨酶是APOBEC3C脱氨酶。在一些实施方案中,脱氨酶是APOBEC3D脱氨酶。在一些实施方案中,脱氨酶是APOBEC3E脱氨酶。在一些实施方案中,脱氨酶是APOBEC3F脱氨酶。在一些实施方案中,脱氨酶是APOBEC3G脱氨酶。在一些实施方案中,脱氨酶是APOBEC3H脱氨酶。在一些实施方案中,脱氨酶是APOBEC4脱氨酶。在一些实施方案中,脱氨酶是活化诱导的脱氨酶(AID)。在一些实施方案中,脱氨酶是大鼠APOBEC1(SEQ ID NO:74)。在一些实施方案中,脱氨酶是人APOBEC1(SEQID NO:76)。在一些实施方案中,脱氨酶是海七鳃鳗胞苷脱氨酶1(pmCDA1)。在一些实施方案中,脱氨酶是人APOBEC3G(SEQ ID NO:60)。在一些实施方案中,脱氨酶是人APOBEC3G的片段(SEQ ID NO:83)。在一些实施方案中,脱氨酶是包含D316R_D317R突变的人APOBEC3G变体(SEQ ID NO:82)。在一些实施方案中,脱氨酶是人APOBEC3G的,并且包含对应于SEQ ID NO:60中的D316R_D317R突变的突变的片段(SEQ ID NO:84)。

[0619] 在一些实施方案中,接头包含$(GGGS)_n$(SEQ ID NO:613),$(GGGGS)_n$(SEQ ID NO:607),$(G)_n$(SEQ ID NO:608),$(EAAAK)_n$(SEQ ID NO:609),$(GGS)_n$(SEQ ID NO:610),SGSETPGTSESATPES(SEQ ID NO:604)或$(XP)_n$(SEQ ID NO:611)基序,或这些的任何组合,其中n独立地是1和30之间的整数。

[0620] 本文提供了合适的UGI蛋白和核苷酸序列,并且另外的合适的UGI序列是本领域技术人员已知的,并且包括例如以下中发表的那些:Wang et al.,Uracil-DNA glycosylase inhibitor gene of bacteriophage PBS2 encodes a binding protein specific for uracil-DNA glycosylase.J.Biol.Chem.264:1163-1171(1989);Lundquist et al.,Site-directed mutagenesis and characterization of uracil-DNA glycosylase inhibitor protein.Role of specific carboxylic amino acids in complex formation with Escherichia coli uracil-DNA glycosylase.J.Biol.Chem.272:21408-21419(1997);Ravishankar et al.,X-ray analysis of a complex of Escherichia coli uracil DNA glycosylase(EcUDG)with a proteinaceous inhibitor.The structure elucidation of a prokaryotic UDG.Nucleic Acids Res.26:4880-4887(1998);和Putnam et al.,Protein mimicry of DNA from crystal structures of the uracil-DNA glycosylase inhibitor protein and its complex with Escherichia coli uracil-DNA

glycosylase.J.Mol.Biol.287:331-346(1999)，其全部内容通过引用并入本文。在一些实施方案中，任选的接头包含(GGS)$_n$(SEQ ID NO:610)基序，其中n为1、2、3、4、5、6、7、8、9、19、11、12、13、14、15、16、17、18、19或20。在一些实施方案中，任选的接头包含(GGS)$_n$(SEQ ID NO:610)基序，其中n为1、3或7。在一些实施方案中，任选的接头包含氨基酸序列SGSETPGTSESATPES(SEQ ID NO:604)，其在实施例中也称为XTEN接头。

[0621] 在一些实施方案中，Cas9切口酶可以进一步促进生物体中非编辑链上碱基的除去，所述生物体的基因组在体内编辑。如本文所述，Cas9切口酶可以包含SEQ ID NO:6中的D10A突变，或任何Cas9蛋白，诸如如本文提供的任一种Cas9氨基酸序列中的相应突变。在一些实施方案中，本公开内容的Cas9切口酶可以包含SEQ ID NO:6的突变840处的组氨酸或任何Cas9蛋白，诸如如本文提供的任一种Cas9氨基酸序列中的相应突变。此类包含Cas9切口酶的融合蛋白可以切割靶DNA序列的单链，例如不被编辑的链。不希望受任何特定理论的束缚，此种切割可以抑制错配修复机制，其回复由脱氨酶产生的C至U编辑物。

[0622] 具有引导RNA的Cas9复合物

[0623] 本公开内容的一些方面提供了复合物，其包含本文提供的任何融合蛋白和与融合蛋白的Cas9域(例如，dCas9，核酸酶活性Cas9或Cas9切口酶)结合的引导RNA。

[0624] 在一些实施方案中，引导RNA长15-100个核苷酸，并且包含与靶序列互补的至少10个连续核苷酸的序列。在一些实施方案中，引导RNA长15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49或50个核苷酸。在一些实施方案中，引导RNA包含15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39或40个与靶序列互补的连续核苷酸的序列。在一些实施方案中，靶序列是DNA序列。在一些实施方案中，靶序列是哺乳动物基因组中的序列。在一些实施方案中，靶序列是人基因组中的序列。在一些实施方案中，靶序列的3'末端直接与规范PAM序列(NGG)相邻。在一些实施方案中，引导RNA与疾病或病症相关的序列互补。

[0625] 使用Cas9融合蛋白的方法

[0626] 本公开内容的一些方面提供了使用本文提供的Cas9蛋白，融合蛋白或复合物的方法。例如，本公开内容的一些方面提供了方法，包括使DNA分子(a)与本文提供的任何Cas9蛋白或融合蛋白以及至少一个引导RNA接触，其中引导RNA长为约15-100个核苷酸并包含与靶序列互补的至少10个连续核苷酸的序列；或(b)与如本文提供的Cas9蛋白，Cas9融合蛋白或Cas9蛋白或具有至少一种gRNA的融合蛋白复合物接触。在一些实施方案中，靶序列的3'末端不直接与规范PAM序列(NGG)相邻。在一些实施方案中，靶序列的3'末端直接与AGC，GAG，TTT，GTG或CAA序列相邻。

[0627] 在一些实施方案中，靶DNA序列包含与疾病或病症相关的序列。在一些实施方案中，靶DNA序列包含与疾病或病症相关的点突变。在一些实施方案中，Cas9蛋白，Cas9融合蛋白或复合物的活性导致点突变的校正。在一些实施方案中，靶DNA序列包含与疾病或病症相关的T→C点突变，并且其中突变体C碱基的脱氨基导致不与疾病或病症相关的序列。在一些实施方案中，靶DNA序列编码蛋白质，并且其中点突变在密码子中，并且与野生型密码子相比，导致由突变体密码子编码的氨基酸发生改变。在一些实施方案中，突变体C的脱氨基导致由突变体密码子编码的氨基酸的变化。在一些实施方案中，突变体C的脱氨基导致编码野

生型氨基酸的密码子。在一些实施方案中，在受试者中体内接触。在一些实施方案中，受试者具有或已经被诊断出疾病或病症。在一些实施方案中，疾病或病症是囊性纤维化(cystic fibrosis)，苯丙酮尿症(phenylketonuria)，表皮松解角化过度症(epidermolytic hyperkeratosis)(EHK)，夏科-马里-图思病(Charcot-Marie-Toot disease)4J型，神经母细胞瘤(neuroblastoma)(NB)，血管性血友病(von Willebrand disease)(vWD)，先天性肌强直(myotonia congenital)，遗传性肾淀粉样变性(hereditary renal amyloidosis)，扩张型心肌病(dilated cardiomyopathy)(DCM)，遗传性淋巴水肿(hereditary lymphedema)，家族性阿尔茨海默氏病，HIV，朊病毒病(Prion disease)，慢性婴儿神经系统皮肤关节综合征(chronic infantile neurologic cutaneous articular syndrome)(CINCA)，结蛋白相关性肌病(desmin-related myopathy)(DRM)，与突变体PI3KCA蛋白，突变体CTNNB1蛋白，突变体HRAS蛋白或突变体p53蛋白相关的肿瘤性疾病(neoplastic disease)。

[0628] 一些实施方案提供了使用本文提供的Cas9 DNA编辑融合蛋白的方法。在一些实施方案中，融合蛋白用于通过使靶核碱基，例如C残基脱氨基而将点突变引入核酸中。在一些实施方案中，靶核碱基的脱氨基导致遗传缺陷的校正，例如，导致基因产物功能丧失的点突变的校正。在一些实施方案中，遗传缺陷与疾病或病症，例如溶酶体贮积症或代谢疾病，诸如例如I型糖尿病有关。在一些实施方案中，本文提供的方法用于将失活点突变引入编码与疾病或病症相关的基因产物的基因或等位基因中。例如，在一些实施方案中，本文提供了采用Cas9 DNA编辑融合蛋白将失活点突变引入癌基因中(例如，在治疗增殖性疾病中)的方法。在一些实施方案中，失活突变可以在编码序列中产生过早的终止密码子，其导致截短的基因产物，例如，缺乏全长蛋白质功能的截短的蛋白质的表达。

[0629] 在一些实施方案中，本文提供的方法的目的是通过基因组编辑来恢复功能障碍的基因的功能。可以例如通过校正人细胞培养物中疾病相关的突变来在体外针对基于基因编辑的人治疗剂验证本文提供的Cas9脱氨酶融合蛋白。熟练技术人员将理解，本文提供的融合蛋白，例如包含Cas9域和核酸脱氨酶域的融合蛋白可用于校正任何单点T->C或A->G突变。在第一种情况下，将突变体C脱氨基变回U可以校正突变，并且在后一种情况下，与突变体G碱基配对的C的脱氨基和接着的一轮复制校正突变。

[0630] 可以通过所提供的融合蛋白在体外或体内校正的示例性疾病相关突变是PI3KCA蛋白中的H1047R(A3140G)多态性。磷酸肌醇-3-激酶，催化性α亚基(PI3KCA)蛋白起磷酸化磷脂酰肌醇的肌醇环的3-OH基团的作用。已发现PI3KCA基因在许多不同的癌中发生了突变，因此认为它是有力的癌基因[37]。实际上，A3140G突变存在于几种NCI-60癌细胞系中，诸如例如HCT116，SKOV3和T47D细胞系，它们从美国典型培养物保藏中心(ATCC)容易获得[38]。

[0631] 在一些实施方案中，使携带要校正的突变的细胞(例如携带点突变，例如在PI3KCA基因的外显子20中的A3140G点突变，导致PI3KCA蛋白中的H1047R取代的细胞)与编码Cas9脱氨酶融合蛋白的表达构建体和适当设计的sgRNA接触，所述sgRNA将融合蛋白靶向至编码PI3KCA基因中的相应突变位点。可以进行对照实验，其中将sgRNA设计为将融合酶靶向到PI3KCA基因内的非C残基。可以提取经处理的细胞的基因组DNA，并将PI3KCA基因的相关序列进行PCR扩增和测序，以评估人细胞培养物中融合蛋白的活性。

[0632] 应当理解的是，提供校正PI3KCA中的点突变的例子是出于说明的目的，并不意味

着限制本公开内容。熟练技术人员将理解，本公开的DNA编辑融合蛋白可用于校正其他点突变和与其他癌症以及与除癌症之外的其他疾病(包括其他增殖性疾病)有关的突变。

[0633] 疾病相关基因和等位基因中点突变的成功校正为治疗学和基础研究中的应用打开了基因校正的新策略。位点特异性单碱基修饰系统，如公开的Cas9和脱氨酶或域的融合物在“逆”基因疗法中也具有应用，其中有意抑制或消除某些基因功能。在这些情况下，可以使用以位点特异性方式将Trp(TGG)，Gln(CAA和CAG)或Arg(CGA)残基突变为提前终止密码子(TAA，TAG，TGA)，以在体外，离体，或体内消除蛋白质功能。

[0634] 本公开内容提供了用于治疗诊断患有与点突变相关的疾病或由该点突变引起的疾病的受试者的方法，所述点突变可以通过本文提供的Cas9 DNA编辑融合蛋白来校正。例如，在一些实施方案中，提供了方法，该方法包括对患有此类疾病(例如，如上所述与PI3KCA点突变相关的癌症)的受试者施用有效量的Cas9脱氨酶融合蛋白，其校正该点突变或将失活突变引入疾病相关基因中。在一些实施方案中，疾病是增殖性疾病。在一些实施方案中，该疾病是遗传疾病。在一些实施方案中，该疾病是新生性疾病。在一些实施方案中，疾病是代谢疾病。在一些实施方案中，疾病是溶酶体贮积病。可以通过校正点突变或将失活突变引入疾病相关基因中来治疗的其他疾病是本领域技术人员已知的，并且本公开内容在此方面不受限制。

[0635] 本公开内容提供了用于治疗另外的疾病或病症的方法，例如与点突变相关或由该点突变引起的疾病或病症，所述点突变可以通过脱氨酶介导的基因编辑来校正。本文中描述了一些此类疾病，并且基于本公开内容，可以用本文提供的策略和融合蛋白治疗的其他合适的疾病对于本领域技术人员将是明显的。在下面列出示例性的合适的疾病和病症。应当理解的是，相应序列中特定位置或残基的编号取决于所使用的特定蛋白质和编号方案。编号可以不同，例如在成熟蛋白的前体和成熟蛋白自身中，并且物种间的序列差异可以影响编号。本领域技术人员将能够通过本领域公知的方法，例如通过序列比对和同源残基的确定，来鉴定任何同源蛋白质和相应编码核酸中的相应残基。示例性的合适疾病和病症包括但不限于囊性纤维化(参见例如Schwank et al.,Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients.Cell stem cell.2013;13:653-658;和Wu et.al.,Correction of a genetic disease in mouse via use of CRISPR-Cas9.Cell stem cell.2013;13:659-662,两者均未使用脱氨酶融合蛋白来校正遗传缺陷)；苯丙酮尿症——例如，苯丙氨酸羟化酶基因中的位置835(小鼠)或240(人)或同源残基处苯丙氨酸至丝氨酸突变(T>C突变)——例如参见McDonald et al.,Genomics.1997;39:402-405;巨血小板综合征(BSS)——例如血小板膜糖蛋白IX中位置55或同源残基处的苯丙氨酸至丝氨酸突变，或残基24或同源残基处的半胱氨酸至精氨酸(T>C突变)——参见例如Noris et al.,British Journal of Haematology.1997;97:312-320,and Ali et al.,Hematol.2014;93:381-384;表皮松解角化过度症(EHK)——例如，角蛋白1中位置160或161(若计算引发剂甲硫氨酸)或同源残基处的亮氨酸至脯氨酸突变(T>C突变)——参见，例如Chipev et al.,Cell.1992;70:821-828,还见www[dot]uniprot[dot]org的UNIPROT数据库中的登录号P04264;慢性阻塞性肺疾病(COPD)——例如，α_1-抗胰蛋白酶的加工形式中的位置54或55(若计算引发剂甲硫氨酸)处或未加工形式中的残基78或同源残基处的亮氨酸至脯氨酸突变(T>C突变)——参见例如

Poller et al.,Genomics.1993;17:740-743,还见UNIPROT数据库中的登录号P01011;4J型夏科-马里-图斯病——例如,FIG4中的位置41或同源残基处的异亮氨酸至苏氨酸突变(T>C突变)——例如参见Lenk et al.,PLoS Genetics.2011;7:e1002104;神经母细胞瘤(NB)——例如,胱天蛋白酶-9中的位置197或同源残基处的亮氨酸至脯氨酸突变(T>C突变)——例如,参见Kundu et al.,3Biotech.2013,3:225-234;血管性血友病(vWD)——例如,von Willebrand因子的加工形式中位置509或同源残基处或von Willebrand因子的未加工形式中位置1272或同源残基处的半胱氨酸至精氨酸突变(T>C突变)——参见例如Lavergne et al.,Br.J.Haematol.1992,还见UNIPROT数据库中的登录号P04275;82:66-72;先天性肌强直——例如,肌肉氯化物通道基因CLCN1中位置277位或同源残基处的半胱氨酸至精氨酸突变(T>C突变)——参见例如Weinberger et al.,The J.of Physiology.2012;590:3449-3464;遗传性肾淀粉样变性病——例如,载脂蛋白AII加工形式中的位置78或同源残基处或未加工形式中的位置101或同源残基处的终止密码子至精氨酸突变(T>C突变)——例如Yazaki et al.,Kidney Int.2003;64:11-16;扩张型心肌病(DCM)——参见例如FOXD4基因中的位置148或同源残基处的色氨酸至精氨酸突变(T>C突变),参见例如Minoretti et.al.,Int.J.of Mol.Med.2007;19:369-372;遗传性淋巴水肿——例如,VEGFR3酪氨酸激酶中的位置1035或同源残基处的组氨酸至精氨酸突变(A>G突变),参见例如Irrthum et al.,Am.J.Hum.Genet.2000;67:295-301;家族性阿尔茨海默氏病——例如,早老蛋白1中位置143或同源残基处的异亮氨酸至缬氨酸突变(A>G突变),参见例如Gallo et.al.,J.Alzheimer's disease.2011;25:425-431;朊病毒疾病——例如,朊病毒蛋白中的位置129或同源残基处的甲硫氨酸至缬氨酸突变(A>G突变)——参见例如Lewis et.al.,J.of General Virology.2006;87:2443-2449;慢性婴儿神经系统皮肤关节综合症(CINCA)——例如,cryopyrin中的位置570或同源残基处的酪氨酸至半胱氨酸突变(A>G突变)——参见例如Fujisawa et.al.Blood.2007;109:2903-2911;和结蛋白相关的肌病(DRM)——例如αβ晶体蛋白的位置120或同源残基处的精氨酸至甘氨酸突变(A>G突变)-例如参见Kumar et al.,J.Biol.Chem.1999;274:24137-24141。所有参考文献和数据库条目的全部内容通过引用并入本文。

[0636] 对于本领域技术人员而言明显的是,为了将如本文公开的Cas9:核酸编辑酶/域融合蛋白靶向到靶位点,例如包含待编辑的点突变的位点,通常有必要将Cas9:核酸编辑酶/域融合蛋白与引导RNA(例如sgRNA)一起共表达。如本文其他地方更详细地解释的,引导RNA通常包含允许Cas9结合的tracrRNA框架和引导序列,其对Cas9:核酸编辑酶/域融合蛋白赋予序列特异性。在一些实施方案中,引导RNA包含结构5'-[引导序列]guuuuagagcuagaaauagcaaguuaaaauaaggcuaguccguuaucaacuugaaaaaguggcaccgagucggugcuuuuu-3'(SEQ ID NO:618),其中引导序列包含与靶序列互补的序列。引导序列通常长为20个核苷酸。基于本公开内容,用于将Cas9:核酸编辑酶/域融合蛋白靶向到特定基因组靶位点的合适的引导RNA的序列对于本领域技术人员将是明显的。此类合适的引导RNA序列通常包含与待编辑靶核苷酸上游或下游50个核苷酸内的核酸序列互补的引导序列。以下提供了一些适用于将Cas9:核酸编辑酶/域融合蛋白靶向到特定靶序列的示例性引导RNA序列。

[0637] 碱基编辑效率

[0638] 本公开内容的一些方面是基于以下认识:本文提供的任何碱基编辑器能够修饰特

定的核苷酸碱基而不产生显著比例的插入/缺失。如本文所用，“插入/缺失”是指核酸内核苷酸碱基的插入或缺失。此类插入或缺失可导致基因编码区内的移码突变。在一些实施方案中，期望产生有效修饰(例如突变或脱氨基)核酸内的特定核苷酸而不在核酸中产生大量插入或缺失(即插入/缺失)的碱基编辑器。在某些实施方案中，相对于插入/缺失，本文提供的任何碱基编辑器能够产生更大比例的意图修饰(例如，点突变或脱氨基)。在一些实施方案中，本文提供的碱基编辑器能够产生大于1:1的意图点突变与插入/缺失的比率。在一些实施方案中，本文提供的碱基编辑器能够产生至少1.5:1、至少2:1、至少2.5:1、至少3:1、至少3.5:1、至少4:1、至少4.5:1、至少5:1、至少5.5:1、至少6:1、至少6.5:1、至少7:1、至少7.5:1、至少8:1、至少10:1、至少12:1、至少15:1、至少20:1、至少25:1、至少30:1、至少40:1、至少50:1、至少100:1、至少200:1、至少300:1、至少400:1、至少500:1、至少600:1、至少700:1、至少800:1、至少900:1、或至少1000:1或更多的意图的点突变与插入/缺失的比率。可以使用任何合适的方法，例如以下实施例中使用的方法，确定意图的突变和插入/缺失的数目。

[0639] 在一些实施方案中，本文提供的碱基编辑器能够限制核酸区域中插入/缺失的形成。在一些实施方案中，该区域在碱基编辑器靶向的核苷酸处或碱基编辑器靶向的核苷酸的2、3、4、5、6、7、8、9或10个核苷酸内的区域。在一些实施方案中，本文提供的任何碱基编辑器能够将核酸区域处插入/缺失的形成限制为小于1%、小于1.5%、小于2%、小于2.5%、小于3%、小于3.5%、小于4%、小于4.5%、小于5%、小于6%、小于7%、小于8%、小于9%、小于10%、小于12%、小于15%、或小于20%。在核酸区域处形成的插入/缺失的数目可以取决于核酸(例如，细胞基因组内的核酸)暴露于碱基编辑器的时间量。在一些实施方案中，在将核酸(例如，细胞基因组内的核酸)暴露于碱基编辑器的至少1小时，至少2小时，至少6小时，至少12小时，至少24小时，至少36小时，至少48小时，至少3天，至少4天，至少5天，至少7天，至少10天或至少14天后确定插入/缺失的数目或比例。

[0640] 本公开内容的一些方面是基于以下的认识：本文提供的任何碱基编辑器能够在核酸(例如受试者的基因组内的核酸)中有效产生意图的突变，例如点突变，而不产生大量非意图突变，例如非意图点突变。在一些实施方案中，意图的突变是由与gRNA结合的特定碱基编辑器产生的突变，其经特别设计以产生意图的突变。在一些实施方案中，意图的突变是与疾病或病症相关的突变。在一些实施方案中，意图的突变是与疾病或病症相关的胞嘧啶(C)至胸腺嘧啶(T)点突变。在一些实施方案中，意图的突变是与疾病或病症相关的鸟嘌呤(G)至腺嘌呤(A)点突变。在一些实施方案中，意图的突变是基因的编码区内的胞嘧啶(C)至胸腺嘧啶(T)点突变。在一些实施方案中，意图的突变是基因的编码区内的鸟嘌呤(G)至腺嘌呤(A)点突变。在一些实施方案中，意图的突变是点突变，其在基因的编码区内产生终止密码子，例如过早的终止密码子。在一些实施方案中，意图的突变是消除终止密码子的突变。在一些实施方案中，意图的突变是改变基因的剪接的突变。在一些实施方案中，意图的突变是改变基因的调控序列(例如，基因启动子或基因阻遏物)的突变。在一些实施方案中，本文提供的任何碱基编辑器能够产生大于1:1的意图突变与非意图突变(例如，意图点突变:非意图点突变)的比率。在一些实施方案中，本文提供的任何碱基编辑器能够产生至少1.5:1、至少2:1、至少2.5:1、至少3:1、至少3.5:1、至少4:1、至少4.5:1、至少5:1、至少5.5:1、至少6:1、至少6.5:1、至少7:1、至少7.5:1、至少8:1、至少10:1、至少12:1、至少15:1、至少20:1、

至少25:1、至少30:1、至少40:1、至少50:1、至少100:1、至少150:1、至少200:1、至少250:1、至少500:1、或至少1000:1或更多的意图突变与非意图突变(例如,意图点突变:非意图点突变)的比率。应当理解,本文"碱基编辑器效率"部分中描述的碱基编辑器的特征可以应用于任何融合蛋白,或使用本文提供的融合蛋白的方法。

[0641] 用于编辑核酸的方法

[0642] 本公开内容的一些方面提供了用于编辑核酸的方法。在一些实施方案中,该方法是用于编辑核酸的核碱基(例如,双链DNA序列的碱基对)的方法。在一些实施方案中,该方法包括以下步骤:a)使核酸的靶区域(例如,双链DNA序列)与包含碱基编辑器(例如,与胞苷脱氨酶域融合的Cas9域)的复合物和引导核酸(例如gRNA)接触,其中靶区域包含靶定的核碱基对,b)诱导所述靶区域的链分离,c)将靶区域的单链中的所述靶核碱基对的第一核碱基转化为第二核碱基;并且d)切割所述靶区域的不超过一条链,其中与第一核碱基碱基互补的第三核碱基由与第二核碱基互补的第四核碱基替换;并且该方法导致核酸中小于20%的插入/缺失形成。应当理解,在一些实施方案中,将步骤b省略。在一些实施方案中,第一核碱基是胞嘧啶。在一些实施方案中,第二核碱基是脱氨基的胞嘧啶或尿嘧啶。在一些实施方案中,第三核碱基是鸟嘌呤。在一些实施方案中,第四核碱基是腺嘌呤。在一些实施方案中,第一核碱基是胞嘧啶,第二核碱基是脱氨基的胞嘧啶或尿嘧啶,第三核碱基是鸟嘌呤,并且第四核碱基是腺嘌呤。在一些实施方案中,该方法导致小于19%,18%,16%,14%,12%,10%,8%,6%,4%,2%,1%,0.5%,0.2%或小于0.1%的插入/缺失形成。在一些实施方案中,该方法进一步包括用与第四核碱基互补的第五核碱基替换第二核碱基,从而产生意图编辑的碱基对(例如,C:G->T:A)。在一些实施方案中,第五核碱基是胸腺嘧啶。在一些实施方案中,编辑至少5%的意图碱基对。在一些实施方案中,编辑至少10%,15%,20%,25%,30%,35%,40%,45%或50%的意图碱基对。

[0643] 在一些实施方案中,靶核苷酸中意图产物与非意图产物的比率为至少2:1,5:1,10:1,20:1,30:1,40:1,50:1,60:1,70:1,80:1,90:1,100:1,或200:1或更多。在一些实施方案中,意图的点突变与插入/缺失的比率大于1:1,10:1,50:1,100:1,500:1,或1000:1或更多。在一些实施方案中,切割的单链(带切口的链)与引导核酸杂交。在一些实施方案中,切割的单链与包含第一核碱基的链相对。在一些实施方案中,碱基编辑器包括Cas9域。在一些实施方案中,第一碱基是胞嘧啶,而第二碱基不是G,C,A或T。在一些实施方案中,第二碱基是尿嘧啶。在一些实施方案中,第一碱基是胞嘧啶。在一些实施方案中,第二碱基不是G,C,A或T。在一些实施方案中,第二碱基是尿嘧啶。在一些实施方案中,碱基编辑器抑制编辑的链的碱基切除修复。在一些实施方案中,碱基编辑器保护或结合未编辑的链。在一些实施方案中,碱基编辑器包括UGI活性。在一些实施方案中,碱基编辑器包括切口酶活性。在一些实施方案中,意图编辑的碱基对在PAM位点的上游。在一些实施方案中,意图编辑的碱基对在PAM位点上游1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,或20个核苷酸。在一些实施方案中,意图编辑的碱基对在PAM位点的下游。在一些实施方案中,意图编辑的碱基对在PAM位点下游的1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,或20个核苷酸。在一些实施方案中,方法不需要规范的(例如,NGG)PAM位点。在一些实施方案中,核碱基编辑器包含接头。在一些实施方案中,接头的长度为1-25个氨基酸。在一些实施方案中,接头的长度为5-20个氨基酸。在一些实施方案中,接头的长度为10、11、12、13、14、15、16、17、18、19

或20个氨基酸。在一些实施方案中，靶区域包括靶窗，其中靶窗包括靶核碱基对。在一些实施方案中，靶窗包含1-10个核苷酸。在一些实施方案中，靶窗的长度为1-9、1-8、1-7、1-6、1-5、1-4、1-3、1-2或1个核苷酸。在一些实施方案中，靶窗长度是1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、17、18、19或20个核苷酸。在一些实施方案中，意图编辑的碱基对在靶窗内。在一些实施方案中，靶窗包括意图编辑的碱基对。在一些实施方案中，使用本文提供的任何碱基编辑器来进行该方法。在一些实施方案中，靶窗是脱氨基窗。

[0644] 在一些实施方案中，本公开内容提供了用于编辑核苷酸的方法。在一些实施方案中，本公开内容提供了用于编辑双链DNA序列的核碱基对的方法。在一些实施方案中，该方法包括a)使双链DNA序列的靶区域与包含碱基编辑器和引导核酸(例如，gRNA)的复合物接触，其中该靶区域包含靶核碱基对，b)诱导所述靶区域的链分离，c)将靶区域的单链中的所述靶核碱基对的第一核碱基转化为第二核碱基，d)切割所述靶区域的不超过一条链，其中与第一核碱基碱基互补的第三核碱基由与第二核碱基互补的第四核碱基替换，并且第二核碱基用与第四核碱基互补的第五核碱基替换，从而产生意图编辑的碱基对，其中产生意图编辑的碱基对的效率碱基对为至少5％。应当理解，在一些实施方案中，将步骤b省略。在一些实施方案中，编辑至少5％的意图碱基对。在一些实施方案中，编辑至少10％，15％，20％，25％，30％，35％，40％，45％或50％的意图碱基对。在一些实施方案中，该方法导致小于19％，18％，16％，14％，12％，10％，8％，6％，4％，2％，1％，0.5％，0.2％或小于0.1％的插入/缺失形成。在一些实施方案中，在靶核苷酸处意图产物与非意图产物的比率为至少2:1，5:1，10:1，20:1，30:1，40:1，50:1，60:1，70:1，80:1，90:1，100:1，或200:1或更多。在一些实施方案中，意图的点突变与插入/缺失形成的比率大于1:1，10:1，50:1，100:1，500:1或1000:1或更多。在一些实施方案中，切割的单链与引导核酸杂交。在一些实施方案中，切割的单链与包含第一核碱基的链相对。在一些实施方案中，第一碱基是胞嘧啶。在一些实施方案中，第二核碱基不是G，C，A或T。在一些实施方案中，第二碱基是尿嘧啶。在一些实施方案中，碱基编辑器抑制编辑的链的碱基切除修复。在一些实施方案中，碱基编辑器保护或结合未编辑的链。在一些实施方案中，核碱基编辑器包含UGI活性。在一些实施方案中，核碱基编辑包括切口酶活性。在一些实施方案中，意图编辑的碱基对在PAM位点的上游。在一些实施方案中，意图编辑的碱基对在PAM位点上游的1，2，3，4，5，6，7，8，9，10，11，12，13，14，15，16，17，18，19或20个核苷酸。在一些实施方案中，意图编辑的碱基对在PAM位点的下游。在一些实施方案中，意图编辑的碱基对在PAM位点下游的1，2，3，4，5，6，7，8，9，10，11，12，13，14，15，16，17，18，19，或20个核苷酸。在一些实施方案中，该方法不需要规范的(例如，NGG)PAM位点。在一些实施方案中，核碱基编辑器包含接头。在一些实施方案中，接头的长度为1-25个氨基酸。在一些实施方案中，接头的长度为5-20个氨基酸。在一些实施方案中，接头的长度为10、11、12、13、14、15、16、17、18、19或20个氨基酸。在一些实施方案中，靶区域包括靶窗，其中靶窗包括靶核碱基对。在一些实施方案中，靶窗包含1-10个核苷酸。在一些实施方案中，靶窗的长度为1-9、1-8、1-7、1-6、1-5、1-4、1-3、1-2或1个核苷酸。在一些实施方案中，靶窗的长度为1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、17、18、19或20个核苷酸。在一些实施方案中，意图编辑的碱基对在靶窗内发生。在一些实施方案中，靶窗包括意图编辑的碱基对。在一些实施方案中，核碱基编辑器是本文提供的任何碱基编辑器。

[0645] 药物组合物

[0646] 在一些实施方案中，本文描述的任何融合蛋白，gRNA和/或复合物作为药物组合物的一部分提供。在一些实施方案中，药物组合物包含本文提供的任何融合蛋白。在一些实施方案中，药物组合物包含本文提供的任何复合物。在一些实施方案中，药物组合物包含核糖核蛋白复合物，其包含与gRNA和阳离子脂质形成复合物的RNA引导的核酸酶(例如，Cas9)。在一些实施方案中，药物组合物包含gRNA，核酸可编程DNA结合蛋白，阳离子脂质和药学上可接受的赋形剂。药物组合物可任选地包含一种或多种其他治疗活性物质。

[0647] 在一些实施方案中，将本文提供的组合物施用于受试者，例如人受试者，以在受试者内实现靶向的基因组修饰。在一些实施方案中，细胞获自受试者，并与本文提供的任何药物组合物接触。在一些实施方案中，将从受试者中取出并与药物组合物离体接触的细胞重新引入受试者中，任选地在细胞中已经实现或检测到期望的基因组修饰之后。递送包含核酸酶的药物组合物的方法是已知的，并且例如记载于美国专利No.6,453,242;6,503,717;6,534,261;6,599,692;6,607,882;6,689,558;6,824,978;6,933,113;6,979,539;7,013,219;和7,163,824，所有专利的公开内容通过引用完整并入本文。尽管本文提供的药物组合物的描述主要针对适合于对人施用的药物组合物，但是熟练技术人员应理解，此类组合物通常适合于对所有种类的动物或生物体施用。为了使组合物适合于对各种动物施用，适合于对人施用的药物组合物的修改是公知的，并且普通技术的兽医药理师可以仅通过普通(若有的话)的实验来设计和/或进行此类修改。意图接受药物组合物施用的受试者包括但不限于人和/或其他灵长类;哺乳动物，家养动物，宠物和商业有关的哺乳动物，例如牛，猪，马，绵羊，猫，狗，小鼠和/或大鼠;和/或禽类，包括商业相关的禽类，例如鸡，鸭，鹅和/或火鸡。

[0648] 本文所述的药物组合物的制剂可以通过药理学领域中已知的或以后开发的任何方法来制备。通常，此类制备方法包括以下步骤:将一种或多种活性成分与赋形剂和/或一种或多种其他辅助成分结合，然后，若必要和/或期望的话，将产品成形和/或包装成期望的单剂量或多剂量单位。

[0649] 药物制剂可另外包含药学上可接受的赋形剂，如本文所用，其包括任何和所有溶剂，分散介质，稀释剂或其他液体媒介物，分散或悬浮助剂，表面活性剂，等张剂，增稠或乳化剂，防腐剂，固体粘合剂，润滑剂等，如适合于期望的特定剂型。Remington's The Science and Practice of Pharmacy,21stEdition,A.R.Gennaro(Lippincott,Williams&Wilkins,Baltimore,MD,2006;其通过引用完整并入)公开了用于配制药物组合物的各种赋形剂及其制备的已知技术。关于用于产生包含核酸酶的药物组合物的其他合适的方法，试剂，赋形剂和溶剂，还参见PCT申请PCT/US2010/055131(公开号WO2011053982A8，于2010年11月2日提交)，其通过引用完整并入本文。除非任何常规的赋形剂介质与某种物质或其衍生物不相容，例如通过产生任何不希望的生物学作用或以有害的方式与药物组合物的任何其他成分相互作用，否则意图其使用在本公开内容的范围内。

[0650] 在一些实施方案中，根据本发明的组合物可以用于治疗多种疾病，病症和/或病况中的任何一种，包括但不限于以下一种或多种:自身免疫性病症(例如糖尿病，狼疮，多发性硬化症，银屑病，类风湿关节炎);炎性病症(例如关节炎，盆腔炎性疾病);传染病(例如病毒感染(例如HIV,HCV,RSV)，细菌感染，真菌感染，败血症);神经系统疾病(例如阿尔茨海默氏病，亨廷顿氏病，自闭症，迪谢内肌营养不良);心血管疾病(例如动脉粥样硬化，高胆固醇血

症,血栓形成,凝固病症,血管生成性病症如黄斑变性);增殖性病症(例如癌症,良性新生物);呼吸系统病症(例如慢性阻塞性肺疾病);消化病症(例如炎性肠病,溃疡);肌肉骨骼病症(例如纤维肌痛,关节炎);内分泌,代谢和营养病症(例如糖尿病,骨质疏松);泌尿病症(例如肾疾病);心理障碍(例如抑郁,精神分裂症);皮肤病症(例如伤口,湿疹);血液和淋巴疾病(例如贫血,血友病);等等

[0651] 试剂盒、载体、细胞

[0652] 本公开内容的一些方面提供了包含核酸构建体的试剂盒,所述核酸构建体包含(a)编码如本文提供的Cas9蛋白或Cas9融合蛋白的核苷酸序列;和(b)驱动(a)序列表达的异源启动子。在一些实施方案中,试剂盒进一步包含编码引导RNA主链的表达构建体,其中该构建体包含克隆位点,其定位成允许将与靶序列相同或互补的核酸序列克隆到引导RNA主链中。

[0653] 本公开内容的一些方面提供了编码如本文提供的融合蛋白的Cas9蛋白的多核苷酸。本公开内容的一些方面提供了包含此类多核苷酸的载体。在一些实施方案中,载体包含驱动多核苷酸表达的异源启动子。

[0654] 本公开内容的一些方面提供了细胞,其包含Cas9蛋白,融合蛋白,编码融合蛋白的核酸分子,包含Cas9蛋白和gRNA的复合物和/或载体,如本文提供。

[0655] 上述报告系统的示例性实施方案的描述仅出于说明目的而提供,并不意味着限制。本公开内容还包括其他报告系统,例如上面详细描述的示例性系统的变型。

实施例

[0656] 实施例1:Cas9脱氨酶融合蛋白

[0657] 产生了许多Cas9:脱氨酶融合蛋白,并且表征了所产生的融合物的脱氨酶活性。测试了以下脱氨酶:

[0658] 人AID(hAID):

[0659] MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPYLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGLLD(SEQ ID NO:49)

[0660] 人AID-DC(hAID-DC,hAID的截短形式,具有7倍增加的活性):

[0661] MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILL(SEQ ID NO:50)

[0662] 大鼠APOBEC1(rAPOBEC1):

[0663] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK(SEQ ID NO:76)

[0664] 人APOBEC1(hAPOBEC1)

[0665] MTSEKGPSTGDPTLRRRIEPWEFDVFYDPRELRKEACLLYEIKWGMSRKIWRSSGKNTTNHVEVNFIKKFTSERDFHPSMSCSITWFLSWSPCWECSQAIREFLSRHPGVTLVIYVARLFWHMDQQNRQGLRDLVNSGVTIQIM

RASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKISRRWQNHLTFFRLHLQNCHYQTIPP HILLATGLIHPSVAWR(SEQ ID NO:74)

[0666] 海七鳃鳗(八目鳗)CDA1(pmCDA1):

[0667] MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQSGTERGIHAE IFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLKIWACKLYYEKNARNQIGLWNLRDN GVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKRRSELSIMIQVKILHTTKSPAV(SEQ ID NO: 81)

[0668] 人APOBEC3G(hAPOBEC3G):

[0669] MELKYHPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPKVTLTIFVARLYYFW DPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEPWNNLPKYYILLHIMLGEILRHSMDPPTFT FNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRV TCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVDHQG CPFQPWDGLDEHSQDLSGRLRAILQNQEN(SEQ ID NO:60)

[0670] 对ssDNA的脱氨酶活性。采用用于脱氨基的基于USER(尿嘧啶特异性切除试剂)酶的测定法来测试单链DNA(ssDNA)底物上各种脱氨基酶的活性。从New England Biolabs获得USER酶。ssDNA底物在不同位置处提供有靶胞嘧啶残基。ssDNA胞嘧啶靶残基的脱氨基导致靶胞嘧啶转化为尿嘧啶。USER酶切除尿嘧啶碱基,并在该位置处切割ssDNA主链,从而将ssDNA底物切成两个较短的DNA片段。在某些测定法中,ssDNA底物在一个末端用染料标记,例如用5’Cy3标记物(以下方案中的*)。在对链进行脱氨基,切除和切割后,可以对底物进行电泳,并且可以通过检测标记物来显现底物和从其中释放的任何片段。在Cy5是图像的情况下,通过成像仅可见具有标记物的片段。

[0671] 在一种USER酶测定法中,使用了与所测试的各种脱氨酶的靶序列相匹配的ssDNA底物。将编码测试的脱氨基酶的表达盒插入先前在实验室中使用过的CMV主链质粒(Addgene质粒52970)。根据制造商的建议,使用TNT快速偶联转录/翻译系统(Promega)表达脱氨酶蛋白。温育90分钟后,将5mL裂解物与5’Cy3标记的ssDNA底物和1个单位的USER酶(NEB)温育3小时。将DNA溶解在10%TBE PAGE凝胶上,并使用Cy染料成像对DNA进行成像。图41中显示了USER酶测定法的示意图。

[0672] 图1显示了测试的脱氨基酶在ssDNA底物如Doench 1,Doench 2,G7□和VEGF靶物2上的脱氨酶活性。rAPOBEC1酶在具有规范NGG PAM,但没有阴性对照非规范NNN PAM的单链DNA底物上表现出大量的脱氨基。

[0673] 产生具有APOBEC家族脱氨基酶的Cas9融合蛋白。构建以下融合结构并在ssDNA上测试:

[0674] rAPOBEC1-GGS-dCas9一级序列

[0675] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTN
KHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLY
HHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRL
YVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKGGS*DKKYSI*
GLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR
TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVD
KLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLI
ALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGY
AGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLG
ELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW
NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEG
MRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASL
GTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMK
QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKED
IQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMAR
ENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDM
YVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMK
NYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSR
MNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTA
LIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANG
EIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSS
FEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKY
VNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVL
SAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQS
ITGLYETRIDLSQLGGD (SEQ ID NO: 149)

[0676] rAPOBEC1-(GGS)$_3$-dCas9一级序列

[0677] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTN
KHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLY
HHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRL
YVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKGGSGGSGG
S*MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET*
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKK
NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLA
AKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF
DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS
EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKV
KYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDS
LTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEV

VKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVA QILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTE ITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSK ESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 150)

dCas9-GGS-rAPOBEC1

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD NSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAK NLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFD QSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIP HQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSE ETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVK YVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDR FNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDD KVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLT FKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVI EMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQN GRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVV KKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQI LDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESI LPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIM ERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELAL PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANL DKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT LIHQSITGLYETRIDLSQLGGDGGSMSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKE TCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCG ECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWR NFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSC HYQRLPPHILWATGLK (SEQ ID NO: 151)

dCas9-GGS_3-rAPOBEC1

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD NSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAK NLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFD QSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIP HQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSE ETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVK YVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDR FNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDD KVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLT

[0678]

[0679] *FKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVI EMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQN GRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVV KKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQI LDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESI LPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIM ERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELAL PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANL DKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT LIHQSITGLYETRIDLSQLGGD*GGSGGSGGSMSSETGPVAVDPTLRRRIEPHEFEVFFDP RELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFL SWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQES GYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTI ALQSCHYQRLPPHILWATGLK (SEQ ID NO: 152)

[0680] rAPOBEC1-XTEN-*dCas9*一级序列

[0681] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTN KHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLY HHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRL YVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPGTS ESATPES*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFD SGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIE GDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYAD LFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYN ELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIS GVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLI HDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHK PENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYL YYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVP SEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITK HVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA YLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFF KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGG FSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKEL LGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQK GNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVI LADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTK EVLDATLIHQSITGLYETRIDLSQLGGD* (SEQ ID NO: 153)

[0682] 图2显示了N端脱氨酶融合物在单链DNA底物上显示出显著的活性。因此，仅选择N端结构进行进一步的实验。

[0683] 图3显示了脱氨酶-dCas9:sgRNA复合物的双链DNA底物的结合。产生了许多双链脱氨酶底物序列。下面提供了序列。在这些序列中鉴定出根据图3的结构(36bp:带下划线，sgRNA靶序列:粗体;PAM:框示;21bp:斜体)。用5'-Cy3标记物标记所有底物：

2:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGTCCCGCGGATTTA
TTTATTTAATGGATGACCTCTGGATCCATGGAC-3' (SEQ ID NO: 85)
3:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCTTCCGCGGATTT
ATTTATTTATGGATGACCTCTGGATCCATGGAC-3' (SEQ ID NO: 86)
4:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCCTTCCGCGGATT
TATTTATTATGGATGACCTCTGGATCCATGGAC-3' (SEQ ID NO: 87)
5:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCCATTCCGCGGAT
TTATTTATTTGGATGACCTCTGGATCCATGGAC-3' (SEQ ID NO: 88)
6:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCCTATTCCGCGGA
TTTATTTATTGGATGACCTCTGGATCCATGGAC-3' (SEQ ID NO: 89)
7:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCCTTATTCCGCGG
ATTTATTTATGGATGACCTCTGGATCCATGGAC-3' (SEQ ID NO: 90)
8:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCCATTATTCCGCG
GATTTATTTGGATGACCTCTGGATCCATGGAC-3' (SEQ ID NO: 91)
9:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCCTATTATTCCGC
GGATTTATTGGATGACCTCTGGATCCATGGAC-3' (SEQ ID NO: 92)
10:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCCATTATATTCCG
[0684] CGGATTTATGGATGACCTCTGGATCCATGGAC-3' (SEQ ID NO: 93)
11:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCCTATTATATTCC
GCGGATTTAGGATGACCTCTGGATCCATGGAC-3' (SEQ ID NO: 94)
12:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCCTTATTATATTC
CGCGGATTTGGATGACCTCTGGATCCATGGAC-3' (SEQ ID NO: 95)
13:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCCATTATTATATT
CCGCGGATTGGATGACCTCTGGATCCATGGAC-3' (SEQ ID NO: 96)
14:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCCTATTATTATAT
TCCGCGGATGGATGACCTCTGGATCCATGGAC-3' (SEQ ID NO: 97)
15:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCCATTATTATTAT
TACCGCGGAGGATGACCTCTGGATCCATGGAC-3' (SEQ ID NO: 98)
18:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCCATTATTATTAT
TATTACCGCGGATGACCTCTGGATCCATGGAC-3' (SEQ ID NO: 99)
"-" :
GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGTAATATTAATTTATT
TATTTAAGGATGACCTCTGGATCCATGGAC-3' (SEQ ID NO: 100)
8U:GTAGGTAGTTAGGATGAATGGAAGGTTGGTGTAGATTATTATCUGC
GGATTTATGGATGACCTCTGGATCCATGGACAT-3' (SEQ ID NO: 101)

[0685] *在除"8U"以外的所有底物中,图3中的顶部链是本文规定的序列的互补物。在"8U"的情况下,存在有与U相对的"G"。

[0686] 图4显示了双链DNA脱氨基测定法的结果。将融合物表达,并通过Ni-NTA和sepharose层析用N末端His6标签纯化。为了评估在dsDNA底物上的脱氨基,将先前图上显示的各种dsDNA底物以1:8dsDNA:融合蛋白比率温育,并在37℃温育2小时。一旦融合物的dCas9部分与DNA结合,它阻断USER酶接近DNA。因此,温育后将融合蛋白变性,并将dsDNA在旋转柱上纯化,然后与USER酶一起温育45分钟,并在10%TBE-尿素凝胶上解析得到的DNA底物和底物片段。

[0687] 图5证明了Cas9融合物可以靶向双链DNA靶序列的位置3-11(根据图3中的示意图

编号)。上凝胶:1μM rAPOBEC1-GGS-dCas9,125nM dsDNA,1eq sgRNA。中间凝胶:1μM rAPOBEC1-$(GGS)_3$-dCas9,125nM dsDNA,1eq sgRNA。下部凝胶:1.85μM rAPOBEC1-XTEN-dCas9,125nM dsDNA,1eq sgRNA。基于来自这些凝胶的数据,位置3-11(根据图3中的编号)充分暴露于要由所测试的融合蛋白靶向的脱氨酶的活性。脱氨酶接近其他位置最可能被dCas9蛋白阻断。

[0688] 数据进一步指示仅3个氨基酸(GGS)的接头对于允许脱氨酶接近DNA的单链部分不是最佳的。9个氨基酸的接头[$(GGS)_3$](SEQ ID NO:610)和最有结构的16个氨基酸的接头(XTEN)允许更有效的脱氨基。

[0689] 图6证明了脱氨酶活性需要正确的引导RNA,例如正确的sgRNA。凝胶显示了将脱氨酶与dCas9融合,脱氨酶变为序列特异性的(例如,使用与eGFPsgRNA的融合不导致脱氨基),并且还赋予脱氨酶使dsDNA脱氨基的能力。脱氨酶的天然底物是ssDNA,并且当不添加sgRNA时就不发生脱氨基。这与APOBEC脱氨酶本身不使dsDNA脱氨基的报告知识相一致。数据指示了Cas9在短窗内打开双链DNA螺旋,暴露单链DNA,其然后可接近APOBEC脱氨酶以进行胞苷脱氨基。下面提供了所用的sgRNA序列。序列(36bp:带下划线,sgRNA靶序列:粗体;PAM:框示;21bp:斜体)

[0690] DNA序列8:

[0691] 5'-Cy3-GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCC**ATTATTCC GCGGATTTATT** TTGG*ATGACCTCTGGATCCATGGAC*-3' (SEQ ID NO: 102)

[0692] 正确sgRNA序列(部分3'序列):

[0693] 5'-**AUUAUUCCGCGGAUUUAUUU**GUUUUAGAGCUAG...-3' (SEQ ID NO: 103)

[0694] eGFP sgRNA序列(部分3'-序列):

[0695] 5'-**CGUAGGCCAGGGUGGUCACG**GUUUUAGAGCUAG...-3' (SEQ ID NO: 104)

[0696] 实施例2:DNA靶序列的脱氨基

[0697] 示例性的脱氨基靶物。可以将本文所述的dCas9:脱氨酶融合蛋白在体外或离体递送至细胞或体内递送至受试者,并且当靶核苷酸就PAM而言位于位置3-11中时,可用于实现C至T或G至A的转变。示例性的脱氨基靶物包括但不限于以下:CCR5截短:可以将编码CCR5的Q93,Q102,Q186,R225,W86,或Q261的任何密码子进行脱氨基以生成终止密码子,这导致CCR5的非功能性截短,在HIV治疗中得到应用。APOE4突变:编码C11R和C57R突变体APOE4蛋白质的突变体密码子可以进行脱氨基以恢复为野生型氨基酸,在阿尔茨海默氏病治疗中得到应用。eGFP截短:编码Q158,Q184,Q185的任何密码子可以脱氨基以产生终止密码子,或者编码M1的密码子可以脱氨基以编码I,它们都导致eGFP荧光丧失,在报告系统中得到应用。eGFP恢复:编码T65A或Y66C突变体GFP(其不表现出实质性荧光)的突变体密码子可以脱氨基以恢复野生型氨基酸并赋予荧光。PIK3CA突变:编码K111E突变体PIK3CA的突变体密码子可以脱氨基以恢复野生型氨基酸残基,在癌症中得到应用。CTNNB1突变:编码T41A突变体CTNNB1的突变体密码子可以脱氨基以恢复野生型氨基酸残基,在癌症中得到应用。HRAS突

变:编码Q61R突变体HRAS的突变体密码子可以脱氨基以恢复野生型氨基酸残基,在癌症中得到应用。P53突变:可以将编码Y163C,Y236C或N239D突变体p53的任何突变体密码子进行脱氨基以编码野生型氨基酸序列,在癌症中得到应用。

[0698] 在图7和8中证明了使双链DNA中的这些靶序列脱氨基的可行性。图7显示了脱氨酶-dCas9:sgRNA复合物对体内靶序列的靶DNA结合的机制。

[0699] 图8显示了示例性疾病相关靶序列的成功脱氨基。顶部凝胶:CCR5 Q93:位置10中的编码链靶物(位置2、5、6、8、9处的潜在脱靶);CCR5 Q102:位置9中的编码链靶物(位置1、12、14处的潜在脱靶);CCR5 Q186:位置9中的编码链靶物(位置1、5、15处的潜在脱靶);CCR5 R225:位置6中的编码链靶物(无潜在脱靶);eGFP Q158:位置5中的编码链靶物(位置1、13、16处的潜在脱靶);eGFP Q184/185:位置4和7中的编码链靶物(位置3、12、14、15、16、17、18处的潜在脱靶);eGFP M1:位置12中的模板链靶物(位置2、3、7、9、11处的潜在脱靶)(在小程度上靶位置7和9);eGFP T65A:位置7中的模板链靶物(位置1、8、17处的潜在脱靶);PIK3CA K111E:位置2中的模板链靶物(位置5、8、10、16、17处的潜在脱靶);PIK3CA K111E:位置13中的模板链靶物(位置11、16、19处的潜在脱靶)X。下部凝胶:CCR5 W86:位置2和3中的模板链靶物(位置1、13处的潜在脱靶)X;APOE4 C11R:位置11中的编码链靶物(位置7、13、16、17处的潜在脱靶);APOE4 C57R:位置5中的编码链靶物)(位置7、8、12处的潜在脱靶);eGFP Y66C:位置11中的模板链靶物(位置1、4,6、8、9、16处的潜在脱靶);eGFP Y66C:位置3中的模板链靶物(位置1、8、17处的潜在脱靶);CCR5 Q261:位置10中的编码链靶物(位置3、5、6、9、18处的潜在脱靶);CTNNB1 T41A:位置7中的模板链靶物(位置1、13、15、16处的潜在脱靶)X;HRAS Q61R:位置6中的模板链靶物(位置1、2、4、5、9、10、13处的潜在脱靶);p53 Y163C:位置6中的模板链靶物(位置2、13、14处的潜在脱靶);p53 Y236C:位置8中的模板链靶物(位置2、4处的潜在脱靶);p53 N239D:位置4中的模板链靶物(位置6、8处的潜在脱靶)。下文提供了疾病靶物的示例性DNA序列(PAMs(5'-NGG-3')并且靶位置是框示的)

CCR5 **Q93**: 5'-Cy3-GTAGGTAGTTAGGATGAATGGAAGGTTGGTAACTAT**GCTGCCGCC**
CAGTGGGACTT[TGG]*AAATACAATGTGTCAACTCTT*-3'(SEQ ID NO: 105)
CCR5 **Q102**: 5'-Cy3-GTAGGTAGTTAGGATGAATGGAAGGTTGGTAAAATA**CAATGTGT**
CAACTCTTGACA[GGG]*CTCTATTTTATAGGCTTCTTC*-3'(SEQ ID NO: 106)
CCR5 **Q186**: 5'-Cy3-GTAGGTAGTTAGGATGAATGGAAGGTTGGTATTTTC**CATACAGT**
CAGTATCAATTC[TGG]*AAGAATTTCCAGACATTAAAG*-3'(SEQ ID NO: 107)
CCR5 **R225**: 5'-Cy3-GTAGGTAGTTAGGATGAATGGAAGGTTGGTAGCTTC**GGTGT**[C]**GA**
AATGAGAAGAAG[AGG]*CACAGGGCTGTGAGGCTTATC*-3'(SEQ ID NO: 108)
CCR5 **W86**: 5'-Cy3-GTAGGTAGTTAGGATGAATGGAAGGTTGGTAGTGAG[C]**CCAGAAGG**
GGACAGTAAGA[AGG]*AAAAACAGGTCAGAGATGGCC*-3'(SEQ ID NO: 109)
CCR5 **Q261**: 5'-Cy3-GTAGGTAGTTAGGATGAATGGAAGGTTGGTATCCTG**AACACCTT**
C[C]**AGGAATTCTT**[TGG]*CCTGAATAATTGCAGTAGCTC*-3'(SEQ ID NO: 110)

[0700]

APOE4 **C11R**: 5'-Cy3-GTAGGTAGTTAGGATGAATGGAAGGTTGGTAGACAT**GGAGGAC**
GTG[C]**GCGGCCGCC**[TGG]*TGCAGTACCGCGGCGAGGTGC*-3'(SEQ ID NO: 111)
APOE4 **C57R**: 5'-Cy3-GTAGGTAGTTAGGATGAATGGAAGGTTGGTACTGCA**GAAG**[C]**GC**
CTGGCAGTGTACC[AGG]*CCGGGGCCCGCGAGGGCGCCG*-3'(SEQ ID NO: 112)
eGFP **Q158**: 5'-Cy3-GTAGGTAGTTAGGATGAATGGAAGGTTGGTAGCCGA**CAAG**[C]**AGA**
AGAACGGCATCA[AGG]*TGAACTTCAAGATCCGCCACA*-3'(SEQ ID NO: 113)
eGFP **Q184/185**:
5'-Cy3-GTAGGTAGTTAGGATGAATGGAAGGTTGGTAACCAC**TAC**[C]**AG**[C]**AGAACACCCCCAT**[CG]
[G]*CGACGGCCCCGTGCTGCTGCC*-3'(SEQ ID NO: 114)
eGFP **M1**: 5'-Cy3-GTAGGTAGTTAGGATGAATGGAAGGTTGGTACCTCG**CCCTTGCTCA**
C[C]**ATCTCGAGT**[CGG]*CCGCCAGTGTGATGGATATCT*-3'(SEQ ID NO: 115)
eGFP **T65A**: 5'-Cy3-GTAGGTAGTTAGGATGAATGGAAGGTTGGTACACGC**GTAGG**[C]**CA**
GGGTGGTCACG[AGG]*GTGGGCCAGGGCACGGGCAGC*-3'(SEQ ID NO: 116)
eGFP **Y66C**: 5'-Cy3-GTAGGTAGTTAGGATGAATGGAAGGTTGGTAAAGCA**CTGCACTC**
CG[C]**AGGTCAGGG**[TGG]*TCACGAGGGTTGGCCAGGGCA*-3'(SEQ ID NO: 117)
eGFP **Y66C**: 5'-Cy3-GTAGGTAGTTAGGATGAATGGAAGGTTGGTACACTC**CG**[C]**AGGTC**
AGGGTGGTCACG[AGG]*GTTGGCCAGGGCACGGGCAGG*-3'(SEQ ID NO: 118)
PIK3CA **K111E**: 5'-Cy3-GTAGGTAGTTAGGATGAATGGAAGGTTGGTAGGATC**T**[C]**TTC**
TTCACGGTTGCCTAC[TGG]*TTCAATTACTTTTAAAAATGG*-3'(SEQ ID NO: 119)
PIK3CA **K111E**: 5'-Cy3-GTAGGTAGTTAGGATGAATGGAAGGTTGGTATTCTC**GATTG**
AGGATC[T]**CTTCTTCA**[CGG]*TTGCCTACTGGTTCAATTACT*-3'(SEQ ID NO: 120)

[0701]

CTNNB1 **T41A**: 5'-Cy3-GTAGGTAGTTAGGATGAATGGAAGGTTGGTAAGGAG**CTGTGG**
CAGTGGCACCAGAA[TGG]*ATTCCAGAGTCCAGGTAAGAC*-3'(SEQ ID NO: 121)
HRAS **Q61R**: 5'-Cy3-GTAGGTAGTTAGGATGAATGGAAGGTTGGTAGTACT**CCTCC**[C]**GG**
CCGGCGGTATCC[AGG]*ATGTCCAACAGGCACGTCTCC*-3'(SEQ ID NO: 122)
p53 **Y163C**: 5'-Cy3-GTAGGTAGTTAGGATGAATGGAAGGTTGGTATGACT**GCTTG**[C]**AG**
ATGGCCATGGCG[CGG]*ACGCGGGTGCCGGGCGGGGGT*-3'(SEQ ID NO: 123)
p53 **Y236C**:
5'-Cy3-GTAGGTAGTTAGGATGAATGGAAGGTTGGTACTGTT**ACACATG**[C]**AGTTGTAGTGGA**[TG]
[G]*TGGTACAGTCAGAGCCAACCT*-3'(SEQ ID NO: 124)
p53 **N239D**: 5'-Cy3-GTAGGTAGTTAGGATGAATGGAAGGTTGGTAGGAAC**TGT**[C]**CACAC**
ATGTAGTTGTAG[TGG]*ATGGTGGTACAGTCAGAGCCA*-3'(SEQ ID NO: 125)

[0702] 实施例3:尿嘧啶糖基化酶抑制剂融合物改善脱氨基效率

[0703] 可以通过将尿嘧啶糖基化酶抑制剂(UGI)与dCas9:脱氨基融合蛋白融合来显著改善dCas9:脱氨基融合蛋白在哺乳动物细胞中的直接可编程核碱基编辑效率。

[0704] 图9显示了使用rAPOBEC1-XTEN-dCas9在人HEK293细胞中的体外C→T编辑效率:

[0705] rAPOBEC1-**XTEN**-[dCas9]-*NLS*一级序列

[0706] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTN
KHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLY
HHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRL
YVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK**SGSETPGTS**
ESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF
DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIE
GDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYAD
LFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFD
NGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMT
RKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELT
KVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGV
EDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLF
DDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ

[0707] NGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHV
AQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFK
TEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFS
KESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKG
NELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILA
DANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEV
LDATLIHQSITGLYETRIDLSQLGGDSGGS*PKKKRKV* (SEQ ID NO: 126)

[0708] 前原间隔物序列如下：

[0709] EMX1: 5'- GAGTC_5C_6GAGC_{10}AGAAGAAGAAGGG -3' (SEQ ID NO: 127)

[0710] FANCF: 5'- GGAAT$C_6C_7C_8$TTC_{11}TGCAGCACCTGG -3' (SEQ ID NO: 128)

[0711] HEK293位点2: 5'- GAAC_4AC_6AAAGC_{11}ATAGACTGCGGG -3' (SEQ ID NO: 129)

[0712] HEK293位点3: 5'- GG$C_3C_4C_5$AGAC_9TGAGCACGTGATGG -3' (SEQ ID NO: 130)

[0713] HEK293位点4: 5'- GGC_3AC_5TGC_8GGC_{11}TGGAGGTGGGGG -3' (SEQ ID NO: 735)

[0714] RNF2: 5'- GT$C_3AT C_6$TTAGTCATTACCTGAGG -3' (SEQ ID NO: 132)

[0715] *PAM是框示的，靶窗(位置3-11)内的C残基是加数字且粗体的。

[0716] 图10证明了当UGI域与rAPOBEC1:dCas9融合蛋白融合时大大增强HEK293T细胞的相同原间隔物序列上的C→T编辑效率。

[0717] rAPOBEC1-**XTEN**-dCas9-**UGI**-*NLS*一级序列

[0718] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTN
KHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLY
HHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRL
YVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK**SGSETPGTS**
ESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF
DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIE
GDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYAD
LFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFD
NGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMT
RKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELT
KVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGV
EDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLF
DDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHV
AQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFK
TEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFS

[0719] KESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKG
NELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILA
DANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEV
LDATLIHQSITGLYETRIDLSQLGGDSGGS**TNLSDIIEKETGKQLVIQESILMLPEEVEE**
VIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGG
S*PKKKRKV* (SEQ ID NO: 133)

[0720] 显示了图9和图10中的百分比，来自对靶序列的两条链进行测序。因为仅链之一是用于脱氨基的底物，此测定法中最大可能的脱氨基值为50％。因此，脱氨基效率是表中显示的百分比的两倍。例如，50％的值指100％的双链靶序列的脱氨基。

[0721] 将尿嘧啶糖基化酶抑制剂(UGI)与dCas9:脱氨酶融合蛋白融合(例如rAPOBEC1-XTEN-dCas9-[UGI]-NLS)时，观察到细胞中编辑效率的显著提高。该结果指示在哺乳动物细胞中，切割U:G碱基对中的尿嘧啶碱基的DNA修复机制是DNA编辑中的限速过程。将UGI栓系到dVas9:脱氨酶融合蛋白大大提高了编辑产率。

[0722] 在没有UGI的情况下，人细胞中的典型编辑效率在约2-14％的产率范围内(图9和图10，“XTEN”条目)。在使用UGI的情况下(图10，“UGI”条目)，可以在约6-40％的范围内观察到编辑。因此，使用UGI融合物比通过HDR校正点突变的当前替代方法(该方法除了校正点突变外还创建过多的插入/缺失)更为有效。没有观察到源自Cas9:脱氨酶:UGI融合物的处理的插入/缺失。

[0723] 实施例4:基因组DNA中靶核苷酸的直接可编程转化而无双链DNA切割

[0724] 当前的基因组编辑技术在感兴趣的靶基因座处引入双链DNA断裂，作为基因校正的第一步[39,40]。尽管大多数遗传疾病源于单一核碱基向不同核碱基的突变，但转化此类变化的当前方法非常低效，并且由于细胞对双链DNA断裂的应答，通常在靶基因座处诱导大量随

机插入和缺失(插入/缺失)[39,40]。本文报告了核碱基编辑的发展,一种新的基因组编辑策略,其实现一种靶核碱基以可编程方式直接转化为另一种碱基,而无需双链DNA主链断裂。工程化改造CRISPR/Cas9融合物,并保留用引导RNA编程的胞苷脱氨酶酶APOBEC1不诱导双链DNA断裂,并介导胞苷直接转化为尿嘧啶,从而实现了DNA复制,DNA修复或转录(若模板链是靶定的)后的C→T(或G→A)取代。所得的"核碱基编辑器"在约五个核苷酸的窗内转换胞苷,并可以在体外有效校正与人疾病相关的多种点突变。在四种转化的人和鼠细胞系中,分别融合尿嘧啶糖基化酶抑制剂(UGI)并使用靶向非编辑链的Cas9切口酶的第二代和第三代核碱基编辑器可以克服对核碱基编辑的细胞DNA修复应答,导致在最小(通常≤1%)插入/缺失形成的情况下人细胞中总细胞DNA的多达37%或(约15-75%)的永久校正。相反,在相同靶物上的规范Cas9介导的HDR以4%插入/缺失形成产生0.7%校正的平均值。核碱基编辑器用于将人乳腺癌和淋巴瘤细胞中的两个致癌p53突变恢复为野生型等位基因,并将ApoE4中与阿尔茨海默氏病相关的Arg密码子转化为小鼠星形胶质细胞中非疾病相关的Cys密码子。碱基编辑扩大了点突变的基因组编辑的范围和效率。

[0725] 聚簇规则间隔短回文重复序列(CRISPR)系统是原核适应性免疫系统,其已经适应于介导多种生物体和细胞系中的基因组工程[41]。CRISPR/Cas9蛋白-RNA复合物通过与引导RNA进行碱基配对定位于靶DNA序列,并在由引导RNA规定的基因座处天然创建DNA双链断裂(DSB)。响应DSB,内源性DNA修复过程通常通过非同源末端连接(NHEJ)导致DNA切割位点处的随机插入或缺失(缺失/缺失)。在存在同源DNA模板的情况下,切割位点周围的DNA可以通过同源性指导修复(HDR)替换。当疾病相关基因的简单破坏是足够的(例如,足以治疗某些功能获得疾病)时,靶向DNA切割和然后的插入/缺失形成可以是有效的。然而,对于大多数已知的遗传疾病,需要校正靶基因座中的点突变,而不是基因的随机破坏,以解决或研究疾病的根本原因[68]。

[0726] 受此需求的启示,研究人员投入了巨大的努力来提高HDR的效率并抑制NHEJ。例如,NHEJ途径中的必需酶连接酶IV的小分子抑制剂已显示提高HDR的效率[42,43]。然而,此种策略在有丝分裂后的细胞(其通常下调HDR)中有挑战性的,并且其治疗相关性受到抑制非靶细胞中连接酶IV的潜在风险的限制。由于HDR效率是高度细胞周期依赖性的,也可以通过将Cas9引导RNA复合物定时递送到化学同步细胞中来实现增强的HDR效率[44]。然而,由于使细胞同步化是高度破坏性的,此类方法仅限于细胞培养的研究应用。尽管取得了这些进展,但目前在大多数情况下使用HDR替换点突变的策略是非常低效的(通常约0.1%至5%)[42,43,45,46,75],尤其是在未修饰的非分裂细胞中。此外,HDR在解决双链断裂期间与NHEJ竞争,而插入/缺失通常是比基因替换更丰富的结果。这些观察结果突出显示了需要开发在基因组DNA中安装特定修饰的替代方法,其不依赖于创建双链DNA断裂。NHEJ途径中的必需酶连接酶IV的小分子抑制剂已显示出增加HDR的效率[42,43]。然而,此种策略在有丝分裂后的细胞(其通常下调HDR)中有挑战性的,并且其治疗相关性受到抑制非靶细胞中连接酶IV的潜在风险的限制。由于HDR效率是高度细胞周期依赖性的,也可以通过将Cas9引导RNA复合物定时递送到化学同步细胞中来实现增强的HDR效率[44]。然而,由于使细胞同步化是高度破坏性的,此类方法仅限于细胞培养的研究应用。在某些情况下,可以设计HDR模板,使得成功的HDR产物在PAM序列中含有突变,因此不再是后续Cas9修饰的底物,从而增加HDR产物的总产率[75],尽管此类方法强加对产物序列的约束。最近,此策略已与使用与非靶链互补的ssDNA供体和高效率核糖

核蛋白(RNP)递送相结合,以显著提高HDR的效率,但即使在这些情况下,HDR与NHEJ结果的比率是相对较低的(<2)[83]。

[0727] 想到在不需要DNA主链切割的情况下在可编程靶基因座处一个核碱基至另一个核碱基转化的直接催化相对于HDR可以提高基因校正的效率,而无需在感兴趣的基因组处引入不想要的随机插入/缺失。催化死亡Cas9(dCas9)(其含有使其核酸酶活性失活的Asp10Ala和His840Ala突变)保留了其以引导RNA编程方式结合DNA的能力,但不切割DNA主链[16,47]。原则上,dCas9与介导一个核碱基直接转化为另一种核碱基的酶促或化学催化剂的缀合可以实现RNA编程的核碱基编辑。胞嘧啶(C)的脱氨基由胞苷脱氨酶催化[29],并且产生尿嘧啶(U),它具有胸腺嘧啶(T)的碱基配对特性。将dCas9与胞苷脱氨酶融合,以测试其在引导RNA规定的DNA基因座处将C转化为U的能力。最已知的胞苷脱氨酶对RNA起作用,而少数已知接受DNA的实例需要单链DNA[48]。最近的关于Cas9-靶DNA复合物的研究表明,替换的DNA片段的至少9个核苷酸在Cas9:引导RNA:DNA“R-环”复合物形成后不配对[12]。实际上,在Cas9 R-环复合物的结构中,替换的DNA链上的原间隔物的前11个核苷酸是无序的,提示了其移动不是高度受约束的[76]。还已经推测非靶链中的胞嘧啶处的Cas9切口酶诱导的突变可以源自细胞胞苷脱氨酶对它们的可及性[77]。最近关于dCas9-靶DNA复合物的研究已经揭示了当Cas9结合其靶DNA序列时,非模板链上的至少26个碱基不配对[49]。推断R环中单链DNA的此区段的子集可以充当dCas9栓系的胞苷脱氨酶的底物,以实现DNA中C至U的直接可编程转化(图11A)。

[0728] 在哺乳动物细胞裂解物衍生的体外转录-翻译系统中表达四种不同的胞苷脱氨酶(hAID,hAPOBEC3G,rAPOBEC1和pmCDA1),并评估ssDNA的脱氨基。在这四种酶中,rAPOBEC1在测试条件下显示出最高的脱氨酶活性,并选择用于dCas9融合物实验(图36A)。虽然将rAPOBEC1附加到dCas9的C末端消除脱氨酶活性,但是与dCas9的N末端融合在与未融合酶的水平相当的水平保持ssDNA上的脱氨酶活性。表达和纯化四种rAPOBEC1-dCas9融合物,具有不同长度和组成的接头(图36B),并且在体外对每种融合物评估单链RNA(sgRNA)编程的dsDNA脱氨基(图11A至11C和图15A至15D)。

[0729] 在体外观察到有效的、序列特异性的、sgRNA依赖的C至U的转化(图11A至11C)。使用长度超过9个氨基酸的rAPOBEC1-dCas9接头,转化效率最大。易脱氨基的位置数目(脱氨基“活性窗”)随着接头长度从3个氨基酸扩展至21个氨基酸而增加(图36C至36F15A至15D)。发现16个残基的XTEN接头[50]在这两个特征之间提供有希望的平衡,有效的脱氨基窗为约5个核苷酸,从原间隔物内的位置4至8起,将原间隔物邻近基序(PAM)远端的末端计为位置1。rAPOBEC1-XTEN-dCas9蛋白充当第一代核碱基编辑器(NBE1)。

[0730] 选出了与人疾病相关的七个突变,它们在理论上可以通过C至T核碱基编辑来校正,合成相应序列的80聚体双链DNA并评估NBE1在体外校正这些突变的能力(图16A至16B)。NBE1以44%的平均表观编辑效率在体外在这7个靶物的6个中产生产物,该产物与靶C的有效编辑或者在存在多个C时活性窗内的至少一个C的有效编辑一致(图16A至16B)。在脱氨基窗中存在多个C的三种情况下,观察到这些胞嘧啶的部分或全部脱氨基的证据。在测试的七种情况的仅一种中观察到经编辑的产物的大量产率(图16A至16B)。尽管据报告APOBEC1底物的优选序列背景为CC或TC[51],但是预计由dCas9结合靶基因座介导的脱氨酶及其单链DNA底物的有效摩尔浓度增加可以放松此限制。为了阐明NBE1的序列背景一般性,测定了其编辑在位置7处含有单一固定C的60聚体双链DNA寡核苷酸,以及所有36个单突变变体的能力,

在所述单突变变体中针对其他三个碱基中的每个个别改变原间隔物碱基1-6和8-13。这37个序列中的每个序列用1.9μM NBE1、1.9μM相应的sgRNA和125nM DNA处理2h，类似于体外Cas9测定法的标准条件[52]。高通量DNA测序(HTS)揭示了靶定链的50％至80％C至U转化(源自两条DNA链的总序列读段的25％至40％，它们之一不是NBE1的底物)(图12A)。靶C周围的核苷酸对编辑效率几乎没有影响，不依赖于序列背景，除非靶C的直接5'的碱基是G，在此情况下，编辑效率实质性较低(图12A至12B)。在原间隔物中位置1至8处的所有四个NC基序上评估了NBE1的体外活性(图12A至12B)。通常，观察到底物上的NBE1活性遵循TC≥CC≥AC>GC的顺序，当靶C处于或接近位置7时实现最大的编辑效率。此外，观察到核碱基编辑器是高度持续合成能力的，并会在5个碱基的活性窗内有效将同一条DNA链上所有C的大多数转化为U(图17)。

[0731] 虽然BE1有效加工试管中的底物，但在细胞中，可能的DNA修复结果树决定了碱基编辑的初始U:G产物的命运(图29A)。为了测试在人细胞中碱基编辑的效率，针对哺乳动物的表达对NBE1密码子选择进行了优化，附加C末端核定位序列(NLS)[53]，并测定其在整个人基因组中的6孔研究的靶位点中在14C上在人细胞中将C转化为T的能力(图37A)[54]。通过将NBE1与对应于六个不同基因组位点的合成80-聚体温育，然后进行HTS在体外在每个原间隔物内确认可编辑的C(图13A至13C，图29B和图25)。接下来，用编码NBE1和六个靶sgRNA之一的质粒转染HEK293T细胞，允许核碱基编辑发生三天，从细胞中提取基因组DNA，并通过HTS分析基因座。尽管对所有六种情况观察到靶基因座处的细胞中的C至T的编辑，但核碱基编辑的效率为总DNA序列的1.1％至6.3％或0.8％-7.7％(对应于靶定链的2.2％至12.6％)，与体外核碱基编辑的效力相比，效率降低6.3倍至37倍或5倍至36倍(图13A至13C，图29B和图25)。观察到当底物C前面有T时位置4至8的典型窗之外的某种碱基编辑，我们将这归因于APOBEC1对TC底物的异常高活性[48]。

[0732] 询问对U:G异双链体DNA的存在的细胞DNA修复应答是否造成细胞中核碱基编辑效率的大幅降低(图29A)。尿嘧啶DNA糖基化酶(UDG)催化从细胞中的DNA除去U并启动碱基切除修复(BER)，最常见的结果是将U:G对恢复为C:G对(图29A)[55]。尿嘧啶DNA糖化酶抑制剂(UGI)(枯草芽孢杆菌噬菌体PBS1的83个残基的蛋白质)有力阻断人UDG活性(IC_{50}＝12pM)[56]。将UGI与NBE1的C末端融合，以创建第二代核碱基编辑器NBE2，并在所有六个基因组基因座上重复编辑测定法。人细胞中的编辑效率在使用NBE2的情况下比在使用NBE1的情况下平均高3倍，从而导致测序的总DNA的多达22.8％的基因转换效率(靶定链的多达45.6％)(图13A至13C和图29B)。为了测试人细胞中的碱基编辑，BE1密码子选择针对哺乳动物表达进行了优化，并附加C末端核定位序列(NLS)[53]。

[0733] 当将过表达UGI的分开质粒与NBE1共转染时，观察到相似的编辑效率(图18A至18H)。然而，虽然UGI与NBE1的直接融合导致在监测的非靶定基因组位置处C至T突变没有显著增加，但未融合的UGI的过表达可检测地增加基因组中其他位置处C至T突变的频率(图18A至18H)。通过评估U2OS细胞中相同的六个基因组靶物的编辑效率，证实了NBE2介导的核碱基编辑的一般性，并观察到与HEK293T细胞中的结果相似的结果(图19)。重要的是，NBE2通常不产生任何可检测的插入/缺失(图13C和图29C)，这与NHEJ对双链DNA断裂的已知机理依赖性一致[57,78]。这些结果共同指示了将UGI与NBE1缀合可以大大增加人细胞中核碱基编辑的效率。

[0734] 通过在测试的基因组基因座的两处监测HEK293T细胞中多个细胞分裂内的编辑效率证实人细胞中核碱基编辑的持久性。在两个时间点收获基因组DNA:用表达NBE2和合适的sgRNA的质粒转染后三天,以及传代细胞并使它们再生长四天后(约五个后续的细胞分裂)。在非传代细胞(对于三个不同的靶C,在靶定链的4.6%至6.6%中观察到编辑)和传代细胞(对于相同的三个靶链,在靶定链的4.6%至6.4%中观察到编辑)之间未观察到编辑效率的显著变化),确认核碱基编辑物在细胞分裂后变为永久性的(图20)。在极少数情况下,插入/缺失源自通过细胞修复过程对U:G损伤的加工,其涉及已知导致插入/缺失的单链断裂中间体[84]。鉴于每个人细胞每天从自发胞苷脱氨酶产生数百个内源性U:G损伤[85],预计在单个靶基因座处,来自U:G损伤修复的总插入/缺失频率不可能从BE1或BE2活性增加。

[0735] 为了进一步提高细胞中核碱基编辑效率,预期由于真核错配修复机制使用链不连续性来指导对错配双链体的任何断裂链的DNA修复,对非编辑的链产生切口可以导致较小分数的经编辑的U被细胞除去(图29A)[58,79,80]。催化性His残基在Cas9 HNH域中的位置840处得到恢复[47,59],导致第三代核碱基编辑器NBE3,其对含有与靶定C相对的G的非编辑链产生缺口,但不切割含有C的靶链。由于NBE3在Cas9中仍含有Asp10Ala突变,因此它不诱导双链DNA切割。相对于NBE2,此种对非编辑链产生切口的策略将人细胞中的核碱基编辑效率再提高1.4至4.8倍,从而导致HEK293T细胞中相同的6个人基因组靶物上含有靶定的C至T转化的总DNA序列的多达36.3%(图13A至13C和图29B)。重要的是,从NBE3处理中仅观察到小的插入/缺失频率,平均为0.8%(对于六个不同基因座,范围为0.2%至1.6%)(图13C,图29C和图34)。相反,当用野生型Cas9、sgRNA和单链DNA供体模板处理细胞以介导这些基因座中的三处的HDR时,观察到C至T转化效率平均仅为0.7%,相对插入/缺失形成高得多,平均为3.9%(图13A至13C和图29C)。等位基因转化与NHEJ结果的比率平均>1,000(对于BE2),23(对于BE3),和>0.17(对于野生型Cas9)(图3c)。我们通过监测HEK293T位点3和4基因组基因座处的HEK293T细胞中多个细胞分裂内的编辑效率确认人细胞中碱基编辑的永久性(图38)。这些结果共同建立了核碱基编辑可以在人细胞中比Cas9介导的HDR实现有效得多的靶向单碱基编辑,且以小得多的插入/缺失形成(NBE3)或无插入/缺失形成(NBE2)。

[0736] 接下来,评估了人细胞中NBE1,NBE2和NBE3的脱靶活性。Cas9,dCas9和Cas9切口酶的脱靶活性已得到广泛研究(图23至24和31至33)[54,60-62]。由于rAPOBEC1的序列偏爱性已显示出不依赖于DNA碱基,来自靶C的超过1个碱基[63],这与图12A至12B中观察到的序列背景独立性一致,假设核碱基编辑器的潜在脱靶活性源自脱靶Cas9结合。由于仅一定分数的Cas9脱靶位点在用于核碱基编辑的活性窗内具有C,因此脱靶核碱基编辑位点应是规范Cas9变体的脱靶位点的子集。对于所研究的六个位点中的每个,对先前使用GUIDE-seq方法确定的人细胞中前10种已知Cas9脱靶基因座进行了测序(图23至27和31至33)[54,61]。观察到仅在已知dCas9脱靶基因座的子集(对于NBE1和NBE2为16/34,47%,和对于NBE3为17/34,50%)处的可检测的脱靶核苷酸编辑。在所有情况下,脱靶碱基编辑底物在五碱基靶窗内包含C。通常,脱靶C至T的转化与脱靶Cas9核酸酶介导的基因组修饰频率平行(图23至27)。还监测了在测试的6个中靶和34个脱靶基因座周围的2,500个不同的胞嘧啶处C至T的转化,代表从约$1.8x10^6$个细胞中衍生的总共14,700,000个序列读段,并且未观察到与未处理的细胞的转化相比NBE1,NBE2或NBE3处理后这些其它位点的任一处的C至T的转化的可检测增加(图28)。这些发现共同指示了核碱基编辑器的脱靶底物包含Cas9脱靶底物的子集,并且人细胞

中的核碱基编辑器不在整个基因组中以可以通过本文使用的方法检测的水平诱导未靶定的C至T转化。在非传代的HEK293T细胞(对于三个靶C用BE2在测序链的1.8％至2.6％中观察到编辑,且用BE3为6.2％至14.3％)和在碱基编辑后已经经历约5个细胞分裂的细胞(对于相同靶C用BE2在测序链的1.9％至2.3％中观察到编辑以及用BE3为6.4％至14.5％)之间的编辑效率中未观察到实质性变化,确认这种细胞中的碱基编辑物是持久的(扩展数据图6)。

[0737] 最后,测试了核碱基编辑在哺乳动物细胞中校正三种与疾病相关的突变的潜力。载脂蛋白E基因变体APOE4编码氨基酸位置112和158处的两个Arg残基,并且是晚期阿尔茨海默氏病的最大且最常见的遗传危险因素[64]。在位置112或158中具有Cys残基的ApoE变体,包括APOE2(Cys112/Cys158),APOE3(Cys112/Arg158)和APOE3’(Arg112/Cys158)已显示[65]或被假设[81]比APOE4赋予实质性更低的阿尔茨海默氏病的风险。受NBE1在体外将APOE4转化为APOE3’的能力的鼓舞(图16A至16B),在永生化的小鼠星形胶质细胞中尝试此种转化,其中内源鼠APOE基因已被人APOE4(Taconic)替换。通过核转染(核转染效率为25％)将编码NBE3和适当sgRNA的DNA传递到这些星形胶质细胞中,两天后从所有处理过的细胞中提取基因组DNA,并通过HTS测量编辑效率。在总DNA测序读段的58-75％(44％的核转染的星形胶质细胞)中观察到了Arg158至Cys158的转化(图14A至14C和图30A)。如预期还观察到密码子158的第三位处的总DNA的36-50％编辑和Leu159的第一位置处总DNA的38-55％编辑,因为这些C的所有三个在活性核碱基编辑窗内。然而,由于TGC和TGT两者均编码Cys,且CTG和TTG两者均编码Leu,其他两个C→T转化均不导致ApoE3’蛋白的氨基酸序列的变化。从源自1x10^{6}个细胞的>1,500,000个测序读段中,观察到在NBE3处理后在靶定基因座处的1.7％插入/缺失的证据(图35)。相比之下,用wt Cas9和供体ssDNA对星形胶质细胞进行相同的处理在靶定基因座处产生0.1-0.3％的APOE4校正和26-40％的插入/缺失,效率与先前使用Cas9和HDR进行单碱基校正的报告一致[45,75](图30A和图40A)。除了用靶向VEGFA基因座的sgRNA以外相同处理的星形胶质细胞未显示出APOE4碱基编辑的证据(图34和图40A)。这些结果证明了核碱基编辑器如何在蛋白质的编码序列中实现精确的单氨基酸变化作为编辑的主要产物,即使当其持续合成能力导致基因组DNA中的超过一个核苷酸变化时。Cas9,dCas9和Cas9切口酶的脱靶活性已得到广泛研究[54,60-62]。通常,BE1,BE2和BE3的脱靶C至T转化与脱靶Cas9核酸酶介导的基因组修饰频率平行。

[0738] Tyr163Cys和Asn239Asp的显性阴性p53突变与几种类型的癌症密切相关[66-67]。这两种突变都可以通过模板链上的C至T转化来校正(图16A至16B)。将p53 Tyr163Cys突变纯合的人乳腺癌细胞系(HCC1954细胞)用编码NBE3和经编程以校正Tyr163Cys的sgRNA的DNA进行核转染。因为HCC1954细胞的核转染效率<10％,所以将表达IRFP的质粒共核转染到这些细胞中,以便在处理后两天通过荧光激活的细胞分选分离出核转染的细胞。基因组DNA的HTS揭示了7.6％的核转染HCC1954细胞中Tyr163Cys突变的校正(图30B和图40A至40B)。也以92％的核转染效率用编码NBE2和经编程为校正Asn239Asp的sgRNA的DNA对在p53 Asn239Asp上杂合的人淋巴瘤细胞系(ST486细胞)进行核转染。在11％的处理过的ST486细胞(12％的核转染ST486细胞)中观察到Asn239Asp突变的校正。与HEK细胞中的发现一致,用NBE2处理ST486细胞未观察到插入/缺失,而用NBE3处理HCC1954细胞则观察到0.6％插入/缺失形成。在源自2x10^{5}个细胞的>2,000,000个测序读段中未以高于未处理对照的频率的频率检测到原间隔物两侧至少50个碱基对内的其他DNA变化(图14A至14C,图30B)。这些结

果共同代表了基因组DNA中三个与疾病相关的等位基因向其野生型形式以某种效率且在缺乏其他基因组修饰事件的情况下转化，据我们所知，这是目前使用其他方法无法实现。

[0739] 为了阐明核碱基编辑器解决人遗传疾病的潜在相关性，在NCBI ClinVar数据库[68]中搜索了已知的遗传病，所述遗传病原则上可以通过此种方法进行校正。通过首先仅检查单核苷酸多态性(SNP)，然后除去任何非致病性变体来过滤ClinVar。在24,670个致病性SNP中，3,956个是由T至C或A至G取代引起的。对该列表进行了进一步过滤，以仅包括具有会将SNP定位在脱氨基活性窗内的附近NGG PAM的变体，从而产生1,089个临床相关的致病基因变体，其原则上可以由本文描述的核碱基编辑器进行校正(图21)。

[0740] 在一些实施方案中，本文中提供的任何碱基编辑器可以用于治疗疾病或病症。例如，本文中提供的任何碱基编辑器可以用于校正与本文中提供的任何疾病或病症相关的一种或多种突变。可以治疗的示例性疾病或病症包括但不限于3-甲基戊烯二酸尿2型、46,XY性腺发育不全、4-Alpha-羟苯丙酮酸羟化酶缺陷、6-丙酮酰-四氢蝶呤合酶缺陷、全色盲、酸不稳定亚基缺陷、肢端骨发育不全、肢端红色角皮症(acroerythrokeratoderma)、ACTH抗性、不依赖于ACTH的大结节肾上腺增生、活化PI3K-delta综合征、急性间歇性卟啉病、急性髓样白血病、Adams-Oliver综合征1/5/6、腺苷酸基琥珀酸裂解酶缺陷、肾上腺脑白质营养不良、成人神经元蜡样质脂褐质沉积症、成人发作性共济失调伴有动眼运用不能、睡眠期提前综合征、年龄相关性黄斑变性、Alagille综合征、亚历山大病、Allan-Herndon-Dudley综合征、奥尔波特综合征、X连锁隐性、儿童交替性偏瘫、肺泡毛细血管发育不良伴有肺静脉未对准、釉质形成不全、淀粉样蛋白形成转甲状腺素蛋白淀粉样变性(Amyloidogenic transthyretin amyloidosis)、肌萎缩侧索硬化、贫血(非球形溶血性、由于G6PD缺陷)、贫血(铁粒幼细胞性、吡多辛难治性、常染色体隐性)、甲缺如、抗凝血酶III缺陷、主动脉瘤、再生障碍性贫血、载脂蛋白C2缺陷、表观盐皮质激素过多、芳香酶缺陷、心律失常性右心室心肌病、家族性肥大性心肌病、肥厚性心肌病、先天性多关节弯曲、天冬氨酰基葡萄糖胺尿、窒息性腔部营养不良、共济失调伴有维生素E缺陷、共济失调(痉挛性)、心房颤动、房间隔缺损、非典型溶血性尿毒综合征、常染色体显性CD11C+/CD1C+树突细胞缺陷、常染色体显性进行性外眼肌麻痹伴有线粒体DNA删除、Baraitser-Winter综合征、巴特综合征、基底神经节钙化、贝-威二氏综合征、良性家族性新生儿癫痫、良性肩关节肌营养不良、Bernard Soulier综合征、Beta中间型地中海贫血、Beta-D-甘露糖苷病、比蒂结晶性角膜视网膜营养不良(Bietti crystalline corneoretinal dystrophy)、胆汁酸吸收不良、生物胞素酶缺陷、Borjeson-Forssman-Lehmann综合征、Boucher Neuhauser综合征、Bowen-Conradi综合征、短指、Brown-Vialetto-Van laere综合征、Brugada综合征、心律不整、心脏脸皮肤(Cardiofaciocutaneous)综合征、心肌病、Carnevale综合征、肉碱软脂酰转移酶II缺陷、Carpenter综合征、白内障、儿茶酚胺依赖性室性心动过速、中央轴空病，染色体1,9和16的着丝粒不稳定性和免疫缺陷、脑常染色体显性动脉病、脑-眼-脸-骨骼综合征、蜡样脂褐质沉积症、夏科-马里-图斯病、胆固烷醇贮积病、软骨钙质沉着症、软骨发育不全、慢性进行性多发性硬化、辅酶Q10缺陷、Cohen综合征、因子V和因子VIII组合缺陷、组合免疫缺陷、组合氧化磷酸化缺陷、组合部分17-alpha-羟化酶/17,20-裂合酶缺陷、补体因子d缺陷，完全组合17-alpha-羟化酶/17,20-裂合酶缺陷、视锥-视杆营养不良、先天性挛缩型蜘蛛状指、先天性糖基化病症、先天性脂肪瘤过度生长、卵巢新生物、PIK3CA相关过度生长谱、先天性长

QT综合征、先天性肌营养不良、先天性肌营养不良-脑综合征、先天性肌无力综合征、先天性肌病伴有纤维型不相称、Eichsfeld型先天性肌营养不良,先天性静止性夜盲、角膜营养不良、Cornelia de Lange综合征、颅骨干骺端发育不全症、Crigler Najjar综合征、Crouzon综合征、皮肤松弛症伴有骨发育不全、发绀、囊性纤维化病、胱氨酸贮积症、细胞色素-c氧化酶缺陷、线粒体复合物I缺陷、D-2-羟基戊二酸尿、Danon病、耳聋伴有迷路不发育小耳畸形和小牙(LAMM)、耳聋、乙酰基-CoA乙酰基转移酶缺陷、亚铁氧化酶缺陷、UDP葡萄糖-己糖-1-磷酸尿苷基转化酶缺陷、Dejerine-Sottas病、Desbuquois综合征、DFNA、2型糖尿病、糖尿病-耳聋综合征、戴-布二氏贫血、畸型发育不全、二氢蝶啶还原酶缺陷、二氢嘧啶酶缺陷、扩张型心肌病、播散性非典型分支杆菌感染、远端关节挛缩、远端遗传性运动神经元病(neuronopathy)、Donnai Barrow综合征、迪谢内肌营养不良、贝克肌营养不良、遗传性泛发性色素异常病、先天性角化不良症、张力失常、早期幼儿癫痫性脑病、Ehlers-Danlos综合征、Eichsfeld型先天性肌营养不良症、Emery-Dreifuss肌营养不良症、Enamel-renal综合征、反向营养不良性大疱性表皮松解(Epidermolysis bullosa dystrophica inversa)、疱疹性大疱性表皮松解(Epidermolysis bullosa herpetiformis)、癫痫、发作性共济失调、变异性红角皮病、红细胞生成性原卟啉病、锻炼不耐性、渗出性玻璃体视网膜病变、法布里病、因子V缺陷、因子VII缺陷、因子xiii缺陷、家族性多发性腺癌、乳腺癌、卵巢癌、冷蛰麻疹、慢性婴儿神经学、皮肤和关节综合征、偏瘫型偏头痛、高胆固醇血症、肥厚性心肌病、低α-脂蛋白血症、低钾血症-低镁血症、幼年痛风、高脂蛋白血症、内脏淀粉样变性、低血磷性维生素D难治性佝偻病、FG综合征、眼外肌纤维化、费氏先天性肾病综合征、局限型癫痫、局灶性节段性肾小球硬化症、额鼻发育异常、额颞痴呆、果糖-二磷酸酶缺陷、Gamstorp-Wohlfart综合征、神经节苷脂唾液酸酶缺陷、GATA-1相关的血小板减少、戈谢病、巨轴索神经病、Glanzmann血小板机能不全、肾小球肾病、肾小球病、糖皮质激素抗性、葡萄糖-6-磷酸转运缺陷、戊二酸尿症、糖原贮积病、Gorlin综合征、前脑无裂畸形、GRACILE综合征、出血性毛细血管扩张、血色素沉着症、血红蛋白H病、溶血性贫血、嗜血细胞性淋巴组织细胞增多症、结肠癌、Myhre综合征、脑白质病、遗传性因子IX缺陷病、遗传性因子VIII缺陷病、遗传性因子XI缺陷病、遗传性果糖尿症、遗传性非息肉病性结肠直肠新生物、遗传性胰腺炎、遗传性焦性异形红细胞病(pyropoikilocytosis)、椭圆形红细胞性贫血、内脏异位、异位、组织细胞性髓性网状细胞增多症、组织细胞增多症-淋巴结病+综合征、由于醛缩酶A缺陷所致的HNSHA、羧化全酶合成酶缺陷、同型半胱氨酸血症、Howel-Evans综合征、葡萄胎、高钙尿高钙血症、高免疫球蛋白D、甲羟戊酸尿、高胰岛素血低血糖症、高钾血性周期性麻痹、vonEulenburg先天性副肌强直、高脂蛋白血症、高锰血症(Hypermanganesemia)、高蛋氨酸血症、高磷酸酯酶血、高血压、低镁血症、低β脂蛋白血症、低钙血症、低促性腺素性功能减退症、低促性腺素性功能减退症、少汗性外胚层发育不良、超-IgM免疫缺陷、少汗性X连锁外胚层发育不良症、低镁血症、甲状旁腺功能减退、特发性纤维化肺泡炎、免疫缺陷、免疫球蛋白A缺陷、婴儿型低磷酸酶症、婴儿型帕金森症-张力失常、胰岛素依赖性糖尿病、中间型枫糖尿病、坐骨髌骨发育异常、胰岛细胞增生、孤立性生长激素缺陷、孤立性促黄体素缺陷、异戊酸血症、Joubert综合征、幼年性息肉病综合征、青年性视网膜劈裂症、Kallmann综合征、Kartagener综合征、库格尔贝格-韦兰德病、格子状角膜变性、利伯先天性黑矇、利伯视神经萎缩、左心室致密化不全(Left ventricular noncompaction)、利氏病、线粒体复合物I缺

陷、矮妖精貌综合征、关节挛缩、前角细胞疾病、白细胞粘附缺陷、脑白质营养不良症、脑白质病、卵巢脑白质营养不良症、L-铁蛋白缺陷、Li-Fraumeni综合征、肢肌营养不良-肌营养不良蛋白聚糖病(dystroglycanopathy)、Loeys-Dietz综合征、长QT综合征、大头/自闭症综合征、斑点状角膜营养不良、黄斑营养不良、恶性体温过高易感性、前列腺恶性肿瘤、枫糖尿病、Marden Walker样综合征、Marfan综合征、Marie Unna遗传性稀毛症、肥大细胞病、胎粪性肠梗阻、中链乙酰辅酶A脱氢酶缺陷、Melnick-Fraser综合征、智力低下、分层蛋白缺陷先天性肌营养不良症、间皮瘤、异染性脑白质营养不良、干骺端软骨发育不全、高铁血红蛋白血症、甲基丙二酸尿症、高胱氨酸尿、小头、脉络膜视网膜病变、淋巴水肿、小眼、轻度非-PKU高苯丙氨酸血症、Mitchell-Riley综合征、线粒体3-羟基-3-甲基戊二酰-CoA合酶缺陷、线粒体复合物I缺陷、线粒体复合物III缺陷、线粒体肌病、粘脂质累积III、粘多糖贮积症、多硫酸酯酶缺陷、肌无力综合征、结核分枝杆菌、髓过氧化物酶缺陷、Myhre综合征、肌阵挛型癫痫、肌纤维肌病、肌红蛋白尿、肌病、近视、先天性肌强直、Navajo神经肝病、线形体肌病、胃新生物、肾性尿崩症、肾痨、肾病综合征、神经纤维瘤病、中性脂质贮积病、尼曼-皮克病、非酮性高甘氨酸血症、Noonan综合征、Noonan综合征样病症、诺伦病、黄斑变性、N-端乙酰转移酶缺陷、眼皮肤白化病、眼齿指发育不良、Ohdo综合征、视神经不发育、鸟氨酸氨甲酰基转移酶缺陷、口面指综合征、成骨不全、骨硬化症、卵巢发育不全、甲肥厚、掌跖角化病、非表皮松解性、Papillon-Lef\xc3\xa8vre综合征、Haim-Munk综合征、牙周炎、Peeling皮肤综合征、Pendred综合征、过氧化氢酶体脂肪酰基-coa还原酶1病症、过氧化物酶体生源(biogenesis)病症、Pfeiffer综合征、苯丙酮尿、苯丙酮尿、高苯丙氨酸血症、非PKU、垂体激素缺陷、毛发红糠疹、结节性多动脉炎、多囊肾病、多囊性脂膜骨发育异常、多小脑回、脑桥小脑低常增生、汗孔角化病、后柱共济失调、原发性红斑性肢痛症、高草酸尿症、进行性家族性肝内胆汁淤积、进行性假类风湿发育异常、丙酸血症、假两性畸形、假性醛固酮减少症、弹性假黄色瘤样病症、嘌呤核苷酸磷酸化酶缺陷、吡哆醛5-磷酸依赖性癫痫、肾发育不良、视网膜色素营养不良、小脑性共济失调、骨骼发育不良、网状细胞发育不全、色素性视网膜炎、Usher综合征、视网膜母细胞瘤、视网膜病变、RRM2B相关线粒体病、Rubinstein-Taybi综合征、施奈德结晶状角膜营养不良、皮脂肿瘤、严重先天性嗜中性白血球减少症、婴儿严重肌阵挛型癫痫、严重X连锁肌管性肌病、甲发育不良、面部畸形、稀毛症、短肋骨胸廓发育不良、唾液酸贮积病、唾液酸沉积症、铁粒幼细胞性贫血、小纤维神经病、Smith-Magenis综合征、Sorsby眼底营养不良、痉挛性共济失调、痉挛性截瘫、精子发生失败、球形细胞增多症、鞘磷脂/胆固醇脂沉积、脊髓小脑性共济失调、手裂/足畸形、脊椎干骺端发育不全、扁椎骨致死性骨骼发育不良(Platyspondylic lethal skeletal dysplasia)、头和颈的鳞状细胞癌、斯塔加特病、蔗糖酶-异麦芽糖酶缺陷、婴儿猝死综合征、主动脉瓣上狭窄(Supravalvar aortic stenosis)、表面活化剂代谢功能障碍、丹吉尔病、Tatton-Brown-rahman综合征、胸主动脉瘤和主动脉夹层、血栓形成倾向、甲状腺激素抗性、TNF受体相关周期热综合征(TRAPS)、牙齿发育不全、扭转性室速、大动脉的移位、Treacher Collins综合征、结节性硬化症综合征、酪氨酸酶阴性眼皮肤白化病、酪氨酸酶阳性眼皮肤白化病、酪氨酸血症、UDP葡萄糖-4-差向异构酶缺陷、Ullrich先天性肌营养不良症、Bethlem肌病Usher综合征、UV-敏感性综合征、Van der Woude综合征、胭窝翼状赘蹼综合征、极长链酰基-CoA脱氢酶缺陷、膀胱输尿管反流、玻璃体视网膜脉络膜病(Vitreoretinochoroidopathy)、Von Hippel-

Lindau综合征、血管性血友病、Waardenburg综合征、Warsaw断裂综合征、WFS1相关病症、Wilson病、着色性干皮病、X连锁丙种球蛋白缺乏血症、X连锁遗传性运动和感觉神经病、X连锁严重组合免疫缺陷和Zellweger综合征。

[0741] 核碱基编辑的发展推进了基因组编辑的范围和效率两者。本文描述的核碱基编辑器为研究人员提供了在几乎没有插入/缺失形成(NBE2)的情况下编辑,或者以较低插入/缺失形成频率(本文通常≤1%)(NBE3)的更有效编辑的选择。根据定义,碱基编辑的产物不再是底物,这可能通过防止后续产物转化而促成编辑效率,所述后续产物转化可以阻碍传统的Cas9应用。通过消除对双链DNA切割和随机DNA修复过程(该过程随细胞状态和细胞类型而极大差异)的依赖,核碱基编辑具有扩大可以干净安装的基因组修饰类型,这些修饰的效率以及适合编辑的细胞类型的潜力。可能的是将最近工程化改造的具有改善的DNA特异性的Cas9变体[69,70,82]或递送方法[71]以及具有改变的PAM特异性的Cas9变体[72]整合到该策略中,以提供具有改进的DNA特异性或可以靶向甚至更宽的疾病相关突变范围的其他核碱基编辑器。这些发现还提示,将dCas9与催化额外核碱基转化的酶的额外融合将增加可通过核碱基编辑进行的可能的DNA碱基变化的分数。这些结果还提示用于融合其他DNA修饰酶(包括甲基化酶和脱甲基纤维素)的结构,其可以使其他类型的可编程基因组和表观基因组碱基编辑成为可能。

[0742] 材料和方法

[0743] 克隆。本文使用的所有构建体和引物的DNA序列均列在补充序列中。包含编码NBE1,NBE2和NBE3的基因的质粒可从Addgene获得。使用VeraSeqULtra DNA聚合酶(Enzymatics)或Q5 Hot Start高保真DNA聚合酶(New England Biolabs)进行PCR。使用USER克隆(New England Biolabs)构建NBE质粒。将脱氨酶基因合成为gBlocks基因片段(Integrated DNA Technologies),并从先前报告的质粒[18]中获得Cas9基因。将脱氨酶和融合基因克隆到pCMV(哺乳动物密码子优化)或pET28b(大肠杆菌密码子优化)主链中。使用定点诱变构建sgRNA表达质粒。简而言之,根据制造商的说明,使用T4多核苷酸激酶(New England Biolabs)将补充序列中列出的引物进行5'磷酸化。接下来,根据制造商的说明,使用Q5 Hot Start高保真聚合酶(New England Biolabs),以磷酸化引物和质粒pFYF1320(EGFP sgRNA表达质粒)为模板,进行PCR。将PCR产物与DpnI(20U,New England Biolabs)在37℃下温育1小时,在QIAprep旋转柱(Qiagen)上纯化,并根据制造商的说明使用QuickLigase(New England Biolabs)连接。使用Mach1感受态细胞(ThermoFisher Scientific)进行DNA载体扩增。

[0744] 在ssDNA上的体外脱氨酶测定法。所有ssDNA底物的序列在补充序列中列出。所有Cy3标记的底物均获自Integrated DNA Technologies(IDT)。使用TNT T7快速偶联转录/翻译试剂盒(Promega),根据制造商的说明,使用1μg质粒在体外表达脱氨酶。蛋白质表达后,将5μL裂解物与CutSmart缓冲液(New England Biolabs)(50mM乙酸钾,29mM Tris乙酸盐,10mM乙酸镁,100ug/mL BSA,pH 7.9)中的35μL ssDNA(1.8μM)和USER酶(1个单位)组合并在37℃下温育2小时。在10%TBE-尿素凝胶(Bio-Rad)上将经切割的含U的底物与全长未修饰的底物解析。

[0745] His_6-rAPOBEC1-接头-dCas9融合物的表达和纯化。用编码带有GGS,$(GGS)_3$,(SEQ ID NO:610)XTEN,或$(GGS)_7$(SEQ ID NO:610)接头的pET28b-His_6-rAPOBEC-接头-dCas9的质

粒转化大肠杆菌BL21 STAR(DE3)感受态细胞(ThermoFisher Scientific)。将所得的表达菌株在含有100μg/mL卡那霉素的Luria-Bertani(LB)肉汤中于37℃培养过夜。将细胞以1:100稀释到相同的生长培养基中,并在37℃下生长至OD_{600}=约0.6。在2h时段内将培养物冷却至4℃,并以0.5mM添加异丙基-β-D-1-硫代半乳糖吡喃糖苷(IPTG)诱导蛋白质表达。约16h后,以4,000g离心收集细胞,并重悬于裂解缓冲液(50mM三(羟甲基)-氨基甲烷(Tris)-HCl,pH 7.0,1M NaCl,20%甘油,10mM三(2-羧乙基)膦(TCEP,Soltec Ventures))中。通过超声处理(20s脉冲打开,20s脉冲关闭,在6W输出下总共8分钟)进行超声处理裂解细胞,并在25,000g离心15分钟后分离出裂解物上清液。将裂解物与His-Pur镍-氮川乙酸(镍-NTA)树脂(ThermoFisher Scientific)在4℃温育1小时,以捕获带His标签的融合蛋白。将树脂转移到柱上,并用40mL的裂解缓冲液洗涤。将带有His标签的融合蛋白在补充有285mM咪唑的裂解缓冲液中洗脱,并通过超滤(Amicon-Millipore,100-kDa分子量截留)浓缩至1mL总体积。在含有50mM三(羟甲基)-氨基甲烷(Tris)-HCl,pH 7.0、0.1M NaCl,20%甘油,10mM TCEP的低盐纯化缓冲液中将蛋白质稀释至20mL,并加载到SP Sepharose Fast Flow树脂(GE Life Sciences)。用40mL此种低盐缓冲液洗涤树脂,并用5mL活性缓冲液洗脱蛋白质,该活性缓冲液包含50mM三(羟甲基)-氨基甲烷(Tris)-HCl,pH 7.0,0.5M NaCl,20%甘油,10mM TCEP。洗脱的蛋白质在SDSPAGE凝胶上定量。

[0746] sgRNA的体外转录。根据制造商的说明,使用TranscriptAid T7高产率转录试剂盒(ThermoFisher Scientific)使用补充序列中列出的引物,体外转录包含T7启动子和后面的20bp sgRNA靶序列的线性DNA片段。根据制造商的说明,使用MEGAclear试剂盒(ThermoFisher Scientific)纯化sgRNA产物,并通过UV吸光度进行定量。

[0747] 经Cy3缀合的dsDNA底物的制备。80个核苷酸的未标记链的序列列于补充序列中,并作为来自IDT的PAGE纯化寡核苷酸订购。补充序列中列出的25nt Cy3标记的引物与每个80-nt底物的3'末端互补。该引物从IDT以HPLC纯化寡核苷酸订购。为了生成Cy3标记的dsDNA底物,将80-nt的链(5μL的100μM溶液)与Cy3标记的引物(5μL的100μM溶液)在具有dNTPs(0.75μL的100mM溶液)的NEBuffer 2(38.25μL 50mM NaCl,10mMTris-HCl,10mM $MgCl_2$,1mM DTT,pH 7.9溶液,New England Biolabs)中组合,并且加热至95℃5分钟,然后以0.1℃/s速率逐渐冷却至45℃。在此退火期后,加入Klenow exo^-(5U,New England Biolabs),并将反应在37℃下温育1h。溶液用PB缓冲液(250μL,Qiagen)和异丙醇(50μL)稀释,并在QIAprep旋转柱(Qiagen)上纯化,用50μLTris缓冲液洗脱。

[0748] dsDNA上的脱氨酶测定法。将纯化的融合蛋白(活性缓冲液中20μL1.9μM)与1当量的适当sgRNA组合,并在环境温度下温育5分钟。加入Cy3标记的dsDNA底物至终浓度125nM,并将所得溶液在37℃下温育2小时。通过添加缓冲液PB(100μL,Qiagen)和异丙醇(25μL)从融合物中分离dsDNA,并在EconoSpin微旋转柱(Epoch Life Science)上纯化,用20μL CutSmart缓冲液(New England Biolabs)洗脱。将USER酶(1U,New England Biolabs)添加到纯化的经编辑的dsDNA,并在37℃下温育1h。通过将5μL反应溶液与15μL基于DMSO的加载缓冲液(5mM Tris,0.5mM EDTA,12.5%甘油,0.02%溴酚蓝,0.02%二甲苯青(xylene cyan),80%DMSO)组合,将Cy3标记的链与其互补物完全变性。在10%TBE-尿素凝胶(Bio-Rad)上,将全长的含C的底物与任何切割的含U的编辑底物分离,并在GEAmersham Typhoon成像仪上成像。

[0749] 用于高通量测序(HTS)的体外编辑的dsDNA的制备。补充序列中列出的寡核苷酸获自IDT。在Tris缓冲液中组合互补序列(5μL的100μM溶液)，并通过加热至95℃5分钟进行退火，然后以0.1℃/s的速率逐渐冷却至45℃以生成60-bp dsDNA底物。将纯化的融合蛋白(活性缓冲液中的20μL 1.9μM)与1当量的适当sgRNA组合，并在环境温度下温育5分钟。加入60聚体dsDNA底物至终浓度125nM，并且将所得溶液在37℃下温育2小时。通过添加缓冲液PB(100μL，Qiagen)和异丙醇(25μL)从融合物中分离dsDNA，并在EconoSpin微旋转柱(Epoch Life Science)上纯化，用20μL Tris缓冲液洗脱。根据制造商的说明，使用补充序列和VeraSeq Ultra(Enzymatics)中规定的HTS引物对，通过PCR以13个扩增循环扩增得到的经编辑的DNA(1μL用作模板)。使用RapidTips(Diffinity Genomics)纯化PCR反应产物，并使用含有测序衔接头的引物通过PCR扩增纯化的DNA，纯化并在MiSeq高通量DNA测序仪(Illumina)上测序，如先前描述[73]。

[0750] 细胞培养。在补充有10％(v/v)胎牛血清(FBS)和青霉素/链霉素(1x，Amresco)的Dulbecco改良Eagle培养基加GlutaMax(ThermoFisher)中，在37℃和5％CO_2维持HEK293T(ATCC CRL-3216)，U2OS(ATCC-HTB-96)和ST486细胞(ATCC)。在如上所述补充的RPMI-1640培养基(ThermoFisher Scientific)中维持HCC1954细胞(ATCC CRL-2338)。将含有APOE基因ApoE4同种型的永生化大鼠星形胶质细胞(Taconic Biosciences)在补充有10％(v/v)胎牛血清(FBS)和200μg/mL遗传霉素(ThermoFisher Scientific)的Dulbecco改良Eagle培养基加GlutaMax(ThermoFisher Scientific)中培养。

[0751] 转染。将HEK293T细胞接种在48孔胶原蛋白包被的BioCoat平板(Corning)上，并以约85％汇合转染。简而言之，根据制造商的方案，每孔使用1.5μl Lipofectamine 2000(ThermoFisher Scientific)转染750ng NBE和250ng sgRNA 表达质粒。根据制造商的说明，使用适当的AMAXA NUCLEOFECTOR™ II程序转染星形胶质细胞，U2OS，HCC1954，HEK293T和ST486细胞。将40ng infrared RFP(Addgene质粒45457)[74]加入到核转染溶液中以评估这些细胞系中的核转染效率。对于星形胶质细胞，U2OS和ST486细胞，核转染效率分别为25％，74％和92％。对于HCC1954细胞，核转染效率<10％。因此，在进行胰蛋白酶消化后，将HCC1954细胞通过40微米滤网(Fisher Scientific)过滤，并使用iRFP信号(abs 643nm，em 670nm)在Beckman Coulter MoFlo XDP细胞分选仪上收集经核转染的HCC1954细胞。在不富集核转染细胞的情况下使用其他细胞。

[0752] 基因组DNA样品的高通量DNA测序。3d后收获转染的细胞，并根据制造商的说明使用Agencourt DNAdvance基因组DNA分离试剂盒(Beckman Coulter)分离基因组DNA。通过PCR，使用补充序列中列出的侧翼HTS引物对，扩增感兴趣的中靶和脱靶基因组区域。按照制造商的说明，使用5ng基因组DNA作为模板使用Phusion高保真DNA聚合酶(ThermoFisher)进行PCR扩增。分别确定每个引物对的循环数，以确保反应在扩增的线性范围内停止(EMX1，FANCF，HEK293位点2，HEK293位点3，HEK293位点4和RNF2引物分别为30、28、28、28、28、32和32个循环)。PCR产物用RapidTips(Diffinity Genomics)纯化。用含有测序衔接头的引物通过PCR扩增纯化的DNA。对产物进行凝胶纯化，并且使用QUANT-IT™ PicoGreen dsDNA测定试剂盒(ThermoFisher)和KAPA Library Quantitation Kit-Illumina(KAPA Biosystems)定量。如前所述[73]，在Illumina MiSeq上对样品进行测序

[0753] 数据分析。测序读段使用MiSeq Reporter(Illumina)自动解复用，并使用补充说

明中提供的自定义Matlab脚本分析了个别FASTQ文件。使用Smith-Waterman算法，将每个读段与适当的参考序列成对比对。Q得分低于31的碱基用N替换，因此在计算核苷酸频率时排除在外。此处理产生约1,000之1的预期MiSeq碱基调用错误率。将读段和参考序列不含缺口的比对序列存储在比对表中，对于每个基因座，可以从所述比对表将碱基频率制成表格。

[0754] 使用先前描述的标准[71]，用补充说明中所示的自定义Matlab脚本对插入/缺失频率进行定量。扫描测序读段以查找与两个10-bp序列的精确匹配，该序列位于可能发生插入/缺失的窗两侧侧翼。若未定位精确匹配，则从分析中排除读段。若此插入/缺失窗的长度与参考序列完全匹配，则将读段分类为不含插入/缺失。若插入/缺失窗比参考序列长或短两个或多个碱基，则测序读段分别分类为插入或缺失。

[0755] 本文(例如在背景技术，发明概述，发明详述，实施例和/或参考文献部分中)提及的所有出版物，专利，专利申请，出版物和数据库条目(例如序列数据库条目)在此通过引用完整并入本文，就像每个单独的出版物，专利，专利申请，出版物和数据库条目均通过引用被具体地和单独地并入本文一样。在冲突的情况下，以本申请(包括本文的任何定义)为准。

[0756] 补充序列

[0757] 用于产生sgRNA转染质粒的引物。在所有情况下都使用rev_sgRNA_plasmid。如材料和方法部分所述，将pFYF1320质粒用作模板。SEQ ID NO:187-196分别从上至下显示。

[0758]

rev_sgRNA_plasmid	GGTGTTTCGTCCTTTCCACAAG
fwd _p53 _Y163C	GCTTGCAGATGGCCATGGCGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC
fwd_p53_N239D	TGTCACACATGTAGTTGTAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC
fwd_APOE4 _C158R	GAAGCGCCTGGCAGTGTACCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC
fwd_EMX1	GAGTCCGAGCAGAAGAAGAAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC
fwd_FANCF	GGAATCCCTTCTGCAGCACCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC
fwd_HEK293_2	GAACACAAAGCATAGACTGCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC
fwd_HEK293_3	GGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC
fwd_HEK293_4	GGCACTGCGGCTGGAGGTGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC
fwd_RNF2	GTCATCTTAGTCATTACCTGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC

[0759] 用于体外脱氨酶测定法的所有ssDNA底物的序列。SEQ ID NO:197-199分别从上至下显示。

[0760] rAPOBEC1底物 Cy3-ATTATTATTATTCCGCGGATTTATTTATTTATTTATTTATTT

[0761] hAID/pmCDA1底物 Cy3-ATTATTATTATTAGCTATTTATTTATTTATTTATTTATTT

[0762] rAPOBEC3G底物 Cy3-ATTATTATTATTCCCGGATTTATTTATTTATTTATTTATTT

[0763] 引物，用于产生PCR产物以充当基于凝胶的脱氨酶测定法的sgRNA的T7转录的底物。在所有情况下使用rev_gRNA_T7。如材料和方法部分所述，将pFYF1320质粒用作模板。SEQ ID NO:200-223分别从上至下显示。

[0764]

rev_sgRNA_T7	AAAAAAAGCACCGACTCGGTG
fwd_sgRNA_T7_dsDNA_2	TAATACGACTCACTATAGGCCGCGGATTTATTTATTTAAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_3	TAATACGACTCACTATAGGTCCGCGGATTTATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_4	TAATACGACTCACTATAGGTTCCGCGGATTTATTTATTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_5	TAATACGACTCACTATAGGATTCCGCGGATTTATTTATTGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_6	TAATACGACTCACTATAGGTATTCCGCGGATTTATTTATGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_7	TAATACGACTCACTATAGGTTATTCCGCGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_8	TAATACGACTCACTATAGGATTATTCCGCGGATTTATTTGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_9	TAATACGACTCACTATAGGTATTATTCCGCGGATTTATTGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_10	TAATACGACTCACTATAGGATTATTATCCGCGGATTTATGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_11	TAATACGACTCACTATAGGTATTATATTCCGCGGATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_12	TAATACGACTCACTATAGGTTATTATATTCCGCGGATTTGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_13	TAATACGACTCACTATAGGATTATTATATTCCGCGGATTGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_14	TAATACGACTCACTATAGGTATTATTATATTCCGCGGATGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_15	TAATACGACTCACTATAGGATTATTATTATTACCGCGGAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_18	TAATACGACTCACTATAGGATTATTATTATTATTACCGCGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_noC	TAATACGACTCACTATAGGATATTAATTTATTTATTTAAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_APOE4_C112R	TAATACGACTCACTATAGGGGAGGACGTGCGCGGCCGCCGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_APOE4_C158R	TAATACGACTCACTATAGGGAAGCGCCTGGCAGTGTACCGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_CTNNB1_T41A	TAATACGACTCACTATAGGCTGTGGCAGTGGCACCAGAAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_HRAS_Q61R	TAATACGACTCACTATAGGCCTCCCGGCCGGCGGTATCCGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_53_Y163C	TAATACGACTCACTATAGGGCTTGCAGATGGCCATGGCGGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_53_Y236C	TAATACGACTCACTATAGGACACATGCAGTTGTAGTGGAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_53_N239D	TAATACGACTCACTATAGGTGTCACACATGTAGTTGTAGGTTTTAGAGCTAGAAATAGCA

[0765]

[0766] 在基于凝胶的dsDNA脱氨酶测定法中使用的80个核苷酸的未标记链和Cy3标记的通用引物的序列。SEQ ID NO:224-248分别从上至下显示。

[0767]

Cy3-primer	Cy3-GTAGGTAGTTAGGATGAATGGAAGGTTGGTA
dsDNA_2	GTCCATGGATCCAGAGGTCATCCATTAAATAAATAAATCCGCGGGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_3	GTCCATGGATCCAGAGGTCATCCATAAATAAATAAATCCGCGGAAGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_4	GTCCATGGATCCAGAGGTCATCCATAATAAATAAATCCGCGGAAGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_5	GTCCATGGATCCAGAGGTCATCCAAATAAATAAATCCGCGGAATGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_6	GTCCATGGATCCAGAGGTCATCCAATAAATAAATCCGCGGAATAGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_7	GTCCATGGATCCAGAGGTCATCCATAAATAAATCCGCGGAATAAGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_8	GTCCATGGATCCAGAGGTCATCCAAAATAAATCCGCGGAATAATGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_9	GTCCATGGATCCAGAGGTCATCCAAATAAATCCGCGGAATAATAGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_10	GTCCATGGATCCAGAGGTCATCCAATAAATCCGCGGATAATAATGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_11	GTCCATGGATCCAGAGGTCATCCATAAATCCGCGGAATATAATAGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_12	GTCCATGGATCCAGAGGTCATCCAAAATCCGCGGAATATAATAAGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_13	GTCCATGGATCCAGAGGTCATCCAAATCCGCGGAATATAATAATGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_14	GTCCATGGATCCAGAGGTCATCCAATCCGCGGAATATAATAATAGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_15	GTCCATGGATCCAGAGGTCATCCATCCGCGGTAATAATAATAATGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_18	GTCCATGGATCCAGAGGTCATCCAGCGGTAATAATAATAATAATGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_noC	GTCCATGGATCCAGAGGTCATCCATTAAATAAATAAATTAATATTACTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_8U	5Cy3-GTAGGTAGTTAGGATGAATGGAAGGTTGGTGTAGATTATTATCUGCGGATTTATTGGATGACCTCTGGATCCATGGACAT
dsDNA_APOE_C112R	GCACCTCGCCGCGGTACTGCACCAGGCGGCCGCGCACGTCCTCCATGTCTACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_APOE_C158R	CGGCGCCCTCGCGGGCCCCGGCCTGGTACACTGCCAGGCGCTTCTGCAGTACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_CTNNB1_T41A	GTCTTACCTGGACTCTGGAATCCATTCTGGTGCCACTGCCACAGCTCCTTACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_HRAS_Q61R	GGAGACGTGCCTGTTGGACATCCTGGATACCGCCGGCCGGGAGGAGTACTACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_p53_Y163C	ACCCCCGCCCGGCACCCGCGTCCGCGCCATGGCCATCTGCAAGCAGTCATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_p53_Y236C	AGGTTGGCTCTGACTGTACCACCATCCACTACAACTGCATGTGTAACAGTACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_p53_N239D	TGGCTCTGACTGTACCACCATCCACTACAACTACATGTGTGACAGTTCCTACCAACCTTCCATTCATCCTAACTACCTAC

[0768] 引物，用于产生PCR产物以充当用于高通量测序的sgRNA的T7转录的底物。在所有情况下使用rev_gRNA_T7(上方)。如材料和方法部分所述，将pFYF1320质粒用作模板。SEQ ID NO:249-300分别从上至下显示。

	fwd_sgRNA_T7_HTS_ base	TAATACGACTCACTATAGGTTATTTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
	fwd_sgRNA_T7_HTS_ 1A	TAATACGACTCACTATAGGATATTTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
	fwd_sgRNA_T7_HTS_ 1C	TAATACGACTCACTATAGGCTATTTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
	fwd_sgRNA_T7_HTS_ 1G	TAATACGACTCACTATAGGGTATTTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
	fwd_sgRNA_T7_HTS_ 2A	TAATACGACTCACTATAGGTAATTTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
	fwd_sgRNA_T7_HTS_ 2C	TAATACGACTCACTATAGGTCATTTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
	fwd_sgRNA_T7_HTS_ 2G	TAATACGACTCACTATAGGTGATTTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
	fwd_sgRNA_T7_HTS_ 3T	TAATACGACTCACTATAGGTTTTTTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
	fwd_sgRNA_T7_HTS_ 3C	TAATACGACTCACTATAGGTTCTTTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
	fwd_sgRNA_T7_HTS_ 3G	TAATACGACTCACTATAGGTTGTTTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
	fwd_sgRNA_T7_HTS_ 4A	TAATACGACTCACTATAGGTTAATTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
	fwd_sgRNA_T7_HTS_ 4C	TAATACGACTCACTATAGGTTACTTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
	fwd_sgRNA_T7_HTS_ 4G	TAATACGACTCACTATAGGTTAGTTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
	fwd_sgRNA_T7_HTS_ 5A	TAATACGACTCACTATAGGTTATATCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
	fwd_sgRNA_T7_HTS_ 5C	TAATACGACTCACTATAGGTTATCTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
	fwd_sgRNA_T7_HTS_ 5G	TAATACGACTCACTATAGGTTATGTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
[0769]	fwd_sgRNA_T7_HTS_ 6A	TAATACGACTCACTATAGGTTATTACGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
	fwd_sgRNA_T7_HTS_ 6C	TAATACGACTCACTATAGGTTATTCCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
	fwd_sgRNA_T7_HTS_ 6G	TAATACGACTCACTATAGGTTATTGCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
	fwd_sgRNA_T7_HTS_ 8A	TAATACGACTCACTATAGGTTATTTCATGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
	fwd_sgRNA_T7_HTS_ 8T	TAATACGACTCACTATAGGTTATTTCTTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
	fwd_sgRNA_T7_HTS_ 8C	TAATACGACTCACTATAGGTTATTTCCTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
	fwd_sgRNA_T7_HTS_ 9A	TAATACGACTCACTATAGGTTATTTCGAGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
	fwd_sgRNA_T7_HTS_ 9C	TAATACGACTCACTATAGGTTATTTCGCGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
	fwd_sgRNA_T7_HTS_ 9G	TAATACGACTCACTATAGGTTATTTCGGGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
	fwd_sgRNA_T7_HTS_ 10A	TAATACGACTCACTATAGGTTATTTCGTAGATTTATTTAGTTTTAGAGCTAGAAATAGCA
	fwd_sgRNA_T7_HTS_ 10T	TAATACGACTCACTATAGGTTATTTCGTTGATTTATTTAGTTTTAGAGCTAGAAATAGCA
	fwd_sgRNA_T7_HTS_ 10C	TAATACGACTCACTATAGGTTATTTCGTCGATTTATTTAGTTTTAGAGCTAGAAATAGCA
	fwd_sgRNA_T7_HTS_ 11A	TAATACGACTCACTATAGGTTATTTCGTGAATTTATTTAGTTTTAGAGCTAGAAATAGCA
	fwd_sgRNA_T7_HTS_ 11T	TAATACGACTCACTATAGGTTATTTCGTGTATTTATTTAGTTTTAGAGCTAGAAATAGCA
	fwd_sgRNA_T7_HTS_ 11C	TAATACGACTCACTATAGGTTATTTCGTGCATTTATTTAGTTTTAGAGCTAGAAATAGCA
	fwd_sgRNA_T7_HTS_ 12T	TAATACGACTCACTATAGGTTATTTCGTGGTTTTATTTAGTTTTAGAGCTAGAAATAGCA
	fwd_sgRNA_T7_HTS_ 12C	TAATACGACTCACTATAGGTTATTTCGTGGCTTTATTTAGTTTTAGAGCTAGAAATAGCA

[0770]

fwd_sgRNA_T7_HTS_ 12G	TAATACGACTCACTATAGGTTATTTCGTGGGTTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 13A	TAATACGACTCACTATAGGTTATTTCGTGGAATTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 13C	TAATACGACTCACTATAGGTTATTTCGTGGACTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 13G	TAATACGACTCACTATAGGTTATTTCGTGGAGTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ multiC	TAATACGACTCACTATAGGTTCCCCCCCCGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ TCGCACCC_odd	TAATACGACTCACTATAGGCGCACCCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ CCTCGCAC_odd	TAATACGACTCACTATAGGCTCGCACGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ ACCCTCGC_odd	TAATACGACTCACTATAGGCCCTCGCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ GCACCCTC_odd	TAATACGACTCACTATAGGCACCCTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ TCGCACCC_even	TAATACGACTCACTATAGGTCGCACCCGTGGATTTATTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ CCTCGCAC_even	TAATACGACTCACTATAGGCCTCGCACGTGGATTTATTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ ACCCTCGC_even	TAATACGACTCACTATAGGACCCTCGCGTGGATTTATTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ GCACCCTC_even	TAATACGACTCACTATAGGGCACCCTCGTGGATTTATTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ EMX1	TAATACGACTCACTATAGGGAGTCCGAGCAGAAGAAGAAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ FANCF	TAATACGACTCACTATAGGGGAATCCCTTCTGCAGCACCGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ HEK293_site2	TAATACGACTCACTATAGGGAACACAAAGCATAGACTGCGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ HEK293_site3	TAATACGACTCACTATAGGGGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ HEK293_site4	TAATACGACTCACTATAGGGGCACTGCGGCTGGAGGTGGGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ RNF2	TAATACGACTCACTATAGGGTCATCTTAGTCATTACCTGGTTTTAGAGCTAGAAATAGCA

[0771] 用于高通量测序(HTS)的体外编辑的dsDNA的序列。显示的是经编辑的链的序列。还获得了所示所有序列的反向互补物。如材料和方法中所述,通过退火互补链获得dsDNA底物。代表EMX1,FANCF,HEK293位点2,HEK293位点3,HEK293位点4和RNF2基因座的寡核苷酸最初设计用于基于凝胶的脱氨酶测定法,因此在其5'-端具有相同的25-nt序列(与Cy3引物匹配)。SEQ ID NO:301-352分别从上至下显示。

[0772] 实施例2:

[0773]

Base sequence	ACGTAAACGGCCACAAGTTCTTATTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
1A	ACGTAAACGGCCACAAGTTCATATTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
1C	ACGTAAACGGCCACAAGTTCCTATTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
1G	ACGTAAACGGCCACAAGTTCGTATTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
2A	ACGTAAACGGCCACAAGTTCTAATTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG

[0774]

2C	ACGTAAACGGCCACAAGTTCTCATTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
2G	ACGTAAACGGCCACAAGTTCTGATTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
3T	ACGTAAACGGCCACAAGTTCTTTTTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
3C	ACGTAAACGGCCACAAGTTCTTCTTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
3G	ACGTAAACGGCCACAAGTTCTTGTTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
4A	ACGTAAACGGCCACAAGTTCTTAATTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
4C	ACGTAAACGGCCACAAGTTCTTACTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
4G	ACGTAAACGGCCACAAGTTCTTAGTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
5A	ACGTAAACGGCCACAAGTTCTTATATCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
5C	ACGTAAACGGCCACAAGTTCTTATCTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
5G	ACGTAAACGGCCACAAGTTCTTATGTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
6A	ACGTAAACGGCCACAAGTTCTTATTACGTGGATTTATTTATGGCATCTTCTTCAAGGACG
6C	ACGTAAACGGCCACAAGTTCTTATTCCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
6G	ACGTAAACGGCCACAAGTTCTTATTGCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
8A	ACGTAAACGGCCACAAGTTCTTATTTCATGGATTTATTTATGGCATCTTCTTCAAGGACG
8T	ACGTAAACGGCCACAAGTTCTTATTTCTTGGATTTATTTATGGCATCTTCTTCAAGGACG
8C	ACGTAAACGGCCACAAGTTCTTATTTCCTGGATTTATTTATGGCATCTTCTTCAAGGACG
9A	ACGTAAACGGCCACAAGTTCTTATTTCGAGGATTTATTTATGGCATCTTCTTCAAGGACG
9C	ACGTAAACGGCCACAAGTTCTTATTTCGCGGATTTATTTATGGCATCTTCTTCAAGGACG
9G	ACGTAAACGGCCACAAGTTCTTATTTCGGGGATTTATTTATGGCATCTTCTTCAAGGACG
10A	ACGTAAACGGCCACAAGTTCTTATTTCGTAGATTTATTTATGGCATCTTCTTCAAGGACG
10T	ACGTAAACGGCCACAAGTTCTTATTTCGTTGATTTATTTATGGCATCTTCTTCAAGGACG
10C	ACGTAAACGGCCACAAGTTCTTATTTCGTCGATTTATTTATGGCATCTTCTTCAAGGACG
11A	ACGTAAACGGCCACAAGTTCTTATTTCGTGAATTTATTTATGGCATCTTCTTCAAGGACG
11T	ACGTAAACGGCCACAAGTTCTTATTTCGTGTATTTATTTATGGCATCTTCTTCAAGGACG
11C	ACGTAAACGGCCACAAGTTCTTATTTCGTGCATTTATTTATGGCATCTTCTTCAAGGACG
12T	ACGTAAACGGCCACAAGTTCTTATTTCGTGGTTTTATTTATGGCATCTTCTTCAAGGACG
12C	ACGTAAACGGCCACAAGTTCTTATTTCGTGGCTTTATTTATGGCATCTTCTTCAAGGACG
12G	ACGTAAACGGCCACAAGTTCTTATTTCGTGGGTTTATTTATGGCATCTTCTTCAAGGACG
13A	ACGTAAACGGCCACAAGTTCTTATTTCGTGGAATTATTTATGGCATCTTCTTCAAGGACG
13C	ACGTAAACGGCCACAAGTTCTTATTTCGTGGACTTATTTATGGCATCTTCTTCAAGGACG
13G	ACGTAAACGGCCACAAGTTCTTATTTCGTGGAGTTATTTATGGCATCTTCTTCAAGGACG
multiC	ACGTAAACGGCCACAAGTTCTTCCCCCCCCGATTTATTTATGGCATCTTCTTCAAGGACG
TCGCACCC_odd	ACGTAAACGGCCACAAGTTTCGCACCCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
CCTCGCAC_odd	ACGTAAACGGCCACAAGTTCCTCGCACGTGGATTTATTTATGGCATCTTCTTCAAGGACG
ACCCTCGC_odd	ACGTAAACGGCCACAAGTTACCCTCGCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
GCACCCTC_odd	ACGTAAACGGCCACAAGTTGCACCCTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
TCGCACCC_even	ACGTAAACGGCCACAAGTATTCGCACCCGTGGATTTATTATGGCATCTTCTTCAAGGACG
CCTCGCAC_even	ACGTAAACGGCCACAAGTATCCTCGCACGTGGATTTATTATGGCATCTTCTTCAAGGACG

[0775]

ACCCTCGC_even	ACGTAAACGGCCACAAGTATACCCTCGCGTGGATTTATTATGGCATCTTCTTCAAGGACG
GCACCCTC_even	ACGTAAACGGCCACAAGTATGCACCCTCGTGGATTTATTATGGCATCTTCTTCAAGGACG
EMX1_invitro	GTAGGTAGTTAGGATGAATGGAAGGTTGGTAGGCCTGAGTCCGAGCAGAAGAAGAAGGGCTCCCATCACATCAACCGGTG
FANCF_invitro	GTAGGTAGTTAGGATGAATGGAAGGTTGGTACTCATGGAATCCCTTCTGCAGCACCTGGATCGCTTTTCCGAGCTTCTGG
HEK293_site2_invitro	GTAGGTAGTTAGGATGAATGGAAGGTTGGTAAACTGGAACACAAAGCATAGACTGCGGGGCGGGCCAGCCTGAATAGCTG
HEK293_site3_invitro	GTAGGTAGTTAGGATGAATGGAAGGTTGGTACTTGGGGCCCAGACTGAGCACGTGATGGCAGAGGAAAGGAAGCCCTGCT
HEK293_site4_invitro	GTAGGTAGTTAGGATGAATGGAAGGTTGGTACCGGTGGCACTGCGGCTGGAGGTGGGGGTTAAAGCGGAGACTCTGGTGC
RNF2_invitro	GTAGGTAGTTAGGATGAATGGAAGGTTGGTATGGCAGTCATCTTAGTCATTACCTGAGGTGTTCGTTGTAACTCATATAA

[0776] 用于体外编辑的dsDNA的HTS的引物。SEQ ID NOs:353-361分别从上到下出现。

[0777]

fwd_invitro_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNACGTAAACGGCCACAA
rev_invitro_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCGTCCTTGAAGAAGATGC
fwd_invitro_HEK_targets	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTAGGTAGTTAGGATGAATGGAA
rev_EMX1_invitro	TGGAGTTCAGACGTGTGCTCTTCCGATCTCACCGGTTGATGTGATGG
rev_FANCF_invitro	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCAGAAGCTCGGAAAAGC
rev_HEK293_site2_invitro	TGGAGTTCAGACGTGTGCTCTTCCGATCTCAGCTATTCAGGCTGGC
rev_HEK293_site3_invitro	TGGAGTTCAGACGTGTGCTCTTCCGATCTAGCAGGGCTTCCTTTC
rev_HEK293_site4_invitro	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCACCAGAGTCTCCG
rev_RNF2_invitro	TGGAGTTCAGACGTGTGCTCTTCCGATCTTTATATGAGTTACAACGAACACC

[0778] 用于来自所有哺乳动物细胞培养实验的中靶和脱靶位点的HTS的引物。SEQ ID NOs:362-469分别从上到下出现。

[0779]

fwd_EMX1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCAGCTCAGCCTGAGTGTTGA
rev_EMX1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTCGTGGGTTTGTGGTTGC
fwd_FANCF_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCATTGCAGAGAGGCGTATCA
rev_FANCF_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGGGGTCCCAGGTGCTGAC
fwd_HEK293_site2_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCAGCCCCATCTGTCAAACT
rev_HEK293_site2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTGAATGGATTCCTTGGAAACAATGA
fwd_HEK293_site3_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNATGTGGGCTGCCTAGAAAGG
rev_HEK293_site3_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCCAGCCAAACTTGTCAACC
fwd_HEK293_site4_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGAACCCAGGTAGCCAGAGAC
rev_HEK293_site4_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCCTTTCAACCCGAACGGAG
fwd_RNF2_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCTCTTCTTTATTTCCAGCAATGT
rev_RNF2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGTTTTCATGTTCTAAAAATGTATCCCA
fwd_p53_Y163C_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTACAGTACTCCCCTGCCCTC
rev_p53_Y163C_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCTGCTCACCATCGCTATCT
fwd_p53_N239D_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCTCATCTTGGGCCTGTGTT
rev_p53_N239D_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTAAATCGGTAAGAGGTGGGCC

	fwd_APOE4_C158R_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGCGGACATGGAGGACGTG
	rev_APOE4_C158R_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTGTTCCACCAGGGGCCC
	fwd_EMX1_off1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTGCCCAATCATTGATGCTTTT
	rev_EMX1_off1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTAGAAACATTTACCATAGACTATCACCT
	fwd_EMX1_off2_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAGTAGCCTCTTTCTCAATGTGC
	rev_EMX1_off2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCTTTCACAAGGATGCAGTCT
	fwd_EMX1_off3_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGAGCTAGACTCCGAGGGGA
	rev_EMX1_off3_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCCTCGTCCTGCTCTCACTT
	fwd_EMX1_off4_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAGAGGCTGAAGAGGAAGACCA
	rev_EMX1_off4_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGGCCCAGCTGTGCATTCTAT
	fwd_EMX1_off6_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCAAGAGGGCCAAGTCCTG
	rev_EMX1_off6_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCAGCGAGGAGTGACAGCC
	fwd_EMX1_off7_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCACTCCACCTGATCTCGGGG
	rev_EMX1_off7_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCGAGGAGGGAGGGAGCAG
	fwd_EMX1_off8_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNACCACAAATGCCCAAGAGAC
	rev_EMX1_off8_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGACACAGTCAAGGGCCGG
	fwd_EMX1_off9_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCCACCTTTGAGGAGGCAAA
	rev_EMX1_off9_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTTCCATCTGAGAAGAGAGTGGT
	fwd_EMX1_off10_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTCATACCTTGGCCCTTCCT
[0780]	rev_EMX1_off10_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCCCTAGGCCCACACCAG
	fwd_FANCF_off1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAACCCACTGAAGAAGCAGGG
	rev_FANCF_off1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGGTGCTTAATCCGGCTCCAT
	fwd_FANCF_off2_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCCAGTGTTTCCATCCCGAA
	rev_FANCF_off2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCTCTGACCTCCACAACTCT
	fwd_FANCF_off3_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCTGGGTACAGTTCTGCGTGT
	rev_FANCF_off3_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCACTCTGAGCATCGCCAAG
	fwd_FANCF_off4_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGGTTTAGAGCCAGTGAACTAGAG
	rev_FANCF_off4_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCAAGACAAAATCCTCTTTATACTTTG
	fwd_FANCF_off5_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGGGAGGGGACGGCCTTAC
	rev_FANCF_off5_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCCTCTGGCGAACATGGC
	fwd_FANCF_off6_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCCTGGTTAAGAGCATGGGC
	rev_FANCF_off6_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGATTGAGTCCCCACAGCACA
	fwd_FANCF_off7_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCAGTGTTTCCCATCCCCAA
	rev_FANCF_off7_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTGACCTCCACAACTGGAAAAT
	fwd_FANCF_off8_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGCTTCCAGACCCACCTGAAG
	rev_FANCF_off8_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTACCGAGGAAAATTGCTTGTCG
	fwd_HEK293_site2_off1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTGTGGAGAGTGAGTAAGCCA
	rev_HEK293_site2_off1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTACGGTAGGATGATTTCAGGCA
	fwd_HEK293_site2_off2_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCACAAAGCAGTGTAGCTCAGG

[0781]

rev_HEK293_site2_off2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTTTTTGGTACTCGAGTGTTATTCAG
fwd_HEK293_site3_off1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCCCCTGTTGACCTGGAGAA
rev_HEK293_site3_off1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCACTGTACTTGCCCTGACCA
fwd_HEK293_site3_off2_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTGGTGTTGACAGGGAGCAA
rev_HEK293_site3_off2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTGAGATGTGGGCAGAAGGG
fwd_HEK293_site3_off3_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTGAGAGGGAACAGAAGGGCT
rev_HEK293_site3_off3_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGTCCAAAGGCCCAAGAACCT
fwd_HEK293_site3_off4_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCCTAGCACTTTGGAAGGTCG
rev_HEK293_site3_off4_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCTCATCTTAATCTGCTCAGCC
fwd_HEK293_site3_off5_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAAAGGAGCAGCTCTTCCTGG
rev_HEK293_site3_off5_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGTCTGCACCATCTCCCACAA
fwd_HEK293_site4_off1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGGCATGGCTTCTGAGACTCA
rev_HEK293_site4_off1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGTCTCCCTTGCACTCCCTGTCTTT
fwd_HEK293_site4_off2_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTTGGCAATGGAGGCATTGG
rev_HEK293_site4_off2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGAAGAGGCTGCCCATGAGAG
fwd_HEK293_site4_off3_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGGTCTGAGGCTCGAATCCTG
rev_HEK293_site4_off3_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTGTGGCCTCCATATCCCTG
fwd_HEK293_site4_off4_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTTCCACCAGAACTCAGCCC
rev_HEK293_site4_off4_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCTCGGTTCCTCCACAACAC
fwd_HEK293_site4_off5_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCACGGGAAGGACAGGAGAAG
rev_HEK293_site4_off5_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCAGGGGAGGGATAAAGCAG
fwd_HEK293_site4_off6_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCACGGGAGATGGCTTATGT
rev_HEK293_site4_off6_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCACATCCTCACTGTGCCACT
fwd_HEK293_site4_off7_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTCAGTCTCGGCCCCTCA
rev_HEK293_site4_off7_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCCACTGTAAAGCTCTTGGG
fwd_HEK293_site4_off8_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAGGGTAGAGGGACAGAGCTG
rev_HEK293_site4_off8_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGGACCCCACATAGTCAGTGC
fwd_HEK293_site4_off9_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGCTGTCAGCCCTATCTCCATC
rev_HEK293_site4_off9_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTGGGCAATTAGGACAGGGAC
fwd_HEK293_site4_off10_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGCAGCGGAGGAGGTAGATTG
rev_HEK293_site4_off10_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTCAGTACCTGGAGTCCCGA
fwd_HEK2_ChIP_off1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGACAGGCTCAGGAAAGCTGT
rev_HEK2_ChIP_off1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTACACAAGCCTTTCTCCAGGG
fwd_HEK2_ChIP_off2_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAATAGGGGGTGAGACTGGGG
rev_HEK2_ChIP_off2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCCTCAGACGAGACTTGAGG
fwd_HEK2_ChIP_off3_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGGCCAGCAGGAAAGGAATCT
rev HEK2 ChIP off3 HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTGACTGCACCTGTAGCCATG
fwd_HEK2_ChIP_off4_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCAAGGAAATCACCCTGCCC
rev_HEK2_ChIP_off4_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTAACTTCCTTGGTGTGCAGCT
fwd_HEK2_ChIP_off5_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNATGGGCTCAGCTACGTCATG
rev HEK2 ChIP off5 HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTAATAGCAGTGTGGTGGGCAA

[0782]

fwd_HEK3_ChIP_off1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCGCACATCCCTTGTCTCTCT
rev_HEK3_ChIP_off1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTACTGGAGCACACCCCAAG
fwd_HEK3_ChIP_off2_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTGGGTCACGTAGCTTTGGTC
rev_HEK3_ChIP_off2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTGGTGGCCATGTGCAACTAA
fwd_HEK3_ChIP_off3_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCTACTACGTGCCAGCTCAGG
rev_HEK3_ChIP_off3_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTACCTCCCCTCCTCACTAACC
fwd_HEK3_ChIP_off4_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGCCTCAGCTCCATTTCCTGT
rev_HEK3_ChIP_off4_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTAACCTTTATGGCACCAGGGG
fwd_HEK3_ChIP_off5_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGAGCTCAGCATTAGCAGGCT
rev_HEK3_ChIP_off5_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTTCCTGGCTTTCCGATTCCC
fwd_HEK4_ChIP_off1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTGCAATTGGAGGAGGAGCT
rev_HEK4_ChIP_off1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCACCAGCTACAGGCAGAACA
fwd_HEK4_ChIP_off3_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCTACCCCCAACACAGATGG
rev_HEK4_ChIP_off3_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCACACAACTCAGGTCCTCC

[0783] HDR研究中使用的单链寡核苷酸供体模板(ssODNs)的序列。

[0784] EMX1有义(SEQ ID NO:470)

[0785] TCATCTGTGCCCCTCCCTCCCTGGCCCAGGTGAAGGTGTGGTTCCAGAACCGGAGGACAAAGTACA AACGGCAGAAGCTGGAGGAGGAAGGGCCTGAGTTTGAGCAGAAGAAGAAGGGCTCCCATCACATC AACCGGTGGCGCATTGCCACGAAGCAGGCCAATGGGGAGGACATCGATGTCACCTCCAATGACTAG GGT

[0786] EMX1反义(SEQ ID NO:471)

[0787] ACCCTAGTCATTGGAGGTGACATCGATGTCCTCCCCATTGGCCTGCTTCGTGGCAATGCGCCACCG GTTGATGTGATGGGAGCCCTTCTTCTTCTGCTCAAACTCAGGCCCTTCCTCCTCCAGCTTCTGCCGT TTGTACTTTGTCCTCCGGTTCTGGAACCACACCTTCACCTGGGCCAGGGAGGGAGGGGCACAGATG A

[0788] HEK293位点3有义(SEQ ID NO:472)

[0789] CATGCAATTAGTCTATTTCTGCTGCAAGTAAGCATGCATTTGTAGGCTTGATGCTTTTTTTCTGCTTCT CCAGCCCTGGCCTGGGTCAATCCTTGGGGCTTAGACTGAGCACGTGATGGCAGAGGAAAGGAAGC CCTGCTTCCTCCAGAGGGCGTCGCAGGACAGCTTTTCCTAGACAGGGGCTAGTATGTGCAGCTCCT

[0790] HEK293 site 3antisense(SEQ ID NO:473)

[0791] AGGAGCTGCACATACTAGCCCCTGTCTAGGAAAAGCTGTCCTGCGACGCCCTCTGGAGGAAGCAGG GCTTCCTTTCCTCTGCCATCACGTGCTCAGTCTAAGCCCCAAGGATTGACCCAGGCCAGGGCTGGA GAAGCAGAAAAAAAGCATCAAGCCTACAAATGCATGCTTACTTGCAGCAGAAATAGACTAATTGCATG

[0792] HEK位点4有义(SEQ ID NO:474)

[0793] GGCTGACAAAGGCCGGGCTGGGTGGAAGGAAGGGAGGAAGGGCGAGGCAGAGGGTCCAAAGCAG GATGACAGGCAGGGGCACCGCGGCGCCCCGGTGGCATTGCGGCTGGAGGTGGGGGTTAAAGCGG AGACTCTGGTGCTGTGTGACTACAGTGGGGGCCCTGCCCTCTCTGAGCCCCCGCCTCCAGGCCTGT GTGTGT

[0794] HEK位点4反义(SEQ ID NO:475)

[0795] ACACACACAGGCCTGGAGGCGGGGGCTCAGAGAGGGCAGGGCCCCCACTGTAGTCACACAGCACC AGAGTCTCCGCTTTAACCCCCACCTCCAGCCGCAATGCCACCGGGGCGCCGCGGTGCCCCTGCCT GTCATCCTGCTTTGGACCCTCTGCCTCGCCCTTCCTCCCTTCCTTCCACCCAGCCCGGCCTTTGTCA GCC

[0796] APOE4有义(SEQ ID NO:476)

[0797] AGCACCGAGGAGCTGCGGGTGCGCCTCGCCTCCCACCTGCGCAAGCTGCGTAAGCGGCTCCTCCG CGATGCCGATGACCTGCAGAAGTGCCTGGCAGTGTACCAGGCCGGGGCCCGCGAGGGCGCCGAG CGCGGCCTCAGCGCCATCCGCGAGCGCCTGGGGCCCCTGGTGGAACAGGGCCGCGTGCGGGCCG CCACTGT

[0798] APOE4反义(SEQ ID NO:477)

[0799] ACAGTGGCGGCCCGCACGCGGCCCTGTTCCACCAGGGGCCCCAGGCGCTCGCGGATGGCGCTGA GGCCGCGCTCGGCGCCCTCGCGGGCCCCGGCCTGGTACACTGCCAGGCACTTCTGCAGGTCATCG GCATCGCGGAGGAGCCGCTTACGCAGCTTGCGCAGGTGGGAGGCGAGGCGCACCCGCAGCTCCT CGGTGCT

[0800] p53 Y163C有义(SEQ ID NO:478)

[0801] ACTCCCCTGCCCTCAACAAGATGTTTTGCCAACTGGCCAAGACCTGCCCTGTGCAGCTGTGGGTTGA TTCCACACCCCCGCCCGGCACCCGCGTCCGCGCCATGGCCATCTACAAGCAGTCACAGCACATGAC GGAGGTTGTGAGGCGCTGCCCCCACCATGAGCGCTGCTCAGATAGCGATGGTGAGCAGCTGGGGC TG

[0802] p53 Y163C反义(SEQ ID NO:479)

[0803] CAGCCCCAGCTGCTCACCATCGCTATCTGAGCAGCGCTCATGGTGGGGGCAGCGCCTCACAACCTC CGTCATGTGCTGTGACTGCTTGTAGATGGCCATGGCGCGGACGCGGGTGCCGGGCGGGGGTGTGG AATCAACCCACAGCTGCACAGGGCAGGTCTTGGCCAGTTGGCAAAACATCTTGTTGAGGGCAGGGG AGT

[0804] 脱氨酶基因gBlocks基因片段

[0805] hAID(SEQ ID NO:169)

[0806] rAPOBEC1(哺乳动物)(SEQ ID NO:170)

[0807] CATCCTTGGTACCGAGCTCGGATCCAGCCACCATGAGCTCAGAGACTGGCCCAGTGGCTGTGGACC CCACATTGAGACGGCGGATCGAGCCCCATGAGTTTGAGGTATTCTTCGATCCGAGAGAGCTCCGCA AGGAGACCTGCCTGCTTTACGAAATTAATTGGGGGGGCCGGCACTCCATTTGGCGACATACATCACA GAACACTAACAAGCACGTCGAAGTCAACTTCATCGAGAAGTTCACGACAGAAAGATATTTCTGTCCG AACACAAGGTGCAGCATTACCTGGTTTCTCAGCTGGAGCCCATGCGGCGAATGTAGTAGGGCCATC ACTGAATTCCTGTCAAGGTATCCCCACGTCACTCTGTTTATTTACATCGCAAGGCTGTACCACCACGC TGACCCCCGCAATCGACAAGGCCTGCGGGATTTGATCTCTTCAGGTGTGACTATCCAAATTATGACT GAGCAGGAGTCAGGATACTGCTGGAGAAACTTTGTGAATTATAGCCCGAGTAATGAAGCCCACTGG CCTAGGTATCCCCATCTGTGGGTACGACTGTACGTTCTTGAACTGTACTGCATCATACTGGGCCTGC CTCCTTGTCTCAACATTCTGAGAAGGAAGCAGCCACAGCTGACATTCTTTACCATCGCTCTTCAGTCT TGTCATTACCAGCGACTGCCCCCACACATTCTCTGGGCCACCGGGTTGAAATGAGCGGCCGCTCGA TTGGTTTGGTGTGGCTCTAA

[0808] pmCDA1(SEQ ID NO:171)

[0809] CATCCTTGGTACCGAGCTCGGATCCAGCCACCATGACAGACGCTGAATATGTTAGGATCCATGAAAA ACTGGATATCTATACATTTAAGAAGCAGTTCTTCAATAACAAAAAGTCAGTATCTCACAGATGCTATGT CCTGTTCGAACTCAAGAGAAGAGGAGAAAGGCGGGCCTGTTTCTGGGGGTACGCGGTTAATAAACC CCAGTCCGGGACCGAGAGGGGGATTCACGCCGAGATCTTTTCAATTAGGAAGGTTGAAGAGTATCT TCGCGACAATCCCGGTCAGTTCACAATTAACTGGTACAGCTCCTGGAGCCCTTGCGCTGATTGCGCC GAGAAAATACTCGAATGGTACAACCAGGAGTTGAGAGGCAATGGCCACACTCTCAAGATTTGGGCTT GCAAGCTTTACTACGAGAAGAACGCGAGAAATCAGATTGGCTTGTGGAACCTCAGGGACAACGGGG TCGGGTTGAATGTTATGGTGTCCGAACATTACCAGTGCTGTAGAAAGATCTTCATTCAGTCCAGTCAC AATCAGCTGAACGAGAACAGATGGCTGGAGAAAACACTGAAACGGGCAGAGAAAAGGCGCTCAGAG CTGAGTATCATGATCCAGGTCAAAATCCTGCATACAACCAAAAGCCCGGCTGTATAAGCGGCCGCTC GATTGGTTTGGTGTGGCTCTAA

[0810] haPOBEC3G(SEQ ID NO:172)

[0811]
CATCCTTGGTACCGAGCTCGGATCCAGCCACCATGGAGCTGAAGTATCACCCTGAGATGCGGTTTTT
CCACTGGTTTAGTAAGTGGCGCAAACTTCATCGGGATCAGGAGTATGAAGTGACCTGGTATATCTCT
TGGTCTCCCTGCACAAAATGTACACGCGACATGGCCACATTTCTGGCCGAGGATCCAAAGGTGACG
CTCACAATCTTTGTGGCCCGCCTGTATTATTTCTGGGACCCGGATTATCAGGAGGCACTTAGGTCAT
TGTGCCAAAAGCGCGACGGACCACGGGCGACTATGAAAATCATGAATTATGACGAATTCCAGCATTG
CTGGAGTAAGTTTGTGTACAGCCAGCGGGAGCTGTTCGAGCCCTGGAACAATCTTCCCAAGTACTAC
ATACTGCTTCACATTATGTTGGGGGAGATCCTTCGGCACTCTATGGATCCTCCTACCTTTACGTTTAA
CTTTAATAATGAGCCTTGGGTTCGCGGGCGCCATGAAACCTATTTGTGCTACGAGGTCGAGCGGATG
CATAATGATACGTGGGTCCTGCTGAATCAGAGGAGGGGGTTTCTGTGTAACCAGGCTCCACATAAAC
ATGGATTTCTCGAGGGGCGGCACGCCGAACTGTGTTTCCTTGATGTGATACCTTTCTGGAAGCTCGA
CCTTGATCAAGATTACAGGGTGACGTGTTTCACCTCCTGGTCACCCTGCTTCAGTTGCGCCCAAGAG
ATGGCTAAATTTATCAGTAAGAACAAGCATGTGTCCCTCTGTATTTTTACAGCCAGAATTTATGATGAC
CAGGGCCGGTGCCAGGAGGGGCTGCGGACACTCGCTGAGGCGGGCGCGAAGATCAGCATAATGA
CATACTCCGAATTCAAACACTGTTGGGACACTTTTGTGGACCACCAGGGCTGCCCATTTCAGCCGTG
GGATGGGCTCGACGAACATAGTCAGGATCTCTCAGGCCGGCTGCGAGCCATATTGCAGAACCAGGA
GAATTAGGCGGCCGCTCGATTGGTTTGGTGTGGCTCTAA

[0812] rAPOBEC1(大肠杆菌)(SEQ ID NO:173)

[0813]
GGCCGGGGATTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGATGTCTTCTGAAA
CCGGTCCGGTTGCGGTTGACCCGACCCTGCGTCGTCGTATCGAACCGCACGAATTCGAAGTTTTCT
TCGACCCGCGTGAACTGCGTAAAGAAACCTGCCTGCTGTACGAAATCAACTGGGGTGGTCGTCACT
CTATCTGGCGTCACACCTCTCAGAACACCAACAAACACGTTGAAGTTAACTTCATCGAAAAATTCACC
ACCGAACGTTACTTCTGCCCGAACACCCGTTGCTCTATCACCTGGTTCCTGTCTTGGTCTCCGTGCG
GTGAATGCTCTCGTGCGATCACCGAATTCCTGTCTCGTTACCCGCACGTTACCCTGTTCATCTACATC
GCGCGTCTGTACCACCACGCGGACCCGCGTAACCGTCAGGGTCTGCGTGACCTGATCTCTTCTGGT
GTTACCATCCAGATCATGACCGAACAGGAATCTGGTTACTGCTGGCGTAACTTCGTTAACTACTCTCC
GTCTAACGAAGCGCACTGGCCGCGTTACCCGCACCTGTGGGTTCGTCTGTACGTTCTGGAACTGTA
CTGCATCATCCTGGGTCTGCCGCCGTGCCTGAACATCCTGCGTCGTAAACAGCCGCAGCTGACCTT

CTTCACCATCGCGCTGCAGTCTTGCCACTACCAGCGTCTGCCGCCGCACATCCTGTGGGCGACCGG
TCTGAAAGGTGGTAGTGGAGGGAGCGGCGGTTCAATGGATAAGAAATAC

[0814] NBE1,NBE2和NBE3的氨基酸序列。

[0815] 用于大肠杆菌表达的NBE1(His$_6$-rAPOBEC1-XTEN-dCas9)(SEQ ID NO:154)

[0816]
MGSSHHHHHHMSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTN
KHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGL
RDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQ
LTFFTIALQSCHYQRLPPHILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSK
KFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRL
EESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSL
GLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSA
SMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELL
VKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQL
LNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLK
SKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIG
KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSS
FEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKL
KGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNL
GAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKV

[0817] 用于哺乳动物表达的NBE1(rAPOBEC1-XTEN-dCas9-NLS)(SEQ ID NO:155)

[0818] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKF
TTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQ
IMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSC
HYQRLPPHILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDR
HSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLF
IQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDL
AEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQ
DLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP
WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQ
[0818] KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILE
DIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVI
EMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI
NRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF
YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIM
NFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKG
YKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTI
DRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKV

[0819] 具有人APOBEC1的用于哺乳动物表达的备选NBE1(hAPOBEC1-XTEN-dCas9-NLS)(SEQ ID NO:158)

[0820] MTSEKGPSTGDPTLRRRIEPWEFDVFYDPRELRKEACLLYEIKWGMSRKIWRSSGKNTTNHVEVNFIK
KFTSERDFHPSMSCSITWFLSWSPCWECSQAIREFLSRHPGVTLVIYVARLFWHMDQQNRQGLRDLVNSGVTIQIM
RASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKISRRWQNHLTFFRLHLQNCHYQTIPP
HILLATGLIHPSVAWRGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKN
LIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI
VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE
NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDD
LDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ
EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNL
PNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSV
EISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRR
RYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGS
PAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNY
WRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITL
KSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATA
KYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESI
LPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKG
YKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQH
KHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTS
TKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKVNBE2(rAPOBEC1-XTEN-dCas9-UGI-NLS)(SEQ ID NO:156)

[0821] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKF
TTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQ
IMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSC
HYQRLPPHILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDR
HSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLF
IQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDL
AEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQ
DLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP
WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQ
KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILE
DIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVI
EMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI
NRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF
YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIM
NFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKG
YKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTI
DRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIG
NKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV

[0822] NBE3(rAPOBEC1-XTEN-Cas9n-UGI-NLS)(SEQ ID NO:157)

[0823] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKF
TTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQ
IMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSC
HYQRLPPHILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDR
HSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLF
IQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDL
AEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQ
DLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP
WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQ
KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILE
DIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVI
EMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI
NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF
YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIM
NFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKG
YKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTI
DRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIG
NKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV

[0824] pmCDA1-XTEN-dCas9-UGI(细菌)(SEQ ID NO:159)

[0825] MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQSGTERGIHAEI
FSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLKIWACKLYYEKNARNQIGLWNLRDNGV
GLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKRRSELSIMIQVKILHTTKSPAVSGSETPGTSESATPE
SDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKN
RICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLI
YLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPG
EKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRV
NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDG
TEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF

AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF
LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDIL
EDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF
MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTT
QKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFL
KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVET
RQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKY
PKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGR
DFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSK
KLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKY
VNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI
IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSMTNLSDIIEKETGK
QLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML

[0826] pmCDA1-XTEN-nCas9-UGI-NLS(哺乳动物构建体)(SEQ ID NO:160)

[0827] MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQSGTERGIHAEI
FSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLKIWACKLYYEKNARNQIGLWNLRDNGV
GLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKRRSELSIMIQVKILHTTKSPAVSGSETPGTSESATPE
SDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKN
RICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLI
YLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPG
EKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRV
NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDG
TEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF
LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDIL
EDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF
MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTT
QKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFL
KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVET
RQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKY
PKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGR
DFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSK
KLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKY
VNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI
IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQ
LVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKK
RKV

[0828] huAPOBEC3G-XTEN-dCas9-UGI(细菌)(SEQ ID NO:161)

[0829] MDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDV

IPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTYSE
FKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDE
YKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRL
EESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN
SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSN
FDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDL
TLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGS
IPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGA
SAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQ
LKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVP
SEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDK
LIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI
DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQ
LFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDR
KRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSMTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPE
SDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML

[0830] huAPOBEC3G-XTEN-nCas9-UGI-NLS(哺乳动物构建体)(SEQ ID NO:162)

[0831] MDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDV
IPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTYSE
FKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDE
YKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRL
EESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN
SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSN
FDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDL
TLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGS
IPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGA
SAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQ
LKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVP
SEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDK
LIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT

GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI
DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQ
LFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDR
KRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPES
DILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV

[0832] huAPOBEC3G(D316R_D317R)-XTEN-nCas9-UGI-NLS(哺乳动物构建体)(SEQ ID NO: 163)

[0833] MDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDV
IPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIYRRQGRCQEGLRTLAEAGAKISIMTYSE
FKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDE
YKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRL
EESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN
SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSN
FDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDL
TLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGS
IPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGA
SAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQ
LKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVP
SEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDK
LIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI
DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQ
LFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDR
KRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPES
DILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV

[0834] BaseCallingMatlabScript

WTnuc='GCGGACATGGAGGACGTGCGCGGCCGCCTGGTGCAGTACCGCGGCGAGGTGCAGGCCATGCTCG
GCCAGA
GCACCGAGGAGCTGCGGGTGCGCCTCGCCTCCCACCTGCGCAAGCTGCGTAAGCGGCTCCTCCGCGATGC
CGATGAC
CTGCAGAAGCGCCTGGCAGTGTACCAGGCCGGGGCCCGCGAGGGCGCCGAGCGCGGCCTCAGCGCCATC
CGCGAGCG CCTGGGGCCCCTGGTGGAACAG'(SEQ ID NO: 164);

```
%cycle through fastq files for different samples  files=dir('*.fastq');
for d=1:20
filename=files(d).name;
%read fastq file
[header,seqs,qscore]=fastqread(filename);
seqsLength=length(seqs);                                    % number of sequences  seqsFile=
strrep(filename,'.fastq','');                               % trims off .fastq
%create a directory with the same name as fastq file  ifexist(seqsFile,'dir');
error('Directory already exists. Please rename or move it before  moving on.');
end
mkdir(seqsFile);                                            % make directory
wtLength=length(WTnuc);                                     % length of wildtype sequence
%% aligning back to the wildtype nucleotide sequence
%
  % AlN is a matrix of the nucleotide alignment  window=1:wtLength;
  sBLength=length(seqs);                                    % number of sequences
  % counts number of skips  nSkips = 0;
  ALN=repmat('',[sBLength wtLength]);
  % iterate through each sequencing read  for i = 1:sBLength
  %If you only have forward read fastq files leave as is
  %If you have R1 foward and R2 is reverse fastq files uncomment the
  %next four lines of code and the subsequent end statement
[0835]  %                 ifmod(d,2)==0;
  %                       reverse=seqrcomplement(seqs{i});
  %                       [score,alignment,start]=
  swalign(reverse,WTnuc,'Alphabet','NT');
  %                 else

  [score,alignment,start]=swalign(seqs{i},WTnuc,'Alphabet','NT');
  %                 end

  % length of the sequencing read  len=
  length(alignment(3,:));
  % if there is a gap in the alignment , skip = 1 and we will
  % throw away the entire read  skip = 0;
  for j = 1:len
  if (alignment(3,j) == '-' || alignment(1,j) == '-')  skip = 1;
                          break;
  end
  %in addition if the qscore for any given base in the read is
  %below 31 the nucleotide is turned into an N (fastq qscores that  are not letters)
  ifisletter(qscore{i}(start(1)+j-1)) else
  alignment(1,j)='N';
  end

  end
  if skip == 0 && len>10
  ALN(i,   start(2):(start(2)+length(alignment)-1))=alignment(1,:);

            end
end
```

```
% with the alignment matrices we can simply tally up the occurrences of
% each nucleotide at each column in the alignment these
% tallies ignore bases annotated as N
% due to low qscores
TallyNTD=zeros(5,wtLength); fori=1:wtLength
```

[0836]

```
TallyNTD(:,i)=[sum(ALN(:,i)=='A'),sum(ALN(:,i)=='C'),sum(ALN(:,i)=='G'),sum(A
LN(:,i)=='T'),sum(ALN(:,i)=='N')];
end
% we then save these tally matrices in the respective folder for
% further processing

save(strcat(seqsFile,'/TallyNTD'),'TallyNTD'); dlmwrite(strcat(seqsFile,'/TallyNTD.txt'),TallyNTD,'precision',
'%.3f','newline','pc'); end
```

[0837] 插入/缺失检测Matlab Script

```
WTnuc='GCGGACATGGAGGACGTGCGCGGCCGCCTGGTGCAGTACCGCGGCGAGGTGCAGGCCATGCTCGGC
CAGA
GCACCGAGGAGCTGCGGGTGCGCCTCGCCTCCCACCTGCGCAAGCTGCGTAAGCGGCTCCTCCGCGATGCCG
ATGAC
CTGCAGAAGCGCCTGGCAGTGTACCAGGCCGGGGCCCGCGAGGGCGCCGAGCGCGGCCTCAGCGCCATCCGC
GAGCG CCTGGGGGCCCCTGGTGGAACAG'(SEQ ID NO: 164);

%cycle through fastq files for different samples  files=dir('*.fastq');
%specify start and width of indel window as well as length of each flank  indelstart=154;
width=30; flank=10;

for d=1:3
filename=files(d).name;
%read fastq file
[header,seqs,qscore]=fastqread(filename);
seqsLength=length(seqs);                                    % number of sequences  seqsFile
=strcat(strrep(filename,'.fastq',''),'_INDELS');
%create a directory with the same name as fastq file+_INDELS  ifexist(seqsFile,'dir');
error('Directory already exists. Please rename or move it before  moving on.');
end
mkdir(seqsFile);                                            % make directory
wtLength=length(WTnuc);                                     % length of wildtype sequence  sBLength=
length(seqs);                                               % number of sequences

% initialize counters and cell arrays
nSkips = 0; notINDEL=0;
ins={};
dels={}; NumIns=0;
NumDels=0;
% iterate through each sequencing read  for i = 1:sBLength
%search for 10BP sequences that should flank both sides of the "INDEL  WINDOW"
windowstart=strfind(seqs{i},WTnuc(indelstart-flank:indelstart));

                    windowend=strfind(seqs{i},WTnuc(indelstart+width:indelstart+width+flank
));
%if the flanks are found proceed
```

```
iflength(windowstart)==1&&length(windowend)==1
%if the sequence length matches the INDEL window length save as
%not INDEL
if  windowend-windowstart==width+flank notINDEL=notINDEL+1;
%if the sequence is two or more bases longer than the INDEL
%window length save as an Insertion
elseif  windowend-windowstart>=width+flank+2 NumIns=NumIns+1;
ins{NumIns}=seqs{i};
%if the sequence is two or more bases shorter than the INDEL
%window length save as a Deletion
elseif  windowend-windowstart<=width+flank-2  NumDels=NumDels+1;
dels{NumDels}=seqs{i};
%keep track of skipped sequences that are either one base
%shorter or longer than the INDEL window width  else
nSkips=nSkips+1;
end
%keep track of skipped sequences that do not possess matching flank
%sequences else
nSkips=nSkips+1;
```

[0839]

```
            end
end

fid=fopen(strcat(seqsFile,'/summary.txt'),'wt');
fprintf(fid,'Skipped reads %i\n not INDEL %i\n Insertions %i\n Deletions
%i\n',[nSkips,notINDEL,NumIns,NumDels]);  fclose(fid);
save(strcat(seqsFile,'/nSkips'),'nSkips');  save(strcat(seqsFile,'/notINDEL'),'notINDEL');
save(strcat(seqsFile,'/NumIns'),'NumIns');  save(strcat(seqsFile,'/NumDels'),'NumDels');
save(strcat(seqsFile,'/dels'),'dels');
C = dels;
fid = fopen(strcat(seqsFile,'/dels.txt'),'wt');  fprintf(fid,'"%s"\n',C{:});
fclose(fid);
save(strcat(seqsFile,'/ins'),'ins');  C = ins;
fid = fopen(strcat(seqsFile,'/ins.txt'),'wt');  fprintf(fid,'"%s"\n',C{:});
fclose(fid);

end
```

[0840] 实施例5:Cas9变体序列

[0841] 本公开提供了Cas9变体,例如来自一种或多种生物体的Cas9蛋白,其可以包含一个或多个突变(例如,以产生dCas9或Cas9切口酶)。在一些实施方案中,可以突变Cas9蛋白的一个或多个氨基酸残基(下文通过星号标识)。在一些实施方案中,SEQ ID NO:6中提供的氨基酸序列的D10和/或H840残基,或任何Cas9,诸如如本文中提供的任一种Cas9氨基酸序列中的相应的突变得以突变。在一些实施方案中,SEQ ID NO:6中提供的氨基酸序列的D10残基,或任何Cas9,诸如如本文中提供的任一种Cas9氨基酸序列中的相应的突变得以突变为除D之外的任何氨基酸残基。在一些实施方案中,SEQ ID NO:6中提供的氨基酸序列的D10残基,或任何Cas9,诸如如本文中提供的任一种Cas9氨基酸序列中的相应的突变得以突变为A。在一些实施方案中,SEQ ID NO:6中提供的氨基酸序列的H840残基,或任何Cas9,诸如如本文中提供的任一种Cas9氨基酸序列中的相应的残基是H。在一些实施方案中,SEQ ID NO:6中提供的氨基酸序列的H840残基,或任何Cas9,诸如如本文中提供的任一种Cas9氨基酸序列中的相应的突变得以突变为除H之外的任何氨基酸残基。在一些实施方案中,SEQ ID NO:6中提供的氨基酸序列的H840残基,或任何Cas9,诸如如本文中提供的任一种Cas9氨基酸序列中的相应的突变得以突变为A。在一些实施方案中,SEQ ID NO:6中提供的氨基酸序

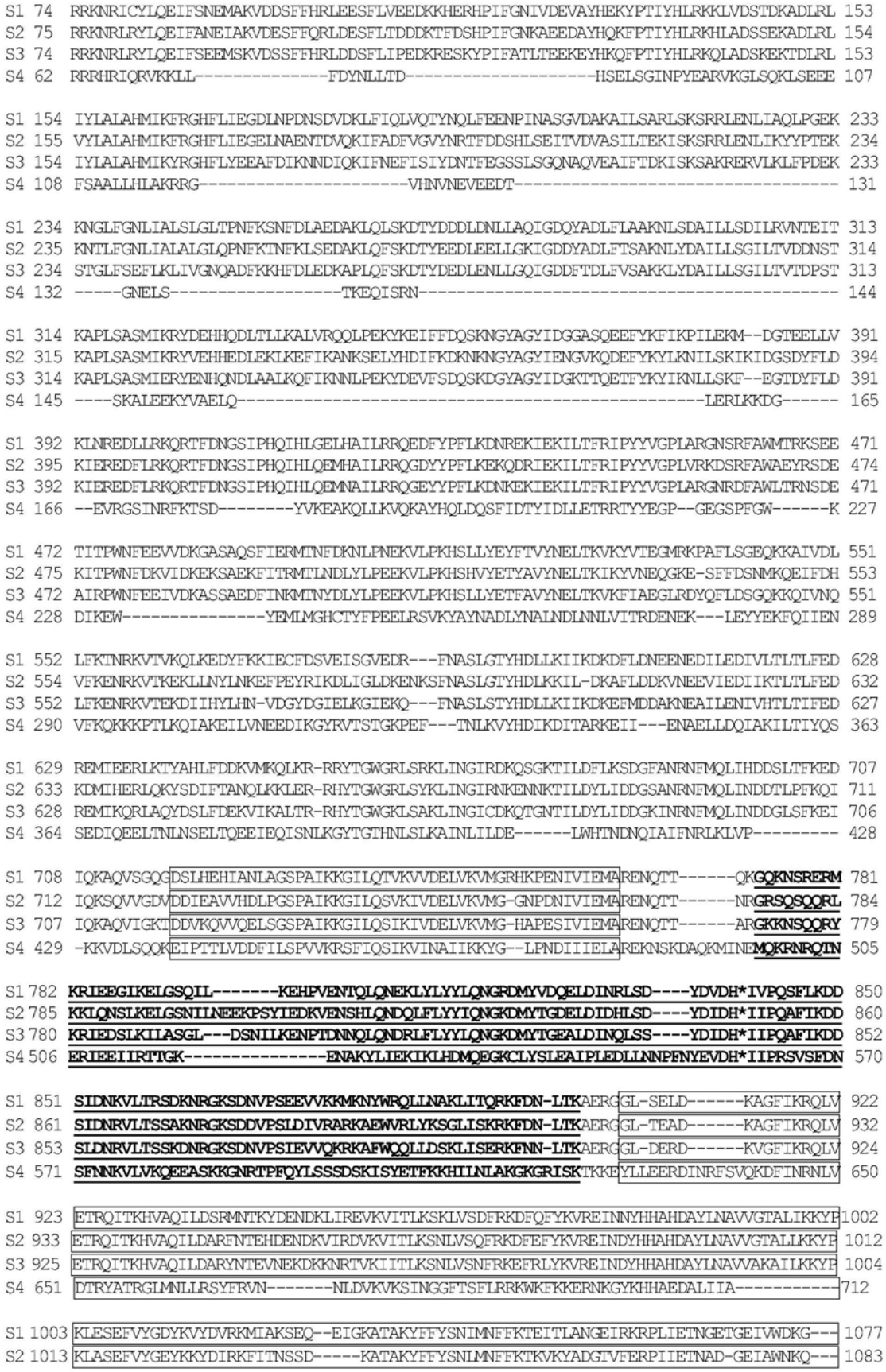
列的D10残基,或任何Cas9,诸如如本文中提供的任一种Cas9氨基酸序列中的相应的残基是D。

[0842] 对来自各种物种的Cas9序列进行比对以确定是否可以在其他Cas9蛋白中鉴定SEQ ID NO:6或SEQ ID NO:567的D10和H840的相应的同源氨基酸残基,从而允许具有同源氨基酸残基的相应的突变的Cas9变体产生。使用NCBI基于约束的多重比对工具(NCBI Constraint-based Multiple Alignment Tool)(COBALT(可在st-va.ncbi.nlm.nih.gov/tools/cobalt获得)在以下参数的情况下进行比对。比对参数:空位罚分-11,-1;末端空位罚分-5,-1。CDD参数:使用RPS BLAST开启(on);Blast E值0.003;查找保守列和重新计算开启。查询聚类参数:使用查询聚类开启;词大小4;最大聚类距离0.8;字母常规。

[0843] 以下提供了四个Cas9序列的示例性比对。比对中的Cas9序列是:序列1(S1)SEQ ID NO:567|WP_010922251|gi 499224711|II型CRISPR RNA引导的内切核酸酶Cas9[酿脓链球菌(Streptococcus pyogenes)];序列2(S2):SEQ ID NO:568|WP_039695303|gi 746743737|II型CRISPR RNA引导的内切核酸酶Cas9[解没食子酸链球菌(Streptococcus gallolyticus)];序列3(S3):SEQ ID NO:569|WP_045635197|gi 782887988|II型CRISPR RNA引导的内切核酸酶Cas9[缓症链球菌(Streptococcus mitis)];序列4(S4):SEQ ID NO:570|5AXW_A|gi 924443546|金黄色葡萄球菌(Staphylococcus aureus)Cas9。针对四个序列中的每一个鉴定HNH域(粗体和下划线)和RuvC域(加框)。S1中的氨基酸残基10和840以及比对序列中的同源氨基酸在相应的氨基酸残基后用星号标识。

[0844]

```
S1 1   --MDKK-YSIGLD*IGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLI--GALLFDSG--ETAEATRLKRTARRRYT 73
S2 1   --MTKKNYSIGLD*IGTNSVGWAVITDDYKVPAKKMKVLGNTDKKYIKKNLL--GALLFDSG--ETAEATRLKRTARRRYT 74
S3 1   --M-KKGYSIGLD*IGTNSVGFAVITDDYKVPSKKMKVLGNTDKRFIKKNLI--GALLFDEG--TTAEARRLKRTARRRYT 73
S4 1   GSHMKRNYILGLD*IGITSVGYGII--DYET----------------RDVIDAGVRLFKEANVENNEGRRSKRGARRLKR 61
```



[0845]

```
S3 1005 KLEPEFVYGEYQKYDLKRYISRSKDPKEVEKATEKYFFYSNLLNFFKEEVHYADGTIVKRENIEYSKDTGEIAWNKE--- 1081
S4  713 --NADFIFKEWKKLDKAKKVMENQM------------------------FEEKQAESMPEIETEQEYKEIFITPHQIK  764

S1 1078 -----RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKD---WDPKKYGGFDSPTVAYSVLVVAKV 1149
S2 1084 -----IDFEKVRKVLSYPQVNIVKKVETQTGGFSKESILPKGDSDKLIPRKTKKVYWDTKKYGGFDSPTVAYSVFVVADV 1158
S3 1082 -----KDFAIIKKVLSLPQVNIVKKREVQTGGFSKESILPKGNSDKLIPRKTKDILLDTTKYGGFDSPVIAYSILLIADI 1156
S4  765 HIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKL----KKLIN-KSP----EKLIMYHH  835

S1 1150 EKGKSKKLKSVKELLGITIMERSSFEKNPI-DFLEAKG-----YKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKG 1223
S2 1159 EKGKAKKLKTVKELVGISIMERSFFEENPV-EFLENKG-----YHNIREDKLIKLPKYSLFEFEGGRRRLIASASELQKG 1232
S3 1157 EKGKAKKLKTVKTLVGITIMEKAAFEENPI-TFLENKG-----YHNVRKENILCLPKYSLFELENGRRRLIASAKELQKG 1230
S4  836 DPQTYQKLK--------LIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKV  907

S1 1224 NELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH------ 1297
S2 1233 NEMVLPGYLVELLYHAHRADNF-----NSTEYLNYVSEHKKEFEKVLSCVEDFANLYVDVEKNLSKIRAVADSM------ 1301
S3 1231 NEIVLPVYLTTLLYHSKNVHKL-----DEPGHLEYIQKHRNEFKDLINLVSEFSQKYVLADANLEKIKSLYADN------ 1299
S4  908 VKLSLKPYRFD-VYLDNGVYKFV-----TVKNLDVIK--KENYYEVNSKAYEEAKKLKKISNQAEFIASFYNNDLIKING  979

S1 1298 RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSIT--------GLYETRI----DLSQL 1365
S2 1302 DNFSIEEISNSFINLLTLTALGAPADFNFLGEKIPRKRYTSTKECLNATLIHQSIT--------GLYETRI----DLSKL 1369
S3 1300 EQADIEILANSFINLLTFTALGAPAAFKFFGKDIDRKRYTTVSEILNATLIHQSIT--------GLYETWI----DLSKL 1367
S4  980 ELYRVIGVNNDLLNRIEVNMIDITYR-EYLENMNDKRPPRIIKTIASKT---QSIKKYSTDILGNLYEVKSKKHPQIIKK 1055

S1 1366 GGD 1368
S2 1370 GEE 1372
S3 1368 GED 1370
S4 1056 G-- 1056
```

[0846]

[0847] 比对证明,可以通过使用本领域已知的比对程序和算法鉴定与参照序列或参照残基比对的氨基酸序列或残基来在Cas9序列变体间鉴定与参照Cas9氨基酸序列或氨基酸残基同源的氨基酸序列和氨基酸残基,所述Cas9序列变体包括但不限于来自不同物种的Cas9序列。本公开提供了Cas9变体,其中如本文所述突变通过SEQ ID NO:567-570中的星号标识(例如,分别为S1、S2、S3和S4)的一个或多个氨基酸残基。SEQ ID NO:6的Cas9中的残基D10和H840(其对应于SEQ ID NO:567-570中通过星号标识的残基)在本文中称为“同源的”或“相应的”残基。此类同源的残基可以通过序列比对来鉴定,例如,如上所述,并通过鉴定与参照序列或残基比对的序列或残基来鉴定。类似地,对应于本文SEQ ID NO:6中鉴定的突变的Cas9序列中的突变,例如SEQ ID NO:6中的残基10和840的突变,在本文中称为“同源的”或“相应的”突变。例如,对应于SEQ ID NO:6或上述四个比对序列的S1(SEQ ID NO:567)中的D10A突变的突变是S2的D11A、S3的D10A和S4的D13A;SEQ ID NO:6或S1(SEQ ID NO:567)中H840A的相应的突变是S2的H850A、S3的H842A和S4的H560A。

[0848] 实施例6:下一代C至T编辑器

[0849] 检查了胞苷脱氨酶的其他家族作为碱基编辑器3(BE3)构建体的替代。开发了不同的C至T编辑器,以具有窄的或不同的编辑窗,交替序列特异性来扩展可靶向的底物并且具有更高的活性。

[0850] 使用实施例4中描述的方法,评估了HeK-3位点处的pmCDA1(来自海七鳃鳗的胞苷脱氨酶1)活性(图42)。pmCDA1-nCas9-UGI-NLS(nCas9指示本文所述的Cas9切口酶)构建体在rAPOBEC1(BE3)不能接近的某些位点(例如,互补链上位置9、5、4和3处的C碱基)有活性。

[0851] 在图43中给出了在HeK-2位点处的pmCDA1活性。pmCDA1-XTEN-nCas9-UGI-NLS构建体在与“G”相邻的位点处是有活性的,而rAPOBEC1类似物(BE3构建体)在与“G”相邻的“C”,例如互补链上位置11处的C碱基上具有低活性。

[0852] 对CDA和APOBEC1(BE3构建体)显示了具有转化为T的靶C(图44),转化为A的C(图

45)和转化为G的C(图46)的总测序读段的百分比。

[0853] 图47中显示了HeK-2位点处的huAPOBEC3G活性。使用了两种构建体:huAPOBEC3G-XTEN-nCas9-UGI-NLS和huAPOBEC3G*(D316R_D317R)-XTEN-nCas9-UGI-NLS。huAPOBEC3G-XTEN-nCas9-UGI-NLS构建体与rAPOBEC1(BE3)具有不同的序列特异性,如图47所示,编辑窗表现为较窄,如APOBEC3G与APOBEC1相比在位置4处的活性降低指示。huAPOBEC3G中发生的突变(D316R和D317R)增加了ssDNA的结合,并导致对扩展编辑的位点的可观察到的效果(比较图47中的APOBEC3G和APOBEC3G_RR)。根据APOBEC3G晶体结构选择突变,参见Holden et al.,Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implication.Nature.(2008);121-4,其全部内容通过引用并入本文。

[0854] 实施例7:pmCDA1/huAPOBEC3G/rAPOBEC1在大肠杆菌中起作用

[0855] 使用具有F质粒上编码的lacZ的细菌菌株进行A至I转化的LacZ选择优化。将关键的谷氨酸残基突变(例如,GAG至GGG,Glu至Gly突变),使得胞苷脱氨酶将G变为A将恢复lacZ活性(图48)。选择菌株CC102用于选择测定法。APOBEC1和CDA构建体用于选择测定法,以优化G至A的转化。

[0856] 为了评估编码脱氨酶构建体的质粒的拷贝数对lacZ回复频率的影响,将CDA和APOBEC1脱氨酶克隆到4个具有不同复制起点(因此具有不同拷贝数)的质粒,SC101,CloDF3,RSF1030,和PUC(拷贝数:PUC>RSF1030>CloDF3>SC101)并置于诱导型启动子下。个别将质粒转化到带有含有突变的LacZ基因的F质粒的大肠杆菌细胞中。诱导脱氨基酶的表达,并检测每种构建体的LacZ活性(图49)。如图49所示,在所有情况下,CDA表现出比APOBEC1显著更高的活性,而与克隆脱氨基酶的质粒拷贝数无关。此外,就拷贝数而言,脱氨酶活性与克隆它们的质粒的拷贝数呈正相关,即PUC>CloDF3>SC101。

[0857] 通过对lacZ基因座处的基因组DNA进行测序来确认LacZ回复。为了获得包含校正的LacZ基因的基因组DNA,在含有X-gal的培养基上培养细胞,其中具有LacZ活性的细胞形成蓝色菌落。选择蓝色菌落并在含乳糖的基本培养基中培养。将离心向下旋转,洗涤,并重新铺在基本培养基板(乳糖)上。然后选择最高稀释度的蓝色菌落,并在lacZ基因座处对其基因组DNA进行测序(图50)。

[0858] 设计氯霉素回复测定法以测试不同胞苷脱氨酶(例如,CDA和APOBEC1)的活性。以RSF1030为复制起点,构建携带突变型CAT1基因的质粒,该基因对细菌赋予氯霉素抗性。突变体CAT1基因编码具有H195R(CAC到CGC)突变的CAT1蛋白,从而使该蛋白无活性(图51)。CGC密码子中与G碱基碱基配对的C的脱氨基将使该密码子转变回CAC密码子,从而恢复蛋白质的活性。如图52所示,在恢复氯霉素抗性基因的活性方面,CDA在大肠杆菌中的表现优于rAPOBEC。使用选择质粒(pNMG_ch_5)在S1030中的氯霉素最小抑制浓度(MIC)为约1μg/mL。rAPOBEC-XTEN-dCas9-UGI和CDA-XTEN-dCas9-UGI两者诱导选择质粒上的DNA校正(图53)。

[0859] 接下来,在相同测定法中测试了huAPOBEC3G-XTEN-dCas9-UGI蛋白。令人感兴趣的是,huAPOBEC3G-XTEN-dCas9-UGI与rAPOBEC1-XTEN-dCas9-UGI融合蛋白表现出不同的序列特异性。与rAPOBEC11-XTEN-dCas9-UGI融合物(其中位置3、6和8被编辑)相比,仅位置8用APOBEC3G-XTEN-dCas9-UGI融合物编辑(图54)。

[0860] 实施例8:具有较小脱靶编辑的C至T碱基编辑器

[0861] 当前的碱基编辑技术允许基因组DNA中C:G碱基对向T:A碱基对的序列特异性转

化。这是通过胞苷脱氨酶将胞嘧啶直接催化转化为尿嘧啶来完成的，因此，与传统的基因组编辑技术不同，不将双链DNA断裂(DSB)引入DNA中作为第一步。见Komor,A.C.,Kim,Y.B.,Packer,M.S.,Zuris,J.A.,and Liu,D.R.(2016),"Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage."Nature 533,420-424;其全部内容通过引用并入本文。取而代之的是，将催化死亡的SpCas9(dCas9)或SpCas9切口酶(dCas9(A840H))栓系到胞苷脱氨酶，例如rAPOBEC1，pmCDA1或hAPOBEC3G。感兴趣的基因组基因座由sgRNA编码，并且融合物的dCas9部分促进DNA结合和局部变性。然而，就像wt dCas9和wt Cas9表现出脱靶DNA结合和切割一样，当前的碱基编辑器在Cas9脱靶基因座上也表现出C至T编辑，这限制了它们的治疗用途。

[0862] 据报告，在SpCas9中仅引入3至4个突变增加了SpCas9的DNA结合特异性，该突变中和蛋白质与其靶DNA的糖-磷酸主链之间的非特异性静电相互作用。参见Kleinstiver,B.P.,Pattanayak,V.,Prew,M.S.,Tsai,S.Q.,Nguyen,N.T.,Zheng,Z.,and Joung,J.K.(2016)"High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects."Nature 529,490-495;和Slaymaker,I.M.,Gao,L.,Zetsche,B.,Scott,D.A.,Yan,W.X.,and Zhang,F.(2015)"Rationally engineered Cas9 nucleases with improved specificity.Science351,84-88;每篇的全部内容通过引用并入本文。因此，将四个报告的中和突变引入最初报告的碱基编辑器BE3(SEQ ID NO:48)中，并且发现该酶的脱靶C至T编辑也大大减少了(图55)，中靶编辑没有降低(图56)。

[0863] 如图55所示，使用Lipofectamine 2000用表达BE3或HF-BE3和与EMX1序列匹配的sgRNA的质粒转染HEK293T细胞，转染后三天，提取基因组DNA，通过PCR扩增，并通过在如先前由Joung和同事使用GUIDE-seq方法确定的EMX1 sgRNA的中靶基因座，加上前10个已知的Cas9脱靶基因座处的高通量DNA测序分析。参见Tsai,S.Q.,Zheng,Z.,Nguyen,N.T.,Liebers,M.,Topkar,V.V.,Thapar,V.,Wyvekens,N.,Khayter,C.,Iafrate,A.J.,Le,L.P.,et al.(2015)"GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases."Nat Biotech 33,187-197;其全部内容通过引用并入本文。EMX1脱靶5位点未扩增并且未显示。显示了中靶和脱靶的原间隔物和原间隔物邻近基序(PAM)的序列(图55)。对BE3和HF-BE3显示了细胞C至T转化百分比，定义为在原间隔物内原始C的每个位置上有T的总DNA测序读段的百分比。

[0864] 在图56中，使用Lipofectamine 2000，用表达BE3或HF-BE3和与指示的基因组基因座匹配的sgRNA的质粒转染HEK293T细胞。转染后三天，提取基因组DNA，通过PCR扩增，并通过中靶基因座处的高通量DNA测序分析。对于用BE3或HF-BE3的处理显示了每个原间隔物内靶C处具有所有四个碱基的总DNA测序读段的百分比(图56)。还显示了插入/缺失形成的频率。

[0865] HF-BE3的一级蛋白序列(SEQ ID NO:48)：

[0866] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYP

TIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKA
ILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYID
GGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIE
KILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTAFDKNLPNEKVLPKHSLLYEYF
TVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH
DLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGALSRKLINGIRD
KQSGKTILDFLKSDGFANRNFMALIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVK
VMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKA
ERGGLSELDKAGFIKRQLVETRAITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI
RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF
DSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRK
RMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLD
KVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL
GGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALV
IQDSNGENKIKMLSGGSPKKKRKV

[0867] 实施例9:开发使用Cas9变体的碱基编辑器以及调节碱基编辑器持续合成能力以增加碱基编辑技术的靶物范围和精确性

[0868] 与传统的基因组编辑平台不同,碱基编辑技术允许DNA中精确的单核苷酸变化而不引起双链断裂(DSB)。参见Komor,A.C.et al.Nature533,420-424(2016)。当前一代的碱基编辑器仅使用NGG PAM。这限制了它在基因组内编辑期望碱基的能力,因为碱基编辑器需要放置在靶碱基位于4碱基区域(“脱氨基窗”)内的精确位置,PAM上游约15个碱基。参见Komor,A.C.etal.Nature533,420-424(2016)。此外,由于胞苷脱氨酶的高持续合成能力,碱基编辑器可以将其脱氨基窗内的所有胞苷转化为胸苷,这可以诱导除研究人员所需的氨基酸变化以外的氨基酸变化。参见Komor,A.C.et al.Nature533,420-424(2016)。

[0869] 通过开发具有Cas9变体的碱基编辑器来扩展碱基编辑的范围

[0870] 将具有不同PAM特异性的Cas9同源物和其他RNA指导的DNA结合剂掺入碱基编辑器结构中。参见Kleinstiver,B.P.et al.Nature523,481-485(2015);Kleinstiver,B.P.et al.Nature Biotechnology33,1293-1298(2015);and Zetsche,B.et al.Cell163,759-771(2015);每篇的全部内容通过引用并入本文。此外,还并入加宽各种Cas9蛋白的PAM特异性的创新,以进一步扩展碱基编辑器的目标范围。参见Kleinstiver,B.P.et al.Nature523,481-485(2015);and Kleinstiver,B.P.et al.Nature Biotechnology33,1293-1298(2015)。表4总结了当前的碱基编辑器选项。

[0871] 表4.从Cas9变体制备的新碱基编辑器

[0872]

物种	PAM	碱基编辑器名称	Cas9变体的参考文献
酿脓链球菌	...NGG	BE3	野生型
	...NGA	VQR BE3或EQR BE3	Kleinstiver, B. P. *et al.*
	...NGCG	VRER BE3	Kleinstiver, B. P. *et al.*
金黄色葡萄球菌	...NNGRRT	SaBE3	野生型
	...NNNRRT	SaKKH BE3	Kleinstiver, B. P. *et al.*
L. bacterium	TTTN...	dCpf1 BE2	Zetsche, B. *et al.*

[0873] 通过rAPOBEC1的定点诱变来调节碱基编辑器的持续合成能力

[0874] 推断可以通过在脱氨酶中进行点突变来调节碱基编辑器的持续合成能力。寻求突变的掺入，所述突变略微降低脱氨酶的催化活性，其中碱基编辑器仍可平均催化一轮胞苷脱氨基，但是不太可能在相关的时间范围内接近并催化另一个脱氨基。实际上，所得的碱基编辑器将具有较窄的脱氨基窗。

[0875] 在这项工作中探查到的rAPOBEC1突变列于表5中。一些突变导致rAPOBEC1催化的轻微的表观损害，这表现出当在脱氨基窗内发现多个胞苷时，一个胞苷比另一个胞苷优先编辑。组合这些突变中的一些具有叠加效果，从而允许碱基编辑器以更高的严格性来区分底物胞苷。一些双重突变体和三重突变体允许选择性编辑彼此紧邻的多个胞苷间的一个胞苷(图57)。

[0876] 表5.研究的rAPOBEC1点突变

[0877]

本工作中研究的rAPOBEC1突变	APOBEC3G中的相应突变	参考文献
H121R / H122R	D315R / D316R	Holden, L. G. *et al.*
R126A	R320A	Chen, K-M. *et al.*
R126E	R320E	Chen, K-M. *et al.*
R118A	R313A	Chen, K-M. *et al.*
W90A	W285A	Chen, K-M. *et al.*
W90Y	W285Y	此工作
R132E	R326E	此工作

[0878] 碱基编辑器PAM扩展和持续合成能力调节

[0879] 下一代碱基编辑器设计为通过使用其他RNA引导的DNA结合剂来扩展基因组中的可编辑胞苷(图58)。使用NGG PAM仅允许在“窗”内的单一靶物，而使用多个不同的PAM允许将Cas9放置在任何位置以实现选择性脱氨基。已经从Cas9变体中创建了许多新的碱基编辑器(图59和表4)。探索了不同的PAM位点(NGA，图60；NGCG，图61；NNGRRT，图62；NNHRRT，图63)。通过动态调节胞苷脱氨酶点诱变成功实现了选择性脱氨基(图65和表5)。

[0880] 然后，在细胞培养物中使用具有多个胞苷的间隔物研究了各种突变对脱氨基窗的影响(图66和67)。

[0881] 此外，检查了各种突变对具有有限数目的胞苷的不同基因组位点的影响(图68至71)。发现在间隔物中的脱氨基窗内将编辑约1个胞苷，而胞苷的剩余部分保持完整。总的来说，编辑的偏爱如下：$C_6>C_5>>C_7\approx C_4$。

[0882] 使用Cpf1的碱基编辑

[0883] Cpf1(Cas9同源物)可以以AsCpf1，LbCpf1或从任何其他物种获得。图73中显示了融合构建体的示意图，包括BE2和BE3等同物。BE2等同物使用无催化活性的Cpf2酶(dCpf1)代替Cas9，而BE3等同物包括Cpf1突变体，该突变体对靶链产生切口。底部示意图描绘了不同的融合物结构，以组合其上方所示的两种创新(图73)。使用不同的间隔物检查了使用Cpf1 BE2的HEK293T细胞TTTN PAM位点的碱基编辑结果(图64A至64C)。在一些实施方案中，在本文提供的任何碱基编辑器中，可以使用Cpf1代替Cas9域。在一些实施方案中，Cpf1是与SEQ ID NO 9至少50％，55％，60％，65％，70％，75％，80％，85％，90％，95％，98％，99％或99.5％相同的蛋白质。

[0884] Cpf1的完整蛋白质序列(SEQ ID NO:9)：

[0885] MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSSV
CISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESDLILWLKQSK
DNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKD
KAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGI
NEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLL
FDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLE
TIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLD
QTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANG
WDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPS
EDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGY
KLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPK
KITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSI
DRGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIA
KLVIEYNAIVVFEDLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQ
TGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASF
GSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGT
ELDYLISPVADVNGNFFDSRQAPKNMPQDADANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0886] 实施例10:碱基编辑的增加的保真性

[0887] 检查BE3和HF-BE3的质粒递送之间的差异，发现两者以相当的效率编辑中靶基因座(图74和75)。然而，HF-BE3编辑的脱靶基因座远小于BE3，这意味着HF-BE3具有比BE3高得多的DNA特异性(图76)。对HEK细胞的脱氨酶蛋白脂质体转染证明，BE3的蛋白质递送比BE3的质粒DNA递送产生相当的中靶活性，但好得多的特异性。使用改良的转染程序和更好的质粒(n＝2)，该实验使用以下条件：蛋白质递送为125nM Cas9:sgRNA复合物，质粒递送为750ng BE3/HF-BE3质粒+250ng sgRNA质粒，并且脂质转染用每孔1.5μL Lipofectamine 2000。对于任何递送方法，EMX-1脱靶位点2和FANCF脱靶位点1与所有测定的脱靶相比显示用BE3的最多脱靶编辑(图77和78)，而HEK-3未显示脱靶处的显著编辑(图79)。HEK-4在中靶

位点处显示了一些C至G编辑，而其脱靶位点1、3和4显示了所有测定位点的最多脱靶编辑(图80)。

[0888] 通过微注射向斑马鱼递送BE3蛋白

[0889] 在体外测定中测试了TYR引导RNA的sgRNA活性(图81和82)。%HTS读段显示在与纯化的BE3蛋白温育2小时和所得产物的PCR期间多少C残基转化为T残基。实验使用80聚体合成DNA底物，在其基因组背景的60bp中具有靶脱氨基位点。这与%编辑的DNA链不同，因为仅对一条链产生切口，因此该产物不通过PCR扩增。编辑的HTS读段的比例等于x/(2-x)，其中x是编辑的THS读段的实际比例。对于60%的编辑，编辑的碱基的实际比例为75%。“脱靶”表示在与脱靶sgRNA结合的情况下与相同的DNA底物一起温育的BE3。发现sgRNAs sgRH_13，sgHR_17和可能的sgHR_16似乎是体内注射实验的有希望的靶物。

[0890] 在体内在斑马鱼中测试了BE3蛋白的递送。给斑马鱼胚胎(n＝16-24)注入杂乱的sgRNA，sgHR_13，sgHR_16或sgHR_17和纯化BE3。独立地(单一胚胎)并针对每种条件分析来自每种条件的三个胚胎，合并所有注射的胚胎，并作为合并物进行测序。在图83至85中显示了结果。

[0891] 实施例11:使用碱基编辑器治疗疾病

[0892] 本文提供的碱基编辑器或复合物(例如，BE3)可用于修饰核酸。例如，可以使用碱基编辑器将核酸(例如DNA)中的胞嘧啶变为胸腺嘧啶。可以进行此类变化，以尤其改变蛋白质的氨基酸序列，以破坏或创建起始密码子，以创建终止密码子，以破坏剪接供体，以破坏剪接受体或编辑调节序列。在图86中显示了可能的核苷酸变化的例子。

[0893] 本文提供的碱基编辑器或复合物(例如，BE3)可用于编辑受试者中载脂蛋白E的同种型。例如，可以编辑载脂蛋白E同种型，以产生与形成阿尔茨海默氏病的较低风险相关的同种型。载脂蛋白E具有四种分别在氨基酸112和158处不同的同种型。APOE4是晚期阿尔茨海默氏病的最大且最常见的遗传风险因素。可以通过使用碱基编辑器(例如BE3)以将CGC核酸序列改变为TGC(其编码在残基158处的半胱氨酸)将由核酸序列CGC编码的APOE4的精氨酸残基158改变为半胱氨酸。此变化产生APOE3r同种型，其与较低的阿尔茨海默氏病风险相关。参见图87。

[0894] 测试了是否可以使用碱基编辑器BE3将小鼠星形胶质细胞中的APOE4编辑为APOE3r(图88)。用Cas9+模板或BE3(靶向编码APOE4精氨酸158的核酸)核转染APOE 4小鼠星形胶质细胞。Cas9+模板在26%插入/缺失的情况下仅产生0.3%的编辑，而BE3在5%插入/缺失的情况下产生75%编辑。另外两个碱基编辑的胞嘧啶是沉默的并且不对氨基酸序列产生变化(图88)。

[0895] 本文提供的碱基编辑器或复合物可用于例如通过将突变引入PRNP基因来治疗蛋白质疾病，例如克罗伊茨费尔特-雅各布病和家族致命性失眠症。恢复PRNP突变可以不产生治疗效果，并且PRNP中的插入/缺失可以是致病的。因此，测试了是否可以使用碱基编辑器(例如BE3)使PRNP突变以在PRNP基因中引入过早的终止密码子。将与其引导RNA缔合的BE3导入HEK细胞或胶质母细胞瘤细胞中，并能够编辑PRNP基因，将残基37处的编码精氨酸改变为终止密码子。BE3产生41%的编辑(图89)。

[0896] 可以编辑的其他基因包括以下者:APOE编辑Arg 112和Arg 158，以治疗增加的阿尔茨海默氏病的风险;APP编辑Ala 673以降低阿尔茨海默氏病的风险;PRNP编辑Arg 37以

家族致命性失眠症和其他朊病毒蛋白病;DMD编辑外显子23和51的剪接位点,以治疗迪谢内肌营养不良;FTO编辑内含子1以治疗肥胖风险;PDS编辑外显子8以治疗彭德莱综合征(遗传性耳聋);TMC1编辑外显子8以治疗先天性听力损失;CYBB编辑各种与患者相关的突变,以治疗慢性肉芽肿病。下表6显示了可以使用本文提供的碱基编辑器治疗的其他疾病。

[0897] UGI也起着关键作用。显示敲除UDG(UGI抑制的)显著改善C至T碱基编辑的清洁度和效率(图90)。此外,显示具有切口酶且不具有UGI的碱基编辑器产生结果混合,具有非常高的插入/缺失率(图91)。

[0898] 实施例12:扩展碱基编辑的靶向范围

[0899] 碱基编辑是基因组编辑的新方法,其使用含有催化缺陷的酿脓链球菌Cas9,胞苷脱氨酶和碱基切除修复抑制剂的融合蛋白来诱导DNA中可编程的单核苷酸C→T(或G→A)的变化而不产生双链DNA断裂,不需要供体DNA模板,且不诱导过多的随机插入和缺失[1]。本文中描述了五种新的C→T(或G→A)碱基编辑器的开发,所述编辑器使用具有不同原间隔物邻近基序(PAM)特异性的天然的和工程化的Cas9变体,以将可以通过碱基编辑靶向的位点数目扩展2.5倍。另外,还工程化改造新的碱基编辑器,该碱基编辑器含有突变的胞苷脱氨酶域,其将表观编辑窗的宽度从约5个核苷酸缩窄为1个或2个核苷酸,从而能够区分以前会以相当效率编辑的邻近C核苷酸。这些开发共同大大增加了碱基编辑的靶向范围。

[0900] CRISPR-Cas9核酸酶已被广泛用于介导靶向的基因组编辑[2]。在大多数基因组编辑应用中,Cas9与单引导RNA(sgRNA)形成复合物,并在sgRNA序列指定的靶位点处诱导双链DNA断裂(DSB)。细胞主要通过非同源末端连接(NHEJ)修复途径对此DSB响应,所述非同源末端连接(NHEJ)修复途径导致随机插入或缺失(插入/缺失),其可以导致破坏基因的移码突变。在存在与DSB侧翼序列具有高度同源性的供体DNA模板的情况下,可以通过称为同源性指导修复(HDR)的备选途径来实现基因校正[3,4]。不幸的是,在大多数非干扰性条件下HDR效率低下,取决于细胞状态和细胞类型,并以较高的插入/缺失频率为主导[3,4]。由于与人疾病相关的大多数已知遗传变异都是点突变[5],因此需要可以更有效且更干净地产生精确点突变的方法。

[0901] 最近已经描述了碱基编辑,该碱基编辑使得能够以可编程的方式用T:A碱基对靶向性替换C:G碱基对而不诱导DSB[1]。碱基编辑使用酿脓链球菌Cas9的催化失活(dCas9)或切口酶形式(SpCas9),胞苷脱氨酶(如APOBEC1)和碱基切除修复抑制剂(如尿嘧啶糖基化酶抑制剂(UGI))之间的融合蛋白,以在sgRNA规定的5个核苷酸窗内将胞苷转化为尿苷[1]。第三代碱基编辑器BE3在多种细胞系中与使用其他基因组编辑方法可以实现的情况相比以更高的效率和更低的插入/缺失频率将C:G碱基对转换为T:A碱基对,包括与疾病相关的点突变[1]。随后的研究已经在多种背景中验证了脱氨酶-dCas9融合方法[6,7]。

[0902] BE3的有效编辑需要NGG PAM的存在,该NGG PAM将靶C置于原间隔物的PAM远端末端附近的5个核苷酸的窗(位置4-8,将PAM计为位置21-23)内[1]。该PAM要求实质上限制了BE3可以有效靶向的人基因组中的位点数目,因为许多感兴趣位点缺少靶C下游13至17个核苷酸的NGG。此外,BE3的高活性和持续合成能力导致将编辑窗内的所有C转换为T,这可以潜在对靶基因座引起不想要的改变。在此,描述了解决这两个限制的新的C:G至T:A碱基编辑器。

[0903] 认为任何结合DNA并形成含有单链DNA气泡的“R环”复合物[8]的Cas9同源物原则上都可以转化为碱基编辑器。这些新的碱基编辑器将通过允许非NGG PAM位点进行编辑而扩

展可靶向基因座的数目。金黄色葡萄球菌的Cas9同源物(SaCas9)比SpCas9小得多(1053对1368个残基),可以介导哺乳动物细胞中的有效基因组编辑,并且需要NNGRRT PAM[9]。在BE3中将SpCas9替换为SaCas9,以生成SaBE3,并用编码SaBE3和靶向6个人基因组基因座的sgRNA的质粒转染HEK293T细胞(图92A和92B)。3d后,对基因组基因座进行高通量DNA测序(HTS),以定量碱基编辑效率。SaBE3实现人细胞中多种基因组位点处的靶C的C至T碱基编辑,以源自靶向位置6-11处C的很高的转化效率(从C转化为T的总DNA序列的约50-75%,而不富集转染细胞)。SaBE3在含有NNGRRT的靶位点上的效率通常超过BE3在含有NGG的靶位点上的效率[1]。也许由于其较高的平均效率,SaBE3还可导致在规范BE3活性窗之外的位置处的靶C处进行可检测的碱基编辑(图92C)。相比之下,根据已知的SpCas9 PAM偏爱,BE3在相同条件下(0-11%)显示出显著减少的编辑(图106A)[10]。这些数据显示了SaBE3可以促进无法接近BE3的位点处的非常有效的碱基编辑。

[0904] 通过应用扩展或改变PAM特异性的最近工程化Cas9变体进一步扩展碱基编辑器的靶向范围。Joung及同事最近报告了三个接受NGA(VQR-Cas9),NGAG(EQR-Cas9)或NGCG(VRER-Cas9)PAM序列的SpCas9突变体[11]。此外,Joung及同事工程化改造SaCas9变体,其含有将其PAM要求放宽至NNNRRT的三个突变(SaKKH-Cas9)[12]。将BE3的SpCas9部分替换为这四个Cas9变体,以产生分别靶向NNNRRT,NGA,NGAG和NGCG PAM的VQR-BE3,EQR-BE3,VRER-BE3和SaKKH-BE3。对于每个新碱基编辑器,用编码这些构建体和靶向六个基因组基因座的sgRNA的质粒转染HEK293T细胞,并使用HTS测量C至T碱基的转化。

[0905] SaKKH-BE3以经处理的非富集细胞的多达62%的效率编辑具有NNNRRT PAM的位点(图92D)。如预期,SaBE3无法有效地编辑含有NNNHRRT(其中H=A,C或T)的PAM的靶物(图92D)。VQR-BE3,EQR-BE3和VRER-BE3以与BE3的情况相似的编辑窗在具有预期PAM要求的基因组基因座处表现出经处理的非富集细胞的多达50%的较为适中但仍实质性的碱基编辑效率(图92E和92F)。通常,VQR-BE3,EQR-BE3和VRER-BE3的碱基编辑效率与相应的Cas9核酸酶的报告的PAM要求非常接近;例如,EQR-BE3无法有效地编辑包含NGAH PAM序列的靶物(图92F)。相比之下,BE3可能由于其PAM限制而无法有效地编辑具有NGA或NGCG PAM的位点(0-3%)(图106B)。

[0906] SaBE3,SaKKH-BE3,VQR-BE3,EQR-BE3和VRER-BE3的特性共同建立碱基编辑器表现出模块性,该模块性促进其利用Cas9同源物和工程化变体的能力。

[0907] 接下来,开发了具有改变的活性窗宽度的碱基编辑器。BE3活性窗内的所有C都可以有效地转换为T[1]。在仅编辑BE3活性窗中存在的C子集重要的情况下,调节此窗宽度的功能将非常有用。

[0908] 先前观察到APOBEC1和dCas9之间的接头长度调节了APOBEC1在体外可接近的碱基数。然而,在HEK293T细胞中,改变接头长度并不能显著调节编辑窗的宽度,这提示在复杂的细胞环境中,dCas9和胞苷脱氨酶的相对方向和柔性不受接头长度的强烈决定(图96)。接下来,认为将sgRNA的5'端截短可通过减少RNA-DNA异双链体形成时脱氨酶可接近的单链DNA的长度来缩窄碱基编辑窗。用编码BE3和不同间隔物长度的sgRNA的质粒共转染HEK293T细胞,所述sgRNA靶向编辑窗内具有多个C的基因座。当使用有17至19个碱基的间隔物的截短sgRNA时,没有观察到碱基编辑宽度的一致变化(图95A至95C)。将sgRNA间隔物截短至小于17个碱基导致活性大量损失(图95A)。

[0909] 作为替代方法，认为对脱氨酶域的突变可通过多种可能的机制使编辑窗的宽度变窄。首先，某些突变可以以降低对脱氨酶活性位点的非最佳C呈现的耐受性的方式，改变底物结合，结合的DNA的构象或底物对活性位点的可及性。其次，由于APOBEC1的高活性可能促成每个DNA结合事件的多个C的脱氨基[1,13,14]，降低碱基编辑器的脱氨酶域的催化效率的突变可以在与DNA解离前阻止它催化连续几轮脱氨基。一旦发生任何C:G至T:A的编辑事件，sgRNA便不再与靶DNA序列完全匹配，并且碱基编辑器与靶基因座的重新结合将不太有利。对这两种策略进行了测试，以发现区分原始编辑窗中的多个胞苷的新的碱基编辑器。

[0910] 鉴于缺乏可用的APOBEC1结构，鉴定出几个突变，先前报告了所述突变调节APOBEC3G的催化活性，所述APOBEC3G是来自与APOBEC1共享含有其活性位点的域的42％序列相似性的同一家族的胞苷脱氨酶[15]。将对应的APOBEC1突变掺入BE3中，并评估它们对HEK293T细胞中在位置3、4、5、6、8、9、10、12、13和14处含有C(位点A)；或在位置5、6、7、8、9、10、11和13处含有C(位点B)的两个富含C的基因组位点处的碱基编辑效率和编辑窗宽度的影响。

[0911] APOBEC1突变R118A和W90A各自导致碱基编辑效率的显著损失(图97C)。R132E导致编辑效率普遍下降，但没有改变(实质性缩短)编辑窗的形状(图97C)。相反，发现了几种突变，它们使编辑窗的宽度变窄，而维持实质性的编辑效率(图93A和97C)。定义“编辑窗宽度”以表示人工计算的窗宽度，在该窗宽度内，编辑效率超过针对该靶物的半最大值。对于所测试的两个富含C的基因组位点，BE3的编辑窗宽度是5.0(位点A)和6.1(位点B)核苷酸。

[0912] 预测APOBEC1中的R126与ssDNA13的磷酸主链相互作用[13]。先前的研究已经显示了将相应的突变引入APOBEC3G将催化降低至少5倍[14]。令人感兴趣的是，当引入BE3中的APOBEC1中时，相对于BE3，R126A和R126E在最强编辑位置(C5，C6和C7)增加或维持活性，而在其他位置处降低编辑活性(图93A和97C)。因此，这两个突变中的每一个都将位点A和位点B处的编辑窗的宽度分别缩窄到4.4和3.4个核苷酸(R126A)或4.2和3.1个核苷酸(R126E)(图93A和97C)。

[0913] 预测APOBEC1中的W90(对应于APOBEC3G中的W285)在APOBEC3G活性位点中形成疏水口袋并帮助底物结合[13]。将该残基突变为Ala消除了APOBEC3G的催化活性[13]。在BE3中，W90A几乎完全消除了碱基编辑效率(图97C)。相比之下，发现W90Y仅适度降低了碱基编辑活性，而将位点A和位点B处的编辑窗宽度分别缩窄到3.8和4.9个核苷酸(图93A)。这些结果证明了对胞苷脱氨酶域的突变可以缩窄相应碱基编辑器的活性窗宽度。

[0914] 将W90Y，R126E和R132E(使编辑窗变窄而不显著降低碱基编辑活性的三个突变)组合成双重和三重突变的碱基编辑器。双重突变体W90Y+R126E产生碱基编辑器(YE1-BE3)，其具有BE3样的最大编辑效率，但实质性变窄的编辑窗宽度(位点A和位点B的宽度分别＝2.9和3.0个核苷酸(图93A)。W90Y+R132E碱基编辑器(YE2-BE3)表现出适度较低的编辑效率(与BE3相比，五个测试位点间的最大编辑产率平均低1.4倍)，以及还有实质性变窄的编辑窗宽度(位点A和位点B的宽度分别＝2.7和2.8个核苷酸)(图97C)。R126E+R132E双重突变体(EE-BE3)显示出与YE2-BE3类似的最大编辑效率和编辑窗宽度(图97C)。三重突变体W90Y+R126E+R132E(YEE-BE3)表现出低2.0倍的平均最大编辑产率，但非常少的超出C6位置的编辑，以及对于位点A和位点B分别为2.1和1.4个核苷酸的编辑窗宽度(图97C)。这些数据共同指示胞苷脱氨酶域中的突变可以强烈影响编辑窗的宽度，在某些情况下对编辑效率具有最小或

仅仅适度的影响。

[0915] 在HEK293T细胞中，进一步比较了BE3，YE1-BE3，YE2-BE3，EE-BE3和YEE-BE3的碱基编辑结果，它们靶向BE3活性窗内包含多个C的四个充分研究的人基因组位点[1]。这些靶基因座在位置4和5(HEK位置3)，位置4和6(HEK位置2)，位置5和6(EMX1)或位置6、7、8和11(FANCF)处包含靶C。对于在位置4-8活性窗中编辑任何C，BE3表现出很少(<1.2倍)的偏爱。相比之下，YE1-BE3在编辑C5相对于C4(HEK位点3)方面表现出1.3倍偏爱，在C6相对于C4(HEK位点2)方面表现出2.6倍偏爱，在C5相对于C6(EMX1)方面表现出2.0倍偏爱以及在C6相对于C7(FANCF)方面表现出1.5倍偏爱(图93B)。与YE1-BE3相比，YE2-BE3和EE-BE3表现出略大的位置特异性(较窄的活性范围)，平均而言在编辑C5相对于C4(HEK位点3)方面2.4倍偏爱，在C6相对于C4(HEK位点2)方面9.5倍偏爱，在C5相对于C6(EMX1)方面2.9倍偏爱以及在C7相对于C6(FANCF)方面2.6倍偏爱(图93B)。YEE-BE3显示出最大的位置选择性，在编辑C5相对于C4(HEK位点3)方面2.9倍偏爱，在C6相对于C4(HEK位点2)方面29.7倍偏爱，在C5相对于C6(EMX1)方面7.9倍偏爱以及在C7相对于C6(FANCF)方面7.9倍偏爱(图93B)。发现建立了即使这两个核苷酸都在BE3编辑窗之内，突变体碱基编辑器也可以区分相邻的C。

[0916] 通过HTS进一步分析这四个突变体和BE3的产物分布，以评估它们的表观持续合成能力。BE3在经处理的HEK293T细胞中主要产生T4-T5(HEK位点3)，T4-T6(HEK位点2)和T5-T6(EMX1)产物，平均而言产生与含有单一T的产物相比多7.4倍的含有2个T的产物。相比之下，YE1-BE3，YE2-BE3，EE-BE3和YEE-BE3对单编辑的C4-T5，C4-T6和T5-C6产物显示出实质性更高的偏爱(图93C)。YE1-BE3产生具有平均单T与双T产物比率1.4的产物。YE2-BE3和EE-BE3分别产生具有平均单T与双T产物比率为4.3和5.1的产物(图93C)。与以上结果一致，YEE-BE3三重突变体在三个基因组基因座间有利于单T产物达14.3倍的平均值。(图93C)。对于靶窗内只有一个C的靶位点(HEK位点4，位于位置C5)，所有四个突变体表现出与BE3相当的编辑效率(图98)。这些发现指示这些BE3突变体具有降低的表观持续合成能力，并且可以有利于在BE3编辑窗内包含多个C的靶位点处的仅单个C的转化。这些数据还提示，这些突变体碱基编辑器的位置偏爱C5>C6>C7≈C4，尽管这种偏爱可以随靶序列而不同。

[0917] 将APOBEC1中的窗调节突变应用于VQR-BE3，从而允许在NGA PAM靶向的位点处的底物的选择性碱基编辑(图107A)。然而，将这些突变应用于SaKKH-BE3时，观察到碱基编辑效率呈线性下降，而底物选择性没有改善，这提示该碱基编辑器与BE3及其变体具有不同的动力学平衡和底物可及性(图107B)。

[0918] 本研究中描述的具有改变的PAM特异性的五个碱基编辑器共同增加了ClinVar数据库中原则上可以通过碱基编辑校正的疾病相关突变的数目达2.5倍(图94A和94B)。类似地，具有变窄的编辑窗的碱基编辑器的开发使具有可以在没有对非靶C的相当多的修改的情况下通过碱基编辑校正的适当定位的NGG PAM的ClinVar条目的分数大致翻倍(从BE3的31％变为BE3的59％YEE-BE3)(图94A和94B)。

[0919] 总之，通过开发使用具有不同PAM特异性的Cas9变体的碱基编辑器，以及通过开发具有不同编辑窗宽度的脱氨酶突变体的集合实质性扩展碱基编辑的靶向范围。从理论上讲，使用其他可编程的DNA结合蛋白(例如Cpf1[16])进行碱基编辑应当是可能的，所述DNA结合蛋白创建可以充当单链特异性核苷酸脱氨酶的底物的单链DNA的气泡。

[0920] 材料和方法

[0921] 克隆。使用Q5热启动高保真DNA聚合酶(New England Biolabs)进行PCR。BE和sgRNA的质粒使用USER克隆(New England Biolabs)构建,获自先前报告的质粒[1]。使用NEB 10beta感受态细胞(New England Biolabs)进行DNA载体扩增。

[0922] 细胞培养。在补充有10%(v/v)胎牛血清(FBS)的Dulbecco的改良Eagle培养基加GlutaMax(ThermoFisher)中于37℃和5%CO_2培养HEK293T(ATCC CRL-3216)。在补充有10%(v/v)胎牛血清(FBS)和200μg/mL遗传霉素(ThermoFisher Scientific)的Dulbecco改良Eagle培养基加GlutaMax(ThermoFisher Scientific)中维持含有APOE基因ApoE4同种型的永生化大鼠星形胶质细胞(Taconic Biosciences)。

[0923] 转染。将HEK293T细胞接种在48孔胶原包被的BioCoat平板(Corning)上,并以约85%汇合度转染。根据制造商的方案,每孔使用1.5μl Lipofectamine 2000(ThermoFisher Scientific)转染750ng BE和250ng sgRNA表达质粒。

[0924] 基因组DNA样品的高通量DNA测序。3d后收获转染的细胞,并根据制造商的说明使用Agencourt DNAdvance基因组DNA分离试剂盒(Beckman Coulter)分离基因组DNA。通过PCR,使用补充序列中列出的侧翼HTS引物对扩增感兴趣的基因组区域。按照生产商的说明,使用Phusion热启动II DNA聚合酶(ThermoFisher)进行PCR扩增。PCR产物用RapidTips(Diffinity Genomics)纯化。进行二次PCR以连接测序衔接头。对产物进行凝胶纯化并且使用KAPA库定量试剂盒-Illumina(KAPA Biosystems)定量。如前所述[1],在Illumina MiSeq上对样品进行测序。

[0925] 数据分析。使用先前描述的MATLAB脚本[1]评估核苷酸频率。简而言之通过Smith-Waterman算法将读段与参考序列进行比对。Q得分低于30的碱基调用用占位核苷酸(N)替换。此质量阈值产生具有预期理论错误率为1000之1的核苷酸频率。

[0926] 使用自定义的python脚本进行碱基编辑持续合成能力的分析。该程序将测序读段修整为20个核苷酸的原间隔物序列,如由应当位于靶位点侧翼的7个核苷酸序列的完全匹配确定。然后将这些靶物合并,并按丰度进行分类,以评估碱基编辑产物的频率。

[0927] 以与先前描述的方式相似但具有少量调整[1]的方式进行人疾病相关突变的ClinVar数据库的生物信息学分析。这些调整实现使用可定制长度和序列的PAM鉴定靶物。此外,此改进的脚本包括对靶C位置的优先级排序(C5>C6>C7>C8≈C4),因此实现靶位点的鉴定,其中中靶C是窗内唯一的胞嘧啶或置于比编辑窗内任何脱靶C具有更高的预测编辑效率的位置处。

[0928] 实施例12的参考文献

[0929] 1 Komor,A.C.et al.Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage.Nature533,420-424(2016).

[0930] 2 Sander,J.D.&Joung,J.K.CRISPR-Cas systems for editing,regulating and targeting genomes.Nature biotechnology32,347-355(2014).

[0931] 3 Cong,L.et al.Multiplex genome engineering using CRISPR/Cas systems.Science339,819-823(2013).

[0932] 4 Ran,F.A.et al.Genome engineering using the CRISPR-Cas9 system.Nat.Protocols8,2281-2308(2013).

[0933] 5 Landrum,M.J.et al.ClinVar:public archive of interpretations of

clinically relevant variants.Nucleic Acids Res.44,D862-D868(2015).

[0934] 6 Nishida,K.et al.Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems.Science353,aaf8729-1-8(2016).

[0935] 7 Ma,Y.et al.Targeted AID-mediated mutagenesis(TAM)enables efficient genomic diversification in mammalian cells.Nat.Methods doi:10.1038/nmeth.4027(2016).

[0936] 8 Jiang,F.et al.Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage.Science351,867-71(2016).

[0937] 9 Ran,F.A.et al.In vivo genome editing using Staphylococcus aureus Cas9.Nature520,186-191(2015).

[0938] 10 Zhang,Y.et al.Comparison of non-canonical PAMs for CRISPR/Cas9-mediated DNA cleavage in human cells.Sci.Rep.4,(2014).

[0939] 11 Kleinstiver,B.P.et.al.Engineered CRISPR-Cas9 nucleases with altered PAMspecificities.Nature523,481-485(2015).

[0940] 12 Kleinstiver,B.P.et.al.Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition.Nat.Biotechnol.33,1293-1298(2015).

[0941] 13 Holden,L.G.et al.Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications.Nature452,121-124(2008).

[0942] 14 Chen,K.-M.et al.Structure of the DNA deaminase domain of the HIV-1restriction factor APOBEC3G.Nature452,116-119(2008).

[0943] 15 Harris,R.S.,Petersen-Mahrt,S.K.&Neuberger,M.S.RNA Editing Enzyme APOBEC1 and Some of Its Homologs Can Act as DNA Mutators.Molecular Cell10,1247-1253(2002).

[0944] 16 Zetsche,B.et al.Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2CRISPR-Cas System.Cell163,759-771(2015).

[0945] 实施例13

[0946] 使用改良的转染程序和更好的质粒,使用生物学重复(n＝3)来将四个HF突变安装到BE3的Cas9部分中。突变不显著影响质粒递送的中靶编辑(图99)。在测试的浓度下,BE3蛋白递送起作用。然而,中靶编辑低于质粒递送的情况(图100)。具有安装的HF突变的BE3的蛋白质递送降低了中靶编辑效率,但仍产生一些经编辑的细胞(图101)。

[0947] 显示了脂转染和安装HF突变两者减少脱靶脱氨基事件。对于图102中所示的四个位点,测定了具有最高的GUIDE-Seq读段和脱氨基事件的脱靶位点(OT)(Komor et al.,Nature,2016)。通过在最接近的相应C处将脱靶编辑除以中靶编辑来计算特异性比率。在检测不到脱靶编辑的情况下,该比率设置为100。因此,较高的特异性比率指示更特异性的构建体。与BE3的蛋白质递送,HF-BE3的质粒递送或HF-BE3的蛋白质递送相比,BE3质粒递送显示出高得多的脱靶/中靶编辑(图102和105)。

[0948] 在体外对纯化的蛋白质HF-BE3和BE3分析它们在最为允许的基序的间隔物中的不

同位置处将C转化为T残基的能力。发现BE3和HF-BE3蛋白都具有用于碱基编辑的相同的“窗”(图103和104)。

[0949] 表8给出了疾病靶物的列表。表8中要编辑的碱基以粗体和下划线指示。

[0950] 表8.碱基编辑器疾病靶物

[0951]

基因	疾病	间隔物	SEQ ID NO	PAM	编辑器	缺陷	细胞
RB1	RETINOBLASTOMA	AATCTAGTAAATAATTGATGT	571	AAAAGT	SAKKH-BE3	剪接损害	J82
PTEN	CANCER	GACCAACGGCTAAGTGAAGA	572	TGA	VQR-BE3	W111R	MC116
PIK3CA	CANCER	TCCTTTCTTCACGGTTGCCT	573	ACTGGT	SAKKH-BE3	K111R	CRL-5853
PIK3CA	CANCER	CTCCTGCTCAGTGATTTCAG	574	AGA	VQR-BE3	Q546R	CRL-2505
TP53	CANCER	TGTCACACATGTAGTTGTAG	575	TGG	YEE-BE3	N239D	SNU475

[0952]

HRAS	CANCER	CCTCCCGGCCGGCGGTATCC	576	AGG	YEE-BE3	Q61R	MC/CAR

[0953] 表6.可以使用碱基编辑器治疗的示例性疾病。sgRNA(PAM)栏中显示了原间隔物和PAM序列(SEQ ID NO:577-589)。PAM序列显示在括号中,并用下划线表示要编辑的碱基。

[0954] 表6

[0955]

疾病靶标	基因符号	变化的碱基	sgRNA (PAM)	基因编辑器
朊病毒病	PRNP	R37*	GGCAGCCGATACCCGGGGCA(GGG)	BE3
			GGGCAGCCGATACCCGGGGC(AGG)	
彭德莱综合征	Slc26a4	c.919-2A>G	TTATTGTCCGAAATAAAAGA(AGA)	BE3
			ATTGTCCGAAATAAAAGAAG(AGG)	(VQR
			TTGTCCGAAATAAAAGAAGA(GGA)	SaCas9)
			GTCCGAAATAAAAGAAGAGGAAAA(AAT)	
			GTCCGAAATAAAAGAAGAGGAAAAA(ATT)	
先天性耳聋	Tmc1	c.545A>G	CAGGAAGCACGAGGCCACTG(AGG)	BE3
			AACAGGAAGCACGAGGCCAC(TGA)	YE-BE3
			AGGAAGCACGAGGCCACTGA(GGA)	YEE-BE3
获得性耳聋	SNHL	S33F	TTGGATTCTGGAATCCATTC(TGG)	BE3
阿尔茨海默氏病	APP	A673T	TCTGCATCCATCTTCACTTC(AGA)	BE3 VQR
尼曼-皮克二氏病C型	NPC1	I1061T	CTTACAGCCAGTAATGTCAC(CGA)	BE3 VQR

[0956] 实施例14:测试碱基编辑构建体

[0957] 对几种碱基编辑构建体(包括BE3,BE4-pmCDA1,BE4-hAID,BE4-3G,BE4-N,BE4-SSB,BE4-$(GGS)_3$,BE4-XTEN,BE4-32aa,BE4-2xUGI,和BE4)测试它们在不同靶序列(即EMX1,

FANCF,HEK2,HEK3,HEK4和RNF2)内编辑胞嘧啶(C)残基的能力。例如,测试了这些构建体是否能够产生C至T突变。碱基编辑构建体的示意图显示在图109中。测试的靶序列也显示在图109中,其中靶定的胞嘧啶有编号并用红色表示。

[0958] 在本实施例的碱基编辑构建物中使用了以下氨基酸序列:

[0959] UGI:TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML(SEQ ID NO:736)

[0960] rAPOBEC1:

[0961] SSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK(SEQ ID NO:737)

[0962] pmCDA1:

[0963] MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQSGTERGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLKIWACKLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKRRSELSIMIQVKILHTTKSPAV(SEQ ID NO:81)

[0964] hAID:

[0965] MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL(SEQ ID NO:49)

[0966] hAPOBEC3G:

[0967] MELKYHPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPKVTLTIFVARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEPWNNLPKYYILLHIMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQNQEN(SEQ ID NO:60)

[0968] SSB(单链DNA结合蛋白):

[0969] ASRGVNKVILVGNLGQDPEVRYMPNGGAVANITLATSESWRDKATGEMKEQTEWHRVVLFGKLAEVASEYLRKGSQVYIEGQLRTRKWTDQSGQDRYTTEVVVNVGGTMQMLGGRQGGGAPAGGNIGGGQPQGGWGQPQQPQGGNQFSGGAQSRPQQSAPAAPSNEPPMDFDDDIPF(SEQ ID NO:590)

[0970] 接头序列:

[0971] XTEN:

[0972] SGSETPGTSESATPES(SEQ ID NO:604)

[0973] 32aa:

[0974] SGGSSGGSSGSETPGTSESATPESSGGSSGGS(SEQ ID NO:605)

[0975] SGGS:

[0976] SGGS(SEQ ID NO:606)

[0977] $(GGS)_3$:

[0978] GGSGGSGGS(SEQ ID NO:610)

[0979] 图109中显示的构建体的氨基酸序列在下文列出：

[0980] BE3:

[0981] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIE
KFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIM
TEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPP
HILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALL
FDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS
GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLA
QIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSK
NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPF
LKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVL
PKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVE
DRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKG
ILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLY
LYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLN
AKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVS
DFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS
NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDE
IIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
ATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDE
STDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV(SEQ ID NO:174)

[0982] BE4-pmCDA1:

[0983] MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQSGTERGIHAE
IFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLKIWACKLYYEKNARNQIGLWNLRDN
GVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKRRSELSIMIQVKILHTTKSPAVSGSETPGTSESA
TPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYT
RRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKA
DLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENL
IAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAIL
LSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVG
PLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKY
VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKD
FLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILD
FLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPE

NIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL
SDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLS
ELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAH
DAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRP
LIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSP
TVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRM
LASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL
GGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWAL
VIQDSNGENKIKMLSGGSPKKKRKV(SEQID NO:175)

[0984] BE4-hAID:

[0985] MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELLFLRYISDW
DLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDY
FYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGLSGSETPGTSESATPESDKKYSI
GLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQ
EIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALA
HMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKN
GLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTE
ITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTE
ELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFA
WMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF
LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDI
LEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR
NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMAREN
QTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVP
QSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR
QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGT
ALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE
IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKG
NELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRD
KPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNL
SDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENK
IKMLSGGSPKKKRKV(SEQ ID NO:176)

[0986] BE4-3G:

[0987] MELKYHPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPKVTLTIFVARLYYFW
DPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEPWNNLPKYYILLHIMLGEILRHSMDPPTFT
FNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRV
TCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVDHQG

CPFQPWDGLDEHSQDLSGRLRAILQNQENSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKF
KVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVE
EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKL
FIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAE
DAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLK
ALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH
QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASA
QSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ
GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG
SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD
ENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK
MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVK
KTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS
SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGS
PEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAF
KYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEV
EEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV(SEQ ID NO:
177)

[0988] BE4-N:

[0989] MTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPW
ALVIQDSNGENKIKMLGGSSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNT
NKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDL
ISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIAL
QSCHYQRLPPHILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHS
IKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQ
LFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDT
YDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEK
YKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
LRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNF
DKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIEC
FDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIAN
LAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVE
NTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVK

VITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIG
KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFS
KESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFL
EAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLF
VEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
RYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKV(SEQ ID NO:178)

[0990] BE4-SSB:

[0991] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIE
KFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIM
TEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPP
HILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALL
FDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS
GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLA
QIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSK
NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPF
LKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVL
PKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVE
DRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKG
ILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLY
LYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLN
AKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVS
DFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS
NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDE
IIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
ATLIHQSITGLYETRIDLSQLGGDSGGSGGSGGSASRGVNKVILVGNLGQDPEVRYMPNGGAVANITLATSESWRD
KATGEMKEQTEWHRVVLFGKLAEVASEYLRKGSQVYIEGQLRTRKWTDQSGQDRYTTEVVVNVGGTMQMLGGRQGG
GAPAGGNIGGGQPQGGWGQPQQPQGGNQFSGGAQSRPQQSAPAAPSNEPPMDFDDDIPFSGGSPKKKRKV(SEQ ID
NO:179)

[0992] BE4-$(GGS)_3$:

[0993] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIE
KFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIM
TEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPP
HILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALL
FDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS

GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLA
QIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSK
NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPF
LKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVL
PKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVE
DRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKG
ILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLY
LYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLN
AKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVS
DFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS
NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDE
IIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
ATLIHQSITGLYETRIDLSQLGGDSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILV
HTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV(SEQ ID NO:180)

[0994] BE4-XTEN:

[0995] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIE
KFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIM
TEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPP
HILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALL
FDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS
GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLA
QIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSK
NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPF
LKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVL
PKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVE
DRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKG
ILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLY
LYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLN
AKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVS
DFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS
NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDE
IIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD

ATLIHQSITGLYETRIDLSQLGGDSGSETPGTSESATPESTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKP
ESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV(SEQ ID NO:181)

[0996] BE4-32aa:

[0997] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIE
KFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIM
TEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPP
HILWATGLKSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLG
NTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQL
VQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKL
QLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVR
QQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHL
GELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFI
ERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDY
FKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL
HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQIL
KEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDK
LIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAK
SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEK
NPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDN
EQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFD
TTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVI
GNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV(SEQ ID NO:182)

[0998] BE4-2XUGI:

[0999] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIE
KFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIM
TEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPP
HILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALL
FDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS
GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLA
QIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSK
NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPF
LKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVL
PKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVE
DRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG

RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKG
ILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLY
LYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLN
AKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVS
DFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS
NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDE
IIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
ATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDE
STDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPE
SDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV(SEQ ID NO:183)

[1000] BE4:

[1001] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIE
KFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIM
TEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPP
HILWATGLKSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLG
NTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQL
VQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKL
QLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVR
QQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHL
GELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFI
ERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDY
FKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL
HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQIL
KEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDK
LIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAK
SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEK
NPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDN
EQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFD
TTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPE
EVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGSTNLSDIIEK
ETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSG
GSPKKKRKV(SEQ ID NO:184)

[1002] 在图110-115中显示了图109的碱基编辑构建体突变EMX1,FANCF,HEK2,HEK3,HEK4

和RNF2的靶胞嘧啶残基的能力。图110-115显示了编辑的靶胞嘧啶的百分比(包括突变为T，A或G的C残基的比例)以及生成的插入/缺失％。使用以下公式计算编辑的靶胞嘧啶的百分比：100-[具有C的测序读段的％]。图110-115的饼形图显示了带有指示的各种碱基的读段的分布，这意味着查看所有经碱基编辑的读段(在指示的碱基上具有C以外的核苷酸的读段)，具有A，G或T的那些读段的何种百分比。下面的表9-14显示了图110-115的饼图中所指示的突变百分比的值。

[1003] 观察到的C至非T编辑可能是由于UDG(尿嘧啶DNA糖基化酶)。例如，一旦C转化为尿嘧啶中间体，UDG便可以将其转化为无碱基位点。然后，该无碱基位点被其他内源性酶处理，并最终导致插入/缺失或替换C的其他碱基(例如G或A)。我们已经显示了在UDG敲除细胞系中显示具有很少插入/缺失或完全没有插入/缺失的增加的C至T编辑。

[1004] 表9 EMX1

[1005]

EMX1		C_5
BE3	A	1.9%
	C	51.9%
	G	6.0%
	T	40.3%
pmCDA1	A	0.2%
	C	88.9%
	G	0.8%
	T	10.0%
hAID	A	0.3%
	C	83.6%
	G	0.6%
	T	15.4%
hAPOBEC3G	A	0.0%
	C	98.9%
	G	0.2%
	T	1.0%

[1006]

BE4-N	A	1.4%
	C	85.6%
	G	3.5%
	T	9.5%
EMX1		
BE4-SSB	A	0.3%
	C	96.7%
	G	0.7%
	T	2.3%
BE4-$(GGS)_3$	A	1.7%
	C	36.8%
	G	4.1%
	T	57.4%
BE4-XTEN	A	2.1%
	C	45.7%
	G	5.1%
	T	47.1%
BE4-32aa	A	1.6%
	C	46.3%
	G	4.6%
	T	47.5%
BE4-2XUGI	A	0.6%
	C	60.2%
	G	1.9%
	T	37.3%

[1007] 表10 FANCF

[1008]

FANCF		C_8
BE3	A	1.6%
	C	74.4%
	G	0.7%
	T	23.3%
pmCDA1	A	0.5%
	C	88.8%
	G	0.3%
	T	10.3%
hAID	A	0.3%
	C	89.6%

[1009]

	G	0.2%
	T	9.9%
hAPOBEC3G	A	2.1%
	C	75.4%
	G	10.4%
	T	12.1%
BE4-N	A	1.8%
	C	83.5%
	G	0.6%
	T	14.2%
BE4-SSB	A	0.3%
	C	97.1%
	G	0.1%
	T	2.6%
BE4-$(GGS)_3$	A	2.3%
	C	45.7%
	G	1.4%
	T	50.6%
BE4-XTEN	A	1.1%
	C	56.3%
	G	0.4%
	T	42.1%
BE4-32aa	A	1.2%
	C	70.0%
	G	0.6%
	T	28.2%
BE4-2XUGI	A	0.9%
	C	57.8%
	G	1.1%
	T	40.2%

[1010] 表11 HEK2

[1011]

HEK2		C_6
BE3	A	0.9%
	C	28.2%
	G	52.9%
	T	18.1%

[1012]

pmCDA1	A	1.8%
	C	73.5%
	G	3.5%
	T	21.2%
hAID	A	2.1%
	C	56.9%
	G	7.7%
	T	33.3%
hAPOBEC3G	A	0.1%
	C	86.5%
	G	9.9%
	T	3.5%
BE4-N	A	1.0%
	C	37.6%
	G	57.0%
	T	4.4%
BE4-SSB	A	0.2%
	C	78.4%
	G	20.0%
	T	1.4%
BE4-$(GGS)_3$	A	0.6%
	C	11.1%
	G	40.6%
	T	47.7%
BE4-XTEN	A	1.2%
	C	24.8%
	G	44.6%
	T	29.4%
BE4-32aa	A	1.1%
	C	26.3%
	G	41.8%
	T	30.7%
BE4-2XUGI	A	0.8%
	C	37.0%
	G	21.6%
	T	40.6%

[1013] 表12 HEK3

[1014]

HEK3		C_5
BE3	A	2.23%
	C	38.06%
	G	12.77%
	T	46.95%
pmCDA1	A	0.21%
	C	76.57%
	G	0.12%
	T	23.09%
hAID	A	0.28%
	C	60.23%
	G	1.03%
	T	38.45%
hAPOBEC3G	A	3.11%
	C	33.89%
	G	28.59%
	T	34.41%
BE4-N	A	2.6%
	C	64.1%
	G	13.5%
	T	19.8%
BE4-SSB	A	0.4%
	C	92.9%
	G	2.8%
	T	3.9%
BE4-$(GGS)_3$	A	1.3%
	C	9.9%
	G	7.9%
	T	80.8%
BE4-XTEN	A	2.3%
	C	15.9%
	G	12.2%
	T	69.6%
BE4-32aa	A	1.3%
	C	14.9%

[1015]

		G	9.9%
		T	73.9%
	BE4-2XUGI	A	0.6%
		C	23.4%
		G	3.8%
		T	72.2%

[1016] 表13 HEK4

[1017]

HEK4		C_5
BE3	A	8.40%
	C	41.89%
	G	24.54%
	T	25.17%
pmCDA1	A	0.50%
	C	87.53%
	G	0.01%
	T	11.95%
hAID	A	0.93%
	C	71.32%
	G	0.69%
	T	27.06%
hAPOBEC3G	A	0.12%
	C	99.37%
	G	0.35%
	T	0.16%
BE4-N	A	7.3%
	C	56.6%
	G	25.7%
	T	10.3%
BE4-SSB	A	2.1%
	C	86.8%
	G	5.8%
	T	5.2%
BE4-$(GGS)_3$	A	6.7%
	C	13.0%
	G	19.8%
	T	60.5%

[1018]

BE4-XTEN	A	7.5%
	C	19.7%
	G	25.4%
	T	47.4%
BE4-32aa	A	7.9%
	C	21.8%
	G	25.1%
	T	45.3%
BE4-2XUGI	A	3.4%
	C	22.2%
	G	12.4%
	T	62.0%

[1019] 表14 RNF2

[1020]

RNF2		C_6
BE3	A	2.46%
	C	46.65%
	G	19.87%
	T	31.03%
pmCDA1	A	0.60%
	C	83.52%
	G	1.33%
	T	14.55%
hAID	A	0.36%
	C	75.03%
	G	3.20%
	T	21.40%
hAPOBEC3G	A	0.10%
	C	86.60%
	G	3.70%
	T	9.59%
BE4-N	A	5.1%
	C	50.0%
	G	28.8%
	T	16.2%
BE4-SSB	A	1.1%
	C	89.9%
	G	4.9%

[1021]

	T	4.1%
BE4-$(GGS)_3$	A	2.0%
	C	23.0%
	G	14.0%
	T	61.0%
BE4-XTEN	A	2.6%
	C	32.4%
	G	16.0%
	T	49.0%
BE4-32aa	A	2.2%
	C	29.2%
	G	18.5%
	T	50.0%
BE4-2XUGI	A	0.7%
	C	45.0%
	G	6.5%
	T	47.8%

[1022] 实施例15:碱基编辑器包括一个切掉了靶链的Cpf1切口酶。

[1023] 如上所讨论,本文提供的任何融合蛋白的核酸可编程DNA结合蛋白(napDNAbp)可以是Cpf1蛋白。在一些实施方案中,Cpf1蛋白是Cpf1切口酶(nCpf1)。切割非靶链的Cpf1切口酶,例如Cpf1切口酶(AsCpf1中的R1225A;LbCpf1中的R1138A)已记载于Yamano et al.,"Crystal structure of Cpf1 in complex with guide RNA and target DNA."Cell(165)2016,p.949-962;其全部内容通过引用并入本文。然而,预期切割靶链的切口酶(例如,碱基编辑器蛋白的Cpf1切口酶)改善碱基编辑效率。

[1024] 使用荧光标记的DNA来鉴定Cpf1突变体,该突变体优先对靶链而不是非靶链产生切口(参见图116)。在图116中,DNA构建体1-3的顶部链是非靶链,而底部链是靶链。使用野生型LbCpf1,R836A(LbCpf1)和R1138A(LbCpf1)进行体外测定法。R836A(LbCpf1)似乎是"残缺"的切口酶,这意味着它切割靶链比切割非靶链更有效。如图117所示,未切割非靶链,并且未观察到荧光350块。两个小时后,这两条链都被切割。不同强度提示比非靶链切割更多的靶链。

[1025] 用AsCp1-BE3建立碱基编辑窗

[1026] 显示具有LbCpf1(R836A)或AsCpf1(R912A)作为napDNAbp的碱基编辑器蛋白(例如BE3(SpCas9-BE3))以低效率(0.1%至0.4%)编辑碱基。具有AsCpf1(R912A)napDNAbp的碱基编辑器可以更有效地突变EMX1,FANCF,HEK3和HEK4位点处的靶C。测试的构建体的编辑窗似乎是从第7个碱基至第11个碱基。数字与使用具有最高数字的BE3和具有较低数字的自我挫败BE(即APOBEC-AsCpf1(R1225A)-UGI,其切割非靶链)的趋势一致。参见图118在EMX1上使用Cas9-BE3的阳性对照:5-6%。AsCpf1的插入/缺失值:>20%。AsCpf1中的R912在Cpf1家

族的许多成员中都是保守的。LbCpf1中的相应残基为R836，当R突变为A时，它被认为是“残缺的”切口酶。

[1027] Cpf1-BE的优化(接头)

[1028] 插入/缺失数据提示Cpf1可以接近DNA靶位点。因此，Cpf1碱基编辑器蛋白的优化已集中在特定的APOBEC蛋白，接头和/或UGI域上。使用LbCpf1(R836A)和AsCpf1(R912A)两者，使用不同的接头测试了图119中所示的构建体。简而言之，测试了POBEC和Cpf1域(例如AsCpf1或LbCpf1)之间不同的接头序列(即XTEN,GGS,$(GGS)_3$(SEQ ID NO:610),$(GGS)_5$(SEQ ID NO:610),和$(GGS)_7$(SEQ ID NO:610))。参见图120。对构建体测试了它们突变HEK3位点的C_8残基的能力，所述HEK3位点为TGCTTCTC$_8$CAGCCCTGGCCTGG(SEQ ID NO:592)。具有AsCpf1的碱基编辑器蛋白的编辑水平达到1％以上，而具有LbCpf1的碱基编辑器蛋白在碱基编辑效率上显示相当的降低。如图121所示，对于AsCpf1-BE3，来自综合生物信息学中心VU(Centre of Integrative Bioinformatics VU)维护的数据库中的接头没有显示与GGS型接头一样显著的改善。图121所示的接头如下所示：

[1029]

PDB_code	长度 (aa)	序列
1au7A_1	10	KRRTTISIAA (SEQ ID NO: 593)
1c1kA_1	19	ALVFYREYIGRLKQIKFKF (SEQ ID NO: 594)
1c20A_1	14	LPIMAKSVLDLYEL (SEQ ID NO: 595)
1ee8A_1	5	LLRLG (SEQ ID NO: 596)
1f1zA_1	15	TDKEINPVVKENIEW (SEQ ID NO: 597)
1ignA_1	8	PPSIKRKF (SEQ ID NO: 598)
1jmcA_1	9	LPTVQFDFT (SEQ ID NO: 599)
1sfe_1	14	LPLDIRGTAFQQQV (SEQ ID NO: 600)

[1030]

2ezx_1	8	AYVVLGQF (SEQ ID NO: 601)
2reb_1	8	INFYGELV (SEQ ID NO: 602)

[1031] Cpf1-BE的优化(方向)

[1032] Cas9在C和N末端之间具有一段氨基酸(参见图123，红色正方形)，而AsCpf1没有(图122)。此外，AsCpf1在N和C末端之间的距离更短。这些观察结果指示APOBEC(在N端)和UGI(在C端)之间存在潜在的干扰，UGI可能通过所述干扰来阻止APOBEC接近非靶链。一种解决方案是将APOBEC和UGI移到相同的末端(N或C)上。因此，测试了具有结构NLS-UGI-APOBEC-XTEN-AsCpf1;UGI-APOBEC-XTEN-AsCpf1-NLS;和AsCpf1-XTEN-APOBEC-NLS的构建体。

[1033] Cpf1-BE的优化(内部截短)

[1034] 没有解析非靶链的已知Cpf1晶体结构(参见图124，青色)。认为编辑窗应位于如图124所示的红色圆圈内。存在可以阻碍APOBEC的螺旋区域(参见图124中的正方形)。该区域包含氨基酸序列K(661)KTGDQK(667)(SEQ ID NO:603)。

[1035] 为了测试去除两个，四个或六个残基是否改善碱基编辑效率，使用HEK3作为靶位点，使用具有AsCpf1(R912A)napDNAbp的碱基编辑器进行实验。编辑水平增加到约2.6％——当删除T663和D665时，自对照水平增加6倍(参见下面的表7)。本实验中使用的构

建体是APOBEC-XTEN-AsCpf1(R912A)-SGGS-UGI

[1036] 表7

[1037]

缺失	C8处的编辑	C9处的编辑
T663,D665	2.59%	1.29%
K662,T663,D665,Q666	0.15%	0.15%
K661,K662,T663,D665,Q666,K667	0.22%	0.21%

[1038] 参考文献

[1039] 1.Humbert O,Davis L,Maizels N.Targeted gene therapies:tools, applications,optimization.Crit Rev Biochem Mol.2012;47(3):264-81.PMID: 22530743.

[1040] 2.Perez-Pinera P,Ousterout DG,Gersbach CA.Advances in targeted genome editing.Curr Opin Chem Biol.2012;16(3-4):268-77.PMID:22819644.

[1041] 3.Urnov FD,Rebar EJ,Holmes MC,Zhang HS,Gregory PD.Genome editing with engineered zinc finger nucleases.Nat Rev Genet.2010;11(9):636-46.PMID: 20717154.

[1042] 4.Joung JK,Sander JD.TALENs:a widely applicable technology for targeted genome editing.Nat Rev Mol Cell Biol.2013;14(1):49-55.PMID:23169466.

[1043] 5.Charpentier E,Doudna JA.Biotechnology:Rewriting a genome.Nature.2013;495,(7439):50-1.PMID:23467164.

[1044] 6.Pan Y,Xia L,Li AS,Zhang X,Sirois P,Zhang J,Li K.Biological and biomedical applications of engineered nucleases.Mol Biotechnol.2013;55(1):54-62.PMID:23089945.

[1045] 7.De Souza,N.Primer:genome editing with engineered nucleases.Nat Methods.2012;9(1):27.PMID:22312638.

[1046] 8.Santiago Y,Chan E,Liu PQ,Orlando S,Zhang L,Urnov FD,Holmes MC, Guschin D,Waite A,Miller JC,Rebar EJ,Gregory PD,Klug A,Collingwood TN.Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases.Proc Natl Acad Sci U S A.2008;105(15):5809-14.PMID:18359850.

[1047] 9.Cargill M,Altshuler D,Ireland J,Sklar P,Ardlie K,Patil N,Lane CR, Lim EP,Kalyanaraman N,Nemesh J,Ziaugra L,Friedland L,Rolfe A,Warrington J, Lipshutz R,Daley GQ,Lander ES.Characterization of single-nucleotide polymorphisms in coding regions of human genes.Nat Genet.1999;22(3):231-8.PMID:10391209.

[1048] 10.Jansen R,van Embden JD,Gaastra W,Schouls LM.Identification of genes that are associated with DNA repeats in prokaryotes.Mol Microbiol.2002; 43(6):1565-75.PMID:11952905.

[1049] 11.Mali P,Esvelt KM,Church GM.Cas9 as a versatile tool for engineering biology.Nat Methods.2013;10(10):957-63.PMID:24076990.

[1050] 12.Jore MM,Lundgren M,van Duijin E,Bultema JB,Westra ER,Waghmare SP,

Wiedenheft B,Pul U,Wurm R,Wagner R,Beijer MR,Barendregt A,Shou K,Snijders AP,Dickman MJ,Doudna JA,Boekema EJ,Heck AJ,van der Oost J,Brouns SJ.Structural basis for CRISPR RNA-guided DNA recognition by Cascade.Nat Struct Mol Biol.2011;18(5):529-36.PMID:21460843.

[1051] 13.Horvath P,Barrangou R.CRISPR/Cas,the immune system of bacteria and archaea.Science.2010;327(5962):167-70.PMID:20056882.

[1052] 14.Wiedenheft B,Sternberg SH,Doudna JA.RNA-guided genetic silencing systems in bacteria and archaea.Nature.2012;482(7385):331-8.PMID:22337052.

[1053] 15.Gasiunas G,Siksnys V.RNA-dependent DNA endonuclease Cas9 of the CRISPR system:Holy Grail of genome editing?Trends Microbiol.2013;21(11):562-7.PMID:24095303.

[1054] 16.Qi LS,Larson MH,Gilbert LA,Doudna JA,Weissman JS,Arkin AP,Lim WA.Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression.Cell.2013;152(5):1173-83.PMID:23452860.

[1055] 17.Perez-Pinera P,Kocak DD,Vockley CM,Adler AF,Kabadi AM,Polstein LR,Thakore PI,Glass KA,Ousterout DG,Leong KW,Guilak F,Crawford GE,Reddy TE,Gersbach CA.RNA-guided gene activation by CRISPR-Cas9-based transcription factors.Nat Methods.2013;10(10):973-6.PMID:23892895.

[1056] 18.Mali P,Aach J,Stranges PB,Esvelt KM,Moosburner M,Kosuri S,Yang L,Church GM.CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering.Nat Biotechnol.2013;31(9):833-8.PMID:23907171.

[1057] 19.Gilbert LA,Larson MH,Morsut L,Liu Z,Brar GA,Torres SE,Stern-Ginossar N,Brandman O,Whitehead EH,Doudna JA,Lim WA,Weissman JS,Qi LS.CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes.Cell.2013;154(2):442-51.PMID:23849981.

[1058] 20.Larson MH,Gilbert LA,Wang X,Lim WA,Weissman JS,Qi LS.CRISPR interference (CRISPRi)for sequence-specific control of gene expression.Nat Protoc.2013;8(11):2180-96.PMID:24136345.

[1059] 21.Mali P,Yang L,Esvelt KM,Aach J,Guell M,DiCarlo JE,Norville JE,Church GM.RNA-guided human genome engineering via Cas9.Science.2013;339(6121):823-6.PMID:23287722.

[1060] 22.Cole-Strauss A,Yoon K,Xiang Y,Byrne BC,Rice MC,Gryn J,Holloman WK,Kmiec EB.Correction of the mutation responsible for sickle cell anemia by an RNA-DNA oligonucleotide.Science.1996;273(5280):1386-9.PMID:8703073.

[1061] 23.Tagalakis AD,Owen JS,Simons JP.Lack of RNA-DNA oligonucleotide (chimeraplast)mutagenic activity in mouse embryos.Mol Reprod Dev.2005;71(2):140-4.PMID:15791601.24.Ray A,Langer M.Homologous recombination:ends as the means.Trends Plant Sci.2002;7(10):435-40.PMID 12399177.

[1062] 25.Britt AB,May GD.Re-engineering plant gene targeting.Trends Plant Sci.2003;8(2):90-5.PMID:12597876.

[1063] 26.Vagner V,Ehrlich SD.Efficiency of homologous DNA recombination varies along the Bacillus subtilis chromosome.J Bacteriol.1988;170(9):3978-82.PMID:3137211.

[1064] 27.Saleh-Gohari N,Helleday T.Conservative homologous recombination preferentially repairs DNA double-strand breaks in the S phase of the cell cycle in human cells.Nucleic Acids Res.2004;32(12):3683-8.PMID:15252152.

[1065] 28.Lombardo A,Genovese P,Beausejour CM,Colleoni S,Lee YL,Kim KA,Ando D,Urnov FD,Galli C,Gregory PD,Holmes MC,Naldini L.Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery.Nat Biotechnol.2007;25(11):1298-306.PMID:17965707.

[1066] 29.Conticello SG.The AID/APOBEC family of nucleic acid mutators.Genome Biol.2008;9(6):229.PMID:18598372.

[1067] 30.Reynaud CA,Aoufouchi S,Faili A,Weill JC.What role for AID:mutator, or assembler of the immunoglobulin mutasome?Nat Immunol.2003;4(7):631-8.

[1068] 31.Bhagwat AS.DNA-cytosine deaminases:from antibody maturation to antiviral defense.DNA Repair(Amst).2004;3(1):85-9.PMID:14697763.

[1069] 32.Navaratnam N,Sarwar R.An overview of cytidine deaminases.Int J Hematol.2006;83(3):195-200.PMID:16720547.

[1070] 33.Holden LG,Prochnow C,Chang YP,Bransteitter R,Chelico L,Sen U, Stevens RC,Goodman MF,Chen XS.Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications.Nature.2008;456(7218):121-4.PMID:18849968.

[1071] 34.Chelico L,Pham P,Petruska J,Goodman MF.Biochemical basis of immunological and retroviral responses to DNA-targeted cytosine deamination by activation-induced cytidine deaminase and APOBEC3G.J Biol Chem.2009;284(41).27761-5.PMID:19684020.

[1072] 35.Pham P,Bransteitter R,Goodman MF.Reward versus risk:DNA cytidine deaminases triggering immunity and disease.Biochemistry.2005;44(8):2703-15.PMID 15723516.

[1073] 36.Chen X,Zaro JL,Shen WC.Fusion protein linkers:property,design and functionality.Adv Drug Deliv Rev.2013;65(10):1357-69.PMID:23026637.

[1074] 37.Lee JW,Soung YH,Kim SY,Lee HW,Park WS,Nam SW,Kim SH,Lee JY,Yoo NJ, Lee SH.PIK3CA gene is frequently mutated in breast carcinomas and hepatocellular carcinomas.Oncogene.2005;24(8):1477-80.PMID:15608678.

[1075] 38.Ikediobi ON,Davies H,Bignell G,Edkins S,Stevens C,O'Meara S, Santarius T,Avis T,Barthorpe S,Brackenbury L,Buck G,Butler A,Clements J,Cole J,Dicks E,Forbes S,Gray K,Halliday K,Harrison R,Hills K,Hinton J,Hunter C,

Jenkinson A,Jones D,Kosmidou V,Lugg R,Menzies A,Mironenko T,Parker A,Perry J,Raine K,Richardson D,Shepherd R,Small A,Smith R,Solomon H,Stephens P,Teaque J,Tofts C,Varian J,Webb T,West S,Widaa S,Yates A,Reinhold W,Weinstein JN,Stratton MR,Futreal PA,Wooster R.Mutation analysis of 24 known cancer genes in the NCI-60 cell line set.Mol Cancer Ther.2006;5(11):2606-12.PMID:17088437.

[1076] 39.Cox,D.B.,Platt,R.J.&Zhang,F.Therapeuticgenomeediting:prospectsandchallenges.Naturemedicine21,121-131,doi:10.1038/nm.3793(2015).

[1077] 40.Hilton,I.B.&Gersbach,C.A.Enablingfunctionalgenomicswithgenomeengineering.Genomeresearch25,1442-1455,doi:10.1101/gr.190124.115(2015).

[1078] 41.Sander,J.D.&Joung,J.K.CRISPR-Cassystemsforediting,regulatingandtargetinggenomes.Naturebiotechnology32,347-355,doi:10.1038/nbt.2842(2014).

[1079] 42.Maruyama,T.etal.Increasingtheefficiencyofprecisegenomeeditingwith CRISPR-Cas9byinhibitionofnonhomologousendjoining.Naturebiotechnology33,538-542,doi:10.1038/nbt.3190(2015).

[1080] 43.Chu,V.T.etal.Increasingtheefficiencyofhomology-directedrepairforCRISPR-Cas9-inducedprecisegeneeditinginmammaliancells.Naturebiotechnology33,543-548,doi:10.1038/nbt.3198(2015).

[1081] 44.Lin,S.,Staahl,B.T.,Alla,R.K.&Doudna,J.A.Enhancedhomology-directed humangenomeengineeringbycontrolledtimingofCRISPR/Cas9delivery.eLife3,e04766,doi:10.7554/eLife.04766(2014).

[1082] 45.Cong,L.etal.MultiplexgenomeengineeringusingCRISPR/Cassystems.Science339,819-823,doi:10.1126/science.1231143(2013).

[1083] 46.Rong,Z.,Zhu,S.,Xu,Y.&Fu,X.Homologousrecombinationinhumanembryonicstemcellsusing CRISPR/Cas9nickaseandalongDNAdonortemplate.Protein&cell5,258-260,doi:10.1007/s13238-014-0032-5(2014).

[1084] 47.Jinek,M.etal.AProgrammableDual-RNA-GuidedDNAEndonucleaseinAdaptiveBacterialImmunity.Science337,816-821,doi:10.1126/science.1225829(2012).

[1085] 48.Harris,R.S.,Petersen-Mahrt,S.K.&Neuberger,M.S.RNAEditingEnzymeAPOBEC1andSomeofItsHomologsCanActasDNAMutators.MolecularCell10,1247-1253(2002).

[1086] 49.Jinek,M.etal.StructuresofCas9endonucleasesrevealRNA-mediatedconformationalactivation.Science343,1247997,doi:10.1126/science.1247997(2014).

[1087] 50.Schellenberger,V.etal.Arecombinantpolypeptideextendstheinvivohalf-lifeofpeptidesandproteinsinatunablemanner.Naturebiotechnology27,1186-1190,doi:10.1038/nbt.1588(2009).

[1088] 51.Saraconi,G.,Severi,F.,Sala,C.,Mattiuz,G.&Conticello,S.G.TheRNAeditingenzymeAP OBEC 1inducessomaticmutationsandacompatiblemutationalsignatureispresentinesophagealadenocarcinomas.Genomebiology15,417-(2014).

[1089] 52.Anders,C.,Niewoehner,O.,Duerst,A.&Jinek,M.StructuralbasisofPAM-dependenttarg etDNArecognitionbytheCas9endonuclease.Nature513,569-573,doi:

10.1038/nature13579(2014).

[1090] 53.Cong,L.etal.MultiplexgenomeengineeringusingCRISPR/Cassystems.Science339,819-823(2013).

[1091] 54.Tsai,S.Q.etal.GUIDE-seqenablesgenome-wideprofilingofoff-targetcleavagebyCRISP R-Casnucleases.Naturebiotechnology33,187-197,doi:10.1038/nbt.3117(2015).

[1092] 55.Kunz,C.,Saito,Y.&Schar,P.DNARepairinmammaliancells:Mismatchedrepair:variationsonatheme.Cellularandmolecularlifesciences:CMLS66,1021-1038,doi:10.1007/s00018-009-8739-9(2009).

[1093] 56.D.,M.C.etal.Crystalstructureofhumanuracil-DNAglycosylaseincomplexwithaproteininhibitor:proteinmimicryofDNA.Cell82,701-708(1995).

[1094] 57.Caldecott,K.W.Single-strandbreakrepairandgeneticdisease.Naturereviews.Genetics9,619-631,doi:10.1038/nrg2380(2008).

[1095] 58.Fukui,K.DNAmismatchrepairineukaryotesandbacteria.Journalofnucleicacids2010,doi:10.4061/2010/260512(2010).

[1096] 59.Gasiunas,G.,Barrangou,R.,Horvath,P.&Siksnys,V.Cas9-crRNAribonucleoproteincomplexmediatesspecificDNAcleavageforadaptiveimmunityinbacteria.ProceedingsoftheNationalAcademyofSciences109,E2579-E2586,doi:10.1073/pnas.1208507109(2012).

[1097] 60.Ran,F.A.etal.InvivogenomeeditingusingStaphylococcusaureusCas9.Nature520,186-191,doi:10.1038/nature14299(2015).

[1098] 61.Kuscu,C.,Arslan,S.,Singh,R.,Thorpe,J.&Adli,M.Genome-wideanalysisrevealscharacteristicsofoff-targetsitesboundbytheCas9endonuclease.Naturebiotechnology32,677-683,doi:10.1038/nbt.2916(2014).

[1099] 62.Wu,X.etal.Genome-widebindingoftheCRISPRendonucleaseCas9inmammaliancells.Naturebiotechnology32,670-676,doi:10.1038/nbt.2889(2014).

[1100] 63.Beale,R.C.L.etal.ComparisonoftheDifferentialContext-dependenceofDNADeaminationbyAPOBECEnzymes:CorrelationwithMutationSpectrainVivo.JournalofMolecularBiology337,585-596,doi:10.1016/j.jmb.2004.01.046(2004).

[1101] 64.Kim,J.,Basak,J.M.&Holtzman,D.M.TheroleofapolipoproteinEinAlzheimer'sdisease.Neuron63,287-303,doi:10.1016/j.neuron.2009.06.026(2009).

[1102] 65.Liu,C.C.,Kanekiyo,T.,Xu,H.&Bu,G.ApolipoproteinEandAlzheimerdisease:risk,mechanismsandtherapy.Naturereviews.Neurology9,106-118,doi:10.1038/nrneurol.2012.263(2013).

[1103] 66.Sjöblom,T.etal.TheConsensusCodingSequencesofHumanBreastandColorectalCancers.Science314,268-274,doi:10.1126/science.1133427(2006).

[1104] 67.Stephens,P.J.etal.Thelandscapeofcancergenesandmutationalprocessesinbreastcancer.Nature486,400-404,doi:10.1038/nature11017(2012).

[1105] 68.Landrum,M.J.etal.ClinVar:publicarchiveofinterpretationsofclinical

lyrelevantvariants.NucleicAcidsResearch,doi:10.1093/nar/gkv1222(2015).

[1106] 69.Slaymaker,I.M.etal.RationallyengineeredCas9nucleaseswithimprovedspecificity.Science,doi:10.1126/science.aad5227(2015).

[1107] 70.Davis,K.M.,Pattanayak,V.,Thompson,D.B.,Zuris,J.A.&Liu,D.R.Smallmolecule-triggeredCas9proteinwithimprovedgenome-editingspecificity.Naturechemicalbiology11,316-318,doi:10.1038/nchembio.1793(2015).

[1108] 71.Zuris,J.A.etal.Cationiclipid-mediateddeliveryofproteinsenableseefficientprotein-based genomeeditinginvitroandinvivo.Naturebiotechnology33,73-80,doi:10.1038/nbt.3081(2015).

[1109] 72.Kleinstiver,B.P.etal.EngineeredCRISPR-Cas9nucleaseswithalteredPAMspecificities.Nature523,481-485,doi:10.1038/nature14592(2015).

[1110] 73.Pattanayak,V.etal.High-throughputprofilingofoff-targetDNAcleavagerevealsRNA-programmedCas9nucleasespecificity.NatureBiotechnology31,839-843,doi:10.1038/nbt.2673(2013).

[1111] 74.Shcherbakova,D.M.&Verkhusha,V.V.Near-infraredfluorescentproteinsformulticolorinvivoimaging.NatureMethods10,751-754,doi:10.1038/nmeth.2521(2013).

[1112] 75.Ran,F.A.et al.Genome engineering using the CRISPR-Cas9 system.Nat.Protocols 8,2281-2308,doi:10.1038/nprot.2013.143(2013).

[1113] 76.Jiang,F.et al.Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage.Science,doi:10.1126/science.aad8282(2016).

[1114] 77.Tsai,S.Q.et al.Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing.Nat Biotech 32,569-576,doi:10.1038/nbt.2908(2014).

[1115] 78.Lieber,M.R.,Ma,Y.,Pannicke,U.&Schwarz,K.Mechanism and regulation of human non-homologous DNA end-joining.Nat Rev Mol Cell Biol 4,712-720(2003).

[1116] 79.Heller,R.C.&Marians,K.J.Replisome assembly and the direct restart of stalled replication forks.Nat Rev Mol Cell Biol 7,932-943(2006).

[1117] 80.Pluciennik,A.et al.PCNA function in the activation and strand direction of MutLαendonuclease in mismatch repair.Proceedings of the National Academy of Sciences of the United States ofAmerica 107,16066-16071,doi:10.1073/pnas.1010662107(2010).

[1118] 81.Seripa,D.et al.The missing ApoE allele.Annals of human genetics 71,496-500,doi:10.1111/j.1469-1809.2006.00344.x(2007).

[1119] 82.Kleinstiver,B.P.et al.High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects.Nature 529,490-495,doi:10.1038/nature16526(2016).

[1120] 83.Richardson,C.D.,Ray,G.J.,DeWitt,M.A.,Curie,G.L.&Corn,J.E.Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-

Cas9 using asymmetric donor DNA.Nat Biotech 34,339-344,doi:10.1038/nbt.3481 (2016).

[1121] 84.Simonelli,V.,Narciso,L.,Dogliotti,E.&Fortini,P.Base excision repair intermediates are mutagenic in mammalian cells.Nucleic acids research 33,4404-4411,doi:10.1093/nar/gki749(2005).

[1122] 85.Barnes,D.E.&Lindahl,T.Repair and Genetic Consequences of Endogenous DNA Base Damage in Mammalian Cells.Annual Review of Genetics 38, 445-476,doi:doi:10.1146/annurev.genet.38.072902.092448(2004).

[1123] 实施例16:通过蛋白质工程和蛋白质递送改善碱基编辑的DNA特异性和适用性

[1124] 开发了碱基编辑,一种基因组编辑方法,其使得能够将一个碱基对可编程转化为另一个碱基对而没有双链DNA切割,过量的随机插入和缺失或对同源性指导的修复的依赖性。脱靶活性和对细胞内DNA传递的依赖限制了碱基编辑的应用。本文描述了解决这些局限性的两个进展。首先,通过将突变安装到第三代碱基编辑器(BE3)中以生成高保真性碱基编辑器(HF-BE3),减少了脱靶碱基编辑。接下来,将BE3和HF-BE3纯化并作为核糖核蛋白(RNP)复合物递送到哺乳动物细胞中,建立无DNA的碱基编辑。甚至与HF-BE3质粒转染相比,BE3的RNP递送赋予更高的特异性,而维持相当的中靶编辑水平。最后,将这些进展应用于将BE3 RNP传递到斑马鱼胚胎和活小鼠的内耳中,以体内实现特异性的无DNA的碱基编辑。

[1125] 导言

[1126] 传统的基因组编辑剂引入双链DNA断裂(DSB)作为基因组编辑的第一步[1-4]。细胞对DSB的响应主要通过非同源末端连接(NHEJ),导致切割位点处的随机插入或缺失(插入/缺失)[1,5]。为了在基因组DNA中产生更精确的变化,可以使用同源性指导的修复(HDR)将切割位点周围的基因组DNA替换为外源提供的DNA供体模板的[6-8]。不幸的是,HDR通常伴随着竞争性NHEJ导致的过多插入/缺失,并且主要限于有丝分裂细胞。此外,大多数基因组编辑方法都依赖于将外源质粒或病毒DNA递送至哺乳动物细胞中,然后在细胞内表达该试剂[9-12]。即使在中靶基因座已被编辑后,这些传递方法仍导致连续的,不受控制的Cas9和sgRNA表达,从而增加了在脱靶基因座处进行基因组编辑的机会[1,13]。

[1127] 最近描述了碱基编辑,一种用于基因组编辑的不同方法,其使得能够将C:G碱基对直接可编程靶向性转化为T:A碱基对[3,14]。第三代碱基编辑器BE3在单一蛋白质中含有(i)催化受损的Cas9,其在引导RNA规定的基因座处打开小的单链DNA气泡,(ii)栓系的单链特异性胞苷脱氨酶,其在单链DNA气泡中约5个核苷酸的窗内将C转化为U,(iii)尿嘧啶糖基化酶抑制剂(UGI),其抑制碱基切除修复,从而改善碱基编辑的效率和产物选择性,以及(iv)切口酶活性,以操作细胞错配修复以替换含G的DNA链。这些成分的组合使得哺乳动物细胞中有效且永久的C至T(或G至A)转化,具有最小的插入/缺失形成。自先前报告[14]以来,其他研究人员已经证实了这种策略和相关方法在哺乳动物细胞[15-17]和植物[18]中促进Cas9指导的C至T转化的能力。

[1128] 本文描述了两项进步,它们大大改善碱基编辑的DNA特异性,并允许在不提供外源DNA的情况下在体外和体内进行碱基编辑,所述外源DNA与同宿主基因组重组的风险和细胞毒性有关[18,19]。首先,工程化改造BE3突变体形式,该突变体掺入已知降低Cas9的DNA亲和力的突变[20],该突变减少脱靶编辑事件,且在中靶编辑活性中仅具有适度降低。接下来,揭示了

与质粒DNA递送相比，脂质介导的与引导RNA复合的碱基编辑器蛋白的递送导致甚至更大的特异性增强，在中靶碱基编辑中没有明显减少。作为RNP的碱基编辑器递送在培养的人和小鼠细胞中通常将脱靶编辑降低到可测量水平以下，即使对于靶向高度重复的基因组DNA序列的众所周知泛宿主性的引导RNA。这些进展使得在小鼠和斑马鱼中证明高特异性无DNA的体内碱基编辑。

[1129] 结果

[1130] 工程化改造高保真性碱基编辑器

[1131] 已经显示Cas9核酸酶及其相关的融合构建体在脱靶基因组基因座处结合并切割DNA[21-24]。Joung其同事开发了HF-Cas9，一种高保真性SpCas9变体，其包含设计为消除Cas9与DNA靶链的磷酸主链之间的非特异性相互作用的四个点突变(N497A，R661A，Q695A，Q926A)(图125A)[20]，与先前消除TALEN中非特异性DNA相互作用一致，该消除大大提高其DNA切割特异性[25]。由于碱基编辑器对由Cas9创建的单链DNA气泡内的非靶链起作用[14]，可以假设将来自HF-Cas9的这四个点突变引入BE3中以生成“HF-BE3”可以减少脱靶碱基编辑而不改变其碱基转换能力(图125A和125C)。

[1132] 在大肠杆菌中过表达编码作为带His_6标签的蛋白质的BE3和HF-BE3的质粒，并首先通过镍亲和层析，然后通过阳离子交换层析纯化(图126A-126B)。在广泛优化表达和纯化条件后，可以以每升培养基约2mg的产率常规生产BE3和HF-BE3蛋白(图126A-126C)。

[1133] 使用纯化的碱基编辑器蛋白在生化上比较HF-BE3和BE3的碱基编辑效率和编辑窗的宽度。在具有由TC重复构成的原间隔物的合成dsDNA 79聚体中体外测量C至U转化效率。将靶dsDNA(250nM)与BE3:sgRNA或HF-BE3:sgRNA(2μM)在37℃下温育30分钟。温育后，使用耐尿嘧啶的聚合酶扩增编辑的DNA，并通过高通量DNA测序(HTS)进行测序。在体外观察到HF-BE3和BE3的相当的编辑效率和活性窗宽度(图125B)。这些发现指示将高保真性突变引入BE3中不损害固有的中靶碱基编辑效率或者在体外不改变所得HF-BE3蛋白的编辑窗的宽度。

[1134] HF-BE3增强DNA转染后的编辑特异性

[1135] 接下来，在质粒DNA转染后在哺乳动物细胞中比较BE3和HF-BE3的碱基编辑效率，特异性和编辑窗宽度。选择了四个经过充分研究的内源基因组基因座(HEK293位点3，FANCF，EMX1和VEGFA位点2)来询问哺乳动物细胞中的中靶和脱靶碱基编辑[14,24]。VEGFA位点2是高度重复的，在20个原间隔物核苷酸中含有14个C，并且与已知脱靶基因组编辑的众所周知高比率相关[20,22,24,26]。选择包括此位点，因为它构成了巨大的特异性挑战。与大多数基于核酸酶的基因组编辑应用相反，碱基编辑依赖于原间隔物的精确位置，以将靶核苷酸置于编辑窗内，通常在选择引导RNA时几乎没有灵活性或没有灵活性。因此，即使对于高度重复的混杂的sgRNA靶物，具有增强的特异性的碱基编辑器的开发也是至关重要的[3,14]。

[1136] 用四个碱基编辑器:sgRNA对中的每个进行质粒转染[24]后，通过PCR扩增中靶基因座和已知脱靶基因座，并通过HTS分析。通过引入HF突变，HEK293T细胞中这四个内源基因组基因座的中靶编辑略有减少；用BE3的编辑平均为29±5%，对于HF-BE3为21±3%(对于n=3个生物学重复，平均值±标准差)(图127A-127D，128A)。

[1137] 对于三个标准的非重复性靶位点(HEK293位点3，FANCF和EMX1)的每个，检查了三个最频繁修饰的脱靶基因座，其含有来自先前报告从用Cas9和相同引导RNA处理修饰的脱

靶基因座的编辑窗内的C(表15)[24]。当用BE3质粒转染细胞时,HEK293位点3,FANCF和EMX1的9个最频繁修饰的Cas9脱靶基因座间的C→T转化平均为1.1±0.3%(图127A-C;对于n=3个生物学重复的平均值±s.d.)。HF突变的安装使平均脱靶编辑的绝对水平降低37倍,至0.03±0.005%,可测量的脱靶C→T转化仅有1个情况(图127A;脱靶1处的EMX1 C_5)。

[1138] 位

[1139] 点

[1140] 序列

[1141]

SEQ ID NO	GUIDE-Seq 计数

[1142]

EMX1 中靶	GAGTC$_5$C$_6$GAGCAGAAGA AGAA**GGG**	480	4,521
EMX1 脱靶 1	GAGTC$_5$TAAGCAGAAGA AGAA**GAG**	481	1,445
EMX1 脱靶 2	GAGGC$_5$C$_6$GAGCAGAAG AAAGA**CGG**	482	700
EMX1 脱靶 3	GAGTC$_5$C$_6$TAGCAGGAGA AGAA**GAG**	483	390
HEK293 位点 3 中靶	GGCC$_4$C$_5$AGACTGAGCAC GTGA**TGG**	484	2,074
HEK293 位点 3 脱靶 1	CACC$_4$C$_5$AGACTGAGCAC GTGC**TGG**	485	327
HEK293 位点 3 脱靶 2	GACAC$_5$AGACTGGGCAC GTGA**GGG**	486	306
HEK293 位点 3 脱靶 3	AGCTC$_5$AGACTGAGCAA GTGA**GGG**	487	136
VEGFA 位点 2 中靶	GAC$_3$C$_4$C$_5$C$_6$C$_7$TC$_9$C$_{10}$ACCC CGCCTC**CGG**	488	540
VEGFA 位点 2 脱靶 1	CTAC$_4$C$_5$C$_6$C$_7$TC$_9$C$_{10}$ACCC CGCCTC**CGG**	489	1,925
VEGFA 位点 2 脱靶 2	ATTC$_4$C$_5$C$_6$C$_7$C$_8$C$_9$C$_{10}$ACCC CGCCTC**AGG**	490	1,549
VEGFA 位点 2 脱靶 3	ACAC$_4$C$_5$C$_6$C$_7$C$_8$C$_9$C$_{10}$ACC CCGCCTC**AGG**	491	1,178
VEGFA 位点 2 脱靶 4	TGC$_3$C$_4$C$_5$C$_6$C$_7$C$_8$C$_9$C$_{10}$ACC CCACCTC**TGG**	492	1,107
FANCF 中靶	GGAATC$_6$C$_7$C$_8$TTC$_{11}$TGCA GCACC**TGG**	493	4,816
FANCF 脱靶 1	GGAAC$_5$C$_6$C$_7$C$_8$GTC$_{11}$TGC AGCACC**AGG**	494	2,099
FANCF 脱靶 2	GGAGTC$_6$C$_7$C$_8$TC$_{10}$C$_{11}$TAC AGCACC**AGG**	495	524

[1143]

FANCF 脱靶 3	AGAGGC$_6$C$_7$C$_8$C$_9$TC$_{11}$TGC AGCACC**AGG**	496	150

[1144] 表15:这项工作中研究的中靶和脱靶人基因组基因座的原间隔物和PAM序列。脱靶

位点是根据其GUIDE-Seq读段计数[45]选择的。对特定sgRNA的编辑窗中的胞嘧啶进行编号。PAM序列以粗体显示。已经对脱靶基因座中的原间隔物碱基加下划线，所述脱靶基因座不同于它们相应的中靶基因座。对于在鼠类样品中询问的基因组序列，参见图132E。

[1145] 为了表征在极富挑战性的位点上的HF-BE3特异性，比较了当靶向高度重复的VEGFA位点2基因座时BE3和HF-BE3脱靶活性。BE3处理导致位于与该sgRNA相关的四个测试的脱靶位点的活性窗中的胞嘧啶的平均编辑15±5％(本段中引用的所有平均值代表n＝3个生物学重复的平均值±s.d.)。相比之下，HF-BE3在相同的脱靶位点处导致绝对脱靶编辑降低3倍(5.0±2.3％)(图127D)。当与BE3转染相比时，HF-BE3显著($p<0.05$，两尾学生t检验)减少了位于脱靶基因座处的57个胞嘧啶的27处的脱靶编辑(表16)，而HF-BE3处理导致16个询问中靶胞嘧啶残基的仅仅3处的中靶编辑的显著降低($p<0.05$，两尾Student t检验)

[1146] 表.16

[1147]

基因座和胞嘧啶位置	P值(student两尾t检验)以在列出的处理间比较									
	质粒 BE3 对 质粒 HF-BE3	质粒 BE3 对 蛋白 BE3	质粒 BE3 对 蛋白 HF-BE3	质粒 BE3 对 对照	质粒 HF-BE3对 蛋白 BE3	质粒 HF-BE3对 蛋白 HF-BE3	质粒 HF-BE3 对 对照	蛋白质 BE3 对 蛋白 HF-BE3	蛋白质 BE3 对 对照	蛋白质 HF-BE3 对 对照
EMX1, C5	0.053	0.000	0.001	0.000	0.416	0.056	0.003	0.023	0.000	0.004
EMX1, C6	0.065	0.000	0.000	0.000	0.445	0.023	0.004	0.001	0.000	0.003
FANCF, C6	0.152	0.017	0.003	0.000	0.137	0.003	0.000	0.002	0.000	0.004
FANCF, C7	0.591	0.554	0.007	0.000	0.914	0.007	0.000	0.008	0.000	0.006
FANCF, C8	0.011	0.026	0.004	0.000	0.958	0.018	0.000	0.023	0.000	0.007
FANCF, C11	0.524	0.948	0.019	0.001	0.363	0.010	0.000	0.010	0.000	0.021
HEK 位点3, C3	0.061	0.001	0.071	0.000	0.002	0.002	0.005	0.00199	0.001	0.003
HEK 位点3, C4	0.048	0.924	0.010	0.004	0.001	0.001	0.001	0.00004	0.000	0.001
HEK 位点3, C5	0.291	0.592	0.016	0.006	0.243	0.243	0.002	0.00022	0.000	0.001
VEGFA 位点2, C3	0.060	0.416	0.239	0.010	0.042	0.280	0.002	0.475	0.002	0.018
VEGFA 位点2, C4	0.036	0.191	0.047	0.004	0.032	0.803	0.002	0.066	0.000	0.005
VEGFA 位点2, C5	0.098	0.650	0.028	0.003	0.044	0.169	0.002	0.004	0.000	0.001
VEGFA 位点2, C6	0.452	0.781	0.118	0.004	0.165	0.239	0.002	0.013	0.000	0.001
VEGFA 位点2, C7	0.401	0.683	0.172	0.003	0.120	0.454	0.002	0.026	0.000	0.001
VEGFA 位点2, C9	0.225	0.308	0.254	0.004	0.504	0.828	0.004	0.624	0.000	0.002
VEGFA 位点2, C10	0.064	0.061	0.023	0.005	0.732	0.257	0.009	0.057	0.000	0.003
EMX1, C5 脱靶 1	0.003	0.003	0.002	0.002	0.119	0.036	0.035	0.269	0.255	0.643
EMX1, C5 脱靶 2	0.013	0.013	0.013	0.013	0.158	0.294	0.521	0.390	0.058	0.054
EMX1, C6 脱靶. 2	0.024	0.024	0.024	0.024	0.285	0.560	0.954	0.420	0.103	0.306
EMX1, C5 脱靶 3	0.022	0.019	0.019	0.019	0.297	0.297	0.300	>0.99999	0.882	0.815
EMX1, C6 脱靶 3	0.017	0.015	0.015	0.015	0.296	0.296	0.328	>0.99999	0.051	0.025
FANCF, C5 脱靶 1	0.031	0.031	0.031	0.031	0.314	0.530	0.333	0.337	0.349	0.618
FANCF, C6 脱靶 1	0.016	0.016	0.016	0.016	0.347	0.786	0.930	0.338	0.344	0.678
FANCF, C7 脱靶 1	0.028	0.028	0.028	0.027	0.374	0.039	0.106	0.353	0.346	0.639
FANCF, C8 脱靶 1	0.014	0.014	0.014	0.014	0.341	0.932	0.685	0.343	0.318	0.605
FANCF, C11 脱靶 1	0.007	0.007	0.007	0.007	0.374	0.001	0.000	>0.99999	0.475	0.016
FANCF, C6 脱靶 2	0.099	0.099	0.032	0.036	0.599	0.475	0.912	0.914	0.393	0.060
FANCF, C7 脱靶 2	0.080	0.080	0.027	0.030	0.898	0.638	0.819	0.539	0.859	0.530
FANCF, C8 脱靶 2	0.123	0.123	0.045	0.050	0.789	0.538	0.960	0.047	0.539	0.252
FANCF, C10 脱靶 2	0.093	0.093	0.029	0.033	0.630	0.509	0.847	0.768	0.670	0.482
FANCF, C11 脱靶 2	0.264	0.264	0.127	0.107	0.599	0.658	0.326	>0.99999	0.047	0.345
FANCF, C6 脱靶 3	0.872	0.872	0.492	0.108	0.239	0.493	0.129	0.469	0.584	0.252
FANCF, C7 脱靶 3	>0.99999	>0.99999	0.859	0.016	0.537	0.866	0.116	0.572	0.272	0.595
FANCF, C8 脱靶 3	0.886	0.886	0.246	0.757	>0.99999	0.001	0.648	0.495	0.780	0.650
FANCF, C10 脱靶 3	0.566	0.566	0.284	0.202	0.053	0.387	0.260	0.453	0.913	0.541
FANCF, C11 脱靶 3	0.422	0.422	0.145	0.145	0.495	0.230	0.230	0.658	0.658	>0.99999
HEK293 位点3, C3 脱靶 1	>0.99999	0.910	0.412	0.326	0.910	0.412	0.326	0.293	0.223	0.480
HEK293 位点3, C4 脱靶 1	>0.99999	0.994	0.437	0.391	0.994	0.437	0.391	0.495	0.451	0.614
HEK293 位点3, C5 脱靶 1	>0.99999	0.616	0.814	0.337	0.616	0.814	0.337	0.459	0.116	0.481
HEK293 位点3, C3 脱靶 2	0.285	0.473	0.100	0.473	0.141	0.735	0.141	0.038	>0.99999	0.038
HEK293 位点3, C5 脱靶 2	0.375	0.294	0.177	0.294	0.687	0.428	0.687	0.064	>0.99999	0.064
HEK293 位点3, C9 脱靶 2	0.053	0.624	0.374	0.624	0.554	0.154	0.554	0.872	>0.99999	0.872
HEK293 位点3, C3 脱靶 3	0.067	0.116	0.768	0.435	0.519	0.230	0.561	0.349	0.768	0.643
HEK293 位点3, C5 脱靶 3	0.011	0.011	0.011	0.011	0.016	0.643	0.435	0.184	0.025	0.346
HEK293 位点3, C9 脱靶 3	>0.99999	0.374	0.652	0.811	0.132	0.539	0.776	0.609	0.643	0.893
VEGFA 位点2, C4 脱靶 1	0.101	0.015	0.014	0.012	0.041	0.032	0.025	0.117	0.012	0.001
VEGFA 位点2, C5 脱靶 1	0.060	0.013	0.012	0.010	0.078	0.062	0.044	0.201	0.009	0.012
VEGFA 位点2, C6 脱靶. 1	0.019	0.005	0.005	0.004	0.080	0.062	0.045	0.087	0.002	0.012
VEGFA 位点2, C7 脱靶 1	0.017	0.004	0.004	0.003	0.080	0.060	0.037	0.076	0.001	0.002
VEGFA 位点2, C9 脱靶 1	0.230	0.088	0.037	0.011	0.667	0.256	0.051	0.134	0.004	0.007
VEGFA 位点2, C10 脱靶 1	0.535	0.136	0.106	0.035	0.283	0.211	0.050	0.717	0.028	0.010
VEGFA 位点2, C4 脱靶 2	0.038	0.004	0.003	0.003	0.087	0.051	0.048	0.063	0.048	0.134
VEGFA 位点2, C5 脱靶 2	0.033	0.004	0.004	0.004	0.078	0.061	0.059	0.028	0.020	0.248
VEGFA 位点2, C6 脱靶 2	0.026	0.005	0.005	0.004	0.051	0.038	0.038	0.043	0.038	0.783
VEGFA 位点2, C7 脱靶 2	0.053	0.006	0.005	0.005	0.072	0.056	0.055	0.078	0.064	0.704
VEGFA 位点2, C8 脱靶 2	0.071	0.006	0.006	0.006	0.079	0.065	0.065	0.118	0.107	0.703
VEGFA 位点2, C9 脱靶 2	0.193	0.008	0.007	0.006	0.103	0.090	0.084	0.068	0.007	0.217
VEGFA 位点2, C10 脱靶 2	0.063	0.003	0.003	0.002	0.116	0.107	0.090	0.545	0.016	0.346
VEGFA 位点2, C4 脱靶 3	0.005	0.003	0.003	0.003	0.091	0.031	0.030	0.116	0.107	0.158
VEGFA 位点2, C5 脱靶 3	0.011	0.007	0.005	0.005	0.211	0.048	0.042	0.220	0.177	0.001
VEGFA 位点2, C6 脱靶 3	0.020	0.005	0.003	0.003	0.142	0.038	0.033	0.193	0.149	0.015
VEGFA 位点2, C7 脱靶 3	0.045	0.006	0.003	0.003	0.101	0.035	0.030	0.093	0.060	0.083
VEGFA 位点2, C8 脱靶 3	0.069	0.007	0.005	0.005	0.087	0.045	0.039	0.120	0.067	0.041
VEGFA 位点2, C9 脱靶 3	0.093	0.006	0.005	0.005	0.041	0.032	0.028	0.396	0.195	0.005
VEGFA 位点2, C10 脱靶 3	0.342	0.011	0.008	0.007	0.109	0.081	0.069	0.273	0.098	0.036
VEGFA 位点2, C3 脱靶 4	0.001	0.001	0.001	0.001	0.374	0.374	0.230	0.271	0.358	0.633
VEGFA 位点2, C4 脱靶 4	0.007	0.006	0.006	0.006	0.137	0.137	0.137	0.592	0.862	0.690
VEGFA 位点2, C5 脱靶 4	0.007	0.007	0.007	0.007	0.026	0.017	0.018	0.461	0.655	0.279
VEGFA 位点2, C6 脱靶 4	0.005	0.004	0.004	0.004	0.021	0.018	0.018	0.398	0.546	0.149
VEGFA 位点2, C7 脱靶 4	0.007	0.006	0.006	0.006	0.051	0.048	0.050	0.373	0.720	0.029
VEGFA 位点2, C8 脱靶 4	0.007	0.006	0.006	0.006	0.092	0.092	0.092	0.325	0.014	0.275
VEGFA 位点2, C9 脱靶 4	0.016	0.007	0.007	0.007	0.150	0.150	0.150	0.502	1.000	0.615
VEGFA 位点2, C10 脱靶 4	0.213	0.009	0.009	0.009	0.261	0.261	0.261	0.653	0.575	0.660

[1148] 表16:本研究中评估的所有基因座处在不同处理条件下碱基编辑差异的P值。如材料和方法中所述,使用学生的两尾t检验计算p值。当p值指示显著差异($p<0.05$)时,突出显示了相应的条目。

[1149] 除了考虑在脱靶基因座处的绝对编辑之间的差异之外,还通过将观察到的中靶效率除以脱靶效率来计算中靶:脱靶编辑特异性比率(图129A-129B)。此度量考虑到与HF突变

的安装相关的中靶编辑的任何减少，并且对于对碱基编辑的效率和特异性两者敏感的应用是有用的。通过HF-BE3进行的脱靶编辑低于几个脱靶基因座的高通量测序的检测限。对于这些情况，假定保守的脱靶编辑效率等于检测上限(0.025％C→T转化;参见方法)。基于该分析，当进行两种构建体的质粒递送时，在本文所检查的所有34个靶胞嘧啶间安装HF突变后，特异性比率的平均改善为19倍。这些结果共同建立了对于非重复位点(图129A)以及高度重复位点(图129B)，与BE3相比，HF-BE3导致碱基编辑特异性大大增强，中靶编辑效率仅适度降低。

[1150] BE3的RNP递送实现无DNA碱基编辑

[1151] 接下来，研究了无DNA的RNP形式的BE3在直接递送至培养的人细胞中时介导碱基编辑的能力。最近已建立阳离子脂质试剂可将带负电荷的蛋白质或蛋白质:核酸复合物有力地递送到包含核糖核蛋白(RNP)复合物的哺乳动物细胞中，并且RNP递送可实质性减少脱靶基因组编辑[27-29]。

[1152] 在与靶向EMX1，HEK293位点3，FANCF或VEGFA位点2基因座的引导RNA预复合后，将市售的阳离子脂质Lipofectamine 2000与纯化的BE3蛋白或HF-BE3蛋白组合，并将得到的脂质:RNP复合物与HEK293T细胞一起温育。72h后，收获基因组DNA，并通过高通量DNA测序分析中靶和脱靶碱基编辑。与所有基于Cas9的技术一样，在不同的基因组基因座处的编辑效率中也观察到实质性变化(图127和130)。为了显示与不同处理之间的中靶编辑效率相关的趋势，在四个测试基因座处计算平均中靶碱基编辑效率(图128A)。BE3(200nM)的蛋白质递送导致与用质粒转染观察到的情况相当的中靶编辑效率(分别为26±4％对29±5％;n＝3个生物学重复的平均值±s.e.m.;图128A)。

[1153] 相比之下，与在所研究的四个基因组基因座处的BE3的蛋白质递送相比，HF-BE3的蛋白质递送减少中靶编辑(平均编辑效率分别为13±3％对26±4％;对于n＝3个生物学重复的平均值±s.e.m.;图128A)。由于HF-BE3和BE3在试管中具有相当的编辑效率(图125B)，并且当从HEK293T细胞中的质粒表达HF-BE3时，编辑仅略有降低(图127A-D)，因此吸引人推测来自HF-BE3蛋白递送的编辑效率降低可以是哺乳动物细胞中HF-BE3稳定性降低的结果。较低的稳定性可以由从质粒的连续表达抵消，而非在一次蛋白质递送后发生。此观察结果与最近的下述报告一致:当核转染到$CD34^+$造血干细胞和祖细胞中时，与纯化的Cas9相比用纯化的HF-Cas9的中靶插入/缺失形成减少[30]。在审查这项工作的情况下，Kim等人证明了使用电穿孔将BE3以RNP递送至小鼠胚胎中[31]。就本发明人所知，本方法是能够在不进行电穿孔的情况下对哺乳动物细胞中的个别核苷酸产生精确变化的第一种无DNA技术，其具有有限的体内治疗相关性。

[1154] 碱基编辑器的RNP递送大大提高了DNA的特异性

[1155] 重要的是，尽管BE3和HF-BE3的RNP传递导致了实质性的中靶碱基编辑，但在与EMX1，FANCF和HEK293位点3相关的9个测试的脱靶基因座中的任一处没有观察到可测量的碱基编辑的情况(<0.025％)(图130A-130C)。相比之下，BE3的质粒递送导致在碱基编辑活性窗内所有测序的胞嘧啶间的1.1±0.3％(n＝3个生物学重复的平均值±s.d.)脱靶编辑的平均值，以及在位于这9个脱靶基因座处的16个脱靶胞嘧啶的11处可检测的脱靶编辑(图127A-127D)。在所测试的三个非重复基因座的脱靶基因座处，BE3蛋白递送导致比质粒递送的平均特异性比率高26倍的平均特异性比率(图127A)。这些结果揭示了碱基编辑器的RNP

递送显著提高了碱基编辑的DNA特异性。

[1156] 与BE3或HF-BE3的质粒递送相比，BE3或HF-BE3的蛋白质递送还导致在高度混杂的VEGFA位点2基因座处极大改善的碱基编辑特异性(比较图127和130；参见表16)。对于BE3(质粒递送：15±4％脱靶编辑；蛋白质递送：1.3±0.4％脱靶编辑；本段中的所有数值代表n＝3个生物学重复的平均值±s.d.)和HF-BE3(质粒递送：5±2％脱靶编辑；蛋白质递送：0.5±0.1％脱靶编辑)两者，与此位点相关的脱靶基因座处的碱基编辑的绝对频率在蛋白质递送后降低至少10倍。在所有四个研究的基因座中，与BE3质粒质粒相比，对于BE3蛋白质递送，中靶:脱靶编辑修饰的碱基编辑特异性比率增加66倍的平均值(图129)。这些结果共同揭示了对于重复的和非重复的靶位点两者，RNP对DNA递送是比高保真性Cas9突变的存在或不存在更强的碱基编辑特异性的决定因素。

[1157] HF突变的引入和递送方法都没有实质性改变与碱基编辑相关的低插入/缺失率。在这项研究中所有处理条件间所有中靶基因座处的插入/缺失频率保持较低(通常≤5％；图131A)，并且在所有测试情况下编辑:插入/缺失率保持比先前使用优化的HDR方案进行的研究中高(通常≥10倍；图131B)[30,32,33]。对于非重复性sgRNA，在脱靶基因座处观察到很少的插入/缺失(图131C)，尽管已注意到BE3的质粒递送对于与VEGFA位点2相关的脱靶基因座产生高达5％的插入/缺失(图131C)。

[1158] 这些结果共同建立了碱基编辑器的蛋白质递送维持中靶碱基编辑效率并且相对于质粒DNA的递送大大增强编辑特异性。

[1159] RNP递送解偶联中靶和脱靶编辑

[1160] 鉴于与BE3的蛋白质递送相关的碱基编辑特异性的显著增强，研究了此种改善是否是递送到细胞中的活性基因组编辑剂的总量减少的结果。使用靶向EMX1的sgRNA，进行了质粒(图128B)和蛋白质递送(图128C)的剂量响应研究。为了使处理条件之间的转染效率最大化，对于所有测试，Lipofectamine2000的体积为1.5μL，并且对于蛋白质递送，碱基编辑器蛋白质:sgRNA摩尔比维持为1:1.1。对于质粒递送，使用1:3(摩尔比约1:1)的sgRNA质粒:BE3质粒的质量比和1.5μL的Lipofectamine 2000。在所有测试的质粒递送条件下观察到脱靶碱基编辑(图128B)，但在所有测试的蛋白质递送条件下实际上未观察到脱靶编辑(图128C)。

[1161] 进行线性回归分析以评估质粒和蛋白质递送的中靶和脱靶编辑之间的关系。对于质粒递送，脱靶编辑与中靶编辑率密切相关(R^2＝0.95，p＝0.0012，对于非零斜率，F检验)，而使用蛋白质递送的脱靶和中靶编辑之间没有显著关联(R^2＝0.078，p＝0.59，对于非零斜率，F检验)。

[1162] 这些数据指示碱基编辑器的蛋白质递送提供了不依赖于剂量的固有的特异性优势。与先前的观察结果[29,34]一起，这些发现支持了以下的模型，在该模型中相比于DNA传递来自蛋白质传递的碱基编辑的更高的DNA特异性源自蛋白质传递避免基因组长时间暴露于碱基编辑器的能力，从而使已经修饰中靶基因座后碱基编辑器加工脱靶基因座的机会最小化。

[1163] 斑马鱼和小鼠中的无DNA碱基编辑

[1164] 以上观察结果提示了BE3的蛋白质递送有望维持中靶碱基编辑，而消除可检测的脱靶碱基编辑。因此，通过将靶向酪氨酸酶基因座的BE3:sgRNA复合物注射到受精的斑马鱼

胚胎中测试了BE3的蛋白质传递是否可用于在斑马鱼中产生特定的点突变。注射后4天，从所得的斑马鱼幼体中收获基因组DNA，并通过高通量测序测量碱基编辑和插入/缺失频率(图132A)。测试的三个BE3:sgRNA复合物中的两个在体内诱导实质性的点突变(TYR1:C_3→$T_3$5.3±1.8%,TYR2:C_4→T_4 4.3±2.1%;n=3个注射胚胎的平均值±s.d.;图132A)。斑马鱼基因座的序列列于表17。

[1165]

HEK 细胞样品

样品描述	重复 1	重复 2	重复 3		扩增子	
蛋白质，BE3	237256	295609	159391	VEGFA	中靶	
蛋白质，HF-BE3	383480	389874	383467	VEGFA	中靶	
质粒，BE3	315213	280891	335668	VEGFA	中靶	
质粒，HF-BE3	196323	251965	369201	VEGFA	中靶	
对照	390748	395523	353614	VEGFA	中靶	
蛋白质，BE3	19280	26472	24799	FANCF	脱靶位点 #1	
蛋白质，HF-BE3	36383	30007	39193	FANCF	脱靶位点 #1	
质粒，BE3	35580	29557	22243	FANCF	脱靶位点 #1	
质粒，HF-BE3	40371	27187	28248	FANCF	脱靶位点 #1	
对照	35106	37274	34939	FANCF	脱靶位点 #1	
蛋白质，BE3	82978	124689	83840	VEGFA	脱靶位点 #1	
蛋白质，HF-BE3	142404	140482	142220	VEGFA	脱靶位点 #1	
质粒，BE3	112027	117071	100894	VEGFA	脱靶位点 #1	
质粒，HF-BE3	114187	98876	122553	VEGFA	脱靶位点 #1	
对照	76854	90547	89271	VEGFA	脱靶位点 #1	
蛋白质，BE3	14514	25515	19325	VEGFA	脱靶位点 #2	
蛋白质，HF-BE3	24678	24363	25312	VEGFA	脱靶位点 #2	
质粒，BE3	16945	19918	10225	VEGFA	脱靶位点 #2	
质粒，HF-BE3	12200	14769	17797	VEGFA	脱靶位点 #2	
对照	8739	13648	7818	VEGFA	脱靶位点 #2	
蛋白质，BE3	17924	11169	13909	VEGFA	脱靶位点 #3	
蛋白质，HF-BE3	24389	30050	27613	VEGFA	脱靶位点 #3	
质粒，BE3	20847	29137	15543	VEGFA	脱靶位点 #3	
质粒，HF-BE3	11717	15403	19915	VEGFA	脱靶位点 #3	
对照	11926	17043	12168	VEGFA	脱靶位点 #3	
蛋白质，BE3	23779	26294	18537	VEGFA	脱靶位点 #4	
蛋白质，HF-BE3	31325	23369	24492	VEGFA	脱靶位点 #4	
质粒，BE3	24309	23031	23442	VEGFA	脱靶位点 #4	
质粒，HF-BE3	17095	16009	14069	VEGFA	脱靶位点 #4	
对照	15869	14872	13727	VEGFA	脱靶位点 #4	
蛋白质，BE3	2868	2823	4331	HEK3	中靶	
蛋白质，HF-BE3	4930	4257	5769	HEK3	中靶	
质粒，BE3	5500	5581	5431	HEK3	中靶	
质粒，HF-BE3	5519	1138	765	HEK3	中靶	
对照	6374	4287	4852	HEK3	中靶	
蛋白质，BE3	10482	18179	16109	HEK3	脱靶位点 # 1	
蛋白质，HF-BE3	20458	17556	17730	HEK3	脱靶位点 # 1	
质粒，BE3	17858	15226	20686	HEK3	脱靶位点 # 1	
质粒，HF-BE3	19129	13842	16078	HEK3	脱靶位点 # 1	
对照	19030	19006	19051	HEK3	脱靶位点 # 1	
蛋白质，BE3	14608	9511	13501	HEK3	脱靶位点 # 2	
蛋白质，HF-BE3	15594	13654	15799	HEK3	脱靶位点 # 2	
质粒，BE3	15003	12843	15890	HEK3	脱靶位点 # 2	
质粒，HF-BE3	37107	12364	14256	HEK3	脱靶位点 # 2	
对照	13032	13454	14183	HEK3	脱靶位点 # 2	
蛋白质，BE3	14505	17533	16183	HEK3	脱靶位点 # 3	
蛋白质，HF-BE3	21233	17899	17988	HEK3	脱靶位点 # 3	
质粒，BE3	18645	16650	8044	HEK3	脱靶位点 # 3	
质粒，HF-BE3	16373	11845	13471	HEK3	脱靶位点 # 3	
对照	13146	13447	15560	HEK3	脱靶位点 # 3	
蛋白质，BE3	41986	61678	67890	FANCF	中靶	
蛋白质，HF-BE3	41057	55850	86411	FANCF	中靶	
质粒，BE3	39114	48575	70074	FANCF	中靶	
质粒，HF-BE3	41617	55638	75718	FANCF	中靶	
对照	68852	59422	81265	FANCF	中靶	
蛋白质，BE3	113462	80529	191344	FANCF	脱靶位点 # 1	
蛋白质，HF-BE3	202662	233981	203024	FANCF	脱靶位点 # 1	
质粒，BE3	208912	202044	107234	FANCF	脱靶位点 # 1	
质粒，HF-BE3	86494	113989	86807	FANCF	脱靶位点 # 1	
对照	92255	72386	56661	FANCF	脱靶位点 # 1	
蛋白质，BE3	96271	117442	84374	FANCF	脱靶位点 # 2	
蛋白质，HF-BE3	105624	102312	105343	FANCF	脱靶位点 # 2	
质粒，BE3	101002	98747	70052	FANCF	脱靶位点 # 2	
质粒，HF-BE3	30896	69787	83184	FANCF	脱靶位点 # 2	
对照	99986	100344	100659	FANCF	脱靶位点 # 2	
蛋白质，BE3	25524	18245	15388	FANCF	脱靶位点 # 3	
蛋白质，HF-BE3	71858	75553	71785	FANCF	脱靶位点 # 3	
质粒，BE3	60980	57360	78169	FANCF	脱靶位点 # 3	
质粒，HF-BE3	68316	34659	85718	FANCF	脱靶位点 # 3	
对照	49685	57388	60418	FANCF	脱靶位点 # 3	
蛋白质，BE3	46981	90793	79439	EMX1	中靶	
蛋白质，HF-BE3	58629	71186	61575	EMX1	中靶	
质粒，BE3	70817	82736	75706	EMX1	中靶	
质粒，HF-BE3	77038	71123	78511	EMX1	中靶	
对照	62183	48574	68439	EMX1	中靶	
蛋白质，BE3	165905	257565	142888	EMX 1	脱靶位点 # 1	
蛋白质，HF-BE3	148339	151300	130712	EMX 1	脱靶位点 # 1	
质粒，BE3	101950	103226	203004	EMX 1	脱靶位点 # 1	
质粒，HF-BE3	167969	175193	97010	EMX 1	脱靶位点 # 1	
对照	101476	150435	102327	EMX 1	脱靶位点 # 1	
蛋白质，BE3	136738	213438	118711	EMX 1	脱靶位点 # 2	
蛋白质，HF-BE3	123413	126114	109375	EMX 1	脱靶位点 # 2	
质粒，BE3	85576	86600	169592	EMX 1	脱靶位点 # 2	
质粒，HF-BE3	140317	145738	137050	EMX 1	脱靶位点 # 2	
对照	84818	125139	85454	EMX 1	脱靶位点 # 2	
蛋白质，BE3	11940	36593	24946	EMX1	脱靶位点 # 3	
蛋白质，HF-BE3	26762	31566	36377	EMX1	脱靶位点 # 3	
质粒，BE3	32420	21547	14659	EMX1	脱靶位点 # 3	
质粒，HF-BE3	31427	16592	17385	EMX1	脱靶位点 # 3	
对照	17385	28128	32717	EMX1	脱靶位点 # 3	

来自NIH 3T3细胞处理的鼠样品

样品描述	重复 1	重复 2	重复 3		扩增子	
质粒，BE3	16641	102216	46361		中靶	VEGFA (Mus)
质粒，HF-BE3	89330	126545	100993		中靶	VEGFA (Mus)
蛋白质，BE3	88998	81697	51124		中靶	VEGFA (Mus)
蛋白质，HF-BE3	128218	29193	131515		中靶	VEGFA (Mus)
对照	18767	38866	58985		中靶	VEGFA (Mus)
质粒，BE3	174782	167504	182565	CFD	脱靶 1	
质粒，HF-BE3	167120	182520	192389	CFD	脱靶 1	
蛋白质，BE3	230569	212605	138144	CFD	脱靶 1	
蛋白质，HF-BE3	228668	211457	183370	CFD	脱靶 1	
对照	171738	191117	20879	CFD	脱靶 1	
质粒，BE3	206475	227332	206089	CFD	脱靶2	
质粒，HF-BE3	213809	203028	199078	CFD	脱靶2	
蛋白质，BE3	215995	275754	249969	CFD	脱靶2	
蛋白质，HF-BE3	250918	272063	241059	CFD	脱靶2	
对照	193760	175959	246963	CFD	脱靶2	
质粒，BE3	60388	126278	7328	CFD	脱靶3	
质粒，HF-BE3	89045	128508	5178	CFD	脱靶3	
蛋白质，BE3	167195	330046	11163	CFD	脱靶3	
蛋白质，HF-BE3	82120	309352	10393	CFD	脱靶3	
对照	83204	176939	5661	CFD	脱靶3	
质粒，BE3	192846	113709	171078	CFD	脱靶4	
质粒，HF-BE3	205601	151434	188943	CFD	脱靶4	
蛋白质，BE3	218194	181993	208398	CFD	脱靶4	
蛋白质，HF-BE3	211966	148976	186838	CFD	脱靶4	
对照	183933	130318	197476	CFD	脱靶4	

小鼠耳蜗样品

样品描述	重复 1	重复 2	重复 3		扩增子	
血管纹	37889	205706	62091		中靶 (VEGFA)	
科尔蒂器	148447	175004	29075		中靶 (VEGFA)	
蜗轴	182806	181382	61269		中靶(VEGFA)	
未注射的对照	228222	241979	272759		中靶(VEGFA)	
血管纹	44457	244487	244646	CFD	脱靶 1	
科尔蒂器	136335	118318	34747	CFD	脱靶 1	
蜗轴	67176	209543	68699	CFD	脱靶 1	
未注射的对照	343100	342717	379015	CFD	脱靶 1	
血管纹	72962	319883	265793	CFD	脱靶 2	
科尔蒂器	198456	131430	60530	CFD	脱靶 2	
蜗轴	92014	251509	81413	CFD	脱靶 2	
未注射的对照	399138	345965	483920	CFD	脱靶 2	
血管纹	8325	80322	142556	CFD	脱靶 3	
科尔蒂器	81014	45976	1810	CFD	脱靶 3	
蜗轴	9928	75555	11341	CFD	脱靶 3	
未注射的对照	399138	345965	483920	CFD	脱靶 3	
血管纹	232194	397770	554054	CFD	脱靶 4	
科尔蒂器	313472	285302	176872	CFD	脱靶 4	
蜗轴	230105	371399	258142	CFD	脱靶 4	
未注射的对照	524503	637946	624709	CFD	脱靶 4	

斑马鱼样品

	1	2	3	扩增子
经处理的斑马鱼	72355	49498	81061	TYR 1
杂乱的 sgRNA	107919	98502	92429	TYR 1
经处理的斑马鱼	51434	48014	41547	TYR 2
杂乱的 sgRNA	61466	62374	66765	TYR 2
经处理的斑马鱼	6487	57247	75883	TYR 3
杂乱的 sgRNA	64596	71234	75624	TYR 3

小鼠耳蜗样品——用无关的sgRNA处理

样品				
血	科尔蒂器	蜗轴	扩增子	
537459	249767	389274	中靶	(VEGFA)

[1166] 表17:与参考序列比对并且通过材料和方法中所述的质量滤器的HTS读段数目。

[1167] 最后，应用这些开发以在小鼠中实现无DNA的高特异性碱基编辑。为了使体内观察到中靶和脱靶碱基编辑的可能性最大化，使用靶向VEGFA位点2的高度重复性sgRNA；方便地，鼠和人基因组在该靶位点处是相同的。

[1168] 使用培养的鼠NIH/3T3细胞，证实了BE3蛋白递送在该基因座处产生了34±11％的有效的中靶碱基编辑(图133A；该段中的所有编辑百分比代表n＝3生物重复的平均值±s.d.)。切割频率决定因素(CFD)算法[29,34]用于预测与VEGFA位点2sgRNA相关的小鼠基因组中的脱靶基因座(表18)。使用培养的NIH/3T3细胞，可以确认在培养的鼠细胞中，BE3的质粒递送确实修饰了前四个预测的脱靶基因座中的两个(CFD脱靶基因座1，9±5％编辑；和CFD脱靶基因座4，3±2％编辑；图133B-133E)。与来自人细胞的结果一致，BE3的蛋白质递送将脱靶编辑降低到与阴性对照的水平相似的水平(图133C和133E)。CFD脱靶基因座1和4的平均碱基编辑特异性比率从BE3质粒递送的28±13增加到BE3蛋白质递送的≥780±300(数值表示平均值±s.e.m.；n＝3个生物学重复)。

[1169]

位点	序列	SEQ ID NO	CFD 得分	基因座描述
中靶	GACCCCCTCCACCCCG CCTCCGG	497		VEGFA 位点 2
脱靶 1	TCCCCCCTCCACCCCA CCTCCGG	498	0.7857	基因间:mmu-mir-21c-Nrp1/Mir1903
脱靶 2	TGCCCACCTCACCCCG CCTCTGG	499	0.65	内含子:Vipr1
脱靶 3	GCCCCTCCCAACCCCA CCTCTGG	500	0.6323	内含子:Nos1ap
脱靶 4	CACCCCCCTCACCCCG CCTCAGG	501	0.625	基因间:Unc5b-mmu-mir-6408

[1170] 表18：与VEGFA位点2sgRNA相关的小鼠基因组中预测的脱靶基因座的原间隔物和PAM序列。使用CRISPOR [47]计算CFD得分[46]。与中靶序列不同的脱靶原间隔物中的位置是加下划线的。

[1171] 为了在小鼠中建立无DNA的碱基编辑，将BE3:sgRNA复合物与Lipofectamine 2000组合(图132B)，并在P1-P2时向小鼠幼鼠进行耳蜗内注射。注射后3-4天收获经注射的耳蜗组织，并且显微切割成每个耳蜗区域的5-7个样品。同时收获未注射的小鼠的对照耳蜗。从收获的组织中提取基因组DNA，通过qPCR扩增至指数后期，然后进行高通量DNA测序以测量C→T转化。虽然由于不可能检索仅来自暴露于碱基编辑器蛋白的细胞的DNA，不可能量化经处理的细胞之间的碱基编辑效率，但是从耳蜗的三个区域中的组织观察到明确的碱基编辑：科尔蒂器的基底末端、血管纹和蜗轴(图132C-132D)。在经处理的组织样品中未检测到显著的插入/缺失形成(<0.1％插入/缺失；图134B)。

[1172] 含有靶C→T转化的耳蜗细胞的百分比(图132C)显著低于在培养物中经处理的NIH/3T3细胞中观察到的百分比(图133A)，这与基于脂质的蛋白质递送的高度局部化性质和无法仅从暴露于碱基编辑器的细胞中分离DNA的能力一致。尽管如此，局部递送为可及的

应用提供了关键优势,包括控制编辑的细胞类型以及简化制备和施用。

[1173] 最后,分析了耳蜗内注射BE3:sgRNA:脂质复合物后的脱靶编辑。对来自注射有BE3:VEGFA位点2sgRNA:脂质复合物的小鼠的耳蜗组织的基因组DNA中的所有四个预测的脱靶基因座(包括已确认的脱靶位点CFD基因座1和CFD基因座4)的分析揭示了对于测试的任何脱靶基因座,无高于未处理的对照样品中观察到的情况的可检测的C→T转化或插入/缺失形成(图134A)。

[1174] 这些体内碱基编辑结果共同建立了一种无病毒,无DNA的策略,用于以高DNA序列特异性将动物的基因组DNA中的个别核苷酸精确转化。

[1175] 讨论

[1176] 在该研究中开发和实施的策略通过除去或减少脱靶碱基编辑并建立支持体内碱基编辑的无DNA的递送方法,扩展了碱基编辑的实用性和适用性。与相同构建体的质粒递送相比,蛋白质递送改善了人和鼠细胞中碱基编辑的特异性(图137、130和133),并实现了斑马鱼和小鼠耳蜗的特异性碱基编辑(图132)。

[1177] 通过将已知增强Cas9的DNA特异性的突变[20]安装到BE3中产生了高保真性碱基编辑器。据报告,将这些突变安装到Cas9中导致在与非重复sgRNA相关的脱靶基因座,包括本文询问的EMX1基因座处检测不到的插入/缺失(图127A)[20]。虽然在HF-BE3中观察到的特异性增强是实质性的,但是是适度的;当作为质粒DNA递送到哺乳动物细胞中时,HF-BE3在重复和非重复基因座两者处表现出可检测到的脱靶碱基编辑(图127A,127D,133C和133E)。吸引人推测此种特异性增强的差异可以源自下述实情:与Cas9介导的插入/缺失形成不同,碱基编辑不需要DNA切割,但仅必需DNA结合和R环形成[14],并且HF-Cas9的某些增强的特异性可以源自已经结合的脱靶基因座处受损的DNA切割。

[1178] 在减少脱靶碱基编辑的第二次尝试中,证明了碱基编辑器的RNP递送导致中靶编辑和脱靶编辑的解偶联(图128B-128C)。RNP递送消除了非重复位点处的脱靶编辑,而维持与质粒递送相当的中靶编辑(图130A-130C和128A),甚至在高度重复的VEGFA位点2处大大降低脱靶编辑(图130D)。对于通常难以控制或表征细胞剂量的体内编辑应用,碱基编辑器的RNP递送可以是特别有用的。

[1179] 与供体DNA模板的递送偶联的Cas9的RNP递送先前已用于在哺乳动物细胞中进行基于HDR的基因组编辑。然而,这些方法仍然受到低效率,细胞状态依赖性和插入/缺失形成效率的限制,所述效率通常超出期望的HDR结果,特别是点突变校正的[29,30,32,35]。相比之下,无DNA碱基编辑在体内和在细胞中产生了相对于随机插入/缺失而言实质性过多的编辑产物(图132A,134A和134B)。就发明人所知,碱基编辑器的RNP递送代表了第一种对基因组DNA进行特异性和精确修饰而无需外源DNA的策略。

[1180] 方法

[1181] 质粒克隆

[1182] 通过USER克隆产生本研究中的质粒。Phusion U热启动聚合酶(Thermo Fisher)用于安装点突变并从先前报告的构建体36构建蛋白质表达质粒。蛋白质序列在补充信息中列出,用于表达BE3和HF-BE3的质粒可从Addgene获得。

[1183] BE3和HF-BE3的表达和纯化

[1184] 用编码具有His_6 N末端纯化标签的细菌密码子优化的碱基编辑器的质粒转化

BL21 Star(DE3)感受态大肠杆菌细胞。在含有50μg mL^{-1}卡那霉素的Luria-Bertani(LB)肉汤中于37℃将单菌落培养过夜。将细胞以1:200稀释到2L的相同培养基中,并在37℃培养直至OD_{600}=0.70-0.75。将培养物在冰上温育60分钟,并用0.5mM异丙基-β-D-1-硫代吡喃半乳糖苷(IPTG,GoldBio)诱导蛋白质表达。在18℃下摇动持续表达14-16小时。随后的纯化步骤在4℃下进行。通过在6,000g离心20分钟收集细胞并重悬于细胞收集缓冲液(100mM三(羟甲基)-氨基甲烷(Tris)-HCl,pH 8.0,1M NaCl,20%甘油,5mM三(2-羧乙基)膦(TCEP;GoldBio),0.4mM苯基甲烷磺酰氟(PMSF;Sigma Aldrich)和每50mL使用的缓冲液1cOmplete,无EDTA的蛋白酶抑制剂团粒(Roche))中。通过超声处理(总共6分钟,打开3秒,关闭3秒)裂解细胞,并通过以25,000g(20分钟)离心使细胞裂解液澄清。

[1185] 将澄清的裂解物与His-Pur镍氮川乙酸(镍-NTA)树脂(每升培养物1mL树脂,Thermo Fisher)在4℃下旋转情况下一起温育60-90分钟。用20柱体积的细胞收集缓冲液洗涤树脂,然后用洗脱缓冲液((100mM三(羟甲基)-氨基甲烷(Tris)-HCl,pH 8.0,0.5M NaCl,20%甘油,5mM三(2-羧乙基)膦(TCEP;GoldBio),200mM咪唑)洗脱结合的蛋白质。使用Akta Pure FPLC在5mL Hi-Trap HP SP(GE Healthcare)阳离子交换柱上进一步纯化所得的蛋白质级分。使用具有100,000kDa截留(Millipore)的柱以3,000g离心浓缩含有蛋白质的级份,并将浓缩的溶液通过.22μm PVDF膜(Millipore)进行无菌过滤。

[1186] 在无菌过滤之后,用还原剂相容双金鸡宁酸(BCA)测定法(Pierce Biotechnology)对蛋白质进行定量,在液氮中速冻,并以等分试样储存在-80℃。表达的蛋白质序列列在补充说明2中。

[1187] sgRNA的体外转录

[1188] 使用如补充信息中列出的引物通过PCR(Q5 Hot Start MasterMix,New England Biolabs)产生包含期望20bp sgRNA原间隔物上游的T7 RNA聚合酶启动子序列和sgRNA主链的线性DNA片段,并在Minelute柱(Qiagen)上浓缩。sgRNA用HiScribe T7高产RNA合成试剂盒(New England Biolabs)在16℃转录14-16h,每20μL反应含1μg线性模板。根据制造商的说明,使用MEGAClear转录清洁试剂盒(Thermo Fisher)纯化sgRNA。将纯化的sgRNA等分保存在-80℃。

[1189] 体外脱氨基测定法

[1190] 在补充说明3中显示了用作体外脱氨基测定法的模板的DNA寡核苷酸的序列。所有寡核苷酸均购自IDT。将合成有互补序列的单链寡核苷酸在pH 8.0的Tris缓冲液中合并(5μL的100μM溶液),并通过加热至95℃进行5分钟退火,然后以0.1℃$秒^{-1}$的速率逐渐冷却至37℃以生成79个碱基对(bp)的dsDNA底物。在室温将新鲜融化的碱基编辑器蛋白(在10μL反应体积中的2μM终浓度)与反应缓冲液(20mM HEPES pH 7.5,150mM KCl,0.5mM DTT,0.1mM EDTA,10mM $MgCl_2$)[37]中指示的sgRNA(2.2μM终浓度)中复合5分钟。然后添加退火的dsDNA底物至终浓度250nM。在37℃进行反应30分钟,然后通过在99℃加热5分钟进行蛋白质变性。PB缓冲液(Qiagen,100μL)和异丙醇(25μL)的添加确保蛋白质与底物DNA分离。用Minelute柱(Qiagen)纯化DNA,并使用耐U的聚合酶(Phusion U Hot Start,ThermoFisher)和补充信息中列出的引物,通过15个qPCR循环(12ng输入DNA,50μL反应体积)将所得产物扩增到线性范围的顶部。使用RapidTip2(Diffinity Genomics)纯化扩增的DNA,并通过第二轮PCR(8个循环,5ng输入)添加条形码,然后制备在Illumina MiSeq上进行测序,如下文所述。

[1191] 基因组DNA的纯化和测序

[1192] 根据制造商的说明，使用Agencourt DNAdvance基因组DNA分离试剂盒(Beckman Coulter)分离基因组DNA。对于第一PCR，根据制造商的说明但添加3％DMSO和SYBR金核酸染色剂(Thermo Fisher)，使用Q5热启动DNA聚合酶(NEB)将DNA扩增到线性范围的顶部。对于所有扩增子，使用的PCR方案是在98℃下进行2分钟的初始加热步骤，然后进行优化的扩增循环数(98℃12s，61℃25s，72℃30s)。对于斑马鱼以及对于经转染的细胞样品，在50μL反应中使用30ng的输入DNA，对于耳蜗样品，在25μL反应中使用20ng。进行qPCR以确定每个扩增子的最佳循环数。使用RapidTip2(Diffinity Genomics)纯化扩增的DNA，并用进一步的PCR(8个循环，5ng输入)添加条形码。在第一轮PCR中使用的独特正向和反向引物包含位于退火区5'的恒定区(正向:5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNN-3'(SEQ ID NO:502)，反向:5'-TGGAGTTCAGACGTGTGCTCTTCCGATCT-3'(SEQ ID NO:503))，其有助于条形码化引物与扩增的DNA结合以用于第二轮PCR。

[1193] 第二轮PCR使用了具有三个区域的引物:5'恒定区，其允许扩增子结合Illumina流动池(斜体)；8-碱基的条形码化区(X)，和3'恒定区，其允许条形码化引物与第一轮PCR扩增子结合(粗体)。引物序列的例子有：

[1194] 正向：

[1195] 5'-*AATGATACGGCGACCACCGAGATCTACAC*XXXXXXXX**ACACTCTTTCCCTACACGAC-3'**(SEQ ID NO: 504)

[1196] 反向：

[1197] 5'-CAAGCAGAAGACGGCATACGAGATXXXXXXX**GTGACTGGAGTTCAGACGTGTGCTCTTC-3'**(SEQ ID NO: 505)

[1198] 测序衔接头和双重条形码化序列基于TruSeq索引化衔接头(Illumina)。合并条形码化样品，并通过凝胶提取(Qiagen)纯化，然后使用Ampure珠(Beckman Coulter)纯化，然后根据制造商的说明使用Qubit dsDNA HS试剂盒(Thermo Fisher)和qPCR(KAPA BioSystems)进行定量。根据制造商的说明，在MiSeq(Illumina)上使用180-250个碱基(取决于扩增子大小)的单端读段进行合并样品的测序。

[1199] 用于PCR扩增的寡核苷酸的序列示于补充说明3中。所有寡核苷酸获自IDT。本研究中每个扩增子的最佳PCR循环次数如下:VEGFA位点2人基因组DNA(所有延伸步骤的退火温度为61℃，持续25秒)：中靶:29个循环，脱靶#1:32个循环，脱靶#2:28个循环，脱靶#3:27个循环，脱靶#4:27个循环，VEGFA位点2鼠基因组DNA:中靶:31个循环，脱靶#1，#2，#3和#4:31个循环。HEK293位点3:脱靶#1:29个循环，脱靶#2:28个循环，脱靶#3:28个循环。FANCF脱靶#1:29个循环，脱靶#2:28个循环，脱靶3:28个循环。EMX1脱靶#1，#2和#3:28个循环。用于扩增斑马鱼DNA的TYR1，TYR2和TYR3 sgRNA:32个循环。如前所述[14]，遵循针对EMX1，FANCF和HEK293位点3基因座的中靶扩增的优化方案。

[1200] 基因组DNA测序读段的分析和比对

[1201] 如先前所述[14]对测序读段进行分析。简而言之，使用MiSeq报告仪(Illumina)对测序读段进行解复用(demultiplex)，并使用先前报告的自定义Matlab脚本[14]对个别FASTQ文件进行分析。使用Smith-Waterman算法将读段与参考序列比对。Q得分低于30的碱基调用用占位核苷酸(N)替换。此质量阈值导致具有预期错误率为1,000之1的核苷酸频率。使用以前

发布的计算下述插入/缺失的自定义Matlab脚本对插入/缺失频率进行定量，所述插入/缺失在nCas9切割位点周围的30个碱基的窗中发生，并且是最小2个碱基的插入或缺失[14]。若在经处理的样品与未处理的对照中插入/缺失的形成之间存在显著差异(Student的两尾t检验，$p<0.05$)，则将插入/缺失定义为可检测。

[1202] 对于与鼠基因组中的VEGFA位点2sgRNA相关的测序扩增子之一，CFD脱靶#3，不可能精确地测量插入/缺失的形成。在该基因座处的原间隔物之前直接有12个鸟嘌呤碱基，这使得该位点的PCR和高通量测序容易产生随机插入或缺失；在多个独立的未处理的对照样品中观察到高达20％的测序读段的缺失率。由于在任何处理条件下在此脱靶基因座处均未检测到显著的碱基编辑(图132和133)，因此怀疑在此基因座处的插入/缺失的形成也可忽略不计。

[1203] phred.II Q30得分对应于碱基调用中估计的99.9％的准确性[38]。若假定基站调用错误随机分布在四个碱基间，则在给定位置处0.1％不正确碱基调用的概率对应于0.1/4＝0.025％的碱基调用下限。低于此阈值的C→T编辑百分比归类为不可检测。PCR期间的自发脱氨基[39]或聚合酶错误也可以引入人工C→T编辑物。为了区分碱基编辑器诱导的C→T编辑与人工C→T编辑速率，对于每个扩增子对未处理的对照细胞测序，并使用Student两尾t检验以$p<0.05$为阈值计算特定条件下C→T编辑是否是统计学显著的。具有统计学显著的编辑率>0.025％的脱靶位点认为是可测量的。表17中包括了每个样品的比对且高质量过滤的读段的数目。

[1204] 基因组DNA序列比对的统计分析

[1205] 除非另有说明，否则全文中引用的平均值代表$n\geq3$个独立的生物学重复，并且已经陈述了平均值±标准差。

[1206] 图2和3中显示的高通量测序数据的统计分析通过比较经处理的样品中的中靶和脱靶编辑百分比与阴性对照样品(未处理)中测量的任何编辑来进行。使用了Student两尾t检验，并且表16中显示了各个p值。*$p\leq0.05$，**$p\leq0.01$和***$p\leq0.001$。当编辑低于检测限(0.025％)时，不计算显著性；所有未处理的对照样品显示检测不到的编辑。

[1207] 对于图128A，通过在碱基编辑活性窗(对于HEK293位点3和EMX1，C_4-C_8，对于FANCF和VEGFA位点2，C_4-C_9)中平均胞嘧啶的编辑来计算平均中靶碱基编辑。

[1208] 为了解决碱基编辑活性中sgRNA依赖性的差异，计算碱基编辑:插入/缺失比率(图130B)。通过用具有C→T转化的HTS读取的百分比(每个位点的碱基编辑窗间的平均值)除以含有插入/缺失的HTS读取的百分比产生此比率。如上所述，若针对特定基因座的脱靶编辑低于检测限，则出于计算特异性比率的目的，保守假定检测方法的估计上限(0.025％)。

[1209] 体外编辑DNA的数据分析

[1210] 测序读段使用MiSeq报告仪(Illumina.)自动解复用。使用在线包usegalaxy.org[40]进行质量过滤。使用FASTQ Groomer，接着使用FASTAMasker[41]，将Illumina质量得分小于或等于30的个别碱基转换为占位核苷酸“N”。随后，使用补充说明1中提供的自定义python脚本对生成的经质量过滤的FASTQ文件进行分析。扫描测序读段以查找与靶DNA序列两侧侧翼的两个14个碱基的序列的精确匹配。若未找到完全匹配，则从分析中排除读段。若找到两个14个碱基的序列并且它们之间的序列长度等于预期的原间隔物长度(20个碱基)，则保存侧翼区域之间找到的原间隔物序列，并计数通过原间隔物内每个位点处的高通量测序调用的

碱基。

[1211] 细胞培养

[1212] 在补充有10％(v/v)胎牛血清(FBS)的Dulbecco改良Eagle培养基加GlutaMax(ThermoFisher)中于37℃和5％CO_2维持HEK293T(ATCCCRL-3216)和NIH/3T3(ATCC CRL-1658)。细胞获自ATCC并且在购买时ATCC验证不含支原体。

[1213] 将碱基编辑器质粒转染到HEK293T细胞中

[1214] 将HEK293T细胞接种在无抗生素培养基中的48孔胶原包被的BioCoat平板(Corning)上，并以约70％汇合度转染。除非另有说明，否则根据制造商的方案，每孔使用1.5μl Lipofectamine 2000(Thermo Fisher)转染750ng BE和250ng sgRNA表达质粒。

[1215] 将碱基编辑器蛋白质转染到HEK293T细胞中

[1216] 将HEK293T细胞接种在250μL无抗生素培养基中的48孔胶原包被的BioCoat平板(Corning)上，并以约70％汇合度转染。将碱基编辑器蛋白在室温与1.1X摩尔过量的必要sgRNA温育5分钟。然后将复合物与1.5μLLipofectamine 2000(Thermo Fisher)一起温育，并按照制造商的规定进行转染，以进行质粒递送。除非另有说明，否则将BE蛋白添加到最终浓度200nM(基于总孔体积275μL)。

[1217] 将碱基编辑器质粒转染到NIH/3T3细胞中

[1218] 将NIH/3T3细胞接种在无抗生素DMEM培养基中的48孔胶原包被的BioCoat平板(Corning)上，并以约75％汇合度转染。除非另有说明，否则根据制造商的方案，每孔使用1.4μL Lipofectamine 3000与1μL P3000试剂(Thermo Fisher)转染600ng BE和200ng sgRNA表达质粒。

[1219] 将碱基编辑器蛋白质转染到NIH/3T3细胞中

[1220] 将NIH/3T3细胞接种在无抗生素DMEM培养基中的48孔胶原包被的BioCoat平板(Corning)上，并以约75％汇合度转染。将碱基编辑器蛋白质与1.1倍摩尔过量的所示sgRNA在25℃温育5分钟。然后将复合物与1.4μL Lipofectamine 3000(Thermo Fisher)一起温育，并根据制造商的规定进行转染，以进行质粒递送。未使用P3000试剂，因为其添加导致蛋白质沉淀和碱基编辑效率降低。除非另有说明，否则将BE蛋白添加到终浓度400nM(基于总孔体积275μL)。

[1221] 阳离子脂质中包囊的BE3蛋白:引导RNA的耳蜗内递送

[1222] 所有动物实验均由麻省眼耳医院的机构动物护理和使用委员会(Institutional Animal Care and the Use Committee of the Massachusetts Eye and Ear Infirmary)批准。如先前所述[42]，在混合遗传背景的P1-P2小鼠中进行耳蜗内递送。在手术程序之前通过降低体温来麻醉小鼠。在右耳附近产生耳廓后切口，提起泡以暴露耳蜗。将BE3蛋白(57.7μM)与sgRNA(100μM)以1:1.1摩尔比预复合，然后以1:1体积比与Lipofectamine 2000(Thermo Fisher)混合。以250nL min^{-1}的速率将所得溶液(1.2-1.5μL)用玻璃移液管(端径，5μm)通过蜗囊注入耳蜗基底转处的耳蜗管(scala media)中，与纳升微量泵(WPI，UMP3+Micro4+NanoFil)连接。注射后，将切口闭合，并且将小鼠置于加热垫上以恢复。3-4天后，将小鼠的耳蜗解剖成科尔蒂器、血管纹和蜗轴。将每个组织进一步显微切割成5-7个分开的片，并对每个样品分别进行DNA提取，然后如上所述进行高通量测序。图132和图134中呈现的数据显示了从每个耳蜗区域提取一个微解剖样品所得的测序数据。

[1223] 微量注射BE3蛋白:引导RNA入斑马鱼胚胎

[1224] 符合波士顿儿童医院的内部法规审查要求将斑马鱼(Tuebingen株)保持在标准条件下。对单细胞阶段的斑马鱼胚胎注射以约1:1的摩尔比(终浓度4.5μM)与合适的sgRNA或无关的sgRNA对照预复合的2nL的BE3蛋白。受精后四天,如前所述[43]于95℃在50mM NaOH中30分钟从幼体中提取DNA,然后用Tris-HCl中和所得的溶液。如上所述,对基因组DNA进行定量,通过PCR扩增和测序。

[1225] 蛋白质凝胶分析

[1226] 显示的所有蛋白质凝胶均为预制的4-12%聚丙烯酰胺Bis-TrisPlus(Thermo Fisher)。它们在MOPS缓冲液(Thermo Fisher)中以180V运行50分钟。通过在100mM DTT和1X十二烷基硫酸锂(LDS)变性样品缓冲液(Thermo Fisher)中加热至99℃达10分钟制备以进行加载。根据制造商的说明,使用Instant Blue Protein Stain(Expedion)对凝胶进行染色。

[1227] 对于细胞裂解物分析,将2mL诱导后的过夜培养物以15,000g沉淀,然后在100μL B-PER(Thermo Fisher)中按照制造商的说明裂解。

[1228] 资料可用性

[1229] 支持本研究结果的高通量测序数据已以登录号SRP097884存储在NCBISequence Read Archive数据库中。编码用于蛋白质表达的HF-BE3和BE3以及用于哺乳动物表达的HF-BE3的质粒可以以登录号87439(pCMV-HF-BE3),87438(pET42b-HF-BE3),87437(pET42b-BE3)从Addgene获得。

[1230] 参考文献

[1231] 1 Cox,D.B.,Platt,R.J.&Zhang,F.Therapeutic genome editing:prospects and challenges.Nature medicine21,121-131,doi:10.1038/nm.3793(2015).

[1232] 2 Doudna,J.A.&Charpentier,E.Genome editing.The new frontier of genome engineering with CRISPR-Cas9.Science346,1258096,doi:10.1126/science.1258096(2014).

[1233] 3 Komor,A.C.,Badran,A.H.&Liu,D.R.CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes.Cell168,20-36,doi:10.1016/j.cell.2016.10.044(2017).

[1234] 4 Sander,J.D.&Joung,J.K.CRISPR-Cas systems for editing,regulating and targeting genomes.Nature biotechnology32,347-355,doi:10.1038/nbt.2842(2014).

[1235] 5 Hilton,I.B.&Gersbach,C.A.Enabling functional genomics with genome engineering.Genome Res25,1442-1455,doi:10.1101/gr.190124.115(2015).

[1236] 6 Wyman,C.&Kanaar,R.DNA double-strand break repair:all's well that ends well.Annual review of genetics40,363-383,doi:10.1146/annurev.genet.40.110405.090451(2006).

[1237] 7 Ran,F.A.et al.Genome engineering using the CRISPR-Cas9 system.Nat Protoc8,2281-2308,doi:10.1038/nprot.2013.143(2013).

[1238] 8 Cong,L.et al.Multiplex genome engineering using CRISPR/Cas systems.Science339,819-823,doi:10.1126/science.1231143(2013).

[1239] 9 Bodles-Brakhop,A.M.,Heller,R.&Draghia-Akli,R.Electroporation for the Delivery of DNA-based Vaccines and Immunotherapeutics:Current Clinical Developments.Mol Ther17,585-592,doi:10.1038/mt.2009.5(2009).

[1240] 10 Kay,M.A.,Glorioso,J.C.&Naldini,L.Viral vectors for gene therapy: the art of turning infectious agents into vehicles of therapeutics.Nature medicine7,33-40,doi:Doi 10.1038/83324(2001).

[1241] 11 Midoux,P.,Pichon,C.,Yaouanc,J.J.&Jaffres,P.A.Chemical vectors for gene delivery:a current review on polymers,peptides and lipids containing histidine or imidazole as nucleic acids carriers.Brit J Pharmacol157,166-178, doi:10.1111/j.1476-5381.2009.00288.x(2009).

[1242] 12 Yin,H.et al.Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype.Nature biotechnology32,551-553,doi:10.1038/ nbt0914-952d(2014).

[1243] 13 Hu,J.H.,Davis,K.M.&Liu,D.R.Chemical Biology Approaches to Genome Editing:Understanding,Controlling,and Delivering Programmable Nucleases.Cell chemical biology23,57-73,doi:10.1016/j.chembiol.2015.12.009(2016).

[1244] 14 Komor,A.C.,Kim,Y.B.,Packer,M.S.,Zuris,J.A.&Liu,D.R.Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage.Nature533,420-424,doi:10.1038/nature17946(2016).

[1245] 15 Hess,G.T.et al.Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells.Nature methods13,1036-1042,doi:10.1038/ nmeth.4038(2016).

[1246] 16 Ma,Y.et al.Targeted AID-mediated mutagenesis(TAM)enables efficient genomic diversification in mammalian cells.Nature methods13,1029-1035,doi: 10.1038/nmeth.4027(2016).

[1247] 17 Nishida,K.et al.Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems.Science353,aaf8729,doi: 10.1126/science.aaf8729(2016).

[1248] 18 Li,J.,Sun,Y.,Du,J.,Zhao,Y.&Xia,L.Generation of targeted point mutations in rice by a modified CRISPR/Cas9 system.Mol PlantIn Press,doi: 10.1016/j.molp.2016.12.001(2016).

[1249] 19 Koyama,S.,Ishii,K.J.,Coban,C.&Akira,S.Innate immune response to viral infection.Cytokine43,336-341,doi:10.1016/j.cyto.2008.07.009(2008).

[1250] 20 Kleinstiver,B.P.et al.High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects.Nature529,490-495,doi:10.1038/ nature16526(2016).

[1251] 21 Hsu,P.D.et al.DNA targeting specificity of RNA-guided Cas9 nucleases.Nature biotechnology31,827-832,doi:10.1038/nbt.2647(2013).

[1252] 22 Fu,Y.F.et al.High-frequency off-target mutagenesis induced by

CRISPR-Cas nucleases in human cells.Nature biotechnology31,822-826,doi:10.1038/nbt.2623(2013).

[1253] 23 Pattanayak,V.et al.High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity.Nature biotechnology31,839-843,doi:10.1038/nbt.2673(2013).

[1254] 24 Tsai,S.Q.et al.GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases.Nature biotechnology33,187-197,doi:10.1038/nbt.3117(2015).

[1255] 25 Guilinger,J.P.et al.Broad specificity profiling of TALENs results in engineered nucleases with improved DNA-cleavage specificity.Nature methods11,429-435,doi:10.1038/nmeth.2845(2014).

[1256] 26 Singh,R.,Kuscu,C.,Quinlan,A.,Qi,Y.J.&Adli,M.Cas9-chromatin binding information enables more accurate CRISPR off-target prediction.Nucleic acids research43,677-683,doi:ARTN e118 10.1093/nar/gkv575(2015).

[1257] 27 Liang,X.Q.et al.Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection.J Biotechnol208,44-53,doi:10.1016/j.jbiotec.2015.04.024(2015).

[1258] 28 Wang,M.et al.Efficient delivery of genome-editing proteins using bioreducible lipid nanoparticles.Proceedings of the National Academy of Sciences of the United States of America113,2868-2873,doi:10.1073/pnas.1520244113(2016).

[1259] 29 Zuris,J.A.et al.Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo.Nature biotechnology33,73-80,doi:10.1038/nbt.3081(2015).

[1260] 30 DeWitt,M.A.et al.Selection-free genome editing of the sickle mutation in human adult hematopoietic stem/progenitor cells.Sci Transl Med8,360ra134,doi:10.1126/scitranslmed.aaf9336(2016).

[1261] 31 Kim,K.et al.Highly efficient RNA-guided base editing in mouse embryos.Nature biotechnology,doi:10.1038/nbt.3816(2017).

[1262] 32 Richardson,C.D.,Ray,G.J.,DeWitt,M.A.,Curie,G.L.&Corn,J.E.Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA.Nature biotechnology34,339-344,doi:10.1038/nbt.3481(2016).

[1263] 33 Lin,Y.C.et al.Genome dynamics of the human embryonic kidney 293lineage in response to cell biology manipulations.Nature communications5,4767,doi:10.1038/ncomms5767(2014).

[1264] 34 Davis,K.M.,Pattanayak,V.,Thompson,D.B.,Zuris,J.A.&Liu,D.R.Small molecule-triggered Cas9 protein with improved genome-editing specificity.Nat Chem Biol11,316-318,doi:10.1038/Nchembio.1793(2015).

[1265] 35 Lin,S.,Staahl,B.,Alla,R.K.&Doudna,J.A.Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery.Elife,e04766,doi:10.7554/eLife.04766(2014).

[1266] 36 Doench,J.G.et al.Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9.Nature biotechnology34,184-191,doi:10.1038/nbt.3437(2016).

[1267] 37 Jinek,M.et al.A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity.Science337,816-821,doi:10.1126/science.1225829(2012).

[1268] 38 Ewing,B.&Green,P.Base-calling of automated sequencer traces using phred.II.Error probabilities.Genome Res8,186-194(1998).

[1269] 39 Liu,M.&Schatz,D.G.Balancing AID and DNA repair during somatic hypermutation.Trends in immunology30,173-181,doi:10.1016/j.it.2009.01.007(2009).

[1270] 40 Afgan,E.et al.The Galaxy platform for accessible,reproducible and collaborative biomedical analyses:2016update.Nucleic acids research44,W3-W10,doi:10.1093/nar/gkw343(2016).

[1271] 41 Blankenberg,D.et al.Manipulation of FASTQ data with Galaxy.Bioinformatics26,1783-1785,doi:10.1093/bioinformatics/btq281(2010).

[1272] 42 Hu,L.X.et al.Diphtheria Toxin-Induced Cell Death Triggers Wnt-Dependent Hair Cell Regeneration in Neonatal Mice.J Neurosci36,9479-9489,doi:10.1523/Jneurosci.2447-15.2016(2016).

[1273] 43 Meeker,N.D.,Hutchinson,S.A.,Ho,L.&Trede,N.S.Method for isolation of PCR-ready genomic DNA from zebrafish tissues.Biotechniques43,610,612,614(2007).

[1274] 44 Zuris,J.A.et al.Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo.Nature biotechnology 33,73-80(2015).

[1275] 45 Tsai,S.Q.et al.GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases.Nature biotechnology 33,187-197(2015).

[1276] 46 Doench,J.G.et al.Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9.Nature biotechnology 34,184-191(2016).

[1277] 47Haeussler,M.et al.Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR.Genome Biol 17,148(2016).

[1278] 补充信息

[1279] 补充说明1:用于分析质量过滤的体外编辑的DNA的Python脚本。

[1280]

```
from __future__ import print_function
from __future__ import division

import Bio #This will import the BioPython suite
from Bio import SeqIO #Necessary to read/write sequence handles
from Bio.Seq import Seq
import os
import collections
import csv

inputfile = "please_specify_your_input_file_here_containing_filtered_reads" #specify the filenames that contain sequences
filenames = []

```

[1281]

```
    for file in os.listdir(inputfile):
        if file.endswith(".fastqsanger"):
            filenames.append(file)

    spacer = []
    list_of_filenames = []

    for file in filenames:
        site = {}
        output = open(file + ".txt", "w")
        list_of_filenames.append(file + ".txt") #allows calling of the txt files that com
e from fastq files later
        for rec in SeqIO.parse(file, "fastq"):
            split1=rec.seq.tostring().split("GTTCGCGGCGATCG") #14-base pair constant_regi
on_before_protospacer
            if len(split1)>=2:
                split2=split1[1].split("TGGATCGCCTGGCA") #14-base pairc constant_region_a
fter_protospacer
                site=split2[0]
                if len(site)==20:
                    output.write(site + "\n")

    BASES = 'ATGCN'
    UNRECOGNIZED = 'X'
    BASE_SEPERATOR = dict(zip(BASES, ',,,,\n'))
    a_index = 0
    t_index = 1
    g_index = 2
    c_index = 3
    n_index = 4

    def get_counts_by_column(base, count, library):
        current_count = library[count]
        if base == 'A':
            current_count[a_index] += 1
        elif base == 'T':
            current_count[t_index] += 1
        elif base == 'G':
            current_count[g_index] += 1
        elif base == 'C':
            current_count[c_index] += 1
        elif base == 'N':
            current_count[n_index] += 1

    def dna_counts(list_of_sequences, sample):
        first_oligo = list_of_sequences[0]
        for i in range (len(first_oligo)):
            sample.append([0,0,0,0,0])
        for j in range(len(first_oligo)):
            for i in range(len(list_of_sequences)):
                get_counts_by_column(list_of_sequences[i][j], j, libname)


    for file in list_of_filenames:
        spacer_list = open(file).read().splitlines()
        output2=[]
        dna_counts(spacer_list, output2)
        with open(file  + ".csv", "wb") as f:
            writer = csv.writer(f)
            writer.writerows(output2)
```

[1282] 补充说明2:本研究中使用的蛋白质序列

[1283] 表达BE3的蛋白质序列

[1284] MGSSHHHHHHSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTN
KHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLI
SSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQ
SCHYQRLPPHILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSI

KKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF
GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQL
FEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY
DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAIL
RRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFD
KNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF
DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQL
KRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL
AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVEN
TQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKM
KNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKV
ITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK
ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSK
ESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLE
AKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFV
EQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR
YTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESD
ILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV(SEQ ID NO:185)

[1285] 表达的HF-BE3的蛋白质序列

[1286] MGSSHHHHHHSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTN
KHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLI
SSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQ
SCHYQRLPPHILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSI
KKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF
GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQL
FEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY
DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAIL
RRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTAFD
KNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF
DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQL
KRRRYTGWGALSRKLINGIRDKQSGKTILDFLKSDGFANRNFMALIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL
AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVEN
TQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKM
KNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRAITKHVAQILDSRMNTKYDENDKLIREVKV
ITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK
ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSK
ESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLE

AKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFV EQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR YTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESD ILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV(SEQ ID NO:186)

[1287] 补充说明3:本研究中使用的核苷酸的序列

[1288] 未发表的引物,用于扩增人细胞中用于HTS的脱靶基因组DNA(SEQ ID NOS:506-513)

[1289]	fwd_VEGFA_site2_off_targe t_1_human	ACACTCTTTCCCTACACGACGCTCTTCCGATCTN NNNTCCTACAAGTAACAGTCCAAGAA
[1290]	rev_VEGFA_site2_off_targe t_1_human	TGGAGTTCAGACGTGTGCTCTTCCGATCTTTCTG CAACTTAACTTACGTGAAA
[1291]	fwd_VEGFA_site2_off_targe t_2_human	ACACTCTTTCCCTACACGACGCTCTTCCGATCTN NNNACCAAGCCCATTTGTCCAGG
[1292]	rev_VEGFA_site2_off_targe t_2_human	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCCTT CTTTTTGAGCTTTGGGC
[1293]	fwd_VEGFA_site2_off_targe t_3_human	ACACTCTTTCCCTACACGACGCTCTTCCGATCTN NNNTCCATACCAGCAGCAGTTCC
[1294]	rev_VEGFA_site2_off_targe t_3_human	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTCAC CTCAGCTCCTGCAC
[1295]	fwd_VEGFA_site2_off_targe t_4_human	ACACTCTTTCCCTACACGACGCTCTTCCGATCTN NNNCCCACTGATTCTACACCATGGT
[1296]	rev_VEGFA_site2_off_targe t_4_human	TGGAGTTCAGACGTGTGCTCTTCCGATCTGGAGT TCCCAACCTTTTTGACA

[1297] 先前发表其它引物(用于与HEK_3,EMX1,FANCF相关的脱靶位点)。

[1298] 引物,用于扩增鼠细胞中用于HTS的脱靶基因组DNA(SEQ ID NOS:514-521)

[1299]	fwd_VEGFA_site2_off_targ et_1_murine	ACACTCTTTCCCTACACGACGCTCTTCCGATCTN NNNCTGGCTGGAGATTCAGAGACAC
[1300]	rev_VEGFA_site2_off_targ et_1_murine	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTGC CCCTTCTGACACACATAC
[1301]	fwd_VEGFA_site2_off_targ et_2_murine	ACACTCTTTCCCTACACGACGCTCTTCCGATCTN NNNACCCCTCAAGGCTTGACATTTC
[1302]	rev_VEGFA_site2_off_targ et_2_murine	TGGAGTTCAGACGTGTGCTCTTCCGATCTTGAAA AGTTGGGAGAGGGGATG
[1303]	fwd_VEGFA_site2_off_targ et_3_murine	ACACTCTTTCCCTACACGACGCTCTTCCGATCTN NNNTTTGTACCCCAGTCCCCTCATC
[1304]	rev_VEGFA_site2_off_targ et_3_murine	TGGAGTTCAGACGTGTGCTCTTCCGATCTTGAAG TTACGGGGATGTCACTTG
[1305]	fwd_VEGFA_site2_off_targ et_4_murine	ACACTCTTTCCCTACACGACGCTCTTCCGATCTN NNNTTAACATCCAGTCTCCCAAACACA

[1306] rev_VEGFA_site2_off_targ et_4_murine TGGAGTTCAGACGTGTGCTCTTCCGATCTACACA CACACACTACTAGGACA

[1307] 引物,用于扩增鼠细胞中用于HTS的中靶基因组DNA(SEQ ID NOS:522-523)

[1308] fwd_VEGFA_site2_on_targ et_murine ACACTCTTTCCCTACACGACGCTCTTCCGATCTN NNNCGCTACTACGGAGCGAGAAG

[1309] rev_VEGFA_site2_on_targ et_murine TGGAGTTCAGACGTGTGCTCTTCCGATCTACAG GGGCAAAGTGAGTGAC

[1310] 引物,用于产生PCR产物以充当sgRNA的T7转录的底物(SEQ ID NOS:524-529)

[1311] rev_sgRNA_T7:用于所有

[1312] 情况 AAAAAAAGCACCGACTCGGTGCCAC TAATACGACTCACTATAGGGAGTCCGAGCAGAA

[1313] fwd_sgRNA_T7_EMX1 GAAGAAGTTTTAGAGCTAGAAATAGCA TAATACGACTCACTATAGGGGAATCCCTTCTGC

[1314] fwd_sgRNA_T7_FANCF AGCACCGTTTTAGAGCTAGAAATAGCA

[1315] fwd_sgRNA_T7_HEK_site _3 TAATACGACTCACTATAGGGGCCCAGACTGAGC ACGTGAGTTTTAGAGCTAGAAATAGCA TAATACGACTCACTATAG

[1316] fwd_sgRNA_T7_VEGFA_ site_2 GACCCCCTCCACCCCGCCTCGTTTTAGAGCTA GAAATAGCA

[1317] fwd_sgRNA_T7_TC_repe at_in_vitro TAATACGACTCACTATAGGTCTCTCTCTCTCTCT CTCTCGTTTTAGAGCTAGAAATAGCA

[1318] 用于产生sgRNA转染质粒的引物(SEQ ID NO:530-531)

[1319] pFYF1320质粒用作模板,如先前描述(Komor et al)。先前报告了其它sgRNA质粒的序列。

[1320] rev_sgRNA_plasmid GGTGTTTCGTCCTTTCCACAAG GACCCCCTCCACCCCGCCTCGTTTTAGAGCTAGAA

[1321] fwd_VEGFA_site_2 ATAGCAAGTTAAAATAAGGC

[1322] 用于体外脱氨酶测定法的ssDNA底物的序列(SEQ ID NOS:532-533)

[1323] fwd_TC_repeat_s ACGTAAACGGCCACAAGTTCGCGGCGATCGTCTCTCTCTC

[1324] ubstrate TCTCTCTCTCTGGATCGCCTGGCATCTTCTTCAAGGACG

[1325] rev_TC_repeat_s ubstrate CGTCCTTGAAGAAGATGCCAGGCGATCCAGAGAGAGAGAG AGAGAGAGACGATCGCCGCGAACTTGTGGCCGTTTACGT

[1326] 用于扩增人细胞中用于HTS的脱靶基因组DNA的先前发表的引物(SEQ IDNOS:534-557)

[1327] fwd_EMX1_HTS ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCA GCTCAGCCTGAGTGTTGA TGGAGTTCAGACGTGTGCTCTTCCGATCTCTCGTGGGTT

[1328] rev_EMX1_HTS TGTGGTTGC ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCAT

[1329]	fwd_FANCF_HTS	TGCAGAGAGGCGTATCA TGGAGTTCAGACGTGTGCTCTTCCGATCTGGGGTCCCAG
[1330]	rev_FANCF_HTS	GTGCTGAC
[1331]	fwd_HEK293_site3_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNATG TGGGCTGCCTAGAAAGG
[1332]	rev_HEK293_site3_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCCAGCCAAA CTTGTCAACC
[1333]	fwd_EMX1_off1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAGT AGCCTCTTTCTCAATGTGC
[1334]	rev_EMX1_off1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCTTTCACAA GGATGCAGTCT
[1335]	fwd_EMX1_off2_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGA GCTAGACTCCGAGGGGA
[1336]	rev_EMX1_off2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCCTCGTCCT GCTCTCACTT
[1337]	fwd_EMX1_off3_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAG AGGCTGAAGAGGAAGACCA
[1338]	rev_EMX1_off3_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGGCCCAGCTG TGCATTCTAT
[1339]	fwd_FANCF_off1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAAC CCACTGAAGAAGCAGGG
[1340]	rev_FANCF_off1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGGTGCTTAAT CCGGCTCCAT
[1341]	fwd_FANCF_off2_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCC AGTGTTTCCATCCCGAA
[1342]	rev_FANCF_off2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCTCTGACCT CCACAACTCT
[1343]	fwd_FANCF_off3_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCT GGGTACAGTTCTGCGTGT
[1344]	rev_FANCF_off3_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCACTCTGAG CATCGCCAAG
[1345]	fwd_HEK293_site3_off1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCC CCTGTTGACCTGGAGAA
[1346]	rev_HEK293_site3_off1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCACTGTACTT GCCCTGACCA
[1347]	fwd_HEK293_site3_off2_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTGGTGTTGACA GGGAGCAA
[1348]	rev_HEK293_site3_off2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTGAGATGTG GGCAGAAGGG

[1349] fwd_HEK293_site3_off3_HTS ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTGAGAGGGAACAGAAGGGCT

[1350] rev_HEK293_site3_off3_HTS TGGAGTTCAGACGTGTGCTCTTCCGATCTGTCCAAAGGCCCAAGAACCT

[1351] 用于扩增斑马鱼中用于HTS的中靶基因组DNA的引物(SEQ ID NOS:558-563)

[1352] fwd_TYR1_zebrafish ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTTCCCCCGAGTCTGCACCT

[1353] rev_TYR1_zebrafish TGGAGTTCAGACGTGTGCTCTTCCGATCTCGAACTTGCATTCGCCGCAA

[1354] fwd_TYR2_zebrafish ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTCTGCCTTGGCATCGGGTG

[1355] rev_TYR2_zebrafish TGGAGTTCAGACGTGTGCTCTTCCGATCTCACCATACCGCCCCTAGAACTAACATTC

[1356] fwd_TYR3_zebrafish ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNACAACTGCTTTCCATGGTGTGT

[1357] rev_TYR3_zebrafish TGGAGTTCAGACGTGTGCTCTTCCGATCTTCCCAGGGCTTTCGTGGAGA

[1358]

位点	序列	SEQ ID NO
TYR1	$GTC_3AGGTC_8GAGGGTTCTGTCAGG$	564
TYR2	$CTTC_4C_5AGGATGAGAACACAGAGG$	565
TYR3	$CAAC_4C_5AC_7TGCTCAAAGATGCTGG$	566

[1359] 补充表1.这项工作中研究的斑马鱼基因组位点的原间隔物和PAM序列。

[1360] 实施例17:具有更高效率和产物纯度的C:G-至-T:A碱基编辑器

[1361] 碱基编辑是使用Cas9变体和胞苷脱氨酶的工程化融合物将靶C:G碱基对可编程转化为T:A碱基对,而不引起双链DNA断裂或不需要同源性指导的修复(1)。第三代碱基编辑器(BE3)和相关技术已被许多研究人员成功地用于一大批生物体中(2-13)。碱基编辑的产物分布(靶C:G碱基对转化为不想要的副产物以及期望的T:A产物的混合物的频率)以靶位点依赖性方式变化(2、3、6-8)。在本文,我们表征人细胞中碱基编辑结果的决定因素,并建立不想要的产物的形成取决于尿嘧啶N-糖基化酶(UNG),并且更有可能发生在碱基编辑活性窗内仅含有单个C的靶位点处。工程化改造使用增加某些序列的碱基编辑效率的胞苷脱氨酶同源物的构建体CDA1-BE3和AID-BE3。此外,还工程化改造第四代酿脓链球菌Cas9衍生的碱基编辑器(BE4),该编辑器更有效地阻断UNG接近经碱基编辑的中间体。与BE3相比,BE4在将不期望的副产物的频率减半的情况下将C:G至T:A碱基编辑的效率提高约50%,。这些改进也适用于产生金黄色葡萄球菌Cas9衍生的BE4(SaBE4),它比BE4小得多,并且具有替代的靶向范围。

[1362] 导言

[1363] 传统的基因组编辑方法在基因组靶基因座处引入双链DNA断裂(DSB)(14)。对DSB损伤的细胞响应主要通过非同源末端连接(NHEJ)和相关过程进行(15)。尽管NHEJ通常重新连接DSB侧翼的两个末端,但在典型的基因组编辑条件下,不断重新引入DSB,最终导致在

DSB的位点处积累插入和缺失(插入/缺失)或易位,并破坏相应的基因组基因座(16)。在与DSB周围的区域含有同源性的供体DNA模板的存在下,活跃分裂的细胞还可以通过启动同源性指导的修复(HDR)来响应DSB,这允许研究人员比经由NHEJ可能的情况更精确和可预测地操作基因组(17)。HDR依赖性的基因组编辑受到与NHEJ结果竞争以及HDR对有丝分裂的依赖性所导致的低效率的限制(18)。

[1364] 已经报告了碱基编辑的发展,其实现以可编程的方式将C:G碱基对直接不可逆转化为T:A碱基对,而无需HDR或引入DSB(1)。碱基编辑器由栓系于催化受损的Cas9蛋白的单链DNA特异性胞苷脱氨酶和碱基切除修复抑制剂组成(1、4、9、10)。Cas9变体结合由相应的引导RNA编程的感兴趣的基因组基因座。蛋白质:RNA:DNA三元"R环"复合物(19)的形成暴露了小的(约5-nt)单链DNA窗,该窗充当栓系的胞苷胞苷脱氨酶的底物。该窗内的任何胞苷均被水解脱氨基为尿嘧啶,从而产生G:U中间体。

[1365] 碱基切除修复(BER)是细胞对G:U错配的主要响应,并且是通过尿嘧啶N-糖基化酶(UNG)切除尿嘧啶而引发的(20)。为了保护经编辑的G:U中间体免于被UNG切除,将83个氨基酸的尿嘧啶糖基化酶抑制剂(UGI)直接与催化死亡的Cas9(dCas9)的C末端融合(1)。为了操作细胞DNA错配修复系统优先将G:U错配中的G替换为A,将dCas9中的Ala 840氨基酸回复为His,使Cas9蛋白能够在与新形成的尿嘧啶相对的DNA链上形成切口,从而有效得多地将G:U中间体转化为期望的A:U和A:T产物(1)。组合这两个工程化成果产生了BE3,由以下组成的单一蛋白质:通过16个氨基酸的接头与酿脓链球菌dCas9(A840H)栓系的APOBEC1胞苷脱氨酶的三部分融合物,所述酿脓链球菌dCas9(A840H)经由4-氨基酸接头与UGI共价连接(1)。BE3和相关的碱基编辑器现已用于极其多种应用,包括植物基因组编辑,体内哺乳动物基因组编辑,靶向诱变和敲除研究(2-13)。最近已经通过报告具有改变的PAM要求(4),变窄的编辑窗(4),减少的脱靶编辑(10)和小分子依赖性(21)的BE3变体扩展了碱基编辑的范围。

[1366] 在某些基因座处,诸如BE3的碱基编辑器产生不想要的副产物,其中靶C:G碱基对转化为G:C或A:T碱基对,而不是期望的T:A产物(2、3、6-8)。在本文,我们阐明了碱基编辑产物纯度的决定因素,并建立了形成不想要的副产物需要UNG活性。已经确定阻断UNG接近尿嘧啶中间体对于靶基因座尤其重要,在所述靶基因座中单个C位于编辑窗内,以使使不想要的产物最小化。工程化改造第四代碱基编辑器BE4和SaBE4,它们与以前描述的碱基编辑器,包括BE3相比以更高的效率进行碱基编辑并大大改善产物纯度。

[1367] 结果

[1368] UNG活性是形成副产物所需要的

[1369] 由于在含尿嘧啶的DNA链内无碱基位点的形成和易错解析,在碱基切除修复期间可以产生不想要的碱基编辑副产物。为了确定缺乏尿嘧啶N-糖基化酶(UNG)的细胞中碱基编辑的产物纯度是否改善,用编码BE3和靶向EMX1,FANCF,HEK2,HEK3,HEK4或RNF2基因座的sgRNA的质粒对HAP1细胞(单倍体人细胞系)和HAP1 UNG^{-}细胞进行核转染(关于靶序列,参见图135B)。核转染后三天,提取基因组DNA,并通过PCR扩增靶基因座,并通过高通量DNA测序(HTS)进行分析。碱基编辑产物纯度定义为靶C被编辑为T的经编辑的测序读段(靶C已转化为A,G或T的读段)的百分比。经BE3处理的HAP1细胞的碱基编辑产物纯度在6个基因座中的12个靶C间平均为68±6%(n=3个生物学重复的平均值±s.d.)。在HAP1 UNG^{-}细胞中,所有测试的12个靶C均经过碱基编辑,产物纯度>98%(图135A)。另外,UNG敲除后,所有六个测

试基因座处的插入/缺失频率降低7至100倍(图135C)。这些数据强烈暗示UNG活性是碱基编辑期间不想要的产物形成必需的,与下述模型一致:在该模型中,无碱基位点形成和随后的用易错聚合酶的碱基切除修复导致靶核苷酸的随机化和导致插入/缺失的偶尔的链断裂。

[1370] 具有多个可编辑C的靶物表现出更高的产物纯度

[1371] 与APOBEC1的已知序列偏爱一致(22),通过BE3进行的碱基编辑效率对于直接在G下游的一些(但非所有)靶C可以是较低的(1)(图136A)。为了有效编辑此类靶标物,产生BE3变体,其中APOBEC1脱氨酶替换为CDA1(以生成CDA1-BE3),AID(以生成AID-BE3)或APOBEC3G(以生成APOBEC3G-BE3),三个具有不同序列偏爱的单链DNA特异性胞苷脱氨酶(23)。用编码这些BE3变体和靶向EMX1,FANCF,HEK2,HEK3,HEK4,或RNF2基因座的sgRNA的质粒转染HEK293T细胞。转染后三天,提取基因组DNA,并通过PCR扩增靶基因座,并使用HTS评估碱基编辑。与BE3相比,使用CDA1-BE3和AID-BE3观察到直接在G之后的靶C的更有效的编辑(图136A,图140A-D和图146-151)。通常,在不在G后的靶C处,CDA1-BE3和AID-BE3比BE3表现出更低的编辑效率(图140A-D)。相比之下,APOBEC3G-BE3表现出不可预测的序列偏爱,与BE3相比更低的C至T编辑的总体产率。这些发现提示了对于某些靶5'-GC-3'序列,CDA1-BE3和AID-BE3可以比BE3提供更高的编辑效率。

[1372] 在分析这些数据时,注意到CDA1-BE3和AID-BE3的产物纯度通常在CDA1-BE3和AID-BE3比BE3编辑更多C的那些位点处高于BE3的产物纯度(图136A-D)。例如,在HEK4基因座处,BE3仅有效编辑单个C(前面没有G的C),但是CDA1-BE3和AID-BE3两者编辑三个C(图140A-C)。在此基因座处BE3的产物纯度为50±7%(n=3个生物学重复的平均值±S.D.),而CDA1-BE3和AID-BE3的产物纯度分别为97±2%和93±2%。此外,分别具有产物纯度84±3%和91±2%的由BE3编辑的EMX1和FANCF含有多个以相当的效率编辑的C(图S2),而分别具有低得多的产物纯度28±3%和64±3%的由BE3编辑的HEK2和RNF2含有以不等效率编辑的多个C(图141A-C)。CDA1-BE3和AID-BE3(它们以相当的效率编辑HEK2基因座内的两个C)在此基因座处显示出高得多的产物纯度(分别为85±5%和81±4%)(图136A-D和图140C)。排除了下述可能性:在HEK2和RNF2位点处,BE3最初以相当的效率将多个C转化为U,然后被DNA修复系统以不同效率处理。鉴于此,当用BE3相对于CDA1-BE3或AID-BE3处理这些位点时预期相似的产物分布,而不是观察到的不同产物分布(图136B和图146-151)。取而代之的是,分离的G:U比G:U损伤簇更容易被UNG处理。可能的是,BE3(1,24)中的胞苷脱氨酶域的持续合成能力可以增加BE3在含有多个可编辑C的基因座处的停留时间,从而比在含有单个可编辑C的基因座处更有效地阻止UNG的接近。

[1373] 进一步分析了产物纯度,各个测序读段中编辑的C数目和UNG活性之间的关系。为了揭示在缺乏明确的UNG抑制的情况下碱基编辑的DNA的命运,将BE3的UGI成分除去以生成BE3B。用编码BE3或BE3B和靶向EMX1,FANCF,HEK2,HEK3,HEK4,或RNF2基因座的sgRNA的质粒转染HEK293T细胞。如预期,鉴于UNG在使上文建立的碱基编辑结果多样化中的作用,与经BE3处理的DNA相比经BE3B处理的DNA中在所有靶C处的产物纯度均大大降低,含非T的编辑产物的分数增加1.8±0.4倍的平均值(图142B)。

[1374] 分析了来自用靶向多C位点HEK2,HEK3和RNF2的sgRNA和BE3或BE3B处理的HEK293T细胞的个别DNA测序读段。对于每个位点,主要靶C指定为最有效修饰的核苷酸。在所有这三个位点间,含有主要靶C的不想要的C至非T编辑的测序读段的80±10%的平均值仅表现出

该单个碱基编辑事件(图142A-D和图143)。相比之下,在相同的三个多C位点间,含有主要靶C的清洁C至T编辑的测序读段的32±4%的低得多的平均值仅表现出该单个清洁碱基编辑事件(图142A-D和图143)。此外,经BE3B处理的HEK4 DNA(在编辑窗内仅含1个C的位点)的产物分布大致遵循A:G:T的1:3:1的比率(图143D)。这些观察结果共同指示当给定靶物中的单个胞苷在不存在UGI的情况下转化为U时,它由UNG引发的BER进行有效处理,从而给出产物的混合物。

[1375] 这些数据与下述模型一致,其中聚簇的G:U错配与分离的G:U错配差异地进行处理,并且更有可能产生清洁的C至T编辑。当只期望单一C至T编辑事件时,上述观察结果提示了UNG抑制对于使不期望的副产物最小化是至关重要的。然而,当使用dCas9-脱氨酶融合物(例如与TAM(8)和CRISPR-X(2))进行靶向随机诱变时,以上观察结果提示了仅具有单一可编辑C的靶位点将使产物混合物最大化。

[1376] 针对改善的产物纯度优化BE3结构

[1377] 将BE3的UGI组分替换为单链DNA结合(SSB)蛋白,以产生SSB-BE3,从而SSB可以阻断UNG接近R环的含有尿嘧啶的ssDNA部分。观察到SSB-BE3的碱基编辑效率大大降低,四个位点间的所有7个C表现出仅1.9±0.5%C至T转化的平均值(图137C)。

[1378] 由于在确定碱基编辑结果的步骤期间APOBEC,UGI和UNG的相对定位是未知的,因此将UGI重定位到BE3的N端(N-UGI-BE3),以改善UNG抑制。将UGI移动到BE3的N端导致与BE3相比2.3±0.6倍的所有7个测试的靶C间C至T编辑百分比的平均降低(图137C),以及与BE3相比所有四个位点处总体产物纯度降低,平均为2.2±0.5倍(图137B)。

[1379] 相比之下,将另外的UGI拷贝附加到BE3的C端(BE3-2xUGI)导致相对于BE3的产物纯度的大幅增加和与BE3的C至T编辑百分比相当的C至T编辑百分比。在测试的四个基因座中,非T编辑产物下降2.2±0.8倍的平均值(图137B)。这些观察结果提示了,添加第二UGI拷贝实质性减少UNG对G:U碱基编辑中间体的接近,从而大大改善了产物纯度。

[1380] 由于上述实验还揭示了碱基编辑结果对组分结构的敏感性,接下来我们优化了BE3组分之间的接头,以进一步提高产物纯度和编辑效率。我们将rAPOBEC1-dCas9(A840H)接头从16个氨基酸(BE3)改变为32个氨基酸(BE3C),并且将dCas9(A840H)-UGI接头从4(BE3)改变至9(BE3D)改变至16个氨基酸(BE3E,图138A)。当dCas9(A840H)-UGI接头的长度为9个氨基酸残基(BE3D)而不是4个氨基酸(BE3)时,非T产物形成平均减少1.3±0.1倍(图138D),C至T的编辑效率没有表观差异(图138C)。将rAPOBEC1-dCas9(A840H)接头从16个氨基酸(BE3)增加到32个氨基酸(BE3C)在HEK2基因座处的C至T编辑效率提高1.2±0.1倍的平均值(图138C)。此基因座以前是测试的最不均匀编辑的多C位点(图141A-C),并且扩展此接头导致C6相对于C4的优先编辑(在C6处编辑的测序读段与C4处编辑的测序读段的百分比比率)从2.6±0.2倍降低到1.8±0.1倍。我们推断此种较长的接头可以使脱氨酶更好接近R环中的ssDNA,并在碱基编辑窗中存在多个靶C时导致更均匀的脱氨基。在其他测试的基因座处,BE3C还表现出相当或改善的碱基编辑效率和产物纯度(图138C-D)。

[1381] BE4,具有增强的效率和产物纯度的C:G至T:A碱基编辑器

[1382] 通过组合所有三个改进来工程化改造碱基编辑器构建体BE4——将rAPOBEC1-dCas9接头延伸至32个氨基酸,将dCas9-UGI接头延伸至9个氨基酸,以及将UGI的第二拷贝附加至具有另一个9-氨基酸接头的构建体的C末端。靶物-AID,一种由Nishida等(9)报告的

替代碱基编辑器也克隆到与BE4相同的质粒主链中。用编码BE3,BE4或靶物-AID和靶向EMX1,FANCF,HEK2,HEK3,HEK4,或RNF2基因座的sgRNA的质粒转染HEK293T细胞。转染后三天,提取基因组DNA,并通过PCR扩增靶基因座并通过HTS分析。观察到相对于BE3,BE4的所有12个经编辑的C间的C至T编辑效率平均增加1.5±0.3倍(图139C)。尽管在相同位置上对靶物-AID进行C至T编辑的平均效率比BE3的平均效率低1.5±0.5倍并且比BE4的平均效率低2.1±0.5倍,但重要的是注意到靶物-AID(其使用CDA1脱氨酶)似乎具有相对于BE3和BE4偏移的编辑窗,在位置C3和C4周围具有最佳编辑(图139C)。此种偏移的编辑窗使靶物-AID与BE3或BE4之间的效率和产物纯度的比较变得困难,因为即使使用相同的引导RNA时,给定的靶C可位于不同编辑窗内的较优化的位置或不太优化的位置。

[1383] 除了更高的C至T编辑效率外,相对于BE3,在所有测试的基因组位点,BE4还表现出实质性改善的产物纯度,非T产物形成的平均减少为2.3±0.3倍(图139D)。如从进一步阻碍碱基切除修复(这可以导致插入/缺失(25))预期,还观察到与BE3相比在BE4处理后所有六个基因座间的平均2.3±1.1倍的插入/缺失率降低(图144A-C)。这些结果共同指示BE4在所有测试的基因座上均提供了高效率的C至T编辑,高产物纯度和低插入/缺失形成率。

[1384] BE4改进与金黄色葡萄球菌Cas9(26)集成以生成SaBE4,该SaBE4用较小的金黄色葡萄球菌Cas9(A580N)替代酿脓链球菌dCas9(A840H),并且由于其替代的PAM要求而可以接近不同靶物。用编码SaBE3(4)或SaBE4和靶向FANCF,HEK3或HEK4基因座的sgRNA的质粒转染HEK293T细胞。与比较BE4和BE3的结果一致,我们观察到相对于SaBE3对于SaBE4的所有10个经的编辑C间C至T编辑效率的1.4±0.2倍的平均增加(图145A),不需要的非T编辑产物的1.8±0.5倍的平均减少(图145B)。这些结果指示,源自BE4增强的碱基编辑效率和产物纯度的增加也适用于源自其他Cas9同源物的碱基编辑器。

[1385] 材料和方法

[1386] 质粒克隆

[1387] 本研究中的所有质粒均通过使用Phusion U热起始聚合酶(Thermo Fisher)的USER克隆产生。将脱氨酶和SSB基因合成为gBlocks基因片段(Integrated DNA Technologies),并从Addgene(质粒#79620)获得靶物-AID。蛋白质序列列在补充说明中。

[1388] 细胞培养

[1389] 在补充有10%(v/v)胎牛血清(FBS)的Dulbecco改良的Eagle培养基加GlutaMax(ThermoFisher)中于37℃和5%CO_2维持HEK293T(ATCC CRL-3216)细胞。在补充有10%(v/v)胎牛血清(FBS)的Iscove改良Dulbecco培养基加GlutaMax(ThermoFisher Scientific)中于37℃和5%CO_2维持HAP1(Horizon Discovery C631)和HAP1 UNG^-(Horizon DiscoveryHZGHC001531c012)。

[1390] 转染

[1391] 将HEK293T细胞接种在48孔胶原包被的BioCoat平板(Corning)上,并以约75%汇合度转染。简而言之,根据制造商的方案,每孔使用1.5μLLipofectamine 2000(ThermoFisherScientific)转染750ng BE和250ng sgRNA表达质粒。

[1392] 根据制造商的方案,使用SE Cell Line 4DNucleofector™X Kit S(Lonza)对HAP1和HAP1 UNG^-细胞进行了核转染。简而言之,使用4DNucleofector™程序DZ-113用300ng BE和100ng sgRNA表达质粒对4x 10^5个细胞进行核转染。

[1393] 基因组DNA样品的高通量DNA测序

[1394] 3天后收获转染的细胞，并通过在37℃将细胞在裂解缓冲液(10mM Tris-HCl pH 8.0,0.05％SDS,25μg/mL蛋白酶K)中温育1小时，然后在80℃下30分钟分离基因组DNA。如前所述(6,1)，使用侧翼HTS引物对通过PCR扩增感兴趣的基因组区域。按照制造商的说明并如先前描述，使用Phusion高保真DNA聚合酶(ThermoFisher)进行PCR扩增。用含有测序衔接头的引物通过PCR扩增纯化的DNA。对产品进行凝胶纯化，并且使用QuantiT™ PicoGreen dsDNA测定试剂盒(ThermoFisher)和KAPA Library Quantitation Kit-Illumina(KAPA Biosystems)定量。如前所述，在Illumina MiSeq上对样品进行测序。

[1395] 数据分析

[1396] 测序读段使用MiSeq Reporter(Illumina)自动解复用，并使用自定义的Matlab脚本分析个别FASTQ文件，如先前所述(1)。使用Smith-Waterman算法，将每个读段成对与适当的参考序列比对。Q得分低于31的碱基用N取代，因此在计算核苷酸频率时排除。此处理产生约1,000之1的预期MiSeq碱基调用错误率。将读段和参考序列不含缺口的比对序列存储在比对表中，对于每个基因座，可以从所述比对表将碱基频率制成表格。

[1397] 使用先前描述的Matlab脚本(5,6,1)，对插入/缺失频率进行定量。扫描测序读段以查找与两个10-bp序列的精确匹配，该序列位于可能发生插入/缺失的窗两侧侧翼。若未定位精确匹配，则从分析中排除读段。若此插入/缺失窗的长度与参考序列完全匹配，则将读段分类为不含插入/缺失。若插入/缺失窗比参考序列长或短两个或多个碱基，则测序读段分别分类为插入或缺失。

[1398] 为了评估在编辑窗内多个靶胞苷处的碱基编辑结果之间的相互依赖性(连锁不平衡)，通过自动以Python脚本(补充说明1)分析了来自经BE处理的细胞的靶位点序列。简而言之，对测序读段进行扫描，以寻找与原间隔物每侧侧翼的两个7-bp序列的精确匹配。若中间区域不与20-bp完全匹配，则将其从进一步分析排除。基于具有最非T编辑结果的位置(主要靶C)处核苷酸的身份，将原间隔物序列进一步过滤成四组。对于这四个组中的每个以及整个合并物，对原间隔物中20个位置的每处的核苷酸丰度进行了计数。

[1399] 实施例18:包含LbCpf1(核酸酶死亡，核酸酶活性和切口酶)的碱基编辑器。

[1400] 如上所讨论，本文提供的任何融合蛋白的核酸可编程DNA结合蛋白(napDNAbp)可以是LbCpf1蛋白。在一些实施方案中，LbCpf1蛋白是无核酸酶活性，核酸酶活性或LbCpf1切口酶。对包含LbCpf1形式的融合蛋白的几种构建体测试了它们在不同靶序列中进行C至T编辑的能力。图152显示了所测试的构建体的示意图。构建体10具有[Apobec]-[LbCpf1]-[UGI]-[UGI]的域排列。构建体11具有[Apobec]-[LbCpf1]-[UGI]的域排列;构建体12具有[UGI]-[Apobec]-[LbCpf1]的域排列;构建体13具有[Apobec]-[UGI]-[LbCpf1]的域排列;构建体14具有[LbCpf1]-[UGI]-[Apobec]的域排列;构建体15具有[LbCpf1]-[Apobec]-[UGI]的域排列。对于每个构建体，使用了三种不同的LbCpf1蛋白(D/N/A，其指核酸酶死亡的LbCpf1(D);LbCpf1切口酶(N)和核酸酶活性LbCpf1(A))。对于这些构建体中的每个，连接域的接头在下文显示，其中XTEN指具有序列SGSETPGTSESATPES(SEQ ID NO:604)的XTEN接头，而BPNLS指具有序列KRTADGSEFEPKKKRKV(SEQ ID NO:740)的核定位序列。从N端(左)到C端(右)显示了构建体。

[1401] 构建体10:

[1402] Apobec-SGGSSGGSXTENSGGSSGGS-LbCpf1-SGGSGGSGGS-UGI-SGGSGGSGGS-UGI-SGGS-BPNLS

[1403] 构建体11:

[1404] Apobec-SGGSSGGSXTENSGGSSGGS-LbCpf1-SGGSGGSGGS-UGI-SGGS-BPNLS

[1405] 构建体12:

[1406] UGI-SGGSGGSGGS-Apobec-SGGSSGGSXTENSGGSSGGS-LbCpf1-SGGS-BPNLS

[1407] 构建体13:

[1408] Apobec-SGGSGGSGGS-UGI-SGGSSGGSXTENSGGSSGGS-LbCpf1-SGGS-BPNLS

[1409] 构建体14:LbCpf1-SGGSGGSGGS-UGI-SGGSGGSGGS-Apobec-SGGS-BPNLS

[1410] 构建体15:LbCpf1-SGGSGGSGGS-Apobec-SGGSGGSGGS-UGI-SGGS-BPNLS

[1411] 在实验中使用的共同指导主链(sgRNA)是GTAATTTCTACTAAGTGTAGAT(SEQ ID NO:741)[引导序列]TTTTTTT,其中SEQ ID NO:741的每个T是尿嘧啶(U),并且其中在括号之间显示了将构建体靶向到特定核苷酸序列的引导序列。在一些实施方案中,本文提供的任何构建体与包含GTAATTTCTACTAAGTGTAGAT(SEQ ID NO:741)的主链序列的sgRNA复合,其中SEQ ID NO:741的每个T是尿嘧啶(U)。在一些实施方案中,本文提供的任何引导RNA包含引导序列,该引导序列包含与序列,例如靶DNA序列完全互补的14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29或30个核苷酸。在进行的实验中,下文显示了测试的引导序列:

[1412] EMX23:TACTTTGTCCTCCGGTTCTGGAA(SEQ ID NO:742)

[1413] EMX20:TACTTTGTCCTCCGGTTCTG(SEQ ID NO:743)

[1414] EMX19:TACTTTGTCCTCCGGTTCT(SEQ ID NO:744)

[1415] EMX18:TACTTTGTCCTCCGGTTC(SEQ ID NO:745)

[1416] EMX17:TACTTTGTCCTCCGGTT(SEQ ID NO:746)

[1417] Hek2_23:CAGCCCGCTGGCCCTGTAAAGGA(SEQ ID NO:747)

[1418] Hek2_20:CAGCCCGCTGGCCCTGTAAA(SEQ ID NO:748)

[1419] Hek2_19:CAGCCCGCTGGCCCTGTAA(SEQ ID NO:749)

[1420] Hek2_18:CAGCCCGCTGGCCCTGTA(SEQ ID NO:750)

[1421] Hek2_17:CAGCCCGCTGGCCCTGT(SEQ ID NO:751)

[1422] 图153-159中显示了证明使用各种构建体和靶序列的C至T碱基对编辑百分比的数据,并且图160-166中显示了编辑百分比值(基于插入/缺失计数进行调节后)以及用于实验的插入/缺失百分比。

[1423] 补充序列

[1424] CDA1-BE3,AID-BE3,BE4和SaBE4融合蛋白的氨基酸序列。

[1425] CDA1-XTEN-dCas9-UGI-NLS一级序列(CDA1-BE3):

MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQSGT
ERGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLKIWA
CKLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRA
*EKRRSELSIMIQVKILHTTKSPAV***SGSETPGTSESATPES**DKKYSIGLAIGTNSVGWAVIT
DEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRIC
YLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLR
KKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFE
ENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFD
LAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKA
PLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY
KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFL
KDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFI
ERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIV
DLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
[1426] NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKL
INGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA
NLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMK
RIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDH
IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKV
ITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDY
KVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIV
WDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKK
YGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK
EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLK
GSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQ
AENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLG
GDSGGS**TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTD**
ENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGS*PKKKRKV* (SEQ ID NO: 165)

[1427] AID-XTEN-dCas9-UGI-NLS一级序列(AID-BE3)：

MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELLF
LRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEP
[1428] *EGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYE*
*VDDLRDAFRTLGL***SGSETPGTSESATPES**DKKYSIGLAIGTNSVGWAVITDEYKVPSKK
FKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEM

AKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDK
ADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVD
AKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQL
SKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRY
DEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKM
DGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKI
LTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKN
LPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILE
DIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIK
KGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKEL
GSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLK
DDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAE
RGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLV
SDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVR
KMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGR
DFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDS
PTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLI
IKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNE
QKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLF
TLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGS**T**
NLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLT
SDAPEYKPWALVIQDSNGENKIKMLSGGS*PKKKRKV* (SEQ ID NO: 166)

[1429]

[1430] rAPOBEC1-接头-dCas9-UGI-UGI-NLS一级序列(BE4)：

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHV
EVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPR
NRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGL
*PPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK***SGGSSGGSSGSETPGTSESAT**
PESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLI
GALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLV
EEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFR
GHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLEN
LIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIG
DQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQ
QLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNS
RFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYF
TVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF
DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEER
LKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN
FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQN
EKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRG
KSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLV
ETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNY
HHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFF
YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKK
TEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKS
KKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQ

[1431]

[1432] ISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTID RKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSGGSGGS**TNLSDIIEKETGK QLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWAL VIQDSNGENKIKML**SGGSGGSGGS***TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNK PESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML***SGGS*PKKKRK* (SEQ ID NO: 167)

[1433] rAPOBEC1-接头-SaCas9d-UGI-UGI-NLS一级序列(SaBE4):

[1434] *MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHV EVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPR NRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGL PPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK***SGGSSGGSSGSETPGTSESAT PESSGGSSGGSG**KRNYILGLAIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRR SKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSA ALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEV RGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFG WKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYY EKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITAR KEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAI NLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVI NAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLI EKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEEN SKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKD FINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKG YKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFI TPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDN DKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSK KDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVT VKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNN DLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKH PQIIKKG***GSPKKKRKVSSDYKDHDGDYKDHDIDYKDDDDKSGGSGGSGGS**TNLSDIIE **KETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEY KPWALVIQDSNGENKIKML**SGGSGGSGGS***TNLSDIIEKETGKQLVIQESILMLPEEVE EVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML***SGGS *PKKKRKV* (SEQ ID NO: 168)

[1435] 实施例17的参考文献

[1436] 1.A.C.Komor,Y.B.Kim,M.S.Packer,J.A.Zuris,D.R.Liu,Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage.Nature 533,420-424(2016).

[1437] 2.G.T.Hess et al.,Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells.Nat Methods 13,1036-1042(2016).

[1438] 3.K.Kim et al.,Highly efficient RNA-guided base editing in mouse embryos.Nat Biotechnol 35,435-437(2017).

[1439] 4.Y.B.Kim et al.,Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions.Nat Biotechnol 35,371-376(2017).

[1440] 5.C.Kuscu et al.,CRISPR-STOP:gene silencing through base-editing-induced nonsense mutations.Nat Meth advance online publication,(2017).

[1441] 6.J.Li,Y.Sun,J.Du,Y.Zhao,L.Xia,Generation of Targeted Point Mutations

in Rice by a Modified CRISPR/Cas9 System.Mol Plant 10,526-529(2017).

[1442] 7.Y.Lu,J.K.Zhu,Precise Editing of a Target Base in the Rice Genome Using a Modified CRISPR/Cas9 System.Mol Plant 10,523-525(2017).

[1443] 8.Y.Ma et al.,Targeted AID-mediated mutagenesis(TAM)enables efficient genomic diversification in mammalian cells.Nat Methods 13,1029-1035(2016).

[1444] 9.K.Nishida et al.,Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems.Science 353,(2016).

[1445] 10.H.A.Rees et al.,Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery.Nat Commun 8,15790(2017).

[1446] 11.L.Yang et al.,Engineering and optimising deaminase fusions for genome editing.Nat Commun 7,13330(2016).

[1447] 12.Y.Zong et al.,Precise base editing in rice,wheat and maize with a Cas9-cytidine deaminase fusion.Nat Biotechnol 35,438-440(2017).

[1448] 13.Z.Shimatani et al.,Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion.Nat Biotech 35,441-443(2017).

[1449] 14.A.C.Komor,A.H.Badran,D.R.Liu,CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes.Cell 168,20-36(2017).

[1450] 15.A.J.Davis,D.J.Chen,DNA double strand break repair via non-homologous end-joining.Translational cancer research 2,130-143(2013).

[1451] 16.M.M.Vilenchik,A.G.Knudson,Endogenous DNA double-strand breaks: Production,fidelity of repair,and induction of cancer.Proceedings of the National Academy of Sciences100,12871-12876(2003).

[1452] 17.F.Liang,M.Han,P.J.Romanienko,M.Jasin,Homology-directed repair is a major doublestrand break repair pathway in mammalian cells.Proceedings of the National Academy of Sciences 95,5172-5177(1998).

[1453] 18.Y.Miyaoka et al.,Systematic quantification of HDR and NHEJ reveals effects of locus,nuclease,and cell type on genome-editing.Scientific Reports 6,23549(2016).

[1454] 19.M.M.Jore et al.,Structural basis for CRISPR RNA-guided DNA recognition by Cascade.Nat Struct Mol Biol 18,529-536(2011).

[1455] 20.L.H.Pearl,Structure and function in the uracil-DNA glycosylase superfamily.Mutation Research/DNA Repair 460,165-181(2000).

[1456] 21.W.Tang,J.H.Hu,D.R.Liu,Aptazyme-embedded guide RNAs enable ligand-responsive genome editing and transcriptional activation.8,15939(2017).

[1457] 22.G.Saraconi,F.Severi,C.Sala,G.Mattiuz,S.G.Conticello,The RNA editing enzyme APOBEC1 induces somatic mutations and a compatible mutational signature is present in esophageal adenocarcinomas.Genome Biology 15,417(2014).

[1458] 23.R.M.Kohli et al.,Local Sequence Targeting in the AID/APOBEC Family Differentially Impacts Retroviral Restriction and Antibody Diversification.Journal of Biological Chemistry285,40956-40964(2010).

[1459] 24.L.Chelico,P.Pham,M.F.Goodman,Stochastic properties of processive cytidine DNA deaminases AID and APOBEC3G.Philosophical Transactions of the Royal Society B:Biological Sciences 364,583-593(2009).

[1460] 25.E.A.Kouzminova,A.Kuzminov,Patterns of chromosomal fragmentation due to uracil-DNA incorporation reveal a novel mechanism of replication-dependent double-stranded breaks.Molecular Microbiology 68,202-215(2008).

[1461] 26.F.A.Ran et al.,In vivo genome editing using Staphylococcus aureus Cas9.Nature 520,186-191(2015).

[1462] 27.F.d.A.di Fagagna,G.R.Weller,A.J.Doherty,S.P.Jackson,The Gam protein of bacteriophage Mu is an orthologue of eukaryotic Ku.EMBO Reports 4, 47-52(2003).

[1463] 28.C.Shee et al.,Engineered proteins detect spontaneous DNA breakage in human and bacterial cells.eLife 2,e01222(2013).

[1464] 等同实施方案和范围

[1465] 本领域技术人员将认识到或能够仅仅使用常规的实验确定本文所述的实施方案的许多等同方案。本发明的范围不意图限于以上说明书,而是如所附权利要求书中所述。

[1466] 在权利要求书中,诸如“一种”、“一个”和“该”的冠词可以表示一个或超出一个,除非相反地指出或者从上下文中显而易见。若存在一个、超出一个或所有组成员,则认为在组中的两个或多个成员之间包括“或”的权利要求或说明书是满足的,除非另有说明或从上下文中明显。包括两个或多个组成员之间的“或”的组的公开提供刚好存在组的一个成员的实施方案、存在组的超过一个成员的实施方案、和存在所有组成员的实施方案。为了简洁,本文尚未个别清楚地说明那些实施方案,但是应当理解本文中提供了这些中的每个实施方案,并且可以具体要求保护或放弃权利。

[1467] 应理解,本发明涵盖所有变型、组合和置换,其中来自一个或多个权利要求或来自说明书的相关部分的一个或多个限制、元素、条款、描述性术语等被引入另一个权利要求中。例如,可以修改依赖于另一个权利要求的任何权利要求以包括在依赖于相同基本权利要求的任何其他权利要求中找到的一个或多个限制。此外,在权利要求叙述组合物的情况下,应当理解包括根据本文中公开的任何制备或使用方法或根据本领域已知的方法(若有的话)的制备或使用组合物的方法,除非另有说明或者除非本领域普通技术人员明白会出现矛盾或不一致。

[1468] 在将元素呈现为列表(例如,以马库什群组格式)的情况下,应当理解,还公开了元素的每个子群,并且可以从群组中移除任何元素或元素子群。还应注意,术语“包含”旨在是开放的并且允许包含另外的元素或步骤。应当理解,通常,在实施方案、产物或方法称为包含特定元素、特征、步骤的情况下,也提供了由此类元素、特征或步骤组成,或基本上由此类元素、特征或步骤组成的实施方案、产物或方法。为了简洁,本文尚未个别清楚地说明那些实施方案,但是应当理解本文中提供了这些中的每个实施方案,并且可以具体要求保护或

放弃权利。

[1469] 在给出范围的情况下，端点包括在内。此外，应当理解，除非另有说明或从上下文和/或本领域普通技术人员的理解中另外明显，否则表示为范围的值可以假定在一些实施方案中的所述范围内的任何特定值，至该范围下限的单位的十分之一，除非上下文另有明确规定。为了简洁，本文尚未个别清楚地说明每个范围中的数值，但是应当理解本文中提供了这些中的每个数值并且可以具体要求保护或放弃权利。还应当理解，除非另有说明或从上下文和/或本领域普通技术人员的理解中另外明显，否则表示为范围的值可以假定给定范围内的任何子范围，其中子范围的端点表示为与范围的下限的单位的十分之一相同的精度。

[1470] 此外，应当理解，本发明的任何具体实施方案可以明确地从任何一个或多个权利要求中排除。在给出范围的情况下，该范围内的任何值可以明确地从任何一个或多个权利要求中排除。本发明的组合物和/或方法的任何实施方案、元素、特征、应用或方面可以从任何一个或多个权利要求中排除。出于简洁的目的，本文未明确阐述其中排除一个或多个元素、特征、目的或方面的所有实施方案。

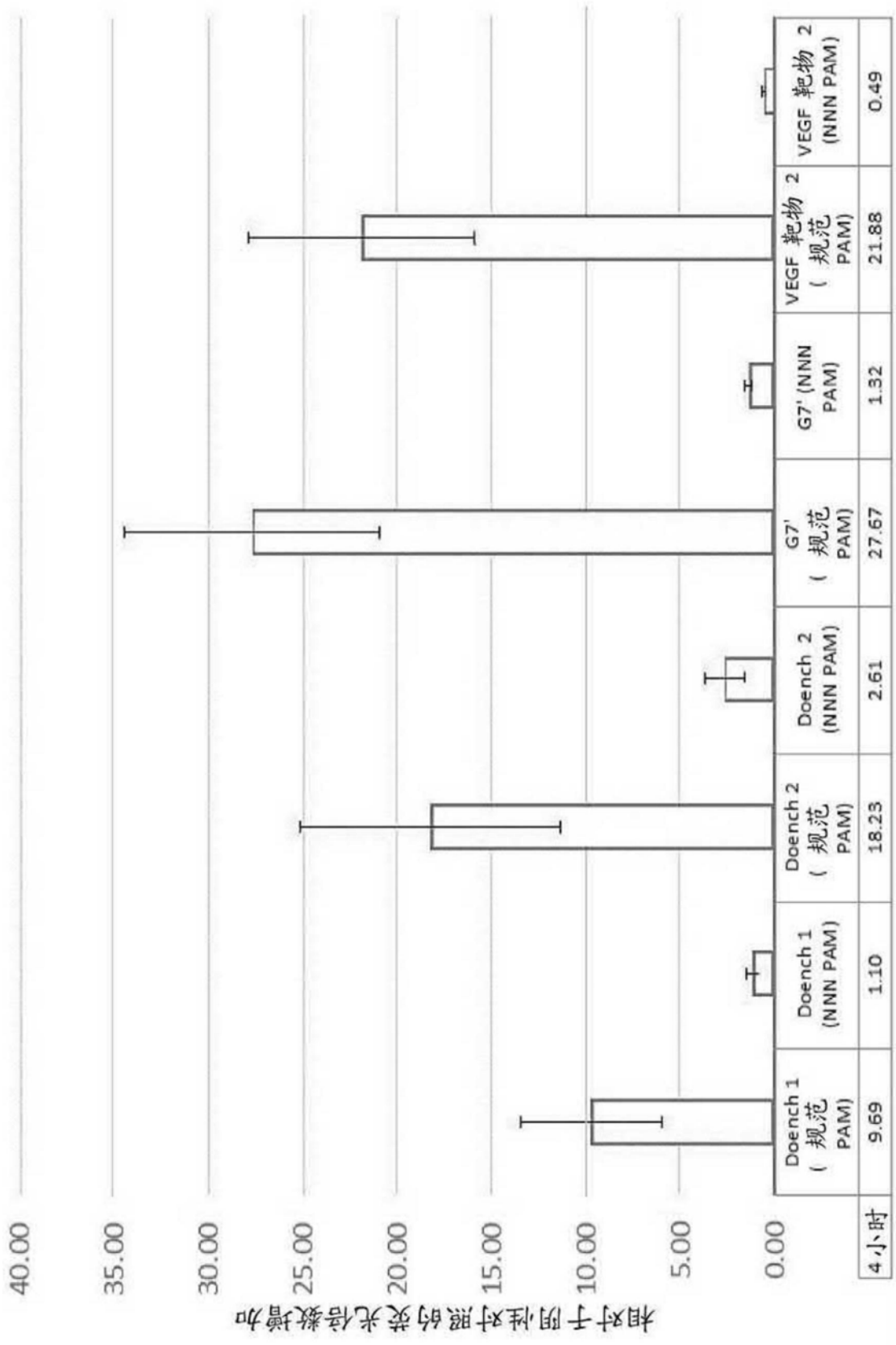

相对于阴性对照的荧光倍数增加
40.00
35.00
30.00
25.00
20.00
15.00
10.00
5.00
0.00
4小时
Doench 1 (规范 PAM)
9.69
Doench 1 (NNN PAM)
1.10
Doench 2 (规范 PAM)
18.23
Doench 2 (NNN PAM)
2.61
G7' (规范 PAM)
27.67
G7' (NNN PAM)
1.32
VEGF 靶物 2 (规范 PAM)
21.88
VEGF 靶物 2 (NNN PAM)
0.49

图1

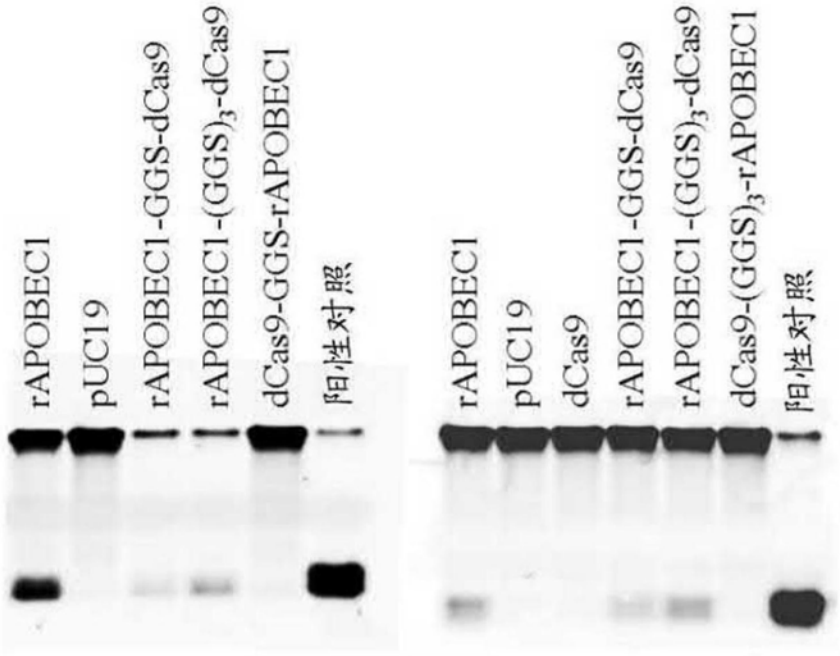

rAPOBEC1
pUC19
rAPOBEC1-GGS-dCas9
rAPOBEC1-(GGS)3-dCas9
dCas9-GGS-rAPOBEC1
阳性对照
rAPOBEC1
pUC19
dCas9
rAPOBEC1-GGS-dCas9
rAPOBEC1-(GGS)3-dCas9
dCas9-(GGS)3-rAPOBEC1
阳性对照

图2

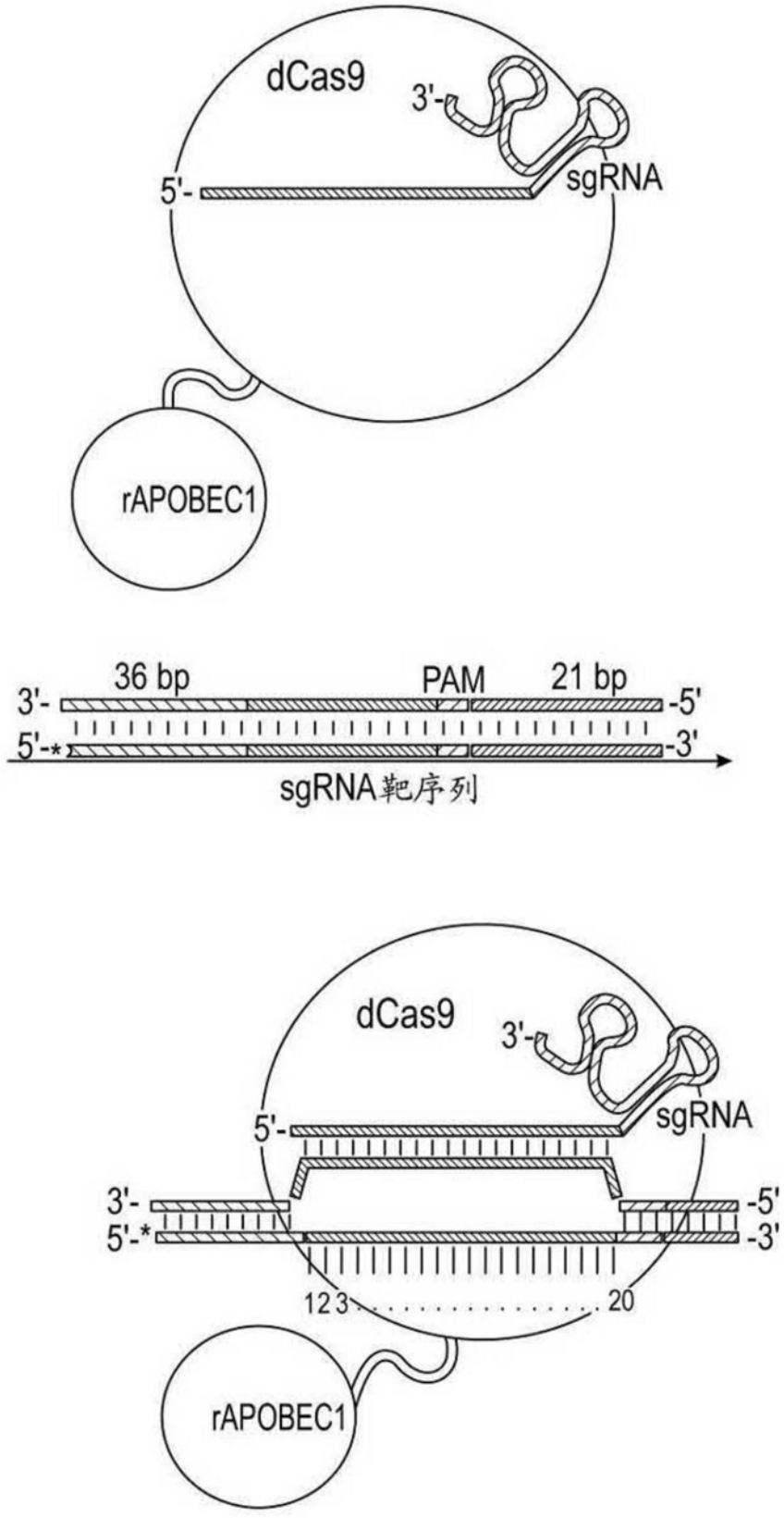

dCas9
3'-
5'-
sgRNA
rAPOBEC1
36 bp
PAM
21 bp
3'-
-5'
5'-*
-3'
sgRNA靶序列
dCas9
3'-
5'-
sgRNA
3'-
-5'
5'-*
-3'
1 2 3 20
rAPOBEC1

图3

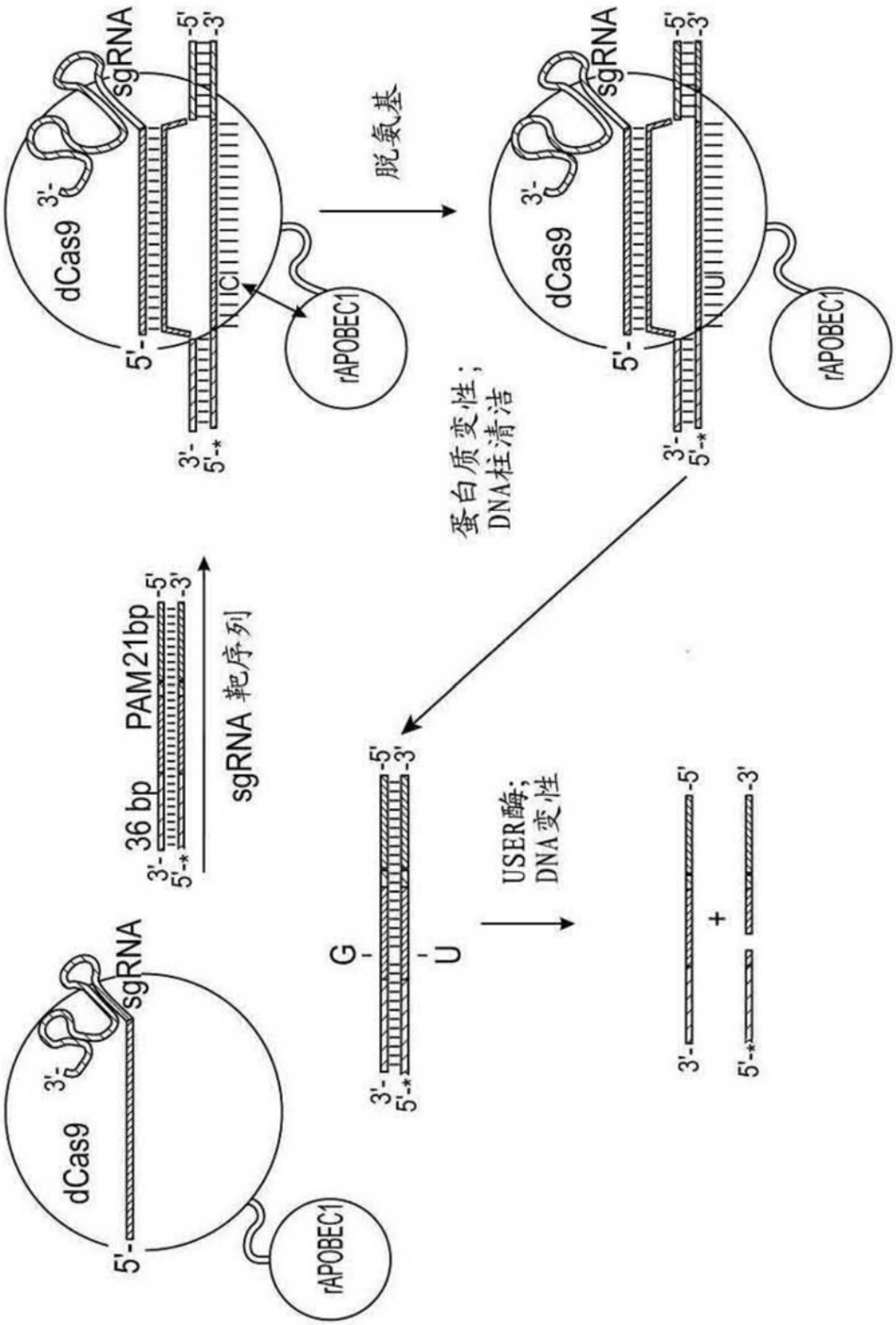

sgRNA
dCas9
rAPOBEC1
脱氨基
蛋白质变性；
DNA柱清洁
sgRNA 靶序列
36 bp
PAM21bp
USER酶；
DNA变性

图4

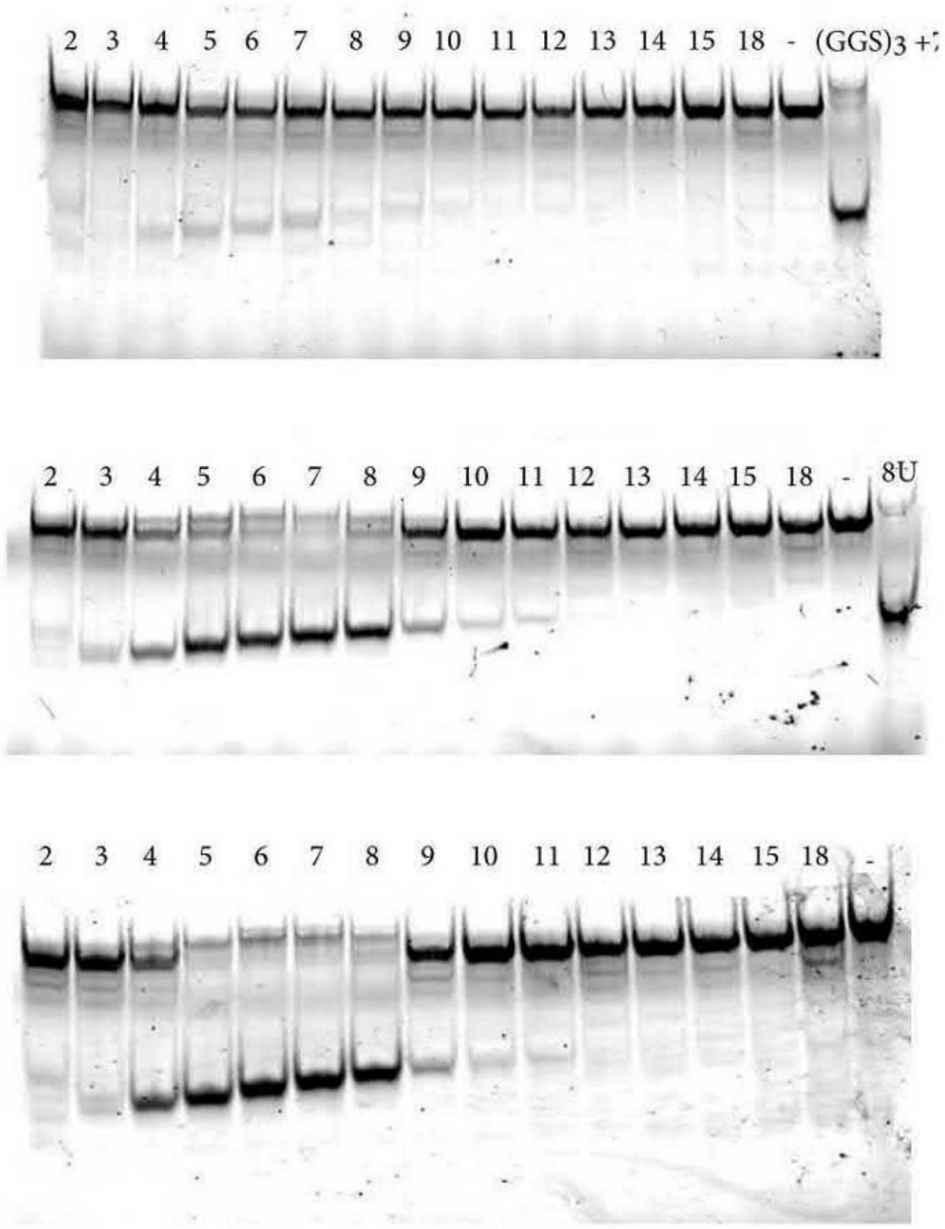

2 3 4 5 6 7 8 9 10 11 12 13 14 15 18 - (GGS)3 +7
2 3 4 5 6 7 8 9 10 11 12 13 14 15 18 - 8U
2 3 4 5 6 7 8 9 10 11 12 13 14 15 18 -

图5

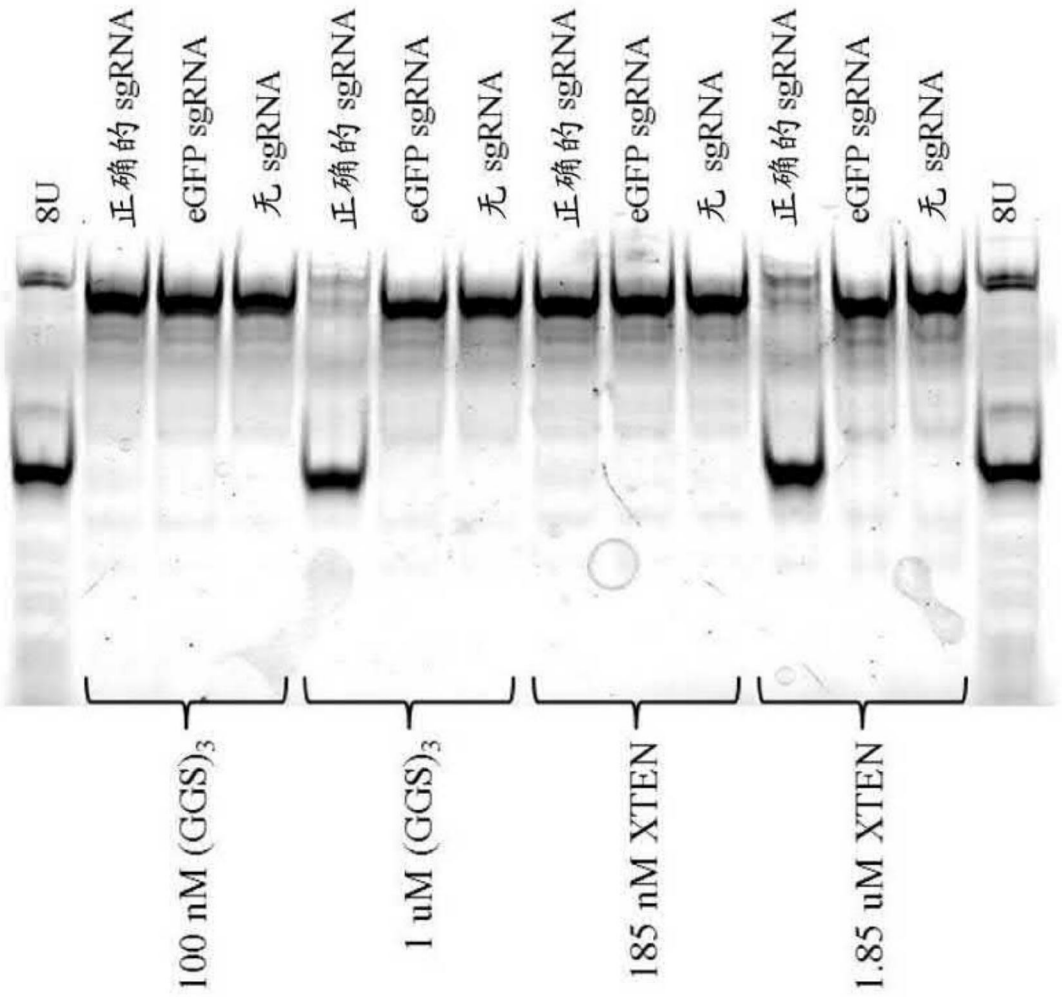

8U
正确的 sgRNA
eGFP sgRNA
无 sgRNA
正确的 sgRNA
eGFP sgRNA
无 sgRNA
正确的 sgRNA
eGFP sgRNA
无 sgRNA
正确的 sgRNA
eGFP sgRNA
无 sgRNA
8U
100 nM (GGS)3
1 uM (GGS)3
185 nM XTEN
1.85 uM XTEN

图6

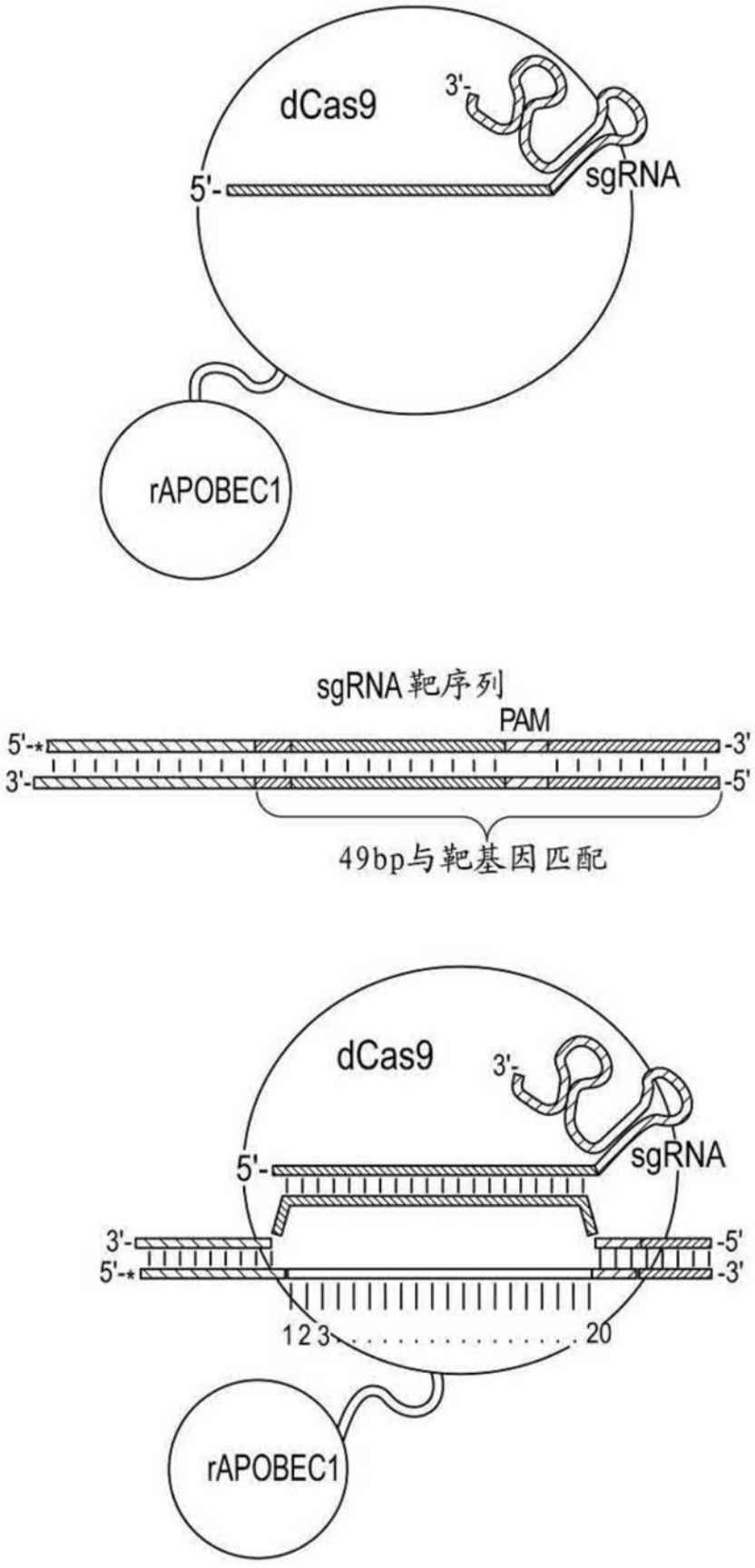

dCas9
3'-
5'-
sgRNA
rAPOBEC1
sgRNA靶序列
PAM
5'-*
-3'
3'-
-5'
49bp与靶基因匹配
dCas9
3'-
5'-
sgRNA
3'-
-5'
5'-*
-3'
1 2 3 20
rAPOBEC1

图7

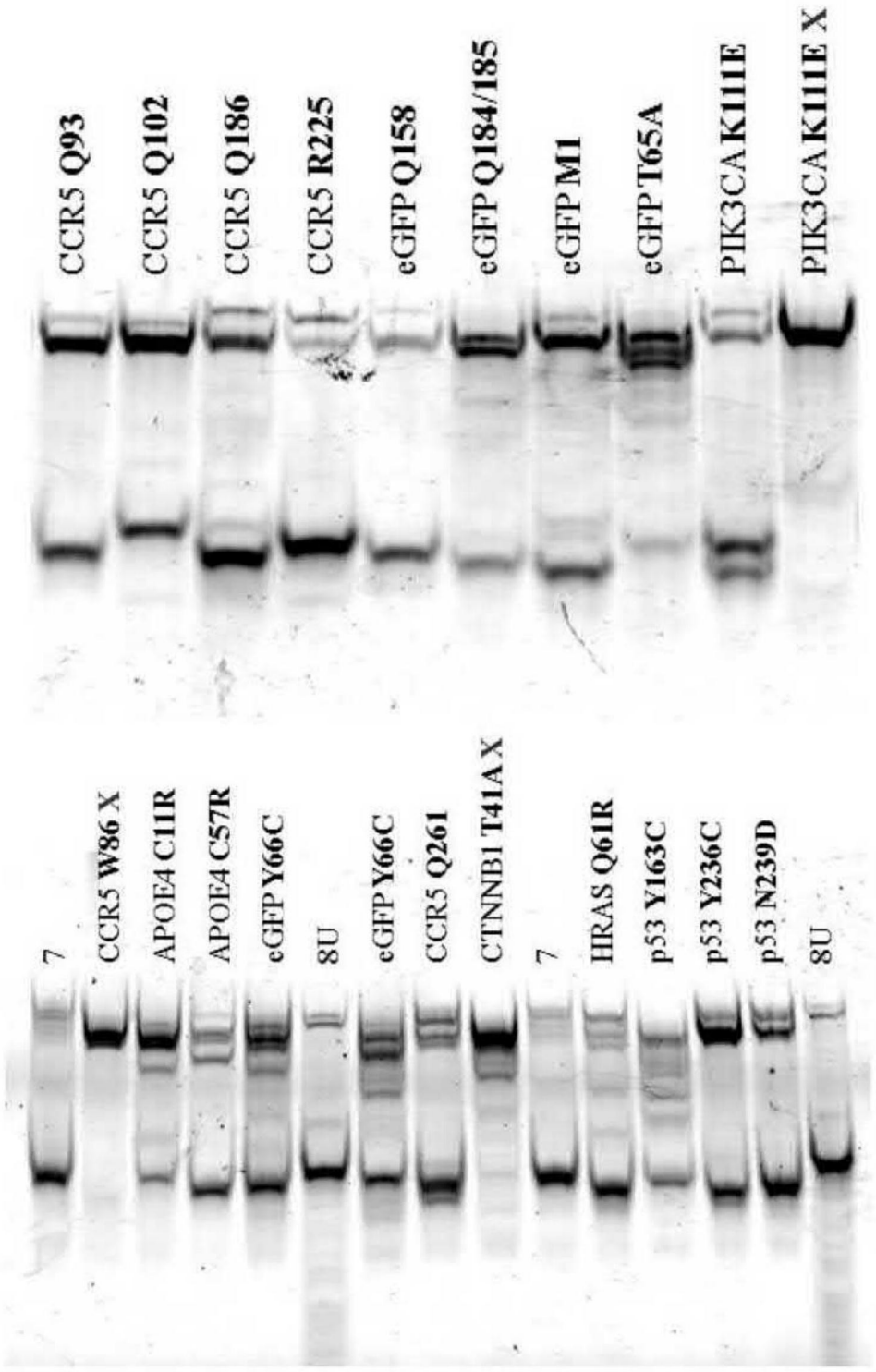

CCR5 Q93
CCR5 Q102
CCR5 Q186
CCR5 R225
eGFP Q158
eGFP Q184/185
eGFP M1
eGFP T65A
PIK3CA K111E
PIK3CA K111E X
7
CCR5 W86 X
APOE4 C11R
APOE4 C57R
eGFP Y66C
8U
eGFP Y66C
CCR5 Q261
CTNNB1 T41A X
7
HRAS Q61R
p53 Y163C
p53 Y236C
p53 N239D
8U

图8

EMX1		C_5	C_6	C_{10}
0 uM XTEN	A	0.1%	0.1%	0.1%
	C	99.8%	99.8%	99.8%
	G	0.0%	0.0%	0.1%
	T	0.0%	0.1%	0.0%
1.85 uM XTEN	A	0.1%	0.0%	0.1%
	C	60.4%	61.0%	99.1%
	G	0.0%	0.0%	0.1%
	T	39.5%	39.0%	0.7%

FANCF		C_6	C_7	C_8	C_{11}
0 uM XTEN	A	0.1%	0.1%	0.1%	0.1%
	C	99.8%	99.8%	99.9%	99.9%
	G	0.0%	0.1%	0.0%	0.0%
	T	0.0%	0.0%	0.0%	0.0%
1.85 uM XTEN	A	0.1%	0.1%	0.1%	0.1%
	C	63.9%	64.7%	65.0%	72.6%
	G	0.0%	0.0%	0.0%	0.0%
	T	36.0%	35.1%	34.9%	27.3%

HEK293 位点2		C_4	C_6	C_{11}
0 uM XTEN	A	0.1%	0.1%	0.1%
	C	99.9%	99.9%	99.9%
	G	0.0%	0.0%	0.0%
	T	0.0%	0.0%	0.1%
1.85 uM XTEN	A	0.1%	0.1%	0.1%
	C	80.6%	76.9%	99.6%
	G	0.0%	0.0%	0.0%
	T	19.3%	22.9%	0.3%

HEK293位点3		C_3	C_4	C_5	C_9
0 uM XTEN	A	0.1%	0.1%	0.0%	0.1%
	C	99.8%	99.9%	99.9%	99.9%
	G	0.0%	0.0%	0.0%	0.0%
	T	0.1%	0.0%	0.0%	0.0%
1.85 uM XTEN	A	0.1%	0.1%	0.0%	0.1%
	C	92.2%	74.8%	71.5%	96.6%
	G	0.0%	0.0%	0.0%	0.0%
	T	7.7%	25.1%	28.5%	3.3%

HEK293 位点4		C_3	C_5	C_8	C_{11}
0 uM XTEN	A	0.1%	0.0%	0.1%	0.0%
	C	99.8%	99.9%	99.8%	99.9%
	G	0.0%	0.0%	0.0%	0.0%
	T	0.0%	0.0%	0.1%	0.0%
1.85 uM XTEN	A	0.1%	0.1%	0.1%	0.1%
	C	98.8%	60.1%	97.0%	99.4%
	G	0.0%	0.0%	0.0%	0.0%
	T	1.1%	39.8%	2.9%	0.5%

RNF2		C_3	C_6
0 uM XTEN	A	0.1%	0.0%
	C	99.9%	99.9%
	G	0.0%	0.0%
	T	0.0%	0.0%
1.85 uM XTEN	A	0.1%	0.0%
	C	59.1%	57.8%
	G	0.0%	0.0%
	T	40.8%	42.1%

图9

EMX1		C_5	C_6	C_{10}
未处理的	A	0.0%	0.0%	0.0%
	C	99.5%	99.7%	100.0%
	G	0.0%	0.1%	0.0%
	T	0.5%	0.2%	0.0%
XTEN	A	0.7%	0.5%	0.0%
	C	93.5%	95.8%	100.0%
	G	2.1%	0.3%	0.0%
	T	3.6%	3.3%	0.0%
XTEN-UGI	A	0.2%	0.0%	0.0%
	C	81.8%	82.5%	100.0%
	G	0.6%	0.3%	0.0%
	T	17.4%	17.1%	0.0%

FANCF		C_6	C_7	C_8	C_{11}
未处理的	A	0.0%	0.0%	0.2%	0.1%
	C	99.9%	99.8%	99.8%	99.9%
	G	0.0%	0.0%	0.0%	0.0%
	T	0.1%	0.1%	0.0%	0.0%
XTEN	A	0.3%	0.1%	0.0%	0.0%
	C	98.1%	99.2%	99.0%	99.8%
	G	0.4%	0.0%	0.0%	0.0%
	T	1.2%	0.7%	1.0%	0.2%
XTEN-UGI	A	0.0%	0.0%	0.1%	0.0%
	C	93.2%	93.5%	93.4%	98.2%
	G	0.0%	0.0%	0.0%	0.0%
	T	6.7%	6.5%	6.5%	1.8%

HEK293位点2		C_4	C_6	C_{11}
未处理的	A	0.3%	0.2%	0.2%
	C	99.7%	99.7%	99.7%
	G	0.0%	0.0%	0.0%
	T	0.0%	0.0%	0.0%
XTEN	A	0.3%	0.3%	0.3%
	C	99.7%	99.4%	99.7%
	G	0.0%	0.3%	0.0%
	T	0.0%	0.0%	0.0%
XTEN-UGI	A	0.3%	0.2%	0.2%
	C	98.8%	98.2%	99.8%
	G	0.0%	0.3%	0.0%
	T	0.9%	1.3%	0.0%

HEK293位点2		C_3	C_4	C_5	C_9
未处理的	A	0.0%	0.0%	0.0%	0.0%
	C	100.0%	100.0%	100.0%	99.9%
	G	0.0%	0.0%	0.0%	0.0%
	T	0.0%	0.0%	0.0%	0.1%
XTEN	A	0.0%	0.6%	0.3%	0.1%
	C	100.0%	95.8%	95.8%	99.2%
	G	0.0%	0.2%	0.7%	0.4%
	T	0.0%	3.4%	3.2%	0.3%
XTEN-UGI	A	0.0%	0.3%	0.3%	0.0%
	C	96.8%	83.0%	79.2%	98.5%
	G	0.0%	0.0%	1.1%	0.2%
	T	3.2%	16.8%	19.4%	1.3%

HEK293位点4		C_3	C_5	C_8	C_{11}
未处理的	A	0.0%	0.4%	0.0%	0.0%
	C	99.8%	97.6%	99.9%	100.0%
	G	0.0%	1.0%	0.0%	0.0%
	T	0.2%	1.0%	0.0%	0.0%
XTEN	A	0.0%	1.1%	0.0%	0.0%
	C	99.6%	92.2%	99.9%	100.0%
	G	0.0%	2.2%	0.0%	0.0%
	T	0.4%	4.5%	0.0%	0.0%
XTEN-UGI	A	0.0%	0.5%	0.0%	0.0%
	C	99.4%	86.7%	99.1%	100.0%
	G	0.0%	1.8%	0.0%	0.0%
	T	0.6%	11.0%	0.9%	0.0%

RNF2		C_3	C_6
未处理的	A	0.0%	0.0%
	C	99.9%	99.5%
	G	0.0%	0.2%
	T	0.0%	0.3%
XTEN	A	0.0%	0.0%
	C	99.8%	99.3%
	G	0.0%	0.2%
	T	0.2%	0.5%
XTEN-UGI	A	0.0%	0.0%
	C	99.6%	99.1%
	G	0.0%	0.4%
	T	0.4%	0.5%

图10

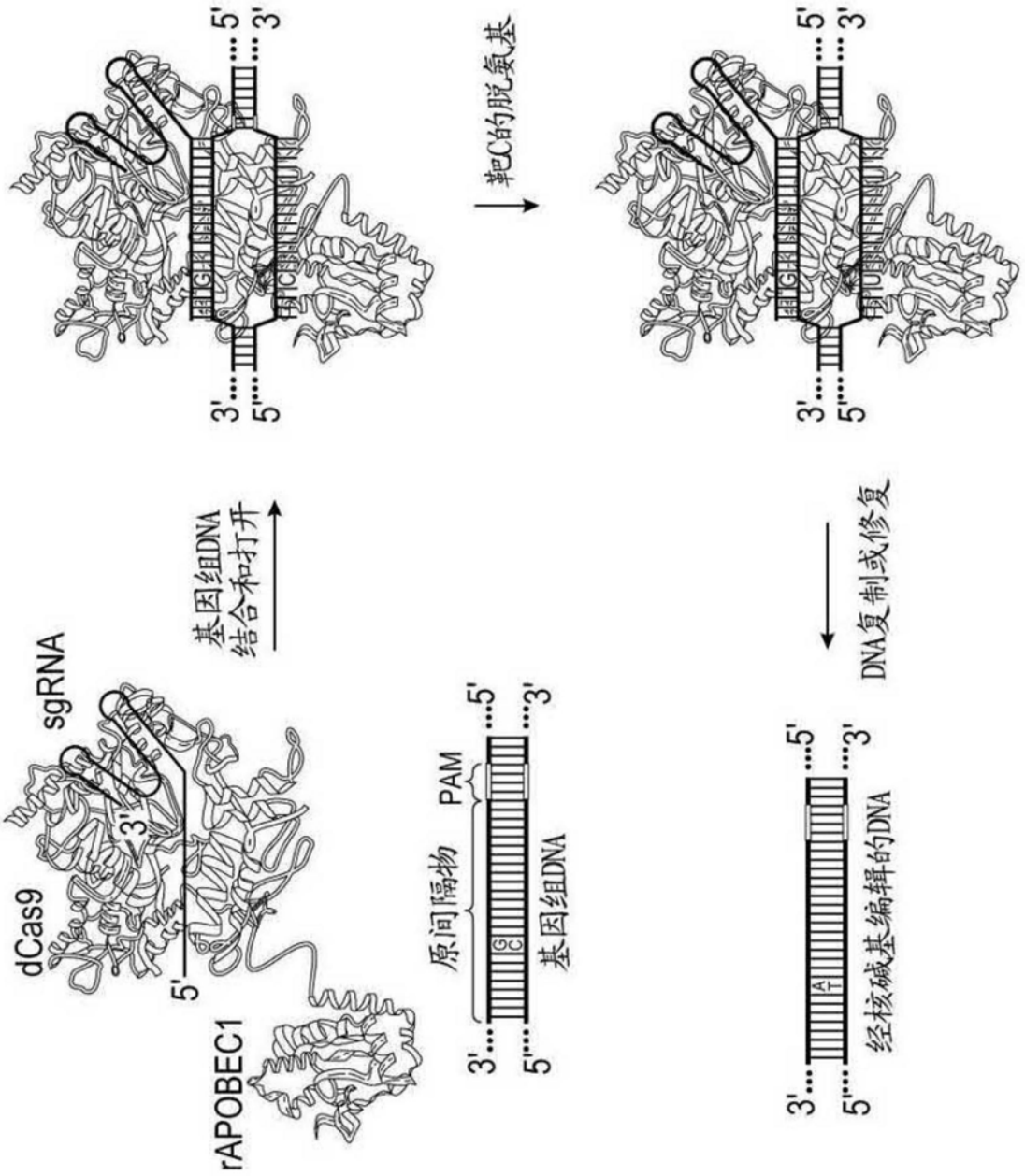

靶C的脱氨基
DNA复制或修复
基因组DNA
结合和打开
sgRNA
dCas9
rAPOBEC1
原间隔物
PAM
基因组DNA
经核碱基编辑的DNA
5'
3'

图11A

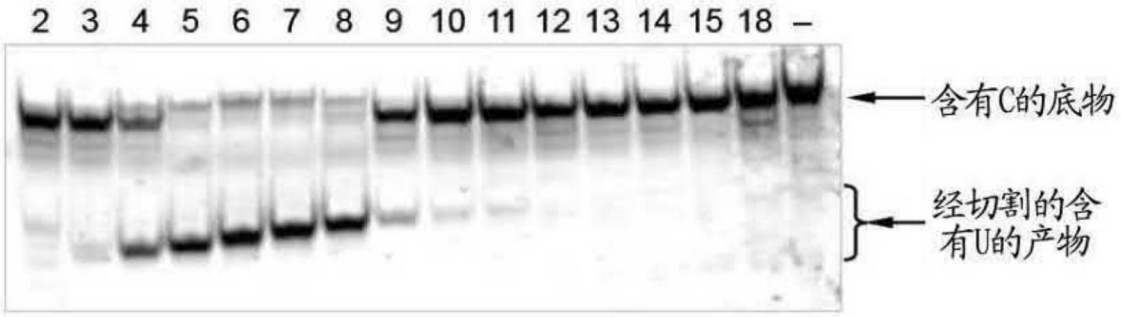

2 3 4 5 6 7 8 9 10 11 12 13 14 15 18 –
含有C的底物
经切割的含有U的产物

图11B

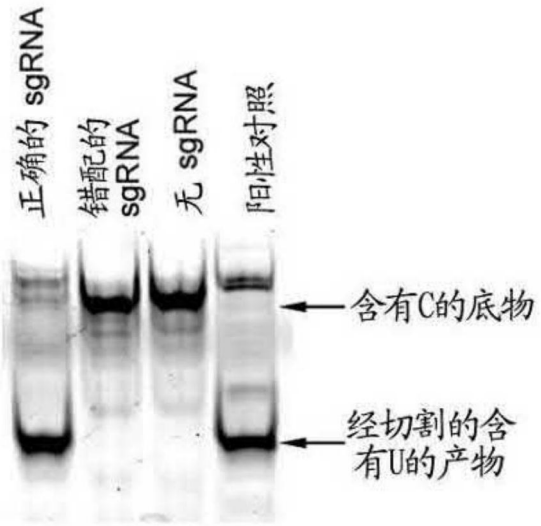


图11C

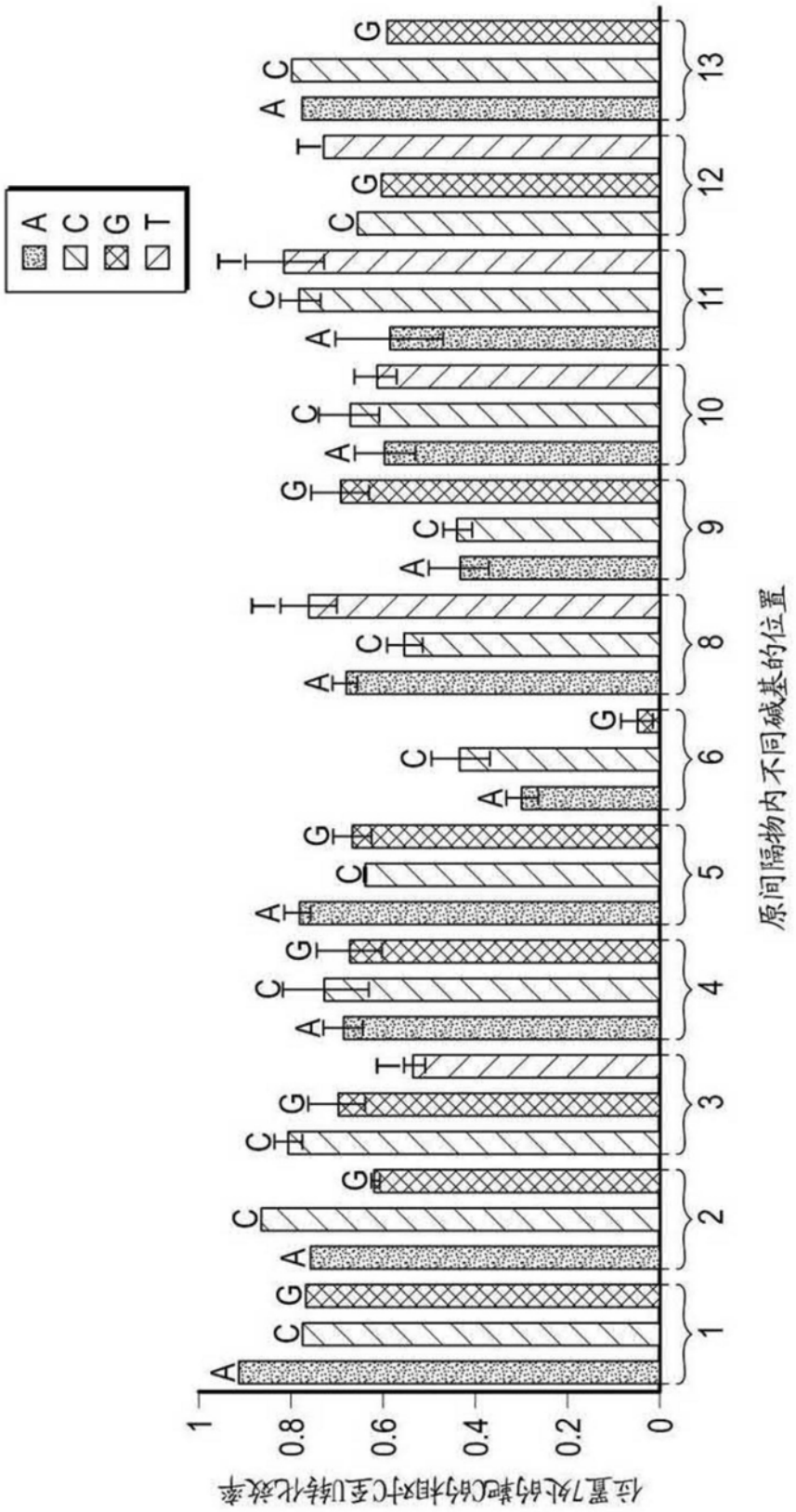


图12A

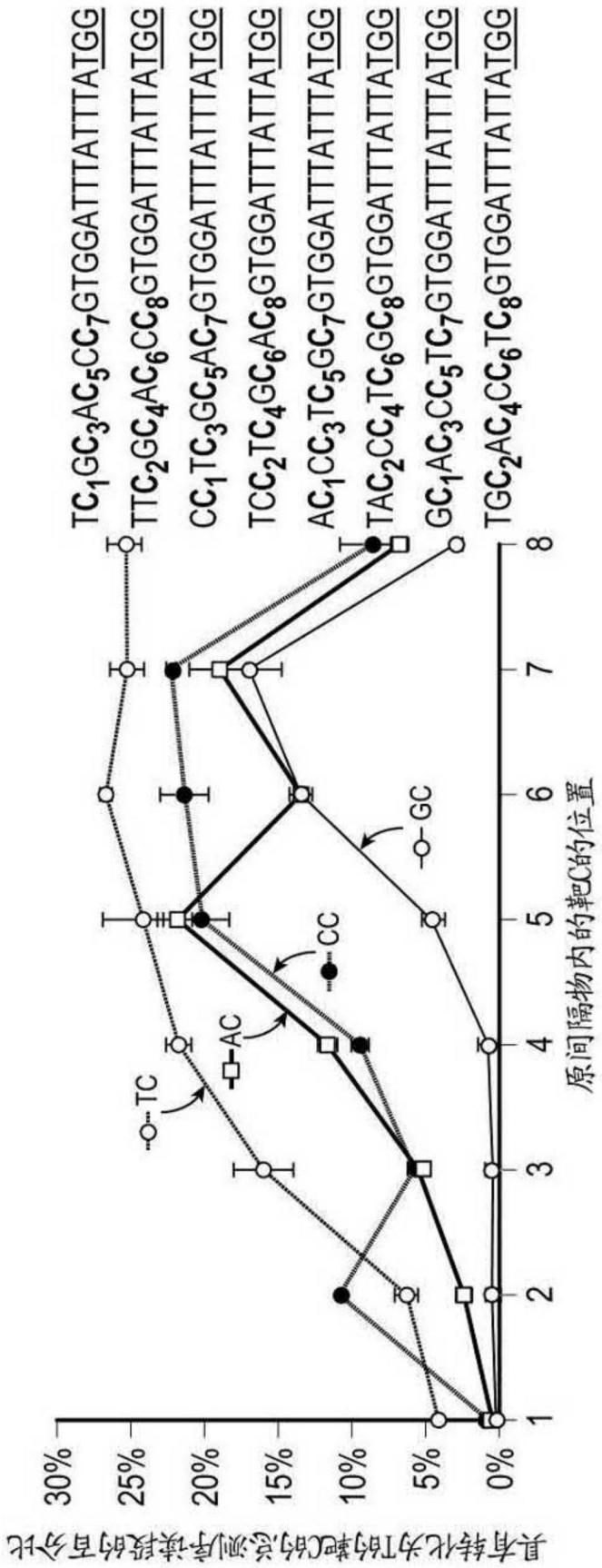

TC1GC3AC5CC7GTGGATTTATTTATGG
TTC2GC4AC6CC8GTGGATTTATTATGG
CC1TC3GC5AC7GTGGATTTATTTATGG
TCC2TC4GC6AC8GTGGATTTATTATGG
AC1CC3TC5GC7GTGGATTTATTTATGG
TAC2CC4TC6GC8GTGGATTTATTATGG
GC1AC3CC5TC7GTGGATTTATTTATGG
TGC2AC4CC6TC8GTGGATTTATTATGG
TC
AC
CC
GC
30%
25%
20%
15%
10%
5%
0%
1
2
3
4
5
6
7
8
原间隔物内的靶C的位置
具有转化为T的靶C的总测序读段的百分比

图12B

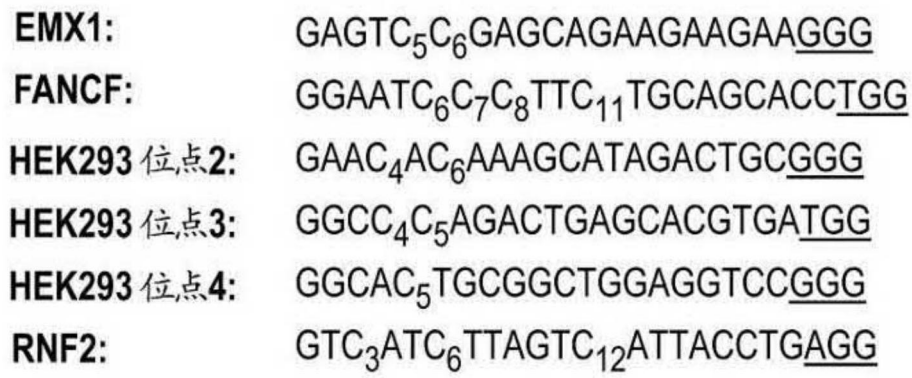

EMX1: GAGTC5C6GAGCAGAAGAAGAAGGG
FANCF: GGAATC6C7C8TTC11TGCAGCACCTGG
HEK293 位点2: GAAC4AC6AAAGCATAGACTGCGGG
HEK293 位点3: GGCC4C5AGACTGAGCACGTGATGG
HEK293 位点4: GGCAC5TGCGGCTGGAGGTCCGGG
RNF2: GTC3ATC6TTAGTC12ATTACCTGAGG

图13A

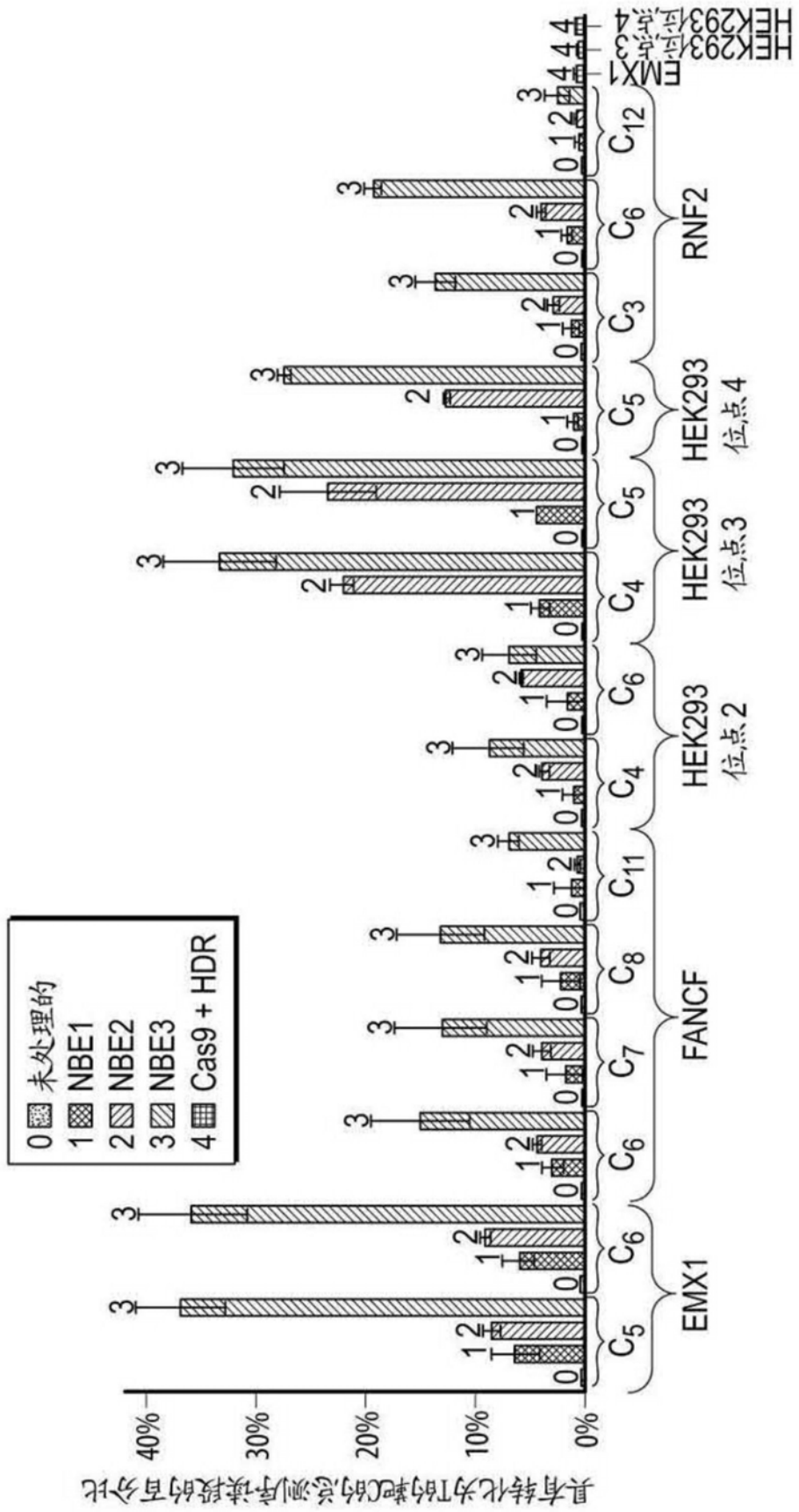

具有转化为T的靶C的总测序读段的百分比
40%
30%
20%
10%
0%
0 未处理的
1 NBE1
2 NBE2
3 NBE3
4 Cas9 + HDR
EMX1
FANCF
HEK293 位点2
HEK293 位点3
HEK293 位点4
RNF2
HEK293 位点1
HEK293 位点4
EMX1
C5
C6
C7
C8
C11
C4
C5
C3
C6
C12

图13B

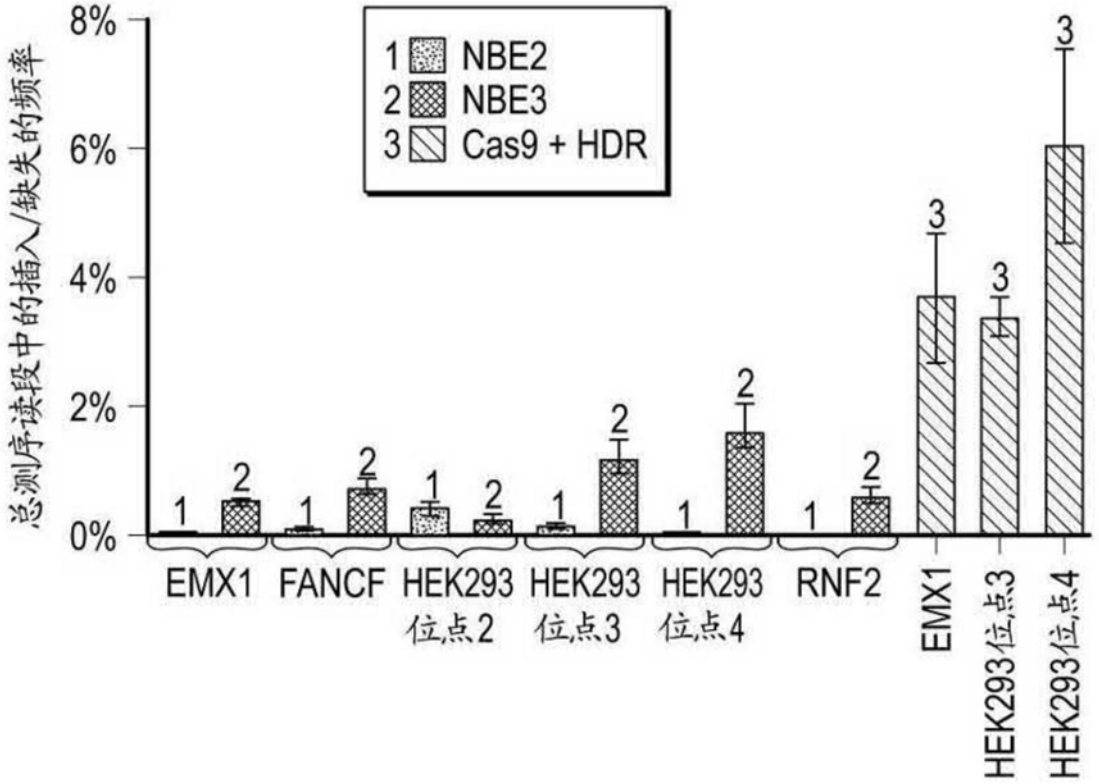
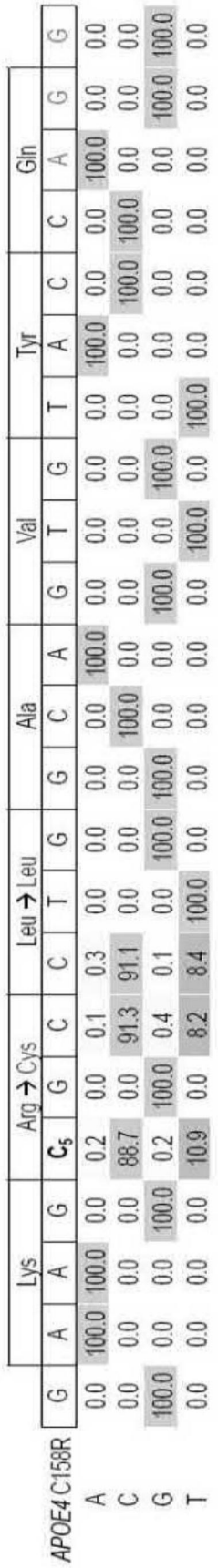
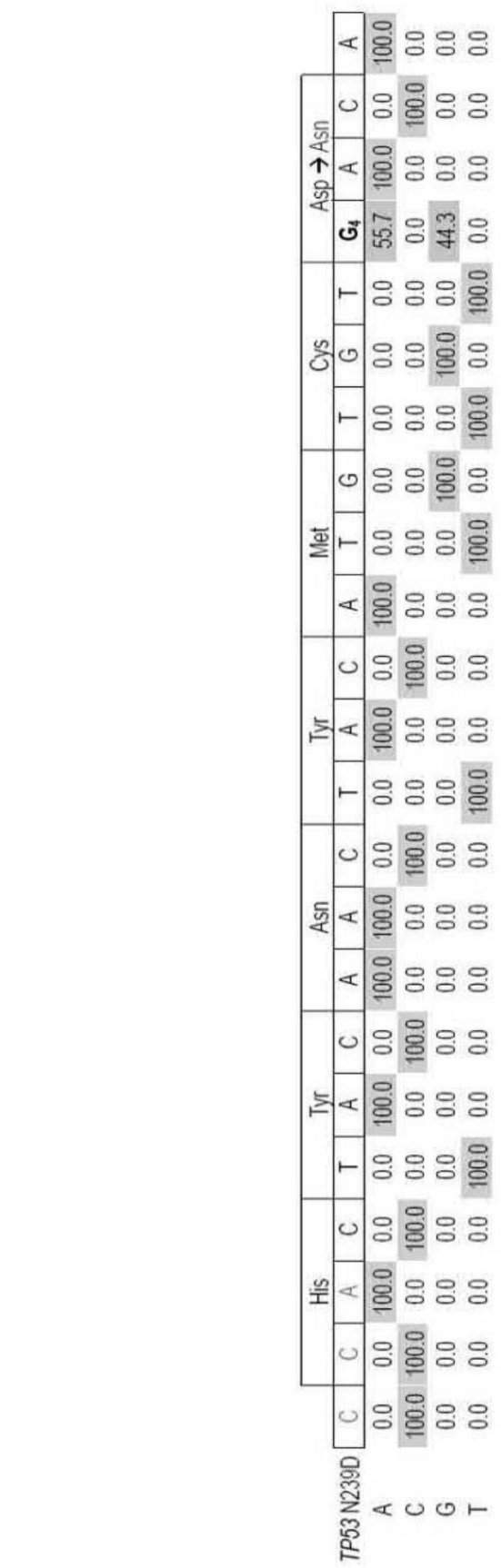

总测序读段中的插入/缺失的频率
8%
6%
4%
2%
0%
1 NBE2
2 NBE3
3 Cas9 + HDR
EMX1
FANCF
HEK293 位点2
HEK293 位点3
HEK293 位点4
RNF2
EMX1
HEK293位点3
HEK293位点4

图13C



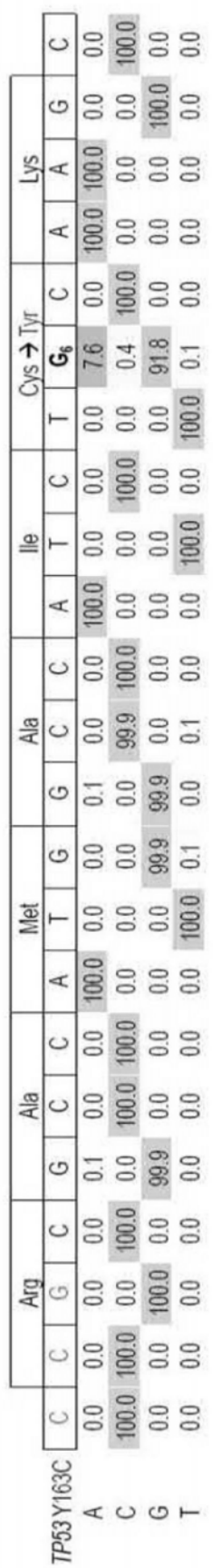
APOE4 C158R	Lys			Arg → Cys			Leu → Leu			Ala			Val			Tyr			Gln		
	G	A	G	C_5	G	C	C	T	G	G	C	A	G	T	G	A	C	C	A	G	G
A	0.0	100.0	100.0	0.0	0.2	0.0	0.1	0.3	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0
C	0.0	0.0	0.0	0.0	88.7	0.0	91.3	91.1	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0
G	100.0	0.0	0.0	100.0	0.2	100.0	0.4	0.1	0.0	100.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	100.0
T	0.0	0.0	0.0	0.0	10.9	0.0	8.2	8.4	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

图14A



TP53 N239D	His			Tyr			Asn			Tyr			Met			Cys			Asp → Asn			
	C	C	A	C	T	A	C	A	A	C	T	A	C	A	T	G	T	G	G_4	A	C	A
A	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	55.7	100.0	0.0	100.0
C	100.0	100.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
G	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	44.3	0.0	0.0	0.0
T	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0

图14B



TP53 Y163C	Arg			Ala			Met			Ala			Ile			Cys → Tyr			Lys		
	C	G	C	G	C	C	A	T	G	G	C	C	A	T	C	G_6	C	A	A	G	C
A	0.0	0.0	0.0	0.1	0.0	0.0	100.0	0.0	0.0	0.1	0.0	0.0	100.0	0.0	0.0	7.6	0.0	100.0	100.0	0.0	0.0
C	100.0	100.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	99.9	100.0	0.0	0.0	100.0	0.4	100.0	0.0	0.0	0.0	100.0
G	0.0	0.0	100.0	99.9	0.0	0.0	0.0	0.0	99.9	99.9	0.0	0.0	0.0	0.0	0.0	91.8	0.0	0.0	0.0	100.0	0.0
T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.1	0.0	0.1	0.0	0.0	100.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0

图14C

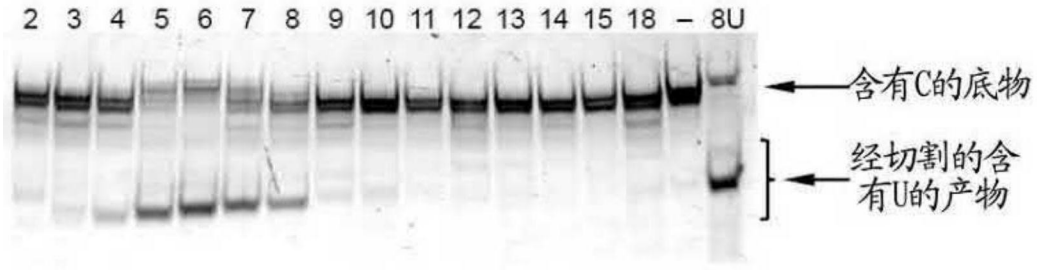


图15A

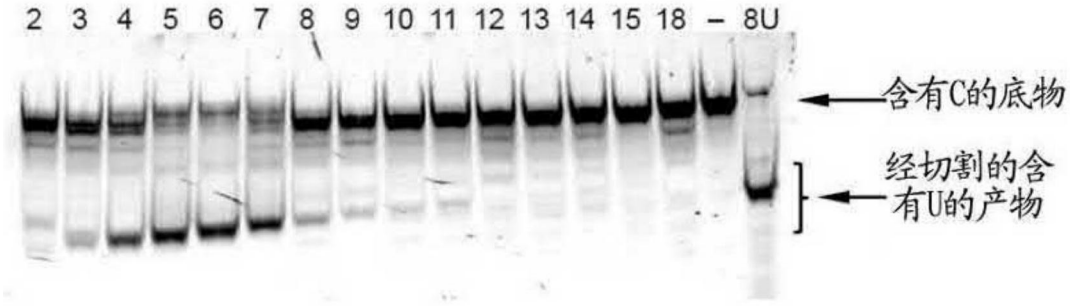


图15B

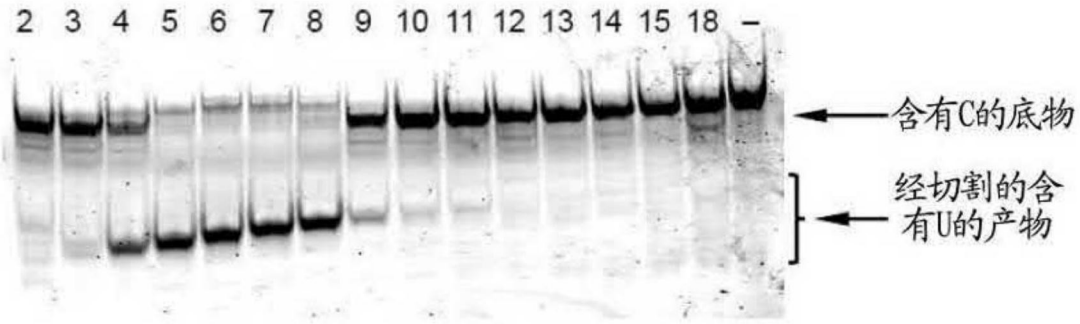


图15C

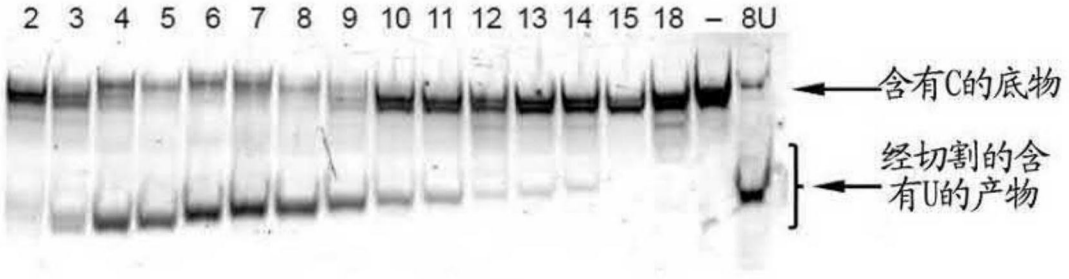


图15D

APOE4 Cys112Arg:	5'-GGAGGACGTGC$_{11}$GCGGCCGCCTGG
APOE4 Cys158Arg:	5'-GAAGC$_5$GCCTGGCAGTGTACCAGG
CTNNB1 Thr41Ala:	5'-CTGTGGC$_7$AGTGGCACCAGAATGG
HRAS Gln61Arg:	5'-CCTCCC$_6$GGCCGGCGGTATCCAGG
p53 Tyr163Cys:	5'-GCTTGC$_6$AGATGGCCATGGCGCGG
p53 Tyr236Cys:	5'-ACACATGC$_8$AGTTGTAGTGGATGG
p53 Asn239Asp:	5'-TGTC$_4$ACACATGTAGTTGTAGTGG

图16A

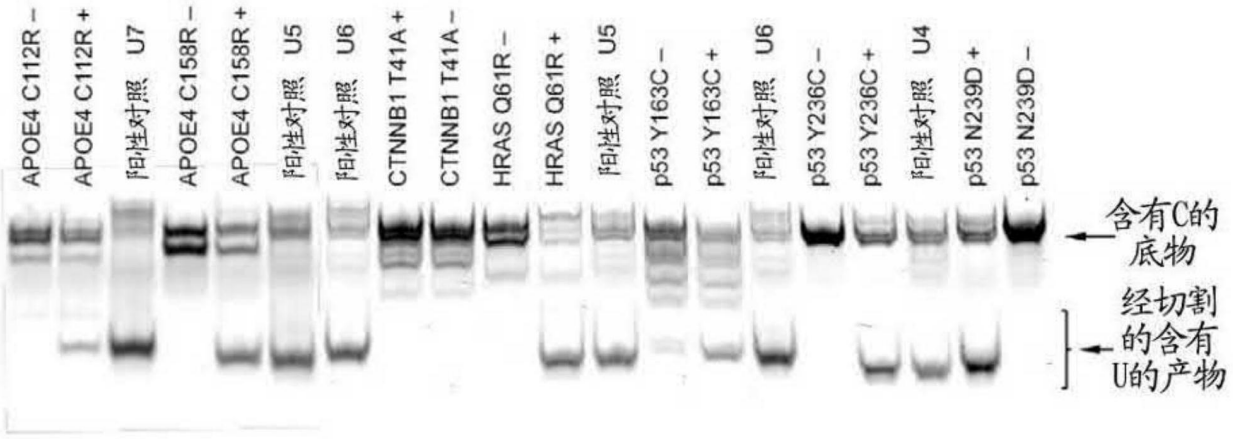


图16B

原间隔物和PAM序列: 5'-TTCCCCCCCCGATTTATTTATGG-3'

序列	总读段的%
CCCCCCCC	62.4
TTTTTTCC	18.2
TTTTTTTC	13.4
TTTTTTTT	3.3
TCCCCCCC	0.8
CCCCTTCC	0.3
CCCTTTCC	0.3
TTTTTCCC	0.3
CCCCTCCC	0.3

图17

EMX1: GAGTC$_5$C$_6$GAGCAGAAGAAGAAGGG
FANCF: GGAATC$_6$C$_7$C$_8$TTC$_{11}$TGCAGCACCTGG
HEK293位点2: GAAC$_4$AC$_6$AAAGCATAGACTGCGGG
HEK293位点3: GGCC$_4$C$_5$AGACTGAGCACGTGATGG
HEK293位点4: GGCAC$_5$TGCGGCTGGAGGTCCGGG
RNF2: GTC$_3$ATC$_6$TTAGTCATTACCTGAGG

图18A

EMX1	C_5	C_6
NBE1	6.2%	6.5%
NBE1 + UGI	9.7%	10.1%
NBE2	8.0%	8.7%

图18B

FANCF	C_6	C_7	C_8	C_{10}
NBE1	3.7%	3.2%	3.4%	2.4%
NBE1 + UGI	7.5%	7.6%	7.5%	1.6%
NBE2	4.7%	4.6%	4.6%	0.8%

图18C

HEK293位点2	C_4	C_6
NBE1	0.4%	0.4%
NBE1 + UGI	1.6%	2.6%
NBE2	3.4%	5.9%

图18D

HEK293 位点 3	C_4	C_5
NBE1	2.0%	1.9%
NBE1 + UGI	6.5%	6.7%
NBE2	10.0%	12.5%

图18E

HEK293 位点 4	C_5
NBE1	1.4%
NBE1 + UGI	5.4%
NBE2	8.2%

图18F

RNF2	C_3	C_6
NBE1	0.7%	1.4%
NBE1 + UGI	3.4%	3.9%
NBE2	2.5%	3.7%

图18G

非原间隔物 Cs	C	T
未处理的	99.93%	0.03%
NBE1	99.95%	0.03%
NBE1 + UGI	99.91%	0.06%
NBE2	99.92%	0.04%

图18H

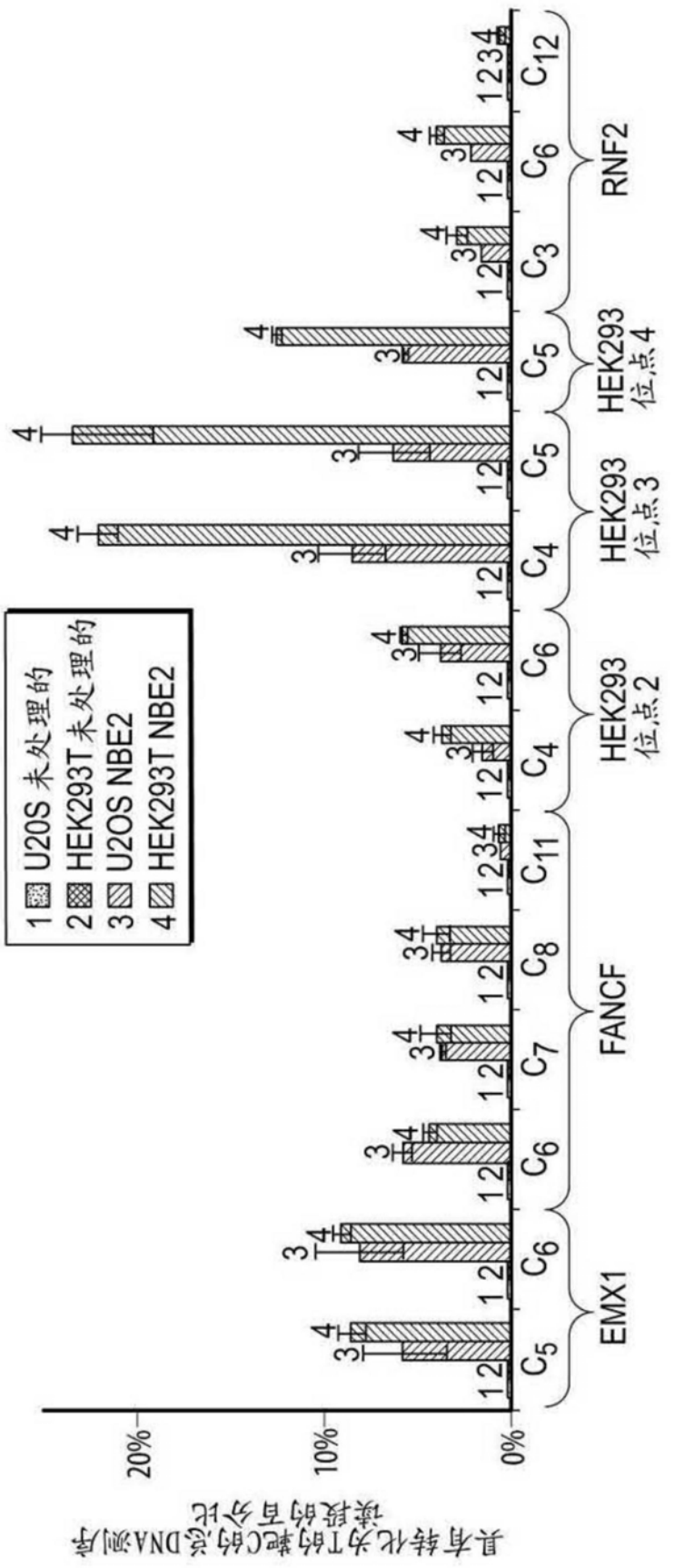

具有转化为T的靶C的总DNA测序
读段的百分比
20%
10%
0%
1 U2OS 未处理的
2 HEK293T 未处理的
3 U2OS NBE2
4 HEK293T NBE2
EMX1
FANCF
HEK293 位点2
HEK293 位点3
HEK293 位点4
RNF2

图19

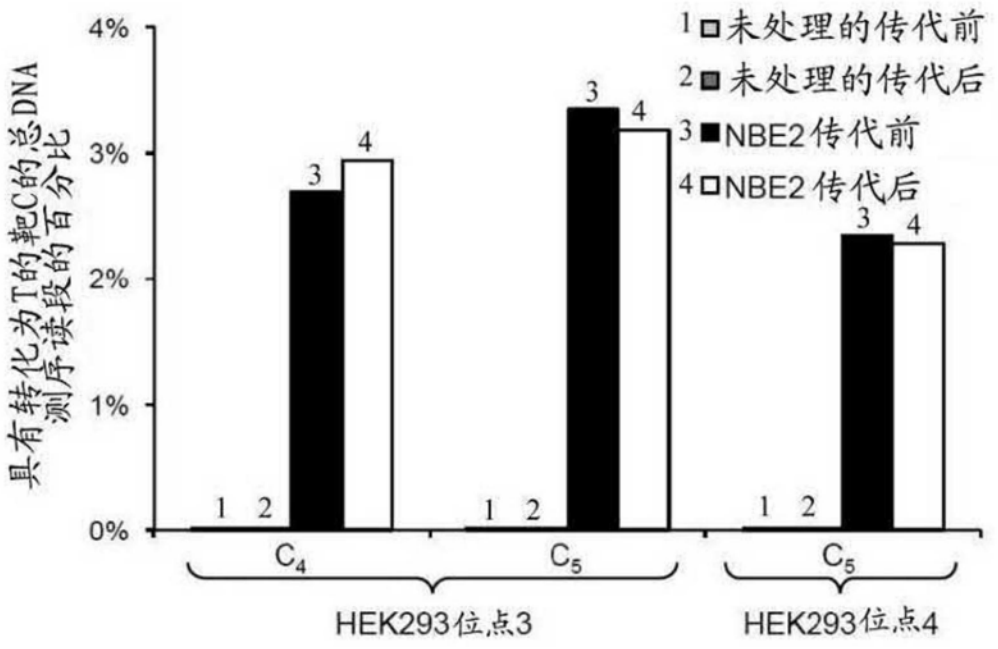

具有转化为T的靶C的总DNA测序读段的百分比
4%
3%
2%
1%
0%
1 未处理的传代前
2 未处理的传代后
3 NBE2传代前
4 NBE2传代后
C4
C5
C5
HEK293位点3
HEK293位点4

图20

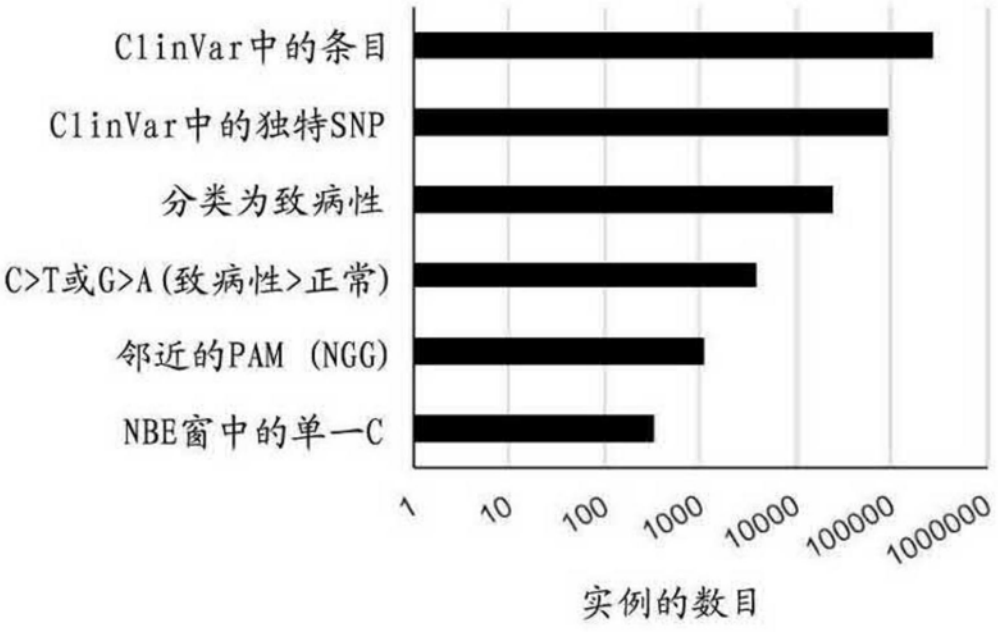

ClinVar中的条目
ClinVar中的独特SNP
分类为致病性
C>T或G>A(致病性>正常)
邻近的PAM (NGG)
NBE窗中的单一C
1
10
100
1000
10000
100000
1000000
实例的数目

图21

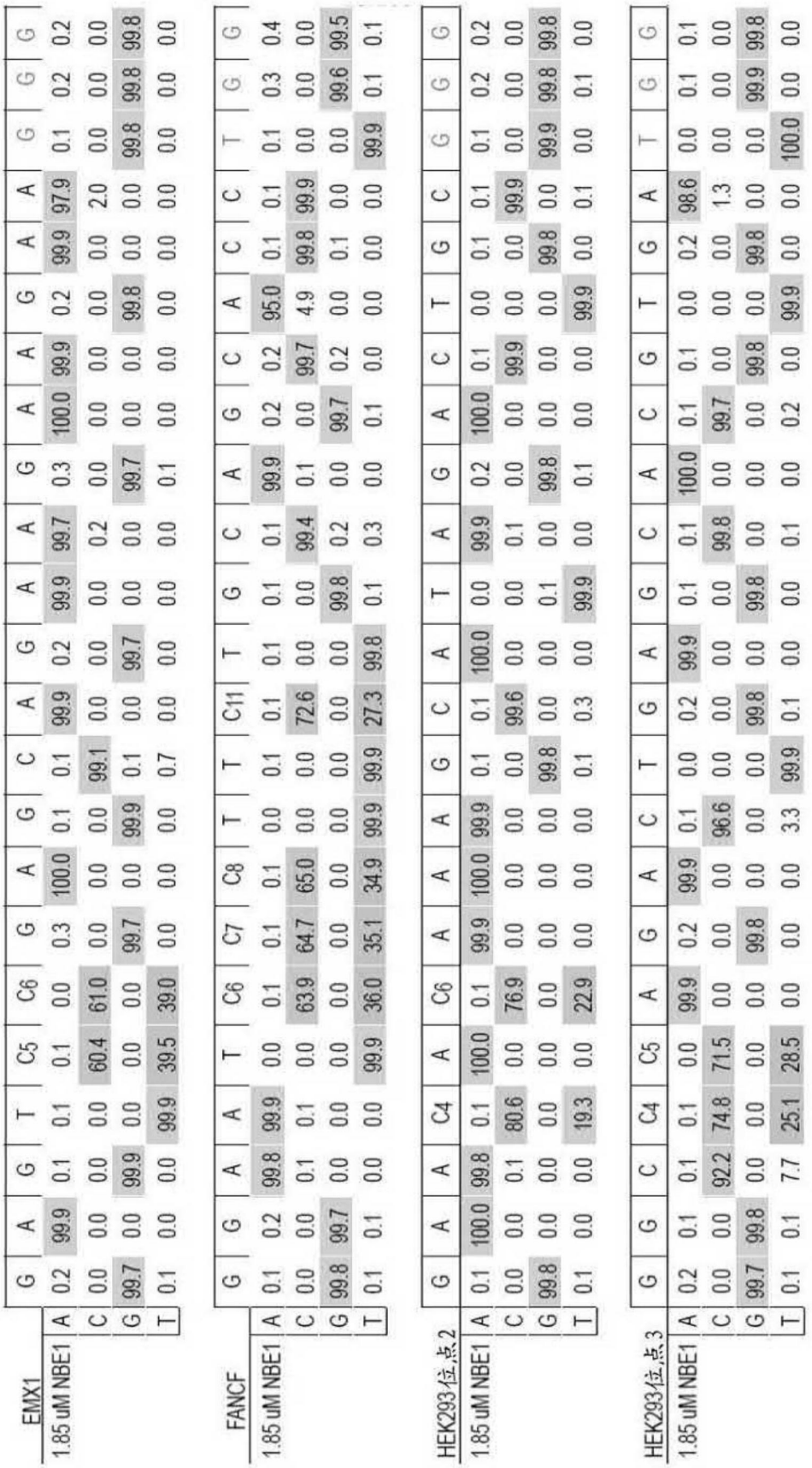

EMX1
1.85 uM NBE1
FANCF
1.85 uM NBE1
HEK293位点2
1.85 uM NBE1
HEK293位点3
1.85 uM NBE1

图22

HEK293位点3		G	G	C	C4	C5	A	G	A	C	T	G	A	T	C	G	C	A	G	C	G	A	T	G	G
1.85 uM NBE1	A	0.2	0.1	0.1	0.1	0.0	99.9	0.2	99.9	0.1	0.0	0.2	99.9	0.1	0.1	100.0	0.1	0.1	0.0	0.2	98.6	0.0	0.1	0.1
	C	0.0	0.0	92.2	74.8	71.5	0.0	0.0	0.0	96.6	0.0	0.0	0.0	0.0	99.8	0.0	99.7	0.0	0.0	0.0	1.3	0.0	0.0	0.0
	G	99.7	99.8	0.0	0.0	0.0	0.0	99.8	0.0	0.0	0.0	99.8	0.0	0.0	0.0	0.0	0.0	99.8	0.0	99.8	0.0	0.0	99.9	99.8
	T	0.1	0.1	7.7	25.1	28.5	0.0	0.0	0.0	3.3	99.9	0.1	0.0	0.0	0.1	0.0	0.2	0.0	99.9	0.0	0.0	100.0	0.0	0.0

HEK293位点4		G	G	C	A	C5	T	G	C	G	T	G	C	T	A	G	G	A	G	G	T	G	G	G	G
1.85 uM NBE1	A	0.3	0.2	0.1	100.0	0.1	0.0	0.2	0.1	0.2	0.2	0.1	0.0	0.3	0.5	100.0	0.1	0.1	0.0	0.1	0.1	0.1	0.2	0.2
	C	0.0	0.0	98.8	0.0	60.1	0.0	0.0	97.0	0.0	0.0	99.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	99.7	99.7	0.0	0.0	0.0	0.0	99.8	0.0	99.8	99.8	0.0	0.0	99.7	99.4	0.0	99.8	99.8	0.1	99.8	99.8	99.8	99.8	99.7
	T	0.0	0.1	1.1	0.0	39.8	100.0	0.1	2.9	0.0	0.1	0.5	100.0	0.0	0.1	0.0	0.0	0.0	99.9	0.0	0.1	0.1	0.0	0.0

RNF2		G	T	C3	A	T	C6	T	T	A	G	T	C12	A	T	T	A	C	C	T	G	A	G	G
1.85 uM NBE1	A	0.1	0.0	0.1	99.9	0.0	0.0	0.0	0.0	99.9	0.1	0.0	0.1	99.9	0.0	0.0	99.9	0.1	0.0	0.0	0.1	98.6	0.0	0.0
	C	0.0	0.0	59.1	0.0	0.0	57.8	0.0	0.0	0.1	0.0	0.0	86.4	0.0	0.0	0.0	0.1	99.9	99.9	0.0	0.0	1.4	0.0	0.0
	G	99.9	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.8	0.0	99.9	99.9
	T	0.0	99.9	40.8	0.0	100.0	42.1	100.0	100.0	0.0	0.1	99.9	13.4	0.0	100.0	99.9	0.0	0.0	0.0	99.9	0.1	0.0	0.1	0.0

图22续

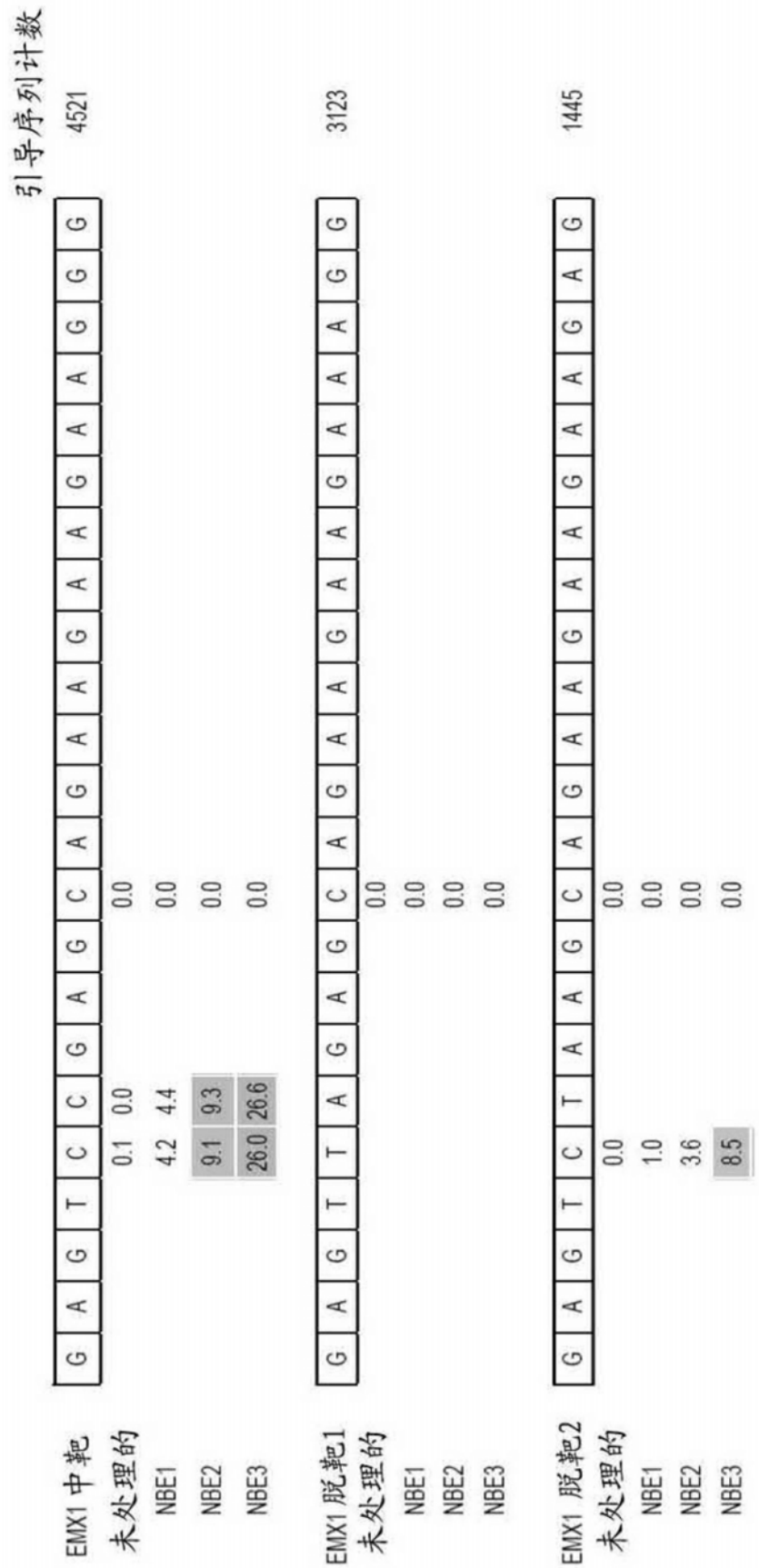

引导序列计数
EMX1 中靶
G A G T C C G A G C A G A A A G A A G G G
4521
未处理的
0.1 0.0 0.0
NBE1
4.2 4.4 0.0
NBE2
9.1 9.3 0.0
NBE3
26.0 26.6 0.0
EMX1 脱靶1
G A G T T A G A G C A G A A A G A A G G G
3123
未处理的
0.0
NBE1
0.0
NBE2
0.0
NBE3
0.0
EMX1 脱靶2
G A G T C T A A G C A G A A A G A A G A G
1445
未处理的
0.0 0.0
NBE1
1.0 0.0
NBE2
3.6 0.0
NBE3
8.5 0.0

图23

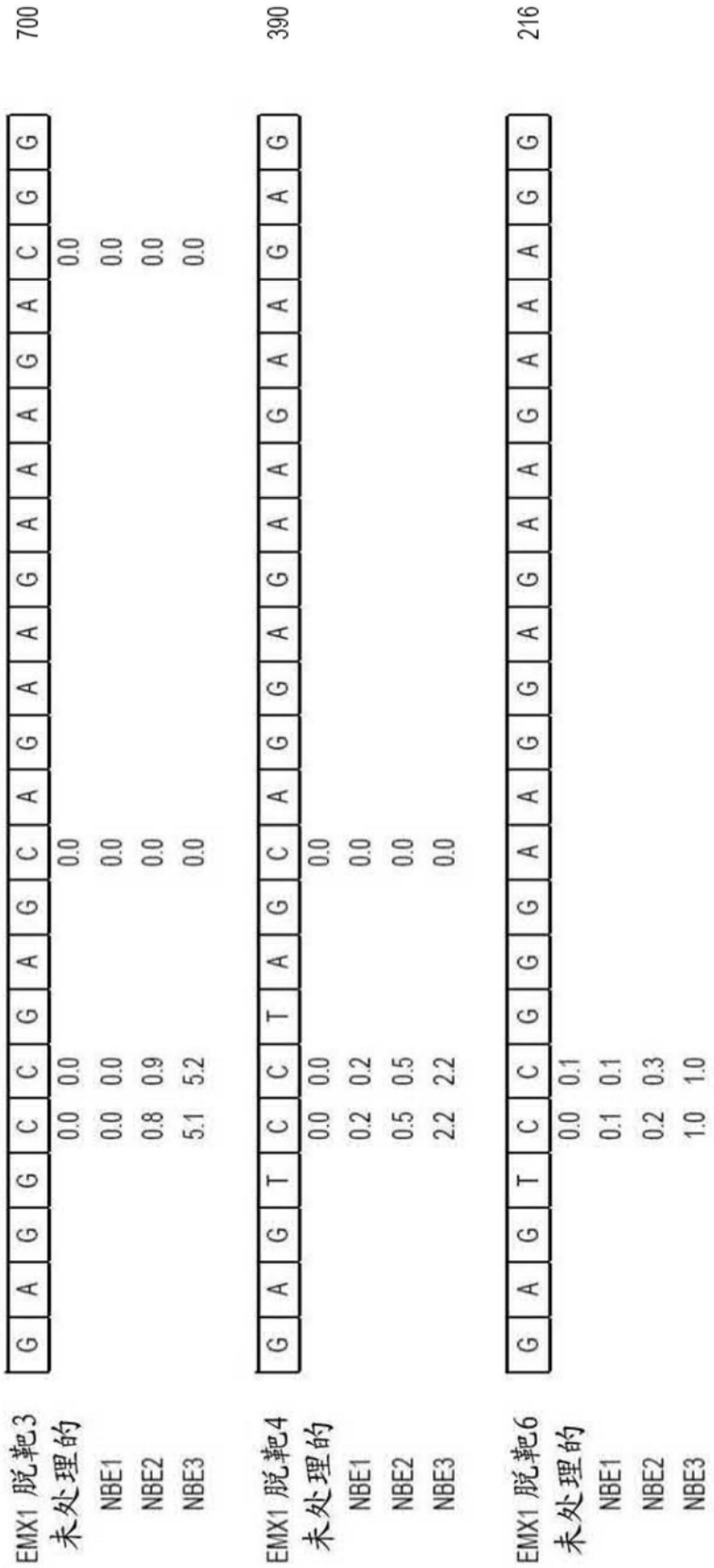

700
G A G G C C G A G C A G A A A A G A C G G
EMX1 脱靶3
未处理的 0.0 0.0 0.0 0.0
NBE1 0.0 0.0 0.0 0.0
NBE2 0.8 0.9 0.0 0.0
NBE3 5.1 5.2 0.0 0.0
390
G A G T C C T A G C A G G A G A A G A G
EMX1 脱靶4
未处理的 0.0 0.0 0.0
NBE1 0.2 0.2 0.0
NBE2 0.5 0.5 0.0
NBE3 2.2 2.2 0.0
216
G A G T C C G G G A G G A G A A A G G
EMX1 脱靶6
未处理的 0.0 0.1
NBE1 0.1 0.1
NBE2 0.2 0.3
NBE3 1.0 1.0

图23续

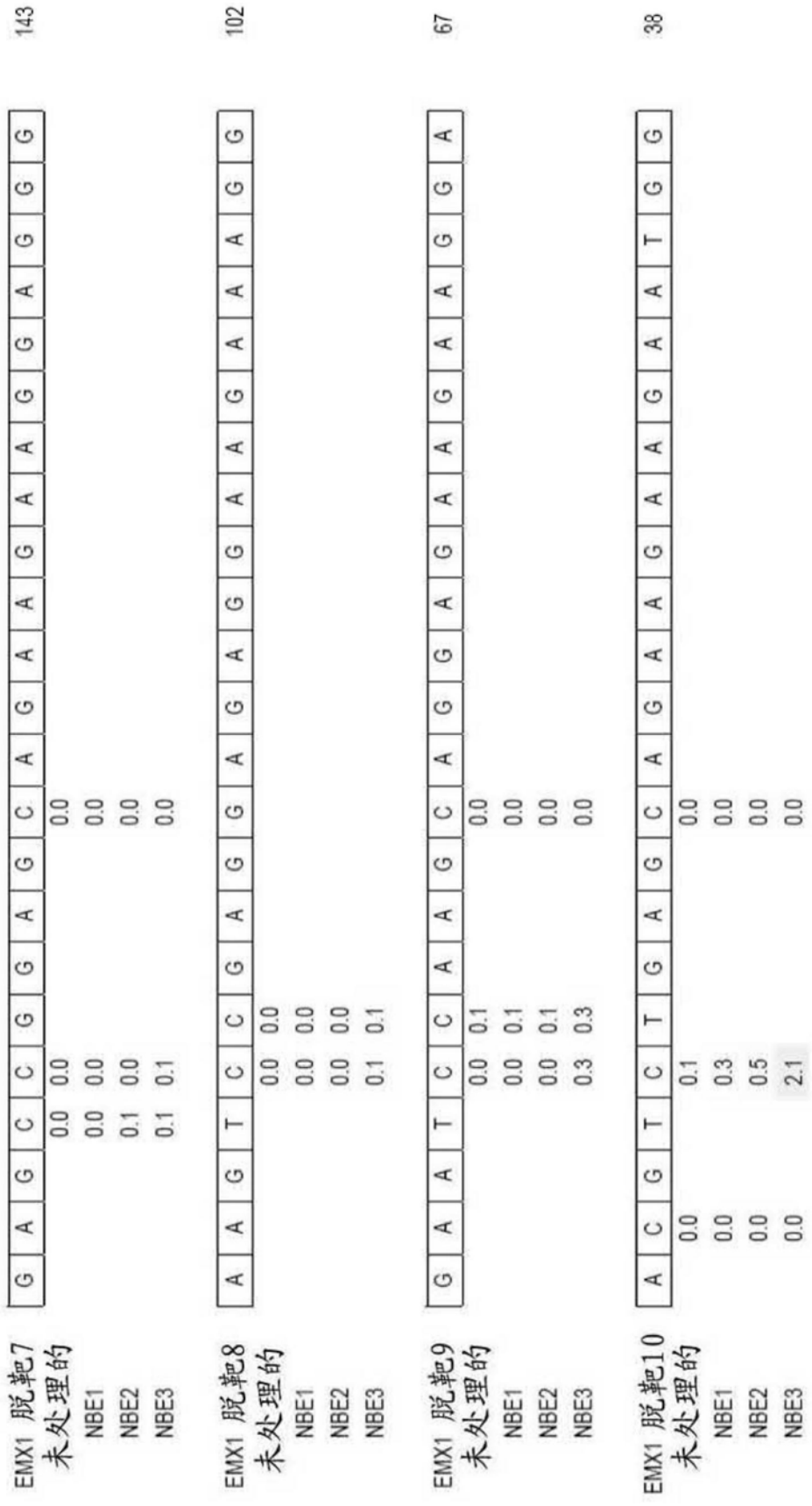
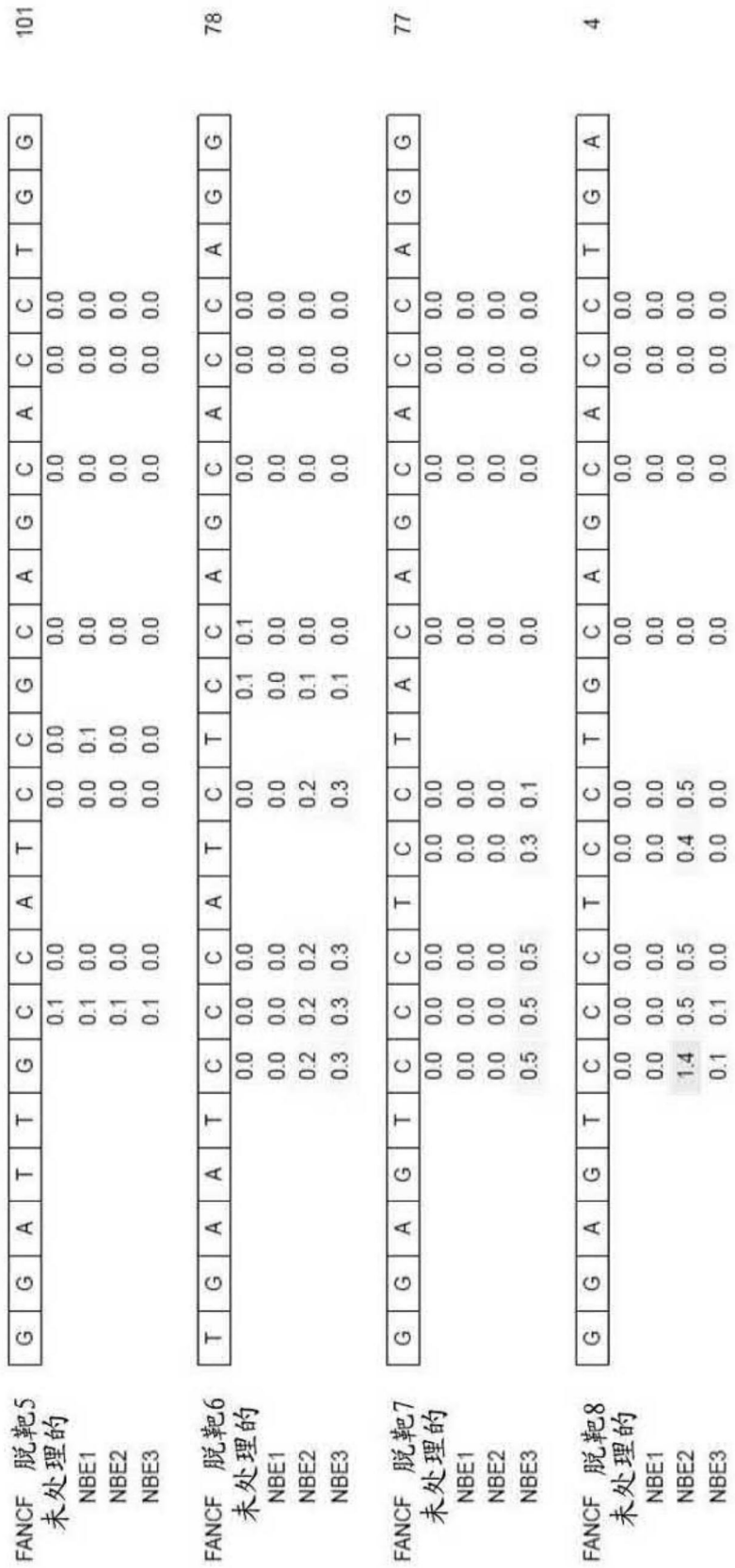

143
102
67
38
EMX1 脱靶7 未处理的
NBE1
NBE2
NBE3
EMX1 脱靶8 未处理的
NBE1
NBE2
NBE3
EMX1 脱靶9 未处理的
NBE1
NBE2
NBE3
EMX1 脱靶10 未处理的
NBE1
NBE2
NBE3

图23续

引导序列计数
FANCF 中靶 未处理的 4816
NBE1
NBE2
NBE3
FANCF 脱靶1 未处理的 2099
NBE1
NBE2
NBE3
FANCF 脱靶2 未处理的 524
NBE1
NBE2
NBE3
FANCF 脱靶3 未处理的 150
NBE1
NBE2
NBE3
FANCF 脱靶4 未处理的 125
NBE1
NBE2
NBE3

图24



	G	G	A	T	T	G	C	C	A	T	C	C	G	C	A	G	C	A	C	T	G	G
FANCF 脱靶5 未处理的							0.1	0.0			0.0	0.0		0.0			0.0		0.0	0.0		101
NBE1							0.1	0.0			0.0	0.1		0.0			0.0		0.0	0.0		
NBE2							0.1	0.0			0.0	0.0		0.0			0.0		0.0	0.0		
NBE3							0.1	0.0			0.0	0.0		0.0			0.0		0.0	0.0		
	T	G	A	A	T	C	C	C	A	T	C	T	C	C	A	G	C	A	C	A	G	G
FANCF 脱靶6 未处理的						0.0	0.0	0.0			0.0		0.1	0.1			0.0		0.0	0.0		78
NBE1						0.0	0.0	0.0			0.0		0.0	0.0			0.0		0.0	0.0		
NBE2						0.2	0.2	0.2			0.2		0.1	0.0			0.0		0.0	0.0		
NBE3						0.3	0.3	0.3			0.3		0.1	0.0			0.0		0.0	0.0		
	G	G	A	G	T	C	C	C	T	C	C	T	A	C	A	G	C	A	C	A	G	G
FANCF 脱靶7 未处理的						0.0	0.0	0.0		0.0	0.0			0.0			0.0		0.0	0.0		77
NBE1						0.0	0.0	0.0		0.0	0.0			0.0			0.0		0.0	0.0		
NBE2						0.0	0.0	0.0		0.0	0.0			0.0			0.0		0.0	0.0		
NBE3						0.5	0.5	0.5		0.3	0.1			0.0			0.0		0.0	0.0		
	G	G	A	G	T	C	C	C	T	C	C	T	G	C	A	G	C	A	C	T	G	A
FANCF 脱靶8 未处理的						0.0	0.0	0.0		0.0	0.0			0.0			0.0		0.0	0.0		4
NBE1						0.0	0.0	0.0		0.0	0.0			0.0			0.0		0.0	0.0		
NBE2						1.4	0.5	0.5		0.4	0.5			0.0			0.0		0.0	0.0		
NBE3						0.1	0.1	0.0		0.0	0.0			0.0			0.0		0.0	0.0		

图24续

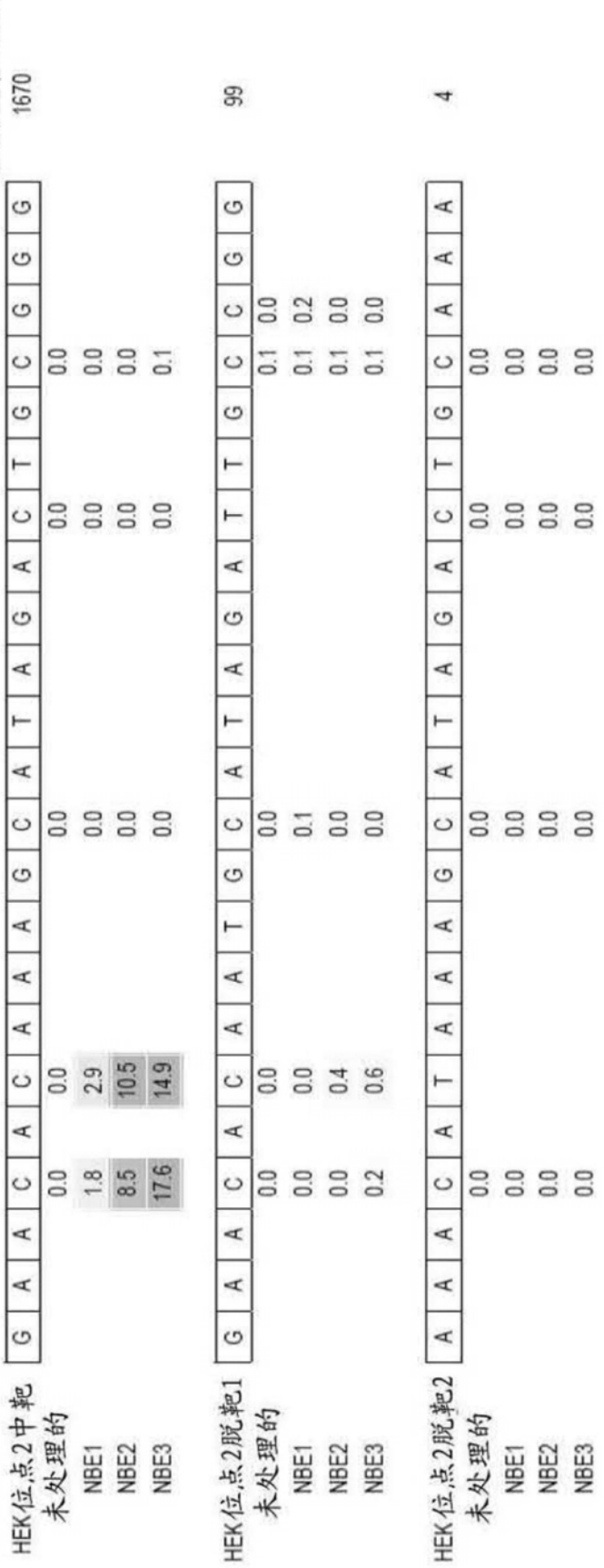

引导序列计数
1670
99
4
HEK位点2中靶
未处理的
NBE1
NBE2
NBE3
HEK位点2脱靶1
未处理的
NBE1
NBE2
NBE3
HEK位点2脱靶2
未处理的
NBE1
NBE2
NBE3

图25

引导序列计数																						
																						2074
HEK位点3中靶未处理的	G	G	C	C	C	A	G	A	C	T	G	A	G	C	A	C	G	T	G	A	T	G	G
			0.1	0.1	0.1				0.0					0.0		0.0							
NBE1			0.1	4.8	4.3				0.0					0.0		0.0							
NBE2			3.6	21.3	20.4				0.0					0.0		0.0							
NBE3			0.5	33.4	33.1				0.5					0.0		0.2							
HEK位点3脱靶1未处理的	C	A	C	C	C	A	G	A	C	T	G	A	G	C	A	C	G	T	G	C	T	G	G
	0.0		0.0	0.0	0.0				0.0					0.0		0.0				0.0			
NBE1	0.0		0.0	0.0	0.0				0.0					0.0		0.0				0.0			
NBE2	0.0		0.0	0.0	0.0				0.0					0.0		0.0				0.0			
NBE3	0.0		0.0	0.0	0.0				0.0					0.0		0.0				0.0			
HEK位点3脱靶2未处理的	G	A	C	A	C	A	G	A	C	T	G	G	G	C	A	C	G	T	G	A	G	G	G
			0.0		0.0				0.0					0.0		0.0							
NBE1			0.0		0.0				0.0					0.0		0.0							
NBE2			0.0		0.0				0.0					0.0		0.0							
NBE3			0.0		0.0				0.0					0.0		0.0							
HEK位点3脱靶3未处理的	A	G	C	T	C	A	G	A	C	T	G	A	G	C	A	A	G	T	G	A	G	G	G
			0.0		0.0				0.0					0.0									
NBE1			0.0		0.0				0.0					0.0									
NBE2			0.0		0.0				0.0					0.0									
NBE3			0.0		0.9				0.0					0.0									
HEK位点3脱靶4未处理的	A	G	A	C	C	A	G	A	C	T	G	A	G	C	A	A	G	A	G	A	G	G	G
				0.0	0.0				0.0					0.0									
NBE1				0.0	0.0				0.0					0.1									
NBE2				0.0	0.0				0.0					0.0									
NBE3				0.0	0.0				0.0					0.0									
HEK位点3脱靶5未处理的	G	A	G	C	C	A	G	A	A	T	G	A	G	C	A	C	G	T	G	A	G	G	G
			0.0	0.0					0.0					0.0		0.0							
NBE1			0.0	0.0					0.0					0.0		0.0							
NBE2			0.0	0.0					0.0					0.0		0.0							
NBE3			0.0	0.0					0.0					0.0		0.0							

图26

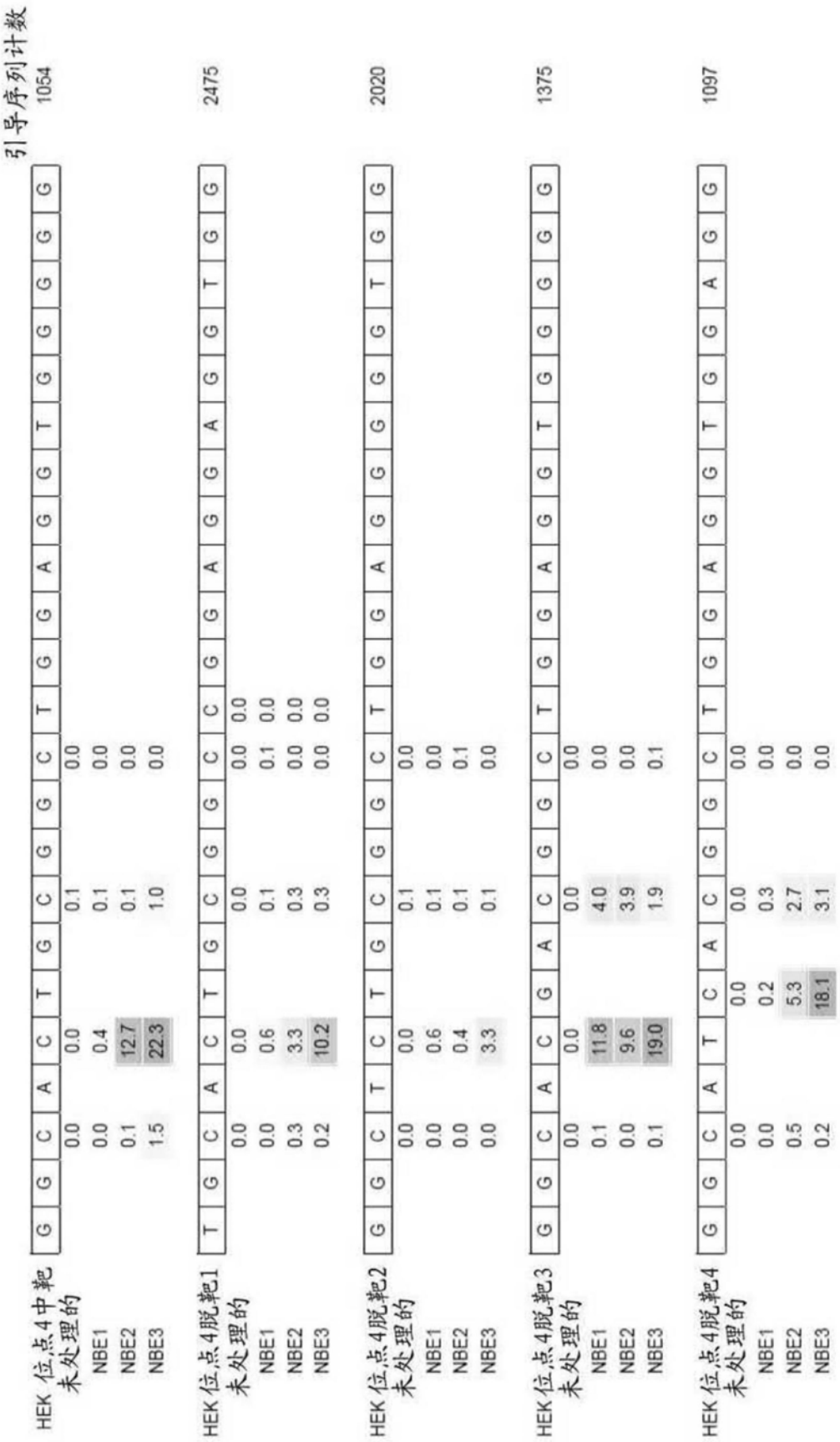

引导序列计数
HEK 位点4中靶 未处理的 1054
G G C A C T G C G G C T G G G A G G T G G G G G
NBE1 0.0 0.0 0.1 0.0
NBE2 0.0 0.4 0.1 0.0
NBE3 0.1 12.7 0.1 0.0
1.5 22.3 1.0 0.0
HEK 位点4脱靶1 未处理的 2475
T G C A C T G C G G C C G G G A G G G A G G T G G
0.0 0.0 0.0 0.0 0.0
NBE1 0.0 0.6 0.1 0.1 0.0
NBE2 0.3 3.3 0.3 0.0 0.0
NBE3 0.2 10.2 0.3 0.0 0.0
HEK 位点4脱靶2 未处理的 2020
G G C T C T G C G G C T G G G A G G G T G G G
0.0 0.0 0.1 0.0
NBE1 0.0 0.6 0.1 0.0
NBE2 0.0 0.4 0.1 0.1
NBE3 0.0 3.3 0.1 0.0
HEK 位点4脱靶3 未处理的 1375
G G C A C G A C G G C T G G G A G G T G G G G G
0.0 0.0 0.0 0.0
NBE1 0.1 11.8 4.0 0.0
NBE2 0.0 9.6 3.9 0.0
NBE3 0.1 19.0 1.9 0.1
HEK 位点4脱靶4 未处理的 1097
G G C A T C A C G G C T G G G A G G T G G G A G G
0.0 0.0 0.0 0.0
NBE1 0.0 0.2 0.3 0.0
NBE2 0.5 5.3 2.7 0.0
NBE3 0.2 18.1 3.1 0.0

图27

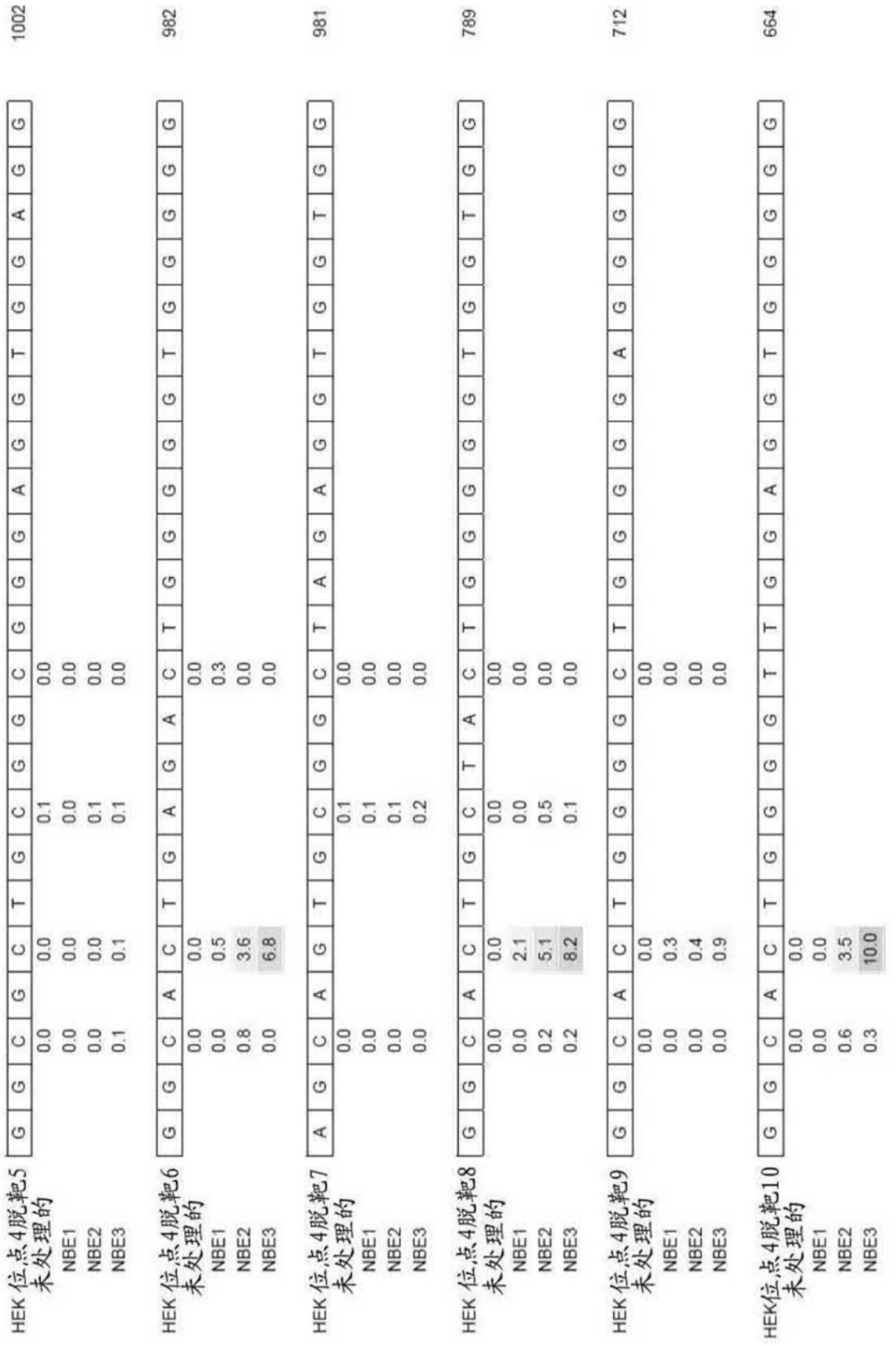

1002
HEK 位点4脱靶5 未处理的
NBE1
NBE2
NBE3
G G C G C T G C G G C G G A G G G A G G
0.0 0.0 0.1 0.0
0.0 0.0 0.0 0.1 0.0 0.0
0.0 0.0 0.1 0.0 0.0 0.0
0.1 0.1 0.1 0.1 0.0 0.0
982
HEK 位点4脱靶6 未处理的
NBE1
NBE2
NBE3
G G C A C T G A G A C T G G G G G T G G G G
0.0 0.0 0.0
0.0 0.5 0.3
0.8 3.6 0.0
0.0 6.8 0.0
981
HEK 位点4脱靶7 未处理的
NBE1
NBE2
NBE3
A G C A G T G C G G C T A G A G G T G T G G
0.0 0.1 0.0
0.0 0.1 0.0
0.0 0.1 0.0
0.0 0.2 0.0
789
HEK 位点4脱靶8 未处理的
NBE1
NBE2
NBE3
G G C A C T G C T A C T G G G G T G T G G G
0.0 0.0 0.0 0.0
0.0 2.1 0.0 0.0 0.0
0.2 5.1 0.5 0.0
0.2 8.2 0.1 0.0
712
HEK 位点4脱靶9 未处理的
NBE1
NBE2
NBE3
G G C A C T G G G G C T G G G G A G G G G G
0.0 0.0 0.0
0.0 0.3 0.0
0.0 0.4 0.0
0.0 0.9 0.0
664
HEK位点4脱靶10 未处理的
NBE1
NBE2
NBE3
G G C A C T G G G G T T G G G A G G T G G G G G
0.0 0.0
0.0 0.0
0.6 3.5
0.3 10.0

图27续

非原间隔物 Cs	C (%)	T (%)
未处理的	99.94	0.04
NBE1	99.92	0.05
NBE2	99.92	0.05
NBE3	99.94	0.03

图28

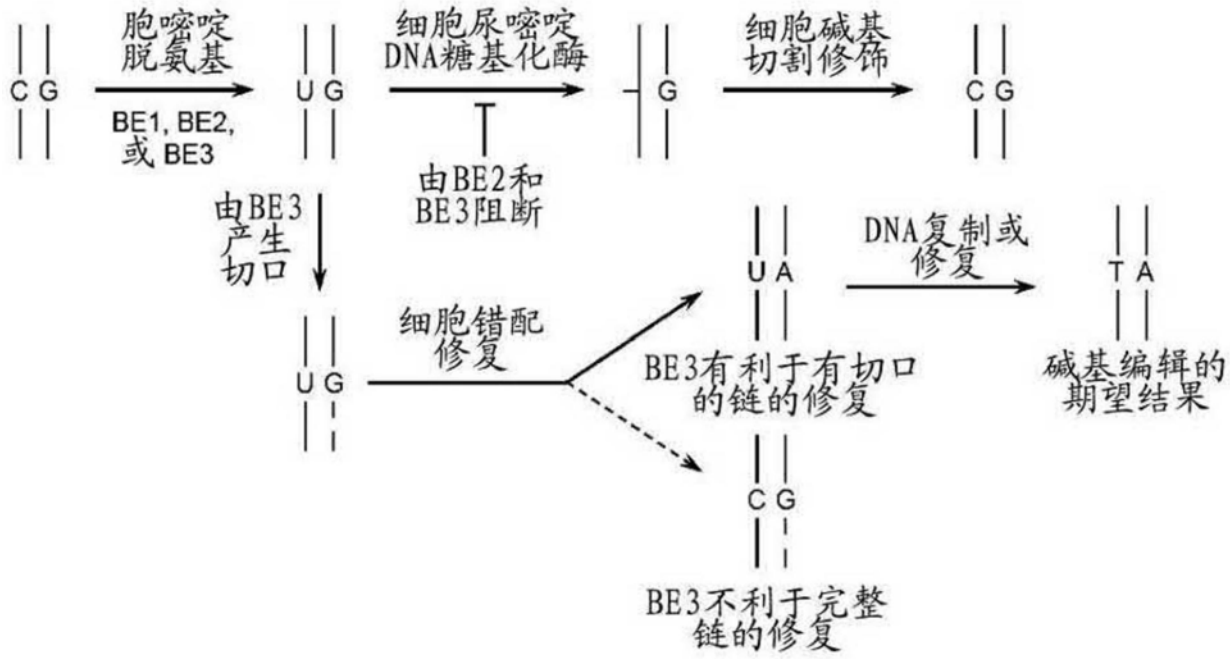


图29A

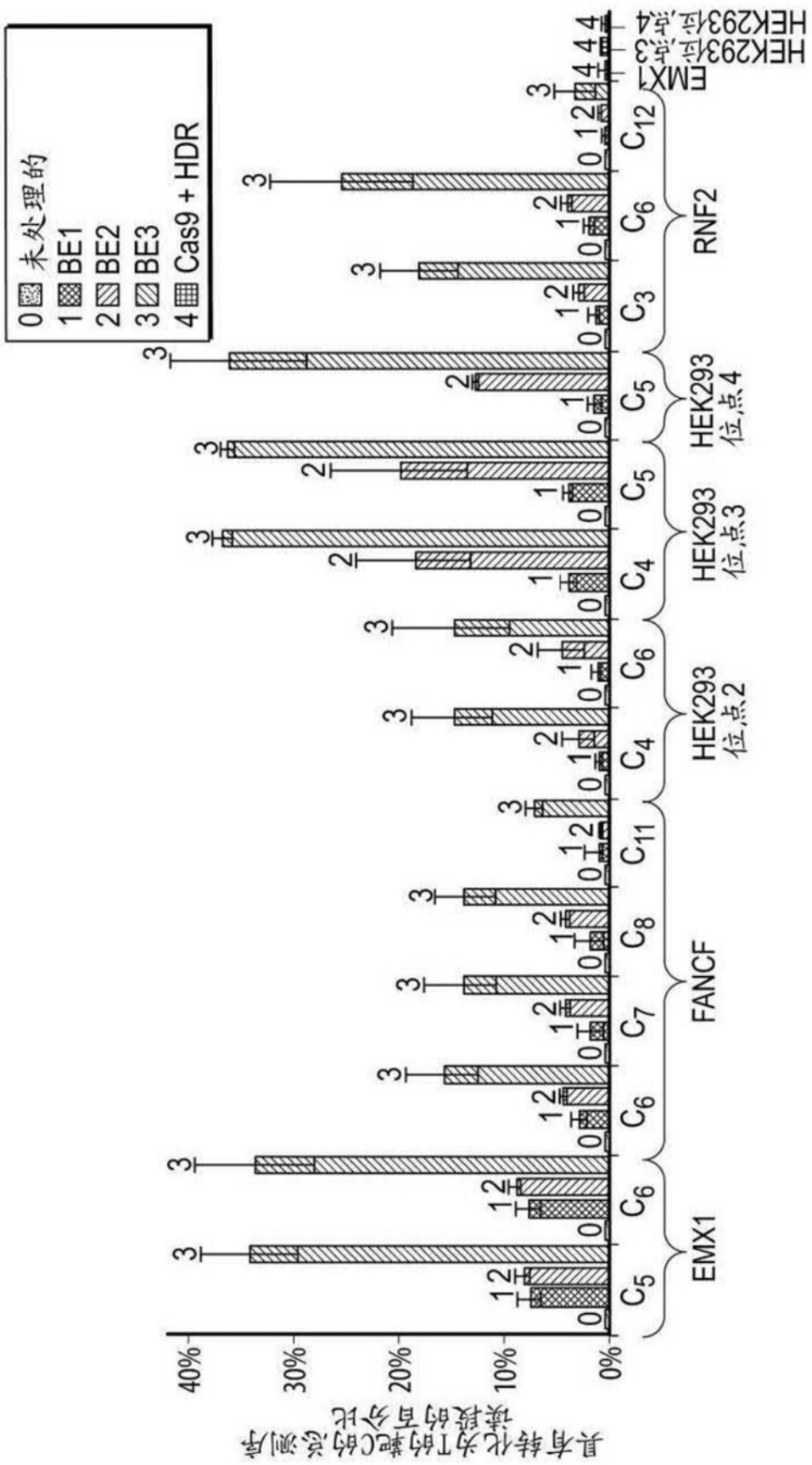

0 未处理的
1 BE1
2 BE2
3 BE3
4 Cas9 + HDR
具有转化为T的靶C的总测序读段的百分比
40%
30%
20%
10%
0%
EMX1
FANCF
HEK293 位点2
HEK293 位点3
HEK293 位点4
RNF2
EMX1
HEK293位点3
HEK293位点4

图29B

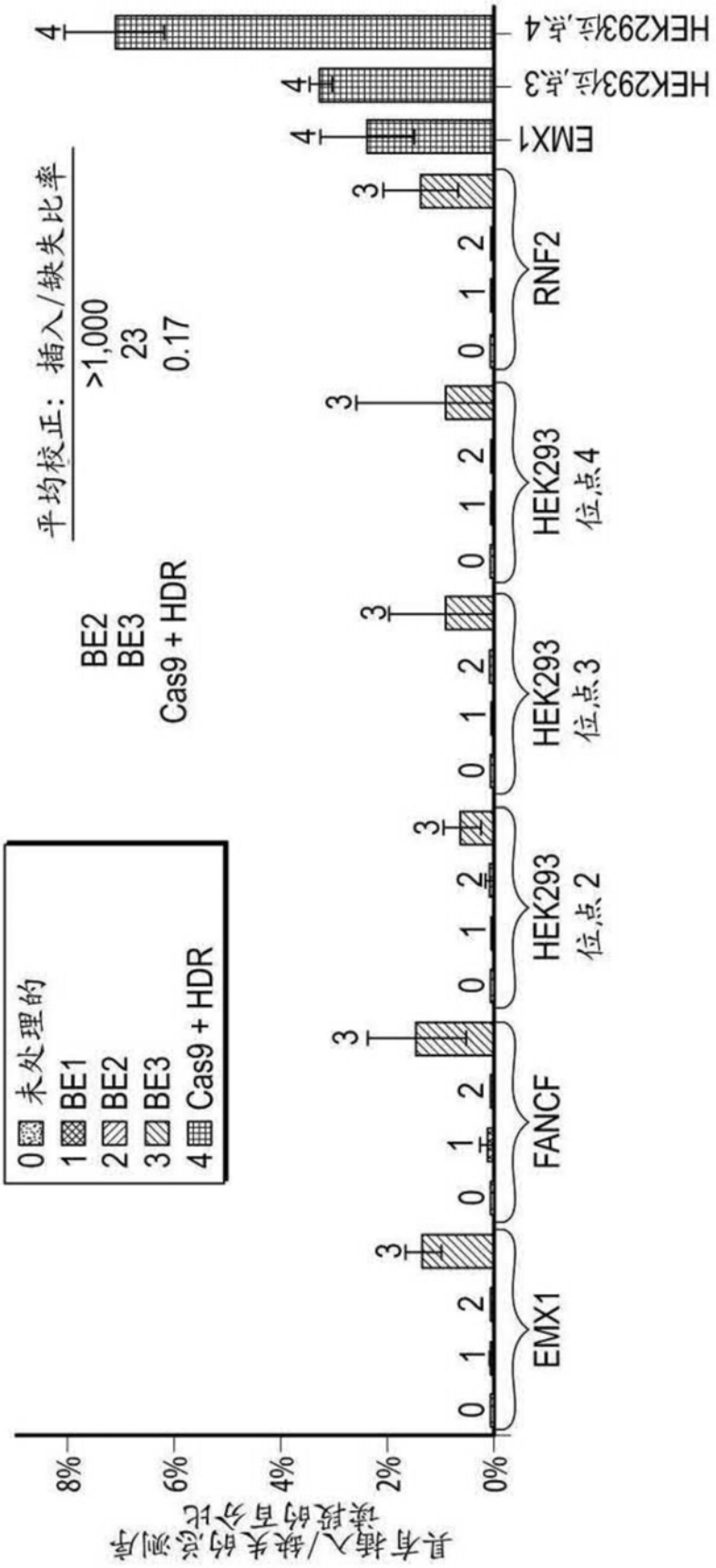
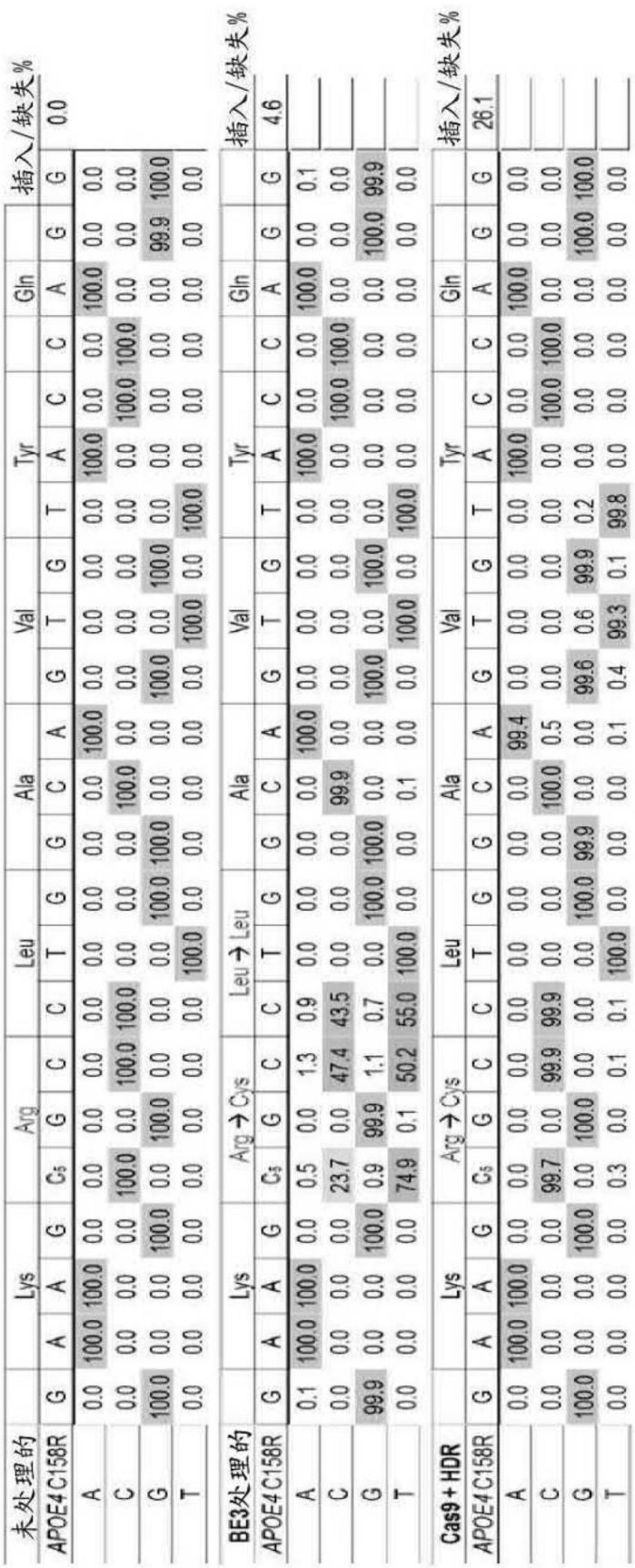

具有插入/缺失的测序读段的百分比
8%
6%
4%
2%
0%
0 未处理的
1 BE1
2 BE2
3 BE3
4 Cas9 + HDR
平均校正：插入/缺失比率
>1,000
23
0.17
BE2
BE3
Cas9 + HDR
EMX1
FANCF
HEK293 位点2
HEK293 位点3
HEK293 位点4
RNF2
EMX1
HEK293位点3
HEK293位点4

图29C

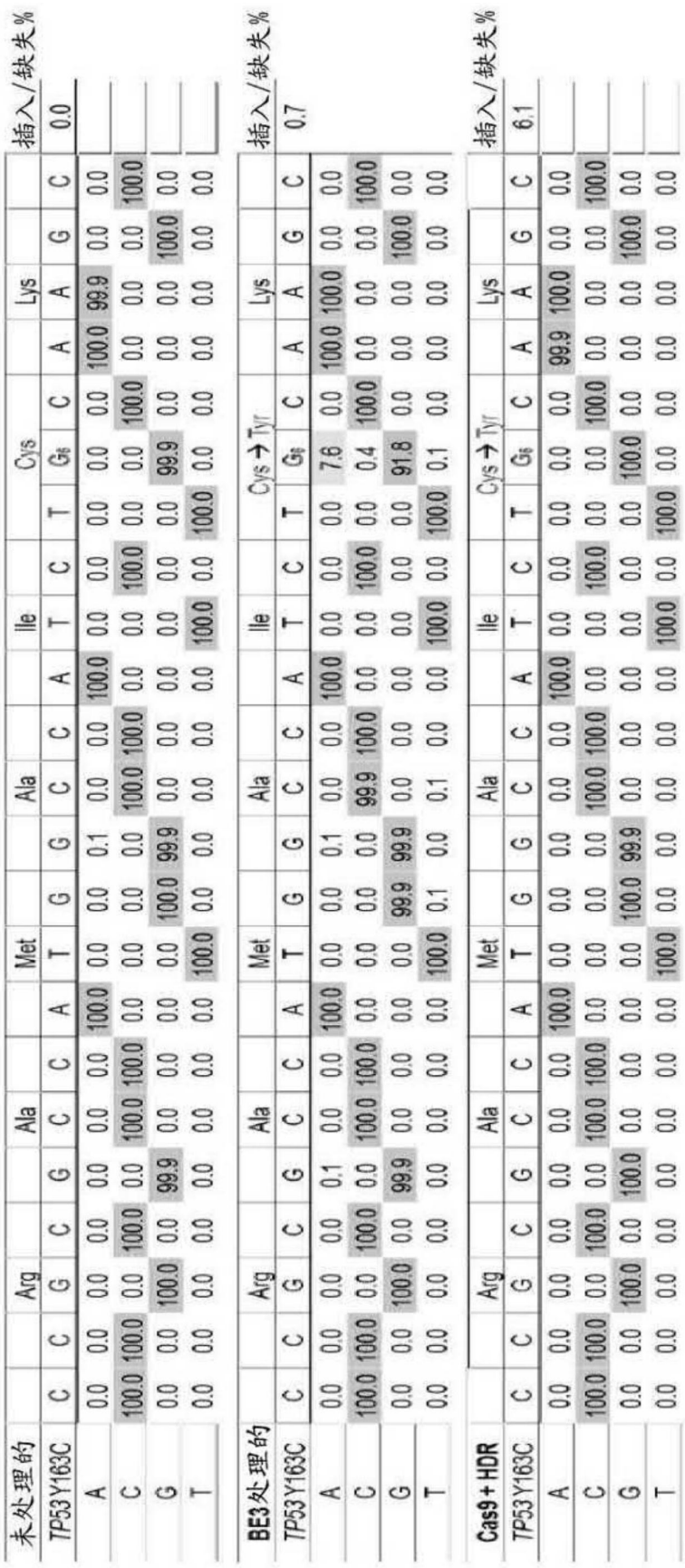


未处理的		Lys			Arg			Leu			Ala			Val			Tyr			Gln			插入/缺失%
APOE4 C158R	G	A	A	G	C_5	G	C	C	T	G	C	A	G	G	T	G	T	A	C	C	A	G	G	0.0
A	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	
C	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	
G	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	100.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	99.9	100.0		
T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0		

BE3处理的		Lys			Arg → Cys			Leu → Leu			Ala			Val			Tyr			Gln			插入/缺失%
APOE4 C158R	G	A	A	G	C_5	G	C	C	T	G	C	A	G	G	T	G	T	A	C	C	A	G	G	4.6
A	0.1	100.0	100.0	0.0	0.5	0.0	1.3	0.9	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.1	
C	0.0	0.0	0.0	0.0	23.7	0.0	47.4	43.5	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	
G	99.9	0.0	0.0	100.0	0.9	99.9	1.1	0.7	0.0	100.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	99.9	
T	0.0	0.0	0.0	0.0	74.9	0.1	50.2	55.0	100.0	0.0	0.0	0.1	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	

Cas9 + HDR		Lys			Arg → Cys			Leu			Ala			Val			Tyr			Gln			插入/缺失%
APOE4 C158R	G	A	A	G	C_5	G	C	C	T	G	C	A	G	G	T	G	T	A	C	C	A	G	G	26.1
A	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.4	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	
C	0.0	0.0	0.0	0.0	99.7	0.0	99.9	99.9	0.0	0.0	100.0	0.5	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	
G	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	100.0	99.9	0.0	0.0	99.6	0.6	99.9	0.2	0.0	0.0	0.0	0.0	100.0	100.0	
T	0.0	0.0	0.0	0.0	0.3	0.0	0.1	0.1	100.0	0.0	0.0	0.1	0.4	99.3	0.1	99.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

图30A



未处理的			Arg			Ala				Met			Ala			Ile			Cys			Lys			插入/缺失%
TP53Y163C	C	C	G	C	G	C	C	A	T	G	G	C	C	A	T	C	T	Gfi	C	A	A	G	C	0.0	
A	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.1	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	99.9	0.0	0.0		
C	100.0	100.0	0.0	100.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0		
G	0.0	0.0	100.0	0.0	99.9	0.0	0.0	0.0	0.0	100.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	0.0	100.0	0.0		
T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0		

BE3处理的			Arg			Ala				Met			Ala			Ile			Cys → Tyr			Lys			插入/缺失%
TP53Y163C	C	C	G	C	G	C	C	A	T	G	G	C	C	A	T	C	T	Gfi	C	A	A	G	C	0.7	
A	0.0	0.0	0.0	0.0	0.1	0.0	0.0	100.0	0.0	0.0	0.1	0.0	0.0	100.0	0.0	0.0	0.0	7.6	0.0	100.0	100.0	0.0	0.0		
C	100.0	100.0	0.0	100.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	99.9	100.0	0.0	0.0	100.0	0.0	0.4	100.0	0.0	0.0	0.0	100.0		
G	0.0	0.0	100.0	0.0	99.9	0.0	0.0	0.0	0.0	99.9	99.9	0.0	0.0	0.0	0.0	0.0	0.0	91.8	0.0	0.0	0.0	100.0	0.0		
T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.1	0.0	0.1	0.0	0.0	100.0	0.0	100.0	0.1	0.0	0.0	0.0	0.0	0.0		

Cas9 + HDR			Arg			Ala				Met			Ala			Ile			Cys → Tyr			Lys			插入/缺失%
TP53Y163C	C	C	G	C	G	C	C	A	T	G	G	C	C	A	T	C	T	Gfi	C	A	A	G	C	6.1	
A	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	99.9	100.0	0.0	0.0		
C	100.0	100.0	0.0	100.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0		
G	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0		
T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0		

图30B

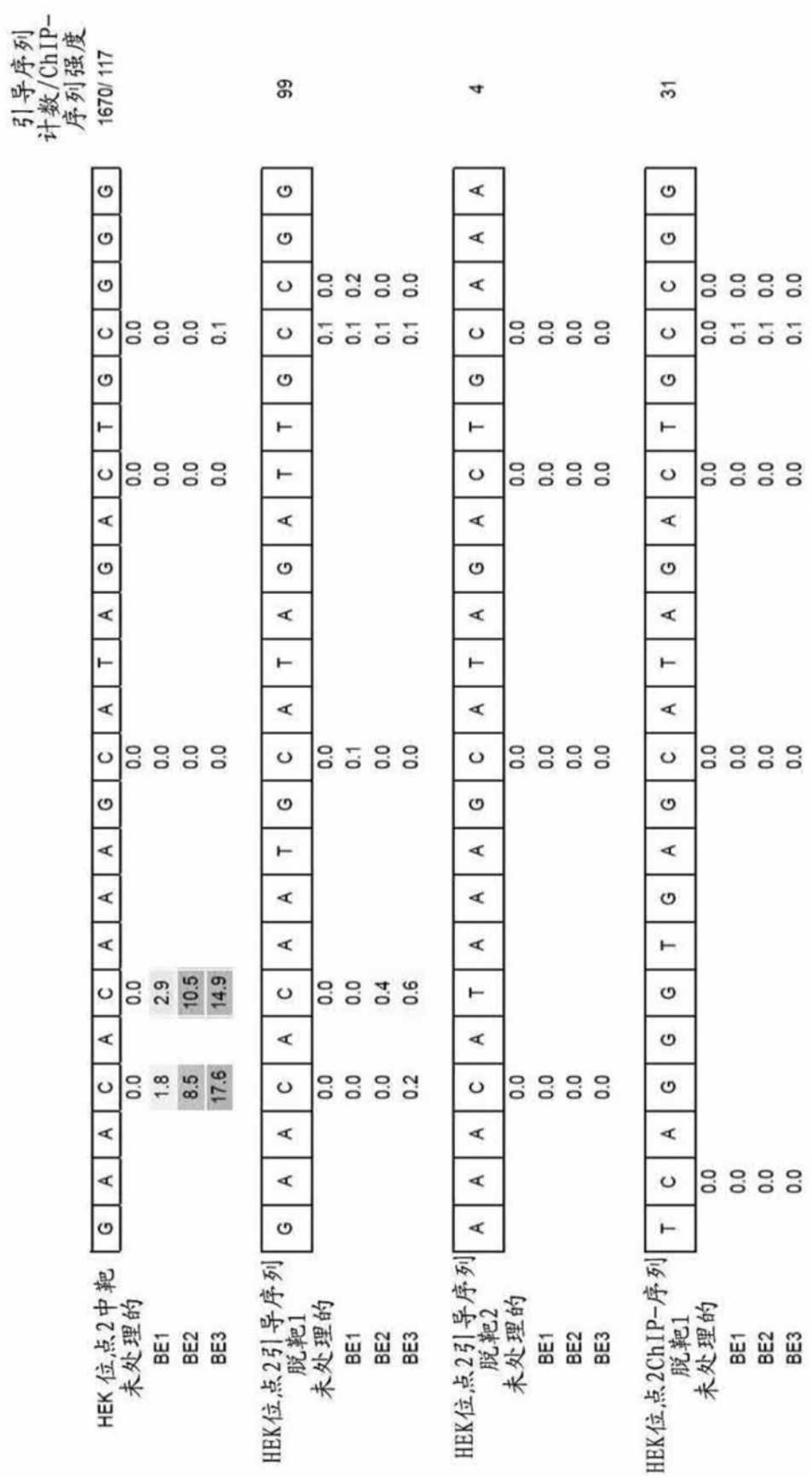

引导序列
计数/ChIP-
序列强度
1670/ 117
HEK 位点2中靶
未处理的
G A A C A C A A A G C A T A G A C T G C G G
0.0 0.0 0.0 0.0 0.0
BE1 1.8 2.9 0.0 0.0 0.0
BE2 8.5 10.5 0.0 0.0 0.0
BE3 17.6 14.9 0.0 0.0 0.1
99
HEK位点2引导序列
脱靶1
未处理的
G A A C A C A A T G C A T A G A T T G C C G
0.0 0.0 0.0 0.1 0.0
BE1 0.0 0.0 0.1 0.1 0.2
BE2 0.0 0.4 0.0 0.1 0.0
BE3 0.2 0.6 0.0 0.1 0.0
4
HEK位点2引导序列
脱靶2
未处理的
A A A C A T A A G C A T A G A C T G C A A A
0.0 0.0 0.0 0.0
BE1 0.0 0.0 0.0 0.0
BE2 0.0 0.0 0.0 0.0
BE3 0.0 0.0 0.0 0.0
31
HEK位点2ChIP-序列
脱靶1
未处理的
T C A G G G T G A G C A T A G A C T G C C G G
0.0 0.0 0.0 0.0
BE1 0.0 0.0 0.0 0.1 0.0
BE2 0.0 0.0 0.0 0.1 0.0
BE3 0.0 0.0 0.0 0.1 0.0

图31

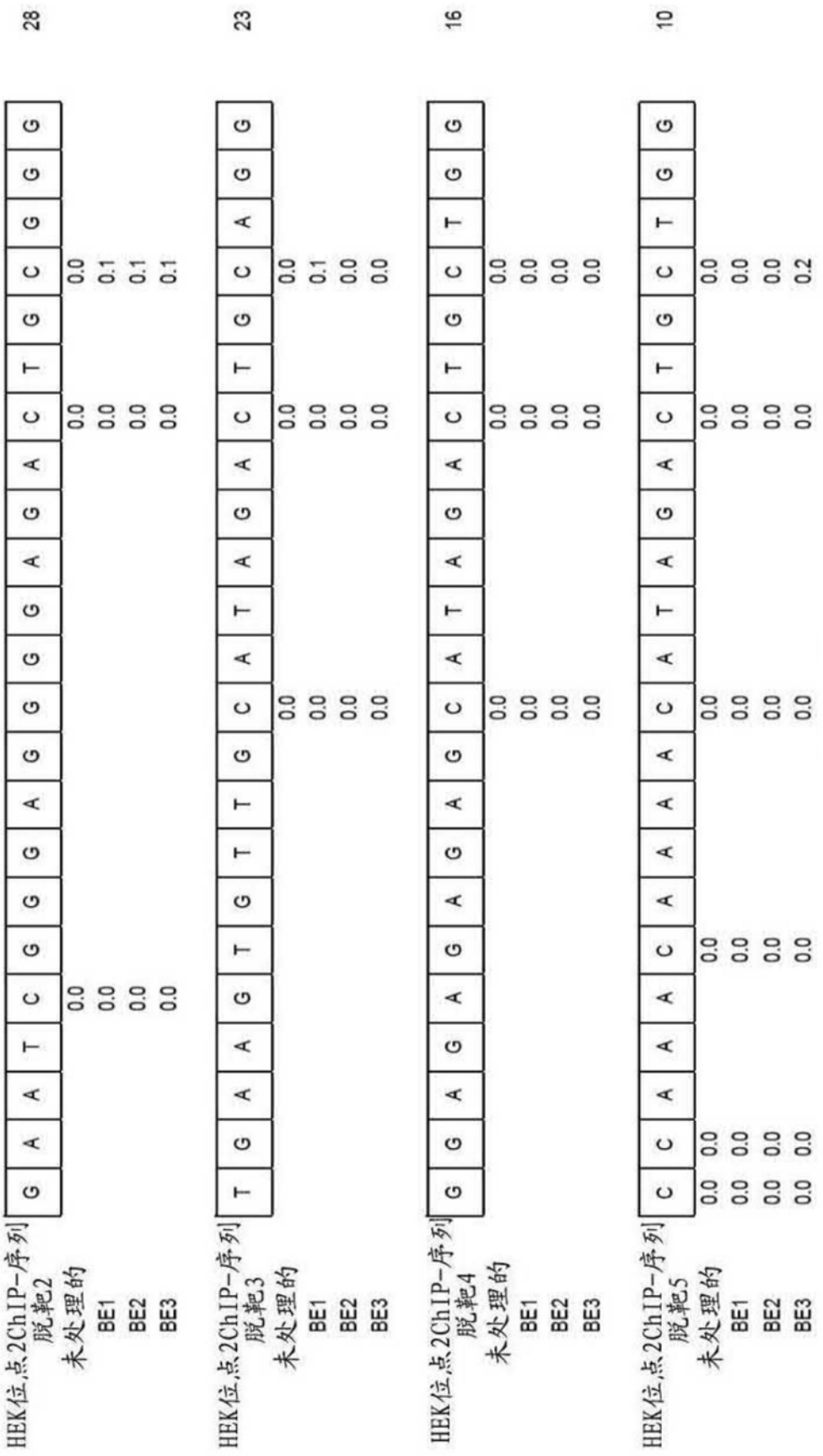

28
23
16
10
HEK位点2ChIP-序列
脱靶2
未处理的
BE1
BE2
BE3
HEK位点2ChIP-序列
脱靶3
未处理的
BE1
BE2
BE3
HEK位点2ChIP-序列
脱靶4
未处理的
BE1
BE2
BE3
HEK位点2ChIP-序列
脱靶5
未处理的
BE1
BE2
BE3

图31续

HEK位点3中靶	G	G	C	C	C	A	G	A	C	T	G	A	G	C	A	C	G	T	G	A	T	G	G	引导序列计数/ChIP-序列强度	
未处理的			0.1	0.1	0.1				0.0					0.0		0.0								2074/163	
BE1			0.1	4.8	4.3				0.0					0.0		0.0									
BE2			3.6	21.3	20.4				0.0					0.0		0.0									
BE3			0.5	33.4	33.1				0.5					0.0		0.2									
HEK位点3引导序列脱靶1	C	A	C	C	C	A	G	A	C	T	G	A	G	C	A	C	G	T	G	C	T	G	G	327	
未处理的	0.0		0.0	0.0	0.0				0.0					0.0		0.0				0.0					
BE1	0.0		0.0	0.0	0.0				0.0					0.0		0.0				0.0					
BE2	0.0		0.0	0.0	0.0				0.0					0.0		0.0				0.0					
BE3	0.0		0.0	0.0	0.0				0.0					0.0		0.0				0.0					
HEK位点3引导序列脱靶2	G	A	C	A	C	A	G	A	C	T	G	G	G	C	A	C	G	T	G	A	G	G	G	306	
未处理的		0.0		0.0					0.0					0.0		0.0									
BE1		0.0		0.0					0.0					0.0		0.0									
BE2		0.0		0.0					0.0					0.0		0.0									
BE3		0.0		0.0					0.0					0.0		0.0									
HEK位点3引导序列脱靶3	A	G	C	T	C	A	G	A	C	T	G	A	G	C	A	A	G	T	G	A	G	G	G	136	
未处理的		0.0		0.0					0.0					0.0											
BE1		0.0		0.0					0.0					0.0											
BE2		0.0		0.0					0.0					0.0											
BE3		0.0		0.9					0.0					0.0											
HEK位点3引导序列脱靶4	A	G	A	C	C	A	G	A	C	T	G	A	G	C	A	A	G	A	G	A	G	G	G	10	
未处理的		0.0		0.0					0.0					0.0											
BE1		0.0		0.0					0.0					0.1											
BE2		0.0		0.0					0.0					0.0											
BE3		0.0		0.0					0.0					0.0											

图32

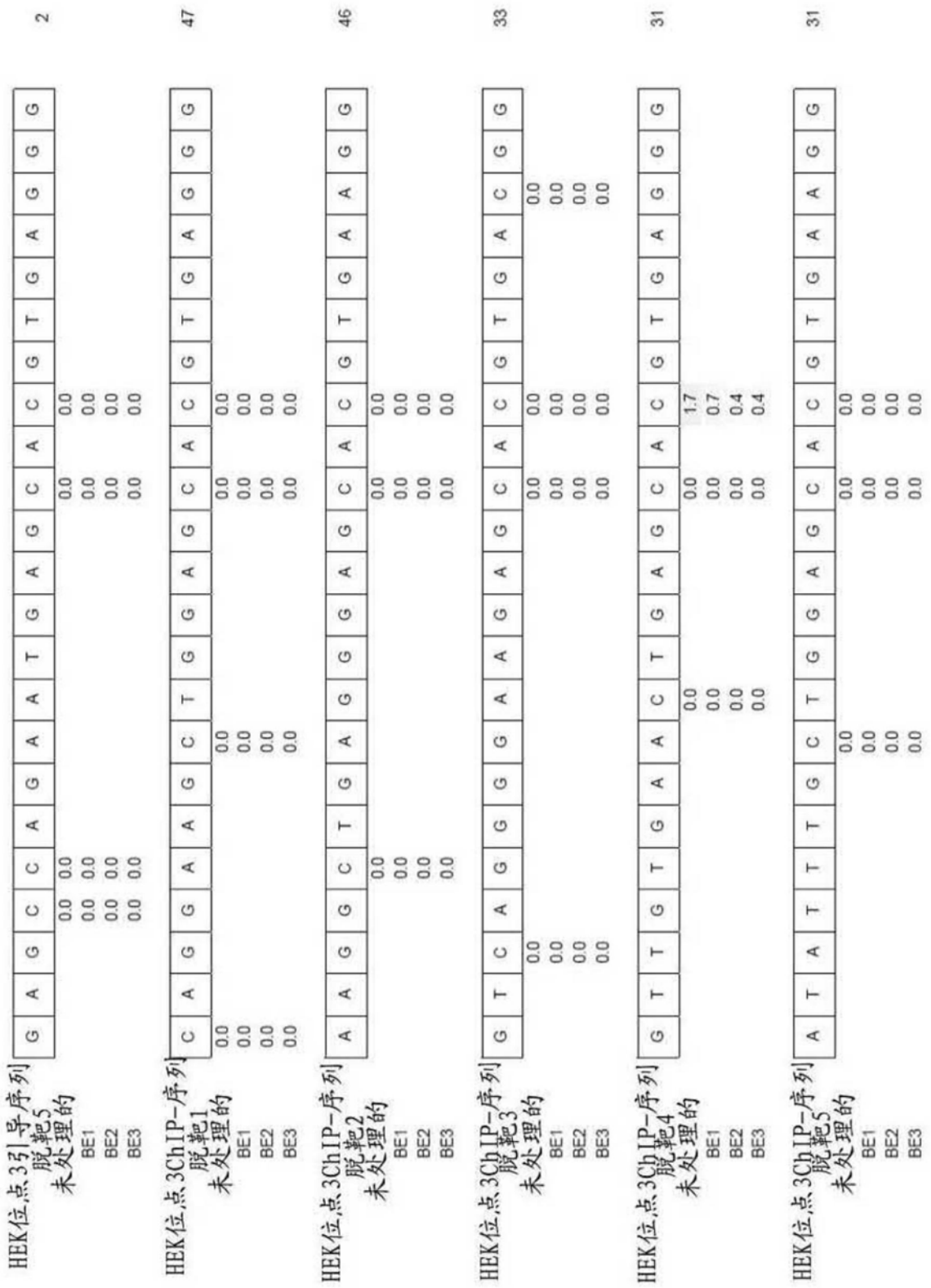

HEK位点3引导序列
脱靶5
未处理的
BE1
BE2
BE3
HEK位点3ChIP-序列
脱靶1
未处理的
BE1
BE2
BE3
HEK位点3ChIP-序列
脱靶2
未处理的
BE1
BE2
BE3
HEK位点3ChIP-序列
脱靶3
未处理的
BE1
BE2
BE3
HEK位点3ChIP-序列
脱靶4
未处理的
BE1
BE2
BE3
HEK位点3ChIP-序列
脱靶5
未处理的
BE1
BE2
BE3

图32续

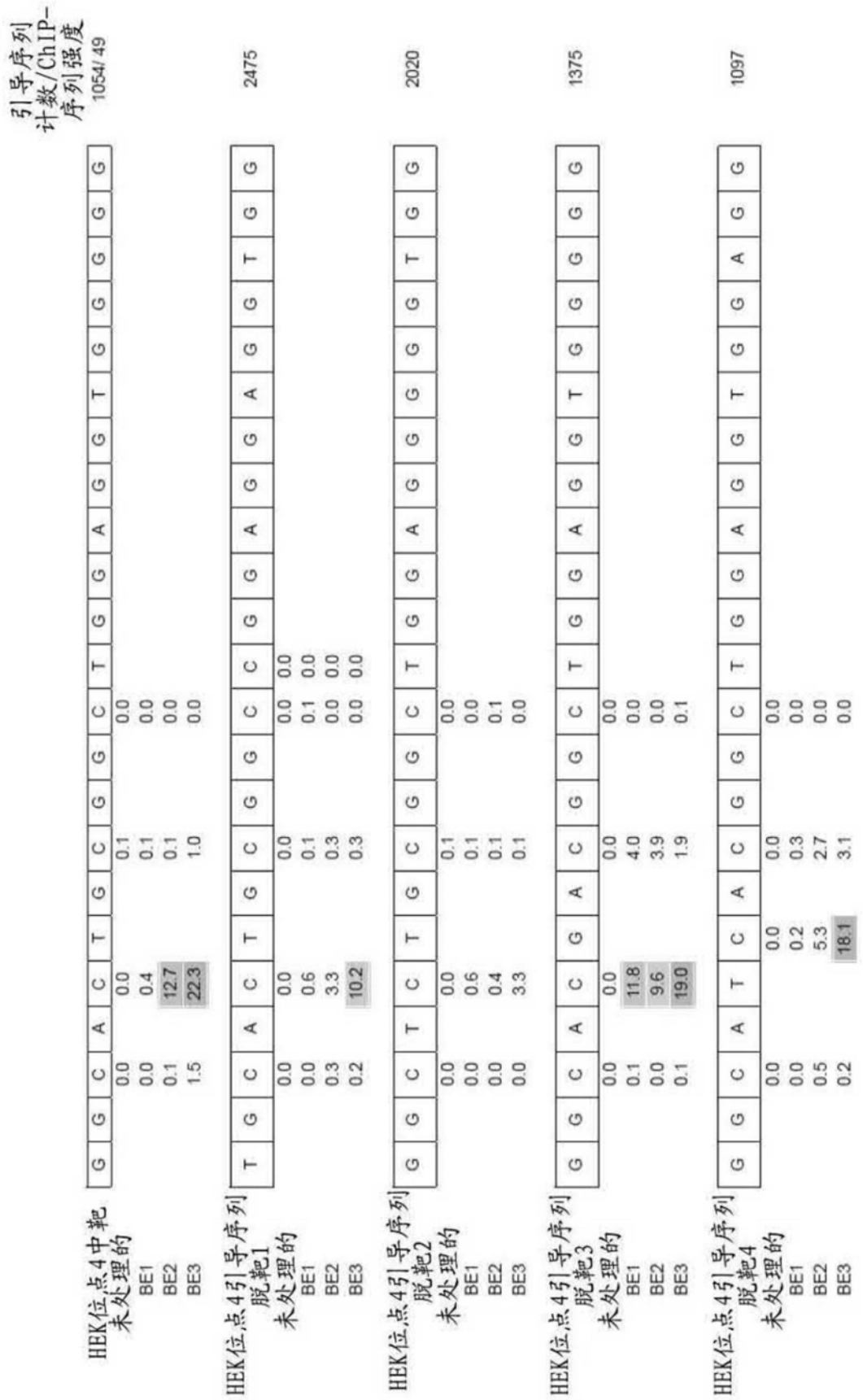

引导序列
计数/ChIP-
序列强度
1054/ 49
HEK位点4中靶
未处理的
G G C A C T G C G G C T G G A G G A T G G G G G
0.0 0.0 0.1 0.0
0.0 0.4 0.1 0.0
0.1 12.7 0.1 0.0
1.5 22.3 1.0 0.0
HEK位点4引导序列
脱靶1
2475
T G C A C T G C G G C C G G A G G A T G G G G G
0.0 0.0 0.0 0.0 0.0
0.0 0.6 0.1 0.1 0.0
0.3 3.3 0.3 0.0 0.0
0.2 10.2 0.3 0.0 0.0
未处理的
BE1
BE2
BE3
HEK位点4引导序列
脱靶2
2020
G G C T C T G C G G C C T G G A G G A G G G T G G G
0.0 0.0 0.1 0.0
0.0 0.6 0.1 0.0
0.0 0.4 0.1 0.1
0.0 3.3 0.1 0.0
未处理的
BE1
BE2
BE3
HEK位点4引导序列
脱靶3
1375
G G C A C G A C G G C T G G A G G T G G G G G G
0.0 0.0 0.0 0.0
0.1 11.8 4.0 0.0
0.0 9.6 3.9 0.0
0.1 19.0 1.9 0.1
未处理的
BE1
BE2
BE3
HEK位点4引导序列
脱靶4
1097
G G C A T C A C G G C T G G A G G T G G A G G G
0.0 0.0 0.0 0.0
0.0 0.2 0.3 0.0
0.5 5.3 2.7 0.0
0.2 18.1 3.1 0.0
未处理的
BE1
BE2
BE3

图33

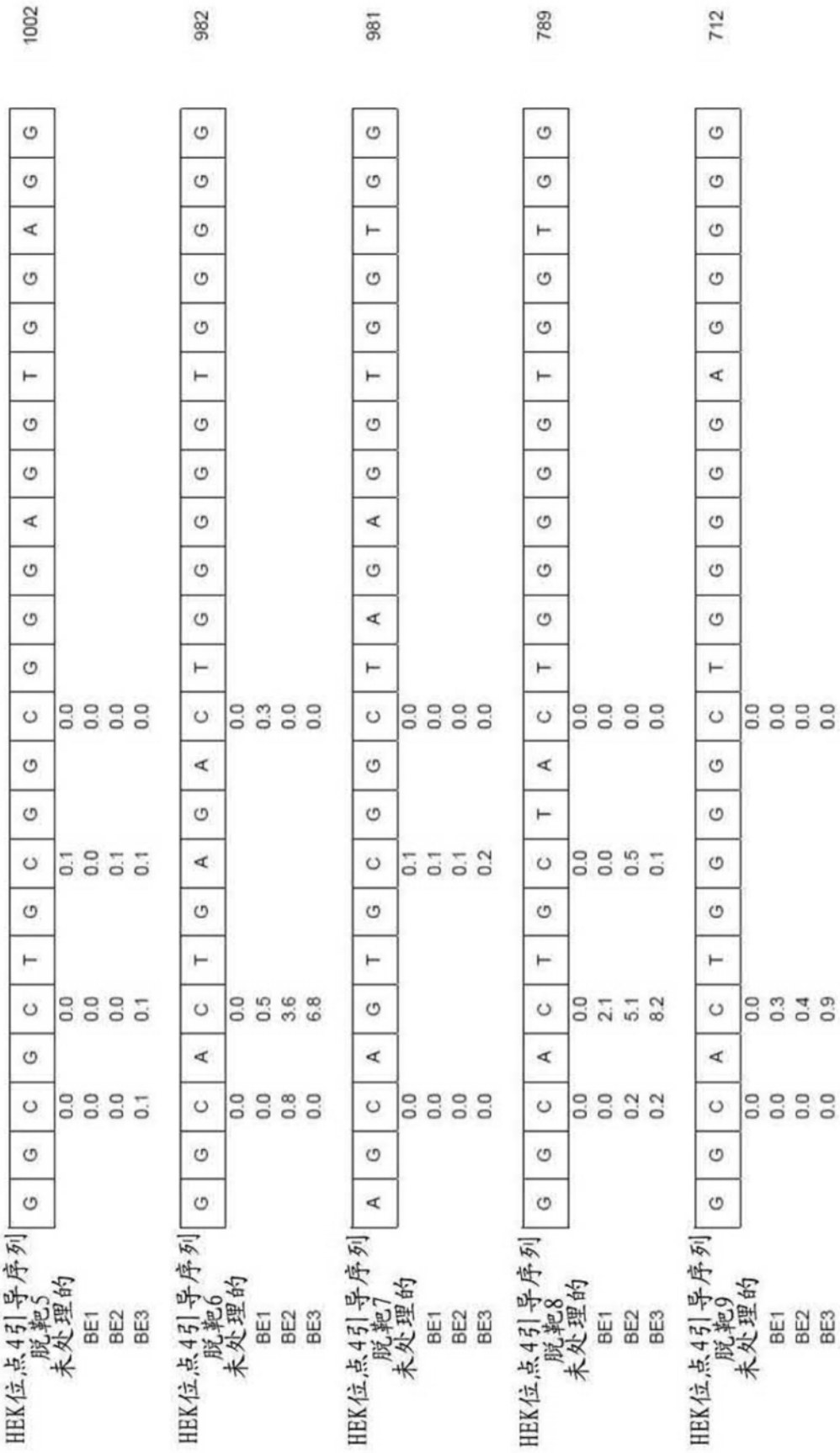

1002
HEK位点4引导序列 G G C G C T G C G G G G A G G T G G A G G
脱靶5
未处理的 0.0 0.0 0.1 0.0
BE1 0.0 0.0 0.0 0.0
BE2 0.0 0.0 0.1 0.0
BE3 0.1 0.1 0.1 0.0
982
HEK位点4引导序列 G G C A C T G A G G G G A G G T G G G G
脱靶6
未处理的 0.0 0.0 0.0
BE1 0.0 0.5 0.3
BE2 0.8 3.6 0.0
BE3 0.0 6.8 0.0
981
HEK位点4引导序列 A G C A G T G C G G G G A G G T G G T G
脱靶7
未处理的 0.0 0.1 0.0
BE1 0.0 0.1 0.0
BE2 0.0 0.1 0.0
BE3 0.0 0.2 0.0
789
HEK位点4引导序列 G G C A C T G C T A C T G G G G T G G T G
脱靶8
未处理的 0.0 0.0 0.0
BE1 0.0 2.1 0.0 0.0
BE2 0.2 5.1 0.5 0.0
BE3 0.2 8.2 0.1 0.0
712
HEK位点4引导序列 G G C A C T G G G G C T G G G G A G G G G
脱靶9
未处理的 0.0 0.0 0.0
BE1 0.0 0.3 0.0
BE2 0.0 0.4 0.0
BE3 0.0 0.9 0.0

图33续

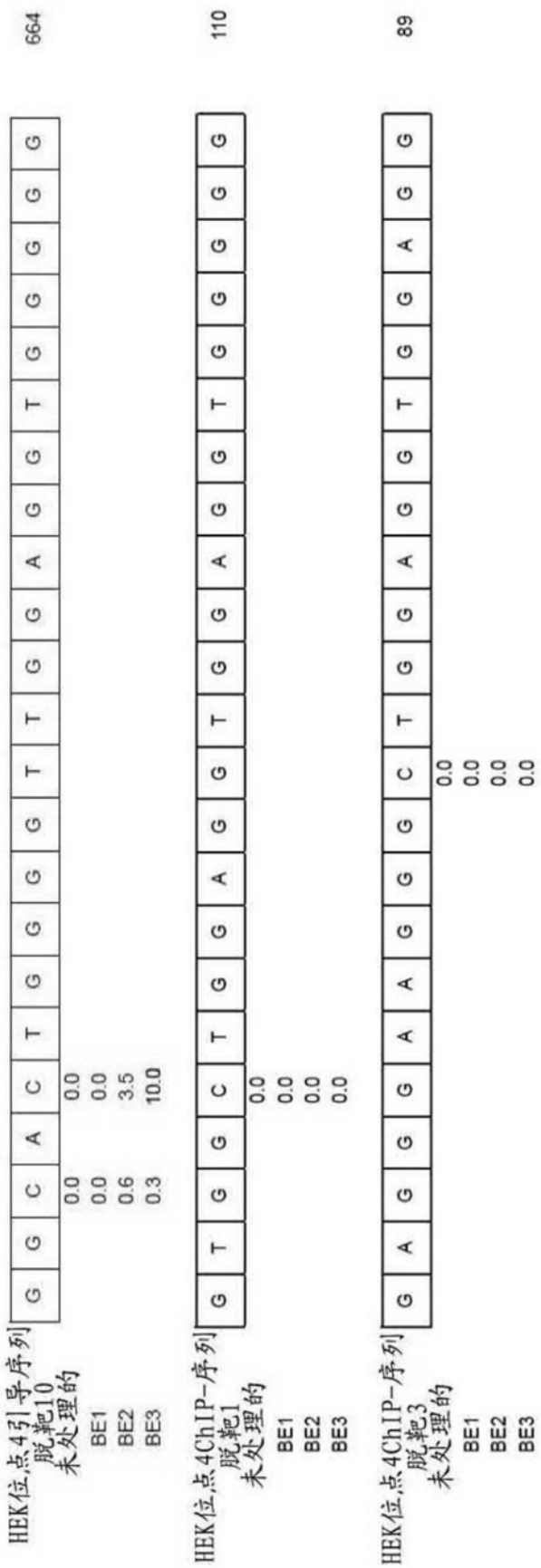

HEK位点4引导序列
脱靶10
未处理的
BE1
BE2
BE3
G G C A C T G G G G T T G G A G G A T G T G G G G G
664
0.0 0.0
0.0 0.0
0.6 3.5
0.3 10.0
HEK位点4ChIP-序列
脱靶1
未处理的
BE1
BE2
BE3
G T G G C T G G A G G T G G A G G T G G G G
110
0.0
0.0
0.0
0.0
HEK位点4ChIP-序列
脱靶3
未处理的
BE1
BE2
BE3
G A G G G A A G G C T G G A G G T G G A G G
89
0.0
0.0
0.0
0.0

图33续

APOE4

未处理的

	C50	A49	C48	C47	T46	G45	C44	G43	C42	A41	A40	G39	C38	T37	G36	C35	G34	T33	A32	A31	G30	C29	G28	G27	C26
A	0.1±0.1	100±0.1	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
C	100±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0±0	0.1±0.1	0±0	100±0.1	0.1±0.1	0±0	100±0.1	0.1±0.1	0.1±0.1	0±0	0.1±0.1	0±0	100±0.1	0.1±0.1	0.1±0.1	100±0.1
G	0.1±0.1	0.1±0.1	0±0	0±0	0.1±0.1	100±0.1	0±0	99.9±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0±0	0±0	100±0.1	0.1±0.1	100±0.1	0±0	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1	0±0
T	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1

BE3+中靶sgRNA

A	0±0	100±0.1	0±0	0.1±0.1	0±0	0.1±0.1	0±0	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0±0	0.1±0.1	0±0	100±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
C	100±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0±0	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0±0	0±0	0±0	0±0	100±0.1
G	0.1±0.1	0.1±0.1	0±0	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0±0	0.1±0.1	0.1±0.1	100±0.1	0±0	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0±0	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0±0	100±0.1	0.1±0.1
T	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1

BE3+脱靶sgRNA

A	0±0	100±0.1	0±0	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0±0	100±0.1	100±0.1	100±0.1	0.1±0.1	0±0	0±0	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0±0
C	100±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0±0	100±0.1	0.1±0.1	100±0	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0±0	100±0.1	0.1±0.1	0±0	100±0.1
G	0±0	0.1±0.1	0±0	0±0	0±0	100±0.1	0.1±0.1	100±0.1	0±0	0.1±0.1	0.1±0.1	100±0	0±0	0±0	100±0.1	0±0	100±0.1	0±0	0.1±0.1	0.1±0.1	100±0.1	0±0	100±0.1	100±0.1	0±0
T	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0±0	0±0	0±0	0±0	0.1±0.1	0±0	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0±0	0.1±0.1	0.1±0.1	0±0	0.1±0.1	0.1±0.1	0.1±0.1

APOE4

未处理的

	T25	C24	C23	T22	C21	C20	G19	C18	G17	A16	T15	G14	C13	C12	G11	A10	T9	G8	A7	C6	C5	T4	G3	C2	A1
A	0.1±0.1	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0±0	100±0.1
C	0.1±0.1	100±0	100±0.1	0.1±0.1	100±0.1	100±0.1	0±0	100±0.1	0.1±0.1	0±0	0.1±0.1	0±0	100±0.1	100±0.1	0±0	0±0	0.1±0.1	0±0	0±0	100±0.1	100±0.1	0.1±0.1	0±0	100±0.1	0.1±0.1
G	0±0	0±0	0±0	0.1±0.1	0±0	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0±0	0±0	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0±0	0±0	0±0	100±0.1	0.1±0.1	0.1±0.1
T	100±0.1	0±0	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1

BE3+中靶sgRNA

A	0±0	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0±0	0.1±0.1	0±0	0.1±0.1	0.2±0.1	100±0.1	0.1±0.1	0.2±0.1	99.9±0.2	0±0	0.1±0.1	0.2±0.2	0.1±0.1	0.1±0.1	99.9±0.2
C	0.1±0.1	99.9±0.1	100±0.1	0.1±0.1	99.5±0.1	100±0.1	0.1±0.1	100±0.1	0±0	0±0	0.1±0.1	0±0	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0±0	100±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.2±0.2	0.2±0.2
G	0±0	0±0	0.1±0.1	0±0	0.1±0.1	0±0	100±0.1	0±0	100±0.1	0.1±0.1	0±0	100±0.1	0.1±0.1	0.1±0.1	99.9±0.1	0.1±0.1	0±0	99.9±0.1	0.2±0.2	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1
T	100±0.1	0.2±0.1	0.1±0.1	100±0.1	0.5±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	99.9±0.2	0.1±0.1	0.1±0.1	0.1±0.1

BE3+脱靶sgRNA

A	0.1±0.1	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0±0	0.1±0.1	0.1±0.1	0±0	0.1±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1
C	0.1±0.1	100±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1	0±0	100±0.1	0.1±0.1	0±0	0.1±0.1	0±0	100±0.1	100±0.1	0±0	0±0	0.1±0.1	0±0	0±0	100±0.1	100±0.1	0.1±0.1	0±0	100±0.1	0.1±0.1
G	0±0	0±0	0±0	0.1±0.1	0±0	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0±0	0±0	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0±0	0±0	100±0.1	0.1±0.1	0.1±0.1
T	100±0.1	0±0	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1

BE3+中靶sgRNA

A	0±0	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0±0	0.1±0.1	0±0	0.1±0.1	0.2±0.1	100±0.1	0.1±0.1	0.2±0.1	99.9±0.2	0±0	0.1±0.1	0.2±0.2	0.1±0.1	0.1±0.1	99.9±0.2
C	0.1±0.1	99.9±0.1	100±0.1	0.1±0.1	99.5±0.1	100±0.1	0.1±0.1	100±0.1	0±0	0±0	0.1±0.1	0±0	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0±0	100±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.2±0.2	0.2±0.2
G	0±0	0±0	0.1±0.1	0±0	0.1±0.1	0±0	100±0.1	0±0	100±0.1	0.1±0.1	0±0	100±0.1	0.1±0.1	0.1±0.1	99.9±0.1	0.1±0.1	0±0	99.9±0.1	0.2±0.2	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1
T	100±0.1	0.2±0.1	0.1±0.1	100±0.1	0.5±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	99.9±0.2	0.1±0.1	0.1±0.1	0.1±0.1

BE3+脱靶sgRNA

A	0.1±0.1	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0±0	0.1±0.1	0.1±0.1	0±0	0.1±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1
C	0.1±0.1	100±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1	0±0	100±0.1	0.1±0.1	0±0	0.1±0.1	0±0	100±0.1	100±0.1	0±0	0±0	0.1±0.1	0±0	0±0	100±0.1	100±0.1	0.1±0.1	0±0	100±0.1	0.1±0.1
G	0±0	0±0	0±0	0.1±0.1	0±0	0.1±0.1	100±0.1	0±0	100±0.1	0±0	0±0	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0±0	0±0	0±0	100±0.1	0.1±0.1	0.1±0.1
T	100±0.1	0±0	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1

图34

APOE4

未处理的

	A-1	G-2	G-3	C-4	C-5	G-6	G-7	G-8	G-9	C-10	C-11	C-12	G-13	C-14	G-15	A-16	G-17	G-18	G-19	C-20	G-21	C-22	C-23	G-24	A-25
A	100±0.1	0.1±0.1	0.1±0.1	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.2±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0±0	0.2±0.1	0.1±0.1	0±0	0.1±0.1	100±0.1
C	0.1±0.1	0.1±0.1	0±0	100±0.1	100±0.1	0.1±0.1	0±0	0.1±0.1	0.1±0.1	100±0.1	100±0.1	100±0.1	0.1±0.1	99.9±0.1	0.1±0.1	0.1±0.1	0±0	0±0	0±0	100±0.1	0.1±0.1	100±0.1	100±0.1	0±0	0.1±0.1
G	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0±0	99.9±0.1	100±0.1	100±0.1	100±0.1	0.1±0.1	0±0	0.1±0.1	99.9±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1	100±0.1	0.1±0.1	99.9±0.1	0.1±0.1	0.1±0.1	100±0.2	0.1±0.1
T	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.2	0.1±0.1

BE3 + 中靶 sgRNA

A	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.2±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.2±0.1	0.2±0.2	0.2±0.1	0.2±0.1	0.1±0.1	0.1±0.1	99.9±0.1
C	0.1±0.1	0±0	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0±0	0.1±0.1	0±0	100±0.1	100±0.1	100±0.1	0.1±0.1	100±0.1	0±0	0±0	0±0	0.1±0.1	0±0	99.8±0.2	0±0	99.9±0.1	100±0.1	0±0	0±0
G	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	99.9±0.1	0±0	100±0.1	0.1±0.1	100±0.1	100±0.1	99.9±0.1	0.2±0.1	99.8±0.2	0±0	0±0	100±0.1	0.1±0.1
T	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.2±0.2	0.1±0.1	0.1±0.1	0.1±0.1	0.2±0.1

BE3 + 脱靶 sgRNA

A	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0±0	0.1±0.1	0±0	0.1±0.1	0.1±0.1	0.2±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1
C	0.1±0.1	0±0	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0±0	0±0	0±0	100±0.1	100±0.1	100±0.1	0±0	100±0.1	0±0	0.1±0.1	0.1±0.1	0±0	0±0	100±0.1	0.1±0.1	100±0.1	100±0.1	0±0	0±0
G	0.1±0.1	100±0.1	100±0.1	0±0	0.1±0.1	100±0.1	100±0.1	100±0.1	100±0.1	0±0	0±0	0±0	99.9±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1	100±0.1	0.1±0.1	99.9±0.1	0±0	0±0	100±0.1	0±0
T	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0±0	0±0	0.1±0.1	0±0	0.1±0.1	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1

APOE4

未处理的

	G-26	C-27	G-28	C-29	G-30	G-31	C-32	C-33	T-34	C-35	A-36	G-37	C-38	G-39	C-40	C-41	A-42	T-43	C-44	C-45	G-46	C-47	G-48	A-49	G-50
A	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0±0	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.2±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1
C	0±0	100±0.1	0.1±0.1	99.9±0.1	0±0	0.1±0.1	100±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	99.8±0.1	0.1±0.1	100±0.1	100±0.1	0±0	0.1±0.1	100±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0±0	0.1±0.1
G	100±0.1	0.1±0.1	99.9±0.2	0.1±0.1	100±0.1	100±0.1	0±0	0±0	0±0	0±0	0.1±0.1	100±0.1	0.1±0.1	99.9±0.1	0±0	0±0	0.1±0.1	0.5±0.1	0.1±0.1	0.1±0.1	99.9±0.2	0±0	99.7±0.1	0.1±0.1	100±0.1
T	0.1±0.1	0.1±0.1	0.2±0.2	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.3±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	99.6±0.1	0.1±0.1	0.1±0.1	0.2±0.2	0.1±0.1	0.3±0.1	0±0	0.1±0.1

BE3 + 中靶 sgRNA

A	0.1±0.1	0.1±0.1	0.1±0.1	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.2±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1
C	0±0	100±0.1	0.1±0.1	100±0.1	0±0	0.1±0.1	100±0.1	100±0.1	0.1±0.1	99.9±0.1	0±0	0±0	99.7±0.1	0.1±0.1	100±0.1	100±0.1	0±0	0.1±0.1	100±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0±0	0.2±0.2
G	100±0.1	0.1±0.1	100±0.1	0±0	100±0.1	100±0.1	0.1±0.1	0±0	0.1±0.1	0±0	0.1±0.1	100±0.1	0.1±0.1	99.8±0.2	0.1±0.1	0±0	0.1±0.1	0.5±0.5	0±0	0±0	100±0.1	0.1±0.1	99.6±0.4	0.1±0.1	99.9±0.2
T	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0±0	100±0.1	0.2±0.1	0.1±0.1	0.1±0.1	0.4±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	99.6±0.5	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.4±0.3	0±0	0±0

BE3 + 脱靶 sgRNA

A	0.1±0.1	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0±0	0.1±0.1	0±0	0±0	100±0.1	0.1±0.1	0.1±0.1	0±0	0±0	100±0.1	0±0	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1
C	0.1±0.1	100±0.1	0±0	100±0.1	0.1±0.1	0±0	100±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	99.8±0.1	0.1±0.1	100±0.1	100±0.1	0±0	0.1±0.1	100±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0±0	0.1±0.1
G	0.1±0.1	0.1±0.1	100±0.1	0±0	0.1±0.1	100±0.1	0±0	0±0	0.1±0.1	0±0	0.1±0.1	0.1±0.1	99.9±0.1	0.1±0.1	0±0	0±0	0.1±0.1	0.4±0.1	0±0	0±0	100±0.1	0.1±0.1	99.7±0.2	0.1±0.1	100±0.1
T	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0±0	0.1±0.1	0±0	0±0	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.2±0.2	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	99.7±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.4±0.2	0±0	0.1±0.1

图34续

TP53 未处理的	C-50	C-49	T-48	G-47	T-46	G-45	C-44	A-43	G-42	C-41	T-40	G-39	T-38	G-37	G-36	G-35	T-34	T-33	G-32	A-31	T-30	T-29	C-28	C-27	A-26
A	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1
C	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1
G	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
T	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1
BE3+ 中靶 sgRNA																									
A	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1
C	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1
G	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	99.8±0.3	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
T	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.3±0.3	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1
BE3+ 脱靶 sgRNA																									
A	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1
C	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1
G	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
T	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1

TP53 未处理的	C-25	A-24	C-23	C-22	C-21	C-20	C-19	G-18	C-17	C-16	C-15	G-14	G-13	C-12	A-11	C-10	C-9	C-8	G-7	C-6	G-5	T-4	C-3	C-2	G-1
A	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
C	100±0.1	0.1±0.1	100±0.1	100±0.1	100±0.1	100±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1
G	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1
T	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1
BE3+ 中靶 sgRNA																									
A	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
C	100±0.1	0.1±0.1	100±0.1	100±0.1	100±0.1	100±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1	99.9±0.2	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1
G	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1
T	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.2±0.2	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1
BE3+ 脱靶 sgRNA																									
A	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.5±0.7	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
C	100±0.1	0.1±0.1	100±0.1	100±0.1	100±0.1	100±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1	99.6±0.6	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1
G	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1
T	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1

图35

TP53 未处理的

	A1	G2	T3	C4	A5	C6	A7	G8	C9	A10	C11	A12	T13	G14	A15	C16	G17	G18	A19	G20	G21	T22	T23	G24	T25
A	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
C	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
G	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1
T	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1

BE3+ 中靶 sgRNA

A	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.3±0.4	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
C	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.5±0.7	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
G	0.1±0.1	99.9±0.3	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	99.8±0.4	0.1±0.1	100±0.1	99.6±0.7	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1
T	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	100±0.1

BE3+ 脱靶 sgRNA

A	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
C	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
G	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1
T	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	100±0.1

	G26	A27	G28	G29	C30	G31	C32	T33	G34	C35	C36	C37	C38	C39	A40	C41	C42	A43	T44	G45	A46	G47	C48	G49	C50
A	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
C	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	100±0.1	100±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1
G	100±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
T	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1

TP53 未处理的

BE3+ 中靶 sgRNA

A	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.2±0.3	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
C	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	99.9±0.3	100±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1
G	10±0.1	0.1±0.1	99.7±0.6	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1
T	0.1±0.1	0.1±0.1	0.4±0.5	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1

BE3+ 脱靶 sgRNA

A	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
C	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	100±0.1
G	100±0.1	0.1±0.1	100±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1
T	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	100±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1	0.1±0.1

图35续

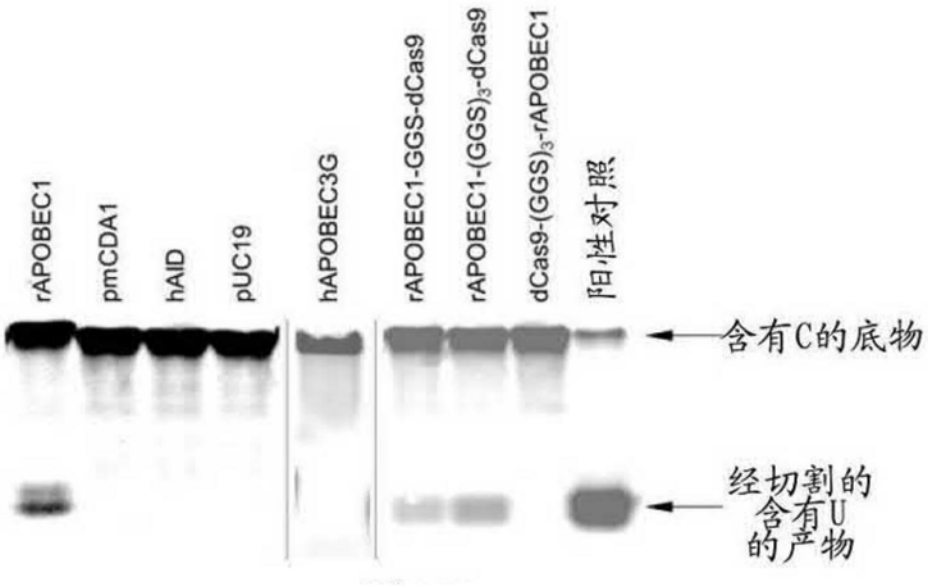

rAPOBEC1
pmCDA1
hAID
pUC19
hAPOBEC3G
rAPOBEC1-GGS-dCas9
rAPOBEC1-(GGS)3-dCas9
dCas9-(GGS)3-rAPOBEC1
阳性对照
含有C的底物
经切割的
含有U
的产物

图36A

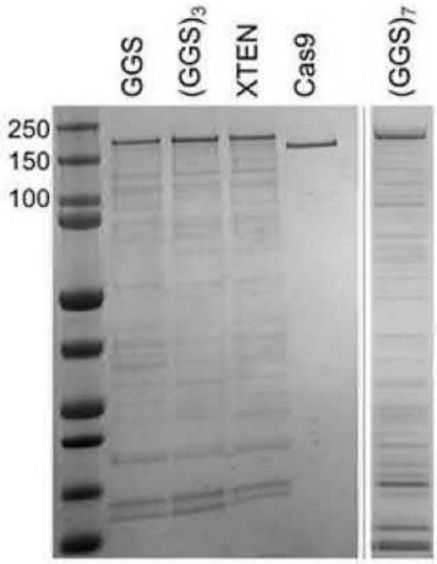

GGS
(GGS)3
XTEN
Cas9
(GGS)7
250
150
100

图36B

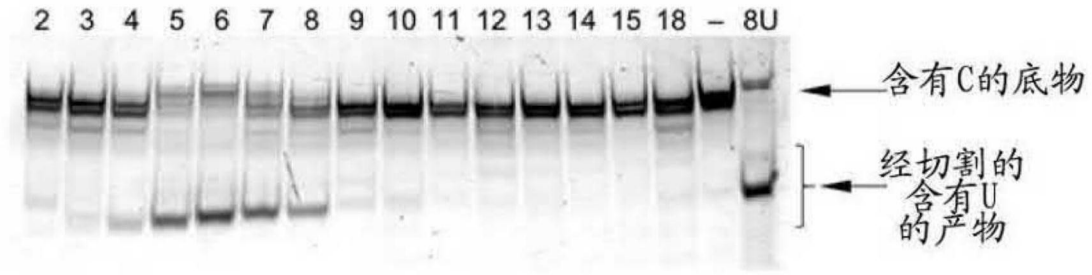


图36C

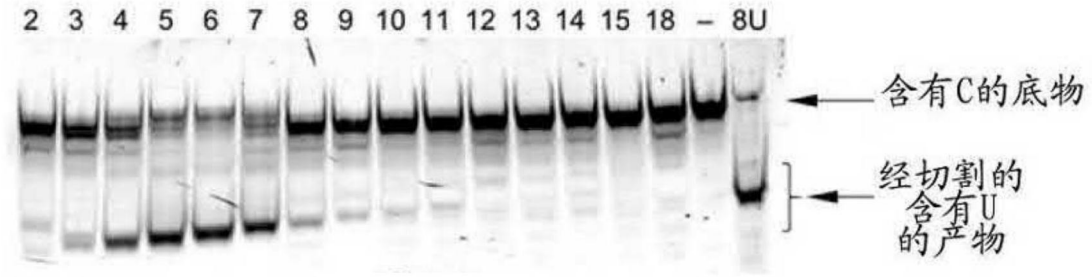


图36D

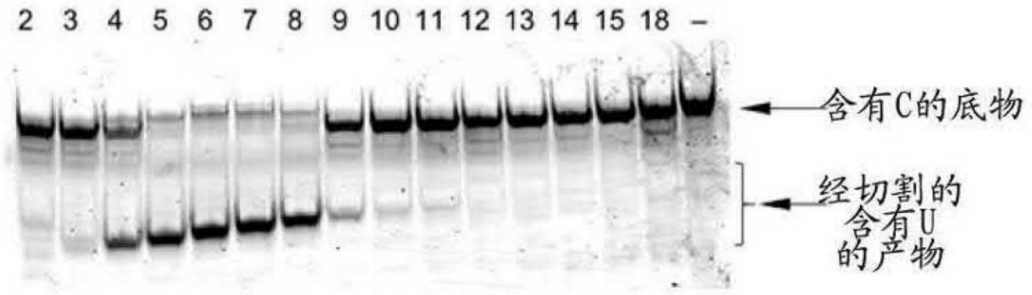


图36E

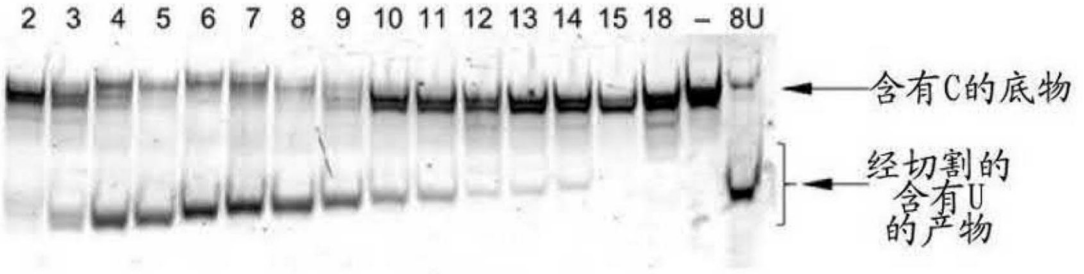


图36F

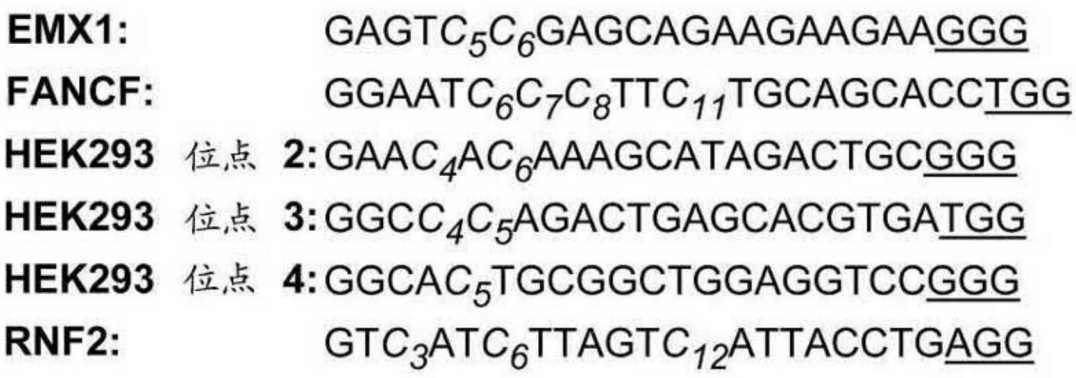
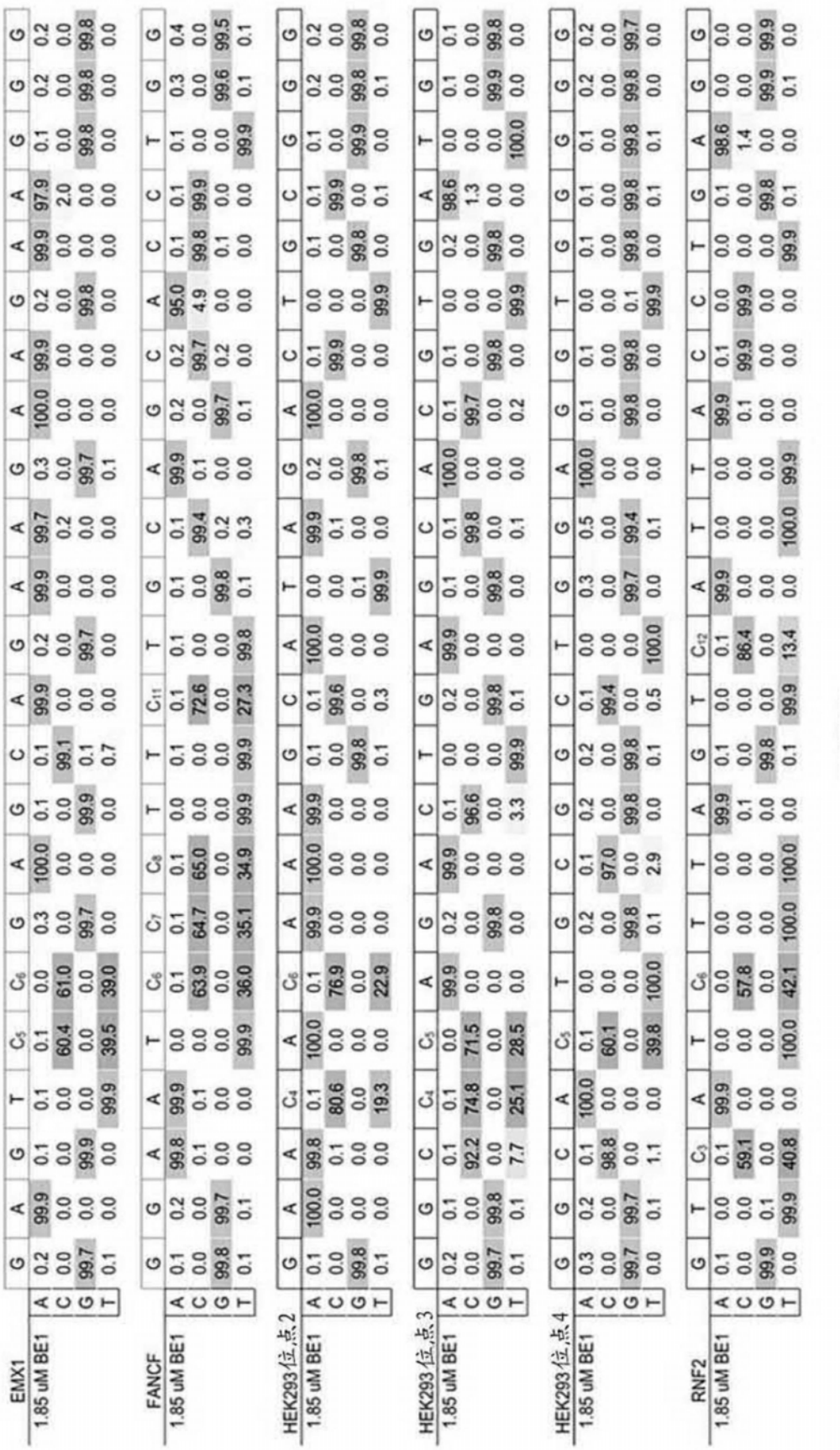

EMX1: GAGTC5C6GAGCAGAAGAAGAAGGG
FANCF: GGAATC6C7C8TTC11TGCAGCACCTGG
HEK293 位点 2: GAAC4AC6AAAGCATAGACTGCGGG
HEK293 位点 3: GGCC4C5AGACTGAGCACGTGATGG
HEK293 位点 4: GGCAC5TGCGGCTGGAGGTCCGGG
RNF2: GTC3ATC6TTAGTC12ATTACCTGAGG

图37A



EMX1		G	A	G	T	C_5	C_6	G	A	G	C	A	G	A	A	G	A	A	A	G	G	G
1.85 uM BE1	A	0.2	99.9	0.1	0.1	0.1	0.0	0.3	100.0	0.1	0.1	99.9	0.2	99.9	99.7	0.3	100.0	99.9	0.2	99.9	97.9	0.1	0.2	0.2
	C	0.0	0.0	0.0	0.0	60.4	61.0	0.0	0.0	0.0	99.1	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	2.0	0.0	0.0	0.0	0.0
	G	99.7	0.0	99.9	0.0	0.0	0.0	99.7	0.0	99.9	0.1	0.0	99.7	0.0	0.0	99.7	0.0	0.0	99.8	0.0	0.0	99.8	99.8	99.8
	T	0.1	0.0	0.0	99.9	39.5	39.0	0.0	0.0	0.0	0.7	0.0	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

FANCF		G	G	A	A	T	C_6	C_7	C_8	T	T	C_{11}	T	G	C	A	G	C	A	C	C	T	G	G
1.85 uM BE1	A	0.1	0.2	99.8	99.9	0.0	0.1	0.1	0.1	0.0	0.1	0.1	0.1	0.1	0.1	99.9	0.2	0.2	95.0	0.1	0.1	0.1	0.3	0.4
	C	0.0	0.0	0.1	0.1	0.0	63.9	64.7	65.0	0.0	0.0	72.6	0.0	0.0	99.4	0.1	0.0	99.7	4.9	99.8	99.9	0.0	0.0	0.0
	G	99.8	99.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.8	0.2	0.0	99.7	0.2	0.0	0.1	0.0	0.0	0.0	99.6	99.5
	T	0.1	0.1	0.0	0.0	99.9	36.0	35.1	34.9	99.9	99.9	27.3	99.8	0.1	0.3	0.0	0.1	0.0	0.0	0.0	0.0	99.9	0.1	0.1

HEK293位点2		G	A	A	C_4	A	C_5	A	A	A	G	C	A	T	A	G	A	C	T	G	C	G	G	G
1.85 uM BE1	A	0.1	100.0	99.8	0.1	100.0	0.1	99.9	100.0	99.9	0.1	0.1	100.0	0.0	99.9	0.2	100.0	0.1	0.0	0.1	0.1	0.1	0.2	0.2
	C	0.0	0.0	0.1	80.6	0.0	76.9	0.0	0.0	0.0	0.0	99.6	0.0	0.0	0.1	0.0	0.0	99.9	0.0	0.0	99.9	0.0	0.0	0.0
	G	99.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.8	0.0	0.0	0.1	0.0	99.8	0.0	0.0	99.8	0.0	99.9	99.8	99.8	
	T	0.1	0.0	0.0	19.3	0.0	22.9	0.0	0.0	0.0	0.1	0.3	0.0	99.9	0.0	0.1	0.0	0.0	99.9	0.0	0.1	0.0	0.1	0.0

HEK293位点3		G	G	C	C_4	C_5	A	G	A	C	T	G	A	G	C	C	A	C	G	T	G	A	T	G	G
1.85 uM BE1	A	0.2	0.1	0.1	0.1	0.0	99.9	0.2	99.9	0.1	0.0	0.2	99.9	0.1	0.1	100.0	0.1	0.1	0.0	0.2	98.6	0.0	0.1	0.1
	C	0.0	0.0	92.2	74.8	71.5	0.0	0.0	0.0	96.6	0.0	0.0	0.0	0.0	99.8	0.0	99.7	0.0	0.0	0.0	1.3	0.0	0.0	0.0
	G	99.7	99.8	0.0	0.0	0.0	0.0	99.8	0.0	0.0	0.0	99.8	0.0	99.8	0.0	0.0	0.0	99.8	0.0	99.8	0.0	0.0	99.9	99.8
	T	0.1	0.1	7.7	25.1	28.5	0.0	0.0	0.0	3.3	99.9	0.1	0.0	0.0	0.1	0.0	0.2	0.0	99.9	0.0	0.0	100.0	0.0	0.0

HEK293位点4		G	G	C	A	C_5	T	G	C	G	G	C	T	G	C	G	A	G	T	G	G	G	G	G
1.85 uM BE1	A	0.3	0.2	0.1	100.0	0.1	0.0	0.2	0.1	0.2	0.2	0.1	0.0	0.3	0.5	100.0	0.1	0.1	0.0	0.1	0.1	0.1	0.2	0.2
	C	0.0	0.0	98.8	0.0	60.1	0.0	0.0	97.0	0.0	0.0	99.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	99.7	99.7	0.0	0.0	0.0	0.0	99.8	0.0	99.8	99.8	0.0	0.0	99.7	99.4	0.0	99.8	99.8	0.1	99.8	99.8	99.8	99.8	99.7
	T	0.0	0.1	1.1	0.0	39.8	100.0	0.1	2.9	0.0	0.1	0.5	100.0	0.0	0.1	0.0	0.0	0.0	99.9	0.0	0.1	0.1	0.0	0.0

RNF2		G	T	C_3	A	T	C_6	T	T	A	G	T	C_{12}	A	T	T	T	A	C	C	T	G	A	G	G
1.85 uM BE1	A	0.1	0.0	0.1	99.9	0.0	0.0	0.0	0.0	99.9	0.1	0.0	0.1	99.9	0.0	0.0	99.9	0.1	0.0	0.0	0.1	98.6	0.0	0.0
	C	0.0	0.0	59.1	0.0	0.0	57.8	0.0	0.0	0.1	0.0	0.0	86.4	0.0	0.0	0.0	0.1	99.9	99.9	0.0	0.0	1.4	0.0	0.0
	G	99.9	0.1	0.0	0.0	0.0	0.0	0.0	0.0	99.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.8	0.0	99.9	99.9
	T	0.0	99.9	40.8	0.0	100.0	42.1	100.0	100.0	0.0	0.1	99.9	13.4	0.0	100.0	99.9	0.0	0.0	0.0	99.9	0.1	0.0	0.1	0.0

图37B

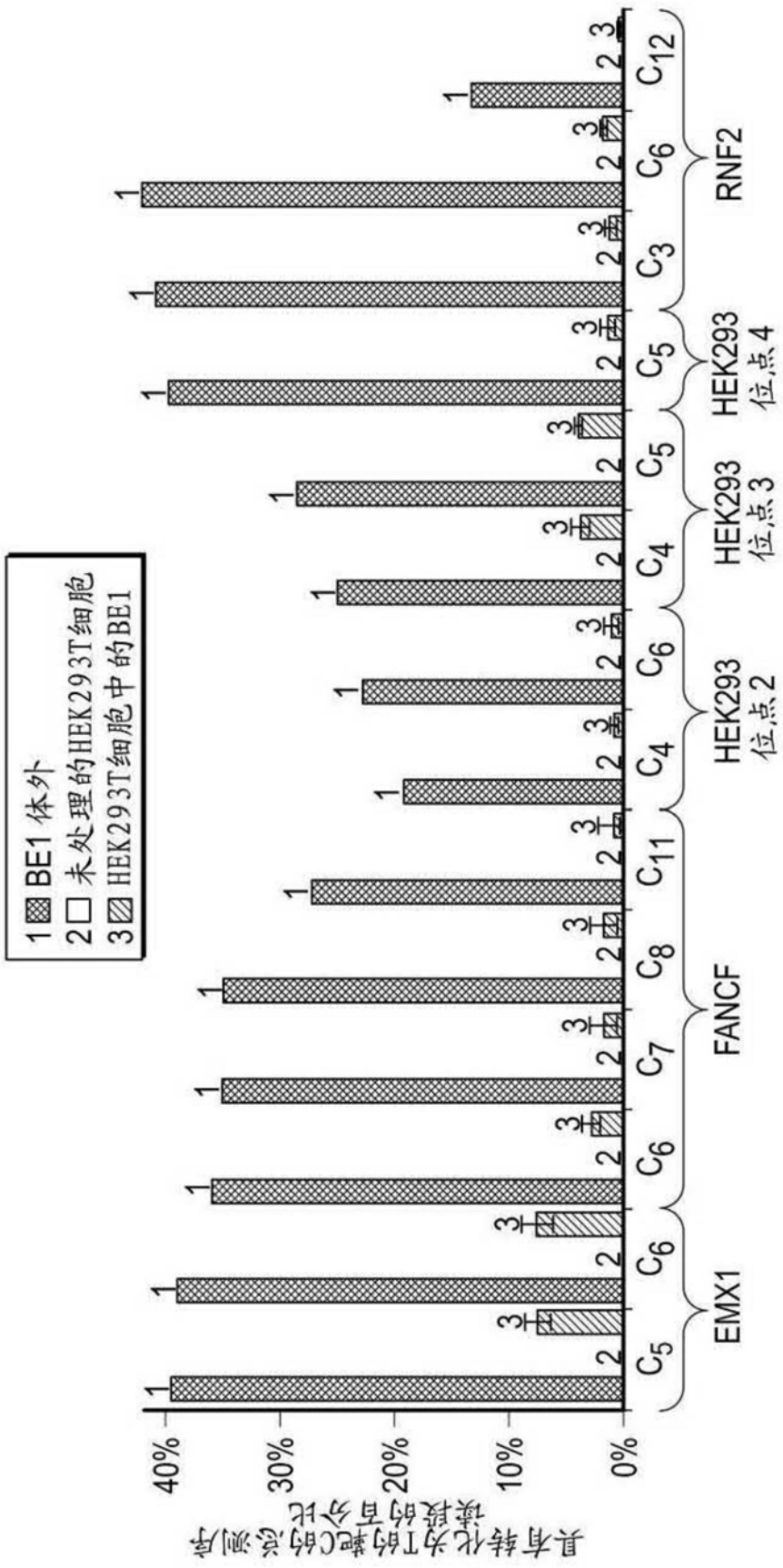

1 BE1 体外
2 未处理的HEK293T细胞
3 HEK293T细胞中的BE1
具有转化为T的靶C的总测序读段的百分比
40%
30%
20%
10%
0%
EMX1
FANCF
HEK293 位点2
HEK293 位点3
HEK293 位点4
RNF2
C5
C6
C7
C8
C11
C12
C3
C4

图37C

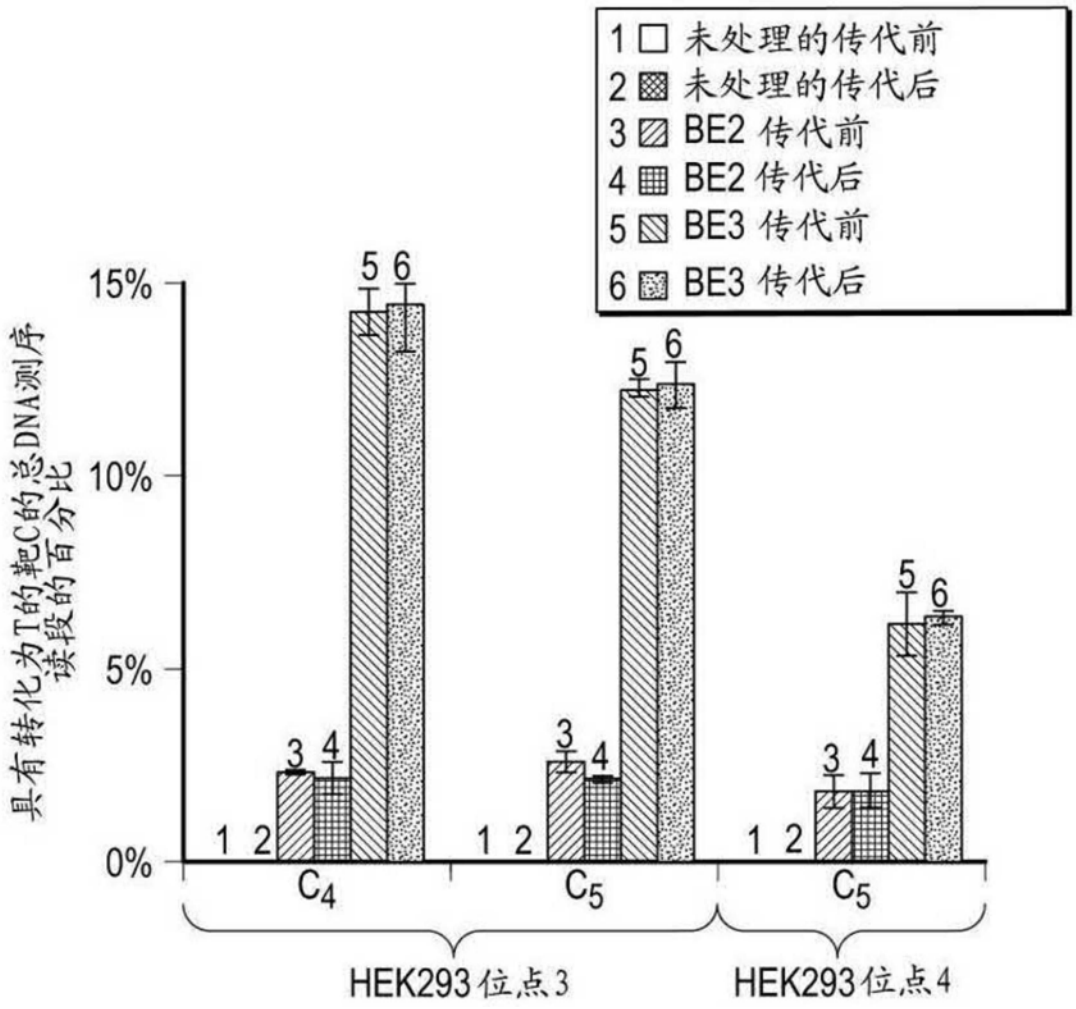

1 未处理的传代前
2 未处理的传代后
3 BE2 传代前
4 BE2 传代后
5 BE3 传代前
6 BE3 传代后
具有转化为T的靶C的总DNA测序读段的百分比
15%
10%
5%
0%
C4
C5
C5
HEK293位点3
HEK293位点4

图38

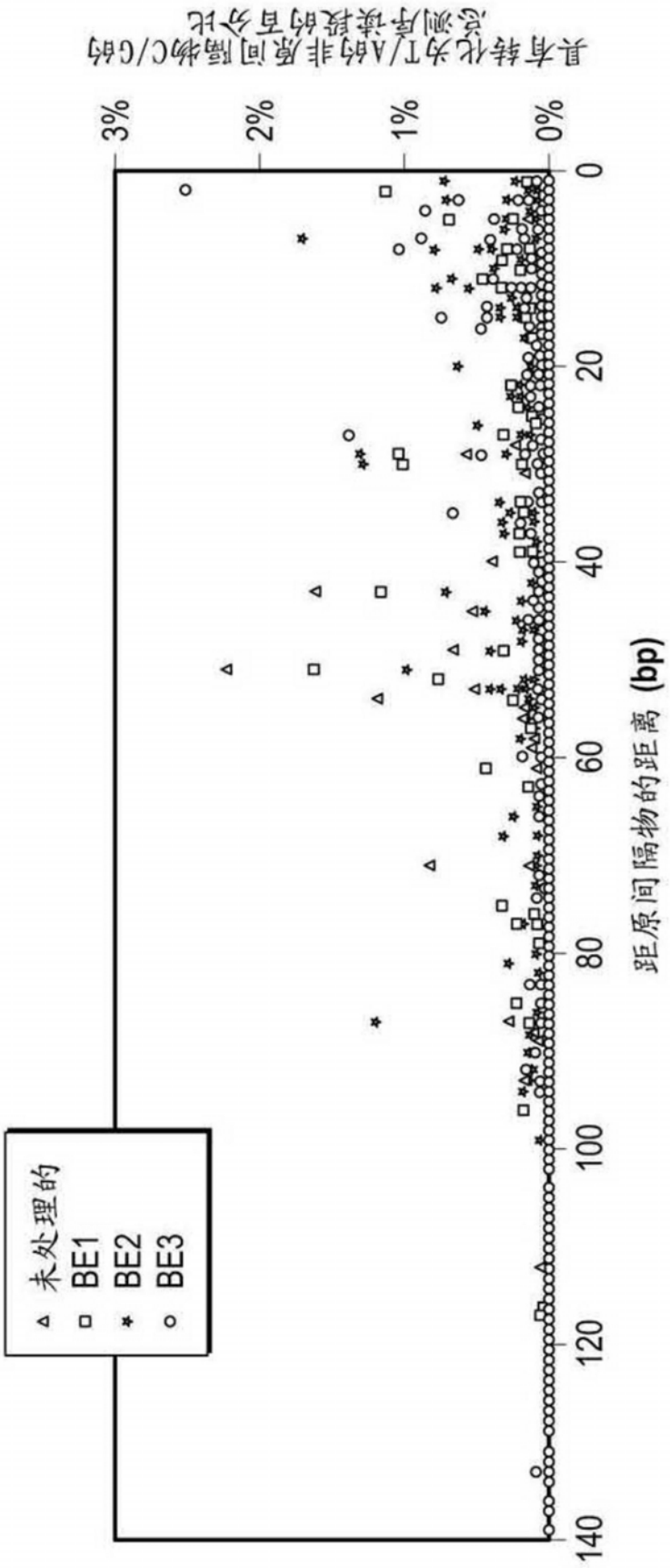

具有转化为T/A的非原间隔物C/G的
总测序读段的百分比
3%
2%
1%
0%
0
20
40
60
80
100
120
140
距原间隔物的距离 (bp)
未处理的
BE1
BE2
BE3

图39A

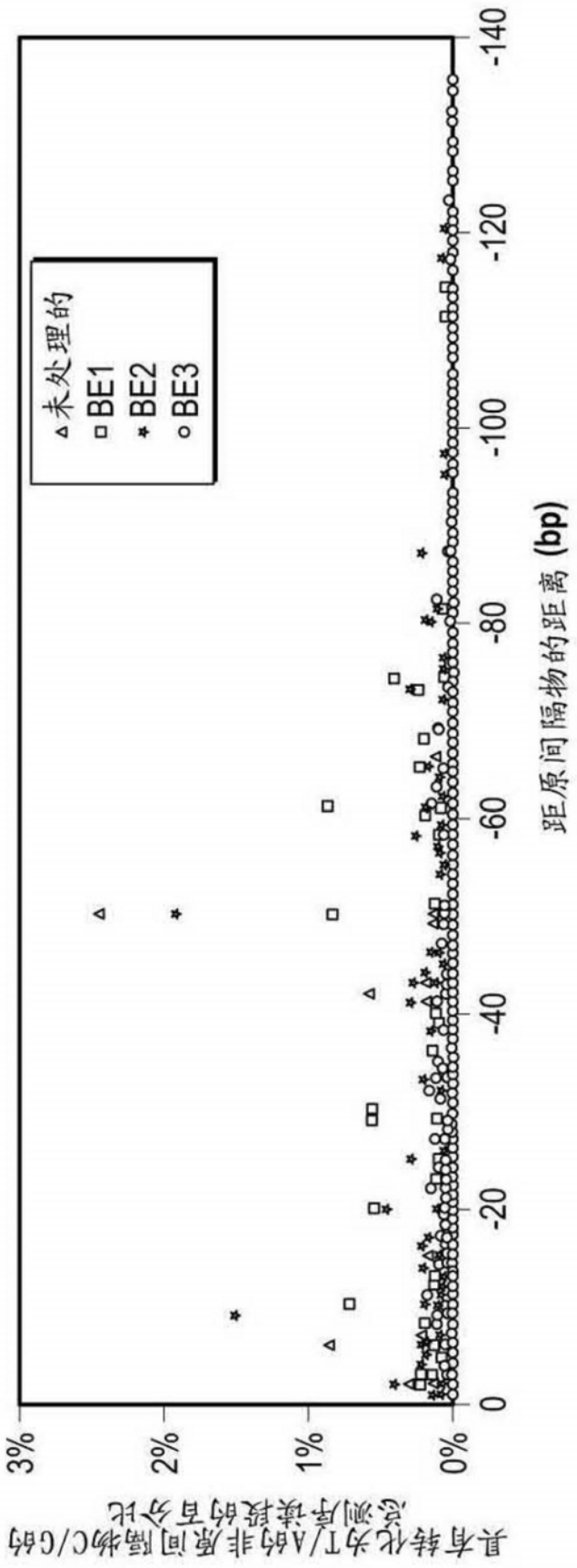

△未处理的
□BE1
*BE2
○BE3
0
-20
-40
-60
-80
-100
-120
-140
距原间隔物的距离 (bp)
0%
1%
2%
3%
具有转化为T/A的非原间隔物C/G的
经测序读段的百分比

图39B

非原间隔物 C/Gs	平均 C/G (%)	平均 T/A (%)	最低 T/A (%)	最高 T/A (%)
未处理的	99.95 ± 0.14	0.02 ± 0.02	0.00	2.44
BE1	99.95 ± 0.24	0.03 ± 0.03	0.00	1.64
BE2	99.95 ± 0.13	0.03 ± 0.03	0.00	1.92
BE3	99.97 ± 0.09	0.02 ± 0.02	0.00	2.52

图39C

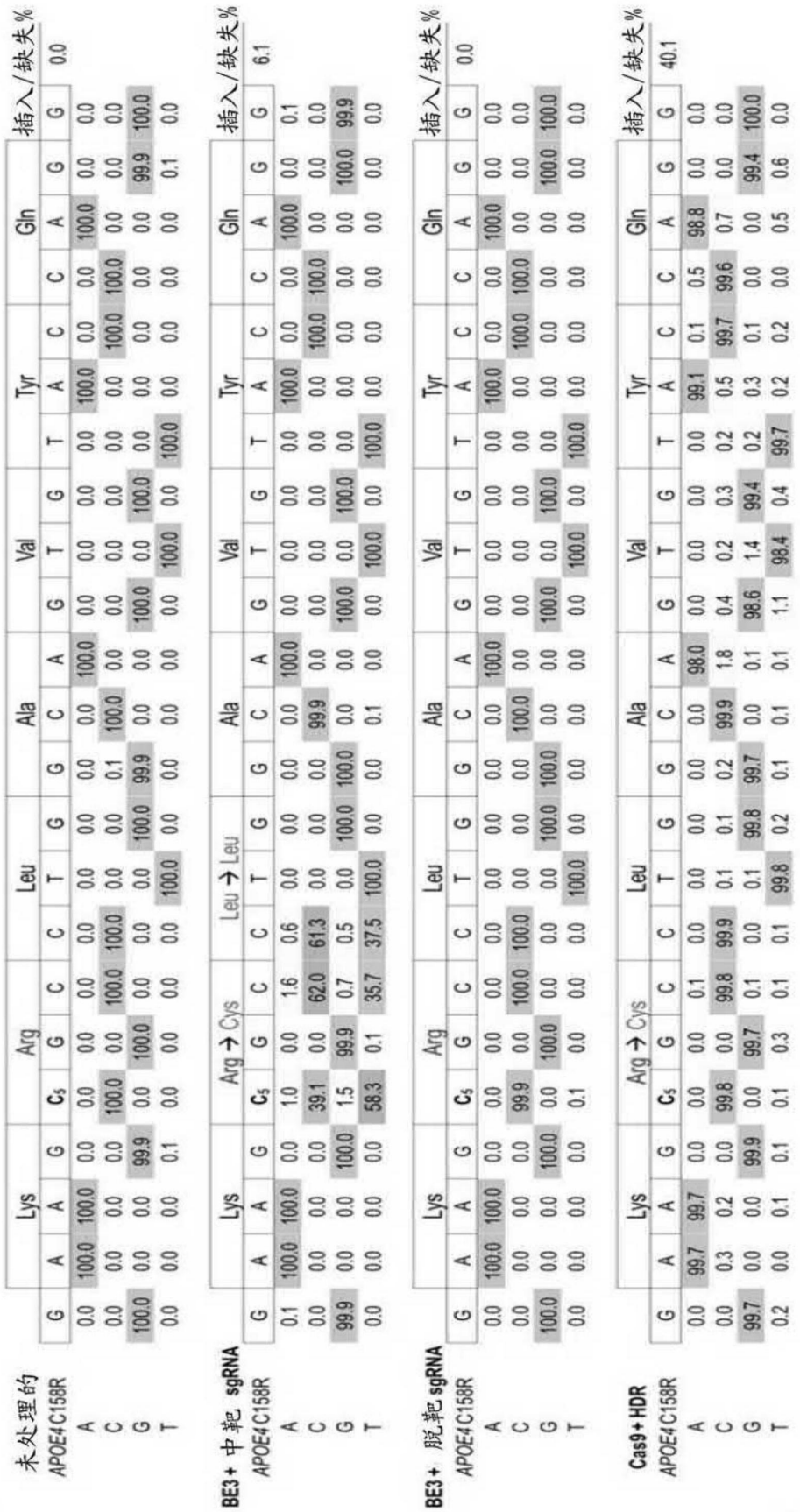

未处理的
APOE4C158R
Lys Arg Leu Ala Val Tyr Gln 插入/缺失%
G A A G C5 G C T G G C A G T A C C A G G
A 0.0 100.0 100.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 100.0 0.0 0.0 0.0 100.0 0.0 0.0 0.0 0.0
C 0.0 0.0 0.0 0.0 100.0 0.0 100.0 100.0 0.0 0.0 0.1 100.0 0.0 0.0 0.0 0.0 0.0 100.0 100.0 0.0 0.0
G 100.0 0.0 0.0 99.9 0.0 100.0 0.0 0.0 0.0 100.0 99.9 0.0 0.0 100.0 0.0 0.0 0.0 0.0 0.0 99.9 100.0
T 0.0 0.0 0.0 0.1 0.0 0.0 0.0 0.0 100.0 0.0 0.0 0.0 0.0 0.0 100.0 0.0 0.0 0.0 0.0 0.1 0.0 0.0
BE3+ 中靶 sgRNA
APOE4C158R
Lys Arg → Cys Leu → Leu Ala Val Tyr Gln 插入/缺失%
G A A G C5 G C C T G G C A G T A C C A G G
A 0.1 100.0 100.0 0.0 1.0 0.0 1.6 0.6 0.0 0.0 0.0 0.0 100.0 0.0 0.0 0.0 100.0 0.0 0.0 0.0 6.1
C 0.0 0.0 0.0 0.0 39.1 0.0 62.0 61.3 0.0 0.0 0.0 99.9 0.0 0.0 0.0 0.0 0.0 100.0 100.0 0.0 0.1
G 99.9 0.0 0.0 100.0 1.5 99.9 0.7 0.5 0.0 100.0 100.0 0.0 0.0 100.0 0.0 0.0 0.0 0.0 0.0 100.0 99.9
T 0.0 0.0 0.0 0.0 58.3 0.1 35.7 37.5 100.0 0.0 0.0 0.1 0.0 0.0 100.0 0.0 0.0 0.0 0.0 0.0 0.0
BE3+ 脱靶 sgRNA
APOE4C158R
Lys Arg Leu Ala Val Tyr Gln 插入/缺失%
G A A G C5 G C C T G G C A G T A C C A G G
A 0.0 100.0 100.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 100.0 0.0 0.0 0.0 100.0 0.0 0.0 0.0 0.0
C 0.0 0.0 0.0 0.0 99.9 0.0 100.0 100.0 0.0 0.0 0.0 100.0 0.0 0.0 0.0 0.0 0.0 100.0 100.0 0.0 0.0
G 100.0 0.0 0.0 100.0 0.0 100.0 0.0 0.0 0.0 100.0 100.0 0.0 0.0 100.0 0.0 0.0 0.0 0.0 0.0 100.0 100.0
T 0.0 0.0 0.0 0.0 0.1 0.0 0.0 0.0 100.0 0.0 0.0 0.0 0.0 0.0 100.0 0.0 0.0 0.0 0.0 0.0 0.0
Cas9+HDR
APOE4C158R
Lys Arg → Cys Leu Ala Val Tyr Gln 插入/缺失%
G A A G C5 G C C T G G C A G T A C C A G G
A 0.0 99.7 99.7 0.0 0.0 0.0 0.1 0.0 0.0 0.0 0.0 0.0 96.0 0.0 0.0 0.0 99.1 0.1 0.5 98.8 0.0 0.0
C 0.0 0.3 0.2 0.0 99.8 0.0 99.8 99.9 0.1 0.1 0.2 99.9 1.8 0.4 0.2 0.3 0.2 0.5 99.7 99.6 0.7 0.0 0.0
G 99.7 0.0 0.0 99.9 0.0 99.7 0.1 0.0 0.1 99.8 99.7 0.0 0.1 98.6 1.4 99.4 0.2 0.3 0.1 0.0 0.0 99.4 100.0
T 0.2 0.0 0.1 0.1 0.1 0.3 0.1 0.1 99.8 0.2 0.1 0.1 0.1 1.1 98.4 0.4 99.7 0.2 0.2 0.0 0.5 0.6 0.0

图40A

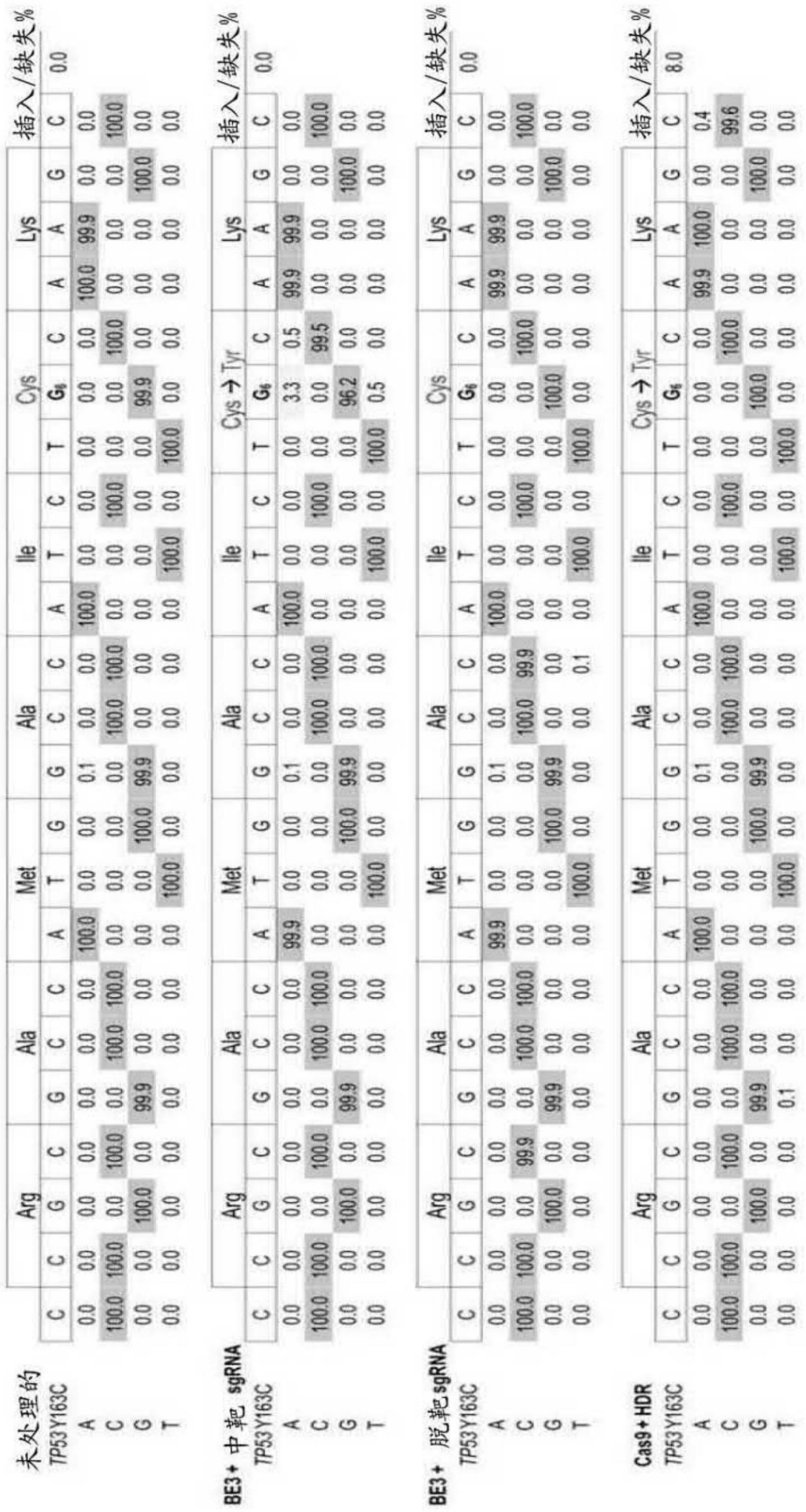

未处理的
TP53Y163C
BE3+ 中靶 sgRNA
TP53Y163C
BE3+ 脱靶 sgRNA
TP53Y163C
Cas9+HDR
TP53Y163C
Arg
Ala
Met
Ile
Cys
Cys → Tyr
Lys
插入/缺失%

图40B

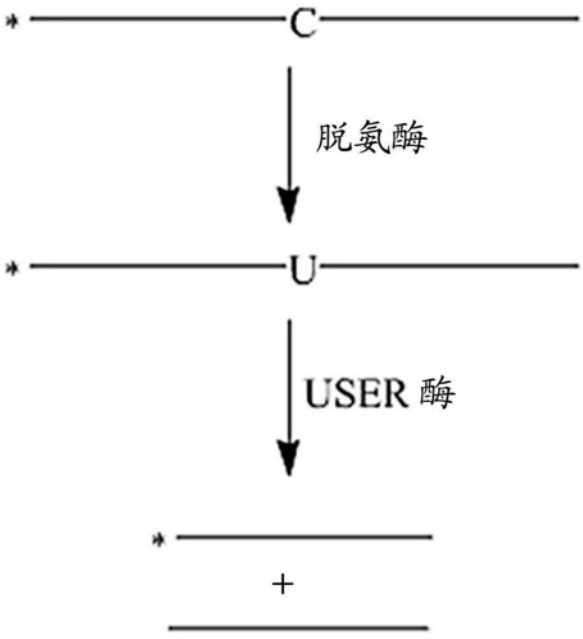


图41

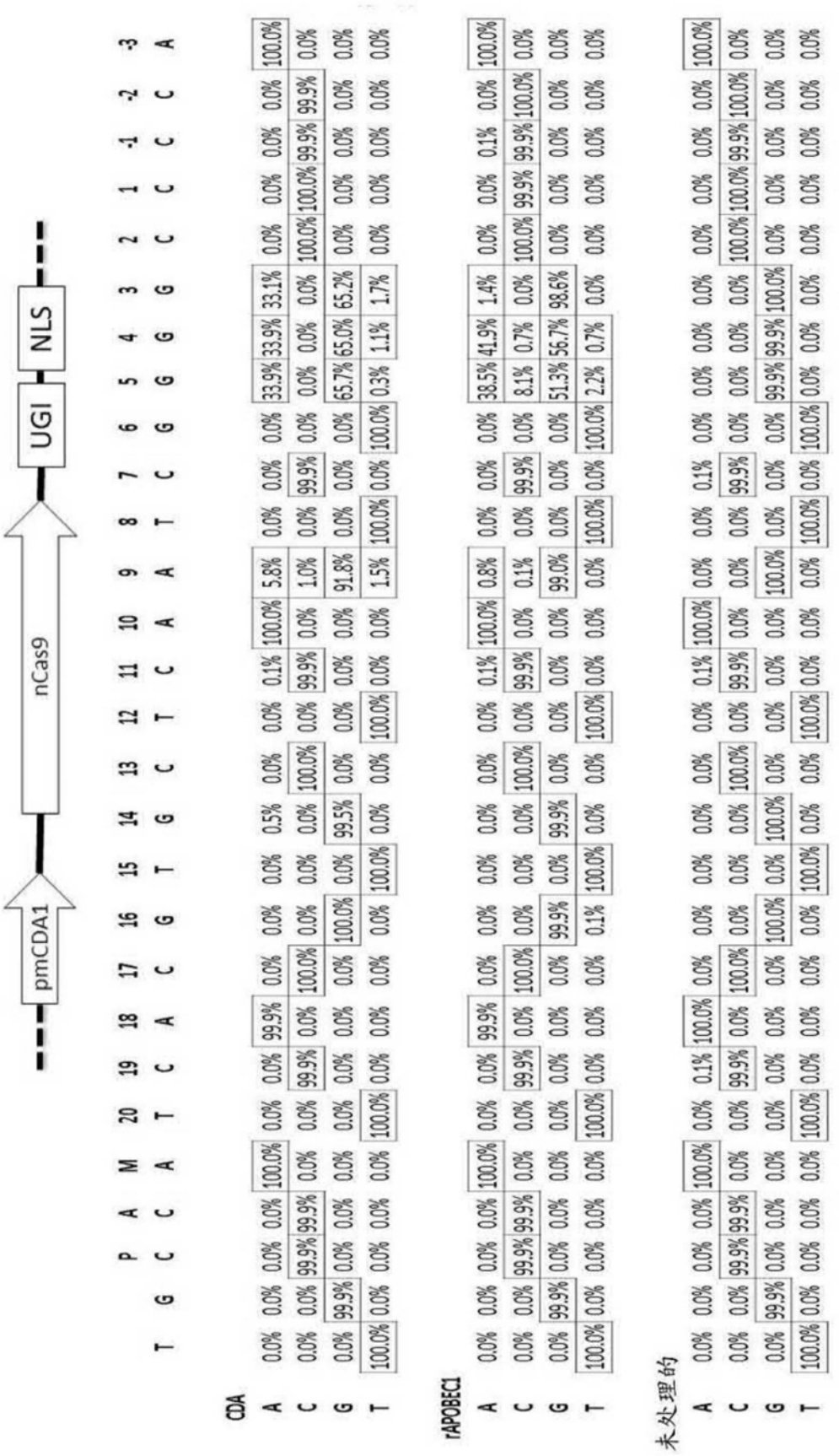

pmCDA1
nCas9
UGI
NLS
P A M 20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1 -1 -2 -3
T G C C A T C A C G T G C T C A A T C G G G G C C C C A
CDA
A 0.0% 0.0% 0.0% 0.0% 100.0% 0.0% 0.0% 99.9% 0.0% 0.0% 0.0% 0.5% 0.0% 0.0% 0.1% 100.0% 5.8% 0.0% 0.0% 0.0% 33.9% 33.9% 33.1% 0.0% 0.0% 0.0% 100.0%
C 0.0% 0.0% 99.9% 99.9% 0.0% 0.0% 99.9% 0.0% 100.0% 0.0% 0.0% 0.0% 100.0% 0.0% 99.9% 0.0% 1.0% 0.0% 99.9% 0.0% 0.0% 0.0% 0.0% 100.0% 100.0% 99.9% 99.9% 0.0%
G 0.0% 99.9% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 100.0% 0.0% 0.0% 99.5% 0.0% 0.0% 0.0% 0.0% 91.8% 0.0% 0.0% 0.0% 65.7% 65.0% 65.2% 0.0% 0.0% 0.0% 0.0%
T 100.0% 0.0% 0.0% 0.0% 0.0% 100.0% 0.0% 0.0% 0.0% 0.0% 100.0% 0.0% 0.0% 100.0% 0.0% 0.0% 1.5% 100.0% 0.0% 100.0% 0.3% 1.1% 1.7% 0.0% 0.0% 0.0% 0.0%
rAPOBEC1
A 0.0% 0.0% 0.0% 0.0% 100.0% 0.0% 0.0% 99.9% 0.0% 0.0% 0.0% 0.0% 0.0% 0.1% 100.0% 0.8% 0.0% 0.0% 0.0% 38.5% 41.9% 1.4% 0.0% 0.0% 0.1% 0.0% 100.0%
C 0.0% 0.0% 99.9% 99.9% 0.0% 0.0% 99.9% 0.0% 100.0% 0.0% 0.0% 0.0% 100.0% 0.0% 99.9% 0.1% 0.0% 99.9% 0.0% 8.1% 0.7% 0.0% 100.0% 99.9% 99.9% 100.0% 0.0%
G 0.0% 99.9% 0.0% 0.0% 0.0% 0.0% 0.0% 99.9% 0.0% 0.0% 0.0% 99.9% 0.0% 0.0% 0.0% 99.0% 0.0% 0.0% 0.0% 51.3% 56.7% 98.6% 0.0% 0.0% 0.0% 0.0% 0.0%
T 100.0% 0.0% 0.0% 0.0% 0.0% 100.0% 0.0% 0.0% 0.1% 100.0% 0.0% 0.0% 100.0% 0.0% 0.0% 0.0% 100.0% 0.0% 100.0% 2.2% 0.7% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0%
未处理的
A 0.0% 0.0% 0.0% 0.0% 100.0% 0.0% 0.1% 100.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.1% 100.0% 0.0% 0.0% 0.1% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 100.0%
C 0.0% 0.0% 99.9% 99.9% 0.0% 0.0% 99.9% 0.0% 100.0% 0.0% 0.0% 0.0% 100.0% 0.0% 99.9% 0.0% 0.0% 99.9% 0.0% 0.0% 0.0% 0.0% 100.0% 100.0% 99.9% 100.0% 0.0%
G 0.0% 99.9% 0.0% 0.0% 0.0% 0.0% 0.0% 100.0% 0.0% 0.0% 0.0% 100.0% 0.0% 0.0% 0.0% 100.0% 0.0% 0.0% 0.0% 99.9% 99.9% 100.0% 0.0% 0.0% 0.0% 0.0% 0.0%
T 100.0% 0.0% 0.0% 0.0% 0.0% 100.0% 0.0% 0.0% 0.0% 100.0% 0.0% 0.0% 100.0% 0.0% 0.0% 0.0% 100.0% 0.0% 100.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0%

图42

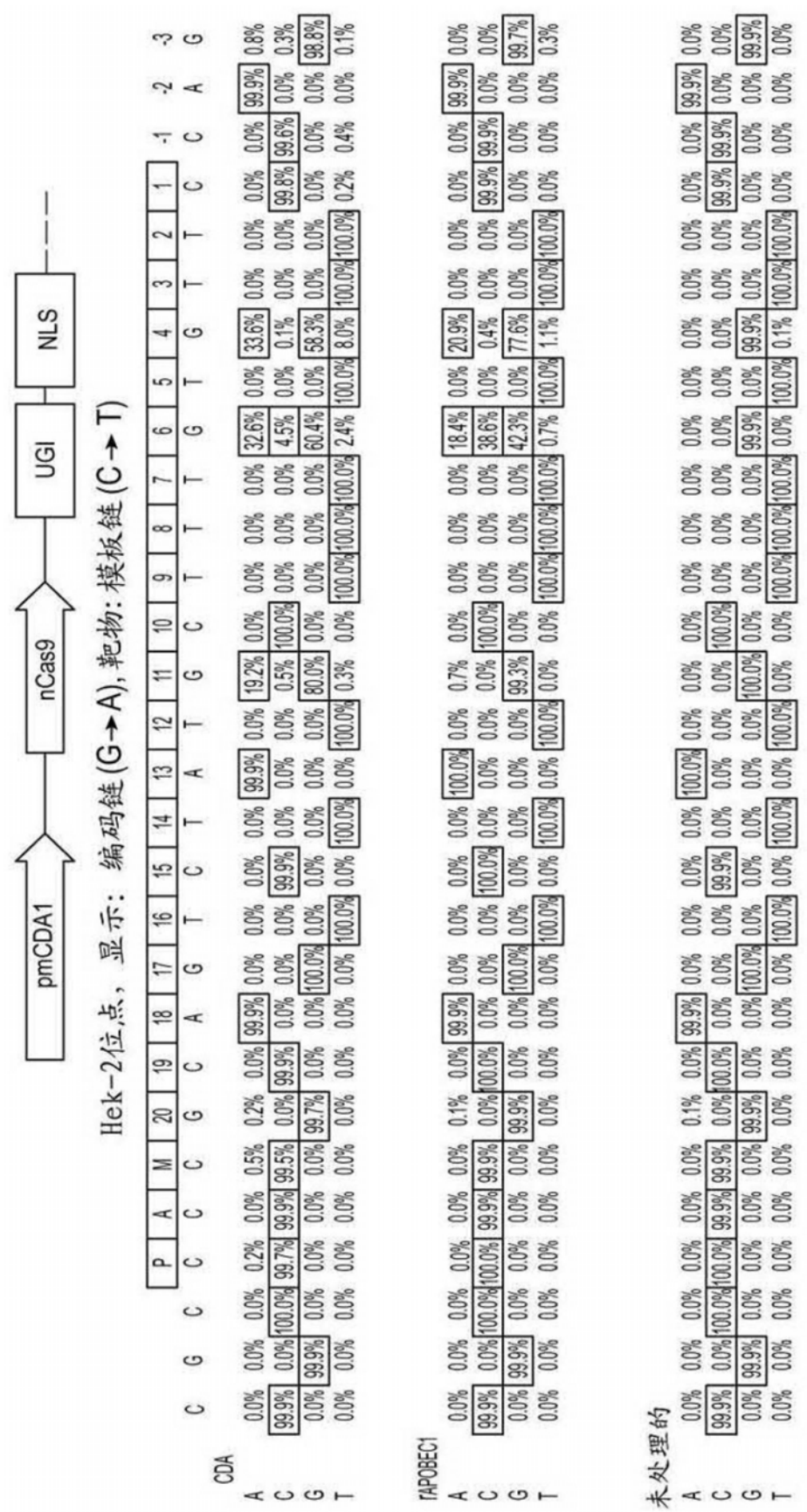

pmCDA1
nCas9
UGI
NLS
Hek-2位点，显示：编码链(G→A)，靶物：模板链(C→T)
P A M 20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1 -1 -2 -3
C G C C C C G C A G T C T A T G C T T T G T T C C A G
CDA
A 0.0% 0.0% 0.0% 0.2% 0.0% 0.5% 0.2% 0.0% 99.9% 0.0% 0.0% 0.0% 99.9% 0.0% 19.2% 0.0% 0.0% 0.0% 0.0% 32.6% 0.0% 33.6% 0.0% 0.0% 0.0% 99.9% 0.8%
C 99.9% 0.0% 100.0% 99.7% 99.9% 99.5% 0.0% 99.9% 0.0% 0.0% 0.0% 0.0% 0.5% 100.0% 0.0% 0.0% 0.0% 4.5% 0.0% 0.1% 0.0% 0.0% 99.8% 99.6% 0.0% 0.3%
G 0.0% 99.9% 0.0% 0.0% 0.0% 0.0% 99.7% 0.0% 0.0% 100.0% 0.0% 0.0% 0.0% 0.0% 80.0% 0.0% 0.0% 0.0% 0.0% 60.4% 0.0% 58.3% 0.0% 0.0% 0.0% 0.0% 98.8%
T 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 100.0% 0.0% 0.0% 100.0% 0.0% 0.3% 0.0% 100.0% 100.0% 100.0% 2.4% 100.0% 8.0% 100.0% 100.0% 0.2% 0.4% 0.0% 0.1%
rAPOBEC1
A 0.0% 0.0% 0.0% 0.0% 0.0% 0.1% 0.0% 99.9% 0.0% 0.0% 0.0% 100.0% 0.0% 0.7% 0.0% 0.0% 0.0% 0.0% 18.4% 0.0% 20.9% 0.0% 0.0% 0.0% 0.0% 99.9% 0.0%
C 99.9% 0.0% 100.0% 100.0% 99.9% 99.9% 0.0% 0.0% 0.0% 100.0% 0.0% 0.0% 100.0% 0.0% 0.0% 100.0% 0.0% 0.0% 38.6% 0.0% 0.4% 0.0% 0.0% 99.9% 99.9% 0.0% 0.0%
G 0.0% 99.9% 0.0% 0.0% 0.0% 0.0% 99.9% 0.0% 0.0% 100.0% 0.0% 0.0% 0.0% 0.0% 99.3% 0.0% 0.0% 0.0% 42.3% 0.0% 77.6% 0.0% 0.0% 0.0% 0.0% 99.7%
T 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 100.0% 0.0% 0.0% 100.0% 0.0% 0.0% 100.0% 0.0% 0.0% 0.7% 100.0% 1.1% 100.0% 100.0% 0.0% 0.0% 0.0% 0.3%
未处理的
A 0.0% 0.0% 0.0% 0.0% 0.0% 0.1% 0.0% 99.9% 0.0% 0.0% 0.0% 100.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 99.9% 0.0%
C 99.9% 0.0% 100.0% 100.0% 99.9% 99.9% 0.0% 0.0% 0.0% 100.0% 0.0% 0.0% 100.0% 0.0% 0.0% 100.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 99.9% 99.9% 0.0% 0.0% 0.0%
G 0.0% 99.9% 0.0% 0.0% 0.0% 0.0% 99.9% 0.0% 0.0% 100.0% 0.0% 0.0% 0.0% 0.0% 100.0% 0.0% 0.0% 0.0% 99.9% 0.0% 99.9% 0.0% 0.0% 0.0% 0.0% 99.9%
T 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 100.0% 0.0% 0.0% 100.0% 0.0% 0.0% 100.0% 0.0% 0.0% 0.0% 100.0% 0.1% 100.0% 100.0% 0.0% 0.0% 0.0% 0.0%

图43

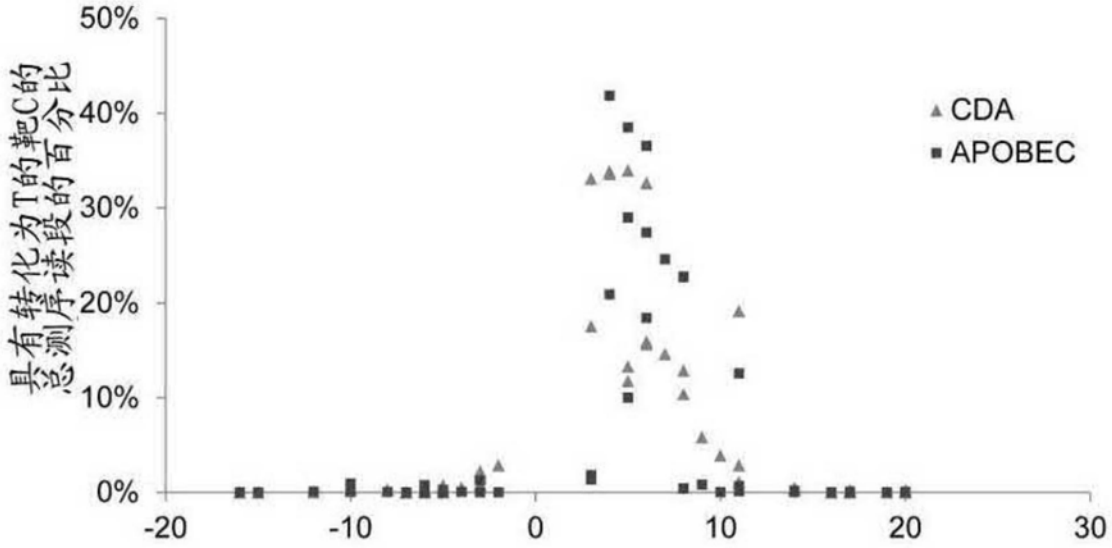

50%
40%
30%
20%
10%
0%
-20
-10
0
10
20
30
具有转化为T的靶C的总测序读段的百分比
CDA
APOBEC

图44

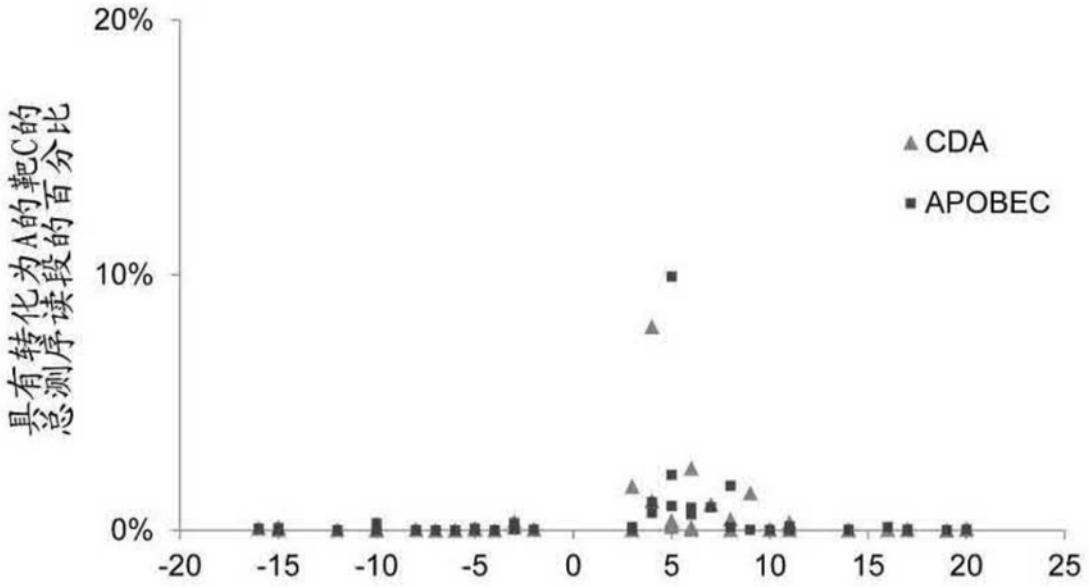

20%
10%
0%
-20
-15
-10
-5
0
5
10
15
20
25
具有转化为A的靶C的总测序读段的百分比
CDA
APOBEC

图45

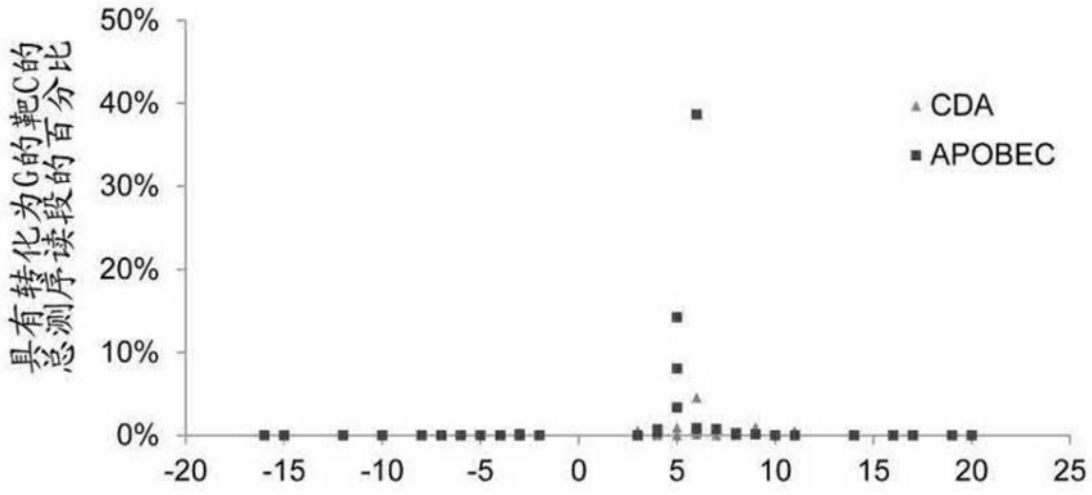

具有转化为G的靶C的总测序读段的百分比
50%
40%
30%
20%
10%
0%
-20
-15
-10
-5
0
5
10
15
20
25
CDA
APOBEC

图46

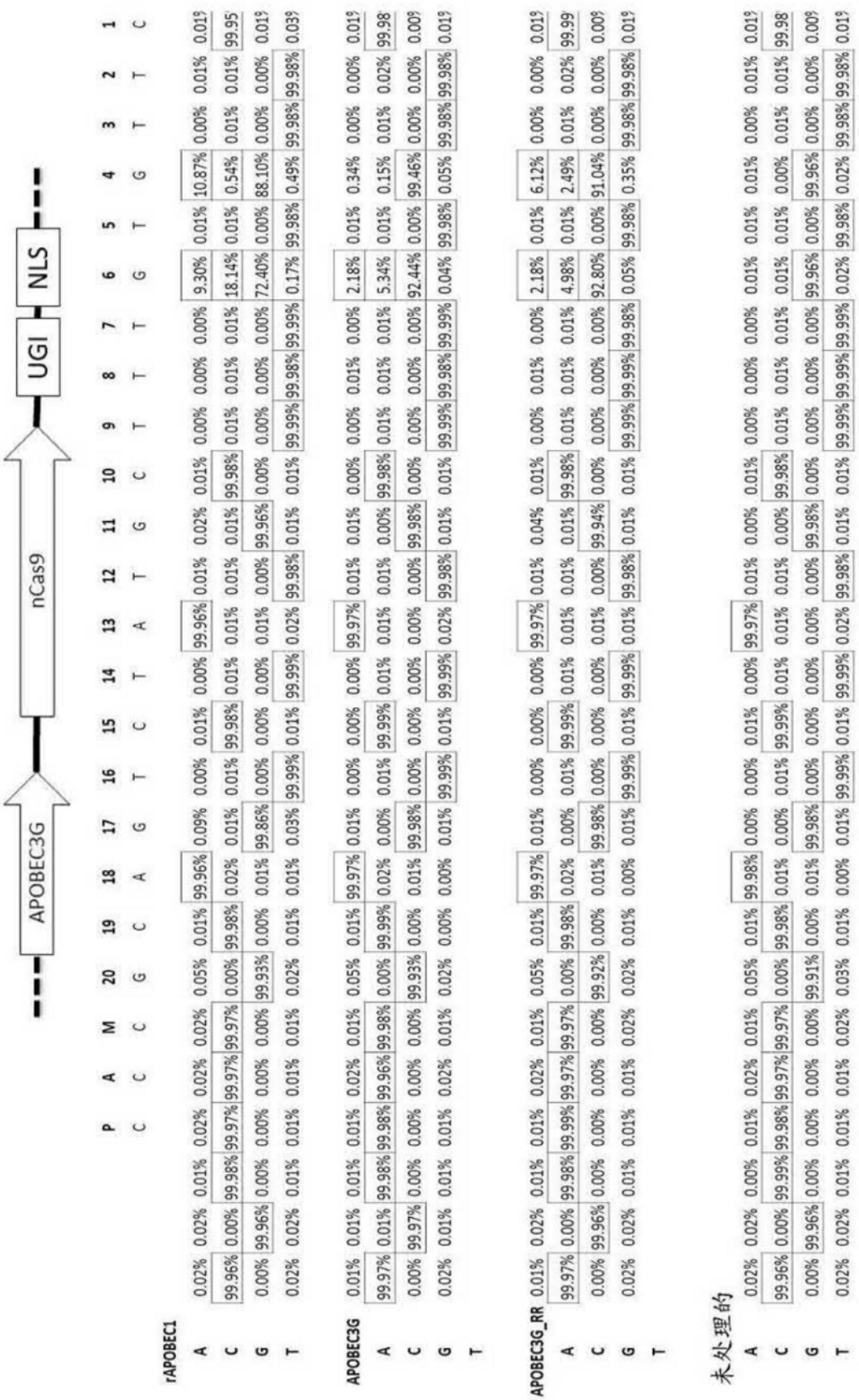

APOBEC3G
nCas9
UGI
NLS
P A M 20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1
C C C G C A G T C T A T G C T T T G T T C
rAPOBEC1
A 0.02% 0.02% 0.01% 0.02% 0.02% 0.05% 0.01% 99.96% 0.09% 0.00% 0.01% 0.00% 99.96% 0.01% 0.02% 0.01% 0.00% 0.00% 0.00% 9.30% 0.01% 10.87% 0.00% 0.01% 0.01%
C 99.96% 0.00% 99.98% 99.97% 99.97% 0.00% 99.98% 0.02% 0.01% 0.01% 99.98% 0.01% 0.01% 0.01% 99.98% 0.01% 0.01% 0.01% 18.14% 0.01% 0.54% 0.01% 0.01% 99.95%
G 0.00% 99.96% 0.00% 0.00% 0.00% 0.00% 99.93% 0.00% 0.01% 99.86% 0.00% 0.00% 0.01% 0.00% 99.96% 0.00% 0.00% 0.00% 72.40% 0.00% 88.10% 0.00% 0.00% 0.01%
T 0.02% 0.02% 0.01% 0.01% 0.01% 0.01% 0.02% 0.01% 0.01% 0.03% 99.99% 0.01% 99.99% 0.02% 0.01% 0.01% 99.98% 99.98% 99.99% 0.17% 99.98% 0.49% 99.98% 99.98% 0.03%
APOBEC3G
A 0.01% 0.01% 0.01% 0.01% 0.02% 0.01% 0.05% 0.01% 99.97% 0.01% 0.00% 0.00% 0.00% 99.97% 0.01% 0.01% 0.00% 0.00% 0.01% 0.00% 2.18% 0.01% 0.34% 0.00% 0.00% 0.01%
C 99.97% 0.01% 99.98% 99.98% 99.96% 99.98% 0.00% 99.99% 0.02% 0.00% 0.01% 99.99% 0.01% 0.01% 0.01% 0.00% 99.98% 0.01% 0.01% 0.01% 5.34% 0.01% 0.15% 0.01% 0.02% 99.98%
G 0.00% 99.97% 0.00% 0.00% 0.00% 0.00% 99.93% 0.00% 0.01% 99.98% 0.00% 0.00% 0.00% 0.00% 99.98% 0.00% 0.00% 0.00% 92.44% 0.00% 99.46% 0.00% 0.00% 0.00%
T 0.02% 0.01% 0.01% 0.01% 0.02% 0.01% 0.02% 0.00% 0.00% 0.01% 99.99% 0.01% 0.01% 0.02% 99.98% 0.01% 0.01% 99.99% 99.98% 99.99% 0.04% 99.98% 0.05% 99.98% 99.98% 0.01%
APOBEC3G_RR 0.01% 0.02% 0.01% 0.01% 0.02% 0.01% 0.05% 0.01% 99.97% 0.01% 0.00% 0.00% 0.00% 99.97% 0.01% 0.04% 0.01% 0.00% 0.01% 0.00% 2.18% 0.01% 6.12% 0.00% 0.00% 0.01%
A 99.97% 0.00% 99.98% 99.99% 99.97% 99.97% 0.00% 99.98% 0.02% 0.00% 0.01% 99.99% 0.01% 0.01% 0.01% 0.01% 99.98% 0.01% 0.01% 0.01% 4.98% 0.01% 2.49% 0.01% 0.02% 99.99%
C 0.00% 99.96% 0.00% 0.00% 0.00% 0.00% 99.92% 0.00% 0.01% 99.98% 0.00% 0.00% 0.00% 0.01% 0.00% 99.94% 0.00% 0.00% 0.00% 0.00% 92.80% 0.00% 91.04% 0.00% 0.00% 0.00%
G 0.02% 0.02% 0.01% 0.01% 0.01% 0.02% 0.02% 0.01% 0.00% 0.01% 99.99% 0.01% 99.99% 0.01% 99.98% 0.01% 0.01% 99.99% 99.99% 99.98% 0.05% 99.98% 0.35% 99.98% 99.98% 0.01%
T
未处理的
A 0.02% 0.02% 0.00% 0.01% 0.02% 0.01% 0.05% 0.01% 99.98% 0.00% 0.00% 0.01% 0.00% 99.97% 0.01% 0.00% 0.00% 0.00% 0.01% 0.01% 0.00% 0.01% 0.01% 0.00% 0.01% 0.01%
C 99.96% 0.00% 99.99% 99.98% 99.97% 0.00% 99.98% 0.01% 0.00% 0.01% 99.99% 0.01% 0.01% 0.01% 99.98% 0.01% 0.01% 0.01% 0.01% 0.01% 0.00% 0.01% 0.01% 0.01% 99.98%
G 0.00% 99.96% 0.00% 0.00% 0.00% 0.00% 99.91% 0.00% 0.01% 99.98% 0.00% 0.00% 0.00% 0.00% 99.98% 0.00% 0.00% 0.00% 99.96% 0.00% 99.96% 0.00% 0.00% 0.00%
T 0.02% 0.02% 0.01% 0.01% 0.01% 0.02% 0.03% 0.01% 0.00% 0.01% 99.99% 0.01% 99.99% 0.02% 0.01% 0.01% 99.98% 99.99% 99.99% 0.02% 99.98% 0.02% 99.98% 99.98% 0.01%

图47

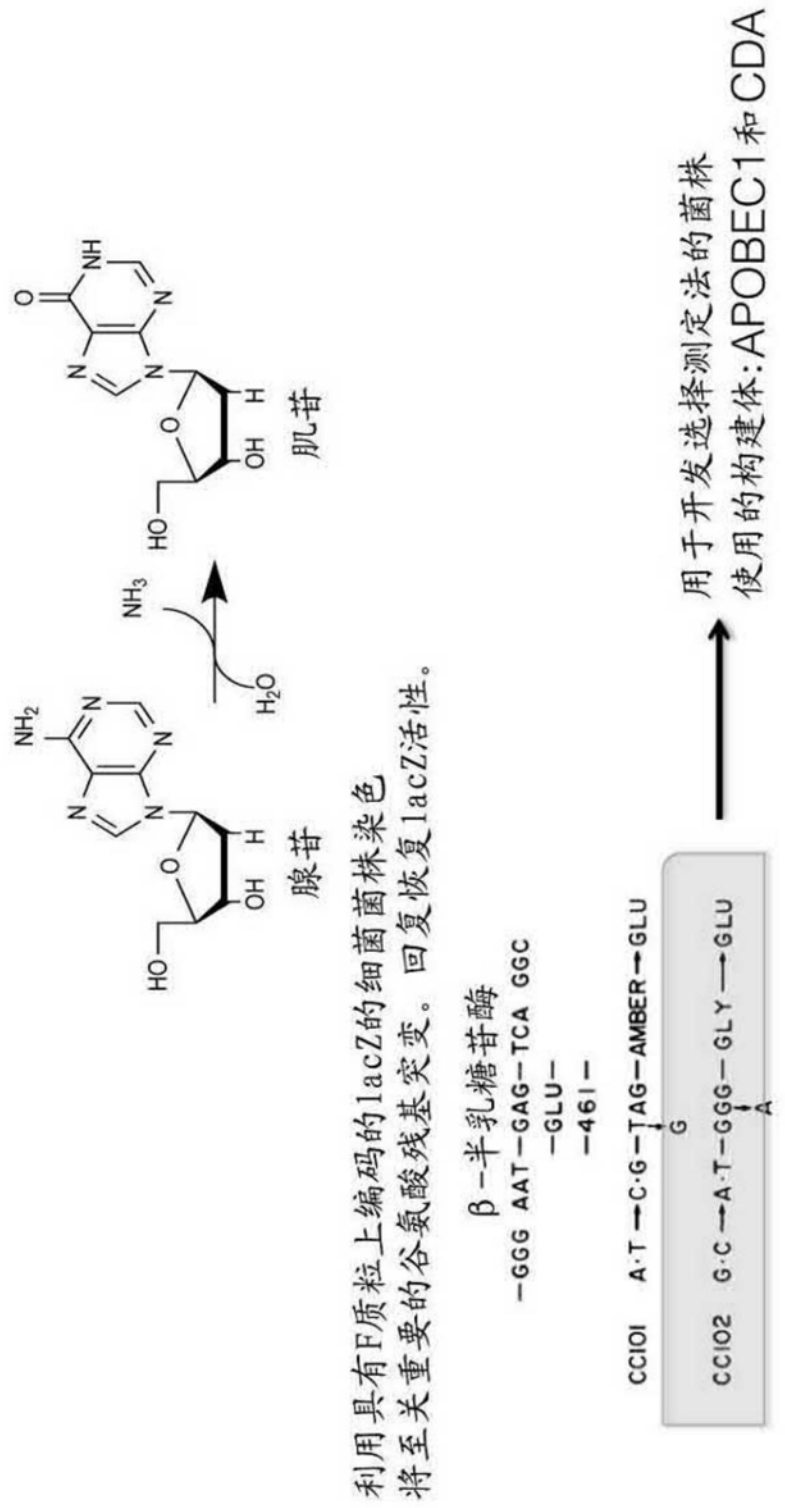

腺苷
肌苷
NH_3
H_2O
利用具有F质粒上编码的lacZ的细菌菌株染色
将至关重要的谷氨酸残基突变。回复恢复lacZ活性。
β-半乳糖苷酶
—GGG AAT—GAG—TCA GGC
—GLU—
—461—
CC101 A·T—C·G—TAG—AMBER—GLU
G
CC102 G·C—A·T—GGG—GLY—GLU
A
用于开发选择测定法的菌株
使用的构建体:APOBEC1和CDA

图48

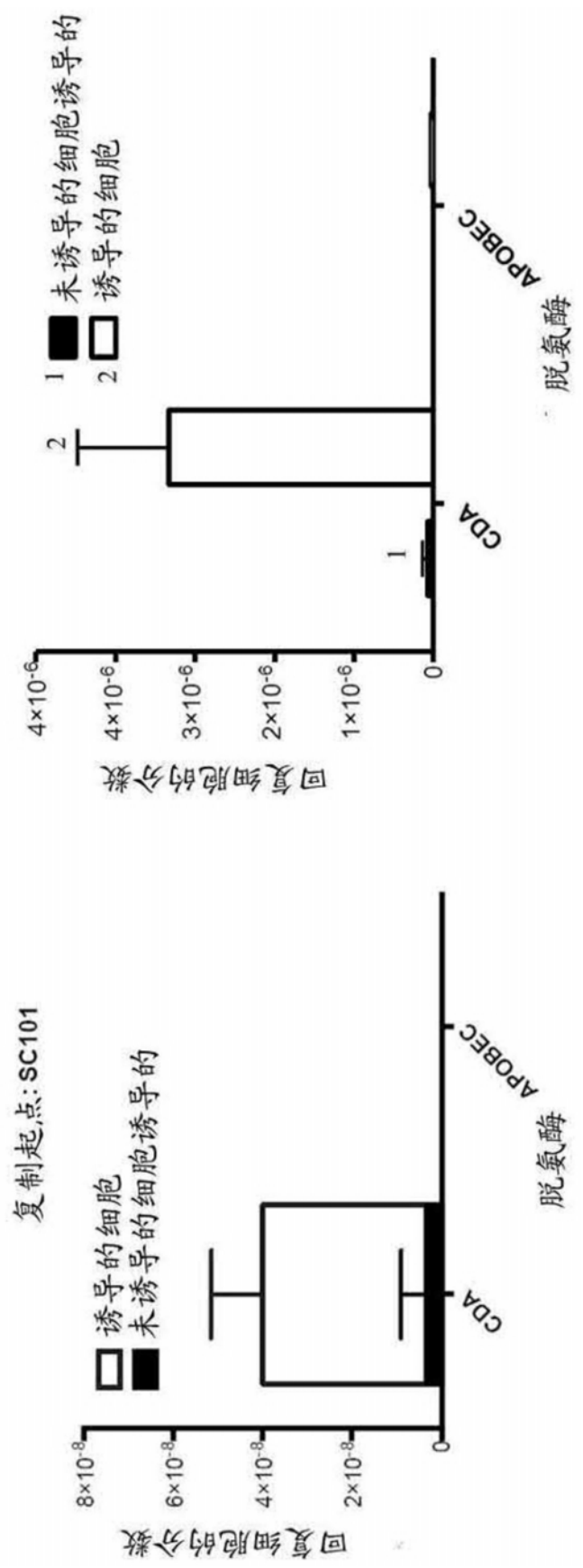

复制起点:SC101
诱导的细胞
未诱导的细胞诱导的
回复细胞的分数
8×10-8
6×10-8
4×10-8
2×10-8
0
CDA
APOBEC
脱氨酶
未诱导的细胞诱导的
诱导的细胞
回复细胞的分数
4×10-6
4×10-6
3×10-6
2×10-6
1×10-6
0
1
2
CDA
APOBEC
脱氨酶

图49

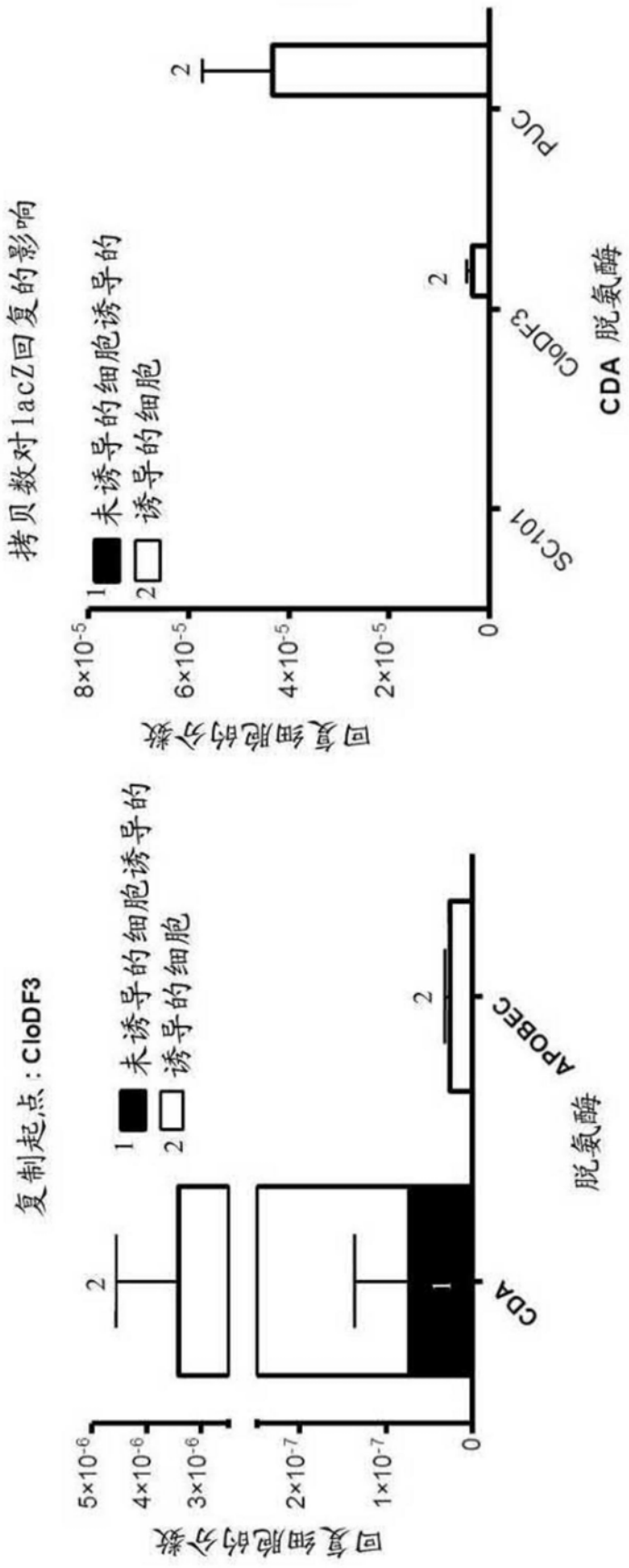

拷贝数对lacZ回复的影响
1 未诱导的细胞诱导的
2 诱导的细胞
回复细胞的分数
8×10-5
6×10-5
4×10-5
2×10-5
0
SC101
CloDF3
PUC
CDA 脱氨酶
复制起点：ClOoDF3
1 未诱导的细胞诱导的
2 诱导的细胞
回复细胞的分数
5×10-6
4×10-6
3×10-6
2×10-7
1×10-7
0
CDA
APOBEC
脱氨酶

图49续

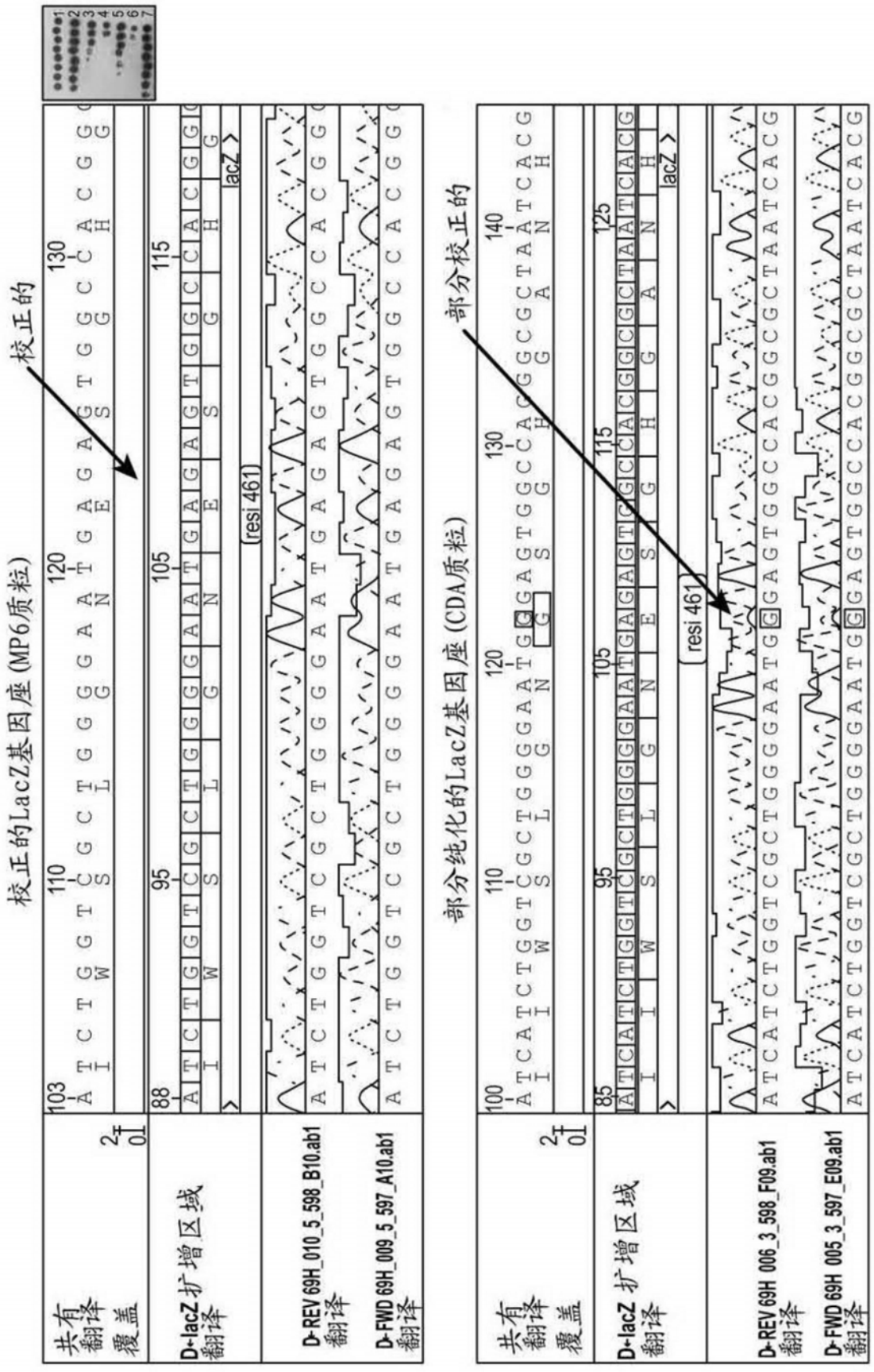

校正的LacZ基因座 (MP6质粒)
校正的
共有翻译覆盖
D-lacZ 扩增区域翻译
D-REV 69H_010_5_598_B10.ab1翻译
D-FWD 69H_009_5_597_A10.ab1翻译
(resi 461)
lacZ
部分纯化的LacZ基因座 (CDA质粒)
部分校正的
共有翻译覆盖
D-lacZ 扩增区域翻译
D-REV 69H 006_3_598_F09.ab1翻译
D-FWD 69H 005_3_597_E09.ab1翻译
(resi 461)
lacZ

图50

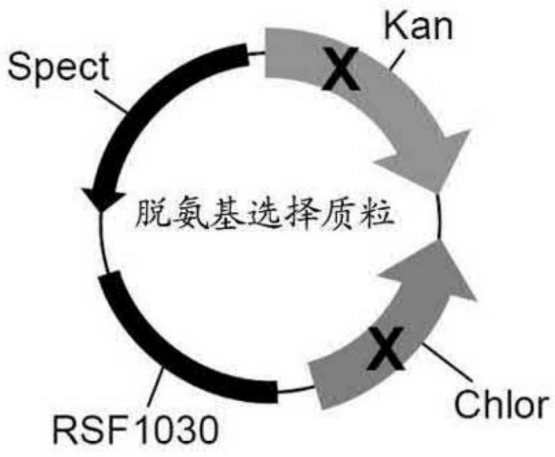
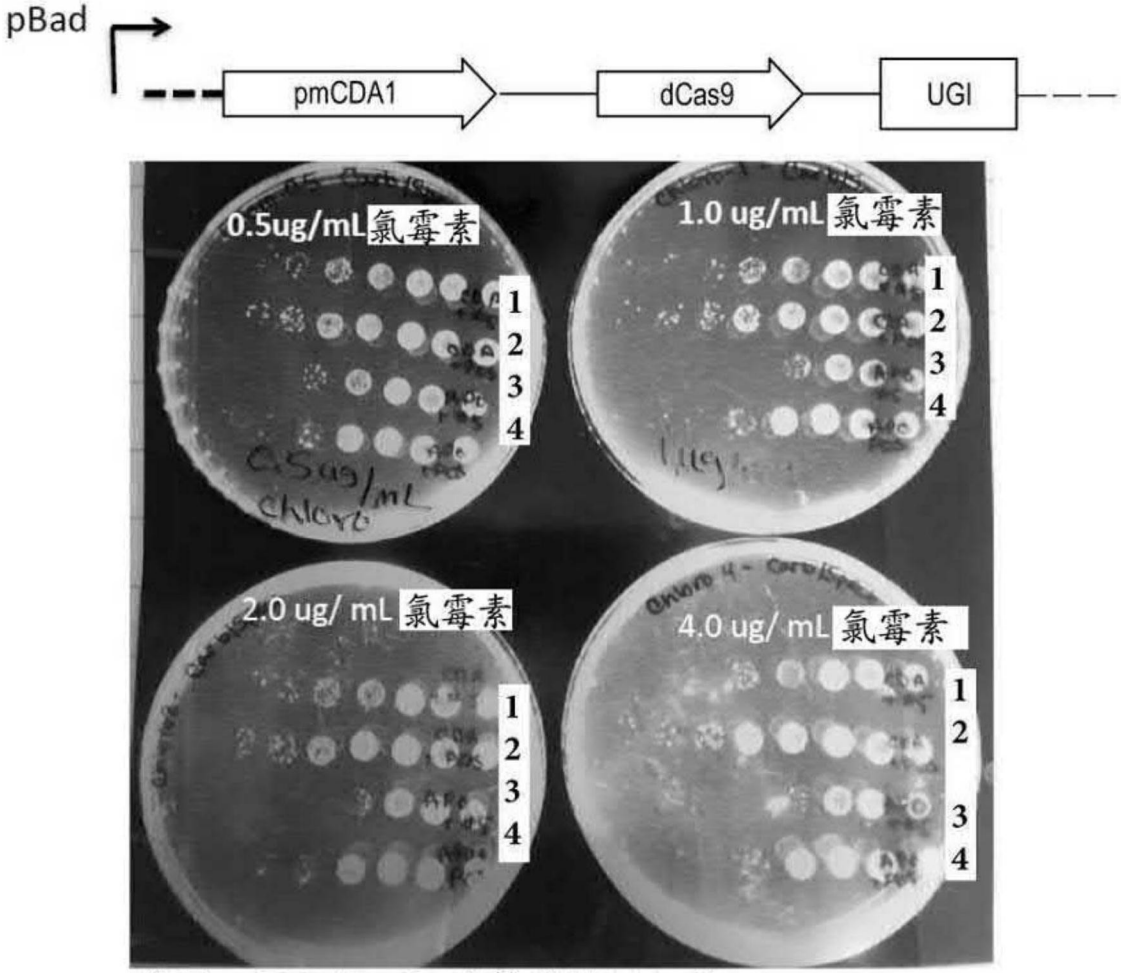


图51



第1行: CDA-dCas9 + 选择质粒 (chlor^S)

第2行: CDA-dCas9 + 阳性对照选择 (chlor^R)

第3行: rAPOBEC-dCas9 + 选择质粒 (chlor^S)

第4行: rAPOBEC-dCas9 + 阳性对照选择 (chlor^R)

图52

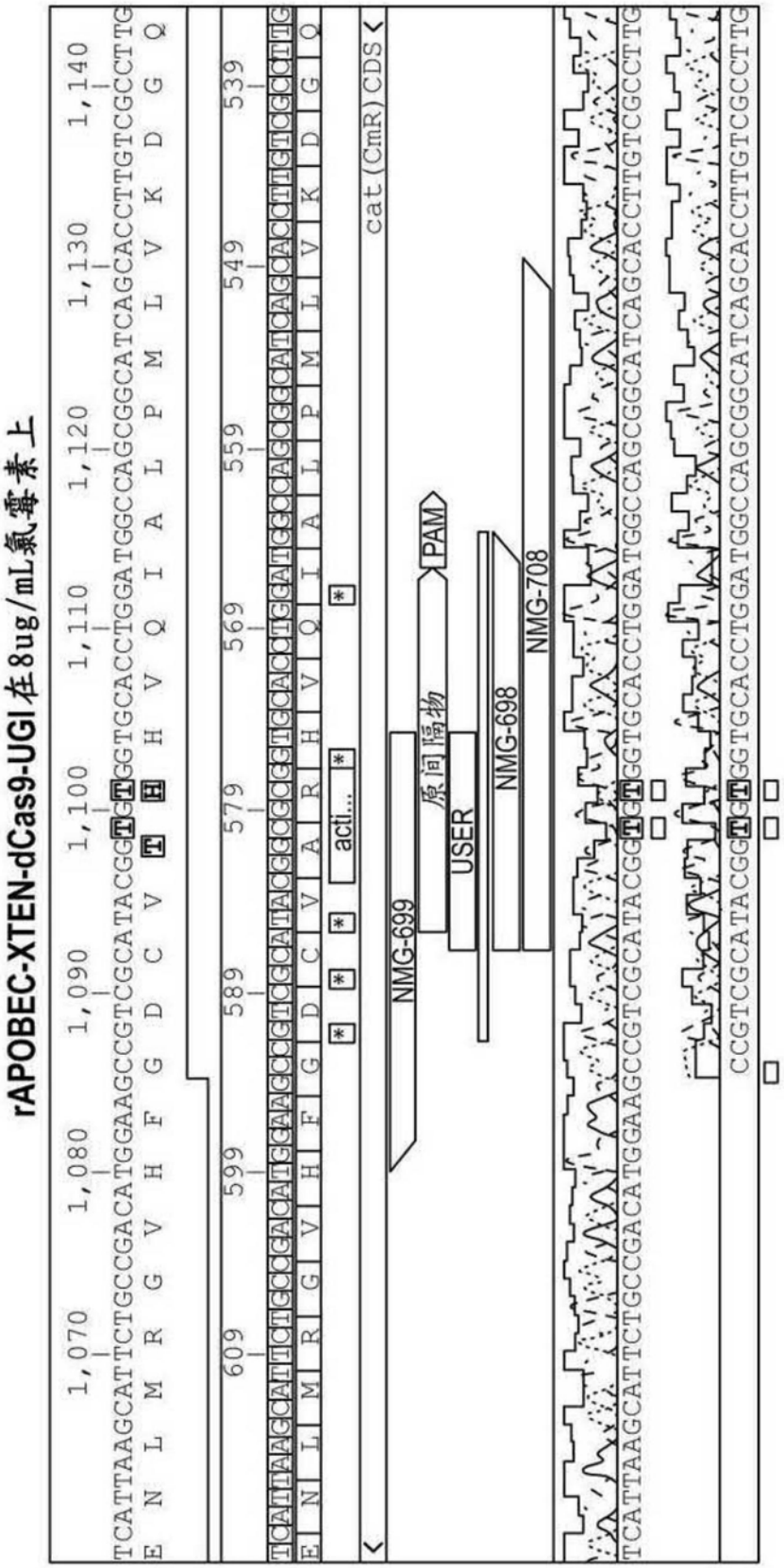

rAPOBEC-XTEN-dCas9-UGI在8ug/mL氯霉素上
NMG-699
原间隔物
PAM
USER
NMG-698
NMG-708
cat(CmR)CDS

图53A

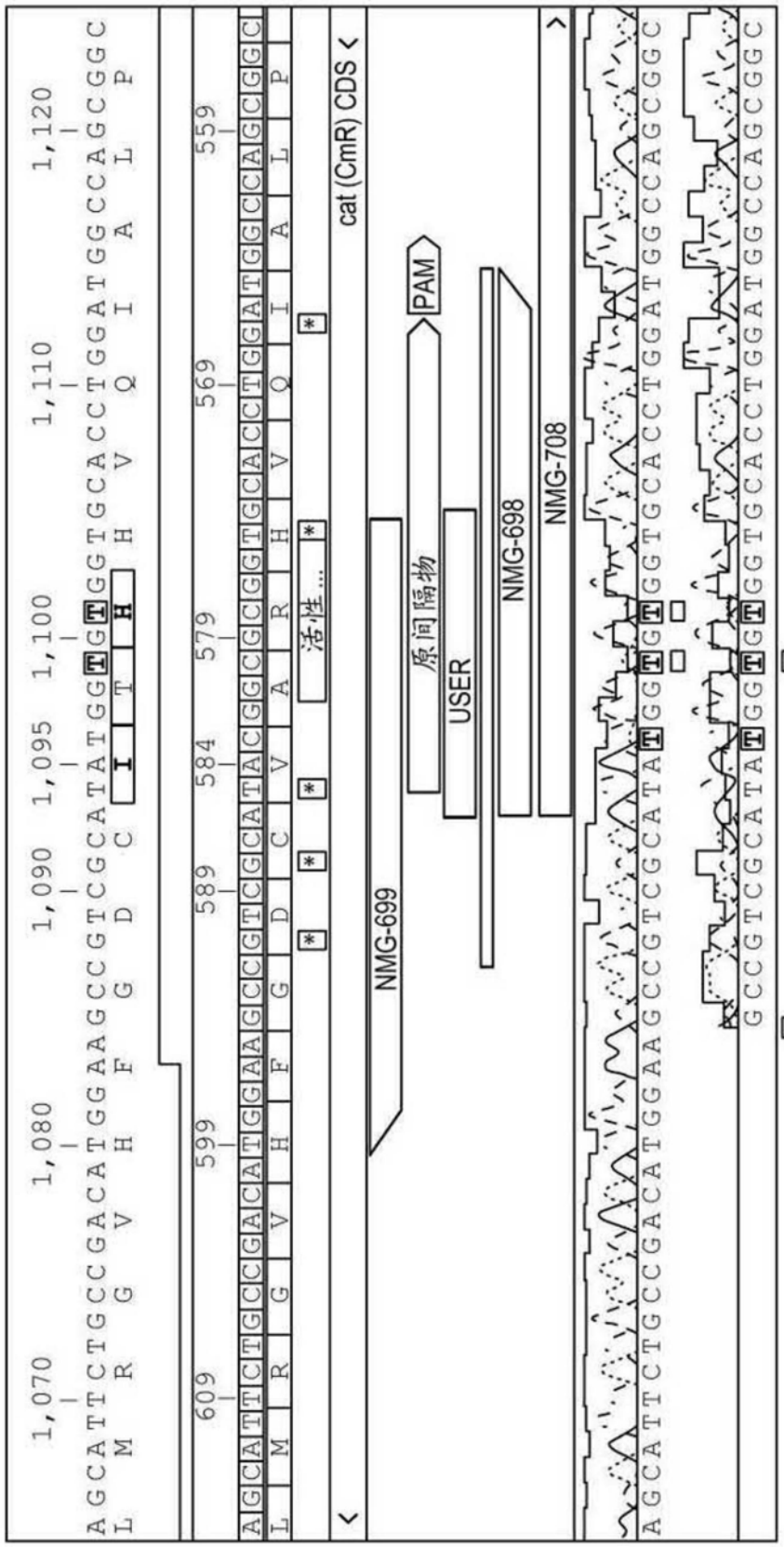

CDA-XTEN-dCas9-UGI: 在8ug/mL氯霉素上存活
cat (CmR) CDS
NMG-699
原间隔物
PAM
USER
NMG-698
NMG-708

图53B

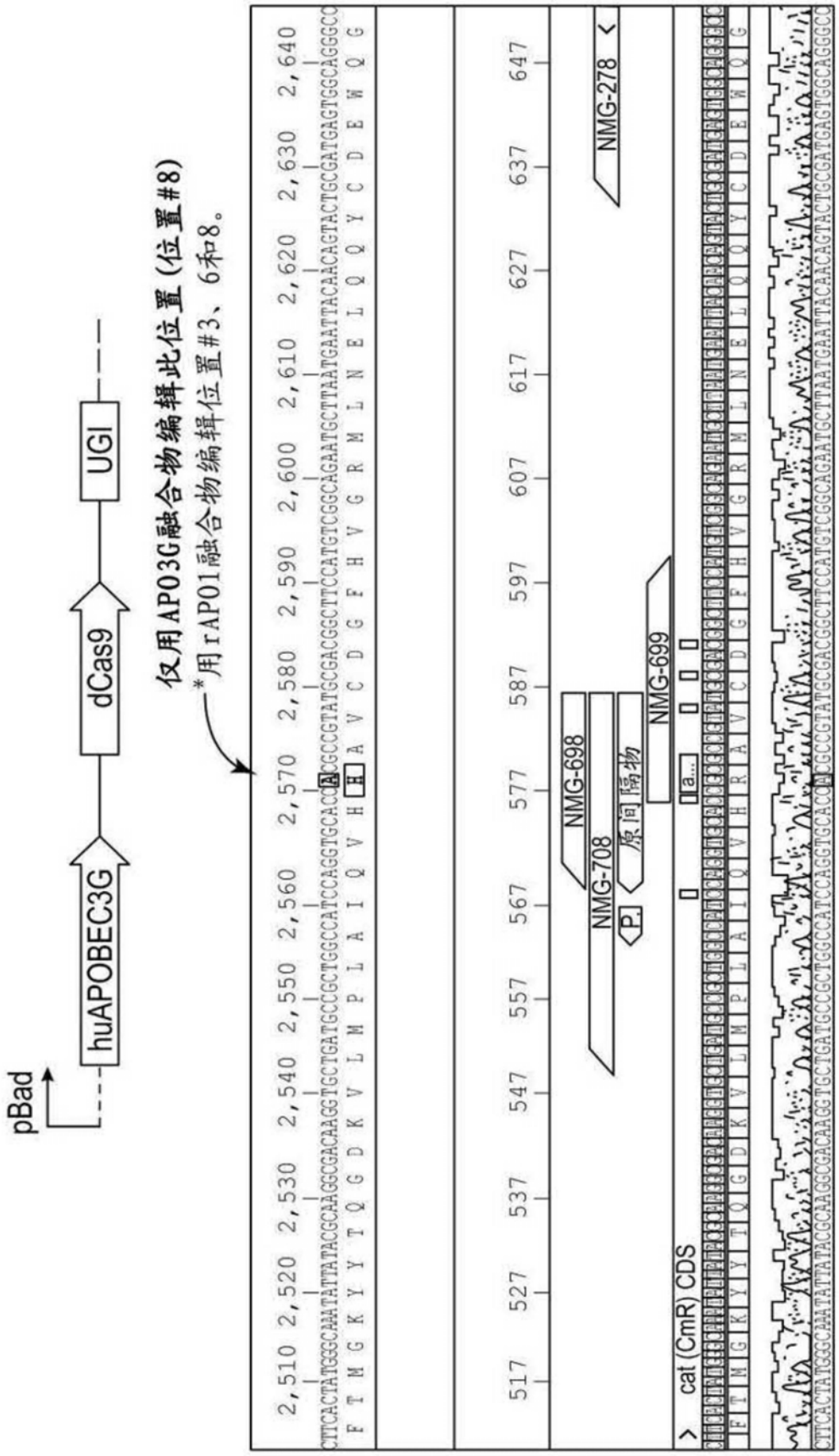

pBad
huAPOBEC3G
dCas9
UGI
仅用AP03G融合物编辑此位置(位置#8)
*用rAP01融合物编辑位置#3、6和8。
NMG-278
NMG-699
NMG-698
NMG-708
原间隔物
cat (CmR) CDS

图54

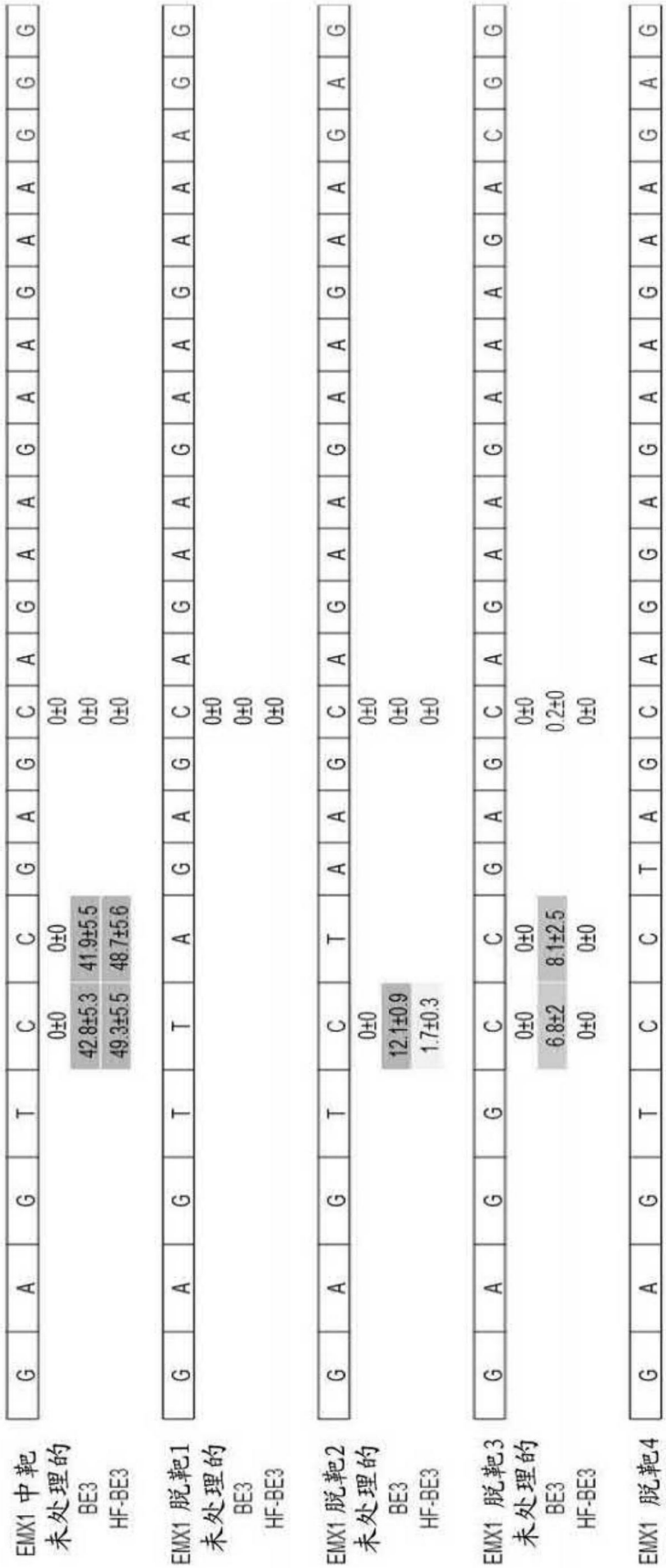

G A G T C C G A G C A G A A G A A A G G G
EMX1 中靶未处理的 0±0 0±0 0±0
BE3 42.8±5.3 41.9±5.5 0±0
HF-BE3 49.3±5.5 48.7±5.6 0±0
G A G T T A G A G C A G A A G A A A G G
EMX1 脱靶1未处理的 0±0
BE3 0±0
HF-BE3 0±0
G A G T C T A A G C A G A A G A A A G A
EMX1 脱靶2未处理的 0±0 0±0
BE3 12.1±0.9 0±0
HF-BE3 1.7±0.3 0±0
G A G G C C G A G C A G A A G A A C G G
EMX1 脱靶3未处理的 0±0 0±0 0±0
BE3 6.8±2 8.1±2.5 0.2±0
HF-BE3 0±0 0±0 0±0
G A G T C C T A G C A G G A A A G A G A
EMX1 脱靶4未处理的 0±0 0±0 0±0
BE3 1.3±0.1 1.3±0.1 0±0
HF-BE3 0.1±0 0.2±0 0±0

图55

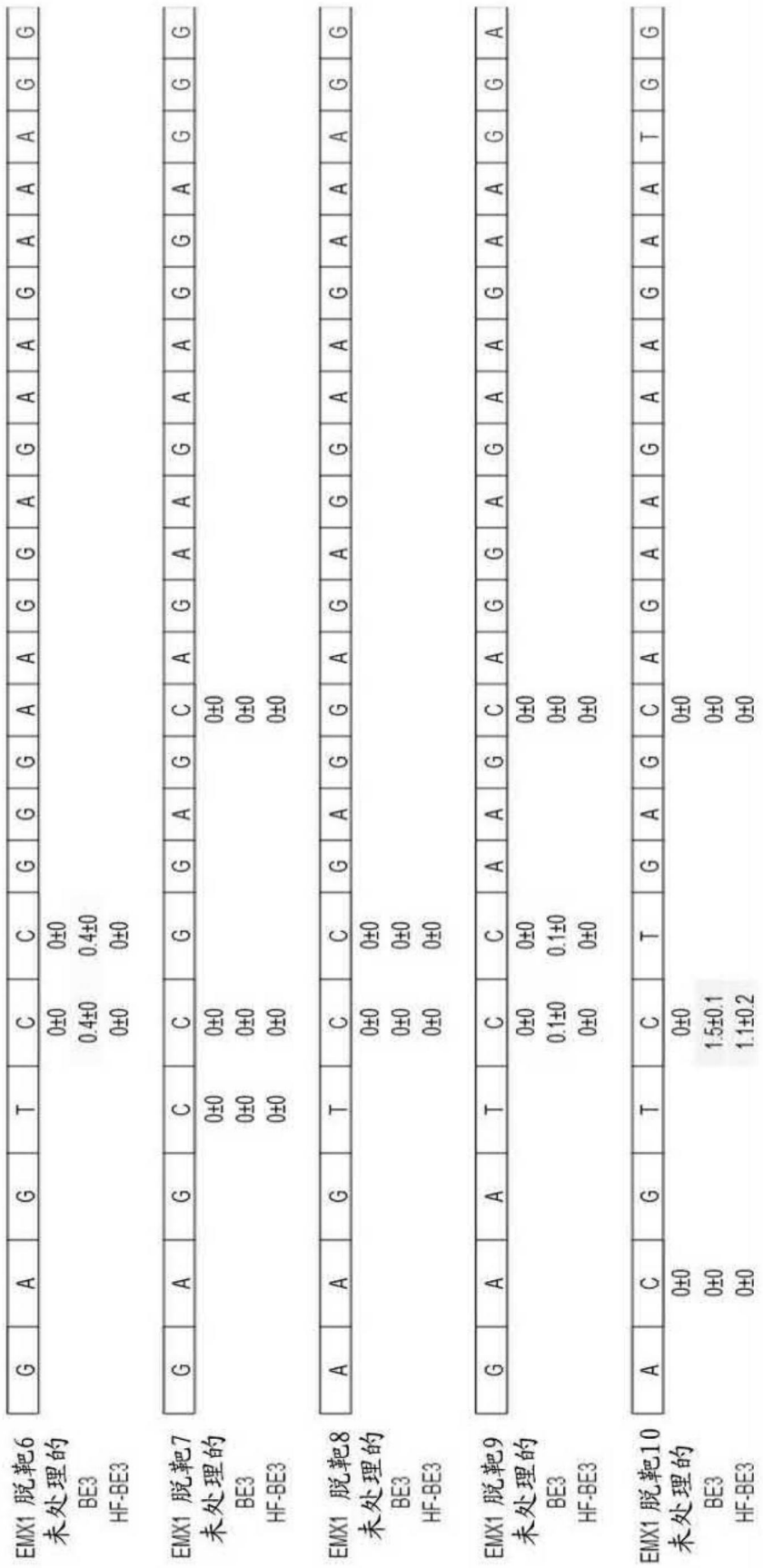

G A G T C C G G G A C A G G A G G A A G G A A A G G
EMX1 脱靶6
未处理的 0±0 0±0
BE3 0.4±0 0.4±0
HF-BE3 0±0 0±0
G A G C C G G A G C A G G A A G G A A G G G
EMX1 脱靶7
未处理的 0±0 0±0 0±0
BE3 0±0 0±0 0±0
HF-BE3 0±0 0±0 0±0
A A G T C C G A G G A G G A A G G A A A G G
EMX1 脱靶8
未处理的 0±0 0±0
BE3 0±0 0±0
HF-BE3 0±0 0±0
G A A T C C A G C A G G A A G G A A G G A
EMX1 脱靶9
未处理的 0±0 0±0 0±0
BE3 0.1±0 0.1±0 0±0
HF-BE3 0±0 0±0 0±0
A C G T C T G A G C A G G A A G G A A T G G
EMX1 脱靶10
未处理的 0±0 0±0 0±0
BE3 0±0 1.5±0.1 0±0
HF-BE3 0±0 1.1±0.2 0±0

图55续

EMX1		C_3	C_6	C_{12}	插入/缺失%
未处理的	A	0±0	0±0	0±0	0±0
	C	99.9±0	99.9±0	99.9±0	
	G	0±0	0±0	0±0	
	T	0±0	0±0	0±0	
BE3	A	1.3±0.1	0.7±0.2	0±0	2.6±0.3
	C	51.7±1	55.2±0.1	99.9±0	
	G	4±0.7	1.9±0.4	0±0	
	T	42.8±0.2	41.9±0.4	0±0	
HF BE3	A	1.5±0.7	0.5±0.1	0±0	1±0.1
	C	44.8±8.5	49.9±6.6	99.9±0	
	G	4.2±1.2	0.6±0.2	0±0	
	T	49.3±6.5	48.7±6.3	0±0	

FANCF		C_6	C_7	C_8	C_{11}	插入/缺失%
未处理的	A	0±0	0±0	0±0	0±0	0±0
	C	99.9±0	99.9±0	99.9±0	99.9±0	
	G	0±0	0±0	0±0	0±0	
	T	0±0	0±0	0±0	0±0	
BE3	A	0.8±0.1	1.2±0.3	1.2±0.1	0.2±0.1	5.8±0.9
	C	64.7±5.9	67.9±5.1	68.9±5.2	85.4±2.7	
	G	0.6±0	0.7±0	0.7±0.1	0.1±0	
	T	33.8±5.7	29.9±4.8	29±5	14.1±2.6	
HF BE3	A	0.9±0.2	2±0.4	0.9±0.2	0.4±0	5.9±0.7
	C	52.5±8.4	57.7±7.6	61.9±6.8	84.7±2.4	
	G	1.4±0.2	0.6±0	0.5±0.1	0.3±0	
	T	45±8	39.5±7.2	36.5±6.5	14.4±2.2	

RNF2		C_3	C_6	C_{12}	插入/缺失%
未处理的	A	0±0	0±0	0±0	0±0
	C	99.9±0	99.9±0	99.9±0	
	G	0±0	0±0	0±0	
	T	0±0	0±0	0±0	
BE3	A	0.4±0	1.5±0.2	0.1±0	2.3±0.3
	C	70.9±2.7	45±0	95±0.2	
	G	0±0	12.6±2.4	0.1±0	
	T	28.5±2.7	40.7±2.7	4.6±0.2	
HF BE3	A	0±0	0.2±0	0±0	0.5±0.1
	C	76.8±1.8	64.5±3.3	95.4±0.1	
	G	0±0	2.4±0.5	0±0	
	T	23±1.8	32.7±2.6	4.5±0.1	

HEK3		C_6	C_7	C_8	C_{11}	插入/缺失%
未处理的	A	0±0	0±0	0±0	0±0	0±0
	C	99.9±0	99.9±0	99.9±0	99.9±0	
	G	0±0	0±0	0±0	0±0	
	T	0±0	0±0	0±0	0±0	
BE3	A	0±0	1.6±0.2	1.5±0.3	0.2±0	2.7±0.4
	C	98.7±0.2	48.6±5.4	40.8±6.7	98.6±0.1	
	G	0±0	1.4±0.2	10.5±1.2	0.3±0	
	T	1.1±0.2	48.2±4.9	47±5.2	0.7±0.1	
HF BE3	A	0±0	0.8±0.2	1.1±0.3	0.4±0	4.1±0.7
	C	97.8±0.4	65±3.8	40.4±6.3	95.9±0.5	
	G	0±0	0.9±0.2	6.1±1.1	0.7±0.2	
	T	2±0.4	33±3.4	52.2±5.2	2.7±0.2	

HEK4		C_6	C_7	C_8	C_{11}	插入/缺失%
未处理的	A	0±0	0±0	0±0	0±0	0±0
	C	99.9±0	99.9±0	99.9±0	99.9±0	
	G	0±0	0±0	0±0	0±0	
	T	0±0	0±0	0±0	0±0	
BE3	A	0±0	5.4±1.3	0±0	0±0	3±0.7
	C	98.7±0	44.5±8.2	98.9±0	99.8±0	
	G	0±0	18.1±2.3	0±0	0±0	
	T	1.1±0.1	31.9±4.8	0.9±0.1	0.1±0	
HF BE3	A	0±0	5.2±0.8	0±0	0±0	1.3±0
	C	99.6±0	36.6±9	98.5±0.1	99.9±0	
	G	0±0	16±2.8	0±0	0±0	
	T	0.2±0	42.1±5.3	1.3±0.1	0±0	

图56

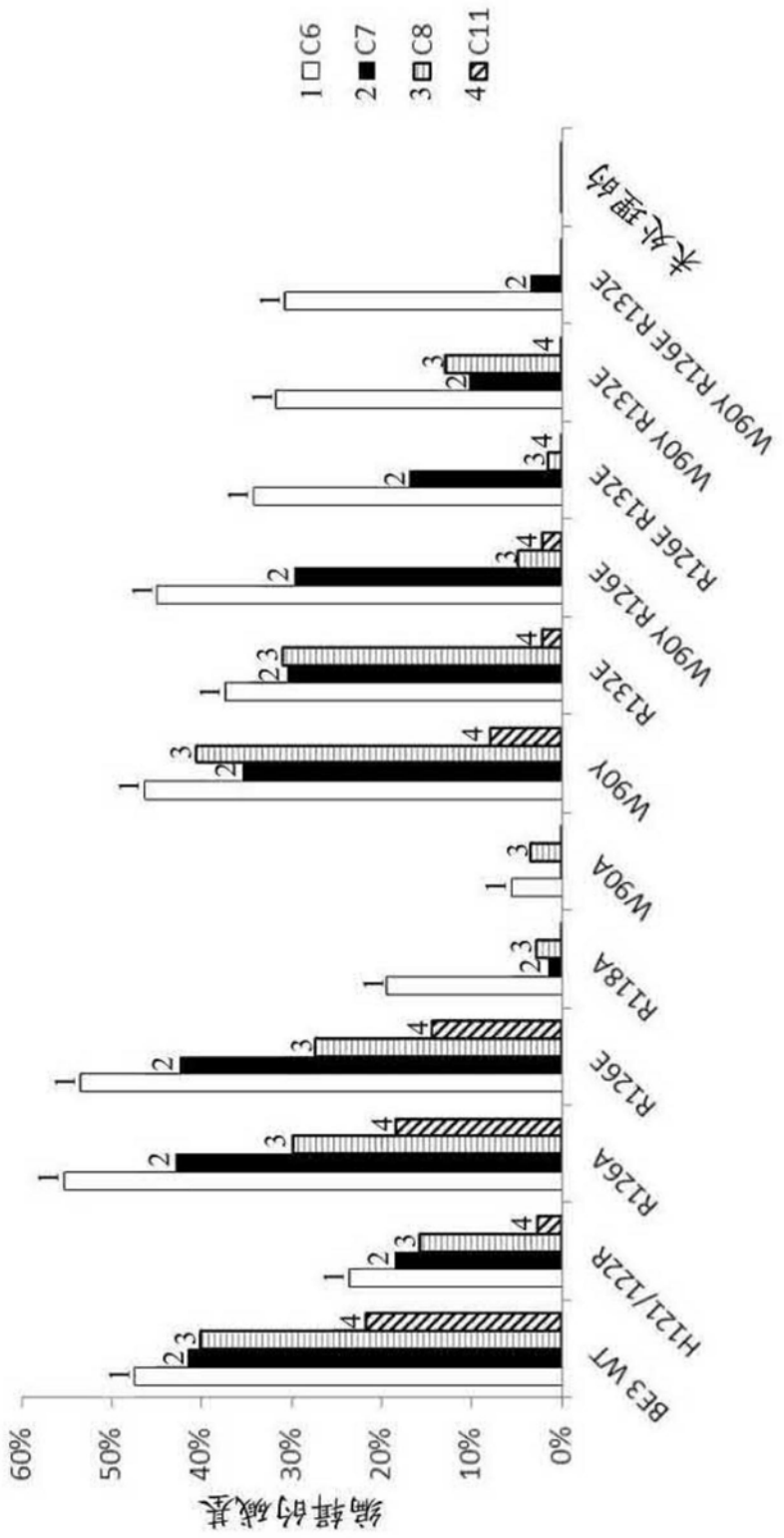

1 C6
2 C7
3 C8
4 C11
未处理的
W90Y R126E R132E
W90Y R132E
R126E R132E
W90Y R126E
R132E
W90Y
W90A
R118A
R126E
R126A
H121/122R
BE3 WT
60%
50%
40%
30%
20%
10%
0%
编辑的碱基

图57

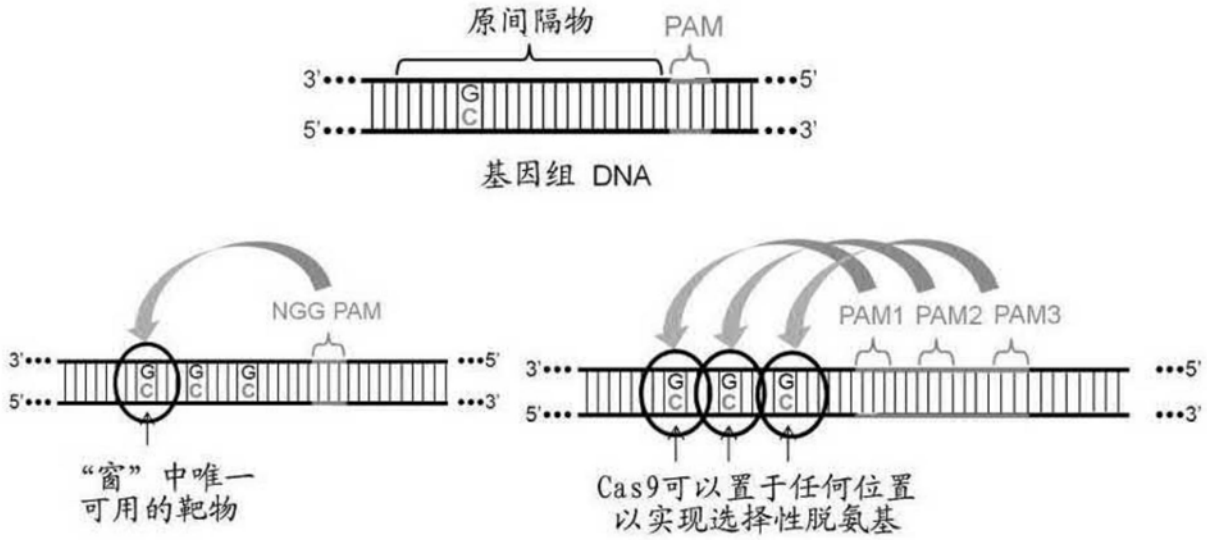

原间隔物
PAM
3'
5'
5'
3'
基因组 DNA
NGG PAM
"窗"中唯一
可用的靶物
PAM1 PAM2 PAM3
Cas9可以置于任何位置
以实现选择性脱氨基

图58

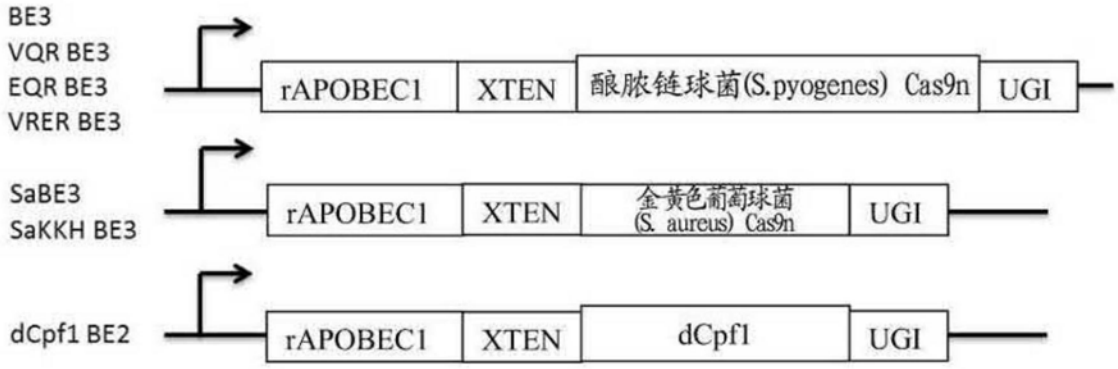

BE3
VQR BE3
EQR BE3
VRER BE3
rAPOBEC1
XTEN
酿脓链球菌(S.pyogenes) Cas9n
UGI
SaBE3
SaKKH BE3
rAPOBEC1
XTEN
金黄色葡萄球菌
(S. aureus) Cas9n
UGI
dCpf1 BE2
rAPOBEC1
XTEN
dCpf1
UGI

图59

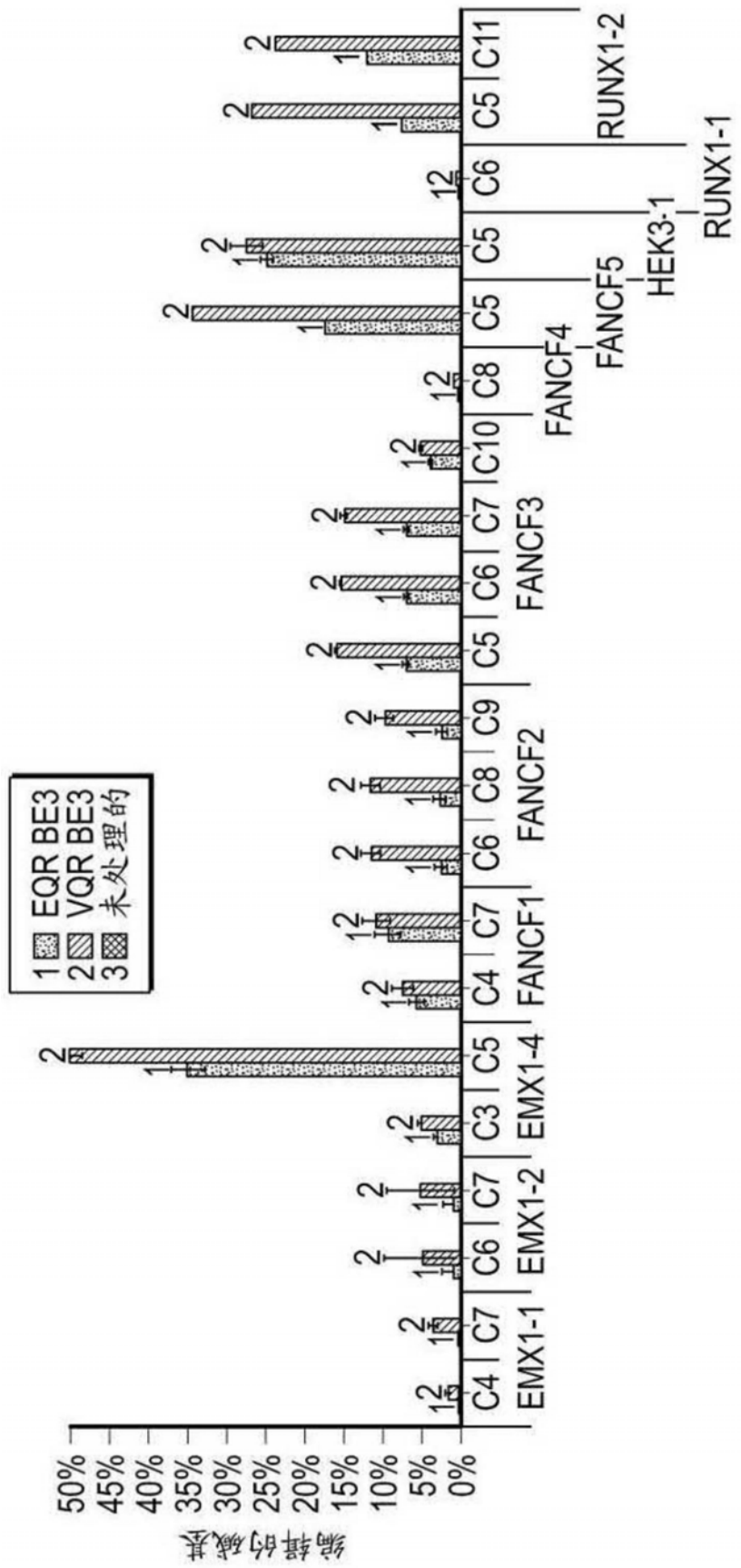

1 EQR BE3
2 VQR BE3
3 未处理的
编辑的碱基
50%
45%
40%
35%
30%
25%
20%
15%
10%
5%
0%
EMX1-1
EMX1-2
EMX1-4
FANCF1
FANCF2
FANCF3
FANCF4
FANCF5
HEK3-1
RUNX1-1
RUNX1-2
C4
C7
C6
C7
C3
C5
C4
C7
C6
C8
C9
C5
C6
C7
C10
C8
C5
C5
C6
C5
C11

图60

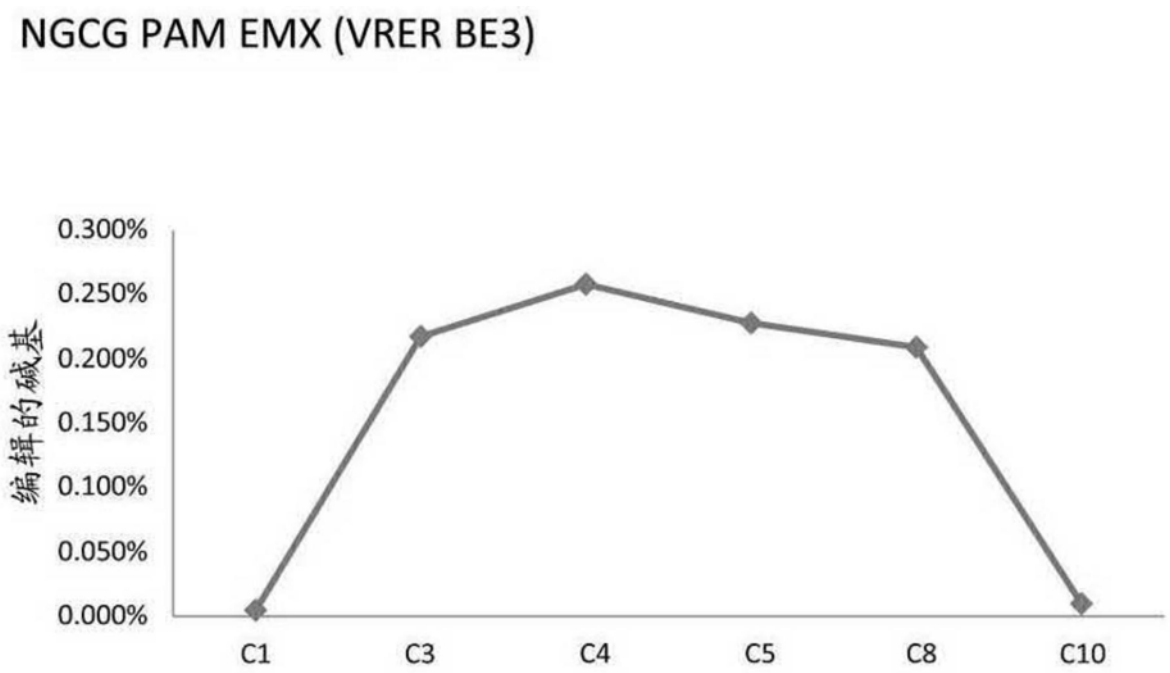

NGCG PAM EMX (VRER BE3)
编辑的碱基
0.300%
0.250%
0.200%
0.150%
0.100%
0.050%
0.000%
C1
C3
C4
C5
C8
C10

图61

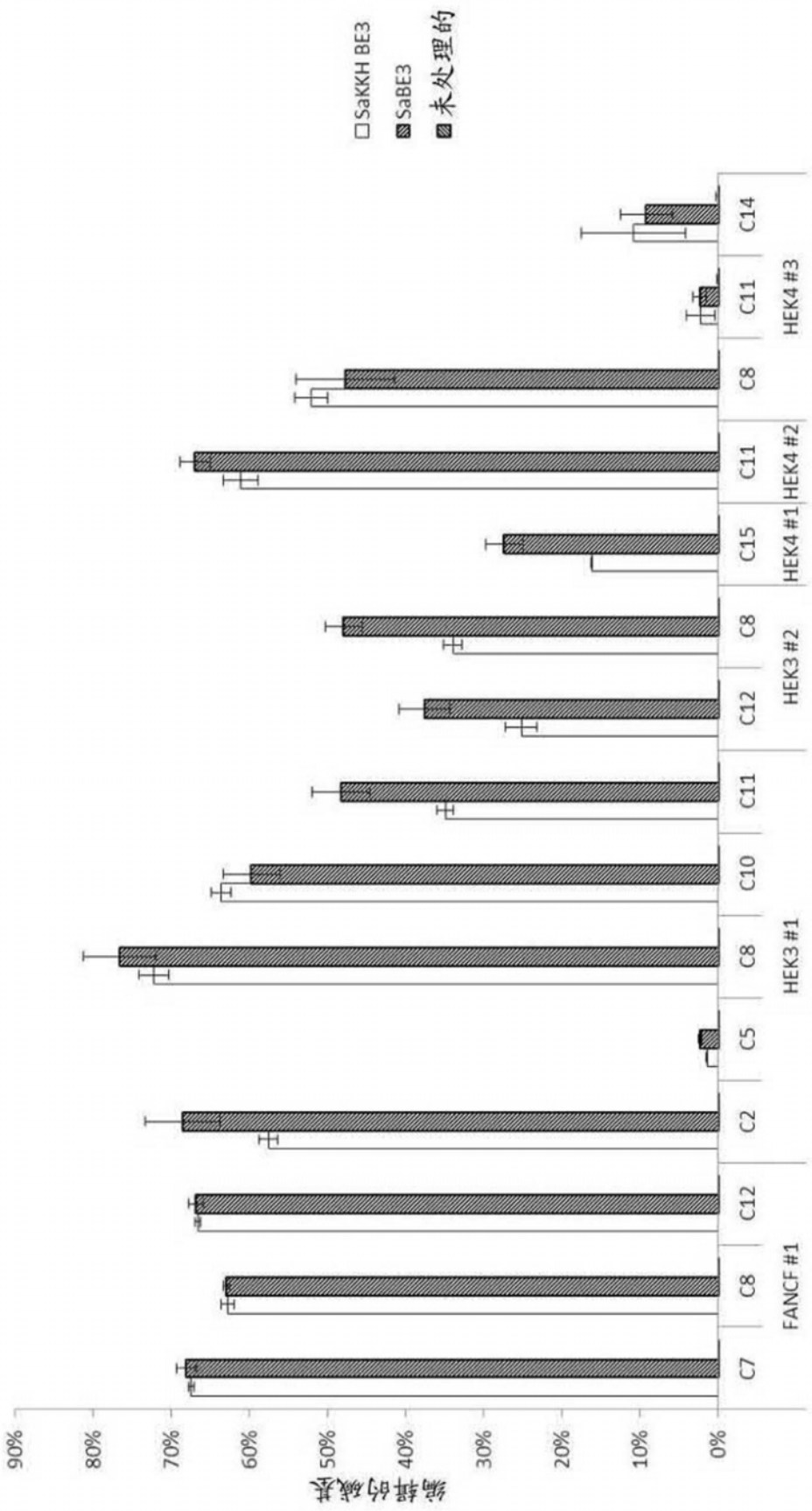

SaKKH BE3
SaBE3
未处理的
编辑的碱基
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
FANCF #1
C7
C8
C12
HEK3 #1
C2
C5
C8
C10
C11
HEK3 #2
C12
C8
HEK4 #1
C15
HEK4 #2
C11
C8
HEK4 #3
C11
C14

图62

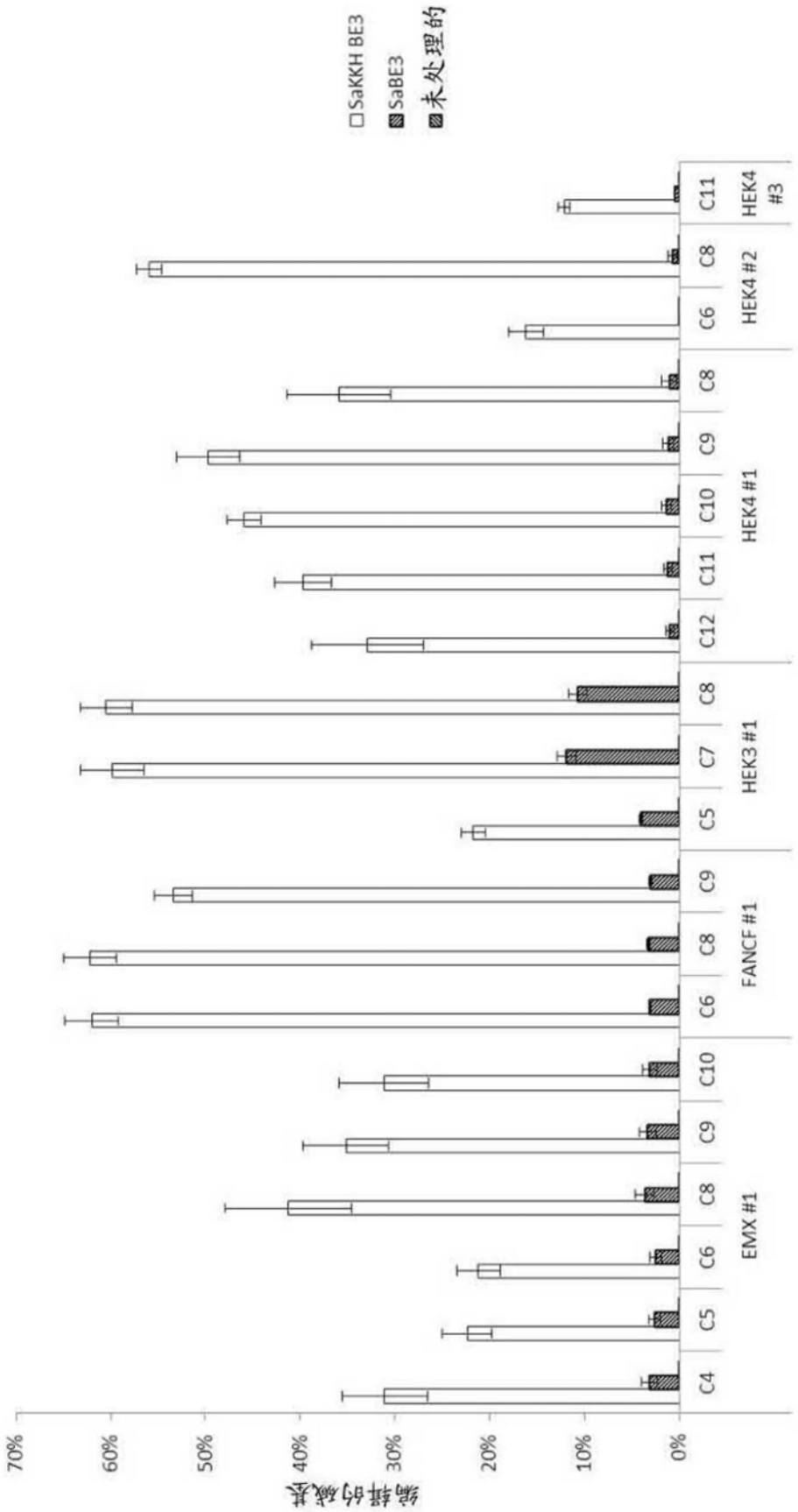

SaKKH BE3
SaBE3
未处理的
编辑的碱基
70%
60%
50%
40%
30%
20%
10%
0%
C4
C5
C6
C8
C9
C10
EMX #1
C6
C8
C9
FANCF #1
C5
C7
C8
HEK3 #1
C12
C11
C10
C9
C8
HEK4 #1
C6
C8
HEK4 #2
C11
HEK4 #3

图63

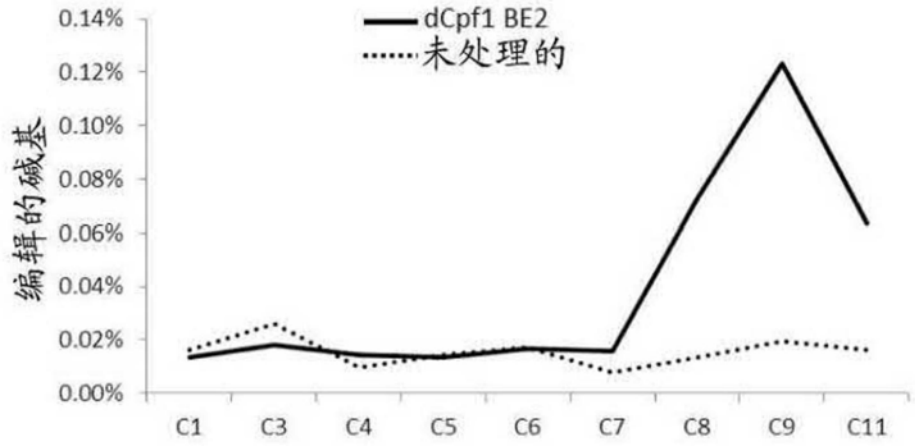


图64A

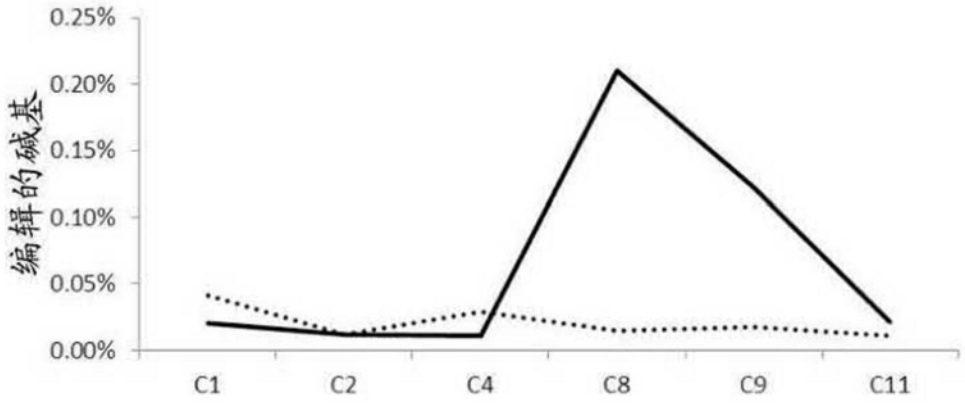


图64B

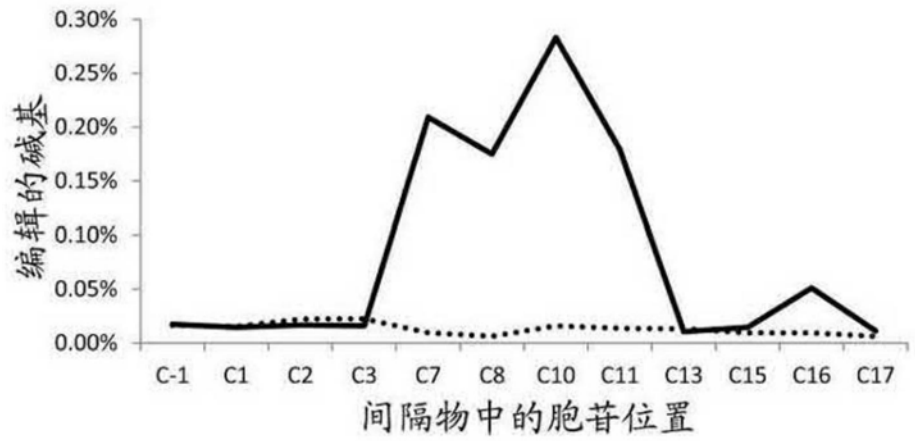


图64C

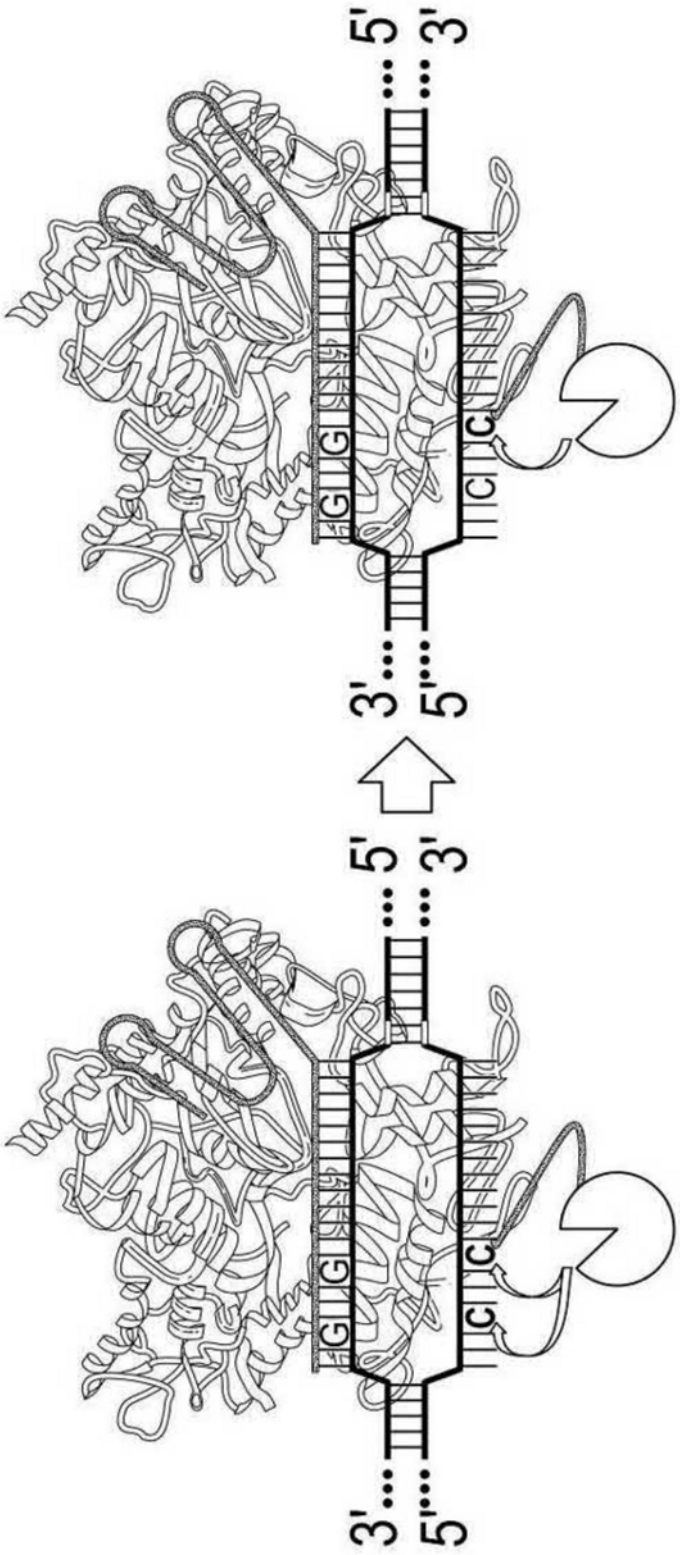


图65

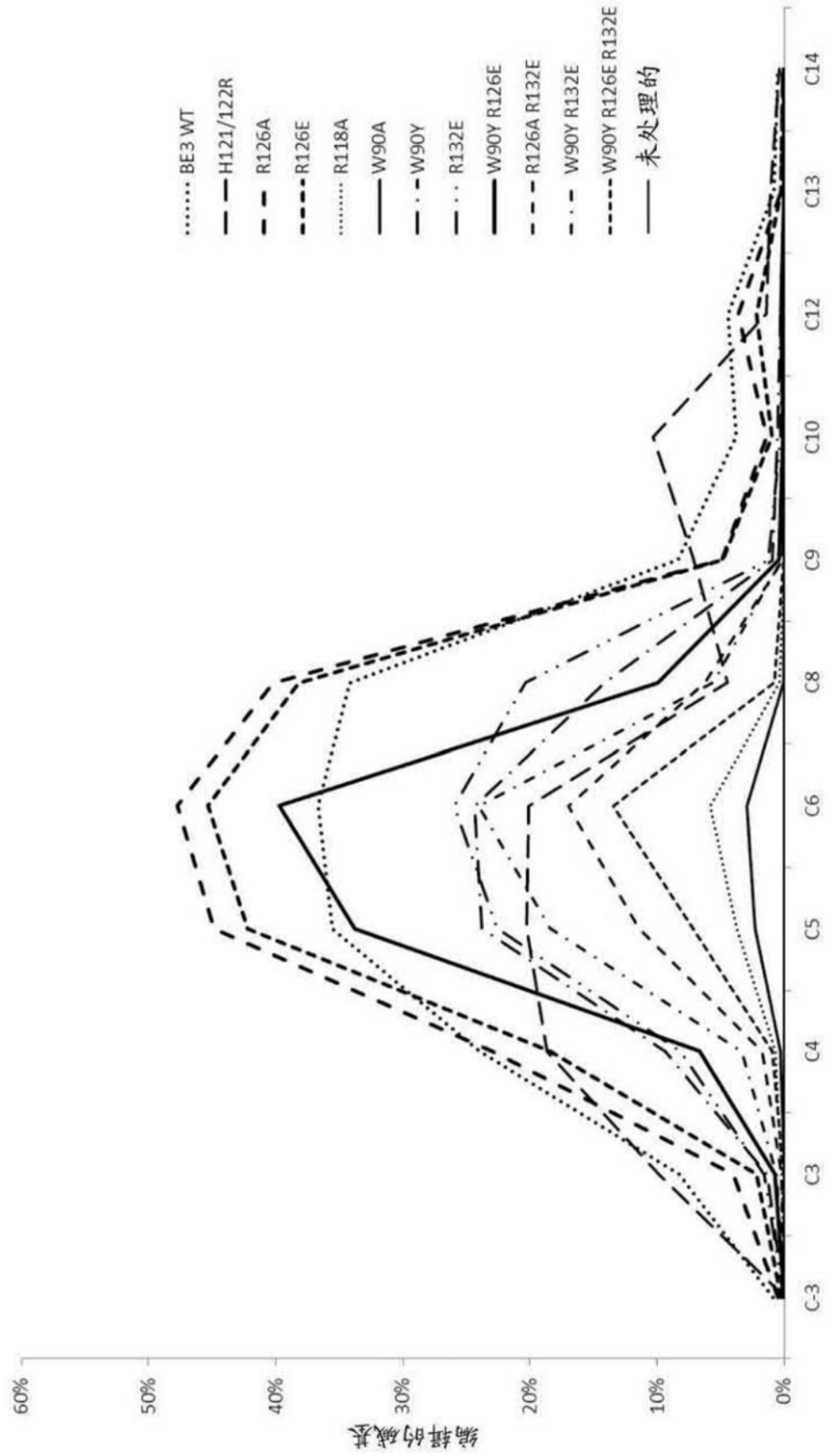

BE3 WT
H121/122R
R126A
R126E
R118A
W90A
W90Y
R132E
W90Y R126E
R126A R132E
W90Y R132E
W90Y R126E R132E
未处理的
C-3
C3
C4
C5
C6
C8
C9
C10
C12
C13
C14
60%
50%
40%
30%
20%
10%
0%
编辑的碱基

图66

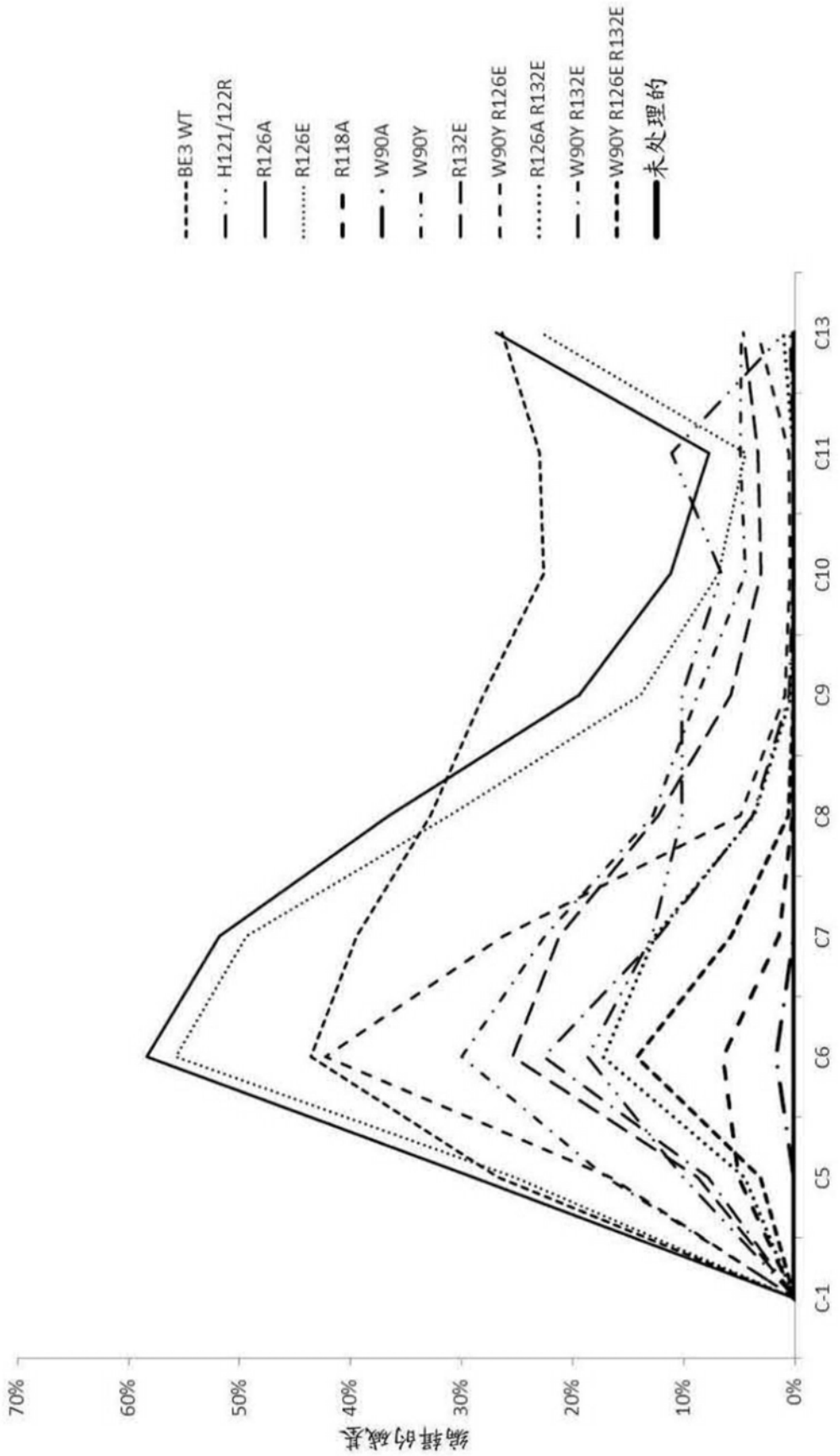

BE3 WT
H121/122R
R126A
R126E
R118A
W90A
W90Y
R132E
W90Y R126E
R126A R132E
W90Y R132E
W90Y R126E R132E
未处理的
编辑的碱基
70%
60%
50%
40%
30%
20%
10%
0%
C-1
C5
C6
C7
C8
C9
C10
C11
C13

图67

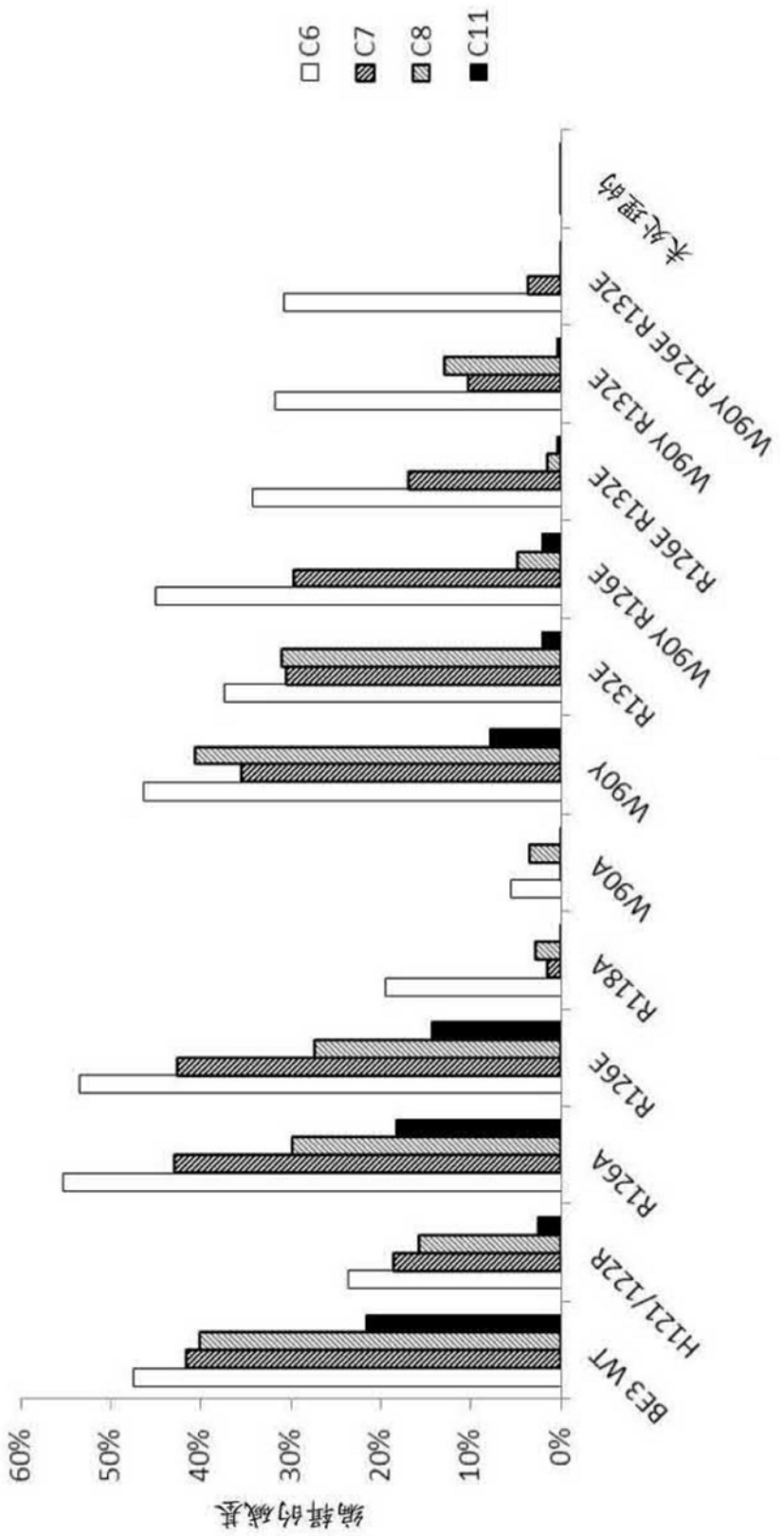

C6
C7
C8
C11
未处理的
W90Y R126E R132E
W90Y R132E
R126E R132E
W90Y R126E
R132E
W90Y
W90A
R118A
R126E
R126A
H121/122R
BE3 WT
60%
50%
40%
30%
20%
10%
0%
编辑的碱基

图68

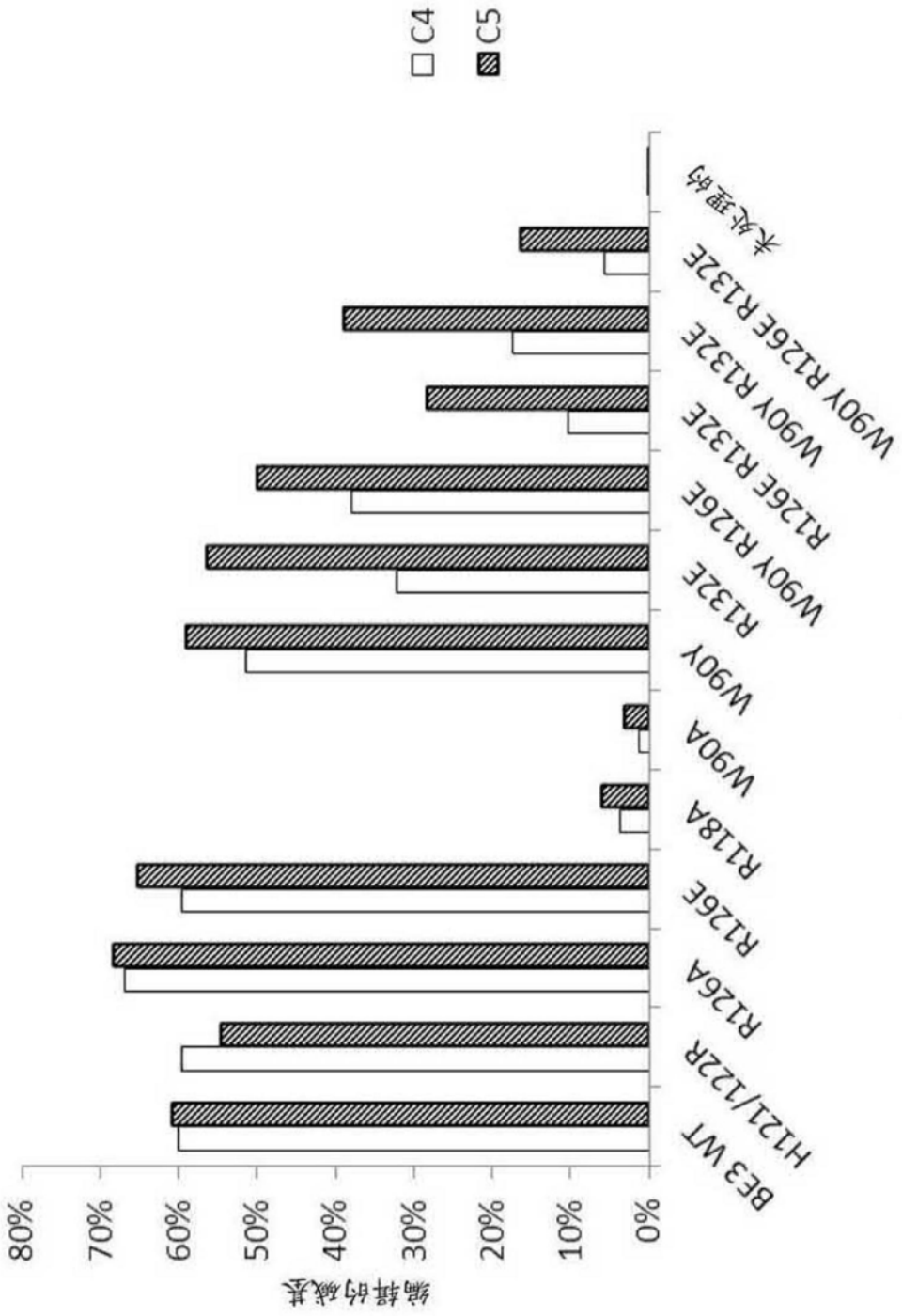

C4
C5
未处理的
W90Y R126E R132E
W90Y R132E
R126E R132E
W90Y R126E
R132E
W90Y
W90A
R118A
R126E
R126A
H121/122R
BE3 WT
80%
70%
60%
50%
40%
30%
20%
10%
0%
编辑的碱基

图69

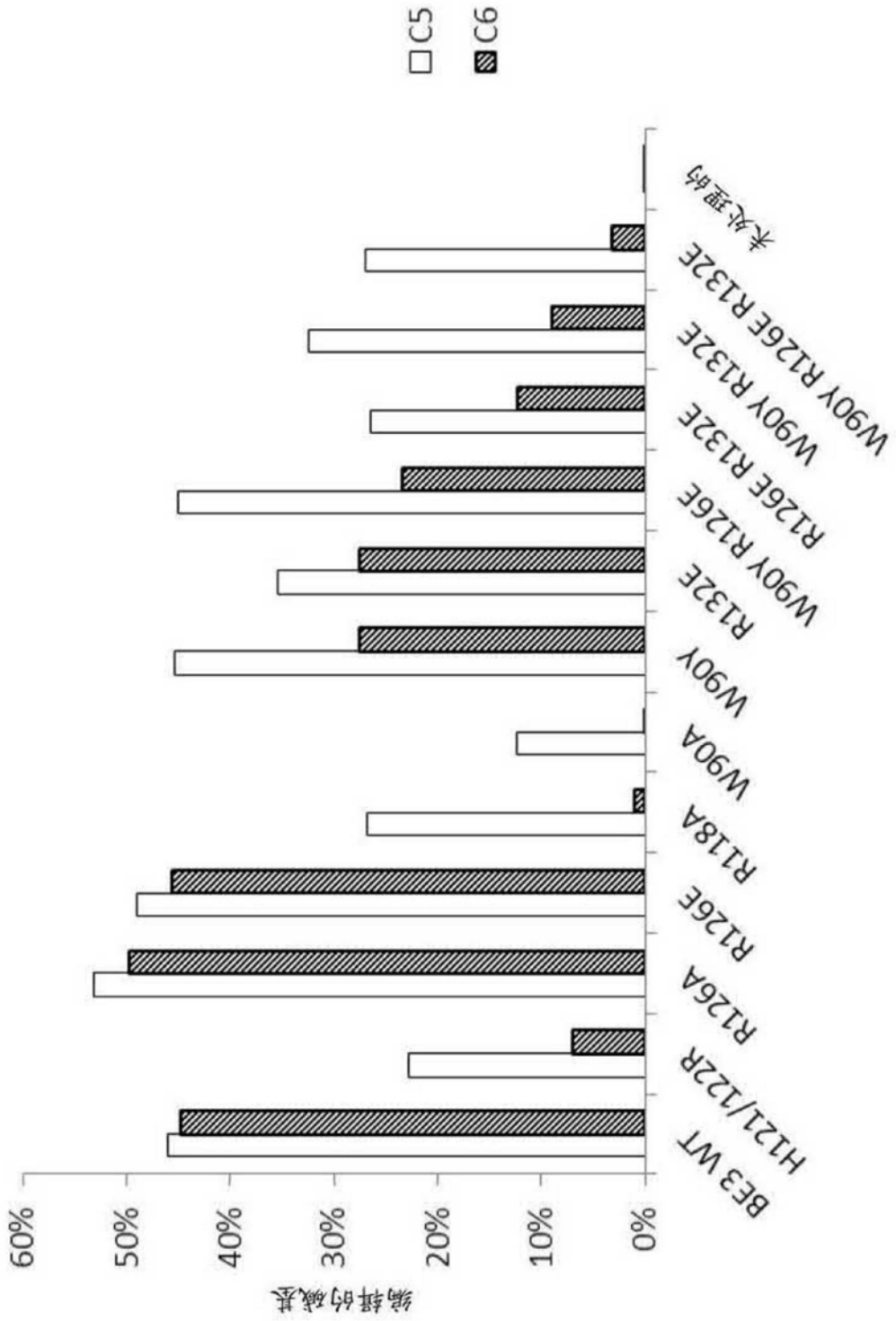

C5
C6
未处理的
W90V R126E R132E
W90V R132E
R126E R132E
W90V R126E
R132E
W90V
W90A
R118A
R126E
R126A
H121/122R
BE3 WT
60%
50%
40%
30%
20%
10%
0%
编辑的碱基

图70

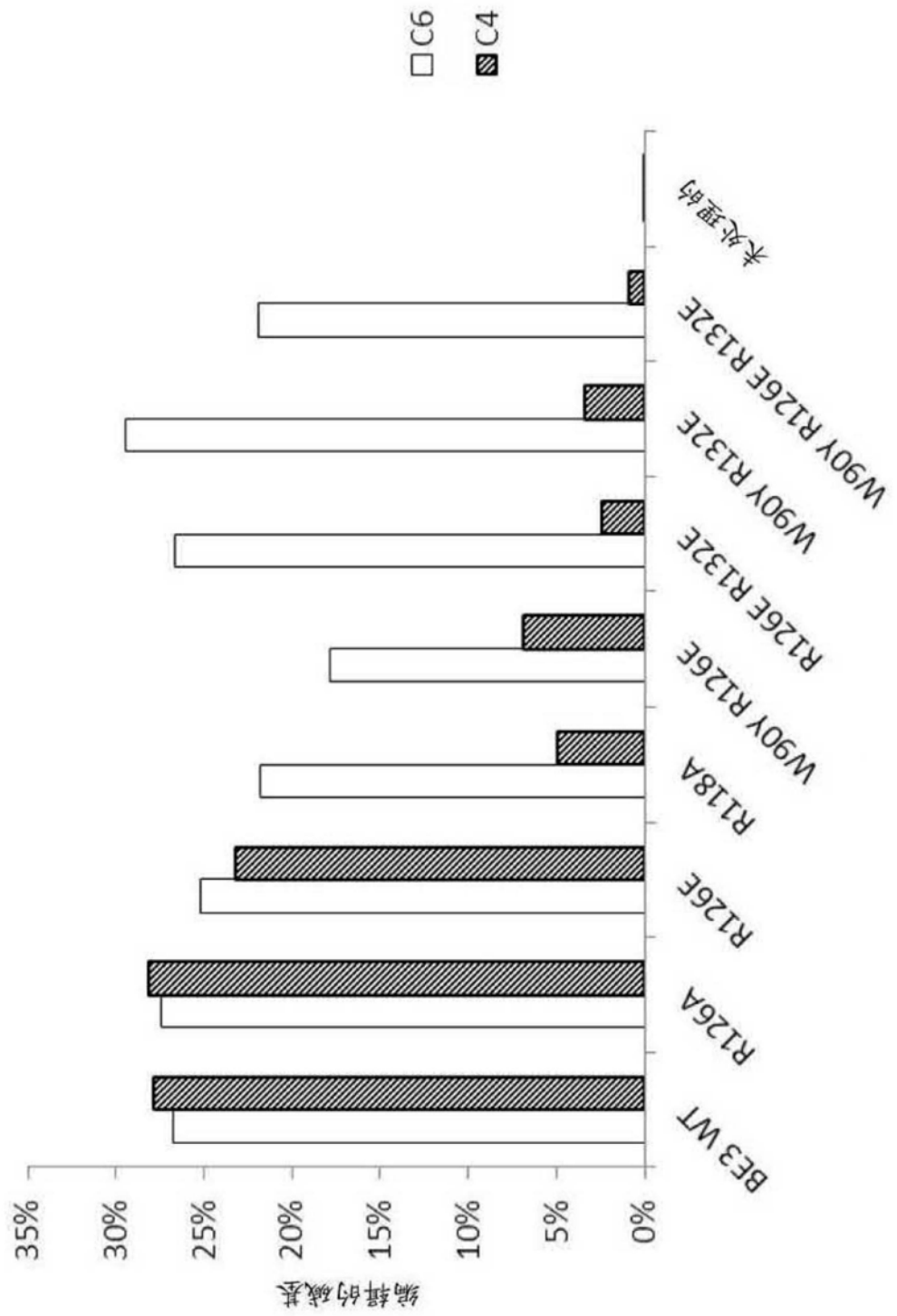

C6
C4
未处理的
W90Y R126E R132E
W90Y R132E
R126E R132E
W90Y R126E
R118A
R126E
R126A
BE3 WT
35%
30%
25%
20%
15%
10%
5%
0%
编辑的碱基

图71

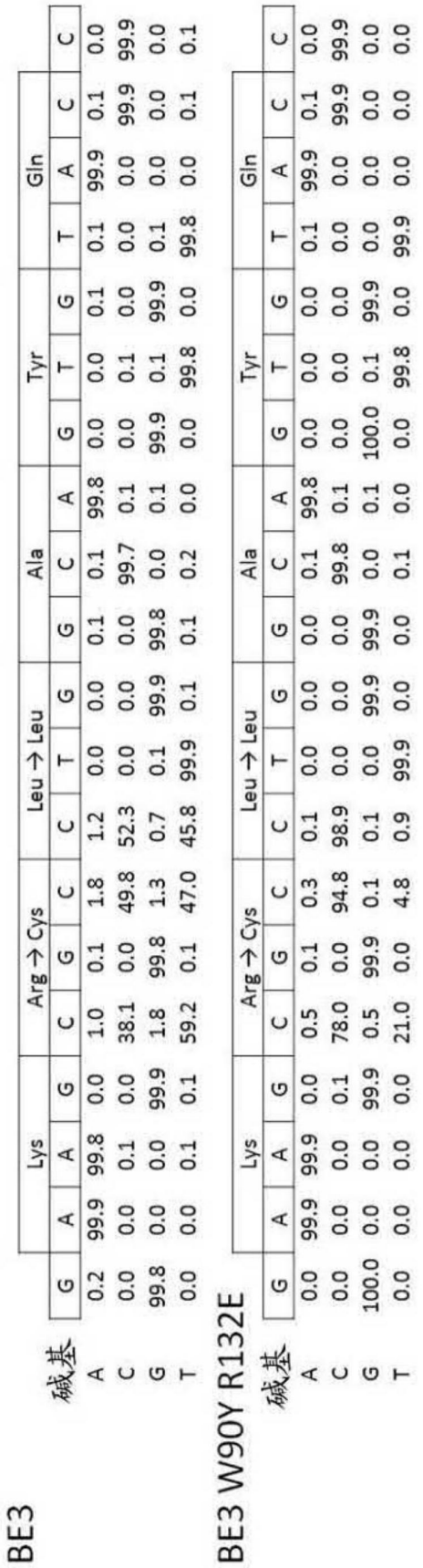

BE3
碱基
Lys
Arg → Cys
Leu → Leu
Ala
Tyr
Gln
G A G C G C C T G G A G T A C
A 0.2 99.9 99.8 0.0 1.0 0.1 1.8 1.2 0.0 0.0 0.1 0.1 99.8 0.0 0.0 0.1 0.1 99.9 0.1 0.0
C 0.0 0.0 0.1 0.0 38.1 0.0 49.8 52.3 0.0 0.0 0.0 99.7 0.1 0.0 0.1 0.0 0.0 0.0 99.9 99.9
G 99.8 0.0 0.0 99.9 1.8 99.8 1.3 0.7 0.1 99.9 99.8 0.0 0.1 99.9 0.1 99.9 0.1 0.0 0.0 0.0
T 0.0 0.0 0.1 0.1 59.2 0.1 47.0 45.8 99.9 0.1 0.1 0.2 0.0 0.0 99.8 0.0 99.8 0.0 0.1 0.1
BE3 W90Y R132E
碱基
Lys
Arg → Cys
Leu → Leu
Ala
Tyr
Gln
G A G C G C C T G G A G T A C
A 0.0 99.9 99.9 0.0 0.5 0.1 0.3 0.1 0.0 0.0 0.0 0.1 99.8 0.0 0.0 0.0 0.1 99.9 0.1 0.0
C 0.0 0.0 0.0 0.1 78.0 0.0 94.8 98.9 0.0 0.0 0.0 99.8 0.1 0.0 0.0 0.0 0.0 0.0 99.9 99.9
G 100.0 0.0 0.0 99.9 0.5 99.9 0.1 0.1 0.0 99.9 99.9 0.0 0.1 100.0 0.1 99.9 0.0 0.0 0.0 0.0
T 0.0 0.0 0.0 0.0 21.0 0.0 4.8 0.9 99.9 0.0 0.0 0.1 0.0 0.0 99.8 0.0 99.9 0.0 0.0 0.0

图72

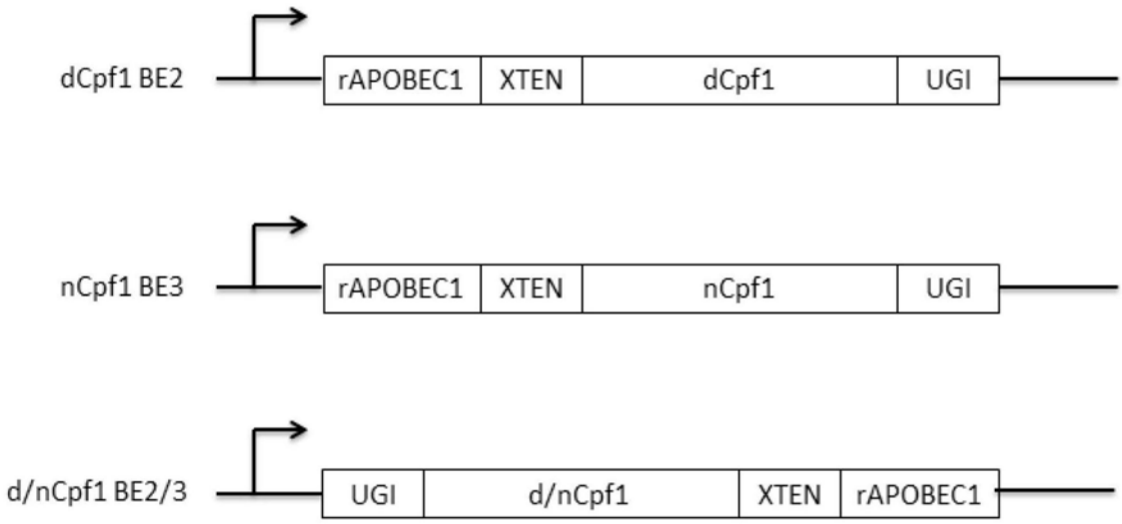

dCpf1 BE2
rAPOBEC1
XTEN
dCpf1
UGI
nCpf1 BE3
rAPOBEC1
XTEN
nCpf1
UGI
d/nCpf1 BE2/3
UGI
d/nCpf1
XTEN
rAPOBEC1

图73

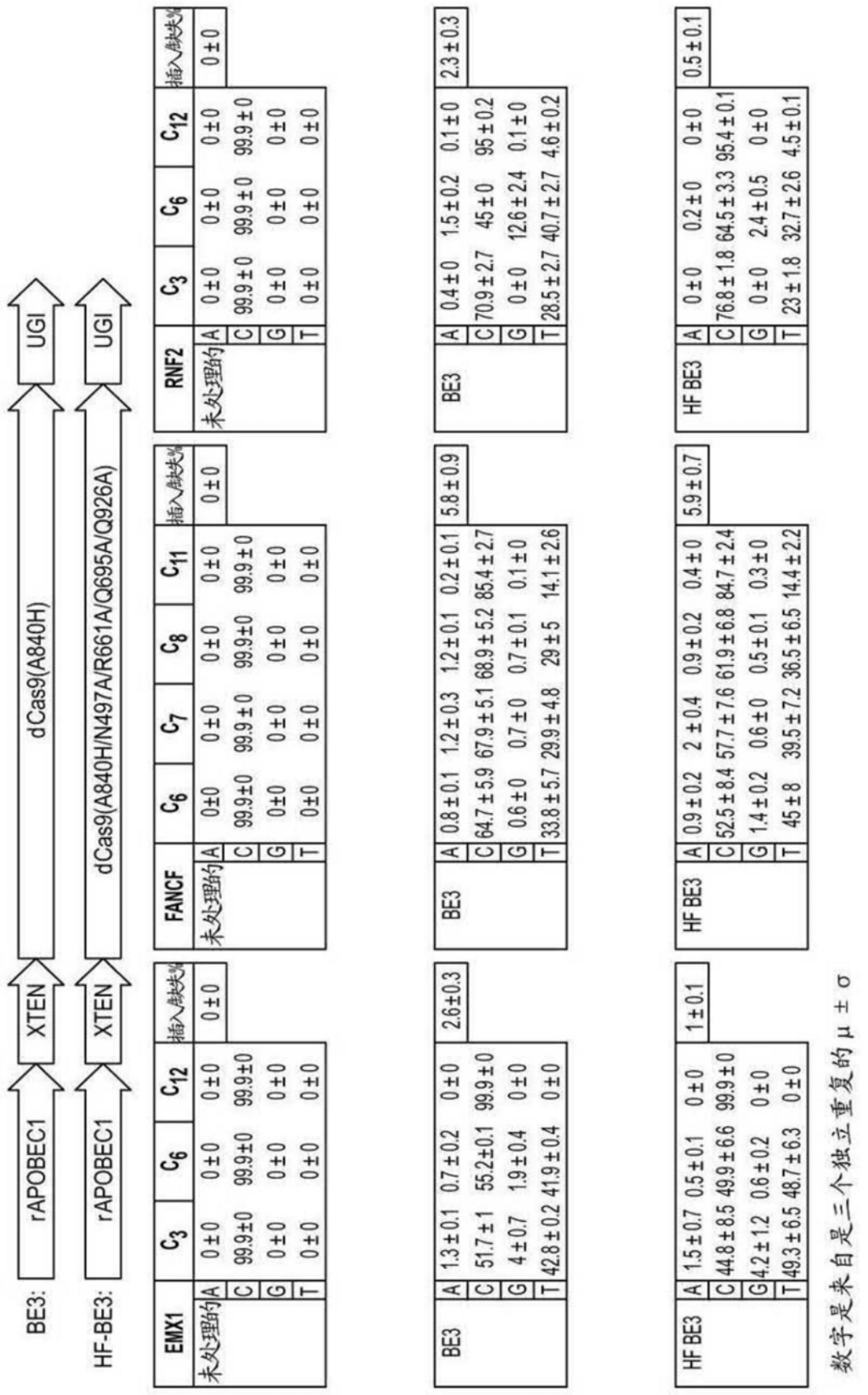
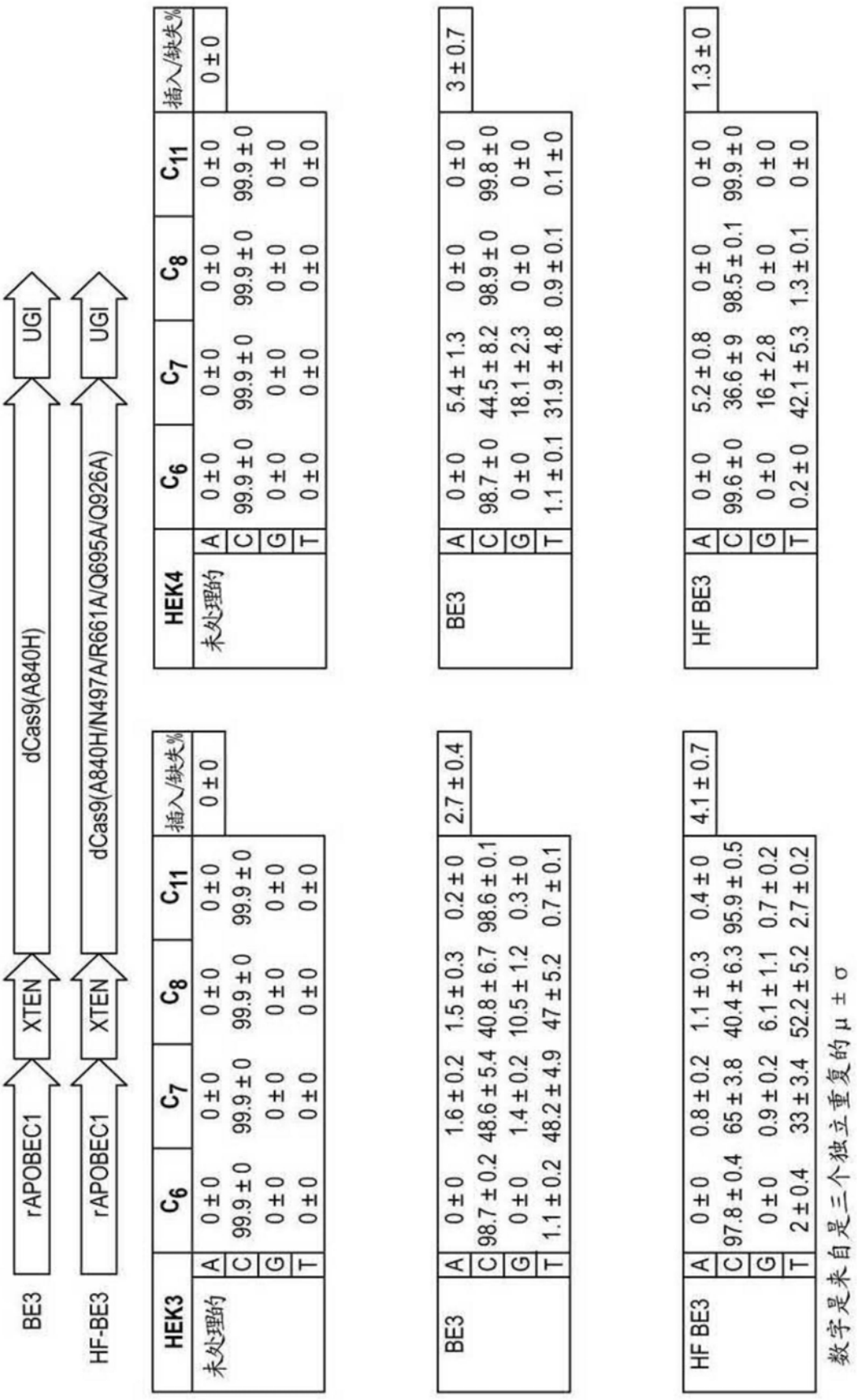

BE3:
rAPOBEC1
XTEN
dCas9(A840H)
UGI
HF-BE3:
rAPOBEC1
XTEN
dCas9(A840H/N497A/R661A/Q695A/Q926A)
UGI
EMX1
FANCF
RNF2
BE3
HF BE3
未处理的
插入/缺失%
数字是来自是三个独立重复的 μ ± σ

图74



HEK3		C_6	C_7	C_8	C_{11}	插入/缺失%
未处理的	A	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
	C	99.9 ± 0	99.9 ± 0	99.9 ± 0	99.9 ± 0	
	G	0 ± 0	0 ± 0	0 ± 0	0 ± 0	
	T	0 ± 0	0 ± 0	0 ± 0	0 ± 0	

HEK4		C_6	C_7	C_8	C_{11}	插入/缺失%
未处理的	A	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
	C	99.9 ± 0	99.9 ± 0	99.9 ± 0	99.9 ± 0	
	G	0 ± 0	0 ± 0	0 ± 0	0 ± 0	
	T	0 ± 0	0 ± 0	0 ± 0	0 ± 0	

BE3						
	A	0 ± 0	1.6 ± 0.2	1.5 ± 0.3	0.2 ± 0	2.7 ± 0.4
	C	98.7 ± 0.2	48.6 ± 5.4	40.8 ± 6.7	98.6 ± 0.1	
	G	0 ± 0	1.4 ± 0.2	10.5 ± 1.2	0.3 ± 0	
	T	1.1 ± 0.2	48.2 ± 4.9	47 ± 5.2	0.7 ± 0.1	

BE3						
	A	0 ± 0	5.4 ± 1.3	0 ± 0	0 ± 0	3 ± 0.7
	C	98.7 ± 0	44.5 ± 8.2	98.9 ± 0	99.8 ± 0	
	G	0 ± 0	18.1 ± 2.3	0 ± 0	0 ± 0	
	T	1.1 ± 0.1	31.9 ± 4.8	0.9 ± 0.1	0.1 ± 0	

HF BE3						
	A	0 ± 0	0.8 ± 0.2	1.1 ± 0.3	0.4 ± 0	4.1 ± 0.7
	C	97.8 ± 0.4	65 ± 3.8	40.4 ± 6.3	95.9 ± 0.5	
	G	0 ± 0	0.9 ± 0.2	6.1 ± 1.1	0.7 ± 0.2	
	T	2 ± 0.4	33 ± 3.4	52.2 ± 5.2	2.7 ± 0.2	

HF BE3						
	A	0 ± 0	5.2 ± 0.8	0 ± 0	0 ± 0	1.3 ± 0
	C	99.6 ± 0	36.6 ± 9	98.5 ± 0.1	99.9 ± 0	
	G	0 ± 0	16 ± 2.8	0 ± 0	0 ± 0	
	T	0.2 ± 0	42.1 ± 5.3	1.3 ± 0.1	0 ± 0	

数字是来自是三个独立重复的 $\mu \pm \sigma$

图75

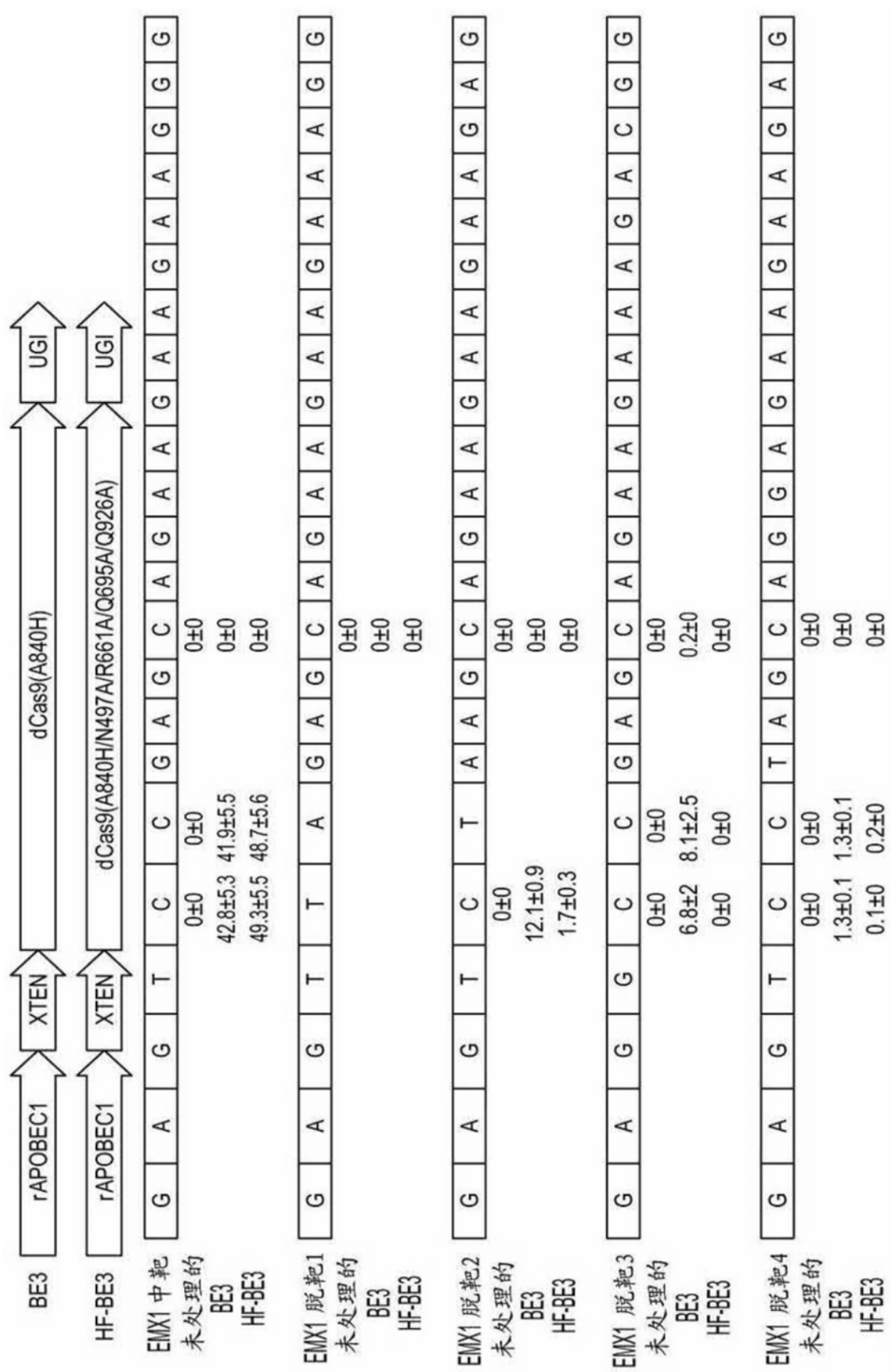

BE3
rAPOBEC1
XTEN
dCas9(A840H)
UGI
HF-BE3
rAPOBEC1
XTEN
dCas9(A840H/N497A/R661A/Q695A/Q926A)
UGI
EMX1 中靶
未处理的
BE3
HF-BE3
EMX1 脱靶1
未处理的
BE3
HF-BE3
EMX1 脱靶2
未处理的
BE3
HF-BE3
EMX1 脱靶3
未处理的
BE3
HF-BE3
EMX1 脱靶4
未处理的
BE3
HF-BE3

图76

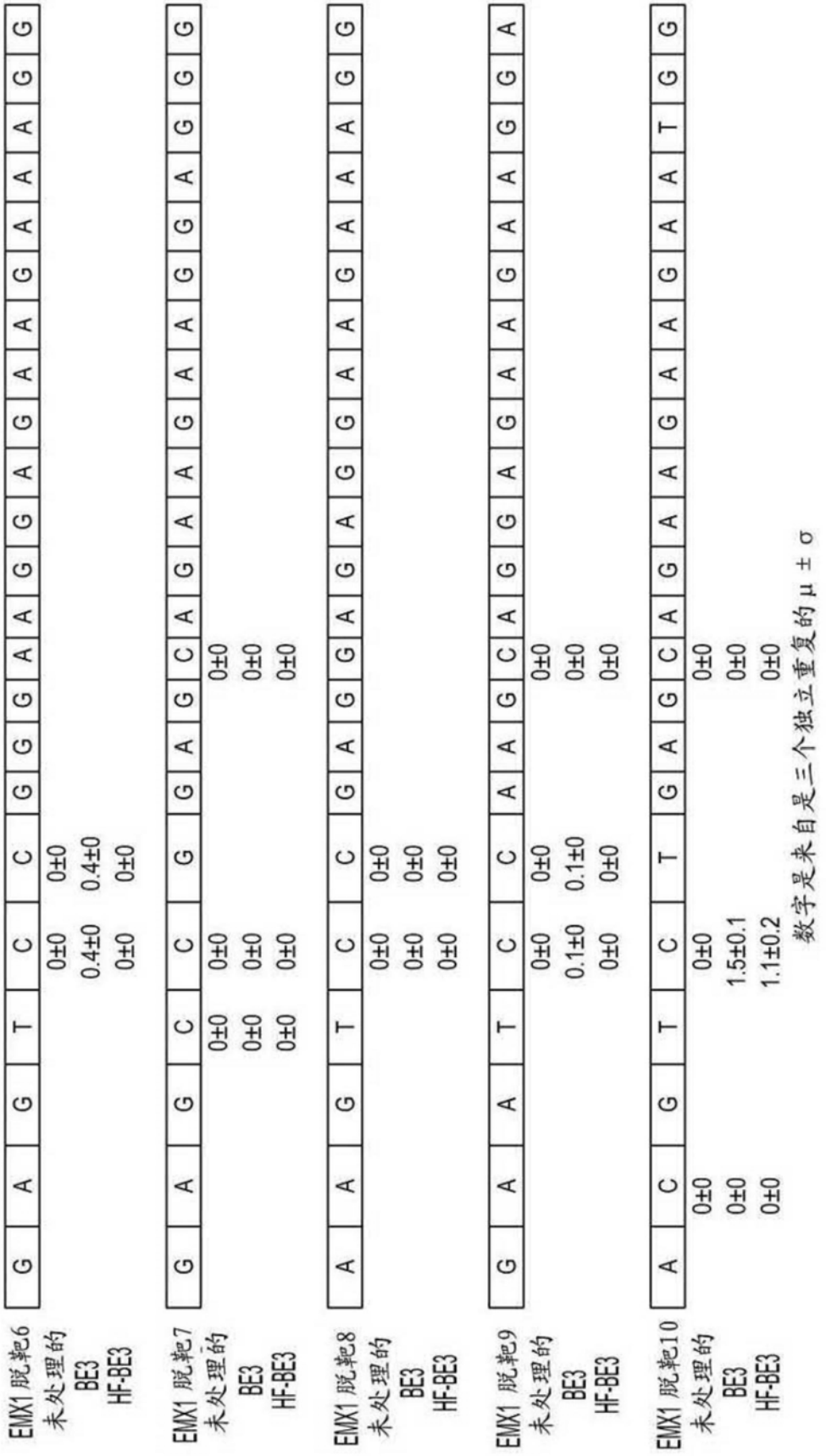

EMX1 脱靶6 G A G T C C G G G A A G G A A G A A A G G
未处理的 0±0 0±0
BE3 0.4±0 0.4±0
HF-BE3 0±0 0±0
EMX1 脱靶7 G A G C C G G A G C A G A A G G A A G G G
未处理的 0±0 0±0 0±0
BE3 0±0 0±0 0±0
HF-BE3 0±0 0±0 0±0
EMX1 脱靶8 A A G T C C G A G G A G G A A G A A A G G
未处理的 0±0 0±0
BE3 0±0 0±0
HF-BE3 0±0 0±0
EMX1 脱靶9 G A A T C C A A G C A G G A G A A G G A A
未处理的 0±0 0±0 0±0
BE3 0.1±0 0.1±0 0±0
HF-BE3 0±0 0±0 0±0
EMX1 脱靶10 A C G T C T G A G C A G A A G A A A T G G
未处理的 0±0 0±0 0±0
BE3 0±0 1.5±0.1 0±0
HF-BE3 0±0 1.1±0.2 0±0
数字是来自是三个独立重复的 μ ± σ

图76续

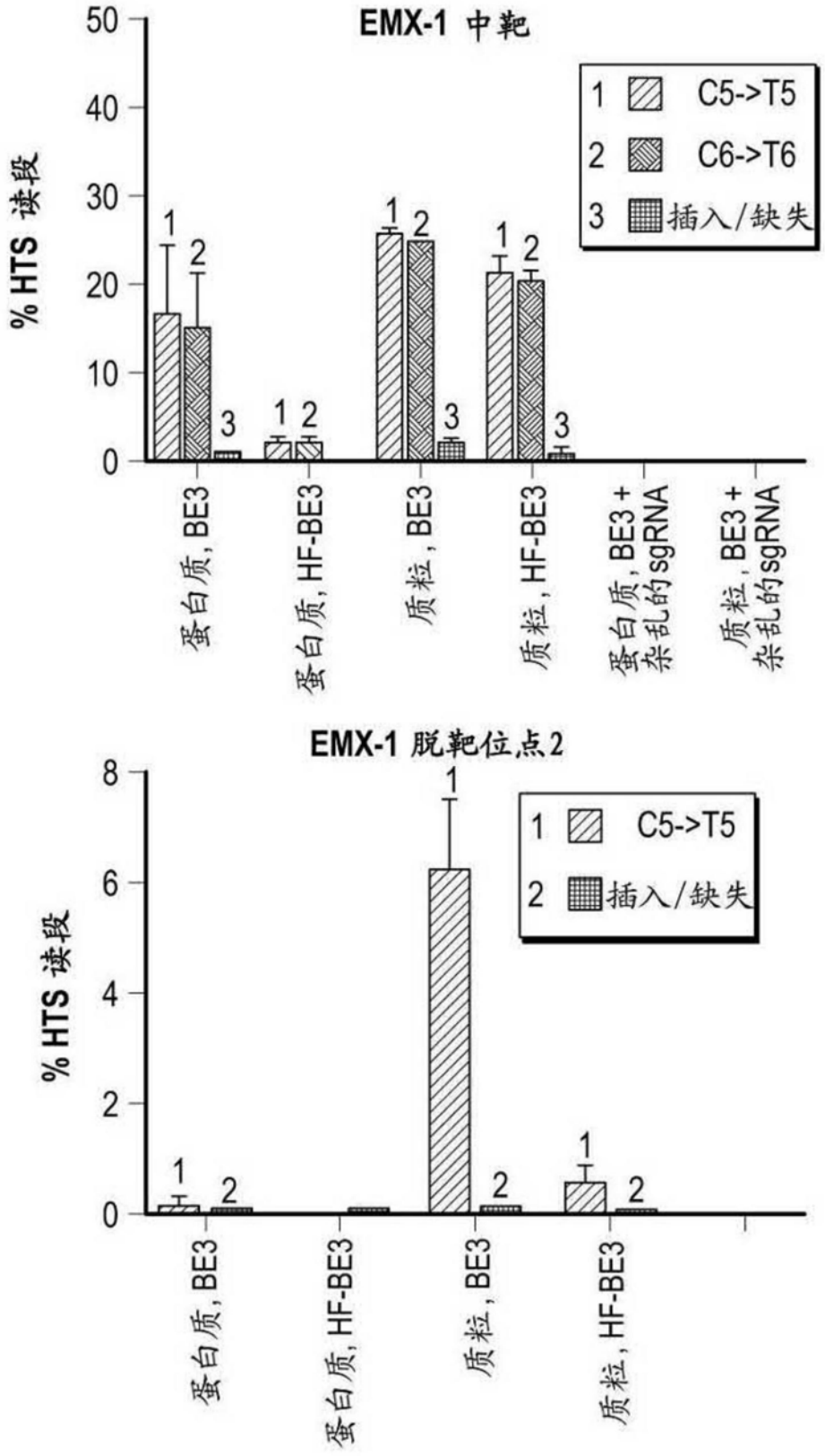

EMX-1 中靶
% HTS 读段
50
40
30
20
10
0
1 C5->T5
2 C6->T6
3 插入/缺失
蛋白质, BE3
蛋白质, HF-BE3
质粒, BE3
质粒, HF-BE3
蛋白质, BE3 + 杂乱的sgRNA
质粒, BE3 + 杂乱的sgRNA
EMX-1 脱靶位点2
% HTS 读段
8
6
4
2
0
1 C5->T5
2 插入/缺失
蛋白质, BE3
蛋白质, HF-BE3
质粒, BE3
质粒, HF-BE3

图77

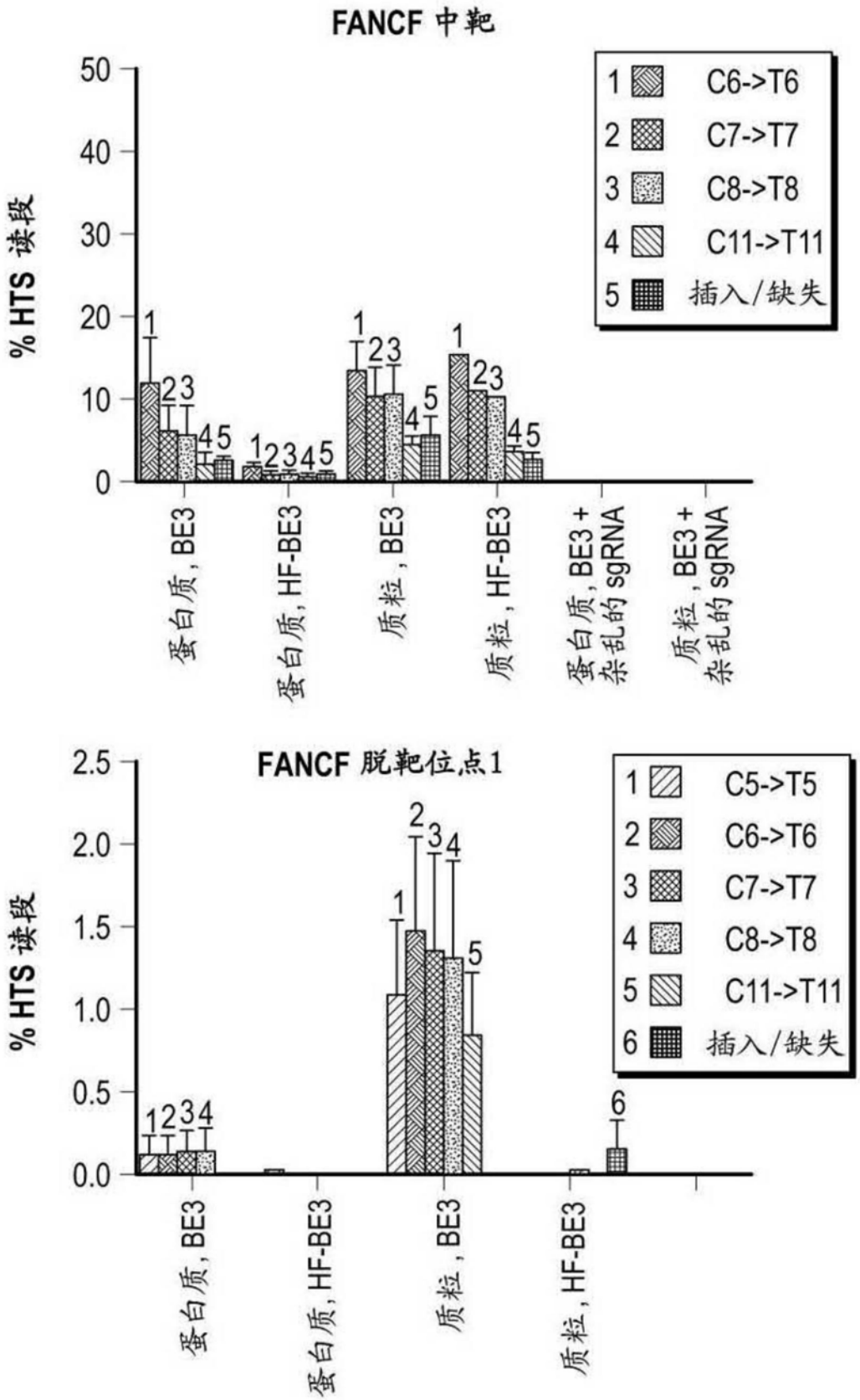

FANCF 中靶
% HTS 读段
50
40
30
20
10
0
1 C6->T6
2 C7->T7
3 C8->T8
4 C11->T11
5 插入/缺失
蛋白质，BE3
蛋白质，HF-BE3
质粒，BE3
质粒，HF-BE3
蛋白质，BE3 + 杂乱的 sgRNA
质粒，BE3 + 杂乱的 sgRNA
FANCF 脱靶位点1
% HTS 读段
2.5
2.0
1.5
1.0
0.5
0.0
1 C5->T5
2 C6->T6
3 C7->T7
4 C8->T8
5 C11->T11
6 插入/缺失
蛋白质，BE3
蛋白质，HF-BE3
质粒，BE3
质粒，HF-BE3

图78

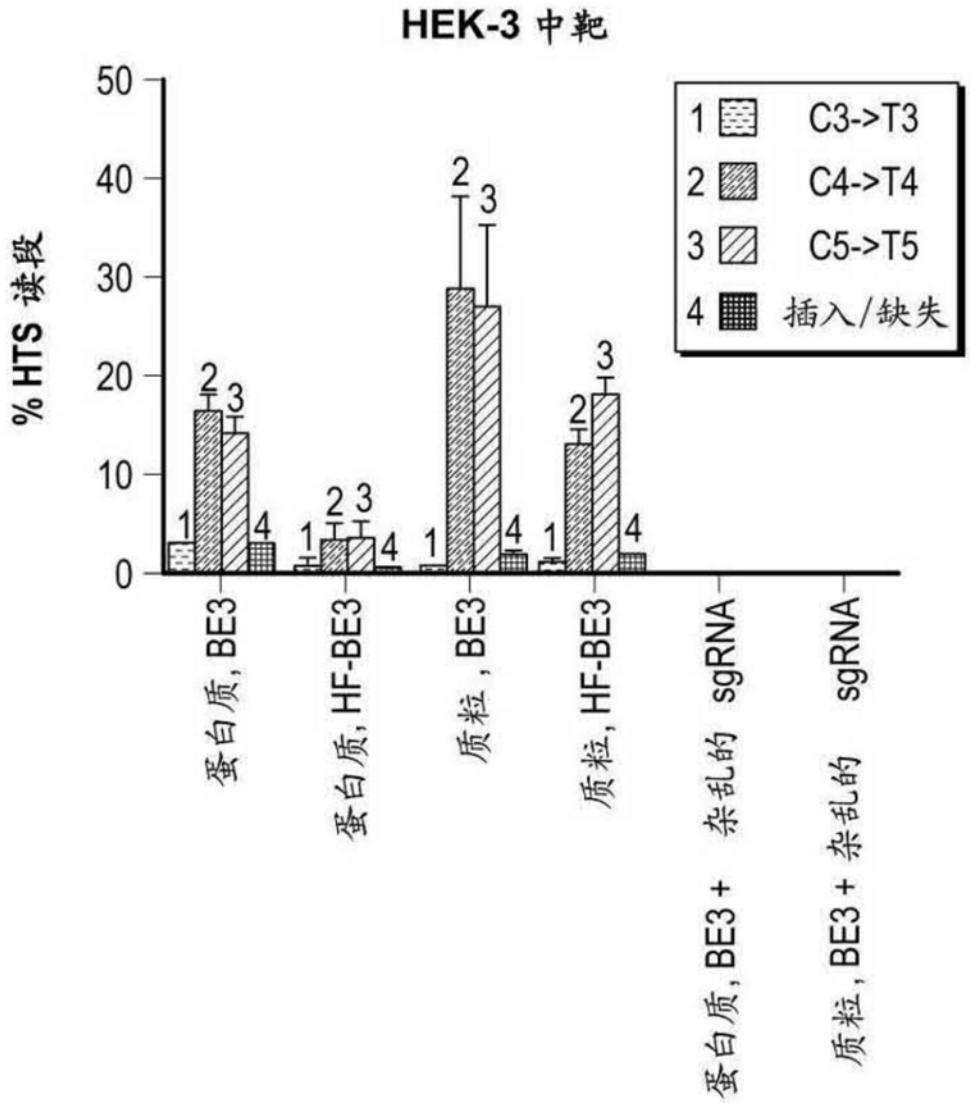

HEK-3 中靶
% HTS 读段
50
40
30
20
10
0
1 C3->T3
2 C4->T4
3 C5->T5
4 插入/缺失
蛋白质, BE3
蛋白质, HF-BE3
质粒 , BE3
质粒, HF-BE3
蛋白质, BE3 + 杂乱的 sgRNA
质粒 , BE3 + 杂乱的 sgRNA

图79

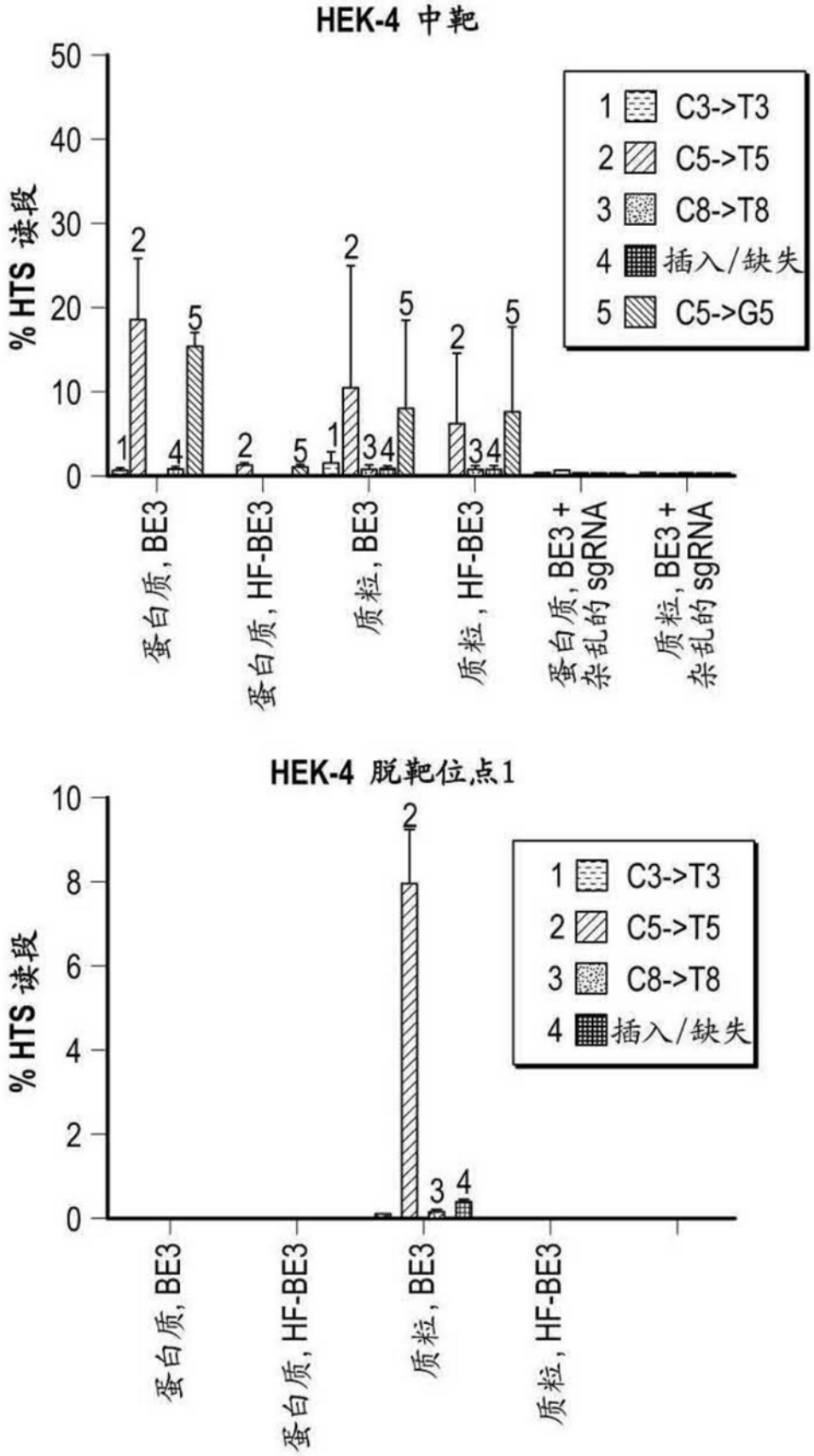

HEK-4 中靶
% HTS 读段
50
40
30
20
10
0
1 C3->T3
2 C5->T5
3 C8->T8
4 插入/缺失
5 C5->G5
蛋白质, BE3
蛋白质, HF-BE3
质粒, BE3
质粒, HF-BE3
蛋白质, BE3 + 杂乱的 sgRNA
质粒, BE3 + 杂乱的 sgRNA
HEK-4 脱靶位点1
% HTS 读段
10
8
6
4
2
0
1 C3->T3
2 C5->T5
3 C8->T8
4 插入/缺失
蛋白质, BE3
蛋白质, HF-BE3
质粒, BE3
质粒, HF-BE3

图80

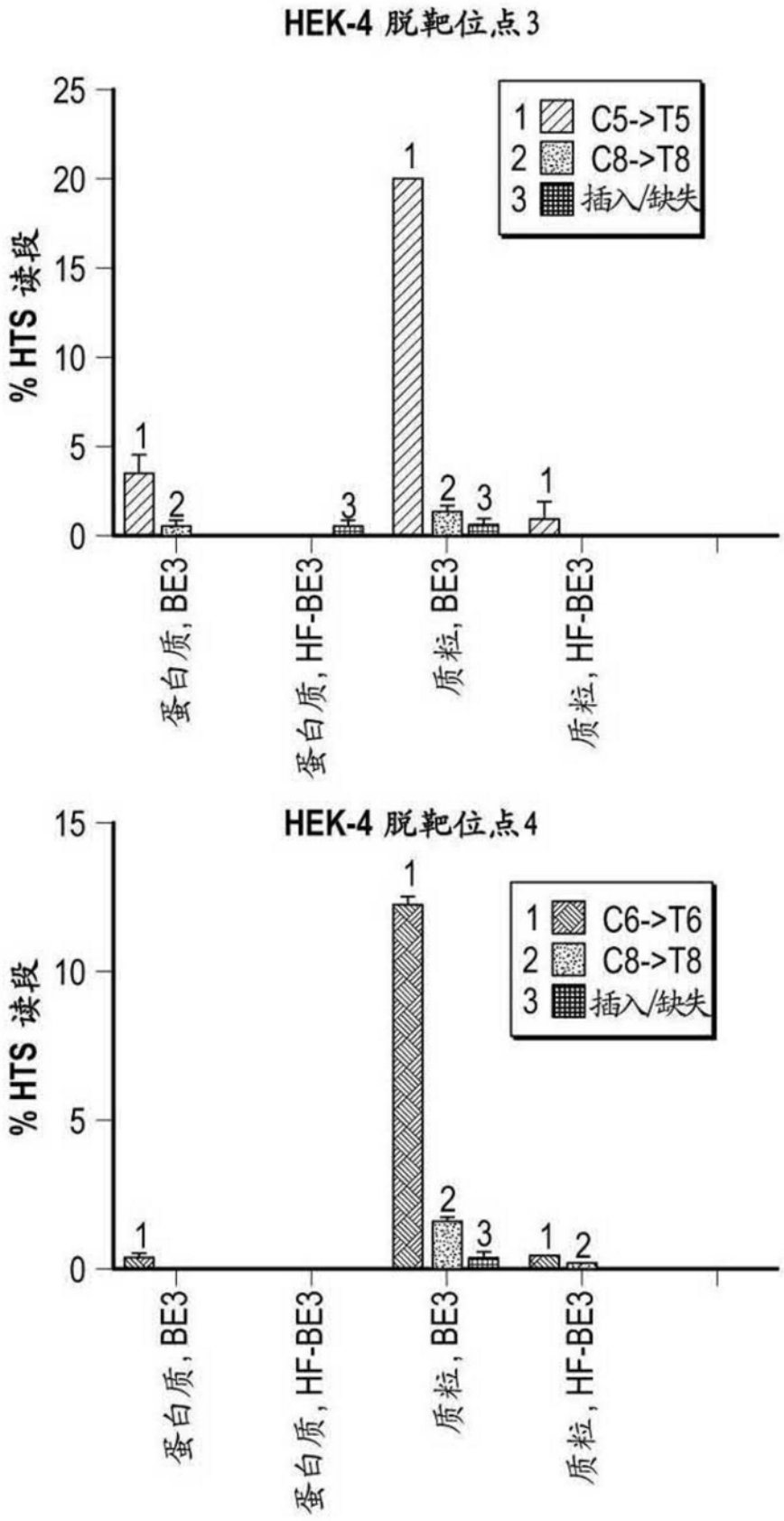

HEK-4 脱靶位点3
% HTS 读段
25
20
15
10
5
0
1 C5->T5
2 C8->T8
3 插入/缺失
蛋白质, BE3
蛋白质, HF-BE3
质粒, BE3
质粒, HF-BE3
HEK-4 脱靶位点4
% HTS 读段
15
10
5
0
1 C6->T6
2 C8->T8
3 插入/缺失
蛋白质, BE3
蛋白质, HF-BE3
质粒, BE3
质粒, HF-BE3

图80续

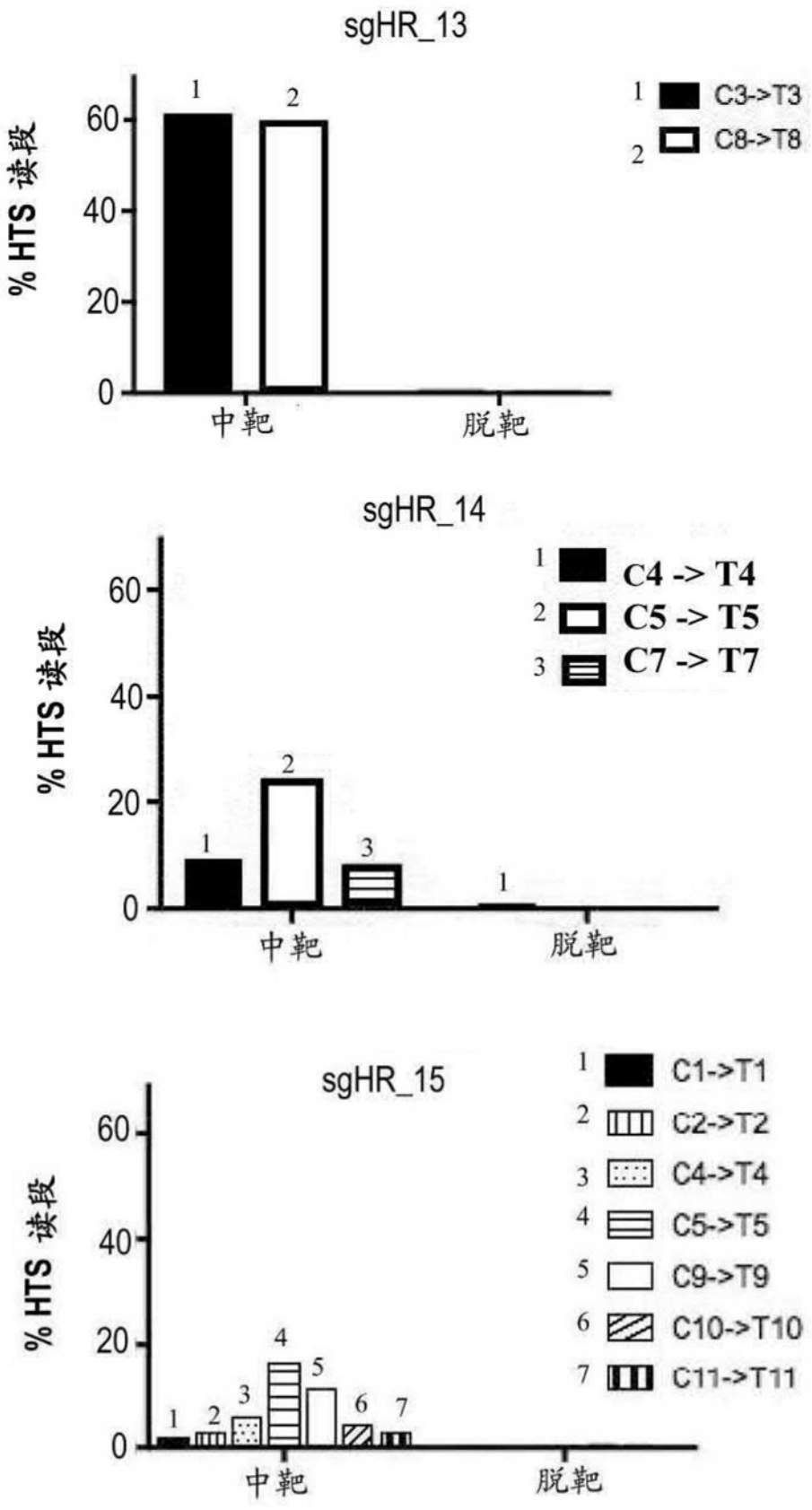

sgHR_13
% HTS 读段
60
40
20
0
1
2
1 C3->T3
2 C8->T8
中靶
脱靶
sgHR_14
% HTS 读段
60
40
20
0
1
2
3
1
1 C4 -> T4
2 C5 -> T5
3 C7 -> T7
中靶
脱靶
sgHR_15
% HTS 读段
60
40
20
0
1
2
3
4
5
6
7
1 C1->T1
2 C2->T2
3 C4->T4
4 C5->T5
5 C9->T9
6 C10->T10
7 C11->T11
中靶
脱靶

图81

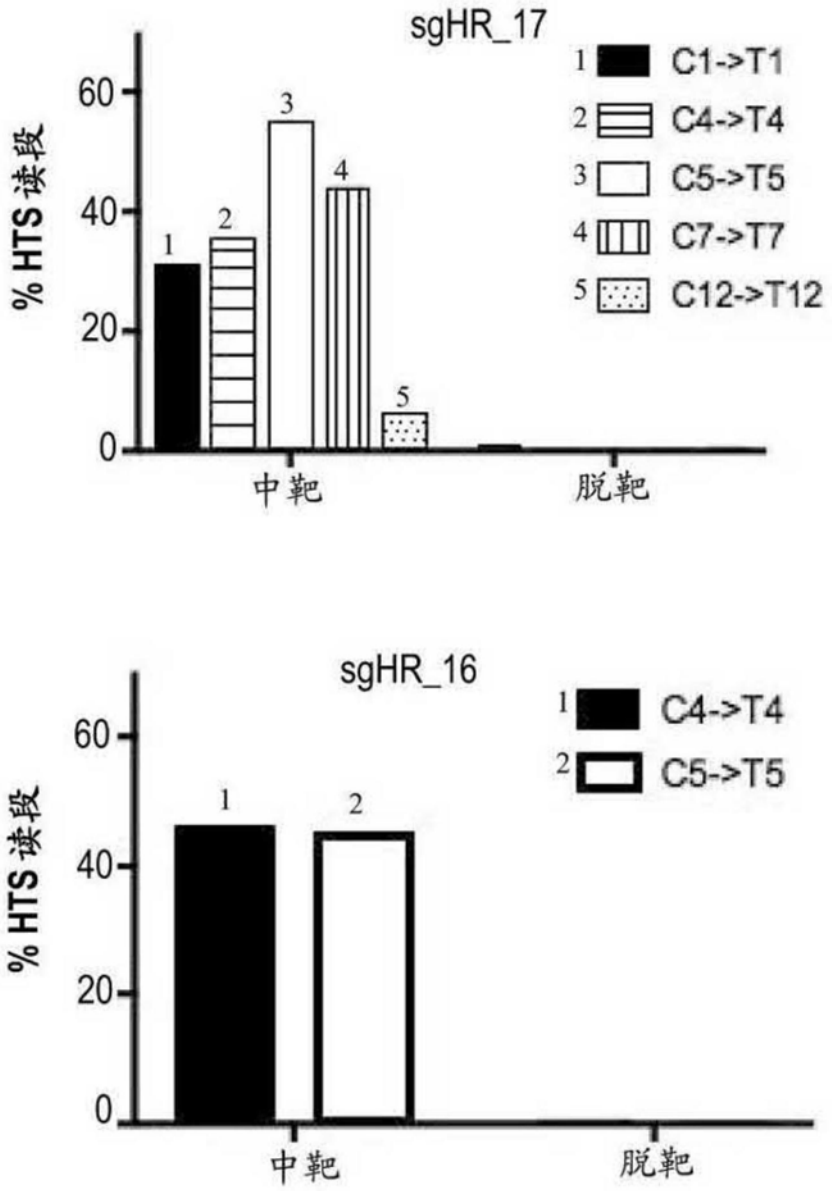

sgHR_17
% HTS 读段
60
40
20
0
1
2
3
4
5
1 C1->T1
2 C4->T4
3 C5->T5
4 C7->T7
5 C12->T12
中靶
脱靶
sgHR_16
% HTS 读段
60
40
20
0
1
2
1 C4->T4
2 C5->T5
中靶
脱靶

图82

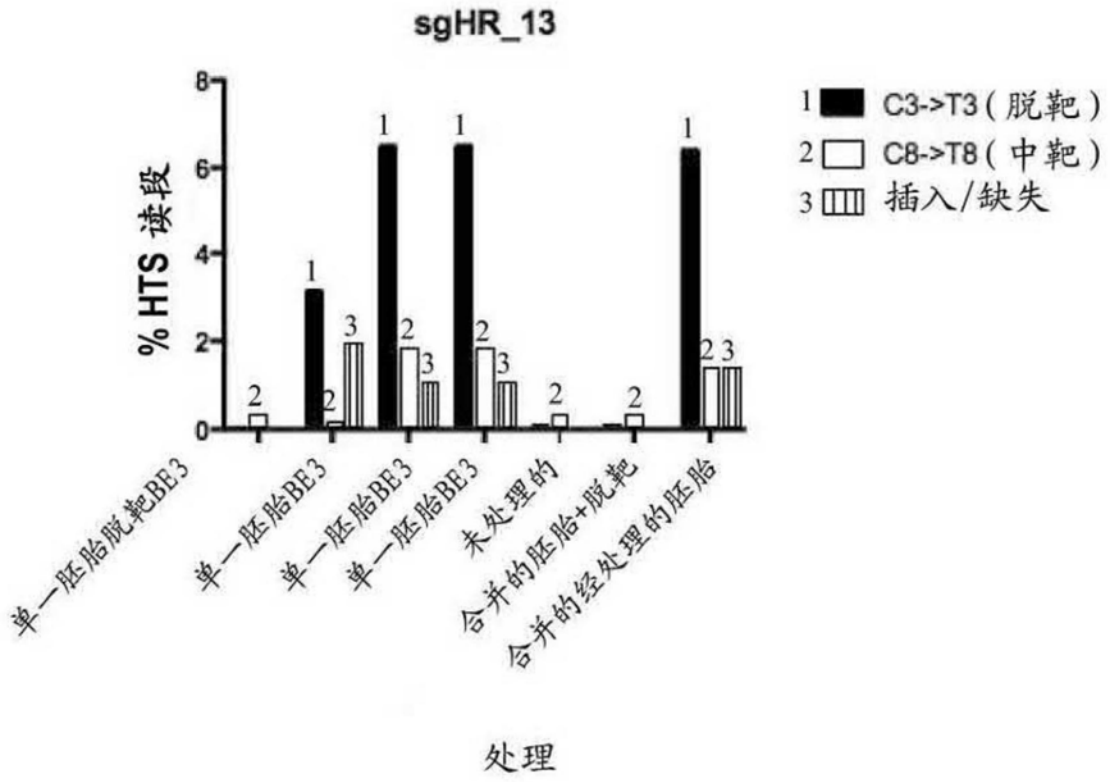

sgHR_13
% HTS 读段
处理
1 C3->T3 (脱靶)
2 C8->T8 (中靶)
3 插入/缺失
单一胚胎脱靶BE3
单一胚胎BE3
单一胚胎BE3
单一胚胎BE3
未处理的
合并的胚胎+脱靶
合并的经处理的胚胎

图83

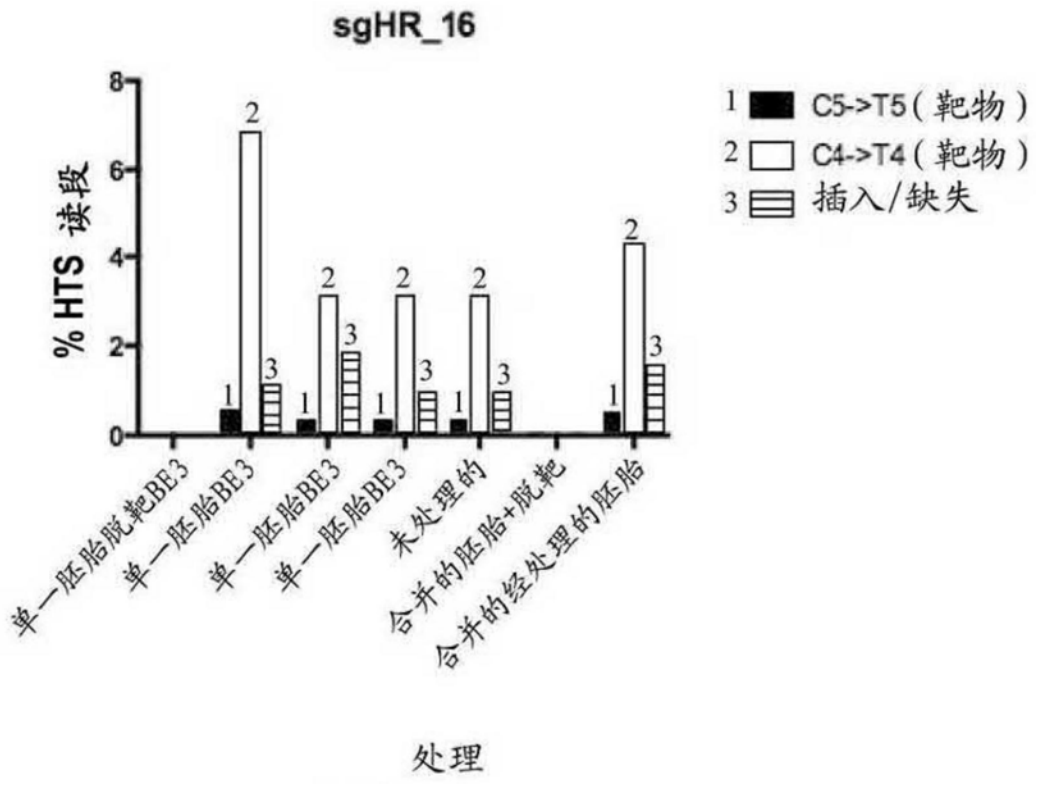

sgHR_16
% HTS 读段
处理
1 C5->T5 (靶物)
2 C4->T4 (靶物)
3 插入/缺失
单一胚胎脱靶BE3
单一胚胎BE3
单一胚胎BE3
单一胚胎BE3
未处理的
合并的胚胎+脱靶
合并的经处理的胚胎

图84

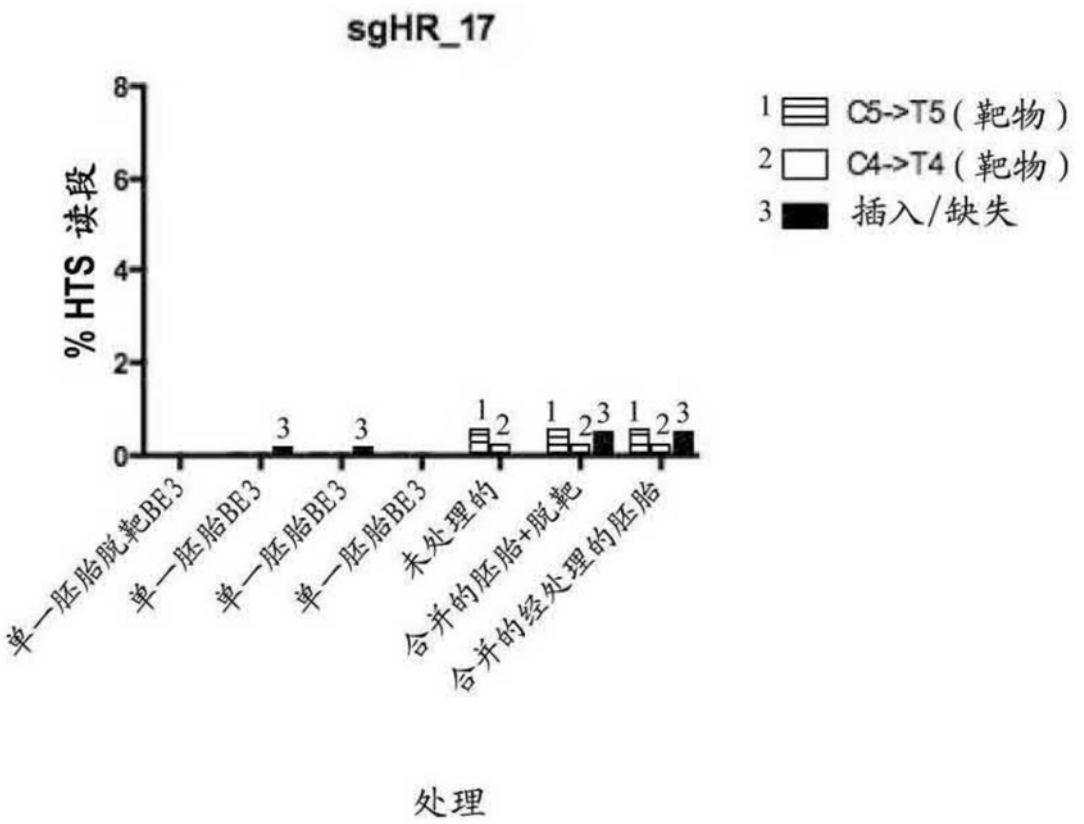

sgHR_17
% HTS 读段
8
6
4
2
0
1 C5->T5(靶物)
2 C4->T4(靶物)
3 插入/缺失
3
3
1 2
1 2 3
1 2 3
单一胚胎脱靶BE3
单一胚胎BB3
单一胚胎BB3
单一胚胎BB3
未处理的
合并的胚胎+脱靶
合并的经处理的胚胎
处理

图85

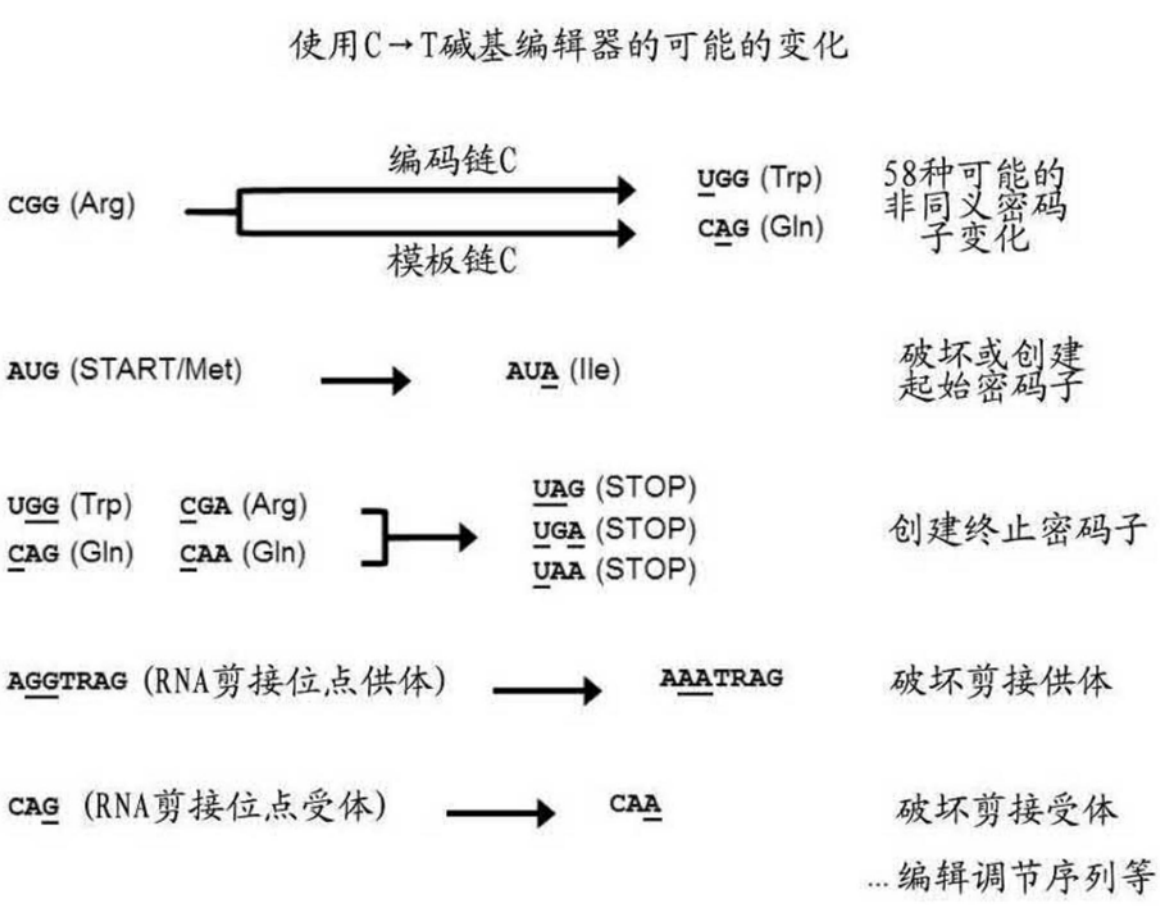

使用C→T碱基编辑器的可能的变化
编码链C
CGG (Arg)
模板链C
UGG (Trp)
CAG (Gln)
58种可能的非同义密码子变化
AUG (START/Met)
AUA (Ile)
破坏或创建起始密码子
UGG (Trp)
CGA (Arg)
CAG (Gln)
CAA (Gln)
UAG (STOP)
UGA (STOP)
UAA (STOP)
创建终止密码子
AGGTRAG (RNA剪接位点供体)
AAATRAG
破坏剪接供体
CAG (RNA剪接位点受体)
CAA
破坏剪接受体
...编辑调节序列等

图86

碱基编辑APOE4至APOE3r
• 载脂蛋白E具有在氨基酸112和158处不同的4种同种型
• APOE4是迟发性阿尔茨海默氏病的最大且最常见的遗传风险因素
等位基因
残基
残基
AD风险(已知的或假设的)
APOE2
Cys TGC
Cys TGC
低
APOE3
Cys TGC
Arg CGC
中性
APOE3r
Arg CGC
Cys TGC
中性
APOE4
Arg CGC
Arg CGC
高
BE3
Arg112
Arg158

图87

未处理的

APOE4 C158R	Lys			Arg 158			Leu			Ala			Val			Tyr			Gln			插入/缺失%
	G	A	A	G	C_5	G	C	C	T	G	G	C	A	G	T	G	T	A	C	C	A	G	G
	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	**0.0**
A	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	
C	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	
G	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	100.0	
T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

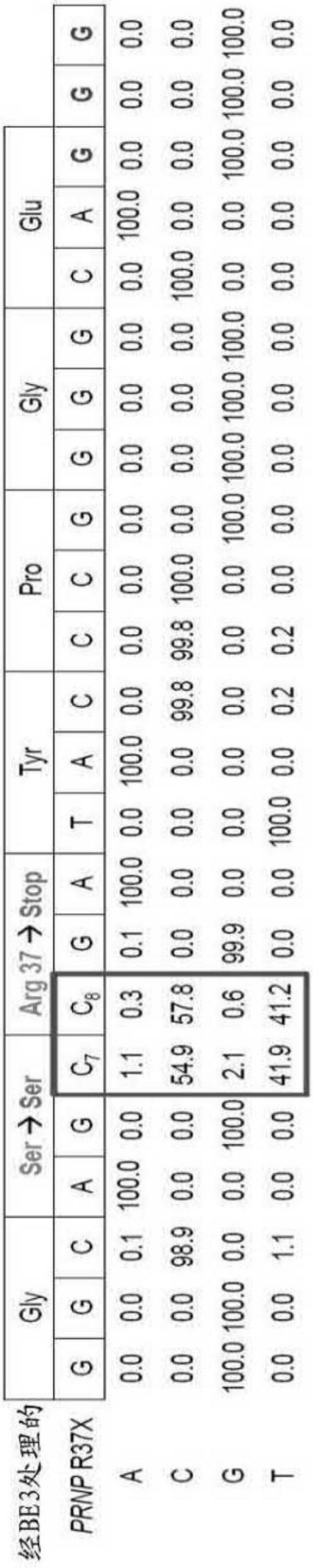
Cas9 + HDR

APOE4 C158R	Lys			Arg 158 → Cys			Leu			Ala			Val			Tyr			Gln			插入/缺失%
	G	A	A	G	C_5	G	C	C	T	G	G	C	A	G	T	G	T	A	C	C	A	G	G
	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.4	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	**26.1**
A	0.0	0.0	0.0	0.0	99.7	0.0	99.9	99.9	0.0	0.0	0.0	100.0	0.5	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	
C	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	100.0	99.9	0.0	0.0	99.6	0.6	99.9	0.2	0.0	0.0	0.0	0.0	100.0	100.0	
G	0.0	0.0	0.0	0.0	**0.3**	0.0	0.1	0.1	100.0	0.0	0.0	0.0	0.1	0.4	99.3	0.1	99.8	0.0	0.0	0.0	0.0	0.0	0.0	
T																								

经BE3处理的

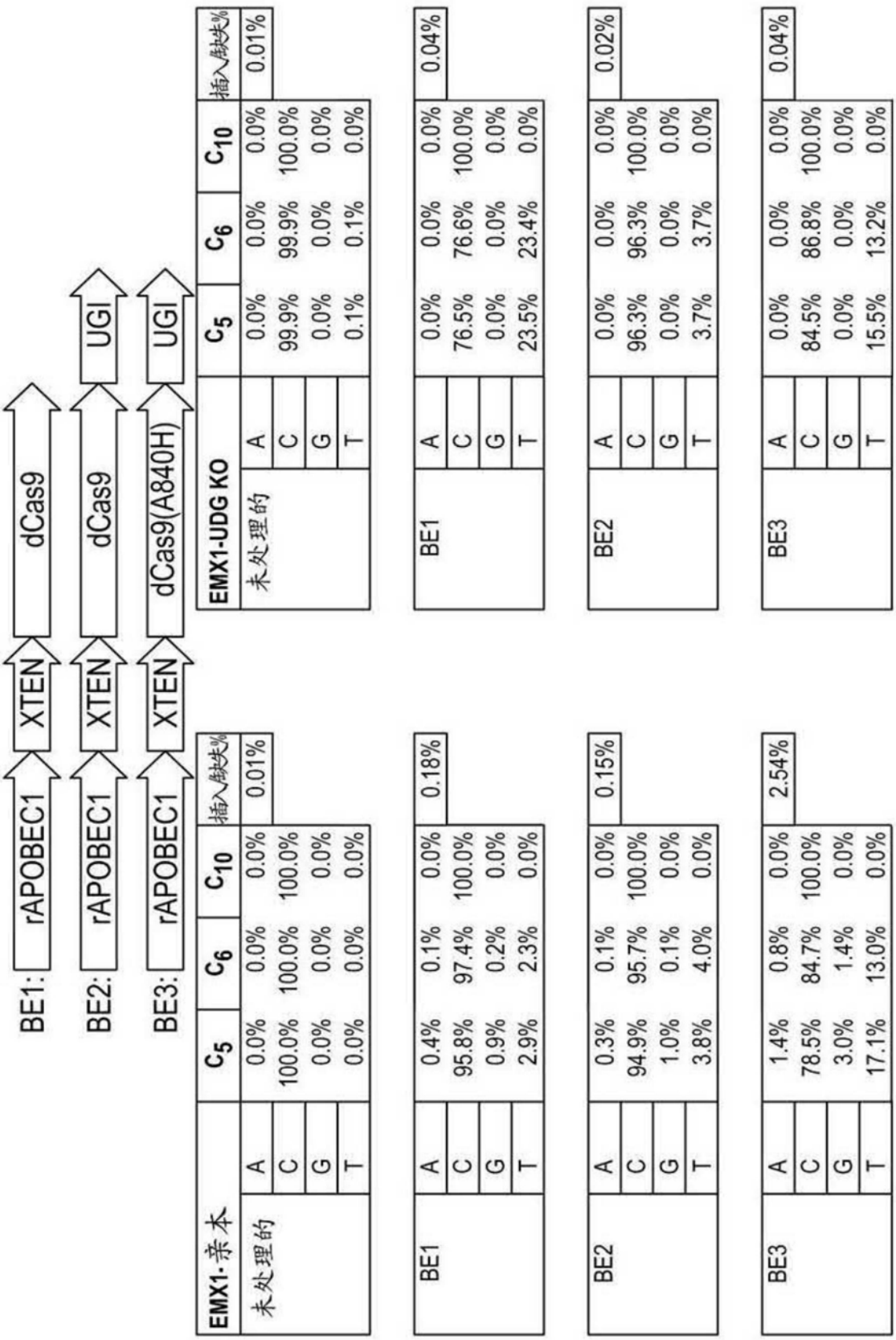
APOE4 C158R	Lys			Arg 158 → Cys			Leu → Leu			Ala			Val			Tyr			Gln			插入/缺失%
	G	A	A	G	C_5	G	C_7	C_8	T	G	G	C	A	G	T	G	T	A	C	C	A	G	G
	0.1	100.0	100.0	0.0	0.5	0.0	1.3	0.9	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	**4.6**
A	0.0	0.0	0.0	0.0	23.7	0.0	47.4	43.5	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.1	
C	99.9	0.0	0.0	100.0	0.9	99.9	1.1	0.7	0.0	100.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	99.9	
G	0.0	0.0	0.0	0.0	**74.9**	0.1	**50.2**	**55.0**	100.0	0.0	0.0	0.1	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	
T																								

图88



经BE3处理的		Gly			Ser → Ser			Arg 37 → Stop			Tyr			Pro			Gly			Glu				
PRNP R37X		G	G	C	A	G	C_7	C_8	G	A	T	A	C	C	C	G	G	G	G	C	A	G	G	G
A		0.0	0.0	0.1	100.0	0.0	1.1	0.3	0.1	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0
C		0.0	0.0	98.9	0.0	0.0	54.9	57.8	0.0	0.0	0.0	0.0	99.8	99.8	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0
G		100.0	100.0	0.0	0.0	100.0	2.1	0.6	99.9	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	100.0	0.0	0.0	100.0	100.0	100.0
T		0.0	0.0	1.1	0.0	0.0	41.9	41.2	0.0	0.0	100.0	0.0	0.2	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

图89



EMX1-亲本		C_5	C_6	C_{10}	插入缺失%
未处理的	A	0.0%	0.0%	0.0%	0.01%
	C	100.0%	100.0%	100.0%	
	G	0.0%	0.0%	0.0%	
	T	0.0%	0.0%	0.0%	

BE1		C_5	C_6	C_{10}	插入缺失%
	A	0.4%	0.1%	0.0%	0.18%
	C	95.8%	97.4%	100.0%	
	G	0.9%	0.2%	0.0%	
	T	2.9%	2.3%	0.0%	

BE2		C_5	C_6	C_{10}	插入缺失%
	A	0.3%	0.1%	0.0%	0.15%
	C	94.9%	95.7%	100.0%	
	G	1.0%	0.1%	0.0%	
	T	3.8%	4.0%	0.0%	

BE3		C_5	C_6	C_{10}	插入缺失%
	A	1.4%	0.8%	0.0%	2.54%
	C	78.5%	84.7%	100.0%	
	G	3.0%	1.4%	0.0%	
	T	17.1%	13.0%	0.0%	

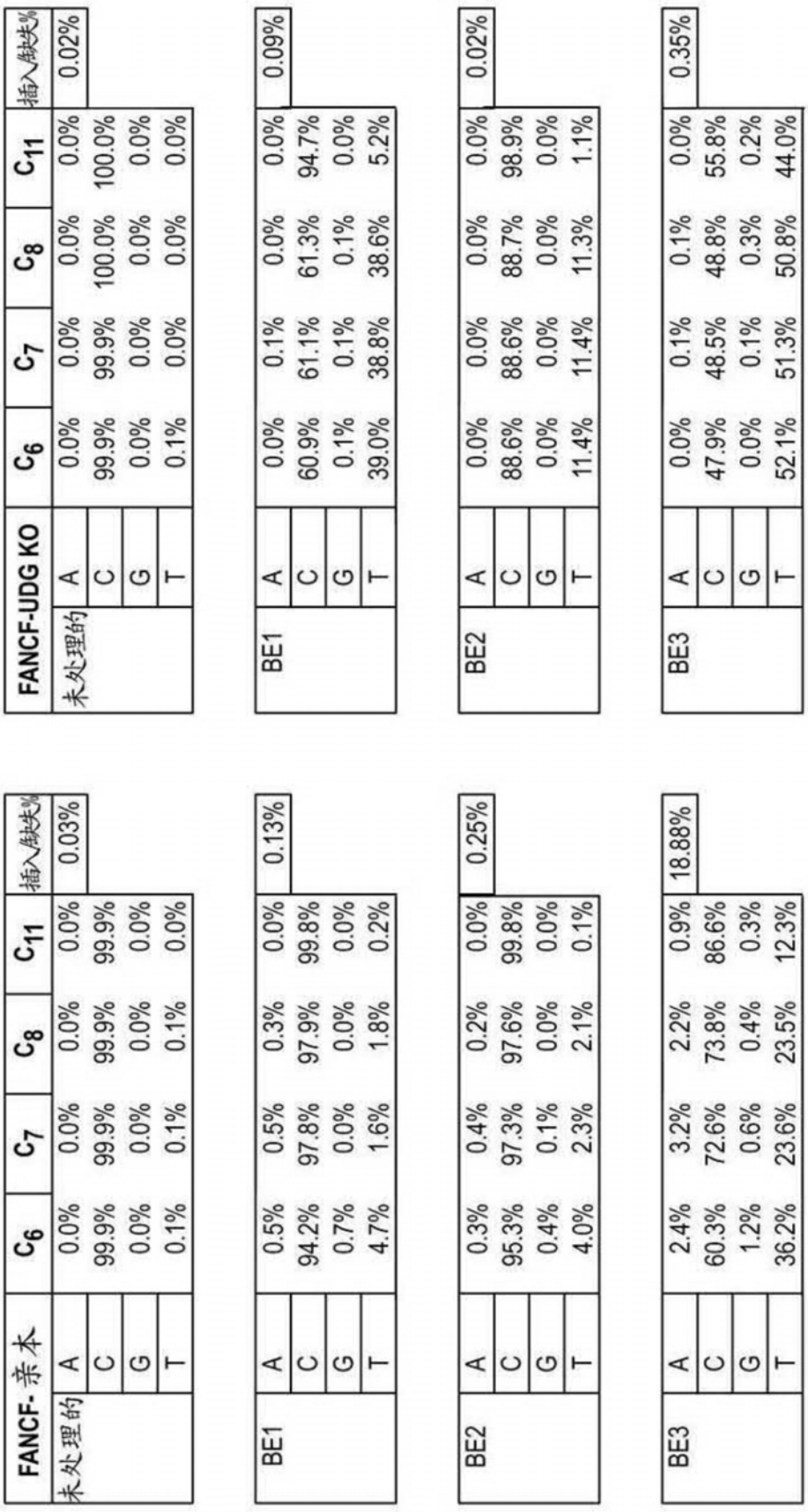
EMX1-UDG KO		C_5	C_6	C_{10}	插入缺失%
未处理的	A	0.0%	0.0%	0.0%	0.01%
	C	99.9%	99.9%	100.0%	
	G	0.0%	0.0%	0.0%	
	T	0.1%	0.1%	0.0%	

BE1		C_5	C_6	C_{10}	插入缺失%
	A	0.0%	0.0%	0.0%	0.04%
	C	76.5%	76.6%	100.0%	
	G	0.0%	0.0%	0.0%	
	T	23.5%	23.4%	0.0%	

BE2		C_5	C_6	C_{10}	插入缺失%
	A	0.0%	0.0%	0.0%	0.02%
	C	96.3%	96.3%	100.0%	
	G	0.0%	0.0%	0.0%	
	T	3.7%	3.7%	0.0%	

BE3		C_5	C_6	C_{10}	插入缺失%
	A	0.0%	0.0%	0.0%	0.04%
	C	84.5%	86.8%	100.0%	
	G	0.0%	0.0%	0.0%	
	T	15.5%	13.2%	0.0%	

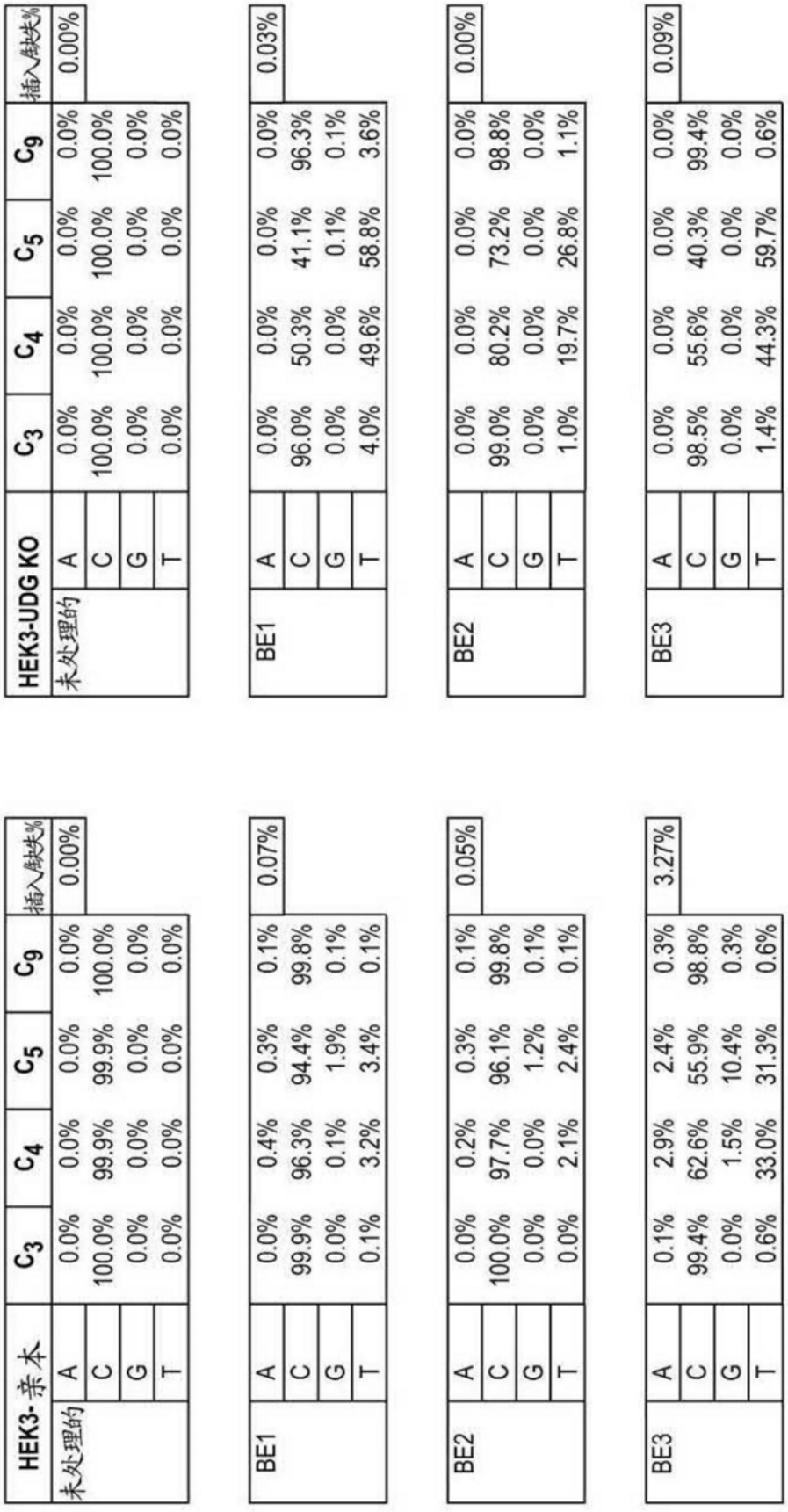
图90



FANCF-亲本		C_6	C_7	C_8	C_{11}	插入缺失%
未处理的	A	0.0%	0.0%	0.0%	0.0%	0.03%
	C	99.9%	99.9%	99.9%	99.9%	
	G	0.0%	0.0%	0.0%	0.0%	
	T	0.1%	0.1%	0.1%	0.0%	
BE1	A	0.5%	0.5%	0.3%	0.0%	0.13%
	C	94.2%	97.8%	97.9%	99.8%	
	G	0.7%	0.0%	0.0%	0.0%	
	T	4.7%	1.6%	1.8%	0.2%	
BE2	A	0.3%	0.4%	0.2%	0.0%	0.25%
	C	95.3%	97.3%	97.6%	99.8%	
	G	0.4%	0.1%	0.0%	0.0%	
	T	4.0%	2.3%	2.1%	0.1%	
BE3	A	2.4%	3.2%	2.2%	0.9%	18.88%
	C	60.3%	72.6%	73.8%	86.6%	
	G	1.2%	0.6%	0.4%	0.3%	
	T	36.2%	23.6%	23.5%	12.3%	

FANCF-UDG KO		C_6	C_7	C_8	C_{11}	插入缺失%
未处理的	A	0.0%	0.0%	0.0%	0.0%	0.02%
	C	99.9%	99.9%	100.0%	100.0%	
	G	0.0%	0.0%	0.0%	0.0%	
	T	0.1%	0.0%	0.0%	0.0%	
BE1	A	0.0%	0.1%	0.0%	0.0%	0.09%
	C	60.9%	61.1%	61.3%	94.7%	
	G	0.1%	0.1%	0.1%	0.0%	
	T	39.0%	38.8%	38.6%	5.2%	
BE2	A	0.0%	0.0%	0.0%	0.0%	0.02%
	C	88.6%	88.6%	88.7%	98.9%	
	G	0.0%	0.0%	0.0%	0.0%	
	T	11.4%	11.4%	11.3%	1.1%	
BE3	A	0.0%	0.1%	0.1%	0.0%	0.35%
	C	47.9%	48.5%	48.8%	55.8%	
	G	0.0%	0.1%	0.3%	0.2%	
	T	52.1%	51.3%	50.8%	44.0%	

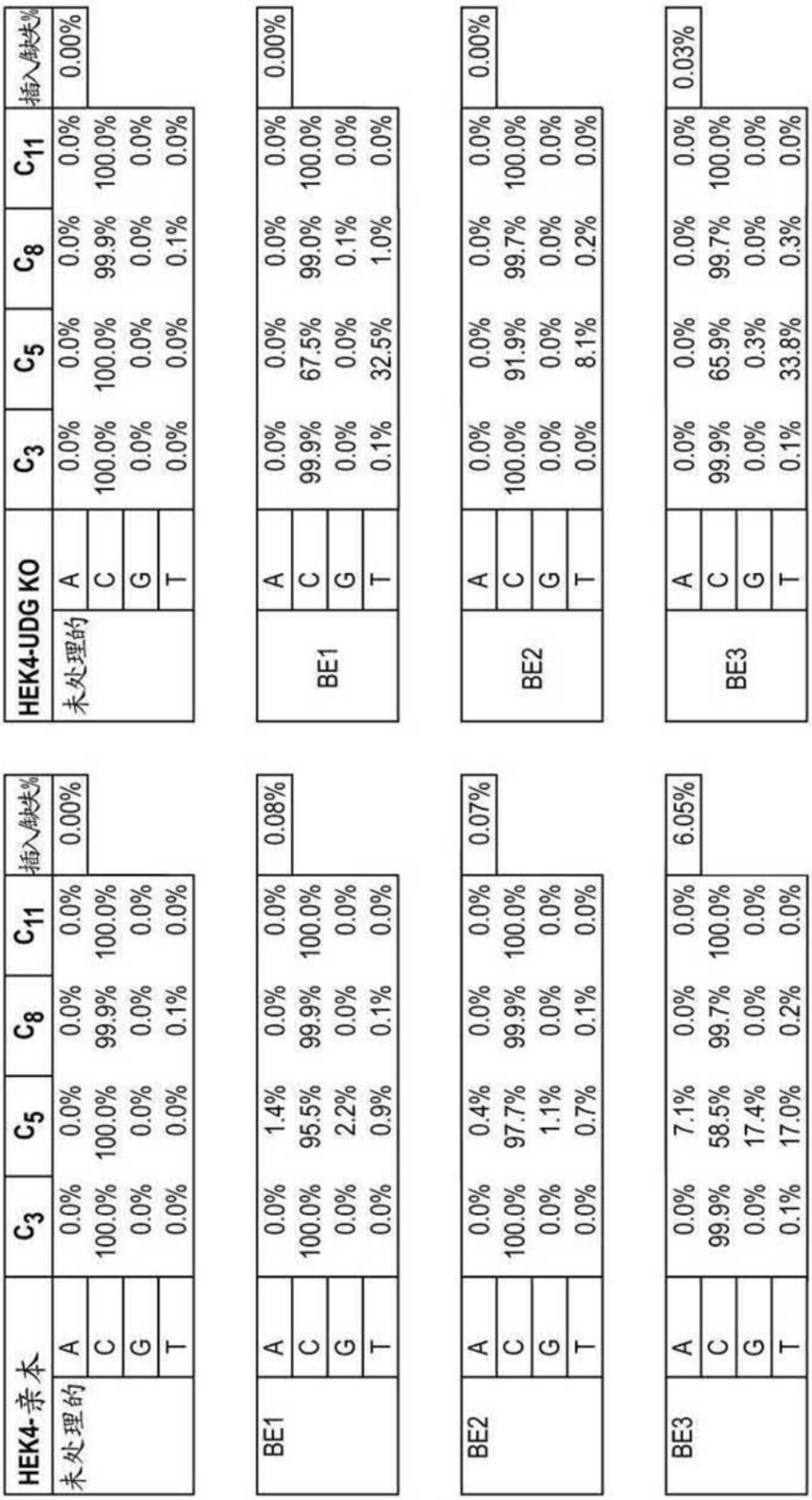
图90续



HEK3-亲本		C_3	C_4	C_5	C_9	插入缺失%
未处理的	A	0.0%	0.0%	0.0%	0.0%	0.00%
	C	100.0%	99.9%	99.9%	100.0%	
	G	0.0%	0.0%	0.0%	0.0%	
	T	0.0%	0.0%	0.0%	0.0%	
BE1	A	0.0%	0.4%	0.3%	0.1%	0.07%
	C	99.9%	96.3%	94.4%	99.8%	
	G	0.0%	0.1%	1.9%	0.1%	
	T	0.1%	3.2%	3.4%	0.1%	
BE2	A	0.0%	0.2%	0.3%	0.1%	0.05%
	C	100.0%	97.7%	96.1%	99.8%	
	G	0.0%	0.0%	1.2%	0.1%	
	T	0.0%	2.1%	2.4%	0.1%	
BE3	A	0.1%	2.9%	2.4%	0.3%	3.27%
	C	99.4%	62.6%	55.9%	98.8%	
	G	0.0%	1.5%	10.4%	0.3%	
	T	0.6%	33.0%	31.3%	0.6%	

HEK3-UDG KO		C_3	C_4	C_5	C_9	插入缺失%
未处理的	A	0.0%	0.0%	0.0%	0.0%	0.00%
	C	100.0%	100.0%	100.0%	100.0%	
	G	0.0%	0.0%	0.0%	0.0%	
	T	0.0%	0.0%	0.0%	0.0%	
BE1	A	0.0%	0.0%	0.0%	0.0%	0.03%
	C	96.0%	50.3%	41.1%	96.3%	
	G	0.0%	0.0%	0.1%	0.1%	
	T	4.0%	49.6%	58.8%	3.6%	
BE2	A	0.0%	0.0%	0.0%	0.0%	0.00%
	C	99.0%	80.2%	73.2%	98.8%	
	G	0.0%	0.0%	0.0%	0.0%	
	T	1.0%	19.7%	26.8%	1.1%	
BE3	A	0.0%	0.0%	0.0%	0.0%	0.09%
	C	98.5%	55.6%	40.3%	99.4%	
	G	0.0%	0.0%	0.0%	0.0%	
	T	1.4%	44.3%	59.7%	0.6%	

图90续



HEK4-亲本		C_3	C_5	C_8	C_{11}	插入缺失%
未处理的	A	0.0%	0.0%	0.0%	0.0%	0.00%
	C	100.0%	100.0%	99.9%	100.0%	
	G	0.0%	0.0%	0.0%	0.0%	
	T	0.0%	0.0%	0.1%	0.0%	

BE1		C_3	C_5	C_8	C_{11}	插入缺失%
	A	0.0%	1.4%	0.0%	0.0%	0.08%
	C	100.0%	95.5%	99.9%	100.0%	
	G	0.0%	2.2%	0.0%	0.0%	
	T	0.0%	0.9%	0.1%	0.0%	

BE2		C_3	C_5	C_8	C_{11}	插入缺失%
	A	0.0%	0.4%	0.0%	0.0%	0.07%
	C	100.0%	97.7%	99.9%	100.0%	
	G	0.0%	1.1%	0.0%	0.0%	
	T	0.0%	0.7%	0.1%	0.0%	

BE3		C_3	C_5	C_8	C_{11}	插入缺失%
	A	0.0%	7.1%	0.0%	0.0%	6.05%
	C	99.9%	58.5%	99.7%	100.0%	
	G	0.0%	17.4%	0.0%	0.0%	
	T	0.1%	17.0%	0.2%	0.0%	

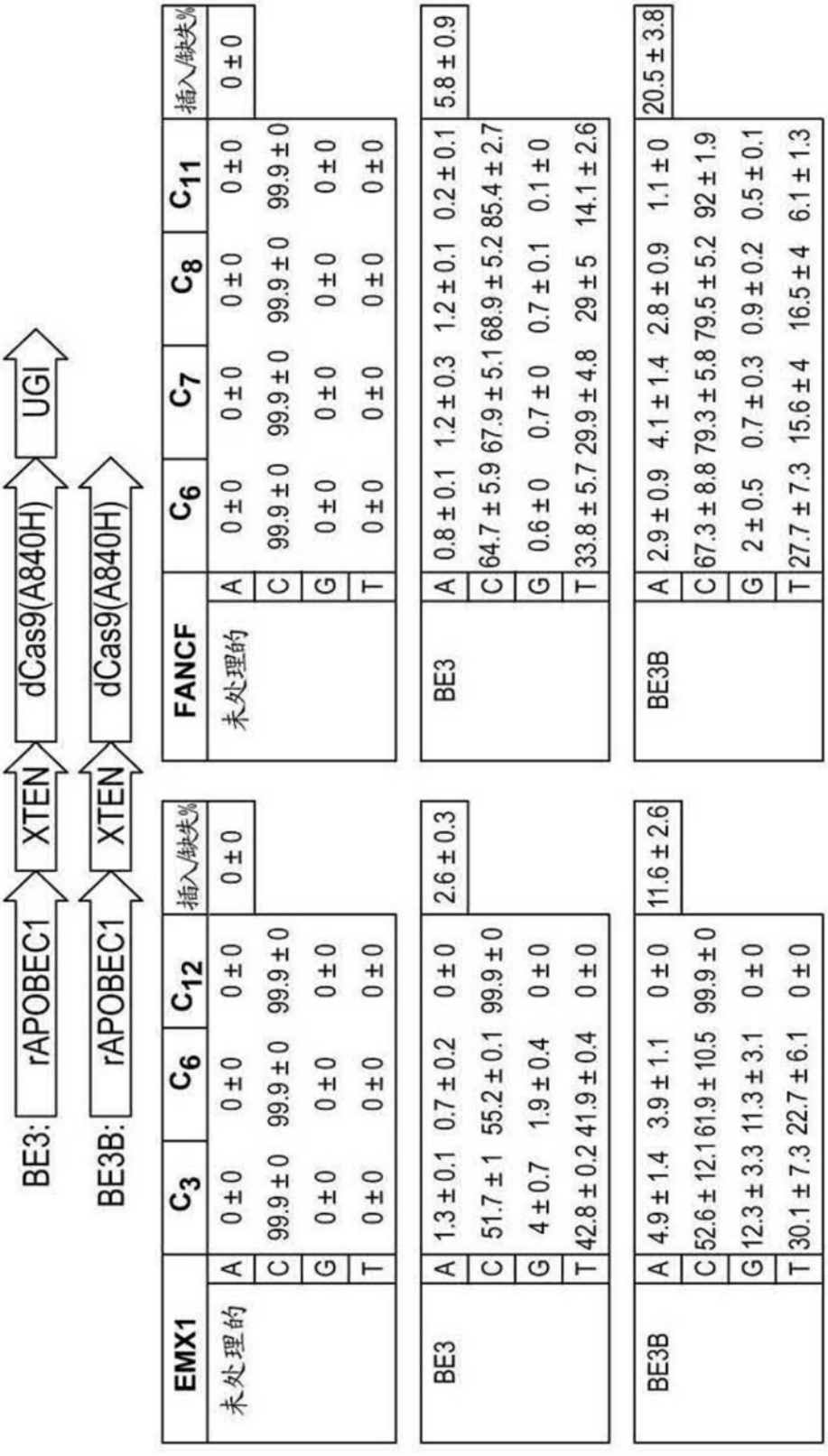
HEK4-UDG KO		C_3	C_5	C_8	C_{11}	插入缺失%
未处理的	A	0.0%	0.0%	0.0%	0.0%	0.00%
	C	100.0%	100.0%	99.9%	100.0%	
	G	0.0%	0.0%	0.0%	0.0%	
	T	0.0%	0.0%	0.1%	0.0%	

BE1		C_3	C_5	C_8	C_{11}	插入缺失%
	A	0.0%	0.0%	0.0%	0.0%	0.00%
	C	99.9%	67.5%	99.0%	100.0%	
	G	0.0%	0.0%	0.1%	0.0%	
	T	0.1%	32.5%	1.0%	0.0%	

BE2		C_3	C_5	C_8	C_{11}	插入缺失%
	A	0.0%	0.0%	0.0%	0.0%	0.00%
	C	100.0%	91.9%	99.7%	100.0%	
	G	0.0%	0.0%	0.0%	0.0%	
	T	0.0%	8.1%	0.2%	0.0%	

BE3		C_3	C_5	C_8	C_{11}	插入缺失%
	A	0.0%	0.0%	0.0%	0.0%	0.03%
	C	99.9%	65.9%	99.7%	100.0%	
	G	0.0%	0.3%	0.0%	0.0%	
	T	0.1%	33.8%	0.3%	0.0%	

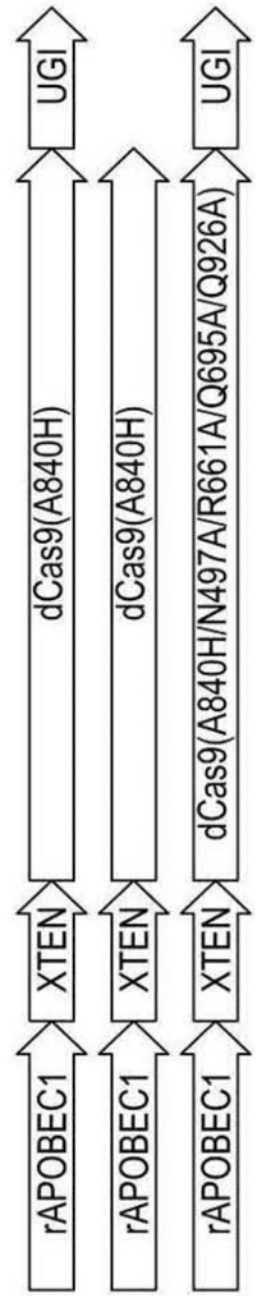
图90续



EMX1		C_3	C_6	C_{12}	插入/缺失%
未处理的	A	0 ± 0	0 ± 0	0 ± 0	0 ± 0
	C	99.9 ± 0	99.9 ± 0	99.9 ± 0	
	G	0 ± 0	0 ± 0	0 ± 0	
	T	0 ± 0	0 ± 0	0 ± 0	
BE3	A	1.3 ± 0.1	0.7 ± 0.2	0 ± 0	2.6 ± 0.3
	C	51.7 ± 1	55.2 ± 0.1	99.9 ± 0	
	G	4 ± 0.7	1.9 ± 0.4	0 ± 0	
	T	42.8 ± 0.2	41.9 ± 0.4	0 ± 0	
BE3B	A	4.9 ± 1.4	3.9 ± 1.1	0 ± 0	11.6 ± 2.6
	C	52.6 ± 12.1	61.9 ± 10.5	99.9 ± 0	
	G	12.3 ± 3.3	11.3 ± 3.1	0 ± 0	
	T	30.1 ± 7.3	22.7 ± 6.1	0 ± 0	

FANCF		C_6	C_7	C_8	C_{11}	插入/缺失%
未处理的	A	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
	C	99.9 ± 0	99.9 ± 0	99.9 ± 0	99.9 ± 0	
	G	0 ± 0	0 ± 0	0 ± 0	0 ± 0	
	T	0 ± 0	0 ± 0	0 ± 0	0 ± 0	
BE3	A	0.8 ± 0.1	1.2 ± 0.3	1.2 ± 0.1	0.2 ± 0.1	5.8 ± 0.9
	C	64.7 ± 5.9	67.9 ± 5.1	68.9 ± 5.2	85.4 ± 2.7	
	G	0.6 ± 0	0.7 ± 0	0.7 ± 0.1	0.1 ± 0	
	T	33.8 ± 5.7	29.9 ± 4.8	29 ± 5	14.1 ± 2.6	
BE3B	A	2.9 ± 0.9	4.1 ± 1.4	2.8 ± 0.9	1.1 ± 0	20.5 ± 3.8
	C	67.3 ± 8.8	79.3 ± 5.8	79.5 ± 5.2	92 ± 1.9	
	G	2 ± 0.5	0.7 ± 0.3	0.9 ± 0.2	0.5 ± 0.1	
	T	27.7 ± 7.3	15.6 ± 4	16.5 ± 4	6.1 ± 1.3	

图91



HEK3		C_6	C_7	C_8	C_{11}	插入/缺失%
未处理的	A	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
	C	99.9 ± 0	99.9 ± 0	99.9 ± 0	99.9 ± 0	
	G	0 ± 0	0 ± 0	0 ± 0	0 ± 0	
	T	0 ± 0	0 ± 0	0 ± 0	0 ± 0	
BE3	A	0 ± 0	1.6 ± 0.2	1.5 ± 0.3	0.2 ± 0	2.7 ± 0.4
	C	98.7 ± 0.2	48.6 ± 5.4	40.8 ± 6.7	98.6 ± 0.1	
	G	0 ± 0	1.4 ± 0.2	10.5 ± 1.2	0.3 ± 0	
	T	1.1 ± 0.2	48.2 ± 4.9	47 ± 5.2	0.7 ± 0.1	
BE3B	A	0.1 ± 0	4.9 ± 0.8	3.8 ± 0.8	0.8 ± 0.1	7.7 ± 1
	C	99.4 ± 0.1	57.6 ± 6.3	44.4 ± 8.4	97.7 ± 0.3	
	G	0 ± 0	3.3 ± 0.7	21.2 ± 2.8	0.7 ± 0.1	
	T	0.4 ± 0.1	33.9 ± 4.7	30.4 ± 4.7	0.6 ± 0	

HEK4		C_6	C_7	C_8	C_{11}	插入/缺失%
未处理的	A	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
	C	99.9 ± 0	99.9 ± 0	99.9 ± 0	99.9 ± 0	
	G	0 ± 0	0 ± 0	0 ± 0	0 ± 0	
	T	0 ± 0	0 ± 0	0 ± 0	0 ± 0	
BE3	A	0 ± 0	5.4 ± 1.3	0 ± 0	0 ± 0	3 ± 0.7
	C	98.7 ± 0	44.5 ± 8.2	98.9 ± 0	99.8 ± 0	
	G	0 ± 0	18.1 ± 2.3	0 ± 0	0 ± 0	
	T	1.1 ± 0.1	31.9 ± 4.8	0.9 ± 0.1	0.1 ± 0	
BE3B	A	0 ± 0	10.6 ± 2.6	0.3 ± 0	0 ± 0	7.6 ± 0.7
	C	99.6 ± 0	45 ± 9.9	99.2 ± 0.1	99.9 ± 0	
	G	0 ± 0	32.8 ± 5.8	0 ± 0	0 ± 0	
	T	0.2 ± 0	11.4 ± 1.6	0.4 ± 0.2	0 ± 0	

图91续

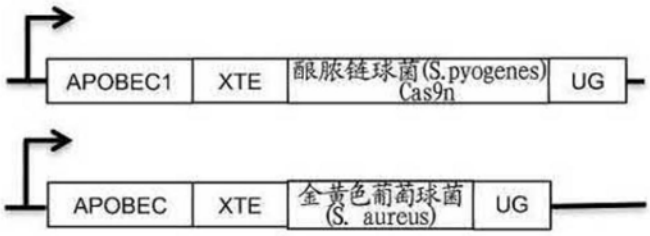


图92A

物种	PAM	碱基编辑器	参照
酿脓链球菌 (S. pyogenes)	NGG	BE3	野生型
	NGA	VQR, EQR BE3	Ref #7
	NGCG	VRER BE3	Ref #7
金黄色葡萄球菌 (S. aureus)	NNGRRT	SaBE3	野生型
	NNNRRT	SaKKHBE3	Ref #8

图92B

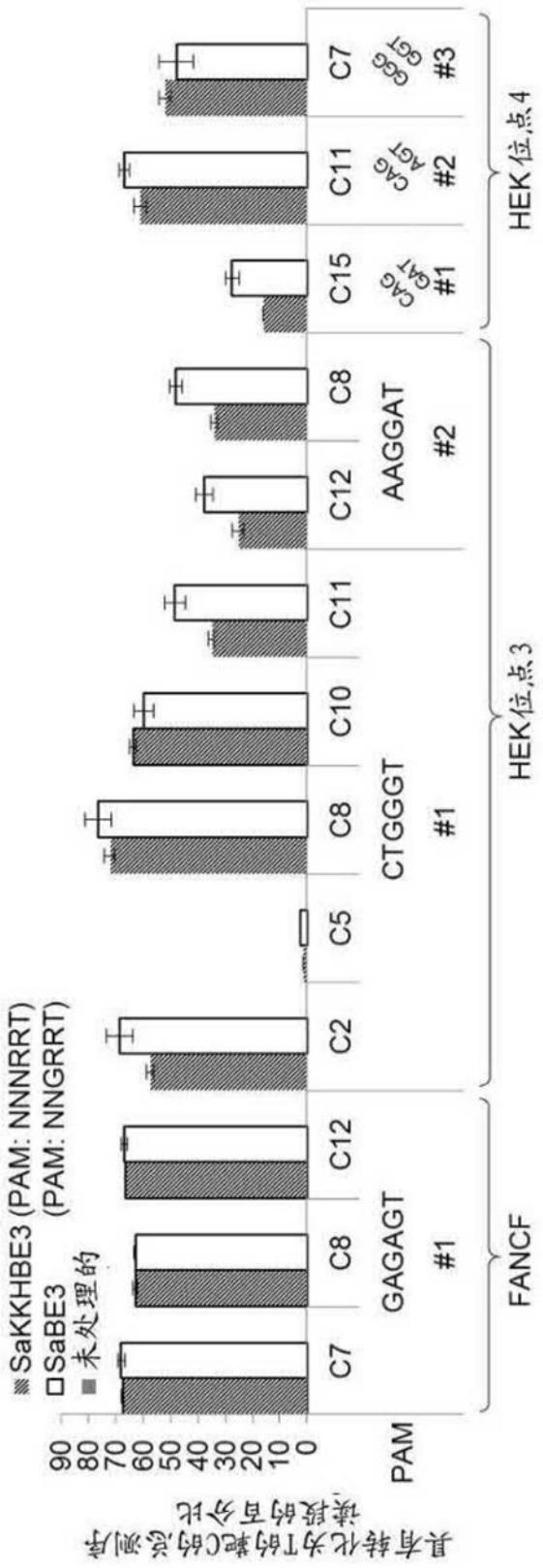

具有转化为T的靶C的总测序读段的百分比
90
80
70
60
50
40
30
20
10
0
SaKKHBE3 (PAM: NNNRRT)
SaBE3 (PAM: NNGRRT)
未处理的
PAM
C7
C8
C12
C2
C5
C8
C10
C11
C12
C8
C15
C11
C7
GAGAGT
#1
CTGGGT
#1
AAGGAT
#2
CAG GAT #1
CAG AGT #2
GGG GGT #3
FANCF
HEK位点3
HEK位点4

图92C

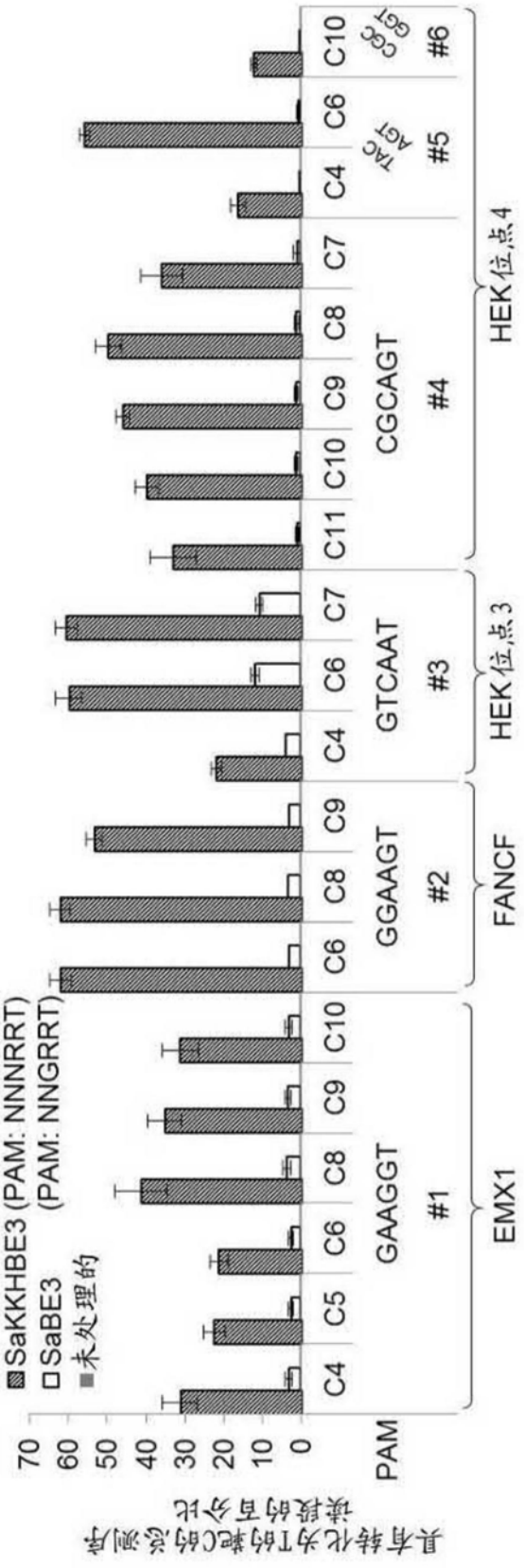

SaKKHBE3 (PAM: NNNRRT)
SaBE3 (PAM: NNGRRT)
未处理的
具有转化为T的靶C的总测序读段的百分比
70
60
50
40
30
20
10
0
PAM
C4
C5
C6
C8
C9
C10
GAAGGT
#1
EMX1
C6
C8
C9
GGAAGT
#2
FANCF
C4
C6
C7
GTCAAT
#3
HEK位点3
C7
C8
C9
C10
C11
CGCAGT
#4
HEK位点4
C4
C6
TAC AGT
#5
C10
GGT CG
#6

图92D

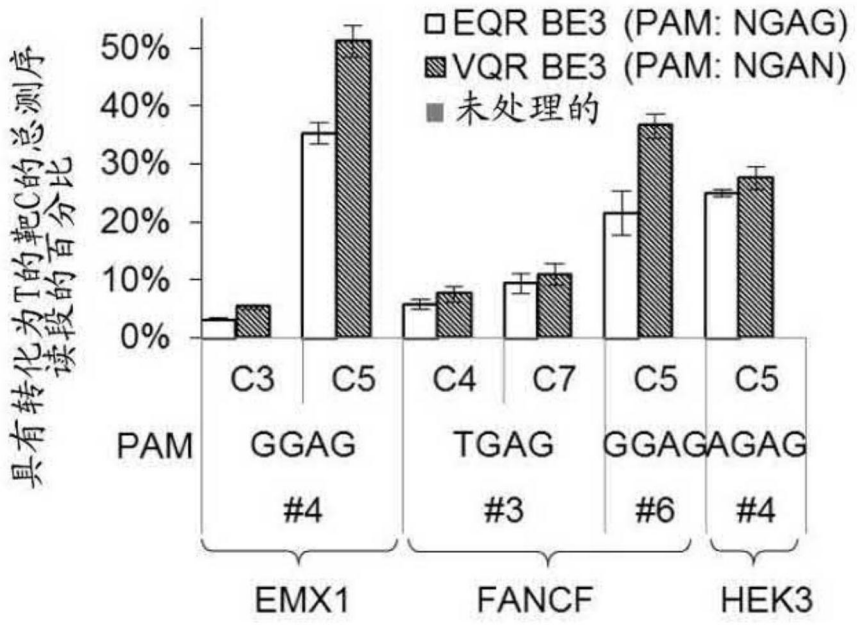

EQR BE3 (PAM: NGAG)
VQR BE3 (PAM: NGAN)
未处理的
具有转化为T的靶C的总测序读段的百分比
50%
40%
30%
20%
10%
0%
C3
C5
C4
C7
C5
C5
PAM
GGAG
TGAG
GGAG
AGAG
#4
#3
#6
#4
EMX1
FANCF
HEK3

图92E

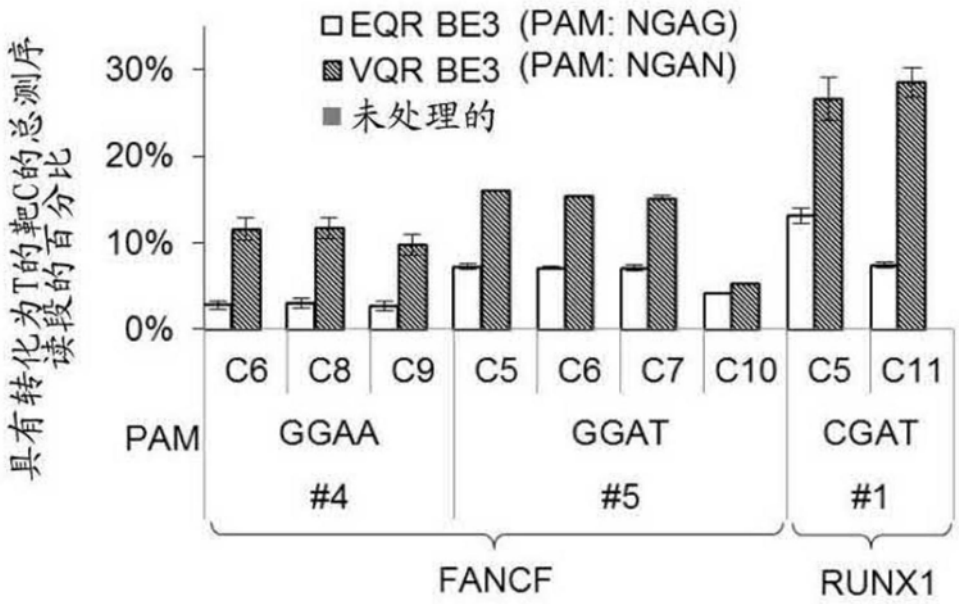

EQR BE3 (PAM: NGAG)
VQR BE3 (PAM: NGAN)
未处理的
具有转化为T的靶C的总测序读段的百分比
30%
20%
10%
0%
C6
C8
C9
C5
C6
C7
C10
C5
C11
PAM
GGAA
GGAT
CGAT
#4
#5
#1
FANCF
RUNX1

图92F

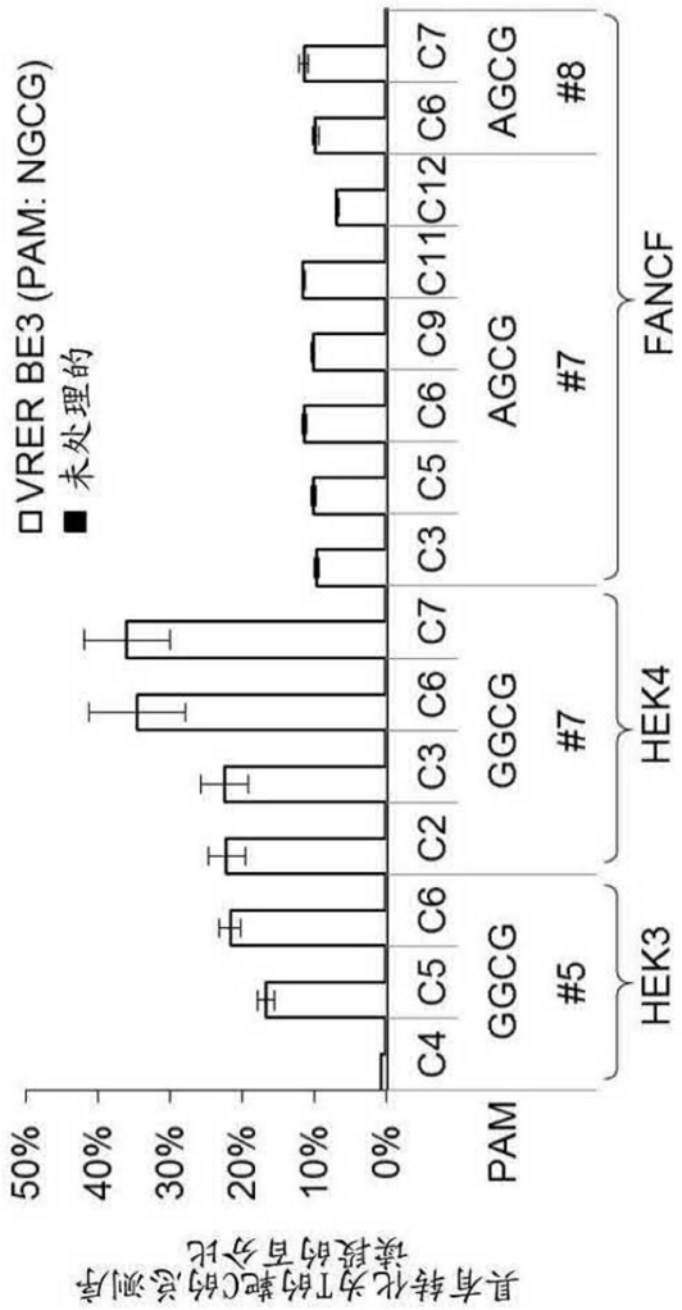

VRER BE3 (PAM: NGCG)
未处理的
具有转化为T的靶C的总测序
读段的百分比
50%
40%
30%
20%
10%
0%
PAM
C4 C5 C6 C2 C3 C6 C7 C3 C5 C6 C9 C11 C12 C6 C7
GGCG #5 HEK3
GGCG #7 HEK4
AGCG #7 FANCF
AGCG #8

图92G

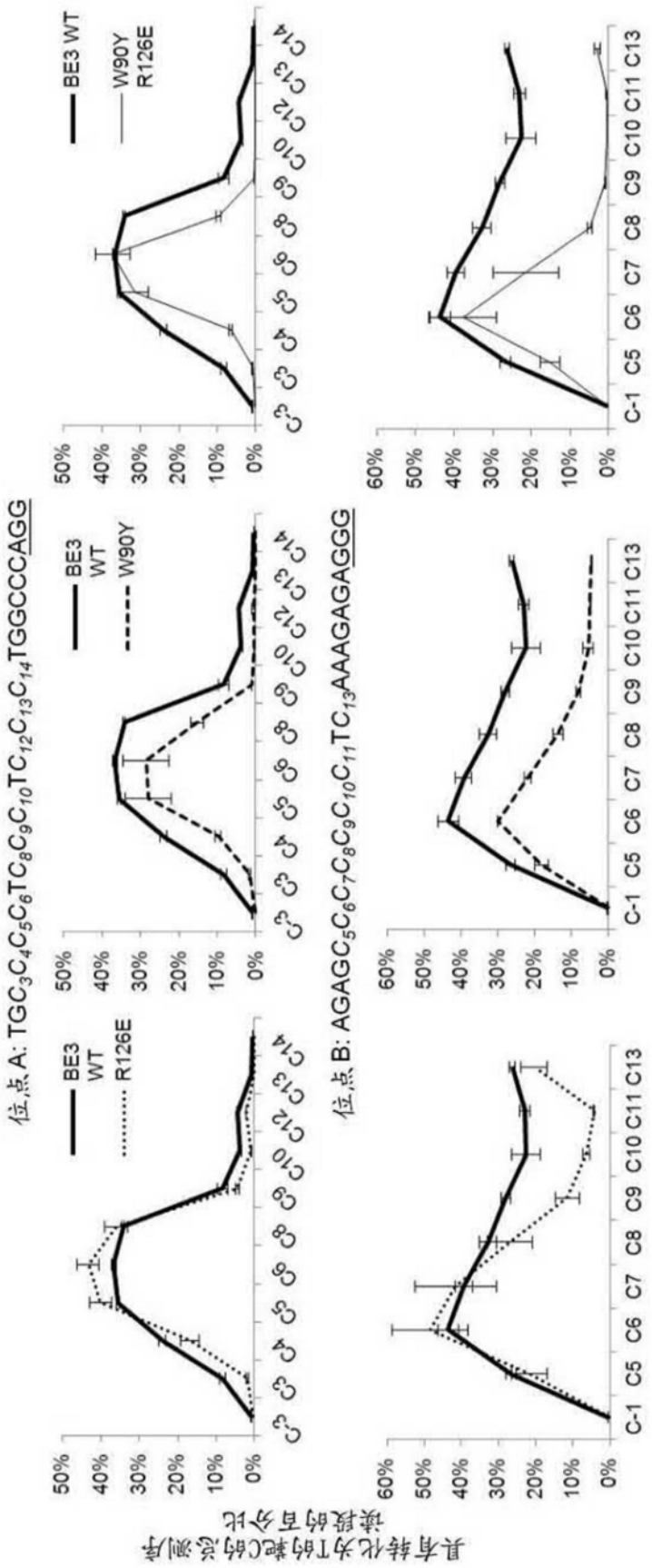

位点A: TGC3C4C5C6TC8C9C10TC12C13C14TGGCCCAGG
位点B: AGAGC5C6C7C8C9C10C11TC13AAAGAGAGGG
具有转化为T的靶C的总测序
读段的百分比
BE3 WT
W90Y
R126E
BE3
WT
R126E
BE3
WT
W90Y
50%
40%
30%
20%
10%
0%
60%
C-1 C5 C6 C7 C8 C9 C10 C11 C13
C3 C4 C12 C14

图93A

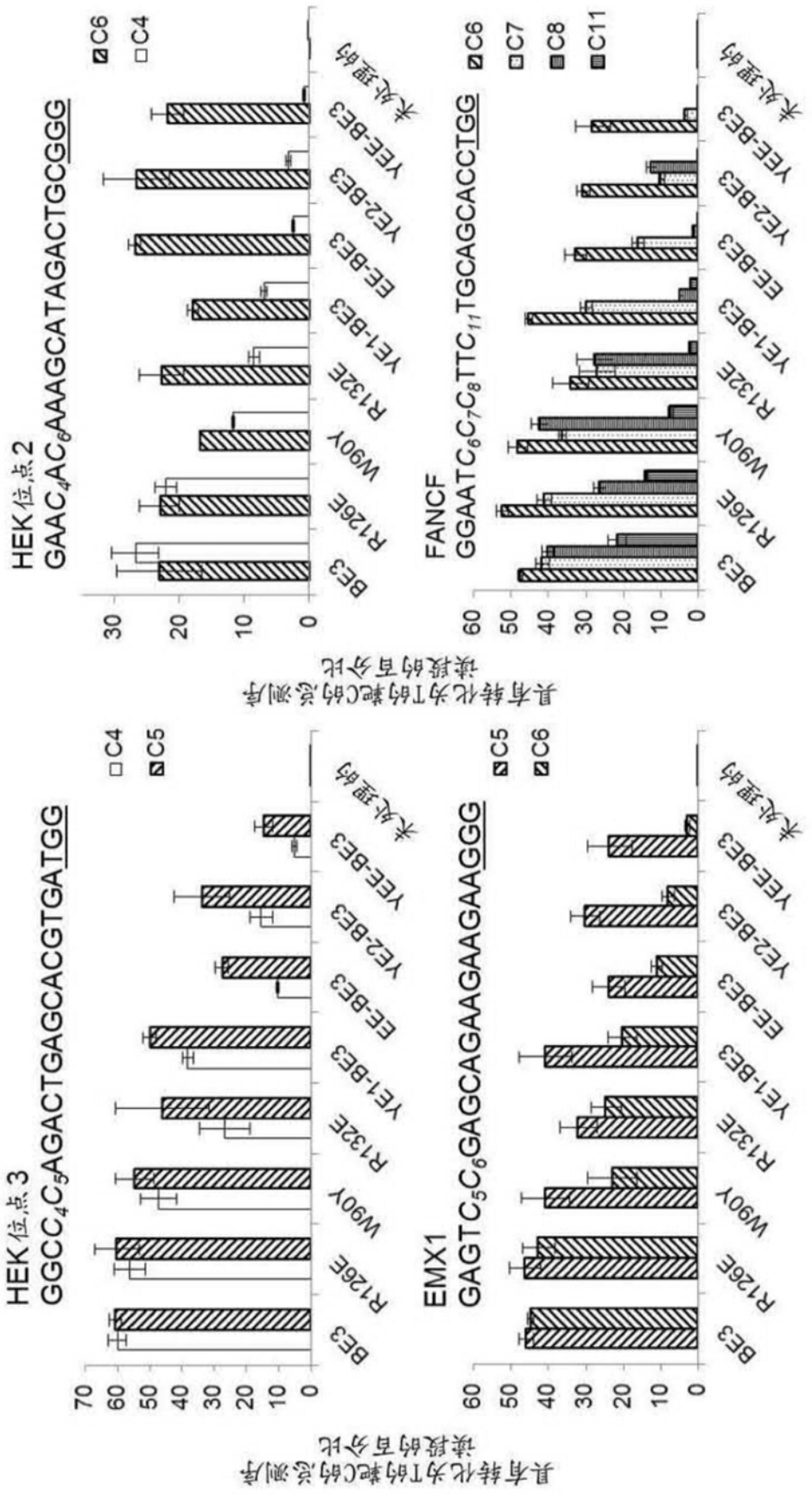

HEK位点2
GAAC4AC6AAAGCATAGACTGCGGG
C6
C4
BE3
R126E
W90Y
R132E
YE1-BE3
EE-BE3
YE2-BE3
YEE-BE3
未处理的
具有转化为T的靶C的总测序
读段的百分比
FANCF
GGAATC6C7C8TTC11TGCAGCACCTGG
C6
C7
C8
C11
HEK位点3
GGCC4C5AGACTGAGCACGTGATGG
C4
C5
EMX1
GAGTC5C6GAGCAGAAGAAGAAGGG
C5
C6

图93B

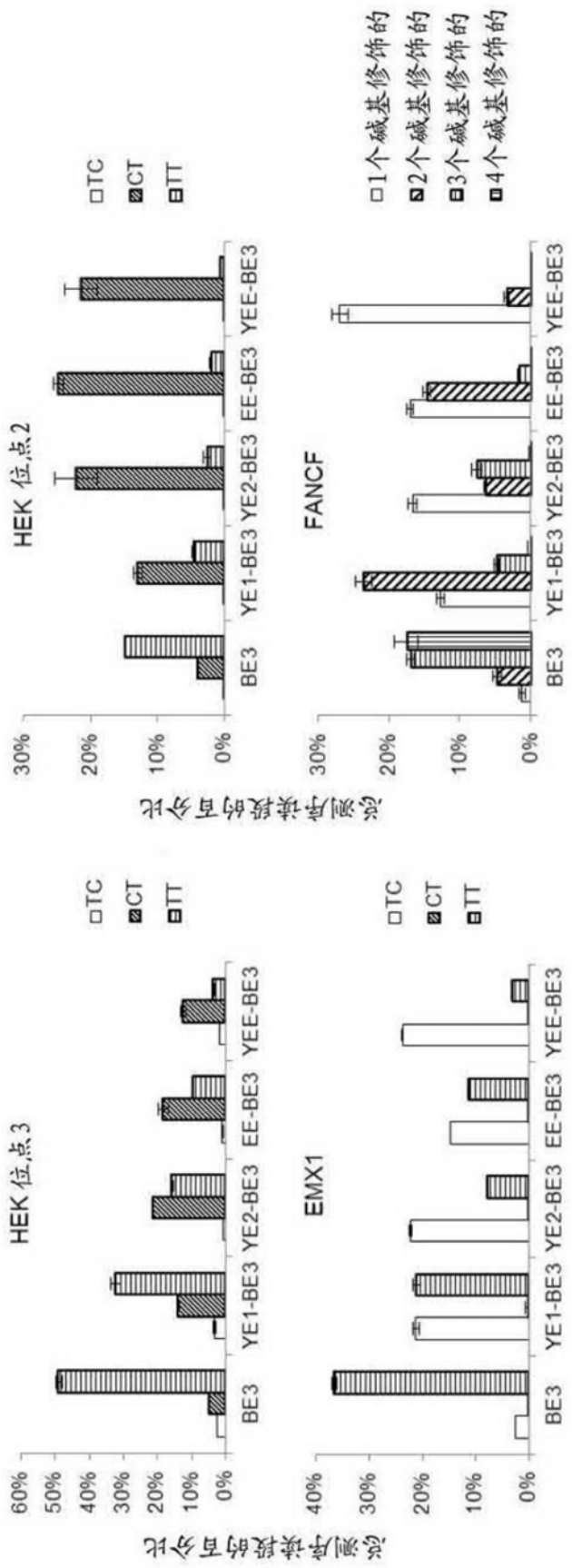

HEK 位点2
FANCF
HEK 位点3
EMX1
□TC
▨CT
▪TT
□1个碱基修饰的
▨2个碱基修饰的
▪3个碱基修饰的
▪4个碱基修饰的
BE3
YE1-BE3
YE2-BE3
EE-BE3
YEE-BE3
总测序读段的百分比
0%
10%
20%
30%
40%
50%
60%

图93C

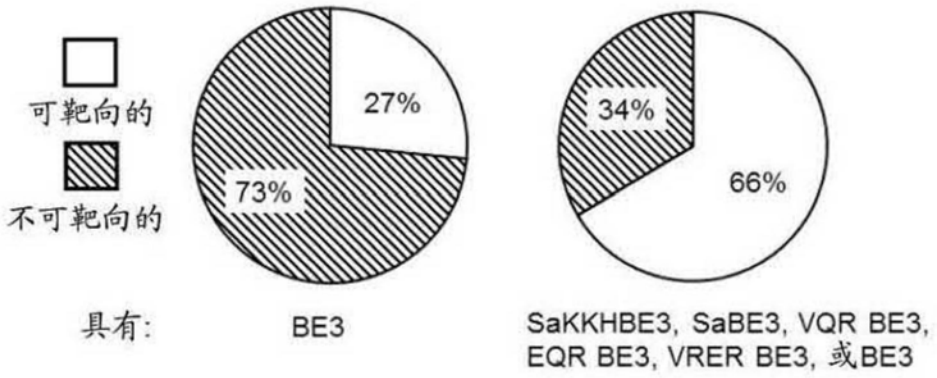

可靶向的
不可靶向的
27%
73%
34%
66%
具有:
BE3
SaKKHBE3, SaBE3, VQR BE3,
EQR BE3, VRER BE3, 或BE3

图94A

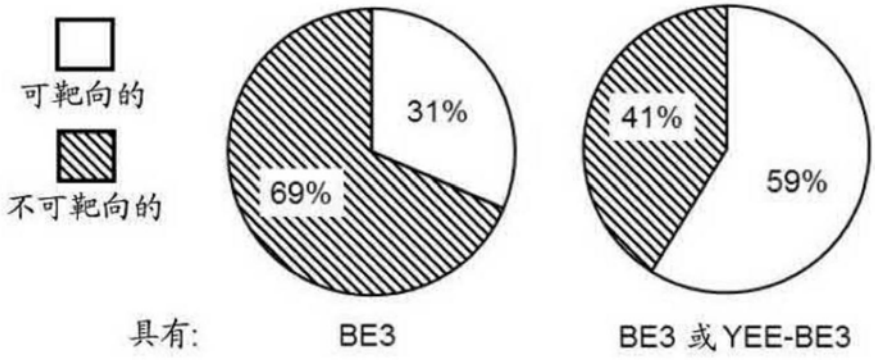

可靶向的
不可靶向的
31%
69%
41%
59%
具有:
BE3
BE3 或YEE-BE3

图94B

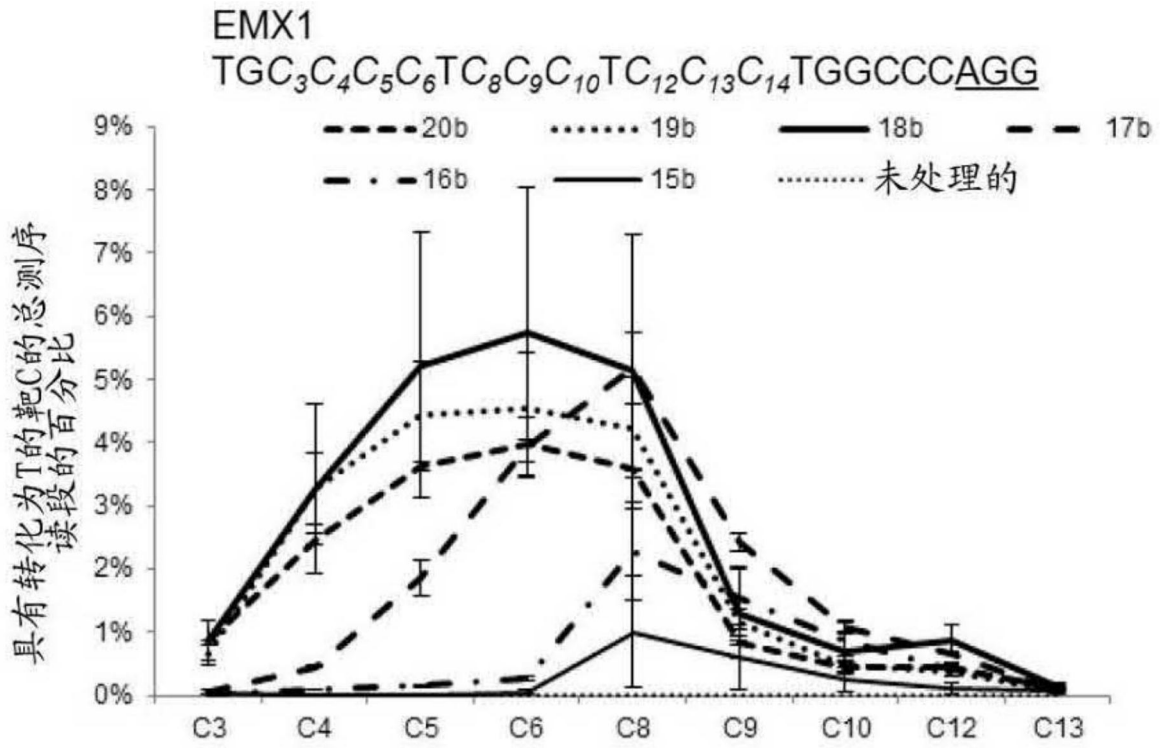

EMX1
TGC3C4C5C6TC8C9C10TC12C13C14TGGCCCAGG
20b
19b
18b
17b
16b
15b
未处理的
具有转化为T的靶C的总测序读段的百分比
9%
8%
7%
6%
5%
4%
3%
2%
1%
0%
C3
C4
C5
C6
C8
C9
C10
C12
C13

图95A

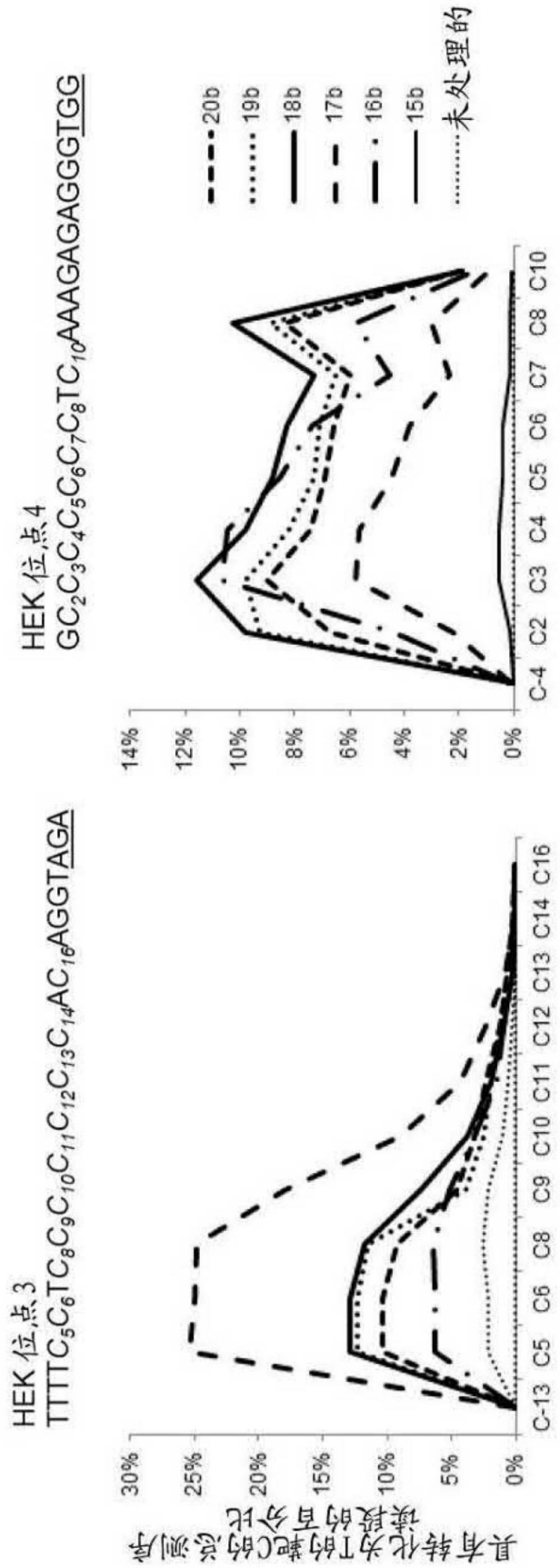

HEK 位点 3
TTTTC5C6TC8C9C10C11C12C13C14AC16AGGTAGA
HEK 位点 4
GC2C3C4C5C6C7C8TC10AAAGAGAGGGTGG
具有转化为T的靶C的总测序读段的百分比
30%
25%
20%
15%
10%
5%
0%
C-13 C5 C6 C8 C9 C10 C11 C12 C13 C14 C16
14%
12%
10%
8%
6%
4%
2%
0%
C-4 C2 C3 C4 C5 C6 C7 C8 C10
20b
19b
18b
17b
16b
15b
未处理的

图95B

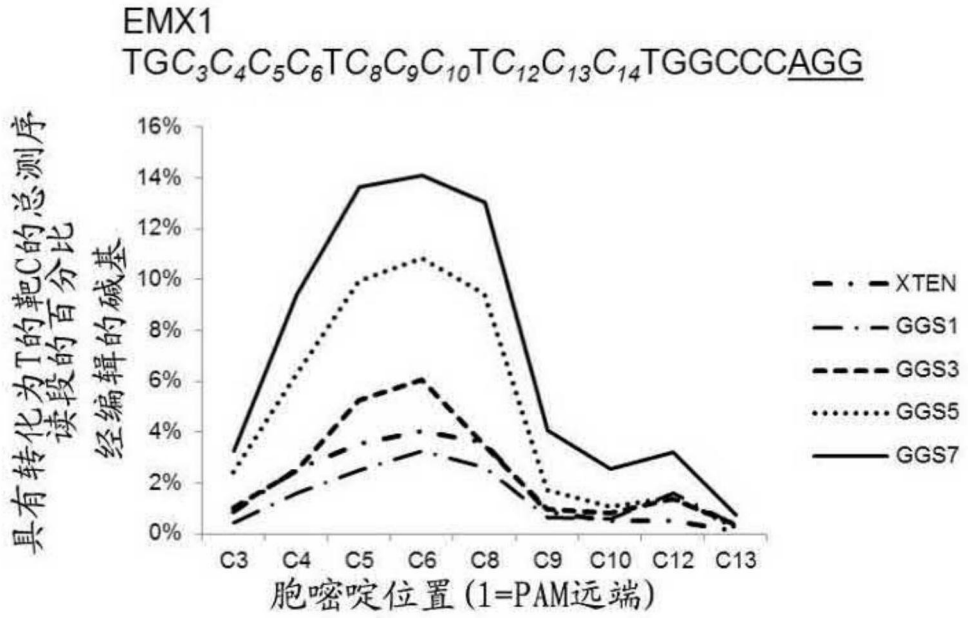

EMX1
TGC3C4C5C6TC8C9C10TC12C13C14TGGCCCAGG
具有转化为T的靶C的总测序读段的百分比
经编辑的碱基
16%
14%
12%
10%
8%
6%
4%
2%
0%
C3
C4
C5
C6
C8
C9
C10
C12
C13
胞嘧啶位置(1=PAM远端)
XTEN
GGS1
GGS3
GGS5
GGS7

图96

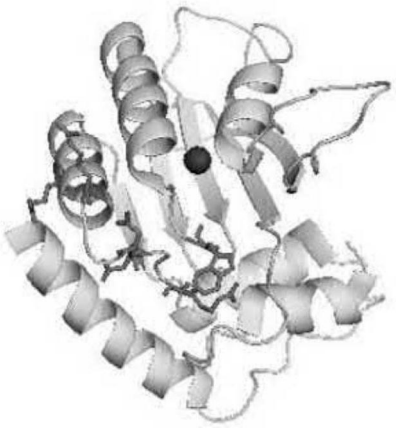


图97A

APOBEC1 突变	APOBEC3G 突变	参照
R126A	R320A	#9,10
R126E	R320E	#9,10
W90A	W285A	#9,10
W90Y	W285Y	此工作
R132E	R326E	此工作

图97B

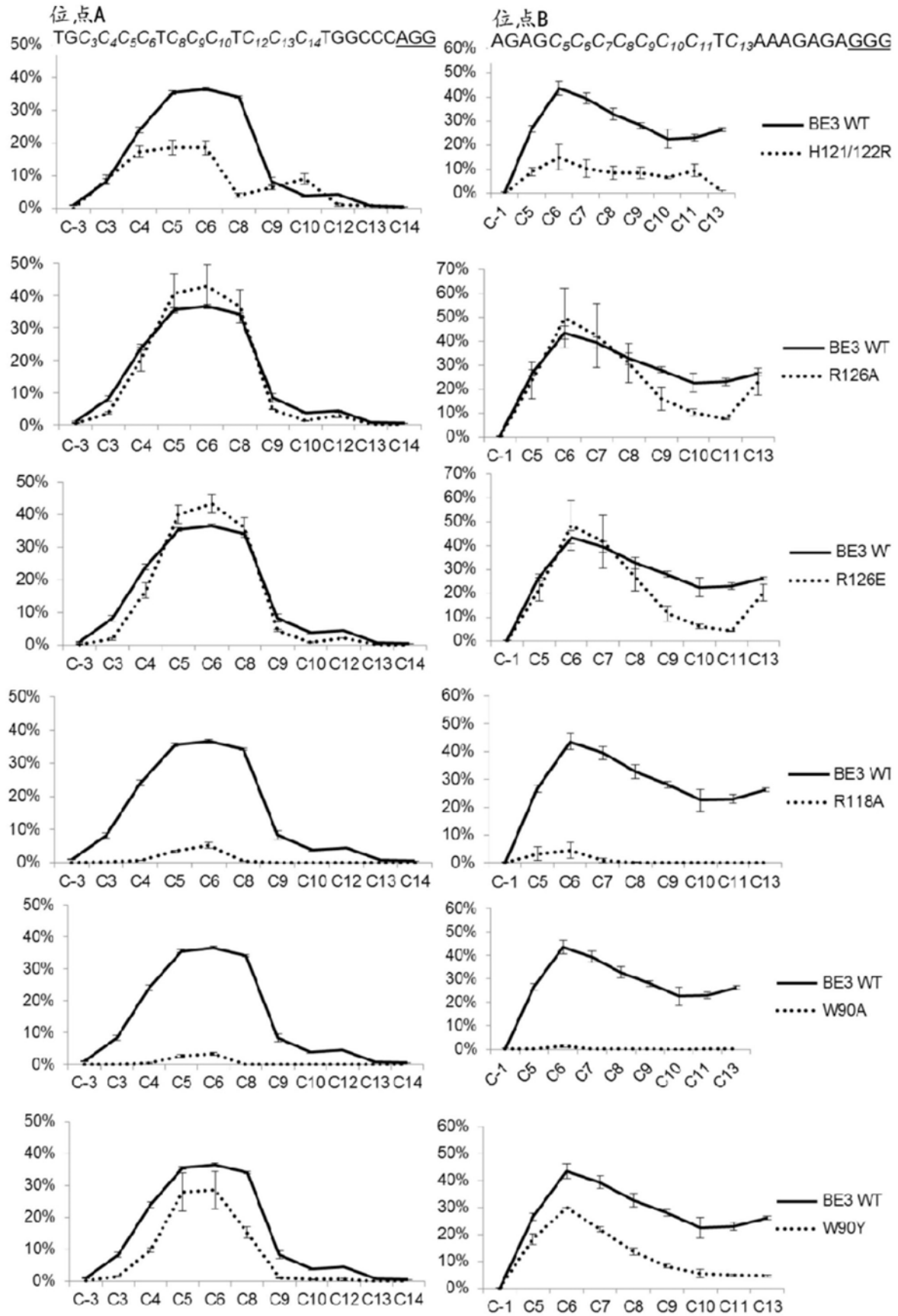
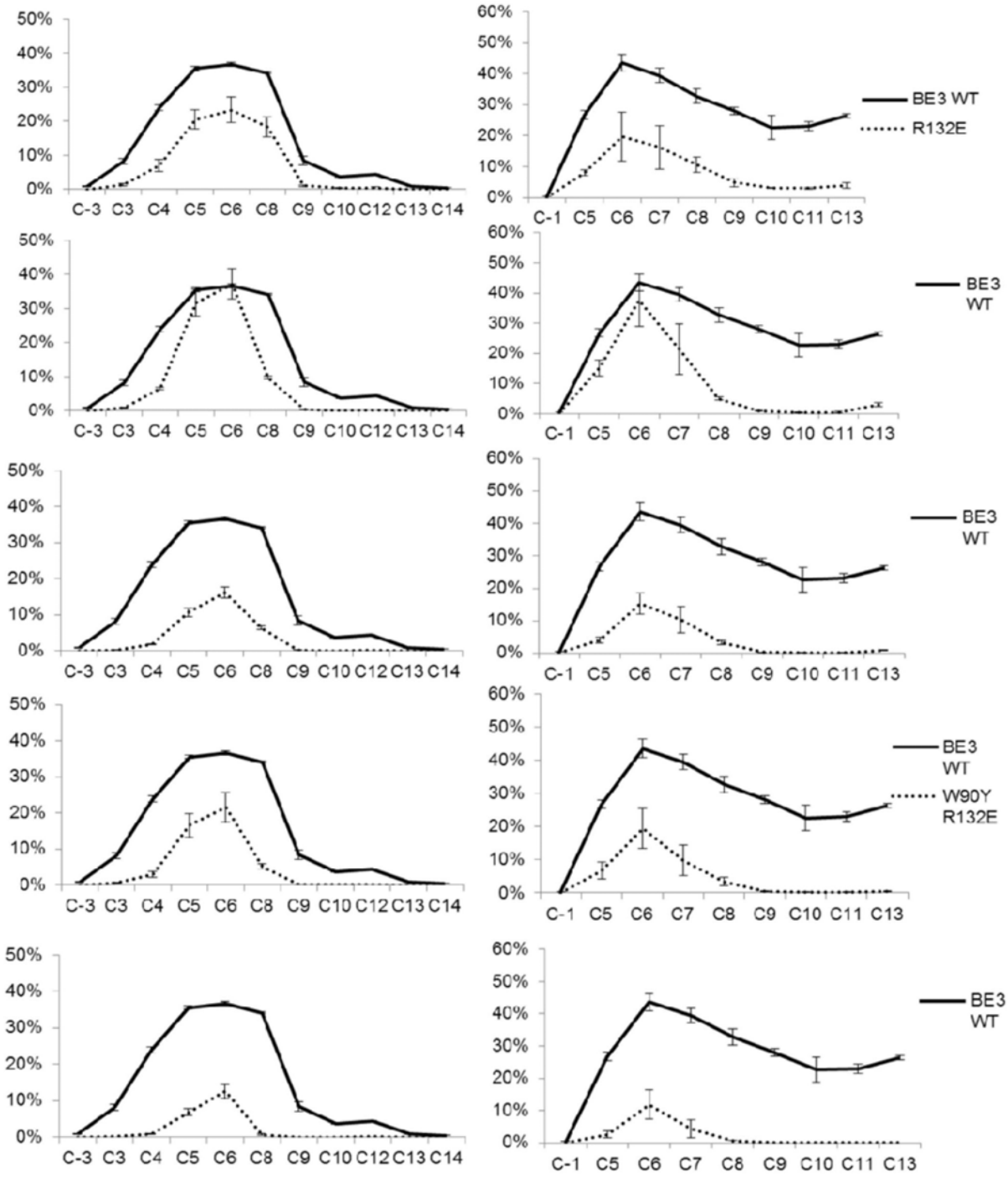

位点A
TGC3C4C5C6TC8C9C10TC12C13C14TGGCCCAGG
位点B
AGAGC5C6C7C8C9C10C11TC13AAAGAGAGGG
BE3 WT
H121/122R
R126A
R126E
R118A
W90A
W90Y
C-3 C3 C4 C5 C6 C8 C9 C10 C12 C13 C14
C-1 C5 C6 C7 C8 C9 C10 C11 C13

图97C



编辑窗宽度

	位点A	位点B		位点A	位点B
BE3 WT	5.0	6.1	W90Y	3.8	4.9
H121/122R	4.2	7.4	R132E	4.0	3.0
R126A	4.4	3.4	W90Y R126E	2.9	3.0
R126E	4.2	3.1	R126E R132E	2.9	3.0
R118A	2.4	3.6	W90Y R132E	2.7	2.8
W90A	2.5	1.1	W90Y R126E R132E	2.1	1.4

图97C续

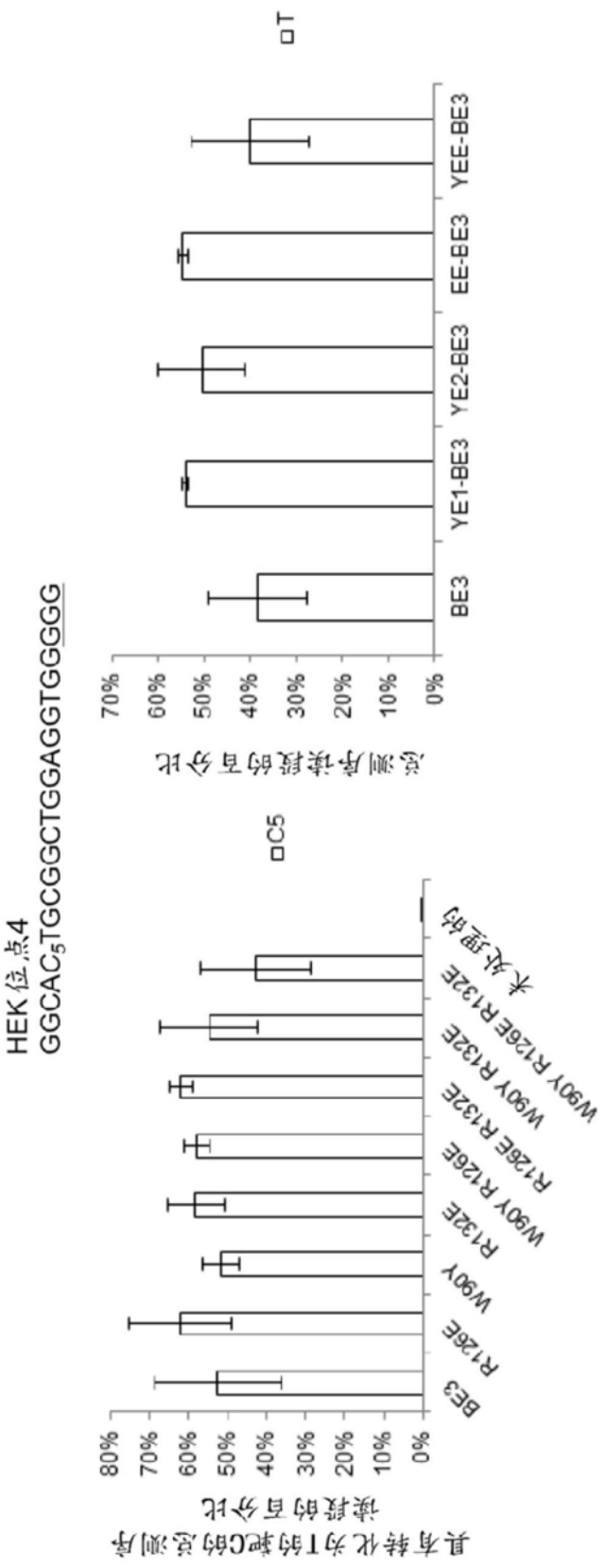

HEK位点4
GGCAC5TGCGGCTGGAGGTGGGGG
□C5
□T
具有转化为T的靶C的总测序读段的百分比
总测序读段的百分比
0%
10%
20%
30%
40%
50%
60%
70%
80%
BE3
R126E
W90Y
R132E
W90Y R126E
R126E R132E
W90Y R132E
W90Y R126E R132E
未处理的
BE3
YE1-BE3
YE2-BE3
EE-BE3
YEE-BE3

图98

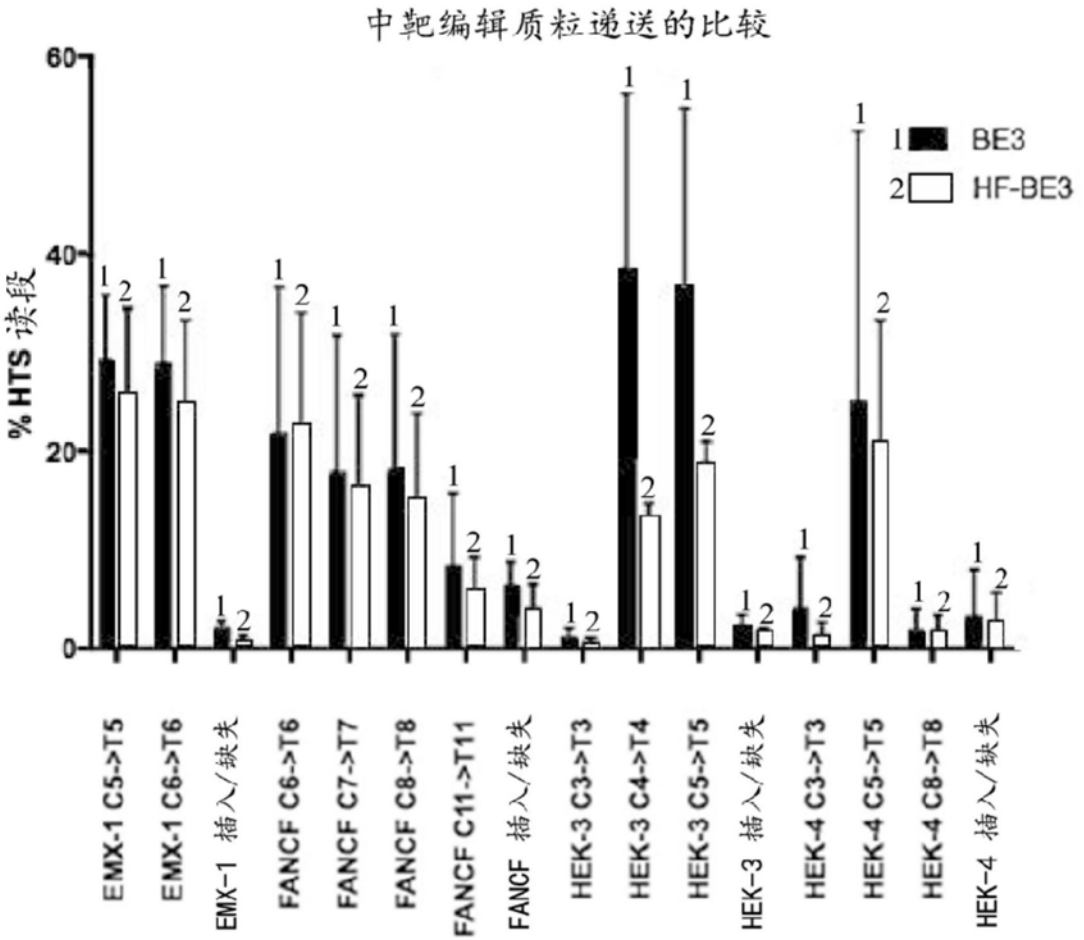

中靶编辑质粒递送的比较
% HTS 读段
60
40
20
0
1 BE3
2 HF-BE3
EMX-1 C5->T5
EMX-1 C6->T6
EMX-1 插入/缺失
FANCF C6->T6
FANCF C7->T7
FANCF C8->T8
FANCF C11->T11
FANCF 插入/缺失
HEK-3 C3->T3
HEK-3 C4->T4
HEK-3 C5->T5
HEK-3 插入/缺失
HEK-4 C3->T3
HEK-4 C5->T5
HEK-4 C8->T8
HEK-4 插入/缺失

图99

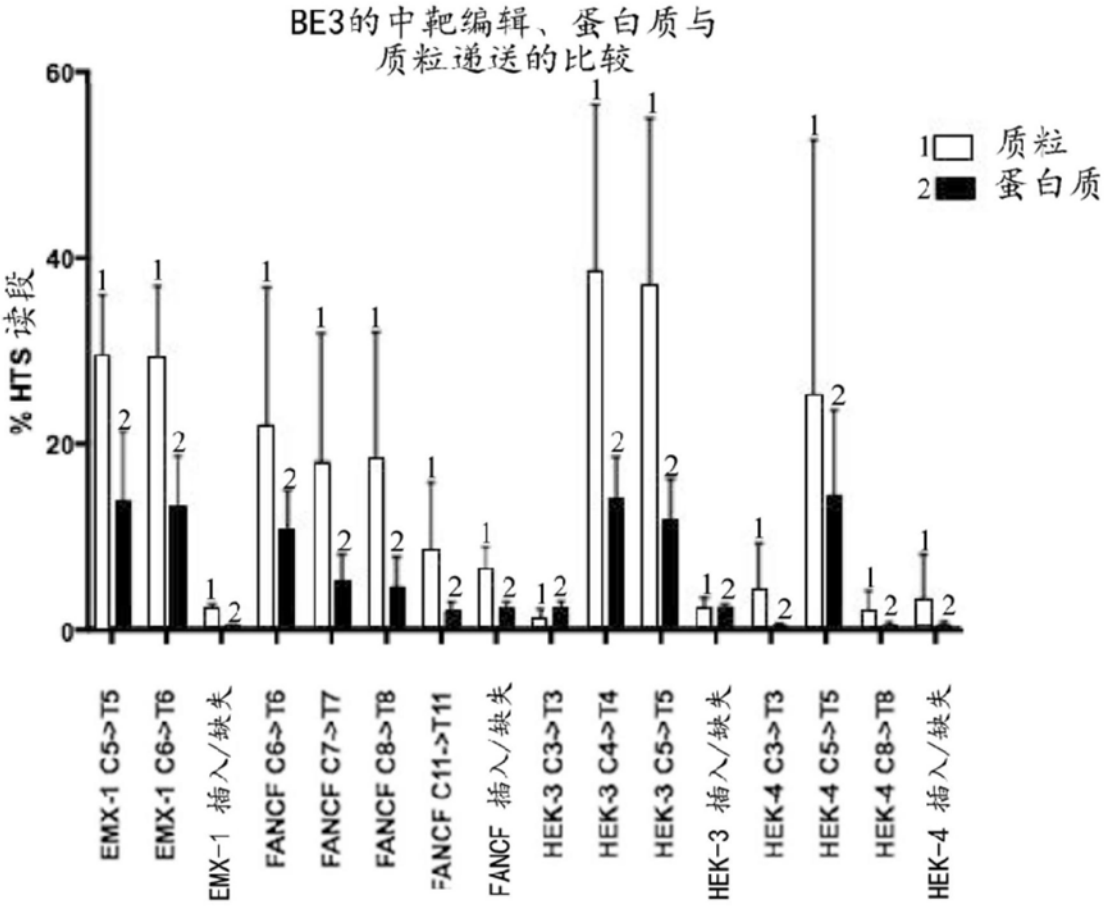

BE3的中靶编辑、蛋白质与
质粒递送的比较
% HTS 读段
60
40
20
0
1 质粒
2 蛋白质
EMX-1 C5->T5
EMX-1 C6->T6
EMX-1 插入/缺失
FANCF C6->T6
FANCF C7->T7
FANCF C8->T8
FANCF C11->T11
FANCF 插入/缺失
HEK-3 C3->T3
HEK-3 C4->T4
HEK-3 C5->T5
HEK-3 插入/缺失
HEK-4 C3->T3
HEK-4 C5->T5
HEK-4 C8->T8
HEK-4 插入/缺失

图100

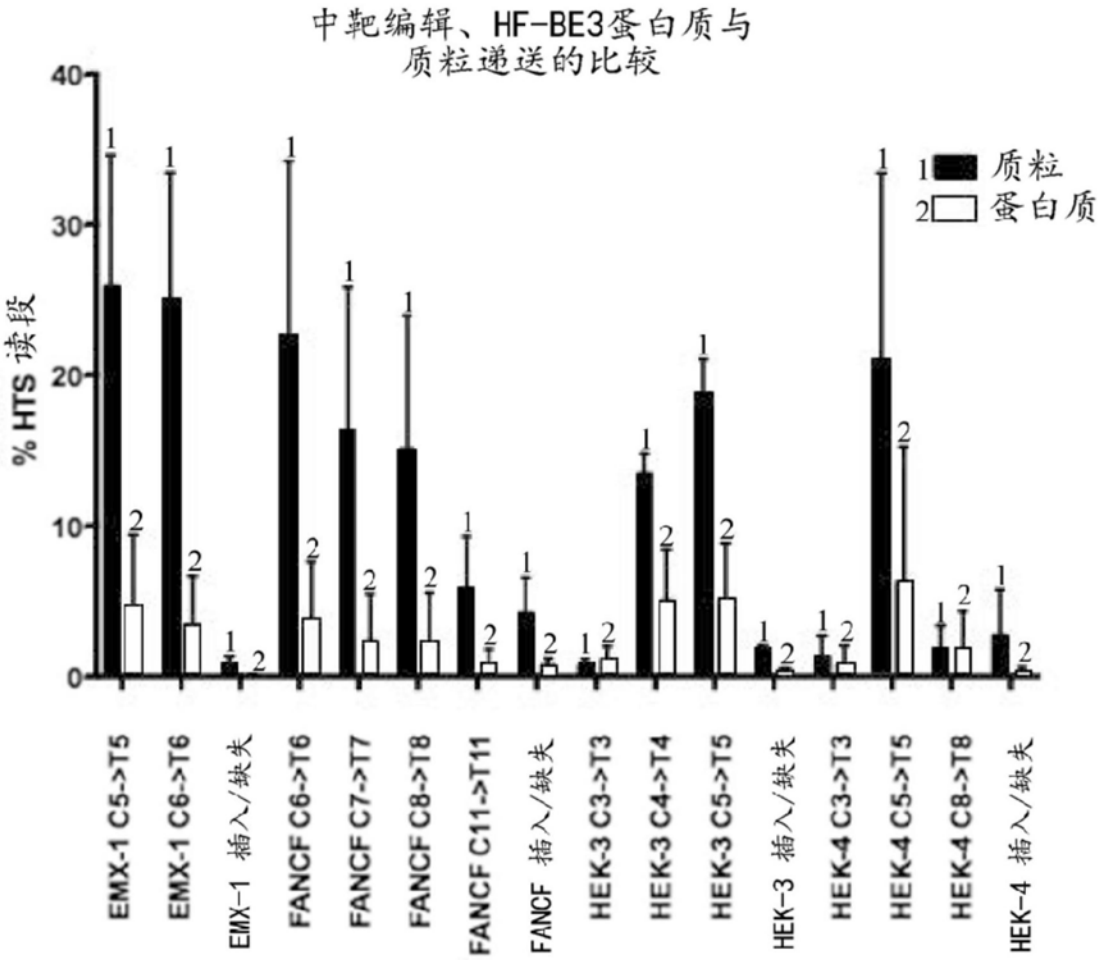

中靶编辑、HF-BE3蛋白质与
质粒递送的比较
% HTS 读段
40
30
20
10
0
1 质粒
2 蛋白质
EMX-1 C5->T5
EMX-1 C6->T6
EMX-1 插入/缺失
FANCF C6->T6
FANCF C7->T7
FANCF C8->T8
FANCF C11->T11
FANCF 插入/缺失
HEK-3 C3->T3
HEK-3 C4->T4
HEK-3 C5->T5
HEK-3 插入/缺失
HEK-4 C3->T3
HEK-4 C5->T5
HEK-4 C8->T8
HEK-4 插入/缺失

图101

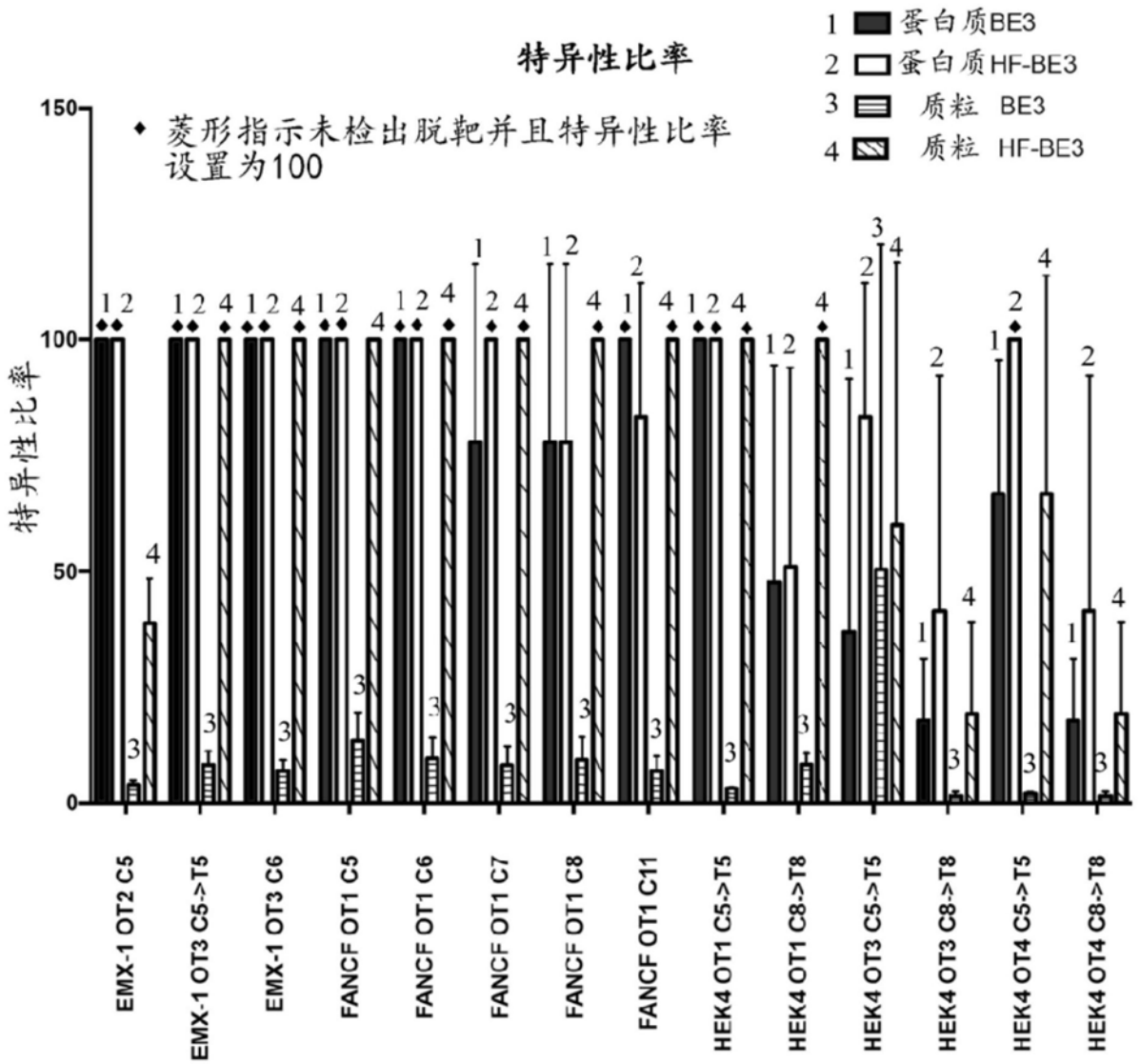

特异性比率
1 蛋白质BE3
2 蛋白质HF-BE3
3 质粒 BE3
4 质粒 HF-BE3
◆ 菱形指示未检出脱靶并且特异性比率设置为100
特异性比率
150
100
50
0
EMX-1 OT2 C5
EMX-1 OT3 C5->T5
EMX-1 OT3 C6
FANCF OT1 C5
FANCF OT1 C6
FANCF OT1 C7
FANCF OT1 C8
FANCF OT1 C11
HEK4 OT1 C5->T5
HEK4 OT1 C8->T8
HEK4 OT3 C5->T5
HEK4 OT3 C8->T8
HEK4 OT4 C5->T5
HEK4 OT4 C8->T8

图102

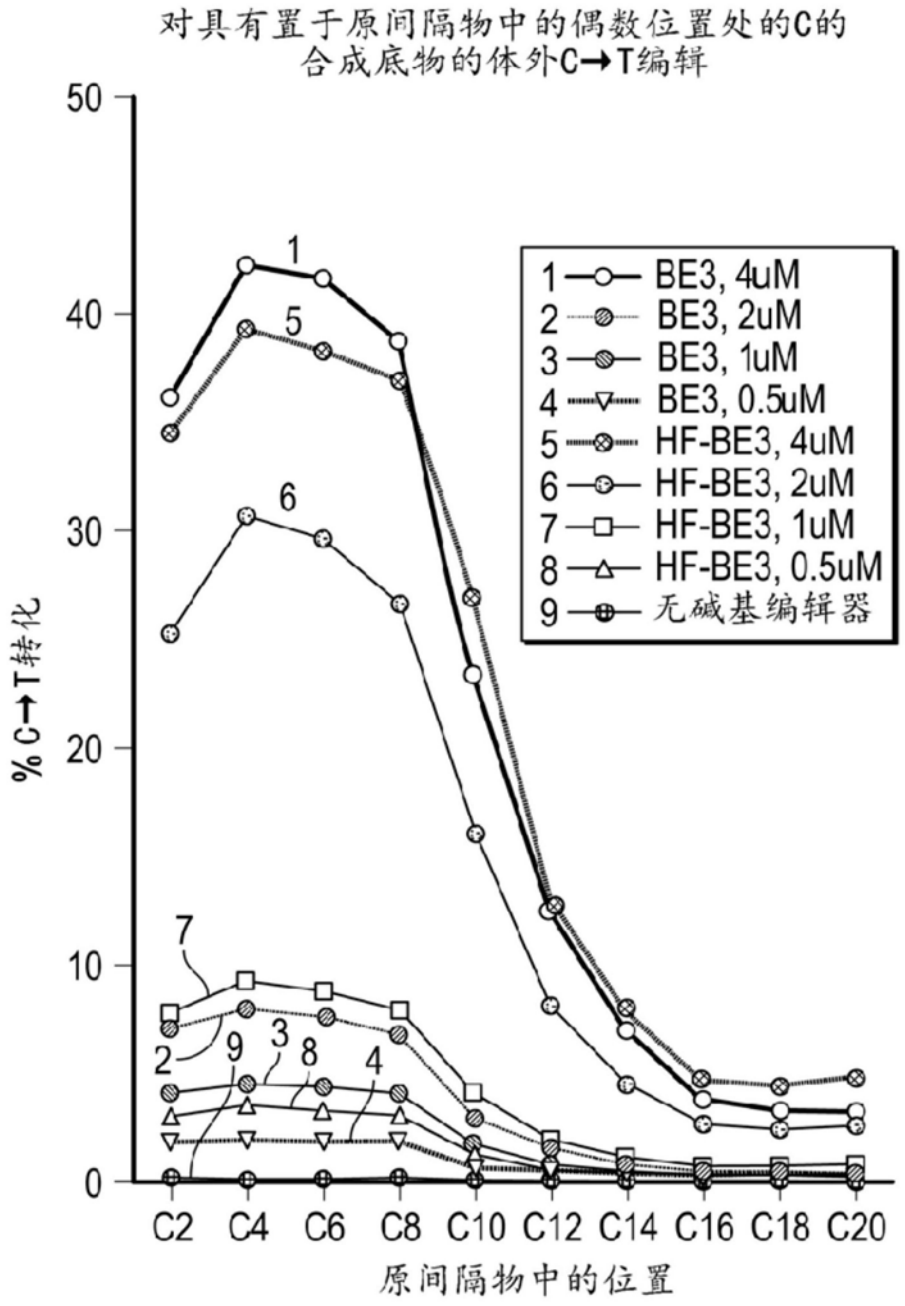

对具有置于原间隔物中的偶数位置处的C的
合成底物的体外C→T编辑
% C→T转化
50
40
30
20
10
0
C2
C4
C6
C8
C10
C12
C14
C16
C18
C20
原间隔物中的位置
1 BE3, 4uM
2 BE3, 2uM
3 BE3, 1uM
4 BE3, 0.5uM
5 HF-BE3, 4uM
6 HF-BE3, 2uM
7 HF-BE3, 1uM
8 HF-BE3, 0.5uM
9 无碱基编辑器

图103

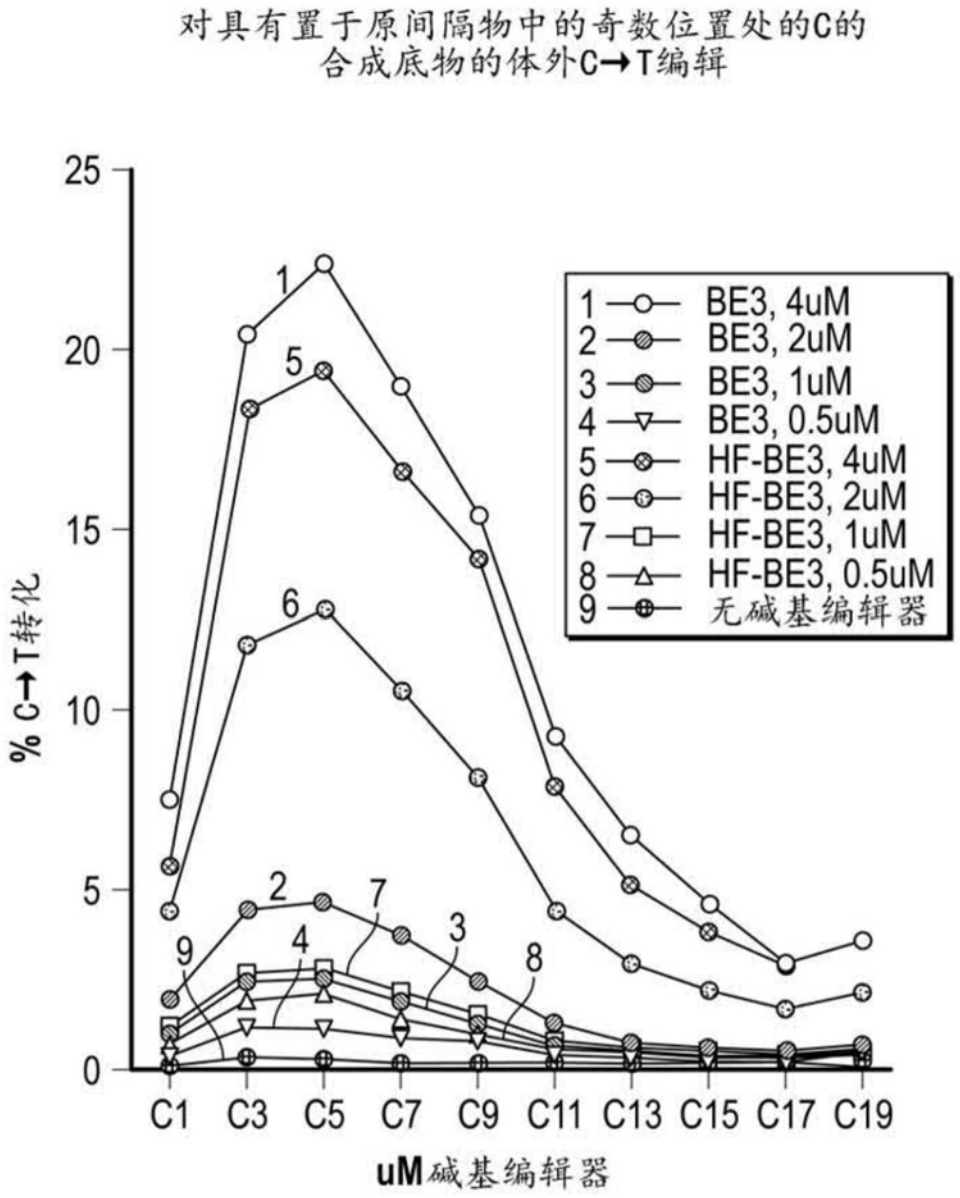

对具有置于原间隔物中的奇数位置处的C的
合成底物的体外C→T编辑
% C→T转化
25
20
15
10
5
0
C1
C3
C5
C7
C9
C11
C13
C15
C17
C19
uM碱基编辑器
1 BE3, 4uM
2 BE3, 2uM
3 BE3, 1uM
4 BE3, 0.5uM
5 HF-BE3, 4uM
6 HF-BE3, 2uM
7 HF-BE3, 1uM
8 HF-BE3, 0.5uM
9 无碱基编辑器

图104

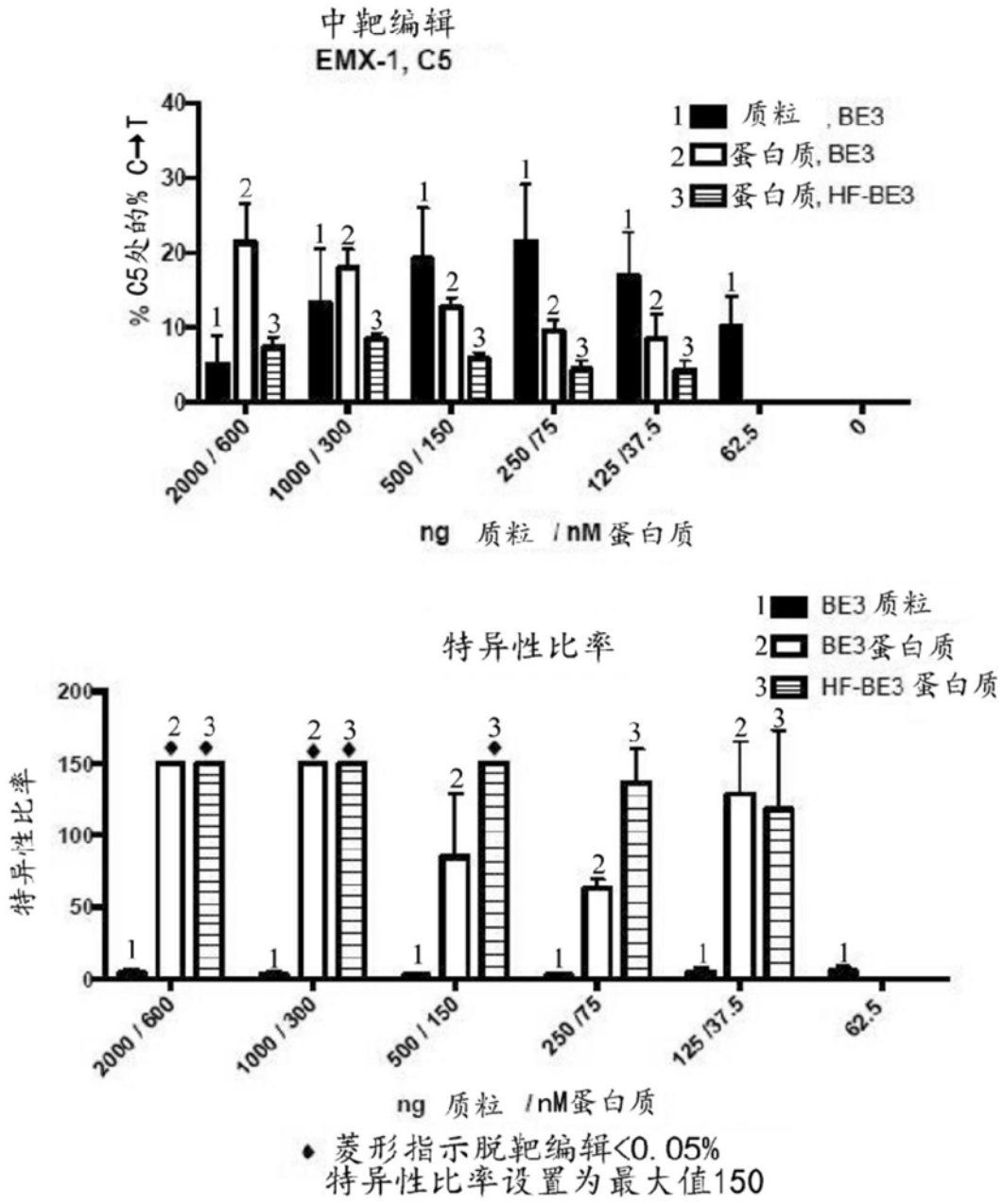

中靶编辑
EMX-1, C5
% C5处的% C→T
1 质粒 , BE3
2 蛋白质, BE3
3 蛋白质, HF-BE3
2000 / 600
1000 / 300
500 / 150
250 /75
125 /37.5
62.5
0
ng 质粒 / nM 蛋白质
特异性比率
1 BE3 质粒
2 BE3蛋白质
3 HF-BE3 蛋白质
特异性比率
ng 质粒 / nM蛋白质
菱形指示脱靶编辑<0.05%
特异性比率设置为最大值150

图105

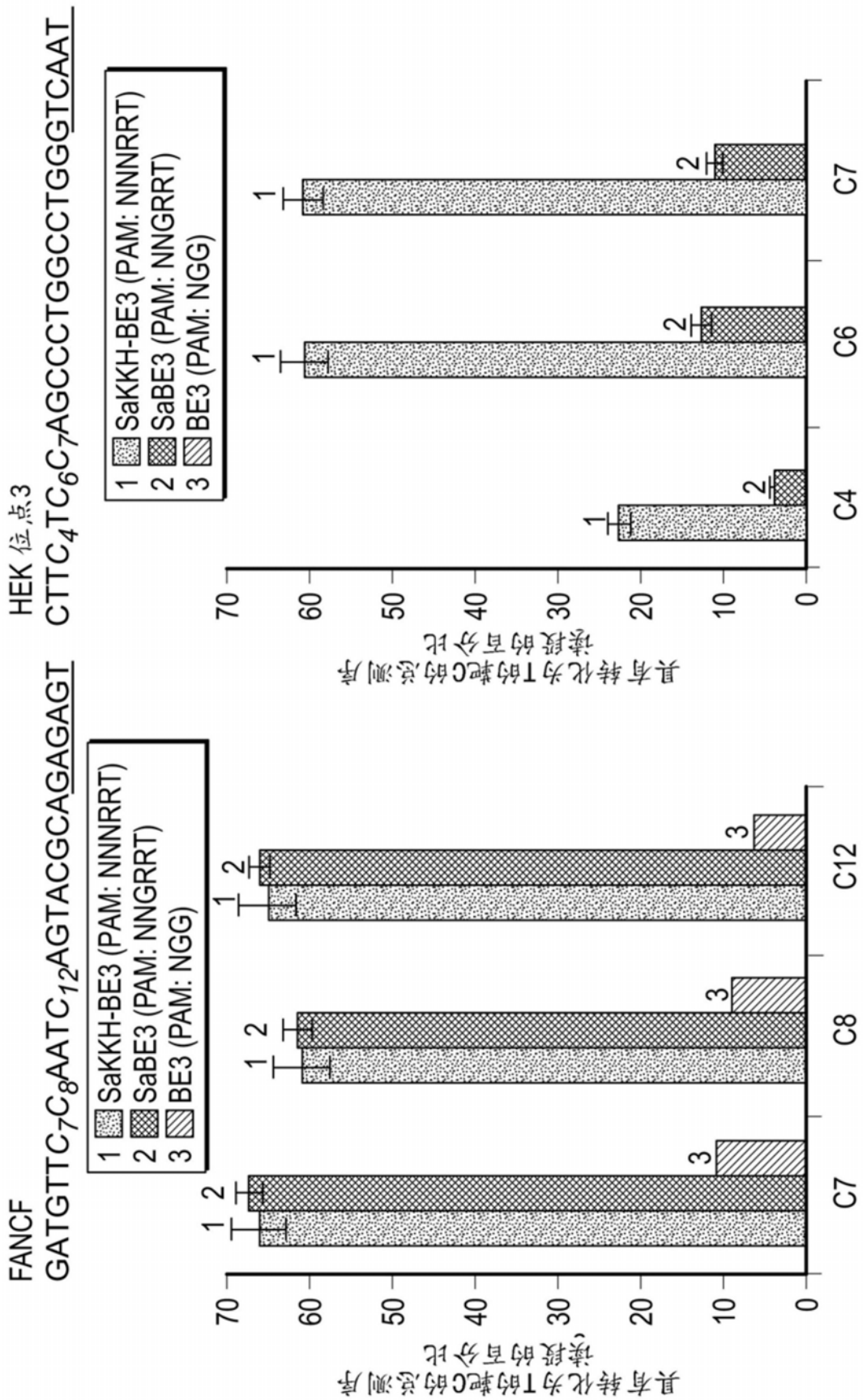
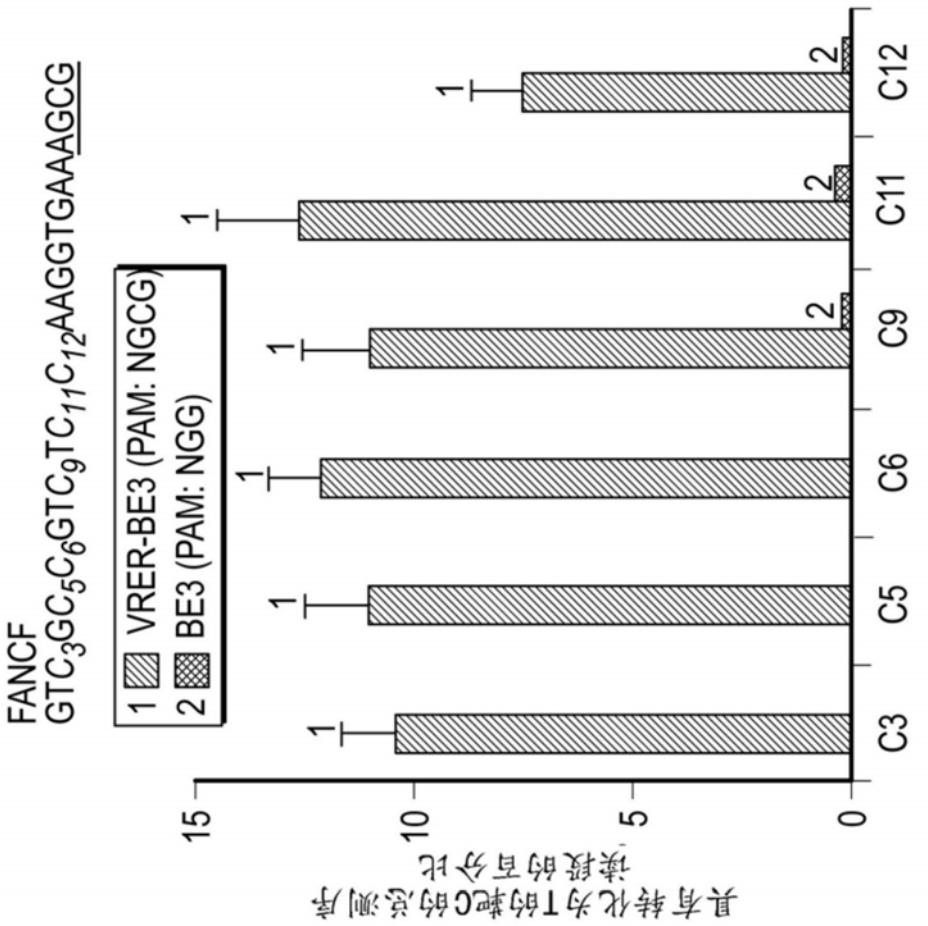

FANCF
GATGTTC7C8AATC12AGTACGCAGAGAGT
HEK 位点3
CTTC4TC6C7AGCCCTGGCCTGGGTCAAT
1 SaKKH-BE3 (PAM: NNNRRT)
2 SaBE3 (PAM: NNGRRT)
3 BE3 (PAM: NGG)
具有转化为T的靶C的总测序读段的百分比
0
10
20
30
40
50
60
70
C4
C6
C7
C8
C12

图106A


FANCF
GTC3GC5C6GTC9TC11C12AAGGTGAAAGCG
1 VRER-BE3 (PAM: NGCG)
2 BE3 (PAM: NGG)
具有转化为T的靶C的总测序
读段的百分比
0
5
10
15
C3
C5
C6
C9
C11
C12

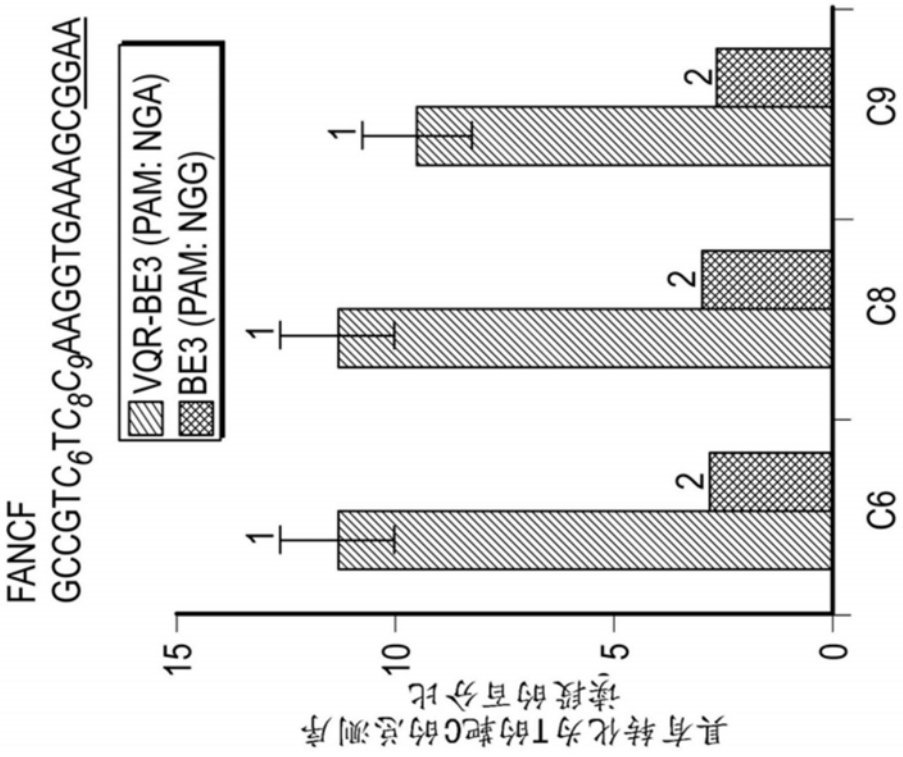

FANCF
GCCGTC6TC8C9AAGGTGAAAGCGGAA
VQR-BE3 (PAM: NGA)
BE3 (PAM: NGG)
具有转化为T的靶C的总测序
读段的百分比
0
5
10
15
C6
C8
C9

图106B

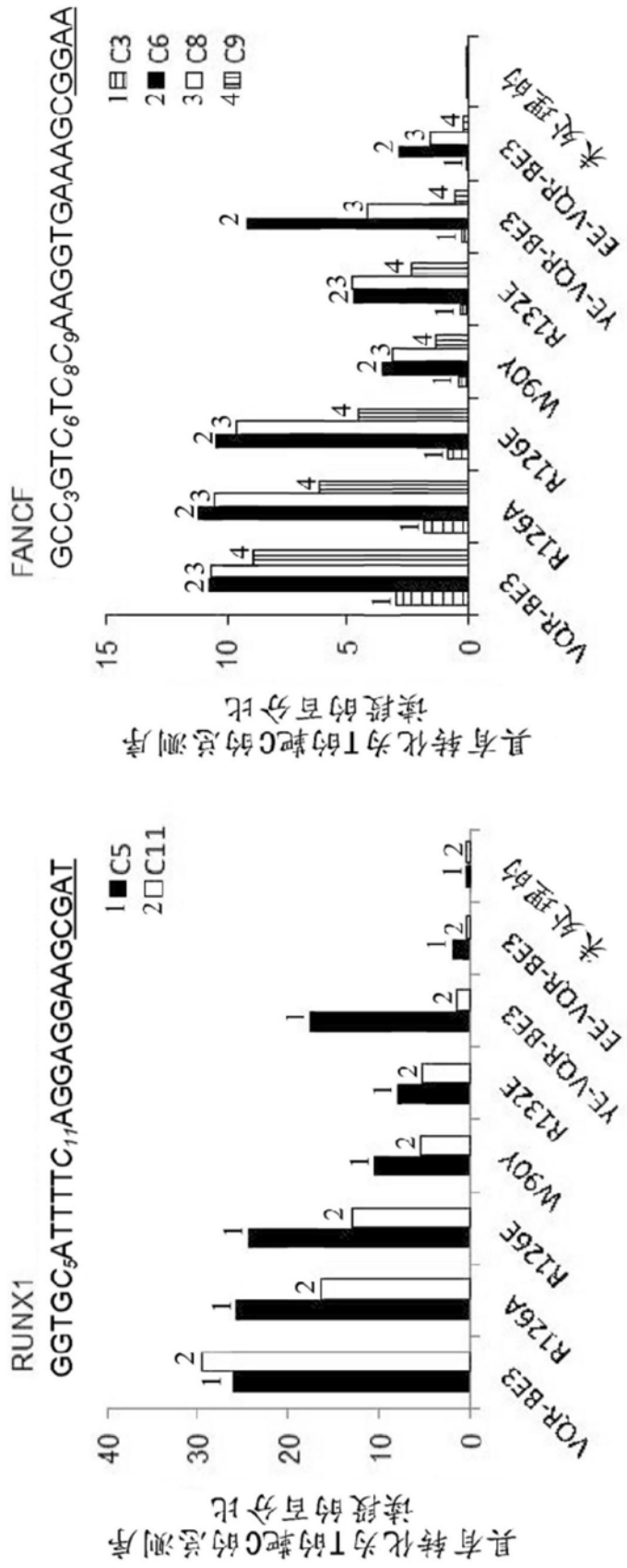

FANCF
GCC3GTC6TC8C9AAGGTGAAAGCGGAA
1 C3
2 C6
3 C8
4 C9
具有转化为T的靶C的总测序读段的百分比
VQR-BE3
R126A
R126E
W90Y
R132E
YE-VQR-BE3
EE-VQR-BE3
未处理的
RUNX1
GGTGC5ATTTTC11AGGAGGAAGCGAT
1 C5
2 C11
具有转化为T的靶C的总测序读段的百分比
VQR-BE3
R126A
R126E
W90Y
R132E
YE-VQR-BE3
EE-VQR-BE3
未处理的

图107A

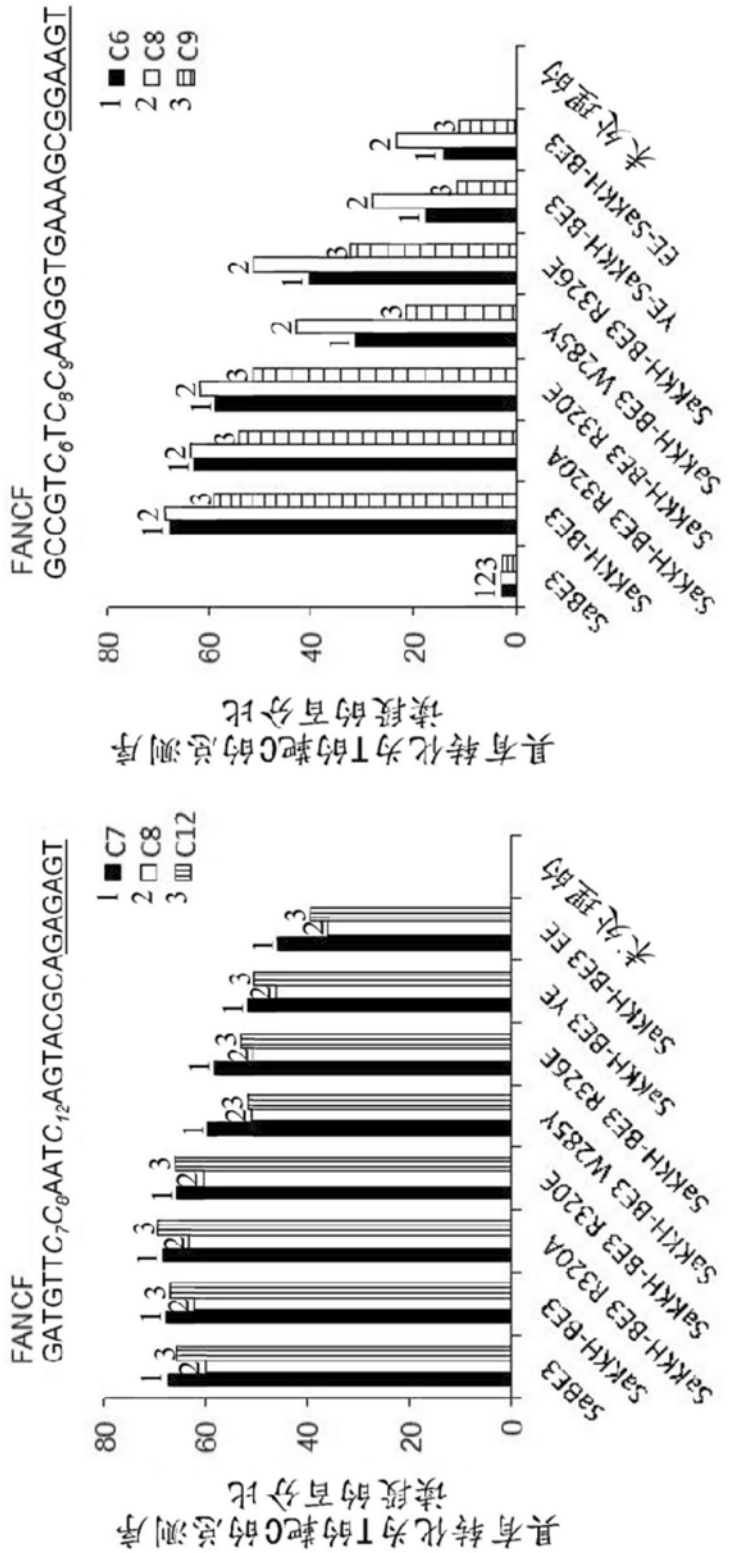

FANCF
GCCGTC6TC8C9AAGGTGAAAGCGGAAGT
1 C6
2 C8
3 C9
具有转化为T的靶C的总测序读段的百分比
0
20
40
60
80
saBE3
saKKH-BE3
saKKH-BE3 R32OA
saKKH-BE3 R32OE
saKKH-BE3 W28SY
saKKH-BE3 R326E
saKKH-BE3 YE
saKKH-BE3 EE
未处理的
FANCF
GATGTTC7C8AATC12AGTACGCAGAGAGT
1 C7
2 C8
3 C12
具有转化为T的靶C的总测序读段的百分比
0
20
40
60
80
saBE3
saKKH-BE3
saKKH-BE3 R32OA
saKKH-BE3 R32OE
saKKH-BE3 W28SY
saKKH-BE3 R326E
saKKH-BE3 YE
saKKH-BE3 EE
未处理的

图107B

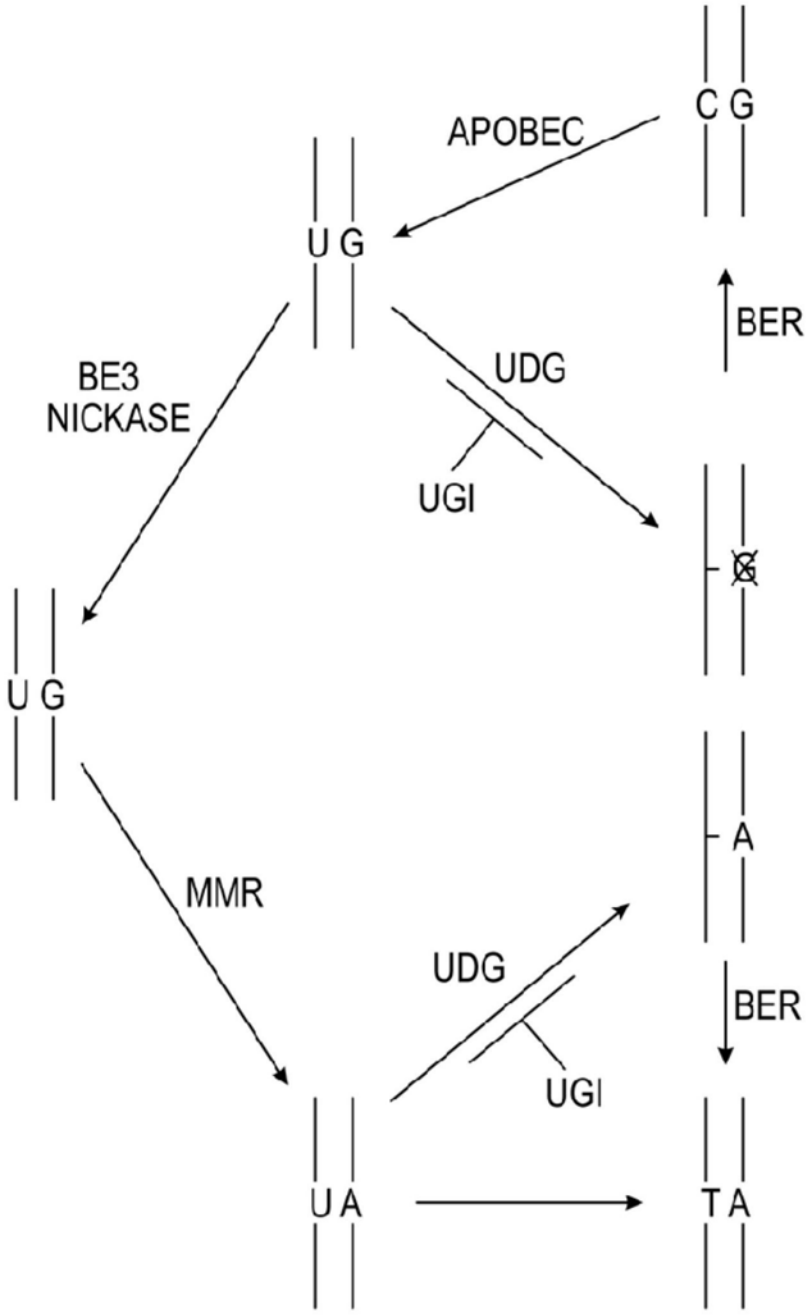


图108

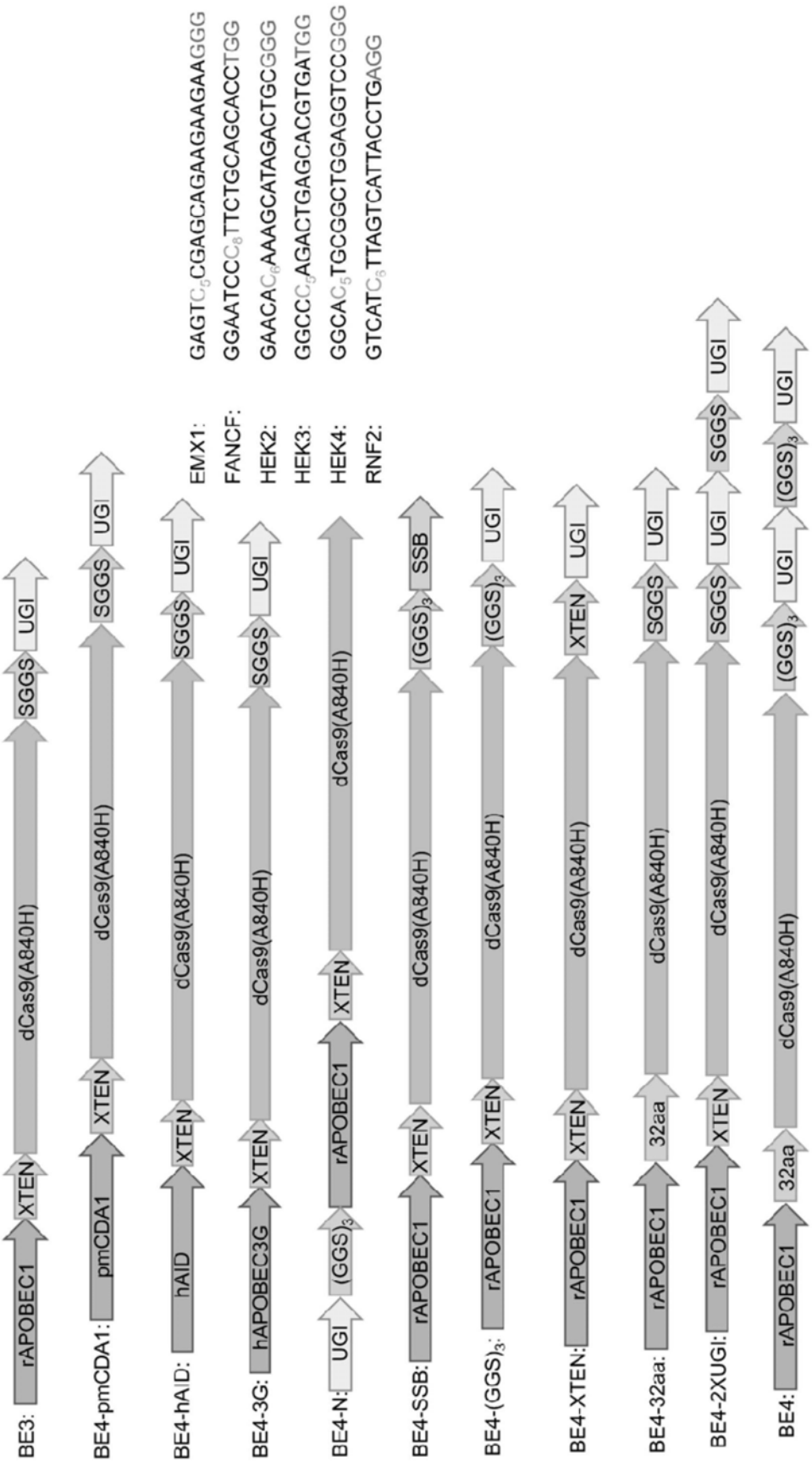

BE3:
rAPOBEC1
XTEN
dCas9(A840H)
SGGS
UGI
BE4-pmCDA1:
pmCDA1
XTEN
dCas9(A840H)
SGGS
UGI
BE4-hAID:
hAID
XTEN
dCas9(A840H)
SGGS
UGI
BE4-3G:
hAPOBEC3G
XTEN
dCas9(A840H)
SGGS
UGI
BE4-N:
UGI
(GGS)3
rAPOBEC1
XTEN
dCas9(A840H)
BE4-SSB:
rAPOBEC1
XTEN
dCas9(A840H)
(GGS)3
SSB
BE4-(GGS)3:
rAPOBEC1
XTEN
dCas9(A840H)
(GGS)3
UGI
BE4-XTEN:
rAPOBEC1
XTEN
dCas9(A840H)
XTEN
UGI
BE4-32aa:
rAPOBEC1
32aa
dCas9(A840H)
SGGS
UGI
BE4-2XUGI:
rAPOBEC1
XTEN
dCas9(A840H)
SGGS
UGI
SGGS
UGI
BE4:
rAPOBEC1
32aa
dCas9(A840H)
(GGS)3
UGI
(GGS)3
UGI
EMX1:
GAGTCCGAGCAGAAGAAGAAGGG
FANCF:
GGAATCCCCTTCTGCAGCACCTGG
HEK2:
GAACACAAAGCATAGACTGCGGG
HEK3:
GGCCCAGACTGAGCACGTGATGG
HEK4:
GGCACTGCGGCTGGAGGTCCGGG
RNF2:
GTCATCTTAGTCATTACCTGAGG

图109

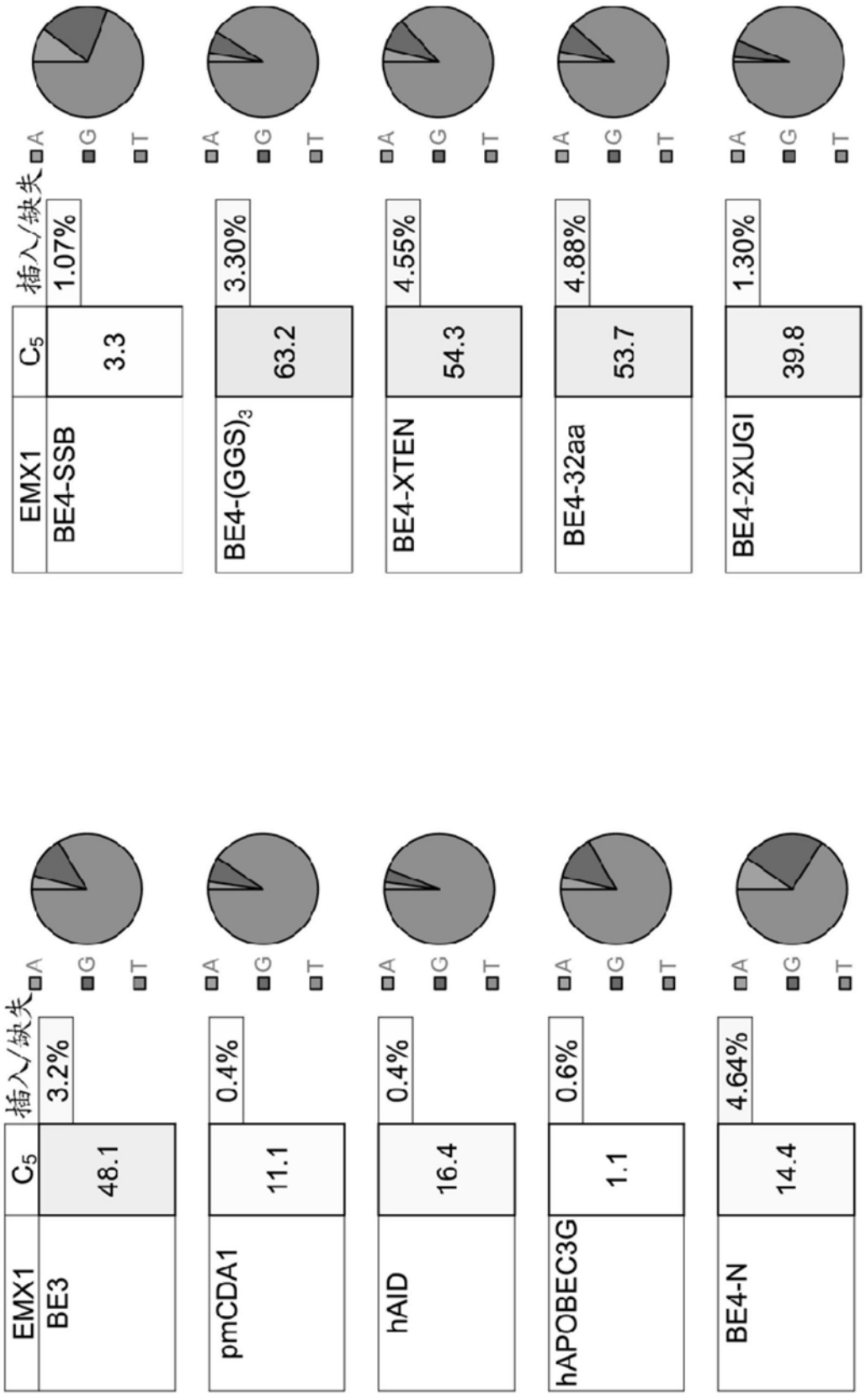

EMX1
BE4-SSB
C5
3.3
插入/缺失
1.07%
BE4-(GGS)3
63.2
3.30%
BE4-XTEN
54.3
4.55%
BE4-32aa
53.7
4.88%
BE4-2XUGI
39.8
1.30%
BE3
48.1
3.2%
pmCDA1
11.1
0.4%
hAID
16.4
0.4%
hAPOBEC3G
1.1
0.6%
BE4-N
14.4
4.64%
A
G
T

图110

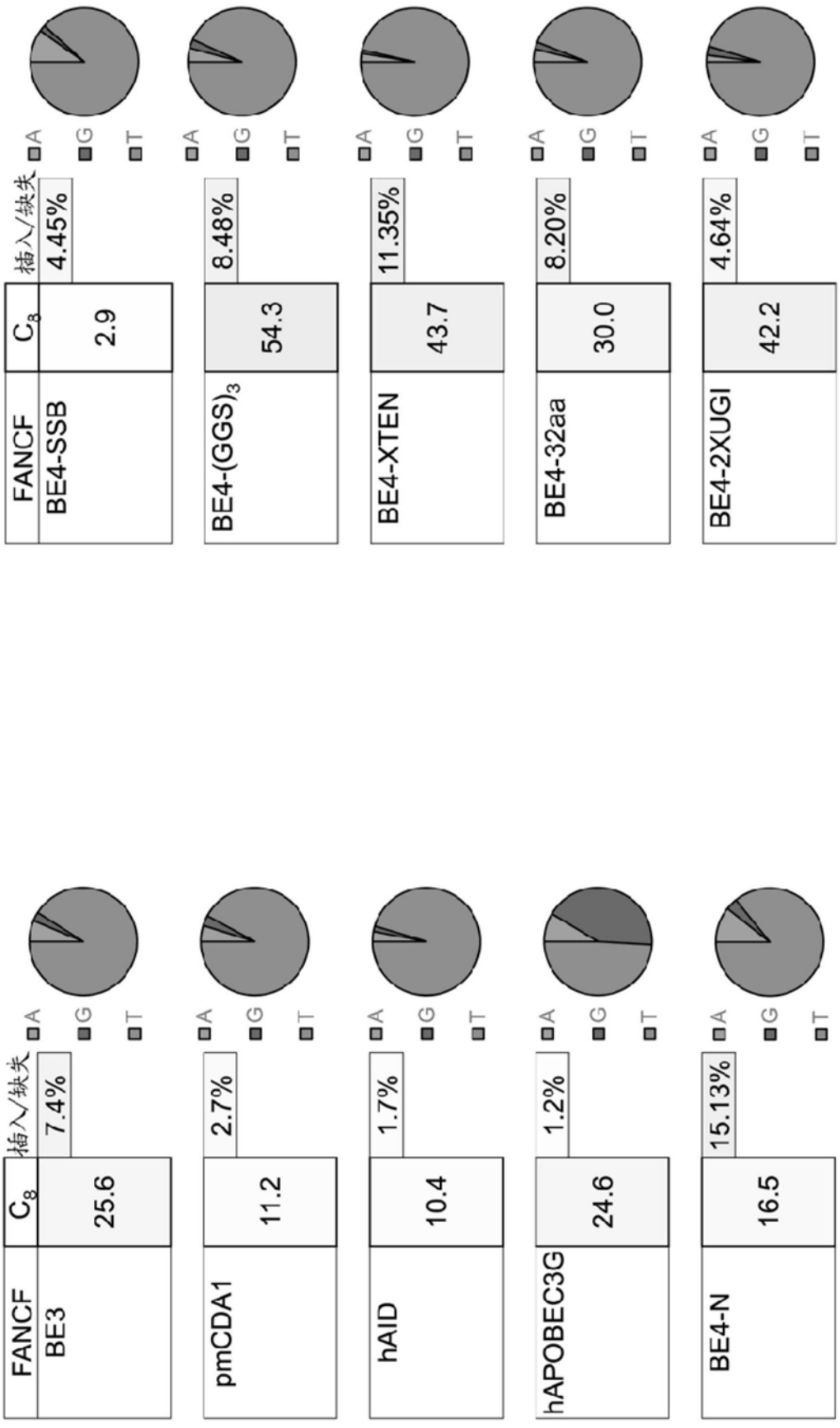

FANCF
BE4-SSB
C_8
2.9
插入/缺失
4.45%
BE4-(GGS)$_3$
54.3
8.48%
BE4-XTEN
43.7
11.35%
BE4-32aa
30.0
8.20%
BE4-2XUGI
42.2
4.64%
BE3
25.6
7.4%
pmCDA1
11.2
2.7%
hAID
10.4
1.7%
hAPOBEC3G
24.6
1.2%
BE4-N
16.5
15.13%
A
G
T

图111

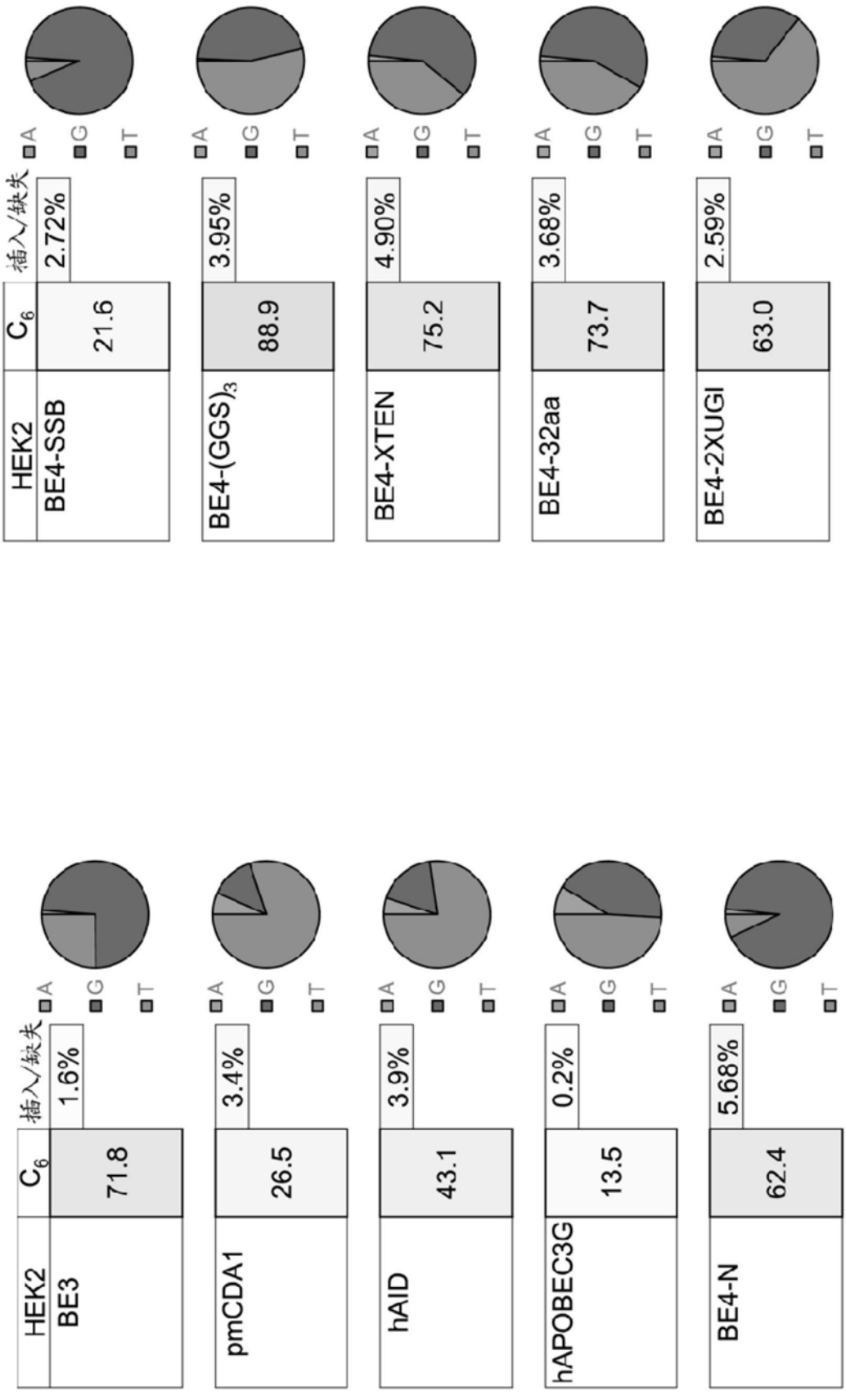

HEK2
BE4-SSB
C6
21.6
插入/缺失
2.72%
BE4-(GGS)3
88.9
3.95%
BE4-XTEN
75.2
4.90%
BE4-32aa
73.7
3.68%
BE4-2XUGI
63.0
2.59%
BE3
71.8
1.6%
pmCDA1
26.5
3.4%
hAID
43.1
3.9%
hAPOBEC3G
13.5
0.2%
BE4-N
62.4
5.68%
A
G
T

图112

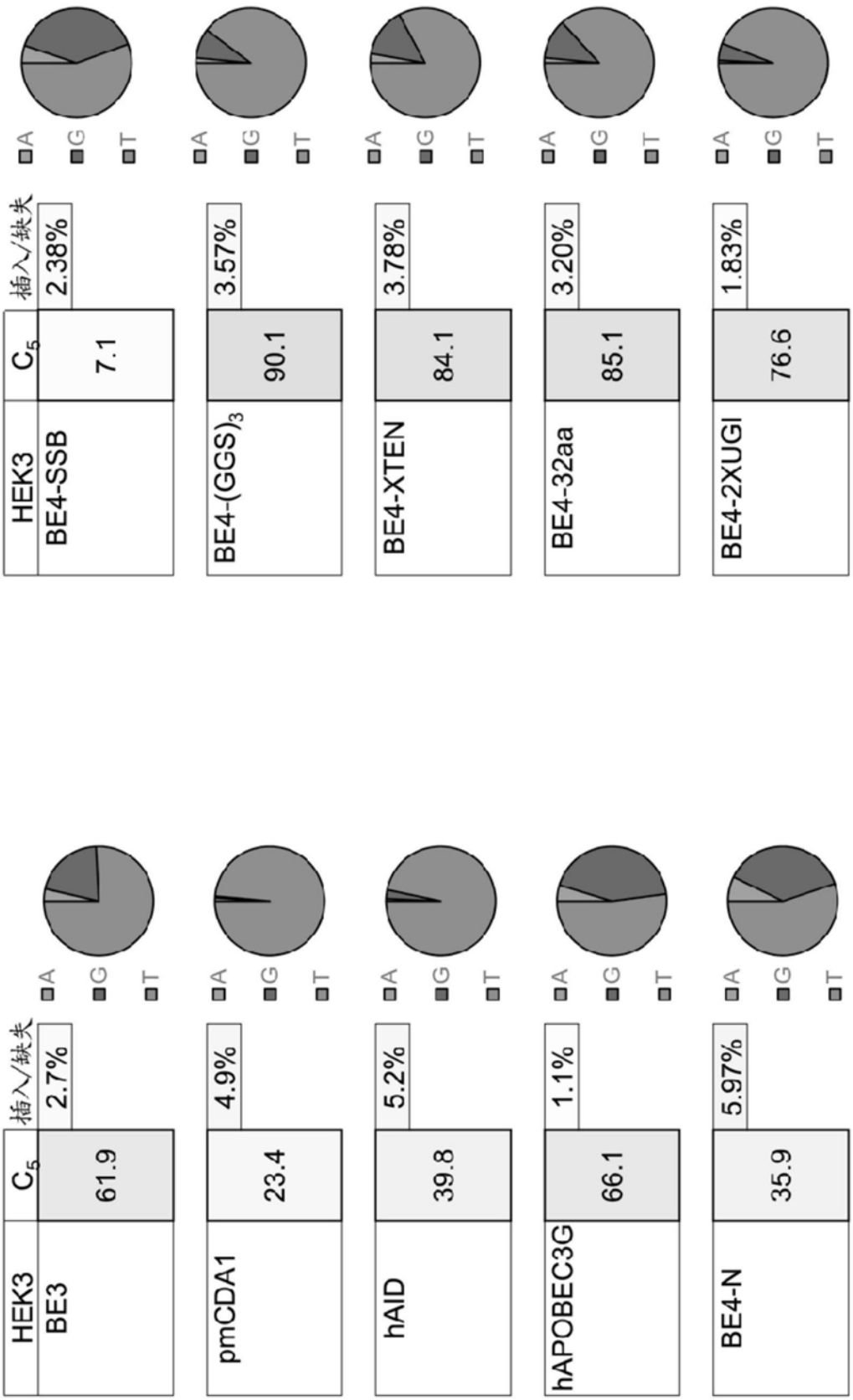

HEK3
BE4-SSB
C5
7.1
插入/缺失
2.38%
BE4-(GGS)3
90.1
3.57%
BE4-XTEN
84.1
3.78%
BE4-32aa
85.1
3.20%
BE4-2XUGI
76.6
1.83%
BE3
61.9
2.7%
pmCDA1
23.4
4.9%
hAID
39.8
5.2%
hAPOBEC3G
66.1
1.1%
BE4-N
35.9
5.97%
A
G
T

图113

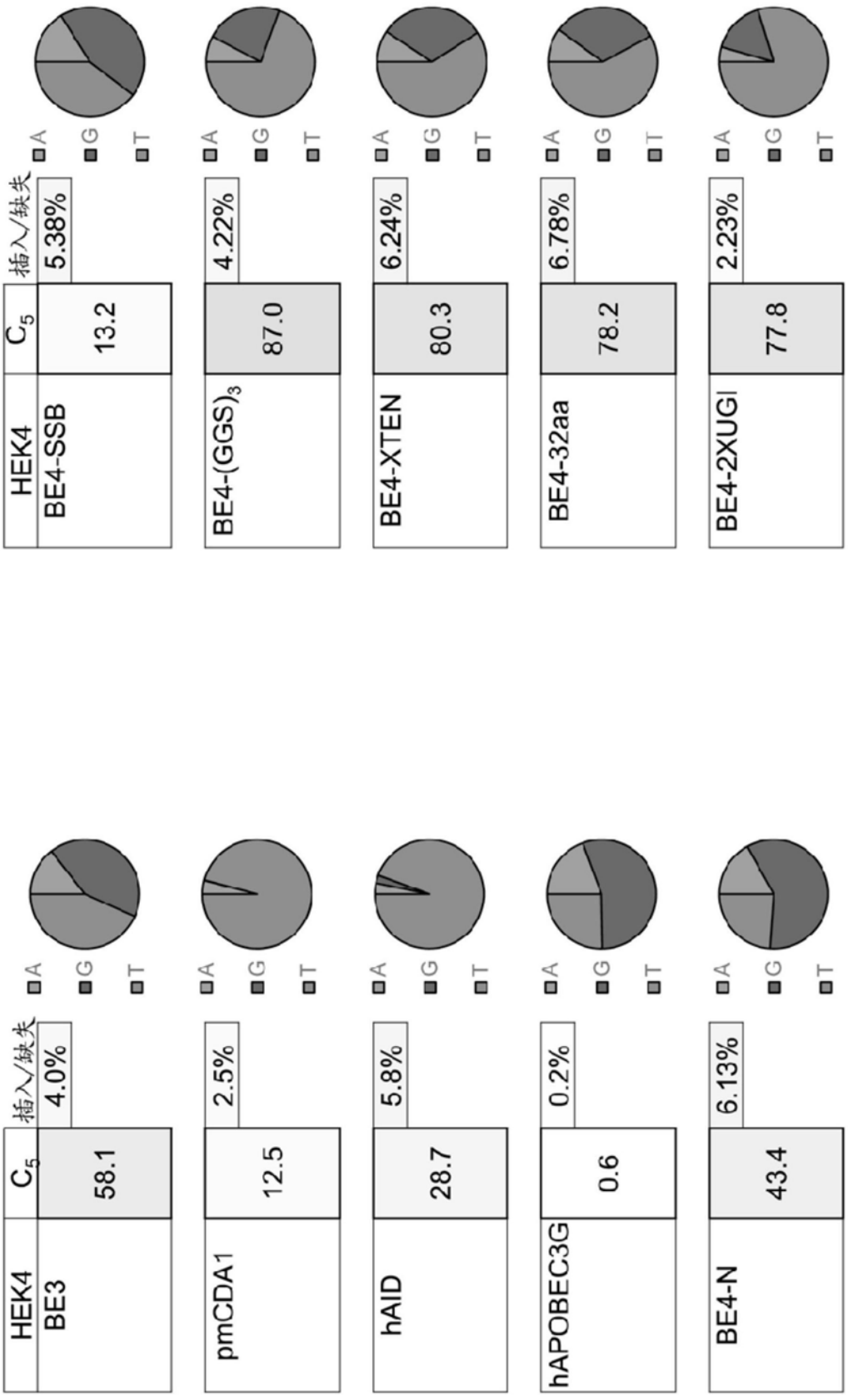

HEK4
BE4-SSB
C5
13.2
插入/缺失
5.38%
BE4-(GGS)3
87.0
4.22%
BE4-XTEN
80.3
6.24%
BE4-32aa
78.2
6.78%
BE4-2XUGI
77.8
2.23%
BE3
58.1
4.0%
pmCDA1
12.5
2.5%
hAID
28.7
5.8%
hAPOBEC3G
0.6
0.2%
BE4-N
43.4
6.13%
A
G
T

图114

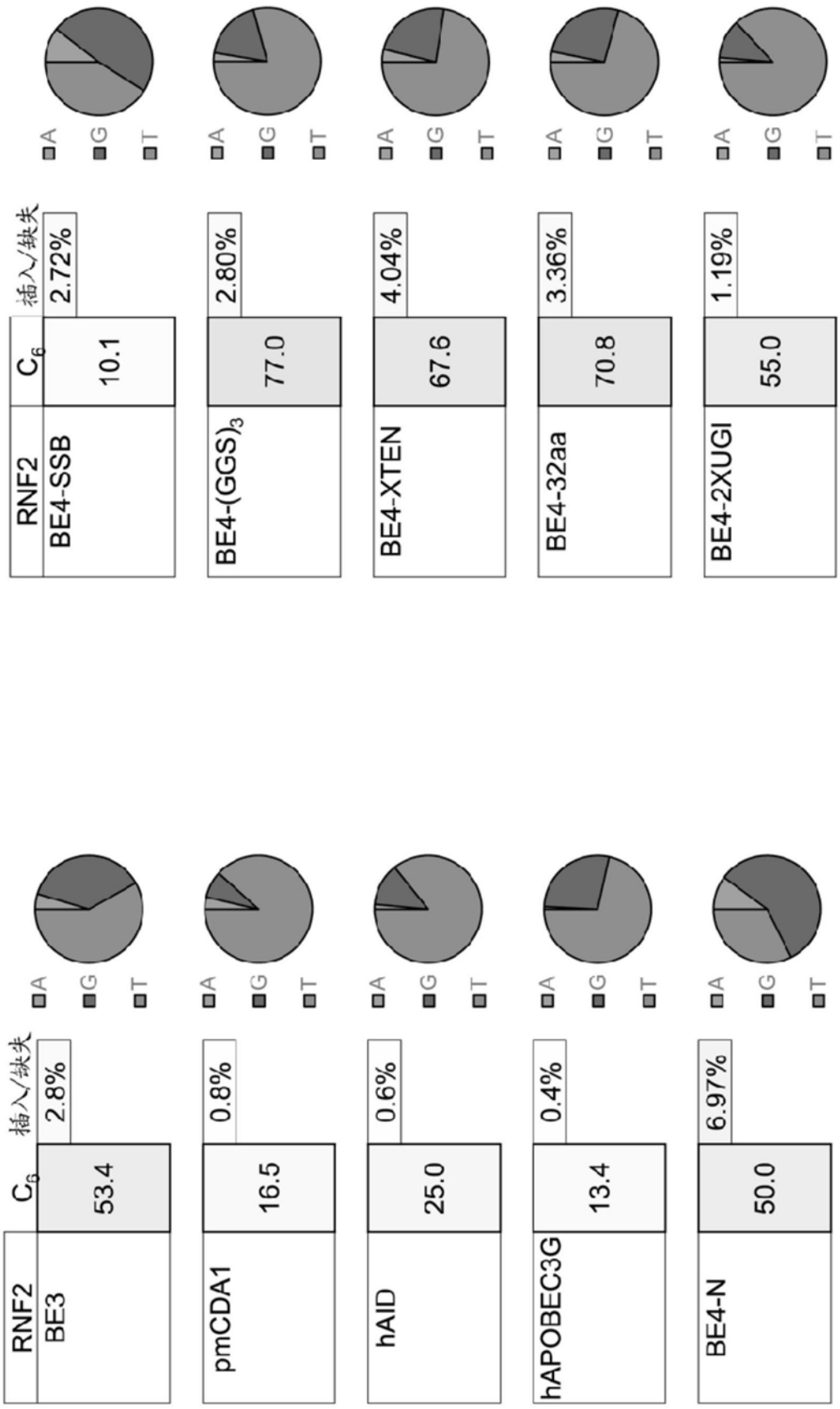

RNF2
BE4-SSB
C6
10.1
插入/缺失
2.72%
BE4-(GGS)3
77.0
2.80%
BE4-XTEN
67.6
4.04%
BE4-32aa
70.8
3.36%
BE4-2XUGI
55.0
1.19%
BE3
53.4
2.8%
pmCDA1
16.5
0.8%
hAID
25.0
0.6%
hAPOBEC3G
13.4
0.4%
BE4-N
50.0
6.97%
A
G
T

图115

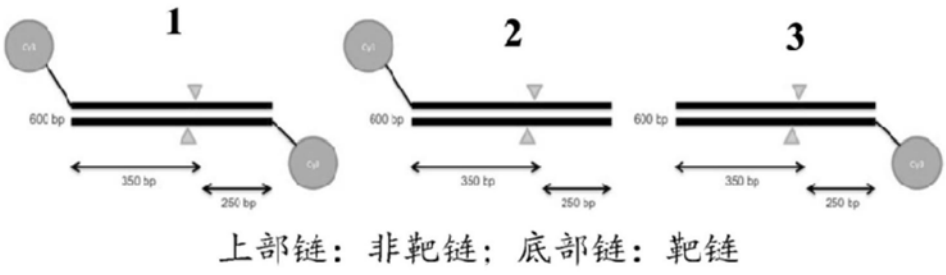

1
2
3
600 bp
350 bp
250 bp
上部链：非靶链；底部链：靶链

图116

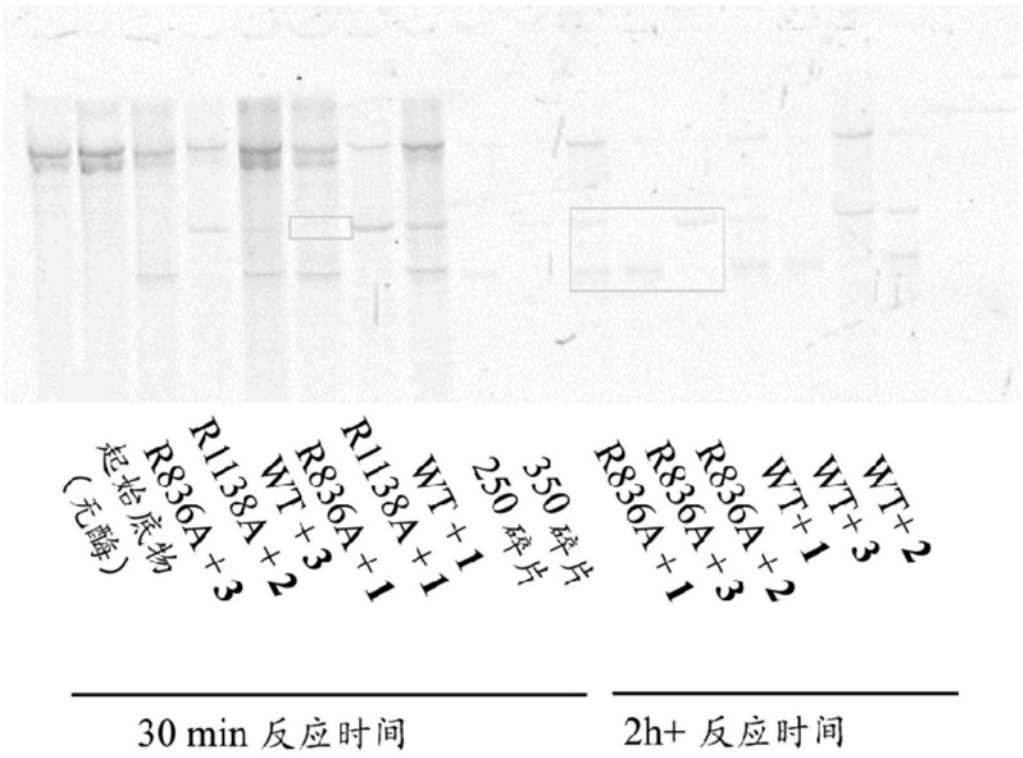

起始底物（无酶）
R836A +3
R1138A +2
WT +3
R836A +1
R1138A +1
WT +1
250 碎片
350 碎片
R836A +1
R836A +3
R836A +2
WT+1
WT+3
WT+2
30 min 反应时间
2h+ 反应时间

图117

PAM

FANCF1	TTTC	G1	C2	G3	G4	G7	C10	C11	C15	G17	C20	C22					
BG		0.01005	0.03022	0.01083	0.00933	0.00502	0.01872	0.02014	0.03740	0.00504	0.01938	0.02099					
AsCpf1		0.00895	0.02769	0.07439	0.00672	0.06495	0.02395	0.01795	0.02691	0.01570	0.02237	0.02406					
APOBEC-AsCpf1(R912A)-UGI		0.00883	0.02606	0.01430	0.00539	0.00490	0.29513	0.28158	0.04373	0.01425	0.02795	0.01631					
APOBEC-AsCpf1(R1225A)-UGI		0.00796	0.03754	0.01131	0.00749	0.01124	0.02769	0.02908	0.03846	0.02016	0.01732	0.01651					
FANCF2	TTTC	C1	G2	G3	C4	C7	G9	G10	C11	G12	G13	C15	C16	G17	C18		
BG		0.02628	0.01963	0.02346	0.00318	0.00798	0.00844	0.00850	0.08758	0.05168	0.00637	0.00499	0.07694	0.01382	0.01805		
AsCpf1		0.03489	0.03325	0.05248	0.00392	0.00722	0.02484	0.00791	0.06139	0.10805	0.01851	0.02139	0.10988	0.00927	0.00992		
APOBEC-AsCpf1(R912A)-UGI		0.02440	0.02677	0.00699	0.00481	0.19399	0.01115	0.00534	0.11463	0.01660	0.00427	0.00722	0.10175	0.01119	0.01065		
APOBEC-AsCpf1(R1225A)-UGI		0.02855	0.04216	0.00322	0.00479	0.00587	0.01004	0.00585	0.06634	0.00586	0.00638	0.00444	0.07056	0.00797	0.01061		
HEK3-3	TTTC	C3	C6	C8	C9	C12	C13	C14	C18	C19							
BG		0.00907	0.00726	0.01440	0.01539	0.01084	0.01181	0.00729	0.01540	0.00634							
AsCpf1		0.01118	0.00670	0.01994	0.00892	0.01334	0.01342	0.01797	0.01118	0.03351							
APOBEC-AsCpf1(R912A)-UGI		0.00647	0.02264	0.44917	0.30191	0.03754	0.12192	0.01513	0.01291	0.00754							
APOBEC-AsCpf1(R1225A)-UGI		0.00770	0.00769	0.10243	0.07236	0.01202	0.00770	0.00551	0.01865	0.00549							
HEK3-4		G23	C20	G18	G16	C15	C13	G12	G10	G7	G6	C5	G1	GAAA			
BG		0.00638	0.01001	0.03459	0.01008	0.01008	0.01285	0.01284	0.00548	0.01458	0.02557	0.01099	0.00636				
AsCpf1		0.00988	0.00862	0.02466	0.01984	0.03343	0.02600	0.00865	0.00617	0.01718	0.02216	0.01232	0.00488				
APOBEC-AsCpf1(R912A)-UGI		0.00323	0.01185	0.03012	0.00759	0.01735	0.01734	0.04550	0.15878	0.00862	0.02266	0.00542	0.00644				
APOBEC-AsCpf1(R1225A)-UGI		0.01142	0.01045	0.03044	0.01435	0.01626	0.02104	0.01051	0.01238	0.01045	0.03428	0.01810	0.00473				

BG－背景； R912A－BE3；R1225A－自我抹败性BE（对非靶链产生切口）

图118



图119

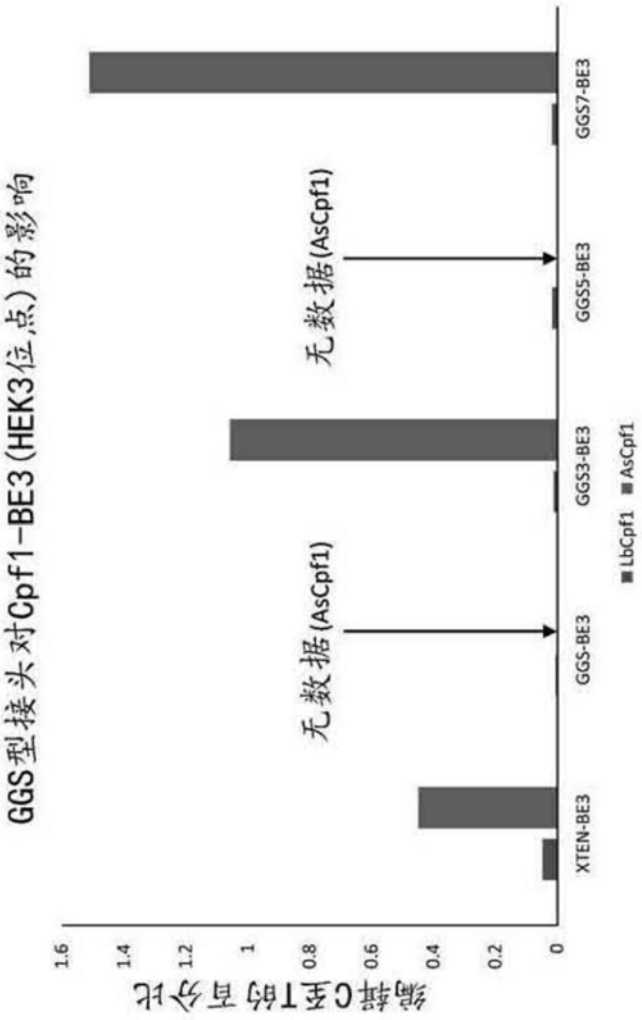

GGS型接头对Cpf1-BE3(HEK3位点)的影响
编辑C的百分比
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
无数据(AsCpf1)
无数据(AsCpf1)
XTEN-BE3
GGS-BE3
GGS3-BE3
GGS5-BE3
GGS7-BE3
LbCpf1
AsCpf1

图120

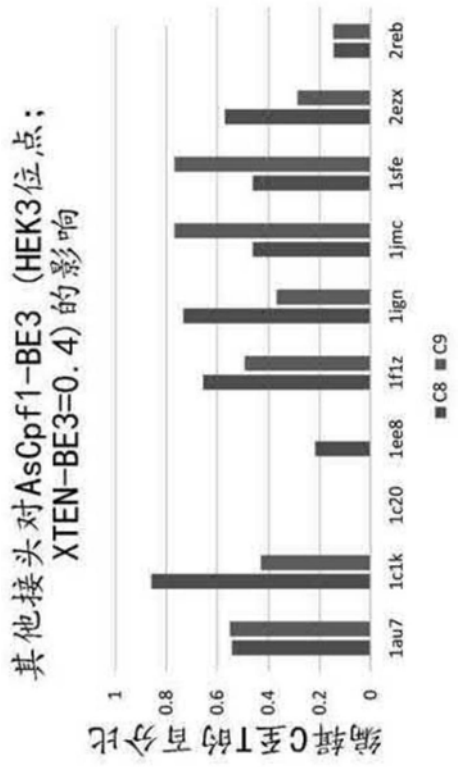

其他接头对AsCpf1-BE3（HEK3位点；
XTEN-BE3=0.4）的影响
编辑C至T的百分比
1
0.8
0.6
0.4
0.2
0
1au7
1c1k
1c20
1ee8
1f1z
1ign
1jmc
1sfe
2ezx
2reb
C8
C9

图121

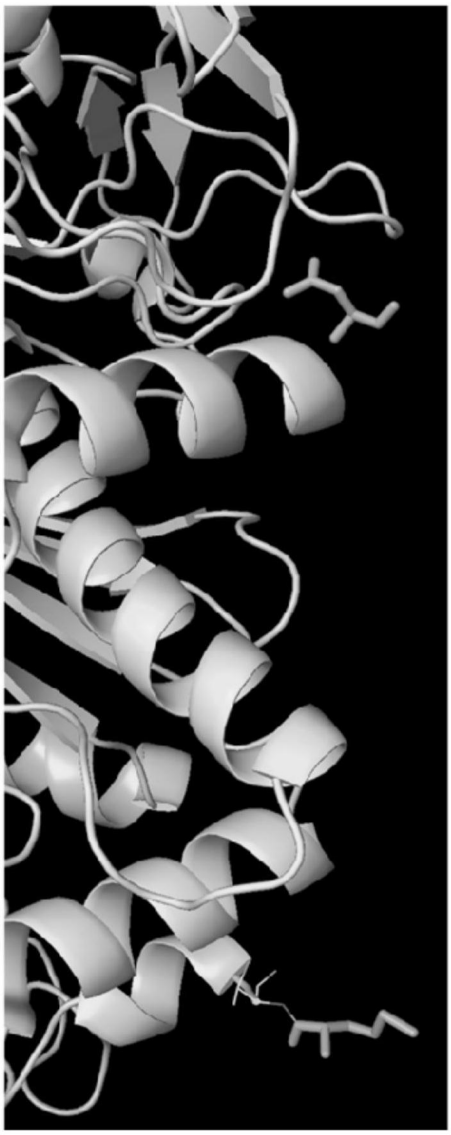


图122

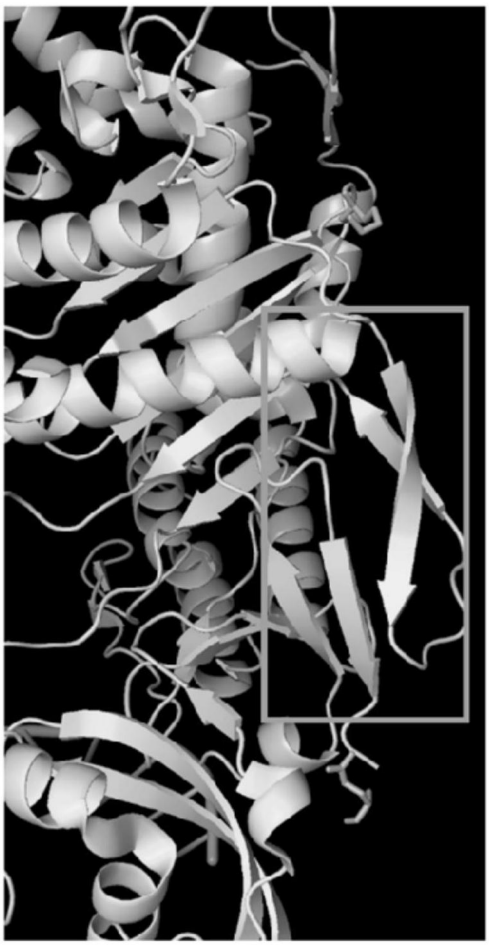


图123



图124

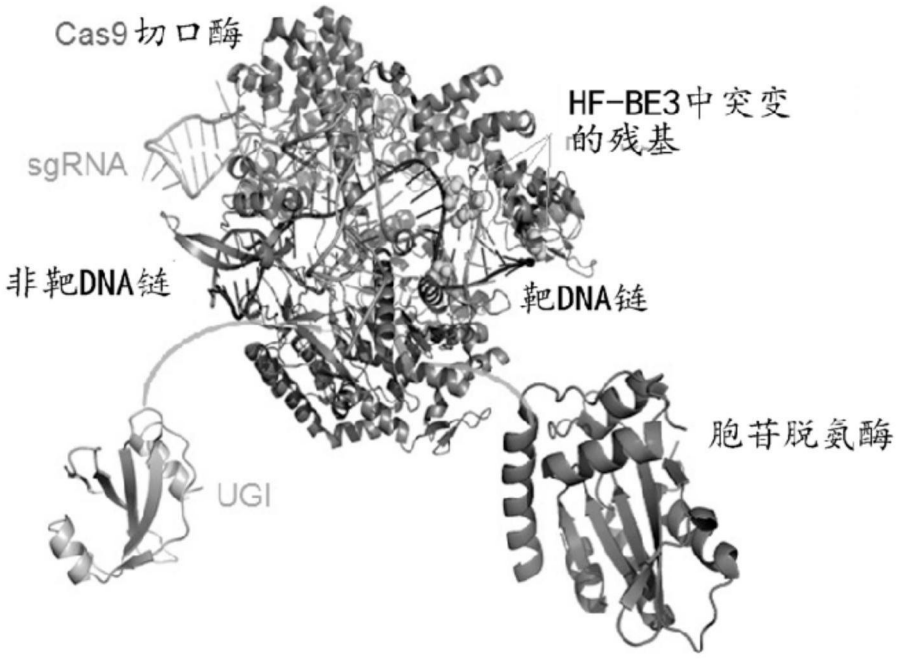

Cas9切口酶
HF-BE3中突变的残基
sgRNA
非靶DNA链
靶DNA链
胞苷脱氨酶
UGI

图125A

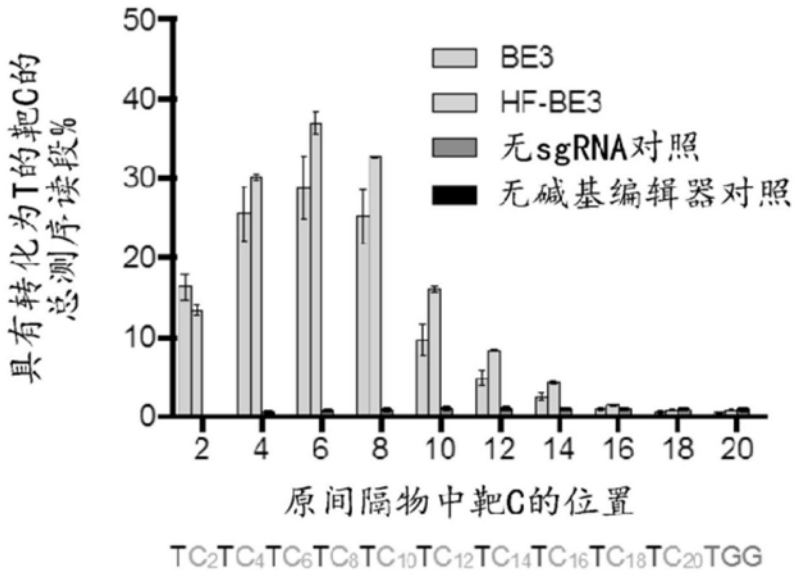

具有转化为T的靶C的总测序读段%
50
40
30
20
10
0
BE3
HF-BE3
无sgRNA对照
无碱基编辑器对照
2 4 6 8 10 12 14 16 18 20
原间隔物中靶C的位置
TC2TC4TC6TC8TC10TC12TC14TC16TC18TC20TGG

图125B

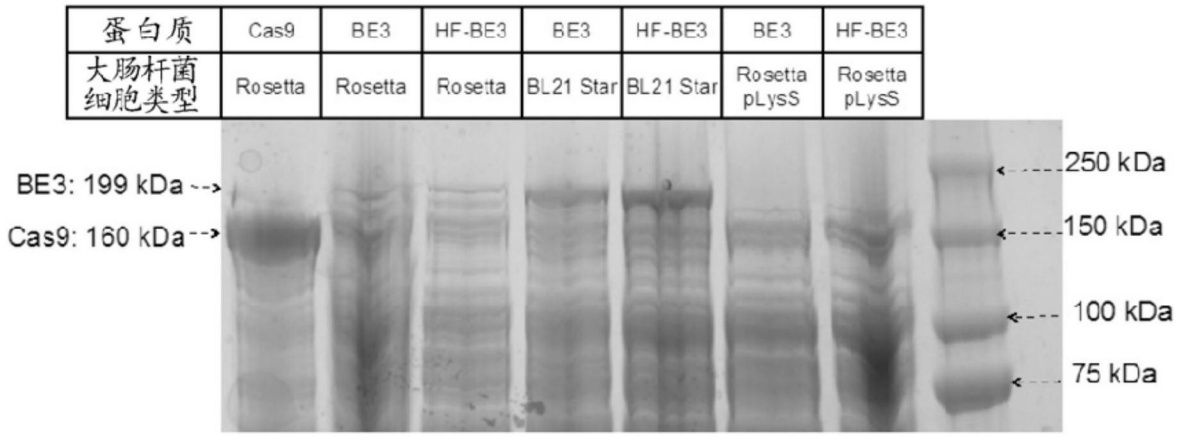

蛋白质
Cas9
BE3
HF-BE3
BE3
HF-BE3
BE3
HF-BE3
大肠杆菌细胞类型
Rosetta
Rosetta
Rosetta
BL21 Star
BL21 Star
Rosetta pLysS
Rosetta pLysS
BE3: 199 kDa
Cas9: 160 kDa
250 kDa
150 kDa
100 kDa
75 kDa

图126A

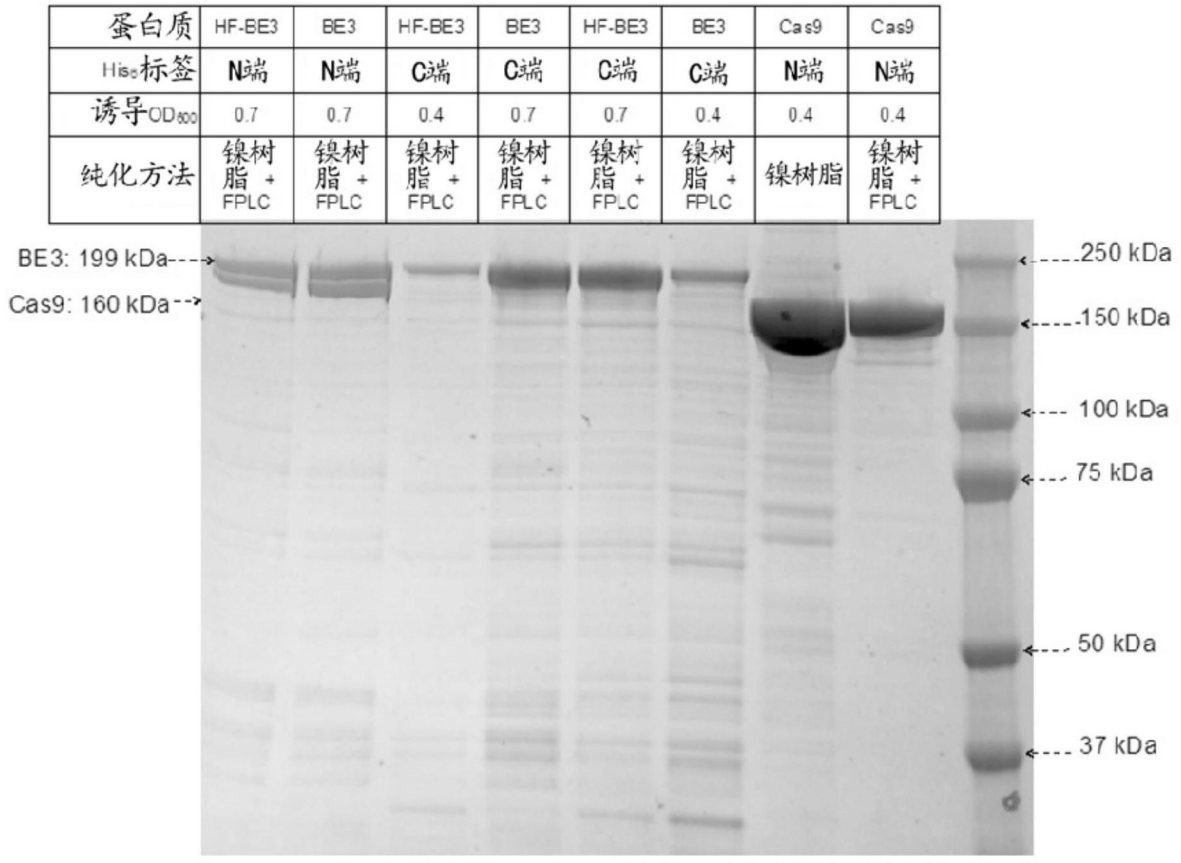

蛋白质
HF-BE3
BE3
HF-BE3
BE3
HF-BE3
BE3
Cas9
Cas9
His6标签
N端
N端
C端
C端
C端
C端
N端
N端
诱导OD600
0.7
0.7
0.4
0.7
0.7
0.4
0.4
0.4
纯化方法
镍树脂 + FPLC
镍树脂 + FPLC
镍树脂 + FPLC
镍树脂 + FPLC
镍树脂 + FPLC
镍树脂 + FPLC
镍树脂
镍树脂 + FPLC
BE3: 199 kDa
Cas9: 160 kDa
250 kDa
150 kDa
100 kDa
75 kDa
50 kDa
37 kDa

图126B

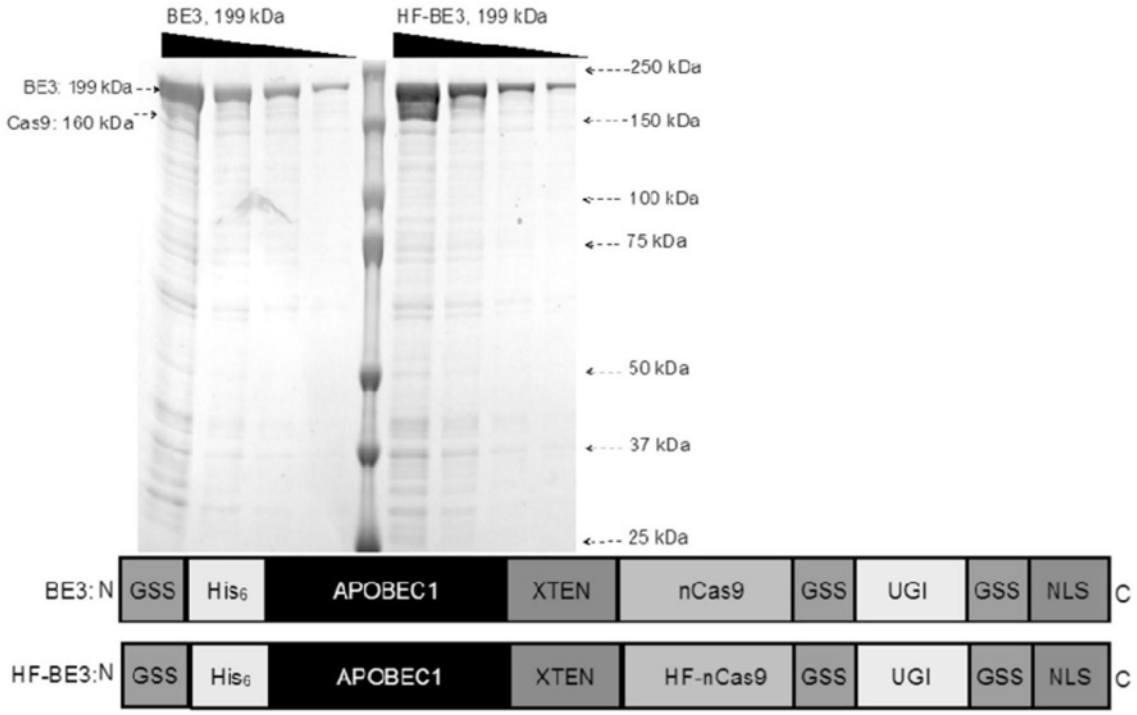

BE3, 199 kDa
HF-BE3, 199 kDa
BE3: 199 kDa
Cas9: 160 kDa
250 kDa
150 kDa
100 kDa
75 kDa
50 kDa
37 kDa
25 kDa
BE3:N
GSS
His6
APOBEC1
XTEN
nCas9
GSS
UGI
GSS
NLS
C
HF-BE3:N
GSS
His6
APOBEC1
XTEN
HF-nCas9
GSS
UGI
GSS
NLS
C

图126C

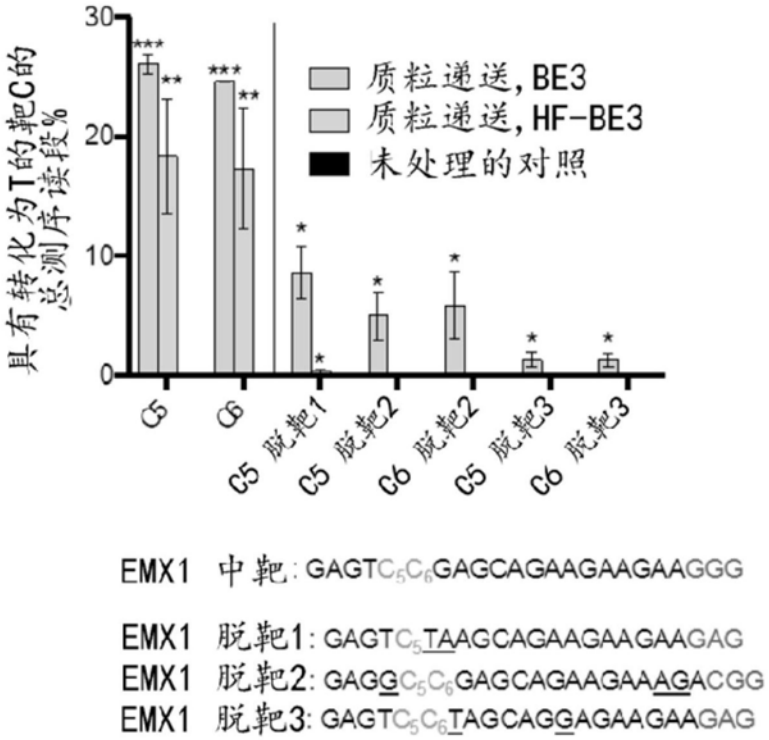

具有转化为T的靶C的总测序读段%
30
20
10
0
质粒递送, BE3
质粒递送, HF-BE3
未处理的对照
C5
C6
C5 脱靶1
C5 脱靶2
C6 脱靶2
C5 脱靶3
C6 脱靶3
EMX1 中靶: GAGTC5C6GAGCAGAAGAAGAAGGG
EMX1 脱靶1: GAGTC5TAAGCAGAAGAAGAAGAG
EMX1 脱靶2: GAGGC5C6GAGCAGAAGAAAGAGACGG
EMX1 脱靶3: GAGTC5C6TAGCAGGAGAAGAAGAG

图127A

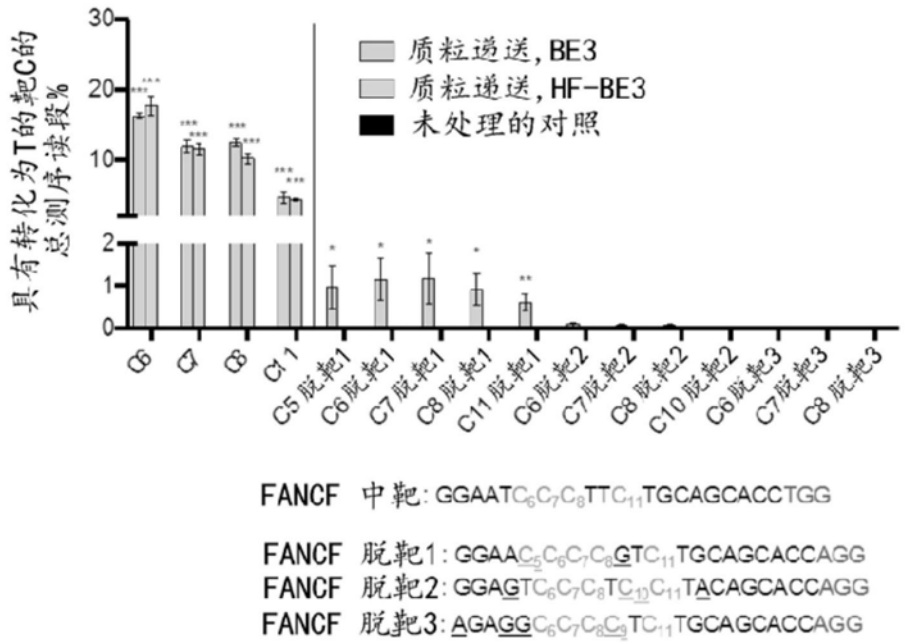

具有转化为T的靶C的总测序读段%
质粒递送,BE3
质粒递送,HF-BE3
未处理的对照
C6
C7
C8
C11
C5 脱靶1
C6 脱靶1
C7 脱靶1
C8 脱靶1
C11 脱靶1
C6 脱靶2
C7 脱靶2
C8 脱靶2
C10 脱靶2
C6 脱靶3
C7 脱靶3
C8 脱靶3
FANCF 中靶: GGAATC6C7C8TTC11TGCAGCACCTGG
FANCF 脱靶1: GGAAC5C6C7C8GTC11TGCAGCACCAGG
FANCF 脱靶2: GGAGTC6C7C8TC10C11TACAGCACCAGG
FANCF 脱靶3: AGAGGC6C7C8C9TC11TGCAGCACCAGG

图127B

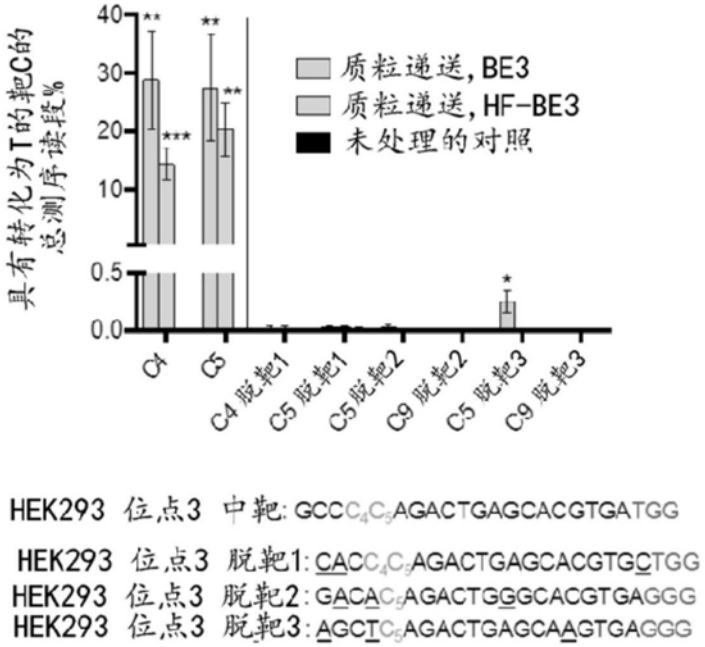
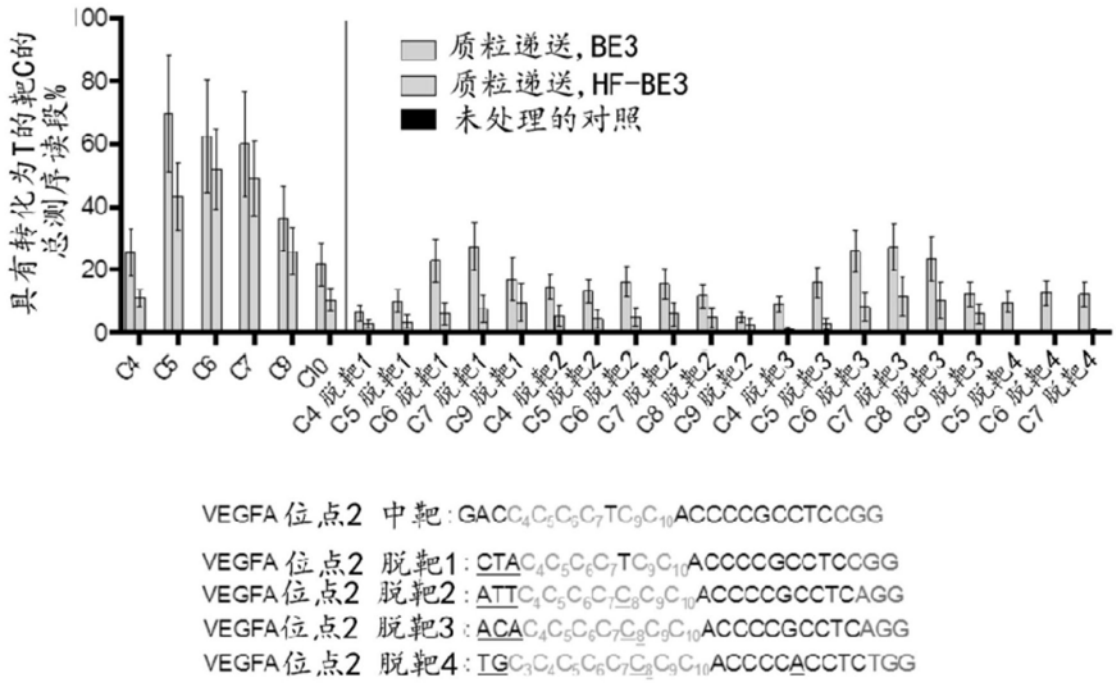

具有转化为T的靶C的总测序读段%
质粒递送,BE3
质粒递送,HF-BE3
未处理的对照
C4
C5
C4 脱靶1
C5 脱靶1
C5 脱靶2
C9 脱靶2
C5 脱靶3
C9 脱靶3
HEK293 位点3 中靶: GCCC4C5AGACTGAGCACGTGATGG
HEK293 位点3 脱靶1: CACC4C5AGACTGAGCACGTGCTGG
HEK293 位点3 脱靶2: GACAC5AGACTGGGCACGTGAGGG
HEK293 位点3 脱靶3: AGCTC5AGACTGAGCAAGTGAGGG

图127C



VEGFA 位点2 中靶 : GACC4C5C6C7TC9C10ACCCCGCCTCCGG

VEGFA 位点2 脱靶1 : CTAC4C5C6C7TC9C10ACCCCGCCTCCGG
VEGFA 位点2 脱靶2 : ATTC4C5C6C7C8C9C10ACCCCGCCTCAGG
VEGFA 位点2 脱靶3 : ACAC4C5C6C7C8C9C10ACCCCGCCTCAGG
VEGFA 位点2 脱靶4 : TGC3C4C5C6C7C8C9C10ACCCCACCTCTGG

图127D

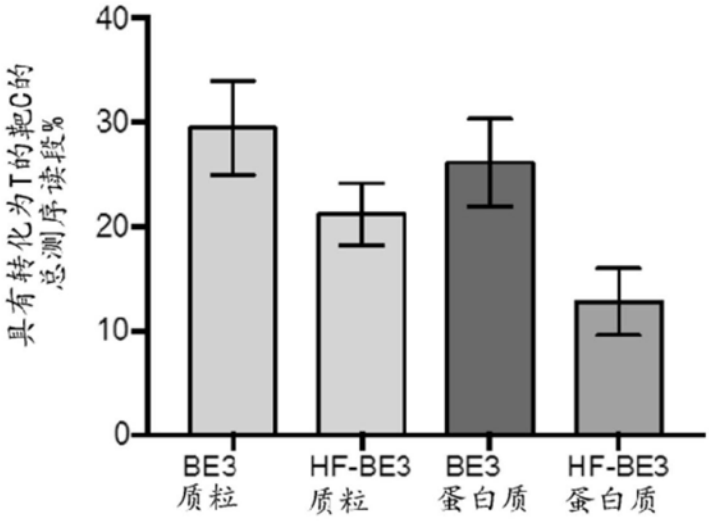
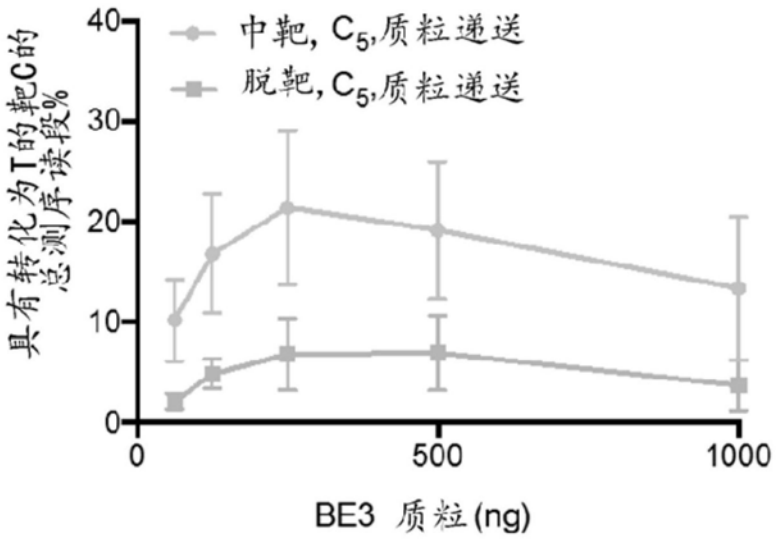


图128A



EMX1: GAGTC5C6GAGCAGAAGAAGAAGGG
EMX1 脱靶1 : GAGTC5TAAGCAGAAGAAGAAGAG

图128B

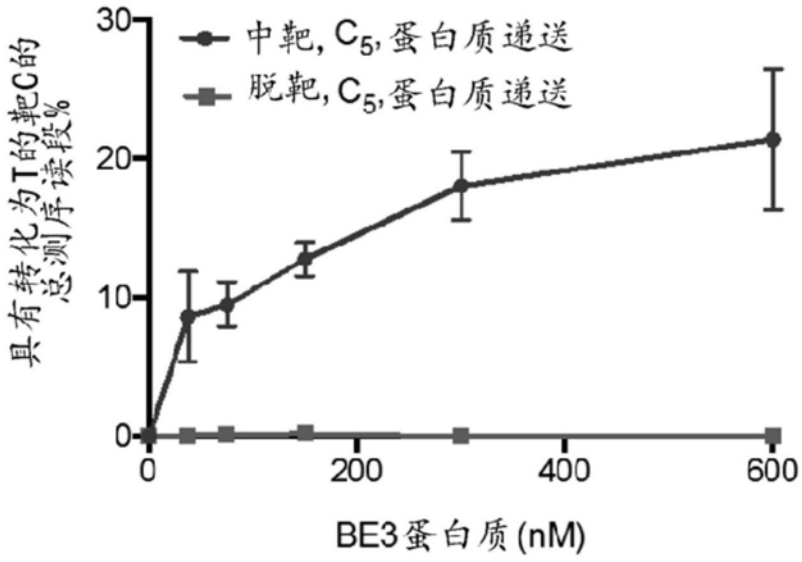


图128C

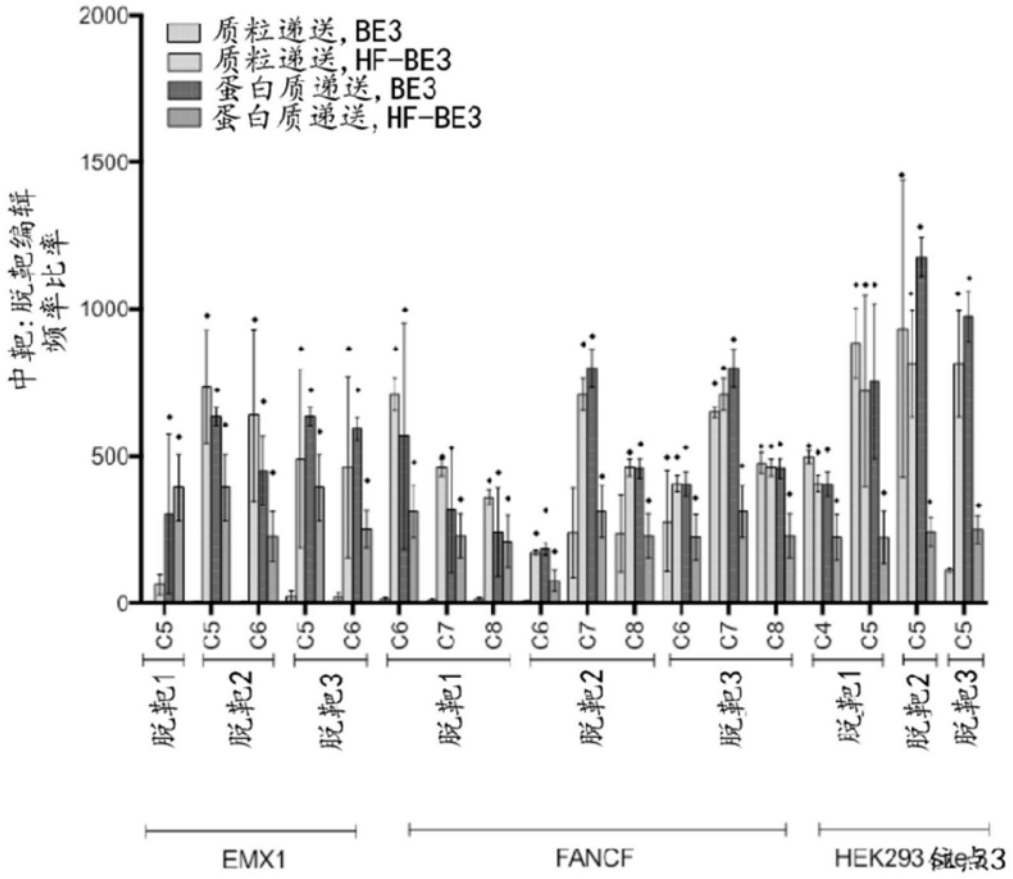

质粒递送, BE3
质粒递送, HF-BE3
蛋白质递送, BE3
蛋白质递送, HF-BE3
中靶:脱靶编辑频率比率
2000
1500
1000
500
0
C5 C5 C6 C5 C6 C6 C7 C8 C6 C7 C8 C6 C7 C8 C4 C5 C5 C5
脱靶1 脱靶2 脱靶3 脱靶1 脱靶2 脱靶3 脱靶1 脱靶2 脱靶3
EMX1
FANCF

图129A

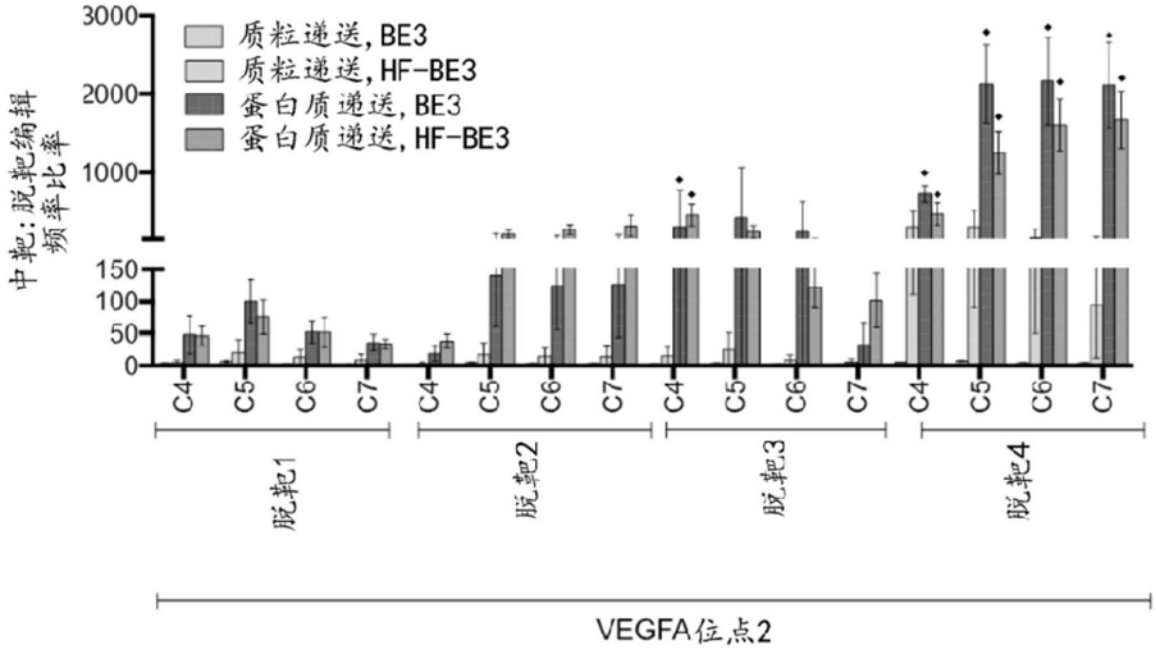
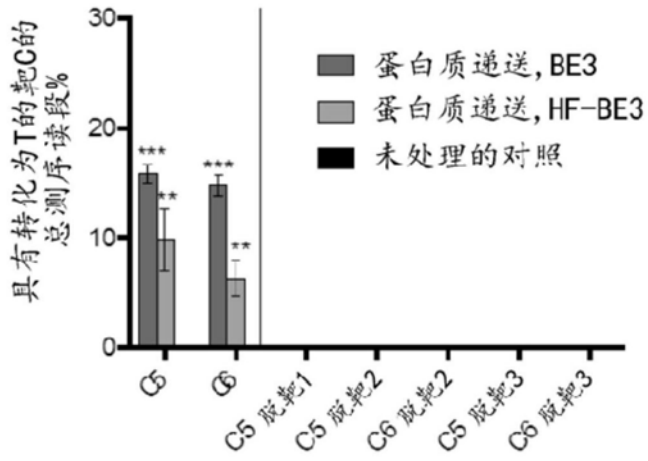

质粒递送, BE3
质粒递送, HF-BE3
蛋白质递送, BE3
蛋白质递送, HF-BE3
中靶:脱靶编辑频率比率
3000
2000
1000
150
100
50
0
C4 C5 C6 C7 C4 C5 C6 C7 C4 C5 C6 C7 C4 C5 C6 C7
脱靶1 脱靶2 脱靶3 脱靶4
VEGFA位点2

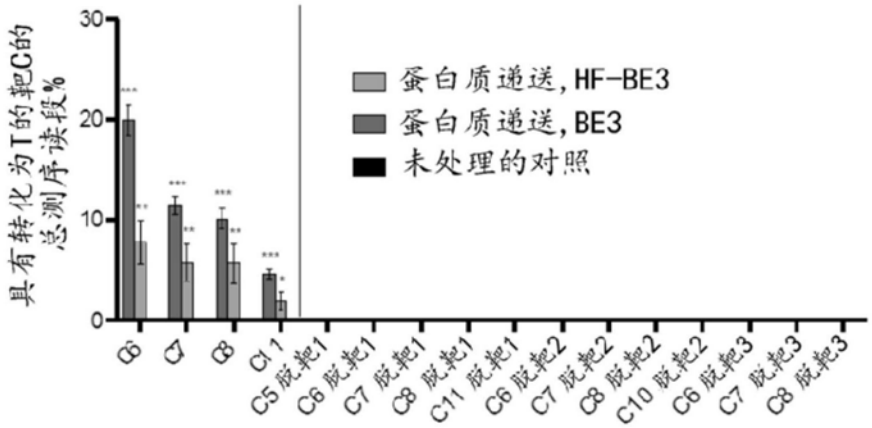
图129B



EMX1 中靶: GAGTC5C6GAGCAGAAGAAGAAGGG

EMX1 脱靶1: GAGTC5TAAGCAGAAGAAGAAGAG
EMX1 脱靶2: GAGGC5C6GAGCAGAAGAAAGACGG
EMX1 脱靶3: GAGTC5C6TAGCAGGAGAAGAAGAG

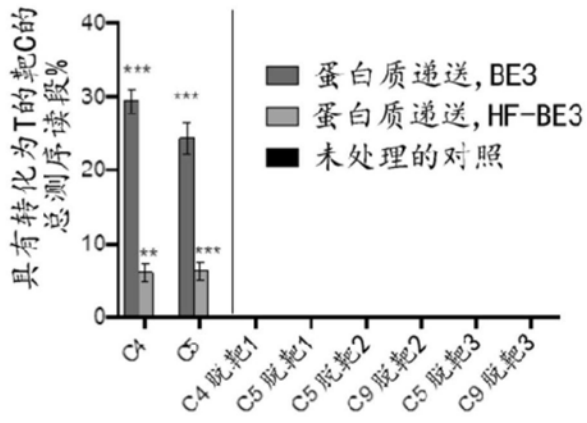
图130A



FANCF 中靶: GGAATC6C7C8TTC11TGCAGCACCTGG

FANCF 脱靶1: GGAAC5C6C7C8GTC11TGCAGCACCAGG
FANCF 脱靶2: GGAGTC6C7C8TC10C11TACAGCACCAGG
FANCF 脱靶3: AGAGGC6C7C8C9TC11TGCAGCACCAGG

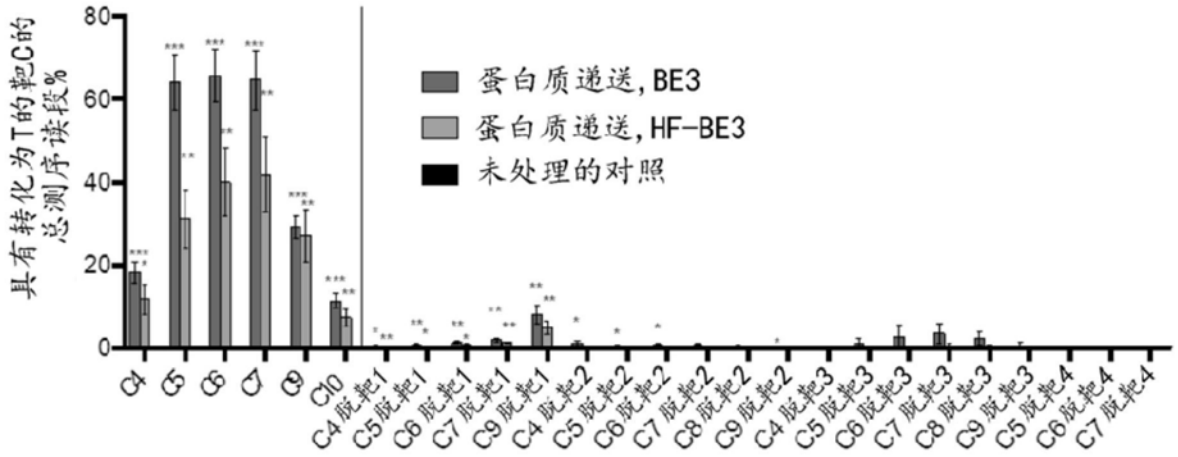
图130B



HEK293 位点3 中靶: GCCC4C5AGACTGAGCACGTGATGG

HEK293 位点3 脱靶1 : CACC4C5AGACTGAGCACGTGCTGG
HEK293 位点3 脱靶2 : GACAC5AGACTGGGCACGTGAGGG
HEK293 位点3 脱靶3 : AGCTC5AGACTGAGCAAGTGAGGG

图130C



VEGFA位点2 中靶: GACC4C5C6C7TC9C10ACCCCGCCTCCGG

VEGFA 位点2 脱靶1: CTAC4C5C6C7TC9C10ACCCCGCCTCCGG
VEGFA 位点2 脱靶2: ATTC4C5C6C7C8C9C10ACCCCGCCTCAGG
VEGFA 位点2 脱靶3: ACAC4C5C6C7C8C9C10ACCCCGCCTCAGG
VEGFA 位点2 脱靶4: TGC3C4C5C6C7C8C9C10ACCCCACCTCTGG

图130D

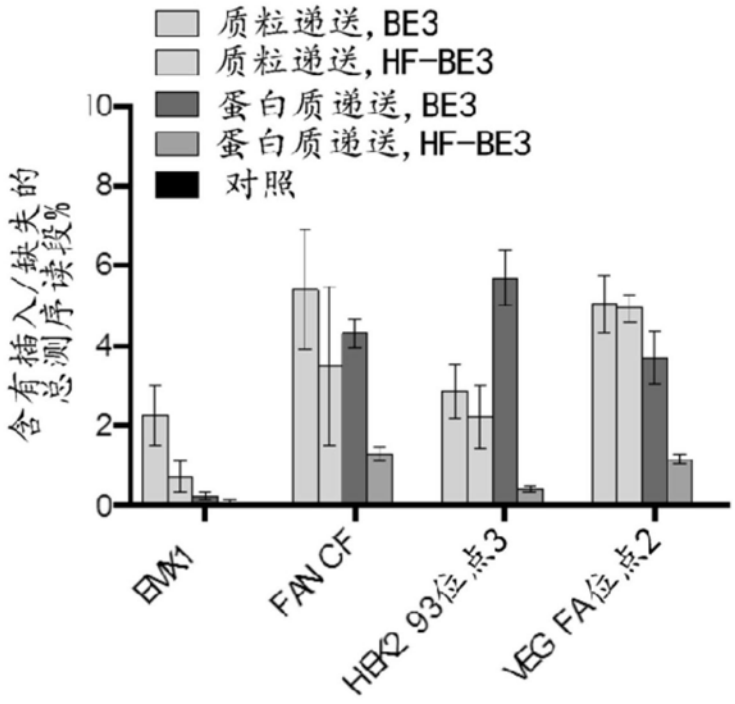

质粒递送, BE3
质粒递送, HF-BE3
蛋白质递送, BE3
蛋白质递送, HF-BE3
对照
含有插入/缺失的总测序读段%
0
2
4
6
8
10
EMX1
FANCF
HEK2 93位点3
VEGFA位点2

图131A

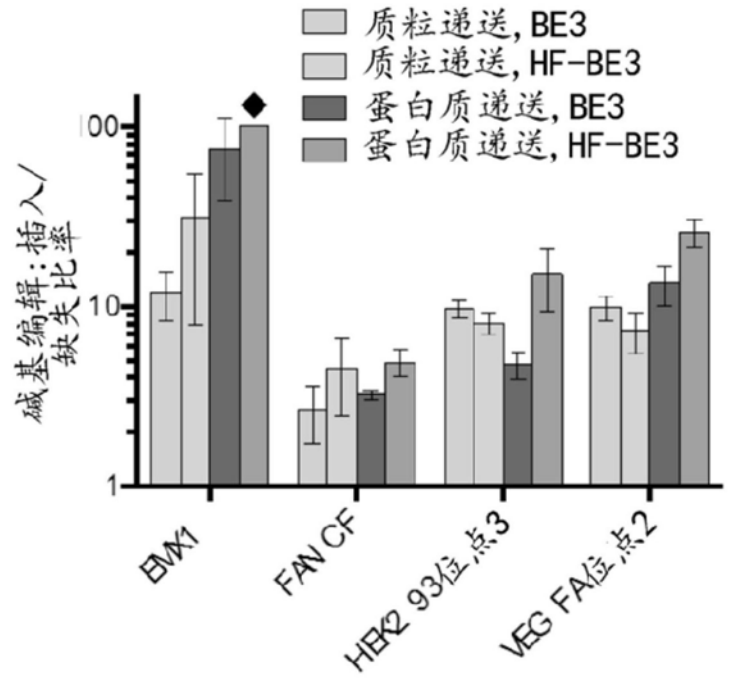

质粒递送, BE3
质粒递送, HF-BE3
蛋白质递送, BE3
蛋白质递送, HF-BE3
碱基编辑:插入/缺失比率
1
10
100
EMX1
FANCF
HEK2 93位点3
VEGFA位点2

图131B

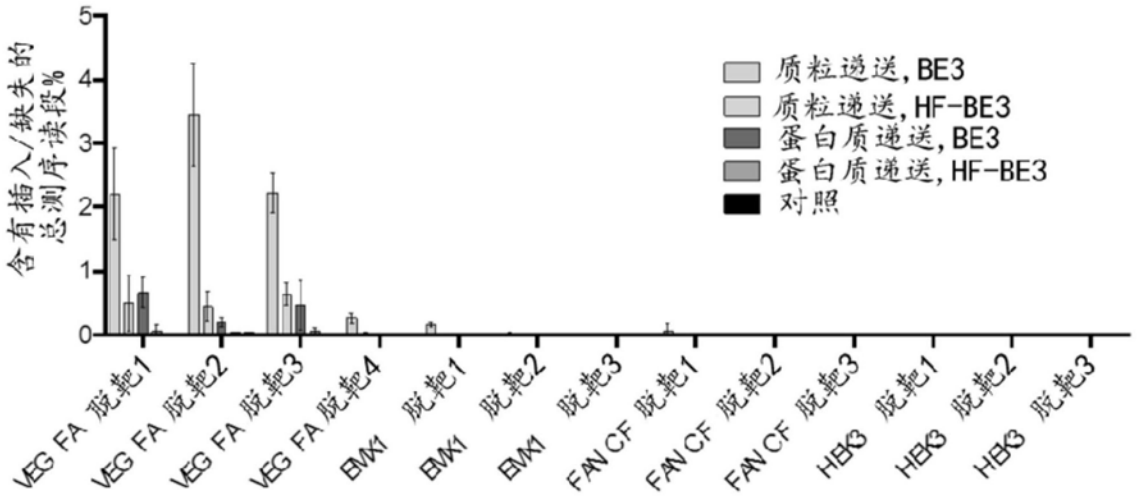

含有插入/缺失的总测序读段%
质粒递送,BE3
质粒递送,HF-BE3
蛋白质递送,BE3
蛋白质递送,HF-BE3
对照
VEGFA 靶点1
VEGFA 靶点2
VEGFA 靶点3
VEGFA 靶点4
EMX1 靶点1
EMX1 靶点2
EMX1 靶点3
FANCF 靶点1
FANCF 靶点2
FANCF 靶点3
HEK3 靶点1
HEK3 靶点2
HEK3 靶点3

图131C

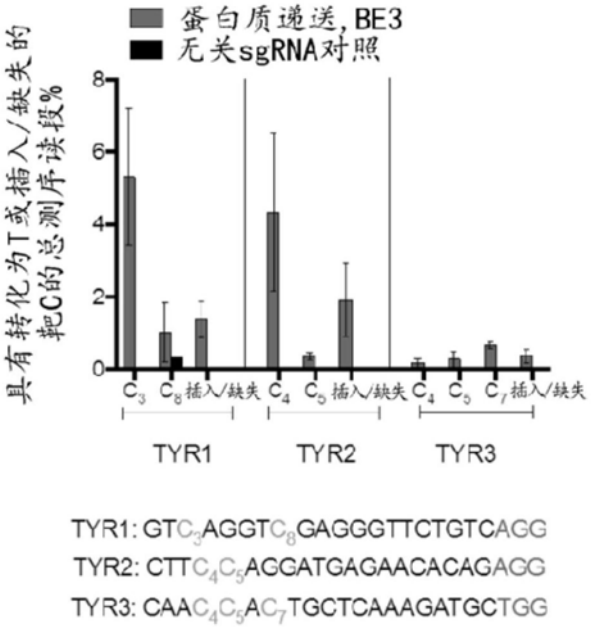

蛋白质递送,BE3
无关sgRNA对照
具有转化为T或插入/缺失的靶C的总测序读段%
C3 C8 插入/缺失 C4 C5 插入/缺失 C4 C5 C7 插入/缺失
TYR1
TYR2
TYR3
TYR1: GTC3AGGTC8GAGGGTTCTGTCAGG
TYR2: CTTC4C5ACGATGAGAACACAGAGG
TYR3: CAAC4C5AC7TGCTCAAAGATGCTGG

图132A

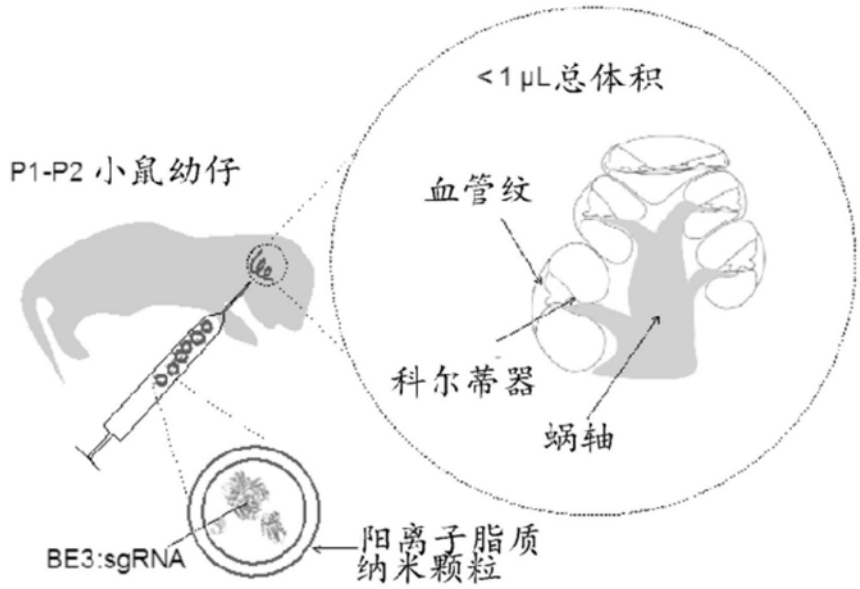
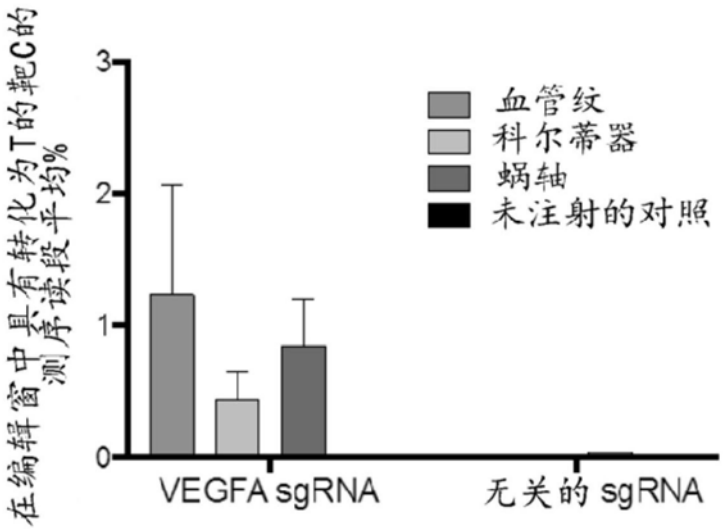


图132B



VEGFA 位点2: GAC$C_4C_5C_6C_7TC_9C_{10}$ACCCCGCCTCCGG

图132C

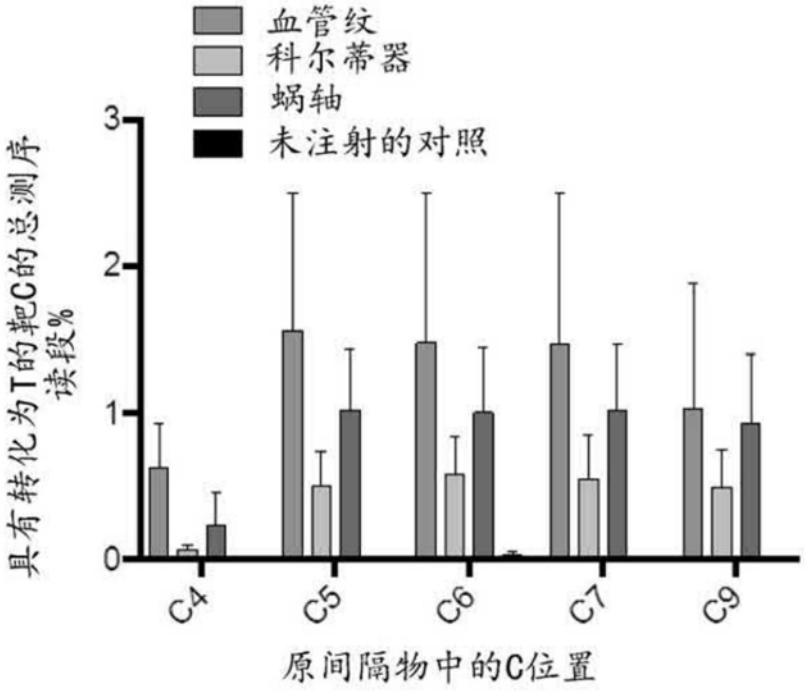

血管纹
科尔蒂器
蜗轴
未注射的对照
具有转化为T的靶C的总测序读段%
3
2
1
0
C4
C5
C6
C7
C9
原间隔物中的C位置

图132D

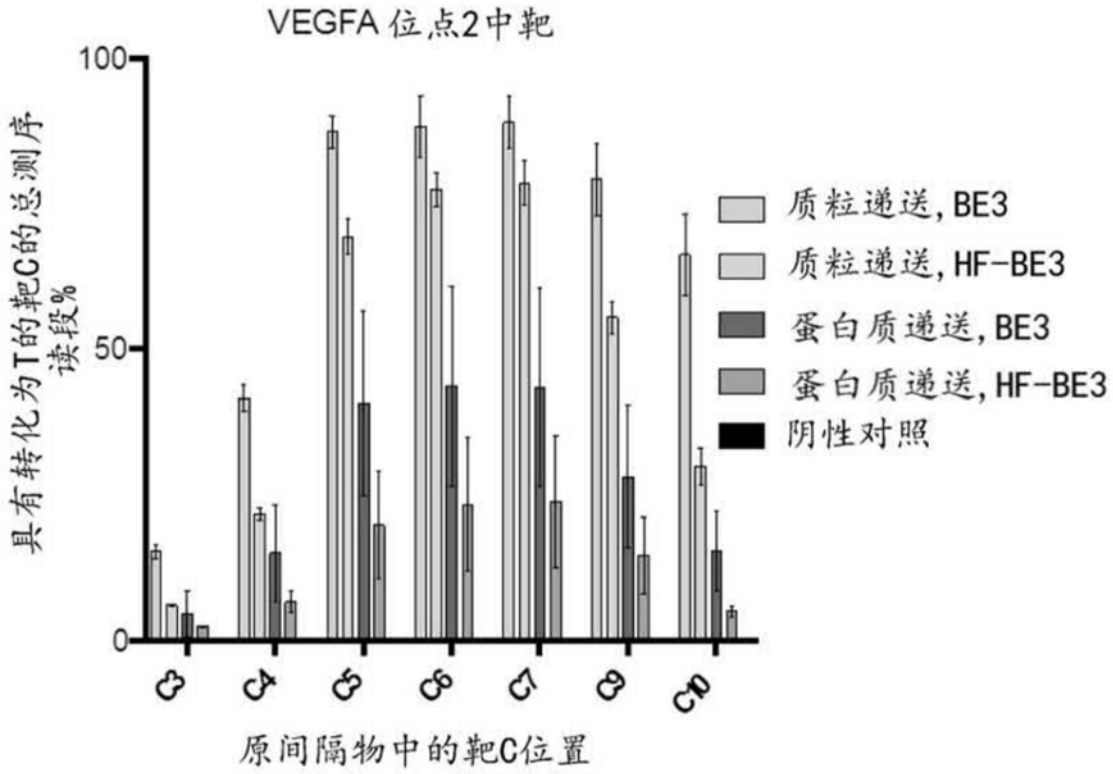

VEGFA 位点2中靶
具有转化为T的靶C的总测序读段%
100
50
0
质粒递送, BE3
质粒递送, HF-BE3
蛋白质递送, BE3
蛋白质递送, HF-BE3
阴性对照
C3
C4
C5
C6
C7
C8
C10
原间隔物中的靶C位置

图133A

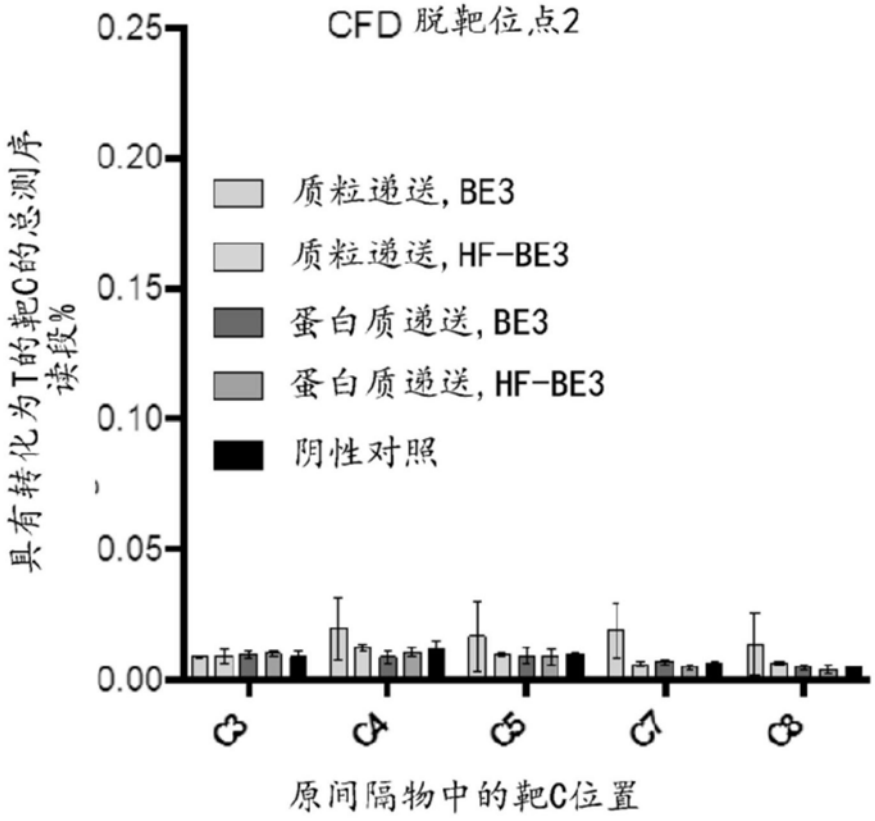

CFD 脱靶位点2
具有转化为T的靶C的总测序读段%
0.25
0.20
0.15
0.10
0.05
0.00
质粒递送, BE3
质粒递送, HF-BE3
蛋白质递送, BE3
蛋白质递送, HF-BE3
阴性对照
3
4
5
6
7
8
原间隔物中的靶C位置

图133B

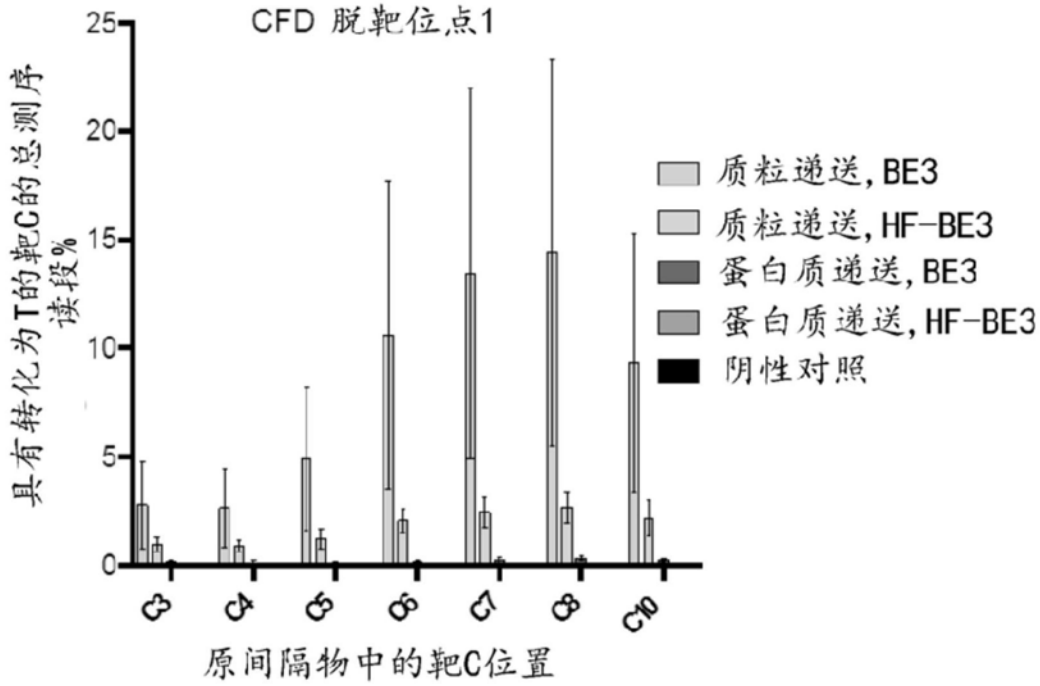

CFD 脱靶位点1
具有转化为T的靶C的总测序读段%
25
20
15
10
5
0
质粒递送, BE3
质粒递送, HF-BE3
蛋白质递送, BE3
蛋白质递送, HF-BE3
阴性对照
3
4
5
6
7
8
10
原间隔物中的靶C位置

图133C

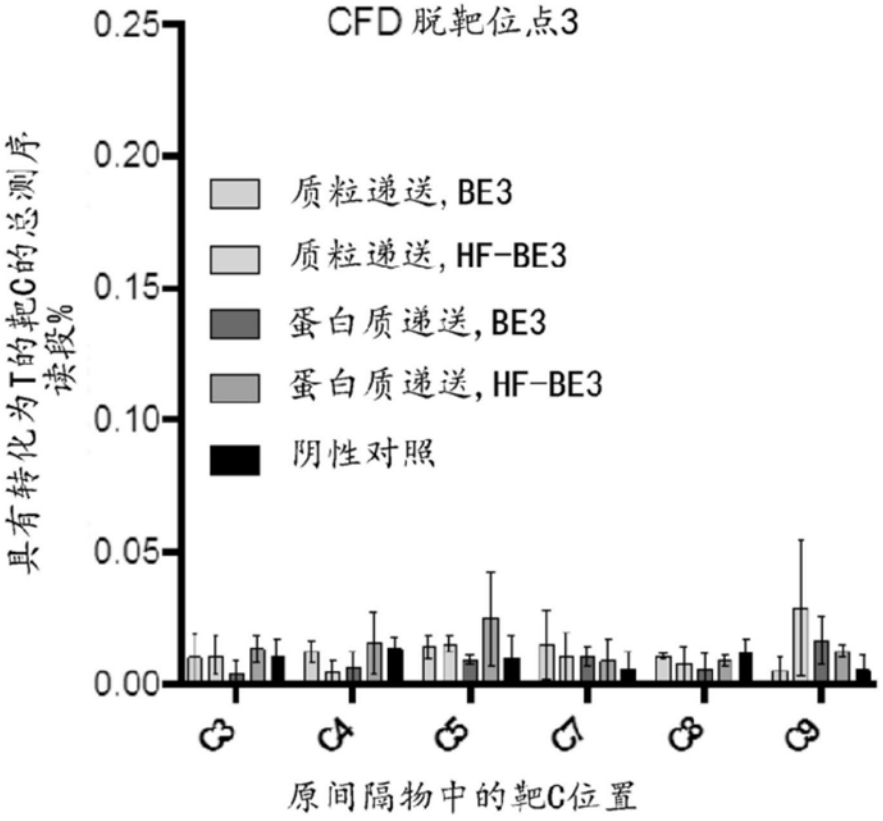

CFD 脱靶位点3
具有转化为T的靶C的总测序读段%
0.25
0.20
0.15
0.10
0.05
0.00
质粒递送, BE3
质粒递送, HF-BE3
蛋白质递送, BE3
蛋白质递送, HF-BE3
阴性对照
3
4
5
6
7
8
原间隔物中的靶C位置

图133D

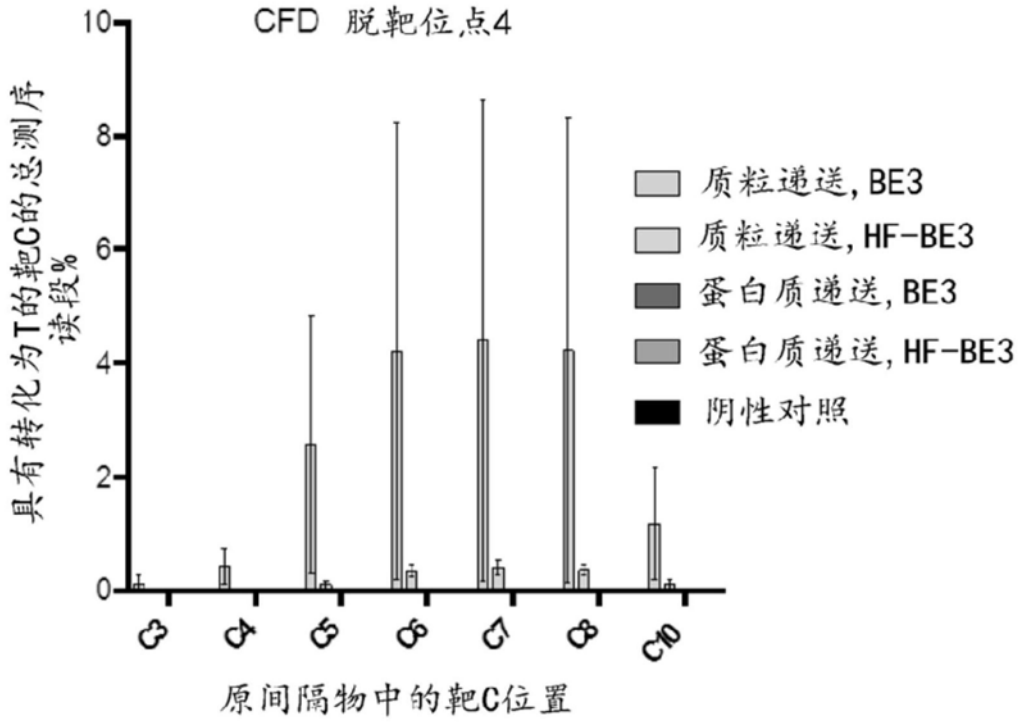

CFD 脱靶位点4
具有转化为T的靶C的总测序读段%
10
8
6
4
2
0
质粒递送, BE3
质粒递送, HF-BE3
蛋白质递送, BE3
蛋白质递送, HF-BE3
阴性对照
3
4
5
6
7
8
9
原间隔物中的靶C位置

图133E

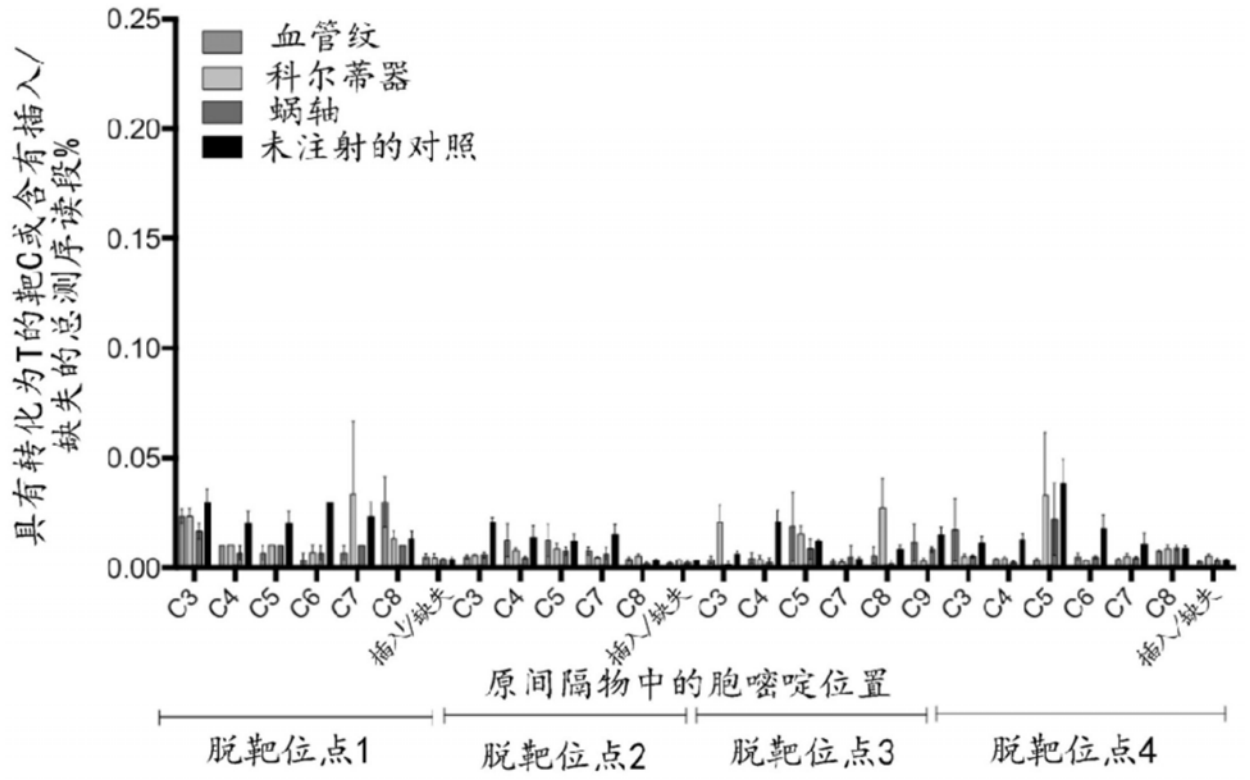

具有转化为T的靶C或含有插入/缺失的总测序读段%
0.25
0.20
0.15
0.10
0.05
0.00
血管纹
科尔蒂器
蜗轴
未注射的对照
C3 C4 C5 C6 C7 C8 插入/缺失 C3 C4 C5 C7 C8 插入/缺失 C3 C4 C5 C7 C8 C9 C3 C4 C5 C6 C7 C8 插入/缺失
原间隔物中的胞嘧啶位置
脱靶位点1
脱靶位点2
脱靶位点3
脱靶位点4

图134A

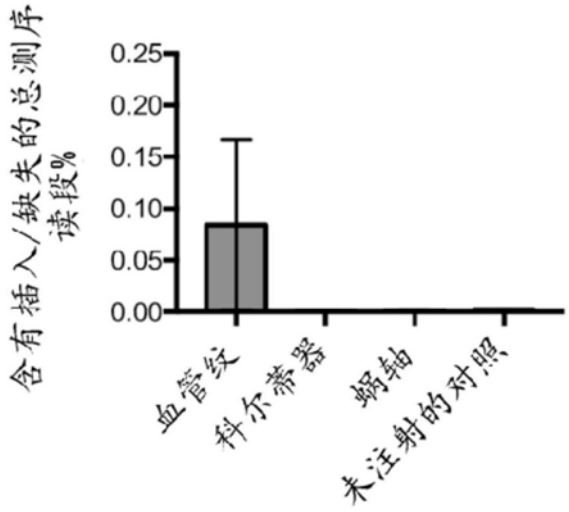

含有插入/缺失的总测序读段%
0.25
0.20
0.15
0.10
0.05
0.00
血管纹
科尔蒂器
蜗轴
未注射的对照

图134B

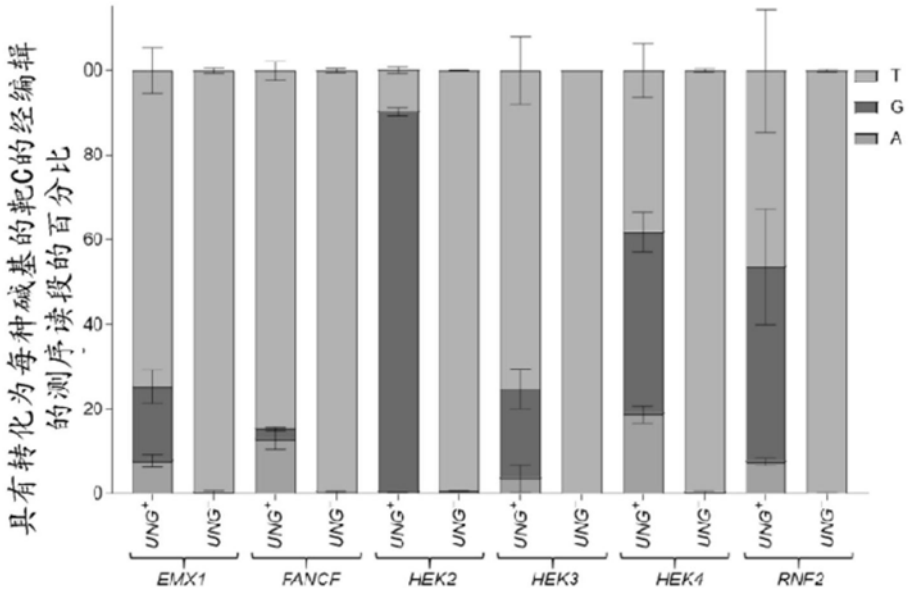


图135A

EMX1: GAGTC$_5$CGAGCAGAAGAAGAAGGG

FANCF: GGAATCC$_7$CTTCTGCAGCACCTGG

HEK2: GAACAC$_6$AAAGCATAGACTGCGGG

HEK3: GGCCC$_5$AGACTGAGCACGTGATGG

HEK4: GGCAC$_5$TGCGGCTGGAGGTCCGGG

RNF2: GTCATC$_6$TTAGTCATTACCTGAGG

图135B

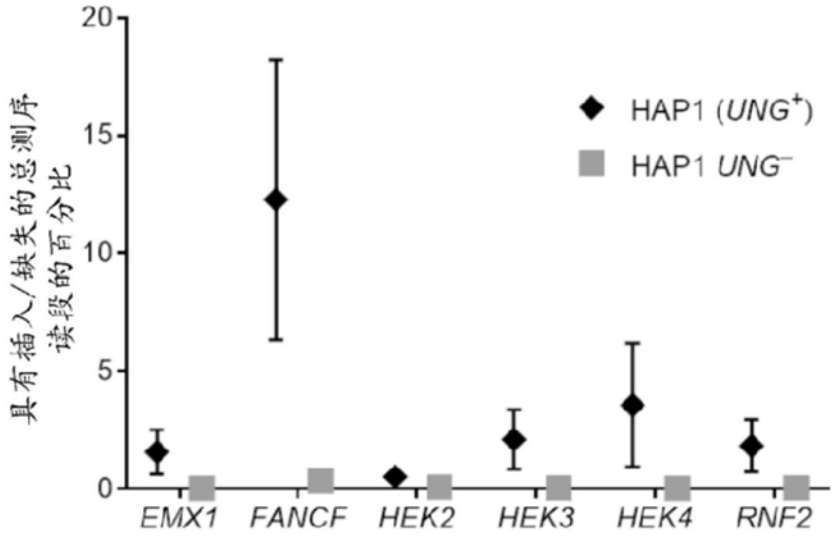
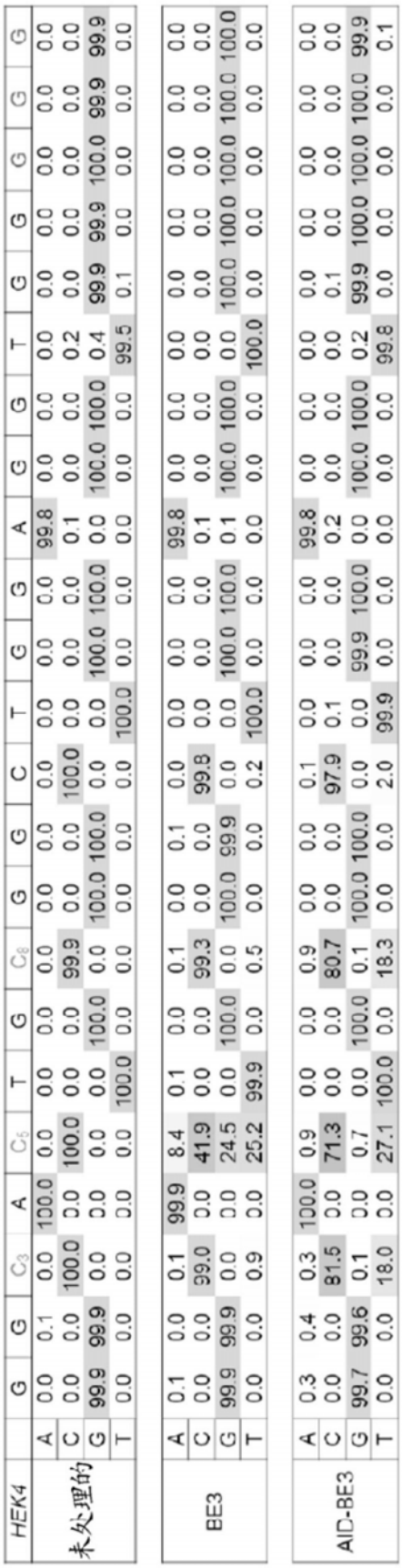

具有插入/缺失的总测序读段的百分比
20
15
10
5
0
HAP1 (UNG+)
HAP1 UNG−
EMX1
FANCF
HEK2
HEK3
HEK4
RNF2

图135C



HEK4		G	G	C3	A	C5	T	G	C8	G	G	C	T	G	G	A	G	G	T	G	G	G	G
未处理的	A	0.0	0.1	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	100.0	0.0	100.0	0.0	0.0	99.9	0.0	0.0	100.0	0.0	0.0	0.0	0.1	0.0	0.0	0.2	0.0	0.0	0.0	0.0
	G	99.9	99.9	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	0.0	100.0	100.0	0.0	100.0	100.0	0.4	99.9	99.9	100.0	99.9	99.9
	T	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	99.5	0.1	0.0	0.0	0.0	0.0
BE3	A	0.1	0.0	0.1	99.9	8.4	0.1	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.0	99.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	99.0	0.0	41.9	0.0	0.0	99.3	0.0	0.0	99.8	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	99.9	99.9	0.0	0.0	24.5	0.0	100.0	0.0	100.0	99.9	0.0	0.0	100.0	100.0	0.1	100.0	100.0	0.0	100.0	100.0	100.0	100.0	100.0
	T	0.0	0.0	0.9	0.0	25.2	99.9	0.0	0.5	0.0	0.0	0.2	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0
AID-BE3	A	0.3	0.4	0.3	100.0	0.9	0.0	0.0	0.9	0.0	0.0	0.1	0.0	0.0	0.0	99.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	81.5	0.0	71.3	0.0	0.0	80.7	0.0	0.0	97.9	0.1	0.0	0.0	0.2	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0
	G	99.7	99.6	0.1	0.0	0.7	0.0	100.0	0.1	100.0	100.0	0.0	0.0	99.9	100.0	0.0	100.0	100.0	0.2	99.9	100.0	100.0	100.0	99.9
	T	0.0	0.0	18.0	0.0	27.1	100.0	0.0	18.3	0.0	0.0	2.0	99.9	0.0	0.0	0.0	0.0	0.0	99.8	0.0	0.0	0.0	0.0	0.1

图136A

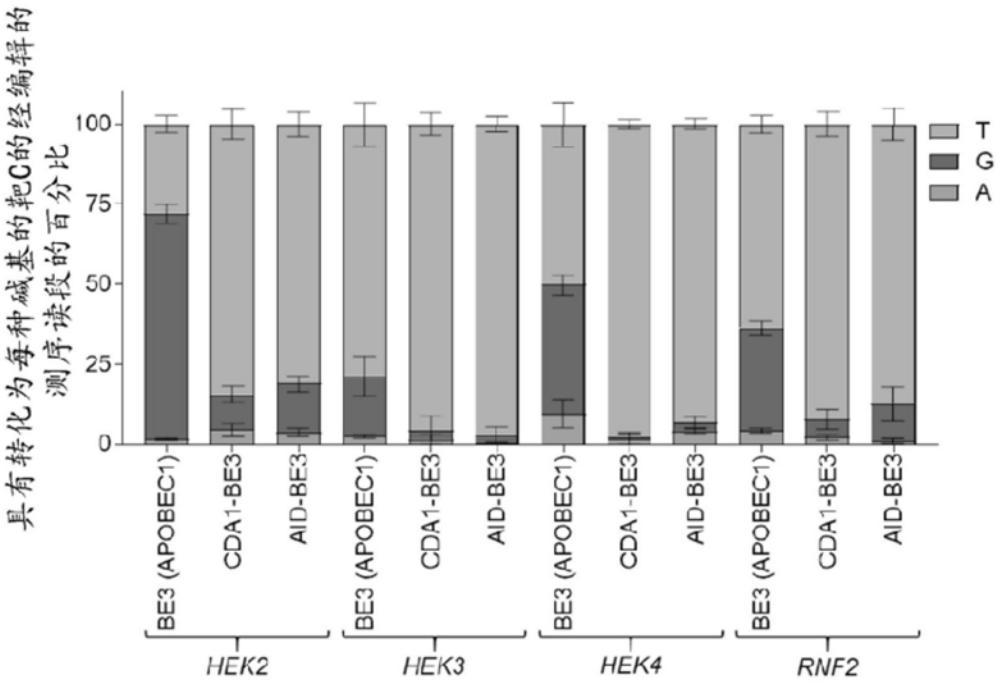


图136B

HEK2: GAACAC$_6$AAAGCATAGACTGCGGG

HEK3: GGCCC$_5$AGACTGAGCACGTGATGG

HEK4: GGCAC$_5$TGCGGCTGGAGGTCCGGG

RNF2: GTCATC$_6$TTAGTCATTACCTGAGG

图136C

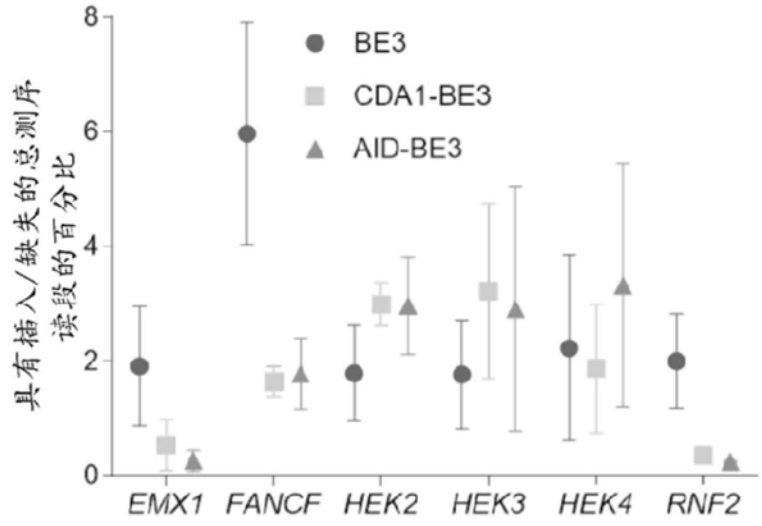


图136D

HEK2: GAAC$_4$AC$_6$AAAGCATAGACTGCGGG

HEK3: GGCC$_4$C$_5$AGACTGAGCACGTGATGG

HEK4: GGCAC$_5$TGCGGCTGGAGGTCCGGG

RNF2: GTC$_3$ATC$_6$TTAGTCATTACCTGAGG

图137A

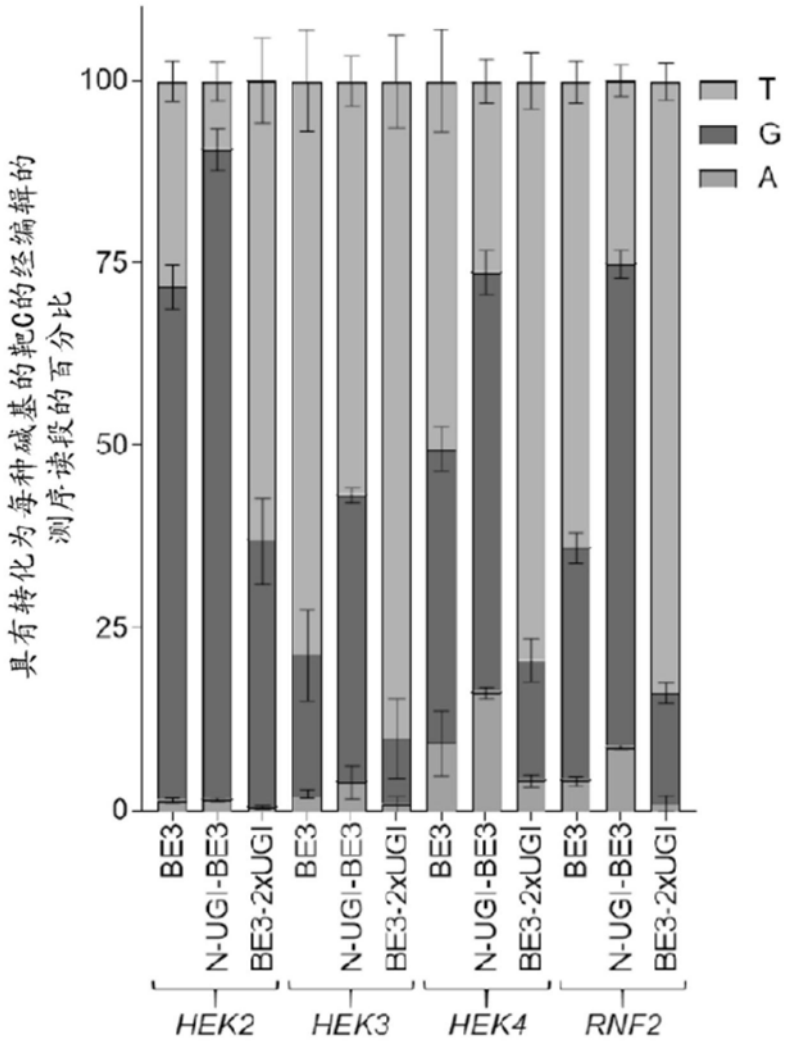


图137B

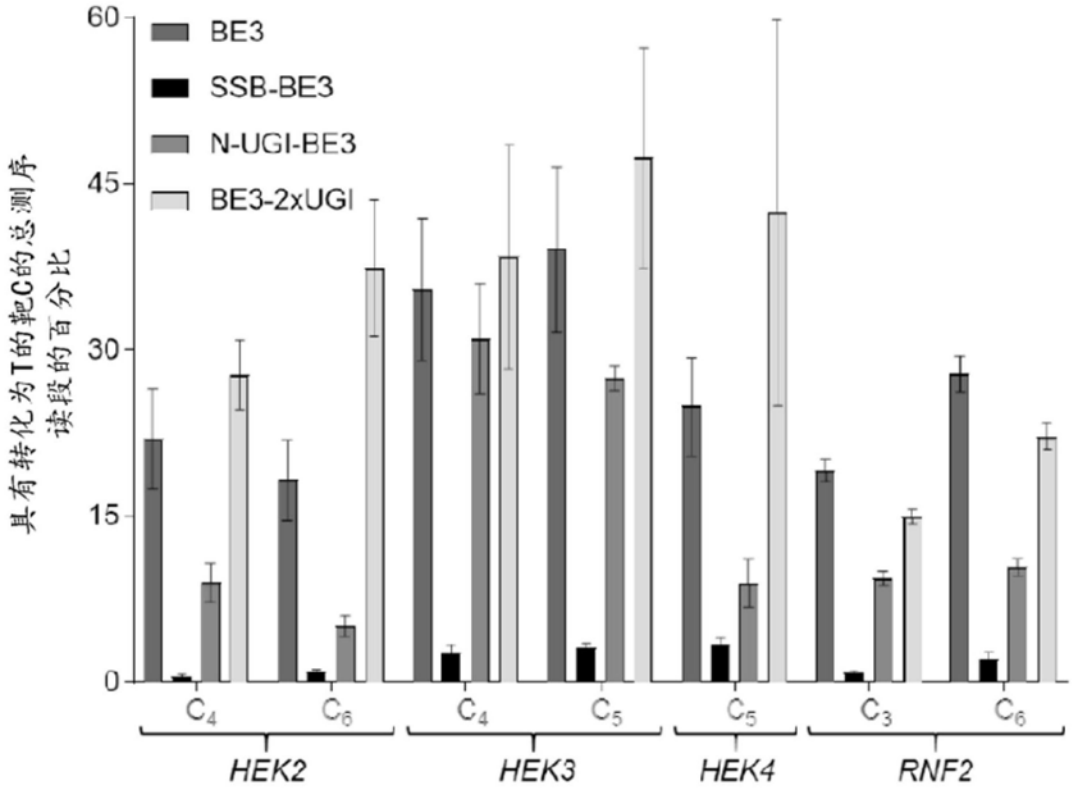


图137C

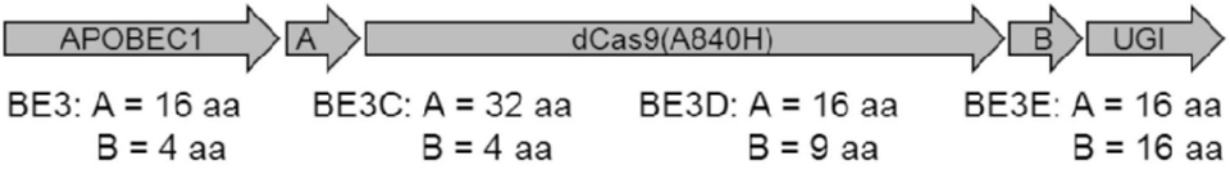


图138A

HEK2: GAAC$_4$AC$_6$AAAGCATAGACTGCGGG

HEK3: GGCC$_4$C$_5$AGACTGAGCACGTGATGG

HEK4: GGCAC$_5$TGCGGCTGGAGGTCCGGG

RNF2: GTC$_3$ATC$_6$TTAGTCATTACCTGAGG

图138B

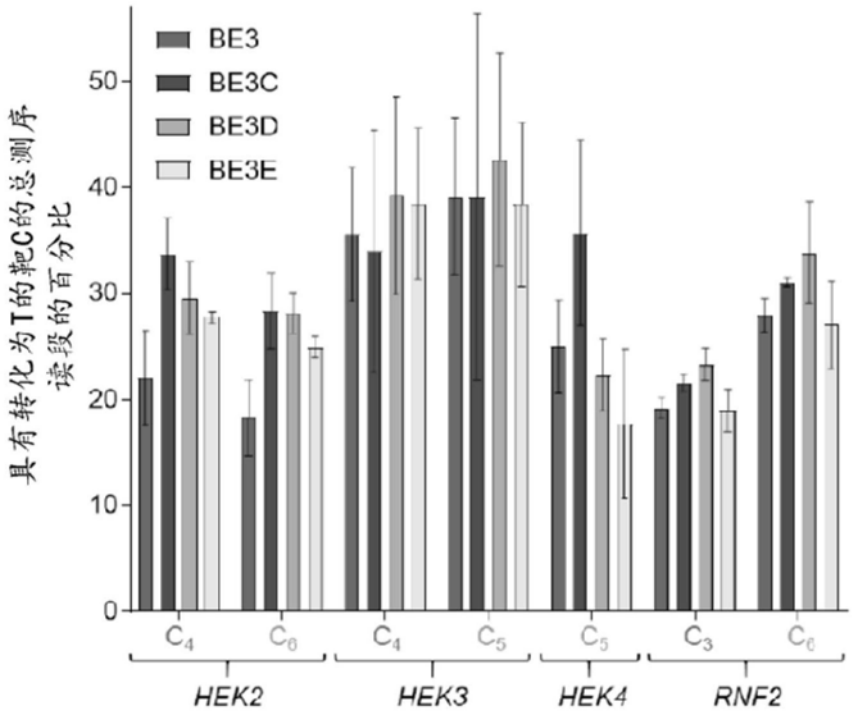

BE3
BE3C
BE3D
BE3E
具有转化为T的靶C的总测序读段的百分比
50
40
30
20
10
0
C4
C6
C4
C5
C5
C3
C6
HEK2
HEK3
HEK4
RNF2

图138C

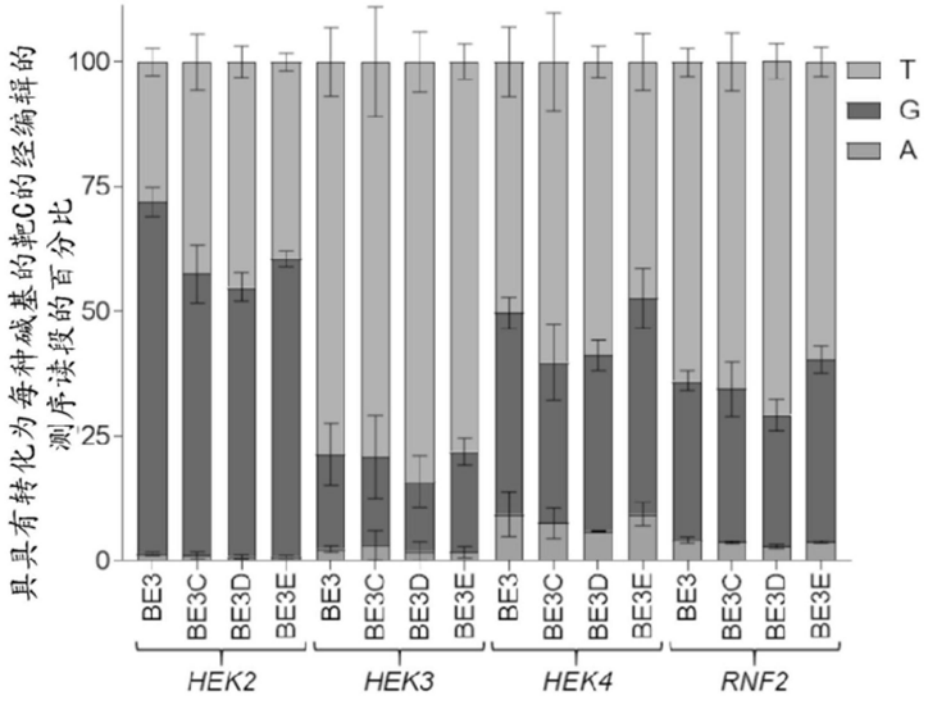

具具有转化为每种碱基的靶C的经编辑的测序读段的百分比
100
75
50
25
0
T
G
A
BE3
BE3C
BE3D
BE3E
HEK2
HEK3
HEK4
RNF2

图138D

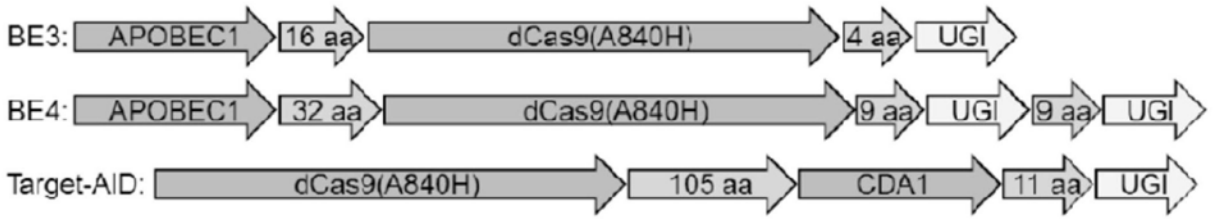


图139A

EMX1: GAGTC$_5$C$_6$GAGCAGAAGAAGAAGGG

FANCF: GGAATC$_6$C$_7$C$_8$TTCTGCAGCACCTGG

HEK2: GAAC$_4$AC$_6$AAAGCATAGACTGCGGG

HEK3: GGCC$_4$C$_5$AGACTGAGCACGTGATGG

HEK4: GGCAC$_5$TGCGGCTGGAGGTCCGGG

RNF2: GTC$_3$ATC$_6$TTAGTCATTACCTGAGG

图139B

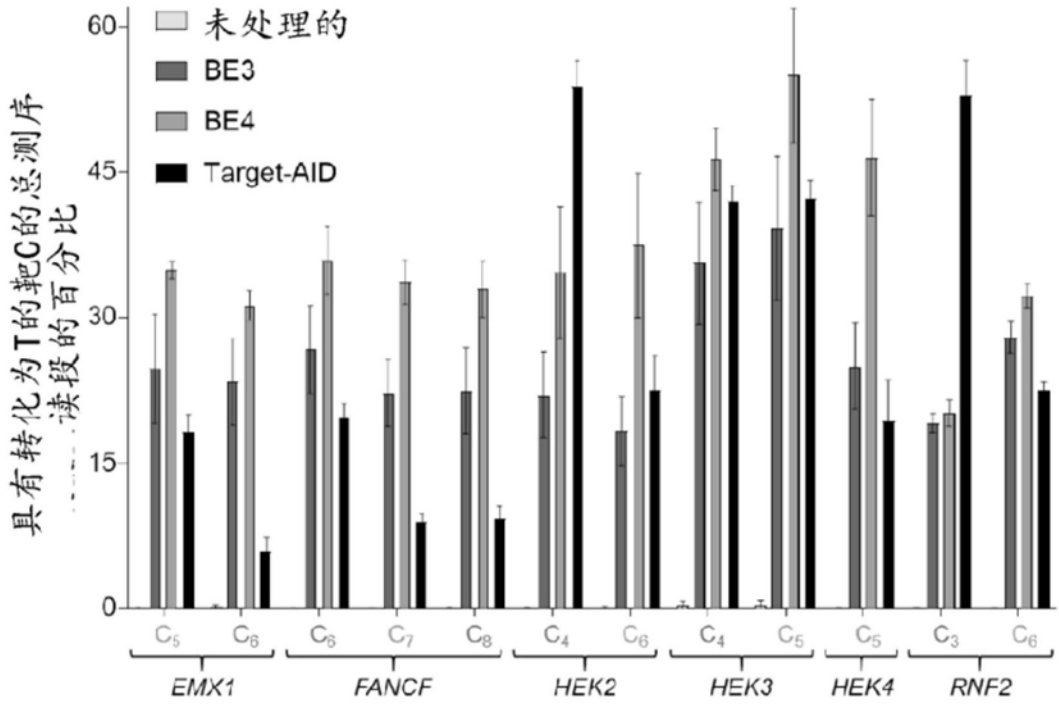


图139C

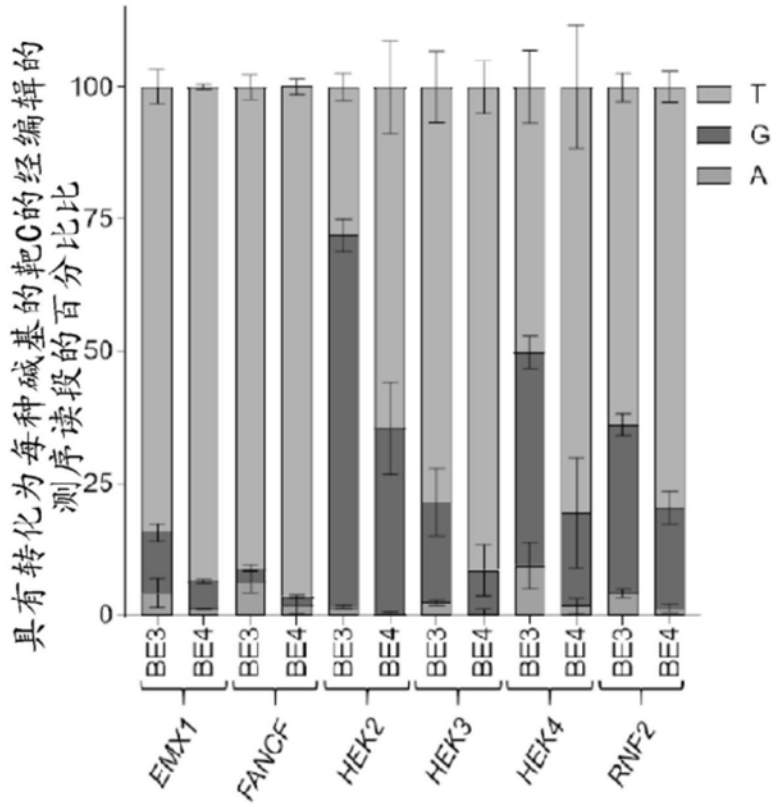


图139D

EMX1: GAGTC$_5$C$_6$GAGCAGAAGAAGAAGGG

FANCF: GGAATC$_6$C$_7$C$_8$TTCTGCAGCACCTGG

HEK2: GAAC$_4$AC$_6$AAAGCATAGACTGCGGG

HEK3: GGC$_3$C$_4$C$_5$AGACTGAGCACGTGATGG

HEK4: GGC$_3$AC$_5$TGC$_8$GGCTGGAGGTCCGGG

RNF2: GTC$_3$ATC$_6$TTAGTCATTACCTGAGG

图140A

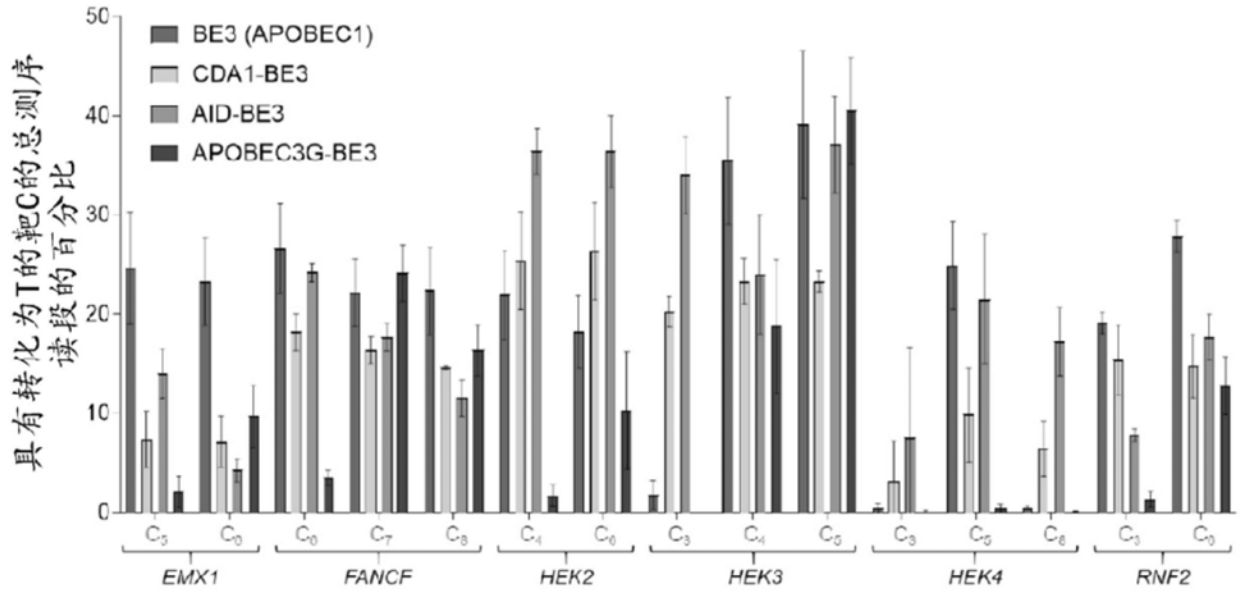


图140B

具有给定序列的干净编辑的读段的百分比

HEK2 原间隔物序列	BE3	CDA1-BE3	AID-BE3
GAAT$_4$AT$_6$AAAGCATAGACTGCGGG	77 ± 1	88 ± 2	92 ± 2
GAAC$_4$AT$_6$AAAGCATAGACTGCGGG	23 ± 1	5 ± 1	4 ± 1
GAAT$_4$AC$_6$AAAGCATAGACTGCGGG	0 ± 0	7 ± 2	4 ± 2

图140C

EMX1: GAGTC$_5$C$_6$GAGCAGAAGAAGAAGGG

FANCF: GGAATC$_6$C$_7$C$_8$TTCTGCAGCACCTGG

HEK2: GAAC$_4$AC$_6$AAAGCATAGACTGCGGG

HEK3: GGCC$_4$C$_5$AGACTGAGCACGTGATGG

HEK4: GGCAC$_5$TGCGGCTGGAGGTCCGGG

RNF2: GTC$_3$ATC$_6$TTAGTCATTACCTGAGG

图141A

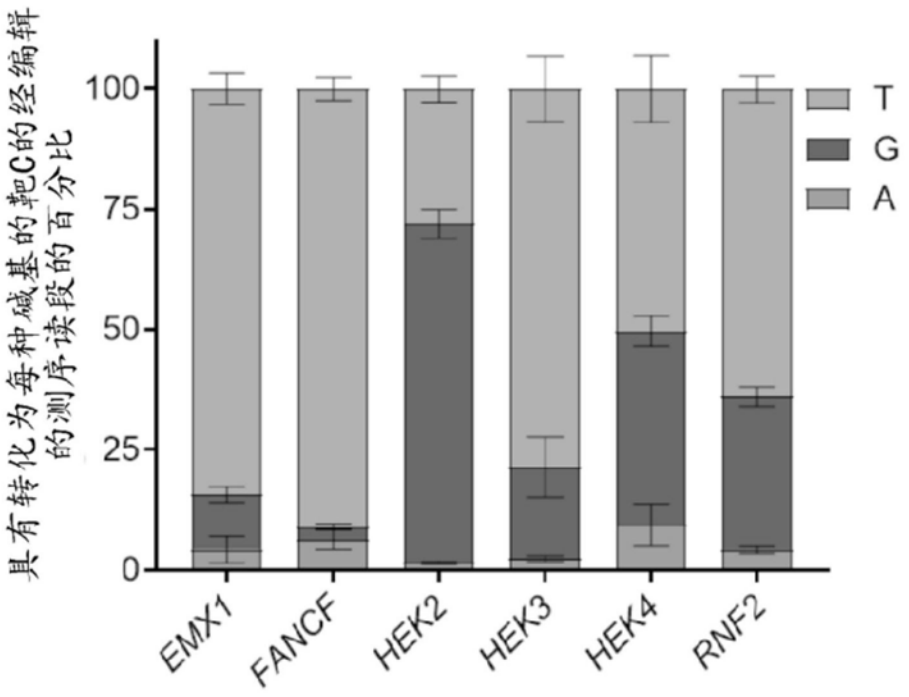

具有转化为每种碱基的靶C的经编辑的测序读段的百分比
100
75
50
25
0
T
G
A
EMX1
FANCF
HEK2
HEK3
HEK4
RNF2

图141B

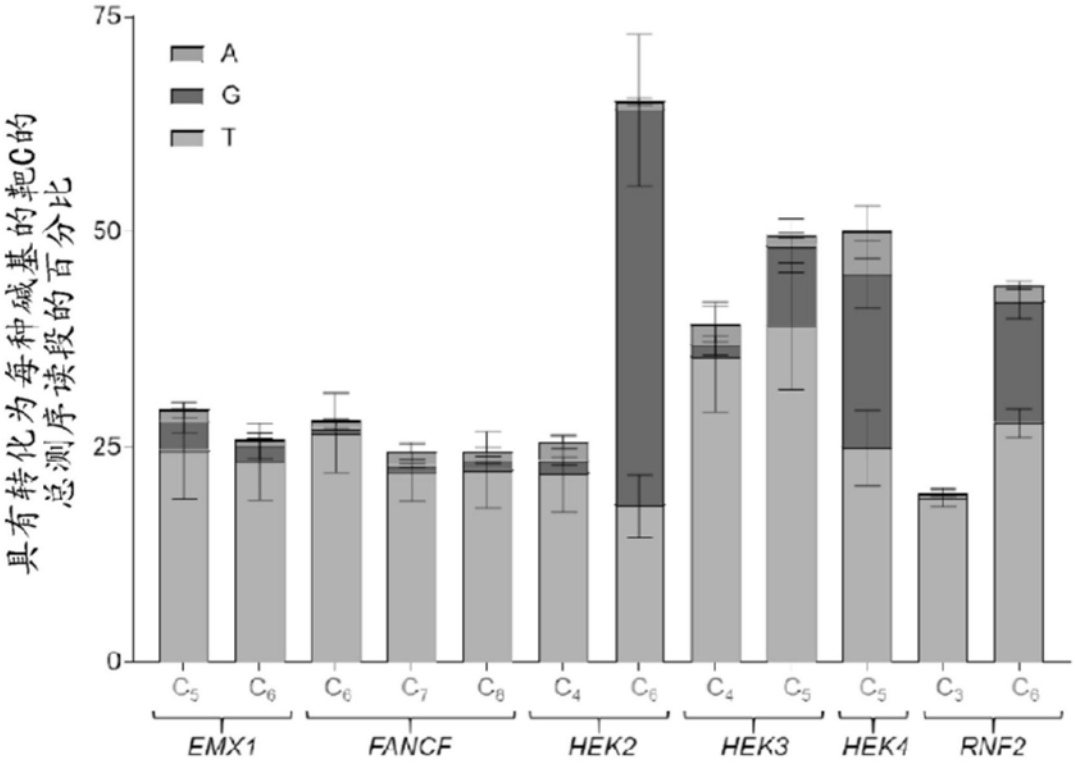

具有转化为每种碱基的靶C的总测序读段的百分比
75
50
25
0
A
G
T
C5
C6
C6
C7
C8
C4
C6
C4
C5
C5
C3
C6
EMX1
FANCF
HEK2
HEK3
HEK1
RNF2

图141C

HEK2: GAAC$_4$AC$_6$AAAGCATAGACTGCGGG

HEK3: GGCC$_4$C$_5$AGACTGAGCACGTGATGG

HEK4: GGCAC$_5$TGCGGCTGGAGGTCCGGG

RNF2: GTC$_3$ATC$_6$TTAGTCATTACCTGAGG

图142A

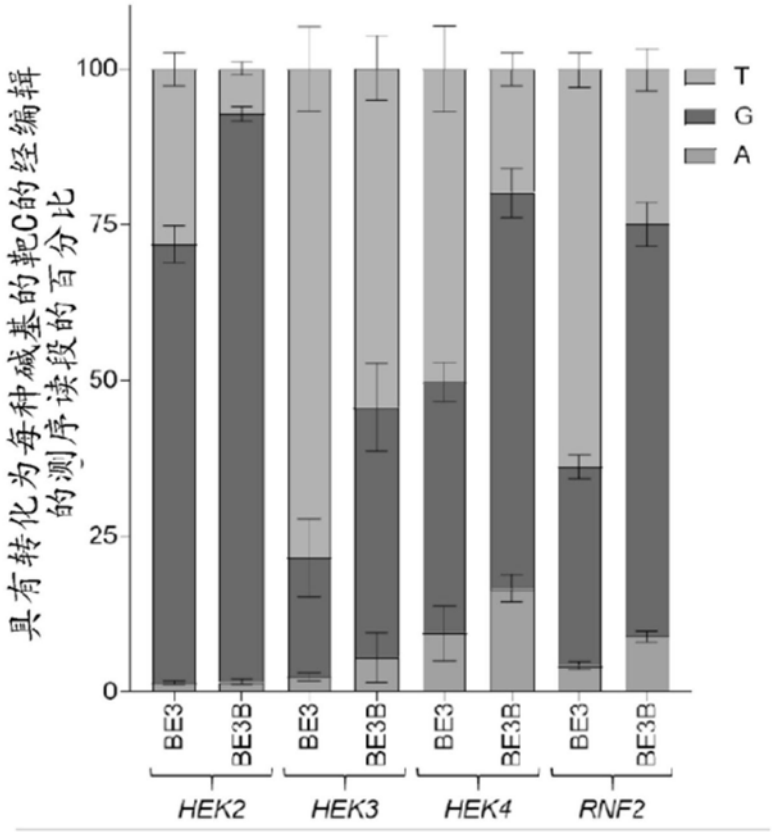


图142B

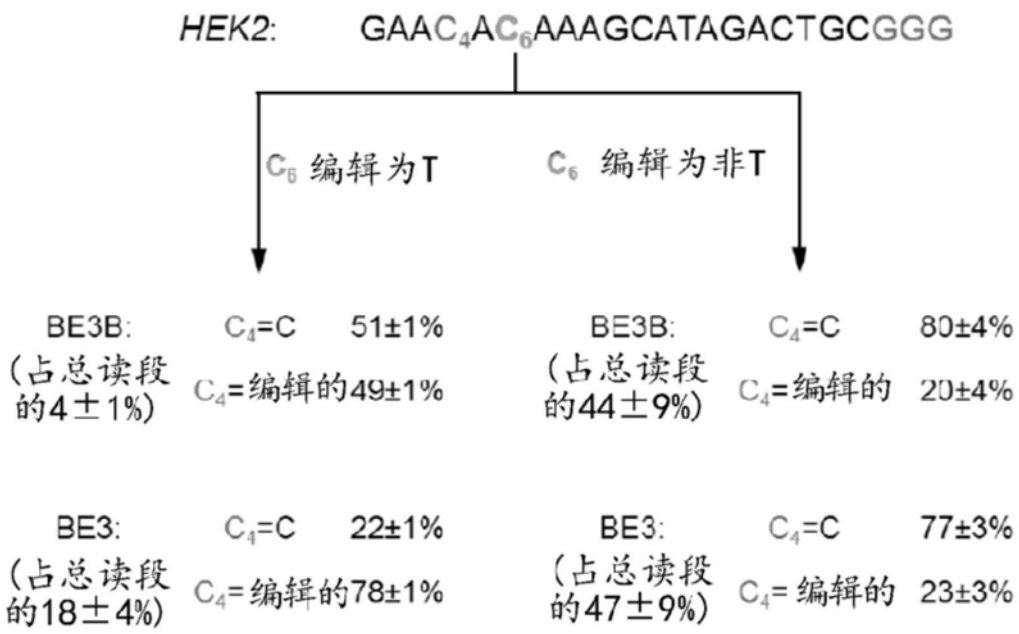

HEK2: GAAC4AC6AAAGCATAGACTGCGGG
C6 编辑为T
C6 编辑为非T
BE3B:
(占总读段的4±1%)
C4=C 51±1%
C4=编辑的49±1%
BE3B:
(占总读段的44±9%)
C4=C 80±4%
C4=编辑的 20±4%
BE3:
(占总读段的18±4%)
C4=C 22±1%
C4=编辑的78±1%
BE3:
(占总读段的47±9%)
C4=C 77±3%
C4=编辑的 23±3%

图142C

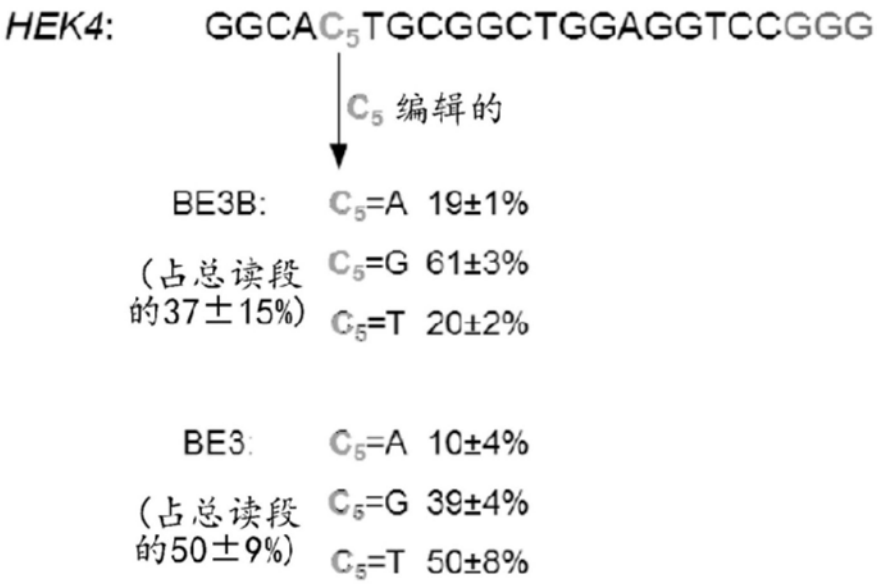

HEK4: GGCAC5TGCGGCTGGAGGTCCGGG
C5 编辑的
BE3B:
(占总读段的37±15%)
C5=A 19±1%
C5=G 61±3%
C5=T 20±2%
BE3:
(占总读段的50±9%)
C5=A 10±4%
C5=G 39±4%
C5=T 50±8%

图142D

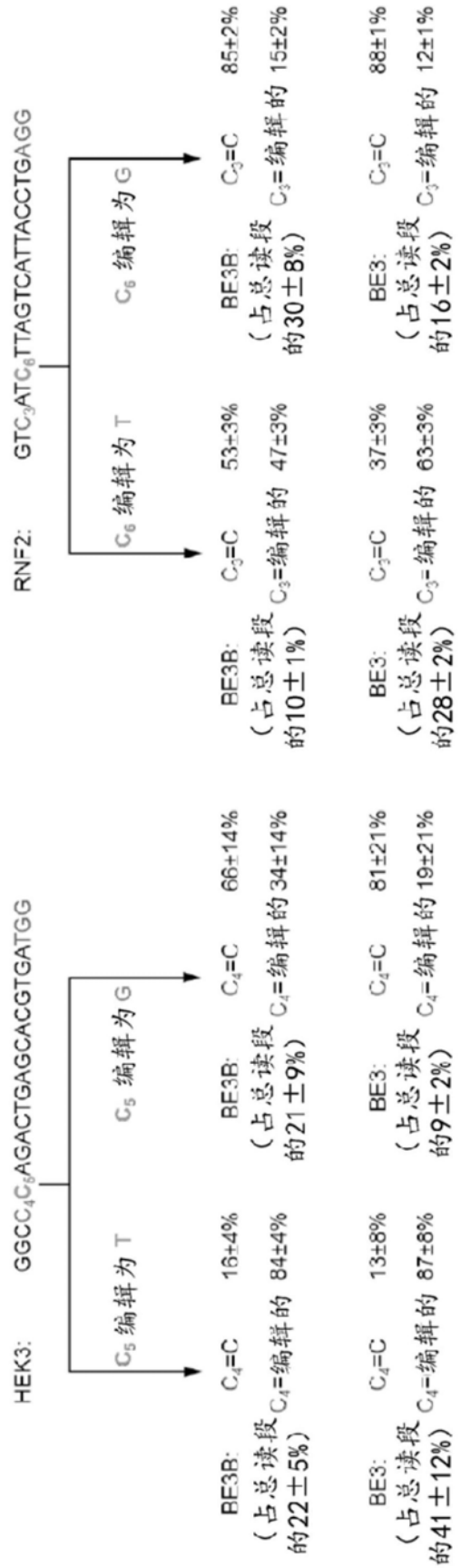

HEK3: GGCC4C5AGACTGAGCACGTGATGG
C5 编辑为 T
C5 编辑为 G
BE3B: C4=C 16±4%
(占总读段的22±5%) C4=编辑的 84±4%
BE3: C4=C 13±8%
(占总读段的41±12%) C4=编辑的 87±8%
BE3B: C4=C 66±14%
(占总读段的21±9%) C4=编辑的 34±14%
BE3: C4=C 81±21%
(占总读段的9±2%) C4=编辑的 19±21%
RN=2: GTC3ATC6TTAGTCATTACCTGAGG
C6 编辑为 T
C6 编辑为 G
BE3B: C3=C 53±3%
(占总读段的10±1%) C3=编辑的 47±3%
BE3: C3=C 37±3%
(占总读段的28±2%) C3=编辑的 63±3%
BE3B: C3=C 85±2%
(占总读段的30±8%) C3=编辑的 15±2%
BE3: C3=C 88±1%
(占总读段的16±2%) C3=编辑的 12±1%

图143

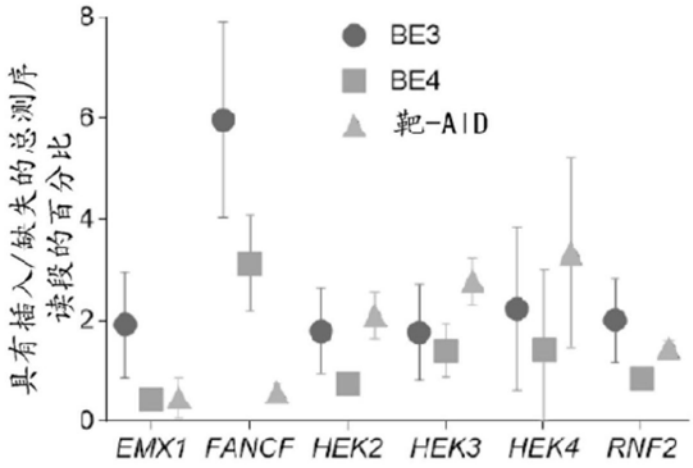

具有插入/缺失的总测序读段的百分比
8
6
4
2
0
BE3
BE4
靶-AID
EMX1
FANCF
HEK2
HEK3
HEK4
RNF2

图144A

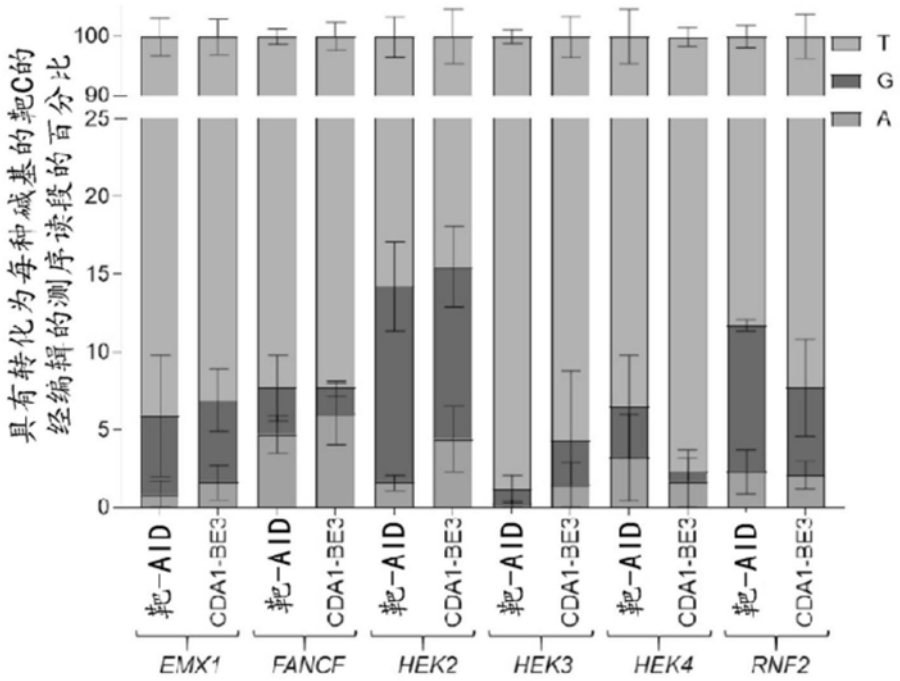

具有转化为每种碱基的靶C的经编辑的测序读段的百分比
100
90
25
20
15
10
5
0
T
G
A
靶-AID
CDA1-BE3
EMX1
FANCF
HEK2
HEK3
HEK4
RNF2

图144B

EMX1: GAGTC$_5$CGAGCAGAAGAAGAAGGG

FANCF: GGAATCC$_7$CTTCTGCAGCACCTGG

HEK2: GAACAC$_6$AAAGCATAGACTGCGGG

HEK3: GGCCC$_5$AGACTGAGCACGTGATGG

HEK4: GGCAC$_5$TGCGGCTGGAGGTCCGGG

RNF2: GTCATC$_6$TTAGTCATTACCTGAGG

图144C

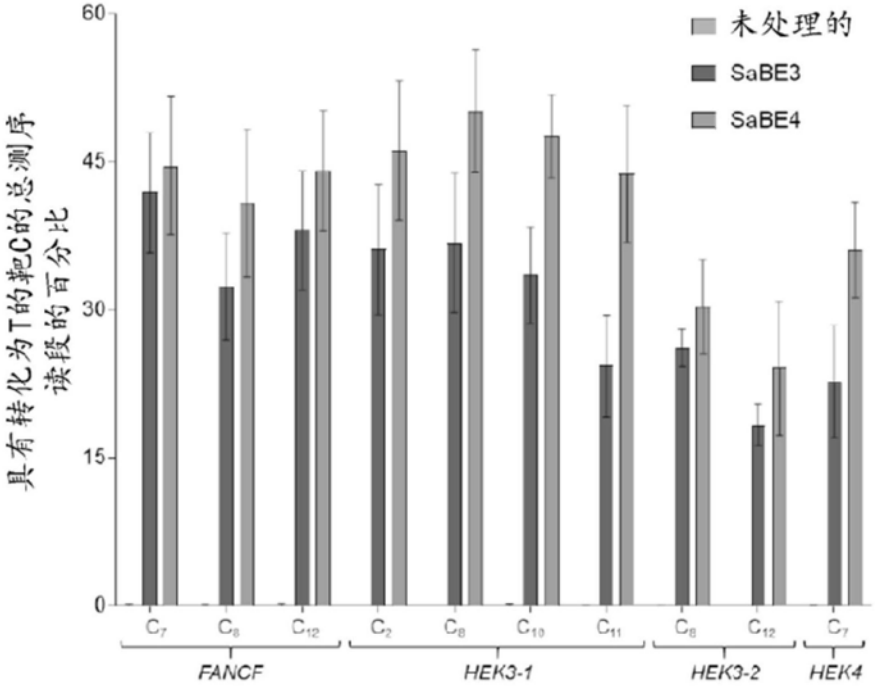


图145A

FANCF: GATGTTC$_7$C$_8$AATC$_{12}$AGTACGCAGAGAGT

HEK3-1: TC$_2$TGCTTC$_8$TC$_{10}$C$_{11}$AGCCCTGGCCTGGGT

HEK3-2: ACGTGCTC$_8$AGTC$_{12}$TGGGCCCCAAGGAT

HEK4: GTGGCAC$_7$TGCGGCTGGAGGTGGGGGT

图145B

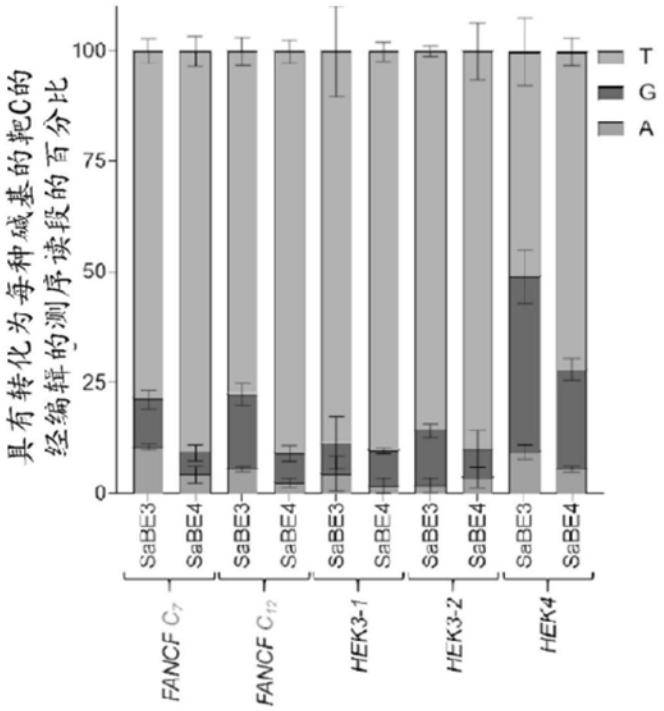

具有转化为每种碱基的靶C的经编辑的测序读段的百分比
100
75
50
25
0
T
G
A
SaBE3
SaBE4
FANCF C7
FANCF C12
HEK3-1
HEK3-2
HEK4

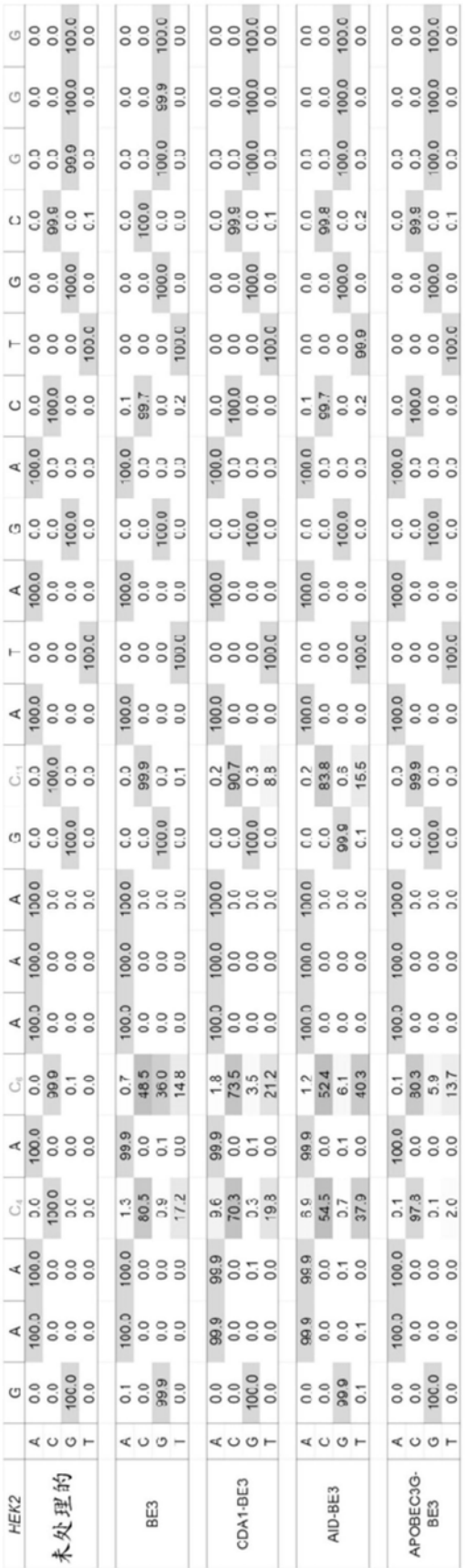
图145C

EMX1		G	A	G	T	C_6	C_5	G	A	G	C_{10}	A	G	A	G	A	A	G	A	G	G	G
未处理的	A	0.0	99.9	0.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	99.9	0.0	99.9	100.0	0.0	100.0	100.0	99.9	99.9	0.0	0.0
	C	0.0	0.0	0.0	0.0	99.9	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.0
	G	100.0	0.1	100.0	0.0	0.0	0.0	100.0	0.0	99.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	99.0	100.0
	T	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0
BE3	A	0.0	100.0	0.0	0.0	2.0	0.8	0.0	99.5	0.0	0.0	100.0	0.0	100.0	100.0	0.0	100.0	100.0	100.0	100.0	0.0	0.0
	C	0.0	0.0	0.0	0.0	72.6	76.2	0.0	0.5	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	100.0	0.0	100.0	0.0	3.0	1.1	100.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	100.0
	T	0.0	0.0	0.0	100.0	22.3	21.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
CDA1 BE3	A	0.1	100.0	0.0	0.0	0.2	0.0	0.1	99.9	0.0	0.0	100.0	0.0	100.0	100.0	0.0	100.0	100.0	100.0	100.0	0.0	0.0
	C	0.0	0.0	0.0	0.0	91.5	92.8	0.0	0.0	0.0	98.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	99.9	0.0	99.9	0.0	0.4	0.1	99.8	0.0	99.9	0.0	0.0	99.9	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	100.0
	T	0.0	0.0	0.0	100.0	7.9	7.1	0.0	0.0	0.0	1.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
AID-BE3	A	0.0	100.0	0.2	0.0	0.2	0.1	0.3	100.0	0.0	0.1	100.0	0.0	100.0	100.0	0.0	100.0	100.0	100.0	100.0	0.0	0.0
	C	0.0	0.0	0.0	0.0	85.8	95.2	0.0	0.0	0.3	94.5	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	99.9	0.0	99.8	0.0	0.7	0.0	99.7	0.0	99.6	0.3	0.0	99.7	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	99.9
	T	0.0	0.0	0.0	100.0	13.4	4.6	0.0	0.0	0.0	5.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
APOBEC3C-BE3	A	0.0	100.0	0.0	0.0	0.0	0.7	0.0	100.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	100.0	100.0	100.0	100.0	0.0	0.0
	C	0.0	0.0	0.0	0.0	98.9	88.0	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	100.0	0.0	100.0	0.0	0.2	2.3	99.9	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	100.0
	T	0.0	0.0	0.0	100.0	1.0	8.9	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

图146

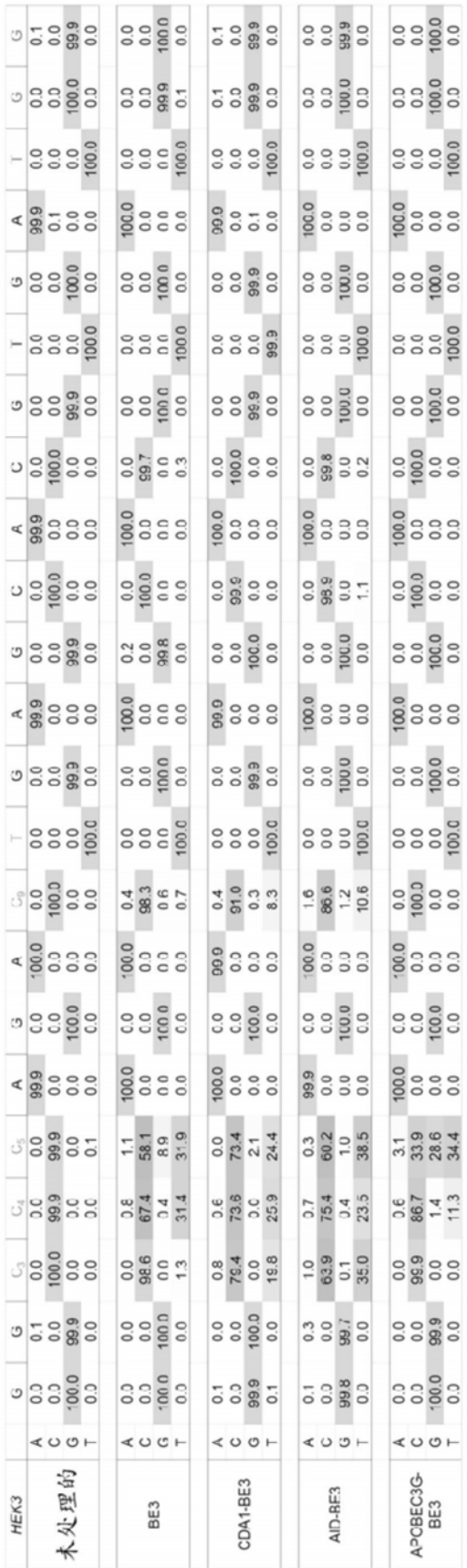
FANCF		G	G	A	A	T	C_5	C_7	C_9	T	T	C_{11}	T	G	C	A	G	C	C	A	C	C	T	G	G
未处理的	A	0.1	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	100.0	0.0	100.0	100.0	0.0	0.0	0.0
	G	99.9	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0
	T	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0
BE3	A	0.0	0.1	100.0	100.0	0.0	1.0	1.4	1.4	0.0	0.0	0.2	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	0.0	0.0	0.0	70.7	75.0	74.4	0.0	0.0	88.8	0.0	0.0	99.9	0.0	0.0	100.0	0.0	0.0	100.0	99.9	0.0	0.0	0.0
	G	100.0	99.9	0.0	0.0	0.0	0.4	0.8	0.9	0.0	0.0	0.1	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	100.0
	T	0.0	0.0	0.0	0.0	100.0	28.0	22.9	23.4	100.0	100.0	10.9	100.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	100.0	0.0	0.0
CDA1-BE3	A	0.1	0.1	100.0	100.0	0.0	0.3	0.8	0.3	0.0	0.0	0.0	0.0	0.1	0.0	100.0	0.1	0.0	100.0	0.0	0.0	0.0	0.0	0.1	0.1
	C	0.0	0.0	0.0	0.0	0.0	82.4	83.5	84.7	0.0	0.0	97.3	0.0	0.0	99.6	0.0	0.0	99.8	0.0	100.0	100.0	0.0	0.0	0.0	0.0
	G	99.9	99.9	0.0	0.0	0.0	0.1	0.3	0.2	0.0	0.0	0.0	0.0	99.9	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	99.9	99.9
	T	0.0	0.0	0.0	0.0	100.0	17.2	15.6	14.7	99.9	100.0	2.6	100.0	0.0	0.3	0.0	0.0	0.2	0.0	0.0	0.0	0.0	100.0	0.0	0.0
AID-BE3	A	0.0	0.0	99.9	99.9	0.1	0.2	0.8	0.3	0.0	0.0	0.0	0.0	0.1	0.0	99.9	0.1	0.1	100.0	0.0	0.0	0.0	0.0	0.1	0.1
	C	0.0	0.0	0.0	0.0	0.0	73.8	80.3	86.6	0.0	0.0	96.7	0.0	0.0	98.8	0.0	0.0	97.0	0.0	99.9	100.0	0.0	0.0	0.0	0.0
	G	99.9	99.9	0.0	0.0	0.0	0.8	0.5	0.4	0.0	0.0	0.0	0.0	99.9	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	99.9	99.9	99.9
	T	0.0	0.0	0.0	0.0	99.9	25.2	18.6	12.7	100.0	100.0	3.2	100.0	0.0	1.2	0.0	0.0	2.9	0.0	0.1	0.0	0.0	100.0	0.0	0.0
APOBEC3G-BE3	A	0.0	0.0	100.0	100.0	0.0	0.0	2.9	1.7	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.1	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.1
	C	0.0	0.0	0.0	0.0	0.0	96.1	88.3	88.3	0.0	0.0	99.8	0.0	0.0	99.9	0.0	0.0	100.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0
	G	100.0	99.9	0.0	0.0	0.0	0.3	1.8	10.8	0.0	0.0	0.0	0.0	100.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	100.0	99.9	99.9
	T	0.0	0.0	0.0	0.0	100.0	3.6	27.2	19.2	100.0	100.0	0.1	100.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0

图147



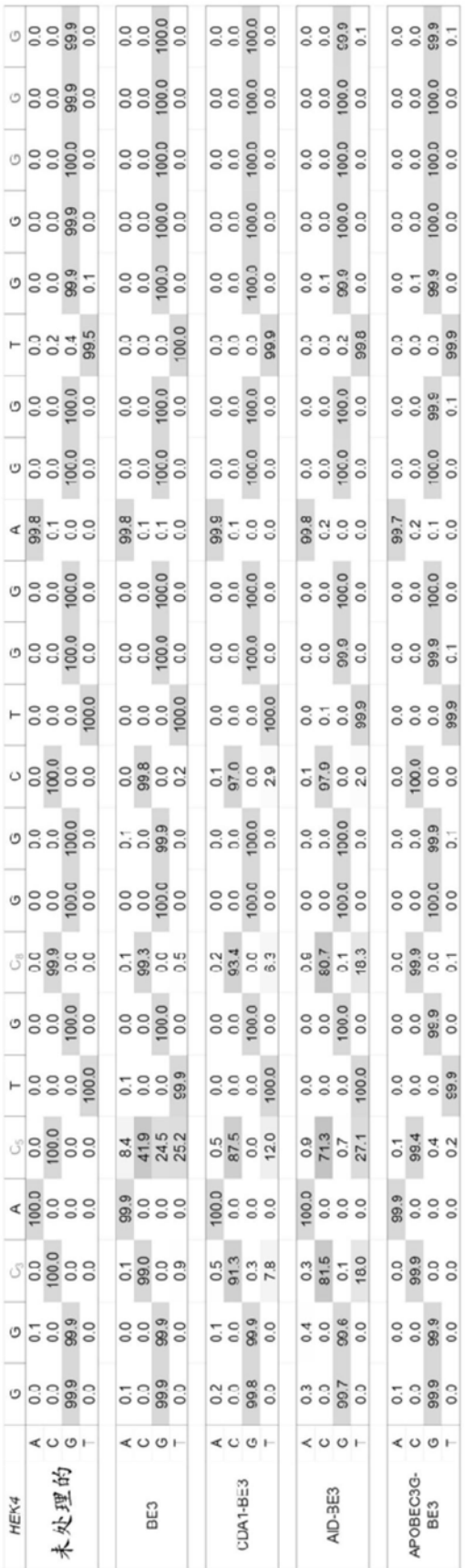
HEK2		G	A	A	C_{14}	A	C_{11}	A	A	A	G	C_{11}	A	T	A	G	A	C	T	G	C	G	G	G
未处理的	A	0.0	100.0	100.0	0.0	100.0	0.0	100.0	100.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	0.0	100.0	0.0	99.9	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	99.9	0.0	0.0	0.0	0.0
	G	100.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	100.0	0.3	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	99.9	100.0	100.0	100.0
	T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.1	0.0	0.0	0.0
BE3	A	0.1	100.0	100.0	1.3	99.9	0.7	100.0	100.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	0.0	80.5	0.0	48.5	0.0	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	99.7	0.0	0.0	100.0	0.0	0.0	0.0
	G	99.9	0.0	0.0	0.9	0.1	36.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	99.9	100.0
	T	0.0	0.0	0.0	17.2	0.0	14.8	0.0	0.0	0.0	0.0	0.1	0.0	100.0	0.0	0.0	0.0	0.2	100.0	0.0	0.0	0.0	0.0	0.0
CDA1-BE3	A	0.0	99.9	99.9	9.6	99.9	1.8	100.0	100.0	100.0	0.0	0.2	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	0.0	70.3	0.0	73.5	0.0	0.0	0.0	0.0	90.7	0.0	0.0	0.0	0.0	100.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0
	G	100.0	0.0	0.1	0.3	0.1	3.5	0.0	0.0	0.0	100.0	0.3	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	100.0
	T	0.0	0.0	0.0	19.3	0.0	21.2	0.0	0.0	0.0	0.0	8.8	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.1	0.0	0.0	0.0
AID-BE3	A	0.0	99.9	99.9	8.9	99.9	1.2	100.0	100.0	100.0	0.0	0.2	100.0	0.0	100.0	0.0	100.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	0.0	54.5	0.0	52.4	0.0	0.0	0.0	0.0	83.8	0.0	0.0	0.0	0.0	0.0	99.7	0.0	99.8	0.0	0.0	0.0	0.0
	G	99.9	0.0	0.1	0.7	0.1	6.1	0.0	0.0	0.0	99.9	0.6	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	100.0	100.0
	T	0.1	0.1	0.0	37.9	0.0	40.3	0.0	0.0	0.0	0.1	15.5	0.0	100.0	0.0	0.0	0.0	0.2	99.9	0.0	0.2	0.0	0.0	0.0
APOBEC3G-BE3	A	0.0	100.0	100.0	0.1	100.0	0.1	100.0	100.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	0.0	97.3	0.0	30.3	0.0	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	100.0	0.0	99.9	0.0	0.0	0.0	0.0
	G	100.0	0.0	0.0	0.1	0.0	5.9	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	100.0	100.0
	T	0.0	0.0	0.0	2.0	0.0	13.7	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.1	0.0	0.0	0.0

图148



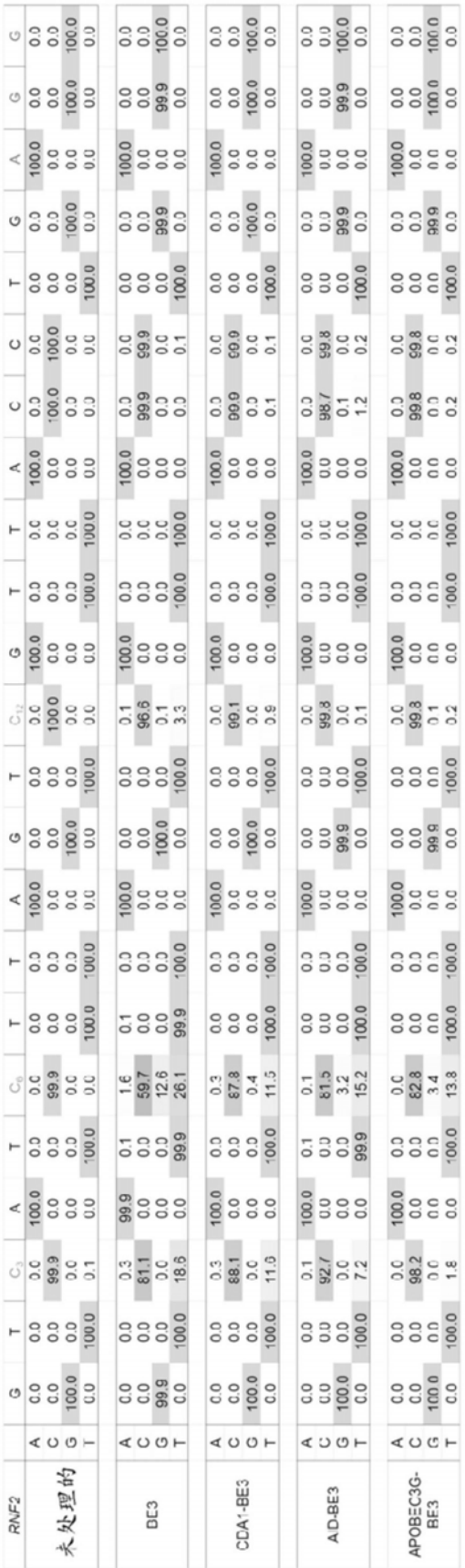
HEK3		G	G	C_3	C_4	C_5	A	G	A	C_9	T	G	A	G	C	A	C	G	T	G	A	T	G	G
未处理的	A	0.0	0.1	0.0	0.0	0.0	99.9	0.0	100.0	0.0	0.0	0.0	99.9	0.0	0.0	99.9	0.0	0.0	0.0	0.0	99.9	0.0	0.0	0.1
	C	0.0	0.0	100.0	99.9	99.9	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0
	G	100.0	99.9	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	99.9	0.0	99.9	0.0	0.0	0.0	99.9	0.0	100.0	0.0	0.0	100.0	99.9
	T	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0
BE3	A	0.0	0.0	0.0	0.8	1.1	100.0	0.0	100.0	0.4	0.0	0.0	100.0	0.2	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0
	C	0.0	0.0	98.6	67.4	58.1	0.0	0.0	0.0	98.3	0.0	0.0	0.0	0.0	100.0	99.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	100.0	100.0	0.0	0.4	8.9	0.0	100.0	0.0	0.6	0.0	100.0	0.0	99.8	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	99.9	100.0
	T	0.0	0.0	1.3	31.4	31.9	0.0	0.0	0.0	0.7	100.0	0.0	0.0	0.0	0.0	0.3	0.0	0.0	100.0	0.0	0.0	100.0	0.1	0.0
CDA1-BE3	A	0.1	0.0	0.8	0.6	0.0	100.0	0.0	99.9	0.4	0.0	0.0	99.9	0.0	0.0	100.0	0.0	0.0	0.0	0.0	99.9	0.0	0.1	0.1
	C	0.0	0.0	79.4	73.6	73.4	0.0	0.0	0.0	91.0	0.0	0.0	0.0	99.9	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	99.9	100.0	0.0	0.0	2.1	0.0	100.0	0.0	0.3	0.0	99.9	0.0	100.0	0.0	0.0	0.0	99.9	0.0	99.9	0.1	0.0	99.9	99.9
	T	0.1	0.0	19.8	25.9	24.4	0.0	0.0	0.0	8.3	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	100.0	0.0	0.0
AID-BE3	A	0.1	0.3	1.0	0.7	0.3	99.9	0.0	100.0	1.6	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0
	C	0.0	0.0	63.9	75.4	60.2	0.0	0.0	0.0	86.6	0.0	0.0	0.0	98.9	0.0	99.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	99.8	99.7	0.1	0.4	1.0	0.0	100.0	0.0	1.2	0.0	100.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	99.9
	T	0.0	0.0	35.0	23.5	38.5	0.0	0.0	0.0	10.6	100.0	0.0	0.0	0.0	1.1	0.0	0.2	0.0	100.0	0.0	0.0	100.0	0.0	0.0
A3CBEC3G-BE3	A	0.0	0.0	0.0	0.6	3.1	100.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0
	C	0.0	0.0	99.9	86.7	33.9	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	100.0	99.9	0.0	1.4	28.6	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	100.0	100.0
	T	0.0	0.0	0.0	11.3	34.4	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0

图149



HEK4		G	G	C_0	A	C_6	T	G	C_9	G	G	C	T	G	G	A	G	G	T	G	G	G	G
未处理的	A	0.0	0.1	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	100.0	0.0	100.0	0.0	0.0	99.9	0.0	0.0	100.0	0.0	0.0	0.0	0.1	0.0	0.0	0.2	0.0	0.0	0.0	0.0
	G	99.9	99.9	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	0.0	100.0	100.0	0.0	100.0	100.0	0.4	99.9	99.9	100.0	99.9	99.9
	T	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	99.5	0.1	0.0	0.0	0.0	0.0
BE3	A	0.1	0.0	0.1	99.9	8.4	0.1	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.0	99.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	99.0	0.0	41.9	0.0	0.0	99.3	0.0	0.0	99.8	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	99.9	99.9	0.0	0.0	24.5	0.0	100.0	0.0	100.0	99.9	0.0	0.0	100.0	100.0	0.1	100.0	100.0	0.0	100.0	100.0	100.0	100.0	100.0
	T	0.0	0.0	0.9	0.0	25.2	99.9	0.0	0.5	0.0	0.0	0.2	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0
CDA1-BE3	A	0.2	0.1	0.5	100.0	0.5	0.0	0.0	0.2	0.0	0.0	0.1	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	91.3	0.0	87.5	0.0	0.0	93.4	0.0	0.0	97.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	99.8	99.9	0.3	0.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	0.0	100.0	100.0	0.0	100.0	100.0	0.0	100.0	100.0	100.0	100.0	100.0
	T	0.0	0.0	7.8	0.0	12.0	100.0	0.0	6.3	0.0	0.0	2.9	100.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0
AID-BE3	A	0.3	0.4	0.3	100.0	0.9	0.0	0.0	0.6	0.0	0.0	0.1	0.0	0.0	0.0	99.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	81.5	0.0	71.3	0.0	0.0	80.7	0.0	0.0	97.0	0.1	0.0	0.0	0.2	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0
	G	99.7	99.6	0.1	0.0	0.7	0.0	100.0	0.1	100.0	100.0	0.0	0.0	99.9	100.0	0.0	100.0	100.0	0.2	99.9	100.0	100.0	100.0	99.9
	T	0.0	0.0	18.0	0.0	27.1	100.0	0.0	18.3	0.0	0.0	2.0	99.9	0.0	0.0	0.0	0.0	0.0	99.8	0.0	0.0	0.0	0.0	0.1
APOBEC3G-BE3	A	0.1	0.0	0.0	99.9	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	99.9	0.0	99.4	0.0	0.0	99.9	0.0	0.0	100.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0
	G	99.9	99.9	0.0	0.0	0.4	0.0	99.9	0.0	100.0	99.9	0.0	0.0	99.9	100.0	0.1	100.0	99.9	0.0	99.9	100.0	100.0	100.0	99.9
	T	0.0	0.0	0.0	0.0	0.2	99.9	0.0	0.1	0.0	0.1	0.0	99.9	0.1	0.0	0.0	0.0	0.1	99.9	0.0	0.0	0.0	0.0	0.1

图150



RNF2		G	T	C_{13}	A	T	C_{16}	T	T	A	G	T	C_{20}	G	T	T	A	C	C	T	G	A	G	G
未处理的	A	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0
	C	0.0	0.0	99.9	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0
	G	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0
	T	0.0	100.0	0.1	0.0	100.0	0.0	100.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0
BE3	A	0.0	0.0	0.3	99.9	0.1	1.6	0.1	0.0	100.0	0.0	0.0	0.1	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0
	C	0.0	0.0	81.1	0.0	0.0	59.7	0.0	0.0	0.0	0.0	0.0	96.6	0.0	0.0	0.0	0.0	99.9	99.9	0.0	0.0	0.0	0.0	0.0
	G	99.9	0.0	0.0	0.0	0.0	12.6	0.0	0.0	0.0	100.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	99.9	100.0
	T	0.0	100.0	18.8	0.0	99.9	26.1	99.9	100.0	0.0	0.0	100.0	3.3	0.0	100.0	100.0	0.0	0.0	0.1	100.0	0.0	0.0	0.0	0.0
CDA1-BE3	A	0.0	0.0	0.3	100.0	0.0	0.3	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0
	C	0.0	0.0	88.1	0.0	0.0	87.8	0.0	0.0	0.0	0.0	0.0	99.1	0.0	0.0	0.0	0.0	99.9	99.9	0.0	0.0	0.0	0.0	0.0
	G	100.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0
	T	0.0	100.0	11.9	0.0	100.0	11.5	100.0	100.0	0.0	0.0	100.0	0.9	0.0	100.0	100.0	0.0	0.1	0.1	100.0	0.0	0.0	0.0	0.0
AD-BE3	A	0.0	0.0	0.1	100.0	0.1	0.1	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0
	C	0.0	0.0	92.7	0.0	0.0	81.5	0.0	0.0	0.0	0.0	0.0	99.8	0.0	0.0	0.0	0.0	98.7	99.8	0.0	0.0	0.0	0.0	0.0
	G	100.0	0.0	0.0	0.0	0.0	3.2	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	99.9	0.0	99.9	100.0
	T	0.0	100.0	7.2	0.0	99.9	15.2	100.0	100.0	0.0	0.0	100.0	0.1	0.0	100.0	100.0	0.0	1.2	0.2	100.0	0.0	0.0	0.0	0.0
APOBEC3G-BE3	A	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0
	C	0.0	0.0	98.2	0.0	0.0	82.8	0.0	0.0	0.0	0.0	0.0	99.3	0.0	0.0	0.0	0.0	99.8	99.8	0.0	0.0	0.0	0.0	0.0
	G	100.0	0.0	0.0	0.0	0.0	3.4	0.0	0.0	0.0	99.9	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	100.0	100.0
	T	0.0	100.0	1.8	0.0	100.0	13.8	100.0	100.0	0.0	0.0	100.0	0.2	0.0	100.0	100.0	0.0	0.2	0.2	100.0	0.0	0.0	0.0	0.0

图151

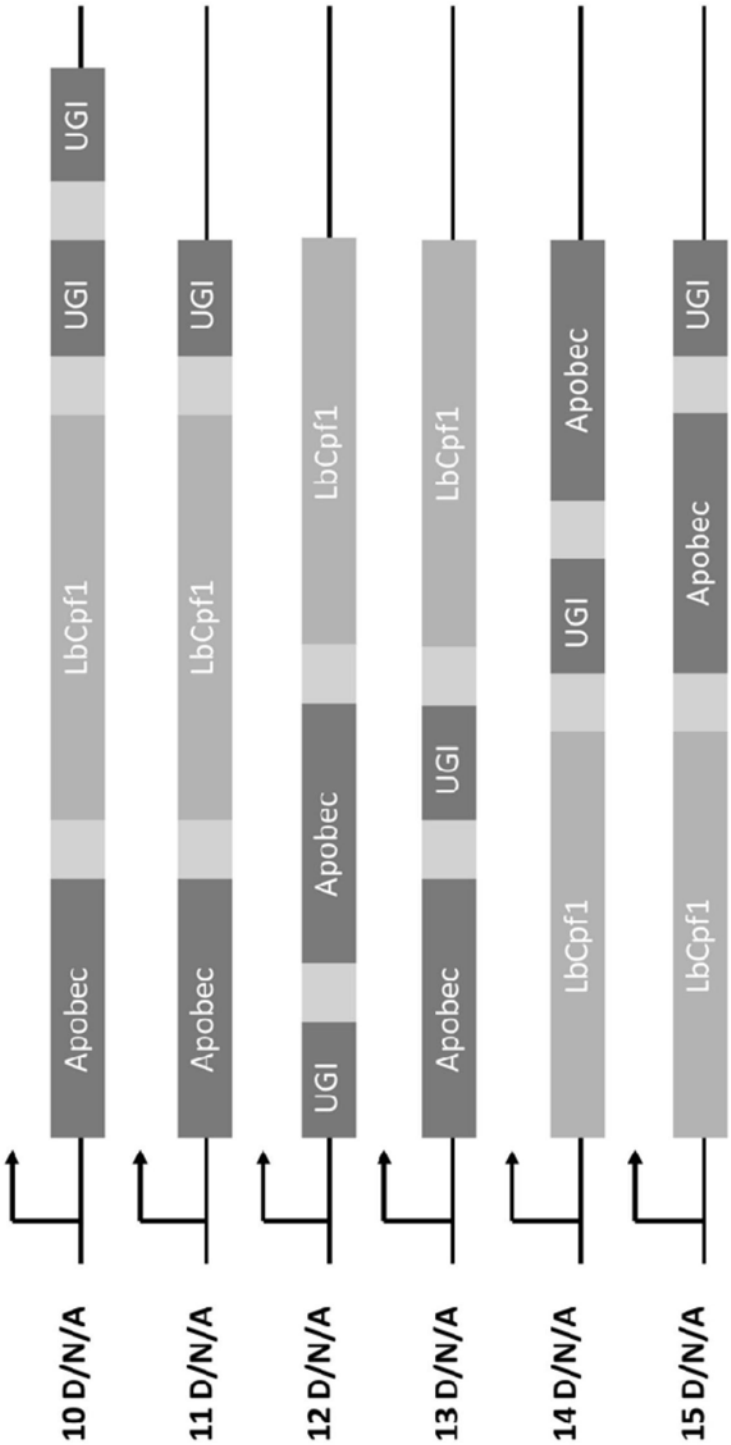

UGI
UGI
LbCpf1
Apobec
10 D/N/A
UGI
LbCpf1
Apobec
11 D/N/A
LbCpf1
Apobec
UGI
12 D/N/A
LbCpf1
UGI
Apobec
13 D/N/A
Apobec
UGI
LbCpf1
14 D/N/A
UGI
Apobec
LbCpf1
15 D/N/A

图152

198/208

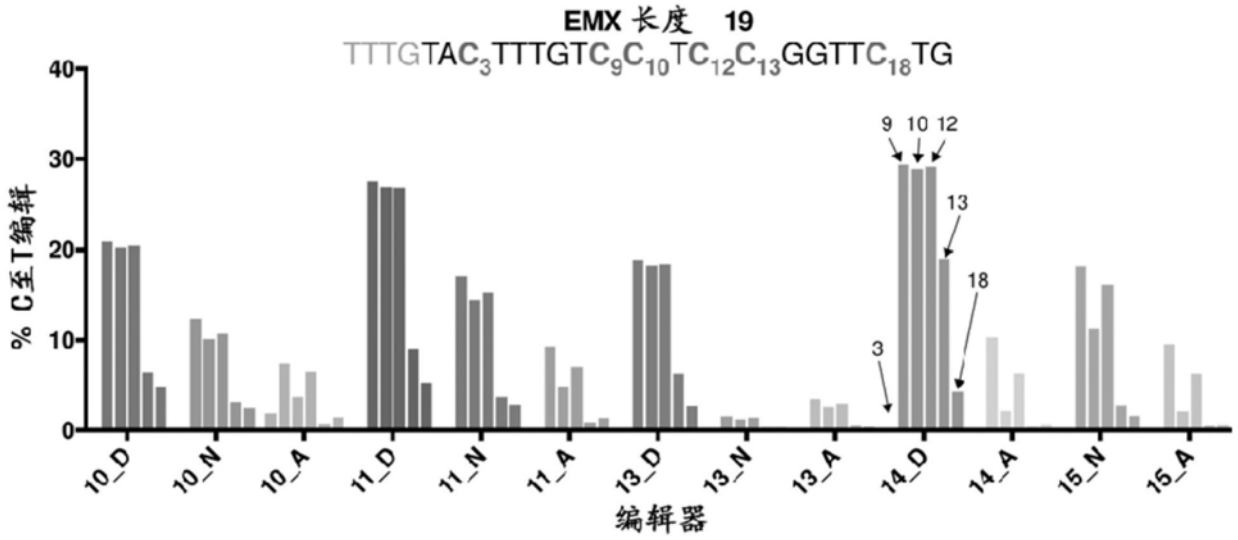


图153

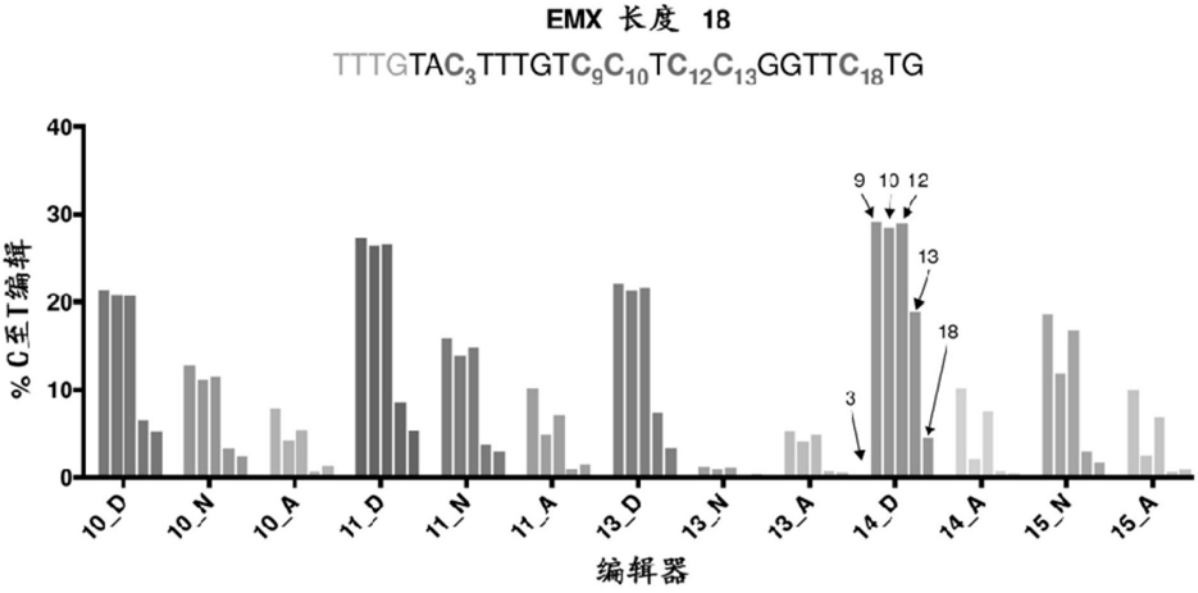


图154

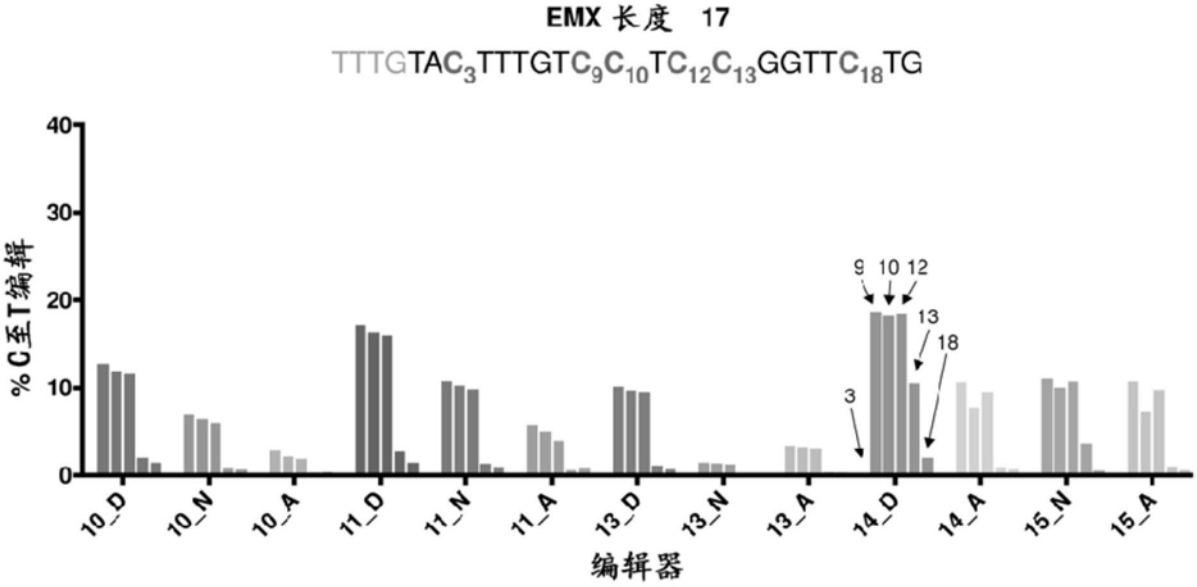

EMX 长度 17
TTTGTAC3TTTGTC9C10TC12C13GGTTC18TG
% C至T编辑
40
30
20
10
0
9 10 12
13
18
3
10_D
10_N
10_A
11_D
11_N
11_A
13_D
13_N
13_A
14_D
14_A
15_N
15_A
编辑器

图155

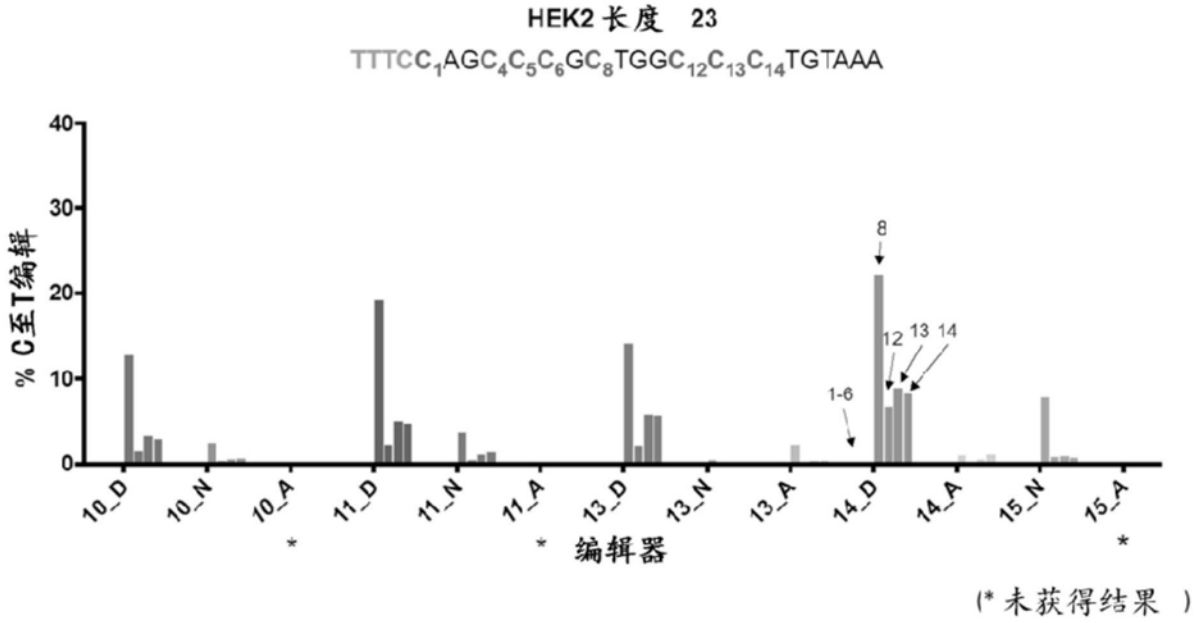

HEK2 长度 23
TTTCC1AGC4C5C6GC8TGGC12C13C14TGTAAA
% C至T编辑
40
30
20
10
0
8
12
13
14
1-6
10_D
10_N
10_A
11_D
11_N
11_A
13_D
13_N
13_A
14_D
14_A
15_N
15_A
*
*
*
编辑器
(* 未获得结果)

图156

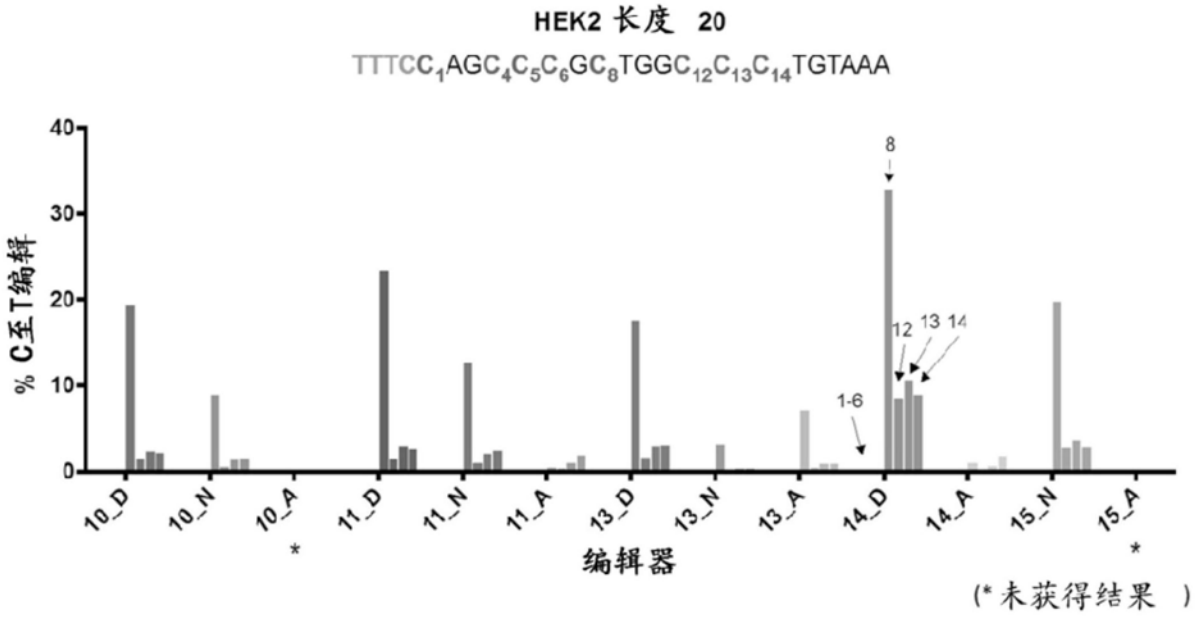

HEK2 长度 20
TTTCC1AGC4C5C6GC8TGGC12C13C14TGTAAA
% C至T编辑
编辑器
10_D
10_N
10_A
11_D
11_N
11_A
13_D
13_N
13_A
14_D
14_A
15_N
15_A
1-6
8
12
13
14
(* 未获得结果)

图157

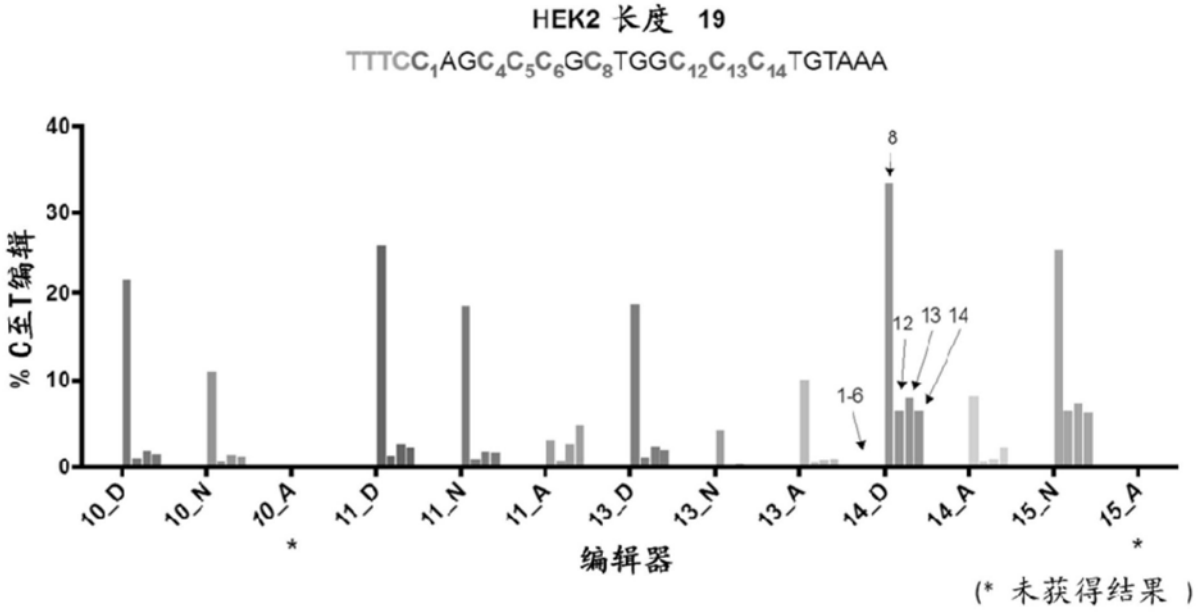

HEK2 长度 19
TTTCC1AGC4C5C6GC8TGGC12C13C14TGTAAA
% C至T编辑
编辑器
10_D
10_N
10_A
11_D
11_N
11_A
13_D
13_N
13_A
14_D
14_A
15_N
15_A
1-6
8
12
13
14
(* 未获得结果)

图158

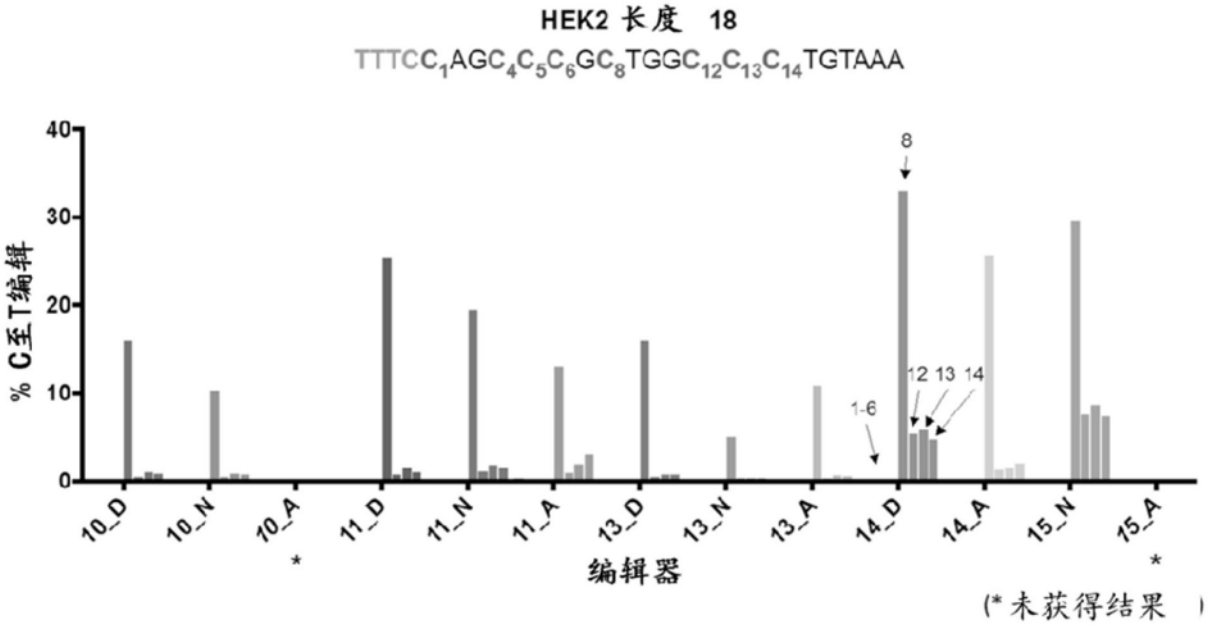

HEK2 长度 18
TTTCC1AGC4C5C6GC8TGGC12C13C14TGTAAA
% C至T编辑
40
30
20
10
0
8
12 13 14
1-6
10_D
10_N
10_A
11_D
11_N
11_A
13_D
13_N
13_A
14_D
14_A
15_N
15_A
*
*
编辑器
(*未获得结果)

图159

EMX 19

	10_D	10_N	10_A	11_D	11_N	11_A	13_D	13_N	13_A	14_D	14_A	15_N	15_A
C3	0.025	0.039	2.662	0.012	0.013	0.041	0.02	0.013	0.012	0.033	0.012	0.02	0.021
C9	20.899	13.15	10.456	27.535	18.921	13.072	19.043	1.591	3.638	29.39	14.271	20.95	13.924
C10	20.226	10.784	5.206	26.909	16.026	6.748	18.445	1.266	2.658	28.933	2.934	13.039	3.051
C12	20.453	11.385	9.095	26.849	16.893	9.983	18.589	1.437	3.127	29.165	8.695	18.597	9.07
C13	6.39	3.396	1.085	8.946	4.126	1.3	6.317	0.257	0.659	18.973	0.547	3.247	0.8
C18	4.757	2.603	2.013	5.304	3.233	1.914	2.805	0.235	0.367	4.302	0.992	1.848	0.994
插入/缺失(%)	0.028979	6.0862	28.9183	0.038623	9.8557	28.6068	0.94296	1.4573	3.78	0.024623	27.6872	13.0741	31.3295
C3	0.02499276	0.03662638	1.89219485	0.01199537	0.01171876	0.02927121	0.01981141	0.01281055	0.0115464	0.03299187	0.00867754	0.01738518	0.01442081
C9	20.8929437	12.3496647	7.43230255	27.5243652	17.056203	9.3325191	18.8634321	1.56781436	3.5004836	29.3827633	10.3197597	18.2109761	9.56168042
C10	20.2201387	10.1276642	3.7005133	26.8986069	14.4465255	4.81761314	18.271071	1.24755058	2.5575276	28.9258758	2.12165755	11.3342681	2.09513696
C12	20.4470729	10.6920861	6.46488062	26.8386301	15.2280766	7.12718316	18.4137132	1.4160586	3.0087994	29.1578187	6.28759796	15.1656096	6.22841435
C13	6.38814824	3.18931265	0.77123645	8.94254479	3.71935382	0.9281116	6.25743322	0.25325474	0.6340898	18.9683283	0.39555102	2.82248397	0.549364
C18	4.75562147	2.44457621	1.43087462	5.30195144	2.91436522	1.36646585	2.77854997	0.23157535	0.3531274	4.30094072	0.71734298	1.60639063	0.68258477

图160

EMX 18

	10_D	10_N	10_A	11_D	11_N	11_A	13_D	13_N	13_A	14_D	14_A	15_N	15_A
C3	0.023	0.008	0.032	0.044	0.035	0.016	0.03	0.016	0.016	0.018	0.02	0.017	0.01
C9	21.333	13.606	10.333	27.294	17.216	14.646	22.315	1.297	5.485	29.094	14.566	21.847	13.66
C10	20.865	11.931	5.528	26.409	15.041	6.931	21.548	0.932	4.309	28.456	2.996	13.762	3.337
C12	20.773	12.165	7.147	26.572	16.178	10.157	21.839	1.202	5.059	28.937	10.88	19.625	9.359
C13	6.538	3.591	0.848	8.562	4.096	1.387	7.426	0.129	0.729	18.993	0.945	3.353	0.809
C18	5.174	2.54	1.788	5.313	3.146	2.196	3.443	0.243	0.565	4.517	0.647	2.03	1.154
插入/缺失(%)	0.039639	5.7886	23.1911	0.035862	7.9573	30.2671	1.1014	1.1121	4.6798	0.040521	29.8821	14.2217	26.5552
C3	0.02299088	0.00753691	0.02457885	0.04398422	0.03221495	0.01115726	0.02966958	0.01582206	0.01525123	0.01799271	0.01402358	0.01458231	0.00734448
C9	21.3245438	12.8184031	7.93656364	27.2842118	15.8460712	10.2130805	22.0692226	1.28257606	5.22831297	29.0822108	10.2133733	18.7399852	10.0325597
C10	20.8567293	11.1461507	4.24595599	26.3995292	13.8441425	4.8331873	21.3106703	0.92163523	4.10734742	28.4444693	2.10073228	11.8048096	2.45085298
C12	20.7647658	11.4608168	5.48953208	26.5624707	14.890668	7.08277065	21.5984653	1.18863256	4.82224892	28.9252744	7.62882752	16.8339914	6.87369883
C13	6.5354084	3.38313137	0.65133947	8.5589295	3.77006899	0.96719532	7.34421004	0.12756539	0.69488426	18.9853038	0.66261416	2.8761464	0.59416843
C18	5.17194908	2.39296956	1.37334313	5.31109465	2.89566334	1.53133448	3.4050788	0.2402976	0.53855913	4.51516967	0.45366281	1.74129949	0.84755299

图161

EMX 17

	10_D	10_N	10_A	11_D	11_N	11_A	13_D	13_N	13_A	14_D	14_A	15_N	15_A
C3	0.016	0.014	0.044	0.012	0.018	0.012	0.013	0.016	0.03	0.014	0.011	0.016	0.005
C9	12.742	6.895	2.862	17.142	10.77	5.941	10.177	1.471	3.395	18.719	11.13	11.103	11.493
C10	11.901	6.427	2.199	16.385	10.277	5.064	9.741	1.393	3.216	18.206	8.1	10.086	7.747
C12	11.548	6.014	1.916	15.933	9.872	4.041	9.468	1.287	2.986	18.492	9.847	10.751	10.523
C13	1.993	0.768	0.252	2.696	1.357	0.599	1.155	0.083	0.285	10.532	0.971	3.674	1.082
C18	1.457	0.613	0.361	1.471	0.987	0.773	0.667	0.043	0.245	1.978	0.729	0.545	0.621
插入/缺失(%)	0.017245	0.052423	1.7787	0.025203	0.11499	2.7876	0.14938	0.019738	0.041079	0.04713	4.2923	0.37931	6.7995
C3	0.01599724	0.01399266	0.04321737	0.01199698	0.0179793	0.01166549	0.01298058	0.01599684	0.02998768	0.0139934	0.01052785	0.01593931	0.00466003
C9	12.7398026	6.89138543	2.81109361	17.1376797	10.7576156	5.77538868	10.1617976	1.47070965	3.39360537	18.7101777	10.652267	11.0608852	10.7115335
C10	11.8989477	6.42363077	2.15988639	16.3808705	10.2651825	4.92283594	9.72644889	1.39272505	3.2146789	18.1974195	7.7523237	10.0477428	7.22024274
C12	11.5460085	6.01084728	1.88192011	15.9289844	9.86064819	3.92835308	9.4538567	1.28674597	2.98477338	18.4832847	9.42433722	10.7102204	9.80748862
C13	1.99265631	0.76759739	0.24751768	2.69532053	1.35543959	0.58230228	1.15327466	0.08298362	0.28188292	10.5270363	0.92932177	3.66006415	1.00842941
C18	1.45674874	0.6126786 5	0.35457889	1.47062926	0.98586505	0.75145185	0.66600364	0.04299151	0.24489936	1.97706777	0.69770913	0.54293276	0.57877511

图162

Hek2 23

	10_D	10_N	10_A	11_D	11_N	11_A	13_D	13_N	13_A	14_D	14_A	15_N	15_A
C1	0.042	0.021	0	0.039	0.084	0	0.045	0.07	0.04	0.045	0.056	0.057	0
C4	0.016	0.018	0	0.023	0.013	0	0.027	0.022	0.016	0.035	0.008	0.016	0
C5	0.031	0.018	0	0.037	0.052	0	0.048	0.021	0.044	0.047	0.157	0.017	0
C6	0.064	0.045	0	0.064	0.02	0	0.077	0.068	0.051	0.078	0.032	0.08	0
C8	12.758	2.444	0	19.198	3.66	0	14.134	0.423	2.22	22.158	0.977	7.763	0
C12	1.418	0.291	0	2.281	0.446	0	2.163	0.101	0.21	6.745	0.235	0.809	0
C13	3.282	0.603	0	5.021	1.121	0	5.713	0.042	0.308	8.933	0.599	0.921	0
C14	2.898	0.641	0	4.719	1.38	0	5.614	0.096	0.38	8.268	1.066	0.746	0

插入/缺失(%)	0.052343	11.0664		0.041898			2.2549	6.3138	13.9681	0.085777	59.2508	11.3052	
C1	0.04197802	0.01867606	0	0.03898366	0.084	0	0.0439853	0.06558034	0.03441276	0.0449614	0.02281955	0.05055604	0
C4	0.01599163	0.01600805	0	0.02299036	0.013	0	0.02639118	0.02061096	0.0137651	0.03496998	0.00325994	0.01419117	0
C5	0.03098377	0.01600805	0	0.0369845	0.052	0	0.04691765	0.0196741	0.03785404	0.04695968	0.06397624	0.01507812	0
C6	0.0639665	0.04002012	0	0.06397319	0.02	0	0.07526373	0.06370662	0.04387627	0.07793309	0.01303974	0.07095584	0
C8	12.7513221	2.17353718	0	19.1899564	3.66	0	13.8152924	0.39529263	1.90990818	22.1389935	0.39811968	6.88537732	0
C12	1.41725778	0.25879678	0	2.28004131	0.446	0	2.11422651	0.09462306	0.18066699	6.73921431	0.09576062	0.71754093	0
C13	3.2802821	0.53626951	0	5.0188963	1.121	0	5.58417756	0.0393482	0.26497825	8.92533754	0.24408771	0.81687911	0
C14	2.8964831	0.57006438	0	4.71702283	1.38	0	5.48740991	0.08993875	0.32692122	8.26090795	0.43438647	0.66166321	0

图163

Hek2 20

	10_D	10_N	10_A	11_D	11_N	11_A	13_D	13_N	13_A	14_D	14_A	15_N	15_A
C1	0.024	0.051	0	0.036	0.116	0.126	0.035	0.022	0.042	0.032	0.02	0.133	0
C4	0.054	0.023	0	0.026	0.011	0.01	0.019	0.019	0.011	0.025	0.016	0.016	0
C5	0.033	0.017	0	0.051	0.054	0.048	0.046	0.02	0.029	0.035	0.075	0.043	0
C6	0.096	0.083	0	0.058	0.032	0.022	0.055	0.016	0.047	0.139	0.166	0.083	0
C8	19.339	9.265	0	23.342	13.199	1.173	17.974	3.225	7.353	32.768	2.835	20.891	0
C12	1.402	0.551	0	1.491	1.036	0.713	1.68	0.168	0.471	8.549	0.3	2.967	0
C13	2.377	1.409	0	2.974	2.174	2.814	3.043	0.356	0.895	10.654	1.81	3.81	0
C14	2.129	1.545	0	2.604	2.523	5.582	3.076	0.316	0.955	8.964	4.97	3.086	0
插入/缺失(%)	0.10237	3.2815	24.3902	0.10375	4.3674	66.4779	1.8097	1.6132	3.0719	0.093678	63.4082	5.4543	
C1	0.02397543	0.04932644	0	0.03596265	0.11093382	0.04223785	0.03436661	0.0216451	0.0407098	0.03197002	0.00731836	0.12574578	0
C4	0.05394472	0.02224526	0	0.02597303	0.01051959	0.00335221	0.01865616	0.01869349	0.01066209	0.02497658	0.00585469	0.01512731	0
C5	0.03296622	0.01644215	0	0.05094709	0.0516416	0.01609061	0.04516754	0.01967736	0.02810915	0.03496721	0.02744385	0.04065465	0
C6	0.09590172	0.08027636	0	0.05793983	0.03060243	0.00737486	0.05400467	0.01574189	0.04555621	0.13886979	0.06074239	0.07847293	0
C8	19.3192027	8.96096503	0	23.3177827	12.6225469	0.39321423	17.6487245	3.1729743	7.12712319	32.7373036	1.03737753	19.7515422	0
C12	1.40056477	0.53291894	0	1.48945309	0.99075374	0.23901257	1.64959704	0.18496718	0.45653135	8.54099147	0.1097754	2.80517092	0
C13	2.37456667	1.36275367	0	2.97091448	2.07905272	0.94331189	2.98793083	0.35025701	0.8675065	10.6440195	0.66231158	3.60219117	0
C14	2.12682054	1.49430083	0	2.60129835	2.4128105	1.87120352	3.02033363	0.31090229	0.92566336	8.9556027	1.81861246	2.9176803	0

图164

Hek2 19

	10_D	10_N	10_A	11_D	11_N	11_A	13_D	13_N	13_A	14_D	14_A	15_N	15_A
C1	0.072	0.031	0	0.039	0.062	0.406	0.023	0.019	0.025	0.06	0.224	0.04	0
C4	0.034	0.019	0	0.021	0.06	0.018	0.039	0.033	0.025	0.012	0.009	0.008	0
C5	0.03	0.037	0	0.044	0.045	0.027	0.018	0.031	0.033	0.024	0.021	0.012	0
C6	0.053	0.066	0	0.042	0.071	0.164	0.033	0.031	0.046	0.098	0.117	0.034	0
C8	21.611	11.001	0	25.556	18.67	6.38	18.944	4.272	10.143	33.441	15.227	25.017	0
C12	1.049	0.704	0	1.345	1.022	1.548	1.154	0.195	0.489	6.54	1.177	6.539	0
C13	1.914	1.39	0	2.629	1.819	5.23	2.301	0.253	0.881	8.065	1.823	7.452	0
C14	1.545	1.212	0	2.238	1.761	9.623	1.98	0.151	0.975	6.517	4.174	6.365	0

插入/缺失(%)	0.055704	0.17936		0.04541	0.43984	50.2032	0.61981	0.18691	0.22381	0.14694	45.816	0.52493	

C1	0.0719517	0.0309444	0	0.03898229	0.0617273	0.20217501	0.02285744	0.01896449	0.02494405	0.05991184	0.12137216	0.03979003	0
C4	0.03398072	0.01896592	0	0.02099046	0.0597361	0.00896342	0.03875827	0.03293832	0.02494405	0.01198237	0.00487656	0.00795801	0
C5	0.02998299	0.03693364	0	0.04398002	0.04480207	0.01344514	0.01788843	0.03094206	0.03292614	0.02396473	0.01137864	0.01193701	0
C6	0.05296995	0.06588162	0	0.04198093	0.07068771	0.08156675	0.03279546	0.03094206	0.04589705	0.097856	0.06339528	0.03382152	0
C8	21.5987457	10.9812686	0	25.544395	18.5878819	3.17703584	18.8265832	4.2640152	10.120299	33.3918618	8.25059768	24.8856783	0
C12	1.04840518	0.70273731	0	1.34438924	1.01750484	0.77085446	1.14684739	0.15463553	0.48790557	6.53039012	0.63774568	6.50467483	0
C13	1.91291469	1.3875069	0	2.62780617	1.81099931	2.60437264	2.28673817	0.25252712	0.87902823	8.05314929	0.98777432	7.41288222	0
C14	1.54412392	1.20982616	0	2.23698372	1.75325442	4.79194606	1.96772776	0.15071777	0.97281785	6.50742392	2.26164016	6.33158821	0

图165

Hek2 18

	10_D	10_N	10_A	11_D	11_N	11_A	13_D	13_N	13_A	14_D	14_A	15_N	15_A
C1	0.037	0.027	0	0.035	0.041	0.315	0.039	0.034	0.034	0.028	0.125	0.074	0
C4	0.014	0.016	0	0.007	0.018	0.017	0.018	0.038	0.019	0.014	0.021	0.035	0
C5	0.052	0.018	0	0.051	0.034	0.045	0.033	0.039	0.022	0.014	0.088	0.035	0
C6	0.056	0.035	0	0.032	0.08	0.047	0.082	0.051	0.07	0.081	0.208	0.107	0
C8	16.035	10.427	0	25.298	19.509	14.808	15.991	5.043	10.926	33.064	28.458	29.702	0
C12	0.571	0.503	0	0.802	1.161	1.138	0.513	0.228	0.317	5.485	1.471	7.611	0
C13	1.12	0.903	0	1.607	1.869	2.299	0.803	0.315	0.709	5.902	1.741	8.752	0
C14	0.877	0.819	0	1.132	1.609	3.468	0.751	0.375	0.636	4.806	2.316	7.414	0
插入/缺失(%)	0.076858	0.17892		0.048785	0.13947	11.5608	0.14868	0.1177	0.076621	0.23744	9.8678	0.50485	
C1	0.03697156	0.02695169	0	0.03498293	0.04094282	0.27858348	0.03894201	0.03395998	0.03397395	0.02793352	0.11266525	0.07362641	0
C4	0.01398924	0.01597137	0	0.00699659	0.0179749	0.01503466	0.01797324	0.03795527	0.01898544	0.01396676	0.01892776	0.0348233	0
C5	0.05196003	0.01796779	0	0.05097512	0.03395258	0.03979764	0.03295094	0.0389541	0.02198314	0.01396676	0.07931634	0.0348233	0
C6	0.05595696	0.03493738	0	0.03198439	0.07988842	0.04156642	0.08187808	0.05093997	0.06994637	0.08080767	0.18747498	0.10645981	0
C8	16.0226758	10.408344	0	25.2856584	19.4817908	13.0960767	15.9672246	5.03706439	10.9176284	32.9854928	25.6498215	29.5520495	0
C12	0.57056114	0.50210003	0	0.80160874	1.15938075	1.0064381	0.51223727	0.22773164	0.31675711	5.47197642	1.32584466	7.57257587	0
C13	1.11913919	0.90138435	0	1.60621603	1.86639331	2.03321721	0.8018061	0.31462925	0.70845676	5.88798629	1.5692016	8.70781553	0
C14	0.87632596	0.81753465	0	1.13144775	1.60675593	3.06707146	0.74988341	0.37455863	0.63551269	4.79458863	2.08746175	7.37657042	0

图166