



(12) 发明专利申请

(10) 申请公布号 CN 111742051 A

(43) 申请公布日 2020.10.02

(21) 申请号 201980014548.2

(22) 申请日 2019.01.23

(30) 优先权数据

10-2018-0008492 2018.01.23 KR

(85) PCT国际申请进入国家阶段日

2020.08.20

(86) PCT国际申请的申请数据

PCT/KR2019/000962 2019.01.23

(87) PCT国际申请的公布数据

WO2019/147014 KO 2019.08.01

(71) 申请人 基础科学研究院

地址 韩国大田市

(72) 发明人 金晋秀 林佳映 姜汎昌 柳锡玟

(74) 专利代理机构 北京坤瑞律师事务所 11494

代理人 陈桢

(51) Int.Cl.

C12N 15/113 (2006.01)

C12N 15/10 (2006.01)

C12N 15/90 (2006.01)

C12N 9/22 (2006.01)

权利要求书2页 说明书15页

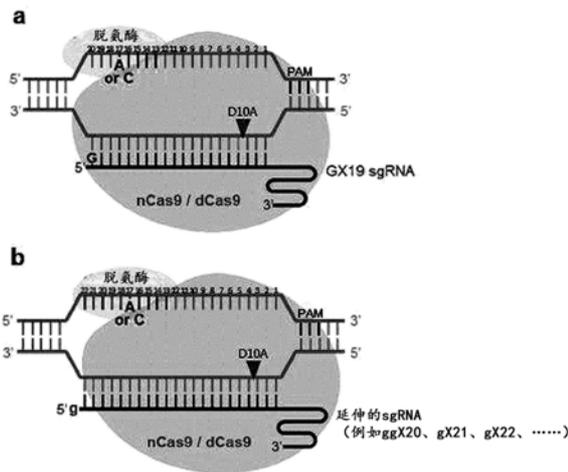
序列表1页 附图15页

(54) 发明名称

延伸的单向导RNA及其用途

(57) 摘要

本发明涉及:延伸的向导RNA和包含所述延伸的向导RNA的用于碱基编辑的组合物;以及用于碱基编辑的方法和用于产生基因修饰动物或植物的方法,所述两种方法均使用所述用于碱基编辑的组合物。



1. 一种可与靶序列杂交的用于碱基编辑的延伸的向导RNA,所述延伸的向导RNA在其5'末端进一步包含1至3个鸟嘌呤(G)和1至10个核苷酸(其中所述核苷酸各自独立地选自A、T、C和G)。
2. 根据权利要求1所述的延伸的向导RNA,其中在所述5'末端的所述鸟嘌呤与所述靶序列互补或不互补。
3. 根据权利要求1所述的延伸的向导RNA,其中所述延伸的向导RNA包含作为可与所述靶序列杂交的位点的CRISPR RNA (crRNA)、以及与RNA导向的工程化核酸酶相互作用的反式激活crRNA (tracrRNA)。
4. 一种用于碱基编辑的组合物,其包含:
 - (i) 脱氨酶或编码所述脱氨酶的基因;
 - (ii) RNA导向的工程化核酸酶或编码所述RNA导向的工程化核酸酶的基因;和
 - (iii) 可与靶序列杂交的延伸的向导RNA或编码所述延伸的向导RNA的基因,其中所述延伸的向导RNA在其5'末端进一步包含1至3个鸟嘌呤(G)和1至10个核苷酸(其中所述核苷酸各自独立地选自A、T、C和G)。
5. 根据权利要求4所述的组合物,其中在所述延伸的向导RNA的5'末端进一步包含的所述鸟嘌呤与所述靶序列互补或不互补。
6. 根据权利要求4所述的组合物,其中所述延伸的向导RNA包含作为可与所述靶序列杂交的位点的CRISPR RNA (crRNA)、以及与所述RNA导向的工程化核酸酶相互作用的反式激活crRNA (tracrRNA)。
7. 根据权利要求4所述的组合物,其中所述向导RNA是双向向导RNA或单向导RNA (sgRNA)。
8. 根据权利要求4所述的组合物,其中所述脱氨酶选自载脂蛋白B编辑复合物1 (APOBEC1)、激活诱导的脱氨酶 (AID) 和tRNA特异性腺苷脱氨酶 (tadA)。
9. 根据权利要求4所述的组合物,其进一步包含尿嘧啶DNA糖基化酶抑制剂 (UGI) 或编码所述尿嘧啶DNA糖基化酶抑制剂的基因。
10. 根据权利要求4所述的组合物,其进一步包含核定位序列 (NLS) 或编码所述核定位序列的基因。
11. 根据权利要求4所述的组合物,其中所述RNA导向的工程化核酸酶是被修饰以切割所述靶基因的一条链的经修饰的Cas9 (CRISPR相关蛋白9) 系统或经修饰的Cpf1 (源自普雷沃菌 (Prevotella) 和弗朗西斯菌 (Francisella) 的CRISPR 1) 系统。
12. 根据权利要求4所述的组合物,其中所述RNA导向的工程化核酸酶是Cas9切口酶 (nCas9) 或催化缺陷型Cas9 (dCas9)。
13. 一种用于碱基编辑的方法,所述方法包括将根据权利要求4所述的用于碱基编辑的组合物引入细胞中。
14. 根据权利要求13所述的方法,其中所述细胞是真核细胞。
15. 根据权利要求14所述的方法,其中所述真核细胞是动物细胞或真核植物细胞。
16. 根据权利要求15所述的方法,其中所述真核细胞是哺乳动物胚胎细胞或真核植物细胞。
17. 一种用于产生除人以外的哺乳动物或真核植物的突变成体的方法,所述方法包括:
 - (a) 将根据权利要求4所述的用于碱基编辑的组合物引入哺乳动物胚胎或真核植物胚

胎中;并且

(b) 使所述胚胎生长以获得成体。

延伸的单向导RNA及其用途

【技术领域】

[0001] 本公开文本涉及延伸的向导RNA、含有所述延伸的向导RNA的用于碱基编辑的组合物、使用所述用于碱基编辑的组合物进行碱基编辑的方法、以及使用所述用于碱基编辑的组合物产生基因修饰动物或植物的方法。

【背景技术】

[0002] 基因编辑技术始于免疫系统,其中通过用噬菌体感染以DNA的形式来记忆噬菌体的片段,然后通过用Cas9 (CRISPR相关蛋白9:RNA导向的DNA内切核酸酶) 切割而去除相应DNA,Cas9是一种在发生继发感染时充当基因剪刀的核酸酶。这已经发展成为一种基因编辑技术,其中当向导RNA (gRNA) 识别特定核苷酸时,Cas9蛋白可以切割相应位点以进行编辑 (Ran F.A.等人,Nat.Protoc.,8:2281-2308,2013;Woo J.W.等人,Nat.Biotechnol.,33:1162-1164,2015)。

[0003] 通过修饰传统的CRISPR-Cas9基因剪刀而形成的用于碱基编辑的基因编辑器是最近引起关注的技术,因为其可以改变特定的碱基而不会切割DNA的两条链。用于碱基编辑的基因编辑器通过具有与靶DNA互补的序列的sgRNA与靶位点结合,然后使用能够对暴露于相对侧的单链DNA起作用的脱氨酶将胞嘧啶 (C) 变为尿嘧啶 (U) 或将腺嘌呤 (A) 变为次黄嘌呤 (I)。在DNA修复和复制过程中,所得的碱基被变为胸腺嘧啶 (T) 和鸟嘌呤 (G),并且因此,可以将特定的DNA碱基分别从胞嘧啶 (C) 编辑为胸腺嘧啶 (T),从腺嘌呤 (A) 编辑为鸟嘌呤 (G)。此时,已知碱基编辑器所操作的碱基编辑范围位于从前间区序列邻近基序 (PAM) 起在前间区序列方向的位置13至17处,并且所述范围之外的效率非常低 (图1A)。

[0004] 因此,为了克服上述问题,本发明人发现,当修改限定用于碱基编辑的基因编辑器的靶标位置的sgRNA的形式和长度时,用于碱基编辑的基因编辑器的操作范围可以进一步扩展。基于此发现,完成了本公开文本。

【发明内容】

[0005] 【技术问题】

[0006] 本公开文本的一个目的是提供一种延伸的向导RNA,其可以进一步扩展用于碱基编辑的基因编辑器的操作范围。

[0007] 本公开文本的另一个目的是提供一种用于碱基编辑的组合物,其包括脱氨酶、靶标特异性核酸酶和延伸的向导RNA,所述用于碱基编辑的组合物能够进一步扩展用于碱基编辑的基因编辑器的操作范围。

[0008] 本公开文本的另一个目的是提供一种使用所述用于碱基编辑的组合物碱基编辑方法。

[0009] 本公开文本的另一个目的是提供一种使用所述用于碱基编辑的组合物来产生基因修饰动物或植物的方法。

[0010] 【技术方案】

[0011] 根据本公开文本的一个方面,上述及其他目的可以通过提供可与靶序列杂交的用于碱基编辑的延伸的向导RNA来实现,所述延伸的向导RNA在其5'末端处进一步包括1至3个鸟嘌呤(G)和1至10个核苷酸。

[0012] 根据本公开文本的另一方面,提供了一种用于碱基编辑的组合物,其包括(i)脱氨酶或编码所述脱氨酶的基因;(ii)RNA导向的工程化核酸酶或编码所述RNA导向的工程化核酸酶的基因;和(iii)可与靶序列杂交的延伸的向导RNA或编码所述延伸的向导RNA的基因,

[0013] 其中所述延伸的向导RNA在5'末端进一步包括1至3个鸟嘌呤(G)和1至10个核苷酸。

[0014] 根据本公开文本的另一方面,提供了一种用于碱基编辑的方法,其包括将所述用于碱基编辑的组合物引入细胞中。

[0015] 根据本公开文本的另一方面,提供了一种用于产生除人以外的哺乳动物或真核植物的突变成体的方法,所述方法包括(a)将所述用于碱基编辑的组合物引入哺乳动物胚胎或真核植物胚胎中并且(b)使所述胚胎生长以获得成体。

【附图说明】

[0016] 图1是展示了碱基编辑器根据sgRNA长度进行的操作的示意图。当使用常规方法GX19 sgRNA(a)时和当使用延伸的sgRNA(b)时,在与靶位置结合后所暴露的单链DNA中可能发生脱氨作用。当使用延伸的sgRNA时,从PAM起在5'方向上暴露的单链DNA延伸,从而引起更广范围的脱氨作用。

[0017] 图2示出了基于在HEK293T细胞系中通过深度测序所测量的每种活性取决于sgRNA长度的碱基编辑窗口的变化,更具体地,图2A是示出取决于在HEK2位点的sgRNA长度在不同碱基位置处的ABE 7.10置换活性的图。图2B是示出与使用GX19 sgRNA的情况相比的相对置换活性[gX20~30活性/GX19活性]的图。图2C示出了最常观察到的突变等位基因,其中在WT序列中引入了突变的部分以红色表示。图2D示出了取决于在HBB位点的sgRNA长度在不同碱基位置处的BE3置换活性。图2E示出了与使用GX19 sgRNA的情况相比的相对置换活性[gX20~30活性/GX19活性]。图2F示出了最常观察到的突变等位基因,其中在WT序列中引入了突变的部分以红色表示,并且已知在使用GX19 sgRNA时有效操作的碱基编辑窗口以浅蓝色表示。

[0018] 图3示出了基于在HEK293T细胞系中通过深度测序所测量的每种活性当使用进一步包括1或2个另外的错配G的sgRNA时碱基编辑窗口的变化,更具体地,图3A和图3C是示出取决于在FANCF位点(a)和HBB位点(c)的sgRNA长度在不同碱基位置处的BE3置换活性的图,并且图3B和3D是示出与当使用GX19 sgRNA时相比在FANCF位点(b)和HBB位点(d)中的相对置换活性[gX20~30活性/GX19活性]的图。

[0019] 图4示出了基于在HEK293T细胞系中通过深度测序所测量的每种活性取决于在四个不同位点的sgRNA长度的碱基编辑窗口的变化,更具体地,图4A是示出与当使用GX19 sgRNA时相比在四个位点的ABE 7.10相对置换活性[gX20~30活性/GX19活性]的图,图4B是示出与当使用GX19 sgRNA时相比的BE3相对置换活性[gX20~30活性/GX19活性]的图,其中已知在使用GX19sgRNA时有效操作的碱基编辑窗口以浅蓝色表示。

[0020] 图5示出了基于通过深度测序所测量的活性在油菜和大豆中取决于sgRNA类型的

碱基编辑窗口的变化,更具体地,图5A示出了当将gX19 sgRNA和gX20 sgRNA与AID2胞嘧啶碱基编辑器一起在油菜原生质体中使用时,取决于胞嘧啶位置的置换效率,图5B示出了根据sgRNA类型引入了最频繁发生的突变的等位基因的变化,其中发现仅当使用gX20 sgRNA时才产生TAG终止密码子,图5C示出了当将gX19 sgRNA和gX20 sgRNA与AID2胞嘧啶碱基编辑器一起在大豆原生质体中使用时,取决于胞嘧啶位置的置换效率,并且图5D示出了根据sgRNA类型引入了最频繁发生的突变的等位基因的变化,其中发现仅当使用gX20 sgRNA时才产生TAG终止密码子。

[0021] 图6示出了在小鼠中取决于sgRNA类型的碱基编辑窗口的变化,更具体地,图6A示出了在将ABE 7.10mRNA与不同类型的sgRNA一起显微注射到小鼠胚胎中之后,在胚胎期通过深度测序所分析的置换活性,并且图6B示出了分析通过使用ABE 7.10mRNA与GX21 sgRNA组合对胚胎进行显微注射而获得的幼崽的结果,其中获得了三个具有所需H420R突变的幼崽。

【具体实施方式】

[0022] 除非另外定义,本文所用的全部技术和科学术语具有与本公开文本所属领域的技术人员所理解的含义。总体上,本文所用的命名法是本领域熟知的并且是通常使用的。

[0023] 本公开文本提出了一种技术,所述技术通过修改限定用于碱基编辑的基因编辑器的靶位置的sgRNA的形式和长度(图1B)来进一步扩大用于碱基编辑的基因编辑器的操作范围。

[0024] 如图1所示,当使用常规方法GX19 sgRNA(a)时和当使用延伸的sgRNA(b)时,在与靶位置结合之后所暴露的单链DNA中可能发生脱氨作用。当使用延伸的sgRNA时,从PAM起在5'方向上暴露的单链DNA延伸,从而引起更广范围的脱氨作用。

[0025] 常规使用的sgRNA是GX19或gX19,使用从PAM起在5'方向上的20个核苷酸(nt)的序列。在一种新型方法中,在ggX20形式(其中在从PAM起在5'方向上的20个核苷酸前添加两个另外的错配鸟嘌呤(G))中使用延伸的sgRNA进行实验,或者在gX21-gX30形式中使用21至30个核苷酸序列进行实验。作为用ABE(腺苷碱基编辑器)和延伸的sgRNA在HEK2位点对HEK293T细胞进行实验的结果,常规GX19 sgRNA在从PAM起的第13至第17个腺苷中显示出突变,并且当使用gX20/gX21/gX22 sgRNA时,第18和第19个腺苷也发生了改变(参见图2A、图2B和图2C)。可以看出,与GX19 sgRNA中显示的效率相比,使用gX20/gX21/gX22 sgRNA引入的第18和19个腺苷突变的效率增加了10倍或更多(图2B)。类似地,作为用CBE(胞嘧啶碱基编辑器)和延伸的sgRNA在HBB位点的观察的结果,发现当使用gX20/gX22 sgRNA时,突变被引入位置20、位置21和位置23处的胞嘧啶中(参见图2D、图2E和图2F)。另外,发现当使用具有另外的错配鸟嘌呤的ggX20 sgRNA时,在位置20至23处的胞嘧啶中通过CBE产生突变的发生率增加了3倍或更多(参见图3)。当将CBE和ABE各自在四个不同的靶位点测试时,使用延伸的sgRNA代替GX19 sgRNA将碱基编辑的操作范围扩展到位置18至23,这是与常规的碱基编辑范围(从PAM起在5'方向的位置13至17)相比更远的区域,并且将效率增加了多达5至60倍(参见图4)。

[0026] 在一个方面,本公开文本涉及可与靶序列杂交的延伸的向导RNA,所述延伸的向导RNA在5'末端处进一步包括1至3个鸟嘌呤(G)和1至10个核苷酸(其中所述核苷酸各自独立

地选自A、T、C和G)。

[0027] 本公开文本的延伸的向导RNA可以是单链形式(单向导RNA;sgRNA)。延伸的向导RNA可以在常规的向导RNA(例如sgRNA)的5'末端进一步包括1至10个核苷酸(其中所述核苷酸中的每一个独立地选自A、T、C和G,例如是与相应的DNA靶序列互补的序列)(靶向序列为20nt;具体而言,在5'末端的第一个核苷酸可以是与相应的DNA靶位点序列匹配(互补)的鸟嘌呤(G),或者可以是与之不匹配(不互补)的鸟嘌呤(G))。与其他类型的sgRNA相比,这种延伸形式的sgRNA可以增加碱基编辑频率和/或编辑效率。

[0028] 另外,延伸的sgRNA可以在5'末端进一步包括一至三个与相应的DNA靶序列匹配(互补)的鸟嘌呤(G),或者一至三个与相应的DNA靶序列不匹配(不互补)的鸟嘌呤(G)。在5'末端另外包括的所述1至10个随机核苷酸可以与相应靶位点的靶DNA序列互补,因此在靶位点处从PAM起在5'方向所暴露的单链DNA的长度可以增加,使基因编辑(脱氨作用)能够在更广的范围内发生(例如,即使在靶位点处从PAM起在5'方向的位置18-30nt或18-22nt也可以引入突变(碱基编辑))(参见图1B)。

[0029] 因此,在一个方面,本公开文本涉及一种用于碱基编辑的组合物,其包括(i)脱氨酶或编码所述脱氨酶的基因;(ii)RNA导向的工程化核酸酶或编码所述RNA导向的工程化核酸酶的基因;和(iii)可与靶序列杂交的延伸的向导RNA或编码所述延伸的向导RNA的基因。

[0030] 在本公开文本的一个实施方案中,用于碱基编辑的组合物包括(1)脱氨酶或编码所述脱氨酶的基因,(2)靶标特异性核酸酶(RNA导向的工程化核酸酶)或编码所述靶标特异性核酸酶的基因,和(3)可与靶基因的靶位点杂交(或具有与之互补的核苷酸序列)的向导RNA,或编码所述向导RNA的DNA(或包括所述DNA的重组载体)。此时,如上所述,向导RNA可以是延伸的向导RNA,其在常规的向导RNA(例如sgRNA)的5'末端进一步包括1至10个核苷酸(其中每个核苷酸独立地选自A、T、C和G,例如,是与相应DNA靶序列互补的序列),并且在sgRNA的5'末端可以进一步包括一至三个匹配或错配的鸟嘌呤(G)。

[0031] 用于碱基编辑的组合物在真核细胞中可以具有碱基编辑(例如碱基置换)活性。真核细胞可以是真核动物的细胞,诸如胚胎细胞,或者可以是真核植物(例如,藻类、单子叶植物、双子叶植物等)的细胞,并且在一个特定例子中,真核细胞可以是哺乳动物细胞(诸如哺乳动物胚胎细胞)、或真核植物细胞。本文所用的编码基因能以cDNA、rDNA、含有它们的重组载体、或mRNA的形式使用。

[0032] “脱氨酶”泛指具有从真核细胞中的特定碱基上去除胺基的活性的酶,并且可以是例如将胞苷转换为尿苷的胞苷脱氨酶和/或腺苷脱氨酶。在一个例子中,脱氨酶可以包括选自载脂蛋白B编辑复合物1(APOBEC1)、激活诱导脱氨酶(AID)和tRNA特异性腺苷脱氨酶(tadA)中的一种或多种,但不限于此。真核细胞中的单核苷酸取代可以通过此类碱基转换(例如,胞苷至尿苷的转换)来诱导。

[0033] 在一个例子中,除了(1)脱氨酶或编码所述脱氨酶的基因(含有mRNA或编码DNA的重组载体)、(2)RNA导向的工程化核酸酶或编码所述RNA导向的工程化核酸酶的基因(含有mRNA或编码DNA的重组载体)和(3)延伸的向导RNA或编码所述延伸的向导RNA的基因(DNA)以外,所述用于碱基编辑的组合物可以进一步包括(4)尿嘧啶DNA糖基化酶抑制剂(UGI)或编码所述尿嘧啶DNA糖基化酶抑制剂的基因和/或(5)核定位序列(NLS)或编码所述核定位序列的基因。

[0034] 在本公开文本的用于碱基编辑的组合物中,当使用脱氨酶、RNA导向的工程化核酸酶、以及任选地UGI和/或NLS连接的融合蛋白或与其编码基因连接的融合基因时,在蛋白质或基因之间,例如在脱氨酶与RNA导向的工程化核酸酶之间、核酸酶与UGI之间、以及在UGI与NLS之间中的一者或多者中可以进一步包括至少一个合适的接头(在融合蛋白的情况下为肽接头(3-30个或3-20个氨基酸),并且在融合基因的情况下为寡核苷酸接头(9-90或9-60nt))。

[0035] 在一个例子中,RNA导向的工程化核酸酶可以是被修饰以消除其基因双链切割活性的经修饰的RNA导向的工程化核酸酶。

[0036] 所述经修饰的RNA导向的工程化核酸酶是被修饰以切割靶基因的一条链(形成切口)的经修饰的Cas9(CRISPR相关蛋白9)系统或经修饰的Cpf1(源自普雷沃菌(*Prevotella*)和弗朗西斯菌(*Francisella*)的CRISPR 1)系统。在一个例子中,经修饰的RNA导向的工程化核酸酶可以选自Cas9切口酶(nCas9)、催化缺陷型Cas9(dCas9)等。

[0037] 在本公开文本中,当用于碱基编辑的组合物包括脱氨酶编码基因和RNA导向的核酸酶编码基因时,所述编码基因可以是编码DNA或mRNA。另外,脱氨酶编码基因和RNA导向的工程化核酸酶编码基因以mRNA的形式、或以在单独载体中包括一个所述基因(DNA)的重组载体(即,包括编码脱氨酶的DNA的重组载体和包括编码DNA导向的核酸酶的DNA的重组载体)的形式、或以在一个载体中包括多个所述基因(DNA)的重组载体的形式被包括。

[0038] 向导RNA可以是双向导RNA,包括CRISPR RNA(crRNA)、反式激活crRNA(tracrRNA)、crRNA和tracrRNA(crRNA和tracrRNA的复合物);或单向导RNA(sgRNA)。在一个例子中,用于碱基编辑的组合物可以包括核糖核蛋白(RNP),所述核糖核蛋白包括编码脱氨酶和经修饰的RNA导向的核酸酶的mRNA、和向导RNA,或者包括脱氨酶和经修饰的RNA导向的核酸酶和向导RNA。所述核糖核蛋白可以包括脱氨酶、经修饰的RNA导向的核酸酶和向导RNA的混合物,或者可以包括脱氨酶、经修饰的RNA导向的核酸酶和向导RNA的复合物。

[0039] 在另一方面,本公开文本提供了一种用于进行碱基编辑的方法,所述方法包括将用于碱基编辑的组合物引入细胞中。

[0040] 在另一方面,本公开文本提供了一种用于进行碱基编辑的方法,所述方法包括将用于碱基编辑的组合物引入细胞中。所述细胞可以是真核细胞,并且所述碱基编辑方法可以通过在真核细胞中进行碱基编辑(例如,碱基置换)来进行。

[0041] 真核细胞可以是真核动物的细胞(诸如真核动物的胚胎细胞)和/或真核植物的细胞,并且在一个特定例子中,真核细胞可以是哺乳动物细胞(诸如哺乳动物胚胎细胞)和/或真核植物细胞。在真核细胞(例如,真核胚胎细胞和/或真核植物细胞)中,所述碱基编辑方法能够获得40%或更高、45%或更高、50%或更高、55%或更高、60%或更高、65%或更高、70%或更高、75%或更高、80%或更高、85%或更高、90%或更高、95%或更高、97%或更高、99%或更高或100%的碱基转换率(碱基置换率)。另外,所述碱基编辑方法可以通过碱基置换在基因(例如,编码序列)中产生终止密码子而敲除所述基因、或通过在不产生蛋白质的非编码DNA序列中引入突变来诱导多种突变体。

[0042] 在本公开文本的一个方面,在油菜(甘蓝型油菜(*Brassica napus*))和大豆(soybean或*Glycine max*)中检查了碱基编辑是否可以通过增加基于其的sgRNA的长度在更广的范围内进行。在源自甘蓝型油菜子叶的原生质体中,将能够靶向ALS基因的gX19和gX20

sgRNA转染到AID2碱基编辑器中,所述ALS基因是除草剂抗性基因。结果,当使用gX20 sgRNA时,位置20处的胞嘧啶变为胸腺嘧啶(图5A)。仅在这种情况下产生终止密码子,从而敲除了相应基因(图5B)。当对从作为另一种作物的大豆的愈伤组织获得的原生质体进行转染以靶向ALS基因时,当使用gX20 sgRNA时,增加了位置20处的胞嘧啶向胸腺嘧啶的转换效率(参见图5C和图5D)。最后,使用ABE将称为引起白化病的突变的H420R置换引入小鼠酪氨酸酶基因中。作为在改变靶向酪氨酸酶的sgRNA形式的同时,在与ABE mRNA一起向小鼠胚胎进行显微注射之后在胚泡期进行分析的结果,发现当使用GX20或GX21 sgRNA而不是gX19时,改变位置18处的腺苷的效率增加(图6A)。为了引入所需的H420R突变,应改变位置18处的腺苷。因此,使用能够以最高效率改变相应位置的GX21 sgRNA获得了小鼠幼崽。其中,发现三只幼崽具有H420R突变(图6B)。如上所述,当有必要在常规碱基编辑窗口之外在5'方向上编辑碱基时,发现使用延伸的sgRNA比一般的GX19 sgRNA更有效。

[0043] 在另一方面,本公开文本提供了一种用于产生除人以外的哺乳动物或真核植物的突变成体的方法,所述方法包括(a)将用于碱基编辑的组合物引入哺乳动物胚胎或真核植物胚胎中并且(b)使所述胚胎生长以获得成体。

[0044] 具体而言,本公开文本的用于碱基编辑的组合物可以有效地应用于哺乳动物或真核植物成体的产生,其中通过将所述组合物应用于哺乳动物胚胎或真核植物胚而灭活所需基因或诱导所需突变。

[0045] 将用于碱基编辑的组合物引入细胞中的步骤包括将脱氨酶或编码脱氨酶的基因、RNA导向的核酸酶或编码RNA导向的核酸酶的基因以及延伸的向导RNA或编码延伸的向导RNA的基因引入细胞中。可以将所述编码基因中的一种或多种以包括在单独的重组载体或单个重组载体中的状态引入。

[0046] 在一个例子中,将用于碱基编辑的组合物引入细胞中的步骤可以以下列方式进行:

[0047] 1) 用包括编码脱氨酶的DNA、编码RNA导向的工程化核酸酶的DNA和编码延伸的向导RNA的基因中的一种或两种或更多种的重组载体转染细胞,

[0048] 2) 将脱氨酶、RNA导向的核酸酶和延伸的向导RNA(例如,含有脱氨酶、RNA导向的核酸酶和延伸的向导RNA的混合物或复合物形式的核糖核蛋白)直接注射到细胞中,或

[0049] 3) 将编码脱氨酶的mRNA、编码RNA导向的核酸酶的mRNA和向导RNA中的每一种或其混合物直接注射到细胞中。

[0050] “直接注射”意指脱氨酶、RNA导向的工程化核酸酶和延伸的向导RNA(例如,2)的含有脱氨酶、RNA导向的工程化核酸酶和延伸的向导RNA的混合物或复合物形式的核糖核蛋白,或3)的编码脱氨酶的mRNA、编码RNA导向的工程化核酸酶的mRNA和延伸的向导RNA)在没有使用重组载体的情况下穿过细胞膜和/或核膜,然后递送至基因组;并且可以通过例如电穿孔、脂质转染、显微注射等进行。

[0051] 在另一方面,本公开文本提供了基因修饰细胞,其包括通过碱基编辑方法编辑的碱基。基因修饰细胞可以是其中在靶基因中由于碱基编辑而发生碱基置换(例如单碱基置换或点突变)的细胞。所述细胞可以是真核细胞。真核细胞可以是真核动物细胞(诸如胚胎细胞)和/或真核植物细胞,并且在一个实施方案中,可以是包括或不包括人的哺乳动物细胞(诸如包括或不包括人的哺乳动物胚胎细胞)和/或真核植物细胞。

[0052] 在另一方面,本公开文本提供了一种用于产生经基因修饰的动物的方法,所述方法包括将注射了用于碱基编辑的组合物的哺乳动物胚胎或包括通过所述碱基编辑方法编辑的碱基的经基因修饰的哺乳动物胚胎移植到哺乳动物的输卵管中以产生基因修饰动物。经基因修饰的哺乳动物可以是源自由于碱基编辑而在靶基因中具有碱基置换(例如,单碱基置换或点突变)的胚胎的动物。

[0053] 向其输卵管中移植了胚胎细胞的哺乳动物可以是与所述胚胎细胞所来源的哺乳动物相同物种的哺乳动物(寄主)。

[0054] 在另一方面,本公开文本提供了源自基因修饰细胞的基因修饰动物。可以通过用于产生基因修饰动物的方法来产生基因修饰动物。所述动物可以是真核动物,诸如哺乳动物,包括人或非人哺乳动物。

[0055] 本文中应用了用于碱基编辑的组合物的细胞可以是真核细胞,诸如真核动物细胞。所述真核动物可以是哺乳动物,包括诸如人的灵长类动物或诸如小鼠的啮齿动物。所述真核动物细胞可以是哺乳动物胚胎。例如,所述胚胎可以通过使雄性哺乳动物与诱导超排卵的雌性哺乳动物(例如,通过注射性激素诸如孕马血清促性腺激素(PMSG)或人绒毛膜促性腺激素(hCG)来诱导超排卵)杂交而获得的受精胚胎,其中可以从雌性哺乳动物的输卵管中收集所述受精胚胎。应用(注射)了用于碱基编辑的组合物的胚胎可以是受精的1-细胞期胚胎(受精卵)。

[0056] 如本文所用,术语“碱基编辑”是指在靶基因内的靶位点引起点突变(诸如由于基因或基因水平的点突变而引起的单个氨基酸突变)的碱基突变(置换、缺失或添加),并且与涉及相对大量的碱基的突变的基因编辑的区别之处在于仅少数碱基(一个或两个碱基,例如一个碱基)被突变。碱基编辑可以不涉及所述基因的双链DNA切割。

[0057] 根据本文提供的用于碱基编辑的组合物或方法,碱基编辑(碱基修饰或碱基置换;通过A或C的脱氨作用而引起的突变)可以发生在与带切口的DNA链(与PAM序列所在的链相对的链、与向导RNA结合(杂交)的链)相对的链(即,PAM序列所在的链)中。例如,当使用具有正常长度的向导RNA时,在从PAM起在5'方向的位置17处的核苷酸中发生碱基编辑(碱基修饰或碱基取代),但是当使用本文提供的延伸的向导RNA时,也可以在从PAM起在5'方向的位置17之后的区域中,例如,在与从PAM起在5'方向的位置18至30、位置18至25或位置18至22相对应的扩展范围内发生碱基编辑。

[0058] 如本文所用,术语“碱基突变(或碱基置换)”意指在包括碱基的核苷酸中已经发生了突变(例如取代),并且可以与“核苷酸突变(或核苷酸取代)”互换使用,并且这样的碱基突变可以在一个或两个等位基因中发生。

[0059] 在一个例子中,碱基突变或涉及碱基突变的碱基编辑可以通过多种方法进行,例如,通过敲除靶基因或通过靶位点产生终止密码子而将突变引入不产生蛋白质的非编码DNA序列中,或产生编码不同于野生型的氨基酸的密码子,但不限于此。

[0060] 在本公开文本中,碱基编辑或碱基突变可以在体外或体内进行。

[0061] 如本文所用,术语“碱基序列”是指含有相应碱基的核苷酸的序列,并且可以与“核苷酸序列”或“核酸序列”互换使用。

[0062] 如本文所用,术语“靶基因”是指作为进行碱基编辑(或碱基突变)的对象的基因,并且术语“靶位点”或“靶区域”意指靶基因中由靶标特异性核酸酶引起碱基编辑的位点。例

如,当靶标特异性核酸酶包括RNA导向的工程化核酸酶(RGEN)时,靶位点意指靶基因中位于由RNA导向的工程化核酸酶(RGEN)识别的序列(PAM序列)的5'末端和/或3'末端附近的基因位点(双链或双链中的任何一条),并且具有约50bp或约40bp的最大长度。

[0063] 在一个例子中,当靶标特异性核酸酶包括RNA导向的工程化核酸酶时,其可以进一步包括包含靶向序列的向导RNA以及RNA导向的工程化核酸酶。术语“靶向序列”可以是指向导RNA的位点,其包括与靶区域中连续的含有约20个核苷酸(nt)的区域的碱基序列互补(可杂交)的碱基序列。本文所述的延伸的向导RNA在5'末端进一步包括1至10个另外的任选核苷酸(其中所述核苷酸选自A、T、C和G;例如,可以与相应的靶序列互补)和/或可以在5'末端包括1-3个另外的匹配或错配鸟嘌呤。在5'末端的所述1至10个另外的任意核苷酸可以是与对应于其的延伸的靶DNA区域的序列互补的序列,由此靶区域中PAM在5'方向上暴露的单链DNA的长度可以增加,以允许在更广的范围内进行基因编辑(脱氨作用)。

[0064] 在本公开文本中,包括与靶向序列互补的碱基序列的靶位点的碱基序列可以被称为“靶序列”,并且所述靶序列可以是长度为约20nt的连续碱基序列或是和与之互补的链相对应的位点,所述位点位于由RNA导向的工程化核酸酶(RGEN)识别的PAM序列的5'末端和/或3'末端附近。

[0065] 脱氨酶是指具有从真核细胞中的特定碱基上去除氨基的活性的酶,并且可以是例如将胞苷转换为尿苷的胞苷脱氨酶和/或腺苷脱氨酶。在一个例子中,脱氨酶可以包括选自APOBEC(载脂蛋白B mRNA编辑酶、催化多肽样酶)、AID(激活诱导的脱氨酶)和tadA(tRNA特异性腺苷脱氨酶)等中的一种或多种,但不限于此。APOBEC1、AID和tadA可以源自原核动物诸如大肠杆菌(*E. coli*),或者可以源自真核动物诸如灵长类动物(包括人)、或哺乳动物诸如啮齿动物(包括小鼠)。

[0066] 可以将脱氨酶以蛋白质、编码所述蛋白质的基因(例如,DNA或mRNA)或含有所述基因的重组载体的形式使用。如本文所用,靶标特异性核酸酶也被称为“基因编辑器”(可编程核酸酶),并且统称为能够识别和切割(单链或双链切割)所需基因组DNA上的特定位置的核酸酶(例如,内切核酸酶)。

[0067] 例如,靶标特异性核酸酶可以是选自识别靶基因的特定序列并且由于其核苷酸切割活性而在靶基因中引起插入和/或缺失(Indel)的所有核酸酶中的一种或多种。

[0068] 例如,靶标特异性核酸酶可以包括选自源自作为微生物免疫系统的CRISPR的RGEN(RNA导向的工程化核酸酶;例如,Cas蛋白(例如,Cas9)、Cpf1等)中的一种或多种,但不限于此。

[0069] 靶标特异性核酸酶可以识别原核细胞和/或动植物细胞(例如,真核细胞)(包括人类细胞)的基因组中的特定核苷酸序列,以引起双链断裂(DSB)。双链断裂可以通过使DNA的双螺旋断裂而形成平端或粘端。DSB可以在细胞中通过同源重组或非同源末端连接(NHEJ)机制得以有效修复。在此过程中,可以将所需突变引入靶位点中。

[0070] 例如,靶标特异性核酸酶可以包括选自II型和/或V型CRISPR系统中所涉及的核酸酶(例如,内切核酸酶)中的一种或多种,所述核酸酶诸如Cas蛋白(例如,Cas9蛋白(CRISPR(成簇规律间隔短回文重复序列)相关蛋白9))和Cpf1蛋白(源自普雷沃菌和弗朗西斯菌的CRISPR 1)。在这种情况下,靶标特异性核酸酶进一步包括靶DNA特异性向导RNA来为基因组DNA的靶位点导向。向导RNA可以在体外转录,并且可以例如从双链寡核苷酸或质粒模板转

录,但不限于此。靶标特异性核酸酶可以在体外或体内递送至细胞之后,通过形成与向导RNA结合的核糖核酸-蛋白质复合物(RNA导向的工程化核酸酶)来充当核糖核酸蛋白(RNP)。Cas蛋白是CRISPR/Cas系统的主要蛋白质组分,并且是能够形成激活的内切核酸酶或切口酶的蛋白质。

[0071] Cas蛋白或基因信息可以从已知数据库获得,所述数据库诸如国家生物技术信息中心(National Center for Biotechnology Information,NCBI)的GenBank。例如,Cas蛋白可以包括选自以下中的一种或多种:源自链球菌属(*Streptococcus*)物种(例如,酿脓链球菌(*Streptococcus pyogenes*))的Cas9蛋白;Cas9蛋白(例如,SwissProt登录号Q99ZW2(NP_269215.1));源自弯曲杆菌属(*Campylobacter*)物种(例如空肠弯曲杆菌(*Campylobacter jejuni*))的Cas蛋白;源自链球菌属(*Streptococcus*)物种(诸如嗜热链球菌(*Streptococcus thermophiles*)或金黄色链球菌(*Streptococcus aureus*))的Cas蛋白;源自脑膜炎奈瑟氏球菌(*Neisseria meningitidis*)的Cas9蛋白;源自巴斯德菌属(*Pasteurella*)物种(诸如多杀巴斯德菌(*Pasteurella multocida*))的Cas蛋白;和源自弗朗西斯菌属物种(例如新凶手弗朗西斯菌(*Francisella novicida*))的Cas9蛋白,但不限于此。

[0072] 在本公开文本中,Cpf1蛋白是与CRISPR/Cas系统有区别的新CRISPR系统的内切核酸酶,与Cas9相比尺寸相对较小,不需要tracrRNA,并且可以通过单向导RNA起作用。另外,Cpf1蛋白识别富含胸腺嘧啶的前间区序列邻近基序(PAM)序列并切割DNA双链以产生粘端(粘性双链断裂)。

[0073] 例如,Cpf1蛋白可以源自Candidatus物种、毛螺菌属(*Lachnospira*)物种、丁酸弧菌属(*Butyrivibrio*)物种、异域菌门(*Peregrinibacteria*)物种、氨基酸球菌属(*Acidaminococcus*)物种、卟啉单胞菌属(*Porphyromonas*)物种、普雷沃菌属物种、弗朗西斯菌属物种、Candidatus *Methanoplasma*或真杆菌属(*Eubacterium*)物种,例如微生物,诸如俭菌总门(*Parcubacteria*)细菌(GWC2011_GWC2_44_17)、毛螺菌科(*Lachnospiraceae*)细菌(MC2017)、解蛋白丁酸弧菌(*Butyrivibrio proteoclasticus*)、异域菌门细菌(GW2011_GWA_33_10)、氨基酸球菌属物种(BV3L6)、猕猴卟啉单胞菌(*Porphyromonas macacae*)、毛螺菌科细菌(ND2006)、狗口腔卟啉单胞菌(*Porphyromonas crevioricanis*)、解糖豚普雷沃菌(*Prevotella disiens*)、牛眼莫拉氏菌(*Moraxella bovoculi*) (237)、史密斯氏菌属(*Smithella*)物种(SC_K08D17)、稻田钩端螺旋体(*Leptospira inadai*)、毛螺菌科细菌(MA2020)、新凶手弗朗西斯菌(U112)、Candidatus *Methanoplasma*termitum、Candidatus *Paceibacter*和挑剔真杆菌(*Eubacterium eligens*),但本公开文本不限于此。

[0074] 靶标特异性核酸酶可以从微生物中分离,或者可以通过诸如重组方法或合成方法的方法人工或非天然获得。靶标特异性核酸酶可以是预转录的mRNA或在体外预先产生的蛋白质,或者可以包括在重组载体中在靶细胞中进行体内表达。在一个例子中,靶标特异性核酸酶(例如,Cas9、Cpf1等)可以由重组DNA(rDNA)产生的重组蛋白。重组DNA是指通过诸如分子克隆的基因重组方法人工产生而包括从各种生物体中获得的异种或同种遗传物质的DNA分子。例如,当重组DNA在适当的生物体中表达以产生靶标特异性核酸酶(在体内或在体外)时,重组DNA可以具有通过在编码待产生的蛋白质的密码子中选择经优化以在生物体中表达的密码子而重新构建的核苷酸序列。

[0075] 本文所用的靶标特异性核酸酶可以是突变的靶标特异性核酸酶的突变形式。突变的靶标特异性核酸酶可以意指被突变而丧失切割DNA双链的内切核酸酶活性的核酸酶,例如,可以包括选自被突变而丧失内切核酸酶活性但具有切口酶活性的突变的靶标特异性核酸酶、和被突变而丧失内切核酸酶活性和切口酶活性两者的突变的靶标特异性核酸酶中的至少一种。

[0076] 当突变的靶标特异性核酸酶具有切口酶活性时,可以同时或依次地向发生碱基编辑的链或与之相对的链中引入切口,无论其序列如何,通过脱氨酶进行碱基转换(例如,胞苷转换为尿苷)(例如,在与PAM所在的链相对的链中,在PAM序列的5'末端方向上的第三个核苷酸和第四个核苷酸之间引入切口)。靶标特异性核酸酶的突变(例如,氨基酸置换)可以至少在核酸酶的催化活性结构域(例如,在Cas9的情况下为RuvC催化结构域)中发生。在一个例子中,当靶标特异性核酸酶是酿脓链球菌来源的Cas9蛋白(SwissProt登录号Q99ZW2(NP_269215.1))时,所述突变可以包括选自以下中的至少一种的置换形式的突变:具有催化活性的催化性天冬氨酸残基(位置10的天冬氨酸(D10))、位置762的谷氨酸(E762)、位置840的组氨酸(H840)、位置854的天冬酰胺(N854)、位置863的天冬酰胺(N863)、位置986的天冬氨酸(D986)等,以及任何其他氨基酸。在这种情况下,经置换的任何其他氨基酸可以是丙氨酸,但不限于此。

[0077] 在另一个例子中,突变的靶标特异性核酸酶可以被突变为识别不同于野生型Cas9蛋白的PAM序列。例如,突变的靶标特异性核酸酶可以通过用其他氨基酸置换源自酿脓链球菌的Cas9蛋白的位置1135的天冬氨酸(D1135)、位置1335的精氨酸(R1335)和位置1337的苏氨酸(T1337)中的至少一个(例如,三个)而被突变为识别与野生型Cas9的PAM序列(NGG)不同的NGA(其中N是选自A、T、G和C的任何碱基)。

[0078] 在一个例子中,突变的靶标特异性核酸酶可以在酿脓链球菌来源的Cas9蛋白的氨基酸序列的以下区域中具有氨基酸置换:

[0079] (1) D10、H840、或D10+H840;

[0080] (2) D1135、R1335、T1337、或D1135+R1335+T1337;或

[0081] (3) 残基(1)和(2)两者。

[0082] 如本文所用,术语“其他氨基酸”意指选自除野生型蛋白在其原始突变位置中所具有的氨基酸以外的氨基酸,包括丙氨酸、异亮氨酸、亮氨酸、甲硫氨酸、苯丙氨酸、脯氨酸、色氨酸、缬氨酸、天冬酰胺、半胱氨酸、谷氨酰胺、甘氨酸、丝氨酸、苏氨酸、酪氨酸、天冬氨酸、谷氨酸、精氨酸、组氨酸、赖氨酸、以及上述氨基酸的所有已知变体。例如,“其他氨基酸”可以是丙氨酸、缬氨酸、谷氨酰胺或精氨酸。

[0083] 在一个例子中,突变的靶标特异性核酸酶可以识别不同于野生型Cas9的PAM序列或丧失内切核酸酶活性(例如,具有切口酶活性,或同时丧失内切核酸酶活性和切口酶活性)的经修饰的Cas9蛋白。例如,经修饰的Cas9蛋白是在源自酿脓链球菌的Cas9蛋白中:

[0084] (1) 由于在位置D10或H840处引入突变(例如,被其他氨基酸置换)而丧失内切核酸酶活性但具有切口酶活性的经修饰的Cas9蛋白,或者由于向源自酿脓链球菌的Cas9蛋白中的位置D10或H840处引入突变(例如,被其他氨基酸置换)而丧失内切核酸酶活性和切口酶活性两者的Cas9蛋白;

[0085] (2) 由于在D1135、R1335和T1337中的一个或多个中引入突变(例如,被其他氨基酸

置换)而识别不同于野生型的PAM序列的经修饰的Cas9蛋白;或

[0086] (3)具有切口酶活性并且识别不同于野生型的PAM序列,或由于(1)和(2)的突变的引入而丧失内切核酸酶活性和切口酶活性两者并且识别不同于野生型的PAM序列的经修饰的Cas9蛋白。

[0087] 例如,CAs9蛋白的位置D10处的突变可以是D10A突变(意指Cas9蛋白氨基酸的第10个氨基酸(D)被(A)置换;下文中,引入Cas9中的突变可以按上述相同的方式表示),位置H840处的突变可以用“H840A突变”表示,并且位置D1135、R1335和T1337处的突变可以分别用D1135V、R1335Q和T1337R表示。

[0088] 如本文所用,除非另有说明,否则术语“核酸酶”意指“靶标特异性核酸酶”,诸如如上所述的Cas9或Cpf1。

[0089] 核酸酶可以从微生物中分离,或者可以通过诸如重组方法或合成方法的方法人工或非天然获得。在一个例子中,核酸酶(例如,Cas9、Cpf1等)可以是由重组DNA产生的重组蛋白。“重组DNA(rDNA)”是指通过诸如分子克隆的基因重组方法人工产生而含有从各种生物体获得的异源或同源遗传物质的DNA分子。例如,当重组DNA在适当的生物体中表达以产生蛋白质(核酸酶)(在体内或在体外)时,重组DNA可以具有通过在编码待产生的蛋白质的密码子中选择经优化以在生物体中表达的密码子而重新构建的核苷酸序列。

[0090] 可以将核酸酶以蛋白质、编码所述蛋白质的核酸分子(例如,DNA或mRNA)、与向导RNA连接的核糖核蛋白、编码所述核糖核蛋白的核酸分子、或含有所述核酸分子的重组载体的形式使用。

[0091] 脱氨酶和核酸酶和/或编码所述脱氨酶和核酸酶的核酸分子可以呈可以递送至细胞核、在细胞核上起作用和/或在细胞核中表达的形式。

[0092] 脱氨酶和核酸酶可以具有使得能够易于引入细胞中的形式。例如,脱氨酶和核酸酶可以与细胞穿透肽和/或蛋白质转导结构域连接。蛋白质转导结构域可以是聚精氨酸或HIV来源的TAT蛋白,但不限于此。由于除了上述例子以外,多种种类的细胞穿透肽或蛋白质转导结构域是本领域已知的,因此本领域技术人员将理解,本公开文本不限于以上例子,并且多种例子可以适用。

[0093] 另外,脱氨酶和核酸酶和/或编码所述脱氨酶和核酸酶的核酸分子可以进一步包括核定位信号(NLS)序列或编码所述核定位信号序列的核酸序列。因此,包括编码脱氨酶的核酸分子和/或编码核酸酶的核酸分子的表达盒可以包括调控序列,诸如用于表达脱氨酶和/或核酸酶的启动子序列,并且任选地可以进一步包括NLS序列(CCCAAGAAGAAGAGGAAAGTC:SEQ ID NO:2)。NLS序列是本领域熟知的。

[0094] 脱氨酶和核酸酶和/或编码所述脱氨酶和核酸酶的核酸分子可以与用于分离和/或纯化的标签或与编码所述标签的核酸序列连接。例如,所述标签可以适当地选自小肽标签(诸如His标签、Flag标签和S标签)、GST(谷胱甘肽S-转移酶)标签和MBP(麦芽糖结合蛋白)标签,但是不限于此。

[0095] 另外,本文所用的用于碱基编辑的组合物可以进一步包括尿嘧啶DNA糖基化酶抑制剂(UGI)或编码所述尿嘧啶DNA糖基化酶抑制剂的基因(含有编码DNA的重组载体形式或体外转录的mRNA形式)。当用于碱基编辑的组合物进一步包括尿嘧啶DNA糖基化酶抑制剂时,使用脱氨酶置换的特定碱基(例如,通过胞嘧啶脱氨酶用T替换C)的比例与没有尿嘧啶

DNA糖基化酶抑制剂的情况相比增加,并且当不进一步包括尿嘧啶DNA糖基化酶抑制剂时,除特定碱基置换(例如,通过胞嘧啶脱氨酶用T置换C)以外的碱基置换的比例增加(即发生多种形式的碱基置换)。

[0096] 在本公开文本中,术语“向导RNA”是指包括如下靶向序列的RNA,所述靶向序列可以与靶基因中的靶位点内的特定碱基序列(靶序列)杂交,并在体外或在体内与核酸酶诸如Cas蛋白或Cpf1结合以将其导向至靶基因(或靶位点)。可以根据形成复合物的核酸酶类型和/或其来源微生物适当地选择向导RNA。

[0097] 例如,向导RNA可以包括选自以下中的至少一种:

[0098] CRISPR RNA (crRNA),其包括可与靶序列杂交的位点(靶序列);

[0099] 反式激活crRNA (tracrRNA),其包括与核酸酶(诸如Cas蛋白或Cpf1)相互作用的位点;和

[0100] 单向导RNA (sgRNA),其与crRNA和tracrRNA的主要位点(例如,含有靶向序列的crRNA位点和与核酸酶相互作用的tracrRNA位点)融合。

[0101] 具体地,向导RNA可以是包括CRISPR RNA (crRNA)和反式激活crRNA (tracrRNA)的双RNA,或者可以是包括crRNA和tracrRNA的主要位点的单向导RNA (sgRNA)。

[0102] sgRNA可以包括与靶序列中的靶基因(靶位点)具有互补序列(靶向序列)的部分(也称为“间隔区”、“靶DNA识别序列”或“碱基配对区”),以及用于Cas蛋白结合的发夹结构。更具体地,sgRNA可以包括具有与靶基因中的靶序列互补的序列(靶向序列)的部分、用于Cas蛋白结合的发夹结构、和终止子序列。所述结构可以从5'到3'存在,但是本公开文本不限于此。只要向导RNA包括crRNA和tracrRNA的主要部分以及与靶DNA的互补部分,任何类型的向导RNA都可以作为上述结构用于本公开文本中。

[0103] 例如,Cas9蛋白需要两个向导RNA来进行靶基因编辑,即具有可与靶基因的靶位点杂交的核苷酸序列的CRISPR RNA (crRNA)和与Cas9蛋白相互作用的反式激活crRNA (tracrRNA),并且这些crRNA和tracrRNA可以以彼此连接的双链crRNA:tracrRNA复合物的形式使用,或通过接头连接的单向导RNA (sgRNA)的形式使用。在一个例子中,当使用源自酿脓链球菌的Cas9蛋白时,sgRNA可以具有由整个crRNA或至少包括crRNA的可杂交的核苷酸序列的部分和整个tracrRNA或至少包括Cas9的tracrRNA的与Cas9蛋白相互作用的位点的部分经由核苷酸接头(在这种情况下,核苷酸接头可以与环结构相对应)形成的发夹(茎-环)结构。

[0104] 向导RNA,具体地crRNA或sgRNA,包括与靶基因中的靶序列互补的序列(靶向序列),并且可以在crRNA或sgRNA上游区域,具体地在双RNA的sgRNA或crRNA的5'末端包括一个或多个,例如1至10个、1至5个或1至3个另外的核苷酸。所述另外的核苷酸可以是鸟嘌呤(G),但不限于此。

[0105] 在另一个例子中,当核酸酶是Cpf1时,向导RNA可以包括crRNA,并且可以根据形成复合物的Cpf1蛋白的类型和/或其来源微生物适当地选择。

[0106] 可以根据核酸酶的类型(Cas9或Cpf1)(即,其来源微生物)适当地选择向导RNA的特异性序列,这可以被本公开文本所属领域的技术人员容易地理解。

[0107] 在一个例子中,当使用源自酿脓链球菌的Cas9蛋白作为靶标特异性核酸酶时,sgRNA可以用以下通式1表示:

[0108] 例如,向导RNA可以由以下通式(1)表示:

[0109] $5'-(N_{\text{cas9}})_1-(\text{GUUUUAGAGCUA})-(\text{寡核苷酸接头})-$

[0110] $(\text{UAGCAAGUUAUUUUUAAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC}(\text{SEQ. ID NO.1})-3'$ (通式1)

[0111] 在上述通式1中,

[0112] 在 $(N_{\text{cas9}})_1$ 中, N_{cas9} 表示与靶基因的靶位点结合(杂交)的靶向序列,并且根据靶位点的序列确定核酸序列(即,可以与靶位点杂交的序列),1表示靶向序列中包括的核苷酸的数量,可以是20,并且从5'末端起的第一个核酸可以是与靶位点序列匹配的鸟嘌呤(以G表示;靶标;当靶位点的相应位置是胞嘧啶(C)时)或与靶位点序列不匹配的鸟嘌呤(以g表示;当靶位点的相应位置不是胞嘧啶(C)时)。

[0113] 寡核苷酸接头可以包括3至5个核苷酸,例如4个核苷酸,并且所述核苷酸可以彼此相同或不同,并且可以各自独立地选自A、U、C和G。

[0114] 例如,其中 N_{cas9} 由总共20个核苷酸组成的情况可以用“X20”表示(X后面的数字(其中X选自A、T、C和G)表示任意核苷酸的数量),或者其中与从5'末端起的第一个核酸匹配的鸟嘌呤所在的情况可以表示为“GX19”,并且其中与从5'末端起的第一个核酸不匹配的鸟嘌呤所在的情况可以用“gX19”表示。

[0115] sgRNA可以进一步包括在3'末端包含5至7个尿嘧啶(U)的终止位点。

[0116] 延伸的向导RNA可以在上述通式1的sgRNA的5'末端进一步包括1至10个核苷酸。进一步包括的核苷酸中的每一个可以独立地选自A、T、C和G。在这种情况下,另外包括的核苷酸可以具有与靶DNA序列的相应位置(延伸位置)处的核苷酸互补的序列。

[0117] 另外,sgRNA可以在5'末端进一步包括1至3个鸟嘌呤(G)。在这种情况下,另外包括的鸟嘌呤中的每一个可以独立地与靶序列的相应位置的核苷酸互补(匹配)或不互补(错配)。

[0118] 如上所述,与上述通式1的sgRNA(例如X20、GX19或gX19)相比,在5'末端进一步包括1至3个鸟嘌呤(G)和/或在crRNA或sgRNA的5'末端进一步包括1至10个核苷酸(其中每个核苷酸可以独立地选自A、T、C和G)的延伸的sgRNA可以增加碱基编辑的频率和/或效率,并在更广泛的区域中诱导碱基编辑。

[0119] 向导RNA的靶序列可以是在靶DNA上的PAM(前间区序列邻近基序)序列(在酿脓链球菌Cas9的情况下为 $5'-\text{NGG}-3'$ (N为A、T、G或C))所在的链或相对链(互补链)上位于PAM的5'末端附近的连续核酸序列。

[0120] 能够与向导RNA的靶序列杂交的向导RNA靶向序列意指与跟DNA链(即,PAM序列($5'-\text{NGG}-3'$ (其中N是A、T、G或C))所在的DNA链或相对链)互补的链的核苷酸序列具有50%或更高、60%或更高、70%或更高、80%或更高、90%或更高、95%或更高、99%或更高或100%的序列同源性并且可以与所述核苷酸序列互补地结合的核苷酸序列。

[0121] 在本说明书中,靶位点的核酸序列由靶基因的相应基因位点的两条DNA链中,PAM序列所在的链的核酸序列表示。此时,向导RNA实际结合的DNA链是与PAM序列所在的链互补的链,因此除了由于RNA的特征T变为U以外,向导RNA中包括的靶向序列具有与靶位点的序列相同的核酸序列。因此,在本说明书中,除了T和U相互互换以外,向导RNA的靶向序列和靶位点的序列(或切割位点的序列)用相同的核酸序列表示。

[0122] 向导RNA可以以RNA形式使用(或存在于组合物中),或者可以以含有编码它的DNA的质粒的形式使用(或存在于组合物中)。

[0123] 实施例

[0124] 在下文中,将参考以下实施例更加详细地描述本公开文本。然而,对于本领域技术人员而言清楚的是,以下实施例仅出于说明本公开文本的目的而提供,而不应当解释为基于本公开文本的主题而限制本公开文本的范围。

[0125] 实施例1:测试取决于sgRNA长度的碱基编辑窗口的变化

[0126] 通过深度测序来测量HEK293T细胞系中的每种活性。结果示于图2A至图2F中。

[0127] 图2A示出了取决于HEK2位点的sgRNA长度在不同碱基位置处的ABE7.10置换活性。图2B示出了与使用GX19 sgRNA的情况相比的相对置换活性[gX20~30活性/GX19活性]。图2C示出了最常观察到的突变等位基因,其中在WT序列中引入了突变的部分以红色表示。图2D示出了取决于在HBB位点的sgRNA长度在不同碱基位置处的BE3置换活性。图2E示出了与使用GX19sgRNA的情况相比的相对置换活性[gX20~30活性/GX19活性]。图2F示出了最常观察到的突变等位基因,其中在WT序列中引入了突变的部分以红色表示。已知在使用GX19 sgRNA时有效操作的碱基编辑窗口以浅蓝色表示。

[0128] 实施例2:测试当使用含有1个或2个另外的错配G的sgRNA时碱基编辑窗口的变化

[0129] 通过深度测序来测量HEK293T细胞系中的每种活性,并且结果示于图3中。

[0130] 示出了在FANCF位点(图3A)和HBB位点(图3C),取决于sgRNA长度在不同碱基位置处的BE3置换活性。示出了与当使用GX19 sgRNA时相比在FANCF位点(图3B)和HBB位点(图3D)的相对置换活性[gX20~30活性/GX19活性]。

[0131] 实施例3:测试取决于四个不同位点处的sgRNA长度的碱基编辑窗口的变化

[0132] 测试了取决于四个不同位点处的sgRNA长度的碱基编辑窗口的变化,并且结果示于图4A和图4B中。

[0133] 与使用GX19 sgRNA时相比,确定了在四个位点的ABE 7.10相对置换活性[gX20~30活性/GX19活性],并且结果示于图4A中。与使用GX19 sgRNA时相比,确定了BE3相对置换活性[gX20~30活性/GX19活性],并且结果示于图4B中。已知在使用GX19 sgRNA时有效操作的碱基编辑窗口以浅蓝色表示。通过深度测序法来测量HEK293T细胞系中的每种活性。

[0134] 实施例4:测试在油菜和大豆中取决于sgRNA类型的碱基编辑窗口的变化

[0135] 测试了在作为真核植物代表的油菜和大豆中取决于sgRNA类型的碱基编辑窗口的变化,并且结果示于图5A至图5D中。

[0136] 通过深度测序来分析每种活性。

[0137] 当将gX19 sgRNA和gX20 sgRNA与AID2胞嘧啶碱基编辑器一起在油菜原生质体中使用,测量取决于胞嘧啶位置的置换效率,并且结果示于图5A中。

[0138] 图5B示出了根据sgRNA类型引入了最频繁发生的突变的等位基因的变化。发现仅当使用gX20 sgRNA时才产生TAG终止密码子。

[0139] 当将gX19 sgRNA和gX20 sgRNA与AID2胞嘧啶碱基编辑器一起在大豆原生质体中使用,测量取决于胞嘧啶位置的置换效率,并且结果示于图5C中。图5D示出了根据sgRNA类型引入了最频繁发生的突变的等位基因的变化。类似地,发现仅当使用gX20 sgRNA时才产生TAG终止密码子。

[0140] 实施例5:测试小鼠中取决于sgRNA类型的碱基编辑窗口的变化

[0141] 测试了在作为真核动物代表的小鼠中取决于sgRNA类型的碱基编辑窗口的变化,并且结果示于图6A和图6B中。

[0142] 将ABE 7.10mRNA和不同类型的sgRNA显微注射到小鼠胚胎中之后,在胚泡期通过深度测序分析置换活性,并且结果示于图6A中。

[0143] 作为分析通过使用ABE 7.10mRNA与GX21 sgRNA组合对胚胎进行显微注射而获得的幼崽的结果,如图6B所示,获得了三个具有所需H420R突变的幼崽。

[0144] **【实用性】**

[0145] 根据本公开文本,通过使用与常规向导RNA相比延伸的向导RNA进行使用脱氨酶的基因碱基编辑,可以提高碱基编辑的频率和/或效率,并且可以使用此技术有效地诱导所需的点突变。

[0146] 尽管已详细描述本公开文本的具体配置,但本领域技术人员将理解,优选实施方案在说明书中出于说明性目的给出,而不应当解释为限制本公开文本的范围。因此,本公开文本的实质范围是由随附权利要求和其等效内容来限定。

[0147] **【序列表独立文本】**

[0148] 附有电子文件。

<110>	基础科学研究院	
<120>	延伸的单向导RNA及其用途	
<130>	PP-B2171	
<150>	KR 10-2018-0008492	
<151>	2018-01-23	
<160>	2	
<170>	KopatentIn 2.0	
<210>	1	
<211>	60	
<212>	RNA	
<213>	人工序列	
<220>		
<223>	向导RNA	
<400>	1	
	uagcaaguua aaauaaggcu aguccguuau caacuugaaa aaguggcacc gagucggugc	60
		60
<210>	2	
<211>	21	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	NLS序列	
<400>	2	
	cccaagaaga agaggaaagt c	21

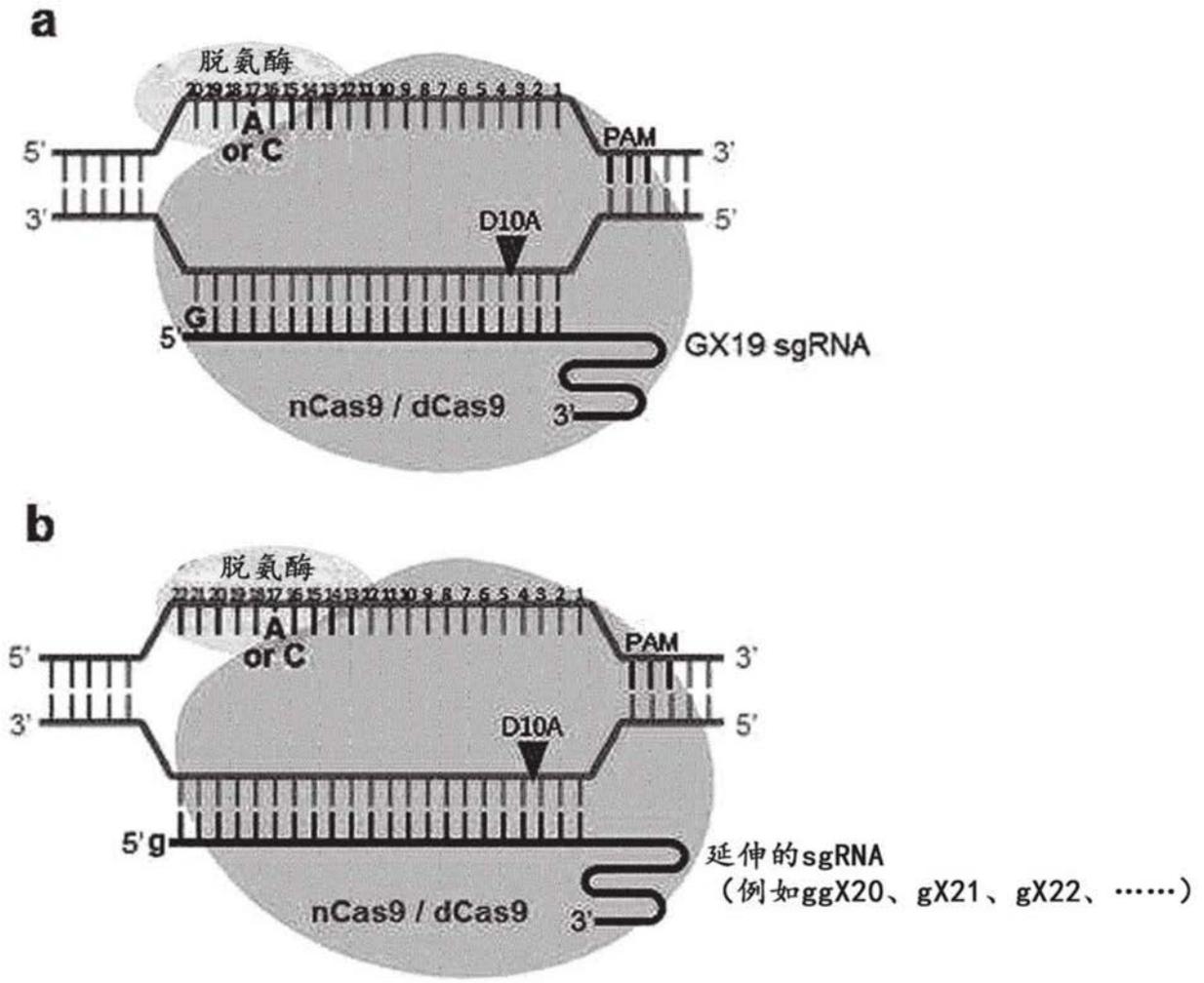


图1

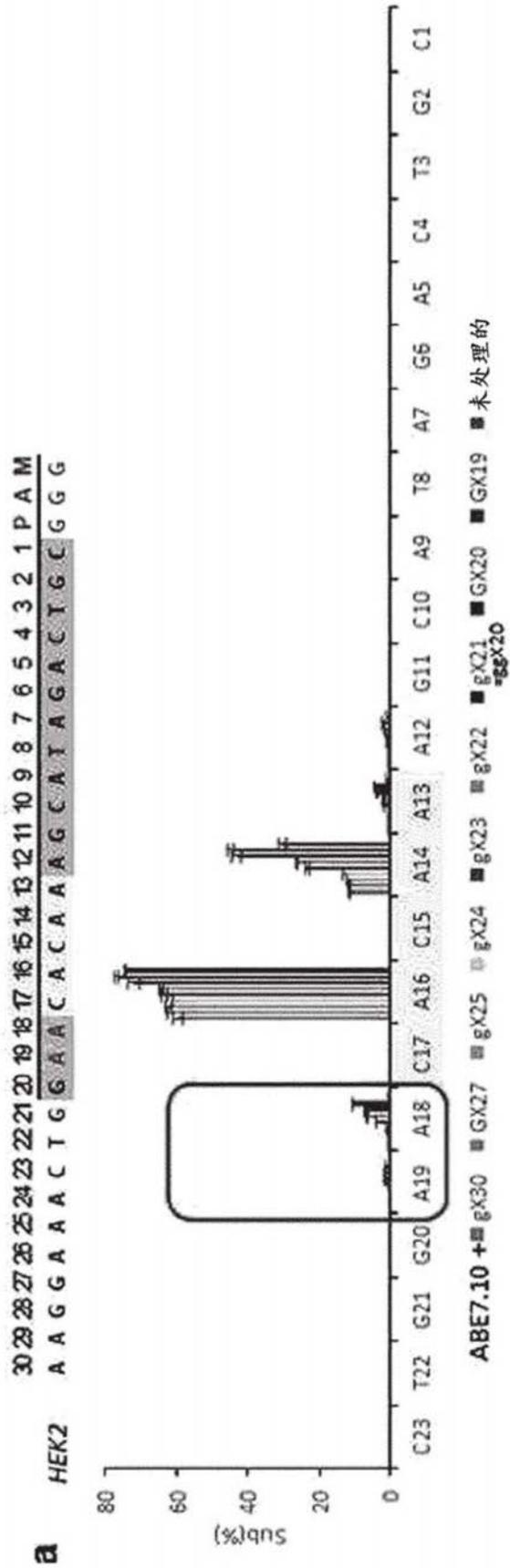


图2a

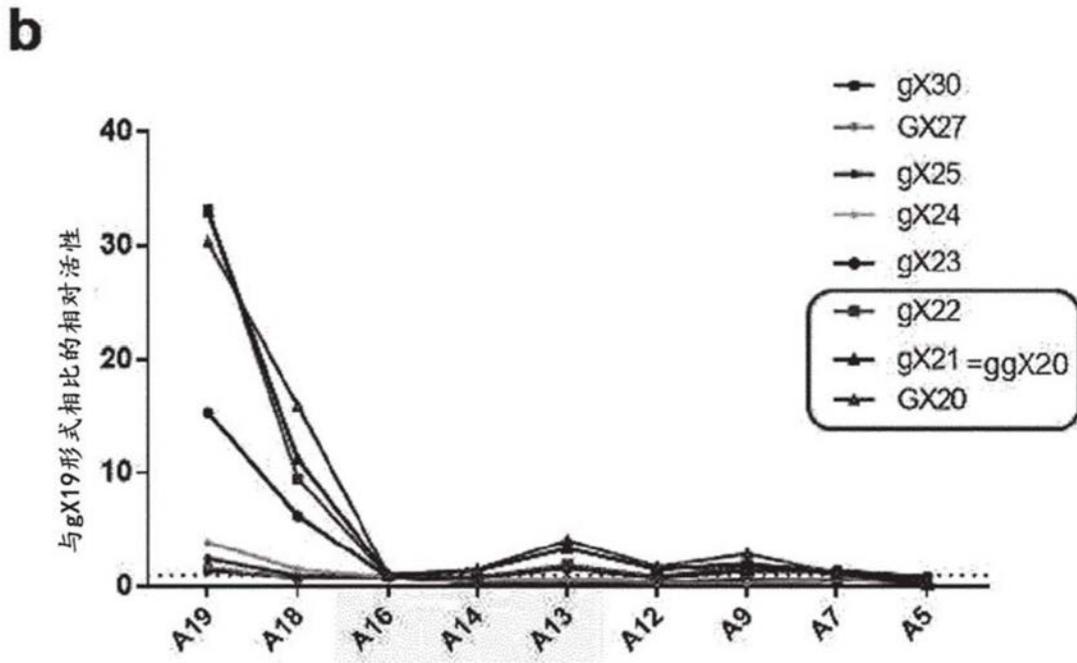


图2b



图2c

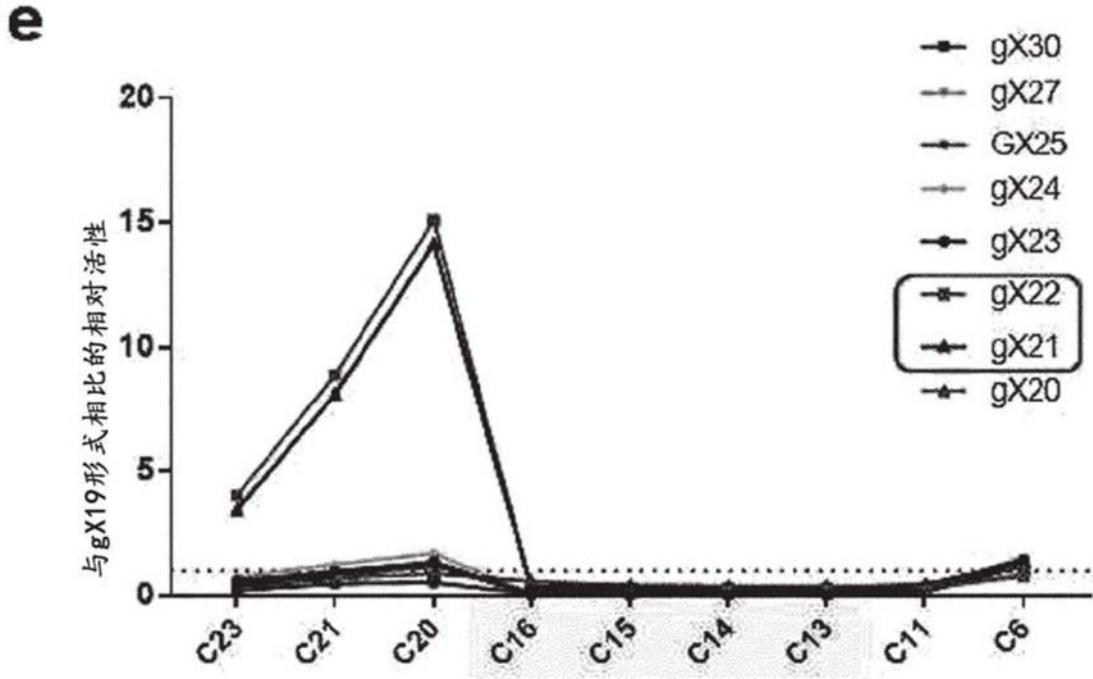


图2e

f

```

野生型 CCACGTTACCTTGCCCCACAGGGCAGTAACGG
|
gX19  CCACGTTACCTTGTTTTACAGGGCAGTAACGG
|
      CCACGTTACCTTGCTTTACAGGGCAGTAACGG
|
      CCACGTTACCTTGCCTTACAGGGCAGTAACGG
      *****
|
gX22  CCACGTTACCTTGTTTTACAGGGCAGTAACGG
|
      CCACGTTACCTTGCTTTACAGGGCAGTAACGG
|
      CCACGTTTATTTTGTTTTACAGGGCAGTAACGG
      ***** *   *** *****
    
```

图2f

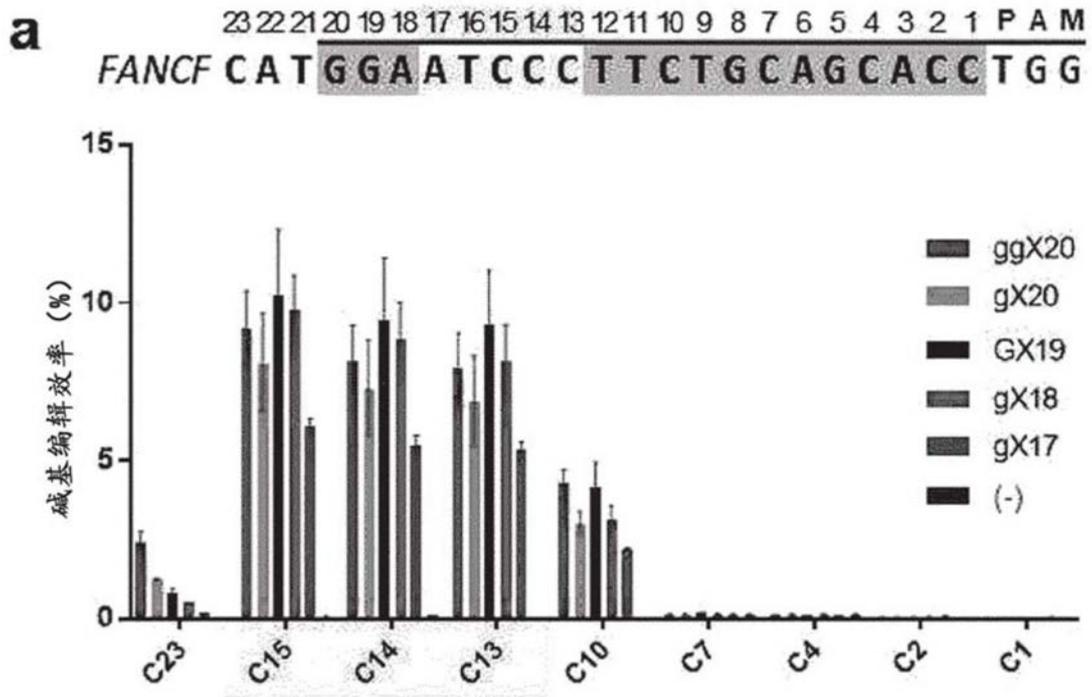


图3a

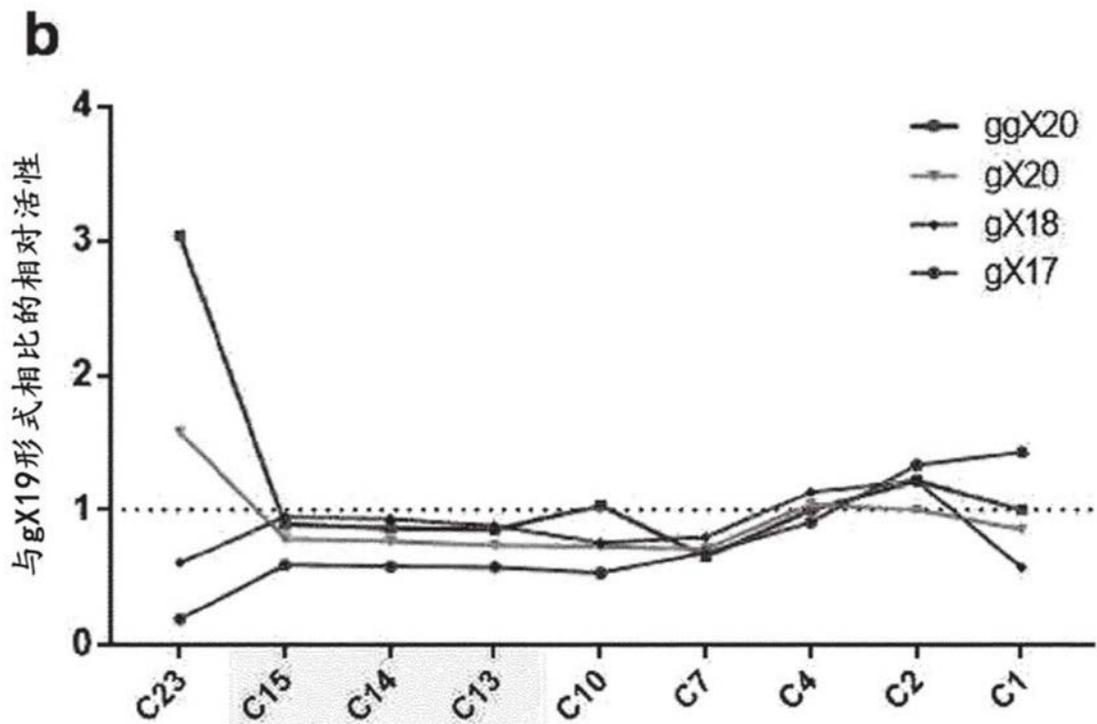


图3b

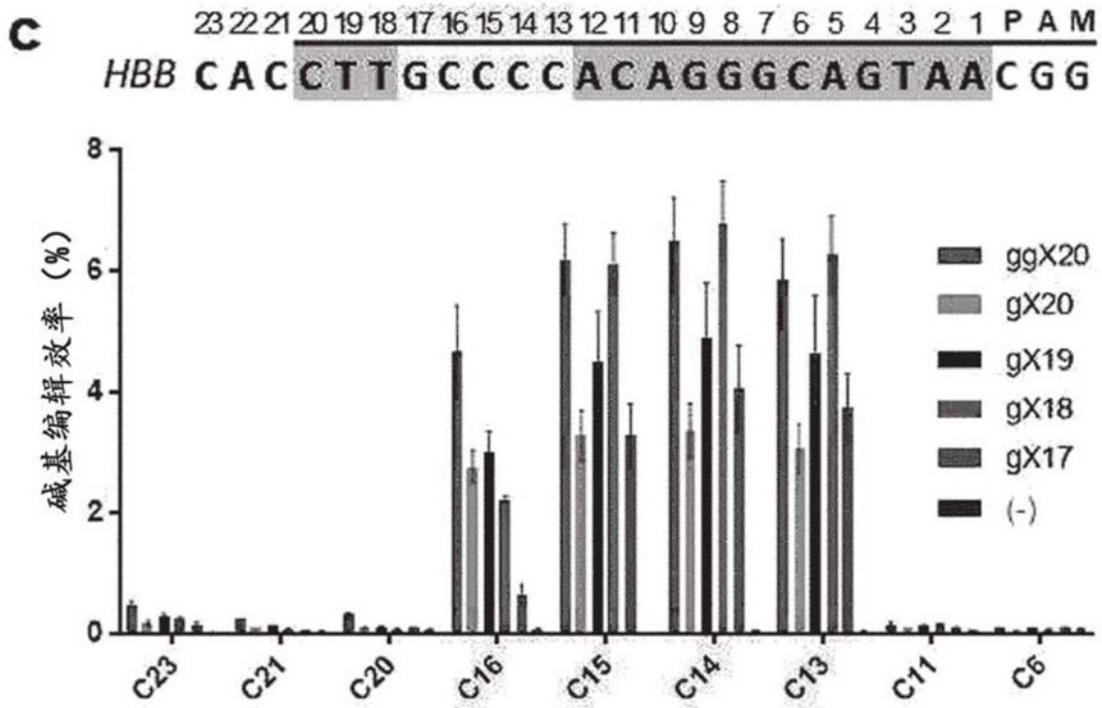


图3c

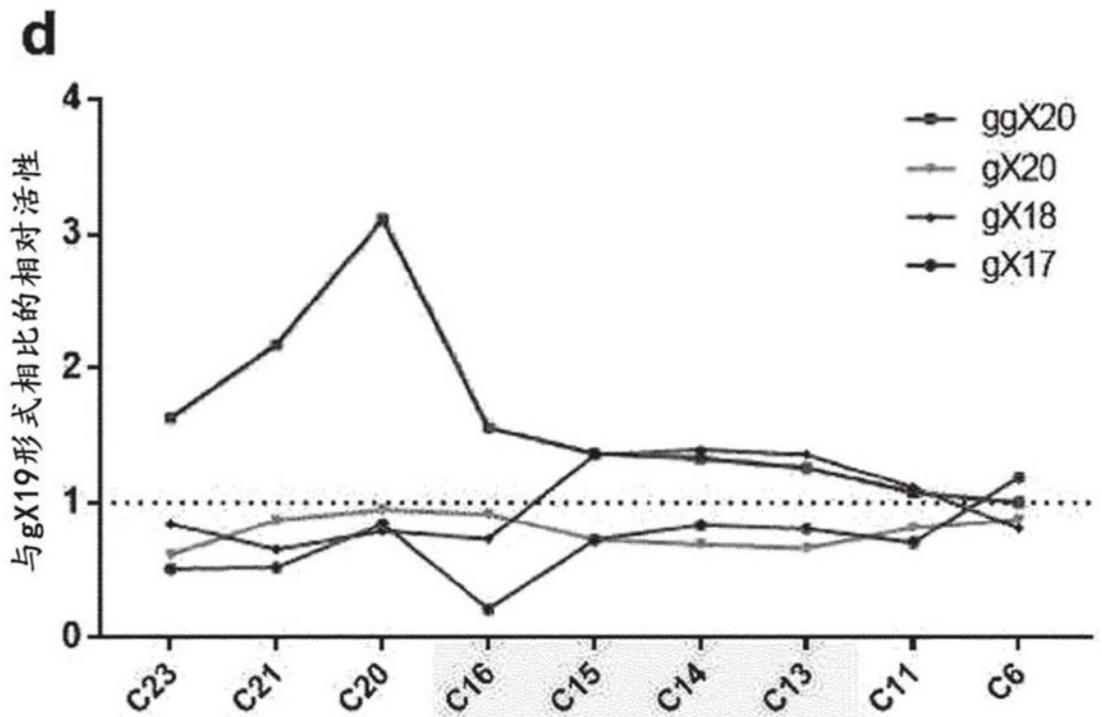


图3d

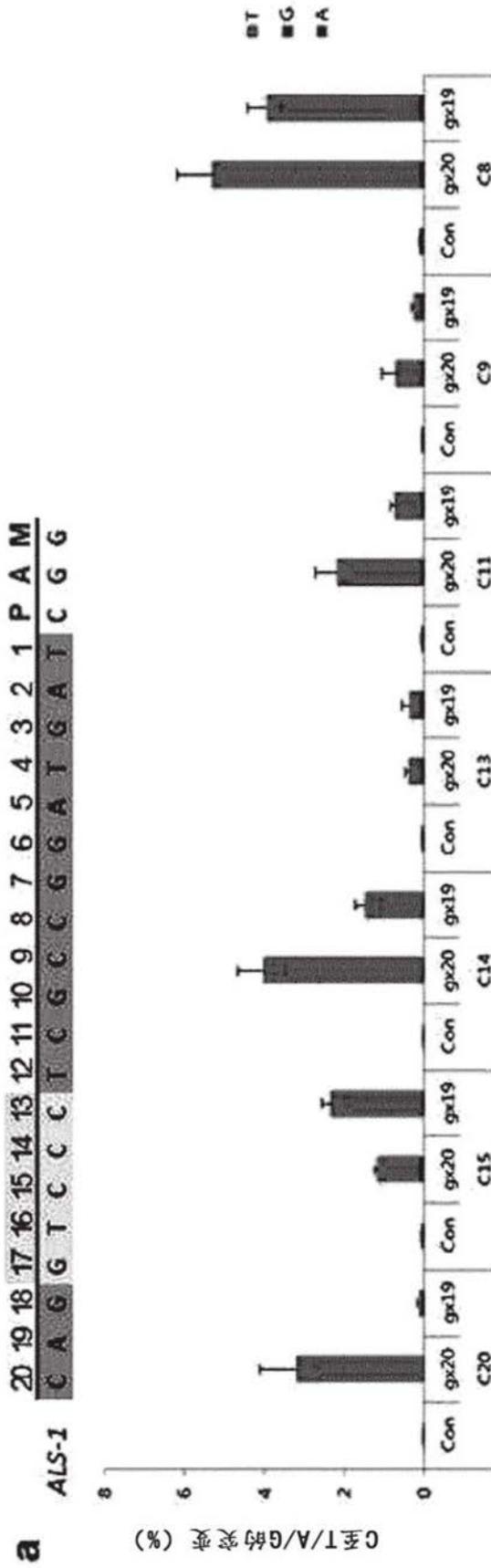


图5a

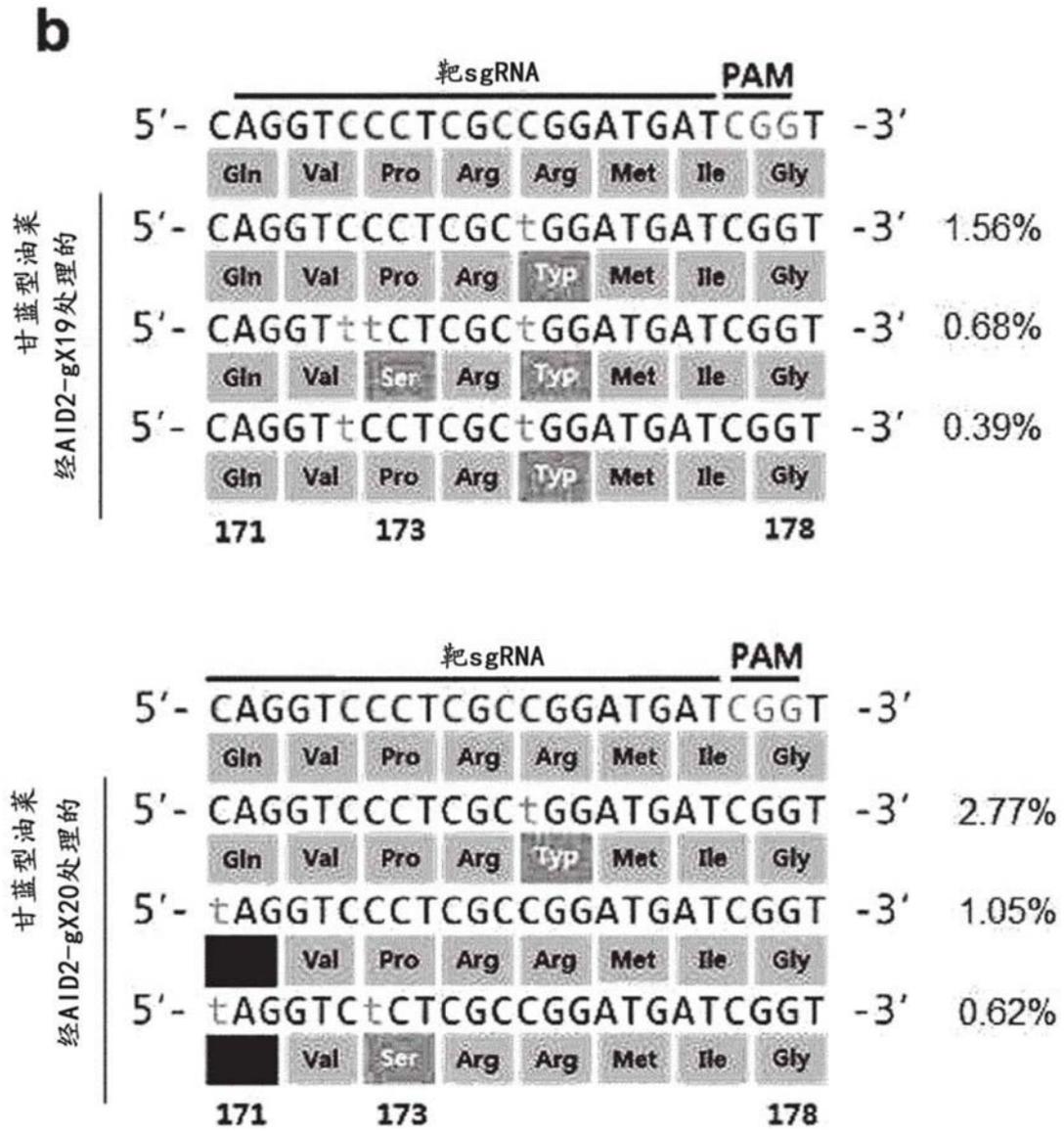


图5b

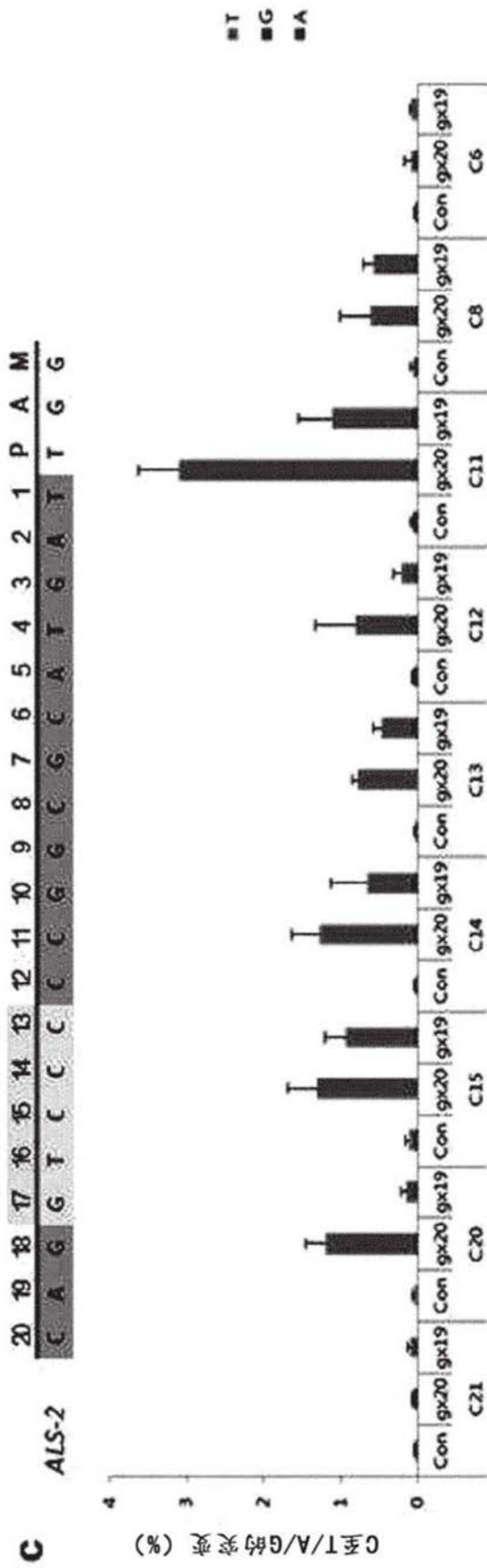


图5c

d

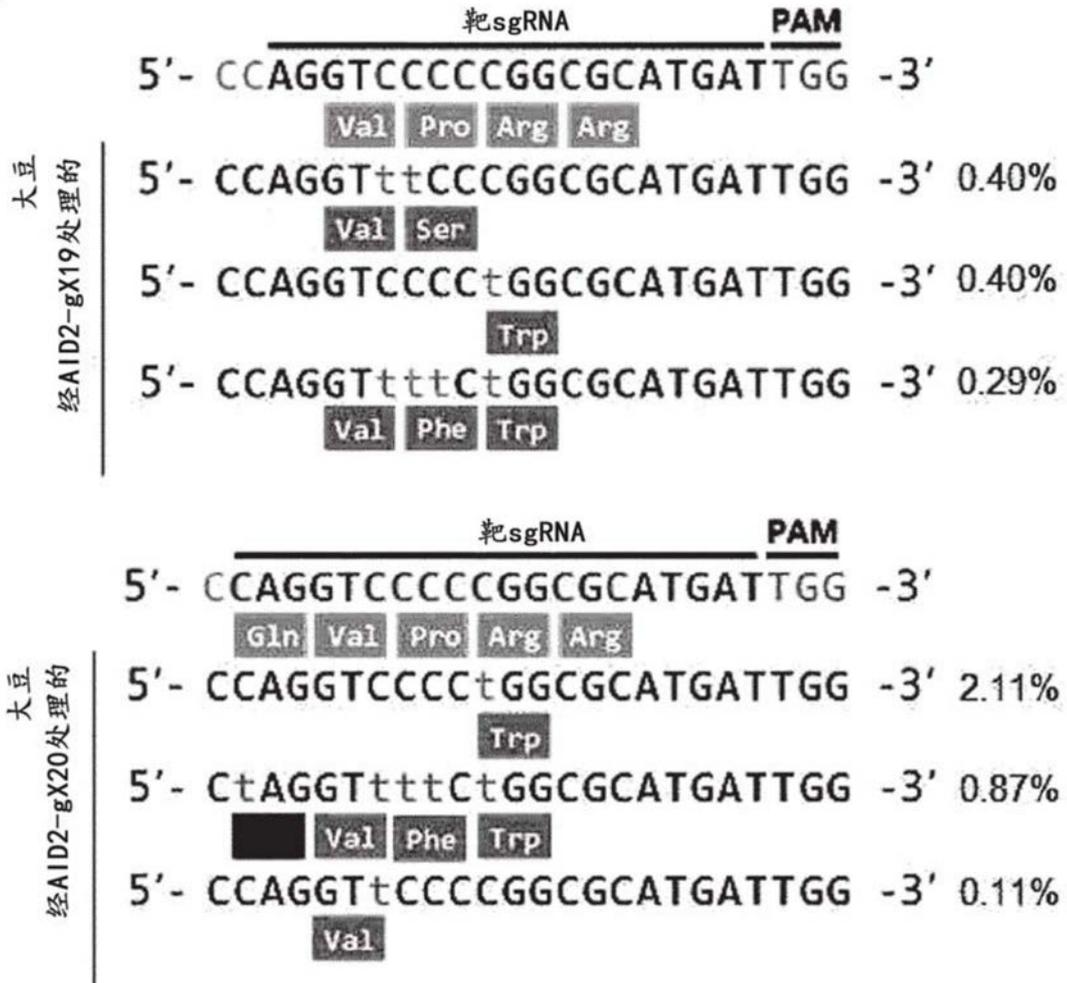


图5d

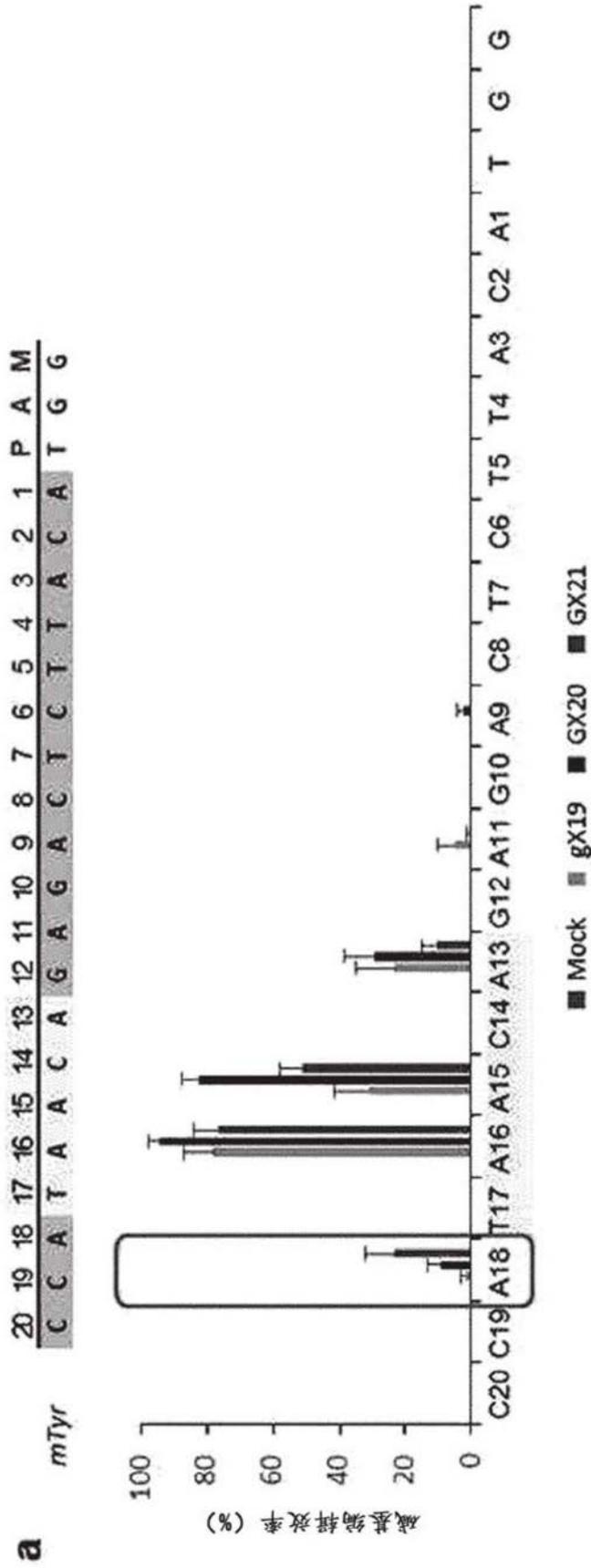


图6a

b	420 421 422 423 424 425 426	
	His Asn Arg Asp Ser Tyr Met	置换 (%)
幼崽#1	CCATAACAGAGACTCTTACATGG	
	CCATGGCAGAGACTCTTACATGG	45.2 (N421G)
	CCGTGACAGAGACTCTTACATGG	25.5 (H420R, N421D)
幼崽#2	CCATGACAGAGACTCTTACATGG	22.1 (N421D)
	CCATGGCAGAGACTCTTACATGG	36.4 (N421D)
	CCGTGGCAGAGACTCTTACATGG	35.9 (N421G)
幼崽#3	CCATGGCAGAGACTCTTACATGG	11.4 (H420R, N421G)
	CCATGGCGGAGACTCTTACATGG	40.0 (N421G, R422G)
	CCGTGACAGAGACTCTTACATGG	27.7 (H420R, N421D)
	CCGTGGCAGAGACTCTTACATGG	19.4 (H420R, N421G)
	CCATGGCAGAGACTCTTACATGG	4.8 (N421G)

图6b