(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 117999351 A
(43)申请公布日 2024.05.07

(21)申请号 202280060888.0
(22)申请日 2022.09.06
(30)优先权数据
63/241,928 2021.09.08 US
(85)PCT国际申请进入国家阶段日
2024.03.08
(86)PCT国际申请的申请数据
PCT/US2022/075992 2022.09.06
(87)PCT国际申请的公布数据
WO2023/039378 EN 2023.03.16
(71)申请人 宏基因组学公司
地址 美国加利福尼亚州
(72)发明人 布莱恩·C·托马斯
克利斯多佛·布朗
辛迪·卡斯泰勒 利萨·亚历山大
利利安娜·冈萨雷斯-奥索里奥
保拉·马瑟斯卡尔内瓦利
多姆·卡斯坦佐
(74)专利代理机构 北京安信方达知识产权代理有限公司 11262
专利代理师 徐爱文 武晶晶
(51)Int.Cl.
C12N 15/113(2006.01)
C12N 9/12(2006.01)

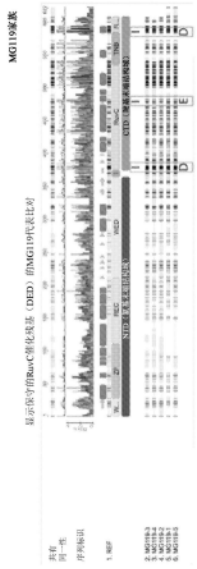
权利要求书10页 说明书62页
序列表（电子公布） 附图31页

(54)发明名称

II类V型CRISPR系统

(57)摘要

本文描述了可用于基因编辑的源自未经培养的微生物的方法、组合物和系统，所述方法、组合物和系统涉及新颖II类V型CRISPR相关核酸内切酶。



CN 117999351 A

1.一种经工程化的核酸酶系统,其包括:

(a)核酸内切酶,所述核酸内切酶与SEQ ID NO:1-325、420-431、476-624或629中的任一者或其变体具有至少75%序列同一性;以及

(b)经工程化的向导RNA,其中所述经工程化的向导RNA被配置成与所述核酸内切酶形成复合物,并且所述经工程化的向导RNA包括间隔子序列,所述间隔子序列被配置成与靶核酸序列杂交。

2.根据权利要求1所述的经工程化的核酸酶系统,其中所述向导RNA包括与SEQ ID NO:410-419、432、434、436、438、440、442、444、446、448、450、452、454、456、458、460、462、464、466、468、470、472和474中的任一者的非简并核苷酸具有至少80%序列同一性的序列。

3.根据权利要求1至2中任一项所述的经工程化的核酸酶系统,其中所述核酸内切酶与SEQ ID NO:30-33、39、48、56、57、61、83、92、100、110、124、136、145、148、424、425、429、476或629中的任一者具有至少约80%、至少约85%、至少约86%、至少约87%、至少约88%、至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%或100%序列同一性。

4.根据权利要求1至3中任一项所述的经工程化的核酸酶系统,其中所述向导RNA包括与SEQ ID NO:414-419、432、434、436、438、440、442、444、446、448、450、452、454、456、458、460、462、464、466、468、470、472和474中的任一者的非简并核苷酸具有至少约85%、至少约86%、至少约87%、至少约88%、至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%或100%序列同一性的序列。

5.一种经工程化的核酸酶系统,其包括:

(a)经工程化的向导RNA,所述经工程化的向导RNA包括与SEQ IDNO:410-419、432、434、436、438、440、442、444、446、448、450、452、454、456、458、460、462、464、466、468、470、472和474中的任一者的非简并核苷酸具有至少80%序列同一性的序列;以及

(b)2类V型Cas核酸内切酶,所述2类V型Cas核酸内切酶被配置成与所述经工程化的向导RNA结合。

6.根据权利要求1至5中任一项所述的经工程化的核酸酶系统,其中所述向导RNA包括与真核、真菌、植物、哺乳动物或人基因组多核苷酸序列互补的序列。

7.根据权利要求1至6中任一项所述的经工程化的核酸酶系统,其中所述向导RNA的长度为30-250个核苷酸。

8.根据权利要求1至7中任一项所述的经工程化的核酸酶系统,其中所述核酸内切酶包括接近所述核酸内切酶的N末端或C末端的一个或多个核定位序列(NLS)。

9.根据权利要求1至8中任一项所述的经工程化的核酸酶系统,其中所述NLS包括与选自由SEQ ID NO:630-645组成的组的序列至少80%相同的序列。

10.根据权利要求1至9中任一项所述的经工程化的核酸酶系统,其进一步包括单链或双链DNA修复模板,所述单链或双链DNA修复模板从5'至3'包括:第一同源臂,所述第一同源臂包括位于所述靶脱氧核糖核酸序列的5'的至少20个核苷酸的序列;至少10个核苷酸的合成DNA序列;以及第二同源臂,所述第二同源臂包括位于所述靶序列的3'的至少20个核苷酸的序列。

11.根据权利要求10所述的经工程化的核酸酶系统,其中所述第一同源臂或所述第二同源臂包括至少40、80、120、150、200、300、500或1,000个核苷酸的序列。

12.根据权利要求10或权利要求11所述的经工程化的核酸酶系统,其中所述第一同源臂和所述第二同源臂与原核生物、细菌、真菌或真核生物的基因组序列同源。

13.根据权利要求10至12所述的经工程化的核酸酶系统,其中所述单链或双链DNA修复模板包括转基因供体。

14.根据权利要求1至13中任一项所述的经工程化的核酸酶系统,其进一步包括DNA修复模板,所述DNA修复模板包括侧接有一个或两个单链DNA区段的双链DNA区段。

15.根据权利要求14所述的经工程化的核酸酶系统,其中所述单链DNA区段与所述双链DNA区段的5'端缀合。

16.根据权利要求14所述的经工程化的核酸酶系统,其中所述单链DNA区段与所述双链DNA区段的3'端缀合。

17.根据权利要求14至16中任一项所述的经工程化的核酸酶系统,其中所述单链DNA区段的长度为4至10个核苷酸碱基。

18.根据权利要求14至17中任一项所述的经工程化的核酸酶系统,其中所述单链DNA区段具有与所述间隔子序列内的序列互补的核苷酸序列。

19.根据权利要求14至18中任一项所述的经工程化的核酸酶系统,其中所述双链DNA序列包括条形码、开放阅读框、增强子、启动子、蛋白质编码序列、miRNA编码序列、RNA编码序列或转基因。

20.根据权利要求14至18中任一项所述的经工程化的核酸酶系统,其中所述双链DNA序列侧接有核酸酶切割位点。

21.根据权利要求20所述的经工程化的核酸酶系统,其中所述核酸酶切割位点包括间隔子和PAM序列。

22.根据权利要求21所述的经工程化的核酸酶系统,其中所述PAM包括SEQ ID NO:433、435、437、439、441、443、445、447、449、451、453、455、457、459、461、463、465、467、469、471、473和475中的任一者的序列。

23.根据权利要求1至22中任一项所述的经工程化的核酸酶系统,其中所述系统进一步包括Mg^{2+}的来源。

24.根据权利要求1至23中任一项所述的经工程化的核酸酶系统,其中所述向导RNA包括发夹,所述发夹包括至少8个、至少10个或至少12个碱基配对的核糖核苷酸。

25.根据权利要求24所述的经工程化的核酸酶系统,其中所述发夹包括10个碱基配对的核糖核苷酸。

26.根据权利要求1至25中任一项所述的经工程化的核酸酶系统,其中:

a)所述核酸内切酶包括与SEQ ID NO:1、6、15、30、151、292或319中的任一者至少75%、80%或90%相同的序列或其变体;并且

b)所述向导RNA结构包括与SEQ ID NO:410-419中的任一者的非简并核苷酸至少80%或90%相同的序列。

27.根据权利要求1至25中任一项所述的经工程化的核酸酶系统,其中:

a)所述核酸内切酶包括与SEQ ID NO:30-33、39、48、56、57、61、83、92、100、110、124、

136、145、148、424、425、429、476或629中的任一者至少约75％、至少约80％、至少约85％、至少至少约80％、至少约85％、至少约86％、至少约87％、至少约88％、至少约89％、至少约90％、至少约91％、至少约92％、至少约93％、至少约94％、至少约95％、至少约96％、至少约97％、至少约98％、至少约99％或100％序列同一性的序列;并且

b)所述向导RNA结构包括与SEQ ID NO:414-419、432、434、436、438、440、442、444、446、448、450、452、454、456、458、460、462、464、466、468、470、472和474中的任一者的非简并核苷酸至少约80％、至少约85％、至少约86％、至少约87％、至少约88％、至少约89％、至少约90％、至少约91％、至少约92％、至少约93％、至少约94％、至少约95％、至少约96％、至少约97％、至少约98％、至少约99％或100％序列同一性的序列。

28.根据权利要求1至27中任一项所述的经工程化的核酸酶系统,其中所述序列同一性是通过BLASTP、CLUSTALW、MUSCLE、MAFFT算法或CLUSTALW算法使用史密斯-沃特曼同源性搜索算法参数(Smith-Waterman homology search algorithm parameter)来确定的。

29.根据权利要求28所述的经工程化的核酸酶系统,其中所述序列同一性是通过所述BLASTP同源性搜索算法使用字长(W)为3、期望值(E)为10的参数以及BLOSUM62评分矩阵将空位罚分设置为存在11,扩展1并且使用条件组成评分矩阵调整来确定的。

30.一种经工程化的向导核糖核酸(RNA)多核苷酸,其包括:

a)DNA靶向区段,所述DNA靶向区段包括与靶DNA分子中的靶序列互补的核苷酸序列;以及

b)蛋白质结合区段,所述蛋白质结合区段包括杂交以形成双链RNA(dsRNA)双链体的两个互补的核苷酸延伸段,

其中所述两个互补的核苷酸延伸段用中间核苷酸彼此共价连接,并且

其中所述经工程化的向导核糖核酸多核苷酸能够与2型V类Cas核酸内切酶形成复合物。

31.根据权利要求30所述的经工程化的向导RNA,其中所述2型V类Cas核酸内切酶源自未经培养的生物体。

32.根据权利要求30或权利要求31所述的经工程化的向导核糖核酸多核苷酸,其中所述Cas核酸内切酶与SEQ ID NO:1-325、420-431、476-624或629中的任一者具有至少75％序列同一性,并且将所述复合物靶向到所述靶DNA分子的所述靶序列。

33.根据权利要求30至32中任一项所述的经工程化的向导核糖核酸多核苷酸,其中所述DNA靶向区段定位在所述两个互补的核苷酸延伸段中的两者的3'。

34.根据权利要求30至33中任一项所述的经工程化的向导核糖核酸多核苷酸,其中所述蛋白质结合区段包括与SEQ ID NO:410-419的非简并核苷酸具有至少70％、至少80％或至少90％同一性的序列。

35.根据权利要求30至34中任一项所述的经工程化的向导核糖核酸多核苷酸,其中所述双链RNA(dsRNA)双链体包括至少5个、至少8个、至少10个或至少12个核糖核苷酸。

36.一种脱氧核糖核酸多核苷酸,其编码根据权利要求30至35中任一项所述的经工程化的向导核糖核酸多核苷酸。

37.一种核酸,其包括经工程化的核酸序列,所述经工程化的核酸序列被优化以在生物体中表达,其中所述核酸编码2类V型Cas核酸内切酶,并且其中所述核酸内切酶源自未经培

养的微生物,其中所述生物体不是所述未经培养的生物体。

38.根据权利要求37所述的核酸,其中所述核酸内切酶包括与SEQ ID NO:1-325、420-431、476-624或629中的任一者具有至少70%或至少80%序列同一性的变体。

39.根据权利要求37或38所述的核酸,其中所述核酸内切酶包括编码接近所述核酸内切酶的N末端或C末端的一个或多个核定位序列(NLS)的序列。

40.根据权利要求39所述的核酸,其中所述NLS包括选自SEQ ID NO:630-645的序列。

41.根据权利要求39或40所述的核酸,其中所述NLS包括SEQ ID NO:631。

42.根据权利要求41所述的核酸,其中所述NLS接近所述核酸内切酶的所述N末端。

43.根据权利要求39或40所述的核酸,其中所述NLS包括SEQ ID NO:630。

44.根据权利要求43所述的核酸,其中所述NLS接近所述核酸内切酶的所述C末端。

45.根据权利要求37至44中任一项所述的核酸,其中所述生物体是原核生物、细菌、真核生物、真菌、植物、哺乳动物、啮齿动物或人。

46.一种经工程化的载体,其包括核酸序列,所述核酸序列编码2类V型Cas核酸内切酶,其中所述核酸内切酶源自未经培养的微生物。

47.一种经工程化的载体,其包括根据权利要求37至45中任一项所述的核酸。

48.一种经工程化的载体,其包括根据权利要求36所述的脱氧核糖核酸多核苷酸。

49.根据权利要求46至48中任一项所述的经工程化的载体,其中所述载体是质粒、微环、CELiD、腺相关病毒(AAV)源性病毒体、慢病毒或腺病毒。

50.一种细胞,其包括根据权利要求46至49中任一项所述的经工程化的载体。

51.一种制备核酸内切酶的方法,所述方法包括培养根据权利要求50所述的所述细胞。

52.一种用于结合、切割、标记或修饰双链脱氧核糖核酸多核苷酸的方法,所述方法包括:

(a)使所述双链脱氧核糖核酸多核苷酸与2类V型Cas核酸内切酶接触,所述2类V型Cas核酸内切酶与经工程化的向导RNA复合,所述经工程化的向导RNA被配置成与所述核酸内切酶和所述双链脱氧核糖核酸多核苷酸结合;

其中所述双链脱氧核糖核酸多核苷酸包括原间隔子相邻基序(PAM);并且

其中所述向导RNA结构包括与SEQ ID NO:410-419中的任一者的非简并核苷酸至少80%或90%相同的序列。

53.根据权利要求52所述的方法,其中所述双链脱氧核糖核酸多核苷酸包括第一链和第二链,所述第一链包括与所述经工程化的向导RNA的序列互补的序列,所述第二链包括所述PAM。

54.根据权利要求53所述的方法,其中所述PAM与和所述经工程化的向导RNA的所述序列互补的所述序列的5'端直接相邻。

55.根据权利要求52至54中任一项所述的方法,其中所述PAM包括SEQ ID NO:433、435、437、439、441、443、445、447、449、451、453、455、457、459、461、463、465、467、469、471、473和475中的任一者的序列。

56.根据权利要求52至55中任一项所述的方法,其中所述2类V型Cas核酸内切酶源自未经培养的微生物。

57.根据权利要求52至56中任一项所述的方法,其中所述2类V型Cas核酸内切酶进一步

包括PAM相互作用结构域。

58.根据权利要求52至57中任一项所述的方法,其中所述双链脱氧核糖核酸多核苷酸是真核生物、植物、真菌、哺乳动物、啮齿动物或人双链脱氧核糖核酸多核苷酸。

59.一种修饰靶核酸基因座的方法,所述方法包括向所述靶核酸基因座递送根据权利要求1至29中任一项所述的所述经工程化的核酸酶系统,其中所述核酸内切酶被配置成与所述经工程化的向导核糖核酸结构形成复合物,并且其中所述复合物被配置成使得在所述复合物与所述靶核酸基因座结合时,所述复合物修饰所述靶核酸基因座。

60.根据权利要求59所述的方法,其中修饰所述靶核酸基因座包括结合、切口、切割或标记所述靶核酸基因座。

61.根据权利要求59或60所述的方法,其中所述靶核酸基因座包括脱氧核糖核酸(DNA)或核糖核酸(RNA)。

62.根据权利要求59所述的方法,其中所述靶核酸包括基因组DNA、病毒DNA、病毒RNA或细菌DNA。

63.根据权利要求59至62中任一项所述的方法,其中所述靶核酸基因座是在体外的。

64.根据权利要求59至62中任一项所述的方法,其中所述靶核酸基因座是在细胞内的。

65.根据权利要求64所述的方法,其中所述细胞是原核细胞、细菌细胞、真核细胞、真菌细胞、植物细胞、动物细胞、哺乳动物细胞、啮齿动物细胞、灵长类动物细胞、人细胞或原代细胞。

66.根据权利要求64或65所述的方法,其中所述细胞是原代细胞。

67.根据权利要求66所述的方法,其中所述原代细胞是T细胞。

68.根据权利要求66所述的方法,其中所述原代细胞是造血干细胞(HSC)。

69.根据权利要求59至68中任一项所述的方法,其中向所述靶核酸基因座递送所述经工程化的核酸酶系统包括递送根据权利要求37至45中任一项所述的核酸或根据权利要求46至49中任一项所述的经工程化的载体。

70.根据权利要求59至69中任一项所述的方法,其中向所述靶核酸基因座递送所述经工程化的核酸酶系统包括递送包括编码所述核酸内切酶的开放阅读框的核酸。

71.根据权利要求70所述的方法,其中所述核酸包括启动子,编码所述核酸内切酶的所述开放阅读框与所述启动子可操作地连接。

72.根据权利要求59至71中任一项所述的方法,其中向所述靶核酸基因座递送所述经工程化的核酸酶系统包括递送包含编码所述核酸内切酶的所述开放阅读框的加帽mRNA。

73.根据权利要求59至72中任一项所述的方法,其中向所述靶核酸基因座递送所述经工程化的核酸酶系统包括递送经翻译的多肽。

74.根据权利要求59至72中任一项所述的方法,其中向所述靶核酸基因座递送所述经工程化的核酸酶系统包括递送编码与核糖核酸(RNA)pol III启动子可操作地连接的所述经工程化的向导RNA的脱氧核糖核酸(DNA)。

75.根据权利要求59至74中任一项所述的方法,其中所述核酸内切酶诱导所述靶基因座处或接近所述靶基因座的单链断裂或双链断裂。

76.根据权利要求75所述的方法,其中所述核酸内切酶诱导所述靶基因座内或位于所述靶基因座的3'的交错单链断裂。

77.一种宿主细胞,其包括编码异源核酸内切酶的开放阅读框,所述异源核酸内切酶与SEQ ID NO:1-325、420-431、476-624或629中的任一者或其变体具有至少75%序列同一性。

78.根据权利要求77所述的宿主细胞,其中所述核酸内切酶与SEQ ID NO:1、6、15、30、151、292或319中的任一者或其变体具有至少75%序列同一性。

79.根据权利要求77所述的宿主细胞,其中所述核酸内切酶与SEQ ID NO:30-33、39、48、56、57、61、83、92、100、110、124、136、145、148、424、425、429、476或629中的任一者具有至少约75%、至少约80%、至少约85%、至少约86%、至少约87%、至少约88%、至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%或100%序列同一性。

80.根据权利要求77至79中任一项所述的宿主细胞,其中所述宿主细胞是大肠杆菌(E.coli)细胞。

81.根据权利要求80所述的宿主细胞,其中所述大肠杆菌细胞是λDE3溶原菌,或者所述大肠杆菌细胞是BL21(DE3)菌株。

82.根据权利要求80或81所述的宿主细胞,其中所述大肠杆菌细胞具有ompT lon基因型。

83.根据权利要求77至82中任一项所述的宿主细胞,其中所述开放阅读框与以下可操作地连接:T7启动子序列、T7-lac启动子序列、lac启动子序列、tac启动子序列、trc启动子序列、ParaBAD启动子序列、PrhaBAD启动子序列、T5启动子序列、cspA启动子序列、$araP_{BAD}$启动子、来自噬菌体λ的强左侧启动子(pL启动子)或其任何组合。

84.根据权利要求77至83中任一项所述的宿主细胞,其中所述开放阅读框包括编码亲和标签的序列,所述亲和标签与编码所述核酸内切酶的序列框内连接。

85.根据权利要求84所述的宿主细胞,其中所述亲和标签是固定化金属亲和色谱法(IMAC)标签。

86.根据权利要求85所述的宿主细胞,其中所述IMAC标签是多组氨酸标签。

87.根据权利要求84所述的宿主细胞,其中所述亲和标签是myc标签、人流感血凝素(HA)标签、麦芽糖结合蛋白(MBP)标签、谷胱甘肽S-转移酶(GST)标签、链霉亲和素标签、FLAG标签或其任何组合。

88.根据权利要求84至87中任一项所述的宿主细胞,其中所述亲和标签经由编码蛋白酶切割位点的接头序列与编码所述核酸内切酶的所述序列框内连接。

89.根据权利要求88所述的宿主细胞,其中所述蛋白酶切割位点是烟草蚀纹病毒(TEV)蛋白酶切割位点、PreScission®蛋白酶(PSP)切割位点、凝血酶切割位点、因子Xa切割位点、肠激酶切割位点或其任何组合。

90.根据权利要求77至89中任一项所述的宿主细胞,其中所述开放阅读框被密码子优化以在所述宿主细胞中表达。

91.根据权利要求77至90中任一项所述的宿主细胞,其中所述开放阅读框被提供在载体上。

92.根据权利要求77至90中任一项所述的宿主细胞,其中所述开放阅读框被整合到所述宿主细胞的基因组中。

93.一种培养物,其包括在相容液体培养基中的根据权利要求77至92中任一项所述的

宿主细胞。

94.一种产生核酸内切酶的方法,所述方法包括在相容液体培养基中培养根据权利要求77至92中任一项所述的宿主细胞。

95.根据权利要求94所述的方法,其进一步包括通过添加另外的化学药剂或增加量的营养物来诱导所述核酸内切酶的表达。

96.根据权利要求95所述的方法,其中另外的化学药剂或增加量的营养物包括异丙基β-D-1-硫代半乳糖苷(IPTG)或另外的量的乳糖。

97.根据权利要求94至96中任一项所述的方法,其进一步包括在所述培养后分离所述宿主细胞并且使所述宿主细胞裂解以产生蛋白质提取物。

98.根据权利要求97所述的方法,其进一步包括使所述蛋白质提取物经受IMAC或离子亲和色谱法。

99.根据权利要求98所述的方法,其中所述开放阅读框包括编码IMAC亲和标签的序列,所述IMAC亲和标签与编码所述核酸内切酶的序列框内连接。

100.根据权利要求99所述的方法,其中所述IMAC亲和标签经由编码蛋白酶切割位点的接头序列与编码所述核酸内切酶的所述序列框内连接。

101.根据权利要求100所述的方法,其中所述蛋白酶切割位点包括烟草蚀纹病毒(TEV)蛋白酶切割位点、PreScission®蛋白酶切割位点、凝血酶切割位点、因子Xa切割位点、肠激酶切割位点或其任何组合。

102.根据权利要求100至101中任一项所述的方法,其进一步包括通过使对应于所述蛋白酶切割位点的蛋白酶与所述核酸内切酶接触来切割所述IMAC亲和标签。

103.根据权利要求102所述的方法,其进一步包括执行减材IMAC亲和色谱法以从包括所述核酸内切酶的组合物中去除所述亲和标签。

104.一种破坏细胞中的基因座的方法,所述方法包括使包括以下的组合物与所述细胞接触:

(a)2类V型Cas核酸内切酶,所述2类V型Cas核酸内切酶与SEQ ID NO:1-325、420-431、476-624或629中的任一者或其变体具有至少75%同一性;以及

(b)经工程化的向导RNA,其中所述经工程化的向导RNA被配置成与所述核酸内切酶形成复合物,并且所述经工程化的向导RNA包括间隔子序列,所述间隔子序列被配置成与所述基因座的区杂交,

其中所述2类V型Cas核酸内切酶具有与所述细胞中的spCas9至少等效的切割活性。

105.根据权利要求104所述的方法,其中所述切割活性是通过将所述核酸内切酶连同相容向导RNA引入到包括所述靶核酸的细胞并检测所述细胞中的所述靶核酸序列的切割来在体外测量的。

106.根据权利要求104或权利要求105所述的方法,其中所述组合物包括20皮摩尔(pmol)或更少的所述2类V型Cas核酸内切酶。

107.根据权利要求106所述的方法,其中所述组合物包括1pmol或更少的所述2类V型Cas核酸内切酶。

108.一种破坏细胞中的白蛋白基因座的方法,所述方法包括使包括以下的组合物与所述细胞接触:

(a)核酸内切酶,所述核酸内切酶与SEQ ID NO:1-325、420-431、476-624或629中的任一者或其变体具有至少75%同一性;以及

(b)经工程化的向导RNA,其中所述经工程化的向导RNA被配置成与所述核酸内切酶形成复合物,并且所述经工程化的向导RNA包括间隔子序列,所述间隔子序列被配置成与所述基因座的区杂交,

其中所述经工程化的向导RNA被配置成与表6中的靶序列中的任一者杂交。

109.根据权利要求108所述的方法,其中所述经工程化的向导RNA包括与SEQ ID NO:414-419432、434、436、438、440、442、444、446、448、450、452、454、456、458、460、462、464、466、468、470、472和474中的任一者的至少18个非简并核苷酸具有至少约80%、至少约85%、至少约86%、至少约87%、至少约88%、至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%或100%序列同一性的序列。

110.根据权利要求108或权利要求109所述的方法,其中所述经工程化的向导RNA包括表6中的单向导RNA(sgRNA)序列中的任一者的经修饰的核苷酸。

111.根据权利要求108至110中任一项所述的方法,其中所述核酸内切酶与SEQ ID NO:30-33、39、48、56、57、61、83、92、100、110、124、136、145、148、424、425、429、476或629中的任一者具有至少约75%、至少约80%、至少约85%、至少约86%、至少约87%、至少约88%、至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%或100%序列同一性。

112.根据权利要求111所述的方法,其中所述核酸内切酶与SEQ ID NO:57具有至少约75%、至少约80%、至少约85%、至少约86%、至少约87%、至少约88%、至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%或100%序列同一性。

113.根据权利要求108至112中任一项所述的方法,其中所述区位于PAM序列的5',所述PAM序列包括SEQ ID NO:433、435、437、439、441、443、445、447、449、451、453、455、457、459、461、463、465、467、469、471、473和475中的任一者。

114.一种分离的RNA分子,其包括与表6中的任何序列至少约80%、至少约85%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%或100%序列同一性的序列。

115.根据权利要求114所述的分离的RNA分子,其进一步包括表6中所述的向导RNA中的任一者中所述的化学修饰模式。

116.一种根据权利要求114或权利要求115所述的RNA分子的用途,其用于修饰细胞的白蛋白基因座。

117.一种经工程化的核酸酶系统,其包括:

(a)核酸内切酶,所述核酸内切酶被配置成对原间隔子相邻基序(PAM)具有选择性,所述PAM包括SEQ ID NO:433、435、437、439、441、443、445、447、449、451、453、455、457、459、461、463、465、467、469、471、473和475中的任一者;以及

(b)经工程化的向导RNA,其中所述经工程化的向导RNA被配置成与所述核酸内切酶形成复合物,并且所述经工程化的向导RNA包括间隔子序列,所述间隔子序列被配置成与靶核

酸序列杂交。

118.根据权利要求117所述的经工程化的核酸酶系统,其中所述核酸内切酶是2类V型Cas核酸内切酶。

119.根据权利要求117或权利要求118所述的经工程化的核酸酶系统,其中所述核酸内切酶不是Cas12a核酸酶。

120.根据权利要求117至119中任一项所述的经工程化的核酸酶系统,其中所述核酸内切酶源自未经培养的生物体。

121.根据权利要求117至120中任一项所述的经工程化的核酸酶系统,其中所述核酸内切酶进一步包括PAM相互作用结构域,所述PAM相互作用结构域被配置成与所述PAM相互作用。

122.根据权利要求117至121中任一项所述的经工程化的核酸酶系统,其中所述核酸内切酶与SEQ ID NO:1-325、420-431、476-624或629中的任一者或其变体具有至少75%序列同一性。

123.根据权利要求122所述的经工程化的核酸酶系统,其中所述核酸内切酶与SEQ ID NO:30-33、39、48、56、57、61、83、92、100、110、124、136、145、148、424、425、429、476或629中的任一者具有至少约75%、至少约80%、至少约85%、至少约86%、至少约87%、至少约88%、至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%或100%序列同一性。

124.一种经工程化的核酸酶系统,其包括:

(a)核酸内切酶,所述核酸内切酶与SEQ ID NO:1-325、420-431、476-624或629中的任一者或其变体具有至少75%序列同一性;以及

(b)DNA甲基转移酶。

125.根据权利要求124所述的经工程化的核酸酶系统,其中所述核酸内切酶与SEQ ID NO:30-33、39、48、56、57、61、83、92、100、110、124、136、145、148、424、425、429、476或629中的任一者具有至少约75%、至少约80%、至少约85%、至少约86%、至少约87%、至少约88%、至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%或100%序列同一性。

126.根据权利要求124或权利要求125所述的经工程化的核酸酶系统,其中所述DNA甲基转移酶与所述核酸内切酶非共价结合。

127.根据权利要求124或权利要求125所述的经工程化的核酸酶系统,其中所述DNA甲基转移酶与单个多肽中的所述核酸内切酶融合。

128.根据权利要求124至127中任一项所述的经工程化的核酸酶系统,其中所述DNA甲基转移酶包括Dmnt3A或Dnmt3L。

129.根据权利要求124至128中任一项所述的经工程化的核酸酶系统,其进一步包括KRAB结构域。

130.根据权利要求129所述的经工程化的核酸酶系统,其中所述KRAB结构域与所述核酸内切酶或所述DNA甲基转移酶非共价结合。

131.根据权利要求129所述的经工程化的核酸酶系统,其中所述KRAB结构域与所述核酸内切酶或所述DNA甲基转移酶共价连接。

132.根据权利要求131所述的经工程化的核酸酶系统,其中所述KRAB结构域与单个多肽中的所述核酸内切酶或所述DNA甲基转移酶融合。

133.根据权利要求124至132中任一项所述的经工程化的核酸酶系统,其中所述核酸内切酶是切口酶或是催化死亡的。

134.根据权利要求124至133中任一项所述的经工程化的核酸酶系统,其进一步包括经工程化的向导RNA结构,所述经工程化的向导RNA结构被配置成与所述核酸内切酶形成复合物,并且其中所述经工程化的向导RNA结构包括间隔子序列,所述间隔子序列被配置成与靶核酸序列杂交。

135.根据权利要求134所述的经工程化的核酸酶系统,其中所述靶核酸序列包括在靶基因组的启动子内或接近所述启动子。

136.根据权利要求134或权利要求135所述的经工程化的核酸酶系统,其中所述经工程化的向导RNA结构包括一个或多个:(a)2'-O-甲基核苷酸;(b)2'-氟核苷酸;或(c)硫代磷酸酯键。

137.根据权利要求134或权利要求135所述的经工程化的核酸酶系统,其中所述经工程化的向导RNA结构包括表6中的单向导RNA中的任一者的经化学修饰的核苷酸的模式。

138.一种修饰靶核酸基因座的方法,所述方法包括向所述靶核酸基因座递送根据权利要求124至137中任一项所述的所述经工程化的核酸酶系统,其中所述核酸内切酶被配置成与所述经工程化的向导RNA结构形成复合物,并且其中所述复合物被配置成使得在所述复合物与所述靶核酸基因座结合时,所述DNA甲基转移酶修饰所述靶核酸基因座。

139.一种根据权利要求124至137中任一项所述的经工程化的核酸酶系统的用途,其用于修饰核酸基因座。

140.根据权利要求139所述的用途,其中修饰所述核酸基因座包括使所述核酸基因座的核苷酸甲基化或去甲基化。

II类V型CRISPR系统

[0001] 相关申请

[0002] 本申请涉及PCT专利申请第PCT/US2021/021259号并且涉及PCT专利申请第PCT/US2022/031849号,所述专利申请中的每一个通过引用以其整体并入本文。

[0003] 交叉引用

[0004] 本申请要求于2021年9月8日提交的标题为“II类V型CRISPR系统(CLASS II,TYPE V CRISPR SYSTEMS)”的美国临时申请第63/241,928号的权益,所述美国临时申请通过引用以其整体并入本文。

背景技术

[0005] Cas酶以及其相关的成簇规则间隔短回文重复序列(CRISPR)向导核糖核酸(RNA)似乎是原核免疫系统的普遍组分(约45%的细菌,约84%的古细菌),用于通过CRISPR-RNA引导的核酸切割来保护此类微生物免受非自身核酸,如传染性病毒和质粒的侵害。虽然编码CRISPR RNA元件的脱氧核糖核酸(DNA)元件在结构和长度上可能相对保守,但其CRISPR相关(Cas)蛋白是高度多样化的,含有多种核酸相互作用结构域。虽然早在1987年就观察到CRISPR DNA元件,但CRISPR/Cas复合物的可编程核酸内切酶切割能力直到最近才被认识到,从而引起重组CRISPR/Cas系统在各种DNA操纵和基因编辑应用中的使用。

[0006] 序列表

[0007] 本申请含有已经以XML格式电子提交的序列表,并且特此通过引用以其整体并入。创建于2022年9月6日的所述XML副本命名为55921-732601_revised_2.xml并且大小为1,114,268字节。

发明内容

[0008] 在一些方面,本公开提供了一种经工程化的核酸酶系统,其包括:核酸内切酶,所述核酸内切酶与SEQ ID NO:1-325、420-431、476-624或629中的任一者或其变体具有至少75%序列同一性;以及经工程化的向导RNA,其中所述经工程化的向导RNA被配置成与所述核酸内切酶形成复合物,并且所述经工程化的向导RNA包括间隔子序列,所述间隔子序列被配置成与靶核酸序列杂交。

[0009] 在一些实施例中,所述向导RNA包括与SEQ ID NO:410-419、432、434、436、438、440、442、444、446、448、450、452、454、456、458、460、462、464、466、468、470、472和474中的任一者的非简并核苷酸具有至少80%序列同一性的序列。在一些实施例中,所述核酸内切酶与SEQ ID NO:30-33、39、48、56、57、61、83、92、100、110、124、136、145、148、424、425、429、476或629中的任一者具有至少约80%、至少约85%、至少约86%、至少约87%、至少约88%、至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%或100%序列同一性。在一些实施例中,所述向导RNA包括与SEQ ID NO:414-419、432、434、436、438、440、442、444、446、448、450、452、454、456、458、460、462、464、466、468、470、472和474中的任一者的非简并核苷酸具有

至少约85％、至少约86％、至少约87％、至少约88％、至少约89％、至少约90％、至少约91％、至少约92％、至少约93％、至少约94％、至少约95％、至少约96％、至少约97％、至少约98％、至少约99％或100％序列同一性的序列。

[0010] 在一些方面,本公开提供了一种经工程化的核酸酶系统,其包括:经工程化的向导RNA,所述经工程化的向导RNA包括与SEQ ID NO:410-419、432、434、436、438、440、442、444、446、448、450、452、454、456、458、460、462、464、466、468、470、472和474中的任一者的非简并核苷酸具有至少80％序列同一性的序列;以及2类V型Cas核酸内切酶,所述2类V型Cas核酸内切酶被配置成与所述经工程化的向导RNA结合。在一些实施例中,所述经工程化的核酸酶系统进一步包括DNA修复模板,所述DNA修复模板包括侧接有一个或两个单链DNA区段的双链DNA区段。在一些实施例中,所述单链DNA区段与所述双链DNA区段的5'端缀合。在一些实施例中,所述单链DNA区段与所述双链DNA区段的3'端缀合。在一些实施例中,所述单链DNA区段的长度为4至10个核苷酸碱基。

[0011] 在一些实施例中,所述单链DNA区段具有与所述间隔子序列内的序列互补的核苷酸序列。在一些实施例中,所述双链DNA序列包括条形码、开放阅读框、增强子、启动子、蛋白质编码序列、miRNA编码序列、RNA编码序列或转基因。在一些实施例中,所述双链DNA序列侧接有核酸酶切割位点。在一些实施例中,所述核酸酶切割位点包括间隔子和PAM序列。在一些实施例中,所述PAM包括SEQ ID NO:433、435、437、439、441、443、445、447、449、451、453、455、457、459、461、463、465、467、469、471、473和475中的任一者的序列。在一些实施例中,所述系统进一步包括Mg^{2+}的来源。在一些实施例中,所述向导RNA包括发夹,所述发夹包括至少8个、至少10个或至少12个碱基配对的核糖核苷酸。在一些实施例中,所述发夹包括10个碱基配对的核糖核苷酸。在一些实施例中,所述核酸内切酶包括与SEQ ID NO:1、6、15、30、151、292或319中的任一者至少75％、80％或90％相同的序列或其变体;并且所述向导RNA结构包括与SEQ ID NO:410-419中的任一者的非简并核苷酸至少80％或90％相同的序列。在一些实施例中,所述核酸内切酶包括与SEQ ID NO:30-33、39、48、56、57、61、83、92、100、110、124、136、145、148、424、425、429、476或629中的任一者至少约75％、至少约80％、至少约85％、至少至少约80％、至少约85％、至少约86％、至少约87％、至少约88％、至少约89％、至少约90％、至少约91％、至少约92％、至少约93％、至少约94％、至少约95％、至少约96％、至少约97％、至少约98％、至少约99％或100％序列同一性的序列;并且所述向导RNA结构包括与SEQ ID NO:414-419、432、434、436、438、440、442、444、446、448、450、452、454、456、458、460、462、464、466、468、470、472和474中的任一者的非简并核苷酸至少约80％、至少约85％、至少约86％、至少约87％、至少约88％、至少约89％、至少约90％、至少约91％、至少约92％、至少约93％、至少约94％、至少约95％、至少约96％、至少约97％、至少约98％、至少约99％或100％序列同一性的序列。在一些实施例中,所述序列同一性是通过BLASTP、CLUSTALW、MUSCLE、MAFFT算法或CLUSTALW算法使用史密斯-沃特曼同源性搜索算法参数(Smith-Waterman homology search algorithm parameter)来确定的。在一些实施例中,所述序列同一性是通过所述BLASTP同源性搜索算法使用字长(W)为3、期望值(E)为10的参数以及BLOSUM62评分矩阵将空位罚分设置为存在11,扩展1并且使用条件组成评分矩阵调整来确定的。

[0012] 在一些方面,本公开提供了一种经工程化的向导核糖核酸(RNA)多核苷酸,其包

括:DNA靶向区段,所述DNA靶向区段包括与靶DNA分子中的靶序列互补的核苷酸序列;以及蛋白质结合区段,所述蛋白质结合区段包括杂交以形成双链RNA(dsRNA)双链体的两个互补的核苷酸延伸段,其中所述两个互补的核苷酸延伸段用中间核苷酸彼此共价连接,并且其中所述经工程化的向导核糖核酸多核苷酸能够与2型V类Cas核酸内切酶形成复合物。在一些实施例中,所述2型V类Cas核酸内切酶源自未经培养的生物体。在一些实施例中,所述Cas核酸内切酶与SEQ ID NO:1-325、420-431、476-624或629中的任一者具有至少75%序列同一性,并且将所述复合物靶向到所述靶DNA分子的所述靶序列。在一些实施例中,所述DNA靶向区段定位在所述两个互补的核苷酸延伸段中的两者的3'。在一些实施例中,所述蛋白质结合区段包括与SEQ ID NO:410-419的非简并核苷酸具有至少70%、至少80%或至少90%同一性的序列。在一些实施例中,所述双链RNA(dsRNA)双链体包括至少5个、至少8个、至少10个或至少12个核糖核苷酸。

[0013] 在一些方面,本公开提供了一种脱氧核糖核酸多核苷酸,其编码本文所公开的经工程化的向导RNA中的任一种。

[0014] 在一些方面,本公开提供了一种核酸,其包括经工程化的核酸序列,所述经工程化的核酸序列被优化以在生物体中表达,其中所述核酸编码2类V型Cas核酸内切酶,并且其中所述核酸内切酶源自未经培养的微生物,其中所述生物体不是所述未经培养的生物体。在一些实施例中,所述核酸内切酶包括与SEQ ID NO:1-325、420-431、476-624或629中的任一者具有至少70%或至少80%序列同一性的变体。在一些实施例中,所述核酸内切酶包括编码接近所述核酸内切酶的N末端或C末端的一个或多个核定位序列(NLS)的序列。在一些实施例中,所述NLS包括选自SEQ ID NO:630-645的序列。在一些实施例中,所述NLS包括SEQ ID NO:631。在一些实施例中,所述NLS接近所述核酸内切酶的所述N末端。在一些实施例中,所述NLS包括SEQ ID NO:630。在一些实施例中,所述NLS接近所述核酸内切酶的所述C末端。在一些实施例中,所述生物体是原核生物、细菌、真核生物、真菌、植物、哺乳动物、啮齿动物或人。

[0015] 在一些方面,本公开提供了一种经工程化的载体,其包括核酸序列,所述核酸序列编码2类V型Cas核酸内切酶,其中所述核酸内切酶源自未经培养的微生物。

[0016] 在一些方面,本公开提供了一种经工程化的载体,其包括本文所公开的任何核酸。在一些实施例中,所述载体是质粒、微环、CELiD、腺相关病毒(AAV)源性病毒体、慢病毒或腺病毒。

[0017] 在一些方面,本公开提供了一种细胞,其包括本文所公开的任何经工程化的载体。

[0018] 在一些方面,本公开提供了一种制备核酸内切酶的方法,所述方法包括培养本文所公开的任何细胞。

[0019] 在一些方面,本公开提供了一种用于结合、切割、标记或修饰双链脱氧核糖核酸多核苷酸的方法,所述方法包括:使所述双链脱氧核糖核酸多核苷酸与2类V型Cas核酸内切酶接触,所述2类V型Cas核酸内切酶与经工程化的向导RNA复合,所述经工程化的向导RNA被配置成与所述核酸内切酶和所述双链脱氧核糖核酸多核苷酸结合;其中所述双链脱氧核糖核酸多核苷酸包括原间隔子相邻基序(PAM);并且其中所述向导RNA结构包括与SEQ ID NO:410-419中的任一者的非简并核苷酸至少80%或90%相同的序列。在一些实施例中,所述双链脱氧核糖核酸多核苷酸包括第一链和第二链,所述第一链包括与所述经工程化的向导

RNA的序列互补的序列,所述第二链包括所述PAM。在一些实施例中,所述PAM与和所述经工程化的向导RNA的所述序列互补的所述序列的5'端直接相邻。在一些实施例中,所述PAM包括SEQ ID NO:433、435、437、439、441、443、445、447、449、451、453、455、457、459、461、463、465、467、469、471、473和475中的任一者的序列。在一些实施例中,所述2类V型Cas核酸内切酶源自未经培养的微生物。在一些实施例中,所述2类V型Cas核酸内切酶进一步包括PAM相互作用结构域。在一些实施例中,所述双链脱氧核糖核酸多核苷酸是真核生物、植物、真菌、哺乳动物、啮齿动物或人双链脱氧核糖核酸多核苷酸。

[0020] 在一些方面,本公开提供了一种修饰靶核酸基因座的方法,所述方法包括向所述靶核酸基因座递送根据根据权利要求1至29中任一项所述的所述经工程化的核酸酶系统,其中所述核酸内切酶被配置成与所述经工程化的向导核糖核酸结构形成复合物,并且其中所述复合物被配置成使得在所述复合物与所述靶核酸基因座结合时,所述复合物修饰所述靶核酸基因座。在一些实施例中,修饰所述靶核酸基因座包括结合、切口、切割或标记所述靶核酸基因座。在一些实施例中,所述靶核酸基因座包括脱氧核糖核酸(DNA)或核糖核酸(RNA)。在一些实施例中,所述靶核酸包括基因组DNA、病毒DNA、病毒RNA或细菌DNA。在一些实施例中,所述靶核酸基因座是在体外的。在一些实施例中,所述靶核酸基因座是在细胞内的。在一些实施例中,所述细胞是原核细胞、细菌细胞、真核细胞、真菌细胞、植物细胞、动物细胞、哺乳动物细胞、啮齿动物细胞、灵长类动物细胞、人细胞或原代细胞。在一些实施例中,所述细胞是原代细胞。在一些实施例中,所述原代细胞是T细胞。在一些实施例中,所述原代细胞是造血干细胞(HSC)。在一些实施例中,向所述靶核酸基因座递送所述经工程化的核酸酶系统包括递送如本文所公开的任何核酸或如本文所公开的任何载体。在一些实施例中,向所述靶核酸基因座递送所述经工程化的核酸酶系统包括递送包括编码所述核酸内切酶的开放阅读框的核酸。在一些实施例中,所述核酸包括启动子,编码所述核酸内切酶的所述开放阅读框与所述启动子可操作地连接。在一些实施例中,向所述靶核酸基因座递送所述经工程化的核酸酶系统包括递送包含编码所述核酸内切酶的所述开放阅读框的加帽mRNA。在一些实施例中,向所述靶核酸基因座递送所述经工程化的核酸酶系统包括递送经翻译的多肽。在一些实施例中,向所述靶核酸基因座递送所述经工程化的核酸酶系统包括递送编码与核糖核酸(RNA)pol III启动子可操作地连接的所述经工程化的向导RNA的脱氧核糖核酸(DNA)。在一些实施例中,所述核酸内切酶诱导所述靶基因座处或接近所述靶基因座的单链断裂或双链断裂。在一些实施例中,所述核酸内切酶诱导所述靶基因座内或位于所述靶基因座的3'的交错单链断裂。

[0021] 在一些方面,本公开提供了一种宿主细胞,其包括编码异源核酸内切酶的开放阅读框,所述异源核酸内切酶与SEQ ID NO:1-325、420-431、476-624或629中的任一者或其变体具有至少75%序列同一性。在一些实施例中,所述核酸内切酶与SEQ ID NO:1、6、15、30、151、292或319中的任一者或其变体具有至少75%序列同一性。在一些实施例中,所述核酸内切酶与SEQ ID NO:30-33、39、48、56、57、61、83、92、100、110、124、136、145、148、424、425、429、476或629中的任一者具有至少约75%、至少约80%、至少约85%、至少约86%、至少约87%、至少约88%、至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%或100%序列同一性。在一些实施例中,所述宿主细胞是大肠杆菌(E.coli)细胞。在一些实施例中,所述大肠杆菌

细胞是λDE3溶原菌,或者所述大肠杆菌细胞是BL21(DE3)菌株。在一些实施例中,所述大肠杆菌细胞具有ompT lon基因型。在一些实施例中,所述开放阅读框与以下可操作地连接:T7启动子序列、T7-lac启动子序列、lac启动子序列、tac启动子序列、trc启动子序列、ParaBAD启动子序列、PrhaBAD启动子序列、T5启动子序列、cspA启动子序列、araPBAD启动子、来自噬菌体λ的强左侧启动子(pL启动子)或其任何组合。在一些实施例中,所述开放阅读框包括编码亲和标签的序列,所述亲和标签与编码所述核酸内切酶的序列框内连接。在一些实施例中,所述亲和标签是固定化金属亲和色谱法(IMAC)标签。在一些实施例中,所述IMAC标签是多组氨酸标签。在一些实施例中,所述亲和标签是myc标签、人流感血凝素(HA)标签、麦芽糖结合蛋白(MBP)标签、谷胱甘肽S-转移酶(GST)标签、链霉亲和素标签、FLAG标签或其任何组合。在一些实施例中,所述亲和标签经由编码蛋白酶切割位点的接头序列与编码所述核酸内切酶的所述序列框内连接。在一些实施例中,所述蛋白酶切割位点是烟草蚀纹病毒(TEV)蛋白酶切割位点、PreScission®蛋白酶(PSP)切割位点、凝血酶切割位点、因子Xa切割位点、肠激酶切割位点或其任何组合。在一些实施例中,所述开放阅读框被密码子优化以在所述宿主细胞中表达。在一些实施例中,所述开放阅读框被提供在载体上。在一些实施例中,所述开放阅读框被整合到所述宿主细胞的基因组中。

[0022] 在一些方面,本公开提供了一种培养物,其包括在相容液体培养基中的本文所公开的任何宿主细胞。

[0023] 在一些方面,本公开提供了一种产生核酸内切酶的方法,所述方法包括在相容液体培养基中培养本文所公开的任何宿主细胞。在一些实施例中,所述方法进一步包括通过添加另外的化学药剂或增加量的营养物来诱导所述核酸内切酶的表达。在一些实施例中,所述方法进一步包括在所述培养之后分离所述宿主细胞并且使所述宿主细胞裂解以产生蛋白质提取物。在一些实施例中,所述方法进一步包括使所述蛋白质提取物经受IMAC或离子亲和色谱法。在一些实施例中,所述方法进一步包括通过使对应于所述蛋白酶切割位点的蛋白酶与所述核酸内切酶接触来切割所述IMAC亲和标签。在一些实施例中,所述方法进一步包括执行减材IMAC亲和色谱法以从包括所述核酸内切酶的组合物中去除所述亲和标签。

[0024] 在一些方面,本公开提供了一种破坏细胞中的基因座的方法,所述方法包括使包括以下的组合物与所述细胞接触:2类V型Cas核酸内切酶,所述2类V型Cas核酸内切酶与SEQ ID NO:1-325、420-431、476-624或629中的任一者或其变体具有至少75%同一性;以及经工程化的向导RNA,其中所述经工程化的向导RNA被配置成与所述核酸内切酶形成复合物,并且所述经工程化的向导RNA包括间隔子序列,所述间隔子序列被配置成与所述基因座的区杂交,其中所述2类V型Cas核酸内切酶具有与所述细胞中的spCas9至少等效的切割活性。在一些实施例中,所述切割活性是通过将所述核酸内切酶连同相容向导RNA引入到包括所述靶核酸的细胞并检测所述细胞中的所述靶核酸序列的切割来在体外测量的。在一些实施例中,所述组合物包括20皮摩尔(pmol)或更少的所述2类V型Cas核酸内切酶。在一些实施例中,所述组合物包括1pmol或更少的所述2类V型Cas核酸内切酶。

[0025] 在一些方面,本公开提供了一种破坏细胞中的白蛋白基因座的方法,所述方法包括使包括以下的组合物与所述细胞接触:核酸内切酶,所述核酸内切酶与SEQ ID NO:1-325、420-431、476-624或629中的任一者或其变体具有至少75%同一性;以及经工程化的向

导RNA,其中所述经工程化的向导RNA被配置成与所述核酸内切酶形成复合物,并且所述经工程化的向导RNA包括间隔子序列,所述间隔子序列被配置成与所述基因座的区杂交,其中所述经工程化的向导RNA被配置成与表6中的靶序列中的任一者杂交。在一些实施例中,所述经工程化的向导RNA包括与SEQ ID NO:414-419432、434、436、438、440、442、444、446、448、450、452、454、456、458、460、462、464、466、468、470、472和474中的任一者的至少18个非简并核苷酸具有至少约80%、至少约85%、至少约86%、至少约87%、至少约88%、至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%或100%序列同一性的序列。在一些实施例中,所述经工程化的向导RNA包括表6中的单向导RNA(sgRNA)序列中的任一者的经修饰的核苷酸。在一些实施例中,所述核酸内切酶与SEQ ID NO:30-33、39、48、56、57、61、83、92、100、110、124、136、145、148、424、425、429、476或629中的任一者具有至少约75%、至少约80%、至少约85%、至少约86%、至少约87%、至少约88%、至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%或100%序列同一性。在一些实施例中,所述核酸内切酶与SEQ ID NO:57具有至少约75%、至少约80%、至少约85%、至少约86%、至少约87%、至少约88%、至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%或100%序列同一性。在一些实施例中,所述区位于PAM序列的5',所述PAM序列包括SEQ ID NO:433、435、437、439、441、443、445、447、449、451、453、455、457、459、461、463、465、467、469、471、473和475中的任一者。

[0026] 在一些方面,本公开提供了一种分离的RNA分子,其包括与表6中的任何序列至少约80%、至少约85%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%或100%序列同一性的序列。在一些实施例中,所述分离的RNA分子进一步包括表6中所述的向导RNA中的任一者中所述的化学修饰模式。

[0027] 在一些方面,本公开提供了一种本文所公开的任何RNA分子的用途,其用于修饰细胞的白蛋白基因座。

[0028] 在一些方面,本公开提供了一种经工程化的核酸酶系统,其包括:核酸内切酶,所述核酸内切酶被配置成对原间隔子相邻基序(PAM)具有选择性,所述PAM包括SEQ ID NO:433、435、437、439、441、443、445、447、449、451、453、455、457、459、461、463、465、467、469、471、473和475中的任一者;以及经工程化的向导RNA,其中所述经工程化的向导RNA被配置成与所述核酸内切酶形成复合物,并且所述经工程化的向导RNA包括间隔子序列,所述间隔子序列被配置成与靶核酸序列杂交。在一些实施例中,所述核酸内切酶是2类V型Cas核酸内切酶。在一些实施例中,所述核酸内切酶不是Cas12a核酸酶。在一些实施例中,所述核酸内切酶源自未经培养的生物体。在一些实施例中,所述核酸内切酶进一步包括PAM相互作用结构域,所述PAM相互作用结构域被配置成与所述PAM相互作用。在一些实施例中,所述核酸内切酶与SEQ ID NO:1-325、420-431、476-624或629中的任一者或其变体具有至少75%序列同一性。在一些实施例中,所述核酸内切酶与SEQ ID NO:30-33、39、48、56、57、61、83、92、100、110、124、136、145、148、424、425、429、476或629中的任一者具有至少约75%、至少约80%、至少约85%、至少约86%、至少约87%、至少约88%、至少约89%、至少约90%、至少约

91％、至少约92％、至少约93％、至少约94％、至少约95％、至少约96％、至少约97％、至少约98％、至少约99％或100％序列同一性。

[0029] 在一些方面,本公开提供了一种经工程化的核酸酶系统,其包括:核酸内切酶,所述核酸内切酶与SEQ ID NO:1-325、420-431、476-624或629中的任一者或其变体具有至少75％序列同一性;以及DNA甲基转移酶。在一些实施例中,所述核酸内切酶与SEQ ID NO:30-33、39、48、56、57、61、83、92、100、110、124、136、145、148、424、425、429、476或629中的任一者具有至少约75％、至少约80％、至少约85％、至少约86％、至少约87％、至少约88％、至少约89％、至少约90％、至少约91％、至少约92％、至少约93％、至少约94％、至少约95％、至少约96％、至少约97％、至少约98％、至少约99％或100％序列同一性。在一些实施例中,所述DNA甲基转移酶与所述核酸内切酶非共价结合。在一些实施例中,所述DNA甲基转移酶与单个多肽中的所述核酸内切酶融合。在一些实施例中,所述DNA甲基转移酶包括Dmnt3A或Dnmt3L。在一些实施例中,所述KRAB结构域与所述核酸内切酶或所述DNA甲基转移酶非共价结合。

[0030] 在一些实施例中,所述KRAB结构域与所述核酸内切酶或所述DNA甲基转移酶共价连接。在一些实施例中,所述KRAB结构域与单个多肽中的所述核酸内切酶或所述DNA甲基转移酶融合。在一些实施例中,所述核酸内切酶是切口酶或是催化死亡的。在一些实施例中,所述经工程化的核酸酶系统进一步包括经工程化的向导RNA结构,所述经工程化的向导RNA结构被配置成与所述核酸内切酶形成复合物,并且其中所述经工程化的向导RNA结构包括间隔子序列,所述间隔子序列被配置成与靶核酸序列杂交。在一些实施例中,所述靶核酸序列包括在靶基因组的启动子内或接近所述启动子。在一些实施例中,所述经工程化的向导RNA结构包括一个或多个:(a)2'-O-甲基核苷酸;(b)2'-氟核苷酸;或(c)硫代磷酸酯键。在一些实施例中,所述经工程化的向导RNA结构包括表6中的单向导RNA中的任一者的经化学修饰的核苷酸的模式。

[0031] 在一些方面,本公开提供了一种修饰靶核酸基因座的方法,所述方法包括向所述靶核酸基因座递送本文所公开的任何经工程化的核酸酶系统,其中所述核酸内切酶被配置成与所述经工程化的向导RNA结构形成复合物,并且其中所述复合物被配置成使得在所述复合物与所述靶核酸基因座结合时,所述DNA甲基转移酶修饰所述靶核酸基因座。

[0032] 在一些方面,本公开提供了一种本文所公开的任何经工程化的核酸酶系统的用途,其用于修饰核酸基因座。在一些实施例中,修饰所述核酸基因座包括使所述核酸基因座的核苷酸甲基化或去甲基化。

[0033] 在一些方面,本公开提供了一种经工程化的核酸酶系统,其包括:(a)核酸内切酶,所述核酸内切酶包括RuvC结构域,其中所述核酸内切酶源自未经培养的微生物,并且其中所述核酸内切酶不是Cas12a核酸酶;以及(b)经工程化的向导RNA,其中所述经工程化的向导RNA被配置成与所述核酸内切酶形成复合物,并且所述经工程化的向导RNA包括间隔子序列,所述间隔子序列被配置成与靶核酸序列杂交。在一些方面,本公开提供了一种经工程化的核酸酶系统,其包括:(a)核酸内切酶,所述核酸内切酶与SEQ ID NO:1-325、420-431、476-624或629中的任一者或其变体具有至少75％序列同一性;以及(b)经工程化的向导RNA,其中所述经工程化的向导RNA被配置成与所述核酸内切酶形成复合物,并且所述经工程化的向导RNA包括间隔子序列,所述间隔子序列被配置成与靶核酸序列杂交。在一些实施

例中,所述核酸内切酶包括RuvCI、II或III结构域。在一些实施例中,所述核酸内切酶与SEQ ID NO:1-325、420-431、476-624或629中的任一者或其变体的RuvCI、II或III结构域具有至少约20%、至少约25%、至少约30%、至少约35%、至少约40%、至少约45%、至少约50%、至少约55%、至少约60%、至少约65%、至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%同一性。在一些实施例中,所述RuvCI结构域包括D催化残基。在一些实施例中,所述RuvCII结构域包括E催化残基。在一些实施例中,所述RuvCIII结构域包括D催化残基。在一些实施例中,所述RuvC结构域不具有核酸酶活性。在一些实施例中,所述核酸内切酶进一步包括WED II结构域,所述WED II结构域与SEQ ID NO:1-325、420-431、476-624或629中的任一者或其变体的WED II结构域具有至少约20%、至少约25%、至少约30%、至少约35%、至少约40%、至少约45%、至少约50%、至少约55%、至少约60%、至少约65%、至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%同一性。在一些实施例中,所述向导RNA包括与SEQ ID NO:410-419中的任一者的非简并核苷酸具有至少80%序列同一性的序列。在一些方面,本公开提供了一种经工程化的核酸酶系统,其包括:(a)经工程化的向导RNA,所述经工程化的向导RNA包括与SEQ ID NO:410-419中的任一者的非简并核苷酸具有至少80%序列同一性的序列;以及(b)2类V型Cas核酸内切酶,所述2类V型Cas核酸内切酶被配置成与所述经工程化的向导RNA结合。在一些实施例中,所述向导RNA包括与真核、真菌、植物、哺乳动物或人基因组多核苷酸序列互补的序列。在一些实施例中,所述向导RNA的长度为30-250个核苷酸。在一些实施例中,所述核酸内切酶包括接近所述核酸内切酶的N末端或C末端的一个或多个核定位序列(NLS)。在一些实施例中,所述NLS包括与选自由SEQ ID NO:630-645组成的组的序列至少80%相同的序列。

[0034] 在一些实施例中,所述经工程化的核酸酶系统进一步包括单链或双链DNA修复模板,所述单链或双链DNA修复模板从5'至3'包括:第一同源臂,所述第一同源臂包括位于所述靶脱氧核糖核酸序列的5'的至少20个核苷酸的序列;至少10个核苷酸的合成DNA序列;以及第二同源臂,所述第二同源臂包括位于所述靶序列的3'的至少20个核苷酸的序列。在一些实施例中,所述第一同源臂或所述第二同源臂包括至少40、80、120、150、200、300、500或1,000个核苷酸的序列。在一些实施例中,所述第一同源臂和所述第二同源臂与原核生物、细菌、真菌或真核生物的基因组序列同源。在一些实施例中,所述单链或双链DNA修复模板包括转基因供体。在一些实施例中,所述经工程化的核酸酶系统进一步包括DNA修复模板,所述DNA修复模板包括侧接有一个或两个单链DNA区段的双链DNA区段。在一些实施例中,所述单链DNA区段与所述双链DNA区段的5'端缀合。在一些实施例中,所述单链DNA区段与所述双链DNA区段的3'端缀合。在一些实施例中,所述单链DNA区段的长度为4至10个核苷酸碱基。在一些实施例中,所述单链DNA区段具有与所述间隔子序列内的序列互补的核苷酸序列。在一些实施例中,所述双链DNA序列包括条形码、开放阅读框、增强子、启动子、蛋白质编码序列、miRNA编码序列、RNA编码序列或转基因。在一些实施例中,所述双链DNA序列侧接有核酸酶切割位点。在一些实施例中,所述核酸酶切割位点包括间隔子和PAM序列。在一些实施例中,所述系统进一步包括Mg^{2+}的来源。在一些实施例中,所述向导RNA包括发夹,所述发

夹包括至少8个、至少10个或至少12个碱基配对的核糖核苷酸。在一些实施例中,所述发夹包括10个碱基配对的核糖核苷酸。在一些实施例中,a)所述核酸内切酶包括与SEQ ID NO:1、6、15、30、151、292或319中的任一者至少75%、80%或90%相同的序列或其变体;以及b)向导RNA结构包括与SEQ ID NO:410-419中的任一者的非简并核苷酸至少80%或90%相同的序列。在一些实施例中,所述序列同一性是通过BLASTP、CLUSTALW、MUSCLE、MAFFT算法或CLUSTALW算法使用史密斯-沃特曼同源性搜索算法参数来确定的。在一些实施例中,所述序列同一性是通过所述BLASTP同源性搜索算法使用字长(W)为3、期望值(E)为10的参数以及BLOSUM62评分矩阵将空位罚分设置为存在11,扩展1并且使用条件组成评分矩阵调整来确定的。

[0035] 在一些方面,本公开提供了一种经工程化的向导RNA,其包括:a)DNA靶向区段,所述DNA靶向区段包括与靶DNA分子中的靶序列互补的核苷酸序列;以及b)蛋白质结合区段,所述蛋白质结合区段包括杂交以形成双链RNA(dsRNA)双链体的两个互补的核苷酸延伸段,其中所述两个互补的核苷酸延伸段用中间核苷酸彼此共价连接,并且其中所述经工程化的向导核糖核酸多核苷酸能够与核酸内切酶形成复合物,所述核酸内切酶与SEQ ID NO:1-325、420-431、476-624或629中的任一者具有至少75%序列同一性,并且将所述复合物靶向到所述靶DNA分子的所述靶序列。在一些实施例中,所述DNA靶向区段定位在所述两个互补的核苷酸延伸段中的两者的3'。在一些实施例中,所述蛋白质结合区段包括与SEQ ID NO:410-419的非简并核苷酸具有至少70%、至少80%或至少90%同一性的序列。在一些实施例中,所述双链RNA(dsRNA)双链体包括至少5个、至少8个、至少10个或至少12个核糖核苷酸。

[0036] 在一些方面,本公开提供了一种脱氧核糖核酸多核苷酸,其编码本文所述的经工程化的向导核糖核酸多核苷酸。

[0037] 在一些方面,本公开提供了一种核酸,其包括经工程化的核酸序列,所述经工程化的核酸序列被优化以在生物体中表达,其中所述核酸编码2类V型Cas核酸内切酶,并且其中所述核酸内切酶源自未经培养的微生物,其中所述生物体不是所述未经培养的生物体。在一些实施例中,所述核酸内切酶包括与SEQ ID NO:1-325、420-431、476-624或629中的任一者具有至少70%或至少80%序列同一性的变体。在一些实施例中,所述核酸内切酶包括编码接近所述核酸内切酶的N末端或C末端的一个或多个核定位序列(NLS)的序列。在一些实施例中,所述NLS包括选自SEQ ID NO:630-645的序列。在一些实施例中,所述NLS包括SEQ ID NO:631。在一些实施例中,所述NLS接近所述核酸内切酶的N末端。在一些实施例中,所述NLS包括SEQ ID NO:630。在一些实施例中,所述NLS接近所述核酸内切酶的C末端。在一些实施例中,所述生物体是原核生物、细菌、真核生物、真菌、植物、哺乳动物、啮齿动物或人。

[0038] 在一些方面,本公开提供了一种经工程化的载体,其包括核酸序列,所述核酸序列编码2类V型Cas核酸内切酶,其中所述核酸内切酶源自未经培养的微生物。

[0039] 在一些方面,本公开提供了一种经工程化的载体,其包括本文所述的核酸。

[0040] 在一些方面,本公开提供了一种经工程化的载体,其包括本文所述的脱氧核糖核酸多核苷酸。在一些实施例中,所述载体是质粒、微环、CELiD、腺相关病毒(AAV)源性病毒体、慢病毒或腺病毒。

[0041] 在一些方面,本公开提供了一种细胞,其包括本文所述的载体。

[0042] 在一些方面,本公开提供了一种制备核酸内切酶的方法,所述方法包括培养本文

所述的宿主细胞中的任何宿主细胞。

[0043] 在一些方面,本公开提供了一种用于结合、切割、标记或修饰双链脱氧核糖核酸多核苷酸的方法,所述方法包括:(a)使所述双链脱氧核糖核酸多核苷酸与2类V型Cas核酸内切酶接触,所述2类V型Cas核酸内切酶与经工程化的向导RNA复合,所述经工程化的向导RNA被配置成与所述核酸内切酶和所述双链脱氧核糖核酸多核苷酸结合;其中所述双链脱氧核糖核酸多核苷酸包括原间隔子相邻基序(PAM);并且其中所述向导RNA结构包括与SEQ ID NO:410-419中的任一者的非简并核苷酸至少80%或90%相同的序列。在一些实施例中,所述双链脱氧核糖核酸多核苷酸包括第一链和第二链,所述第一链包括与所述经工程化的向导RNA的序列互补的序列,所述第二链包括所述PAM。在一些实施例中,所述PAM与和所述经工程化的向导RNA的所述序列互补的所述序列的5'端直接相邻。在一些实施例中,所述2类V型Cas核酸内切酶源自未经培养的微生物。在一些实施例中,所述双链脱氧核糖核酸多核苷酸是真核生物、植物、真菌、哺乳动物、啮齿动物或人双链脱氧核糖核酸多核苷酸。

[0044] 在一些方面,本公开提供了一种修饰靶核酸基因座的方法,所述方法包括向所述靶核酸基因座递送本文所述的经工程化的核酸酶系统,其中所述核酸内切酶被配置成与所述经工程化的向导核糖核酸结构形成复合物,并且其中所述复合物被配置成使得在所述复合物与所述靶核酸基因座结合时,所述复合物修饰所述靶核酸基因座。在一些实施例中,修饰所述靶核酸基因座包括结合、切口、切割或标记所述靶核酸基因座。在一些实施例中,所述靶核酸基因座包括脱氧核糖核酸(DNA)或核糖核酸(RNA)。在一些实施例中,所述靶核酸包括基因组DNA、病毒DNA、病毒RNA或细菌DNA。在一些实施例中,所述靶核酸基因座是在体外的。在一些实施例中,所述靶核酸基因座是在细胞内的。在一些实施例中,所述细胞是原核细胞、细菌细胞、真核细胞、真菌细胞、植物细胞、动物细胞、哺乳动物细胞、啮齿动物细胞、灵长类动物细胞、人细胞或原代细胞。在一些实施例中,所述细胞是原代细胞。在一些实施例中,所述原代细胞是T细胞。在一些实施例中,所述原代细胞是造血干细胞(HSC)。在一些实施例中,向所述靶核酸基因座递送所述经工程化的核酸酶系统包括递送本文所述的核酸或本文所述的载体。在一些实施例中,向所述靶核酸基因座递送所述经工程化的核酸酶系统包括递送包括编码所述核酸内切酶的开放阅读框的核酸。在一些实施例中,所述核酸包括启动子,编码所述核酸内切酶的所述开放阅读框与所述启动子可操作地连接。在一些实施例中,向所述靶核酸基因座递送所述经工程化的核酸酶系统包括递送包含编码所述核酸内切酶的所述开放阅读框的加帽mRNA。在一些实施例中,向所述靶核酸基因座递送所述经工程化的核酸酶系统包括递送经翻译的多肽。在一些实施例中,向所述靶核酸基因座递送所述经工程化的核酸酶系统包括递送编码与核糖核酸(RNA)pol III启动子可操作地连接的所述经工程化的向导RNA的脱氧核糖核酸(DNA)。在一些实施例中,所述核酸内切酶诱导所述靶基因座处或接近所述靶基因座的单链断裂或双链断裂。在一些实施例中,所述核酸内切酶诱导所述靶基因座内或位于所述靶基因座的3'的交错单链断裂。

[0045] 在一些方面,本公开提供了一种宿主细胞,其包括编码异源核酸内切酶的开放阅读框,所述异源核酸内切酶与SEQ ID NO:1-325、420-431、476-624或629中的任一者或其变体具有至少75%序列同一性。在一些实施例中,所述核酸内切酶与SEQ ID NO:1、6、15、30、151、292或319中的任一者或其变体具有至少75%序列同一性。在一些实施例中,所述宿主细胞是大肠杆菌细胞或哺乳动物细胞。在一些实施例中,所述宿主细胞是大肠杆菌细胞。在

一些实施例中,所述大肠杆菌细胞是λDE3溶原菌,或者所述大肠杆菌细胞是BL21(DE3)菌株。在一些实施例中,所述大肠杆菌细胞具有ompT lon基因型。在一些实施例中,所述开放阅读框与以下可操作地连接:T7启动子序列、T7-lac启动子序列、lac启动子序列、tac启动子序列、trc启动子序列、ParaBAD启动子序列、PrhaBAD启动子序列、T5启动子序列、cspA启动子序列、$araP_{BAD}$启动子、来自噬菌体λ的强左侧启动子(pL启动子)或其任何组合。在一些实施例中,所述开放阅读框包括编码亲和标签的序列,所述亲和标签与编码所述核酸内切酶的序列框内连接。在一些实施例中,所述亲和标签是固定化金属亲和色谱法(IMAC)标签。在一些实施例中,所述IMAC标签是多组氨酸标签。在一些实施例中,所述亲和标签是myc标签、人流感血凝素(HA)标签、麦芽糖结合蛋白(MBP)标签、谷胱甘肽S-转移酶(GST)标签、链霉亲和素标签、FLAG标签或其任何组合。在一些实施例中,所述亲和标签经由编码蛋白酶切割位点的接头序列与编码所述核酸内切酶的所述序列框内连接。在一些实施例中,所述蛋白酶切割位点是烟草蚀纹病毒(TEV)蛋白酶切割位点、PreScission®蛋白酶切割位点、凝血酶切割位点、因子Xa切割位点、肠激酶切割位点或其任何组合。在一些实施例中,开放阅读框被密码子优化以在所述宿主细胞中表达。在一些实施例中,所述开放阅读框被提供在载体上。在一些实施例中,所述开放阅读框被整合到所述宿主细胞的基因组中。

[0046] 在一些方面,本公开提供了一种培养物,其包括在相容液体培养基中的本文所述的宿主细胞中的任何宿主细胞。

[0047] 在一些方面,本公开提供了一种产生核酸内切酶的方法,所述方法包括在相容液体培养基中培养本文所述的宿主细胞中的任何宿主细胞。在一些实施例中,所述方法进一步包括通过添加另外的化学药剂或增加量的营养物来诱导所述核酸内切酶的表达。在一些实施例中,所述另外的化学药剂或增加量的营养物包括异丙基β-D-1-硫代半乳糖苷(IPTG)或另外的量的乳糖。在一些实施例中,所述方法进一步包括在所述培养之后分离所述宿主细胞并且使所述宿主细胞裂解以产生蛋白质提取物。在一些实施例中,所述方法进一步包括使所述蛋白质提取物经受IMAC或离子亲和色谱法。在一些实施例中,所述开放阅读框包括编码IMAC亲和标签的序列,所述IMAC亲和标签与编码所述核酸内切酶的序列框内连接。在一些实施例中,所述IMAC亲和标签经由编码蛋白酶切割位点的接头序列与编码所述核酸内切酶的所述序列框内连接。在一些实施例中,所述蛋白酶切割位点包括烟草蚀纹病毒(TEV)蛋白酶切割位点、PreScission®蛋白酶切割位点、凝血酶切割位点、因子Xa切割位点、肠激酶切割位点或其任何组合。在一些实施例中,所述方法进一步包括通过使对应于所述蛋白酶切割位点的蛋白酶与所述核酸内切酶接触来切割所述IMAC亲和标签。在一些实施例中,所述方法进一步包括执行减材IMAC亲和色谱法以从包括所述核酸内切酶的组合物中去除所述亲和标签。

[0048] 在一些方面,本公开提供了一种破坏细胞中的基因座的方法,所述方法包括使包括以下的组合物与所述细胞接触:(a)2类V型Cas核酸内切酶,所述2类V型Cas核酸内切酶与SEQ ID NO:1-325、420-431、476-624或629中的任一者或其变体具有至少75%同一性;以及(b)经工程化的向导RNA,其中所述经工程化的向导RNA被配置成与所述核酸内切酶形成复合物,并且所述经工程化的向导RNA包括间隔子序列,所述间隔子序列被配置成与所述基因座的区杂交,其中所述2类V型Cas核酸内切酶具有与所述细胞中的spCas9至少等效的切割活性。在一些实施例中,所述切割活性是通过将所述核酸内切酶连同相容向导RNA引入到包

括所述靶核酸的细胞并检测所述细胞中的所述靶核酸序列的切割来在体外测量的。在一些实施例中,所述组合物包括20pmole或更少的所述2类V型Cas核酸内切酶。在一些实施例中,所述组合物包括1pmol或更少的所述2类V型Cas核酸内切酶。

[0049] 根据以下具体实施方式,本领域的技术人员将显而易知本公开的另外的方面和优势,在具体实施方式中仅示出和描述本公开的说明性实施例。如将认识到,本公开能够具有其它并且不同的实施例,并且其若干细节能够在各种明显的方面进行修改,所有这些都不脱离本公开。因此,附图和说明书本质上被视为是说明性的而非限制性的。

[0050] 通过引用并入

[0051] 本说明书中所提到的所有公开、专利和专利申请均通过以相同的程度引用并入本文,如同特定且单独地指示每个单独的公开、专利或专利申请通过引用并入。

附图说明

[0052] 本发明的新颖特征在所附权利要求书中具体阐述。通过参考阐述了说明性实施例的以下具体实施方式,将获得对本发明的特征和优点的更好理解,在所述实施例中利用了本发明的原理,并且在其附图中:

[0053] 图1描绘了在本公开之前先前描述的不同类别和类型的CRISPR/Cas基因座的典型组织。

[0054] 图2A-2D描绘了MG119家族的概述。图2A描绘了MG119效应子图示的多重比对,示出了对于双链DNA切割活性的功能至关重要的RuvC催化残基的结构域组成和保守性。图2B描绘了其中基因组背景围绕CRISPR阵列和Cas效应子(MG119-1的实例)的含有CRISPR的重叠群的图示。图2C描绘了MG119-1的直接重复序列的折叠。图2D描绘了设计用于MG119-1的单向导RNA。

[0055] 图3A-3C描绘了MG90家族的概述。图3A描绘了MG90效应子图示的多重比对,示出了对于双链DNA切割活性的功能至关重要的RuvC催化残基的结构域组成和保守性。图3B描绘了其中基因组背景围绕CRISPR阵列和Cas效应子(MG90-5的实例)的含有CRISPR的重叠群的图示。图3C描绘了MG90-5的直接重复序列的折叠。

[0056] 图4A-4C描绘了MG126家族的概述。图4A描绘了MG126效应子图示的多重比对,示出了对于双链DNA切割活性的功能至关重要的RuvC催化残基的结构域组成和保守性。图4B描绘了其中基因组背景围绕CRISPR阵列和Cas效应子(MG126-4的实例)的含有CRISPR的重叠群的图示。图4C描绘了MG126-4的直接重复序列的折叠。

[0057] 图5A-5C描绘了MG118家族的概述。图5A描绘了MG118效应子图示的多重比对,示出了对于双链DNA切割活性的功能至关重要的RuvC催化残基的结构域组成和保守性。图5B描绘了其中基因组背景围绕CRISPR阵列和Cas效应子(MG118-1的实例)的含有CRISPR的重叠群的图示。图5C描绘了MG118-1的直接重复序列的折叠。

[0058] 图6A-6C描绘了MG122家族的概述。图6A描绘了MG122效应子图示的多重比对,示出了对于双链DNA切割活性的功能至关重要的RuvC催化残基的结构域组成和保守性。图6B描绘了其中基因组背景围绕CRISPR阵列和Cas效应子(MG122-4的实例)的含有CRISPR的重叠群的图示。图6C描绘了MG122-4的直接重复序列的折叠。

[0059] 图7A-7C描绘了MG120家族的概述。图7A描绘了MG120效应子图示的多重比对,示出

了对于双链DNA切割活性的功能至关重要的RuvC催化残基的结构域组成和保守性。图7B描绘了其中基因组背景围绕CRISPR阵列和Cas效应子(MG120-1的实例)的含有CRISPR的重叠群的图示。图7C描绘了MG120-1的直接重复序列的折叠。

[0060] 图8A-8D描绘了MG91家族的概述。图8A描绘了其中基因组背景围绕CRISPR阵列和Cas效应子(MG91B-24的实例)的含有CRISPR的重叠群的图示。图8B描绘了MG91B-24的直接重复序列的折叠。图8C描绘了其中基因组背景围绕CRISPR阵列和Cas效应子(MG91C-10的实例)的含有CRISPR的重叠群的图示。图8D描绘了MG91C-10的直接重复序列的折叠。

[0061] 图9描绘了使用TXTL测定的MG119-2的体外活性。用来自MG119-2重叠群的两个基因间序列、含有正向或反向取向的重复序列的最小阵列(MA)序列以及PAM文库靶质粒测试MG119-2的dsDNA切割。使用基因间(IG)序列1以及具有正向取向的重复序列的最小阵列,在泳道1中观察到阳性基因间富集,作为扩增的切割产物。泳道3和7是省略IG的阴性对照,并且泳道4是省略阵列和IG两者的第三阴性对照。

[0062] 图10A描绘了经由对从体外切割测定获得的切割产物进行的下一代测序(NGS)确定的MG119-2 PAM(5'-nTnn-3')的SeqLogo。图10B描绘了割点的直方图(距PAM 23bd)。

[0063] 图11A和11B描绘了活性MG119核酸酶及其sgRNA设计的实例。图11A描绘了不具有间隔子的单向导RNA序列的预测折叠。蓝色圆圈表示tracrRNA的第一5'核苷酸,并且红色圆圈表示重复序列的3'核苷酸。TracrRNA和重复序列用GAAA四环来环成环。重复反重复序列(anti-repeat)折叠位于每个结构的3'端上。描绘了同一家族内的活性向导的三种不同RNA结构。从左到右:MG119-28向导具有四个发夹,在5'端上有三个较小的发夹,并且在重复反重复序列折叠旁边有具有两个凸起的非常长的发夹。MG119-83 sgRNA具有三个小发夹,并且重复反重复序列具有两个凸起。MG119-118具有四个发夹,第二发夹从5'端分支成三个发夹,而第三发夹和重复反重复序列具有一个凸起。此向导还具有在tracr的5'端与重复序列的3'端之间的一些配对核苷酸。图11B描绘了2%琼脂糖凝胶上的体外切割测定扩增产物。低分子量DNA序列梯(NEB)在泳道1、7和11中。从左到右的其它泳道内容:(2)仅MG119-28核酸酶、MG119-28核酸酶加(3)具有U67间隔子的sgRNA1、(4)具有U40间隔子的sgRNA1、(5)具有U67间隔子的sgRNA2和(6)具有U40间隔子的sgRNA2;(8)仅MG119-83核酸酶、MG119-83核酸酶加(9)具有U67间隔子的sgRNA1和(10)具有U40间隔子的sgRNA1;(12)仅MG119-118核酸酶、MG119-118核酸酶加(13)具有U67间隔子的sgRNA1和(14)具有U40间隔子的sgRNA1。所得扩增子产物在用携带U67间隔子的向导时为188bp,或者在用携带U40间隔子的向导时为205bp。

[0064] 图12描绘了活性MG119核酸酶的原间隔子相邻基序(PAM)的序列标志。

[0065] 图13A-13F描绘了蛋白质纯化步骤的示例SDS-PAGE凝胶和尺寸排除色谱法(SEC)A280迹线。图13A描绘了用从(1)超声后裂解、(2)澄清后离心、(3)Ni-NTA重力柱流出物、(4)来自Ni-NTA树脂的洗脱液、(5)经浓缩的样品回收的样品进行的MG119-28Δ净化。图13B描绘了S200i 10/300GL柱SEC A280迹线。合并峰级分并浓缩。图13C和13D描绘了用从(1)超声后裂解、(2)澄清后离心、(3)Ni-NTA重力柱流出物、(4)来自Ni-NTA树脂的洗脱液、(5)经浓缩的蛋白质、(6)用TEV蛋白酶隔夜切割的经浓缩的蛋白质、(7)以及离心(21,000x g,4℃,10分钟)成小球聚集体、(8)直链淀粉柱流出物、(9)流出物离心(21,000x g,4℃,10分钟)成小球聚集体以及(10)浓缩流出物回收的样品进行的经MBP标记/切割的MG119-28Δ净化。图

13E描绘了S200i 10/300GL柱SEC A280迹线。图13F描绘的数据证实,在pMGB和pMGBΔ表达载体两者中表达的五种MG119候选物中,所述候选物在pMGBΔ载体中均显示出更高的收率。

[0066] 图14A和14B描绘了用经纯化的蛋白质体外切割效率的实例。图14A描绘了示出了RNP:底物比滴定和在更高的比率下增加的底物切割的琼脂糖凝胶。图14B描绘了使用密度测定法测定每个泳道的切割底物百分比。在Prism8中绘制切割分数,并且使用线性切割范围的斜率来计算蛋白质活性分数。此测定使用在pMGBΔ主链中表达的MG119-28。

[0067] 图15A和15B描绘了小鼠Hepa1-6细胞DNA的体外切割和编辑效率。图15A描绘了具有靶向内含子1处的小鼠白蛋白基因的四种经化学修饰的向导的MG119-28的切割百分比(表6)。测试两种浓度的核酸酶15.6nM(黑色条)和7.8nM(白色条)。将切割相对于非靶向对照归一化。MG119-28可以在15.6nM RNP下用sgRNA4将Hepa 1-6gDNA切割至多平均60%,并且在7.8nM RNP下切割至多33%。图15B描绘了在相对于apo反应归一化的Hepa 1-6细胞中通过MG119-28产生的INDEL百分比。每种条件进行三次重复。平均25.12%的测序读数用sgRNA3进行编辑。如图所示,sgRNA3在体外和细胞中始终具有活性。细胞中的下一个最佳向导是平均编辑率为4.11%的sgRNA4。观察到的编辑在很大程度上是在4-24bp之间的缺失。

[0068] 序列表简要说明

[0069] 随此提交的序列表提供了用于根据本公开的方法、组合物和系统的示例性多核苷酸和多肽序列。下文是其中的序列的示例性描述。

[0070] MG122

[0071] SEQ ID NO:1-5示出了MG122核酸酶的全长肽序列。

[0072] MG120

[0073] SEQ ID NO:6-14示出了MG120核酸酶的全长肽序列。

[0074] SEQ ID NO:333-335和355-357示出了源自与MG120 Cas效应子的相同基因座的MG120 tracrRNA的核苷酸序列。

[0075] SEQ ID NO:374-375和389-390示出了MG120最小阵列的核苷酸序列。

[0076] MG118

[0077] SEQ ID NO:15示出了MG118核酸酶的全长肽序列。

[0078] SEQ ID NO:376示出了MG118最小阵列的核苷酸序列。

[0079] SEQ ID NO:391示出了MG118最小阵列的核苷酸序列。

[0080] SEQ ID NO:400-401示出了MG118靶CRISPR重复序列的核苷酸序列。

[0081] SEQ ID NO:410-411示出了MG118 crRNA的核苷酸序列。

[0082] MG90

[0083] SEQ ID NO:16-29示出了MG90核酸酶的全长肽序列。

[0084] SEQ ID NO:346-347和368-369示出了源自与MG90 Cas效应子的相同基因座的MG90 tracrRNA的核苷酸序列。

[0085] SEQ ID NO:383-384和398-399示出了MG90最小阵列的核苷酸序列。

[0086] SEQ ID NO:402-403示出了MG90靶CRISPR重复序列的核苷酸序列。

[0087] SEQ ID NO:412-413示出了MG90 sgRNA的核苷酸序列。

[0088] MG119

[0089] SEQ ID NO:30-150、420-431、476-624和629示出了MG119核酸酶的全长肽序列。

[0090] SEQ ID NO:326-332、336-345、348-354和358-367示出了源自与MG119 Cas效应子的相同基因座的MG119 tracrRNA的核苷酸序列。

[0091] SEQ ID NO:370-373、377-382、385-388和392-397示出了MG119最小阵列的核苷酸序列。

[0092] SEQ ID NO:404-409示出了MG119靶CRISPR重复序列的核苷酸序列。

[0093] SEQ ID NO:414-419、432、434、436、438、440、442、444、446、448、450、452、454、456、458、460、462、464、466、468、470、472和474示出了MG119 sgRNA的核苷酸序列。

[0094] SEQ ID NO:433、435、437、439、441、443、445、447、449、451、453、455、457、459、461、463、465、467、469、471、473和475示出了MG119 PAM的核苷酸序列。

[0095] MG91B

[0096] SEQ ID NO:151-291示出了MG91B核酸酶的全长肽序列。

[0097] MG91C

[0098] SEQ ID NO:292-318示出了MG91C核酸酶的全长肽序列。

[0099] MG91A

[0100] SEQ ID NO:319示出了MG91A核酸酶的全长肽序列。

[0101] MG126

[0102] SEQ ID NO:320-325示出了MG126核酸酶的全长肽序列。

具体实施方式

[0103] 虽然本文中已经示出并描述了本发明的各种实施例,但是对于本领域的技术人员显而易见的是,此类实施例仅作为实例提供。在不脱离本发明的情况下,本领域技术人员可以想到多种变化、改变和替换。应当理解,可以采用本文所描述的本发明的实施例的各种替代方案。

[0104] 除非另有指示,否则本文所公开的一些方法的实践采用免疫学、生物化学、化学、分子生物学、微生物学、细胞生物学、基因组学和重组DNA的技术。参见例如,Sambrook和Green,《分子克隆:实验室手册(Molecular Cloning:A Laboratory Manual)》,第4版(2012);丛书《当代分子生物学实验指南(Current Protocols in Molecular Biology)》(F.M.Ausubel等人编辑);丛书《酶学方法(Methods In Enzymology)》(学术出版社公司(Academic Press,Inc.)),《PCR 2:实用方法(PCR 2:A Practical Approach)》(M.J.MacPherson,B.D.Hames和G.R.Taylor编辑(1995));Harlow和Lane编辑(1988)《抗体:实验室手册(Antibodies,ALaboratory Manual)》以及《动物细胞培养:基础技术和专门应用手册(Culture of Animal Cells:A Manual of Basic Technique and Specialized Applications)》,第6版(R.I.Freshney编辑(2010))(所述文献通过引用整体并入本文)。

[0105] 如本文所使用的,除非上下文另外清楚地指示,否则单数形式“一个(a)”、“一种(an)”和“所述(the)”旨在也包含复数形式。此外,在具体实施方式和/或权利要求中使用术语“包含(including)”、“包含(include)”、“具有(having)”、“具有(has)”、“具有(with)”或其变体的情况下,此类术语旨在以类似于术语“包括”的方式是包含性的。

[0106] 术语“约”或“大约”意指在如由本领域普通技术人员确定的特定值的可接受误差范围内,这将部分地取决于值是如何测量或确定的,即,测量系统的局限性。例如,“约”可以

意指按照本领域的实践在一个或超过一个标准差内。可替代地,“约”可以意指给定值的至多20%、至多15%、至多10%、至多5%或至多1%的范围。

[0107] 如本文所使用的,“细胞”通常是指生物细胞。细胞可以是活生物体的基本结构、功能和/或生物单位。细胞可以源自具有一个或多个细胞的任何生物体。一些非限制性实例包含:原核细胞、真核细胞、细菌细胞、古细菌细胞,单细胞真核生物体的细胞、原生动物细胞、来自植物的细胞(例如来自种植农作物、水果、蔬菜、谷物、大豆、玉米、玉蜀黍、小麦、种子、西红柿、大米、木薯、甘蔗、南瓜、干草、土豆、棉花、大麻、烟草、开花植物、针叶树、裸子植物、蕨类植物、石松、角苔纲、苔类、苔藓的细胞)、藻细胞(例如,布朗葡萄藻(Botryococcus braunii)、莱茵衣藻(Chlamydomonas reinhardtii)、拟微球藻(Nannochloropsis gaditana)、蛋白核小球藻(Chlorella pyrenoidosa)、亨氏马尾藻(Sargassum patens C.Agardh)等)、海藻(例如海带)、真菌细胞(例如酵母细胞,来自蘑菇的细胞)、动物细胞、来自无脊椎动物(例如果蝇、刺胞动物、棘皮动物、线虫等)的细胞、来自脊椎动物(例如鱼、两栖动物、爬行动物、鸟、哺乳动物)的细胞、来自哺乳动物(例如猪、牛、山羊、绵羊、啮齿动物、大鼠、小鼠、非人灵长类动物、人等)的细胞等。有时,细胞并非源自天然生物体(例如,细胞可以是合成制造的,有时被称为人工细胞)。

[0108] 如本文所使用的,术语“核苷酸”通常是指碱基-糖-磷酸组合。核苷酸可以包括合成核苷酸。核苷酸可以包括合成核苷酸类似物。核苷酸可以是核酸序列(例如脱氧核糖核酸(DNA)和核糖核酸(RNA))的单体单元。术语核苷酸可以包含:核糖核苷三磷酸,腺苷三磷酸(ATP)、尿苷三磷酸(UTP)、胞嘧啶三磷酸(CTP)、鸟苷三磷酸(GTP);以及脱氧核糖核苷三磷酸,如dATP、dCTP、dITP、dUTP、dGTP、dTTP或其衍生物。此类衍生物可以包含例如[αS]dATP、7-脱氮-dGTP和7-脱氮-dATP,以及赋予含有其的核酸分子核酸酶抗性的核苷酸衍生物。如本文所使用的,术语核苷酸可以是指双脱氧核糖核苷三磷酸(ddNTP)及其衍生物。双脱氧核糖核苷三磷酸的说明性实例可以包含但不限于:ddATP、ddCTP、ddGTP、ddITP和ddTTP。核苷酸可以是未经标记的或经可检测标记的,如使用包括光学可检测部分(例如荧光团)的部分。也可以用量子点进行标记。可检测标记可以包含例如放射性同位素、荧光标记、化学发光标记、生物发光标记和酶标记。核苷酸的荧光标记可以包含但不限于:荧光素、5-羧基荧光素(FAM)、2’7’-二甲氧基-4’5-二氯-6-羧基荧光素(JOE)、罗丹明、6-羧基罗丹明(R6G)、N,N,N’,N’-四甲基-6-羧基罗丹明(TAMRA)、6-羧基-X-罗丹明(ROX)、4-(4’二甲氨基苯偶氮基)苯甲酸(DABCYL)、瀑布蓝、俄勒冈绿、德克萨斯红、青色素和5-(2’-氨乙基)氨基萘-1-磺酸(EDANS)。经荧光标记的核苷酸的具体实例可以包含:可从加利福尼亚州福斯特市的铂金埃尔默公司(Perkin Elmer、Foster City、Calif)获得的[R6G]dUTP、[TAMRA]dUTP、[R110]dCTP、[R6G]dCTP、[TAMRA]dCTP、[JOE]ddATP、[R6G]ddATP、[FAM]ddCTP、[R110]ddCTP、[TAMRA]ddGTP、[ROX]ddTTP、[dR6G]ddATP、[dR110]ddCTP、[dTAMRA]ddGTP和[dROX]ddTTP;可从伊利诺伊州阿灵顿高地的安玛西亚公司(Amersham,Arlington Heights,Il.)获得的FluoroLink脱氧核苷酸、FluoroLink Cy3-dCTP、FluoroLink Cy5-dCTP、FluoroLink Fluor X-dCTP、FluoroLink Cy3-dUTP和FluoroLink Cy5-dUTP;可从印第安纳州印第安纳波利斯的宝灵曼公司(Boehringer Mannheim,Indianapolis,Ind.)获得的荧光素-15-dATP、荧光素-12-dUTP、四甲基-罗丹明-6-dUTP、IR770-9-dATP、荧光素-12-ddUTP、荧光素-12-UTP和荧光素-15-2’-dATP;以及可从俄勒冈州尤金的分子探针公司(Molecular Probes,Eugene,

Oreg)获得的经染色体标记的核苷酸、BODIPY-FL-14-UTP、BODIPY-FL-4-UTP、BODIPY-TMR-14-UTP、BODIPY-TMR-14-dUTP、BODIPY-TR-14-UTP、BODIPY-TR-14-dUTP、瀑布蓝-7-UTP、瀑布蓝-7-dUTP、荧光素-12-UTP、荧光素-12-dUTP、俄勒冈绿488-5-dUTP、罗丹明绿-5-UTP、罗丹明绿-5-dUTP、四甲基罗丹明-6-UTP、四甲基罗丹明-6-dUTP、德克萨斯红-5-UTP、德克萨斯红-5-dUTP和德克萨斯红-12-dUTP。核苷酸也可以通过化学修饰进行标记或标记。经化学修饰的单核苷酸可以是生物素-dNTP。经生物素化的dNTP的一些非限制性实例可以包含生物素-dATP(例如,bio-N6-ddATP、生物素-14-dATP)、生物素-dCTP(例如,生物素-11-dCTP、生物素-14-dCTP)和生物素-dUTP(例如生物素-11-dUTP、生物素-16-dUTP、生物素-20-dUTP)。

[0109] 术语“多核苷酸”、“寡核苷酸”和“核酸”可互换使用以通常指代具有任何长度的核苷酸,即脱氧核糖核苷酸或核糖核苷酸或其类似物的聚合形式,呈单链、双链或多链形式。多核苷酸对于细胞可以是外源性的或内源性的。多核苷酸可以存在于无细胞的环境中。多核苷酸可以是基因或其片段。多核苷酸可以是DNA。多核苷酸可以是RNA。多核苷酸可以具有任何三维结构,并且可以进行任何功能。多核苷酸可以包括一种或多种类似物(例如,改变的主链、糖或核碱基)。如果存在,则可以在组装聚合物之前或之后赋予对核苷酸结构的修饰。类似物的一些非限制性实例包含:5-溴尿嘧啶、肽核酸、异源核酸、吗啉代、锁核酸、甘油核酸、苏糖核酸、双脱氧核苷酸、虫草素、7-脱氮-GTP、荧光团(例如,与糖连接的罗丹明或荧光素)、含硫醇的核苷酸、生物素连接的核苷酸、荧光碱基类似物、CpG岛、甲基-7-鸟苷、甲基化核苷酸、肌苷、硫尿苷、假尿苷、二氢尿苷、辫苷和怀俄苷。多核苷酸的非限制性实例包含基因或基因片段的编码或非编码区域、根据连接分析定义的多个基因座(基因座)、外显子、内含子、信使RNA(mRNA)、转移RNA(tRNA)、核糖体RNA(rRNA)、短干扰RNA(siRNA)、短发夹RNA(shRNA)、微RNA(miRNA)、核酶、cDNA、重组多核苷酸、分支多核苷酸、质粒、载体、任何序列的分离的DNA、任何序列的分离的RNA、包含无细胞DNA(cfDNA)和无细胞RNA(cfRNA)的无细胞多核苷酸、核酸探针和引物。核苷酸的序列可以间杂有非核苷酸组分。

[0110] 术语“转染(transfection)”或“转染(transfected)”通常指通过非病毒或基于病毒的方法将核酸引入细胞中。核酸分子可以是编码完整蛋白或其功能部分的基因序列。参见例如Sambrook等人1989,《分子克隆:实验室手册》,18.1-18.88(所述文献通过引用整体并入本文)。

[0111] 术语“肽”、“多肽”和“蛋白质”在本文中可互换使用以通常指代至少两个通过肽键连接的氨基酸残基的聚合物。此术语不表示聚合物的具体长度,也不旨在暗示或区分肽是使用重组技术、化学或酶促合成产生的还是天然存在的。所述术语适用于天然存在的氨基酸聚合物以及包括至少一种经修饰的氨基酸的氨基酸聚合物。在一些情况下,聚合物可以间杂有非氨基酸。所述术语包含任何长度的氨基酸链,包含全长蛋白质以及具有或不具有二级和/或三级结构(例如,结构域)的蛋白质。术语还涵盖已被修饰的氨基酸聚合物;例如通过二硫键形成、糖基化、脂化、乙酰化、磷酸化、氧化和任何其它操作,如与标记组分缀合。如本文所使用的,术语“氨基酸(amino acid)”和“氨基酸(amino acids)”通常是指天然和非天然氨基酸,包含但不限于经修饰的氨基酸和氨基酸类似物。经修饰的氨基酸可以包含已被化学修饰以包含非天然存在于氨基酸上的基团或化学部分的天然氨基酸和非天然氨基酸。氨基酸类似物可以指氨基酸衍生物。术语“氨基酸”包含D-氨基酸和L-氨基酸。

[0112] 如本文所使用的,“非天然”通常可以指在天然核酸或蛋白质中未发现的核酸或多肽序列。非天然可以指亲和标签。非天然可以指融合物。非天然可以指天然存在的包括突变、插入和/或缺失的核酸或多肽序列。非天然序列可以表现出和/或编码也可以由与非天然序列融合的核酸和/或多肽序列表现出的活性(例如,酶活性、甲基转移酶活性、乙酰转移酶活性、激酶活性、泛素化活性等)。非天然核酸或多肽序列可以通过基因工程化与天然存在的核酸或多肽序列(或其变体)连接以产生嵌合核酸和/或编码嵌合核酸和/或多肽的多肽序列。

[0113] 如本文所使用的,术语“启动子”通常是指控制基因转录或表达并且可以位于与启动RNA转录的核苷酸或核苷酸区域相邻或重叠的调节DNA区域。启动子可以含有结合蛋白质因子(通常称为转录因子)的特定DNA序列,其促进RNA聚合酶与DNA的结合,从而导致基因转录。‘基础启动子’,也称为‘核心启动子’,通常可以指含有促进可操作连接的多核苷酸转录表达的所有基本必需元件的启动子。真核基础启动子通常(尽管不一定)含有TATA盒和/或CAAT盒。

[0114] 如本文所使用的,术语“表达”通常是指从DNA模板转录核酸序列或多核苷酸(如转录为mRNA或其它RNA转录本)的过程和/或随后将经转录的mRNA翻译为肽、多肽、或蛋白质的过程。转录物和编码的多肽可以统称为“基因产物”。如果多核苷酸源自基因组DNA,则表达可以包含在真核细胞中剪接mRNA。

[0115] 如本文所使用的,“可操作地连接”、“可操作连接”、“操作性地连接”或其语法等效物通常是指遗传元件,例如启动子、增强子、聚腺苷酸化序列等的并置,其中所述元件处于允许其以预期方式操作的关系中。例如,如果调节元件有助于启动编码序列的转录,则可以包括启动子和/或增强子序列的调节元件与编码区可操作地连接。只要维持这种功能关系,调节元件与编码区之间就会存在插入残基。

[0116] 如本文所使用的,“载体”通常是指包括多核苷酸或与多核苷酸缔合并且可以被用于介导多核苷酸到细胞的递送的大分子或大分子的缔合物。载体的实例包含质粒、病毒载体、脂质体和其它基因递送媒剂。载体通常包括遗传元件(例如调节元件),其与基因可操作地连接以促进基因在靶标中的表达。

[0117] 如本文所使用的,“表达盒”和“核酸盒”通常可互换使用以指代一起表达或可操作地连接用于表达的核酸序列或元件的组合。在一些情况下,表达盒是指调节元件和其可操作地连接用于表达的一个或多个基因的组合。

[0118] DNA或蛋白质序列的“功能片段”通常是指保留与全长DNA或蛋白质序列的生物活性基本上类似的(功能或结构)生物活性的片段。DNA序列的生物活性可能是其以归因于全长序列的已知方式影响表达的能力。

[0119] 如本文所使用的,“经工程化的”对象通常表明所述对象已通过人为干预进行修饰。根据非限制性实例:核酸可以通过将其序列改变成自然界中不存在的序列来修饰;核酸可以通过将其连接到自然界中不与其缔合的核酸来修饰,使得连接产物具有原始核酸中不存在的功能;经工程化的核酸可以用自然界不存在的序列在体外合成;蛋白质可以通过将其氨基酸序列改变成自然界中不存在的序列来修饰;经工程化的蛋白质可以获得新的功能或特性。“经工程化的”系统包括至少一个经工程化的组分。

[0120] 如本文所使用的,“合成的”和“人工的”通常可以互换使用以指代与天然存在的人

蛋白质具有低序列同一性(例如小于50％序列同一性、小于25％序列同一性、小于10％序列同一性、小于5％序列同一性、小于1％序列同一性)的蛋白质或其结构域。例如,VPR和VP64结构域是合成的反式激活结构域。

[0121] 如本文所使用的,术语“Cas12a”通常指属于2类V-A型Cas核酸内切酶的Cas核酸内切酶家族,并且(a)使用相对较小的向导RNA(约42-44个核苷酸),其在从CRISPR阵列转录后由核酸酶自身处理,以及(b)切割DNA以留下交错切割位点。该酶家族的另外的特征可见于例如Zetsche B,Heidenreich M,Mohanraju P等人《自然生物技术(Nat Biotechnol)》2017;35:31-34和Zetsche B,Gootenberg JS,Abudayyeh OO等人《细胞(Cell)》2015;163:759-771,所述文献通过引用并入本文。

[0122] 如本文所使用的,“向导核酸”通常可以指可以与另一个核酸杂交的核酸。向导核酸可以是RNA。向导核酸可以是DNA。向导核酸可以被编程成与核酸序列位点特异性结合。待靶向的核酸或靶核酸可以包括核苷酸。向导核酸可以包括核苷酸。靶核酸的一部分可以与向导核酸的一部分互补。与向导核酸互补并杂交的双链靶多核苷酸的链可以被称为互补链。双链靶多核苷酸的与互补链互补并且因此可能不与向导核酸互补的链可以被称为非互补链。向导核酸可以包括多核苷酸链,并且可以被称为“单向导核酸”。向导核酸可以包括两条多核苷酸链,并且可以称为“双向导核酸”。如果没有另外说明,则术语“向导核酸”可以是包含性的,指的是单向导核酸和双向导核酸两者。向导核酸可以包括可以被称为“核酸靶向区段”或“核酸靶向序列”或“间隔子序列”的区段。核酸靶向区段可以包括子区段,所述子区段可以被称为“蛋白质结合区段”或“蛋白质结合序列”或“Cas蛋白结合区段”。

[0123] 在两个或更多个核酸或多肽序列的上下文中,术语“序列同一性”或“百分比同一性”通常是指当在局部或全局比较窗内进行比较和比对以获得最大对应性时,两个(例如,在成对比对中)或更多个(例如,在多序列比对中)序列相同或具有相同特定百分比的氨基酸残基或核苷酸,如使用序列比较算法测量的。用于多肽序列的合适的序列比较算法包含例如:用于长于30个残基的多肽序列的使用字长(W)为3、期望值(E)为10的参数以及BLOSUM62评分矩阵将空位罚分设置为存在11,扩展1并且使用条件组成评分矩阵调整的BLASTP;用于少于30个残基的序列的使用字长(W)为2、期望值(E)为1000000的参数以及PAM30评分矩阵将空位罚分设置为空位打开为9以及扩展空位为1的BLASTP(这些是BLAST套件中BLASTP的默认参数,可在https://blast.ncbi.nlm.nih.gov获得);使用匹配为2、失配为-1并且空位为-1的史密斯-沃特曼同源性搜索算法参数的CLUSTALW;使用默认参数的MUSCLE;使用retree为2并且最大迭代为1000的参数的MAFFT;使用默认参数的Novafold;使用默认参数的HMMER hmmalign。

[0124] 在两个或更多个核酸或多肽序列的上下文中,术语“最佳比对”通常是指已经与氨基酸残基或核苷酸的最大对应性比对的两个(例如,在成对比对中)或更多个(例如,在多序列比对中)序列,例如,如通过产生最高或“优化”百分比同一性评分的比对确定的。

[0125] 本公开包含本文所描述的具有一个或多个保守氨基酸取代的酶中的任何酶的变体。此类保守取代可以在多肽的氨基酸序列中进行,而不破坏多肽的三维结构或功能。保守取代可以通过具有类似疏水性、极性和R链长度的氨基酸彼此取代来完成。另外或可替代地,通过比较来自不同物种的同源蛋白质的比对序列,可以通过定位物种之间已突变的氨基酸残基(例如非保守残基)而不改变经编码的蛋白质的基本功能来鉴定保守取代。此类保

守取代的变体可以包含与本文所述的核酸内切酶蛋白质序列中的任何一个核酸内切酶蛋白质序列(例如,本文所述的MG90、MG91A、MG91B、MG91C、MG118、MG119、MG120、MG122或MG126家族核酸内切酶,或本文所描述的任何其它家族核酸酶)具有至少约20%、至少约25%、至少约30%、至少约35%、至少约40%、至少约45%、至少约50%、至少约55%、至少约60%、至少约65%、至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%同一性的变体。在一些实施例中,此类经保守取代的变体是功能变体。此类功能变体可以涵盖具有取代的序列,使得核酸内切酶的一个或多个关键活性位点残基或向导RNA结合残基的活性不被破坏。在一些实施例中,本文所述的蛋白质中的任何蛋白质的功能变体缺乏图2A、3A、4A、5A或6A中所指示的保守或功能残基中的至少一个的取代。在一些实施例中,本文所述的蛋白质中的任何蛋白质的功能变体缺乏图2A、3A、4A、5A或6A中所指示的全部保守或功能残基的取代。

[0126] 本公开还包含本文所述的酶中的任何酶的变体,其取代一个或多个催化残基以降低或消除酶的活性(例如,活性降低的变体)。在一些实施例中,作为本文所述的蛋白质的活性降低的变体包括图2A、3A、4A、5A或6A中所指示的至少一个、至少两个或所有三个催化残基的破坏性取代。

[0127] 提供功能类似氨基酸的保守取代表可从各种参考文献中获得(参见例如Creighton,《蛋白质:结构与分子特性(Proteins:Structures and Molecular Properties)》(W H弗里曼出版社(W H Freeman&Co.);第2版(1993年12月)))。以下八个基团各自含有彼此保守取代的氨基酸:

[0128] 1)丙氨酸(A)、甘氨酸(G);

[0129] 2)天冬氨酸(D)、谷氨酸(E);

[0130] 3)天冬酰胺(N)、谷氨酰胺(Q);

[0131] 4)精氨酸(R)、赖氨酸(K);

[0132] 5)异亮氨酸(I)、亮氨酸(L)、甲硫氨酸(M)、缬氨酸(V);

[0133] 6)苯丙氨酸(F)、酪氨酸(Y)、色氨酸(W);

[0134] 7)丝氨酸(S)、苏氨酸(T);以及

[0135] 8)半胱氨酸(C)、甲硫氨酸(M)

[0136] 概述

[0137] 具有独特功能和结构的新Cas酶的发现可能会提供进一步破坏脱氧核糖核酸(DNA)编辑技术的可能性,从而提高速度、特异性、功能和易用性。相对于微生物中成簇的规则间隔的短回文重复序列(CRISPR)系统的预测流行率和微生物物种的纯粹多样性,文献中存在功能表征的CRISPR/Cas酶相对较少。这部分是因为大量的微生物物种可能不容易在实验室条件下培养。对含有大量微生物物种的自然环境生态位进行宏基因组测序可以提供大幅增加已知的新的CRISPR/Cas系统的数量以及加速新寡核苷酸编辑功能的发现的可能性。这种方法富有成效的最近的实例通过2016年通过对天然微生物群落的宏基因组分析发现CasX/CasY CRISPR系统证明。

[0138] CRISPR/Cas系统是RNA引导的核酸酶复合物,其已描述为在微生物中充当适应性免疫系统。在CRISPR/Cas系统的自然环境中,所述CRISPR/Cas系统出现在CRISPR(成簇的规

则间隔的短回文重复序列)操纵子或基因座中,其通常包括两部分:(i)由同样短的间隔子序列分隔的短重复序列阵列(30-40bp),其编码基于RNA的靶向元件;以及(ii)编码Cas的ORF,所述Cas编码由基于RNA的靶向元件以及辅助蛋白/酶引导的核酸酶多肽。特定靶核酸序列的有效核酸酶靶向通常需要以下两者:(i)靶标(靶种子)的前6-8个核酸与crRNA向导之间的互补杂交;以及(ii)靶种子的定义附近内存在原间隔子相邻基序(PAM)序列(PAM通常是宿主基因组内不常表示的序列)。根据系统的确切功能和组织,CRISPR-Cas系统通常基于共享的功能特性和进化类似性分为2种类别、5种类型和16种亚型(参见图1)。

[0139] I类CRISPR-Cas系统具有大的多亚基效应子复合物,并且包括I、III和IV型。II类CRISPR-Cas系统通常具有单多肽多结构域核酸酶效应子,并且包括II型、V型和VI型。

[0140] II型CRISPR-Cas系统在组分方面被认为是最简单的。在II型CRISPR-Cas系统中,将CRISPR阵列处理成成熟的crRNA不需要存在特殊的核酸内切酶亚基,而是需要小的反式编码的crRNA(tracrRNA),其区域与阵列重复序列互补;tracrRNA与其对应的效应子核酸酶(例如Cas9)和重复序列相互作用以形成前体dsRNA结构,所述前体dsRNA结构被内源性RNA酶III切割,从而生成负载tracrRNA和crRNA两者的成熟的效应子酶。Cas II核酸酶被称为DNA核酸酶。2型效应子通常表现出由RuvC样核酸内切酶结构域组成的结构,所述RuvC样核酸内切酶结构域采用RNA酶H折叠,其中RuvC样核酸酶结构域的所述折叠内插入有不相关的HNH核酸酶结构域。RuvC样结构域负责靶(例如crRNA互补)DNA链的切割,而HNH结构域负责置换的DNA链的切割。

[0141] V型CRISPR-Cas系统的特征在于与包括RuvC样结构域的II型效应子的结构类似的核酸酶效应子(例如Cas12)结构。类似于II型,大多数(但不是全部)V型CRISPR系统使用tracrRNA将crRNA前体处理成成熟的crRNA;然而,与需要RNA酶III将crRNA前体切割成多个crRNA的II型系统不同,V型系统能够使用效应子核酸酶本身来切割crRNA前体。与II型CRISPR-Cas系统一样,V型CRISPR-Cas系统再次被称为DNA核酸酶。与II型CRISPR-Cas系统不同,一些V型酶(例如Cas12a)似乎具有强大的由双链靶序列的第一个crRNA定向切割激活的单链非特异性脱氧核糖核酸酶活性。

[0142] CRISPR-Cas系统由于其靶向性和易用性,近年来已成为首选的基因编辑技术。最常用的系统是2类、II型SpCas9和2类、V-A型Cas12a(先前的Cpf1)。特别是V-A型系统正变得越来越广泛地使用,因为据其在细胞中的报告的特异性高于其它核酸酶,具有更少或没有脱靶作用。V-A系统的优点还在于,向导RNA小(42-44个核苷酸,相比之下SpCas9大约为100nt),并且在从CRISPR阵列转录后由核酸酶自身处理,从而简化了多基因编辑的多重应用。此外,V-A系统具有交错切割位点,这可能有助于定向修复途径,如微同源性依赖性靶向整合(MITI)。

[0143] 最常用的V-A型酶需要在所选靶位点旁边有一个5'原间隔子相邻基序(PAM):针对毛螺菌科(Lachnospiraceae)细菌ND2006 LbCas12a和氨基酸球菌属(Acidaminococcus sp.)AsCas12a的5'-TTTV-3';以及针对新凶手弗朗西斯菌(Francisella novicida)FnCas12a的5'-TTV-3'。最近对直链同源物的探索揭示了具有较少限制性PAM序列的蛋白质,其在哺乳动物细胞培养例如,YTV、YYN或TTN中也具有活性。然而,这些酶并不完全涵盖V型生物多样性和靶向性,并且可能不代表所有可能的活性和PAM序列要求。在此,从大量的V型核酸酶的宏基因组中提取了数千个基因组片段。已知的V酶的多样性可能已经扩大,并且

新的系统可能已经发展成为高度靶向、紧凑和精确的基因编辑药剂。

[0144] MG酶

[0145] V-A型CRISPR系统正快速用于各种基因组编辑应用中。这些可编程核酸酶是适应性微生物免疫系统的一部分,其天然多样性在很大程度上尚未被探索。通过对从各种复杂环境中收集的宏基因组进行大规模分析来鉴定V-A型CRISPR酶的新家族,并将这些系统的代表开发到基因编辑平台中。这些系统中的大多数来自未经培养的生物体,其中一些编码同一CRISPR操纵子内的发散V型效应子。

[0146] 在一些方面,本公开提供了新V型候选物。这些候选物可以表示一种或多种新型亚型,并且可能已经鉴定出一些亚家族。这些核酸酶的长度为小于约900个氨基酸。这些新型亚型可以在与已知V型效应子相同的CRISPR基因座中找到。RuvC催化残基可能已经鉴定为新型V型候选物,并且这些新型V型候选物可能不需要tracrRNA。

[0147] 在一些方面,本公开提供了较小的V型效应子。此类效应子可以是小的推定效应子。这些效应子可以简化递送并且可以延伸治疗性应用。

[0148] 在一些方面,本公开提供了新型V型效应子。此类效应子可以是如本文所述的MG90(参见图3A-3C)。此类效应子可以是如本文所述的MG91(参见图8A-8B)。此类效应子可以是如本文所述的MG118(参见图5A-5C)。此类效应子可以是如本文所述的MG119(参见图2A-2D)。此类效应子可以是如本文所述的MG120(参见图7A-7C)。此类效应子可以是如本文所述的MG122(参见图6A-6C)。此类效应子可以是如本文所述的MG126(参见图4A-4C)。

[0149] 一方面,本公开提供了一种通过宏基因组测序发现的经工程化的核酸酶系统。在一些情况下,对样品进行宏基因组测序。在一些情况下,可以从各种环境中收集样品。此类环境可以是人微生物组、动物微生物组、高温环境、低温环境。此类环境可以包含沉积物。

[0150] 一方面,本公开提供了一种经工程化的核酸酶系统,其包括核酸内切酶。在一些情况下,所述核酸内切酶是Cas核酸内切酶。在一些情况下,所述核酸内切酶是2类V型Cas核酸内切酶。在一些情况下,所述核酸内切酶是新型亚型的2类V型Cas核酸内切酶。在一些情况下,所述核酸内切酶源自未经培养的微生物。所述核酸内切酶可以包括RuvC结构域。在一些情况下,所述经工程化的核酸酶系统包括经工程化的向导RNA。在一些情况下,所述经工程化的向导RNA被配置成与核酸内切酶形成复合物。在一些情况下,所述经工程化的向导RNA包括间隔子序列。在一些情况下,所述间隔子序列被配置成与靶核酸序列杂交。

[0151] 一方面,本公开提供了一种经工程化的核酸酶系统,其包括核酸内切酶。在一些情况下,所述核酸内切酶与SEQ ID NO:1-325、420-431、476-624或629中的任一者具有至少约70%的序列同一性。在一些情况下,所述核酸内切酶与SEQ ID NO:1-325、420-431、476-624或629中的任一者具有至少约20%、至少约25%、至少约30%、至少约35%、至少约40%、至少约45%、至少约50%、至少约55%、至少约60%、至少约65%、至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%同一性。

[0152] 在一些情况下,所述核酸内切酶包括与SEQ ID NO:1-325、420-431、476-624或629中的任一者具有至少约20%、至少约25%、至少约30%、至少约35%、至少约40%、至少约45%、至少约50%、至少约55%、至少约60%、至少约65%、至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约

95％、至少约96％、至少约97％、至少约98％或至少约99％同一性的变体。在一些情况下,所述核酸内切酶可以与SEQ ID NO:1-325、420-431、476-624或629中的任一者基本上相同。

[0153] 在一些情况下,所述经工程化的核酸酶系统包括经工程化的向导RNA。在一些情况下,所述经工程化的向导RNA被配置成与核酸内切酶形成复合物。在一些情况下,所述经工程化的向导RNA包括间隔子序列。在一些情况下,所述间隔子序列被配置成与靶核酸序列杂交。在一些情况下,所述核酸内切酶被配置成与原间隔子相邻基序(PAM)序列结合。

[0154] 在一些情况下,所述核酸内切酶不是Cpf1或Cms1核酸内切酶。

[0155] 在一些情况下,所述向导RNA包括与SEQ ID NO:410-419的前19个核苷酸或非简并核苷酸具有至少80％序列同一性的序列。在一些情况下,所述向导RNA包括与SEQ ID NO:410-419的前19个核苷酸或非简并核苷酸具有至少约20％、至少约25％、至少约30％、至少约35％、至少约40％、至少约45％、至少约50％、至少约55％、至少约60％、至少约65％、至少约70％、至少约75％、至少约80％、至少约85％、至少约90％、至少约91％、至少约92％、至少约93％、至少约94％、至少约95％、至少约96％、至少约97％、至少约98％或至少约99％同一性的序列。在一些情况下,所述向导RNA包括与SEQ ID NO:410-419的前19个核苷酸或非简并核苷酸具有至少约20％、至少约25％、至少约30％、至少约35％、至少约40％、至少约45％、至少约50％、至少约55％、至少约60％、至少约65％、至少约70％、至少约75％、至少约80％、至少约85％、至少约90％、至少约91％、至少约92％、至少约93％、至少约94％、至少约95％、至少约96％、至少约97％、至少约98％或至少约99％同一性的变体。在一些情况下,所述向导RNA包括与SEQ ID NO:410-419的前19个核苷酸或非简并核苷酸基本上相同的序列。

[0156] 在一些情况下,所述向导RNA包括与SEQ ID NO:410-419的前19个核苷酸或非简并核苷酸具有至少约20％、至少约25％、至少约30％、至少约35％、至少约40％、至少约45％、至少约50％、至少约55％、至少约60％、至少约65％、至少约70％、至少约75％、至少约80％、至少约85％、至少约90％、至少约91％、至少约92％、至少约93％、至少约94％、至少约95％、至少约96％、至少约97％、至少约98％或至少约99％同一性的序列。在一些情况下,所述核酸内切酶被配置成与经工程化的向导RNA结合。在一些情况下,所述Cas核酸内切酶被配置成与经工程化的向导RNA结合。在一些情况下,所述2类Cas核酸内切酶被配置成与经工程化的向导RNA结合。在一些情况下,所述2类V型Cas核酸内切酶被配置成与经工程化的向导RNA结合。在一些情况下,所述2类V型新型亚型Cas核酸内切酶被配置成与经工程化的向导RNA结合。

[0157] 在一些情况下,所述向导RNA包括与真核、真菌、植物、哺乳动物或人基因组多核苷酸序列互补的序列。在一些情况下,所述向导RNA包括与真核基因组多核苷酸序列互补的序列。在一些情况下,所述向导RNA包括与真菌基因组多核苷酸序列互补的序列。在一些情况下,所述向导RNA包括与植物基因组多核苷酸序列互补的序列。在一些情况下,所述向导RNA包括与哺乳动物基因组多核苷酸序列互补的序列。在一些情况下,所述向导RNA包括与人基因组多核苷酸序列互补的序列。

[0158] 在一些情况下,所述向导RNA的长度为30-250个核苷酸。在一些情况下,所述向导RNA的长度为42-44个核苷酸。在一些情况下,所述向导RNA的长度为42个核苷酸。在一些情况下,所述向导RNA的长度为43个核苷酸。在一些情况下,所述向导RNA的长度为44个核苷酸。在一些情况下,所述向导RNA的长度为85-245个核苷酸。在一些情况下,所述向导RNA的

长度为大于90个核苷酸。在一些情况下,所述向导RNA的长度为小于245个核苷酸。

[0159] 在一些情况下,所述核酸内切酶可以包括具有一个或多个核定位序列(NLS)的变体。所述NLS可以接近所述核酸内切酶的N末端或C末端。所述NLS可以被附加到SEQ ID NO:630-645中的任一者的N末端或C末端,或者被附加到与SEQ ID NO:630-645中的任一者具有至少约20%、至少约25%、至少约30%、至少约35%、至少约40%、至少约45%、至少约50%、至少约55%、至少约60%、至少约65%、至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%同一性的变体。在一些情况下,所述NLS可以包括与SEQ ID NO:630-645中的任一者基本上相同的序列。

[0160] 表1:可以与根据本公开的Cas效应子一起使用的示例NLS序列。

[0161]

来源	NLS 氨基酸序列	SEQ ID NO
SV40	PKKKRKV	630
核质蛋白二分 NLS	KRPAATKKAGQAKKKK	631
c-myc NLS	PAAKRVKLD	632
c-myc NLS	RQRRNELKRSP	633
hRNPA1 M9 NLS	NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY	634
输入蛋白-α IBB 结构域	RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKR RNV	635
肌瘤 T 蛋白	VSRKRPRP	636
肌瘤 T 蛋白	PPKKARED	637
p53	PQPKKKPL	638
小鼠 c-abl IV	SALIKKKKKMAP	639
流感病毒 NS1	DRLRR	640
流感病毒 NS1	PKQKKRK	641
肝炎病毒 δ 抗原	RKLKKKIKKL	642
小鼠 Mx1 蛋白	REKKKFLKRR	643
人聚(ADP-核糖)聚合酶	KRKGDEVDGVDEVAKKKSKK	644
类固醇激素受体(人)糖皮质激素	RKCLQAGMNLEARKTKK	645

[0162] 在一些情况下,所述经工程化的核酸酶系统进一步包括单链或双链DNA修复模板。在一些情况下,所述经工程化的核酸酶系统进一步包括单链DNA修复模板。在一些情况下,所述经工程化的核酸酶系统进一步包括双链DNA修复模板。在一些情况下,所述单链或双链DNA修复模板可以从5'至3'包括:第一同源臂,所述第一同源臂包括位于所述靶脱氧核糖核酸序列的5'的至少20个核苷酸的序列;至少10个核苷酸的合成DNA序列;以及第二同源臂,所述第二同源臂包括位于所述靶序列的3'的至少20个核苷酸的序列。

[0163] 在一些情况下,所述第一同源臂包括至少40个、至少50个、至少60个、至少70个、至少80个、至少90个、至少100个、至少110个、至少120个、至少130个、至少140个、至少150个、至少175、至少200个、至少250个、至少300个、至少400个、至少500个、至少750个或至少1000个核苷酸的序列。在一些情况下,所述第二同源臂包括至少40个、至少50个、至少60个、至少70个、至少80个、至少90个、至少100个、至少110个、至少120个、至少130个、至少140个、至少150个、至少175、至少200个、至少250个、至少300个、至少400个、至少500个、至少750个或至

少1000个核苷酸的序列。

[0164] 在一些情况下,所述第一同源臂和所述第二同源臂与原核生物的基因组序列同源。在一些情况下,所述第一同源臂和所述第二同源臂与细菌的基因组序列同源。在一些情况下,所述第一同源臂和所述第二同源臂与真菌的基因组序列同源。在一些情况下,所述第一同源臂和所述第二同源臂与真核生物的基因组序列同源。

[0165] 在一些情况下,所述经工程化的核酸酶系统进一步包括DNA修复模板。DNA修复模板可以包括双链DNA区段。双链DNA区段可以侧接一个单链DNA区段。双链DNA区段可以侧接两个单链DNA区段。在一些情况下,所述单链DNA区段与所述双链DNA区段的5'端缀合。在一些情况下,所述单链DNA区段与所述双链DNA区段的3'端缀合。

[0166] 在一些情况下,所述单链DNA区段的长度为1至15个核苷酸碱基。在一些情况下,所述单链DNA区段的长度为4至10个核苷酸碱基。在一些情况下,所述单链DNA区段的长度为4个核苷酸碱基。在一些情况下,所述单链DNA区段的长度为5个核苷酸碱基。在一些情况下,所述单链DNA区段的长度为6个核苷酸碱基。在一些情况下,所述单链DNA区段的长度为7个核苷酸碱基。在一些情况下,所述单链DNA区段的长度为8个核苷酸碱基。在一些情况下,所述单链DNA区段的长度为9个核苷酸碱基。在一些情况下,所述单链DNA区段的长度为10个核苷酸碱基。

[0167] 在一些情况下,所述单链DNA区段具有与所述间隔子序列内的序列互补的核苷酸序列。在一些情况下,所述双链DNA序列包括条形码、开放阅读框、增强子、启动子、蛋白质编码序列、miRNA编码序列、RNA编码序列或转基因。

[0168] 在一些情况下,所述经工程化的核酸酶系统进一步包括Mg^{2+}.的来源。

[0169] 在一些情况下,所述向导RNA包括发夹,所述发夹包括至少8个碱基配对的核糖核苷酸。在一些情况下,所述向导RNA包括发夹,所述发夹包括至少9个碱基配对的核糖核苷酸。在一些情况下,所述向导RNA包括发夹,所述发夹包括至少10个碱基配对的核糖核苷酸。在一些情况下,所述向导RNA包括发夹,所述发夹包括至少11个碱基配对的核糖核苷酸。在一些情况下,所述向导RNA包括发夹,所述发夹包括至少12个碱基配对的核糖核苷酸。

[0170] 在一些情况下,所述核酸内切酶包括与SEQ ID NO:1、6、15、30、151、292或319中的任一者或其变体至少70%相同的序列。在一些情况下,所述核酸内切酶包括与SEQ ID NO:1、6、15、30、151、292或319中的任一者或其变体至少75%相同的序列。在一些情况下,所述核酸内切酶包括与SEQ ID NO:1、6、15、30、151、292或319中的任一者或其变体至少80%相同的序列。在一些情况下,所述核酸内切酶包括与SEQ ID NO:1、6、15、30、151、292或319中的任一者或其变体至少85%相同的序列。在一些情况下,所述核酸内切酶包括与SEQ ID NO:1、6、15、30、151、292或319中的任一者或其变体至少90%相同的序列。在一些情况下,所述核酸内切酶包括与SEQ ID NO:1、6、15、30、151、292或319中的任一者或其变体至少95%相同的序列。

[0171] 在一些情况下,序列可以通过BLASTP、CLUSTALW、MUSCLE、MAFFT算法或CLUSTALW算法使用史密斯-沃特曼同源性搜索算法参数来确定。序列同一性可以通过所述BLASTP同源性搜索算法使用字长(W)为3、期望值(E)为10的参数以及BLOSUM62评分矩阵将空位罚分设置为存在11,扩展1并且使用条件组成评分矩阵调整来确定。

[0172] 一方面,本公开提供了一种经工程化的向导RNA,其包括DNA靶向区段。在一些情况

下,所述DNA靶向区段包括与靶序列互补的核苷酸序列。在一些情况下,所述靶序列在靶DNA分子中。在一些情况下,所述经工程化的向导RNA包括蛋白质结合区段。在一些情况下,所述蛋白质结合区段包括两个互补的核苷酸延伸段。在一些情况下,两个互补的核苷酸延伸段杂交以形成双链RNA(dsRNA)双链体。在一些情况下,两个互补的核苷酸延伸段用中间核苷酸彼此共价连接。在一些情况下,所述经工程化的向导核糖核酸多核苷酸能够与核酸内切酶形成复合物。在一些情况下,所述核酸内切酶与SEQ ID NO:1-325、420-431、476-624或629中的任一者具有至少约20%、至少约25%、至少约30%、至少约35%、至少约40%、至少约45%、至少约50%、至少约55%、至少约60%、至少约65%、至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%同一性。在一些情况下,所述复合物靶向靶DNA分子的靶序列。在一些情况下,所述DNA靶向区段定位在所述两个互补的核苷酸延伸段中的两者的3'。

[0173] 在一些情况下,所述双链RNA(dsRNA)双链体包括至少8个核糖核苷酸。在一些情况下,所述双链RNA(dsRNA)双链体包括至少9个核糖核苷酸。在一些情况下,所述双链RNA(dsRNA)双链体包括至少10个核糖核苷酸。在一些情况下,所述双链RNA(dsRNA)双链体包括至少11个核糖核苷酸。在一些情况下,所述双链RNA(dsRNA)双链体包括至少12个核糖核苷酸。

[0174] 在一些情况下,所述脱氧核糖核酸多核苷酸编码所述经工程化的向导核糖核酸多核苷酸。

[0175] 一方面,本公开提供了一种核酸,其包括经工程化的核酸序列。在一些情况下,所述经工程化的核酸序列被优化以在生物体中表达。在一些情况下,所述核酸编码核酸内切酶。在一些情况下,所述核酸内切酶是Cas核酸内切酶。在一些情况下,所述核酸内切酶是2类核酸内切酶。在一些情况下,所述核酸内切酶是2类V型Cas核酸内切酶。在一些情况下,所述核酸内切酶是2类V型新型亚型Cas核酸内切酶。在一些情况下,所述核酸内切酶源自未经培养的微生物。在一些情况下,所述生物体不是未经培养的生物体。

[0176] 在一些情况下,所述核酸内切酶包括与SEQ ID NO:1-325、420-431、476-624或629中的任一者具有至少约20%、至少约25%、至少约30%、至少约35%、至少约40%、至少约45%、至少约50%、至少约55%、至少约60%、至少约65%、至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的变体。

[0177] 在一些情况下,所述核酸内切酶可以包括具有一个或多个核定位序列(NLS)的变体。所述NLS可以接近所述核酸内切酶的N末端或C末端。所述NLS可以被附加到SEQ ID NO:630-645中的任一者的N末端或C末端,或者被附加到与SEQ ID NO:630-645中的任一者具有至少约20%、至少约25%、至少约30%、至少约35%、至少约40%、至少约45%、至少约50%、至少约55%、至少约60%、至少约65%、至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的变体。

[0178] 在一些情况下,所述生物体是原核细胞。在一些情况下,所述生物体是细菌。在一些情况下,所述生物体是真核生物。在一些情况下,所述生物体是真菌。在一些情况下,所述

生物体是植物。在一些情况下,所述生物体是哺乳动物。在一些情况下,所述生物体是啮齿动物。在一些情况下,所述生物体是人。

[0179] 一方面,本公开提供了一种经工程化的载体。在一些情况下,所述经工程化的载体包括编码核酸内切酶的核酸序列。在一些情况下,所述核酸内切酶是Cas核酸内切酶。在一些情况下,所述核酸内切酶是2类Cas核酸内切酶。在一些情况下,所述核酸内切酶是2类V型Cas核酸内切酶。在一些情况下,所述核酸内切酶是2类V型新型亚型Cas核酸内切酶。在一些情况下,所述核酸内切酶源自未经培养的微生物。

[0180] 在一些情况下,所述经工程化的载体包括本文所述的核酸。在一些情况下,本文所描述的核酸是本文所描述的脱氧核糖核酸多核苷酸。在一些情况下,所述载体是质粒、微环、CELiD、腺相关病毒(AAV)源性病毒体、慢病毒或慢病毒。

[0181] 一方面,本公开提供了一种细胞,其包括本文所述的载体。

[0182] 一方面,本公开提供了一种制备核酸内切酶的方法。在一些情况下,所述方法包括培养细胞。

[0183] 一方面,本公开提供了一种用于结合、切割、标记或修饰双链脱氧核糖核酸多核苷酸的方法。所述方法可以包括使双链脱氧核糖核酸多核苷酸与核酸内切酶接触。在一些情况下,所述核酸内切酶是Cas核酸内切酶。在一些情况下,所述核酸内切酶是2类Cas核酸内切酶。在一些情况下,所述核酸内切酶是2类V型Cas核酸内切酶。在一些情况下,所述核酸内切酶是2类V型新型亚型Cas核酸内切酶。在一些情况下,所述核酸内切酶与经工程化的向导RNA复合。在一些情况下,所述经工程化的向导RNA被配置成与核酸内切酶结合。在一些情况下,所述经工程化的向导RNA被配置成与双链脱氧核糖核酸多核苷酸结合。在一些情况下,所述经工程化的向导RNA被配置成与核酸内切酶结合并且与双链脱氧核糖核酸多核苷酸结合。在一些情况下,所述双链脱氧核糖核酸多核苷酸包括原间隔子相邻基序(PAM)。

[0184] 在一些情况下,所述双链脱氧核糖核酸多核苷酸包括第一链和第二链,所述第一链包括与所述经工程化的向导RNA的序列互补的序列,所述第二链包括所述PAM。在一些情况下,所述PAM与和所述经工程化的向导RNA的所述序列互补的所述序列的5'端直接相邻。在一些情况下,所述核酸内切酶不是Cpf1核酸内切酶或Cms1核酸内切酶。在一些情况下,所述核酸内切酶源自未经培养的微生物。在一些情况下,所述双链脱氧核糖核酸多核苷酸是真核生物、植物、真菌、哺乳动物、啮齿动物或人双链脱氧核糖核酸多核苷酸。

[0185] 一方面,本公开提供了一种修饰靶核酸基因座的方法。所述方法可以包括将本文所述的经工程化的核酸酶系统递送到靶核酸基因座。在一些情况下,所述核酸内切酶被配置成与经工程化的向导核糖核酸结构形成复合物。在一些情况下,所述复合物被配置成使得在所述复合物与所述靶核酸基因座结合时,所述复合物修饰所述靶核酸基因座。

[0186] 在一些情况下,修饰所述靶核酸基因座包括结合、切口、切割或标记所述靶核酸基因座。在一些情况下,所述靶核酸基因座包括脱氧核糖核酸(DNA)或核糖核酸(RNA)。在一些情况下,所述靶核酸包括基因组DNA、病毒DNA、病毒RNA或细菌DNA。在一些情况下,所述靶核酸基因座是在体外的。在一些情况下,所述靶核酸基因座在细胞内。在一些情况下,所述细胞是原核细胞、细菌细胞、真核细胞、真菌细胞、植物细胞、动物细胞、哺乳动物细胞、啮齿动物细胞、灵长类动物细胞或人细胞。

[0187] 在一些情况下,向所述靶核酸基因座递送所述经工程化的核酸酶系统包括递送本

文所述的核酸或本文所述的载体。在一些情况下,向所述靶核酸基因座递送经工程化的核酸酶系统包括递送包括编码所述核酸内切酶的开放阅读框的核酸。在一些情况下,所述核酸包括启动子。在一些情况下,编码核酸内切酶的开放阅读框与启动子可操作地连接。

[0188] 在一些情况下,向所述靶核酸基因座递送所述经工程化的核酸酶系统包括递送包含编码所述核酸内切酶的开放阅读框的加帽mRNA。在一些情况下,向所述靶核酸基因座递送所述经工程化的核酸酶系统包括递送经翻译的多肽。在一些情况下,向所述靶核酸基因座递送所述经工程化的核酸酶系统包括递送编码与核糖核酸(RNA)pol III启动子可操作地连接的经工程化的向导RNA的脱氧核糖核酸(DNA)。

[0189] 在一些情况下,所述核酸内切酶诱导所述靶基因座处或接近所述靶基因座的单链断裂或双链断裂。在一些情况下,所述核酸内切酶诱导所述靶基因座内或位于所述靶基因座的3'的交错单链断裂。

[0190] 在一些情况下,效应子重复序列基序用于为MG核酸酶的向导设计提供信息。例如,V型系统中的经处理的gRNA包括CRISPR重复序列的最后20-22个核苷酸。此序列可以合成为crRNA(连同间隔子),并与合成的核酸酶一起在体外进行测试,以在可能的靶标文库上进行切割。使用这种方法,可以确定PAM。在一些情况下,V型酶可以使用"通用"gRNA。在一些情况下,V型酶可能需要独特gRNA。

[0191] 本公开的系统可以用于各种应用,例如核酸编辑(例如基因编辑)、与核酸分子结合(例如序列特异性结合)。此类系统可以用于例如解决(例如,去除或替换)可能引起受试者的疾病的遗传突变,使基因灭活以便确定其在细胞中的功能,作为检测致病遗传元件的诊断工具(例如通过切割逆转录病毒RNA或编码致病突变的扩增DNA序列),作为灭活酶与探针结合以靶向和检测特定核苷酸序列(例如编码细菌抗生素抗性的序列),通过靶向病毒基因组使病毒灭活或无法感染宿主细胞,添加基因或修改代谢途径来对生物体进行工程化以产生有价值的小分子、大分子或次级代谢物,建立用于进化选择的基因驱动元件,作为生物传感器检测外来小分子和核苷酸对细胞的干扰。

[0192] 实例

[0193] 根据IUPAC惯例,在各个实例中使用以下缩写:

[0194] A=腺嘌呤

[0195] C=胞嘧啶

[0196] G=鸟嘌呤

[0197] T=胸腺嘧啶

[0198] R=腺嘌呤或鸟嘌呤

[0199] Y=胞嘧啶或胸腺嘧啶

[0200] S=鸟嘌呤或胞嘧啶

[0201] W=腺嘌呤或胸腺嘧啶

[0202] K=鸟嘌呤或胸腺嘧啶

[0203] M=腺嘌呤或胞嘧啶

[0204] B=C、G或T

[0205] D=A、G或T

[0206] H=A、C或T;

[0207] V＝A、C或G

[0208] 实例1-新蛋白质的宏基因组分析方法

[0209] 从沉积物、土壤和动物收集宏基因组样品。用Zymobiomics DNA微型制备试剂盒提取脱氧核糖核酸(DNA)并在Illumina HiSeq®2500上测序。在产权所有者同意的情况下收集样品。来自公共来源的另外的原始序列数据包含动物微生物群、沉积物、土壤、温泉、深海热泉、海洋、泥炭沼泽、永久冻土和污水序列。使用基于包含II类V型Cas效应蛋白在内的已知Cas蛋白序列生成的隐马尔可夫模型(Hidden Markov Model)检索宏基因组序列数据以鉴定新的Cas效应子。通过搜索鉴定的新型效应蛋白与已知的蛋白质对比以鉴定潜在的活性位点。此宏基因组工作流导致本文所述的MG90、MG91A、MG91B、MG91C、MG118、MG119、MG120、MG122和MG126家族的描绘。

[0210] 实例2-CRISPR系统的MG90、MG91A、MG91B、MG91C、MG118、MG119、MG120、MG122和MG126家族的发现

[0211] 对来自实例1的宏基因组分析的数据的分析揭示了先前未描述的推定CRISPR系统的新簇,其包括9个家族(MG90、MG91A、MG91B、MG91C、MG118、MG119、MG120、MG122和MG126)。这些新酶及其示例亚结构域的对应蛋白质和核酸序列如SEQ ID NO:1-325、420-431、476-624或629所示。

[0212] 实例3-用于转录和翻译的模板DNA

[0213] 在具有T7启动子的质粒中排序有所有MG、VU和CasPhi核酸酶的经大肠杆菌密码子优化的序列(特维斯特生物科技公司(Twist Biosciences))。通过PCR从质粒扩增线性模板以包含T7和核酸酶序列。从由T7启动子、天然重复、通用间隔子和天然重复组成的序列扩增最小阵列线性模板,侧接用于扩增的衔接子序列。通用间隔子与8N靶文库中的间隔子匹配,其中存在与间隔子相邻的8N个混合碱基以用于PAM确定。从宏基因组重叠群鉴定ORF或CRISPR阵列附近的三个基因间序列,并且将其排序为具有用于扩增的侧接衔接子序列的gBlock(整合DNA技术)。

[0214] 实例4-crRNA、最小阵列和sgRNA的体外转录

[0215] 使用HiScribe™T7高收率RNA合成试剂盒通过体外转录产生RNA,并使用Monarch®RNA纯化试剂盒(纽英伦生物技术公司(New England Biolabs Inc.))纯化。T7转录的模板不同。对于crRNA,DNA寡核苷酸被设计成具有T7启动子、修剪的天然重复序列和通用间隔子。对于最小阵列,使用与上述相同的模板。对于sgRNA,DNA超聚体被设计成具有T7启动子、修剪的tracrRNA、GAAA四环、修剪的天然重复序列和通用间隔子。用衔接子引物扩增最小阵列模板。将crRNA和sgRNA模板作为反向补体排序,并且在95℃下用具有T7启动子序列的引物在1X IDT双链体缓冲液中退火两分钟,随后以0.1℃/秒冷却至22℃,以产生适于转录的混合ds/ssDNA底物。在转录之后,但在清洁之前,将每个反应用DNA酶I处理并在37℃下温育15分钟。经由RNA TapeStation或经由变性尿素PAGE凝胶验证所有转录产物的收率和纯度。

[0216] 实例5-TXTL表达

[0217] 使用myTXTL®Sigma 70主混合物试剂盒(Arbor生物科学公司(Arbor Biosciences))在转录翻译反应混合物中表达核酸酶、基因间序列和最小阵列。最终反应混

合物含有5nM核酸酶DNA模板、12nM基因间DNA模板、15nM最小阵列DNA模板、0.1nM pTXTL-P70a-T7rnap和1X的myTXTL®Sigma 70主混合物。将反应在29℃下温育16小时,然后在4℃下储存。

[0218] 实例6-PURExpress表达

[0219] 将10nM的核酸酶PCR模板用PURExpress®体外蛋白质合成试剂盒(纽英伦生物技术公司)在37℃下表达3小时以用经体外转录的RNA进行切割。这些反应用于按照与切割反应部分中所述的相同程序测试具有50nM sgRNA或最小阵列RNA的体外切割。

[0220] 实例7-大肠杆菌表达

[0221] 将编码效应子的质粒、来自基因组重叠群的基因间序列、天然重复序列和具有T7启动子的通用间隔子序列转化成BL21 DE3或T7表达lysY/Iq,并在37℃下在补充有100μg/mL的氨苄西林(ampicillin)的60mL极品肉汤培养基中培养。在培养物达到0.5的OD_{600nm}并在16℃下温育过夜后,用0.4mM IPTG诱导表达。将25mL的细胞通过离心沉淀,并重悬于1.5mL的裂解缓冲液(20mM Tris-HCl、500mM NaCl、1mM TCEP、5%甘油、pH为7.5的10mM MgCl2和Pierce蛋白酶抑制剂(Thermo Scientific™))中。然后通过超声处理使细胞裂解。通过离心分离上清液和细胞碎片。

[0222] 实例8-切割反应

[0223] 质粒文库DNA切割反应通过在37℃下混合5nM的靶文库、5倍稀释的TXTL或PURExpress表达、10nM Tris-HCl、10nM $MgCl_2$和100mM NaCl进行2小时。对于与大肠杆菌表达的反应,添加10μL的澄清裂解物。停止反应并用HighPrep™PCR纯化微球(MAGBIO基因组学公司(MAGBIO Genomics,Inc.))清洁,并在pH为8.0的Tris EDTA缓冲液中洗脱。将3nM的切割产物端用3.33μM dNTP、1X T4 DNA连接酶缓冲液和0.167U/μL的Klenow片段(纽英伦生物技术公司)在25℃下钝化15分钟。将1.5nM的切割产物用150nM衔接子、1X T4 DNA连接酶缓冲液(纽英伦生物技术公司)、20U/μL的T4 DNA连接酶(纽英伦生物技术公司)在室温下钝化20分钟。将连接产物用NGS引物通过PCR扩增,并且通过NGS测序以得到PAM。MG119-2的体外活性描绘于图9,而MG119-2的PAM确定描绘于图10。

[0224] 实例9-从TXTL和大肠杆菌裂解物中制备基因间富集的RNAseq文库

[0225] 从TXTL提取RNA,并且进行细胞裂解物表达,随后是Quick-RNA™微量制备型试剂盒(Zymo研究公司(Zymo Research)),并在30-50μL的水中洗脱。在Nanodrop、Tapestation和Qubit上测量转录物的总浓度。

[0226] 使用用于因美纳(Illumina)的NEBNext小型RNA文库制备套件(纽英伦生物技术公司)制备来自每个样品的100ng-1ug的总RNA以进行RNA测序。将150-300bp之间的扩增子通过Tapestation和Qubit定量,并合并到4nM的最终浓度。将12.5pM的最终浓度上样到MiSeq V3试剂盒中,并在Miseq系统(因美纳公司)中测序176个总循环。RNAseq读段被用于鉴定基因的tracr序列。

[0227] 实例10-预测的RNA折叠

[0228] 使用Andronescu 2007的方法在37℃下计算活性单个RNA序列的预测RNA折叠。碱基的阴影对应于所述碱基的碱基配对概率。

[0229] 实例11-体外切割效率(预测)

[0230] 在T7诱导型启动子下在大肠杆菌蛋白酶缺陷型B菌株中表达蛋白质,使用超声处

理裂解细胞,并且使用HisTrap FF(通用生命科学公司(GE Lifescience))Ni-NTA亲和色谱法在AKTA Avant FPLC(通用生命科学公司)上纯化所关注的经His标记的蛋白质。使用在SDS-PAGE和InstantBlue超高速(西格玛-奥德里奇公司(Sigma-Aldrich))经考马斯染色的丙烯酰胺凝胶(伯乐公司(Bio-Rad))上解析的蛋白质带的ImageLab软件(伯乐公司)中的密度测定法确定纯度。将蛋白质在由pH为7.5的50mM Tris-HCl、300mM NaCl、1mM TCEP、5%甘油组成的储存缓冲液中脱盐;并在-80℃下储存。

[0231] 构建含有间隔子序列和通过NGS确定的PAM的靶DNA。在PAM中的简并碱基的情况下,选择单个代表性PAM进行测试。靶DNA是通过PCR扩增衍生自质粒的2200bp的线性DNA。PAM和间隔子位于距一端700bp处。成功的切割产生700和1500bp的片段。

[0232] 将靶DNA、经体外转录的单RNA和经纯化的重组蛋白在含有过量蛋白质和RNA的切割缓冲液(10mM Tris、100mM NaCl、10mM MgCl2)中合并,并温育5'至3小时,通常为1小时。通过添加RNA酶A并在60°下温育来停止反应。将反应在1.2% TAE琼脂糖凝胶上解析,并且在ImageLab软件中定量经切割的靶DNA的分数。

[0233] 实例12-大肠杆菌中的活性(预测)

[0234] 为了测试细菌细胞中的核酸酶活性,用含有具有对所关注的酶具有特异性的靶间隔子和对应PAM序列的基因组序列构建菌株。然后用所关注的核酸酶转化经工程化的菌株,并且随后使转化子具有化学能力,并且用50ng的对(靶上)靶序列具有特异性或对(脱靶)靶标不具有特异性的单向导进行转化。在热冲击之后,在37℃下在SOC中将转化回收2小时,并且通过在诱导培养基上生长的5倍稀释系列来确定核酸酶效率。稀释系列一式三份地定量菌落。

[0235] 实例13-哺乳动物细胞中的活性(预测)

[0236] 为了示出在哺乳动物细胞中的靶向和切割活性,将蛋白质序列克隆到2个哺乳动物表达载体中,一个具有C末端SV40 NLS和2A-GFP标签以及一个没有GFP标签和2个NLS序列(一个在N末端上并且一个在C末端上)。也可以使用替代的NLS序列。蛋白质的DNA序列可以是天然序列、经大肠杆菌密码子优化的序列或经哺乳动物密码子优化的序列。具有所关注的基因靶标的单向导RNA序列也克隆到哺乳动物表达载体中。将两种质粒共转染到HEK293T细胞中。将表达质粒和sgRNA靶向质粒共转染到HEK293T细胞中72小时后,提取DNA并用于制备NGS文库。在靶位点的测序中通过插入缺失测量NHEJ百分比,以证明酶在哺乳动物细胞中的靶向效率。选择至少10个不同的靶位点来测试每种蛋白质的活性。

[0237] 实例14-MG119家族中紧凑V型核酸酶的表征

[0238] MG119家族中新型紧凑V型核酸酶的计算机鉴定

[0239] 基于同源搜索,在紧凑V型核酸酶的MG119家族中发现与核酸酶序列相关的预测蛋白质。使用HMMER软件(http://hmmer.org/)执行搜索。如果V型核酸酶序列命中符合以下标准,则将其保留:(i)hmmsearch e值≤10^{-5};(ii)编码核酸酶的基因在距CRISPR阵列1kb以内,以及(iii)氨基酸序列长度的范围为350至700aa。MMSeqs2(https://github.com/soedinglab/MMseqs2)用于以100%氨基酸同一性聚集序列,具有覆盖模式1和靶序列的80%覆盖度(参数--cov-模式1-c 0.8--min-seq-id 1.0)。选择序列代表以使用MAFFT(https://mafft.cbrc.jp/alignment/software/)和Needleman-Wunsch算法来构建多序列比对以用于全局比对,并且使用FastTree(https://doi.org/10.1371/

journal.pone.0009490)来构建系统发育树。仔细检查系统发育树上的单个分支,包含核酸酶基因的基因组背景,导致MG119家族中的若干新型紧凑V型核酸酶序列的鉴定(SEQ ID NO:476-624和629)。

[0240] 体外表征以鉴定推定tracrRNA

[0241] 为了鉴定推定tracrRNA序列(例如,核酸酶MG119-2),使用myTXTL®Sigma 70主混合物试剂盒(Arbor生物科学公司)在转录翻译反应混合物中表达相邻的基因间序列和最小阵列。最终反应混合物含有5nM核酸酶DNA模板、12nM基因间DNA模板、15nM最小阵列DNA模板、0.1nM pTXTL-P70a-T7rnap和1X的myTXTL®Sigma 70主混合物。将反应在29℃下温育16小时,然后在4℃下储存。

[0242] 经由体外切割反应测试核糖核蛋白复合物。质粒DNA文库切割反应通过在37℃下混合代表所有可能的8N PAM的5nM的靶质粒DNA文库、5倍稀释的TXTL表达、10nM Tris-HCl、10nM $MgCl_2$和100mM NaCl进行2小时。停止反应并用HighPrep™PCR纯化微球(MAGBIO基因组学公司)清洁,并在pH为8.0的Tris EDTA缓冲液中洗脱。

[0243] 为了获得PAM序列,将3nM的切割产物端用3.33μM dNTP、1X T4 DNA连接酶缓冲液和0.167U/μL的Klenow片段(纽英伦生物技术公司)在25℃下钝化15分钟。将1.5nM的切割产物用150nM衔接子、1X T4 DNA连接酶缓冲液(纽英伦生物技术公司)和20U/μL的T4 DNA连接酶(纽英伦生物技术公司)在室温下钝化20分钟。将连接产物用NGS引物通过PCR扩增,并且通过NGS测序。

[0244] 为了获得tracrRNA和crRNA的序列,按照Quick-RNA™最小制备试剂盒(Zymo研究公司)从TXTL裂解液中提取RNA,并在30-50μL的水中洗脱。使用用于因美纳的NEBNext小型RNA文库制备套件(纽英伦生物技术公司)制备来自每个样品的100ng-1μg的总RNA以进行RNA测序。将150-300bp之间的扩增子通过Tapestation和Qubit定量,并合并到4nM的最终浓度。将12.5pM的最终浓度上样到MiSeq V3试剂盒中,并在Miseq系统(因美纳公司)中测序176个总循环。RNAseq读段被用于通过映射回原始序列来鉴定基因的tracr序列。

[0245] 计算机搜索新的tracrRNA序列

[0246] 为了鉴定含有潜在tracrRNA的另外的非编码区,活性tracrRNA的序列被映射到含有相同核酸酶家族中的核酸酶的其它重叠群(例如,MG119-1和MG119-3)。新鉴定的序列用于产生协方差模型以预测另外的tracrRNA。协方差模型由活性和预测的tracrRNA序列的多序列比对(MSA)构建。MSA的次级结构用RNAalifold(维也纳包装公司(Vienna Package))获得,并且协方差模型用推断包装(http://eddylab.org/infernal/)构建。使用具有推断命令‘cmsearch’的协方差模型搜索含有候选物核酸酶的其它重叠群。在体外测试TracrRNA候选物(见下文),并且在迭代过程中,使用来自活性候选物的序列来改进协方差模型,并在与其它核酸酶候选物相关的基因间区域中寻找另外的tracrRNA。

[0247] sgRNA设计

[0248] 从协方差模型获得的预测tracrRNA及其相关的CRISPR重复序列被修饰以产生sgRNA(图11A)如下:对预测的tracrRNA序列的3’端和重复序列的5’端进行修剪,并且然后用GAAA四环连接。

[0249] 体外切割反应以确认核酸酶活性并进行PAM测定

[0250] 将5nM的经核酸酶扩增的DNA模板和经25nM sgRNA扩增的DNA模板(包含表2中列出的间隔子序列之一)用PURExpress®体外蛋白质合成试剂盒(纽英伦生物技术公司)在37℃下表达3小时。质粒文库DNA切割反应通过在37℃下混合代表所有可能的8N PAM的5nM的靶文库、5倍稀释的PURExpress表达、pH为7.9的10mM Tris-HCl、10mM $MgCl_2$、100μg/mL BSA和50mM NaCl(NEB 2.1缓冲液,NEB公司(NEB Inc.))进行2小时。停止反应并用HighPrep™PCR纯化微球(MAGBIO基因组学公司)清洁,并在pH为8.0的Tris EDTA缓冲液中洗脱。将3nM的切割产物端用3.33μMdNTP、1X T4 DNA连接酶缓冲液和0.167U/μL的Klenow片段(纽英伦生物技术公司)在25℃下钝化15分钟。将1.5nM的切割产物用150nM衔接子、1X T4 DNA连接酶缓冲液(纽英伦生物技术公司)和20U/μL的T4 DNA连接酶(纽英伦生物技术公司)在室温下钝化20分钟。将连接产物用NGS引物通过PCR扩增,并且通过NGS测序以得到PAM。成功切割PAM文库的活性蛋白在琼脂糖凝胶中产生约188或205bp的条带,具体取决于sgRNA中编码的靶位点(图11B)。

[0251] 表2:用于测试向导的间隔子序列

[0252]

代码	序列
U67间隔子	GTCGAGGCTTGCGACGTGGT
U40间隔子	TGGAGATATCTTGAACCTTG

[0253] 通过MG119核酸酶识别的PAM显示为用Seqlog maker制成的序列标记(图12)。表3列出了与U40间隔子互补的原间隔子序列的靶链上的优选切割位置。

[0254] 表3:原间隔子序列中的MG119核酸酶优选切割位点

[0255]

核酸酶	**sgRNA**	**切割位点**
119-1	MG119-1_sgRNA1	20 & 23
119-2	MG119-2_sgRNA1_突变体 1	22
119-3	MG119-3_sgRNA1_突变体 1	22-23
119-4	MG119-4_sgRNA1	22-23
119-10	MG119-10_sgRNA1	22-23

[0256]

119-19	MG119-19_sgRNA1	23
119-27	MG119-27_sgRNA2_突变体 2	22-23
119-28	MG119-28_sgRNA2	22-23
119-32	MG119-32_sgRNA1	23
119-54	MG119-54_sgRNA1	22
119-64	MG119-64_sgRNA2	20
119-72	MG119-72_sgRNA1	23
119-83	MG119-83_sgRNA1	23
119-97	MG119-97_sgRNA1_突变体 1	22
119-109	MG119-109_sgRNA1	24-25
119-118	MG119-118_sgRNA1_突变体 2	23
119-121	MG119-121_sgRNA1_突变体 1	20 & 22
119-125	MG119-125_sgRNA1	22-23
119-128	MG119-128_sgRNA2_突变体 1	22
119-129	MG119-129_sgRNA1_突变体 1	22-23
119-133	MG119-133_sgRNA1_突变体 1	22
119-136	MG119-136_sgRNA1_突变体 2	23
119-137	MG119-137_sgRNA1	22-23

[0257] 蛋白质表达和纯化

[0258] 分离纯蛋白和功能蛋白对于生化特性的广泛体外分析和机制研究至关重要。优化MG119候选物的表达和纯化以获得足够量和质量的蛋白质以用于此类表征。所有构建体均在大肠杆菌(NEBExpress I^qCompetent大肠杆菌,NEB C3037I)中表达。构建体在pMGB表达载体(MBP融合)、pMGBΔ表达载体(无融合蛋白)或两者中表达。

[0259] 蛋白质表达

[0260] pMGB和pMGBΔ构建体的蛋白质表达方案是相同的。将培养物在37℃下在具有100μg/L羧苄西林(Carbenicillin)的2xYT培养基(1.6％胰蛋白胨、1％酵母提取物、0.5％NaCl)或TB培养基(Teknova T0690)中生长。在OD600≈0.8-1.2下,将培养物用0.5mM IPTG(GoldBio I2481)诱导并在18℃下温育过夜或在24℃下温育4-6小时,具体取决于构建体。然后通过以6,000x g离心10分钟收获培养物,并将沉淀物重悬于Nickel_A缓冲液(pH为7.5的50mM Tris,750mM NaCl,10mM $MgCl_2$,20mM咪唑,0.5mM EDTA,5％甘油,0.5mM TCEP)+蛋白酶抑制剂(Pierce蛋白酶抑制剂片剂,无EDTA,赛默飞世尔公司(ThermoFisher)A32965)中并在-80℃下储存。

[0261] 蛋白质纯化-pMGBΔ表达载体

[0262] 在此载体中表达的蛋白质具有以下序列结构:6xHis-(GS)2-PSP-核质蛋白二分

NLS-(GGS)1-(GS)1-MG119-X-(GGS)3-SV40 NLS(表5)。在此载体中表达的蛋白质表示为MG119-XΔ。解冻细胞沉淀物,并将体积用Cf＝0.5％n-辛基-β-D-葡糖苷洗涤剂(P212121,CI-00234)补充至120mL。使用15秒开启/45秒关闭循环在冰水浴中以75％振幅对样品进行超声处理,总处理时间为3分钟。将裂解物通过以30,000x g离心25分钟来澄清,并且将上清液分批结合到5mL Ni-NTA树脂(HisPur Ni-NTA树脂,赛默飞世尔公司88223)≥20分钟。将样品上样到重力柱上并用30CV Nickel_A缓冲液洗涤,然后在4CV Nickel_B缓冲液(Nickel_A缓冲液+250mM咪唑)中洗脱,然后在50kDa MWCO浓缩器(Amicon Ultra-15,密理博西格玛公司(MilliporeSigma)UFC9050)中浓缩。在整个纯化过程中采集样品,并在SDS-PAGE蛋白质凝胶(伯乐公司#4568126)上运行,在UV激活5分钟后在无染色通道中在ChemiDoc上成像(图13A)。然后将ΔMBP构建体上样到S200i 10/300GL柱(Cytiva 28-9909-44)上并运行到缓冲液Nickel_A中(图13B)。合并峰级分并在50kDa MWCO浓缩器中浓缩。pMGBΔ载体中表达的蛋白质的纯化通常每升表达培养物产生25-125nmol蛋白质(图13F)。

[0263] 蛋白质纯化-pMGB表达载体

[0264] 在此载体中表达的蛋白质具有以下序列结构:6xHis-(GS)1-MBP-(GS)1-TEV-核质蛋白二分NLS-(GGGGS)3-(GS)1-MG119-X-(GGS)3-SV40 NLS(表5)。MBP融合构建体通过裂解、澄清、亲和纯化和Nickel_B中洗脱与pMGBΔ蛋白相同地纯化(图13C)。在50kDa MWCO浓缩器中蛋白质浓缩之后,将TEV蛋白酶(GenScript Z03030)添加到每个样品(Cf＝1UI/μL)中并在4℃下温育过夜,轻轻地翻滚(end-over-end)旋转。将样品离心(21,000x g,4℃,10分钟)以沉淀聚集体,并且然后将上清液在4℃下分批结合到3mL淀粉树脂(NEB E8021L)中30分钟,然后上样到重力柱上。收集流出物并将其在50kDa MWCO浓缩器中浓缩(图13D)。再次,将样品离心(21,000x g,4℃,10分钟)以沉淀聚集物,然后加载到S200i 10/300GL柱上并运行到Nickel_A缓冲液中(图13E)。合并峰级分并在50kDa MWCO浓缩器中浓缩。在整个纯化过程中采集样品,并在SDS-PAGE蛋白质凝胶(伯乐公司#4568126)上运行,在UV激活5分钟后在无染色通道中在ChemiDoc上成像(图13D)。

[0265] 从pMGB和pMGBΔ表达载体两者纯化选定的几个MG119候选物。最终蛋白质收率的比较(相对于初始表达培养物体积归一化)显示出来自pMGBΔ的载体表达收率较高的趋势(图13E)。pMGBΔ载体中表达的蛋白质的纯化通常每升表达培养物产生2-15nmol蛋白质(图13E)。表4中示出了蛋白质纯化收率。

[0266] 表4:蛋白质纯化收率

[0267]

核酸酶	表达载体	表达培养基	收率（nmol）
MG119-1	pMGB	TB	8.8
MG119-1	pMGBΔ	TB	105.2
MG119-2	pMGB	2xYT	5.0
MG119-2	pMGBΔ	2xYT	95.0
MG119-3	pMGB	2xYT	7.6
MG119-3	pMGBΔ	2xYT	78.1
MG119-27	pMGB	2xYT	11.9
MG119-28	pMGB	2xYT	11.6
MG119-28	pMGBΔ	2xYT	102.2
MG119-32	pMGB	2xYT	4.3
MG119-54	pMGB	2xYT	5.6
MG119-64	pMGB	2xYT	2.1
MG119-97	pMGB	2xYT	4.3
MG119-109	pMGB	2xYT	4.1
MG119-121	pMGB	2xYT	8.2
MG119-128	pMGB	2xYT	9.9
MG119-128	pMGBΔ	2xYT	37.0

[0268]

核酸酶	表达载体	表达培养基	收率（nmol）
MG119-129	pMGB	2xYT	2.8
MG119-136	pMGB	2xYT	14.3
MG119-137	pMGB	2xYT	10.4

[0269] 表5:序列元素术语表

[0270]

元素名称	元素氨基酸序列
6xHis	HHHHHH
$(GS)_n$	GS
$(GGS)_n$	GGS
$(GGGGS)_n$	GGGGS
PSP	LEVQFQGP
TEV	ENLYFQG
核质蛋白二分NLS	KRPAATKKAGQAKKKK
SV40NLS	PKKKRKV

[0271] 纯化蛋白的体外切割效率

[0272] 在线性DNA底物切割测定中测定蛋白质等分试样的活性分数。将效应蛋白与2倍摩尔过量的sgRNA在室温下预温育20分钟以形成核糖核蛋白复合物(RNP)。使用25nM DNA底物和来自0.25X至10X摩尔过量的底物的RNP滴定来设置反应。反应缓冲液组成为pH为7.5的10mM Tris、10mM $MgCl_2$和100mM NaCl。DNA底物长522bp。成功的切割产生172和350bp的片段。将反应在37℃下温育60分钟,然后在75℃下温育10分钟。将RNase(NEB T3018)添加到每个反应中(Cf=0.33μg/μL),并将样品在37℃下温育10分钟。每个反应中添加蛋白酶K(NEB P8107)(Cf=60单位/mL),并将样品在55℃下温育15分钟。然后将每个反应全部在具有GelGreen染料(Biotium,#41005)的1.5%琼脂糖凝胶上进行(图14A)并在GelGreen通道中在ChemiDoc上成像。使用伯乐公司的图像实验室软件(6.1.0版构建体7)通过密度测定分析计算每个泳道的切割底物百分比。活性分数由切割线性范围的斜率确定(图14B)。

[0273] 用纯化的蛋白质对纯化的Hepa1-6基因组DNA进行体外切割

[0274] 为了评估纯化的小鼠Hepa1-6基因组DNA(gDNA)的切割,将小鼠白蛋白基因靶向在内含子1(表6)处。按照PurelinkTM基因组DNA迷你试剂盒(英杰公司(Invitrogen))从具有800万个细胞的Hepa1-6细胞沉淀物中提取gDNA,并在pH 8下在10mM TrisHCl中洗脱。sgRNA从整合DNA技术公司(Integrated DNA technologies)(IDT)以2nmol订购,然后以20μM重悬于10mM Tris EDTA缓冲液中(表6)。核糖核蛋白(RNP)通过在室温下在1X效应子缓冲液(pH 7.5下的100mM NaCl、10mM $MgCl_2$、10mM Tris HCl)中将核酸酶与靶向或非靶向向导以1:2摩尔比预温育30分钟来制备。所有反应均重复三次进行,包含无sgRNA的阴性对照。在RNP形成后,将RNP添加到含有20ng/μL纯化的gDNA的1X效应子缓冲液中的消化反应中,并在37℃下温育1小时。以7.8和15.6nM两种最终浓度测试核酸酶。通过将目标浓度除以每种核酸酶的活性分数来将这些浓度归一化。温育后,将这些反应立即移到4℃,用水稀释30倍,然后在含有1X PrimeTime®基因表达主混合物、10μM正向引物、10μM反向引物和5μM 5'-FAM和ZEN/Iowa Black荧光猝灭剂Taqman探针(IDT)的主混合物中制备用于qPCR(表7)。AriaMx实时PCR系统(安捷伦公司(Agilent))与以下循环一起使用:1)在95℃下持续15分钟,2)在95℃下持续5秒,以及3)在60℃下持续1分钟,其中步骤2-3重复40X。使用Cq值计算在切割百分比方程(下文)之后每个反应的gDNA切割百分比。将所有相对于非靶向对照反应归一化。图15A示出了在所使用的较高浓度的蛋白质下,通过MG119-28和sgRNA3切割的平均60%gDNA以及用sgRNA2切割的21%的实例。

[0275] 切割百分比方程

[0276] 切割%=100-($2^{-(Cq(实验)-Cq(非靶向对照))}$ x 100)

[0277] 表6:小鼠白蛋白内含子1和经化学修饰的sgRNA(IDT)中的靶向序列

[0278]

sgRNA 名称	序列 (5'-3')	小鼠白蛋白靶标 (5'-3')
119-28 sgRNA1_小鼠_Alb	mU*mU*mG*rArArArUrArArArArUrGrArArUrUrUrCrArArArCrCrCrCrUrUrCrGrGrGrGrGrArGrGrGrCrGrCrGrUrUrGrGrArGrCrGrCrCrUrUrArGrUrUrUrGrArGrGrUrGrCrArGrArArUrCrArArArArArArArArCrUrGrCrGrArCrGrArUrGrGrAr	AAGATTGATGAAGACAACTA

sgRNA 名称	序列（5'-3'）	小鼠白蛋白靶标（5'-3'）
	GrGrUrCrGrUrUrUrCrArGrUrCrU rCrUrGrUrArCrArCrUrCrArArAr ArArArUrUrCrArCrUrUrGrArGrA rArArUrCrArArGrUrGrArArUrAr UrCrCrArArCrArArGrArUrUrGrA rUrGrArArGrArCrArA*mC*mU* mA	
119-28 sgRNA2_ 小鼠_Alb	mU*mU*mG*rArArArUrArArArA rUrGrArArUrUrUrCrArArArCrCr CrCrUrUrCrGrGrGrGrGrArGrGrG rCrGrCrGrUrUrGrGrArGrCrGrCrC rUrUrArGrUrUrUrGrArGrGrUrGr CrArGrArArUrCrArArArArArArA rCrUrGrCrGrArCrGrArUrGrGrAr GrGrUrCrGrUrUrUrCrArGrUrCrU rCrUrGrUrArCrArCrUrCrArArAr ArArArUrUrCrArCrUrUrGrArGrA rArArUrCrArArGrUrGrArArUrAr UrCrCrArArCrGrGrUrCrArGrUrG rArArGrArGrArArGrA*mA*mC* mA	GGTCAGTGAAGAGAAGAACA
119-28 sgRNA3_ 小鼠_Alb	mU*mU*mG*rArArArUrArArArA rUrGrArArUrUrUrCrArArArCrCr CrCrUrUrCrGrGrGrGrGrArGrGrG rCrGrCrGrUrUrGrGrArGrCrGrCrC rUrUrArGrUrUrUrGrArGrGrUrGr CrArGrArArUrCrArArArArArArA rCrUrGrCrGrArCrGrArUrGrGrAr GrGrUrCrGrUrUrUrCrArGrUrCrU rCrUrGrUrArCrArCrUrCrArArAr ArArArUrUrCrArCrUrUrGrArGrA rArArUrCrArArGrUrGrArArUrAr UrCrCrArArCrGrGrUrCrArGrUrG rArArGrArGrArArGrA*mC*mC* mA	AGTGTAGCAGAGAGGAACCA
119-28 sgRNA4_ 小鼠_Alb	mU*mU*mG*rArArArUrArArArA rUrGrArArUrUrUrCrArArArCrCr CrCrUrUrCrGrGrGrGrGrArGrGrG rCrGrCrGrUrUrGrGrArGrCrGrCrC rUrUrArGrUrUrUrGrArGrGrUrGr CrArGrArArUrCrArArArArArArA rCrUrGrCrGrArCrGrArUrGrGrAr GrGrUrCrGrUrUrUrCrArGrUrCrU rCrUrGrUrArCrArCrUrCrArArAr ArArArUrUrCrArCrUrUrGrArGrA rArArUrCrArArGrUrGrArArUrAr	TCTGTGGAAACAGGGAGAGA

[0279]

[0280]

sgRNA 名称	序列（5'-3'）	小鼠白蛋白靶标（5'-3'）
	UrCrCrArArCrUrCrUrGrUrGrGrArArArArCrArGrGrGrArG*mA*mG*mA	

[0281] 表7:用于qPCR的DNA寡核苷酸

[0282]

寡核苷酸名称	寡核苷酸序列
611F_HE	TGCACAGATATAAACACTTAACGGG
869R_HE	GGGCGATCTCACTCTTGTCT
680_HE Taqman探针	5'-FAM-AGCAGAGAGGAACCATTGCCACCTTCAG

[0283] 使用纯化蛋白对Hepa 1-6细胞中的基因组DNA进行体内切割

[0284] 细胞内编辑用靶向内含子1处的小鼠白蛋白基因的核酸酶和向导的RNP复合物进行演示(表6)。将Hepa1-6细胞解冻、洗涤并重悬于Dulbecco的改性伊格尔培养基(DMEM,10% FBS和1% Pen-strep)中。在37℃下在30mL的培养基中将细胞以每15cm培养皿$4x10^6$个细胞的密度接种。两天后,当细胞达到70-80%汇合时,将细胞分裂。将细胞用0.25%胰蛋白酶进行胰蛋白酶化,然后在37℃下温育30秒。添加DMEM,然后分成3mL,并用27mL培养基进一步稀释。将分裂的细胞再温育两天。在核转染之前,从板中抽吸培养基,并且在胰蛋白酶化之前用pH为7.2的1X磷酸盐缓冲盐水(PBS,Gibco™)洗涤细胞。中和胰蛋白酶,并用DMEM重悬细胞。用Countess3FL(英杰公司)对细胞悬浮液中的细胞进行计数,以计算要沉淀的细胞体积。下游的每个处理总共需要100,000个细胞。将细胞在sorvall X Pro系列离心机(赛默飞世尔公司)中以300x g离心7分钟,然后在pH为7.2的PBS中洗涤,然后重悬于来自Amaxa™4D-Nucleofector™试剂盒(龙沙公司(Lonza))的Nucleofector™溶液中。

[0285] 通过将120pmol的核酸酶与120pmol的向导在室温下温育90分钟来单独制备RNP复合物。将20μL的制备细胞添加到RNP中。在4D-Nucleofector™系统(龙沙公司)中按照Amaxa™4D-Nucleofector™方案的建议进行核转染。将核转染的细胞从核转染盒转移到24孔板,每个孔含有500μL培养基。在温育两天后,使用以下循环用QuickExtract(龙沙公司)萃取来自所有处理的gDNA:1)在65℃下15分钟;2)在68℃下15分钟;以及3)在98℃下10分钟,并且然后在4℃下保持直到使用。使用以下循环将用Phusion Flash高保真PCR主混合物(赛默飞世尔公司)所得的提取gDNA扩增317bp的靶向窗口:1)在98℃下10秒;2)在98℃下1秒;3)在63℃下5秒;4)在72℃下15秒;以及5)在72℃下重复步骤2-5持续30个循环1分钟,然后在4℃下保持。在清洁之前,将扩增子在2%琼脂糖凝胶上可视化,并用具有1.8X的珠体积的HighPrep磁珠(MagBio基因组学公司)浓缩以取样。将样品在水中洗脱。将INDEL通过NGS在具有v3试剂盒(600个循环;表8)和用于2x301bp双端读数的5%phiX的MiSeq上测序,每个样品至少20,000个读数。用修改后的CRISPResso2程序进行INDEL分析(Clement等人,2019;https://doi.org/10.1038/s41587-019-0032-3),并且结果显示在表9和图15B中。

[0286] 表8:用于NGS PCR1的寡核苷酸

[0287]

寡核苷酸名称	寡核苷酸序列(5'-3')
611F_NGS	GCTCTTCCGATCTNNNNNTGCACAGATATAAACACTTAACGGG
927R_NGS	GCTCTTCCGATCTNNNNNTTCAGCATTATAACTTACAGGCCT

[0288] 表9:相对于Apo条件归一化的INDEL百分比

[0289]

RNP	复制品 1	复制品 2	复制品 3	平均
	% INDEL	% INDEL	% INDEL	% INDEL
119-28 sgRNA1_小鼠_Alb	0.70	0.23	0.87	0.60
119-28 sgRNA2_小鼠_Alb	0.32	0.48	0.38	0.39
119-28 sgRNA3_小鼠_Alb	50.57	13.12	11.68	25.12
119-28 sgRNA4_小鼠_Alb	9.23	2.52	0.60	4.12

[0290] 实例15-MG119蛋白纯化的缓冲液优化(预测)

[0291] 迄今为止,MG119蛋白已在Nickel_A缓冲液中纯化。由于其高盐度,Nickel_A缓冲液与下游体内测定不相容,并且快速稀释到低盐溶液中诱导蛋白质沉淀。为了优化蛋白质稳定性和下游测定相容性的缓冲液,MG119核酸酶最初在高盐缓冲液(750mM NaCl)中纯化,并逐渐洗涤成具有200mM NaCl和两性离子氨基酸L-精氨酸(50mM)和L-谷氨酸盐(50mM)的Nickel_A缓冲液变体。在经验基础上,还将各种稳定糖(核糖、山梨糖醇、甘露糖醇、木糖醇)添加到缓冲液中以增强低盐缓冲液中的蛋白质稳定性。

[0292] 实例16-核酸酶活性的基于荧光的测量(预测)

[0293] 新型细胞系工程化

[0294] 用于测量体内(即哺乳动物细胞系中的)核酸酶活性的目前测定需要大量的数据分析和长达一周的周转时间。为了加快体内核酸酶活性的评估,永生化哺乳动物细胞系经工程化提供基因组DNA编辑的即时数据。在IMDM(Gibco#12440053)+10% FBS(Corning™常规胎牛血清,MT35011CV)中生长的K562哺乳动物细胞用于此测定。K562哺乳动物细胞用12pmol Cas9蛋白(IDT#1081058)、60pmol sgRNA(Mali等人《科学(Science)》,2013年2月15日;339(6121):823-6.),以及含有用于mMBP-(GGS)3-eGFP蛋白的表达序列的1200ng质粒(pUC主链)进行转染。此构建体的基因组整合导致在合成的MND启动子下的组成型表达。使细胞生长6天,每3天传代一次。通过使用Sony MA900细胞分选器将表达GFP的单个细胞分选到96孔板中,从单细胞分离单基因细胞系。

[0295] 基于荧光的体内核酸酶活性筛选

[0296] 适当的sgRNA被设计成沿着mMBP和eGFP基因引导核酸酶切割,使得插入缺失形成产生移码突变,从而导致荧光丧失。通过将100pmol蛋白和200pmol sgRNA合并,并且在室温下在5μL的最终体积中温育≥20分钟来形成MG119 RNP复合物。将K562细胞在1x PBS中洗涤,并重悬于核载体溶液(SF细胞系96孔Nucleofector™溶液)中,每孔大约200,000个细胞。将细胞和RNP以25μL的最终体积在Lonza 96孔核转染板(SF细胞系96孔Nucleofector™试剂盒,V4SC-2096)中合并,进行核转染(K562细胞,FF-120),并在IMDM+10% FBS培养基中回收。使细胞在37℃下恢复2-3天。为了进行分析,将细胞用1x PBS洗涤两次,然后在室温下用1x PBS+LIVE/DEAD可固定近IR死细胞染色试剂盒染料(赛默飞世尔公司L10119)染色20分钟。在重悬于1x PBS中之前,将细胞用1x PBS再洗涤一次,并上样到Attune NxT声聚焦流式细胞仪(型号AFC2)中以用于荧光分析。未经编辑的阳性对照(无RNP的核转染)和阴性对照(无荧光K562细胞)用于建立阳性和阴性荧光门,并分析细胞群的GFP通道中的荧光损失,以

评估体内核酸酶活性。

[0297] 实例17-用于表观基因组编辑(预测)

[0298] 表观基因组编辑是基因调节技术,其包括组成型地或暂时地打开或关闭基因。此类技术可以使用与3种蛋白质融合的经催化死亡的Cas9(dCas9):Dnmt3A、Dnmt3L和KRAB(例如,如Nuñez等人《细胞(Cell)》2021,184(9),2503-2519中所述,所述文献通过引用以其整体并入)。Dnmt3A和Dnmt3L是DNA甲基转移酶。KRAB结构域介导组蛋白甲基化。启动子区域中的DNA和组蛋白的甲基化介导组成型基因抑制。dCas9和向导RNA可以将DNA和组蛋白甲基化复合物募集到启动子区域,不需要核酸酶活性。Dnmt3A、Dnmt3L和KRAB一起为579aa,并且dCas9为1,368aa。融合蛋白由1,947aa或5,841个核苷酸组成,超过腺相关病毒载体(AAV)包装极限(4.7Kb)。因此,需要创建更紧凑的表观基因组编辑器。来自MG119家族的紧凑V型核酸酶代表了在表观基因组编辑技术中用作死核酸酶合作伙伴的良好候选物。由于其尺寸较小,范围为350至700aa,当与DNA和组蛋白甲基化复合物融合时,融合蛋白的尺寸可在例如约929至约1,279aa或约2787至约3837个核苷酸的范围内,从而允许容易包装在AAV中。

[0299] 为了测试MG119融合蛋白作为表观基因组编辑,通过慢病毒转导产生在嵌合启动子(GAPDH-Srnpn)下表达GFP的HEK293T细胞。设计靶向嵌合启动子的MG119家族向导RNA。从IDT订购引导物,用3个2'-O-甲基取代基和3个硫代磷酸酯键修饰5'和3'核苷酸以获得稳定性。将MG119核酸酶的死形式与DNA和组蛋白甲基化复合物(MG119表观基因组编辑器)融合。将融合蛋白克隆在CMV启动子下方的哺乳动物表达质粒中。用化学合成的向导和表达MG119表观基因组编辑的质粒转染表达GFP的HEK293T细胞。通过流式细胞术分析转染的细胞。成功的MG119表观基因组编辑物由转染细胞中GFP荧光的损失确定。然后使用MG119表观基因组编辑来靶向治疗关注的基因。

[0300] 表10-本文提及的蛋白质和核酸序列

[0301]

目录号	SEQ ID NO	描述	类型
MG122 效应子	1	MG122-1 效应子	蛋白质
MG122 效应子	2	MG122-2 效应子	蛋白质
MG122 效应子	3	MG122-3 效应子	蛋白质
MG122 效应子	4	MG122-4 效应子	蛋白质
MG122 效应子	5	MG122-5 效应子	蛋白质
MG120 效应子	6	MG120-1 效应子	蛋白质

[0302]

目录号	SEQ ID NO	描述	类型
MG120 效应子	7	MG120-2 效应子	蛋白质
MG120 效应子	8	MG120-3 效应子	蛋白质
MG120 效应子	9	MG120-4 效应子	蛋白质
MG120 效应子	10	MG120-5 效应子	蛋白质
MG120 效应子	11	MG120-6 效应子	蛋白质
MG120 效应子	12	MG120-7 效应子	蛋白质
MG120 效应子	13	MG120-8 效应子	蛋白质
MG120 效应子	14	MG120-9 效应子	蛋白质
MG118 效应子	15	MG118-1 效应子	蛋白质
MG90 效应子	16	MG90-3 效应子	蛋白质
MG90 效应子	17	MG90-5 效应子	蛋白质
MG90 效应子	18	MG90-6 效应子	蛋白质
MG90 效应子	19	MG90-7 效应子	蛋白质
MG90 效应子	20	MG90-8 效应子	蛋白质
MG90 效应子	21	MG90-16 效应子	蛋白质
MG90 效应子	22	MG90-17 效应子	蛋白质
MG90 效应子	23	MG90-18 效应子	蛋白质
MG90 效应子	24	MG90-19 效应子	蛋白质
MG90 效应子	25	MG90-20 效应子	蛋白质
MG90 效应子	26	MG90-21 效应子	蛋白质
MG90 效应子	27	MG90-22 效应子	蛋白质
MG90 效应子	28	MG90-23 效应子	蛋白质
MG90 效应子	29	MG90-24 效应子	蛋白质
MG119 效应子	30	MG119-1 效应子	蛋白质
MG119 效应子	31	MG119-2 效应子	蛋白质
MG119 效应子	32	MG119-3 效应子	蛋白质
MG119 效应子	33	MG119-4 效应子	蛋白质
MG119 效应子	34	MG119-5 效应子	蛋白质
MG119 效应子	35	MG119-6 效应子	蛋白质
MG119 效应子	36	MG119-7 效应子	蛋白质
MG119 效应子	37	MG119-8 效应子	蛋白质
MG119 效应子	38	MG119-9 效应子	蛋白质
MG119 效应子	39	MG119-10 效应子	蛋白质
MG119 效应子	40	MG119-11 效应子	蛋白质
MG119 效应子	41	MG119-12 效应子	蛋白质
MG119 效应子	42	MG119-13 效应子	蛋白质

[0303]

目录号	SEQ ID NO	描述	类型
MG119 效应子	43	MG119-14 效应子	蛋白质
MG119 效应子	44	MG119-15 效应子	蛋白质
MG119 效应子	45	MG119-16 效应子	蛋白质
MG119 效应子	46	MG119-17 效应子	蛋白质
MG119 效应子	47	MG119-18 效应子	蛋白质
MG119 效应子	48	MG119-19 效应子	蛋白质
MG119 效应子	49	MG119-20 效应子	蛋白质
MG119 效应子	50	MG119-21 效应子	蛋白质
MG119 效应子	51	MG119-22 效应子	蛋白质
MG119 效应子	52	MG119-23 效应子	蛋白质
MG119 效应子	53	MG119-24 效应子	蛋白质
MG119 效应子	54	MG119-25 效应子	蛋白质
MG119 效应子	55	MG119-26 效应子	蛋白质
MG119 效应子	56	MG119-27 效应子	蛋白质
MG119 效应子	57	MG119-28 效应子	蛋白质
MG119 效应子	58	MG119-29 效应子	蛋白质
MG119 效应子	59	MG119-30 效应子	蛋白质
MG119 效应子	60	MG119-31 效应子	蛋白质
MG119 效应子	61	MG119-32 效应子	蛋白质
MG119 效应子	62	MG119-33 效应子	蛋白质
MG119 效应子	63	MG119-34 效应子	蛋白质
MG119 效应子	64	MG119-35 效应子	蛋白质
MG119 效应子	65	MG119-36 效应子	蛋白质
MG119 效应子	66	MG119-37 效应子	蛋白质
MG119 效应子	67	MG119-38 效应子	蛋白质
MG119 效应子	68	MG119-39 效应子	蛋白质
MG119 效应子	69	MG119-40 效应子	蛋白质
MG119 效应子	70	MG119-41 效应子	蛋白质
MG119 效应子	71	MG119-42 效应子	蛋白质
MG119 效应子	72	MG119-43 效应子	蛋白质
MG119 效应子	73	MG119-44 效应子	蛋白质
MG119 效应子	74	MG119-45 效应子	蛋白质
MG119 效应子	75	MG119-46 效应子	蛋白质
MG119 效应子	76	MG119-47 效应子	蛋白质
MG119 效应子	77	MG119-48 效应子	蛋白质
MG119 效应子	78	MG119-49 效应子	蛋白质

[0304]

目录号	SEQ ID NO	描述	类型
MG119 效应子	79	MG119-50 效应子	蛋白质
MG119 效应子	80	MG119-51 效应子	蛋白质
MG119 效应子	81	MG119-52 效应子	蛋白质
MG119 效应子	82	MG119-53 效应子	蛋白质
MG119 效应子	83	MG119-54 效应子	蛋白质
MG119 效应子	84	MG119-55 效应子	蛋白质
MG119 效应子	85	MG119-56 效应子	蛋白质
MG119 效应子	86	MG119-57 效应子	蛋白质
MG119 效应子	87	MG119-58 效应子	蛋白质
MG119 效应子	88	MG119-59 效应子	蛋白质
MG119 效应子	89	MG119-61 效应子	蛋白质
MG119 效应子	90	MG119-62 效应子	蛋白质
MG119 效应子	91	MG119-63 效应子	蛋白质
MG119 效应子	92	MG119-64 效应子	蛋白质
MG119 效应子	93	MG119-65 效应子	蛋白质
MG119 效应子	94	MG119-66 效应子	蛋白质
MG119 效应子	95	MG119-67 效应子	蛋白质
MG119 效应子	96	MG119-68 效应子	蛋白质
MG119 效应子	97	MG119-69 效应子	蛋白质
MG119 效应子	98	MG119-70 效应子	蛋白质
MG119 效应子	99	MG119-71 效应子	蛋白质
MG119 效应子	100	MG119-72 效应子	蛋白质
MG119 效应子	101	MG119-73 效应子	蛋白质
MG119 效应子	102	MG119-74 效应子	蛋白质
MG119 效应子	103	MG119-75 效应子	蛋白质
MG119 效应子	104	MG119-76 效应子	蛋白质
MG119 效应子	105	MG119-77 效应子	蛋白质
MG119 效应子	106	MG119-78 效应子	蛋白质
MG119 效应子	107	MG119-79 效应子	蛋白质
MG119 效应子	108	MG119-80 效应子	蛋白质
MG119 效应子	109	MG119-81 效应子	蛋白质
MG119 效应子	110	MG119-83 效应子	蛋白质
MG119 效应子	111	MG119-84 效应子	蛋白质
MG119 效应子	112	MG119-85 效应子	蛋白质
MG119 效应子	113	MG119-86 效应子	蛋白质
MG119 效应子	114	MG119-87 效应子	蛋白质

[0305]

目录号	SEQ ID NO	描述	类型
MG119 效应子	115	MG119-88 效应子	蛋白质
MG119 效应子	116	MG119-89 效应子	蛋白质
MG119 效应子	117	MG119-90 效应子	蛋白质
MG119 效应子	118	MG119-91 效应子	蛋白质
MG119 效应子	119	MG119-92 效应子	蛋白质
MG119 效应子	120	MG119-93 效应子	蛋白质
MG119 效应子	121	MG119-94 效应子	蛋白质
MG119 效应子	122	MG119-95 效应子	蛋白质
MG119 效应子	123	MG119-96 效应子	蛋白质
MG119 效应子	124	MG119-97 效应子	蛋白质
MG119 效应子	125	MG119-98 效应子	蛋白质
MG119 效应子	126	MG119-99 效应子	蛋白质
MG119 效应子	127	MG119-100 效应子	蛋白质
MG119 效应子	128	MG119-101 效应子	蛋白质
MG119 效应子	129	MG119-102 效应子	蛋白质
MG119 效应子	130	MG119-103 效应子	蛋白质
MG119 效应子	131	MG119-104 效应子	蛋白质
MG119 效应子	132	MG119-105 效应子	蛋白质
MG119 效应子	133	MG119-106 效应子	蛋白质
MG119 效应子	134	MG119-107 效应子	蛋白质
MG119 效应子	135	MG119-108 效应子	蛋白质
MG119 效应子	136	MG119-109 效应子	蛋白质
MG119 效应子	137	MG119-110 效应子	蛋白质
MG119 效应子	138	MG119-111 效应子	蛋白质
MG119 效应子	139	MG119-112 效应子	蛋白质
MG119 效应子	140	MG119-113 效应子	蛋白质
MG119 效应子	141	MG119-114 效应子	蛋白质
MG119 效应子	142	MG119-115 效应子	蛋白质
MG119 效应子	143	MG119-116 效应子	蛋白质
MG119 效应子	144	MG119-117 效应子	蛋白质
MG119 效应子	145	MG119-118 效应子	蛋白质
MG119 效应子	146	MG119-119 效应子	蛋白质
MG119 效应子	147	MG119-120 效应子	蛋白质
MG119 效应子	148	MG119-121 效应子	蛋白质
MG119 效应子	149	MG119-122 效应子	蛋白质
MG119 效应子	150	MG119-123 效应子	蛋白质

[0306]

目录号	SEQ ID NO	描述	类型
MG91B 效应子	151	MG91B-1 效应子	蛋白质
MG91B 效应子	152	MG91B-2 效应子	蛋白质
MG91B 效应子	153	MG91B-3 效应子	蛋白质
MG91B 效应子	154	MG91B-4 效应子	蛋白质
MG91B 效应子	155	MG91B-5 效应子	蛋白质
MG91B 效应子	156	MG91B-6 效应子	蛋白质
MG91B 效应子	157	MG91B-7 效应子	蛋白质
MG91B 效应子	158	MG91B-8 效应子	蛋白质
MG91B 效应子	159	MG91B-9 效应子	蛋白质
MG91B 效应子	160	MG91B-10 效应子	蛋白质
MG91B 效应子	161	MG91B-11 效应子	蛋白质
MG91B 效应子	162	MG91B-12 效应子	蛋白质
MG91B 效应子	163	MG91B-13 效应子	蛋白质
MG91B 效应子	164	MG91B-14 效应子	蛋白质
MG91B 效应子	165	MG91B-15 效应子	蛋白质
MG91B 效应子	166	MG91B-16 效应子	蛋白质
MG91B 效应子	167	MG91B-17 效应子	蛋白质
MG91B 效应子	168	MG91B-18 效应子	蛋白质
MG91B 效应子	169	MG91B-19 效应子	蛋白质
MG91B 效应子	170	MG91B-20 效应子	蛋白质
MG91B 效应子	171	MG91B-21 效应子	蛋白质
MG91B 效应子	172	MG91B-22 效应子	蛋白质
MG91B 效应子	173	MG91B-23 效应子	蛋白质
MG91B 效应子	174	MG91B-24 效应子	蛋白质
MG91B 效应子	175	MG91B-25 效应子	蛋白质
MG91B 效应子	176	MG91B-26 效应子	蛋白质
MG91B 效应子	177	MG91B-27 效应子	蛋白质
MG91B 效应子	178	MG91B-28 效应子	蛋白质
MG91B 效应子	179	MG91B-29 效应子	蛋白质
MG91B 效应子	180	MG91B-30 效应子	蛋白质
MG91B 效应子	181	MG91B-31 效应子	蛋白质
MG91B 效应子	182	MG91B-32 效应子	蛋白质
MG91B 效应子	183	MG91B-33 效应子	蛋白质
MG91B 效应子	184	MG91B-34 效应子	蛋白质
MG91B 效应子	185	MG91B-35 效应子	蛋白质
MG91B 效应子	186	MG91B-36 效应子	蛋白质

[0307]

目录号	SEQ ID NO	描述	类型
MG91B 效应子	187	MG91B-37 效应子	蛋白质
MG91B 效应子	188	MG91B-38 效应子	蛋白质
MG91B 效应子	189	MG91B-39 效应子	蛋白质
MG91B 效应子	190	MG91B-40 效应子	蛋白质
MG91B 效应子	191	MG91B-41 效应子	蛋白质
MG91B 效应子	192	MG91B-42 效应子	蛋白质
MG91B 效应子	193	MG91B-43 效应子	蛋白质
MG91B 效应子	194	MG91B-44 效应子	蛋白质
MG91B 效应子	195	MG91B-45 效应子	蛋白质
MG91B 效应子	196	MG91B-46 效应子	蛋白质
MG91B 效应子	197	MG91B-47 效应子	蛋白质
MG91B 效应子	198	MG91B-48 效应子	蛋白质
MG91B 效应子	199	MG91B-49 效应子	蛋白质
MG91B 效应子	200	MG91B-50 效应子	蛋白质
MG91B 效应子	201	MG91B-51 效应子	蛋白质
MG91B 效应子	202	MG91B-52 效应子	蛋白质
MG91B 效应子	203	MG91B-53 效应子	蛋白质
MG91B 效应子	204	MG91B-54 效应子	蛋白质
MG91B 效应子	205	MG91B-55 效应子	蛋白质
MG91B 效应子	206	MG91B-56 效应子	蛋白质
MG91B 效应子	207	MG91B-57 效应子	蛋白质
MG91B 效应子	208	MG91B-58 效应子	蛋白质
MG91B 效应子	209	MG91B-59 效应子	蛋白质
MG91B 效应子	210	MG91B-60 效应子	蛋白质
MG91B 效应子	211	MG91B-61 效应子	蛋白质
MG91B 效应子	212	MG91B-62 效应子	蛋白质
MG91B 效应子	213	MG91B-63 效应子	蛋白质
MG91B 效应子	214	MG91B-64 效应子	蛋白质
MG91B 效应子	215	MG91B-65 效应子	蛋白质
MG91B 效应子	216	MG91B-66 效应子	蛋白质
MG91B 效应子	217	MG91B-67 效应子	蛋白质
MG91B 效应子	218	MG91B-68 效应子	蛋白质
MG91B 效应子	219	MG91B-69 效应子	蛋白质
MG91B 效应子	220	MG91B-70 效应子	蛋白质
MG91B 效应子	221	MG91B-71 效应子	蛋白质
MG91B 效应子	222	MG91B-72 效应子	蛋白质

[0308]

目录号	SEQ ID NO	描述	类型
MG91B 效应子	223	MG91B-73 效应子	蛋白质
MG91B 效应子	224	MG91B-74 效应子	蛋白质
MG91B 效应子	225	MG91B-75 效应子	蛋白质
MG91B 效应子	226	MG91B-76 效应子	蛋白质
MG91B 效应子	227	MG91B-77 效应子	蛋白质
MG91B 效应子	228	MG91B-78 效应子	蛋白质
MG91B 效应子	229	MG91B-79 效应子	蛋白质
MG91B 效应子	230	MG91B-80 效应子	蛋白质
MG91B 效应子	231	MG91B-81 效应子	蛋白质
MG91B 效应子	232	MG91B-82 效应子	蛋白质
MG91B 效应子	233	MG91B-83 效应子	蛋白质
MG91B 效应子	234	MG91B-84 效应子	蛋白质
MG91B 效应子	235	MG91B-85 效应子	蛋白质
MG91B 效应子	236	MG91B-86 效应子	蛋白质
MG91B 效应子	237	MG91B-87 效应子	蛋白质
MG91B 效应子	238	MG91B-88 效应子	蛋白质
MG91B 效应子	239	MG91B-89 效应子	蛋白质
MG91B 效应子	240	MG91B-90 效应子	蛋白质
MG91B 效应子	241	MG91B-91 效应子	蛋白质
MG91B 效应子	242	MG91B-92 效应子	蛋白质
MG91B 效应子	243	MG91B-93 效应子	蛋白质
MG91B 效应子	244	MG91B-94 效应子	蛋白质
MG91B 效应子	245	MG91B-95 效应子	蛋白质
MG91B 效应子	246	MG91B-96 效应子	蛋白质
MG91B 效应子	247	MG91B-97 效应子	蛋白质
MG91B 效应子	248	MG91B-98 效应子	蛋白质
MG91B 效应子	249	MG91B-99 效应子	蛋白质
MG91B 效应子	250	MG91B-100 效应子	蛋白质
MG91B 效应子	251	MG91B-101 效应子	蛋白质
MG91B 效应子	252	MG91B-102 效应子	蛋白质
MG91B 效应子	253	MG91B-103 效应子	蛋白质
MG91B 效应子	254	MG91B-104 效应子	蛋白质
MG91B 效应子	255	MG91B-105 效应子	蛋白质
MG91B 效应子	256	MG91B-106 效应子	蛋白质
MG91B 效应子	257	MG91B-107 效应子	蛋白质
MG91B 效应子	258	MG91B-108 效应子	蛋白质

[0309]

目录号	SEQ ID NO	描述	类型
MG91B 效应子	259	MG91B-109 效应子	蛋白质
MG91B 效应子	260	MG91B-110 效应子	蛋白质
MG91B 效应子	261	MG91B-111 效应子	蛋白质
MG91B 效应子	262	MG91B-112 效应子	蛋白质
MG91B 效应子	263	MG91B-113 效应子	蛋白质
MG91B 效应子	264	MG91B-114 效应子	蛋白质
MG91B 效应子	265	MG91B-115 效应子	蛋白质
MG91B 效应子	266	MG91B-116 效应子	蛋白质
MG91B 效应子	267	MG91B-117 效应子	蛋白质
MG91B 效应子	268	MG91B-118 效应子	蛋白质
MG91B 效应子	269	MG91B-119 效应子	蛋白质
MG91B 效应子	270	MG91B-120 效应子	蛋白质
MG91B 效应子	271	MG91B-121 效应子	蛋白质
MG91B 效应子	272	MG91B-122 效应子	蛋白质
MG91B 效应子	273	MG91B-123 效应子	蛋白质
MG91B 效应子	274	MG91B-124 效应子	蛋白质
MG91B 效应子	275	MG91B-125 效应子	蛋白质
MG91B 效应子	276	MG91B-126 效应子	蛋白质
MG91B 效应子	277	MG91B-127 效应子	蛋白质
MG91B 效应子	278	MG91B-128 效应子	蛋白质
MG91B 效应子	279	MG91B-129 效应子	蛋白质
MG91B 效应子	280	MG91B-130 效应子	蛋白质
MG91B 效应子	281	MG91B-131 效应子	蛋白质
MG91B 效应子	282	MG91B-132 效应子	蛋白质
MG91B 效应子	283	MG91B-133 效应子	蛋白质
MG91B 效应子	284	MG91B-134 效应子	蛋白质
MG91B 效应子	285	MG91B-135 效应子	蛋白质
MG91B 效应子	286	MG91B-136 效应子	蛋白质
MG91B 效应子	287	MG91B-137 效应子	蛋白质
MG91B 效应子	288	MG91B-138 效应子	蛋白质
MG91B 效应子	289	MG91B-139 效应子	蛋白质
MG91B 效应子	290	MG91B-140 效应子	蛋白质
MG91B 效应子	291	MG91B-141 效应子	蛋白质
MG91C 效应子	292	MG91C-1 效应子	蛋白质
MG91C 效应子	293	MG91C-2 效应子	蛋白质
MG91C 效应子	294	MG91C-3 效应子	蛋白质

[0310]

目录号	SEQ ID NO	描述	类型
MG91C 效应子	295	MG91C-4 效应子	蛋白质
MG91C 效应子	296	MG91C-5 效应子	蛋白质
MG91C 效应子	297	MG91C-6 效应子	蛋白质
MG91C 效应子	298	MG91C-7 效应子	蛋白质
MG91C 效应子	299	MG91C-8 效应子	蛋白质
MG91C 效应子	300	MG91C-9 效应子	蛋白质
MG91C 效应子	301	MG91C-10 效应子	蛋白质
MG91C 效应子	302	MG91C-11 效应子	蛋白质
MG91C 效应子	303	MG91C-12 效应子	蛋白质
MG91C 效应子	304	MG91C-13 效应子	蛋白质
MG91C 效应子	305	MG91C-14 效应子	蛋白质
MG91C 效应子	306	MG91C-15 效应子	蛋白质
MG91C 效应子	307	MG91C-16 效应子	蛋白质
MG91C 效应子	308	MG91C-17 效应子	蛋白质
MG91C 效应子	309	MG91C-18 效应子	蛋白质
MG91C 效应子	310	MG91C-19 效应子	蛋白质
MG91C 效应子	311	MG91C-20 效应子	蛋白质
MG91C 效应子	312	MG91C-21 效应子	蛋白质
MG91C 效应子	313	MG91C-22 效应子	蛋白质
MG91C 效应子	314	MG91C-23 效应子	蛋白质
MG91C 效应子	315	MG91C-24 效应子	蛋白质
MG91C 效应子	316	MG91C-25 效应子	蛋白质
MG91C 效应子	317	MG91C-26 效应子	蛋白质
MG91C 效应子	318	MG91C-27 效应子	蛋白质
MG91A 效应子	319	MG91A-1 效应子	蛋白质
MG126 效应子	320	MG126-3 效应子	蛋白质
MG126 效应子	321	MG126-4 效应子	蛋白质
MG126 效应子	322	MG126-5 效应子	蛋白质
MG126 效应子	323	MG126-6 效应子	蛋白质
MG126 效应子	324	MG126-7 效应子	蛋白质
MG126 效应子	325	MG126-8 效应子	蛋白质
编码潜在 tracrRNA 的 MG119-3 效应子基因间区	326	MG119-3_IG1	核苷酸
编码潜在 tracrRNA 的 MG119-3 效应子基因间区	327	MG119-3_IG2	核苷酸
编码潜在 tracrRNA 的 MG119-3 效应子基因间区	328	MG119-3_IG3	核苷酸

[0311]

目录号	SEQ ID NO	描述	类型
编码潜在 tracrRNA 的 MG119-4 效应子基因间区	329	MG119-4_IG1	核苷酸
编码潜在 tracrRNA 的 MG119-4 效应子基因间区	330	MG119-4_IG2	核苷酸
编码潜在 tracrRNA 的 MG119-4 效应子基因间区	331	MG119-4_IG3	核苷酸
编码潜在 tracrRNA 的 MG119-4 效应子基因间区	332	MG119-4_IG4	核苷酸
编码潜在 tracrRNA 的 MG120-1 效应子基因间区	333	MG120-1_IG1	核苷酸
编码潜在 tracrRNA 的 MG120-1 效应子基因间区	334	MG120-1_IG2	核苷酸
编码潜在 tracrRNA 的 MG120-1 效应子基因间区	335	MG120-1_IG3	核苷酸
编码潜在 tracrRNA 的 MG119-1 效应子基因间区	336	MG119-1_IG1	核苷酸
编码潜在 tracrRNA 的 MG119-1 效应子基因间区	337	MG119-1_IG2	核苷酸
编码潜在 tracrRNA 的 MG119-1 效应子基因间区	338	MG119-1_IG3	核苷酸
编码潜在 tracrRNA 的 MG119-1 效应子基因间区	339	MG119-1_IG4	核苷酸
编码潜在 tracrRNA 的 MG119-1 效应子基因间区	340	MG119-1_IG5	核苷酸
编码潜在 tracrRNA 的 MG119-2 效应子基因间区	341	MG119-2_IG1	核苷酸
编码潜在 tracrRNA 的 MG119-2 效应子基因间区	342	MG119-2_IG2	核苷酸
编码潜在 tracrRNA 的 MG119-5 效应子基因间区	343	MG119-5_IG1	核苷酸
编码潜在 tracrRNA 的 MG119-5 效应子基因间区	344	MG119-5_IG2	核苷酸
编码潜在 tracrRNA 的 MG119-5 效应子基因间区	345	MG119-5_IG3	核苷酸
编码潜在 tracrRNA 的 MG90-3 效应子基因间区	346	MG90-3_IG1	核苷酸
编码潜在 tracrRNA 的 MG90-3 效应子基因间区	347	MG90-3_IG2	核苷酸
编码潜在 tracrRNA 加衔接子的 MG119-3 效应子基因间区	348	MG119-3_IG1_衔接子	核苷酸

[0312]

目录号	SEQ ID NO	描述	类型
编码潜在 tracrRNA 加衔接子的 MG119-3 效应子基因间区	349	MG119-3_IG2_衔接子	核苷酸
编码潜在 tracrRNA 加衔接子的 MG119-3 效应子基因间区	350	MG119-3_IG3_衔接子	核苷酸
编码潜在 tracrRNA 加衔接子的 MG119-4 效应子基因间区	351	MG119-4_IG1_衔接子	核苷酸
编码潜在 tracrRNA 加衔接子的 MG119-4 效应子基因间区	352	MG119-4_IG2_衔接子	核苷酸
编码潜在 tracrRNA 加衔接子的 MG119-4 效应子基因间区	353	MG119-4_IG3_衔接子	核苷酸
编码潜在 tracrRNA 加衔接子的 MG119-4 效应子基因间区	354	MG119-4_IG4_衔接子	核苷酸
编码潜在 tracrRNA 加衔接子的 MG120-1 效应子基因间区	355	MG120-1_IG1_衔接子	核苷酸
编码潜在 tracrRNA 加衔接子的 MG120-1 效应子基因间区	356	MG120-1_IG2_衔接子	核苷酸
编码潜在 tracrRNA 加衔接子的 MG120-1 效应子基因间区	357	MG120-1_IG3_衔接子	核苷酸
编码潜在 tracrRNA 加衔接子的 MG119-1 效应子基因间区	358	MG119-1_IG1_衔接子	核苷酸
编码潜在 tracrRNA 加衔接子的 MG119-1 效应子基因间区	359	MG119-1_IG2_衔接子	核苷酸
编码潜在 tracrRNA 加衔接子的 MG119-1 效应子基因间区	360	MG119-1_IG3_衔接子	核苷酸
编码潜在 tracrRNA 加衔接子的 MG119-1 效应子基因间区	361	MG119-1_IG4_衔接子	核苷酸
编码潜在 tracrRNA 加衔接子的 MG119-1 效应子基因间区	362	MG119-1_IG5_衔接子	核苷酸
编码潜在 tracrRNA 加衔接子的 MG119-2 效应子基因间区	363	MG119-2_IG1_衔接子	核苷酸
编码潜在 tracrRNA 加衔接子的 MG119-2 效应子基因间区	364	MG119-2_IG2_衔接子	核苷酸
编码潜在 tracrRNA 加衔接子的 MG119-5 效应子基因间区	365	MG119-5_IG1_衔接子	核苷酸
编码潜在 tracrRNA 加衔接子的 MG119-5 效应子基因间区	366	MG119-5_IG2_衔接子	核苷酸
编码潜在 tracrRNA 加衔接子的 MG119-5 效应子基因间区	367	MG119-5_IG3_衔接子	核苷酸
编码潜在 tracrRNA 加衔接子的 MG90-3 效应子基因间区	368	MG90-3_IG1_衔接子	核苷酸

[0313]

目录号	SEQ ID NO	描述	类型
编码潜在 tracrRNA 加衔接子的 MG90-3 效应子基因间区	369	MG90-3_IG2_衔接子	核苷酸
具有 T7 启动子、正向取向的两个重复序列和一个间隔子的 MG119-3 最小阵列	370	119-3_5U40_31_F	核苷酸
具有 T7 启动子、反向取向的两个重复序列和一个间隔子的 MG119-3 最小阵列	371	119-3_5U40_31_R	核苷酸
具有 T7 启动子、正向取向的两个重复序列和一个间隔子的 MG119-4 最小阵列	372	119-4_5U40_31_F	核苷酸
具有 T7 启动子、反向取向的两个重复序列和一个间隔子的 MG119-4 最小阵列	373	119-4_5U40_31_R	核苷酸
具有 T7 启动子、正向取向的两个重复序列和一个间隔子的 MG120-1 最小阵列	374	120-1_5U40_37_F	核苷酸
具有 T7 启动子、反向取向的两个重复序列和一个间隔子的 MG120-1 最小阵列	375	120-1_5U40_37_R	核苷酸
具有 T7 启动子、反向取向的两个重复序列和一个间隔子的 MG118-1 最小阵列	376	118-1_5U40_38_R	核苷酸
具有 T7 启动子、正向取向的两个重复序列和一个间隔子的 MG119-1 最小阵列	377	119-1_5U67_32_F	核苷酸
具有 T7 启动子、反向取向的两个重复序列和一个间隔子的 MG119-1 最小阵列	378	119-1_5U67_32_R	核苷酸
具有 T7 启动子、正向取向的两个重复序列和一个间隔子的 MG119-2 最小阵列	379	119-2_5U40_28_F	核苷酸
具有 T7 启动子、反向取向的两个重复序列和一个间隔子的 MG119-2 最小阵列	380	119-2_5U40_28_R	核苷酸
具有 T7 启动子、正向取向的两个重复序列和一个间隔子的 MG119-5 最小阵列	381	119-5_5U40_31_F	核苷酸
具有 T7 启动子、反向取向的两个重复序列和一个间隔子的 MG119-5 最小阵列	382	119-5_5U40_31_R	核苷酸

[0314]

目录号	SEQ ID NO	描述	类型
具有 T7 启动子、正向取向的两个重复序列和一个间隔子的 MG90-3 最小阵列	383	90-3_5U67_37_F	核苷酸
具有 T7 启动子、反向取向的两个重复序列和一个间隔子的 MG90-3 最小阵列	384	90-3_5U67_37_R	核苷酸
具有 T7 启动子、正向取向的两个重复序列和一个间隔子加衔接子的 MG119-3 最小阵列	385	119-3_5U40_31_F_衔接子	核苷酸
具有 T7 启动子、反向取向的两个重复序列和一个间隔子加衔接子的 MG119-3 最小阵列	386	119-3_5U40_31_R_衔接子	核苷酸
具有 T7 启动子、正向取向的两个重复序列和一个间隔子加衔接子的 MG119-4 最小阵列	387	119-4_5U40_31_F_衔接子	核苷酸
具有 T7 启动子、反向取向的两个重复序列和一个间隔子加衔接子的 MG119-4 最小阵列	388	119-4_5U40_31_R_衔接子	核苷酸
具有 T7 启动子、正向取向的两个重复序列和一个间隔子加衔接子的 MG120-1 最小阵列	389	120-1_5U40_37_F_衔接子	核苷酸
具有 T7 启动子、反向取向的两个重复序列和一个间隔子加衔接子的 MG120-1 最小阵列	390	120-1_5U40_37_R_衔接子	核苷酸
具有 T7 启动子、反向取向的两个重复序列和一个间隔子加衔接子的 MG118-1 最小阵列	391	118-1_5U40_38_R_衔接子	核苷酸
具有 T7 启动子、正向取向的两个重复序列和一个间隔子加衔接子的 MG119-1 最小阵列	392	119-1_5U67_32_F_衔接子	核苷酸
具有 T7 启动子、反向取向的两个重复序列和一个间隔子加衔接子的 MG119-1 最小阵列	393	119-1_5U67_32_R_衔接子	核苷酸
具有 T7 启动子、正向取向的两个重复序列和一个间隔子加衔接子的 MG119-2 最小阵列	394	119-2_5U40_28_F_衔接子	核苷酸
具有 T7 启动子、反向取向的两个重复序列和一个间隔子加衔接子的 MG119-2 最小阵列	395	119-2_5U40_28_R_衔接子	核苷酸
具有 T7 启动子、正向取向的两个重复序列和一个间隔子加衔	396	119-5_5U40_31_F_衔接子	核苷酸

[0315]

目录号	SEQ ID NO	描述	类型
接子的 MG119-5 最小阵列			
具有 T7 启动子、反向取向的两个重复序列和一个间隔子加衔接子的 MG119-5 最小阵列	397	119-5_5U40_31_R_衔接子	核苷酸
具有 T7 启动子、正向取向的两个重复序列和一个间隔子加衔接子的 MG90-3 最小阵列	398	90-3_5U67_37_F_衔接子	核苷酸
具有 T7 启动子、反向取向的两个重复序列和一个间隔子加衔接子的 MG90-3 最小阵列	399	90-3_5U67_37_R_衔接子	核苷酸
具有修剪重复序列和 18 nt 通用间隔子、靶序列的 MG118-1 crRNA	400	MG118-1_U40_18nt_靶标	核苷酸
具有修剪重复序列和 18 nt 通用间隔子、靶序列的 MG118-1 crRNA	401	MG118-1_U67_18nt_靶标	核苷酸
具有 10 RAR 和 24 nt 通用间隔子、靶序列的 MG90-3 sgRNA	402	90-3_sgRNA_10bp_RAR_U40_24_靶标	核苷酸
具有 16 RAR 和 24 nt 通用间隔子、靶序列的 MG90-3 sgRNA	403	90-3_sgRNA_16bp_RAR_U40_24_靶标	核苷酸
具有 10 RAR 和 24 nt 通用间隔子、靶序列的 MG119-1 sgRNA	404	119-1_sgRNA_10bp_RAR_U40_24_靶标	核苷酸
具有 15 RAR 和 24 nt 通用间隔子、靶序列的 MG119-1 sgRNA	405	119-1_sgRNA_15bp_RAR_U40_24_靶标	核苷酸
具有 10 RAR 和 24 nt 通用间隔子、靶序列的 MG119-2 sgRNA	406	119-2_sgRNA_10bp_RAR_U40_24_靶标	核苷酸
具有 16 RAR 和 24 nt 通用间隔子、靶序列的 MG119-2 sgRNA	407	119-2_sgRNA_16bp_RAR_U40_24_靶标	核苷酸
具有 9 RAR 和 24 nt 通用间隔子、靶序列的 MG119-5 sgRNA	408	119-5_sgRNA_9bp_RAR_U40_24_靶标	核苷酸
具有 16 RAR 和 24 nt 通用间隔子、靶序列的 MG119-5 sgRNA	409	119-5_sgRNA_16bp_RAR_U40_24_靶标	核苷酸
具有修剪重复序列和 18 nt 通用间隔子的 MG118-1 crRNA	410	MG118-1_U40_18nt	核苷酸
具有修剪重复序列和 18 nt 通用间隔子的 MG118-1 crRNA	411	MG118-1_U67_18nt	核苷酸
具有 10 RAR 和 24 nt 通用间隔子的 MG90-3 sgRNA	412	90-3_sgRNA_10bp_RAR_U40_24	核苷酸
具有 16 RAR 和 24 nt 通用间隔子的 MG90-3 sgRNA	413	90-3_sgRNA_16bp_RAR_U40_24	核苷酸

[0316]

目录号	SEQ ID NO	描述	类型
具有 10 RAR 和 24 nt 通用间隔子的 MG119-1 sgRNA	414	119-1_sgRNA_10bp_RAR_U40_24	核苷酸
具有 15 RAR 和 24 nt 通用间隔子的 MG119-1 sgRNA	415	119-1_sgRNA_15bp_RAR_U40_24	核苷酸
具有 10 RAR 和 24 nt 通用间隔子的 MG119-2 sgRNA	416	119-2_sgRNA_10bp_RAR_U40_24	核苷酸
具有 16 RAR 和 24 nt 通用间隔子的 MG119-2 sgRNA	417	119-2_sgRNA_16bp_RAR_U40_24	核苷酸
具有 9 RAR 和 24 nt 通用间隔子的 MG119-5 sgRNA	418	119-5_sgRNA_9bp_RAR_U40_24	核苷酸
具有 16 RAR 和 24 nt 通用间隔子的 MG119-5 sgRNA	419	119-5_sgRNA_16bp_RAR_U40_24	核苷酸
MG119 效应子	420	MG119-124 效应子	蛋白质
MG119 效应子	421	MG119-125 效应子	蛋白质
MG119 效应子	422	MG119-126 效应子	蛋白质
MG119 效应子	423	MG119-127 效应子	蛋白质
MG119 效应子	424	MG119-128 效应子	蛋白质
MG119 效应子	425	MG119-129 效应子	蛋白质
MG119 效应子	426	MG119-130 效应子	蛋白质
MG119 效应子	427	MG119-131 效应子	蛋白质
MG119 效应子	428	MG119-132 效应子	蛋白质
MG119 效应子	429	MG119-133 效应子	蛋白质
MG119 效应子	430	MG119-134 效应子	蛋白质
MG119 效应子	431	MG119-135 效应子	蛋白质
MG119-1 效应子 sgRNA1	432	MG119-1_sgRNA1	核苷酸
MG119-1 效应子 PAM（5'）	433	MG119-1 PAM（5'）	核苷酸
MG119-2 效应子 sgRNA1	434	MG119-2_sgRNA1_突变体 1	核苷酸
MG119-2 效应子 PAM（5'）	435	MG119-2 PAM（5'）	核苷酸
MG119-3 效应子 sgRNA1	436	MG119-3_sgRNA1_突变体 1	核苷酸
MG119-3 效应子 PAM（5'）	437	MG119-3 PAM（5'）	核苷酸
MG119-4 效应子 sgRNA1	438	MG119-4_sgRNA1	核苷酸
MG119-4 效应子 PAM（5'）	439	MG119-4 PAM（5'）	核苷酸
MG119-10 效应子 sgRNA1	440	MG119-10_sgRNA1	核苷酸
MG119-10 效应子 PAM（5'）	441	MG119-10 PAM（5'）	核苷酸
MG119-19 效应子 sgRNA1	442	MG119-19_sgRNA1	核苷酸
MG119-19 效应子 PAM（5'）	443	MG119-19 PAM（5'）	核苷酸
MG119-27 效应子 sgRNA2	444	MG119-27_sgRNA2_突变体 2	核苷酸

[0317]

目录号	SEQ ID NO	描述	类型
MG119-27 效应子 PAM（5'）	445	MG119-27 PAM（5'）	核苷酸
MG119-28 效应子 sgRNA2	446	MG119-28_sgRNA2	核苷酸
MG119-28 效应子 PAM（5'）	447	MG119-28 PAM（5'）	核苷酸
MG119-32 效应子 sgRNA1	448	MG119-32_sgRNA1	核苷酸
MG119-32 效应子 PAM（5'）	449	MG119-32 PAM（5'）	核苷酸
MG119-54 效应子 sgRNA1	450	MG119-54_sgRNA1	核苷酸
MG119-54 效应子 PAM（5'）	451	MG119-54 PAM（5'）	核苷酸
MG119-64 效应子 sgRNA2	452	MG119-64_sgRNA2	核苷酸
MG119-64 效应子 PAM（5'）	453	MG119-64 PAM（5'）	核苷酸
MG119-72 效应子 sgRNA1	454	MG119-72_sgRNA1	核苷酸
MG119-72 效应子 PAM（5'）	455	MG119-72 PAM（5'）	核苷酸
MG119-83 效应子 sgRNA1	456	MG119-83_sgRNA1	核苷酸
MG119-83 效应子 PAM（5'）	457	MG119-83 PAM（5'）	核苷酸
MG119-97 效应子 sgRNA1	458	MG119-97_sgRNA1_突变体 1	核苷酸
MG119-97 效应子 PAM（5'）	459	MG119-97 PAM（5'）	核苷酸
MG119-109 效应子 sgRNA1	460	MG119-109_sgRNA1	核苷酸
MG119-109 效应子 PAM（5'）	461	MG119-109 PAM（5'）	核苷酸
MG119-118 效应子 sgRNA1	462	MG119-118_sgRNA1_突变体 2	核苷酸
MG119-118 效应子 PAM（5'）	463	MG119-118 PAM（5'）	核苷酸
MG119-121 效应子 sgRNA1	464	MG119-121_sgRNA1_突变体 1	核苷酸
MG119-121 效应子 PAM（5'）	465	MG119-121 PAM（5'）	核苷酸
MG119-125 效应子 sgRNA1	466	MG119-125_sgRNA1	核苷酸
MG119-125 效应子 PAM（5'）	467	MG119-125 PAM（5'）	核苷酸
MG119-128 效应子 sgRNA1	468	MG119-128_sgRNA2_突变体 1	核苷酸
MG119-128 效应子 PAM（5'）	469	MG119-128 PAM（5'）	核苷酸
MG119-129 效应子 sgRNA1	470	MG119-129_sgRNA1_突变体 1	核苷酸
MG119-129 效应子 PAM（5'）	471	MG119-129 PAM（5'）	核苷酸
MG119-133 效应子 sgRNA1	472	MG119-133_sgRNA1_突变体 1	核苷酸
MG119-133 效应子 PAM（5'）	473	MG119-133 PAM（5'）	核苷酸
MG119-136 效应子 sgRNA1	474	MG119-136_sgRNA1_突变体 2	核苷酸
MG119-136 效应子 PAM（5'）	475	MG119-136 PAM（5'）	核苷酸
MG119-136 活性效应子	476	MG119-136 效应子	蛋白质
MG119-139 效应子	477	MG119-139 效应子	蛋白质
MG119-140 效应子	478	MG119-140 效应子	蛋白质
MG119-141 效应子	479	MG119-141 效应子	蛋白质
MG119-142 效应子	480	MG119-142 效应子	蛋白质

[0318]

目录号	SEQ ID NO	描述	类型
MG119-143 效应子	481	MG119-143 效应子	蛋白质
MG119-144 效应子	482	MG119-144 效应子	蛋白质
MG119-145 效应子	483	MG119-145 效应子	蛋白质
MG119-146 效应子	484	MG119-146 效应子	蛋白质
MG119-147 效应子	485	MG119-147 效应子	蛋白质
MG119-148 效应子	486	MG119-148 效应子	蛋白质
MG119-149 效应子	487	MG119-149 效应子	蛋白质
MG119-150 效应子	488	MG119-150 效应子	蛋白质
MG119-151 效应子	489	MG119-151 效应子	蛋白质
MG119-152 效应子	490	MG119-152 效应子	蛋白质
MG119-153 效应子	491	MG119-153 效应子	蛋白质
MG119-154 效应子	492	MG119-154 效应子	蛋白质
MG119-155 效应子	493	MG119-155 效应子	蛋白质
MG119-156 效应子	494	MG119-156 效应子	蛋白质
MG119-157 效应子	495	MG119-157 效应子	蛋白质
MG119-158 效应子	496	MG119-158 效应子	蛋白质
MG119-159 效应子	497	MG119-159 效应子	蛋白质
MG119-160 效应子	498	MG119-160 效应子	蛋白质
MG119-161 效应子	499	MG119-161 效应子	蛋白质
MG119-162 效应子	500	MG119-162 效应子	蛋白质
MG119-163 效应子	501	MG119-163 效应子	蛋白质
MG119-164 效应子	502	MG119-164 效应子	蛋白质
MG119-165 效应子	503	MG119-165 效应子	蛋白质
MG119-166 效应子	504	MG119-166 效应子	蛋白质
MG119-167 效应子	505	MG119-167 效应子	蛋白质
MG119-168 效应子	506	MG119-168 效应子	蛋白质
MG119-169 效应子	507	MG119-169 效应子	蛋白质
MG119-170 效应子	508	MG119-170 效应子	蛋白质
MG119-171 效应子	509	MG119-171 效应子	蛋白质
MG119-172 效应子	510	MG119-172 效应子	蛋白质
MG119-173 效应子	511	MG119-173 效应子	蛋白质
MG119-174 效应子	512	MG119-174 效应子	蛋白质
MG119-175 效应子	513	MG119-175 效应子	蛋白质
MG119-176 效应子	514	MG119-176 效应子	蛋白质
MG119-177 效应子	515	MG119-177 效应子	蛋白质
MG119-178 效应子	516	MG119-178 效应子	蛋白质

[0319]

目录号	SEQ ID NO	描述	类型
MG119-179 效应子	517	MG119-179 效应子	蛋白质
MG119-180 效应子	518	MG119-180 效应子	蛋白质
MG119-181 效应子	519	MG119-181 效应子	蛋白质
MG119-182 效应子	520	MG119-182 效应子	蛋白质
MG119-183 效应子	521	MG119-183 效应子	蛋白质
MG119-184 效应子	522	MG119-184 效应子	蛋白质
MG119-185 效应子	523	MG119-185 效应子	蛋白质
MG119-186 效应子	524	MG119-186 效应子	蛋白质
MG119-187 效应子	525	MG119-187 效应子	蛋白质
MG119-188 效应子	526	MG119-188 效应子	蛋白质
MG119-189 效应子	527	MG119-189 效应子	蛋白质
MG119-190 效应子	528	MG119-190 效应子	蛋白质
MG119-191 效应子	529	MG119-191 效应子	蛋白质
MG119-192 效应子	530	MG119-192 效应子	蛋白质
MG119-193 效应子	531	MG119-193 效应子	蛋白质
MG119-194 效应子	532	MG119-194 效应子	蛋白质
MG119-195 效应子	533	MG119-195 效应子	蛋白质
MG119-196 效应子	534	MG119-196 效应子	蛋白质
MG119-197 效应子	535	MG119-197 效应子	蛋白质
MG119-198 效应子	536	MG119-198 效应子	蛋白质
MG119-199 效应子	537	MG119-199 效应子	蛋白质
MG119-200 效应子	538	MG119-200 效应子	蛋白质
MG119-201 效应子	539	MG119-201 效应子	蛋白质
MG119-202 效应子	540	MG119-202 效应子	蛋白质
MG119-203 效应子	541	MG119-203 效应子	蛋白质
MG119-204 效应子	542	MG119-204 效应子	蛋白质
MG119-205 效应子	543	MG119-205 效应子	蛋白质
MG119-206 效应子	544	MG119-206 效应子	蛋白质
MG119-207 效应子	545	MG119-207 效应子	蛋白质
MG119-208 效应子	546	MG119-208 效应子	蛋白质
MG119-209 效应子	547	MG119-209 效应子	蛋白质
MG119-210 效应子	548	MG119-210 效应子	蛋白质
MG119-211 效应子	549	MG119-211 效应子	蛋白质
MG119-212 效应子	550	MG119-212 效应子	蛋白质
MG119-213 效应子	551	MG119-213 效应子	蛋白质
MG119-214 效应子	552	MG119-214 效应子	蛋白质

[0320]

目录号	SEQ ID NO	描述	类型
MG119-215 效应子	553	MG119-215 效应子	蛋白质
MG119-216 效应子	554	MG119-216 效应子	蛋白质
MG119-217 效应子	555	MG119-217 效应子	蛋白质
MG119-218 效应子	556	MG119-218 效应子	蛋白质
MG119-219 效应子	557	MG119-219 效应子	蛋白质
MG119-220 效应子	558	MG119-220 效应子	蛋白质
MG119-221 效应子	559	MG119-221 效应子	蛋白质
MG119-222 效应子	560	MG119-222 效应子	蛋白质
MG119-223 效应子	561	MG119-223 效应子	蛋白质
MG119-224 效应子	562	MG119-224 效应子	蛋白质
MG119-225 效应子	563	MG119-225 效应子	蛋白质
MG119-226 效应子	564	MG119-226 效应子	蛋白质
MG119-227 效应子	565	MG119-227 效应子	蛋白质
MG119-228 效应子	566	MG119-228 效应子	蛋白质
MG119-229 效应子	567	MG119-229 效应子	蛋白质
MG119-230 效应子	568	MG119-230 效应子	蛋白质
MG119-231 效应子	569	MG119-231 效应子	蛋白质
MG119-232 效应子	570	MG119-232 效应子	蛋白质
MG119-233 效应子	571	MG119-233 效应子	蛋白质
MG119-234 效应子	572	MG119-234 效应子	蛋白质
MG119-235 效应子	573	MG119-235 效应子	蛋白质
MG119-236 效应子	574	MG119-236 效应子	蛋白质
MG119-237 效应子	575	MG119-237 效应子	蛋白质
MG119-238 效应子	576	MG119-238 效应子	蛋白质
MG119-239 效应子	577	MG119-239 效应子	蛋白质
MG119-240 效应子	578	MG119-240 效应子	蛋白质
MG119-241 效应子	579	MG119-241 效应子	蛋白质
MG119-242 效应子	580	MG119-242 效应子	蛋白质
MG119-243 效应子	581	MG119-243 效应子	蛋白质
MG119-244 效应子	582	MG119-244 效应子	蛋白质
MG119-245 效应子	583	MG119-245 效应子	蛋白质
MG119-246 效应子	584	MG119-246 效应子	蛋白质
MG119-247 效应子	585	MG119-247 效应子	蛋白质
MG119-248 效应子	586	MG119-248 效应子	蛋白质
MG119-249 效应子	587	MG119-249 效应子	蛋白质
MG119-250 效应子	588	MG119-250 效应子	蛋白质

[0321]

目录号	SEQ ID NO	描述	类型
MG119-251 效应子	589	MG119-251 效应子	蛋白质
MG119-252 效应子	590	MG119-252 效应子	蛋白质
MG119-253 效应子	591	MG119-253 效应子	蛋白质
MG119-254 效应子	592	MG119-254 效应子	蛋白质
MG119-255 效应子	593	MG119-255 效应子	蛋白质
MG119-256 效应子	594	MG119-256 效应子	蛋白质
MG119-257 效应子	595	MG119-257 效应子	蛋白质
MG119-258 效应子	596	MG119-258 效应子	蛋白质
MG119-259 效应子	597	MG119-259 效应子	蛋白质
MG119-260 效应子	598	MG119-260 效应子	蛋白质
MG119-261 效应子	599	MG119-261 效应子	蛋白质
MG119-262 效应子	600	MG119-262 效应子	蛋白质
MG119-263 效应子	601	MG119-263 效应子	蛋白质
MG119-264 效应子	602	MG119-264 效应子	蛋白质
MG119-265 效应子	603	MG119-265 效应子	蛋白质
MG119-266 效应子	604	MG119-266 效应子	蛋白质
MG119-267 效应子	605	MG119-267 效应子	蛋白质
MG119-268 效应子	606	MG119-268 效应子	蛋白质
MG119-269 效应子	607	MG119-269 效应子	蛋白质
MG119-270 效应子	608	MG119-270 效应子	蛋白质
MG119-271 效应子	609	MG119-271 效应子	蛋白质
MG119-272 效应子	610	MG119-272 效应子	蛋白质
MG119-273 效应子	611	MG119-273 效应子	蛋白质
MG119-274 效应子	612	MG119-274 效应子	蛋白质
MG119-275 效应子	613	MG119-275 效应子	蛋白质
MG119-276 效应子	614	MG119-276 效应子	蛋白质
MG119-277 效应子	615	MG119-277 效应子	蛋白质
MG119-278 效应子	616	MG119-278 效应子	蛋白质
MG119-279 效应子	617	MG119-279 效应子	蛋白质
MG119-280 效应子	618	MG119-280 效应子	蛋白质
MG119-281 效应子	619	MG119-281 效应子	蛋白质
MG119-282 效应子	620	MG119-282 效应子	蛋白质
MG119-283 效应子	621	MG119-283 效应子	蛋白质
MG119-284 效应子	622	MG119-284 效应子	蛋白质
MG119-285 效应子	623	MG119-285 效应子	蛋白质
MG119-286 效应子	624	MG119-286 效应子	蛋白质

[0322]

目录号	SEQ ID NO	描述	类型
sgRNA	625	119-28 sgRNA1_小鼠_Alb	核苷酸（RNA）
sgRNA	626	119-28 sgRNA2_小鼠_Alb	核苷酸（RNA）
sgRNA	627	119-28 sgRNA3_小鼠_Alb	核苷酸（RNA）
sgRNA	628	119-28 sgRNA4_小鼠_Alb	核苷酸（RNA）
MG119 效应子	629	MG119-137 效应子	蛋白质

[0323] 表11-本文提及的蛋白质和核酸序列

[0324]

目录号	SEQ ID NO	描述	类型	生物体	其它信息	序列
MG119 效应子	629	MG119-137 效应子	蛋白质	未知	未经培养的生物体	MSKDKYVITRKIKLLP VGGENEVDRVYDFIRN GQYSQYQALNLLMGQ LASKYYDCKKDLSSAE FKDAQKSILSNSNPNLC DIEFVKGCDTKSAVVQ KVRQDFSTAIKNGLPR GERNITNYKRTVPLITR GRDLVFVHGYENYTEF LDNLYTDRNLKVFIKW VNKIQFKIVFGNPYKSA ELRSVVQNIFEERYKIN GSSICIDDDDIILNLSLT MPKEIKELDESKVVGV DLGIAIPAVCALNTNSY SRKSIGSADDFLRVRTK IRAQRRRLQKSLSQTSG GHGRKKKLRALDKFSE YEKHWVQNYNHYVSK QVVDFAIKNNAKYINL EDLEGYGEEEKNKFILS NWSYYQLQQYIAYKAE KYGIEVRKINPYHTSQV CSCCGHWESGQRVNQ KTFICKNPECENFGEEV NADFNAARNIALSTNW SDIDEKKNKKNKKK

[0325] 虽然本文已经示出和描述本发明的优选实施例，但对于本领域技术人员明显的是此类实施例仅借助于实例提供。本发明不旨在受说明书中提供的具体实施例的限制。虽然已参考前述说明书描述本发明，但本文实施例的描述和说明不打算以限制性意义进行。在不背离本发明的情况下，本领域技术人员现在将想到许多变化、改变和替代。此外，应当理解，本发明的全部方面不限于本文所阐述的具体描述、配置或相对比例，其取决于各种条件和变量。应当理解，在实践本发明时可以采用本文所述的本发明的实施例的各种替代方案。因此，经考虑本发明应同样涵盖任何这类替代方案、修改、变型或等效物。以下权利要求书旨在限定本发明的范围以及由此覆盖在这些权利要求和其等同物的范围内的方法和结构。

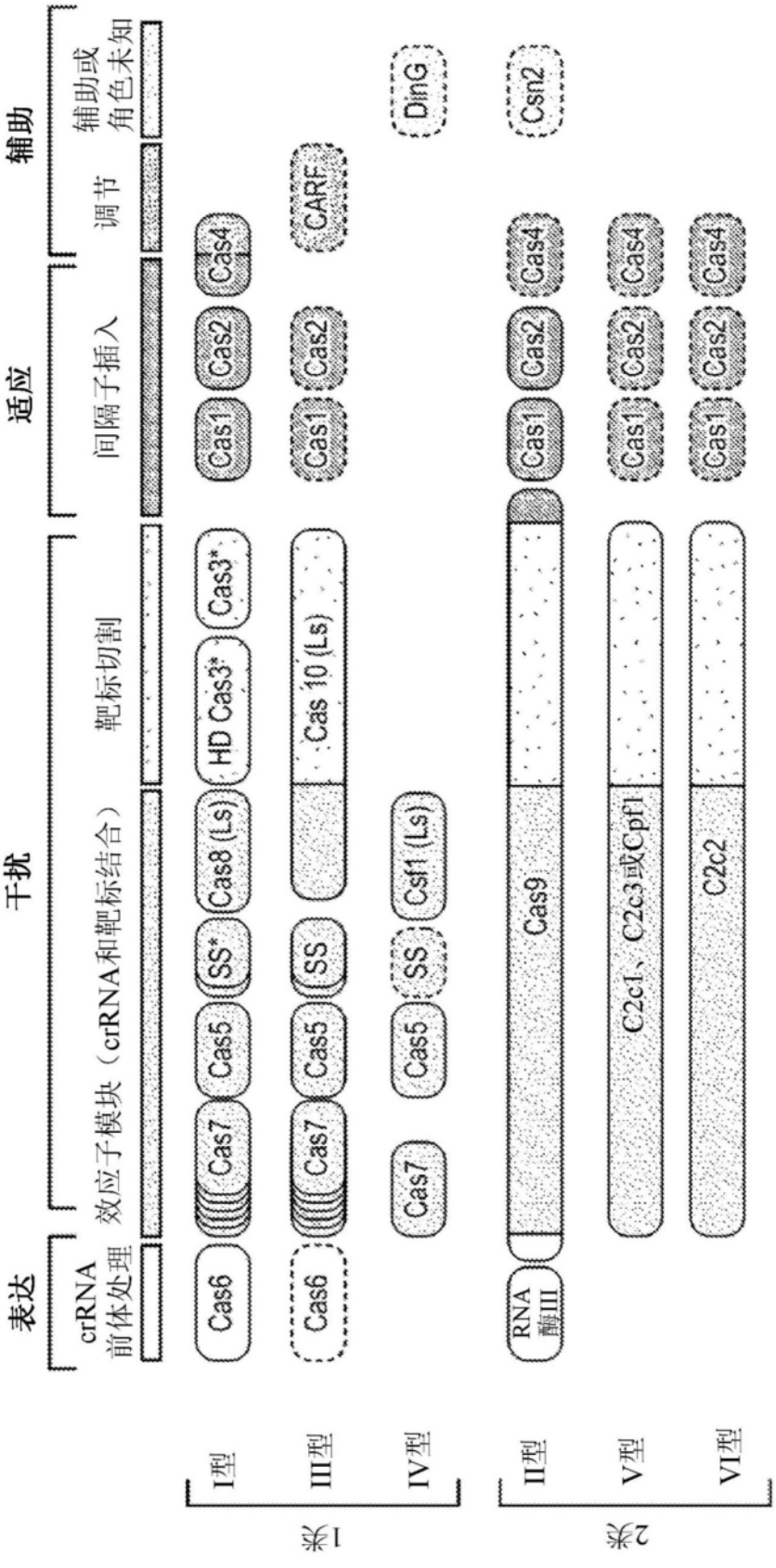

表达
干扰
适应
辅助
crRNA
前体处理
效应子模块（crRNA和靶标结合）
靶标切割
间隔子插入
调节
辅助或
角色未知
Cas6
Cas7
Cas5
SS*
Cas8 (Ls)
HD Cas3*
Cas3*
Cas1
Cas2
Cas4
Cas6
Cas7
Cas5
SS
Cas 10 (Ls)
Cas1
Cas2
CARF
Cas7
Cas5
SS
Csf1 (Ls)
DinG
RNA
酶III
Cas9
Cas1
Cas2
Cas4
Csn2
C2c1、C2c3或Cpf1
Cas1
Cas2
Cas4
C2c2
Cas1
Cas2
Cas4
I型
III型
IV型
II型
V型
VI型
1类
2类

图1

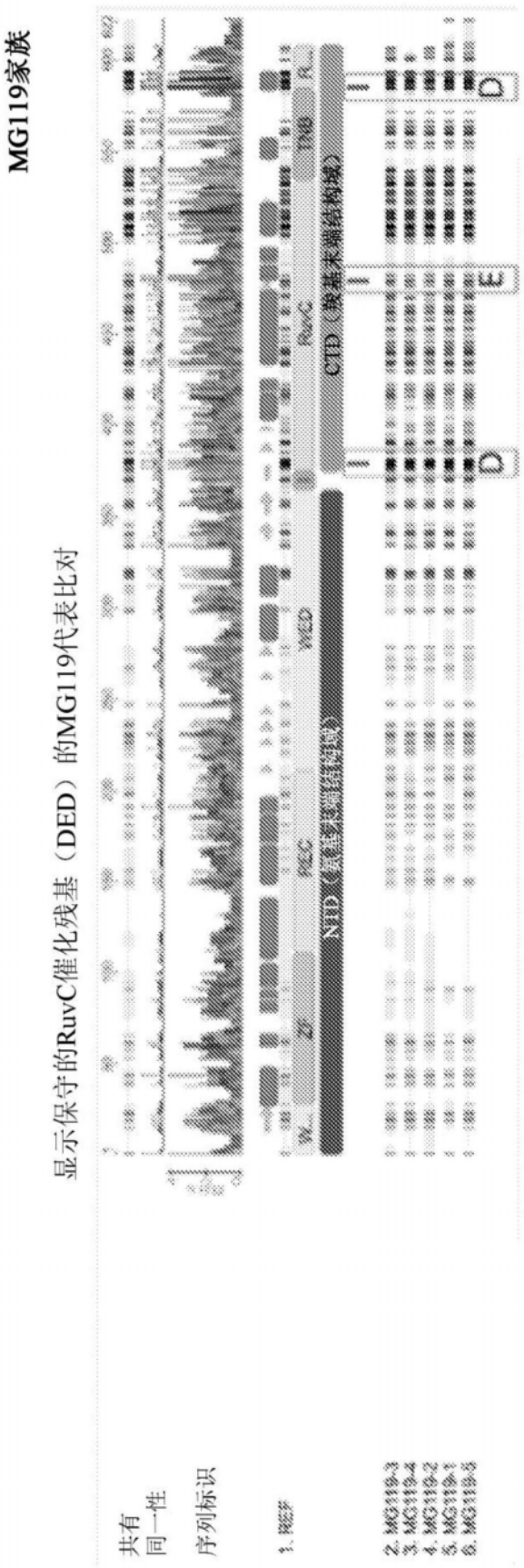

MG119家族
显示保守的RuvC催化残基（DED）的MG119代表比对
共有
同一性
序列标识
1. 靶DNA
2. MG119-3
3. MG119-4
4. MG119-2
5. MG119-1
6. MG119-5
NTD（氨基末端结构域）
CTD（羧基末端结构域）
RuvC
WED
PI
REC
ZF
TNB
D
E

图2A

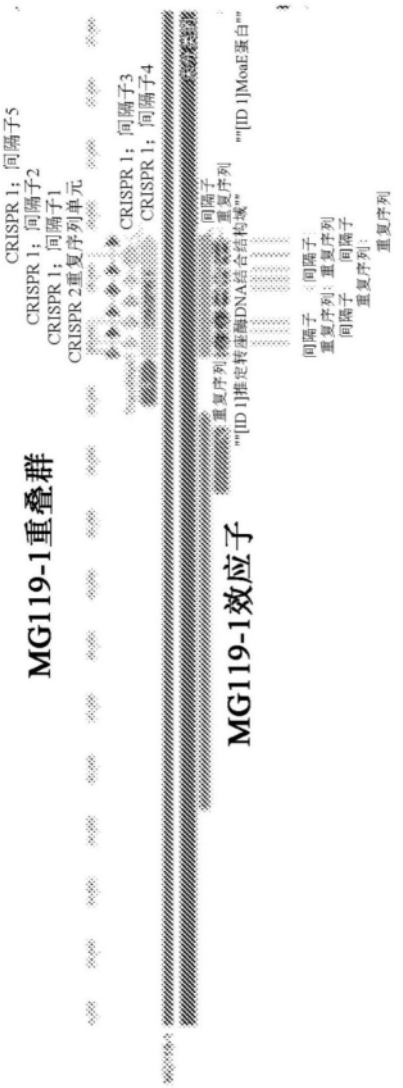

MG119-1重叠群
CRISPR 1；间隔子5
CRISPR 1；间隔子2
CRISPR 1；间隔子1
CRISPR 2重复序列单元
CRISPR 1；间隔子3
CRISPR 1；间隔子4
MG119-1效应子
""[ID 1]推定转座酶DNA结合结构域""
""[ID 1]MoaE蛋白""

图2B

MG119-1直接重复序列

图2C

图2D

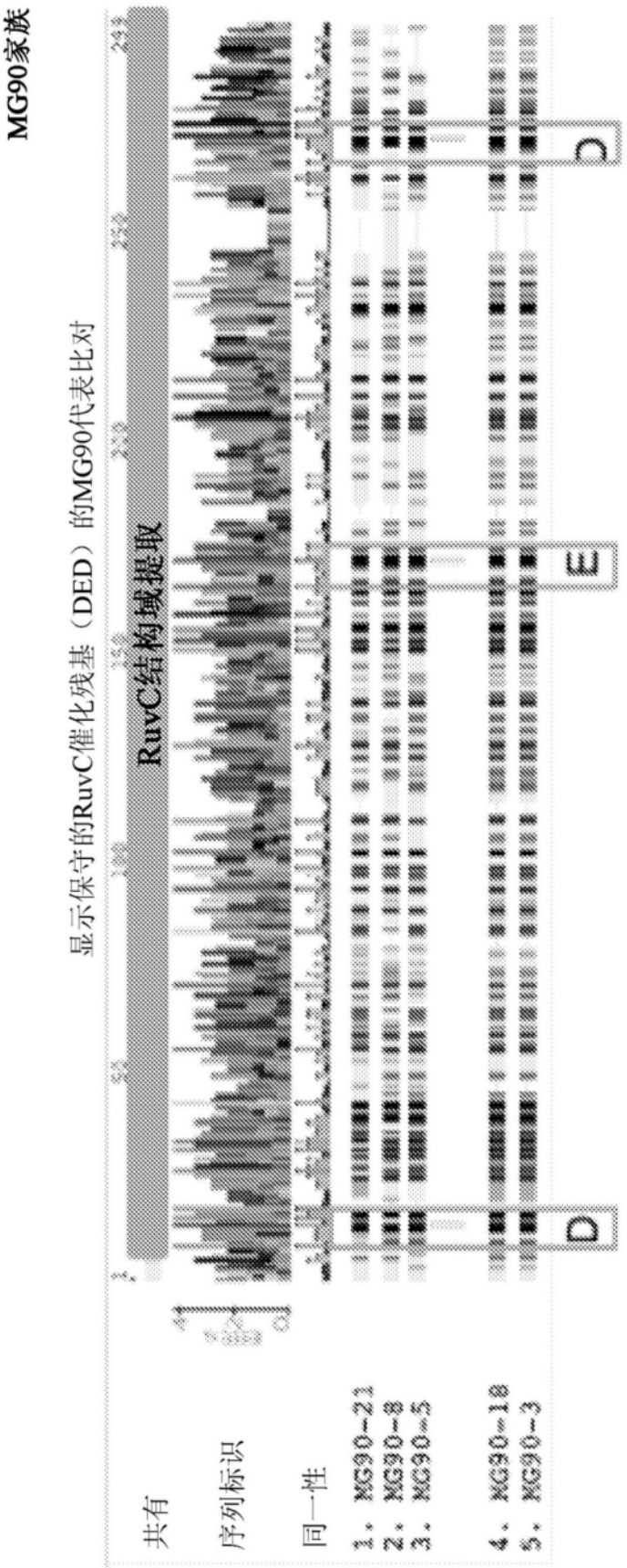

MG90家族
显示保守的RuvC催化残基（DED）的MG90代表比对
RuvC结构域提取
共有
序列标识
同一性
1. MG90-21
2. MG90-8
3. MG90-5
4. MG90-18
5. MG90-3
D
E
D

图3A

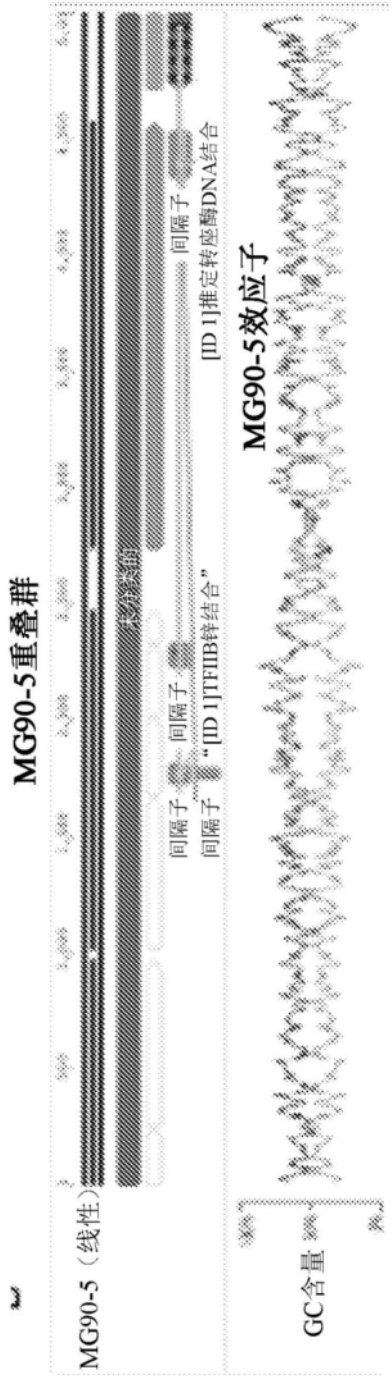

MG90-5重叠群
MG90-5（线性）
间隔子
间隔子
“[ID 1]TFIIB锌结合”
间隔子
间隔子
[ID 1]推定转座酶DNA结合
MG90-5效应子
GC含量

图3B

图3C

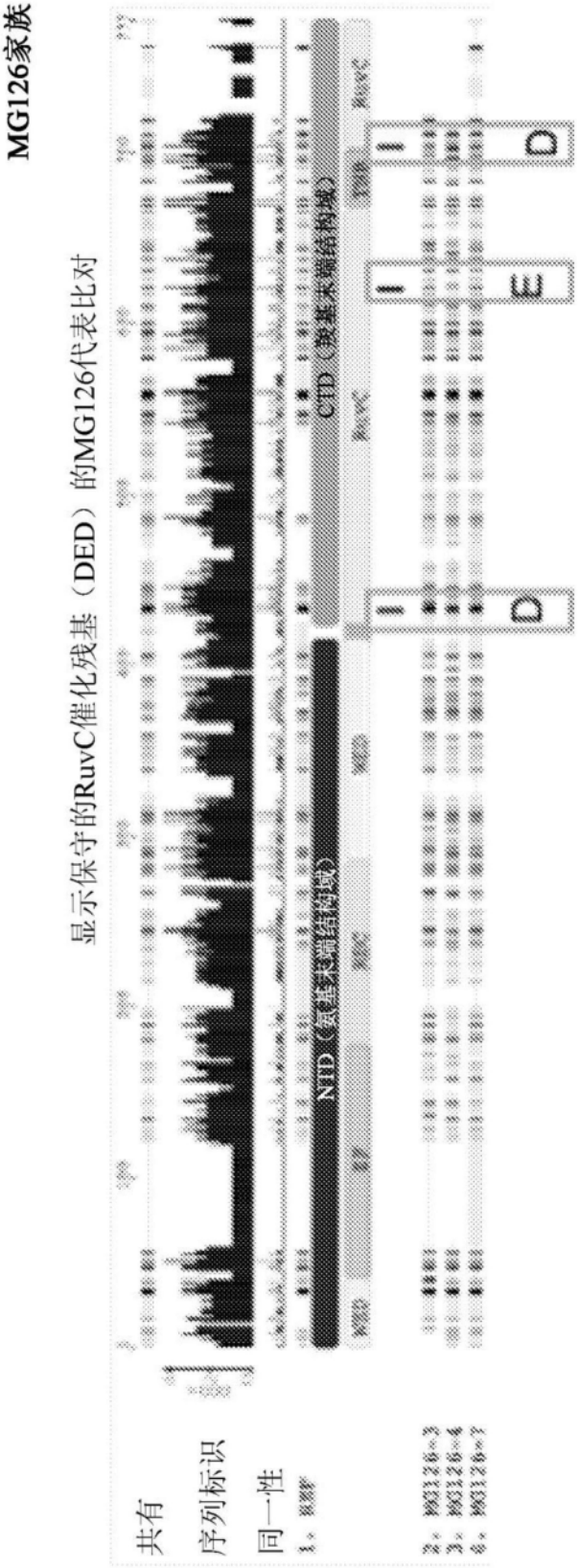

MG126家族
显示保守的RuvC催化残基（DED）的MG126代表比对
共有
序列标识
同一性
NTD（氨基末端结构域）
CTD（羧基末端结构域）
D
E
D

图4A

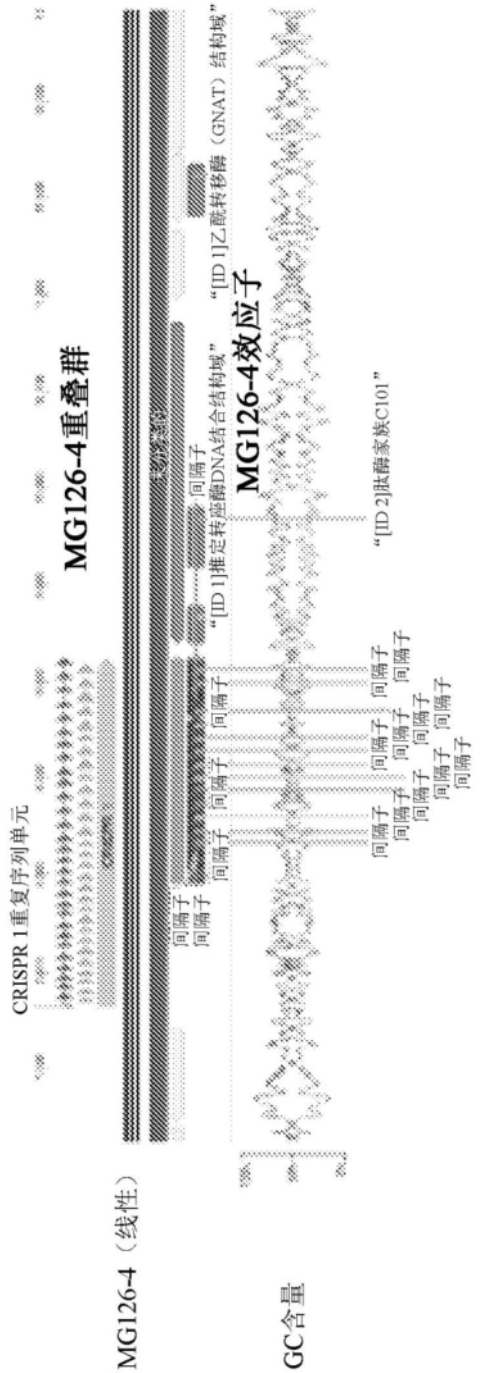

MG126-4（线性）
GC含量
CRISPR 1重复序列单元
MG126-4重叠群
间隔子
“[ID 1]推定转座酶DNA结合结构域”
“[ID 1]乙酰转移酶（GNAT）结构域”
MG126-4效应子
“[ID 2]肽酶家族C101”

图4B

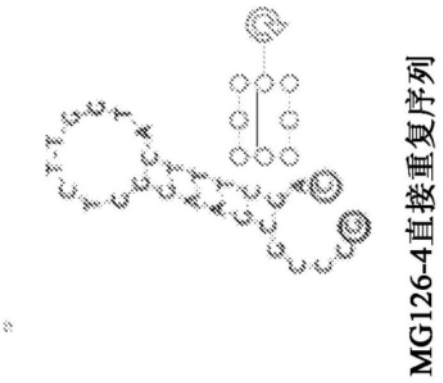


图4C

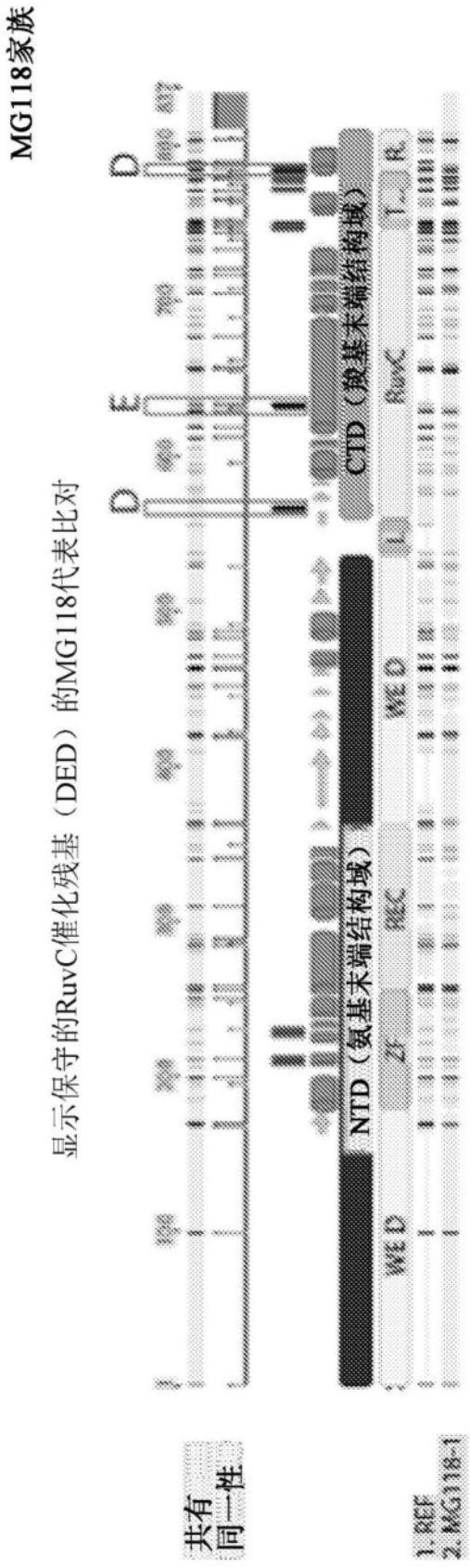

MG118家族
显示保守的RuvC催化残基（DED）的MG118代表比对
NTD（氨基末端结构域）
CTD（酸基末端结构域）
WED
ZF
REC
RuvC
共有
同一性
1. REF
2. MG118-1

图5A

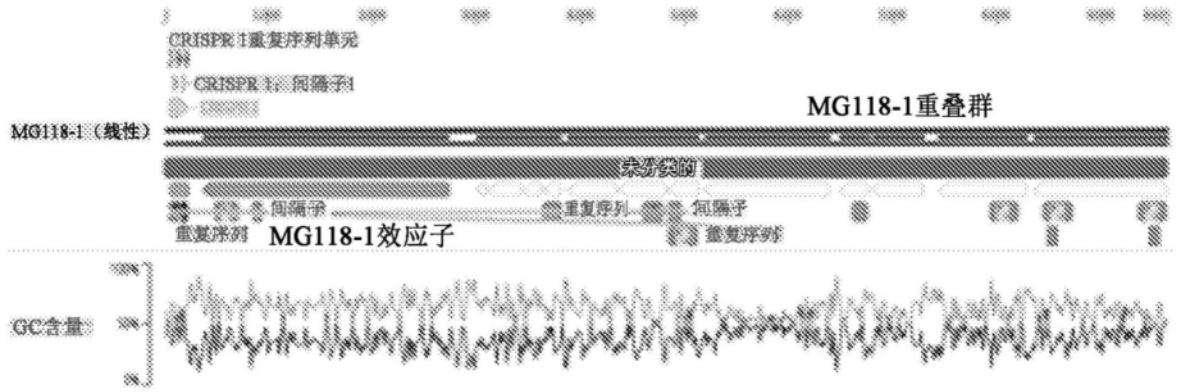


图5B

MG118-1直接重复序列

图5C

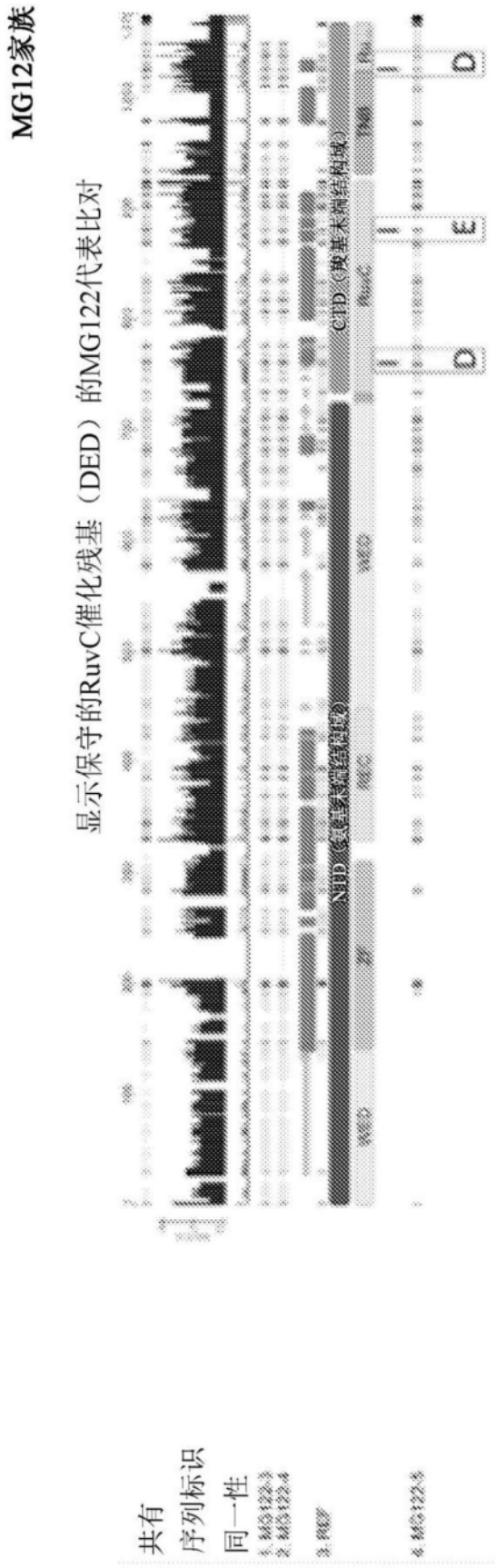

MG12家族
显示保守的RuvC催化残基（DED）的MG122代表比对
共有
序列标识
同一性
NTD（无臂末端结构域）
CTD（楔基末端结构域）
RuvC
D
E
D

图6A

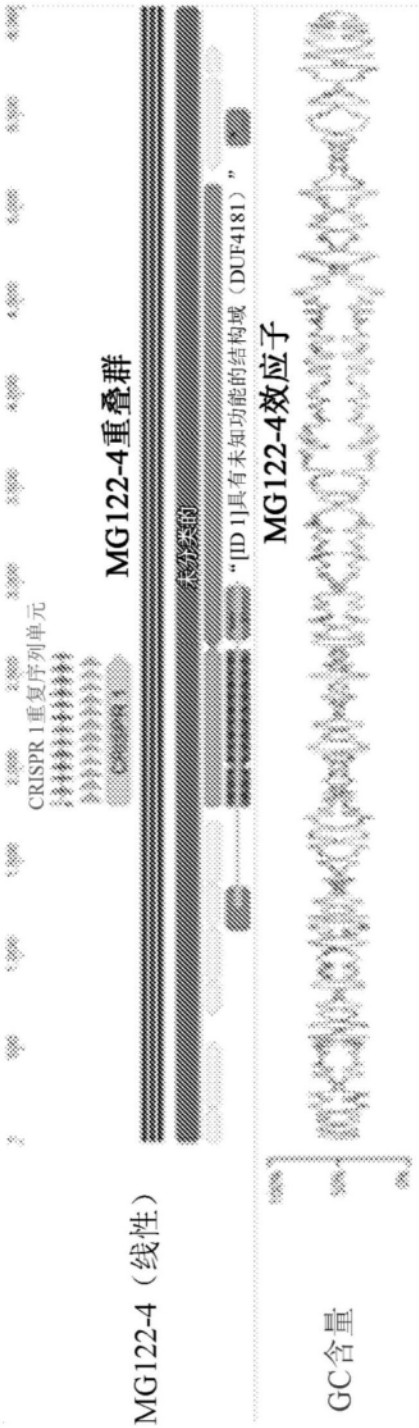

MG122-4（线性）
GC含量
CRISPR 1重复序列单元
MG122-4重叠群
“[ID 1]具有未知功能的结构域（DUF4181）”
MG122-4效应子

图6B

图6C

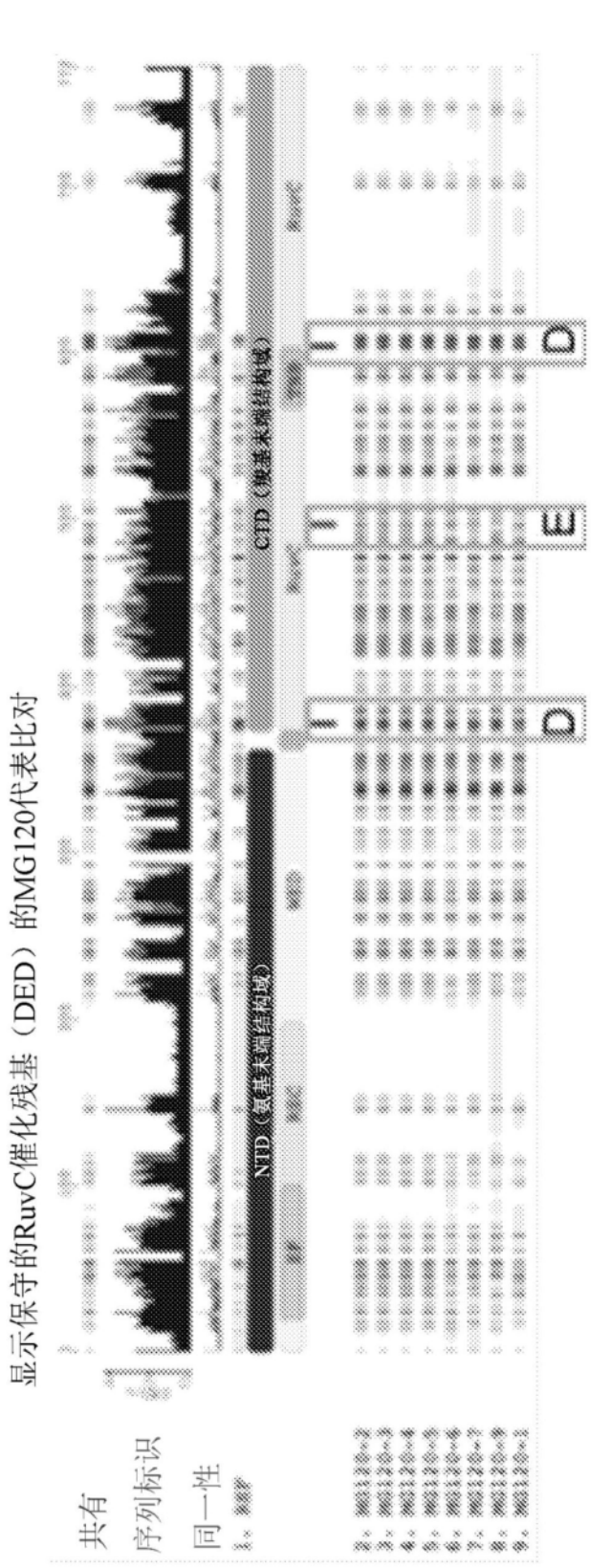

MG120家族
显示保守的RuvC催化残基（DED）的MG120代表比对
共有
序列标识
同一性
NTD（氨基末端结构域）
CTD（羧基末端结构域）
RuvC
D
E
D

图7A

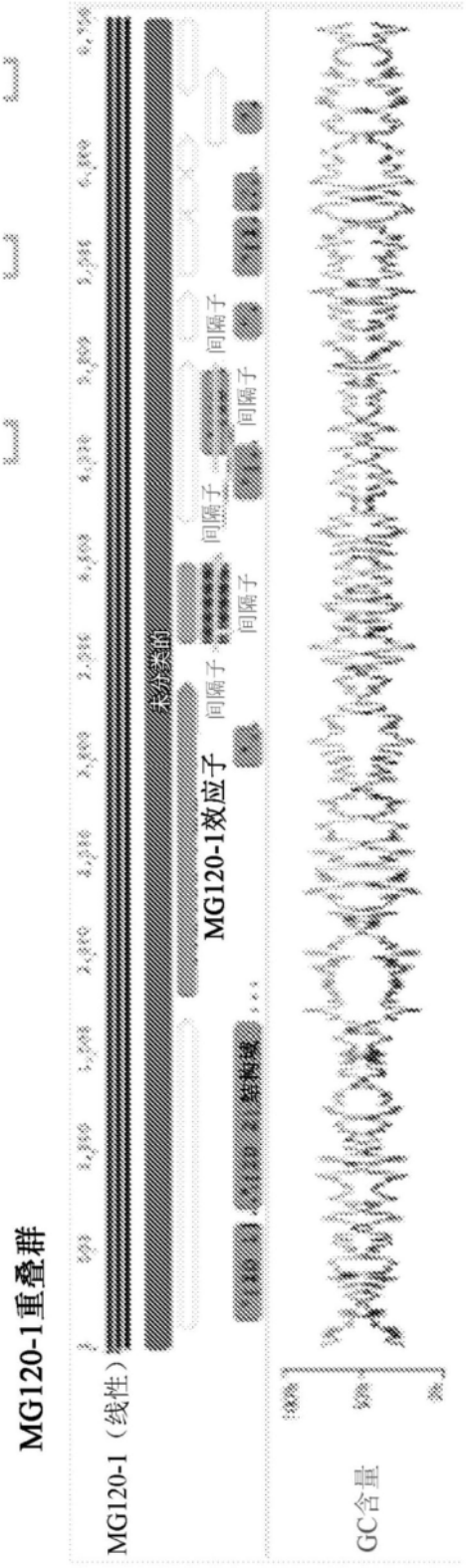


图7B

图7C

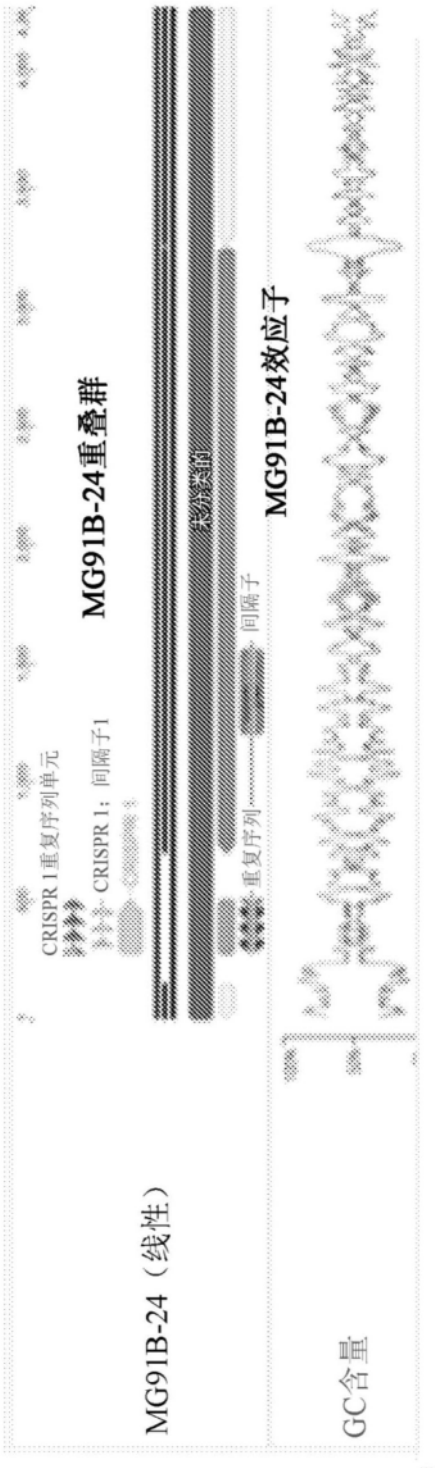

MG91B-24（线性）
GC含量
CRISPR 1重复序列单元
CRISPR 1：间隔子1
MG91B-24重叠群
未分类的
重复序列
间隔子
MG91B-24效应子

图8A

图8B

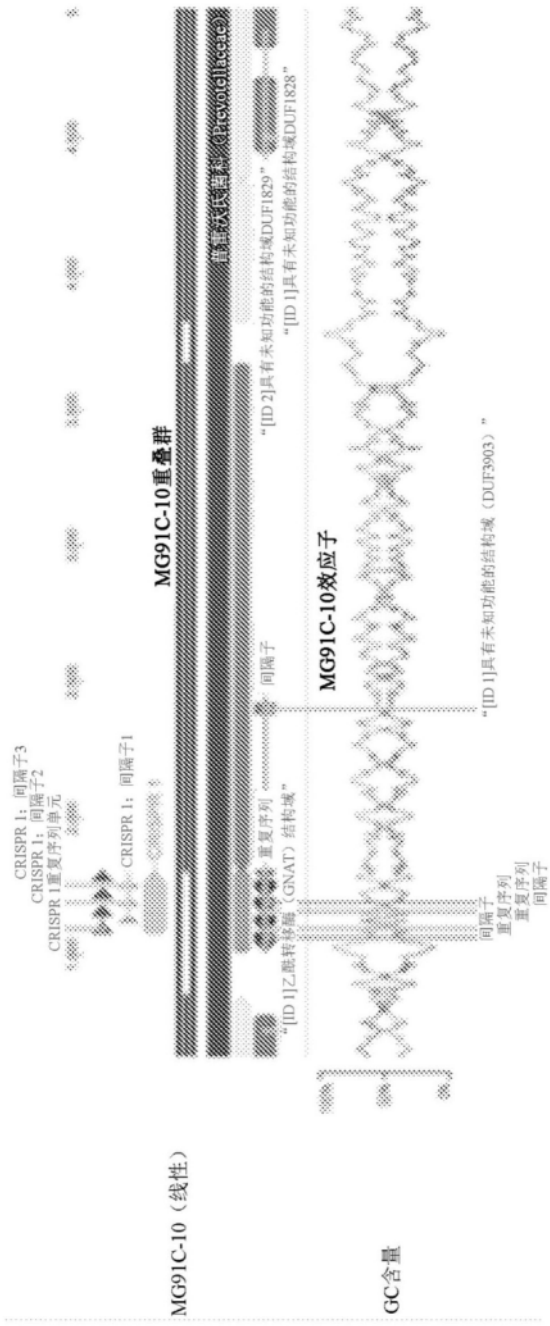

MG91C-10重叠群
普雷沃氏菌科（Prevotellaceae）
CRISPR 1：间隔子3
CRISPR 1：间隔子2
CRISPR 1重复序列单元
CRISPR 1：间隔子1
"[ID 1]乙酰转移酶（GNAT）结构域"
重复序列
间隔子
"[ID 2]具有未知功能的结构域DUF1829"
"[ID 1]具有未知功能的结构域DUF1828"
MG91C-10效应子
"[ID 1]具有未知功能的结构域（DUF3903）"
间隔子
重复序列
重复序列
间隔子
MG91C-10（线性）
GC含量

图8C

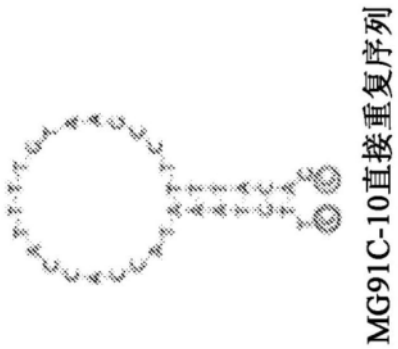

MG91C-10直接重复序列

图8D

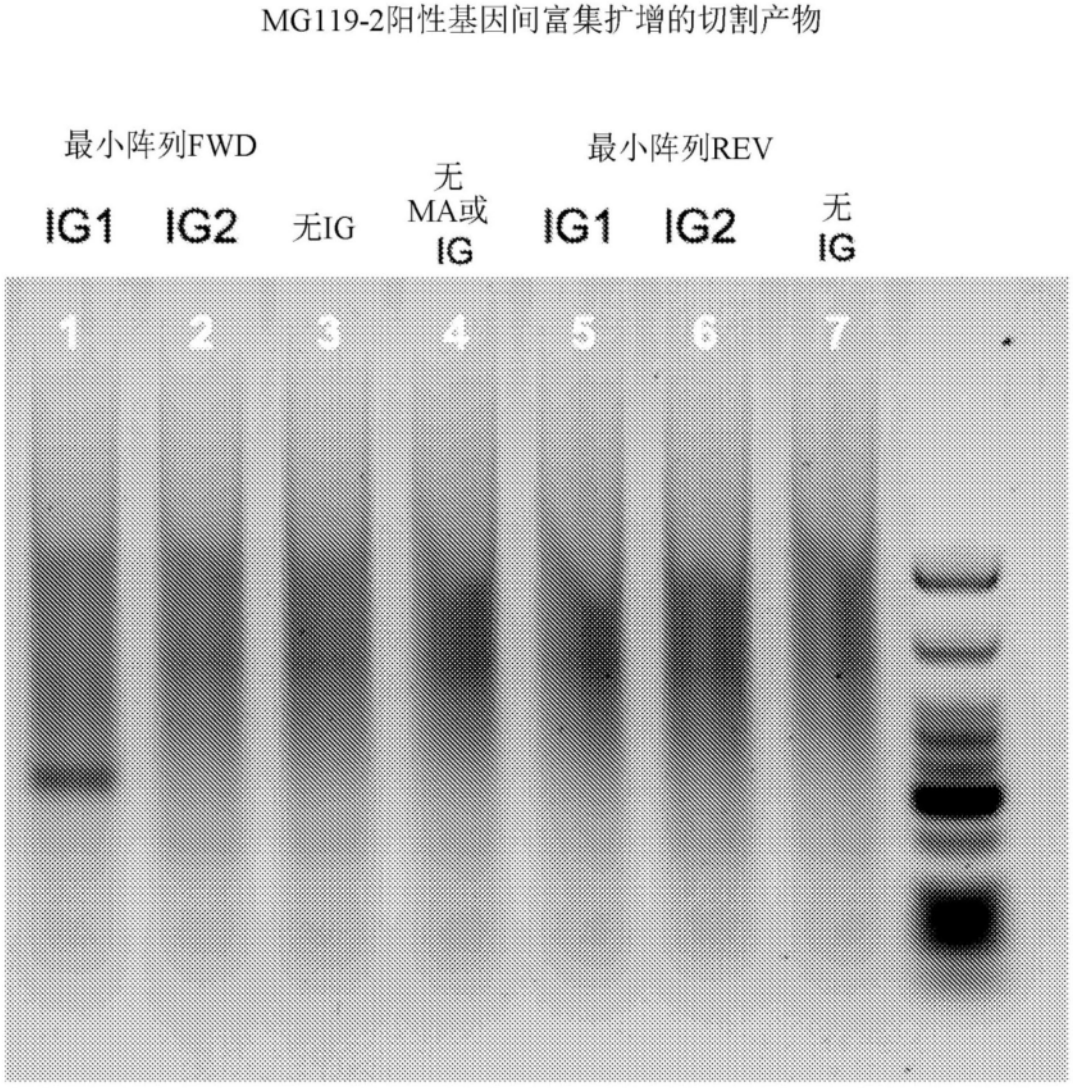

MG119-2阳性基因间富集扩增的切割产物
最小阵列FWD
最小阵列REV
IG1
IG2
无IG
无MA或IG
IG1
IG2
无IG
1
2
3
4
5
6
7

图9

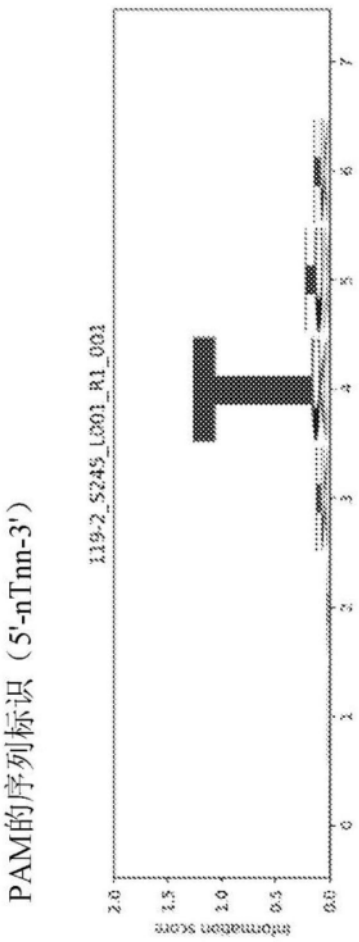

PAM的序列标识（5'-nTnn-3'）
119-2_S245_L001_R1_001
information score
2.0
1.5
1.0
0.5
0.0
0
1
2
3
4
5
6
7

图10A

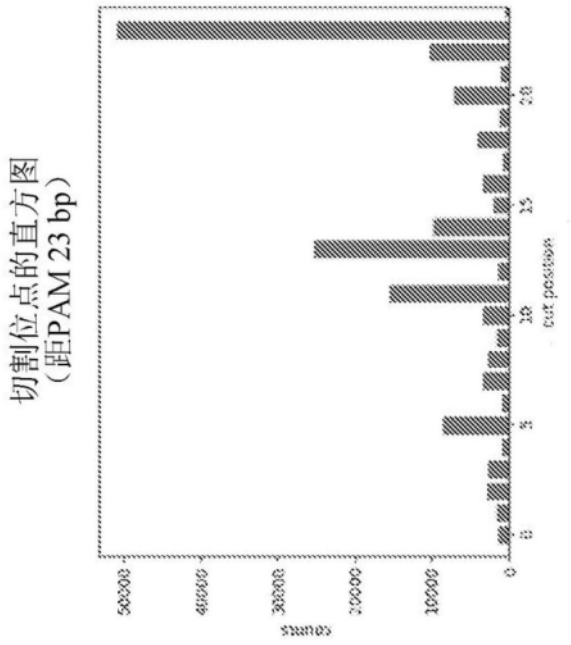

切割位点的直方图
（距PAM 23 bp）
counts
cut position
500000
400000
300000
200000
100000
0
5
10
15
20

图10B

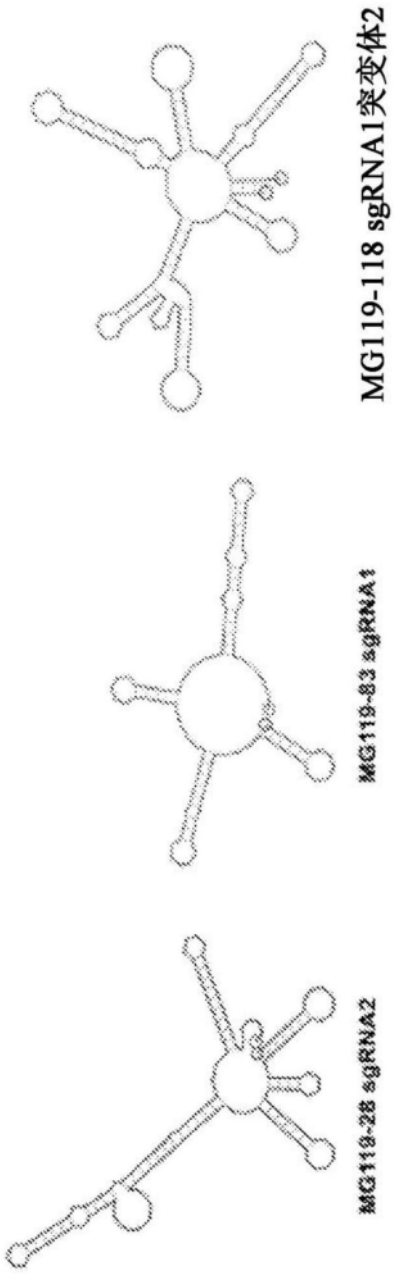


图11A

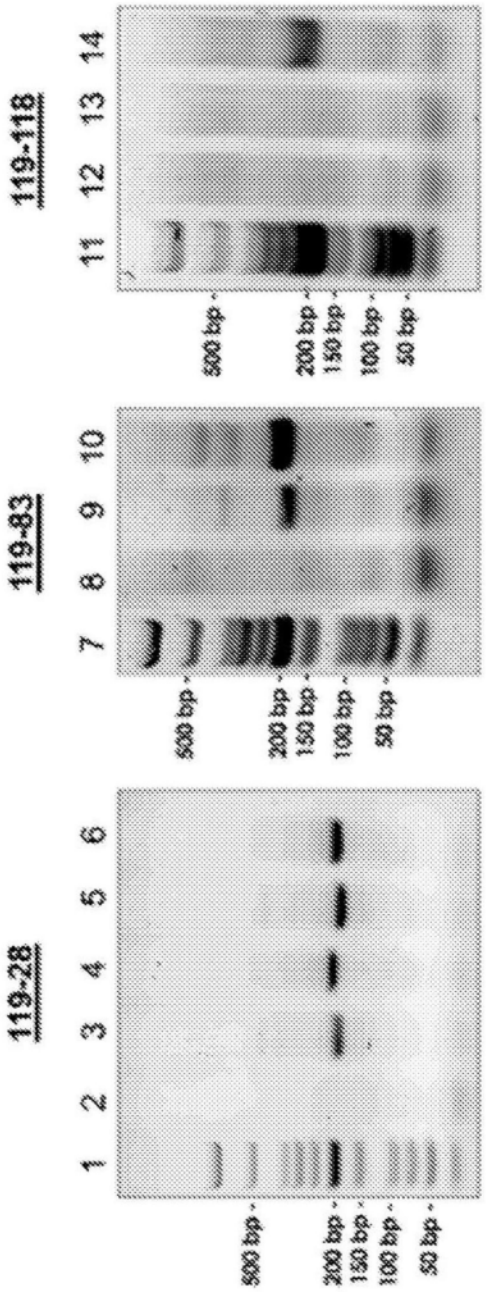

119-118
11
12
13
14
500 bp
200 bp
150 bp
100 bp
50 bp
119-83
7
8
9
10
500 bp
200 bp
150 bp
100 bp
50 bp
119-28
1
2
3
4
5
6
500 bp
200 bp
150 bp
100 bp
50 bp

图11B

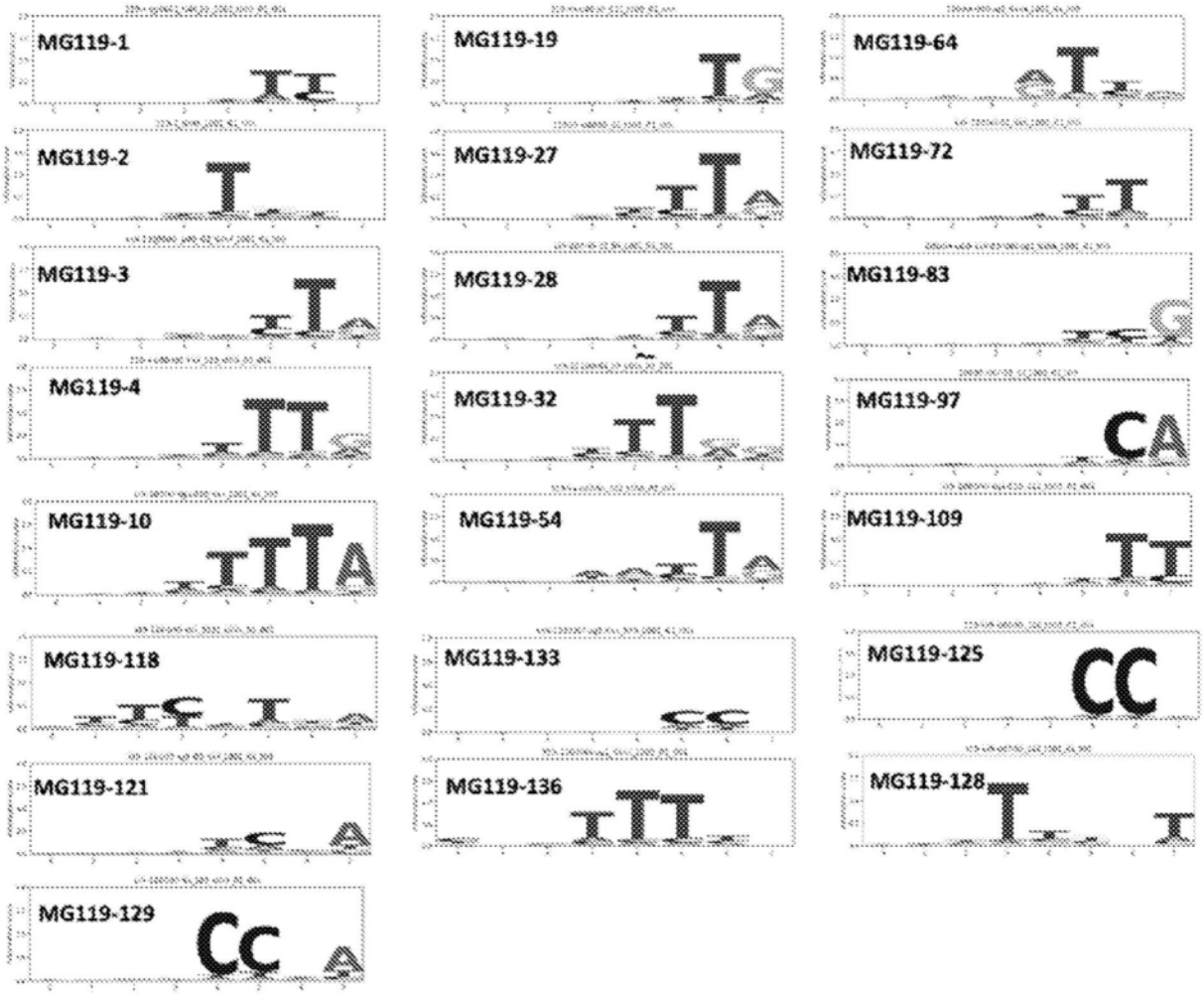

MG119-1
MG119-19
MG119-64
MG119-2
MG119-27
MG119-72
MG119-3
MG119-28
MG119-83
MG119-4
MG119-32
MG119-97
MG119-10
MG119-54
MG119-109
MG119-118
MG119-133
MG119-125
MG119-121
MG119-136
MG119-128
MG119-129

图12

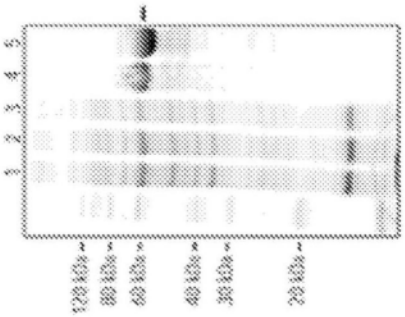


图13A

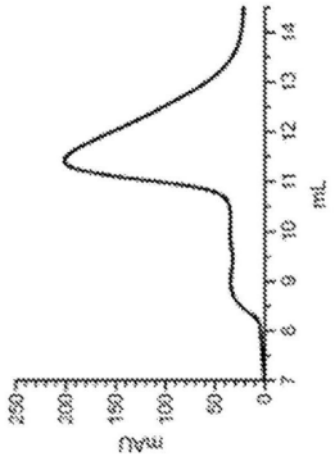


图13B

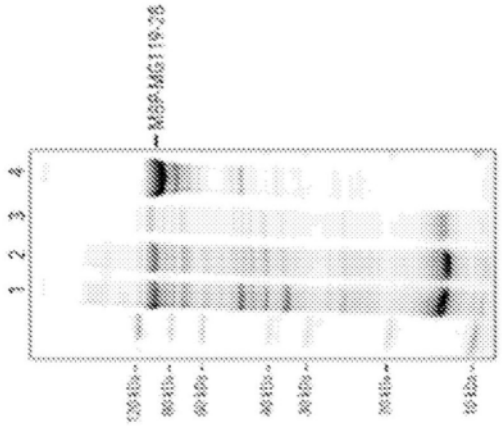


图13C

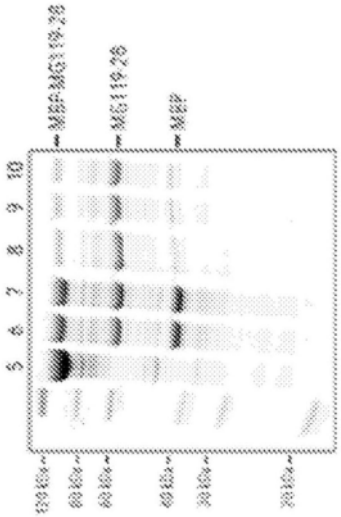


图13D

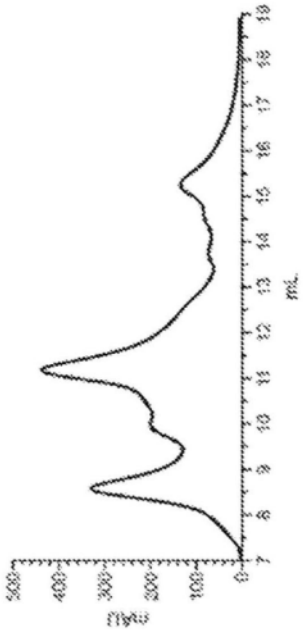


图13E

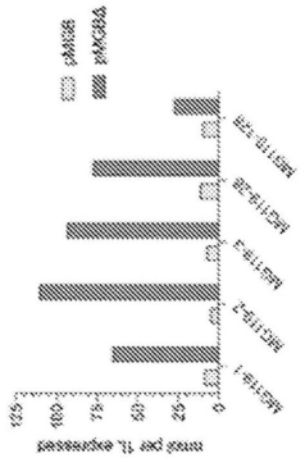


图13F

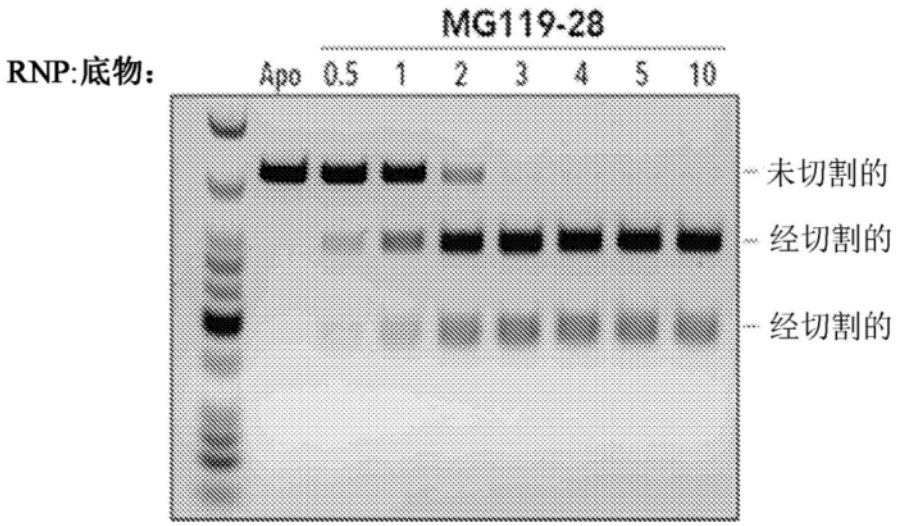

MG119-28
RNP:底物：
Apo 0.5 1 2 3 4 5 10
未切割的
经切割的
经切割的

图14A

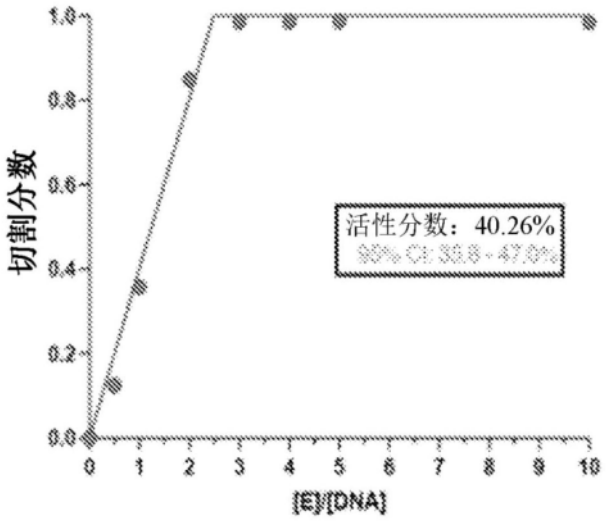

切割分数
1.0
0.8
0.6
0.4
0.2
0.0
0 1 2 3 4 5 6 7 8 9 10
[E]/[DNA]
活性分数：40.26%
95% CI: 33.8 - 47.0%

图14B

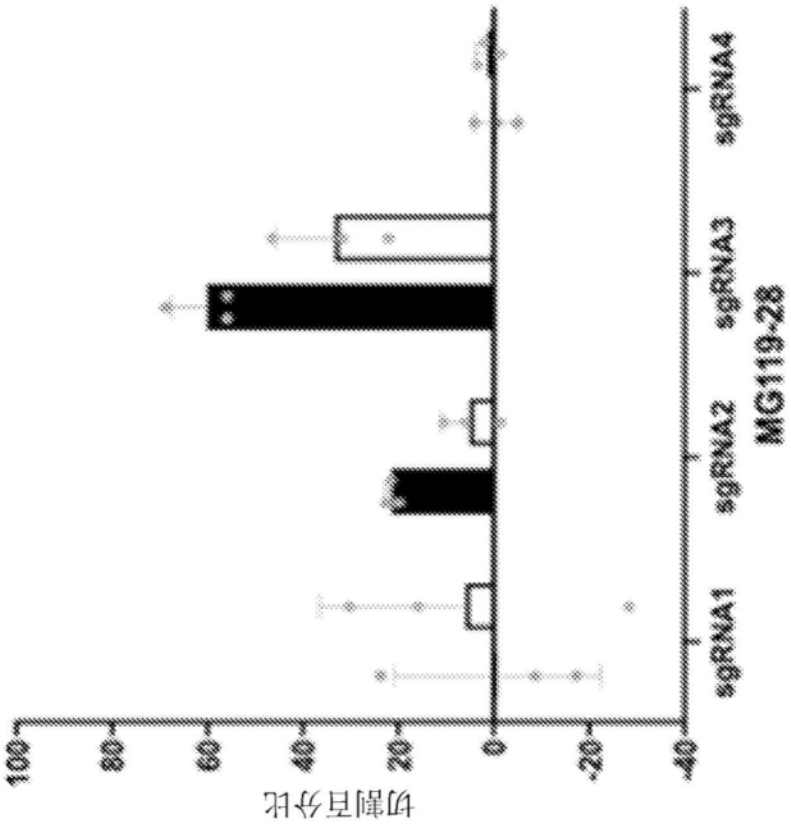

sgRNA4
sgRNA3
sgRNA2
sgRNA1
MG119-28
100
80
60
40
20
0
20
40
切割百分比

图15A

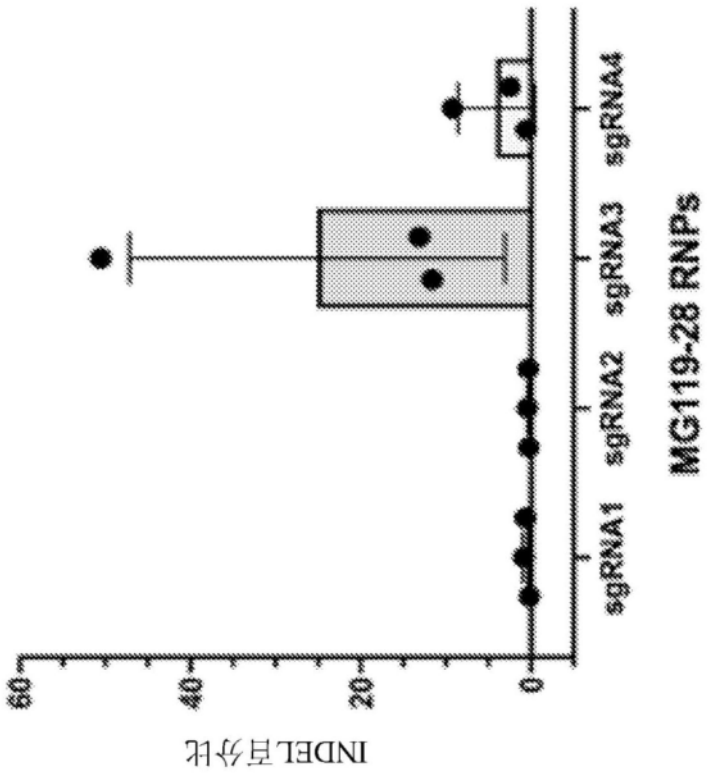

INDEL百分比
60
40
20
0
sgRNA1
sgRNA2
sgRNA3
sgRNA4
MG119-28 RNPs

图15B