(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 116096885 A
(43)申请公布日 2023.05.09

(21)申请号 202180034169.7

(22)申请日 2021.03.17

(30)优先权数据
62/991,403 2020.03.18 US

(85)PCT国际申请进入国家阶段日
2022.11.09

(86)PCT国际申请的申请数据
PCT/US2021/022840 2021.03.17

(87)PCT国际申请的公布数据
WO2021/188729 EN 2021.09.23

(71)申请人 斯克里贝治疗公司
地址 美国加利福尼亚州

(72)发明人 B•奥克斯 H•施平纳 S•丹尼 B•T•斯特尔 K•泰勒 K•巴尼 I•科林 M•阿迪尔 C•乌尔内斯 S•希金斯

(74)专利代理机构 北京坤瑞律师事务所 11494
专利代理师 封新琴

(51)Int.Cl.
C12N 15/113(2006.01)
C12N 9/22(2006.01)

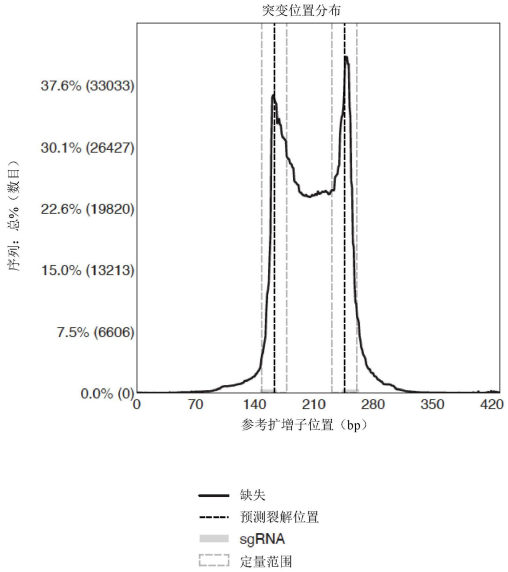
权利要求书12页 说明书192页
序列表（电子公布） 附图18页

(54)发明名称

用于靶向C9orf72的组合物和方法

(57)摘要

本文提供能够用于修饰C9orf72基因的2类V型系统，其包含核酸酶、引导核酸(gNA)和任选的供体模板核酸。所述系统还能够用于引入到细胞，例如具有所述C9orf72基因中的突变或重复的真核细胞中。还提供使用所述系统来修饰具有所述突变或重复的细胞的方法。



CN 116096885 A

1.一种系统,其包含2类V型CRISPR蛋白质和引导核酸(gNA),其中所述gNA包含与包含染色体9开放阅读框架72(C9orf72)基因的靶核酸序列互补的靶向序列。

2.根据权利要求1所述的系统,其中所述C9orf72基因包含一个或多个突变。

3.根据权利要求1所述的系统,其中所述C9orf72基因突变在六核苷酸重复序列扩增段(HRS)中包含六核苷酸重复序列GGGGCC的多于30、多于100、多于500、多于700、多于1000或多于1600个副本。

4.根据权利要求2或权利要求3所述的系统,其中所述突变为功能丧失突变。

5.根据权利要求2或权利要求3所述的系统,其中所述突变为功能增强突变。

6.根据前述权利要求中任一项所述的系统,其中所述gNA为引导RNA(gRNA)。

7.根据权利要求1到5中任一项所述的系统,其中所述gNA为引导DNA(gDNA)。

8.根据权利要求1到5中任一项所述的系统,其中所述gNA为包含DNA和RNA的嵌合体。

9.根据权利要求1到8中任一项所述的系统,其中所述gNA为单分子gNA(sgNA)。

10.根据权利要求1到8中任一项所述的系统,其中所述gNA为双分子gNA(dgNA)。

11.根据权利要求1到10中任一项所述的系统,其中所述gNA的所述靶向序列包含选自由以下组成的组的序列:SEQ ID NO:309-343、363-2100和2295-21835,或与其具有至少约65%、至少约75%、至少约85%或至少约95%同一性的序列。

12.根据权利要求1到10中任一项所述的系统,其中所述gNA的所述靶向序列包含选自由以下组成的组的序列:SEQ ID NO:309-343、363-2100和2295-21835。

13.根据权利要求1到10中任一项所述的系统,其中所述gNA的所述靶向序列包含选自由以下组成的组的序列:SEQ ID NO:309-343、363-2100和2295-21835,其中从所述序列的3'端去除单个核苷酸。

14.根据权利要求1到10中任一项所述的系统,其中所述gNA的所述靶向序列包含选自由以下组成的组的序列:SEQ ID NO:309-343、363-2100和2295-21835,其中从所述序列的3'端去除两个核苷酸。

15.根据权利要求1到10中任一项所述的系统,其中所述gNA的所述靶向序列包含选自由以下组成的组的序列:SEQ ID NO:309-343、363-2100和2295-21835,其中从所述序列的3'端去除三个核苷酸。

16.根据权利要求1到10中任一项所述的系统,其中所述gNA的所述靶向序列包含选自由以下组成的组的序列:SEQ ID NO:309-343、363-2100和2295-21835,其中从所述序列的3'端去除四个核苷酸。

17.根据权利要求1到10中任一项所述的系统,其中所述gNA的所述靶向序列包含选自由以下组成的组的序列:SEQ ID NO:309-343、363-2100和2295-21835,其中从所述序列的3'端去除五个核苷酸。

18.根据权利要求1到17中任一项所述的系统,其中所述gNA的所述靶向序列包含具有相对于选自由SEQ ID NO:309-343、363-2100和2295-21835组成的组的序列的一种或多种单核苷酸多态性(SNP)的序列。

19.根据权利要求1到18中任一项所述的系统,其中所述gNA的所述靶向序列与所述C9orf72基因的非编码区互补。

20.根据权利要求1到18中任一项所述的系统,其中所述gNA的所述靶向序列与所述

C9orf72基因的蛋白质编码区互补。

21.根据权利要求1到18中任一项所述的系统,其中所述gNA的所述靶向序列与C9orf72外显子的序列互补。

22.根据权利要求1到18中任一项所述的系统,其中所述gNA的所述靶向序列与C9orf72内含子的序列互补。

23.根据权利要求1到18中任一项所述的系统,其中所述gNA的所述靶向序列与C9orf72内含子-外显子接界的序列互补。

24.根据权利要求1到18中任一项所述的系统,其中所述gNA的所述靶向序列与C9orf72调节元件的序列互补。

25.根据权利要求1到18中任一项所述的系统,其中所述gNA的所述靶向序列与所述C9orf72基因的基因间区的序列互补。

26.根据权利要求1到18中任一项所述的系统,其中所述gNA的所述靶向序列与在所述HRS的5'处的序列互补。

27.根据权利要求26所述的系统,其中所述gNA的所述靶向序列与所述C9orf72基因的内含子1或启动子的序列互补。

28.根据权利要求1到27中任一项所述的系统,其进一步包含第二gNA,其中所述第二gNA具有与以下互补的靶向序列:所述靶核酸序列相较于所述gNA的所述靶向序列的不同或重叠部分。

29.根据权利要求27所述的系统,其中所述第二gNA的所述靶向序列与在所述HRS的5'或3'处的序列互补。

30.根据权利要求27所述的系统,其中第一gNA的所述靶向序列为针对在所述HRS的5'处的序列,且所述第二gNA的所述靶向序列与在所述HRS的3'处的序列互补。

31.根据权利要求29所述的系统,其中所述gNA的所述靶向序列与所述C9orf72基因的内含子1的序列互补。

32.根据权利要求1到31中任一项所述的系统,其中所述gNA具有支架,所述支架包含选自由以下组成的组的序列:SEQ ID NO:4-16和2101-2294,或与其具有至少约50%、至少约60%、至少约70%、至少约80%、至少约90%、至少约95%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的序列。

33.根据权利要求1到31中任一项所述的系统,其中所述gNA具有支架,所述支架包含选自由以下组成的组的序列:SEQ ID NO:2101-2294。

34.根据权利要求1到31中任一项所述的系统,其中所述gNA具有支架,所述支架包含相对于参考gNA序列具有至少一种修饰的序列,所述参考gNA序列选自由SEQ ID NO:4-16的序列组成的组。

35.根据权利要求34所述的系统,其中所述参考gNA的所述至少一种修饰包含至少一个gNA序列的核苷酸的取代、缺失或插入。

36.根据权利要求1到35中任一项所述的系统,其中所述gNA经化学修饰。

37.根据权利要求1到36中任一项所述的系统,其中所述2类V型CRISPR蛋白质包含具有SEQ ID NO:1-3中任一个的序列的参考CasX蛋白,具有SEQ ID NO:49-150、233-235、238-252、272-281中任一个的序列的CasX变异蛋白质,或与其具有至少约50%、至少约60%、至

少约70％、至少约80％、至少约90％、至少约95％、至少约96％、至少约97％、至少约98％或至少约99％序列同一性的序列。

38.根据权利要求1到36中任一项所述的系统，其中所述2类V型CRISPR蛋白质包含CasX变异蛋白质，所述CasX变异蛋白质包含相对于具有选自SEQ ID NO:1-3的序列的参考CasX蛋白的至少一种修饰。

39.根据权利要求38所述的系统，其中所述至少一种修饰包含相对于所述参考CasX蛋白的所述CasX变异蛋白质的结构域中的至少一个氨基酸取代、缺失或插入。

40.根据权利要求39所述的系统，其中所述结构域选自由非靶链结合(NTSB)结构域、靶链负载(TSL)结构域、螺旋形I结构域、螺旋形II结构域、寡核苷酸结合结构域(OBD)和RuvC DNA裂解结构域组成的组。

41.根据权利要求37到40中任一项所述的系统，其中所述CasX蛋白进一步包含一个或多个核定位信号(NLS)。

42.根据权利要求41所述的系统，其中所述一个或多个NLS选自由以下组成的序列的组：PKKKRKV(SEQ ID NO:165)、KRPAATKKAGQAKKKK(SEQ ID NO:166)、PAAKRVKLD(SEQ ID NO:167)、RQRRNELKRSP(SEQ ID NO:168)、NQSSNFGPMKGGNFGGRSSGP YGGGGQYFAKPRNQGGY(SEQ ID NO:169)、RMRIZFKNKGKDTAELRRRRVEVSVEL RKAKKDEQILKRRNV(SEQ ID NO:170)、VSRKRPRP(SEQ ID NO:171)、PPKKARED(SEQ ID NO:172)、PQPKKKPL(SEQ ID NO:173)、SALIKKKKKMAP(SEQ ID NO:174)、DRLRR(SEQ ID NO:175)、PKQKKRK(SEQ ID NO:176)、RKLKKKIKKL(SEQ ID NO:177)、REKKKFLKRR(SEQ ID NO:178)、KRKGDEVDGVDEVAKKKSKK(SEQ ID NO:179)、RKCLQAGMNLEARKTKK(SEQ ID NO:180)、PRPRKIPR(SEQ ID NO:181)、PPRKKRTVV(SEQ ID NO:182)、NLSKKKKRKREK(SEQ ID NO:183)、RRPSRPFRKP(SEQ ID NO:184)、KRPRSPSS(SEQ ID NO:185)、KRGINDRNFWRGENERKTR(SEQ ID NO:186)、PRPPKMARYDN(SEQ ID NO:187)、KRSFSKAF(SEQ ID NO:188)、KLKIKRPVK(SEQ ID NO:189)、PKTRRRPRRSQRKRPPT(SEQ ID NO:191)、RRKKRRPRRKKRR(SEQ ID NO:194)、PKKKSRKPKKKSRK(SEQ ID NO:195)、HKKKHPDASVNFSEFSK(SEQ ID NO:196)、QRPGPYDRPQRPGPYDRP(SEQ ID NO:197)、LSPSLSPLLSPSLSPL(SEQ ID NO:198)、RGKGGKGLGKGGAKRHRK(SEQ ID NO:199)、PKRGRGRPKRGRGR(SEQ ID NO:200)、MSRRRKANPTKLSENAKKLAKEVEN(SEQ ID NO:192)、PKKKRKVPPPPAAKRVKLD(SEQ ID NO:190)、PKKKRKVPPPPKKKRKV(SEQ ID NO:201)、PAKRARRGYKC(SEQ ID NO:202)、KLGPRKATGRW(SEQ ID NO:203)和PRRKREE(SEQ ID NO:204)。

43.根据权利要求41或权利要求42所述的系统，其中所述一个或多个NLS在所述CasX蛋白的C端处或附近。

44.根据权利要求41或权利要求42所述的系统，其中所述一个或多个NLS在所述CasX蛋白的N端处或附近。

45.根据权利要求41或权利要求42所述的系统，其中所述CasX蛋白包含至少两个NLS，所述NLS在所述CasX蛋白的N端和C端处或附近。

46.根据权利要求37到45中任一项所述的系统，其中所述2类V型CRISPR蛋白质能够与所述gNA形成核糖核蛋白质复合物(RNP)。

47.根据权利要求37到46中任一项所述的系统，其中所述CasX变异蛋白质和gNA变异体

展现相比于SEQ ID NO:1-3中任一个的参考CasX蛋白和SEQ ID NO:4-16中任一个的gNA的至少一种或更多种改进的特征。

48.根据权利要求47所述的系统,其中所述改进的特征选自由以下组成的组:CasX变异体的改进的折叠;对引导核酸(gNA)的改进的结合亲和力;对靶DNA的改进的结合亲和力;在靶DNA的编辑中利用较大范围的一种或多种包括ATC、CTC、GTC或TTC的PAM序列的改进的能力;所述靶DNA的改进的解旋;增加的编辑活性;改进的编辑效率;改进的编辑特异性;增加的核酸酶活性;增加的用于双链裂解的靶链负载;减少的用于单链切割的靶链负载;减少的脱靶裂解;改进的非靶DNA链的结合;改进的蛋白质稳定性;改进的蛋白质溶解度;改进的蛋白质:gNA复合物(RNP)稳定性;改进的蛋白质:gNA复合物溶解度;改进的蛋白质产率;改进的蛋白质表达;以及改进的融合特征。

49.根据权利要求47或权利要求48所述的系统,其中所述CasX变异蛋白质的所述改进的特征相对于SEQ ID NO:1、SEQ ID NO:2或SEQ ID NO:3的所述参考CasX蛋白改进至少约1.1到约100,000倍。

50.根据权利要求47或权利要求48所述的系统,其中所述CasX变异蛋白质的所述改进的特征相对于SEQ ID NO:1、SEQ ID NO:2或SEQ ID NO:3的所述参考CasX蛋白改进至少约10倍、至少约100倍、至少约1,000倍或至少约10,000倍。

51.根据权利要求47到50中任一项所述的系统,其中所述改进的特征包含编辑效率,且所述CasX变异蛋白质与所述gNA变异体的所述RNP包含相比于SEQ ID NO:2的所述参考CasX蛋白与包含SEQ ID NO:4-16中的任一个的所述gNA的所述RNP的1.1到100倍编辑效率改进。

52.根据权利要求47到51中任一项所述的系统,其中在细胞分析系统中,当所述PAM序列TTC、ATC、GTC或CTC中的任一个定位于与所述gNA的所述靶向序列具有同一性的非靶链序列5'的1个核苷酸处时,包含所述CasX变异体和所述gNA变异体的所述RNP,相比于类似分析系统中包含参考CasX蛋白和参考gNA的RNP的编辑效率和/或结合,展现更高的靶DNA中编辑效率和/或靶序列结合。

53.根据权利要求52所述的系统,其中所述PAM序列为TTC。

54.根据权利要求53所述的系统,其中所述gNA的所述靶向序列包含选自由以下组成的组的序列:SEQ ID NO:5427-12893。

55.根据权利要求52所述的系统,其中所述PAM序列为ATC。

56.根据权利要求55所述的系统,其中所述gNA的所述靶向序列包含选自由以下组成的组的序列:SEQ ID NO:363-2100和2295-5426。

57.根据权利要求52所述的系统,其中所述PAM序列为CTC。

58.根据权利要求57所述的系统,其中所述gNA的所述靶向序列包含选自由以下组成的组的序列:SEQ ID NO:16203-21835。

59.根据权利要求52所述的系统,其中所述PAM序列为GTC。

60.根据权利要求59所述的系统,其中所述gNA的所述靶向序列包含选自由以下组成的组的序列:SEQ ID NO:12894-16202。

61.根据权利要求52到60中任一项所述的系统,其中相比于对于所述PAM序列的SEQ ID NO:1-3的所述CasX蛋白中的任一个的结合亲和力,对于一个或多个PAM序列的增加的结合亲和力为至少1.5倍更大。

62.根据权利要求52到61中任一项所述的系统，其中与包含SEQ ID NO:4-16中的任一个的所述参考CasX与所述参考gNA的RNP相比，所述RNP具有至少5%、至少10%、至少15%或至少20%更高百分比的裂解潜能RNP。

63.根据权利要求37到62中任一项所述的系统，其中所述CasX变异蛋白质包含具有切口酶活性的核酸酶结构域。

64.根据权利要求63所述的系统，其中所述CasX变异体能够仅裂解双链靶核酸分子的一条链。

65.根据权利要求37到62中任一项所述的系统，其中所述CasX变异蛋白质包含具有双链裂解活性的核酸酶结构域。

66.根据权利要求37到62中任一项所述的系统，其中所述CasX蛋白为非催化活性CasX(dCasX)蛋白质，并且其中所述dCasX和所述gNA保持与所述靶核酸序列结合的能力。

67.根据权利要求66所述的系统，其中所述dCasX包含以下残基处的突变：

a.对应于SEQ ID NO:1的所述参考CasX蛋白的D672、E769和/或D935;或

b.对应于SEQ ID NO:2的所述参考CasX蛋白的D659、E756和/或D922。

68.根据权利要求67所述的系统，其中所述突变为丙氨酸取代所述残基。

69.根据权利要求1到65中任一项所述的系统，其进一步包含供体模板核酸。

70.根据权利要求69所述的系统，其中所述供体模板包含核酸，所述核酸包含所述C9orf72基因的至少一部分，其中所述C9orf72基因部分选自由以下组成的组:C9orf72外显子、C9orf72内含子、C9orf72内含子-外显子接界、C9orf72调节元件、或其组合。

71.根据权利要求69所述的系统，其中所述供体模板包含GGGGCC序列的多个六核苷酸重复序列，其中重复序列的数目在10到约30个重复序列的范围内。

72.根据权利要求69到71中任一项所述的系统，其中所述供体模板包含与侧接所述靶核酸中的核酸酶裂解位点的序列互补的同源臂。

73.根据权利要求69到72中任一项所述的系统，其中所述供体模板包含与野生型C9orf72基因相比的一个或多个突变。

74.根据权利要求69到73中任一项所述的系统，其中所述供体模板包含与野生型C9orf72基因相比的异源序列。

75.根据权利要求69到72中任一项所述的系统，其中所述供体模板包含野生型C9orf72基因的全部或一部分。

76.根据权利要求69到75所述的系统，其中所述供体模板的大小在10到15,000个核苷酸的范围内。

77.根据权利要求69到76中任一项所述的系统，其中所述供体模板为单链DNA模板或单链RNA模板。

78.根据权利要求69到76中任一项所述的系统，其中所述供体模板为双链DNA模板。

79.一种核酸，其包含编码根据权利要求1到36中任一项所述的gNA、根据权利要求37到68中任一项所述的CasX、根据权利要求69到78中任一项所述的供体模板或其组合的序列。

80.根据权利要求79所述的核酸，其中编码所述CasX蛋白的序列经密码子优化以在真核细胞中表达。

81.一种载体，其包含根据权利要求79或权利要求80所述的核酸。

82.根据权利要求81所述的载体,其中所述载体进一步包含启动子。

83.根据权利要求81或权利要求82所述的载体,其中所述载体选自由以下组成的组:逆转录病毒载体、慢病毒载体、腺病毒载体、腺相关病毒(AAV)载体、单纯疱疹病毒(HSV)载体、病毒样粒子(VLP)、质粒、小环、纳米质粒和RNA载体。

84.根据权利要求83所述的载体,其中所述载体为AAV载体。

85.根据权利要求84所述的载体,其中所述AAV载体选自AAV1、AAV2、AAV3、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAV10、AAV11、AAV12、AAV 44.9、AAV-Rh74或AAVRh10。

86.根据权利要求83所述的载体,其中所述载体为逆转录病毒载体。

87.根据权利要求83所述的载体,其中所述载体为包含gag多聚蛋白的一个或多个组分的VLP载体。

88.根据权利要求87所述的载体,其中所述gag多聚蛋白的所述一种或多种组分选自由以下组成的组:基质蛋白质(MA)、核衣壳蛋白质(NC)、衣壳蛋白质(CA)、p1肽、p6肽、P2A肽、P2B肽、P10肽、p12肽、PP21/24肽、P12/P3/P8肽和P20肽。

89.根据权利要求87或权利要求88所述的载体,其中编码所述VLP的所述载体包含一种或多种编码所述gag多聚蛋白、所述CasX蛋白和所述gNA的核酸。

90.根据权利要求89所述的载体,其中所述CasX蛋白和所述gNA在RNP中结合在一起。

91.根据权利要求87到90中任一项所述的载体,其进一步包含所述供体模板。

92.根据权利要求87到91中任一项所述的载体,其进一步包含提供所述VLP与靶细胞的结合和融合的假型化病毒包膜糖蛋白或抗体片段。

93.一种宿主细胞,其包含根据权利要求81到92中任一项所述的载体。

94.根据权利要求93所述的宿主细胞,其中所述宿主细胞选自由以下组成的组:BHK、HEK293、HEK293T、NS0、SP2/0、YO骨髓瘤细胞、P3X63小鼠骨髓瘤细胞、PER、PER.C6、NIH3T3、COS、HeLa、CHO和酵母细胞。

95.一种修饰细胞群体中的C9orf72靶核酸序列的方法,所述方法包含将以下引入到所述群体的细胞中:

a.根据权利要求1到78中任一项所述的系统;

b.根据权利要求79或权利要求80所述的核酸;

c.根据权利要求81到86中任一项所述的载体;

d.根据权利要求87到92中任一项所述的VLP;或

e.其组合,

其中由第一gNA靶向的所述细胞的C9orf72基因靶核酸序列由所述CasX蛋白修饰。

96.根据权利要求95所述的方法,其中所述CasX蛋白和所述gNA在核糖核蛋白质复合物(RNP)中结合在一起。

97.根据权利要求95或权利要求96所述的方法,其进一步包含第二gNA或编码所述第二gNA的核酸,其中所述第二gNA具有与所述靶核酸序列的不同部分互补的靶向序列。

98.根据权利要求95到97中任一项所述的方法,其中所述C9orf72基因包含突变。

99.根据权利要求98所述的方法,其中所述突变为功能增强突变。

100.根据权利要求98所述的方法,其中所述突变为功能丧失突变。

101.根据权利要求98所述的方法,其中所述C9orf72基因突变包含六核苷酸重复序列

GGGGCC的多于30、多于100、多于500、多于700、多于1000或多于1600个副本。

102.根据权利要求95到101中任一项所述的方法，其中所述修饰包含在所述靶核酸序列中引入单链断裂。

103.根据权利要求94到100中任一项所述的方法，其中所述修饰包含在所述靶核酸序列中引入双链断裂。

104.根据权利要求95到103中任一项所述的方法，其中所述修饰包含引入所述靶核酸序列中的一个或多个核苷酸的插入、缺失、取代、重复或倒位。

105.根据权利要求95到104中任一项所述的方法，其中所述修饰包含修饰所述HRS。

106.根据权利要求105所述的方法，其中删除所述HRS的一部分。

107.根据权利要求105或权利要求106所述的方法，其中经修饰的HRS包含GGGGCC序列的10个至30个重复序列。

108.根据权利要求105或权利要求106所述的方法，其中所述经修饰的HRS由GGGGCC序列的10个至30个重复序列组成。

109.根据权利要求95到108中任一项所述的方法，其中所述靶核酸序列的所述修饰在体外或离体发生。

110.根据权利要求95到109中任一项所述的方法，其中所述靶核酸序列的所述修饰在细胞内部发生。

111.根据权利要求95到108中任一项所述的方法，其中所述靶核酸序列的所述修饰在体内发生。

112.根据权利要求95到111中任一项所述的方法，其中所述细胞为真核细胞。

113.根据权利要求112所述的方法，其中所述真核细胞选自由以下组成的组：啮齿动物细胞、小鼠细胞、大鼠细胞、猪细胞、灵长类动物细胞和非人类灵长类动物细胞。

114.根据权利要求112所述的方法，其中所述真核细胞为人类细胞。

115.根据权利要求95到114中任一项所述的方法，其中所述细胞选自由以下组成的组：普尔基涅(Purkinje)细胞、额皮质神经元、运动皮质神经元、海马神经元、小脑神经元、上运动神经元、脊髓神经元、脊髓运动神经元、神经胶质细胞和星形胶质细胞。

116.根据权利要求95到115中任一项所述的方法，其中所述方法进一步包含使所述靶核酸序列与供体模板接触，所述供体模板包含与侧接所述系统所靶向的所述靶核酸中的裂解位点的序列互补的同源臂。

117.根据权利要求116所述的方法，其中所述供体模板包含相比于所述野生型C9orf72基因序列的一个或多个突变，并且其中所述插入引起所述C9orf72基因的敲减或敲除。

118.根据权利要求116所述的方法，其中插入所述供体模板置换所述C9orf72基因的所述HRS中的一些或全部。

119.根据权利要求118所述的方法，其中插入所述供体模板产生具有所述GGGGCC序列的10到约30个重复序列的HRS。

120.根据权利要求116所述的方法，其中所述供体模板包含野生型C9orf72基因序列的全部或一部分，其中所述插入校正所述C9orf72基因的一个或多个突变。

121.根据权利要求116到120中任一项所述的方法，其中所述供体模板的大小在10到15,000个核苷酸的范围内。

122.根据权利要求116到120中任一项所述的方法,其中所述供体模板的大小在100到1,000个核苷酸的范围内。

123.根据权利要求116到122中任一项所述的方法,其中所述供体模板为单链DNA模板或单链RNA模板。

124.根据权利要求116到122中任一项所述的方法,其中所述供体模板为双链DNA模板。

125.根据权利要求116到124中任一项所述的方法,其中所述供体模板通过同源定向修复(HDR)插入。

126.根据权利要求95到125中任一项所述的方法,其中所述靶核酸已经修饰,使得所述群体的所述细胞的所述HRS或二肽重复蛋白质(DPR)表达相较于其中所述靶核酸尚未修饰的细胞降低至少约10%、至少约20%、至少约30%、至少约40%、至少约50%、至少约60%、至少约70%、至少约80%或至少约90%。

127.根据权利要求95到125中任一项所述的方法,其中所述细胞已经修饰,使得所述细胞不表达能够检测水平的所述二肽重复蛋白质(DPR)。

128.根据权利要求95到125中任一项所述的方法,其中所述靶核酸已经修饰,使得功能C9orf72蛋白质的表达相较于其中所述靶核酸尚未修饰的细胞增加至少约10%、至少约20%、至少约30%、至少约40%、至少约50%、至少约60%、至少约70%、至少约80%或至少约90%。

129.根据权利要求95到128中任一项所述的方法,其中所述细胞为真核细胞。

130.根据权利要求129所述的方法,其中所述真核细胞选自由以下组成的组:啮齿动物细胞、小鼠细胞、大鼠细胞和非人类灵长类动物细胞。

131.根据权利要求129所述的方法,其中所述真核细胞为人类细胞。

132.根据权利要求129到131中任一项所述的方法,其中所述真核细胞选自由以下组成的组:普尔基涅细胞、额皮质神经元、运动皮质神经元、海马神经元、小脑神经元、上运动神经元、脊髓神经元、脊髓运动神经元、神经胶质细胞和星形胶质细胞。

133.根据权利要求95到132中任一项所述的方法,其中所述细胞群体的所述C9orf72基因靶核酸序列的所述修饰在体外或离体发生。

134.根据权利要求95到132所述的方法,其中所述细胞群体的所述C9orf72基因靶核酸序列的所述修饰在受试者体内发生。

135.根据权利要求134所述的方法,其中所述受试者选自由以下组成的组:啮齿动物、小鼠、大鼠和非人类灵长类动物。

136.根据权利要求134所述的方法,其中所述受试者为人类。

137.根据权利要求134到136中任一项所述的方法,其中所述方法包含向所述受试者施用治疗有效剂量的AAV载体。

138.根据权利要求137所述的方法,其中所述AAV载体以下列剂量向所述受试者施用:至少约1×10^{8}载体基因组(vg)、至少约1×10^{5}载体基因组/kg(vg/kg)、至少约1×10^{6}vg/kg、至少约1×10^{7}vg/kg、至少约1×10^{8}vg/kg、至少约1×10^{9}vg/kg、至少约1×10^{10}vg/kg、至少约1×10^{11}vg/kg、至少约1×10^{12}vg/kg、至少约1×10^{13}vg/kg、至少约1×10^{14}vg/kg、至少约1×10^{15}vg/kg或至少约1×10^{16}vg/kg。

139.根据权利要求137所述的方法,其中所述AAV载体以下列剂量向所述受试者施用:

至少约1×10^{5}vg/kg到约1×10^{16}vg/kg、至少约1×10^{6}vg/kg到约1×10^{15}vg/kg或至少约1×10^{7}vg/kg到约1×10^{14}vg/kg。

140.根据权利要求134到136中任一项所述的方法，其中所述方法包含向所述受试者施用治疗有效剂量的VLP。

141.根据权利要求140所述的方法，其中所述VLP以下列剂量向所述受试者施用：至少约1×10^{5}个粒子/kg、至少约1×10^{6}个粒子/kg、至少约1×10^{7}个粒子/kg、至少约1×10^{8}个粒子/kg、至少约1×10^{9}个粒子/kg、至少约1×10^{10}个粒子/kg、至少约1×10^{11}个粒子/kg、至少约1×10^{12}个粒子/kg、至少约1×10^{13}个粒子/kg、至少约1×10^{14}个粒子/kg、至少约1×10^{15}个粒子/kg、至少约1×10^{16}个粒子/kg。

142.根据权利要求140所述的方法，其中所述VLP以下列剂量向所述受试者施用：至少约1×10^{5}个粒子/kg到约1×10^{16}个粒子/kg、或至少约1×10^{6}个粒子/kg到约1×10^{15}个粒子/kg、或至少约1×10^{7}个粒子/kg到约1×10^{14}个粒子/kg。

143.根据权利要求137到142中任一项所述的方法，其中通过选自以下的施用途径向所述受试者施用所述载体或所述VLP：皮下、皮内、神经内、结节内、髓内、肌肉内、腰髓内、鞘内、蛛网膜下、室内、囊内、静脉内、淋巴内或腹膜内途径，并且其中施用方法为注射、输注或植入。

144.根据权利要求95到143中任一项所述的方法，其包含进一步使所述靶核酸序列与额外CRISPR核酸酶或编码所述额外CRISPR核酸酶的多核苷酸接触。

145.根据权利要求144所述的方法，其中所述额外CRISPR核酸酶为具有不同于根据前述权利要求中任一项所述的CasX蛋白的序列的CasX蛋白。

146.根据权利要求144所述的方法，其中所述额外CRISPR核酸酶不是CasX蛋白。

147.一种通过根据权利要求95到146中任一项所述的方法修饰的细胞群体，其中所述细胞已经修饰，使得经修饰的细胞的至少10％、至少约20％、至少约30％、至少约40％、至少约50％、至少约60％、至少约70％、至少约80％或至少约90％不表达能够检测水平的DPR。

148.一种通过根据权利要求95到146中任一项所述的方法修饰的细胞群体，其中所述细胞已经修饰，使得功能C9orf72蛋白质的表达相较于其中所述C9orf72基因尚未修饰的细胞增加至少约10％、至少约20％、至少约30％、至少约40％、至少约50％、至少约60％、至少约70％、至少约80％或至少约90％。

149.一种通过根据权利要求95到146中任一项所述的方法修饰的细胞群体，其中所述C9orf72基因的所述突变在所述群体的所述经修饰的细胞中得到校正，使得所述经修饰的细胞表达功能C9orf72蛋白质。

150.根据权利要求147到149中任一项所述的细胞群体，其中所述细胞为非灵长类哺乳动物细胞、非人类灵长类动物细胞或人类细胞。

151.根据权利要求147到150所述的细胞群体，其中所述细胞选自由以下组成的组：普尔基涅细胞、额皮质神经元、运动皮质神经元、海马神经元、小脑神经元、上运动神经元、脊髓神经元、脊髓运动神经元、神经胶质细胞和星形胶质细胞。

152.一种治疗有需要的受试者的C9orf72相关病症的方法，其包含施用治疗有效量的根据权利要求147到151中任一项所述的细胞。

153.根据权利要求152所述的方法，其中所述C9orf72相关病症为肌肉萎缩性侧索硬化

(ALS)或额颞叶型痴呆(FTD)。

154.根据权利要求152或权利要求153所述的方法,其中所述细胞相对于待施用所述细胞的所述受试者为自体的。

155.根据权利要求152或权利要求153所述的方法,其中所述细胞相对于待施用所述细胞的所述受试者为同种异体的。

156.根据权利要求152到155中任一项所述的方法,其中所述方法进一步包含施用化学治疗剂。

157.根据权利要求152到156中任一项所述的方法,其中所述受试者选自由以下组成的组:啮齿动物、小鼠、大鼠和非人类灵长类动物。

158.根据权利要求152到156中任一项所述的方法,其中所述受试者为人类。

159.一种治疗有需要的受试者的C9orf72相关病症的方法,其包含修饰所述受试者的细胞中的C9orf72基因,所述修饰包含使所述细胞与治疗有效剂量的以下接触:

a.根据权利要求1到78中任一项所述的系统;

b.根据权利要求79或权利要求80所述的核酸;

c.根据权利要求81到86中任一项所述的载体;

d.根据权利要求87到90中任一项所述的VLP;或

e.其组合,

其中由所述第一gNA靶向的所述细胞的所述C9orf72基因由所述CasX蛋白修饰。

160.根据权利要求159所述的方法,其中所述C9orf72相关病症为肌肉萎缩性侧索硬化(ALS)或额颞叶型痴呆(FTD)。

161.根据权利要求159或权利要求160所述的系统,其中所述第一gNA的所述靶向序列与在所述C9orf72基因的所述HRS的5'处的序列互补。

162.根据权利要求159到161中任一项所述的方法,其进一步包含第二gNA或编码所述第二gNA的核酸,其中所述第二gNA具有与以下互补的靶向序列:与所述第一gNA相比所述靶核酸序列的不同或重叠部分。

163.根据权利要求162所述的系统,其中所述第二gNA的所述靶向序列与在所述C9orf72基因的内含子1中并且在所述HRS的3'处的序列互补。

164.根据权利要求159到163中任一项所述的方法,其中所述方法包含将所述供体模板插入到所述细胞的所述C9orf72基因靶核酸序列的一个或多个断裂位点中。

165.根据权利要求164所述的方法,其中所述供体模板的所述插入由同源定向修复(HDR)或同源独立靶向整合(HITI)介导。

166.根据权利要求164或权利要求165所述的方法,其中所述供体模板的插入引起所述受试者的所述经修饰的细胞中的所述C9orf72基因中的一个或多个所述突变的校正。

167.根据权利要求166所述的方法,其中所述突变的校正使得所述受试者的所述经修饰的细胞表达功能C9orf72蛋白质。

168.根据权利要求159所述的方法,其中所述载体为AAV。

169.根据权利要求168所述的方法,其中所述AAV载体以下列剂量向所述受试者施用:至少约1×10^{8}载体基因组(vg)、至少约1×10^{5}载体基因组/kg(vg/kg)、至少约1×10^{6}vg/kg、至少约1×10^{7}vg/kg、至少约1×10^{8}vg/kg、至少约1×10^{9}vg/kg、至少约1×10^{10}vg/kg、至少约

1×10^{11}vg/kg、至少约1×10^{12}vg/kg、至少约1×10^{13}vg/kg、至少约1×10^{14}vg/kg、至少约1×10^{15}vg/kg或至少约1×10^{16}vg/kg。

170.根据权利要求168所述的方法，其中所述AAV载体以下列剂量向所述受试者施用：至少约1×10^{5}vg/kg到约1×10^{16}vg/kg、至少约1×10^{6}vg/kg到约1×10^{15}vg/kg或至少约1×10^{7}vg/kg到约1×10^{14}vg/kg。

171.根据权利要求159所述的方法，其中所述VLP以下列剂量向所述受试者施用：至少约1×10^{5}个粒子/kg、至少约1×10^{6}个粒子/kg、至少约1×10^{7}个粒子/kg、至少约1×10^{8}个粒子/kg、至少约1×10^{9}个粒子/kg、至少约1×10^{10}个粒子/kg、至少约1×10^{11}个粒子/kg、至少约1×10^{12}个粒子/kg、至少约1×10^{13}个粒子/kg、至少约1×10^{14}个粒子/kg、至少约1×10^{15}个粒子/kg、至少约1×10^{16}个粒子/kg。

172.根据权利要求159所述的方法，其中所述VLP以下列剂量向所述受试者施用：至少约1×10^{5}个粒子/kg到约1×10^{16}个粒子/kg、或至少约1×10^{6}个粒子/kg到约1×10^{15}个粒子/kg、或至少约1×10^{7}个粒子/kg到约1×10^{14}个粒子/kg。

173.根据权利要求168到172中任一项所述的方法，其中通过选自以下的施用途径向所述受试者施用所述载体或所述VLP：皮下、皮内、神经内、结节内、髓内、肌肉内、腰髓内、鞘内、蛛网膜下、室内、囊内、静脉内、淋巴内或腹膜内途径，并且其中施用方法为注射、输注或植入。

174.根据权利要求159到173中任一项所述的方法，其中所述经修饰的细胞的所述C9orf72基因表达增加水平的功能C9orf72蛋白质，其中所述增加是相较于C9orf72基因尚未修饰的细胞的至少约10％、至少约20％、至少约30％、至少约40％、至少约50％、至少约60％、至少约70％、至少约80％或至少约90％。

175.根据权利要求159到174中任一项所述的方法，其中至少约10％、至少约20％、至少约30％、至少约40％、至少约50％、至少约60％、至少约70％、至少约80％或至少约90％的所述经修饰的细胞不表达能够检测水平的二肽重复蛋白质(DPR)。

176.根据权利要求159到174中任一项所述的方法，其中所述修饰引入所述C9orf72基因中的一个或多个突变，或其中所述HRS和/或所述DPR的表达相较于尚未修饰的细胞降低至少约10％、至少约20％、至少约30％、至少约40％、至少约50％、至少约60％、至少约70％、至少约80％或至少约90％。

177.根据权利要求159到176中任一项所述的方法，其中所述细胞选自由以下组成的组：普尔基涅细胞、额皮质神经元、运动皮质神经元、海马神经元、小脑神经元、上运动神经元、脊髓神经元、脊髓运动神经元、神经胶质细胞和星形胶质细胞。

178.根据权利要求159到177中任一项所述的方法，其中所述受试者选自由以下组成的组：小鼠、大鼠、猪和非人类灵长类动物。

179.根据权利要求159到177中任一项所述的方法，其中所述受试者为人类。

180.根据权利要求159到179中任一项所述的方法，其包含进一步使所述靶核酸序列与额外CRISPR核酸酶或编码额外CRISPR蛋白质的多核苷酸接触。

181.根据权利要求180所述的方法，其中所述额外CRISPR核酸酶为具有不同于根据前述权利要求中任一项所述的CasX的序列的CasX蛋白。

182.根据权利要求180所述的方法，其中所述额外CRISPR核酸酶不是CasX蛋白。

183.根据权利要求159到181中任一项所述的方法,其中所述方法进一步包含施用化学治疗剂。

184.根据权利要求159到183中任一项所述的方法,其中所述方法引起至少一种选自由以下组成的组的临床上相关参数的改进:神经元细胞死亡、神经发炎、TDP-43相关病变、轴突和神经肌肉接头(NMJ)异常、前额皮质处的树突状脊密度变化、新生皮质神经元中的电生理缺陷、预测缓慢肺活量(SVC)百分比相对于基线的变化、肌力相对于基线的变化、延髓强度相对于基线的变化、ALS功能评级量表(ALSFRS-(R))、功能和存活的组合评估、反应持续时间、到死亡的时间、到气管切开术的时间、到持续辅助通气的时间(DTP)、用力肺活量(FVC%);徒手肌力测试、最大自主等长收缩、反应持续时间、无发展存活期、到疾病进展的时间和到治疗失效的时间。

185.根据权利要求159到183中任一项所述的方法,其中所述方法引起至少两种选自由以下组成的组的临床上相关参数的改进:神经元细胞死亡、神经发炎、TDP-43相关病变、轴突和神经肌肉接头(NMJ)异常、前额皮质处的树突状脊密度变化、新生皮质神经元中的电生理缺陷、预测缓慢肺活量(SVC)百分比相对于基线的变化、肌力相对于基线的变化、延髓强度相对于基线的变化、ALS功能评级量表(ALSFRS-(R))、功能和存活的组合评估、反应持续时间、到死亡的时间、到气管切开术的时间、到持续辅助通气的时间(DTP)、用力肺活量(FVC%);徒手肌力测试、最大自主等长收缩、反应持续时间、无发展存活期、到疾病进展的时间和到治疗失效的时间。

186.根据权利要求1到78中任一项所述的系统,其中所述靶核酸序列与定位于前间隔区邻近基序(PAM)序列的3'的1个核苷酸处的非靶链序列互补。

187.根据权利要求186所述的系统,其中所述PAM序列包含TC基序。

188.根据权利要求187所述的系统,其中所述PAM序列包含ATC、GTC、CTC或TTC。

189.根据权利要求186到188中任一项所述的系统,其中所述2类V型CRISPR蛋白质包含RuvC结构域。

190.根据权利要求189所述的系统,其中所述RuvC结构域在所述靶核酸序列中产生交错双链断裂。

191.根据权利要求186到190中任一项所述的系统,其中所述2类V型CRISPR蛋白质不包含HNH核酸酶结构域。

192.一种用于治疗有需要的受试者的C9orf72相关病症的方法的组合物,其包含施用治疗有效量的根据权利要求147到151中任一项所述的细胞。

193.一种用于治疗有需要的受试者的C9orf72相关病症的方法的组合物,其包含修饰所述受试者的细胞中的C9orf72基因,所述修饰包含使所述细胞与治疗有效剂量的以下接触:

a.根据权利要求1到78中任一项所述的系统;

b.根据权利要求79或权利要求80所述的核酸;

c.根据权利要求81到86中任一项所述的载体;

d.根据权利要求87到90中任一项所述的VLP;或

e.其组合,

其中由第一gNA靶向的所述细胞的所述C9orf72基因由所述CasX蛋白修饰。

用于靶向C9orf72的组合物和方法

[0001] 相关申请案的交互参考

[0002] 本申请要求2020年3月18日提交的美国临时专利申请第62/991,403号的优先权，其内容以全文引用的方式并入本文中。

[0003] 通过引用并入序列表

[0004] 与本申请一起以电子方式提交的文本文件的内容以全文引用的方式并入本文中：序列表的计算机可读格式副本(文件名:SCRB-025-01WO_SeqList_ST25.txt;记录日期:2021年3月12日;文件大小:5.61兆字节)。

背景技术

[0005] 肌肉萎缩性侧索硬化(ALS)和额颞叶型痴呆(FTD)是具有破坏性结果的进行性神经疾病形式。ALS是一种致命的神经退化性疾病，在临床上特征在于进行性麻痹，典型地在症状初发的两年到三年内引起呼吸衰竭死亡，并且为西方世界第三大最常见神经退化性疾病(Rowland和Shneider,《新英格兰医学杂志(N.Engl.J.Med.)》,2001,344,1688-1700;Hirtz等人,《神经病学(Neurology)》,2007,68,326-337)。FTD为早老性痴呆的第二最常见原因，其中脑部的额叶和颞叶的退化导致人格、行为和语言的进行性变化，而感知和记忆相对保留(Graff-Radford,N和Woodruff,B.《额颞叶型痴呆(Frontotemporal dementia)》.《神经学研讨会(Semin.Neurol.)》27(1):48(2007))。

[0006] 染色体9开放阅读框架72蛋白质为由C9orf72基因(有时也称为C9orf72-SMCR8复合子单元)编码的蛋白质。与C9orf72基因中的突变或异常有关的疾病形式包括FTD和ALS。所述蛋白质发现于脑部的许多区域中，包括神经元的细胞质，以及突触前末梢中。特别地，与FTD和ALS有关的C9orf72基因中的相关突变为核苷酸六字母串GGGGCC的六核苷酸重复序列扩增段，其出现于C9orf72基因的内含子1中介于两个5'-非翻译区(5'-UTR)外显子之间或启动子区中(DeJesus-Hernandez,M.等人《C9ORF72非编码区域中的扩增GGGGCC六核苷酸重复序列引起染色体9p相关的FTD和ALS(Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS)》.《神经元(Neuron)》72:245(2011);Niblock,M.等人《C9orf72 mRNA中含六核苷酸重复序列的内含子的保留:对ALS/FTD发病机制的影响(Retention of hexanucleotide repeat-containing intron in C9orf72 mRNA:implications for the pathogenesis of ALS/FTD)》.《神经病理学通讯学报(Acta Neuropathologica Communications)》4:18(2016))。因此，六核苷酸重复序列(HRS)扩增段的存在不改变所得C9orf72蛋白质的编码序列。在健康个体中，存在较少此六核苷酸的重复序列，典型地30个或更少，但在具有患病表型的人中，重复单元在大约700到1600个范围内(Mori K.等人《C9orf72 GGGGCC重复序列被翻译成FTLD/ALS中的聚集二肽重复蛋白(The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS)》.《科学(Science)》.339:1335(2013))。据信，重复六核苷酸扩增段使得损失一种交替剪接的C9orf72转录物，并且通过与重复序列相关的非AUG起始(RAN)翻译，形成和积聚不溶性二肽重复蛋白聚集体，大多含有聚-(Gly-Ala)，

并且在较小程度上,含有聚-(Gly-Pro)和聚-(Gly-Arg)二肽重复蛋白(DPR),其具有极强的疏水性并且可在FTD-ALS患者中是病原性的(Mori K.等人2013;Niblock,M.等人2016)。此外,已提出三种主要疾病机制:C9orf72蛋白质的功能丧失,由C9orf72重复RNA通过在额皮质和脊髓中积聚含有重复序列以及反义GGCCCC RNA的RNA转录物而获得毒性功能,或由通过与重复序列相关的非ATG翻译产生的DPR积聚而获得毒性功能(Balendra R,Isaacs AM.《C9orf72介导的ALS和FTD:疾病的多个路径(C9orf72-mediated ALS and FTD:multiple pathways to disease)》.《自然评论神经病学(Nat Rev Neurol.)》14:544(2018))。C9orf72突变的遗传是常染色体显性遗传(Iyer等人《C9orf72,一种与肌肉萎缩性侧索硬化(ALS)相关的蛋白质,是鸟嘌呤核苷酸交换因子(C9orf72,a protein associated with amyotrophic lateral sclerosis(ALS)is a guanine nucleotide exchange factor)》《同行杂志(PeerJ)》6:e5815(2018))。

[0007] CRISPR/Cas系统的出现和这些最小系统的可编程性质促进了它们作为基因组操纵和工程改造的通用技术的用途。然而,通过基因工程改造来校正如FTD和ALS的C9orf72相关疾病的工作受到的关注有限。因此,需要调节患有C9orf72相关疾病的受试者中的C9orf72的组合物和方法。本文提供了用于靶向C9orf72基因以满足此需求的组合物和方法。

发明内容

[0008] 本公开提供用于编辑染色体9开放阅读框架72(C9orf72)基因靶核酸序列的经修饰的2类V型CRISPR蛋白质和引导核酸的组合物。2类V型CRISPR蛋白质和引导核酸经修饰以被动进入到靶细胞中。2类V型CRISPR蛋白质和引导核酸可用于C9orf72相关疾病的靶核酸修饰的多种方法,也提供所述方法。

[0009] 在一个方面,本公开涉及用于改变细胞中的包含具有一个或多个突变或包含六核苷酸重复序列扩增段(HRS)的C9orf72基因的靶核酸的CasX:引导核酸系统(CasX:gNA系统)和方法。在本公开的一些实施例中,CasX:gNA系统在敲减或敲除具有一个或多个突变或包含六核苷酸重复序列扩增段(HRS)的C9orf72基因,以便减少或消除患有C9orf72相关疾病的受试者中的C9orf72基因产物的表达、来自HRS的RNA和/或DPR的积聚方面具有效用。在其它实施例中,CasX:gNA系统在校正包含HRS的C9orf72基因方面具有效用。

[0010] 在系统的一些实施例中,gNA为gRNA,或gDNA,或RNA和DNA的嵌合体,且可为单分子gNA或双分子gNA。在其它实施例中,CasX:gNA系统gNA具有与包含C9orf72基因内的区域的靶核酸序列互补的靶向序列。在一些实施例中,gNA的靶向序列选自由以下组成的组:SEQ ID NO:309-343、363-2100、2295-2185或与其具有至少约65%、至少约75%、至少约85%或至少约95%同一性的序列。gNA可包含含有14至30个连续核苷酸的靶向序列。在一些实施例中,gNA的靶向序列由21个核苷酸组成。在其它实施例中,gNA的靶向序列由20个核苷酸组成。在其它实施例中,靶向序列由19个核苷酸组成,gNA的靶向序列具有选自由以下组成的组的序列:SEQ ID NO:309-343、363-2100、2295-2185,其中从序列的3'端去除单个核苷酸。在其它实施例中,靶向序列由18个核苷酸组成,具有选自由以下组成的组的序列:SEQ ID NO:309-343、363-2100、2295-2185,其中从序列的3'端去除两个核苷酸。在其它实施例中,靶向序列由17个核苷酸组成,具有选自由以下组成的组的序列:SEQ ID NO:309-343、363-

2100、2295-2185，其中从序列的3'端去除三个核苷酸。在其它实施例中，靶向序列由16个核苷酸组成，具有选自由以下组成的组的序列：SEQ ID NO:309-343、363-2100、2295-2185，其中从序列的3'端去除四个核苷酸。在其它实施例中，靶向序列由15个核苷酸组成，具有选自由以下组成的组的序列：SEQ ID NO:309-343、363-2100、2295-2185，其中从序列的3'端去除五个核苷酸。

[0011] 在所述系统的一些实施例中，所述gNA具有支架，所述支架包含选自由以下组成的组的序列：SEQ ID NO:4-16和2101-2294，或与其具有至少约50％、至少约60％、至少约70％、至少约80％、至少约90％、至少约95％、至少约95％、至少约96％、至少约97％、至少约98％或至少约99％序列同一性的序列。

[0012] 在所述系统的一些实施例中，2类V型CRISPR蛋白质包含具有SEQ ID NO:1-3中的任一个的序列的参考CasX蛋白、具有选自由SEQ ID NO:49-150、233-235、238-239、240-242和272-281组成的组的序列的CasX变异蛋白质，或与其具有至少约50％、至少约60％、至少约70％、至少约80％、至少约90％、至少约95％、至少约96％、至少约97％、至少约98％或至少约99％序列同一性的序列。在这些实施例中，相对于参考CasX蛋白，CasX变异体表现出一种或多种改进的特征。在一些实施例中，CasX蛋白对选自由TTC、ATC、GTC和CTC组成的组的前间隔区邻近基序(PAM)序列具有结合亲和力。在一些实施例中，与SEQ ID NO:1-3的任何一种CasX蛋白对选自由TTC、ATC、GTC和CTC组成的组的PAM序列的结合亲和力相比，所述CasX蛋白对PAM序列的结合亲和力至少高1.5倍。

[0013] 在系统的一些实施例中，CasX分子和gNA分子在核糖核蛋白质复合物(RNP)中结合在一起。在特定实施例中，在细胞分析系统中，当所述PAM序列TTC、ATC、GTC或CTC中的任一个定位于与所述gNA的所述靶向序列具有同一性的非靶链序列5'的1个核苷酸处时，包含所述CasX变异体和所述gNA变异体的所述RNP，相比于类似分析系统中包含参考CasX蛋白和参考gNA的RNP的编辑效率和/或结合，展现更高的靶DNA中编辑效率和/或靶序列结合。

[0014] 在一些实施例中，所述系统进一步包含供体模板，其包含核酸，所述核酸包含C9orf72基因的至少一部分，其中所述C9orf72基因部分选自由以下组成的组：C9orf72外显子、C9orf72内含子、C9orf72内含子-外显子接界、C9orf72调节元件或其组合，其中所述供体模板用于敲减或敲除所述C9orf72基因或用于校正所述C9orf72基因中的突变。在一些实施例中，供体模板包含GGGGCC序列的六核苷酸重复序列，其中重复序列的数目在10到约30个重复序列范围内，并且用于置换突变C9orf72基因的六核苷酸重复序列扩增段。在一些情况下，供体序列为单链DNA模板或单链RNA模板。在其它情况下，供体模板是双链DNA模板。

[0015] 在其它实施例中，本公开涉及编码本文所述的任何实施例的系统的核酸，以及包含所述核酸的载体。在一些实施例中，载体选自由逆转录病毒载体、慢病毒载体、腺病毒载体、腺相关病毒(AAV)载体、单纯疱疹病毒(HSV)载体、质粒、小环、纳米质粒和RNA载体组成的组。在其它实施例中，载体为病毒样粒子(VLP)，其包含本文所述的实施例中的任一个的CasX与gNA的RNP，和任选地，供体模板核酸和靶向部分，例如病毒来源的糖蛋白。

[0016] 在其它实施例中，本公开提供一种修饰群体的细胞的C9orf72靶核酸序列的方法，其中所述方法包含将以下引入到所述细胞中：a)本文所公开的实施例中的任一个的CasX:gNA系统；b)本文所公开的实施例中的任一个的核酸；c)本文所公开的实施例中的任一个的载体；d)本文所公开的实施例中的任一个的VLP；或e)前述的组合，其中由所述第一gNA靶向

的所述细胞的所述C9orf72基因靶核酸序列通过所述CasX蛋白修饰，在所述靶核酸序列中引入单链或双链断裂。在所述方法的一些实施例中，所述方法进一步包含第二gNA或编码所述第二gNA的核酸，其中所述第二gNA具有与所述靶核酸序列的不同部分互补的靶向序列。在所述方法的一些实施例中，修饰包含与野生型序列相比，在靶核酸序列中引入一个或多个核苷酸的插入、缺失、取代、重复或倒位。在一些情况下，所述方法还包含使靶核酸与本文公开的任一实施例的供体模板核酸接触。在一些实施例中，用于修饰的靶C9orf72基因包含六核苷酸重复序列GGGGCC的多于30、多于100、多于500、多于700、多于1000或多于1600个副本。在所述方法的一些实施例中，供体模板包含核酸，所述核酸包含用于校正(通过敲入)C9orf72基因的突变的C9orf72基因的至少一部分，或包含含有用于敲减或敲除突变C9orf72的突变或异源序列的序列，使得群体的细胞的HRS或DPR的表达相较于尚未修饰的细胞降低至少约10％、至少约20％、至少约30％、至少约40％、至少约50％、至少约60％、至少约70％、至少约80％或至少约90％。在一些情况下，靶核酸序列的修饰在体内发生。在一些实施例中，细胞为选自由以下组成的组的真核细胞：啮齿动物细胞、小鼠细胞、大鼠细胞、灵长类动物细胞、和非人类灵长类动物细胞。在一些实施例中，细胞为人类细胞。在一些实施例中，细胞选自由以下组成的组：普尔基涅(Purkinje)细胞、额皮质神经元、运动皮质神经元、海马神经元、小脑神经元、上运动神经元、脊髓神经元、脊髓运动神经元、神经胶质细胞和星形胶质细胞。

[0017] 在其它实施例中，本公开提供修饰受试者的细胞群体中的C9orf72靶核酸的方法，其中使用编码CasX蛋白和一种或多种包含与C9orf72基因互补的靶向序列的gNA，并且任选地进一步包含供体模板的载体接触靶细胞。在一些情况下，载体为选自AAV1、AAV2、AAV3、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAV10、AAV11、AAV12、AAV 44.9、AAV-Rh74或AAVRh10的腺相关病毒(AAV)载体。在其它情况下，载体是慢病毒载体。在其它实施例中，本公开提供其中使用载体接触靶细胞的方法，其中载体为病毒样粒子(VLP)，其包含本文所述的任何实施例的CasX与gNA的RNP和任选的供体模板核酸。在所述方法的一些实施例中，将载体以治疗有效剂量施用于受试者。受试者可以是小鼠、大鼠、猪、非人灵长类动物或人类。剂量可以通过选自由以下组成的组的施用途径施用：皮下、皮内、神经内、结节内、髓内、肌肉内、腰髓内、鞘内、蛛网膜下、室内、囊内、静脉内、淋巴内或腹膜内途径，其中施用方法为注射、输注或植入。

[0018] 在其它实施例中，本公开提供一种治疗受试者的C9orf72相关病症的方法，其包含修饰所述受试者的细胞中编码C9orf72基因的基因，所述修饰包含使所述细胞与以下接触：a)本文所公开的实施例中的任一个的CasX:gNA系统；b)本文所公开的实施例中的任一个的核酸；c)本文所公开的实施例中的任一个的载体；d)本文所公开的实施例中的任一个的VLP；或e)前述的组合，其中由所述第一gNA靶向的所述细胞的所述C9orf72基因由所述CasX蛋白修饰。在一些实施例中，受试者选自由以下组成的组：小鼠、大鼠、猪、非人类灵长类动物和人类。在一些实施例中，C9orf72相关病症为ALS或FTD。在一些情况下，治疗患有C9orf72相关疾病的受试者的方法引起至少一种临床上相关参数的改进。在其它情况下，治疗患有C9orf72相关疾病的受试者的方法引起至少两种临床上相关参数的改进。

[0019] 在其它实施例中，本公开提供用于治疗受试者的C9orf72相关病症的方法的组合物。在一些实施例中，所述方法包含修饰受试者的细胞中编码C9orf72基因的基因，所述修

饰包含使所述细胞与选自以下的组合物接触:a)本文所公开的实施例中的任一个的CasX:gNA系统;b)本文所公开的实施例中的任一个的核酸;c)本文所公开的实施例中的任一个的载体;d)本文所公开的实施例中的任一个的VLP;或e)前述的组合,其中由所述第一gNA靶向的所述细胞的所述C9orf72基因由所述CasX蛋白修饰。在一些实施例中,受试者选自由以下组成的组:小鼠、大鼠、猪、非人类灵长类动物和人类。在一些实施例中,C9orf72相关病症为ALS或FTD。在一些情况下,治疗患有C9orf72相关疾病的受试者的方法引起至少一种临床上相关参数的改进。在其它情况下,治疗患有C9orf72相关疾病的受试者的方法引起至少两种临床上相关参数的改进。

[0020] 通过引用的并入

[0021] 本说明书中所提及的所有公开案、专利及专利申请案均以引用的方式并入本文中,其引用的程度如每个个别公开案、专利或专利申请案经特定及个别地指示以引用的方式并入一般。公开CasX变异体和gNA变异体的2020年6月5日提交的WO 2020/247882和2020年12月3日提交的美国临时申请63/121,196和2021年3月17日提交的63/162,346的内容特此以全文引用的方式并入。

附图说明

[0022] 本公开的新颖特征在随附权利要求书中细致阐述。将参考阐述利用本公开原理的说明性实施例及其附图的以下详细描述来获得对本公开的特征及优势的更好理解:

[0023] 图1示出了如实例1中所述的通过胶体考马斯染色观测的CasX StX2纯化级分的SDS-PAGE凝胶。

[0024] 图2示出了如实例1中所述的使用Superdex 200 16/600pg凝胶过滤对CasX StX2进行的尺寸排阻色谱分析的色谱图。

[0025] 图3示出了如实例1中所述的通过胶体考马斯染色观测的CasX StX2纯化级分的SDS-PAGE凝胶。

[0026] 图4为示出了如实例2中所述的用于组装CasX构建体的pSTX34质粒中的组分的组织的示意图。

[0027] 图5为示出了如实例2中所述的产生具有CasX 119变异体的pSTX34质粒的步骤的示意图。

[0028] 图6示出了如实例2中所述的在Bio-Rad Stain-Free™凝胶上观测的纯化样品的SDS-PAGE凝胶。

[0029] 图7示出了如实例2中所述的Superdex 200 16/600pg凝胶过滤的色谱图。

[0030] 图8示出了如实例2中所述的通过胶体考马斯染色的凝胶过滤样品的SDS-PAGE凝胶。

[0031] 图9为如实例13中所述的由sgRNA174和CasX变异体119、457、488和491形成的RNP的活性分率的定量分析的结果的图示。等摩尔量的RNP及靶标经共培育且在指定时间点测定裂解靶标的量。针对每个时间点显示三个独立重复样的平均值及标准差。展示合并重复样的双相拟合。“2”是指SEQ ID NO:2的参考CasX蛋白。

[0032] 图10示出了如实例13中所述的由CasX2(SEQ ID NO:2的参考CasX蛋白)和经修饰的sgRNA形成的RNP的活性分率的定量。等摩尔量的RNP及靶标经共培育且在指定时间点测

定裂解靶标的量。针对每个时间点显示三个独立重复样的平均值及标准差。展示合并重复样的双相拟合。

[0033] 图11示出了如实例13中所述的在引导限制条件下由CasX 491和修饰的sgRNA形成的RNP的活性分率的定量。等摩尔量的RNP及靶标经共培育且在指定时间点测定裂解靶标的量。示出了数据的双相拟合。

[0034] 图12示出了如实例13中所述的由sgRNA174和CasX变异体形成的RNP的裂解速率的定量。靶DNA与20倍过量的指定RNP一起培育,且在指定时间点测定裂解靶标的量。示出了每个时间点的三个独立重复样的平均值和标准差,除了示出了单个重复样的488和491。展示合并重复样的单相拟合。

[0035] 图13示出了如实例13中所述的由CasX2和sgRNA变异体形成的RNP的裂解速率的定量。靶DNA与20倍过量的指定RNP一起培育,且在指定时间点测定裂解靶标的量。针对每个时间点显示三个独立重复样的平均值及标准差。展示合并重复样的单相拟合。

[0036] 图14示出了如实例13中所述的由CasX2和sgRNA变异体形成的RNP的初始速度的定量。将前述裂解实验之前两个时间点与线性模型拟合以确定初始裂解速度。

[0037] 图15示出了如实例13中所述的由CasX491和sgRNA变异体形成的RNP的裂解速率的定量。靶DNA与20倍过量的指定RNP在10℃下一起培育,且在指定时间点测定裂解靶标的量。示出了时间点的单相拟合。

[0038] 图16A-16D示出了如实例14中所述的CasX变异体在NTC PAM上的裂解速率的定量。具有相同间隔区和指定的PAM序列的靶DNA底物与20倍过量的指定RNP在37℃下一起培育,且在指定时间点测定裂解靶的量。示出了单个重复样的单相拟合。图16A示出了具有TTC PAM的序列的结果。图16B示出了具有CTC PAM的序列的结果。图16C示出了具有GTC PAM的序列的结果。图16D示出了具有ATC PAM的序列的结果。

[0039] 图17为展示如实例23中所述的用于封装于腺相关病毒(AAV)中的CasX蛋白和支架DNA序列的实例的示意图。由编码CasX的DNA及其启动子,及编码支架的DNA及其启动子构成的AAV反向末端重复序列(ITR)之间的DNA区段在AAV生产期间变为封装于AAV衣壳内。

[0040] 图18示出了编辑分析的结果,所述编辑分析比较从Ai9-tdtomato转基因小鼠分离的小鼠神经祖细胞(mNPC)中的gRNA支架229-237与支架174。细胞用指定剂量的靶向mRHO的编码CasX 491、支架和间隔区11.30(5'AAGGGGCUCCGCACCACGCC 3',SEQ ID NO:361)的p59质粒进行核转染。通过NGS在转染后5天评估mRHO基因座处的编辑,并且展示用具有支架230、231、234和235的构建体编辑在两个剂量下展现相比于具有支架174的构建体更大的编辑。

[0041] 图19示出了在mNPC细胞中比较gRNA支架229-237与支架174的编辑分析的结果。细胞用指定剂量的编码CasX 491、支架和间隔区12.7(5'CUGCAUUCUAGUUGUGGUUU 3',SEQ ID NO:362)的p59质粒进行核转染,所述间隔区靶向预防tdtomato荧光蛋白质表达的重复元件。转染后5天通过FACS评估编辑,以定量tdTomato阳性细胞的分率。细胞用相较于具有支架174的构建体在高剂量下显示大约35%更大编辑,且在低剂量下显示大约25%更大编辑的支架231-235进行核转染。

[0042] 图20示出了在定制HEK293细胞系PASS_V1.01中比较CasX核酸酶2、119、491、515、527、528、529、530和531的编辑分析的结果。细胞用2μg编码所指示的CasX蛋白的p67质粒进

行脂转染。在五天之后，提取细胞基因组DNA。进行PCR扩增和二代测序以分离和定量在定制设计中靶编辑位点处经编辑细胞的分率。对于每一样品，在由以下PAM序列组成的靶位点(个别点)处评价编辑：48TTC、14ATC、22CTC、11GTC个别位点，并且将编辑百分比相对于媒剂对照组归一化。除了CasX 528，用任何核酸酶脂转染的细胞显示与野生型核酸酶CasX 2相比在TTC PAM靶位点(水平条)处更高的平均编辑。任何给定核酸酶对于四种不同PAM序列的相对偏好也由小提琴图表示。特别是，CasX核酸酶527、528和529展现与野生型核酸酶CasX 2显著不同的PAM偏好。

[0043] 图21示出了在定制HEK293细胞系PASS_V1.01中比较改进的CasX核酸酶491与改进的核酸酶532和533的编辑分析的结果。细胞一式两份地用2μg编码所指示的CasX蛋白和嘌呤霉素抗性基因的p67质粒进行脂转染，并且在嘌呤霉素选择下生长。三天后，提取细胞基因组DNA。进行PCR扩增和二代测序以分离和定量在定制设计中靶编辑位点处经编辑细胞的分率。对于每一样品，在由以下PAM序列组成的靶位点处评价编辑：48TTC、14ATC、22CTC、11GTC个别位点，并且将编辑分率相对于媒剂对照组归一化。除了CasX 533在TTC PAM靶位点处，在PAM序列中的每一个处，用CasX 532或533脂转染的细胞显示的平均编辑比Cas 491更高。误差条表示n＝2个生物样品的平均值的标准误差。

[0044] 图22为C9orf72基因座的5'区的一部分的示意图。顶部图示出了外显子1a和外显子1b的相对位置，其侧接六核苷酸重复序列元件(HRE)，而空心框指示下游外显子。下部图示出了如实例18中所描述的由(与)表15的引导RNA的靶向区段(间隔区)靶向(互补)的基因座的区域。

[0045] 图23为展示如实例18中所描述的使用靶向序列164在外显子1a中引入编辑的单一切割实验的结果的图式。黑色删除迹线指示扩增子中的每个位置和在所述位置处具有删除的读数的分率。在图式的底部的灰色条指示sgRNA结合位点位置。定量范围指示用于对删除进行定量的区域。预测裂解位置为CasX诱导的双链断裂的位置。删除迹线说明由递送的单一引导序列产生的基因删除的比率和程度，产生65.4％的总删除效率。数据表示针对单一切割观察到的结果(表15)。

[0046] 图24为展示使用靶向序列(间隔区)138和151的双切割实验的结果的图式，在参考扩增子中在位置193-248处所述靶向序列侧接六核苷酸重复序列元件(HRE，在本文中有时也称为六核苷酸重复序列扩增段或HRS)，如实例18中所描述。黑色删除迹线指示扩增子中的每个位置，在所述位置处具有删除的读数的分率。在图式的底部的灰色条指示sgRNA结合位点位置。定量范围指示用于对删除进行定量的区域。预测裂解位置为CasX诱导的双链断裂的位置。在此实验中，总删除效率为45.4％，且表示针对双切割观察到的结果(表16)，其支持在实验条件下，HRE可使用双切割设计删除。

[0047] 图25为如实例26中所描述的测试间隔区长度对编辑Jurkat细胞中的靶核酸的能力的影响的实验的一对图式。结果表明，在由RNP进行的离体编辑中，相比于20个碱基的间隔区，18或19个碱基的更短间隔区支持增加的活性。

具体实施方式

[0048] 虽然本文中已显示及描述示范性实施例，但所属领域技术人员将显而易知此类实施例仅作为实例提供。在不背离本文所要求保护的本发明的情况下，所属领域技术人员现

将想到许多变化、改变及替代。应理解，本文所述的实施例的各种替代例可用于实践本公开的实施例。预期权利要求书界定本发明的范围，且因此涵盖这些权利要求书及其等效物的范围内的方法及结构。

[0049] 除非另外定义，否则本文中所用的所有技术及科学术语均具有与本发明所属领域的普通技术人员通常所理解相同的含义。尽管与本文所述的方法及材料类似或等效的方法及材料可用于实践或测试本发明实施例，但下文描述适合的方法及材料。在冲突的情况下，将以专利说明书(包括定义)为准。另外，所述材料、方法及实例仅为说明性的且不旨在为限制性的。在不脱离本发明的情况下，所属领域技术人员现将想到诸多变化、改变及取代。

[0050] 定义

[0051] 术语“多核苷酸”及“核酸”在本文中可互换使用，是指任何长度的核苷酸(核糖核苷酸或脱氧核糖核苷酸)的聚合形式。因此，术语“多核苷酸”及“核酸”涵盖单链DNA；双链DNA；多链DNA；单链RNA；双链RNA；多链RNA；基因组DNA；cDNA；DNA-RNA杂合体；及包含嘌呤及嘧啶碱基或其它天然、经化学或生物化学修饰、非天然或衍生化核苷酸碱基的聚合物。

[0052] “可杂交”或“互补”可互换使用，意味着核酸(例如RNA、DNA)包含使其能够在适当的温度及溶液离子强度的体外和/或体内条件下，以序列特异性、反向平行方式(即，核酸与互补核酸特异性结合)与另一核酸非共价结合(即形成沃森-克里克(Watson-Crick)碱基对和/或G/U碱基对)、“退火(anneal)”或“杂交”的核苷酸序列。应当理解，多核苷酸的序列不必与待特异性杂交的靶核酸序列100％互补；其可以具有至少约70％、至少约80％、或至少约90％、或至少约95％序列同一性且仍与靶核酸序列杂交。此外，多核苷酸可在一个或多个区段上杂交以使得中间或邻近区段不参与杂交事件(例如环结构或发夹结构、‘凸起’、‘气泡’及其类似物)。

[0053] 出于本公开的目的，“基因”包括编码基因产物(例如蛋白质、RNA)的DNA区域以及调节基因产物生产的所有DNA区域，无论此类调节序列是否邻近于编码和/或经转录序列。因此，基因可以包括调节元件序列，其包括但不必限于启动子序列、终止子、翻译调节序列(如核糖体结合位点和内部核糖体进入位点)、增强子、沉默子、绝缘子、边界元件、复制起点、基质附着位点和基因座控制区。编码序列在转录或转录及翻译后编码基因产物；本公开的编码序列可包含片段且不必含有全长开放阅读框架。基因可以包括转录的链以及含有反密码子的互补链。

[0054] 术语“下游”是指位于参考核苷酸序列的3’处的核苷酸序列。在某些实施例中，下游核苷酸序列与转录起始点之后的序列相关。举例来说，基因的翻译起始密码子位于转录起始位点下游。

[0055] 术语“上游”是指位于参考核苷酸序列的5’处的核苷酸序列。在某些实施例中，上游核苷酸序列与位于编码区或转录起始点的5’侧上的序列相关。举例来说，大部分启动子位于转录起始位点上游。

[0056] 术语“调节元件”在本文中可与术语“调节序列”互换使用，且打算包括启动子、增强子及其它表达调节元件(例如转录终止信号，例如多聚腺苷酸化信号及聚-U序列)。示范性调节元件包括转录启动子，例如但不限于CMV、CMV+、内含子A、SV40、RSV、HIV-Ltr、延伸因子1α(EF1α)、MMLV-ltr、内部核糖体进入位点(IRES)或P2A肽以准许自单一转录物、金属硫蛋白、转录增强子元件、转录终止信号、聚腺苷酸化序列、用于使翻译起始优化的序列及翻

译终止序列翻译多个基因。在用于外显子跳跃的系统的情况下，调节元件包括外显子剪接增强子。应了解，适当调节元件的选择将取决于待表达的经编码组分(例如蛋白质或RNA)或核酸是否包含多个需要不同聚合酶或不打算表达为融合蛋白的组分。

[0057] 术语“启动子”是指含有RNA聚合酶结合位点、转录起始位点、TATA盒和/或B识别元件且有助于或促进相关可转录多核苷酸序列和/或基因(或转殖基因)的转录及表达的DNA序列。启动子可以合成方式产生或可衍生自已知或天然存在的启动子序列或另一启动子序列。启动子可在待转录的基因近端或远程。启动子亦可包括嵌合启动子，其包含两种或更多种异源序列的组合以赋予某些特性。本公开的启动子可包括与本文已知或提供的其它启动子序列在组成上类似，但与其不相同的启动子序列的变异体。启动子可根据与相关编码或可转录序列或基因的表达模式相关的标准分类，所述序列或基因可操作地连接于启动子，例如组成性、发育性、组织特异性、诱导性启动子等。

[0058] 术语“增强子”是指当与称为转录因子的特异性蛋白质结合时，调节相关基因的表达的调节DNA序列。增强子可位于基因的内含子中，或基因的编码序列的5’或3’处。增强子可在基因近端(即，在启动子的几十或数百个碱基对(bp)内)，或可位于基因远程(即，与启动子相距数千个bp、数十万个bp或甚至数百万个bp)。单一基因可通过超过一种增强子调节，其均被设想为在本公开的范围内。

[0059] 如本文所用，“重组”意指特定核酸(DNA或RNA)为克隆、限制和/或连接步骤的各种组合的产物，产生具有与天然系统中发现的内源核酸可区分的结构性编码或非编码序列的构建体。一般来说，编码结构性编码序列的DNA序列可组装自cDNA片段及短寡核苷酸接头，或组装自一系列合成寡核苷酸，以得到能够自细胞或游离转录及翻译系统中所含的重组转录单元表达的合成核酸。此类序列可以未经内部非翻译序列，或内含子(其典型地存在于真核基因中)间断的开放阅读框架的形式提供。包含相关序列的基因组DNA亦可用于形成重组基因或转录单元。非翻译DNA的序列可存在于开放阅读框架的5’或3’，其中此类序列不干扰编码区的操纵或表达，且可实际上用于通过各种机制调节所需产物的生产(参见上文的“增强子”及“启动子”)。

[0060] 术语“重组多核苷酸”或“重组核酸”是指不天然存在的多核苷酸或核酸，例如经由人工干预由序列的两个另外分离区段的人工组合制得。此人工组合通常通过化学合成手段或通过人工操纵核酸的分离区段，例如通过基因工程改造技术来实现。可进行此类操作以用编码相同或保守氨基酸，同时典型地引入或去除序列识别位点的冗余密码子来替换密码子。或者，进行其以将具有所需功能的核酸区段连接在一起以产生功能的所需组合。此人工组合通常通过化学合成手段或通过人工操纵核酸的分离区段，例如通过基因工程改造技术来实现。

[0061] 类似地，术语”重组多肽”或”重组蛋白”是指并非天然存在的多肽或蛋白质，例如通过经由人工干预将氨基序列的两个另外分离的区段人工组合而制得。因此，例如包含异源氨基酸序列的蛋白质为重组的。

[0062] 如本文所用，术语“接触”意指在两个或更多个实体之间建立物理连接。例如，使靶核酸序列与引导核酸接触意味着使靶核酸序列和引导核酸共享物理连接；例如，在序列共享序列类似性时可以杂交。

[0063] “解离常量”或“K_d”可互换使用且意谓配位体“L”与蛋白质“P”之间的亲和力；即配

位体与特定蛋白质结合的紧密程度。其可使用式K_d＝[L][P]/[LP]计算，其中[P]、[L]及[LP]分别表示蛋白质、配位体及复合物的摩尔浓度。

[0064] 本公开提供适用于编辑靶核酸序列的组合物及方法。如本文所用，“编辑”可与“修饰”互换使用，且包括但不限于裂解、切割、删除、敲入、敲除等。

[0065] 术语“敲除”是指基因的消除或基因的表达。例如，可以通过缺失或添加导致阅读框破坏的核苷酸序列来敲除基因。作为另一实例，可以通过用不相关的序列替换基因的一部分来敲除基因。如本文所用，术语“敲减”是指基因或其基因产物的表达减少。作为基因敲减的结果，蛋白质活性或功能可能会减弱，或者蛋白质水平可能会降低或消除。

[0066] 如本文所用，“同源定向修复”(HDR)是指在修复细胞中的双链断裂期间发生的DNA修复形式。此过程需要核苷酸序列同源性，并且使用供体模板修复或基因敲除靶DNA，并且使得将遗传信息从供体(例如供体模板)转移到靶，产生所关注的转基因。如果供体模板不同于靶DNA序列并且供体模板的部分或全部序列并入到靶DNA中，那么同源定向修复可以通过插入、缺失或突变引起靶核酸序列的序列的改变。

[0067] 如本文所用，“非同源末端连接”(NHEJ)是指通过断裂末端彼此直接连接而修复DNA中的双链断裂，而无需同源模板(相比于同源定向修复，其需要同源序列来引导修复)。NHEJ通常引起双链断裂位点附近核苷酸序列的丢失(缺失)。

[0068] 如本文所用，“微同源性介导的末端连接”(MMEJ)是指诱变DSB修复机制，其始终与侧接断裂位点的缺失结合，而无需同源模板(相比于同源定向修复，其需要同源序列来引导修复)。MMEJ通常引起双链断裂位点附近核苷酸序列的损失(缺失)。

[0069] 多核苷酸或多肽(或蛋白质)与另一多核苷酸或多肽具有某一百分比“序列类似性”或“序列同一性”意指当比对时，碱基或氨基酸的百分比相同，且当比较两个序列时在相同的相对位置。序列类似性(可互换地称为百分比类似性、百分比同一性或同源性)可以多种不同方式确定。为了确定序列类似性，序列可使用所属领域中已知的方法及计算机程序比对，包括在万维网上于ncbi.nlm.nih.gov/BLAST可用的BLAST。核酸内的核酸序列的特定伸长部之间的百分比互补性可使用任何便利方法确定。实例性方法包括BLAST程序(基本局部比对搜索工具)及PowerBLAST程序(Altschul等人，《分子生物学杂志(J.Mol.Biol.)》，1990,215,403-410;Zhang及Madden，《基因组研究(Genome Res.)》,1997,7,649-656)或通过使用Gap程序(Wisconsin Sequence Analysis Package，用于Unix的版本8，威斯康星州麦迪逊(Madison Wis.)University Research Park的Genetics Computer Group)，例如使用默认设定，其使用史密斯及沃特曼算法(algorithm of Smith and Waterman)(《应用数学进展(Adv.Appl.Math.)》,1981,2,482-489)。

[0070] 术语“多肽”及“蛋白质”在本文中可互换使用，且是指任何长度的氨基酸的聚合形式，其可包括编码及非编码氨基酸、化学或生物化学修饰或衍生的氨基酸及具有经修饰肽主链的多肽。所述术语包括融合蛋白，包括但不限于具有异源氨基酸序列的融合蛋白。

[0071] “载体”或“表达载体”为复制子，如质粒、噬菌体、病毒、病毒样粒子或粘质粒，另一DNA区段，即“插入物”，可与其连接，以引起细胞中经连接区段的复制或表达。

[0072] 应用于核酸、多肽、细胞或生物体的如本文所用的术语”天然存在的”或”未修饰”或”野生型”是指自然界中发现的核酸、多肽、细胞或生物体。因此，“野生型”可以指核酸、多肽、细胞或生物体的多于一种天然存在的变异体。关于基因，“野生型”也可以用于指天然存

在的非致病的基因变异体。

[0073] 如本文所用,“突变”是指相比于野生型或参考氨基酸序列或野生型或参考核苷酸序列,一个或多个氨基酸或核苷酸的插入、缺失、取代、重复或倒位。

[0074] 如本文所用,术语“经分离”意在描述处于与多核苷酸、多肽或细胞天然存在的环境不同的环境中的多核苷酸、多肽或细胞。经分离的遗传修饰宿主细胞可存在于遗传修饰宿主细胞的混合群体中。

[0075] 如本文所用,“宿主细胞”指示真核细胞、原核细胞或来自以单细胞实体培养的多细胞生物体(例如细胞系)的细胞,所述真核细胞或原核细胞用作核酸的接受体(例如表达载体),且包括已通过核酸遗传修饰的原始细胞的后代。应理解,单细胞的后代可因天然、偶发或故意突变而不一定与原始亲本具有完全相同的形态或基因组或总DNA互补序列。“重组宿主细胞”(亦称为“遗传修饰宿主细胞”)为其中已引入异源核酸,例如表达载体的宿主细胞。

[0076] 术语“保守氨基酸取代”是指具有类似侧链的氨基酸残基的蛋白质中的互换性。举例来说,具有脂肪族侧链的一组氨基酸由甘氨酸、丙氨酸、缬氨酸、亮氨酸及异亮氨酸组成;具有脂肪族-羟基侧链的一组氨基酸由丝氨酸及苏氨酸组成;具有含酰胺侧链的一组氨基酸由天冬酰胺及谷氨酰胺组成;具有芳香族侧链的一组氨基酸由苯丙氨酸、酪氨酸及色氨酸组成;具有碱性侧链的一组氨基酸由赖氨酸、精氨酸及组氨酸组成;且具有含硫侧链的一组氨基酸由半胱氨酸及甲硫氨酸组成。示范性保守氨基酸取代组为:缬氨酸-亮氨酸-异亮氨酸、苯丙氨酸-酪氨酸、氨酸赖氨酸-精氨酸、丙氨酸-缬氨酸及天冬酰胺-氨酸谷氨酰胺。

[0077] 如本文所用,”治疗(treatment)”或”治疗(treating)”在本文中可互换使用,且是指获得有益或所需结果,包括但不限于治疗益处和/或预防益处的方法。治疗效益意指根除或改善所治疗的潜在病症或疾病。治疗益处也可通过与基础疾病相关的一种或多种症状的根除或改善或一种或多种临床参数的改进,使得在受试者中观测到改进来达成,尽管受试者仍可能罹患基础疾病。

[0078] 如本文所用,术语“治疗有效量”和“治疗有效剂量”是指单独或作为组合物一部分的药物或生物制剂的量,其当以一个或重复剂量向如人类或实验动物的受试者施用时,能够对疾病状态或病况的任何症状、方面、测量参数或特征具有任何可检测的有益影响。此类效应不必绝对有益。

[0079] 如本文所用,“施用”意指向受试者提供一定剂量的化合物(例如,本公开的组合物)或组合物(例如,药物组合物)的方法。

[0080] 如本文所用,“受试者”为哺乳动物。哺乳动物包括但不限于驯养动物、非人灵长类动物、人类、兔子、小鼠、大鼠及其它啮齿动物。

[0081] I.通用方法

[0082] 除非另外规定,否则本发明的实践采用免疫学、生物化学、化学、分子生物学、微生物学、细胞生物学、基因组学及重组DNA的常规技术,其可见于例如以下的标准教科书:Molecular Cloning:A Laboratory Manual,第3版(Sambrook等人,冷泉港实验室出版社(HaRBor Laboratory Press)2001);《分子生物学的简短协议(Short Protocols in Molecular Biology)》,第4版(Ausubel等人编,John Wiley&Sons 1999);Protein Methods(Bollag等人,John Wiley&Sons 1996);Nonviral Vectors for Gene Therapy(Wagner等

人编,Academic Press 1999);Viral Vectors(Kaplift及Loewy编,Academic Press 1995);Immunology Methods Manual(I.Lefkovits编,Academic Press 1997);及Cell and Tissue Culture:Laboratory Procedures in Biotechnology(Doyle及Griffiths,John Wiley&Sons 1998),所述文献的公开内容以引用的方式并入本文中。

[0083] 当提供值范围时,应理解除非上下文另外明确指出,否则包括端点且在所述范围的上限与下限之间的各个中间值(至下限的单位的十分之一)及在所述规定范围内的任何其它指定值或中间值均被涵盖。这些较小范围的上限及下限可独立地包括于较小范围中,且亦被涵盖,在所述范围内受到任何特定排他性限制。当所述范围包括限值中的一个或两个时,也包括排除那些所包括限值的任一个或两个的范围。

[0084] 除非另外规定,否则本文中所用的所有技术及科学术语具有与本发明所属领域的普通技术人员通常所理解相同的含义。本文中所提及的所有公开案以引用的方式并入本文中以结合所述公开案所列举的方法和/或材料来公开且描述。

[0085] 必须注意,除非上下文另有明确规定,否则如在本文中及所附权利要求书中所使用,单数形式“一个”、“一种”及“所述”包括多个指示物。

[0086] 应了解,出于明晰的目的而在独立实施例的背景下描述的本公开的某些特征亦可以组合形式提供于单一实施例中。在其它情况下,为简洁起见而在单一实施例的背景下描述的本公开的各种特征亦可分别或以任何适合的子组合提供。关于本公开的实施例的所有组合旨在由本公开特定涵盖且在本文中公开,如同单独且明确地公开每一组合一般。另外,各种实施例及其要素的所有子组合亦由本公开特定涵盖且在本文中公开,如同单独且明确地在本文中公开每一此类子组合一般。

[0087] II.用于C9orf72基因的基因编辑的系统

[0088] 在第一方面中,本公开提供包含2类V型CRISPR核酸酶蛋白质和一种或多种用于修饰具有一种或多种突变或包含HRS的C9orf72基因的引导核酸(gNA),以便减少或消除C9orf72基因产物、来自HRS转录的RNA和/或DPR(在本文中统称为“靶核酸”,包括编码和非编码区域)的表达的系统。

[0089] 人类C9orf72基因(HGNC:28337)编码具有以下序列的蛋白质(Q01453):MSTLCPPPSPAVAKTEIALSGKSPLLAATFAYWDNILGPRVRHIWAPKTEQVLLSDGEITFLANHTLNGEILRNAESGAIDVKFFVLSEKGVIIVSLIFDGNWNGDRSTYGLSIILPQTELSFYLPLHRVCVDRLTHIIRKGRIWMHKERQENVQKIILEGTERMEDQGQSIIPMLTGEVIPVMELLSSMKSHSVPEEIDIADTVLNDDDIGDSCHEGFLLNAISSHLQTCGCSVVVGSSAEKVNKIVRTLCLFLTPAERKCSRLCEAESSFKYESGLFVQGLLKDSTGSFVLPFRQVMYAPYPTTHIDVDVNTVKQMPPCHEHIYNQRRYMRSELTAFWRATSEEDMAQDTIIYTDESFTPDLNIFQDVLHRDTLVKAFLDQVFQLKPGLSLRSTFLAQFLLVLHRKALTLIKYIEDDTQKGKKPFKSLRNLKIDLDLTAEGDLNIIMALAEKIKPGLHSFIFGRPFYTSVQERDVLMTF(SEQ ID NO:227)。C9orf72基因被定义为跨越染色体9上人类基因组的chr9:27,546,546-27,573,866(智人更新注释发布109.20191205,GRCh38.p13(NCBI))的序列。人类C9orf72基因部分地在NCBI数据库(ncbi.nlm.nih.gov)中描述为参考序列NC_000009.12,所述参考序列通过引用并入本文。C9orf72基因座含有12个外显子,包括2个交替的非编码第一外显子(外显子1a和1b)(DeJesus-Hernandez,M.等人2011)。在六核苷酸重复序列的情况下,所翻译的DPR蛋白质包括聚-(Gly-Ala),并且在较小程度上,包括聚-(Gly-Pro)和聚-(Gly-Arg)。较短同功异构物b(NP_659442.2)具有序列MSTLCPPPSPAVAKTE

IALSGKSPLLAATFAYWDNILGPRVRHIWAPKTEQVLLSDGEITFLANHTLNGEILRNAESGAIDVKFFVLSEKGV IIVSLIFDGNWNGDRSTYGLSIILPQTELSFYLPLHRVCVDRLTHIIRKGRIWMHKERQENVQKIILEGTERMEDQ GQSIIPMLTGEVIPVMELLSSMKSHSVPEEIDIADTVLNDDDIGDSCHEGFLLK(SEQ ID NO:228)。

[0090] 在一些实施例中,本公开提供经专门设计以修饰真核细胞中的C9orf72基因的系统。在一些情况下,系统经设计以敲减或敲除C9orf72基因。在其它情况下,系统经设计以校正C9orf72基因中的一个或多个突变。在一些实施例中,系统经设计以切除六核苷酸重复序列并且恢复细胞表达功能C9orf72蛋白质的能力。在一些实施例中,系统经设计以校正编码HRS和/或DPR的RNA转录物的C9orf72基因的六核苷酸重复序列GGGGCC突变并且恢复细胞表达功能C9orf72蛋白质的能力。

[0091] 一般来说,C9orf72基因的任何部分可以使用本文所提供的可编程组合物和方法靶向。在一些实施例中,CRISPR核酸酶是2类V型核酸酶。在一些实施例中,2类V型核酸酶选自由以下组成的组:Cas12a、Cas12b、Cas12c、Cas12d(CasY)、Cas12J和CasX。在一些实施例中,2类V型核酸酶为CasX。在一些实施例中,本公开提供包含一种或多种CasX蛋白和一种或多种引导核酸(gNA)作为CasX:gNA系统和任选地一种或多种供体模板核酸的系统。下文描述这些组分中的每一个和其在C9orf72基因的编辑中的用途。

[0092] 在一些实施例中,本公开提供了本文所述的任一实施例的CasX和gNA的基因编辑对,其能够在其用于基因编辑之前结合在一起并且因此"预复合"为核糖核蛋白复合物(RNP)。使用预复合的RNP在将系统组分递送至细胞或靶核酸序列以编辑靶核酸序列方面赋予优势。在一些实施例中,功能性RNP可以通过电泳或化学手段离体递送至细胞。在其它实施例中,功能性RNP可通过载体以其功能形式离体或体内递送,或表达且接着复合在一起成为RNP。gNA可以通过包括具有与靶核酸序列互补的核苷酸序列的靶向序列(或"间隔区")来为复合物提供靶特异性,而预复合的CasX:gNA的CasX蛋白提供了位点特异性活性,诸如靶序列的裂解或切割,所述活性由于其与gNA的关联被引导至靶核酸序列(例如,待修饰的C9orf72基因)内的靶位点(例如,在靶位点稳定化)。下文更完整地描述CasX:gNA系统的CasX蛋白和gNA组分和其序列、特征和功能。

[0093] 在一些实施例中,用于编辑C9orf72基因的CasX:gNA系统可以任选地进一步包含:包含全部或至少一部分编码C9orf72蛋白质的基因的供体模板、非编码区或C9orf72调节元件,其中所述供体模板包含用于插入的相较于野生型C9orf72基因的一个或多个突变,以基因敲除或基因敲减(下文更完整地描述)具有一个或多个突变或HRS的靶核酸序列。在其它情况下,CasX:gNA系统可任选地进一步包含用于引入(或敲入)编码生理学上正常数目的六核苷酸重复序列的基因的全部或一部分,或用于产生野生型C9orf72蛋白质的序列(SEQ ID NO:227或228),或用于在靶细胞中产生生理学上正常水平的C9orf72的序列的供体模板。在一些实施例中,供体模板包含野生型C9orf72基因的至少约20、至少约50、至少约100、至少约200、至少约300、至少约400、至少约500、至少约600、至少约700、至少约800、至少约900、至少约1000、至少约10,000、至少约15,000或至少约25,000个核苷酸,其中C9orf72基因部分选自由以下组成的组:C9orf72外显子、C9orf72内含子、C9orf72内含子-外显子接界、C9orf72调节元件、C9orf72编码区、C9orf72非编码区或C9orf72基因的全部。在一些实施例中,C9orf72基因部分包含C9orf72外显子序列、C9orf72内含子序列、C9orf72内含子-外显子接界序列、C9orf72非编码区或C9orf72调节元件序列中的任一个的组合。在具体实施例

中,供体模板包含具有生理学上正常数目的GGGGCC序列的六核苷酸重复序列的序列,其中在供体模板插入后,置换C9orf72基因的六核苷酸重复序列扩增段。在其它实施例中,供体多核苷酸包含野生型C9orf72基因的至少约10到约15,000个核苷酸、至少约100到约10,000个核苷酸、至少约400到约6000个核苷酸、至少约600到约4000个核苷酸或至少约1000到约2000个核苷酸。在一些实施例中,供体模板为单链DNA模板或单链RNA模板。在其它实施例中,供体模板为双链DNA模板。

[0094] III.遗传编辑系统的引导核酸

[0095] 在另一方面,本公开涉及包含与C9orf72基因的靶核酸序列互补的靶向序列的引导核酸(gNA),其中gNA能够与CRISPR蛋白质形成复合物,所述CRISPR蛋白质具有对包含互补非靶链中的TC基序的前间隔区邻近基序(PAM)序列的特异性,并且其中所述PAM序列位于与靶核酸的靶链中的靶核酸序列互补的非靶链中的序列的5'的1个核苷酸处。在一些实施例中,gNA能够与2类V型CRISPR核酸酶形成复合物。在一个具体的实施例中,gNA能够与CasX核酸酶形成复合物。

[0096] 在一些实施例中,本公开提供在CasX:gNA系统中利用的gNA,其在细胞中的基因组编辑中具有效用,在编辑C9orf72基因中具有效用。本公开提供了特别设计的引导核酸("gNA"),其具有与C9orf72基因互补(并因此能够与之杂交)的靶向序列作为基因编辑CasX:gNA系统的组分。可在实施例的gNA中使用的针对C9orf72靶核酸的靶向序列的代表性但非限制性实例被呈现为SEQ ID NO:309-343、363-2100和2295-21835。在一些实施例中,gNA为脱氧核糖核酸分子("gDNA");在一些实施例中,gNA为核糖核酸分子("gRNA"),且在其它实施例中,gNA为嵌合体,且包含DNA及RNA两者。如本文所用,术语gNA、gRNA和gDNA涵盖天然存在的分子,以及序列变异体。

[0097] 设想在一些实施例中,多种gNA(例如多种gRNA)在CasX:gNA系统中递送,以用于修饰编码C9orf72蛋白质的一个或多个区域的基因、C9orf72基因的非编码区域或C9orf72调节元件。例如,当需要删除基因的调节元件或HRS时,可以使用一对具有针对靶核酸序列的不同或重叠区域的靶向序列的gNA,以便在基因内的两个不同或重叠位点结合和裂解。在要删除六核苷酸重复序列区域的其它情况下,可以使用一对gNA以便在C9orf72基因内六核苷酸重复序列的5'和3'处的两个不同位点处结合和裂解,使得切除HRS,其随后通过非同源末端连接(NHEJ)、同源定向修复(HDR)、同源独立靶向整合(HITI)、微同源介导末端连接(MMEJ)、单链退火(SSA)或碱基切除修复(BER)编辑。可用于编辑HRS的示范性gNA对展示于以下表16中,且下文实例18中描述用以实现编辑的示范性方法。

[0098] a.参考gNA和gNA变异体.

[0099] 在一些实施例中,本公开的gNA包含天然存在的gNA("参考gNA")的序列。在其它情况下,本公开的参考gNA可经受一种或多种突变诱发方法,例如本文所述的突变诱发方法,其可包括深度突变进化(DME)、深度突变扫描(DMS)、易错PCR、卡匣突变诱发、随机突变诱发、交错延伸PCR、基因改组或结构域交换,以便产生一个或多个具有相对于参考gNA增强或改变的特性的gNA变异体。gNA变异体亦包括包含一个或多个外源序列,例如与5'或3'端融合,或插入内部的变异体。参考gNA的活性可用作与gNA变异体的活性进行比较的基准,由此测量gNA变异体的功能或其它特性的改进。在其它实施例中,参考gNA可经受一个或多个有意的特异性靶向突变以产生gNA变异体,例如合理设计的变异体。

[0100] 本公开的gNA包含两个区段:靶向序列和蛋白结合区段。gNA的靶向区段包括核苷酸序列(可互换地称为引导序列、间隔区、靶向子或靶向序列),其与靶核酸序列(例如,靶ssRNA、靶ssDNA、双链靶DNA的一条链等)内的特定序列(靶位点)互补(且因此与其杂交),在下文更充分描述。gNA的靶向序列能够结合至靶核酸序列,包括编码序列、编码序列的互补序列、非编码序列,且结合至调节元件。蛋白质结合区段(或"激活因子"或"蛋白质结合序列")作为复合物与CasX蛋白相互作用(例如,结合),形成RNP(下文更全面地描述)。

[0101] 在双引导RNA(dgRNA)的情况下,靶向子和活化子部分各自具有双螺旋体形成区段,其中靶向子的双螺旋体形成区段和活化子的双螺旋体形成区段彼此具有互补性,并彼此杂交以形成双链双螺旋体(gRNA的dsRNA双螺旋体)。当所述gNA是gRNA,术语"靶向子"或"靶向子RNA"在本文中用于指CasX双引导RNA(并且因此当"激活因子"和"靶向子"例如通过插入核苷酸连接在一起时,CasX单引导RNA)的crRNA样分子(crRNA:"CRISPR RNA")。crRNA有一个5'区域,所述区域与tracrRNA退火,然后是靶向序列的核苷酸。因此,举例来说,引导RNA(dgRNA或sgRNA)包含引导序列及crRNA的双链体形成区段,其亦可称为crRNA重复序列。对应tracrRNA样分子(活化子)也包含核苷酸的双螺旋体形成段,其形成引导RNA的蛋白结合区段的dsRNA双螺旋体的另一半。因此,靶向子和激活因子作为相应的一对杂交以形成双引导NA,在本文中称为"双引导NA"、"双分子gNA"、"dgNA"、"双分子引导NA"或"二分子引导NA"。CasX蛋白对靶核酸序列(例如,基因组DNA)的位点特异性结合和/或裂解可以发生在由gNA的靶向序列和靶核酸序列之间的碱基配对互补性测定的一个或多个位置(例如,靶核酸的序列)。因此,例如,本公开的gNA具有与靶核酸互补的序列,因此可以与靶核酸杂交,所述靶核酸与与TC PAM基序或PAM序列(诸如ATC、CTC、GTC或TTC)互补的序列相邻。由于引导序列的靶向序列与靶核酸序列的序列杂交,因此只要考虑到PAM序列的位置,靶向子就可以由用户修饰以与特定靶核酸序列杂交。因此,在一些情况下,靶向子的序列可以为非天然存在的序列。在其它情况下,靶向子的序列可以是天然存在的序列,源自待编辑的基因。在其它实施例中,gNA的活化子和靶向子彼此共价连接(而不是彼此杂交)且包含单分子,在本文中称为"单分子gNA"、"一分子引导NA"、"单引导NA"、"单引导RNA"、"单分子引导RNA"、"一分子引导RNA"、"单引导DNA"、"单分子DNA"或"一分子引导DNA"("sgNA"、"sgRNA"或"sgDNA")。在一些实施例中,sgNA包括"激活因子"或"靶向子"且因此可分别为"激活因子-RNA"及"靶向子-RNA"。

[0102] 总的来说,本公开的组装gNA包含四个不同的区域,或结构域:RNA三链体、支架茎、延伸茎和靶向序列,在本公开的实施例中,它们对靶核酸具有特异性且定位于gNA的3'端。RNA三链体、支架茎和延伸茎在一起被称为gNA的"支架"。

[0103] b.RNA三螺旋体

[0104] 在本文提供的引导NA(包括参考sgNA)的一些实施例中,存在RNA三螺旋体,并且RNA三螺旋体包含UUU--nX(～4-15)--UUU茎环(SEQ ID NO:19)的序列,其在2个中间茎环(支架茎环和延伸茎环)之后以AAAG结束,形成也可延伸穿过三螺旋体进入双螺旋假结中的假结。三螺旋体的UU-UUU-AAA序列形成为间隔区、支架茎和延伸茎之间的连接。在示范性参考CasX sgNA中,首先对UUU-环-UUU区域进行编码,然后是支架茎环,且接着是延伸茎环,其由四环连接,且接着AAAG封闭三链体,随后变为间隔区。

[0105] c.支架茎环

[0106] 在本公开的sgNA的一些实施例中，三螺旋区之后是支架茎环。支架茎环为与CasX蛋白(例如参考或CasX变异蛋白)结合的gNA区域。在一些实施例中，支架茎环为相当短且稳定的茎环。在一些情况下，支架茎环不耐受许多变化，且需要一些形式的RNA气泡。在一些实施例中，支架茎是CasX sgNA功能所需的。尽管CasX sgNA的支架茎可能与Cas9的连接茎类似地作为重要茎环，但在一些实施例中，其具有与CRISPR/Cas系统中发现的许多其它茎环不同的所需凸起(RNA气泡)。在一些实施例中，这个凸起的存在在与不同CasX蛋白相互作用的sgNA中是保守的。gNA的支架茎环序列的示范性序列包含序列CCAGCGACUAUGUCGUAUGG(SEQ ID NO:20)。在其它实施例中，本公开提供了gNA变异体，其中支架茎环被来自具有近端5'和3'端的异源性RNA源的RNA茎环序列取代，例如但不限于选自MS2、Qβ、U1发夹II、Uvsx或PP7茎环的茎环序列。在一些情况下，gNA的异源性RNA茎环能够结合蛋白质、RNA结构、DNA序列或小分子。

[0107] d.延伸茎环

[0108] 在本公开的CasX sgNA的一些实施例中，支架茎环之后是延伸茎环。在一些实施例中，延伸茎包含很大程度上未经CasX蛋白结合的合成tracr及crRNA融合物。在一些实施例中，延伸茎环可为高度展性的。在一些实施例中，通过延伸茎环中tracrRNA与crRNA之间的GAAA四环接头或GAGAAA接头制得单引导gRNA。在一些情况下，CasX sgNA的靶向子和活化子通过中间核苷酸彼此连接，且接头的长度可为3至20个核苷酸。在本公开的CasX sgNA的一些实施例中，延伸茎为位于核糖核蛋白复合物中的CasX蛋白外部的大型32-bp环。sgNA的延伸茎环序列的示范性序列包含序列GCGCUUAUUUAUCGGAGAGAAAUCCGAUAAAUAAGAAGC(SEQ ID NO:21)。在一些实施例中，延伸茎环包含GAGAAA间隔区序列。在一些实施例中，本公开提供gNA变异体，其中延伸茎环被替换为来自具有近端5'及3'端的异源性RNA源的RNA茎环序列，例如但不限于选自MS2、Qβ、U1发夹II、Uvsx或PP7茎环的茎环序列。在此类状况下，异源性RNA茎环增加gNA的稳定性。在其它实施例中，本公开提供了具有延伸茎环区的gNA变异体，所述茎环区包含至少10、至少100、至少500、至少1000或至少10,000个核苷酸，或至少10-10,000、至少10-1000或至少10-100个核苷酸。

[0109] e.靶向序列

[0110] 在本公开的gNA的一些实施例中，延伸茎环之后是形成三螺旋体的部分的区域，且接着是靶向序列。靶向序列将CasX核糖核蛋白整体复合物靶向至C9orf72基因的靶核酸序列的特定区域。因此，例如，当TC PAM基序或PAM序列TTC、ATC、GTC或CTC中的任一个位于与靶序列互补的非靶链序列的5'的1个核苷酸处时，本公开的CasX gNA靶向序列，作为RNP的组分，与真核细胞中的核酸中的C9orf72基因的一部分(例如真核染色体、染色体序列、真核RNA等)具有序列互补性，且因此可与其杂交。可以修饰gNA的靶向序列，使得gNA可以靶向任何所需靶核酸序列的所需序列，只要考虑到PAM序列位置即可。在一些实施例中，gNA支架在靶向序列的5'端，靶向序列位于gNA的3'端。在一些实施例中，被RNP的核酸酶识别的PAM基序序列是TC。在其它实施例中，被RNP的核酸酶识别的PAM序列是NTC。

[0111] 在一些实施例中，gNA的靶向序列对编码包含一个或多个突变的C9orf72蛋白质的基因的一部分具有特异性。在一些实施例中，gNA的靶向序列对C9orf72外显子具有特异性。在一些实施例中，gNA的靶向序列对C9orf72内含子具有特异性。在一些实施例中，gNA的靶向序列对C9orf72内含子-外显子接界具有特异性。在一些实施例中，gNA的靶向序列具有与

C9orf72调节元件、C9orf72编码区、C9orf72非编码区或其组合杂交的序列。在具体实施例中，gNA的靶向序列与在HRS的5'处的序列杂交。在使用两种或更多种gNA的一些实施例中，gNA的第一gNA靶向序列与在HRS的5'处的序列杂交并且第二gNA与在HRS的3'处的序列杂交。在一些实施例中，gNA的靶向序列与包含C9orf72基因的一种或多种单核苷酸多态性(SNP)的序列或其互补序列互补。在C9orf72编码序列内或在C9orf72非编码序列内的SNP皆在本公开的范围内。在其它实施例中，gNA的靶向序列与C9orf72基因的基因间区域的序列或与C9orf72基因的基因间区域互补的序列互补。

[0112] 在一些实施例中，gNA的靶向序列对调节C9orf72表达的调节元件具有特异性。此类C9orf72调节元件包括但不限于启动子区、增强子区、基因间区、5'非翻译区(5'UTR)、3'非翻译区(3'UTR)、基因间区、基因增强子元件、保守元件及包含顺式调节元件的区域。启动子区旨在涵盖C9orf72起始点100kb内的核苷酸，或者在基因增强子元件或保守元件的情况下，可以在C9orf72基因1Mb或更远的地方。在一些实施例中，本公开提供了具有与C9orf72调节元件杂交的靶向序列的gNA。在前述内容中，靶标是其中靶标的编码基因打算经敲除或敲减，以使得包含C9orf72基因产物的突变或六核苷酸重复的C9orf72蛋白质在细胞中不表达或以较低水平表达。在一些实施例中，本公开提供了CasX:gNA系统，其中gNA的靶向序列(或间隔区)与编码C9orf72的核酸序列、C9orf72蛋白的一部分、C9orf72调节元件的一部分或C9orf72基因的一部分的互补序列互补。在一些实施例中，gNA的靶向序列具有14至35个连续核苷酸。在一些实施例中，靶向序列具有14、15、16、18、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34或35个连续核苷酸。在一些实施例中，靶向序列由21个连续核苷酸组成。在一些实施例中，靶向序列由20个连续核苷酸组成。在一些实施例中，靶向序列由19个连续核苷酸组成。在一些实施例中，靶向序列由18个连续核苷酸组成。在一些实施例中，靶向序列由17个连续核苷酸组成。在一些实施例中，靶向序列由16个连续核苷酸组成。在一些实施例中，靶向序列由15个连续核苷酸组成。在一些实施例中，靶向序列具有14、15、16、17、18、19、20或21个连续核苷酸，且靶向序列可包含0至5、0至4、0至3或0至2个相对于靶核酸序列的失配且保留足够结合特异性，以使得包含含有靶向序列的gNA的RNP可与靶核酸形成互补键。

[0113] 野生型C9orf72核酸的靶向序列的代表性但非限制性实例作为SEQ ID NO:309-343、363-2100和2295-21835呈现在表3和15中。在一些实施例中，本公开提供靶向序列，其包含与表3和15中作为SEQ ID NO:309-343、363-2100和2295-21835的序列具有至少50％同一性、至少55％同一性、至少60％同一性、至少65％同一性、至少70％同一性、至少75％同一性、至少80％同一性、至少85％同一性、至少90％同一性、至少95％同一性或100％同一性的序列。在一些实施例中，gNA的靶向序列包含从序列的3'端去除单个核苷酸的SEQ ID NO:2281-159093的序列。在其它实施例中，gNA的靶向序列包含从序列的3'端去除两个核苷酸的SEQ ID NO:309-343、363-2100和2295-21835的序列。在其它实施例中，gNA的靶向序列包含从序列的3'端去除三个核苷酸的SEQ ID NO:309-343、363-2100和2295-21835的序列。在其它实施例中，gNA的靶向序列包含从序列的3'端去除四个核苷酸的SEQ ID NO:309-343、363-2100和2295-21835的序列。在其它实施例中，gNA的靶向序列包含从序列的3'端去除五个核苷酸的SEQ ID NO:309-343、363-2100和2295-21835的序列。在段落的前述实施例中，胸腺嘧啶(T)核苷酸可以取代任何靶向序列中的一个或多个或所有尿嘧啶(U)核苷酸，使得

gNA可以是gDNA或gRNA，或RNA和DNA的嵌合体，或在间隔区的编码序列并入表达载体中的那些情况下。在一些实施例中，SEQ ID NO:309-343、363-2100和2295-21835的靶向序列具有至少1、2、3、4、5或6个或更多个取代尿嘧啶核苷酸的胸腺嘧啶核苷酸。在其它实施例中，本公开的gNA、gRNA或gDNA包含SEQ ID NO:309-343、363-2100和2295-21835的1、2、3或更多个靶向序列，或与SEQ ID NO:309-343、363-2100和2295-21835的一个或多个序列具有至少50％同一性、至少55％同一性、至少60％同一性、至少65％同一性、至少70％同一性、至少75％同一性、至少80％同一性、至少85％同一性、至少90％同一性、至少95％同一性或100％同一性的靶向序列。

[0114] 在一些实施例中，靶向序列与编码SEQ ID NO:227或228的C9orf72蛋白的突变或破坏C9orf72蛋白的功能或表达的六核苷酸重复的核酸序列互补。

[0115] 在一些实施例中，CasX:gNA系统包含第一gNA且进一步包含第二(和任选地第三、第四或第五)gNA，其中第二gNA或额外gNA具有与相较于第一gNA的靶向序列的靶核酸序列或其互补序列的不同部分互补的靶向序列；例如，第一gNA靶向六核苷酸重复序列5'且第二gNA靶向六核苷酸重复序列3'。通过选择gNA的靶向序列，可以使用本文所述的CasX:gNA系统修饰或编辑靶核酸序列的指定区域。

[0116] f.gNA支架

[0117] 在一些实施例中，CasX参考gRNA包含分离或衍生自δ变形菌纲(Deltaproteobacteria)的序列。在一些实施例中，序列为CasX tracrRNA序列。分离或衍生自δ变形菌纲的示范性CasX参考tracrRNA序列可以包括：ACAUCUGGCGCGUUUAUUCCAUUACUUUGGAGCCAGUCCCAGCGACUAUGUCGUAUGGACGAAGCGCUUAUUUAUCGGAGA(SEQ ID NO:22)和ACAUCUGGCGCGUUUAUUCCAUUACUUUGGAGCCAGUCCCAGCGACUAUGUCGUAUGGACGAAGCGCUUAUUUAUCGG(SEQ ID NO:23)。分离或衍生自δ变形菌纲的示范性crRNA序列可以包含CCGAUAAGUAAAACGCAUCAAAG(SEQ ID NO:24)的序列。在一些实施例中，CasX参考gNA包含与分离或衍生自δ变形菌纲的序列具有至少60％同一性、至少65％同一性、至少70％同一性、至少75％同一性、至少80％同一性、至少81％同一性、至少82％同一性、至少83％同一性、至少84％同一性、至少85％同一性、至少86％同一性、至少86％同一性、至少87％同一性、至少88％同一性、至少89％同一性、至少89％同一性、至少90％同一性、至少91％同一性、至少92％同一性、至少93％同一性、至少94％同一性、至少95％同一性、至少96％同一性、至少97％同一性、至少98％同一性、至少99％同一性、至少99.5％同一性或100％同一性的序列。

[0118] 在一些实施例中，CasX参考引导RNA包含分离或衍生自浮霉菌门(Planctomycetes)的序列。在一些实施例中，序列为CasX tracrRNA序列。分离或衍生自δ变形菌纲的示范性CasX参考tracrRNA序列可以包括：UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGA(SEQ ID NO:25)和UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGG(SEQ ID NO:26)。分离或衍生自浮霉菌门的示范性crRNA序列可以包含UCUCCGAUAAAUAAGAAGCAUCAAAG(SEQ ID NO:27)的序列。在一些实施例中，CasX参考gNA包含与分离或衍生自浮霉菌门的序列具有至少60％同一性、至少65％同一性、至少70％同一性、至少75％同一性、至少80％同一性、至少81％同一性、至少82％同一性、至少83％同一性、至少84％同一性、至少85％同一性、至少86％同一性、至少86％同一性、至少87％同一性、至少88％同一性、至少89％同一

性、至少89%同一性、至少90%同一性、至少91%同一性、至少92%同一性、至少93%同一性、至少94%同一性、至少95%同一性、至少96%同一性、至少97%同一性、至少98%同一性、至少99%同一性、至少99.5%同一性或100%同一性的序列。

[0119] 在一些实施例中，CasX参考gNA包含分离或衍生自宋氏细菌暂定种(Candidatus Sungbacteria)的序列。在一些实施例中，序列为CasX tracrRNA序列。分离或衍生自宋氏细菌暂定种的示范性CasX参考tracrRNA序列可以包含以下序列：GUUUACACACUCCCUCUCAUAGGGU(SEQ ID NO:28)、GUUUACACACUCCCUCUCAUGAGGU(SEQ ID NO:29)、UUUUACAUACCCCCUCUCAUGGGAU(SEQ ID NO:30)和GUUUACACACUCCCUCUCAUGGGGG(SEQ ID NO:31)。在一些实施例中，CasX参考引导RNA包含与分离或衍生自宋氏细菌暂定种的序列具有至少60%同一性、至少65%同一性、至少70%同一性、至少75%同一性、至少80%同一性、至少81%同一性、至少82%同一性、至少83%同一性、至少84%同一性、至少85%同一性、至少86%同一性、至少86%同一性、至少87%同一性、至少88%同一性、至少89%同一性、至少89%同一性、至少90%同一性、至少91%同一性、至少92%同一性、至少93%同一性、至少94%同一性、至少95%同一性、至少96%同一性、至少97%同一性、至少98%同一性、至少99%同一性、至少99.5%同一性或100%同一性的序列。

[0120] 表1提供了参考gRNA tracr的序列和支架序列。在一些实施例中，本公开提供了gNA序列，其中gNA具有支架，其包含相对于具有表1的SEQ ID NO:4-16中的任一个的序列的参考gNA序列具有至少一个核苷酸修饰的序列。应了解，在其中载体包含用于gNA的DNA编码序列或其中gNA为gDNA或RNA及DNA的嵌合体的那些实施例中，胸腺嘧啶(T)碱基可取代本文所述的gNA序列实施例(包括表1和表2的序列)中的任一个的尿嘧啶(U)碱基。

[0121] 表1.参考gRNA tracr和支架序列

[0122]

SEQ ID NO.	核苷酸序列
4	ACAUCUGGCGCGUUUAUUCCAUUACUUUGGAGCCAGUCCCAGCGACUAUGUCGUAUGGACGAAGCGCUUAUUUAUCGGAGAGAAACCGAUAAGUAAAACGCAUCAAAG
5	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGAAAUCCGAUAAAUAAGAAGCAUCAAAG
6	ACAUCUGGCGCGUUUAUUCCAUUACUUUGGAGCCAGUCCCAGCGACUAUGUCGUAUGGACGAAGCGCUUAUUUAUCGGAGA

[0123]

7	ACAUCUGGCGCGUUUAUUCCAUUACUUUGGAGCCAGUCCCAGCGACUAUGUCGUAUGGACGAAGCGCUUAUUUAUCGG
8	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGA
9	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGG
10	GUUUACACACUCCCUCUCAUAGGGU
11	GUUUACACACUCCCUCUCAUGAGGU
12	UUUUACAUACCCCCUCUCAUGGGAU
13	GUUUACACACUCCCUCUCAUGGGGG
14	CCAGCGACUAUGUCGUAUGG
15	GCGCUUAUUUAUCGGAGAGAAAUCCGAUAAAUAAGAAGC
16	GGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGA

[0124] g.gNA变异体

[0125] 在另一方面中，本公开涉及引导核酸变异体(替代地，在本文中称为“gNA变异体”

或“gRNA变异体”),其包含相对于参考gRNA支架的一个或多个修饰。如本文所用,“支架”是指除间隔序列的外的gNA功能所需的gNA的所有部分。

[0126] 在一些实施例中,gNA变异体包含相对于本公开的参考gRNA序列的一个或多个核苷酸取代、插入、缺失或交换或替换区域。在一些实施例中,突变可发生于参考gRNA的任何区中以产生gNA变异体。在一些实施例中,gNA变异序列的支架与SEQ ID NO:4或SEQ ID NO:5的序列具有至少20%、至少30%、至少40%、至少50%、至少60%、或至少70%、至少80%、至少85%、至少约90%、至少约95%、至少约96%、至少约97%、至少约98%、或至少约99%同一性。

[0127] 在一些实施例中,gNA变异体包含参考gRNA的一个或多个区域内的一个或多个核苷酸变化,所述变化改进了参考gRNA的特征。示范性区包括RNA三螺旋体、假结、支架茎环及延伸茎环。在一些情况下,变异体支架茎进一步包含气泡。在其它情况下,变异体支架进一步包含三螺旋环区。在其它情况下,变异体支架进一步包含5’非结构化区。在一些实施例中,gNA变异体支架包含与SEQ ID NO:14具有至少60%序列同一性的支架茎环。在其它实施例中,gNA变异体包含具有CCAGCGACUAUGUCGUAGUGG(SEQ ID NO:32)的序列的支架茎环。在其它实施例中,本公开提供了相对于SEQ ID NO:5包含C18G取代、G55插入、U1缺失和经修饰的延伸茎环的gNA支架,其中原始6nt环和13个在环最近端的碱基对(总共32个核苷酸)经Uvsx发夹(4nt环和5个环近侧碱基对;总共14个核苷酸)取代,且延伸茎的环远侧碱基通过A99的缺失和G64U的取代而转化为与新Uvsx发夹邻接的完全碱基配对茎。在前述实施例中,gNA支架包含序列

[0128] ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAGCAUCAAAG(SEQ ID NO:33)。

[0129] 当将变异体gNA与本文描述的参考gRNA进行比较时,具有一个或多个改进功能或特征,或添加一种或多种新功能的所有gNA变异体均被设想为在本公开的范围内。这种gNA变异体的代表性实例是引导序列174(SEQ ID NO:2238),其设计描述于实例中。在一些实施例中,gNA变异体向包含gNA变异体的RNP添加新功能。在一些实施例中,gNA变异体具有选自以下的改进特征:改进的稳定性;改进的溶解度;改进的gNA转录;改进的核酸酶活性抗性;增加的gNA折叠速率;折叠期间减少的副产物形成;增加的生产性折叠;改进的与CasX蛋白的结合亲和力;当与CasX蛋白复合时改进的与靶DNA的结合亲和力;当与CasX蛋白复合时改进的基因编辑;当与CasX蛋白复合时改进的编辑特异性;以及当与CasX蛋白复合时改进的在靶DNA的编辑中利用较大范围的一个或多个PAM序列,包括ATC、CTC、GTC或TTC的能力,及其任何组合。在一些情况下,gNA变异体的改进特征中的一个或多个是相对于SEQ ID NO:4或SEQ ID NO:5的参考gNA改进至少约1.1至约100,000倍。在其它情况下,gNA变异体的一种或多种改进特征是相对于SEQ ID NO:4或SEQ ID NO:5的参考gNA改进至少约1.1、至少约10、至少约100、至少约1000、至少约10,000、至少约100,000倍或更大改进。在其它情况下,gNA变异体的改进特征中的一个或多个是相对于SEQ ID NO:4或SEQ ID NO:5的参考gNA改进约1.1至100,00倍、约1.1至10,00倍、约1.1至1,000倍、约1.1至500倍、约1.1至100倍、约1.1至50倍、约1.1至20倍、约10至100,00倍、约10至10,00倍、约10至1,000倍、约10至500倍、约10至100倍、约10至50倍、约10至20倍、约2至70倍、约2至50倍、约2至30倍、约2至20倍、约2至10倍、约5至50倍、约5至30倍、约5至10倍、约100至100,00倍、约100至10,00倍、约100至1,

000倍、约100至500倍、约500至100,00倍、约500至10,00倍、约500至1,000倍、约500至750倍、约1,000至100,00倍、约10,000至100,00倍、约20至500倍、约20至250倍、约20至200倍、约20至100倍、约20至50倍、约50至10,000倍、约50至1,000倍、约50至500倍、约50至200倍或约50至100倍。在其它情况下,gNA变异体的一种或多种改进特征是相对于SEQ ID NO:4或SEQ ID NO:5的参考gNA改进约1.1倍、1.2倍、1.3倍、1.4倍、1.5倍、1.6倍、1.7倍、1.8倍、1.9倍、2倍、3倍、4倍、5倍、6倍、7倍、8倍、9倍、10倍、11倍、12倍、13倍、14倍、15倍、16倍、17倍、18倍、19倍、20倍、25倍、30倍、40倍、45倍、50倍、55倍、60倍、70倍、80倍、90倍、100倍、110倍、120倍、130倍、140倍、150倍、160倍、170倍、180倍、190倍、200倍、210倍、220倍、230倍、240倍、250倍、260倍、270倍、280倍、290倍、300倍、310倍、320倍、330倍、340倍、350倍、360倍、370倍、380倍、390倍、400倍、425倍、450倍、475倍或500倍。

[0130] 在一些实施例中,可如下创建gNA变异体:通过使参考gRNA经受一种或多种诱变方法,如下文描述的诱变方法,其可以包括深度突变进化(DME)、深度突变扫描(DMS)、易错PCR、盒式诱变、随机诱变、交错延伸PCR、基因改组或结构域交换,以便产生本公开的gNA变异体。参考gRNA的活性可用作与gNA变异体的活性进行比较的基准,从而衡量gNA变异体功能的改进。在其它实施例中,参考gRNA可经受一个或多个有意的靶向突变、取代或结构域交换以产生gNA变异体,例如合理设计的变异体。由此类方法产生的示范性gRNA变异体描述于实例中,且gNA支架的代表性序列呈现于表2中。

[0131] 在一些实施例中,gNA变异体包含相比于参考引导核酸支架序列的一个或多个修饰,其中一个或多个修饰选自:gNA变异体区中的至少一个核苷酸取代;gNA变异体区中的至少一个核苷酸缺失;gNA变异体区中的至少一个核苷酸插入;gNA变异体区的全部或一部分的取代;gNA变异体区的全部或一部分的缺失;或前述的任何组合。在一些情况下,修饰是在一个或多个区中取代gNA变异体中的1至15个连续或非连续核苷酸。在其它情况下,修饰是在一个或多个区中缺失gNA变异体中的1至10个连续或非连续核苷酸。在其它情况下,修饰是在一个或多个区中插入gNA变异体中的1至10个连续或非连续核苷酸。在其它情况下,修饰为通过来自具有近端5'及3'端的异源性RNA源的RNA茎环序列取代支架茎环或延伸茎环。在一些情况下,本公开的gNA变异体在一个区中包含两个或更多个修饰。在其它情况下,本公开的gNA变异体在两个或更多个区中包含修饰。在其它情况下,gNA变异体包含此段中所述的前述修饰的任何组合。

[0132] 在一些实施例中,将5'G添加到gNA变异序列以用于体内表达,因为当+1核苷酸为G时,从U6启动子的转录更高效且相对于起始位点更一致。在其它实施例中,将两个5'G添加到gNA变异序列用于体外转录以提高生产效率,因为T7聚合酶强烈偏好+1位置中的G和+2位置中的嘌呤。在一些情况下,将5'G碱基添加至表1的参考支架。在其它情况下,将5'G碱基添加至表2的变异体支架。

[0133] 表2提供了示范性gNA变异体支架序列。在表2中,(-)表示在相对于SEQ ID NO:5的参考序列的指定位置处的缺失,(+)表示在相对于SEQ ID NO:5的指示位置处插入指定碱基,(:)表示相对于SEQ ID NO:5的缺失或取代的指定起始:终止坐标处的碱基范围,且多个插入、缺失或取代通过逗号分隔;例如,A14C,U17G。在一些实施例中,gNA变异体支架包含表2中所列为SEQ ID NO:2101-2294的序列中的任一个,或与其具有至少约50%、至少约60%、至少约70%、至少约80%、至少约90%、至少约95%、至少约95%、至少约96%、至少约97%、

至少约98％、至少约99％序列同一性的序列。应了解，在其中载体包含用于gNA的DNA编码序列或其中gNA为gDNA或RNA及DNA的嵌合体的那些实施例中，胸腺嘧啶(T)碱基可取代本文所述的gNA序列实施例中的任一个的尿嘧啶(U)碱基。

[0134] 表2.示范性gNA支架序列

[0135]

SEQ ID NO:	名称或修饰	核苷酸序列
2101	噬菌体复制稳定	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCAGGUGGGACGACCUCUCG GUCGUCCUAUCUGAAGCAUCAAAG
2102	吻合环_b1	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUGCUCGACGCGUCCUCGA GCAGAAGCAUCAAAG
2103	吻合环_a	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUGCUCGCUCCGUUCGAGC AGAAGCAUCAAAG
2104	32: uvsX 发夹	GUACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAG CGACUAUGUCGUAUGGGUAAAGCGCCCUCUUCGGAGGGAAGC AUCAAAG
2105	PP7	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCAGGAGUUUCUAUGGAAAC CCUGAAGCAUCAAAG
2106	64: 三重突变，延伸茎截断	GUACUGGCGCCUUUAUCUCAUUACUUUGAGAGCCAUCACCAG CGACUAUGUCGUAUGGGUAAAGCGCUUACGGACUUCGGUCCG UAAGAAGCAUCAAAG
2107	超稳定四环	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUGCGCUUGCGCAGAAGCA UCAAAG
2108	C18G	UACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGAAA UCCGAUAAAUAAGAAGCAUCAAAG
2109	U17G	UACUGGCGCUUUUAUCGCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGAAA UCCGAUAAAUAAGAAGCAUCAAAG
2110	CUUCGG 环	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGACUUC GGUCCGAUAAAUAAGAAGCAUCAAAG
2111	MS2	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCACAUGAGGAUUACCCAUG UGAAGCAUCAAAG
2112	-1, A2G, -78, G77U	GCUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGUGAGAAAUC CGAUAAAUAAGAAGCAUCAAAG

[0136]

SEQ ID NO:	名称或修饰	核苷酸序列
2113	QB	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUGCAUGUCUAAGACAGCA GAAGCAUCAAAG
2114	45,44 发夹	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCAGGGCUUCGGCCGAAGCA UCAAAG
2115	U1A	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCAAUCCAUUGCACUCCGGA UUGAAGCAUCAAAG
2116	A14C, U17G	UACUGGCGCUUUUCUCGCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGAAA UCCGAUAAAUAAGAAGCAUCAAAG
2117	CUUCGG 环经修饰	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGACUUCGG UCCGAUAAAUAAGAAGCAUCAAAG
2118	吻合环_b2	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUGCUCGUUUGCGGCUACG AGCAGAAGCAUCAAAG
2119	-76:78, -83:87	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGAGAGAUAA AUAAGAAGCAUCAAAG
2120	-4	UACGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGAAAU CCGAUAAAUAAGAAGCAUCAAAG
2121	延伸茎截短	UACUGGCGCCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAG CGACUAUGUCGUAUGGGUAAAGCGCUUACGGACUUCGGUCCG UAAGAAGCAUCAAAG
2122	C55	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUCGGUAAAGCGCUUAUUUAUCGGAGAGAAA UCCGAUAAAUAAGAAGCAUCAAAG
2123	trip mut	UACUGGCGCCUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGACUUCGG UCCGAUAAAUAAGAAGCAUCAAAG
2124	-76:78	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGAGAAAUCC GAUAAAUAAGAAGCAUCAAAG
2125	-1:5	GCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCGACUA UGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGAAAUCCGA UAAAUAAGAAGCAUCAAAG
2126	-83:87	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGAGA UAAAUAAGAAGCAUCAAAG
2127	=+G28, A82U, -84,	UACUGGCGCUUUUAUCUCAUUACUUUGGAGAGCCAUCACCAG CGACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGUA UCCGAUAAAUAAGAAGCAUCAAAG
2128	=+51U	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGAA AUCCGAUAAAUAAGAAGCAUCAAAG

[0137]

SEQ ID NO:	名称或修饰	核苷酸序列
2129	-1:4, +G5A, +G86,	AGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCGACU AUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGAAAUGCC GAUAAAUAAGAAGCAUCAAAG
2130	=+A94	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGAAA UCCGAUAAAAUAAGAAGCAUCAAAG
2131	=+G72	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUAUUGUAUCGGAGAGAA AUCCGAUAAAUAAGAAGCAUCAAAG
2132	缩短前沿, CUUCGG 环经修饰. 延伸经延伸	GCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCGACUA UGUCGUAUGGGUAAAGCGCUUAUUUAUCGGACUUCGGUCCGA UAAAUAAGCGCAUCAAAG
2133	A14C	UACUGGCGCUUUUCUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGAAA UCCGAUAAAUAAGAAGCAUCAAAG
2134	-1:3, +G3	GUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCGA CUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGAAAUC CGAUAAAUAAGAAGCAUCAAAG
2135	=+C45, +U46	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACCUUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGA AAUCCGAUAAAUAAGAAGCAUCAAAG
2136	CUUCGG 环经修饰, 功能启动（fun start）	GAUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGACUUCGGU CCGAUAAAUAAGAAGCAUCAAAG
2137	-93:94	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGAAA UCCGAUAAAAGAAGCAUCAAAG
2138	=+U45	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GAUCUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGAA AUCCGAUAAAUAAGAAGCAUCAAAG
2139	-69, -94	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGGCUUAUUUAUCGGAGAGAAAU CCGAUAAAAAGAAGCAUCAAAG
2140	-94	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGAAA UCCGAUAAAAAGAAGCAUCAAAG
2141	修饰后的 CUUCGG，在第一个三螺旋体中减去 U	UACUGGCGCUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGACUUCGGU CCGAUAAAUAAGAAGCAUCAAAG
2142	-1:4, +C4, A14C, U17G, +G72, -76:78, -83:87	CGGCGCUUUUCUCGCAUUACUUUGAGAGCCAUCACCAGCGAC UAUGUCGUAUGGGUAAAGCGCUUAUUGUAUCGAGAGAUAAAU AAGAAGCAUCAAAG
2143	U1C, -73	CACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUAUUUUCGGAGAGAAAU CCGAUAAAUAAGAAGCAUCAAAG

[0138]

SEQ ID NO:	名称或修饰	核苷酸序列
2144	支架 uuCG, 茎 uuCG. 茎交换, t 缩短	UACUGGCGCUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCG ACUUCGGUCGUAUGGGUAAAGCGCUUAUGUAUCGGCUUCGGC CGAUACAUAAGAAGCAUCAAAG
2145	支架 uuCG, 茎 uuCG. 茎交换.	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUUCGGUCGUAUGGGUAAAGCGCUUAUGUAUCGGCUUCGG CCGAUACAUAAGAAGCAUCAAAG
2146	=+G60	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUGAAAGCGCUUAUUUAUCGGAGAGAA AUCCGAUAAAUAAGAAGCAUCAAAG
2147	无茎支架 uuCG	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUUCGGUCGUAUGGGUAAAG
2148	无茎支架 uuCG, 功能启动	GAUGGGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCGA CUUCGGUCGUAUGGGUAAAG
2149	支架 uuCG, 茎 uuCG, 功能启动	GAUGGGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCGA CUUCGGUCGUAUGGGUAAAGCGCUUAUUUAUCGGCUUCGGCC GAUAAAUAAGAAGCAUCAAAG
2150	假结	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUACACUGGGAUCGCUGAA UUAGAGAUCGGCGUCCUUUCAUUCUAUAUACUUUGGAGUUUU AAAAUGUCUCUAAGUACAGAAGCAUCAAAG
2151	支架 uuCG, 茎 uuCG	GGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCGACU UCGGUCGUAUGGGUAAAGCGCUUAUUUAUCGGCUUCGGCCGA UAAAUAAGAAGCAUCAAAG
2152	支架 uuCG, 茎 uuCG, 无启动	GCUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCG ACUUCGGUCGUAUGGGUAAAGCGCUUAUUUAUCGGCUUCGGC CGAUAAAUAAGAAGCAUCAAAG
2153	支架 uuCG	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUUCGGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGAA AUCCGAUAAAUAAGAAGCAUCAAAG
2154	=+GCUC36	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUGCUCCAC CAGCGACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGA GAAAUCCGAUAAAUAAGAAGCAUCAAAG
2155	G 四螺旋体端粒篮+末端	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGGGGUUAGGGUUAGGGUUAG GGAAGCAUCAAAG
2156	G 四螺旋体 M3q	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGGAGGGAGGGAGGGAGAGGG AAAGCAUCAAAG
2157	G 四螺旋体端粒篮, 无末端	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGUUGGGUUAGGGUUAGGGUU AGGGAAAAGCAUCAAAG
2158	45,44 发夹（老版）	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGC--------AGGGCUUCGG CCG---------GAAGCAUCAAAG
2159	帚曲菌素-蓖麻毒素环	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCCUGCUCAGUACGAGAGGA ACCGCAGGAAGCAUCAAAG

[0139]

SEQ ID NO:	名称或修饰	核苷酸序列
2160	uvsX, C18G	UACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCCCUCUUCGGAGGGAAGCA UCAAAG
2161	截短茎环，C18G，trip mut（U10C）	UACUGGCGCCUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUACGGACUUCGGUCCGU AAGAAGCAUCAAAG
2162	短噬菌体 rep, C18G	UACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCGGACGACCUCUCGGUCGU CCGAAGCAUCAAAG
2163	噬菌体 rep 环, C18G	UACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCAGGUGGGACGACCUCUCG GUCGUCCUAUCUGAAGCAUCAAAG
2164	=+G18, 堆叠至 64 上	UACUGGCGCCUUUAUCUGCAUUACUUUGAGAGCCAUCACCAG CGACUAUGUCGUAUGGGUAAAGCGCUUACGGACUUCGGUCCG UAAGAAGCAUCAAAG
2165	截短茎环, C18G, -1 A2G	GCUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCGCUUACGGACUUCGGUCCGUA AGAAGCAUCAAAG
2166	噬菌体 rep 环，C18G，trip mut（U10C）	UACUGGCGCCUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCAGGUGGGACGACCUCUCG GUCGUCCUAUCUGAAGCAUCAAAG
2167	短噬菌体 rep，C18G，trip mut（U10C）	UACUGGCGCCUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCGGACGACCUCUCGGUCGU CCGAAGCAUCAAAG
2168	uvsX, trip mut (U10C)	UACUGGCGCCUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCCCUCUUCGGAGGGAAGCA UCAAAG
2169	截短茎环	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUACGGACUUCGGUCCGU AAGAAGCAUCAAAG
2170	=+A17, 堆叠至 64 上	UACUGGCGCCUUUAUCAUCAUUACUUUGAGAGCCAUCACCAG CGACUAUGUCGUAUGGGUAAAGCGCUUACGGACUUCGGUCCG UAAGAAGCAUCAAAG
2171	3' HDV 基因组核酶	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGAAA UCCGAUAAAUAAGAAGCAUCAAAGGGCCGGCAUGGUCCCAGC CUCCUCGCUGGCGCCGGCUGGGCAACAUUCCGAGGGGACCGU CCCCUCGGUAAUGGCGAAUGGGACCC
2172	噬菌体 rep 环，trip mut（U10C）	UACUGGCGCCUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCAGGUGGGACGACCUCUCG GUCGUCCUAUCUGAAGCAUCAAAG
2173	-79:80	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAAAUC CGAUAAAUAAGAAGCAUCAAAG
2174	短噬菌体 rep，trip mut（U10C）	UACUGGCGCCUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCGGACGACCUCUCGGUCGU CCGAAGCAUCAAAG

[0140]

SEQ ID NO:	名称或修饰	核苷酸序列
2175	额外截短茎环	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCCGGACUUCGGUCCGGAAG CAUCAAAG
2176	U17G, C18G	UACUGGCGCUUUUAUCGGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGAAA UCCGAUAAAUAAGAAGCAUCAAAG
2177	短噬菌体 rep	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCGGACGACCUCUCGGUCGU CCGAAGCAUCAAAG
2178	uvsX, C18G, -1 A2G	GCUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCGCCCUCUUCGGAGGGAAGCAU CAAAG
2179	uvsX, C18G, trip mut (U10C), -1 A2G, HDV -99 G65U	GCUGGCGCCUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCUCCCCUCUUCGGAGGGAGCAUC AAAG
2180	3' HDV 反基因组核酶	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGAAA UCCGAUAAAUAAGAAGCAUCAAAGGGGUCGGCAUGGCAUCUC CACCUCCUCGCGGUCCGACCUGGGCAUCCGAAGGAGGACGCA CGUCCACUCGGAUGGCUAAGGGAGAGCCA
2181	uvsX, C18G, trip mut (U10C), -1 A2G, HDV AA(98:99)C	GCUGGCGCCUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCGCCCUCUUCGGAGGGCGCAUC AAAG
2182	3' HDV 核酶（Lior Nissim, Timothy Lu）	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGAAA UCCGAUAAAUAAGAAGCAUCAAAGUUUUGGCCGGCAUGGUCC CAGCCUCCUCGCUGGCGCCGGCUGGGCAACAUGCUUCGGCAU GGCGAAUGGGACCCCGGG
2183	TAC（1:3）GA，堆叠至 64 上	GAUGGCGCCUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCGCUUACGGACUUCGGUCCGUA AGAAGCAUCAAAG
2184	uvsX, -1 A2G	GCUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCGCCCUCUUCGGAGGGAAGCAU CAAAG
2185	截短茎环, C18G, trip mut（U10C）,-1 A2G, HDV -99 G65U	GCUGGCGCCUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCUCUUACGGACUUCGGUCCGUA AGAGCAUCAAAG
2186	短噬菌体 rep, C18G, trip mut（U10C）, -1 A2G, HDV -99 G65U	GCUGGCGCCUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCUCGGACGACCUCUCGGUCGUC CGAGCAUCAAAG
2187	3' sTRSV WT 病毒锤头核酶	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGAAA UCCGAUAAAUAAGAAGCAUCAAAGCCUGUCACCGGAUGUGCU UUCCGGUCUGAUGAGUCCGUGAGGACGAAACAGG

[0141]

SEQ ID NO:	名称或修饰	核苷酸序列
2188	短噬菌体 rep, C18G, -1 A2G	GCUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCGCGGACGACCUCUCGGUCGUC CGAAGCAUCAAAG
2189	短噬菌体 rep, C18G, trip mut（U10C）, -1 A2G, 3'基因组 HDV	GCUGGCGCCUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCGCGGACGACCUCUCGGUCGUC CGAAGCAUCAAAG
2190	噬菌体 rep 环, C18G, trip mut（U10C）, -1 A2G, HDV -99 G65U	GCUGGCGCCUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCUCAGGUGGGACGACCUCUCGG UCGUCCUAUCUGAGCAUCAAAG
2191	3' HDV 核酶（Owen Ryan, Jamie Cate）	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGAAA UCCGAUAAAUAAGAAGCAUCAAAGGAUGGCCGGCAUGGUCCC AGCCUCCUCGCUGGCGCCGGCUGGGCAACACCUUCGGGUGGC GAAUGGGAC
2192	噬菌体 rep 环, C18G, -1 A2G	GCUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCGCAGGUGGGACGACCUCUCGG UCGUCCUAUCUGAAGCAUCAAAG
2193	0.14	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUACUGGGUAAAGCGCUUAUUUAUCGGAGAGAA AUCCGAUAAAUAAGAAGCAUCAAAG
2194	-78, G77U	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGUGAGAAAU CCGAUAAAUAAGAAGCAUCAAAG
2195		GUACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAG CGACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGAA AUCCGAUAAAUAAGAAGCAUCAAAG
2196	短噬菌体 rep, -1 A2G	GCUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCGCGGACGACCUCUCGGUCGUC CGAAGCAUCAAAG
2197	截短茎环, C18G, trip mut（U10C）, -1 A2G	GCUGGCGCCUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCGCUUACGGACUUCGGUCCGUA AGAAGCAUCAAAG
2198	-1, A2G	GCUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGAAAU CCGAUAAAUAAGAAGCAUCAAAG
2199	截短茎环, trip mut（U10C）, -1 A2G	GCUGGCGCCUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCGCUUACGGACUUCGGUCCGUA AGAAGCAUCAAAG
2200	uvsX, C18G, trip mut (U10C), -1 A2G	GCUGGCGCCUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCGCCCUCUUCGGAGGGAAGCAU CAAAG
2201	噬菌体 rep 环, -1 A2G	GCUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCGCAGGUGGGACGACCUCUCGG UCGUCCUAUCUGAAGCAUCAAAG
2202	噬菌体 rep 环, trip mut（U10C）, -1 A2G	GCUGGCGCCUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCGCAGGUGGGACGACCUCUCGG UCGUCCUAUCUGAAGCAUCAAAG

[0142]

SEQ ID NO:	名称或修饰	核苷酸序列
2203	噬菌体 rep 环, C18G, trip mut（U10C）, -1 A2G	GCUGGCGCCUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCGCAGGUGGGACGACCUCUCGG UCGUCCUAUCUGAAGCAUCAAAG
2204	截短茎环, C18G	UACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUACGGACUUCGGUCCGU AAGAAGCAUCAAAG
2205	uvsX, trip mut (U10C), -1 A2G	GCUGGCGCCUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCGCCCUCUUCGGAGGGAAGCAU CAAAG
2206	截短茎环, -1 A2G	GCUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCGCUUACGGACUUCGGUCCGUA AGAAGCAUCAAAG
2207	短噬菌体 rep, trip mut（U10C）, -1 A2G	GCUGGCGCCUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCGCGGACGACCUCUCGGUCGUC CGAAGCAUCAAAG
2208	5'HDV 核酶（Owen Ryan, Jamie Cate）	GAUGGCCGGCAUGGUCCCAGCCUCCUCGCUGGCGCCGGCUGG GCAACACCUUCGGGUGGCGAAUGGGACUACUGGCGCUUUUAU CUCAUUACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAUGG GUAAAGCGCUUAUUUAUCGGAGAGAAAUCCGAUAAAUAAGAA GCAUCAAAG
2209	5'HDV 基因组核酶	GGCCGGCAUGGUCCCAGCCUCCUCGCUGGCGCCGGCUGGGCA ACAUUCCGAGGGGACCGUCCCCUCGGUAAUGGCGAAUGGGAC CCUACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCA GCGACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGA AAUCCGAUAAAUAAGAAGCAUCAAAG
2210	截短茎环, C18G, trip mut（U10C）, -1 A2G, HDV AA（98:99）C	GCUGGCGCCUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCGCUUACGGACUUCGGUCCGUA AGCGCAUCAAAG
2211	5'env25 手枪核酶（具有添加的 CUUCGG 环）	CGUGGUUAGGGCCACGUUAAAUAGUUGCUUAAGCCCUAAGCG UUGAUCUUCGGAUCAGGUGCAAUACUGGCGCUUUUAUCUCAU UACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAUGGGUAAA GCGCUUAUUUAUCGGAGAGAAAUCCGAUAAAUAAGAAGCAUC AAAG
2212	5'HDV 反基因组核酶	GGGUCGGCAUGGCAUCUCCACCUCCUCGCGGUCCGACCUGGG CAUCCGAAGGAGGACGCACGUCCACUCGGAUGGCUAAGGGAG AGCCAUACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCA CCAGCGACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAG AGAAAUCCGAUAAAUAAGAAGCAUCAAAG
2213	3'锤头核酶（Lior Nissim, Timothy Lu）引导支架疤痕	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGAAA UCCGAUAAAUAAGAAGCAUCAAAGCCAGUACUGAUGAGUCCG UGAGGACGAAACGAGUAAGCUCGUCUACUGGCGCUUUUAUCU CAU
2214	=+A27, 堆叠至 64 上	UACUGGCGCCUUUAUCUCAUUACUUUAGAGAGCCAUCACCAG CGACUAUGUCGUAUGGGUAAAGCGCUUACGGACUUCGGUCCG UAAGAAGCAUCAAAG

[0143]

SEQ ID NO:	名称或修饰	核苷酸序列
2215	5'锤头核酶（Lior Nissim, Timothy Lu）较小疤痕	CGACUACUGAUGAGUCCGUGAGGACGAAACGAGUAAGCUCGU CUAGUCGUACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAU CACCAGCGACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGG AGAGAAAUCCGAUAAAUAAGAAGCAUCAAAG
2216	噬菌体 rep 环, C18G, trip mut（U10C）, -1 A2G, HDV AA（98:99）C	GCUGGCGCCUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCGCAGGUGGGACGACCUCUCGG UCGUCCUAUCUGCGCAUCAAAG
2217	-27, 堆叠至 64 上	UACUGGCGCCUUUAUCUCAUUACUUUAGAGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCGCUUACGGACUUCGGUCCGUA AGAAGCAUCAAAG
2218	3'手斧	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGAAA UCCGAUAAAUAAGAAGCAUCAAAGCAUUCCUCAGAAAAUGAC AAACCUGUGGGGCGUAAGUAGAUCUUCGGAUCUAUGAUCGUG CAGACGUUAAAAUCAGGU
2219	3'锤头核酶（Lior Nissim, Timothy Lu）	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGAAA UCCGAUAAAUAAGAAGCAUCAAAGCGACUACUGAUGAGUCCG UGAGGACGAAACGAGUAAGCUCGUCUAGUCGCGUGUAGCGAA GCA
2220	5'手斧	CAUUCCUCAGAAAAUGACAAACCUGUGGGGCGUAAGUAGAUC UUCGGAUCUAUGAUCGUGCAGACGUUAAAAUCAGGUUACUGG CGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCGACUAU GUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGAAAUCCGAU AAAUAAGAAGCAUCAAAG
2221	5' HDV 核酶（Lior Nissim, Timothy Lu）	UUUUGGCCGGCAUGGUCCCAGCCUCCUCGCUGGCGCCGGCUG GGCAACAUGCUUCGGCAUGGCGAAUGGGACCCCGGGUACUGG CGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCGACUAU GUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGAAAUCCGAU AAAUAAGAAGCAUCAAAG
2222	5'锤头核酶（Lior Nissim, Timothy Lu）	CGACUACUGAUGAGUCCGUGAGGACGAAACGAGUAAGCUCGU CUAGUCGCGUGUAGCGAAGCAUACUGGCGCUUUUAUCUCAUU ACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAUGGGUAAAG CGCUUAUUUAUCGGAGAGAAAUCCGAUAAAUAAGAAGCAUCA AAG
2223	3' HH15 最小锤头核酶	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGAAA UCCGAUAAAUAAGAAGCAUCAAAGGGGAGCCCCGCUGAUGAG GUCGGGGAGACCGAAAGGGACUUCGGUCCCUACGGGGCUCCC
2224	5' RBMX 募集基序	CCACCCCCACCACCACCCCCACCCCCACCACCACCCUACUGG CGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCGACUAU GUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGAAAUCCGAU AAAUAAGAAGCAUCAAAG

[0144]

SEQ ID NO:	名称或修饰	核苷酸序列
2225	3'锤头核酶（Lior Nissim, Timothy Lu）较小疤痕	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGAAA UCCGAUAAAUAAGAAGCAUCAAAGCGACUACUGAUGAGUCCG UGAGGACGAAACGAGUAAGCUCGUCUAGUCG
2226	3' env25 手枪核酶(具有添加的 CUUCGG 环）	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGAAA UCCGAUAAAUAAGAAGCAUCAAAGCGUGGUUAGGGCCACGUU AAAUAGUUGCUUAAGCCCUAAGCGUUGAUCUUCGGAUCAGGU GCAA
2227	3' Env-9 绞扭器	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGAAA UCCGAUAAAUAAGAAGCAUCAAAGGGCAAUAAAGCGGUUACA AGCCCGCAAAAAUAGCAGAGUAAUGUCGCGAUAGCGCGGCAU UAAUGCAGCUUUAUUG
2228	=+AUUAUCUCAUU ACU25	UACUGGCGCUUUUAUCUCAUUACUAUUAUCUCAUUACUUUGA GAGCCAUCACCAGCGACUAUGUCGUAUGGGUAAAGCGCUUAU UUAUCGGAGAGAAAUCCGAUAAAUAAGAAGCAUCAAAG
2229	5' Env-9 绞扭器	GGCAAUAAAGCGGUUACAAGCCCGCAAAAAUAGCAGAGUAAU GUCGCGAUAGCGCGGCAUUAAUGCAGCUUUAUUGUACUGGCG CUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCGACUAUGU CGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGAAAUCCGAUAA AUAAGAAGCAUCAAAG
2230	3'扭曲姊妹 1	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGAAA UCCGAUAAAUAAGAAGCAUCAAAGACCCGCAAGGCCGACGGC AUCCGCCGCCGCUGGUGCAAGUCCAGCCGCCCCUUCGGGGGC GGGCGCUCAUGGGUAAC
2231	无茎	UACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAUGGGUAAAG
2232	5' HH15 最小锤头核酶	GGGAGCCCCGCUGAUGAGGUCGGGGAGACCGAAAGGGACUUC GGUCCCUACGGGGCUCCCUACUGGCGCUUUUAUCUCAUUACU UUGAGAGCCAUCACCAGCGACUAUGUCGUAUGGGUAAAGCGC UUAUUUAUCGGAGAGAAAUCCGAUAAAUAAGAAGCAUCAAAG
2233	5'锤头核酶（Lior Nissim, Timothy Lu）引导支架疤痕	CCAGUACUGAUGAGUCCGUGAGGACGAAACGAGUAAGCUCGU CUACUGGCGCUUUUAUCUCAUUACUGGCGCUUUUAUCUCAUU ACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAUGGGUAAAG CGCUUAUUUAUCGGAGAGAAAUCCGAUAAAUAAGAAGCAUCA AAG
2234	5'扭曲姊妹 1	ACCCGCAAGGCCGACGGCAUCCGCCGCCGCUGGUGCAAGUCC AGCCGCCCCUUCGGGGGCGGGCGCUCAUGGGUAACUACUGGC GCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCGACUAUG UCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGAAAUCCGAUA AAUAAGAAGCAUCAAAG
2235	5' sTRSV WT 病毒锤头核酶	CCUGUCACCGGAUGUGCUUUCCGGUCUGAUGAGUCCGUGAGG ACGAAACAGGUACUGGCGCUUUUAUCUCAUUACUUUGAGAGC CAUCACCAGCGACUAUGUCGUAUGGGUAAAGCGCUUAUUUAU CGGAGAGAAAUCCGAUAAAUAAGAAGCAUCAAAG

[0145]

SEQ ID NO:	名称或修饰	核苷酸序列
2236	148: =+G55, 堆叠至64上	GUACUGGCGCCUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAGUGGGUAAAGCGCUUACGGACUUCGGUCCGUAAGAAGCAUCAAAG
2237	158: 103+148(+G55)-99, G65U	GUACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAGUGGGUAAAGCUCCCCUCUUCGGAGGGAGCAUCAAAG
2238	174: Uvsx 延伸茎伴以[A99] G65U），C18G,^G55, [GU-1]	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAGUGGGUAAAGCUCCCCUCUUCGGAGGGAGCAUCAAAG
2239	175: 延伸茎突变, U10C, [GU-1]	ACUGGCGCCUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAUGGGUAAAGCGCUUACGGACUUCGGUCCGUAAGAAGCAUCAAAG
2240	176: 174，用A1G代替T7转录	GCUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAGUGGGUAAAGCUCCCCUCUUCGGAGGGAGCAUCAAAG
2241	177: 174，去除气泡（+G55）	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAUGGGUAAAGCUCCCCUCUUCGGAGGGAGCAUCAAAG
2242	181: 茎42（截短茎环）；U10C,C18G,[GU-1] (95+[GU-1])	ACUGGCGCCUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAUGGGUAAAGCGCUUACGGACUUCGGUCCGUAAGAAGCAUCAAAG
2243	182: 茎42（截短茎环）; C18G,[GU-1]	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAUGGGUAAAGCGCUUACGGACUUCGGUCCGUAAGAAGCAUCAAAG
2244	183: 茎42（截短茎环）；C18G,^G55,[GU-1]	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAGUGGGUAAAGCGCUUACGGACUUCGGUCCGUAAGAAGCAUCAAAG
2245	184: 茎48（uvsx, -99 g65t）；C18G,^T55,[GU-1]	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAUUGGGUAAAGCUCCCCUCUUCGGAGGGAGCAUCAAAG
2246	185: 茎42（截短茎环）；C18G,^U55,[GU-1]	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAUUGGGUAAAGCGCUUACGGACUUCGGUCCGUAAGAAGCAUCAAAG
2247	186: 茎42（截短茎环）;U10C,^A17,[GU-1]	ACUGGCGCCUUUAUCAUCAUUACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAUGGGUAAAGCGCUUACGGACUUCGGUCCGUAAGAAGCAUCAAAG
2248	187: 茎46（uvsx）；C18G,^G55,[GU-1]	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAGUGGGUAAAGCGCCCUCUUCGGAGGGAAGCAUCAAAG
2249	188: 茎50 (ms2 U15C, -99, g65t); C18G,^G55,[GU-1]	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAGUGGGUAAAGCUCACAUGAGGAUCACCCAUGUGAGCAUCAAAG

[0146]

SEQ ID NO:	名称或修饰	核苷酸序列
2250	189: 174 + G8A;U15C;U35A	ACUGGCACUUUUACCUGAUUACUUUGAGAGCCAACACCAGCG ACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAGCAU CAAAG
2251	190: 174 + G8A	ACUGGCACUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAGCAU CAAAG
2252	191: 174 + G8C	ACUGGCCCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAGCAU CAAAG
2253	192: 174 + U15C	ACUGGCGCUUUUACCUGAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAGCAU CAAAG
2254	193, 174 + U35A	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAACACCAGCG ACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAGCAU CAAAG
2255	195: 175 + C18G + G8A;U15C;U35A	ACUGGCACCUUUACCUGAUUACUUUGAGAGCCAACACCAGCG ACUAUGUCGUAUGGGUAAAGCGCUUACGGACUUCGGUCCGUA AGAAGCAUCAAAG
2256	196: 175 + C18G + G8A	ACUGGCACCUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCGCUUACGGACUUCGGUCCGUA AGAAGCAUCAAAG
2257	197: 175 + C18G + G8C	ACUGGCCCCUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCGCUUACGGACUUCGGUCCGUA AGAAGCAUCAAAG
2258	198: 175 + C18G + U35A	ACUGGCGCCUUUAUCUGAUUACUUUGAGAGCCAACACCAGCG ACUAUGUCGUAUGGGUAAAGCGCUUACGGACUUCGGUCCGUA AGAAGCAUCAAAG
2259	199: 174 + A2G (在开始时测试 G 的转录; ccGCT...)	GCUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAGCAU CAAAG
2260	200: 174 + ^G1 (ccGACU...)	GACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAGCA UCAAAG
2261	201: 174 + U10C;^G28	ACUGGCGCCUUUAUCUGAUUACUUUGGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAGCA UCAAAG
2262	202: 174 + U10A;A28U	ACUGGCGCAUUUAUCUGAUUACUUUGUGAGCCAUCACCAGCG ACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAGCAU CAAAG
2263	203: 174 + U10C	ACUGGCGCCUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAGCAU CAAAG
2264	204: 174 + ^G28	ACUGGCGCUUUUAUCUGAUUACUUUGGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAGCA UCAAAG

[0147]

SEQ ID NO:	名称或修饰	核苷酸序列
2265	205: 174 + U10A	ACUGGCGCAUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAGCAU CAAAG
2266	206, 174 + A28U	ACUGGCGCUUUUAUCUGAUUACUUUGUGAGCCAUCACCAGCG ACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAGCAU CAAAG
2267	207: 174 + ^U15	ACUGGCGCUUUUAUUCUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAGCA UCAAAG
2268	208: 174 + [U4]	ACGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCGA CUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAGCAUC AAAG
2269	209: 174 + C16A	ACUGGCGCUUUUAUAUGAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAGCAU CAAAG
2270	210: 174 + ^U17	ACUGGCGCUUUUAUCUUGAUUACUUUGAGAGCCAUCACCAGC GACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAGCA UCAAAG
2271	211: 174 + U35G（与上文的 174 + U35A 相比）	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAGCACCAGCG ACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAGCAU CAAAG
2272	212: 174 +U11G, A105G (A86G), U26C	ACUGGCGCUGUUAUCUGAUUACUUCGAGAGCCAUCACCAGCG ACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAGCAU CGAAG
2273	213: 174 +U11C, A105G (A86G), U26C	ACUGGCGCUCUUAUCUGAUUACUUCGAGAGCCAUCACCAGCG ACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAGCAU CGAAG
2274	214: 174+U12G; A106G (A87G), U25C	ACUGGCGCUUGUAUCUGAUUACUCUGAGAGCCAUCACCAGCG ACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAGCAU CAGAG
2275	215: 174+U12C; A106G (A87G), U25C	ACUGGCGCUUCUAUCUGAUUACUCUGAGAGCCAUCACCAGCG ACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAGCAU CAGAG
2276	216: 174_tx_11.G,87.G,22.C	ACUGGCGCUUUGAUCUGAUUACCUUGAGAGCCAUCACCAGCG ACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAGCAU CAAGG
2277	217: 174_tx_11.C,87.G,22.C	ACUGGCGCUUUCAUCUGAUUACCUUGAGAGCCAUCACCAGCG ACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAGCAU CAAGG
2278	218: 174 +U11G	ACUGGCGCUGUUAUCUGAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAGCAU CAAAG
2279	219: 174 +A105G (A86G)	ACUGGCGCUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAGCAU CGAAG

[0148]

SEQ ID NO:	名称或修饰	核苷酸序列
2280	220: 174 +U26C	ACUGGCGCUUUUAUCUGAUUACUUCGAGAGCCAUCACCAGCG ACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAGCAU CAAAG
2281	221: 182: + G7A, T11C, T24C, A95G + ^C63, A88G	ACUGGCACUUCUAUCUGAUUACUCUGAGAGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCCGCUUACGGACUUCGGUCCGU AAGAGGCAUCAGAG
2282	222: 174: + G7A, T11C, T24C, A87G	ACUGGCACUUCUAUCUGAUUACUCUGAGAGCCAUCACCAGCG ACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAGCAU CAGAG
2283	223: 175: + G7A, C17G + ^C63, A88G	ACUGGCACCUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCCGCUUACGGACUUCGGUCCGU AAGAGGCAUCAAAG
2284	224: 182: + G7A, T11G, T24C, A95G + ^C63, A88G	ACUGGCACUUGUAUCUGAUUACUCUGAGAGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCCGCUUACGGACUUCGGUCCGU AAGAGGCAUCAGAG
2285	225: 174: + G7A, T11G, T24C, A87G	ACUGGCACUUGUAUCUGAUUACUCUGAGAGCCAUCACCAGCG ACUAUGUCGUAGUGGGUAAAGCUCCCUCUUCGGAGGGAGCAU CAGAG
2286	229: 196 + C9T	ACUGGCACUUUUAUCUGAUUACUUUGAGAGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCGCUUACGGACUUCGGUCCGUA AGAAGCAUCAAAG
2287	230: 221 + [C64],G89A	ACUGGCACUUCUAUCUGAUUACUCUGAGAGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCGCUUACGGACUUCGGUCCGUA AGAAGCAUCAGAG
2288	231: 221 + A7G	ACUGGCGCUUCUAUCUGAUUACUCUGAGAGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCCGCUUACGGACUUCGGUCCGU AAGAGGCAUCAGAG
2289	232: 221 + A29C	ACUGGCACUUCUAUCUGAUUACUCUGAGCGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCCGCUUACGGACUUCGGUCCGU AAGAGGCAUCAGAG
2290	233: 221 + A7G,A29C	ACUGGCGCUUCUAUCUGAUUACUCUGAGCGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCCGCUUACGGACUUCGGUCCGU AAGAGGCAUCAGAG
2291	234: 221 + A7G,A29C + CG64GC,AG88GA	ACUGGCGCUUCUAUCUGAUUACUCUGAGCGCCAUCACCAGCG ACUAUGUCGUAUGGGUAAAGCGCCUUACGGACUUCGGUCCGU AAGGAGCAUCAGAG
2292	235: 221 + A7G,A29C,^G54	ACUGGCGCUUCUAUCUGAUUACUCUGAGCGCCAUCACCAGCG ACUAUGUCGUAGUGGGUAAAGCCGCUUACGGACUUCGGUCCG UAAGAGGCAUCAGAG
2293	236: 221 +假结 GGGACT_AGTCCC	ACGGGACUUUCUAUCUGAUUACUCUGAAGUCCCUCACCAGCG ACUAUGUCGUAUGGGUAAAGCCGCUUACGGACUUCGGUCCGU AAGAGGCAUCAGAG
2294	237: 221 +假结 CTGTAG_CTACAG	ACCUGUAGUUCUAUCUGAUUACUCUGACUACAGUCACCAGCG ACUAUGUCGUAUGGGUAAAGCCGCUUACGGACUUCGGUCCGU AAGAGGCAUCAGAG

[0149] 在一些实施例中，gNA变异体包含tracrRNA茎环，其包含序列-UUU-N4-25-UUU-(SEQ ID NO:34)。举例来说，gNA变异体包含支架茎环或其替代物，经两个促进三螺旋区的三联体U基序侧接。在一些实施例中，支架茎环或其替代物包含至少4个核苷酸、至少5个核苷酸、至少6个核苷酸、至少7个核苷酸、至少7个核苷酸、至少8个核苷酸、至少9个核苷酸、至少10个核苷酸、至少11个核苷酸、至少12个核苷酸、至少13个核苷酸、至少14个核苷酸、至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸、至少22个核苷酸、至少23个核苷酸、至少24个核苷酸或至少25

个核苷酸。

[0150] 在一些实施例中，gNA变异体包含在间隔区的5'端的位置具有-AAAG-的crRNA序列。在一些实施例中，-AAAG-序列紧靠间隔区的5'端。

[0151] 在一些实施例中，对参考gNA的至少一个核苷酸修饰以产生gNA变异体包含CasX变异gNA中相对于参考gRNA的至少一个核苷酸缺失。在一些实施例中，gNA变异体包含相对于参考gNA缺失1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个连续或非连续核苷酸。在一些实施例中，至少一个缺失包含相对于参考gNA缺失1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个或更多个连续核苷酸。在一些实施例中，gNA变异体包含相对于参考gNA的2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个或更多个核苷酸缺失，且所述缺失不在连续核苷酸中。在其中gNA变异体中存在两个或更多个相对于参考gRNA的非连续缺失的那些实施例中，如本文所述的任何缺失长度及缺失长度的任何组合涵盖于本公开的范围内。在一些实施例中，gNA变异体包含参考gRNA的不同区中的至少两个缺失。在一些实施例中，gNA变异体包含参考gRNA的相同区中的至少两个缺失。举例来说，所述区可为gNA变异体的延伸茎环、支架茎环、支架茎气泡、三螺旋环、假结、三螺旋体或5'端。参考gRNA中任何核苷酸的缺失涵盖于本公开的范围内。

[0152] 在一些实施例中，参考gRNA的至少一个核苷酸修饰以产生gNA变异体包含至少一个核苷酸插入。在一些实施例中，gNA变异体包含相对于参考gRNA插入1、2、3、4、5、6、7、8、9或10个连续或非连续核苷酸。在一些实施例中，至少一个核苷酸插入包含相对于参考gRNA插入1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个或更多个连续核苷酸。在一些实施例中，gNA变异体包含相对于参考gRNA的2个或更多个插入，且所述插入不连续。在其中gNA变异体中存在两个或更多个相对于参考gRNA的非连续插入的那些实施例中，如本文所述的任何插入长度及插入长度的任何组合涵盖于本公开的范围内。举例来说，在一些实施例中，gNA变异体可包含一个核苷酸的第一插入，及两个核苷酸的第二插入，且所述两个插入不连续。在一些实施例中，gNA变异体包含参考gRNA的不同区中的至少两个插入。在一些实施例中，gNA变异体包含参考gRNA的相同区中的至少两个插入。举例来说，所述区可为gNA变异体的延伸茎环、支架茎环、支架茎气泡、三螺旋环、假结、三螺旋体或5'端。在参考gRNA中的任何位置插入任何A、G、C、U(或T，于对应DNA中)或其组合涵盖于本公开的范围内。

[0153] 在一些实施例中，参考gRNA的至少一个核苷酸修饰以产生gNA变异体包含至少一个核酸取代。在一些实施例中，相对于参考gRNA，gNA变异体包含1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个或更多个连续或非连续经取代核苷酸。在一些实施例中，相对于参考gRNA，gNA变异体包含1-4个核苷酸取代。在一些实施例中，至少一个取代包含相对于参考gRNA取代1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个或更多个连续核苷酸。在一些实施例中，gNA变异体包含相对于参考gRNA的2个或更多个取代，且所述取代不连续。在其中gNA变异体中存在两个或更多个相对于参考gRNA的非连续取代的那些实施例中，如本文所述的任何经取代核苷酸长度及经取代核苷酸长度的任何组合涵盖于本公开的范围内。举例来说，在一些实施例中，gNA变异体可包含一个核苷酸的第一取代，及两个核苷酸的第二取代，且所述两个取代不连续。在一些实施例中，gNA变异体包含参考gRNA的不同区中的至少两个取代。在一些实施例中，gNA变异体包含参考gRNA的相同区中的

至少两个取代。举例来说，所述区可为gNA变异体的三螺旋体、延伸茎环、支架茎环、支架茎气泡、三螺旋环、假结、三螺旋体或5'端。在参考gRNA中的任何位置取代任何A、G、C、U(或T，于对应DNA中)或其组合涵盖于本公开的范围内。

[0154] 本文所述的取代、插入及缺失中的任一个可经合并以产生本公开的gNA变异体。举例来说，gNA变异体可包含相对于参考gRNA的至少一个取代及至少一个缺失、相对于参考gRNA的至少一个取代及至少一个插入、相对于参考gRNA的至少一个插入及至少一个缺失或相对于参考gRNA的至少一个取代、一个插入及一个缺失。

[0155] 在一些实施例中，gNA变异体包含与SEQ ID NO:4-16中的任一个具有至少20％同一性、至少30％同一性、至少40％同一性、至少50％同一性、至少60％同一性、至少65％同一性、至少70％同一性、至少75％同一性、至少80％同一性、至少85％同一性、至少90％同一性、至少91％同一性、至少92％同一性、至少93％同一性、至少94％同一性、至少95％同一性、至少96％同一性、至少97％同一性、至少98％同一性或至少99％同一性的支架区。在一些实施例中，gNA变异体包含与SEQ ID NO:4-16中的任一个具有至少60％同源(或同一性)的支架区。

[0156] 在一些实施例中，gNA变异体包含与SEQ ID NO:14具有至少60％同一性、至少65％同一性、至少70％同一性、至少75％同一性、至少80％同一性、至少85％同一性、至少90％同一性或至少95％同一性的tracr茎环。在一些实施例中，gNA变异体包含与SEQ ID NO:14具有至少60％同源(或同一性)的tracr茎环。

[0157] 在一些实施例中，gNA变异体包含与SEQ ID NO:15具有至少60％同一性、至少65％同一性、至少70％同一性、至少75％同一性、至少80％同一性、至少85％同一性、至少90％同一性或至少95％同一性的延伸茎环。在一些实施例中，gNA变异体包含与SEQ ID NO:15具有至少60％同源(或同一性)的延伸茎环。

[0158] 在一些实施例中，gNA变异体包含外源延伸茎环，其中与参考gNA的此类差异描述如下。在一些实施例中，外源延伸茎环与本文公开的参考茎环区(例如，SEQ ID NO:15)几乎没有或没有同一性。在一些实施例中，外源茎环为至少10bp、至少20bp、至少30bp、至少40bp、至少50bp、至少60bp、至少70bp、至少80bp、至少90bp、至少100bp、至少200bp、至少300bp、至少400bp、至少500bp、至少600bp、至少700bp、至少800bp、至少900bp、至少1,000bp、至少2,000bp、至少3,000bp、至少4,000bp、至少5,000bp、至少6,000bp、至少7,000bp、至少8,000bp、至少9,000bp、至少10,000bp、至少12,000bp、至少15,000bp或至少20,000bp。在一些实施例中，gNA变异体含有包含至少10、至少100、至少500、至少1000或至少10,000个核苷酸的延伸茎环区。在一些实施例中，异源性茎环增加gNA的稳定性。在一些实施例中，异源性RNA茎环能够结合蛋白质、RNA结构、DNA序列或小分子。在一些实施例中，替代茎环的外源茎环区包含RNA茎环或发夹，其中所得gNA具有增加的稳定性，且取决于环的选择，可与某些细胞蛋白质或RNA相互作用。此类外源性延伸茎环可以包含例如热稳定RNA，如MS2(ACAUGAGGAUCACCCAUGU(SEQ ID NO:35))、Qβ(UGCAUGUCUAAGACAGCA(SEQ ID NO:36))、U1发夹II(AAUCCAUUGCACUCCGGAUU(SEQ ID NO:37))、Uvsx(CCUCUUCGGAGG(SEQ ID NO:38))、PP7(AGGAGUUUCUAUGGAAACCCU(SEQ ID NO:39))、噬菌体复制环(AGGUGGGACGACCUCUCGGUCGUCCUAUCU(SEQ ID NO:40))、吻合环_a(UGCUCGCUCCGUUCGAGCA(SEQ ID NO:41))、吻合环_b1(UGCUCGACGCGUCCUCGAGCA(SEQ ID NO:42))、吻合环_b2

(UGCUCGUUUGCGGCUACGAGCA(SEQ ID NO:43))、G四螺旋体M3q(AGGGAGGGAGGGAGAGG(SEQ ID NO:44))、G四螺旋体端粒篮(GGUUAGGGUUAGGGUUAGG(SEQ ID NO:45))、帚曲菌素-蓖麻毒素环(CUGCUCAGUACGAGAGGAACCGCAG(SEQ ID NO:46))或假结(UACACUGGGAUCGCUGAAUUAGAGAUCGGCGUCCUUUCAUUCUAUAUACUUUGGAGUUUUAAAAUGUCUCUAAGUACA(SEQ ID NO:47))。在一些实施例中,外源茎环包含长非编码RNA(lncRNA)。如本文所用,lncRNA是指长度长于大约200bp的非编码RNA。在一些实施例中,外源茎环的5'及3'端碱基配对,即相互作用以形成双链体RNA区域。在一些实施例中,外源茎环的5'及3'端碱基配对,且外源茎环的5'与3'端之间的一个或多个区域不碱基配对。在一些实施例中,至少一个核苷酸修饰包含:(a)在一个或多个区中取代gNA变异体的1至15个连续或非连续核苷酸;(b)在一个或多个区中缺失gNA变异体的1至10个连续或非连续核苷酸;(c)在一个或多个区中插入gNA变异体的1至10个连续或非连续核苷酸;(d)经来自具有近端5'及3'端的异源性RNA源的RNA茎环序列取代支架茎环或延伸茎环;或(a)-(d)的任何组合。

[0159] 在一些实施例中,gNA变异体包含CCAGCGACUAUGUCGUAGUGG(SEQ ID NO:32)的支架茎环序列。在一些实施例中,gNA变异体包含与其具有至少1、2、3、4或5个失配的CCAGCGACUAUGUCGUAGUGG(SEQ ID NO:32)的支架茎环序列。

[0160] 在一些实施例中,gNA变异体含有包含小于32个核苷酸、小于31个核苷酸、小于30个核苷酸、小于29个核苷酸、小于28个核苷酸、小于27个核苷酸、小于26个核苷酸、小于25个核苷酸、小于24个核苷酸、小于23个核苷酸、小于22个核苷酸、小于21个核苷酸或小于20个核苷酸的延伸茎环区。在一些实施例中,gNA变异体含有包含小于32个核苷酸的延伸茎环区。在一些实施例中,gNA变异体进一步包含热稳定茎环。

[0161] 在一些实施例中,sgRNA变异体包含以下序列:SEQ ID NO:2104、SEQ ID NO:2106、SEQ ID NO:2163、SEQ ID NO:2107、SEQ ID NO:2164、SEQ ID NO:2165、SEQ ID NO:2166、SEQ ID NO:2103、SEQ ID NO:2167、SEQ ID NO:2105、SEQ ID NO:2108、SEQ ID NO:2112、SEQ ID NO:2160、SEQ ID NO:2170、SEQ ID NO:2114、SEQ ID NO:2171、SEQ ID NO:2112、SEQ ID NO:2173、SEQ ID NO:2102、SEQ ID NO:2174、SEQ ID NO:2175、SEQ ID NO:2109、SEQ ID NO:2176、SEQ ID NO:2238、SEQ ID NO:2239、SEQ ID NO:2240、SEQ ID NO:2241、SEQ ID NO:2256、SEQ ID NO:2274、SEQ ID NO:2275、SEQ ID NO:2279或SEQ ID NO:2281。在一些实施例中,sgRNA变异体包含SEQ ID NO:2238、2246、2256、2274或2275的序列。

[0162] 在一些实施例中,gNA变异体包含SEQ ID NO:2236、2237、2238、2241、2244、2248、2249、2256或2259到2294中的任一个的序列,或与其具有至少约80%、至少约90%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%的同一性的序列。在一些实施例中,gNA变异体包含SEQ ID NO:2201到2294中任一个的序列的一个或多个额外变化。在一些实施例中,gNA变异体包含SEQ ID NO:2236、2237、2238、2241、2244、2248、2249或2259-2294中的任一个的序列。

[0163] 在一些实施例中,sgRNA变异体包含对以下序列的一个或多个额外改变:SEQ ID NO:2104、SEQ ID NO:2163、SEQ ID NO:2107、SEQ ID NO:2164、SEQ ID NO:2165、SEQ ID NO:2166、SEQ ID NO:2103、SEQ ID NO:2167、SEQ ID NO:2105、SEQ ID NO:2108、SEQ ID NO:2112、SEQ ID NO:2160、SEQ ID NO:2170、SEQ ID NO:2114、SEQ ID NO:2171、SEQ ID NO:2112、SEQ ID NO:2173、SEQ ID NO:2102、SEQ ID NO:2174、SEQ ID NO:2175、SEQ ID

NO:2109、SEQ ID NO:2176、SEQ ID NO:2238、SEQ ID NO:2239、SEQ ID NO:2240、SEQ ID NO:2241、SEQ ID NO:2243、SEQ ID NO:2256、SEQ ID NO:2274、SEQ ID NO:2275、SEQ ID NO:2279或SEQ ID NO:2281。

[0164] 在本公开的gNA变异体的一些实施例中,gNA变异体包含至少一个修饰,其中相比于SEQ ID NO:5的参考引导支架的至少一个修饰选自以下中的一个或多个:(a)三螺旋环中的C18G取代;(b)茎气泡中的G55插入;(c)U1缺失;(d)延伸茎环的修饰,其中(i)6nt环和13个环近侧碱基对经Uvsx发夹替换;且(ii)A99的缺失和G65U的取代产生经完全碱基配对的环远侧碱基。在此类实施例中,gNA变异体包含SEQ ID NO:2236、2237、2238、2241、2244、2248、2249、2256或2259-2294中的任一个的序列。

[0165] 在gNA变异体的实施例中,gNA变异体进一步包含位于gNA的3'端处的间隔区(或靶向序列),其对C9orf72序列具有特异性。示范性间隔区和其同源PAM序列展示于下表3中。

[0166] 表3.C9orf72基因的gNA靶向序列

[0167]

PAM序列	SEQ ID NO
ATC	363-2100,2295-5426
TTC	5427-12893
GTC	12894-16202
CTC	16203-21835

[0168] 在gNA变异体的实施例中,gNA变异体进一步包含更充分描述于上文中的位于gNA的3'端的间隔区(或靶向序列),其包含至少14至约35个核苷酸,其中间隔区被设计成具有与靶核酸互补的序列。在一些实施例中,gNA变异体包含与靶核酸互补的至少10至30个核苷酸的靶向序列。在一些实施例中,靶向序列具有14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34或35个核苷酸。在一些实施例中,gNA变异体包含具有20个核苷酸的靶向序列。在一些实施例中,靶向序列具有25个核苷酸。在一些实施例中,靶向序列具有24个核苷酸。在一些实施例中,靶向序列具有23个核苷酸。在一些实施例中,靶向序列具有22个核苷酸。在一些实施例中,靶向序列具有21个核苷酸。在一些实施例中,靶向序列具有19个核苷酸。在一些实施例中,靶向序列具有18个核苷酸。在一些实施例中,靶向序列具有17个核苷酸。在一些实施例中,靶向序列具有16个核苷酸。在一些实施例中,靶向序列具有15个核苷酸。在一些实施例中,靶向序列具有14个核苷酸。在一些实施例中,本公开提供包括在本公开的gNA变异体中的靶向序列,其包含与SEQ ID NO:309-343、363-2100和2295-21835的序列具有至少50%同一性、至少55%同一性、至少60%同一性、至少65%同一性、至少70%同一性、至少75%同一性、至少80%同一性、至少85%同一性、至少90%同一性、至少95%同一性或100%同一性的序列。在一些实施例中,gNA变异体的靶向序列包含从序列的3'端去除单个核苷酸的SEQ ID NO:309-343、363-2100和2295-21835的序列。在其它实施例中,gNA变异体的靶向序列包含从序列的3'端去除两个核苷酸的SEQ ID NO:309-343、363-2100和2295-21835的序列。在其它实施例中,gNA变异体的靶向序列包含从序列的3'端去除三个核苷酸的SEQ ID NO:309-343、363-2100和2295-21835的序列。在其它实施例中,gNA变异体的靶向序列包含从序列的3'端去除四个核苷酸的SEQ ID NO:309-343、363-2100和2295-21835的序列。在其它实施例中,gNA变异体的靶向序列包含从序列的3'端去除五个核苷酸的SEQ ID NO:309-343、363-2100和2295-21835的序列。

[0169] 在一些实施例中,gNA变异体进一步包含位于gNA 3'端的间隔区(靶向)区,其中间隔区被设计成具有与靶核酸互补的序列。在一些实施例中,靶核酸包含位于间隔区5'的PAM序列,其中至少单个核苷酸将PAM与间隔区的第一核苷酸分开。在一些实施例中,PAM位于靶区的非靶向链上,即与靶核酸互补的链。在一些实施例中,PAM序列为ATC。在一些实施例中,ATC PAM的靶向序列包含SEQ ID NO:363-2100或2295-5426,或与SEQ ID NO:363-2100或2295-5426具有至少50%同一性、至少55%同一性、至少60%同一性、至少65%同一性、至少70%同一性、至少75%同一性、至少80%同一性、至少85%同一性、至少90%同一性、至少95%同一性或至少99%同一性的序列。在一些实施例中,ATC PAM的靶向序列选自由SEQ ID NO:363-2100或2295-5426组成的组。在一些实施例中,PAM序列为CTC。在一些实施例中,CTC PAM的靶向序列包含SEQ ID NO:16203-21835,或与SEQ ID NO:16203-21835具有至少50%同一性、至少55%同一性、至少60%同一性、至少65%同一性、至少70%同一性、至少75%同一性、至少80%同一性、至少85%同一性、至少90%同一性、至少95%同一性或至少99%同一性的序列。在一些实施例中,CTC PAM的靶向序列选自由SEQ ID NO:16203-21835组成的组。在一些实施例中,PAM序列为GTC。在一些实施例中,GTC PAM的靶向序列包含SEQ ID NO:12894-16202,或与SEQ ID NO:12894-16202具有至少50%同一性、至少55%同一性、至少60%同一性、至少65%同一性、至少70%同一性、至少75%同一性、至少80%同一性、至少85%同一性、至少90%同一性、至少95%同一性或至少99%同一性的序列。在一些实施例中,GTC PAM的靶向序列选自由SEQ ID NO:12894-16202组成的组。在一些实施例中,PAM序列为TTC。在一些实施例中,TTC PAM的靶向序列包含SEQ ID NO:5427-12893,或与SEQ ID NO:5427-12893具有至少50%同一性、至少55%同一性、至少60%同一性、至少65%同一性、至少70%同一性、至少75%同一性、至少80%同一性、至少85%同一性、至少90%同一性、至少95%同一性或至少99%同一性的序列。在一些实施例中,TTC PAM的靶向序列选自由SEQ ID NO:5427-12893组成的组。

[0170] 在一些实施例中,gNA变异体的支架是RNP的一部分,其具有包含SEQ ID NO:1、SEQ ID NO:2或SEQ ID NO:3的参考CasX蛋白。在其它实施例中,gNA变异体的支架是RNP的一部分,所述RNP具有包含表4、6、7、8或10的序列中的任一个,或与其具有至少约50%、至少约60%、至少约70%、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%同一性的序列的CasX变异蛋白质。在前述实施例中,gNA进一步包含间隔区序列。

[0171] 在一些实施例中,gNA变异体的支架是包含对包含SEQ ID NO:4或SEQ ID NO:5的参考gRNA的序列的一个或多个额外改变的变异体。在其中参考gRNA的支架衍生自SEQ ID NO:4或SEQ ID NO:5的那些实施例中,gNA变异体的一个或多个改进或增加的特征相比于SEQ ID NO:4或SEQ ID NO:5中的相同特征得到了改进。

[0172] m.与CasX蛋白形成复合物

[0173] 在一些实施例中,当相比于参考gRNA时,gNA变异体具有改进的与CasX蛋白(例如参考CasX或CasX变异蛋白)形成复合物的能力。在一些实施例中,当相比于参考gRNA时,gNA变异体具有改进的针对CasX蛋白(例如参考或变异蛋白)的亲和力,由此改进其与CasX蛋白形成核糖核蛋白(RNP)复合物的能力,如实例中所述。在一些实施例中,改进核糖核蛋白复合物形成可提高组装功能性RNP的效率。在一些实施例中,大于90%、大于93%、大于95%、

大于96％、大于97％、大于98％或大于99％的包含gNA变异体及其间隔区的RNP有能力对靶核酸进行基因编辑。

[0174] h.与CasX蛋白形成复合物

[0175] 在一些实施例中，当相比于参考gRNA时，gNA变异体具有改进的与CasX蛋白(例如参考CasX或CasX变异蛋白)形成复合物的能力。在一些实施例中，当相比于参考gRNA时，gNA变异体具有改进的针对CasX蛋白(例如参考或变异蛋白)的亲和力，由此改进其与CasX蛋白形成核糖核蛋白(RNP)复合物的能力，如实例中所述。在一些实施例中，改进核糖核蛋白复合物形成可提高组装功能性RNP的效率。在一些实施例中，大于90％、大于93％、大于95％、大于96％、大于97％、大于98％或大于99％的包含gNA变异体及其间隔区的RNP有能力对靶核酸进行基因编辑。

[0176] 在一些实施例中，可改进gNA变异体与CasX蛋白形成复合物的能力的示范性核苷酸变化可包括以热稳定茎环替换支架茎。不希望受任何理论束缚，以热稳定茎环替换支架茎可增加gNA变异体与CasX蛋白的总体结合稳定性。或者或另外，去除一大段茎环可改变gNA变异体折叠动力学，且使得功能性折叠gNA更容易且更快速地结构组装，例如通过减轻gNA变异体自身可变得“缠结”的程度。在一些实施例中，支架茎环序列的选择可随着用于gNA的不同间隔区而改变。在一些实施例中，支架序列可适于间隔区且因此适于靶序列。生物化学分析可用于评估CasX蛋白与gNA变异体结合以形成RNP的结合亲和力，包括实例的分析。举例来说，普通技术人员可测量结合至固定CasX蛋白的荧光标记gNA的量的变化，作为对增加额外未标记的“冷竞争者”gNA的浓度的反应。或者或另外，可监测荧光信号或查看其如何变化，因为不同量的经荧光标记的gNA流经固定CasX蛋白。或者，可使用体外裂解分析相对于界定靶核酸序列评估形成RNP的能力。

[0177] i.gNA稳定性

[0178] 在一些实施例中，当相比于参考gRNA时，gNA变异体具有改进的稳定性。在一些实施例中，增加的稳定性及有效折叠可增加gNA变异体持续存在于靶细胞内部的程度，其可由此提高形成能够执行CasX功能(例如基因编辑)的功能性RNP的概率。在一些实施例中，增加的gNA变异体稳定性亦可允许在向细胞递送较低量gNA的情况下的类似结果，其可转而降低基因编辑期间的脱靶效应的概率。可以多种方式评估引导RNA稳定性，包括例如在体外通过组装所述引导、在模拟细胞内环境的溶液中培育不同时段并且然后经由本文所述的体外裂解分析来测量功能活性。或者或另外，gNA可在初始转染/转导gNA之后的不同时间点自细胞收获，以确定gNA变异体相对于参考gRNA保持的时长。

[0179] j.溶解度

[0180] 在一些实施例中，当相比于参考gRNA时，gNA变异体具有改进的溶解度。在一些实施例中，当相比于参考gRNA时，gNA变异体具有改进的CasX蛋白:gNA RNP溶解度。在一些实施例中，CasX蛋白:gNA RNP的溶解度通过将核酶序列添加至gNA变异体的5’或3’端，例如参考sgRNA的5’或3’来改进。一些核酶，例如M1核酶可经由RNA介导的蛋白质折叠增加蛋白质的溶解度。包含如本文所述的gNA变异体的CasX RNP的增加的溶解度可经由所属领域技术人员已知的多种方法评估，例如通过在表达CasX及gNA变异体的裂解大肠杆菌的可溶部分的凝胶上获取密度测定法读数。

[0181] k.核酸酶活性抗性

[0182] 在一些实施例中，与参考gRNA相比，gNA变异体具有改进的核酸酶活性抗性，这可例如增加变异体gNA在细胞内环境中的持久性，从而改进基因编辑。核酸酶活性抗性可经由所属领域技术人员已知的多种方法来评估。举例来说，测量核酸酶活性抗性的体外方法可包括例如使参考gNA与具有一种或多种示范性RNA核酸酶的变异体接触及测量降解。或者或另外，使用本文所述的方法测量gNA变异体于细胞环境中的续存可指示gNA变异体的核酸酶抵抗性程度。

[0183] l.对靶DNA的结合亲和力

[0184] 在一些实施例中，相对于参考gRNA，gNA变异体具有对靶DNA的改进亲和力。在某些实施例中，相对于包含参考gRNA的RNP的亲和力，包含gNA变异体的核糖核蛋白复合物对靶DNA的亲和力有所提高。在一些实施例中，RNP对靶DNA的改进亲和力包括对靶序列的改进亲和力、对PAM序列的改进亲和力、RNP搜索用于靶序列的DNA的改进能力或其任何组合。在一些实施例中，对靶DNA的改进亲和力为增加的总体DNA结合亲和力的结果。

[0185] 在不希望受到理论约束的情况下，gNA变异体中影响CasX蛋白中OBD的功能的核苷酸变化可能会增加CasX变异蛋白与前间隔区邻近基序(PAM)结合的亲和力，以及结合或利用除了由SEQ ID NO:2的参考CasX蛋白识别的典型TTC PAM以外更多PAM序列(包括选自由TTC、ATC、GTC和CTC组成的组的PAM序列)，从而增加CasX变异蛋白对靶DNA序列的亲和力和多样性，使得与参考CasX相比，可编辑和/或结合的靶核酸序列大大增加。如下文更充分地描述，相比于参考CasX，增加可编辑的靶核酸的序列是指PAM及前间隔区序列及其根据非靶链定向的方向性。这不意味着由非靶链，而不是靶链的PAM序列决定裂解或在机制上涉及靶识别。举例来说，当参考TTC PAM时，其可实际上为靶标裂解所需的互补GAA序列，或其可为来自两条链的核苷酸的某一组合。在本文公开的CasX蛋白的情况下，PAM位于前间隔区的5'端，其中至少单个核苷酸将PAM与前间隔区的第一核苷酸分离。或者或另外，影响增加CasX变异蛋白对靶DNA链的亲和力的螺旋形I和/或螺旋形II结构域的功能的gNA的变化可增加包含变异gNA的CasX RNP对靶DNA的亲和力。

[0186] m.添加或改变gNA功能

[0187] 在一些实施例中，相对于参考gRNA，gNA变异体可包含改变gNA变异体的拓扑结构的较大结构变化，由此允许不同gNA功能。举例来说，在一些实施例中，gNA变异体用先前鉴别的稳定RNA结构或茎环交换参考gRNA支架的内源茎环，所述RNA结构或茎环可与蛋白质或RNA结合伴侣相互作用以将额外部分募集至CasX或将CasX募集至特定位置，例如具有与所述RNA结构的结合伴侣的病毒衣壳内部。在其它情境下，RNA可彼此补充(如在吻合环中)，使得两个CasX蛋白可共定位以在靶DNA序列处更有效地基因编辑。此类RNA结构可包括MS2、Qβ、U1发夹II、Uvsx、PP7、噬菌体复制环、吻合环_a、吻合环_b1、吻合环_b2、G四螺旋体M3q、G四螺旋体端粒篮、帚曲菌素-蓖麻毒素环或假结。

[0188] 在一些实施例中，gNA变异体包含末端融合伴侣。示范性末端融合物可包括gRNA与自裂解核酶或蛋白质结合基序的融合物。如本文所用，“核酶”是指具有一种或多种与蛋白质酶类似的催化活性的RNA或其区段。示范性核酶催化活性可包括例如RNA的裂解和/或连接、DNA的裂解和/或连接或肽键形成。在一些实施例中，此类融合可改进支架折叠或募集DNA修复机构。举例来说，在一些实施例中，gRNA可与丁型肝炎病毒(HDV)反基因组核酶、HDV基因组核酶、手斧核酶(来自宏基因组数据)、env25手枪核酶(代表物来自Aliistipes

putredinis)、HH15最小锤头核酶、烟草环斑病毒(TRSV)核酶、WT病毒锤头核酶(及合理变异体)或扭曲姊妹1或RBMX募集基序融合。锤头核酶为在RNA分子内的特定位点处催化可逆裂解及连接反应的RNA基序。锤头核酶包括I型、II型及III型锤头核酶。HDV、手枪及手斧核酶具有自裂解活性。包含一种或多种核酶的gNA变异体可允许相比于gRNA参考物扩展的gNA功能。举例来说,在一些实施例中,包含自裂解核酶的gNA可转录及加工为成熟gNA,作为多顺反子转录物的一部分。此类融合物可出现于gNA的5'或3'端。在一些实施例中,gNA变异体在5'及3'端处均包含融合物,其中各融合物独立地如本文所述。在一些实施例中,gNA变异体包含噬菌体复制环或四环。在一些实施例中,gNA包含能够结合蛋白质的发夹环。举例来说,在一些实施例中,发夹环为MS2、Qβ、U1发夹II、Uvsx或PP7发夹环。

[0189] 在一些实施例中,gNA变异体包含一个或多个RNA适体。如本文所用,"RNA适体"是指以高亲和力及高特异性结合靶标的RNA分子。在一些实施例中,gNA变异体包含一个或多个核糖开关。如本文所用,"核糖开关"是指在结合小分子时改变状态的RNA分子。在一些实施例中,gNA变异体进一步包含一个或多个蛋白质结合基序。在一些实施例中,将蛋白质结合基序添加到本公开的参考gRNA或gNA变异体可允许CasX RNP与额外蛋白质缔合,其可例如将那些蛋白质的功能添加到CasX RNP。

[0190] n.化学修饰的gNA

[0191] 在一些实施例中,本公开涉及化学修饰的gNA。在一些实施例中,本公开提供了一种化学修饰的gNA,其具有引导RNA功能且降低了对通过核酸酶裂解的易感性。包含除四种典型核糖核苷酸A、C、G及U或脱氧核苷酸以外的任何核苷酸的gNA为经化学修饰的gNA。在一些情况下,经化学修饰的gNA包含除天然磷酸二酯核苷酸间键以外的任何主链或核苷酸间键。在某些实施例中,保留功能包括经修饰gNA结合至本文所描述的任一实施例的CasX的能力。在某些实施例中,所保留的功能性包括经修饰gNA与C9orf72靶核酸序列结合的能力。在某些实施例中,保留功能包括靶向CasX蛋白或预复合CasX蛋白gNA结合到靶核酸序列的能力。在某些实施例中,保留功能包括通过CasX-gNA切割靶多核苷酸的能力。在某些实施例中,保留功能包括通过CasX-gNA裂解靶核酸序列的能力。在某些实施例中,保留功能是gNA于具有本公开实施例中含有CasX蛋白的CasX系统中gNA的任何其它已知功能。

[0192] 在一些实施例中,本公开提供了一种化学修饰的gNA,其中核苷酸糖修饰并入至选自由以下组成的组的gNA中:2′-O—C_{1-4}烷基(如2′-O-甲基(2′-OMe))、2'-脱氧基(2′-H)、2′-O—C_{1-3}烷基-O—C_{1-3}烷基(如2′-甲氧基乙基("2′-MOE"))、2'-氟基("2'-F")、2'-氨基("2'-NH_2")、2'-阿拉伯糖基("2'-阿糖")核苷酸、2'-F-阿拉伯糖基("2'-F-阿糖")核苷酸、2'-锁定核酸("LNA")核苷酸、2'-解锁核酸("ULNA")核苷酸、L形式的糖("L-糖")和4'-硫代核糖基核苷酸。在其它实施例中,并入引导RNA的核苷酸间连键修饰选自由以下组成的组:硫代磷酸酯"P(S)"(P(S))、膦酰基羧酸酯(P(CH_2)$_n$COOR)(如膦酰基乙酸酯"PACE"(P(CH_2COO^-)))、硫代膦酸羧酸酯((S)P(CH_2)$_n$COOR)(如硫代膦酸乙酸酯"thioPACE"((S)P(CH_2)$_n$$COO^-$)))、烷基膦酸酯(P($C_{1-3}$烷基)(如甲基膦酸酯-P($CH_3$))、硼烷膦酸酯(P($BH_3$))和二硫代磷酸酯(P(S)$_2$)。

[0193] 在某些实施例中,本公开提供经化学修饰的gNA,其中核碱基("碱基")修饰并入至选自由以下组成的组的gNA中:2-硫尿嘧啶("2-thioU")、2-硫胞嘧啶("2-thioC")、4-硫尿嘧啶("4-thioU")、6-硫鸟嘌呤("6-thioG")、2-氨基腺嘌呤("2-aminoA")、2-氨基嘌呤、假

尿嘧啶、次黄嘌呤、7-去氮鸟嘌呤、7-去氮-8-氮杂鸟嘌呤、7-去氮腺嘌呤、7-去氮-8-氮杂腺嘌呤、5-甲基胞嘧啶(“5-methylC”)、5-甲基尿嘧啶(“5-methylU”)、5-羟甲基胞嘧啶、5-羟甲基尿嘧啶、5,6-去氢尿嘧啶、5-丙炔基胞嘧啶、5-丙炔基尿嘧啶、5-乙炔基胞嘧啶、5-乙炔基尿嘧啶、5-烯丙基尿嘧啶(“5-allylU”)、5-烯丙基胞嘧啶(“5-allylC”)、5-氨基烯丙基尿嘧啶(“5-aminoallylU”)、5-氨基烯丙基-胞嘧啶(“5-aminoallylC”)、无碱基核苷酸、Z碱基、P碱基、非结构化核酸(“UNA”)、异鸟嘌呤(“isoG”)、异胞嘧啶(“isoC”)、5-甲基-2-嘧啶、x(A、G、C、T)及y(A、G、C、T)。

[0194] 在其它实施例中,本公开提供经化学修饰的gNA,其中在核苷酸糖、核碱基、磷酸二酯键和/或磷酸核苷酸,包括包含一个或多个^{15}N、^{13}C、^{14}C、氘、^{3}H、^{32}P、^{125}I、^{131}I原子或其它用作示踪剂的原子或元素的核苷酸上引入一个或多个同位素修饰。

[0195] 在一些实施例中,并入至gNA中的“末端”修饰选自由以下组成的组:聚乙二醇(PEG);烃接头(包括:杂原子(O、S、N)取代的烃间隔基;卤基取代的烃间隔基;含酮基、羧基、酰氨基、亚硫酰基、氨甲酰基、硫羰氨甲酰基的烃间隔基);精胺接头;包括附接到例如6-荧光素-己基的接头的荧光染料(例如荧光素、若丹明、花青)的染料;淬灭剂(例如dabcyl、BHQ);和其它标记(例如生物素、地高辛(digoxigenin)、吖啶、抗生蛋白链菌素、抗生物素蛋白、肽和/或蛋白质)。在一些实施例中,“末端”修饰包含将gNA结合(或连接)至包含脱氧核苷酸和/或核糖核苷酸的寡核苷酸的另一分子、肽、蛋白质、糖、寡糖、类固醇、脂质、叶酸、维生素和/或其它分子。在某些实施例中,本公开提供经化学修饰的gNA,其中“末端”修饰(上文所述)经由例如2-(4-丁基酰氨基荧光素)丙烷-1,3-二醇双(磷酸二酯)接头的接头定位于gNA序列内部,所述接头以磷酸二酯键形式并入且可并入gNA中的两个核苷酸之间的任何位置。

[0196] 在一些实施例中,本公开提供具有末端修饰的经化学修饰的gNA,所述末端修饰包含末端官能团,例如胺、硫醇(或巯基)、羟基、羧基、羰基、亚硫酰基、硫羰基、胺甲酰基、胺(硫甲酰)基、磷酰基、烯烃、炔烃、卤素或官能团封端的接头,其可随后结合至选自由以下组成的组的所需部分:荧光染料、非荧光标记、标签(例如^{14}C、生物素、抗生物素蛋白、抗生蛋白链菌素或含有同位素标记,例如^{15}N、^{13}C、氘、^{3}H、^{32}P、^{125}I等的部分)、寡核苷酸(包含脱氧核苷酸和/或核糖核苷酸,包括适体)、氨基酸、肽、蛋白质、糖、寡糖、类固醇、脂质、叶酸及维生素。共轭采用所属领域中熟知的标准化学方法,包括但不限于经由N-羟基丁二酰亚胺、异硫氰酸酯、DCC(或DCI)偶合,和/或如出版社爱思唯尔科学公司(Eslsevier Science)Greg T.Hermanson在《Bioconjugate Techniques》第3版(2013)中所述的任何其它标准方法,所述文献的内容以全文引用的方式并入本文中。

[0197] IV.用于修饰靶核酸的蛋白质

[0198] 本公开提供了包含CRISPR核酸酶的系统,其可用于真核细胞的基因组编辑。在一些实施例中,基因组编辑系统中采用的CRISPR核酸酶为2类V型核酸酶。尽管2类V型CRISPR-Cas系统的成员具有差异,但它们具有一些共同特征,这些特征将它们与Cas9系统区分开来。首先,V型核酸酶具有单RNA引导的含RuvC结构域的效应子,但没有HNH结构域,并且它们识别非靶向链上的靶区上游5’的富含T的PAM,这不同于依赖靶序列3’侧富含G的PAM的Cas9系统。V型核酸酶在PAM序列的远端产生交错的双链断裂,这不同于Cas9,后者在靠近PAM的近端位点产生一个平端。此外,当被顺式结合的靶dsDNA或ssDNA激活时,V型核酸酶反式降

解ssDNA。在一些实施例中，实施例的V型核酸酶识别5′-TC PAM基序，并产生仅被RuvC结构域裂解的交错末端。在一些实施例中，V型核酸酶选自由Cas12a、Cas12b、Cas12c、Cas12d(CasY)和CasX组成的组。在一些实施例中，V型核酸酶为CasX核酸酶。在一些实施例中，本公开提供了包含CasX蛋白和一种或多种gNA酸的系统(CasX:gNA系统)，这些系统被专门设计成修饰真核细胞中的靶核酸序列。

[0199] 如本文所用，术语“CasX蛋白”是指蛋白质家族，且涵盖所有天然存在的CasX蛋白、与天然存在的CasX蛋白具有至少50％同一性的蛋白质以及相对于天然存在的参考CasX蛋白呈现一种或多种改进特征的CasX变异体。

[0200] CasX变异体实施例的示范性改进特征包括但不限于改进的变异体折叠、改进的对gNA的结合亲和力、改进的对靶核酸的结合亲和力、改进的利用较大范围的PAM序列编辑和/或结合靶DNA的能力、改进的靶DNA退绕、增加的编辑活性、改进的编辑效率、改进的编辑特异性、增加的可有效编辑的真核基因组的百分比、增加的核酸酶活性、增加的用于双链裂解的靶链负载、减少的用于单链切割的靶链负载、减少的脱靶裂解、改进的DNA的非靶链的结合、改进的蛋白质稳定性、改进的蛋白质:gNA(RNP)复合物稳定性、改进的蛋白质溶解度、改进的蛋白质:gNA(RNP)复合物溶解度、改进的蛋白质产率、改进的蛋白质表达及改进的熔融特征，如下文更充分地描述。在一些实施例中，当以可比较的方式分析时，CasX变异体和gNA变异体的RNP表现出一种或多种改进特征，其相对于SEQ ID NO:1、SEQ ID NO:2或SEQ ID NO:3的参考CasX蛋白和表1的gNA的RNP改进至少约1.1至约100,000倍。在其它情况下，CasX变异体和gNA变异体的RNP的一种或多种改进特征是相对于SEQ ID NO:1、SEQ ID NO:2或SEQ ID NO:3的参考CasX蛋白和表1的gNA的RNP改进至少约1.1、至少约10、至少约100、至少约1000、至少约10,000、至少约100,000倍或更多。在其它情况下，当以可比较的方式分析时，CasX变异体和gNA变异体的RNP的改进特征中的一个或多个是相对于SEQ ID NO:1、SEQ ID NO:2或SEQ ID NO:3的参考CasX蛋白和表1的gNA的RNP改进约1.1至100,00倍、约1.1至10,00倍、约1.1至1,000倍、约1.1至500倍、约1.1至100倍、约1.1至50倍、约1.1至20倍、约10至100,00倍、约10至10,00倍、约10至1,000倍、约10至500倍、约10至100倍、约10至50倍、约10至20倍、约2至70倍、约2至50倍、约2至30倍、约2至20倍、约2至10倍、约5至50倍、约5至30倍、约5至10倍、约100至100,00倍、约100至10,00倍、约100至1,000倍、约100至500倍、约500至100,00倍、约500至10,00倍、约500至1,000倍、约500至750倍、约1,000至100,00倍、约10,000至100,00倍、约20至500倍、约20至250倍、约20至200倍、约20至100倍、约20至50倍、约50至10,000倍、约50至1,000倍、约50至500倍、约50至200倍或约50至100倍。在其它情况下，当以可比较的方式分析时，CasX变异体和gNA变异体的RNP的一种或多种改进特征是相对于SEQ ID NO:1、SEQ ID NO:2或SEQ ID NO:3的参考CasX蛋白和表1的gNA的RNP改进约1.1倍、1.2倍、1.3倍、1.4倍、1.5倍、1.6倍、1.7倍、1.8倍、1.9倍、2倍、3倍、4倍、5倍、6倍、7倍、8倍、9倍、10倍、11倍、12倍、13倍、14倍、15倍、16倍、17倍、18倍、19倍、20倍、25倍、30倍、40倍、45倍、50倍、55倍、60倍、70倍、80倍、90倍、100倍、110倍、120倍、130倍、140倍、150倍、160倍、170倍、180倍、190倍、200倍、210倍、220倍、230倍、240倍、250倍、260倍、270倍、280倍、290倍、300倍、310倍、320倍、330倍、340倍、350倍、360倍、370倍、380倍、390倍、400倍、425倍、450倍、475倍或500倍。

[0201] 术语“CasX变异体”包括为融合蛋白的变异体；即CasX“融合至”异源序列。此包括

包含CasX变异体序列及CasX与异源蛋白或其结构域的N端、C端或内部融合物的CasX变异体。

[0202] 本公开的CasX蛋白包含以下结构域中的至少一个:非靶链结合(NTSB)结构域、靶链负载(TSL)结构域、螺旋形I结构域、螺旋形II结构域、寡核苷酸结合结构域(OBD)及RuvC DNA裂解结构域(其中的最后一个可在催化死亡的CasX变异体中经修饰或缺失),下文将更充分描述。另外,与参考CasX蛋白和参考gNA的RNP相比,本公开的CasX变异蛋白在与gNA复合成为RNP时,利用PAM TC基序(包括选自TTC、ATC、GTC或CTC的PAM序列),具有增强的高效编辑和/或结合靶DNA的能力。在前文中,与包含参考CasX蛋白和参考gNA的RNP在可比分析系统中的编辑效率和/或结合相比,PAM序列位于与分析系统中gNA的靶向序列具有同一性的前间隔区的非靶链的5′端至少1个核苷酸处。在一个实施例中,CasX变异体和gNA变异体的RNP在可比较的分析系统中与包含参考CasX蛋白和参考gNA的RNP相比表现出更高的编辑效率和/或靶DNA中靶序列的结合,其中靶DNA的PAM序列是TTC。在另一实施例中,CasX变异体和gNA变异体的RNP在可比较的分析系统中与包含参考CasX蛋白和参考gNA的RNP相比表现出更高的编辑效率和/或靶DNA中靶序列的结合,其中靶DNA的PAM序列是ATC。在另一实施例中,CasX变异体和gNA变异体的RNP在可比较的分析系统中与包含参考CasX蛋白和参考gNA的RNP相比表现出更高的编辑效率和/或靶DNA中靶序列的结合,其中靶DNA的PAM序列是CTC。在另一实施例中,CasX变异体和gNA变异体的RNP在可比较的分析系统中与包含参考CasX蛋白和参考gNA的RNP相比表现出更高的编辑效率和/或靶DNA中靶序列的结合,其中靶DNA的PAM序列是GTC。在前述实施例中,与SEQ ID NO:1-3的任何一种CasX蛋白和表1的gNA的RNP对PAM序列的编辑效率和/或结合亲和力相比,一种或多种PAM序列的增加的编辑效率和/或结合亲和力至少大1.5倍或更多。

[0203] 在一些实施例中,CasX蛋白可结合和/或修饰(例如裂解、切割、甲基化、去甲基等)靶核酸和/或与靶核酸相关的多肽(例如组蛋白尾的甲基化或乙酰化)。在一些实施例中,CasX蛋白为催化死亡的(dCasX),但保留结合靶核酸的能力。示范性催化死亡的CasX蛋白包含CasX蛋白的RuvC结构域的活性位点中的一个或多个突变。在一些实施例中,催化死亡的CasX蛋白包含SEQ ID NO:1的残基672、769和/或935处的取代。在一个实施例中,催化死亡的CasX蛋白包含SEQ ID NO:1的参考CasX蛋白中D672A、E769A和/或D935A取代。在其它实施例中,催化死亡的CasX蛋白包含SEQ ID NO:2的参考CasX蛋白中氨基酸659、756和/或922处的取代。在一些实施例中,催化死亡的CasX蛋白包含SEQ ID NO:2的参考CasX蛋白中D659A、E756A和/或D922A取代。在其它实施例中,催化死亡的CasX蛋白包含CasX蛋白的全部或一部分RuvC结构域的缺失。应了解,相同的前述取代可类似地引入至本公开的CasX变异体中,产生dCasX变异体。在一个实施例中,全部或一部分RuvC结构域自CasX变异体缺失,产生dCasX变异体。在一些实施例中,无催化活性的dCasX变异蛋白可用于碱基编辑或表观遗传修饰。在对DNA的较高亲和力下,在一些实施例中,相对于催化活性CasX,无催化活性的dCasX变异蛋白可以更快地发现其靶核酸、与靶核酸保持结合的时间更长、以更稳定方式结合靶核酸或其组合,从而与保留其裂解能力的CasX变异体相比,改善催化死亡的CasX变异蛋白的这些功能。

[0204] a.非靶链结合结构域

[0205] 本公开的参考CasX蛋白包含非靶链结合结构域(NTSBD)。NTSBD为先前未发现于任

何Cas蛋白中的结构域;举例来说,此结构域不存在于Cas蛋白,例如Cas9、Cas12a/Cpf1、Cas13、Cas14、CASCADE、CSM或CSY中。不受理论或机制束缚,CasX中的NTSBD允许结合至非靶DNA链且可帮助非靶及靶链的解旋。NTSBD被认为负责非靶DNA链的解旋或呈解旋状态的非靶DNA链的捕捉。NTSBD与迄今为止派生的CryoEM模型结构中的非靶链直接接触,且可含有非典型锌指结构域。NTSBD亦可在解旋、引导RNA侵入及R环形成期间于稳定DNA中起作用。在一些实施例中,示范性NTSBD包含SEQ ID NO:1的氨基酸101-191或SEQ ID NO:2的氨基酸103-192。在一些实施例中,参考CasX蛋白的NTSBD包含四链β折叠。

[0206] b.靶链负载结构域

[0207] 本公开的参考CasX蛋白包含靶链负载(TSL)结构域。TSL结构域为某些Cas蛋白,例如Cas9、CASCADE、CSM或CSY中未发现的结构域。不希望受理论或机制所束缚,认为TSL结构域负责辅助将靶DNA链负载至CasX蛋白的RuvC活性位点中。在一些实施例中,TSL用以放置或捕捉呈折叠状态的靶链,其将靶链DNA主链的易切断磷酸置于RuvC活性位点中。TSL包含由TSL的主体分隔的cys4(CXXC、CXXC锌指/带结构域(SEQ ID NO:48))。在一些实施例中,示范性TSL包含SEQ ID NO:1的氨基酸825-934或SEQ ID NO:2的氨基酸813-921。

[0208] c.螺旋形I结构域

[0209] 本公开的参考CasX蛋白包含螺旋形I结构域。除CasX以外的某些Cas蛋白具有可以类似方式命名的结构域。然而,在一些实施例中,相比于非CasX蛋白,CasX蛋白的螺旋形I结构域包含一个或多个独特结构特征,或包含独特序列,或其组合。举例来说,在一些实施例中,相比于可具有类似名称的其它Cas蛋白中的结构域,CasX蛋白的螺旋形I结构域包含一个或多个独特二级结构。举例来说,在一些实施例中,相比于其它CRISPR蛋白,CasX蛋白中的螺旋形I结构域包含布置、数目及长度独特的结构及序列的一个或多个α螺旋。在某些实施例中,螺旋形I结构域负责与引导RNA的结合DNA及间隔区相互作用。不希望受理论所束缚,认为在一些情况下,螺旋形I结构域可促进前间隔区邻近基序(PAM)的结合。在一些实施例中,示范性螺旋形I结构域包含SEQ ID NO:1的氨基酸57-100和192-332,或SEQ ID NO:2的氨基酸59-102和193-333。在一些实施例中,参考CasX蛋白的螺旋形I结构域包含一个或多个α螺旋。

[0210] d.螺旋形II结构域

[0211] 本公开的参考CasX蛋白包含螺旋形II结构域。除CasX以外的某些Cas蛋白具有可以类似方式命名的结构域。然而,在一些实施例中,相比于可具有类似名称的其它Cas蛋白中的结构域,CasX蛋白的螺旋形II结构域包含一个或多个独特结构特征,或独特序列,或其组合。举例来说,在一些实施例中,螺旋形II结构域包含沿靶DNA:引导RNA通道对准的一个或多个独特结构性α螺旋束。在一些实施例中,在包含螺旋形II结构域的CasX中,靶链及引导RNA与螺旋形II(且在一些实施例中,螺旋形I结构域)相互作用,以允许RuvC结构域接近靶DNA。螺旋形II结构域负责结合至引导RNA支架茎环以及结合DNA。在一些实施例中,示范性螺旋形II结构域包含SEQ ID NO:1的氨基酸333-509,或SEQ ID NO:2的氨基酸334-501。

[0212] e.寡核苷酸结合结构域

[0213] 本公开的参考CasX蛋白包含寡核苷酸结合结构域(OBD)。除CasX以外的某些Cas蛋白具有可以类似方式命名的结构域。然而,在一些实施例中,OBD包含一种或多种独特功能特征,或包含相对于CasX蛋白独特的序列,或其组合。举例来说,在一些实施例中,桥连螺旋

(BH)、螺旋形I结构域、螺旋形II结构域及寡核苷酸结合结构域(OBD)一起负责将CasX蛋白结合至引导RNA。因此,举例来说,在一些实施例中,OBD相对于CasX蛋白的独特的处在于其与螺旋形I结构域,或螺旋形II结构域或两者功能上相互作用,所述结构域各自可相对于如本文所述的CasX蛋白为独特的。具体地,在CasX中,OBD很大程度上结合引导RNA支架的RNA三螺旋体。OBD亦可负责结合至前间隔区邻近基序(PAM)。示范性OBD结构域包含SEQ ID NO:1的氨基酸1-56和510-660,或SEQ ID NO:2的氨基酸1-58和502-647。

[0214] f.RuvC DNA裂解结构域

[0215] 本公开的参考CasX蛋白包含RuvC结构域,其包括2个部分RuvC结构域(RuvC-I及RuvC-II)。RuvC结构域为所有12型CRISPR蛋白的祖先结构域。RuvC结构域源自TNPB(转座酶B)样转座酶。与其它RuvC结构域类似,CasX RuvC结构域具有负责配位镁(Mg)离子及裂解DNA的DED催化三联体。在一些实施例中,RuvC具有负责裂解DNA的两条链(一个接一个地,最可能首先为靶向序列中11-14个核苷酸(nt)处的非靶链,且接着随后为靶序列之后2-4个核苷酸附近的靶链)的DED基序活性位点。具体地,在CasX中,RuvC结构域的独特之处在于其亦负责结合对CasX功能重要的引导RNA支架茎环。示范性RuvC结构域包含SEQ ID NO:1的氨基酸661-824和935-986,或SEQ ID NO:2的氨基酸648-812和922-978。

[0216] g.参考CasX蛋白

[0217] 本公开提供了天然存在的CasX蛋白(在本文中称为“参考CasX蛋白”),其充当催化靶向双链DNA(dsDNA)中特定序列处的双链断裂的核酸内切酶。序列特异性由其所复合的相关gNA的靶向序列提供,所述靶向序列与靶核酸内的靶序列杂交。举例来说,参考CasX蛋白可自天然存在的原核生物,例如δ变形菌纲、浮霉菌门或宋氏细菌暂定种物种分离。参考CasX蛋白(有时在本文中称为参考CasX蛋白)为V型CRISPR/Cas核酸内切酶,其属于能够与引导NA相互作用以形成核糖核蛋白(RNP)复合物的CasX(有时称为Cas12e)蛋白家族。在一些实施例中,包含参考CasX蛋白的RNP复合物可经由gNA的靶向序列(或间隔区)与靶核酸中的靶序列之间的碱基配对靶向至靶核酸中的特定位点。在一些实施例中,包含参考CasX蛋白的RNP能够裂解靶DNA。在一些实施例中,包含参考CasX蛋白的RNP能够切割靶DNA。在一些实施例中,包含参考CasX蛋白的RNP能够编辑靶DNA,例如在如下那些实施例中,其中参考CasX蛋白能够裂解或切割DNA,接着为非同源末端连接(NHEJ)、同源定向修复(HDR)、同源独立性靶向集成(HITI)、微同源性介导的末端连接(MMEJ)、单链退火(SSA)或碱基切除修复(BER)。在一些实施例中,包含CasX蛋白的RNP为催化死亡(无催化活性或基本上无裂解活性)CasX蛋白(dCasX),但保留结合靶DNA的能力,更充分描述于前文。

[0218] 在一些情况下,参考CasX蛋白分离或衍生自δ变形菌纲。在一些实施例中,CasX蛋白包含与以下序列具有至少50%同一性、至少60%同一性、至少65%同一性、至少70%同一性、至少75%同一性、至少80%同一性、至少81%同一性、至少82%同一性、至少83%同一性、至少84%同一性、至少85%同一性、至少86%同一性、至少86%同一性、至少87%同一性、至少88%同一性、至少89%同一性、至少89%同一性、至少90%同一性、至少91%同一性、至少92%同一性、至少93%同一性、至少94%同一性、至少95%同一性、至少96%同一性、至少97%同一性、至少98%同一性、至少99%同一性、至少99.5%同一性或100%同一性的序列:

```
   1 MEKRINKIRK KLSADNATKP VSRSGPMKTL LVRVMTDDLK KRLEKRRKKP EVMPQVISNN
  61 AANNLRMLLD DYTKMKEAIL QVYWQEFKDD HVGLMCKFAQ PASKKIDQNK LKPEMDEKGN
 121 LTTAGFACSQ CGQPLFVYKL EQVSEKGKAY TNYFGRCNVA EHEKLILLAQ LKPEKDSDEA
 181 VTYSLGKFGQ RALDFYSIHV TKESTHPVKP LAQIAGNRYA SGPVGKALSD ACMGTIASFL
 241 SKYQDIIIEH QKVVKGNQKR LESLRELAGK ENLEYPSVTL PPQPHTKEGV DAYNEVIARV
 301 RMWVNLNLWQ KLKLSRDDAK PLLRLKGFPS FPVVERRENE VDWWNTINEV KKLIDAKRDM
 361 GRVFWSGVTA EKRNTILEGY NYLPNENDHK KREGSLENPK KPAKRQFGDL LLYLEKKYAG
 421 DWGKVFDEAW ERIDKKIAGL TSHIEREEAR NAEDAQSKAV LTDWLRAKAS FVLERLKEMD
[0219] 481 EKEFYACEIQ LQKWYGDLRG NPFAVEAENR VVDISGFSIG SDGHSIQYRN LLAWKYLENG
 541 KREFYLLMNY GKKGRIRFTD GTDIKKSGKW QGLLYGGGKA KVIDLTFDPD DEQLIILPLA
 601 FGTRQGREFI WNDLLSLETG LIKLANGRVI EKTIYNKKIG RDEPALFVAL TFERREVVDP
 661 SNIKPVNLIG VDRGENIPAV IALTDPEGCP LPEFKDSSGG PTDILRIGEG YKEKQRAIQA
 721 AKEVEQRRAG GYSRKFASKS RNLADDMVRN SARDLFYHAV THDAVLVFEN LSRGFGRQGK
 781 RTFMTERQYT KMEDWLTAKL AYEGLTSKTY LSKTLAQYTS KTCSNCGFTI TTADYDGMLV
 841 RLKKTSDGWA TTLNNKELKA EGQITYYNRY KRQTVEKELS AELDRLSEES GNNDISKWTK
 901 GRRDEALFLL KKRFSHRPVQ EQFVCLDCGH EVHADEQAAL NIARSWLFLN SNSTEFKSYK
 961 SGKQPFVGAW QAFYKRRLKE VWKPNA（SEQ ID NO: 1）。
```

[0220] 在一些情况下，参考CasX蛋白分离或衍生自浮霉菌门。在一些实施例中，CasX蛋白包含与以下序列具有至少50％同一性、至少60％同一性、至少65％同一性、至少70％同一性、至少75％同一性、至少80％同一性、至少81％同一性、至少82％同一性、至少83％同一性、至少84％同一性、至少85％同一性、至少86％同一性、至少86％同一性、至少87％同一性、至少88％同一性、至少89％同一性、至少89％同一性、至少90％同一性、至少91％同一性、至少92％同一性、至少93％同一性、至少94％同一性、至少95％同一性、至少96％同一性、至少97％同一性、至少98％同一性、至少99％同一性、至少99.5％同一性或100％同一性的序列：

```
   1 MQEIKRINKI RRRLVKDSNT KKAGKTGPMK TLLVRVMTPD LRERLENLRK KPENIPQPIS
  61 NTSRANLNKL LTDYTEMKKA ILHVYWEEFQ KDPVGLMSRV AQPAPKNIDQ RKLIPVKDGN
 121 ERLTSSGFAC SQCCQPLYVY KLEQVNDKGK PHTNYFGRCN VSEHERLILL SPHKPEANDE
 181 LVTYSLGKFG QRALDFYSIH VTRESNHPVK PLEQIGGNSC ASGPVGKALS DACMGAVASF
 241 LTKYQDIILE HQKVIKKNEK RLANLKDIAS ANGLAFPKIT LPPQPHTKEG IEAYNNVVAQ
 301 IVIWVNLNLW QKLKIGRDEA KPLQRLKGFP SFPLVERQAN EVDWWDMVCN VKKLINEKKE
 361 DGKVFWQNLA GYKRQEALLP YLSSEEDRKK GKKFARYQFG DLLLHLEKKH GEDWGKVYDE
 421 AWERIDKKVE GLSKHIKLEE ERRSEDAQSK AALTDWLRAK ASFVIEGLKE ADKDEFCRCE
[0221] 481 LKLQKWYGDL RGKPFAIEAE NSILDISGFS KQYNCAFIWQ KDGVKKLNLY LIINYFKGGK
 541 LRFKKIKPEA FEANRFYTVI NKKSGEIVPM EVNFNFDDPN LIILPLAFGK RQGREFIWND
 601 LLSLETGSLK LANGRVIEKT LYNRRTRQDE PALFVALTFE RREVLDSSNI KPMNLIGIDR
 661 GENIPAVIAL TDPEGCPLSR FKDSLGNPTH ILRIGESYKE KQRTIQAAKE VEQRRAGGYS
 721 RKYASKAKNL ADDMVRNTAR DLLYYAVTQD AMLIFENLSR GFGRQGKRTF MAERQYTRME
 781 DWLTAKLAYE GLPSKTYLSK TLAQYTSKTC SNCGFTITSA DYDRVLEKLK KTATGWMTTI
 841 NGKELKVEGQ ITYYNRYKRQ NVVKDLSVEL DRLSEESVNN DISSWTKGRS GEALSLLKKR
 901 FSHRPVQEKF VCLNCGFETH ADEQAALNIA RSWLFLRSQE YKKYQTNKTT GNTDKRAFVE
 961 TWQSFYRKKL KEVWKPAV（SEQ ID NO: 2）。
```

[0222] 在一些实施例中，CasX蛋白包含SEQ ID NO:2，或与其至少60％类似的序列。在一些实施例中，CasX蛋白包含SEQ ID NO:2，或与其至少80％类似的序列。在一些实施例中，CasX蛋白包含SEQ ID NO:2，或与其至少90％类似的序列。在一些实施例中，CasX蛋白包含SEQ ID NO:2，或与其至少95％类似的序列。在一些实施例中，CasX蛋白由SEQ ID NO:2的序列组成。在一些实施例中，CasX蛋白包含相对于SEQ ID NO:2的序列具有至少1个、至少2个、

至少3个、至少4个、至少5个、至少6个、至少7个、至少8个、至少9个、至少10个、至少20个、至少30个、至少40个或至少50个突变的序列或由其组成。这些突变可为插入、缺失、氨基酸取代或其任何组合。

[0223] 在一些情况下，参考CasX蛋白分离或衍生自宋氏细菌暂定种。在一些实施例中，CasX蛋白包含与以下序列具有至少50％同一性、至少60％同一性、至少65％同一性、至少70％同一性、至少75％同一性、至少80％同一性、至少81％同一性、至少82％同一性、至少83％同一性、至少84％同一性、至少85％同一性、至少86％同一性、至少86％同一性、至少87％同一性、至少88％同一性、至少89％同一性、至少89％同一性、至少90％同一性、至少91％同一性、至少92％同一性、至少93％同一性、至少94％同一性、至少95％同一性、至少96％同一性、至少97％同一性、至少98％同一性、至少99％同一性、至少99.5％同一性或100％同一性的序列：

[0224]
```
  1 MDNANKPSTK SLVNTTRISD HFGVTPGQVT RVFSFGIIPT KRQYAIIERW FAAVEAARER
 61 LYGMLYAHFQ ENPPAYLKEK FSYETFFKGR PVLNGLRDID PTIMTSAVFT ALRHKAEGAM
121 AAFHTNHRRL FEEARKKMRE YAECLKANEA LLRGAADIDW DKIVNALRTR LNTCLAPEYD
181 AVIADFGALC AFRALIAETN ALKGAYNHAL NQMLPALVKV DEPEEAEESP RLRFFNGRIN
241 DLPKFPVAER ETPPDTETII RQLEDMARVI PDTAEILGYI HRIRHKAARR KPGSAVPLPQ
301 RVALYCAIRM ERNPEEDPST VAGHFLGEID RVCEKRRQGL VRTPFDSQIR ARYMDIISFR
361 ATLAHPDRWT EIQFLRSNAA SRRVRAETIS APFEGFSWTS NRTNPAPQYG MALAKDANAP
421 ADAPELCICL SPSSAAFSVR EKGGDLIYMR PTGGRRGKDN PGKEITWVPG SFDEYPASGV
481 ALKLRLYFGR SQARRMLTNK TWGLLSDNPR VFAANAELVG KKRNPQDRWK LFFHMVISGP
541 PPVEYLDFSS DVRSRARTVI GINRGEVNPL AYAVVSVEDG QVLEEGLLGK KEYIDQLIET
601 RRRISEYQSR EQTPPRDLRQ RVRHLQDTVL GSARAKIHSL IAFWKGILAI ERLDDQFHGR
661 EQKIIPKKTY LANKTGFMNA LSFSGAVRVD KKGNPWGGMI EIYPGGISRT CTQCGTVWLA
721 RRPKNPGHRD AMVVIPDIVD DAAATGFDNV DCDAGTVDYG ELFTLSREWV RLTPRYSRVM
781 RGTLGDLERA IRQGDDRKSR QMLELALEPQ PQWGQFFCHR CGFNGQSDVL AATNLARRAI
841 SLIRRLPDTD TPPTP（SEQ ID NO: 3）。
```

[0225] 在一些实施例中，CasX蛋白包含SEQ ID NO:3，或与其至少60％类似的序列。在一些实施例中，CasX蛋白包含SEQ ID NO:3，或与其至少80％类似的序列。在一些实施例中，CasX蛋白包含SEQ ID NO:3，或与其至少90％类似的序列。在一些实施例中，CasX蛋白包含SEQ ID NO:3，或与其至少95％类似的序列。在一些实施例中，CasX蛋白由SEQ ID NO:3的序列组成。在一些实施例中，CasX蛋白包含相对于SEQ ID NO:3的序列具有至少1个、至少2个、至少3个、至少4个、至少5个、至少6个、至少7个、至少8个、至少9个、至少10个、至少20个、至少30个、至少40个或至少50个突变的序列或由其组成。这些突变可为插入、缺失、氨基酸取代或其任何组合。

[0226] h.CasX变异蛋白

[0227] 本公开提供了参考CasX蛋白(在本文中可互换地称为“CasX变异体”或“CasX变异蛋白”)的变异体，其中CasX变异体在参考CasX蛋白的至少一个结构域中包含至少一个修饰，包括SEQ ID NO:1 -3的序列。在一些实施例中，相比于参考CasX蛋白，CasX变异体展现至少一种改进特征。当相比于本文所述的参考CasX蛋白时改进CasX变异蛋白的一种或多种功能或特征的所有变异体都被设想为在本公开的范围内。在一些实施例中，修饰为参考CasX的一个或多个氨基酸中的突变。在其它实施例中，修饰为参考CasX的一个或多个结构域经来自不同CasX的一个或多个结构域取代。在一些实施例中，插入包括插入来自不同CasX蛋白的部分或所有结构域。突变可出现于参考CasX蛋白的任何一个或多个结构域中，

且可包括例如一个或多个结构域的一部分或全部的缺失,或参考CasX蛋白的任何结构域中的一个或多个氨基酸取代、缺失或插入。CasX蛋白的结构域包括非靶链结合(NTSB)结构域、靶链负载(TSL)结构域、螺旋形I结构域、螺旋形II结构域、寡核苷酸结合结构域(OBD)及RuvC DNA裂解结构域。将引起CasX蛋白的特征改进的参考CasX蛋白的任何氨基酸序列变化视为本公开的CasX变异蛋白。举例来说,相对于参考CasX蛋白序列,CasX变异体可包含一个或多个氨基酸取代、插入、缺失或交换结构域,或其任何组合。

[0228] 在一些实施例中,CasX变异蛋白在参考CasX蛋白的两个结构域的至少每一个中包含至少一个修饰,包括SEQ ID NO:1-3的序列。在一些实施例中,CasX变异蛋白包含参考CasX蛋白的至少2个结构域、至少3个结构域、至少4个结构域或至少5个结构域中的至少一个修饰。在一些实施例中,CasX变异蛋白包含参考CasX蛋白的至少一个结构域中的两个或更多个修饰。在一些实施例中,CasX变异蛋白包含参考CasX蛋白的至少一个结构域中的至少两个修饰、参考CasX蛋白的至少一个结构域中的至少三个修饰或参考CasX蛋白的至少一个结构域中的至少四个修饰。在一些实施例中,其中与参考CasX蛋白相比,CasX变异体包含两个或更多个修饰,每个修饰在独立地选自由NTSBD、TSLD、螺旋形I结构域、螺旋形II结构域、OBD和RuvC DNA裂解结构域组成的组中的结构域中进行。

[0229] 在一些实施例中,CasX变异蛋白的至少一个修饰包含参考CasX蛋白的一个结构域的至少一部分的缺失,包括SEQ ID NO:1-3的序列。在一些实施例中,缺失在NTSBD、TSLD、螺旋形I结构域、螺旋形II结构域、OBD或RuvC DNA裂解结构域中。

[0230] 适用于产生本公开的CasX变异蛋白的诱变方法可包括例如深度突变进化(DME)、深度突变扫描(DMS)、易错PCR、盒式诱变、随机诱变、交错延伸PCR、基因改组或结构域交换。在一些实施例中,CasX变异体例如通过选择参考CasX中的一个或多个所需突变而设计。在某些实施例中,参考CasX蛋白的活性係用作比较一种或多种CasX变异体的活性,由此测量CasX变异体的功能改进的基准。CasX变异体的示范性改进包括但不限于改进的变异体折叠、改进的对gNA的结合亲和力、改进的对靶DNA的结合亲和力、改变的对一个或多个PAM序列的结合亲和力、改进的靶DNA解旋、增加的活性、改进的编辑效率、改进的编辑特异性、增加的核酸酶活性、增加的用于双链裂解的靶链负载、减少的用于单链切割的靶链负载、减少的脱靶裂解、改进的DNA的非靶链的结合、改进的蛋白质稳定性、改进的蛋白质:gNA复合物稳定性、改进的蛋白质溶解度、改进的蛋白质:gNA复合物溶解度、改进的蛋白质产率、改进的蛋白质表达及改进的熔融特征,如下文更充分地描述。

[0231] 在本文所述的CasX变异体的一些实施例中,至少一个修饰包含:(a)与SEQ ID NO:1、SEQ ID NO:2或SEQ ID NO:3的参考CasX相比,CasX变异体中1至100个连续或非连续氨基酸的取代;(b)与参考CasX相比,CasX变异体中1至100个连续或非连续氨基酸的缺失;(c)与参考CasX相比,CasX中1至100个连续或非连续氨基酸的插入;或(d)(a)-(c)的任何组合。在一些实施例中,至少一个修饰包含:(a)与SEQ ID NO:1、SEQ ID NO:2或SEQ ID NO:3的参考CasX相比,CasX变异体中5至10个连续或非连续氨基酸的取代;(b)与参考CasX相比,CasX变异体中1至5个连续或非连续氨基酸的缺失;(c)与参考CasX相比,CasX中1至5个连续或非连续氨基酸的插入;或(d)(a)-(c)的任何组合。

[0232] 在一些实施例中,CasX变异蛋白包含相对于SEQ ID NO:1、SEQ ID NO:2或SEQ ID NO:3的序列具有至少1个、至少2个、至少3个、至少4个、至少5个、至少6个、至少7个、至少8

个、至少9个、至少10个、至少20个、至少30个、至少40个或至少50个突变的序列或由其组成。这些突变可为插入、缺失、氨基酸取代或其任何组合。

[0233] 在一些实施例中，CasX变异蛋白包含参考CasX蛋白的至少一个结构域中的至少一个氨基酸取代。在一些实施例中，相对于参考CasX蛋白，CasX变异蛋白包含至少约1-4个氨基酸取代、1-10个氨基酸取代、1-20个氨基酸取代、1-30个氨基酸取代、1-40个氨基酸取代、1-50个氨基酸取代、1-60个氨基酸取代、1-70个氨基酸取代、1-80个氨基酸取代、1-90个氨基酸取代、1-100个氨基酸取代、2-10个氨基酸取代、2-20个氨基酸取代、2-30个氨基酸取代、3-10个氨基酸取代、3-20个氨基酸取代、3-30个氨基酸取代、4-10个氨基酸取代、4-20个氨基酸取代、3-300个氨基酸取代、5-10个氨基酸取代、5-20个氨基酸取代、5-30个氨基酸取代、10-50个氨基酸取代或20-50个氨基酸取代，这些氨基酸取代可以是连续的或非连续的或位于不同结构域中。如本文所用，“连续氨基酸”是指在多肽的一级序列中连续的氨基酸。在一些实施例中，相对于参考CasX蛋白，CasX变异蛋白包含至少约100个或更多个氨基酸取代。在一些实施例中，氨基酸取代为保守取代。在其它实施例中，取代为非保守的；例如极性氨基酸取代非极性氨基酸，或反之亦然。

[0234] 任何氨基酸可在本文所述的取代中取代任何其它氨基酸。取代可为保守取代(例如碱性氨基酸取代另一碱性氨基酸)。取代可为非保守取代(例如碱性氨基酸取代酸性氨基酸，或反的亦然)。举例来说，参考CasX蛋白中的脯氨酸可取代以下中的任一个以产生本公开的CasX变异蛋白：精氨酸、组氨酸、氨酸赖氨酸、天冬氨酸、麸氨酸、丝氨酸、苏氨酸、天冬酰胺、氨酸谷氨酰胺、半胱氨酸、甘氨酸、丙氨酸、异亮氨酸、亮氨酸、甲硫氨酸、苯丙氨酸、色氨酸、酪氨酸或缬氨酸。

[0235] 在一些实施例中，CasX变异蛋白包含相对于参考CasX蛋白的至少一个氨基酸缺失。在一些实施例中，相对于参考CasX蛋白，CasX变异蛋白包含1-4个氨基酸、1-10个氨基酸、1-20个氨基酸、1-30个氨基酸、1-40个氨基酸、1-50个氨基酸、1-60个氨基酸、1-70个氨基酸、1-80个氨基酸、1-90个氨基酸、1-100个氨基酸、2-10个氨基酸、2-20个氨基酸、2-30个氨基酸、3-10个氨基酸、3-20个氨基酸、3-30个氨基酸、4-10个氨基酸、4-20个氨基酸、3-300个氨基酸、5-10个氨基酸、5-20个氨基酸、5-30个氨基酸、10-50个氨基酸或20-50个氨基酸的缺失。在一些实施例中，相对于参考CasX蛋白，CasX蛋白包含至少约100个连续氨基酸的缺失。在一些实施例中，相对于参考CasX蛋白，CasX变异蛋白包含至少1、2、3、4、5、6、7、8、9、10、20、30、40、50或100个连续氨基酸的缺失。在一些实施例中，CasX变异蛋白包含1、2、3、4、5、6、7、8、9或10个连续氨基酸的缺失。

[0236] 在一些实施例中，CasX变异蛋白包含相对于参考CasX蛋白的两个或更多个缺失，且所述两个或更多个缺失不为连续氨基酸。举例来说，第一缺失可在参考CasX蛋白的第一结构域中，且第二缺失可在参考CasX蛋白的第二结构域中。在一些实施例中，相对于参考CasX蛋白，CasX变异蛋白包含2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个非连续缺失。在一些实施例中，相对于参考CasX蛋白，CasX变异蛋白包含至少20个非连续缺失。各非连续缺失可具有本文所述的氨基酸的任何长度，例如1-4个氨基酸、1-10个氨基酸等。

[0237] 在一些实施例中，CasX变异蛋白包含相对于SEQ ID NO:1、2或3的序列的一个或多个氨基酸插入。在一些实施例中，相对于参考CasX蛋白，CasX变异蛋白包含1个氨基酸的插

入、2-3个连续氨基酸或非连续氨基酸、2-4个连续氨基酸或非连续氨基酸、2-5个连续氨基酸或非连续氨基酸、2-6个连续氨基酸或非连续氨基酸、2-7个连续氨基酸或非连续氨基酸、2-8个连续氨基酸或非连续氨基酸、2-9个连续氨基酸或非连续氨基酸、2-10个连续氨基酸或非连续氨基酸、2-20个连续氨基酸或非连续氨基酸、2-30个连续氨基酸或非连续氨基酸、2-40个连续氨基酸或非连续氨基酸、2-50个连续氨基酸或非连续氨基酸、2-60个连续氨基酸或非连续氨基酸、2-70个连续氨基酸或非连续氨基酸、2-80个连续氨基酸或非连续氨基酸、2-90个连续氨基酸或非连续氨基酸、2-100个连续氨基酸或非连续氨基酸、3-10个连续氨基酸或非连续氨基酸、3-20个连续氨基酸或非连续氨基酸、3-30个连续氨基酸或非连续氨基酸、4-10个连续氨基酸或非连续氨基酸、4-20个连续氨基酸或非连续氨基酸、3-300个连续氨基酸或非连续氨基酸、5-10个连续氨基酸或非连续氨基酸、5-20个连续氨基酸或非连续氨基酸、5-30个连续氨基酸或非连续氨基酸、10-50个连续氨基酸或非连续氨基酸或20-50个连续氨基酸或非连续氨基酸的插入。在一些实施例中，CasX变异蛋白包含2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个连续或非连续氨基酸的插入。在一些实施例中，CasX变异蛋白包含至少约100个连续或非连续氨基酸的插入。任何氨基酸或氨基酸的组合可以插入本文所述的插入物中以产生CasX变异蛋白。

[0238] 本文所述的取代、插入及缺失的实施例的任何排列可经组合以产生本公开的CasX变异蛋白。举例来说，CasX变异蛋白可包含相对于参考CasX蛋白序列的至少一个取代及至少一个缺失、相对于参考CasX蛋白序列的至少一个取代及至少一个插入、相对于参考CasX蛋白序列的至少一个插入及至少一个缺失或相对于参考CasX蛋白序列的至少一个取代、一个插入及一个缺失。

[0239] 在一些实施例中，CasX变异蛋白与SEQ ID NO:2或其部分具有至少约60％序列类似性。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的Y789T的取代、SEQ ID NO:2的P793的缺失、SEQ ID NO:2的Y789D的取代、SEQ ID NO:2的T72S的取代、SEQ ID NO:2的I546V的取代、SEQ ID NO:2的E552A的取代、SEQ ID NO:2的A636D的取代，SEQ ID NO:2的F536S的取代、SEQ ID NO:2的A708K的取代、SEQ ID NO:2的Y797L的取代、SEQ ID NO:2的L792G的取代、SEQ ID NO:2的A739V的取代、SEQ ID NO:2的G791M的取代、在SEQ ID NO:2的位置661处的插入A、SEQ ID NO:2的A788W的取代、SEQ ID NO:2的K390R的取代、SEQ ID NO:2的A751S的取代、SEQ ID NO:2的E385A的取代、在SEQ ID NO:2的位置696处的插入P、在SEQ ID NO:2的位置773处的插入M、SEQ ID NO:2的G695H的取代、在SEQ ID NO:2的位置793处的插入AS、在SEQ ID NO:2的位置795处的插入AS、SEQ ID NO:2的C477R的取代、SEQ ID NO:2的C477K的取代、SEQ ID NO:2的C479A的取代、SEQ ID NO:2的C479L的取代、SEQ ID NO:2的I55F的取代、SEQ ID NO:2的K210R的取代、SEQ ID NO:2的C233S的取代、SEQ ID NO:2的D231N的取代、SEQ ID NO:2的Q338E的取代、SEQ ID NO:2的Q338R的取代、SEQ ID NO:2的L379R的取代、SEQ ID NO:2的K390R的取代、SEQ ID NO:2的L481Q的取代、SEQ ID NO:2的F495S的取代、SEQ ID NO:2的D600N的取代、SEQ ID NO:2的T886K的取代、SEQ ID NO:2的A739V的取代、SEQ ID NO:2的K460N的取代、SEQ ID NO:2的I199F的取代、SEQ ID NO:2的G492P的取代、SEQ ID NO:2的T153I的取代、SEQ ID NO:2的R591I的取代、在SEQ ID NO:2的位置795处的插入AS、在SEQ ID NO:2的位置796处的插入AS、在SEQ ID NO:2的位置889处的插入L、SEQ ID NO:2的E121D的取代、SEQ ID NO:2的S270W的取代、SEQ ID NO:2的E712Q的

取代、SEQ ID NO:2的K942Q的取代、SEQ ID NO:2的E552K的取代、SEQ ID NO:2的K25Q的取代、SEQ ID NO:2的N47D的取代、在SEQ ID NO:2的位置696处的插入T、SEQ ID NO:2的L685I的取代、SEQ ID NO:2的N880D的取代、SEQ ID NO:2的Q102R的取代、SEQ ID NO:2的M734K的取代、SEQ ID NO:2的A724S的取代、SEQ ID NO:2的T704K的取代、SEQ ID NO:2的P224K的取代、SEQ ID NO:2的K25R的取代、SEQ ID NO:2的M29E的取代、SEQ ID NO:2的H152D的取代、SEQ ID NO:2的S219R的取代、SEQ ID NO:2的E475K的取代、SEQ ID NO:2的G226R的取代、SEQ ID NO:2的A377K的取代、SEQ ID NO:2的E480K的取代、SEQ ID NO:2的K416E的取代、SEQ ID NO:2的H164R的取代、SEQ ID NO:2的K767R的取代、SEQ ID NO:2的I7F的取代、SEQ ID NO:2的M29R的取代、SEQ ID NO:2的H435R的取代、SEQ ID NO:2的E385Q的取代、SEQ ID NO:2的E385K的取代、SEQ ID NO:2的I279F的取代、SEQ ID NO:2的D489S的取代、SEQ ID NO:2的D732N的取代、SEQ ID NO:2的A739T的取代、SEQ ID NO:2的W885R的取代、SEQ ID NO:2的E53K的取代、SEQ ID NO:2的A238T的取代、SEQ ID NO:2的P283Q的取代、SEQ ID NO:2的E292K的取代、SEQ ID NO:2的Q628E的取代、SEQ ID NO:2的R388Q的取代、SEQ ID NO:2的G791M的取代、SEQ ID NO:2的L792K的取代、SEQ ID NO:2的L792E的取代、SEQ ID NO:2的M779N的取代、SEQ ID NO:2的G27D的取代、SEQ ID NO:2的K955R的取代、SEQ ID NO:2的S867R的取代、SEQ ID NO:2的R693I的取代、SEQ ID NO:2的F189Y的取代、SEQ ID NO:2的V635M的取代、SEQ ID NO:2的F399L的取代、SEQ ID NO:2的E498K的取代、SEQ ID NO:2的E386R的取代、SEQ ID NO:2的V254G的取代、SEQ ID NO:2的P793S的取代、SEQ ID NO:2的K188E的取代、SEQ ID NO:2的QT945KI的取代、SEQ ID NO:2的T620P的取代、SEQ ID NO:2的T946P的取代、SEQ ID NO:2的TT949PP的取代、SEQ ID NO:2的N952T的取代、SEQ ID NO:2的K682E的取代、SEQ ID NO:2的K975R的取代、SEQ ID NO:2的L212P的取代、SEQ ID NO:2的E292R的取代、SEQ ID NO:2的I303K的取代、SEQ ID NO:2的C349E的取代、SEQ ID NO:2的E385P的取代、SEQ ID NO:2的E386N的取代、SEQ ID NO:2的D387K的取代、SEQ ID NO:2的L404K的取代、SEQ ID NO:2的E466H的取代、SEQ ID NO:2的C477Q的取代、SEQ ID NO:2的C477H的取代、SEQ ID NO:2的C479A的取代、SEQ ID NO:2的D659H的取代、SEQ ID NO:2的T806V的取代、SEQ ID NO:2的K808S的取代、在SEQ ID NO:2的位置797处的插入AS、SEQ ID NO:2的V959M的取代、SEQ ID NO:2的K975Q的取代、SEQ ID NO:2的W974G的取代、SEQ ID NO:2的A708Q的取代、SEQ ID NO:2的V711K的取代、SEQ ID NO:2的D733T的取代、SEQ ID NO:2的L742W的取代、SEQ ID NO:2的V747K的取代、SEQ ID NO:2的F755M的取代、SEQ ID NO:2的M771A的取代、SEQ ID NO:2的M771Q的取代、SEQ ID NO:2的W782Q的取代、SEQ ID NO:2的G791F的取代、SEQ ID NO:2的L792D的取代、SEQ ID NO:2的L792K的取代、SEQ ID NO:2的P793Q的取代、SEQ ID NO:2的P793G的取代、SEQ ID NO:2的Q804A的取代、SEQ ID NO:2的Y966N的取代、SEQ ID NO:2的Y723N的取代、SEQ ID NO:2的Y857R的取代、SEQ ID NO:2的S890R的取代、SEQ ID NO:2的S932M的取代、SEQ ID NO:2的L897M的取代、SEQ ID NO:2的R624G的取代、SEQ ID NO:2的S603G的取代、SEQ ID NO:2的N737S的取代、SEQ ID NO:2的L307K的取代、SEQ ID NO:2的I658V的取代、在SEQ ID NO:2的位置688处的插入PT、在SEQ ID NO:2的位置794处的插入SA、SEQ ID NO:2的S877R的取代、SEQ ID NO:2的N580T的取代、SEQ ID NO:2的V335G的取代、SEQ ID NO:2的T620S的取代、SEQ ID NO:2的W345G的取代、SEQ ID NO:2的T280S的取代、SEQ ID NO:2的L406P的取代、SEQ ID NO:2的A612D的取

代、SEQ ID NO:2的A751S的取代、SEQ ID NO:2的E386R的取代、SEQ ID NO:2的V351M的取代、SEQ ID NO:2的K210N的取代、SEQ ID NO:2的D40A的取代、SEQ ID NO:2的E773G的取代、SEQ ID NO:2的H207L的取代、SEQ ID NO:2的T62A的取代、SEQ ID NO:2的T287P的取代、SEQ ID NO:2的T832A的取代、SEQ ID NO:2的A893S的取代、在SEQ ID NO:2的位置14处的插入V、在SEQ ID NO:2的位置13处的插入AG、SEQ ID NO:2的R11V的取代、SEQ ID NO:2的R12N的取代、SEQ ID NO:2的R13H的取代、在SEQ ID NO:2的位置13处的插入Y、SEQ ID NO:2的R12L的取代、在SEQ ID NO:2的位置13处的插入Q、SEQ ID NO:2的V15S的取代、在SEQ ID NO:2的位置17处的插入D或其组合。

[0240] 在一些实施例中,CasX变异体包含NTSB结构域中的至少一个修饰。

[0241] 在一些实施例中,CasX变异体包含TSL结构域中的至少一个修饰。在一些实施例中,TSL结构域中的至少一个修饰包含SEQ ID NO:2的氨基酸Y857、S890或S932中的一个或多个的氨基酸取代。

[0242] 在一些实施例中,CasX变异体包含螺旋形I结构域中的至少一个修饰。在一些实施例中,螺旋形I结构域中的至少一个修饰包含SEQ ID NO:2的氨基酸S219、L249、E259、Q252、E292、L307或D318中的一个或多个的氨基酸取代。

[0243] 在一些实施例中,CasX变异体包含螺旋形II结构域中的至少一个修饰。在一些实施例中,螺旋形II结构域中的至少一个修饰包含SEQ ID NO:2的氨基酸D361、L379、E385、E386、D387、F399、L404、R458、C477或D489中的一个或多个的氨基酸取代。

[0244] 在一些实施例中,CasX变异体包含OBD结构域中的至少一个修饰。在一些实施例中,OBD中的至少一个修饰包含SEQ ID NO:2的氨基酸F536、E552、T620或I658中的一个或多个的氨基酸取代。

[0245] 在一些实施例中,CasX变异体包含RuvC DNA裂解结构域中的至少一个修饰。在一些实施例中,RuvC DNA裂解结构域中的至少一个修饰包含SEQ ID NO:2的氨基酸K682、G695、A708、V711、D732、A739、D733、L742、V747、F755、M771、M779、W782、A788、G791、L792、P793、Y797、M799、Q804、S819或Y857中的一个或多个的氨基酸取代或氨基酸P793的缺失。

[0246] 在一些实施例中,与SEQ ID NO:2的参考CasX序列相比,CasX变异体包含选自以下中的一个或多个的至少一个修饰:(a)L379R的氨基酸取代;(b)A708K的氨基酸取代;(c)T620P的氨基酸取代;(d)E385P的氨基酸取代;(e)Y857R的氨基酸取代;(f)I658V的氨基酸取代;(g)F399L的氨基酸取代;(h)Q252K的氨基酸取代;(i)L404K的氨基酸取代;和(j)P793的氨基酸缺失。

[0247] 在一些实施例中,CasX变异体包含选自由以下组成的组的参考CasX变异蛋白质的序列的至少两个氨基酸变化:SEQ ID NO:2的Y789T的取代、SEQ ID NO:2的P793的缺失、SEQ ID NO:2的Y789D的取代、SEQ ID NO:2的T72S的取代、SEQ ID NO:2的I546V的取代、SEQ ID NO:2的E552A的取代、SEQ ID NO:2的A636D的取代,SEQ ID NO:2的F536S的取代、SEQ ID NO:2的A708K的取代、SEQ ID NO:2的Y797L的取代、SEQ ID NO:2的L792G的取代、SEQ ID NO:2的A739V的取代、SEQ ID NO:2的G791M的取代、在SEQ ID NO:2的位置661处的插入A、SEQ ID NO:2的A788W的取代、SEQ ID NO:2的K390R的取代、SEQ ID NO:2的A751S的取代、SEQ ID NO:2的E385A的取代、在SEQ ID NO:2的位置696处的插入P、在SEQ ID NO:2的位置773处的插入M、SEQ ID NO:2的G695H的取代、在SEQ ID NO:2的位置793处的插入AS、在SEQ

ID NO:2的位置795处的插入AS、SEQ ID NO:2的C477R的取代、SEQ ID NO:2的C477K的取代、SEQ ID NO:2的C479A的取代、SEQ ID NO:2的C479L的取代、SEQ ID NO:2的I55F的取代、SEQ ID NO:2的K210R的取代、SEQ ID NO:2的C233S的取代、SEQ ID NO:2的D231N的取代、SEQ ID NO:2的Q338E的取代、SEQ ID NO:2的Q338R的取代、SEQ ID NO:2的L379R的取代、SEQ ID NO:2的K390R的取代、SEQ ID NO:2的L481Q的取代、SEQ ID NO:2的F495S的取代、SEQ ID NO:2的D600N的取代、SEQ ID NO:2的T886K的取代、SEQ ID NO:2的A739V的取代、SEQ ID NO:2的K460N的取代、SEQ ID NO:2的I199F的取代、SEQ ID NO:2的G492P的取代、SEQ ID NO:2的T153I的取代、SEQ ID NO:2的R591I的取代、在SEQ ID NO:2的位置795处的插入AS、在SEQ ID NO:2的位置796处的插入AS、在SEQ ID NO:2的位置889处的插入L、SEQ ID NO:2的E121D的取代、SEQ ID NO:2的S270W的取代、SEQ ID NO:2的E712Q的取代、SEQ ID NO:2的K942Q的取代、SEQ ID NO:2的E552K的取代、SEQ ID NO:2的K25Q的取代、SEQ ID NO:2的N47D的取代、在SEQ ID NO:2的位置696处的插入T、SEQ ID NO:2的L685I的取代、SEQ ID NO:2的N880D的取代、SEQ ID NO:2的Q102R的取代、SEQ ID NO:2的M734K的取代、SEQ ID NO:2的A724S的取代、SEQ ID NO:2的T704K的取代、SEQ ID NO:2的P224K的取代、SEQ ID NO:2的K25R的取代、SEQ ID NO:2的M29E的取代、SEQ ID NO:2的H152D的取代、SEQ ID NO:2的S219R的取代、SEQ ID NO:2的E475K的取代、SEQ ID NO:2的G226R的取代、SEQ ID NO:2的A377K的取代、SEQ ID NO:2的E480K的取代、SEQ ID NO:2的K416E的取代、SEQ ID NO:2的H164R的取代、SEQ ID NO:2的K767R的取代、SEQ ID NO:2的I7F的取代、SEQ ID NO:2的M29R的取代、SEQ ID NO:2的H435R的取代、SEQ ID NO:2的E385Q的取代、SEQ ID NO:2的E385K的取代、SEQ ID NO:2的I279F的取代、SEQ ID NO:2的D489S的取代、SEQ ID NO:2的D732N的取代、SEQ ID NO:2的A739T的取代、SEQ ID NO:2的W885R的取代、SEQ ID NO:2的E53K的取代、SEQ ID NO:2的A238T的取代、SEQ ID NO:2的P283Q的取代、SEQ ID NO:2的E292K的取代、SEQ ID NO:2的Q628E的取代、SEQ ID NO:2的R388Q的取代、SEQ ID NO:2的G791M的取代、SEQ ID NO:2的L792K的取代、SEQ ID NO:2的L792E的取代、SEQ ID NO:2的M779N的取代、SEQ ID NO:2的G27D的取代、SEQ ID NO:2的K955R的取代、SEQ ID NO:2的S867R的取代、SEQ ID NO:2的R693I的取代、SEQ ID NO:2的F189Y的取代、SEQ ID NO:2的V635M的取代、SEQ ID NO:2的F399L的取代、SEQ ID NO:2的E498K的取代、SEQ ID NO:2的E386R的取代、SEQ ID NO:2的V254G的取代、SEQ ID NO:2的P793S的取代、SEQ ID NO:2的K188E的取代、SEQ ID NO:2的QT945KI的取代、SEQ ID NO:2的T620P的取代、SEQ ID NO:2的T946P的取代、SEQ ID NO:2的TT949PP的取代、SEQ ID NO:2的N952T的取代、SEQ ID NO:2的K682E的取代、SEQ ID NO:2的K975R的取代、SEQ ID NO:2的L212P的取代、SEQ ID NO:2的E292R的取代、SEQ ID NO:2的I303K的取代、SEQ ID NO:2的C349E的取代、SEQ ID NO:2的E385P的取代、SEQ ID NO:2的E386N的取代、SEQ ID NO:2的D387K的取代、SEQ ID NO:2的L404K的取代、SEQ ID NO:2的E466H的取代、SEQ ID NO:2的C477Q的取代、SEQ ID NO:2的C477H的取代、SEQ ID NO:2的C479A的取代、SEQ ID NO:2的D659H的取代、SEQ ID NO:2的T806V的取代、SEQ ID NO:2的K808S的取代、在SEQ ID NO:2的位置797处的插入AS、SEQ ID NO:2的V959M的取代、SEQ ID NO:2的K975Q的取代、SEQ ID NO:2的W974G的取代、SEQ ID NO:2的A708Q的取代、SEQ ID NO:2的V711K的取代、SEQ ID NO:2的D733T的取代、SEQ ID NO:2的L742W的取代、SEQ ID NO:2的V747K的取代、SEQ ID NO:2的F755M的取代、SEQ ID NO:2的M771A的取代、

SEQ ID NO:2的M771Q的取代、SEQ ID NO:2的W782Q的取代、SEQ ID NO:2的G791F的取代、SEQ ID NO:2的L792D的取代、SEQ ID NO:2的L792K的取代、SEQ ID NO:2的P793Q的取代、SEQ ID NO:2的P793G的取代、SEQ ID NO:2的Q804A的取代、SEQ ID NO:2的Y966N的取代、SEQ ID NO:2的Y723N的取代、SEQ ID NO:2的Y857R的取代、SEQ ID NO:2的S890R的取代、SEQ ID NO:2的S932M的取代、SEQ ID NO:2的L897M的取代、SEQ ID NO:2的R624G的取代、SEQ ID NO:2的S603G的取代、SEQ ID NO:2的N737S的取代、SEQ ID NO:2的L307K的取代、SEQ ID NO:2的I658V的取代、在SEQ ID NO:2的位置688处的插入PT、在SEQ ID NO:2的位置794处的插入SA、SEQ ID NO:2的S877R的取代、SEQ ID NO:2的N580T的取代、SEQ ID NO:2的V335G的取代、SEQ ID NO:2的T620S的取代、SEQ ID NO:2的W345G的取代、SEQ ID NO:2的T280S的取代、SEQ ID NO:2的L406P的取代、SEQ ID NO:2的A612D的取代、SEQ ID NO:2的A751S的取代、SEQ ID NO:2的E386R的取代、SEQ ID NO:2的V351M的取代、SEQ ID NO:2的K210N的取代、SEQ ID NO:2的D40A的取代、SEQ ID NO:2的E773G的取代、SEQ ID NO:2的H207L的取代、SEQ ID NO:2的T62A的取代、SEQ ID NO:2的T287P的取代、SEQ ID NO:2的T832A的取代、SEQ ID NO:2的A893S的取代、在SEQ ID NO:2的位置14处的插入V、在SEQ ID NO:2的位置13处的插入AG、SEQ ID NO:2的R11V的取代、SEQ ID NO:2的R12N的取代、SEQ ID NO:2的R13H的取代、在SEQ ID NO:2的位置13处的插入Y、SEQ ID NO:2的R12L的取代、在SEQ ID NO:2的位置13处的插入Q、SEQ ID NO:2的V15S的取代、和在SEQ ID NO:2的位置17处的插入D。在一些实施例中，参考CasX蛋白的至少两个氨基酸变化选自表4中所阐述的SEQ ID NO:49到150的序列中所公开的氨基酸变化。在一些实施例中，CasX变异体包含此段之前述实施例的任何组合。

[0248] 在一些实施例中，CasX变异蛋白包含对参考CasX蛋白氨基酸序列的超过一个取代、插入和/或缺失。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的S794R的取代和Y797L的取代。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的K416E的取代和A708K的取代。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的A708K的取代和P793的缺失。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的P793的缺失和在位置795处的插入AS。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的Q367K的取代和I425S的取代。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的A708K的取代、位置793处P的缺失和A793V的取代。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的Q338R的取代和A339E的取代。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的Q338R的取代和A339K的取代。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的S507G的取代和G508R的取代。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的L379R的取代、A708K的取代和位置793处P的缺失。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的C477K的取代、A708K的取代和位置793处P的缺失。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的L379R的取代、C477K的取代、A708K的取代和位置793处P的缺失。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的L379R的取代、A708K的取代、位置793处P的缺失和A739V的取代。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的C477K的取代、A708K的取代、位置793处P的缺失和A739V的取代。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的L379R的取代、C477K的取代、A708K的取代、位置793处P的缺失和A739V的取代。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的L379R的取代、A708K的取代、位置793处P的缺失和M779N的取代。在一些实施例中，CasX变异蛋白包含SEQ

ID NO:2的L379R的取代、A708K的取代、位置793处P的缺失和M771N的取代。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的L379R的取代、708K的取代、位置793处P的缺失和D489S的取代。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的L379R的取代、A708K的取代、位置793处P的缺失和A739T的取代。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的L379R的取代、A708K的取代、位置793处P的缺失和D732N的取代。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的L379R的取代、A708K的取代、位置793处P的缺失和G791M的取代。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的L379R的取代、708K的取代、位置793处P的缺失和Y797L的取代。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的L379R的取代、C477K的取代、A708K的取代、位置793处P的缺失和M779N的取代。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的L379R的取代、C477K的取代、A708K的取代、位置793处P的缺失和M771N的取代。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的L379R的取代、C477K的取代、A708K的取代、位置793处P的缺失和D489S的取代。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的L379R的取代、C477K的取代、A708K的取代、位置793处P的缺失和A739T的取代。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的L379R的取代、C477K的取代、A708K的取代、位置793处P的缺失和D732N的取代。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的L379R的取代、C477K的取代、A708K的取代、位置793处P的缺失和G791M的取代。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的L379R的取代、C477K的取代、A708K的取代、位置793处P的缺失和Y797L的取代。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的L379R的取代、C477K的取代、A708K的取代、位置793处P的缺失和T620P的取代。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的A708K的取代、位置793处P的缺失和E386S的取代。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的E386R的取代、F399L的取代和位置793处P的缺失。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的R581I和A739V的取代。在一些实施例中，CasX变异体包含此段之前述实施例的任何组合。

[0249] 在一些实施例中，CasX变异蛋白包含对参考CasX蛋白氨基酸序列的超过一个取代、插入和/或缺失。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的A708K的取代、位置793处P的缺失和A739V的取代。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的L379R的取代、A708K的取代和位置793处P的缺失。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的C477K的取代、A708K的取代和位置793处P的缺失。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的L379R的取代、C477K的取代、A708K的取代和位置793处P的缺失。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的L379R的取代、A708K的取代、位置793处P的缺失和A739V的取代。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的C477K的取代、A708K的取代、位置793处P的缺失和A739的取代。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的L379R的取代、C477K的取代、A708K的取代、位置793处P的缺失和A739V的取代。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的L379R的取代、C477K的取代、A708K的取代、位置793处P的缺失和T620P的取代。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的M771A的取代。在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的L379R的取代、A708K的取代、位置793处P的缺失和D732N的取代。在一些实施例中，CasX变异体包含此段之前述实施例的任何组合。

[0250] 在一些实施例中，CasX变异蛋白包含SEQ ID NO:2的W782Q的取代。在一些实施例

中,CasX变异蛋白包含SEQ ID NO:2的M771Q的取代。在一些实施例中,CasX变异蛋白包含SEQ ID NO:2的R458I的取代和A739V的取代。在一些实施例中,CasX变异蛋白包含SEQ ID NO:2的L379R的取代、A708K的取代、位置793处P的缺失和M771N的取代。在一些实施例中,CasX变异蛋白包含SEQ ID NO:2的L379R的取代、A708K的取代、位置793处P的缺失和A739T的取代。在一些实施例中,CasX变异蛋白包含SEQ ID NO:2的L379R的取代、C477K的取代、A708K的取代、位置793处P的缺失和D489S的取代。在一些实施例中,CasX变异蛋白包含SEQ ID NO:2的L379R的取代、C477K的取代、A708K的取代、位置793处P的缺失和D732N的取代。在一些实施例中,CasX变异蛋白包含SEQ ID NO:2的V711K的取代。在一些实施例中,CasX变异蛋白包含SEQ ID NO:2的L379R的取代、C477K的取代、A708K的取代、位置793处P的缺失和Y797L的取代。在一些实施例中,CasX变异蛋白包含SEQ ID NO:2的L379R的取代、A708K的取代和位置793处P的缺失。在一些实施例中,CasX变异蛋白包含SEQ ID NO:2的L379R的取代、C477K的取代、A708K的取代、位置793处P的缺失和M771N的取代。在一些实施例中,CasX变异蛋白包含SEQ ID NO:2的A708K的取代、位置793处P的取代和E386S的取代。在一些实施例中,CasX变异蛋白包含SEQ ID NO:2的L379R的取代、C477K的取代、A708K的取代和位置793处P的缺失。在一些实施例中,CasX变异蛋白包含SEQ ID NO:2的L792D的取代。在一些实施例中,CasX变异蛋白包含SEQ ID NO:2的G791F的取代。在一些实施例中,CasX变异蛋白包含SEQ ID NO:2的A708K的取代、位置793处P的缺失和A739V的取代。在一些实施例中,CasX变异蛋白包含SEQ ID NO:2的L379R的取代、A708K的取代、位置793处P的缺失和A739V的取代。在一些实施例中,CasX变异蛋白包含SEQ ID NO:2的C477K的取代、A708K的取代和位置793处P的取代。在一些实施例中,CasX变异蛋白包含SEQ ID NO:2的L249I的取代和M771N的取代。在一些实施例中,CasX变异蛋白包含SEQ ID NO:2的V747K的取代。在一些实施例中,CasX变异蛋白包含SEQ ID NO:2的L379R的取代、C477的取代、A708K的取代、位置793处P的缺失和M779N的取代。在一些实施例中,CasX变异蛋白包含F755M的取代。在一些实施例中,CasX变异体包含此段之前述实施例的任何组合。

[0251] 在一些实施例中,CasX变异蛋白与SEQ ID NO:2的参考CasX序列相比包含至少一个修饰,其中至少一个修饰选自以下中的一个或多个:L379R的氨基酸取代;A708K的氨基酸取代;T620P的氨基酸取代;E385P的氨基酸取代;Y857R的氨基酸取代;I658V的氨基酸取代;F399L的氨基酸取代;Q252K的氨基酸取代;以及[P793]的氨基酸缺失。在一些实施例中,CasX变异蛋白与SEQ ID NO:2的参考CasX序列相比包含至少一个修饰,其中至少一个修饰选自以下中的一个或多个:L379R的氨基酸取代;A708K的氨基酸取代;T620P的氨基酸取代;E385P的氨基酸取代;Y857R的氨基酸取代;I658V的氨基酸取代;F399L的氨基酸取代;Q252K的氨基酸取代;L404K的氨基酸取代;以及[P793]的氨基酸缺失。在其它实施例中,CasX变异蛋白与SEQ ID NO:2的参考CasX序列相比包含前述取代或缺失的任何组合。在其它实施例中,除了前述取代或缺失以外,CasX变异蛋白可进一步包含来自SEQ ID NO:1的参考CasX的NTSB和/或螺旋形1b结构域的取代。

[0252] 在一些实施例中,CasX变异蛋白包含400至2000个氨基酸、500至1500个氨基酸、700至1200个氨基酸、800至1100个氨基酸或900至1000个氨基酸。

[0253] 在一些实施例中,CasX变异蛋白在形成发生gNA:靶DNA复合的通道的非连续残基区域中包含一个或多个修饰。在一些实施例中,CasX变异蛋白包含一个或多个修饰,其包含

形成与gNA结合的界面的非连续残基区域。举例来说,在参考CasX蛋白的一些实施例中,螺旋形I、螺旋形II及OBD结构域全部接触或邻近gNA:靶DNA复合物,且对此些结构域中的任一个内的非连续残基的一个或多个修饰可改进CasX变异蛋白的功能。

[0254] 在一些实施例中,CasX变异蛋白在形成与非靶链DNA结合的通道的非连续残基区域中包含一个或多个修饰。举例来说,CasX变异蛋白可包含对NTSBD的非连续残基的一个或多个修饰。在一些实施例中,CasX变异蛋白在形成与PAM结合的界面的非连续残基区域中包含一个或多个修饰。举例来说,CasX变异蛋白可包含对螺旋形I结构域或OBD的非连续残基的一个或多个修饰。在一些实施例中,CasX变异蛋白含有包含非连续表面暴露残基区域的一个或多个修饰。如本文所用,“表面暴露残基”是指CasX蛋白的表面上的氨基酸,或其中氨基酸的至少一部分,例如主链或一部分侧链在蛋白质的表面上的氨基酸。例如CasX的细胞蛋白质的表面暴露残基(其暴露于水性细胞内环境)经常选自带正电亲水性氨基酸,例如精氨酸、天冬酰胺、天冬氨酸、谷氨酰胺、谷氨酸、组氨酸、赖氨酸、丝氨酸及苏氨酸。因此,举例来说,在本文提供的变异体的一些实施例中,相比于参考CasX蛋白,表面暴露残基的区包含一个或多个插入、缺失或取代。在一些实施例中,一个或多个带正电残基取代一个或多个其它带正电残基,或带负电残基,或不带电残基,或其任何组合。在一些实施例中,一个或多个取代氨基酸残基接近结合核酸,例如RuvC结构域或螺旋形I结构域中接触靶DNA的残基,或OBD或螺旋形II结构域中结合gNA的残基可取代一个或多个带正电或极性氨基酸。

[0255] 在一些实施例中,CasX变异蛋白包含非连续残基区域中的一个或多个修饰,所述非连续残基经由参考CasX蛋白的结构域中的疏水性填充形成核。不希望受任何理论束缚,经由疏水性填充形成核心的区域富含疏水性氨基酸,例如缬氨酸、异亮氨酸、亮氨酸、甲硫氨酸、苯丙氨酸、色氨酸及半胱氨酸。举例来说,在一些参考CasX蛋白中,RuvC结构域包含邻近于活性位点的疏水袋。在一些实施例中,所述区域的2至15个残基为带电、极性或碱基堆叠的。带电氨基酸(有时在本文中称为残基)可包括例如精氨酸、赖氨酸、天冬氨酸和谷氨酸,且这些氨基酸的侧链可形成盐桥,前提是亦存在桥连伴侣(bridge partner)。极性氨基酸可包括例如氨酸谷氨酰胺、天冬酰胺、组氨酸、丝氨酸、苏氨酸、酪氨酸及半胱氨酸。在一些实施例中,极性氨基酸可取决于其侧链标识而形成质子供体或受体形式的氢键。如本文所用,“碱基堆叠”包括氨基酸残基(例如色氨酸、酪氨酸、苯丙氨酸或组氨酸)的芳香族侧链与核酸中的堆叠核苷酸碱基的相互作用。在空间上紧邻以形成CasX变异蛋白的功能性部分的对非连续氨基酸区的任何修饰被设想为在本公开的范围内。

[0256] i.具有来自多种源蛋白质的结构域的CasX变异蛋白

[0257] 在某些实施例中,本公开提供了嵌合CasX蛋白,其包含来自两种或更多种不同CasX蛋白的蛋白结构域,如两种或更多种参考CasX蛋白,或两种或更多种如本文所述的CasX变异蛋白序列。如本文所用,“嵌合CasX蛋白”是指含有至少两个分离或衍生自不同来源,例如两种天然存在的蛋白质的结构域的CasX,在一些实施例中,所述两种蛋白质可分离自不同物种。举例来说,在一些实施例中,嵌合CasX蛋白包含来自第一CasX蛋白的第一结构域及来自不同的第二CasX蛋白的第二结构域。在一些实施例中,第一结构域可选自由以下组成的组:NTSB、TSL、螺旋形I、螺旋形II、OBD和RuvC结构域。在一些实施例中,第二结构域选自由以下组成的组:NTSB、TSL、螺旋形I、螺旋形II、OBD和RuvC结构域,其中第二结构域不同于前述第一结构域。例如,嵌合CasX蛋白可包含来自SEQ ID NO:2的CasX蛋白的NTSB、

TSL、螺旋形I、螺旋形II、OBD结构域，以及来自SEQ ID NO:1的CasX蛋白的RuvC结构域，或反之亦然。作为另一实例，嵌合CasX蛋白可包含来自SEQ ID NO:2的CasX蛋白的NTSB、TSL、螺旋形II、OBD和RuvC结构域，以及来自SEQ ID NO:1的CasX蛋白的螺旋形I结构域，或反之亦然。因此，在某些实施例中，嵌合CasX蛋白可包含来自第一CasX蛋白的NTSB、TSL、螺旋形II、OBD和RuvC结构域，以及来自第二CasX蛋白的螺旋形I结构域。在嵌合CasX蛋白的一些实施例中，第一CasX蛋白的结构域衍生自SEQ ID NO:1、SEQ ID NO:2或SEQ ID NO:3的序列，且第二CasX蛋白的结构域衍生自SEQ ID NO:1、SEQ ID NO:2或SEQ ID NO:3的序列，并且第一CasX蛋白和第二CasX蛋白不相同。在一些实施例中，第一CasX蛋白的结构域包含衍生自SEQ ID NO:1的序列，且第二CasX蛋白的结构域包含衍生自SEQ ID NO:2的序列。在一些实施例中，第一CasX蛋白的结构域包含衍生自SEQ ID NO:1的序列，且第二CasX蛋白的结构域包含衍生自SEQ ID NO:3的序列。在一些实施例中，第一CasX蛋白的结构域包含衍生自SEQ ID NO:2的序列，且第二CasX蛋白的结构域包含衍生自SEQ ID NO:3的序列。在一些实施例中，CasX变异体包含SEQ ID NO:130到138或141到144，其序列阐述于表4中。在一些实施例中，CasX变异体包含SEQ ID NO:72、94、113、135、138、144、239、277或280的序列。在一些实施例中，CasX变异体包含SEQ ID NO:94、72、138、144或280的序列。在一些实施例中，CasX变异蛋白包含至少一个嵌合结构域，其包含来自第一CasX蛋白的第一部分及来自不同的第二CasX蛋白的第二部分。如本文所用，“嵌合结构域”是指含有至少两个分离或衍生自不同来源，例如两种天然存在的蛋白质的部分的结构域，或来自两种参考CasX蛋白的结构域部分。至少一个嵌合结构域可为如本文所述的NTSB、TSL、螺旋形I、螺旋形II、OBD或RuvC结构域中的任一个。在一些实施例中，CasX结构域的第一部分包含SEQ ID NO:1的序列，且CasX结构域的第二部分包含SEQ ID NO:2的序列。在一些实施例中，CasX结构域的第一部分包含SEQ ID NO:1的序列，且CasX结构域的第二部分包含SEQ ID NO:3的序列。在一些实施例中，CasX结构域的第一部分包含SEQ ID NO:2的序列，且CasX结构域的第二部分包含SEQ ID NO:3的序列。在一些实施例中，至少一个嵌合结构域包含嵌合RuvC结构域。作为前述的实例，嵌合RuvC结构域包含SEQ ID NO:1的氨基酸661至824和SEQ ID NO:2的氨基酸922至978。作为前述的替代实例，嵌合RuvC结构域包含SEQ ID NO:2的氨基酸648至812和SEQ ID NO:1的氨基酸935至986。在一些实施例中，CasX蛋白包含来自第一CasX蛋白的第一结构域及来自第二CasX蛋白的第二结构域，及至少一个嵌合结构域，其包含使用此段中所述的实施例的方法自不同CasX蛋白分离的至少两个部分。在前述实施例中，具有衍生自SEQ ID NO:1、2和3的结构域或结构域部分的嵌合CasX蛋白可进一步包含本文公开的任一实施例的氨基酸插入、缺失或取代。

[0258] 在一些实施例中，CasX变异蛋白质包含表4、6、7、8或10中阐述的序列。在一些实施例中，CasX变异蛋白由表4中阐述的序列组成。在其它实施例中，CasX变异蛋白包含与表4、6、7、8或10中所阐述的SEQ ID NO:49-150、233-235、238-252或272-281的序列具有至少60％同一性、至少65％同一性、至少70％同一性、至少75％同一性、至少80％同一性、至少81％同一性、至少82％同一性、至少83％同一性、至少84％同一性、至少85％同一性、至少86％同一性、至少86％同一性、至少87％同一性、至少88％同一性、至少89％同一性、至少89％同一性、至少90％同一性、至少91％同一性、至少92％同一性、至少93％同一性、至少94％同一性、至少95％同一性、至少96％同一性、至少97％同一性、至少98％同一性、至少

99%同一性、至少99.5%同一性的序列。在其它实施例中，CasX变异蛋白质包含如表4中所阐述的SEQ ID NO:49到150的序列，并且进一步包含本文中所公开的在N端、C端或这两者处或附近的一个或多个NLS。应了解，在一些情况下，表中的CasX变异体的N端甲硫氨酸在翻译后修饰期间自经表达CasX变异体去除。

[0259] 表4:CasX变异体序列

[0260]

描述*	氨基酸序列
来自 SEQ ID NO: 2 的 TSL、螺旋形 I、螺旋形 II、OBD 和 RuvC 结构域，以及来	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPASK KIDQNKLKPEMDEKGNLTTAGFACSQCGQPLFVYKLEQVSEKGKAYTNYFGRC NVAEHEKLILLAQLKPEKDSDEAVTYSLGKFGQRALDFYSIHVTRESNHPVKP LEQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLAN LKDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGR

[0261]

描述*	氨基酸序列
自 SEQ ID NO: 1 的 NTSB 结构域	DEAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLA GYKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWE RIDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEF CRCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNL YLIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLI ILPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFV ALTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNP THILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARD LLYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSK TYLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVE GQITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRF SHRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTD KRAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO: 49）。
来自 SEQ ID NO: 2 的 NTSB、螺旋形 I、螺旋形 II、OBD 和 RuvC 结构域，以及来自 SEQ ID NO: 1 的 TSL 结构域	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITTADYDGMLVRLKKTSDGWATTLNNKELKAEG QITYYNRYKRQTVEKELSAELDRLSEESGNNDISKWTKGRRDEALFLLKKRFS HRPVQEQFVCLDCGHEVHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO: 50）。
来自 SEQ ID NO: 1 的 TSL、螺旋形 I、螺旋形 II、OBD 和 RuvC 结构域，以及来自 SEQ ID NO: 2 的 NTSB 结构域	MEKRINKIRKKLSADNATKPVSRSGPMKTLLVRVMTDDLKKRLEKRRKKPEVM PQVISNNAANNLRMLLDDYTKMKEAILQVYWQEFKDDHVGLMCKFAQPAPKNI DQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRCNV SEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTKESTHPVKPLAQ IAGNRYASGPVGKALSDACMGTIASFLSKYQDIIIEHQKVVKGNQKRLESLRE LAGKENLEYPSVTLPPQPHTKEGVDAYNEVIARVRMWVNLNLWQKLKLSRDDA KPLLRLKGFPSFPVVERRENEVDWWNTINEVKKLIDAKRDMGRVFWSGVTAEK RNTILEGYNYLPNENDHKKREGSLENPKKPAKRQFGDLLLYLEKKYAGDWGKV FDEAWERIDKKIAGLTSHIEREEARNAEDAQSKAVLTDWLRAKASFVLERLKE MDEKEFYACEIQLQKWYGDLRGNPFAVEAENRVVDISGFSIGSDGHSIQYRNL LAWKYLENGKREFYLLMNYGKKGRIRFTDGTDIKKSGKWQGLLYGGGKAKVID LTFDPDDEQLIILPLAFGTRQGREFIWNDLLSLETGLIKLANGRVIEKTIYNK KIGRDEPALFVALTFERREVVDPSNIKPVNLIGVDRGENIPAVIALTDPEGCP LPEFKDSSGGPTDILRIGEGYKEKQRAIQAAKEVEQRRAGGYSRKFASKSRNL ADDMVRNSARDLFYHAVTHDAVLVFENLSRGFGRQGKRTFMTERQYTKMEDWL TAKLAYEGLTSKTYLSKTLAQYTSKTCSNCGFTITTADYDGMLVRLKKTSDGW ATTLNNKELKAEGQITYYNRYKRQTVEKELSAELDRLSEESGNNDISKWTKGR

[0262]

描述*	氨基酸序列
	RDEALFLLKKRFSHRPVQEQFVCLDCGHEVHADEQAALNIARSWLFLNSNSTE FKSYKSGKQPFVGAWQAFYKRRLKEVWKPNA（SEQ ID NO: 51）
来自 SEQ ID NO: 1 的 NTSB、螺旋形 I、螺旋形 II、OBD 和 RuvC 结构域，以及来自 SEQ ID NO: 2 的 TSL 结构域	MEKRINKIRKKLSADNATKPVSRSGPMKTLLVRVMTDDLKKRLEKRRKKPEVM PQVISNNAANNLRMLLDDYTKMKEAILQVYWQEFKDDHVGLMCKFAQPASKKI DQNKLKPEMDEKGNLTTAGFACSQCGQPLFVYKLEQVSEKGKAYTNYFGRCNV AEHEKLILLAQLKPEKDSDEAVTYSLGKFGQRALDFYSIHVTKESTHPVKPLA QIAGNRYASGPVGKALSDACMGTIASFLSKYQDIIIEHQKVVKGNQKRLESLR ELAGKENLEYPSVTLPPQPHTKEGVDAYNEVIARVRMWVNLNLWQKLKLSRDD AKPLLRLKGFPSFPVVERRENEVDWWNTINEVKKLIDAKRDMGRVFWSGVTAE KRNTILEGYNYLPNENDHKKREGSLENPKKPAKRQFGDLLLYLEKKYAGDWGK VFDEAWERIDKKIAGLTSHIEREEARNAEDAQSKAVLTDWLRAKASFVLERLK EMDEKEFYACEIQLQKWYGDLRGNPFAVEAENRVVDISGFSIGSDGHSIQYRN LLAWKYLENGKREFYLLMNYGKKGRIRFTDGTDIKKSGKWQGLLYGGGKAKVI DLTFDPDDEQLIILPLAFGTRQGREFIWNDLLSLETGLIKLANGRVIEKTIYN KKIGRDEPALFVALTFERREVVDPSNIKPVNLIGVDRGENIPAVIALTDPEGC PLPEFKDSSGGPTDILRIGEGYKEKQRAIQAAKEVEQRRAGGYSRKFASKSRN LADDMVRNSARDLFYHAVTHDAVLVFENLSRGFGRQGKRTFMTERQYTKMEDW LTAKLAYEGLTSKTYLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATG WMTTINGKELKVEGQITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKG RSGEALSLLKKRFSHRPVQEKFVCLNCGFETHADEQAALNIARSWLFLNSNST EFKSYKSGKQPFVGAWQAFYKRRLKEVWKPNA（SEQ ID NO: 52）。
来自 SEQ ID NO: 2 的 NTSB、TSL、螺旋形 I、螺旋形 II 和 OBD 结构域，以及来自第二 CasX 蛋白的外源 RuvC 结构域或其部分。	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPVNLIGVDRGENIPAVIALTDPEGCPLPEFKDSSGGPT DILRIGEGYKEKQRAIQAAKEVEQRRAGGYSRKFASKSRNLADDMVRNSARDL FYHAVTHDAVLVFENLSRGFGRQGKRTFMTERQYTKMEDWLTAKLAYEGLTSK TYLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVE GQITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRF SHRPVQEKFVCLNCGFETHA（SEQ ID NO: 53）
	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC

[0263]

描述*	氨基酸序列
	RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHA（SEQ ID NO：54）
来自 SEQ ID NO: 2 的 NTSB、TSL、螺旋形 II、OBD 和 RuvC 结构域，以及来自 SEQ ID NO: 1 的螺旋形 I 结构域	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNNAANNLRMLLDDYTKMKEAILQVYWQEFKDDHVGLMCKFAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTKESTHPVKPL AQIAGNRYASGPVGKALSDACMGTIASFLSKYQDIIIEHQKVVKGNQKRLESL RELAGKENLEYPSVTLPPQPHTKEGVDAYNEVIARVRMWVNLNLWQKLKLSRD DAKPLLRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：55）
来自 SEQ ID NO: 2 的 NTSB、TSL、螺旋形 I、OBD 和 RuvC 结构域，以及来自 SEQ ID NO: 1 的螺旋形 II 结构域	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPVVERRENEVDWWNTINEVKKLIDAKRDMGRVFWSGVTA EKRNTILEGYNYLPNENDHKKREGSLENPKKPAKRQFGDLLLYLEKKYAGDWG KVFDEAWERIDKKIAGLTSHIEREEARNAEDAQSKAVLTDWLRAKASFVLERL KEMDEKEFYACEIQLQKWYGDLRGNPFAVEAENSILDISGFSKQYNCAFIWQK DGVKKLNLYLIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNF NFDDPNLIILPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTR QDEPALFVALTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSR FKDSLGNPTHILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADD MVRNTARDLLYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAK LAYEGLSKTYLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTI NGKELKVEGQITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEA LSLLKKRFSHRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQT NKTTGNTDKRAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：56）
来自第一 CasX	MISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPKNIDQ

[0264]

描述*	氨基酸序列
蛋白的 NTSB、TSL、螺旋形 I、螺旋形 II 及 RuvC 结构域，及来自第二 CasX 蛋白的外源 OBD 或其部分	RKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRCNVSE HERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPLEQIG GNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANLKDIA SANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRDEAKP LQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAGYKRQ EALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWERIDKK VEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFCRCEL KLQKWYGDLRGKPFAIEAENRVVDISGFSIGSDGHSIQYRNLLAWKYLENGKR EFYLLMNYGKKGRIRFTDGTDIKKSGKWQGLLYGGGKAKVIDLTFDPDDEQLI ILPLAFGTRQGREFIWNDLLSLETGLIKLANGRVIEKTIYNKKIGRDEPALFV ALTFERREVVDPSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNP THILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARD LLYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSK TYLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVE GQITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRF SHRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTD KRAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO: 57）
	MEKRINKIRKKLSADNATKPVSRSGPMKTLLVRVMTDDLKKRLEKRRKKPEVM PQVISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPKNI DQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRCNV SEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPLEQ IGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANLKD IASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRDEA KPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAGYK RQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWERID KKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFCRC ELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLYLI INYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLIILP LAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVALT FERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPTHI LRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDLLY YAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKTYL SKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEGQI TYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFSHR PVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDKRA FVETWQSFYRKKLKEVWKPAV（SEQ ID NO: 58）
	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENRVVDISGFSIGSDGHSIQYRNLLAWKYLE NGKREFYLLMNYGKKGRIRFTDGTDIKKSGKWQGLLYGGGKAKVIDLTFDPDD EQLIILPLAFGTRQGREFIWNDLLSLETGLIKLANGRVIEKTIYNKKIGRDEP ALFVALTFERREVVDPSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDS

[0265]

描述*	氨基酸序列
	LGNPTHILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRN TARDLLYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYE GLSKTYLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKE LKVEGQITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLL KKRFSHRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTT GNTDKRAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：59）
SEQ ID NO：2 的 L379R 的取代、C477K 的取代、A708K 的取代、位置 793 处 P 的缺失及 T620P 的取代	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFK RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKPLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：60）
SEQ ID NO：2 的 M771A 的取代。	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALLPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAAKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFAAERQYTRMEDWLTAKLAYEGLPSK TYLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVE GQITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRF SHRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTD KRAFVETWQSFYRKKLKEVWKPA（SEQ ID NO：61）
SEQ ID NO：2 的 L379R 的取代、A708K 的取代、位置 793 处 P 的	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL

[0266]

描述*	氨基酸序列
缺失及D732N的取代。	EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLANDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：62）
SEQ ID NO：2的W782Q的取代。	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALLPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAAKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDQLTAKLAYEGLPSK TYLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVE GQITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRF SHRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTD KRAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：63）
SEQ ID NO：2的M771Q的取代	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALLPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAAKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFQAERQYTRMEDWLTAKLAYEGLPSK

描述*	氨基酸序列
	TYLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVE GQITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRF SHRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTD KRAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：64）
SEQ ID NO：2 的 R458I 的取代及 A739V 的取代。	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALLPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLIAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAAKEVEQRRAGGYSRKYASKAKNLADDMVRNTVRDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLPSK TYLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVE GQITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRF SHRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTD KRAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：65）
[0267] SEQ ID NO：2 的 L379R、A708K 的取代、位置 793 处 P 的缺失及 M771N 的取代	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFNAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：66）
SEQ ID NO：2 的 L379R 的取代、A708K 的取代、位置 793 处 P 的缺失及 A739T 的取代	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD

[0268]

描述*	氨基酸序列
	EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTTRDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：67）
SEQ ID NO：2 的 L379R 的取代、C477K 的取代、A708K 的取代、位置 793 处 P 的缺失及 D489S 的取代。	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFK RCELKLQKWYGSLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：68）
SEQ ID NO：2 的 L379R 的取代、C477K 的取代、A708K 的取代、位置 793 处 P 的缺失及 D732N 的取代。	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFK RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLANDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS

[0269]

描述*	氨基酸序列
	HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：69）
SEQ ID NO：2 的 V711K 的取代。	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALLPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAAKEKEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLPSK TYLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVE GQITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRF SHRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTD KRAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：70）
SEQ ID NO：2 的 L379R 的取代、C477K 的取代、A708K 的取代、位置 793 处 P 的缺失及 Y797L 的取代。	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFK RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT LLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：71）
119：SEQ ID NO：2 的 L379R 的取代、A708K 的取代和位置 793 处 P 的缺失。	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER

[0270]

描述*	氨基酸序列
	IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：72）
SEQ ID NO：2 的 L379R 的取代、C477K 的取代、A708K 的取代、位置 793 处 P 的缺失及 M771N 的取代。	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFK RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFNAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：73）
SEQ ID NO：2 的 A708K、位置 793 处 P 的缺失及 E386S 的取代。	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALLPYLSSESDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK

[0271]

描述*	氨基酸序列
	RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO: 74）
SEQ ID NO: 2 的 L379R 的取代、C477K 的取代、A708K 的取代和位置 793 处 P 的缺失。	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFK RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO: 75）
SEQ ID NO: 2 的 L792D 的取代。	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALLPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAAKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGDPSK TYLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVE GQITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRF SHRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTD KRAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO: 76）
SEQ ID NO: 2 的 G791F 的取代。	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALLPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC

[0272]

描述*	氨基酸序列
	RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAAKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEFLPSK TYLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVE GQITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRF SHRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTD KRAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：77）
SEQ ID NO: 2 的A708K、位置 793处 P 的缺失及A739V 的取代。	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALLPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTVRDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：78）
SEQ ID NO: 2 的L379R 的取代、A708K 的取代、位置 793 处 P 的缺失及A739V的取代。	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTVRDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：79）

[0273]

描述*	氨基酸序列
SEQ ID NO: 2 的 C477K 的取代、A708K 的取代和位置 793 处 P 的缺失。	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALLPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFK RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO: 80）
SEQ ID NO: 2 的 L249I 的取代及 M771N 的取代。	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIIIEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALLPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAAKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFNAERQYTRMEDWLTAKLAYEGLPSK TYLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVE GQITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRF SHRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTD KRAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO: 81）
SEQ ID NO: 2 的 V747K 的取代。	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALLPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII

[0274]

描述*	氨基酸序列
	LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAAKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAKTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLPSK TYLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVE GQITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRF SHRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTD KRAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：82）
SEQ ID NO：2 的 L379R 的取代、C477K 的取代、A708K 的取代、位置 793 处 P 的缺失及 M779N 的取代。	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFK RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRNEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：83）
L379R, F755M	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALLPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAAKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIMENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLPSK TYLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVE GQITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRF SHRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTD KRAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：84）
429：L379R, A708K, P793_,	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK

[0275]

描述*	氨基酸序列
Y857R	NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRRKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：85）
430：L379R， A708K，P793_， Y857R，I658V	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRRKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：86）
431：L379R， A708K，P793_， Y857R，I658V， E386N	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSENDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNPT

[0276]

描述*	氨基酸序列
	HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRRKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO: 87）
432: L379R, A708K, P793_, Y857R, I658V, L404K	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQFGDLLKHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRRKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO: 88）
433: L379R, A708K, P793_, Y857R, I658V, ^V192	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQVRALDFYSIHVTRESNHPVKP LEQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLAN LKDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGR DEAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLA GYKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWE RIDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEF CRCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNL YLIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLI ILPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFV ALTFERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNP THILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARD LLYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSK TYLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVE GQITYYNRRKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRF SHRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTD KRAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO: 89）
434: L379R, A708K, P793_, Y857R, I658V, L404K, E386N	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL

[0277]

描述*	氨基酸序列
	EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSENDRKKGKKFARYQFGDLLKHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRRKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：90）
435：L379R， A708K，P793_， Y857R，I658V， F399L	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQLGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRRKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：91）
436：L379R， A708K，P793_， Y857R，I658V， F399L，E386N	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSENDRKKGKKFARYQLGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT

[0278]

描述*	氨基酸序列
	YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRRKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：92）
437：L379R， A708K，P793_， Y857R，I658V， F399L，C477S	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQLGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFS RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRRKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：93）
438：L379R， A708K，P793_， Y857R，I658V， F399L，L404K	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQLGDLLKHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRRKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：94）
439：L379R， A708K，P793_， Y857R，I658V， F399L，E386N， C477S，L404K	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD

[0279]

描述*	氨基酸序列
	EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSENDRKKGKKFARYQLGDLLKHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFS RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRRKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO: 95）
440: L379R, A708K, P793_, Y857R, I658V, F399L, Y797L	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQLGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT LLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRRKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO: 96）
441: L379R, A708K, P793_, Y857R, I658V, F399L, Y797L, E386N	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSENDRKKGKKFARYQLGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT LLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRRKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS

[0280]

描述*	氨基酸序列
	HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：97）
442：L379R, A708K, P793_, Y857R, I658V, F399L, Y797L, E386N, C477S, L404K	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSENDRKKGKKFARYQLGDLLKHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFS RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT LLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRRKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：98）
443：L379R, A708K, P793_, Y857R, I658V, Y797L	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT LLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRRKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：99）
444：L379R, A708K, P793_, Y857R, I658V, Y797L, L404K	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQFGDLLKHLEKKHGEDWGKVYDEAWER

[0281]

描述*	氨基酸序列
	IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT LLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRRKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：100）
445：L379R， A708K，P793_， Y857R，I658V， Y797L，E386N	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSENDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT LLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRRKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：101）
446：L379R， A708K，P793_， Y857R，I658V， Y797L，E386N， C477S，L404K	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSENDRKKGKKFARYQFGDLLKHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFS RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT LLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRRKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK

[0282]

描述*	氨基酸序列
	RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：102）
447：L379R，A708K，P793_，Y857R，E386N	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSENDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRRKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：103）
448：L379R，A708K，P793_，Y857R，E386N，L404K	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSENDRKKGKKFARYQFGDLLKHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRRKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：104）
449：L379R，A708K，P793_，D732N，E385P，Y857R	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSPEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC

[0283]

描述*	氨基酸序列
	RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLANDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRRKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：105）
450：L379R， A708K，P793_， D732N，E385P， Y857R，I658V	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSPEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLANDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRRKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：106）
451：L379R， A708K，P793_， D732N，E385P， Y857R，I658V， F399L	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSPEDRKKGKKFARYQLGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLANDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRRKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：107）

[0284]

描述*	氨基酸序列
452: L379R, A708K, P793_, D732N, E385P, Y857R, I658V, E386N	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSPNDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLANDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRRKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：108）
453: L379R, A708K, P793_, D732N, E385P, Y857R, I658V, L404K	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSPEDRKKGKKFARYQFGDLLKHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLANDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRRKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：109）
454: L379R, A708K, P793_, T620P, E385P, Y857R, Q252K	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHKKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSPEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII

[0285]

描述*	氨基酸序列
	LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKPLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRRKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：110）
455：L379R，A708K，P793_，T620P，E385P，Y857R，I658V，Q252K	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHKKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSPEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKPLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRRKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：111）
456：L379R，A708K，P793_，T620P，E385P，Y857R，I658V，E386N，Q252K	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHKKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSPNDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKPLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRRKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：112）
457：L379R，A708K，P793_，	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK

[0286]

描述*	氨基酸序列
T620P, E385P, Y857R, I658V, F399L, Q252K	NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHKKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSPEDRKKGKKFARYQLGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKPLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRRKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：113）
458: L379R, A708K, P793_, T620P, E385P, Y857R, I658V, L404K, Q252K	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHKKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSPEDRKKGKKFARYQFGDLLKHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKPLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRRKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：114）
459: L379R, A708K, P793_, T620P, Y857R, I658V, E386N	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSENDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKPLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNPT

[0287]

描述*	氨基酸序列
	HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRRKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：115）
460：L379R, A708K, P793_, T620P, E385P, Q252K	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHKKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSPEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKPLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：116）
278	QEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPEN IPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPKN IDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRCN VSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPLE QIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANLK DIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRDE AKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAGY KRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWERI DKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFCR CELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLYL IINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLIIL PLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVAL TFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPTH ILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDLL YYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKTY LSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEGQ ITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFSH RPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDKR AFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：117）
279	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL

[0288]

描述*	氨基酸序列
	EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：118）
280	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：119）
285	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT

[0289]

描述*	氨基酸序列
	YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：120）
286	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：121）
287	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：122）
288	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTMSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGR CNVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKP LEQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLAN LKDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGR

[0290]

描述*	氨基酸序列
	DEAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLA GYKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWE RIDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEF CRCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNL YLIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLI ILPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFV ALTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNP THILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARD LLYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSK TYLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVE GQITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRF SHRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTD KRAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：123）
290	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：124）
291	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS

[0291]

描述*	氨基酸序列
	HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：125）
293	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：126）
300	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：127）
492	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER

[0292]

描述*	氨基酸序列
	IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：128）
493	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：129）
387: 来自 SEQ ID NO:1 的 NTSB 交换	QEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPEN IPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPASKK IDQNKLKPEMDEKGNLTTAGFACSQCGQPLFVYKLEQVSEKGKAYTNYFGRCN VAEHEKLILLAQLKPEKDSDEAVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK

[0293]

描述*	氨基酸序列
	RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO: 130）
395: 来自 SEQ ID NO:1 的螺旋形 1B 交换	QEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPEN IPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPKN IDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRCN VSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTKESTHPVKPLA QIAGNRYASGPVGKALSDACMGTIASFLSKYQDIIIEHQKVVKGNQKRLESLR ELAGKENLEYPSVTLPPQPHTKEGVDAYNEVIARVRMWVNLNLWQKLKLSRDD AKPLLRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAGY KRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWERI DKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFCR CELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLYL IINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLIIL PLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVAL TFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPTH ILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDLL YYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKTY LSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEGQ ITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFSH RPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDKR AFVETWQSFYRKKLKEVWKPAV（SEQ ID NO: 131）
485: 来自 SEQ ID NO:1 的螺旋形 1B 交换	QEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPEN IPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPKN IDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRCN VSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTKESTHPVKPLA QIAGNRYASGPVGKALSDACMGTIASFLSKYQDIIIEHQKVVKGNQKRLESLR ELAGKENLEYPSVTLPPQPHTKEGVDAYNEVIARVRMWVNLNLWQKLKLSRDD AKPLLRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAGY KRQEALRPYLSSEEDRKKGKKFARYQLGDLLLHLEKKHGEDWGKVYDEAWERI DKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFCR CELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLYL IINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLIIL PLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVAL TFERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNPTH ILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDLL YYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKTY LSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEGQ ITYYNRRKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFSH RPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDKR AFVETWQSFYRKKLKEVWKPAV（SEQ ID NO: 132）
486: 来自 SEQ ID NO:1 的螺旋形 1B 交换	QEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPEN IPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPKN IDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRCN VSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTKESTHPVKPLA QIAGNRYASGPVGKALSDACMGTIASFLSKYQDIIIEHQKVVKGNQKRLESLR ELAGKENLEYPSVTLPPQPHTKEGVDAYNEVIARVRMWVNLNLWQKLKLSRDD AKPLLRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAGY KRQEALRPYLSSEEDRKKGKKFARYQLGDLLKHLEKKHGEDWGKVYDEAWERI DKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFCR

[0294]

描述*	氨基酸序列
	CELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLYL IINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLIIL PLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVAL TFERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNPTH ILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDLL YYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKTY LSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEGQ ITYYNRRKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFSH RPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDKR AFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：133）
487: 来自 SEQ ID NO:1 的螺旋形 1B 交换	QEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPEN IPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPKN IDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRCN VSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTKESTHPVKPLA QIAGNRYASGPVGKALSDACMGTIASFLSKYQDIIIEHQKVVKGNQKRLESLR ELAGKENLEYPSVTLPPQPHTKEGVDAYNEVIARVRMWVNLNLWQKLKLSRDD AKPLLRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAGY KRQEALRPYLSSEEDRKKGKKFARYQLGDLLLHLEKKHGEDWGKVYDEAWERI DKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFCR CELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLYL IINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLIIL PLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVAL TFERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNPTH ILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDLL YYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKTY LSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEGQ ITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFSH RPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDKR AFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：134）
488: 来自 SEQ ID NO:1 的 NTSB 和螺旋形 1B 交换	QEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPEN IPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPASKK IDQNKLKPEMDEKGNLTTAGFACSQCGQPLFVYKLEQVSEKGKAYTNYFGRCN VAEHEKLILLAQLKPEKDSDEAVTYSLGKFGQRALDFYSIHVTKESTHPVKPL AQIAGNRYASGPVGKALSDACMGTIASFLSKYQDIIIEHQKVVKGNQKRLESL RELAGKENLEYPSVTLPPQPHTKEGVDAYNEVIARVRMWVNLNLWQKLKLSRD DAKPLLRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：135）

[0295]

描述*	氨基酸序列
489: 来自 SEQ ID NO:1 的 NTSB 和螺旋形 1B 交换	QEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPEN IPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPASKK IDQNKLKPEMDEKGNLTTAGFACSQCGQPLFVYKLEQVSEKGKAYTNYFGRCN VAEHEKLILLAQLKPEKDSDEAVTYSLGKFGQRALDFYSIHVTKESTHPVKPL AQIAGNRYASGPVGKALSDACMGTIASFLSKYQDIIIEHQKVVKGNQKRLESL RELAGKENLEYPSVTLPPQPHTKEGVDAYNEVIARVRMWVNLNLWQKLKLSRD DAKPLLRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQLGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRRKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：136）
490: 来自 SEQ ID NO:1 的 NTSB 和螺旋形 1B 交换	QEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPEN IPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPASKK IDQNKLKPEMDEKGNLTTAGFACSQCGQPLFVYKLEQVSEKGKAYTNYFGRCN VAEHEKLILLAQLKPEKDSDEAVTYSLGKFGQRALDFYSIHVTKESTHPVKPL AQIAGNRYASGPVGKALSDACMGTIASFLSKYQDIIIEHQKVVKGNQKRLESL RELAGKENLEYPSVTLPPQPHTKEGVDAYNEVIARVRMWVNLNLWQKLKLSRD DAKPLLRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQLGDLLKHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRRKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：137）
491: 来自 SEQ ID NO:1 的 NTSB 和螺旋形 1B 交换	QEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPEN IPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPASKK IDQNKLKPEMDEKGNLTTAGFACSQCGQPLFVYKLEQVSEKGKAYTNYFGRCN VAEHEKLILLAQLKPEKDSDEAVTYSLGKFGQRALDFYSIHVTKESTHPVKPL AQIAGNRYASGPVGKALSDACMGTIASFLSKYQDIIIEHQKVVKGNQKRLESL RELAGKENLEYPSVTLPPQPHTKEGVDAYNEVIARVRMWVNLNLWQKLKLSRD DAKPLLRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQLGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII

[0296]

描述*	氨基酸序列
	LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：138）
494: 来自 SEQ ID NO:1 的 NTSB 交换	QEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPEN IPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPASKK IDQNKLKPEMDEKGNLTTAGFACSQCGQPLFVYKLEQVSEKGKAYTNYFGRCN VAEHEKLILLAQLKPEKDSDEAVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQLGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDK RAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：139）
328: S867G	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALLPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAAKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLPSK TYLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVE GQITYYNRYKRQNVVKDLGVELDRLSEESVNNDISSWTKGRSGEALSLLKKRF SHRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTD KRAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：140）
388: L379R+A708K+	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK

[0297]

描述*	氨基酸序列
[P793]+X1 螺旋形 2 交换	NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPVVERRENEVDWWNTINEVKKLIDAKRDMGRVFWSGVTA EKRNTILEGYNYLPNENDHKKREGSLENPKKPAKRQFGDLLLYLEKKYAGDWG KVFDEAWERIDKKIAGLTSHIEREEARNAEDAQSKAVLTDWLRAKASFVLERL KEMDEKEFYACEIQLQKWYGDLRGNPFAVEAENSILDISGFSKQYNCAFIWQK DGVKKLNLYLIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNF NFDDPNLIILPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTR QDEPALFVALTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSR FKDSLGNPTHILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADD MVRNTARDLLYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAK LAYEGLSKTYLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTI NGKELKVEGQITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEA LSLLKKRFSHRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQT NKTTGNTDKRAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：141）
389: L379R+A708K+ [P793]+X1 RuvC1 交换	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPVNLIGVDRGENIPAVIALTDPEGCPLPEFKDSSGGPT DILRIGEGYKEKQRAIQAAKEVEQRRAGGYSRKFASKSRNLADDMVRNSARDL FYHAVTHDAVLVFENLSRGFGRQGKRTFMTERQYTKMEDWLTAKLAYEGLTSK TYLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVE GQITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRF SHRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTD KRAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：142）
390: L379R+A708K+ [P793]+X1 RuvC2 交换	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRC NVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPL EQIGGNSCASGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANL KDIASANGLAFPKITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRD EAKPLQRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPT

[0298]

描述*	氨基酸序列
	HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKT YLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEG QITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLNSNSTEFKSYKSGKQPFVG AWQAFYKRRLKEVWKPNA（SEQ ID NO：143）
527: 491 中的^R26	QEIKRINKIRRRLVKDSNTKKAGKTRGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPASK KIDQNKLKPEMDEKGNLTTAGFACSQCGQPLFVYKLEQVSEKGKAYTNYFGRC NVAEHEKLILLAQLKPEKDSDEAVTYSLGKFGQRALDFYSIHVTKESTHPVKP LAQIAGNRYASGPVGKALSDACMGTIASFLSKYQDIIIEHQKVVKGNQKRLES LRELAGKENLEYPSVTLPPQPHTKEGVDAYNEVIARVRMWVNLNLWQKLKLSR DDAKPLLRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLA GYKRQEALRPYLSSEEDRKKGKKFARYQLGDLLLHLEKKHGEDWGKVYDEAWE RIDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEF CRCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNL YLIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLI ILPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFV ALTFERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNP THILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARD LLYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSK TYLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVE GQITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRF SHRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTD KRAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：144）
528: 515 中的 G223Y	QEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPEN IPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPASKK IDQNKLKPEMDEKGNLTTAGFACSQCGQPLFVYKLEQVSEKGKAYTNYFGRCN VAEHEKLILLAQLKPEKDSDEAVTYSLGKFGQRALDFYSIHVTKESTHPVKPL AQIAGNRYASYPVGKALSDACMGTIASFLSKYQDIIIEHQKVVKGNQKRLESL RELAGKENLEYPSVTLPPQPHTKEGVDAYNEVIARVRMWVNLNLWQKLKLSRD DAKPLLRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQLGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLPSK TYLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVE GQITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRF SHRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTD KRAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：145）
529: 515 中的 G223N	QEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPEN IPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPASKK IDQNKLKPEMDEKGNLTTAGFACSQCGQPLFVYKLEQVSEKGKAYTNYFGRCN VAEHEKLILLAQLKPEKDSDEAVTYSLGKFGQRALDFYSIHVTKESTHPVKPL

[0299]

描述*	氨基酸序列
	AQIAGNRYASNPVGKALSDACMGTIASFLSKYQDIIIEHQKVVKGNQKRLESL RELAGKENLEYPSVTLPPQPHTKEGVDAYNEVIARVRMWVNLNLWQKLKLSRD DAKPLLRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQLGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLII LPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVA LTFERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNPT HILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDL LYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLPSK TYLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVE GQITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRF SHRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTD KRAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO:146）
530: 515 中的^W539	QEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPEN IPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPASKK IDQNKLKPEMDEKGNLTTAGFACSQCGQPLFVYKLEQVSEKGKAYTNYFGRCN VAEHEKLILLAQLKPEKDSDEAVTYSLGKFGQRALDFYSIHVTKESTHPVKPL AQIAGNRYASGPVGKALSDACMGTIASFLSKYQDIIIEHQKVVKGNQKRLESL RELAGKENLEYPSVTLPPQPHTKEGVDAYNEVIARVRMWVNLNLWQKLKLSRD DAKPLLRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQLGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGWGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLI ILPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFV ALTFERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNP THILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARD LLYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLPS KTYLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKV EGQITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKR FSHRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNT DKRAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO: 147）
531: 515 中的^Y539	QEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPEN IPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPASKK IDQNKLKPEMDEKGNLTTAGFACSQCGQPLFVYKLEQVSEKGKAYTNYFGRCN VAEHEKLILLAQLKPEKDSDEAVTYSLGKFGQRALDFYSIHVTKESTHPVKPL AQIAGNRYASGPVGKALSDACMGTIASFLSKYQDIIIEHQKVVKGNQKRLESL RELAGKENLEYPSVTLPPQPHTKEGVDAYNEVIARVRMWVNLNLWQKLKLSRD DAKPLLRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAG YKRQEALRPYLSSEEDRKKGKKFARYQLGDLLLHLEKKHGEDWGKVYDEAWER IDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFC RCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLY LIINYFKGYGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLI ILPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFV ALTFERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNP THILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARD LLYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLPS

[0300]

描述*	氨基酸序列
	KTYLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKV EGQITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKR FSHRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNT DKRAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：147）
532： 515中的^R26	QEIKRINKIRRRLVKDSNTKKAGKTRGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPASK KIDQNKLKPEMDEKGNLTTAGFACSQCGQPLFVYKLEQVSEKGKAYTNYFGRC NVAEHEKLILLAQLKPEKDSDEAVTYSLGKFGQRALDFYSIHVTKESTHPVKP LAQIAGNRYASGPVGKALSDACMGTIASFLSKYQDIIIEHQKVVKGNQKRLES LRELAGKENLEYPSVTLPPQPHTKEGVDAYNEVIARVRMWVNLNLWQKLKLSR DDAKPLLRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLA GYKRQEALRPYLSSEEDRKKGKKFARYQLGDLLLHLEKKHGEDWGKVYDEAWE RIDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEF CRCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNL YLIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLI ILPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFV ALTFERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNP THILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARD LLYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLPS KTYLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKV EGQITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKR FSHRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNT DKRAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：149）
533： 515中的^R26和 G223Y	QEIKRINKIRRRLVKDSNTKKAGKTRGPMKTLLVRVMTPDLRERLENLRKKPE NIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPASK KIDQNKLKPEMDEKGNLTTAGFACSQCGQPLFVYKLEQVSEKGKAYTNYFGRC NVAEHEKLILLAQLKPEKDSDEAVTYSLGKFGQRALDFYSIHVTKESTHPVKP LAQIAGNRYASYPVGKALSDACMGTIASFLSKYQDIIIEHQKVVKGNQKRLES LRELAGKENLEYPSVTLPPQPHTKEGVDAYNEVIARVRMWVNLNLWQKLKLSR DDAKPLLRLKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLA GYKRQEALRPYLSSEEDRKKGKKFARYQLGDLLLHLEKKHGEDWGKVYDEAWE RIDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEF CRCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNL YLIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLI ILPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFV ALTFERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGNP THILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARD LLYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLPS KTYLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKV EGQITYYNRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKR FSHRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNT DKRAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：150）

[0301] 在一些实施例中，CasX变异蛋白质包含选自由以下组成的组的序列：SEQ ID NO:49-150、233-235、238-252、272-281。

[0302] 在一些实施例中，当与参考CasX蛋白相比时，例如与SEQ ID NO:1、SEQ ID NO:2或SEQ ID NO:3的参考蛋白相比时，CasX变异蛋白具有一个或多个CasX蛋白改进特征。在一些实施例中，CasX变异体的至少一个改进特征是相对于参考蛋白改进至少约1.1至约100,000倍。在一些实施例中，CasX变异体的至少一种改进特征是相比于参考CasX蛋白改进至少约1.1至约10,000倍、改进至少约1.1至约1,000倍、改进至少约1.1至约500倍、改进至少约1.1至约400倍、改进至少约1.1至约300倍、改进至少约1.1至约200倍、改进至少约1.1至约100倍、改进至少约1.1至约50倍、改进至少约1.1至约40倍、改进至少约1.1至约30倍、改进至少约1.1至约20倍、改进至少约1.1至约10倍、改进至少约1.1至约9倍、改进至少约1.1至约8

倍、改进至少约1.1至约7倍、改进至少约1.1至约6倍、改进至少约1.1至约5倍、改进至少约1.1至约4倍、改进至少约1.1至约3倍、改进至少约1.1至约2倍、改进至少约1.1至约1.5倍、改进至少约1.5至约3倍、改进至少约1.5至约4倍、改进至少约1.5至约5倍、改进至少约1.5至约10倍、改进至少约5至约10倍、改进至少约10至约20倍、改进至少10至约30倍、改进至少10至约50倍或改进至少10至约100倍。在一些实施例中,CasX变异体的至少一种改进特征是相对于参考CasX蛋白改进至少约10至约1000倍。

[0303] 在一些实施例中,CasX变异蛋白质的一种或多种改进的特征为相对于SEQ ID NO:1、SEQ ID NO:2或SEQ ID NO:3的参考CasX蛋白改进至少约1.1、至少约5、至少约10、至少约20、至少约30、至少约40、至少约50、至少约60、至少约70、至少约80、至少约90、至少约100、至少约250、至少约500、或至少约1000、至少约5,000、至少约10,000或至少约100,000倍。在其它情况下,CasX变异体的一个或多个改进特征是相对于SEQ ID NO:1、SEQ ID NO:2或SEQ ID NO:3的参考CasX改进约1.1至100,00倍、约1.1至10,00倍、约1.1至1,000倍、约1.1至500倍、约1.1至100倍、约1.1至50倍、约1.1至20倍、约10至100,00倍、约10至10,00倍、约10至1,000倍、约10至500倍、约10至100倍、约10至50倍、约10至20倍、约2至70倍、约2至50倍、约2至30倍、约2至20倍、约2至10倍、约5至50倍、约5至30倍、约5至10倍、约100至100,00倍、约100至10,00倍、约100至1,000倍、约100至500倍、约500至100,00倍、约500至10,00倍、约500至1,000倍、约500至750倍、约1,000至100,00倍、约10,000至100,00倍、约20至500倍、约20至250倍、约20至200倍、约20至100倍、约20至50倍、约50至10,000倍、约50至1,000倍、约50至500倍、约50至200倍或约50至100倍。

[0304] 可以相对于参考CasX蛋白中的相同特征在CasX变异蛋白质中改进的示范性特征包括但不限于:改进的变异体折叠、对gNA的改进的结合亲和力、对更广范围的PAM序列的改进的结合亲和力、改进的靶DNA解旋、增加的活性、改进的编辑效率、改进的编辑特异性、增加的核酸酶活性、增加的用于双链裂解的靶链负载、减少的用于单链切割的靶链负载、减少的脱靶裂解、改进的DNA非靶链的结合、改进的蛋白质稳定性、改进的蛋白质:gNA复合物稳定性、改进的蛋白质溶解度、改进的蛋白质:gNA复合物溶解度、改进的蛋白质产率、改进的蛋白质表达和改进的融合特征。在一些实施例中,变异体包含至少一种改进特征。在其它实施例中,变异体包含至少两种改进特征。在其它实施例中,变异体包含至少三种改进特征。在一些实施例中,变异体包含至少四种改进特征。在其它实施例中,变异体包含至少五种、至少六种、至少七种、至少八种、至少九种、至少十种、至少十一种、至少十二种、至少十三种或更多种改进特征。这些改进特征更详细描述于下文中。

[0305] j.蛋白质稳定性

[0306] 在一些实施例中,本公开提供相对于参考CasX蛋白具有经改进稳定性的CasX变异蛋白。在一些实施例中,CasX变异蛋白的经改进稳定性引起较高稳态蛋白质的表达,其提高编辑效率。在一些实施例中,CasX变异蛋白的经改进稳定性使得较大分率的CasX蛋白以功能性构象保持折叠,且提高编辑效率或改进纯化能力以用于制造目的。如本文所用,“功能性构象”是指构象为其中蛋白质能够结合gNA及靶DNA的CasX蛋白。在其中CasX变异体不携有一个或多个使其催化死亡的突变的实施例中,CasX变异体能够裂解、切割或以其它方式修饰靶DNA。举例来说,在一些实施例中,功能性CasX变异体可用于基因编辑,且功能性构象是指“编辑潜能”构象。在一些示范性实施例,包括其中CasX变异蛋白产生较大分率的以功

能性构象保持折叠的CasX蛋白的那些实施例中，相比于参考CasX蛋白，例如基因编辑的应用需要较低浓度的CasX变异体。因此，在一些实施例中，相比于参考CasX，具有经改进稳定性的CasX变异体在一种或多种基因编辑背景下具有经改进效率。

[0307] 在一些实施例中，本公开提供相对于参考CasX蛋白具有经改进热稳定性的CasX变异蛋白。在一些实施例中，CasX变异蛋白在特定温度范围内具有改进的CasX变异蛋白热稳定性。不希望受任何理论束缚，一些参考CasX蛋白天然地在生态栖位处于地下水及沉积物中的生物体中起作用；因此，一些参考CasX蛋白可能已进化为在比某些应用可能需要的温度更低或更高的温度下展现最优选功能。举例来说，CasX变异蛋白的一种应用为哺乳动物细胞的基因编辑，其典型地在约37℃下进行。在一些实施例中，相比于参考CasX蛋白，如本文所述的CasX变异蛋白在至少16℃、至少18℃、至少20℃、至少22℃、至少24℃、至少26℃、至少28℃、至少30℃、至少32℃、至少34℃、至少35℃、至少36℃、至少37℃、至少38℃、至少39℃、至少40℃、至少41℃、至少42℃、至少44℃、至少46℃、至少48℃、至少50℃、至少52℃或更高温度下具有改进的热稳定性。在一些实施例中，相比于参考CasX蛋白，CasX变异蛋白具有改进的热稳定性及功能，产生改进的基因编辑功能，例如哺乳动物基因编辑应用，其可包括人类基因编辑应用。可以通过所属领域的技术人员已知的多种方法评价核酸酶的改进的热稳定性。

[0308] 在一些实施例中，本公开提供了相对于参考CasX蛋白:gNA复合物具有改进的CasX变异蛋白:gNA复合物稳定性的CasX变异蛋白，使得RNP保持于功能形式。稳定性改进可包括增加的热稳定性；蛋白水解降解抗性；增强的药物动力学特性；跨越一系列pH条件、盐条件及张力的稳定性。在一些实施例中，复合物改进的稳定性使得编辑效率提高。在一些实施例中，CasX变异体和gNA变异体的RNP与SEQ ID NO:1-3的参考CasX的RNP和表1的SEQ ID NO:4-16中任一项的gNA的RNP相比，具有百分比高至少5％、至少10％、至少15％、或至少20％、或至少5-20％的裂解潜能RNP。实例中提供增加的裂解潜能RNP的示范性数据。

[0309] 在一些实施例中，本公开提供相对于参考CasX蛋白:gNA复合物具有改进的CasX变异蛋白:gNA复合物热稳定性的CasX变异蛋白。在一些实施例中，CasX变异蛋白相对于参考CasX蛋白具有改进的热稳定性。在一些实施例中，CasX变异蛋白:gNA复合物在至少16℃、至少18℃、至少20℃、至少22℃、至少24℃、至少26℃、至少28℃、至少30℃、至少32℃、至少34℃、至少35℃、至少36℃、至少37℃、至少38℃、至少39℃、至少40℃、至少41℃、至少42℃、至少44℃、至少46℃、至少48℃、至少50℃、至少52℃或更高的温度下相对于包含参考CasX蛋白的复合物具有改进的热稳定性。在一些实施例中，与参考CasX蛋白:gNA复合物相比，CasX变异蛋白具有改进的CasX变异蛋白:gNA复合物热稳定性，其使得针对基因编辑应用，如哺乳动物基因编辑应用(其可包括人类基因编辑应用)的功能改进。可以通过所属领域的技术人员已知的多种方法评价RNP的改进的热稳定性。

[0310] 在一些实施例中，CasX变异蛋白改进的稳定性和/或热稳定性包含CasX变异蛋白相对于参考CasX蛋白更快的折叠动力学、CasX变异蛋白相对于参考CasX蛋白更慢的去折叠动力学、CasX变异蛋白相对于参考CasX蛋白在折叠时更大的自由能释放、相对于参考CasX蛋白更高的50％的CasX变异蛋白未折叠的温度(Tm)或其任何组合。这些特征可改进大范围的值；例如相比于参考CasX蛋白改进至少1.1、至少1.5、至少10、至少50、至少100、至少500、至少1,000、至少5,000或至少10,000倍。在一些实施例中，CasX变异蛋白改进的热稳定性包

含CasX变异蛋白相对于参考CasX蛋白更高的Tm。在一些实施例中，CasX变异蛋白的Tm为约20℃至约30℃、约30℃至约40℃、约40℃至约50℃、约50℃至约60℃、约60℃至约70℃、约70℃至约80℃、约80℃至约90℃或约90℃至约100℃。热稳定性係通过测量“熔融温度”(Tm)来测定，熔融温度定义为一半分子变性的温度。测量蛋白质稳定性的特征，例如Tm及去折叠自由能的方法为所属领域所属领域的一般技术人员所知，且可使用标准生物化学技术在体外测量。举例来说，Tm可使用差示扫描热测量定来测量，差示扫描热测量定为一种热分析技术，其中测量增加样品及参考的温度所需的热量差作为温度的函数(Chen等人(2003)Pharm Res 20:1952-60;Ghirlando等人(1999)Immunol Lett 68:47-52)。或者或另外，CasX变异蛋白Tm可使用市售方法，例如ThermoFisher Protein Thermal Shift system系统来测量。或者或另外，圆二色性可用于测量折叠及去折叠的动力学，以及Tm(Murray等人(2002)J.Chromatogr Sci 40:343-9)。圆二色性(CD)依赖于左手侧及右手侧圆偏振光被例如蛋白质的不对称分子不等地吸收。蛋白质的某些结构，例如α螺旋及β折叠具有特征性CD光谱。因此，在一些实施例中，CD可用于确定CasX变异蛋白的二级结构。

[0311] 在一些实施例中，CasX变异蛋白改进的稳定性和/或热稳定性包含CasX变异蛋白相对于参考CasX蛋白改进的折叠动力学。在一些实施例中，相对于参考CasX蛋白，CasX变异蛋白的折叠动力学改进至少约5倍、至少约10倍、至少约50倍、至少约100倍、至少约500倍、至少约1,000倍、至少约2,000倍、至少约3,000倍、至少约4,000倍、至少约5,000倍或至少约10,000倍改进。在一些实施例中，相对于参考CasX蛋白，CasX变异蛋白的折叠动力学改进至少约1kJ/mol、至少约5kJ/mol、至少约10kJ/mol、至少约20kJ/mol、至少约30kJ/mol、至少约40kJ/mol、至少约50kJ/mol、至少约60kJ/mol、至少约70kJ/mol、至少约80kJ/mol、至少约90kJ/mol、至少约100kJ/mol、至少约150kJ/mol、至少约200kJ/mol、至少约250kJ/mol、至少约300kJ/mol、至少约350kJ/mol、至少约400kJ/mol、至少约450kJ/mol或至少约500kJ/mol。

[0312] 相对于参考CasX蛋白，可增加CasX变异蛋白的稳定性的示范性氨基酸变化可包括但不限于以下氨基酸变化：增加CasX变异蛋白内的氢键数目、增加CasX变异蛋白内的二硫桥键数目、增加CasX变异蛋白内的盐桥数目、增强CasX变异蛋白的部分之间的相互作用、增加CasX变异蛋白的埋入疏水表面积或其任何组合。

[0313] k.蛋白质产率

[0314] 在一些实施例中，本公开提供相对于参考CasX蛋白，在表达及纯化期间具有改进产率的CasX变异蛋白。在一些实施例中，相对于参考CasX蛋白，自细菌或真核宿主细胞纯化的CasX变异蛋白的产率经改进。在一些实施例中，细菌宿主细胞为大肠杆菌细胞。在一些实施例中，真核细胞是酵母、植物(例如烟草)、昆虫(例如草地贪夜蛾(Spodoptera frugiperda)sf9细胞)、小鼠、大鼠、仓鼠、天竺鼠、猴子或人类细胞。在一些实施例中，真核宿主细胞是哺乳动物细胞，包括但不限于人胚肾293(HEK293)细胞、HEK292T细胞、幼仓鼠肾(BHK)细胞、NS0细胞、SP2/0细胞、YO骨髓瘤细胞、P3X63小鼠骨髓瘤细胞、PER细胞、PER.C6细胞、杂交瘤细胞、NIH3T3细胞、COS、HeLa或中国仓鼠卵巢(CHO)细胞。

[0315] 在一些实施例中，CasX变异蛋白的改进产率经由密码子优化达成。细胞使用64种不同的密码子，其中的61种编码20种标准氨基酸，而另外3种充当终止密码子。在一些情况下，单一氨基酸由超过一个密码子编码。对于相同的天然存在的氨基酸，不同生物体展现朝向使用不同密码子的偏移。因此，蛋白编码序列中密码子的选择，及将密码子选择与蛋白质

将表达的生物体匹配可在一些情况下显著影响蛋白质翻译且因此影响蛋白质表达量。在一些实施例中,CasX变异蛋白由已经密码子优化的核酸编码。在一些实施例中,编码CasX变异蛋白的核酸已经密码子优化以表达于细菌细胞、酵母细胞、昆虫细胞、植物细胞或哺乳动物细胞中。在一些实施例中,哺乳动物细胞为小鼠、大鼠、仓鼠、天竺鼠、猴或人类。在一些实施例中,CasX变异蛋白由已经密码子优化以表达于人类细胞中的核酸编码。在一些实施例中,CasX变异蛋白由已去除降低原核生物及真核生物中的翻译速率的核苷酸序列的核酸编码。举例来说,大于三个胸腺嘧啶残基成一列的运行可降低某些生物体中的翻译速率,或内部聚腺苷酸化信号可减少翻译。

[0316] 在一些实施例中,如本文所述的溶解度及稳定性的改进使得CasX变异蛋白的产率相对于参考CasX蛋白改进。

[0317] 可通过所属领域中已知的方法评估表达及纯化期间改进的蛋白质产率。例如,可如下地测定CasX变异蛋白的量:通过在SDS-page凝胶上运行蛋白质,且将CasX变异蛋白与事先已知其量或浓度的对照进行比较,以确定蛋白质的绝对含量。或者或另外,纯化CasX变异蛋白可在SDS-page凝胶上紧邻经历相同纯化过程的参考CasX蛋白运行,以确定CasX变异蛋白产率的相对改进。或者或另外,蛋白质含量可使用免疫组织化学方法,例如通过针对CasX的抗体的蛋白质印迹或ELISA,或通过HPLC来测量。对于溶液中的蛋白质,可通过测量蛋白质的内在UV吸光度,或通过使用蛋白质依赖性颜色变化的方法,例如劳立分析(Lowry assay)、史密斯铜/双金鸡纳酸分析(Smith copper/bicinchoninic assay)或布拉福染料分析(Bradford dye assay)来确定浓度。此类方法可用于计算在某些条件下通过表达获得的总蛋白质(如总可溶性蛋白质)产率。举例来说,此可与参考CasX蛋白在类似表达条件下的蛋白质产率比较。

[0318] 1.蛋白质溶解度

[0319] 在一些实施例中,CasX变异蛋白相对于参考CasX蛋白具有改进的溶解度。在一些实施例中,相对于包含参考CasX蛋白的核糖核蛋白复合物,CasX变异蛋白具有改进的CasX:gNA核糖核蛋白复合物变异体溶解度。

[0320] 在一些实施例中,蛋白质溶解度的改进使得自蛋白质纯化技术,例如自大肠杆菌纯化的蛋白质产率较高。在一些实施例中,CasX变异蛋白改进的溶解度可使得细胞中的活性能够更高效,因为更可溶的蛋白质不大可能在细胞中聚集。蛋白质聚集体可在某些实施例中对细胞为毒性或繁重的,且不希望受任何理论束缚,增加CasX变异蛋白的溶解度可改善此蛋白质聚集结果。另外,CasX变异蛋白改进的溶解度可允许增强的配制物,准许递送更高有效剂量的功能蛋白,例如在所需基因编辑应用中。在一些实施例中,CasX变异蛋白相对于参考CasX蛋白改进的溶解度使得CasX变异蛋白在纯化期间的产率改进,产率大至少约5倍、至少约10倍、至少约20倍、至少约30倍、至少约40倍、至少约50倍、至少约60倍、至少约70倍、至少约80倍、至少约90倍、至少约100倍、至少约250倍、至少约500倍或至少约1000倍。在一些实施例中,CasX变异蛋白相对于参考CasX蛋白改进的溶解度将CasX变异蛋白在细胞中的活性改进了至少约1.1倍、至少约1.2倍、至少约1.3倍、至少约1.4倍、至少约1.5倍、至少约1.6倍、至少约1.7倍、至少约1.8倍、至少约1.9倍、至少约2倍、至少约2.1倍、至少约2.2倍、至少约2.3倍、至少约2.4倍、至少约2.5倍、至少约2.6倍、至少约2.7倍、至少约2.8倍、至少约2.9倍、至少约3倍、至少约3.5倍、至少约4倍、至少约4.5倍、至少约5倍、至少约5.5倍、

至少约6倍、至少约6.5倍、至少约7.0倍、至少约7.5倍、至少约8倍、至少约8.5倍、至少约9倍、至少约9.5倍、至少约10倍、至少约11倍、至少约12倍、至少约13倍、至少约14倍或至少约15倍的更大活性。可以通过所属领域的技术人员已知的多种方法评价核酸酶的改进的溶解度，包括通过获取溶解的大肠杆菌的可溶分率的凝胶上密度测量法读数。替代地或另外，可通过测量可溶性蛋白质产物在整个蛋白质纯化过程中的维持情况来测量CasX变异蛋白溶解度的改进。例如，可在凝胶亲和纯化、标签裂解、阳离子交换纯化、在施胶柱上运行蛋白质的一个或多个步骤中测量可溶性蛋白质产物。在一些实施例中，在纯化过程的各步骤之后读取凝胶上的每一蛋白质带的密度测定值。在一些实施例中，当相比于参考CasX蛋白时，具有改进溶解度的CasX变异蛋白可在蛋白质纯化过程的一个或多个步骤处维持较高浓度，同时不溶性蛋白质变异体可由于缓冲液交换、过滤步骤、与纯化柱的相互作用等而在一个或多个步骤处损失。

[0321] 在一些实施例中，当相比于参考CasX蛋白时，改进CasX变异蛋白的溶解度产生就蛋白质纯化期间蛋白质的mg/L而言较高的产率。

[0322] 在一些实施例中，当在编辑分析，例如本文所述的EGFP破坏分析中评估时，改进CasX变异蛋白的溶解度使得相比于较不可溶的蛋白质，编辑事件的量能够更大。

[0323] m.对gNA的蛋白质亲和力

[0324] 在一些实施例中，相对于参考CasX蛋白，CasX变异蛋白对gNA的亲和力改进，使得形成核糖核蛋白复合物。CasX变异蛋白对gNA增加的亲和力可例如针对RNP复合物生成产生更低Kd，其可在一些情况下使得核糖核蛋白复合物形成更稳定。在一些实施例中，CasX变异蛋白对gNA增加的亲和力使得核糖核蛋白复合物在递送至人类细胞时的稳定性增加。此增加的稳定性可影响复合物于受试者细胞中的功能及效用，以及使得当递送至受试者时改进血液中的药物动力学特性。在一些实施例中，CasX变异蛋白增加的亲和力，以及由此产生的核糖核蛋白复合物增加的稳定性允许向受试者或细胞递送较低剂量的CasX变异蛋白，同时仍具有所需活性，例如体内或体外基因编辑。

[0325] 在一些实施例中，当CasX变异蛋白及gNA均保持于RNP复合物中时，CasX变异蛋白对gNA的较高亲和力(更紧密结合)允许编辑事件的量更大。可使用编辑分析，如本文所述的EGFP破坏测定来评估增加的编辑事件。

[0326] 在一些实施例中，相对于参考CasX蛋白，CasX变异蛋白针对gNA的K_d增加至少约1.1倍、至少约1.2倍、至少约1.3倍、至少约1.4倍、至少约1.5倍、至少约1.6倍、至少约1.7倍、至少约1.8倍、至少约1.9倍、至少约2倍、至少约3倍、至少约4倍、至少约5倍、至少约6倍、至少约7倍、至少约8倍、至少约9倍、至少约10倍、至少约15倍、至少约20倍、至少约25倍、至少约30倍、至少约35倍、至少约40倍、至少约45倍、至少约50倍、至少约60倍、至少约70倍、至少约80倍、至少约90倍或至少约100倍。在一些实施例中，相比于SEQ ID NO:2的参考CasX蛋白，CasX变异体对gNA的结合亲和力增加约1.1至约10倍。

[0327] 在不希望受理论约束的情况下，在一些实施例中，螺旋形I结构域中的氨基酸变化可增加CasX变异蛋白与gNA靶向序列的结合亲和力，而螺旋形II结构域中的变化可增加CasX变异蛋白与gNA支架茎环的结合亲和力，且寡核苷酸结合结构域(OBD)中的变化增加CasX变异蛋白与gRNA三螺旋体的结合亲和力。

[0328] 测量CasX蛋白对CasX gNA的结合亲和力的方法包括使用纯化的CasX蛋白及gNA的

体外方法。如果gNA或CasX蛋白用荧光团标记，那么可通过荧光偏振测量对参考CasX及变异蛋白的结合亲和力。或者或另外，可通过生物层干涉测量术、电泳迁移率变动分析(EMSA)或过滤结合来测量结合亲和力。定量RNA结合蛋白，例如本公开的参考CasX及变异蛋白对特定gNA，例如参考gNA及其变异体的绝对亲和力的额外标准技术包括但不限于等温量热法(ITC)及表面等离子体子共振(SPR)，以及实例的方法。

[0329] n.对靶核酸的亲和力

[0330] 在一些实施例中，CasX变异蛋白质相对于参考CasX蛋白对靶核酸序列的亲和力，对靶核酸序列的结合亲和力改进。在一些实施例中，对靶核酸序列的改进的亲和力包含对靶核酸序列的改进的亲和力、对更广范围的PAM序列的改进的结合亲和力、改进的在DNA中搜索靶核酸序列的能力或其任何组合。在不希望受到理论约束的情况下，认为如CasX的CRISPR/Cas系统蛋白质可通过沿DNA分子的一维扩散发现其靶核酸序列。认为这个方法包括(1)核糖核蛋白与DNA分子结合，接着为(2)在靶核酸序列处停顿，在一些实施例中，其中的任一个可受CasX蛋白对靶核酸序列改进的亲和力影响，从而相比于参考CasX蛋白改进了CasX变异蛋白的功能。

[0331] 在一些实施例中，具有改进的靶核酸序列亲和力的CasX变异蛋白质对于DNA具有增加的总体亲和力。在一些实施例中，具有改进的靶核酸亲和力的CasX变异蛋白质对除由SEQ ID NO:1或2的参考CasX蛋白识别的典型TTC PAM以外的特定PAM序列具有增加的亲和力，包括对选自由TTC、ATC、GTC和CTC组成的组的PAM序列的结合亲和力。在不希望受到理论约束的情况下，这些蛋白质变异体可总体上更强有力地与DNA相互作用，且由于能够结合野生型Cas X以外的额外PAM序列，因此能够更有效地搜索CasX蛋白中的靶序列，从而能够更有效地访问和编辑靶DNA内的序列。在一些实施例中，对DNA的较高总体亲和力亦可增加CasX蛋白可有效地起始及完成结合及退绕步骤的频率，由此促进靶链侵入及R环形成，且最终促进靶核酸序列裂解。

[0332] 不希望受理论束缚，可能的是增加非靶DNA链的解旋或呈解旋状态的非靶DNA链的捕捉效率的NTSBD中的氨基酸变化可增加CasX变异蛋白对靶DNA的亲和力。或者或另外，增加NTSBD在解旋期间稳定DNA的能力的NTSBD中的氨基酸变化可增加CasX变异蛋白对靶DNA的亲和力。或者或另外，OBD中的氨基酸变化可增加CasX变异蛋白结合至前间隔区邻近基序(PAM)的亲和力，由此增加CasX变异蛋白对靶核酸序列的亲和力。替代地或另外，螺旋形I和/或II、RuvC及TSL结构域中增加CasX变异蛋白对靶核酸链的亲和力的氨基酸变化可增加CasX变异蛋白对靶核酸序列的亲和力。

[0333] 在一些实施例中，相比于SEQ ID NO:1、SEQ ID NO:2或SEQ ID NO:3的参考蛋白质，CasX变异蛋白质对靶核酸序列的结合亲和力增加。在一些实施例中，相对于参考CasX蛋白，本公开的CasX变异蛋白对靶核酸分子的结合亲和力增加至少约1.1倍、至少约1.2倍、至少约1.3倍、至少约1.4倍、至少约1.5倍、至少约1.6倍、至少约1.7倍、至少约1.8倍、至少约1.9倍、至少约2倍、至少约3倍、至少约4倍、至少约5倍、至少约6倍、至少约7倍、至少约8倍、至少约9倍、至少约10倍、至少约15倍、至少约20倍、至少约25倍、至少约30倍、至少约35倍、至少约40倍、至少约45倍、至少约50倍、至少约60倍、至少约70倍、至少约80倍、至少约90倍或至少约100倍。

[0334] 在一些实施例中，CasX变异蛋白对靶核酸的非靶链的结合亲和力改进。如本文所

用,术语“非靶链”是指不与gNA中的靶向序列形成沃森及克里克(Watson and Crick)碱基对,且与靶链互补的DNA靶核酸序列的链。

[0335] 测量CasX蛋白(如参考或变异体)对靶核酸分子的亲和力的方法可包括电泳迁移率变动分析(EMSA)、过滤结合、等温量热法(ITC)及表面等离子体子共振(SPR)、荧光偏振及生物层干涉测量法(BLI)。测量CasX蛋白对靶标的亲和力的其它方法包括测量随时间推移的DNA裂解事件的体外生物化学分析。

[0336] 在一些实施例中,相比于对靶核酸序列不具有增加的亲和力的参考CasX蛋白,对其靶核酸序列具有较高亲和力的CasX变异蛋白可更快速地裂解靶核酸序列。

[0337] 在一些实施例中,CasX变异蛋白是催化死亡的(dCasX)。在一些实施例中,本公开提供了包含催化死亡的CasX蛋白的RNP,其保留结合靶DNA的能力。示范性催化死亡的CasX变异蛋白包含CasX蛋白的RuvC结构域的活性位点中的一个或多个突变。在一些实施例中,催化死亡的CasX变异蛋白包含SEQ ID NO:1的残基672、769和/或935处的取代。在一些实施例中,催化死亡的CasX变异蛋白包含SEQ ID NO:1的参考CasX蛋白中的D672A、E769A和/或D935A的取代。在一些实施例中,催化死亡的CasX蛋白包含SEQ ID NO:2的氨基酸659、765和/或922处的取代。在一些实施例中,催化死亡的CasX蛋白包含SEQ ID NO:2的参考CasX蛋白中的D659A、E756A和/或D922A取代。在其它实施例中,催化死亡的参考CasX蛋白包含参考CasX蛋白的RuvC结构域全部或一部分的缺失。

[0338] 在一些实施例中,CasX变异蛋白对DNA的改进亲和力也改进了CasX变异蛋白的无催化活性形式的功能。在一些实施例中,CasX变异蛋白的无催化活性形式包含RuvC中的DED基序中的一个或多个突变。在一些实施例中,催化死亡的CasX变异蛋白可用于碱基编辑或表观遗传修饰。在一些实施例中,在对DNA的较高亲和力下,相对于催化活性CasX,催化死亡的CasX变异蛋白可更快地发现其靶DNA、与靶DNA保持结合的时间更长、以更稳定方式结合靶DNA,或其组合,从而改进催化死亡的CasX变异蛋白的功能。

[0339] o.对靶位点改进的特异性

[0340] 在一些实施例中,相对于参考CasX蛋白,CasX变异蛋白对靶核酸序列的特异性改进。如本文所用,“特异性”(可互换地称为“靶特异性”)是指CRISPR/Cas系统核糖核蛋白复合物裂解与靶核酸序列类似,但不相同的脱靶序列的程度;例如,相对于参考CasX蛋白,具有较高特异性程度的CasX变异体RNP将展现减少的序列脱靶裂解。CRISPR/Cas系统蛋白质的特异性及潜在有害的脱靶效应的减少可为极其重要的,以便达成用于哺乳动物受试者的可接受治疗指数。

[0341] 在一些实施例中,CasX变异蛋白对与gNA的靶向序列互补的靶核酸序列内的靶位点的特异性改进。

[0342] 不希望受理论束缚,有可能,螺旋形I和II结构域中增加CasX变异蛋白对靶核酸链的特异性的氨基酸变化可增加CasX变异蛋白对靶核酸序列的总体特异性。在一些实施例中,增加CasX变异蛋白对靶核酸序列的特异性的氨基酸变化也可使得CasX变异蛋白对DNA的亲和力降低。

[0343] 测试CasX蛋白(例如变异体或参考)靶特异性的方法可包括引导及环化以通过测序体外报道裂解效应(CIRCLE-seq),或类似方法。简言之,在CIRCLE-seq技术中,基因组DNA经剪切且通过连接茎-环衔接子而环化,所述衔接子在茎-环区中带切口以暴露4个核苷酸

的回文突出端。此后为其余线性DNA的分子内连接及降解。含有CasX裂解位点的环状DNA分子随后经CasX线性化，且衔接子连接至暴露末端，接着进行高通量测序以产生含有关于脱靶位点的信息的配对末端读段。可用于检测脱靶事件，且因此检测CasX蛋白特异性的额外分析包括用于检测及定量那些所选脱靶位点处形成的插入缺失(插入及缺失)的分析，例如失配检测核酸酶分析及二代测序(NGS)。示范性失配检测分析包括核酸酶分析，其中来自用CasX和sgRNA处理的细胞的基因组DNA经PCR扩增、变性和再杂交以形成杂双螺旋DNA，其含有一条野生型链和一条具有插入缺失的链。失配经失配检测核酸酶，例如Surveyor核酸酶或T7核酸内切酶I识别及裂解。

[0344] p.DNA的退绕

[0345] 在一些实施例中，CasX变异蛋白相对于参考CasX蛋白具有改进的解旋DNA的能力。先前已显示不佳dsDNA解旋会削弱或阻止CRISPR/Cas系统蛋白质AnaCas9或Cas14s裂解DNA的能力。因此，不希望受任何理论束缚，可能的是通过一些本公开的CasX变异蛋白增加的DNA裂解活性至少部分归因于增强的发现及解旋靶位点处的dsDNA的能力。

[0346] 不希望受理论束缚，认为NTSB结构域中的氨基酸变化可产生具有增加的DNA解旋特征的CasX变异蛋白。或者或另外，与PAM相互作用的OBD或螺旋形结构域区域中的氨基酸变化亦可产生具有增加的DNA退绕特征的CasX变异蛋白。

[0347] 测量CasX蛋白(例如变异体或参考)退绕DNA的能力的方法包括但不限于观测荧光偏振或生物层干涉测量术中dsDNA靶增加的缔合速率的体外分析。

[0348] q.催化活性

[0349] 本文公开的CasX:gNA系统的核糖核蛋白复合物包含结合靶核酸序列并裂解靶核酸序列的参考CasX蛋白或其变异体。在一些实施例中，CasX变异蛋白相对于参考CasX蛋白具有改进的催化活性。不希望受理论束缚，认为在一些情况下，靶链裂解可为Cas12样分子产生dsDNA断裂中的限制因素。在一些实施例中，CasX变异蛋白改进DNA的靶链的弯曲及此链的裂解，使得通过CasX核糖核蛋白复合物裂解dsDNA的总效率改进。

[0350] 在一些实施例中，CasX变异蛋白相比于参考CasX蛋白具有增加的核酸酶活性。具有增加的核酸酶活性的变异体可例如经由RuvC核酸酶结构域中的氨基酸变化来产生。在一些实施例中，CasX变异体包含具有切口酶活性的核酸酶结构域。在前述内容中，CasX:gNA系统的CasX切口酶在非靶链中PAM位点3'的10-18个核苷酸内产生单链断裂。在其它实施例中，CasX变异体包含具有双链裂解活性的核酸酶结构域。在前述内容中，CasX:gNA系统的CasX在靶链上的PAM位点5'的18-26个核苷酸和非靶链上3'的10-18个核苷酸内产生双链断裂。可通过多种方法，包括实例的那些方法分析核酸酶活性。在一些实施例中，CasX变异体的$K_{裂解}$常数与参考CasX相比大至少2倍，或至少3倍，或至少4倍，或至少5倍，或至少6倍，或至少7倍，或至少8倍，或至少9倍，或至少10倍。

[0351] 在一些实施例中，与参考CasX相比，CasX变异蛋白具有增加的用于双链裂解的靶链负载。具有增加的靶链负载活性的变异体可例如经由TLS结构域中的氨基酸变化来产生。

[0352] 不希望受理论束缚，TSL结构域中的氨基酸变化可产生具有改进的催化活性的CasX变异蛋白。或者或另外，RNA:DNA双螺旋的结合通道周围的氨基酸变化亦可改进CasX变异蛋白的催化活性。

[0353] 在一些实施例中，CasX变异蛋白相比于参考CasX蛋白具有增加的附带裂解活性。

如本文所用,“附带裂解活性”是指在识别及裂解靶核酸序列之后,核酸的额外非靶向裂解。在一些实施例中,CasX变异蛋白相比于参考CasX蛋白具有减少的附带裂解活性。

[0354] 在一些实施例,例如涵盖其中靶核酸序列的裂解并非所需结果的应用的那些实施例中,改进CasX变异蛋白的催化活性包括改变、降低或消除CasX变异蛋白的催化活性。在一些实施例中,包含dCasX变异蛋白的核糖核蛋白复合物结合至靶核酸序列且不裂解靶核酸。

[0355] 在一些实施例中,包含CasX变异蛋白的CasX核糖核蛋白复合物结合靶DNA,但在靶DNA中产生单链切口。在一些实施例,尤其是其中CasX蛋白为切口酶的那些实施例中,CasX变异蛋白具有减少的针对单链切口的靶链负载。具有减少的靶链负载的变异体可例如经由TSL结构域中的氨基酸变化来产生。

[0356] 用于表征CasX蛋白的催化活性的示范性方法可包括但不限于体外裂解分析,包括以下实例的那些。在一些实施例中,DNA产物于琼脂糖凝胶上的电泳可查询链裂解的动力学。

[0357] r.对C9orf72靶DNA和RNA的亲和力

[0358] 在一些实施例中,包含参考CasX蛋白或CasX变异蛋白质的核糖核蛋白复合物与靶C9orf72 DNA结合并且使靶核酸序列裂解。在一些实施例中,核糖核蛋白复合物在靶核酸中产生双链断裂。在其它实施例中,核糖核蛋白复合物在靶核酸中产生单链断裂。在一些实施例中,当与参考CasX蛋白相比时,参考CasX蛋白的变异体增加CasX变异蛋白对靶C9orf72RNA的特异性,且增加CasX变异蛋白相对于靶RNA的活性。举例来说,当相比于参考CasX蛋白时,CasX变异蛋白可显示增加的对靶RNA的结合亲和力,或增加的靶RNA裂解。在一些实施例中,包含CasX变异蛋白的核糖核蛋白复合物结合至靶RNA和/或裂解靶RNA。在一些实施例中,相比于SEQ ID NO:1、SEQ ID NO:2或SEQ ID NO:3的参考蛋白,CasX变异体对C9orf72靶RNA的结合亲和力增加至少约二倍至约10倍。

[0359] s.突变组合

[0360] 本公开提供了Cas X变异体,其是来自单独CasX变异蛋白的突变的组合。在一些实施例中,本文描述的任何结构域的任何变异体可以与本文描述的其它变异体组合。在一些实施例中,本文描述的任何结构域内的任何变异体可以与本文描述的其它变异体在同一结构域中组合。在一些实施例中,不同氨基酸变化的组合可以产生新的优化变异体,其功能通过氨基酸变化的组合得到进一步改善。在一些实施例中,组合氨基酸变化对CasX蛋白功能的影响是线性的。如本文所用,线性组合是指当单独分析时其对功能的影响等于每个单独氨基酸变化的影响的总和的组合。在一些实施例中,组合氨基酸变化对CasX蛋白功能的影响是协同的。如本文所用,具有协同作用的变异体组合是指当单独分析时其对功能的影响大于每个单独氨基酸变化的影响的总和的组合。在一些实施例中,组合氨基酸变化产生CasX变异蛋白,其中CasX蛋白的一种以上功能相对于参考CasX蛋白得到改善。

[0361] t.CasX融合蛋白

[0362] 在一些实施例中,本公开提供包含与CasX融合的异源蛋白的CasX蛋白。在一些情况下,CasX为参考CasX蛋白。在其它情况下,CasX为本文所述的任一实施例的CasX变异体。

[0363] 在一些实施例中,CasX变异蛋白与具有感兴趣的不同的活性的一种或多种蛋白或其结构域融合,产生融合蛋白。例如,在一些实施例中,CasX变异蛋白与抑制转录、修饰靶核酸序列或修饰与核酸相关的多肽(例如,组蛋白修饰)的蛋白质(或其结构域)融合。

[0364] 在一些实施例中，异源多肽(或异源氨基酸，例如半胱氨酸残基或非天然氨基酸)可插入CasX蛋白内的一个或多个位置以产生CasX融合蛋白。在其它实施例中，半胱氨酸残基可插入CasX蛋白内的一个或多个位置，接着结合下文所述的异源多肽。在一些替代实施例中，异源多肽或异源氨基酸可在参考或CasX变异蛋白的N端或C端处添加。在其它实施例中，异源多肽或异源氨基酸可插入CasX蛋白的序列内部。

[0365] 在一些实施例中，参考CasX或变异融合蛋白保留RNA引导序列特异性靶核酸结合及裂解活性。在一些情况下，参考CasX或变异融合蛋白具有(保留)不具有异源蛋白插入的对应参考CasX或变异蛋白的活性(例如裂解和/或结合活性)的50％或更大。在一些情况下，参考CasX融合蛋白或CasX变异融合蛋白保持不具有异源蛋白插入的对应CasX蛋白的至少约60％、或至少约70％或更大、至少约80％、或至少约90％、或至少约92％、或至少约95％、或至少约98％、或至少约100％的活性(例如，裂解和/或结合活性)。

[0366] 在一些情况下，相对于不具有插入的异源氨基酸或异源多肽的CasX蛋白的活性，参考CasX或变异融合蛋白保持(具有)靶核酸结合活性。在一些情况下，参考CasX或变异融合蛋白保持不具有异源蛋白插入的对应CasX蛋白的至少约60％、或至少约70％或更大、至少约80％、或至少约90％、或至少约92％、或至少约95％、或至少约98％、或至少约100％的结合活性。

[0367] 在一些情况下，相对于不具有插入的异源氨基酸或异源多肽的亲本CasX蛋白的活性，参考CasX或变异融合蛋白保持(具有)靶核酸结合和/或裂解活性。例如，在一些情况下，参考CasX或变异融合蛋白具有(保持)对应亲本CasX蛋白(不具有插入的CasX蛋白)的50％或更大的结合和/或裂解活性。例如，在一些情况下，参考CasX或变异融合蛋白具有(保持)对应CasX亲本蛋白质(不具有插入的CasX蛋白)的60％或更大(70％或更大、80％或更大、90％或更大、92％或更大、95％或更大、98％或更大、或100％)的结合和/或裂解活性。测量CasX蛋白和/或CasX融合蛋白的裂解和/或结合活性的方法是所属领域普通技术人员已知的并且可以使用任何方便的方法。

[0368] 多种异源多肽适合包括于本公开的参考CasX或CasX变异融合蛋白中。在一些情况下，融合伴侣可调节靶DNA的转录(例如抑制转录、增加转录)。举例来说，在一些情况下，融合伴侣为抑制转录的蛋白质(或来自蛋白质的结构域)(例如转录抑制因子，一种经由募集转录抑制剂蛋白、修饰靶DNA(例如甲基化)、募集DNA修饰剂、调节与靶DNA相关的组蛋白、募集组蛋白修饰剂(例如修饰组蛋白的乙酰化和/或甲基化的那些)等起作用的蛋白质)。在一些情况下，融合伴侣为增加转录的蛋白质(或来自蛋白质的结构域)(例如，转录活化因子，一种经由募集转录激活因子蛋白、修饰靶DNA(例如，去甲基化)、募集DNA修饰剂、调节与靶DNA相关的组蛋白、募集组蛋白修饰剂(例如，修饰组蛋白的乙酰化和/或甲基化的那些)等起作用的蛋白质)。

[0369] 在一些情况下，融合伴侣具有修饰靶核酸序列的酶活性；例如，核酸酶活性、甲基转移酶活性、去甲基酶活性、DNA修复活性、DNA损伤活性、脱氨基活性、岐化酶活性、烷基化活性、脱嘌呤活性、氧化活性、嘧啶二聚体形成活性、整合酶活性、转座酶活性、重组酶活性、聚合酶活性、连接酶活性、解螺旋酶活性、光裂合酶活性或糖基化酶活性。

[0370] 在一些情况下，融合伴侣具有修饰与靶核酸相关的多肽(例如，组蛋白)的酶活性(例如，甲基转移酶活性、脱甲基酶活性、乙酰转移酶活性、脱乙酰基酶活性、激酶活性、磷酸

酶活性、泛素蛋白连接酶活性、去泛素化活性、腺苷酸化活性、去腺苷酸化活性、SUMO化活性、去SUMO化活性、核糖基化活性、去核糖基化活性、豆蔻酰化活性或去豆蔻酰化活性)。

[0371] 可用作融合伴侣以增加转录的蛋白质(或其片段)的实例包括但不限于:转录活化子,如VP16、VP64、VP48、VP160、p65子结构域(例如,来自NFkB)和EDLL的激活结构域和/或TAL激活结构域(例如,用于植物中的活性);组蛋白氨酸赖氨酸甲基转移酶,如含有SET结构域的1A,组蛋白氨酸赖氨酸甲基转移酶(SET1A)、含有SET结构域的1B,组蛋白氨酸赖氨酸甲基转移酶(SET1B)、氨酸赖氨酸甲基转移酶2A(MLL1至5、ASCL1(ASH1)无刚毛鳞甲(achaete-scute)家族bHLH转录因子1(ASH1)、含有SET和MYND结构域2(SYMD2)、核受体结合SET结构域蛋白1(NSD1)及其类似物;组蛋白氨酸赖氨酸去甲基酶,如氨酸赖氨酸去甲基酶3A(JHDM2a)/氨酸赖氨酸特异性去甲基酶3B(JHDM2b)、氨酸赖氨酸去甲基酶6A(UTX)、氨酸赖氨酸去甲基酶6B(JMJD3)及其类似物;组蛋白乙酰基转移酶,如氨酸赖氨酸乙酰转移酶2A(GCN5)、氨酸赖氨酸乙酰转移酶2B(PCAF)、CREB结合蛋白(CBP)、E1A结合蛋白p300(p300)、TATA-盒结合蛋白相关因子1(TAF1)、氨酸赖氨酸乙酰转移酶5(TIP60/PLIP)、氨酸赖氨酸乙酰转移酶6A(MOZ/MYST3)、氨酸赖氨酸乙酰转移酶6B(MORF/MYST4)、SRC原癌基因、非受体酪氨酸激酶(SRC1)、核受体共活化剂3(ACTR)、MYB结合蛋白1a(P160)、时钟昼夜节律调节因子(CLOCK)及其类似物;及DNA去甲基酶,如十-十一易位(TET)双加氧酶1(TET1CD)、tet甲基胞嘧啶双加氧酶1(TET1)、demeter(DME)、demeter类似物1(DML1)、demeter类似物2(DML2)、蛋白质ROS1(ROS1)及其类似物。

[0372] 可用作融合伴侣以减少转录的蛋白质(或其片段)的实例包括但不限于:转录抑制子,如Kruppel相关盒(KRAB或SKD);KOX1抑制结构域;Mad mSIN3相互作用结构域(SID);ERF抑制子结构域(ERD)、SRDX抑制结构域(例如,用于植物中的抑制)及其类似物;组蛋白氨酸赖氨酸甲基转移酶,如含PR/SET结构域的蛋白质(Pr-SET7/8)、氨酸赖氨酸甲基转移酶5B(SUV4-20H1)、PR/SET结构域2(RIZ1)及其类似物;组蛋白氨酸赖氨酸去甲基酶,如氨酸赖氨酸去甲基酶4A(JMJD2A/JHDM3A)、氨酸赖氨酸去甲基酶4B(JMJD2B)、氨酸赖氨酸去甲基酶4C(JMJD2C/GASC1)、氨酸赖氨酸去甲基酶4D(JMJD2D)、氨酸赖氨酸去甲基酶5A(JARID1A/RBP2)、氨酸赖氨酸去甲基酶5B(JARID1B/PLU-1)、氨酸赖氨酸去甲基酶5C(JARID 1C/SMCX)、氨酸赖氨酸去甲基酶5D(JARID1D/SMCY)及其类似物;组蛋白氨酸赖氨酸脱乙酰基酶,如组蛋白脱乙酰基酶1(HDAC1)、HDAC2、HDAC3、HDAC8、HDAC4、HDAC5、HDAC7、HDAC9、长寿蛋白1(SIRT1)、SIRT2、HDAC11及其类似物;DNA甲基化酶,如HhaI DNA m5c-甲基转移酶(M.HhaI)、DNA甲基转移酶1(DNMT1)、DNA甲基转移酶3a(DNMT3a)、DNA甲基转移酶3b(DNMT3b)、甲基转移酶1(MET1)、S-腺苷-L-甲硫氨酸依赖性甲基转移酶超家族蛋白(DRM3)(植物)、DNA胞嘧啶甲基转移酶MET2a(ZMET2)、染色质甲基化酶1(CMT1)、染色质甲基化酶2(CMT2)(植物)及其类似物;及边缘募集元件,如核纤层蛋白A、核纤层蛋白B及其类似物。

[0373] 在一些情况下,融合伴侣具有修饰靶核酸序列(例如,ssRNA、dsRNA、ssDNA、dsDNA)的酶活性。可由融合伴侣提供的酶活性的实例包括但不限于:核酸酶活性,例如由限制酶(例如,FokI核酸酶)提供;甲基转移酶活性,例如由甲基转移酶(例如,HhaI DNA m5c-甲基转移酶(M.HhaI)、DNA甲基转移酶1(DNMT1)、DNA甲基转移酶3a(DNMT3a)、DNA甲基转移酶3b(DNMT3b)、METI、DRM3(植物)、ZMET2、CMT1、CMT2(植物)等)提供;去甲基酶活性,例如由去甲基酶(例如,十-十一易位(TET)双加氧酶1(TET 1CD)、TET1、DME、DML1、DML2、ROS1等)提供;

DNA修复活性;DNA损伤活性;脱氨基活性,如由脱氨酶(例如,胞嘧啶脱氨酶,例如APOBEC蛋白,如大鼠APOBEC1)提供;岐化酶活性;烷基化活性;脱嘌呤活性;氧化活性;嘧啶二聚体形成活性;整合酶活性,例如由整合酶和/或解离酶(例如,Gin转化酶,如Gin转化酶的高度活化突变体GinH106Y;人类免疫缺陷病毒1型整合酶(IN);Tn3解离酶等)提供;转座酶活性;重组酶活性,例如由重组酶(例如,Gin重组酶的催化结构域)提供;聚合酶活性;连接酶活性;解螺旋酶活性;光裂合酶活性及糖基化酶活性)。

[0374] 在一些情况下,本公开的参考CasX或CasX变异蛋白质与选自以下的多肽融合:增加转录的结构域(例如,VP16结构域、VP64结构域)、减少转录的结构域(例如,KRAB结构域,例如来自Kox1蛋白)、组蛋白乙酰转移酶(例如,组蛋白乙酰转移酶p300)的核心催化结构域、提供可检测信号的蛋白质/结构域(例如,荧光蛋白,如GFP)、核酸酶结构域(例如,Fok1核酸酶)或碱基编辑剂(例如,胞苷脱氨酶,如APOBEC1)。

[0375] 在一些情况下,融合伴侣具有修饰与靶核酸(例如,ssRNA、dsRNA、ssDNA、dsDNA)相关的蛋白质(例如,组蛋白、RNA结合蛋白、DNA结合蛋白等)的酶活性。可由由融合伴侣提供的酶活性(修饰与靶核酸相关的蛋白质)的实施例包括但不限于:甲基转移酶活性,如由组蛋白甲基转移酶(HMT)(例如,杂色抑制子3-9同源物1(SUV39H1,亦称为KMT1A)、常染色质组蛋白赖氨酸甲基转移酶2(G9A,亦称为KMT1C和EHMT2)、SUV39H2、ESET/SETDB 1及其类似物、SET1A、SET1B、MLL1至5、ASH1、SYMD2、NSD1、DOT1L,Pr-SET7/8,SUV4-20H1,EZH2,RIZ1提供;去甲基酶活性,如由组蛋白去甲基酶(例如,氨酸赖氨酸去甲基酶1A(KDM1A,亦称为LSD1)、JHDM2a/b、JMJD2A/JHDM3A、JMJD2B、JMJD2C/GASC1、JMJD2D、JARID1A/RBP2、JARID1B/PLU-1、JARID1C/SMCX、JARID1D/SMCY、UTX、JMJD3及其类似物)提供;乙酰转移酶活性,如由组蛋白乙酰酶转移酶(例如,人类乙酰转移酶p300、GCN5、PCAF、CBP、TAF1、TIP60/PLIP、MOZ/MYST3、MORF/MYST4、HBO1/MYST2、HMOF/MYST1、SRC1、ACTR、P160、CLOCK及其类似物的催化核/片段)提供;脱乙酰基酶活性,如由组蛋白脱乙酰基酶(例如,HDAC1、HDAC2、HDAC3、HDAC8、HDAC4、HDAC5、HDAC7、HDAC9、SIRT1、SIRT2、HDAC11及其类似物)提供;激酶活性;磷酸酶活性;泛素蛋白连接酶活性;去泛素化活性;腺苷酸化活性;去腺苷酸化活性;SUMO化活性;去SUMO化活性;核糖基化活性;去核糖基化活性;豆蔻酰化活性;及去豆蔻酰化活性。

[0376] 适合的融合伴侣的额外实例为(i)二氢叶酸还原酶(DHFR)去稳定化结构域(例如以生成化学可控的主题RNA引导多肽或条件活性RNA引导多肽),以及(ii)叶绿体转运肽。

[0377] 适合的叶绿体转运肽包括但不限于:

[0378] MASMISSSAVTTVSRASRGQSAAMAPFGGLKSMTGFPVRKVNTDITSITSNGGRVKCMQVWPPIGKKKFETLSYLPPLTRDSRA(SEQ ID NO:151);

[0379] MASMISSSAVTTVSRASRGQSAAMAPFGGLKSMTGFPVRKVNTDITSITSNGGRVKS(SEQ ID NO:152);

[0380] MASSMLSSATMVASPAQATMVAPFNGLKSSAAFPATRKANNDITSITSNGGRVNCMQVWPPIEKKKFETLSYLPDLTDSGGRVNC(SEQ ID NO:153);

[0381] MAQVSRICNGVQNPSLISNLSKSSQRKSPLSVSLKTQQHPRAYPISSSWGLKKSGMTLIGSELRPLKVMSSVSTAC(SEQ ID NO:154);

[0382] MAQVSRICNGVWNPSLISNLSKSSQRKSPLSVSLKTQQHPRAYPISSSWGLKKSGMTLIGSELRPLKVMSSVSTAC(SEQ ID NO:155);

[0383] MAQINNMAQGIQTLNPNSNFHKPQVPKSSSFLVFGSKKLKNSANSMLVLKKDSIFMQLFCSFRISASVATAC(SEQ ID NO:156)；

[0384] MAALVTSQLATSGTVLSVTDRFRRPGFQGLRPRNPADAALGMRTVGASAAPKQSRKPHRFDRRCLSMVV(SEQ ID NO:157)；

[0385] MAALTTSQLATSATGFGIADRSAPSSLLRHGFQGLKPRSPAGGDATSLSVTTSARATPKQQRSVQRGSRRFPSVVVC(SEQ ID NO:158)；

[0386] MASSVLSSAAVATRSNVAQANMVAPFTGLKSAASFPVSRKQNLDITSIASNGGRVQC(SEQ ID NO:159)；

[0387] MESLAATSVFAPSRVAVPAARALVRAGTVVPTRRTSSTSGTSGVKCSAAVTPQASPVISRSAAAA(SEQ ID NO:160)；和

[0388] MGAAATSMQSLKFSNRLVPPSRRLSPVPNNVTCNNLPKSAAPVRTVKCCASSWNSTINGAAATTNGASAASS(SEQ ID NO:161)。

[0389] 在一些情况下，本公开的参考CasX或变异多肽可包括内体逃逸肽。在一些情况下，内体逃逸多肽包含氨基酸序列GLFXALLXLLXSLWXLLLXA(SEQ ID NO:162)，其中每个X独立地选自赖氨酸、组氨酸和精氨酸。在一些情况下，内体逃逸多肽包含氨基酸序列GLFHALLHLLHSLWHLLLHA(SEQ ID NO:163)或HHHHHHHHH(SEQ ID NO:164)。

[0390] 当靶向ssRNA靶核酸序列时使用的融合伴侣的非限制性实例包括但不限于：剪接因子(例如，RS结构域)；蛋白质翻译组分(例如，翻译起始、伸长和/或释放因子；例如eIF4G)；RNA甲基化酶；RNA编辑酶(例如，RNA脱氨基酶，例如作用于RNA的腺苷脱胺酶(ADAR)，包括A至I和/或C至U编辑酶)；解螺旋酶；RNA结合蛋白；及其类似物。应理解，异源多肽可包括整个蛋白质，或在一些情况下可包括蛋白质片段(例如功能结构域)。

[0391] 融合伴侣可为能够与ssRNA相互作用的任何结构域(出于本公开的目的，其包括分子内和/或分子间二级结构，例如双链RNA双螺旋体，如发夹、茎环等)，无论是瞬时的还是不可逆的、直接的或间接的，包括但不限于选自由以下构成的组的效应结构域：核酸内切酶(例如来自SMG5和SMG6等蛋白质的RNase III、CRR22 DYW结构域、Dicer和PIN(PilT N-末端)结构域)；负责刺激RNA裂解的蛋白及蛋白结构域(例如CPSF、CstF、CFIm和CFIIm)；核酸外切酶(例如XRN-1或核酸外切酶T)；脱腺苷化酶(例如HNT3)；负责无义介导的RNA衰减的蛋白及蛋白结构域(例如UPF1、UPF2、UPF3、UPF3b、RNP SI、Y14、DEK、REF2和SRm160)；负责稳定RNA的蛋白及蛋白结构域(例如PABP)；负责抑制翻译的蛋白及蛋白结构域(例如Ago2和Ago4)；负责刺激翻译的蛋白及蛋白结构域(例如Staufen)；负责(例如，能够)调节翻译的蛋白及蛋白结构域(例如，翻译因子，如起始因子、伸长因子、释放因子等，例如，eIF4G)；负责RNA的聚腺苷酸化的蛋白及蛋白结构域(例如PAP1、GLD-2和Star-PAP)；负责RNA的聚尿苷化的蛋白及蛋白结构域(例如CI D1和末端尿苷酸转移酶)；负责RNA定位的蛋白及蛋白结构域(例如来自IMP1、ZBP1、She2p、She3p和Bicaudal-D)；负责RNA的核保留的蛋白及蛋白结构域(例如Rrp6)；负责RNA的核输出的蛋白及蛋白结构域(例如TAP、NXF1、THO、TREX、REF和Aly)；负责抑制RNA剪接的蛋白及蛋白结构域(例如PTB、Sam68和hnRNP A1)；负责刺激RNA剪接的蛋白及蛋白结构域(例如富含丝氨酸/精氨酸(SR)的结构域)；负责降低转录效率的蛋白及蛋白结构域(例如FUS(TLS))；及负责刺激转录的蛋白及蛋白结构域(例如CDK7和HIV Tat)。替代地，效应子结构域可选自包含以下的组：核酸内切酶；能够刺激RNA裂解的蛋白和蛋白

结构域;核酸外切酶;脱腺苷化酶;具有无义介导的RNA衰减活性的蛋白和蛋白结构域;能够稳定化RNA的蛋白和蛋白结构域;能够抑制翻译的蛋白和蛋白结构域;能够刺激翻译的蛋白和蛋白结构域;能够调节翻译的蛋白和蛋白结构域(例如翻译因子,例如起始因子、伸长因子、释放因子等,例如eIF4G);能够对RNA进行聚腺苷酸化的蛋白和蛋白结构域;能够对RNA进行聚尿苷化的蛋白和蛋白结构域;具有RNA定位活性的蛋白和蛋白结构域;能够对RNA进行核保留的蛋白和蛋白结构域;具有RNA核导出活性的蛋白和蛋白结构域;能够抑制RNA剪接的蛋白和蛋白结构域;能够刺激RNA剪接的蛋白和蛋白结构域;能够降低转录效率的蛋白和蛋白结构域;和能够刺激转录的蛋白和蛋白结构域。另一适合的异源多肽为PUF RNA结合结构域,其更详细地描述于以全文引用的方式并入本文中的WO2012068627中。

[0392] 可用作(以整体或其片段形式)融合伴侣的RNA剪接因子具有模块化组织,其具有独立的序列特异性RNA结合模块及剪接效应子结构域。举例来说,富含丝氨酸/精氨酸(SR)的蛋白质家族的成员含有结合至前mRNA中的外显子剪接增强子(ESE)的N端RNA识别基序(RRM)及促进外显子包涵的C端RS结构域。作为另一实例,hnRNP蛋白hnRNP A1经由其RRM结构域结合至外显子剪接沉默子(ESS),且经由C-末端富含甘氨酸的结构域抑制外显子包含。一些剪接因子可通过结合至两个替代位点之间的调节序列而调节剪接位点的替代使用。例如,ASF/SF2可识别ESE且促进使用内含子近侧位点,而hnRNP AI可结合至ESS且使剪接转向使用内含子远侧位点。此类因子的一种应用为产生调节内源基因,尤其是疾病相关基因的替代性剪接的ESF。例如,Bcl-x前mRNA产生两种剪接异构体,具有两个可选的5'剪接位点,以编码功能相反的蛋白质。长剪接同功异型物Bcl-xL为强力细胞凋亡抑制剂,其表达于长寿命的有丝分裂后细胞中且在许多癌细胞中上调,保护细胞免受凋亡信号影响。短同功异型物Bcl-xS为促细胞凋亡同功异型物,且在具有高周转率(例如产生淋巴细胞)的细胞中以高表达量表达。通过位于核外显子区域或外显子延伸区域(即,在两个替代5'剪接位点之间)中的多个cc-元件调节两种Bcl-x剪接的比。关于更多实例,参见WO2010075303,其以全文引用的方式并入本文中。

[0393] 其它适合的融合伴侣包括但不限于作为边界元件的蛋白质(或其片段)(例如,CTCF)、提供边缘募集的蛋白质及其片段(例如,核纤层蛋白A、核纤层蛋白B等)及蛋白质对接元件(例如,FKBP/FRB、Pill/Abyl等)。

[0394] 在一些情况下,异源多肽(融合伴侣)提供亚细胞定位,即异源多肽含有亚细胞定位序列(例如,用于靶向至细胞核的核定位信号(NLS);保持融合蛋白在细胞核的外的序列,例如核输出序列(NES);保持融合蛋白滞留于细胞质中的序列;用于靶向至线粒体的线粒体定位信号;用于靶向至叶绿体的叶绿体定位信号;ER滞留信号;等)。在一些实施例中,主题RNA引导多肽或条件活性RNA引导多肽和/或主题CasX融合蛋白不包括NLS,以使得蛋白质不靶向至细胞核(这可能是有利的;例如当靶核酸序列是存在于细胞溶质中的RNA时)。在一些实施例中,融合伴侣可提供标签(即,异源多肽为可检测标记)以易于跟踪和/或纯化(例如荧光蛋白,例如绿色荧光蛋白(GFP)、黄色荧光蛋白(YFP)、红色荧光蛋白(RFP)、强化型蓝荧光蛋白(CFP)、mCherry、tdTomato及其类似物;组氨酸标签,例如6×His标签;血凝素(HA)标签;FLAG标签;Myc标签;及其类似物)。

[0395] 在一些情况下,参考或CasX变异多肽包括(融合至)核定位信号(NLS)(例如,在一些情况下,2个或更多、3个或更多、4个或更多、5个或更多、6个或更多、7个或更多、8个或更

多NLS)。因此,在一些情况下,参考或CasX变异多肽包括一个或多个NLS(例如,2个或更多、3个或更多、4个或更多、或5个或更多NLS)。在一些情况下,一个或多个NLS(2个或更多个、3个或更多个、4个或更多个、或5个或更多个NLS)位于N端和/或C端处或附近(例如在其50个氨基酸内)。在一些情况下,一个或多个NLS(2个或更多个、3个或更多个、4个或更多个、或5个或更多个NLS)位于N端处或附近(例如,在其50个氨基酸内)。在一些情况下,一个或多个NLS(2个或更多个、3个或更多个、4个或更多个、或5个或更多个NLS)位于C端处或附近(例如,在其50个氨基酸内)。在一些情况下,一个或多个NLS(3个或更多个、4个或更多个、或5个或更多个NLS)位于N-末端及C-末端处或其附近(例如,在其50个氨基酸内)。在一些情况下,一个NLS位于N端且一个NLS位于C端。在一些情况下,参考或CasX变异多肽包括(融合至)1至10个NLS(例如,1至9、1至8、1至7、1至6、1至5、2至10、2至9、2至8、2至7、2至6或2至5个NLS)。在一些情况下,参考或CasX变异多肽包括(融合至)2至5个NLS(例如,2至4或2至3个NLS)。

[0396] NLS的非限制性实例包括衍生自以下的序列:SV40病毒大T-抗原的NLS,具有氨基酸序列PKKKRKV(SEQ ID NO:165);来自核质蛋白的NLS(例如,具有序列KRPAATKKAGQAKKKK(SEQ ID NO:166)的双分型核质蛋白NLS;具有氨基酸序列PAAKRVKLD(SEQ ID NO:167)或RQRRNELKRSP(SEQ ID NO:168)的c-myc NLS;hRNPA1M9 NLS,其具有序列NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY(SEQ ID NO:169);来自输入蛋白-α的IBB结构域的序列RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV(SEQ ID NO:170);肌瘤T蛋白的序列VSRKRPRP(SEQ ID NO:171)及PPKKARED(SEQ ID NO:172);人类p53的序列PQPKKKPL(SEQ ID NO:173);小鼠c-abl IV的序列SALIKKKKKMAP(SEQ ID NO:174);流感病毒NS1的序列DRLRR(SEQ ID NO:175)及PKQKKRK(SEQ ID NO:176);丁型肝炎病毒抗原的序列RKLKKKIKKL(SEQ ID NO:177);小鼠Mx1蛋白的序列REKKKFLKRR(SEQ ID NO:178);人类聚(ADP-核糖)聚合酶的序列KRKGDEVDGVDEVAKKKSKK(SEQ ID NO:179);类固醇激素受体(人类)糖皮质激素的序列RKCLQAGMNLEARKTKK(SEQ ID NO:180);博纳病(Borna disease)病毒P蛋白(BDV-P1)的序列PRPRKIPR(SEQ ID NO:181);丙型肝炎病毒非结构蛋白(HCV-NS5A)的序列PPRKKRTVV(SEQ ID NO:182);LEF1的序列NLSKKKKRKREK(SEQ ID NO:183);ORF57 simirae的序列RRPSRPFRKP(SEQ ID NO:184);EBV LANA的序列KRPRSPSS(SEQ ID NO:185);甲型流感蛋白的序列KRGINDRNFWRGENERKTR(SEQ ID NO:186);人类RNA解螺旋酶A(RHA)的序列PRPPKMARYDN(SEQ ID NO:187);核仁RNA解螺旋酶II的序列KRSFSKAF(SEQ ID NO:188);TUS-蛋白的序列KLKIKRPVK(SEQ ID NO:189);与输入蛋白-α相关的序列PKKKRKVPPPPAAKRVKLD(SEQ ID NO:190);来自HTLV-1中的Rex蛋白的序列PKTRRRPRRSQRKRPPT(SEQ ID NO:191);来自秀丽隐杆线虫(Caenorhabditis elegans)的EGL-13蛋白的序列MSRRRKANPTKLSENAKKLAKEVEN(SEQ ID NO:192);及序列KTRRRPRRSQRKRPPT(SEQ ID NO:193)、RRKKRRPRRKKRR(SEQ ID NO:194)、PKKKSRKPKKKSRK(SEQ ID NO:195)、HKKKHPDASVNFSEFSK(SEQ ID NO:196)、QRPGPYDRPQRPGPYDRP(SEQ ID NO:197)、LSPSLSPLLSPSLSPL(SEQ ID NO:198)、RGKGGKGLGKGGAKRHRK(SEQ ID NO:199)、PKRGRGRPKRGRGR(SEQ ID NO:200)、PKKKRKVPPPPAAKRVKLD(SEQ ID NO:190)、PKKKRKVPPPPKKKRKV(SEQ ID NO:201)、来自CPV的序列PAKRARRGYKC(SEQ ID NO:202)、来自B19的序列KLGPRKATGRW(SEQ ID NO:203)和来自hBOV的序列PRRKREE(SEQ ID NO:204)。一般来说,NLS(或多个NLS)具有足以在真核细胞的细胞核中驱动参考或CasX变异融合蛋白的

积聚的强度。可通过任何适合的技术进行细胞核中的积聚的检测。举例来说，可检测标记物可与参考或CasX变异融合蛋白融合，使得可观测到细胞内的位置。细胞核亦可自细胞分离，可接着通过任何适合于检测蛋白质的方法，例如免疫组织化学、蛋白质印迹或酶活性分析来分析其内容。亦可确定细胞核中的积聚。

[0397] 在一些情况下，参考或CasX变异体融合蛋白包括“蛋白质转导结构域”或PTD(亦称为CPP-细胞穿透肽)，其是指促进穿越脂质双层、胶束、细胞膜、细胞器膜或囊泡膜的蛋白质、多核苷酸、碳水化合物或有机或无机化合物。连接至另一分子(其可在小极性分子至大型大分子和/或纳米粒子范围内)的PTD促进分子穿越膜，例如自细胞外空间进入细胞内空间，或自胞溶质进入细胞器内。在一些实施例中，PTD共价连接至参考或CasX变异体融合蛋白的氨基末端。在一些实施例中，PTD共价连接至参考或CasX变异体融合蛋白的羧基末端。在一些情况下，PTD在适合的插入位点处插入参考或CasX变异体融合蛋白的序列内部。在一些情况下，参考或CasX变异体融合蛋白包括(缀合至、融合至)一个或多个PTD(例如两个或更多个、三个或更多个、四个或更多个PTD)。在一些情况下，PTD包括一个或多个核定位信号(NLS)。PTD的实例包括但不限于包含YGRKKRRQRRR(SEQ ID NO:205)、RKKRRQRR(SEQ ID NO:206)、YARAAARQARA(SEQ ID NO:207)、THRLPRRRRRR(SEQ ID NO:208)及GGRRARRRRRR(SEQ ID NO:209)的HIV TAT的肽转导结构域；包含足以直接进入细胞的多个精氨酸(例如，3、4、5、6、7、8、9、10或10至50个精氨酸(SEQ ID NO:210))的聚精氨酸序列；VP22结构域(Zender等人(2002)《癌症基因治疗(Cancer Gene Ther.)》9(6):489-96)；果蝇触角足蛋白转导结构域(Noguchi等人(2003)《糖尿病(Diabetes)》52(7):1732-1737)；截短人类降钙素肽(Trehin等人(2004)《药学研究(Pharm.Research)》21:1248-1256)；聚赖氨酸(Wender等人(2000)《美国国立科学院院报(Proc.Natl.Acad.Sci.USA)97:13003-13008)；RRQRRTSKLMKR(SEQ ID NO:211)；运输蛋白GWTLNSAGYLLGKINLKALAALAKKIL(SEQ ID NO:212)；KALAWEAKLAKALAKALAKHLAKALAKALKCEA(SEQ ID NO:213)；及RQIKIWFQNRRMKWKK(SEQ ID NO:214)。在一些实施例中，PTD为可活化CPP(ACPP)(Aguilera等人(2009)《综合生物学(Integr Biol(Camb))》6月;1(5-6):371-381)。ACPP包含经由可裂解接头连接至匹配的聚阴离子(例如Glu9或“E9”)的聚阳离子CPP(例如Arg9或“R9”)，其将净电荷降至接近零且因此抑制黏附及吸收至细胞中。在接头裂解之后，聚阴离子释放，局部揭露聚精氨酸及其固有黏附性，因此“活化”ACPP以穿过膜。

[0398] 在一些实施例中，参考或CasX变异体融合蛋白可包括经由接头多肽(例如一个或多个接头多肽)连接至内部插入的异源氨基酸或异源多肽(异源氨基酸序列)的CasX蛋白。在一些实施例中，参考或CasX变异体融合蛋白可经由接头多肽(例如一个或多个接头多肽)在C末端和/或N末端连接至异源多肽(融合伴侣)。接头多肽可具有多个氨基酸序列中的任一个。蛋白质可通过一般具有柔性性质之间隔肽连接，但不排除其它化学键。适合的接头包括长度为4个氨基酸至40个氨基酸，或长度为4个氨基酸至25个氨基酸的多肽。这些接头一般通过使用合成、编码接头的寡核苷酸产生以偶联蛋白质。可使用具有一定程度的柔性的肽接头。连接肽可具有几乎任何氨基酸序列，应记住，优选接头将具有产生总体柔性肽的序列。使用小氨基酸，例如甘氨酸及丙氨酸在产生柔性肽中有用。产生此类序列对于所属领域技术人员为常规的。多种不同接头为市售的且被视为适合使用。示例接头多肽包括甘氨酸聚合物(G)n、甘氨酸-丝氨酸聚合物(包括例如，(GS)n、GSGGSn(SEQ ID NO:215)、GGSGGSn

(SEQ ID NO:216)及GGGSn(SEQ ID NO:217),其中n为至少一的整数)、甘氨酸-丙氨酸聚合物、丙氨酸-丝氨酸聚合物、甘氨酸-脯氨酸聚合物、脯氨酸聚合物及脯氨酸-丙氨酸聚合物。示例接头可包含氨基酸序列,其包括但不限于GGSG(SEQ ID NO:218)、GGSGG(SEQ ID NO:219)、GSGSG(SEQ ID NO:220)、GSGGG(SEQ ID NO:221)、GGGSG(SEQ ID NO:222)、GSSSG(SEQ ID NO:223)、GPGP(SEQ ID NO:224)、GGP、PPP、PPAPPA(SEQ ID NO:225)、PPPGPPP(SEQ ID NO:226)等。普通技术人员应认识到,结合至上文所述的任何元件的肽的设计可包括完全或部分柔性的接头,以使得接头可包括柔性接头以及一个或多个赋予较不可挠结构的部分。

[0399] V.用于修饰C9orf72基因的CasX:gNA系统和方法

[0400] 本文提供的CasX蛋白、引导核酸及其变异体可用于各种应用,包括作为治疗学、诊断学和用于研究。为了实现本公开的基因编辑方法,本文提供了可编程的CasX:gNA系统。本文所提供的CasX:gNA系统的可编程性质允许精确靶向,以在编码C9orf72蛋白质的靶核酸序列、C9orf72调节元件、C9orf72基因的非编码区或这两者中的预定关注的一个或多个区处实现所需作用(切割、裂解等)。在一些实施例中,本文所提供的CasX:gNA系统包含表4、6、7、8或10中所阐述的SEQ ID NO:49-150、233-235、238-252或272-281中的任一个的CasX变异体,或与其具有至少60%同一性、至少70%同一性、至少80%同一性、至少81%同一性、至少82%同一性、至少83%同一性、至少84%同一性、至少85%同一性、至少86%同一性、至少86%同一性、至少87%同一性、至少88%同一性、至少89%同一性、至少89%同一性、至少90%同一性、至少91%同一性、至少92%同一性、至少93%同一性、至少94%同一性、至少95%同一性、至少96%同一性、至少97%同一性、至少98%同一性、至少99%同一性或至少99.5%同一性的变异序列,gNA支架包含如表2中所阐述的SEQ ID NO:2101-2294中的任一个的序列,或与其具有至少65%同一性、至少70%同一性、至少75%同一性、至少80%同一性、至少81%同一性、至少82%同一性、至少83%同一性、至少84%同一性、至少85%同一性、至少86%同一性、至少86%同一性、至少87%同一性、至少88%同一性、至少89%同一性、至少89%同一性、至少90%同一性、至少91%同一性、至少92%同一性、至少93%同一性、至少94%同一性、至少95%同一性、至少96%同一性、至少97%同一性、至少98%同一性、至少99%同一性、至少99.5%同一性的序列,并且gNA包含SEQ ID NO:309-343、363-2100或2295-21835中的任一个的靶向序列,或与其具有至少65%同一性、至少70%同一性、至少75%同一性、至少80%同一性、至少85%同一性、至少90%同一性或至少95%同一性,并且具有15与30个之间的核苷酸的序列。在一些实施例中,gNA的靶向序列与编码SEQ ID NO:227或228的C9orf72蛋白质的一个或多个突变,或破坏C9orf72蛋白质的功能或表达的一个或多个突变的靶核酸序列杂交。在另一实施例中,gNA的靶向序列与包含在六核苷酸重复序列GGGGCC或其互补序列的5'或3'处的序列的靶核酸序列杂交。在其它实施例中,gNA的靶向序列与包含C9orf72基因的调节元件的靶核酸序列杂交。在一些实施例中,gNA的靶向序列具有与C9orf72外显子序列杂交的序列。在一些实施例中,gNA的靶向序列具有与C9orf72内含子序列杂交的序列。在一些实施例中,gNA的靶向序列具有与C9orf72基因的内含子1杂交的序列。在一些实施例中,多种gNA的靶向序列具有与C9orf72内含子-外显子接界序列、C9orf72调节元件、C9orf72编码区、C9orf72非编码区或其组合杂交的序列。在所述方法的一些实施例中,gNA经化学修饰。在其它实施例中,本公开提供了编码前述CasX变异蛋白和gNA的一种或多种多核苷酸。在一些情况下,CasX:gNA系统进一步包含供体模板核

酸,其中供体模板可以通过宿主细胞的HDR或HITI修复机制插入以敲减或敲除C9orf72基因,或在其它情况下校正突变;例如通过删除突变HRS重复序列和插入具有GGGGCC序列的10与30个之间的重复序列的HRS。

[0401] 在一些实施例中,本文所提供的CasX:gNA系统包含CasX蛋白和gNA,或编码CasX蛋白和gNA的一种或多种多核苷酸,其中gNA的靶向序列与编码C9orf72蛋白质的靶核酸序列、C9orf72调节元件、C9orf72基因的非编码区(例如内含子1)、桥接这些区域的序列互补,并且因此能够与其杂交,或能够与或与与其互补的序列杂交。在具体实施例中,gNA的靶向序列与HRS或在HRS的5'或3'处的区域内的序列互补,并且因此能够与其杂交。在另一特定实施例中,gNA的靶向序列与C9orf72的启动子内的序列互补,并且因此能够与其杂交。可用于靶向C9orf72 HRS的示范性但非限制性靶向序列包括如表15中所阐述的SEQ ID NO:309-343。在一些实施例中,靶向序列包含SEQ ID NO:309-343的序列。在一些实施例中,CasX:gNA系统包含选自SEQ ID NO:309-343的两个靶向序列,且两个靶向序列不相同。在一些实施例中,CasX:gNA系统包含两个靶向序列,其中第一靶向序列包含SEQ ID NO:310且第二靶向序列选自由SEQ ID NO:321-324组成的组。在一些实施例中,CasX:gNA系统包含两个靶向序列,其中第一靶向序列包含SEQ ID NO:319且第二靶向序列选自由SEQ ID NO:321-325组成的组。在一些实施例中,CasX:gNA系统包含两个靶向序列,其中第一靶向序列包含SEQ ID NO:320且第二靶向序列选自由SEQ ID NO:321-325组成的组。在一些实施例中,两个靶向序列包含SEQ ID NO:310和321、SEQ ID NO:310和322、SEQ ID NO:310和323、SEQ ID NO:310和324、SEQ ID NO:319和321、SEQ ID NO:319和322、SEQ ID NO:319和323、SEQ ID NO:319和324、SEQ ID NO:319和325、SEQ ID NO:320和321、SEQ ID NO:320和322、SEQ ID NO:320和323、SEQ ID NO:320和324、或SEQ ID NO:320和325。

[0402] 在体外条件下将包含编码本公开的CasX:gNA系统(和任选的供体模板序列)的序列的重组表达载体引入细胞中可以在促进细胞存活和CasX:gNA产生的任何合适的培养基中和任何合适的培养条件下进行。将重组表达载体引入靶细胞中可在体内、体外或离体进行。在所述方法的一些实施例中,载体可直接提供至靶宿主细胞。例如,可使细胞与载体接触,使得载体被细胞摄取,所述载体具有编码本文所述的任一实施例的CasX和gNA的核酸,并且任选地具有供体模板序列。使细胞与为质粒的核酸载体接触的方法包括电穿孔、氯化钙转染、显微注射、转导及脂转染,为所属领域中熟知的。对于病毒载体递送,可使细胞与病毒粒子接触,所述病毒粒子包含主题病毒表达载体及编码CasX和gNA的核酸以及任选的供体模板。在一些实施例中,载体是腺相关病毒(AAV)载体,其中AAV选自AAV1、AAV2、AAV3、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAV10、AAV11、AAV12、AAV 44.9、AAV-Rh74或AAVRh10。下文更完整地描述AAV载体的实施例。在其它实施例中,载体是慢病毒载体。逆转录病毒,例如慢病毒,可适用于本公开的方法。常用的逆转录病毒载体是"缺陷性的",例如不能产生生产性感染所需的病毒蛋白,且通常被称为病毒样粒子。确切而言,载体的复制需要包装细胞系中的生长。下文更完整地描述逆转录病毒载体的实施例。

[0403] 在其它实施例中,本公开提供了使用本文所述任一实施例的CasX:gNA系统修饰靶核酸序列的方法,并且所述方法还包含使靶核酸序列与额外的CRISPR蛋白或编码额外的CRISPR蛋白的多核苷酸接触。在一些实施例中,额外的CRISPR蛋白是CasX蛋白,其序列不同于CasX:gNA系统的CasX。在一些实施例中,额外的CRISPR蛋白不是CasX蛋白;例如,额外的

CRISPR蛋白可以是Cpf1、Cas9、Cas12a或Cas13a。

[0404] 在一些实施例中，可能需要在包含突变或重复，例如引起肌肉萎缩性侧索硬化(ALS)和额颞叶型痴呆(FTD)的显性突变或重复的受试者中敲减或敲除C9orf72基因的表达。术语“敲除”是指基因的消除或基因的表达。例如，可以通过缺失或添加导致阅读框破坏的核苷酸序列来敲除基因。作为另一实例，可以通过用不相关或异源的序列替换基因的一部分来敲除基因。如本文所用，术语“敲减”是指基因或其基因产物的表达减少。作为基因敲减的结果，蛋白质活性或功能可能会减弱，或者蛋白质水平可能会降低或消除。在此类实施例中，可以使用具有对编码C9orf72蛋白质的基因的一部分或C9orf72调节元件具有特异性的靶向序列的gNA。根据所使用的CasX蛋白和gNA，所述事件可能是一个裂解事件，允许敲减/敲除表达。在一些实施例中，C9orf72基因表达可通过引入随机插入或缺失(插入缺失)而破坏或消除，例如通过利用不精确的非同源DNA末端连接(NHEJ)修复路径。在此类实施例中，C9orf72的靶向区包括C9orf72基因的编码序列(外显子)，因为在编码序列内插入或删除核苷酸可以产生移码突变。此方法还可用于非编码区(例如内含子)或调节元件中以干扰C9orf72基因的表达。因此，在一些实施例中，本公开提供一种用于改变细胞的一种或多种靶核酸序列的方法中的CasX:gNA系统，所述方法包含使所述细胞与包含本文所述的实施例的CasX蛋白和gNA的CasX:gNA系统接触，其中gRNA包含针对基因组靶的靶向序列，其与编码C9orf72蛋白质的序列、在HRS的5’或3’处的序列、C9orf72调节元件或这些序列的互补序列互补并且因此能够与其杂交。在其它实施例中，本公开提供改变细胞的靶核酸序列的方法，其包含使所述细胞与编码包含本文所描述的实施例的CasX蛋白和gNA的CasX:gNA系统的核酸接触，其中gRNA包含针对基因组靶的靶向序列，其与编码C9orf72蛋白质的序列、在HRS的5’或3’处的序列、C9orf72调节元件或这些序列的互补序列互补并且因此能够与其杂交。在其它实施例中，本公开提供改变细胞的靶核酸序列的方法，其包含使所述细胞与本文描述的实施例的包含编码包含CasX蛋白和gNA的CasX:gNA系统的核酸的载体接触，其中gNA包含靶向序列，其与编码C9orf72蛋白质的序列、在HRS的5’或3’处的序列、C9orf72调节元件或这些序列的互补序列互补并且因此能够与其杂交。在一些实施例中，本公开提供敲减或敲除两个C9orf72等位基因的细胞表达的方法和CasX:gNA系统。在一些实施例中，本公开提供敲减或敲除单个C9orf72等位基因的细胞表达的方法和CasX:gNA系统。在所述方法的其它实施例中，CasX:gNA系统进一步包含对应于C9orf72基因的全部或至少一部分的供体模板核酸，其中所述供体模板核酸包含异源序列，或相比于编码C9orf72的所述部分的基因组核酸序列的一个或多个核苷酸的缺失、插入或突变，其中所述接触引起C9orf72的基因敲减或敲除。在前述中，细胞已经修饰，使得与尚未修饰的细胞相比，HRS的表达降低了至少约10％、至少约20％、至少约30％、至少约40％、至少约50％、至少约60％、至少约70％、至少约80％或至少约90％。在所述方法的其它实施例中，细胞已经修饰，使得细胞不表达可检测水平的HRS RNA或DPR蛋白质。在所述方法的又其它实施例中，供体模板核酸包含校正性序列，所述校正性序列在通过CasX:gNA系统插入到靶核酸中之后，可以表达功能C9orf72蛋白质或生理学上正常水平的C9orf72。

[0405] 因此，本文所述的CasX:gNA系统和方法可以与常规分子生物学方法组合使用以修饰细胞群体(其实例更完整地描述于下文)以产生具有产生功能C9ord72蛋白质的能力的细胞。因此，此方法可以用于产生可以向患有如ALS或FTD的疾病的受试者施用的细胞群体。在

其它实施例中，本文所述的CasX:gNA系统和方法可以用于通过施用编码CasX:gNA组分的系统的组分或载体以修饰受试者的靶细胞的C9orf72基因来治疗受试者。

[0406] VI.多核苷酸和载体

[0407] 在其它实施例中，本公开提供本文所述的编码V型核酸酶蛋白质的多核苷酸和gNA的多核苷酸。在一些实施例中，本公开提供本文所描述的CasX:gNA系统实施例中的任一个的编码CasX蛋白的多核苷酸和gNA(例如gDNA和gRNA)的多核苷酸，以及与编码CasX蛋白和gNA的多核苷酸实施例互补的序列。在额外实施例中，本公开提供编码C9orf72基因的部分或全部的供体模板多核苷酸。在一些情况下，供体模板的C9orf72基因包含用于敲减或敲除靶核酸中的C9orf72基因的突变或异源序列。在其它情况下，供体模板包含用于敲入功能C9orf72基因或其部分的校正性序列。在另外其它实施例中，本公开涉及包含本文所述的编码CasX蛋白和CasX gNA的多核苷酸的载体。在另外其它实施例中，本公开涉及包含含有本文所述的供体模板的多核苷酸的载体。

[0408] 在一些实施例中，本公开提供了编码SEQ ID NO:1-3的参考CasX的多核苷酸序列。在其它实施例中，本公开提供编码本文所述的任何实施例的CasX变异体的多核苷酸序列，包括如表4、6-8和10中所阐述的SEQ ID NO:49-150、233-235、238-252、272-281的CasX蛋白变异体，或与表4的序列具有至少约50％、至少约60％、至少约70％、至少约80％、至少约90％、至少约95％、至少约96％、至少约97％、至少约98％或至少约99％序列同一性的序列，或编码所述变异体的多核苷酸序列的互补序列。在一些实施例中，本公开提供编码本文所描述的任何实施例的gNA序列的经分离多核苷酸序列，包括表1和2的序列，或与其具有至少约50％、至少约60％、至少约70％、至少约80％、至少约90％、至少约95％、至少约96％、至少约97％、至少约98％、至少约99％序列同一性的序列的支架。在一些实施例中，多核苷酸编码选自由以下组成的组的gNA支架序列:SEQ ID NO:2101-2294，或与其具有至少约50％、至少约60％、至少约70％、至少约80％、至少约90％、至少约95％、至少约96％、至少约97％、至少约98％、至少约99％序列同一性的序列。

[0409] 在一些实施例中，本公开提供编码gNA支架的多核苷酸，并且多核苷酸进一步包含与支架3'连接的靶向序列多核苷酸，其具有SEQ ID NO:309-343、363-2100或2295-21835的序列，或与其具有至少约65％、至少约75％、至少约85％或至少约95％同一性的序列，所述序列与C9orf72基因互补并且因此与其杂交。在其它实施例中，本公开提供具有15个核苷酸、16个核苷酸、17个核苷酸、18个核苷酸、19个核苷酸、20个核苷酸或21个核苷酸的靶向多核苷酸。在一些情况下，多核苷酸序列编码包含与C9orf72外显子杂交的靶向序列的gNA。在其它情况下，多核苷酸序列编码包含与C9orf72内含子杂交的靶向序列的gNA。在其它情况下，多核苷酸序列编码包含与C9orf72内含子-外显子接界杂交的靶向序列的gNA。在其它情况下，多核苷酸序列编码包含与C9orf72基因的基因间区杂交的靶向序列的gNA。在其它情况下，多核苷酸序列编码包含与位于HRS的5'处的序列杂交的靶向序列的gNA。在其它情况下，多核苷酸序列编码包含与位于HRS的3'处的序列杂交的靶向序列的gNA。在其它实施例中，本公开提供编码两种或更多种gNA的多核苷酸序列，所述gNA各自具有支架和靶向序列，所述靶向序列共同地与位于HRS的5'处的序列和位于HRS的3'处的序列杂交。在其它实施例中，多核苷酸序列编码包含与C9orf72调节元件杂交的靶向序列的gNA。在一些情况下，C9orf72调节元件为C9orf72启动子或增强子。在一些情况下，C9orf72调节元件位于

C9orf72转录起始位点的5'处、C9orf72转录起点的3'处或C9orf72内含子中。在一些情况下,C9orf72调节元件处于C9orf72基因的内含子中。在其它情况下,C9orf72调节元件包含C9orf72基因的5'UTR。在又其它情况下,C9orf72调节元件包含C9orf72基因的3'UTR。

[0410] 在其它实施例中,本公开提供供体模板核酸,其中供体模板包含与C9orf72靶核酸序列具有同源性但不与打算进行基因编辑的靶核酸的靶序列具有完全同一性的核苷酸序列。在一些实施例中,C9orf72供体模板意图用于基因编辑并且包含C9orf72基因的全部或至少一部分。在一些实施例中,C9orf72供体模板包含与C9orf72基因杂交的序列。在其它实施例中,C9orf72供体序列包含编码C9orf72外显子的至少一部分的序列。在其它实施例中,C9orf72供体模板具有编码C9orf72内含子的至少一部分的序列。在其它实施例中,C9orf72供体模板具有编码C9orf72内含子-外显子接界的至少一部分的序列。在其它实施例中,C9orf72供体模板具有编码C9orf72基因的基因间区的至少一部分的序列。在其它实施例中,C9orf72供体模板具有编码C9orf72调节元件的至少一部分的序列。在一些情况下,C9orf72供体模板是编码SEQ ID NO:227或228的全部或一部分的野生型序列。在其它情况下,C9orf72供体模板序列包含相对于野生型C9orf72基因的一个或多个突变,且可含有相对于基因组序列的一个或多个单碱基变化、插入、缺失、倒位或重排,其条件是与靶序列存在足够的同源性以支持同源定向修复,或供体模板具有同源臂,因此插入可使得剪接出包含例如六核苷酸重复序列的区域,以使得可表达功能C9orf72蛋白质。在特定实施例中,C9orf72供体模板序列包含六核苷酸重复序列GGGGCC的10到约30个副本。在前述实施例中,供体模板的大小可以在10到10,000个核苷酸的范围内。在一些实施例中,供体模板是单链DNA模板。在其它实施例中,供体模板是单链RNA模板。在其它实施例中,供体模板是双链DNA模板。

[0411] 在一些实施例中,本公开涉及产生编码本文描述的任一实施例的参考CasX、CasX变异体或gNA的多核苷酸序列(包括其变异体)的方法,以及表达由多核苷酸序列表达的蛋白质或转录的RNA的方法。通常,所述方法包括产生编码本文描述的任一实施例的参考CasX、CasX变异体或gNA的多核苷酸序列,并将编码基因掺入适合宿主细胞的表达载体中。为了产生本文描述的任一实施例的编码的参考CasX、CasX变异体或gNA,所述方法包括用包含编码多核苷酸的表达载体转化适当的宿主细胞,并在引起或允许本文描述的任何实施例的所得参考CasX、CasX变异体或gNA在转化的宿主细胞中表达或转录的条件下培养宿主细胞,从而产生参考CasX、CasX变异体或gNA,其通过本文描述的方法或通过所属领域中已知的标准纯化方法,包括实例的方法回收。分子生物学中的标准重组技术用于制备本公开的多核苷酸和表达载体。

[0412] 根据本公开,本文所述的任何实施例的编码参考CasX、CasX变异体或gNA的多核苷酸序列用于产生引导适当宿主细胞中的表达的重组DNA分子。几种克隆策略适用于实施本公开,其中许多用于生成包含编码本公开组合物或其互补序列的基因的构建体。在一些实施例中,克隆策略用于创建编码构建体的基因,所述构建体包含编码参考CasX、CasX变异体或gNA的核苷酸并用于转化宿主细胞以表达组合物。

[0413] 在一种方法中,首先制备含有编码参考CasX、CasX变异体或gNA的DNA序列的构建体。用于制备此类构建体的示范性方法在实例中进行了描述。然后将所述构建体用于创建适合转化宿主细胞(如原核或真核宿主细胞)的表达载体,以表达和恢复多肽构建体。必要

时，宿主细胞为大肠杆菌细胞。在其它实施例中，宿主细胞选自BHK细胞、HEK293细胞、HEK293T细胞、Lenti-X HEK293细胞、NS0细胞、SP2/0细胞、YO骨髓瘤细胞、P3X63小鼠骨髓瘤细胞、PER细胞、PER.C6细胞、杂交瘤细胞、NIH3T3细胞、COS、HeLa、CHO或酵母细胞。实例中描述了用于创建表达载体、宿主细胞转化以及参考CasX、CasX变异体或gNA的表达和回收的示范性方法。

[0414] 编码参考CasX、CasX变异体或gNA构建体的一种或多种基因可以在一个或多个步骤中制造，其采用完全合成的方式或通过合成以及酶工艺，如限制酶介导的克隆、PCR和重叠延伸，包括实例中更完整地描述的方法。例如，本文公开的方法可用于连接编码所需序列的各种组分(例如，CasX和gNA)基因的多核苷酸序列。使用基因合成的标准技术从寡核苷酸组装编码多肽组合物的基因。

[0415] 在一些实施例中，编码CasX蛋白的核苷酸序列经密码子优化。这种类型的优化可能需要对编码核苷酸序列进行突变，以模拟预期宿主生物体或细胞在编码相同CasX蛋白时的密码子偏好。因此，密码子可改变，但经编码蛋白质保持不变。例如，如果CasX蛋白的预期靶细胞是人类细胞，则可以使用人类密码子优化的CasX编码核苷酸序列。作为另一非限制性实例，如果预期宿主细胞是小鼠细胞，则可以生成小鼠密码子优化的CasX编码核苷酸序列。作为另一非限制性实例，如果预期宿主细胞是植物细胞，则可以生成植物密码子优化的编码CasX蛋白变异体的核苷酸序列。作为另一非限制性实例，如果预期宿主细胞是昆虫细胞，则可以生成昆虫密码子优化的CasX蛋白编码核苷酸序列。可以使用优化密码子使用和氨基酸组成的算法进行基因设计，所述算法适用于生产参考CasX、CasX变异体或gNA时使用的宿主细胞。在本公开的一种方法中，创建编码构建体的组分的多核苷酸文库，然后进行组装，如上所述。然后组装所得基因，并且所得基因用于转化宿主细胞并产生和恢复参考CasX、CasX变异体或gNA组合物以评估其性质，如本文所述。

[0416] 在一些实施例中，编码gNA的核苷酸序列可操作地连接到控制元件，例如转录控制元件，如启动子。在一些实施例中，编码CasX蛋白的核苷酸序列可操作地连接到控制元件，例如转录控制元件，如启动子。

[0417] 转录控制元件可以是启动子。在一些情况下，启动子为组成性活化启动子。在一些情况下，启动子为可调节启动子。在一些情况下，启动子为诱导型启动子。在一些情况下，启动子为组织特异性启动子。在一些情况下，启动子为细胞类型特异性启动子。在一些情况下，转录控制元件(例如启动子)在靶细胞类型或靶细胞群体中起作用。举例来说，在一些情况下，转录控制元件可在真核细胞中具有功能性，例如细胞选自由以下组成的组：普尔基涅细胞、额皮质神经元、运动皮质神经元、海马神经元、小脑神经元、上运动神经元、脊髓神经元、脊髓运动神经元、神经胶质细胞和星形胶质细胞。

[0418] 真核启动子(在真核细胞中起作用的启动子)的非限制性实例包括EF1α、EF1α核启动子、来自巨细胞病毒(CMV)即刻早期的那些、单纯疱疹病毒(HSV)胸苷激酶、早期及晚期SV40、来自逆转录病毒的长末端重复序列(LTR)及小鼠金属硫蛋白-I。真核启动子的其它非限制性实例包括CMV启动子全长启动子、最小CMV启动子、鸡β-肌动蛋白启动子、hPGK启动子、HSV TK启动子、Mini-TK启动子、赋予神经元特异性表达的人类突触蛋白I启动子、选择性表达于神经元中的Mecp2启动子、最小IL-2启动子、劳氏肉瘤病毒增强子/启动子(单一)、形成脾脏病灶的病毒长末端重复序列(LTR)启动子、SV40启动子、SV40增强子及早期启动

子、TBG启动子:来自人类甲状腺素结合球蛋白基因的启动子(肝脏特异性)、PGK启动子、人类泛素C启动子、UCOE启动子(HNRPA2B1-CBX3的启动子)、组蛋白H2启动子、组蛋白H3启动子、U1a1小核RNA启动子(226nt)、U1b2小核RNA启动子(246nt)26、TTR最小增强子/启动子、b-驱动蛋白启动子、人类eIF4A1启动子、ROSA26启动子及3-磷酸甘油醛脱氢酶(GAPDH)启动子。

[0419] 合适的载体和启动子的选择完全在所属领域普通技术的水平内,因为它涉及控制表达,例如,用于修饰参与抗原加工、抗原呈递、抗原识别和/或抗原反应的蛋白质和/或其调节元件。表达载体亦可含有用于翻译起始及转录终止的核糖体结合位点。表达载体亦可包括用于扩增表达的适合的序列。表达载体也可包括编码蛋白质标签(例如,6xHis标签、血凝素标签、FLAG标签、荧光蛋白等)的核苷酸序列,这些蛋白质标签可与CasX蛋白融合,从而产生用于纯化或检测的嵌合CasX蛋白。

[0420] 在一些实施例中,编码每种gNA变异体或CasX蛋白的核苷酸序列与诱导型启动子、组成型活性启动子、空间受限启动子(即,转录控制元件、增强子、组织特异性启动子、细胞类型特异性启动子等)或时间受限启动子可操作地连接。在其它实施例中,将编码gNA或CasX的单个核苷酸序列与前述启动子类别之一连接,然后通过下述常规方法将其引入待修饰的细胞中。

[0421] 在某些实施例中,适合的启动子可衍生自病毒且可因此称为病毒启动子,或其可衍生自任何生物体,包括原核或真核生物体。适合的启动子可用于通过任何RNA聚合酶(例如pol I、pol II、pol III)驱动表达。示范性启动子包括但不限于SV40早期启动子、小鼠乳房肿瘤病毒长末端重复序列(LTR)启动子;腺病毒主要晚期启动子(Ad MLP);单纯疱疹病毒(HSV)启动子;巨细胞病毒(CMV)启动子,例如CMV即刻早期启动子区域(CMVIE)、劳氏肉瘤病毒(RSV)启动子、人类U6小核启动子(U6)、增强型U6启动子、人类H1启动子(H1)、POL1启动子、7SK启动子、tRNA启动子等。

[0422] 在一些实施例中,一种或多种编码CasX和gNA并且任选地包含供体模板的核苷酸序列各自可操作地连接到可在真核细胞中操作的启动子(在其控制下)。诱导型启动子的实例可以包括但不限于T7 RNA聚合酶启动子、T3 RNA聚合酶启动子、异丙基-β-D-硫代吡喃半乳糖苷(IPTG)-调节的启动子、乳糖诱导的启动子、热休克启动子、四环素调节的启动子、类固醇调节的启动子、金属调节的启动子、雌激素受体调节的启动子等。因此,在一些实施例中,诱导型启动子可以由包括但不限于强力霉素;雌激素和/或雌激素类似物;IPTG;等的分子调节。

[0423] 在某些实施例中,适合使用的诱导型启动子可包括本文所述或所属领域所属领域的一般技术人员已知的任何诱导型启动子。诱导型启动子的实例包括但不限于化学/生物化学调节和物理调节的启动子,如醇调节启动子、四环素调节启动子(例如,无水四环素(aTc)反应性启动子及其它四环素反应性启动子系统,其包括四环素抑制蛋白(tetR)、四环素操纵序列(tetO)和四环素反式激活融合蛋白(tTA)、类固醇调节启动子(例如,基于大鼠糖皮质激素受体、人类雌激素受体、蛾蜕皮激素受体的启动子,以及来自类固醇/类视黄素/甲状腺受体超家族的启动子)、金属调节启动子(例如,衍生自来自酵母、小鼠及人类的金属硫蛋白(结合和螯合金属离子的蛋白)基因的启动子)、发病机制调节启动子(例如,由水杨酸、乙烯或苯并噻二唑(BTH)诱导)、温度/热诱导性启动子(例如,热休克启动子)以及光调

节启动子(例如,来自植物细胞的光反应性启动子)。

[0424] 在一些情况下,启动子是空间受限启动子(即,细胞类型特异性启动子、组织特异性启动子等),使得在多细胞生物体中,启动子在特定细胞的子集中是活跃的(即,“开启”)。空间受限启动子亦可称为增强子、转录控制元件、控制序列等。可使用任何便利的空间受限启动子,只要启动子在靶向宿主细胞(例如真核细胞;原核细胞)中起作用。

[0425] 在一些情况下,启动子为可逆启动子。适合的可逆启动子,包括可逆诱导型启动子为所属领域中已知的。此类可逆启动子可分离及衍生自多种生物体,例如真核生物及原核生物。衍生自用于第二生物体的第一生物体(例如第一原核生物及第二真核生物、第一真核生物及第二原核生物等)的可逆启动子的修饰在所属领域中为熟知的。此类可逆启动子及基于此类可逆启动子的系统但也包含额外对照蛋白,包括但不限于醇调节启动子(例如,醇脱氢酶I(alcA)基因启动子、对醇反式激活蛋白(AlcR)有反应的启动子等)、四环素调节启动子(例如,包括Tet活化子、TetON、TetOFF等的启动子系统)、类固醇调节启动子(例如,大鼠糖皮质激素受体启动子系统、人类雌激素受体启动子系统、类视黄素启动子系统、甲状腺启动子系统、蜕皮激素启动子系统、米非司酮启动子系统等)、金属调节启动子(例如,金属硫蛋白启动子系统等)、发病机制相关调节启动子(例如,水杨酸调节启动子、乙烯调节启动子、苯并噻二唑调节启动子等)、温度调节启动子(例如,热休克诱导性启动子(例如,HSP-70、HSP-90、大豆热休克启动子等)、光调节启动子、合成诱导型启动子及其类似物。

[0426] 本公开的重组表达载体还可包含促进本公开的CasX蛋白和gNA的稳健表达的元件。举例来说,重组表达载体可包括以下中的一或多个:聚腺苷酸化信号(PolyA)、内含子序列或转录后调节元件,例如土拔鼠肝炎转录后调节元件(WPRE)。示范性polyA序列包括hGHpoly(A)信号(短)、HSVTKpoly(A)信号、合成聚腺苷酸化信号、SV40 poly(A)信号、β-血球蛋白poly(A)信号及其类似物。所属领域普通技术人员将能够选择要包括于本文所述的重组表达载体中的适合元件。

[0427] 可接着将编码参考CasX、CasX变异体和gNA序列的多核苷酸个别地克隆到一个或多个表达载体中。在一些实施例中,本公开提供了包含多核苷酸的载体,其选自由以下组成的组:逆转录病毒载体、慢病毒载体、腺病毒载体、腺相关病毒(AAV)载体、病毒样粒子(VLP)、单纯疱疹病毒(HSV)载体、质粒、小环、纳米质粒、DNA载体和RNA载体。在一些实施例中,载体是重组表达载体,其包含编码CasX蛋白的核苷酸序列。在其它实施例中,本公开提供了包含编码CasX蛋白的核苷酸序列和编码gNA的核苷酸序列的重组表达载体。在一些情况下,编码CasX蛋白变异体的核苷酸序列和/或编码gNA的核苷酸序列可操作地连接到在所选细胞类型中可操作的启动子。在其它实施例中,编码CasX蛋白变异体的核苷酸序列和编码gNA的核苷酸序列在可操作地连接到启动子的单独载体中提供。

[0428] 在一些实施例中,本文提供了一种或多种重组表达载体,其包含以下中的一个或多个:(i)供体模板核酸的核苷酸序列,其中供体模板包含与靶核酸(例如,靶基因组)的靶序列具有同源性的核苷酸序列;(ii)编码gNA的核苷酸序列,所述gNA与靶向的基因组的基因座的靶序列杂交(例如,被配置为单或双引导RNA)并与在如真核细胞的靶细胞中可操作的启动子可操作地连接;和(iii)编码CasX蛋白的核苷酸序列,其与在如真核细胞的靶细胞中可操作的启动子可操作地连接。在一些实施例中,编码供体模板、gNA和CasX蛋白的序列在不同重组表达载体中,并且在其它实施例中,一种或多种多核苷酸序列(针对供体模板、

CasX和gNA)在相同重组表达载体中。在其它情况下,CasX和gNA作为RNP(例如,通过电穿孔或化学方式)递送到靶细胞,并且供体模板通过载体递送。

[0429] 通过多种程序将多核苷酸序列插入载体中。通常,使用所属领域已知的技术将DNA插入适当的限制性核酸内切酶位点。载体组分通常包括但不限于信号序列、复制起点、一种或多种标记基因、增强子元件、启动子和转录终止序列中的一种或多种。含有一种或多种这些组分的合适载体的构建采用所属领域技术人员已知的标准连接技术。此类技术在所属领域中是众所周知的并且在科学和专利文献中得到充分描述。各种载体是公开的。例如,载体可以是质粒、粘粒、病毒粒子或噬菌体的形式,它们可以方便地进行重组DNA程序,并且载体的选择通常取决于将其引入的宿主细胞。因此,载体可以是自主复制载体,即作为染色体外实体存在的载体,其复制独立于染色体复制,例如质粒。替代地,所述载体可以是当被引入宿主细胞时,被整合到宿主细胞基因组中并与已整合入的染色体一起复制的载体。一旦引入合适的宿主细胞,参与抗原加工、抗原呈递、抗原识别和/或抗原反应的蛋白质的表达可以使用所属领域已知的任何核酸或蛋白质测定来确定。例如,参考CasX的转录mRNA或CasX变异体的存在可以通过常规杂交测定(例如,Northern印迹分析)、扩增程序(例如RT-PCR)、SAGE(美国专利号5,695,937)和基于阵列的技术(参见例如,美国专利号5,405,783、5,412,087和5,445,934),使用与多核苷酸的任何区域互补的探针进行检测和/或量化。

[0430] 本公开提供了含有复制和控制序列的质粒表达载体的用途,所述复制和控制序列与宿主细胞相容并被宿主细胞识别并且可操作地连接到编码多肽的基因以用于多肽的受控表达或RNA的转录。此类载体序列对于多种细菌、酵母和病毒是众所周知的。可以使用的有用的表达载体包括例如染色体、非染色体和合成DNA序列的段。“表达载体”是指含有DNA序列的DNA构建体,所述DNA序列可操作地连接到合适的控制序列,所述控制序列能够影响编码多肽的DNA在合适的宿主中的表达。要求是载体在所选宿主细胞中是可复制的和可行的。可以根据需要使用低拷贝数或高拷贝数载体。载体的控制序列包括影响转录的启动子、控制这种转录的任选操纵子序列、编码合适的mRNA核糖体结合位点的序列以及控制转录和翻译终止的序列。启动子可以是在所选宿主细胞中显示转录活性的任何DNA序列,并且可以衍生自编码与宿主细胞同源或异源的蛋白质的基因。

[0431] 可以通过多种方法将多核苷酸和重组表达载体递送至靶宿主细胞。此类方法包括但不限于病毒感染、转染、脂转染、电穿孔、磷酸钙沉淀、聚乙烯亚胺(PEI)介导的转染、DEAE-葡聚糖介导的转染、显微注射、脂质体介导的转染、粒子枪技术、核转染、通过与供体DNA融合或募集的细胞穿透CasX蛋白直接添加、细胞挤压、磷酸钙沉淀、直接显微注射、纳米粒子介导的核酸递送,以及使用Qiagen的市售TransMessenger®试剂、Stemgent的StemfectTM RNA转染试剂盒和来自Mirus Bio LLC的TransIT®-mRNA转染试剂盒、Lonza核转染、Maxagen电穿孔等。

[0432] 根据本公开,对本文中所描述的任何实施例的参考CasX、CasX变异体或gNA进行编码的核酸序列(或其互补序列)用于产生引导适当宿主细胞中的表达的重组DNA分子。几种克隆策略适用于实施本公开,其中许多用于生成包含编码本公开组合物或其互补序列的基因的构建体。在一些实施例中,克隆策略用于创建编码构建体的基因,所述构建体包含编码参考CasX、CasX变异体或gNA的核苷酸并用于转化宿主细胞以表达组合物。

[0433] 在一种方法中,首先制备含有编码参考CasX、CasX变异体或gNA的DNA序列的构建

体。用于制备此类构建体的示范性方法在实例中进行了描述。然后将所述构建体用于创建适合转化宿主细胞(如原核或真核宿主细胞)的表达载体,以表达和恢复多肽构建体。必要时,宿主细胞为大肠杆菌细胞。在其它实施例中,宿主细胞选自BHK细胞、HEK293细胞、HEK293T细胞、Lenti-X HEK293细胞、NSO细胞、SP2/0细胞、YO骨髓瘤细胞、P3X63小鼠骨髓瘤细胞、PER细胞、PER.C6细胞、杂交瘤细胞、NIH3T3细胞、COS、HeLa、CHO或酵母细胞。实例中描述了用于创建表达载体、宿主细胞转化以及参考CasX、CasX变异体或gNA的表达和回收的示范性方法。

[0434] 编码参考CasX、CasX变异体或gNA构建体的一种或多种基因可以在一个或多个步骤中制造,其采用完全合成的方式或通过合成以及酶工艺,如限制酶介导的克隆、PCR和重叠延伸,包括实例中更完整地描述的方法。例如,本文公开的方法可用于连接编码所需序列的各种组分(例如,CasX和gNA)基因的多核苷酸序列。使用基因合成的标准技术从寡核苷酸组装编码多肽组合物的基因。

[0435] 在一些实施例中,编码CasX蛋白的核苷酸序列经密码子优化。这种类型的优化可能需要对编码核苷酸序列进行突变,以模拟预期宿主生物体或细胞在编码相同CasX蛋白时的密码子偏好。因此,密码子可改变,但经编码蛋白质保持不变。例如,如果CasX蛋白的预期靶细胞是人类细胞,则可以使用人类密码子优化的CasX编码核苷酸序列。作为另一非限制性实例,如果预期宿主细胞是小鼠细胞,则可以生成小鼠密码子优化的CasX编码核苷酸序列。作为另一非限制性实例,如果预期宿主细胞是植物细胞,则可以生成植物密码子优化的编码CasX蛋白变异体的核苷酸序列。作为另一非限制性实例,如果预期宿主细胞是昆虫细胞,则可以生成昆虫密码子优化的CasX蛋白编码核苷酸序列。可以使用优化密码子使用和氨基酸组成的算法进行基因设计,所述算法适用于生产参考CasX、CasX变异体或gNA时使用的宿主细胞。在本公开的一种方法中,创建编码构建体的组分的多核苷酸文库,然后进行组装,如上所述。然后组装所得基因,并且所得基因用于转化宿主细胞并产生和恢复参考CasX、CasX变异体或gNA组合物以评估其性质,如本文所述。

[0436] 在一些实施例中,编码gNA的核苷酸序列可操作地连接到控制元件,例如转录控制元件,如启动子。在一些实施例中,编码CasX蛋白的核苷酸序列可操作地连接到控制元件,例如转录控制元件,如启动子。

[0437] 转录控制元件可以是启动子。在一些情况下,启动子为组成性活化启动子。在一些情况下,启动子为可调节启动子。在一些情况下,启动子为诱导型启动子。在一些情况下,启动子为组织特异性启动子。在一些情况下,启动子为细胞类型特异性启动子。在一些情况下,转录控制元件(例如启动子)在靶细胞类型或靶细胞群体中起作用。举例来说,在一些情况下,转录控制元件可在真核细胞中具有功能性,例如细胞选自由以下组成的组:普尔基涅细胞、额皮质神经元、运动皮质神经元、海马神经元、小脑神经元、上运动神经元、脊髓神经元、脊髓运动神经元、神经胶质细胞和星形胶质细胞。

[0438] 真核启动子(在真核细胞中起作用的启动子)的非限制性实例包括EF1α、EF1α核启动子、来自巨细胞病毒(CMV)即刻早期的那些、单纯疱疹病毒(HSV)胸苷激酶、早期及晚期SV40、来自逆转录病毒的长末端重复序列(LTR)及小鼠金属硫蛋白-I。真核启动子的其它非限制性实例包括CMV启动子全长启动子、最小CMV启动子、鸡β-肌动蛋白启动子、RSV启动子、HIV-Ltr启动子、hPGK启动子、HSV TK启动子、微型-TK启动子、赋予神经元特异性表达的人

类突触蛋白I启动子、选择性表达于神经元中的Mecp2启动子、最小IL-2启动子、劳氏肉瘤病毒增强子/启动子(单一)、形成脾脏病灶的病毒长末端重复序列(LTR)启动子、SV40启动子、SV40增强子及早期启动子、TBG启动子:来自人类甲状腺素结合球蛋白基因的启动子(肝脏特异性)、PGK启动子、人类泛素C启动子、UCOE启动子(HNRPA2B1-CBX3的启动子)、组蛋白H2启动子、组蛋白H3启动子、U1a1小核RNA启动子(226nt)、U1b2小核RNA启动子(246nt)26、TTR最小增强子/启动子、b-驱动蛋白启动子、人类eIF4A1启动子、ROSA26启动子及甘油醛3-磷酸脱氢酶(GAPDH)启动子。在一些实施例中,在gNA构建体中使用的启动子为U6(Kunkel,GR等人《U6小核RNA通过RNA聚合酶III转录(U6 small nuclear RNA is transcribed by RNA polymerase III)》《美国国家科学院院刊》83(22):8575(1986))。

[0439] 适当载体和启动子的选择完全在所属领域的一般技术人员的能力范围之内,因为其涉及控制表达,例如,用于修饰C9orf72基因。表达载体亦可含有用于翻译起始及转录终止的核糖体结合位点。表达载体亦可包括用于扩增表达的适合的序列。表达载体也可包括编码蛋白质标签(例如,6xHis标签、血凝素标签、荧光蛋白等)的核苷酸序列,其可与CasX蛋白融合,从而产生用于纯化或检测的嵌合CasX蛋白。

[0440] 在一些实施例中,编码每种gNA变异体或CasX蛋白的核苷酸序列与诱导型启动子、组成型活性启动子、空间受限启动子(即,转录控制元件、增强子、组织特异性启动子、细胞类型特异性启动子等)或时间受限启动子可操作地连接。在其它实施例中,将编码gNA或CasX的单个核苷酸序列与前述启动子类别之一连接,然后通过下述常规方法将其引入待修饰的细胞中。

[0441] 在某些实施例中,适合的启动子可衍生自病毒且可因此称为病毒启动子,或其可衍生自任何生物体,包括原核或真核生物体。适合的启动子可用于通过任何RNA聚合酶(例如pol I、pol II、pol III)驱动表达。示范性启动子包括但不限于SV40早期启动子、小鼠乳房肿瘤病毒长末端重复序列(LTR)启动子;腺病毒主要晚期启动子(Ad MLP);单纯疱疹病毒(HSV)启动子;巨细胞病毒(CMV)启动子,例如CMV即刻早期启动子区域(CMVIE)、劳氏肉瘤病毒(RSV)启动子、人类U6小核启动子(U6)、增强型U6启动子、人类H1启动子(H1)、POL1启动子、7SK启动子、tRNA启动子等。

[0442] 在一些实施例中,一种或多种编码CasX和gNA并且任选地包含供体模板的核苷酸序列各自可操作地连接到可在真核细胞中操作的启动子(在其控制下)。诱导型启动子的实例可以包括但不限于T7 RNA聚合酶启动子、T3 RNA聚合酶启动子、异丙基-β-D-硫代吡喃半乳糖苷(IPTG)-调节的启动子、乳糖诱导的启动子、热休克启动子、四环素调节的启动子、类固醇调节的启动子、金属调节的启动子、雌激素受体调节的启动子等。因此,在一些实施例中,诱导型启动子可以由包括但不限于强力霉素;雌激素和/或雌激素类似物;IPTG;等的分子调节。

[0443] 在某些实施例中,适合使用的诱导型启动子可包括本文所述或所属领域所属领域的一般技术人员已知的任何诱导型启动子。诱导型启动子的实例包括但不限于化学/生物化学调节和物理调节的启动子,如醇调节启动子、四环素调节启动子(例如,无水四环素(aTc)反应性启动子及其它四环素反应性启动子系统,其包括四环素抑制蛋白(tetR)、四环素操纵序列(tetO)和四环素反式激活融合蛋白(tTA)、类固醇调节启动子(例如,基于大鼠糖皮质激素受体、人类雌激素受体、蛾蜕皮激素受体的启动子,以及来自类固醇/类视黄素/

甲状腺受体超家族的启动子)、金属调节启动子(例如,衍生自来自酵母、小鼠及人类的金属硫蛋白(结合和螯合金属离子的蛋白)基因的启动子)、发病机制调节启动子(例如,由水杨酸、乙烯或苯并噻二唑(BTH)诱导)、温度/热诱导性启动子(例如,热休克启动子)以及光调节启动子(例如,来自植物细胞的光反应性启动子)。

[0444] 在一些情况下,启动子是空间受限启动子(即,细胞类型特异性启动子、组织特异性启动子等),使得在多细胞生物体中,启动子在特定细胞的子集中是活跃的(即,“开启”)。空间受限启动子亦可称为增强子、转录控制元件、控制序列等。可使用任何便利的空间受限启动子,只要启动子在靶向宿主细胞(例如真核细胞;原核细胞)中起作用。

[0445] 在一些情况下,启动子为可逆启动子。适合的可逆启动子,包括可逆诱导型启动子为所属领域中已知的。此类可逆启动子可分离及衍生自多种生物体,例如真核生物及原核生物。衍生自用于第二生物体的第一生物体(例如第一原核生物及第二真核生物、第一真核生物及第二原核生物等)的可逆启动子的修饰在所属领域中为熟知的。此类可逆启动子及基于此类可逆启动子的系统但也包含额外对照蛋白,包括但不限于醇调节启动子(例如,醇脱氢酶I(alcA)基因启动子、对醇反式激活蛋白(AlcR)有反应的启动子等)、四环素调节启动子(例如,包括Tet活化子、TetON、TetOFF等的启动子系统)、类固醇调节启动子(例如,大鼠糖皮质激素受体启动子系统、人类雌激素受体启动子系统、类视黄素启动子系统、甲状腺启动子系统、蜕皮激素启动子系统、米非司酮启动子系统等)、金属调节启动子(例如,金属硫蛋白启动子系统等)、发病机制相关调节启动子(例如,水杨酸调节启动子、乙烯调节启动子、苯并噻二唑调节启动子等)、温度调节启动子(例如,热休克诱导性启动子(例如,HSP-70、HSP-90、大豆热休克启动子等)、光调节启动子、合成诱导型启动子及其类似物。

[0446] 本公开的重组表达载体还可包含促进本公开的CasX蛋白和gNA的稳健表达的元件。举例来说,重组表达载体可包括以下中的一或多个:聚腺苷酸化信号(PolyA)、内含子序列或转录后调节元件,例如土拨鼠肝炎转录后调节元件(WPRE)。示范性polyA序列包括hGHpoly(A)信号(短)、HSVTKpoly(A)信号、合成聚腺苷酸化信号、SV40 poly(A)信号、β-血球蛋白poly(A)信号及其类似物。所属领域普通技术人员将能够选择要包括于本文所述的重组表达载体中的适合元件。

[0447] 可接着将编码参考CasX、CasX变异体和gNA序列的多核苷酸个别地克隆到一个或多个表达载体中。在一些实施例中,本公开提供了包含多核苷酸的载体,其选自由以下组成的组:逆转录病毒载体、慢病毒载体、腺病毒载体、腺相关病毒(AAV)载体、病毒样粒子(VLP)、单纯疱疹病毒(HSV)载体、质粒、小环、纳米质粒、DNA载体和RNA载体。在一些实施例中,载体是重组表达载体,其包含编码CasX蛋白的核苷酸序列。在其它实施例中,本公开提供了包含编码CasX蛋白的核苷酸序列和编码gNA的核苷酸序列的重组表达载体。在一些情况下,编码CasX蛋白变异体的核苷酸序列和/或编码gNA的核苷酸序列可操作地连接到在所选细胞类型中可操作的启动子。在其它实施例中,编码CasX蛋白变异体的核苷酸序列和编码gNA的核苷酸序列在可操作地连接到启动子的单独载体中提供。

[0448] 在一些实施例中,本文提供了一种或多种重组表达载体,其包含以下中的一个或多个:(i)供体模板核酸的核苷酸序列,其中供体模板包含与靶核酸(例如,靶基因组)的靶序列具有同源性的核苷酸序列;(ii)编码gNA的核苷酸序列,所述gNA与靶向的基因组的基因座的靶序列杂交(例如,被配置为单或双引导RNA)并与在如真核细胞的靶细胞中可操作

的启动子可操作地连接;和(iii)编码CasX蛋白的核苷酸序列,其与在如真核细胞的靶细胞中可操作的启动子可操作地连接。在一些实施例中,编码供体模板、gNA和CasX蛋白的序列在不同重组表达载体中,并且在其它实施例中,一种或多种多核苷酸序列(针对供体模板、CasX和gNA)在相同重组表达载体中。在其它情况下,CasX和gNA作为RNP(例如,通过电穿孔或化学方式)递送到靶细胞,并且供体模板通过载体递送。

[0449] 通过多种程序将多核苷酸序列插入载体中。通常,使用所属领域已知的技术将DNA插入适当的限制性核酸内切酶位点。载体组分通常包括但不限于信号序列、复制起点、一种或多种标记基因、增强子元件、启动子和转录终止序列中的一种或多种。含有一种或多种这些组分的合适载体的构建采用所属领域技术人员已知的标准连接技术。此类技术在所属领域中是众所周知的并且在科学和专利文献中得到充分描述。各种载体是公开的。例如,载体可以是质粒、粘粒、病毒粒子或噬菌体的形式,它们可以方便地进行重组DNA程序,并且载体的选择通常取决于将其引入的宿主细胞。因此,载体可以是自主复制载体,即作为染色体外实体存在的载体,其复制独立于染色体复制,例如质粒。替代地,所述载体可以是当被引入宿主细胞时,被整合到宿主细胞基因组中并与已整合入的染色体一起复制的载体。一旦引入合适的宿主细胞,参与抗原加工、抗原呈递、抗原识别和/或抗原反应的蛋白质的表达可以使用所属领域已知的任何核酸或蛋白质测定来确定。例如,参考CasX的转录mRNA或CasX变异体的存在可以通过常规杂交测定(例如,Northern印迹分析)、扩增程序(例如RT-PCR)、SAGE(美国专利号5,695,937)和基于阵列的技术(参见例如,美国专利号5,405,783、5,412,087和5,445,934),使用与多核苷酸的任何区域互补的探针进行检测和/或量化。

[0450] 本公开提供了含有复制和控制序列的质粒表达载体的用途,所述复制和控制序列与宿主细胞相容并被宿主细胞识别并且可操作地连接到编码多肽的基因以用于多肽的受控表达或RNA的转录。此类载体序列对于多种细菌、酵母和病毒是众所周知的。可以使用的有用的表达载体包括例如染色体、非染色体和合成DNA序列的段。“表达载体”是指含有DNA序列的DNA构建体,所述DNA序列可操作地连接到合适的控制序列,所述控制序列能够影响编码多肽的DNA在合适的宿主中的表达。要求是载体在所选宿主细胞中是可复制的和可行的。可以根据需要使用低拷贝数或高拷贝数载体。载体的控制序列包括影响转录的启动子、控制这种转录的任选操纵子序列、编码合适的mRNA核糖体结合位点的序列以及控制转录和翻译终止的序列。启动子可以是在所选宿主细胞中显示转录活性的任何DNA序列,并且可以衍生自编码与宿主细胞同源或异源的蛋白质的基因。

[0451] 可以通过多种方法将多核苷酸和重组表达载体递送至靶宿主细胞。此类方法包括但不限于病毒感染、转染、脂转染、电穿孔、磷酸钙沉淀、聚乙烯亚胺(PEI)介导的转染、DEAE-葡聚糖介导的转染、显微注射、脂质体介导的转染、粒子枪技术、核转染、通过与供体DNA融合或募集的细胞穿透CasX蛋白直接添加、细胞挤压、磷酸钙沉淀、直接显微注射、纳米粒子介导的核酸递送,以及使用Qiagen的市售TransMessenger®试剂、Stemgent的StemfectTM RNA转染试剂盒和来自Mirus Bio LLC的TransIT®-mRNA转染试剂盒、Lonza核转染、Maxagen电穿孔等。

[0452] 重组表达载体序列可封装至病毒或病毒样粒子(在本文中亦称为“VLP”或“病毒粒子”)中,用于随后离体、体外或体内的细胞感染和转化。此类VLP或病毒粒子将典型地包括包裹或包装载体基因组的蛋白质。合适的表达载体可以包括基于牛痘病毒的病毒表达载

体;脊髓灰质炎病毒;腺病毒;逆转录病毒载体(例如,鼠白血病病毒)、脾坏死病毒和衍生自逆转录病毒的载体,如劳斯肉瘤病毒、哈维肉瘤病毒、禽白血病病毒、逆转录病毒、慢病毒、人类免疫缺陷病毒、骨髓增殖性肉瘤病毒和乳腺肿瘤病毒;等等。

[0453] 在一些实施例中,本公开的重组表达载体为重组腺相关病毒(AAV)载体。在一个具体实施例中,本公开的重组表达载体是重组逆转录病毒载体。在另一具体实施例中,本公开的重组表达载体是重组慢病毒载体。

[0454] AAV是一种小型(20nm)非致病性病毒,在使用病毒载体输送到细胞(如真核细胞)的情况下,可用于治疗人类疾病,无论是在体内还是体外,用于制备向受试者施用的细胞。产生构建体,其例如编码如本文所述的CasX蛋白和gNA实施例中的任一种,和任选的供体模板,并且可以与AAV反向末端重复序列(ITR)侧接,由此使AAV载体能够封装到AAV病毒粒子中。

[0455] "AAV"载体可指天然存在的野生型病毒自身或其衍生物。所述术语涵盖所有亚型、血清型及假型,及天然存在的及重组形式,除了另外要求时。如本文所用,术语"血清型"是指基于衣壳蛋白与界定抗血清的反应性鉴别且区别于其它AAV的AAV,例如存在许多已知的灵长类动物AAV血清型。在一些实施例中,AAV载体係选自AAV1、AAV2、AAV3、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAV10、AAV-Rh74(恒河猴源性AAV)及AAVRh10,及此些血清型的经修饰衣壳。举例来说,血清型AAV-2用于指含有自AAV-2的cap基因编码的衣壳蛋白及含有来自相同AAV-2血清型的5'及3'ITR序列的基因组的AAV。假型AAV是指含有来自一种血清型的衣壳蛋白及包括第二血清型的5'-3'ITR的病毒基因组的AAV。将预期假型rAAV具有衣壳血清型的细胞表面结合特性及与ITR血清型一致的遗传特性。假型重组AAV(rAAV)係使用所属领域中描述的标准技术产生。如本文所用,举例来说,rAAV1可用于指衣壳蛋白及5'-3'ITR均来自相同血清型的AAV,或其可指具有来自血清型1的衣壳蛋白及来自不同AAV血清型(例如AAV血清型2)的5'-3'ITR的AAV。对于本文中说明的各实例,载体设计及生产的规格描述衣壳及5'-3'ITR序列的血清型。

[0456] "AAV病毒"或"AAV病毒粒子"是指由至少一种AAV衣壳蛋白(优选野生型AAV的所有衣壳蛋白)及衣壳化多核苷酸构成的病毒粒子。如果粒子另外包含异源多核苷酸(即,除了递送至哺乳动物细胞的野生型AAV基因组以外的多核苷酸),那么其典型地称为"rAAV"。示范性异源多核苷酸是包含本文所述的任一实施例的CasX蛋白和/或sgNA及任选地供体模板的多核苷酸。

[0457] "腺相关病毒反向末端重复"或"AAVITR"意谓发现于AAV基因组的每一端处的技术识别的区域,其以顺式在一起起作用,作为DNA复制起点及病毒的封装信号。AAV ITR连同AAV rep编码区提供自插入两个侧接ITR之间的核苷酸序列的有效切除及解救,及将所述核苷酸序列整合至哺乳动物细胞基因组中。

[0458] AAV ITR区的核苷酸序列为已知的。参见,例如Kotin,R.M.(1994)Human Gene Therapy5:793-801;Berns,K.I."Parvoviridae and their Replication",Fundamental Virology,第2版,(B.N.Fields及D.M.Knipe编)。如本文所用,AAV ITR不必描绘野生型核苷酸序列,而是可经改变,例如通过核苷酸的插入、缺失或取代。另外,AAV ITR可衍生自若干AAV血清型中的任一个,包括但不限于AAV1、AAV2、AAV3、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAV10、AAV-Rh74及AAVRh10,及此些血清型的经修饰衣壳。此外,侧接AAV载体中的所选核苷

酸序列的5'及3'ITR不必相同或衍生自相同AAV血清型或分离株,只要其如所预期地起作用,即允许自宿主细胞基因组或载体切除及解救所关注序列,及允许将异源序列集成至受体细胞基因组中(当AAVRep基因产物存在于细胞中时)。使用AAV血清型将异源序列集成至宿主细胞中为所属领域中已知的(参见例如,WO2018195555A1和US20180258424A1,其通过引用并入本文)。

[0459] "AAV rep编码区"意谓编码复制蛋白Rep 78、Rep 68、Rep 52及Rep 40的AAV基因组区域。已显示这些Rep表达产物具有许多功能,包括识别、结合及切割AAV的DNA复制起点、DNA解螺旋酶活性及调节自AAV(或其它异源)启动子的转录。复制AAV基因组总体需要Rep表达产物。

[0460] "AAV cap编码区"意指编码衣壳蛋白VP1、VP2及VP3,或其功能同源物的AAV基因组区。这些Cap表达产物提供包装病毒基因组总体需要的包装功能。

[0461] 在一些实施例中,用于将CasX、gNA和任选的供体模板核苷酸递送到宿主细胞的AAV衣壳可来源于数种AAV血清型中的任一个,包括但不限于AAV1、AAV2、AAV3、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAV10、AAV-Rh74(恒河猕猴源性AAV)和AAVRh10,且AAV ITR来源于AAV血清型2。在特定实施例中,利用AAV1、AAV7、AAV6、AAV8或AAV9将CasX、gNA和任选的供体模板核苷酸递送到宿主肌肉细胞。

[0462] 为了产生rAAV病毒粒子,使用已知技术,例如通过转染将AAV表达载体引入至适合的宿主细胞中。包装细胞典型地用于形成病毒粒子;此类细胞包括封装腺病毒的HEK293或HEK293T细胞(和本文中所描述或所属领域中已知的其它细胞)。多种转染技术为所属领域中总体已知的;参见例如Sambrook等人(1989)MolecularCloning,alaboratorymanual,ColdSpringHarborLaboratories,NewYork。尤其适合的转染方法包括磷酸钙共沉淀、直接显微注射至经培养细胞中、电穿孔、脂质粒介导的基因转移、脂质介导的转导及使用高速微弹的核酸递送。

[0463] 在一些实施例中,经上述AAV表达载体转染的宿主细胞使得能够提供AAV辅助功能,以便复制及衣壳化由AAV ITR侧接的核苷酸序列,以产生rAAV病毒粒子。AAV辅助功能一般为AAV源性编码序列,其可经表达以得到AAV基因产物,所述产物转而以反式起作用以进行生产性AAV复制。AAV辅助功能在本文中用于补充自AAV表达载体缺失的所需AAV功能。因此,AAV辅助功能包括一种或两种编码rep及cap编码区的AAV ORF(开放阅读框架),或其功能同源物。可使用所属领域技术人员已知的方法将辅助功能引入至宿主细胞中且接着表达于宿主细胞中。通常,辅助功能通过用无关的辅助病毒感染宿主细胞来提供。在一些实施例中,辅助功能係使用辅助功能载体提供。取决于所利用的宿主/载体系统,多种适合的转录及翻译控制元件(包括组成型及诱导型启动子、转录增强子元件、转录终止子等)中的任一个可用于表达载体中。在一些实施例中,本公开提供包含本文所公开的实施例的AAV载体的宿主细胞。

[0464] 在其它实施例中,适合的载体可以包括病毒样粒子(VLP)。病毒样粒子(VLP)是与病毒非常相似的粒子,但不含病毒遗传物质,因此不具有传染性。在一些实施例中,VLP包含编码所关注的转基因,例如CasX蛋白和/或gNA实施例中的任一个的多核苷酸,和本文所述的任选的供体模板多核苷酸,其用一种或多种病毒结构蛋白质封装。

[0465] 在其它实施例中,本公开提供体外产生的VLP,其包含CasX:gNA RNP复合物和任选

的供体模板。来自不同病毒的结构蛋白质的组合可以用于产生VLP,包括来自病毒科的组分,所述病毒科包括细小病毒科(例如,腺相关病毒)、逆转录病毒科(例如α逆转录病毒、β逆转录病毒、γ逆转录病毒、δ逆转录病毒、ε逆转录病毒或慢病毒)、黄病毒科(例如,丙型肝炎病毒)、副粘病毒科(例如,尼帕病毒(Nipah))和噬菌体(例如,Qβ、AP205)。在一些实施例中,本公开提供了使用逆转录病毒组分设计的VLP系统,包括慢病毒(如HIV)和α逆转录病毒、β逆转录病毒、γ逆转录病毒、δ逆转录病毒、ε逆转录病毒,其中将包含编码各种组分的多核苷酸的个别质粒引入包装细胞中,继而产生VLP。在一些实施例中,本公开提供VLP,其包含一种或多种以下组分:i)蛋白酶,ii)蛋白酶裂解位点,iii)选自以下的gag多聚蛋白的一种或多种组分:基质蛋白质(MA)、核衣壳蛋白质(NC)、衣壳蛋白质(CA)、p1肽、p6肽、P2A肽、P2B肽、P10肽、p12肽、PP21/24肽、P12/P3/P8肽和P20肽;v)CasX;vi)gNA,和vi)靶向糖蛋白或抗体片段,其中所得VLP粒子衣壳化CasX:gNA RNP。靶向糖蛋白或抗体片段在表面上提供VLP到靶细胞的向性,其中在施用和进入到靶细胞中之后,RNP分子自由运输到细胞的细胞核中。在其它实施例中,本公开提供前述VLP并且进一步包含一种或多种pol多聚蛋白(例如蛋白酶)组分和任选的第二CasX或供体模板。前述提供优于所属领域其它载体的优势在于病毒转导至分裂和非分裂细胞是有效的,并且VLP递送有效且短寿命的RNP,其逃避原本会检测到外来蛋白质的受试者的免疫监视机制。

[0466] 在一些实施例中,本公开提供包含编码一种或多种选自以下的组分的多核苷酸或载体的宿主细胞:i)一种或多种gag多聚蛋白组分(其组分如上文所列举);ii)本文所描述的实施例中的任一个的CasX蛋白;iii)蛋白酶裂解位点;iv)蛋白酶;v)本文所述的任何实施例的引导RNA;vi)pol多聚蛋白或其部分(例如蛋白酶);vii)假型化糖蛋白或抗体片段,其提供VLP与靶细胞的结合和融合;以及viii)供体模板。本公开涵盖经编码组分的布置的多种配置,包括一些经编码组分的重复。包膜糖蛋白可以衍生自所属领域中已知的赋予VLP向性的任何包膜病毒,包括但不限于由以下组成的组:阿根廷出血热病毒、澳大利亚蝙蝠病毒、苜蓿丫纹夜蛾(Autographa californica)多核多角体病毒、禽白血病病毒、狒狒内源性病毒、玻利维亚出血热病毒、博纳病(Borna disease)病毒、布雷达(Breda)病毒、布尼亚姆韦拉(Bunyamwera)病毒、昌迪普拉(Chandipura)病毒、基孔肯雅(Chikungunya)病毒、克里米亚-刚果出血热病毒、登革热病毒、杜文哈奇(Duvenhage)病毒、东部马脑炎病毒、埃博拉出血热病毒、埃博拉扎伊尔(Zaire)病毒、肠道腺病毒、暂时热病毒、艾-巴二氏病毒(Epstein-Bar virus,EBV)、欧洲蝙蝠病毒1、欧洲蝙蝠病毒2、Fug合成gP融合病毒、长臂猿白血病病毒、汉坦病毒(Hantavirus)、亨德拉(Hendra)病毒、甲型肝炎病毒、乙型肝炎病毒、丙型肝炎病毒、丁型肝炎病毒、戊型肝炎病毒、庚型肝炎病毒(GB病毒C)、单纯疱疹病毒1型、单纯疱疹病毒2型、人类巨细胞病毒(HHV5)、人类泡沫病毒、人类疱疹病毒(HHV)、人类疱疹病毒7、人类疱疹病毒6型、人类疱疹病毒8型、人类免疫缺陷病毒1(HIV-1)、人类偏肺病毒(metapneumovirus)、人类嗜T淋巴细胞病毒1、甲型流感、乙型流感、丙型流感病毒、日本脑炎病毒、卡波西肉瘤(Kaposi's sarcoma)相关疱疹病毒(HHV8)、凯萨努尔森林病(Kaysanur Forest disease)病毒、拉克罗斯(La Crosse)病毒、拉各斯(Lagos)蝙蝠病毒、拉沙热(Lassa fever)病毒、淋巴细胞脉络丛脑膜炎病毒(LCMV)、马丘波(Machupo)病毒、马尔堡(Marburg)出血热病毒、麻疹病毒、中东呼吸综合征相关冠状病毒、莫科拉(Mokola)病毒、莫洛尼(Moloney)鼠类白血病病毒、猴痘病毒、小鼠乳房肿瘤病毒、腮腺炎病毒、鼠类丙型疱疹

病毒、新城疫病毒、尼帕病毒、尼帕病毒、诺沃克(Norwalk)病毒、鄂木斯克(Omsk)出血热病毒、乳头状瘤病毒、细小病毒、伪狂犬病病毒、夸兰菲尔(Quaranfil)病毒、狂犬病病毒、RD114内源性猫逆转录病毒、呼吸道合胞病毒(RSV)、裂谷热病毒、罗斯河(Ross River)病毒、r轮状病毒(rRotavirus)、劳氏肉瘤(Rous sarcoma)病毒、风疹病毒、萨比亚(Sabia)相关出血热病毒、SARS相关冠状病毒(SARS-CoV)、仙台病毒、塔卡里伯(Tacaribe)病毒、索戈托病毒(Thogotovirus)、蜱传致脑炎病毒、水痘带状疱疹病毒(HHV3)、水痘带状疱疹病毒(HHV3)、重型天花病毒、轻型天花病毒、委内瑞拉马脑炎病毒、委内瑞拉出血热病毒、水疱性口炎病毒(VSV)、VSV-G、水疱病毒、西尼罗河病毒、西部马脑炎病毒和寨卡(Zika)病毒。在一些实施例中,用于产生VLP的包装细胞选自由以下组成的组:HEK293细胞、Lenti-X HEK293T细胞、BHK细胞、HepG2细胞、Saos-2细胞、HuH7细胞、NSO细胞、SP2/0细胞、YO骨髓瘤细胞、A549细胞、P3X63小鼠骨髓瘤细胞、PER细胞、PER.C6细胞、杂交瘤细胞、VERO细胞、NIH3T3细胞、COS细胞、WI38细胞、MRC5细胞、A549细胞、HeLa细胞、CHO细胞或HT1080细胞。

[0467] 在生产和回收包含本文所述任一实施例的CasX:gNA RNP的VLP后,VLP可用于通过施用这种VLP来编辑受试者的靶细胞的方法,如下文更充分描述。

[0468] VII.细胞

[0469] 在另外其它方面,本文提供包含通过本文所述的CasX:gNA系统实施例中的任一个修饰的C9orf72基因的细胞。在一些情况下,可以出于如基因疗法的目的向受试者施用已以此方式遗传修饰的细胞,例如以治疗与C9orf72基因中的缺陷相关的疾病。在其它情况下,细胞在患有C9orf72相关疾病的受试者体内经修饰。在一些实施例中,本公开提供已经修饰以切除C9orf72基因的六核苷酸重复序列扩增区域,使得表达功能C9orf72蛋白质的细胞群体。在前述的一些情况下,待修饰的细胞包含C9orf72基因中的一个或多个突变,其破坏C9orf72蛋白质的功能或表达。在前述的其它情况下,待修饰的细胞包含C9orf72基因中的HRS扩增段,使得产生过量RNA或DPR蛋白质且并入到细胞中。在前述的其它情况下,待修饰的细胞包含SEQ ID NO:227或228的C9orf72蛋白质的一个或多个突变或截短。

[0470] 在一些实施例中,细胞群体通过V型Cas核酸酶和一个或多个靶向至接近与C9orf72靶核酸的六核苷酸重复序列扩增区域结合的序列的引导序列修饰。在一些实施例中,本公开提供通过向群体的每个细胞中引入以下来修饰细胞的方法和群体:i)CasX:gNA系统,其包含本文所述的实施例中的任一个的CasX和gNA;ii)CasX:gNA系统,其包含本文所述的实施例中的任一个的CasX、gNA和供体模板;iii)编码CasX和gNA,并且任选地包含供体模板的核酸;iv)选自由以下组成的组的载体:逆转录病毒载体、慢病毒载体、腺病毒载体、腺相关病毒(AAV)载体和单纯疱疹病毒(HSV)载体,并且所述载体包含以上(iii)的核酸;v)VLP,其包含本文所述的实施例中的任一个的CasX:gNA系统;或vi)(i)到(v)中的两个或更多个的组合,其中通过gNA靶向的细胞的靶核酸序列通过CasX蛋白和任选的供体模板修饰。在前述中,供体模板包含C9orf72基因的至少一部分,其中C9orf72基因部分选自C9orf72外显子、C9orf72内含子、C9orf72内含子-外显子接界、C9orf72调节元件(例如启动子)、C9orf72编码区、C9orf72非编码区或其组合,或C9orf72基因的全部,并且对细胞的修饰使得突变校正成野生型序列、置换六核苷酸重复序列扩增区域的全部或一部分或敲减或敲除C9orf72基因。在一些情况下,供体模板可包含编码SEQ ID NO:227或228的序列的全部或一部分的核酸,或包含跨越人类基因组的chr9:27,546,546到27,573,866(GRCh37/hg19)(记

法是指染色体4(chr4),在所述染色体的27,546,546bp开始,并且延伸到所述染色体的27,573,866bp)或其一部分的全部或一部分的多核苷酸序列。在其它情况下,供体模板可包含相较于野生型C9orf72基因的异源序列,以便敲减或敲除基因。在又其它情况下,供体模板包含GGGGCC序列的六核苷酸重复序列,其中重复序列的数目在10到约30个重复序列的范围内。在前述中,供体模板将用于置换具有数百到数千个六核苷酸重复序列的细胞的有缺陷的序列。供体模板将进一步包含在核酸酶引入的裂解位点的5'和3'处的同源臂,以促进其通过HDR的插入。供体模板的大小可在10到30,000个核苷酸或20到10,000个核苷酸或100到1000个核苷酸的范围内。在一些情况下,供体模板是单链DNA模板或单链RNA模板。在其它情况下,供体模板是双链DNA模板。在一些情况下,使细胞与CasX和至少第一gNA接触,其中gNA为引导RNA(gRNA)。在一些情况下,使细胞与CasX和至少第一和第二gNA接触,其中gNA为引导RNA(gRNA)。在其它情况下,细胞与CasX和gNA接触,其中gNA是引导DNA(gDNA)。在其它情况下,细胞与CasX和gNA接触,其中gNA是包含DNA和RNA的嵌合体。如本文所述,在任一种组合的实施例中,所述gNA分子中的每一个(支架和靶向序列的组合,其可以被配置为sgRNA或dgRNA)可以与本文所述的CasX实施例一起作为RNP提供,以并入到实施例的细胞中。在一些实施例中,群体的细胞与包含如表4、6-8和10中所阐述的SEQ ID NO:49-150、233-235、238-252或272-281的序列,或与其具有至少65%同一性、至少70%同一性、至少75%同一性、至少80%同一性、至少81%同一性、至少82%同一性、至少83%同一性、至少84%同一性、至少85%同一性、至少86%同一性、至少86%同一性、至少87%同一性、至少88%同一性、至少89%同一性、至少89%同一性、至少90%同一性、至少91%同一性、至少92%同一性、至少93%同一性、至少94%同一性、至少95%同一性、至少96%同一性、至少97%同一性、至少98%同一性、至少99%同一性或至少99.5%同一性的序列的CasX的RNP接触,gNA支架包含如表2中所阐述的SEQ ID NO:2101-2294的序列,或与其具有至少65%同一性、至少70%同一性、至少75%同一性、至少80%同一性、至少81%同一性、至少82%同一性、至少83%同一性、至少84%同一性、至少85%同一性、至少86%同一性、至少86%同一性、至少87%同一性、至少88%同一性、至少89%同一性、至少89%同一性、至少90%同一性、至少91%同一性、至少92%同一性、至少93%同一性、至少94%同一性、至少95%同一性、至少96%同一性、至少97%同一性、至少98%同一性、至少99%同一性、至少99.5%同一性的序列,并且gNA包含SEQ ID NO:309-343、363-2100和2295-21835的靶向序列,或与其具有至少65%同一性、至少70%同一性、至少75%同一性、至少80%同一性、至少85%同一性、至少90%同一性或至少95%同一性并且具有15与21个之间的氨基酸的序列。

[0471] 在一些实施例中,经修饰的细胞的经修饰的C9orf72基因包含单链断裂,产生通过细胞的修复机制的突变、插入或缺失。在其它实施例中,细胞的经修饰的C9orf72基因包含双链断裂,产生通过细胞的修复机制的突变、插入或缺失。举例来说,CasX:gNA系统可在C9orf72基因的起始点处或附近将例如移码突变的插入缺失引入到细胞中。在一些实施例中,细胞通过与以下接触来修饰:CasX、和靶向六核苷酸重复序列扩增区域的5'处的靶核酸的第一gNA、和靶向六核苷酸重复序列扩增区域的3'处的靶核酸的第二gNA,其中六核苷酸重复序列扩增区域从C9orf72基因切除,其中所述修饰使得细胞能够产生野生型或功能C9orf72蛋白质。在一些实施例中,细胞群体已经修饰,使得与尚未修饰的细胞相比,六核苷酸转录物RNA或DPR的表达降低了至少约10%、至少约20%、至少约30%、至少约40%、至少

约50％、至少约60％、至少约70％、至少约80％、至少约90％、或至少约95％。在其它实施例中，至少30％、至少40％、至少50％、至少60％、至少05％、至少75％、至少80％、至少85％、至少90％或至少95％的经修饰细胞不表达可检测水平的六核苷酸转录物RNA或DPR。在一些实施例中，第一gNA靶向序列选自由SEQ ID NO:310和319-320组成的组，且第二gNA靶向序列选自由SEQ ID NO:321-325组成的组。六核苷酸转录物RNA或DPR的表达的降低或消除可以通过ELISA或电化学发光分析测量，有义G4C2重复转录物可以通过RNA荧光原位杂交(FISH)分析(Batra,R和Lee,C.《肌肉萎缩性侧索硬化/额颞叶型痴呆中C9orf72六核苷酸重复序列扩增段的小鼠模型(Mouse Models of C9orf72 Hexanucleotide Repeat Expansion in Amyotrophic Lateral Sclerosis/Frontotemporal Dementia)》《细胞神经科学前沿(Frontiers Cell.Neurosci.)》11:196(2017))或所属领域中已知的其它方法，或如实例中所述分析。在一些实施例中，本公开提供细胞群体，其经修饰以使得与尚未修饰的细胞相比，功能C9orf72蛋白质的表达增加了至少约10％、至少约20％、至少约30％、至少约40％、至少约50％、至少约60％、至少约70％、至少约80％、或至少约90％。

[0472] 在一些情况下，体外进行包含一个或多个突变或重复的细胞的C9orf72基因的修饰。在此类情况下，接着可以向受试者施用经修饰的细胞群体。可以通过任何适合的方法将RNP引入待修饰的细胞中，包括通过电穿孔、注射、核转染、通过脂质体递送、通过纳米粒子递送或使用与CasX:gNA的一种或多种组分结合的蛋白质转导结构域(PTD)。在其它情况下，CasX和一种或多种gNA使用载体作为编码多核苷酸引入到细胞群体中；本文中描述其实施例。使用CasX:gNA系统组分修饰细胞的额外方法包括病毒感染、转染、缀合、原生质体融合、粒子枪技术、磷酸钙沉淀、直接显微注射等。方法的选择通常取决于被转化细胞的类型和发生转化的环境；例如，体外、离体或体内。对这些方法的一般性讨论可见于Ausubel等人，《分子生物学简短协议(Short Protocols in Molecular Biology)》，第3版，Wiley&Sons出版社，1995。

[0473] 在其它情况下，在体内进行包含一个或多个突变或重复的细胞的C9orf72基因的修饰。在此类情况下，向受试者施用CasX和gNA以及任选的供体模板。在其它情况下，CasX和gNA以及任选的供体模板在编码CasX和一种或多种gNA并且任选地含有供体模板的载体内施用到受试者。在又其它情况下，CasX和gNA以及任选的供体模板在如衣壳化RNP的VLP并且任选地含有供体模板的载体内施用到受试者。在前述中，修饰校正一个或多个突变，或在替代方案中，修饰是抑制或遏制六核苷酸转录物RNA或DPR的表达、表达功能C9orf72蛋白质，或表达野生型或功能C9orf72蛋白质。

[0474] 可以充当本公开的CasX蛋白和/或gNA，和/或包含编码CasX蛋白和/或CasX gNA变异体的核苷酸序列和任选的供体模板的核酸的接受者的细胞可以是多种细胞中的任一种，包括例如体外细胞；体内细胞；离体细胞；原代细胞；癌细胞；动物细胞等。细胞可以是本公开的CasX RNP的接受者。细胞可为本公开的CasX系统的单一组分的受体。在某些实施例中，如本文所提供，细胞可以是体外细胞(例如确立的培养细胞系，包括但不限于HEK293细胞、HEK293-F细胞、BHK细胞、NS0细胞、SP2/0细胞、YO骨髓瘤细胞、P3X63小鼠骨髓瘤细胞、PER细胞、PER.C6细胞、杂交瘤细胞、NIH3T3细胞、COS、HeLa或CHO细胞)。细胞可以是离体细胞(来自个体的培养细胞)。细胞可为体内细胞(例如个体中的细胞)。细胞可为经分离细胞。细胞可为生物体内的细胞。细胞可为生物体。细胞可以是细胞培养物(例如，体外细胞培养物)中

的细胞。细胞可为细胞集合中的一个。细胞可为动物细胞或衍生自动物细胞。细胞可为脊椎动物细胞或可衍生自脊椎动物细胞。细胞可为哺乳动物细胞或衍生自哺乳动物细胞。细胞可为啮齿动物细胞或衍生自啮齿动物细胞。细胞可以是非人类灵长类动物细胞或衍生自非人类灵长类动物细胞。细胞可为人类细胞或衍生自人类细胞。

[0475] 在一些实施例中,经修饰细胞为真核细胞,其中真核细胞选自由以下组成的组:啮齿动物细胞、小鼠细胞、大鼠细胞、灵长类动物细胞、非人类灵长类动物细胞和人类细胞。在一些实施例中,经修饰的细胞为人类细胞。在其它实施例中,细胞相对于待施用细胞的受试者为自体的。在其它实施例中,细胞相对于待施用细胞的受试者为同种异体的。在一些实施例中,经修饰细胞为中枢神经系统(CNS)的细胞。在一些实施例中,经修饰的细胞选自由以下组成的组:普尔基涅细胞、额皮质神经元、运动皮质神经元、海马神经元、小脑神经元、上运动神经元、脊髓神经元、脊髓运动神经元、神经胶质细胞和星形胶质细胞。在前述中,细胞群体在治疗C9orf72相关疾病中具有效用,其中细胞群体向患有C9orf72相关疾病的受试者施用。在前述的一些情况下,待修饰的细胞包含C9orf72基因中的一个或多个突变,其破坏C9orf72蛋白质的功能或表达。在前述的其它情况下,待修饰的细胞包含C9orf72基因中的HRS扩增段,使得产生过量RNA或DPR蛋白质且并入到细胞中。在前述的其它情况下,待修饰的细胞包含SEQ ID NO:227或228的C9orf72蛋白质的一个或多个突变或截短。

[0476] 在其它实施例中,本公开提供用于患有C9orf72相关疾病的受试者的经修饰细胞的群体。在一些实施例中,本公开提供一种治疗患有C9orf72相关疾病的受试者的方法,所述方法包含向所述受试者施用有效量的本文所描述的实施例中的任一个的多个经修饰细胞,其中所述经修饰细胞表达生理学上正常水平的C9orf72。在一些实施例中,C9orf72相关疾病选自由以下组成的组:肌肉萎缩性侧索硬化(ALS)和额颞叶型痴呆(FTD)。

[0477] VIII.应用

[0478] 本文所提供的包含CasX蛋白、引导序列和其变异体的CasX:gNA系统可用于在包括治疗、诊断和研究的各种应用中修饰C9orf72靶核酸序列的方法。

[0479] 在本文所述的修饰细胞中的C9orf72靶核酸序列的方法中,所述方法利用本文所述的CasX:gNA系统的实施例中的任一个,并且任选地包括本文所述的供体模板。在一些情况下,所述方法敲减突变C9orf72的表达。在其它情况下,所述方法敲除突变C9orf72的表达。在又其它情况下,所述方法产生功能C9orf72蛋白质的表达。

[0480] 在一些实施例中,所述方法包含使靶核酸序列与CasX蛋白和包含靶向序列的引导核酸(gNA)接触,其中所述接触引起通过CasX蛋白对靶核酸序列的修饰。在一些实施例中,所述方法包含将CasX蛋白或编码CasX蛋白的核酸和gNA或编码gNA的核酸引入到细胞中,其中靶核酸序列包含C9orf72基因,并且其中靶向序列包含与编码C9orf72蛋白质的C9orf72基因的一部分、C9orf72调节元件或C9orf72编码序列和C9orf72调节元件两者互补的序列,其中所述接触引起C9orf72基因的修饰。在一些实施例中,gNA的靶向序列包含SEQ ID NO:309-343、363-2100和2295-21835的序列,或与其具有至少约65%、至少约75%、至少约85%或至少约95%同一性的序列。在一些实施例中,gNA的支架包含SEQ ID NO:4、5或2101-2294的序列,或与其具有至少约50%、至少约60%、至少约70%、至少约80%、至少约90%、至少约95%、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,CasX蛋白为本文所述的任何实施例的CasX变异蛋白质,或参考

CasX蛋白SEQ ID NO:1、SEQ ID NO:2或SEQ ID NO:3。

[0481] 在一些实施例中,经修饰的细胞的经修饰的C9orf72基因包含单链断裂,产生通过细胞的修复机制的突变、插入或缺失。在其它实施例中,经修饰的细胞的经修饰的C9orf72基因包含双链断裂,产生通过细胞的修复机制的突变、插入或缺失。举例来说,CasX:gNA系统可在C9orf72基因的起始点处或附近将例如移码突变的插入缺失引入到细胞中。在其它实施例中,细胞的经修饰的C9orf72基因已经通过插入其中C9orf72基因已经敲减或敲除的供体模板来修饰。在前述中,细胞已经修饰,使得与尚未修饰的细胞相比,HRS或DPR蛋白质的表达降低了至少约10%、至少约20%、至少约30%、至少约40%、至少约50%、至少约60%、至少约70%、至少约80%或至少约90%。在其它实施例中,至少70%、至少75%、至少80%、至少85%、至少90%或至少95%的经修饰的细胞不表达可检测水平的HRS RNA或DPR。HRS RNA或DPR蛋白质表达的降低或消除可以通过ELISA或电化学发光分析(Mcdonald,D.等人《用于总计和聚谷氨酰胺扩增亨廷顿蛋白的定量分析(Quantification Assays for Total and Polyglutamine-Expanded Huntingtin Proteins)》PLoS ONE 9(5):e96854(2014))或所属领域中已知的其它方法,或如实例中所述测量。

[0482] 在修饰C9orf72靶核酸序列的方法的一些实施例中,靶核酸序列包含具有一个或多个突变或重复的C9orf72基因,并且gNA的靶向序列具有与C9orf72基因互补并且因此可以与其杂交的序列。在一些情况下,C9orf72基因具有野生型核酸序列。在其它实施例中,所述方法包含使靶核酸序列与靶向到具有一个或多个突变或重复的C9orf72基因的不同或重叠区域的多种(例如,两种或更多种)gNA接触。在所述方法的一些实施例中,靶核酸为DNA。在所述方法的一些实施例中,靶核酸为RNA。在一些实施例中,gNA为引导RNA(gRNA)。在一些实施例中,gNA为引导DNA(gDNA)。在一些实施例中,gNA为单分子gNA(sgNA)。在其它实施例中,gNA为双分子gNA(dgNA)。在一些实施例中,gNA为嵌合gRNA-gDNA。在一些实施例中,所述方法包含使靶核酸序列与预复合CasX蛋白-gNA(即,RNP)接触。在一些实施例中,C9orf72基因包含突变或重复并且所述修饰包含在靶核酸中引入单链断裂。在其它实施例中,C9orf72基因包含突变或重复并且所述修饰包含在靶核酸中引入双链断裂。

[0483] 在前述中,所得修饰可以是与野生型序列相比的一个或多个核苷酸的插入、缺失、取代、重复或倒位。在一些实施例中,所述修饰校正功能增加突变。在其它实施例中,所述修饰校正功能缺失突变。待修饰的突变可以包含破坏C9orf72蛋白质的功能或表达的一个或多个突变或重复。

[0484] 在一些实施例中,修饰靶核酸序列的方法包含使C9orf72基因与CasX蛋白和gNA对和供体模板接触,所述供体模板包含可以在CasX引入的断裂位点处插入或敲入的校正性序列。举例来说,可以包含待整合的校正性序列(或缺失或插入,以敲除有缺陷的序列)的外源性供体模板侧接与靶核酸序列具同源性以促进其引入到细胞中的上游序列和下游序列(例如同源臂)。在一些实施例中,供体模板的大小范围为10-10,000个核苷酸。在其它实施例中,供体模板的大小范围为100-1,000个核苷酸。在一些实施例中,供体模板为单链DNA模板或单链RNA模板。在其它实施例中,供体模板是双链DNA模板。

[0485] 在所述方法的一些实施例中,CasX为非催化活性CasX(dCasX)蛋白质,其保留与gNA和包含突变的靶核酸序列结合的能力,进而干扰突变C9orf72的转录。在一些实施例中,所述方法包含使C9orf72基因与CasX蛋白和gNA接触并且不包含使靶核酸序列与供体模板

多核苷酸接触，并且靶核酸序列通过CasX核酸酶裂解并且经修饰使得靶核酸序列内的核苷酸根据细胞自身的修复路径删除或插入。在一些实施例中，编辑在体内在细胞内部发生，例如在生物体或受试者的细胞中发生。在一些实施例中，所述细胞为真核细胞。示范性真核细胞可以包括选自由以下组成的组的细胞：啮齿动物细胞、小鼠细胞、大鼠细胞、灵长类动物细胞、非人类灵长类动物细胞和人类细胞。在一些实施例中，细胞为人类细胞。在一些实施例中，细胞为非人类灵长类细胞。在所述方法的一些实施例中，所述细胞选自由以下组成的组：普尔基涅细胞、额皮质神经元、运动皮质神经元、海马神经元、小脑神经元、上运动神经元、脊髓神经元、脊髓运动神经元、神经胶质细胞和星形胶质细胞。

[0486] 将核酸(例如，包含供体多核苷酸序列的核酸、编码CasX蛋白和/或gNA的一种或多种核酸)引入到细胞中的方法为所属领域中已知的，且可以使用任何方便的方法将核酸(例如，表达构建体)引入到细胞中。适合的方法包括例如病毒感染或与具有对于靶细胞的向性的病毒样粒子(VLP)接触。逆转录病毒，例如慢病毒，可适用于本公开的方法。常用的逆转录病毒载体是“缺陷性的”，例如不能产生生产性感染所需的病毒蛋白。确切而言，载体的复制需要包装细胞系中的生长。为了产生包含所关注核酸的病毒粒子，通过包装细胞系将包含所述核酸的逆转录病毒核酸包装至病毒衣壳中。不同包装细胞系提供并入至衣壳中的不同包膜蛋白(亲嗜性、双嗜性或嗜异性)，且此包膜蛋白决定病毒粒子对细胞的特异性(对鼠类及大鼠为亲嗜性的；对大部分哺乳动物细胞类型，包括人类、狗及小鼠为双嗜性的；且对除了鼠类细胞的大部分哺乳动物细胞类型为嗜异性的)。适当包装细胞系可用于确保细胞经包装病毒粒子靶向。将主题载体表达载体引入包装细胞系中以及收集通过包装细胞系产生的病毒粒子的方法在所属领域中众所周知，包括美国专利号5,173,414;Tratschin等人，《分子与细胞生物学(Mol.Cell.Biol.)》5:3251-3260(1985)；Tratschin等人，《分子与细胞生物学》4:2072-2081(1984)；Hermonat&Muzyczka,PNAS 81:6466-6470(1984)；及Samulski等人，《病毒学杂志(J.Virol.)》63:03822-3828(1989)。也可以通过直接微量注射(例如，RNA注射)引入核酸。

[0487] 在其它实施例中，本公开涉及产生本文所述的任何实施例的CasX蛋白和编码CasX组合物的核酸，或与多核苷酸序列互补的序列，包括其同源变异体的方法，以及表达由多核苷酸序列表达的CasX蛋白的方法。本公开的CasX蛋白可以通过真核细胞或原核细胞在体外产生。对于由宿主细胞产生，一般来说，方法包括产生编码本文所描述的任何实施例的CasX蛋白的多核苷酸序列，并且将编码基因并入适合于宿主细胞的表达载体中。为了产生本文所述的任何实施例的所编码CasX蛋白，所述方法包括用表达载体转化适合的宿主细胞，并且在引起或允许所得CasX蛋白在经转化的宿主细胞中表达的条件下培养所述宿主细胞，从而产生CasX蛋白，所述CasX蛋白通过本文所述的方法或所属领域中已知的标准蛋白质纯化方法回收。分子生物学中的标准重组技术用于制备本公开的多核苷酸和表达载体。

[0488] 在改变细胞的C9orf72靶核酸序列或诱导靶核酸序列的裂解的方法的一些实施例中，本公开的CasX gNA和/或CasX蛋白和/或供体模板序列(无论其作为核酸或多肽引入)通过本文所述的实施例的载体或粒子提供到细胞。向细胞提供载体或粒子可以约每天到约每4天的频率，例如每1.5天、每2天、每3天，或约每天到约每四天的任何其它频率，或每周或每月重复。药剂可以向主题细胞提供一次或多次，例如一次、两次、三次或超过三次。

[0489] 在将两种或更多种不同靶向复合物提供到细胞(例如具有不同靶向序列的两种

CasX gNA)的实施例中,复合物可以同时提供(例如作为两种多肽和/或核酸),或同时递送。或者,其可连续提供,例如首先提供靶向复合物,接着提供第二靶向复合物等,或反之亦然。

[0490] 为了改进DNA载体向靶细胞的递送,可保护DNA免受损伤且例如通过使用脂质复合体及聚合复合体促进其进入细胞。因此,在一些情况下,本公开的核酸(例如本公开的重组表达载体)可以经如微胞或脂质体的组织结构中的脂质覆盖。当组织化结构与DNA复合时,其係称为脂质复合体。存在三种类型的脂质:阴离子型(带负电)、中性或阳离子型(带正电)。已证明利用阳离子型脂质的脂质复合体适用于基因转移。阳离子型脂质由于其正电荷而天然地与带负电DNA复合。并且,由于其电荷,其与细胞膜相互作用。接着发生脂质复合体的内饮作用,且DNA释放至细胞质中。阳离子型脂质亦保护DNA免于被细胞降解。

[0491] 聚合物与DNA的复合物係称为聚合复合体。大多数聚合复合体由阳离子聚合物组成,且其产生通过离子相互作用调节。聚合复合物和脂质体复合物的作用方法之间的一个较大差异为聚合复合物无法将其DNA负载释放到细胞质中,因此为此目的,必须用内体溶解剂(以溶解在内饮作用期间制造的内体),如灭活腺病毒进行共转染。然而,情况并非始终如此;例如聚乙烯亚胺的聚合物自身具有内体破坏方法,聚葡萄胺糖及三甲基聚葡萄胺糖亦如此。

[0492] 树枝状聚合物(具有球形的高度分支大分子)亦可用于遗传修饰干细胞。树枝状聚合物粒子的表面可经官能化以改变其特性。特定言的,有可能构建阳离子树枝状聚合物(即,具有正表面电荷的树枝状聚合物)。在例如DNA质粒的遗传物质存在下时,电荷互补使得核酸与阳离子树枝状聚合物暂时缔合。到达其目的地时,树枝状聚合物-核酸复合物可通过内饮作用溶解于细胞中。

[0493] IX.治疗方法

[0494] 本公开提供治疗有需要的受试者的C9orf72相关疾病的方法,其包括但不限于肌肉萎缩性侧索硬化(ALS)和额颞叶型痴呆(FTD)。在一些实施例中,本公开的方法可以通过向受试者施用本公开组合物来预防、治疗和/或改善受试者的C9orf72相关疾病。许多治疗策略已经用于设计用于治疗患有C9orf72相关疾病的受试者的方法的组合物。另外,所述方法可以用于在C9orf72相关疾病的任何症状之前治疗受试者。因此,预防施用经修饰的细胞群体或治疗有效量的实施例的CasX:gNA系统组合物或编码CasX:gNA系统的多核酸可用于预防C9orf72相关疾病。在一些实施例中,向受试者施用的组合物还包含药学上可接受的载剂、稀释剂或赋形剂。

[0495] 在一些情况下,受试者的C9orf72基因的等位基因中的一个包含HRS。在一些情况下,受试者的C9orf72基因的一个或两个等位基因包含突变。在其它情况下,受试者的C9orf72基因的一个或两个等位基因包含C9orf72基因的至少一部分的重复。在其它情况下,受试者的C9orf72基因的一个或两个等位基因包含C9orf72基因的重复。在其它情况下,C9orf72基因编码改变C9orf72蛋白质的功能或表达的突变,例如但不限于相比于野生型序列的一个或多个核苷酸的取代、缺失或插入。

[0496] 在一些实施例中,本公开提供了治疗有需要的受试者的C9orf72或相关疾病的方法,所述方法包含修饰受试者细胞中的C9orf72基因,所述修饰包含使所述细胞与治疗有效剂量的以下物质接触:i)包含本文描述的任一实施例的CasX和gNA的组合物;ii)包含本文描述的任一实施例的CasX、gNA和供体模板的组合物;iii)编码或包含(i)或(ii)的组合物

的一种或多种核酸;iv)载体,其选自由逆转录病毒载体、慢病毒载体、腺病毒载体、腺相关病毒(AAV)载体、单纯疱疹病毒(HSV)载体组成的组,并且包含(iii)的核酸;v)包含(i)或(ii)的组合物的VLP;或vi)(i)-(v)中两个或更多个的组合,其中细胞的C9orf72基因被CasX蛋白和任选的供体模板修饰,使得表达野生型或功能C9orf72蛋白质。在治疗受试者的C9orf72相关疾病的方法的一些实施例中,利用第二gNA,其中第二gNA具有与相较于第一gNA的靶核酸的不同或重叠部分互补的靶向序列(例如,在六核苷酸重复序列扩增段的5'和3'处),引起受试者的细胞的C9orf72靶核酸中的额外断裂。在前述中,基因可以通过NHEJ宿主修复机制修饰,或与通过HDR或HITI机制插入以切除、校正或补偿突变的供体模板结合使用,使得经修饰的细胞中的野生型或功能C9orf72蛋白质的表达相较于尚未修饰的细胞增加至少约50%、至少约60%、至少约70%、至少约80%、至少约90%、或至少约95%。在一些实施例中,通过施用上文(i)-(v)的模态治疗的方法引起C9orf72基因的敲减或敲除,使得经修饰的细胞中的HRS RNA和/或DPR的表达相较于尚未修饰的细胞降低至少约10%、至少约20%、至少约30%、至少约40%、至少约50%、至少约60%、至少约70%、至少约80%、或至少约90%。在治疗方法的实施例中,C9orf72相关疾病包括由于受试者中HRS RNA和/或DPR的表达、C9orf72的突变、C9orf72基因的重复或C9orf72的异常表达而出现的所有疾病。下文更完整地详述段落的实施例。

[0497] 在一些实施例中,所述方法包含施用载体,所述载体包含或编码CasX和靶向到C9orf72基因中的不同位置的多种gNA,其中受试者的细胞与CasX:gNA复合物的接触引起细胞的靶核酸的修饰。

[0498] 在一些实施例中,以治疗有效剂量向受试者施用实施例的载体。在一个特定实施例中,载体为本文所述的实施例的AAV,其编码CasX:gNA系统的组分和任选的供体模板。在前文中,AAV载体选自AAV1、AAV2、AAV3、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAV10、AAV11、AAV12、AAV 44.9、AAV-Rh74或AAVRh10。在一些实施例中,AAV载体以下列剂量向受试者施用:至少约1×10^{5}载体基因组/kg(vg/kg)、至少约1×10^{6}vg/kg、至少约1×10^{7}vg/kg、至少约1×10^{8}vg/kg、至少约1×10^{9}vg/kg、至少约1×10^{10}vg/kg、至少约1×10^{11}vg/kg、至少约1×10^{12}vg/kg、至少约1×10^{13}vg/kg、至少约1×10^{14}vg/kg、至少约1×10^{15}vg/kg、或至少约1×10^{16}vg/kg。在一些实施例中,AAV载体以下列剂量向受试者施用:至少约1×10^{5}vg/kg到约1×10^{16}vg/kg、至少约1×10^{6}vg/kg到约1×10^{15}vg/kg、或至少约1×10^{7}vg/kg到约1×10^{14}vg/kg。在其它实施例中,所述方法包含向受试者施用治疗有效剂量的本文所描述的实施例的VLP,所述VLP包含CasX:gNA系统的组分和任选的供体模板。在一些实施例中,VLP以下列剂量向受试者施用:至少约1×10^{5}个粒子/kg、至少约1×10^{6}个粒子/kg、至少约1×10^{7}个粒子/kg至少约1×10^{8}个粒子/kg、至少约1×10^{9}个粒子/kg、至少约1×10^{10}个粒子/kg、至少约1×10^{11}个粒子/kg、至少约1×10^{12}个粒子/kg、至少约1×10^{13}个粒子/kg、至少约1×10^{14}个粒子/kg、至少约1×10^{15}个粒子/kg、或至少约1×10^{16}个粒子/kg。在一些实施例中,VLP以下列剂量向受试者施用:至少约1×10^{5}个粒子/kg到约1×10^{16}个粒子/kg、或至少约1×10^{6}个粒子/kg到约1×10^{15}个粒子/kg、或至少约1×10^{7}个粒子/kg到约1×10^{14}个粒子/kg。载体或VLP可以根据下文中所公开的任何治疗方案施用。

[0499] 在一些实施例中,向受试者施用本公开的C9orf72靶向载体组合物将CasX:gNA组合物递送到所述受试者的细胞,引起所述细胞中的C9orf72靶核酸的编辑。被治疗的受试者

的经修饰的细胞可以是选自由啮齿动物细胞、小鼠细胞、大鼠细胞、灵长类动物细胞、非人灵长类动物细胞和人类细胞组成的组的真核细胞。在一些实施例中，被治疗的受试者的真核细胞是人类细胞。在一些实施例中，所述细胞为选自由以下组成的组的细胞：普尔基涅细胞、额皮质神经元、运动皮质神经元、海马神经元、小脑神经元、上运动神经元、脊髓神经元、脊髓运动神经元、神经胶质细胞和星形胶质细胞。在一些实施例中，所述细胞包含细胞中的C9orf72基因的至少一个经修饰等位基因，其中所述修饰用于校正或补偿受试者中的C9orf72基因的一部分的突变或重复；例如HRS。在其它实施例中，所述细胞包含细胞中的C9orf72基因的至少一个经修饰等位基因，其中所述修饰用于敲减或敲除受试者中的C9orf72基因。

[0500] 在治疗方法的其它实施例中，所述方法包含进一步向受试者施用额外的CRISPR蛋白或编码额外的CRISPR蛋白的多核苷酸。在前述实施例中，额外的CRISPR蛋白具有不同于所述方法的第一CasX蛋白的序列。在一些实施例中，额外CRISPR蛋白质不是CasX蛋白；即，是Cpf1、Cas9、Cas10、Cas12a或Cas13a。在一些情况下，治疗方法中使用的gNA是单分子gNA(sgNA)。在其它情况下，gNA是双分子gNA(dgNA)。在又其它情况下，所述方法包含使靶核酸序列与靶向到C9orf72基因的不同或重叠序列的多种gNA接触。

[0501] 在一些实施例中，所述治疗方法包含通过选自由以下组成的组的施用途径向所述受试者施用CasX:gNA组合物或载体：皮下、皮内、神经内、结节内、髓内、肌肉内、腰髓内、鞘内、蛛网膜下、室内、囊内、静脉内、淋巴内或腹膜内途径，其中施用方法包含注射、输注或植入。在治疗受试者的C9orf72相关疾病的方法的一些实施例中，所述受试者选自由以下组成的组：小鼠、大鼠、猪、非人类灵长类动物和人类。在具体实施例中，受试者为人类。在一些实施例中，待通过本公开的方法修饰的受试者的细胞是选自由以下组成的组的细胞：普尔基涅细胞、额皮质神经元、运动皮质神经元、海马神经元、小脑神经元、上运动神经元、脊髓神经元、脊髓运动神经元、神经胶质细胞和星形胶质细胞。

[0502] 许多治疗策略已经用于设计用于治疗患有C9orf72相关疾病的受试者的方法的组合物。在一些实施例中，本发明提供一种治疗患有C9orf72相关疾病的受试者的方法，所述方法包含根据包含使用治疗有效剂量的一个或多个连续剂量的治疗方案，向所述受试者施用本文所公开的实施例中的任一个的CasX:gNA组合物或载体。在治疗方案的一些实施例中，组合物或载体的治疗有效剂量以单次剂量施用。在治疗方案的其它实施例中，治疗有效剂量以经至少两周、或至少一个月、或至少两个月、或至少三个月、或至少四个月、或至少五个月、或至少六个月的时段的两个或更多个剂量向受试者施用。在治疗方案的一些实施例中，有效剂量通过选自由以下组成的组的途径施用：皮下、皮内、神经内、结节内、髓内、肌肉内、腰髓内、鞘内、蛛网膜下、室内、囊内、静脉内、淋巴内或腹膜内途径，其中施用方法为注射、输注或植入。

[0503] 在一些实施例中，施用治疗有效量的CasX:gNA模态或包含编码本文所公开的CasX蛋白和引导核酸的多核苷酸的载体，以在患有C9orf72相关疾病的受试者中敲减或敲除C9orf72的表达，其中所述修饰使得预防或改善基础C9orf72相关疾病，使得在所述受试者中观察到改进，尽管所述受试者仍可能罹患基础疾病。在其它实施例中，向患有C9orf72相关疾病的受试者施用治疗有效量的CasX:gNA模态或包含编码本文所公开的CasX蛋白和引导核酸的多核苷酸的载体校正或补偿突变，以使得野生型或功能C9orf72蛋白质的表达使

得预防或改善基础C9orf72相关疾病，使得在所述受试者中观察到改进，尽管所述受试者仍可能罹患基础疾病。在一些实施例中，施用治疗有效量的CasX-gNA模态引起C9orf72相关疾病的至少一个临床上相关参数的改进，所述参数包括但不限于神经元细胞死亡、神经发炎、TDP-43相关病变、轴突和神经肌肉接头(NMJ)异常、前额皮质处的树突状脊密度变化、新生皮质神经元中的电生理缺陷、预测缓慢肺活量(SVC)百分比相对于基线的变化、肌力相对于基线的变化、延髓强度相对于基线的变化、ALS功能评级量表(ALSFRS-(R))、功能和存活的组合评估、反应持续时间、到死亡的时间、到气管切开术的时间、到持续辅助通气的时间(DTP)、用力肺活量(FVC%)；徒手肌力测试、最大自主等长收缩、反应持续时间、无发展存活期、到疾病进展的时间和到治疗失效的时间。在一些实施例中，施用治疗有效量的CasX-gNA模态引起C9orf72相关疾病治疗的至少两个临床上相关参数的改进。C9orf72相关疾病可以是FTD、ALS或这两者。在治疗方法的一些实施例中，受试者选自小鼠、大鼠、猪、犬、非人类灵长类动物和人类。

[0504] 在一些实施例中，所述治疗方法包含施用治疗有效剂量的经修饰以校正或补偿C9orf72基因的突变的细胞群体。用于修饰此类细胞群体的方法描述于上文中。通过治疗方法，经修饰的细胞的施用使得受试者中表达野生型或功能C9orf72蛋白质。在治疗方法的一些实施例中，总细胞的剂量在等于或约10^4到等于或约10^9个细胞/千克(kg)体重，如10^5到10^6个细胞/kg体重的范围内，例如等于或约1×10^5个细胞/kg、1.5×10^5个细胞/kg、2×10^5个细胞/kg或1×10^6个细胞/kg体重。例如，在一些实施例中，以等于或约10^4至或约10^9细胞/千克(kg)体重，如10^5和10^6细胞/kg体重之间，例如等于或约1×10^5细胞/kg、1.5×10^5细胞/kg、2×10^5细胞/kg，或1×10^6细胞/kg体重，或在某个误差范围内施用细胞。在一个实施例中，细胞相对于待施用细胞的受试者为自体的。在另一实施例中，细胞相对于待施用细胞的受试者为同种异体的。

[0505] 在一些实施例中，所述治疗方法进一步包含施用化学治疗剂，其中所述药剂可有效地改进与C9orf72相关疾病相关的病征或症状，包括但不限于利鲁唑(riluzole)、雷诺嗪(ranolazine)、拉蒂卡瓦(radicava)和右美沙芬(dextromethorphan)HBr与硫酸奎尼丁(quinidine sulfate)组合。

[0506] 从被治疗的受试者获得用于分析以确定治疗效果的样品(例如体液或组织)的方法，以及制备允许分析的样品的方法是所属领域技术人员所熟知的。分析RNA和蛋白质水平的方法已在上文讨论，并且为所属领域技术人员所熟知。还可以通过所属领域已知的常规临床方法，从与本发明的一种或多种化合物接触的动物收集的上述流体、组织或器官中测量与靶基因表达相关的生物标记物，从而评估治疗效果。C9orf72疾病的生物标记物包括但不限于C9orf72水平、C9orf72 RNA、含GGGGCC重复序列的RNA物质(以及反义GGCCCC RNA)、保留含六核苷酸重复序列的内含子的聚腺苷酸化C9orf72 RNA物质、DPR水平和DPR RNA水平。

[0507] 存在C9orf72六核苷酸重复序列扩增段的数种小鼠模型，并且适合于评价实施例的治疗方法(Batra R和Lee CW.《肌肉萎缩性侧索硬化/额颞叶型痴呆中C9orf72六核苷酸重复序列扩增段的小鼠模型》《细胞神经科学前沿》2017;11:196(2017))。

[0508] X.试剂盒及制品

[0509] 在其它实施例中，本文提供包含以下的试剂盒：CasX蛋白、本公开的任何实施例的

一种或多种包含对C9orf72基因具有特异性的靶向序列的CasX gNA，和适合容器(例如管、小瓶或板)。

[0510] 在一些实施例中，试剂盒进一步包含缓冲剂、核酸酶抑制剂、蛋白酶抑制剂、脂质体、治疗剂、标记、标记显色剂、或前述的任何组合。在一些实施例中，试剂盒进一步包含药学上可接受的载剂、稀释剂或赋形剂。

[0511] 在一些实施例中，试剂盒包含用于基因修饰应用的适当对照组合物，及使用说明书。

[0512] 在一些实施例中，所述试剂盒包含载体，所述载体包含编码本公开的CasX蛋白、本公开的CasX gNA、任选的供体模板或其组合的序列，并且所述试剂盒进一步包含药学上可接受的载剂、稀释剂或赋形剂。

[0513] 本说明书阐述大量示范性配置、方法、参数等。然而，应认识到，此类描述并不旨在作为本公开的范围的限制，而是替代地作为示范性实施例的描述而提供。

[0514] 说明性实施例

[0515] 可以参考以下所说明的所列举的实施例理解本发明：

[0516] 1.一种CasX:gNA系统，其包含CasX蛋白和引导核酸(gNA)，其中所述gNA包含与包含染色体9开放阅读框架72(C9orf72)基因的靶核酸序列互补的靶向序列。

[0517] 2.根据实施例1所述的CasX:gNA系统，其中所述C9orf72基因包含一个或多个突变。

[0518] 3.根据实施例1所述的CasX:gNA系统，其中所述C9orf72基因突变包含六核苷酸重复序列(HRS)GGGGCC的多于30、多于100、多于500、多于700、多于1000或多于1600个副本。

[0519] 4.根据实施例2或实施例3所述的CasX:gNA系统，其中所述突变为功能丧失突变。

[0520] 5.根据实施例2或实施例3所述的CasX:gNA系统，其中所述突变为功能增强突变。

[0521] 6.根据前述实施例中任一项所述的CasX:gNA系统，其中所述gNA为引导RNA(gRNA)。

[0522] 7.根据实施例1到5中任一项所述的CasX:gNA系统，其中所述gNA为引导DNA(gDNA)。

[0523] 8.根据实施例1到5中任一项所述的CasX:gNA系统，其中所述gNA为包含DNA和RNA的嵌合体。

[0524] 9.根据实施例1到8中任一项所述的CasX:gNA系统，其中所述gNA为单分子gNA(sgNA)。

[0525] 10.根据实施例1到8中任一项所述的CasX:gNA系统，其中所述gNA为双分子gNA(dgNA)。

[0526] 11.根据实施例1到10中任一项所述的CasX:gNA系统，其中所述gNA的所述靶向序列与包含所述C9orf72基因的一种或多种单核苷酸多态性(SNP)的序列互补。

[0527] 12.根据实施例1到10中任一项所述的CasX:gNA系统，其中所述gNA的所述靶向序列包含选自由表3中所阐述的序列组成的组的序列。

[0528] 13.根据实施例1到10中任一项所述的CasX:gNA系统，其中所述gNA的所述靶向序列包含表3的序列，其中从所述序列的3'端去除单个核苷酸。

[0529] 14.根据实施例1到10中任一项所述的CasX:gNA系统，其中所述gNA的所述靶向序

列包含表3的序列,其中从所述序列的3'端去除两个核苷酸。

[0530] 15.根据实施例1到10中任一项所述的CasX:gNA系统,其中所述gNA的所述靶向序列包含表3的序列,其中从所述序列的3'端去除三个核苷酸。

[0531] 16.根据实施例1到10中任一项所述的CasX:gNA系统,其中所述gNA的所述靶向序列包含表3的序列,其中从所述序列的3'端去除四个核苷酸。

[0532] 17.根据实施例1到10中任一项所述的CasX:gNA系统,其中所述gNA的所述靶向序列包含表3的序列,其中从所述序列的3'端去除五个核苷酸。

[0533] 18.根据实施例1到10中任一项所述的CasX:gNA系统,其中所述gNA的所述靶向序列包含与选自表3中所阐述的序列组成的组的序列具有至少约65%、至少约75%、至少约85%或至少约95%同一性的序列。

[0534] 19.根据实施例1到10中任一项所述的CasX:gNA系统,其中所述gNA的所述靶向序列包含相对于表3中所提供的序列具有一种或多种单核苷酸多态性(SNP)的序列。

[0535] 20.根据实施例1到19中任一项所述的CasX:gNA系统,其中所述gNA的所述靶向序列与所述C9orf72基因的非编码区互补。

[0536] 21.根据实施例1到19中任一项所述的CasX:gNA系统,其中所述gNA的所述靶向序列与所述C9orf72基因的编码区互补。

[0537] 22.根据实施例1到19中任一项所述的CasX:gNA系统,其中所述gNA的所述靶向序列与C9orf72外显子的序列互补。

[0538] 23.根据实施例1到19中任一项所述的CasX:gNA系统,其中所述gNA的所述靶向序列与C9orf72内含子的序列互补。

[0539] 24.根据实施例1到19中任一项所述的CasX:gNA系统,其中所述gNA的所述靶向序列与C9orf72内含子-外显子接界的序列互补。

[0540] 25.根据实施例1到19中任一项所述的CasX:gNA系统,其中所述gNA的所述靶向序列与C9orf72调节元件的序列互补。

[0541] 26.根据实施例1到19中任一项所述的CasX:gNA系统,其中所述gNA的所述靶向序列与所述C9orf72基因的基因间区的序列互补。

[0542] 27.根据实施例1到19中任一项所述的CasX:gNA系统,其中所述gNA的所述靶向序列与在所述HRS的5'处的序列互补。

[0543] 28.根据实施例27所述的CasX:gNA系统,其中所述gNA的所述靶向序列与所述C9orf72基因的内含子1或启动子的序列互补。

[0544] 29.根据实施例1到28中任一项所述的CasX:gNA系统,其进一步包含第二gNA,其中所述第二gNA具有与以下互补的靶向序列:所述靶核酸序列相较于根据前述实施例中任一项所述的gNA的所述靶向序列的不同或重叠部分。

[0545] 30.根据实施例28所述的CasX:gNA系统,其中所述第二gNA gNA的所述靶向序列与在所述HRS的5'或3'处的序列互补。

[0546] 31.根据实施例30所述的CasX:gNA系统,其中所述gNA的所述靶向序列与所述C9orf72基因的内含子1的序列互补。

[0547] 32.根据实施例1到31中任一项所述的CasX:gNA系统,其中所述gNA具有支架,所述支架包含选自由以下组成的组的序列:表1和表2中所阐述的序列,或与其具有至少约50%、

至少约60％、至少约70％、至少约80％、至少约90％、至少约95％、至少约95％、至少约96％、至少约97％、至少约98％或至少约99％序列同一性的序列。

[0548] 33.根据实施例1到31中任一项所述的CasX:gNA系统，其中所述gNA具有支架，所述支架包含相对于参考gNA序列具有至少一种修饰的序列，所述参考gNA序列选自由SEQ ID NO:4-16的序列组成的组。

[0549] 34.根据实施例33所述的CasX:gNA系统，其中所述参考gNA的所述至少一种修饰包含至少一个gNA序列的核苷酸的取代、缺失或插入。

[0550] 35.根据实施例1到34中任一项所述的CasX:gNA系统，其中所述gNA经化学修饰。

[0551] 36.根据实施例1到35中任一项所述的CasX:gNA系统，其中所述CasX蛋白包含具有SEQ ID NO:1-3中任一个的序列的参考CasX蛋白，具有表4的序列的CasX变异蛋白质，或与其具有至少约50％、至少约60％、至少约70％、至少约80％、至少约90％、至少约95％、至少约96％、至少约97％、至少约98％或至少约99％序列同一性的序列。

[0552] 37.根据实施例36所述的CasX:gNA系统，其中所述CasX变异蛋白质包含相对于具有选自SEQ ID NO:1-3的序列的参考CasX蛋白的至少一种修饰。

[0553] 38.根据实施例37所述的CasX:gNA系统，其中所述至少一种修饰包含在所述CasX变异蛋白质的结构域中的相对于所述参考CasX蛋白的至少一个氨基酸取代、缺失或插入。

[0554] 39.根据实施例38所述的CasX:gNA系统，其中所述结构域选自由非靶链结合(NTSB)结构域、靶链负载(TSL)结构域、螺旋形I结构域、螺旋形II结构域、寡核苷酸结合结构域(OBD)和RuvC DNA裂解结构域组成的组。

[0555] 40.根据实施例36到39中任一项所述的CasX:gNA系统，其中所述CasX蛋白进一步包含一个或多个核定位信号(NLS)。

[0556] 41.根据实施例40所述的CasX:gNA系统，其中所述一个或多个NLS选自由以下组成的序列的组：PKKKRKV(SEQ ID NO:165)、KRPAATKKAGQAKKKK(SEQ ID NO:166)、PAAKRVKLD(SEQ ID NO:167)、RQRRNELKRSP(SEQ ID NO:168)、NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY(SEQ ID NO:169)、RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV(SEQ ID NO:170)、VSRKRPRP(SEQ ID NO:171)、PPKKARED(SEQ ID NO:172)、PQPKKKPL(SEQ ID NO:173)、SALIKKKKKMAP(SEQ ID NO:174)、DRLRR(SEQ ID NO:175)、PKQKKRK(SEQ ID NO:176)、RKLKKKIKKL(SEQ ID NO:177)、REKKKFLKRR(SEQ ID NO:178)、KRKGDEVDGVDEVAKKKSKK(SEQ ID NO:179)、RKCLQAGMNLEARKTKK(SEQ ID NO:180)、PRPRKIPR(SEQ ID NO:181)、PPRKKRTVV(SEQ ID NO:182)、NLSKKKKRKREK(SEQ ID NO:183)、RRPSRPFRKP(SEQ ID NO:184)、KRPRSPSS(SEQ ID NO:185)、KRGINDRNFWRGENERKTR(SEQ ID NO:186)、PRPPKMARYDN(SEQ ID NO:187)、KRSFSKAF(SEQ ID NO:188)、KLKIKRPVK(SEQ ID NO:189)、PKTRRRPRRSQRKRPPT(SEQ ID NO:191)、RRKKRRPRRKKRR(SEQ ID NO:194)、PKKKSRKPKKKSRK(SEQ ID NO:195)、HKKKHPDASVNFSEFSK(SEQ ID NO:196)、QRPGPYDRPQRPGPYDRP(SEQ ID NO:197)、LSPSLSPLLSPSLSPL(SEQ ID NO:198)、RGKGGKGLGKGGAKRHRK(SEQ ID NO:199)、PKRGRGRPKRGRGR(SEQ ID NO:200)、MSRRRKANPTKLSENAKKLAKEVEN(SEQ ID NO:192)、PKKKRKVPPPPAAKRVKLD(SEQ ID NO:190)和PKKKRKVPPPPKKKRKV(SEQ ID NO:201)。

[0557] 42.根据实施例40或实施例41所述的CasX:gNA系统，其中所述一个或多个NLS在所述CasX蛋白的C端。

[0558] 43.根据实施例40或实施例41所述的CasX:gNA系统,其中所述一个或多个NLS在所述CasX蛋白的N端。

[0559] 44.根据实施例40或实施例41所述的CasX:gNA系统,其中所述一个或多个NLS在所述CasX蛋白的N端和C端。

[0560] 45.根据实施例36到44中任一项所述的CasX:gNA系统,其中所述CasX变异蛋白质和所述gNA相比于表1的参考CasX蛋白和gNA展现至少一种或更多种改进的特征。

[0561] 46.根据实施例45所述的CasX:gNA系统,其中所述改进的特征选自由以下组成的组:改进的所述CasX蛋白的折叠、所述CasX蛋白对所述gNA的改进的结合亲和力、改进的核糖核蛋白质复合物(RNP)形成、更高百分比的裂解潜能RNP、对所述靶核酸序列的改进的结合亲和力、对PAM序列的改进的结合亲和力、改进的所述靶核酸序列的解链、增加的活性、增加的靶核酸序列裂解速率、改进的编辑效率、改进的编辑特异性、增加的核酸酶活性、增加的用于双链裂解的靶链负载、减少的用于单链切割的靶链负载、减少的脱靶裂解、改进的DNA非靶链的结合、改进的CasX蛋白稳定性、改进的蛋白质:引导RNA复合物稳定性、改进的蛋白质溶解度、改进的蛋白质:gNA复合物溶解度、改进的蛋白质产率、改进的蛋白质表达和改进的融合特征。

[0562] 47.根据实施例45或实施例46所述的CasX:gNA系统,其中所述CasX变异蛋白质的所述改进的特征相对于SEQ ID NO:1、SEQ ID NO:2或SEQ ID NO:3的所述参考CasX蛋白改进至少约1.1到约100,000倍。

[0563] 48.根据实施例45或实施例46所述的CasX:gNA系统,其中所述CasX变异蛋白质的所述改进的特征相对于SEQ ID NO:1、SEQ ID NO:2或SEQ ID NO:3的所述参考CasX蛋白改进至少约10倍、至少约100倍、至少约1,000倍或至少约10,000倍。

[0564] 49.根据实施例46到48中任一项所述的CasX:gNA系统,其中所述改进的特征是改进的与所述靶核酸序列的结合亲和力。

[0565] 50.根据实施例46到48中任一项所述的CasX:gNA系统,其中所述改进的特征是增加的靶核酸序列裂解速率。

[0566] 51.根据实施例46到48中任一项所述的CasX:gNA系统,其中所述改进的特征是增加的与一个或多个PAM序列的结合亲和力,其中所述一个或多个PAM序列选自由TTC、ATC、GTC和CTC组成的组。

[0567] 52.根据实施例51所述的CasX:gNA系统,其中增加的对一个或多个PAM序列的结合亲和力是相比于SEQ ID NO:1-3的CasX蛋白中的任一个对于PAM序列的结合亲和力,大至少1.5倍。

[0568] 53.根据前述实施例中任一项所述的CasX:gNA系统,其中所述CasX变异蛋白质和所述gNA在RNP中结合在一起。

[0569] 54.根据实施例52所述的CasX:gNA系统,其中与表1的参考CasX和gNA的RNP相比,所述RNP具有至少5%、至少10%、至少15%或至少20%更高百分比的裂解潜能RNP。

[0570] 55.根据实施例39到54中任一项所述的CasX:gNA系统,其中所述CasX变异蛋白质包含具有切口酶活性的核酸酶结构域。

[0571] 56.根据实施例55所述的CasX:gNA系统,其中所述CasX变异体能够仅裂解双链靶核酸分子的一条链。

[0572] 57.根据实施例1到54中任一项所述的CasX:gNA系统，其中所述CasX变异蛋白质包含具有双链裂解活性的核酸酶结构域。

[0573] 58.根据实施例1到44中任一项所述的CasX:gNA系统，其中所述CasX蛋白为非催化活性CasX(dCasX)蛋白质，并且其中所述dCasX和所述gNA保持与所述靶核酸序列结合的能力。

[0574] 59.根据实施例58所述的CasX:gNA系统，其中所述dCasX包含以下残基处的突变：

[0575] a.对应于SEQ ID NO:1的所述参考CasX蛋白的D672、E769和/或D935；或

[0576] b.对应于SEQ ID NO:2的所述参考CasX蛋白的D659、E756和/或D922。

[0577] 60.根据实施例59所述的CasX:gNA系统，其中所述突变为丙氨酸取代所述残基。

[0578] 61.根据实施例1到57中任一项所述的CasX:gNA系统，其进一步包含供体模板核酸。

[0579] 62.根据实施例61所述的CasX:gNA系统，其中所述供体模板包含核酸，所述核酸包含所述C9orf72基因的至少一部分，其中所述C9orf72基因部分选自由以下组成的组：C9orf72外显子、C9orf72内含子、C9orf72内含子-外显子接界、C9orf72调节元件、或其组合。

[0580] 63.根据实施例61或实施例62所述的CasX:gNA系统，其中供体模板包含与侧接靶核酸中的裂解位点的序列互补的同源臂。

[0581] 64.根据实施例61到63所述的CasX:gNA系统，其中供体模板的大小在10到15,000个核苷酸的范围内。

[0582] 65.根据实施例61到64中任一项所述的CasX:gNA系统，其中所述供体模板为单链DNA模板或单链RNA模板。

[0583] 66.根据实施例61到64中任一项所述的CasX:gNA系统，其中所述供体模板为双链DNA模板。

[0584] 67.根据实施例61到66中任一项所述的CasX:gNA系统，其中所述供体模板包含与野生型C9orf72基因相比的一个或多个突变。

[0585] 68.根据实施例61到66中任一项所述的CasX:gNA系统，其中所述供体模板包含与野生型C9orf72基因相比的异源序列。

[0586] 69.根据实施例61到66中任一项所述的CasX:gNA系统，其中所述供体模板包含野生型C9orf72基因的全部或一部分。

[0587] 70.一种核酸，其包含编码根据实施例1到60中任一项所述的CasX:gNA系统的序列。

[0588] 71.根据实施例70所述的核酸，其中编码所述CasX蛋白和gNA的序列经密码子优化以在真核细胞中表达。

[0589] 72.一种载体，其包含根据实施例70或实施例71所述的核酸。

[0590] 73.根据实施例72所述的载体，其中所述载体进一步包含启动子。

[0591] 74.一种包含供体模板的载体，其中所述供体模板包含核酸，所述核酸包含C9orf72基因的至少一部分，其中所述C9orf72基因部分选自由以下组成的组：C9orf72外显子、C9orf72内含子、C9orf72内含子-外显子接界和C9orf72调节元件。

[0592] 75.根据实施例74所述的载体，其中所述供体模板包含与野生型C9orf72基因相比

的一个或多个突变,或包含与在C9orf72靶核酸中的裂解位点的5'和3'处的序列互补的两个同源臂侧接的异源序列。

[0593] 76.根据实施例74或实施例75所述的载体,其进一步包含根据实施例70或实施例71所述的核酸。

[0594] 77.根据实施例72到76中任一项所述的载体,其中所述载体选自由以下组成的组:逆转录病毒载体、慢病毒载体、腺病毒载体、腺相关病毒(AAV)载体、单纯疱疹病毒(HSV)载体、病毒样粒子(VLP)、质粒、小环、纳米质粒和RNA载体。

[0595] 78.根据实施例77所述的载体,其中所述载体为AAV载体。

[0596] 79.根据实施例78所述的载体,其中所述AAV载体选自AAV1、AAV2、AAV3、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAV10、AAV-Rh74或AAVRh10。

[0597] 80.根据实施例77所述的载体,其中所述载体为逆转录病毒载体。

[0598] 81.根据实施例77所述的载体,其中编码VLP的所述载体包含一种或多种编码gag多聚蛋白、根据实施例36到60中任一项所述的CasX蛋白和根据实施例1到35中任一项所述的gNA的核酸。

[0599] 82.一种病毒样粒子(VLP),其包含根据实施例36到60中任一项所述的CasX蛋白和根据实施例1到35中任一项所述的gNA。

[0600] 83.根据实施例82所述的VLP,其中所述CasX蛋白和所述gNA在RNP中结合在一起。

[0601] 84.根据实施例82或实施例83所述的VLP,其进一步包含提供所述VLP与靶细胞的结合和融合的假型化病毒包膜糖蛋白或抗体片段。

[0602] 85.一种修饰C9orf72靶核酸序列的方法,所述方法包含使所述靶核酸序列与CasX蛋白和包含靶向序列的引导核酸(gNA)接触,其中所述接触包含向细胞中引入:

[0603] a.根据实施例1到69中任一项所述的CasX:gNA系统;

[0604] b.根据实施例70或实施例71所述的核酸;

[0605] c.根据实施例72到81中任一项中的载体;

[0606] d.根据实施例82到84中任一项所述的VLP;或

[0607] e.其组合,

[0608] 其中所述接触使得所述C9orf72靶核酸序列由所述CasX蛋白修饰。

[0609] 86.根据实施例85所述的方法,其中所述CasX蛋白和所述gNA在核糖核蛋白质复合物(RNP)中结合在一起。

[0610] 87.根据实施例85或实施例86所述的方法,其进一步包含第二gNA或编码所述第二gNA的核酸,其中所述第二gNA具有与以下互补的靶向序列:与实施例85的引导序列相比所述靶核酸序列的不同部分或其互补序列。

[0611] 88.根据实施例85到87中任一项所述的方法,其中所述C9orf72基因包含突变。

[0612] 89.根据实施例88所述的方法,其中所述突变为功能增强突变。

[0613] 90.根据实施例88所述的方法,其中所述突变为功能丧失突变。

[0614] 91.根据实施例88所述的方法,其中所述C9orf72基因突变包含六核苷酸重复序列GGGGCC的多于30、多于100、多于500、多于700、多于1000或多于1600个副本。

[0615] 92.根据实施例85到90中任一项所述的方法,其中所述修饰包含在所述靶核酸序列中引入单链断裂。

[0616] 93.根据实施例85到90中任一项所述的方法,其中所述修饰包含在所述靶核酸序列中引入双链断裂。

[0617] 94.根据实施例85到93中任一项所述的方法,其中所述修饰包含引入所述靶核酸序列中的一个或多个核苷酸的插入、缺失、取代、重复或倒位。

[0618] 95.根据实施例85到94中任一项所述的方法,其中所述靶核酸序列的所述修饰在体外或离体发生。

[0619] 96.根据实施例85到95中任一项所述的方法,其中所述靶核酸序列的所述修饰在细胞内部发生。

[0620] 97.根据实施例85到95中任一项所述的方法,其中所述靶核酸序列的所述修饰在体内发生。

[0621] 98.根据实施例85到97中任一项所述的方法,其中所述细胞为真核细胞。

[0622] 99.根据实施例98所述的方法,其中所述真核细胞选自由以下组成的组:啮齿动物细胞、小鼠细胞、大鼠细胞、猪细胞、灵长类动物细胞和非人类灵长类动物细胞。

[0623] 100.根据实施例98所述的方法,其中所述真核细胞为人类细胞。

[0624] 101.根据实施例85到100中任一项所述的方法,其中所述细胞选自由以下组成的组:普尔基涅(Purkinje)细胞、额皮质神经元、运动皮质神经元、海马神经元、小脑神经元、上运动神经元、脊髓神经元、脊髓运动神经元、神经胶质细胞和星形胶质细胞。

[0625] 102.根据实施例85到101中任一项所述的方法,其中所述方法进一步包含使所述靶核酸序列与供体模板接触,所述供体模板包含与侧接根据实施例1到57中任一项所述的CasX:gNA系统所靶向的所述靶核酸中的裂解位点的序列互补的同源臂。

[0626] 103.根据实施例102所述的方法,其中所述供体模板包含相比于所述野生型C9orf72基因序列的一个或多个突变,并且其中所述插入引起所述C9orf72基因的敲减或敲除。

[0627] 104.根据实施例102所述的方法,其中插入所述供体模板置换所述C9orf72基因的所述HRS中的一些或全部。

[0628] 105.根据实施例102所述的方法,其中所述供体模板包含野生型C9orf72基因序列的全部或一部分,其中所述插入校正所述C9orf72基因的一个或多个突变。

[0629] 106.根据实施例102到104中任一项所述的方法,其中所述供体模板的大小在10到15,000个核苷酸的范围内。

[0630] 107.根据实施例102到104中任一项所述的方法,其中所述供体模板的大小在100到1,000个核苷酸的范围内。

[0631] 108.根据实施例102到107中任一项所述的方法,其中所述供体模板为单链DNA模板或单链RNA模板。

[0632] 109.根据实施例102到107中任一项所述的方法,其中所述供体模板为双链DNA模板。

[0633] 110.根据实施例102到109中任一项所述的方法,其中所述供体模板通过同源定向修复(HDR)插入。

[0634] 111.根据实施例85到110中任一项所述的方法,其中所述靶核酸已经修饰,使得HRS或DPR的表达相较于尚未修饰的靶核酸降低至少约10%、至少约20%、至少约30%、至少

约40％、至少约50％、至少约60％、至少约70％、至少约80％或至少约90％。

[0635] 112.根据实施例85到111中任一项所述的方法，其中所述载体以治疗有效剂量向受试者施用。

[0636] 113.根据实施例112所述的方法，其中所述受试者选自由以下组成的组：小鼠、大鼠、猪和非人类灵长类动物。

[0637] 114.根据实施例112所述的方法，其中所述受试者为人类。

[0638] 115.根据实施例85到114中任一项所述的方法，其中载体以下列剂量施用：至少约1×108载体基因组(vg)、至少约1×10^{9}vg、至少约1×10^{10}vg、至少约1×10^{11}vg、或至少约1×10^{12}vg、或至少约1×10^{13}vg、或至少约1×10^{14}vg、或至少约1×10^{15}vg、或至少约1×10^{16}vg。

[0639] 116.根据实施例111到115中任一项所述的方法，其中所述载体通过选自由以下组成的组的施用途径施用：皮下、皮内、神经内、结节内、髓内、肌肉内、腰髓内、鞘内、蛛网膜下、室内、囊内、静脉内、淋巴内或腹膜内途径，其中施用方法为注射、输注或植入。

[0640] 117.根据实施例85到116中任一项所述的方法，其包含进一步使所述靶核酸序列与额外CRISPR核酸酶或编码所述额外CRISPR核酸酶的多核苷酸接触。

[0641] 118.根据实施例117所述的方法，其中所述额外CRISPR核酸酶为具有不同于根据前述实施例中任一项所述的CasX蛋白的序列的CasX蛋白。

[0642] 119.根据实施例117所述的方法，其中所述额外CRISPR核酸酶不是CasX蛋白。

[0643] 120.一种改变细胞的C9orf72靶核酸序列的方法，其包含使所述细胞与以下接触：

[0644] a)根据实施例1到69中任一项所述的CasX:gNA系统；

[0645] b)根据实施例70或实施例71所述的核酸；

[0646] c)根据实施例72到81中任一项所述的载体；

[0647] d)根据实施例82到84中任一项所述的VLP;或

[0648] e)其组合，

[0649] 其中所述接触使得所述C9orf72靶核酸序列由所述CasX蛋白修饰。

[0650] 121.根据实施例120所述的方法，其中所述细胞已经修饰，使得所述HRS和/或所述DPR的表达相较于尚未修饰的细胞降低至少约10％、至少约20％、至少约30％、至少约40％、至少约50％、至少约60％、至少约70％、至少约80％或至少约90％。

[0651] 122.根据实施例120或实施例121所述的方法，其中所述细胞已经修饰，使得所述细胞不表达能够检测水平的二肽重复蛋白质(DPR)。

[0652] 123.一种通过根据实施例120或实施例121所述的方法修饰的细胞群体，其中所述细胞已经修饰，使得经修饰的细胞的至少10％、至少约20％、至少约30％、至少约40％、至少约50％、至少约60％、至少约70％、至少约80％或至少约90％不表达能够检测水平的DPR。

[0653] 124.根据实施例123所述的细胞群体，其中所述细胞为非灵长类哺乳动物细胞、非人类灵长类动物细胞或人类细胞。

[0654] 125.根据实施例123或实施例124所述的细胞群体，其中所述细胞选自由以下组成的组：普尔基涅细胞、额皮质神经元、运动皮质神经元、海马神经元、小脑神经元、上运动神经元、脊髓神经元、脊髓运动神经元、神经胶质细胞和星形胶质细胞。

[0655] 126.一种治疗有需要的受试者的C9orf72相关病症的方法，其包含修饰所述受试者的细胞中的C9orf72基因，所述修饰包含使所述细胞与以下接触：

[0656] a.根据实施例1到69中任一项所述的CasX:gNA系统；

[0657] b.根据实施例70或实施例71所述的核酸；

[0658] c.根据实施例72到81中任一项中的载体；

[0659] d.根据实施例82到84中任一项所述的VLP;或

[0660] e.其组合，

[0661] 其中所述接触使得所述C9orf72靶核酸序列由所述CasX蛋白修饰。

[0662] 127.根据实施例126所述的方法,其中所述C9orf72相关病症为肌肉萎缩性侧索硬化(ALS)或额颞叶型痴呆(FTD)。

[0663] 128.根据实施例126所述的CasX:gNA系统,其中所述gNA的所述靶向序列与在所述C9orf72基因的所述HRS的5'处的序列互补。

[0664] 129.根据实施例126到128中任一项所述的方法,其进一步包含第二gNA或编码所述第二gNA的核酸,其中所述第二gNA具有与以下互补的靶向序列:与根据实施例126所述的gNA相比所述靶核酸序列的不同或重叠部分。

[0665] 130.根据实施例129所述的CasX:gNA系统,其中所述第二gNA的所述靶向序列与在所述C9orf72基因的内含子1中并且在所述HRS的3'处的序列互补。

[0666] 131.根据实施例126到130中任一项所述的方法,其中所述修饰引入所述C9orf72基因中的一个或多个突变,或其中所述HRS和/或所述DPR的表达相较于尚未修饰的细胞降低至少约10％、至少约20％、至少约30％、至少约40％、至少约50％、至少约60％、至少约70％、至少约80％或至少约90％。

[0667] 132.根据实施例126到130中任一项所述的方法,其中所述方法包含使所述细胞与根据实施例61到69中任一项所述的供体模板接触。

[0668] 133.根据实施例126到132中任一项所述的方法,其中所述细胞选自由以下组成的组:普尔基涅细胞、额皮质神经元、运动皮质神经元、海马神经元、小脑神经元、上运动神经元、脊髓神经元、脊髓运动神经元、神经胶质细胞和星形胶质细胞。

[0669] 134.根据实施例126到133中任一项所述的方法,其中所述受试者选自由以下组成的组:小鼠、大鼠、猪、非人类灵长类动物和人类。

[0670] 135.根据实施例134所述的方法,其中所述受试者为人类。

[0671] 136.根据实施例126到135中任一项所述的方法,其中所述载体以治疗有效剂量向受试者施用。

[0672] 137.根据实施例126到136中任一项所述的方法,其中所述载体以下列剂量施用到所述受试者:至少约1×10^{10}载体基因组(vg)、或至少约1×10^{11}vg、或至少约1×10^{12}vg、或至少约1×10^{13}vg、或至少约1×10^{14}vg、或至少约1×10^{15}vg、或至少约1×10^{16}vg。

[0673] 138.根据实施例126到136中任一项所述的方法,其中所述载体通过选自由以下组成的组的施用途径施用:皮下、皮内、神经内、结节内、髓内、肌肉内、腰髓内、鞘内、蛛网膜下、室内、囊内、静脉内、淋巴内或腹膜内途径,其中施用方法为注射、输注或植入。

[0674] 139.根据实施例126到138中任一项所述的方法,其包含进一步使所述靶核酸序列与额外CRISPR核酸酶或编码额外CRISPR蛋白质的多核苷酸接触。

[0675] 140.根据实施例139所述的方法,其中所述额外CRISPR核酸酶为具有不同于根据前述实施例中任一项所述的CasX的序列的CasX蛋白。

[0676] 141.根据实施例140所述的方法,其中所述额外CRISPR核酸酶不是CasX蛋白。

[0677] 142.根据实施例126到141中任一项所述的方法,其中所述方法进一步包含施用化学治疗剂。

[0678] 143.根据实施例126到142中任一项所述的方法,其中所述方法引起至少一种选自由以下组成的组的临床上相关参数的改进:神经元细胞死亡、神经发炎、TDP-43相关病变、轴突和神经肌肉接头(NMJ)异常、前额皮质处的树突状脊密度变化、新生皮质神经元中的电生理缺陷、预测缓慢肺活量(SVC)百分比相对于基线的变化、肌力相对于基线的变化、延髓强度相对于基线的变化、ALS功能评级量表(ALSFRS-(R))、功能和存活的组合评估、反应持续时间、到死亡的时间、到气管切开术的时间、到持续辅助通气的时间(DTP)、用力肺活量(FVC%);徒手肌力测试、最大自主等长收缩、反应持续时间、无发展存活期、到疾病进展的时间和到治疗失效的时间。

[0679] 144.根据实施例126到142中任一项所述的方法,其中所述方法引起至少两种选自由以下组成的组的临床上相关参数的改进:神经元细胞死亡、神经发炎、TDP-43相关病变、轴突和神经肌肉接头(NMJ)异常、前额皮质处的树突状脊密度变化、新生皮质神经元中的电生理缺陷、预测缓慢肺活量(SVC)百分比相对于基线的变化、肌力相对于基线的变化、延髓强度相对于基线的变化、ALS功能评级量表(ALSFRS-(R))、功能和存活的组合评估、反应持续时间、到死亡的时间、到气管切开术的时间、到持续辅助通气的时间(DTP)、用力肺活量(FVC%);徒手肌力测试、最大自主等长收缩、反应持续时间、无发展存活期、到疾病进展的时间和到治疗失效的时间。

[0680] 实例

[0681] 实例1:CasX Stx2的产生、表达及纯化

[0682] 1.生长和表达

[0683] 源自浮霉菌门的CasX Stx2(本文也称为CasX2)的表达构建体(具有SEQ ID NO:2的CasX氨基酸序列并由下表5中的序列编码)由针对大肠杆菌密码子优化的基因片段(Twist Biosciences)构建。组装的构建体含有TEV-可裂解、C端、TwinStrep标签,且克隆至含有氨苄青霉素抗性基因的pBR322衍生的质粒主链中。将表达构建体转化至化学感受态BL21*(DE3)大肠杆菌中,且起子培养物在37℃、200RPM下在UltraYield烧瓶(Thomson Instrument Company)中于补充有羧苄青霉素的LB培养液中生长过夜。第二天,此培养物以1:100比率(起子培养物:表达培养物)用于种子表达培养物。将表达培养物接种至补充有羧苄青霉素的Terrific Broth(Novagen)中且在37℃、200RPM下于UltraYield烧瓶中生长。一旦培养物达到2的光密度(OD),便将其冷却至16℃,且从1M原料中添加IPTG(异丙基β-D-1-硫代半乳糖苷)至1mM的最终浓度。培养物在16℃、200RPM下诱导20小时,随后通过在4℃下以4,000xg离心15分钟收获。将细胞浆料称重且以每克细胞浆料5mL溶解缓冲液的比率再悬浮于溶解缓冲液(50mM HEPES-NaOH,250mM NaCl,5mM $MgCl_2$,1mM TCEP,1mM苯甲脒-HCL,1mM PMSF,0.5% CHAPS,10%甘油,pH 8)中。一旦再悬浮,便将样品冷冻于-80℃直至纯化。

[0684] 表5:CasX Stx2构建体的DNA序列

[0685]

构建体	DNA 序列
SV40 NLS-CasX-SV40 NLS-TEV 裂解位点 – TwinStrep 标签	ATGGCTCCGAAGAAGAAGCGAAAGGTCAGCCAGGAAATTAAACGCATCAAC AAGATCCGCCGTCGTCTGGTAAAAGACAGCAATACGAAAAAAGCCGGAAAA ACCGGTCCGATGAAAACGCTGCTGGTGCGCGTGATGACGCCGGATCTCCGC GAACGTCTTGAGAATTTGCGTAAGAAACCTGAAAATATTCCGCAACCGATT TCTAACACCTCGCGCGCCAATCTGAATAAACTGCTGACCGATTACACCGAA ATGAAGAAAGCGATTCTGCACGTTTACTGGGAAGAGTTCCAGAAAGACCCG GTCGGTCTGATGAGCCGCGTTGCGCAACCTGCGCCGAAAAATATCGATCAG CGCAAGTTAATCCCGGTTAAAGATGGTAATGAACGTTTAACCTCCAGCGGC TTTGCCTGCAGTCAGTGCTGCCAGCCACTTTATGTTTATAAACTTGAACAG GTTAACGATAAAGGGAAACCCCATACCAATTATTTCGGCCGCTGCAATGTC AGCGAACATGAACGCCTGATTTTGTTAAGCCCGCATAAACCGGAAGCGAAT GACGAACTGGTGACCTATTCCCTGGGTAAATTTGGTCAGCGGGCGCTGGAT TTTTACAGCATTCATGTGACGCGGGAAAGTAACCATCCGGTAAAGCCACTG GAACAAATCGGCGGTAACAGCTGCGCCTCTGGCCCGGTTGGCAAAGCGCTT AGCGATGCCTGTATGGGCGCGGTGGCGAGCTTTCTGACAAAATACCAGGAT ATTATCCTGGAGCATCAGAAGGTGATCAAAAAGAACGAGAAACGTCTGGCA AATTTAAAGGATATTGCCTCCGCTAACGGCCTGGCGTTCCCGAAGATTACC TTACCGCCGCAGCCGCACACCAAAGAAGGTATCGAAGCGTATAACAACGTT GTTGCCCAGATCGTCATCTGGGTGAATCTCAACCTGTGGCAAAAACTGAAA ATTGGTCGTGATGAAGCAAAACCGTTGCAGCGACTGAAAGGATTCCCGTCG TTTCCGCTGGTTGAACGACAGGCGAACGAAGTGGATTGGTGGGATATGGTT TGTAACGTCAAAAAATTGATCAACGAAAAAAAGGAAGATGGCAAAGTTTTC TGGCAAAATCTGGCGGGTTACAAACGTCAGGAGGCGTTGCTTCCGTATCTC TCTTCAGAAGAAGATCGCAAAAAAGGCAAGAAGTTTGCTCGCTATCAGTTT GGCGATTTATTACTGCATCTGGAAAAAAAACACGGCGAAGACTGGGGCAAA GTGTACGATGAAGCCTGGGAGCGTATCGACAAAAAAGTGGAAGGTTTGTCG AAACATATTAAACTCGAAGAAGAGCGCCGCAGTGAAGATGCGCAGTCAAAA

[0686]

	GCAGCGCTGACGGACTGGTTACGTGCGAAAGCCAGTTTTGTGATTGAAGGA TTAAAAGAAGCTGATAAAGATGAATTTTGCCGTTGCGAACTGAAACTGCAA AAATGGTATGGCGACCTGCGCGGCAAACCGTTCGCCATTGAGGCAGAAAAT AGCATCCTTGATATCTCCGGTTTCAGCAAACAATATAACTGCGCGTTTATT TGGCAGAAAGACGGCGTGAAAAAGCTTAACCTGTATCTGATCATTAACTAT TTTAAAGGCGGGAAACTGCGTTTCAAGAAAATCAAGCCGGAAGCATTTGAA GCCAATCGTTTTTATACCGTTATTAATAAAAAAAGCGGTGAAATCGTGCCG ATGGAAGTTAATTTTAACTTTGATGATCCGAACTTGATTATTCTGCCGCTG GCATTCGGTAAACGGCAGGGCCGTGAGTTTATCTGGAACGACCTGTTATCG CTGGAAACGGGCAGCCTGAAATTAGCCAACGGTCGCGTCATTGAAAAAACG CTCTACAACCGCCGCACCCGCCAGGATGAGCCGGCACTGTTTGTCGCGCTG ACCTTTGAACGGCGTGAAGTCCTCGATAGCAGCAACATCAAACCAATGAAC CTTATCGGTATTGATCGTGGTGAAAACATTCCTGCCGTTATCGCCCTGACT GATCCAGAAGGCTGCCCGCTTTCTCGCTTCAAAGATTCACTGGGCAACCCG ACCCATATCCTCCGTATTGGCGAGAGCTACAAAGAGAAACAGCGTACCATT CAGGCAGCCAAAGAAGTGGAGCAGCGTCGCGCGGGCGGCTATAGCCGTAAA TATGCCAGCAAAGCTAAAAACCTGGCGGATGACATGGTGCGTAACACGGCG CGCGATTTGCTGTACTACGCCGTCACCCAGGACGCGATGCTGATTTTTGAG AACCTCTCCCGCGGTTTTGGGCGTCAGGGTAAACGCACGTTTATGGCGGAA CGCCAGTATACGCGTATGGAGGACTGGCTGACCGCGAAGCTGGCCTATGAA GGCTTGCCGTCTAAAACTTACCTGAGCAAGACCCTGGCTCAGTACACCAGT AAAACCTGTAGTAATTGCGGCTTTACCATCACCAGCGCCGATTATGACCGC GTGCTGGAAAAGCTGAAGAAAACCGCCACCGGCTGGATGACCACCATCAAT GGTAAAGAGCTTAAAGTCGAAGGGCAGATTACTTATTACAACCGTTATAAG CGGCAAAACGTGGTGAAAGATCTGTCGGTTGAGCTGGACCGTTTGTCTGAA GAAAGCGTGAACAATGATATCAGCTCCTGGACCAAAGGTCGTTCCGGCGAA GCGTTAAGTCTGTTGAAAAAGCGCTTTAGCCATCGCCCGGTGCAGGAAAAA TTCGTTTGCCTGAACTGTGGCTTCGAAACCCACGCCGACGAGCAAGCGGCG CTCAATATTGCGCGTAGCTGGCTGTTCCTGCGCAGCCAGGAATATAAAAAA TATCAAACCAACAAAACAACTGGCAATACCGACAAGCGTGCCTTTGTTGAA ACCTGGCAGAGCTTCTATCGCAAAAAACTGAAAGAGGTCTGGAAACCGGCG GTAGCGCCAAAGAAAAAACGCAAAGTGAGCGAAAATCTTTATTTTCAAGGT AGCGCATGGAGTCATCCTCAATTCGAGAAAGGTGGAGGTTCTGGCGGTGGA TCGGGAGGTTCAGCGTGGAGCCACCCGCAGTTCGAAAAAGGAAGGGGATCC GGCTGCTAA（SEQ ID NO: 229）

[0687] 2.纯化

[0688] 冷冻样品在4℃下在磁力搅拌下解冻过夜。通过超声处理降低所得裂解物的黏度，且通过使用Emulsiflex C3(Avestin)在17k PSI下分三次均质化来完成裂解。通过在4℃下以50,000x g离心30分钟来澄清裂解物且收集上清液。通过重力流将澄清的上清液上样到Heparin 6Fast Flow柱(GE Life Sciences)。用5CV肝素缓冲液A(50mM HEPES-NaOH,250mM NaCl,5mM $MgCl_2$,1mM TCEP,10％甘油,pH 8)洗涤柱子，接着用5CV肝素缓冲液B(NaCl浓度调节至500mM的缓冲液A)洗涤。用5CV的肝素缓冲液C(将NaCl浓度调整到1M的缓冲液A)洗脱蛋白质，收集级分。通过Bradford Assay分析各级分中的蛋白质，并将含蛋白质的级分汇集起来。通过重力流将合并的肝素洗脱液应用于Strep-Tactin XT Superflow柱(IBA Life Sciences)。用5CV的Strep缓冲液(50mM HEPES-NaOH,500mM NaCl,5mM $MgCl_2$,1mM TCEP,10％甘油,pH 8)洗涤柱子。使用添加50mM D-生物素的5CV Strep缓冲液自柱子洗脱蛋白质且收集级分。将含有CasX的级分合并，使用30kDa截止自旋浓缩器在4℃下浓缩，并在Superdex 200pg柱(GE Life Sciences)上通过尺寸排阻色谱法纯化。柱子用SEC缓冲液(25mM磷酸钠,300mM NaCl,1mM TCEP,10％甘油,pH 7.25)平衡，所述缓冲液由AKTA纯FPLC系统(GE Life Sciences)操作。在适当分子量下洗脱的含CasX的级分经汇集，使用30kDa截止旋转浓缩器在4℃下浓缩，等分，且在液氮中急冻，随后存储于-80℃下。

[0689] 3.结果

[0690] 来自整个纯化程序的样品通过SDS-PAGE解析且通过胶体考马斯染色观测,如图1和图3中所示。在图1中,自左至右,通道为:分子量标准,集结粒:细胞裂解之后的不溶性部分,裂解物:细胞裂解之后的可溶性部分,流经:不结合肝素柱的蛋白质,洗涤:洗涤缓冲液中自柱洗脱的蛋白质,洗脱:自肝素柱洗脱的蛋白质伴以洗脱缓冲液,流经:不结合StrepTactinXT柱的蛋白质,洗脱:自StrepTactin XT柱洗脱的蛋白质伴以洗脱缓冲液,注射:注射至s200凝胶过滤柱上的浓缩蛋白质,冷冻:已浓缩及冷冻的来自s200洗脱的汇集级分。在图3中,从右到左的通道是注射(将蛋白质样品注入凝胶过滤柱)分子量标记物,通道3-9是指示洗脱体积的样品。来自凝胶过滤的结果如图2所示。68.36mL峰对应于CasX的表观分子量且含有大多数CasX蛋白。通过胶体考马斯染色评估,每升培养物的平均产量为0.75mg纯化CasX蛋白,纯度为75%。

[0691] 实例2:CasX构建体119、438和457

[0692] 为了生成CasX 119、438及457构建体(表6中的序列),经密码子优化的CasX 37构建体(基于实例1的CasX Stx2构建体,编码浮霉菌门CasX SEQ ID NO:2,具有经融合NLS进行的A708K取代及[P793]缺失,及连接的引导序列及非靶向序列)使用标准克隆方法克隆至哺乳动物表达质粒(pStX;参见图4)中。为了建立CasX 119,在两个反应中使用Q5 DNA聚合酶(新英格兰生物实验室目录号M0491L),根据制造商的方案,分别使用引物oIC539及oIC88以及oIC87及oIC540对CasX 37构建体DNA进行PCR扩增(参见图5)。为了构建CasX 457,在四个反应中使用Q5 DNA聚合酶并分别使用引物oIC539和oIC212、oIC211和oIC376、oIC375和oIC551以及oIC550和oIC540对CasX 365构建体DNA进行PCR扩增。为了构建CasX 438,在四个反应中使用Q5 DNA聚合酶并分别使用引物oIC539和oIC689、oIC688和oIC376、oIC375和oIC551以及oIC550和oIC540对CasX 119构建体DNA进行PCR扩增。所得PCR扩增产物接着使用Zymoclean DNA清洁及浓缩器(Zymo Research目录号4014),根据制造商的方案进行纯化。使用XbaI及SpeI消化pStX主链以去除质粒pStx34中的两个位点之间的DNA的2931个碱基对的片段。消化的主链片段係通过使用Zymoclean凝胶DNA回收试剂盒(Zymoclean Gel DNA Recovery Kit)(Zymo Research目录号D4002),根据制造商的方案自1%琼脂糖凝胶(Gold Bio目录号A-201-500)进行凝胶提取而纯化。然后,按照制造商的协议,使用Gibson组装技术(新英格兰生物实验室Cat#E2621S)将三个片段拼接在一起。将pStx34中的装配产物转化至化学感受态或电感受态Turbo感受态大肠杆菌细菌细胞,接种于含有羧苄青霉素的LB-琼脂板(LB:Teknova目录号L9315,琼脂:Quartzy目录号214510)上。选取个别菌落且使用Qiagen Qiaprep spin Miniprep Kit(Qiagen目录号27104)遵循制造商的方案进行小规模纯化。所得质粒是使用桑格测序法测序以确保适当组装。pStX34包括用于蛋白质的EF-1α启动子以及用于嘌呤霉素及羧苄青霉素两者的选择标记物。编码靶向所关注基因的靶向序列的序列係基于CasX PAM位置而设计。靶向序列DNA以由靶向序列及所述序列的反向互补序列组成的单链DNA(ssDNA)寡核苷酸(Integrated DNA Technologies)形式订购。将这两个寡核苷酸退火在一起,并使用T4 DNA连接酶(新英格兰生物实验室目录号M0202L)和合适的质粒限制酶,通过Golden Gate组装技术,单独或批量克隆到pStX中。将Golden Gate产物转化为化学或电感受态细胞,如NEB Turbo感受态大肠杆菌(NEB目录号C2984I)中,将其涂于含有羧苄青霉素的LB-琼脂板上。选取个别菌落且使用Qiagen Qiaprep spin

Miniprep Kit(Qiagen目录号27104)且遵循制造商的方案进行小规模纯化。使用桑格测序法对所得质粒进行测序,以确保正确的连接。SaCas9及SpyCas9对照质粒係与上文所述的pStX质粒类似地制备,其中pStX的蛋白质及引导区交换对应蛋白质及引导序列。SaCas9及SpyCas9的靶向序列获自文献或根据确立方法合理地设计。CasX 119和457蛋白的表达和恢复使用实例1的一般方法进行(但是DNA序列为在大肠杆菌中表达进行了密码子优化)。CasX 119的分析测定结果如在图6至8所示。通过胶体考马斯染色评估,在纯度为75%的条件下,每升培养物中的CasX 119的平均产量为1.56mg纯化的CasX蛋白。图6示出了纯化样品的SDS-PAGE凝胶,在Bio-Rad Stain-Free™凝胶上可视化,如上所述。自左至右,通道为:集结粒:细胞裂解之后的不溶性部分,裂解物:细胞裂解之后的可溶性部分,流经:不结合肝素柱的蛋白质,洗涤:洗涤缓冲液中自柱洗脱的蛋白质,洗脱:自肝素柱洗脱的蛋白质伴以洗脱缓冲液,流经:不结合StrepTactinXT柱的蛋白质,洗脱:自StrepTactin XT柱洗脱的蛋白质伴以洗脱缓冲液,注射:注射至s200凝胶过滤柱上的浓缩蛋白质,冷冻:已浓缩及冷冻的来自s200洗脱的汇集级分。

[0693] 图7示出了Superdex 200 16/600pg凝胶过滤的色谱图,如所述。CasX变异体119蛋白的凝胶过滤运行绘制为280nm吸光度与洗脱体积的关系。65.77mL峰对应于CasX变异体119的表观分子量且含有大多数CasX变异体119蛋白。图8示出了凝胶过滤样品的SDS-PAGE凝胶,如所述,用胶体考马斯染色。来自指定级分的样品通过SDS-PAGE解析且通过胶体考马斯染色。自右向左,注射:注射至凝胶过滤柱上的蛋白质样品、分子量标记物、通道3-10:来自指定洗脱体积的样品。

[0694] 表6:CasX 119、438和457的序列

[0695]

构建体	DNA 序列	氨基酸序列
CasX 119	（SEQ ID NO：230）	QEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRK KPENIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSR VAQPAPKNIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDK GKPHTNYFGRCNVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFY SIHVTRESNHPVKPLEQIGGNSCASGPVGKALSDACMGAVASFLTKYQD IILEHQKVIKKNEKRLANLKDIASANGLAFPKITLPPQPHTKEGIEAYN NVVAQIVIWVNLNLWQKLKIGRDEAKPLQRLKGFPSFPLVERQANEVDW WDMVCNVKKLINEKKEDGKVFWQNLAGYKRQEALRPYLSSEEDRKKGKK FARYQFGDLLLHLEKKHGEDWGKVYDEAWERIDKKVEGLSKHIKLEEER RSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFCRCELKLQKWYGDLR GKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLYLIINYFKGGK LRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLIILPLAFG KRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVALT FERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGN PTHILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVR NTARDLLYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTA KLAYEGLSKTYLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATG WMTTINGKELKVEGQITYYNRYKRQNVVKDLSVELDRLSEESVNNDISS WTKGRSGEALSLLKKRFSHRPVQEKFVCLNCGFETHADEQAALNIARSW LFLRSQEYKKYQTNKTTGNTDKRAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO：233）
CasX 457	（SEQ ID NO：231）	QEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRK KPENIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSR VAQPAPKNIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDK GKPHTNYFGRCNVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFY SIHVTRESNHPVKPLEQIGGNSCASGPVGKALSDACMGAVASFLTKYQD IILEHKKVIKKNEKRLANLKDIASANGLAFPKITLPPQPHTKEGIEAYN NVVAQIVIWVNLNLWQKLKIGRDEAKPLQRLKGFPSFPLVERQANEVDW WDMVCNVKKLINEKKEDGKVFWQNLAGYKRQEALRPYLSSPEDRKKGKK FARYQLGDLLLHLEKKHGEDWGKVYDEAWERIDKKVEGLSKHIKLEEER RSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFCRCELKLQKWYGDLR GKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLYLIINYFKGGK

[0696]

		LRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLIILPLAFG KRQGREFIWNDLLSLETGSLKLANGRVIEKPLYNRRTRQDEPALFVALT FERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGN PTHILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVR NTARDLLYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTA KLAYEGLSKTYLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATG WMTTINGKELKVEGQITYYNRRKRQNVVKDLSVELDRLSEESVNNDISS WTKGRSGEALSLLKKRFSHRPVQEKFVCLNCGFETHADEQAALNIARSW LFLRSQEYKKYQTNKTTGNTDKRAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO: 234）
CasX 438	（SEQ ID NO: 232）	QEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRK KPENIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSR VAQPAPKNIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDK GKPHTNYFGRCNVSEHERLILLSPHKPEANDELVTYSLGKFGQRALDFY SIHVTRESNHPVKPLEQIGGNSCASGPVGKALSDACMGAVASFLTKYQD IILEHQKVIKKNEKRLANLKDIASANGLAFPKITLPPQPHTKEGIEAYN NVVAQIVIWVNLNLWQKLKIGRDEAKPLQRLKGFPSFPLVERQANEVDW WDMVCNVKKLINEKKEDGKVFWQNLAGYKRQEALRPYLSSEEDRKKGKK FARYQLGDLLKHLEKKHGEDWGKVYDEAWERIDKKVEGLSKHIKLEEER RSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFCRCELKLQKWYGDLR GKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLNLYLIINYFKGGK LRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPNLIILPLAFG KRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVALT FERREVLDSSNIKPMNLIGVDRGENIPAVIALTDPEGCPLSRFKDSLGN PTHILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVR NTARDLLYYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTA KLAYEGLSKTYLSKTLAQYTSKTCSNCGFTITSADYDRVLEKLKKTATG WMTTINGKELKVEGQITYYNRRKRQNVVKDLSVELDRLSEESVNNDISS WTKGRSGEALSLLKKRFSHRPVQEKFVCLNCGFETHADEQAALNIARSW LFLRSQEYKKYQTNKTTGNTDKRAFVETWQSFYRKKLKEVWKPAV（SEQ ID NO: 235）

[0697] 实例3:CasX构建体488和491

[0698] 为了生成CasX 488构建体(表7中的序列),经密码子优化的CasX 119构建体(基于实例1的CasX Stx2构建体,编码浮霉菌门CasX SEQ ID NO:2,具有经融合NLS进行的A708K取代、L379R取代及[P793]缺失,及连接的引导序列及非靶向序列)使用标准克隆方法克隆至哺乳动物表达质粒(pStX;参见图4)中。构建体CasX 1(基于实例1的CasX Stx1构建体,编码CasX SEQ ID NO:1)使用标准克隆方法克隆到目的载体中。为了建立CasX 488,使用Q5 DNA聚合酶,使用引物oIC765和oIC762对CasX 119构建体DNA进行PCR扩增(见图5)。使用Q5 DNA聚合酶,使用引物oIC766和oIC784对CasX 1构建体进行PCR扩增。通过使用Zymoclean凝胶DNA回收试剂盒,自1％琼脂糖凝胶进行凝胶提取来纯化PCR产物。随后使用吉布森组装(Gibson assembly)将两个片段拼接在一起。将pStx1中的组装产物转化到化学感受态的Turbo感受态大肠杆菌细菌细胞中,涂铺于含有卡那霉素(kanamycin)的LB-琼脂板上。挑取个别菌落且使用Qiagen Qiaprep旋转小规模制备试剂盒进行小规模制备。使用桑格测序法对所得质粒进行测序,以确保正确的组装。然后使用限制酶克隆将正确的克隆亚克隆到哺乳动物表达载体pStx34中。pStx1中的pStx34主链和CasX 488克隆分别用XbaI和BamHI消化。通过使用Zymoclean凝胶DNA回收试剂盒,从1％琼脂糖凝胶进行凝胶提取来纯化消化的主链和插入片段。然后根据制造商的方案使用T4连接酶(新英格兰生物实验室目录号M0202L)将干净的主链和插入物连接在一起。将连接的产物转化到化学感受态的Turbo感受态大肠杆菌细菌细胞中,涂铺于含有羧苄青霉素的LB-琼脂板上。挑取个别菌落且使用

Qiagen Qiaprep旋转小规模制备试剂盒进行小规模制备。使用桑格测序法对所得质粒进行测序,以确保正确的组装。

[0699] 为了产生CasX 491(表7中的序列),使用Q5 DNA聚合酶,使用引物oIC765和oIC762对CasX 484构建体DNA进行PCR扩增(见图5)。使用Q5 DNA聚合酶,使用引物oIC766和oIC784对CasX 1构建体进行PCR扩增。通过使用Zymoclean凝胶DNA回收试剂盒,自1%琼脂糖凝胶进行凝胶提取来纯化PCR产物。随后使用吉布森组装将两个片段拼接在一起。将pStx1中的组装产物转化到化学感受态的Turbo感受态大肠杆菌细菌细胞中,涂铺于含有卡那霉素的LB-琼脂板上。挑取个别菌落且使用Qiagen Qiaprep旋转小规模制备试剂盒进行小规模制备。使用桑格测序法对所得质粒进行测序,以确保正确的组装。然后使用限制酶克隆将正确的克隆亚克隆到哺乳动物表达载体pStx34中。pStx1中的pStx34主链和CasX 491克隆分别用XbaI和BamHI消化。通过使用Zymoclean凝胶DNA回收试剂盒,从1%琼脂糖凝胶进行凝胶提取来纯化消化的主链和插入片段。然后根据制造商的方案使用T4连接酶(新英格兰生物实验室目录号M0202L)将干净的主链和插入物连接在一起。将连接的产物转化到化学感受态的Turbo感受态大肠杆菌细菌细胞中,涂铺于含有羧苄青霉素的LB-琼脂板上。挑取个别菌落且使用Qiagen Qiaprep旋转小规模制备试剂盒进行小规模制备。所得质粒是使用桑格测序法测序以确保适当组装。pStX34包括用于蛋白质的EF-1α启动子以及用于嘌呤霉素及羧苄青霉素两者的选择标记物。编码靶向所关注基因的靶向序列的序列係基于CasX PAM位置而设计。靶向序列DNA以由靶向序列及所述序列的反向互补序列组成的单链DNA(ssDNA)寡核苷酸(Integrated DNA Technologies)形式订购。将这两个寡核苷酸退火在一起,并使用T4 DNA连接酶和适当的质粒限制酶通过Golden Gate组装单独或批量克隆到pStX中。将Golden Gate产物转化到化学或电感受态细胞,如NEB Turbo感受态大肠杆菌中,涂铺于含有羧苄青霉素的LB-琼脂板上。挑取个别菌落且使用Qiagen Qiaprep旋转小规模制备试剂盒进行小规模制备。使用桑格测序法对所得质粒进行测序,以确保正确的连接。SaCas9及SpyCas9对照质粒係与上文所述的pStX质粒类似地制备,其中pStX的蛋白质及引导区交换对应蛋白质及引导序列。SaCas9及SpyCas9的靶向序列获自文献或根据确立方法合理地设计。使用实例1和实例2的通用方法进行CasX构建体的表达和回收,获得了相似的结果。

[0700] 表7:CasX 488和491的序列

[0701]

构建体	DNA 序列	氨基酸序列
CasX 488	(SEQ ID NO: 236)	QEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLR KKPENIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLM SRVAQPASKKIDQNKLKPEMDEKGNLTTAGFACSQCGQPLFVYKLEQV SEKGKAYTNYFGRCNVAEHEKLILLAQLKPEKDSDEAVTYSLGKFGQR ALDFYSIHVTKESTHPVKPLAQIAGNRYASGPVGKALSDACMGTIASF LSKYQDIIIEHQKVVKGNQKRLESLRELAGKENLEYPSVTLPPQPHTK EGVDAYNEVIARVRMWVNLNLWQKLKLSRDDAKPLLRLKGFPSFPLVE RQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAGYKRQEALRPYLSS EEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAWERIDKKVEGL SKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFCRCE LKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLN LYLIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNF DDPNLIILPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNR RTRQDEPALFVALTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTD PEGCPLSRFKDSLGNPTHILRIGESYKEKQRTIQAKKEVEQRRAGGYS RKYASKAKNLADDMVRNTARDLLYYAVTQDAMLIFENLSRGFGRQGKR TFMAERQYTRMEDWLTAKLAYEGLSKTYLSKTLAQYTSKTCSNCGFTI TSADYDRVLEKLKKTATGWMTTINGKELKVEGQITYYNRYKRQNVVKD LSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFSHRPVQEKFVCL NCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDKRAFVET WQSFYRKKLKEVWKPAV (SEQ ID NO: 238)
CasX 491	(SEQ ID NO: 237)	QEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLR KKPENIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLM SRVAQPASKKIDQNKLKPEMDEKGNLTTAGFACSQCGQPLFVYKLEQV SEKGKAYTNYFGRCNVAEHEKLILLAQLKPEKDSDEAVTYSLGKFGQR ALDFYSIHVTKESTHPVKPLAQIAGNRYASGPVGKALSDACMGTIASF LSKYQDIIIEHQKVVKGNQKRLESLRELAGKENLEYPSVTLPPQPHTK EGVDAYNEVIARVRMWVNLNLWQKLKLSRDDAKPLLRLKGFPSFPLVE RQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAGYKRQEALRPYLSS EEDRKKGKKFARYQLGDLLLHLEKKHGEDWGKVYDEAWERIDKKVEGL SKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKEADKDEFCRCE LKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKDGVKKLN LYLIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNF DDPNLIILPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNR RTRQDEPALFVALTFERREVLDSSNIKPMNLIGVDRGENIPAVIALTD PEGCPLSRFKDSLGNPTHILRIGESYKEKQRTIQAKKEVEQRRAGGYS RKYASKAKNLADDMVRNTARDLLYYAVTQDAMLIFENLSRGFGRQGKR TFMAERQYTRMEDWLTAKLAYEGLSKTYLSKTLAQYTSKTCSNCGFTI TSADYDRVLEKLKKTATGWMTTINGKELKVEGQITYYNRYKRQNVVKD

[0702]

		LSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFSHRPVQEKFVCL NCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDKRAFVET WQSFYRKKLKEVWKPAV (SEQ ID NO: 239)

[0703] 实例4:CasX构建体278-280、285-288、290、291、293、300、492和493的设计及产生

[0704] 为了生成CasX 278-280、285-288、290、291、293、300、492和493构建体(表8中的序列),哺乳动物表达载体中的经密码子优化的CasX 119构建体(基于实例2的CasX Stx37构建体,编码浮霉菌门CasX SEQ ID NO:2,具有经融合NLS进行的A708K取代及[P793]缺失,及连接的引导序列及非靶向序列)的N端及C端经操纵以删除或添加NLS序列(表9中的序列)。构建体278、279及280为仅使用SV40 NLS序列的N端及C端操纵。构建体280在N端上不具有NLS且在C端上添加两个SV40 NLS,在两个SV40 NLS序列之间具有三重脯氨酸接头。通过用Q5 DNA聚合酶,对于第一片段各自使用引物oIC527及oIC528、oIC730及oIC522以及oIC730及oIC530,且为了产生第二片段各自使用oIC529及oIC520、oIC519及oIC731以及oIC529及oIC731扩增pStx34.119.174.NT而制得构建体278、279及280。通过使用Zymoclean凝胶DNA

回收试剂盒,自1%琼脂糖凝胶进行凝胶提取来纯化这些片段。使用吉布森组装将对应片段克隆在一起。将pStx34中的组装产物转化到化学感受态的Turbo感受态大肠杆菌细菌细胞中,涂铺于含有羧苄青霉素的LB-琼脂板上且在37℃下培育。挑取个别菌落且使用Qiagen Qiaprep旋转小规模制备试剂盒进行小规模制备。使用桑格测序法对所得质粒进行测序,以确保正确的组装。编码靶向所关注基因的靶向序列的序列係基于CasX PAM位置而设计。靶向序列DNA以由靶向序列及所述序列的反向互补序列组成的单链DNA(ssDNA)寡核苷酸(Integrated DNA Technologies)形式订购。将这两个寡核苷酸退火在一起,并使用T4 DNA连接酶和适当的质粒限制酶通过Golden Gate组装单独或批量克隆到pStX中。将Golden Gate产物转化到化学或电感受态细胞,如NEB Turbo感受态大肠杆菌中,涂铺于含有羧苄青霉素的LB-琼脂板上并且在37℃下培育。挑取个别菌落且使用Qiagen Qiaprep旋转小规模制备试剂盒进行小规模制备。使用桑格测序法对所得质粒进行测序,以确保正确的连接。

[0705] 为了产生构建体285-288、290、291、293及300,嵌套式PCR方法用于克隆。使用的主链载体及PCR模板为构建体pStx34 279.119.174.NT,其具有CasX 119、引导序列174及非靶向间隔区。构建体278具有配置SV40 NLS-CasX119。构建体279具有配置CasX119-SV40NLS。构建体280具有配置CasX119-SV40 NLS-PPP接头-SV40 NLS。构建体285具有配置CasX119-SV40 NLS-PPP接头-SynthNLS3。构建体286具有配置CasX119-SV40 NLS-PPP接头-SynthNLS4。构建体287具有配置CasX119-SV40 NLS-PPP接头-SynthNLS5。构建体288具有配置CasX119-SV40 NLS-PPP接头-SynthNLS6。构建体290具有配置CasX119-SV40NLS-PPP接头-EGL-13NLS。构建体291具有配置CasX119-SV40 NLS-PPP接头-c-Myc NLS。构建体293具有配置CasX119-SV40 NLS-PPP接头-核仁RNA解螺旋酶II NLS。构建体300具有配置CasX119-SV40 NLS-PPP接头-A型流感蛋白NLS。构建体492具有配置SV40NLS-CasX119-SV40NLS-PPP接头-SV40NLS。构建体493具有配置SV40NLS-CasX119-SV40NLS-PPP接头-c-Myc NLS。各变异体具有一组三个PCR;其中的两者为嵌套式的,通过凝胶提取纯化,消化,且接着连接至经消化及纯化的主链。将pStx34中的组装产物转化到化学感受态的Turbo感受态大肠杆菌细菌细胞中,涂铺于含有羧苄青霉素的LB-琼脂板上且在37℃下培育。挑取个别菌落且使用Qiagen Qiaprep旋转小规模制备试剂盒进行小规模制备。使用桑格测序法对所得质粒进行测序,以确保正确的组装。编码靶向所关注基因的靶向序列的序列係基于CasX PAM位置而设计。靶向序列DNA以由靶向序列及所述序列的反向互补序列组成的单链DNA(ssDNA)寡核苷酸(Integrated DNA Technologies)形式订购。将这两个寡核苷酸退火在一起,并使用T4 DNA连接酶和对于质粒适当的限制酶,通过Golden Gate组装个别或批量克隆到所得pStX中。将Golden Gate产物转化到化学或电感受态细胞,如NEB Turbo感受态大肠杆菌中,涂铺于含有羧苄青霉素的LB-琼脂板上并且在37℃下培育。挑取个别菌落且使用Qiagen Qiaprep旋转小规模制备试剂盒进行小规模制备。使用桑格测序法对所得质粒进行测序,以确保正确的连接。

[0706] 为了产生构建体492及493,使用XbaI及BamHI(NEB#R0145S及NEB#R3136S)根据制造商的方案消化构建体280及291。接下来,其通过使用Zymoclean凝胶DNA回收试剂盒,自1%琼脂糖凝胶进行凝胶提取来纯化。最后,使用T4DNA连接酶(NEB#M0202S)根据制造商的方案,使用XbaI及BamHI及Zymoclean凝胶DNA回收试剂盒将其连接至经消化及纯化的pStx34.119.174.NT中。将pStx34中的组装产物转化到化学感受态的Turbo感受态大肠杆菌

细菌细胞中，涂铺于含有羧苄青霉素的LB-琼脂板上且在37℃下培育。挑取个别菌落且使用Qiagen Qiaprep旋转小规模制备试剂盒进行小规模制备。使用桑格测序法对所得质粒进行测序，以确保正确的组装。编码靶向所关注基因的靶向间隔序列的序列係基于CasX PAM位置而设计。靶向序列DNA排序为由此序列的靶向间隔序列及反向互补序列组成的单链DNA(ssDNA)寡核苷酸(Integrated DNA Technologies)。将这两个寡核苷酸退火在一起，并使用T4 DNA连接酶和对于对应质粒适当的限制酶，通过Golden Gate组装个别或批量克隆到每个pStX中。将Golden Gate产物转化到化学或电感受态细胞，如NEB Turbo感受态大肠杆菌中，涂铺于含有羧苄青霉素的LB-琼脂板上并且在37℃下培育。挑取个别菌落且使用Qiagen Qiaprep旋转小规模制备试剂盒进行小规模制备。使用桑格测序法对所得质粒进行测序，以确保正确的连接。利用实例1和2的通用方法，使用质粒生产和回收CasX蛋白。

[0707] 表8:CasX 278-280、285-288、290、291、293、300、492和493序列

构建体	氨基酸序列
278	MAPKKKRKVSRQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENL RKKPENIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPK NIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRCNVSEH ERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPLEQIGGNSCAS GPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANLKDIASANGLAFPKIT LPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRDEAKPLQRLKGFPSFPLVERQ ANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAGYKRQEALRPYLSSEEDRKKGKKFAR YQFGDLLLHLEKKHGEDWGKVYDEAWERIDKKVEGLSKHIKLEEERRSEDAQSKAALT DWLRAKASFVIEGLKEADKDEFCRCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQ YNCAFIWQKDGVKKLNLYLIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPM EVNFNFDDPNLIILPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQ DEPALFVALTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLG NPTHILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDLLY YAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKTYLSKTLA QYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEGQITYYNRYKRQN VVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFSHRPVQEKFVCLNCGFET HADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDKRAFVETWQSFYRKKLKEVWKPA V (SEQ ID NO: 240)
279	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPENIPQP ISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPKNIDQRKLIPV KDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRCNVSEHERLILLSPHK PEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPLEQIGGNSCASGPVGKALSDA CMGAVASFLTKYQDIILEHQKVIKKNEKRLANLKDIASANGLAFPKITLPPQPHTKEG IEAYNNVVAQIVIWVNLNLWQKLKIGRDEAKPLQRLKGFPSFPLVERQANEVDWWDMV CNVKKLINEKKEDGKVFWQNLAGYKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHL EKKHGEDWGKVYDEAWERIDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFV IEGLKEADKDEFCRCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKD GVKKLNLYLIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPN LIILPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVALT FERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPTHILRIGE SYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDLLYYAVTQDAMLI FENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKTYLSKTLAQYTSKTCSNC GFTITSADYDRVLEKLKKTATGWMTTINGKELKVEGQITYYNRYKRQNVVKDLSVELD RLSEESVNNDISSWTKGRSGEALSLLKKRFSHRPVQEKFVCLNCGFETHADEQAALNI ARSWLFLRSQEYKKYQTNKTTGNTDKRAFVETWQSFYRKKLKEVWKPAVTSPKKKRKV (SEQ ID NO: 241)
280	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPENIPQP ISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPKNIDQRKLIPV

[0708]

[0709]

构建体	氨基酸序列
	KDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRCNVSEHERLILLSPHK PEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPLEQIGGNSCASGPVGKALSDA CMGAVASFLTKYQDIILEHQKVIKKNEKRLANLKDIASANGLAFPKITLPPQPHTKEG IEAYNNVVAQIVIWVNLNLWQKLKIGRDEAKPLQRLKGFPSFPLVERQANEVDWWDMV CNVKKLINEKKEDGKVFWQNLAGYKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHL EKKHGEDWGKVYDEAWERIDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFV IEGLKEADKDEFCRCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKD GVKKLNLYLIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPN LIILPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVALT FERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPTHILRIGE SYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDLLYYAVTQDAMLI FENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKTYLSKTLAQYTSKTCSNC GFTITSADYDRVLEKLKKTATGWMTTINGKELKVEGQITYYNRYKRQNVVKDLSVELD RLSEESVNNDISSWTKGRSGEALSLLKKRFSHRPVQEKFVCLNCGFETHADEQAALNI ARSWLFLRSQEYKKYQTNKTTGNTDKRAFVETWQSFYRKKLKEVWKPAVTSPKKKRKV PPPPKKKRKV（SEQ ID NO：242）
285	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPENIPQP ISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPKNIDQRKLIPV KDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRCNVSEHERLILLSPHK PEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPLEQIGGNSCASGPVGKALSDA CMGAVASFLTKYQDIILEHQKVIKKNEKRLANLKDIASANGLAFPKITLPPQPHTKEG IEAYNNVVAQIVIWVNLNLWQKLKIGRDEAKPLQRLKGFPSFPLVERQANEVDWWDMV CNVKKLINEKKEDGKVFWQNLAGYKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHL EKKHGEDWGKVYDEAWERIDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFV IEGLKEADKDEFCRCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKD GVKKLNLYLIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPN LIILPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVALT FERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPTHILRIGE SYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDLLYYAVTQDAMLI FENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKTYLSKTLAQYTSKTCSNC GFTITSADYDRVLEKLKKTATGWMTTINGKELKVEGQITYYNRYKRQNVVKDLSVELD RLSEESVNNDISSWTKGRSGEALSLLKKRFSHRPVQEKFVCLNCGFETHADEQAALNI ARSWLFLRSQEYKKYQTNKTTGNTDKRAFVETWQSFYRKKLKEVWKPAVTSPKKKRKV PPPHKKKHPDASVNFSEFSK（SEQ ID NO：243）
286	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPENIPQP ISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPKNIDQRKLIPV KDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRCNVSEHERLILLSPHK PEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPLEQIGGNSCASGPVGKALSDA CMGAVASFLTKYQDIILEHQKVIKKNEKRLANLKDIASANGLAFPKITLPPQPHTKEG IEAYNNVVAQIVIWVNLNLWQKLKIGRDEAKPLQRLKGFPSFPLVERQANEVDWWDMV CNVKKLINEKKEDGKVFWQNLAGYKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHL EKKHGEDWGKVYDEAWERIDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFV IEGLKEADKDEFCRCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKD GVKKLNLYLIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPN LIILPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVALT FERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPTHILRIGE SYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDLLYYAVTQDAMLI FENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKTYLSKTLAQYTSKTCSNC GFTITSADYDRVLEKLKKTATGWMTTINGKELKVEGQITYYNRYKRQNVVKDLSVELD

[0710]

构建体	氨基酸序列
	RLSEESVNNDISSWTKGRSGEALSLLKKRFSHRPVQEKFVCLNCGFETHADEQAALNI ARSWLFLRSQEYKKYQTNKTTGNTDKRAFVETWQSFYRKKLKEVWKPAVTSPKKKRKV PPPQRPGPYDRPQRPGPYDRP（SEQ ID NO：244）
287	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPENIPQP ISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPKNIDQRKLIPV KDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRCNVSEHERLILLSPHK PEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPLEQIGGNSCASGPVGKALSDA CMGAVASFLTKYQDIILEHQKVIKKNEKRLANLKDIASANGLAFPKITLPPQPHTKEG IEAYNNVVAQIVIWVNLNLWQKLKIGRDEAKPLQRLKGFPSFPLVERQANEVDWWDMV CNVKKLINEKKEDGKVFWQNLAGYKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHL EKKHGEDWGKVYDEAWERIDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFV IEGLKEADKDEFCRCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKD GVKKLNLYLIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPN LIILPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVALT FERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPTHILRIGE SYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDLLYYAVTQDAMLI FENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKTYLSKTLAQYTSKTCSNC GFTITSADYDRVLEKLKKTATGWMTTINGKELKVEGQITYYNRYKRQNVVKDLSVELD RLSEESVNNDISSWTKGRSGEALSLLKKRFSHRPVQEKFVCLNCGFETHADEQAALNI ARSWLFLRSQEYKKYQTNKTTGNTDKRAFVETWQSFYRKKLKEVWKPAVTSPKKKRKV PPPLSPSLSPLLSPSLSPL（SEQ ID NO：245）
288	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPENIPQP ISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPKNIDQRKLIPV KDGNERLTMSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRCNVSEHERLILLSPH KPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPLEQIGGNSCASGPVGKALSD ACMGAVASFLTKYQDIILEHQKVIKKNEKRLANLKDIASANGLAFPKITLPPQPHTKE GIEAYNNVVAQIVIWVNLNLWQKLKIGRDEAKPLQRLKGFPSFPLVERQANEVDWWDM VCNVKKLINEKKEDGKVFWQNLAGYKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLH LEKKHGEDWGKVYDEAWERIDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASF VIEGLKEADKDEFCRCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQK DGVKKLNLYLIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDP NLIILPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVAL TFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPTHILRIG ESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDLLYYAVTQDAML IFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKTYLSKTLAQYTSKTCSN CGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEGQITYYNRYKRQNVVKDLSVEL DRLSEESVNNDISSWTKGRSGEALSLLKKRFSHRPVQEKFVCLNCGFETHADEQAALN IARSWLFLRSQEYKKYQTNKTTGNTDKRAFVETWQSFYRKKLKEVWKPAVTSPKKKRK VPPPRGKGGKGLGKGGAKRHRK（SEQ ID NO：246）
290	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPENIPQP ISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPKNIDQRKLIPV KDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRCNVSEHERLILLSPHK PEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPLEQIGGNSCASGPVGKALSDA CMGAVASFLTKYQDIILEHQKVIKKNEKRLANLKDIASANGLAFPKITLPPQPHTKEG IEAYNNVVAQIVIWVNLNLWQKLKIGRDEAKPLQRLKGFPSFPLVERQANEVDWWDMV CNVKKLINEKKEDGKVFWQNLAGYKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHL EKKHGEDWGKVYDEAWERIDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFV IEGLKEADKDEFCRCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKD

[0711]

构建体	氨基酸序列
	GVKKLNLYLIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPN LIILPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVALT FERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPTHILRIGE SYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDLLYYAVTQDAMLI FENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKTYLSKTLAQYTSKTCSNC GFTITSADYDRVLEKLKKTATGWMTTINGKELKVEGQITYYNRYKRQNVVKDLSVELD RLSEESVNNDISSWTKGRSGEALSLLKKRFSHRPVQEKFVCLNCGFETHADEQAALNI ARSWLFLRSQEYKKYQTNKTTGNTDKRAFVETWQSFYRKKLKEVWKPAVTSPKKKRKV PPPSRRRKANPTKLSENAKKLAKEVEN（SEQ ID NO：247）
291	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPENIPQP ISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPKNIDQRKLIPV KDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRCNVSEHERLILLSPHK PEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPLEQIGGNSCASGPVGKALSDA CMGAVASFLTKYQDIILEHQKVIKKNEKRLANLKDIASANGLAFPKITLPPQPHTKEG IEAYNNVVAQIVIWVNLNLWQKLKIGRDEAKPLQRLKGFPSFPLVERQANEVDWWDMV CNVKKLINEKKEDGKVFWQNLAGYKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHL EKKHGEDWGKVYDEAWERIDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFV IEGLKEADKDEFCRCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKD GVKKLNLYLIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPN LIILPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVALT FERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPTHILRIGE SYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDLLYYAVTQDAMLI FENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKTYLSKTLAQYTSKTCSNC GFTITSADYDRVLEKLKKTATGWMTTINGKELKVEGQITYYNRYKRQNVVKDLSVELD RLSEESVNNDISSWTKGRSGEALSLLKKRFSHRPVQEKFVCLNCGFETHADEQAALNI ARSWLFLRSQEYKKYQTNKTTGNTDKRAFVETWQSFYRKKLKEVWKPAVTSPKKKRKV PPPPAAKRVKLD（SEQ ID NO：248）
293	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPENIPQP ISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPKNIDQRKLIPV KDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRCNVSEHERLILLSPHK PEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPLEQIGGNSCASGPVGKALSDA CMGAVASFLTKYQDIILEHQKVIKKNEKRLANLKDIASANGLAFPKITLPPQPHTKEG IEAYNNVVAQIVIWVNLNLWQKLKIGRDEAKPLQRLKGFPSFPLVERQANEVDWWDMV CNVKKLINEKKEDGKVFWQNLAGYKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHL EKKHGEDWGKVYDEAWERIDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFV IEGLKEADKDEFCRCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKD GVKKLNLYLIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPN LIILPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVALT FERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPTHILRIGE SYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDLLYYAVTQDAMLI FENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKTYLSKTLAQYTSKTCSNC GFTITSADYDRVLEKLKKTATGWMTTINGKELKVEGQITYYNRYKRQNVVKDLSVELD RLSEESVNNDISSWTKGRSGEALSLLKKRFSHRPVQEKFVCLNCGFETHADEQAALNI ARSWLFLRSQEYKKYQTNKTTGNTDKRAFVETWQSFYRKKLKEVWKPAVTSPKKKRKV PPPKRSFSKAF（SEQ ID NO：249）
300	MQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLENLRKKPENIPQP ISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAPKNIDQRKLIPV KDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRCNVSEHERLILLSPHK

[0712]

构建体	氨基酸序列
	PEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPLEQIGGNSCASGPVGKALSDA CMGAVASFLTKYQDIILEHQKVIKKNEKRLANLKDIASANGLAFPKITLPPQPHTKEG IEAYNNVVAQIVIWVNLNLWQKLKIGRDEAKPLQRLKGFPSFPLVERQANEVDWWDMV CNVKKLINEKKEDGKVFWQNLAGYKRQEALRPYLSSEEDRKKGKKFARYQFGDLLLHL EKKHGEDWGKVYDEAWERIDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFV IEGLKEADKDEFCRCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFIWQKD GVKKLNLYLIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVPMEVNFNFDDPN LIILPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVALT FERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPTHILRIGE SYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDLLYYAVTQDAMLI FENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKTYLSKTLAQYTSKTCSNC GFTITSADYDRVLEKLKKTATGWMTTINGKELKVEGQITYYNRYKRQNVVKDLSVELD RLSEESVNNDISSWTKGRSGEALSLLKKRFSHRPVQEKFVCLNCGFETHADEQAALNI ARSWLFLRSQEYKKYQTNKTTGNTDKRAFVETWQSFYRKKLKEVWKPAVTSPKKKRKV PPPKRGINDRNFWRGENERKTR（SEQ ID NO：250）
492	MAPKKKRKVSRMQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLEN LRKKPENIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAP KNIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRCNVSE HERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPLEQIGGNSCA SGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANLKDIASANGLAFPKI TLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRDEAKPLQRLKGFPSFPLVER QANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAGYKRQEALRPYLSSEEDRKKGKKFA RYQFGDLLLHLEKKHGEDWGKVYDEAWERIDKKVEGLSKHIKLEEERRSEDAQSKAAL TDWLRAKASFVIEGLKEADKDEFCRCELKLQKWYGDLRGKPFAIEAENSILDISGFSK QYNCAFIWQKDGVKKLNLYLIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVP MEVNFNFDDPNLIILPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTR QDEPALFVALTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSL GNPTHILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDLL YYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKTYLSKTL AQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEGQITYYNRYKRQ NVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFSHRPVQEKFVCLNCGFE THADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDKRAFVETWQSFYRKKLKEVWKP AVTSPKKKRKVPPPPKKKRKV（SEQ ID NO：251）
493	MAPKKKRKVSRMQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMTPDLRERLEN LRKKPENIPQPISNTSRANLNKLLTDYTEMKKAILHVYWEEFQKDPVGLMSRVAQPAP KNIDQRKLIPVKDGNERLTSSGFACSQCCQPLYVYKLEQVNDKGKPHTNYFGRCNVSE HERLILLSPHKPEANDELVTYSLGKFGQRALDFYSIHVTRESNHPVKPLEQIGGNSCA SGPVGKALSDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANLKDIASANGLAFPKI TLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRDEAKPLQRLKGFPSFPLVER QANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAGYKRQEALRPYLSSEEDRKKGKKFA RYQFGDLLLHLEKKHGEDWGKVYDEAWERIDKKVEGLSKHIKLEEERRSEDAQSKAAL TDWLRAKASFVIEGLKEADKDEFCRCELKLQKWYGDLRGKPFAIEAENSILDISGFSK QYNCAFIWQKDGVKKLNLYLIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGEIVP MEVNFNFDDPNLIILPLAFGKRQGREFIWNDLLSLETGSLKLANGRVIEKTLYNRRTR QDEPALFVALTFERREVLDSSNIKPMNLIGIDRGENIPAVIALTDPEGCPLSRFKDSL GNPTHILRIGESYKEKQRTIQAKKEVEQRRAGGYSRKYASKAKNLADDMVRNTARDLL YYAVTQDAMLIFENLSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKTYLSKTL AQYTSKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEGQITYYNRYKRQ NVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFSHRPVQEKFVCLNCGFE

[0713]

构建体	氨基酸序列
	THADEQAALNIARSWLFLRSQEYKKYQTNKTTGNTDKRAFVETWQSFYRKKLKEVWKP AVTSPKKKRKVPPPPAAKRVKLD（SEQ ID NO：252）

[0714] 表9:核定位序列列表

[0715]

CasX	NLS	DNA 序列	氨基酸序列
278, 279, 280, 492, 493	SV40	CCAAAGAAGAAGCGGAAGGTC (SEQ ID NO: 253)	PKKKRKV(SEQ ID NO: 165)
285	SynthNLS3	CACAAGAAGAAACATCCAGACGCATCAGTCAACTTTAGCGAGTTCAGTAAA (SEQ ID NO: 254)	HKKKHPDASVNFSEFSK (SEQ ID NO: 196)
286	SynthNLS4	CAGCGCCCTGGGCCTTACGATAGGCCGCAAAGACCCGGACCGTATGATCGCCCT (SEQ ID NO: 255)	QRPGPYDRPQRPGPYDRP (SEQ ID NO: 197)
287	SynthNLS5	CTCAGCCCGAGTCTTAGTCCACTGCTTTCCCCGTCCCTGTCTCCACTG (SEQ ID NO: 256)	LSPSLSPLLSPSLSPL (SEQ ID NO: 198)
288	SynthNLS6	CGGGGCAAGGGTGGCAAGGGGCTTGGCAAGGGGGGGGCAAAGAGGCACAGGAAG (SEQ ID NO: 257)	RGKGGKGLGKGGAKRHRK (SEQ ID NO: 199)
290	EGL-13	AGCCGCCGCAGAAAAGCCAATCCTACAAAACTGTCAGAAAATGCGAAAAAACTTGCTAAGGAGGTGGAAAAC (SEQ ID NO: 258)	SRRRKANPTKLSENAKKLAKEVEN (SEQ ID NO: 192)
291	c-Myc	CCTGCCGCAAAGCGAGTGAAATTGGAC (SEQ ID NO: 259)	PAAKRVKLD (SEQ ID NO: 167)
293	核仁RNA解螺旋酶II	AAGCGGTCCTTCAGTAAGGCCTTT (SEQ ID NO: 260)	KRSFSKAF (SEQ ID NO: 188)
300	甲型流感蛋白	AAACGGGGAATAAACGACCGGAACTTCTGGCGCGGGGAAAACGAGCGCAAAACCCGA (SEQ ID NO: 261)	KRGINDRNFWRGENERKTR(SEQ ID NO: 186)

[0716] 实例5:CasX构建体387、395、485-491及494的设计及产生

[0717] 为了生呈CasX395、CasX485、CasX486、CasX487,经密码子优化的CasX 119(基于实例2的CasX 37构建体,编码浮霉菌门CasX SEQ ID NO:2,具有经融合NLS进行的A708K取代及[P793]缺失,及连接引导及非靶向序列)、CasX435、CasX438及CasX484(各自基于实例2的CasX 119构建体,编码浮霉菌门CasX SEQ ID NO:2,具有经融合NLS进行的L379R取代、A708K取代及[P793]缺失,及连接引导及非靶向序列)分别使用标准克隆方法克隆至包含KanR标记物、colE1 ori及具有融合NLS的CasX(pStx1)的4kb分段载体中。吉布森引物经设计以自其自身载体中的氨基酸192-331扩增CasX SEQ ID NO:1螺旋形I结构域,以分别替换pStx1中的CasX119、CasX435、CasX438及CasX484上的此对应区域(aa 193-332)。使用Q5 DNA聚合酶,根据制造商的方案,通过引物oIC768及oIC784扩增来自CasX SEQ ID NO:1的螺旋形I结构域。使用Q5 DNA聚合酶,根据制造商的方案,通过引物oIC765及oIC764扩增含有所需CasX变异体的目的载体。通过使用Zymoclean凝胶DNA回收试剂盒,自1%琼脂糖凝胶进行凝胶提取来纯化两个片段。随后使用吉布森组装将插入物和主链片段拼接在一起。将pStx1分段载体中的组装产物转化到化学感受态的Turbo感受态大肠杆菌细菌细胞中,涂铺于含有卡那霉素的LB-琼脂板上且在37℃下培育。挑取个别菌落且使用Qiagen Qiaprep旋转小规模制备试剂盒进行小规模制备。使用桑格测序法对所得质粒进行测序,以确保正确的组装。然后使用标准克隆方法将正确的克隆剪切并粘贴到哺乳动物表达质粒中(见图5)。使用桑格测序法对所得质粒进行测序,以确保正确的组装。编码靶向所关注基因的靶向间隔序列的序列係基于CasX PAM位置而设计。靶向间隔序列DNA排序为由此序列的靶向序列及反向互补序列组成的单链DNA(ssDNA)寡核苷酸(Integrated DNA Technologies)。将这

两个寡核苷酸退火在一起,并使用T4 DNA连接酶和适当的质粒限制酶通过Golden Gate组装单独或批量克隆到pStX中。将Golden Gate产物转化到化学或电感受态细胞,如NEB Turbo感受态大肠杆菌中,涂铺于含有羧苄青霉素的LB-琼脂板上并且在37℃下培育。挑取个别菌落且使用Qiagen Qiaprep旋转小规模制备试剂盒进行小规模制备。使用桑格测序法对所得质粒进行测序,以确保正确的连接。

[0718] 为了生成CasX 488、CasX 489、CasX 490及CasX 491(表10中的序列),经密码子优化的CasX 119(基于实例2的CasX 37构建体,编码浮霉菌门CasX SEQ ID NO:2,具有经融合NLS进行的A708K取代及[P793]缺失,及连接的引导序列及非靶向序列)、CasX 435、CasX438及CasX 484(各自基于实例2的CasX 119构建体,编码浮霉菌门CasX SEQ ID NO:2,具有经融合NLS进行的L379R取代、A708K取代及[P793]缺失,及连接的引导序列及非靶向序列)分别使用标准克隆方法克隆至由KanR标记物、colE1 ori及具有融合NLS的STX(pStx1)构成的4kb分段载体中。吉布森引物经设计以扩增其自身载体中的来自氨基酸101-191的CasX Stx1 NTSB结构域及来自氨基酸192-331的螺旋形I结构域,以分别替换pStx1中的CasX119、CasX435、CasX438及CasX484上的此类似区域(aa 103-332)。使用Q5 DNA聚合酶,根据制造商的方案,通过引物oIC766及oIC784扩增来自CasX SEQ ID NO:1的NTSB及螺旋形I结构域。使用Q5 DNA聚合酶,根据制造商的方案,通过引物oIC762及oIC765扩增含有所需CasX变异体的目的载体。通过使用Zymoclean凝胶DNA回收试剂盒,自1%琼脂糖凝胶进行凝胶提取来纯化两个片段。随后使用吉布森组装将插入物和主链片段拼接在一起。将pStx1分段载体中的组装产物转化到化学感受态的Turbo感受态大肠杆菌细菌细胞中,涂铺于含有卡那霉素的LB-琼脂板上且在37℃下培育。挑取个别菌落且使用Qiagen Qiaprep旋转小规模制备试剂盒进行小规模制备。使用桑格测序法对所得质粒进行测序,以确保正确的组装。然后使用标准克隆方法将正确的克隆剪切并粘贴到哺乳动物表达质粒中(见图5)。使用桑格测序法对所得质粒进行测序,以确保正确的组装。编码靶向所关注基因的靶向间隔序列的序列係基于CasX PAM位置而设计。靶向间隔序列DNA排序为由此序列的靶向序列及反向互补序列组成的单链DNA(ssDNA)寡核苷酸(Integrated DNA Technologies)。将这两个寡核苷酸退火在一起,并使用T4 DNA连接酶和适当的质粒限制酶通过Golden Gate组装单独或批量克隆到pStX中。将Golden Gate产物转化到化学或电感受态细胞,如NEB Turbo感受态大肠杆菌中,涂铺于含有羧苄青霉素的LB-琼脂板上并且在37℃下培育。挑取个别菌落且使用Qiagen Qiaprep旋转小规模制备试剂盒进行小规模制备。使用桑格测序法对所得质粒进行测序,以确保正确的连接。

[0719] 为了生成CasX 387及CasX 494(表10中的序列),经密码子优化的CasX 119(基于实例2的CasX 37构建体,编码浮霉菌门CasX SEQ ID NO:2,具有经融合NLS进行的A708K取代及[P793]缺失,及连接的引导序列及非靶向序列)及CasX 484(基于实例2的CasX119构建体,编码浮霉菌门CasX SEQ ID NO:2,具有经融合NLS进行的L379R取代、A708K取代及[P793]缺失,及连接的引导序列及非靶向序列)分别使用标准克隆方法克隆至由KanR标记物、colE1 ori及具有融合NLS的STX(pStx1)构成的4kb分段载体中。吉布森引物经设计以扩增其自身载体中来自氨基酸101-191的CasX Stx1 NTSB结构域,以分别替换pStx1中的CasX119及CasX484上的此类似区域(aa 103-192)。使用Q5 DNA聚合酶,根据制造商的方案,通过引物oIC766及oIC767扩增来自CasX Stx1的NTSB结构域。使用Q5 DNA聚合酶,根据制造

商的方案，通过引物oIC763及oIC762扩增含有所需CasX变异体的目的载体。通过使用Zymoclean凝胶DNA回收试剂盒，自1％琼脂糖凝胶进行凝胶提取来纯化两个片段。随后使用吉布森组装将插入物和主链片段拼接在一起。将pStx1分段载体中的组装产物转化到化学感受态的Turbo感受态大肠杆菌细菌细胞中，涂铺于含有卡那霉素的LB-琼脂板上且在37℃下培育。挑取个别菌落且使用Qiagen Qiaprep旋转小规模制备试剂盒进行小规模制备。使用桑格测序法对所得质粒进行测序，以确保正确的组装。然后使用标准克隆方法将正确的克隆剪切并粘贴到哺乳动物表达质粒中(见图5)。使用桑格测序法对所得质粒进行测序，以确保正确的组装。编码靶向所关注基因的靶向序列的序列係基于CasX PAM位置而设计。靶向序列DNA以由靶向序列及所述序列的反向互补序列组成的单链DNA(ssDNA)寡核苷酸(Integrated DNA Technologies)形式订购。将这两个寡核苷酸退火在一起，并使用T4 DNA连接酶和适当的质粒限制酶通过Golden Gate组装单独或批量克隆到pStX中。将Golden Gate产物转化到化学或电感受态细胞，如NEB Turbo感受态大肠杆菌中，涂铺于含有羧苄青霉素的LB-琼脂板上并且在37℃下培育。挑取个别菌落且使用Qiagen Qiaprep旋转小规模制备试剂盒进行小规模制备。使用桑格测序法对所得质粒进行测序，以确保正确的连接。所得构建体的序列在表10中列出。

[0720] 表10:CasX 395和485-491的序列

[0721]

构建体	DNA 序列	氨基酸序列
CasX 387	(SEQ ID NO: 262)	MAPKKKRKVSRQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMT PDLRERLENLRKKPENIPQPISNTSRANLNKLLTDYTEMKKAILHVYW EEFQKDPVGLMSRVAQPASKKIDQNKLKPEMDEKGNLTTAGFACSQCG QPLFVYKLEQVSEKGKAYTNYFGRCNVAEHEKLILLAQLKPEKDSDEA VTYSLGKFGQRALDFYSIHVTRESNHPVKPLEQIGGNSCASGPVGKAL SDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANLKDIASANGLAFP KITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRDEAKPLQR LKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAGYK RQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEA WERIDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLK EADKDEFCRCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAF IWQKDGVKKLNLYLIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSG EIVPMEVNFNFDDPNLIILPLAFGKRQGREFIWNDLLSLETGSLKLAN GRVIEKTLYNRRTRQDEPALFVALTFERREVLDSSNIKPMNLIGIDRG ENIPAVIALTDPEGCPLSRFKDSLGNPTHILRIGESYKEKQRTIQAKK EVEQRRAGGYSRKYASKAKNLADDMVRNTARDLLYYAVTQDAMLIFEN LSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKTYLSKTLAQYT SKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEGQITYY NRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTT GNTDKRAFVETWQSFYRKKLKEVWKPAVTSPKKKRKV (SEQ ID NO: 272)
CasX 395	(SEQ ID NO: 263)	MAPKKKRKVSRQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMT PDLRERLENLRKKPENIPQPISNTSRANLNKLLTDYTEMKKAILHVYW EEFQKDPVGLMSRVAQPAPKNIDQRKLIPVKDGNERLTSSGFACSQCC QPLYVYKLEQVNDKGKPHTNYFGRCNVSEHERLILLSPHKPEANDELV TYSLGKFGQRALDFYSIHVTKESTHPVKPLAQIAGNRYASGPVGKALS DACMGTIASFLSKYQDIIIEHQKVVKGNQKRLESLRELAGKENLEYPS VTLPPQPHTKEGVDAYNEVIARVRMWVNLNLWQKLKLSRDDAKPLLRL KGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAGYKR QEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEAW ERIDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKE ADKDEFCRCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFI

[0722]

构建体	DNA 序列	氨基酸序列
		WQKDGVKKLNLYLIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGE IVPMEVNFNFDDPNLIILPLAFGKRQGREFIWNDLLSLETGSLKLANG RVIEKTLYNRRTRQDEPALFVALTFERREVLDSSNIKPMNLIGIDRGE NIPAVIALTDPEGCPLSRFKDSLGNPTHILRIGESYKEKQRTIQAKKE VEQRRAGGYSRKYASKAKNLADDMVRNTARDLLYYAVTQDAMLIFENL SRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKTYLSKTLAQYTS KTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEGQITYYN RYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFSH RPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTG NTDKRAFVETWQSFYRKKLKEVWKPAVTSPKKKRKVTSPKKKRKV (SEQ ID NO: 273)
CasX 485	(SEQ ID NO: 264)	MAPKKKRKVSRQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMT PDLRERLENLRKKPENIPQPISNTSRANLNKLLTDYTEMKKAILHVYW EEFQKDPVGLMSRVAQPAPKNIDQRKLIPVKDGNERLTSSGFACSQCC QPLYVYKLEQVNDKGKPHTNYFGRCNVSEHERLILLSPHKPEANDELV TYSLGKFGQRALDFYSIHVTKESTHPVKPLAQIAGNRYASGPVGKALS DACMGTIASFLSKYQDIIIEHQKVVKGNQKRLESLRELAGKENLEYPS VTLPPQPHTKEGVDAYNEVIARVRMWVNLNLWQKLKLSRDDAKPLLRL KGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAGYKR QEALRPYLSSEEDRKKGKKFARYQLGDLLLHLEKKHGEDWGKVYDEAW ERIDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKE ADKDEFCRCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFI WQKDGVKKLNLYLIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGE IVPMEVNFNFDDPNLIILPLAFGKRQGREFIWNDLLSLETGSLKLANG RVIEKTLYNRRTRQDEPALFVALTFERREVLDSSNIKPMNLIGVDRGE NIPAVIALTDPEGCPLSRFKDSLGNPTHILRIGESYKEKQRTIQAKKE VEQRRAGGYSRKYASKAKNLADDMVRNTARDLLYYAVTQDAMLIFENL SRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKTYLSKTLAQYTS KTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEGQITYYN RRKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFSH RPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTG NTDKRAFVETWQSFYRKKLKEVWKPAVTSPKKKRKV (SEQ ID NO: 274)
CasX 486	(SEQ ID NO: 265)	MAPKKKRKVSRQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMT PDLRERLENLRKKPENIPQPISNTSRANLNKLLTDYTEMKKAILHVYW EEFQKDPVGLMSRVAQPAPKNIDQRKLIPVKDGNERLTSSGFACSQCC QPLYVYKLEQVNDKGKPHTNYFGRCNVSEHERLILLSPHKPEANDELV TYSLGKFGQRALDFYSIHVTKESTHPVKPLAQIAGNRYASGPVGKALS DACMGTIASFLSKYQDIIIEHQKVVKGNQKRLESLRELAGKENLEYPS VTLPPQPHTKEGVDAYNEVIARVRMWVNLNLWQKLKLSRDDAKPLLRL KGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAGYKR QEALRPYLSSEEDRKKGKKFARYQLGDLLKHLEKKHGEDWGKVYDEAW ERIDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKE ADKDEFCRCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFI WQKDGVKKLNLYLIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGE IVPMEVNFNFDDPNLIILPLAFGKRQGREFIWNDLLSLETGSLKLANG RVIEKTLYNRRTRQDEPALFVALTFERREVLDSSNIKPMNLIGVDRGE NIPAVIALTDPEGCPLSRFKDSLGNPTHILRIGESYKEKQRTIQAKKE VEQRRAGGYSRKYASKAKNLADDMVRNTARDLLYYAVTQDAMLIFENL

[0723]

构建体	DNA 序列	氨基酸序列
		SRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKTYLSKTLAQYTS KTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEGQITYYN RRKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFSH RPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTG NTDKRAFVETWQSFYRKKLKEVWKPAVTSPKKKRKV（SEQ ID NO: 275）
CasX 487	（SEQ ID NO：266）	MAPKKKRKVSRQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMT PDLRERLENLRKKPENIPQPISNTSRANLNKLLTDYTEMKKAILHVYW EEFQKDPVGLMSRVAQPAPKNIDQRKLIPVKDGNERLTSSGFACSQCC QPLYVYKLEQVNDKGKPHTNYFGRCNVSEHERLILLSPHKPEANDELV TYSLGKFGQRALDFYSIHVTKESTHPVKPLAQIAGNRYASGPVGKALS DACMGTIASFLSKYQDIIIEHQKVVKGNQKRLESLRELAGKENLEYPS VTLPPQPHTKEGVDAYNEVIARVRMWVNLNLWQKLKLSRDDAKPLLRL KGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAGYKR QEALRPYLSSEEDRKKGKKFARYQLGDLLLHLEKKHGEDWGKVYDEAW ERIDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLKE ADKDEFCRCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAFI WQKDGVKKLNLYLIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSGE IVPMEVNFNFDDPNLIILPLAFGKRQGREFIWNDLLSLETGSLKLANG RVIEKTLYNRRTRQDEPALFVALTFERREVLDSSNIKPMNLIGVDRGE NIPAVIALTDPEGCPLSRFKDSLGNPTHILRIGESYKEKQRTIQAKKE VEQRRAGGYSRKYASKAKNLADDMVRNTARDLLYYAVTQDAMLIFENL SRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKTYLSKTLAQYTS KTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEGQITYYN RYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFSH RPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTTG NTDKRAFVETWQSFYRKKLKEVWKPAVTSPKKKRKV（SEQ ID NO: 276）
CasX 488	（SEQ ID NO：267）	MAPKKKRKVSRQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMT PDLRERLENLRKKPENIPQPISNTSRANLNKLLTDYTEMKKAILHVYW EEFQKDPVGLMSRVAQPASKKIDQNKLKPEMDEKGNLTTAGFACSQCG QPLFVYKLEQVSEKGKAYTNYFGRCNVAEHEKLILLAQLKPEKDSDEA VTYSLGKFGQRALDFYSIHVTKESTHPVKPLAQIAGNRYASGPVGKAL SDACMGTIASFLSKYQDIIIEHQKVVKGNQKRLESLRELAGKENLEYP SVTLPPQPHTKEGVDAYNEVIARVRMWVNLNLWQKLKLSRDDAKPLLR LKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAGYK RQEALRPYLSSEEDRKKGKKFARYQFGDLLLHLEKKHGEDWGKVYDEA WERIDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLK EADKDEFCRCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAF IWQKDGVKKLNLYLIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSG EIVPMEVNFNFDDPNLIILPLAFGKRQGREFIWNDLLSLETGSLKLAN GRVIEKTLYNRRTRQDEPALFVALTFERREVLDSSNIKPMNLIGIDRG ENIPAVIALTDPEGCPLSRFKDSLGNPTHILRIGESYKEKQRTIQAKK EVEQRRAGGYSRKYASKAKNLADDMVRNTARDLLYYAVTQDAMLIFEN LSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKTYLSKTLAQYT SKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEGQITYY NRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTT

[0724]

构建体	DNA 序列	氨基酸序列
		GNTDKRAFVETWQSFYRKKLKEVWKPAVTSPKKKRKV（SEQ ID NO: 277）
CasX 489	（SEQ ID NO: 268）	MAPKKKRKVSRQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMT PDLRERLENLRKKPENIPQPISNTSRANLNKLLTDYTEMKKAILHVYW EEFQKDPVGLMSRVAQPASKKIDQNKLKPEMDEKGNLTTAGFACSQCG QPLFVYKLEQVSEKGKAYTNYFGRCNVAEHEKLILLAQLKPEKDSDEA VTYSLGKFGQRALDFYSIHVTKESTHPVKPLAQIAGNRYASGPVGKAL SDACMGTIASFLSKYQDIIIEHQKVVKGNQKRLESLRELAGKENLEYP SVTLPPQPHTKEGVDAYNEVIARVRMWVNLNLWQKLKLSRDDAKPLLR LKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAGYK RQEALRPYLSSEEDRKKGKKFARYQLGDLLLHLEKKHGEDWGKVYDEA WERIDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLK EADKDEFCRCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAF IWQKDGVKKLNLYLIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSG EIVPMEVNFNFDDPNLIILPLAFGKRQGREFIWNDLLSLETGSLKLAN GRVIEKTLYNRRTRQDEPALFVALTFERREVLDSSNIKPMNLIGVDRG ENIPAVIALTDPEGCPLSRFKDSLGNPTHILRIGESYKEKQRTIQAKK EVEQRRAGGYSRKYASKAKNLADDMVRNTARDLLYYAVTQDAMLIFEN LSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKTYLSKTLAQYT SKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEGQITYY NRRKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTT GNTDKRAFVETWQSFYRKKLKEVWKPAVTSPKKKRKV（SEQ ID NO: 278）
CasX 490	（SEQ ID NO: 269）	MAPKKKRKVSRQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMT PDLRERLENLRKKPENIPQPISNTSRANLNKLLTDYTEMKKAILHVYW EEFQKDPVGLMSRVAQPASKKIDQNKLKPEMDEKGNLTTAGFACSQCG QPLFVYKLEQVSEKGKAYTNYFGRCNVAEHEKLILLAQLKPEKDSDEA VTYSLGKFGQRALDFYSIHVTKESTHPVKPLAQIAGNRYASGPVGKAL SDACMGTIASFLSKYQDIIIEHQKVVKGNQKRLESLRELAGKENLEYP SVTLPPQPHTKEGVDAYNEVIARVRMWVNLNLWQKLKLSRDDAKPLLR LKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAGYK RQEALRPYLSSEEDRKKGKKFARYQLGDLLKHLEKKHGEDWGKVYDEA WERIDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLK EADKDEFCRCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAF IWQKDGVKKLNLYLIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSG EIVPMEVNFNFDDPNLIILPLAFGKRQGREFIWNDLLSLETGSLKLAN GRVIEKTLYNRRTRQDEPALFVALTFERREVLDSSNIKPMNLIGVDRG ENIPAVIALTDPEGCPLSRFKDSLGNPTHILRIGESYKEKQRTIQAKK EVEQRRAGGYSRKYASKAKNLADDMVRNTARDLLYYAVTQDAMLIFEN LSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKTYLSKTLAQYT SKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEGQITYY NRRKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTT GNTDKRAFVETWQSFYRKKLKEVWKPAVTSPKKKRKV（SEQ ID NO: 279）
CasX 491	（SEQ ID NO: 270）	MAPKKKRKVSRQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMT PDLRERLENLRKKPENIPQPISNTSRANLNKLLTDYTEMKKAILHVYW

构建体	DNA 序列	氨基酸序列
		EEFQKDPVGLMSRVAQPASKKIDQNKLKPEMDEKGNLTTAGFACSQCG QPLFVYKLEQVSEKGKAYTNYFGRCNVAEHEKLILLAQLKPEKDSDEA VTYSLGKFGQRALDFYSIHVTKESTHPVKPLAQIAGNRYASGPVGKAL SDACMGTIASFLSKYQDIIIEHQKVVKGNQKRLESLRELAGKENLEYP SVTLPPQPHTKEGVDAYNEVIARVRMWVNLNLWQKLKLSRDDAKPLLR LKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAGYK RQEALRPYLSSEEDRKKGKKFARYQLGDLLLHLEKKHGEDWGKVYDEA WERIDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLK EADKDEFCRCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAF IWQKDGVKKLNLYLIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSG EIVPMEVNFNFDDPNLIILPLAFGKRQGREFIWNDLLSLETGSLKLAN GRVIEKTLYNRRTRQDEPALFVALTFERREVLDSSNIKPMNLIGVDRG ENIPAVIALTDPEGCPLSRFKDSLGNPTHILRIGESYKEKQRTIQAKK EVEQRRAGGYSRKYASKAKNLADDMVRNTARDLLYYAVTQDAMLIFEN LSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKTYLSKTLAQYT SKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEGQITYY NRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTT GNTDKRAFVETWQSFYRKKLKEVWKPAVTSPKKKRKV（SEQ ID NO: 280）
[0725] CasX 494	（SEQ ID NO：271）	MAPKKKRKVSRQEIKRINKIRRRLVKDSNTKKAGKTGPMKTLLVRVMT PDLRERLENLRKKPENIPQPISNTSRANLNKLLTDYTEMKKAILHVYW EEFQKDPVGLMSRVAQPASKKIDQNKLKPEMDEKGNLTTAGFACSQCG QPLFVYKLEQVSEKGKAYTNYFGRCNVAEHEKLILLAQLKPEKDSDEA VTYSLGKFGQRALDFYSIHVTRESNHPVKPLEQIGGNSCASGPVGKAL SDACMGAVASFLTKYQDIILEHQKVIKKNEKRLANLKDIASANGLAFP KITLPPQPHTKEGIEAYNNVVAQIVIWVNLNLWQKLKIGRDEAKPLQR LKGFPSFPLVERQANEVDWWDMVCNVKKLINEKKEDGKVFWQNLAGYK RQEALRPYLSSEEDRKKGKKFARYQLGDLLLHLEKKHGEDWGKVYDEA WERIDKKVEGLSKHIKLEEERRSEDAQSKAALTDWLRAKASFVIEGLK EADKDEFCRCELKLQKWYGDLRGKPFAIEAENSILDISGFSKQYNCAF IWQKDGVKKLNLYLIINYFKGGKLRFKKIKPEAFEANRFYTVINKKSG EIVPMEVNFNFDDPNLIILPLAFGKRQGREFIWNDLLSLETGSLKLAN GRVIEKTLYNRRTRQDEPALFVALTFERREVLDSSNIKPMNLIGVDRG ENIPAVIALTDPEGCPLSRFKDSLGNPTHILRIGESYKEKQRTIQAKK EVEQRRAGGYSRKYASKAKNLADDMVRNTARDLLYYAVTQDAMLIFEN LSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLSKTYLSKTLAQYT SKTCSNCGFTITSADYDRVLEKLKKTATGWMTTINGKELKVEGQITYY NRYKRQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFS HRPVQEKFVCLNCGFETHADEQAALNIARSWLFLRSQEYKKYQTNKTT GNTDKRAFVETWQSFYRKKLKEVWKPAVTSPKKKRKV（SEQ ID NO: 281）

[0726] 实例6:RNA引导序列的产生

[0727] 为了产生RNA单引导序列及间隔区,通过用Q5聚合酶(NEB M0491)根据推荐方案,通过用于各主链的模板寡核苷酸及具有T7启动子及间隔区序列的扩增引物进行PCR来产生用于体外转录的模板。用于引导序列及间隔区的T7启动子、引导序列及间隔区的DNA引物序列呈现于下表11中。对于各支架标记为“正向主链”及“反向主链”的模板寡核苷酸以各20nM的最终浓度包括在内,且扩增引物(T7启动子及独特间隔区引物)以各1μM的最终浓度包括在内。sg2、sg32、sg64及sg174引导序列分别对应于SEQ ID NO:5、2104、2106及2238,不同之处在于sg2、sg32及sg64经额外5’G修饰以提高转录效率(比较表11与表2中的序列)。7.37间隔区靶向β2-微球蛋白(B2M)。在PCR扩增之后,模板经清洁且通过酚-氯仿-异戊醇萃取分离,接着进行乙醇沉淀。

[0728] 在含有50mM Tris pH 8.0、30mM $MgCl_2$、0.01% Triton X-100、2mM亚精胺、20mM DTT、5mM NTP、0.5μM模板及100μg/mL T7 RNA聚合酶的缓冲液中进行体外转录。将反应物在37℃下培育过夜。每1mL转录体积添加20单位的DNA酶I(Promega#M6101))且培育一小时。RNA产物经由变性PAGE纯化、经乙醇沉淀且再悬浮于1×磷酸盐缓冲盐水中。为了折叠sgRNA,将样品加热至70℃后维持5分钟且接着冷却至室温。将反应物补充至1mM最终$MgCl_2$浓度,加热至50℃后维持5分钟且接着冷却至室温。将最终RNA引导序列产物存储于-80℃。

[0729] 表11:用于产生引导RNA的序列

[0730]

引物	引物序列	RNA 产物
T7 启动子引物	GAAATTAATACGACTCACTATA (SEQ ID NO: 282)	用于所有
sg2 正向主链	GAAATTAATACGACTCACTATAGGTACTGGCGCTTTTATCTCATTACTTTGAGAGCCATCACCAGCGACTATGTCGTATGGGTAAAG (SEQ ID NO: 283)	GGUACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAUGGGUAAAGCGCUUAUUUAUCGGAGAGAAAUCCGAUAAAUAAGAAGCAUCAAAGGGCCGAGAUGUCUCGCUCCG (SEQ ID NO: 295)
sg2 反向主链	CTTTGATGCTTCTTATTTATCGGATTTCTCTCCGATAAATAAGCGCTTTACCCATACGACATAGTCGCTGGTGATGGC (SEQ ID NO: 284)	
sg2.7.37 间隔区引物	CGGAGCGAGACATCTCGGCCCTTTGATGCTTCTTATTTATCGGATTTCTCTCCG (SEQ ID NO: 285)	
sg32 正向主链	GAAATTAATACGACTCACTATAGGTACTGGCGCTTTTATCTCATTACTTTGAGAGCCATCACCAGCGACTATGTCGTATGGGTAAAGCGC (SEQ ID NO: 286)	GGUACUGGCGCUUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAUGGGUAAAGCGCCCUCUUCGGAGGGAAGCAUCAAAGGGCCGAGAUGUCUCG (SEQ ID NO: 296)
sg32 反向主链	CTTTGATGCTTCCCTCCGAAGAGGGCGCTTTACCCATACGACATAG (SEQ ID NO: 287)	

[0731]

sg32.7.37 间隔区引物	CGGAGCGAGACATCTCGGCCCTTTGATGCTTCCCTCCGAAGAG (SEQ ID NO: 288)	
sg64 正向主链	GAAATTAATACGACTCACTATAGGTACTGGCGCCTTTTATCTCATTACTTTGAGAGCCATCACCAGCGACTATGTCGTATGGGTAAAGCGC (SEQ ID NO: 289)	GGUACUGGCGCCUUUAUCUCAUUACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAUGGGUAAAGCGCUUACGGACUUCGGUCCGUAAGAAGCAUCAAAGGGCCGAGAUGUCUCGCUCCG (SEQ ID NO: 297)
sg64 反向主链	CTTTGATGCTTCTTACGGACCGAAGTCCGTAAGCGCTTTACCCATACGACATAG (SEQ ID NO: 290)	
sg64.7.37 间隔区引物	CGGAGCGAGACATCTCGGCCCTTTGATGCTTCTTACGGACCGAAG (SEQ ID NO: 291)	
sg174 正向主链	GAAATTAATACGACTCACTATAACTGGCGCTTTTATCTGATTACTTTGAGAGCCATCACCAGCGACTATGTCGTAGTGGGTAAAGCT (SEQ ID NO: 292)	ACUGGCGCUUUUAUCUgAUUACUUUGAGAGCCAUCACCAGCGACUAUGUCGUAgUGGGUAAAGCUCCCUCUUCGGAGGGAGCAUCAAAGGGCCGAGAUGUCUCGCUCCG (SEQ ID NO: 298)
sg174 反向主链	CTTTGATGCTCCCTCCGAAGAGGGAGCTTTACCCACTACGACATAGTCGC(SEQ ID NO: 293)	
sg174.7.37 间隔区引物	CGGAGCGAGACATCTCGGCCCTTTGATGCTCCCTCC (SEQ ID NO: 294)	

[0732] 实例7:RNP组装

[0733] CasX及单引导RNA(sgRNA)的纯化野生型及RNP在即将进行实验之前制备,或经制备且在液氮中急冻且存储于-80℃以便后续使用。为了制备RNP复合物,将CasX蛋白与sgRNA以1:1.2摩尔比一起培育。简言之,将sgRNA添加至缓冲液#1(25mM NaPi、150mM NaCl、200mM海藻糖、1mM MgCl2)中,接着将CasX在涡旋下缓慢添加至sgRNA溶液中,且在37℃下培育10分钟以形成RNP复合物。RNP复合物在使用之前经由用200μl缓冲液#1预润湿的0.22μm Costar 8160过滤器过滤。必要时,RNP样品用0.5ml Ultra 100-Kd截止过滤器(Millipore零件号UFC510096)浓缩,直至获得所需体积。如实例13中所述评估潜能RNP的形成。

[0734] 实例8:评估对引导RNA的结合亲和力

[0735] 纯化野生型及改进CasX将在含有氯化镁以及肝素的低盐缓冲液中与含有3'Cy7.5部分的合成单引导RNA一起培育,以防止非特异性结合及聚集。sgRNA将维持于10pM的浓度,而蛋白质将在独立结合反应中自1pM滴定至100μM。在允许反应达到平衡之后,样品将穿过具有硝化纤维素膜及带正电尼龙膜的真空歧管过滤器-结合分析,所述膜分别结合蛋白质及核酸。膜将经成像以鉴别引导RNA,且将通过针对各蛋白质浓度在硝化纤维素相对于尼龙膜上的荧光的量来确定结合相对于未结合RNA的分率,以计算蛋白质-sgRNA复合物的解离常数。亦将通过sgRNA的改进变异体进行实验,以确定这些突变是否亦影响引导对于野生型及突变蛋白的亲和力。我们亦将进行电迁移率变动分析以与过滤器-结合分析定性比较,及确认可溶性结合而非聚集为蛋白质-RNA结合的主要贡献因素。

[0736] 实例9:评估对靶DNA的结合亲和力

[0737] 纯化野生型及改进CasX将与携有与靶核酸互补的靶向序列的单引导RNA复合。RNP复合物将与含有PAM及适当靶核酸序列(在靶链上具有5'Cy7.5标记)的双链靶DNA在含有氯化镁以及肝素的低盐缓冲液中一起培育,以防止非特异性结合及聚集。靶DNA将维持于1nM的浓度,而RNP将在独立结合反应中自1pM滴定至100μM。在允许反应达到平衡之后,样品将在天然5%聚丙烯酰胺凝胶上运行以分离结合及未结合靶DNA。凝胶将经成像以鉴别靶DNA的迁移率变动,且将对于各蛋白质浓度计算结合相对于未结合DNA的分率,以确定RNP-靶DNA三元复合物的解离常数。

[0738] 实例10:体外评估差异性PAM识别

[0739] 纯化野生型及工程改造CasX变异体将与携有固定靶向序列的单引导RNA复合。RNP复合物将以100nM的最终浓度添加至含有MgCl2的缓冲液,且以10nM的浓度与5'Cy7.5标记的双链靶DNA一起培育。将通过含有与靶核酸序列邻近的不同PAM的不同DNA底物进行独立反应。将在固定时间点获取反应物的等分试样且通过添加等体积的50mM EDTA及95%甲酰胺淬灭。样品将在变性聚丙烯酰胺凝胶上运行,以分离裂解及未裂解的DNA底物。将观测结果且将测定非典型PAM通过CasX变异体的裂解速率。

[0740] 实例11:评估针对双链裂解的核酸酶活性

[0741] 纯化的野生型及工程改造CasX变异体将与携有固定HRS靶向序列的单引导RNA复合。RNP复合物将以100nM的最终浓度添加至含有MgCl2的缓冲液且以10nM的浓度与在靶或非靶链上具有5'Cy7.5标记的双链靶DNA一起培育。将在固定时间点获取反应物的等分试样且通过添加等体积的50mMEDTA及95%甲酰胺淬灭。样品将在变性聚丙烯酰胺凝胶上运行,以分离裂解及未裂解的DNA底物。将观测结果且将确定靶及非靶链通过野生型及工程改造变异体的裂解速率。为了更明显地区分靶结合相对于核分解反应自身的催化速率的变化,

将在10nM至1μM范围内滴定蛋白质浓度,且将在各浓度下确定裂解速率以产生假-米曼氏拟合(pseudo-Michaelis-Mentenfit)且确定kcat*及KM*。KM*的变化指示改变的结合,而kcat*的变化指示改变的催化。

[0742] 实例12:评估针对裂解的靶链负载

[0743] 纯化的野生型及工程改造CasX 491将与携有固定HRS靶向序列的单引导RNA复合。RNP复合物将以100nM的最终浓度添加至含有MgCl2的缓冲液,且以10nM的浓度与在靶链上具有5'Cy7.5标记且在非靶链上具有5'Cy5标记的双链靶DNA一起培育。将在固定时间点获取反应物的等分试样且通过添加等体积的50mM EDTA及95%甲酰胺淬灭。样品将在变性聚丙烯酰胺凝胶上运行,以分离裂解及未裂解的DNA底物。将观测结果且将确定两个链通过变异体的裂解速率。靶链裂解但并非非靶链裂解的速率的变化将指示裂解活性位点中靶链负载的改进。此活性可通过用在非靶链上具有间隙的dsDNA底物(模拟预裂解底物)重复所述分析而进一步分离。在此情形下改进的非靶链裂解将进一步证明靶链的负载及裂解。

[0744] 实例13:CasX:gNA体外裂解分析

[0745] 1.与野生型参考CasX相比,确定蛋白质变异体的裂解潜能分率

[0746] 使用体外裂解分析确定相比于参考CasX,CasX变异体形成活性RNP的能力。如下产生用于裂解分析的β-2微球蛋白(B2M)7.37靶标。具有序列TGAAGCTGACAGCATTCGGGCCGAGATGTCTCGCTCCGTGGCCTTAGCTGTGCTCGCGC T(非靶链,NTS(SEQ ID NO:299))及TGAAGCTGACAGCATTCGGGCCGAGATGTCTC GCTCCGTGGCCTTAGCTGTGCTCGCGCT(靶链,TS(SEQ ID NO:300))的DNA寡核苷酸与5'荧光标记(分别为LI-COR IRDye 700和800)一起购买。如下地形成dsDNA靶标:通过在1×裂解缓冲液(20mM Tris HCl pH 7.5,150mM NaCl,1mM TCEP,5%甘油,10mM $MgCl_2$)中以1:1比率混合寡核苷酸,加热至95℃后保持10分钟,且使溶液冷却至室温。

[0747] CasX RNP于37℃下使用在1×裂解缓冲液(20mM Tris HCl pH 7.5,150mM NaCl,1mM TCEP,5%甘油,10mM $MgCl_{2)}$中1μM最终浓度的指示CasX及引导序列(参见图表)复原10分钟(其中除非另外指明,否则指示引导序列为1.5倍过量),随后移至冰上直至准备使用。使用7.37靶标,以及具有与7.37靶标互补的间隔区的sgRNA。

[0748] 制备最终RNP浓度为100nM且最终靶浓度为100nM的裂解反应物。在37℃下进行反应且通过添加7.37靶DNA起始。在5、10、30、60及120分钟处获取等分试样且通过添加至95%甲酰胺,20mM EDTA中淬灭。样品通过在95℃下加热10分钟变性,且在10%脲-PAGE凝胶上运行。使用LI-COR Odyssey CLx对凝胶进行成像并使用LI-COR Image Studio软件进行定量,或者施用Cytiva Typhoon对凝胶进行成像并使用Cytiva IQTL软件进行定量。使用Prism绘制及分析所得数据。我们假设CasX在分析条件下基本上以单周转酶形式起作用,如由以下观测结果指示:亚化学计算量的酶即使在扩展时间标度下亦无法裂解大于化学计算量的靶标,且替代地接近随着存在的酶的量缩放的平稳段。因此,靶标在长时间标度内通过等摩尔量的RNP裂解的分率指示RNP的何种分率为恰当形成的且对于裂解具活性。用双相速率模型拟合裂解迹线,因为裂解反应在此浓度范围内明显偏离单相,且对于三个独立重复样中的每一个确定平稳段。计算平均值及标准差以确定活性分率(表12)。图式在图9中示出。

[0749] 对于针对CasX2+引导序列174+7.37间隔区、CasX119+引导序列174+7.37间隔区、CasX457+引导序列174+7.37间隔区、CasX488+引导序列174+7.37间隔区及CasX491+引导序列174+7.37间隔区形成的RNP确定表观活性(潜能)分率。确定的活性分率展示于表12中。所

有CasX变异体均具有高于野生型CasX2的活性分率，表明相比于野生型CasX，工程改造CasX变异体在测试条件下与相同引导序列形成显著更具活性且稳定的RNP。这可归因于对sgRNA增加的亲和力、在sgRNA存在下增加的稳定性或溶解度、或工程改造CasX:sgRNA复合物的裂解潜能构象的更大稳定性。与CasX2相比，向sgRNA中添加CasX457、CasX488或CasX491时，观察到的沉淀物显著减少，表明RNP的溶解度增加。

[0750] 2.体外裂解分析-确定CasX变异体相比于野生型参考CasX的$k_{裂解}$

[0751] 对于CasX2.2.7.37、CasX2.32.7.37、CasX2.64.7.37及CasX2.174.7.37，也使用相同的方案测定裂解潜能分率为16±3%、13±3%、5±2%及22±5%，如图10和表12中所示。

[0752] 第二组引导序列在不同条件下进行了测试，以更好地隔离引导序列对RNP形成的贡献。将174、175、185、186、196、214和215个带有7.37间隔区的引导序列与CasX491混合，最终浓度为1μM的引导序列和1.5μM的蛋白质，而不是像以前那样使用过量的引导序列。结果示于图11和表12。与174相比，这些引导序列中的许多都表现出额外的改进，其中185和196分别实现了91±4%和91±1%的潜能分率，而在这些引导序列限制条件下，174为80±9%。

[0753] 数据指示相比于野生型CasX及野生型sgRNA，CasX变异体及sgRNA变异体均能够通过引导RNA形成较高程度的活性RNP。

[0754] 与野生型参考CasX相比，CasX变异体119、457、488和491的表观裂解速率是通过体外荧光分析来测定的，用于裂解靶7.37。

[0755] CasX RNP于37℃下使用在1×裂解缓冲液(20mM Tris HCl pH 7.5，150mM NaCl，1mM TCEP，5%甘油，10mM $MgCl_2$)中1μM最终浓度的指示CasX(参见图12)和1.5倍过量的指示引导序列复原10分钟，随后移至冰上直至准备使用。以200nM的最终RNP浓度及10nM的最终靶浓度建立裂解反应。在37℃下进行反应且通过添加靶DNA起始。在0.25、0.5、1、2、5及10分钟处获取等分试样且通过添加至95%甲酰胺，20mM EDTA中淬灭。样品通过在95℃下加热10分钟变性，且在10%脲-PAGE凝胶上运行。凝胶用LI-COR Odyssey CLx成像，并使用LI-COR Image Studio软件进行定量，或用Cytiva Typhone成像，并使用Cytiva IQTL软件进行定量。使用Prism绘制及分析所得数据，且针对各CasX:sgRNA组合重复样个别地确定非靶链裂解的表观一阶速率常数($k_{裂解}$)。具有独立拟合的三个重复样的平均值和标准差如表12所示，且裂解迹线如图12所示。

[0756] 测定野生型CasX2和CasX变异体119、457、488和491的表观裂解速率常数，每个分析中使用引导序列174和间隔区7.37(见表12和图12)。相对于野生型CasX2，所有CasX变异体都提高了裂解速率。CasX457的裂解速度比119慢，尽管如上所确定具有更高的潜能分率。CasX488和CasX491具有较大幅度的最高裂解速率；由于靶标在第一个时间点几乎完全被裂解，真正的裂解速率超过了所述分析的分辨率，报道的$k_{裂解}$应作为下限。

[0757] 数据表明，相比于野生型CasX2，CasX变异体具有较高活性水准，其中$k_{裂解}$速率至少高出30倍。

[0758] 3.体外裂解测定：比较引导变异体与野生型引导序列

[0759] 亦通过野生型参考CasX2及参考引导序列2相比于引导变异体32、64及174进行裂解测定，以确定变异体是否改进裂解。如上文所述地进行实验。由于许多所得RNP在测试时间内未接近靶标的完全裂解，我们确定初始反应速度(V_0)而非一阶速率常数。前两个时间点(15及30秒)与各CasX:sgRNA组合及重复样的线拟合。确定三个重复样的斜率的平均值及

标准差(图13)。

[0760] 在分析条件下,CasX2在引导序列2、32、64及174的情况下的V_0为20.4±1.4nM/min、18.4±2.4nM/min、7.8±1.8nM/min及49.3±1.4nM/min(参见表12及图13和图14)。引导序列174表明,所得RNP的裂解速率大幅改进(相对于2为约2.5倍,参见图14),而引导序列32及64的表现与引导序列2类似或比其更差。值得注意的是,引导序列64支持比引导序列2更低的裂解速率,但在体内的表现好得多(数据未示出)。产生引导序列64的一些序列改变可能以参与三螺旋体形成的核苷酸为代价来改进体内转录。引导序列64改进的表达可能解释其改进的体内活性,而其降低的稳定性可导致不当体外折叠。

[0761] 使用具有间隔区7.37的引导序列174、175、185、186、196、214和215和CasX491进行了额外的实验,以确定相对裂解速率。为了将裂解动力学降低到用我们的分析可测量的范围,裂解反应在10℃下培育。结果在图15和表12中。在这些条件下,215是唯一支持比174更快的裂解速率的引导序列。196在引导限制条件下表现出最高的RNP活性分率,其动力学与174基本相同,再次强调了不同的变异体导致不同特征的改进。

[0762] 这些数据支持,在所述分析的条件下,使用带有CasX的大多数引导变异体导致RNP的活性水平高于使用野生型引导变异体的RNP,初始裂解速度的改善范围为约2倍至>6倍。表12中的数值从左到右表示RNP构建体的CasX变异体、sgRNA支架和间隔区序列。在下表的RNP构建体名称中,从左至右表示CasX蛋白变异体、引导支架和间隔区。

[0763] 表12:裂解及RNP形成分析的结果

[0764]

RNP 构建体	$k_{裂解}$*	初始速度*	潜能分率
2.2.7.37		20.4 ± 1.4 nM/min	16 ± 3%
2.32.7.37		18.4 ± 2.4 nM/min	13 ± 3%
2.64.7.37		7.8 ± 1.8 nM/min	5 ± 2%
2.174.7.37	0.51 ± 0.01 min^{-1}	49.3 ± 1.4 nM/min	22 ± 5%
119.174.7.37	6.29 ± 2.11 min^{-1}		35 ± 6%
457.174.7.37	3.01 ± 0.90 min^{-1}		53 ± 7%
488.174.7.37	15.19 min^{-1}		67%
491.174.7.37	16.59 min^{-1} / 0.293 min^{-1} (10℃)		83% / 17%(引导序列限制)
491.175.7.37	0.089 min^{-1}(10℃)		5%(引导序列限制)
491.185.7.37	0.227 min^{-1}(10℃)		44%(引导序列限制)
491.186.7.37	0.099 min^{-1}(10℃)		11%(引导序列限制)
491.196.7.37	0.292 min^{-1}(10℃)		46%(引导序列限制)
491.214.7.37	0.284 min^{-1}(10℃)		30%(引导序列限制)
491.215.7.37	0.398 min^{-1}(10℃)		38%(引导序列限制)

[0765] *平均值及标准差

[0766] 实例14:体外评估差异性PAM识别

[0767] 基本上如实例13中所述,使用与sg174.7.37复合的CasX2、CasX119和CasX438进行体外裂解分析。使用带有7.37间隔区和TTC、CTC、GTC或ATC PAM的荧光标记的dsDNA靶标(序列在表13中)。在0.25、0.5、1、2、5、10、30及60分钟处取时间点。凝胶通过CytivaTyphoon成像且使用IQTL8.2软件定量。针对各靶标上的各Casx:sgRNA复合物确定非靶链裂解的表观一阶速率常量($k_{裂解}$)。将非TTC PAM靶标的速率常数与TTC PAM靶标的速率常数进行比较,以确定在给定的蛋白质变异体中是否改变了对每种PAM的相对偏好。

[0768] 对于所有变异体，TTC靶标支持最高的裂解速率，其次是ATC，然后是CTC，最后是GTC靶标(图16A-16D，表14)。对于CasX变异体和NTC PAM的每种组合，显示了裂解速率$k_{裂解}$。对于所有非NTC PAM，相比于所述变异体的TTC速率的相对裂解速率在圆括号中示出。所有非TTC PAM显示出显著降低的裂解速率(所有>10倍)。特定变异体的非TTC PAM和TTC PAM的裂解速率之比在所有变异体中保持一致。相对于TTC靶标的速度，CTC靶标以3.5-4.3％支持裂解；GTC靶标以1.0-1.4％支持裂解；且ATC靶标以6.5-8.3％支持裂解。例外情况是491，其中TTC PAM的裂解动力学太快，无法进行精确测量，这人为地减小了TTC PAM与非TTC PAM之间的表观差异。比较GTC、CTC和ATC PAM上491的相对比率(这些比率在可测量范围内)，得出的比率与跨非ATC PAM进行比较时的其它变异体的比率相当，与比率协同地增加一致。总体而言，变异体之间的差异不足以表明对各种NTC PAM的相对偏好已经改变。然而，变异体较高的基础裂解速率允许ATC或CTC PAM的靶标在10分钟内几乎完全被裂解，表观$k_{裂解}$与TTC PAM上CasX2的$k_{裂解}$相当或更大(表14)。这种增加的裂解速率可能会超过在人类细胞中进行有效基因组编辑所必需的阈值，这解释了这些变异体的PAM灵活性明显增加。

[0769] 表13.用于体外PAM裂解测定的DNA底物序列。

[0770]

引导序列*	DNA 序列	SEQ ID NO
7.37 TTC PAM TS	AGCGCGAGCACAGCTAAGGCCACGGAGCGAGACATCTCGGCCC**G AA**TGCTGTCAGCTTCA	301
7.37 TTC PAM NTS	TGAAGCTGACAGCA**TTC**GGGCCGAGATGTCTCGCTCCGTGGCCT TAGCTGTGCTCGCGCT	302
7.37 CTC PAM TS	AGCGCGAGCACAGCTAAGGCCACGGAGCGAGACATCTCGGCCC**G AG**TGCTGTCAGCTTCA	303
7.37 CTC PAM NTS	TGAAGCTGACAGCA**CTC**GGGCCGAGATGTCTCGCTCCGTGGCCT TAGCTGTGCTCGCGCT	304
7.37 GTC PAM TS	AGCGCGAGCACAGCTAAGGCCACGGAGCGAGACATCTCGGCCC**G AC**TGCTGTCAGCTTCA	305
7.37 GTC PAM NTS	TGAAGCTGACAGCA**GTC**GGGCCGAGATGTCTCGCTCCGTGGCCT TAGCTGTGCTCGCGCT	305
7.37 ATC PAM TS	AGCGCGAGCACAGCTAAGGCCACGGAGCGAGACATCTCGGCCC**G AT**TGCTGTCAGCTTCA	307
7.37 ATC PAM NTS	TGAAGCTGACAGCA**ATC**GGGCCGAGATGTCTCGCTCCGTGGCCT TAGCTGTGCTCGCGCT	308

[0771] *每个序列的PAM序列都用粗体显示。TS-靶链。NTS-非靶链。

[0772] 表14.CasX变异体对NTC PAM的表观裂解率。

[0773]

变异体	TTC	CTC	GTC	ATC
2	0.267 min^{-1}	9.29E-3 min^{-1} (0.035)	3.75E-3 min^{-1} (0.014)	1.87E-2 min^{-1} (0.070)
119	8.33 min^{-1}	0.303 min^{-1} (0.036)	8.64E-2 min^{-1} (0.010)	0.540 min^{-1} (0.065)

[0774]

438	4.94 min^{-1}	0.212 min^{-1} (0.043)	1.31E-2 min^{-1} (0.013)	0.408 min^{-1} (0.083)
491	16.42 min^{-1}	8.605 min^{-1} (0.524)	2.447 min^{-1} (0.149)	11.33 min^{-1} (0.690)

[0775] 实例15：切口变异体的识别

[0776] 纯化的修饰CasX变异体将与带有固定靶向序列的单引导RNA复合。RNP复合物将以100nM的最终浓度添加至含有$MgCl_2$的缓冲液，且以10nM的浓度与在靶链上具有5'荧光素标

记且在非靶链上具有5'Cy5标记的双链靶DNA一起培育。将在固定时间点获取反应物的等分试样且通过添加等体积的50mM EDTA及95％甲酰胺淬灭。样品将在变性聚丙烯酰胺凝胶上运行,以分离裂解及未裂解的DNA底物。一条链而不是另一条链的有效裂解表明所述变异体具有单链切口酶活性。

[0777] 实例16:评估用于RNP生产的CasX变异体的改进表达和溶解度特征

[0778] 野生型和修饰的CasX变异体将在相同条件下在BL21(DE3)大肠杆菌中表达。所有蛋白质都将处于IPTG诱导型T7启动子的控制之下。细胞将在37℃的TB培养基中生长至OD为0.6,此时生长温度将降低至16℃,并通过添加0.5mM IPTG诱导表达。细胞将在表达18小时后收获。将提取可溶性蛋白质级分并在SDS-PAGE凝胶上进行分析。可溶性CasX表达的相对水平将通过考马斯染色来鉴定。蛋白质将根据上述方案平行纯化,并比较纯蛋白质的最终产量。为了确定纯化蛋白质的溶解度,构建体将在储存缓冲液中浓缩,直到蛋白质开始沉淀。将通过离心去除沉淀的蛋白质,并测量可溶性蛋白质的最终浓度以确定每种变异体的最大溶解度。最后,CasX变异体将与单引导RNA复合并浓缩,直到开始沉淀。沉淀的RNP将通过离心去除,并测量可溶性RNP的最终浓度以确定每个变异体与引导RNA结合时的最大溶解度。

[0779] 实例17:C9orf72的CasX:gNA编辑

[0780] 此实例阐述了用以制造和测试能够修饰C9orf72基因座的组合物的参数。

[0781] 实验设计:

[0782] A)修饰C9orf72的间隔区的选择过程:

[0783] 20bp XTC PAM间隔区将被设计成靶向人类基因组中的以下区:

[0784] (a)C9orf72顺式增强子元件

[0785] (b)在脊椎动物中高度保守的C9orf72近端非编码遗传元件(UCSC基因组浏览器)

[0786] (c)C9orf72基因组基因座。C9orf72基因被定义为跨越染色体9上人类基因组的chr9:27,546,546-27,573,866(智人更新注释发布109.20191205,GRCh38.p13(NCBI))的序列。人类C9orf72基因部分地在NCBI数据库(ncbi.nlm.nih.gov)中描述为参考序列NC_000009.12,所述参考序列通过引用并入本文。靶向C9orf72的间隔区可以类似地由其它基因组组装。

[0787] B)产生靶向C9orf72的构建体的方法:

[0788] 为了产生靶向C9orf72的构建体,将靶向C9orf72的间隔区克隆到基础哺乳动物表达质粒构建体(pStX)中,所述构建体由以下组分构成:经密码子优化的CasX(构建体CasX 491分子和rRNA引导序列174(491.174);序列见表)+NLS;和哺乳动物选择标记物嘌呤霉素。间隔区序列DNA将以由间隔区序列和所述序列的反向互补序列组成的单链DNA(ssDNA)寡核苷酸形式从Integrated DNA Technologies(IDT)订购。将这两个寡核苷酸退火在一起,并使用T4 DNA连接酶和对于质粒适当的限制酶,通过Golden Gate组装个别或批量克隆到pStX中。将组装产物转化到化学或电感受态细菌细胞中,涂铺于含有羧苄青霉素的LB-琼脂板上,并培育直至出现菌落。将挑取个别菌落且使用Qiagen Qiaprep旋转小规模制备试剂盒(Qiagen目录号27104)根据制造商的方案进行小规模制备。将使用Sanger测序法对所得质粒进行测序以确保正确连接。SaCas9和SpyCas9对照质粒(具有基于Cas蛋白特异性PAM选择的间隔区)将以与上述pStX质粒类似的方式制备。

[0789] C)产生C9orf72报道系的方法：

[0790] 在HEPG2细胞系中，荧光编码DNA(例如，GFP)将被敲入在最后一个C9orf72外显子的3'端。经修饰细胞将每3-5天通过连续继代扩增，且维持于由以下各者组成的成纤维细胞(FB)培养基中：达尔伯克氏改进伊格尔培养基(DMEM;Corning Cellgro,#10-013-CV)，补充有10％胎牛血清(FBS;Seradigm,#1500-500)，或其它适当培养基，及100单位/毫升青霉素及100mg/ml链霉素(100×-青霉素-链霉素;GIBCO#15140-122)，且可另外包括丙酮酸钠(100×,Thermofisher#11360070)、非必需氨基酸(100×Thermofisher#11140050)、HEPES缓冲液(100×Thermofisher#15630080)及2-巯基乙醇(1000×Thermofisher#21985023)。将在37℃和5％ CO2下培育细胞。1-2周后，将单个GFP+细胞分选到FB或其它合适的培养基中。报道系克隆将通过每3-5天连续继代扩增且在37℃及5％ CO2下维持于培育箱中的FB培养基中。将通过基因组测序以及使用靶向C9orf72的分子对C9orf72基因座的功能修饰来表征这些细胞系。最佳报道系将被鉴别为如下细胞系：i)具有正确整合在靶C9orf72基因座的单一GFP副本，ii)保持与未修饰细胞相当的倍增时间，iii)当使用下述方法分析时，C9orf72基因破坏后导致GFP荧光减少。

[0791] D)评估C9orf72-GFP报道细胞系中的C9orf72修饰活性的方法：

[0792] C9orf72报道细胞将以20-40k个细胞/孔接种在96孔板的100μl FB(或其它适当)培养基中，且在具有5％ CO2的37℃培育箱中培养。第二天，将检查经接种细胞的汇合度。理想情况下，转染时细胞应该达到约75％的汇合度。如果细胞将处于适当的汇合度，则将进行转染。

[0793] 按照制造商的方案，使用Lipofectamine 3000以每孔100-500ng转染具有靶向C9orf72的适当间隔区的每个CasX构建体(CasX 491和引导序列174，序列见表)，每个构建体使用3个孔作为重复样。靶向C9orf72的SaCas9和SpyCas9将用作基准对照。对于各Cas蛋白类型，非靶向质粒将用作阴性对照。

[0794] 在以0.3-3μg/ml进行嘌呤霉素选择24-48小时以选择成功转染的细胞，接着在FB或其它合适的培养基中恢复24-48小时之后，经由流式细胞术分析经转染细胞中的荧光。在此方法中，细胞针对适当正向及侧向散射进行设门，针对单细胞进行选择且接着针对报道子表达进行设门(Attune Nxt Flow Cytometer,Thermo Fisher Scientific)，以定量荧光团的表达水平。将对各样品收集至少10,000个事件。然后将数据用于计算抗体标记阴性(编辑)细胞的百分比。

[0795] 将裂解来自实例的每个样品的细胞亚群，并按照制造商的方案使用快速提取(Quick extract)溶液提取基因组。将使用T7E1分析来分析编辑。简而言之，将在热循环仪上使用PCR程序，使用引物(例如，预期靶标周围500bp的区)扩增靶向编辑位点的基因组基因座。随后PCR扩增子将在热循环仪上按照杂交程序进行杂交，且随后用T7核酸内切酶在37℃下处理30分钟。然后将在2％琼脂糖凝胶上或在片段分析仪(Fragment Analyzer)上分析样品以观察DNA条带。

[0796] 实例18：评估HEK293T细胞中的C9orf72六核苷酸重复序列扩增段(HRE)修饰活性的方法

[0797] HEK293T(加州大学伯克利分校细胞培养设施(Cell Culture Facility,UC Berkeley))以30k个细胞/孔接种在96孔板的100μl FB培养基中，且在具有5％ CO2的37℃

培育箱中培养。第二天,将检查经接种细胞的汇合度。在转染时达至少约75%汇合度的细胞用于转染。

[0798] 将编码具有靶向C9orf72 HRE区域5'到3'的序列的引导序列174和适当间隔区(表15,其中基因座中靶向位置在图22中示意性地展示)的CasX构建体491的质粒p59.491,174,29.X按照制造商的方案,使用Lipofectamine 3000以每孔100ng脂转染,并且每个构建体置于3个孔中作为重复样。非靶向质粒用作阴性对照。使用1-3μg/ml的嘌呤霉素选择来选择经成功转染的细胞。在4天之后,收获样品用于gDNA提取并且扩增以用于NGS分析。

[0799] 结果:

[0800] 对于单一切割,编辑百分比展示于表15中,且单一间隔区的结果展示于图23中。多种PAM(即,ATC、GTC和TTC)观察到有效编辑,其中TTC中见到的编辑效率最高。

[0801] 对于HRE区的双重切割和漏失(drop-out),间隔区的不同组合展现有效编辑以及介入HRE序列的对应缺失,平均大约36%(表16)。图24中展示编辑的代表图。

[0802] 结果表明,在实验条件下,可以利用具有单一引导序列的CasX系统来直接编辑顺式调节元件以及HRE,同时可以利用两个引导序列来成功地切除HRE区。

[0803] 表15:C9orf72间隔区和编辑百分比

[0804]

间隔区#	间隔区序列	编辑%	SEQ ID NO
137	CUUGUCCCUGCGCCGCCGCC	32	309
138	CCACCGCCUGCGCCUCCGCC	-	310
140	CCUCUAGCGACUGGUGGAAU	-	311
141	ACCAGUCGCUAGAGGCGAAA	-	312
142	CUCUCAUUUCUCUGACCGAA	-	313
143	ACCCGCGCCUCUUCCCGGCA	51.86	314
144	GCUGCCGGGAAGAGGCGCGG	48.05	315
145	CGGCAGCCGAACCCCAAACA	52.8	316
146	UGGCGGGUGGCUGUUUGGGG	56.23	317
147	UCACUCACCCACUCGCCACC	49.93	318
148	GCCGCCGCGGGCGCAGGCAC	36.45	319
149	GGGCGGGCCCGGGGGCGGGC	20.76	320
151	GGAGUCGCGCGCUAGGGGCC	53.12	321
153	UGAGUUCCAGAGCUUGCUAC	41.34	322
154	AGAGCUUGCUACAGGCUGCG	50.55	323
156	UCUGGUUAAUCUUUAUCAGG	44.25	324
158	UGUUCACCCUCAGCGAGUAC	59.26	325
159	CCCUCAGCGAGUACUGUGAG	50.88	326
161	UAAACCCACACCUGCUCUUG	27.45	327
162	UGCUAGACCCCGCCCCCAAA	-	328
163	CUUUUGGGGGCGGGGUCUAG	11.03	329
164	GCCCACGUAAAAGAUGACGC	65.42	330

165	UUUACGUGGGCGGAACUUGU	12.67	331
166	CUUUCCUAGCGGGACACCGU	24.08	332
167	UUCCUAGCGGGACACCGUAG	27.93	333
168	UAGCGGGACACCGUAGGUUA	54.33	334
170	AUGUGCGAUGACGUUUUCUC	-	335
171	CUAGCCUCGUGAGAAAACGU	-	336
172	CACGAGGCUAGCGAAAUGGG	-	337
173	UUUAUCUUAAGACCCGCUCU	-	338
175	UCCAGAGCGGGUCUUAAGAU	-	339
178	UAAGACCCGCUCUGGAGGAG	-	340
180	CCAGAUCCAGCAGCCUCCCC	-	341
181	AAACCACGAAAUCGUCUUCA	42.86	342
183	UCCACCAAUUAUGUGUUGAG	1.2	343

[0805]

[0806] 表16:用2个间隔区进行双重切割的删除编辑

[0807]

间隔区组合	删除编辑%
138/151	45.43
138/153	43.88
138/154	33.31
138/156	33.87
148/151	32.98
148/153	32.94
148/154	26.38
148/156	34.7
148/158	40.54
149/151	37.94
149/153	31.69
149/154	27.83
149/156	37.81
149/158	40.15

[0808] 实例19:将靶向C9orf72的CasX构建体封装于慢病毒载体中的方法

[0809] 封装靶向C9orf72的CasX:gNA构建体(例如CasX 491和引导序列174)的靶向C9orf72的慢病毒粒子将通过以下产生:使用编码CasX、引导RNA的转基因质粒、慢病毒封装质粒和VSV-G包膜质粒的基于聚乙烯亚胺的转染,转染70%到90%汇合度的HEK293。对于慢病毒粒子产生,将在转染后12小时更换培养基,并且将在转染后36到48小时收获病毒。病毒上清液将使用0.45μm膜滤器过滤,并在适当情况下在FB培养基(成纤维细胞培养基,其由以下成分构成:含Glutamax的DMEM(Gibco 10566-016),补充有MEM-NEAA(Thermo11140050)、丙酮酸钠(Thermo 11360070)、HEPES(Thermo 15630080)、2-巯基乙醇(Gibco21985023)、青霉素/链霉素(Thermo 15140122)和10%体积分率的胎牛血清(FBS,VWR#97068-085))中稀释。

[0810] 实例20:通过慢病毒筛选评估C9orf72修饰的方法

[0811] 将慢病毒质粒如上文所描述且根据标准克隆程序克隆,使得每个慢病毒质粒具有一个靶向C9orf72的间隔区-引导序列支架,和一个携带CasX分子的密码子优化的NLS(例如具有嘌呤霉素选择标记物的CasX 491分子及rRNA引导序列174构建体(491.174)。进行克隆,使得最终滴度涵盖全部文库大小的>100倍的靶向所有已知PAM的所有可能的C9orf72间隔区以及它们在C9orf72基因中的对应间隔区。如果文库大小为约5,000,则被评估的文库将>$5x10^5$。

[0812] 通过使用包含间隔区文库的质粒、慢病毒包装质粒和VSV-G包膜质粒的基于聚乙烯亚胺的转染来转染70%-90%汇合度的HEK293T,从而产生慢病毒粒子。为了产生粒子,在转染后12小时更换培养基,并在转染后36-48小时收获病毒。

[0813] 病毒上清液使用0.45μm膜滤器过滤,在适当情况下在FB培养基中稀释,并添加到靶细胞(在这种情况下为C9orf72-GFP报道细胞系)中。如果需要,以5-20μg/ml添加补充聚凝胺以增强转导效率。在转导后24-48小时在FB培养基中使用0.3-3μg/ml的嘌呤霉素选择经转导的细胞,并在具有5% CO_2的37℃培育箱中在FB或其它适当的培养基中生长7-10天。

[0814] 在SH-100或MA900 SONY分选仪上分选细胞。在这个过程中,细胞针对适当正向和侧向散射进行设门,针对单细胞进行选择且接着针对报道子表达进行设门。基于荧光水平来建立不同的细胞分选门(关=完全敲除,中等=部分破坏或敲减(KD),高=无编辑,极高=增强子),以区分和收集i)高功能C9orf72破坏分子、ii)仅降低表达的分子和iii)增加表达的分子编辑的细胞。如果在人类患者细胞中使用两种颜色,也可以运行所述分析来识别等位基因特异性引导序列。按照制造商的推荐方案,使用快速提取(Quick Extract)(Lucigen目录号QE09050)溶液从每组分选的细胞中收集基因组DNA。

[0815] 然后通过PCR直接从基因组中扩增来自每个收集库的间隔区文库,并收集用于在Miseq上进行深度测序。根据特定活性的门和丰度来进行间隔区分析;间隔区命中的NGS分析的详细方法见下文。

[0816] 然后,重新克隆从每个分选组中选择的引导序列,并通过流式细胞术和T7E1分析和/或蛋白质印迹,在报道细胞系和原代人细胞系中个别地验证其活性,并通过NGS分析来评估插入缺失谱。接下来的步骤可能类似于在评估报道细胞系中C9orf72修饰活性的方法中提供的描述。

[0817] 间隔区命中的NGS分析的方法

[0818] 此处提供关于如何分析来自上述慢病毒筛选的二代测序(NGS)数据的方法。使用二代测序(NGS)评估每个间隔区破坏C9orf72基因的能力。通过含有间隔区的慢病毒主链的特异性扩增来产生NGS文库。为每个分选的群体产生不同的文库(对应于低、中、高C9orf72表达的GFP高、中、低,等等),且随后用Illumina Hiseq进行评估。

[0819] 来自Illumina Hiseq的测序读段针对衔接子序列和低测序质量区进行了修剪。基于它们的重叠序列来合并成对的末端读段,以形成每个测序片段的单一共有序列。使用bowtie2将共有序列与设计的间隔区序列进行比对。与超过一个设计的间隔区序列比对的读段被丢弃。

[0820] 每个间隔区序列的"丰度"被定义为与所述序列比对的读段的数量。将每个测序文库的丰度制成表格,形成计数表,给出每个测序文库(即,分选群体)中每个间隔区序列的丰

度。最后,接着通过以下方式将丰度数归一化以考虑每个文库的不同测序深度:除以所述文库中的总读段计数,再乘以文库间的平均读段计数。归一化计数表用于确定每个门(高、中、低,等等)中每个间隔区的活性。

[0821] 通过在内源性人C9orf72基因座敲入GFP来构建C9orf72-GFP报道系。偶联至与gRNA间隔区互补的gRNA靶向序列的报道子(例如,GFP报道子)整合至报道细胞系中。细胞经CasX蛋白和/或sgRNA变异体转化或转染,其中sgRNA的间隔区基序与报道子的gRNA靶序列互补且靶向gRNA靶序列。通过FACS分析CasX:sgRNA核糖核蛋白复合物裂解靶核酸序列的能力。丧失报道子表达的细胞指示发生CasX:sgRNA核糖核蛋白复合物介导的裂解及插入缺失形成。报道系统基于成功修饰(编辑)C9orf72基因座后通过流式细胞术检测的GFP荧光减少。

[0822] 在初始筛选中,将测试两个gNA的C9orf72间隔区。将在报道细胞系中用CasX蛋白(CasX491与gNA 174的构建体)测试间隔区,使用SaCas9和SpyCas9作为对照。将在C9orf72-GFP报道细胞中评价GFP荧光的减少和编辑,使用嘌呤霉素选择成功进行脂转染的细胞并随后通过FACS分析GFP破坏。预计CasX 491和引导序列174可以编辑至少5-10%的细胞,证明CasX可以修饰内源性C9orf72基因座,并且比SaCas9和SpyCas9系统更有效。将进行T7E1分析或蛋白质印迹法以分析C9orf72-GFP报道细胞系中的基因编辑。CasX 491和具有靶向C9orf72的间隔区和非靶向对照物(NT)的引导序列174将被脂转染入C9orf72-GFP报道细胞,使用嘌呤霉素选择成功进行脂转染的细胞,并随后在T7E1分析中分析基因编辑,证明C9orf72基因座的成功编辑。

[0823] 实例21:采用慢病毒构建体使用CasX以等位基因特异性方式编辑C9orf72基因的方法

[0824] 所述实例展示CasX编辑C9orf72基因座的能力。永久治疗C9orf72相关疾病的一种策略是特异性破坏基因的突变副本,同时保留WT等位基因。具有两个野生型等位基因的HEK293细胞应可由WT CasX间隔区来编辑,但不能由突变CasX间隔区来编辑。所述实例将额外证明CasX间隔区区分相差单个核苷酸的中靶和脱靶等位基因的能力。HEK293细胞以20-40k个细胞/孔接种在96孔板的100μl FB培养基中,且在具有5% CO_2的37℃培育箱中培养。第二天,检查经接种细胞的汇合度,以确保细胞在转染时将达到约75%的汇合度。如果细胞处于适当的汇合度,则使用实例19的病毒上清液(具有CasX 491和引导序列174)进行转染,每个构建体使用3个孔作为重复样。靶向C9orf72的SaCas9和SpyCas9用作基准对照。对于各Cas蛋白类型,非靶向质粒用作阴性对照。将用0.3-3μg/ml的嘌呤霉素选择成功转染的细胞24-48小时,然后在FB培养基中恢复24-48小时。将裂解来自实验的每个样品的细胞亚群,并将按照制造商的方案使用快速提取(Quick extract)溶液提取基因组。将使用T7E1分析来分析编辑。简而言之,在热循环仪上使用PCR程序,使用引物(例如,预期靶标周围500bp的区)扩增靶向编辑位点的基因组基因座。随后PCR扩增子在热循环仪上按照杂交程序进行杂交,且随后用T7核酸内切酶在37℃下处理30分钟。然后在2%琼脂糖凝胶上或在片段分析仪(Fragment Analyzer)上分析样品以观察DNA条带。

[0825] 实例22:在源自常染色体显性C9orf72患者的负载六核苷酸重复序列扩增段的细胞系中证明等位基因特异性编辑的方法

[0826] 将获得来源于具有C9orf72中的HRS的患者的细胞并且在供应商建议的条件下培

养。将使用Lipofectamine 3000按照制造商的方案用CasX构建体(例如,CasX 491与间隔区174)转染细胞,或使用Lonza nucleofector试剂盒根据制造商的方案对细胞进行核转染,并接种在96孔板中进行培育和生长。或者,CasX构建体可以封装在慢病毒中,并且用于转导患者来源的细胞。将使用含有0.3-3μg/ml嘌呤霉素的培养基选择成功进行脂转染或核转染或慢病毒转导的细胞2-4天或更长时间,然后在不含嘌呤霉素的培养基中恢复2天或更长时间。C9orf72基因座的编辑可以在基因组、转录组和蛋白质组层级进行评估。在选择和恢复期结束时,将裂解来自实验的每个样品的细胞亚群,并且使用快速提取(QE)溶液按照制造商的方案提取基因组;将在RIPA细胞溶解缓冲液中裂解另一细胞亚群以用于蛋白质组分析;另一细胞亚群可以继代以用于在稍后时间点分析。一部分经QE处理的样品将用于使用T7E1分析来评估编辑。简而言之,将在热循环仪上使用PCR程序,使用引物(例如,预期靶标周围500bp的区)扩增靶向编辑位点的基因组基因座。随后PCR扩增子将在热循环仪上按照杂交程序进行杂交,且随后用T7核酸内切酶在37℃下处理30分钟。随后将在2%琼脂糖凝胶上或在片段分析仪(Fragment Analyzer)上分析样品以观察DNA条带,从而证实CasX构建体可以成功编辑C9orf72突变。另一部分经QE处理的样品将用于使用NGS来评估C9orf72基因座处的编辑。

[0827] 蛋白质组分析将通过蛋白质印迹进行。在RIPA缓冲液中裂解的样品将首先根据制造商的方案使用比色蛋白质定量分析(如BCA(Pierce)或Bradford(BioRad))对蛋白质含量进行定量。在定量之后,将使用β-巯基乙醇补充的Laemmli缓冲液中的稀释样品负载每孔2.5-20μg总蛋白质。样品将在95℃到100℃下热变性5-10分钟,且随后冷却到室温。然后样品将被上样到聚丙烯酰胺凝胶上并在其上运行。一旦凝胶已运行足够长时间,蛋白质将被转移到PVDF膜上,在室温下阻断至少1小时,并用抗C9orf72的初级抗体和适当的内参考物(loading control)进行标记。在室温下,在摇床上用PBST(补充有0.1v/v%Triton X100的PBS)洗涤印迹三次,每次洗涤五分钟。然后,在室温下,使用合适的与报道子结合的二抗来标记一抗1小时。在室温下,在摇床上用PBST(补充有0.1v/v%Triton X100的PBS)洗涤印迹三次,每次洗涤五分钟。随后将添加任何必要的底物,根据需要进行淬灭,并在凝胶成像仪上成像。将使用适当的软件按照制造商的方案对条带强度进行定量。

[0828] 实例23:通过AAV递送靶向C9orf72的构建体的方法:用编码的CasX系统制造和回收AAV

[0829] 本实例描述产生及表征封装编码CasX分子及引导RNA的序列的AAV载体所遵循的典型方案。

[0830] 材料及方法:

[0831] 对于AAV生产,使用三质粒转染方法,并且需要三种必需质粒:携带待包装于AAV中的靶向所关注C9orf72基因的CasX:gRNA的pTransgene、pRC和pHelper。将编码CasX及引导RNA的DNA克隆至AAV转基因盒中的ITR之间,以产生pTransgene质粒,其示意图展示于图17中。构建的转基因质粒通过全长质粒测序、限制消化及功能测试(包括哺乳动物细胞的体外转染)来验证。AAV生产所需的额外质粒(pRC质粒及pHelper质粒)购自商业供应商(Aldevron,Takara)。

[0832] 对于AAV生产,HEK293细胞在具有5% CO_2的37℃培育箱中于FB培养基中培养。HEK293细胞的10-40个15cm培养皿用于单批病毒生产。对于单个15cm培养皿,将45-60μg质

粒以1:1:1摩尔比在4ml FB培养基中混合在一起,并在室温下与聚乙烯亚胺(PEI)(即以3μg PEI/μg DNA)复合10分钟(注意:所用三种质粒的比率可以变化以优化病毒生产)。接着将PEI-DNA复合物缓慢滴至HEK293细胞的15cm板上,且将经转染细胞的板移回至培育箱中。第二天,可将培养基换成含2% FBS的FB(而不是10% FBS;在适当情况下,成纤维细胞培养基由以下成分构成:含Glutamax的DMEM(Gibco 10566-016),补充有MEM-NEAA(Thermo 11140050)、丙酮酸钠(Thermo 11360070)、HEPES(Thermo 15630080)、2-巯基乙醇(Gibco 21985023)、青霉素/链霉素(Thermo 15140122)和10%体积分率的胎牛血清(FBS,VWR# 97068-085))。在质粒初始转染后48-120小时之间的任何时间,可从上清液、或从细胞集结粒、或从上清液和细胞集结粒的组合中收获AAV。

[0833] 如果在转染后72小时后收获病毒,可在此时收集来自细胞的培养基以增加病毒产率。在转染后2-5天,收集培养基和细胞(注意:收获的时机可以变化以优化病毒产率)。通过离心来集结细胞,并从顶部收集培养基。细胞在37℃下于具有高盐含量和高盐活性核酸酶的缓冲液中溶解1小时(注意:细胞还可以使用额外方法溶解,如连续冻融或清洁剂化学溶解)。收获时收集的培养基,及在更早时间点收集的任何培养基用含有40% PEG8000及2.5M NaCl的溶液的1:5稀释液处理,且在冰上培育2小时,以使AAV沉淀(注意:还可在4℃下进行过夜培育)。来自培养基的AAV沉淀物通过离心来集结,再悬浮于具有高盐活性核酸酶的高盐含量缓冲液中且与裂解的细胞集结粒合并。合并的细胞裂解物接着通过离心和通过0.45μm过滤器过滤来澄清,且在AAV Poros亲和力树脂柱(Thermofisher Scientific)上纯化。病毒从柱洗脱到中和溶液中(注意:在此阶段,病毒可进行额外轮次的纯化以提高病毒制剂质量)。洗脱的病毒接着通过qPCR滴定以定量病毒产率。对于滴定,病毒样品首先用DNA酶消化以去除任何非包装病毒DNA,DNA酶经去活化,且接着通过蛋白酶K进行病毒衣壳破坏以暴露包装的病毒基因组,以用于滴定。

[0834] 预计将从使用此处所述方法生产的一批病毒获得约$1x10^{12}$个病毒基因组。

[0835] 实例24:在具有HRS扩增段的小鼠模型中的C9orf72编辑的体内评价

[0836] C9-BAC小鼠模型(O'Rourke等人《C9orf72 BAC转基因小鼠显示ALS/FTD的典型病理性特征(C9orf72 BAC transgenic mice display typical pathologic features of ALS/FTD)》《神经元(Neuron)》88:892(2015))携带人类染色体9开放阅读框架72基因(C9orf72),在交替剪接的非编码第一外显子1a和1b之间的内含子中具有六核苷酸重复序列扩增段(GGGGCC),所述模型将用于评价使用作为RNP或由AAV载体递送的CasX:gNA系统的C9orf72基因编辑。

[0837] 方法:

[0838] 对于含有CasX和具有靶向C9orf72基因的间隔区的gNA或使用非靶向gNA作为阴性对照的AAV或RNP的注射,小鼠将被麻醉且置于啮齿动物立体定位设备上,随后将1-5μl体积中不同剂量的CasX:gNA的病毒粒子或RNP(用NLS或Lipofectamine2000配制用于细胞递送)注射到其侧脑室、海马区、纹状体、初级躯体感觉皮质(S1)和/或初级视觉皮质V1中的一个中。另一组动物将被鞘内注射病毒或RNP。使用例如旋转棒测试、握力强度测试、平衡木测试、足迹测试和旷场测试的分析监测各组小鼠的体重、存活以及行为和神经肌肉变化(Hao,Z.等人《表达聚-PR的C9orf72小鼠品系中的运动功能障碍和神经退化(Motor dysfunction and neurodegeneration in a C9orf72 mouse line expressing poly-PR)》《自然通讯

(Nat.Commun.)》10:2906(2019))。将以介于1到24个月范围内的预定间隔将额外组的小鼠安乐死,并经心脏灌注生理盐水。脑将从颅骨中取出,一个半球加工以进行组织学分析,另一半球解剖并快速冷冻以进行生化和遗传分析。类似地,脊髓将被解剖且快速冷冻或加工以用于组织学分析(颈/胸)。用于组织学的脑部和脊髓将在4%多聚甲醛中滴式固定(drop-fixed)24小时,随后转移到30%蔗糖中24到48小时,并且在液氮中冷冻以连续切片,接着使用适当的抗体或杂交探针分析/检查,以观察聚GP(DPR)和RNA转录物团簇。

[0839] 对于聚GP(DPR)定量,脑部和脊髓样品将在RIPA(50mM Tris,150mM NaCl,0.5% DOC,1% NP40,0.1% SDS和Complete™,pH 8.0)中均质化,接着离心且集结粒再悬浮于5M胍-HCl中。聚GP将使用多克隆抗体AB1358(Millipore Sigma)在使用聚GP标准作为对照的96孔形式分析中针对捕捉和检测抗体定量,或通过qPCR转录物分析定量。另外,RIPA匀浆将用于C9orf72蛋白质的表达,其中水平由蛋白质印迹法测定。简单来说,来自RIPA提取物的蛋白质将通过4-12% SDS-PAGE进行大小分级且转移到PVDF膜上。为了检测C9orf72,将使用小鼠单克隆抗C9orf72抗体GT779(加利福尼亚州尔湾(Irvine,CA)的Gene Tex),接着二级染料结合抗体进行膜免疫印迹。将使用Odyssey/Li-Cor成像系统进行可视化。

[0840] 实例25:编辑后的C9orf72 BAC杂合小鼠模型中的认知行为的体内评价

[0841] 将评价具有GGGGCC重复序列的C9orf72 BAC小鼠模型在使用实例17、18、22和23中所描述的CasX:gNA系统编辑C9orf72基因之后的认知测试中的改进。

[0842] 方法:

[0843] 在使用CasX和具有靶向C9orf72基因的间隔区的gNA或使用非靶向gNA作为阴性编辑对照来在小鼠中编辑C9orf72基因之后,加上相同背景下的普通未处理小鼠,将在注射后1个、2个和3个月使用认知测试来评价各组小鼠。所述测试将包括巴恩斯迷宫测试(Barnes maze test)、径向臂迷宫测试、埋珠测试和高架十字迷宫测试(Jiang,J.等人《通过反义寡核苷酸靶向含GGGGCC的RNA,来自C9ORF72中的ALS/FTD相关重复序列扩增段的毒性增加得到缓解(Gain of Toxicity from ALS/FTD-Linked Repeat Expansions in C9ORF72 Is Alleviated by Antisense Oligonucleotides Targeting GGGGCC-Containing RNAs)》《神经元》90:535(2016)。

[0844] 实例26:评价作为RNP递送时间隔区长度对细胞中编辑的影响

[0845] CasX变异体491如上文所描述经纯化。通过体外转录(IVT)制备具有支架174的引导RNA。IVT模板是通过PCR,根据建议的方案使用Q5聚合酶(NEB M0491)、用于每个支架主链的模板寡核苷酸,和具有T7启动子和全长(20个核苷酸)或对应间隔区3'端截短一个或两个核苷酸的15.3(CAAACAAATGTGTCACAAAG,SEQ ID NO:344)或15.5(GGAATAATGCTGTTGTTGAA,SEQ ID NO:345)间隔区(序列在表18中)的扩增引物产生。用于产生IVT模板的引物的序列展示于表17中。所得模板接着与T7 RNA聚合酶一起使用以根据标准方案产生RNA引导序列。引导序列使用变性聚丙烯酰胺凝胶电泳纯化并且在使用之前再折叠。通过在含有25mM磷酸钠缓冲剂(pH 7.25)、300mM NaCl、1mM MgCl2和200mM海藻糖的缓冲液中将蛋白质与1.2倍摩尔过量的引导序列混合来组装个别RNP。在37℃下培育RNP 10分钟,且随后通过尺寸排阻色谱纯化,并且更换到含有25mM磷酸钠缓冲剂(pH 7.25)、150mM NaCl、1mM MgCl2和200mM海藻糖的缓冲液(缓冲液1)中。在纯化后使用Pierce 660nm蛋白质分析法测定RNP的浓度。

[0846] 测试纯化的RNP在Jurkat细胞中的T细胞受体α(TCRα)基因座处的编辑。使用

Lonza4-D nucleofector系统通过电穿孔递送RNP。将700,000个细胞再悬浮于20μL Lonza缓冲液SE中,且将其添加到在缓冲液1中稀释到适当浓度且最终体积为2μL的RNP中。使用Lonza96孔穿梭系统使用方案CL-120对细胞进行电穿孔。在37℃下在预平衡的RPMI中回收细胞,且随后将每个电穿孔条件分成96孔板的三个孔。在核转染之后一天将细胞更换到新鲜RPMI中。在核转染后第三天,细胞用针对TCRα/β的Alexa Fluor 647标记的抗体(BioLegend)染色并且使用Attune Nxt流式细胞仪评估表面TCRα/β的损失。一部分Jurkat细胞在不存在编辑的情况下对于TCRα/β染色非阳性。为了考虑到这一点并且估计TCRα通过编辑敲除的细胞的实际百分比,应用式$TCR_{KO}=(TCR-_{观测}-TCR-_{阴性})/(1-TCR-_{阴性})$,其中$TCR_{KO}$为TCRα的估计敲除率,$TCR-_{观测}$为实验样品中TCR染色阴性的观测细胞分率,并且$TCR-_{阴性}$为无RNP对照样品中TCR染色阴性的细胞分率。此式假设表达和不表达TCRα/β的细胞以相同速率被编辑。TCRα敲除细胞的经校正分率使用Prism相对于RNP浓度标绘。对于每一间隔区,使用除EC50以外的共用参数将三个间隔区长度与剂量反应曲线拟合。所报道的p值是20nt间隔区的剂量曲线和比较截短间隔区的剂量曲线可以用相同EC50参数建模的概率。

[0847] 表17:用于产生IVT模板的寡核苷酸

[0848]

构建体	DNA 序列	SEQ ID NO
支架 174 正向模板	GAAATTAATACGACTCACTATAACTGGCGCTTTTATC TGATTACTTTGAGAGCCATCACCAGCGACTATGTCGT AGTGGGTAAAGCT	346
支架 174 反向模板	CTTTGATGCTCCCTCCGAAGAGGGAGCTTTACCCACT ACGACATAGTCGC	347
T7 扩增引物	GAAATTAATACGACTCACTATA	348
15.3 20-nt 间隔区引物	CTTTGTGACACATTTGTTTGCTTTGATGCTCCCTCC	349
15.3 19-nt 间隔区引物	TTTGTGACACATTTGTTTGCTTTGATGCTCCCTCC	350
15.3 18-nt 间隔区引物	TTGTGACACATTTGTTTGCTTTGATGCTCCCTCC	351
15.5 20-nt 间隔区引物	TTCAACAACAGCATTATTCCCTTTGATGCTCCCTCC	352
15.5 19-nt 间隔区引物	TCAACAACAGCATTATTCCCTTTGATGCTCCCTCC	353
15.5 18-nt 间隔区引物	CAACAACAGCATTATTCCCTTTGATGCTCCCTCC	354

[0849] 表18:间隔区序列

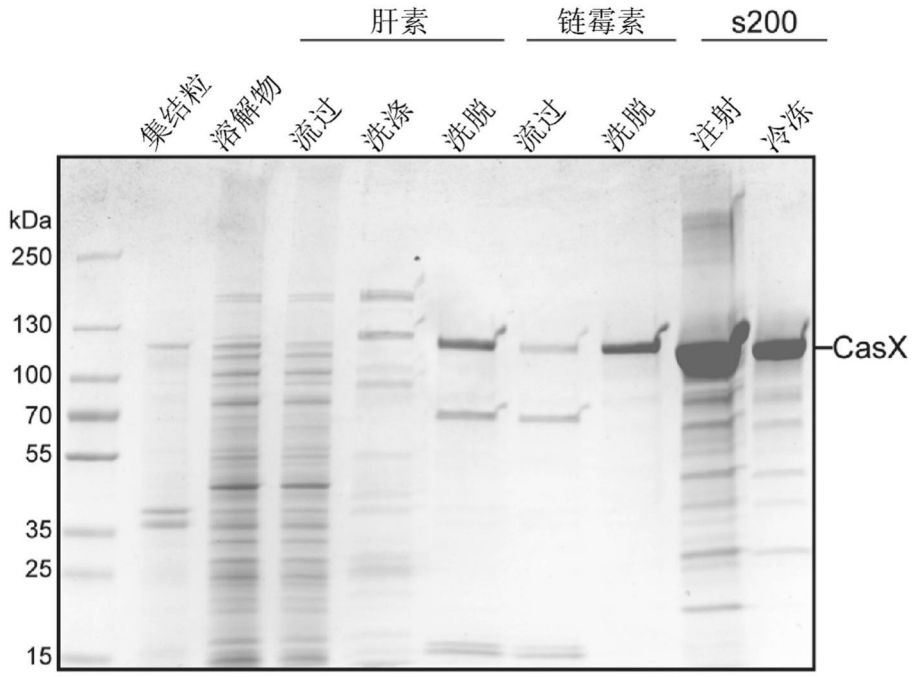
[0850]

构建体	RNA序列	SEQ ID NO
15.3 20-nt间隔区	CAAACAAAUGUGUCACAAAG	355
15.3 19-nt间隔区	CAAACAAAUGUGUCACAAA	356
15.3 18-nt间隔区	CAAACAAAUGUGUCACAA	357
15.5 20-nt间隔区	GGAAUAAUGCUGUUGUUGAA	358
15.5 19-nt间隔区	GGAAUAAUGCUGUUGUUGA	359
15.5 18-nt间隔区	GGAAUAAUGCUGUUGUUG	360

[0851] 结果

[0852] 使用CasX变异体491和由具有间隔区15.3或15.5的支架174构成的引导序列组装CasX RNP,所述两个间隔区均靶向TCRα基因的恒定区。测试具有全长20nt间隔区以及截短19nt和18nt间隔区的引导序列以确定当预组装的RNP核转染以用于离体编辑时,使用较短间隔区是否支持编辑增加。在22μL核转染反应中在0.3125μM到2.5μM范围内的2倍稀释下测试RNP。通过在核转染之后三天的流式细胞术评估编辑。对于两个间隔区序列,跨越剂量范围,与具有20nt间隔区的RNP相比,具有截短间隔区的RNP编辑大多更有效(图25,剂量-反应

曲线)。对于间隔区15.3,相比于20nt间隔区的1.414μM,18nt和19nt间隔区的EC50值分别为0.225μM和0.299μM(对于两个截短序列p<0.0001;额外平方和F检验)。对于间隔区15.5,18nt间隔区的EC50为0.519μM,相对于20nt间隔区的0.938μM(p=0.0001),而19nt间隔区更类似于20nt间隔区,EC50为0.808μM(p=0.0762)。尽管19nt 15.3间隔区具有类似于18nt间隔区的编辑,同时19nt 15.5间隔区更紧密类似于对应20nt间隔区,然而,对于所测试的两个间隔区,趋势的方向保持一致,且表明使用具有18nt间隔区的引导序列可为用于当CasX编辑分子作为预组装的RNP递送时来增加编辑的可推广策略。将进行利用基于细胞的分析的额外实验以确认这些发现。


肝素
链霉素
s200
集结粒
溶解物
流过
洗涤
洗脱
流过
洗脱
注射
冷冻
kDa
250
130
100
70
55
35
25
15
CasX

胶体考马斯

图1

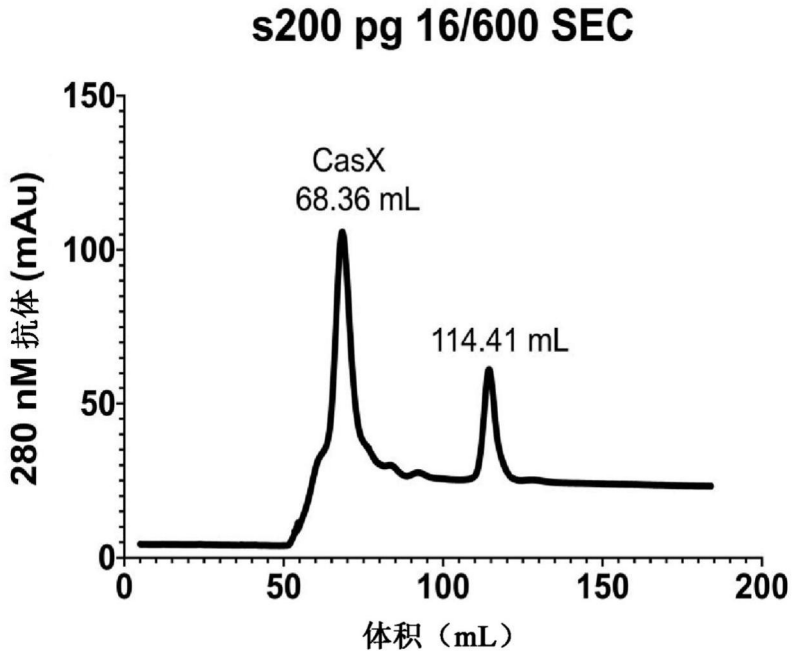

s200 pg 16/600 SEC
280 nM 抗体 (mAu)
150
100
50
0
CasX
68.36 mL
114.41 mL
0
50
100
150
200
体积（mL）

图2

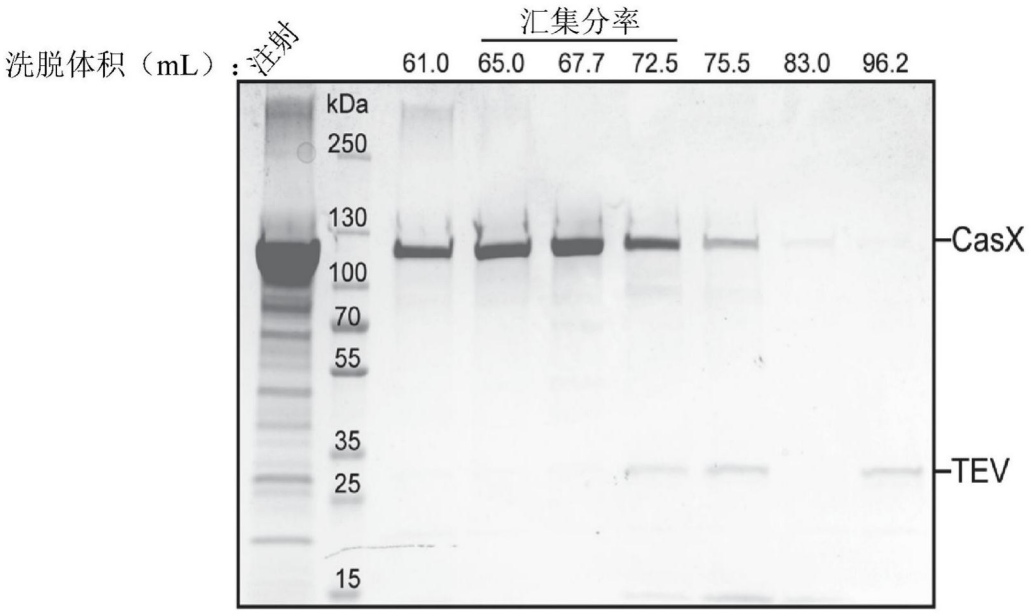

汇集分率
洗脱体积（mL）：
注射
61.0
65.0
67.7
72.5
75.5
83.0
96.2
kDa
250
130
100
70
55
35
25
15
CasX
TEV
胶体考马斯

图3

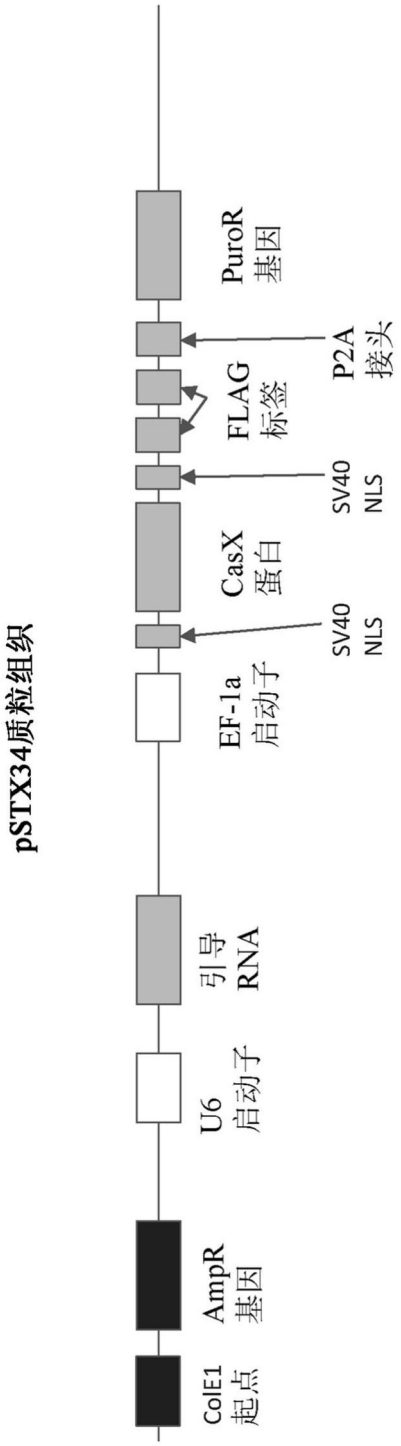

pSTX34质粒组织
ColE1 起点
AmpR 基因
U6 启动子
引导 RNA
EF-1a 启动子
SV40 NLS
CasX 蛋白
SV40 NLS
FLAG 标签
P2A 接头
PuroR 基因

图4

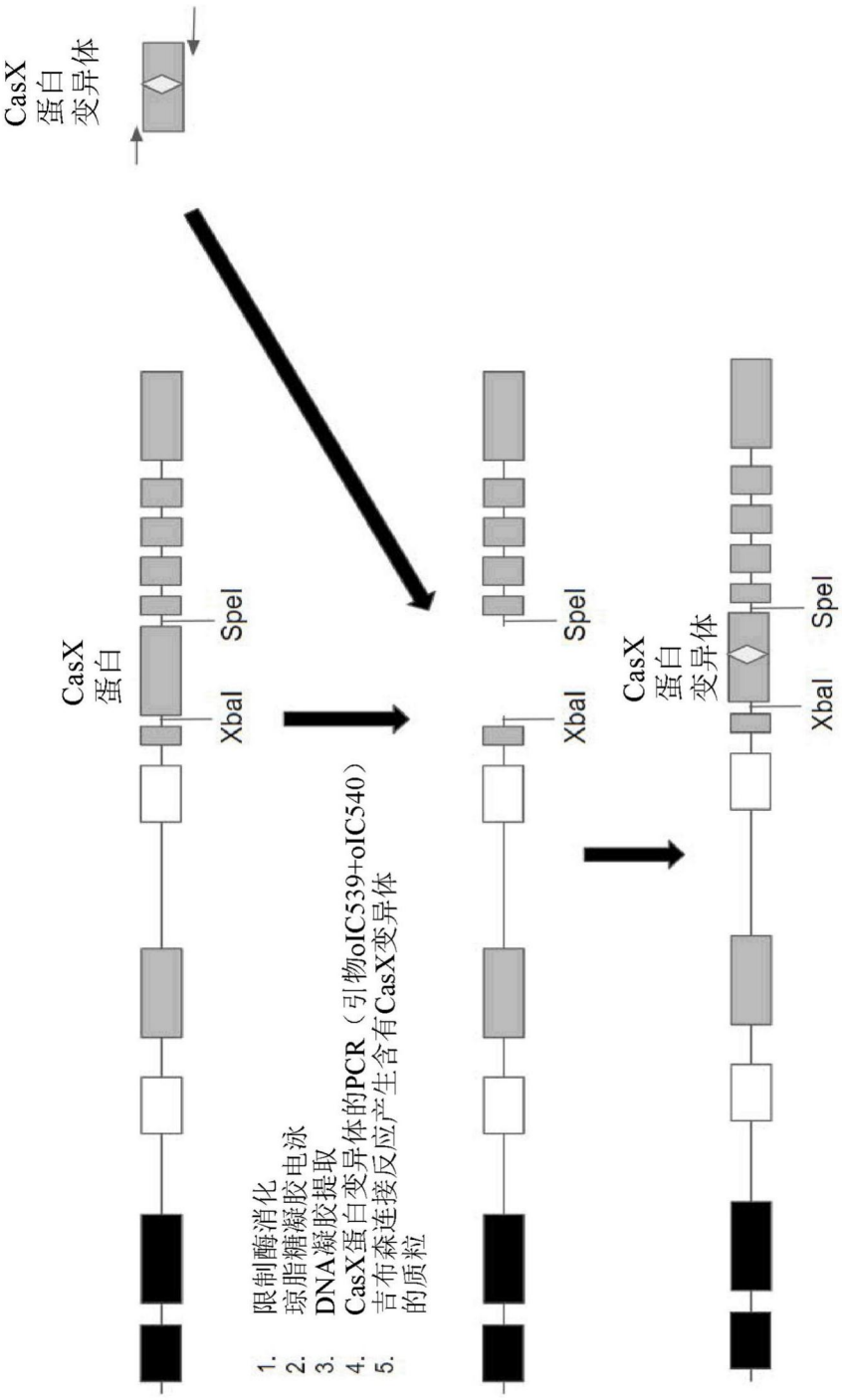

CasX
蛋白
变异体
CasX
蛋白
SpeI
XbaI
1. 限制酶消化
2. 琼脂糖凝胶电泳
3. DNA凝胶提取
4. CasX蛋白变异体的PCR（引物oIC539+oIC540）
5. 吉布森连接反应产生含有CasX变异体的质粒
SpeI
XbaI
CasX
蛋白
变异体
SpeI
XbaI

图5

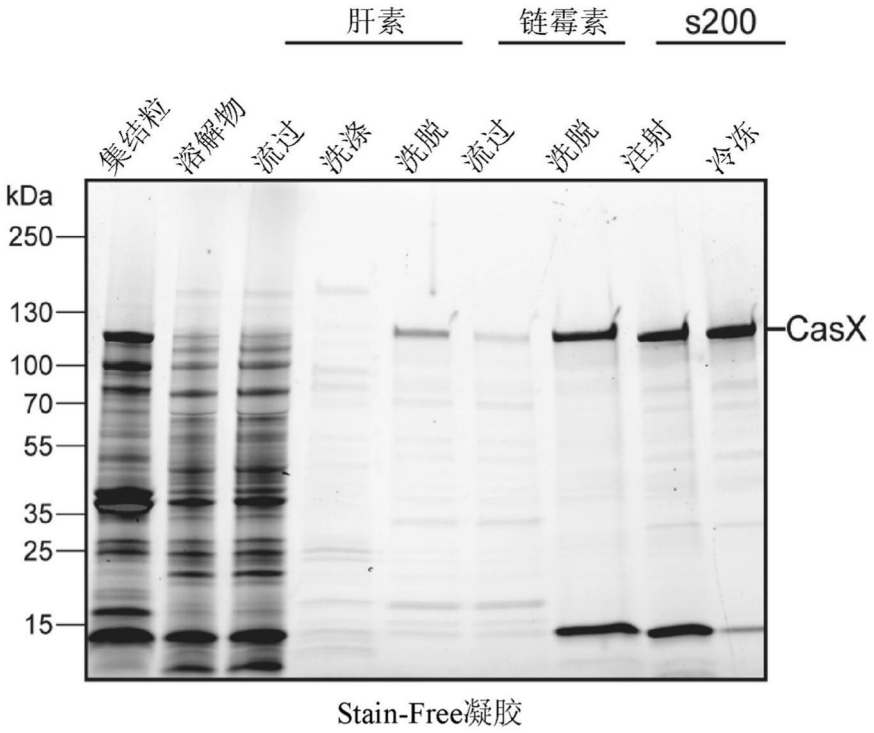

肝素
链霉素
s200
集结粒
溶解物
流过
洗涤
洗脱
流过
洗脱
注射
冷冻
kDa
250
130
100
70
55
35
25
15
CasX
Stain-Free凝胶

图6

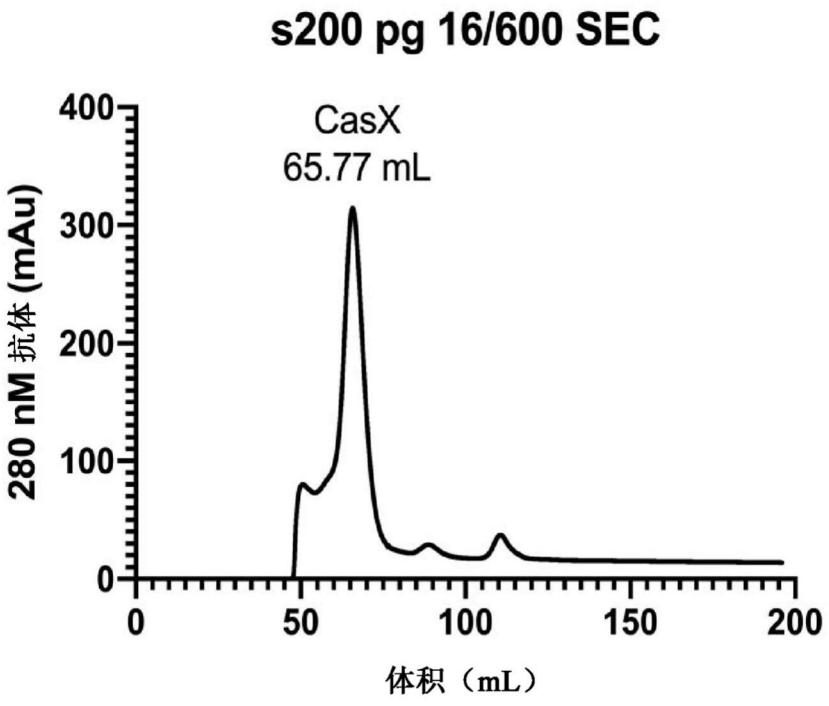
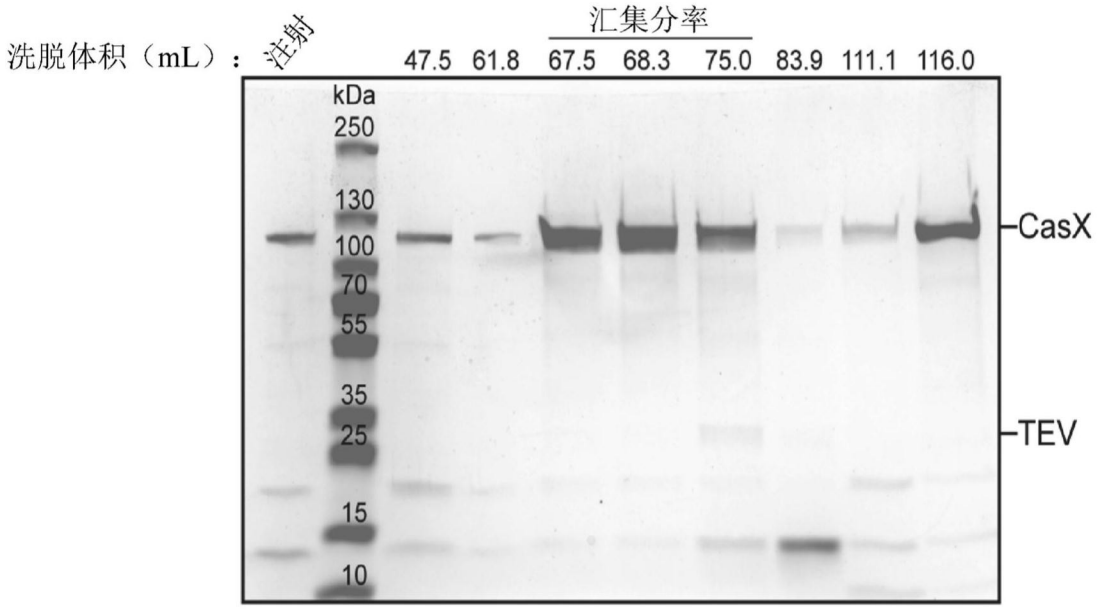

s200 pg 16/600 SEC
CasX
65.77 mL
280 nM 抗体 (mAu)
400
300
200
100
0
0
50
100
150
200
体积（mL）

图7


洗脱体积（mL）：
注射
汇集分率
47.5
61.8
67.5
68.3
75.0
83.9
111.1
116.0
kDa
250
130
100
70
55
35
25
15
10
CasX
TEV

胶体考马斯

图8

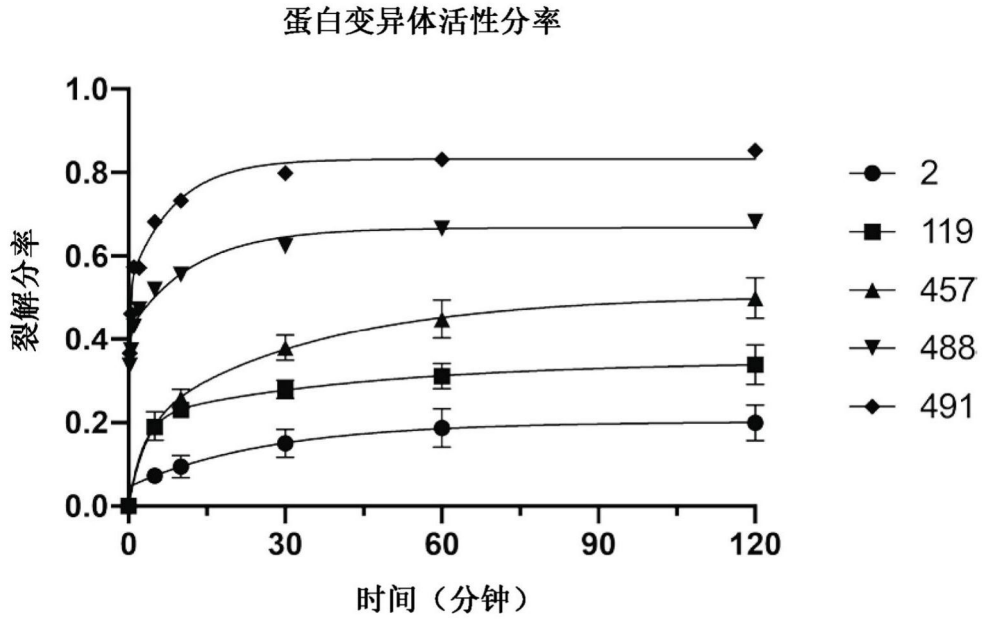

蛋白变异体活性分率
裂解分率
1.0
0.8
0.6
0.4
0.2
0.0
0
30
60
90
120
时间（分钟）
2
119
457
488
491

图9

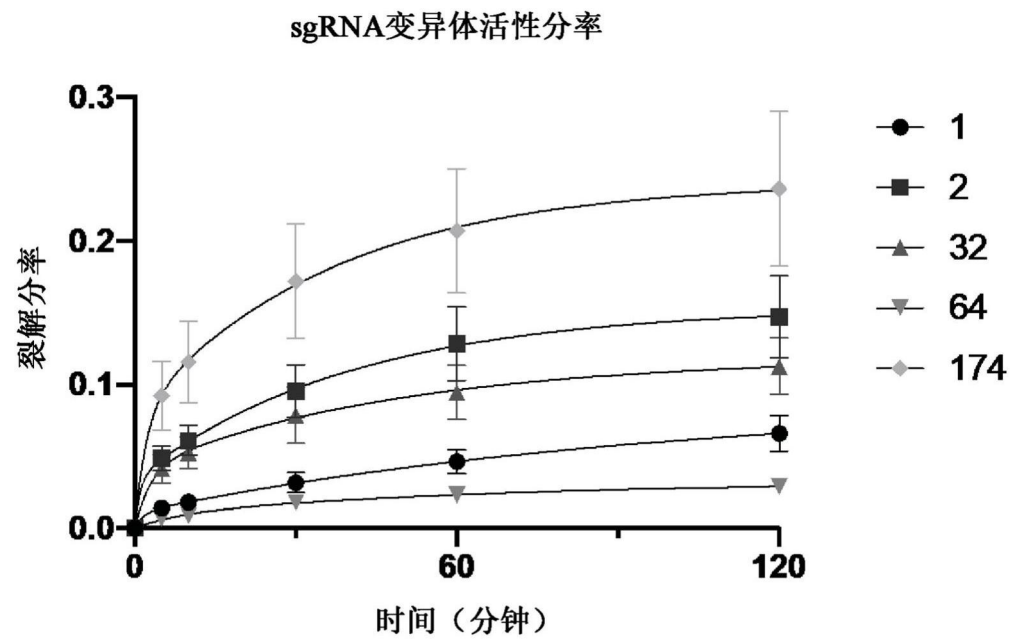

sgRNA变异体活性分率
裂解分率
0.3
0.2
0.1
0.0
0
60
120
时间（分钟）
1
2
32
64
174

图10

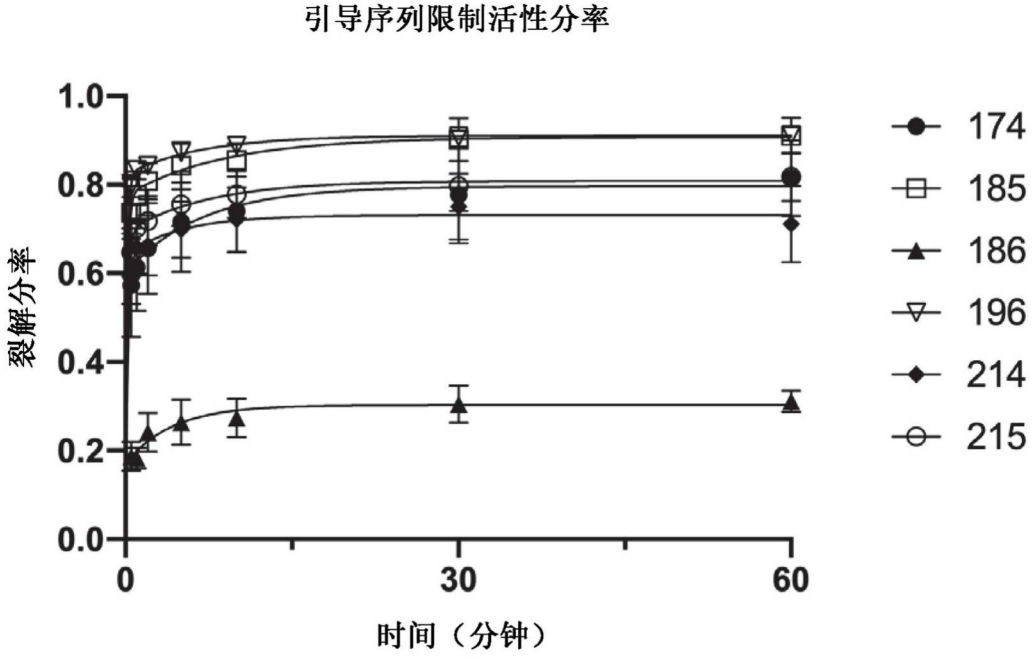

引导序列限制活性分率
裂解分率
1.0
0.8
0.6
0.4
0.2
0.0
0
30
60
时间（分钟）
174
185
186
196
214
215

图11

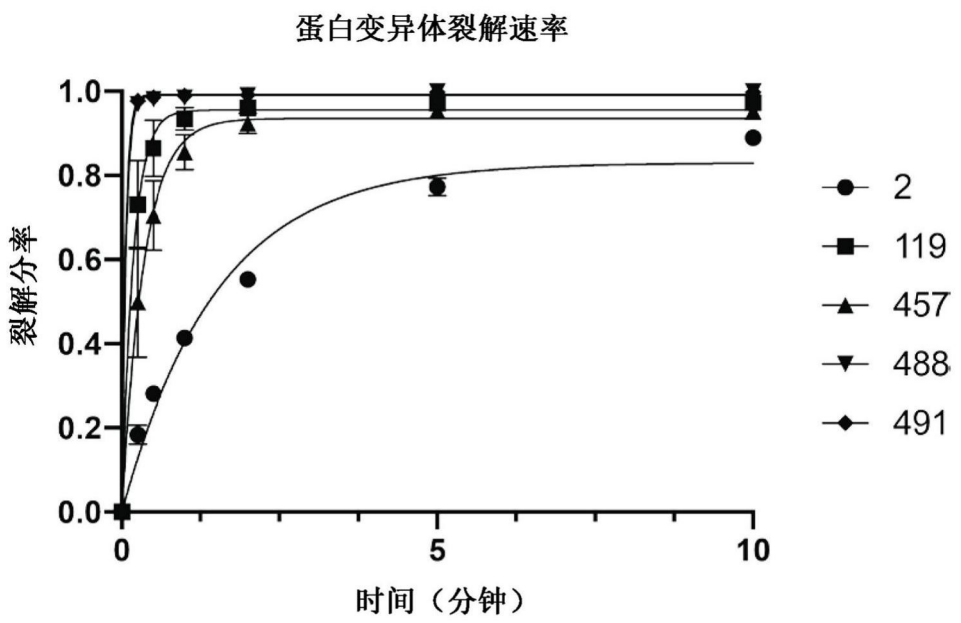

蛋白变异体裂解速率
裂解分率
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
时间（分钟）
2
119
457
488
491

图12

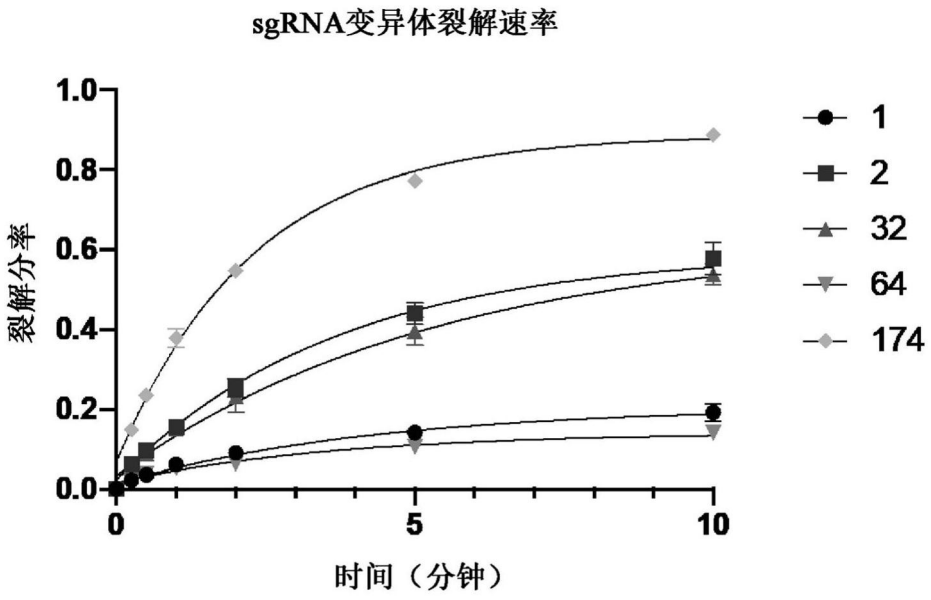

sgRNA变异体裂解速率
裂解分率
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
时间（分钟）
1
2
32
64
174

图13

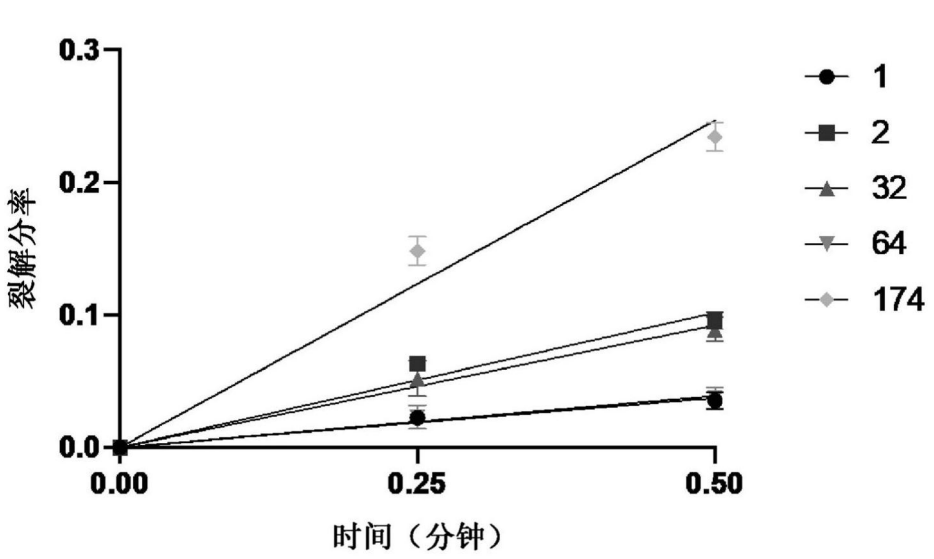

sgRNA变异体初始速度
0.3
0.2
0.1
0.0
裂解分率
0.00
0.25
0.50
时间（分钟）
1
2
32
64
174

图14

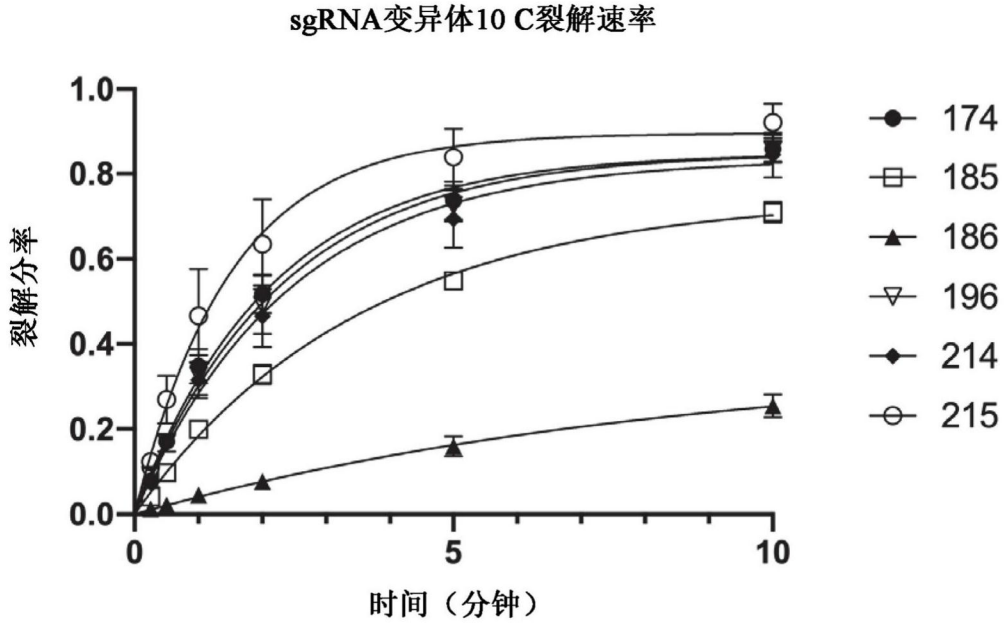

sgRNA变异体10 C裂解速率
1.0
0.8
0.6
0.4
0.2
0.0
裂解分率
0
5
10
时间（分钟）
174
185
186
196
214
215

图15

NTC PAM上的裂解速率

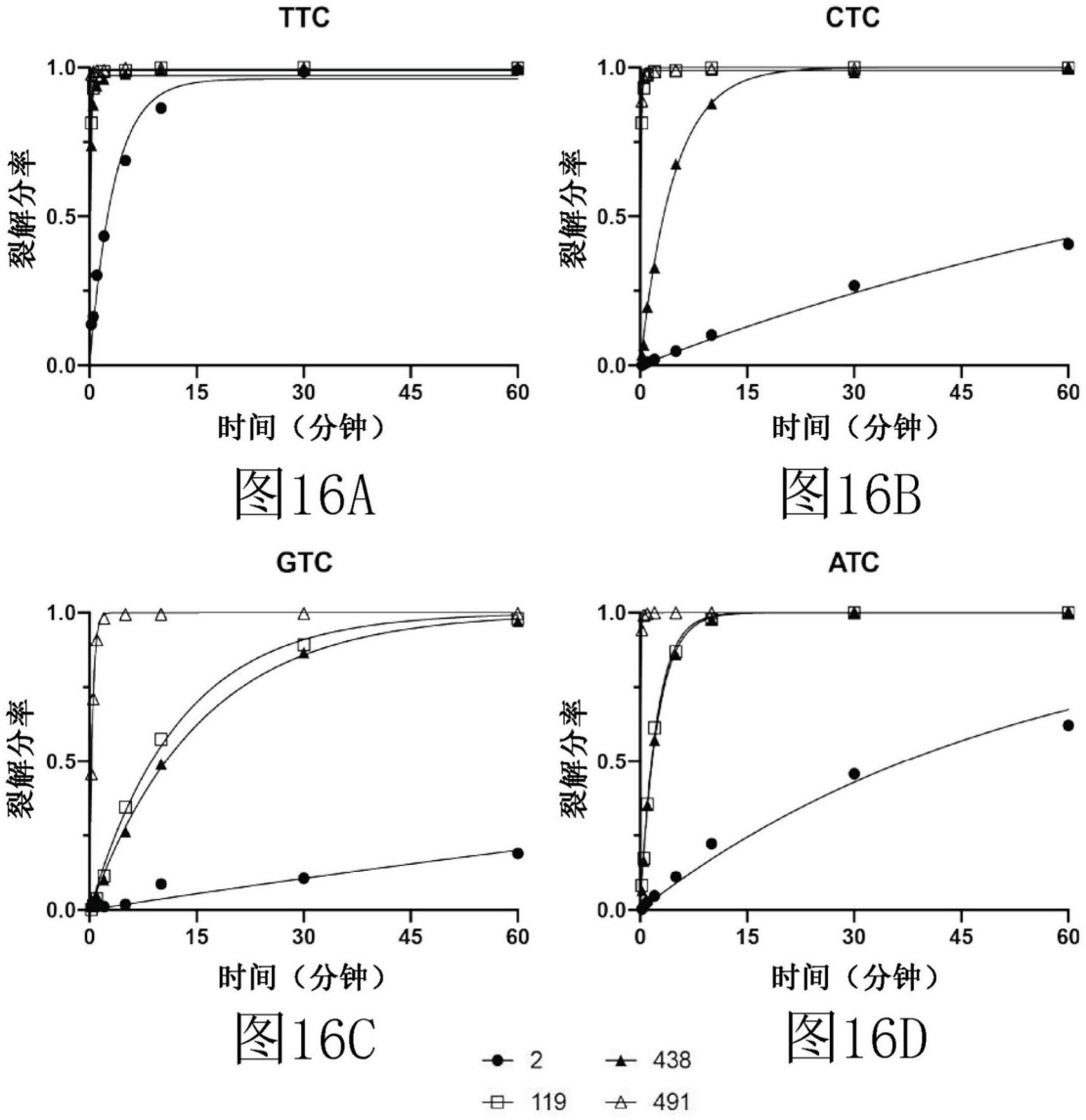


图16A 图16B

图16C 图16D

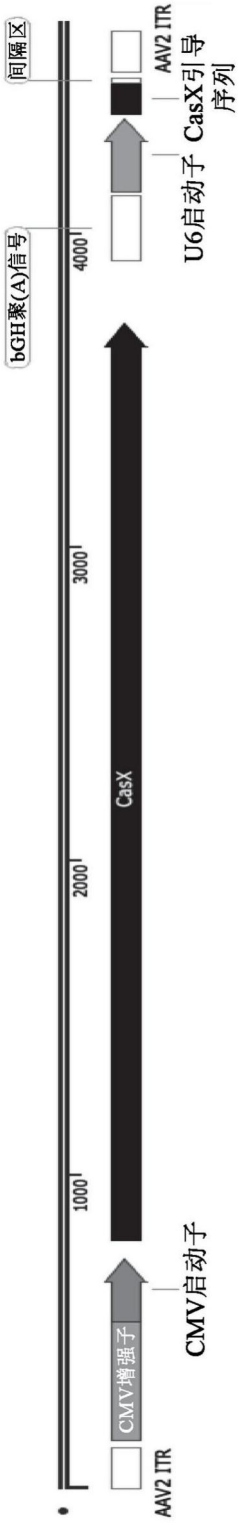

间隔区
bGH聚(A)信号
4000
3000
2000
1000
AAV2 ITR
CasX引导序列
U6启动子
CasX
CMV启动子
CMV增强子
AAV2 ITR

图17

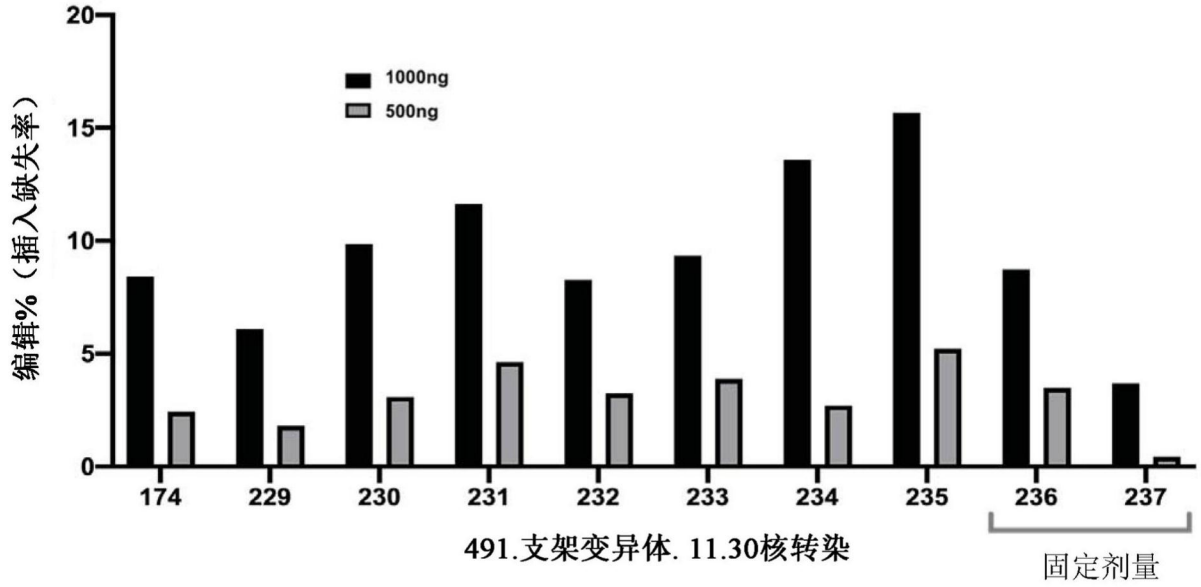

20
15
10
5
0
编辑%（插入缺失率）
1000ng
500ng
174
229
230
231
232
233
234
235
236
237
491.支架变异体. 11.30核转染
固定剂量

图18

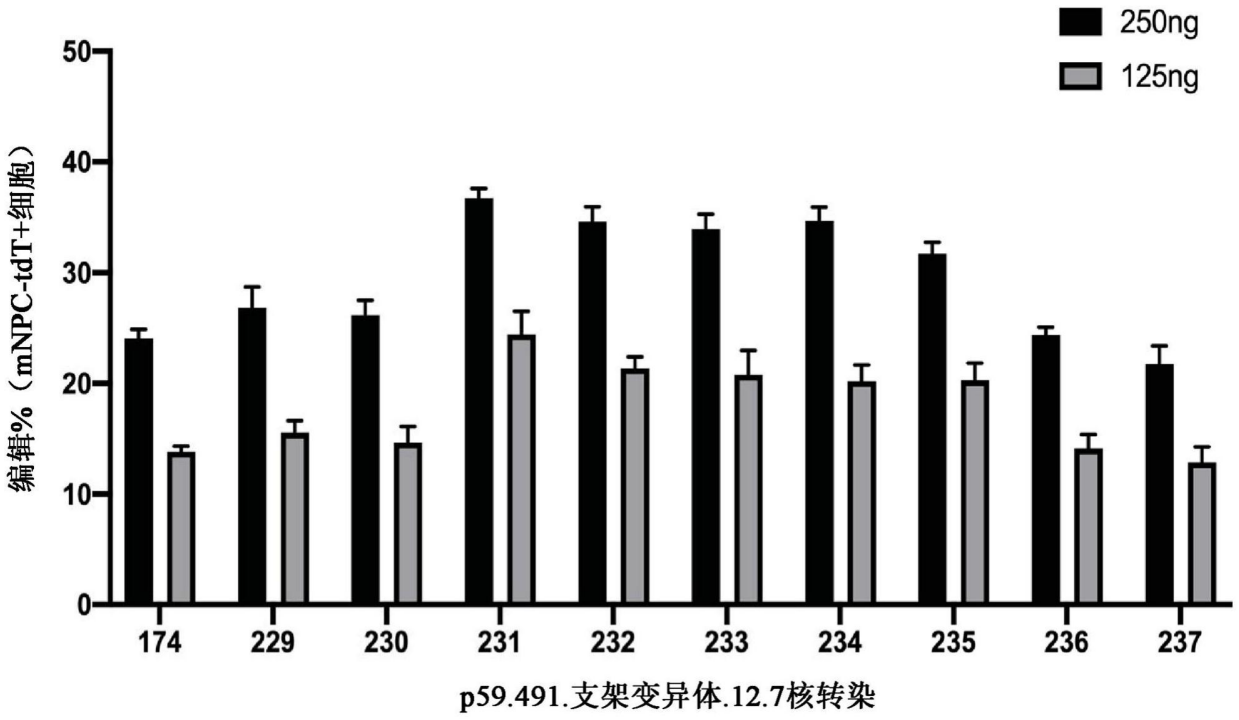

250ng
125ng
50
40
30
20
10
0
编辑%（mNPC-tdT+细胞）
174
229
230
231
232
233
234
235
236
237
p59.491.支架变异体.12.7核转染

图19

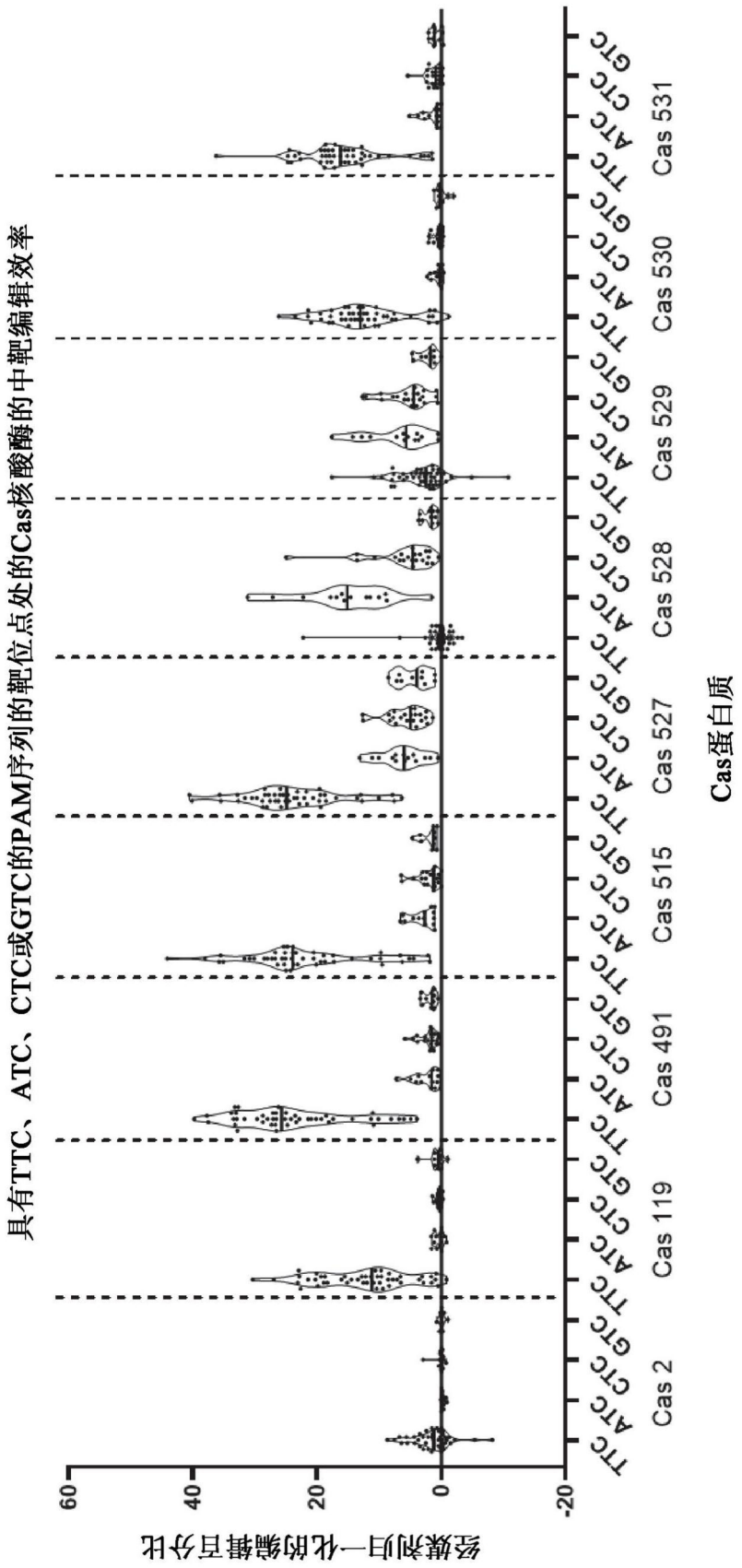

具有TTC、ATC、CTC或GTC的PAM序列的靶位点处的Cas核酸酶的中靶编辑效率
经媒剂归一化的编辑百分比
60
40
20
0
-20
TTC ATC CTC GTC
Cas 2
Cas 119
Cas 491
Cas 515
Cas 527
Cas 528
Cas 529
Cas 530
Cas 531
Cas蛋白质

图20

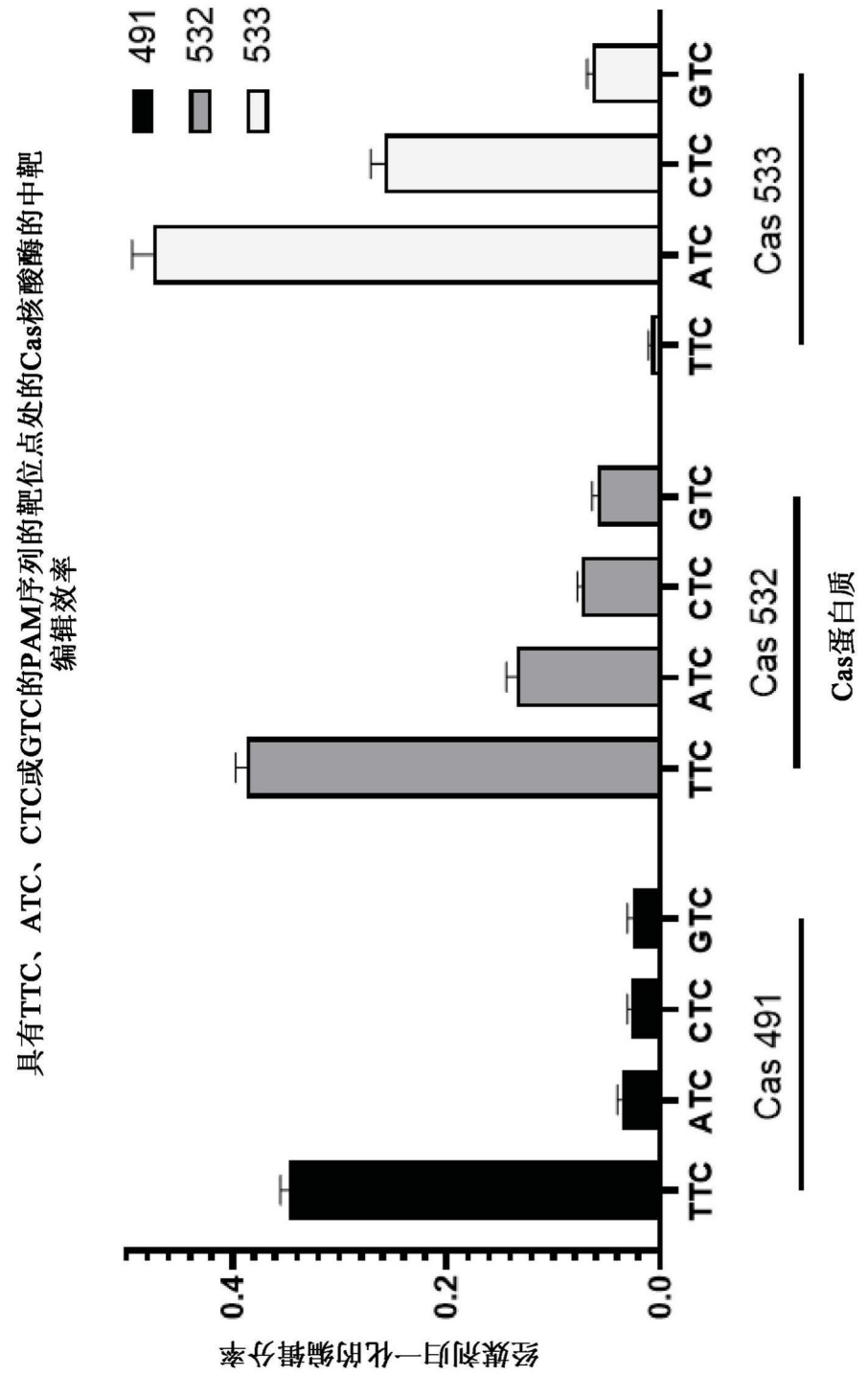

具有TTC、ATC、CTC或GTC的PAM序列的靶位点处的Cas核酸酶的中靶编辑效率
491
532
533
经媒剂归一化的编辑分率
0.0
0.2
0.4
TTC ATC CTC GTC
TTC ATC CTC GTC
TTC ATC CTC GTC
Cas 491
Cas 532
Cas 533
Cas蛋白质

图21

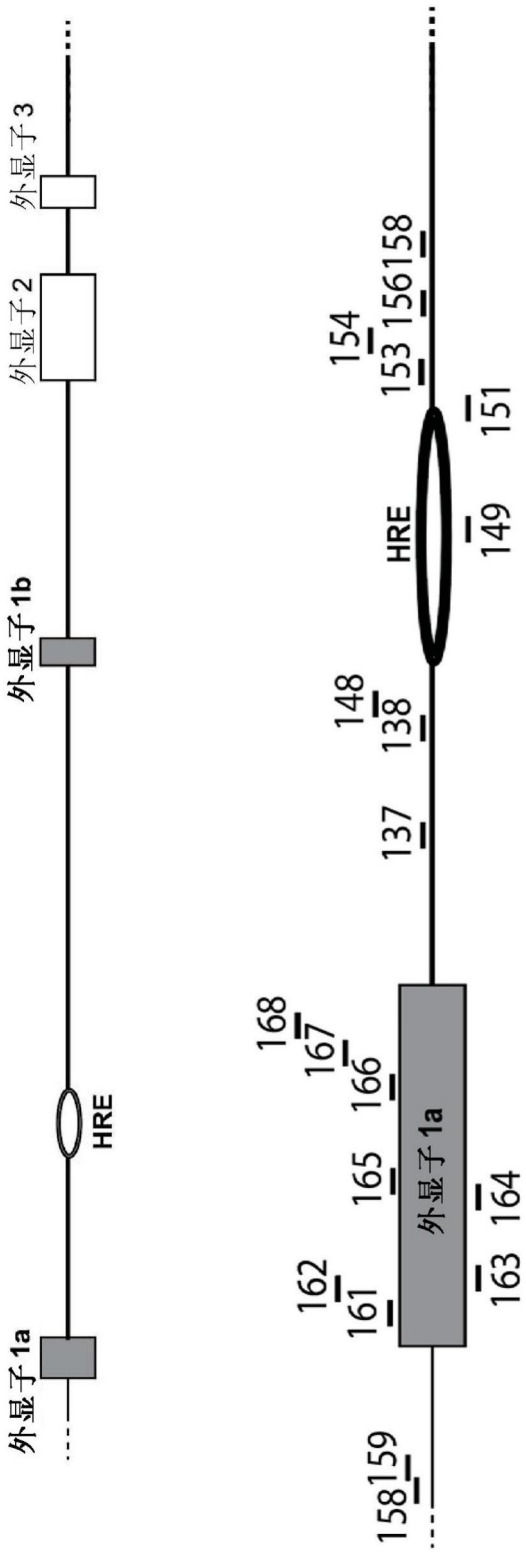

外显子3
外显子2
外显子1b
HRE
外显子1a
154
153
156
158
151
149
HRE
148
138
137
168
167
166
165
外显子1a
164
163
162
161
158
159

图22

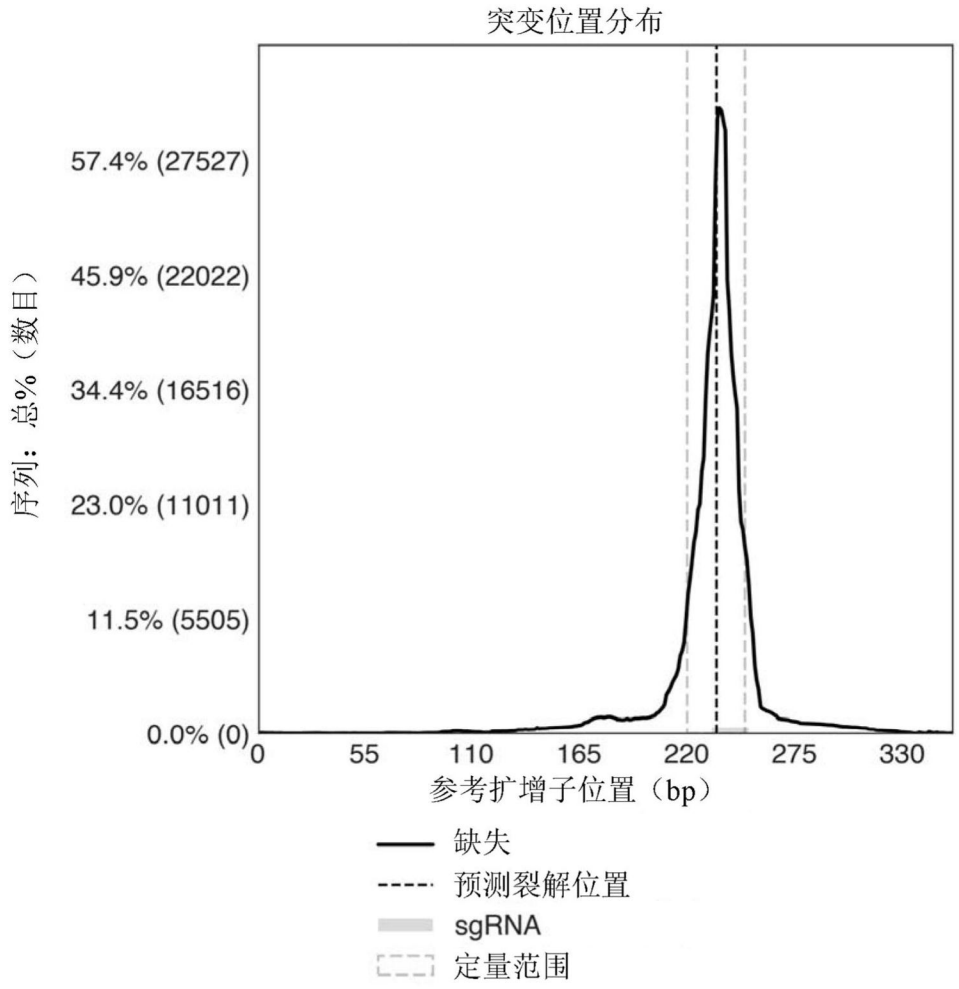

突变位置分布
57.4% (27527)
45.9% (22022)
34.4% (16516)
23.0% (11011)
11.5% (5505)
0.0% (0)
序列：总%（数目）
0
55
110
165
220
275
330
参考扩增子位置（bp）
缺失
预测裂解位置
sgRNA
定量范围

图23

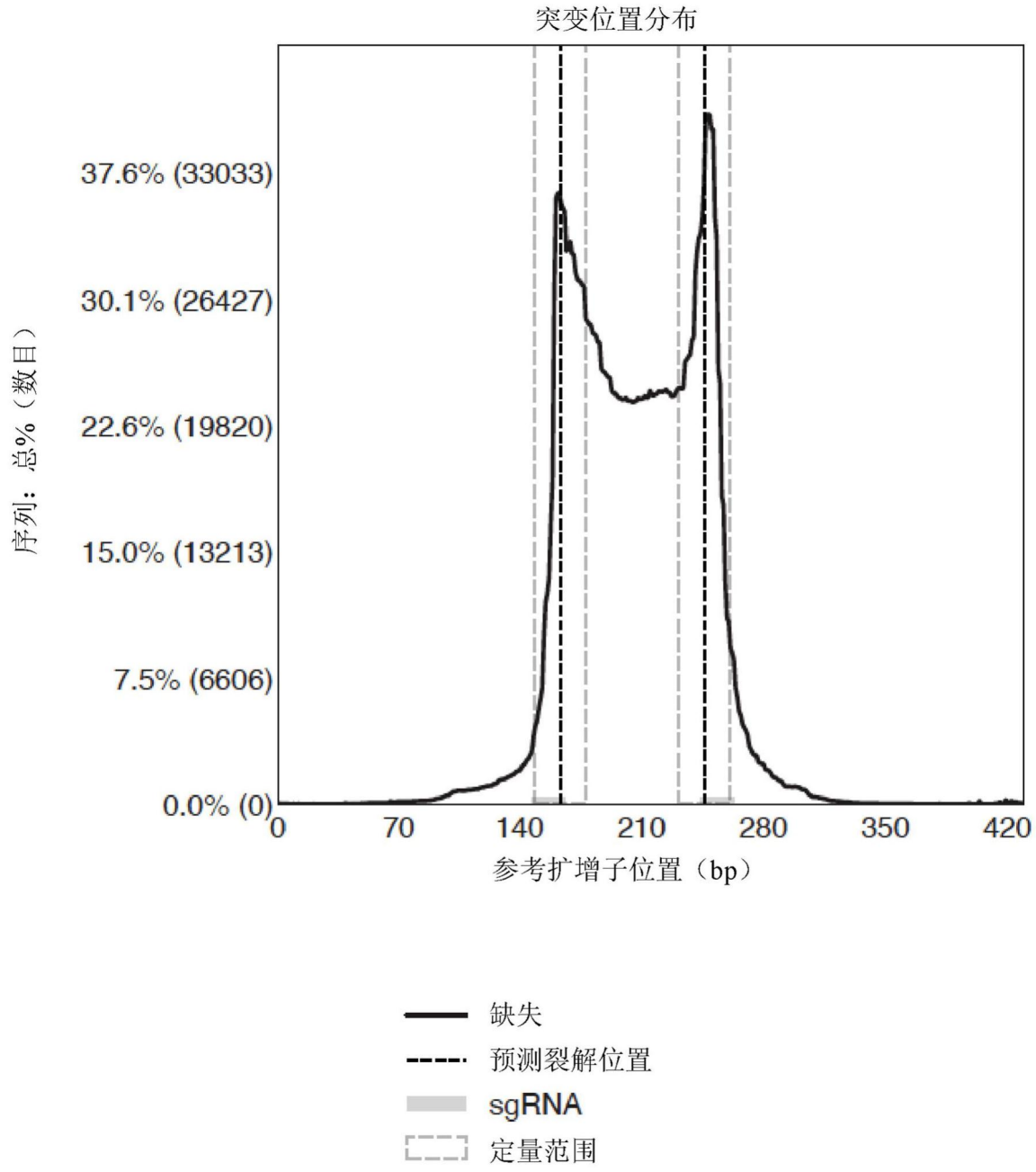

突变位置分布
37.6% (33033)
30.1% (26427)
22.6% (19820)
15.0% (13213)
7.5% (6606)
0.0% (0)
序列：总%（数目）
0
70
140
210
280
350
420
参考扩增子位置（bp）
缺失
预测裂解位置
sgRNA
定量范围

图24

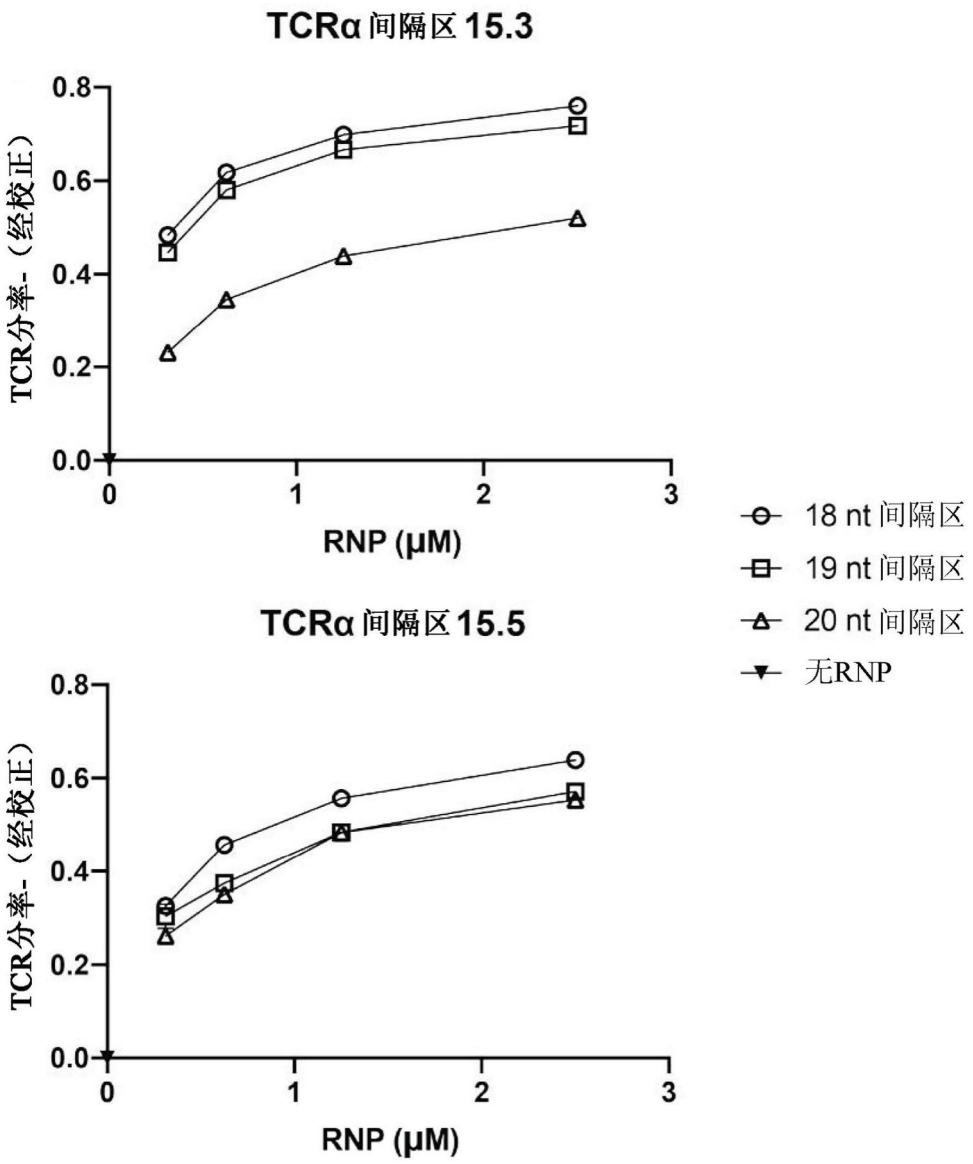

TCRα 间隔区 15.3
TCR分率-（经校正）
0.8
0.6
0.4
0.2
0.0
0
1
2
3
RNP (μM)
TCRα 间隔区 15.5
TCR分率-（经校正）
0.8
0.6
0.4
0.2
0.0
0
1
2
3
RNP (μM)
18 nt 间隔区
19 nt 间隔区
20 nt 间隔区
无RNP

图25