(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 118369429 A
(43)申请公布日 2024.07.19

(21)申请号 202280067009.7
(22)申请日 2022.09.08
(30)优先权数据
63/241,964 2021.09.08 US
(85)PCT国际申请进入国家阶段日
2024.04.02
(86)PCT国际申请的申请数据
PCT/US2022/076106 2022.09.08
(87)PCT国际申请的公布数据
WO2023/039468 EN 2023.03.16
(71)申请人 比姆医疗股份有限公司
地址 美国马萨诸塞州剑桥市兰兹当街26号2楼
申请人 冉飞 林启宇
(72)发明人 冉飞 林启宇
(74)专利代理机构 北京恒律知识产权代理有限公司 11416
专利代理师 戴嵩玮
(51)Int.Cl.
C12N 15/113(2006.01)
C12N 15/11(2006.01)
C12N 15/86(2006.01)

权利要求书12页 说明书376页 附图16页

(54)发明名称

病毒引导RNA递送

(57)摘要

本文提供了重组负链RNA病毒基因组(例如重组狂犬病病毒基因组)和重组负链RNA病毒(例如重组狂犬病病毒)以及使用它们将引导RNA和任选地转基因递送至靶细胞中的方法。还提供了包装系统和使用所述包装系统产生重组负链RNA病毒的方法。

CN 118369429 A

1.一种重组负链RNA病毒基因组,其包含

编码第一引导RNA(gRNA)的核酸,所述编码第一引导RNA(gRNA)的核酸包含5'端和3'端;和

编码第一转移RNA(tRNA)的核酸,所述编码第一转移RNA(tRNA)的核酸位于编码所述第一gRNA的所述核酸的所述3'端或编码所述第一gRNA的所述核酸的所述5'端中的一者或两者处。

2.如权利要求1所述的重组负链RNA病毒基因组,其包含编码第二tRNA的核酸。

3.如权利要求2所述的重组负链RNA病毒基因组,其中编码所述第一tRNA的所述核酸位于编码所述第一gRNA的所述核酸的所述3'端;并且编码所述第二tRNA的所述核酸位于编码所述第一gRNA的所述核酸的所述5'端。

4.如权利要求3所述的重组负链RNA病毒基因组,其中所述第一tRNA的核苷酸序列和所述第二tRNA的核苷酸序列是至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%,或100%同一的。

5.如权利要求3或4所述的重组负链RNA病毒基因组,其中所述第一tRNA和所述第二tRNA指定相同的氨基酸。

6.如权利要求3或4所述的重组负链RNA病毒基因组,其中所述第一tRNA和所述第二tRNA指定不同的氨基酸。

7.如权利要求1-6中任一项所述的重组负链RNA病毒基因组,其包含编码所述第一tRNA的两种核酸。

8.如权利要求1或2所述的重组负链RNA病毒基因组,其包含编码所述第一tRNA的三种核酸。

9.如权利要求1-8中任一项所述的重组负链RNA病毒基因组,其包含编码第二gRNA的核酸。

10.如权利要求9所述的重组负链RNA病毒基因组,其中所述两种或更多种核酸编码同一的gRNA。

11.如权利要求9所述的重组负链RNA病毒基因组,其中所述两种或更多种核酸编码至少一种不同的gRNA。

12.如权利要求9所述的重组负链RNA病毒基因组,其中所述第一gRNA的核苷酸序列和所述第二gRNA的核苷酸序列是至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%,或100%同一的。

13.如权利要求9所述的重组负链RNA病毒基因组,其中所述第一gRNA和所述第二gRNA与相同的靶核酸序列特异性地杂交。

14.如权利要求9所述的重组负链RNA病毒基因组,其中所述第一gRNA和所述第二gRNA与不同的靶核酸序列特异性地杂交。

15.如权利要求1-14中任一项所述的重组负链RNA病毒基因组,其中所述第一tRNA和/或所述第二tRNA各自选自由以下组成的组:tRNA-ala、tRNA-arg、tRNA-asn、tRNA-asp、tRNA-cys、tRNA-gln、tRNA-gly、tRNA-his、tRNA-ile、tRNA-leu、tRNA-lys、tRNA-met、tRNA-phe、tRNA-pro、tRNA-pyl、tRNA-sec、tRNA-ser、tRNA-thr、tRNA-trp、tRNA-tyr和tRNA-val。

16.如权利要求1-15中任一项所述的重组负链RNA病毒基因组,其中所述编码第一tRNA

和/或第二tRNA的核酸包含以下中的任一种:

GGCTCGTTGGTCTAGGGGTATGATTCTCGCTTAGGGTGCG AGAGGTCCCGGGTTCAAATCCCGGACGAGCCC(tRNA-pro;SEQ ID NO:4011),或与其至少90%同一的序列;

GGCTCCATAGCTCAGGGGTTAGAGCACTGGTCTTGTAAAC CAGGGGTCGCGAGTTCAATTCTCGCTGGGGCTT(tRNA-thr;SEQ ID NO:4012),或与其至少90%同一的序列;

GCGTTGGTGGTATAGTGGTGAGCATAGCTGCCTTCCAAGC AGTTGACCCGGGTTCGATTCCCGGCCAACGCA(tRNA-gly G8;SEQ ID NO:4013),或与其至少90%同一的序列;

GCATGGGTGGTTCAGTGGTAGAATTCTCGCCTGCCACGCG GGAGGCCCGGGTTCGATTCCCGGCCCATGCA(tRNA-gly G27;SEQ ID NO:4014),或与其至少90%同一的序列;

GTCAGGATGGCCGAGCGGTCTAAGGCGCTGCGTTCAGGTC GCAGTCTCCCCTAGAGGCGTGGGTTCGAATCCCACTCCTGACA(tRNA-leu;SEQ ID NO:4015),或与其至少90%同一的序列;

GCTCCAGTGGCGCAATCGGTTAGCGCGCGGTACTTATAAG ACAGTGCACCTGTGAGCAATGCCGAGGTTGTGAGTTCAAGCCT CACCTGGAGCA(tRNA-ile;SEQ ID NO:4016),或与其至少90%同一的序列;

GAAAAAGTCATGGAGGCCATGGGGTTGGCTTGAAACCAG CTTTGGGGGGTTCGATTCCTTCCTTTTTTGTCT(tRNA-ser;SEQ ID NO:4017),或与其至少90%同一的序列;

GGGCCAGTGGCGCAATGGATAACGCGTCTGACTACGGATC AGAAGATTCCAGGTTCGACTCCTGGCTGGCTCGGTGTA(tRNA-arg;SEQ ID NO:4018),或与其至少90%同一的序列;

AAACAAGCGCAAGTGGTTTAGTGGTAAAATCCAACGTTGC CATCGTTGGGCCCCCGGTTCGATTCCGGGCTTGCGCA(tRNA-asp1;SEQ ID NO:4019),或与其至少90%同一的序列;

AACAAAGCACCAGTGGTCTAGTGGTAGAATAGTACCCTGC CACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCA(tRNA-asp2;SEQ ID NO:4020),或与其至少90%同一的序列;或

TCCTCGTTAGTATAGTGGTGAGTATCCCCGCCTGTCACGCG GGAGACCGGGGTTCGATTCCCCGACGGGGAG(tRNA-asp D15;SEQ ID NO:4021),或与其至少90%同一的序列。

17.如权利要求1-14中任一项所述的重组负链RNA病毒基因组,其中所述第一tRNA和/或所述第二tRNA包含tRNA样结构。

18.如权利要求17所述的重组负链RNA病毒基因组,其中所述tRNA样结构包含MALAT1相关的小细胞质RNA(mascRNA)。

19.如权利要求18所述的重组负链RNA病毒基因组,其中所述mascRNA由包含以下中的任一种的核酸编码:

AAAAGCAAAAGATGCTGGTGGTTGGCACTCCTGGTTTCCA GGACGGGGTTCAAATCCCTGCGGCGTCTTTGCTTT

(masc_Malat1;SEQ ID NO:4022),或与其至少90%同一的序列;

AAAGACGCTGGTGGTTGGTGTTTCCAGGACGGGGTTCAAG TCCCTGCGGCGTCCTCGC(masc_liz38;SEQ ID NO:4023),或与其至少90%同一的序列;

GGCTCTGGTGGCTTCCAGGACGGGGTTCAAGTCCCTGCAG TGCCCTTGCTGA(masc_liz40;SEQ ID NO:4024),或与其至少90%同一的序列;

AAAGGCGCTGGTGGTGGCACTCCCAGCGGGACGGGGTTCG AATCCCCGCGGCGCCTCTGC(masc_turk;SEQ ID NO:4025),或与其至少90%同一的序列;

GAAGGTTTTTCTTTTCCTGAGAAAACAACACGTATTGTTTTCTCAGGTTTTGCTTTTTGGCCTTTTTCTAGC

TTAAAAAAAAAAGCAAAAGATGCTGGTGGTTGGCACTCCTGGTTTCCAGGACGGGGTTCAAATCCCTGCGGCGTCTTTGCTTT(hMALAT1.1;SEQ ID NO:4026),或与其至少90%同一的序列;

GCAGGTGTTTCTTTTACTGAGTGCAGCCCATGGCCGCACTCAGGTTTTGCTTTTCACCTTCCCATCTGTGAAAGAGTGAGCAGGAAAAAGCAAAAGGCGCTGGTGGTGGCACGTCCAGCACGGCTGGGCCGGGGTTCGAGTCCCCGCAGTGTTGCTGC(hMALAT1.2;SEQ ID NO:4027),或与其至少90%同一的序列;

GAAGGTTTTTCTTTTCCTGAGAAAACAACACGTTTTGTTTTCTCAGGTTTTGCTTTTTGGCCTTTTTCTAGCTTAAAAAAAAAAAGCAAAAGATGCTGGTGGTTGGCACTCCTGGTTTCCAGGACAGGGTTCAAATCCCTGCGGCGTCTTTGCTTT(黑猩猩1;SEQ ID NO:4028),或与其至少90%同一的序列;

AAAGCAAAAGATGCTGGTGGTTGGCACTCCTGGTTTCCAG GACAGGGTTCAAATCCCTGCGGCGTCTTTGCTTT(黑猩猩1短;SEQ ID NO:4029),或与其至少90%同一的序列;

GCAGGTGTTTCTTTTACTGAGTGCAGCCCATGGCCGCACTCAGGTTTTGCTTTTCACCTTCCCATCTGTGAAAGAGTGAGCAGGAAAAAGCAAAAGGCGCTGGTGGTGGCACGTCCAGCACGGCTGGGCCGGGGTTCGAGTCCCCGCAGTGTTGCTGC(黑猩猩2;SEQ ID NO:4030),或与其至少90%同一的序列;

AAAGGTTTTTCTTTTCCTGAGAAAACAACCTTTTGTTTTCTCAGGTTTTGCTTTTTGGCCTTTCCCTAGCTTTAAAAAAAAAAGCAAAAGACGCTGGTGGCTGGCACTCCTGGTTTCCAGGACGGGGTTCAAGTCCCTGCGGTGTCTTTGC(MoTse.1;SEQ ID NO:4031),或与其至少90%同一的序列;

AAAGCAAAAGACGCTGGTGGCTGGCACTCCTGGTTTCCAG GACGGGGTTCAAGTCCCTGCGGTGTCTTTGCTTGAC(MoTse.1短;SEQ ID NO:4032),或与其至少90%同一的序列;或

GCAGGTGTTTCTTTTCCTGACCGCGGCTCATGGCCGCGCTCAGGTTTTGCTTTTCACCTTTGTCTGAGAGAACGAACGTGAGCAGGAAAAAGCAAAAGGCACTGGTGGCGGCACGCCCGCACCTCGGGCCAGGGTTCGAGTCCCTGCAGTACCGTGC(MoTse.2;SEQ ID NO:4033),或与其至少90%同一的序列。

20.如权利要求17所述的重组负链RNA病毒基因组,其中所述tRNA样结构包含tRNA变体。

21.如权利要求20所述的重组负链RNA病毒基因组,其中所述tRNA变体包含用G或C核苷酸对一个或多个A和/或T核苷酸进行的取代。

22.如权利要求20所述的重组负链RNA病毒基因组,其中所述tRNA变体包含相对于野生型tRNA较低的A和/或T核苷酸含量。

23.如权利要求20-22中任一项所述的重组负链RNA病毒基因组,其中所述tRNA变体由包含以下中的任一种的核酸编码:

GGCTCGTTGGCCTAGGGGTATGGCTCCCGCTTAGGGTGCG GGAGGTCCCGGGTTCAAATCCCGGACGAGCC(tRNA-pro var1;SEQ ID NO:4034),或与其至少90%同一的序列;

GGCTCGTTGGCCTAGGGGTATGGCTGAAAAGGTCCCGGGT TCAAATCCCGGACGAGCC(tRNA-pro var2;SEQ ID NO:4035),或与其至少90%同一的序列;

GGCTCGTTGAAAGAAAAGGTCCCGGGTTCAAATCCCGGAC GAGCC(tRNA-pro var3;SEQ ID NO:4036),或与其至少90%同一的序列;

GGCTCCATAGCGCAGGGGTTAGCGCACCGGTCTTGTAAAC CGGGGGTCGCGAGTTCAATTCTCGCTGGGGCTT(tRNA-thr var1;SEQ ID NO:4037),或与其至少90%同一的序列;

GGCTCCATAGCGCAGGGGTTAGCGCAGAAAGGGTCGCGA GTTCAATTCTCGCTGGGGCTT(tRNA-thr var2;SEQ ID NO:4038),或与其至少90%同一的序列;或

GGCTCCATAGAAAGAAAGAAAGGGTCGCGAGTTCAATTCT CGCTGGGGCTT(tRNA-thr var3;SEQ ID NO:4039),或与其至少90％同一的序列。

24.如权利要求17所述的重组负链RNA病毒基因组,其中所述tRNA样结构包含tRNA片段。

25.如权利要求17所述的重组负链RNA病毒基因组,其中所述tRNA样结构包含病毒tRNA样结构(vtRNA)。

26.如权利要求25所述的重组负链RNA病毒基因组,其中所述vtRNA由包含以下中的任一种的核酸编码:

GCCAGAGTAGCTCAATTGGTAGAGCAACAGGTCACCGATC CTGGTGGTTCTCGGTTCAAGTCCGAGCTCTG GTC(vtRNA-1;SEQ ID NO:4040),或与其至少90％同一的序列;

GCCAGGGTAGCTCAATCGGTAGAGCAGCGGTTCCTGGAGT CCGCTGGTTCTCGGTTCAAGCCCGAGCCCTG GTTG(vtRNA-2;SEQ ID NO:4041),或与其至少90％同一的序列;

GTCGGGGTAGCTCAAATGGTAGAGTGGCAGGCCAACATA GCCAGCAGATCTCGGTTCAAACCCGAGCCCTG ACCA(vtRNA-3;SEQ ID NO:4042),或与其至少90％同一的序列;

GTCGGGGTAGCTCAATTGGTAGAGCGGCAGGCTCATCCCC TGCAGGTTCTCGGTTCAATCCCGGGTCCCGA CGC(vtRNA-4;SEQ ID NO:4043),或与其至少90％同一的序列;

GCCAGGGTAGCTCAATTGGTAGAGCATCAGGCTAGTATCC TGTCGGTTCCGGTTCAAGTCCGGGCCCTGGT TA(vtRNA-5;SEQ ID NO:4044),或与其至少90％同一的序列;

GCCAGCGTAGCTCAATTGTTAGAGCAGCGGCCACCAAGCC TGCAGGTTCTCGGTTCAAGTCCGGGCGCTGG CAT(vtRNA-6;SEQ ID NO:4045),或与其至少90％同一的序列;

GCCAGGGTAGCTCAATTGGTAGAGCGGCAGACACCACCTA CGTGGTCTAGTCTGTGGATCTCGGTTCAAGT CCGAGTCCTGGCC A(vtRNA-7;SEQ ID NO:4046),或与其至少90％同一的序列;

GCCAGGGTAGCTCAATTGGTAGAGCGGCAGACACCACCTA CGTGGTCTAGTCTGTGGATCTCGGTTCAAGT CCGAGTCCTGGCC A(vtRNA-7;SEQ ID NO:4047),或与其至少90％同一的序列;或

ACCAGAGTGGCTCACCTGGTAGAGCACCAGGCTGCCCATC CTGTTGGTTCTCGGTTCAAATCCGAGCTCTG GTGA(vtRNA-8;SEQ ID NO:4048),或与其至少90％同一的序列。

27.如权利要求1-26中任一项所述的重组负链RNA病毒基因组,其包含编码负链RNA病毒基因的核酸。

28.如权利要求1-27中任一项所述的重组负链RNA病毒基因组,其包含编码转基因的核酸。

29.如权利要求27所述的重组负链RNA病毒基因组,其中编码所述第一gRNA的所述核酸和编码所述第一tRNA的所述核酸位于各自编码负链RNA病毒基因的两种核酸之间。

30.如权利要求28所述的重组负链RNA病毒基因组,其中编码所述第一gRNA的所述核酸和编码所述第一tRNA的所述核酸位于各自编码转基因的两种核酸之间。

31.如权利要求28所述的重组负链RNA病毒基因组,其中编码所述第一gRNA的所述核酸和编码所述第一tRNA的所述核酸位于编码负链RNA病毒基因的核酸和编码转基因的核酸之间。

32.如权利要求1-31中任一项所述的重组负链RNA病毒基因组,其包含gRNA表达盒,所述gRNA表达盒从3’至5’包含负链RNA病毒转录起始信号、编码tRNA的核酸、编码gRNA的核

酸、和转录终止多聚腺苷酸化信号。

33.如权利要求1-31中任一项所述的重组负链RNA病毒基因组,其包含gRNA表达盒,所述gRNA表达盒从3'至5'包含负链RNA病毒转录起始信号、编码所述第一tRNA的核酸、编码所述第一gRNA的核酸、编码第二tRNA的核酸和转录终止多聚腺苷酸化信号。

34.如权利要求1-31中任一项所述的重组负链RNA病毒基因组,其包含gRNA表达盒,所述gRNA表达盒从3'至53'包含负链RNA病毒转录起始信号、编码所述第一tRNA的核酸、编码所述第一gRNA的核酸、编码第二tRNA的核酸、编码第二gRNA的核酸和转录终止多聚腺苷酸化信号。

35.如权利要求1-31中任一项所述的重组负链RNA病毒基因组,其包含gRNA表达盒,所述gRNA表达盒从5'至3'包含负链RNA病毒转录起始信号、编码所述第一tRNA的核酸、编码所述第一gRNA的核酸、编码第二tRNA的核酸、编码第二gRNA的核酸、编码第三tRNA的核酸和转录终止多聚腺苷酸化信号。

36.如权利要求32-35中任一项所述的重组负链RNA病毒基因组,其中编码所述第一tRNA、第二tRNA和/或第三tRNA的所述核酸是同一的。

37.如权利要求32-35中任一项所述的重组负链RNA病毒基因组,其中编码所述第一tRNA、第二tRNA和/或第三tRNA的所述核酸是不同的。

38.如权利要求32-35中任一项所述的重组负链RNA病毒基因组,其中所述第一tRNA的核苷酸序列和所述第二tRNA的核苷酸序列是至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%,或100%同一的。

39.如权利要求32-35中任一项所述的重组负链RNA病毒基因组,其中所述第一tRNA和所述第二tRNA指定相同的氨基酸。

40.如权利要求32-35中任一项所述的重组负链RNA病毒基因组,其中所述第一tRNA和所述第二tRNA指定不同的氨基酸。

41.如权利要求32-35中任一项所述的重组负链RNA病毒基因组,其中编码所述第一gRNA和/或第二gRNA的所述核酸是同一的。

42.如权利要求32-35中任一项所述的重组负链RNA病毒基因组,其中编码所述第一gRNA和/或第二gRNA的所述核酸是不同的。

43.如权利要求32-35中任一项所述的重组负链RNA病毒基因组,其中所述第一gRNA的核苷酸序列和所述第二gRNA的核苷酸序列是至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%,或100%同一的。

44.如权利要求32-35中任一项所述的重组负链RNA病毒基因组,其中所述第一gRNA和所述第二gRNA与相同的靶核酸序列特异性地杂交。

45.如权利要求32-35中任一项所述的重组负链RNA病毒基因组,其中所述第一gRNA和所述第二gRNA与不同的靶核酸序列特异性地杂交。

46.如权利要求32-35中任一项所述的重组负链RNA病毒基因组,其中所述转录终止多聚腺苷酸化信号包含内源转录终止多聚腺苷酸化信号。

47.如权利要求32-35中任一项所述的重组负链RNA病毒基因组,其中所述转录终止多聚腺苷酸化信号包含异源转录终止多聚腺苷酸化信号。

48.如权利要求1-47中任一项所述的重组负链RNA病毒基因组,其中所述负链RNA病毒

基因组是重组狂犬病毒属基因组。

49.如权利要求48所述的重组负链RNA病毒基因组,其中所述重组狂犬病毒属基因组是重组狂犬病病毒基因组。

50.如权利要求49所述的重组负链RNA病毒基因组,其中所述重组狂犬病病毒基因组包含编码治疗性转基因的核酸,其中:

所述基因组缺乏编码狂犬病病毒糖蛋白或其功能变体的G基因;

所述基因组缺乏编码狂犬病病毒聚合酶或其功能变体的L基因;并且/或者

所述基因组缺乏编码狂犬病病毒基质蛋白或其功能变体的M基因。

51.如权利要求50所述的重组负链RNA病毒基因组,其中所述基因组缺乏编码狂犬病病毒糖蛋白或其功能变体的G基因。

52.如权利要求50所述的重组负链RNA病毒基因组,其中所述基因组缺乏编码狂犬病病毒糖蛋白或其功能变体的G基因,并且其中所述基因组缺乏编码狂犬病病毒聚合酶或其功能变体的L基因。

53.如权利要求50所述的重组负链RNA病毒基因组,其中所述基因组缺乏编码狂犬病病毒糖蛋白或其功能变体的G基因,所述基因组缺乏编码狂犬病病毒聚合酶或其功能变体的L基因,并且所述基因组缺乏编码狂犬病病毒基质蛋白或其功能变异体的M基因。

54.如权利要求50-52中任一项所述的重组负链RNA病毒基因组,其中所述基因组包含:

编码狂犬病病毒核蛋白或其功能变体的N基因;

编码狂犬病病毒磷蛋白或其功能变体的P基因;和

编码狂犬病病毒基质蛋白或其功能变体的M基因。

55.如权利要求1-54中任一项所述的重组负链RNA病毒基因组,其中所述重组负链RNA病毒基因组包含正链反基因组,所述正链反基因组包含:

编码第一引导RNA(gRNA)的核酸,所述编码第一引导RNA(gRNA)的核酸包含5'端和3'端;和

编码第一转移RNA(tRNA)的核酸,所述编码第一转移RNA(tRNA)的核酸位于编码所述第一gRNA的所述核酸的所述5'端或编码所述第一gRNA的所述核酸的所述3'端中的一者或两者处。

56.一种源自权利要求1-54中任一项所述的重组负链RNA病毒基因组的正链反基因组,其中所述正链反基因组包含:

编码第一引导RNA(gRNA)的核酸,所述编码第一引导RNA(gRNA)的核酸包含5'端和3'端;和

编码第一转移RNA(tRNA)的核酸,所述编码第一转移RNA(tRNA)的核酸位于编码所述第一gRNA的所述核酸的所述5'端或编码所述第一gRNA的所述核酸的所述3'端中的一者或两者处。

57.如权利要求56所述的正链反基因组,其中所述正链反基因组由RNA依赖性RNA聚合酶和如权利要求1-54中任一项所述的重组负链RNA病毒基因组合成。

58.一种重组狂犬病病毒颗粒,其包含狂犬病病毒糖蛋白和如权利要求50-54中任一项所述的重组狂犬病病毒基因组。

59.一种重组狂犬病病毒颗粒,其包含:

狂犬病病毒糖蛋白;和

重组狂犬病病毒基因组,所述重组狂犬病病毒基因组包含编码第一引导RNA(gRNA)的核酸,所述编码第一引导RNA(gRNA)的核酸包含5'端和3'端;和

编码第一转移RNA(tRNA)的核酸,所述编码第一转移RNA(tRNA)的核酸位于编码所述第一gRNA的所述核酸的所述3'端或编码所述第一gRNA的所述核酸的所述5'端中的一者或两者处。

60.如权利要求59所述的重组病毒颗粒,其中:

所述基因组缺乏编码狂犬病病毒糖蛋白或其功能变体的G基因;

所述基因组缺乏编码狂犬病病毒聚合酶或其功能变体的L基因;并且/或

所述基因组缺乏编码狂犬病病毒基质蛋白或其功能变体的M基因。

61.如权利要求60所述的重组病毒颗粒,其中所述基因组缺乏编码狂犬病病毒糖蛋白或其功能变体的G基因。

62.如权利要求60所述的重组病毒颗粒,其中所述基因组缺乏编码狂犬病病毒糖蛋白或其功能变体的G基因,并且其中所述基因组缺乏编码狂犬病病毒聚合酶或其功能变体的L基因。

63.如权利要求60所述的重组病毒颗粒,其中所述基因组缺乏编码狂犬病病毒糖蛋白或其功能变体的G基因,所述基因组缺乏编码狂犬病病毒聚合酶或其功能变体的L基因,并且所述基因组缺乏编码狂犬病病毒基质蛋白或其功能变异体的M基因。

64.如权利要求59-63中任一项所述的重组病毒颗粒,其中所述基因组包含:

编码狂犬病病毒核蛋白或其功能变体的N基因;和

编码狂犬病病毒磷蛋白或其功能变体的P基因。

65.如权利要求59-62中任一项所述的重组病毒颗粒,其中所述基因组包含:

编码狂犬病病毒核蛋白或其功能变体的N基因;

编码狂犬病病毒磷蛋白或其功能变体的P基因;和

编码狂犬病病毒基质蛋白或其功能变体的M基因。

66.如权利要求50-54和60-65中任一项所述的重组基因组或病毒颗粒,其中所述基因中的每个基因均可操作地连接至转录调控元件。

66.如权利要求50-54和60-65中任一项所述的重组基因组或病毒颗粒,其中所述转录调控元件包含转录起始信号。

67.如权利要求66所述的重组基因组或病毒颗粒,其中所述转录起始信号对于所述狂犬病病毒来说是外源的。

68.如权利要求66所述的重组基因组或病毒颗粒,其中所述转录起始信号对于所述狂犬病病毒来说是内源的。

69.如权利要求50-54和60-66中任一项所述的重组基因组或病毒颗粒,其中所述基因中的每个基因均可操作地连接至转录终止多聚腺苷酸化信号。

70.如权利要求50-54和60-68中任一项所述的重组基因组或病毒颗粒,其中所述治疗性转基因包括基因编辑系统或基因编辑蛋白。

71.如权利要求70所述的重组基因组或病毒颗粒,其中所述基因编辑系统选自由以下组成的组:成簇规则间隔短回文重复序列(CRISPR)系统、锌指核酸酶(ZFN)、大范围核酸酶

和基于转录激活子样效应物的核酸酶(TALEN)。

72.如权利要求70或71所述的重组基因组或病毒颗粒,其中所述基因编辑系统是CRISPR系统。

73.如权利要求72所述的重组基因组或病毒颗粒,其中所述CRISPR系统包含核碱基编辑器,所述核碱基编辑器包含多核苷酸可编程核苷酸结合结构域和核碱基编辑结构域。

74.如权利要求73所述的重组基因组或病毒颗粒,其中所述核碱基编辑结构域是腺苷脱氨酶、胞苷脱氨酶或其功能变体。

75.如权利要求73或74所述的重组基因组或病毒颗粒,其中所述核碱基编辑结构域是腺苷脱氨酶。

76.如权利要求75所述的重组基因组或病毒颗粒,其中所述腺苷脱氨酶是ABE7.10或ABE8.20。

77.如权利要求73-76中任一项所述的重组基因组或病毒颗粒,其中所述DNA结合结构域是Cas9多肽、Cas12多肽或其功能变体。

78.如权利要求71-77中任一项所述的重组基因组或病毒颗粒,其中所述CRISPR系统进一步包含引导RNA(gRNA)。

79.如权利要求50-54和60-69中任一项所述的重组基因组或病毒颗粒,其中所述治疗性转基因包含治疗性多肽和/或治疗性核酸。

80.如权利要求79所述的重组基因组或病毒颗粒,其中所述治疗性多肽和/或所述治疗性核酸是经分泌的。

81.如权利要求50-54和60-80中任一项所述的重组基因组或病毒颗粒,其中所述治疗性转基因可操作地连接至转录调控元件。

82.如权利要求50-54和60-81中任一项所述的重组基因组或病毒颗粒,其中所述转录调控元件包含转录起始信号。

83.如权利要求82所述的重组基因组或病毒颗粒,其中所述转录起始信号对于所述狂犬病病毒来说是外源的。

84.如权利要求82所述的重组基因组或病毒颗粒,其中所述转录起始信号对于所述狂犬病病毒来说是内源的。

85.如权利要求50-54和60-84中任一项所述的重组基因组或病毒颗粒,其中所述治疗性转基因可操作地连接至转录终止多聚腺苷酸化信号。

86.一种药物组合物,其包含权利要求58-85中任一项所述的重组病毒颗粒。

87.一种用于使治疗性转基因在靶细胞中表达的方法,其包括用权利要求58-85中任一项所述的重组病毒颗粒转导靶细胞。

88.一种用于使核碱基编辑器和引导RNA(gRNA)在靶细胞中表达的方法,其包括用重组狂犬病病毒颗粒转导靶细胞,其中所述重组病毒颗粒包含:

狂犬病病毒糖蛋白;和

重组狂犬病病毒基因组,所述重组狂犬病病毒基因组包含:

编码核碱基编辑器的核酸,所述核碱基编辑器包含多核苷酸可编程核苷酸结合结构域和核碱基编辑结构域;

编码第一gRNA的核酸,所述编码第一gRNA的核酸包含5'端和3'端;和

编码第一tRNA的核酸,所述编码第一tRNA的核酸位于编码所述第一gRNA的所述核酸的所述3’端或编码所述第一gRNA的所述核酸的所述5’端中的一者或两者处。

89.如权利要求88所述的方法,其中:

所述基因组缺乏编码狂犬病病毒糖蛋白或其功能变体的G基因;

所述基因组缺乏编码狂犬病病毒聚合酶或其功能变体的L基因;并且/或

所述基因组缺乏编码狂犬病病毒基质蛋白或其功能变体的M基因。

90.如权利要求88或89所述的方法,其中所述基因组包含:

编码狂犬病病毒核蛋白或其功能变体的N基因;

编码狂犬病病毒磷蛋白或其功能变体的P基因;和

编码狂犬病病毒基质蛋白或其功能变体的M基因。

91.如权利要求88-90中任一项所述的方法,其中所述基因和/或核酸中的每个基因和/或核酸均可操作地连接至转录调控元件。

92.如权利要求91所述的方法,其中所述转录调控元件包含转录起始信号。

93.如权利要求92所述的方法,其中所述转录起始信号对于所述狂犬病病毒来说是外源的。

94.如权利要求92所述的方法,其中所述转录起始信号对于所述狂犬病病毒来说是内源的。

95.如权利要求88-94中任一项所述的方法,其中所述基因和/或核酸中的每个基因和/或核酸均可操作地连接至转录终止多聚腺苷酸化信号。

96.如权利要求88-94中任一项所述的方法,其中所述核碱基编辑结构域是腺苷脱氨酶、胞苷脱氨酶或其功能变体。

97.如权利要求96所述的方法,其中所述碱基编辑器是腺苷脱氨酶。

98.如权利要求96或97所述的方法,其中所述腺苷脱氨酶是ABE7.10或ABE8.20。

99.如权利要求88-98中任一项所述的方法,其中所述DNA结合结构域是Cas9多肽、Cas12多肽或其功能变体。

100.如权利要求88-99中任一项所述的方法,其中所述gRNA能够靶向所述靶细胞的基因组基因座。

101.如权利要求88-100中任一项所述的方法,其中所述靶细胞是经体外转导的。

102.如权利要求101所述的方法,其中所述靶细胞是人细胞。

103.如权利要求101或102所述的方法,其中所述靶细胞是从人获得的。

104.如权利要求101-103中任一项所述的方法,其中所述靶细胞对于所述人来说是自体的。

105.如权利要求101-103中任一项所述的方法,其中所述靶细胞对于所述人来说是同种异体的。

106.如权利要求88-100中任一项所述的方法,其中所述靶细胞是经体内转导的。

107.如权利要求106所述的方法,其中所述靶细胞是人细胞。

108.如权利要求106或107所述的方法,其中所述靶细胞是神经元细胞、上皮细胞或肝细胞。

109.如权利要求106-108中任一项所述的方法,其中所述靶细胞是处于人中。

110.一种用于重组制备狂犬病病毒颗粒的包装系统,其中所述包装系统包括:

编码狂犬病病毒核蛋白或其功能变体的N基因;

编码狂犬病病毒磷蛋白或其功能变体的P基因;

编码狂犬病病毒聚合酶或其功能变体的L基因;和

重组狂犬病病毒基因组,其中:

所述基因组包含编码第一引导RNA(gRNA)的核酸,所述编码第一引导RNA(gRNA)的核酸包含5'端和3'端;并且

所述基因组包含编码第一转移RNA(tRNA)的核酸,所述编码第一转移RNA(tRNA)的核酸位于编码所述第一gRNA的所述核酸的所述3'端或编码所述第一gRNA的所述核酸的所述5'端中的一者或两者处。

111.如权利要求110所述的包装系统,其中:

所述基因组缺乏编码狂犬病病毒糖蛋白或其功能变体的G基因;

所述基因组缺乏编码狂犬病病毒聚合酶或其功能变体的L基因;并且/或

所述基因组缺乏编码狂犬病病毒基质或其功能变体的M基因。

112.如权利要求110或111所述的包装系统,其中所述重组狂犬病病毒基因组进一步包含编码转基因或治疗性转基因的核酸。

113.如权利要求110-112中任一项所述的包装系统,其中所述重组狂犬病病毒基因组包含在病毒基因组载体内。

114.如权利要求110-113中任一项所述的包装系统,其中所述N、P和L基因各自包含在单独的载体内。

115.如权利要求114所述的包装系统,其中所述N、P和L基因中的每一者均可操作地连接至转录调控元件。

116.如权利要求115所述的包装系统,其中所述转录调控元件包含启动子和/或增强子。

117.如权利要求116所述的包装系统,其中所述启动子是组成型启动子。

118.如权利要求116或117所述的包装系统,其中所述启动子是延伸因子1α启动子。

119.如权利要求113-118中任一项所述的包装系统,其中所述单独的载体各自包含在单独的转染质粒内。

120.如权利要求113-119中任一项所述的包装系统,其中所述N、P和L基因包含在单个载体内。

121.如权利要求120所述的包装系统,其中所述单个载体包含含有所述N和P基因的第一表达盒和含有所述L基因的第二表达盒。

122.如权利要求121所述的包装系统,其中所述第一表达盒从5′至3′包含:

转录调控元件;

所述P基因;和

所述N基因。

123.如权利要求121或122所述的包装系统,其中所述第一表达盒从5′至3′包含:

转录调控元件;

所述P基因;

核糖体跳跃元件;和

所述N基因。

124.如权利要求123所述的包装系统,其中所述核糖体跳跃元件是IRES元件。

125.如权利要求123所述的包装系统,其中所述核糖体跳跃元件是2A元件。

126.如权利要求121-125中任一项所述的包装系统,其中所述第二表达盒从5′至3′包含:

转录调控元件;和

所述L基因。

127.如权利要求122-126中任一项所述的包装系统,其中所述转录调控元件包含启动子和/或增强子。

128.如权利要求127所述的包装系统,其中所述启动子是组成型启动子。

129.如权利要求127或128所述的包装系统,其中所述启动子是延伸因子1α启动子。

130.如权利要求121-129中任一项所述的包装系统,其中所述第一表达盒和所述第二表达盒在所述载体中处于相反的取向。

131.如权利要求122-130中任一项所述的包装系统,其中所述单个载体包含在单个转染质粒内。

132.如权利要求110-131中任一项所述的包装系统,其进一步包含编码狂犬病病毒基质蛋白或其功能变体的M基因。

133.如权利要求132所述的包装系统,其中所述M基因包含在载体内。

134.如权利要求132或133所述的包装系统,其中所述M基因可操作地连接至转录调控元件。

135.如权利要求134所述的包装系统,其中所述转录调控元件包含启动子和/或增强子。

136.如权利要求133-135中任一项所述的包装系统,其中包含所述M基因的所述载体包含在转染质粒内。

137.如权利要求110-136中任一项所述的包装系统,其进一步包含编码狂犬病病毒糖蛋白或其功能变体的G基因。

138.如权利要求137所述的包装系统,其中所述G基因包含在载体内。

139.如权利要求137或138所述的包装系统,其中所述G基因可操作地连接至转录调控元件。

140.如权利要求139所述的包装系统,其中所述转录调控元件包含启动子和/或增强子。

141.如权利要求137-140中任一项所述的包装系统,其中包含所述G基因的所述载体包含在转染质粒内。

142.一种用于产生重组狂犬病病毒颗粒的方法,所述方法包括将权利要求110-141中任一项所述的包装系统在可有效包裹重组狂犬病病毒基因组的条件下引入到细胞中以形成所述重组狂犬病病毒颗粒。

143.如权利要求142所述的方法,其中所述引入是由电穿孔、核转染或脂转染介导的。

144.一种重组狂犬病病毒颗粒包装细胞,其包含权利要求110-143中任一项所述的包

装系统。

145.一种治疗受试者的疾病或病症的方法,所述方法包括向所述受试者施用权利要求58-85中任一项所述的重组狂犬病病毒颗粒或权利要求86所述的药物组合物。

146.如权利要求145所述的方法,其中所述疾病或病症是神经疾病或病症。

147.如权利要求145所述的方法,其中所述疾病或病症是眼科疾病或病症。

148.权利要求58-85中任一项所述的重组狂犬病病毒或权利要求86所述的药物组合物在制造用于治疗受试者的疾病或病症的药物中的用途。

病毒引导RNA递送

[0001] 相关申请

[0002] 本申请要求2021年9月8日提交的美国临时专利申请序列号63/241,964的权益,该临时专利申请的全部公开内容特此以引用的方式并入本文。

背景技术

[0003] 传统上,基于病毒的引导RNA(gRNA)递送是用DNA病毒(例如腺病毒)介导的,其中所述gRNA是从DNA病毒基因组转录而来。这些系统可以利用良好表征的表达系统,如U6(Pol III启动子)-或T7体外系统。然而,用负链RNA病毒(如狂犬病病毒)递送gRNA的实例有限,并且用带负链RNA病毒的侧接tRNA递送gRNA还没有被报道。

[0004] 负链RNA病毒gRNA的递送呈现出独特的挑战。负链RNA病毒在其生命周期中没有DNA阶段,因此不能使用基于DNA的启动子。负链RNA病毒基因组中的每个转录盒均由RNA依赖性RNA聚合酶(RdRp)读取。产生的转录物总是具有5'帽和polyA尾,这可能会干扰gRNA的活性。

[0005] 因此,需要比当前病毒系统具有优势的新型病毒gRNA递送系统。

发明内容

[0006] 本文提供了重组负链RNA病毒基因组(例如重组狂犬病病毒基因组)和包含所述重组负链RNA病毒基因组的重组病毒颗粒(例如重组狂犬病病毒颗粒),其可用于转导靶细胞和在其中表达引导RNA(gRNA)。本公开提供的重组RNA病毒基因组和病毒可用作有效的病毒gRNA和转基因(例如,核碱基编辑器)递送系统。还提供了病毒包装系统和产生本文所述的重组病毒的方法。

[0007] 在一方面,本发明提供了一种重组负链RNA病毒基因组,其包含编码第一引导RNA(gRNA)的核酸,该编码第一引导RNA(gRNA)的核酸包含5'端和3'端;和编码第一转移RNA(tRNA)的核酸,该编码第一转移RNA(tRNA)的核酸位于编码第一gRNA的核酸的3'端或编码第一gRNA的核酸的5'端中的一者或两者处。

[0008] 在某些实施方案中,重组负链RNA病毒基因组包含编码第二tRNA的核酸。

[0009] 在某些实施方案中,编码第一tRNA的核酸位于编码第一gRNA的核酸的3'端;并且编码第二tRNA的核酸位于编码第一gRNA的核酸的5'端。

[0010] 在某些实施方案中,第一tRNA的核苷酸序列和第二tRNA的核苷酸序列是至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%,或100%同一的。

[0011] 在某些实施方案中,第一tRNA和第二tRNA指定相同的氨基酸。在某些实施方案中,第一tRNA和第二tRNA指定不同的氨基酸。

[0012] 在某些实施方案中,重组负链RNA病毒基因组包含编码第一tRNA的两种核酸。在某些实施方案中,重组负链RNA病毒基因组包含编码第一tRNA的三种核酸。

[0013] 在某些实施方案中,重组负链RNA病毒基因组包含编码第二gRNA的核酸。在某些实施方案中,该两种或更多种核酸编码同一的gRNA。在某些实施方案中,两种或更多种核酸编

码至少一种不同的gRNA。在某些实施方案中,第一gRNA的核苷酸序列和第二gRNA的核苷酸序列是至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%,或100%同一的。在某些实施方案中,第一gRNA和第二gRNA与相同的靶核酸序列特异性地杂交。在某些实施方案中,第一gRNA和第二gRNA与不同的靶核酸序列特异性地杂交。

[0014] 在某些实施方案中,第一tRNA和/或第二tRNA各自选自由以下组成的组:tRNA-ala、tRNA-arg、tRNA-asn、tRNA-asp、tRNA-cys、tRNA-gln、tRNA-gly、tRNA-his、tRNA-ile、tRNA-leu、tRNA-lys、tRNA-met、tRNA-phe、tRNA-pro、tRNA-pyl、tRNA-sec、tRNA-ser、tRNA-thr、tRNA-trp、tRNA-tyr和tRNA-val。

[0015] 在某些实施方案中,编码第一tRNA和/或第二tRNA的核酸包含以下中的任一种:

[0016] GGCTCGTTGGTCTAGGGGTATGATTCTCGCTTAGGGTGCG AGAGGTCCCGGGTTCAAATCCCGGACG AGCCC(tRNA-pro;SEQ ID NO:4011),或与其至少90%同一的序列;

[0017] GGCTCCATAGCTCAGGGGTTAGAGCACTGGTCTTGTAAAC CAGGGGTCGCGAGTTCAATTCTCGCTG GGGCTT(tRNA-thr;SEQ ID NO:4012),或与其至少90%同一的序列;

[0018] GCGTTGGTGGTATAGTGGTGAGCATAGCTGCCTTCCAAGC AGTTGACCCGGGTTCGATTCCCGGCCA ACGCA(tRNA-gly G8;SEQ ID NO:4013),或与其至少90%同一的序列;

[0019] G C A T G G G T G G T T C A G T G G T A G A A T T C T C G C C T G C C A C G C G GGAGGCCCGGGTTCGATTCCCGGCCCATGCA(tRNA-gly G27;SEQ ID NO:4014),或与其至少90%同一的序列;

[0020] GTCAGGATGGCCGAGCGGTCTAAGGCGCTGCGTTCAGGTC GCAGTCTCCCCTAGAGGCGTGGGTTCG AATCCCACTCCTGACA(tRNA-leu;SEQ ID NO:4015),或与其至少90%同一的序列;

[0021] GCTCCAGTGGCGCAATCGGTTAGCGCGCGGTACTTATAAG ACAGTGCACCTGTGAGCAATGCCGAGG TTGTGAGTTCAAGCCT CACCTGGAGCA(tRNA-ile;SEQ ID NO:4016),或与其至少90%同一的序列;

[0022] GAAAAAGTCATGGAGGCCATGGGGTTGGCTTGAAACCAG CTTTGGGGGGTTCGATTCCTTCCTTTTT TGTCT(tRNA-ser;SEQ ID NO:4017),或与其至少90%同一的序列;

[0023] GGGCCAGTGGCGCAATGGATAACGCGTCTGACTACGGATC AGAAGATTCCAGGTTCGACTCCTGGCT GGCTCGGTGTA(tRNA-arg;SEQ ID NO:4018),或与其至少90%同一的序列;

[0024] AAACAAGCGCAAGTGGTTTAGTGGTAAAATCCAACGTTGC CATCGTTGGGCCCCCGGTTCGATTCCG GGCTTGCGCA(tRNA-asp1;SEQ ID NO:4019),或与其至少90%同一的序列;

[0025] AACAAAGCACCAGTGGTCTAGTGGTAGAATAGTACCCTGC CACGGTACAGACCCGGGTTCGATTCCC GGCTGGTGCA(tRNA-asp2;SEQ ID NO:4020),或与其至少90%同一的序列;或

[0026] T C C T C G T T A G T A T A G T G G T G A G T A T C C C C G C C T G T C A C G C G GGAGACCGGGGTTCGATTCCCCGACGGGGAG(tRNA-asp D15;SEQ ID NO:4021),或与其至少90%同一的序列。

[0027] 在某些实施方案中,第一tRNA和/或第二tRNA包含tRNA样结构。

[0028] 在某些实施方案中,tRNA样结构包含MALAT1相关的小细胞质RNA(mascRNA)。

[0029] 在某些实施方案中,mascRNA由包含以下中的任一种的核酸编码:

[0030] AAAAGCAAAAGATGCTGGTGGTTGGCACTCCTGGTTTCCA GGACGGGGTTCAAATCCCTGCGGCGTC TTTGCTTT(masc_Malat1;SEQ ID NO:X),或与其至少90%同一的序列;

[0031] AAAGACGCTGGTGGTTGGTGTTTCCAGGACGGGGTTCAAG TCCCTGCGGCGTCCTCGC(masc_liz38;SEQ ID NO:X),或与其至少90%同一的序列;

[0032] GGCTCTGGTGGCTTCCAGGACGGGGTTCAAGTCCCTGCAG TGCCCTTGCTGA(masc_liz40;SEQ ID NO:X),或与其至少90%同一的序列;

[0033] AAAGGCGCTGGTGGTGGCACTCCCAGCGGGACGGGGTTCG AATCCCCGCGGCGCCTCTGC(masc_turk;SEQ ID NO:X),或与其至少90%同一的序列;

[0034] GAAGGTTTTTCTTTTCCTGAGAAAACAACACGTATTGTTTTCTCAGGTTTTGCTTTTTGGCCTTTTTCTAGCTTAAAAAAAAAAGCAAAAGATGCTGGTGGTTGGCACTCCTGGTTTCCAGGACGGGGTTCAAATCCCTGCGGCGTCTTTGCTTT(hMALAT1.1;SEQ ID NO:X),或与其至少90%同一的序列;

[0035] GCAGGTGTTTCTTTTACTGAGTGCAGCCCATGGCCGCACTCAGGTTTTGCTTTTCACCTTCCCATCTGTGAAAGAGTGAGCAGGAAAAAGCAAAAGGCGCTGGTGGTGGCACGTCCAGCACGGCTGGGCCGGGGTTCGAGTCCCCGCAGTGTTGCTGC(hMALAT1.2;SEQ ID NO:X),或与其至少90%同一的序列;

[0036] GAAGGTTTTTCTTTTCCTGAGAAAACAACACGTTTTGTTTTCTCAGGTTTTGCTTTTTGGCCTTTTTCTAGCTTAAAAAAAAAAGCAAAAGATGCTGGTGGTTGGCACTCCTGGTTTCCAGGACAGGGTTCAAATCCCTGCGGCGTCTTTGCTTT(黑猩猩1;SEQ ID NO:X),或与其至少90%同一的序列;

[0037] AAAGCAAAAGATGCTGGTGGTTGGCACTCCTGGTTTCCAG GACAGGGTTCAAATCCCTGCGGCGTCTTTGCTTT(黑猩猩1短;SEQ ID NO:X),或与其至少90%同一的序列;

[0038] GCAGGTGTTTCTTTTACTGAGTGCAGCCCATGGCCGCACTCAGGTTTTGCTTTTCACCTTCCCATCTGTGAAAGAGTGAGCAGGAAAAAGCAAAAGGCGCTGGTGGTGGCACGTCCAGCACGGCTGGGCCGGGGTTCGAGTCCCCGCAGTGTTGCTGC(chimp.2;SEQ ID NO:X),或与其至少90%同一的序列;

[0039] AAAGGTTTTTCTTTTCCTGAGAAAACAACCTTTTGTTTTCTCAGGTTTTGCTTTTTGGCCTTTCCCTAGCTTTAAAAAAAAAAGCAAAAGACGCTGGTGGCTGGCACTCCTGGTTTCCAGGACGGGGTTCAAGTCCCTGCGGTGTCTTTGC(MoTse.1;SEQ ID NO:X),或与其至少90%同一的序列;

[0040] AAAGCAAAAGACGCTGGTGGCTGGCACTCCTGGTTTCCAG GACGGGGTTCAAGTCCCTGCGGTGTCTTTGCTTGAC(MoTse.1短;SEQ ID NO:X),或与其至少90%同一的序列;或

[0041] GCAGGTGTTTCTTTTCCTGACCGCGGCTCATGGCCGCGCTCAGGTTTTGCTTTTCACCTTTGTCTGAGAGAACGAACGTGAGCAGGAAAAAGCAAAAGGCACTGGTGGCGGCACGCCCGCACCTCGGGCCAGGGTTCGAGTCCCTGCAGTACCGTGC(MoTse.2;SEQ ID NO:X),或与其至少90%同一的序列。

[0042] 在某些实施方案中,tRNA样结构包含tRNA变体。

[0043] 在某些实施方案中,tRNA变体包含用G或C核苷酸对一个或多个A和/或T核苷酸进行的取代。

[0044] 在某些实施方案中,tRNA变体包含相对于野生型tRNA较低的A和/或T核苷酸含量。

[0045] 在某些实施方案中,tRNA变体由包含以下中的任一种的核酸编码:

[0046] G G C T C G T T G G C C T A G G G G T A T G G C T C C C G C T T A G G G T G C GGGAGGTCCCGGGTTCAAATCCCGGACGAGCC(tRNA-pro var1;SEQ ID NO:X),或与其至少90%同一的序列;

[0047] GGCTCGTTGGCCTAGGGGTATGGCTGAAAAGGTCCCGGGT TCAAATCCCGGACGAGCC(tRNA-pro var2;SEQ ID NO:X),或与其至少90%同一的序列;

[0048] GGCTCGTTGAAAGAAAAGGTCCCGGGTTCAAATCCCGGAC GAGCC(tRNA-pro var3;SEQ ID

NO:X),或与其至少90%同一的序列;

[0049] GGCTCCATAGCGCAGGGGTTAGCGCACCGGTCTTGTAAAC CGGGGGTCGCGAGTTCAATTCTCGCTGGGGCTT(tRNA-thr var1;SEQ ID NO:X),或与其至少90%同一的序列;

[0050] GGCTCCATAGCGCAGGGGTTAGCGCAGAAAGGGTCGCGA GTTCAATTCTCGCTGGGGCTT(tRNA-thr var2;SEQ ID NO:X),或与其至少90%同一的序列;或

[0051] GGCTCCATAGAAAGAAAGAAAGGGTCGCGAGTTCAATTCT CGCTGGGGCTT(tRNA-thr var3;SEQ ID NO:X),或与其至少90%同一的序列。

[0052] 在某些实施方案中,tRNA样结构包含tRNA片段。

[0053] 在某些实施方案中,tRNA样结构包含病毒tRNA样结构(vtRNA)。

[0054] 在某些实施方案中,vtRNA由包含以下中的任一种的核酸编码:

[0055] GCCAGAGTAGCTCAATTGGTAGAGCAACAGGTCACCGATCCTGGTGGTTCTCGGTTCAAGTCCGAGCTCTGGTC(vtRNA-1;SEQID NO:X),或与其至少90%同一的序列;

[0056] GCCAGGGTAGCTCAATCGGTAGAGCAGCGGTTCCTGGAGTCCGCTGGTTCTCGGTTCAAGCCCGAGCCCTGGTTG(vtRNA-2;SEQ ID NO:X),或与其至少90%同一的序列;

[0057] GTCGGGGTAGCTCAAATGGTAGAGTGGCAGGCCAACATAGCCAGCAGATCTCGGTTCAAACCCGAGCCCTGACCA(vtRNA-3;SEQ ID NO:X),或与其至少90%同一的序列;

[0058] GTCGGGGTAGCTCAATTGGTAGAGCGGCAGGCTCATCCCCTGCAGGTTCTCGGTTCAATCCCGGGTCCCGACGC(vtRNA-4;SEQ ID NO:X),或与其至少90%同一的序列;

[0059] GCCAGGGTAGCTCAATTGGTAGAGCATCAGGCTAGTATCCTGTCGGTTCCGGTTCAAGTCCGGGCCCTGGTTA(vtRNA-5;SEQ ID NO:X),或与其至少90%同一的序列;

[0060] GCCAGCGTAGCTCAATTGTTAGAGCAGCGGCCACCAAGCCTGCAGGTTCTCGGTTCAAGTCCGGGCGCTGGCAT(vtRNA-6;SEQ ID NO:X),或与其至少90%同一的序列;

[0061] GCCAGGGTAGCTCAATTGGTAGAGCGGCAGACACCACCTACGTGGTCTAGTCTGTGGATCTCGGTTCAAGTCCGAGTCCTGGCCA(vtRNA-7;SEQ ID NO:X),或与其至少90%同一的序列;

[0062] GCCAGGGTAGCTCAATTGGTAGAGCGGCAGACACCACCTACGTGGTCTAGTCTGTGGATCTCGGTTCAAGTCCGAGTCCTGGCCA(vtRNA-7;SEQ ID NO:X),或与其至少90%同一的序列;或

[0063] ACCAGAGTGGCTCACCTGGTAGAGCACCAGGCTGCCCATCCTGTTGGTTCTCGGTTCAAATCCGAGCTCTGGTGA(vtRNA-8;SEQ ID NO:X),或与其至少90%同一的序列。

[0064] 在某些实施方案中,重组负链RNA病毒基因组包含编码负链RNA病毒基因的核酸。

[0065] 在某些实施方案中,重组负链RNA病毒基因组进一步包含编码转基因的核酸。

[0066] 在某些实施方案中,编码第一gRNA的核酸和编码第一tRNA的核酸位于各自编码负链RNA病毒基因的两种核酸之间。

[0067] 在某些实施方案中,编码第一gRNA的核酸和编码第一tRNA的核酸位于各自编码转基因的两种核酸之间。

[0068] 在某些实施方案中,编码第一gRNA的核酸和编码第一tRNA的核酸位于编码负链RNA病毒基因的核酸和编码转基因的核酸之间。

[0069] 在某些实施方案中,重组负链RNA病毒基因组包含gRNA表达盒,该gRNA表达盒从3’至5’包含负链RNA病毒转录起始信号、编码tRNA的核酸、编码gRNA的核酸、和转录终止多聚腺苷酸化信号。

[0070] 在某些实施方案中,重组负链RNA病毒基因组包含gRNA表达盒,该gRNA表达盒从3'至5'包含负链RNA病毒转录起始信号、编码第一tRNA的核酸、编码第一gRNA的核酸、编码第二tRNA的核酸和转录终止多聚腺苷酸化信号。

[0071] 在某些实施方案中,重组负链RNA病毒基因组包含gRNA表达盒,该gRNA表达盒从3'至5'包含负链RNA病毒转录起始信号、编码第一tRNA的核酸、编码第一gRNA的核酸、编码第二tRNA的核酸、编码第二gRNA的核酸和转录终止多聚腺苷酸化信号。

[0072] 在某些实施方案中,重组负链RNA病毒基因组包含gRNA表达盒,该gRNA表达盒从3'至5'包含负链RNA病毒转录起始信号、编码第一tRNA的核酸、编码第一gRNA的核酸、编码第二tRNA的核酸、编码第二gRNA的核酸和转录终止多聚腺苷酸化信号。

[0073] 在某些实施方案中,重组负链RNA病毒基因组包含gRNA表达盒,该gRNA表达盒从3'至5'包含负链RNA病毒转录起始信号、编码第一tRNA的核酸、编码第一gRNA的核酸、编码第二tRNA的核酸、编码第二gRNA的核酸、编码第三tRNA的核酸和转录终止多聚腺苷酸化信号。

[0074] 在gRNA表达盒的某些实施方案中,编码第一tRNA、第二tRNA和/或第三tRNA的核酸是同一的。在gRNA表达盒的某些实施方案中,编码第一tRNA、第二tRNA和/或第三tRNA的核酸是不同的。在gRNA表达盒的某些实施方案中,第一tRNA的核苷酸序列和第二tRNA的核苷酸序列是至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%,或100%同一的。在gRNA表达盒的某些实施方案中,第一tRNA和第二tRNA指定相同的氨基酸。在gRNA表达盒的某些实施方案中,第一tRNA和第二tRNA指定不同的氨基酸。在gRNA表达盒的某些实施方案中,编码第一gRNA和/或第二gRNA的核酸是同一的。在gRNA表达盒的某些实施方案中,编码第一gRNA和/或第二gRNA的核酸是不同的。在gRNA表达盒的某些实施方案中,第一gRNA的核苷酸序列和第二gRNA的核苷酸序列是至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%,或100%同一的。在gRNA表达盒的某些实施方案中,第一gRNA和第二gRNA与相同的靶核酸序列特异性地杂交。在gRNA表达盒的某些实施方案中,第一gRNA和第二gRNA与不同的靶核酸序列特异性地杂交。在gRNA表达盒的某些实施方案中,转录终止多聚腺苷酸化信号包含内源转录终止多聚腺苷酸化信号。在gRNA表达盒的某些实施方案中,转录终止多聚腺苷酸化信号包含异源转录终止多聚腺苷酸化信号。

[0075] 在某些实施方案中,负链RNA病毒基因组是重组狂犬病毒属基因组。

[0076] 在某些实施方案中,重组狂犬病毒属基因组是重组狂犬病病毒基因组。

[0077] 在一方面,本发明提供了一种重组负链RNA病毒基因组,其包含编码第一引导RNA(gRNA)的核酸,该编码第一引导RNA(gRNA)的核酸包含5'端和3'端;编码第一转移RNA(tRNA)的核酸,该编码第一转移RNA(tRNA)的核酸位于编码第一gRNA的核酸的3'端或编码第一gRNA的核酸的5'端中的一者或两者处;和编码转基因(例如治疗性转基因)的核酸。

[0078] 在某些实施方案中,该转基因包含核碱基编辑器。

[0079] 在某些实施方案中,重组狂犬病病毒基因组包含编码治疗性转基因的核酸,其中:该基因组缺乏编码狂犬病病毒糖蛋白或其功能变体的G基因;并且/或者该基因组缺乏编码狂犬病病毒聚合酶或其功能变体的L基因。在某些实施方案中,该基因组缺乏编码狂犬病病毒糖蛋白或其功能变体的G基因。在某些实施方案中,该基因组缺乏编码狂犬病病毒糖蛋白或其功能变体的G基因,并且其中该基因组缺乏编码狂犬病病毒聚合酶或其功能变体的L基因。

[0080] 在某些实施方案中,该基因组包含:编码狂犬病病毒核蛋白或其功能变体的N基因;编码狂犬病病毒磷蛋白或其功能变体的P基因;和编码狂犬病病毒基质蛋白或其功能变体的M基因。

[0081] 在一方面,本公开提供了由上文所述的重组负链RNA病毒基因组表达的信使RNA(mRNA)。

[0082] 在某些实施方案中,mRNA包含第一引导RNA(gRNA),该第一引导RNA(gRNA)包含5'端和3'端;和第一转移RNA(tRNA),该第一转移RNA(tRNA)位于第一gRNA的3'端或第一gRNA的5'端中的一者或两者处。

[0083] 在另一方面,本公开提供了重组狂犬病病毒颗粒,其包含狂犬病病毒糖蛋白和上文所述的重组狂犬病病毒基因组。

[0084] 在另一个方面,本公开提供了一种重组狂犬病病毒颗粒,其包含:狂犬病病毒糖蛋白;和重组狂犬病病毒基因组,该重组狂犬病病毒基因组包含编码第一引导RNA(gRNA)的核酸,该编码第一引导RNA(gRNA)的核酸包含5'端和3'端;和编码第一转移RNA(tRNA)的核酸,该编码第一转移RNA(tRNA)的核酸位于编码第一gRNA的核酸的3'端或编码第一gRNA的核酸的5'端中的一者或两者处。

[0085] 在某些实施方案中,该基因组缺乏编码狂犬病病毒糖蛋白或其功能变体的G基因;并且/或者该基因组缺乏编码狂犬病病毒聚合酶或其功能变体的L基因。在某些实施方案中,该基因组缺乏编码狂犬病病毒糖蛋白或其功能变体的G基因。在某些实施方案中,该基因组缺乏编码狂犬病病毒糖蛋白或其功能变体的G基因,并且其中该基因组缺乏编码狂犬病病毒聚合酶或其功能变体的L基因。

[0086] 在某些实施方案中,该基因组包含:编码狂犬病病毒核蛋白或其功能变体的N基因;编码狂犬病病毒磷蛋白或其功能变体的P基因;和编码狂犬病病毒基质蛋白或其功能变体的M基因。

[0087] 在某些示实施方案中,基因中的每个基因均可操作地连接至转录调控元件。在某些实施方案中,转录调控元件包含转录起始信号。在某些实施方案中,转录起始信号对于狂犬病病毒来说是外源的。在某些实施方案中,转录起始信号对于狂犬病病毒来说是内源的。

[0088] 在某些实施方案中,基因中的每个基因均可操作地连接至转录终止多聚腺苷酸化信号。

[0089] 在某些实施方案中,治疗性转基因包含基因编辑系统或基因编辑蛋白质。

[0090] 在某些实施方案中,基因编辑系统选自由以下组成的组:成簇规则间隔短回文重复序列(CRISPR)系统、锌指核酸酶(ZFN)、大范围核酸酶和基于转录激活子样效应物的核酸酶(TALEN)。在某些实施方案中,基因编辑系统是CRISPR系统。

[0091] 在某些实施方案中,CRISPR系统包含核碱基编辑器,该核碱基编辑器包含多核苷酸可编程核苷酸结合结构域和核碱基编辑结构域。

[0092] 在某些实施方案中,核碱基编辑结构域是腺苷脱氨酶、胞苷脱氨酶或其功能变体。在某些实施方案中,核碱基编辑结构域是腺苷脱氨酶。在某些实施方案中,腺苷脱氨酶是ABE7.10或ABE8.20。

[0093] 在某些实施方案中,DNA结合结构域是Cas9多肽、Cas12多肽或其功能变体。

[0094] 在某些实施方案中,CRISPR系统进一步包含引导RNA(gRNA)。

[0095] 在某些实施方案中,治疗性转基因包含治疗性多肽和/或治疗性核酸。

[0096] 在某些实施方案中,治疗性多肽和/或治疗性核酸是经分泌的。

[0097] 在某些实施方案中,治疗性转基因可操作地连接至转录调控元件。在某些实施方案中,转录调控元件包含转录起始信号。在某些实施方案中,转录起始信号对于狂犬病病毒来说是外源的。在某些实施方案中,转录起始信号对于狂犬病病毒来说是内源的。在某些实施方案中,治疗性转基因可操作地连接至转录终止多聚腺苷酸化信号。

[0098] 在一方面,本公开提供了一种药物组合物,其包含如上文所述的重组病毒颗粒。

[0099] 在一方面,本公开提供了一种用于使治疗性转基因在靶细胞中表达的方法,其包括用如上文所述的重组病毒颗粒转导靶细胞。

[0100] 在一方面,本公开提供了一种用于使核碱基编辑器和引导RNA(gRNA)在靶细胞中表达的方法,其包括用重组狂犬病病毒颗粒转导靶细胞,其中该重组病毒颗粒包含:狂犬病病毒糖蛋白;和重组狂犬病病毒基因组,该重组狂犬病病毒基因组包含:编码包含多核苷酸可编程核苷酸结合结构域和核碱基编辑结构域的核碱基编辑器的核酸;编码第一gRNA的核酸,该编码第一gRNA的核酸包含5'端和3'端;和编码第一tRNA的核酸,该编码第一tRNA的核酸位于编码第一gRNA的核酸的3'端或编码第一gRNA的核酸的5'端中的一者或两者处。

[0101] 在该方法的某些实施方案中,该基因组缺乏编码狂犬病病毒糖蛋白或其功能变体的G基因;并且/或者该基因组缺乏编码狂犬病病毒聚合酶或其功能变体的L基因。

[0102] 在该方法的某些实施方案中,该基因组包含:编码狂犬病病毒核蛋白或其功能变体的N基因;编码狂犬病病毒磷蛋白或其功能变体的P基因;和编码狂犬病病毒基质蛋白或其功能变体的M基因。

[0103] 在该方法的某些示实施方案中,基因和/或核酸中的每个基因和/或核酸均可操作地连接至转录调控元件。在该方法的某些实施方案中,转录调控元件包含转录起始信号。在该方法的某些实施方案中,转录起始信号对于狂犬病病毒来说是外源的。在该方法的某些实施方案中,该转录起始信号对于狂犬病病毒来说是内源的。在该方法的某些实施方案中,基因和/或核酸中的每个基因和/或核酸均可操作地连接至转录终止多聚腺苷酸化信号。

[0104] 在该方法的某些实施方案中,核碱基编辑结构域是腺苷脱氨酶、胞苷脱氨酶或其功能变体。

[0105] 在该方法的某些实施方案中,碱基编辑器是腺苷脱氨酶。在该方法的某些实施方案中,腺苷脱氨酶是ABE7.10或ABE8.20。

[0106] 在该方法的某些实施方案中,DNA结合结构域是Cas9多肽、Cas12多肽或其功能变体。

[0107] 在该方法的某些示例性实施方案中,gRNA能够靶向靶细胞的基因组基因座。

[0108] 在该方法的某些实施方案中,靶细胞是经离体转导的。在该方法的某些实施方案中,靶细胞是人细胞。在该方法的某些实施方案中,靶细胞是从人获得的。在该方法的某些实施方案中,靶细胞对于该人来说是自体的。在该方法的某些实施方案中,靶细胞对于该人来说是同种异体的。

[0109] 在该方法的某些实施方案中,靶细胞是经体内转导的。在该方法的某些实施方案中,靶细胞是人细胞。在该方法的某些实施方案中,靶细胞是神经元细胞、上皮细胞或肝细胞。在该方法的某些实施方案中,靶细胞是处于人中。

[0110] 在一方面,本公开提供了一种用于重组制备狂犬病病毒颗粒的包装系统,其中该包装系统包含:编码狂犬病病毒核蛋白或其功能变体的N基因;编码狂犬病病毒磷蛋白或其功能变体的P基因;编码狂犬病病毒聚合酶或其功能变体的L基因;和重组狂犬病病毒基因组,其中:该基因组包含编码第一引导gRNA(gRNA)的核酸,该编码第一引导gRNA(gRNA)的核酸包含5'端和3'端;并且该基因组包含编码第一转移RNA(tRNA)的核酸,该编码第一转移RNA(tRNA)的核酸位于编码第一gRNA的核酸的3'端或编码第一gRNA的核酸的5'端中的一者或两者处。

[0111] 在该包装系统的某些实施方案中,该基因组缺乏编码狂犬病病毒糖蛋白或其功能变体的G基因;并且/或者该基因组缺乏编码狂犬病病毒聚合酶或其功能变体的L基因。

[0112] 在该包装系统的某些实施方案中,重组狂犬病病毒基因组进一步包含编码转基因或治疗性转基因的核酸。

[0113] 在该包装系统的某些实施方案中,重组狂犬病病毒基因组被包含在病毒基因组载体内。

[0114] 在该包装系统的某些实施方案中,N、P和L基因各自被包含在单独的载体内。

[0115] 在该包装系统的某些实施方案中,N、P和L基因中的每一者均可操作地连接至转录调控元件。在该包装系统的某些实施方案中,转录调控元件包含启动子和/或增强子。在该包装系统的某些实施方案中,启动子是组成型启动子。在该包装系统的某些实施方案中,启动子是延伸因子1α启动子。

[0116] 在该包装系统的某些实施方案中,单独的载体各自被包含在单独的转染质粒内。

[0117] 在该包装系统的某些实施方案中,N、P和L基因被包含在单个载体内。

[0118] 在该包装系统的某些实施方案中,单个载体包含含有N和P基因的第一表达盒和含有L基因的第二表达盒。

[0119] 在该包装系统的某些实施方案中,第一表达盒从5'至3'包含:转录调控元件、P基因,和N基因。

[0120] 在该包装系统的某些实施方案中,第一表达盒从5'至3'包含:转录调控元件、P基因、核糖体跳跃元件,和N基因。

[0121] 在该包装系统的某些实施方案中,核糖体跳跃元件是IRES元件。在该包装系统的某些实施方案中,核糖体跳跃元件是2A元件。

[0122] 在该包装系统的某些实施方案中,第二表达盒从5'至3'包含:转录调控元件,和L基因。

[0123] 在该包装系统的某些实施方案中,转录调控元件包含启动子和/或增强子。在该包装系统的某些实施方案中,启动子是组成型启动子。在该包装系统的某些实施方案中,启动子是延伸因子1α启动子。

[0124] 在该包装系统的某些实施方案中,第一和第二表达盒在载体中处于相反的取向。

[0125] 在该包装系统的某些实施方案中,单个载体被包含在单个转染质粒内。

[0126] 在该包装系统的某些实施方案中,包装系统进一步包含编码狂犬病病毒基质蛋白或其功能变体的M基因。在该包装系统的某些实施方案中,M基因被包含在载体内。在该包装系统的某些实施方案中,M基因可操作地连接至转录调控元件。在该包装系统的某些实施方案中,转录调控元件包含启动子和/或增强子。在该包装系统的某些实施方案中,包含M基因

的载体被包含在转染质粒内。

[0127] 在该包装系统的某些实施方案中,包装系统进一步包含编码狂犬病病毒糖蛋白或其功能变体的G基因。在该包装系统的某些实施方案中,G基因被包含在载体内。在该包装系统的某些实施方案中,G基因可操作地连接至转录调控元件。在该包装系统的某些实施方案中,转录调控元件包含启动子和/或增强子。在该包装系统的某些实施方案中,包含G基因的载体被包含在转染质粒内。

[0128] 在一方面,本公开提供了一种用于产生重组狂犬病病毒颗粒的方法,该方法包括将如上文所述的包装系统在可有效包裹重组狂犬病病毒基因组的条件下引入到细胞中以形成重组狂犬病病毒颗粒。

[0129] 在该方法的某些实施方案中,引入是通过电穿孔、核转染或脂转染介导的。

[0130] 在一方面,本公开提供了一种重组狂犬病病毒颗粒包装细胞,其包含如上文所述的包装系统。

[0131] 在一方面,本公开提供了治疗受试者的疾病或病症的方法,该方法包括向受试者施用上文所述的重组狂犬病病毒颗粒或上文所述的药物组合物。在该方法的某些实施方案中,疾病或病症是神经疾病或病症。在该方法的某些实施方案中,疾病或病症是眼科疾病或病症。

[0132] 在一方面,本公开提供了所描述的重组狂犬病病毒或所描述的药物组合物在制造用于治疗受试者的疾病或病症的药物中的用途。

附图说明

[0133] 图1是示出来自在指示天数从各种稳定细胞系收获的等体积的含病毒上清液的293T细胞的相对感染性的图表。

[0134] 图2A是描绘VIR218复制子的示意图。

[0135] 图2B是描绘重组狂犬病病毒颗粒介导的基因递送的产生和感染方案的示意图。

[0136] 图2C是描绘包含编码核碱基编辑器的重组狂犬病病毒基因组的重组狂犬病病毒颗粒可以实现靶序列的基因编辑的图表。

[0137] 图3A是描绘包含gRNA、多核苷酸可编程核苷酸结合结构域和核碱基编辑器的重组狂犬病病毒基因组的组织的示意图。

[0138] 图3B是描绘编码两个tRNA序列之间的gRNA的gRNA-tRNA表达盒的示意图,其中箭头指示RNA的切割位点。

[0139] 图3C是描绘编码gRNA(第一gRNA和第二gRNA)的gRNA-tRNA表达盒的示意图,其中第一gRNA位于第一tRNA与第二tRNA之间,随后是第二gRNA。

[0140] 图3D是描绘编码gRNA(第一gRNA和第二gRNA)的gRNA-tRNA表达盒的示意图,其中第一gRNA位于第一tRNA与第二tRNA之间,并且第二gRNA位于第二tRNA与第三tRNA之间。

[0141] 图3E是描绘用重组狂犬病病毒颗粒转导的HEK细胞中的感染%和A>G碱基编辑%的图表,该重组狂犬病病毒颗粒包含编码核碱基编辑器的重组狂犬病病毒基因组和在多个tRNA之间编码的gRNA。在被Hek2靶向gRNA和IEDG靶向gRNA靶向的Hek2位点和IEDG位点处测量碱基编辑%。

[0142] 图4A是描绘用表达核碱基编辑器的载体和表达介于侧接tRNA之间的gRNA(在代表

tRNA-gRNA-tRNA格式的数据中称为“侧翼(flank)”)或非经侧接的gRNA(即tRNA-gRNA)的载体共转染的293T细胞中的A>G碱基编辑%的图表。碱基编辑%是在被Hek2靶向gRNA靶向的Hek2位点处测量的。

[0143] 图4B是描绘用表达核碱基编辑器的载体和表达与源自多种物种的MALAT1相关小细胞质RNA(mascRNA)连接的gRNA的载体共转染的293T细胞中的A>G碱基编辑%的图表。碱基编辑%是在被Hek2靶向gRNA靶向的Hek2位点处测量的。

[0144] 图4C是描绘用表达核碱基编辑器的载体和表达tRNA-gRNA变体的载体共转染的293T细胞中的A>G碱基编辑%的图表。碱基编辑%是在被Hek2靶向gRNA靶向的Hek2位点处测量的。

[0145] 图4D是描绘用表达核碱基编辑器的载体和表达与gRNA连接的tRNA片段、RnaseZ或RnaseP底物的载体共转染的293T细胞中的A>G碱基编辑%的图表。碱基编辑%是在被Hek2靶向gRNA靶向的Hek2位点处测量的。

[0146] 图5是描绘用表达核碱基编辑器的载体和表达连接到gRNA的来自γ疱疹病毒(GHV68)的病毒tRNA样结构(vtRNA)的载体共转染的293T细胞中的A>G碱基编辑%的图表。碱基编辑%是在被Hek2靶向gRNA靶向的Hek2位点、被SOD1靶向gRNA靶向的SOD1位点以及被ALAS1靶向gRNA靶向的ALAS1位点处测量的。

[0147] 图6A是描绘tRNA-gRNA盒在共表达核碱基编辑器的不同RABV基因组架构内的放置的示意图。

[0148] 图6B是描绘用重组狂犬病病毒颗粒转导的293T细胞中的A>G碱基编辑%的图表,该重组狂犬病病毒颗粒包含编码核碱基编辑器的重组狂犬病病毒基因组和插入在不同RABV基因组架构中的几个位置处的tRNA(Gly)-gRNA盒。碱基编辑%是在ALAS1位点和SOD1位点处测量的。

具体实施方式

[0149] 本文提供了一种重组负链RNA病毒基因组,其包含编码第一引导RNA(gRNA)的核酸,该编码第一引导RNA(gRNA)的核酸包含5’端和3’端;和编码第一转移RNA(tRNA)的核酸,该编码第一转移RNA(tRNA)的核酸位于编码第一gRNA的核酸的3’端或编码第一gRNA的核酸的5’端中的一者或两者处。

[0150] 应当理解,本文描述的方法不限于本文公开的特定方法和实验条件,因为这样的方法和条件可以变化。还应理解,本文所使用的术语仅是用于描述特定实施方案的目的,并且不意图具限制性。本文描述的方法使用普通技术人员的技术范围内的常规分子和细胞生物学和免疫学技术。此类技术对于本领域技术人员来说是众所周知的并且在科学文献中进行了解释。

[0151] A.定义

[0152] 除非另外定义,否则本文所使用的所有技术性和科学性术语具有本发明所属领域中的技术人员通常所理解的含义。以下参考文献向技术人员提供了在本发明中使用的许多术语的一般定义:Singleton等人,Dictionary of Microbiology and Molecular Biology(第2版1994);The Cambridge Dictionary of Science and Technology(Walker编,1988);The Glossary of Genetics,第5版,R.Rieger等人(编),Springer Verlag(1991);

以及Hale和Marham,The Harper Collins Dictionary of Biology(1991)。如本文所用,除非另外指明,否则以下术语具有下面赋予它们的含义。

[0153] “腺苷脱氨酶”是指能够催化腺嘌呤或腺苷水解脱氨的多肽或其片段。在一些实施方案中,脱氨酶或脱氨酶结构域是催化腺苷水解脱氨为肌苷或者催化脱氧腺苷水解脱氨为脱氧肌苷的腺苷脱氨酶。在一些实施方案中,腺苷脱氨酶催化脱氧核糖核酸(DNA)中的腺嘌呤或腺苷的水解脱氨。本文提供的腺苷脱氨酶(例如经工程化的腺苷脱氨酶、进化的腺苷脱氨酶)可以来自任何生物体,诸如细菌。

[0154] “腺苷脱氨酶碱基编辑器8(ABE8)多肽”或“ABE8”意指如本文所定义的碱基编辑器,该碱基编辑器包含腺苷脱氨酶变体,该变体包含位于以下参考序列的氨基酸位置82和/或166处的改变:MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD(SEQ ID NO:8)。

[0155] 在一些实施方案中,ABE8包含如本文所述的相对于参考序列的进一步的改变。

[0156] “腺苷脱氨酶碱基编辑器8(ABE8)多核苷酸”意指编码ABE8的多核苷酸。

[0157] “施用”在本文中是指向患者或受试者提供一种或多种本文所述的组合物。

[0158] “剂”意指任何小分子化学化合物、抗体、核酸分子,或多肽,或其片段。

[0159] “改变”意指分析物、基因或多肽的水平、结构或活性的变化(增加或减少),如通过标准领域(standard art)已知的方法如本文所述的那些方法所检测。如本文中所用,改变包括表达水平的10%变化、25%的变化、40%的变化,以及50%或更高的变化。在一些实施方案中,改变包括核碱基或氨基酸的插入、缺失或取代。

[0160] “改善”意指减少、阻抑、衰减、缩减、迟滞或稳定化疾病的发展或进展。

[0161] “类似物”意指不同一但具有类似功能或结构特征的分子。例如,多肽类似物保留了相应的天然存在的多肽的生物学活性,同时具有相对于天然存在的多肽增强类似物的功能的某些生化修饰。此类生化修饰可以增加类似物的蛋白酶抗性、膜渗透性或半衰期,而不改变例如配体结合。类似物可包括非天然氨基酸。

[0162] “碱基编辑器(BE)”或“核碱基编辑器多肽(NBE)”意指结合多核苷酸并且具有核碱基修饰活性的剂。在各种实施方案中,碱基编辑器包含核碱基修饰多肽(例如,脱氨酶)和与引导多核苷酸(例如,引导RNA(gRNA))联合的多核苷酸可编程核苷酸结合结构域(例如,Cas9或Cpf1)。碱基编辑器的代表性核酸和蛋白质序列在序列表中作为SEQ ID NO:274-283提供。

[0163] “碱基编辑活性”意指化学改变多核苷酸内的碱基的作用。在一个实施方案中,第一碱基被转化为第二碱基。在一个实施方案中,碱基编辑活性是例如将靶C·G转化为T·A的胞苷脱氨酶活性。在另一个实施方案中,碱基编辑活性是例如,将A·T转化为G·C的腺苷或腺嘌呤脱氨酶活性。

[0164] 术语“碱基编辑器系统”是指用于编辑靶核苷酸序列的核碱基的分子间复合物。在各种实施方案中,碱基编辑器(BE)系统包含(1)用于使靶核苷酸序列中的核碱基脱氨的多核苷酸可编程核苷酸结合结构域、脱氨酶结构域(例如,胞苷脱氨酶或腺苷脱氨酶);和(2)与多核苷酸可编程核苷酸结合结构域联合的一种或多种引导多核苷酸(例如,引导RNA)。在各种实施方案中,碱基编辑器(BE)系统包含选自腺苷脱氨酶或胞苷脱氨酶的核碱基编辑器

结构域,以及具有核酸序列特异性结合活性的结构域。在一些实施方案中,碱基编辑器系统包含(1)碱基编辑器(BE),该碱基编辑器包含用于将靶核苷酸序列中的一个或多个核碱基脱氨的多核苷酸可编程DNA结合结构域和脱氨酶结构域;以及(2)与多核苷酸可编程DNA结合结构域联合的一种或多种引导RNA。在一些实施方案中,多核苷酸可编程核苷酸结合结构域是多核苷酸可编程DNA结合结构域。在一些实施方案中,碱基编辑器是胞苷碱基编辑器(CBE)。在一些实施方案中,该碱基编辑器是腺嘌呤或腺苷碱基编辑器(ABE)。在一些实施方案中,碱基编辑器是腺嘌呤或腺苷碱基编辑器(ABE)或胞苷碱基编辑器(CBE)。

[0165] “碱基编辑活性”意指化学改变多核苷酸内的碱基的作用。在一个实施方案中,第一碱基被转化为第二碱基。在另一个实施方案中,碱基编辑活性是例如,将靶C·G转化为T·A的胞苷脱氨酶活性。在另一个实施方案中,碱基编辑活性是例如,将A·T转化为G·C的腺苷脱氨酶活性。

[0166] 术语“Cas9”或“Cas9结构域”是指RNA引导的核酸酶,该核酸酶包含Cas9蛋白或其片段(例如,包含Cas9的有活性、无活性或有部分活性的DNA切割结构域的蛋白,和/或Cas9的gRMA结合结构域)。Cas9核酸酶有时也称为casn1核酸酶或CRISPR(成簇规则间隔短回文重复序列)相关核酸酶。

[0167] 术语“保守氨基酸取代”或“保守突变”是指一种氨基酸被另一种具有共同特性的氨基酸替换。一种定义单个氨基酸之间的共同特性的功能性途径是分析同源生物体的相应蛋白质之间的氨基酸变化的归一化频率(Schulz,G.E.和Schirmer,R.H.,Principles of Protein Structure,Springer-Verlag,New York(1979))。根据此类分析,可定义多组氨基酸,其中一个组内的氨基酸优先彼此交换,并因此在它们对总体蛋白质结构的影响方面彼此最相似(Schulz,G.E.和Schirmer,R.H.,同上)。保守突变的非限制性示例包括以下氨基酸的氨基酸取代:例如,赖氨酸取代精氨酸且反之亦然,使得正电荷可得以保持;谷氨酸取代天冬氨酸且反之亦然,使得负电荷可得以保持;丝胺酸取代苏氨酸,使得游离-OH可得以保持;以及谷氨酰胺取代天冬酰胺,使得游离-NH_2可得以保持。

[0168] 如本文中可互换使用的术语“编码序列”或“蛋白质编码序列”是指编码蛋白质的多核苷酸节段。编码序列也可称为开放阅读框。该区域或序列在更靠近5’端的地方结合有起始密码子,并且在更靠近3’端的地方结合有终止密码子。对本文所述的碱基编辑器有用的终止密码子包括以下:

[0169]

谷氨酰胺	**CAG → TAG** 终止密码子
	CAA → TAA
精氨酸	**CGA → TGA**
色氨酸	**TGG → TGA**
	TGG → TAG
	TGG → TAA

[0170] “胞苷脱氨酶”是指能够催化将氨基转化为羰基的脱氨反应的多肽或其片段。在一个实施方案中,胞苷脱氨酶将胞嘧啶转化为尿嘧啶或者将5-甲基胞嘧啶转化为胸腺嘧啶。源自自海七鳃鳗(海七鳃鳗胞嘧啶脱氨酶1,“PmCDA1”)的PmCDA1(SEQ ID NO:41-42),源自自哺乳动物(例如,人、猪、牛、马、猴等)的AID(激活诱导的胞苷脱氨酶;AICDA)(示例性AID

多肽序列在序列表中作为SEQ ID NO:43-44、1372和1374-1377提供),和APOBEC是示例性胞苷脱氨酶(示例性APOBEC多肽序列在序列表中作为SEQ ID NO:1378-1416、1421和1422提供)。进一步的示例性胞苷脱氨酶(CDA)序列在序列表中作为SEQ ID NO:1373、1417-1420提供。包括APOBEC多肽序列在内的另外的示例性胞苷脱氨酶序列在序列表中作为SEQ ID NO:1378-1422提供。

[0171] 如本文所用的术语"脱氨酶"或"脱氨酶结构域"是指催化脱氨反应的蛋白或酶。

[0172] "检测"是指鉴定待检测分析物的存在、不存在或量。在一个实施方案中,检测多核苷酸或多肽中的序列改变。在另一个实施方案中,检测插入/缺失的存在。

[0173] "可检测标记"意指这样的组合物,该组合物在与关注的分子连接时使得该关注的分子能够经由光谱、光化学、生物化学、免疫化学或化学手段检测。例如,有用的标记包括放射性同位素、磁珠、金属珠、胶体颗粒、荧光染料、电子致密试剂、酶(例如,如通常用于酶联免疫吸附测定(ELISA))、生物素、地高辛或半抗原。

[0174] "疾病"意指损害或干扰细胞、组织或器官的正常功能的任何疾患或病症。示例性疾病包括神经疾病和眼科疾病。

[0175] "有效量"意指相对于未经治疗的患者或没有疾病的个体(即,健康个体)改善疾病症状所需的剂或活性化合物(例如本文所述的碱基编辑器)的量,或者是足以引发所需生物反应的剂或活性化合物的量。用于实践本发明以治疗性地治疗疾病的一种或多种活性化合物的有效量根据施用方式,受试者的年龄、体重和总体健康状况而变化。最终,主治医师或兽医将决定适当的量和剂量方案。此种量被称为"有效"量。在一个实施方案中,有效量是足以在细胞(例如,体外或体内细胞)中的关注的基因中引入改变的本发明的碱基编辑器的量。在一个实施方案中,有效量是达到治疗效果所需的碱基编辑器的量。此种治疗效果不需要足以改变受试者、组织或器官的所有细胞中的致病基因,而仅需改变存在于受试者、组织或器官中约1%、5%、10%、25%、50%、75%或更多的细胞中的致病基因。在一个实施方案中,有效量足以改善疾病的一种或多种症状。

[0176] 术语"核酸外切酶"是指能够从游离末端消化核酸(例如,RNA或DNA)的蛋白质或多肽。

[0177] 术语"核酸内切酶"是指能够催化(例如,切割)核酸(例如,DNA或RNA)中的内部区域的蛋白或多肽。

[0178] "片段"意指多肽或核酸分子的一部分。此部分含有参考核酸分子或多肽的整个长度的至少约10%、20%、30%、40%、50%、60%、70%、80%或90%。片段可含有10、20、30、40、50、60、70、80、90或100、200、300、400、500、600、700、800、900或1000个核苷酸或氨基酸。

[0179] "引导RNA"或"gRNA"意指对靶序列具有特异性并且可以与多核苷酸可编程核苷酸结合结构域蛋白(例如,Cas9或Cpf1)形成复合物的多核苷酸或多核苷酸复合物。在一个实施方案中,引导多核苷酸是引导RNA(gRNA)。gRNA可以作为两种或更多种RNA的复合物存在,或者作为单个RNA分子存在。

[0180] "tRNA"或"转移RNA"意指由RNA核苷酸组成的多核苷酸,该多核苷酸充当衔接子从而充当mRNA与由所述mRNA编码的蛋白质的氨基酸序列之间的物理链路。"tRNA"或"转移RNA"还指包含二级结构的RNA分子,该二级结构可以充当参与tRNA成熟的细胞RNase(诸如RNase P或RNase Z)的底物。tRNA通常包含三叶草结构,该三叶草结构可以包括受体茎区,

和包括TψC环、可变环、反密码子环和D环在内的几个环中的至少一个环。术语“tRNA样结构”也被术语tRNA涵盖并且包括tRNA变体、tRNA片段、病毒tRNA和mascRNA。tRNA成熟过程包括tRNA结构的识别、和切割。切割可以例如通过RNase如RNase P或RNase Z发生。因此,位于gRNA的5’端或gRNA的3’端中的一者或两者处的tRNA或tRNA样结构将在所述tRNA的切割后释放所述gRNA。在负链基因组的背景下,tRNA或tRNA样结构位于gRNA的3’端或gRNA的5’端中的一者或两者处。

[0181] “杂交”意指互补核碱基之间的氢键合,该氢键合可以是Watson-Crick、Hoogsteen或反向Hoogsteen氢键合。例如,腺嘌呤和胸腺嘧啶是通过形成氢键配对的互补核碱基。

[0182] “增加”意指至少10%、25%、50%、75%或100%的正向改变。

[0183] 术语“碱基修复的抑制剂(inhibitor of base repair)”、“碱基修复抑制剂(base repair inhibitor)”“IBR”或它们的语法等同物是指能够抑制核酸修复酶(例如碱基切除修复酶)的活性的蛋白质。

[0184] 术语“经分离的(isolated)”、“经纯化的”或“生物学纯的”是指这样的材料,该材料在不同程度上不含如在它的原生(native)状态下发现的通常伴随它的组分。“分离”表示与原始来源或周围环境的分离程度。“纯化”表示高于分离的分离程度。“经纯化的”或“生物学纯的”蛋白质充分不含其他材料,使得任何杂质不会实质性地影响蛋白质的生物学特性或引起其他不利后果。即,如果本发明的核酸或肽在通过重组DNA技术产生时基本上不含细胞材料、病毒材料或培养基,或者在化学合成时基本上不含化学前体或其他化学品,则该核酸或肽是经纯化的。纯度和均匀性通常使用分析化学技术,例如聚丙烯酰胺凝胶电泳或高效液相色谱确定。术语“经纯化的”可以表示核酸或蛋白质在电泳凝胶中产生基本上一个条带。对于可以经受修饰(例如磷酸化或糖基化)的蛋白,不同的修饰可能会产生不同的经分离的蛋白,这些蛋白可以被单独地纯化。

[0185] “经分离的多核苷酸”意指这样的核酸(例如,DNA),该核酸不含在本发明的核酸分子所来源的生物体的天然存在的基因组中侧接该基因的基因。因此,该术语包括,例如,掺入到载体中的重组DNA;掺入到自主复制的质粒或病毒中的重组DNA;或掺入到原核生物或真核生物的基因组DNA中的重组DNA;或作为独立于其他序列的单独分子(例如,通过PCR或限制性核酸内切酶消化产生的cDNA或基因组或cDNA片段)存在的重组DNA。此外,该术语包括从DNA分子转录的RNA分子,以及作为编码另外的多肽序列的杂合基因的一部分的重组DNA。

[0186] “经分离的多肽”意指这样的本发明的多肽,该多肽已与天然伴随它的组分分离。通常,当多肽的至少60重量%不含与其天然关联的蛋白质和天然存在的有机分子时,则该多肽是经分离的。优选地,制备物(preparation)是至少75重量%、更优选至少90重量%,且最优选至少99重量%的本发明的多肽。本发明的经分离的多肽可以,例如,通过从天然来源中提取,通过使编码此种多肽的重组核酸表达;或通过化学合成蛋白来获得。纯度可以通过任何合适的方法,例如柱色谱、聚丙烯酰胺凝胶电泳,或通过HPLC分析测量。

[0187] 如本文所用的术语“突变”是指序列(例如核酸或氨基酸序列)内的残基被另一残基取代,或序列内一个或多个残基的缺失或插入。突变在本文中通常通过鉴定原始残基以及随后鉴定该残基在序列内的位置以及通过新取代的残基的身份来描述。用于制作本文提供的氨基酸取代(突变)的各种方法是本领域众所周知的,并且由例如Green和Sambrook,

Molecular Cloning:A Laboratory Manual(第4版,Cold Spring Harbor Laboratory Press,Cold Spring Harbor,N.Y.(2012))提供。

[0188] 如本文所用的术语“核酸”和“核酸分子”是指包含核碱基和酸性部分的化合物,例如核苷、核苷酸或核苷酸的聚合物。通常,聚合性核酸,例如,包含三个或更多个核苷酸的核酸分子为线性分子,其中相邻核苷酸经由磷酸二酯连键彼此连接。在一些实施方案中,“核酸”是指单个核酸残基(例如核苷酸和/或核苷)。在一些实施方案中,“核酸”是指包含三个或更多个单个核苷酸残基的寡核苷酸链。如本文所用,术语“寡核苷酸”和“多核苷酸”可互换地用来指核苷酸的聚合物(例如,至少三个核苷酸的串)。在一些实施方案中,“核酸”涵盖RNA以及单链和/或双链DNA。核酸可以是天然存在的,例如在基因组、转录物、mRNA、tRNA、rRNA、siRNA、snRNA、质粒、粘粒、染色体、染色单体或其他天然存在的核酸分子的背景下。在另一方面,核酸分子可以是非天然存在的分子,例如重组DNA或RNA、人工染色体、经工程化的基因组,或其片段,或者合成的DNA、RNA、DNA/RNA杂合体,或者包括非天然存在的核苷酸或核苷。此外,术语“核酸”、“DNA”、“RNA”和/或类似术语包括核酸类似物,例如具有除磷酸二酯骨架以外的骨架的类似物。核酸可由天然来源纯化、使用重组表达系统产生并任选地被纯化、化学合成等。在适当的情况下,例如,在化学合成的分子的情况下,核酸可包含核苷类似物如具有经化学修饰的碱基或糖,以及骨架修饰的类似物。除非另有说明,否则核酸序列以5’到3’方向呈现。在一些实施方案中,核酸是或包含天然核苷(例如腺苷、胸苷、鸟苷、胞苷、尿苷、脱氧腺苷、脱氧胸苷、脱氧鸟苷和脱氧胞苷);核苷类似物(例如,2-氨基腺苷、2-硫胸苷、肌苷、吡咯并嘧啶、3-甲基腺苷、5-甲基胞苷、2-氨基腺苷、C5-溴尿苷、C5-氟尿苷、C5-碘尿苷、C5-丙炔基-尿苷、C5-丙炔基-胞苷、C5-甲基胞苷、2-氨基腺苷、7-脱氮腺苷、7-脱氮鸟苷、8-氧代腺苷、8-氧代鸟苷、O(6)-甲基鸟苷和2-硫代胞苷);经化学修饰的碱基;经生物修饰的碱基(例如甲基化碱基);经插入的碱基;经修饰的糖(2’-例如氟核糖、核糖、2’-脱氧核糖、阿拉伯糖和己糖);和/或经修饰的磷酸酯基团(例如,硫代磷酸酯和5’-N-亚磷酰胺连键)。

[0189] 术语“核定位序列”、“核定位信号”或“NLS”是指促进蛋白质输入到细胞核中的氨基酸序列。核定位序列是本领域已知的,并且在例如Plank等人在2000年11月23日提交的国际PCT申请PCT/EP2000/011690(其在2001年5月31日以WO/2001/038547公布)中有所描述,该专利申请的内容中的公开示例性核定位序列的内容以引用的方式并入本文。在其他实施方案中,NLS是例如由Koblan等人,Nature Biotech.2018doi:10.1038/nbt.4172所描述的经优化的NLS。在一些实施方案中,NLS包含氨基酸序列KRTADGSEFESPKKKRKV(SEQ ID NO:84)、KRPAATKKAGQAKKKK(SEQ ID NO:85)、KKTEL QTTNAENKTKKL(SEQ ID NO:86)、KRGINDRNFWRGENGRKT R(SEQ ID NO:87)、RKSGKIAAIVVKRPRK(SEQ ID NO:88)、PKKKRKV(SEQ ID NO:89),或MDSLLMNRRKFLYQFKNVRWA KGRRETYLC(SEQ ID NO:90)。

[0190] 术语“核碱基”、“含氮碱基”或“碱基”在本文中可互换使用,是指形成核苷的含氮生物化合物,核苷又是核苷酸的组分。核碱基形成碱基对并相互堆叠的能力直接导致长链螺旋结构,诸如核糖核酸(RNA)和脱氧核糖核酸(DNA)。五种核碱基——腺嘌呤(A)、胞嘧啶(C)、鸟嘌呤(G)、胸腺嘧啶(T)和尿嘧啶(U)——被称为初级碱基或典型碱基。腺嘌呤和鸟嘌呤衍生自嘌呤,并且胞嘧啶、尿嘧啶和胸腺嘧啶衍生自嘧啶。DNA和RNA也可以含有被修饰的其他(非初级)碱基。非限制性示例性的经修饰的核碱基可包括次黄嘌呤、黄嘌呤、7-甲基鸟

嘌呤、5,6-二氢尿嘧啶、5-甲基胞嘧啶(m5C)和5-氢甲基胞嘧啶。次黄嘌呤和黄嘌呤可以通过诱变剂的存在产生,二者都是通过脱氨(用羰基替换胺基团)产生的。次黄嘌呤可以从腺嘌呤修饰而成。黄嘌呤可以从鸟嘌呤修饰而成。尿嘧啶可以从胞嘧啶的脱氨产生。“核苷”由核碱基和五碳糖(核糖或脱氧核糖)组成。核苷的实例包括腺苷、鸟苷、尿苷、胞苷、5-甲基尿苷(m5U)、脱氧腺苷、脱氧鸟苷、胸苷、脱氧尿苷和脱氧胞苷。具有经修饰的核碱基的核苷的实例包括肌苷(I)、黄苷(X)、7-甲基鸟苷(m7G)、二氢尿苷(D)、5-甲基胞苷(m5C)和假尿苷(Ψ)。“核苷酸”由核碱基、五碳糖(核糖或脱氧核糖)和至少一个磷酸酯基团组成。

[0191] 如本文所用的术语“核酸”和“核酸分子”是指包含核碱基和酸性部分的化合物,例如核苷、核苷酸或核苷酸的聚合物。

[0192] 如本文所用,术语“寡核苷酸”和“多核苷酸”可互换地用来指核苷酸的聚合物。

[0193] 术语“核酸可编程DNA结合蛋白”或“napDNAbp”可以与“多核苷酸可编程核苷酸结合结构域”互换地用来指与核酸(例如,DNA或RNA)缔合的蛋白,诸如将napDNAbp引导至特定核酸序列的引导核酸或引导多核苷酸(例如,gRNA)。在一些实施方案中,多核苷酸可编程核苷酸结合结构域是多核苷酸可编程DNA结合结构域。在一些实施方案中,多核苷酸可编程核苷酸结合结构域是多核苷酸可编程RNA结合结构域。在一些实施方案中,多核苷酸可编程核苷酸结合结构域是Cas9蛋白。Cas9蛋白可以与引导RNA缔合,该引导RNA将Cas9蛋白引导至与该引导RNA互补的特定DNA序列。在一些实施方案中,napDNAbp是Cas9结构域,例如核酸酶有活性的Cas9、Cas9切口酶(nCas9),或核酸酶无活性的Cas9(dCas9)。核酸可编程DNA结合蛋白的非限制性实例包括Cas9(例如,dCas9和nCas9)、Cas12a/Cpf1、Cas12b/C2c1、Cas12c/C2c3、Cas12d/CasY、Cas12e/CasX、Cas12g,Cas12h、Cas12i,和Cas12j/CasΦ(Cas12j/Casphi)。Cas酶的非限制性实例包括Cas1、Cas1B、Cas2、Cas3、Cas4、Cas5、Cas5d、Cas5t、Cas5h、Cas5a、Cas6、Cas7、Cas8、Cas8a、Cas8b、Cas8c、Cas9(也称为Csn1或Csx12)、Cas10、Cas10d、Cas12a/Cpf1、Cas12b/C2c1、Cas12c/C2c3、Cas12d/CasY、Cas12e/CasX、Cas12g、Cas12h、Cas12i、Cas12j/CasΦ、Cpf1、Csy1、Csy2、Csy3、Csy4、Cse1、Cse2、Cse3、Cse4、Cse5e、Csc1、Csc2、Csa5、Csn1、Csn2、Csm1、Csm2、Csm3、Csm4、Csm5、Csm6、Cmr1、Cmr3、Cmr4、Cmr5、Cmr6、Csb1、Csb2、Csb3、Csx17、Csx14、Csx10、Csx16、CsaX、Csx3、Csx1、Csx1S、Csx11、Csf1、Csf2、Cs0、Csf4、Csd1、Csd2、Cst1、Cst2、Csh1、Csh2、Csa1、Csa2、Csa3、Csa4、Csa5、II型Cas效应蛋白、V型Cas效应蛋白、VI型Cas效应蛋白、CARF、DinG、它们的同源物或它们的经修饰的或经工程化的型式。其他核酸可编程DNA结合蛋白也在本公开的范围内,尽管它们可能未在本公开中具体列出。参见例如Makarova等人“Classification and Nomenclature of CRISPR-Cas Systems:Where from Here?”CRISPR J.2018年10月;1:325-336.doi:10.1089/crispr.2018.0033;Yan等人,“Functionally diverse type V CRISPR-Cas systems”Science.2019年1月4日;363(6422):88-91.doi:10.1126/science.aav7271,其全部内容以引用方式并入本文。示例性核酸可编程DNA结合蛋白和编码核酸可编程DNA结合蛋白的核酸序列在序列表中作为SEQ ID NO:223、230-232、235-242、246-256和285-294提供。

[0194] 如本文中所用的术语“核碱基编辑结构域”或“核碱基编辑蛋白”是指这样的蛋白或酶,该蛋白或酶可以催化RNA或DNA中的核碱基修饰,诸如胞嘧啶(或胞苷)到尿嘧啶(或尿苷)或胸腺嘧啶(或胸苷)以及腺嘌呤(或腺苷)到次黄嘌呤(或肌苷)脱氨,以及非模板化的核苷酸加成和插入。在一些实施方案中,核碱基编辑结构域是脱氨酶结构域(例如腺嘌呤脱

氨酶或腺苷脱氨酶;或胞苷脱氨酶或胞嘧啶脱氨酶)。

[0195] 如本文中所用,如“获得剂”中的“获得”包括合成、购买或以其它方法取得该剂。

[0196] 如本文中所用的“患者”或“受试者”是指被诊断出患有疾病或病症、处于患有或患上疾病或病症的风险中,或者被怀疑患有或患上疾病或病症的哺乳动物受试者或个体。在一些实施方案中,术语“患者”是指患上疾病或病症的可能性高于平均可能性的哺乳动物受试者。示例性患者可以是可受益于本文所公开的疗法的人、非人灵长动物、猫、狗、猪、牛、猫、马、骆驼、骆马、山羊、绵羊、啮齿动物(例如,小鼠、兔、大鼠或豚鼠)和其他哺乳动物。示例性人类患者可以是男性和/或女性。

[0197] “有需要的患者”或“有需要的受试者”在本文中是指被诊断患有疾病或病症、处于患有疾病或病症的风险中、预先确定患有疾病或病症,或者被怀疑患有疾病或病症的患者。

[0198] 术语“致病突变”、“致病变体”、“造成疾病的突变”、“造成疾病的变体”、“有害突变”或“诱发性突变(predisposing mutation)”是指增加个体对某种疾病或病症的易感性(susceptibility)或倾向性(predisposition)的基因改变或突变。在一些实施方案中,致病突变包含在由基因编码的蛋白中至少一个野生型氨基酸被至少一个致病氨基酸取代。

[0199] 术语“蛋白”、“肽”、“多肽”及其语法等效物在本文中可互换地使用,并且是指通过肽(酰胺)键连接在一起的氨基酸残基的聚合物。蛋白、肽或多肽可以是天然存在的、重组的或合成的,或它们的任何组合。

[0200] 如本文所用的术语“融合蛋白”是指一种杂合多肽,其包含来自至少两种不同蛋白的蛋白结构域。

[0201] 如本文在蛋白或核酸的上下文中使用的术语“重组”是指不存在于自然界中而是人类工程化产物的蛋白或核酸。例如,在一些实施方案中,重组蛋白或核酸分子包含含有至少一个、至少两个、至少三个、至少四个、至少五个、至少六个或至少七个与任何天然存在的序列相比的突变的氨基酸或核苷酸序列。

[0202] “降低”意指至少10%、25%、50%、75%或100%的负向改变。

[0203] “参考序列”是定义的用作序列比较的基础的序列。参考序列可以是指定序列的子集或整体;例如,全长cDNA或基因序列的节段,或完整的cDNA或基因序列。对于多肽,参考多肽序列的长度将一般为至少约16个氨基酸、至少约20个氨基酸、至少约25个氨基酸、约35个氨基酸、约50个氨基酸或约100个氨基酸。对于核酸,参考核酸序列的长度将一般为至少约50个核苷酸、至少约60个核苷酸、至少约75个核苷酸、约100个核苷酸或约300个核苷酸或它们附近或它们之间的任何整数。在一些实施方案中,参考序列是关注的蛋白的野生型序列。在其他实施方案中,参考序列是编码野生型蛋白的多核苷酸序列。

[0204] 术语“RNA可编程核酸酶”和“RNA引导的核酸酶”与一种或多种不是切割靶标的RNA一起使用(例如,结合或缔合)。在一些实施方案中,当与RNA形成复合物时,RNA可编程核酸酶可以被称为核酸酶:RNA复合物。通常,一种或多种被结合的RNA被称为引导RNA(gRNA)。在一些实施方案中,RNA可编程核酸酶是(CRISPR相关系统)Cas9核酸内切酶,例如,来自酿脓链球菌的Cas9(Casn1)。

[0205] 术语“单核苷酸多态性(SNP)”是发生在基因组中的具体位置处的单个核苷酸的变异,其中每种变异在群体内以某种可感知的程度(例如,>1%)存在。

[0206] “特异性地结合”意指识别并且结合本发明的多肽和/或核酸分子,但基本上不识

别并且结合样品(例如生物样品)中的其他分子的核酸分子、多肽、多肽/多核苷酸复合物、化合物或分子。

[0207] “实质上同一的”是指与参考氨基酸序列表现出至少50％同一性的多肽或核酸分子。在一个实施方案中,参考序列是野生型氨基酸或核酸序列。在另一个实施方案中,参考序列是本文描述的氨基酸或核酸序列中的任一种。在一个实施方案中,此序列在氨基酸水平或核酸水平上与用于比较的序列是至少60％,更优选80％或85％,且更优选90％、95％或甚至99％同一的。

[0208] 序列同一性通常使用序列分析软件(例如,威斯康星大学生物技术中心的遗传学计算机组(Genetics Computer Group,University of Wisconsin Biotechnology Center,1710University Avenue,Madison,Wis.53705)的序列分析软件包、BLAST、BESTFIT、GAP,或PILEUP/PRETTYBOX程序)测量。此种软件通过为各种取代、缺失和/或其他修饰分配同源性程度来匹配同一或相似的序列。保守性取代通常包括以下组内的取代:甘氨酸、丙氨酸;缬氨酸、异亮氨酸、亮氨酸;天冬氨酸、谷氨酸、天冬酰胺、谷氨酰胺;丝氨酸、苏氨酸;赖氨酸、精氨酸;和苯丙氨酸、酪氨酸。在用于确定同一性程度的示例性方法中,可以使用BLAST程序,其中e^{-3}和e^{-100}之间的概率得分指示密切相关的序列。

[0209] 例如,COBALT与以下参数一起使用:

[0210] a)比对参数:空位罚分为-11、-1,并且末端空位罚分为-5、-1,

[0211] b)CDD参数:使用RPS BLAST开启(on);Blast E-值0.003;寻找保守的列并重新计算(Find Conserved columns and Recompute)开启,和

[0212] c)查询聚簇分析参数:使用查询聚簇开启;字长4;最大聚簇间距离0.8;常规字符(Alphabet Regular)。

[0213] 例如,EMBOSS Needle与以下参数一起使用:

[0214] a)矩阵:BLOSUM62;

[0215] b)空位开放(GAP OPEN):10;

[0216] c)空位扩展(GAP EXTEND):0.5;

[0217] d)输出格式:配对(pair);

[0218] e)末端空位罚分(END GAP PENALTY):假(false);

[0219] f)末端空位开放(END GAP OPEN):10;以及

[0220] g)末端空位扩展(END GAP EXTEND):0.5。

[0221] 可用于本发明方法的核酸分子包括编码本发明多肽或其片段的任何核酸分子。此类核酸分子不需要与内源核酸序列100％同一,但通常将表现出实质的同一性。与内源序列具有“实质的同一性”的多核苷酸通常能够与双链核酸分子的至少一条链杂交。可用于本发明方法的核酸分子包括编码本发明多肽或其片段的任何核酸分子。此类核酸分子不需要与内源核酸序列100％同一,但通常将表现出实质的同一性。与内源序列具有“实质的同一性”的多核苷酸通常能够与双链核酸分子的至少一条链杂交。“杂交”意指在多种严格性条件下配对以在互补多核苷酸序列(例如,本文所述的基因)或其部分之间形成双链分子。(参见例如,Wahl,G.M.和S.L.Berger(1987)Methods Enzymol.152:399;Kimmel,A.R.(1987)Methods Enzymol.152:507)。

[0222] 例如,严格盐浓度将通常低于约750mM NaCl和75mM柠檬酸三钠,优选低于约500mM

NaCl和50mM柠檬酸三钠,并且更优选低于约250mM NaCl和25mM柠檬酸三钠。低严格性杂交可以在不存在有机溶剂例如甲酰胺的情况下获得,而高严格性杂交可以在存在至少约35%甲酰胺,且更优选至少约50%甲酰胺的情况下获得。严格的温度条件将通常包括至少约30℃,更优选至少约37℃,且最优选至少约42℃的温度。可变的另外的因素,诸如杂交时间、洗涤剂(如十二烷基硫酸钠(SDS))的浓度,以及载剂DNA的包含或排除是本领域技术人员众所周知的。通过按需要组合这些各种条件来实施各种水平的严格性。在一个优选的实施方案中,杂交将在30℃下在750mM NaCl、75mM柠檬酸三钠和1%SDS中发生。在一个更优选的实施方案中,杂交将在37℃下在500mM NaCl、50mM柠檬酸三钠、1%SDS、35%甲酰胺和100μg/ml变性鲑鱼精子DNA(ssDNA)中发生。在一个优选的实施方案中,杂交将在42℃下在250mM NaCl、25mM柠檬酸三钠、1%SDS、50%甲酰胺和200μg/ml ssDNA中发生。这些条件的有用变化对于本领域技术人员来说是十分显而易见的。

[0223] 对于大多数应用,杂交后的洗涤步骤在严格性方面也会有所不同。洗涤严格性条件可以通过盐浓度和通过温度来限定。如上文,可以通过降低盐浓度或提高温度来增加洗涤严格性。例如,洗涤步骤的严格盐浓度将优选小于约30mM NaCl和3mM柠檬酸三钠,并且最优选小于约15mM NaCl和1.5mM柠檬酸三钠。洗涤步骤的严格温度条件将通常包括至少约25℃,更优选至少约42℃,并且甚至更优选至少约68℃的温度。在一个实施方案中,洗涤步骤将在25℃下在30mM NaCl、3mM柠檬酸三钠和0.1%SDS中发生。在另一个实施方案中,洗涤步骤将在42℃下在15mM NaCl、1.5mM柠檬酸三钠和0.1%SDS中发生。在更优选的实施方案中,洗涤步骤将在68℃下在15mM NaCl、1.5mM柠檬酸三钠和0.1%SDS中发生。这些条件的另外的变化对本领域技术人员来说是十分显而易见的。杂交技术是本领域技术人员众所周知的并且描述于例如Benton和Davis(Science196:180,1977);Grunstein和Hogness(Proc.Natl.Acad.Sci.,USA 72:3961,1975);Ausubel等人(Current Protocols in Molecular Biology,Wiley Interscience,New York,2001);Berger和Kimmel(Guide to Molecular Cloning Techniques,1987,Academic Press,New York);以及Sambrook等人,Molecular Cloning:A Laboratory Manual,Cold Spring Harbor Laboratory Press,New York。

[0224] “分裂”是指分成两个或更多个片段。

[0225] “分裂Cas9蛋白”或“分裂Cas9”是指作为由两个单独核苷酸序列编码的N端片段和C端片段提供的Cas9蛋白。对应于Cas9蛋白的N末端部分和C末端部分的多肽可以被剪接从而形成“经重构的”Cas9蛋白。

[0226] 术语“靶位点”是指核酸分子内被脱氨酶(例如胞苷或腺嘌呤脱氨酶)或包含脱氨酶的融合蛋白(例如本文公开的dCas9-腺苷脱氨酶融合蛋白或碱基编辑器)脱氨的序列。

[0227] 如本文所用,术语“治疗(treat)”、“治疗(treating)”、“治疗(treatment)”等是指减少或改善病症和/或与其相关的症状或者获得期望的药理学和/或生理学效果。应当理解,尽管不排除,治疗病症或疾患并不要求完全消除病症、疾患或与其相关的症状。在一些实施方案中,作用是治疗性的,即但不限于,该作用部分或完全减少、减弱、消除、减缓、缓和、降低疾病和/或可归因于该疾病的不利症状的强度或者治愈该疾病和/或不利症状。在一些实施方案中,作用是预防性的,即该作用保护或防止疾病或疾患的发生或复发。为此,目前公开的方法包括施用治疗有效量的如本文所述的组合物。

[0228] “尿嘧啶糖苷酶抑制剂”或“UGI”意指抑制尿嘧啶切除修复系统的剂。包含胞苷脱氨酶的碱基编辑器将胞嘧啶转化为尿嘧啶,尿嘧啶然后通过DNA复制或修复被转化为胸腺嘧啶。在碱基编辑器中纳入尿嘧啶DNA糖苷酶(UGI)抑制剂可防止使U变回C的碱基切除修复。

[0229] 示例性UGI包含如下氨基酸序列:

[0230] >splP14739IUNGI_BPPB2尿嘧啶-DNA糖苷酶抑制剂

[0231] MTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVH TAYDESTDENVMLLTSD APEYKPWALVIQDSNGENKIKML(SEQ ID NO:106)。

[0232] 本文提供的范围应理解为该范围内所有值的简写。例如,1至50的范围应理解为包括来自由1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49,或50组成的组的任何数字、数字组合,或亚范围。

[0233] 对本文中变量的任何定义中的化学基团列表的列举包括对此种作为任何单个基团或所列基团的组合的变量的定义。对本文中变量或方面的实施方案的列举包括此种作为任何单一实施方案或与任何其它实施方案或其部分组合的实施方案。

[0234] 所有术语都旨在按照它们被本领域技术人员所理解的那样来理解。除非另有定义,本文使用的所有技术和科学术语均具有与本公开所属领域的普通技术人员通常理解的相同的含义。

[0235] 在本申请中,除非另外确切说明,否则单数的使用包括复数。必须指出的是,除非上下文另有明确规定,否则如说明书中所用,单数形式“一种/个(a)”、“一种/个(an)”和“该/所述(the)”包括复数指示物。在本申请中,除非另有说明,否则″或″的使用是指″和/或″。此外,术语“包括(including)”以及其它形式,诸如“包括(includes)”和“包括(included)”的使用不受限制。

[0236] 如本文在说明书和权利要求书中使用的词“包含(comprising)”(以及包含的任何形式,诸如“包含(comprise)”和“包含(comprises)”)、“具有(having)”(以及具有的任何形式,诸如“具有(have)”和“具有(has)”)、“包括(including)”(以及包括的任何形式,诸如“包括(includes)”和“包括(include)”),或“含有(containing)”(以及含有的任何形式,诸如“包含(contains)”和“含有(contain)”)为包含性的或开放式的并且不排除附加的、未提及的要素或方法步骤。可以设想本说明书中讨论的任何实施方案可以以本公开的任何方法或组合物实现,且反之亦然。此外,本公开的组合物可用来实现本公开的方法。

[0237] 术语“约”或“大约”意指处于如由本领域技术人员所确定的特定数值的可接受的误差范围内,该误差范围将部分地取决于该数值是如何测量或确定的,即,测量系统的限制。例如,根据本领域的实践,“约”可以意指处于1个或超过1个标准偏差内。可替代地,“约”可以意指给定值的至多20%、至多10%、至多5%或至多1%的范围。可替代地,特别是关于生物系统或过程,该术语可以意指处于数值的数量级内,例如处于数值的5倍内、处于数值的2倍内。如果在申请书和权利要求书中描述特定数值,除非另外指定,否则应假设术语“约”意指处于该特定数值的可接受的误差范围内。

[0238] 说明书中对“某些实施方案”、“一些实施方案”、“实施方案”、“一个实施方案”或“其他实施方案”的引用意味着,结合该实施方案描述的特定特征、结构或特点包括在本公

开的至少一些实施方案中,但不必包括在本公开的全部实施方案中。

[0239] B.重组负链RNA病毒

[0240] 本文提供了可用于转导靶细胞和递送引导RNA(gRNA)的重组负链RNA病毒(例如狂犬病病毒)。在一方面,本公开的重组负链RNA病毒包含负链RNA病毒糖蛋白和重组负链RNA病毒基因组。在某些实施方案中,重组负链RNA病毒基因组包含编码gRNA(即,第一gRNA)的核酸,该编码gRNA(即,第一gRNA)的核酸包含5'端和3'端。在某些实施方案中,重组负链RNA病毒基因组包含编码tRNA的核酸,该编码tRNA的核酸位于编码gRNA的核酸的3'端和编码gRNA的核酸的5'端中的一者或两者处。

[0241] 在某些实施方案中,重组负链RNA病毒基因组进一步包含编码治疗性转基因的核酸。因此,本公开的重组负链RNA病毒可被用在用于转导靶细胞的方法中,其中该重组负链RNA病毒包含负链RNA病毒糖蛋白和重组负链RNA病毒基因组,该重组负链RNA病毒基因组包含编码gRNA的核酸和任选的转基因(例如,治疗性转基因,如核碱基编辑器)。在转导靶细胞后,重组负链RNA病毒基因组内包含的gRNA被表达并产生gRNA。

[0242] 如本文所用,术语"负链RNA病毒"或"负义单链RNA病毒"是指负RNA病毒门。负链RNA病毒包含充当由病毒酶RNA依赖性RNA聚合酶(RdRp)(例如,由狂犬病病毒的L基因编码的聚合酶)从中合成信使RNA(mRNA)的互补链的基因组。在病毒基因组复制期间,RdRp合成正义反基因组,该正义反基因组用作用于创建基因组负义RNA的模板。因此,本领域技术人员将容易理解,当从负链基因组被提及时,表达元件可以从3'至5'取向,而不是从5'至3'取向。就负链基因组而言,编码本公开的tRNA-gRNA盒的核酸将从3'到5'包含第一tRNA、第一gRNA和任选的第二tRNA。从所述tRNA-gRNA盒表达的mRNA将从5'到3'包含第一tRNA、第一gRNA和任选的第二tRNA。

[0243] 如本文所用,术语"狂犬病毒属"是指属于弹状病毒科的负义单链RNA病毒属。狂犬病毒属颗粒是具有圆柱形形态的有包膜病毒,其宽约75nm且长约180nm。该结构包括由围绕螺旋核糖核蛋白核心的糖蛋白G组成的脂蛋白包膜。狂犬病毒属基因组含有编码促进基因组的转录和复制的蛋白质以及构成病毒的结构组分的蛋白质的五个基因。这五个基因是:编码狂犬病毒属核蛋白的N基因;编码狂犬病毒属磷蛋白的P基因;编码狂犬病毒属基质蛋白的M基因;编码狂犬病毒属包膜蛋白(也称为糖蛋白)的G基因;以及编码狂犬病毒属聚合酶的L基因。病毒基因组RNA和核蛋白一起形成用作狂犬病毒属聚合酶(一种依赖于RNA的RNA聚合酶)复制和转录的模板的核糖核蛋白。示例性狂犬病毒属包括但不限于狂犬病病毒(RABV)、莫科拉病毒(mokola virus,MOKV)、杜文哈基病毒(duvenhage virus,DUVV)、拉各斯蝙蝠病毒(lagos bat virus,LBV)和西高加索蝙蝠病毒(west caucasian bat virus,WCBV)。

[0244] 狂犬病病毒也称为狂犬病狂犬病毒属(Rabies lyssavirus),是弹状病毒科的狂犬病毒属的负义单链RNA病毒。狂犬病病毒具有圆柱形形态,并且结构包括由糖蛋白G组成的脂蛋白包膜,该脂蛋白包膜围绕螺旋核糖核蛋白芯。狂犬病病毒基因组含有五个基因,这些基因编码促进基因组转录和复制的蛋白以及构成该病毒的结构组分的蛋白。这五个基因是:编码狂犬病病毒核蛋白的N基因;编码狂犬病病毒磷蛋白的P基因;编码狂犬病病毒基质蛋白的M基因;编码狂犬病病毒糖蛋白的G基因;以及编码狂犬病病毒聚合酶的L基因。病毒基因组RNA和核蛋白一起形成充当狂犬病病毒聚合酶(RNA依赖性RNA聚合酶)复制和转录的

模板的核糖核蛋白。

[0245] 在某些实施方案中,本公开的重组狂犬病病毒基因组已被去除了一种或多种狂犬病病毒基因。例如,该重组狂犬病病毒基因组中可以不存在N基因、P基因、M基因、L基因和/或G基因。在某些实施方案中,重组狂犬病病毒基因组缺乏编码狂犬病病毒糖蛋白或其功能变体的G基因。缺乏编码狂犬病病毒糖蛋白的G基因的重组狂犬病病毒基因组阻止该病毒内源性地产生糖蛋白。由于该糖蛋白仅是病毒生命周期的最后步骤所必需的,因此这种缺失防止病毒扩散到最初被感染的细胞之外,但它并不能阻止病毒完成它在该点之前的整个复制周期。在某些实施方案中,重组狂犬病病毒基因组缺乏编码狂犬病病毒聚合酶或其功能变体的L基因。L基因产物是病毒基因转录和病毒基因组复制所必需的,并且L基因的缺失可能导致靶转导细胞的细胞毒性较低。参见例如Chatterjee等人,Nat.Neurosci.(2018)21(4):638-646,其公开内容通过引用整体并入本文。在某些实施方案中,重组狂犬病病毒基因组缺乏编码狂犬病病毒糖蛋白或其功能变体的G基因,并且缺乏编码狂犬病病毒聚合酶或其功能变体的L基因。

[0246] 本领域普通技术人员容易理解,如本文所述的缺乏狂犬病病毒基因的重组狂犬病病毒基因组是指缺乏狂犬病病毒基因的全部或一部分的狂犬病病毒基因组。例如,缺乏G基因的重组狂犬病病毒基因组可以缺乏G基因的全部或一部分,其中该G基因的该部分是G基因产物的功能所必需的。在某些实施方案中,缺乏G基因产物功能所必需的G基因的一部分可能导致产生截短的、无功能的糖蛋白。在某些实施方案中,缺乏L基因的重组狂犬病病毒基因组可以缺乏L基因的全部或一部分,其中该L基因的该部分是L基因产物的功能所必需的。在某些实施方案中,缺乏L基因产物功能所必需的L基因的一部分可能导致产生截短的、非功能性RNA依赖性RNA聚合酶。

[0247] 在某些实施方案中,本公开的重组狂犬病病毒基因组包含编码gRNA的核酸,该编码gRNA的核酸包含5'端和3'端。在某些实施方案中,重组狂犬病病毒基因组进一步包含编码转移RNA(tRNA)的核酸,该编码转移RNA(tRNA)的核酸位于编码gRNA的核酸的3'端或编码gRNA核酸的5'端。

[0248] 在某些实施方案中,本公开的重组狂犬病病毒基因组进一步包含编码转基因的核酸。在某些实施方案中,包含转基因的核酸替换了被去除的一种或多种狂犬病病毒基因,如本文所述。例如,包含转基因的核酸可以替换狂犬病病毒基因的全部或一部分。在某些实施方案中,包含转基因的核酸替换了G基因的全部或一部分,其中该G基因的该部分是G基因产物的功能所必需的。在某些实施方案中,包含转基因的核酸替换了L基因的全部或一部分,其中该L基因的该部分是L基因产物的功能所必需的。在某些实施方案中,包含转基因的核酸替换了L基因的全部或一部分,其中该L基因的该部分是L基因产物的功能所必需的;以及G基因的全部或一部分,其中该G基因的该部分是G基因产物的功能所必需的。

[0249] 在某些实施方案中,本公开的重组狂犬病病毒基因组编码包含转基因的核酸,其中该转基因替换了去除的一种或多种狂犬病病毒基因,如本文所述。在某些实施方案中,重组狂犬病病毒基因组包含编码狂犬病病毒核蛋白或其功能变体的N基因、编码狂犬病病毒磷蛋白或其功能变体的P基因,和/或编码狂犬病病毒基质蛋白或其功能变体的M基因。

[0250] N、P、M、L和G基因的示例性核酸序列以及这些基因的基因产物的氨基酸序列提供于表1中。

[0251] 表1:N、P、M、L和G的示例性序列

[0252]

SEQ ID NO:	序列
SEQ ID NO: 4001 N 基因 (核酸)	atggatgccgacaagattgtattcaaagtcaataatcaggtggtctctttgaagcctga gattatcgtggatcaatatgagtacaagtaccctgccatcaaagatttgaaaaagccct gtataaccctaggaaaggctcccgatttaaataaagcatacaagtcagttttgtcaggc atgagcgccgccaaacttaatcctgacgatgtatgttcctatttggcagcggcaatgc agtttttgaggggacatgtccggaagactggaccagctatggaattgtgattgcacg aaaaggagataagatcaccccaggttctctggtggagataaaacgtactgatgtaga agggaattgggctctgacaggaggcatggaactgacaagagaccccactgtccctg agcatgcgtccttagtcggtcttctcttgagtctgtataggttgagcaaaatatccgggc aaaacactggtaactataagacaaacattgcagacaggatagagcagatttttgagac agccccttttgttaaaatcgtggaacaccatactctaatgacaactcacaaaatgtgtg ctaattggagtactataccaaacttcagatttttggccggaacctatgacatgtttttctcc cggattgagcatctatattcagcaatcagagtgggcacagttgtcactgcttatgaaga ctgttcaggactggtatcatttactgggttcataaaacaaatcaatctcaccgctagaga ggcaatactatatttcttccacaagaactttgaggaagagataagaagaatgtttgagc cagggcaggagacagctgttcctcactcttatttcatccacttccgttcactaggcttga gtgggaaatctccttattcatcaaatgctgttggtcacgtgttcaatctcattcactttgta ggatgctatatgggtcaagtcagatccctaaatgcaacggttattgctgcatgtgctcc tcatgaaatgtctgttctaggggctatctgggagaggaattcttcgggaaagggaca tttgaaagaagattcttcagagatgagaaagaacttcaagaatacgaggcggctgaa ctgacaaagactgacgtagcactggcagatgatggaactgtcaactctgacgacga

[0253]

	ggactacttttcaggtgaaaccagaagtccggaggctgtttatactcgaatcatgatga atggaggtcgactaaagagatctcacatacggagatatgtctcagtcagttccaatcat caagcccgtccaaactcattcgccgagtttctaaacaagacatattcgagtgactca
SEQ ID NO: 4002 N 基因 (氨基酸)	MDADKIVFKVNNQVVSLKPEIIVDQYEYKYPAIKDLK KPCITLGKAPDLNKAYKSVLSGMSAAKLNPDDVCSY LAAAMQFFEGTCPEDWTSYGIVIARKGDKITPGSLVEI KRTDVEGNWALTGGMELTRDPTVPEHASLVGLLLSL YRLSKISGQNTGNYKTNIADRIEQIFETAPFVKIVEHHT LMTTHKMCANWSTIPNFRFLAGTYDMFFSRIEHLYSA IRVGTVVTAYEDCSGLVSFTGFIKQINLTAREAILYFFH KNFEEEIRRMFEPGQETAVPHSYFIHFRSLGLSGKSPY SSNAVGHVFNLIHFVGCYMGQVRSLNATVIAACAPH EMSVLGGYLGEEFFGKGTFERRFFRDEKELQEYEAAE LTKTDVALADDGTVNSDDEDYFSGETRSPEAVYTRIM MNGGRLKRSHIRRYVSVSSNHQARPNSFAEFLNKTYS SDS
SEQ ID NO: 4003 L 基因 (核酸)	ctcgatcctggagaggtctatgatgaccctattgacccaatcgagttagaggctgaac ccagaggaacccccattgtccccaacatcttgaggaactctgactacaatctcaactc tcctttgatagaagatcctgctagactaatgttagaatggttaaaaacagggaatagac cttatcggatgactctaacagacaattgctccaggtctttcagagttttgaaagattattt caagaaggtagatttgggttctctcaaggtgggcggaatggctgcacagtcaatgatt tctctctggttatatggtgcccactctgaatccaacaggagccggagatgtataacag acttggcccatttctattccaagtcgtcccccatagagaagctgttgaatctcacgctag gaaatagagggctgagaatcccccccagagggagtgttaagttgccttgagagggtt gattatgataatgcatttggaaggtatcttgccaacacgtattcctcttacttgttcttccat gtaatcaccttatacatgaacgccctagactgggatgaagaaaagaccatcctagcat tatggaaagatttaacctcagtggacatcgggaaggacttggtaaagttcaaagacca aatatggggactgctgatcgtgacaaaggactttgtttactcccaaagttccaattgtct ttttgacagaaactacacacttatgctaaaagatcttttcttgtctcgcttcaactccttaat ggtcttgctctctcccccagagccccgatactcagatgacttgatatctcaactatgcc agctgtacattgctggggatcaagtcttgtctatgtgtggaaactccggctatgaagtc atcaaaatattggagccatatgtcgtgaatagtttagtccagagagcagaaaagtttag gcctctcattcattccttgggagactttcctgtatttataaaagacaaggtaagtcaactt gaagagacgttcggtccctgtgcaagaaggttctttagggctctggatcaattcgaca acatacatgacttggtttttgtgtttggctgttacaggcattgggggcacccatatatag attatcgaaagggtctgtcaaaactatatgatcaggttcaccttaaaaaaatgatagata agtcctaccaggagtgcttagcaagcgacctagccaggaggatccttagatggggtt ttgataagtactccaagtggtatctggattcaagattcctagcccgagaccacccctttg actccttatatcaaaacccaaacatggccacccaaacatattgtagacttggtgggggg atacatggcacaagctcccgatcacgcagatctttgagattcctgaatcaatggatcc gtcagaaatattggatgacaaatcacattctttcaccagaacgagactagcttcttggct gtcagaaaaccgaggggggcctgttcctagcgaaaaagttattatcacggccctgtc taagccgcctgtcaatccccgagagtttctgaggtctatagacctcggaggattgcca

[0254]

gatgaagacttgataattggcctcaagccaaaggaacgggaattgaagattgaaggt
cgattctttgctctaatgtcatggaatctaagattgtattttgtcatcactgaaaaactcttg
gccaactacatcttgccacttttgacgcgctgactatgacagacaacctgaacaagg
tgtttaaaaagctgatcgacagggtcaccgggcaagggcttttggactattcaagggt
cacatatgcatttcacctggactatgaaaagtggaacaaccatcaaagattagagtca
acagaggatgtattttctgtcctagatcaagtgtttggattgaagagagtgttttctagaa
cacacgagtttttttcaaaaggcctggatctattattcagacagatcagacctcatcggg
ttacgggaggatcaaatatactgcttagatgcgtccaacggcccaacctgttggaatg
gccaggatggcgggctagaaggcttacggcagaagggctggagtctagtcagctta
ttgatgatagatagagaatctcaaatcaggaacacaagaaccaaaatactagctcaa
ggagacaaccaggttttatgtccgacatacatgttgtcgccagggctatctcaagagg
ggctcctctatgaattggagagaatatcaaggaatgcactttcgatatacagagccgt
cgaggaaggggcatctaagctagggctgatcatcaagaaagaagagaccatgtgta
gttatgacttcctcatctatggaaaaacccctttgtttagaggtaacatattggtgcctga
gtccaaaagatgggccagagtctcttgcgtctctaatgaccaaatagtcaacctcgcc
aatataatgtcgacagtgtccaccaatgcgctaacagtggcacaacactctcaatcttt
gatcaaaccgatgagggattttctgctcatgtcagtacaggcagtctttcactacctgct
atttagcccaatcttaaagggaagagtttacaagattctgagcgctgaaggggagag
ctttctcctagccatgtcaaggataatctatctagatccttctttgggagggatatctgga
atgtccctcggaagattccatatacgacagttctcagaccctgtctctgaagggttatc
cttctggagagagatctggttaagctcccaagagtcctggattcacgcgttgtgtcaa
gaggctggaaacccagatcttggagagagaacactcgagagcttcactcgccttcta
gaagatccgaccaccttaaatatcagaggaggggccagtcctaccattctactcaag
gatgcaatcagaaaggctttatatgacgaggtggacaaggtggaaaattcagagtttc
gagaggcaatcctgttgtccaagacccatagagataattttatactcttcttaatatctgtt
gagcctctgtttcctcgatttctcagtgagctattcagttcgtctttttgggaatccccga
gtcaatcattggattgatacaaaactcccgaacgataagaaggcagtttagaaagagt
ctctcaaaaactttagaagaatccttctacaactcagagatccacgggattagtcggat
gacccagacacctcagagggttgggggggtgtggccttgctcttcagagagggca
gatctacttagggagatctcttggggaagaaaagtggtaggcacgacagttcctcac
ccttctgagatgttgggattacttcccaagtcctctatttcttgcacttgtggagcaacag
gaggaggcaatcctagagtttctgtatcagtactcccgtcctttgatcagtcatttttttca
cgaggccccctaaagggatacttgggctcgtccacctctatgtcgacccagctattcc
atgcatgggaaaaagtcactaatgttcatgtggtgaagagagctctatcgttaaaaga
atctataaactggttcattactagagattccaacttggctcaagctctaattaggaacatt
atgtctctgacaggccctgatttccctctagaggaggcccctgtcttcaaaaggacgg
ggtcagccttgcataggttcaagtctgccagatacagcgaaggagggtattcttctgt
ctgcccgaacctcctctctcatatttctgttagtacagacaccatgtctgatttgacccaa
gacgggaagaactacgatttcatgttccagccattgatgctttatgcacagacatgga
catcagagctggtacagagagacacaaggctaagagactctacgtttcattggcacc
tccgatgcaacaggtgtgtgagacccattgacgacgtgaccctggagacctctcaga
tcttcgagtttccggatgtgtcgaaaagaatatccagaatggtttctggggctgtgcctc
acttccagaggcttcccgatatccgtctgagaccaggagattttgaatctctaagcggt

[0255]

	agagaaaagtctcaccatatcggatcagctcaggggctcttatactcaatcttagtggc aattcacgactcaggatacaatgatggaaccatcttccctgtcaacatatacggcaag gtttcccctagagactatttgagagggctcgcaaggggagtattgataggatcctcga tttgcttcttgacaagaatgacaaatatcaatattaatagacctcttgaattggtctcagg ggtaatctcatatattctcctgaggctagataaccatccctccttgtacataatgctcaga gaaccgtctcttagaggagagatattttctatccctcagaaaatccccgccgcttatcc aaccactatgaaagaaggcaacagatcaatcttgtgttatctccaacatgtgctacgct atgagcgagagataatcacggcgtctccagagaatgactggctatggatcttttcaga ctttagaagtgccaaaatgacgtacctatccctcattacttaccagtctcatcttctactc cagagggttgagagaaacctatctaagagtatgagagataacctgcgacaattgagt tctttgatgaggcaggtgctgggcgggcacggagaagataccttagagtcagacga caacattcaacgactgctaaaagactctttacgaaggacaagatgggtggatcaaga ggtgcgccatgcagctagaaccatgactggagattacagccccaacaagaaggtgt cccgtaaggtaggatgttcagaatgggtctgctctgctcaacaggttgcagtctctacc tcagcaaacccggcccctgtctcggagcttgacataagggccctctctaagaggttc cagaaccctttgatctcgggcttgagagtggttcagtgggcaaccggtgctcattata agcttaagcctattctagatgatctcaatgttttcccatctctctgccttgtagttggggac gggtcaggggggatatcaagggcagtcctcaacatgtttccagatgccaagcttgtg ttcaacagtcttttagaggtgaatgacctgatggcttccggaacacatccactgcctcc ttcagcaatcatgaggggaggaaatgatatcgtctccagagtgatagatcttgactca atctgggaaaaaccgtccgacttgagaaacttggcaacctggaaatacttccagtca gtccaaaagcaggtcaacatgtcctatgacctcattatttgcgatgcagaagttactga cattgcatctatcaaccggatcaccctgttaatgtccgattttgcattgtctatagatgga ccactctatttggtcttcaaaacttatgggactatgctagtaaatccaaactacaaggct attcaacacctgtcaagagcgttcccctcggtcacagggtttatcacccaagtaacttc gtctttttcatctgagctctacctccgattctccaaacgagggaagtttttcagagatgct gagtacttgacctcttccacccttcgagaaatgagccttgtgttattcaattgtagcagc cccaagagtgagatgcagagagctcgttccttgaactatcaggatcttgtgagaggat ttcctgaagaaatcatatcaaatccttacaatgagatgatcataactctgattgacagtg atgtagaatcttttctagtccacaagatggttgatgatcttgagttacagaggggaactc tgtctaaagtggctatcattatagccatcatgatagttttctccaacagagtcttcaacgtt tccaaacccctaactgacccctcgttctatccaccgtctgatcccaaaatcctgaggca cttcaacatatgttgcagtactatgatgtatctatctactgctttaggtgacgtccctagct tcgcaagacttcacgacctgtataacagacctataacttattacttcagaaagcaagtc attcgagggaacgtttatctatcttggagttggtccaacgacacctcagtgttcaaaag ggtagcctgtaattctagcctgagtctgtcatctcactggatcaggttgatttacaagat agtgaagactaccagactcgttggcagcatcaaggatctatccagagaagtggaaa gacaccttcataggtacaacaggtggatcaccctagaggatatcagatctagatcatc cctactagactacagttgcctg
SEQ ID NO: 4004	LDPGEVYDDPIDPIELEAEPRGTPIVPNILRNSDYNLNS PLIEDPARLMLEWLKTGNRPYRMTLTDNCSRSFRVLK DYFKKVDLGSLKVGGMAAQSMISLWLYGAHSESNRS RRCITDLAHFYSKSSPIEKLLNLTLGNRGLRIPPEGVLS

[0256]

L基因(氨基酸)	CLERVDYDNAFGRYLANTYSSYLFFHVITLYMNALD WDEEKTILALWKDLTSVDIGKDLVKFKDQIWGLLIVT KDFVYSQSSNCLFDRNYTLMLKDLFLSRFNSLMVLLS PPEPRYSDDLISQLCQLYIAGDQVLSMCGNSGYEVIKI LEPYVVNSLVQRAEKFRPLIHSLGDFPVFIKDKVSQLE ETFGPCARRFFRALDQFDNIHDLVFVFGCYRHWGHPY IDYRKGLSKLYDQVHLKKMIDKSYQECLASDLARRIL RWGFDKYSKWYLDSRFLARDHPLTPYIKTQTWPPKHI VDLVGDTWHKLPITQIFEIPESMDPSEILDDKSHSFTRT RLASWLSENRGGPVPSEKVIITALSKPPVNPREFLRSID LGGLPDEDLIIGLKPKERELKIEGRFFALMSWNLRLYF VITEKLLANYILPLFDALTMTDNLNKVFKKLIDRVTG QGLLDYSRVTYAFHLDYEKWNNHQRLESTEDVFSVL DQVFGLKRVFSRTHEFFQKAWIYYSDRSDLIGLREDQI YCLDASNGPTCWNGQDGGLEGLRQKGWSLVSLLMI DRESQIRNTRTKILAQGDNQVLCPTYMLSPGLSQEGL LYELERISRNALSIYRAVEEGASKLGLIIKKEETMCSY DFLIYGKTPLFRGNILVPESKRWARVSCVSNDQIVNLA NIMSTVSTNALTVAQHSQSLIKPMRDFLLMSVQAVFH YLLFSPILKGRVYKILSAEGESFLLAMSRIIYLDPSLGGI SGMSLGRFHIRQFSDPVSEGLSFWREIWLSSQESWIHA LCQEAGNPDLGERTLESFTRLLEDPTTLNIRGGASPTIL LKDAIRKALYDEVDKVENSEFREAILLSKTHRDNFILF LISVEPLFPRFLSELFSSSFLGIPESIIGLIQNSRTIRRQFR KSLSKTLEESFYNSEIHGISRMTQTPQRVGGVWPCSSE RADLLREISWGRKVVGTTVPHPSEMLGLLPKSSISCTC GATGGGNPRVSVSVLPSFDQSFFSRGPLKGYLGSSTS MSTQLFHAWEKVTNVHVVKRALSLKESINWFITRDS NLAQALIRNIMSLTGPDFPLEEAPVFKRTGSALHRFKS ARYSEGGYSSVCPNLLSHISVSTDTMSDLTQDGKNYD FMFQPLMLYAQTWTSELVQRDTRLRDSTFHWHLRCN RCVRPIDDVTLETSQIFEFPDVSKRISRMVSGAVPHFQ RLPDIRLRPGDFESLSGREKSHHIGSAQGLLYSILVAIH DSGYNDGTIFPVNIYGKVSPRDYLRGLARGVLIGSSIC FLTRMTNININRPLELVSGVISYILLRLDNHPSLYIMLR EPSLRGEIFSIPQKIPAAYPTTMKEGNRSILCYLQHVLR YEREIITASPENDWLWIFSDFRSAKMTYLSLITYQSHL LLQRVERNLSKSMRDNLRQLSSLMRQVLGGHGEDTL ESDDNIQRLLKDSLRRTRWVDQEVRHAARTMTGDYS PNKKVSRKVGCSEWVCSAQQVAVSTSANPAPVSELDI RALSKRFQNPLISGLRVVQWATGAHYKLKPILDDLNV FPSLCLVVGDGSGGISRAVLNMFPDAKLVFNSLLEVN DLMASGTHPLPPSAIMRGGNDIVSRVIDLDSIWEKPSD

[0257]

	LRNLATWKYFQSVQKQVNMSYDLIICDAEVTDIASIN RITLLMSDFALSIDGPLYLVFKTYGTMLVNPNYKAIQ HLSRAFPSVTGFITQVTSSFSSELYLRFSKRGKFFRDAE YLTSSTLREMSLVLFNCSSPKSEMQRARSLNYQDLVR GFPEEIISNPYNEMIITLIDSDVESFLVHKMVDDLELQR GTLSKVAIIIAIMIVFSNRVFNVSKPLTDPSFYPPSDPKI LRHFNICCSTMMYLSTALGDVPSFARLHDLYNRPITY YFRKQVIRGNVYLSWSWSNDTSVFKRVACNSSLSLSS HWIRLIYKIVKTTRLVGSIKDLSREVERHLHRYNRWIT LEDIRSRSSLLDYSCL
SEQ ID NO: 4005 M 基因 (核酸)	ttctagaagcagagaggaatctttgtcctcttcggacctttgtgtctgaagagacatgtc agaccatagttgacatgctctcgggttcatgttgatacaccagactctgccctggatat gacactgttttgcaatcactcttatttgcaatccgacgaactcagtatcatcatcccaagt gatctcctgagagtattccaactcctcccttcaagagggcccctggaatcagcccac tggaagataaaggttctcctcaatttgtatacccagttcaggccctcagggactggag atcctgacaaagccagtccaataaccactttgactaacccgatcatcctatgattccca gaatatatctcgtcgaatgatttcagaatgtgccgcaggatcctgaacgagtaaccatt cgggctacacactttaacccttccgttgatacaaaagttcctcatgttcttcttgcctgta agttctttcagcgggacgtattcaggggggtggaagccacaagtcatcgtcatccaga ggggctgacgcgggagaggatttttgagtgtcctcgtccctgcggtttttcactatctta cgtaggaggtt
SEQ ID NO: 4006 M 基因 (氨基酸)	NLLRKIVKNRRDEDTQKSSPASAPLDDDDLWLPPPEY VPLKELTGKKNMRNFCINGRVKVCSPNGYSFRILRHIL KSFDEIYSGNHRMIGLVKVVIGLALSGSPVPEGLNWV YKLRRTFIFQWADSRGPLEGEELEYSQEITWDDDTEF VGLQIRVIAKQCHIQGRVWCINMNPRACQLWSDMSL QTQRSEEDKDSSLLLE
SEQ ID NO: 4007 P 基因 (核酸)	agcaagatctttgtcaatcctagtgctattagagccggtctggccgatcttgagatggc tgaagaaactgttgatctgatcaatagaaatatcgaagacaatcaggctcatctccaa ggggaacccatagaggtggacaatctccctgaggatatggggcgacttcacctgga tgatggaaaatcgcccaaccatggtgagatagccaaggtgggagaaggcaagtatc gagaggactttcagatggatgaaggagaggatcctagcttcctgttccagtcatacct ggaaaatgttggagtccaaatagtcagacaaatgaggtcaggagagagatttctcaa gatatggtcacagaccgtagaagagattatatcctatgtcgcggtcaactttcccaacc ctccaggaaagtcttcagaggataaatcaacccagactactggccgagagctcaag aaggagacaacacccactccttctcagagagaaagccaatcatcgaaagccaggat ggcggctcaaattgcttctggccctccagcccttgaatggtcggctaccaatgaaga ggatgatctatcagtggaggctgagatcgctcaccagattgcagaaagtttctccaaa aaatataagtttccctctcgatcctcagggatactcttgtataattttgagcaattgaaaat gaaccttgatgatatagttaaagaggcaaaaaatgtaccaggtgtgacccgtttagcc catgacgggtccaaactccccctaagatgtgtactgggatgggtcgctttggccaact

[0258]

	ctaagaaattccagttgttagtcgaatccgacaagctgagtaaaatcatgcaagatga cttgaatcgctatacatcttgc
SEQ ID NO: 4008 P 基因 (氨基酸)	SKIFVNPSAIRAGLADLEMAEETVDLINRNIEDNQAHL QGEPIEVDNLPEDMGRLHLDDGKSPNHGEIAKVGEG KYREDFQMDEGEDPSFLFQSYLENVGVQIVRQMRSG ERFLKIWSQTVEEIISYVAVNFPNPPGKSSEDKSTQTT GRELKKETTPTPSQRESQSSKARMAAQIASGPPALEW SATNEEDDLSVEAEIAHQIAESFSKKYKFPSRSSGILLY NFEQLKMNLDDIVKEAKNVPGVTRLAHDGSKLPLRC VLGWVALANSKKFQLLVESDKLSKIMQDDLNRYTSC
SEQ ID NO: 4009 G 基因 (核酸)	atggttcctcaggctctcctgtttgtacccttctggttttccattgtgttttgggaaattc cctatttacacgataccagacaagcttggtccctggagtccgattgacatacatcacct cagctgcccaaacaatttggtagtggaggacgaaggatgcaccaacctgtcagggtt ctcctacatggaacttaaagttggatacatcttagccataaaagtgaacgggttcactt gcacaggcgttgtgacggaggctgaaacctacactaacttcgttggttatgtcacaac cacgttcaaaagaaagcatttccgcccaacaccagatgcatgtagagccgcgtacaa ctggaagatggccggtgaccccagatatgaagagtctctacacaatccgtaccctga ctaccgctggcttcgaactgtaaaaaccaccaaggagtctctcgttatcatatctccaa gtgtggcagatttggacccatatgacagatcccttcactcgagggtcttccctagcgg gaagtgctcaggagtagcggtgtcttctacctactgctccactaaccacgattacacc atttggatgcccgagaatccgagactagggatgtcttgtgacatttttaccaatagtag agggaagagagcatccaaagggagtgagacttgcggctttgtagatgaaagaggc ctatataagtctttaaaaggagcatgcaaactcaagttatgtggagttctaggacttaga cttatggatggaacatgggtctcgatgcaaacatcaaatgaaaccaaatggtgccctc ccgataagttggtgaacctgcacgactttcgctcagacgaaattgagcaccttgttgta gaggagttggtcaggaagagagaggagtgtctggatgcactagagtccatcatgac aaccaagtcagtgagtttcagacgtctcagtcatttaagaaaacttgtccctgggtttgg aaaagcatataccatattcaacaagaccttgatggaagccgatgctcactacaagtca gtcagaacttggaatgagatcctcccttcaaaagggtgtttaagagttgggggaggt gtcatcctcatgtgaacggggtgtttttcaatggtataatattaggacctgacggcaatg tcttaatcccagagatgcaatcatccctcctccagcaacatatggagttgttggaatcct cggttatccccccttgtgcacccccctggcagacccgtctaccgttttcaaggacggtga cgaggctgaggattttgttgaagttcaccttcccgatgtgcacaatcaggtctcagga gttgacttgggtctcccgaactgggggaagtatgtattactgagtgcaggggccctga ctgccttgatgttgataattttcctgatgacatgttgtagaagagtcaatcgatcagaac ctacgcaacacaatctcagagggacagggagggaggtgtcagtcactccccaaag cgggaagatcatatcttcatgggaatcacacaagagtgggggtgagaccagactg
SEQ ID NO: 4010 G 基因	MVPQALLFVPLLVFPLCFGKFPIYTIPDKLGPWSPIDIH HLSCPNNLVVEDEGCTNLSGFSYMELKVGYILAIKVN GFTCTGVVTEAETYTNFVGYVTTTFKRKHFRPTPDAC RAAYNWKMAGDPRYEESLHNPYPDYRWLRTVKTTK ESLVIISPSVADLDPYDRSLHSRVFPSGKCSGVAVSST YCSTNHDYTIWMPENPRLGMSCDIFTNSRGKRASKGS

[0259]

(氨基酸)	ETCGFVDERGLYKSLKGACKLKLCGVLGLRLMDGT WVSMQTSNETKWCPPDKLVNLHDFRSDEIEHLVVEE LVRKREECLDALESIMTTKSVSFRRLSHLRKLVPGFG KAYTIFNKTLMEADAHYKSVRTWNEILPSKGCLRVG GRCHPHVNGVFFNGIILGPDGNVLIPEMQSSLLQQHM ELLESSVIPLVHPLADPSTVFKDGDEAEDFVEVHLPDV HNQVSGVDLGLPNWGKYVLLSAGALTALMLIIFLMT CCRRVNRSEPTQHNLRGTGREVSVTPQSGKIISSWESH KSGGETRL

[0260] 在某些实施方案中,重组狂犬病病毒基因组包含具有与SEQ ID NO:4001中所示的核酸序列约60%、约65%、约70%、约75%、约80%、约81%、约82%、约83%、约84%、约85%、约86%、约87%、约88%、约89%、约90%、约91%、约92%、约93%、约94%、约95%、约96%、约97%、约98%、约99%同一的核酸序列的N基因。在某些实施方案中,重组狂犬病病毒基因组包含具有与SEQ ID NO:4001中所示的核酸序列至少60%、至少65%、约70%、至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%同一的核酸序列的N基因。在某些实施方案中,重组狂犬病病毒基因组包含含有SEQ ID NO:4001中所示的核酸序列的N基因。在某些实施方案中,重组狂犬病病毒基因组包含由SEQ ID NO:4001中所示的核酸序列组成的N基因。在某些实施方案中,N基因编码与SEQ ID NO:4002中所示的氨基酸序列约80%、约81%、约82%、约83%、约84%、约85%、约86%、约87%、约88%、约89%、约90%、约91%、约92%、约93%、约94%、约95%、约96%、约97%、约98%、约99%同一的氨基酸序列。在某些实施方案中,N基因编码与SEQ ID NO:4002中所示的氨基酸序列至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%同一的氨基酸序列。在某些实施方案中,N基因编码包含SEQ IDNO:4002中所示的氨基酸序列的氨基酸序列。在某些实施方案中,N基因编码由SEQ ID NO:4002中所示的氨基酸序列组成的氨基酸序列。

[0261] 在某些实施方案中,重组狂犬病病毒基因组包含具有与SEQ ID NO:4003中所示的核酸序列约60%、约65%、约70%、约75%、约80%、约81%、约82%、约83%、约84%、约85%、约86%、约87%、约88%、约89%、约90%、约91%、约92%、约93%、约94%、约95%、约96%、约97%、约98%、约99%同一的核酸序列的L基因。在某些实施方案中,重组狂犬病病毒基因组包含具有与SEQ ID NO:4003中所示的核酸序列至少60%、至少65%、约70%、至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%同一的核酸序列的L基因。在某些实施方案中,重组狂犬病病毒基因组包含含有SEQ ID NO:4003中所示的核酸序列的L基因。在某些实施方案中,重组狂犬病病毒基因组包含由SEQ ID NO:4003中所示的核酸序列组成的L基因。在某些实施方案中,L基因编码与SEQ ID NO:4004中所示的氨基酸序列约80%、约81%、约82%、约83%、约84%、约85%、约86%、约87%、约88%、约89%、约90%、约91%、约92%、约93%、约

94％、约95％、约96％、约97％、约98％、约99％同一的氨基酸序列。在某些实施方案中,L基因编码与SEQ ID NO:4004中所示的氨基酸序列至少80％、至少81％、至少82％、至少83％、至少84％、至少85％、至少86％、至少87％、至少88％、至少89％、至少90％、至少91％、至少92％、至少93％、至少94％、至少95％、至少96％、至少97％、至少98％、至少99％同一的氨基酸序列。在某些实施方案中,L基因编码包含SEQ ID NO:4004中所示的氨基酸序列的氨基酸序列。在某些实施方案中,L基因编码由SEQ ID NO:4004中所示的氨基酸序列组成的氨基酸序列。

[0262] 在某些实施方案中,重组狂犬病病毒基因组包含具有与SEQ ID NO:4005中所示的核酸序列约60％、约65％、约70％、约75％、约80％、约81％、约82％、约83％、约84％、约85％、约86％、约87％、约88％、约89％、约90％、约91％、约92％、约93％、约94％、约95％、约96％、约97％、约98％、约99％同一的核酸序列的M基因。在某些实施方案中,重组狂犬病病毒基因组包含具有与SEQ ID NO:4005中所示的核酸序列至少60％、至少65％、约70％、至少75％、至少80％、至少81％、至少82％、至少83％、至少84％、至少85％、至少86％、至少87％、至少88％、至少89％、至少90％、至少91％、至少92％、至少93％、至少94％、至少95％、至少96％、至少97％、至少98％、至少99％同一的核酸序列的M基因。在某些实施方案中,重组狂犬病病毒基因组包含含有SEQ ID NO:4005中所示的核酸序列的M基因。在某些实施方案中,重组狂犬病病毒基因组包含由SEQ ID NO:4005中所示的核酸序列组成的M基因。在某些实施方案中,M基因编码与SEQ ID NO:4006中所示的氨基酸序列约80％、约81％、约82％、约83％、约84％、约85％、约86％、约87％、约88％、约89％、约90％、约91％、约92％、约93％、约94％、约95％、约96％、约97％、约98％、约99％同一的氨基酸序列。在某些实施方案中,M基因编码与SEQ ID NO:4006中所示的氨基酸序列至少80％、至少81％、至少82％、至少83％、至少84％、至少85％、至少86％、至少87％、至少88％、至少89％、至少90％、至少91％、至少92％、至少93％、至少94％、至少95％、至少96％、至少97％、至少98％、至少99％同一的氨基酸序列。在某些实施方案中,M基因编码包含SEQ ID NO:4006中所示的氨基酸序列的氨基酸序列。在某些实施方案中,M基因编码由SEQ ID NO:4006中所示的氨基酸序列组成的氨基酸序列。

[0263] 在某些实施方案中,重组狂犬病病毒基因组包含具有与SEQ ID NO:4007中所示的核酸序列约60％、约65％、约70％、约75％、约80％、约81％、约82％、约83％、约84％、约85％、约86％、约87％、约88％、约89％、约90％、约91％、约92％、约93％、约94％、约95％、约96％、约97％、约98％、约99％同一的核酸序列的P基因。在某些实施方案中,重组狂犬病病毒基因组包含具有与SEQ ID NO:4007中所示的核酸序列至少60％、至少65％、约70％、至少75％、至少80％、至少81％、至少82％、至少83％、至少84％、至少85％、至少86％、至少87％、至少88％、至少89％、至少90％、至少91％、至少92％、至少93％、至少94％、至少95％、至少96％、至少97％、至少98％、至少99％同一的核酸序列的P基因。在某些实施方案中,重组狂犬病病毒基因组包含含有SEQ ID NO:4007中所示的核酸序列的P基因。在某些实施方案中,重组狂犬病病毒基因组包含由SEQ ID NO:4007中所示的核酸序列组成的P基因。在某些实施方案中,P基因编码与SEQ ID NO:4008中所示的氨基酸序列约80％、约81％、约82％、约83％、约84％、约85％、约86％、约87％、约88％、约89％、约90％、约91％、约92％、约93％、约94％、约95％、约96％、约97％、约98％、约99％同一的氨基酸序列。在某些实施方案中,P基

因编码与SEQ ID NO:4008中所示的氨基酸序列至少80％、至少81％、至少82％、至少83％、至少84％、至少85％、至少86％、至少87％、至少88％、至少89％、至少90％、至少91％、至少92％、至少93％、至少94％、至少95％、至少96％、至少97％、至少98％、至少99％同一的氨基酸序列。在某些实施方案中,P基因编码包含SEQ ID NO:4008中所示的氨基酸序列的氨基酸序列。在某些实施方案中,P基因编码由SEQ ID NO:4008中所示的氨基酸序列组成的氨基酸序列。

[0264] 在某些实施方案中,重组狂犬病病毒基因组包含具有与SEQ ID NO:4009中所示的核酸序列约60％、约65％、约70％、约75％、约80％、约81％、约82％、约83％、约84％、约85％、约86％、约87％、约88％、约89％、约90％、约91％、约92％、约93％、约94％、约95％、约96％、约97％、约98％、约99％同一的核酸序列的G基因。在某些实施方案中,重组狂犬病病毒基因组包含具有与SEQ ID NO:4009中所示的核酸序列至少60％、至少65％、约70％、至少75％、至少80％、至少81％、至少82％、至少83％、至少84％、至少85％、至少86％、至少87％、至少88％、至少89％、至少90％、至少91％、至少92％、至少93％、至少94％、至少95％、至少96％、至少97％、至少98％、至少99％同一的核酸序列的G基因。在某些实施方案中,重组狂犬病病毒基因组包含含有SEQ ID NO:4009中所示的核酸序列的G基因。在某些实施方案中,重组狂犬病病毒基因组包含由SEQ ID NO:4009中所示的核酸序列组成的G基因。在某些实施方案中,G基因编码与SEQ ID NO:4010中所示的氨基酸序列约80％、约81％、约82％、约83％、约84％、约85％、约86％、约87％、约88％、约89％、约90％、约91％、约92％、约93％、约94％、约95％、约96％、约97％、约98％、约99％同一的氨基酸序列。在某些实施方案中,G基因编码与SEQ ID NO:4010中所示的氨基酸序列至少80％、至少81％、至少82％、至少83％、至少84％、至少85％、至少86％、至少87％、至少88％、至少89％、至少90％、至少91％、至少92％、至少93％、至少94％、至少95％、至少96％、至少97％、至少98％、至少99％同一的氨基酸序列。在某些实施方案中,G基因编码包含SEQ ID NO:4010中所示的氨基酸序列的氨基酸序列。在某些实施方案中,G基因编码由SEQ ID NO:4010中所示的氨基酸序列组成的氨基酸序列。

[0265] 本公开的重组狂犬病病毒基因组内包含的基因中的每个基因均可以可操作地连接至转录调控元件。在其中该基因连接在单个表达盒上的某些实施方案中,单个转录调控元件可能能够控制该基因的表达。在某些实施方案中,每个基因均可操作地连接至单独的转录调控元件。在某些实施方案中,每个基因的转录调控元件可以是相同的。在某些实施方案中,每个基因的转录调控元件可以是不同的。

[0266] 在某些实施方案中,基因中的每个基因可操作地连接至转录调控元件,其中该转录调控元件能够控制与其可操作地连接的基因的表达。在某些实施方案中,转录调控元件包含转录起始信号。该转录起始信号对于狂犬病病毒来说可以是内源的或外源的。在某些实施方案中,转录起始信号是合成的转录起始信号。在某些实施方案中,编码转基因的核酸进一步可操作地连接至转录终止多聚腺苷酸化信号。该转录终止多聚腺苷酸化信号对于狂犬病病毒来说可以是内源的或外源的。在某些实施方案中,转录终止多聚腺苷酸化信号是合成的转录终止多聚腺苷酸化信号。合适的转录起始信号和转录终止多聚腺苷酸化信号的实例是本领域普通技术人员已知的,并且描述于例如Albertini等人,Adv.Virus.Res.(2011)79:1-22;Ogino和Green,Viruses(2019)11(6):504;Ogino等人,Nucl.Acids.Res.

(2019)47(1):299-309;以及Ogino和Green,Front.Microbiol.(2019)10:1490,这些文献的公开内容以引用的方式整体并入本文。

[0267] C.引导RNA和编码引导RNA的重组负链RNA病毒基因组

[0268] 在一方面,本公开提供了一种重组负链RNA病毒基因组,其包含编码第一引导RNA(gRNA)的核酸,该核酸包含5'端和3'端;和编码第一转移RNA(tRNA)的核酸,该核酸位于编码第一gRNA的核酸的3'端或编码第一gRNA的核酸的5'端中的一者或两者处。

[0269] 在某些实施方案中,重组负链RNA病毒基因组包含编码第二tRNA的核酸。在某些实施方案中,重组负链RNA病毒基因组包含编码第三tRNA的核酸。在某些实施方案中,重组负链RNA病毒基因组包含编码第四tRNA的核酸。在某些实施方案中,重组负链RNA病毒基因组包含编码第五tRNA的核酸。

[0270] 在某些实施方案中,编码第一tRNA的核酸位于编码第一gRNA的核酸的3'端;并且编码第二tRNA的核酸位于编码第一gRNA的核酸的5'端。

[0271] 在某些实施方案中,第一tRNA的核苷酸序列和第二tRNA、第三tRNA、第四tRNA和/或第五tRNA的核苷酸序列是至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一的。

[0272] 在某些实施方案中,第一tRNA和第二tRNA、第三tRNA、第四tRNA和/或第五tRNA指定相同的氨基酸。例如,第一tRNA和第二tRNA具有不同的反密码子环序列,每个反密码子环序列对应于相同的氨基酸(例如,具有包含5'GGC 3'的反密码子环序列的第一tRNA指定Ala,并且具有包含5'AGC 3'的反密码子环序列的第二tRNA也指定Ala)。

[0273] 在某些实施方案中,第一tRNA和第二tRNA、第三tRNA、第四tRNA和/或第五tRNA指定不同的氨基酸。例如,第一tRNA和第二tRNA具有不同的反密码子环序列,每个反密码子环序列对应于不同的氨基酸(例如,具有包含5'GGC 3'的反密码子环序列的第一tRNA指定Ala,并且具有包含5'AAA 3'的反密码子环序列的第二tRNA指定Phe)。

[0274] 在某些实施方案中,重组负链RNA病毒基因组包含编码第一tRNA、第二tRNA、第三tRNA、第四tRNA和/或第五tRNA的两种或更多种核酸。在某些实施方案中,重组负链RNA病毒基因组包含编码第一tRNA、第二tRNA、第三tRNA、第四tRNA和/或第五tRNA的两种核酸。在某些实施方案中,重组负链RNA病毒基因组包含编码第一tRNA、第二tRNA、第三tRNA、第四tRNA和/或第五tRNA的三种核酸。在某些实施方案中,重组负链RNA病毒基因组包含编码第一tRNA、第二tRNA、第三tRNA、第四tRNA和/或第五tRNA的四种核酸。在某些实施方案中,重组负链RNA病毒基因组包含编码第一tRNA、第二tRNA、第三tRNA、第四tRNA和/或第五tRNA的五种核酸。

[0275] 在某些实施方案中,重组负链RNA病毒基因组包含编码第二gRNA、第三gRNA、第四gRNA和/或第五gRNA的核酸。

[0276] 在某些实施方案中,该两种或更多种核酸编码同一的gRNA。在某些实施方案中,两种或更多种核酸编码至少一种不同的gRNA。在某些实施方案中,第一gRNA的核苷酸序列和第二gRNA、第三gRNA、第四gRNA和/或第五gRNA的核苷酸序列是至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一的。

[0277] 在某些实施方案中,第一gRNA和第二gRNA、第三gRNA、第四gRNA和/或第五gRNA与相同的靶核酸序列特异性地杂交。在某些实施方案中,第一gRNA和第二gRNA、第三gRNA、第

四gRNA和/或第五gRNA与不同的靶核酸序列特异性地杂交。

[0278] 在某些实施方案中,第一tRNA、第二tRNA、第三tRNA、第四tRNA和/或第五tRNA各自选自由以下组成的组:tRNA-ala、tRNA-arg、tRNA-asn、tRNA-asp、tRNA-cys、tRNA-gln、tRNA-gly、tRNA-his、tRNA-ile、tRNA-leu、tRNA-lys、tRNA-met、tRNA-phe、tRNA-pro、tRNA-pyl、tRNA-sec、tRNA-ser、tRNA-thr、tRNA-trp、tRNA-tyr和tRNA-val。

[0279] 在某些实施方案中,编码第一tRNA、第二tRNA、第三tRNA、第四tRNA和/或第五tRNA的核酸包括以下中的任一种:

[0280] GGCTCGTTGGTCTAGGGGTATGATTCTCGCTTAGGGTGCG AGAGGTCCCGGGTTCAAATCCCGGACG AGCCC(tRNA-pro;SEQ ID NO:4011),或与其至少90%(例如,90%、91%、92%、93%、94%、95%、96%、97%、98%或99%)同一的序列;

[0281] GGCTCCATAGCTCAGGGGTTAGAGCACTGGTCTTGTAAAC CAGGGGTCGCGAGTTCAATTCTCGCTG GGGCTT(tRNA-thr;SEQ ID NO:4012),或与其至少90%(例如,90%、91%、92%、93%、94%、95%、96%、97%、98%或99%)同一的序列;

[0282] GCGTTGGTGGTATAGTGGTGAGCATAGCTGCCTTCCAAGC AGTTGACCCGGGTTCGATTCCCGGCCA ACGCA(tRNA-gly G8;SEQ ID NO:4013),或与其至少90%(例如,90%、91%、92%、93%、94%、95%、96%、97%、98%或99%)同一的序列;

[0283] GCATGGGTGGTTCAGTGGTAGAATTCTCGCCTGCCACGCG GGAGGCCCGGGTTCGATTCCCGGCCCATGCA(tRNA-gly G27;SEQ ID NO:4014),或与其至少90%(例如,90%、91%、92%、93%、94%、95%、96%、97%、98%或99%)同一的序列;

[0284] GTCAGGATGGCCGAGCGGTCTAAGGCGCTGCGTTCAGGTC GCAGTCTCCCCTAGAGGCGTGGGTTCG AATCCCACTCCTGACA(tRNA-leu;SEQ ID NO:4015),或与其至少90%(例如,90%、91%、92%、93%、94%、95%、96%、97%、98%或99%)同一的序列;

[0285] GCTCCAGTGGCGCAATCGGTTAGCGCGCGGTACTTATAAG ACAGTGCACCTGTGAGCAATGCCGAGG TTGTGAGTTCAAGCCT CACCTGGAGCA(tRNA-ile;SEQ ID NO:4016),或与其至少90%(例如,90%、91%、92%、93%、94%、95%、96%、97%、98%或99%)同一的序列;

[0286] GAAAAAGTCATGGAGGCCATGGGGTTGGCTTGAAACCAG CTTTGGGGGGTTCGATTCCTTCCTTTTT TGTCT(tRNA-ser;SEQ ID NO:4017),或与其至少90%(例如,90%、91%、92%、93%、94%、95%、96%、97%、98%或99%)同一的序列;

[0287] GGGCCAGTGGCGCAATGGATAACGCGTCTGACTACGGATC AGAAGATTCCAGGTTCGACTCCTGGCT GGCTCGGTGTA(tRNA-arg;SEQ ID NO:4018),或与其至少90%(例如,90%、91%、92%、93%、94%、95%、96%、97%、98%或99%)同一的序列;

[0288] AAACAAGCGCAAGTGGTTTAGTGGTAAAATCCAACGTTGC CATCGTTGGGCCCCCGGTTCGATTCCG GGCTTGCGCA(tRNA-asp1;SEQ ID NO:4019),或与其至少90%(例如,90%、91%、92%、93%、94%、95%、96%、97%、98%或99%)同一的序列;

[0289] AACAAAGCACCAGTGGTCTAGTGGTAGAATAGTACCCTGC CACGGTACAGACCCGGGTTCGATTCCC GGCTGGTGCA(tRNA-asp2;SEQ ID NO:4020),或与其至少90%(例如,90%、91%、92%、93%、94%、95%、96%、97%、98%或99%)同一的序列;或

[0290] TCCTCGTTAGTATAGTGGTGAGTATCCCCGCCTGTCACGCG GGAGACCGGGGTTCGATTCCCCGACGGGGAG(tRNA-asp D15;SEQ ID NO:4021),或与其至少90%

(例如,90％、91％、92％、93％、94％、95％、96％、97％、98％或99％)同一的序列。

[0291] 在某些实施方案中,重组负链RNA病毒基因组包含编码负链RNA病毒基因的核酸。

[0292] 在某些实施方案中,重组负链RNA病毒基因组包含编码转基因(例如核碱基编辑器)的核酸。

[0293] 在某些实施方案中,编码第一gRNA的核酸和编码第一tRNA的核酸位于各自编码负链RNA病毒基因的两种核酸之间。

[0294] 在某些实施方案中,编码第一gRNA的核酸和编码第一tRNA的核酸位于各自编码转基因的两种核酸之间。

[0295] 在某些实施方案中,编码第一gRNA的核酸和编码第一tRNA的核酸位于编码负链RNA病毒基因的核酸和编码转基因的核酸之间。

[0296] 在某些实施方案中,重组负链RNA病毒基因组包含gRNA表达盒,该gRNA表达盒从3'至5'包含负链RNA病毒转录起始信号、编码tRNA的核酸、编码gRNA的核酸和转录终止多聚腺苷酸化信号。

[0297] 在某些实施方案中,重组负链RNA病毒基因组包含gRNA表达盒,该gRNA表达盒从3'至5'包含负链RNA病毒转录起始信号、编码第一tRNA的核酸、编码第一gRNA的核酸、编码第二tRNA的核酸和转录终止多聚腺苷酸化信号。

[0298] 在某些实施方案中,重组负链RNA病毒基因组包含gRNA表达盒,该gRNA表达盒从3'至5'包含负链RNA病毒转录起始信号、编码第一tRNA的核酸、编码第一gRNA的核酸、编码第二tRNA的核酸、编码第二gRNA的核酸和转录终止多聚腺苷酸化信号。

[0299] 在某些实施方案中,重组负链RNA病毒基因组包含gRNA表达盒,该gRNA表达盒从3'至5'包含负链RNA病毒转录起始信号、编码第一tRNA的核酸、编码第一gRNA的核酸、编码第二tRNA的核酸、编码第二gRNA的核酸和转录终止多聚腺苷酸化信号。

[0300] 在某些实施方案中,重组负链RNA病毒基因组包含gRNA表达盒,该gRNA表达盒从3'至5'包含负链RNA病毒转录起始信号、编码第一tRNA的核酸、编码第一gRNA的核酸、编码第二tRNA的核酸、编码第二gRNA的核酸、编码第三tRNA的核酸和转录终止多聚腺苷酸化信号。

[0301] 在gRNA表达盒的某些实施方案中,编码第一tRNA、第二tRNA和/或第三tRNA的核酸是同一的。在gRNA表达盒的某些实施方案中,编码第一tRNA、第二tRNA和/或第三tRNA的核酸是不同的。在gRNA表达盒的某些实施方案中,第一tRNA的核苷酸序列和第二tRNA的核苷酸序列是至少90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％同一的。在gRNA表达盒的某些实施方案中,第一tRNA和第二tRNA指定相同的氨基酸。在gRNA表达盒的某些实施方案中,第一tRNA和第二tRNA指定不同的氨基酸。在gRNA表达盒的某些实施方案中,编码第一gRNA和/或第二gRNA的核酸是同一的。在gRNA表达盒的某些实施方案中,编码第一gRNA和/或第二gRNA的核酸是不同的。在gRNA表达盒的某些实施方案中,第一gRNA的核苷酸序列和第二gRNA的核苷酸序列是至少90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％同一的。在gRNA表达盒的某些实施方案中,第一gRNA和第二gRNA与相同的靶核酸序列特异性地杂交。在gRNA表达盒的某些实施方案中,第一gRNA和第二gRNA与不同的靶核酸序列特异性地杂交。在gRNA表达盒的某些实施方案中,转录终止多聚腺苷酸化信号包含内源转录终止多聚腺苷酸化信号。在gRNA表达盒的某些实施方案中,转录终止多聚腺苷化信号包含异源转录终止多聚腺苷化信号。

[0302] 在某些实施方案中,本公开的tRNA(例如,第一tRNA、第二tRNA、第三tRNA、第四tRNA或第五tRNA)包含tRNA样结构。tRNA样结构以与上文所述的tRNA类似的方式运作。具体地,tRNA样结构是包含二级结构的RNA分子,该二级结构可以充当参与tRNA成熟的细胞RNase(诸如RNase P或RNase Z)的底物。在某些实施方案中,tRNA样结构包含tRNA变体、tRNA片段、病毒tRNA或mascRNA。

[0303] MALAT1相关的小细胞质RNA(mascRNA):

[0304] MALAT1相关的小细胞质RNA(mascRNA)是在细胞溶胶中发现的非编码RNA。它们由酶RNase P从较长的非编码RNA(称为MALAT1)加工而成。MascRNA在结构上与tRNA相似(包括由Rnase P进行的加工),但未被氨酰化。MascRNA更详细地描述于Wilusz等人(Cell.2008年11月28日;135(5):919-932)中,其全部内容通过引用并入本文。

[0305] 在某些实施方案中,mascRNA由包含以下中的任一种的核酸编码:

[0306] AAAAGCAAAAGATGCTGGTGGTTGGCACTCCTGGTTTCCAGGACGGGGTTCAAATCCCTGCGGCGTCTTTGCTTT(masc_Malat1;SEQ ID NO:X),或与其至少90%(例如,90%、91%、92%、93%、94%、95%、96%、97%、98%或99%)同一的序列;

[0307] AAAGACGCTGGTGGTTGGTGTTTCCAGGACGGGGTTCAAGTCCCTGCGGCGTCCTCGC(masc_liz38;SEQ ID NO:X),或与其至少90%(例如,90%、91%、92%、93%、94%、95%、96%、97%、98%或99%)同一的序列;

[0308] GGCTCTGGTGGCTTCCAGGACGGGGTTCAAGTCCCTGCAG TGCCCTTGCTGA(masc_liz40;SEQ ID NO:X),或与其至少90%(例如,90%、91%、92%、93%、94%、95%、96%、97%、98%或99%)同一的序列;

[0309] AAAGGCGCTGGTGGTGGCACTCCCAGCGGGACGGGGTTCG AATCCCCGCGGCGCCTCTGC(masc_turk;SEQ ID NO:X),或与其至少90%(例如,90%、91%、92%、93%、94%、95%、96%、97%、98%或99%)同一的序列;

[0310] GAAGGTTTTTCTTTTCCTGAGAAAACAACACGTATTGTTTTCTCAGGTTTTGCTTTTTGGCCTTTTTCTAGCTTAAAAAAAAAAGCAAAAGATGCTGGTGGTTGGCACTCCTGGTTTCCAGGACGGGGTTCAAATCCCTGCGGCGTCTTTGCTTT(hMALAT1.1;SEQ ID NO:X),或与其至少90%(例如,90%、91%、92%、93%、94%、95%、96%、97%、98%或99%)同一的序列;

[0311] GCAGGTGTTTCTTTTACTGAGTGCAGCCCATGGCCGCACTCAGGTTTTGCTTTTCACCTTCCCATCTGTGAAAGAGTGAGCAGGAAAAAGCAAAAGGCGCTGGTGGTGGCACGTCCAGCACGGCTGGGCCGGGGTTCGAGTCCCCGCAGTGTTGCTGC(hMALAT1.2;SEQ ID NO:X),或与其至少90%(例如,90%、91%、92%、93%、94%、95%、96%、97%、98%或99%)同一的序列;

[0312] GAAGGTTTTTCTTTTCCTGAGAAAACAACACGTTTTGTTTTCTCAGGTTTTGCTTTTTGGCCTTTTTCTAGCTTAAAAAAAAAAGCAAAAGATGCTGGTGGTTGGCACTCCTGGTTTCCAGGACAGGGTTCAAATCCCTGCGGCGTCTTTGCTTT(黑猩猩1;SEQ ID NO:X),或与其至少90%(例如,90%、91%、92%、93%、94%、95%、96%、97%、98%或99%)同一的序列;

[0313] AAAGCAAAAGATGCTGGTGGTTGGCACTCCTGGTTTCCAGGACAGGGTTCAAATCCCTGCGGCGTCTTTGCTTT(黑猩猩1短;SEQ ID NO:X),或与其至少90%(例如,90%、91%、92%、93%、94%、95%、96%、97%、98%或99%)同一的序列;

[0314] GCAGGTGTTTCTTTTACTGAGTGCAGCCCATGGCCGCACTCAGGTTTTGCTTTTCACCTTCCCATCTG

TGAAAGAGTGAGCAGGAAAAAGCAAAAGGCGCTGGTGGTGGCACGTCCAGCACGGCTGGGCCGGGGTTCGAGTCCCCGCAGTGTTGCTGC(chimp.2;SEQ ID NO:X),或与其至少90％(例如,90％、91％、92％、93％、94％、95％、96％、97％、98％或99％)同一的序列;

[0315] AAAGGTTTTTCTTTTCCTGAGAAAACAACCTTTTGTTTTCTCAGGTTTTGCTTTTTGGCCTTTCCCTAGCTTTAAAAAAAAAAGCAAAAGACGCTGGTGGCTGGCACTCCTGGTTTCCAGGACGGGGTTCAAGTCCCTGCGGTGTCTTTGC(MoTse.1;SEQ ID NO:X),或与其至少90％(例如,90％、91％、92％、93％、94％、95％、96％、97％、98％或99％)同一的序列;

[0316] AAAGCAAAAGACGCTGGTGGCTGGCACTCCTGGTTTCCAGGACGGGGTTCAAGTCCCTGCGGTGTCTTTGCTTGAC(MoTse.1短;SEQ ID NO:X),或与其至少90％(例如,90％、91％、92％、93％、94％、95％、96％、97％、98％或99％)同一的序列;或

[0317] GCAGGTGTTTCTTTTCCTGACCGCGGCTCATGGCCGCGCTCAGGTTTTGCTTTTCACCTTTGTCTGAGAGAACGAACGTGAGCAGGAAAAAGCAAAAGGCACTGGTGGCGGCACGCCCGCACCTCGGGCCAGGGTTCGAGTCCCTGCAGTACCGTGC(MoTse.2;SEQ ID NO:X),或与其至少90％(例如,90％、91％、92％、93％、94％、95％、96％、97％、98％或99％)同一的序列。

[0318] 转移RNA变体:

[0319] tRNA变体是包含相对于野生型tRNA或未取代的tRNA的一个或多个核苷酸取代或缺失的tRNA。该取代可被用于增强tRNA变体相对于相应野生型或未取代的tRNA的稳定性。在某些实施方案中,tRNA变体包含用G或C核苷酸对一个或多个A和/或T核苷酸进行的取代。在某些实施方案中,tRNA变体包含相对于野生型tRNA较低的A和/或T核苷酸含量。

[0320] 在某些实施方案中,tRNA变体由包含以下中的任一种的核酸编码:

[0321] GGCTCGTTGGCCTAGGGGTATGGCTCCCGCTTAGGGTGCGGGAGGTCCCGGGTTCAAATCCCGGACGAGCC(tRNA-pro var1;SEQ ID NO:X),或与其至少90％(例如,90％、91％、92％、93％、94％、95％、96％、97％、98％或99％)同一的序列;

[0322] GGCTCGTTGGCCTAGGGGTATGGCTGAAAAGGTCCCGGGTTCAAATCCCGGACGAGCC(tRNA-pro var2;SEQ ID NO:X),或与其至少90％(例如,90％、91％、92％、93％、94％、95％、96％、97％、98％或99％)同一的序列;

[0323] GGCTCGTTGAAAGAAAAGGTCCCGGGTTCAAATCCCGGACGAGCC(tRNA-pro var3;SEQ ID NO:X),或与其至少90％(例如,90％、91％、92％、93％、94％、95％、96％、97％、98％或99％)同一的序列;

[0324] GGCTCCATAGCGCAGGGGTTAGCGCACCGGTCTTGTAAACCGGGGGTCGCGAGTTCAATTCTCGCTGGGGCTT(tRNA-thr var1;SEQ ID NO:X),或与其至少90％(例如,90％、91％、92％、93％、94％、95％、96％、97％、98％或99％)同一的序列;

[0325] GGCTCCATAGCGCAGGGGTTAGCGCAGAAAGGGTCGCGAGTTCAATTCTCGCTGGGGCTT(tRNA-thr var2;SEQ ID NO:X),或与其至少90％(例如,90％、91％、92％、93％、94％、95％、96％、97％、98％或99％)同一的序列;或

[0326] GGCTCCATAGAAAGAAAGAAAGGGTCGCGAGTTCAATTCTCGCTGGGGCTT(tRNA-thr var3;SEQ ID NO:X),或与其至少90％(例如,90％、91％、92％、93％、94％、95％、96％、97％、98％或99％)同一的序列。

[0327] 转移RNA片段:

[0328] tRNA片段是包含相对于野生型tRNA或未取代的tRNA的截短的tRNA。在某些实施方案中,tRNA片段包含分裂tRNA,该分裂tRNA包含能够杂交从而形成完整tRNA的两个单独的tRNA部分。tRNA片段(包括分裂tRNA)保留了Rnase P切割能力。

[0329] 病毒tRNA样结构:

[0330] 病毒tRNA样结构(vtRNA)由病毒基因组表达,并且由很像内源tRNA的细胞机器加工。vtRNA更详细地描述于Bowden等人(.Gen Virol.78:1675-1687.1997)和Dreher(Wiley Interdiscip Rev RNA.1(3):402-14.2010)中,其中的每一篇均通过引用并入本文。

[0331] 在某些实施方案中,vtRNA源自γ-疱疹病毒(GHV68)。

[0332] 在某些实施方案中,vtRNA由包含以下中的任一种的核酸编码:

[0333] GCCAGAGTAGCTCAATTGGTAGAGCAACAGGTCACCGATCCTGGTGGTTCTCGGTTCAAGTCCGAGCTCTGGTC(vtRNA-1;SEQ ID NO:X),或与其至少90%(例如,90%、91%、92%、93%、94%、95%、96%、97%、98%或99%)同一的序列;

[0334] GCCAGGGTAGCTCAATCGGTAGAGCAGCGGTTCCTGGAGTCCGCTGGTTCTCGGTTCAAGCCCGAGCCCTGGTTG(vtRNA-2;SEQ ID NO:X),或与其至少90%(例如,90%、91%、92%、93%、94%、95%、96%、97%、98%或99%)同一的序列;

[0335] GTCGGGGTAGCTCAAATGGTAGAGTGGCAGGCCAACATAGCCAGCAGATCTCGGTTCAAACCCGAGCCCTGACCA(vtRNA-3;SEQ ID NO:X),或与其至少90%(例如,90%、91%、92%、93%、94%、95%、96%、97%、98%或99%)同一的序列;

[0336] GTCGGGGTAGCTCAATTGGTAGAGCGGCAGGCTCATCCCCTGCAGGTTCTCGGTTCAATCCCGGGTCCCGACGC(vtRNA-4;SEQ ID NO:X),或与其至少90%(例如,90%、91%、92%、93%、94%、95%、96%、97%、98%或99%)同一的序列;

[0337] GCCAGGGTAGCTCAATTGGTAGAGCATCAGGCTAGTATCCTGTCGGTTCCGGTTCAAGTCCGGGCCCTGGTTA(vtRNA-5;SEQ ID NO:X),或与其至少90%(例如,90%、91%、92%、93%、94%、95%、96%、97%、98%或99%)同一的序列;

[0338] GCCAGCGTAGCTCAATTGTTAGAGCAGCGGCCACCAAGCCTGCAGGTTCTCGGTTCAAGTCCGGGCGCTGGCAT(vtRNA-6;SEQ ID NO:X),或与其至少90%(例如,90%、91%、92%、93%、94%、95%、96%、97%、98%或99%)同一的序列;

[0339] GCCAGGGTAGCTCAATTGGTAGAGCGGCAGACACCACCTACGTGGTCTAGTCTGTGGATCTCGGTTCAAGTCCGAGTCCTGGCCA(vtRNA-7;SEQ ID NO:X),或与其至少90%(例如,90%、91%、92%、93%、94%、95%、96%、97%、98%或99%)同一的序列;

[0340] GCCAGGGTAGCTCAATTGGTAGAGCGGCAGACACCACCTACGTGGTCTAGTCTGTGGATCTCGGTTCAAGTCCGAGTCCTGGCCA(vtRNA-7;SEQ ID NO:X),或与其至少90%(例如,90%、91%、92%、93%、94%、95%、96%、97%、98%或99%)同一的序列;或

[0341] ACCAGAGTGGCTCACCTGGTAGAGCACCAGGCTGCCCATCCTGTTGGTTCTCGGTTCAAATCCGAGCTCTGGTGA(vtRNA-8;SEQ ID NO:X),或与其至少90%(例如,90%、91%、92%、93%、94%、95%、96%、97%、98%或99%)同一的序列。

[0342] 在某些实施方案中,负链RNA病毒基因组是重组弹状病毒基因组。

[0343] 在某些实施方案中,负链RNA病毒基因组是重组狂犬病毒属基因组。在某些实施方案中,重组狂犬病毒属基因组是重组狂犬病病毒基因组。

[0344] D.治疗性转基因

[0345] 在某些实施方案中,本公开的重组狂犬病病毒基因组编码包含治疗性转基因的核酸。如本文所用,术语“治疗性”是指治疗和/或预防。如本文所用,术语“治疗性转基因”是指编码能够实现对有需要的受试者的治疗和/或预防的转基因产物的转基因。在某些实施方案中,治疗效果是通过阻抑、缓解或根除受试者所遭受的疾病状态来实现的。该治疗性转基因可以编码能够在有需要的受试者中实现治疗和/或预防、从而导致受试者中的疾病状态的阻抑、缓解或根除的任何治疗剂。在某些实施方案中,治疗性转基因编码转基因产物的前体,该前体一旦被加工,例如在细胞中被加工,就能够实现对有需要的受试者的治疗和/或预防。

[0346] 在某些实施方案中,编码治疗性转基因的核酸大于:约300bp、约400bp、约500bp、约600bp、约700bp、约800bp、约900bp、约1,000bp、约1,100bp、约1,200bp、约1,300bp、约1,400bp、约1,500bp、约1,600bp、约1,700bp、约1,800bp、约1,900bp、约2,000bp、约2,100bp、约2,200bp、约2,300bp、约2,400bp、约2,500bp、约2,600bp、约2,700bp、约2,800bp、约2,900bp,或约3,000bp。

[0347] 在某些实施方案中,编码治疗性转基因的核酸大于约300bp(例如,该治疗性转基因为约350bp、约400bp、约450bp、约500bp、约550bp、约600bp,或约650bp)。在某些实施方案中,编码治疗性转基因的核酸大于约650bp(例如,该治疗性转基因为约700bp、约750bp、约800bp、约850bp、约900bp、约950bp或约1,000bp)。在某些实施方案中,编码治疗性转基因的核酸大于约1,000bp(例如,该治疗性转基因为约1,500bp、约2,000bp、约2,500bp或约3,000bp)。在某些实施方案中,编码治疗性转基因的核酸大于约3,000bp(例如,该治疗性转基因为约3,500bp、约4,000bp或约4,500bp)。

[0348] 在某些实施方案中,编码治疗性转基因的核酸大于约4,500bp(例如,该治疗性转基因为约5,000bp、约5,500bp、约6,000bp、约6,500bp、约7,000bp、约7,500bp、约8,000bp或约8,500bp)。

[0349] 在某些实施方案中,编码治疗性转基因的核酸大于约8,500bp(例如,该治疗性转基因为约9,000bp、约9,500bp或约10,000bp)。

[0350] 在某些实施方案中,编码治疗性转基因的核酸大于约10,000bp(例如,该治疗性转基因为约10,500bp、约11,000bp、约11,500bp、约12,000bp、约12,500bp、约13,000bp、约13,500bp、约14,000bp、约14,500bp,或约15,000bp)。

[0351] 在某些实施方案中,编码治疗性转基因的核酸是在约4,000bp和约6,000bp之间(例如,该治疗性转基因是约4,000bp、约4,500bp、约5,000bp、约5,500bp或约6,000bp)。

[0352] 在某些实施方案中,治疗性转基因编码治疗性核酸。该治疗性转基因可编码本领域已知的任何治疗性核酸,例如但不限于任何反义RNA(单链RNA)、任何小干扰RNA(双链RNA)、任何RNA适体和/或任何信使RNA(mRNA)。例如,该治疗性转基因可以编码但不限于miRNA、miRNA模拟物、siRNA、shRNA、gRNA、长非编码RNA、增强子RNA、RNA适体酶、RNA适体、antagomiR和/或合成的RNA。在某些实施方案中,治疗性核酸可以是RNA结合位点,例如miRNA结合位点。各种其他类型的治疗性核酸是本领域普通技术人员已知的。

[0353] 在某些实施方案中,治疗性转基因编码治疗性多肽。该治疗性转基因可以编码本领域已知的任何治疗性多肽,例如但不限于可替换有缺陷的或异常的蛋白质的治疗性多

肽;可增强现有途径的治疗性多肽;可提供新功能或活性的治疗性多肽(例如,可提供对缺乏其的受试者有益的新功能或活性的治疗性多肽);干扰分子或生物体(例如,与宿留(host)靶细胞的生物体不同的生物体)的治疗性多肽;和/或递送其他化合物或蛋白质(例如放射性核素、细胞毒性药物和/或效应蛋白)的治疗性多肽。例如,该治疗性转基因可以编码但不限于核酸修饰蛋白(例如腺嘌呤或胞苷碱基编辑器)或系统、抗体或基于抗体的药物、抗凝剂、血液因子、骨形态发生蛋白、经工程化的蛋白支架、酶、Fc融合蛋白、生长因子、激素、干扰素、白细胞介素和/或溶栓剂。各种其他类型的治疗性多肽是本领域普通技术人员已知的。

[0354] 在某些实施方案中,治疗性转基因编码核酸修饰蛋白。在一些实施方案中,治疗性转基因编码包含核酸结合蛋白(例如,锌指、TALE或核酸可编程核酸结合蛋白,诸如Cas9)的蛋白。在一些实施方案中,核酸编辑系统组分是引导RNA(gRNA)。

[0355] 在一些实施方案中,治疗性转基因编码CRISPR系统。在一些实施方案中,CRISPR系统包含核碱基编辑器,该核碱基编辑器包含多核苷酸可编程核苷酸结合结构域和核碱基编辑结构域。在一些实施方案中,核碱基编辑结构域是腺苷脱氨酶、胞苷脱氨酶、胞嘧啶脱氨酶或其功能变体(例如,能够使诸如DNA或RNA的核酸分子中的核碱基脱氨的功能变体)。在一些实施方案中,核碱基编辑结构域是腺苷脱氨酶。在一些实施方案中,腺苷脱氨酶是ABE7.10。在一些实施方案中,多核苷酸可编程核苷酸结合结构域是Cas9多肽、Cas12多肽或其功能变体。在一些实施方案中,CRISPR系统进一步包含引导RNA(gRNA)或编码gRNA的核酸。

[0356] 在一些实施方案中,治疗性转基因编码核碱基修饰蛋白(例如,碱基编辑器蛋白)。在一些实施方案中,治疗性转基因编码腺苷碱基编辑器(例如ABE7.10)。在一些实施方案中,治疗性转基因编码胞苷碱基编辑器。在一些实施方案中,治疗性转基因编码能够使DNA或RNA中的胞嘧啶脱氨的胞嘧啶碱基编辑器。

[0357] 在某些实施方案中,治疗性转基因编码基因编辑系统,例如本文进一步描述的碱基编辑器系统。

[0358] 对于本领域普通技术人员来说十分显而易见的是,本文所述的本公开的重组狂犬病病毒基因组编码包含治疗性转基因的核酸,其中该治疗性转基因编码治疗性多肽和/或治疗性核酸,例如,在某些实施方案中,治疗性转基因编码治疗性多肽和治疗性核酸的组合。在某些实施方案中,治疗性转基因编码一种或多种治疗性多肽。在某些实施方案中,治疗性转基因编码一种或多种治疗性核酸。在某些实施方案中,治疗性转基因编码一种或多种治疗性多肽和一种或多种治疗性核酸的组合。将治疗性多肽和治疗性核酸的组合递送至靶细胞中可用于本领域普通技术人员已知的各种目的。在某些实施方案中,可以将治疗性多肽递送至靶细胞,其中该递送脱靶(detargeted)至某些其他细胞类型。例如,治疗性转基因可以编码治疗性多肽和/或治疗性核酸,并且还包含miRNA结合位点。该miRNA结合位点可对细胞类型脱靶起作用。例如,可针对肝细胞脱靶采用仅在肝脏中表达的miRNA122a。参见例如Dhungel等人,Molecules(2018)23(7):1500。

[0359] 在某些实施方案中,治疗性转基因进一步编码一种或多种报告序列(reporter sequence)。报告序列在靶细胞中表达时产生直接或间接可检测的信号。合适的报告序列的实例包括但不限于编码荧光蛋白(例如,GFP、RFP、YFP)、碱性磷酸酶、胸苷激酶、氯霉素乙酰

转移酶(CAT)、荧光素酶、β-半乳糖苷酶(LacZ)和β-内酰胺酶的序列。编码细胞表面膜结合蛋白的序列也可适合作为报告序列,例如高亲和力抗体所结合至的膜结合蛋白,例如流感血凝素蛋白(HA)、CD2、CD4、CD8和本领域普通技术人员已知的其他蛋白,包括例如用抗原结构域(例如HA标签、FLAG标签、Myc标签、多组氨酸标签)标识的膜结合蛋白。

[0360] 在某些实施方案中,治疗性转基因编码治疗性多肽和/或治疗性核酸,其中该治疗性多肽和/或该治疗性核酸是经分泌的。例如,本文所述的本公开的重组狂犬病病毒基因组可被引入到靶细胞中,其中该重组狂犬病病毒基因组编码包含治疗性转基因的核酸,并且其中该治疗性转基因编码分泌的治疗性多肽和/或治疗性核酸(例如,可分泌的治疗性转基因和/或可分泌的治疗性核酸)。该治疗性多肽和/或核酸在表达后可以分泌到靶细胞外。在某些实施方案中,治疗性多肽和/或核酸在表达后借助于驻存于该治疗性多肽和/或核酸上的内源元件(例如,指导细胞外分泌的内源信号肽)来分泌。在某些实施方案中,治疗性多肽和/或核酸在表达后借助于驻存于该治疗性多肽和/或核酸上的外源元件(例如,指导细胞外分泌的外源信号肽)来分泌。可分泌的治疗性多肽和/或核酸的递送可用于治疗某些疾病。例如,由溶酶体代谢功能障碍引起的溶酶体贮积症(LSD)包含独特的交叉校正(cross-correction)特征,该特征使得特定的细胞外LSD酶能够被吸收并且被靶向于酶有缺陷的或酶异常的细胞的溶酶体。某些酶的交叉校正特征形成了被称为酶替换疗法的已批准疗法的基础。参见例如Rastall和Amalfitano,Appl.Clin.Genet.(2015)8:157-169。

[0361] 在某些实施方案中,本公开的重组狂犬病病毒基因组包含可操作地连接至编码转基因的核酸的转录调控元件。该转录调控元件能够控制与其可操作地连接的转基因的表达(例如,所编码的治疗性多肽和/或核酸的表达)。在某些实施方案中,转录调控元件包含转录起始信号。该转录起始信号对于狂犬病病毒来说可以是内源的或外源的。在某些实施方案中,转录起始信号是合成的转录起始信号。在某些实施方案中,编码转基因的核酸进一步可操作地连接至转录终止多聚腺苷酸化信号。该转录终止多聚腺苷酸化信号对于狂犬病病毒来说可以是内源的或外源的。在某些实施方案中,转录终止多聚腺苷酸化信号是合成的转录终止多聚腺苷酸化信号。合适的转录起始信号和转录终止多聚腺苷酸化信号的实例是本领域普通技术人员已知的,并且描述于例如Albertini等人,Adv.Virus.Res.(2011)79:1-22;Ogino和Green,Viruses(2019)11(6):504;以及Ogino和Green,Front.Microbiol.(2019)10:1490,这些文献的公开内容以引用的方式整体并入本文。

[0362] 包含含有治疗性转基因的核酸的本公开的重组狂犬病病毒基因组可进一步包含本领域普通技术人员已知的有助于和/或增强治疗性转基因的表达的任何元件。

[0363] 通过本文描述的方法将本公开的重组狂犬病病毒基因组掺入到重组狂犬病病毒颗粒中。在某些实施方案中,本公开的重组狂犬病病毒颗粒包含狂犬病病毒糖蛋白和包含含有如本文所述的治疗性转基因的核酸的重组狂犬病病毒基因组。在某些实施方案中,重组狂犬病病毒颗粒包含:狂犬病病毒糖蛋白和包含含有治疗性转基因的核酸的重组狂犬病病毒基因组,其中该基因组缺乏编码狂犬病病毒糖蛋白的内源G基因。在某些实施方案中,重组狂犬病病毒颗粒包含:狂犬病病毒糖蛋白和包含含有治疗性转基因的核酸的重组狂犬病病毒基因组,其中该基因组缺乏编码狂犬病病毒糖蛋白的内源G基因;并且其中该基因组缺乏编码狂犬病病毒聚合酶的内源L基因。

[0364] 重组负链病毒基因组(例如,狂犬病病毒基因组)及在重组负链病毒基因组(例如,

狂犬病病毒基因组)中编码的治疗性转基因进一步详细描述于2022年2月18日提交的PCT/US2022/017075中,其全部内容通过引用并入本文。

[0365] E.核碱基编辑器

[0366] 在某些示例性实施方案中,可用于本文所述的方法和组合物的治疗性转基因是编辑、修饰或改变多核苷酸的靶核苷酸序列的核碱基编辑器。本文描述的核碱基编辑器通常包括多核苷酸可编程核苷酸结合结构域和核碱基编辑结构域(例如,腺苷脱氨酶或胞苷脱氨酶)。多核苷酸可编程核苷酸结合结构域在与被结合的引导多核苷酸(例如,gRNA)联合时可以特异性地结合至靶多核苷酸序列,并且从而将碱基编辑器定位于需要被编辑的靶核酸序列。

[0367] 多核苷酸可编程核苷酸结合结构域

[0368] 多核苷酸可编程核苷酸结合结构域结合多核苷酸(例如RNA、DNA)。碱基编辑器的多核苷酸可编程核苷酸结合结构域本身可以包含一个或多个结构域(例如,一个或多个核酸酶结构域)。在一些实施方案中,多核苷酸可编程核苷酸结合结构域的核酸酶结构域可包含核酸内切酶或核酸外切酶。核酸内切酶可以切割双链核酸的单链或双链核酸分子的两条链。在一些实施方案中,多核苷酸可编程核苷酸结合结构域的核酸酶结构域可以切割靶多核苷酸的零、一条或两条链。

[0369] 可被掺入到碱基编辑器中的多核苷酸可编程核苷酸结合结构域的非限制性实例包括CRISPR蛋白衍生结构域、限制性核酸酶、大范围核酸酶、TAL核酸酶(TALEN)和锌指核酸酶(ZFN)。在一些实施方案中,碱基编辑器包含多核苷酸可编程核苷酸结合结构域,该多核苷酸可编程核苷酸结合结构域包含经由被结合的引导核酸能够在CRISPR(即,成簇规则间隔短回文重复序列)介导的核酸修饰期间结合至核酸序列的天然或经修饰的蛋白或其部分。此种蛋白在本文中被称为“CRISPR蛋白”。因此,本文公开了碱基编辑器,该碱基编辑器包含含有CRISPR蛋白的全部或一部分的多核苷酸可编程核苷酸结合结构域(即,包含CRISPR蛋白的全部或一部分作为结构域的碱基编辑器,也称为碱基编辑器的“CRISPR蛋白衍生结构域”)。掺入到碱基编辑器中的CRISPR蛋白衍生结构域与CRISPR蛋白的野生型或天然型式相比可以是经修饰的。例如,如下所述,CRISPR蛋白衍生结构域可包含相对于CRISPR蛋白的野生型或天然型式的一个或多个突变、插入、缺失、重排和/或重组。

[0370] 可用于本文的Cas蛋白包括第1类和第2类。Cas蛋白的非限制性实例包括Cas1、Cas1B、Cas2、Cas3、Cas4、Cas5、Cas5d、Cas5t、Cas5h、Cas5a、Cas6、Cas7、Cas8、Cas9(也称为Csn1或Csx12)、Cas10、Csy1、Csy2、Csy3、Csy4、Cse1、Cse2、Cse3、Cse4、Cse5e、Csc1、Csc2、Csa5、Csn1、Csn2、Csm1、Csm2、Csm3、Csm4、Csm5、Csm6、Cmr1、Cmr3、Cmr4、Cmr5、Cmr6、Csb1、Csb2、Csb3、Csx17、Csx14、Csx10、Csx16、CsaX、Csx3、Csx1、Csx1S、Csf1、Csf2、Cs0、Csf4、Csd1、Csd2、Cst1、Cst2、C sh1、Csh2、Csa1、Csa2、Csa3、Csa4、Csa5、Cas12a/Cpf1、Cas12b/C2c1(例如,SEQ ID NO:258)、Cas12c/C2c3、Cas12d/CasY、Cas12e/CasX、Cas12g、Cas12h、Cas12i和Cas12j/CasΦ、CARF、DinG、它们的同源物,或它们的经修饰型式。CRISPR酶可以指导切割靶序列处(例如在靶序列内和/或在靶序列的互补序列内)的一条链或两条链。例如,CRISPR酶可以指导切割距靶序列的第一个或最后一个核苷酸约1、2、3、4、5、6、7、8、9、10、15、20、25、50、100、200、500或更多个碱基对内的一条或两条链。

[0371] 可以使用编码CRISPR酶的载体,该CRISPR酶相对于相应的野生型酶被突变成使得

突变的CRISPR酶缺乏切割含有靶序列的靶多核苷酸的一条或两条链的能力。Cas蛋白(例如,Cas9、Cas12)或Cas结构域(例如,Cas9、Cas12)可以指与野生型示例性Cas多肽或Cas结构域具有至少或至少约50％、60％、70％、80％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％序列同一性和/或序列同源性的多肽或结构域。Cas(例如,Cas9、Cas12)可以指Cas蛋白的野生型或经修饰形式,其可以包含氨基酸变化,诸如缺失、插入、取代、变体、突变、融合、嵌合体,或它们的任何组合。

[0372] 在一些实施方案中,碱基编辑器的CRISPR蛋白衍生结构域可以包括来自溃疡棒状杆菌(NCBI Refs:NC_015683.1,NC_017317.1);白喉棒杆菌(Corynebacterium diphtheria)(NCBI Ref:NC_016782.1,NC_016786.1);梅毒螺旋体(Spiroplasma syrphidicola)(NCBI Ref:NC_021284.1);中间普雷沃菌(Prevotella intermedia)(NCBI Ref:NC_017861.1);台湾螺旋体(Spiroplasma taiwanense)(NCBI Ref:NC_021846.1);海豚链球菌(Streptococcus iniae)(NCBI Ref:NC_021314.1);波罗的海贝尔氏菌(Belliella baltica)(NCBI Ref:NC_018010.1);扭曲冷弯曲菌(Psychroflexus torquis)(NCBI Ref:NC_018721.1);嗜热链球菌(Streptococcus thermophilus)(NCBI Ref:YP_820832.1);无害李斯特菌(Listeria innocua)(NCBI Ref:NP_472073.1);贾氏弯曲杆菌(Campylobacter jejuni)(NCBI Ref:YP_002344900.1);脑膜炎双球菌(Neisseria meningitidis)(NCBI Ref:YP_002342100.1),酿脓链球菌(Streptococcus pyogenes),或金黄色葡萄球菌(Staphylococcus aureus)的Cas9的全部或一部分。

[0373] Cas9核酸酶序列和结构是本领域技术人员众所周知的(参见,如“Complete genome sequence of an M1 strain of Streptococcus pyogenes.”Ferretti等人,Proc.Natl.Acad.Sci.U.S.A.98:4658-4663(2001);“CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.”Deltcheva E.等人,Nature 471:602-607(2011);以及“A programmabledual-RNA-guided DNA endonuclease in adaptive bacterial immunity,”Jinek M.等人,Science 337:816-821(2012),这些中的每一者的全部内容以引用的方式并入本文)。Cas9直向同源物已在各种物种中得到描述,这些物种包括但不限于酿脓链球菌和嗜热链球菌。另外的合适的Cas9核酸酶和序列基于本公开对于本领域技术人员来说将是显而易见的,并且此类Cas9核酸酶和序列包括来自Chylinski,Rhun,and Charpentier,“The tracrRNA and Cas9 families of type IICRISPR-Cas immunity systems”(2013)RNA Biology 10:5,726-737中公开的生物体和基因座的Cas9序列;该文献的全部内容以引用的方式并入本文。

[0374] 高保真(Fidelity)Cas9结构域

[0375] 本公开的一些方面提供了高保真度Cas9结构域。高保真Cas9结构域在本领域中是已知的并且描述于例如,Kleinstiver,B.P.等人“High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects.”Nature 529,490-495(2016);以及Slaymaker,I.M.等人“Rationally engineered Cas9 nucleases with improved specificity.”Science 351,84-88(2015)中,这些文献中的每一篇的全部内容以引用的方式并入本文。示例性高保真Cas9结构域在序列表中作为SEQ ID NO:1423提供。在一些实施方案中,高保真Cas9结构域是经工程化的Cas9结构域,该结构域包含相对于相应的野生型Cas9结构域而言减少Cas9结构域和DNA的糖-磷酸酯骨架之间的静电相互作用的一个或多

个突变。与DNA的糖-磷酸酯骨架的静电相互作用减少的高保真Cas9结构域具有较少的脱靶效应。在一些实施方案中,Cas9结构域(例如,野生型Cas9结构域(SEQ ID NO:223和233))包含减少Cas9结构域和DNA的糖-磷酸酯骨架之间的关联的一个或多个突变。在一些实施方案中,Cas9结构域包含一个或多个突变,该突变将Cas9结构域和DNA的糖-磷酸酯骨架之间的关联减少至少1%、至少2%、至少3%、至少4%、至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%或至少70%。

[0376] 在一些实施方案中,本文提供的Cas9融合蛋白中的任何Cas9融合蛋白包含D10A、N497X、R661X、Q695X和/或Q926X突变,或本文提供的氨基酸序列中的任何氨基酸序列中的相应突变中的一者或多者,其中X是任何氨基酸。在一些实施方案中,高保真Cas9酶是SpCas9(K855A)、eSpCas9(1.1)、SpCas9-HF1或超精确Cas9变体(HypaCas9)。在一些实施方案中,经修饰的Cas9 eSpCas9(1.1)含有丙氨酸取代,该丙氨酸取代削弱了HNH/RuvC槽和非靶DNA链之间的相互作用,从而防止链分离并且在脱靶位点处切割。类似地,SpCas9-HF1通过破坏Cas9和DNA磷酸酯骨架相互作用的丙氨酸取代来减少脱靶编辑。HypaCas9在REC3结构域中含有突变(SpCas9 N692A/M694A/Q695A/H698A),该突变增加Cas9校对和靶识别。与野生型Cas9相比,所有三种高保真酶产生更少的脱靶编辑。

[0377] 排他性减小的Cas9结构域

[0378] 通常,Cas9蛋白(诸如来自酿脓链球菌的Cas9(spCas9))需要“原型间隔区相邻基序(PAM)”或PAM样基序,其是紧接在被CRISPR细菌适应性免疫系统中的Cas9核酸酶靶向的DNA序列之后的2-6个碱基对的DNA序列。NGG PAM序列的存在是结合特定核酸区域所必需的,其中“NGG”中的“N”是腺苷(A)、胸苷(T)或胞嘧啶(C),并且G是鸟苷。这可能会限制在基因组内编辑所需碱基的能力。在一些实施方案中,本文提供的碱基编辑融合蛋白可能需要放置在精确区位(location),例如包含位于PAM上游的靶碱基的区域处。参见例如Komor,A.C.等人,“Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage”Nature 533,420-424(2016),其全部内容以引用方式并入本文。能够结合PAM序列的spCas9蛋白的示例性多肽序列在序列表中作为SEQ ID NO:223、234,和1304-1307提供。因此,在一些实施方案中,本文提供的融合蛋白中的任何融合蛋白可以含有Cas9结构域,该结构域能够结合不含规范(例如NGG)PAM序列的核苷酸序列。结合至非规范PAM序列的Cas9结构域已在本领域中进行了描述,并且对于本领域技术人员来说将是显而易见的。例如,结合非规范PAM序列的Cas9结构域已描述于Kleinstiver,B.P.等人,“Engineered CRISPR-Cas9 nucleases with altered PAM specificities”Nature 523,481-485(2015);以及Kleinstiver,B.P.等人,“Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition”Nature Biotechnology 33,1293-1298(2015);每一篇的全部内容均以引用方式并入本文。

[0379] 切口酶

[0380] 在一些实施方案中,多核苷酸可编程核苷酸结合结构域可以包含切口酶结构域。在本文中,术语“切口酶”是指包含核酸酶结构域的多核苷酸可编程核苷酸结合结构域,该核酸酶结构域能够仅切割双链核酸分子(例如,DNA)中两条链中的一条链。在一些实施方案中,切口酶可以通过将一个或多个突变引入到活性多核苷酸可编程核苷酸结合结构域中而

衍生自多核苷酸可编程核苷酸结合结构域的完全催化活性(例如,天然)形式。例如,当多核苷酸可编程核苷酸结合结构域包含衍生自Cas9的切口酶结构域时,Cas9衍生的切口酶结构域在位置840处可以包含D10A突变和组氨酸。在此类实施方案中,残基H840保留催化活性并因此可以切割核酸双链体的单链。在另一个实例中,Cas9衍生的切口酶结构域可以包含H840A突变,而位置10处的氨基酸残基仍然是D。在一些实施方案中,切口酶可以通过去除切口酶活性不需要的核酸酶结构域的全部或一部分而衍生自多核苷酸可编程核苷酸结合结构域的完全催化活性(例如,天然)形式。例如,在多核苷酸可编程核苷酸结合结构域包含衍生自Cas9的切口酶结构域的情况下,Cas9衍生的切口酶结构域可以包含RuvC结构域或HNH结构域的全部或一部分的缺失。

[0381] 在一些实施方案中,野生型Cas9对应于或包含以下氨基酸序列:

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRH
SIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS
NEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY
PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN
SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLEN
LIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDT

[0382] YDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAP
LSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAG
YIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDN
GSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPL
ARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFD
KNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGE
QKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFN

ASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEER
LKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTI
LDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHI
ANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQT
TQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLY
YLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLT
RSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT
KAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEN
[0383] DKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNA
VVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAK
YFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFA
TVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDW
DPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMER
SSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA
GELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ
HKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAE
NIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITG
[0384] LYETRIDLSQLGGD (SEQ ID NO:223) (单下划线:HNH结构域;双下划线:RuvC结构域)。

[0385] 在一些实施方案中,被包含切口酶结构域(例如,Cas9衍生的切口酶结构域、Cas12衍生的切口酶结构域)的碱基编辑器切割的核酸双链体靶多核苷酸序列的链是未被碱基编辑器编辑的链(即,被碱基编辑器切割的链与包含待编辑的碱基的链相反)。在其他实施方案中,包含切口酶结构域(例如,Cas9衍生的切口酶结构域、Cas12衍生的切口酶结构域)的碱基编辑器可以切割被靶向用于编辑的DNA分子的链。在此类实施方案中,非被靶向的链不被切割。

[0386] 在一些实施方案中,Cas9核酸酶具有无活性(例如,经灭活的)DNA切割结构域,即,Cas9是切口酶,其被称为“nCas9”蛋白(对于“切口酶”Cas9)。Cas9切口酶可以是能够仅切割双链核酸分子(例如,双链DNA分子)的一条链的Cas9蛋白。在一些实施方案中,Cas9切口酶切割双链核酸分子的靶链,意味着Cas9切口酶切割与结合至Cas9的gRNA(例如,sgRNA)碱基配对(互补)的链。在一些实施方案中,Cas9切口酶包含D10A突变并且在位置840处具有组氨酸。在一些实施方案中,Cas9切口酶切割双链核酸分子的非靶、非碱基编辑的链,意味着Cas9切口酶切割与结合至Cas9的gRNA(例如,sgRNA)非碱基配对的链。在一些实施方案中,Cas9切口酶包含H840A突变并且在位置10处具有天冬氨酸残基,或相应的突变。在一些实施方案中,Cas9切口酶包含与本文提供的Cas9切口酶中的任何一种Cas9切口酶至少60%、至

少65％、至少70％、至少75％、至少80％、至少85％、至少90％、至少95％、至少96％、至少97％、至少98％、至少99％或至少99.5％同一的氨基酸序列。另外的合适的Cas9切口酶基于本公开和本领域知识对于本领域技术人员将是显而易见的,并且在本公开的范围内。

[0387] 示例性催化Cas9切口酶(nCas9)的氨基酸序列如下:

[0388] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEER

[0389] LKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTI

[0390] LDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHI

[0391] ANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQT

[0392] TQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLY

[0393] YLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLT

[0394] RSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT

[0395] KAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEN

[0396] DKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNA

[0397] VVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAK

[0398] YFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFA

[0399] TVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDW

[0400] DPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMER

[0401] SSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA

[0402] GELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ

[0403] HKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAE

[0404] NIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITG

[0405] LYETRIDLSQLGGD(SEQ ID NO:234)

[0406] Cas9核酸酶具有两个功能性核酸内切酶结构域:RuvC和HNH。Cas9在将核酸酶结构域定位成切割靶DNA的相反链的靶标结合后经历构象变化。Cas9介导的DNA切割的最终结果是靶DNA(PAM序列上游约3-4个核苷酸)内的双链断裂(DSB)。然后通过以下两种一般修复途径之一修复所得DSB:(1)有效率但容易出错的非同源末端联接(NHEJ)途径;或(2)效率较低但高保真的同源性指导的修复(HDR)途径。

[0407] 非同源末端联接(NHEJ)和/或同源性指导的修复(HDR)的“效率”可以通过任何方便的方法计算。例如,在一些实施方案中,效率可以用成功的HDR的百分比来表示。例如,surveyor核酸酶测定可以用于产生切割产物,并且产物与底物的比率可以用于计算该百分比。例如,作为成功的HDR的结果,可以使用直接切割含有新整合的限制性序列的DNA的

surveyor核酸酶。更多切割的底物指示更高的HDR百分比(更高的HDR效率)。作为说明性实施例,可以使用以下方程[(切割产物)/(底物加切割产物)](例如,(b+c)/(a+b+c),其中“a”是DNA底物的谱带强度,并且“b”和“c”是切割产物)计算HDR的分数(百分比)。

[0408] 在一些实施方案中,效率可以以成功的NHEJ的百分比来表示。例如,T7核酸内切酶I测定可以用于产生切割产物,并且产物与底物的比率可以用于计算NHEJ百分比。T7核酸内切酶I切割由野生型和突变DNA链杂交产生的错配异源双链DNA(NHEJ在原始断裂的位点处产生小的随机插入或缺失(插入/缺失))。更多的切割指示更高的NHEJ百分比(更高的NHEJ效率)。作为说明性实施例,可以使用以下方程:$(1-(1-(b+c)/(a+b+c))^{1/2})\times 100$计算NHEJ的分数(百分比),其中“a”是DNA底物的谱带强度,并且“b”和“c”是切割产物(Ran等人,Cell.2013年9月12日;154(6):1380-9;以及Ran等人,Nat Protoc.2013年11月;8(11):2281-2308)。

[0409] NHEJ修复途径是活性最强的修复机制,并且它经常在DSB位点处导致小的核苷酸插入或缺失(插入/缺失)。NHEJ介导的DSB修复的随机性具有重要的实际意义,因为表达Cas9和gRNA或引导多核苷酸的细胞群体可以导致各种各样的突变。在大多数实施方案中,NHEJ在靶DNA中产生小的插入/缺失,该插入/缺失导致氨基酸缺失、插入或移码突变,从而导致被靶向基因的开放阅读框(ORF)内的提前终止密码子。理想的最终结果是被靶向基因内的功能丧失突变。

[0410] 虽然NHEJ介导的DSB修复常常会破坏基因的开放阅读框,但同源性指导的修复(HDR)可以用于产生特定的核苷酸变化,其范围为从单个核苷酸变化到诸如荧光团或标签的添加的大的插入。

[0411] 为了利用HDR进行基因编辑,可以使用一种或多种gRNA和Cas9或Cas9切口酶将含有所需序列的DNA修复模板递送到关注的细胞类型中。该修复模板可以含有所需的编辑以及紧邻靶标上游和下游的另外的同源序列(称为左同源臂和右同源臂)。每个同源臂的长度可以取决于引入的变化的大小,其中较大的插入需要更长的同源臂。修复模板可以是单链寡核苷酸、双链寡核苷酸或双链DNA质粒。HDR的效率一般较低(<经修饰的等位基因的10%),即使在表达Cas9、gRNA和外源修复模板的细胞中也是如此。HDR的效率可以通过同步细胞来增强,因为HDR发生在细胞周期的S期和G2期。参与NHEJ的化学或遗传抑制基因也可以增加HDR频率。

[0412] 在一些实施方案中,Cas9是经修饰的Cas9。给定的gRNA靶向序列可以在存在部分同源性的整个基因组中具有另外的位点。这些位点称为脱靶位点,并且在设计gRNA时需要考虑。除了优化gRNA设计之外,还可以通过对Cas9的修饰来增加CRISPR的特异性。Cas9通过两个核酸酶结构域RuvC和HNH的组合活性产生双链断裂(DSB)。Cas9切口酶(SpCas9的D10A突变体)保留一个核酸酶结构域并且产生DNA切口而不是DSB。切口酶系统还可以与HDR介导的基因编辑相组合,以进行特定的基因编辑。

[0413] 催化失效(Catalytically Dead)的核酸酶

[0414] 本文还提供了碱基编辑器,其包含催化失效(即,不能切割靶多核苷酸序列)的多核苷酸可编程核苷酸结合结构域。在本文中,术语“催化失效”和“核酸酶失效”可以互换地用来指具有导致其不能切割核酸链的一个或多个突变和/或缺失的多核苷酸可编程核苷酸结合结构域。在一些实施方案中,催化失效的多核苷酸可编程核苷酸结合结构域碱基编辑

器可能由于一个或多个核酸酶结构域中的特定点突变而缺乏核酸酶活性。例如,在包含Cas9结构域的碱基编辑器的情况下,Cas9可以包含D10A突变和H840A突变。此类突变使这两个核酸酶结构域灭活,从而导致核酸酶活性丧失。在其他实施方案中,催化失效的多核苷酸可编程核苷酸结合结构域可以包含催化结构域(例如,RuvC1和/或HNH结构域)的全部或一部分的一个或多个缺失。在进一步的实施方案中,催化失效的多核苷酸可编程核苷酸结合结构域包含点突变(例如,D10A或H840A)以及核酸酶结构域的全部或一部分的缺失。dCas9结构域在本领域中是已知的并且描述于例如Qi等人,"Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression."Cell.2013;152(5):1173-83中,该文献的全部内容以引用的方式并入本文。

[0415] 另外的合适的核酸酶无活性的dCas9结构域基于本公开和本领域知识对于本领域技术人员将是显而易见的,并且在本公开的范围内。此类另外的示例性合适的核酸酶无活性的Cas9结构域包括但不限于D10A/H840A、D10A/D839A/H840A和D10A/D839A/H840A/N863A突变结构域(参见例如,Prashant等人,CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering.Nature Biotechnology.2013;31(9):833-838,该文献的全部内容以引用的方式并入本文)。

[0416] 在一些实施方案中,dCas9对应于或者部分地或整体地包含具有一个或多个使Cas9核酸酶活性灭活的突变的Cas9氨基酸序列。在一些实施方案中,核酸酶无活性的dCas9结构域包含本文阐述的氨基酸序列的D10X突变和H840X突变,或本文提供的氨基酸序列中的任何氨基酸序列中的相应突变,其中X是任何氨基酸变化。在一些实施方案中,核酸酶无活性的dCas9结构域包含本文阐述的氨基酸序列的D10A突变和H840A突变,或本文提供的氨基酸序列中的任何氨基酸序列中的相应突变。在一些实施方案中,核酸酶无活性的Cas9结构域包含克隆载体pPlatTET-gRNA2(登录号BAV54124)中阐述的氨基酸序列。

[0417] 在一些实施方案中,变体Cas9蛋白可以切割引导靶序列的互补链,但切割双链引导靶序列的非互补链的能力减小。例如,变体Cas9蛋白可以具有减小RuvC结构域的功能的突变(氨基酸取代)。作为非限制性实例,在一些实施方案中,变体Cas9蛋白具有D10A(氨基酸位置10处的天冬氨酸变为丙氨酸),并且因此可以切割双链引导靶序列的互补链,但切割双链引导靶序列的非互补链的能力减小(因此在变体Cas9蛋白切割双链靶核酸时导致单链断链(SSB)而不是双链断链(DSB))(参见例如Jinek等人,Science.2012年8月17日;337(6096):816-21)。

[0418] 在一些实施方案中,变体Cas9蛋白可以切割双链引导靶序列的非互补链,但切割引导靶序列的互补链的能力减小。例如,变体Cas9蛋白可以具有减小HNH结构域(RuvC/HNH/RuvC结构域基序)的功能的突变(氨基酸取代)。作为非限制性实例,在一些实施方案中,变体Cas9蛋白具有H840A(氨基酸位置840处的组氨酸变为丙氨酸)突变,并且因此可以切割引导靶序列的非互补链,但切割引导靶序列的互补链的能力减小(因此在变体Cas9蛋白切割双链引导靶序列时导致SSB而不是DSB)。此种Cas9蛋白切割引导靶序列(例如,单链引导靶序列)的能力减小,但保留了结合引导靶序列(例如,单链引导靶序列)的能力。

[0419] 作为另一个非限制性实例,在一些实施方案中,变体Cas9蛋白携有(harbor)W476A和W1126A突变,使得多肽切割靶DNA的能力减小。此种Cas9蛋白切割靶DNA(例如,单链靶

DNA)的能力减小,但保留了结合靶DNA(例如,单链靶DNA)的能力。

[0420] 作为另一个非限制性实例,在一些实施方案中,变体Cas9蛋白携有P475A、W476A、N477A、D1125A、W1126A和D1127A突变,使得多肽切割靶DNA的能力减小。此种Cas9蛋白切割靶DNA(例如,单链靶DNA)的能力减小,但保留了结合靶DNA(例如,单链靶DNA)的能力。

[0421] 作为另一个非限制性实例,在一些实施方案中,变体Cas9蛋白携有H840A、W476A和W1126A突变,使得多肽切割靶DNA的能力减小。此种Cas9蛋白切割靶DNA(例如,单链靶DNA)的能力减小,但保留了结合靶DNA(例如,单链靶DNA)的能力。作为另一个非限制性实例,在一些实施方案中,变体Cas9蛋白携有H840A、D10A、W476A和W1126A突变,使得多肽切割靶DNA的能力减小。此种Cas9蛋白切割靶DNA(例如,单链靶DNA)的能力减小,但保留了结合靶DNA(例如,单链靶DNA)的能力。在一些实施方案中,变体Cas9已恢复Cas9 HNH结构域(A840H)中位置840处的催化His残基。

[0422] 作为另一个非限制性实例,在一些实施方案中,变体Cas9蛋白携有H840A、P475A、W476A、N477A、D1125A、W1126A和D1127A突变,使得多肽切割靶DNA的能力减小。此种Cas9蛋白切割靶DNA(例如,单链靶DNA)的能力减小,但保留了结合靶DNA(例如,单链靶DNA)的能力。作为另一个非限制性实例,在一些实施方案中,变体Cas9蛋白携有D10A、H840A、P475A、W476A、N477A、D1125A、W1126A和D1127A突变,使得多肽切割靶DNA的能力减小。此种Cas9蛋白切割靶DNA(例如,单链靶DNA)的能力减小,但保留了结合靶DNA(例如,单链靶DNA)的能力。在一些实施方案中,当变体Cas9蛋白携有W476A和W1126A突变或当变体Cas9蛋白携有P475A、W476A、N477A、D1125A、W1126A和D1127A突变时,变体Cas9蛋白不能有效率地结合至PAM序列。因此,在一些此类实施方案中,当此种变体Cas9蛋白用于结合的方法时,所述方法不需要PAM序列。换言之,在一些实施方案中,当此种变体Cas9蛋白用于结合的方法时,该方法可以包括引导RNA,但是该方法可以在不存在PAM序列的情况下进行(并且因此由引导RNA的靶向节段提供结合的特异性)。可以使其它残基突变以实现以上作用(即,使一个或另一个核酸酶部分灭活)。作为非限制性实例,残基D10、G12、G17、E762、H840、N854、N863、H982、H983、A984、D986和/或A987可以被改变(即,取代)。同样,除了丙氨酸取代以外的突变也是合适的。

[0423] 在具有减小的催化活性的变体Cas9蛋白(例如,当Cas9蛋白具有D10、G12、G17、E762、H840、N854、N863、H982、H983、A984、D986和/或A987突变,例如D10A、G12A、G17A、E762A、H840A、N854A、N863A、H982A、H983A、A984A和/或D986A时)的一些实施方案中,变体Cas9蛋白仍可以以位点特异性方式结合至靶DNA(因为它仍被引导DNA引导至靶DNA序列),只要该变体Cas9蛋白保留了与引导RNA相互作用的能力即可。

[0424] 在一些实施方案中,变体Cas蛋白可以是spCas9、spCas9-VRQR、spCas9-VRER、xCas9(sp)、saCas9、saCas9-KKH、spCas9-MQKSER、spCas9-LRKIQK,或spCas9-LRVSQL。

[0425] 在一些实施方案中,Cas9结构域是来自金黄色葡萄球菌的Cas9结构域(SaCas9)。在一些实施方案中,SaCas9结构域是核酸酶有活性的SaCas9、核酸酶无活性的SaCas9(SaCas9d)或SaCas9切口酶(SaCas9n)。在一些实施方案中,SaCas9包含N579A突变,或在随同提交的序列表中提供的氨基酸序列中的任何氨基酸序列中的相应突变。

[0426] 在一些实施方案中,SaCas9结构域、SaCas9d结构域或SaCas9n结构域可以结合至具有非规范PAM的核酸序列。在一些实施方案中,SaCas9结构域、SaCas9d结构域或SaCas9n

结构域可以结合至具有NNGRRT或NNGRRV PAM序列的核酸序列。在一些实施方案中,SaCas9结构域包含E781X、N967X和R1014X突变或本文提供的氨基酸序列中的任何氨基酸序列中的相应突变中的一者或多者,其中X是任何氨基酸。在一些实施方案中,SaCas9结构域包含E781K、N967K和R1014H突变,或本文提供的氨基酸序列中的任何氨基酸序列中的一个或多个相应突变中的一者或多者。在一些实施方案中,SaCas9结构域包含E781K、N967K和R1014H突变或本文提供的氨基酸序列中的任何氨基酸序列中的相应突变。

[0427] 在一些实施方案中,融合蛋白中存在的Cas9结构域之一可以被对PAM序列没有要求的引导核苷酸序列可编程DNA结合蛋白结构域替换。在一些实施方案中,Cas9是SaCas9。SaCas9的残基A579可以从N579突变而得以产生SaCas9切口酶。残基K781、K967和H1014可以从E781、N967和R1014突变而得以产生SaKKH Cas9。

[0428] 在一些实施方案中,使用包括氨基酸取代D1135M、S1136Q、G1218K、E1219F、A1322R、D1332A、R1335E和T1337R(SpCas9-MQKFRAER)并且对改变的PAM5'-NGC-3'具有特异性的经修饰的SpCas9。

[0429] 酿脓链球菌Cas9的替代方案可以包括来自在哺乳动物细胞中显示出切割活性的Cpf1家族的RNA引导的核酸内切酶。来自普氏菌属和弗朗西斯氏菌属1的CRISPR(CRISPR/Cpf1)是类似于CRISPR/Cas9系统的DNA编辑技术。Cpf1是II类CRISPR/Cas系统的RNA引导的核酸内切酶。这种获得性免疫机制存在于普氏菌属和弗朗西斯氏菌属的细菌中。Cpf1基因与编码使用引导RNA来寻找和切割病毒DNA的核酸内切酶的CRISPR基因座相关。Cpf1是比Cas9更小、更简单的核酸内切酶,克服了CRISPR/Cas9系统限制中的一些CRISPR/Cas9系统限制。与Cas9核酸酶不同,Cpf1介导的DNA切割的结果是具有短3'悬突的双链断裂。Cpf1的交错切割模式可以开辟类似于传统限制酶克隆的定向基因转移的可能性,该定向基因转移可以提高基因编辑的效率。与上述Cas9变体和直向同源物一样,Cpf1也可以将可被CRISPR靶向的位点数目扩展至缺乏SpCas9所偏爱的NGG PAM位点的富含AT的区域或富含AT的基因组。Cpf1基因座含有混合α/β结构域、RuvC-I(后面是螺旋区域)、RuvC-II和锌指样结构域。Cpf1蛋白具有与Cas9的RuvC结构域类似的RuvC样核酸内切酶结构域。

[0430] 此外,与Cas9不同,Cpf1没有HNH核酸内切酶结构域,并且Cpf1的N端没有Cas9的α-螺旋识别叶。Cpf1 CRISPR-Cas结构域架构显示Cpf1在功能上是独特的,被归类为2类V型CRISPR系统。Cpf1基因座编码Cas1、Cas2和Cas4蛋白,该Cas1、Cas2和Cas4蛋白相比于II型系统与I型和III型系统更相似。功能性Cpf1不需要反式激活性CRISPR RNA(tracrRNA),因此,只需要CRISPR(crRNA)。这有利于基因组编辑,因为Cpf1不仅比Cas9小,而且它具有更小的sgRNA分子(大约是Cas9的核苷酸的一半)。与被Cas9靶向的富含G的PAM相比,Cpf1-crRNA复合物通过鉴定原型间隔区相邻基序5'-YTN-3'或5'-TTN-3'来切割靶DNA或RNA。在鉴定出PAM后,Cpf1引入了具有4或5个核苷酸悬突的粘性末端样DNA双链断裂。

[0431] 在一些实施方案中,Cas9是对改变的PAM序列具有特异性的Cas9变体。在一些实施方案中,另外的Cas9变体和PAM序列描述于Miller,S.M.等人Continuous evolution of SpCas9 variants compatible with non-G PAMs,Nat.Biotechnol.(2020),其全部内容以引用方式并入本文。在一些实施方案中,Cas9变体没有特定的PAM要求。在一些实施方案中,Cas9变体如SpCas9变体对NRNH PAM具有特异性,其中R是A或G并且H是A、C或T。在一些实施方案中,SpCas9变体对PAM序列AAA、TAA、CAA、GAA、TAT、GAT或CAC具有特异性。在一些实施方

案中,SpCas9变体包含位于位置1114、1134、1135、1137、1139、1151、1180、1188、1211、1218、1219、1221、1249、1256、1264、1290、1318、1317、1320、1321、1323、1332、1333、1335、1337或1339或它们的相应位置处的氨基酸取代。在一些实施方案中,SpCas9变体包含位于位置1114、1135、1218、1219、1221、1249、1320、1321、1323、1332、1333、1335或1337或它们的相应位置处的氨基酸取代。在一些实施方案中,SpCas9变体包含位于位置1114、1134、1135、1137、1139、1151、1180、1188、1211、1219、1221、1256、1264、1290、1318、1317、1320、1323、1333或它们的相应位置处的氨基酸取代。在一些实施方案中,SpCas9变体包含位于位置1114、1131、1135、1150、1156、1180、1191、1218、1219、1221、1227、1249、1253、1286、1293、1320、1321、1332、1335、1339或它们的相应位置处的氨基酸取代。在一些实施方案中,SpCas9变体包含位于位置1114、1127、1135、1180、1207、1219、1234、1286、1301、1332、1335、1337、1338或1349或它们的相应位置处的氨基酸取代。SpCas9变体的示例性氨基酸取代和PAM特异性在表2A至表2D中示出。

[0432] 表2A SpCas9变体

SpCas9	SpCas9 氨基酸位置												
	1114	1135	1218	1219	1221	1249	1320	1321	1323	1332	1333	1335	1337
	R	D	G	E	Q	P	A	P	A	D	R	R	T
AAA		N		V	H						G		
AAA		N		V	H						G		
AAA				V							G		
TAA	G	N		V							I		
TAA		N		V							I		A
TAA	G	N		V							I		A
CAA				V							K		
CAA		N		V							K		
CAA		N		V							K		
GAA				V	H		V				K		
GAA		N		V			V				K		
GAA				V	H		V				K		
TAT			S	V	H	S		S				L	
TAT			S	V	H	S		S				L	
TAT			S	V	H	S		S				L	
GAT				V							I		
GAT				V					D			Q	
GAT				V					D			Q	
CAC				V						N		Q	N
CAC		N		V								Q	N
CAC				V						N		Q	N

[0433]

[0434]

表 2B

SpCas9	SpCas9 氨基酸位置																	
	1114	1134	1135	1137	1139	1151	1180	1181	1211	1219	1221	1256	1264	1290	1318	1317	1320	1323	1333
	R	F	D	P	V	K	D	K	K	E	Q	Q	H	V	L	N	A	A	R
GAA										V	H						V		K
GAA			N	S						V			Y				V	D	K
GAA			N							V	H		Y				V		K
CAA			N							V	H		Y				V		K
CAA	G		N	S						V	H		Y				V		K
CAA			N					R		V	H						V		K
CAA			N				G		R	V	H		Y				V		K
CAA			N							V	H		Y				V		K
AAA			N				G			V	H	R	Y				V	D	K
CAA	G		N				G			V	H		Y				V	D	K
CAA		L	N				G			V	H		Y			T	V	D	K
TAA	G		N				G			V	H		Y	G	S		V	D	K
TAA	G		N			E	G			V	H		Y		S		V		K
TAA	G		N				G			V	H		Y		S		V	D	K
TAA	G		N				G		R	V	H						V		K
TAA			N				G		R	V	H		Y				V		K
TAA	G		N		A		G			V	H						V		K
TAA	G		N							V	H						V		K

[0435]

表 2C

SpCas9	SpCas9 氨基酸位置																			
	1114	1131	1135	1150	1156	1180	1190	1191	1218	1219	1348	1219	1221	1227	1249	1253	1283	1291	1320	1321	1332	1335	1339
	R	Y	D	E	K	D	K	G	D	E	N	E	Q	A	P	E	N	A	A	P	D	R	T
SacB.TAT			N				N					V	H		S				V	S		L	
SacB.TAT			N					S				V	H		S					S	G	L	
AAT			N					S				V	H	V	S		K	T		S	G	L	I
TAT	G		N				G	S				V	H		S	K				S	G	L	
TAT	G		N				G	S				V	H		S					S	G	L	
TAT	G	C	N				G	S				V	H		S					S	G	L	
TAT	G	C	N				G	S				V	H		S					S	G	L	
TAT	G	C	N				G	S				V	H		S					S	G	L	
TAT	G	C	N		E	G		S				V	H		S					S	G	L	
TAT	G	C	N	V		G		S				V	H		S					S	G	L	
TAT		C	N			G		S				V	H		S					S	G	L	
TAT	G	C	N			G		S				V	H		S					S	G	L	

表 2D

SpCas9	SpCas9 氨基酸位置																			
	1114	1127	1135	1180	1207	1219	1234	1286	1301	1322	1332	1335	1337	1338	1349					
	R	D	D	D	E	E	N	N	P	D	D	R	T	S	H					
SacB.CAC			N			V					N	Q	N							
AAC	G		N			V					N	Q	N							
AAC	G		N			V					N	Q	N							
TAC	G		N			V					N	Q	N							
TAC	G		N			V		H			N	Q	N							
TAC	G		N		G	V	D	H			N	Q	N							

[0436]

SpCas9	SpCas9氨基酸位置													
	1114	1127	1135	1180	1207	1219	1234	1286	1301	1332	1335	1337	1338	1349
	R	D	D	D	E	E	N	N	P	D	R	T	S	H
TAC	G		N			V				N	Q	N		
TAC	G	G	N	E		V		H		N	Q	N		
TAC	G		N			V		H		N	Q	N		
TAC	G		N			V				N	Q	N	T	R

[0437] 在一些实施方案中,核酸可编程DNA结合蛋白(napDNAbp)是微生物CRISPR-Cas系统的单效应子。微生物CRISPR-Cas系统的单效应子包括但不限于Cas9、Cpf1、Cas12b/C2c1和Cas12c/C2c3。通常,微生物CRISPR-Cas系统分为1类和2类系统。1类系统具有多亚基效应子复合物,而2类系统具有单蛋白质效应子。例如,Cas9和Cpf1是2类效应子。除了Cas9和Cpf1,三种不同的2类CRISPR-Cas系统(Cas12b/C2c1和Cas12c/C2c3)已经描述于Shmakov等人,“Discovery and Functional Characterization of Diverse Class 2CRISPR Cas Systems”,Mol.Cell,2015年11月5日;60(3):385-397,其全部内容据此以引用方式并入。这

些系统中的两个系统的效应子Cas12b/C2c1和Cas12c/C2c3含有与Cpf1相关的RuvC样核酸内切酶结构域。第三系统含有具有两个预测的HEPN RNA酶结构域的效应子。成熟CRISPR RNA的产生不依赖于tracrRNA,与通过Cas12b/C2c1产生CRISPR RNA不同。Cas12b/C2c1依赖于用于DNA切割的CRISPR RNA和tracrRNA。

[0438] 在一些实施方案中,napDNAbp是环状排列变换物(例如,SEQ ID NO:257)。

[0439] 据报道,酸土脂环酸芽孢杆菌(Alicyclobaccillus acidoterrastris)Cas12b/C2c1(AacC2c1)的晶体结构与嵌合单分子引导RNA(sgRNA)复合。参见例如,Liu等人,"C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism",Mol.Cell,2017年1月19日;65(2):310-322,其全部内容据此以引用方式并入。还报道了与靶DNA结合为三元复合物的酸土脂环酸芽孢杆菌(Alicyclobacillus acid oterrestris)C2c1中的晶体结构。参见例如Yang等人,"PAM-depend ent Target DNA Recognition and Cleavage by C2C1 CRISPR-Cas endonuclease",Cell,2016年12月15日;167(7):1814-1828,其全部内容以引用方式并入本文。AacC2c1的具有催化能力的构象(具有靶DN A链和非靶DNA链)已被独立捕获,定位在单个RuvC催化口袋内,具有Cas12b/C2c1介导的导致靶DNA的交错的7个核苷酸断裂的切割。Cas12b/C2c1三元复合物和先前鉴定的Cas9和Cpf1对应物之间的结构比较证明了CRISPR-Cas9系统使用的机制的多样性。

[0440] 在一些实施方案中,本文提供的融合蛋白中的任何融合蛋白的核酸可编程DNA结合蛋白(napDNAbp)可以是Cas12b/C2c1或Cas12c/C2c3蛋白。在一些实施方案中,napDNAbp是Cas12b/C2c1蛋白。在一些实施方案中,napDNAbp是Cas12c/C2c3蛋白。在一些实施方案中,napDNAbp包含与天然存在的Cas12b/C2c1或Cas12c/C2c3蛋白至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%同一的氨基酸序列。在一些实施方案中,napDNAbp是天然存在的Cas12b/C2c1或者Cas12c/C2c3蛋白。在一些实施方案中,napDNAbp包含与本文提供的napDNAbp序列中的任何一个napDNAbp序列至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%同一的氨基酸序列。应当理解,根据本公开也可以使用来自其他细菌物种的Cas12b/C2c1或Cas12c/C2c3。

[0441] 在一些实施方案中,napDNAbp是指Cas12c。在一些实施方案中,Cas12c蛋白是Cas12c1(SEQ ID NO:266)或Cas12c1的变体。在一些实施方案中,Cas12蛋白是Cas12c2(SEQ ID NO:267)或Cas12c2的变体。在一些实施方案中,Cas12蛋白是来自嗜油菌属(Oleiphilus)种HI0009(即,OspCas12c;SEQ ID NO:268)的Cas12c蛋白或OspCas12c的变体。这些Cas12c分子已描述于Yan等人,"Functionally Diverse Type V CRISPR-Cas Systems,"Science,2019年1月4日;363:88-91;其全部内容以引用方式并入本文。在一些实施方案中,napDNAbp包含与天然存在的Cas12c1、Cas12c2或OspCas12c蛋白至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%同一的氨基酸序列。在一些实施方案中,napDNAbp是天然存在的Cas12c1、Cas12c2或OspCas12c蛋白。在一些实施方案中,napDNAbp包含与本文提供的任何Cas12c1、Cas12c2或OspCas12c蛋白至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%同一的氨基酸序列。应当理解,根据本公开也可以使用来自其他细菌物种的Cas12c1、Cas12c2或OspCas12c。

[0442] 在一些实施方案中,napDNAbp是指Cas12g、Cas12h或Cas12i,它们已描述于例如Yan等人,“Functionally Diverse Type V CRISPR-Cas Systems,”Science,2019年1月4日;363:88-91;每一篇的全部内容均据此以引用方式并入。示例性Cas12g、Cas12h和Cas12i多肽序列在序列表中作为SEQ ID NO:269-272提供。通过聚集超过10垓字节的序列数据,鉴定了V型Cas蛋白的新分类,这些分类与先前表征的V类蛋白(包括Cas12g、Cas12h和Cas12i)表现出弱类似性。在一些实施方案中,Cas12蛋白是Cas12g或Cas12g的变体。在一些实施方案中,Cas12蛋白是Cas12h或Cas12h的变体。在一些实施方案中,Cas12蛋白是Cas12i或Cas12i的变体。应当理解,其他RNA引导的DNA结合蛋白可以用作napDNAbp,并且在本公开的范围内。在一些实施方案中,napDNAbp包含与天然存在的Cas12g、Cas12h或Cas12i蛋白至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%同一的氨基酸序列。在一些实施方案中,napDNAbp是天然存在的Cas12g、Cas12h或Cas12i蛋白。在一些实施方案中,napDNAbp包含与本文提供的任何Cas12g、Cas12h或Cas12i蛋白至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%同一的氨基酸序列。应当理解,根据本公开也可以使用来自其他细菌物种的Cas12g、Cas12h或Cas12i。在一些实施方案中,Cas12i是Cas12i1或Cas12i2。

[0443] 在一些实施方案中,本文提供的融合蛋白中的任何融合蛋白的核酸可编程DNA结合蛋白(napDNAbp)可以是Cas12j/CasΦ蛋白。Cas12j/CasΦ描述于Pausch等人,“CRISPR-CasΦfrom huge phages is a hypercompact genome editor,”Science,2020年7月17日,第369卷,第6501期,第333-337页,其以引用方式整体并入本文。在一些实施方案中,napDNAbp包含与天然存在的Cas12j/CasΦ蛋白至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%同一的氨基酸序列。在一些实施方案中,napDNAbp是天然存在的Cas12j/CasΦ蛋白。在一些实施方案中,napDNAbp是核酸酶无活性的(“失效的”)Cas12j/CasΦ蛋白。应当理解,根据本公开也可以使用来自其他物种的Cas12j/CasΦ。

[0444] 具有内部插入的融合蛋白

[0445] 本文提供了包含与核酸可编程核酸结合蛋白(例如,napDNAbp)融合的异源多肽的融合蛋白。异源多肽可以是在原生或野生型napDNAbp多肽序列中未发现的多肽。异源多肽可以在napDNAbp的C端、napDNAbp的N端融合到napDNAbp,或插入在napDNAbp的内部区位处。在一些实施方案中,异源多肽是脱氨酶(例如胞苷或腺苷脱氨酶)或其功能片段。例如,融合蛋白可以包含侧接有Cas9或Cas12(例如,Cas12b/C2c1)多肽的N端片段和C端片段的脱氨酶。在一些实施方案中,胞苷脱氨酶是APOBEC脱氨酶(例如APOBEC1)。在一些实施方案中,腺苷脱氨酶是TadA(例如,TadA*7.10或TadA*8)。在一些实施方案中,TadA是TadA*8或TadA*9。如本文所述的TadA序列(例如,TadA7.10或TadA*8)是用于上述融合蛋白的合适的脱氨酶。

[0446] 在一些实施方案中,融合蛋白包含以下结构:

[0447] NH2-[napDNAbp的N端片段]-[脱氨酶]-[napDNAbp的C端片段]-COOH;

[0448] NH2-[Cas9的N端片段]-[腺苷脱氨酶]-[Cas9的C端片段]-COOH;

[0449] NH2-[Cas12的N端片段]-[腺苷脱氨酶]-[Cas12的C端片段]-COOH;

[0450] NH2-[Cas9的N端片段]-[胞苷脱氨酶]-[Cas9的C端片段]-COOH;

[0451] NH2-[Cas12的N端片段]-[胞苷脱氨酶]-[Cas12的C端片段]-COOH;

[0452] 其中每个"]-["实例是任选的接头。

[0453] 脱氨酶可以是环状排列变换脱氨酶。例如,脱氨酶可以是环状排列变换腺苷脱氨酶。在一些实施方案中,脱氨酶是在如TadA参考序列中编号的氨基酸残基116、136或65处发生环状排列变换的环状排列变换TadA。

[0454] 融合蛋白可以包含多于一种脱氨酶。融合蛋白可以包含例如1种、2种、3种、4种、5种或更多种脱氨酶。在一些实施方案中,融合蛋白包含一种或两种脱氨酶。融合蛋白中的两种或更多种脱氨酶可以是腺苷脱氨酶、胞苷脱氨酶或其组合。该两种或更多种脱氨酶可以是同二聚体或异二聚体。该两种或更多种脱氨酶可以串联插入在napDNAbp中。在一些实施方案中,两种或更多种脱氨酶在napDNAbp中可以不是串联的。

[0455] 在一些实施方案中,融合蛋白中的napDNAbp是Cas9多肽或其片段。Cas9多肽可以是变体Cas9多肽。在一些实施方案中,Cas9多肽是Cas9切口酶(nCas9)多肽或其片段。在一些实施方案中,Cas9多肽是核酸酶失效的Cas9(dCas9)多肽或其片段。融合蛋白中的Cas9多肽可以是全长Cas9多肽。在一些情况下,融合蛋白中的Cas9多肽可以不是全长Cas9多肽。Cas9多肽可以例如在相对于天然存在的Cas9蛋白的N端或C端处被截短。Cas9多肽可以是发生环状排列变换的Cas9蛋白。Cas9多肽可以是仍然能够结合靶多核苷酸和引导核酸序列的Cas9多肽的片段、部分或结构域。

[0456] 在一些实施方案中,Cas9多肽是酿脓链球菌Cas9(SpCas9)、金黄色葡萄球菌Cas9(SaCas9)、嗜热链球菌1Cas9(St1Cas9),或本文所述的Cas9多肽中的任何Cas9多肽的片段或变体。

[0457] 在一些实施方案中,融合蛋白包含插入在Cas9内的腺苷脱氨酶结构域和胞苷脱氨酶结构域。在一些实施方案中,腺苷脱氨酶融合在Cas9内,并且胞苷脱氨酶融合至C端。在一些实施方案中,腺苷脱氨酶融合在Cas9内,并且胞苷脱氨酶融合至N端。在一些实施方案中,胞苷脱氨酶融合在Cas9内,并且腺苷脱氨酶融合至C端。在一些实施方案中,胞苷脱氨酶融合在Cas9内,并且腺苷脱氨酶融合至N端。

[0458] 具有腺苷脱氨酶和胞苷脱氨酶和Cas9的融合蛋白的示例性结构提供如下:

[0459] NH2-[Cas9(腺苷脱氨酶)]-[胞苷脱氨酶]-COOH;

[0460] NH2-[胞苷脱氨酶]-[Cas9(腺苷脱氨酶)]-COOH;

[0461] NH2-[Cas9(胞苷脱氨酶)]-[腺苷脱氨酶]-COOH;或

[0462] NH2-[腺苷脱氨酶]-[Cas9(胞苷脱氨酶)]-COOH。

[0463] 在一些实施方案中,上方通用架构中使用的"-"指示存在任选的接头。

[0464] 在各种实施方案中,催化结构域具有DNA修饰活性(例如脱氨酶活性),诸如腺苷脱氨酶活性。在一些实施方案中,腺苷脱氨酶是TadA(例如,TadA*7.10)。在一些实施方案中,TadA是TadA*8。在一些实施方案中,TadA*8融合在Cas9内,并且胞苷脱氨酶融合至C端。在一些实施方案中,TadA*8融合在Cas9内,并且胞苷脱氨酶融合至N端。在一些实施方案中,胞苷脱氨酶融合在Cas9内,并且TadA*8融合至C端。在一些实施方案中,胞苷脱氨酶融合在Cas9内,并且TadA*8融合至N端。具有TadA*8和胞苷脱氨酶和Cas9的融合蛋白的示例性结构提供如下:

[0465] NH2-[Cas9(TadA*8)]-[胞苷脱氨酶]-COOH;

[0466] NH2-[胞苷脱氨酶]-[Cas9(TadA*8)]-COOH;

[0467] NH2-[Cas9(胞苷脱氨酶)]-[TadA*8]-COOH;或

[0468] NH2-[TadA*8]-[Cas9(胞苷脱氨酶)]-COOH。

[0469] 在一些实施方案中,上方通用架构中使用的“-”指示存在任选的接头。

[0470] 异源多肽(例如,脱氨酶)可以插入在napDNAbp(例如,Cas9或Cas12(例如,Cas12b/C2c1))中的合适的区位处,例如,使得napDNAbp保留其结合靶多核苷酸和引导核酸的能力。脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)可以插入到napDNAbp中而不损害脱氨酶的功能(例如,碱基编辑活性)或napDNAbp的功能(例如,结合至靶核酸和引导核酸的能力)。脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)可以插入在napDNAbp中的例如无序区域或如由晶体学研究所示的包含高温因子或B因子的区域处。不太有序、无序或非结构化的蛋白区域,例如溶剂暴露区域和环,可以用于插入而不损害结构或功能。脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)可以插入在napDNAbp中的柔性环区域或溶剂暴露区域中。在一些实施方案中,将脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)插入在Cas9的柔性环或Cas12b/C2c1多肽中。

[0471] 在一些实施方案中,通过Cas9多肽的晶体结构的B因子分析来确定脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)的插入区位。在一些实施方案中,将脱氨酶(例如腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)插入在包含高于平均水平的B因子(例如,相比于包含无序区域的总蛋白或蛋白结构域更高的B因子)的Cas9多肽区域中。B因子或温度因子可以指示原子相对于其平均位置的波动(例如,由于晶格中的温度依赖性原子振动或静态无序)。骨架原子的高B因子(例如,高于平均水平的B因子)可以指示具有相对高局部迁移率的区域。此种区域可以用于插入脱氨酶而不损害结构或功能。脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)可以插入在带有具有Cα原子的残基的区位处,该区位具有比总蛋白的平均B因子多50%、60%、70%、80%、90%、100%、110%、120%、130%、140%、150%、160%、170%、180%、190%、200%或高于200%的B因子。脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)可以插入在带有具有Cα原子的残基的区位处,该区位具有比包含该残基的Cas9蛋白结构域的平均B因子多50%、60%、70%、80%、90%、100%、110%、120%、130%、140%、150%、160%、170%、180%、190%、200%或高于200%的B因子。包含高于平均水平的B因子的Cas9多肽位置可以包括例如如上面Cas9参考序列中编号的残基768、792、1052、1015、1022、1026、1029、1067、1040、1054、1068、1246、1247和1248。包含高于平均水平的B因子的Cas9多肽区域可以包括例如如上面Cas9参考序列中编号的残基792-872、792-906和2-791。

[0472] 异源多肽(例如,脱氨酶)可以被插入在napDNAbp中的选自由以下组成的组的氨基酸残基:如上面Cas9参考序列中编号的768、791、792、1015、1016、1022、1023、1026、1029、1040、1052、1054、1067、1068、1069、1246、1247和1248,或另一种Cas9多肽中的相应氨基酸残基处。在一些实施方案中,异源多肽被插入在如上面Cas9参考序列中编号的氨基酸位置768-769、791-792、792-793、1015-1016、1022-1023、1026-1027、1029-1030、1040-1041、1052-1053、1054-1055、1067-1068、1068-1069、1247-1248或1248-1249或它们的相应氨基酸位置之间。在一些实施方案中,异源多肽被插入在如上面Cas9参考序列中编号的氨基酸

位置769-770、792-793、793-794、1016-1017、1023-1024、1027-1028、1030-1031、1041-1042、1053-1054、1055-1056、1068-1069、1069-1070、1248-1249或1249-1250或它们的相应氨基酸位置之间。在一些实施方案中,异源多肽替换选自由以下组成的组的氨基酸残基:如上面Cas9参考序列中编号的768、791、792、1015、1016、1022、1023、1026、1029、1040、1052、1054、1067、1068、1069、1246、1247和1248,或另一种Cas9多肽中的相应氨基酸残基。应当理解,对上面Cas9参考序列的关于插入位置的引用是出于说明性目的。如本文所讨论的插入不限于上面Cas9参考序列的Cas9多肽序列,而是包括在变体Cas9多肽(例如Cas9切口酶(nCas9)、核酸酶失效的Cas9(dCas9)、缺少核酸酶结构域的Cas9变体、截短的Cas9,或缺少部分或完整HNH结构域的Cas9结构域)中的相应区位处的插入。

[0473] 异源多肽(例如,脱氨酶)可以被插入在napDNAbp中的选自由以下组成的组的氨基酸残基:如上面Cas9参考序列中编号的768、792、1022、1026、1040、1068和1247,或另一种Cas9多肽中的相应氨基酸残基处。在一些实施方案中,异源多肽被插入在如上面Cas9参考序列中编号的氨基酸位置768-769、792-793、1022-1023、1026-1027、1029-1030、1040-1041、1068-1069或1247-1248或它们的相应氨基酸位置之间。在一些实施方案中,异源多肽被插入在如上面Cas9参考序列中编号的氨基酸位置769-770、793-794、1023-1024、1027-1028、1030-1031、1041-1042、1069-1070或1248-1249或它们的相应氨基酸位置之间。在一些实施方案中,异源多肽替换选自由以下组成的组的氨基酸残基:如上面Cas9参考序列中编号的768、792、1022、1026、1040、1068和1247,或另一种Cas9多肽中的相应氨基酸残基。

[0474] 异源多肽(例如,脱氨酶)可以被插入在napDNAbp中的如本文所述的氨基酸残基,或另一种Cas9多肽中的相应氨基酸残基处。在一个实施方案中,异源多肽(例如-脱氨酶)可以被插入在napDNAbp中的选自由以下组成的组的氨基酸残基:如上面Cas9参考序列中编号的1002、1003、1025、1052-1056、1242-1247、1061-1077、943-947、686-691、569-578、530-539和1060-1077,或另一种Cas9多肽中的相应氨基酸残基处。脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)可以被插入在该残基的N端或C端处或者替换该残基。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)被插入在该残基的C端处。

[0475] 在一些实施方案中,腺苷脱氨酶(例如,TadA)被插入在选自由以下组成的组的氨基酸残基:如上面Cas9参考序列中编号的1015、1022、1029、1040、1068、1247、1054、1026、768、1067、1248、1052和1246,或另一种Cas9多肽中的相应氨基酸残基处。在一些实施方案中,腺苷脱氨酶(例如,TadA)被插入在如上面Cas9参考序列中编号的残基792-872、792-906或2-791,或另一种Cas9多肽中的相应氨基酸残基的位置中。在一些实施方案中,腺苷脱氨酶被插入在选自由以下组成的组的氨基酸的N端:如上面Cas9参考序列中编号的1015、1022、1029、1040、1068、1247、1054、1026、768、1067、1248、1052和1246,或另一种Cas9多肽中的相应氨基酸残基处。在一些实施方案中,腺苷脱氨酶被插入在选自由以下组成的组的氨基酸的C端:如上面Cas9参考序列中编号的1015、1022、1029、1040、1068、1247、1054、1026、768、1067、1248、1052和1246,或另一种Cas9多肽中的相应氨基酸残基处。在一些实施方案中,腺苷脱氨酶被插入以替换选自由以下组成的组的氨基酸:如上面Cas9参考序列中编号的1015、1022、1029、1040、1068、1247、1054、1026、768、1067、1248、1052和1246,或另一种Cas9多肽中的相应氨基酸残基。

[0476] 在一些实施方案中,胞苷脱氨酶(例如,APOBEC1)被插入在选自由以下组成的组的氨基酸残基:如上面Cas9参考序列中编号的1016、1023、1029、1040、1069和1247,或另一种Cas9多肽中的相应氨基酸残基处。在一些实施方案中,胞苷脱氨酶被插入在选自由以下组成的组的氨基酸的N端:如上面Cas9参考序列中编号的1016、1023、1029、1040、1069和1247,或另一种Cas9多肽中的相应氨基酸残基。在一些实施方案中,胞苷脱氨酶被插入在选自由以下组成的组的氨基酸的C端:如上面Cas9参考序列中编号的1016、1023、1029、1040、1069和1247,或另一种Cas9多肽中的相应氨基酸残基处。在一些实施方案中,胞苷脱氨酶被插入以替换选自由以下组成的组的氨基酸:如上面Cas9参考序列中编号的1016、1023、1029、1040、1069和1247,或另一种Cas9多肽中的相应氨基酸残基。

[0477] 在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)被插入在如上面Cas9参考序列中编号的氨基酸残基768,或另一种Cas9多肽中的相应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)被插入在如上面Cas9参考序列中编号的氨基酸残基768的N端,或另一种Cas9多肽中的相应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)被插入在如上面Cas9参考序列中编号的氨基酸残基768的C端,或另一种Cas9多肽中的相应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)被插入以替换如上面Cas9参考序列中编号的氨基酸残基768,或另一种Cas9多肽中的相应氨基酸残基。

[0478] 在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)被插入在如上面Cas9参考序列中编号的氨基酸残基791或氨基酸残基792,或另一种Cas9多肽中的相应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)被插入在如上面Cas9参考序列中编号的氨基酸残基791的N端或氨基酸792的N端,或另一种Cas9多肽中的相应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)被插入在如上面Cas9参考序列中编号的氨基酸791的C端或氨基酸792的N端,或另一种Cas9多肽中的相应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)被插入以替换如上面Cas9参考序列中编号的氨基酸791或氨基酸792,或另一种Cas9多肽中的相应氨基酸残基。

[0479] 在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)被插入在如上面Cas9参考序列中编号的氨基酸残基1016,或另一种Cas9多肽中的相应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)被插入在如上面Cas9参考序列中编号的氨基酸残基1016的N端,或另一种Cas9多肽中的相应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)被插入在如上面Cas9参考序列中编号的氨基酸残基1016的C端,或另一种Cas9多肽中的相应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)被插入以替换如上面Cas9参考序列中编号的氨基酸残基1016,或另一种Cas9多肽中的相应氨基酸残基。

[0480] 在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)被插入在如上面Cas9参考序列中编号的氨基酸残基1022或氨基酸残基1023,或另

一种Cas9多肽中的相应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)被插入在如上面Cas9参考序列中编号的氨基酸残基1022的N端或氨基酸残基1023的N端,或另一种Cas9多肽中的相应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)被插入在如上面Cas9参考序列中编号的氨基酸残基1022的C端或氨基酸残基1023的C端,或另一种Cas9多肽中的相应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)被插入以替换如上面Cas9参考序列中编号的氨基酸残基1022或氨基酸残基1023,或另一种Cas9多肽中的相应氨基酸残基。

[0481] 在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)被插入在如上面Cas9参考序列中编号的氨基酸残基1026或氨基酸残基1029,或另一种Cas9多肽中的相应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)被插入在如上面Cas9参考序列中编号的氨基酸残基1026的N端或氨基酸残基1029的N端,或另一种Cas9多肽中的相应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)被插入在如上面Cas9参考序列中编号的氨基酸残基1026的C端或氨基酸残基1029的C端,或另一种Cas9多肽中的相应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)被插入以替换如上面Cas9参考序列中编号的氨基酸残基1026或氨基酸残基1029,或另一种Cas9多肽中的相应氨基酸残基。

[0482] 在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)被插入在如上面Cas9参考序列中编号的氨基酸残基1040,或另一种Cas9多肽中的相应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)被插入在如上面Cas9参考序列中编号的氨基酸残基1040的N端,或另一种Cas9多肽中的相应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)被插入在如上面Cas9参考序列中编号的氨基酸残基1040的C端,或另一种Cas9多肽中的相应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)被插入以替换如上面Cas9参考序列中编号的氨基酸残基1040,或另一种Cas9多肽中的相应氨基酸残基。

[0483] 在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)被插入在如上面Cas9参考序列中编号的氨基酸残基1052或氨基酸残基1054,或另一种Cas9多肽中的相应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)被插入在如上面Cas9参考序列中编号的氨基酸残基1052的N端或氨基酸残基1054的N端,或另一种Cas9多肽中的相应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)被插入在如上面Cas9参考序列中编号的氨基酸残基1052的C端或氨基酸残基1054的C端,或另一种Cas9多肽中的相应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)被插入以替换如上面Cas9参考序列中编号的氨基酸残基1052或氨基酸残基1054,或另一种Cas9多肽中的相应氨基酸残基。

[0484] 在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)被插入在如上面Cas9参考序列中编号的氨基酸残基1067或氨基酸残基1068或氨基

酸残基1069,或另一种Cas9多肽中的相应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)被插入在如上面Cas9参考序列中编号的氨基酸残基1067的N端或氨基酸残基1068的N端或氨基酸残基1069的N端,或另一种Cas9多肽中的相应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)被插入在如上面Cas9参考序列中编号的氨基酸残基1067的C端或氨基酸残基1068的C端或氨基酸残基1069的C端,或另一种Cas9多肽中的相应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)被插入以替换如上面Cas9参考序列中编号的氨基酸残基1067或氨基酸残基1068或氨基酸残基1069,或另一种Cas9多肽中的相应氨基酸残基。

[0485] 在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)被插入在如上面Cas9参考序列中编号的氨基酸残基1246或氨基酸残基1247或氨基酸残基1248,或另一种Cas9多肽中的相应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)被插入在如上面Cas9参考序列中编号的氨基酸残基1246的N端或氨基酸残基1247的N端或氨基酸残基1248的N端,或另一种Cas9多肽中的相应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)被插入在如上面Cas9参考序列中编号的氨基酸残基1246的C端或氨基酸残基1247的C端或氨基酸残基1248的C端,或另一种Cas9多肽中的相应氨基酸残基处。在一些实施方案中,脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)被插入以替换如上面Cas9参考序列中编号的氨基酸残基1246或氨基酸残基1247或氨基酸残基1248,或另一种Cas9多肽中的相应氨基酸残基。

[0486] 在一些实施方案中,异源多肽(例如,脱氨酶)被插入在Cas9多肽的柔性环中。柔性环部分可以选自由以下组成的组:如上面Cas9参考序列中编号的530-537、569-570、686-691、943-947、1002-1025、1052-1077、1232-1247或1298-1300,或另一种Cas9多肽中的相应氨基酸残基。柔性环部分可以选自由以下组成的组:如上面Cas9参考序列中编号的1-529、538-568、580-685、692-942、948-1001、1026-1051、1078-1231或1248-1297,或另一种Cas9多肽中的相应氨基酸残基。

[0487] 异源多肽(例如,腺嘌呤脱氨酶)可以被插入到对应于氨基酸残基:如上面Cas9参考序列中编号的1017-1069、1242-1247、1052-1056、1060-1077、1002-1003、943-947、530-537、568-579、686-691、1242-1247、1298-1300、1066-1077、1052-1056或1060-1077,或另一种Cas9多肽中的相应氨基酸残基的Cas9多肽区域中。

[0488] 异源多肽(例如腺嘌呤脱氨酶)可以被插入在Cas9多肽的缺失的区域的位置中。该缺失的区域可以对应于Cas9多肽的N端或C端部分。在一些实施方案中,缺失的区域对应于如上面Cas9参考序列中编号的残基792-872,或另一种Cas9多肽中的相应氨基酸残基。在一些实施方案中,缺失的区域对应于如上面Cas9参考序列中编号的残基792-906,或另一种Cas9多肽中的相应氨基酸残基。在一些实施方案中,缺失的区域对应于如上面Cas9参考序列中编号的残基2-791,或另一种Cas9多肽中的相应氨基酸残基。在一些实施方案中,缺失的区域对应于如上面Cas9参考序列中编号的残基1017-1069,或其对应的氨基酸残基。

[0489] 示例性内部融合碱基编辑器提供于下方表3中:

[0490] 表3:Cas9蛋白中的插入基因座

[0491]

BE ID	修饰	其他ID
IBE001	Cas9 TadA ins 1015	ISLAY01
IBE002	Cas9 TadA ins 1022	ISLAY02
IBE003	Cas9 TadA ins 1029	ISLAY03
IBE004	Cas9 TadA ins 1040	ISLAY04
IBE005	Cas9 TadA ins 1068	ISLAY05
IBE006	Cas9 TadA ins 1247	ISLAY06
IBE007	Cas9 TadA ins 1054	ISLAY07
IBE008	Cas9 TadA ins 1026	ISLAY08
IBE009	Cas9 TadA ins 768	ISLAY09
IBE020	ΔHNH TadA 792	ISLAY20
IBE021	N端融合单TadA螺旋截短165末端	ISLAY21
IBE029	TadA-环状排列变换116 ins 1067	ISLAY29
IBE031	TadA-环状排列变换136 ins 1248	ISLAY31
IBE032	TadA-环状排列变换136ins 1052	ISLAY32
IBE035	Δ792-872 TadA ins	ISLAY35
IBE036	Δ792-906 TadA ins	ISLAY36
IBE043	TadA-环状排列变换65 ins 1246	ISLAY43
IBE044	TadA ins C端截短2 791	ISLAY44

[0492] 异源多肽(例如,脱氨酶)可以被插入在Cas9多肽的结构性或功能性结构域内。异源多肽(例如,脱氨酶)可以被插入在Cas9多肽的两个结构性或功能性结构域之间。异源多肽(例如脱氨酶)可以被插入在Cas9多肽的结构性或功能性结构域的位置中,例如在使该结构域从Cas9多肽缺失之后。Cas9多肽的结构性或功能性结构域可以包括例如RuvC I、RuvC II、RuvC III、Rec1、Rec2、PI或HNH。

[0493] 在一些实施方案中,Cas9多肽缺乏一个或多个选自由以下组成的组的结构域:RuvC I、RuvC II、RuvC III、Rec1、Rec2、PI或HNH结构域。在一些实施方案中,Cas9多肽缺乏核酸酶结构域。在一些实施方案中,Cas9多肽缺乏HNH结构域。在一些实施方案中,Cas9多肽缺乏部分HNH结构域,使得Cas9多肽具有减小的或消除的HNH活性。在一些实施方案中,Cas9多肽包含核酸酶结构域的缺失,并且脱氨酶被插入以替换该核酸酶结构域。在一些实施方案中,使HNH结构域缺失并且在其位置中插入脱氨酶。在一些实施方案中,使RuvC结构域中的一个或多个RuvC结构域缺失并且在其位置中插入脱氨酶。

[0494] 包含异源多肽的融合蛋白可以侧接有napDNAbp的N端和C端片段。在一些实施方案中,融合蛋白包含侧接有Cas9多肽的N端片段和C端片段的脱氨酶。该N端片段或该C端片段可以结合靶多核苷酸序列。该N端片段的C端或该C端片段的N端可以包含Cas9多肽的柔性环的一部分。该N端片段的C端或该C端片段的N端可以包含Cas9多肽的α-螺旋结构的一部分。该N端片段或该C端片段可以包含DNA结合结构域。该N端片段或该C端片段可以包含RuvC结构域。该N端片段或该C端片段可以包含HNH结构域。在一些实施方案中,N端片段和该C端片段都不包含HNH结构域。

[0495] 在一些实施方案中,N端Cas9片段的C端包含当融合蛋白使靶核碱基脱氨时接近靶

核碱基的氨基酸。在一些实施方案中,C端Cas9片段的N端包含当融合蛋白使靶核碱基脱氨时接近靶核碱基的氨基酸。不同脱氨酶的插入区位可以不同,以便使靶核碱基与该N端Cas9片段的C端或C端Cas9片段的N端中的氨基酸接近。例如,脱氨酶的插入位置可以位于选自由以下组成的组的氨基酸残基:如上面Cas9参考序列中编号的1015、1022、1029、1040、1068、1247、1054、1026、768、1067、1248、1052和1246,或另一种Cas9多肽中的相应氨基酸残基处。

[0496] 融合蛋白的N端Cas9片段(即融合蛋白中侧接脱氨酶的N端Cas9片段)可以包含Cas9多肽的N端。融合蛋白的N端Cas9片段可以包含至少约100个、200个、300个、400个、500个、600个、700个、800个、900个、1000个、1100个、1200个或1300个氨基酸的长度。融合蛋白的N端Cas9片段可以包含对应于氨基酸残基:如上面Cas9参考序列中编号的1-56、1-95、1-200、1-300、1-400、1-500、1-600、1-700、1-718、1-765、1-780、1-906、1-918或1-1100,或另一种Cas9多肽中的相应氨基酸残基的序列。N端Cas9片段可以包含与氨基酸残基:如上面Cas9参考序列中编号的1-56、1-95、1-200、1-300、1-400、1-500、1-600、1-700、1-718、1-765、1-780、1-906、1-918或1-1100,或另一种Cas9多肽中的相应氨基酸残基包含至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%序列同一性的序列。

[0497] 融合蛋白的C端Cas9片段(即融合蛋白中侧接脱氨酶的C端Cas9片段)可以包含Cas9多肽的C端。融合蛋白的C端Cas9片段可以包含至少约100个、200个、300个、400个、500个、600个、700个、800个、900个、1000个、1100个、1200个或1300个氨基酸的长度。融合蛋白的C端Cas9片段可以包含对应于氨基酸残基:如上面Cas9参考序列中编号的1099-1368、918-1368、906-1368、780-1368、765-1368、718-1368、94-1368或56-1368,或另一种Cas9多肽中的相应氨基酸残基的序列。N端Cas9片段可以包含与氨基酸残基:如上面Cas9参考序列中编号的1099-1368、918-1368、906-1368、780-1368、765-1368、718-1368、94-1368或56-1368,或另一种Cas9多肽中的相应氨基酸残基包含至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%序列同一性的序列。

[0498] 融合蛋白的N端Cas9片段和C端Cas9片段合在一起可能不对应于例如如上面Cas9参考序列中所阐述的全长天然存在的Cas9多肽序列。

[0499] 本文所述的融合蛋白可以实现被靶向脱氨,同时减少非靶位点(例如脱靶位点)处的脱氨,诸如减少全基因组的假脱氨。本文所述的融合蛋白可以实现被靶向脱氨,同时减少非靶位点处的旁观者脱氨(bystander deamination)。与例如包含与Cas9多肽的N端或C端融合的脱氨酶的末端融合蛋白相比,不需要的脱氨或脱靶脱氨可以减少至少30%、至少40%、至少50%、至少60%、至少70%、至少80%、至少90%、至少95%或至少99%。与例如包含与Cas9多肽的N端或C端融合的脱氨酶的末端融合蛋白相比,不需要的脱氨或脱靶脱氨可以减少至少一倍、至少二倍、至少三倍、至少四倍、至少五倍、至少十倍、至少十五倍、至少二十倍、至少三十倍、至少四十倍、至少五十倍、至少60倍、至少70倍、至少80倍、至少90倍或至少一百倍。

[0500] 在一些实施方案中,融合蛋白的脱氨酶(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)使R环范围内的不超过两个核碱基脱氨。在一些实施方案中,融合蛋白的脱氨酶使R环范围内的不超过三个核碱基脱氨。在一些实施方案中,融合蛋白的脱氨酶使

R环范围内的不超过2个、3个、4个、5个、6个、7个、8个、9个或10个核碱基脱氨。R环是包含DNA:RNA杂合体、DNA:DNA或RNA:RNA互补结构并与单链DNA缔合的三链核酸结构。如本文所用,当靶多核苷酸与CRISPR复合物或碱基编辑复合物接触时可以形成R环,其中引导多核苷酸(例如引导RNA)的一部分与靶多核苷酸的一部分杂交并且用靶多核苷酸(例如靶DNA)的一部分置换。在一些实施方案中,R环包含间隔区序列和靶DNA互补序列的杂交区域。R环区域的长度可以是约5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49或50个核碱基对。在一些实施方案中,R环区域的长度为约20个核碱基对。应当理解,如本文所用,R环区域不限于与引导多核苷酸杂交的靶DNA链。例如,R环区域内的靶核碱基的编辑可以针对包含引导RNA互补链的DNA链,或者可以针对作为引导RNA互补链的相反链的DNA链。在一些实施方案中,R环的区域中的编辑包括将非互补链(原型间隔链)上的核碱基编辑为靶DNA序列中的引导RNA。

[0501] 本文所述的融合蛋白可以在不同于规范碱基编辑的编辑窗口中实现靶脱氨。在一些实施方案中,靶核碱基是靶多核苷酸序列中PAM序列上游的约1至约20个碱基。在一些实施方案中,靶核碱基是靶多核苷酸序列中PAM序列上游的约2至约12个碱基。在一些实施方案中,靶核碱基是远离PAM序列或在PAM序列上游的约1至9个碱基对、约2至10个碱基对、约3至11个碱基对、约4至12个碱基对、约5至13个碱基对、约6至14个碱基对、约7至15个碱基对、约8至16个碱基对、约9至17个碱基对、约10至18个碱基对、约11至19个碱基对、约12至20个碱基对、约1至7个碱基对、约2至8个碱基对、约3至9个碱基对、约4至10个碱基对、约5至11个碱基对、约6至12个碱基对、约7至13个碱基对、约8至14个碱基对、约9至15个碱基对、约10至16个碱基对、约11至17个碱基对、约12至18个碱基对、约13至19个碱基对、约14至20个碱基对、约1至5个碱基对、约2至6个碱基对、约3至7个碱基对、约4至8个碱基对、约5至9个碱基对、约6至10个碱基对、约7至11个碱基对、约8至12个碱基对、约9至13个碱基对、约10至14个碱基对、约11至15个碱基对、约12至16个碱基对、约13至17个碱基对、约14至18个碱基对、约15至19个碱基对、约16至20个碱基对、约1至3个碱基对、约2至4个碱基对、约3至5个碱基对、约4至6个碱基对、约5至7个碱基对、约6至8个碱基对、约7至9个碱基对、约8至10个碱基对、约9至11个碱基对、约10至12个碱基对、约11至13个碱基对、约12至14个碱基对、约13至15个碱基对、约14至16个碱基对、约15至17个碱基对、约16至18个碱基对、约17至19个碱基对、约18至20个碱基对。在一些实施方案中,靶核碱基是远离PAM序列或在PAM序列上游的约1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个或更多个碱基对。在一些实施方案中,靶核碱基是PAM序列上游的约1个、2个、3个、4个、5个、6个、7个、8个或9个碱基对。在一些实施方案中,靶核碱基是PAM序列上游的约2个、3个、4个或6个碱基对。

[0502] 融合蛋白可以包含多于一种异源多肽。例如,融合蛋白可以另外包含一个或多个UGI结构域和/或一种或多种核定位信号。该两个或更多个异源结构域可以串联插入。该两个或更多个异源结构域可以插入在使得它们在NapDNAbp中不串联的区位处。

[0503] 融合蛋白可以在脱氨酶和napDNAbp多肽之间包含接头。接头可以是肽或非肽接头。例如,接头可以是XTEN、(GGGS)n(SEQ ID NO:1308)、(GGGGS)n(SEQ ID NO:109)、(G)n、(EAAAK)n(SEQ ID NO:1309)、(GGS)n、SGSETPGTSESATPES(SEQ ID NO:56)。在一些实施方案

中,融合蛋白在N端Cas9片段和脱氨酶之间包含接头。在一些实施方案中,融合蛋白在C端Cas9片段和脱氨酶之间包含接头。在一些实施方案中,napDNAbp的N端和C端片段在有接头的情况下与脱氨酶连接。在一些实施方案中,N端和C端片段在没有接头的情况下与脱氨酶结构域联接。在一些实施方案中,融合蛋白在N端Cas9片段和脱氨酶之间包含接头,但在C端Cas9片段和脱氨酶之间不包含接头。在一些实施方案中,融合蛋白在C端Cas9片段和脱氨酶之间包含接头,但在N端Cas9片段和脱氨酶之间不包含接头。

[0504] 在一些实施方案中,融合蛋白中的napDNAbp是Cas12多肽(例如,Cas12b/C2c1)或其片段。Cas12多肽可以是变体Cas12多肽。在其他实施方案中,Cas12多肽的N端或C端片段包含核酸可编程DN A结合结构域或RuvC结构域。在其他实施方案中,融合蛋白在Cas12多肽和催化结构域之间含有接头。在其他实施方案中,接头的氨基酸序列是GGSGGS(SEQ ID NO:273)或GSSGSETPGTSESATPE SSG(SEQ ID NO:1310)。在其他实施方案中,接头是刚性接头。在以上方面的其他实施方案中,接头由GGAGGCTCTGGAGGAAGC(SEQ ID NO:1311)或GGCTCTTCTGGATCTGAAACACCTGGCACA AGCGAGAGCGCCACCCCTGAGAGCTCTGGC(SEQ ID NO:1312)编码。

[0505] 包含侧接有Cas12多肽的N-和C-末端片段的异源催化结构域的融合蛋白对如本文所述方法中的碱基编辑也是有用的。包含Cas12和一个或多个脱氨酶结构域(例如腺苷脱氨酶)或者包含侧接有Cas12序列的腺苷脱氨酶结构域的融合蛋白对于靶序列的高度特异性且有效率的碱基编辑也是有用的。在一个实施方案中,嵌合Cas12融合蛋白含有插入在Cas12多肽内的异源催化结构域(例如,腺苷脱氨酶、胞苷脱氨酶,或腺苷脱氨酶和胞苷脱氨酶)。在一些实施方案中,融合蛋白包含插入在Cas12内的腺苷脱氨酶结构域和胞苷脱氨酶结构域。在一些实施方案中,腺苷脱氨酶融合在Cas12内,并且胞苷脱氨酶融合至C端。在一些实施方案中,腺苷脱氨酶融合在Cas12内,并且胞苷脱氨酶融合至N端。在一些实施方案中,胞苷脱氨酶融合在Cas12内,并且腺苷脱氨酶融合至C端。在一些实施方案中,胞苷脱氨酶融合在Cas12内,并且腺苷脱氨酶融合至N端。具有腺苷脱氨酶和胞苷脱氨酶和Cas12的融合蛋白的示例性结构提供如下:

[0506] NH2-[Cas12(腺苷脱氨酶)]-[胞苷脱氨酶]-COOH;

[0507] NH2-[胞苷脱氨酶]-[Cas12(腺苷脱氨酶)]-COOH;

[0508] NH2-[Cas12(胞苷脱氨酶)]-[腺苷脱氨酶]-COOH;或

[0509] NH2-[腺苷脱氨酶]-[Cas12(胞苷脱氨酶)]-COOH;

[0510] 在一些实施方案中,上方通用架构中使用的“-”指示存在任选的接头。

[0511] 在各种实施方案中,催化结构域具有DNA修饰活性(例如脱氨酶活性),诸如腺苷脱氨酶活性。在一些实施方案中,腺苷脱氨酶是TadA(例如,TadA*7.10)。在一些实施方案中,TadA是TadA*8。在一些实施方案中,TadA*8融合在Cas12内,并且胞苷脱氨酶融合至C端。在一些实施方案中,TadA*8融合在Cas12内,并且胞苷脱氨酶融合至N端。在一些实施方案中,胞苷脱氨酶融合在Cas12内,并且TadA*8融合至C端。在一些实施方案中,胞苷脱氨酶融合在Cas12内,并且TadA*8融合至N端。具有TadA*8和胞苷脱氨酶和Cas12的融合蛋白的示例性结构提供如下:

[0512] N-[Cas12(TadA*8)]-[胞苷脱氨酶]-C;

[0513] N-[胞苷脱氨酶]-[Cas12(TadA*8)]-C;

[0514] N-[Cas12(胞苷脱氨酶)]-[TadA*8]-C;或

[0515] N-[TadA*8]-[Cas12(胞苷脱氨酶)]-C。

[0516] 在一些实施方案中,上方通用架构中使用的“-”指示存在任选的接头。

[0517] 在其他实施方案中,融合蛋白含有一个或多个催化结构域。在其他实施方案中,一个或多个催化结构域中的至少一个催化结构域被插入在Cas12多肽内或融合在Cas12 N端或C端处。在其他实施方案中,一个或多个催化结构域中的至少一个催化结构域插入在Cas12多肽的环、α螺旋区、非结构化部分或溶剂可及部分内。在其他实施方案中,Cas12多肽是Cas12a、Cas12b、Cas12c、Cas12d、Cas12e、Cas12g、Cas12h、Cas12i或Cas12j/CasΦ。在其他实施方案中,Cas12多肽与外村尚芽孢杆菌(Bacillus hisashii)Cas12b、嗜热淀粉芽孢杆菌(Bacillus thermoamylovorans)Cas12b、芽孢杆菌属V3-13 Cas12b,或嗜酸脂环杆菌(Alicyclob acillusacidiphilu)Cas12b(SEQ ID NO:259)具有至少约85%的氨基酸序列同一性。在其他实施方案中,Cas12多肽与外村尚芽孢杆菌Cas12b(SEQ ID NO:260)、嗜热淀粉芽孢杆菌Cas12b、芽孢杆菌属V3-13 Cas12b或嗜酸脂环杆菌Cas12b具有至少约90%的氨基酸序列同一性。在其他实施方案中,Cas12多肽与外村尚芽孢杆菌Cas12b、嗜热淀粉芽孢杆菌Cas12b(SEQ ID NO:265)、芽孢杆菌属V3-13 Cas12b(SEQ ID NO:264)或嗜酸脂环杆菌Cas12b具有至少约95%的氨基酸序列同一性。在其他实施方案中,Cas12多肽含有或基本上由以下组成:外村尚芽孢杆菌Cas12b、嗜热淀粉芽孢杆菌Cas12b、芽孢杆菌属V3-13 Cas12b或嗜酸脂环杆菌Cas12b的片段。在实施方案中,Cas12多肽含有BvCas12b(V4),其在一些实施方案中表达为5’mRNA帽---5’UTR---bhCas12b---停止序列---3’UTR---120polyA尾(SEQ ID NO:261-263)。

[0518] 在其他实施方案中,催化结构域插入在BhCas12b或者Cas12a、Cas12c、Cas12d、Cas12e、Cas12g、Cas12h、Cas12i或Cas12j/CasΦ的相应氨基酸残基的氨基酸位置153-154、255-256、306-307、980-981、1019-1020、534-535、604-605或344-345之间。在其他实施方案中,催化结构域插入在BhCas12b的氨基酸P153和S154之间。在其他实施方案中,催化结构域插入在BhCas12b的氨基酸K255和E256之间。在其他实施方案中,催化结构域插入在BhCas12b的氨基酸D980和G981之间。在其他实施方案中,催化结构域插入在BhCas12b的氨基酸K1019和L1020之间。在其他实施方案中,催化结构域插入在BhCas12b的氨基酸F534和P535之间。在其他实施方案中,催化结构域插入在BhCas12b的氨基酸K604和G605之间。在其他实施方案中,催化结构域插入在BhCas12b的氨基酸H344和F345之间。在其他实施方案中,催化结构域插入在BvCas12b或者Cas12a、Cas12c、Cas12d、Cas12e、Cas12g、Cas12h、Cas12i或Cas12j/CasΦ的相应氨基酸残基的氨基酸位置147和148、248和249、299和300、991和992或1031和1032之间。在其他实施方案中,催化结构域插入在BvCas12b的氨基酸P147和D148之间。在其他实施方案中,催化结构域插入在BvCas12b的氨基酸G248和G249之间。在其他实施方案中,催化结构域插入在BvCas12b的氨基酸P299和E300之间。在其他实施方案中,催化结构域插入在BvCas12b的氨基酸G991和E992之间。在其他实施方案中,催化结构域插入在BvCas12b的氨基酸K1031和M1032之间。在其他实施方案中,催化结构域插入在AaCas12b或者Cas12a、Cas12c、Cas12d、Cas12e、Cas12g、Cas12h、Cas12i或Cas12j/CasΦ的相应氨基酸残基的氨基酸位置157和158、258和259、310和311、1008和1009或1044和1045之间。在其他实施方案中,催化结构域插入在AaCas12b的氨基酸P157和G158之间。在其他实施方案中,催

化结构域插入在AaCas12b的氨基酸V258和G259之间。在其他实施方案中,催化结构域插入在AaCas12b的氨基酸D310和P311之间。在其他实施方案中,催化结构域插入在AaCas12b的氨基酸G1008和E1009之间。在其他实施方案中,催化结构域插入在AaCas12b的氨基酸G1044和K1045之间。

[0519] 在其他实施方案中,融合蛋白含有核定位信号(例如,双组分(bipartite)核定位信号)。在其他实施方案中,核定位信号的氨基酸序列是MAPKKKRKVGIHGVPAA(SEQ ID NO:1313)。在以上方面的其他实施方案中,核定位信号由以下序列编码:

[0520] ATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGA GTCCCAGCAGCC(SEQ ID NO:1314)。在其他实施方案中,Cas12b多肽含有使RuvC结构域的催化活性沉默的突变。在其他实施方案中,Cas12b多肽含有D574A、D829A和/或D952A突变。在其他实施方案中,融合蛋白进一步含有标签(例如,流感血凝素标签)。

[0521] 在一些实施方案中,融合蛋白包含具有内部融合的核碱基编辑结构域(例如,全部或部分脱氨酶结构域,例如腺苷脱氨酶结构域)的napDNAbp结构域(例如,Cas12衍生的结构域)。在一些实施方案中,napDNAbp是Cas12b。在一些实施方案中,碱基编辑器包含BhCas12b结构域,该结构域具有插入在下表4中提供的基因座处的内部融合的TadA*8结构域。

[0522] 表4:Cas12b蛋白中的插入基因座

[0523]

BhCas12b	插入位点	插入在aa之间
位置1	153	PS
位置2	255	KE
位置3	306	DE
位置4	980	DG
位置5	1019	KL
位置6	534	FP
位置7	604	KG
位置8	344	HF
BvCas12b	插入位点	插入在aa之间
位置1	147	PD
位置2	248	GG
位置3	299	PE
位置4	991	GE
位置5	1031	KM
AaCas12b	插入位点	插入在aa之间
位置1	157	PG
位置2	258	VG
位置3	310	DP
位置4	1008	GE
位置5	1044	GK

[0524] 作为非限制性实例,腺苷脱氨酶(例如,TadA*8.13)可以被插入到BhCas12b中以产生有效编辑核酸序列的融合蛋白(例如,TadA*8.13-BhCas12b)。

[0525] 在一些实施方案中,本文所述的碱基编辑系统是具有插入到Cas9中的TadA的ABE。具有插入到Cas9中的TadA的有关ABE的多肽序列在所附序列表中作为SEQ ID NO:1315-1360提供。

[0526] 在一些实施方案中,生成腺苷脱氨酶碱基编辑器以将TadA或其变体插入到Cas9多肽中鉴定的位置处。

[0527] 示例性但非限制性的融合蛋白描述于国际PCT申请号PCT/US2020/016285和美国临时申请号62/852,228和62/852,224中,它们的内容以引用方式整体并入本文。

[0528] A到G的编辑

[0529] 在一些实施方案中,本文所述的碱基编辑器包含腺苷脱氨酶结构域。碱基编辑器的此种腺苷脱氨酶结构域可以促进将腺嘌呤(A)核碱基编辑为鸟嘌呤(G)核碱基,这通过使A脱氨从而形成表现出G的碱基配对特性的肌苷(I)来实现。腺苷脱氨酶能够使脱氧核糖核酸(DNA)中的脱氧腺苷残基的腺嘌呤脱氨(即,去除胺基)。在一些实施方案中,A到G碱基编辑器进一步包含肌苷碱基切除修复抑制剂,例如,尿嘧啶糖苷酶抑制剂(UGI)结构域或无催化活性的肌苷特异性核酸酶。不受任何特定理论的束缚,UGI结构域或无催化活性的肌苷特异性核酸酶可以抑制或阻止经脱氨的腺苷残基(例如肌苷)的碱基切除修复,这可以提高碱基编辑器的活性或效率。

[0530] 包含腺苷脱氨酶的碱基编辑器可以作用于任何多核苷酸,包括DNA、RNA和DNA-RNA杂合体。在某些实施方案中,包含腺苷脱氨酶的碱基编辑器可以使包含RNA的多核苷酸的靶A脱氨。例如,碱基编辑器可以包含能够使RNA多核苷酸和/或DNA-RNA杂合多核苷酸的靶A脱氨的腺苷脱氨酶结构域。在一个实施方案中,掺入到碱基编辑器中的腺苷脱氨酶包含作用于RNA(ADAR,例如ADAR1或ADAR2)或tRNA(ADAT)的全部或部分腺苷脱氨酶。包含腺苷脱氨酶结构域的碱基编辑器也能够使DNA多核苷酸的A核碱基脱氨。在一个实施方案中,碱基编辑器的腺苷脱氨酶结构域包含全部或部分ADAT,该ADAT包含允许ADAT使DNA中的靶A脱氨的一个或多个突变。例如,碱基编辑器可以包含全部或部分的来自大肠埃希氏菌的ADAT(EcTadA),其包含以下突变中的一者或多者:D108N、A106V、D147Y、E155V、L84F、H123Y、I156F,或在另一个腺苷脱氨酶中的相应突变。示例性ADAT同系物多肽序列在序列表中作为SEQ ID NO:1363-1370提供。

[0531] 腺苷脱氨酶可以源自自任何合适的生物体(例如,大肠埃希氏菌)。在一些实施方案中,腺苷脱氨酶来自原核生物。在一些实施方案中,腺苷脱氨酶来自细菌。在一些实施方案中,腺苷脱氨酶来自大肠埃希氏菌、金黄色葡萄球菌、伤寒沙门氏菌、腐败希瓦氏菌、流感嗜血杆菌、新月柄杆菌或枯草芽孢杆菌。在一些实施方案中,腺苷脱氨酶来自大肠埃希氏菌。在一些实施方案中,腺嘌呤脱氨酶是天然存在的腺苷脱氨酶,该腺苷脱氨酶包括对应于本文提供的突变中的任何突变(例如,ecTadA中的突变)的一个或多个突变。任何同源蛋白中的相应残基可以通过例如序列比对和同源残基的确定来鉴定。可以相应地产生对应于本文描述的突变中的任何突变(例如,在ecTadA中鉴定的突变中的任何突变)的任何天然存在的腺苷脱氨酶(例如,与ecTadA具有同源性)中的突变。

[0532] 在一些实施方案中,腺苷脱氨酶包含与本文提供的腺苷脱氨酶中的任何腺苷脱氨

酶中阐述的氨基酸序列中的任何一个氨基酸序列至少60％、至少65％、至少70％、至少75％、至少80％、至少85％、至少90％、至少95％、至少96％、至少97％、至少98％、至少99％或至少99.5％同一的氨基酸序列。应当理解，本文提供的腺苷脱氨酶可以包含一个或多个突变(例如，本文提供的突变中的任何突变)。本公开提供了具有一定百分比同一性加上本文所述的突变或其组合中的任何突变或组合的任何脱氨酶结构域。在一些实施方案中，腺苷脱氨酶包含与参考序列或本文提供的腺苷脱氨酶中的任何腺苷脱氨酶相比具有1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、21个、24个、25个、26个、27个、28个、29个、30个、31个、32个、33个、34个、35个、36个、37个、38个、39个、40个、41个、42个、43个、44个、45个、46个、47个、48个、49个、50个或更多个突变的氨基酸序列。在一些实施方案中，腺苷脱氨酶包含与本领域已知或本文描述的氨基酸序列中的任一种氨基酸序列相比具有至少5个、至少10个、至少15个、至少20个、至少25个、至少30个、至少35个、至少40个、至少45个、至少50个、至少60个、至少70个、至少80个、至少90个、至少100个、至少110个、至少120个、至少130个、至少140个、至少150个、至少160个或至少170个同一的连续氨基酸残基的氨基酸序列。

[0533] 应当理解，本文提供的突变中的任何突变(例如，基于TadA参考序列)可以被引入到其他腺苷脱氨酶，诸如大肠埃希氏菌TadA(ecTadA)、金黄色葡萄球菌TadA(saTadA)或其他腺苷脱氨酶(例如，细菌腺苷脱氨酶)中。对于本领域技术人员显而易见的是，可以类似地比对另外的脱氨酶以鉴定可以如本文提供的那样突变的同源氨基酸残基。因此，在TadA参考序列中鉴定出的突变中的任何突变可以被制作在具有同源氨基酸残基的其他腺苷脱氨酶(例如，ecTada)中。还应理解，本文提供的突变中的任何突变可以单独或以任何组合被制作在TadA参考序列或另一种腺苷脱氨酶中。

[0534] 在一些实施方案中，腺苷脱氨酶包含TadA参考序列中的D108X突变，或另一种腺苷脱氨酶中的相应突变，其中X表示除野生型腺苷脱氨酶中相应氨基酸之外的任何氨基酸。在一些实施方案中，腺苷脱氨酶包含TadA参考序列中的D108G、D108N、D108V、D108A或D108Y突变，或另一种腺苷脱氨酶中的相应突变。然而，应当理解，可以类似地比对另外的脱氨酶以鉴定可以如本文提供的那样突变的同源氨基酸残基。

[0535] 在一些实施方案中，腺苷脱氨酶包含TadA参考序列中的A106X突变，或另一种腺苷脱氨酶中的相应突变，其中X表示除野生型腺苷脱氨酶中相应氨基酸之外的任何氨基酸。在一些实施方案中，腺苷脱氨酶包含TadA参考序列中的A106V突变，或另一种腺苷脱氨酶(例如，ecTadA)中的相应突变。

[0536] 在一些实施方案中，腺苷脱氨酶包含TadA参考序列中的E155X突变，或另一种腺苷脱氨酶中的相应突变，其中存在X表示除野生型腺苷脱氨酶中相应氨基酸之外的任何氨基酸。在一些实施方案中，腺苷脱氨酶包含TadA参考序列中的E155D、E155G或E155V突变，或另一种腺苷脱氨酶(例如，ecTadA)中的相应突变。

[0537] 在一些实施方案中，腺苷脱氨酶包含TadA参考序列中的D147X突变，或另一种腺苷脱氨酶中的相应突变，其中存在X表示除野生型腺苷脱氨酶中相应氨基酸之外的任何氨基酸。在一些实施方案中，腺苷脱氨酶包含TadA参考序列中的D147Y突变，或另一种腺苷脱氨酶(例如，ecTadA)中的相应突变。

[0538] 在一些实施方案中，腺苷脱氨酶包含TadA参考序列中的A106X、E155X或D147X突

变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应突变,其中X表示除野生型腺苷脱氨酶中相应氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含E155D、E155G或E155V突变。在一些实施方案中,腺苷脱氨酶包含D147Y。

[0539] 还应理解,本文提供的突变中的任何突变可以单独或以任何组合被制作在ecTadA或另一种腺苷脱氨酶中。例如,腺苷脱氨酶可以含有TadA参考序列中的D108N、A106V、E155V和/或D147Y突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应突变。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的以下突变组(突变组由";"分隔),或另一种腺苷脱氨酶中的相应突变:D108N和A106V;D108N和E155V;D108N和D147Y;A106V和E155V;A106V和D147Y;E155V和D147Y;D108N、A106V和E155V;D108N、A106V和D147Y;D108N、E155V和D147Y;A106V、E155V、D147Y;以及D108N、A106V、E155V和D147Y。然而,应当理解,本文提供的相应突变的任何组合可以被制作在腺苷脱氨酶(例如,ecTadA)中。

[0540] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H8X、T17X、L18X、W23X、L34X、W45X、R51X、A56X、E59X、E85X、M94X、I95X、V102X、F104X、A106X、R107X、D108X、K110X、M118X、N127X、A138X、F149X、M151X、R153X、Q154X、I156X和/或K157X突变中的一者或多者,或另一种腺苷脱氨酶中的一个或多个相应突变,其中X的存在表示除野生型腺苷脱氨酶中相应氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H8Y、T17S、L18E、W23L、L34S、W45L、R51H、A56E,或A56S、E59G、E85K,或E85G、M94L、I95L、V102A、F104L、A106V、R107C,或R107H,或R107P,D108G,或D108N,或D108V,或D108A,或D108Y、K110I、M118K、N127S、A138V、F149Y、M151V、R153C、Q154L、I156D和/或K157R突变中的一者或多者,或另一种腺苷脱氨酶中的一个或多个相应突变。

[0541] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H8X、D108X和/或N127X突变中的一者或多者,或另一种腺苷脱氨酶中的一个或多个相应突变,其中X表示存在任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H8Y、D108N和/或N127S突变中的一者或多者,或另一种腺苷脱氨酶中的一个或多个相应突变。

[0542] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H8X、R26X、M61X、L68X、M70X、A106X、D108X、A109X、N127X、D147X、R152X、Q154X、E155X、K161X、Q163X和/或T166X突变中的一者或多者,或另一种腺苷脱氨酶中的一个或多个相应突变,其中X表示存在除野生型腺苷脱氨酶中相应氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H8Y、R26W、M61I、L68Q、M70V、A106T、D108N、A109T、N127S、D147Y、R152C、Q154H或Q154R、E155G或E155V或E155D、K161Q、Q163H和/或T166P突变中的一者或多者,或另一种腺苷脱氨酶中的一个或多个相应突变。

[0543] 在一些实施方案中,腺苷脱氨酶包含一个、两个、三个、四个、五个或六个选自由TadA参考序列中的H8X、D108X、N127X、D147X、R152X和Q154X组成的组的突变,或另一种腺苷脱氨酶(例如,ecTadA)中的一个或多个相应突变,其中X表示存在除野生型腺苷脱氨酶中相应氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含一个、两个、三个、四个、五个、六个、七个或八个选自由TadA参考序列中的H8X、M61X、M70X、D108X、N127X、Q154X、E155X和Q163X组成的组的突变,或另一种腺苷脱氨酶(例如,ecTadA)中的一个或多个相应突变,其中X表示存在除野生型腺苷脱氨酶中相应氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含一个、两个、三个、四个或五个选自由TadA参考序列中的H8X、D108X、

N127X、E155X和T166X组成的组的突变,或另一种腺苷脱氨酶(例如,ecTadA)中的一个或多个相应突变,其中X表示存在除野生型腺苷脱氨酶中相应氨基酸之外的任何氨基酸。

[0544] 在一些实施方案中,腺苷脱氨酶包含一个、两个、三个、四个、五个或六个选自由H8X、A106X和D108X组成的组的突变,或另一种腺苷脱氨酶中的一个或多个相应突变,其中X表示存在除野生型腺苷脱氨酶中相应氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含一个、两个、三个、四个、五个、六个、七个或八个选自由H8X、R26X、L68X、D108X、N127X、D147X和E155X组成的组的突变,或另一种腺苷脱氨酶中的一个或多个相应突变,其中X表示存在除野生型腺苷脱氨酶中相应氨基酸之外的任何氨基酸。

[0545] 在一些实施方案中,腺苷脱氨酶包含一个、两个、三个、四个、五个、六个或七个选自由TadA参考序列中的H8X、R126X、L68X、D108X、N127X、D147X和E155X组成的组的突变,或另一种腺苷脱氨酶中的一个或多个相应突变,其中X表示存在除野生型腺苷脱氨酶中相应氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含一个、两个、三个、四个或五个选自由TadA参考序列中的H8X、D108X、A109X、N127X和E155X组成的组的突变,或另一种腺苷脱氨酶中的一个或多个相应突变,其中X表示存在除野生型腺苷脱氨酶中相应氨基酸之外的任何氨基酸。

[0546] 在一些实施方案中,腺苷脱氨酶包含一个、两个、三个、四个、五个或六个选自由TadA参考序列中的H8Y、D108N、N127S、D147Y、R152C和Q154H组成的组的突变,或另一种腺苷脱氨酶(例如,ecTadA)中的一个或多个相应突变。在一些实施方案中,腺苷脱氨酶包含一个、两个、三个、四个、五个、六个、七个或八个选自由TadA参考序列中的H8Y、M61I、M70V、D108N、N127S、Q154R、E155G和Q163H组成的组的突变,或另一种腺苷脱氨酶(例如,ecTadA)中的一个或多个相应突变。在一些实施方案中,腺苷脱氨酶包含一个、两个、三个、四个或五个选自由TadA参考序列中的H8Y、D108N、N127S、E155V和T166P组成的组的突变,或另一种腺苷脱氨酶(例如,ecTadA)中的一个或多个相应突变。在一些实施方案中,腺苷脱氨酶包含一个、两个、三个、四个、五个或六个选自由TadA参考序列中的H8Y、A106T、D108N、N127S、E155D和K161Q组成的组的突变,或另一种腺苷脱氨酶(例如,ecTadA)中的一个或多个相应突变。在一些实施方案中,腺苷脱氨酶包含一个、两个、三个、四个、五个、六个、七个或八个选自由TadA参考序列中的H8Y、R26W、L68Q、D108N、N127S、D147Y和E155V组成的组的突变,或另一种腺苷脱氨酶(例如,ecTadA)中的一个或多个相应突变。在一些实施方案中,腺苷脱氨酶包含一个、两个、三个、四个或五个选自由TadA参考序列中的H8Y、D108N、A109T、N127S和E155G组成的组的突变,或另一种腺苷脱氨酶(例如,ecTadA)中的一个或多个相应突变。

[0547] 在一些实施方案中,腺苷脱氨酶包含中的一者或多者或另一种腺苷脱氨酶中的一个或多个相应突变。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的D108N、D108G或D108V突变,或另一种腺苷脱氨酶中的相应突变。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的A106V和D108N突变,或另一种腺苷脱氨酶中的相应突变。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的R107C和D108N突变,或另一种腺苷脱氨酶中的相应突变。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H8Y、D108N、N127S、D147Y和Q154H突变,或另一种腺苷脱氨酶中的相应突变。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H8Y、D108N、N127S、D147Y和E155V突变,或另一种腺苷脱氨酶中的相应突变。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的D108N、D147Y和E155V突变,或另一

种腺苷脱氨酶中的相应突变。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H8Y、D108N和N127S突变,或另一种腺苷脱氨酶中的相应突变。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的A106V、D108N、D147Y和E155V突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应突变。

[0548] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的S2X、H8X、I49X、L84X、H123X、N127X、I156X和/或K160X突变中的一者或多者,或另一种腺苷脱氨酶中的一个或多个相应突变,其中存在X表示除野生型腺苷脱氨酶中相应氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的S2A、H8Y、I49F、L84F、H123Y、N127S、I156F和/或K160S突变中的一者或多者,或另一种腺苷脱氨酶(例如,ecTadA)中的一个或多个相应突变。

[0549] 在一些实施方案中,腺苷脱氨酶包括L84X突变腺苷脱氨酶,其中X表示除野生型腺苷脱氨酶中相应氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的L84F突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应突变。

[0550] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H123X突变,或另一种腺苷脱氨酶中的相应突变,其中X表示除野生型腺苷脱氨酶中相应氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H123Y突变,或另一种腺苷脱氨酶中的相应突变。

[0551] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的I156X突变,或另一种腺苷脱氨酶中的相应突变,其中X表示除野生型腺苷脱氨酶中相应氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的I156F突变,或另一种腺苷脱氨酶中的相应突变。

[0552] 在一些实施方案中,腺苷脱氨酶包含一个、两个、三个、四个、五个、六个或七个选自由TadA参考序列中的L84X、A106X、D108X、H123X、D147X、E155X和I156X组成的组的突变,或另一种腺苷脱氨酶中的一个或多个相应突变,其中X表示存在除野生型腺苷脱氨酶中相应氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含一个、两个、三个、四个、五个或六个选自由TadA参考序列中的S2X、I49X、A106X、D108X、D147X和E155X组成的组的突变,或另一种腺苷脱氨酶中的一个或多个相应突变,其中X表示存在除野生型腺苷脱氨酶中相应氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含一个、两个、三个、四个或五个选自由TadA参考序列中的H8X、A106X、D108X、N127X和K160X组成的组的突变,或另一种腺苷脱氨酶中的一个或多个相应突变,其中X表示存在除野生型腺苷脱氨酶中相应氨基酸之外的任何氨基酸。

[0553] 在一些实施方案中,腺苷脱氨酶包含一个、两个、三个、四个、五个、六个或七个选自由TadA参考序列中的L84F、A106V、D108N、H123Y、D147Y、E155V和I156F组成的组的突变,或另一种腺苷脱氨酶中的一个或多个相应突变。在一些实施方案中,腺苷脱氨酶包含一个、两个、三个、四个、五个或六个选自由TadA参考序列中的S2A、I49F、A106V、D108N、D147Y和E155V组成的组的突变。

[0554] 在一些实施方案中,腺苷脱氨酶包含一个、两个、三个、四个或五个选自由TadA参考序列中的H8Y、A106T、D108N、N127S和K160S组成的组的突变,或另一种腺苷脱氨酶中的一个或多个相应突变。

[0555] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的E25X、R26X、R107X、A142X和/或A143X突变中的一者或多者,或另一种腺苷脱氨酶中的一个或多个相应突变,其中存在X表示除野生型腺苷脱氨酶中相应氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的E25M、E25D、E25A、E25R、E25V、E25S、E25Y、R26G、R26N、R26Q、R26C、R26L、R26K、R107P、R107K、R107A、R107N、R107W、R107H、R107S、A142N、A142D、A142G、A143D、A143G、A143E、A143L、A143W、A143M、A143S、A143Q和/或A143R突变中的一者或多者,或另一种腺苷脱氨酶中的一个或多个相应突变。在一些实施方案中,腺苷脱氨酶包含本文所述的对应于TadA参考序列的突变中的一个或多个突变,或另一种腺苷脱氨酶中的一个或多个相应突变。

[0556] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的E25X突变,或另一种腺苷脱氨酶中的相应突变,其中X表示除野生型腺苷脱氨酶中相应氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的E25M、E25D、E25A、E25R、E25V、E25S或E25Y突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应突变。

[0557] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的R26X突变,或另一种腺苷脱氨酶中的相应突变,其中X表示除野生型腺苷脱氨酶中相应氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的R26G、R26N、R26Q、R26C、R26L或R26K突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应突变。

[0558] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的R107X突变,或另一种腺苷脱氨酶中的相应突变,其中X表示除野生型腺苷脱氨酶中相应氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的R107P、R107K、R107A、R107N、R107W、R107H或R107S突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应突变。

[0559] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的A142X突变,或另一种腺苷脱氨酶中的相应突变,其中X表示除野生型腺苷脱氨酶中相应氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的A142N、A142D、A142G突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应突变。

[0560] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的A143X突变,或另一种腺苷脱氨酶中的相应突变,其中X表示除野生型腺苷脱氨酶中相应氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的A143D、A143G、A143E、A143L、A143W、A143M、A143S、A143Q和/或A143R突变,或另一种腺苷脱氨酶(例如,ecTadA)中的相应突变。

[0561] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H36X、N37X、P48X、I49X、R51X、M70X、N72X、D77X、E134X、S146X、Q154X、K157X和/或K161X突变中的一者或多者,或另一种腺苷脱氨酶中的一个或多个相应突变,其中存在X表示除野生型腺苷脱氨酶中相应氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H36L、N37T、N37S、P48T、P48L、I49V、R51H、R51L、M70L、N72S、D77G、E134G、S146R、S146C、Q154H、K157N和/或K161T突变中的一者或多者,或另一种腺苷脱氨酶(例如,ecTadA)中的一个或多个相应突变。

[0562] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H36X突变,或另一种腺苷脱氨酶中的相应突变,其中X表示除野生型腺苷脱氨酶中相应氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的H36L突变,或另一种腺苷脱氨酶中的

相应突变。

[0563] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的N37X突变,或另一种腺苷脱氨酶中的相应突变,其中X表示除野生型腺苷脱氨酶中相应氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的N37T或N37S突变,或另一种腺苷脱氨酶中的相应突变。

[0564] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的P48X突变,或另一种腺苷脱氨酶中的相应突变,其中X表示除野生型腺苷脱氨酶中相应氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的P48T或P48L突变,或另一种腺苷脱氨酶中的相应突变。

[0565] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的R51X突变,或另一种腺苷脱氨酶中的相应突变,其中X表示除野生型腺苷脱氨酶中相应氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的R51H或R51L突变,或另一种腺苷脱氨酶中的相应突变。

[0566] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的S146X突变,或另一种腺苷脱氨酶中的相应突变,其中X表示除野生型腺苷脱氨酶中相应氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的S146R或S146C突变,或另一种腺苷脱氨酶中的相应突变。

[0567] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的K157X突变,或另一种腺苷脱氨酶中的相应突变,其中X表示除野生型腺苷脱氨酶中相应氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的K157N突变,或另一种腺苷脱氨酶中的相应突变。

[0568] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的P48X突变,或另一种腺苷脱氨酶中的相应突变,其中X表示除野生型腺苷脱氨酶中相应氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的P48S、P48T或P48A突变,或另一种腺苷脱氨酶中的相应突变。

[0569] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的A142X突变,或另一种腺苷脱氨酶中的相应突变,其中X表示除野生型腺苷脱氨酶中相应氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的A142N突变,或另一种腺苷脱氨酶中的相应突变。

[0570] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的W23X突变,或另一种腺苷脱氨酶中的相应突变,其中X表示除野生型腺苷脱氨酶中相应氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的W23R或W23L突变,或另一种腺苷脱氨酶中的相应突变。

[0571] 在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的R152X突变,或另一种腺苷脱氨酶中的相应突变,其中X表示除野生型腺苷脱氨酶中相应氨基酸之外的任何氨基酸。在一些实施方案中,腺苷脱氨酶包含TadA参考序列中的R152P或R52H突变,或另一种腺苷脱氨酶中的相应突变。

[0572] 在一个实施方案中,腺苷脱氨酶可以包含突变H36L、R51L、L84F、A106V、D108N、H123Y、S146C、D147Y、E155V、I156F和K157N。在一些实施方案中,腺苷脱氨酶包含以下相对

于TadA参考序列的突变组合,其中组合的每个突变由“_”分隔,并且每个突变组合均在括号之间:

[0573] (A106V_D108N)、

[0574] (R107C_D108N)、

[0575] (H8Y_D108N_N127S_D147Y_Q154H)、

[0576] (H8Y_D108N_N127S_D147Y_E155V)、

[0577] (D108N_D147Y_E155V)、

[0578] (H8Y_D108N_N127S)、

[0579] (H8Y_D108N_N127S_D147Y_Q154H)、

[0580] (A106V_D108N_D147Y_E155V)、

[0581] (D108Q_D147Y_E155V)、

[0582] (D108M_D147Y_E155V)、

[0583] (D108L_D147Y_E155V)、

[0584] (D108K_D147Y_E155V)、

[0585] (D108I_D147Y_E155V)、

[0586] (D108F_D147Y_E155V)、

[0587] (A106V_D108N_D147Y)、

[0588] (A106V_D108M_D147Y_E155V)、

[0589] (E59A_A106V_D108N_D147Y_E155V)、

[0590] (E59A催化失效_A106V_D108N_D147Y_E155V)、(L84F_A106V_D108N_H123Y_D147Y_E155V_I156Y)、(L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)、(D103A_D104N)、

[0591] (G22P_D103A_D104N)、

[0592] (D103A_D104N_S138A)、

[0593] (R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F)、

[0594] (E25G_R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F)、

[0595] (E25D_R26G_L84F_A106V_R107K_D108N_H123Y_A142N_A143G_D147Y_E155V_I156F)、

[0596] (R26Q_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F)、

[0597] (E25M_R26G_L84F_A106V_R107P_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F)、

[0598] (R26C_L84F_A106V_R107H_D108N_H123Y_A142N_D147Y_E155V_I156F)、

[0599] (L84F_A106V_D108N_H123Y_A142N_A143L_D147Y_E155V_I156F)、

[0600] (R26G_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F)、

[0601] (E25A_R26G_L84F_A106V_R107N_D108N_H123Y_A142N_A143E_D147Y_E155V_I156F)、

[0602] (R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F)、

[0603] (A106V_D108N_A142N_D147Y_E155V)、

[0604] (R26G_A106V_D108N_A142N_D147Y_E155V)、

[0605] (E25D_R26G_A106V_R107K_D108N_A142N_A143G_D147Y_E155V)、

[0606] (R26G_A106V_D108N_R107H_A142N_A143D_D147Y_E155V)、

[0607] (E25D_R26G_A106V_D108N_A142N_D147Y_E155V)、

[0608] (A106V_R107K_D108N_A142N_D147Y_E155V)、

[0609] (A106V_D108N_A142N_A143G_D147Y_E155V)、

[0610] (A106V_D108N_A142N_A143L_D147Y_E155V)、

[0611] (H36L_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N)、

[0612] (N37T_P48T_M70L_L84F_A106V_D108N_H123Y_D147Y_I49V_E155V_I156F)、

[0613] (N37S_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K161T)、

[0614] (H36L_L84F_A106V_D108N_H123Y_D147Y_Q154H_E155V_I156F)、

[0615] (N72S_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F)、

[0616] (H36L_P48L_L84F_A106V_D108N_H123Y_E134G_D147Y_E155V_I156F)、

[0617] (H36L_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K157N)

[0618] (H36L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F)、

[0619] (L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F_K161T)、

[0620] (N37S_R51H_D77G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)、

[0621] (R51L_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K157N)、

[0622] (D24G_Q71R_L84F_H96L_A106V_D108N_H123Y_D147Y_E155V_I156F_K160E)、

[0623] (H36L_G67V_L84F_A106V_D108N_H123Y_S146T_D147Y_E155V_I156F)、

[0624] (Q71L_L84F_A106V_D108N_H123Y_L137M_A143E_D147Y_E155V_I156F)、

[0625] (E25G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_Q159L)、

[0626] (L84F_A91T_F104I_A106V_D108N_H123Y_D147Y_E155V_I156F)、

[0627] (N72D_L84F_A106V_D108N_H123Y_G125A_D147Y_E155V_I156F)、

[0628] (P48S_L84F_S97C_A106V_D108N_H123Y_D147Y_E155V_I156F)、

[0629] (W23G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)、

[0630] (D24G_P48L_Q71R_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_Q159L)、

[0631] (L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F)、

[0632] (H36L_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F_K157N)、

[0633] (N37S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_K161T)、

[0634] (L84F_A106V_D108N_D147Y_E155V_I156F)、

[0635] (R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K161T)、

[0636] (L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K161T)、

[0637] (L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K160E_K161T)、

[0638] (L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K160E)、

[0639] (R74Q_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)、

[0640] (R74A_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)、

[0641] (L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)、

[0642] (R74Q_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F)、

[0643] (L84F_R98Q_A106V_D108N_H123Y_D147Y_E155V_I156F)、

[0644] (L84F_A106V_D108N_H123Y_R129Q_D147Y_E155V_I156F)、

[0645] (P48S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F)、

[0646] (P48S_A142N)、

[0647] (P48T_I49V_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_L157N)、

[0648] (P48T_I49V_A142N)、

[0649] (H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N)、

[0650] (H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S146C_A142N_D147Y_E155V_I156F(H36L_P48T_I49V_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N)、

[0651] (H36L_P48T_I49V_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F_K157N)、

[0652] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N)、

[0653] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F_K157N)、

[0654] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_A142N_D147Y_E155V_I156F_K157N)、

[0655] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N)、

[0656] (W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N)、

[0657] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F_K161T)、

[0658] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152H_E155V_I156F_K157N)、

[0659] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F_K157N)、

[0660] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F_K157N)、

[0661] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142A_S146C_D147Y_E155V_I156F_K157N)、

[0662] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142A_S146C_D147Y_R152P_E155V_I156F_K157N)、

[0663] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F_K161T)、

[0664] (W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F_K157N)、

[0665] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_R152P_E155V_

I156F_K157N)。

[0666] 在一些实施方案中,TadA脱氨酶是TadA变体。在一些实施方案中,TadA变体是TadA*7.10。在特定实施方案中,融合蛋白包含单个TadA*7.10结构域(例如,作为单体提供)。在其他实施方案中,融合蛋白包含能够形成异二聚体的TadA*7.10和TadA(wt)。在一个实施方案中,本发明的融合蛋白包含与TadA*7.10连接的野生型TadA,该TadA*7.10与Cas9切口酶连接。

[0667] 在一些实施方案中,TadA*7.10包含至少一个改变。在一些实施方案中,腺苷脱氨酶包含以下序列中的改变:

[0668] TadA*7.10

[0669] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD(SEQ ID NO:8)

[0670] 在一些实施方案中,TadA*7.10包含氨基酸82和/或166处的改变。在特定实施方案中,TadA*7.10包含以下改变中的一者或多者:Y147T、Y147R、Q154S、Y123H、V82S、T166R,和/或Q154R。在其他实施方案中,TadA*7.10的变体包含选自由以下组成的组的改变的组合:Y147T+Q154R;Y147T+Q154S;Y147R+Q154S;V82S+Q154S;V82S+Y147R;V82S+Q154R;V82S+Y123H;I76Y+V82S;V82S+Y123H+Y147T;V82S+Y123H+Y147R;V82S+Y123H+Q154R;Y147R+Q154R+Y123H;Y147R+Q154R+I76Y;Y147R+Q154R+T166R;Y123H+Y147R+Q154R+I76Y;V82S+Y123H+Y147R+Q154R;和I76Y+V82S+Y123H+Y147R+Q154R。

[0671] 在一些实施方案中,腺苷脱氨酶变体(例如,TadA*8)包含缺失。在一些实施方案中,腺苷脱氨酶变体包含C端的缺失。在特定实施方案中,腺苷脱氨酶变体包含相对于TadA*7.10(TadA参考序列)从残基149、150、151、152、153、154、155、156和157处开始的C端缺失,或另一个TadA中的相应突变。

[0672] 在其他实施方案中,腺苷脱氨酶变体(例如,TadA*8)是单体,该单体包含以下相对于TadA*7.10(TadA参考序列)的改变中的一者或多者:Y147T、Y147R、Q154S、Y123H、V82S、T166R和/或Q154R,或另一个TadA中的相应突变。在其他实施方案中,腺苷脱氨酶变体(TadA*8)是单体,该单体包含选自由以下组成的组的相对于TadA*7.10(TadA参考序列)的改变的组合:Y147T+Q154R;Y147T+Q154S;Y147R+Q154S;V82S+Q154S;V82S+Y147R;V82S+Q154R;V82S+Y123H;I76Y+V82S;V82S+Y123H+Y147T;V82S+Y123H+Y147R;V82S+Y123H+Q154R;Y147R+Q154R+Y123H;Y147R+Q154R+I76Y;Y147R+Q154R+T166R;Y123H+Y147R+Q154R+I76Y;V82S+Y123H+Y147R+Q154R;和I76Y+V82S+Y123H+Y147R+Q154R,或另一个TadA中的相应突变。

[0673] 在其他实施方案中,腺苷脱氨酶变体是包含两个腺苷脱氨酶结构域(例如,TadA*8)的同二聚体,该两个腺苷脱氨酶结构域各自具有以下相对于TadA*7.10(TadA参考序列)的改变中的一者或多者:Y147T、Y147R、Q154S、Y123H、V82S、T166R和/或Q154R,或另一个TadA中的相应突变。在其他实施方案中,腺苷脱氨酶变体是包含两个腺苷脱氨酶结构域(例如,TadA*8)的同二聚体,该两个腺苷脱氨酶结构域各自具有选自由以下组成的组的相对于TadA*7.10(TadA参考序列)的改变的组合:Y147T+Q154R;Y147T+Q154S;Y147R+Q154S;V82S+Q154S;V82S+Y147R;V82S+Q154R;V82S+Y123H;I76Y+V82S;V82S+Y123H+Y147T;V82S+Y123H+

Y147R;V82S+Y123H+Q154R;Y147R+Q154R+Y123H;Y147R+Q154R+I76Y;Y147R+Q154R+T166R;Y123H+Y147R+Q154R+I76Y;V82S+Y123H+Y147R+Q154R;和I76Y+V82S+Y123H+Y147R+Q154R,或另一个TadA中的相应突变。

[0674] 在其他实施方案中,腺苷脱氨酶变体是野生型腺苷脱氨酶结构域和腺苷脱氨酶变体结构域(例如,TadA*8)的异二聚体,该腺苷脱氨酶变体结构域包含以下相对于TadA*7.10(TadA参考序列)的改变中的一者或多者:Y147T、Y147R、Q154S、Y123H、V82S、T166R和/或Q154R,或另一个TadA中的相应突变。在其他实施方案中,腺苷脱氨酶变体是野生型腺苷脱氨酶结构域和腺苷脱氨酶变体结构域(例如,TadA*8)的异二聚体,该腺苷脱氨酶变体结构域包含选自由以下组成的组的相对于TadA*7.10(TadA参考序列)的改变的组合:Y147T+Q154R;Y147T+Q154S;Y147R+Q154S;V82S+Q154S;V82S+Y147R;V82S+Q154R;V82S+Y123H;I76Y+V82S;V82S+Y123H+Y147T;V82S+Y123H+Y147R;V82S+Y123H+Q154R;Y147R+Q154R+Y123H;Y147R+Q154R+I76Y;Y147R+Q154R+T166R;Y123H+Y147R+Q154R+I76Y;V82S+Y123H+Y147R+Q154R;和I76Y+V82S+Y123H+Y147R+Q154R,或另一个TadA中的相应突变。

[0675] 在其他实施方案中,腺苷脱氨酶变体是TadA*7.10结构域和腺苷脱氨酶变体结构域(例如,TadA*8)的异二聚体,该腺苷脱氨酶变体结构域包含以下相对于TadA*7.10(TadA参考序列)的改变中的一者或多者:Y147T、Y147R、Q154S、Y123H、V82S、T166R和/或Q154R,或另一个TadA中的相应突变。在其他实施方案中,腺苷脱氨酶变体是TadA*7.10结构域和腺苷脱氨酶变体结构域(例如,TadA*8)的异二聚体,该腺苷脱氨酶变体结构域包含选自由以下组成的组的相对于TadA*7.10(TadA参考序列)的改变的组合:Y147T+Q154R;Y147T+Q154S;Y147R+Q154S;V82S+Q154S;V82S+Y147R;V82S+Q154R;V82S+Y123H;I76Y+V82S;V82S+Y123H+Y147T;V82S+Y123H+Y147R;V82S+Y123H+Q154R;Y147R+Q154R+Y123H;Y147R+Q154R+I76Y;Y147R+Q154R+T166R;Y123H+Y147R+Q154R+I76Y;V82S+Y123H+Y147R+Q154R;和I76Y+V82S+Y123H+Y147R+Q154R,或另一个TadA中的相应突变。

[0676] 在特定实施方案中,腺苷脱氨酶异二聚体包含TadA*8结构域和腺苷脱氨酶结构域,该腺苷脱氨酶结构域选自金黄色葡萄球菌(S.aureus)TadA、枯草芽孢杆菌(B.subtilis)TadA、鼠伤寒沙门氏菌(S.typhimurium)TadA、腐败希瓦氏菌(S.putrefaciens)TadA、流感嗜血杆菌F3031(H.influenzae)TadA、新月柄杆菌(C.crescentus)TadA、硫还原地杆菌(G.sulfurreducens)TadA,或TadA*7.10。

[0677] 在一些实施方案中,腺苷脱氨酶是TadA*8。在一个实施方案中,腺苷脱氨酶是TadA*8,该TadA*8包含以下具有腺苷脱氨酶活性的序列或其片段,或基本上由以下具有腺苷脱氨酶活性的序列或其片段组成:

[0678] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD(SEQ ID NO:12)

[0679] 在一些实施方案中,TadA*8被截短。在一些实施方案中,截短的TadA*8相对于全长TadA*8丢失1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、6个、17个、18个、19个或20个N端氨基酸残基。在一些实施方案中,截短的TadA*8相对于全长TadA*8丢失1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、6个、17个、18个、19个或20个C端氨基酸残基。在一些实施方案中,腺苷脱氨酶变体是全长

TadA*8。

[0680] 在一些实施方案中,TadA*8是TadA*8.1、TadA*8.2、TadA*8.3、TadA*8.4、TadA*8.5、TadA*8.6、TadA*8.7、TadA*8.8、TadA*8.9、TadA*8.10、TadA*8.11、TadA*8.12、TadA*8.13、TadA*8.14、TadA*8.15、TadA*8.16、TadA*8.17、TadA*8.18、TadA*8.19、TadA*8.20、TadA*8.21、TadA*8.22、TadA*8.23或TadA*8.24。

[0681] 在其他实施方案中,本公开的碱基编辑器包含腺苷脱氨酶变体(例如,TadA*8)单体,该单体包含以下相对于TadA*7.10(TadA参考序列)的改变中的一者或多者:R26C、V88A、A109S、T111R、D119N、H122N、Y147D、F149Y、T166I和/或D167N,或另一个TadA中的相应突变。在其他实施方案中,腺苷脱氨酶变体(TadA*8)单体包含选自由以下组成的组的相对于TadA*7.10(TadA参考序列)的改变的组合:R26C+A109S+T111R+D119N+H122N+Y147D+F149Y+T166I+D167N;V88A+A109S+T111R+D119N+H122N+F149Y+T166I+D167N;R26C+A109S+T111R+D119N+H122N+F149Y+T166I+D167N;V88A+T111R+D119N+F149Y;以及A109S+T111R+D119N+H122N+Y147D+F149Y+T166I+D167N,或另一个TadA中的相应突变。

[0682] 在其他实施方案中,碱基编辑器包含野生型腺苷脱氨酶结构域和腺苷脱氨酶变体结构域(例如,TadA*8)的异二聚体,该腺苷脱氨酶变体结构域包含以下相对于TadA*7.10(TadA参考序列)的改变中的一者或多者:R26C、V88A、A109S、T111R、D119N、H122N、Y147D、F149Y、T166I和/或D167N,或另一个TadA中的相应突变。在其他实施方案中,碱基编辑器包含野生型腺苷脱氨酶结构域和腺苷脱氨酶变体结构域(例如,TadA*8)的异二聚体,该腺苷脱氨酶变体结构域包含选自由以下组成的组的相对于TadA*7.10(TadA参考序列)的改变的组合:R26C+A109S+T111R+D119N+H122N+Y147D+F149Y+T166I+D167N;V88A+A109S+T111R+D119N+H122N+F149Y+T166I+D167N;R26C+A109S+T111R+D119N+H122N+F149Y+T166I+D167N;V88A+T111R+D119N+F149Y;以及A109S+T111R+D119N+H122N+Y147D+F149Y+T166I+D167N,或另一个TadA中的相应突变。

[0683] 在其他实施方案中,碱基编辑器包含TadA*7.10结构域和腺苷脱氨酶变体结构域(例如,TadA*8)的异二聚体,该腺苷脱氨酶变体结构域包含以下相对于TadA*7.10(TadA参考序列)的改变中的一者或多者:R26C、V88A、A109S、T111R、D119N、H122N、Y147D、F149Y、T166I和/或D167N,或另一个TadA中的相应突变。在其他实施方案中,碱基编辑器包含TadA*7.10结构域和腺苷脱氨酶变体结构域(例如,TadA*8)的异二聚体,该腺苷脱氨酶变体结构域包含选自由以下组成的组的相对于TadA*7.10(TadA参考序列)的改变的组合:R26C+A109S+T111R+D119N+H122N+Y147D+F149Y+T166I+D167N;V88A+A109S+T111R+D119N+H122N+F149Y+T166I+D167N;R26C+A109S+T111R+D119N+H122N+F149Y+T166I+D167N;V88A+T111R+D119N+F149Y;以及A109S+T111R+D119N+H122N+Y147D+F149Y+T166I+D167N,或另一个TadA中的相应突变。

[0684] 在一些实施方案中,TadA*8是如表5所示的变体。表5示出了TadA氨基酸序列中的某些氨基酸位置编号以及TadA-7.10腺苷脱氨酶中这些位置中存在的氨基酸。表5还示出了在噬菌体辅助非连续进化(PANCE)和噬菌体辅助连续进化(PACE)后TadA变体中相对于TadA-7.10的氨基酸变化,如M.Richter等人,2020,Nature Biotechnology,doi.org/10.1038/s41587-020-0453-z中所述,该文献的全部内容以引用方式并入本文。在一些实施方案中,TadA*8是TadA*8a、TadA*8b、TadA*8c、TadA*8d或TadA*8e。在一些实施方案中,

TadA*8是TadA*8e。

[0685] 表5.选择TadA*8变体

[0686]

		TadA氨基酸编号									
	TadA	26	88	109	111	119	122	147	149	166	167
	TadA-7.10	R	V	A	T	D	H	Y	F	T	D
PANCE 1					R						
PANCE 2				S/T	R						
	TadA-8a	C		S	R	N	N	D	Y	I	N
	TadA-8b		A	S	R	N	N		Y	I	N
PACE	TadA-8c	C		S	R	N	N		Y	I	N
	TadA-8d		A		R	N			Y		
	TadA-8e			S	R	N	N	D	Y	I	N

[0687] 在一个实施方案中,本发明的融合蛋白包含与本文所述的腺苷脱氨酶变体(例如,TadA*8)连接的野生型TadA,该腺苷脱氨酶变体与Cas9切口酶连接。在特定实施方案中,融合蛋白包含单个TadA*8结构域(例如,作为单体提供)。在其他实施方案中,融合蛋白包含能够形成异二聚体的TadA*8和TadA(wt)。

[0688] 在一些实施方案中,腺苷脱氨酶包含与本文提供的腺苷脱氨酶中的任何腺苷脱氨酶中阐述的氨基酸序列中的任何一个氨基酸序列至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%同一的氨基酸序列。应当理解,本文提供的腺苷脱氨酶可以包括一个或多个突变(例如,本文提供的突变中的任何突变)。本公开提供了具有一定百分比同一性加上本文所述的突变或其组合中的任何突变或组合的任何脱氨酶结构域。在一些实施方案中,腺苷脱氨酶包含与参考序列或本文提供的腺苷脱氨酶中的任何腺苷脱氨酶相比具有1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、21个、24个、25个、26个、27个、28个、29个、30个、31个、32个、33个、34个、35个、36个、37个、38个、39个、40个、41个、42个、43个、44个、45个、46个、47个、48个、49个、50个或更多个突变的氨基酸序列。在一些实施方案中,腺苷脱氨酶包含与本领域已知或本文描述的氨基酸序列中的任一种氨基酸序列相比具有至少5个、至少10个、至少15个、至少20个、至少25个、至少30个、至少35个、至少40个、至少45个、至少50个、至少60个、至少70个、至少80个、至少90个、至少100个、至少110个、至少120个、至少130个、至少140个、至少150个、至少160个或至少170个同一的连续氨基酸残基的氨基酸序列。

[0689] 在特定实施方案中,TadA*8包含在以下以粗体示出的位置中的任何位置处的一个或多个突变。在其他实施方案中,TadA*8包含在用下划线示出的位置中的任何位置处的一个或多个突变:

[0690] MSEVEFSHEY WMRHALTLAK RARDEREVPV GAVLVLNNRV IGEGWNRAIG ^{50}LHDPTAHAEI MALRQGGLVM QNYRLIDATL YVTFEPCVMC AGAMIHSRIG ^{100}RVVFGVRNAK TGAAGSLMDV LHYPGMNHRV EITEGILADE CAALLCYFFR ^{150}MPRQVFNAQK KAQSSTD (SEQ ID NO: 8)

[0691] 例如,TadA*8包含相对于TadA*7.10(TadA参考序列)的单独的或与以下Y147T、Y147R、Q154S、Y123H和/或Q154R中的任何一者或多者组合的位于氨基酸位置82和/或166处的改变(例如,V82S、T166R),或另一个TadA中的相应突变。在特定实施方案中,改变的组合选自由以下组成的组:相对于TadA*7.10(TadA参考序列)的Y147T+Q154R;Y147T+Q154S;

Y147R+Q154S;V82S+Q154S;V82S+Y147R;V82S+Q154R;V82S+Y123H;I76Y+V82S;V82S+Y123H+Y147T;V82S+Y123H+Y147R;V82S+Y123H+Q154R;Y147R+Q154R+Y123H;Y147R+Q154R+I76Y;Y147R+Q154R+T166R;Y123H+Y147R+Q154R+I76Y;V82S+Y123H+Y147R+Q154R;以及I76Y+V82S+Y123H+Y147R+Q154R,或另一个TadA中的相应突变。

[0692] 在一些实施方案中,TadA*8被截短。在一些实施方案中,经截短的TadA*8相对于全长TadA*8丢失1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、6个、17个、18个、19个或20个N端氨基酸残基。在一些实施方案中,经截短的TadA*8相对于全长TadA*8丢失1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、6个、17个、18个、19个或20个C端氨基酸残基。在一些实施方案中,腺苷脱氨酶变体是全长TadA*8。

[0693] 在一个实施方案中,本发明的融合蛋白包含与本文所述的腺苷脱氨酶变体(例如,TadA*8)连接的野生型TadA,该腺苷脱氨酶变体与Cas9切口酶连接。在特定实施方案中,融合蛋白包含单个TadA*8结构域(例如,作为单体提供)。在其他实施方案中,碱基编辑器包含能够形成异二聚体的TadA*8和TadA(wt)。

[0694] 在特定实施方案中,融合蛋白包含单个(例如,作为单体提供的)TadA*8。在一些实施方案中,TadA*8与Cas9切口酶连接。在一些实施方案中,本发明的融合蛋白包含与TadA*8连接的野生型TadA(TadA(wt))的异二聚体。在其他实施方案中,本发明的融合蛋白包含与TadA*8连接的TadA*7.10的异二聚体。在一些实施方案中,碱基编辑器是包含TadA*8变体单体的ABE8。在一些实施方案中,碱基编辑器是包含TadA*8和TadA(wt)的异二聚体的ABE8。在一些实施方案中,碱基编辑器是包含TadA*8和TadA*7.10的异二聚体的ABE8。在一些实施方案中,碱基编辑器是包含TadA*8的异二聚体的ABE8。在一些实施方案中,TadA*8选自表11、13或14。在一些实施方案中,ABE8选自表13、14或16。

[0695] 在一些实施方案中,腺苷脱氨酶是TadA*9变体。在一些实施方案中,腺苷脱氨酶是选自下述变体并且参考以下序列(称为TadA*7.10)的TadA*9变体:

[0696] MSEVEFSHEY WMRHALTLAK RARDEREVPV GAVLVLNNRV IGEGWNRAIG LHDPTAHAEI MALRQGGLVM QNYRLIDATL YVTFEPCVMC AGAMIHSRIG RVVFGVRNAK TGAAGSLMDV LHYPGMNHRV EITEGILADE CAALLCYFFR MPRQVFNAQK KAQSSTD (SEQ ID NO: 8)。

[0697] 在一些实施方案中,腺苷脱氨酶包含以下改变中的一者或多者:R21N、R23H、E25F、N38G、L51W、P54C、M70V、Q71M、N72K、Y73S、V82T、M94V、P124W、T133K、D139L、D139M、C146R和A158K。该一个或多个改变在上面的序列中以下划线和粗字体示出。

[0698] 在一些实施方案中,腺苷脱氨酶包含以下改变组合中的一者或多者:V82S+Q154R+Y147R;V82S+Q154R+Y123H;V82S+Q154R+Y147R+Y123H;Q154R+Y147R+Y123H+I76Y+V82S;V82S+I76Y;V82S+Y147R;V82S+Y147R+Y123H;V82S+Q154R+Y123H;Q154R+Y147R+Y123H+I76Y;V82S+Y147R;V82S+Y147R+Y123H;V82S+Q154R+Y123H;V82S+Q154R+Y147R;V82S+Q154R+Y147R;Q154R+Y147R+Y123H+I76Y;Q154R+Y147R+Y123H+I76Y+V82S;I76Y_V82S_Y123H_Y147R_Q154R;Y147R+Q154R+H123H;和V82S+Q154R。

[0699] 在一些实施方案中,腺苷脱氨酶包含以下改变组合中的一者或多者:E25F+V82S+Y123H,T133K+Y147R+Q154R;E25F+V82S+Y123H+Y147R+Q154R;L51W+V82S+Y123H+C146R+

Y147R+Q154R;Y73S+V82S+Y123H+Y147R+Q154R;P54C+V82S+Y123H+Y147R+Q154R;N38G+V82T+Y123H+Y147R+Q154R;N72K+V82S+Y123H+D139L+Y147R+Q154R;E25F+V82S+Y123H+D139M+Y147R+Q154R;Q71M+V82S+Y123H+Y147R+Q154R;E25F+V82S+Y123H+T133K+Y147R+Q154R;E25F+V82S+Y123H+Y147R+Q154R;V82S+Y123H+P124W+Y147R+Q154R;L51W+V82S+Y123H+C146R+Y147R+Q154R;P54C+V82S+Y123H+Y147R+Q154R;Y73S+V82S+Y123H+Y147R+Q154R;N38G+V82T+Y123H+Y147R+Q154R;R23H+V82S+Y123H+Y147R+Q154R;R21N+V82S+Y123H+Y147R+Q154R;V82S+Y123H+Y147R+Q154R+A158K;N72K+V82S+Y123H+D139L+Y147R+Q154R;E25F+V82S+Y123H+D139M+Y147R+Q154R;和M70V+V82S+M94V+Y123H+Y147R+Q154R。

[0700] 在一些实施方案中,腺苷脱氨酶包含以下改变组合中的一者或多者:Q71M+V82S+Y123H+Y147R+Q154R;E25F+I76Y+V82S+Y123H+Y147R+Q154R;I76Y+V82T+Y123H+Y147R+Q154R;N38G+I76Y+V82S+Y123H+Y147R+Q154R;R23H+I76Y+V82S+Y123H+Y147R+Q154R;P54C+I76Y+V82S+Y123H+Y147R+Q154R;R21N+I76Y+V82S+Y123H+Y147R+Q154R;I76Y+V82S+Y123H+D139M+Y147R+Q154R;Y73S+I76Y+V82S+Y123H+Y147R+Q154R;E25F+I76Y+V82S+Y123H+Y147R+Q154R;I76Y+V82T+Y123H+Y147R+Q154R;N38G+I76Y+V82S+Y123H+Y147R+Q154R;R23H+I76Y+V82S+Y123H+Y147R+Q154R;P54C+I76Y+V82S+Y123H+Y147R+Q154R;R21N+I76Y+V82S+Y123H+Y147R+Q154R;I76Y+V82S+Y123H+D139M+Y147R+Q154R;Y73S+I76Y+V82S+Y123H+Y147R+Q154R;和V82S+Q154R;N72K_V82S+Y123H+Y147R+Q154R;Q71M_V82S+Y123H+Y147R+Q154R;V82S+Y123H+T133K+Y147R+Q154R;V82S+Y123H+T133K+Y147R+Q154R+A158K;M70V+Q71M+N72K+V82S+Y123H+Y147R+Q154R;N72K_V82S+Y123H+Y147R+Q154R;Q71M_V82S+Y123H+Y147R+Q154R;M70V+V82S+M94V+Y123H+Y147R+Q154R;V82S+Y123H+T133K+Y147R+Q154R;V82S+Y123H+T133K+Y147R+Q154R+A158K;和M70V+Q71M+N72K+V82S+Y123H+Y147R+Q154R。在一些实施方案中,腺苷脱氨酶是作为单体表达的。在其他实施方案中,腺苷脱氨酶是作为异二聚体表达的。在一些实施方案中,脱氨酶或其他多肽序列缺乏甲硫氨酸,例如当被包含作为融合蛋白的组分时。这可以改变位置的编号。然而,本领域技术人员将理解此类相应突变是指相同的突变,例如,Y73S和Y72S以及D139M和D138M。

[0701] 在一些实施方案中,TadA*9变体包含如本文所述的表17中所述的改变。在一些实施方案中,TadA*9变体是单体。在一些实施方案中,TadA*9变体是具有野生型TadA腺苷脱氨酶的异二聚体。在一些实施方案中,TadA*9变体是具有另一个TadA变体(例如,TadA*8、TadA*9)的异二聚体。TadA*9腺苷脱氨酶的另外的细节描述于国际PCT申请号PCT/2020/049975中,其以引用方式整体并入本文。

[0702] 本文提供的突变中的任何突变和任何另外的突变(例如,基于ecTadA氨基酸序列)可以被引入到任何其他的腺苷脱氨酶中。本文提供的突变中的任何突变可以单独或以任何组合被制作在TadA参考序列或另一种腺苷脱氨酶(例如,ecTadA)中。

[0703] A到G核碱基编辑蛋白的细节描述于国际PCT申请号PCT/2017/045381(WO2018/027078)和Gaudelli,N.M.等人,"Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage"Nature,551,464-471(2017),其全部内容据此以引用方式并入。

[0704] C到T编辑

[0705] 在一些实施方案中,本文公开的碱基编辑器包含融合蛋白,该融合蛋白包含胞苷

脱氨酶,该胞苷脱氨酶能够使多核苷酸的靶胞苷(C)碱基脱氨以产生尿苷(U),该尿苷具有胸腺嘧啶的碱基配对特性。在一些实施方案中,例如在多核苷酸是双链(例如,DNA)的情况下,尿苷碱基然后可以(例如,通过细胞修复机制)被胸苷碱基取代从而产生C:G至T:A转变。在其他实施方案中,通过碱基编辑器将核酸中的C脱氨为U不能伴随着U到T的取代。

[0706] 多核苷酸中的靶C脱氨从而产生U是可由本文描述的碱基编辑器执行的碱基编辑类型的非限制性实例。在另一个实例中,包含胞苷脱氨酶结构域的碱基编辑器可以介导胞嘧啶(C)碱基向鸟嘌呤(G)碱基的转化。例如,由碱基编辑器的胞苷脱氨酶结构域对胞苷的脱氨产生的多核苷酸的U可以通过碱基切除修复机制(例如,通过尿嘧啶DNA糖苷酶(UDG)结构域)从多核苷酸被切除,从而产生脱碱基位点(abasic site)。然后,与该脱碱基位点相对的核碱基可以(例如,通过碱基修复机制)被例如跨损伤聚合酶取代为另一个碱基(诸如C)。尽管与脱碱基位点相对的核碱基通常被C取代,但也可能发生其他取代(例如A、G或T)。

[0707] 因此,在一些实施方案中,本文所述的碱基编辑器包含能够将多核苷酸中的靶C脱氨为U的脱氨基结构域(例如,胞苷脱氨酶结构域)。进一步地,如下所述,该碱基编辑器可以包含促进由脱氨产生的U在一些实施方案中转化为T或G的另外的结构域。例如,包含胞苷脱氨酶结构域的碱基编辑器可以进一步包含用于介导T对U的取代从而完成C到T碱基编辑事件的尿嘧啶糖苷酶抑制剂(UGI)结构域。在另一个实例中,碱基编辑器可以掺入有跨损伤聚合酶以提高C到G碱基编辑的效率,因为跨损伤聚合酶可以促进与脱碱基位点相对的C的掺入(即,导致G在脱碱基位点处掺入,从而完成C到G碱基编辑事件)。

[0708] 包含胞苷脱氨酶作为结构域的碱基编辑器可以使包括DNA、RNA和DNA-RNA杂合体在内的任何多核苷酸中的靶C脱氨。通常,胞苷脱氨酶催化定位于多核苷酸单链部分的环境中的C核碱基。在一些实施方案中,包含靶C的完整多核苷酸可以是单链的。例如,掺入到碱基编辑器中的胞苷脱氨酶可以使单链RNA多核苷酸中的靶C脱氨。在其他实施方案中,包含胞苷脱氨酶结构域的碱基编辑器可以作用于双链多核苷酸,但是靶C可以定位于在脱氨反应时处于单链状态的多核苷酸的一部分中。例如,在NAGPB结构域包含Cas9结构域的实施方案中,几个核苷酸在Cas9-gRNA-靶DNA复合物的形成期间可以保持不配对,从而导致Cas9“R环复合物”的形成。这些不配对的核苷酸可以形成单链DNA泡,其可以充当单链特异性核苷酸脱氨酶(例如胞苷脱氨酶)的底物。

[0709] 在一些实施方案中,碱基编辑器的胞苷脱氨酶可包含载脂蛋白BmRNA编辑复合物(APOBEC)家族脱氨酶的全部或一部分。APOBEC是进化上保守的胞苷脱氨酶家族。此家族的成员是C到U编辑酶。APOBEC样蛋白的N端结构域是催化结构域,而C端结构域是假催化结构域。更具体地,该催化结构域是锌依赖性胞苷脱氨酶结构域并且对于胞苷脱氨作用很重要。APOBEC家族成员包括APOBEC1、APOBEC2、APOBEC3A、APOBEC3B、APOBEC3C、APOBEC3D(“APOBEC3E”现在指的是这个)、APOBEC3F、APOBEC3G、APOBEC3H、APOBEC4和激活诱导的(胞苷)脱氨酶。在一些实施方案中,掺入到碱基编辑器中的脱氨酶包含APOBEC1脱氨酶的全部或一部分。在一些实施方案中,掺入到碱基编辑器中的脱氨酶包含APOBEC2脱氨酶的全部或一部分。在一些实施方案中,掺入到碱基编辑器中的脱氨酶包含APOBEC3脱氨酶的全部或一部分。在一些实施方案中,掺入到碱基编辑器中的脱氨酶包含APOBEC3A脱氨酶的全部或一部分。在一些实施方案中,掺入到碱基编辑器中的脱氨酶包含APOBEC3B脱氨酶的全部或一部分。在一些实施方案中,掺入到碱基编辑器中的脱氨酶包含APOBEC3C脱氨酶的全部或一

部分。在一些实施方案中,掺入到碱基编辑器中的脱氨酶包含APOBEC3D脱氨酶的全部或一部分。在一些实施方案中,掺入到碱基编辑器中的脱氨酶包含APOBEC3E脱氨酶的全部或一部分。在一些实施方案中,掺入到碱基编辑器中的脱氨酶包含APOBEC3F脱氨酶的全部或一部分。在一些实施方案中,掺入到碱基编辑器中的脱氨酶包含APOBEC3G脱氨酶的全部或一部分。在一些实施方案中,掺入到碱基编辑器中的脱氨酶包含APOBEC3H脱氨酶的全部或一部分。在一些实施方案中,掺入到碱基编辑器中的脱氨酶包含APOBEC4脱氨酶的全部或一部分。在一些实施方案中,掺入到碱基编辑器中的脱氨酶包含激活诱导的脱氨酶(AID)的全部或一部分。在一些实施方案中,掺入到碱基编辑器中的脱氨酶包含胞苷脱氨酶1(CDA1)的全部或一部分。应当理解,碱基编辑器可以包含来自任何合适的生物体(例如,人或大鼠)的脱氨酶。在一些实施方案中,碱基编辑器的脱氨酶结构域来自人、黑猩猩、大猩猩、猴、牛、狗、大鼠或小鼠。在一些实施方案中,碱基编辑器的脱氨酶结构域源自自大鼠(例如,大鼠APOBEC1)。在一些实施方案中,碱基编辑器的脱氨酶结构域是人APOBEC1。在一些实施方案中,碱基编辑器的脱氨酶结构域是pmCDA1。

[0710] 下面提供了根据本公开的方面可以与Cas9融合的其他示例性脱氨酶。在实施方案中,脱氨酶是激活诱导的脱氨酶(AID)。应当理解,在一些实施方案中,可以使用相应序列的活性结构域,例如没有定位信号的结构域(没有核输出信号、细胞质定位信号的核定位序列)。

[0711] 本公开的一些方面基于以下认识:调节本文所述的融合蛋白中的任何融合蛋白的脱氨酶结构域催化活性(例如通过在脱氨酶结构域中制作点突变)能影响融合蛋白(例如,碱基编辑器)的持续合成能力。例如,降低但不消除碱基编辑融合蛋白内脱氨酶结构域的催化活性的突变可以使得脱氨酶结构域将不太可能催化与靶残基相邻的残基的脱氨,从而缩小脱氨窗口。缩小脱氨基窗口的能力可以防止与特定靶残基相邻的残基发生不需要的脱氨,这可以减少或防止脱靶效应。

[0712] 例如,在一些实施方案中,掺入到碱基编辑器中的APOBEC脱氨酶可包含选自由rAPOBEC1的H121X、H122X、R126X、R126X、R118X、W90X、W90X和R132X组成的组的一个或多个突变,或另一种APOBEC脱氨酶中的一个或多个相应突变,其中X是任何氨基酸。在一些实施方案中,掺入到碱基编辑器中的APOBEC脱氨酶可包含选自由rAPOBEC1的H121R,H122R、R126A、R126E、R118A、W90A、W90Y和R132E组成的组的一个或多个突变,或另一种APOBEC脱氨酶中的一个或多个相应突变。

[0713] 在一些实施方案中,掺入到碱基编辑器中的APOBEC脱氨酶可包含选自由hAPOBEC3G的D316X、D317X、R320X、R320X、R313X、W285X、W285X、R326X组成的组的一个或多个突变,或另一种APOBEC脱氨酶中的一个或多个相应突变,其中X是任何氨基酸。在一些实施方案中,本文提供的融合蛋白中的任何融合蛋白包含APOBEC脱氨酶,该APOBEC脱氨酶包含选自由hAPOBEC3G的D316R、D317R、R320A、R320E、R313A、W285A、W285Y、R326E组成的组的一个或多个突变,或另一种APOBEC脱氨酶中的一个或多个相应突变。

[0714] 在一些实施方案中,掺入到碱基编辑器中的APOBEC脱氨酶可包含rAPOBEC1的H121R和H122R突变,或另一种APOBEC脱氨酶中的一个或多个相应突变。在一些实施方案中,掺入到碱基编辑器中的APOBEC脱氨酶可包含含有rAPOBEC1的R126A突变或另一种APOBEC脱氨酶中的一个或多个相应突变的APOBEC脱氨酶。在一些实施方案中,掺入到碱基编辑器中

的APOBEC脱氨酶可包含含有rAPOBEC1的R126E突变或另一种APOBEC脱氨酶中的一个或多个相应突变的APOBEC脱氨酶。在一些实施方案中,掺入到碱基编辑器中的APOBEC脱氨酶可包含含有rAPOBEC1的R118A突变或另一种APOBEC脱氨酶中的一个或多个相应突变的APOBEC脱氨酶。在一些实施方案中,掺入到碱基编辑器中的APOBEC脱氨酶可包含含有rAPOBEC1的W90A突变或另一种APOBEC脱氨酶中的一个或多个相应突变的APOBEC脱氨酶。在一些实施方案中,掺入到碱基编辑器中的APOBEC脱氨酶可包含含有rAPOBEC1的W90Y突变或另一种APOBEC脱氨酶中的一个或多个相应突变的APOBEC脱氨酶。在一些实施方案中,掺入到碱基编辑器中的APOBEC脱氨酶可包含含有rAPOBEC1的R132E突变或另一种APOBEC脱氨酶中的一个或多个相应突变的APOBEC脱氨酶。在一些实施方案中,掺入到碱基编辑器中的APOBEC脱氨酶可包含含有rAPOBEC1的W90Y和R126E突变或另一种APOBEC脱氨酶中的一个或多个相应突变的APOBEC脱氨酶。在一些实施方案中,掺入到碱基编辑器中的APOBEC脱氨酶可包含含有rAPOBEC1的R126E和R132E突变或另一种APOBEC脱氨酶中的一个或多个相应突变的APOBEC脱氨酶。在一些实施方案中,掺入到碱基编辑器中的APOBEC脱氨酶可包含含有rAPOBEC1的W90Y和R132E突变或另一种APOBEC脱氨酶中的一个或多个相应突变的APOBEC脱氨酶。在一些实施方案中,掺入到碱基编辑器中的APOBEC脱氨酶可包含含有rAPOBEC1的W90Y、R126E和R132E突变或另一种APOBEC脱氨酶中的一个或多个相应突变的APOBEC脱氨酶。

[0715] 在一些实施方案中,掺入到碱基编辑器中的APOBEC脱氨酶可包含含有hAPOBEC3G的D316R和D317R突变或另一种APOBEC脱氨酶中的一个或多个相应突变的APOBEC脱氨酶。在一些实施方案中,本文提供的融合蛋白中的任何融合蛋白包含APOBEC脱氨酶,该APOBEC脱氨酶包含hAPOBEC3G的R320A突变或另一种APOBEC脱氨酶中的一个或多个相应突变。在一些实施方案中,掺入到碱基编辑器中的APOBEC脱氨酶可包含含有hAPOBEC3G的R320E突变或另一种APOBEC脱氨酶中的一个或多个相应突变的APOBEC脱氨酶。在一些实施方案中,掺入到碱基编辑器中的APOBEC脱氨酶可包含含有hAPOBEC3G的R313A突变或另一种APOBEC脱氨酶中的一个或多个相应突变的APOBEC脱氨酶。在一些实施方案中,掺入到碱基编辑器中的APOBEC脱氨酶可包含含有hAPOBEC3G的W285A突变或另一种APOBEC脱氨酶中的一个或多个相应突变的APOBEC脱氨酶。在一些实施方案中,掺入到碱基编辑器中的APOBEC脱氨酶可包含含有hAPOBEC3G的W285Y突变或另一种APOBEC脱氨酶中的一个或多个相应突变的APOBEC脱氨酶。在一些实施方案中,掺入到碱基编辑器中的APOBEC脱氨酶可包含含有hAPOBEC3G的R326E突变或另一种APOBEC脱氨酶中的一个或多个相应突变的APOBEC脱氨酶。在一些实施方案中,掺入到碱基编辑器中的APOBEC脱氨酶可包含含有hAPOBEC3G的W285Y和R320E突变或另一种APOBEC脱氨酶中的一个或多个相应突变的APOBEC脱氨酶。在一些实施方案中,掺入到碱基编辑器中的APOBEC脱氨酶可包含含有hAPOBEC3G的R320E和R326E突变或另一种APOBEC脱氨酶中的一个或多个相应突变的APOBEC脱氨酶。在一些实施方案中,掺入到碱基编辑器中的APOBEC脱氨酶可包含含有hAPOBEC3G的W285Y和R326E突变或另一种APOBEC脱氨酶中的一个或多个相应突变的APOBEC脱氨酶。在一些实施方案中,掺入到碱基编辑器中的APOBEC脱氨酶可包含含有hAPOBEC3G的W285Y、R320E和R326E突变或另一种APOBEC脱氨酶中的一个或多个相应突变的APOBEC脱氨酶。

[0716] 许多经修饰的胞苷脱氨酶是可商购的,包括但不限于可从Addgene获得的SaBE3、

SaKKH-BE3、VQR-BE3、EQR-BE3、VRER-BE3、YE1-BE3、EE-BE3、YE2-BE3和YEE-BE3(质粒85169、85170、85171、85172、85173、85174、85175、85176、85177)。在一些实施方案中,掺入到碱基编辑器中的脱氨酶包含APOBEC1脱氨酶的全部或一部分。

[0717] C到T核碱基编辑蛋白的细节描述于国际PCT申请号PCT/US2016/058344(WO2017/070632)和Komor,A.C.等人,"Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage"Nature 533,420-424(2016),其全部内容以引用方式并入本文。

[0718] 胞苷脱氨酶

[0719] 在一些实施方案中,本发明的融合蛋白包含一个或多个胞苷脱氨酶结构域。在一些实施方案中,本文提供的胞苷脱氨酶能够将胞嘧啶或5-甲基胞嘧啶脱氨为尿嘧啶或胸腺嘧啶。在一些实施方案中,本文提供的胞苷脱氨酶能够将DNA中的胞嘧啶脱氨。胞苷脱氨酶可以源自自任何合适的生物体。在一些实施方案中,胞苷脱氨酶是天然存在的胞苷脱氨酶,其包含对应于本文提供的突变中的任何突变的一个或多个突变。本领域技术人员将能够例如通过序列比对和同源残基的确定来鉴定任何同源蛋白质中的相应残基。因此,本领域技术人员将能够在任何天然存在的胞苷脱氨酶中产生对应于本文描述的突变中的任何突变的突变。在一些实施方案中,胞苷脱氨酶来自原核生物。在一些实施方案中,胞苷脱氨酶来自细菌。在一些实施方案中,胞苷脱氨酶来自哺乳动物(例如,人)。

[0720] 在一些实施方案中,胞苷脱氨酶包含与本文阐述的胞苷脱氨酶氨基酸序列中的任一种胞苷脱氨酶氨基酸序列至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%同一的氨基酸序列。应当理解,本文提供的胞苷脱氨酶可以包含一种或多种突变(例如,本文提供的突变中的任何突变)。一些实施方案提供了编码任何前述方面或如本文所描述的胞苷脱氨酶核碱基编辑器多肽的多核苷酸分子。在一些实施方案中,多核苷酸是经密码子优化的。

[0721] 本公开提供了具有一定百分比同一性加上本文所述的突变或其组合中的任何突变或组合的任何脱氨酶结构域。在一些实施方案中,胞苷脱氨酶包含与参考序列或本文提供的胞苷脱氨酶中的任何胞苷脱氨酶相比具有1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、21、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50或更多个突变的氨基酸序列。在一些实施方案中,胞苷脱氨酶包含与本领域已知或本文描述的氨基酸序列中的任一种氨基酸序列相比具有至少5个、至少10个、至少15个、至少20个、至少25个、至少30个、至少35个、至少40个、至少45个、至少50个、至少60个、至少70个、至少80个、至少90个、至少100个、至少110个、至少120个、至少130个、至少140个、至少150个、至少160个或至少170个同一的连续氨基酸残基的氨基酸序列。

[0722] 本发明第二蛋白质的融合蛋白包含两个或更多个核酸编辑结构域。

[0723] 引导多核苷酸

[0724] 当与被结合的引导多核苷酸(例如,gRNA)联合时,多核苷酸可编程核苷酸结合结构域可以特异性地结合至靶多核苷酸序列(即,经由被结合的引导核酸的碱基和靶多核苷酸序列的碱基之间的互补碱基配对),并且从而将碱基编辑器定位到需要被编辑的靶核酸序列。在一些实施方案中,靶多核苷酸序列包括单链DNA或双链DNA。在一些实施方案中,靶

多核苷酸序列包括RNA。在一些实施方案中,靶多核苷酸序列包括DNA-RNA杂合体。

[0725] CRISPR是适应性免疫系统,其提供针对可移动遗传元件(病毒、可转座元件和接合质粒)的保护。CRISPR簇含有间隔区、与先行可移动元件互补的序列和靶入侵核酸。CRISPR簇被转录并加工成CRISP R RNA(crRNA)。在II型CRISPR系统中,正确加工pre-crRNA需要反式编码的小RNA(tracrRNA)、内源核糖核酸酶3(rnc)和Cas9蛋白。tracrRNA作为核糖核酸酶3辅助加工pre-crRNA的引导。随后,Cas9/crRNA/tracrRNA核酸内切地切割与间隔区互补的线性或环状dsDNA靶标。与crRNA不互补的靶链首先被核酸内切地切割,然后被3'-5'核酸外切地修剪。在自然界中,DNA结合和切割通常需要蛋白质和这两种RNA。然而,可以对单引导RNA("sgRNA",或简称"gNRA")进行工程化,以将crRNA和tracrRNA二者的各个方面掺入到单个RNA物种中。参见例如Jinek M.,Chylinski K.,Fonfara I.,Hauer M.,Doudna J.A.,Charpentier E.Science 337:816-821(2012),其全部内容以引用方式并入本文。Cas9识别CRISPR重复序列中的一个短基序(PAM或原型间隔区相邻基序),以帮助区分自身与非自身。参见例如,"Complete genome sequence of an M1 strain of Str eptococcus pyogenes."Ferretti,J.J.等人,Natl.Acad.Sci.U.S.A.98:4658-4663(2001);"CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III."Deltcheva E.等人,Nature 471:602-607(2011);以及"Programmabledual-RNA-guided DNA endonucleas e in adaptive bacterial immunity."Jinek M.等人,Science 337:816-821(2012),这些中的每一者的全部内容以引用的方式并入本文)。

[0726] PAM序列可以是本领域已知的任何PAM序列。合适的PAM序列包括但不限于NGG、NGA、NGC、NGN、NGT、NGCG、NGAG、NGAN、NGNG、NGCN、NGCG、NGTN、NNGRRT、NNNRRT、NNGRR(N)、TTTV、TYCV、TYCV、TATV、NNNNGATT、NNAGAAW,或NAAAAC。Y是嘧啶;N是任何核苷酸碱基;W是A或T。

[0727] 在一个实施方案中,本文所述的引导多核苷酸可以是RNA或DNA。在一个实施方案中,引导多核苷酸是gRNA。RNA/Cas复合物可以辅助将Cas蛋白"引导"到靶DNA。Cas9/crRNA/tracrRNA核酸内切地切割与间隔区互补的线性或环状dsDNA靶标。与crRNA不互补的靶链首先被核酸内切地切割,然后被3'-5'核酸外切修剪。在自然界中,DNA结合和切割通常需要蛋白质和这两种RNA。然而,可以对单引导RNA("sgRNA",或简称"gNRA")进行工程化,以将crRNA和tracrRNA二者的各个方面掺入到单个RNA物种中。参见例如Jinek M.等人,Science 337:816-821(2012),其全部内容以引用方式并入本文。

[0728] 在一些实施方案中,引导多核苷酸是至少一个单引导RNA("sgRNA"或"gNRA")。在一些实施方案中,引导多核苷酸包含两个或更多个单独的多核苷酸,它们可以经由例如互补碱基配对(例如,双引导多核苷酸、双gRNA)彼此相互作用。例如,引导多核苷酸可以包含CRISPR RNA(crRNA)和反式激活性CRISPR RNA(tracrRNA)或者可以包含一种或多种反式激活性CRISPR RNA(tracrRNA)。

[0729] 在一些实施方案中,引导多核苷酸是至少一种tracrRNA。在一些实施方案中,引导多核苷酸不需要PAM序列来将多核苷酸可编程DNA结合结构域(例如,Cas9或Cpf1)引导至靶核苷酸序列。

[0730] 引导多核苷酸可以包含天然的或非天然的(不天然的)核苷酸(例如,肽核酸或核苷酸类似物)。在一些情况下,引导核酸序列的靶向性区域的长度可以是至少15个、16个、17

个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个或30个核苷酸。引导核酸的靶向性区域的长度可以在10-30个核苷酸之间,或在15-25个核苷酸之间,或在15-20个核苷酸之间。

[0731] 在一些实施方案中,本文提供的碱基编辑器利用一个或多个引导多核苷酸(例如,多个gRNA)。在一些实施方案中,单引导多核苷酸用于本文所述的不同碱基编辑器。例如,单引导多核苷酸可以用于胞苷碱基编辑器和腺苷碱基编辑器。

[0732] 在一些实施方案中,本文所述的方法可以利用经工程化的Cas蛋白。引导RNA(gRNA)是短合成RNA,其由Cas结合所必需的支架序列和使用者定义的约20个核苷酸的间隔区组成,该间隔区限定了待修饰的基因组靶标。示例性gRNA支架序列在序列表中作为SEQ ID NO:224-230、223、3000,和243-245提供。因此,熟练的技术人员可以改变Cas蛋白特异性的基因组靶标,部分取决于与基因组的其余部分相比,gRNA靶向序列对基因组靶标的特异性。

[0733] 在其他实施方案中,引导多核苷酸可以在单个分子(即,单分子引导核酸)中包含核酸的多核苷酸靶向部分和核酸的支架部分。例如,单分子引导多核苷酸可以是单引导RNA(sgRNA或gRNA)。在本文中,术语引导多核苷酸序列设想了能够与碱基编辑器相互作用并将碱基编辑器指导至靶多核苷酸序列的任何单分子、双分子或多分子核酸。

[0734] 通常,引导多核苷酸(例如,crRNA/trRNA复合物或gRNA)包含含有能够识别并结合至靶多核苷酸序列的序列的“多核苷酸靶向节段”,以及在碱基编辑器的多核苷酸可编程核苷酸结合结构域组分内稳定引导多核苷酸的“蛋白结合节段”。在一些实施方案中,引导多核苷酸的多核苷酸靶向节段识别并结合至DNA多核苷酸,从而促进DNA中碱基的编辑。在其他情况下,引导多核苷酸的多核苷酸靶向节段识别并结合至RNA多核苷酸,从而促进RNA中碱基的编辑。在本文中,“节段”是指分子的部分或区域,例如,引导多核苷酸中的一段连续的核苷酸。节段还可以指复合物的区域/部分,使得节段可以包含多于一个分子的区域。例如,在引导多核苷酸包含多个核酸分子的情况下,其蛋白质结合节段可以包含例如沿着互补区域杂交的多个单独分子的全部或一部分。在一些实施方案中,包含两个单独分子的靶向DNA的RNA的蛋白质结合节段可以包含(i)长度为100个碱基对的第一RNA分子的碱基对40-75;和(ii)长度为50个碱基对的第二RNA分子的碱基对10-25。除非另外在特定背景下确切地定义,否则“节段”的定义不限于特定数目的总碱基对,不限于来自给定RNA分子的任何具体数目的碱基对,不限于复合物内的具体数目的单独分子,并且可以包括具有任何总长度的RNA分子的区域并且可以包括与其他分子具有互补性的区域。

[0735] 引导多核苷酸可以化学、酶促或二者的组合合成。例如,可以使用标准的基于亚磷酰胺的固相合成方法合成gRNA。可替代地,可以通过将编码gRNA的DNA可操作地连接到由噬菌体RNA聚合酶识别的启动子控制序列来体外合成gRNA。合适的噬菌体启动子序列的实例包括T7、T3、SP6启动子序列或其变体。在gRNA包含两个单独分子(例如,crRNA和tracrRNA)的实施方案中,crRNA可以是化学合成的并且tracrRNA可以是酶促合成的。

[0736] gRNA分子可以在体外转录。

[0737] 引导多核苷酸可以例如通过编码gRNA的DNA(例如,包含编码gRNA的序列的DNA载体)来表达。gRNA可以单独地被编码,或与编码的碱基编辑器一起被编码。此类DNA序列可以一起或单独地被引入到表达系统,例如细胞中。例如,可以将编码多核苷酸可编程核苷酸结

合结构域和gRNA的DNA序列引入到细胞中,每个DNA序列可以是单独分子的一部分(例如,含有多核苷酸可编程核苷酸结合结构域编码序列的一个载体和含有gRNA编码序列的第二载体),或两者可以是同一分子的一部分(例如,含有多核苷酸可编程核苷酸结合结构域和gRNA两者的编码(和调控)序列的一个载体)。RNA可以从合成的DNA分子(例如gBlocks®基因片段)转录。

[0738] gRNA或引导多核苷酸可以包含三个区域:位于可以与染色体序列中的靶位点互补的5'末端处的第一区域、可以形成茎环结构的第二内部区域,以及可以是单链的第三3'区域。每个gRNA的第一区域也可以不同,使得每个gRNA将融合蛋白引导至特定靶位点。进一步地,每个gRNA的第二和第三区域在所有gRNA中可以是同一的。

[0739] gRNA或引导多核苷酸的第一区域可以与染色体序列中靶位点处的序列互补,使得gRNA的第一区域可以与靶位点碱基配对。在一些情况下,gRNA的第一区域可以包含10个核苷酸至25个核苷酸或者约10个核苷酸至25个核苷酸(即,10个核苷酸至核苷酸;或约10个核苷酸至约25个核苷酸;或10个核苷酸至约25个核苷酸;或约10个核苷酸至25个核苷酸)或更多核苷酸。例如,gRNA的第一区域和染色体序列中的靶位点之间的碱基配对区域的长度可以是或可以是约10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、22个、23个、24个、25个或更多个核苷酸。有时,gRNA的第一区域的长度可以是或可以是约19个、20个或21个核苷酸。

[0740] gRNA或引导多核苷酸还可以包含形成二级结构的第二区域。例如,由gRNA形成的二级结构可以包含茎(或发夹)和环。环和茎的长度可以变化。例如,环的长度可以在3至10个核苷酸或者约3至10个核苷酸的范围内,并且茎的长度可以在6至20个碱基对或者约6至20个碱基对的范围内。茎可以包含1至10个或约10个核苷酸的一个或多个凸起。第二区域的总长度可以在16至60个核苷酸或者约16至60个核苷酸的范围内。例如,环的长度可以是4个核苷酸或者可以是约4个核苷酸,并且茎可以是12个碱基对或者可以是约12个碱基对。

[0741] gRNA或引导多核苷酸还可以在3'末端包含第三区域,该第三区域基本上可以是单链的。例如,第三区域有时与关注的细胞中的任何染色体序列不互补,并且有时与gRNA的其余部分不互补。进一步地,第三区域的长度可以变化。第三区域的长度可以多于4个或多于约4个核苷酸。例如,第三区域的长度可以在5至60个核苷酸或者约5至60个核苷酸的范围内。

[0742] gRNA或引导多核苷酸可以靶向基因靶标的任何外显子或内含子。在一些情况下,引导物(guide)可以靶向基因的外显子1或2;在其他情况下,引导物可以靶向基因的外显子3或4。在一些实施方案中,组合物包含全部靶向相同外显子的多个gRNA或靶向不同外显子的多个gRNA。可以靶向基因的外显子和/或内含子。

[0743] gRNA或引导多核苷酸可以靶向约20个核苷酸或少于约20个核苷酸(例如,至少约5个、10个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、30个核苷酸)或约1-100个之间任何数量的核苷酸(例如,5个、10个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、30个、40个、50个、60个、70个、80个、90个、100个)的核酸序列。靶核酸序列可以是紧邻PAM的第一个核苷酸的5'的20个碱基或者可以是紧邻PAM的第一个核苷酸的5'的约20个碱基。gRNA可以靶向核酸序列。靶核酸可以是至少或至少约1-10个、1-20个、1-30个、1-40个、1-50个、1-60个、1-70个、1-80个、1-90或1-100个核苷酸。

[0744] 用于选择、设计和验证引导多核苷酸例如gRNA和靶向序列的方法在本文中进行了描述并且为本领域技术人员所知。例如,为了最小化核碱基编辑器系统中脱氨酶结构域(例如,AID结构域)的潜在底物混乱的影响,可以使可能无意中被靶向以进行脱氨的残基(例如,可潜在地驻存在靶核酸基因座内的单链DNA上的脱靶C残基)的数目最小化。此外,软件工具可以用于优化对应于靶核酸序列的gRNA,例如,用于最小化整个基因组的总脱靶活性。例如,对于使用酿脓链球菌Cas9的每个可能的靶向结构域选择,可以横跨含有多达一定数目(例如,1、2、3、4、5、6、7、8、9或10)的错配的碱基对的基因组鉴定所有脱靶序列(在选择的PAM之前,例如NAG或NGG)。可以鉴定与靶位点互补的gRNA的第一区域,并且可以根据其总预测脱靶得分对所有第一区域(例如,crRNA)排名;排名靠前的靶向性结构域代表那些可能具有最大靶上和最小脱靶活性的结构域。可以通过使用本领域已知和/或如本文所阐述的方法对候选靶向性gRNA进行功能评估。

[0745] 作为非限制性实例,gRNA的crRNA中用于与Cas9一起使用的靶DNA杂交序列可以使用DNA序列搜索算法来鉴定。gRNA设计使用基于如以下文献中所述的公共工具Cas-OFFinder的自定义gRNA设计软件进行:Bae S.、Park J.和Kim J.-S.Cas-OFFinder:A fast and versatile algorithm that searches for potential off-target sites of Cas9RNA-guided endonucleases.Bioinformatics 30,1473-1475(2014)。此软件在计算引导物的全基因组脱靶倾向后对引导物进行评分。对于长度在17至24范围内的引导物(guide),通常会考虑在从完全匹配到7个错配的范围内的匹配。一旦计算确定脱靶位点,就会为每个引导物计算总得分,并使用Web界面以表格输出形式汇总总得分。除了鉴定与PAM序列相邻的潜在靶位点外,该软件还鉴定与选定靶位点相差1个、2个、3个或超过3个核苷酸的所有PAM相邻序列。可以获得靶核酸序列(例如,靶基因)的基因组DNA序列,并且使用公开可用的工具(例如RepeatMasker程序)可以筛选重复元件。RepeatMasker在输入DNA序列中搜索重复元件和低复杂性区域。输出是给定查询序列中存在的重复序列的详细注释。

[0746] 鉴定后,基于以下对gRNA的第一区域(例如crRNA)进行分级:该第一区域与靶位点的距离、该第一区域的正交性以及与相关PAM序列紧密匹配的5'核苷酸(例如,基于含有相关PAM的人类基因组中密切匹配的鉴定的5'G,例如,酿脓链球菌的NGG PAM、金黄色葡萄球菌的NNGRRT或NNGRRV PAM)的存在。如本文所用,正交性是指人类基因组中含有最少数目的与靶序列的错配的序列的数目。"高水平正交性"或"良好正交性"可以例如指不具有人类基因组中除预期靶标之外的同一序列,也不具有在靶序列中含有一个或两个错配的任何序列的20-mer靶向性结构域。可以选择具有良好正交性的靶向性结构域以最小化脱靶DNA切割。

[0747] 然后可以将gRNA作为RNA分子或非RNA核酸分子(例如,DNA分子)引入到细胞或胚胎中。在一个实施方案中,编码gRNA的DNA可操作地连接至启动子控制序列,以用于在关注的细胞或胚胎中表达该gRNA。RNA编码序列可以可操作地连接至由RNA聚合酶III(Pol III)识别的启动子序列。可用于表达gRNA的质粒载体包括但不限于px330载体和px333载体。在一些情况下,质粒载体(例如,px333载体)可以包含至少两个编码gRNA的DNA序列。进一步地,载体可以包含另外的表达控制序列(例如,增强子序列、科扎克(Kozak)序列、多聚腺苷酸化序列、转录终止序列等)、可选择标志物序列(例如,GFP或抗生素抗性基因诸如嘌呤霉素)、复制起点等。编码gRNA的DNA分子还可以是线性的。编码gRNA或引导多核苷酸的DNA分子还可以是环状的。

[0748] 在一些实施方案中,报告系统用于检测碱基编辑活性和测试候选引导多核苷酸。在一些实施方案中,报告系统包括基于报告基因的测定,在该测定中碱基编辑活性导致报告基因的表达。例如,报告系统可以包括包含经去激活的起始密码子(例如,模板链上从3'-TAC-5'到3'-CAC-5'的突变)的报告基因。在靶C成功脱氨后,相应的mRNA将被转录为5'-AUG-3'而不是5'-GUG-3',从而实现报告基因的翻译。合适的报告基因对于本领域技术人员将是显而易见的。报告基因的非限制性实例包括编码绿色荧光蛋白(GFP)、红色荧光蛋白(RFP)、萤光素酶、分泌型碱性磷酸酶(SEAP)的基因,或者其表达对于本领域技术人员来说是可检测的和显而易见的任何其他基因。报告系统可以用于测试许多不同的gRNA,例如,以便确定相应脱氨酶将靶向关于靶DNA序列的哪个(哪些)残基。还可以测试靶向非模板链的sgRNA,以便评估特定碱基编辑蛋白(例如Cas9脱氨酶融合蛋白)的脱靶效应。在一些实施方案中,此类gRNA可以被设计成使得突变的起始密码子将不会与gRNA碱基配对。引导多核苷酸可以包括标准核糖核苷酸、经修饰的核糖核苷酸(例如,假尿苷)、核糖核苷酸异构体和/或核糖核苷酸类似物。在一些实施方案中,引导多核苷酸可以包含至少一种可检测标记。可检测标记可以是荧光团(例如,FAM、TMR、Cy3、Cy5、Texas Red、Oregon Green、Alexa Fluors、Halo标签或合适的荧光染料)、检测标签(例如,生物素、地高辛等)、量子点或金粒子。

[0749] 在一些实施方案中,碱基编辑器系统可以包含多个引导多核苷酸,例如gRNA。例如,gRNA可以靶向于一个或多个包含在碱基编辑器系统中的靶基因座(例如,至少1个gRNA、至少2个gRNA、至少5个gRNA、至少10个gRNA、至少20个gRNA、至少30个g RNA、至少50个gRNA)。该多个gRNA序列可以串联排列并且优选地被直接重复序列分开。

[0750] 引导多核苷酸可以包含一种或多种修饰以提供具有新的或增强的特征的核酸。引导多核苷酸可以包含核酸亲和标签。引导多核苷酸可以包含合成核苷酸、合成核苷酸类似物、核苷酸衍生物和/或经修饰的核苷酸。

[0751] 在一些情况下,gRNA或引导多核苷酸可以包括修饰。可以在gRNA或引导多核苷酸的任何区位处制作修饰。可以对单个gRNA或引导多核苷酸制作超过一种修饰。gRNA或引导多核苷酸可以在修饰后经受质量控制。在一些情况下,质量控制可以包括PAGE、HPLC、MS或它们的任何组合。

[0752] gRNA或引导多核苷酸的修饰可以是取代、插入、缺失、化学修饰、物理修饰、稳定化、纯化或它们的任何组合。

[0753] gRNA或引导多核苷酸也可以被以下修饰:5'腺苷酸、5'鸟苷-三磷酸酯帽、5'N7-甲基鸟苷-三磷酸酯帽、5'三磷酸酯帽、3'磷酸酯、3'硫代磷酸酯、5'磷酸酯、5'硫代磷酸酯、Cis-Syn胸苷二聚体、三聚体、C12间隔区、C3间隔区、C6间隔区、dSpacer、PC间隔区、rSpacer、间隔区18、间隔区9、3'-3'修饰、5'-5'修饰、脱碱基、吖啶、偶氮苯、生物素、生物素BB、生物素TEG、胆固醇TEG、脱硫生物素TEG、DNP TEG、DNP-X、DOTA、dT-生物素、双生物素、PC生物素、补骨脂素C2、补骨脂素C6、TINA、3'DABCYL、黑洞淬灭剂1、黑洞淬灭剂2、DABCYL SE、dT-DABCYL、IRDye QC-1、QSY-21、QSY-35、QSY-7、QSY-9、羧基接头、硫醇接头、2'-脱氧核糖核苷类似物嘌呤、2'-脱氧核糖核苷类似物嘧啶、核糖核苷类似物、2'-O-甲基核糖核苷类似物、糖修饰类似物、摇摆/通用碱基、荧光染料标记、2'-氟RNA、2'-O-甲基RNA、甲基膦酸酯、磷酸二酯DNA、磷酸二酯RNA、硫代磷酸酯DNA、硫代磷酸酯RNA、UNA、假尿苷-5'-三磷酸酯、5'-甲基胞苷-5'-三磷酸酯,或它们的任何组合。

[0754] 在一些情况下,修饰是永久性的。在其他情况下,修饰是瞬时的。在一些情况下,对gRNA或引导多核苷酸制作多种修饰。gRNA或引导多核苷酸修饰可以改变核苷酸的物理化学特性,诸如它们的构象、极性、疏水性、化学反应性、碱基配对相互作用或它们的任何组合。

[0755] 通过用经分离的gRNA或包含编码引导RNA的序列和启动子的质粒DNA转染细胞,可以将引导多核苷酸转移至该细胞中。gRNA或引导多核苷酸也可以通过其他方式(诸如使用病毒介导的基因递送)转移到细胞中。可以分离gRNA或引导多核苷酸。例如,gRNA可以以经分离的RNA的形式转染到细胞或生物体中。可以使用本领域已知的任何体外转录系统通过体外转录制备gRNA。gRNA可以以经分离的RNA的形式而不是以包含gRNA编码序列的质粒的形式转移到细胞中。

[0756] 修饰也可以是硫代磷酸酯取代物。在一些情况下,天然磷酸二酯键可能容易被细胞核酸酶快速降解;并且使用硫代磷酸酯(PS)键取代物对核苷酸间连键进行的修饰对于细胞降解引起的水解可以更稳定。修饰可以增加gRNA或引导多核苷酸的稳定性。修饰还可以增强生物活性。在一些情况下,硫代磷酸酯增强的RNA gRNA可以抑制RNA酶A、RNA酶T1、小牛血清核酸酶或它们的任何组合。这些特性可以允许PS-RNA gRNA用于在体内或体外暴露于核酸酶的可能性较高的应用中。例如,可以在gRNA的5'或5''末端的最后3-5个核苷酸之间引入硫代磷酸酯(PS)键,其可以抑制核酸外切酶降解。在一些情况下,可以在整个gRNA中添加硫代磷酸酯键以减少核酸内切酶的攻击。

[0757] 在一些实施方案中,引导RNA被设计成破坏剪接位点(即,剪接受体(SA)或剪接供体(SD))。在一些实施方案中,引导RNA被设计成使得碱基编辑导致提前终止密码子。

[0758] 原型间隔区相邻基序

[0759] 术语"原型间隔区相邻基序(PAM)"或PAM样基序是指紧接在被CRISPR细菌适应性免疫系统中Cas9核酸酶靶向的DNA序列之后的2-6个碱基对的DNA序列。在一些实施方案中,PAM可以是5'PAM(即,位于原型间隔区5'末端的上游)。在其他实施方案中,PAM可以是3'PAM(即,位于原型间隔区5'末端的下游)。PAM序列对于靶标结合至关重要,但确切的序列取决于Cas蛋白的类型。PAM序列可以是本领域已知的任何PAM序列。合适的PAM序列包括但不限于NGG、NGA、NGC、NGN、NGT、NGTT、NGCG、NGAG、NGAN、NGNG、NGCN、NGCG、NGTN、NNGRRT、NNNRRT、NNGRR(N)、TTTV、TYCV、TYCV、TATV、NNNNGATT、NNAGAAW,或NAAAAC。Y是嘧啶;N是任何核苷酸碱基;W是A或T。

[0760] 本文提供的碱基编辑器可以包含CRISPR蛋白衍生结构域,该结构域能够结合含有规范或非规范原型间隔区相邻基序(PAM)序列的核苷酸序列。PAM位点是接近靶多核苷酸序列的核苷酸序列。本公开的一些方面提供了碱基编辑器,其包含具有不同PAM特异性的CRISPR蛋白的全部或一部分。

[0761] 例如,Cas9蛋白,诸如来自酿脓链球菌的Cas9(spCas9),通常需要规范的NGG PAM序列来结合特定的核酸区域,其中"NGG"中的"N"是腺嘌呤(A)、胸腺嘧啶(T)、鸟嘌呤(G)或胞嘧啶(C),并且G是鸟嘌呤。PAM可以是CRISPR蛋白特异性的,并且在包含不同CRISPR蛋白衍生结构域的不同碱基编辑器之间可以不同。PAM可以是靶序列的5'或3'。PAM可以位于靶序列的上游或下游。PAM的长度可以是1个、2个、3个、4个、5个、6个、7个、8个、9个、10个或更多个核苷酸。通常,PAM的长度在2-6个核苷酸之间。

[0762] 在一些实施方案中,PAM是"NRN"PAM,其中"NRN"中的"N"是腺嘌呤(A)、胸腺嘧啶

(T)、鸟嘌呤(G)或胞嘧啶(C),并且R是腺嘌呤(A)或鸟嘌呤(G);或PAM为“NYN”PAM,其中NYN中的“N”是腺嘌呤(A)、胸腺嘧啶(T)、鸟嘌呤(G)或胞嘧啶(C),并且Y是胞苷(C)或胸腺嘧啶(T),例如,如R.T.Walton等人,2020,Science,10.1126/science.aba8853(2020)中描述,其全部内容以引用方式并入本文。

[0763] 下面的表6中描述了几种PAM变体。

[0764] 表6.Cas9蛋白和相应的PAM序列

[0765]

变体	PAM
spCas9	NGG
spCas9-VRQR	NGA
spCas9-VRER	NGCG
xCas9(sp)	NGN
saCas9	NNGRRT
saCas9-KKH	NNNRRT
spCas9-MQKSER	NGCG
spCas9-MQKSER	NGCN
spCas9-LRKIQK	NGTN
spCas9-LRVSQK	NGTN
spCas9-LRVSQL	NGTN
spCas9-MQKFRAER	NGC
Cpf1	5’(TTTV)
SpyMac	5’-NAA-3’

[0766] 在一些实施方案中,PAM是NGC。在一些实施方案中,NGC PAM被Cas9变体识别。在一些实施方案中,NGC PAM变体包含一个或多个选自D1135M、S1136Q、G1218K、E1219F、A1322R、D1332A、R1335E和T1337R(统称为“MQKFRAER”)的氨基酸取代。

[0767] 在一些实施方案中,PAM是NGT。在一些实施方案中,NGT PAM被Cas9变体识别。在一些实施方案中,NGT PAM变体是通过在一个或多个残基1335、1337、1135、1136、1218和/或1219处的被靶向突变产生的。在一些实施方案中,NGT PAM变体是通过在一个或多个残基1219、1335、1337、1218处的被靶向突变产生的。在一些实施方案中,NGT PAM变体是通过在一个或多个残基1135、1136、1218、1219和或1335处的被靶向突变产生的。在一些实施方案中,NGT PAM变体选自下表7A和7B中提供的被靶向突变组。

[0768] 表7A:在残基1219、1335、1337、1218处的NGT PAM变体突变

[0769]

变体	E1219V	R1335Q	T1337	G1218
1	F	V	T	
2	F	V	R	
3	F	V	Q	
4	F	V	L	
5	F	V	T	R
6	F	V	R	R
7	F	V	Q	R

8	F	V	L	R
9	L	L	T	
10	L	L	R	
11	L	L	Q	
12	L	L	L	
13	F	I	T	
14	F	I	R	
15	F	I	Q	
16	F	I	L	
17	F	G	C	
18	H	L	N	
19	F	G	C	A
20	H	L	N	V
21	L	A	W	
22	L	A	F	
23	L	A	Y	
24	I	A	W	
25	I	A	F	
26	I	A	Y	

[0770] 表7B:在残基1135、1136、1218、1219和1335处的NGT PAM变体突变

[0771]

变体	D1135L	S1136R	G1218S	E1219V	R1335Q
27	G				
28	V				
29	I				
30		A			
31		W			
32		H			
33		K			
34			K		
35			R		
36			Q		
37			T		
38			N		
39				I	
40				A	
41				N	
42				Q	
43				G	
44				L	

45				S	
46				T	
47					L
48					I
49					V
50					N
51					S
52					T
53					F
54					Y
55	N1286Q	I1331F			

[0772] 在一些实施方案中,NGT PAM变体选自表7A和表7B中的变体5、7、28、31或36。在一些实施方案中,变体具有改进的NGT PAM识别。

[0773] 在一些实施方案中,NGT PAM变体在残基1219、1335、1337和/或1218处具有突变。在一些实施方案中,从下表8中提供的变体中选择具有用于改进识别的突变的NGT PAM变体。

[0774] 表8:在残基1219、1335、1337和1218处的NGT PAM变体突变

[0775]

变体	E1219V	R1335Q	T1337	G1218
1	F	V	T	
2	F	V	R	
3	F	V	Q	
4	F	V	L	
5	F	V	T	R
6	F	V	R	R
7	F	V	Q	R
8	F	V	L	R

[0776] 在一些实施方案中,NGT PAM选自下表9中提供的变体。

[0777] 表9.NGT PAM变体

[0778]

	NGTN 变体	D1135	S1136	G1218	E1219	A1322R	R1335	T1337
变体1	LRKIQK	L	R	K	I	-	Q	K
变体2	LRSVQK	L	R	S	V	-	Q	K
变体3	LRSVQL	L	R	S	V	-	Q	L
变体4	LRKIRQK	L	R	K	I	R	Q	K
变体5	LRSVRQK	L	R	S	V	R	Q	K
变体6	LRSVRQL	L	R	S	V	R	Q	L

[0779] 在一些实施方案中,NGTN变体是变体1。在一些实施方案中,NGTN变体是变体2。在一些实施方案中,NGTN变体是变体3。在一些实施方案中,NGTN变体是变体4。在一些实施方案中,NGTN变体是变体5。在一些实施方案中,NGTN变体是变体6。

[0780] 在一些实施方案中,Cas9结构域是来自酿脓链球菌的Cas9结构域(SpCas9)。在一些实施方案中,SpCas9结构域是核酸酶有活性的SpCas9、核酸酶无活性的SpCas9(SpCas9d)或SpCas9切口酶(SpCas9n)。在一些实施方案中,SpCas9包含D9X突变,或本文提供的氨基酸

序列中的任何氨基酸序列中的相应突变,其中X是除D之外的任何氨基酸。在一些实施方案中,SpCas9包含D9A突变,或本文提供的氨基酸序列中的任何氨基酸序列中的相应突变。在一些实施方案中,SpCas9结构域、SpCas9d结构域或SpCas9n结构域可以结合至具有非规范PAM的核酸序列。在一些实施方案中,SpCas9结构域、SpCas9d结构域或SpCas9n结构域可以结合至具有NGG、NGA或NGCG PAM序列的核酸序列。

[0781] 在一些实施方案中,SpCas9结构域包含D1135X、R1335X和T1337X突变中的一者或多者,或本文提供的氨基酸序列中的任何氨基酸序列中的相应突变,其中X是任何氨基酸。在一些实施方案中,SpCas9结构域包含D1135E、R1335Q和T1337R突变中的一者或多者,或本文提供的氨基酸序列中的任何氨基酸序列中的相应突变。在一些实施方案中,SpCas9结构域包含D1135E、R1335Q和T1337R突变,或本文提供的氨基酸序列中的任何氨基酸序列中的相应突变。在一些实施方案中,SpCas9结构域包含D1135X、R1335X和T1337X突变中的一者或多者,或本文提供的氨基酸序列中的任何氨基酸序列中的相应突变,其中X是任何氨基酸。在一些实施方案中,SpCas9结构域包含D1135V、R1335Q和T1337R突变中的一者或多者,或本文提供的氨基酸序列中的任何氨基酸序列中的相应突变。在一些实施方案中,SpCas9结构域包含D1135V、R1335Q和T1337R突变,或本文提供的氨基酸序列中的任何氨基酸序列中的相应突变。在一些实施方案中,SpCas9结构域包含D1135X、G1218X、R1335X和T1337X突变中的一者或多者,或本文提供的氨基酸序列中的任何氨基酸序列中的相应突变,其中X是任何氨基酸。在一些实施方案中,SpCas9结构域包含D1135V、G1218R、R1335Q和T1337R突变中的一者或多者,或本文提供的氨基酸序列中的任何氨基酸序列中的相应突变。在一些实施方案中,SpCas9结构域包含D1135V、G1218R、R1335Q和T1337R突变,或本文提供的氨基酸序列中的任何氨基酸序列中的相应突变。

[0782] 在一些实例中,可将由本文公开的碱基编辑器的CRISPR蛋白衍生结构域识别的PAM在编码碱基编辑器的插入物(例如,AAV插入物)的单独寡核苷酸上提供给细胞。在此类实施方案中,在单独寡核苷酸上提供PAM可以允许切割否则将不能被切割的靶序列,因为在与靶序列相同的多核苷酸上不存在相邻的PAM。

[0783] 在一个实施方案中,酿脓链球菌Cas9(SpCas9)可以用作用于基因组工程化的CRISPR核酸内切酶。然而,也可以使用其他的。在一些实施方案中,可以使用不同的核酸内切酶来靶向某些基因组靶标。在一些实施方案中,可以使用具有非NGG PAM序列的合成SpCas9衍生变体。此外,已经鉴定了来自不同物种的其他Cas9直向同源物,并且这些“非SpCas9”可以结合也可用于本公开的多种PAM序列。例如,相对大的尺寸的SpCas9(大约4kb编码序列)可以导致携带不能在细胞中有效率地表达的SpCas9 cDNA的质粒。相反,金黄色葡萄球菌Cas9(SaCas9)的编码序列比SpCas9短大约1千碱基,从而可能使它能够在细胞中有效率地表达。与SpCas9类似,SaCas9核酸内切酶能够体外修饰哺乳动物细胞中的靶基因和体内修饰小鼠中的靶基因。在一些实施方案中,Cas蛋白可以靶向不同的PAM序列。在一些实施方案中,靶基因可以与例如Cas9 PAM、5’-NGG相邻。在其他实施方案中,其他Cas9直向同源物可以具有不同的PAM需求。例如,其他PAM,诸如嗜热链球菌(CRISPR1的5’-NNAGAA和CRISPR3的5’-NGGNG)和脑膜炎奈瑟菌(5’-NNNNGATT)的PAM,也可以被发现与靶基因相邻。

[0784] 在一些实施方案中,对于酿脓链球菌系统,靶基因序列可以在5’-NGG PAM之前(即,是5’-NGG PAM的5’),并且20-nt的引导RNA序列可以与相反链碱基配对以介导与PAM相

邻的Cas9切割。在一些实施方案中,相邻的切口可以是PAM上游的3个或约3个碱基对。在一些实施方案中,相邻的切口可以是PAM上游的10个或约10个碱基对。在一些实施方案中,相邻的切口可以是PAM上游的0-20个或约0-20个碱基对。例如,相邻的切口可以紧挨着PAM上游1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个或30个碱基对。相邻的切口也可以在PAM下游1到30个碱基对旁边。能够结合PAM序列的示例性SpCas9蛋白的序列如下:

[0785] 在一些实施方案中,经工程化的SpCas9变体能够识别侧接有3'H(非G PAM)的原型间隔区相邻基序(PAM)序列(参见表2A-2D)。在一些实施方案中,SpCas9变体识别NRNH PAM(其中R是A或G并且H是A、C或T)。在一些实施方案中,非G PAM是NRRH、NRTH或NRCH(参见例如Miller,S.M.等人Continuous evolution of SpCas9 variants compatible with non-G PAMs,Nat.Biotechnol.(2020),其全部内容以引用方式并入本文)。

[0786] 在一些实施方案中,Cas9结构域是重组Cas9结构域。在一些实施方案中,重组Cas9结构域是SpyMacCas9结构域。在一些实施方案中,SpyMacCas9结构域是核酸酶有活性的SpyMacCas9、核酸酶无活性的SpyMacCas9(SpyMacCas9d)或SpyMacCas9切口酶(SpyMacCas9n)。在一些实施方案中,SaCas9结构域、SaCas9d结构域或SaCas9n结构域可以结合至具有非规范PAM的核酸序列。在一些实施方案中,SpyMacCas9结构域、SpCas9d结构域或SpCas9n结构域可以结合至具有NAA PAM序列的核酸序列。

[0787] 猕猴链球菌(Streptococcus macacae)中具有原生5'-NAAN-3'PA M特异性的SpyCas 9的示例性Cas9 A同系物的序列是本领域已知的并且例如由Jakimo等人(www.biorxiv.org/content/biorxiv/early/2018/09/27/429654.full.pdf)描述,并且在序列表中作为SEQ ID NO:1307提供。

[0788] 在一些实施方案中,变体Cas9蛋白具有H840A、P475A、W476A、N477A、D1125A、W1126A和D1218A突变,使得多肽切割靶DNA或RNA的能力减小。此种Cas9蛋白切割靶DNA(例如,单链靶DNA)的能力减小,但保留了结合靶DNA(例如,单链靶DNA)的能力。作为另一个非限制性实例,在一些实施方案中,变体Cas9蛋白具有D10A、H840A、P475A、W476A、N477A、D1125A、W1126A和D1218A突变,使得多肽切割靶DNA的能力减小。此种Cas9蛋白切割靶DNA(例如,单链靶DNA)的能力减小,但保留了结合靶DNA(例如,单链靶DNA)的能力。在一些实施方案中,当变体Cas9蛋白携有W476A和W1126A突变或当变体Cas9蛋白携有P475A、W476A、N477A、D1125A、W1126A和D1218A突变时,变体Cas9蛋白不能有效率地结合至PAM序列。因此,在一些此类情况下,当此种变体Cas9蛋白用于结合的方法时,该方法不需要PAM序列。换言之,在一些实施方案中,当此种变体Cas9蛋白用于结合的方法时,该方法可以包括引导RNA,但是该方法可以在不存在PAM序列的情况下进行(并且因此由引导RNA的靶向节段提供结合的特异性)。可以使其它残基突变以实现以上作用(即,使一个或另一个核酸酶部分灭活)。作为非限制性实例,残基D10、G12、G17、E762、H840、N854、N863、H982、H983、A984、D986和/或A987可以被改变(即,取代)。同样,除了丙氨酸取代以外的突变也是合适的。

[0789] 在一些实施方案中,碱基编辑器的CRISPR蛋白衍生结构域可以包含具有规范PAM序列(NGG)的Cas9蛋白的全部或一部分。在其他实施方案中,碱基编辑器的Cas9衍生结构域可以采用非规范PAM序列。此类序列已在本领域中进行了描述并且对于本领域技术人员来

说是显而易见的。例如,结合非规范PAM序列的Cas9结构域已描述于Kleinstiver,B.P.等人,“Engineered CRISPR-Cas9 nucleases with altered PAM specificities”Nature 523,481-485(2015);以及Kleinstiver,B.P.等人,“Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition”Nature Biotechnology 33,1293-1298(2015);R.T.Walton等人“Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants”Science10.1126/science.aba8853(2020);Hu等人“Evolved Cas9 variants with broad PAM compatibility and high DNA specificity,”Nature,2018年4月5日,556(7699),57-63;Miller等人,“Continuous evolution of SpCas9variants compatible with non-G PAMs”Nat.Biotechnol.,2020年4月;38(4):471-481;每一篇的全部内容以引用方式并入本文。

[0790] 包含NapDNAbp和胞苷脱氨酶和/或腺苷脱氨酶的融合蛋白

[0791] 本公开的一些方面提供了包含Cas9结构域或其他核酸可编程DNA结合蛋白(例如,Cas12)和一个或多个胞苷脱氨酶或腺苷脱氨酶结构域的融合蛋白。应当理解,Cas9结构域可以是本文提供的Cas9结构域或Cas9蛋白中的任何Cas9结构域或Cas9蛋白(例如,dCas9或nCas9)。在一些实施方案中,本文提供的Cas9结构域或Cas9蛋白中的任何Cas9结构域或Cas9蛋白(例如,dCas9或nCas9)可以与本文提供的胞苷脱氨酶和/或腺苷脱氨酶中的任何胞苷脱氨酶和/或腺苷脱氨酶融合。本文公开的碱基编辑器的结构域可以以任何顺序排列。

[0792] 在一些实施方案中,融合蛋白包含以下结构域A-C、A-D或A-E:

[0793] NH_2-[A-B-C]-COOH;

[0794] NH_2-[A-B-C-D]-COOH;或

[0795] NH_2-[A-B-C-D-E]-COOH;

[0796] 其中A和C或A、C和E各自包含以下中的一种或多种:

[0797] 腺苷脱氨酶结构域或其活性片段,

[0798] 胞苷脱氨酶结构域或其活性片段,并且

[0799] 其中B或B和D各自包含一个或多个具有核酸序列特异性结合活性的结构域。

[0800] 在一些实施方案中,融合蛋白包含以下结构:

[0801] NH_2-[A_n-B_o-C_n]-COOH;

[0802] NH_2-[A_n-B_o-C_n-D_o]-COOH;或

[0803] NH_2-[A_n-B_o-C_p-D_o-E_q]-COOH;

[0804] 其中A和C或A、C和E各自包含以下中的一种或多种:

[0805] 腺苷脱氨酶结构域或其活性片段,

[0806] 胞苷脱氨酶结构域或其活性片段,并且

[0807] 其中n是整数:1、2、3、4或5,其中p是整数:0、1、2、3、4或5;其中q是整数0、1、2、3、4,或5;并且其中B或B和D各自包含具有核酸序列特异性结合活性的结构域;并且其中o是整数:1、2、3、4或5。

[0808] 例如但不限于,在一些实施方案中,融合蛋白包含以下结构:

[0809] NH2-[腺苷脱氨酶]-[Cas9结构域]-COOH;

[0810] NH2-[Cas9结构域]-[腺苷脱氨酶]-COOH;

[0811] NH2-[胞苷脱氨酶]-[Cas9结构域]-COOH;

[0812] NH2-[Cas9结构域]-[胞苷脱氨酶]-COOH;

[0813] NH2-[胞苷脱氨酶]-[Cas9结构域]-[腺苷脱氨酶]-COOH;

[0814] NH2-[腺苷脱氨酶]-[Cas9结构域]-[胞苷脱氨酶]-COOH;

[0815] NH2-[腺苷脱氨酶]-[胞苷脱氨酶]-[Cas9结构域]-COOH;

[0816] NH2-[胞苷脱氨酶]-[腺苷脱氨酶]-[Cas9结构域]-COOH;

[0817] NH2-[Cas9结构域]-[腺苷脱氨酶]-[胞苷脱氨酶]-COOH;或

[0818] NH2-[Cas9结构域]-[胞苷脱氨酶]-[腺苷脱氨酶]-COOH。

[0819] 在一些实施方案中,本文提供的Cas12结构域或Cas12蛋白中的任何Cas12结构域或Cas12蛋白可以与本文提供的胞苷或腺苷脱氨酶中的任何胞苷或腺苷脱氨酶融合。例如但不限于,在一些实施方案中,融合蛋白包含以下结构:

[0820] NH2-[腺苷脱氨酶]-[Cas12结构域]-COOH;

[0821] NH2-[Cas12结构域]-[腺苷脱氨酶]-COOH;

[0822] NH2-[胞苷脱氨酶]-[Cas12结构域]-COOH;

[0823] NH2-[Cas12结构域]-[胞苷脱氨酶]-COOH;

[0824] NH2-[胞苷脱氨酶]-[Cas12结构域]-[腺苷脱氨酶]-COOH;

[0825] NH2-[腺苷脱氨酶]-[Cas12结构域]-[胞苷脱氨酶]-COOH;

[0826] NH2-[腺苷脱氨酶]-[胞苷脱氨酶]-[Cas12结构域]-COOH;

[0827] NH2-[胞苷脱氨酶]-[腺苷脱氨酶]-[Cas12结构域]-COOH;

[0828] NH2-[Cas12结构域]-[腺苷脱氨酶]-[胞苷脱氨酶]-COOH;或

[0829] NH2-[Cas12结构域]-[胞苷脱氨酶]-[腺苷脱氨酶]-COOH。

[0830] 在一些实施方案中,腺苷脱氨酶是TadA*8。示例性融合蛋白结构包括以下:

[0831] NH2-[TadA*8]-[Cas9结构域]-COOH;

[0832] NH2-[Cas9结构域]-[TadA*8]-COOH;

[0833] NH2-[TadA*8]-[Cas12结构域]-COOH;或

[0834] NH2-[Cas12结构域]-[TadA*8]-COOH。

[0835] 在一些实施方案中,融合蛋白的腺苷脱氨酶包括TadA*8和胞苷脱氨酶和/或腺苷脱氨酶。在一些实施方案中,TadA*8是TadA*8.1、TadA*8.2、TadA*8.3、TadA*8.4、TadA*8.5、TadA*8.6、TadA*8.7、TadA*8.8、TadA*8.9、TadA*8.10、TadA*8.11、TadA*8.12、TadA*8.13、TadA*8.14、TadA*8.15、TadA*8.16、TadA*8.17、TadA*8.18、TadA*8.19、TadA*8.20、TadA*8.21、TadA*8.22、TadA*8.23或TadA*8.24。

[0836] 示例性融合蛋白结构包括以下:

[0837] NH2-[TadA*8]-[Cas9/Cas12]-[腺苷脱氨酶]-COOH;

[0838] NH2-[腺苷脱氨酶]-[Cas9/Cas12]-[TadA*8]-COOH;

[0839] NH2-[TadA*8]-[Cas9/Cas12]-[胞苷脱氨酶]-COOH;或

[0840] NH2-[胞苷脱氨酶]-[Cas9/Cas12]-[TadA*8]-COOH。

[0841] 在一些实施方案中,融合蛋白的腺苷脱氨酶包括TadA*9和胞苷脱氨酶和/或腺苷脱氨酶。示例性融合蛋白结构包括以下:

[0842] NH2-[TadA*9]-[Cas9/Cas12]-[腺苷脱氨酶]-COOH;

[0843] NH2-[腺苷脱氨酶]-[Cas9/Cas12]-[TadA*9]-COOH;

[0844] NH2-[TadA*9]-[Cas9/Cas12]-[胞苷脱氨酶]-COOH;或

[0845] NH2-[胞苷脱氨酶]-[Cas9/Cas12]-[TadA*9]-COOH。

[0846] 在一些实施方案中,融合蛋白可以包含侧接有Cas9或Cas12多肽的N端片段和C端片段的脱氨酶。在一些实施方案中,融合蛋白包含侧接有Cas9或Cas12多肽的N端片段和C端片段的胞苷脱氨酶。在一些实施方案中,融合蛋白包含侧接有Cas9或Cas 12多肽的N端片段和C端片段的腺苷脱氨酶。

[0847] 在一些实施方案中,包含胞苷脱氨酶或腺苷脱氨酶和napDNAbp(例如,Cas9或Cas12结构域)的融合蛋白不包含接头序列。在一些实施方案中,接头存在于胞苷或腺苷脱氨酶与napDNAbp之间。在一些实施方案中,上方通用架构中使用的“-”指示存在任选的接头。在一些实施方案中,胞苷或腺苷脱氨酶和napDNAbp经由本文提供的接头中的任何接头融合。例如,在一些实施方案中,胞苷或腺苷脱氨酶和napDNAbp经由本文提供的接头中的任何接头融合。

[0848] 应当理解,本公开的融合蛋白可以包含一个或多个另外的特征。例如,在一些实施方案中,融合蛋白可以包含抑制剂,细胞质定位序列,输出序列(诸如核输出序列)或其他定位序列,以及可用于溶解、纯化或检测融合蛋白的序列标签。本文提供的合适的蛋白质标签包括但不限于生物素羧化酶载剂蛋白(BCCP)标签、myc标签、钙调蛋白标签、FLAG标签、血凝素(HA)标签、多组氨酸标签(也称为组氨酸标签或His标签)、麦芽糖结合蛋白(MBP)标签、nus标签、谷胱甘肽-S-转移酶(GST)标签、绿色荧光蛋白(GFP)标签、硫氧还蛋白标签、S标签、Softag(例如,Softag 1、Softag 3)、链球菌标签、生物素连接酶标签、FlAsH标签、V5标签和SBP标签。另外的合适的序列对于本领域技术人员将是显而易见的。在一些实施方案中,融合蛋白包含一个或多个His标签。

[0849] 示例性但非限制性融合蛋白描述于国际PCT申请号PCT/2017/044935、PCT/US2019/044935和PCT/US2020/016288,其中的每一篇均以引用方式整体并入本文。

[0850] 包含核定位序列(NLS)的融合蛋白

[0851] 在一些实施方案中,本文提供的融合蛋白进一步包含一个或多个(例如2个、3个、4个、5个)核靶向序列,例如核定位序列(NLS)。在一个实施方案中,使用双组分NLS。在一些实施方案中,NLS包含有助于将包含NLS的蛋白质输入到细胞核中(例如,通过核转运)的氨基酸序列。在一些实施方案中,NLS与融合蛋白的N端或C端融合。在一些实施方案中,NLS融合到nCas9结构域或dCas9结构域的C端或N端。在一些实施方案中,NLS融合到Cas12结构域的N端或C端。在一些实施方案中,NLS融合到胞苷或腺苷脱氨酶的N端或C端。在一些实施方案中,NLS经由一个或多个接头与融合蛋白融合。在一些实施方案中,NLS在没有接头的情况下与融合蛋白融合。在一些实施方案中,NLS包含本文提供或引用的NLS序列中的任何一种NLS序列的氨基酸序列。另外的核定位序列在本领域中是已知的并且对于技术人员来说将是显而易见的。例如,NLS序列描述于Plank等人,PCT/EP2000/011690,其内容中公开示例性核定位序列的内容以引用方式并入本文。在一些实施方案中,NLS包含氨基酸序列PKKKRKVEGADKRTADGSEFESPKKKRKV(SEQ ID NO:83)、KRTADGSEFESPKKKRKV(SEQ ID NO:84)、KRPAATKKAG QAKKKK(SEQ ID NO:85)、KKTELQTTNAENKTKKL(SEQ ID NO:86)、KRGINDRNFWRGENGRKTR(SEQ ID NO:87)、RKSG KIAAIVVKRPRKPKKKRKV(SEQ ID NO:1424),或MDSLLMNR RKFLYQFKNVRWAKGRRETYLC(SEQ ID NO:90)。

[0852] 在一些实施方案中,包含胞苷或腺苷脱氨酶、Cas9结构域和NLS的融合蛋白不包含接头序列。在一些实施方案中,存在在该结构域或蛋白质中的一个或多个结构域或蛋白质(例如,胞苷或腺苷脱氨酶、Cas9结构域或NLS)之间的接头序列。在一些实施方案中,接头存在于胞苷脱氨酶和腺苷脱氨酶结构域与napDNAbp之间。在一些实施方案中,下方通用架构中使用的“-”指示存在任选的接头。在一些实施方案中,胞苷脱氨酶和腺苷脱氨酶和napDNAbp经由本文提供的接头中的任何接头融合。例如,在一些实施方案中,胞苷脱氨酶和腺苷脱氨酶和napDNAbp经由本文提供的接头中的任何接头融合。

[0853] 在一些实施方案中,具有胞苷或腺苷脱氨酶和napDNAbp(例如,Cas9或Cas12)结构域的示例性napDNAbp(例如,Cas9或Cas12)融合蛋白的一般架构包含以下结构中的任何一种,其中NLS是核定位序列(例如,本文提供的任何NLS),NH_2是融合蛋白的N端,并且COOH是融合蛋白的C端:

[0854] NH_2-NLS-[胞苷脱氨酶]-[napDNAbp结构域]-COOH;

[0855] NH_2-NLS[napDNAbp结构域]-[胞苷脱氨酶]-COOH;

[0856] NH_2-[胞苷脱氨酶]-[napDNAbp结构域]-NLS-COOH;

[0857] NH_2-[napDNAbp结构域]-[胞苷脱氨酶]-NLS-COOH;

[0858] NH_2-NLS-[腺苷脱氨酶]-[napDNAbp结构域]-COOH;

[0859] NH_2-NLS[napDNAbp结构域]-[腺苷脱氨酶]-COOH;

[0860] NH_2-[腺苷脱氨酶]-[napDNAbp结构域]-NLS-COOH;

[0861] NH_2-[napDNAbp结构域]-[腺苷脱氨酶]-NLS-COOH;

[0862] NH_2-NLS-[胞苷脱氨酶]-[napDNAbp结构域]-[腺苷脱氨酶]-COOH;

[0863] NH_2-NLS-[腺苷脱氨酶]-[napDNAbp结构域]-[胞苷脱氨酶]-COOH;

[0864] NH_2-NLS-[腺苷脱氨酶][胞苷脱氨酶]-[napDNAbp结构域]-COOH;

[0865] NH_2-NLS-[胞苷脱氨酶]-[腺苷脱氨酶]-[napDNAbp结构域]-COOH;

[0866] NH_2-NLS-[napDNAbp结构域]-[腺苷脱氨酶]-[胞苷脱氨酶]-COOH;

[0867] NH_2-NLS-[napDNAbp结构域]-[胞苷脱氨酶]-[腺苷脱氨酶]-COOH;

[0868] NH_2-[胞苷脱氨酶]-[napDNAbp结构域]-[腺苷脱氨酶]-NLS-COOH;

[0869] NH_2-[腺苷脱氨酶]-[napDNAbp结构域]-[胞苷脱氨酶]-NLS-COOH;

[0870] NH_2-[腺苷脱氨酶][胞苷脱氨酶]-[napDNAbp结构域]-NLS-COOH;

[0871] NH_2-[胞苷脱氨酶]-[腺苷脱氨酶]-[napDNAbp结构域]-NLS-COOH;

[0872] NH_2-[napDNAbp结构域]-[腺苷脱氨酶]-[胞苷脱氨酶]-NLS-COOH;或

[0873] NH_2-[napDNAbp结构域]-[胞苷脱氨酶]-[腺苷脱氨酶]-NLS-COOH。在一些实施方案中,NLS存在于接头中或者NLS侧接有接头,例如本文所述。双组分NLS包含两个碱性氨基酸簇,这两个碱性氨基酸簇由相对短的间隔区序列分隔(因此双组分-2个部分,而单组分(monopartite)NLS不是)。核质蛋白的NLS,KR[PAATKKAGQA]KKKK(SEQ ID NO:85),是普遍存在的双组分信号:两个由约10个氨基酸的间隔区分隔的碱性氨基酸簇的原型。示例性双组分NLS的序列如下:

[0874] PKKKRKVEGADKRTADGSEFESPKKKRKV(SEQ ID NO:83)

[0875] 可以使用编码包含一个或多个核定位序列(NLS)的CRISPR酶的载体。例如,可以使用或使用约1个、2个、3个、4个、5个、6个、7个、8个、9个、10个NLS。CRISPR酶可以包含位于氨

基端处或附近的NLS、位于羧基端处或附近的约或多于约1个、2个、3个、4个、5个、6个、7个、8个、9个、10个NLS,或它们的任何组合(例如,位于氨基端的一个或多个NLS和位于羧基端的一个或多个NLS)。当存在多于一个NLS时,每个NLS均可以独立于其他NLS来选择,使得单个NLS可以存在于多于一个拷贝中和/或与一个或多个其他NLS组合存在于一个或多个拷贝中。

[0876] 该方法中使用的CRISPR酶可以包含约6个NLS。当离NLS最近的氨基酸位于沿着多肽链的距N端或C端的约50个氨基酸范围内(例如,在1个、2个、3个、4个、5个、10个、15个、20个、25个、30个、40个或50个氨基酸范围内)时,NLS被认为靠近N端或C端。

[0877] 另外的结构域

[0878] 本文所述的碱基编辑器可以包含有助于促进多核苷酸的核碱基编辑、修饰或改变的任何结构域。在一些实施方案中,碱基编辑器包含多核苷酸可编程核苷酸结合结构域(例如,Cas9)、核碱基编辑结构域(例如,脱氨酶结构域)和一个或多个另外的结构域。在一些实施方案中,另外的结构域可以促进碱基编辑器的酶或催化功能、碱基编辑器的结合功能,或者是可能干扰所需碱基编辑结果的细胞机制抑制剂(例如,酶)。在一些实施方案中,碱基编辑器可以包含核酸酶、切口酶、重组酶、脱氨酶、甲基转移酶、甲基化酶、乙酰化酶、乙酰转移酶、转录激活子或转录阻抑因子结构域。

[0879] 在一些实施方案中,碱基编辑器可以包含尿嘧啶糖苷酶抑制剂(UGI)结构域。在一些实施方案中,对存在U:G异源双链DNA的细胞DNA修复反应可以导致细胞中核碱基编辑效率的降低。在此类实施方案中,尿嘧啶DNA糖苷酶(UDG)可以催化从细胞中的DNA中去除U,这可以启动主要导致U:G对逆转为C:G对的碱基切除修复(BER)。在此类实施方案中,可以在包含结合单链、阻断编辑的碱基、抑制UGI、抑制BER、保护编辑的碱基以及/或者促进未经编辑的链的修复的一个或多个结构域的碱基编辑器中抑制BER。因此,本公开设想了包含UGI结构域的碱基编辑器融合蛋白。

[0880] 在一些实施方案中,碱基编辑器包含作为结构域的全部或一部分双链断裂(DSB)结合蛋白。例如,DSB结合蛋白可以包含噬菌体Mu的Gam蛋白,该Gam蛋白可以结合到DSB的末端并且可以保护DSB免于降解。参见Komor,A.C.等人,“Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity”Science Advances3:eaao4774(2017),其全部内容以引用方式并入本文。

[0881] 此外,在一些实施方案中,Gam蛋白可以融合到碱基编辑器的N端。在一些实施方案中,Gam蛋白可以融合到碱基编辑器的C端。噬菌体Mu的Gam蛋白可以结合到双链断链(DSB)的末端并保护双链断链免于降解。在一些实施方案中,使用Gam结合DSB的自由末端可以减少碱基编辑过程中的插入/缺失形成。在一些实施方案中,174个残基的Gam蛋白融合到碱基编辑器的N端。参见Komor,A.C.等人,“Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity”Science Advances 3:eaao4774(2017)。在一些实施方案中,一个或多个突变可以相对于野生型结构域改变碱基编辑器结构域的长度。例如,至少一个结构域中的至少一个氨基酸的缺失可以减少碱基编辑器的长度。在另一种情况下,一个或多个突变不会相对于野生型结构域改变结构域的长度。例如,任何结构域中的取代都不会改变碱基编辑器的长度。

[0882] 此类碱基编辑器的非限制性实例(其中所有结构域的长度与野生型结构域相同)可以包括:

[0883] NH2-[核碱基编辑结构域]-接头1-[APOBEC1]-接头2-[核碱基编辑结构域]-[COOH];

[0884] NH2-[核碱基编辑结构域]-接头1-[APOBEC1]-[核碱基编辑结构域]-COOH;

[0885] NH2-[核碱基编辑结构域]-[APOBEC1]-接头2-[核碱基编辑结构域]-COOH;

[0886] NH2-[核碱基编辑结构域]-[APOBEC1]-[核碱基编辑结构域]-COOH;

[0887] NH2-[核碱基编辑结构域]-接头1-[APOBEC1]-接头2-[核碱基编辑结构域]-[UGI]-COOH;

[0888] NH2-[核碱基编辑结构域]-接头1-[APOBEC1]-[核碱基编辑结构域]-[UGI]-COOH;

[0889] NH2-[核碱基编辑结构域]-[APOBEC1]-接头2-[核碱基编辑结构域]-[UGI]-COOH;

[0890] NH2-[核碱基编辑结构域]-[APOBEC1]-[核碱基编辑结构域]-[UGI]-COOH;

[0891] NH2-[UGI]-[核碱基编辑结构域]-接头1-[APOBEC1]-接头2-[核碱基编辑结构域]-COOH;

[0892] NH2-[UGI]-[核碱基编辑结构域]-接头1-[APOBEC1]-[核碱基编辑结构域]-COOH;

[0893] NH2-[UGI]-[核碱基编辑结构域]-[APOBEC1]-接头2-[核碱基编辑结构域]-COOH;或

[0894] NH2-[UGI]-[核碱基编辑结构域]-[APOBEC1]-[核碱基编辑结构域]-COOH。

[0895] F.碱基编辑器系统

[0896] 本文提供了用于使用碱基编辑器系统编辑核碱基的系统、组合物和方法。在一些实施方案中,碱基编辑器系统包含(1)碱基编辑器(BE),该碱基编辑器包含多核苷酸可编程核苷酸结合结构域和用于编辑核碱基的核碱基编辑结构域(例如脱氨酶结构域);和(2)与该多核苷酸可编程核苷酸结合结构域联合的引导多核苷酸(例如,引导RNA)。在一些实施方案中,碱基编辑器系统是胞苷碱基编辑器(CBE)或腺苷碱基编辑器(ABE)。在一些实施方案中,多核苷酸可编程核苷酸结合结构域是多核苷酸可编程DNA或RNA结合结构域。在一些实施方案中,核碱基编辑结构域是脱氨酶结构域。在一些实施方案中,脱氨酶结构域可以是胞苷脱氨酶或胞嘧啶脱氨酶。在一些实施方案中,脱氨酶结构域可以是腺嘌呤脱氨酶或腺苷脱氨酶。在一些实施方案中,腺苷碱基编辑器可以使DNA中的腺嘌呤脱氨。在一些实施方案中,碱基编辑器能够使DNA中的胞苷脱氨。

[0897] 在一些实施方案中,如本文提供的碱基编辑系统提供了用于基因组编辑的新方法,该方法使用含有催化缺陷的酿脓链球菌Cas9、脱氨酶(例如,胞苷或腺苷脱氨酶)和碱基切除修复抑制剂的融合蛋白诱导DNA中可编程的单核苷酸(C→T或A→G)变化而不会产生双链DNA断裂,不需要供体DNA模板,并且不会诱导过多的随机插入和缺失。

[0898] 核碱基编辑蛋白的细节描述于国际PCT申请号PCT/2017/045381(WO2018/027078)和PCT/US2016/058344(WO2017/070632),其中的每一篇都以引用方式整体并入本文。另外参见Komor,A.C.等人,“Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage”Nature 533,420-424(2016);Gaudelli,N.M.等人,“Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage”Nature 551,464-471(2017);以及Komor,A.C.等人,“Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base

editors with higher efficiency and product purity"Science Advances 3:eaao4774 (2017),其全部内容据此以引用方式并入本文。

[0899] 本文提供的碱基编辑器系统的使用包括以下步骤:(a)使受试者的多核苷酸(例如双链或单链DNA或RNA)的靶核苷酸序列与包含核碱基编辑器(例如,腺苷碱基编辑器或胞苷碱基编辑器)和引导多核酸(例如gRNA)的碱基编辑器系统接触,其中该靶核苷酸序列包含被靶向的核碱基对;(b)诱导所述靶区域的链分离;(c)将靶区域单链中的所述靶核碱基对的第一核碱基转化为第二核碱基;以及(d)切割不超过一条所述靶区域的链,其中与第一核碱基互补的第三核碱基被与第二核碱基互补的第四核碱基替换。应当理解,在一些实施方案中,步骤(b)被省略。在一些实施方案中,所述被靶向的核碱基对是一个或多个基因中的多个核碱基对。在一些实施方案中,本文提供的碱基编辑器系统能够对一个或多个基因中的多个核碱基对进行多重编辑。在一些实施方案中,多个核碱基对位于同一基因中。在一些实施方案中,多个核碱基对位于一个或多个基因中,其中至少一个基因位于不同的基因座中。

[0900] 在一些实施方案中,切口单链(切口链)与引导核酸杂交。在一些实施方案中,切口单链与包含第一核碱基的链相反。在一些实施方案中,碱基编辑器包含Cas9结构域。在一些实施方案中,第一碱基是腺嘌呤,并且第二碱基不是G、C、A或T。在一些实施方案中,第二碱基是肌苷。

[0901] 在一些实施方案中,可以利用单引导多核苷酸使脱氨酶靶向于靶核酸序列。在一些实施方案中,可以利用单对引导多核苷酸使不同的脱氨酶靶向于靶核酸序列。

[0902] 碱基编辑器系统的核碱基组分和多核苷酸可编程核苷酸结合组分可以彼此共价或非共价缔合。例如,在一些实施方案中,多核苷酸可编程核苷酸结合结构域可以使脱氨酶结构域靶向于靶核苷酸序列。在一些实施方案中,多核苷酸可编程核苷酸结合结构域可以与脱氨酶结构域融合或连接。在一些实施方案中,多核苷酸可编程核苷酸结合结构域可以通过与脱氨酶结构域非共价相互作用或缔合来使脱氨酶结构域靶向于靶核苷酸序列。例如,在一些实施方案中,核碱基编辑组分,例如脱氨酶组分可以包含另外的异源部分或结构域,该另外的异源部分或结构域能够与作为多核苷酸可编程核苷酸结合结构域的一部分的另外的异源部分或结构域相互作用、缔合,或能够与作为多核苷酸可编程核苷酸结合结构域的一部分的另外的异源部分或结构域形成复合物。在一些实施方案中,另外的异源部分可能能够与多肽结合、相互作用、缔合,或与多肽形成复合物。在一些实施方案中,另外的异源部分可能能够与多肽结合、相互作用、缔合,或与多肽形成复合物。在一些实施方案中,另外的异源部分可能能够结合至引导多核苷酸。在一些实施方案中,另外的异源部分可能能够结合至多肽接头。在一些实施方案中,另外的异源部分可能能够结合至多核苷酸接头。该另外的异源部分可以是蛋白质结构域。在一些实施方案中,另外的异源部分可以是K同源性(KH)结构域、MS2外壳蛋白结构域、PP7外壳蛋白结构域、SfMu Com外壳蛋白结构域、不育α基序、端粒酶Ku结合基序和Ku蛋白、端粒酶Sm7结合基序和Sm7蛋白,或RNA识别基序。

[0903] 碱基编辑器系统可以进一步包含引导多核苷酸组分。应当理解,碱基编辑器系统的组分可以经由共价键、非共价相互作用,或其缔合和相互作用的任何组合相互缔合。在一些实施方案中,引导多核苷酸可以使脱氨酶结构域靶向于靶核苷酸序列。例如,在一些实施方案中,碱基编辑器系统的核碱基编辑组分,例如脱氨酶组分,可以包含能够与引导多核苷酸的部分或节段(例如,多核苷酸基序)相互作用、缔合,或能够与引导多核苷酸的部分或节

段(例如,多核苷酸基序)形成复合物的另外的异源部分或结构域(例如,多核苷酸结合结构域诸如RNA或DNA结合蛋白)。在一些实施方案中,另外的异源部分或结构域(例如,多核苷酸结合结构域,诸如RNA或DNA结合蛋白)可以与脱氨酶结构域融合或连接。在一些实施方案中,另外的异源部分可能能够与多肽结合、相互作用、缔合,或与多肽形成复合物。在一些实施方案中,另外的异源部分可能能够与多肽结合、相互作用、缔合,或与多肽形成复合物。在一些实施方案中,另外的异源部分可能能够结合至引导多核苷酸。在一些实施方案中,另外的异源部分可能能够结合至多肽接头。在一些实施方案中,另外的异源部分可能能够结合至多核苷酸接头。该另外的异源部分可以是蛋白质结构域。在一些实施方案中,另外的异源部分可以是K同源性(KH)结构域、MS2外壳蛋白结构域、PP7外壳蛋白结构域、SfMu Com外壳蛋白结构域、不育α基序、端粒酶Ku结合基序和Ku蛋白、端粒酶Sm7结合基序和Sm7蛋白,或RNA识别基序。

[0904] 在一些实施方案中,碱基编辑器系统可以进一步包含碱基切除修复(BER)组分的抑制剂。应当理解,碱基编辑器系统的组分可以经由共价键、非共价相互作用,或它们的缔合和相互作用的任何组合相互缔合。BER组分的抑制剂可以包括碱基切除修复抑制剂。在一些实施方案中,碱基切除修复抑制剂可以是尿嘧啶DNA糖苷酶抑制剂(UGI)。在一些实施方案中,碱基切除修复抑制剂可以是肌苷碱基切除修复抑制剂。在一些实施方案中,多核苷酸可编程核苷酸结合结构域可以使碱基切除修复抑制剂靶向于靶核苷酸序列。在一些实施方案中,多核苷酸可编程核苷酸结合结构域可以与碱基切除修复抑制剂融合或连接。在一些实施方案中,多核苷酸可编程核苷酸结合结构域可以与脱氨酶结构域和碱基切除修复抑制剂融合或连接。在一些实施方案中,多核苷酸可编程核苷酸结合结构域可以通过与碱基切除修复抑制剂非共价相互作用或缔合使碱基切除修复抑制剂靶向于靶核苷酸序列。例如,在一些实施方案中,碱基切除修复抑制剂组分可以包含另外的异源部分或结构域,该异源部分或结构域能够与另外的异源部分或结构域相互作用、缔合或者能够与另外的异源部分或结构域形成复合物,该另外的异源部分或结构域是多核苷酸可编程核苷酸结合结构域的一部分。在一些实施方案中,引导多核苷酸可以使碱基切除修复抑制剂靶向于靶核苷酸序列。例如,在一些实施方案中,碱基切除修复抑制剂可以包含能够与引导多核苷酸的部分或节段(例如,多核苷酸基序)相互作用、缔合或者能够与引导多核苷酸的部分或节段(例如,多核苷酸基序)形成复合物的另外的异源部分或结构域(例如,多核苷酸结合结构域诸如RNA或DNA结合蛋白)。在一些实施方案中,引导多核苷酸的另外的异源部分或结构域(例如,多核苷酸结合结构域,诸如RNA或DNA结合蛋白)可以与碱基切除修复抑制剂融合或连接。在一些实施方案中,另外的异源部分可能能够与多肽结合、相互作用、缔合,或与多肽形成复合物。在一些实施方案中,另外的异源部分可能能够结合至引导多核苷酸。在一些实施方案中,另外的异源部分可能能够结合至多肽接头。在一些实施方案中,另外的异源部分可能能够结合至多核苷酸接头。该另外的异源部分可以是蛋白质结构域。在一些实施方案中,另外的异源部分可以是K同源性(KH)结构域、MS2外壳蛋白结构域、PP7外壳蛋白结构域、SfMu Com外壳蛋白结构域、不育α基序、端粒酶Ku结合基序和Ku蛋白、端粒酶Sm7结合基序和Sm7蛋白,或RNA识别基序。

[0905] 在一些实施方案中,碱基编辑器抑制经编辑链的碱基切除修复(BER)。在一些实施方案中,碱基编辑器保护或结合未经编辑的链。在一些实施方案中,碱基编辑器包括UGI活

性。在一些实施方案中,碱基编辑器包含无催化活性的肌苷特异性核酸酶。在一些实施方案中,碱基编辑器包括切口酶活性。在一些实施方案中,碱基对的预期编辑位于PAM位点的上游。在一些实施方案中,碱基对的预期编辑位于PAM位点上游1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个或20个核苷酸处。在一些实施方案中,碱基对的预期编辑位于PAM位点的下游。在一些实施方案中,碱基对的预期编辑位于PAM位点下游1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个或20个核苷酸处。

[0906] 在一些实施方案中,方法不需要规范(例如,NGG)PAM位点。在一些实施方案中,核碱基编辑器包含接头或间隔区。在一些实施方案中,接头或间隔区的长度为1-25个氨基酸。在一些实施方案中,接头或间隔区的长度为5-20个氨基酸。在一些实施方案中,接头或间隔区的长度为10个、11个、12个、13个、14个、15个、16个、17个、18个、19个或20个氨基酸。

[0907] 在一些实施方案中,本文提供的碱基编辑融合蛋白需要定位在精确区位处,例如,在该精确区位中靶碱基被放置在限定区域(例如,“脱氨窗口”)内。在一些实施方案中,靶标可以位于4个碱基区域内。在一些实施方案中,此种限定靶区域可以为PAM上游的大约15个碱基。参见Komor,A.C.等人,“Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage”Nature 533,420-424(2016);Gaudelli,N.M.等人,“Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage”Nature 551,464-471(2017);以及Komor,A.C.等人,“Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity”Science Advances3:eaao4774(2017),其全部内容据此以引用方式并入本文。

[0908] 在一些实施方案中,靶区域包含靶窗口,其中该靶窗口包含靶核碱基对。在一些实施方案中,靶窗口包含1-10个核苷酸。在一些实施方案中,靶窗口长度为1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个或20个核苷酸。在一些实施方案中,碱基对的预期编辑位于靶窗口内。在一些实施方案中,靶窗口包括碱基对的预期编辑。在一些实施方案中,使用本文提供的碱基编辑器中的任何碱基编辑器来执行该方法。在一些实施方案中,靶窗口是脱氨窗口。脱氨窗口可以是碱基编辑器作用于靶核苷酸并使靶核苷酸脱氨的限定区域。在一些实施方案中,脱氨窗口位于2个、3个、4个、5个、6个、7个、8个、9个或10个碱基区域内。在一些实施方案中,脱氨窗口是PAM上游的5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个或25个碱基。

[0909] 本公开的碱基编辑器可以包含促进靶多核苷酸序列的编辑的任何结构域、特征或氨基酸序列。例如,在一些实施方案中,碱基编辑器包含核定位序列(NLS)。在一些实施方案中,碱基编辑器的NLS位于脱氨酶结构域和多核苷酸可编程核苷酸结合结构域之间。在一些实施方案中,碱基编辑器的NLS位于多核苷酸可编程核苷酸结合结构域的C端。

[0910] 可以存在于如本文公开的碱基编辑器中的其他示例性特征是定位序列,诸如细胞质定位序列、输出序列(诸如核输出序列),或其他定位序列,以及可用于溶解、纯化或检测融合蛋白的序列标签。本文提供的合适的蛋白质标签包括但不限于生物素羧化酶载剂蛋白(BCCP)标签、myc标签、钙调蛋白标签、FLAG标签、血凝素(HA)标签、多组氨酸标签(也称为组

氨酸标签或His标签)、麦芽糖结合蛋白(MBP)标签、nus标签、谷胱甘肽-S-转移酶(GST)标签、绿色荧光蛋白(GFP)标签、硫氧还蛋白标签、S标签、Softag(例如,Softag 1、Softag 3)、链球菌标签、生物素连接酶标签、FlAsH标签、V5标签和SBP标签。另外的合适的序列对于本领域技术人员将是显而易见的。在一些实施方案中,融合蛋白包含一个或多个His标签。

[0911] 在一些实施方案中,非限制性示例性胞苷碱基编辑器(CBE)包括BE1(APOBEC1-XTEN-dCas9)、BE2(APOBEC1-XTEN-dCas9-UGI)、BE3(APOBEC1-XTEN-dCas9(A840H)-UGI)、BE3-Gam、saBE3、saBE4-Gam、BE4、BE4-Gam、saBE4,或saB4E-Gam。BE4将APOBEC1-Cas9n(D10A)接头扩展到32个氨基酸,将Cas9n-UGI接头扩展到9个氨基酸,并且将UGI的第二个拷贝附连到构建体的C端,并将另一个9个氨基酸的接头附连到单个碱基编辑器构建体中。碱基编辑器saBE3和saBE4将酿脓链球菌Cas9n(D10A)替换为较小的金黄色葡萄球菌Cas9n(D10A)。BE3-Gam、saBE3-Gam、BE4-Gam和saBE4-Gam具有经由16个氨基酸的XTEN接头与BE3、saBE3、BE4和saBE4的N端融合的174个Gam蛋白残基。

[0912] 在一些实施方案中,腺苷碱基编辑器(ABE)可以使DNA中的腺嘌呤脱氨。在一些实施方案中,ABE是通过用天然的或经工程化的大肠埃希氏菌TadA、人类ADAR2、小鼠ADA或人类ADAT2替换BE3的APOBEC1组分而产生的。在一些实施方案中,ABE包含进化的TadA变体。在一些实施方案中,ABE是ABE1.2(TadA*-XTEN-nCas9-NLS)。在一些实施方案中,TadA*包含A106V和D108N突变。

[0913] 在一些实施方案中,ABE是第二代ABE。在一些实施方案中,ABE是ABE2.1,其在TadA*(TadA*2.1)中包含另外的突变D147Y和E155V。在一些实施方案中,ABE是与人类烷基腺嘌呤DNA糖苷酶的经催化灭活的型式(具有E125Q突变的AAG)融合的ABE2.2、ABE2.1。在一些实施方案中,ABE是与大肠埃希氏菌Endo V的经催化灭活的型式(用D35A突变灭活)融合的ABE2.3、ABE2.1。在一些实施方案中,ABE是ABE2.6,其具有长度为ABE2.1中的接头的两倍(32个氨基酸,$(SGGS)_2$(SEQ ID NO:1425)-XTEN-$(SGGS)_2$(SEQ ID NO:1425))的接头。在一些实施方案中,ABE是ABE2.7,其是与另外的野生型TadA单体栓连的ABE2.1。在一些实施方案中,ABE是ABE2.8,其是与另外的TadA*2.1单体栓连的ABE2.1。在一些实施方案中,ABE是ABE2.9,其是进化的TadA(TadA*2.1)与ABE2.1的N端的直接融合。在一些实施方案中,ABE是ABE2.10,其是野生型TadA与ABE2.1的N端的直接融合。在一些实施方案中,ABE是ABE2.11,其是在TadA*单体的N端具有灭活性E59A突变的ABE2.9。在一些实施方案中,ABE是ABE2.12,其是在内部TadA*单体中具有灭活性E59A突变的ABE2.9。

[0914] 在一些实施方案中,ABE是第三代ABE。在一些实施方案中,ABE是ABE3.1,其是具有三个另外的TadA突变(L84F、H123Y和I156F)的ABE2.3。

[0915] 在一些实施方案中,ABE是第四代ABE。在一些实施方案中,ABE是ABE4.3,其是具有另外的TadA突变A142N(TadA*4.3)的ABE3.1。

[0916] 在一些实施方案中,ABE是第五代ABE。在一些实施方案中,ABE是ABE5.1,其是通过将来自存活克隆的一组共有突变(H36L、R51L、S146C和K157N)输入到ABE3.1中而产生的。在一些实施方案中,ABE是ABE5.3,其具有含有与内部进化的TadA*融合的野生型大肠埃希氏菌TadA的异二聚体构建体。在一些实施方案中,ABE是ABE5.2、ABE5.4、ABE5.5、ABE5.6、ABE5.7、ABE5.8、ABE5.9、ABE5.10、ABE5.11、ABE5.12、ABE5.13或ABE5.14,如下表10所示。在一些实施方案中,ABE是第六代ABE。在一些实施方案中,ABE是ABE6.1、ABE6.2、ABE6.3、

ABE6.4、ABE6.5或ABE6.6,如下表10所示。在一些实施方案中,ABE是第七代ABE。在一些实施方案中,ABE是ABE7.1、ABE7.2、ABE7.3、ABE7.4、ABE7.5、ABE7.6、ABE7.7、ABE7.8、ABE7.9或ABE7.10,如如下表10所示。

[0917] 表10.ABE的基因型

[0918]

	23	26	36	37	48	49	51	72	84	87	106	108	123	125	142	146	147	152	155	156	157	161
ABE0.1	W	R	H	N	P		R	N	L	S	A	D	H	G	A	S	D	R	E	I	K	K
ABE0.2	W	R	H	N	P		R	N	L	S	A	D	H	G	A	S	D	R	E	I	K	K
ABE1.1	W	R	H	N	P		R	N	L	S	A	N	H	G	A	S	D	R	E	I	K	K
ABE1.2	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	D	R	E	I	K	K
ABE2.1	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.2	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.3	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.4	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.5	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.6	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.7	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.8	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.9	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.10	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.11	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.12	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE3.1	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.2	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.3	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.4	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.5	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.6	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.7	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.8	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE4.1	W	R	H	N	P		R	N	L	S	V	N	H	G	N	S	Y	R	V	I	K	K
ABE4.2	W	G	H	N	P		R	N	L	S	V	N	H	G	N	S	Y	R	V	I	K	K
ABE4.3	W	R	H	N	P		R	N	F	S	V	N	Y	G	N	S	Y	R	V	F	K	K
ABE5.1	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.2	W	R	H	S	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	T
ABE5.3	W	R	L	N	P		L	N	I	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.4	W	R	H	S	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	T
ABE5.5	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.6	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.7	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.8	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.9	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.10	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.11	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.12	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.13	W	R	H	N	P		L	D	F	S	V	N	Y	A	A	S	Y	R	V	F	K	K
ABE5.14	W	R	H	N	S		L	N	F	C	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE6.1	W	R	H	N	S		L	N	F	S	V	N	Y	G	N	S	Y	R	V	F	K	K
ABE6.2	W	R	H	N	T	V	L	N	F	S	V	N	Y	G	N	S	Y	R	V	F	N	K
ABE6.3	W	R	L	N	S		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE6.4	W	R	L	N	S		L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
ABE6.5	W	R	L	N	T	V	L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE6.6	W	R	L	N	T	V	L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
ABE7.1	W	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE7.2	W	R	L	N	A		L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
ABE7.3	L	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE7.4	R	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE7.5	W	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	H	V	F	N	K

[0919]

	23	26	36	37	48	49	51	72	84	87	106	108	123	125	142	146	147	152	155	156	157	161
ABE7.6	W	R	L	N	A		L	N	I	S	V	N	Y	G	A	C	Y	P	V	F	N	K
ABE7.7	L	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	P	V	F	N	K
ABE7.8	L	R	L	N	A		L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
ABE7.9	L	R	L	N	A		L	N	F	S	V	N	Y	G	N	C	Y	P	V	F	N	K
ABE7.10	R	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	P	V	F	N	K

[0920] 在一些实施方案中,碱基编辑器是第八代ABE(ABE8)。在一些实施方案中,ABE8含有TadA*8变体。在一些实施方案中,ABE8具有含有TadA*8变体的单体构建体(“ABE8.x-m”)。在一些实施方案中,ABE8是具有单体构建体的ABE8.1-m,该单体构建体含有具有Y147T突变的TadA*7.10(TadA*8.1)。在一些实施方案中,ABE8是具有单体构建体的ABE8.2-m,该单体构建体含有具有Y147R突变的TadA*7.10(TadA*8.2)。在一些实施方案中,ABE8是具有单体构建体的ABE8.3-m,该单体构建体含有具有Q154S突变的TadA*7.10(TadA*8.3)。在一些实施方案中,ABE8是具有单体构建体的ABE8.4-m,该单体构建体含有具有Y123H突变的TadA*7.10(TadA*8.4)。在一些实施方案中,ABE8是具有单体构建体的ABE8.5-m,该单体构建体含有具有V82S突变的TadA*7.10(TadA*8.5)。在一些实施方案中,ABE8是具有单体构建体的ABE8.6-m,该单体构建体含有具有T166R突变的TadA*7.10(TadA*8.6)。在一些实施方案中,ABE8是具有单体构建体的ABE8.7-m,该单体构建体含有具有Q154R突变的TadA*7.10(TadA*8.7)。在一些实施方案中,ABE8是具有单体构建体的ABE8.8-m,该单体构建体含有具有Y147R、Q154R和Y123H突变的TadA*7.10(TadA*8.8)。在一些实施方案中,ABE8是具有单体构建体的ABE8.9-m,该单体构建体含有具有Y147R、Q154R和I76Y突变的TadA*7.10(TadA*8.9)。在一些实施方案中,ABE8是具有单体构建体的ABE8.10-m,该单体构建体含有具有Y147R、Q154R和T166R突变的TadA*7.10(TadA*8.10)。在一些实施方案中,ABE8是具有单体构建体的ABE8.11-m,该单体构建体含有具有Y147T和Q154R突变的TadA*7.10(TadA*8.11)。在一些实施方案中,ABE8是具有单体构建体的ABE8.12-m,该单体构建体含有具有Y147T和Q154S突变的TadA*7.10(TadA*8.12)。

[0921] 在一些实施方案中,ABE8是具有单体构建体的ABE8.13-m,该单体构建体含有具有Y123H(从H123Y复原的Y123H)、Y147R、Q154R和I76Y突变的TadA*7.10(TadA*8.13)。在一些实施方案中,ABE8是具有单体构建体的ABE8.14-m,该单体构建体含有具有I76Y和V82S突变的TadA*7.10(TadA*8.14)。在一些实施方案中,ABE8是具有单体构建体的ABE8.15-m,该单体构建体含有具有V82S和Y147R突变的TadA*7.10(TadA*8.15)。在一些实施方案中,ABE8是具有单体构建体的ABE8.16-m,该单体构建体含有具有V82S、Y123H(从H123Y复原的Y123H)和Y147R突变的TadA*7.10(TadA*8.16)。在一些实施方案中,ABE8是具有单体构建体的ABE8.17-m,该单体构建体含有具有V82S和Q154R突变的TadA*7.10(TadA*8.17)。在一些实施方案中,ABE8是具有单体构建体的ABE8.18-m,该单体构建体含有具有V82S、Y123H(从H123Y复原的Y123H)和Q154R突变的TadA*7.10(TadA*8.18)。在一些实施方案中,ABE8是具有单体构建体的ABE8.19-m,该单体构建体含有具有V82S、Y123H(从H123Y复原的Y123H)、Y147R和Q154R突变的TadA*7.10(TadA*8.19)。在一些实施方案中,ABE8是具有单体构建体的ABE8.20-m,该单体构建体含有具有I76Y、V82S、Y123H(从H123Y复原的Y123H)、Y147R和Q154R突变的TadA*7.10(TadA*8.20)。在一些实施方案中,ABE8是具有单体构建体的ABE8.21-m,该单体构建体含有具有Y147R和Q154S突变的TadA*7.10(TadA*8.21)。在一些实

施方案中,ABE8是具有单体构建体的ABE8.22-m,该单体构建体含有具有V82S和Q154S突变的TadA*7.10(TadA*8.22)。在一些实施方案中,ABE8是具有单体构建体的ABE8.23-m,该单体构建体含有具有V82S和Y123H(从H123Y复原的Y123H)突变的TadA*7.10(TadA*8.23)。在一些实施方案中,ABE8是具有单体构建体的ABE8.24-m,该单体构建体含有具有V82S、Y123H(从H123Y复原的Y123H)和Y147T突变的TadA*7.10(TadA*8.24)。

[0922] 在一些实施方案中,ABE8具有含有与TadA*8变体融合的野生型大肠埃希氏菌TadA的异二聚体构建体(“ABE8.x-d”)。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.1-d,该异二聚体构建体含有与具有Y147T突变的TadA*7.10(TadA*8.1)融合的野生型大肠埃希氏菌TadA。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.2-d,该异二聚体构建体含有与具有Y147R突变的TadA*7.10(TadA*8.2)融合的野生型大肠埃希氏菌TadA。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.3-d,该异二聚体构建体含有与具有Q154S突变的TadA*7.10(TadA*8.3)融合的野生型大肠埃希氏菌TadA。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.4-d,该异二聚体构建体含有与具有Y123H突变的TadA*7.10(TadA*8.4)融合的野生型大肠埃希氏菌TadA。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.5-d,该异二聚体构建体含有与具有V82S突变的TadA*7.10(TadA*8.5)融合的野生型大肠埃希氏菌TadA。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.6-d,该异二聚体构建体含有与具有T166R突变的TadA*7.10(TadA*8.6)融合的野生型大肠埃希氏菌TadA。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.7-d,该异二聚体构建体含有与具有Q154R突变的TadA*7.10(TadA*8.7)融合的野生型大肠埃希氏菌TadA。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.8-d,该异二聚体构建体含有与具有Y147R、Q154R和Y123H突变的TadA*7.10(TadA*8.8)融合的野生型大肠埃希氏菌TadA。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.9-d,该异二聚体构建体含有与具有Y147R、Q154R和I76Y突变的TadA*7.10(TadA*8.9)融合的野生型大肠埃希氏菌TadA。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.10-d,该异二聚体构建体含有与具有Y147R、Q154R和T166R突变的TadA*7.10(TadA*8.10)融合的野生型大肠埃希氏菌TadA。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.11-d,该异二聚体构建体含有与具有Y147T和Q154R突变的TadA*7.10(TadA*8.11)融合的野生型大肠埃希氏菌TadA。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.12-d,该异二聚体构建体含有与具有Y147T和Q154S突变的TadA*7.10(TadA*8.12)融合的野生型大肠埃希氏菌TadA。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.13-d,该异二聚体构建体含有与具有Y123H(从H123Y复原的Y123H)、Y147R、Q154R和I76Y突变的TadA*7.10(TadA*8.13)融合的野生型大肠埃希氏菌TadA。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.14-d,该异二聚体构建体含有与具有I76Y和V82S突变的TadA*7.10(TadA*8.14)融合的野生型大肠埃希氏菌TadA。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.15-d,该异二聚体构建体含有与具有V82S和Y147R突变的TadA*7.10(TadA*8.15)融合的野生型大肠埃希氏菌TadA。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.16-d,该异二聚体构建体含有与具有V82S、Y123H(从H123Y复原的Y123H)和Y147R突变的TadA*7.10(TadA*8.16)融合的野生型大肠埃希氏菌TadA。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.17-d,该异二聚体构建体含有与具有V82S和Q154R突变的TadA*7.10(TadA*8.17)融合的野生型大肠埃希氏菌

TadA。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.18-d,该异二聚体构建体含有与具有V82S、Y123H(从H123Y复原的Y123H)和Q154R突变的TadA*7.10(TadA*8.18)融合的野生型大肠埃希氏菌TadA。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.19-d,该异二聚体构建体含有与具有V82S、Y123H(从H123Y复原的Y123H)、Y147R和Q154R突变的TadA*7.10(TadA*8.19)融合的野生型大肠埃希氏菌TadA。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.20-d,该异二聚体构建体含有与具有I76Y、V82S、Y123H(从H123Y复原的Y123H)、Y147R和Q154R突变的TadA*7.10(TadA*8.20)融合的野生型大肠埃希氏菌TadA。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.21-d,该异二聚体构建体含有与具有Y147R和Q154S突变的TadA*7.10(TadA*8.21)融合的野生型大肠埃希氏菌TadA。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.22-d,该异二聚体构建体含有与具有V82S和Q154S突变的TadA*7.10(TadA*8.22)融合的野生型大肠埃希氏菌TadA。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.23-d,该异二聚体构建体含有与具有V82S和Y123H(从H123Y复原的Y123H)突变的TadA*7.10(TadA*8.23)融合的野生型大肠埃希氏菌TadA。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.24-d,该异二聚体构建体含有与具有V82S、Y123H(从H123Y复原的Y123H)和Y147T突变的TadA*7.10(TadA*8.24)融合的野生型大肠埃希氏菌TadA。

[0923] 在一些实施方案中,ABE8具有含有与TadA*8变体融合的TadA*7.10的异二聚体构建体(“ABE8.x-7”)。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.1-7,该异二聚体构建体含有与具有Y147T突变的TadA*7.10(TadA*8.1)融合的TadA*7.10。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.2-7,该异二聚体构建体含有与具有Y147R突变的TadA*7.10(TadA*8.2)融合的TadA*7.10。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.3-7,该异二聚体构建体含有与具有Q154S突变的TadA*7.10(TadA*8.3)融合的TadA*7.10。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.4-7,该异二聚体构建体含有与具有Y123H突变的TadA*7.10(TadA*8.4)融合的TadA*7.10。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.5-7,该异二聚体构建体含有与具有V82S突变的TadA*7.10(TadA*8.5)融合的TadA*7.10。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.6-7,该异二聚体构建体含有与具有T166R突变的TadA*7.10(TadA*8.6)融合的TadA*7.10。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.7-7,该异二聚体构建体含有与具有Q154R突变的TadA*7.10(TadA*8.7)融合的TadA*7.10。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.8-7,该异二聚体构建体含有与具有Y147R、Q154R和Y123H突变的TadA*7.10(TadA*8.8)融合的TadA*7.10。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.9-7,该异二聚体构建体含有与具有Y147R、Q154R和I76Y突变的TadA*7.10(TadA*8.9)融合的TadA*7.10。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.10-7,该异二聚体构建体含有与具有Y147R、Q154R和T166R突变的TadA*7.10(TadA*8.10)融合的TadA*7.10。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.11-7,该异二聚体构建体含有与具有Y147T和Q154R突变的TadA*7.10(TadA*8.11)融合的TadA*7.10。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.12-7,该异二聚体构建体含有与具有Y147T和Q154S突变的TadA*7.10(TadA*8.12)融合的TadA*7.10。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.13-7,该异二聚体构建体含有与具有Y123H(从H123Y复原的Y123H)、Y147R、

Q154R和I76Y突变的TadA*7.10(TadA*8.13)融合的TadA*7.10。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.14-7,该异二聚体构建体含有与具有I76Y和V82S突变的TadA*7.10(TadA*8.14)融合的TadA*7.10。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.15-7,该异二聚体构建体含有与具有V82S和Y147R突变的TadA*7.10(TadA*8.15)融合的TadA*7.10。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.16-7,该异二聚体构建体含有与具有V82S、Y123H(从H123Y复原的Y123H)和Y147R突变的TadA*7.10(TadA*8.16)融合的TadA*7.10。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.17-7,该异二聚体构建体含有与具有V82S和Q154R突变的TadA*7.10(TadA*8.17)融合的TadA*7.10。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.18-7,该异二聚体构建体含有与具有V82S、Y123H(从H123Y复原的Y123H)和Q154R突变的TadA*7.10(TadA*8.18)融合的TadA*7.10。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.19-7,该异二聚体构建体含有与具有V82S、Y123H(从H123Y复原的Y123H)、Y147R和Q154R突变的TadA*7.10(TadA*8.19)融合的TadA*7.10。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.20-7,该异二聚体构建体含有与具有I76Y、V82S、Y123H(从H123Y复原的Y123H)、Y147R和Q154R突变的TadA*7.10(TadA*8.20)融合的TadA*7.10。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.21-7,该异二聚体构建体含有与具有Y147R和Q154S突变的TadA*7.10(TadA*8.21)融合的TadA*7.10。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.22-7,该异二聚体构建体含有与具有V82S和Q154S突变的TadA*7.10(TadA*8.22)融合的TadA*7.10。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.23-7,该异二聚体构建体含有与具有V82S和Y123H(从H123Y复原的Y123H)突变的TadA*7.10(TadA*8.23)融合的TadA*7.10。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8.24-7,该异二聚体构建体含有与具有V82S、Y123H(从H123Y复原的Y123H)和Y147T突变的TadA*7.10(TadA*8.24)融合的TadA*7.10。

[0924] 在一些实施方案中,ABE是ABE8.1-m、ABE8.2-m、ABE8.3-m、ABE8.4-m、ABE8.5-m、ABE8.6-m、ABE8.7-m、ABE8.8-m、ABE8.9-m、ABE8.10-m、ABE8.11-m、ABE8.12-m、ABE8.13-m、ABE8.14-m、ABE8.15-m、ABE8.16-m、ABE8.17-m、ABE8.18-m、ABE8.19-m、ABE8.20-m、ABE8.21-m、ABE8.22-m、ABE8.23-m、ABE8.24-m、ABE8.1-d、ABE8.2-d、ABE8.3-d、ABE8.4-d、ABE8.5-d、ABE8.6-d、ABE8.7-d、ABE8.8-d、ABE8.9-d、ABE8.10-d、ABE8.11-d、ABE8.12-d、ABE8.13-d、ABE8.14-d、ABE8.15-d、ABE8.16-d、ABE8.17-d、ABE8.18-d、ABE8.19-d、ABE8.20-d、ABE8.21-d、ABE8.22-d、ABE8.23-d或ABE8.24-d,如下表11所示。

[0925] 表11:腺苷脱氨酶碱基编辑器8(ABE8)变体

[0926]

ABE8	腺苷脱氨酶	腺苷脱氨酶描述
ABE8.1-m	TadA*8.1	单体_TadA*7.10 + Y147T
ABE8.2-m	TadA*8.2	单体_TadA*7.10 + Y147R

[0927]

ABE8.3-m	TadA*8.3	单体_TadA*7.10 + Q154S
ABE8.4-m	TadA*8.4	单体_TadA*7.10 + Y123H
ABE8.5-m	TadA*8.5	单体_TadA*7.10 + V82S
ABE8.6-m	TadA*8.6	单体_TadA*7.10 + T166R
ABE8.7-m	TadA*8.7	单体_TadA*7.10 + Q154R
ABE8.8-m	TadA*8.8	单体_TadA*7.10 + Y147R_Q154R_Y123H
ABE8.9-m	TadA*8.9	单体_TadA*7.10 + Y147R_Q154R_I76Y
ABE8.10-m	TadA*8.10	单体_TadA*7.10 + Y147R_Q154R_T166R
ABE8.11-m	TadA*8.11	单体_TadA*7.10 + Y147T_Q154R
ABE8.12-m	TadA*8.12	单体_TadA*7.10 + Y147T_Q154S
ABE8.13-m	TadA*8.13	单体_TadA*7.10 + Y123H_Y147R_Q154R_I76Y
ABE8.14-m	TadA*8.14	单体_TadA*7.10 + I76Y_V82S
ABE8.15-m	TadA*8.15	单体_TadA*7.10 + V82S_Y147R
ABE8.16-m	TadA*8.16	单体_TadA*7.10 + V82S_Y123H_Y147R
ABE8.17-m	TadA*8.17	单体_TadA*7.10 + V82S_Q154R
ABE8.18-m	TadA*8.18	单体_TadA*7.10 + V82S_Y123H_Q154R
ABE8.19-m	TadA*8.19	单体_TadA*7.10 + V82S_Y123H_Y147R_Q154R
ABE8.20-m	TadA*8.20	单体_TadA*7.10 + I76Y_V82S_Y123H_Y147R_Q154R
ABE8.21-m	TadA*8.21	单体_TadA*7.10 + Y147R_Q154S
ABE8.22-m	TadA*8.22	单体_TadA*7.10 + V82S_Q154S
ABE8.23-m	TadA*8.23	单体_TadA*7.10 + V82S_Y123H
ABE8.24-m	TadA*8.24	单体_TadA*7.10 + V82S_Y123H_Y147T
ABE8.1-d	TadA*8.1	异二聚体_(WT) + (TadA*7.10 + Y147T)
ABE8.2-d	TadA*8.2	异二聚体_(WT) + (TadA*7.10 + Y147R)
ABE8.3-d	TadA*8.3	异二聚体_(WT) + (TadA*7.10 + Q154S)
ABE8.4-d	TadA*8.4	异二聚体_(WT) + (TadA*7.10 + Y123H)
ABE8.5-d	TadA*8.5	异二聚体_(WT) + (TadA*7.10 + V82S)
ABE8.6-d	TadA*8.6	异二聚体_(WT) + (TadA*7.10 + T166R)
ABE8.7-d	TadA*8.7	异二聚体_(WT) + (TadA*7.10 + Q154R)
ABE8.8-d	TadA*8.8	异二聚体_(WT) + (TadA*7.10 + Y147R_Q154R_Y123H)
ABE8.9-d	TadA*8.9	异二聚体_(WT) + (TadA*7.10 + Y147R_Q154R_I76Y)
ABE8.10-d	TadA*8.10	异二聚体_(WT) + (TadA*7.10 + Y147R_Q154R_T166R)
ABE8.11-d	TadA*8.11	异二聚体_(WT) + (TadA*7.10 + Y147T_Q154R)

[0928]

ABE8.12-d	TadA*8.12	异二聚体_(WT)+(TadA*7.10+Y147T_Q154S)
ABE8.13-d	TadA*8.13	异二聚体_(WT)+(TadA*7.10+Y123H_Y147T_Q154R_I76Y)
ABE8.14-d	TadA*8.14	异二聚体_(WT)+(TadA*7.10+I76Y_V82S)
ABE8.15-d	TadA*8.15	异二聚体_(WT)+(TadA*7.10+V82S_Y147R)
ABE8.16-d	TadA*8.16	异二聚体_(WT)+(TadA*7.10+V82S_Y123H_Y147R)
ABE8.17-d	TadA*8.17	异二聚体_(WT)+(TadA*7.10+V82S_Q154R)
ABE8.18-d	TadA*8.18	异二聚体_(WT)+(TadA*7.10+V82S_Y123H_Q154R)
ABE8.19-d	TadA*8.19	异二聚体_(WT)+(TadA*7.10+V82S_Y123H_Y147R_Q154R)
ABE8.20-d	TadA*8.20	异二聚体_(WT)+(TadA*7.10+I76Y_V82S_Y123H_Y147R_Q154R)
ABE8.21-d	TadA*8.21	异二聚体_(WT)+(TadA*7.10+Y147R_Q154S)
ABE8.22-d	TadA*8.22	异二聚体_(WT)+(TadA*7.10+V82S_Q154S)
ABE8.23-d	TadA*8.23	异二聚体_(WT)+(TadA*7.10+V82S_Y123H)
ABE8.24-d	TadA*8.24	异二聚体_(WT)+(TadA*7.10+V82S_Y123H_Y147T)

[0929] 在一些实施方案中,ABE8是具有单体构建体的ABE8a-m,该单体构建体含有具有R26C、A109S、T111R、D119N、H122N、Y147D、F149Y、T166I和D167N突变的TadA*7.10(TadA*8a)。在一些实施方案中,ABE8是具有单体构建体的ABE8b-m,该单体构建体含有具有V88A、A109S、T111R、D119N、H122N、F149Y、T166I和D167N突变的TadA*7.10(TadA*8b)。在一些实施方案中,ABE8是具有单体构建体的ABE8c-m,该单体构建体含有具有R26C、A109S、T111R、D119N、H122N、F149Y、T166I和D167N突变的TadA*7.10(TadA*8c)。在一些实施方案中,ABE8是具有单体构建体的ABE8d-m,该单体构建体含有具有V88A、T111R、D119N和F149Y突变的TadA*7.10(TadA*8d)。在一些实施方案中,ABE8是具有单体构建体的ABE8e-m,该单体构建体含有具有A109S、T111R、D119N、H122N、Y147D、F149Y、T166I和D167N突变的TadA*7.10(TadA*8e)。

[0930] 在一些实施方案中,ABE8是具有异二聚体构建体的ABE8a-d,该异二聚体构建体含有与具有R26C、A109S、T111R、D119、H122N、Y147D、F149Y、T166I和D167N突变的TadA*7.10(TadA*8a)融合的野生型大肠埃希氏菌TadA。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8b-d,该异二聚体构建体含有与具有V88A、A109S、T111R、D119N、H122N、F149Y、T166I和D167N突变的TadA*7.10(TadA*8b)融合的野生型大肠埃希氏菌TadA。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8c-d,该异二聚体构建体含有与具有R26C、A109S、

T111R、D119N、H122N、F149Y、T166I和D167N突变的TadA*7.10(TadA*8c)融合的野生型大肠埃希氏菌TadA。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8d-d,该异二聚体构建体含有与具有V88A、T111R、D119N和F149Y突变的TadA*7.10(TadA*8d)融合的野生型大肠埃希氏菌TadA。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8e-d,该异二聚体构建体含有与具有A109S、T111R、D119N、H122N、Y147D、F149Y、T166I和D167N突变的TadA*7.10(TadA*8e)融合的野生型大肠埃希氏菌TadA。

[0931] 在一些实施方案中,ABE8是具有异二聚体构建体的ABE8a-7,该异二聚体构建体含有与具有R26C、A109S、T111R、D119、H122N、Y147D、F149Y、T166I和D167N突变的TadA*7.10(TadA*8a)融合的TadA*7.10。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8b-7,该异二聚体构建体含有与具有V88A、A109S、T111R、D119N、H122N、F149Y、T166I和D167N突变的TadA*7.10(TadA*8b)融合的TadA*7.10。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8c-7,该异二聚体构建体含有与具有R26C、A109S、T111R、D119N、H122N、F149Y、T166I和D167N突变的TadA*7.10(TadA*8c)融合的TadA*7.10。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8d-7,该异二聚体构建体含有与具有V88A、T111R、D119N和F149Y突变的TadA*7.10(TadA*8d)融合的TadA*7.10。在一些实施方案中,ABE8是具有异二聚体构建体的ABE8e-7,该异二聚体构建体含有与具有A109S、T111R、D119N、H122N、Y147D、F149Y、T166I和D167N突变的TadA*7.10(TadA*8e)融合的TadA*7.10。

[0932] 在一些实施方案中,ABE是ABE8a-m、ABE8b-m、ABE8c-m、ABE8d-m、ABE8e-m、ABE8a-d、ABE8b-d、ABE8c-d、ABE8d-d或ABE8e-d,如下表12所示。在一些实施方案中,ABE是ABE8e-m或ABE8e-d。当ABE8e与除SpCas9以外的Cas同源物(例如SaCas9、SaCas9-KKH、Cas12a同源物,例如LbCas12a、enAs-Cas12a、SpCas9-NG和环状排列变换的CP1028-SpCas9和CP1041-SpCas9)一起使用时显示出有效率的腺嘌呤碱基编辑活性和低插入/缺失形成。除了表12中针对ABE8e示出的突变外,通过将V106W取代引入到TadA结构域中减少了脱靶RNA和DNA编辑(如描述于M.Richter等人,2020,Nature Biotechnology,doi.org/10.1038/s41587-020-0453-z,其全部内容以引用方式并入本文)。

[0933] 表12:另外的腺苷脱氨酶碱基编辑器8变体。在该表中,“单体”指示包含单个TadA*7.10的ABE,该TadA*7.10包含指示的改变,并且“异二聚体”指示包含与大肠埃希氏菌TadA腺苷脱氨酶融合的TadA*7.10的ABE,该TadA*7.10包含指示的改变。

[0934]

ABE8 碱基编辑器	腺苷脱氨酶	腺苷脱氨酶描述
ABE8a-m	TadA*8a	单体_TadA*7.10 + R26C + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N
ABE8b-m	TadA*8b	单体_TadA*7.10 + V88A + A109S + T111R + D119N + H122N + F149Y + T166I + D167N
ABE8c-m	TadA*8c	单体_TadA*7.10 + R26C + A109S + T111R

[0935]

ABE8碱基编辑器	腺苷脱氨酶	腺苷脱氨酶描述
		+ D119N + H122N + F149Y + T166I + D167N
ABE8d-m	TadA*8d	单体_TadA*7.10 + V88A + T111R + D119N + F149Y
ABE8e-m	TadA*8e	单体_TadA*7.10 + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N
ABE8a-d	TadA*8a	单体_(WT) + (TadA*7.10 + R26C + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N)
ABE8b-d	TadA*8b	异二聚体_(WT) + (TadA*7.10 + V88A + A109S + T111R + D119N + H122N + F149Y + T166I + D167N)
ABE8c-d	TadA*8c	异二聚体_(WT) + (TadA*7.10 + R26C + A109S + T111R + D119N + H122N + F149Y + T166I + D167N)
ABE8d-d	TadA*8d	异二聚体_(WT) + (TadA*7.10 + V88A + T111R + D119N + F149Y)
ABE8e-d	TadA*8e	异二聚体_(WT) + (TadA*7.10 + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N)

[0936] 在一些实施方案中,碱基编辑器(例如,ABE8)通过将腺苷脱氨酶变体(例如,TadA*8)克隆到包括环状排列变换Cas9(例如,CP5或CP6)和双组分核定位序列的支架中而产生。在一些实施方案中,碱基编辑器(例如,ABE7.9、ABE7.10或ABE8)是NGC PAM CP5变体(酿脓链球菌Cas9或spVRQR Cas9)。在一些实施方案中,碱基编辑器(例如,ABE7.9、ABE7.10或ABE8)是AGA PAM CP5变体(酿脓链球菌Cas9或spVRQR Cas9)。在一些实施方案中,碱基编辑器(例如,ABE7.9、ABE7.10或ABE8)是NGC PAM CP6变体(酿脓链球菌Cas9或spVRQR Cas9)。在一些实施方案中,碱基编辑器(例如,ABE7.9、ABE7.10或ABE8)是AGA PAM CP6变体(酿脓链球菌Cas9或spVRQR Cas9)。

[0937] 在一些实施方案中,ABE具有如下表13中所示的基因型。

[0938] 表13.ABE的基因型

[0939]

	23	26	36	37	48	49	51	72	84	87	105	108	123	125	142	145	147	152	155	156	157	161
ABE 7.9	L	R	L	N	A		L	N	F	S	V	N	Y	G	N	C	Y	P	V	F	N	K
ABE 7.10	R	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	P	V	F	N	K

[0940] 如下表14所示,描述了40个ABE8的基因型。指示了ABE的进化的大肠埃希氏菌TadA部分中的残基位置。当与ABE7.10突变不同时,示出了ABE8中的突变变化。在一些实施方案

中,ABE具有如下表14中所示的ABE之一的基因型。

[0941] 表14.进化的TadA中的残基身份

[0942]

	23	36	48	51	76	82	84	106	108	123	146	147	152	154	155	156	157	166
ABE7.10	R	L	A	L	I	V	F	V	N	Y	C	Y	P	Q	V	F	N	T
ABE8.1-m												T						
ABE8.2-m												R						
ABE8.3-m														S				
ABE8.4-m										H								
ABE8.5-m						S												
ABE8.6-m																		R
ABE8.7-m														R				
ABE8.8-m										H		R		R				
ABE8.9-m					Y							R		R				
ABE8.10-m												R		R				R
ABE8.11-m												T		R				
ABE8.12-m												T		S				
ABE8.13-m					Y					H		R		R				
ABE8.14-m					Y	S												
ABE8.15-m						S						R						
ABE8.16-m						S				H		R						
ABE8.17-m						S								R				
ABE8.18-m						S				H				R				
ABE8.19-m						S				H		R		R				
ABE8.20-m					Y	S				H		R		R				
ABE8.21-m												R		S				
ABE8.22-m						S								S				
ABE8.23-m						S				H								
ABE8.24-m						S				H		T						

[0943]

ABE8.1-d											T						
ABE8.2-d											R						
ABE8.3-d													S				
ABE8.4-d									H								
ABE8.5-d						S											
ABE8.6-d																	R
ABE8.7-d													R				
ABE8.8-d									H		R		R				
ABE8.9-d					Y						R		R				
ABE8.10-d											R		R				R
ABE8.11-d											T		R				
ABE8.12-d											T		S				
ABE8.13-d					Y				H		R		R				
ABE8.14-d					Y	S											
ABE8.15-d						S					R						
ABE8.16-d						S			H		R						
ABE8.17-d						S							R				
ABE8.18-d						S			H				R				
ABE8.19-d						S			H		R		R				
ABE8.20-d					Y	S			H		R		R				
ABE8.21-d											R		S				
ABE8.22-d						S							S				
ABE8.23-d						S			H								
ABE8.24-d						S			H		T						

[0944] 在一些实施方案中,碱基编辑器是ABE8.1,其包含以下具有腺苷脱氨酶活性的序列或其片段,或者基本上由以下具有腺苷脱氨酶活性的序列或其片段组成:

[0945] ABE8.1_Y147T_CP5_NGC PAM_单体

[0946]

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEI
MALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDV
LHYPGMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD***SGGSSGGSSGSET***
PGTSESATPESSGGSSGGSEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKK
YGGFMQPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVK
KDLIIKLPKYSLFELENGRKRMLASAKFLQKGNELALPSKYVNFLYLASHYEKLKGSPEDN
EQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTL

TNLGAPRAFKYFDTTIARKEYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD*GGSGGSGG*
*SGGSGGSGGSGGM*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLI
GALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEE
DKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEK
KNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAA
KNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS
KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHL
GELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNF
[0947] EEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPA
FLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKII
KDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRL
SRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHI
ANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRI
EEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQS
FLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAER
GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFR
KDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQ
EGADKRTADGSEFESPKKKRKV (SEQ ID NO: 1426)

[0948] 在上面的序列中,纯文本表示腺苷脱氨酶序列,粗体序列表示衍生自Cas9的序列,斜体序列表示接头序列,并且加下划线的序列表示双组分核定位序列。其他ABE8序列提供于所附的序列表中(SEQ ID NO:1427-1449)。

[0949] 在一些实施方案中,碱基编辑器是第九代ABE(ABE9)。在一些实施方案中,ABE9含有TadA*9变体。ABE9碱基编辑器包含腺苷脱氨酶变体,该腺苷脱氨酶变体包含含有如本文所述的相对于ABE 7*10参考序列的改变的氨基酸序列。示例性ABE9变体列于表15中。ABE9碱基编辑器的细节描述于国际PCT申请号PCT/2020/049975,其以引用方式整体并入本文。

[0950] 表15.腺苷脱氨酶碱基编辑器9(ABE9)变体。在该表中,“单体”指示包含单个TadA*7.10的ABE,该TadA*7.10包含指示的改变,并且“异二聚体”指示包含与大肠埃希氏菌TadA腺苷脱氨酶融合的TadA*7.10的ABE,该TadA*7.10包含指示的改变。

[0951]

ABE9 描述	改变
ABE9.1_单体	E25F、V82S、Y123H、T133K、Y147R、Q154R
ABE9.2_单体	E25F、V82S、Y123H、Y147R、Q154R
ABE9.3_单体	V82S、Y123H、P124W、Y147R、Q154R
ABE9.4_单体	L51W、V82S、Y123H、C146R、Y147R、Q154R
ABE9.5_单体	P54C、V82S、Y123H、Y147R、Q154R
ABE9.6_单体	Y73S、V82S、Y123H、Y147R、Q154R
ABE9.7_单体	N38G、V82T、Y123H、Y147R、Q154R
ABE9.8_单体	R23H、V82S、Y123H、Y147R、Q154R
ABE9.9_单体	R21N、V82S、Y123H、Y147R、Q154R
ABE9.10_单体	V82S、Y123H、Y147R、Q154R、A158K
ABE9.11_单体	N72K、V82S、Y123H、D139L、Y147R、Q154R,
ABE9.12_单体	E25F、V82S、Y123H、D139M、Y147R、Q154R
ABE9.13_单体	M70V、V82S、M94V、Y123H、Y147R、Q154R
ABE9.14_单体	Q71M、V82S、Y123H、Y147R、Q154R
ABE9.15_异二聚体	E25F、V82S、Y123H、T133K、Y147R、Q154R
ABE9.16_异二聚体	E25F、V82S、Y123H、Y147R、Q154R
ABE9.17_异二聚体	V82S、Y123H、P124W、Y147R、Q154R
ABE9.18_异二聚体	L51W、V82S、Y123H、C146R、Y147R、Q154R
ABE9.19_异二聚体	P54C、V82S、Y123H、Y147R、Q154R
ABE9.2_异二聚体	Y73S、V82S、Y123H、Y147R、Q154R
ABE9.21_异二聚体	N38G、V82T、Y123H、Y147R、Q154R
ABE9.22_异二聚体	R23H、V82S、Y123H、Y147R、Q154R
ABE9.23_异二聚体	R21N、V82S、Y123H、Y147R、Q154R
ABE9.24_异二聚体	V82S、Y123H、Y147R、Q154R、A158K
ABE9.25_异二聚体	N72K、V82S、Y123H、D139L、Y147R、Q154R,
ABE9.26_异二聚体	E25F、V82S、Y123H、D139M、Y147R、Q154R
ABE9.27_异二聚体	M70V、V82S、M94V、Y123H、Y147R、Q154R
ABE9.28_异二聚体	Q71M、V82S、Y123H、Y147R、Q154R
ABE9.29_单体	E25F_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.30_单体	I76Y_V82T_Y123H_Y147R_Q154R
ABE9.31_单体	N38G_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.32_单体	N38G_I76Y_V82T_Y123H_Y147R_Q154R
ABE9.33_单体	R23H_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.34_单体	P54C_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.35_单体	R21N_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.36_单体	I76Y_V82S_Y123H_D138M_Y147R_Q154R

[0952]

ABE9.37_单体	Y72S_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.38_异二聚体	E25F_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.39_异二聚体	I76Y_V82T_Y123H_Y147R_Q154R
ABE9.40_异二聚体	N38G_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.41_异二聚体	N38G_I76Y_V82T_Y123H_Y147R_Q154R
ABE9.42_异二聚体	R23H_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.43_异二聚体	P54C_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.44_异二聚体	R21N_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.45_异二聚体	I76Y_V82S_Y123H_D138M_Y147R_Q154R
ABE9.46_异二聚体	Y72S_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.47_单体	N72K_V82S、Y123H、Y147R、Q154R
ABE9.48_单体	Q71M_V82S、Y123H、Y147R、Q154R
ABE9.49_单体	M70V、V82S、M94V、Y123H、Y147R、Q154R
ABE9.50_单体	V82S、Y123H、T133K、Y147R、Q154R
ABE9.51_单体	V82S、Y123H、T133K、Y147R、Q154R、A158K
ABE9.52_单体	M70V、Q71M、N72K、V82S、Y123H、Y147R、Q154R
ABE9.53_异二聚体	N72K_V82S、Y123H、Y147R、Q154R
ABE9.54_异二聚体	Q71M_V82S、Y123H、Y147R、Q154R
ABE9.55_异二聚体	M70V、V82S、M94V、Y123H、Y147R、Q154R
ABE9.56_异二聚体	V82S、Y123H、T133K、Y147R、Q154R
ABE9.57_异二聚体	V82S、Y123H、T133K、Y147R、Q154R、A158K
ABE9.58_异二聚体	M70V、Q71M、N72K、V82S、Y123H、Y147R、Q154R

[0953] 在一些实施方案中,碱基编辑器包含含有全部或一部分尿嘧啶糖苷酶抑制剂(UGI)的结构域。在一些实施方案中,碱基编辑器包含含有全部或一部分核酸聚合酶的结构域。在一些实施方案中,碱基编辑器可以包含作为结构域的全部或一部分核酸聚合酶(NAP)。例如,碱基编辑器可以包含全部或一部分真核生物NAP。在一些实施方案中,掺入到碱基编辑器中的NAP或其部分是DNA聚合酶。在一些实施方案中,掺入到碱基编辑器中的NAP或其部分具有跨损伤聚合酶活性。在一些实施方案中,掺入到碱基编辑器中的NAP或其部分是跨损伤DNA聚合酶。在一些实施方案中,掺入到碱基编辑器中的NAP或其部分是Rev7、Rev1复合物、聚合酶ι、聚合酶κ或聚合酶η。在一些实施方案中,掺入到碱基编辑器中的NAP或其部分是真核聚合酶α、β、γ、δ、ε、γ、η、ι、κ、λ、μ或ν组分。在一些实施方案中,掺入到碱基编辑器中的NAP或其部分包含与核酸聚合酶(例如,跨损伤DNA聚合酶)至少75%、80%、85%、90%、95%、96%、97%、98%、99%或99.5%同一的氨基酸序列。在一些实施方案中,掺入到碱基编辑器中的核酸聚合酶或其部分是跨损伤DNA聚合酶。

[0954] 在一些实施方案中,碱基编辑器的结构域可以包含多个结构域。例如,包含衍生自

Cas9的多核苷酸可编程核苷酸结合结构域的碱基编辑器可以包含对应于野生型或天然Cas9的REC叶和NUC叶的REC叶和NUC叶。在另一个实例中,基础编辑器可以包含RuvCI结构域、BH结构域、REC1结构域、REC2结构域、RuvCII结构域、L1结构域、HNH结构域、L2结构域、RuvCIII结构域、WED结构域、TOPO结构域或CTD结构域中的一者或多者。在一些实施方案中,碱基编辑器的一个或多个结构域包含相对于包含该结构域的多肽的野生型型式的突变(例如,取代、插入、缺失)。例如,多核苷酸可编程DNA结合结构域的HNH结构域可以包含H840A取代。在另一个实例中,多核苷酸可编程DNA结合结构域的RuvCI结构域可以包含D10A取代。

[0955] 本文公开的碱基编辑器的不同结构域(例如,相邻结构域)可以在使用或不使用一个或多个接头结构域(例如,XTEN接头结构域)的情况下彼此连接。在一些实施方案中,接头结构域可以是键(例如,共价键)、化学基团,或连接两个分子或部分(例如融合蛋白的两个结构域,诸如例如第一结构域(例如,Cas9衍生结构域)和第二结构域(例如,腺苷脱氨酶结构域或胞苷脱氨酶结构域))的分子。在一些实施方案中,接头是共价键(例如,碳-碳键、二硫键、碳-杂原子键等)。在某些实施方案中,接头是酰胺连键的碳氮键。在某些实施方案中,接头是环状或无环、取代或未取代、支化或非支化的脂族或杂脂族接头。在某些实施方案中,接头是聚合性的(例如,聚乙烯、聚乙二醇、聚酰胺、聚酯等)。在某些实施方案中,接头包含氨基链烷酸的单体、二聚体或聚合物。在一些实施方案中,接头包含氨基烷酸(例如,甘氨酸、乙酸、丙氨酸、β-丙氨酸、3-氨基丙酸、4-氨基丁酸、5-戊酸等)。在一些实施方案中,接头包含氨基已酸(Ahx)的单体、二聚体或聚合物。在某些实施方案中,接头基于碳环部分(例如,环戊烷、环己烷)。在其他实施方案中,接头包含聚乙二醇部分(PEG)。在某些实施方案中,接头包含芳基或杂芳基部分。在某些实施方案中,接头基于苯环。接头可以包含功能化部分以促进亲核体(例如,硫醇、氨基)从肽附接到接头。任何亲电体都可以用作接头的一部分。示例性亲电体包括但不限于经激活的酯、经激活的酰胺、迈克尔受体、卤代烷、芳基卤、酰卤和异硫氰酸酯。在一些实施方案中,接头联接RNA可编程核酸酶的gRNA结合结构域,包括Cas9核酸酶结构域和核酸编辑蛋白的催化结构域。在一些实施方案中,接头联接dCas9和第二结构域(例如,UGI等)。

[0956] 接头

[0957] 在某些实施方案中,接头可以用于连接本发明的肽或肽结构域中的任何肽或肽结构域。接头可以像共价键一样简单,或者它可以是长度为许多原子的聚合性接头。在某些实施方案中,接头是多肽或基于氨基酸。在其他实施方案中,接头不是肽样的。在某些实施方案中,接头是共价键(例如,碳-碳键、二硫键、碳-杂原子键等)。在某些实施方案中,接头是酰胺连键的碳-氮键。在某些实施方案中,接头是环状或无环、取代或未取代、支化或非支化的脂族或杂脂族接头。在某些实施方案中,接头是聚合性的(例如,聚乙烯、聚乙二醇、聚酰胺、聚酯等)。在某些实施方案中,接头包含氨基烷酸的单体、二聚体或聚合物。在某些实施方案中,接头包含氨基烷酸(例如,甘氨酸、乙酸、丙氨酸、β-丙氨酸、3-氨基丙酸、4-氨基丁酸、5-戊酸等)。在某些实施方案中,接头包含氨基已酸(Ahx)的单体、二聚体或聚合物。在某些实施方案中,接头基于碳环部分(例如,环戊烷、环己烷)。在其他实施方案中,接头包含聚乙二醇部分(PEG)。在其他实施方案中,接头包含氨基酸。在某些实施方案中,接头包含肽。在某些实施方案中,接头包含芳基或杂芳基部分。在某些实施方案中,接头基于苯环。接头可以包含功能化部分以促进亲核体(例如,硫醇、氨基)从肽附接到接头。任何亲电体都可以

用作接头的一部分。示例性亲电体包括但不限于经激活的酯、经激活的酰胺、迈克尔受体、卤代烷、芳基卤、酰卤和异硫氰酸酯。

[0958] 通常,接头位于两个基团、分子或其他部分之间或侧接有两个基团、分子或其他部分,并经由共价键连接到每一者,从而将两者连接起来。在一些实施方案中,接头是一个氨基酸或多个氨基酸(例如,肽或蛋白质)。在一些实施方案中,接头是有机分子、基团、聚合物或化学部分。在一些实施方案中,接头长度为2-100个氨基酸,例如长度为2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个、30-35个、35-40个、40-45个、45-50个、50-60个、60-70个、70-80个、80-90个、90-100个、100-150个或150-200个氨基酸。在一些实施方案中,接头长度为约3个至约104个(例如,5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22、23个、24个、25个、26个、27个、28个、29个、30个、31个、32个、33个、34个、35个、36个、37个、38个、39个、40个、41个、42个、43个、44个、45个、46个、47个、48个、49个、50个、55个、60个、65个、70个、75个、80个、85个、90个、95个或100个)氨基酸。也设想了更长或更短的接头。

[0959] 在一些实施方案中,本文提供的融合蛋白中的任何融合蛋白包含经由接头彼此融合的胞苷或腺苷脱氨酶和Cas9结构域。可以使用胞苷或腺苷脱氨酶与Cas9结构域之间的各种接头长度和柔性(例如,范围为从(GGGS)n(SEQ ID NO:1308)、(GGGGS)n(SEQ ID NO:109)和(G)n形式的非常柔性的接头到(EAAAK)n(SEQ ID NO:1309)、(SGGS)n(SEQ ID NO:57)、SGSETPGTSESATPES(SEQ ID NO:56)(参见例如,Guilinger JP等人,Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification.Nat.Biotechnol.2014;32(6):577-82;全部内容以引用方式并入本文)和(XP)n形式的更刚性的接头)以获得胞苷或腺苷脱氨酶核碱基编辑器的最佳活性长度。在一些实施方案中,n是1、2、3、4、5、6、7、8、9、10、11、12、13、14或15。在一些实施方案中,接头包含(GGS)n基序,其中n是1、3或7。在一些实施方案中,本文提供的融合蛋白中的任何融合蛋白的胞苷脱氨酶或腺苷脱氨酶和Cas9结构域经由包含氨基酸序列SGSETPGTSESATPES(SEQ ID NO:56)的接头(也可称为XTEN接头)融合。在一些实施方案中,接头包含多个脯氨酸残基并且长度为5-21个、5-14个、5-9个、5-7个氨基酸,例如PAPAP(SEQ ID NO:65)、PAPAPA(SEQ ID NO:66)、PAPAPAP(SEQ ID NO:67)、PAPAPAPA(SEQ ID NO:68)、P(AP)4(SEQ ID NO:69)、P(AP)7(SEQ ID NO:70)、P(AP)10(SEQ ID NO:71)(参见例如Tan J,Zhang F,Karcher D,Bock R.Engineering of high-precision base editors for site-specific single nucleotide replacement.Nat Commun.2019年1月25日;10(1):439;全部内容以引用方式并入本文)。此类富含脯氨酸的接头也称为“刚性”接头。

[0960] 在另一个实施方案中,碱基编辑器系统包含与脱氨酶(DNA脱氨酶)例如腺苷或胞苷脱氨酶非共价相互作用的组分(蛋白质),并将腺苷或胞苷脱氨酶瞬时吸引至靶多核苷酸序列中的靶核碱基以进行特定编辑,具有最小的或减少的旁观者或靶相邻(target-adjacent)效应。此种涉及脱氨酶相互作用蛋白的非共价系统和方法起到将DNA脱氨酶吸引到特定的基因组靶核碱基的作用,并解耦靶上和靶相邻编辑事件,从而增强更精确的单碱基取代突变的实现。在一个实施方案中,脱氨酶相互作用蛋白与脱氨酶(例如,腺苷脱氨酶或胞苷脱氨酶)结合,而不阻断或干扰脱氨酶的活性(催化)位点与靶核碱基(例如,分别为

腺苷或胞苷)接合。诸如,称为“MagnEdit”的系统包括与Cas9和gRNA复合物栓连的相互作用蛋白,并且可以吸引共表达的腺苷或胞苷脱氨酶(外源性或内源性)以编辑特定的基因组靶位点,并描述于McCann,J.等人,2020,“MagnEdit-interacting factors that recruit DNA-editing enzymes to single base targets,”Life-Science-Alliance,第3卷,第4期(e201900606),(doi 10.26508/lsa.201900606),其内容以引用方式整体并入本文。在一个实施方案中,DNA脱氨酶是如本文所述的腺苷脱氨酶变体(例如,TadA*8)。

[0961] 在另一个实施方案中,称为“Suntag”的系统涉及用于将碱基编辑器的蛋白质(例如,腺苷脱氨酶或胞苷脱氨酶)组分或其多个拷贝募集到多核苷酸靶位点以实现在此位点处具有减少的相邻靶编辑的碱基编辑的非共价相互作用组分,例如,如Tanenbaum,M.E.等人,“Aprotein tagging system for signal amplification in gene expression and fluorescence imaging,”Cell.2014年10月23日;159(3):635-646.doi:10.1016/j.cell.2014.09.039;和Huang,Y.-H.等人,2017,“DNA epigenome editing using CRISPR-Cas SunTag-directed DNMT3A,”Genome Biol 18:176.doi:10.1186/s13059-017-1306-z中所描述,其中每一篇的内容均以引用方式整体并入本文。在一个实施方案中,DNA脱氨酶是如本文所述的腺苷脱氨酶变体(例如,TadA*8)。

[0962] 具有引导RNA的核酸可编程DNA结合蛋白

[0963] 本文提供了用于在细胞中进行碱基编辑的组合物和方法。本文进一步提供了组合物,其包含引导多核酸序列,例如引导RNA序列,或如本文提供的2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个或更多个引导RNA的组合。在一些实施方案中,如本文提供的用于碱基编辑的组合物进一步包含编码碱基编辑器,例如C-碱基编辑器或A-碱基编辑器的多核苷酸。例如,用于碱基编辑的组合物可以包含编码BE、BE4、ABE和如所提供的一种或多种引导RNA的组合的mRNA序列。用于碱基编辑的组合物可以包含碱基编辑器多肽和本文提供的任何引导RNA中的一种或多种引导RNA的组合。此种组合物可以用于通过不同的递送途径(例如电穿孔、核转染、病毒转导或转染)在细胞中实现碱基编辑。在一些实施方案中,用于碱基编辑的组合物包含本文提供的用于电穿孔的编码碱基编辑器的mRNA序列和一种或多种引导RNA序列的组合。

[0964] 本公开的一些方面提供了复合物,该复合物包含本文提供的融合蛋白中的任何融合蛋白,以及与融合蛋白的核酸可编程DNA结合蛋白(napDNAbp)结构域(例如,Cas9(例如,dCas9、核酸酶有活性的Cas9,或Cas9切口酶)或Cas12)结合的引导RNA。这些复合物也称为核糖核蛋白(RNP)。在一些实施方案中,引导核酸(例如,引导RNA)为15-100个核苷酸长,并且包含与靶序列互补的至少10个连续核苷酸的序列。在一些实施方案中,引导RNA为15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个、31个、32个、33个、34个、35个、36个、37个、38个、39个、40个、41个、42个、43个、44个、45个、46个、47个、48个、49或50个核苷酸长。在一些实施方案中,引导RNA包含15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个、31个、32个、33个、34个、35个、36个、37个、38个、39个或40个与靶序列互补的连续核苷酸的序列。在一些实施方案中,靶序列是DNA序列。在一些实施方案中,靶序列是RNA序列。在一些实施方案中,靶序列是细菌、酵母、真菌、昆虫、植物或动物的基因组中的序列。在一些实施方案中,靶序列是人类基因组中的序列。在一些实施方案中,靶序列的3’末端紧邻规范PAM序列(NGG)。在一

些实施方案中,靶序列的3'末端紧邻非规范PAM序列(例如,表6中列出的序列或5'-NAA-3')。在一些实施方案中,引导核酸(例如,引导RNA)与关注的基因(例如,与疾病或病症相关的基因)中的序列互补。

[0965] 本公开的一些方面提供了使用本文提供的融合蛋白或复合物的方法。例如,本公开的一些方面提供的方法包括使DNA分子与本文提供的融合蛋白中的任何融合蛋白以及与至少一种引导RNA接触,其中引导RNA为约15-100个核苷酸长并且包含与靶序列互补的至少10个连续核苷酸的序列。在一些实施方案中,靶序列的3'末端紧邻AGC、GAG、TTT、GTG或CAA序列。在一些实施方案中,靶序列的3'末端紧邻NGA、NGCG、NGN、NNGRRT、NNNRRT、NGCG、NGCN、NGTN、NGTN、NGTN或5'(TTTV)序列。在一些实施方案中,靶序列的3'末端紧邻例如TTN、DTTN、GTTN、ATTN、ATTC、DTTNT、WTTN、HATY、TTTN、TTTV、TTTC、TG、RTR或YTN PAM位点。

[0966] 应当理解,各个序列中特定位置或残基的编号取决于所使用的特定蛋白质和编号方案。编号可能不同,例如,在成熟蛋白质的前体和成熟蛋白质本身上不同,并且不同物种之间的序列差异可能会影响编号。本领域技术人员将能够通过本领域众所周知的方法,例如通过序列比对和同源残基的测定,鉴定任何同源蛋白质和相应编码核酸中的相应残基。

[0967] 对于本领域技术人员显而易见的是,为了使本文公开的融合蛋白中的任何融合蛋白靶向于靶位点,例如包含待编辑的突变的位点,通常需要使融合蛋白与引导RNA一起共表达。如本文别处更详细解释,引导RNA通常包含允许napDNAbp(例如,Cas9或Cas12)结合的tracrRNA框架,和引导序列,该引导序列赋予napDNAbp:核酸编辑酶/结构域融合蛋白序列特异性。可替代地,可以单独提供引导RNA和tracrRNA,作为两个核酸分子。在一些实施方案中,引导RNA包含一种结构,其中引导序列包含与靶序列互补的序列。引导序列通常为20个核苷酸长。基于本公开,用于将napDNAbp:核酸编辑酶/结构域融合蛋白靶向于特定基因组靶位点的合适的引导RNA的序列对于本领域技术人员是显而易见的。此类合适的引导RNA序列通常包含与待编辑的靶核苷酸上游或下游50个核苷酸内的核酸序列互补的引导序列。本文提供了一些示例性引导RNA序列,这些引导RNA序列适用于将所提供的融合蛋白中的任何融合蛋白靶向于特定的靶序列。

[0968] 预计sgRNA的不同部分会形成与Cas9(例如,SpyCas9)和/或DNA靶标相互作用的各种特征。已在指导Cas9核酸内切酶活性的原生crRNA:tracrRNA双链体和单引导RNA(sgRNA)内鉴定出六个保守模块(参见Briner等人,Guide RNA Functional Modules Direct Cas9Activity and Orthogonality Mol Cell.2014年10月23日;56(2):333-339)。这六个模块包括负责DNA靶向的间隔区,由CRISPR重复:tracrRNA双链体形成的上位茎、凸起、下位茎,来自tracrRNA 3'末端的连结点(nexus)和发夹。上位茎和下位茎主要通过与磷酸酯骨架的序列非依赖性相互作用与Cas9相互作用。在一些实施方案中,上位茎是可有可无的。在一些实施方案中,下位茎基部处的保守尿嘧啶核苷酸序列是可有可无的。凸起参与与Cas9的Rec1结构域的特定侧链相互作用。U44的核碱基与Tyr 325和His 328的侧链相互作用,而G43与Tyr 329相互作用。连结点形成sgRNA:Cas9相互作用的核心,并且位于sgRNA与Cas9和靶DNA之间的交叉点处。A51和A52的核碱基与Phe1105的侧链相互作用;U56与Arg 457和Asn 459相互作用;U59的核碱基插入到由Arg 74、Asn 77、Pro 475、Leu 455、Phe 446和Ile 448的侧链限定的疏水口袋中;C60与Leu 455、Ala 456和Asn 459相互作用,并且C61与Arg 70的侧链相互作用,Arg 70的侧链又与C15相互作用。在一些实施方案中,这些突变中的一个

或多个突变被制作在Cas9(例如spyCas9)的sgRNA的凸起和/或连结点中以优化sgRNA:Cas9相互作用。

[0969] 此外,tracrRNA连结点和发夹对Cas9配对至关重要,并且可以被互换以跨越分离完全不同的Cas9蛋白的正交屏障,这有助于进一步利用正交Cas9蛋白。在一些实施方案中,连结点和发夹被互换以靶向正交Cas9蛋白。在一些实施方案中,sgRNA被免除了上位茎、发夹1,和/或下位茎的序列灵活性以设计更紧凑且构象稳定的引导RNA。在一些实施方案中,使用具有各种嵌合引导物的单个Cas9或通过同时使用具有不同嵌合sgRNA组合的正交系统对模块进行修饰以优化多重编辑。关于引导功能性模块及其方法的细节描述于例如,Briner等人,Guide RNA Functional Modules Direct Cas9 Activity and Orthogonality Mol Cell.2014年10月23日;56(2):333-339,其内容以引用方式整体并入本文。

[0970] 本文公开的碱基编辑器的结构域可以以任何顺序排列。包含含有例如多核苷酸可编程核苷酸结合结构域(例如,Cas9或Cas12)和脱氨酶结构域(例如,胞苷或腺苷脱氨酶)的融合蛋白的碱基编辑器的非限制性实例可以排列如下:

[0971] NH2-[核碱基编辑结构域]-接头1-[核碱基编辑结构域]-COOH;

[0972] NH2-[脱氨酶]-接头1-[核碱基编辑结构域]-COOH;

[0973] NH2-[脱氨酶]-接头1-[核碱基编辑结构域]-接头2-[UGI]-COOH;

[0974] NH2-[脱氨酶]-接头1-[核碱基编辑结构域]-COOH;

[0975] NH2-[腺苷脱氨酶]-接头1-[核碱基编辑结构域]-COOH;

[0976] NH2-[核碱基编辑结构域]-[脱氨酶]-COOH;

[0977] NH2-[脱氨酶]-[核碱基编辑结构域]-[肌苷BER抑制剂]-COOH;

[0978] NH2-[脱氨酶]-[肌苷BER抑制剂]-[核碱基编辑结构域]-COOH;

[0979] NH2-[肌苷BER抑制剂]-[脱氨酶]-[核碱基编辑结构域]-COOH;

[0980] NH2-[核碱基编辑结构域]-[脱氨酶]-[肌苷BER抑制剂]-COOH;

[0981] NH2-[核碱基编辑结构域]-[肌苷BER抑制剂]-[脱氨酶]-COOH;

[0982] NH2-[肌苷BER抑制剂]-[核碱基编辑结构域]-[脱氨酶]-COOH;

[0983] NH2-[核碱基编辑结构域]-接头1-[脱氨酶]-接头2-[核碱基编辑结构域]-COOH;

[0984] NH2-[核碱基编辑结构域]-接头1-[脱氨酶]-[核碱基编辑结构域]-COOH;

[0985] NH2-[核碱基编辑结构域]-[脱氨酶]-接头2-[核碱基编辑结构域]-COOH;

[0986] NH2-[核碱基编辑结构域]-[脱氨酶]-[核碱基编辑结构域]-COOH;

[0987] NH2-[核碱基编辑结构域]-接头1-[脱氨酶]-接头2-[核碱基编辑结构域]-[肌苷BER抑制剂]-COOH;

[0988] NH2-[核碱基编辑结构域]-接头1-[脱氨酶]-[核碱基编辑结构域]-[肌苷BER抑制剂]-COOH;

[0989] NH2-[核碱基编辑结构域]-[脱氨酶]-接头2-[核碱基编辑结构域]-[肌苷BER抑制剂]-COOH;

[0990] NH2-[核碱基编辑结构域]-[脱氨酶]-[核碱基编辑结构域]-[肌苷BER抑制剂]-COOH;

[0991] NH2-[肌苷BER抑制剂]-[核碱基编辑结构域]-接头1-[脱氨酶]-接头2-[核碱基编

辑结构域]-COOH;

[0992] NH2-[肌苷BER抑制剂]-[核碱基编辑结构域]-接头1-[脱氨酶]-[核碱基编辑结构域]-COOH;

[0993] NH2-[肌苷BER抑制剂]-[核碱基编辑结构域]-[脱氨酶]-接头2-[核碱基编辑结构域]-COOH;或

[0994] NH2-[肌苷BER抑制剂]NH2-[核碱基编辑结构域]-[脱氨酶]-[核碱基编辑结构域]-COOH。

[0995] 在一些实施方案中,本文提供的碱基编辑融合蛋白需要定位在精确区位处,例如,在该精确区位中靶碱基被放置在限定区域(例如,“脱氨窗口”)内。在一些实施方案中,靶标可以位于4个碱基区域内。在一些实施方案中,此种限定靶区域可以为PAM上游的大约15个碱基。另外参见Komor,A.C.等人,“Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage”Nature 533,420-424(2016);Gaudelli,N.M.等人,“Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage”Nature 551,464-471(2017);以及Komor,A.C.等人,“Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity”Science Advances3:eaao4774(2017),其全部内容以引用方式并入本文。

[0996] 限定的靶区域可以是脱氨窗口。脱氨窗口可以是碱基编辑器作用于靶核苷酸并使靶核苷酸脱氨的限定区域。在一些实施方案中,脱氨窗口位于2个、3个、4个、5个、6个、7个、8个、9个或10个碱基区域内。在一些实施方案中,脱氨窗口是PAM上游的5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个或25个碱基。

[0997] 本公开的碱基编辑器可以包含促进靶多核苷酸序列的编辑的任何结构域、特征或氨基酸序列。例如,在一些实施方案中,碱基编辑器包含核定位序列(NLS)。在一些实施方案中,碱基编辑器的NLS位于脱氨酶结构域和napDNAbp结构域之间。在一些实施方案中,碱基编辑器的NLS位于napDNAbp结构域的C端。

[0998] 可以包含在融合蛋白中的蛋白质结构域的非限制性实例包括脱氨酶结构域(例如,腺苷脱氨酶或胞苷脱氨酶)、尿嘧啶糖苷酶抑制剂(UGI)结构域、表位标签、报告基因序列和/或具有本文所述的活性中的一种或多种活性的蛋白质结构域。

[0999] 可以用表位标签、报告蛋白、其他结合结构域检测或标记结构域。表位标签的非限制性实例包括组氨酸(His)标签、V5标签、FLAG标签、流感血凝素(HA)标签、Myc标签、VSV-G标签和硫氧还蛋白(Trx)标签。报告基因的实例包括但不限于谷胱甘肽-5-转移酶(GST)、辣根过氧化物酶(HRP)、氯霉素乙酰转移酶(CAT)β-半乳糖苷酶、β-葡萄糖醛酸酶、荧光素酶、绿色荧光蛋白(GFP)、HcRed、DsRed、青色荧光蛋白(CFP)、黄色荧光蛋白(YFP)和自体荧光蛋白,包括蓝色荧光蛋白(BFP)。另外的蛋白质序列可以包括结合DNA分子或结合其他细胞分子的氨基酸序列,包括但不限于麦芽糖结合蛋白(MBP)、S标签、Lex A DNA结合结构域(DBD)融合体、GAL4 DNA结合结构域融合体和单纯疱疹病毒(HSV)BP16蛋白融合体。

[1000] 使用包含胞苷或腺苷脱氨酶和Cas9结构域的融合蛋白的方法

[1001] 本公开的一些方面提供了使用本文提供的融合蛋白或复合物的方法。例如,本公

开的一些方面提供了包括使DNA分子与本文提供的融合蛋白中的任何融合蛋白以及与本文所述的至少一种引导RNA接触的方法。

[1002] 在一些实施方案中,本发明的融合蛋白用于编辑关注的靶基因。具体而言,本文所述的胞苷脱氨酶或腺苷脱氨酶核碱基编辑器能够在靶序列内制作多个突变。这些突变可能会影响靶标的功能。例如,当使用胞苷脱氨酶或腺苷脱氨酶核碱基编辑器靶向调控区时,调控区的功能被改变并且下游蛋白质的表达被减少或消除。

[1003] 应当理解,各个序列中特定位置或残基的编号取决于所使用的特定蛋白质和编号方案。编号可能不同,例如,成熟蛋白质的前体和成熟蛋白质本身的编号不同,并且物种之间的序列差异可能会影响编号。本领域技术人员将能够通过本领域众所周知的方法,例如通过序列比对和同源残基的测定,鉴定任何同源蛋白质和相应编码核酸中的相应残基。

[1004] 对于本领域技术人员显而易见的是,为了使本如文公开的包含Cas9结构域和胞苷或腺苷脱氨酶的融合蛋白中的任何融合蛋白靶向于靶位点,例如包含待编辑的突变的位点,可使引导RNA(例如sgRNA)共表达。如本文别处更详细解释,引导RNA通常包含允许Cas9结合的tracrRNA框架,和引导序列,该引导序列赋予Cas9:核酸编辑酶/结构域融合蛋白序列特异性。可替代地,可以单独提供引导RNA和tracrRNA,作为两个核酸分子。在一些实施方案中,引导RNA包含一种结构,其中引导序列包含与靶序列互补的序列。引导序列通常为20个核苷酸长。基于本公开,用于使Cas9:核酸编辑酶/结构域融合蛋白靶向于特定基因组靶位点的合适的引导RNA的序列对于本领域技术人员是显而易见的。此类合适的引导RNA序列通常包含与待编辑的靶核苷酸上游或下游50个核苷酸内的核酸序列互补的引导序列。本文提供了一些示例性引导RNA序列,这些引导RNA序列适用于将所提供的融合蛋白中的任何融合蛋白靶向于特定的靶序列。

[1005] 碱基编辑器效率

[1006] 在一些实施方案中,本文提供的方法的目的是经由基因编辑改变基因和/或基因产物。本文提供的核碱基编辑蛋白可以用于体外或体内基于基因编辑的人类治疗剂。本领域技术人员将理解,本文提供的核碱基编辑蛋白,例如,包含多核苷酸可编程核苷酸结合结构域(例如,Cas9)和核碱基编辑结构域(例如,腺苷脱氨酶结构域或胞苷脱氨酶结构域)的融合蛋白可以用于从A到G或C到T地编辑核苷酸。

[1007] 有利地,如本文提供的碱基编辑系统提供了基因组编辑而不产生双链DNA断裂,不需要供体DNA模板,并且不会像CRISPR可能那样诱导过量的随机插入和缺失。在一些实施方案中,本公开提供了碱基编辑器,该碱基编辑器在核酸(诸如,受试者基因组内的核酸)中有效率地产生预期突变,诸如终止密码子,而不产生大量的非预期突变,诸如非预期的点突变。在一些实施方案中,预期突变是由与引导多核苷酸(例如,gRNA)结合的特定碱基编辑器(例如,腺苷碱基编辑器或胞苷碱基编辑器)产生的突变,该碱基编辑器被专门设计用于产生预期突变。在一些实施方案中,预期突变位于与靶抗原相关的基因中,该靶抗原与疾病或病症(例如神经或眼科疾病或病症)相关。在一些实施方案中,预期突变是与靶抗原相关的基因中的腺嘌呤(A)到鸟嘌呤(G)点突变(例如,SNP),该靶抗原与疾病或病症(例如神经或眼科疾病或病症)相关。在一些实施方案中,预期突变是基因的编码区或非编码区(例如,调控区或元件)内的腺嘌呤(A)到鸟嘌呤(G)点突变。在一些实施方案中,预期突变是与靶抗原相关的基因中的胞嘧啶(C)到胸腺嘧啶(T)点突变(例如,SNP),该靶抗原与疾病或病症(例

如神经或眼科疾病或病症)相关。在一些实施方案中,预期突变是基因的编码区或非编码区(例如,调控区或元件)内的胞嘧啶(C)到胸腺嘧啶(T)点突变。在一些实施方案中,预期突变是产生终止密码子(例如基因编码区内的提前终止密码子)的点突变。在一些实施方案中,预期突变是消除终止密码子的突变。

[1008] 本发明的碱基编辑器有利地修饰编码蛋白质的特定核苷酸碱基而不产生显著比例的插入/缺失。如本文所用的"插入/缺失"是指核苷酸碱基在核酸内的插入或缺失。此类插入或缺失可以导致基因编码区内的框移突变。在一些实施方案中,需要产生有效率地修饰(例如,突变)核酸内的特定核苷酸而不在核酸中产生大量插入或缺失(即,插入/缺失)的碱基编辑器。在一些实施方案中,需要产生有效率地修饰(例如,突变或甲基化)核酸内的特定核苷酸而不在核酸中产生大量插入或缺失(即,插入/缺失)的碱基编辑器。在某些实施方案中,本文提供的碱基编辑器中的任何碱基编辑器可以产生相对于插入/缺失更大比例的预期修饰(例如,甲基化)。在某些实施方案中,本文提供的碱基编辑器中的任何碱基编辑器可以产生相对于插入/缺失更大比例的预期修饰(例如,突变)。

[1009] 在一些实施方案中,本文提供的碱基编辑器能够产生大于1:1的预期突变与插入/缺失比率(即,预期点突变:非预期点突变)。在一些实施方案中,本文提供的碱基编辑器能够产生至少1.5:1、至少2:1、至少2.5:1、至少3:1、至少3.5:1、至少4:1、至少4.5:1、至少5:1、至少5.5:1、至少6:1、至少6.5:1、至少7:1、至少7.5:1、至少8:1、至少10:1、至少12:1、至少15:1、至少20:1、至少25:1、至少30:1、至少40:1、至少50:1、至少100:1、至少200:1、至少300:1、至少400:1、至少500:1、至少600:1、至少700:1、至少800:1、至少900:1,或至少1000:1或更大的预期突变与插入/缺失比率。可以使用任何合适的方法确定预期突变和插入/缺失的数目。

[1010] 在一些实施方案中,本文提供的碱基编辑器可以限制核酸区域中插入/缺失的形成。在一些实施方案中,区域位于被碱基编辑器靶向的核苷酸处或被碱基编辑器靶向的核苷酸的2、3、4、5、6、7、8、9或10个核苷酸内的区域处。在一些实施方案中,本文提供的碱基编辑器中的任何碱基编辑器可以将核酸区域处插入/缺失的形成限制为小于1%、小于1.5%、小于2%、小于2.5%、小于3%、小于3.5%、小于4%、小于4.5%、小于5%、小于6%、小于7%、小于8%、小于9%、小于10%、小于12%、小于15%,或小于20%。在核酸区域处形成的插入/缺失的数目可以取决于核酸(例如,细胞基因组内的核酸)暴露于碱基编辑器的时间量。在一些实施方案中,在将核酸(例如,细胞基因组内的核酸)暴露于碱基编辑器至少1小时、至少2小时、至少6小时、至少12小时、至少24小时、至少36小时、至少48小时、至少3天、至少4天、至少5天、至少7天、至少10天或至少14天后确定插入/缺失的数目或比例。

[1011] 本公开的一些方面基于以下认识:本文提供的碱基编辑器中的任何碱基编辑器能够有效率地在核酸(例如,受试者基因组内的核酸)中产生预期突变,而不会产生大量非预期突变(例如,伪脱靶编辑或旁观者编辑)。在一些实施方案中,预期突变是由与gRNA结合的特定碱基编辑器产生的突变,该碱基编辑器被专门设计用于产生预期突变。在一些实施方案中,预期突变是产生终止密码子(例如基因编码区内的提前终止密码子)的突变。在一些实施方案中,预期突变是消除终止密码子的突变。在一些实施方案中,预期突变是改变基因剪接的突变。在一些实施方案中,预期突变是改变基因的调控序列(例如基因启动子或基因阻抑因子)的突变。在一些实施方案中,本文提供的碱基编辑器中的任何碱基编辑器能够产

生大于1:1的预期突变与非预期突变比率(例如,预期突变:非预期突变)。在一些实施方案中,本文提供的碱基编辑器中的任何碱基编辑器能够产生至少1.5:1、至少2:1、至少2.5:1、至少3:1、至少3.5:1、至少4:1、至少4.5:1、至少5:1、至少5.5:1、至少6:1、至少6.5:1、至少7:1、至少7.5:1、至少8:1、至少10:1、至少12:1、至少15:1、至少20:1、至少25:1、至少30:1、至少40:1、至少50:1、至少100:1、至少150:1、至少200:1、至少250:1、至少500:1或至少1000:1或更大的预期突变与非预期突变比率。应当理解,本文所述的碱基编辑器的特征可以应用于本文提供的融合蛋白或使用该融合蛋白的方法中的任何一种。

[1012] 碱基编辑常常被称为"修饰",诸如遗传修饰、基因修饰和核酸序列的修饰,并且基于该修饰是碱基编辑修饰的上下文可以被清楚地理解。因此,碱基编辑修饰是处于核苷酸碱基水平下的修饰(例如作为在整个公开中讨论的脱氨酶活性的结果),该修饰然后导致基因序列的变化,并且可能影响基因产物。因此,本质上,本文所述的基因编辑修饰可以导致基因的在结构上和/或功能上的修饰,其中基因产物的表达可以被修饰,例如基因的表达被敲除;或相反,被增强,或者在一些情形下,基因功能或活性可以被修饰。使用本文公开的方法,碱基编辑效率可以被确定为进行碱基编辑的基因的敲低效率,其中碱基编辑旨在敲低基因的表达。敲低水平可以通过经由任何检测测定,诸如用于蛋白质表达水平的测定,例如经由流式细胞术;用于检测RNA表达的测定,诸如定量RT-PCR、RNA印迹分析,或任何其他合适的测定(诸如焦磷酸法测序)确定表达水平来定量地验证,并且可以通过核苷酸测序反应来定性地验证。

[1013] 在一些实施方案中,修饰(例如单碱基编辑)导致基因靶向表达降低至少10%。在一些实施方案中,碱基编辑效率可以导致基因靶向表达降低至少10%。在一些实施方案中,碱基编辑效率可以导致基因靶向表达降低至少20%。在一些实施方案中,碱基编辑效率可以导致基因靶向表达降低至少30%。在一些实施方案中,碱基编辑效率可以导致基因靶向表达降低至少40%。在一些实施方案中,碱基编辑效率可以导致基因靶向表达降低至少50%。在一些实施方案中,碱基编辑效率可以导致被靶向基因的表达降低至少60%。在一些实施方案中,碱基编辑效率可以导致被靶向基因的表达降低至少70%。在一些实施方案中,碱基编辑效率可以导致被靶向基因的表达降低至少80%。在一些实施方案中,碱基编辑效率可以导致被靶向基因的表达降低至少90%。在一些实施方案中,碱基编辑效率可以导致被靶向基因的表达降低至少91%。在一些实施方案中,碱基编辑效率可以导致被靶向基因的表达降低至少92%。在一些实施方案中,碱基编辑效率可以导致被靶向基因的表达降低至少93%。在一些实施方案中,碱基编辑效率可以导致被靶向基因的表达降低至少94%。在一些实施方案中,碱基编辑效率可以导致被靶向基因的表达降低至少95%。在一些实施方案中,碱基编辑效率可以导致被靶向基因的表达降低至少96%。在一些实施方案中,碱基编辑效率可以导致被靶向基因的表达降低至少97%。在一些实施方案中,碱基编辑效率可以导致被靶向基因的表达降低至少98%。在一些实施方案中,碱基编辑效率可以导致被靶向基因的表达降低至少99%。在一些实施方案中,碱基编辑效率可以导致被靶向的基因的敲除(基因表达的100%敲低)。

[1014] 在一些实施方案中,本文提供的碱基编辑器系统中的任何碱基编辑器系统导致靶多核苷酸序列中少于50%、少于40%、少于30%、少于20%、少于19%、少于18%、少于17%、少于16%、少于15%、少于14%、少于13%、少于12%、少于11%、少于10%、少于9%、少于

8%、少于7%、少于6%、少于5%、少于4%、少于3%、少于2%、少于1%、少于0.9%、少于0.8%、少于0.7%、少于0.6%、少于0.5%、少于0.4%、少于0.3%、少于0.2%、少于0.1%、少于0.09%、少于0.08%、少于0.07%、少于0.06%、少于0.05%、少于0.04%、少于0.03%、少于0.02%或少于0.01%的插入/缺失形成。

[1015] 在一些实施方案中,被靶向修饰(例如单碱基编辑)用于同时靶向至少4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个、31个、32个、33个、34个、35个、36个、37个、38个、39个、40个、41个、42个、43个、44个、45个、46个、47个、48个、49个或50个不同的内源序列,以用于使用不同的引导RNA进行碱基编辑。在一些实施方案中,被靶向修饰(例如单碱基编辑)用于连续靶向至少4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个、31个、32个、33个、34个、35个、36个、37个、38个、39个、40个、41个、42个、43个、44个、45个、46个、47个、48个、49个、50个或更多个不同的内源基因序列,以用于使用不同的引导RNA进行碱基编辑。

[1016] 本公开的一些方面基于以下认识:本文提供的碱基编辑器中的任何碱基编辑器能够在核酸(例如,受试者基因组内的核酸)中有效率地产生预期突变诸如点突变,而不会产生大量非预期突变,诸如非预期点突变(即,旁观者突变)。在一些实施方案中,本文提供的碱基编辑器中的任何碱基编辑器能够产生至少0.01%的预期突变(即,至少0.01%的碱基编辑效率)。在一些实施方案中,本文提供的碱基编辑器中的任何碱基编辑器能够产生至少0.01%、1%、2%、3%、4%、5%、10%、15%、20%、25%、30%、40%、45%、50%、60%、70%、80%、90%、95%或99%的预期突变。

[1017] 在一些实施方案中,包含本文所述的ABE8碱基编辑器变体之一的碱基编辑器系统中的任何碱基编辑器系统导致靶多核苷酸序列中少于50%、少于40%、少于30%、少于20%、少于19%、少于18%、少于17%、少于16%、少于15%、少于14%、少于13%、少于12%、少于11%、少于10%、少于9%、少于8%、少于7%、少于6%、少于5%、少于4%、少于3%、少于2%、少于1%、少于0.9%、少于0.8%、少于0.7%、少于0.6%、少于0.5%、少于0.4%、少于0.3%、少于0.2%、少于0.1%、少于0.09%、少于0.08%、少于0.07%、少于0.06%、少于0.05%、少于0.04%、少于0.03%、少于0.02%或少于0.01%的插入/缺失形成。在一些实施方案中,包含本文所述的ABE8碱基编辑器变体之一的碱基编辑器系统中的任何碱基编辑器系统导致靶多核苷酸序列中少于0.8%的插入/缺失形成。在一些实施方案中,包含本文所述的ABE8碱基编辑器变体之一的碱基编辑器系统中的任何碱基编辑器系统导致靶多核苷酸序列中至多0.8%的插入/缺失形成。在一些实施方案中,包含本文所述的ABE8碱基编辑器变体之一的碱基编辑器系统中的任何碱基编辑器系统导致靶多核苷酸序列中少于0.3%的插入/缺失形成。在一些实施方案中,与包含ABE7碱基编辑器之一的碱基编辑器系统相比,包含所述的ABE8碱基编辑器变体之一的碱基编辑器系统中的任何碱基编辑器系统导致靶多核苷酸序列中较低的插入/缺失形成。在一些实施方案中,与包含ABE7.10的碱基编辑器系统相比,包含本文所述的ABE8碱基编辑器变体之一的的碱基编辑器系统中的任何碱基编辑器系统导致靶多核苷酸序列中较低的插入/缺失形成。

[1018] 在一些实施方案中,与包含ABE7碱基编辑器之一的碱基编辑器系统相比,包含本

文所述的ABE8碱基编辑器变体之一的碱基编辑器系统中的任何碱基编辑器系统在插入/缺失频率上有所降低。在一些实施方案中,与包含ABE7碱基编辑器之一的碱基编辑器系统相比,包含本文所述的ABE8碱基编辑器变体之一的碱基编辑器系统中的任何碱基编辑器系统在插入/缺失频率上降低了至少0.01%、至少1%、至少2%、至少3%、至少4%、至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%或至少95%。在一些实施方案中,与包含ABE7.10的碱基编辑器系统相比,包含本文所述的ABE8碱基编辑器变体之一的碱基编辑器系统在插入/缺失频率上降低了至少0.01%、至少1%、至少2%、至少3%、至少4%、至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%或至少95%。

[1019] 本发明提供了具有增加的效率和特异性的腺苷脱氨酶变体(例如,ABE8变体)。具体而言,本文所述的腺苷脱氨酶变体更有可能编辑多核苷酸内的所需碱基,并且不太可能编辑非预期改变的碱基(例如,“旁观者”)。

[1020] 在一些实施方案中,包含本文所述的ABE8碱基编辑器变体之一的碱基编辑系统中的任何碱基编辑系统具有降低的旁观者编辑或突变。在一些实施方案中,非预期编辑或突变是旁观者突变或旁观者编辑,例如,在靶核苷酸序列的靶窗口中的非预期或非靶位置中的靶碱基(例如,A或C)的碱基编辑。在一些实施方案中,与包含ABE7碱基编辑器(例如ABE7.10)的碱基编辑器系统相比,包含本文所述的ABE8碱基编辑器变体之一的碱基编辑系统中的任何碱基编辑系统具有降低的旁观者编辑或突变。在一些实施方案中,与包含ABE7碱基编辑器(例如ABE7.10)的碱基编辑器系统相比,包含本文所述的ABE8碱基编辑器变体之一的碱基编辑系统中的任何碱基编辑系统具有降低了至少1%、至少2%、至少3%、至少4%、至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%或至少99%的旁观者编辑或突变。在一些实施方案中,与包含ABE7碱基编辑器(例如ABE7.10)的碱基编辑器系统相比,包含本文所述的ABE8碱基编辑器变体之一的碱基编辑系统中的任何碱基编辑系统具有降低了至少1.1倍、至少1.2倍、至少1.3倍、至少1.4倍、至少1.5倍、至少1.6倍、至少1.7倍、至少1.8倍、至少1.9倍、至少2.0倍、至少2.1倍、至少2.2倍、至少2.3倍、至少2.4倍、至少2.5倍、至少2.6倍、至少2.7倍、至少2.8倍、至少2.9倍或至少3.0倍的旁观者编辑或突变。

[1021] 在一些实施方案中,包含本文所述的ABE8碱基编辑器变体之一的碱基编辑系统中的任何碱基编辑系统具有减少的伪编辑(spurious editing)。在一些实施方案中,非预期编辑或突变是伪突变或伪编辑,例如基因组的非预期或非靶区域中的靶碱基(例如,A或C)的非特异性编辑或引导独立编辑。在一些实施方案中,与包含ABE7碱基编辑器(例如ABE7.10)的碱基编辑器系统相比,包含本文所述的ABE8碱基编辑器变体之一的碱基编辑系统中的任何碱基编辑系统具有降低的伪编辑。在一些实施方案中,与包含ABE7碱基编辑器(例如ABE7.10)的碱基编辑器系统相比,包含本文所述的ABE8碱基编辑器变体之一的碱基编辑系统中的任何碱基编辑系统具有降低了至少1%、至少2%、至少3%、至少4%、至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、

至少50％、至少55％、至少60％、至少65％、至少70％、至少75％、至少80％、至少85％、至少90％、至少95％或至少99％的伪编辑。在一些实施方案中，与包含ABE7碱基编辑器(例如ABE7.10)的碱基编辑器系统相比，包含本文所述的ABE8碱基编辑器变体之一的碱基编辑系统中的任何碱基编辑系统具有降低了至少1.1倍、至少1.2倍、至少1.3倍、至少1.4倍、至少1.5倍、至少1.6倍、至少1.7倍、至少1.8倍、至少1.9倍、至少2.0倍、至少2.1倍、至少2.2倍、至少2.3倍、至少2.4倍、至少2.5倍、至少2.6倍、至少2.7倍、至少2.8倍、至少2.9倍或至少3.0倍的伪编辑。

[1022] 在一些实施方案中，本文所述的ABE8碱基编辑器变体中的任何ABE8碱基编辑器变体具有至少0.01％、至少1％、至少2％、至少3％、至少4％、至少5％、至少10％、至少15％、至少20％、至少25％、至少30％、至少35％、至少40％、至少45％、至少50％、至少55％、至少60％、至少65％、至少70％、至少75％、至少80％、至少85％、至少90％或至少95％或至少99％的碱基编辑效率。在一些实施方案中，碱基编辑效率可以通过计算细胞群体中编辑的核碱基的百分比来测量。在一些实施方案中，本文所述的ABE8碱基编辑器变体中ABE8碱基编辑器变体具有如根据细胞群体中经编辑的核碱基测量的至少0.01％、至少1％、至少2％、至少3％、至少4％、至少5％、至少10％、至少15％、至少20％、至少25％、至少30％、至少35％、至少40％、至少45％、至少50％、至少55％、至少60％、至少65％、至少70％、至少75％、至少80％、至少85％、至少90％或至少95％或至少99％的碱基编辑效率。

[1023] 在一些实施方案中，与ABE7碱基编辑器相比，本文所述的ABE8碱基编辑器变体中的任何ABE8碱基编辑器变体具有更高的碱基编辑效率。在一些实施方案中，与ABE7碱基编辑器(例如ABE7.10)相比，本文所述的ABE8碱基编辑器变体中的任何ABE8碱基编辑器变体具有高至少1％、至少2％、至少3％、至少4％、至少5％、至少10％、至少15％、至少20％、至少25％、至少30％、至少35％、至少40％、至少45％、至少50％、至少55％、至少60％、至少65％、至少70％、至少75％、至少80％、至少85％、至少90％、至少95％、至少99％、至少100％、至少105％、至少110％、至少115％、至少120％、至少125％、至少130％、至少135％、至少140％、至少145％、至少150％、至少155％、至少160％、至少165％、至少170％、至少175％、至少180％、至少185％、至少190％、至少195％、至少200％、至少210％、至少220％、至少230％、至少240％、至少250％、至少260％、至少270％、至少280％、至少290％、至少300％、至少310％、至少320％、至少330％、至少340％、至少350％、至少360％、至少370％、至少380％、至少390％、至少400％、至少450％或至少500％的碱基编辑效率。

[1024] 在一些实施方案中，与ABE7碱基编辑器(例如ABE7.10)相比，本文所述的ABE8碱基编辑器变体中的任何ABE8碱基编辑器变体具有高至少1.1倍、至少1.2倍、至少1.3倍、至少1.4倍、至少1.5倍、至少1.6倍、至少1.7倍、至少1.8倍、至少1.9倍、至少2.0倍、至少2.1倍、至少2.2倍、至少2.3倍、至少2.4倍、至少2.5倍、至少2.6倍、至少2.7倍、至少2.8倍、至少2.9倍、至少3.0倍、至少3.1倍、至少3.2倍、至少3.3倍、至少3.4倍、至少3.5倍、至少3.6倍、至少3.7倍、至少3.8倍、至少3.9倍、至少4.0倍、至少4.1倍、至少4.2倍、至少4.3倍、至少4.4倍、至少4.5倍、至少4.6倍、至少4.7倍、至少4.8倍、至少4.9倍或至少5.0倍的碱基编辑效率。

[1025] 在一些实施方案中，本文所述的ABE8碱基编辑器变体中的任何ABE8碱基编辑器变体具有至少0.01％、至少1％、至少2％、至少3％、至少4％、至少5％、至少10％、至少15％、至少20％、至少25％、至少30％、至少35％、至少40％、至少45％、至少50％、至少55％、至少

60％、至少65％、至少70％、至少75％、至少80％、至少85％、至少90％或至少95％或至少99％的中靶碱基编辑效率。在一些实施方案中,本文所述的ABE8碱基编辑器变体中的任何ABE8碱基编辑器变体具有如根据细胞群体中经编辑的靶核碱基测量的至少0.01％、至少1％、至少2％、至少3％、至少4％、至少5％、至少10％、至少15％、至少20％、至少25％、至少30％、至少35％、至少40％、至少45％、至少50％、至少55％、至少60％、至少65％、至少70％、至少75％、至少80％、至少85％、至少90％或至少95％或至少99％的中靶碱基编辑效率。

[1026] 在一些实施方案中,与ABE7碱基编辑器相比,本文所述的ABE8碱基编辑器变体中的任何ABE8碱基编辑器变体具有更高的中靶碱基编辑效率。在一些实施方案中,与ABE7碱基编辑器(例如ABE7.10)相比,本文所述的ABE8碱基编辑器变体中的任何ABE8碱基编辑器变体具有高至少1％、至少2％、至少3％、至少4％、至少5％、至少10％、至少15％、至少20％、至少25％、至少30％、至少35％、至少40％、至少45％、至少50％、至少55％、至少60％、至少65％、至少70％、至少75％、至少80％、至少85％、至少90％、至少95％、至少99％、至少100％、至少105％、至少110％、至少115％、至少120％、至少125％、至少130％、至少135％、至少140％、至少145％、至少150％、至少155％、至少160％、至少165％、至少170％、至少175％、至少180％、至少185％、至少190％、至少195％、至少200％、至少210％、至少220％、至少230％、至少240％、至少250％、至少260％、至少270％、至少280％、至少290％、至少300％、至少310％、至少320％、至少330％、至少340％、至少350％、至少360％、至少370％、至少380％、至少390％、至少400％、至少450％或至少500％的中靶碱基编辑效率。

[1027] 在一些实施方案中,与ABE7碱基编辑器(例如ABE7.10)相比,本文所述的ABE8碱基编辑器变体中的任何ABE8碱基编辑器变体具有高至少1.1倍、至少1.2倍、至少1.3倍、至少1.4倍、至少1.5倍、至少1.6倍、至少1.7倍、至少至少1.8倍、至少1.9倍、至少2.0倍、至少2.1倍、至少2.2倍、至少2.3倍、至少2.4倍、至少2.5倍、至少2.6倍、至少2.7倍、至少2.8倍、至少2.9倍、至少3.0倍、至少3.1倍、至少3.2倍、至少3.3倍、至少3.4倍、至少3.5倍、至少3.6倍、至少3.7倍、至少3.8倍、至少3.9倍、至少4.0倍、至少4.1倍、至少4.2倍、至少4.3倍、至少4.4倍、至少4.5倍、至少4.6倍、至少4.7倍、至少4.8倍、至少4.9倍或至少5.0倍的中靶碱基编辑效率。

[1028] 本文所述的ABE8碱基编辑器变体可以经由质粒、载体、LNP复合物或mRNA递送至宿主细胞。在一些实施方案中,本文所述的ABE8碱基编辑器变体中的任何ABE8碱基编辑器变体作为mRNA递送至宿主细胞。在一些实施方案中,经由基于核酸的递送系统(例如,mRNA)递送的ABE8碱基编辑器具有如根据经编辑的核碱基测量的至少1％、至少2％、至少3％、至少4％、至少5％、至少10％、至少15％、至少20％、至少25％、至少30％、至少35％、至少40％、至少45％、至少50％、至少55％、至少60％、至少65％、至少70％、至少75％、至少80％、至少85％、至少90％、至少95％或至少99％的中靶编辑效率。在一些实施方案中,与通过质粒或载体系统递送的ABE8碱基编辑器相比,由mRNA系统递送的ABE8碱基编辑器具有更高的碱基编辑效率。在一些实施方案中,与通过质粒或载体系统递送时相比,本文所述的ABE8碱基编辑器变体中的任何ABE8碱基编辑器变体在通过mRNA系统递送时具有至少1％、至少2％、至少3％、至少4％、至少5％、至少10％、至少15％、至少20％、至少25％、至少30％、至少35％、至少40％、至少45％、至少50％、至少55％、至少60％、至少65％、至少70％、至少75％、至少80％、至少85％、至少90％、至少95％、至少99％、至少100％、至少105％、至少110％、至少

115%、至少120%、至少125%、至少130%、至少135%、至少140%、至少145%、至少150%、至少155%、至少160%、至少165%、至少170%、至少175%、至少180%、至少185%、至少190%、至少195%、至少200%、至少210%、至少220%、至少230%、至少240%、至少250%、至少260%、至少270%、至少280%、至少290%、至少300%、至少310%、至少320%、至少330%、至少340%、至少350%、至少360%、至少370%、至少380%、至少390%、至少400%、至少450%,或至少500%的中靶编辑效率。在一些实施方案中,与通过质粒或载体系统递送时相比,本文所述的ABE8碱基编辑器变体中的任何ABE8碱基编辑器变体在通过mRNA系统递送时具有高至少1.1倍、至少1.2倍、至少1.3倍、至少1.4倍、至少1.5倍、至少1.6倍、至少1.7倍、至少至少1.8倍、至少1.9倍、至少2.0倍、至少2.1倍、至少2.2倍、至少2.3倍、至少2.4倍、至少2.5倍、至少2.6倍、至少2.7倍、至少2.8倍、至少2.9倍、至少3.0倍、至少3.1倍、至少3.2倍、至少3.3倍、至少3.4倍、至少3.5倍、至少3.6倍、至少3.7倍、至少3.8倍、至少3.9倍、至少4.0倍、至少4.1倍、至少4.2倍、至少4.3倍、至少4.4倍、至少4.5倍、至少4.6倍、至少4.7倍、至少4.8倍、至少4.9倍或至少5.0倍的中靶编辑效率。

[1029] 在一些实施方案中,包含本文所述的ABE8碱基编辑器变体之一的碱基编辑器系统中的任何碱基编辑器系统导致靶多核苷酸序列中少于50%、少于40%、少于30%、少于20%、少于19%、少于18%、少于17%、少于16%、少于15%、少于14%、少于13%、少于12%、少于11%、少于10%、少于9%、少于8%、少于7%、少于6%、少于5%、少于4%、少于3%、少于2%、少于1%、少于0.9%、少于0.8%、少于0.7%、少于0.6%、少于0.5%、少于0.4%、少于0.3%、少于0.2%、少于0.1%、少于0.09%、少于0.08%、少于0.07%、少于0.06%、少于0.05%、少于0.04%、少于0.03%、少于0.02%或少于0.01%的脱靶编辑。

[1030] 在一些实施方案中,与通过质粒或载体系统递送时相比,本文所述的ABE8碱基编辑器变体中的任何ABE8碱基编辑器变体在通过mRNA系统递送时具有较低的受引导脱靶编辑效率。在一些实施方案中,与通过质粒或载体系统递送时相比,本文所述的ABE8碱基编辑器变体中的任何ABE8碱基编辑器变体在通过mRNA系统递送时具有低至少1%、至少2%、至少3%、至少4%、至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%或至少99%的受引导脱靶编辑效率。在一些实施方案中,与通过质粒或载体系统递送时相比,本文所述的ABE8碱基编辑器变体中的任何ABE8碱基编辑器变体在通过mRNA系统递送时具有低至少1.1倍、至少1.2倍、至少1.3倍、至少1.4倍、至少1.5倍、至少1.6倍、至少1.7倍、至少至少1.8倍、至少1.9倍、至少2.0倍、至少2.1倍、至少2.2倍、至少2.3倍、至少2.4倍、至少2.5倍、至少2.6倍、至少2.7倍、至少2.8倍、至少2.9倍或至少3.0倍的受引导脱靶编辑效率。在一些实施方案中,与通过质粒或载体系统递送时相比,本文所述的ABE8碱基编辑器变体中的任何ABE8碱基编辑器变体在通过mRNA系统递送时具有降低至少约2.2倍的受引导脱靶编辑效率。

[1031] 在一些实施方案中,与通过质粒或载体系统递送时相比,本文所述的ABE8碱基编辑器变体中的任何ABE8碱基编辑器变体在通过mRNA系统递送时具有较低的引导非依赖性脱靶编辑效率。在一些实施方案中,与通过质粒或载体系统递送时相比,本文所述的ABE8碱基编辑器变体中的任何ABE8碱基编辑器变体在通过mRNA系统递送时具有低至少1%、至少2%、至少3%、至少4%、至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少

35％、至少40％、至少45％、至少50％、至少55％、至少60％、至少65％、至少70％、至少75％、至少80％、至少85％、至少90％、至少95％或至少99％的引导非依赖性脱靶编辑效率。在一些实施方案中，与通过质粒或载体系统递送时相比，本文所述的ABE8碱基编辑器变体中的任何ABE8碱基编辑器变体在通过mRNA系统递送时具有低至少1.1倍、至少1.2倍、至少1.3倍、至少1.4倍、至少1.5倍、至少1.6倍、至少1.7倍、至少1.8倍、至少1.9倍、至少2.0倍、至少2.1倍、至少2.2倍、至少2.3倍、至少2.4倍、至少2.5倍、至少2.6倍、至少2.7倍、至少2.8倍、至少2.9倍、至少3.0倍、至少5.0倍、至少10.0倍、至少20.0倍、至少50.0倍、至少70.0倍、至少100.0倍、至少120.0倍、至少130.0倍，或至少150.0倍的引导非依赖性脱靶编辑效率。在一些实施方案中，与通过质粒或载体系统递送时相比，本文所述的ABE8碱基编辑器变体在通过mRNA系统递送时具有降低134.0倍的引导非依赖性脱靶编辑效率(例如，伪RNA脱氨作用)。在一些实施方案中，本文所述的ABE8碱基编辑器变体不增加整个基因组的引导非依赖性突变率。

[1032] 在一些实施方案中，单个基因递送事件(例如，通过转导、转染、电穿孔或任何其他方法)可以用于靶向细胞基因组内5个序列的碱基编辑。在一些实施方案中，单个基因递送事件可以用于靶向细胞基因组内6个序列的碱基编辑。在一些实施方案中，单个基因递送事件可以用于靶向细胞基因组内7个序列的碱基编辑。在一些实施方案中，单个电穿孔事件可以用于靶向细胞基因组内8个序列的碱基编辑。在一些实施方案中，单个基因递送事件可以用于靶向细胞基因组内9个序列的碱基编辑。在一些实施方案中，单个基因递送事件可以用于靶向细胞基因组内10个序列的碱基编辑。在一些实施方案中，单个基因递送事件可以用于靶向细胞基因组内20个序列的碱基编辑。在一些实施方案中，单个基因递送事件可以用于靶向细胞基因组内30个序列的碱基编辑。在一些实施方案中，单个基因递送事件可以用于靶向细胞基因组内40个序列的碱基编辑。在一些实施方案中，单个基因递送事件可以用于靶向细胞基因组内50个序列的碱基编辑。

[1033] 在一些实施方案中，本文所述的方法，例如碱基编辑方法具有最小化到没有的脱靶效应。

[1034] 在一些实施方案中，本文所述的碱基编辑方法导致至少50％的已被成功编辑的细胞群体(即，已被成功工程化的细胞)。在一些实施方案中，本文所述的碱基编辑方法导致至少55％的已被成功编辑的细胞群体。在一些实施方案中，本文所述的碱基编辑方法导致至少60％的已被成功编辑的细胞群体。在一些实施方案中，本文所述的碱基编辑方法导致至少65％的已被成功编辑的细胞群体。在一些实施方案中，本文所述的碱基编辑方法导致至少70％的已被成功编辑的细胞群体。在一些实施方案中，本文所述的碱基编辑方法导致至少75％的已被成功编辑的细胞群体。在一些实施方案中，本文所述的碱基编辑方法导致至少80％的已被成功编辑的细胞群体。在一些实施方案中，本文所述的碱基编辑方法导致至少85％的已被成功编辑的细胞群体。在一些实施方案中，本文所述的碱基编辑方法导致至少90％的已被成功编辑的细胞群体。在一些实施方案中，本文所述的碱基编辑方法导致至少95％的已被成功编辑的细胞群体。在一些实施方案中，本文所述的碱基编辑方法导致约91％、92％、93％、94％、95％、96％、97％、98％、99％或100％的已被成功编辑的细胞群体。

[1035] 在一些实施方案中，碱基编辑干预后的活细胞回收率大于碱基编辑事件发生时的起始细胞群体的至少60％、70％、80％、90％。在一些实施方案中，如上所述的活细胞回收率

为约70％。在一些实施方案中,如上所述的活细胞回收率为约75％。在一些实施方案中,如上所述的活细胞回收率为约80％。在一些实施方案中,如上所述的活细胞回收率为约85％。在一些实施方案中,如上所述的活细胞回收率为碱基编辑事件发生时的群体中的细胞的约90％,或约91％、92％、93％、94％、95％、96％、97％、98％,或99％,或100％。

[1036] 在一些实施方案中,经工程化的细胞群体可以在体外被进一步扩增约2倍、约3倍、约4倍、约5倍、约6倍、约7倍、约8倍、约9倍、约10倍、约15倍、约20倍、约25倍、约30倍、约35倍、约40倍、约45倍、约50倍或约100倍。

[1037] 可以使用任何合适的方法来确定预期突变和插入/缺失的数目,该方法例如,如描述于国际PCT申请号PCT/2017/045381(WO2018/027078)和PCT/US2016/058344(WO2017/070632);Komor,A.C.等人,"Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage"Nature 533,420-424(2016);Gaudelli,N.M.等人,"Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage"Nature 551,464-471(2017);以及Komor,A.C.等人,"Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity"Science Advances 3:eaao4774(2017);其全部内容据此以引用方式并入本文。

[1038] 在一些实施方案中,为了计算插入/缺失频率,扫描测序读段以获得与侧接可以出现插入/缺失的窗口的两侧的两个10-bp序列的精确匹配。如果没有定位到精确匹配,则从分析中排除该读段。如果此插入/缺失窗口的长度与参考序列精确匹配,则该读段被分类为不含插入/缺失。如果插入/缺失窗口比参考序列长或短两个或更多个碱基,则测序读段分别被分类为插入或缺失。在一些实施方案中,本文提供的碱基编辑器可以限制核酸区域中插入/缺失的形成。在一些实施方案中,区域位于被碱基编辑器靶向的核苷酸处或被碱基编辑器靶向的核苷酸的2、3、4、5、6、7、8、9或10个核苷酸内的区域处。

[1039] 在靶核苷酸区域处形成的插入/缺失的数目可以取决于核酸(例如,细胞基因组内的核酸)暴露于碱基编辑器的时间量。在一些实施方案中,在将靶核苷酸序列(例如,细胞基因组内的核酸)暴露于碱基编辑器至少1小时、至少2小时、至少6小时、至少12小时、至少24小时、至少36小时、至少48小时、至少3天、至少4天、至少5天、至少7天、至少10天或至少14天后确定插入/缺失的数目或比例。应当理解,如本文所述的碱基编辑器的特征可以应用于本文提供的融合蛋白或使用该融合蛋白的方法中的任何一种。

[1040] 碱基编辑器效率的细节描述于国际PCT申请号PCT/2017/045381(WO 201 8/027078)和PCT/US2016/058344(WO 2017/070632),其中的每一篇均以引用方式整体并入本文。另外参见Komor,A.C.等人,"Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage"Nature 533,420-424(2016);Gaudelli,N.M.等人,"Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage"Nature 551,464-471(2017);以及Komor,A.C.等人,"Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity"Science Advances 3:eaao4774(2017),其全部内容以引用方式并入本文。在一些实施方案中,使用本文提供的方法编辑一个或多个基因中的多个核碱基对导致形成至少一种预期突变。在一些实施方案

中,所述至少一种预期突变的所述形成导致基因正常功能的破坏。在一些实施方案中,所述至少一种预期突变的所述形成减少或消除了由基因编码的蛋白质的表达。应当理解,可以使用本文提供的任何方法或方法的组合来完成多重编辑。

[1041] 经工程化的核酸酶

[1042] 在一些实施方案中,基因编辑系统包含经工程化的核酸酶(例如,大范围核酸酶、锌指核酸酶(ZFN)、转录激活子样效应物核酸酶(TALEN)或Cas核酸酶。在一些实施方案中,基因编辑系统包含ZFN。ZFN是包含锌指DNA结合结构域("ZF")和核酸酶结构域的融合蛋白。每个天然存在的ZF均可以结合至三个连续的碱基对(DNA三联体),并且ZF重复序列被组合以识别DNA靶序列并提供足够的亲和力。因此,经工程化的ZF重复序列被组合以识别更长的DNA序列,诸如9个碱基对、12个碱基对、15个碱基对、18个碱基对等。在一些实施方案中,ZFN包含与来自限制性核酸内切酶(例如FokI)的核酸酶结构域融合的ZF。在一些实施方案中,核酸酶结构域包含二聚化结构域(诸如当核酸酶二聚化为有活性时),并且包含ZF重复序列和该核酸酶结构域的ZFN对被设计用于靶向靶序列,该靶序列包含被DNA分子的相对链上的每个ZF重复序列识别的两个半靶序列,该两个半靶序列之间具有互连序列(有时在文献中称为间隔区)。例如,该互连序列的长度是5至7个碱基对。当该ZFN对的两个ZFN结合时,核酸酶结构域可二聚化并且在互连序列内引入DSB。在一些实施方案中,核酸酶结构域的二聚化结构域包含旋钮-入-孔基序(knob-into-hole motif)以促进二聚化。

[1043] 在一些实施方案中,基因编辑系统包含TALEN。TALEN的DNA结合结构域通常包含可变数目的34或35个氨基酸重复序列("模块"或"TAL模块"),每个模块结合至单个DNA碱基对A、T、G或C。每个模块的第12和13位处的相邻残基("重复可变双残基"或RVD)指定了该模块所结合至的单个DNA碱基对。在一些实施方案中,TALEN可以包含来自限制性核酸内切酶(例如,FokI)的核酸酶结构域。在一些实施方案中,核酸酶结构域可以二聚化为有活性,并且TALEN对被设计用于靶向靶序列,该靶序列包含由DNA分子的相对链上的每个DNA结合结构域识别的两个半靶序列,该两个半靶序列之间具有互连序列。例如,每个半靶序列的长度为10至20个碱基对,互连序列的长度为12至19个碱基对。当该对的两个TALEN结合时,该核酸酶结构域可二聚化并且在互连序列内引入双链断裂。在一些实施方案中,核酸酶结构域的二聚化结构域可以包含旋钮-入-孔基序以促进二聚化。

[1044] 在一些实施方案中,基因编辑系统包含大范围核酸酶。天然存在的大范围核酸酶识别并切割约12至40个碱基对的双链DNA序列,并且通常分为五个家族。在一些实施方案中,大范围核酸酶选自LAGLIDADG家族、GIY-YIG家族、HNH家族、His-Cys盒家族和PD-(D/E)XK家族。在一些实施方案中,大范围核酸酶的DNA结合结构域被工程化以识别并结合至除其同源靶序列之外的序列。在一些实施方案中,大范围核酸酶的DNA结合结构域与异源核酸酶结构域融合。在一些实施方案中,大范围核酸酶诸如归巢核酸内切酶与TAL模块融合以产生杂合蛋白,诸如"megaTAL"蛋白。该megaTAL蛋白可以通过识别该大范围核酸酶的DNA结合结构域和TAL模块二者的靶序列来提高DNA靶向特异性。

[1045] G.药物组合物和制剂

[1046] 本文提供了包含本文所述的重组狂犬病病毒基因组和重组狂犬病病毒中的任何一者的组合物(例如,药物组合物)。如本文所用的术语"药物组合物"是指被配制用于药物用途的组合物。在某些实施方案中,药物组合物进一步包含药学上可接受的载剂。在某些实

施方案中,药物组合物包含另外的剂(例如,用于特异性递送、增加半衰期,或其他治疗性化合物)。

[1047] 如本文所用,术语“药学上可接受的载剂”是指参与将化合物(例如,本文所述的重组狂犬病病毒基因组或重组狂犬病病毒)从身体的一个部位(例如,递送部位)携带或运输至另一部位(例如,靶器官、组织,或身体的一部分)的药学上可接受的材料、组合物或媒介物,诸如液体或固体填充剂、稀释剂、赋形剂、制造助剂(例如润滑剂、滑石、镁、钙或锌硬脂酸盐,或硬脂酸),或溶剂封装材料。药学上可接受的载剂在与制剂的其他成分相容并且对受试者的组织无害的意义上是“可接受的”(例如,生理相容的、无菌的、生理pH等)。

[1048] 可以用作药学上可接受的载剂的物质的一些非限制性实例包括:(1)糖,诸如乳糖、葡萄糖和蔗糖;(2)淀粉,诸如玉米淀粉和土豆淀粉;(3)纤维素和其衍生物,诸如羧甲基纤维素钠、甲基纤维素、乙基纤维素、微晶纤维素和乙酸纤维素;(4)粉末状黄蓍胶;(5)麦芽;(6)明胶;(7)润滑剂,诸如硬脂酸镁、月桂基磺酸钠和滑石;(8)赋形剂,诸如可可油和栓剂蜡;(9)油,诸如花生油、棉籽油、红花油、芝麻油、橄榄油、玉米油和大豆油;(10)二醇,诸如丙二醇;(11)多元醇,诸如甘油、山梨醇、甘露醇和聚乙二醇(PEG);(12)酯,诸如油酸乙酯和月桂酸乙酯;(13)琼脂;(14)缓冲剂,诸如氢氧化镁和氢氧化铝;(15)藻酸;(16)不含热原的水;(17)等渗盐水;(18)林格氏液(Ringer'ssolution);(19)乙醇;(20)pH缓冲溶液;(21)聚酯、聚碳酸酯和/或聚酐;(22)填充剂,诸如多肽和氨基酸;(23)血清醇,诸如乙醇;以及(23)药物制剂中采用的其他无毒相容物质。制剂中还可以存在润湿剂、着色剂、脱模剂、涂层剂、甜味剂、矫味剂、芳香剂、防腐剂和抗氧化剂。术语如“赋形剂”、“载剂”、“药学上可接受的载剂”、“媒介物”等在本文中可互换使用。

[1049] 药物组合物可以包含一种或多种pH缓冲化合物以将制剂的pH维持在反映生理pH的预定水平,诸如在约5.0至约8.0的范围内。用于水性液体制剂的pH缓冲化合物可以是氨基酸,诸如组氨酸或氨基酸(诸如组氨酸和甘氨酸)混合物。可替代地,pH缓冲化合物优选为将制剂的pH维持在预定水平(诸如在约5.0至约8.0的范围内),并且不螯合钙离子的剂。此类pH缓冲化合物的说明性实例包括但不限于咪唑和乙酸根离子。pH缓冲化合物可以以适合于将制剂的pH维持在预定水平的任何量存在。

[1050] 药物组合物还可以含有一种或多种渗透调节剂,即将制剂的渗透特性(例如,张力、渗透度和/或渗透压)调节到受体个体的血流和血细胞可以接受的水平的化合物。渗透调节剂可以是不螯合钙离子的剂。渗透调节剂可以是本领域技术人员已知或可获得的调节制剂的渗透特性的任何化合物。本领域技术人员可以凭经验确定给定渗透调节剂在本发明制剂中的适用性。合适的渗透调节剂类型的说明性实例包括但不限于:盐类,诸如氯化钠和乙酸钠;糖类,诸如蔗糖、右旋糖和甘露醇;氨基酸,诸如甘氨酸;以及这些剂和/或这些类型的剂中的一种或多种的混合物。该一种或多种渗透调节剂可以以足以调节制剂的渗透特性的任何浓度存在。

[1051] 在某些实施方案中,药物组合物被配制用于递送至受试者,例如被配制用于基因疗法。施用本文所述药物组合物的合适途径包括但不限于:外用(topical)、皮下、透皮、真皮内、病灶内、关节内、腹膜内、膀胱内、经粘膜、牙龈、牙内、耳蜗内、经鼓膜、器官内、硬膜外、鞘内、肌肉内、静脉内、血管内、骨内、眼周、瘤内、脑内和脑室内施用。

[1052] 在某些实施方案中,将本文所述的药物组合物局部施用至患病部位(例如肿瘤部

位)。在某些实施方案中,本文所述的药物组合物通过注射、借助导管、借助栓剂或借助植入物施用于受试者,该植入物是多孔、无孔或凝胶状材料(包括膜,诸如硅橡胶膜(silastic membrane)或纤维)。

[1053] 在某些实施方案中,本文所述的药物组合物在控释系统中递送。在某些实施方案中,可以使用泵(参见例如,Langer,1990,Science 249:1527-1533;Sefton,1989,CRC Crit.Ref.Biomed.Eng.14:201;Buchwald等人,1980,Surgery 88:507;Saudek等人,1989,N.Engl.J.Med.321:574)。在某些实施方案中,可以使用聚合性材料。参见例如,Medical Applications of Controlled Release(Langer和Wise编,CRC Press,Boca Raton,Fla.,1974);Controlled Drug Bioavailability,Drug Product Design and Performance(Smolen和Ball编,Wiley,New York,1984);Ranger和Peppas,1983,Macromol.Sci.Rev.Macromol.Chem.23:61。还可参见Levy等人,1985,Science 228:190;During等人,1989,Ann.Neurol.25:351;Howard等人,1989,J.Neurosurg.71:105。其他控释系统例如在Langer(同上)中有讨论。

[1054] 在某些实施方案中,药物组合物根据常规程序被配制成适于静脉内或皮下施用至受试者(诸如人)的组合物。在某些实施方案中,用于通过注射施用的药物组合物是无菌等渗溶液,其用作增溶剂和局部麻醉剂(诸如用于舒缓注射部位处的疼痛的利多卡因)。

[1055] 一般来说,成分单独地或混合在一起以单位剂型(例如作为干燥冻干粉末或无水浓缩物)提供于指示活性剂的量的密闭容器(诸如安瓿或药囊)中。

[1056] 在药物将通过输注施用的情况下,组合物可以用含有无菌药用级水或盐水的输注瓶来分配。在药物组合物通过注射施用的情况下,注射用无菌水或盐水的安瓿可以被提供成使得成分可以在施用之前被混合。

[1057] 用于全身施用的药物组合物可以是液体,例如无菌盐水、乳酸林格氏(Ringer's)溶液或汉克氏(Hank's)溶液。此外,药物组合物可以呈固体形式并在临使用之前被再溶解或悬浮。还设想了冻干形式。药物组合物可以包含在脂质颗粒或囊泡(诸如脂质体或微晶)中,其也适用于胃肠外施用。该颗粒可以具有任何合适的结构,诸如单层或多层,只要其中含有组合物即可。化合物可以被包埋在含有融合脂质二油酰磷脂酰乙醇胺(DOPE)(低水平的(5-10mol%)阳离子脂质)的"经稳定化的质粒脂质颗粒"(SPLP)中,并且通过聚乙二醇(PEG)涂层而稳定化(参见例如Zhang Y.P.等人,Gene Ther.1999,6:1438-47)。带正电荷的脂质,诸如1,2-二油酰基-3-三甲基铵-丙烷,或"DOTAP"对于此类颗粒和囊泡是特别优选的。此类脂质颗粒的制备是众所周知的。参见例如,美国专利号4,880,635、4,906,477、4,911,928、4,917,951、4,920,016、4,921,757;每一篇均以引用方式并入本文。

[1058] 本文所述的药物组合物可以以单位剂量施用或封装。术语"单位剂量"在提及本公开的药物组合物时被使用时是指适合作为用于受试者的单位剂量的物理上离散的单位,每个单位均含有经计算与所需稀释剂(即,载剂或媒介物)联合产生所期望的治疗效果的预定量的活性物质。

[1059] 进一步地,药物组合物可以作为药物试剂盒提供,该试剂盒包括(a)含有呈冻干形式的本发明化合物的容器和(b)含有药学上可接受的稀释剂(例如,用于重构或稀释本发明的冻干化合物的无菌稀释剂)的第二容器。任选地,与此类容器相伴随的可以是由管制医药或生物产品的制造、使用或销售的政府机构规定的形式的通知,该通知反映了制造、使用和

销售机构对人类施用的批准。

[1060] 在另一方面,包括含有可用于治疗上述疾病的材料的制品。在某些实施方案中,制品包括容器和标签。适合容器包括例如瓶、小瓶、注射器和试管。该容器可以由诸如玻璃或塑料的各种材料形成。在某些实施方案中,容器容纳有效治疗疾病的组合物(例如本文所述的重组狂犬病病毒基因组或重组狂犬病病毒)并且可以具有无菌入口。例如,该容器可以是具有皮下注射针可刺穿的塞子的静脉溶液袋或小瓶。组合物中的活性剂是本公开的化合物(例如重组狂犬病病毒基因组或重组狂犬病病毒)。在某些实施方案中,容器上或与容器相伴随的标签指示组合物用于治疗选择的疾病。制品可以进一步包括第二容器,其包含药学上可接受的缓冲剂诸如磷酸盐缓冲盐水、林格氏液及右旋糖溶液。它可以进一步包括从商业和使用者观点来说所需的其他材料,包括其它缓冲剂、稀释剂、过滤器、针、注射器以及带有使用说明的包装插页。

[1061] 在一些实施方案中,本文所述的重组狂犬病病毒基因组或重组狂犬病病毒中的任何一种作为药物组合物的一部分提供。在一些实施方案中,药物组合物包含本文所述的重组狂犬病病毒基因组或重组狂犬病病毒中的任何一种。在一些实施方案中,药物组合物包含本文提供的复合物中的任何复合物。

[1062] 在一些实施方案中,将本文提供的组合物施用于受试者,例如施用于人受试者,以在受试者内实现被靶向的基因组修饰。在一些实施方案中,从受试者获得细胞并且使其与本文提供的药物组合物中的任何药物组合物接触。在一些实施方案中,将从受试者取出并且离体与药物组合物接触的细胞重新引入到受试者中(任选地在已在该细胞中实现或检测到所需的基因组修饰之后)。递送包含核酸酶的药物组合物的方法是已知的,并且描述于例如美国专利号6,453,242、6,503,717、6,534,261、6,599,692、6,607,882、6,689,558、6,824,978、6,933,113、6,979,539、7,013,219和7,163,824,所有这些的公开内容以引用方式整体并入本文。虽然本文提供的药物组合物的描述主要针对适合于向人施用的药物组合物,但是技术人员将理解此类组合物一般适合于向所有种类的动物或生物体施用。对适合于向人施用的药物组合物进行的以使该组合物适合于向各种动物施用的修改是众所周知的,并且普通兽医药理学家可以仅通过普通实验(如果有的话)来设计和/或进行此种修改。所设想的药物组合物可施用于的受试者包括但不限于人类和/或其他灵长类动物;哺乳动物,家养动物、宠物和商业上相关的哺乳动物诸如牛、猪、马、绵羊、猫、犬、小鼠和/或大鼠;和/或鸟类,包括商业上相关的鸟类,诸如鸡、鸭、鹅和/或火鸡。

[1063] 本文所述的药物组合物的制剂可以通过药理学领域中已知的或以后开发的任何方法来制备。一般来说,此类制备方法包括以下步骤:将一种或多种活性成分与赋形剂和/或一种或多种其他辅助成分联合,以及然后如果必要和/或需要,将产品成形和/或包装为所需的单剂量或多剂量单位。

[1064] 药物制剂可以另外包含药学上可接受的赋形剂,如本文所使用,药学上可接受的赋形剂包括适合于所需的特定剂型的任何和所有溶剂、分散介质、稀释剂或其它液体媒介物、分散或悬浮助剂、表面活性剂、等渗剂、增稠剂或乳化剂、防腐剂、固体粘合剂、润滑剂等。Remington的The Science and Practice of Pharmacy,第21版,A.R.Gennaro(Lippincott,Williams&Wilkins,Baltimore,MD,2006;其以引用方式整体并入本文)公开了用于配制药物组合物的各种赋形剂和用于药物组合物的制备的已知技术。对于用于生产

包含核酸酶的药物组合物的另外的合适的方法、试剂、赋形剂和溶剂,还可参见PCT申请PCT/US2010/055131(公布号WO2011053982 A8,提交于2010年11月2日),其以引用方式整体并入本文。除非任何常规赋形剂介质诸如由于产生任何不需要的生物效应或另外以有害的方式与药物组合物的任何一种或多种其它组分相互作用而与物质或其衍生物不相容,否则常规赋形剂介质的使用被设想了在本公开的范围内。在某些实施方案中,根据本发明的组合物可用于治疗多种疾病、病症和/或疾患中的任何一种。

[1065] 除非另有说明,否则本公开的各个方面采用分子生物学(包括重组技术)、微生物学、细胞生物学、生物化学和免疫学的常规技术,这些技术完全在技术人员的技术范围内。此类技术在诸如以下的文献中得到了充分的解释:"Molecular Cloning:A Laboratory Manual",第二版(Sambrook,1989);"Oligonucleotide Synthesis"(Gait,1984);"Animal Cell Culture"(Freshney,1987);"Methods in Enzymology"和"Handbook of Experimental Immunology"(Weir,1996);"Gene Transfer Vectors for Mammalian Cells"(Miller和Calos,1987);"Current Protocols in Molecular Biology"(Ausubel,1987);"PCR:The Polymerase Chain Reaction",(Mullis,1994);"Current Protocols in Immunology"(Coligan,1991)。这些技术可应用于本公开的各个方面的生产,因此,在制作和实践本公开时可以考虑这些技术。

[1066] H.多核苷酸、载体和细胞

[1067] 本文提供了多核苷酸,其包含:(i)本文所述的重组狂犬病病毒基因组;(ii)编码狂犬病病毒核蛋白或其功能变体的N基因;(iii)编码狂犬病病毒磷蛋白或其功能变体的P基因;(iv)编码狂犬病病毒聚合酶(例如RNA依赖性RNA聚合酶)或其功能变体的L基因;(v)编码狂犬病病毒糖蛋白或其功能变体的G基因;和/或(vi)编码狂犬病病毒基质蛋白或其功能变体的M基因。

[1068] 本文描述的多核苷酸可以通过本领域已知的任何方法,诸如通过化学合成DNA链、通过PCR或通过Gibson组装方法获得。通过化学合成,或者PCR方法或Gibson组装方法的组合构建全长DNA的优点是可以优化密码子以确保融合蛋白在宿主细胞中以高水平表达。可以使用Kazusa DNA研究所的主页中公开的遗传密码使用频率数据库(http://www.kazusa.or.jp/codon/index.html)来选择经优化的密码子。在某些实施方案中,多核苷酸是经密码子优化的。在某些实施方案中,可以通过RNA优化来优化多核苷酸。为了增加重组表达的稳定性而可以包括另外的优化方法,包括例如用外源信号序列替换信号序列、去除不稳定性元件、去除抑制区、去除潜在剪接位点以及本领域普通技术人员已知的其他优化方法。参见例如美国专利号6,794,498,其公开内容通过引用以整体并入本文。

[1069] 一旦获得,本公开的多核苷酸就可被掺入到合适的表达载体中。因此,本公开还提供了包含单独或呈组合形式的本文公开的多核苷酸中的任何多核苷酸的载体。合适的载体包括质粒、病毒、人工染色体、杆粒、粘粒和本领域普通技术人员已知的其他载体。在某些实施方案中,载体是表达载体。

[1070] 合适的表达载体包括源自大肠埃希氏菌的质粒(例如,pBR322、pBR325、pUC12、pUC13);源自枯草芽孢杆菌的质粒(例如,pUB110、pTP5、pC194);源自酵母的质粒(例如pSH19、pSH15);适合在昆虫细胞中表达的质粒(例如,pFast-Bac);适合在哺乳动物细胞中表达的质粒(例如,pXT1、pRc/CMV、pRc/RSV、pcDNA1/Neo);还有噬菌体,诸如λ噬菌体等;可

以使用的其他载体包括昆虫病毒载体,诸如杆状病毒等(例如BmNPV、AcNPV);和适合在哺乳动物细胞中表达的病毒载体,诸如逆转录病毒、牛痘病毒、腺病毒等。

[1071] 包含多核苷酸和载体的基因和/或转基因通常在转录调控元件的控制下表达。在某些实施方案中,转录调控元件可包含一种或多种增强子元件、内含子元件和/或启动子元件。在某些实施方案中,转录调控元件包含组成型启动子。转录调控元件的实例包括包含以下的那些:CMV启动子(来自人巨细胞病毒的启动子),任选地包括CMV增强子、EF1α启动子(来自人延伸因子1α的启动子)、CBA启动子(包括CMV早期增强子和鸡β-肌动蛋白启动子)、CAG启动子(包括CBA启动子和兔β-球蛋白内含子)、CAGGS启动子(包括CMV增强子、CBA启动子和鸡β-肌动蛋白内含子1/外显子1)、PGK启动子(来自磷酸甘油酸激酶的启动子)、U6启动子(U6核启动子)、Ubc启动子(来自人泛素C的启动子)、CASI启动子(包括CMV增强子、泛素C增强子和鸡β-肌动蛋白启动子),和CALM1启动子(来自钙调蛋白1的启动子)。各种组成型转录调控元件是本领域普通技术人员已知的。

[1072] 在某些实施方案中,转录调控元件包含诱导型启动子。例如,转录调控元件可包含诱导型TRE启动子(四环素反应元件启动子)。此类诱导型启动子可以是正诱导型,其中启动子是无活性的,因为激活子蛋白不能与启动子结合;或者是负诱导型,其中阻抑因子与启动子结合从而防止转录。诱导型启动子的实例包括化学诱导型启动子,例如四环素ON(Tet-On)启动子系统、lac阻抑因子系统、pBad原核启动子和其他(诸如醇或类固醇调控的启动子)。诱导型启动子可以是温度诱导型,例如热或冷诱导型启动子(例如,Hsp70或Hsp90衍生的启动子),以及光诱导型,在这种情况下光可用于调控转录。在某些实施方案中,转录调控元件包含阻抑型启动子。各种诱导型转录调控元件是本领域普通技术人员已知的。

[1073] 在某些实施方案中,转录调控元件包含对于基因或转基因来说是外源的启动子。在某些实施方案中,转录调控元件包含合成型启动子。

[1074] 合适的启动子可以根据其用于在所需宿主细胞中表达的用途来选择。例如,当宿主是动物细胞时,使用以下启动子中的任一种:SR-α启动子、SV40启动子、LTR启动子、CMV(巨细胞病毒)启动子、RSV(劳斯肉瘤病毒)启动子、MoMuLV(莫洛尼小鼠白血病病毒)LTR、HSV-TK(简单疱疹病毒胸苷激酶)启动子等。在某些实施方案中,启动子是CMV启动子或SRα启动子。在某些实施方案中,启动子是延伸因子1-α(EF1α)启动子。当宿主细胞是大肠埃希氏菌时,可以使用以下启动子中的任何启动子:trp启动子、lac启动子、recA启动子、λPL启动子、lpp启动子、T7启动子等。当宿主为芽孢杆菌属时,可以使用以下启动子中的任何启动子:SPO1启动子、SPO2启动子、penP启动子等。当宿主是酵母时,可以使用以下启动子中的任何启动子:Gal1/10启动子、PHO5启动子、PGK启动子、GAP启动子、ADH启动子等。当宿主是昆虫细胞时,可以使用以下启动子中的任何启动子:多角体启动子、P10启动子等。当宿主是植物细胞时,可以使用任何以下启动子中的任何启动子:CaMV35S启动子、CaMV19S启动子、NOS启动子等。

[1075] 如果需要,表达载体还包括增强子、剪接信号、终止子、多聚腺苷酸化信号、选择标志物(例如,药物抗性基因、营养缺陷型互补基因等)或复制起点中的任何一种或多种。

[1076] 可以使用本领域技术人员已知的常规方法将本公开的多核苷酸引入到几乎任何关注的宿主细胞中,该宿主细胞包括但不限于细菌、酵母、真菌、昆虫、植物和动物细胞。

[1077] 埃希氏菌属包括大肠埃希氏菌K12/DH1、大肠埃希氏菌JM103、大肠埃希氏菌

JA221、大肠埃希氏菌HB101、大肠埃希氏菌C600等。芽孢杆菌属包括枯草芽孢杆菌Ml 114、枯草芽孢杆菌207-21等。

[1078] 可用于宿留本公开的多核苷酸的酵母包括酿酒酵母(Saccharomyces cerevisiae)AH22、AH22 R⁻、NA87-11A、DKD-5D、20B-12、粟酒裂殖酵母(Schizosaccharomyces pombe)NCYC1913、NCYC2036、巴斯德毕赤酵母(Pichia pastoris)KM71等。

[1079] 可以使用例如病毒载体如AcNPV将本公开的多核苷酸引入到昆虫细胞中。昆虫宿主细胞包括以下细胞系中的任何细胞系:源自夜盗蛾(cabbage armyworm)幼虫的已确立系(草地贪夜蛾(Spodoptera frugiperda)细胞;Sf细胞)、源自粉纹夜蛾(Trichoplusiani)的中肠的MG1细胞、源自粉纹夜蛾的卵的High Five细胞、源自甘蓝夜蛾(Mamestra brassicae)的细胞、源自盐泽顶灯蛾(Estigmena acrea)的细胞等。当病毒是BmNPV时,使用源自家蚕(Bombyx mori)的细胞系(家蚕N细胞;BmN细胞)等。Sf细胞包括例如Sf9细胞(ATCC CRL1711)、Sf21s细胞等。

[1080] 可以使用哺乳动物细胞系,包括但不限于猴COS-7细胞、猴Vero细胞、中国仓鼠卵巢(CHO)细胞、dhfr基因缺陷的CHO细胞、小鼠L细胞、小鼠AtT-20细胞、小鼠骨髓瘤细胞、大鼠GH3细胞、人FL细胞、人胚肾(HEK)细胞(例如HEK293、HEK293T)、COS细胞(例如COS1或COS)、BHK细胞、MDCK细胞、NSO细胞、PER.C6细胞、CRL7O3O细胞、HsS78Bst细胞、HeLa细胞、NIH 3T3细胞、HepG2细胞、SP210细胞、R1.1细胞、B-W细胞、L-M细胞、BSC1细胞、BSC40细胞、YB/20细胞和BMT10细胞等。

[1081] 在某些实施方案中,合适的细胞是哺乳动物、细菌或昆虫来源的。在某些实施方案中,细胞选自由HEK293细胞、HEK293T细胞、VERO细胞、BHK细胞和BSR细胞组成的组。

[1082] 所有上面提到的宿主细胞都可以是半倍体(haploid)(单倍体(monoploid)),或多倍体(例如二倍体、三倍体、四倍体等)。

[1083] 将本公开的多核苷酸引入到本文所述的宿主细胞中的各种方法是本领域普通技术人员已知的。例如,此类方法可包括使用本领域已知的任何转染方法(例如,使用溶菌酶、PEG、CaCl2共沉淀、电穿孔、显微注射、基因枪、脂转染、农杆菌等)。根据待转染的宿主细胞选择转染方法。大肠埃希氏菌可以根据例如Proc.Natl.Acad.Sci.USA,69,2110(1972),Gene,17,107(1982)中描述的方法进行转化。用于转导芽孢杆菌属的方法描述于,例如Molecular&General Genetics,168,111(1979)。使用例如Methods in Enzymology,194,182-187(1991),Proc.Natl.Acad.Sci.USA,75,1929(1978)等中描述的方法转导酵母细胞。使用例如Bio/Technology,6,47-55(1988)等中描述的方法转染昆虫细胞。使用例如Cell Engineering增刊第8卷、New Cell Engineering Experiment Protocol,263-267(1995)(Shujunsha出版)和Virology,52,456(1973)中描述的方法转染哺乳动物细胞。

[1084] 包含本公开的表达载体的细胞根据已知方法进行培养,该方法根据宿主而变化。例如,当培养大肠埃希氏菌或芽孢杆菌属细胞时,使用液体培养基。该培养基优选含有碳源、氮源、无机物和转化体生长所必需的其他组分。碳源的实例包括葡萄糖、糊精、可溶性淀粉、蔗糖等;氮源的实例包括无机或有机物质,诸如铵盐、硝酸盐、玉米浆(corn steep liquor)、蛋白胨、酪蛋白、肉提取物、豆饼、马铃薯提取物等;以及无机物的实例包括氯化钙、磷酸二氢钠、氯化镁等。培养基还可以含有酵母提取物、维生素、生长促进因子等。培养

基的pH优选为约5至约8。作为用于培养大肠埃希氏菌的培养基,使用例如含有葡萄糖和酪蛋白氨基酸的M9培养基(参见例如,Journal of Experiments in Molecular Genetics,431-433,Cold Spring Harbour Laboratory,New York 1972)。大肠埃希氏菌一般在约15℃至约43℃下培养。根据需要,可以进行通气和搅拌。芽孢杆菌属一般在约30℃至约40℃下培养。根据需要,进行通气和搅拌。

[1085] 适合于培养酵母的培养基的实例包括Burkholder最低限度培养基、含有0.5%酪蛋白氨基酸的SD培养基等。培养基的pH优选为约5-约8。培养一般在约20℃至约35℃下进行。根据需要,可以进行通气和搅拌。

[1086] 作为用于培养昆虫细胞或昆虫的培养基,使用含有诸如经灭活的10%牛血清等添加剂的Grace's昆虫培养基。培养基的pH优选为约6.2至约6.4。细胞在约27℃下培养。根据需要,可以进行通气和搅拌。

[1087] 例如,在含有约5%至约20%胎牛血清的最低限度必需培养基(MEM)、Dulbecco's改良Eagle培养基(DMEM)、RPMI 1640培养基、199培养基等中的任一种中培养哺乳动物细胞。培养基的pH优选为约6至约8。培养在约30℃至约40℃下进行。根据需要,可以进行通气和搅拌。

[1088] I.包装系统及其方法

[1089] 本公开提供了可用于重组制备本文所述的狂犬病病毒颗粒的包装系统。特别地,该包装系统提供了制备本文所述的狂犬病病毒颗粒所需的必要组分。在某些实施方案中,包装系统可用于重组制备包含重组狂犬病病毒基因组的狂犬病病毒颗粒,其中该基因组缺乏编码狂犬病病毒糖蛋白或其功能变体的G基因;并且/或者该基因组缺乏编码狂犬病病毒聚合酶或其功能变体的L基因。在某些实施方案中,包装系统可用于重组制备包含重组狂犬病病毒基因组的狂犬病病毒颗粒,其中该基因组缺乏编码狂犬病病毒糖蛋白或其功能变体的G基因。在某些实施方案中,包装系统可用于重组制备包含重组狂犬病病毒基因组的狂犬病病毒颗粒,其中该基因组缺乏编码狂犬病病毒糖蛋白或其功能变体的G基因;并且该基因组缺乏编码狂犬病病毒聚合酶或其功能变体的L基因。

[1090] 本文所述的包装系统一般包含以下或由以下组成:(i)编码狂犬病病毒核蛋白或其功能变体的N基因;(ii)编码狂犬病病毒磷蛋白或其功能变体的P基因;和(iii)编码狂犬病病毒聚合酶或其功能变体的L基因。在某些实施方案中,包装系统进一步包含编码狂犬病病毒基质蛋白或其功能变体的M基因。在某些实施方案中,包装系统进一步包含编码狂犬病病毒糖蛋白或其功能变体的G基因。

[1091] 包装系统的N、P和L基因可以在一种或多种载体(例如,转染质粒)中提供。例如,包装系统可以包含用于N、P和L基因中的每一者的单独的转染质粒,例如,包含编码狂犬病病毒核蛋白或其功能变体的N基因的第一转染质粒;包含编码狂犬病病毒磷蛋白或其功能变体的P基因的第二转染质粒;和包含编码狂犬病病毒聚合酶或其功能变体的L基因的第三转染质粒。在某些实施方案中,单个转染质粒包含N、P和L基因中的两种或更多种。例如,包装系统可以包含转染质粒,该转染质粒包含编码狂犬病病毒核蛋白或其功能变体的N基因和编码狂犬病病毒磷蛋白或其功能变体的P基因;包装系统可以包含转染质粒,该转染质粒包含编码狂犬病病毒核蛋白或其功能变体的N基因和编码狂犬病病毒聚合酶或其功能变体的L基因;包装系统可以包含转染质粒,该转染质粒包含编码狂犬病病毒磷蛋白或其功能变体

的P基因和编码狂犬病病毒聚合酶或其功能变体的L基因。在某些实施方案中,包装系统可以包含转染质粒,该转染质粒包含编码狂犬病病毒核蛋白或其功能变体的N基因、编码狂犬病病毒磷蛋白或其功能变体的P基因,和编码狂犬病病毒聚合酶或其功能变体的L基因。

[1092] 包装系统的M和G基因可以在一种或多种转染质粒中提供。在某些实施方案中,包装系统包含用于M和G基因的单独的转染质粒。例如在某些实施方案中,包装系统可进一步包含转染质粒,该转染质粒包含编码狂犬病病毒基质蛋白或其功能变体的M基因。在某些实施方案中,包装系统可进一步包含转染质粒,该转染质粒包含编码狂犬病病毒糖蛋白或其功能变体的G基因。该M和/或G基因还可以被组合到包含如本文所述的N、P和/或L基因的转染质粒中。例如,单个转染质粒可包含编码狂犬病病毒核蛋白或其功能变体的N基因、编码狂犬病病毒磷蛋白或其功能变体的P基因、编码狂犬病病毒聚合酶或其功能变体的L基因、编码狂犬病病毒基质蛋白或其功能变体的M基因,和编码狂犬病病毒糖蛋白或其功能变体的G基因。本领域普通技术人员可以容易地理解各种其他组合。

[1093] N、P、L、M和/或G基因均可处于一种或多种转录调控元件的控制之下。在某些实施方案中,转录调控元件包含启动子和/或增强子序列。在某些实施方案中,转录调控元件包含EF1α启动子。各种启动子和/或增强子序列是本领域已知的并且在本文中作为实例进行了描述,并且本领域普通技术人员将能够根据其需要选择合适的启动子和/或增强子序列。

[1094] 当N、P、L、M和/或G基因中的两者或更多者位于同一载体上时,该两个或更多个基因可以存在于一个或多个表达盒中。例如,N、P、L、M和/或G基因中的每一者均可以位于它们自己的表达盒内,每个表达盒均包含转录调控元件和/或转录终止元件。

[1095] 当两个或更多个基因位于同一表达盒中时,这些基因可以被接头序列分开。在某些实施方案中,接头序列是核糖体跳跃元件,该核糖体跳跃元件包含编码内部核糖体进入位点(IRES)的核酸序列。如本文所用,“内部核糖体进入位点”或“IRES”是指促进内部核糖体直接进入蛋白质编码区的起始密码子(诸如ATG),从而导致基因的帽独立翻译的元件。各种内部核糖体进入位点是本领域技术人员已知的,包括但不限于可从病毒或细胞mRNA来源获得的IRES,例如免疫球蛋白重链结合蛋白(BiP);血管内皮生长因子(VEGF);成纤维细胞生长因子2;胰岛素样生长因子;翻译起始因子eIF4G;酵母转录因子TFIID和HAP4;以及可从例如心脏病毒、鼻病毒、口疮病毒、HCV、Friend鼠白血病病毒(FrMLV)和莫洛尼鼠白血病病毒(MoMLV)获得的IRES。在某些实施方案中,接头序列是包含编码自切割肽的核酸序列的核糖体跳跃元件。如本文所用,“自切割肽”或“2A肽”是指允许多种蛋白质被编码为多蛋白的寡肽,该多蛋白在翻译后解离成组分蛋白。术语“自切割”的使用并不旨在暗示蛋白水解切割反应。各种自切割肽或2A肽是本领域技术人员已知的,包括但不限于在小核糖核酸病毒科病毒家族成员,例如口蹄疫病毒(FMDV)、马鼻炎A病毒(ERAVO、明脉扁刺蛾病毒(Thatea asigna)病毒(TaV)和猪捷申病毒-1(porcine tescho virus-1,PTV-1);以及卡里奥病毒(cariovirus)如泰勒病毒(theilovirus)和脑心肌炎病毒中发现的那些肽。衍生自FMDV、ERAV、PTV-1和TaV的2A肽在本文中分别被称为“F2A”、“E2A”、“P2A”和“T2A”。本领域技术人员将能够根据他们的需要选择合适的接头序列。

[1096] 在某些实施方案中,单一载体(例如转染质粒)包含含有N和P基因的第一表达盒和含有L基因的第二表达盒。在某些实施方案中,第一表达盒从5’至3’包含:转录调控元件、P基因,和N基因。在某些实施方案中,第一表达盒从5’至3’包含:转录调控元件、P基因、核糖

体跳跃元件,和N基因。在某些实施方案中,第二表达盒从5'至3'包含:转录调控元件,和L基因。在某些实施方案中,第一表达盒和第二表达盒在载体内可以处于相同的取向。在某些实施方案中,第一表达盒和第二表达盒在载体内可以处于相反的取向。

[1097] 因此,本发明的包装系统包含:(i)重组狂犬病病毒基因组载体(例如,病毒基因组转染质粒);(ii)包含N、P、L、M和/或G基因的一种或多种转染质粒。可以使用本领域普通技术人员已知的各种方法将该包含N、P、L、M和/或G基因的一种或多种转染质粒引入到宿主细胞(例如重组狂犬病病毒颗粒包装细胞)中。例如,可以通过电穿孔、核转染或脂转染将该一种或多种转染质粒引入到合适的宿主细胞中。

[1098] 本公开还提供了一种用于重组制备狂犬病病毒颗粒的方法,其中该方法包括将本文所述的包装系统在可有效包裹重组狂犬病病毒基因组的条件下引入到细胞中以形成重组狂犬病病毒颗粒。在某些实施方案中,可以用该包含N、P、L、M和/或G基因的一种或多种转染质粒瞬时转染宿主包装细胞。在某些实施方案中,可以用该包含N、P、L、M和/或G基因的一种或多种转染质粒转染宿主包装细胞,其中宿主包装细胞进一步被制成稳定细胞系。用于产生稳定细胞系的各种方法是本领域普通技术人员已知的。一般而言,将关注的基因(例如,N、P、L、M和/或G基因)引入到细胞中,并且然后引入到细胞核中,并且最后整合到细胞的基因组中。染色体整合事件很少见,并且必须选择和培养稳定整合的细胞系。多种选择系统是本领域已知的,包括对诸如新霉素磷酸转移酶的抗生素的抗性,赋予对G418、二氢叶酸还原酶(DHFR)或谷氨酰胺合成酶的抗性。用于产生稳定细胞系的其他方法包括使用如实验实施例中所述的睡美人(Sleeping Beauty,SB)系统。简而言之,包含关注的整合体(integrant)的转座子被设计为具有侧接反向重复/同向重复序列,该反向重复/同向重复序列导致精确整合到TA二核苷酸中。用于产生基于SB转座子的稳定细胞系的方法是本领域已知的,参见例如Davidson等人,Cold Spring Harb Protoc.(2009)4(8):1018-1023。还可以经由使用慢病毒载体产生稳定细胞系,参见例如Tandon等人,Bio Protoc.(2018)8(21):e3073。

[1099] 然后将重组狂犬病病毒基因组载体(例如,病毒基因组转染质粒)引入到其中具有稳定整合或瞬时转染的N、P、L、M和/或G基因的宿主包装细胞中。

[1100] 因此,在某些实施方案中,用于重组制备狂犬病病毒颗粒的方法包括引入(i)重组狂犬病病毒基因组载体(例如,病毒基因组转染质粒);以及(ii)将包含N、P、L、M和/或G基因的一种或多种转染质粒至宿主包装细胞中。在某些实施方案中,用于重组制备狂犬病病毒颗粒的方法包括将重组狂犬病病毒基因组载体(例如,病毒基因组转染质粒)引入到宿主包装细胞中,其中该宿主包装细胞包含稳定地整合于其中的N、P、L、M和/或G基因。用于制备重组狂犬病病毒颗粒的方法是本领域已知的,参见例如Trabelsi等人,Vaccine(2019)37(47):7052-7060;Wickersham等人,Nature Protoc.(2010)5(3):595-606;Ghanem等人,Eur.J.Cell Biol.(2012)91:10-16;Osakada和Wickersham,Nature Protoc.(2013)8(8):1583-1601;以及Sullivan和Wickersham,Cold Spring Harb Protoc.(2015)4:386-91,其公开内容通过引用整体并入本文。

[1101] 在某些实施方案中,使用本文所述的生产方法获得的重组狂犬病病毒颗粒滴度大于约1E8转导单位(TU)/mL。例如,在某些实施方案中,获得的重组狂犬病病毒颗粒滴度为约8E7 TU/mL、约9E7 TU/mL、约1E8 TU/mL、约1.1E8 TU/mL、约1.2E8 TU/mL、约1.3E8 TU/mL、

约1.4E8 TU/mL、约1.5E8 TU/mL、约1.6E8 TU/mL、约1.7E8 TU/mL、约1.8E8 TU/mL、约1.9E8 TU/mL、约2E8 TU/mL、约2.5E8 TU/mL、约3E8 TU/mL、约3.5E8 TU/mL、约4E8 TU/mL、约4.5E8 TU/mL、约5E8 TU/mL、约5.5E8 TU/mL、约6E8 TU/mL、约6.5E8 TU/mL、约7E8 TU/mL、约7.5E8 TU/mL、约8E8 TU/mL、约8.5E8 TU/mL、约9E8 TU/mL、约9.1E8 TU/mL、约9.2E8 TU/mL、约9.3E8 TU/mL、约9.4E8 TU/mL、约9.5E8 TU/mL、约9.6E8 TU/mL、约9.7E8 TU/mL、约9.8E8 TU/mL、约9.9E8 TU/mL、约1E9 TU/mL、约1.1E9 TU/mL、约1.2E9 TU/mL,或前面提到的滴度之间的任何值。在某些实施方案中,获得的重组狂犬病病毒颗粒滴度为约1E8 TU/mL至约1E9 TU/mL,例如8E7 TU/mL至1.2E9 TU/mL,及其间的任何范围。

[1102] J.基因疗法方法

[1103] 本文提供了使用本文所述的重组狂犬病病毒颗粒的基因疗法的方法。在某些实施方案中,提供了用于使治疗性转基因在靶细胞中表达的方法。在某些实施方案中,提供了用于使碱基编辑器在靶细胞中表达的方法。

[1104] 在某些实施方案中,用于使治疗性转基因在靶细胞中表达的方法包括用如本文所述的重组狂犬病病毒颗粒转导靶细胞。例如,用于使治疗性转基因在靶细胞中表达的方法包括用重组狂犬病病毒颗粒转导靶细胞,该重组狂犬病病毒颗粒包含狂犬病病毒糖蛋白和包含编码治疗性转基因的核酸的重组狂犬病病毒基因组,其中:该基因组缺乏编码狂犬病病毒糖蛋白或其功能变体的G基因;并且/或者该基因组缺乏编码狂犬病病毒聚合酶或其功能变体的L基因。在某些实施方案中,方法包括用重组狂犬病病毒颗粒转导靶细胞,该重组狂犬病病毒颗粒包含狂犬病病毒糖蛋白和包含编码治疗性转基因的核酸的重组狂犬病病毒基因组,其中:该基因组缺乏编码狂犬病病毒糖蛋白或其功能变体的G基因。在某些实施方案中,方法包括用重组狂犬病病毒颗粒转导靶细胞,该重组狂犬病病毒颗粒包含狂犬病病毒糖蛋白和包含编码治疗性转基因的核酸的重组狂犬病病毒基因组,其中:该基因组缺乏编码狂犬病病毒糖蛋白或其功能变体的G基因;并且该基因组缺乏编码狂犬病病毒聚合酶或其功能变体的L基因。

[1105] 用重组病毒颗粒转导靶细胞的各种方法是本领域普通技术人员已知的。例如,可以使靶细胞与重组病毒颗粒接触,从而导致受体介导的病毒颗粒的附着,随后病毒颗粒通过网格蛋白依赖性内吞作用进入该细胞中。

[1106] 在某些实施方案中,提供了用于使核碱基编辑器在靶细胞中表达的方法。例如,此类方法包括用重组狂犬病病毒颗粒转导靶细胞,其中该重组病毒颗粒包含:狂犬病病毒糖蛋白和包含编码核碱基编辑器的核酸的重组狂犬病病毒基因组,该核碱基编辑器包含多核苷酸可编程核苷酸结合结构域和核碱基编辑结构域,其中:该基因组缺乏编码狂犬病病毒糖蛋白或其功能变体的G基因;并且/或者该基因组缺乏编码狂犬病病毒聚合酶或其功能变体的L基因。在某些实施方案中,方法包括用重组狂犬病病毒颗粒转导靶细胞,其中该重组病毒颗粒包含:狂犬病病毒糖蛋白和包含编码核碱基编辑器的核酸的重组狂犬病病毒基因组,该核碱基编辑器包含多核苷酸可编程核苷酸结合结构域和核碱基编辑结构域,其中:该基因组缺乏编码狂犬病病毒糖蛋白或其功能变体的G基因。在某些实施方案中,方法包括用重组狂犬病病毒颗粒转导靶细胞,其中该重组病毒颗粒包含:狂犬病病毒糖蛋白和包含编码核碱基编辑器的核酸的重组狂犬病病毒基因组,该核碱基编辑器包含多核苷酸可编程核苷酸结合结构域和核碱基编辑结构域,其中:该基因组缺乏编码狂犬病病毒糖蛋白或其功

能变体的G基因;并且该基因组缺乏编码狂犬病病毒聚合酶或其功能变体的L基因。

[1107] 当该方法用于使核碱基编辑器在靶细胞中表达时,多核苷酸可编程核苷酸结合结构域在与被结合的引导多核苷酸(例如gRNA)联合时,可以特异性地结合至靶多核苷酸序列,并且从而将碱基编辑器定位于需要编辑的靶核酸序列。

[1108] 在某些实施方案中,gRNA以顺式提供给靶细胞。例如,gRNA可以被包含在重组狂犬病病毒基因组内。gRNA可以被包含在重组狂犬病病毒基因组内的任何区位处,例如,在一种或多种狂犬病病毒基因(例如,N基因或P基因)与编码核碱基编辑器的核酸之间,或在两个狂犬病病毒基因之间,或在重组狂犬病病毒基因组的末端(例如,5'端或3'端)处。

[1109] 在某些实施方案中,gRNA以反式提供给靶细胞(例如,外源地提供)。例如,gRNA可以被包含在重组狂犬病病毒颗粒外部的单独载体内。合适的载体包括但不限于病毒载体、质粒和本领域技术人员已知的其他载体。在gRNA以反式提供给靶细胞的实施方案中,gRNA载体经由本领域技术人员已知的各种方法(例如但不限于电穿孔)被引入到靶细胞中。

[1110] 还提供了用于将治疗性转基因(例如核碱基编辑器)递送至受试者的方法。在某些实施方案中,方法包括向受试者施用重组狂犬病病毒颗粒,其中该重组病毒颗粒包含:狂犬病病毒糖蛋白和包含编码治疗性转基因(例如包含多核苷酸可编程核苷酸结合结构域和核碱基编辑结构域的核碱基编辑器)的核酸的重组狂犬病病毒基因组,其中:该基因组缺乏编码狂犬病病毒糖蛋白或其功能变体的G基因;并且/或者该基因组缺乏编码狂犬病病毒聚合酶或其功能变体的L基因。在某些实施方案中,方法包括向受试者施用重组狂犬病病毒颗粒,其中该重组病毒颗粒包含:狂犬病病毒糖蛋白和包含编码治疗性转基因(例如包含多核苷酸可编程核苷酸结合结构域和核碱基编辑结构域的核碱基编辑器)的核酸的重组狂犬病病毒基因组,其中:该基因组缺乏编码狂犬病病毒糖蛋白或其功能变体的G基因。在某些实施方案中,方法包括向受试者施用重组狂犬病病毒颗粒,其中该重组病毒颗粒包含:狂犬病病毒糖蛋白和包含编码治疗性转基因(例如包含多核苷酸可编程核苷酸结合结构域和核碱基编辑结构域的核碱基编辑器)的核酸的重组狂犬病病毒基因组,其中:该基因组缺乏编码狂犬病病毒糖蛋白或其功能变体的G基因;并且该基因组缺乏编码狂犬病病毒聚合酶或其功能变体的L基因。

[1111] 递送方法和/或使治疗性转基因(例如,包含多核苷酸可编程核苷酸结合结构域和核碱基编辑结构域的核碱基编辑器)表达的方法可应用于治疗疾病或病症。在某些实施方案中,治疗受试者的疾病或病症的方法包括施用本文所述的重组狂犬病病毒颗粒或本文所述的药物组合物。在某些实施方案中,疾病或病症是神经疾病或病症。在某些实施方案中,疾病或病症是眼科疾病或病症。

[1112] 本文所设想的药物组合物的施用可以使用常规技术进行,该技术包括但不限于输注、输血(transfusion)或肠胃外施用。在一些实施方案中,肠胃外施用包括血管内、静脉内、肌内、动脉内、鞘内、肿瘤内、皮内、腹膜内、经气管、皮下(subcutaneously)、角质层下(subcuticularly)、关节内、囊下、蛛网膜下和胸骨内输注或注射。

[1113] 除非另有说明,否则本发明的实践采用分子生物学(包括重组技术)、微生物学、细胞生物学、生物化学和免疫学的常规技术,这些技术完全在技术人员的技术范围内。此类技术在诸如以下的文献中得到了充分的解释:"Molecular Cloning:A Laboratory Manual",第二版(Sambrook,1989);"Oligonucleotide Synthesis"(Gait,1984);"Animal Cell

Culture"(Freshney,1987);"Methods in Enzymology""Handbook of Experimental Immunology"(Weir,1996);"Gene Transfer Vectors for Mammalian Cells"(Miller和Calos,1987);"Current Protocols in Molecular Biology"(Ausubel,1987);"PCR:The Polymerase Chain Reaction",(Mullis,1994);"Current Protocols in Immunology"(Coligan,1991)。这些技术适用于本发明的多核苷酸和多肽的生产,并且因此可被考虑用于制作和实践本发明。对于特定实施方案特别有用的技术将在下面的部分中进行讨论。

[1114] 提出以下实施例是为了向本领域普通技术人员提供如何进行和使用本发明的测定、筛选和治疗性方法的完整公开和描述,并且不旨在限制发明者认为是其发明的范围。

[1115] K.实验实施例

[1116] 实施例1:稳定细胞系的产生

[1117] 产生了下表16中描述的稳定细胞系:

[1118] 表16:稳定细胞系

[1119]

细胞系名称	描述	整合载体	选择标志物
RABV-G	稳定地表达狂犬病病毒G基因的HEK293T细胞系	VIR120	杀稻瘟菌素
CA3.11	稳定地表达狂犬病病毒N、P和L基因的HEK293T细胞系	VIR069	杀稻瘟菌素
CA4.27	稳定地表达狂犬病病毒N、P和L基因的HEK293T细胞系	VIR071	吉欧霉素(Zeocin)

[1120] 将睡美人转座酶系统相容性整合载体VIR120、VIR069和VIR071与睡美人转座酶SB100X一起共转染至HEK293T细胞中。VIR120含有包含处于EF1-α启动子控制下的狂犬病病毒G基因的表达盒;VIR069含有从5'至3'包含EF1-α启动子、狂犬病病毒N基因、T2A肽、狂犬病病毒P基因、P2A肽和狂犬病病毒L基因的表达盒;并且VIR071含有从5'至3'包含EF1-α启动子、狂犬病病毒M基因、P2A肽、狂犬病病毒P基因、IRES和狂犬病病毒N基因的第一表达盒,以及从5'至3'包含RPBSA启动子和狂犬病病毒L基因的第二表达盒,其中第一表达盒和第二表达盒处于相反的取向。

[1121] 在共转染一天后,根据所用的整合载体,使用杀稻瘟菌素或吉欧霉素开始选择。根据需要,在共转染后第2天至第7天继续进行选择。到第14天,所有存活的细胞都具有稳定整合的转基因。

[1122] 实施例2:重组狂犬病病毒颗粒的生产

[1123] 对于初级生产,在第0天,使用脂质转染胺(lipofectamine)3000将(i)2ug表达载体的补体质粒混合物和(ii)1ug编码狂犬病复制子的质粒转染至稳定细胞系中。根据表17进行转染:

[1124] 表17:转染混合物

[1125]

稳定细胞系	补体质粒混合物	复制子

RABV-G	DNA52	VIR045“G-缺失”
RABV-G	DNA52	VIR092“G/L-缺失”
CA3.11	VIR11+DNA52	VIR045“G-缺失”
CA3.11	VIR11+DNA52	VIR092“G/L-缺失”
CA4.27	VIR11+DNA52	VIR045“G-缺失”
CA4.27	VIR11+DNA52	VIR092“G/L-缺失”

[1126] VIR045复制子含有缺失了G基因的狂犬病SAD L16完整复制子。VIR092复制子源自自进一步缺失了L基因的VIR045。VIR045和VIR092均含有编码GFP的序列。DNA52是包含编码T7 RNA聚合酶的序列的表达载体。VIR11是包含狂犬病病毒G基因的表达载体。

[1127] 第1天,将培养基更换为OptiMem+5% FBS(“05”)。丢弃第1天的培养基。从第3天开始,每天收获病毒上清液并且将培养基用新鲜的05培养基替换。汇集来自第3-7天的病毒上清液并且将其储存在4℃下。

[1128] 通过以4000rpm离心15分钟来使汇集的病毒上清液变澄清以去除细胞碎片。按照Lenti-X浓缩器(Takara Bio)的方案沉淀和浓缩病毒颗粒。除去上清液,并将沉淀重悬于05培养基中以产生浓缩的病毒储液。使用该浓缩的病毒储液对随后的扩增传代物(passages)进行接种。

[1129] 如下进行次级病毒扩增。第0天,将病毒储液添加到稳定细胞系中。如有必要,在转导时将另外的质粒共转染至稳定细胞系中。对于使用VIR045复制子产生的病毒储液,当在RABV-G稳定细胞系中扩增时不需要任何另外的东西。对于使用VIR092复制子产生的病毒储液,扩增按以下方式进行,括号中示出效率-“+”越多表示效率越高:(1)与含有N、P和L基因的质粒共转染的RABV-G稳定细胞系(+);(2)与含有G基因的质粒共转染的CA4.27稳定细胞系(++);和(3)具有进一步稳定整合的G基因的CA4.27稳定细胞系(+++)。

[1130] 第1天,将培养基更换为为05培养基。丢弃第1天的培养基。第2至7天,收获并汇集病毒上清液。

[1131] 在另一项实验中,比较了用VIR045或VIR092复制子转染的原代转染细胞系HEK293T对照细胞、RABV-G、CA3.11和CA4.27之间的GFP表达。将全补体质粒混合物共转染到每个细胞系中。表18示出了基于初级转染后8天拍摄的图像的GFP表达的定性水平,其中“+”越多表明GFP表达越高。

[1132] 表18:原代转染细胞系中的GFP表达

[1133]

复制子	HEK293T	RABV-G	CA3.11	CA4.27
VIR045	+	+++++	+	+++
VIR092	-	-	-	++

[1134] 第2至4天每天收集病毒上清液,将其汇集并通过Lenti-X浓缩器浓缩。将浓缩的VIR045病毒上清液添加到RABV-G细胞中,并将浓缩的VIR092病毒上清液添加到用含有N、P和L基因的质粒转染的RABV-G细胞中。基于转染后2天拍摄的图像的指示重组狂犬病病毒颗粒的产生的GFP表达的定性水平示于表19中,其中“+”越多表明GFP表达越高。

[1135] 表19:第一次扩增中的GFP表达

[1136]

复制子	HEK293T	RABV-G	CA3.11	CA4.27
VIR045	+	+++++	+	+

[1137]

VIR092	+++	+++	+++	+++

[1138] 在另一个实验中,对从使用不同的稳定细胞系获得的病毒上清液测定了重组狂犬病病毒的相对感染性(图1)。

[1139] 稳定细胞系c1、c8、c39、c40、c53和c54是源自自CA3.11稳定细胞系(“本体(bulk)”)的克隆细胞系。还产生了使用整合载体VIR069(“BHK”)和整合载体VIR120(“BHK-G”)的BHK细胞系。将CA4.27细胞以每孔0.4、0.6、0.8或1百万个细胞铺板。

[1140] 在不同的日子(D2或D3)收获病毒上清液,随后使用该病毒上清液以图1上的体积指示值(5uL或30uL)感染初始HEK293T细胞。滴定通过流式细胞术进行,该流式细胞术示出了如通过GFP表达确定的被感染的细胞的百分比。

[1141] 实施例3:重组狂犬病病毒颗粒基因递送

[1142] 为了研究重组狂犬病病毒颗粒是否可用于基因递送,产生了复制子VIR218。VIR218源自自添加了编码腺苷脱氨酶ABE7.10的序列的VIR092;图2A是VIR218的示意图。图2B是示出随后的生产和扩增方案的示意图。通过将VIR218与全补体质粒混合物共转染至初始HEK293T细胞中来进行初级生产。通过在RABV-G细胞系上另外转染含有N、P和L基因的质粒来进行二级和三级扩增。如上所述收集和浓缩病毒上清液以产生病毒储液。然后将病毒储液添加至初始293T细胞中,同时经由脂转染转染包含靶向HEK2-2(gaacacaaagcatagactgc;SEQ ID NO:4011)的gRNA的质粒,并任选地与包含L基因(“补加的L”)的质粒共转染。提取基因组DNA并进行标准PCR/文库制备以扩增出基因组靶标并评估编辑(图2C)。如图2C所示,在受感染的HEK293T细胞中检测到A>G编辑。

[1143] 实施例4:利用切割tRNA将gRNA编码到狂犬病基因组中

[1144] 为了研究gRNA是否可以在狂犬病病毒基因组中被编码,在图3A所示的组织中产生了复制子VIR621。VIR621源自自编码两个侧接切割tRNA和一个间插gRNA(图3B)的DNA538,其中添加了编码ABE8中包含的多核苷酸可编程核苷酸结合结构域和腺苷脱氨酶的序列以及缺乏G基因的病毒基因组(图3A)。多个靶tRNA也在不同的tRNA组合之间或之后被编码,从而允许多重化(图3C、图3D)。针对图3E中的编辑效率测试了表20中列出的tRNA和gRNA的几种组合。如图3E所示,在受感染的HEK293T细胞中用以下病毒复制子检测到HEK2和IEDG基因的A>G编辑:不含gRNA的病毒复制子(VIR596)、含有靶向HEK2的单个gRNA的病毒复制子(VIR621、VIR622)、含有靶向IEDG的单个gRNA的病毒复制子(VIR712、VIR713),或含有靶向同一病毒复制子中的HEK2和IEDG的多路复用的(multiplexed)多个gRNA的病毒复制子(VIR714、VIR715、VIR717、VIR718、VIR719、VIR720、VIR627、VIR628、VIR629)。

[1145] 表20:tRNA和gRNA复制子

[1146]

载体名称	载体描述	插入物名称	插入物序列
VIR621	SynV ΔGABE8-20-2a-GFP tRNA Pro-Hek2 gRNA	tRNA Pro 3'释放 粗体加下划线文本中的tRNA-pro 序列	gtacaagTAAGAAGTTGAATAACAAAATGCCGGAAATCTACGGATTGTGTATATCCATCATGAAAAAAACTAACACCCCTCCTTTCGAACCATCCCAAAC**ggctcgttggtctaggggtatgattctcgcttagggtgcgagaggtcccgggttcaaatcccggacgagccc**GGAACACAAAGCATAGACTGCgttttagagctaGAAAtagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgcttttCGAGGAAGGAGGTCTGAGGAGGTCACTGcgaaccagtttgtgtc**ggctcgttggtctaggggtatgattctcgcttagggtgcgagaggtcccgggttcaaatcccggacgagccc**tctagaagtgctgggtcatcta

[1147]

VIR622	SynV ΔGABE8-20-2a-GFP tRNA Ile-Hek2 gRNA	tRNA Ile 3'释放 粗体加下划线文本中的tRNA-ile	gtacaagTAAGAAGTTGAATAACAAAATGCCGGAAATCTACGGATTGTGTATATCCATCATGAAAAAAACTAACACCCCTCCTTTCGAACCATCCCAAAC**gctccagtggcgcaatcggttagcgcgcggtacttataagacagtgcacctgtgagcaatgccgaggttgtgagttcaagcctcacctggagca**GGAACACAAAGCATAGACTGCgttttagagctaGAAAtagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgcttCACACACACAAgctccagtggcgcaatcggttagcgcgcggtacttataagacagtgcaGCCgCGAGGAAGGAGGTCTGAGGAGGTCACTGcGGCcctgtgagcaatgccgaggttgtgagttcaagcctcacctggagcata
VIR623	SynV ΔGABE8-20-2a-GFP tRNA Ile-顶端 Hek2 gRNA	tRNA 顶端释放	gtacaagTAAGAAGTTGAATAACAAAATGCCGGAAATCTACGGATTGTGTATATCCATCATGAAAAAAACTAACACCCCTCCTTTCGAACCATCCCAAACgctccagtggcgcaatcggttagcgcgcggtacttataagacagtgcaGAACACAAAGCATAGACTGCgttttagagctaCCGAAAGGtagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgcttcacacacacacaCGAGGAAGGAGGTCTGAGGAGGTCACTGcgcctgtgagcaatgccgaggttgtgagttcaagcctcacctggagcata
VIR624	具有长接头的SynV ΔGABE8-20-2a-GFP tRNA Ile-	具有长接头的tRNA 顶端释放	gtacaagTAAGAAGTTGAATAACAAAATGCCGGAAATCTACGGATTGTGTATATCCATCATGAAAAAAACTAACACCCCTCCTTTCGAACCATCCCAAACgctccagtggcgcaatcggttagcgcgcggtacttataaga

[1148]

	顶端 Hek2 gRNAr		cagtgcagGAACACAAAGCA TAGACTGCgttttagagctaCCG AAAGGtagcaagttaaaaCaaggc tagtccgttatcaacttgaaaaagtggca ccgagtcggtgctttGGCCCGAG GAAGGAGGTCTGAGGAG GTCACTGGGCCAAAACA ACAACCCAACCAACAAA CCAACACCAAACAACAA ACCAAACCCCAACAAAC AACCACCAACCCAAACA Acctgtgagcaatgccgaggttgtgagt tcaagcctcacctggagcata
VIR625	具有长接头的 SynV ΔG ABE8-20 -2a-GFP tRNA Ile-顶端 Hek2 gRNA	经稳定化的 tRNA 顶端释放	gtacaagTAAGAAGTTGAAT AACAAAATGCCGGAAAT CTACGGATTGTGTATATC CATCATGAAAAAAACTA ACACCCCTCCTTTCGAAC CATCCCAAACgctccagtggcg caatcggttagcgcgcggtacttataaga cagtgcaGGAGCCCGAACAC AAAGCATAGACTGCgtttta gagctaGGCCCGAGGAAGG AGGTCTGAGGAGGTCAC TGGGCCtagcaagttaaaataaggc tagtccgttatcaacttgaaaaagtggca ccgagtcggtgcttAAAACAACA ACCCAACCAACAAACCA ACACCAAACAACAAACC AAACCCCAACAAACAAC CACCAACCCAAACAAGG GCTCCcctgtgagcaatgccgaggtt gtgagttcaagcctcacctggagcata
VIR626	SynV ΔG ABE8-20 -2a-GFP tRNA Ile 置换的 Hek2 gRNA	经置换的 tRNA Ile	gtacaagTAAGAAGTTGAAT AACAAAATGCCGGAAAT CTACGGATTGTGTATATC CATCATGAAAAAAACTA ACACCCCTCCTTTCGAAC CATCCCAAACGGGCTCCc ctgtgagcaatgccgaggttgtgagttca agcctcacctggagcaGAAAgctcc agtggcgcaatcggttagcgcgcggtac ttataagacagtgcaGGAGCCCG

[1149]

			AACACAAAGCATAGACT GCgtttagagctaGGCCCGAG GAAGGAGGTCTGAGGAG GTCACTGGGCCtagcaagttaa aataaggctagtccgttatcaacttgaaa aagtggcaccgagtcggtgcttAAA ACAACAACCCAACCAAC AAACCAACACCAAACAA CAAACCAAACCCCAACA AACAACCACCAACCCAA ACAAta
VIR627	SynRV tRNA-Pro-Thr IEDG Hek2 ΔG Abe820m-T2a-mScarlet	P-IEDG-T-Hek2 粗体加下划线文本中的tRNA-pro 序列 粗体斜体文本中的 tRNA-thr 序列	AACATCCCTCAAAagactca aggaaag**<u>ggctcgttggtctaggggta tgattctcgcttagggtgcgagaggtc ccgggttcaaatcccggacgagccc</u>**G cgtGtAgggTaaccatgaacGTTT TAGAGCTAGAAATAGCA AGTTAAAATAAGGCTAG TCCGTTATCAACTTGAAA AAGTGGCACCGAGTCGG TGCTTTTTTCACACACAC AA***ggctccatagctcaggggttagag cactggtcttgtaaaccaggggtcgcg agttcaattctcgctggggctt***GGAA CACAAAGCATAGACTGC gttttagagctaGCCgCGAGGAA GGAGGTCTGAGGAGGTC ACTGcGGCtagcaagttaaaataa ggctagtccgttatcaacttgaaaaagtg gcaccgagtcggtgcttttaaTTAAc cgagaaaaaaa**
VIR628	SynRV tRNA-Val-Lys IEDG Hek2 ΔG Abe820m-T2a-mScarlet	V-IEDG-K-Hek2	AACATCCCTCAAAagactca aggaaaggtttccgtagtgtagtggttatc acgttcgcctcacacgcgaaaggtcccc ggttcgaaaccgggcggaaacaGcgt GtAgggTaaccatgaacGTTTTA GAGCTAGAAATAGCAAG TTAAAATAAGGCTAGTC CGTTATCAACTTGAAAA AGTGGCACCGAGTCGGT GCTTTTTTCACACACACA Agcccggctagctcagtcggtagagca tgagactcttaatctcagggtcgtgggttc

[1150]

			gagccccacgttgggcgGGAACA CAAAGCATAGACTGCgtttt agagctaGCCgCGAGGAAGG AGGTCTGAGGAGGTCAC TGcGGCtagcaagttaaaataaggct agtccgttatcaacttgaaaaagtggcac cgagtcggtgctttttaaTTAAccgag aaaaaaa
VIR629	SynRV tRNA-Asp-Gly-Glu IEDG Hek2 ΔG Abe820m-T2a-mScarlet	D-IEDG-G-Hek2-Q 粗体斜体文本中的 tRNA-asp D15 粗体加下划线文本中的 tRNA-gly G8	AACATCCCTCAAAagactca aggaaag***tcctcgttagtatagtggtga gtatccccgcctgtcacgcgggagacc ggggttcgattccccgacggggag***Gc gtGtAgggTaaccatgaacGTTTT AGAGCTAGAAATAGCAA GTTAAAATAAGGCTAGT CCGTTATCAACTTGAAAA AGTGGCACCGAGTCGGT GCTTTTTTCACACACACA A**gcgttggtggtatagtggtgagcat agctgccttccaagcagttgacccggg ttcgattcccggccaacgca**GGAA CACAAAGCATAGACTGC gttttagagctaGCCgCGAGGAA GGAGGTCTGAGGAGGTC ACTGcGGCtagcaagttaaaataa ggctagtccgttatcaacttgaaaaagtg gcaccgagtcggtgctttCACACA CACAAtccttggtggtctagtggttag gattcggcgctctcaccgccgcggccc gggttcgattcccggtcagggaattaaT TAAccgagaaaaaaa
VIR712	SynRV tRNA-Ile-Ile(corn) IEDG PacI ΔG Abe820m-T2a-mScarlet	VIR622 插入物 粗体加下划线文本中的 tRNA-ile	AACATCCCTCAAAagactca aggaaag**gctccagtggcgcaatcgg ttagcgcgcggtacttataagacagtg cacctgtgagcaatgccgaggttgtga gttcaagcctcacctggagca**GcgtG tAgggTaaccatgaacGTTTTAG AGCTAGAAATAGCAAGT TAAAATAAGGCTAGTCC GTTATCAACTTGAAAAA GTGGCACCGAGTCGGTG CTTTTTTCACACACACAA

[1151]

			gctccagtggcgcaatcggttagcgcgc ggtacttataagacagtgcaGCCgC GAGGAAGGAGGTCTGAG GAGGTCACTGcGGCcctgtg agcaatgccgaggttgtgagttcaagcct cacctggagcaTTAATTAAtccg agaaaaaaa
VIR713	SynRV tRNA-Ile IEDG PacI ΔG Abe820m -T2a-mScarlet	5'Ile 至 PacI 粗体加下划线文本中的tRNA-ile	AACATCCCTCAAAagactca aggaaag**gctccagtggcgcaatcgg ttagcgcgcggtacttataagacagtg cacctgtgagcaatgccgaggttgtga gttcaagcctcacctggagca**GcgtG tAgggTaaccatgaacGTTTTAG AGCTAGAAATAGCAAGT TAAAATAAGGCTAGTCC GTTATCAACTTGAAAAA GTGGCACCGAGTCGGTG CTTTTTTaattaacgagaaaaaaa
VIR714	SynRV tRNA-Ile -Ile(corn) IEDG Hek2 ΔG Abe820m -T2a-mScarlet	I-IEDG-I-Hek2 粗体加下划线文本中的tRNA-ile	AACATCCCTCAAAagactca aggaaag**gctccagtggcgcaatcgg ttagcgcgcggtacttataagacagtg cacctgtgagcaatgccgaggttgtga gttcaagcctcacctggagca**GcgtG tAgggTaaccatgaacGTTTTAG AGCTAGAAATAGCAAGT TAAAATAAGGCTAGTCC GTTATCAACTTGAAAAA GTGGCACCGAGTCGGTG CTTTTTTCACACACACAA gctccagtggcgcaatcggttagcgcgc ggtacttataagacagtgcaGCCgC GAGGAAGGAGGTCTGAG GAGGTCACTGcGGCcctgtg agcaatgccgaggttgtgagttcaagcct cacctggagcaGGAACACAAA GCATAGACTGCgttttagagcta GCCgCGAGGAAGGAGGT CTGAGGAGGTCACTGcGG Ctagcaagttaaaataaggctagtccgtt atcaacttgaaaaagtggcaccgagtcg gtgctttttccgagaaaaaaa
VIR715	SynRV tRNA-Ile	I-IEDG-G-Hek2	AACATCCCTCAAAagactca aggaaag***gctccagtggcgcaatcggt***

[1152]

	-Gly IEDG Hek2 ΔG Abe820m -T2a-mScarlet	粗体斜体文本中的 tRNA-ile 粗体加下划线文本中的 tRNA-gly G8	***tagcgcgcggtacttataagacagtgcacctgtgagcaatgccgaggttgtgagttcaagcctcacctggagca***GcgtGtAgggTaaccatgaacGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTCACACACACAA**gcgttggtggtatagtggtgagcatagctgccttccaagcagttgacccgggttcgattcccggccaacgca**GGAACACAAAGCATAGACTGCgtttagagctaGCCgCGAGGAAGGAGGTCTGAGGAGGTCACTGcGGCtagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgctttttaaTTAAccgagaaaaaaa
VIR716	SynRV tRNA-Ile-Lys IEDG Hek2 ΔG Abe820m -T2a-mScarlet	I-IEDG-K-Hek2 粗体加下划线文本中的 tRNA-ile	AACATCCCTCAAAagactcaaggaaag**gctccagtggcgcaatcggttagcgcgcggtacttataagacagtgcacctgtgagcaatgccgaggttgtgagttcaagcctcacctggagca**GcgtGtAgggTaaccatgaacGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTCACACACACAAgcccggctagctcagtcggtagagcatgagactcttaatctcagggtcgtgggttcgagccccacgttgggcgGGAACACAAAGCATAGACTGCgtttagagctaGCCgCGAGGAAGGAGGTCTGAGGAGGTCACTGcGGCtagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgctttttaaTTAAccgagaaaaaaa
VIR717	SynRV tRNA-Ile-Leu	I-IEDG-L-Hek2 粗体加下划线	AACATCCCTCAAAagactcaaggaaag**gctccagtggcgcaatcggttagcgcgcggtacttataagacagtg**

[1153]

	IEDG Hek2 ΔG Abe820m -T2a-mSc arlet	文本中的 tRNA-ile	**cacctgtgagcaatgccgaggttgtga gttcaagcctcacctggagca**GcgtG tAgggTaaccatgaacGTTTTAG AGCTAGAAATAGCAAGT TAAAATAAGGCTAGTCC GTTATCAACTTGAAAAA GTGGCACCGAGTCGGTG CTTTTTTCACACACACAA ggtagcgtggccgagcggtctaaggcg ctggattaaggctccagtctcttcgggggg cgtgggttcgaatcccaccgctgccaG GAACACAAAGCATAGAC TGCgttttagagctaGCCgCGAG GAAGGAGGTCTGAGGAG GTCACTGcGGCtagcaagttaaa ataaggctagtccgttatcaacttgaaaa agtggcaccgagtcggtgcttttttaaTT AAccgagaaaaaaa		VIR718	SynRV tRNA-Ile -Pro IEDG Hek2 ΔG Abe820m -T2a-mSc arlet	I-IEDG-P-Hek2 粗体斜体文本中的 tRNA-ile 粗体加下划线文本中的 tRNA-pro 序列	AACATCCCTCAAAagactca aggaaag***gctccagtggcgcaatcggt tagcgcgcggtacttataagacagtgc acctgtgagcaatgccgaggttgtgagt tcaagcctcacctggagca***GcgtGtA gggTaaccatgaacGTTTTAGA GCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCG TTATCAACTTGAAAAAGT GGCACCGAGTCGGTGCT TTTTTCACACACACAA**ggc tcgttggtctaggggtatgattctcgctt agggtgcgagaggtcccgggttcaaa tcccggacgagccc**GGAACACA AAGCATAGACTGCgttttaga gctaGCCgCGAGGAAGGAG GTCTGAGGAGGTCACTGc GGCtagcaagttaaaataaggctagtc cgttatcaacttgaaaaagtggcaccga gtcggtgctttttaaTTAA
	VIR719	SynRV tRNA-Ile -Thr IEDG Hek2 ΔG	I-IEDG-T-Hek2 粗体加下划线	AACATCCCTCAAAagactca aggaaag**gctccagtggcgcaatcgg ttagcgcgcggtacttataagacagtg cacctgtgagcaatgccgaggttgtga gttcaagcctcacctggagca**GcgtG

	Abe820m-T2a-mScarlet	文本中的tRNA-ile 粗体斜体文本中的 tRNA-thr 序列	tAgggTaaccatgaacGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTCACACACACAA***ggctccatagctcaggggttagagcacttggtcttgtaaaccaggggtcgcgagttcaattctcgctggggctt***GGAACACAAAGCATAGACTGCgttttagagctaGCCgCGAGGAAGGAGGTCTGAGGAGGTCACTGcGGCtagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgcttttttaaTTAAccgagaaaaaaa
[1154]	VIR720	SynRV tRNA-Ile-Val IEDG Hek2 ΔG Abe820m-T2a-mScarlet	I-IEDG-V-Hek2 粗体加下划线文本中的 tRNA-ile	AACATCCCTCAAAagactcaaggaaag**<u>gctccagtggcgcaatcggttagcgcgcggtacttataagacagtgcacctgtgagcaatgccgaggttgtgagttcaagcctcacctggagca</u>**GcgtGtAgggTaaccatgaacGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTCACACACACAAgtttccgtagtgtagtggttatcacgttcgcctcacacgcgaaaggtccccggttcgaaaccgggcggaaacaGGAACACAAAGCATAGACTGCgtttttagagctaGCCgCGAGGAAGGAGGTCTGAGGAGGTCACTGcGGCtagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgcttttttaaTTAAccgagaaaaaaa

[1155] <u>DNA538序列：</u>

[1156]

DNA538	EFS-tRNA-Pro-HEK2	tRNA-Pro-HEK2 gRNA	**<u>ggctcgttggtctaggggtatgattctcgctt agggtgcgagaggtcccgggttcaaatccc ggacgagccc</u>**GAACACAAAGCAT

[1157]

	gRNA	粗体加下划线文本中的tRNA-pro序列	AGACTGCgtCttagagctaGGCCC GAGGAAGGAGGTCTGAGGA GGTCACTGGGCCtagcaagttaaGa taaggctagtccgttatcaacttgaaaaagtgg caccgagtcggtgcttaaccagtttgtgtc**ggc tcgttggtctaggggtatgattctcgcttagg gtgcgagaggtcccgggttcaaatcccgga cgagccc**

[1158] VIR622序列：

[1159]

VIR622	VIR622 SynV delG ABE8-20-2a-GFP tRNA Ile-Hek2 gRNA 粗体加下划线文本中的tRNA-ile	ABE8-20-2a-GFP tRNA Ile-Hek2 gRNA	ATGtccgaagtcgagttttcccatgagtactggatg agacacgcattgactctcgcaaagagggctcgagat gaacgcgaggtgcccgtgggggcagtactcgtgct caacaatcgcgtaatcggcgaaggttggaataggg caatcggactccacgaccccactgcacatgcggaa atcatggcccttcgacagggagggcttgtgatgcag aattatcgacttatcgatgcgacgctgtacgtcacgttt gaaccttgcgtaatgtgcgcgggagctatgattcact cccgcattggacgagttgtattcggtgttcgcaacgc caagacgggtgccgcaggttcactgatggacgtgct gcattacccaggcatgaaccaccgggtagaaatcac agaaggcatattggcggacgaatgtgcggcgctgtt gtgttacttttttcgcatgcccaggcgtgtctttaacgc ccagaaaaaagcacaatcctctactgactctggtggt tcttctggtggttctagcggcagcgagactcccggg acctcagagtccgccacacccgaaagttctggtggtt cttctggtggttctgacaagaagtacagcatcggcct ggccatcggcaccaactctgtgggctgggccgtgat caccgacgagtacaaggtgcccagcaagaaattca aggtgctgggcaacaccgaccggcacagcatcaa gaagaacctgatcggagccctgctgttcgacagcg gcgaaacagccgaggccacccggctgaagagaac cgccagaagaagatacaccagacggaagaaccgg atctgctatctgcaagagatcttcagcaacgagatgg ccaaggtggacgacagcttcttccacagactggaag agtccttcctggtggaagaggataagaagcacgag cggcaccccatcttcggcaacatcgtggacgaggtg gcctaccacgagaagtaccccaccatctaccacctg agaaagaaactggtggacagcaccgacaaggccg acctgcggctgatctatctggccctggcccacatgat caagttccggggccacttcctgatcgagggcgacct gaaccccgacaacagcgacgtggacaagctgttca

[1160]

			tccagctggtgcagacctacaaccagctgttcgagg aaaaccccatcaacgccagcggcgtggacgccaa ggccatcctgtctgccagactgagcaagagcagac ggctggaaaatctgatcgcccagctgcccggcgag aagaagaatggcctgttcggaaacctgattgccctg agcctgggcctgacccccaacttcaagagcaacttc gacctggccgaggatgccaaactgcagctgagcaa ggacacctacgacgacgacctggacaacctgctgg cccagatcggcgaccagtacgccgacctgtttctgg ccgccaagaacctgtccgacgccatcctgctgagc gacatcctgagagtgaacaccgagatcaccaaggc ccccctgagcgcctctatgatcaagagatacgacga gcaccaccaggacctgaccctgctgaaagctctcgt gcggcagcagctgcctgagaagtacaaagagatttt cttcgaccagagcaagaacggctacgccggctaca ttgacggcggagccagccaggaagagttctacaag ttcatcaagcccatcctggaaaagatggacggcacc gaggaactgctcgtgaagctgaacagagaggacct gctgcggaagcagcggaccttcgacaacggcagc atcccccaccagatccacctgggagagctgcacgc cattctgcggcggcaggaagatttttacccattcctga aggacaaccgggaaaagatcgagaagatcctgacc ttccgcatcccctactacgtgggccctctggccagg ggaaacagcagattcgcctggatgaccagaaagag cgaggaaaccatcacccctggaacttcgaggaag tggtggacaagggcgcttccgcccagagcttcatcg agcggatgaccaacttcgataagaacctgcccaacg agaaggtgctgcccaagcacagcctgctgtacgagt acttcaccgtgtataacgagctgaccaaagtgaaata cgtgaccgagggaatgagaaagcccgccttcctga gcggcgagcagaaaaaggccatcgtggacctgctg ttcaagaccaaccggaaagtgaccgtgaagcagct gaaagaggactacttcaagaaaatcgagtgcttcga ctccgtggaaatctccggcgtggaagatcggttcaa cgcctccctgggcacataccacgatctgctgaaaatt atcaaggacaaggacttcctggacaatgaggaaaa cgaggacattctggaagatatcgtgctgaccctgac actgtttgaggacagagagatgatcgaggaacggct gaaaacctatgcccacctgttcgacgacaaagtgat gaagcagctgaagcggcggagatacaccggctgg ggcaggctgagccggaagctgatcaacggcatccg ggacaagcagtccggcaagacaatcctggatttcct gaagtccgacggcttcgccaacagaaacttcatgca

[1161]

			gctgatccacgacgacagcctgacctttaaagagga catccagaaagcccaggtgtccggccagggcgata gcctgcacgagcacattgccaatctggccggcagc cccgccattaagaagggcatcctgcagacagtgaa ggtggtggacgagctcgtgaaagtgatgggccggc acaagcccgagaacatcgtgatcgaaatggccaga gagaaccagaccacccagaagggacagaagaaca gccgcgagagaatgaagcggatcgaagagggcat caaagagctgggcagccagatcctgaaagaacacc ccgtggaaaacacccagctgcagaacgagaagctg tacctgtactacctgcagaatgggcgggatatgtacg tggaccaggaactggacatcaaccggctgtccgact acgatgtggaccatatcgtgcctcagagctttctgaa ggacgactccatcgacaacaaggtgctgaccagaa gcgacaagaaccggggcaagagcgacaacgtgcc ctccgaagaggtcgtgaagaagatgaagaactactg gcggcagctgctgaacgccaagctgattacccaga gaaagttcgacaatctgaccaaggccgagagaggc ggcctgagcgaactggataaggccggcttcatcaa gagacagctggtggaaacccggcagatcacaaag cacgtggcacagatcctggactcccggatgaacact aagtacgacgagaatgacaagctgatccgggaagt gaaagtgatcaccctgaagtccaagctggtgtccga tttccggaaggatttccagttttacaaagtgcgcgaga tcaacaactaccaccacgcccacgacgcctacctga acgccgtcgtgggaaccgccctgatcaaaaagtac cctaagctggaaagcgagttcgtgtacggcgactac aaggtgtacgacgtgcggaagatgatcgccaagag cgagcaggaaatcggcaaggctaccgccaagtact tcttctacagcaacatcatgaacttttttcaagaccgag attaccctggccaacggcgagatccggaagcggcc tctgatcgagacaaacggcgaaaccggggagatcg tgtgggataagggccgggattttgccaccgtgcgga aagtgctgagcatgccccaagtgaatatcgtgaaaa agaccgaggtgcagacaggcggcttcagcaaaga gtctatcctgcccaagaggaacagcgataagctgat cgccagaaagaaggactgggaccctaagaagtac ggcggcttcgacagccccaccgtggcctattctgtg ctggtggtggccaaagtggaaaagggcaagtccaa gaaactgaagagtgtgaaagagctgctggggatca ccatcatggaaagaagcagcttcgagaagaatccca tcgactttctggaagccaagggctacaaagaagtga aaaaggacctgatcatcaagctgcctaagtactccct

[1162]

			gttcgagctggaaaacggccggaagagaatgctgg cctctgccggcgaactgcagaagggaaacgaactg gccctgccctccaaatatgtgaacttcctgtacctggc cagccactatgagaagctgaagggctccccgagg ataatgagcagaaacagctgtttgtggaacagcaca agcactacctggacgagatcatcgagcagatcagc gagttctccaagagagtgatcctggccgacgctaat ctggacaaagtgctgtccgcctacaacaagcaccg ggataagcccatcagagagcaggccgagaatatca tccacctgtttaccctgaccaatctgggagccctgc cgccttcaagtactttgacaccaccatcgaccggaa gaggtacaccagcaccaaagaggtgctggacgcc accctgatccaccagagcatcaccggcctgtacgag acacggatcgacctgtctcagctgggaggtgacga gggagctgataagcgcaccgccgatggttccgagtt cgaaagccccaagaagaagaggaaagtcGAAT TCGGCAGTGGAGAGGGCAGAGG gtccCTGCTAACATGCGGTGACGT CGAGGAGAATCCtGGcCCaatggtga gcaagggcgaggagctgttcaccggggtggtgccc atcctggtcgagctggacggcgacgtaaacggcca caagttcagcgtgtccggcgagggcgagggcgatg ccacctacggcaagctgaccctgaagttcatctgca ccaccggcaagctgcccgtgccctggcccaccctc gtgaccaccctgacctacggcgtgcagtgcttcagc cgctaccccgaccacatgaagcagcacgacttcttc aagtccgccatgcccgaaggctacgtccaggagcg caccatcttcttcaaggacgacggcaactacaagac ccgcgccgaggtgaagttcgagggcgacaccctg gtgaaccgcatcgagctgaagggcatcgacttcaag gaggacggcaacatcctggggcacaagctggagta caactacaacagccacaacgtctatatcatggccga caagcagaagaacggcatcaaggtgaacttcaaga tccgccacaacatcgaggacggcagcgtgcagctc gccgaccactaccagcagaacaccccatcggcga cggccccgtgctgctgcccgacaaccactacctga gcacccagtccgccctgagcaaagaccccaacga gaagcgcgatcacatggtcctgctggagttcgtgac cgccgccgggatcactctcggcatggacgagctgt acaagTAAGAAGTTGAATAACAAA ATGCCGGAAATCTACGGATTGT GTATATCCATCATGAAAAAAAC TAACACCCCTCCTTTCGAACCAT

[1163]

			CCCAAACgctccagtggcgcaatcggttagc gcgcggtacttataagacagtgcacctgtgagca atgccgaggttgtgagttcaagcctcacctggag caGGAACACAAAGCATAGACTGC gttttagagctaGAAAtagcaagttaaaataaggc tagtccgttatcaacttgaaaaagtggcaccgagtcg gtgcttCACACACACAAgctccagtggcg caatcggttagcgcgcggtacttataagacagtgca GCCgCGAGGAAGGAGGTCTGAG GAGGTCACTGcGGCcctgtgagcaatgc cgaggttgtgagttcaagcctcacctggagcata

[1164] tRNA-gRNA-tRNA盒(在VIR622中):

[1165] gctccagtggcgcaatcggttagcgcgcggtacttataagacagtgcacctgtgagcaatgccgaggttgtgagttcaagcctcacctggagcaGGAACACAAAGCATAGACTGCgttttagagctaGAAAtagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgcttCACACACACAAgctccagtggcgcaatcggttagcgcgcggtacttataagacagtgcaGCCgCGAGGAAGGAGGTCTGAGGAGGTCACTGcGGCcctgtgagcaatgccgaggttgtgagttcaagcctcacctggagca

[1166] 实施例5:使用tRNA和tRNA样分子进行初始gRNA释放筛选

[1167] 在初始筛选中测试了另外的tRNA和tRNA样分子,以确定相邻gRNA被加工并最终被用于介导碱基编辑的能力。对于实施例5中描述的每个实验,用编码碱基编辑器(ABE8.20)的载体和编码tRNA-gRNA盒的载体共转染293T细胞。每个tRNA-gRNA盒均在EFS启动子的控制下。具体而言,在转染前一天将1.3e4 293T细胞接种到96孔板的每个孔中。使用Lipofectamine 3000将50ng碱基编辑器载体和50ng gRNA载体共转染到每个孔中。转染后4天对样品进行测序以进行编辑。结果绘制为A>G编辑%。除非另外指出,否则如上所述,所用的gRNA靶向HEK位点。

[1168] 经侧接的gRNA与非经侧接的gRNA和最小Rnase P或Rnase Z底物:

[1169] 测试了经侧接的gRNA和非经侧接的gRNA之间的差异。经侧接的gRNA从5'到3'包含tRNA、gRNA和tRNA。例如,"tRNA-Pro"意指脯氨酸tRNA位于gRNA的5',而"tRNA-Pro-侧翼"意指脯氨酸tRNA位于gRNA的5'和3'。如图4A所示,无论gRNA是否被侧接,都会发生稳健的编辑。编辑通常等于或优于U6启动子驱动的无tRNA的gRNA的控制(U6::HEK2)。此外,还使用了多种类型的tRNA,每一种均允许gRNA介导稳健的碱基编辑。具体地,测试了tRNA-arg、tRNA-asp、tRNA-gly、tRNA-ile、tRNA-pro、tRNA-ser和tRNA-thr。

[1170] 除了上文所述的tRNA-gRNA盒外,还测试了Rnase P和Rnase Z的几种最小底物。所测试的最小底物是ATM5 ATSer和miniEGS,每个底物均由U6启动子驱动。各种最小底物进一步描述于Nashimoto等人(Biochemistry.38:12089-12096.1999;describing ATM5)和Kikovska等人(Nucleic Acids Research.33(6):2012-2021.2005;describing ATSer)中,其中的每一篇均通过引用并入本文。编码该最小底物的核酸序列列举如下:

[1171] GATCTGAATGGAGAGAGGGGGTTCAAATCCCCCTCTCTCCGC(ATSer;SEQ ID NO:4049);

[1172] GGGCCAGCCAGGTTCGACTCCTGGCTGGCTCGGTGTATTT(ATM5;SEQ ID NO:4050);

[1173] GGTGGGGCCAGCTCCTGAAGGTTCGAATCCTTCCCCCACC(miniEGS;SEQ ID NO:4051)。

[1174] 如图4A所示,几种最小底物可有效释放gRNA以介导碱基编辑。

[1175] tRNA样结构:

[1176] tRNA样结构是至少具有二级结构的RNA,该二级结构可以被加工(例如,切割)以释放与所述tRNA样结构连接的相邻gRNA。MALAT1相关的小细胞质RNA(mascRNA)是在细胞溶胶中发现的非编码RNA。它们由酶RNase P从较长的非编码RNA(称为MALAT1)加工而成。为了测试mascRNA递送表达的gRNA以进行碱基编辑的能力,测试了来自数个不同物种的各种mascRNA。如图4B所示,mascRNA-gRNA盒的碱基编辑虽然低,但高于背景。

[1177] tRNA变体:

[1178] 在与上述类似的tRNA-gRNA盒中测试了tRNA变体。具体来说,测试了几种tRNA-pro和tRNA-thr变体,并将其与表达gRNA或U6驱动的无tRNA的gRNA的稳定细胞系进行比较。如图4C所示,tRNA-pro和tRNA-thr变体可有效介导稳健的碱基编辑。

[1179] tRNA片段和其他RnaseZ或RnaseP底物:

[1180] 在与上述类似的tRNA-gRNA盒中测试了tRNA片段和其他RnaseZ或RnaseP底物。对于片段,tRNA被分裂成两个半部,同时维持Rnase加工位点并连接到gRNA。作为替代方案,通过在其间插入gRNA来分裂tRNA。如图4D所示,所测试的tRNA片段-gRNA盒的碱基编辑虽然低,但高于背景。

[1181] 病毒tRNA样结构(vtRNA):

[1182] 本实验中使用的vtRNA源自γ-疱疹病毒(GHV68)。这些vtRNA由病毒基因组表达,并且由很像内源tRNA的细胞机器加工。vtRNA更详细地描述于Bowden等人(J.Gen Virol.78:1675-1687.1997),其通过引用并入本文。每个gRNA表达盒均构建如下:从5'到3',EFS(Pol II启动子)-狂犬病转录起始序列-tRNA-gRNA-poly A。单独的驱动gRNA的EFS启动子通常不会导致编辑(EFS控制),而在tRNA存在的情况下,就会发生编辑。如图5所示,所有所测试的vtRNA(vt_1至vt_8)在三个不同的目标位点(HEK2、SOD1和ALAS1)处都产生可检测到的碱基编辑。本实验中使用了之前测试过的另外的非病毒tRNA。P对应于tRNA-pro,T对应于tRNA-thr,G8对应于tRNA-gly,G27对应于不同的tRNA-gly,L对应于tRNA-leu,并且D15对应于tRNA-Asp。每种非病毒tRNA也表现出稳健的碱基编辑。

[1183] 使用的SOD1和ALAS1 gRNA间隔区序列如下:

[1184] SOD1:UAAAUAGGCUGUACCAGUGC(SEQ ID NO:4052)

[1185] ALAS1:CAGGAUCCGCACAGACUCCA(SEQ ID NO:4053)

[1186] 实施例6:各种RABV基因组架构中的tRNA-gRNA盒

[1187] 接下来将几个tRNA-gRNA盒插入到不同的RABV基因组架构中以测试碱基编辑。如图6A所示,tRNA-gRNA盒被放置在具有共表达核碱基编辑器的ΔG、ΔGL和ΔMGL RABV基因组的几个位置中。使用以下狂犬病病毒复制子:

[1188]

构建体	复制子类型	靶标	位置
VIR1001	ΔG	ALAS1	M 后
VIR1002	ΔG	ALAS1	P 后
VIR1003	ΔG	ALAS1	N 后
VIR1004	ΔGL	ALAS1	M 后
VIR1005	ΔGL	ALAS1	P 后
VIR1006	ΔGL	ALAS1	N 后
VIR1007	ΔMGL	ALAS1	P 后
VIR1008	ΔMGL	ALAS1	N 后
VIR1017	ΔG	SOD1	M 后
VIR1018	ΔG	SOD1	P 后
VIR1019	ΔG	SOD1	N 后
VIR1020	ΔGL	SOD1	M 后
VIR1021	ΔGL	SOD1	P 后
VIR1022	ΔGL	SOD1	N 后
VIR1023	ΔMGL	SOD1	P 后
VIR1024	ΔMGL	SOD1	N 后

[1189] 将复制子转染至狂犬病生产细胞中并收集病毒上清液。在感染后4天时收获来自生产细胞的基因组DNA,并在指示的基因座(SOD1或ALAS1)处进行序列编辑。如图6B所示,在所有所测试的RABV基因组架构中都检测到碱基编辑,这证明了tRNA-gRNA盒在负链RNA病毒(例如狂犬病)中递送gRNA的有效性。

[1190] 其它实施方案

[1191] 根据前述描述,将显而易知的是可以对本文所述的发明作出变化和修改以使其适于各种用途和条件。此类实施方案也在所附权利要求书的范围内。

[1192] 本文变量的任何定义中的要素清单的列举包括将那个变量定义为任何单一要素或所列要素的组合(或亚组合)。本文中实施方案的列举包括那个实施方案呈任何单一实施方案形式或与任何其它实施方案或其部分组合。

[1193] 本说明书中提及的所有专利和公布以引用方式并入本文,引用程度如同明确地且单独地指示每篇独立专利和公布以引用方式并入一样。

[1194] L.序列表

[1195]

描述	SEQ ID NO:	序列
腺苷脱氨酶参考序列	8	MSEVEFSHEYWMRHALTLAKRARDEREV PVGAVLVLNNRVIGEGWNRAIGLHDPTAH AEIMALRQGGLVMQNYRLIDATLYVTFEP CVMCAGAMIHSRIGRVVFGVRNAKTGAA GSLMDVLHYPGMNHRVEITEGILADECAA LLCYFFRMPRQVFNAQKKAQSSTD
位于P153多核苷酸处的BhCas12b GGSGGS	274	GCCACCATGGCCCCAAAGAAGAAGCGG AAGGTCGGTATCCACGGAGTCCCAGCAG CCGCCACCAGATCCTTCATCCTGAAGAT CGAGCCCAACGAGGAAGTGAAGAAAGG CCTCTGGAAAACCCACGAGGTGCTGAAC

[1196]

-ABE8-Xten20		CACGGAATCGCCTACTACATGAATATCC TGAAGCTGATCCGGCAAGAGGCCATCTA CGAGCACCACGAGCAGGACCCCAAGAA TCCCAAGAAGGTGTCCAAGGCCGAGATC CAGGCCGAGCTGTGGGATTTCGTGCTGA AGATGCAGAAGTGCAACAGCTTCACACA CGAGGTGGACAAGGACGAGGTGTTCAA CATCCTGAGAGAGCTGTACGAGGAACTG GTGCCCAGCAGCGTGGAAAAGAAGGGC GAAGCCAACCAGCTGAGCAACAAGTTTC TGTACCCTCTGGTGGACCCCAACAGCCA GTCTGGAAAGGGAACAGCCAGCAGCGG CAGAAAGCCCAGATGGTACAACCTGAA GATTGCCGGCGATCCCGGAGGCTCTGGA GGAAGCTCCGAAGTCGAGTTTTCCCATG AGTACTGGATGAGACACGCATTGACTCT CGCAAAGAGGGCTCGAGATGAACGCGA GGTGCCCGTGGGGGCAGTACTCGTGCTC AACAATCGCGTAATCGGCGAAGGTTGGA ATAGGGCAATCGGACTCCACGACCCCAC TGCACATGCGGAAATCATGGCCCTTCGA CAGGGAGGGCTTGTGATGCAGAATTATC GACTTTATGATGCGACGCTGTACGTCAC GTTTGAACCTTGCGTAATGTGCGCGGGA GCTATGATTCACTCCCGCATTGGACGAG TTGTATTCGGTGTTCGCAACGCCAAGAC GGGTGCCGCAGGTTCACTGATGGACGTG CTGCATCATCCAGGCATGAACCACCGGG TAGAAATCACAGAAGGCATATTGGCGGA CGAATGTGCGGCGCTGTTGTGTCGTTTTT TTCGCATGCCCAGGCGGGTCTTTAACGC CCAGAAAAAAGCACAATCCTCTACTGAC GGCTCTTCTGGATCTGAAACACCTGGCA CAAGCGAGAGCGCCACCCCTGAGAGCTC TGGCTCCTGGGAAGAAGAGAAGAAGAA GTGGGAAGAAGATAAGAAAAAGGACCC GCTGGCCAAGATCCTGGGCAAGCTGGCT GAGTACGGACTGATCCCTCTGTTCATCC CCTACACCGACAGCAACGAGCCCATCGT GAAAGAAATCAAGTGGATGGAAAAGTC CCGGAACCAGAGCGTGCGGCGGCTGGAT AAGGACATGTTCATTCAGGCCCTGGAAC GGTTCCTGAGCTGGGAGAGCTGGAACCT

[1197]

		GAAAGTGAAAGAGGAATACGAGAAGGT CGAGAAAGAGTACAAGACCCTGGAAGA GAGGATCAAAGAGGACATCCAGGCTCTG AAGGCTCTGGAACAGTATGAGAAAGAG CGGCAAGAACAGCTGCTGCGGGACACCC TGAACACCAACGAGTACCGGCTGAGCAA GAGAGGCCTTAGAGGCTGGCGGGAAAT CATCCAGAAATGGCTGAAAATGGACGA GAACGAGCCCTCCGAGAAGTACCTGGAA GTGTTCAAGGACTACCAGCGGAAGCACC CTAGAGAGGCCGGCGATTACAGCGTGTA CGAGTTCCTGTCCAAGAAAGAGAACCAC TTCATCTGGCGGAATCACCCTGAGTACC CCTACCTGTACGCCACCTTCTGCGAGAT CGACAAGAAAAAGAAGGACGCCAAGCA GCAGGCCACCTTCACACTGGCCGATCCT ATCAATCACCCTCTGTGGGTCCGATTCG AGGAAAGAAGCGGCAGCAACCTGAACA AGTACAGAATCCTGACCGAGCAGCTGCA CACCGAGAAGCTGAAGAAAAAGCTGAC AGTGCAGCTGGACCGGCTGATCTACCCT ACAGAATCTGGCGGCTGGGAAGAGAAG GGCAAAGTGGACATTGTGCTGCTGCCCA GCCGGCAGTTCTACAACCAGATCTTCCT GGACATCGAGGAAAAGGGCAAGCACGC CTTCACCTACAAGGATGAGAGCATCAAG TTCCCTCTGAAGGGCACACTCGGCGGAG CCAGAGTGCAGTTCGACAGAGATCACCT GAGAAGATACCCTCACAAGGTGGAAAG CGGCAACGTGGGCAGAATCTACTTCAAC ATGACCGTGAACATCGAGCCTACAGAGT CCCCAGTGTCCAAGTCTCTGAAGATCCA CCGGGACGACTTCCCCAAGGTGGTCAAC TTCAAGCCCAAAGAACTGACCGAGTGGA TCAAGGACAGCAAGGGCAAGAAACTGA AGTCCGGCATCGAGTCCCTGGAAATCGG CCTGAGAGTGATGAGCATCGACCTGGGA CAGAGACAGGCCGCTGCCGCCTCTATTT TCGAGGTGGTGGATCAGAAGCCCGACAT CGAAGGCAAGCTGTTTTTCCCAATCAAG GGCACCGAGCTGTATGCCGTGCACAGAG CCAGCTTCAACATCAAGCTGCCCGGCGA GACACTGGTCAAGAGCAGAGAAGTGCT

[1198]

		GCGGAAGGCCAGAGAGGACAATCTGAA ACTGATGAACCAGAAGCTCAACTTCCTG CGGAACGTGCTGCACTTCCAGCAGTTCG AGGACATCACCGAGAGAGAGAAGCGGG TCACCAAGTGGATCAGCAGACAAGAGA ACAGCGACGTGCCCCTGGTGTACCAGGA TGAGCTGATCCAGATCCGCGAGCTGATG TACAAGCCTTACAAGGACTGGGTCGCCT TCCTGAAGCAGCTCCACAAGAGACTGGA AGTCGAGATCGGCAAAGAAGTGAAGCA CTGGCGGAAGTCCCTGAGCGACGGAAG AAAGGGCCTGTACGGCATCTCCCTGAAG AACATCGACGAGATCGATCGGACCCGGA AGTTCCTGCTGAGATGGTCCCTGAGGCC TACCGAACCTGGCGAAGTGCGTAGACTG GAACCCGGCCAGAGATTCGCCATCGACC AGCTGAATCACCTGAACGCCCTGAAAGA AGATCGGCTGAAGAAGATGGCCAACAC CATCATCATGCACGCCCTGGGCTACTGC TACGACGTGCGGAAGAAGAAATGGCAG GCTAAGAACCCCGCCTGCCAGATCATCC TGTTCGAGGATCTGAGCAACTACAACCC CTACGAGGAAAGGTCCCGCTTCGAGAAC AGCAAGCTCATGAAGTGGTCCAGACGCG AGATCCCCAGACAGGTTGCACTGCAGGG CGAGATCTATGGCCTGCAAGTGGGAGAA GTGGGCGCTCAGTTCAGCAGCAGATTCC ACGCCAAGACAGGCAGCCCTGGCATCAG ATGTAGCGTCGTGACCAAAGAGAAGCTG CAGGACAATCGGTTCTTCAAGAATCTGC AGAGAGAGGGCAGACTGACCCTGGACA AAATCGCCGTGCTGAAAGAGGGCGATCT GTACCCAGACAAAGGCGGCGAGAAGTT CATCAGCCTGAGCAAGGATCGGAAGTGC GTGACCACACACGCCGACATCAACGCCG CTCAGAACCTGCAGAAGCGGTTCTGGAC AAGAACCCACGGCTTCTACAAGGTGTAC TGCAAGGCCTACCAGGTGGACGGCCAGA CCGTGTACATCCCTGAGAGCAAGGACCA GAAGCAGAAGATCATCGAAGAGTTCGG CGAGGGCTACTTCATTCTGAAGGACGGG GTGTACGAATGGGTCAACGCCGGCAAGC TGAAAATCAAGAAGGGCAGCTCCAAGC

[1199]

		AGAGCAGCAGCGAGCTGGTGGATAGCG ACATCCTGAAAGACAGCTTCGACCTGGC CTCCGAGCTGAAAGGCGAAAAGCTGATG CTGTACAGGGACCCCAGCGGCAATGTGT TCCCCAGCGACAAATGGATGGCCGCTGG CGTGTTCTTCGGAAAGCTGGAACGCATC CTGATCAGCAAGCTGACCAACCAGTACT CCATCAGCACCATCGAGGACGACAGCAG CAAGCAGTCTATGAAAAGGCCGGCGGCC ACGAAAAAGGCCGGCCAGGCAAAAAAG AAAAAGGGATCCTACCCATACGATGTTC CAGATTACGCTTATCCCTACGACGTGCC TGATTATGCATACCCATATGATGTCCCC GACTATGCCTAA
位于P153多肽处的BhCas12bGGSGGS-ABE8-Xten20	275	MAPKKKRKVGIHGVPAAATRSFILKIEPNE EVKKGLWKTHEVLNHGIAYYMNILKLIRQ EAIYEHHEQDPKNPKKVSKAEIQAELWDF VLKMQKCNSFTHEVDKDEVFNILRELYEE LVPSSVEKKGEANQLSNKFLYPLVDPNSQ SGKGTASSGRKPRWYNLKIAGDPGGSGGS SEVEFSHEYWMRHALTLAKRARDEREVP VGAVLVLNNRVIGEGWNRAIGLHDPTAH AEIMALRQGGLVMQNYRLYDATLYVTFE PCVMCAGAMIHSRIGRVVFGVRNAKTGA AGSLMDVLHHPGMNHRVEITEGILADECA ALLCRFFRMPRRVFNAQKKAQSSTDGSSG SETPGTSESATPESSGSWEEEKKKWEEDK KKDPLAKILGKLAEYGLIPLFIPYTDSNEPI VKEIKWMEKSRNQSVRRLDKDMFIQALER FLSWESWNLKVKEEYEKVEKEYKTLEERI KEDIQALKALEQYEKERQEQLLRDTLNTN EYRLSKRGLRGWREIIQKWLKMDENEPSE KYLEVFKDYQRKHPREAGDYSVYEFLSKK ENHFIWRNHPEYPYLYATFCEIDKKKKDA KQQATFTLADPINHPLWVRFEERSGSNLN KYRILTEQLHTEKLKKKLTVQLDRLIYPTE SGGWEEKGKVDIVLLPSRQFYNQIFLDIEE KGKHAFTYKDESIKFPLKGTLGGARVQFD RDHLRRYPHKVESGNVGRIYFNMTVNIEP TESPVSKSLKIHRDDFPKVVNFKPKELTEW IKDSKGKKLKSGIESLEIGLRVMSIDLGQR QAAAASIFEVVDQKPDIEGKLFFPIKGTEL YAVHRASFNIKLPGETLVKSREVLRKARE

[1200]

		DNLKLMNQKLNFLRNVLHFQQFEDITERE KRVTKWISRQENSDVPLVYQDELIQIRELM YKPYKDWVAFLKQLHKRLEVEIGKEVKH WRKSLSDGRKGLYGISLKNIDEIDRTRKFL LRWSLRPTEPGEVRRLEPGQRFAIDQLNHL NALKEDRLKKMANTIIMHALGYCYDVRK KKWQAKNPACQIILFEDLSNYNPYEERSRF ENSKLMKWSRREIPRQVALQGEIYGLQVG EVGAQFSSRFHAKTGSPGIRCSVVTKEKLQ DNRFFKNLQREGRLTLDKIAVLKEGDLYP DKGGEKFISLSKDRKCVTTHADINAAQNL QKRFWTRTHGFYKVYCKAYQVDGQTVYI PESKDQKQKIIEEFGEGYFILKDGVYEWVN AGKLKIKKGSSKQSSSELVDSDILKDSFDL ASELKGEKLMLYRDPSGNVFPSDKWMAA GVFFGKLERILISKLTNQYSISTIEDDSSKQS MKRPAATKKAGQAKKKKGSYPYDVPDY AYPYDVPDYAYPYDVPDYA
位于K255多核苷酸处的BhCas12bGGSGGS-ABE8-Xten20	276	GCCACCATGGCCCCAAAGAAGAAGCGG AAGGTCGGTATCCACGGAGTCCCAGCAG CCGCCACCAGATCCTTCATCCTGAAGAT CGAGCCCAACGAGGAAGTGAAGAAAGG CCTCTGGAAAACCCACGAGGTGCTGAAC CACGGAATCGCCTACTACATGAATATCC TGAAGCTGATCCGGCAAGAGGCCATCTA CGAGCACCACGAGCAGGACCCCAAGAA TCCCAAGAAGGTGTCCAAGGCCGAGATC CAGGCCGAGCTGTGGGATTTCGTGCTGA AGATGCAGAAGTGCAACAGCTTCACACA CGAGGTGGACAAGGACGAGGTGTTCAA CATCCTGAGAGAGCTGTACGAGGAACTG GTGCCCAGCAGCGTGGAAAAGAAGGGC GAAGCCAACCAGCTGAGCAACAAGTTTC TGTACCCTCTGGTGGACCCCAACAGCCA GTCTGGAAAGGGAACAGCCAGCAGCGG CAGAAAGCCCAGATGGTACAACCTGAA GATTGCCGGCGATCCCTCCTGGGAAGAA GAGAAGAAGAAGTGGGAAGAAGATAAG AAAAAGGACCCGCTGGCCAAGATCCTGG GCAAGCTGGCTGAGTACGGACTGATCCC TCTGTTCATCCCCTACACCGACAGCAAC GAGCCCATCGTGAAAGAAATCAAGTGG ATGGAAAAGTCCCGGAACCAGAGCGTG

[1201]

		CGGCGGCTGGATAAGGACATGTTCATTC AGGCCCTGGAACGGTTCCTGAGCTGGGA GAGCTGGAACCTGAAAGTGAAAGAGGA ATACGAGAAGGTCGAGAAAGAGTACAA GACCCTGGAAGAGAGGATCAAAGGAGG CTCTGGAGGAAGCTCCGAAGTCGAGTTT TCCCATGAGTACTGGATGAGACACGCAT TGACTCTCGCAAAGAGGGCTCGAGATGA ACGCGAGGTGCCCGTGGGGGCAGTACTC GTGCTCAACAATCGCGTAATCGGCGAAG GTTGGAATAGGGCAATCGGACTCCACGA CCCCACTGCACATGCGGAAATCATGGCC CTTCGACAGGGAGGGCTTGTGATGCAGA ATTATCGACTTTATGATGCGACGCTGTA CGTCACGTTTGAACCTTGCGTAATGTGC GCGGGAGCTATGATTCACTCCCGCATTG GACGAGTTGTATTCGGTGTTCGCAACGC CAAGACGGGTGCCGCAGGTTCACTGATG GACGTGCTGCATCATCCAGGCATGAACC ACCGGGTAGAAATCACAGAAGGCATATT GGCGGACGAATGTGCGGCGCTGTTGTGT CGTTTTTTTCGCATGCCCAGGCGGGTCTT TAACGCCCAGAAAAAAGCACAATCCTCT ACTGACGGCTCTTCTGGATCTGAAACAC CTGGCACAAGCGAGAGCGCCACCCCTGA GAGCTCTGGCGAGGACATCCAGGCTCTG AAGGCTCTGGAACAGTATGAGAAAGAG CGGCAAGAACAGCTGCTGCGGGACACCC TGAACACCAACGAGTACCGGCTGAGCAA GAGAGGCCTTAGAGGCTGGCGGGAAAT CATCCAGAAATGGCTGAAAATGGACGA GAACGAGCCCTCCGAGAAGTACCTGGAA GTGTTCAAGGACTACCAGCGGAAGCACC CTAGAGAGGCCGGCGATTACAGCGTGTA CGAGTTCCTGTCCAAGAAAGAGAACCAC TTCATCTGGCGGAATCACCCTGAGTACC CCTACCTGTACGCCACCTTCTGCGAGAT CGACAAGAAAAAGAAGGACGCCAAGCA GCAGGCCACCTTCACACTGGCCGATCCT ATCAATCACCCTCTGTGGGTCCGATTCG AGGAAAGAAGCGGCAGCAACCTGAACA AGTACAGAATCCTGACCGAGCAGCTGCA CACCGAGAAGCTGAAGAAAAAGCTGAC

[1202]

		AGTGCAGCTGGACCGGCTGATCTACCCT ACAGAATCTGGCGGCTGGGAAGAGAAG GGCAAAGTGGACATTGTGCTGCTGCCCA GCCGGCAGTTCTACAACCAGATCTTCCT GGACATCGAGGAAAAGGGCAAGCACGC CTTCACCTACAAGGATGAGAGCATCAAG TTCCCTCTGAAGGGCACACTCGGCGGAG CCAGAGTGCAGTTCGACAGAGATCACCT GAGAAGATACCCTCACAAGGTGGAAAG CGGCAACGTGGGCAGAATCTACTTCAAC ATGACCGTGAACATCGAGCCTACAGAGT CCCCAGTGTCCAAGTCTCTGAAGATCCA CCGGGACGACTTCCCCAAGGTGGTCAAC TTCAAGCCCAAAGAACTGACCGAGTGGA TCAAGGACAGCAAGGGCAAGAAACTGA AGTCCGGCATCGAGTCCCTGGAAATCGG CCTGAGAGTGATGAGCATCGACCTGGGA CAGAGACAGGCCGCTGCCGCCTCTATTT TCGAGGTGGTGGATCAGAAGCCCGACAT CGAAGGCAAGCTGTTTTTCCCAATCAAG GGCACCGAGCTGTATGCCGTGCACAGAG CCAGCTTCAACATCAAGCTGCCCGGCGA GACACTGGTCAAGAGCAGAGAAGTGCT GCGGAAGGCCAGAGAGGACAATCTGAA ACTGATGAACCAGAAGCTCAACTTCCTG CGGAACGTGCTGCACTTCCAGCAGTTCG AGGACATCACCGAGAGAGAGAAGCGGG TCACCAAGTGGATCAGCAGACAAGAGA ACAGCGACGTGCCCCTGGTGTACCAGGA TGAGCTGATCCAGATCCGCGAGCTGATG TACAAGCCTTACAAGGACTGGGTCGCCT TCCTGAAGCAGCTCCACAAGAGACTGGA AGTCGAGATCGGCAAAGAAGTGAAGCA CTGGCGGAAGTCCCTGAGCGACGGAAG AAAGGGCCTGTACGGCATCTCCCTGAAG AACATCGACGAGATCGATCGGACCCGGA AGTTCCTGCTGAGATGGTCCCTGAGGCC TACCGAACCTGGCGAAGTGCGTAGACTG GAACCCGGCCAGAGATTCGCCATCGACC AGCTGAATCACCTGAACGCCCTGAAAGA AGATCGGCTGAAGAAGATGGCCAACAC CATCATCATGCACGCCCTGGGCTACTGC TACGACGTGCGGAAGAAGAAATGGCAG

[1203]

		GCTAAGAACCCCGCCTGCCAGATCATCC TGTTCGAGGATCTGAGCAACTACAACCC CTACGAGGAAAGGTCCCGCTTCGAGAAC AGCAAGCTCATGAAGTGGTCCAGACGCG AGATCCCCAGACAGGTTGCACTGCAGGG CGAGATCTATGGCCTGCAAGTGGGAGAA GTGGGCGCTCAGTTCAGCAGCAGATTCC ACGCCAAGACAGGCAGCCCTGGCATCAG ATGTAGCGTCGTGACCAAAGAGAAGCTG CAGGACAATCGGTTCTTCAAGAATCTGC AGAGAGAGGGCAGACTGACCCTGGACA AAATCGCCGTGCTGAAAGAGGGCGATCT GTACCCAGACAAAGGCGGCGAGAAGTT CATCAGCCTGAGCAAGGATCGGAAGTGC GTGACCACACACGCCGACATCAACGCCG CTCAGAACCTGCAGAAGCGGTTCTGGAC AAGAACCCACGGCTTCTACAAGGTGTAC TGCAAGGCCTACCAGGTGGACGGCCAGA CCGTGTACATCCCTGAGAGCAAGGACCA GAAGCAGAAGATCATCGAAGAGTTCGG CGAGGGCTACTTCATTCTGAAGGACGGG GTGTACGAATGGGTCAACGCCGGCAAGC TGAAAATCAAGAAGGGCAGCTCCAAGC AGAGCAGCAGCGAGCTGGTGGATAGCG ACATCCTGAAAGACAGCTTCGACCTGGC CTCCGAGCTGAAAGGCGAAAAGCTGATG CTGTACAGGGACCCCAGCGGCAATGTGT TCCCCAGCGACAAATGGATGGCCGCTGG CGTGTTCTTCGGAAAGCTGGAACGCATC CTGATCAGCAAGCTGACCAACCAGTACT CCATCAGCACCATCGAGGACGACAGCAG CAAGCAGTCTATGAAAAGGCCGGCGGCC ACGAAAAAGGCCGGCCAGGCAAAAAAG AAAAAGGGATCCTACCCATACGATGTTC CAGATTACGCTTATCCCTACGACGTGCC TGATTATGCATACCCATATGATGTCCCC GACTATGCCTAA
位于K255多肽处的BhCas12b GGSGGS	277	MAPKKKRKVGIHGVPAAATRSFILKIEPNE EVKKGLWKTHEVLNHGIAYYMNILKLIRQ EAIYEHHEQDPKNPKKVSKAEIQAELWDF VLKMQKCNSFTHEVDKDEVFNILRELYEE LVPSSVEKKGEANQLSNKFLYPLVDPNSQ SGKGTASSGRKPRWYNLKIAGDPSWEEEK

[1204]

-ABE8-Xten20		KKWEEDKKKDPLAKILGKLAEYGLIPLFIP YTDSNEPIVKEIKWMEKSRNQSVRRLDKD MFIQALERFLSWESWNLKVKEEYEKVEKE YKTLEERIKGGSGGGSSEVEFSHEYWMRHA LTLAKRARDEREVPVGAVLVLNNRVIGEG WNRAIGLHDPTAHAEIMALRQGGLVMQN YRLYDATLYVTFEPCVMCAGAMIHSRIGR VVFGVRNAKTGAAGSLMDVLHHPGMNH RVEITEGILADECAALLCRFFRMPRRVFNA QKKAQSSTDGSSGSETPGTSESATPESSGE DIQALKALEQYEKERQEQLLRDTLNTNEY RLSKRGLRGWREIIQKWLKMDENEPSEKY LEVFKDYQRKHPREAGDYSVYEFLSKKEN HFIWRNHPEYPYLYATFCEIDKKKKDAKQ QATFTLADPINHPLWVRFEERSGSNLNKY RILTEQLHTEKLKKKLTVQLDRLIYPTESG GWEEKGKVDIVLLPSRQFYNQIFLDIEEKG KHAFTYKDESIKFPLKGTLGGARVQFDRD HLRRYPHKVESGNVGRIYFNMTVNIEPTES PVSKSLKIHRDDFPKVVNFKPKELTEWIKD SKGKKLKSGIESLEIGLRVMSIDLGQRQAA AASIFEVVDQKPDIEGKLFFPIKGTELYAV HRASFNIKLPGETLVKSREVLRKAREDNL KLMNQKLNFLRNVLHFQQFEDITEREKRV TKWISRQENSDVPLVYQDELIQIRELMYKP YKDWVAFLKQLHKRLEVEIGKEVKHWRK SLSDGRKGLYGISLKNIDEIDRTRKFLLRW SLRPTEPGEVRRLEPGQRFAIDQLNHLNAL KEDRLKKMANTIIMHALGYCYDVRKKKW QAKNPACQIILFEDLSNYNPYEERSRFENS KLMKWSRREIPRQVALQGEIYGLQVGEVG AQFSSRFHAKTGSPGIRCSVVTKEKLQDNR FFKNLQREGRLTLDKIAVLKEGDLYPDKG GEKFISLSKDRKCVTTHADINAAQNLQKR FWTRTHGFYKVYCKAYQVDGQTVYIPES KDQKQKIIEEFGEGYFILKDGVYEWVNAG KLKIKKGSSKQSSSELVDSDILKDSFDLASE LKGEKLMLYRDPSGNVFPSDKWMAAGVF FGKLERILISKLTNQYSISTIEDDSSKQSMK RPAATKKAGQAKKKKGSYPYDVPDYAYP YDVPDYAYPYDVPDYA
位于D306多	278	GCCACCATGGCCCCAAAGAAGAAGCGG AAGGTCGGTATCCACGGAGTCCCAGCAG

[1205]

核苷酸处的 BhCas12b GGSGGS -ABE8-Xten20		CCGCCACCAGATCCTTCATCCTGAAGAT CGAGCCCAACGAGGAAGTGAAGAAAGG CCTCTGGAAAACCCACGAGGTGCTGAAC CACGGAATCGCCTACTACATGAATATCC TGAAGCTGATCCGGCAAGAGGCCATCTA CGAGCACCACGAGCAGGACCCCAAGAA TCCCAAGAAGGTGTCCAAGGCCGAGATC CAGGCCGAGCTGTGGGATTTCGTGCTGA AGATGCAGAAGTGCAACAGCTTCACACA CGAGGTGGACAAGGACGAGGTGTTCAA CATCCTGAGAGAGCTGTACGAGGAACTG GTGCCCAGCAGCGTGGAAAAGAAGGGC GAAGCCAACCAGCTGAGCAACAAGTTTC TGTACCCTCTGGTGGACCCCAACAGCCA GTCTGGAAAGGGAACAGCCAGCAGCGG CAGAAAGCCCAGATGGTACAACCTGAA GATTGCCGGCGATCCCTCCTGGGAAGAA GAGAAGAAGAAGTGGGAAGAAGATAAG AAAAAGGACCCGCTGGCCAAGATCCTGG GCAAGCTGGCTGAGTACGGACTGATCCC TCTGTTCATCCCCTACACCGACAGCAAC GAGCCCATCGTGAAAGAAATCAAGTGG ATGGAAAAGTCCCGGAACCAGAGCGTG CGGCGGCTGGATAAGGACATGTTCATTC AGGCCCTGGAACGGTTCCTGAGCTGGGA GAGCTGGAACCTGAAAGTGAAAGAGGA ATACGAGAAGGTCGAGAAAGAGTACAA GACCCTGGAAGAGAGGATCAAAGAGGA CATCCAGGCTCTGAAGGCTCTGGAACAG TATGAGAAAGAGCGGCAAGAACAGCTG CTGCGGGACACCCTGAACACCAACGAGT ACCGGCTGAGCAAGAGAGGCCTTAGAG GCTGGCGGGAAATCATCCAGAAATGGCT GAAAATGGACGGAGGCTCTGGAGGAAG CTCCGAAGTCGAGTTTTCCCATGAGTAC TGGATGAGACACGCATTGACTCTCGCAA AGAGGGCTCGAGATGAACGCGAGGTGC CCGTGGGGGCAGTACTCGTGCTCAACAA TCGCGTAATCGGCGAAGGTTGGAATAGG GCAATCGGACTCCACGACCCCACTGCAC ATGCGGAAATCATGGCCCTTCGACAGGG AGGGCTTGTGATGCAGAATTATCGACTT TATGATGCGACGCTGTACGTCACGTTTG

[1206]

		AACCTTGCGTAATGTGCGCGGGAGCTAT GATTCACTCCCGCATTGGACGAGTTGTA TTCGGTGTTCGCAACGCCAAGACGGGTG CCGCAGGTTCACTGATGGACGTGCTGCA TCATCCAGGCATGAACCACCGGGTAGAA ATCACAGAAGGCATATTGGCGGACGAAT GTGCGGCGCTGTTGTGTCGTTTTTTTCGC ATGCCCAGGCGGGTCTTTAACGCCCAGA AAAAAGCACAATCCTCTACTGACGGCTC TTCTGGATCTGAAACACCTGGCACAAGC GAGAGCGCCACCCCTGAGAGCTCTGGCG AGAACGAGCCCTCCGAGAAGTACCTGGA AGTGTTCAAGGACTACCAGCGGAAGCAC CCTAGAGAGGCCGGCGATTACAGCGTGT ACGAGTTCCTGTCCAAGAAAGAGAACCA CTTCATCTGGCGGAATCACCCTGAGTAC CCCTACCTGTACGCCACCTTCTGCGAGA TCGACAAGAAAAAGAAGGACGCCAAGC AGCAGGCCACCTTCACACTGGCCGATCC TATCAATCACCCTCTGTGGGTCCGATTCG AGGAAAGAAGCGGCAGCAACCTGAACA AGTACAGAATCCTGACCGAGCAGCTGCA CACCGAGAAGCTGAAGAAAAAGCTGAC AGTGCAGCTGGACCGGCTGATCTACCCT ACAGAATCTGGCGGCTGGGAAGAGAAG GGCAAAGTGGACATTGTGCTGCTGCCCA GCCGGCAGTTCTACAACCAGATCTTCCT GGACATCGAGGAAAAGGGCAAGCACGC CTTCACCTACAAGGATGAGAGCATCAAG TTCCCTCTGAAGGGCACACTCGGCGGAG CCAGAGTGCAGTTCGACAGAGATCACCT GAGAAGATACCCTCACAAGGTGGAAAG CGGCAACGTGGGCAGAATCTACTTCAAC ATGACCGTGAACATCGAGCCTACAGAGT CCCCAGTGTCCAAGTCTCTGAAGATCCA CCGGGACGACTTCCCCAAGGTGGTCAAC TTCAAGCCCAAAGAACTGACCGAGTGGA TCAAGGACAGCAAGGGCAAGAAACTGA AGTCCGGCATCGAGTCCCTGGAAATCGG CCTGAGAGTGATGAGCATCGACCTGGGA CAGAGACAGGCCGCTGCCGCCTCTATTT TCGAGGTGGTGGATCAGAAGCCCGACAT CGAAGGCAAGCTGTTTTTCCCAATCAAG

[1207]

		GGCACCGAGCTGTATGCCGTGCACAGAG CCAGCTTCAACATCAAGCTGCCCGGCGA GACACTGGTCAAGAGCAGAGAAGTGCT GCGGAAGGCCAGAGAGGACAATCTGAA ACTGATGAACCAGAAGCTCAACTTCCTG CGGAACGTGCTGCACTTCCAGCAGTTCG AGGACATCACCGAGAGAGAGAAGCGGG TCACCAAGTGGATCAGCAGACAAGAGA ACAGCGACGTGCCCCTGGTGTACCAGGA TGAGCTGATCCAGATCCGCGAGCTGATG TACAAGCCTTACAAGGACTGGGTCGCCT TCCTGAAGCAGCTCCACAAGAGACTGGA AGTCGAGATCGGCAAAGAAGTGAAGCA CTGGCGGAAGTCCCTGAGCGACGGAAG AAAGGGCCTGTACGGCATCTCCCTGAAG AACATCGACGAGATCGATCGGACCCGGA AGTTCCTGCTGAGATGGTCCCTGAGGCC TACCGAACCTGGCGAAGTGCGTAGACTG GAACCCGGCCAGAGATTCGCCATCGACC AGCTGAATCACCTGAACGCCCTGAAAGA AGATCGGCTGAAGAAGATGGCCAACAC CATCATCATGCACGCCCTGGGCTACTGC TACGACGTGCGGAAGAAGAAATGGCAG GCTAAGAACCCCGCCTGCCAGATCATCC TGTTCGAGGATCTGAGCAACTACAACCC CTACGAGGAAAGGTCCCGCTTCGAGAAC AGCAAGCTCATGAAGTGGTCCAGACGCG AGATCCCCAGACAGGTTGCACTGCAGGG CGAGATCTATGGCCTGCAAGTGGGAGAA GTGGGCGCTCAGTTCAGCAGCAGATTCC ACGCCAAGACAGGCAGCCCTGGCATCAG ATGTAGCGTCGTGACCAAAGAGAAGCTG CAGGACAATCGGTTCTTCAAGAATCTGC AGAGAGAGGGCAGACTGACCCTGGACA AAATCGCCGTGCTGAAAGAGGGCGATCT GTACCCAGACAAAGGCGGCGAGAAGTT CATCAGCCTGAGCAAGGATCGGAAGTGC GTGACCACACACGCCGACATCAACGCCG CTCAGAACCTGCAGAAGCGGTTCTGGAC AAGAACCCACGGCTTCTACAAGGTGTAC TGCAAGGCCTACCAGGTGGACGGCCAGA CCGTGTACATCCCTGAGAGCAAGGACCA GAAGCAGAAGATCATCGAAGAGTTCGG

[1208]

		CGAGGGCTACTTCATTCTGAAGGACGGG GTGTACGAATGGGTCAACGCCGGCAAGC TGAAAATCAAGAAGGGCAGCTCCAAGC AGAGCAGCAGCGAGCTGGTGGATAGCG ACATCCTGAAAGACAGCTTCGACCTGGC CTCCGAGCTGAAAGGCGAAAAGCTGATG CTGTACAGGGACCCCAGCGGCAATGTGT TCCCCAGCGACAAATGGATGGCCGCTGG CGTGTTCTTCGGAAAGCTGGAACGCATC CTGATCAGCAAGCTGACCAACCAGTACT CCATCAGCACCATCGAGGACGACAGCAG CAAGCAGTCTATGAAAAGGCCGGCGGCC ACGAAAAAGGCCGGCCAGGCAAAAAAG AAAAAGGGATCCTACCCATACGATGTTC CAGATTACGCTTATCCCTACGACGTGCC TGATTATGCATACCCATATGATGTCCCC GACTATGCCTAA
位于D306多肽处的BhCas12b GGSGGS-ABE8-Xten20	279	MAPKKKRKVGIHGVPAAATRSFILKIEPNE EVKKGLWKTHEVLNHGIAYYMNILKLIRQ EAIYEHHEQDPKNPKKVSKAEIQAELWDF VLKMQKCNSFTHEVDKDEVFNILRELYEE LVPSSVEKKGEANQLSNKFLYPLVDPNSQ SGKGTASSGRKPRWYNLKIAGDPSWEEEK KKWEEDKKKDPLAKILGKLAEYGLIPLFIP YTDSNEPIVKEIKWMEKSRNQSVRRLDKD MFIQALERFLSWESWNLKVKEEYEKVEKE YKTLEERIKEDIQALKALEQYEKERQEQLL RDTLNTNEYRLSKRGLRGWREIIQKWLKM DGGSGGSSEVEFSHEYWMRHALTLAKRA RDEREVPVGAVLVLNNRVIGEGWNRAIGL HDPTAHAEIMALRQGGLVMQNYRLYDAT LYVTFEPCVMCAGAMIHSRIGRVVFGVRN AKTGAAGSLMDVLHHPGMNHRVEITEGIL ADECAALLCRFFRMPRRVFNAQKKAQSST DGSSGSETPGTSESATPESSGENEPSEKYLE VFKDYQRKHPREAGDYSVYEFLSKKENHF IWRNHPEYPYLYATFCEIDKKKKDAKQQA TFTLADPINHPLWVRFEERSGSNLNKYRIL TEQLHTEKLKKKLTVQLDRLIYPTESGGW EEKGKVDIVLLPSRQFYNQIFLDIEEKGKH AFTYKDESIKFPLKGTLGGARVQFDRDHL RRYPHKVESGNVGRIYFNMTVNIEPTESPV SKSLKIHRDDFPKVVNFKPKELTEWIKDSK

			GKKLKSGIESLEIGLRVMSIDLGQRQAAAA SIFEVVDQKPDIEGKLFFPIKGTELYAVHR ASFNIKLPGETLVKSREVLRKAREDNLKL MNQKLNFLRNVLHFQQFEDITEREKRVTK WISRQENSDVPLVYQDELIQIRELMYKPYK DWVAFLKQLHKRLEVEIGKEVKHWRKSL SDGRKGLYGISLKNIDEIDRTRKFLLRWSL RPTEPGEVRRLEPGQRFAIDQLNHLNALKE DRLKKMANTIIMHALGYCYDVRKKKWQ AKNPACQIILFEDLSNYNPYEERSRFENSK LMKWSRREIPRQVALQGEIYGLQVGEVGA QFSSRFHAKTGSPGIRCSVVTKEKLQDNRF FKNLQREGRLTLDKIAVLKEGDLYPDKGG EKFISLSKDRKCVTTHADINAAQNLQKRF WTRTHGFYKVYCKAYQVDGQTVYIPESK DQKQKIIEEFGEGYFILKDGVYEWVNAGK LKIKKGSSKQSSSELVDSDILKDSFDLASEL KGEKLMLYRDPSGNVFPSDKWMAAGVFF GKLERILISKLTNQYSISTIEDDSSKQSMKR PAATKKAGQAKKKKGSYPYDVPDYAYPY DVPDYAYPYDVPDYA
[1209]	位于D980多核苷酸处的BhCas12bGGSGGS-ABE8-Xten20	280	GCCACCATGGCCCCAAAGAAGAAGCGG AAGGTCGGTATCCACGGAGTCCCAGCAG CCGCCACCAGATCCTTCATCCTGAAGAT CGAGCCCAACGAGGAAGTGAAGAAAGG CCTCTGGAAAACCCACGAGGTGCTGAAC CACGGAATCGCCTACTACATGAATATCC TGAAGCTGATCCGGCAAGAGGCCATCTA CGAGCACCACGAGCAGGACCCCAAGAA TCCCAAGAAGGTGTCCAAGGCCGAGATC CAGGCCGAGCTGTGGGATTTCGTGCTGA AGATGCAGAAGTGCAACAGCTTCACACA CGAGGTGGACAAGGACGAGGTGTTCAA CATCCTGAGAGAGCTGTACGAGGAACTG GTGCCCAGCAGCGTGGAAAAGAAGGGC GAAGCCAACCAGCTGAGCAACAAGTTTC TGTACCCTCTGGTGGACCCCAACAGCCA GTCTGGAAAGGGAACAGCCAGCAGCGG CAGAAAGCCCAGATGGTACAACCTGAA GATTGCCGGCGATCCCTCCTGGGAAGAA GAGAAGAAGAAGTGGGAAGAAGATAAG AAAAAGGACCCGCTGGCCAAGATCCTGG GCAAGCTGGCTGAGTACGGACTGATCCC

[1210]

		TCTGTTCATCCCCTACACCGACAGCAAC GAGCCCATCGTGAAAGAAATCAAGTGG ATGGAAAAGTCCCGGAACCAGAGCGTG CGGCGGCTGGATAAGGACATGTTCATTC AGGCCCTGGAACGGTTCCTGAGCTGGGA GAGCTGGAACCTGAAAGTGAAAGAGGA ATACGAGAAGGTCGAGAAAGAGTACAA GACCCTGGAAGAGAGGATCAAAGAGGA CATCCAGGCTCTGAAGGCTCTGGAACAG TATGAGAAAGAGCGGCAAGAACAGCTG CTGCGGGACACCCTGAACACCAACGAGT ACCGGCTGAGCAAGAGAGGCCTTAGAG GCTGGCGGGAAATCATCCAGAAATGGCT GAAAATGGACGAGAACGAGCCCTCCGA GAAGTACCTGGAAGTGTTCAAGGACTAC CAGCGGAAGCACCCTAGAGAGGCCGGC GATTACAGCGTGTACGAGTTCCTGTCCA AGAAAGAGAACCACTTCATCTGGCGGAA TCACCCTGAGTACCCCTACCTGTACGCC ACCTTCTGCGAGATCGACAAGAAAAAGA AGGACGCCAAGCAGCAGGCCACCTTCAC ACTGGCCGATCCTATCAATCACCCTCTGT GGGTCCGATTCGAGGAAAGAAGCGGCA GCAACCTGAACAAGTACAGAATCCTGAC CGAGCAGCTGCACACCGAGAAGCTGAA GAAAAAGCTGACAGTGCAGCTGGACCG GCTGATCTACCCTACAGAATCTGGCGGC TGGGAAGAGAAGGGCAAAGTGGACATT GTGCTGCTGCCCAGCCGGCAGTTCTACA ACCAGATCTTCCTGGACATCGAGGAAAA GGGCAAGCACGCCTTCACCTACAAGGAT GAGAGCATCAAGTTCCCTCTGAAGGGCA CACTCGGCGGAGCCAGAGTGCAGTTCGA CAGAGATCACCTGAGAAGATACCCTCAC AAGGTGGAAAGCGGCAACGTGGGCAGA ATCTACTTCAACATGACCGTGAACATCG AGCCTACAGAGTCCCCAGTGTCCAAGTC TCTGAAGATCCACCGGGACGACTTCCCC AAGGTGGTCAACTTCAAGCCCAAAGAAC TGACCGAGTGGATCAAGGACAGCAAGG GCAAGAAACTGAAGTCCGGCATCGAGTC CCTGGAAATCGGCCTGAGAGTGATGAGC ATCGACCTGGGACAGAGACAGGCCGCTG

[1211]

		CCGCCTCTATTTTCGAGGTGGTGGATCA GAAGCCCGACATCGAAGGCAAGCTGTTT TTCCCAATCAAGGGCACCGAGCTGTATG CCGTGCACAGAGCCAGCTTCAACATCAA GCTGCCCGGCGAGACACTGGTCAAGAGC AGAGAAGTGCTGCGGAAGGCCAGAGAG GACAATCTGAAACTGATGAACCAGAAGC TCAACTTCCTGCGGAACGTGCTGCACTT CCAGCAGTTCGAGGACATCACCGAGAGA GAGAAGCGGGTCACCAAGTGGATCAGC AGACAAGAGAACAGCGACGTGCCCCTG GTGTACCAGGATGAGCTGATCCAGATCC GCGAGCTGATGTACAAGCCTTACAAGGA CTGGGTCGCCTTCCTGAAGCAGCTCCAC AAGAGACTGGAAGTCGAGATCGGCAAA GAAGTGAAGCACTGGCGGAAGTCCCTGA GCGACGGAAGAAAGGGCCTGTACGGCA TCTCCCTGAAGAACATCGACGAGATCGA TCGGACCCGGAAGTTCCTGCTGAGATGG TCCCTGAGGCCTACCGAACCTGGCGAAG TGCGTAGACTGGAACCCGGCCAGAGATT CGCCATCGACCAGCTGAATCACCTGAAC GCCCTGAAAGAAGATCGGCTGAAGAAG ATGGCCAACACCATCATCATGCACGCCC TGGGCTACTGCTACGACGTGCGGAAGAA GAAATGGCAGGCTAAGAACCCCGCCTGC CAGATCATCCTGTTCGAGGATCTGAGCA ACTACAACCCCTACGAGGAAAGGTCCCG CTTCGAGAACAGCAAGCTCATGAAGTGG TCCAGACGCGAGATCCCCAGACAGGTTG CACTGCAGGGCGAGATCTATGGCCTGCA AGTGGGAGAAGTGGGCGCTCAGTTCAGC AGCAGATTCCACGCCAAGACAGGCAGCC CTGGCATCAGATGTAGCGTCGTGACCAA AGAGAAGCTGCAGGACAATCGGTTCTTC AAGAATCTGCAGAGAGAGGGCAGACTG ACCCTGGACAAAATCGCCGTGCTGAAAG AGGGCGATCTGTACCCAGACAAAGGCG GCGAGAAGTTCATCAGCCTGAGCAAGGA TCGGAAGTGCGTGACCACACACGCCGAC ATCAACGCCGCTCAGAACCTGCAGAAGC GGTTCTGGACAAGAACCCACGGCTTCTA CAAGGTGTACTGCAAGGCCTACCAGGTG

[1212]

		GACGGAGGCTCTGGAGGAAGCTCCGAA GTCGAGTTTTCCCATGAGTACTGGATGA GACACGCATTGACTCTCGCAAAGAGGGC TCGAGATGAACGCGAGGTGCCCGTGGGG GCAGTACTCGTGCTCAACAATCGCGTAA TCGGCGAAGGTTGGAATAGGGCAATCGG ACTCCACGACCCCACTGCACATGCGGAA ATCATGGCCCTTCGACAGGGAGGGCTTG TGATGCAGAATTATCGACTTTATGATGC GACGCTGTACGTCACGTTTGAACCTTGC GTAATGTGCGCGGGAGCTATGATTCACT CCCGCATTGGACGAGTTGTATTCGGTGT TCGCAACGCCAAGACGGGTGCCGCAGGT TCACTGATGGACGTGCTGCATCATCCAG GCATGAACCACCGGGTAGAAATCACAG AAGGCATATTGGCGGACGAATGTGCGGC GCTGTTGTGTCGTTTTTTTCGCATGCCCA GGCGGGTCTTTAACGCCCAGAAAAAAGC ACAATCCTCTACTGACGGCTCTTCTGGAT CTGAAACACCTGGCACAAGCGAGAGCG CCACCCCTGAGAGCTCTGGCGGCCAGAC CGTGTACATCCCTGAGAGCAAGGACCAG AAGCAGAAGATCATCGAAGAGTTCGGC GAGGGCTACTTCATTCTGAAGGACGGGG TGTACGAATGGGTCAACGCCGGCAAGCT GAAAATCAAGAAGGGCAGCTCCAAGCA GAGCAGCAGCGAGCTGGTGGATAGCGA CATCCTGAAAGACAGCTTCGACCTGGCC TCCGAGCTGAAAGGCGAAAAGCTGATGC TGTACAGGGACCCCAGCGGCAATGTGTT CCCCAGCGACAAATGGATGGCCGCTGGC GTGTTCTTCGGAAAGCTGGAACGCATCC TGATCAGCAAGCTGACCAACCAGTACTC CATCAGCACCATCGAGGACGACAGCAGC AAGCAGTCTATGAAAAGGCCGGCGGCC ACGAAAAAGGCCGGCCAGGCAAAAAAG AAAAAGGGATCCTACCCATACGATGTTC CAGATTACGCTTATCCCTACGACGTGCC TGATTATGCATACCCATATGATGTCCCC GACTATGCCTAA
位于 D980 多	281	MAPKKKRKVGIHGVPAAATRSFILKIEPNE EVKKGLWKTHEVLNHGIAYYMNILKLIRQ EAIYEHHEQDPKNPKKVSKAEIQAELWDF

[1213]

肽处的BhCas12bGGSGGS-ABE8-Xten20		VLKMQKCNSFTHEVDKDEVFNILRELYEE LVPSSVEKKGEANQLSNKFLYPLVDPNSQ SGKGTASSGRKPRWYNLKIAGDPSWEEEK KKWEEDKKKDPLAKILGKLAEYGLIPLFIP YTDSNEPIVKEIKWMEKSRNQSVRRLDKD MFIQALERFLSWESWNLKVKEEYEKVEKE YKTLEERIKEDIQALKALEQYEKERQEQLL RDTLNTNEYRLSKRGLRGWREIIQKWLKM DENEPSEKYLEVFKDYQRKHPREAGDYSV YEFLSKKENHFIWRNHPEYPYLYATFCEID KKKKDAKQQATFTLADPINHPLWVRFEER SGSNLNKYRILTEQLHTEKLKKKLTVQLD RLIYPTESGGWEEKGKVDIVLLPSRQFYNQ IFLDIEEKGKHAFTYKDESIKFPLKGTLGG ARVQFDRDHLRRYPHKVESGNVGRIYFN MTVNIEPTESPVSKSLKIHRDDFPKVVNFK PKELTEWIKDSKGKKLKSGIESLEIGLRVM SIDLGQRQAAAASIFEVVDQKPDIEGKLFF PIKGTELYAVHRASFNIKLPGETLVKSREV LRKAREDNLKLMNQKLNFLRNVLHFQQF EDITEREKRVTKWISRQENSDVPLVYQDEL IQIRELMYKPYKDWVAFLKQLHKRLEVEI GKEVKHWRKSLSDGRKGLYGISLKNIDEI DRTRKFLLRWSLRPTEPGEVRRLEPGQRF AIDQLNHLNALKEDRLKKMANTIIMHALG YCYDVRKKKWQAKNPACQIILFEDLSNYN PYEERSRFENSKLMKWSRREIPRQVALQG EIYGLQVGEVGAQFSSRFHAKTGSPGIRCS VVTKEKLQDNRFFKNLQREGRLTLDKIAV LKEGDLYPDKGGEKFISLSKDRKCVTTHA DINAAQNLQKRFWTRTHGFYKVYCKAYQ VDGGSGGSSEVEFSHEYWMRHALTLAKR ARDEREVPVGAVLVLNNRVIGEGWNRAIG LHDPTAHAEIMALRQGGLVMQNYRLYDA TLYVTFEPCVMCAGAMIHSRIGRVVFGVR NAKTGAAGSLMDVLHHPGMNHRVEITEGI LADECAALLCRFFRMPRRVFNAQKKAQSS TDGSSGSETPGTSESATPESSGGQTVYIPES KDQKQKIIEEFGEGYFILKDGVYEWVNAG KLKIKKGSSKQSSSELVDSDILKDSFDLASE LKGEKLMLYRDPSGNVFPSDKWMAAGVF FGKLERILISKLTNQYSISTIEDDSSKQSMK RPAATKKAGQAKKKKGSYPYDVPDYAYP

[1214]

		YDVPDYAYPYDVPDYA
位于K1019多核苷酸处的BhCas12b GGSGGS-ABE8-Xten20	282	GCCACCATGGCCCCAAAGAAGAAGCGG AAGGTCGGTATCCACGGAGTCCCAGCAG CCGCCACCAGATCCTTCATCCTGAAGAT CGAGCCCAACGAGGAAGTGAAGAAAGG CCTCTGGAAAACCCACGAGGTGCTGAAC CACGGAATCGCCTACTACATGAATATCC TGAAGCTGATCCGGCAAGAGGCCATCTA CGAGCACCACGAGCAGGACCCCAAGAA TCCCAAGAAGGTGTCCAAGGCCGAGATC CAGGCCGAGCTGTGGGATTTCGTGCTGA AGATGCAGAAGTGCAACAGCTTCACACA CGAGGTGGACAAGGACGAGGTGTTCAA CATCCTGAGAGAGCTGTACGAGGAACTG GTGCCCAGCAGCGTGGAAAAGAAGGGC GAAGCCAACCAGCTGAGCAACAAGTTTC TGTACCCTCTGGTGGACCCCAACAGCCA GTCTGGAAAGGGAACAGCCAGCAGCGG CAGAAAGCCCAGATGGTACAACCTGAA GATTGCCGGCGATCCCTCCTGGGAAGAA GAGAAGAAGAAGTGGGAAGAAGATAAG AAAAAGGACCCGCTGGCCAAGATCCTGG GCAAGCTGGCTGAGTACGGACTGATCCC TCTGTTCATCCCCTACACCGACAGCAAC GAGCCCATCGTGAAAGAAATCAAGTGG ATGGAAAAGTCCCGGAACCAGAGCGTG CGGCGGCTGGATAAGGACATGTTCATTC AGGCCCTGGAACGGTTCCTGAGCTGGGA GAGCTGGAACCTGAAAGTGAAAGAGGA ATACGAGAAGGTCGAGAAAGAGTACAA GACCCTGGAAGAGAGGATCAAAGAGGA CATCCAGGCTCTGAAGGCTCTGGAACAG TATGAGAAAGAGCGGCAAGAACAGCTG CTGCGGGACACCCTGAACACCAACGAGT ACCGGCTGAGCAAGAGAGGCCTTAGAG GCTGGCGGGAAATCATCCAGAAATGGCT GAAAATGGACGAGAACGAGCCCTCCGA GAAGTACCTGGAAGTGTTCAAGGACTAC CAGCGGAAGCACCCTAGAGAGGCCGGC GATTACAGCGTGTACGAGTTCCTGTCCA AGAAAGAGAACCACTTCATCTGGCGGAA TCACCCTGAGTACCCCTACCTGTACGCC ACCTTCTGCGAGATCGACAAGAAAAAGA

[1215]

		AGGACGCCAAGCAGCAGGCCACCTTCAC ACTGGCCGATCCTATCAATCACCCTCTGT GGGTCCGATTCGAGGAAAGAAGCGGCA GCAACCTGAACAAGTACAGAATCCTGAC CGAGCAGCTGCACACCGAGAAGCTGAA GAAAAAGCTGACAGTGCAGCTGGACCG GCTGATCTACCCTACAGAATCTGGCGGC TGGGAAGAGAAGGGCAAAGTGGACATT GTGCTGCTGCCCAGCCGGCAGTTCTACA ACCAGATCTTCCTGGACATCGAGGAAAA GGGCAAGCACGCCTTCACCTACAAGGAT GAGAGCATCAAGTTCCCTCTGAAGGGCA CACTCGGCGGAGCCAGAGTGCAGTTCGA CAGAGATCACCTGAGAAGATACCCTCAC AAGGTGGAAAGCGGCAACGTGGGCAGA ATCTACTTCAACATGACCGTGAACATCG AGCCTACAGAGTCCCCAGTGTCCAAGTC TCTGAAGATCCACCGGGACGACTTCCCC AAGGTGGTCAACTTCAAGCCCAAAGAAC TGACCGAGTGGATCAAGGACAGCAAGG GCAAGAAACTGAAGTCCGGCATCGAGTC CCTGGAAATCGGCCTGAGAGTGATGAGC ATCGACCTGGGACAGAGACAGGCCGCTG CCGCCTCTATTTTCGAGGTGGTGGATCA GAAGCCCGACATCGAAGGCAAGCTGTTT TTCCCAATCAAGGGCACCGAGCTGTATG CCGTGCACAGAGCCAGCTTCAACATCAA GCTGCCCGGCGAGACACTGGTCAAGAGC AGAGAAGTGCTGCGGAAGGCCAGAGAG GACAATCTGAAACTGATGAACCAGAAGC TCAACTTCCTGCGGAACGTGCTGCACTT CCAGCAGTTCGAGGACATCACCGAGAGA GAGAAGCGGGTCACCAAGTGGATCAGC AGACAAGAGAACAGCGACGTGCCCCTG GTGTACCAGGATGAGCTGATCCAGATCC GCGAGCTGATGTACAAGCCTTACAAGGA CTGGGTCGCCTTCCTGAAGCAGCTCCAC AAGAGACTGGAAGTCGAGATCGGCAAA GAAGTGAAGCACTGGCGGAAGTCCCTGA GCGACGGAAGAAAGGGCCTGTACGGCA TCTCCCTGAAGAACATCGACGAGATCGA TCGGACCCGGAAGTTCCTGCTGAGATGG TCCCTGAGGCCTACCGAACCTGGCGAAG

[1216]

		TGCGTAGACTGGAACCCGGCCAGAGATT CGCCATCGACCAGCTGAATCACCTGAAC GCCCTGAAAGAAGATCGGCTGAAGAAG ATGGCCAACACCATCATCATGCACGCCC TGGGCTACTGCTACGACGTGCGGAAGAA GAAATGGCAGGCTAAGAACCCCGCCTGC CAGATCATCCTGTTCGAGGATCTGAGCA ACTACAACCCCTACGAGGAAAGGTCCCG CTTCGAGAACAGCAAGCTCATGAAGTGG TCCAGACGCGAGATCCCCAGACAGGTTG CACTGCAGGGCGAGATCTATGGCCTGCA AGTGGGAGAAGTGGGCGCTCAGTTCAGC AGCAGATTCCACGCCAAGACAGGCAGCC CTGGCATCAGATGTAGCGTCGTGACCAA AGAGAAGCTGCAGGACAATCGGTTCTTC AAGAATCTGCAGAGAGAGGGCAGACTG ACCCTGGACAAAATCGCCGTGCTGAAAG AGGGCGATCTGTACCCAGACAAAGGCG GCGAGAAGTTCATCAGCCTGAGCAAGGA TCGGAAGTGCGTGACCACACACGCCGAC ATCAACGCCGCTCAGAACCTGCAGAAGC GGTTCTGGACAAGAACCCACGGCTTCTA CAAGGTGTACTGCAAGGCCTACCAGGTG GACGGCCAGACCGTGTACATCCCTGAGA GCAAGGACCAGAAGCAGAAGATCATCG AAGAGTTCGGCGAGGGCTACTTCATTCT GAAGGACGGGGTGTACGAATGGGTCAA CGCCGGCAAGGGAGGCTCTGGAGGAAG CTCCGAAGTCGAGTTTTCCCATGAGTAC TGGATGAGACACGCATTGACTCTCGCAA AGAGGGCTCGAGATGAACGCGAGGTGC CCGTGGGGGCAGTACTCGTGCTCAACAA TCGCGTAATCGGCGAAGGTTGGAATAGG GCAATCGGACTCCACGACCCCACTGCAC ATGCGGAAATCATGGCCCTTCGACAGGG AGGGCTTGTGATGCAGAATTATCGACTT TATGATGCGACGCTGTACGTCACGTTTG AACCTTGCGTAATGTGCGCGGGAGCTAT GATTCACTCCCGCATTGGACGAGTTGTA TTCGGTGTTCGCAACGCCAAGACGGGTG CCGCAGGTTCACTGATGGACGTGCTGCA TCATCCAGGCATGAACCACCGGGTAGAA ATCACAGAAGGCATATTGGCGGACGAAT

[1217]

		GTGCGGCGCTGTTGTGTCGTTTTTTTCGC ATGCCCAGGCGGGTCTTTAACGCCCAGA AAAAAGCACAATCCTCTACTGACGGCTC TTCTGGATCTGAAACACCTGGCACAAGC GAGAGCGCCACCCCTGAGAGCTCTGGCC TGAAAATCAAGAAGGGCAGCTCCAAGC AGAGCAGCAGCGAGCTGGTGGATAGCG ACATCCTGAAAGACAGCTTCGACCTGGC CTCCGAGCTGAAAGGCGAAAAGCTGATG CTGTACAGGGACCCCAGCGGCAATGTGT TCCCCAGCGACAAATGGATGGCCGCTGG CGTGTTCTTCGGAAAGCTGGAACGCATC CTGATCAGCAAGCTGACCAACCAGTACT CCATCAGCACCATCGAGGACGACAGCAG CAAGCAGTCTATGAAAAGGCCGGCGGCC ACGAAAAAGGCCGGCCAGGCAAAAAAG AAAAAGGGATCCTACCCATACGATGTTC CAGATTACGCTTATCCCTACGACGTGCC TGATTATGCATACCCATATGATGTCCCC GACTATGCCTAA
位于K1019多肽处的BhCas12bGGSGGS-ABE8-Xten20	283	MAPKKKRKVGIHGVPAAATRSFILKIEPNE EVKKGLWKTHEVLNHGIAYYMNILKLIRQ EAIYEHHEQDPKNPKKVSKAEIQAELWDF VLKMQKCNSFTHEVDKDEVFNILRELYEE LVPSSVEKKGEANQLSNKFLYPLVDPNSQ SGKGTASSGRKPRWYNLKIAGDPSWEEEK KKWEEDKKKDPLAKILGKLAEYGLIPLFIP YTDSNEPIVKEIKWMEKSRNQSVRRLDKD MFIQALERFLSWESWNLKVKEEYEKVEKE YKTLEERIKEDIQALKALEQYEKERQEQLL RDTLNTNEYRLSKRGLRGWREIIQKWLKM DENEPSEKYLEVFKDYQRKHPREAGDYSV YEFLSKKENHFIWRNHPEYPYLYATFCEID KKKKDAKQQATFTLADPINHPLWVRFEER SGSNLNKYRILTEQLHTEKLKKKLTVQLD RLIYPTESGGWEEKGKVDIVLLPSRQFYNQ IFLDIEEKGKHAFTYKDESIKFPLKGTLGG ARVQFDRDHLRRYPHKVESGNVGRIYFN MTVNIEPTESPVSKSLKIHRDDFPKVVNFK PKELTEWIKDSKGKKLKSGIESLEIGLRVM SIDLGQRQAAAASIFEVVDQKPDIEGKLFF PIKGTELYAVHRASFNIKLPGETLVKSREV LRKAREDNLKLMNQKLNFLRNVLHFQQF

[1218]

		EDITEREKRVTKWISRQENSDVPLVYQDEL IQIRELMYKPYKDWVAFLKQLHKRLEVEI GKEVKHWRKSLSDGRKGLYGISLKNIDEI DRTRKFLLRWSLRPTEPGEVRRLEPGQRF AIDQLNHLNALKEDRLKKMANTIIMHALG YCYDVRKKKWQAKNPACQIILFEDLSNYN PYEERSRFENSKLMKWSRREIPRQVALQG EIYGLQVGEVGAQFSSRFHAKTGSPGIRCS VVTKEKLQDNRFFKNLQREGRLTLDKIAV LKEGDLYPDKGGEKFISLSKDRKCVTTHA DINAAQNLQKRFWTRTHGFYKVYCKAYQ VDGQTVYIPESKDQKQKIIEEFGEGYFILK DGVYEWVNAGKGGSGGSSEVEFSHEYW MRHALTLAKRARDEREVPVGAVLVLNNR VIGEGWNRAIGLHDPTAHAEIMALRQGGL VMQNYRLYDATLYVTFEPCVMCAGAMIH SRIGRVVFGVRNAKTGAAGSLMDVLHHP GMNHRVEITEGILADECAALLCRFFRMPR RVFNAQKKAQSSTDGSSGSETPGTSESATP ESSGLKIKKGSSKQSSSELVDSDILKDSFDL ASELKGEKLMLYRDPSGNVFPSDKWMAA GVFFGKLERILISKLTNQYSISTIEDDSSKQS MKRPAATKKAGQAKKKKGSYPYDVPDY AYPYDVPDYAYPYDVPDYA
tr|A5H718|A5H718_PETMA 胞嘧啶脱氨酶 OS=海七鳃鳗 OX=7757 PE=2 SV=1 氨基酸序列；PmCDA1 氨基酸序列	41	MTDAEYVRIHEKLDIYTFKKQFFNNKKSV SHRCYVLFELKRRGERRACFWGYAVNKP QSGTERGIHAEIFSIRKVEEYLRDNPGQFTI NWYSSWSPCADCAEKILEWYNQELRGNG HTLKIWACKLYYEKNARNQIGLWNLRDN GVGLNVMVSEHYQCCRKIFIQSSHNQLNE NRWLEKTLKRAEKRRSELSIMIQVKILHTT KSPAV
EF094822.1 海七鳃鳗分离物	42	TGACACGACACAGCCGTGTATATGAGGA AGGGTAGCTGGATGGGGGGGGGGGGAA TACGTTCAGAGAGGACATTAGCGAGCGT CTTGTTGGTGGCCTTGAGTCTAGACACCT

[1219]

PmCDA.21 胞嘧啶脱氨酶 mRNA，完整 cd；PmCDA1 氨基酸序列		GCAGACATGACCGACGCTGAGTACGTGA GAATCCATGAGAAGTTGGACATCTACAC GTTTAAGAAACAGTTTTTCAACAACAAA AAATCCGTGTCGCATAGATGCTACGTTC TCTTTGAATTAAAACGACGGGGTGAACG TAGAGCGTGTTTTTGGGGCTATGCTGTG AATAAACCACAGAGCGGGACAGAACGT GGAATTCACGCCGAAATCTTTAGCATTA GAAAAGTCGAAGAATACCTGCGCGACA ACCCCGGACAATTCACGATAAATTGGTA CTCATCCTGGAGTCCTTGTGCAGATTGC GCTGAAAAGATCTTAGAATGGTATAACC AGGAGCTGCGGGGGAACGGCCACACTTT GAAAATCTGGGCTTGCAAACTCTATTAC GAGAAAAATGCGAGGAATCAAATTGGG CTGTGGAACCTCAGAGATAACGGGGTTG GGTTGAATGTAATGGTAAGTGAACACTA CCAATGTTGCAGGAAAATATTCATCCAA TCGTCGCACAATCAATTGAATGAGAATA GATGGCTTGAGAAGACTTTGAAGCGAGC TGAAAAACGACGGAGCGAGTTGTCCATT ATGATTCAGGTAAAAATACTCCACACCA CTAAGAGTCCTGCTGTTTAAGAGGCTAT GCGGATGGTTTTC
tr|Q6QJ80|Q6QJ80_HUMAN 激活诱导的胞苷脱氨酶 OS=智人 OX=9606 GN=AICDA PE=2 SV=1；AID 氨基酸序列	43	MDSLLMNRRKFLYQFKNVRWAKGRRETY LCYVVKRRDSATSFSLDFGYLRNKNGCHV ELLFLRYISDWDLDPGRCYRVTWFTSWSP CYDCARHVADFLRGNPNLSLRIFTARLYF CEDRKAEPEGLRRLHRAGVQIAIMTFKAP V
NG_011588.1:5001-15681 智人激活诱	44	AGAGAACCATCATTAATTGAAGTGAGAT TTTTCTGGCCTGAGACTTGCAGGGAGGC AAGAAGACACTCTGGACACCACTATGGA CAGGTAAAGAGGCAGTCTTCTCGTGGGT GATTGCACTGGCCTTCCTCTCAGAGCAA

[1220]

导的胞苷脱氨酶(AICDA)，12号染色体上的RefSeqGene (LRG_17)；人AID的CDS的核酸序列		ATCTGAGTAATGAGACTGGTAGCTATCC CTTTCTCTCATGTAACTGTCTGACTGATA AGATCAGCTTGATCAATATGCATATATA TTTTTTGATCTGTCTCCTTTTCTTCTATTC AGATCTTATACGCTGTCAGCCCAATTCTT TCTGTTTCAGACTTCTCTTGATTTCCCTC TTTTTCATGTGGCAAAAGAAGTAGTGCG TACAATGTACTGATTCGTCCTGAGATTTG TACCATGGTTGAAACTAATTTATGGTAA TAATATTAACATAGCAAATCTTTAGAGA CTCAAATCATGAAAAGGTAATAGCAGTA CTGTACTAAAAACGGTAGTGCTAATTTT CGTAATAATTTTGTAAATATTCAACAGT AAAACAACTTGAAGACACACTTTCCTAG GGAGGCGTTACTGAAATAATTTAGCTAT AGTAAGAAAATTTGTAATTTTAGAAATG CCAAGCATTCTAAATTAATTGCTTGAAA GTCACTATGATTGTGTCCATTATAAGGA GACAAATTCATTCAAGCAAGTTATTTAA TGTTAAAGGCCCAATTGTTAGGCAGTTA ATGGCACTTTTACTATTAACTAATCTTTC CATTTGTTCAGACGTAGCTTAACTTACCT CTTAGGTGTGAATTTGGTTAAGGTCCTC ATAATGTCTTTATGTGCAGTTTTTGATAG GTTATTGTCATAGAACTTATTCTATTCCT ACATTTATGATTACTATGGATGTATGAG AATAACACCTAATCCTTATACTTTACCTC AATTTAACTCCTTTATAAAGAACTTACAT TACAGAATAAAGATTTTTTAAAAATATA TTTTTTTGTAGAGACAGGGTCTTAGCCCA GCCGAGGCTGGTCTCTAAGTCCTGGCCC AAGCGATCCTCCTGCCTGGGCCTCCTAA AGTGCTGGAATTATAGACATGAGCCATC ACATCCAATATACAGAATAAAGATTTTT AATGGAGGATTTAATGTTCTTCAGAAAA TTTTCTTGAGGTCAGACAATGTCAAATG TCTCCTCAGTTTACACTGAGATTTTGAAA ACAAGTCTGAGCTATAGGTCCTTGTGAA GGGTCCATTGGAAATACTTGTTCAAAGT AAAATGGAAAGCAAAGGTAAAATCAGC AGTTGAAATTCAGAGAAAGACAGAAAA GGAGAAAAGATGAAATTCAACAGGACA GAAGGGAAATATATTATCATTAAGGAGG

[1221]

		ACAGTATCTGTAGAGCTCATTAGTGATG GCAAAATGACTTGGTCAGGATTATTTTT AACCCGCTTGTTTCTGGTTTGCACGGCTG GGGATGCAGCTAGGGTTCTGCCTCAGGG AGCACAGCTGTCCAGAGCAGCTGTCAGC CTGCAAGCCTGAAACACTCCCTCGGTAA AGTCCTTCCTACTCAGGACAGAAATGAC GAGAACAGGGAGCTGGAAACAGGCCCC TAACCAGAGAAGGGAAGTAATGGATCA ACAAAGTTAACTAGCAGGTCAGGATCAC GCAATTCATTTCACTCTGACTGGTAACAT GTGACAGAAACAGTGTAGGCTTATTGTA TTTTCATGTAGAGTAGGACCCAAAAATC CACCCAAAGTCCTTTATCTATGCCACATC CTTCTTATCTATACTTCCAGGACACTTTT TCTTCCTTATGATAAGGCTCTCTCTCTCT CCACACACACACACACACACACACACAC ACACACACACACACACACAAACACACA CCCCGCCAACCAAGGTGCATGTAAAAAG ATGTAGATTCCTCTGCCTTTCTCATCTAC ACAGCCCAGGAGGGTAAGTTAATATAAG AGGGATTTATTGGTAAGAGATGATGCTT AATCTGTTTAACACTGGGCCTCAAAGAG AGAATTTCTTTTCTTCTGTACTTATTAAG CACCTATTATGTGTTGAGCTTATATATAC AAAGGGTTATTATATGCTAATATAGTAA TAGTAATGGTGGTTGGTACTATGGTAAT TACCATAAAAATTATTATCCTTTTAAAAT AAAGCTAATTATTATTGGATCTTTTTTAG TATTCATTTTATGTTTTTTATGTTTTTGAT TTTTTAAAAGACAATCTCACCCTGTTACC CAGGCTGGAGTGCAGTGGTGCAATCATA GCTTTCTGCAGTCTTGAACTCCTGGGCTC AAGCAATCCTCCTGCCTTGGCCTCCCAA AGTGTTGGGATACAGTCATGAGCCACTG CATCTGGCCTAGGATCCATTTAGATTAA AATATGCATTTTAAATTTTAAAATAATAT GGCTAATTTTTACCTTATGTAATGTGTAT ACTGGCAATAAATCTAGTTTGCTGCCTA AAGTTTAAAGTGCTTTCCAGTAAGCTTC ATGTACGTGAGGGGAGACATTTAAAGTG AAACAGACAGCCAGGTGTGGTGGCTCAC GCCTGTAATCCCAGCACTCTGGGAGGCT

[1222]

		GAGGTGGGTGGATCGCTTGAGCCCTGGA GTTCAAGACCAGCCTGAGCAACATGGCA AAACGCTGTTTCTATAACAAAAATTAGC CGGGCATGGTGGCATGTGCCTGTGGTCC CAGCTACTAGGGGGCTGAGGCAGGAGA ATCGTTGGAGCCCAGGAGGTCAAGGCTG CACTGAGCAGTGCTTGCGCCACTGCACT CCAGCCTGGGTGACAGGACCAGACCTTG CCTCAAAAAAATAAGAAGAAAAATTAA AAATAAATGGAAACAACTACAAAGAGC TGTTGTCCTAGATGAGCTACTTAGTTAG GCTGATATTTTGGTATTTAACTTTTAAAG TCAGGGTCTGTCACCTGCACTACATTATT AAAATATCAATTCTCAATGTATATCCAC ACAAAGACTGGTACGTGAATGTTCATAG TACCTTTATTCACAAAACCCCAAAGTAG AGACTATCCAAATATCCATCAACAAGTG AACAAATAAACAAAATGTGCTATATCCA TGCAATGGAATACCACCCTGCAGTACAA AGAAGCTACTTGGGGATGAATCCCAAAG TCATGACGCTAAATGAAAGAGTCAGACA TGAAGGAGGAGATAATGTATGCCATACG AAATTCTAGAAAATGAAAGTAACTTATA GTTACAGAAAGCAAATCAGGGCAGGCA TAGAGGCTCACACCTGTAATCCCAGCAC TTTGAGAGGCCACGTGGGAAGATTGCTA GAACTCAGGAGTTCAAGACCAGCCTGGG CAACACAGTGAAACTCCATTCTCCACAA AAATGGGAAAAAAAGAAAGCAAATCAG TGGTTGTCCTGTGGGGAGGGGAAGGACT GCAAAGAGGGAAGAAGCTCTGGTGGGG TGAGGGTGGTGATTCAGGTTCTGTATCC TGACTGTGGTAGCAGTTTGGGGTGTTTA CATCCAAAAATATTCGTAGAATTATGCA TCTTAAATGGGTGGAGTTTACTGTATGT AAATTATACCTCAATGTAAGAAAAAATA ATGTGTAAGAAAACTTTCAATTCTCTTGC CAGCAAACGTTATTCAAATTCCTGAGCC CTTTACTTCGCAAATTCTCTGCACTTCTG CCCCGTACCATTAGGTGACAGCACTAGC TCCACAAATTGGATAAATGCATTTCTGG AAAAGACTAGGGACAAAATCCAGGCAT CACTTGTGCTTTCATATCAACCATGCTGT

[1223]

		ACAGCTTGTGTTGCTGTCTGCAGCTGCA ATGGGGACTCTTGATTTCTTTAAGGAAA CTTGGGTTACCAGAGTATTTCCACAAAT GCTATTCAAATTAGTGCTTATGATATGC AAGACACTGTGCTAGGAGCCAGAAAAC AAAGAGGAGGAGAAATCAGTCATTATGT GGGAACAACATAGCAAGATATTTAGATC ATTTTGACTAGTTAAAAAAGCAGCAGAG TACAAAATCACACATGCAATCAGTATAA TCCAAATCATGTAAATATGTGCCTGTAG AAAGACTAGAGGAATAAACACAAGAAT CTTAACAGTCATTGTCATTAGACACTAA GTCTAATTATTATTATTAGACACTATGAT ATTTGAGATTTAAAAAATCTTTAATATTT TAAAATTTAGAGCTCTTCTATTTTTCCAT AGTATTCAAGTTTGACAATGATCAAGTA TTACTCTTTCTTTTTTTTTTTTTTTTTTT TTTTGAGATGGAGTTTTGGTCTTGTTGCC CATGCTGGAGTGGAATGGCATGACCATA GCTCACTGCAACCTCCACCTCCTGGGTTC AAGCAAAGCTGTCGCCTCAGCCTCCCGG GTAGATGGGATTACAGGCGCCCACCACC ACACTCGGCTAATGTTTGTATTTTTAGTA GAGATGGGGTTTCACCATGTTGGCCAGG CTGGTCTCAAACTCCTGACCTCAGAGGA TCCACCTGCCTCAGCCTCCCAAAGTGCT GGGATTACAGATGTAGGCCACTGCGCCC GGCCAAGTATTGCTCTTATACATTAAAA AACAGGTGTGAGCCACTGCGCCCAGCCA GGTATTGCTCTTATACATTAAAAAATAG GCCGGTGCAGTGGCTCACGCCTGTAATC CCAGCACTTTGGGAAGCCAAGGCGGGCA GAACACCCGAGGTCAGGAGTCCAAGGC CAGCCTGGCCAAGATGGTGAAACCCCGT CTCTATTAAAAATACAAACATTACCTGG GCATGATGGTGGGCGCCTGTAATCCCAG CTACTCAGGAGGCTGAGGCAGGAGGATC CGCGGAGCCTGGCAGATCTGCCTGAGCC TGGGAGGTTGAGGCTACAGTAAGCCAAG ATCATGCCAGTATACTTCAGCCTGGGCG ACAAAGTGAGACCGTAACAAAAAAAAA AAAATTTAAAAAAAGAAATTTAGATCAA GATCCAACTGTAAAAAGTGGCCTAAACA

[1224]

		CCACATTAAAGAGTTTGGAGTTTATTCT GCAGGCAGAAGAGAACCATCAGGGGGT CTTCAGCATGGGAATGGCATGGTGCACC TGGTTTTTGTGAGATCATGGTGGTGACA GTGTGGGGAATGTTATTTTGGAGGGACT GGAGGCAGACAGACCGGTTAAAAGGCC AGCACAACAGATAAGGAGGAAGAAGAT GAGGGCTTGGACCGAAGCAGAGAAGAG CAAACAGGGAAGGTACAAATTCAAGAA ATATTGGGGGGTTTGAATCAACACATTT AGATGATTAATTAAATATGAGGACTGAG GAATAAGAAATGAGTCAAGGATGGTTCC AGGCTGCTAGGCTGCTTACCTGAGGTGG CAAAGTCGGGAGGAGTGGCAGTTTAGG ACAGGGGGCAGTTGAGGAATATTGTTTT GATCATTTTGAGTTTGAGGTACAAGTTG GACACTTAGGTAAAGACTGGAGGGGAA ATCTGAATATACAATTATGGGACTGAGG AACAAGTTTATTTTATTTTTTGTTTCGTTT TCTTGTTGAAGAACAAATTTAATTGTAA TCCCAAGTCATCAGCATCTAGAAGACAG TGGCAGGAGGTGACTGTCTTGTGGGTAA GGGTTTGGGGTCCTTGATGAGTATCTCTC AATTGGCCTTAAATATAAGCAGGAAAAG GAGTTTATGATGGATTCCAGGCTCAGCA GGGCTCAGGAGGGCTCAGGCAGCCAGC AGAGGAAGTCAGAGCATCTTCTTTGGTT TAGCCCAAGTAATGACTTCCTTAAAAAG CTGAAGGAAAATCCAGAGTGACCAGATT ATAAACTGTACTCTTGCATTTTCTCTCCC TCCTCTCACCCACAGCCTCTTGATGAACC GGAGGAAGTTTCTTTACCAATTCAAAAA TGTCCGCTGGGCTAAGGGTCGGCGTGAG ACCTACCTGTGCTACGTAGTGAAGAGGC GTGACAGTGCTACATCCTTTTCACTGGA CTTTGGTTATCTTCGCAATAAGGTATCAA TTAAAGTCGGCTTTGCAAGCAGTTTAAT GGTCAACTGTGAGTGCTTTTAGAGCCAC CTGCTGATGGTATTACTTCCATCCTTTTT TGGCATTTGTGTCTCTATCACATTCCTCA AATCCTTTTTTTTATTTCTTTTTCCATGTC CATGCACCCATATTAGACATGGCCCAAA ATATGTGATTTAATTCCTCCCCAGTAATG

[1225]

		CTGGGCACCCTAATACCACTCCTTCCTTC AGTGCCAAGAACAACTGCTCCCAAACTG TTTACCAGCTTTCCTCAGCATCTGAATTG CCTTTGAGATTAATTAAGCTAAAAGCAT TTTTATATGGGAGAATATTATCAGCTTGT CCAAGCAAAAATTTTAAATGTGAAAAAC AAATTGTGTCTTAAGCATTTTTGAAAATT AAGGAAGAAGAATTTGGGAAAAAATTA ACGGTGGCTCAATTCTGTCTTCCAAATG ATTTCTTTTCCCTCCTACTCACATGGGTC GTAGGCCAGTGAATACATTCAACATGGT GATCCCCAGAAAACTCAGAGAAGCCTCG GCTGATGATTAATTAAATTGATCTTTCGG CTACCCGAGAGAATTACATTTCCAAGAG ACTTCTTCACCAAAATCCAGATGGGTTT ACATAAACTTCTGCCCACGGGTATCTCC TCTCTCCTAACACGCTGTGACGTCTGGG CTTGGTGGAATCTCAGGGAAGCATCCGT GGGGTGGAAGGTCATCGTCTGGCTCGTT GTTTGATGGTTATATTACCATGCAATTTT CTTTGCCTACATTTGTATTGAATACATCC CAATCTCCTTCCTATTCGGTGACATGACA CATTCTATTTCAGAAGGCTTTGATTTTAT CAAGCACTTTCATTTACTTCTCATGGCAG TGCCTATTACTTCTCTTACAATACCCATC TGTCTGCTTTACCAAAATCTATTTCCCCT TTTCAGATCCTCCCAAATGGTCCTCATAA ACTGTCCTGCCTCCACCTAGTGGTCCAG GTATATTTCCACAATGTTACATCAACAG GCACTTCTAGCCATTTTCCTTCTCAAAAG GTGCAAAAAGCAACTTCATAAACACAAA TTAAATCTTCGGTGAGGTAGTGTGATGC TGCTTCCTCCCAACTCAGCGCACTTCGTC TTCCTCATTCCACAAAAACCCATAGCCTT CCTTCACTCTGCAGGACTAGTGCTGCCA AGGGTTCAGCTCTACCTACTGGTGTGCT CTTTTGAGCAAGTTGCTTAGCCTCTCTGT AACACAAGGACAATAGCTGCAAGCATCC CCAAAGATCATTGCAGGAGACAATGACT AAGGCTACCAGAGCCGCAATAAAAGTC AGTGAATTTTAGCGTGGTCCTCTCTGTCT CTCCAGAACGGCTGCCACGTGGAATTGC TCTTCCTCCGCTACATCTCGGACTGGGAC

[1226]

		CTAGACCCTGGCCGCTGCTACCGCGTCA CCTGGTTCACCTCCTGGAGCCCCTGCTAC GACTGTGCCCGACATGTGGCCGACTTTC TGCGAGGGAACCCCAACCTCAGTCTGAG GATCTTCACCGCGCGCCTCTACTTCTGTG AGGACCGCAAGGCTGAGCCCGAGGGGC TGCGGCGGCTGCACCGCGCCGGGGTGCA AATAGCCATCATGACCTTCAAAGGTGCG AAAGGGCCTTCCGCGCAGGCGCAGTGCA GCAGCCCGCATTCGGGATTGCGATGCGG AATGAATGAGTTAGTGGGGAAGCTCGAG GGGAAGAAGTGGGCGGGGATTCTGGTTC ACCTCTGGAGCCGAAATTAAAGATTAGA AGCAGAGAAAAGAGTGAATGGCTCAGA GACAAGGCCCCGAGGAAATGAGAAAAT GGGGCCAGGGTTGCTTCTTTCCCCTCGAT TTGGAACCTGAACTGTCTTCTACCCCCAT ATCCCCGCCTTTTTTTCCTTTTTTTTTTT TGAAGATTATTTTTACTGCTGGAATACTT TTGTAGAAAACCACGAAAGAACTTTCAA AGCCTGGGAAGGGCTGCATGAAAATTCA GTTCGTCTCTCCAGACAGCTTCGGCGCA TCCTTTTGGTAAGGGGCTTCCTCGCTTTT TAAATTTTCTTTCTTTCTCTACAGTCTTTT TTGGAGTTTCGTATATTTCTTATATTTTC TTATTGTTCAATCACTCTCAGTTTTCATC TGATGAAAACTTTATTTCTCCTCCACATC AGCTTTTTCTTCTGCTGTTTCACCATTCA GAGCCCTCTGCTAAGGTTCCTTTTCCCTC CCTTTTCTTTCTTTTGTTGTTTCACATCTT TAAATTTCTGTCTCTCCCCAGGGTTGCGT TTCCTTCCTGGTCAGAATTCTTTTCTCCT TTTTTTTTTTTTTTTTTTTTTTTTTTAAACA AACAAACAAAAAACCCAAAAAAACTCT TTCCCAATTTACTTTCTTCCAACATGTTA CAAAGCCATCCACTCAGTTTAGAAGACT CTCCGGCCCCACCGACCCCCAACCTCGT TTTGAAGCCATTCACTCAATTTGCTTCTC TCTTTCTCTACAGCCCCTGTATGAGGTTG ATGACTTACGAGACGCATTTCGTACTTT GGGACTTTGATAGCAACTTCCAGGAATG TCACACACGATGAAATATCTCTGCTGAA GACAGTGGATAAAAAACAGTCCTTCAAG

[1227]

		TCTTCTCTGTTTTTATTCTTCAACTCTCAC TTTCTTAGAGTTTACAGAAAAAATATTT ATATACGACTCTTTAAAAAGATCTATGT CTTGAAAATAGAGAAGGAACACAGGTCT GGCCAGGGACGTGCTGCAATTGGTGCAG TTTTGAATGCAACATTGTCCCCTACTGGG AATAACAGAACTGCAGGACCTGGGAGC ATCCTAAAGTGTCAACGTTTTTCTATGAC TTTTAGGTAGGATGAGAGCAGAAGGTAG ATCCTAAAAAGCATGGTGAGAGGATCAA ATGTTTTTATATCAACATCCTTTATTATT TGATTCATTTGAGTTAACAGTGGTGTTA GTGATAGATTTTTCTATTCTTTTCCCTTG ACGTTTACTTTCAAGTAACACAAACTCTT CCATCAGGCCATGATCTATAGGACCTCC TAATGAGAGTATCTGGGTGATTGTGACC CCAAACCATCTCTCCAAAGCATTAATAT CCAATCATGCGCTGTATGTTTTAATCAGC AGAAGCATGTTTTTATGTTTGTACAAAA GAAGATTGTTATGGGTGGGGATGGAGGT ATAGACCATGCATGGTCACCTTCAAGCT ACTTTAATAAAGGATCTTAAAATGGGCA GGAGGACTGTGAACAAGACACCCTAATA ATGGGTTGATGTCTGAAGTAGCAAATCT TCTGGAAACGCAAACTCTTTTAAGGAAG TCCCTAATTTAGAAACACCCACAAACTT CACATATCATAATTAGCAAACAATTGGA AGGAAGTTGCTTGAATGTTGGGGAGAGG AAAATCTATTGGCTCTCGTGGGTCTCTTC ATCTCAGAAATGCCAATCAGGTCAAGGT TTGCTACATTTTGTATGTGTGTGATGCTT CTCCCAAAGGTATATTAACTATATAAGA GAGTTGTGACAAAACAGAATGATAAAG CTGCGAACCGTGGCACACGCTCATAGTT CTAGCTGCTTGGGAGGTTGAGGAGGGAG GATGGCTTGAACACAGGTGTTCAAGGCC AGCCTGGGCAACATAACAAGATCCTGTC TCTCAAAAAAAAAAAAAAAAAAAAGAA AGAGAGAGGGCCGGGCGTGGTGGCTCA CGCCTGTAATCCCAGCACTTTGGGAGGC CGAGCCGGGCGGATCACCTGTGGTCAGG AGTTTGAGACCAGCCTGGCCAACATGGC AAAACCCCGTCTGTACTCAAAATGCAAA

[1228]

		AATTAGCCAGGCGTGGTAGCAGGCACCT GTAATCCCAGCTACTTGGGAGGCTGAGG CAGGAGAATCGCTTGAACCCAGGAGGTG GAGGTTGCAGTAAGCTGAGATCGTGCCG TTGCACTCCAGCCTGGGCGACAAGAGCA AGACTCTGTCTCAGAAAAAAAAAAAAA AAAGAGAGAGAGAGAGAAAGAGAACAA TATTTGGGAGAGAAGGATGGGGAAGCA TTGCAAGGAAATTGTGCTTTATCCAACA AAATGTAAGGAGCCAATAAGGGATCCCT ATTTGTCTCTTTTGGTGTCTATTTGTCCCT AACAACTGTCTTTGACAGTGAGAAAAAT ATTCAGAATAACCATATCCCTGTGCCGT TATTACCTAGCAACCCTTGCAATGAAGA TGAGCAGATCCACAGGAAAACTTGAATG CACAACTGTCTTATTTTAATCTTATTGTA CATAAGTTTGTAAAAGAGTTAAAAATTG TTACTTCATGTATTCATTTATATTTTATA TTATTTTGCGTCTAATGATTTTTTATTAA CATGATTTCCTTTTCTGATATATTGAAAT GGAGTCTCAAAGCTTCATAAATTTATAA CTTTAGAAATGATTCTAATAACAACGTA TGTAATTGTAACATTGCAGTAATGGTGC TACGAAGCCATTTCTCTTGATTTTTAGTA AACTTTTATGACAGCAAATTTGCTTCTGG CTCACTTTCAATCAGTTAAATAAATGAT AAATAATTTTGGAAGCTGTGAAGATAAA ATACCAAATAAAATAATATAAAAGTGAT TTATATGAAGTTAAAATAAAAAATCAGT ATGATGGAATAAACTTG
犬 AID (clAID)多肽序列	1374	MDSLLMKQRKFLYHFKNVRWAKGRHET YLCYVVKRRDSATSFSLDFGHLRNKSGCH VELLFLRYISDWDLDPGRCYRVTWFTSWS PCYDCARHVADFLRGYPNLSLRIFAARLY FCEDRKAEPEGLRRLHRAGVQIAIMTFKD YFYCWNTFVENREKTFKAWEGLHENSVR LSRQLRRILLPLYEVDDLRDAFRTLGL
牛 AID (btAID)多肽序列	1375	MDSLLKKQRQFLYQFKNVRWAKGRHETY LCYVVKRRDSPTSFSLDFGHLRNKAGCHV ELLFLRYISDWDLDPGRCYRVTWFTSWSP CYDCARHVADFLRGYPNLSLRIFTARLYF CDKERKAEPEGLRRLHRAGVQIAIMTFKD YFYCWNTFVENHERTFKAWEGLHENSVR

[1229]

		LSRQLRRILLPLYEVDDLRDAFRTLGL
大鼠 AID 多肽序列	1376	MAVGSKPKAALVGPHWERERIWCFLCST GLGTQQTGQTSRWLRPAATQDPVSPPRSL LMKQRKFLYHFKNVRWAKGRHETYLCY VVKRRDSATSFSLDFGYLRNKSGCHVELL FLRYISDWDLDPGRCYRVTWFTSWSPCYD CARHVADFLRGNPNLSLRIFTARLTGWGA LPAGLMSPARPSDYFYCWNTFVENHERTF KAWEGLHENSVRLSRRLRRILLPLYEVDD LRDAFRTLGL
小鼠 (mAID) AID 多肽序列	1377	MDSLLMNRRKFLYQFKNVRWAKGRRETY LCYVVKRRDSATSFSLDFGYLRNKNGCHV ELLFLRYISDWDLDPGRCYRVTWFTSWSP CYDCARHVADFLRGNPNLSLRIFTARLYF CEDRKAEPEGLRRLHRAGVQIAIMTFKDY FYCWNTFVENHERTFKAWEGLHENSVRL SRQLRRILLPLYEVDDLRDAFRTLGL
rAPOBEC-1 多肽序列	1378	MSSETGPVAVDPTLRRRIEPHEFEVFFDPR ELRKETCLLYEINWGGRHSIWRHTSQNTN KHVEVNFIEKFTTERYFCPNTRCSITWFLS WSPCGECSRAITEFLSRYPHVTLFIYIARLY HHADPRNRQGLRDLISSGVTIQIMTEQESG YCWRNFVNYSPSNEAHWPRYPHLWVRLY VLELYCIILGLPPCLNILRRKQPQLTFFTIAL QSCHYQRLPPHILWATGLK
maAPOBEC-1 多肽序列	1379	MSSETGPVVVDPTLRRRIEPHEFDAFFDQG ELRKETCLLYEIRWGGRHNIWRHTGQNTS RHVEINFIEKFTSERYFYPSTRCSIVWFLSW SPCGECSKAITEFLSGHPNVTLFIYAARLY HHTDQRNRQGLRDLISRGVTIRIMTEQEYC YCWRNFVNYPPSNEVYWPRYPNLWMRL YALELYCIHLGLPPCLKIKRRHQYPLTFFR LNLQSCHYQRIPPHILWATGFI
ppAPOBEC-1 多肽序列	1380	MTSEKGPSTGDPTLRRRIESWEFDVFYDPR ELRKETCLLYEIKWGMSRKIWRSSGKNTT NHVEVNFIKKFTSERRFHSSISCSITWFLSW SPCWECSQAIREFLSQHPGVTLVIYVARLF WHMDQRNRQGLRDLVNSGVTIQIMRASE YYHCWRNFVNYPPGDEAHWPQYPPLWM MLYALELHCIILSLPPCLKISRRWQNHLAF FRLHLQNCHYQTIPPHILLATGLIHPSVTW R
ocAPOB	1381	MASEKGPSNKDYTLRRRIEPWEFEVFFDP

[1230]

EC1 多肽序列		QELRKEACLLYEIKWGASSKTWRSSGKNT TNHVEVNFLEKLTSEGRLGPSTCCSITWFL SWSPCWECSMAIREFLSQHPGVTLIIFVAR LFQHMDRRNRQGLKDLVTSGVTVRVMSV SEYCYCWENFVNYPPGKAAQWPRYPPRW MLMYALELYCIILGLPPCLKISRRHQKQLT FFSLTPQYCHYKMIPPYILLATGLLQPSVP WR
mdAPOBEC-1 多肽序列	1382	MNSKTGPSVGDATLRRRIKPWEFVAFFNP QELRKETCLLYEIKWGNQNIWRHSNQNTS QHAEINFMEKFTAERHFNSSVRCSITWFLS WSPCWECSKAIRKFLDHYPNVTLAIFISRL YWHMDQQHRQGLKELVHSGVTIQIMSYS EYHYCWRNFVDYPQGEEDYWPKYPYLWI MLYVLELHCIILGLPPCLKISGSHSNQLALF SLDLQDCHYQKIPYNVLVATGLVQPFVTW R
ppAPOBEC-2 多肽序列	1383	MAQKEEAAAATEAASQNGEDLENLDDPE KLKELIELPPFEIVTGERLPANFFKFQFRNV EYSSGRNKTFLCYVVEAQGKGGQVQASR GYLEDEHAAAHAEEAFFNTILPAFDPALR YNVTWYVSSSPCAACADRIIKTLSKTKNL RLLILVGRLFMWEELEIQDALKKLKEAGC KLRIMKPQDFEYVWQNFVEQEEGESKAFQ PWEDIQENFLYYEEKLADILK
btAPOBEC-2 多肽序列	1384	MAQKEEAAAAAEPASQNGEEVENLEDPE KLKELIELPPFEIVTGERLPAHYFKFQFRNV EYSSGRNKTFLCYVVEAQSKGGQVQASR GYLEDEHATNHAEEAFFNSIMPTFDPALR YMVTWYVSSSPCAACADRIVKTLNKTKN LRLLILVGRLFMWEEPEIQAALRKLKEAG CRLRIMKPQDFEYIWQNFVEQEEGESKAF EPWEDIQENFLYYEEKLADILK
mAPOBEC-3-(1)多肽序列	1385	MQPQRLGPRAGMGPFCLGCSHRKCYSPIR NLISQETFKFHFKNLGYAKGRKDTFLCYE VTRKDCDSPVSLHHGVFKNKDNIHAEICFL YWFHDKVLKVLSPREEFKITWYMSWSPCF ECAEQIVRFLATHHNLSLDIFSSRLYNVQD PETQQNLCRLVQEGAQVAAMDLYEFKKC WKKFVDNGGRRFRPWKRLLTNFRYQDSK LQEILRPCYISVPSSSSSTLSNICLTKGLPET RFWVEGRRMDPLSEEEFYSQFYNQRVKHL CYYHRMKPYLCYQLEQFNGQAPLKGCLL

			SEKGKQHAEILFLDKIRSMELSQVTITCYL TWSPCPNCAWQLAAFKRDRPDLILHIYTS RLYFHWKRPFQKGLCSLWQSGILVDVMD LPQFTDCWTNFVNPKRPFWPWKGLEIISRR TQRRLRRIKESWGLQDLVNDFGNLQLGPP MS
	APOBEC-3-(2) (小鼠 APOBEC-3)多肽序列	1386	MGPFCLGCSHRKCYSPIRNLISQETFKFHF KNLGYAKGRKDTFLCYEVTRKDCDSPVSL HHGVFKNKDNIHAEICFLYWFHDKVLKVL SPREEFKITWYMSWSPCFECAEQIVRFLAT HHNLSLDIFSSRLYNVQDPETQQNLCRLV QEGAQVAAMDLYEFKKCWKKFVDNGGR RFRPWKRLLTNFRYQDSKLQEILRPCYIPV PSSSSSTLSNICLTKGLPETRFCVEGRRMDP LSEEEFYSQFYNQRVKHLCYYHRMKPYLC YQLEQFNGQAPLKGCLLSEKGKQHAEILF LDKIRSMELSQVTITCYLTWSPCPNCAWQ LAAFKRDRPDLILHIYTSRLYFHWKRPFQK GLCSLWQSGILVDVMDLPQFTDCWTNFV NPKRPFWPWKGLEIISRRTQRRLRRIKESW GLQDLVNDFGNLQLGPPMS
[1231]	APOBEC-3 多肽序列	1387	MGPFCLGCSHRKCYSPIRNLISQETFKFHF KNRLRYAIDRKDTFLCYEVTRKDCDSPVS LHHGVFKNKDNIHAEICFLYWFHDKVLKV LSPREEFKITWYMSWSPCFECAEQVLRFLA THHNLSLDIFSSRLYNIRDPENQQNLCRLV QEGAQVAAMDLYEFKKCWKKFVDNGGR RFRPWKKLLTNFRYQDSKLQEILRPCYIPV PSSSSSTLSNICLTKGLPETRFCVERRRVHL LSEEEFYSQFYNQRVKHLCYYHGVKPYLC YQLEQFNGQAPLKGCLLSEKGKQHAEILF LDKIRSMELSQVIITCYLTWSPCPNCAWQL AAFKRDRPDLILHIYTSRLYFHWKRPFQKG LCSLWQSGILVDVMDLPQFTDCWTNFVNP KRPFWPWKGLEIISRRTQRRLHRIKESWGL QDLVNDFGNLQLGPPMS
	hAPOBEC-3A 多肽序列	1388	MEASPASGPRHLMDPHIFTSNFNNGIGRHK TYLCYEVERLDNGTSVKMDQHRGFLHNQ AKNLLCGFYGRHAELRFLDLVPSLQLDPA QIYRVTWFISWSPCFSWGCAGEVRAFLQE NTHVRLRIFAARIYDYDPLYKEALQMLRD AGAQVSIMTYDEFKHCWDTFVDHQGCPF QPWDGLDEHSQALSGRLRAILQNQGN

[1232]

hAPOBEC-3F 多肽序列	1389	MKPHFRNTVERMYRDTFSYNFYNRPILSR RNTVWLCYEVKTKGPSRPRLDAKIFRGQV YSQPEHHAEMCFLSWFCGNQLPAYKCFQI TWFVSWTPCPDCVAKLAEFLAEHPNVTLT ISAARLYYYWERDYRRALCRLSQAGARV KIMDDEEFAYCWENFVYSEGQPFMPWYK FDDNYAFLHRTLKEILRNPMEAMYPHIFYF HFKNLRKAYGRNESWLCFTMEVVKHHSP VSWKRGVFRNQVDPETHCHAERCFLSWF CDDILSPNTNYEVTWYTSWSPCPECAGEV AEFLARHSNVNLTIFTARLYYFWDTDYQE GLRSLSQEGASVEIMGYKDFKYCWENFVY NDDEPFKPWKGLKYNFLFLDSKLQEILE
恒河猴 APOBEC-3G 多肽序列	1390	MVEPMDPRTFVSNFNNRPILSGLNTVWLC CEVKTKDPSGPPLDAKIFQGKVYSKAKYH PEMRFLRWFHKWRQLHHDQEYKVTWYV SWSPCTRCANSVATFLAKDPKVTLTIFVAR LYYFWKPDYQQALRILCQKRGGPHATMKI MNYNEFQDCWNKFVDGRGKPFKPRNNLP KHYTLLQATLGELLRHLMDPGTFTSNFNN KPWVSGQHETYLCYKVERLHNDTWVPLN QHRGFLRNQAPNIHGFPKGRHAELCFLDLI PFWKLDGQQYRVTCFTSWSPCFSCAQEM AKFISNNEHVSLCIFAARIYDDQGRYQEGL RALHRDGAKIAMMNYSEFEYCWDTFVDR QGRPFQPWDGLDEHSQALSGRLRAI
黑猩猩 APOBEC-3G 多肽序列	1391	MKPHFRNPVERMYQDTFSDNFYNRPILSH RNTVWLCYEVKTKGPSRPPLDAKIFRGQV YSKLKYHPEMRFFHWFSKWRKLHRDQEY EVTWYISWSPCTKCTRDVATFLAEDPKVT LTIFVARLYYFWDPDYQEALRSLCQKRDG PRATMKIMNYDEFQHCWSKFVYSQRELFE PWNNLPKYYILLHIMLGEILRHSMDPPTFT SNFNNELWVRGRHETYLCYEVERLHNDT WVLLNQRRGFLCNQAPHKHGFLEGRHAE LCFLDVIPFWKLDLHQDYRVTCFTSWSPC FSCAQEMAKFISNNKHVSLCIFAARIYDDQ GRCQEGLRTLAKAGAKISIMTYSEFKHCW DTFVDHQGCPFQPWDGLEEHSQALSGRLR AILQNQGN
绿猴 APOBEC	1392	MNPQIRNMVEQMEPDIFVYYFNNRPILSG RNTVWLCYEVKTKDPSGPPLDANIFQGKL YPEAKDHPEMKFLHWFRKWRQLHRDQEY

[1233]

-3G 多肽序列		EVTWYVSWSPCTRCANSVATFLAEDPKVT LTIFVARLYYFWKPDYQQALRILCQERGG PHATMKIMNYNEFQHCWNEFVDGQGKPF KPRKNLPKHYTLLHATLGELLRHVMDPGT FTSNFNNKPWVSGQRETYLCYKVERSHND TWVLLNQHRGFLRNQAPDRHGFPKGRHA ELCFLDLIPFWKLDDQQYRVTCFTSWSPCF SCAQKMAKFISNNKHVSLCIFAARIYDDQ GRCQEGLRTLHRDGAKIAVMNYSEFEYC WDTFVDRQGRPFQPWDGLDEHSQALSGR LRAI
人 APOBEC -3G 多肽序列	1393	MKPHFRNTVERMYRDTFSYNFYNRPILSR RNTVWLCYEVKTKGPSRPPLDAKIFRGQV YSELKYHPEMRFFHWFSKWRKLHRDQEY EVTWYISWSPCTKCTRDMATFLAEDPKVT LTIFVARLYYFWDPDYQEALRSLCQKRDG PRATMKIMNYDEFQHCWSKFVYSQRELFE PWNNLPKYYILLHIMLGEILRHSMDPPTFT FNFNNEPWVRGRHETYLCYEVERMHNDT WVLLNQRRGFLCNQAPHKHGFLEGRHAE LCFLDVIPFWKLDLDQDYRVTCFTSWSPC FSCAQEMAKFISKNKHVSLCIFTARIYDDQ GRCQEGLRTLAEAGAKISIMTYSEFKHCW DTFVDHQGCPFQPWDGLDEHSQDLSGRLR AILQNQEN
人 APOBEC -3B 多肽序列	1394	MNPQIRNPMERMYRDTFYDNFENEPILYG RSYTWLCYEVKIKRGRSNLLWDTGVFRG QVYFKPQYHAEMCFLSWFCGNQLPAYKC FQITWFVSWTPCPDCVAKLAEFLSEHPNV TLTISAARLYYYWERDYRRALCRLSQAGA RVTIMDYEEFAYCWENFVYNEGQQFMPW YKFDENYAFLHRTLKEILRYLMDPDTFTF NFNNDPLVLRRRQTYLCYEVERLDNGTW VLMDQHMGFLCNEAKNLLCGFYGRHAEL RFLDLVPSLQLDPAQIYRVTWFISWSPCFS WGCAGEVRAFLQENTHVRLRIFAARIYDY DPLYKEALQMLRDAGAQVSIMTYDEFEY CWDTFVYRQGCPFQPWDGLEEHSQALSG RLRAILQNQGN
大鼠 APOBEC -3B 多肽	1395	MQPQGLGPNAGMGPVCLGCSHRRPYSPIR NPLKKLYQQTFYFHFKNVRYAWGRKNNF LCYEVNGMDCALPVPLRQGVFRKQGHIH AELCFIYWFHDKVLRVLSPMEEFKVTWY

[1234]

序列		MSWSPCSKCAEQVARFLAAHRNLSLAIFS SRLYYYLRNPNYQQKLCRLIQEGVHVAA MDLPEFKKCWNKFVDNDGQPFRPWMRLR INFSFYDCKLQEIFSRMNLLREDVFYLQFN NSHRVKPVQNRYYRRKSYLCYQLERANG QEPLKGYLLYKKGEQHVEILFLEKMRSME LSQVRITCYLTWSPCPNCARQLAAFKKDH PDLILRIYTSRLYFWRKKFQKGLCTLWRSG IHVDVMDLPQFADCWTNFVNPQRPFRPW NELEKNSWRIQRRLRRIKESWGL
牛 APOBEC -3B 多肽序列	1396	DGWEVAFRSGTVLKAGVLGVSMTEGWA GSGHPGQGACVWTPGTRNTMNLLREVLF KQQFGNQPRVPAPYYRRKTYLCYQLKQR NDLTLDRGCFRNKKQRHAERFIDKINSLDL NPSQSYKIICYITWSPCPNCANELVNFITRN NHLKLEIFASRLYFHWIKSFKMGLQDLQN AGISVAVMTHTEFEDCWEQFVDNQSRPFQ PWDKLEQYSASIRRRLQRILTAPI
黑猩猩 APOBEC -3B 多肽序列	1397	MNPQIRNPMEWMYQRTFYYNFENEPILYG RSYTWLCYEVKIRRGHSNLLWDTGVFRG QMYSQPEHHAEMCFLSWFCGNQLSAYKC FQITWFVSWTPCPDCVAKLAKFLAEHPNV TLTISAARLYYYWERDYRRALCRLSQAGA RVKIMDDEEFAYCWENFVYNEGQPFMPW YKFDDNYAFLHRTLKEIIRHLMDPDTFTFN FNNDPLVLRRHQTYLCYEVERLDNGTWV LMDQHMGFLCNEAKNLLCGFYGRHAELR FLDLVPSLQLDPAQIYRVTWFISWSPCFSW GCAGQVRAFLQENTHVRLRIFAARIYDYD PLYKEALQMLRDAGAQVSIMTYDEFEYC WDTFVYRQGCPFQPWDGLEEHSQALSGR LRAILQVRASSLCMVPHRPPPPPQSPGPCLP LCSEPPLGSLLPTGRPAPSLPFLLTASFSFPP PASLPPLPSLSLSPGHLPVPSFHSLTSCSIQP PCSSRIRETEGWASVSKEGRDLG
人 APOBEC -3C 多肽序列	1398	MNPQIRNPMKAMYPGTFYFQFKNLWEAN DRNETWLCFTVEGIKRRSVVSWKTGVFRN QVDSETHCHAERCFLSWFCDDILSPNTKY QVTWYTSWSPCPDCAGEVAEFLARHSNV NLTIFTARLYYFQYPCYQEGLRSLSQEGVA VEIMDYEDFKYCWENFVYNDNEPFKPWK GLKTNFRLLKRRLRESLQ
大猩猩	1399	MNPQIRNPMKAMYPGTFYFQFKNLWEAN

APOBEC-3C 多肽序列		DRNETWLCFTVEGIKRRSVVSWKTGVFRNQVDSETHCHAERCFLSWECDDILSPNTNYQVTWYTSWSPCPECAGEVAEFLARHSNVNLTIFTARLYYFQDTDYQEGLRSLSQEGVAVKIMDYKDFKYCWENFVYNDDEPFKPWKGLKYNFRFLKRRLQEILE
恒河猴 APOBEC-3A 多肽序列	1400	MDGSPASRPRHLMDPNTFTFNFNNDLSVRGRHQTYLCYEVERLDNGTWVPMDERRGFLCNKAKNVPCGDYGCHVELRFLCEVPSWQLDPAQTYRVTWFISWSPCFRRGCAGQVRVFLQENKHVRLRIFAARIYDYDPLYQEALRTLRDAGAQVSIMTYEEFKHCWDTFVDRQGRPFQPWDGLDEHSQALSGRLRAILQNQGN
牛 APOBEC-3A 多肽序列	1401	MDEYTFTENFNNQGWPSKTYLCYEMERLDGDATIPLDEYKGFVRNKGLDQPEKPCHAELYFLGKIHSWNLDRNQHYRLTCFISWSPCYDCAQKLTTFLKENHHISLHILASRIYTHNRFGCHQSGLCELQAAGARITIMTFEDFKHCWETFVDHKGKPFQPWEGLNVKSQALCTELQAILKTQQN
人 APOBEC-3H 多肽序列	1402	MALLTAETFRLQFNNKRRLRRPYYPRKALLCYQLTPQNGSTPTRGYFENKKKCHAEICFINEIKSMGLDETQCYQVTCYLTWSPCSSCAWELVDFIKAHDHLNLGIFASRLYYHWCKPQQKGLRLLCGSQVPVEVMGFPKFADCWENFVDHEKPLSFNPYKMLEELDKNSRAIKRRLERIKIPGVRAQGRYMDILCDAEV
恒河猴 APOBEC-3H 多肽序列	1403	MALLTAKTFSLQFNNKRRVNKPYYPRKALLCYQLTPQNGSTPTRGHLKNKKKDHAEIRFINKIKSMGLDETQCYQVTCYLTWSPCPSCAGELVDFIKAHRHLNLRIFASRLYYHWRPNYQEGLLLLCGSQVPVEVMGLPEFTDCWENFVDHKEPPSFNPSEKLEELDKNSQAIKRRLERIKSRSVDVLENGLRSLQLGPVTPSSSIRNSR
人 APOBEC-3D 多肽序列	1404	MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGPVLPKRQSNHRQEVYFRFENHAEMCFLSWFCGNRLPANRRFQITWFVSWNPCLPCVVKVTKFLAEHPNVTLTISAARLYYYRDRDWRWVLLRLHKAGARVKIMDYEDFAYCWENFVCNEGQPFMPWYKFDDNYASLHRTLKEIL

[1235]

[1236]

		RNPMEAMYPHIFYFHFKNLLKACGRNES WLCFTMEVTKHHSAVFRKRGVFRNQVDP ETHCHAERCFLSWFCDDILSPNTNYEVTW YTSWSPCPECAGEVAEFLARHSNVNLTIFT ARLCYFWDTDYQEGLCSLSQEGASVKIMG YKDFVSCWKNFVYSDDEPFKPWKGLQTN FRLLKRRLREILQ
人 APOBEC -1 多肽序列	1405	MTSEKGPSTGDPTLRRRIEPWEFDVFYDPR ELRKEACLLYEIKWGMSRKIWRSSGKNTT NHVEVNFIKKFTSERDFHPSMSCSITWFLS WSPCWECSQAIREFLSRHPGVTLVIYVARL FWHMDQQNRQGLRDLVNSGVTIQIMRAS EYYHCWRNFVNYPPGDEAHWPQYPPLW MMLYALELHCIILSLPPCLKISRRWQNHLT FFRLHLQNCHYQTIPPHILLATGLIHPSVA WR
小鼠 APOBEC -1 多肽序列	1406	MSSETGPVAVDPTLRRRIEPHEFEVFFDPR ELRKETCLLYEINWGGRHSVWRHTSQNTS NHVEVNFLEKFTTERYFRPNTRCSITWFLS WSPCGECSRAITEFLSRHPYVTLFIYIARLY HHTDQRNRQGLRDLISSGVTIQIMTEQEYC YCWRNFVNYPPSNEAYWPRYPHLWVKLY VLELYCIILGLPPCLKILRRKQPQLTFFTITL QTCHYQRIPPHLLWATGLK
人 APOBEC -2 多肽序列	1407	MAQKEEAAVATEAASQNGEDLENLDDPE KLKELIELPPFEIVTGERLPANFFKFQFRNV EYSSGRNKTFLCYVVEAQGKGGQVQASR GYLEDEHAAAHAEEAFFNTILPAFDPALR YNVTWYVSSSPCAACADRIIKTLSKTKNL RLLILVGRLFMWEEPEIQAALKKLKEAGC KLRIMKPQDFEYVWQNFVEQEEGESKAFQ PWEDIQENFLYYEEKLADILK
小鼠 APOBEC -2 多肽序列	1408	MAQKEEAAEAAAPASQNGDDLENLEDPE KLKELIDLPPFEIVTGVRLPVNFFKFQFRNV EYSSGRNKTFLCYVVEVQSKGGQAQATQ GYLEDEHAGAHAEEAFFNTILPAFDPALK YNVTWYVSSSPCAACADRILKTLSKTKNL RLLILVSRLFMWEEPEVQAALKKLKEAGC KLRIMKPQDFEYIWQNFVEQEEGESKAFEP WEDIQENFLYYEEKLADILK
大鼠 APOBEC	1409	MAQKEEAAEAAAPASQNGDDLENLEDPE KLKELIDLPPFEIVTGVRLPVNFFKFQFRNV EYSSGRNKTFLCYVVEAQSKGGQVQATQ

[1237]

-2 多肽序列		GYLEDEHAGAHAEEAFFNTILPAFDPALK YNVTWYVSSSPCAACADRILKTLSKTKNL RLLILVSRLFMWEEPEVQAALKKLKEAGC KLRIMKPQDFEYLWQNFVEQEEGESKAFE PWEDIQENFLYYEEKLADILK
海七鳃鳗 CDA1 (pmCDA1)多肽序列	1410	MTDAEYVRIHEKLDIYTFKKQFFNNKKSV SHRCYVLFELKRRGERRACFWGYAVNKP QSGTERGIHAEIFSIRKVEEYLRDNPGQFTI NWYSSWSPCADCAEKILEWYNQELRGNG HTLKIWACKLYYEKNARNQIGLWNLRDN GVGLNVMVSEHYQCCRKIFIQSSHNQLNE NRWLEKTLKRAEKRRSELSFMIQVKILHTT KSPAV
人 APOBEC 3G D316R D317R 多肽序列	1411	MKPHFRNTVERMYRDTFSYNFYNRPILSR RNTVWLCYEVKTKGPSRPPLDAKIFRGQV YSELKYHPEMRFFHWFSKWRKLHRDQEY EVTWYISWSPCTKCTRDMATFLAEDPKVT LTIFVARLYYFWDPDYQEALRSLCQKRDG PRATMKFNYDEFQHCWSKFVYSQRELFEP WNNLPKYYILLHFMLGEILRHSMDPPTFTF NFNNEPWVRGRHETYLCYEVERMHNDT WVLLNQRRGFLCNQAPHKHGFLEGRHAE LCFLDVIPFWKLDLDQDYRVTCFTSWSPC FSCAQEMAKFISKKHVSLCIFTARIYRRQG RCQEGLRTLAEAGAKISFTYSEFKHCWDT FVDHQGCPFQPWDGLDEHSQDLSGRLRAI LQNQEN
人 APOBEC 3G 链 A 多肽序列	1412	MDPPTFTFNFNNEPWWGRHETYLCYEVE RMHNDTWVLLNQRRGFLCNQAPHKHGFL EGRHAELCFLDVIPFWKLDLDQDYRVTCF TSWSPCFSCAQEMAKFISKNKHVSLCIFTA RIYDDQGRCQEGLRTLAEAGAKISFTYSEF KHCWDTFVDHQGCPFQPWDGLDEHSQDL SGRLRAILQ
人 APOBEC 3G 链 A D120R D121R 多肽序列	1414	MDPPTFTFNFNNEPWVRGRHETYLCYEVE RMHNDTWVLLNQRRGFLCNQAPHKHGFL EGRHAELCFLDVIPFWKLDLDQDYRVTCF TSWSPCFSCAQEMAKFISKNKHVSLCIFTA RIYRRQGRCQEGLRTLAEAGAKISFMTYSE FKHCWDTFVDHQGCPFQPWDGLDEHSQD LSGRLRAILQ
hAPOBEC-4 多肽	1415	MEPIYEEYLANHGTIVKPYYWLSFSLDCSN CPYHIRTGEEARVSLTEFCQIFGFPYGTTFP

[1238]

序列		QTKHLTFYELKTSSGSLVQKGHASSCTGN YIHPESMLFEMNGYLDSAIYNNDSIRHIILY SNNSPCNEANHCCISKMYNFLITYPGITLSI YFSQLYHTEMDFPASAWNREALRSLASL WPRVVLSPISGGIWHSVLHSFISGVSGSHV FQPILTGRALADRHNAYEINAITGVKPYFT DVLLQTKRNPNTKAQEALESYPLNNAFPG QFFQMPSGQLQPNLPPDLRAPVVFVLVPL RDLPPMHMGQNPNKPRNIVRHLNMPQMS FQETKDLGRLPTGRSVEIVEITEQFASSKEA DEKKKKKGKK
rAPOBEC-4 多肽序列	1416	MEPLYEEYLTHSGTIVKPYYWLSVSLNCT NCPYHIRTGEEARVPYTEFHQTFGFPWSTY PQTKHLTFYELRSSSGNLIQKGLASNCTGS HTHPESMLFERDGYLDSLIFHDSNIRHIILY SNNSPCDEANHCCISKMYNFLMNYPEVTL SVFFSQLYHTENQFPTSAWNREALRGLAS LWPQVTLSAISGGIWQSILETFVSGISEGLT AVRPFTAGRTLTDRYNAYEINCITEVKPYF TDALHSWQKENQDQKVWAASENQPLHN TTPAQWQPDMSQDCRTPAVFMLVPYRDL PPIHVNPSPQKPRTVVRHLNTLQLSASKVK ALRKSPSGRPVKKEEARKGSTRSQEANET NKSKWKKQTLFIKSNICHLLEREQKKIGIL SSWSV
rAPOBEC-1 (Δ177-186)多肽序列	1421	MSSETGPVAVDPTLRRRIEPHEFEVFFDPR ELRKETCLLYEINWGGRHSIWRHTSQNTN KHVEVNFIEKFTTERYFCPNTRCSITWFLS WSPCGECSRAITEFLSRYPHVTLFIYIARLY HHADPRNRQGLRDLISSGVTIQIMTEQESG YCWRNFVNYSPSNEAHWPRYPHLWVRGL PPCLNILRRKQPQLTFFTIALQSCHYQRLPP HILWATGLK
rAPOBEC-1 (Δ202-213)多肽序列	1422	MSSETGPVAVDPTLRRRIEPHEFEVFFDPR ELRKETCLLYEINWGGRHSIWRHTSQNTN KHVEVNFIEKFTTERYFCPNTRCSITWFLS WSPCGECSRAITEFLSRYPHVTLFIYIARLY HHADPRNRQGLRDLISSGVTIQIMTEQESG YCWRNFVNYSPSNEAHWPRYPHLWVRLY VLELYCIILGLPPCLNILRRKQPQHYQRLPP HILWATGLK
多肽序列AID	1373	MDSLLMKQKKFLYHFKNVRWAKGRHET YLCYVVKRRDSATSCSLDFGHLRNKSGCH

[1239]

(mAPOBEC-4)多肽序列		VELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVAEFLRWNPNLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIGIMTFKDYFYCWNTFVENRERTFKAWEGLHENSVRLTRQLRRILLPLYEVDDLRDAFRMLGF
pmCDA-1 多肽序列	1417	MAGYECVRVSEKLDFDTFEFQFENLHYATERHRTYVIFDVKPQSAGGRSRRLWGYIINNPNVCHAELILMSMIDRHLESNPGVYAMTWYMSWSPCANCSSKLNPWLKNLLEEQGHTLTMHFSRIYDRDREGDHRGLRGLKHVSNSFRMGVVGRAEVKECLAEYVEASRRTLTWLDTTESMAAKMRRKLFCILVRCAGMRESGIPLHLFTLQTPLLSGRVVWWRV
pmCDA-2 多肽序列	1418	MELREVVDCALASCVRHEPLSRVAFLRCFAAPSQKPRGTVILFYVEGAGRGVTGGHAVNYNKQGTSIHAEVLLLSAVRAALLRRRRCEDGEEATRGCTLHCYSTYSPCRDCVEYIQEFGASTGVRVVIHCCRLYELDVNRRRSEAEGVLRSLSRLGRDFRLMGPRDAIALLLGGRLANTADGESGASGNAWVTETNVVEPLVDMTGFGDEDLHAQVQRNKQIREAYANYASAVSLMLGELHVDPDKFPFLAEFLAQTSVEPSGTPRETRGRPRGASSRGPEIGRQRPADFERALGAYGLFLHPRIVSREADREEIKRDLIVVMRKHNYQGP
pmCDA-5 多肽序列	1419	MAGDENVRVSEKLDFDTFEFQFENLHYATERHRTYVIFDVKPQSAGGRSRRLWGYIINNPNVCHAELILMSMIDRHLESNPGVYAMTWYMSWSPCANCSSKLNPWLKNLLEEQGHTLMMHFSRIYDRDREGDHRGLRGLKHVSNSFRMGVVGRAEVKECLAEYVEASRRTLTWLDTTESMAAKMRRKLFCILVRCAGMRESGMPLHLFT
yCD 多肽序列	1420	MVTGGMASKWDQKGMDIAYEEAALGYKEGGVPIGGCLINNKDGSVLGRGHNMRFQKGSATLHGEISTLENCGRLEGKVYKDTTLYTTLSPCDMCTGAIIMYGIPRCVVGENVNFKSKGEKYLQTRGHEVVVVDDERCKKIMKQFIDERPQDWFEDIGE
NLS	84	KRTADGSEFESPKKKRKV
NLS	85	KRPAATKKAGQAKKKK
NLS	86	KKTELQTTNAENKTKKL
NLS	87	KRGINDRNFWRGENGRKTR

[1240]

NLS	88	RKSGKIAAIVVKRPRK
NLS	89	PKKKRKV
NLS	90	MDSLLMNRRKFLYQFKNVRWAKGRRETY LC
WT cas9 结构域	223	MDKKYSIGLDIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFDSPTVAYSVLVVAKVEKG

[1241]

		KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIE QISEFSKRVILADANLDKVLSAYNKHRDK PIREQAENIIHLFTLTNLGAPAAFKYFDTTI DRKRYTSTKEVLDATLIHQSITGLYETRID LSQLGGD
gRNA 支架核苷酸序列	230	GUUUUAGAGCUAGAAAUAGCAAGUUAA AAUAAGGCUAGUCCGUUAUCAACUUGA AAAAGUGGGACCGAGUCGGUGCUUUU
野生型 spCas9 多核苷酸序列	231	ATGGATAAGAAATACTCAATAGGCTTAG ATATCGGCACAAATAGCGTCGGATGGGC GGTGATCACTGATGATTATAAGGTTCCG TCTAAAAAGTTCAAGGTTCTGGGAAATA CAGACCGCCACAGTATCAAAAAAAATCT TATAGGGGCTCTTTTATTTGGCAGTGGA GAGACAGCGGAAGCGACTCGTCTCAAAC GGACAGCTCGTAGAAGGTATACACGTCG GAAGAATCGTATTTGTTATCTACAGGAG ATTTTTTCAAATGAGATGGCGAAAGTAG ATGATAGTTTCTTTCATCGACTTGAAGA GTCTTTTTTGGTGGAAGAAGACAAGAAG CATGAACGTCATCCTATTTTTGGAAATAT AGTAGATGAAGTTGCTTATCATGAGAAA TATCCAACTATCTATCATCTGCGAAAAA AATTGGCAGATTCTACTGATAAAGCGGA TTTGCGCTTAATCTATTTGGCCTTAGCGC ATATGATTAAGTTTCGTGGTCATTTTTTG ATTGAGGGAGATTTAAATCCTGATAATA GTGATGTGGACAAACTATTTATCCAGTT GGTACAAATCTACAATCAATTATTTGAA GAAAACCCTATTAACGCAAGTAGAGTAG ATGCTAAAGCGATTCTTTCTGCACGATT GAGTAAATCAAGACGATTAGAAAATCTC ATTGCTCAGCTCCCCGGTGAGAAGAGAA ATGGCTTGTTTGGGAATCTCATTGCTTTG TCATTGGGATTGACCCCTAATTTTAAATC AAATTTTGATTTGGCAGAAGATGCTAAA TTACAGCTTTCAAAAGATACTTACGATG ATGATTTAGATAATTTATTGGCGCAAAT TGGAGATCAATATGCTGATTTGTTTTTGG

[1242]

		CAGCTAAGAATTTATCAGATGCTATTTT ACTTTCAGATATCCTAAGAGTAAATAGT GAAATAACTAAGGCTCCCCTATCAGCTT CAATGATTAAGCGCTACGATGAACATCA TCAAGACTTGACTCTTTTAAAAGCTTTAG TTCGACAACAACTTCCAGAAAAGTATAA AGAAATCTTTTTTGATCAATCAAAAAAC GGATATGCAGGTTATATTGATGGGGGAG CTAGCCAAGAAGAATTTTATAAATTTAT CAAACCAATTTTAGAAAAAATGGATGGT ACTGAGGAATTATTGGTGAAACTAAATC GTGAAGATTTGCTGCGCAAGCAACGGAC CTTTGACAACGGCTCTATTCCCCATCAA ATTCACTTGGGTGAGCTGCATGCTATTTT GAGAAGACAAGAAGACTTTTATCCATTT TTAAAAGACAATCGTGAGAAGATTGAAA AAATCTTGACTTTTCGAATTCCTTATTAT GTTGGTCCATTGGCGCGTGGCAATAGTC GTTTTGCATGGATGACTCGGAAGTCTGA AGAAACAATTACCCCATGGAATTTTGAA GAAGTTGTCGATAAAGGTGCTTCAGCTC AATCATTTATTGAACGCATGACAAACTT TGATAAAAATCTTCCAAATGAAAAAGTA CTACCAAAACATAGTTTGCTTTATGAGT ATTTTACGGTTTATAACGAATTGACAAA GGTCAAATATGTTACTGAGGGAATGCGA AAACCAGCATTTCTTTCAGGTGAACAGA AGAAAGCCATTGTTGATTTACTCTTCAA AACAAATCGAAAAGTAACCGTTAAGCA ATTAAAAGAAGATTATTTCAAAAAAATA GAATGTTTTGATAGTGTTGAAATTTCAG GAGTTGAAGATAGATTTAATGCTTCATT AGGCGCCTACCATGATTTGCTAAAAATT ATTAAAGATAAAGATTTTTTGGATAATG AAGAAAATGAAGATATCTTAGAGGATAT TGTTTTAACATTGACCTTATTTGAAGATA GGGGGATGATTGAGGAAAGACTTAAAA CATATGCTCACCTCTTTGATGATAAGGT GATGAAACAGCTTAAACGTCGCCGTTAT ACTGGTTGGGGACGTTTGTCTCGAAAAT TGATTAATGGTATTAGGGATAAGCAATC TGGCAAAACAATATTAGATTTTTTGAAA TCAGATGGTTTTGCCAATCGCAATTTTAT

[1243]

		GCAGCTGATCCATGATGATAGTTTGACA TTTAAAGAAGATATTCAAAAAGCACAGG TGTCTGGACAAGGCCATAGTTTACATGA ACAGATTGCTAACTTAGCTGGCAGTCCT GCTATTAAAAAAGGTATTTTACAGACTG TAAAAATTGTTGATGAACTGGTCAAAGT AATGGGGCATAAGCCAGAAAATATCGTT ATTGAAATGGCACGTGAAAATCAGACAA CTCAAAAGGGCCAGAAAAATTCGCGAG AGCGTATGAAACGAATCGAAGAAGGTA TCAAAGAATTAGGAAGTCAGATTCTTAA AGAGCATCCTGTTGAAAATACTCAATTG CAAAATGAAAAGCTCTATCTCTATTATC TACAAAATGGAAGAGACATGTATGTGGA CCAAGAATTAGATATTAATCGTTTAAGT GATTATGATGTCGATCACATTGTTCCAC AAAGTTTCATTAAAGACGATTCAATAGA CAATAAGGTACTAACGCGTTCTGATAAA AATCGTGGTAAATCGGATAACGTTCCAA GTGAAGAAGTAGTCAAAAAGATGAAAA ACTATTGGAGACAACTTCTAAACGCCAA GTTAATCACTCAACGTAAGTTTGATAAT TTAACGAAAGCTGAACGTGGAGGTTTGA GTGAACTTGATAAAGCTGGTTTTATCAA ACGCCAATTGGTTGAAACTCGCCAAATC ACTAAGCATGTGGCACAAATTTTGGATA GTCGCATGAATACTAAATACGATGAAAA TGATAAACTTATTCGAGAGGTTAAAGTG ATTACCTTAAAATCTAAATTAGTTTCTGA CTTCCGAAAAGATTTCCAATTCTATAAA GTACGTGAGATTAACAATTACCATCATG CCCATGATGCGTATCTAAATGCCGTCGT TGGAACTGCTTTGATTAAGAAATATCCA AAACTTGAATCGGAGTTTGTCTATGGTG ATTATAAAGTTTATGATGTTCGTAAAAT GATTGCTAAGTCTGAGCAAGAAATAGGC AAAGCAACCGCAAAATATTTCTTTTACT CTAATATCATGAACTTCTTCAAAACAGA AATTACACTTGCAAATGGAGAGATTCGC AAACGCCCTCTAATCGAAACTAATGGGG AAACTGGAGAAATTGTCTGGGATAAAGG GCGAGATTTTGCCACAGTGCGCAAAGTA TTGTCCATGCCCCAAGTCAATATTGTCA

[1244]

		AGAAAACAGAAGTACAGACAGGCGGAT TCTCCAAGGAGTCAATTTTACCAAAAAG AAATTCGGACAAGCTTATTGCTCGTAAA AAAGACTGGGATCCAAAAAAATATGGT GGTTTTGATAGTCCAACGGTAGCTTATTC AGTCCTAGTGGTTGCTAAGGTGGAAAAA GGGAAATCGAAGAAGTTAAAATCCGTTA AAGAGTTACTAGGGATCACAATTATGGA AAGAAGTTCCTTTGAAAAAAATCCGATT GACTTTTTAGAAGCTAAAGGATATAAGG AAGTTAAAAAAGACTTAATCATTAAACT ACCTAAATATAGTCTTTTTGAGTTAGAA AACGGTCGTAAACGGATGCTGGCTAGTG CCGGAGAATTACAAAAAGGAAATGAGC TGGCTCTGCCAAGCAAATATGTGAATTT TTTATATTTAGCTAGTCATTATGAAAAGT TGAAGGGTAGTCCAGAAGATAACGAAC AAAAACAATTGTTTGTGGAGCAGCATAA GCATTATTTAGATGAGATTATTGAGCAA ATCAGTGAATTTTCTAAGCGTGTTATTTT AGCAGATGCCAATTTAGATAAAGTTCTT AGTGCATATAACAAACATAGAGACAAA CCAATACGTGAACAAGCAGAAAATATTA TTCATTTATTTACGTTGACGAATCTTGGA GCTCCCGCTGCTTTTAAATATTTTGATAC AACAATTGATCGTAAACGATATACGTCT ACAAAAGAAGTTTTAGATGCCACTCTTA TCCATCAATCCATCACTGGTCTTTATGAA ACACGCATTGATTTGAGTCAGCTAGGAG GTGACTGA
spCas9 多肽序列	232	MDKKYSIGLDIGTNSVGWAVITDDYKVPS KKFKVLGNTDRHSIKKNLIGALLFGSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLADSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQIYNQLFEENPINASRVDA KAILSARLSKSRRLENLIAQLPGEKRNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNSEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE

[1245]

		ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGAY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDRGMIEERLKTYAHLFDDKVMKQLKR RRYTGWGRLSRKLINGIRDKQSGKTILDFL KSDGFANRNFMQLIHDDSLTFKEDIQKAQ VSGQGHSLHEQIANLAGSPAIKKGILQTVK IVDELVKVMGHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFIKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFDSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIE QISEFSKRVILADANLDKVLSAYNKHRDK PIREQAENIIHLFTLTNLGAPAAFKYFDTTI DRKRYTSTKEVLDATLIHQSITGLYETRID LSQLGGD
野生型 Cas9 多核苷酸序列	235	ATGGATAAAAAGTATTCTATTGGTTTAG ACATCGGCACTAATTCCGTTGGATGGGC TGTCATAACCGATGAATACAAAGTACCT TCAAAGAAATTTAAGGTGTTGGGGAACA CAGACCGTCATTCGATTAAAAAGAATCT TATCGGTGCCCTCCTATTCGATAGTGGC GAAACGGCAGAGGCGACTCGCCTGAAA CGAACCGCTCGGAGAAGGTATACACGTC GCAAGAACCGAATATGTTACTTACAAGA

[1246]

		AATTTTTAGCAATGAGATGGCCAAAGTT GACGATTCTTTCTTTCACCGTTTGGAAGA GTCCTTCCTTGTCGAAGAGGACAAGAAA CATGAACGGCACCCCATCTTTGGAAACA TAGTAGATGAGGTGGCATATCATGAAAA GTACCCAACGATTTATCACCTCAGAAAA AAGCTAGTTGACTCAACTGATAAAGCGG ACCTGAGGTTAATCTACTTGGCTCTTGCC CATATGATAAAGTTCCGTGGGCACTTTC TCATTGAGGGTGATCTAAATCCGGACAA CTCGGATGTCGACAAACTGTTCATCCAG TTAGTACAAACCTATAATCAGTTGTTTG AAGAGAACCCTATAAATGCAAGTGGCGT GGATGCGAAGGCTATTCTTAGCGCCCGC CTCTCTAAATCCCGACGGCTAGAAAACC TGATCGCACAATTACCCGGAGAGAAGAA AAATGGGTTGTTCGGTAACCTTATAGCG CTCTCACTAGGCCTGACACCAAATTTTA AGTCGAACTTCGACTTAGCTGAAGATGC CAAATTGCAGCTTAGTAAGGACACGTAC GATGACGATCTCGACAATCTACTGGCAC AAATTGGAGATCAGTATGCGGACTTATT TTTGGCTGCCAAAAACCTTAGCGATGCA ATCCTCCTATCTGACATACTGAGAGTTA ATACTGAGATTACCAAGGCGCCGTTATC CGCTTCAATGATCAAAAGGTACGATGAA CATCACCAAGACTTGACACTTCTCAAGG CCCTAGTCCGTCAGCAACTGCCTGAGAA ATATAAGGAAATATTCTTTGATCAGTCG AAAAACGGGTACGCAGGTTATATTGACG GCGGAGCGAGTCAAGAGGAATTCTACA AGTTTATCAAACCCATATTAGAGAAGAT GGATGGGACGGAAGAGTTGCTTGTAAAA CTCAATCGCGAAGATCTACTGCGAAAGC AGCGGACTTTCGACAACGGTAGCATTCC ACATCAAATCCACTTAGGCGAATTGCAT GCTATACTTAGAAGGCAGGAGGATTTTT ATCCGTTCCTCAAAGACAATCGTGAAAA GATTGAGAAAATCCTAACCTTTCGCATA CCTTACTATGTGGGACCCCTGGCCCGAG GGAACTCTCGGTTCGCATGGATGACAAG AAAGTCCGAAGAAACGATTACTCCATGG AATTTTGAGGAAGTTGTCGATAAAGGTG

[1247]

		CGTCAGCTCAATCGTTCATCGAGAGGAT GACCAACTTTGACAAGAATTTACCGAAC GAAAAAGTATTGCCTAAGCACAGTTTAC TTTACGAGTATTTCACAGTGTACAATGA ACTCACGAAAGTTAAGTATGTCACTGAG GGCATGCGTAAACCCGCCTTTCTAAGCG GAGAACAGAAGAAAGCAATAGTAGATC TGTTATTCAAGACCAACCGCAAAGTGAC AGTTAAGCAATTGAAAGAGGACTACTTT AAGAAAATTGAATGCTTCGATTCTGTCG AGATCTCCGGGGTAGAAGATCGATTTAA TGCGTCACTTGGTACGTATCATGACCTCC TAAAGATAATTAAAGATAAGGACTTCCT GGATAACGAAGAGAATGAAGATATCTTA GAAGATATAGTGTTGACTCTTACCCTCTT TGAAGATCGGGAAATGATTGAGGAAAG ACTAAAAACATACGCTCACCTGTTCGAC GATAAGGTTATGAAACAGTTAAAGAGGC GTCGCTATACGGGCTGGGGACGATTGTC GCGGAAACTTATCAACGGGATAAGAGA CAAGCAAAGTGGTAAAACTATTCTCGAT TTTCTAAAGAGCGACGGCTTCGCCAATA GGAACTTTATGCAGCTGATCCATGATGA CTCTTTAACCTTCAAAGAGGATATACAA AAGGCACAGGTTTCCGGACAAGGGGACT CATTGCACGAACATATTGCGAATCTTGC TGGTTCGCCAGCCATCAAAAAGGGCATA CTCCAGACAGTCAAAGTAGTGGATGAGC TAGTTAAGGTCATGGGACGTCACAAACC GGAAAACATTGTAATCGAGATGGCACGC GAAAATCAAACGACTCAGAAGGGGCAA AAAAACAGTCGAGAGCGGATGAAGAGA ATAGAAGAGGGTATTAAAGAACTGGGC AGCCAGATCTTAAAGGAGCATCCTGTGG AAAATACCCAATTGCAGAACGAGAAACT TTACCTCTATTACCTACAAAATGGAAGG GACATGTATGTTGATCAGGAACTGGACA TAAACCGTTTATCTGATTACGACGTCGA TCACATTGTACCCCAATCCTTTTTGAAGG ACGATTCAATCGACAATAAAGTGCTTAC ACGCTCGGATAAGAACCGAGGGAAAAG TGACAATGTTCCAAGCGAGGAAGTCGTA AAGAAAATGAAGAACTATTGGCGGCAG

[1248]

		CTCCTAAATGCGAAACTGATAACGCAAA GAAAGTTCGATAACTTAACTAAAGCTGA GAGGGGTGGCTTGTCTGAACTTGACAAG GCCGGATTTATTAAACGTCAGCTCGTGG AAACCCGCCAAATCACAAAGCATGTTGC ACAGATACTAGATTCCCGAATGAATACG AAATACGACGAGAACGATAAGCTGATTC GGGAAGTCAAAGTAATCACTTTAAAGTC AAAATTGGTGTCGGACTTCAGAAAGGAT TTTCAATTCTATAAAGTTAGGGAGATAA ATAACTACCACCATGCGCACGACGCTTA TCTTAATGCCGTCGTAGGGACCGCACTC ATTAAGAAATACCCGAAGCTAGAAAGTG AGTTTGTGTATGGTGATTACAAAGTTTAT GACGTCCGTAAGATGATCGCGAAAAGCG AACAGGAGATAGGCAAGGCTACAGCCA AATACTTCTTTTATTCTAACATTATGAAT TTCTTTAAGACGGAAATCACTCTGGCAA ACGGAGAGATACGCAAACGACCTTTAAT TGAAACCAATGGGGAGACAGGTGAAAT CGTATGGGATAAGGGCCGGGACTTCGCG ACGGTGAGAAAAGTTTTGTCCATGCCCC AAGTCAACATAGTAAAGAAAACTGAGG TGCAGACCGGAGGGTTTTCAAAGGAATC GATTCTTCCAAAAAGGAATAGTGATAAG CTCATCGCTCGTAAAAAGGACTGGGACC CGAAAAAGTACGGTGGCTTCGATAGCCC TACAGTTGCCTATTCTGTCCTAGTAGTGG CAAAAGTTGAGAAGGGAAAATCCAAGA AACTGAAGTCAGTCAAAGAATTATTGGG GATAACGATTATGGAGCGCTCGTCTTTT GAAAAGAACCCCATCGACTTCCTTGAGG CGAAAGGTTACAAGGAAGTAAAAAAGG ATCTCATAATTAAACTACCAAAGTATAG TCTGTTTGAGTTAGAAAATGGCCGAAAA CGGATGTTGGCTAGCGCCGGAGAGCTTC AAAAGGGGAACGAACTCGCACTACCGTC TAAATACGTGAATTTCCTGTATTTAGCGT CCCATTACGAGAAGTTGAAAGGTTCACC TGAAGATAACGAACAGAAGCAACTTTTT GTTGAGCAGCACAAACATTATCTCGACG AAATCATAGAGCAAATTTCGGAATTCAG TAAGAGAGTCATCCTAGCTGATGCCAAT

[1249]

		CTGGACAAAGTATTAAGCGCATACAACA AGCACAGGGATAAACCCATACGTGAGC AGGCGGAAAATATTATCCATTTGTTTAC TCTTACCAACCTCGGCGCTCCAGCCGCA TTCAAGTATTTTGACACAACGATAGATC GCAAACGATACACTTCTACCAAGGAGGT GCTAGACGCGACACTGATTCACCAATCC ATCACGGGATTATATGAAACTCGGATAG ATTTGTCACAGCTTGGGGGTGACGGATC CCCCAAGAAGAAGAGGAAAGTCTCGAG CGACTACAAAGACCATGACGGTGATTAT AAAGATCATGACATCGATTACAAGGATG ACGATGACAAGGCTGCAGGA
野生型 Cas9 多肽序列	236	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL

[1250]

		NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFDSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIE QISEFSKRVILADANLDKVLSAYNKHRDK PIREQAENIIHLFTLTNLGAPAAFKYFDTTI DRKRYTSTKEVLDATLIHQSITGLYETRID LSQLGGD
来自酿脓链球菌(NCBI Ref. Seq.: NC_002737.2)多核苷酸序列的 Cas9	237	ATGGATAAGAAATACTCAATAGGCTTAG ATATCGGCACAAATAGCGTCGGATGGGC GGTGATCACTGATGAATATAAGGTTCCG TCTAAAAAGTTCAAGGTTCTGGGAAATA CAGACCGCCACAGTATCAAAAAAAATCT TATAGGGGCTCTTTTATTTGACAGTGGA GAGACAGCGGAAGCGACTCGTCTCAAAC GGACAGCTCGTAGAAGGTATACACGTCG GAAGAATCGTATTTGTTATCTACAGGAG ATTTTTTCAAATGAGATGGCGAAAGTAG ATGATAGTTTCTTTCATCGACTTGAAGA GTCTTTTTTGGTGGAAGAAGACAAGAAG CATGAACGTCATCCTATTTTTGGAAATAT AGTAGATGAAGTTGCTTATCATGAGAAA TATCCAACTATCTATCATCTGCGAAAAA AATTGGTAGATTCTACTGATAAAGCGGA TTTGCGCTTAATCTATTTGGCCTTAGCGC ATATGATTAAGTTTCGTGGTCATTTTTTG ATTGAGGGAGATTTAAATCCTGATAATA GTGATGTGGACAAACTATTTATCCAGTT GGTACAAACCTACAATCAATTATTTGAA GAAAACCCTATTAACGCAAGTGGAGTAG ATGCTAAAGCGATTCTTTCTGCACGATT GAGTAAATCAAGACGATTAGAAAATCTC ATTGCTCAGCTCCCCGGTGAGAAGAAAA ATGGCTTATTTGGGAATCTCATTGCTTTG

[1251]

		TCATTGGGTTTGACCCCTAATTTTAAATC AAATTTTGATTTGGCAGAAGATGCTAAA TTACAGCTTTCAAAAGATACTTACGATG ATGATTTAGATAATTTATTGGCGCAAAT TGGAGATCAATATGCTGATTTGTTTTTGG CAGCTAAGAATTTATCAGATGCTATTTT ACTTTCAGATATCCTAAGAGTAAATACT GAAATAACTAAGGCTCCCCTATCAGCTT CAATGATTAAACGCTACGATGAACATCA TCAAGACTTGACTCTTTTAAAAGCTTTAG TTCGACAACAACTTCCAGAAAAGTATAA AGAAATCTTTTTTGATCAATCAAAAAAC GGATATGCAGGTTATATTGATGGGGGAG CTAGCCAAGAAGAATTTTATAAATTTAT CAAACCAATTTTAGAAAAAATGGATGGT ACTGAGGAATTATTGGTGAAACTAAATC GTGAAGATTTGCTGCGCAAGCAACGGAC CTTTGACAACGGCTCTATTCCCCATCAA ATTCACTTGGGTGAGCTGCATGCTATTTT GAGAAGACAAGAAGACTTTTATCCATTT TTAAAAGACAATCGTGAGAAGATTGAAA AAATCTTGACTTTTCGAATTCCTTATTAT GTTGGTCCATTGGCGCGTGGCAATAGTC GTTTTGCATGGATGACTCGGAAGTCTGA AGAAACAATTACCCCATGGAATTTTGAA GAAGTTGTCGATAAAGGTGCTTCAGCTC AATCATTTATTGAACGCATGACAAACTT TGATAAAAATCTTCCAAATGAAAAAGTA CTACCAAAACATAGTTTGCTTTATGAGT ATTTTACGGTTTATAACGAATTGACAAA GGTCAAATATGTTACTGAAGGAATGCGA AAACCAGCATTTCTTTCAGGTGAACAGA AGAAAGCCATTGTTGATTTACTCTTCAA AACAAATCGAAAAGTAACCGTTAAGCA ATTAAAAGAAGATTATTTCAAAAAAATA GAATGTTTTGATAGTGTTGAAATTTCAG GAGTTGAAGATAGATTTAATGCTTCATT AGGTACCTACCATGATTTGCTAAAAATT ATTAAAGATAAAGATTTTTTGGATAATG AAGAAAATGAAGATATCTTAGAGGATAT TGTTTTAACATTGACCTTATTTGAAGATA GGGAGATGATTGAGGAAAGACTTAAAA CATATGCTCACCTCTTTGATGATAAGGT

[1252]

		GATGAAACAGCTTAAACGTCGCCGTTAT ACTGGTTGGGGACGTTTGTCTCGAAAAT TGATTAATGGTATTAGGGATAAGCAATC TGGCAAAACAATATTAGATTTTTTGAAA TCAGATGGTTTTGCCAATCGCAATTTTAT GCAGCTGATCCATGATGATAGTTTGACA TTTAAAGAAGACATTCAAAAAGCACAAG TGTCTGGACAAGGCGATAGTTTACATGA ACATATTGCAAATTTAGCTGGTAGCCCT GCTATTAAAAAAGGTATTTTACAGACTG TAAAAGTTGTTGATGAATTGGTCAAAGT AATGGGGCGGCATAAGCCAGAAAATAT CGTTATTGAAATGGCACGTGAAAATCAG ACAACTCAAAAGGGCCAGAAAAATTCG CGAGAGCGTATGAAACGAATCGAAGAA GGTATCAAAGAATTAGGAAGTCAGATTC TTAAAGAGCATCCTGTTGAAAATACTCA ATTGCAAAATGAAAAGCTCTATCTCTAT TATCTCCAAAATGGAAGAGACATGTATG TGGACCAAGAATTAGATATTAATCGTTT AAGTGATTATGATGTCGATCACATTGTT CCACAAAGTTTCCTTAAAGACGATTCAA TAGACAATAAGGTCTTAACGCGTTCTGA TAAAAATCGTGGTAAATCGGATAACGTT CCAAGTGAAGAAGTAGTCAAAAAGATG AAAAACTATTGGAGACAACTTCTAAACG CCAAGTTAATCACTCAACGTAAGTTTGA TAATTTAACGAAAGCTGAACGTGGAGGT TTGAGTGAACTTGATAAAGCTGGTTTTA TCAAACGCCAATTGGTTGAAACTCGCCA AATCACTAAGCATGTGGCACAAATTTTG GATAGTCGCATGAATACTAAATACGATG AAAATGATAAACTTATTCGAGAGGTTAA AGTGATTACCTTAAAATCTAAATTAGTTT CTGACTTCCGAAAAGATTTCCAATTCTAT AAAGTACGTGAGATTAACAATTACCATC ATGCCCATGATGCGTATCTAAATGCCGT CGTTGGAACTGCTTTGATTAAGAAATAT CCAAAACTTGAATCGGAGTTTGTCTATG GTGATTATAAAGTTTATGATGTTCGTAA AATGATTGCTAAGTCTGAGCAAGAAATA GGCAAAGCAACCGCAAAATATTTCTTTT ACTCTAATATCATGAACTTCTTCAAAAC

[1253]

		AGAAATTACACTTGCAAATGGAGAGATT CGCAAACGCCCTCTAATCGAAACTAATG GGGAAACTGGAGAAATTGTCTGGGATAA AGGGCGAGATTTTGCCACAGTGCGCAAA GTATTGTCCATGCCCCAAGTCAATATTGT CAAGAAAACAGAAGTACAGACAGGCGG ATTCTCCAAGGAGTCAATTTTACCAAAA AGAAATTCGGACAAGCTTATTGCTCGTA AAAAAGACTGGGATCCAAAAAAATATG GTGGTTTTGATAGTCCAACGGTAGCTTA TTCAGTCCTAGTGGTTGCTAAGGTGGAA AAAGGGAAATCGAAGAAGTTAAAATCC GTTAAAGAGTTACTAGGGATCACAATTA TGGAAAGAAGTTCCTTTGAAAAAAATCC GATTGACTTTTTAGAAGCTAAAGGATAT AAGGAAGTTAAAAAAGACTTAATCATTA AACTACCTAAATATAGTCTTTTTGAGTTA GAAAACGGTCGTAAACGGATGCTGGCTA GTGCCGGAGAATTACAAAAAGGAAATG AGCTGGCTCTGCCAAGCAAATATGTGAA TTTTTTATATTTAGCTAGTCATTATGAAA AGTTGAAGGGTAGTCCAGAAGATAACG AACAAAAACAATTGTTTGTGGAGCAGCA TAAGCATTATTTAGATGAGATTATTGAG CAAATCAGTGAATTTTCTAAGCGTGTTA TTTTAGCAGATGCCAATTTAGATAAAGT TCTTAGTGCATATAACAAACATAGAGAC AAACCAATACGTGAACAAGCAGAAAAT ATTATTCATTTATTTACGTTGACGAATCT TGGAGCTCCCGCTGCTTTTAAATATTTTG ATACAACAATTGATCGTAAACGATATAC GTCTACAAAAGAAGTTTTAGATGCCACT CTTATCCATCAATCCATCACTGGTCTTTA TGAAACACGCATTGATTTGAGTCAGCTA GGAGGTGACTGA
无催化活性的 Cas9 (dCas9)多肽序列	238	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF

[1254]

		GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDAIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFDSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIE QISEFSKRVILADANLDKVLSAYNKHRDK PIREQAENIIHLFTLTNLGAPAAFKYFDTTI DRKRYTSTKEVLDATLIHQSITGLYETRID LSQLGGD
tr|F0NN87|F0NN87_SULIH CRISPR	239	MEVPLYNIFGDNYIIQVATEAENSTIYNNK VEIDDEELRNVLNLAYKIAKNNEDAAAER RGKAKKKKGEEGETTTSNIILPLSGNDKNP WTETLKCYNFPTTVALSEVFKNFSQVKEC

[1255]

相关 Casx 蛋白 OS =冰岛硫化叶菌 (Sulfolobus islandicus) (菌株 HVE10/4) GN = SiH_0402 PE=4 SV=1); CasX 多肽序列		EEVSAPSFVKPEFYEFGRSPGMVERTRRVK LEVEPHYLIIAAAGWVLTRLGKAKVSEGD YVGVNVFTPTRGILYSLIQNVNGIVPGIKPE TAFGLWIARKVVSSVTNPNVSVVRIYTISD AVGQNPTTINGGFSIDLTKLLEKRYLLSER LEAIARNALSISSNMRERYIVLANYIYEYLT GSKRLEDLLYFANRDLIMNLNSDDGKVRD LKLISAYVNGELIRGEG
tr|F0NH53|F0NH53_SULIR CRISPR 相关蛋白，Casx OS =冰岛硫化叶菌 (菌株 REY15A) GN=SiRe_0771 PE=4 SV=1); CasX 多肽序列	240	MEVPLYNIFGDNYIIQVATEAENSTIYNNK VEIDDEELRNVLNLAYKIAKNNEDAAAER RGKAKKKKGEEGETTTSNIILPLSGNDKNP WTETLKCYNFPTTVALSEVFKNFSQVKEC EEVSAPSFVKPEFYKFGRSPGMVERTRRV KLEVEPHYLIMAAAGWVLTRLGKAKVSE GDYVGVNVFTPTRGILYSLIQNVNGIVPGI KPETAFGLWIARKVVSSVTNPNVSVVSIYT ISDAVGQNPTTINGGFSIDLTKLLEKRDLLS ERLEAIARNALSISSNMRERYIVLANYIYE YLTGSKRLEDLLYFANRDLIMNLNSDDGK VRDLKLISAYVNGELIRGEG
CasX 多肽序列	241	MEKRINKIRKKLSADNATKPVSRSGPMKT LLVRVMTDDLKKRLEKRRKKPEVMPQVIS NNAANNLRMLLDDYTKMKEAILQVYWQ EFKDDHVGLMCKFAQPASKKIDQNKLKPE MDEKGNLTTAGFACSQCGQPLFVYKLEQ VSEKGKAYTNYFGRCNVAEHEKLILLAQL KPVKDSDEAVTYSLGKFGQRALDFYSIHV TKESTHPVKPLAQIAGNRYASGPVGKALS DACMGTIASFLSKYQDIIIEHQKVVKGNQK RLESLRELAGKENLEYPSVTLPPQPHTKEG

[1256]

		VDFAYNEVIARVRMWVNLNLWQKLKLSR DDAKPLLRLKGFPSFPVVERRENEVDWW NTINEVKKLIDAKRDMGRVFWSGVTAEK RNTILEGYNYLPNENDHKKREGSLENPKK PAKRQFGDLLLYLEKKYAGDWGKVFDEA WERIDKKIAGLTSHIEREEARNAEDAQSKA VLTDWLRAKASFVLERLKEMDEKEFYAC EIQLQKWYGDLRGNPFAVEAENRVVDISG FSIGSDGHSIQYRNLLAWKYLENGKREFY LLMNYGKKGRIRFTDGTDIKKSGKWQGLL YGGGKAKVIDLTFDPDDEQLIILPLAFGTR QGREFIWNDLLSLETGLIKLANGRVIEKTI YNKKIGRDEPALFVALTFERREVVDPSNIK PVNLIGVARGENIPAVIALTDPEGCPLPEFK DSSGGPTDILRIGEGYKEKQRAIQAAKEVE QRRAGGYSRKFASKSRNLADDMVRNSAR DLFYHAVTHDAVLVFANLSRGFGRQGKR TFMTERQYTKMEDWLTAKLAYEGLTSKT YLSKTLAQYTSKTCSNCGFTITYADMDVM LVRLKKTSDGWATTLNNKELKAEYQITYY NRYKRQTVEKELSAELDRLSEESGNNDISK WTKGRRDEALFLLKKRFSHRPVQEQFVCL DCGHEVHAAEQAALNIARSWLFLNSNSTE FKSYKSGKQPFVGAWQAFYKRRLKEVWK PNA
APG80656.1 CRISPR相关蛋白CasY [未培养的俭菌超门(parcubacteria)群细菌]; CasY多肽序列	**242**	MSKRHPRISGVKGYRLHAQRLEYTGKSGA MRTIKYPLYSSPSGGRTVPREIVSAINDDY VGLYGLSNFDDLYNAEKRNEEKVYSVLDF WYDCVQYGAVFSYTAPGLLKNVAEVRGG SYELTKTLKGSHLYDELQIDKVIKFLNKKE ISRANGSLDKLKKDIIDCFKAEYRERHKDQ CNKLADDIKNAKKDAGASLGERQKKLFR DFFGISEQSENDKPSFTNPLNLTCCLLPFDT VNNNRNRGEVLFNKLKEYAQKLDKNEGS LEMWEYIGIGNSGTAFSNFLGEGFLGRLRE NKITELKKAMMDITDAWRGQEQEEELEK RLRILAALTIKLREPKFDNHWGGYRSDING KLSSWLQNYINQTVKIKEDLKGHKKDLKK AKEMINRFGESDTKEEAVVSSLLESIEKIVP DDSADDEKPDIPAIAIYRRFLSDGRLTLNRF VQREDVQEALIKERLEAEKKKKPKKRKKK SDAEDEKETIDFKELFPHLAKPLKLVPNFY GDSKRELYKKYKNAAIYTDALWKAVEKI

[1257]

		YKSAFSSSLKNSFFDTDFDKDFFIKRLQKIF SVYRRFNTDKWKPIVKNSFAPYCDIVSLAE NEVLYKPKQSRSRKSAAIDKNRVRLPSTE NIAKAGIALARELSVAGFDWKDLLKKEEH EEYIDLIELHKTALALLLAVTETQLDISALD FVENGTVKDFMKTRDGNLVLEGRFLEMFS QSIVFSELRGLAGLMSRKEFITRSAIQTMN GKQAELLYIPHEFQSAKITTPKEMSRAFLD LAPAEFATSLEPESLSEKSLLKLKQMRYYP HYFGYELTRTGQGIDGGVAENALRLEKSP VKKREIKCKQYKTLGRGQNKIVLYVRSSY YQTQFLEWFLHRPKNVQTDVAVSGSFLID EKKVKTRWNYDALTVALEPVSGSERVFVS QPFTIFPEKSAEEEGQRYLGIDIGEYGIAYT ALEITGDSAKILDQNFISDPQLKTLREEVK GLKLDQRRGTFAMPSTKIARIRESLVHSLR NRIHHLALKHKAKIVYELEVSRFEEGKQKI KKVYATLKKADVYSEIDADKNLQTTVWG KLAVASEISASYTSQFCGACKKLWRAEMQ VDETITTQELIGTVRVIKGGTLIDAIKDFMR PPIFDENDTPFPKYRDFCDKHHISKKMRGN SCLFICPFCRANADADIQASQTIALLRYVK EEKKVEDYFERFRKLKNIKVLGQMKKI
野生型 Cpf1 多肽序列	246	MSIYQEFVNKYSLSKTLRFELIPQGKTLENI KARGLILDDEKRAKDYKKAKQIIDKYHQF FIEEILSSVCISEDLLQNYSDVYFKLKKSDD DNLQKDFKSAKDTIKKQISEYIKDSEKFKN LFNQNLIDAKKGQESDLILWLKQSKDNGIE LFKANSDITDIDEALEIIKSFKGWTTYFKGF HENRKNVYSSNDIPTSIIYRIVDDNLPKFLE NKAKYESLKDKAPEAINYEQIKKDLAEEL TFDIDYKTSEVNQRVFSLDEVFEIANFNNY LNQSGITKFNTIIGGKFVNGENTKRKGINE YINLYSQQINDKTLKKYKMSVLFKQILSDT ESKSFVIDKLEDDSDVVTTMQSFYEQIAAF KTVEEKSIKETLSLLFDDLKAQKLDLSKIY FKNDKSLTDLSQQVFDDYSVIGTAVLEYIT QQIAPKNLDNPSKKEQELIAKKTEKAKYLS LETIKLALEEFNKHRDIDKQCRFEEILANFA AIPMIFDEIAQNKDNLAQISIKYQNQGKKD LLQASAEDDVKAIKDLLDQTNNLLHKLKI FHISQSEDKANILDKDEHFYLVFEECYFEL ANIVPLYNKIRNYITQKPYSDEKFKLNFEN

[1258]

		STLANGWDKNKEPDNTAILFIKDDKYYLG VMNKKNNKIFDDKAIKENKGEGYKKIVY KLLPGANKMLPKVFFSAKSIKFYNPSEDIL RIRNHSTHTKNGSPQKGYEKFEFNIEDCRK FIDFYKQSISKHPEWKDFGFRFSDTQRYNSI DEFYREVENQGYKLTFENISESYIDSVVNQ GKLYLFQIYNKDFSAYSKGRPNLHTLYWK ALFDERNLQDVVYKLNGEAELFYRKQSIP KKITHPAKEAIANKNKDNPKKESVFEYDLI KDKRFTEDKFFFHCPITINFKSSGANKFND EINLLLKEKANDVHILSIDRGERHLAYYTL VDGKGNIIKQDTFNIIGNDRMKTNYHDKL AAIEKDRDSARKDWKKINNIKEMKEGYLS QVVHEIAKLVIEYNAIVVFEDLNFGFKRGR FKVEKQVYQKLEKMLIEKLNYLVFKDNEF DKTGGVLRAYQLTAPFETFKKMGKQTGII YYVPAGFTSKICPVTGFVNQLYPKYESVSK SQEFFSKFDKICYNLDKGYFEFSFDYKNFG DKAAKGKWTIASFGSRLINFRNSDKNHNW DTREVYPTKELEKLLKDYSIEYGHGECIKA AICGESDKKFFAKLTSVLNTILQMRNSKTG TELDYLISPVADVNGNFFDSRQAPKNMPQ DADANGAYHIGLKGLMLLGRIKNNQEGK KLNLVIKNEEYFEFVQNRNN
Cpf1 D917A 多肽序列	247	MSIYQEFVNKYSLSKTLRFELIPQGKTLENI KARGLILDDEKRAKDYKKAKQIIDKYHQF FIEEILSSVCISEDLLQNYSDVYFKLKKSDD DNLQKDFKSAKDTIKKQISEYIKDSEKFKN LFNQNLIDAKKGQESDLILWLKQSKDNGIE LFKANSDITDIDEALEIIKSFKGWTTYFKGF HENRKNVYSSNDIPTSIIYRIVDDNLPKFLE NKAKYESLKDKAPEAINYEQIKKDLAEEL TFDIDYKTSEVNQRVFSLDEVFEIANFNNY LNQSGITKFNTIIGGKFVNGENTKRKGINE YINLYSQQINDKTLKKYKMSVLFKQILSDT ESKSFVIDKLEDDSDVVTTMQSFYEQIAAF KTVEEKSIKETLSLLFDDLKAQKLDLSKIY FKNDKSLTDLSQQVFDDYSVIGTAVLEYIT QQIAPKNLDNPSKKEQELIAKKTEKAKYLS LETIKLALEEFNKHRDIDKQCRFEEILANFA AIPMIFDEIAQNKDNLAQISIKYQNQGKKD LLQASAEDDVKAIKDLLDQTNNLLHKLKI FHISQSEDKANILDKDEHFYLVFEECYFEL

[1259]

		ANIVPLYNKIRNYITQKPYSDEKFKLNFEN STLANGWDKNKEPDNTAILFIKDDKYYLG VMNKKNNKIFDDKAIKENKGEGYKKIVY KLLPGANKMLPKVFFSAKSIKFYNPSEDIL RIRNHSTHTKNGSPQKGYEKFEFNIEDCRK FIDFYKQSISKHPEWKDFGFRFSDTQRYNSI DEFYREVENQGYKLTFENISESYIDSVVNQ GKLYLFQIYNKDFSAYSKGRPNLHTLYWK ALFDERNLQDVVYKLNGEAELFYRKQSIP KKITHPAKEAIANKNKDNPKKESVFEYDLI KDKRFTEDKFFFHCPITINFKSSGANKFND EINLLLKEKANDVHILSIARGERHLAYYTL VDGKGNIIKQDTFNIIGNDRMKTNYHDKL AAIEKDRDSARKDWKKINNIKEMKEGYLS QVVHEIAKLVIEYNAIVVFEDLNFGFKRGR FKVEKQVYQKLEKMLIEKLNYLVFKDNEF DKTGGVLRAYQLTAPFETFKKMGKQTGII YYVPAGFTSKICPVTGFVNQLYPKYESVSK SQEFFSKFDKICYNLDKGYFEFSFDYKNFG DKAAKGKWTIASFGSRLINFRNSDKNHNW DTREVYPTKELEKLLKDYSIEYGHGECIKA AICGESDKKFFAKLTSVLNTILQMRNSKTG TELDYLISPVADVNGNFFDSRQAPKNMPQ DADANGAYHIGLKGLMLLGRIKNNQEGK KLNLVIKNEEYFEFVQNRNN
Cpf1 E1006A 多肽序列	248	MSIYQEFVNKYSLSKTLRFELIPQGKTLENI KARGLILDDEKRAKDYKKAKQIIDKYHQF FIEEILSSVCISEDLLQNYSDVYFKLKKSDD DNLQKDFKSAKDTIKKQISEYIKDSEKFKN LFNQNLIDAKKGQESDLILWLKQSKDNGIE LFKANSDITDIDEALEIIKSFKGWTTYFKGF HENRKNVYSSNDIPTSIIYRIVDDNLPKFLE NKAKYESLKDKAPEAINYEQIKKDLAEEL TFDIDYKTSEVNQRVFSLDEVFEIANFNNY LNQSGITKFNTIIGGKFVNGENTKRKGINE YINLYSQQINDKTLKKYKMSVLFKQILSDT ESKSFVIDKLEDDSDVVTTMQSFYEQIAAF KTVEEKSIKETLSLLFDDLKAQKLDLSKIY FKNDKSLTDLSQQVFDDYSVIGTAVLEYIT QQIAPKNLDNPSKKEQELIAKKTEKAKYLS LETIKLALEEFNKHRDIDKQCRFEEILANFA AIPMIFDEIAQNKDNLAQISIKYQNQGKKD LLQASAEDDVKAIKDLLDQTNNLLHKLKI

[1260]

		FHISQSEDKANILDKDEHFYLVFEECYFEL ANIVPLYNKIRNYITQKPYSDEKFKLNFEN STLANGWDKNKEPDNTAILFIKDDKYYLG VMNKKNNKIFDDKAIKENKGEGYKKIVY KLLPGANKMLPKVFFSAKSIKFYNPSEDIL RIRNHSTHTKNGSPQKGYEKFEFNIEDCRK FIDFYKQSISKHPEWKDFGFRFSDTQRYNSI DEFYREVENQGYKLTFENISESYIDSVVNQ GKLYLFQIYNKDFSAYSKGRPNLHTLYWK ALFDERNLQDVVYKLNGEAELFYRKQSIP KKITHPAKEAIANKNKDNPKKESVFEYDLI KDKRFTEDKFFFHCPITINFKSSGANKFND EINLLLKEKANDVHILSIDRGERHLAYYTL VDGKGNIIKQDTFNIIGNDRMKTNYHDKL AAIEKDRDSARKDWKKINNIKEMKEGYLS QVVHEIAKLVIEYNAIVVFADLNFGFKRGR FKVEKQVYQKLEKMLIEKLNYLVFKDNEF DKTGGVLRAYQLTAPFETFKKMGKQTGII YYVPAGFTSKICPVTGFVNQLYPKYESVSK SQEFFSKFDKICYNLDKGYFEFSFDYKNFG DKAAKGKWTIASFGSRLINFRNSDKNHNW DTREVYPTKELEKLLKDYSIEYGHGECIKA AICGESDKKFFAKLTSVLNTILQMRNSKTG TELDYLISPVADVNGNFFDSRQAPKNMPQ DADANGAYHIGLKGLMLLGRIKNNQEGK KLNLVIKNEEYFEFVQNRNN
Cpf1 D1255A 多肽序列	249	MSIYQEFVNKYSLSKTLRFELIPQGKTLENI KARGLILDDEKRAKDYKKAKQIIDKYHQF FIEEILSSVCISEDLLQNYSDVYFKLKKSDD DNLQKDFKSAKDTIKKQISEYIKDSEKFKN LFNQNLIDAKKGQESDLILWLKQSKDNGIE LFKANSDITDIDEALEIIKSFKGWTTYFKGF HENRKNVYSSNDIPTSIIYRIVDDNLPKFLE NKAKYESLKDKAPEAINYEQIKKDLAEEL TFDIDYKTSEVNQRVFSLDEVFEIANFNNY LNQSGITKFNTIIGGKFVNGENTKRKGINE YINLYSQQINDKTLKKYKMSVLFKQILSDT ESKSFVIDKLEDDSDVVTTMQSFYEQIAAF KTVEEKSIKETLSLLFDDLKAQKLDLSKIY FKNDKSLTDLSQQVFDDYSVIGTAVLEYIT QQIAPKNLDNPSKKEQELIAKKTEKAKYLS LETIKLALEEFNKHRDIDKQCRFEEILANFA AIPMIFDEIAQNKDNLAQISIKYQNQGKKD

[1261]

		LLQASAEDDVKAIKDLLDQTNNLLHKLKI FHISQSEDKANILDKDEHFYLVFEECYFEL ANIVPLYNKIRNYITQKPYSDEKFKLNFEN STLANGWDKNKEPDNTAILFIKDDKYYLG VMNKKNNKIFDDKAIKENKGEGYKKIVY KLLPGANKMLPKVFFSAKSIKFYNPSEDIL RIRNHSTHTKNGSPQKGYEKFEFNIEDCRK FIDFYKQSISKHPEWKDFGFRFSDTQRYNSI DEFYREVENQGYKLTFENISESYIDSVVNQ GKLYLFQIYNKDFSAYSKGRPNLHTLYWK ALFDERNLQDVVYKLNGEAELFYRKQSIP KKITHPAKEAIANKNKDNPKKESVFEYDLI KDKRFTEDKFFFHCPITINFKSSGANKFND EINLLLKEKANDVHILSIDRGERHLAYYTL VDGKGNIIKQDTFNIIGNDRMKTNYHDKL AAIEKDRDSARKDWKKINNIKEMKEGYLS QVVHEIAKLVIEYNAIVVFEDLNFGFKRGR FKVEKQVYQKLEKMLIEKLNYLVFKDNEF DKTGGVLRAYQLTAPFETFKKMGKQTGII YYVPAGFTSKICPVTGFVNQLYPKYESVSK SQEFFSKFDKICYNLDKGYFEFSFDYKNFG DKAAKGKWTIASFGSRLINFRNSDKNHNW DTREVYPTKELEKLLKDYSIEYGHGECIKA AICGESDKKFFAKLTSVLNTILQMRNSKTG TELDYLISPVADVNGNFFDSRQAPKNMPQ DAAANGAYHIGLKGLMLLGRIKNNQEGK KLNLVIKNEEYFEFVQNRNN
Cpf1 D917A/E 1006A 多肽序列	250	MSIYQEFVNKYSLSKTLRFELIPQGKTLENI KARGLILDDEKRAKDYKKAKQIIDKYHQF FIEEILSSVCISEDLLQNYSDVYFKLKKSDD DNLQKDFKSAKDTIKKQISEYIKDSEKFKN LFNQNLIDAKKGQESDLILWLKQSKDNGIE LFKANSDITDIDEALEIIKSFKGWTTYFKGF HENRKNVYSSNDIPTSIIYRIVDDNLPKFLE NKAKYESLKDKAPEAINYEQIKKDLAEEL TFDIDYKTSEVNQRVFSLDEVFEIANFNNY LNQSGITKFNTIIGGKFVNGENTKRKGINE YINLYSQQINDKTLKKYKMSVLFKQILSDT ESKSFVIDKLEDDSDVVTTMQSFYEQIAAF KTVEEKSIKETLSLLFDDLKAQKLDLSKIY FKNDKSLTDLSQQVFDDYSVIGTAVLEYIT QQIAPKNLDNPSKKEQELIAKKTEKAKYLS LETIKLALEEFNKHRDIDKQCRFEEILANFA

[1262]

		AIPMIFDEIAQNKDNLAQISIKYQNQGKKD LLQASAEDDVKAIKDLLDQTNNLLHKLKI FHISQSEDKANILDKDEHFYLVFEECYFEL ANIVPLYNKIRNYITQKPYSDEKFKLNFEN STLANGWDKNKEPDNTAILFIKDDKYYLG VMNKKNNKIFDDKAIKENKGEGYKKIVY KLLPGANKMLPKVFFSAKSIKFYNPSEDIL RIRNHSTHTKNGSPQKGYEKFEFNIEDCRK FIDFYKQSISKHPEWKDFGFRFSDTQRYNSI DEFYREVENQGYKLTFENISESYIDSVVNQ GKLYLFQIYNKDFSAYSKGRPNLHTLYWK ALFDERNLQDVVYKLNGEAELFYRKQSIP KKITHPAKEAIANKNKDNPKKESVFEYDLI KDKRFTEDKFFFHCPITINFKSSGANKFND EINLLLKEKANDVHILSIARGERHLAYYTL VDGKGNIIKQDTFNIIGNDRMKTNYHDKL AAIEKDRDSARKDWKKINNIKEMKEGYLS QVVHEIAKLVIEYNAIVVFADLNFGFKRGR FKVEKQVYQKLEKMLIEKLNYLVFKDNEF DKTGGVLRAYQLTAPFETFKKMGKQTGII YYVPAGFTSKICPVTGFVNQLYPKYESVSK SQEFFSKFDKICYNLDKGYFEFSFDYKNFG DKAAKGKWTIASFGSRLINFRNSDKNHNW DTREVYPTKELEKLLKDYSIEYGHGECIKA AICGESDKKFFAKLTSVLNTILQMRNSKTG TELDYLISPVADVNGNFFDSRQAPKNMPQ DADANGAYHIGLKGLMLLGRIKNNQEGK KLNLVIKNEEYFEFVQNRNN
Cpf1 D917A/D 1255A 多肽序列	251	MSIYQEFVNKYSLSKTLRFELIPQGKTLENI KARGLILDDEKRAKDYKKAKQIIDKYHQF FIEEILSSVCISEDLLQNYSDVYFKLKKSDD DNLQKDFKSAKDTIKKQISEYIKDSEKFKN LFNQNLIDAKKGQESDLILWLKQSKDNGIE LFKANSDITDIDEALEIIKSFKGWTTYFKGF HENRKNVYSSNDIPTSIIYRIVDDNLPKFLE NKAKYESLKDKAPEAINYEQIKKDLAEEL TFDIDYKTSEVNQRVFSLDEVFEIANFNNY LNQSGITKFNTIIGGKFVNGENTKRKGINE YINLYSQQINDKTLKKYKMSVLFKQILSDT ESKSFVIDKLEDDSDVVTTMQSFYEQIAAF KTVEEKSIKETLSLLFDDLKAQKLDLSKIY FKNDKSLTDLSQQVFDDYSVIGTAVLEYIT QQIAPKNLDNPSKKEQELIAKKTEKAKYLS

[1263]

		LETIKLALEEFNKHRDIDKQCRFEEILANFA AIPMIFDEIAQNKDNLAQISIKYQNQGKKD LLQASAEDDVKAIKDLLDQTNNLLHKLKI FHISQSEDKANILDKDEHFYLVFEECYFEL ANIVPLYNKIRNYITQKPYSDEKFKLNFEN STLANGWDKNKEPDNTAILFIKDDKYYLG VMNKKNNKIFDDKAIKENKGEGYKKIVY KLLPGANKMLPKVFFSAKSIKFYNPSEDIL RIRNHSTHTKNGSPQKGYEKFEFNIEDCRK FIDFYKQSISKHPEWKDFGFRFSDTQRYNSI DEFYREVENQGYKLTFENISESYIDSVVNQ GKLYLFQIYNKDFSAYSKGRPNLHTLYWK ALFDERNLQDVVYKLNGEAELFYRKQSIP KKITHPAKEAIANKNKDNPKKESVFEYDLI KDKRFTEDKFFFHCPITINFKSSGANKFND EINLLLKEKANDVHILSIARGERHLAYYTL VDGKGNIIKQDTFNIIGNDRMKTNYHDKL AAIEKDRDSARKDWKKINNIKEMKEGYLS QVVHEIAKLVIEYNAIVVFEDLNFGFKRGR FKVEKQVYQKLEKMLIEKLNYLVFKDNEF DKTGGVLRAYQLTAPFETFKKMGKQTGII YYVPAGFTSKICPVTGFVNQLYPKYESVSK SQEFFSKFDKICYNLDKGYFEFSFDYKNFG DKAAKGKWTIASFGSRLINFRNSDKNHNW DTREVYPTKELEKLLKDYSIEYGHGECIKA AICGESDKKFFAKLTSVLNTILQMRNSKTG TELDYLISPVADVNGNFFDSRQAPKNMPQ DAAANGAYHIGLKGLMLLGRIKNNQEGK KLNLVIKNEEYFEFVQNRNN
Cpf1 E1006A/ D1255A 多肽序列	252	MSIYQEFVNKYSLSKTLRFELIPQGKTLENI KARGLILDDEKRAKDYKKAKQIIDKYHQF FIEEILSSVCISEDLLQNYSDVYFKLKKSDD DNLQKDFKSAKDTIKKQISEYIKDSEKFKN LFNQNLIDAKKGQESDLILWLKQSKDNGIE LFKANSDITDIDEALEIIKSFKGWTTYFKGF HENRKNVYSSNDIPTSIIYRIVDDNLPKFLE NKAKYESLKDKAPEAINYEQIKKDLAEEL TFDIDYKTSEVNQRVFSLDEVFEIANFNNY LNQSGITKFNTIIGGKFVNGENTKRKGINE YINLYSQQINDKTLKKYKMSVLFKQILSDT ESKSFVIDKLEDDSDVVTTMQSFYEQIAAF KTVEEKSIKETLSLLFDDLKAQKLDLSKIY FKNDKSLTDLSQQVFDDYSVIGTAVLEYIT

[1264]

		QQIAPKNLDNPSKKEQELIAKKTEKAKYLS LETIKLALEEFNKHRDIDKQCRFEEILANFA AIPMIFDEIAQNKDNLAQISIKYQNQGKKD LLQASAEDDVKAIKDLLDQTNNLLHKLKI FHISQSEDKANILDKDEHFYLVFEECYFEL ANIVPLYNKIRNYITQKPYSDEKFKLNFEN STLANGWDKNKEPDNTAILFIKDDKYYLG VMNKKNNKIFDDKAIKENKGEGYKKIVY KLLPGANKMLPKVFFSAKSIKFYNPSEDIL RIRNHSTHTKNGSPQKGYEKFEFNIEDCRK FIDFYKQSISKHPEWKDFGFRFSDTQRYNSI DEFYREVENQGYKLTFENISESYIDSVVNQ GKLYLFQIYNKDFSAYSKGRPNLHTLYWK ALFDERNLQDVVYKLNGEAELFYRKQSIP KKITHPAKEAIANKNKDNPKKESVFEYDLI KDKRFTEDKFFFHCPITINFKSSGANKFND EINLLLKEKANDVHILSIDRGERHLAYYTL VDGKGNIIKQDTFNIIGNDRMKTNYHDKL AAIEKDRDSARKDWKKINNIKEMKEGYLS QVVHEIAKLVIEYNAIVVFADLNFGFKRGR FKVEKQVYQKLEKMLIEKLNYLVFKDNEF DKTGGVLRAYQLTAPFETFKKMGKQTGII YYVPAGFTSKICPVTGFVNQLYPKYESVSK SQEFFSKFDKICYNLDKGYFEFSFDYKNFG DKAAKGKWTIASFGSRLINFRNSDKNHNW DTREVYPTKELEKLLKDYSIEYGHGECIKA AICGESDKKFFAKLTSVLNTILQMRNSKTG TELDYLISPVADVNGNFFDSRQAPKNMPQ DAAANGAYHIGLKGLMLLGRIKNNQEGK KLNLVIKNEEYFEFVQNRNN
Cpf1 D917A/E 1006A/D 1255A 多肽序列	253	MSIYQEFVNKYSLSKTLRFELIPQGKTLENI KARGLILDDEKRAKDYKKAKQIIDKYHQF FIEEILSSVCISEDLLQNYSDVYFKLKKSDD DNLQKDFKSAKDTIKKQISEYIKDSEKFKN LFNQNLIDAKKGQESDLILWLKQSKDNGIE LFKANSDITDIDEALEIIKSFKGWTTYFKGF HENRKNVYSSNDIPTSIIYRIVDDNLPKFLE NKAKYESLKDKAPEAINYEQIKKDLAEEL TFDIDYKTSEVNQRVFSLDEVFEIANFNNY LNQSGITKFNTIIGGKFVNGENTKRKGINE YINLYSQQINDKTLKKYKMSVLFKQILSDT ESKSFVIDKLEDDSDVVTTMQSFYEQIAAF KTVEEKSIKETLSLLFDDLKAQKLDLSKIY

[1265]

		FKNDKSLTDLSQQVFDDYSVIGTAVLEYIT QQIAPKNLDNPSKKEQELIAKKTEKAKYLS LETIKLALEEFNKHRDIDKQCRFEEILANFA AIPMIFDEIAQNKDNLAQISIKYQNQGKKD LLQASAEDDVKAIKDLLDQTNNLLHKLKI FHISQSEDKANILDKDEHFYLVFEECYFEL ANIVPLYNKIRNYITQKPYSDEKFKLNFEN STLANGWDKNKEPDNTAILFIKDDKYYLG VMNKKNNKIFDDKAIKENKGEGYKKIVY KLLPGANKMLPKVFFSAKSIKFYNPSEDIL RIRNHSTHTKNGSPQKGYEKFEFNIEDCRK FIDFYKQSISKHPEWKDFGFRFSDTQRYNSI DEFYREVENQGYKLTFENISESYIDSVVNQ GKLYLFQIYNKDFSAYSKGRPNLHTLYWK ALFDERNLQDVVYKLNGEAELFYRKQSIP KKITHPAKEAIANKNKDNPKKESVFEYDLI KDKRFTEDKFFFHCPITINFKSSGANKFND EINLLLKEKANDVHILSIARGERHLAYYTL VDGKGNIIKQDTFNIIGNDRMKTNYHDKL AAIEKDRDSARKDWKKINNIKEMKEGYLS QVVHEIAKLVIEYNAIVVFADLNFGFKRGR FKVEKQVYQKLEKMLIEKLNYLVFKDNEF DKTGGVLRAYQLTAPFETFKKMGKQTGII YYVPAGFTSKICPVTGFVNQLYPKYESVSK SQEFFSKFDKICYNLDKGYFEFSFDYKNFG DKAAKGKWTIASFGSRLINFRNSDKNHNW DTREVYPTKELEKLLKDYSIEYGHGECIKA AICGESDKKFFAKLTSVLNTILQMRNSKTG TELDYLISPVADVNGNFFDSRQAPKNMPQ DAAANGAYHIGLKGLMLLGRIKNNQEGK KLNLVIKNEEYFEFVQNRNN
合成多肽	254	KRNYILGLDIGITSVGYGIIDYETRDVIDAG VRLFKEANVENNEGRRSKRGARRLKRRR RHRIQRVKKLLFDYNLLTDHSELSGINPYE ARVKGLSQKLSEEEFSAALLHLAKRRGVH NVNEVEEDTGNELSTKEQISRNSKALEEK YVAELQLERLKKDGEVRGSINRFKTSDYV KEAKQLLKVQKAYHQLDQSFIDTYIDLLE TRRTYYEGPGEGSPFGWKDIKEWYEMLM GHCTYFPEELRSVKYAYNADLYNALNDL NNLVITRDENEKLEYYEKFQIIENVFKQKK KPTLKQIAKEILVNEEDIKGYRVTSTGKPE FTNLKVYHDIKDITARKEIIENAELLDQIAK

[1266]

		ILTIYQSSEDIQEELTNLNSELTQEEIEQISN LKGYTGTHNLSLKAINLILDELWHTNDNQI AIFNRLKLVPKKVDLSQQKEIPTTLVDDFIL SPVVKRSFIQSIKVINAIIKKYGLPNDIIIELA REKNSKDAQKMINEMQKRNRQTNERIEEII RTTGKENAKYLIEKIKLHDMQEGKCLYSL EAIPLEDLLNNPFNYEVDHIIPRSVSFDNSF NNKVLVKQEENSKKGNRTPFQYLSSSDSK ISYETFKKHILNLAKGKGRISKTKKEYLLE ERDINRFSVQKDFINRNLVDTRYATRGLM NLLRSYFRVNNLDVKVKSINGGFTSFLRR KWKFKKERNKGYKHHAEDALIIANADFIF KEWKKLDKAKKVMENQMFEEKQAESMP EIETEQEYKEIFITPHQIKHIKDFKDYKYSH RVDKKPNRELINDTLYSTRKDDKGNTLIV NNLNGLYDKDNDKLKKLINKSPEKLLMY HHDPQTYQKLKLIMEQYGDEKNPLYKYY EETGNYLTKYSKKDNGPVIKKIKYYGNKL NAHLDITDDYPNSRNKVVKLSLKPYRFDV YLDNGVYKFVTVKNLDVIKKENYYEVNS KCYEEAKKLKKISNQAEFIASFYNNDLIKI NGELYRVIGVNNDLLNRIEVNMIDITYREY LENMNDKRPPRIIKTIASKTQSIKKYSTDIL GNLYEVKSKKHPQIIKKG
SaCas9n 多肽序列	255	KRNYILGLDIGITSVGYGIIDYETRDVIDAG VRLFKEANVENNEGRRSKRGARRLKRRR RHRIQRVKKLLFDYNLLTDHSELSGINPYE ARVKGLSQKLSEEEFSAALLHLAKRRGVH NVNEVEEDTGNELSTKEQISRNSKALEEK YVAELQLERLKKDGEVRGSINRFKTSDYV KEAKQLLKVQKAYHQLDQSFIDTYIDLLE TRRTYYEGPGEGSPFGWKDIKEWYEMLM GHCTYFPEELRSVKYAYNADLYNALNDL NNLVITRDENEKLEYYEKFQIIENVFKQKK KPTLKQIAKEILVNEEDIKGYRVTSTGKPE FTNLKVYHDIKDITARKEIIENAELLDQIAK ILTIYQSSEDIQEELTNLNSELTQEEIEQISN LKGYTGTHNLSLKAINLILDELWHTNDNQI AIFNRLKLVPKKVDLSQQKEIPTTLVDDFIL SPVVKRSFIQSIKVINAIIKKYGLPNDIIIELA REKNSKDAQKMINEMQKRNRQTNERIEEII RTTGKENAKYLIEKIKLHDMQEGKCLYSL EAIPLEDLLNNPFNYEVDHIIPRSVSFDNSF

			NNKVLVKQEEASKKGNRTPFQYLSSSDSK ISYETFKKHILNLAKGKGRISKTKKEYLLE ERDINRFSVQKDFINRNLVDTRYATRGLM NLLRSYFRVNNLDVKVKSINGGFTSFLRR KWKFKKERNKGYKHHAEDALIIANADFIF KEWKKLDKAKKVMENQMFEEKQAESMP EIETEQEYKEIFITPHQIKHIKDFKDYKYSH RVDKKPNRELINDTLYSTRKDDKGNTLIV NNLNGLYDKDNDKLKKLINKSPEKLLMY HHDPQTYQKLKLIMEQYGDEKNPLYKYY EETGNYLTKYSKKDNGPVIKKIKYYGNKL NAHLDITDDYPNSRNKVVKLSLKPYRFDV YLDNGVYKFVTVKNLDVIKKENYYEVNS KCYEEAKKLKKISNQAEFIASFYNNDLIKI NGELYRVIGVNNDLLNRIEVNMIDITYREY LENMNDKRPPRIIKTIASKTQSIKKYSTDIL GNLYEVKSKKHPQIIKKG
[1267]	SaKKH Cas9 多肽序列	**256**	KRNYILGLDIGITSVGYGIIDYETRDVIDAG VRLFKEANVENNEGRRSKRGARRLKRRR RHRIQRVKKLLFDYNLLTDHSELSGINPYE ARVKGLSQKLSEEEFSAALLHLAKRRGVH NVNEVEEDTGNELSTKEQISRNSKALEEK YVAELQLERLKKDGEVRGSINRFKTSDYV KEAKQLLKVQKAYHQLDQSFIDTYIDLLE TRRTYYEGPGEGSPFGWKDIKEWYEMLM GHCTYFPEELRSVKYAYNADLYNALNDL NNLVITRDENEKLEYYEKFQIIENVFKQKK KPTLKQIAKEILVNEEDIKGYRVTSTGKPE FTNLKVYHDIKDITARKEIIENAELLDQIAK ILTIYQSSEDIQEELTNLNSELTQEEIEQISN LKGYTGTHNLSLKAINLILDELWHTNDNQI AIFNRLKLVPKKVDLSQQKEIPTTLVDDFIL SPVVKRSFIQSIKVINAIIKKYGLPNDIIIELA REKNSKDAQKMINEMQKRNRQTNERIEEII RTTGKENAKYLIEKIKLHDMQEGKCLYSL EAIPLEDLLNNPFNYEVDHIIPRSVSFDNSF NNKVLVKQEEASKKGNRTPFQYLSSSDSK ISYETFKKHILNLAKGKGRISKTKKEYLLE ERDINRFSVQKDFINRNLVDTRYATRGLM NLLRSYFRVNNLDVKVKSINGGFTSFLRR KWKFKKERNKGYKHHAEDALIIANADFIF KEWKKLDKAKKVMENQMFEEKQAESMP EIETEQEYKEIFITPHQIKHIKDFKDYKYSH

[1268]

		RVDKKPNRKLINDTLYSTRKDDKGNTLIV NNLNGLYDKDNDKLKKLINKSPEKLLMY HHDPQTYQKLKLIMEQYGDEKNPLYKYY EETGNYLTKYSKKDNGPVIKKIKYYGNKL NAHLDITDDYPNSRNKVVKLSLKPYRFDV YLDNGVYKFVTVKNLDVIKKENYYEVNS KCYEEAKKLKKISNQAEFIASFYKNDLIKI NGELYRVIGVNNDLLNRIEVNMIDITYREY LENMNDKRPPHIIKTIASKTQSIKKYSTDIL GNLYEVKSKKHPQIIKKG
Casphi-1 多肽序列	285	MADTPTLFTQFLRHHLPGQRFRKDILKQA GRILANKGEDATIAFLRGKSEESPPDFQPP VKCPIIACSRPLTEWPIYQASVAIQGYVYG QSLAEFEASDPGCSKDGLLGWFDKTGVCT DYFSVQGLNLIFQNARKRYIGVQTKVTNR NEKRHKKLKRINAKRIAEGLPELTSDEPES ALDETGHLIDPPGLNTNIYCYQQVSPKPLA LSEVNQLPTAYAGYSTSGDDPIQPMVTKD RLSISKGQPGYIPEHQRALLSQKKHRRMR GYGLKARALLVIVRIQDDWAVIDLRSLLR NAYWRRIVQTKEPSTITKLLKLVTGDPVL DATRMVATFTYKPGIVQVRSAKCLKNKQ GSKLFSERYLNETVSVTSIDLGSNNLVAVA TYRLVNGNTPELLQRFTLPSHLVKDFERY KQAHDTLEDSIQKTAVASLPQGQQTEIRM WSMYGFREAQERVCQELGLADGSIPWNV MTATSTILTDLFLARGGDPKKCMFTSEPK KKKNSKQVLYKIRDRAWAKMYRTLLSKE TREAWNKALWGLKRGSPDYARLSKRKEE LARRCVNYTISTAEKRAQCGRTIVALEDL NIGFFHGRGKQEPGWVGLFTRKKENRWL MQALHKAFLELAHHRGYHVIEVNPAYTS QTCPVCRHCDPDNRDQHNREAFHCIGCGF RGNADLDVATHNIAMVAITGESLKRARGS VASKTPQPLAAE
Casphi-2 多肽序列	286	MPKPAVESEFSKVLKKHFPGERFRSSYMK RGGKILAAQGEEAVVAYLQGKSEEEPPNF QPPAKCHVVTKSRDFAEWPIMKASEAIQR YIYALSTTERAACKPGKSSESHAAWFAAT GVSNHGYSHVQGLNLIFDHTLGRYDGVLK KVQLRNEKARARLESINASRADEGLPEIKA EEEEVATNETGHLLQPPGINPSFYVYQTISP QAYRPRDEIVLPPEYAGYVRDPNAPIPLGV

			VRNRCDIQKGCPGYIPEWQREAGTAISPKT GKAVTVPGLSPKKNKRMRRYWRSEKEKA QDALLVTVRIGTDWVVIDVRGLLRNARW RTIAPKDISLNALLDLFTGDPVIDVRRNIVT FTYTLDACGTYARKWTLKGKQTKATLDK LTATQTVALVAIDLGQTNPISAGISRVTQE NGALQCEPLDRFTLPDDLLKDISAYRIAW DRNEEELRARSVEALPEAQQAEVRALDGV SKETARTQLCADFGLDPKRLPWDKMSSNT TFISEALLSNSVSRDQVFFTPAPKKGAKKK APVEVMRKDRTWARAYKPRLSVEAQKLK NEALWALKRTSPEYLKLSRRKEELCRRSIN YVIEKTRRRTQCQIVIPVIEDLNVRFFHGSG KRLPGWDNFFTAKKENRWFIQGLHKAFSD LRTHRSFYVFEVRPERTSITCPKCGHCEVG NRDGEAFQCLSCGKTCNADLDVATHNLT QVALTGKTMPKREEPRDAQGTAPARKTK KASKSKAPPAEREDQTPAQEPSQTS
[1269]	Casphi-3 多肽序列	287	MEKEITELTKIRREFPNKKFSSTDMKKAGK LLKAEGPDAVRDFLNSCQEIIGDFKPPVKT NIVSISRPFEEWPVSMVGRAIQEYYFSLTK EELESVHPGTSSEDHKSFFNITGLSNYNYT SVQGLNLIFKNAKAIYDGTLVKANNKNKK LEKKFNEINHKRSLEGLPIITPDFEEPFDEN GHLNNPPGINRNIYGYQGCAAKVFVPSKH KMVSLPKEYEGYNRDPNLSLAGFRNRLEI PEGEPGHVPWFQRMDIPEGQIGHVNKIQRF NFVHGKNSGKVKFSDKTGRVKRYHHSKY KDATKPYKFLEESKKVSALDSILAIITIGDD WVVFDIRGLYRNVFYRELAQKGLTAVQL LDLFTGDPVIDPKKGVVTFSYKEGVVPVFS QKIVPRFKSRDTLEKLTSQGPVALLSVDLG QNEPVAARVCSLKNINDKITLDNSCRISFL DDYKKQIKDYRDSLDELEIKIRLEAINSLET NQQVEIRDLDVFSADRAKANTVDMFDIDP NLISWDSMSDARVSTQISDLYLKNGGDES RVYFEINNKRIKRSDYNISQLVRPKLSDST RKNLNDSIWKLKRTSEEYLKLSKRKLELS RAVVNYTIRQSKLLSGINDIVIILEDLDVKK KFNGRGIRDIGWDNFFSSRKENRWFIPAFH KAFSELSSNRGLCVIEVNPAWTSATCPDCG FCSKENRDGINFTCRKCGVSYHADIDVAT LNIARVAVLGKPMSGPADRERLGDTKKPR

[1270]

		VARSRKTMKRKDISNSTVEAMVTA
Casphi-4 多肽序列	288	MYSLEMADLKSEPSLLAKLLRDRFPGKY WLPKYWKLAEKKRLTGGEEAACEYMAD KQLDSPPPNFRPPARCVILAKSRPFEDWPV HRVASKAQSFVIGLSEQGFAALRAAPPSTA DARRDWLRSHGASEDDLMALEAQLLETI MGNAISLHGGVLKKIDNANVKAAKRLSG RNEARLNKGLQELPPEQEGSAYGADGLLV NPPGLNLNIYCRKSCCPKPVKNTARFVGH YPGYLRDSDSILISGTMDRLTIIEGMPGHIP AWQREQGLVKPGGRRRRLSGSESNMRQK VDPSTGPRRSTRSGTVNRSNQRTGRNGDP LLVEIRMKEDWVLLDARGLLRNLRWRES KRGLSCDHEDLSLSGLLALFSGDPVIDPVR NEVVFLYGEGIIPVRSTKPVGTRQSKKLLE RQASMGPLTLISCDLGQTNLIAGRASAISL THGSLGVRSSVRIELDPEIIKSFERLRKDAD RLETEILTAAKETLSDEQRGEVNSHEKDSP QTAKASLCRELGLHPPSLPWGQMGPSTTFI ADMLISHGRDDDAFLSHGEFPTLEKRKKF DKRFCLESRPLLSSETRKALNESLWEVKRT SSEYARLSQRKKEMARRAVNFVVEISRRK TGLSNVIVNIEDLNVRIFHGGGKQAPGWD GFFRPKSENRWFIQAIHKAFSDLAAHHGIP VIESDPQRTSMTCPECGHCDSKNRNGVRF LCKGCGASMDADFDAACRNLERVALTGK PMPKPSTSCERLLSATTGKVCSDHSLSHDA IEKAS
Casphi-5 多肽序列	289	MSSLPTPLELLKQKHADLFKGLQFSSKDN KMAGKVLKKDGEEAALAFLSERGVSRGE LPNFRPPAKTLVVAQSRPFEEFPIYRVSEAI QLYVYSLSVKELETVPSGSSTKKEHQRFFQ DSSVPDFGYTSVQGLNKIFGLARGIYLGVI TRGENQLQKAKSKHEALNKKRRASGEAE TEFDPTPYEYMTPERKLAKPPGVNHSIMC YVDISVDEFDFRNPDGIVLPSEYAGYCREI NTAIEKGTVDRLGHLKGGPGYIPGHQRKE STTEGPKINFRKGRIRRSYTALYAKRDSRR VRQGKLALPSYRHHMMRLNSNAESAILA VIFFGKDWVVFDLRGLLRNVRWRNLFVD GSTPSTLLGMFGDPVIDPKRGVVAFCYKE QIVPVVSKSITKMVKAPELLNKLYLKSEDP LVLVAIDLGQTNPVGVGVYRVMNASLDY

[1271]

		EVVTRFALESELLREIESYRQRTNAFEAQIR AETFDAMTSEEQEEITRVRAFSASKAKEN VCHRFGMPVDAVDWATMGSNTIHIAKWV MRHGDPSLVEVLEYRKDNEIKLDKNGVPK KVKLTDKRIANLTSIRLRFSQETSKHYNDT MWELRRKHPVYQKLSKSKADFSRRVVNSI IRRVNHLVPRARIVFIIEDLKNLGKVFHGS GKRELGWDSYFEPKSENRWFIQVLHKAFS ETGKHKGYYIIECWPNWTSCTCPKCSCCD SENRHGEVFRCLACGYTCNTDFGTAPDNL VKIATTGKGLPGPKKRCKGSSKGKNPKIA RSSETGVSVTESGAPKVKKSSPTQTSQSSS QSAP
Casphi-6 多肽序列	290	MNKIEKEKTPLAKLMNENFAGLRFPFAIIK QAGKKLLKEGELKTIEYMTGKGSIEPLPNF KPPVKCLIVAKRRDLKYFPICKASCEIQSY VYSLNYKDFMDYFSTPMTSQKQHEEFFKK SGLNIEYQNVAGLNLIFNNVKNTYNGVIL KVKNRNEKLKKKAIKNNYEFEEIKTFNDD GCLINKPGINNVIYCFQSISPKILKNITHLPK EYNDYDCSVDRNIIQKYVSRLDIPESQPGH VPEWQRKLPEFNNTNNPRRRRKWYSNGR NISKGYSVDQVNQAKIEDSLLAQIKIGEDW IILDIRGLLRDLNRRELISYKNKLTIKDVLG FFSDYPIIDIKKNLVTFCYKEGVIQVVSQKS IGNKKSKQLLEKLIENKPIALVSIDLGQTNP VSVKISKLNKINNKISIESFTYRFLNEEILKE IEKYRKDYDKLELKLINEA
Casphi-7 多肽序列	291	MSNTAVSTREHMSNKTTPPSPLSLLLRAHF PGLKFESQDYKIAGKKLRDGGPEAVISYLT GKGQAKLKDVKPPAKAFVIAQSRPFIEWD LVRVSRQIQEKIFGIPATKGRPKQDGLSET AFNEAVASLEVDGKSKLNEETRAAFYEVL GLDAPSLHAQAQNALIKSAISIREGVLKKV ENRNEKNLSKTKRRKEAGEEATFVEEKAH DERGYLIHPPGVNQTIPGYQAVVIKSCPSD FIGLPSGCLAKESAEALTDYLPHDRMTIPK GQPGYVPEWQHPLLNRRKNRRRRDWYSA SLNKPKATCSKRSGTPNRKNSRTDQIQSGR FKGAIPVLMRFQDEWVIIDIRGLLRNARYR KLLKEKSTIPDLLSLFTGDPSIDMRQGVCT FIYKAGQACSAKMVKTKNAPEILSELTKS GPVVLVSIDLGQTNPIAAKVSRVTQLSDGQ

[1272]

		LSHETLLRELLSNDSSDGKEIARYRVASDR LRDKLANLAVERLSPEHKSEILRAKNDTPA LCKARVCAALGLNPEMIAWDKMTPYTEF LATAYLEKGGDRKVATLKPKNRPEMLRR DIKFKGTEGVRIEVSPEAAEAYREAQWDL QRTSPEYLRLSTWKQELTKRILNQLRHKA AKSSQCEVVVMAFEDLNIKMMHGNGKW ADGGWDAFFIKKRENRWFMQAFHKSLTE LGAHKGVPTIEVTPHRTSITCTKCGHCDKA NRDGERFACQKCGFVAHADLEIATDNIER VALTGKPMPKPESERSGDAKKSVGARKA AFKPEEDAEAAE
Casphi-8 多肽序列	292	MIKPTVSQFLTPGFKLIRNHSRTAGLKLKN EGEEACKKFVRENEIPKDECPNFQGGPAIA NIIAKSREFTEWEIYQSSLAIQEVIFTLPKDK LPEPILKEEWRAQWLSEHGLDTVPYKEAA GLNLIIKNAVNTYKGVQVKVDNKNKNNL AKINRKNEIAKLNGEQEISFEEIKAFDDKG YLLQKPSPNKSIYCYQSVSPKPFITSKYHN VNLPEEYIGYYRKSNEPIVSPYQFDRLRIPI GEPGYVPKWQYTFLSKKENKRRKLSKRIK NVSPILGIICIKKDWCVFDMRGLLRTNHW KKYHKPTDSINDLFDYFTGDPVIDTKANV VRFRYKMENGIVNYKPVREKKGKELLENI CDQNGSCKLATVDVGQNNPVAIGLFELKK VNGELTKTLISRHPTPIDFCNKITAYRERYD KLESSIKLDAIKQLTSEQKIEVDNYNNNFT PQNTKQIVCSKLNINPNDLPWDKMISGTHF ISEKAQVSNKSEIYFTSTDKGKTKDVMKS DYKWFQDYKPKLSKEVRDALSDIEWRLR RESLEFNKLSKSREQDARQLANWISSMCD VIGIENLVKKNNFFGGSGKREPGWDNFYK PKKENRWWINAIHKALTELSQNKGKRVIL LPAMRTSITCPKCKYCDSKNRNGEKFNCL KCGIELNADIDVATENLATVAITAQSMPKP TCERSGDAKKPVRARKAKAPEFHDKLAPS YTVVLREAV
Casphi-9 多肽序列	293	MRSSREIGDKILMRQPAEKTAFQVFRQEVI GTQKLSGGDAKTAGRLYKQGKMEAARE WLLKGARDDVPPNFQPPAKCLVVAVSHPF EEWDISKTNHDVQAYIYAQPLQAEGHLNG LSEKWEDTSADQHKLWFEKTGVPDRGLP VQAINKIAKAAVNRAFGVVRKVENRNEK

[1273]

		RRSRDNRIAEHNRENGLTEVVREAPEVAT NADGFLLHPPGIDPSILSYASVSPVPYNSSK HSFVRLPEEYQAYNVEPDAPIPQFVVEDRF AIPPGQPGYVPEWQRLKCSTNKHRRMRQ WSNQDYKPKAGRRAKPLEFQAHLTRERA KGALLVVMRIKEDWVVFDVRGLLRNVEW RKVLSEEAREKLTLKGLLDLFTGDPVIDTK RGIVTFLYKAEITKILSKRTVKTKNARDLL LRLTEPGEDGLRREVGLVAVDLGQTHPIA AAIYRIGRTSAGALESTVLHRQGLREDQKE KLKEYRKRHTALDSRLRKEAFETLSVEQQ KEIVTVSGSGAQITKDKVCNYLGVDPSTLP WEKMGSYTHFISDDFLRRGGDPNIVHFDR QPKKGKVSKKSQRIKRSDSQWVGRMRPR LSQETAKARMEADWAAQNENEEYKRLAR SKQELARWCVNTLLQNTRCITQCDEIVVVI EDLNVKSLHGKGAREPGWDNFFTPKTENR WFIQILHKTFSELPKHRGEHVIEGCPLRTSI TCPACSYCDKNSRNGEKFVCVACGATFHA DFEVATYNLVRLATTGMPMPKSLERQGG GEKAGGARKARKKAKQVEKIVVQANAN VTMNGASLHSP
Casphi-10 多肽序列	294	MDMLDTETNYATETPAQQQDYSPKPPKK AQRAPKGFSKKARPEKKPPKPITLFTQKHF SGVRFLKRVIRDASKILKLSESRTITFLEQAI ERDGSAPPDVTPPVHNTIMAVTRPFEEWPE VILSKALQKHCYALTKKIKIKTWPKKGPG KKCLAAWSARTKIPLIPGQVQATNGLFDRI GSIYDGVEKKVTNRNANKKLEYDEAIKEG RNPAVPEYETAYNIDGTLINKPGYNPNLYI TQSRTPRLITEADRPLVEKILWQMVEKKT QSRNQARRARLEKAAHLQGLPVPKFVPEK VDRSQKIEIRIIDPLDKIEPYMPQDRMAIKA SQDGHVPYWQRPFLSKRRNRRVRAGWGK QVSSIQAWLTGALLVIVRLGNEAFLADIRG ALRNAQWRKLLKPDATYQSLFNLFTGDPV VNTRTNHLTMAYREGVVNIVKSRSFKGRQ TREHLLTLLGQGKTVAGVSFDLGQKHAA GLLAAHFGLGEDGNPVFTPIQACFLPQRYL DSLTNYRNRYDALTLDMRRQSLLALTPAQ QQEFADAQRDPGGQAKRACCLKLNLNPD EIRWDLVSGISTMISDLYIERGGDPRDVHQ QVETKPKGKRKSEIRILKIRDGKWAYDFRP

[1274]

		KIADETRKAQREQLWKLQKASSEFERLSR YKINIARAIANWALQWGRELSGCDIVIPVL EDLNVGSKFFDGKGKWLLGWDNRFTPKK ENRWFIKVLHKAVAELAPHRGVPVYEVM PHRTSMTCPACHYCHPTNREGDRFECQSC HVVKNTDRDVAPYNILRVAVEGKTLDRW QAEKKPQAEPDRPMILIDNQES
>splP14739IUNGI_BPPB2 尿嘧啶-DNA 糖苷酶抑制剂	106	MTNLSDIIEKETGKQLVIQESILMLPEEVEE VIGNKPESDILVHTAYDESTDENVMLLTSD APEYKPWALVIQDSNGENKIKML
Cas12b/C2c1	258	MAVKSIKVKLRLDDMPEIRAGLWKLHKE VNAGVRYYTEWLSLLRQENLYRRSPNGD GEQECDKTAEECKAELLERLRARQVENGH RGPAGSDDELLQLARQLYELLVPQAIGAK GDAQQIARKFLSPLADKDAVGGLGIAKAG NKPRWVRMREAGEPGWEEEKEKAETRKS ADRTADVLRALADFGLKPLMRVYTDSEM SSVEWKPLRKGQAVRTWDRDMFQQAIER MMSWESWNQRVGQEYAKLVEQKNRFEQ KNFVGQEHLVHLVNQLQQDMKEASPGLE SKEQTAHYVTGRALRGSDKVFEKWGKLA PDAPFDLYDAEIKNVQRRNTRRFGSHDLF AKLAEPEYQALWREDASFLTRYAVYNSIL RKLNHAKMFATFTLPDATAHPIWTRFDKL GGNLHQYTFLFNEFGERRHAIRFHKLLKV ENGVAREVDDVTVPISMSEQLDNLLPRDP NEPIALYFRDYGAEQHFTGEFGGAKIQCRR DQLAHMHRRRGARDVYLNVSVRVQSQSE ARGERRPPYAAVFRLVGDNHRAFVHFDK LSDYLAEHPDDGKLGSEGLLSGLRVMSVD LGLRTSASISVFRVARKDELKPNSKGRVPF FFPIKGNDNLVAVHERSQLLKLPGETESKD LRAIREERQRTLRQLRTQLAYLRLLVRCGS EDVGRRERSWAKLIEQPVDAANHMTPDW REAFENELQKLKSLHGICSDKEWMDAVYE SVRRVWRHMGKQVRDWRKDVRSGERPKI RGYAKDVVGGNSIEQIEYLERQYKFLKSW SFFGKVSGQVIRAEKGSRFAITLREHIDHA

[1275]

		KEDRLKKLADRIIMEALGYVYALDERGKG KWVAKYPPCQLILLEELSEYQFNNDRPPSE NNQLMQWSHRGVFQELINQAQVHDLLVG TMYAAFSSRFDARTGAPGIRCRRVPARCT QEHNPEPFPWWLNKFVVEHTLDACPLRA DDLIPTGEGEIFVSPFSAEEGDFHQIHADLN AAQNLQQRLWSDFDISQIRLRCDWGEVDG ELVLIPRLTGKRTADSYSNKVFYTNTGVT YYERERGKKRRKVFAQEKLSEEEAELLVE ADEAREKSVVLMRDPSGIINRGNWTRQKE FWSMVNQRIEGYLVKQIRSRVPLQDSACE NTGDI
高保真 Cas9 多肽序列	1423	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTAFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGALSRKLINGIRDKQSGKTILDFLK SDGFANRNFMALIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI

[1276]

		KRQLVETRAITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFDSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIE QISEFSKRVILADANLDKVLSAYNKHRDK PIREQAENIIHLFTLTNLGAPAAFKYFDTTI DRKRYTSTKEVLDATLIHQSITGLYETRID LSQLGGD
	Wt Cas9 结构域	233	ATGGATAAAAAGTATTCTATTGGTTTAG ACATCGGCACTAATTCCGTTGGATGGGC TGTCATAACCGATGAATACAAAGTACCT TCAAAGAAATTTAAGGTGTTGGGGAACA CAGACCGTCATTCGATTAAAAAGAATCT TATCGGTGCCCTCCTATTCGATAGTGGC GAAACGGCAGAGGCGACTCGCCTGAAA CGAACCGCTCGGAGAAGGTATACACGTC GCAAGAACCGAATATGTTACTTACAAGA AATTTTTAGCAATGAGATGGCCAAAGTT GACGATTCTTTCTTTCACCGTTTGGAAGA GTCCTTCCTTGTCGAAGAGGACAAGAAA CATGAACGGCACCCCATCTTTGGAAACA TAGTAGATGAGGTGGCATATCATGAAAA GTACCCAACGATTTATCACCTCAGAAAA AAGCTAGTTGACTCAACTGATAAAGCGG ACCTGAGGTTAATCTACTTGGCTCTTGCC CATATGATAAAGTTCCGTGGGCACTTTC TCATTGAGGGTGATCTAAATCCGGACAA CTCGGATGTCGACAAACTGTTCATCCAG TTAGTACAAACCTATAATCAGTTGTTTG AAGAGAACCCTATAAATGCAAGTGGCGT GGATGCGAAGGCTATTCTTAGCGCCCGC CTCTCTAAATCCCGACGGCTAGAAAACC TGATCGCACAATTACCCGGAGAGAAGAA AAATGGGTTGTTCGGTAACCTTATAGCG CTCTCACTAGGCCTGACACCAAATTTTA

[1277]

		AGTCGAACTTCGACTTAGCTGAAGATGC CAAATTGCAGCTTAGTAAGGACACGTAC GATGACGATCTCGACAATCTACTGGCAC AAATTGGAGATCAGTATGCGGACTTATT TTTGGCTGCCAAAAACCTTAGCGATGCA ATCCTCCTATCTGACATACTGAGAGTTA ATACTGAGATTACCAAGGCGCCGTTATC CGCTTCAATGATCAAAAGGTACGATGAA CATCACCAAGACTTGACACTTCTCAAGG CCCTAGTCCGTCAGCAACTGCCTGAGAA ATATAAGGAAATATTCTTTGATCAGTCG AAAAACGGGTACGCAGGTTATATTGACG GCGGAGCGAGTCAAGAGGAATTCTACA AGTTTATCAAACCCATATTAGAGAAGAT GGATGGGACGGAAGAGTTGCTTGTAAAA CTCAATCGCGAAGATCTACTGCGAAAGC AGCGGACTTTCGACAACGGTAGCATTCC ACATCAAATCCACTTAGGCGAATTGCAT GCTATACTTAGAAGGCAGGAGGATTTTT ATCCGTTCCTCAAAGACAATCGTGAAAA GATTGAGAAAATCCTAACCTTTCGCATA CCTTACTATGTGGGACCCCTGGCCCGAG GGAACTCTCGGTTCGCATGGATGACAAG AAAGTCCGAAGAAACGATTACTCCATGG AATTTTGAGGAAGTTGTCGATAAAGGTG CGTCAGCTCAATCGTTCATCGAGAGGAT GACCAACTTTGACAAGAATTTACCGAAC GAAAAAGTATTGCCTAAGCACAGTTTAC TTTACGAGTATTTCACAGTGTACAATGA ACTCACGAAAGTTAAGTATGTCACTGAG GGCATGCGTAAACCCGCCTTTCTAAGCG GAGAACAGAAGAAAGCAATAGTAGATC TGTTATTCAAGACCAACCGCAAAGTGAC AGTTAAGCAATTGAAAGAGGACTACTTT AAGAAAATTGAATGCTTCGATTCTGTCG AGATCTCCGGGGTAGAAGATCGATTTAA TGCGTCACTTGGTACGTATCATGACCTCC TAAAGATAATTAAAGATAAGGACTTCCT GGATAACGAAGAGAATGAAGATATCTTA GAAGATATAGTGTTGACTCTTACCCTCTT TGAAGATCGGGAAATGATTGAGGAAAG ACTAAAAACATACGCTCACCTGTTCGAC GATAAGGTTATGAAACAGTTAAAGAGGC

[1278]

		GTCGCTATACGGGCTGGGGACGATTGTC GCGGAAACTTATCAACGGGATAAGAGA CAAGCAAAGTGGTAAAACTATTCTCGAT TTTCTAAAGAGCGACGGCTTCGCCAATA GGAACTTTATGCAGCTGATCCATGATGA CTCTTTAACCTTCAAAGAGGATATACAA AAGGCACAGGTTTCCGGACAAGGGGACT CATTGCACGAACATATTGCGAATCTTGC TGGTTCGCCAGCCATCAAAAAGGGCATA CTCCAGACAGTCAAAGTAGTGGATGAGC TAGTTAAGGTCATGGGACGTCACAAACC GGAAAACATTGTAATCGAGATGGCACGC GAAAATCAAACGACTCAGAAGGGGCAA AAAAACAGTCGAGAGCGGATGAAGAGA ATAGAAGAGGGTATTAAAGAACTGGGC AGCCAGATCTTAAAGGAGCATCCTGTGG AAAATACCCAATTGCAGAACGAGAAACT TTACCTCTATTACCTACAAAATGGAAGG GACATGTATGTTGATCAGGAACTGGACA TAAACCGTTTATCTGATTACGACGTCGA TCACATTGTACCCCAATCCTTTTTGAAGG ACGATTCAATCGACAATAAAGTGCTTAC ACGCTCGGATAAGAACCGAGGGAAAAG TGACAATGTTCCAAGCGAGGAAGTCGTA AAGAAAATGAAGAACTATTGGCGGCAG CTCCTAAATGCGAAACTGATAACGCAAA GAAAGTTCGATAACTTAACTAAAGCTGA GAGGGGTGGCTTGTCTGAACTTGACAAG GCCGGATTTATTAAACGTCAGCTCGTGG AAACCCGCCAAATCACAAAGCATGTTGC ACAGATACTAGATTCCCGAATGAATACG AAATACGACGAGAACGATAAGCTGATTC GGGAAGTCAAAGTAATCACTTTAAAGTC AAAATTGGTGTCGGACTTCAGAAAGGAT TTTCAATTCTATAAAGTTAGGGAGATAA ATAACTACCACCATGCGCACGACGCTTA TCTTAATGCCGTCGTAGGGACCGCACTC ATTAAGAAATACCCGAAGCTAGAAAGTG AGTTTGTGTATGGTGATTACAAAGTTTAT GACGTCCGTAAGATGATCGCGAAAAGCG AACAGGAGATAGGCAAGGCTACAGCCA AATACTTCTTTTATTCTAACATTATGAAT TTCTTTAAGACGGAAATCACTCTGGCAA

[1279]

		ACGGAGAGATACGCAAACGACCTTTAAT TGAAACCAATGGGGAGACAGGTGAAAT CGTATGGGATAAGGGCCGGGACTTCGCG ACGGTGAGAAAAGTTTTGTCCATGCCCC AAGTCAACATAGTAAAGAAAACTGAGG TGCAGACCGGAGGGTTTTCAAAGGAATC GATTCTTCCAAAAAGGAATAGTGATAAG CTCATCGCTCGTAAAAAGGACTGGGACC CGAAAAAGTACGGTGGCTTCGATAGCCC TACAGTTGCCTATTCTGTCCTAGTAGTGG CAAAAGTTGAGAAGGGAAAATCCAAGA AACTGAAGTCAGTCAAAGAATTATTGGG GATAACGATTATGGAGCGCTCGTCTTTT GAAAAGAACCCCATCGACTTCCTTGAGG CGAAAGGTTACAAGGAAGTAAAAAAGG ATCTCATAATTAAACTACCAAAGTATAG TCTGTTTGAGTTAGAAAATGGCCGAAAA CGGATGTTGGCTAGCGCCGGAGAGCTTC AAAAGGGGAACGAACTCGCACTACCGTC TAAATACGTGAATTTCCTGTATTTAGCGT CCCATTACGAGAAGTTGAAAGGTTCACC TGAAGATAACGAACAGAAGCAACTTTTT GTTGAGCAGCACAAACATTATCTCGACG AAATCATAGAGCAAATTTCGGAATTCAG TAAGAGAGTCATCCTAGCTGATGCCAAT CTGGACAAAGTATTAAGCGCATACAACA AGCACAGGGATAAACCCATACGTGAGC AGGCGGAAAATATTATCCATTTGTTTAC TCTTACCAACCTCGGCGCTCCAGCCGCA TTCAAGTATTTTGACACAACGATAGATC GCAAACGATACACTTCTACCAAGGAGGT GCTAGACGCGACACTGATTCACCAATCC ATCACGGGATTATATGAAACTCGGATAG ATTTGTCACAGCTTGGGGGTGACGGATC CCCCAAGAAGAAGAGGAAAGTCTCGAG CGACTACAAAGACCATGACGGTGATTAT AAAGATCATGACATCGATTACAAGGATG ACGATGACAAGGCTGCAGGA
野生型 Cas9 多肽序列	234	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK

[1280]

		ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFDSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIE QISEFSKRVILADANLDKVLSAYNKHRDK PIREQAENIIHLFTLTNLGAPAAFKYFDTTI DRKRYTSTKEVLDATLIHQSITGLYETRID LSQLGGD
结合	1304	MDKKYSIGLAIGTNSVGWAVITDEYKVPS

[1281]

PAM 的 SpEQR Cas9 多肽序列		KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESVLVEEDKKHERHPI FGNIVDEVAYHEKYPTIYHLRKKLVDSTD KADLRLIYLALAHMIKFRGHFLIEGDLNPD NSDVDKLFIQLVQTYNQLFEENPINASGVD AKAILSARLSKSRRLENLIAQLPGEKKNGL FGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKN LSDAILLSDILRVNTEITKAPLSASMIKRYD EHHQDLTLLKALVRQQLPEKYKEIFFDQS KNGYAGYIDGGASQEEFYKFIKPILEKMD GTEELLVKLNREDLLRKQRTFDNGSIPHQI HLGELHAILRRQEDFYPFLKDNREKIEKILT FRIPYYVGPLARGNSRFAWMTRKSEETITP WNFEEVVDKGASAQSFIERMTNFDKNLPN EKVLPKHSLLYEYFTVYNELTKVKYVTEG MRKPAFLSGEQKKAIVDLLFKTNRKVTVK QLKEDYFKKIECFDSVEISGVEDRFNASLG TYHDLLKIIKDKDFLDNEENEDILEDIVLTL TLFEDREMIEERLKTYAHLFDDKVMKQLK RRRYTGWGRLSRKLINGIRDKQSGKTILDF LKSDGFANRNFMQLIHDDSLTFKEDIQKA QVSGQGDSLHEHIANLAGSPAIKKGILQTV KVVDELVKVMGRHKPENIVIEMARENQTT QKGQKNSRERMKRIEEGIKELGSQILKEHP VENTQLQNEKLYLYYLQNGRDMYVDQEL DINRLSDYDVDHIVPQSFLKDDSIDNKVLT RSDKNRGKSDNVPSEEVVKKMKNYWRQL LNAKLITQRKFDNLTKAERGGLSELDKAG FIKRQLVETRQITKHVAQILDSRMNTKYDE NDKLIREVKVITLKSKLVSDFRKDFQFYKV REINNYHHAHDAYLNAVVGTALIKKYPKL ESEFVYGDYKVYDVRKMIAKSEQEIGKAT AKYFFYSNIMNFFKTEITLANGEIRKRPLIE TNGETGEIVWDKGRDFATVRKVLSMPQV NIVKKTEVQTGGFSKESILPKRNSDKLIAR KKDWDPKKYGGFESPTVAYSVLVVAKVE KGKSKKLKSVKELLGITIMERSSFEKNPIDF LEAKGYKEVKKDLIIKLPKYSLFELENGRK RMLASAGELQKGNELALPSKYVNFLYLAS HYEKLKGSPEDNEQKQLFVEQHKHYLDEI IEQISEFSKRVILADANLDKVLSAYNKHRD

[1282]

		KPIREQAENIIHLFTLTNLGAPAAFKYFDTT IDRKQYRSTKEVLDATLIHQSITGLYETRID LSQLGGD
结合 PAM 的 SpVQR Cas9 多肽序列	1305	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFVSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE

[1283]

		AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIE QISEFSKRVILADANLDKVLSAYNKHRDK PIREQAENIIHLFTLTNLGAPAAFKYFDTTI DRKQYRSTKEVLDATLIHQSITGLYETRID LSQLGGD
SpVQR Cas9 多肽序列	1306	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET

[1284]

		NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFVSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASARELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIE QISEFSKRVILADANLDKVLSAYNKHRDK PIREQAENIIHLFTLTNLGAPAAFKYFDTTI DRKEYRSTKEVLDATLIHQSITGLYETRID LSQLGGD
SpyMacCas9 多肽序列	1307	MDKKYSIGLDIGTNSVGWAVITDDYKVPS KKFKVLGNTDRHSIKKNLIGALLFGSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLADSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQIYNQLFEENPINASRVDA KAILSARLSKSRRLENLIAQLPGEKRNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNSEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGAY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDRGMIEERLKTYAHLFDDKVMKQLKR RRYTGWGRLSRKLINGIRDKQSGKTILDFL KSDGFANRNFMQLIHDDSLTFKEDIQKAQ VSGQGHSLHEQIANLAGSPAIKKGILQTVK IVDELVKVMGHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFIKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN

[1285]

		DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEIQTVGQNGGLFDDNPKSPLEVTPS KLVPLKKELNPKKYGGYQKPTTAYPVLLI TDTKQLIPISVMNKKQFEQNPVKFLRDRG YQQVGKNDFIKLPKYTLVDIGDGIKRLWA SSKEIHKGNQLVVSKKSQILLYHAHHLDS DLSNDYLQNHNQQFDVLFNEIISFSKKCKL GKEHIQKIENVYSNKKNSASIEELAESFIKL LGFTQLGATSPFNFLGVKLNQKQYKGKKD YILPCTEGTLIRQSITGLYETRVDLSKIGED
CP5 多肽序列	257	EIGKATAKYFFYSNIMNFFKTEITLANGEIR KRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSD KLIARKKDWDPKKYGGFMQPTVAYSVLV VAKVEKGKSKKLKSVKELLGITIMERSSFE KNPIDFLEAKGYKEVKKDLIIKLPKYSLFE LENGRKRMLASAKFLQKGNELALPSKYV NFLYLASHYEKLKGSPEDNEQKQLFVEQH KHYLDEIIEQISEFSKRVILADANLDKVLSA YNKHRDKPIREQAENIIHLFTLTNLGAPRA FKYFDTTIARKEYRSTKEVLDATLIHQSITG LYETRIDLSQLGGDGGSGGSGGSGGSGGS GGSGGMDKKYSIGLAIGTNSVGWAVITDE YKVPSKKFKVLGNTDRHSIKKNLIGALLFD SGETAEATRLKRTARRRYTRRKNRICYLQ EIFSNEMAKVDDSFFHRLEESFLVEEDKKH ERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLRLIYLALAHMIKFRGHFLIEGD LNPDNSDVDKLFIQLVQTYNQLFEENPINA SGVDAKAILSARLSKSRRLENLIAQLPGEK KNGLFGNLIALSLGLTPNFKSNFDLAEDAK LQLSKDTYDDDLDNLLAQIGDQYADLFLA AKNLSDAILLSDILRVNTEITKAPLSASMIK RYDEHHQDLTLLKALVRQQLPEKYKEIFF DQSKNGYAGYIDGGASQEEFYKFIKPILEK MDGTEELLVKLNREDLLRKQRTFDNGSIP HQIHLGELHAILRRQEDFYPFLKDNREKIE KILTFRIPYYVGPLARGNSRFAWMTRKSEE TITPWNFEEVVDKGASAQSFIERMTNFDK

[1286]

		NLPNEKVLPKHSLLYEYFTVYNELTKVKY VTEGMRKPAFLSGEQKKAIVDLLFKTNRK VTVKQLKEDYFKKIECFDSVEISGVEDRFN ASLGTYHDLLKIIKDKDFLDNEENEDILEDI VLTLTLFEDREMIEERLKTYAHLFDDKVM KQLKRRRYTGWGRLSRKLINGIRDKQSGK TILDFLKSDGFANRNFMQLIHDDSLTFKED IQKAQVSGQGDSLHEHIANLAGSPAIKKGI LQTVKVVDELVKVMGRHKPENIVIEMARE NQTTQKGQKNSRERMKRIEEGIKELGSQIL KEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDN KVLTRSDKNRGKSDNVPSEEVVKKMKNY WRQLLNAKLITQRKFDNLTKAERGGLSEL DKAGFIKRQLVETRQITKHVAQILDSRMN TKYDENDKLIREVKVITLKSKLVSDFRKDF QFYKVREINNYHHAHDAYLNAVVGTALI KKYPKLESEFVYGDYKVYDVRKMIAKSE QEGADKRTADGSEFESPKKKRKV
Cas12c1 多肽序列	266	MQTKKTHLHLISAKASRKYRRTIACLSDT AKKDLERRKQSGAADPAQELSCLKTIKFK LEVPEGSKLPSFDRISQIYNALETIEKGSLS YLLFALILSGFRIFPNSSAAKTFASSSCYKN DQFASQIKEIFGEMVKNFIPSELESILKKGR RKNNKDWTEENIKRVLNSEFGRKNSEGSS ALFDSFLSKFSQELFRKFDSWNEVNKKYL EAAELLDSMLASYGPFDSVCKMIGDSDSR NSLPDKSTIAFTNNAEITVDIESSVMPYMAI AALLREYRQSKSKAAPVAYVQSHLTTTNG NGLSWFFKFGLDLIRKAPVSSKQSTSDGSK SLQELFSVPDDKLDGLKFIKEACEALPEAS LLCGEKGELLGYQDFRTSFAGHIDSWVAN YVNRLFELIELVNQLPESIKLPSILTQKNHN LVASLGLQEAEVSHSLELFEGLVKNVRQT LKKLAGIDISSSPNEQDIKEFYAFSDVLNRL GSIRNQIENAVQTAKKDKIDLESAIEWKE WKKLKKLPKLNGLGGGVPKQQELLDKAL ESVKQIRHYQRIDFERVIQWAVNEHCLET VPKFLVDAEKKKINKESSTDFAAKENAVR FLLEGIGAAARGKTDSVSKAAYNWFVVN NFLAKKDLNRYFINCQGCIYKPPYSKRRSL AFALRSDNKDTIEVVWEKFETFYKEISKEI EKFNIFSQEFQTFLHLENLRMKLLLRRIQK

			PIPAEIAFFSLPQEYYDSLPPNVAFLALNQE ITPSEYITQFNLYSSFLNGNLILLRRSRSYLR AKFSWVGNSKLIYAAKEARLWKIPNAYW KSDEWKMILDSNVLVFDKAGNVLPAPTLK KVCEREGDLRLFYPLLRQLPHDWCYRNPF VKSVGREKNVIEVNKEGEPKVASALPGSL FRLIGPAPFKSLLDDCFFNPLDKDLRECMLI VDQEISQKVEAQKVEASLESCTYSIAVPIR YHLEEPKVSNQFENVLAIDQGEAGLAYAV FSLKSIGEAETKPIAVGTIRIPSIRRLIHSVST YRKKKQRLQNFKQNYDSTAFIMRENVTG DVCAKIVGLMKEFNAFPVLEYDVKNLESG SRQLSAVYKAVNSHFLYFKEPGRDALRKQ LWYGGDSWTIDGIEIVTRERKEDGKEGVE KIVPLKVFPGRSVSARFTSKTCSCCGRNVF DWLFTEKKAKTNKKFNVNSKGELTTADG VIQLFEADRSKGPKFYARRKERTPLTKPIA KGSYSLEEIERRVRTNLRRAPKSKQSRDTS QSQYFCVYKDCALHFSGMQADENAAINIG RRFLTALRKNRRSDFPSNVKISDRLLDN
[1287]	Cas12c2 多肽序列	267	MTKHSIPLHAFRNSGADARKWKGRIALLA KRGKETMRTLQFPLEMSEPEAAAINTTPFA VAYNAIEGTGKGTLFDYWAKLHLAGFRFF PSGGAATIFRQQAVFEDASWNAAFCQQSG KDWPWLVPSKLYERFTKAPREVAKKDGS KKSIEFTQENVANESHVSLVGASITDKTPE DQKEFFLKMAGALAEKFDSWKSANEDRI VAMKVIDEFLKSEGLHLPSLENIAVKCSVE TKPDNATVAWHDAPMSGVQNLAIGVFAT CASRIDNIYDLNGGKLSKLIQESATTPNVT ALSWLFGKGLEYFRTTDIDTIMQDFNIPAS AKESIKPLVESAQAIPTMTVLGKKNYAPFR PNFGGKIDSWIANYASRLMLLNDILEQIEP GFELPQALLDNETLMSGIDMTGDELKELIE AVYAWVDAAKQGLATLLGRGGNVDDAV QTFEQFSAMMDTLNGTLNTISARYVRAVE MAGKDEARLEKLIECKFDIPKWCKSVPKL VGISGGLPKVEEEIKVMNAAFKDVRARMF VRFEEIAAYVASKGAGMDVYDALEKREL EQIKKLKSAVPERAHIQAYRAVLHRIGRA VQNCSEKTKQLFSSKVIEMGVFKNPSHLN NFIFNQKGAIYRSPFDRSRHAPYQLHADKL LKNDWLELLAEISATLMASESTEQMEDAL

			RLERTRLQLQLSGLPDWEYPASLAKPDIEV EIQTALKMQLAKDTVTSDVLQRAFNLYSS VLSGLTFKLLRRSFSLKMRFSVADTTQLIY VPKVCDWAIPKQYLQAEGEIGIAARVVTE SSPAKMVTEVEMKEPKALGHFMQQAPHD WYFDASLGGTQVAGRIVEKGKEVGKERK LVGYRMRGNSAYKTVLDKSLVGNTELSQ CSMIIEIPYTQTVDADFRAQVQAGLPKVSI NLPVKETITASNKDEQMLFDRFVAIDLGER GLGYAVFDAKTLELQESGHRPIKAITNLLN RTHHYEQRPNQRQKFQAKFNVNLSELREN TVGDVCHQINRICAYYNAFPVLEYMVPDR LDKQLKSVYESVTNRYIWSSTDAHKSARV QFWLGGETWEHPYLKSAKDKKPLVLSPG RGASGKGTSQTCSCCGRNPFDLIKDMKPR AKIAVVDGKAKLENSELKLFERNLESKDD MLARRHRNERAGMEQPLTPGNYTVDEIK ALLRANLRRAPKNRRTKDTTVSEYHCVFS DCGKTMHADENAAVNIGGKFIADIEK
[1288]	OspCas12c 多肽序列	268	MTKLRHRQKKLTHDWAGSKKREVLGSNG KLQNPLLMPVKKGQVTEFRKAFSAYARA TKGEMTDGRKNMFTHSFEPFKTKPSLHQC ELADKAYQSLHSYLPGSLAHFLLSAHALG FRIFSKSGEATAFQASSKIEAYESKLASELA CVDLSIQNLTISTLFNALTTSVRGKGEETSA DPLIARFYTLLTGKPLSRDTQGPERDLAEV ISRKIASSFGTWKEMTANPLQSLQFFEEEL HALDANVSLSPAFDVLIKMNDLQGDLKNR TIVFDPDAPVFEYNAEDPADIIIKLTARYAK EAVIKNQNVGNYVKNAITTTNANGLGWL LNKGLSLLPVSTDDELLEFIGVERSHPSCH ALIELIAQLEAPELFEKNVFSDTRSEVQGMI DSAVSNHIARLSSSRNSLSMDSEELERLIKS FQIHTPHCSLFIGAQSLSQQLESLPEALQSG VNSADILLGSTQYMLTNSLVEESIATYQRT LNRINYLSGVAGQINGAIKRKAIDGEKIHL PAAWSELISLPFIGQPVIDVESDLAHLKNQ YQTLSNEFDTLISALQKNFDLNFNKALLNR TQHFEAMCRSTKKNALSKPEIVSYRDLLA RLTSCLYRGSLVLRRAGIEVLKKHKIFESN SELREHVHERKHFVFVSPLDRKAKKLLRL TDSRPDLLHVIDEILQHDNLENKDRESLWL VRSGYLLAGLPDQLSSSFINLPIITQKGDRR

			LIDLIQYDQINRDAFVMLVTSAFKSNLSGL QYRANKQSFVVTRTLSPYLGSKLVYVPKD KDWLVPSQMFEGRFADILQSDYMVWKDA GRLCVIDTAKHLSNIKKSVFSSEEVLAFLR ELPHRTFIQTEVRGLGVNVDGIAFNNGDIP SLKTFSNCVQVKVSRTNTSLVQTLNRWFE GGKVSPPSIQFERAYYKKDDQIHEDAAKR KIRFQMPATELVHASDDAGWTPSYLLGID PGEYGMGLSLVSINNGEVLDSGFIHINSLIN FASKKSNHQTKVVPRQQYKSPYANYLEQS KDSAAGDIAHILDRLIYKLNALPVFEALSG NSQSAADQVWTKVLSFYTWGDNDAQNSI RKQHWFGASHWDIKGMLRQPPTEKKPKP YIAFPGSQVSSYGNSQRCSCCGRNPIEQLR EMAKDTSIKELKIRNSEIQLFDGTIKLFNPD PSTVIERRRHNLGPSRIPVADRTFKNISPSSL EFKELITIVSRSIRHSPEFIAKKRGIGSEYFC AYSDCNSSLNSEANAAANVAQKFQKQLFF EL
[1289]	Cas12g1 多肽序列	269	MAQASSTPAVSPRPRPRYREERTLVRKLLP RPGQSKQEFRENVKKLRKAFLQFNADVSG VCQWAIQFRPRYGKPAEPTETFWKFFLEPE TSLPPNDSRSPEFRRLQAFEAAAGINGAAA LDDPAFTNELRDSILAVASRPKTKEAQRLF SRLKDYQPAHRMILAKVAAEWIESRYRRA HQNWERNYEEWKKEKQEWEQNHPELTPE IREAFNQIFQQLEVKEKRVRICPAARLLQN KDNCQYAGKNKHSVLCNQFNEFKKNHLQ GKAIKFFYKDAEKYLRCGLQSLKPNVQGP FREDWNKYLRYMNLKEETLRGKNGGRLP HCKNLGQECEFNPHTALCKQYQQQLSSRP DLVQHDELYRKWRREYWREPRKPVFRYP SVKRHSIAKIFGENYFQADFKNSVVGLRLD SMPAGQYLEFAFAPWPRNYRPQPGETEISS VHLHFVGTRPRIGFRFRVPHKRSRFDCTQE ELDELRSRTFPRKAQDQKFLEAARKRLLE TFPGNAEQELRLLAVDLGTDSARAAFFIGK TFQQAFPLKIVKIEKLYEQWPNQKQAGDR RDASSKQPRPGLSRDHVGRHLQKMRAQA SEIAQKRQELTGTPAPETTTDQAAKKATL QPFDLRGLTVHTARMIRDWARLNARQIIQ LAEENQVDLIVLESLRGFRPPGYENLDQEK KRRVAFFAHGRIRRKVTEKAVERGMRVV

[1290]

		TVPYLASSKVCAECRKKQKDNKQWEKNK KRGLFKCEGCGSQAQVDENAARVLGRVF WGEIELPTAIP
Cas12h1 多肽序列	270	MKVHEIPRSQLLKIKQYEGSFVEWYRDLQ EDRKKFASLLFRWAAFGYAAREDDGATYI SPSQALLERRLLLGDAEDVAIKFLDVLFKG GAPSSSCYSLFYEDFALRDKAKYSGAKRE FIEGLATMPLDKIIERIRQDEQLSKIPAEEW LILGAEYSPEEIWEQVAPRIVNVDRSLGKQ LRERLGIKCRRPHDAGYCKILMEVVARQL RSHNETYHEYLNQTHEMKTKVANNLTNE FDLVCEFAEVLEEKNYGLGWYVLWQGVK QALKEQKKPTKIQIAVDQLRQPKFAGLLT AKWRALKGAYDTWKLKKRLEKRKAFPY MPNWDNDYQIPVGLTGLGVFTLEVKRTE VVVDLKEHGKLFCSHSHYFGDLTAEKHPS RYHLKFRHKLKLRKRDSRVEPTIGPWIEA ALREITIQKKPNGVFYLGLPYALSHGIDNF QIAKRFFSAAKPDKEVINGLPSEMVVGAA DLNLSNIVAPVKARIGKGLEGPLHALDYG YGELIDGPKILTPDGPRCGELISLKRDIVEIK SAIKEFKACQREGLTMSEETTTWLSEVESP SDSPRCMIQSRIADTSRRLNSFKYQMNKEG YQDLAEALRLLDAMDSYNSLLESYQRMH LSPGEQSPKEAKFDTKRASFRDLLRRRVA HTIVEYFDDCDIVFFEDLDGPSDSDSRNNA LVKLLSPRTLLLYIRQALEKRGIGMVEVAK DGTSQNNPISGHVGWRNKQNKSEIYFYED KELLVMDADEVGAMNILCRGLNHSVCPY SFVTKAPEKKNDEKKEGDYGKRVKRFLK DRYGSSNVRFLVASMGFVTVTTKRPKDAL VGKRLYYHGGELVTHDLHNRMKDEIKYL VEKEVLARRVSLSDSTIKSYKSFAHV
Cas12i1 多肽序列	271	MSNKEKNASETRKAYTTKMIPRSHDRMK LLGNFMDYLMDGTPIFFELWNQFGGGIDR DIISGTANKDKISDDLLLAVNWFKVMPINS KPQGVSPSNLANLFQQYSGSEPDIQAQEYF ASNFDTEKHQWKDMRVEYERLLAELQLS RSDMHHDLKLMYKEKCIGLSLSTAHYITS VMFGTGAKNNRQTKHQFYSKVIQLLEEST QINSVEQLASIILKAGDCDSYRKLRIRCSRK GATPSILKIVQDYELGTNHDDEVNVPSLIA NLKEKLGRFEYECEWKCMEKIKAFLASKV

[1291]

		GPYYLGSYSAMLENALSPIKGMTTKNCKF VLKQIDAKNDIKYENEPFGKIVEGFFDSPY FESDTNVKWVLHPHHIGESNIKTLWEDLN AIHSKYEEDIASLSEDKKEKRIKVYQGDVC QTINTYCEEVGKEAKTPLVQLLRYLYSRK DDIAVDKIIDGITFLSKKHKVEKQKINPVIQ KYPSFNFGNNSKLLGKIISPKDKLKHNLKC NRNQVDNYIWIEIKVLNTKTMRWEKHHY ALSSTRFLEEVYYPATSENPPDALAARFRT KTNGYEGKPALSAEQIEQIRSAPVGLRKVK KRQMRLEAARQQNLLPRYTWGKDFNINIC KRGNNFEVTLATKVKKKKEKNYKVVLGY DANIVRKNTYAAIEAHANGDGVIDYNDLP VKPIESGFVTVESQVRDKSYDQLSYNGVK LLYCKPHVESRRSFLEKYRNGTMKDNRG NNIQIDFMKDFEAIADDETSLYYFNMKYC KLLQSSIRNHSSQAKEYREEIFELLRDGKLS VLKLSSLSNLSFVMFKVAKSLIGTYFGHLL KKPKNSKSDVKAPPITDEDKQKADPEMFA LRLALEEKRLNKVKSKKEVIANKIVAKAL ELRDKYGPVLIKGENISDTTKKGKKSSTNS FLMDWLARGVANKVKEMVMMHQGLEFV EVNPNFTSHQDPFVHKNPENTFRARYSRC TPSELTEKNRKEILSFLSDKPSKRPTNAYY NEGAMAFLATYGLKKNDVLGVSLEKFKQI MANILHQRSEDQLLFPSRGGMFYLATYKL DADATSVNWNGKQFWVCNADLVAAYNV GLVDIQKDFKKK
Cas12i2 多肽序列	272	MSSAIKSYKSVLRPNERKNQLLKSTIQCLE DGSAFFFKMLQGLFGGITPEIVRFSTEQEK QQQDIALWCAVNWFRPVSQDSLTHTIASD NLVEKFEEYYGGTASDAIKQYFSASIGESY YWNDCRQQYYDLCRELGVEVSDLTHDLE ILCREKCLAVATESNQNNSIISVLFGTGEKE DRSVKLRITKKILEAISNLKEIPKNVAPIQEI ILNVAKATKETFRQVYAGNLGAPSTLEKFI AKDGQKEFDLKKLQTDLKKVIRGKSKERD WCCQEELRSYVEQNTIQYDLWAWGEMFN KAHTALKIKSTRNYNFAKQRLEQFKEIQSL NNLLVVKKLNDFFDSEFFSGEETYTICVHH LGGKDLSKLYKAWEDDPADPENAIVVLC DDLKNNFKKEPIRNILRYIFTIRQECSAQDI LAAAKYNQQLDRYKSQKANPSVLGNQGF

			TWTNAVILPEKAQRNDRPNSLDLRIWLYL KLRHPDGRWKKHHIPFYDTRFFQEIYAAG NSPVDTCQFRTPRFGYHLPKLTDQTAIRVN KKHVKAAKTEARIRLAIQQGTLPVSNLKIT EISATINSKGQVRIPVKFDVGRQKGTLQIG DRFCGYDQNQTASHAYSLWEVVKEGQYH KELGCFVRFISSGDIVSITENRGNQFDQLSY EGLAYPQYADWRKKASKFVSLWQITKKN KKKEIVTVEAKEKFDAICKYQPRLYKFNK EYAYLLRDIVRGKSLVELQQIRQEIFRFIEQ DCGVTRLGSLSLSTLETVKAVKGIIYSYFS TALNASKNNPISDEQRKEFDPELFALLEKL ELIRTRKKKQKVERIANSLIQTCLENNIKFI RGEGDLSTTNNATKKKANSRSMDWLARG VFNKIRQLAPMHNITLFGCGSLYTSHQDPL VHRNPDKAMKCRWAAIPVKDIGDWVLRK LSQNLRAKNIGTGEYYHQGVKEFLSHYEL QDLEEELLKWRSDRKSNIPCWVLQNRLAE KLGNKEAVVYIPVRGGRIYFATHKVATGA VSIVFDQKQVWVCNADHVAAANIALTVK GIGEQSSDEENPDGSRIKLQLTS
[1292]	接头	1308	(GGGS)N
	接头	109	(GGGGS)N
	接头	1309	(EAAAK)N
	接头	56	SGSETPGTSESATPES
		57	(SGGS)N
	接头	273	GGSGGS
	接头	1310	GSSGSETPGTSESATPESSG
	接头	1311	GGAGGCTCTGGAGGAAGC
	接头	1312	GGCTCTTCTGGATCTGAAACACCTGGCA CAAGCGAGAGCGCCACCCCTGAGAGCTC TGGC
	AacCas12b 多肽序列	259	MAVKSMKVKLRLDNMPEIRAGLWKLHTE VNAGVRYYTEWLSLLRQENLYRRSPNGD GEQECYKTAEECKAELLERLRARQVENGH CGPAGSDDELLQLARQLYELLVPQAIGAK GDAQQIARKFLSPLADKDAVGGLGIAKAG NKPRWVRMREAGEPGWEEEKAKAEARKS TDRTADVLRALADFGLKPLMRVYTDSDM SSVQWKPLRKGQAVRTWDRDMFQQAIER MMSWESWNQRVGEAYAKLVEQKSRFEQ KNFVGQEHLVQLVNQLQQDMKEASHGLE

[1293]

		SKEQTAHYLTGRALRGSDKVFEKWEKLD PDAPFDLYDTEIKNVQRRNTRRFGSHDLF AKLAEPKYQALWREDASFLTRYAVYNSIV RKLNHAKMFATFTLPDATAHPIWTRFDKL GGNLHQYTFLFNEFGEGRHAIRFQKLLTV EDGVAKEVDDVTVPISMSAQLDDLLPRDP HELVALYFQDYGAEQHLAGEFGGAKIQY RRDQLNHLHARRGARDVYLNLSVRVQSQ SEARGERRPPYAAVFRLVGDNHRAFVHFD KLSDYLAEHPDDGKLGSEGLLSGLRVMSV DLGLRTSASISVFRVARKDELKPNSEGRVP FCFPIEGNENLVAVHERSQLLKLPGETESK DLRAIREERQRTLRQLRTQLAYLRLLVRC GSEDVGRRERSWAKLIEQPMDANQMTPD WREAFEDELQKLKSLYGICGDREWTEAV YESVRRVWRHMGKQVRDWRKDVRSGER PKIRGYQKDVVGGNSIEQIEYLERQYKFLK SWSFFGKVSGQVIRAEKGSRFAITLREHID HAKEDRLKKLADRIIMEALGYVYALDDER GKGKWVAKYPPCQLILLEELSEYQFNNDR PPSENNQLMQWSHRGVFQELLNQAQVHD LLVGTMYAAFSSRFDARTGAPGIRCRRVP ARCAREQNPEPFPWWLNKFVAEHKLDGC PLRADDLIPTGEGEFFVSPFSAEEGDFHQIH ADLNAAQNLQRRLWSDFDISQIRLRCDWG EVDGEPVLIPRTTGKRTADSYGNKVFYTK TGVTYYERERGKKRRKVFAQEELSEEEAE LLVEADEAREKSVVLMRDPSGIINRGDWT RQKEFWSMVNQRIEGYLVKQIRSRVRLQE SACENTGDI
BhCas12b 多肽序列	260	MAPKKKRKVGIHGVPAAATRSFILKIEPNE EVKKGLWKTHEVLNHGIAYYMNILKLIRQ EAIYEHHEQDPKNPKKVSKAEIQAELWDF VLKMQKCNSFTHEVDKDEVFNILRELYEE LVPSSVEKKGEANQLSNKFLYPLVDPNSQ SGKGTASSGRKPRWYNLKIAGDPSWEEEK KKWEEDKKKDPLAKILGKLAEYGLIPLFIP YTDSNEPIVKEIKWMEKSRNQSVRRLDKD MFIQALERFLSWESWNLKVKEEYEKVEKE YKTLEERIKEDIQALKALEQYEKERQEQLL RDTLNTNEYRLSKRGLRGWREIIQKWLKM DENEPSEKYLEVFKDYQRKHPREAGDYSV YEFLSKKENHFIWRNHPEYPYLYATFCEID

[1294]

		KKKKDAKQQATFTLADPINHPLWVRFEER SGSNLNKYRILTEQLHTEKLKKKLTVQLD RLIYPTESGGWEEKGKVDIVLLPSRQFYNQ IFLDIEEKGKHAFTYKDESIKFPLKGTLGG ARVQFDRDHLRRYPHKVESGNVGRIYFN MTVNIEPTESPVSKSLKIHRDDFPKVVNFK PKELTEWIKDSKGKKLKSGIESLEIGLRVM SIDLGQRQAAAASIFEVVDQKPDIEGKLFF PIKGTELYAVHRASFNIKLPGETLVKSREV LRKAREDNLKLMNQKLNFLRNVLHFQQF EDITEREKRVTKWISRQENSDVPLVYQDEL IQIRELMYKPYKDWVAFLKQLHKRLEVEI GKEVKHWRKSLSDGRKGLYGISLKNIDEI DRTRKFLLRWSLRPTEPGEVRRLEPGQRF AIDQLNHLNALKEDRLKKMANTIIMHALG YCYDVRKKKWQAKNPACQIILFEDLSNYN PYEERSRFENSKLMKWSRREIPRQVALQG EIYGLQVGEVGAQFSSRFHAKTGSPGIRCS VVTKEKLQDNRFFKNLQREGRLTLDKIAV LKEGDLYPDKGGEKFISLSKDRKCVTTHA DINAAQNLQKRFWTRTHGFYKVYCKAYQ VDGQTVYIPESKDQKQKIIEEFGEGYFILK DGVYEWVNAGKLKIKKGSSKQSSSELVDS DILKDSFDLASELKGEKLMLYRDPSGNVFP SDKWMAAGVFFGKLERILISKLTNQYSIST IEDDSSKQSMKRPAATKKAGQAKKKK
BvCas12b(芽孢杆菌属种(Bacillus sp.)V3-13)多核苷酸序列	264	MAIRSIKLKMKTNSGTDSIYLRKALWRTH QLINEGIAYYMNLLTLYRQEAIGDKTKEA YQAELINIIRNQQRNNGSSEEHGSDQEILAL LRQLYELIIPSSIGESGDANQLGNKFLYPLV DPNSQSGKGTSNAGRKPRWKRLKEEGNP DWELEKKKDEERKAKDPTVKIFDNLNKY GLLPLFPLFTNIQKDIEWLPLGKRQSVRKW DKDMFIQAIERLLSWESWNRRVADEYKQL KEKTESYYKEHLTGGEEWIEKIRKFEKERN MELEKNAFAPNDGYFITSRQIRGWDRVYE KWSKLPESASPEELWKVVAEQQNKMSEG FGDPKVFSFLANRENRDIWRGHSERIYHIA AYNGLQKKLSRTKEQATFTLPDAIEHPLWI RYESPGGTNLNLFKLEEKQKKNYYVTLSK IIWPSEEKWIEKENIEIPLAPSIQFNRQIKLK QHVKGKQEISFSDYSSRISLDGVLGGSRIQF NRKYIKNHKELLGEGDIGPVFFNLVVDVA

			PLQETRNGRLQSPIGKALKVISSDFSKVIDY KPKELMDWMNTGSASNSFGVASLLEGMR VMSIDMGQRTSASVSIFEVVKELPKDQEQ KLFYSINDTELFAIHKRSFLLNLPGEVVTK NNKQQRQERRKKRQFVRSQIRMLANVLR LETKKTPDERKKAIHKLMEIVQSYDSWTA SQKEVWEKELNLLTNMAAFNDEIWKESL VELHHRIEPYVGQIVSKWRKGLSEGRKNL AGISMWNIDELEDTRRLLISWSKRSRTPGE ANRIETDEPFGSSLLQHIQNVKDDRLKQM ANLIIMTALGFKYDKEEKDRYKRWKETYP ACQIILFENLNRYLFNLDRSRRENSRLMK WAHRSIPRTVSMQGEMFGLQVGDVRSEY SSRFHAKTGAPGIRCHALTEEDLKAGSNTL KRLIEDGFINESELAYLKKGDIIPSQGGELF VTLSKRYKKDSDNNELTVIHADINAAQNL QKRFWQQNSEVYRVPCQLARMGEDKLYI PKSQTETIKKYFGKGSFVKNNTEQEVYKW EKSEKMKIKTDTTFDLQDLDGFEDISKTIE LAQEQQKKYLTMFRDPSGYFFNNETWRP QKEYWSIVNNIIKSCLKKKILSNKVEL
[1295]	BTCas12b.BTCas12b 多肽序列	265	MATRSFILKIEPNEEVKKGLWKTHEVLNH GIAYYMNILKLIRQEAIYEHHEQDPKNPKK VSKAEIQAELWDFVLKMQKCNSFTHEVD KDVVFNILRELYEELVPSSVEKKGEANQLS NKFLYPLVDPNSQSGKGTASSGRKPRWYN LKIAGDPSWEEEKKKWEEDKKKDPLAKIL GKLAEYGLIPLFIPFTDSNEPIVKEIKWMEK SRNQSVRRLDKDMFIQALERFLSWESWNL KVKEEYEKVEKEHKTLEERIKEDIQAFKSL EQYEKERQEQLLRDTLNTNEYRLSKRGLR GWREIIQKWLKMDENEPSEKYLEVFKDYQ RKHPREAGDYSVYEFLSKKENHFIWRNHP EYPYLYATFCEIDKKKKDAKQQATFTLAD PINHPLWVRFEERSGSNLNKYRILTEQLHT EKLKKKLTVQLDRLIYPTESGGWEEKGKV DIVLLPSRQFYNQIFLDIEEKGKHAFTYKD ESIKFPLKGTLGGARVQFDRDHLRRYPHK VESGNVGRIYFNMTVNIEPTESPVSKSLKI HRDDFPKFVNFKPKELTEWIKDSKGKKLK SGIESLEIGLRVMSIDLGQRQAAAASIFEVV DQKPDIEGKLFFPIKGTELYAVHRASFNIK LPGETLVKSREVLRKAREDNLKLMNQKL

[1296]

		NFLRNVLHFQQFEDITEREKRVTKWISRQE NSDVPLVYQDELIQIRELMYKPYKDWVAF LKQLHKRLEVEIGKEVKHWRKSLSDGRK GLYGISLKNIDEIDRTRKFLLRWSLRPTEPG EVRRLEPGQRFAIDQLNHLNALKEDRLKK MANTIIMHALGYCYDVRKKKWQAKNPAC QIILFEDLSNYNPYEERSRFENSKLMKWSR REIPRQVALQGEIYGLQVGEVGAQFSSRFH AKTGSPGIRCSVVTKEKLQDNRFFKNLQR EGRLTLDKIAVLKEGDLYPDKGGEKFISLS KDRKLVTTHADINAAQNLQKRFWTRTHG FYKVYCKAYQVDGQTVYIPESKDQKQKII EEFGEGYFILKDGVYEWGNAGKLKIKKGS SKQSSSELVDSDILKDSFDLASELKGEKLM LYRDPSGNVFPSDKWMAAGVFFGKLERIL ISKLTNQYSISTIEDDSSKQSM
5'UTR	261	GGGAAATAAGAGAGAAAAGAAGAGTAA GAAGAAATATAAGAGCCACC
3'UTR (TriLink 标准 UTR)	262	GCTGGAGCCTCGGTGGCCATGCTTCTTG CCCCTTGGGCCTCCCCCCAGCCCCTCCTC CCCTTCCTGCACCCGTACCCCCGTGGTCT TTGAATAAAGTCTGA
bhCas12b (V4)多核苷酸序列	263	ATGGCCCCAAAGAAGAAGCGGAAGGTC GGTATCCACGGAGTCCCAGCAGCCGCCA CCAGATCCTTCATCCTGAAGATCGAGCC CAACGAGGAAGTGAAGAAAGGCCTCTG GAAAACCCACGAGGTGCTGAACCACGG AATCGCCTACTACATGAATATCCTGAAG CTGATCCGGCAAGAGGCCATCTACGAGC ACCACGAGCAGGACCCCAAGAATCCCA AGAAGGTGTCCAAGGCCGAGATCCAGG CCGAGCTGTGGGATTTCGTGCTGAAGAT GCAGAAGTGCAACAGCTTCACACACGAG GTGGACAAGGACGAGGTGTTCAACATCC TGAGAGAGCTGTACGAGGAACTGGTGCC CAGCAGCGTGGAAAAGAAGGGCGAAGC CAACCAGCTGAGCAACAAGTTTCTGTAC CCTCTGGTGGACCCCAACAGCCAGTCTG GAAAGGGAACAGCCAGCAGCGGCAGAA AGCCCAGATGGTACAACCTGAAGATTGC CGGCGATCCCTCCTGGGAAGAAGAGAA GAAGAAGTGGGAAGAAGATAAGAAAAA GGACCCGCTGGCCAAGATCCTGGGCAAG

[1297]

		CTGGCTGAGTACGGACTGATCCCTCTGT TCATCCCCTACACCGACAGCAACGAGCC CATCGTGAAAGAAATCAAGTGGATGGA AAAGTCCCGGAACCAGAGCGTGCGGCG GCTGGATAAGGACATGTTCATTCAGGCC CTGGAACGGTTCCTGAGCTGGGAGAGCT GGAACCTGAAAGTGAAAGAGGAATACG AGAAGGTCGAGAAAGAGTACAAGACCC TGGAAGAGAGGATCAAAGAGGACATCC AGGCTCTGAAGGCTCTGGAACAGTATGA GAAAGAGCGGCAAGAACAGCTGCTGCG GGACACCCTGAACACCAACGAGTACCGG CTGAGCAAGAGAGGCCTTAGAGGCTGGC GGGAAATCATCCAGAAATGGCTGAAAAT GGACGAGAACGAGCCCTCCGAGAAGTA CCTGGAAGTGTTCAAGGACTACCAGCGG AAGCACCCTAGAGAGGCCGGCGATTACA GCGTGTACGAGTTCCTGTCCAAGAAAGA GAACCACTTCATCTGGCGGAATCACCCT GAGTACCCCTACCTGTACGCCACCTTCT GCGAGATCGACAAGAAAAAGAAGGACG CCAAGCAGCAGGCCACCTTCACACTGGC CGATCCTATCAATCACCCTCTGTGGGTCC GATTCGAGGAAAGAAGCGGCAGCAACC TGAACAAGTACAGAATCCTGACCGAGCA GCTGCACACCGAGAAGCTGAAGAAAAA GCTGACAGTGCAGCTGGACCGGCTGATC TACCCTACAGAATCTGGCGGCTGGGAAG AGAAGGGCAAAGTGGACATTGTGCTGCT GCCCAGCCGGCAGTTCTACAACCAGATC TTCCTGGACATCGAGGAAAAGGGCAAGC ACGCCTTCACCTACAAGGATGAGAGCAT CAAGTTCCCTCTGAAGGGCACACTCGGC GGAGCCAGAGTGCAGTTCGACAGAGATC ACCTGAGAAGATACCCTCACAAGGTGGA AAGCGGCAACGTGGGCAGAATCTACTTC AACATGACCGTGAACATCGAGCCTACAG AGTCCCCAGTGTCCAAGTCTCTGAAGAT CCACCGGGACGACTTCCCCAAGGTGGTC AACTTCAAGCCCAAAGAACTGACCGAGT GGATCAAGGACAGCAAGGGCAAGAAAC TGAAGTCCGGCATCGAGTCCCTGGAAAT CGGCCTGAGAGTGATGAGCATCGACCTG

[1298]

		GGACAGAGACAGGCCGCTGCCGCCTCTA TTTTCGAGGTGGTGGATCAGAAGCCCGA CATCGAAGGCAAGCTGTTTTTCCCAATC AAGGGCACCGAGCTGTATGCCGTGCACA GAGCCAGCTTCAACATCAAGCTGCCCGG CGAGACACTGGTCAAGAGCAGAGAAGT GCTGCGGAAGGCCAGAGAGGACAATCT GAAACTGATGAACCAGAAGCTCAACTTC CTGCGGAACGTGCTGCACTTCCAGCAGT TCGAGGACATCACCGAGAGAGAGAAGC GGGTCACCAAGTGGATCAGCAGACAAG AGAACAGCGACGTGCCCCTGGTGTACCA GGATGAGCTGATCCAGATCCGCGAGCTG ATGTACAAGCCTTACAAGGACTGGGTCG CCTTCCTGAAGCAGCTCCACAAGAGACT GGAAGTCGAGATCGGCAAAGAAGTGAA GCACTGGCGGAAGTCCCTGAGCGACGGA AGAAAGGGCCTGTACGGCATCTCCCTGA AGAACATCGACGAGATCGATCGGACCCG GAAGTTCCTGCTGAGATGGTCCCTGAGG CCTACCGAACCTGGCGAAGTGCGTAGAC TGGAACCCGGCCAGAGATTCGCCATCGA CCAGCTGAATCACCTGAACGCCCTGAAA GAAGATCGGCTGAAGAAGATGGCCAAC ACCATCATCATGCACGCCCTGGGCTACT GCTACGACGTGCGGAAGAAGAAATGGC AGGCTAAGAACCCCGCCTGCCAGATCAT CCTGTTCGAGGATCTGAGCAACTACAAC CCCTACGAGGAAAGGTCCCGCTTCGAGA ACAGCAAGCTCATGAAGTGGTCCAGACG CGAGATCCCCAGACAGGTTGCACTGCAG GGCGAGATCTATGGCCTGCAAGTGGGAG AAGTGGGCGCTCAGTTCAGCAGCAGATT CCACGCCAAGACAGGCAGCCCTGGCATC AGATGTAGCGTCGTGACCAAAGAGAAG CTGCAGGACAATCGGTTCTTCAAGAATC TGCAGAGAGAGGGCAGACTGACCCTGG ACAAAATCGCCGTGCTGAAAGAGGGCG ATCTGTACCCAGACAAAGGCGGCGAGA AGTTCATCAGCCTGAGCAAGGATCGGAA GTGCGTGACCACACACGCCGACATCAAC GCCGCTCAGAACCTGCAGAAGCGGTTCT GGACAAGAACCCACGGCTTCTACAAGGT

		GTACTGCAAGGCCTACCAGGTGGACGGC CAGACCGTGTACATCCCTGAGAGCAAGG ACCAGAAGCAGAAGATCATCGAAGAGT TCGGCGAGGGCTACTTCATTCTGAAGGA CGGGGTGTACGAATGGGTCAACGCCGGC AAGCTGAAAATCAAGAAGGGCAGCTCC AAGCAGAGCAGCAGCGAGCTGGTGGAT AGCGACATCCTGAAAGACAGCTTCGACC TGGCCTCCGAGCTGAAAGGCGAAAAGCT GATGCTGTACAGGGACCCCAGCGGCAAT GTGTTCCCCAGCGACAAATGGATGGCCG CTGGCGTGTTCTTCGGAAAGCTGGAACG CATCCTGATCAGCAAGCTGACCAACCAG TACTCCATCAGCACCATCGAGGACGACA GCAGCAAGCAGTCTATGAAAAGGCCGG CGGCCACGAAAAAGGCCGGCCAGGCAA AAAAGAAAAAG
NLS	1313	MAPKKKRKVGIHGVPAA
NLS	1314	ATGGCCCCAAAGAAGAAGCGGAAGGTC GGTATCCACGGAGTCCCAGCAGCC
101 Cas9 TadAins 1015 多肽序列	1315	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK

[1299]

[1300]

		SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVGSSGSETPGTSESATPESSG SEVEFSHEYWMRHALTLAKRARDEREVP VGAVLVLNNRVIGEGWNRAIGLHDPTAH AEIMALRQGGLVMQNYRLIDATLYVTFEP CVMCAGAMIHSRIGRVVFGVRNAKTGAA GSLMDVLHYPGMNHRVEITEGILADECAA LLCYFFRMPRQVFNAQKKAQSSTDYDVR KMIAKSEQEIGKATAKYFFYSNIMNFFKTE ITLANGEIRKRPLIETNGETGEIVWDKGRD FATVRKVLSMPQVNIVKKTEVQTGGFSKE SILPKRNSDKLIARKKDWDPKKYGGFDSP TVAYSVLVVAKVEKGKSKKLKSVKELLGI TIMERSSFEKNPIDFLEAKGYKEVKKDLIIK LPKYSLFELENGRKRMLASAGELQKGNEL ALPSKYVNFLYLASHYEKLKGSPEDNEQK QLFVEQHKHYLDEIIEQISEFSKRVILADAN LDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTIDRKRYTSTKEVLDA TLIHQSITGLYETRIDLSQLGGD
102 Cas9 TadAins 1022 多肽序列	1316	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE

[1301]

		ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIGSSGSETPGTSE SATPESSGSEVEFSHEYWMRHALTLAKRA RDEREVPVGAVLVLNNRVIGEGWNRAIGL HDPTAHAEIMALRQGGLVMQNYRLIDATL YVTFEPCVMCAGAMIHSRIGRVVFGVRNA KTGAAGSLMDVLHYPGMNHRVEITEGILA DECAALLCYFFRMPRQVFNAQKKAQSSTD AKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFAT VRKVLSMPQVNIVKKTEVQTGGFSKESILP KRNSDKLIARKKDWDPKKYGGFDSPTVA YSVLVVAKVEKGKSKKLKSVKELLGITIM ERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK YSLFELENGRKRMLASAGELQKGNELALP SKYVNFLYLASHYEKLKGSPEDNEQKQLF VEQHKHYLDEIIEQISEFSKRVILADANLD KVLSAYNKHRDKPIREQAENIIHLFTLTNL GAPAAFKYFDTTIDRKRYTSTKEVLDATLI HQSITGLYETRIDLSQLGGD
103 Cas9 TadAins 1029 多肽	1317	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE

[1302]

序列		MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEIGSSGS ETPGTSESATPESSGSEVEFSHEYWMRHAL TLAKRARDEREVPVGAVLVLNNRVIGEG WNRAIGLHDPTAHAEIMALRQGGLVMQN YRLIDATLYVTFEPCVMCAGAMIHSRIGR VVFGVRNAKTGAAGSLMDVLHYPGMNH RVEITEGILADECAALLCYFFRMPRQVFNA QKKAQSSTDGKATAKYFFYSNIMNFFKTEI TLANGEIRKRPLIETNGETGEIVWDKGRDF ATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFDSPT VAYSVLVVAKVEKGKSKKLKSVKELLGIT

[1303]

		IMERSSFEKNPIDFLEAKGYKEVKKDLIIKL PKYSLFELENGRKRMLASAGELQKGNELA LPSKYVNFLYLASHYEKLKGSPEDNEQKQ LFVEQHKHYLDEIIEQISEFSKRVILADANL DKVLSAYNKHRDKPIREQAENIIHLFTLTN LGAPAAFKYFDTTIDRKRYTSTKEVLDAT LIHQSITGLYETRIDLSQLGGD
103 Cas9 TadAins 1040 多肽序列	1318	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSGSSGSETPGTSESATPESSGSEVEF

			SHEYWMRHALTLAKRARDEREVPVGAVL VLNNRVIGEGWNRAIGLHDPTAHAEIMAL RQGGLVMQNYRLIDATLYVTFEPCVMCA GAMIHSRIGRVVFGVRNAKTGAAGSLMD VLHYPGMNHRVEITEGILADECAALLCYF FRMPRQVFNAQKKAQSSTDNIMNFFKTEI TLANGEIRKRPLIETNGETGEIVWDKGRDF ATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFDSPT VAYSVLVVAKVEKGKSKKLKSVKELLGIT IMERSSFEKNPIDFLEAKGYKEVKKDLIIKL PKYSLFELENGRKRMLASAGELQKGNELA LPSKYVNFLYLASHYEKLKGSPEDNEQKQ LFVEQHKHYLDEIIEQISEFSKRVILADANL DKVLSAYNKHRDKPIREQAENIIHLFTLTN LGAPAAFKYFDTTIDRKRYTSTKEVLDAT LIHQSITGLYETRIDLSQLGGD
[1304]	105 Cas9 TadAins 1068 多肽序列	1319	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK

[1305]

		GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGEGSSGSETPGTSESATPESSGSEVEFSHE YWMRHALTLAKRARDEREVPVGAVLVLN NRVIGEGWNRAIGLHDPTAHAEIMALRQG GLVMQNYRLIDATLYVTFEPCVMCAGAM IHSRIGRVVFGVRNAKTGAAGSLMDVLHY PGMNHRVEITEGILADECAALLCYFFRMP RQVFNAQKKAQSSTDTGEIVWDKGRDFA TVRKVLSMPQVNIVKKTEVQTGGFSKESIL PKRNSDKLIARKKDWDPKKYGGFDSPTVA YSVLVVAKVEKGKSKKLKSVKELLGITIM ERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK YSLFELENGRKRMLASAGELQKGNELALP SKYVNFLYLASHYEKLKGSPEDNEQKQLF VEQHKHYLDEIIEQISEFSKRVILADANLD KVLSAYNKHRDKPIREQAENIIHLFTLTNL GAPAAFKYFDTTIDRKRYTSTKEVLDATLI HQSITGLYETRIDLSQLGGD
106 Cas9 TadAins 1247 多肽序列	1320	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN

[1306]

		FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFDSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASAGELQKGNELALPSKYVNFLYLASH YEKLKGGSSGSETPGTSESATPESSGSEVEF SHEYWMRHALTLAKRARDEREVPVGAVL VLNNRVIGEGWNRAIGLHDPTAHAEIMAL RQGGLVMQNYRLIDATLYVTFEPCVMCA GAMIHSRIGRVVFGVRNAKTGAAGSLMD VLHYPGMNHRVEITEGILADECAALLCYF FRMPRQVFNAQKKAQSSTDSPEDNEQKQL FVEQHKHYLDEIIEQISEFSKRVILADANLD KVLSAYNKHRDKPIREQAENIIHLFTLTNL GAPAAFKYFDTTIDRKRYTSTKEVLDATLI HQSITGLYETRIDLSQLGGD
107 Cas9 TadAins 1054 多肽序列	1321	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN

[1307]

		SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGSSGSETPGTS ESATPESSGSEVEFSHEYWMRHALTLAKR ARDEREVPVGAVLVLNNRVIGEGWNRAIG LHDPTAHAEIMALRQGGLVMQNYRLIDAT LYVTFEPCVMCAGAMIHSRIGRVVFGVRN AKTGAAGSLMDVLHYPGMNHRVEITEGIL ADECAALLCYFFRMPRQVFNAQKKAQSST DGEIRKRPLIETNGETGEIVWDKGRDFATV RKVLSMPQVNIVKKTEVQTGGFSKESILPK RNSDKLIARKKDWDPKKYGGFDSPTVAYS VLVVAKVEKGKSKKLKSVKELLGITIMER SSFEKNPIDFLEAKGYKEVKKDLIIKLPKYS LFELENGRKRMLASAGELQKGNELALPSK YVNFLYLASHYEKLKGSPEDNEQKQLFVE

		QHKHYLDEIIEQISEFSKRVILADANLDKV LSAYNKHRDKPIREQAENIIHLFTLTNLGA PAAFKYFDTTIDRKRYTSTKEVLDATLIHQ SITGLYETRIDLSQLGGD
108 Cas9 TadAins 1026 多肽序列	1322	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEGSSGSETP GTSESATPESSGSEVEFSHEYWMRHALTL AKRARDEREVPVGAVLVLNNRVIGEGWN RAIGLHDPTAHAEIMALRQGGLVMQNYR LIDATLYVTFEPCVMCAGAMIHSRIGRVVF

[1308]

[1309]

		GVRNAKTGAAGSLMDVLHYPGMNHRVEI TEGILADECAALLCYFFRMPRQVFNAQKK AQSSTDQEIGKATAKYFFYSNIMNFFKTEI TLANGEIRKRPLIETNGETGEIVWDKGRDF ATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFDSPT VAYSVLVVAKVEKGKSKKLKSVKELLGIT IMERSSFEKNPIDFLEAKGYKEVKKDLIIKL PKYSLFELENGRKRMLASAGELQKGNELA LPSKYVNFLYLASHYEKLKGSPEDNEQKQ LFVEQHKHYLDEIIEQISEFSKRVILADANL DKVLSAYNKHRDKPIREQAENIIHLFTLTN LGAPAAFKYFDTTIDRKRYTSTKEVLDAT LIHQSITGLYETRIDLSQLGGD
	109 Cas9 TadAins 768 多肽序列	1323	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQGSSG SETPGTSESATPESSGSEVEFSHEYWMRHA LTLAKRARDEREVPVGAVLVLNNRVIGEG WNRAIGLHDPTAHAEIMALRQGGLVMQN

[1310]

		YRLIDATLYVTFEPCVMCAGAMIHSRIGR VVFGVRNAKTGAAGSLMDVLHYPGMNH RVEITEGILADECAALLCYFFRMPRTTQKG QKNSRERMKRIEEGIKELGSQILKEHPVEN TQLQNEKLYLYYLQNGRDMYVDQELDIN RLSDYDVDHIVPQSFLKDDSIDNKVLTRSD KNRGKSDNVPSEEVVKKMKNYWRQLLN AKLITQRKFDNLTKAERGGLSELDKAGFIK RQLVETRQITKHVAQILDSRMNTKYDEND KLIREVKVITLKSKLVSDFRKDFQFYKVRE INNYHHAHDAYLNAVVGTALIKKYPKLES EFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFDSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIE QISEFSKRVILADANLDKVLSAYNKHRDK PIREQAENIIHLFTLTNLGAPAAFKYFDTTI DRKRYTSTKEVLDATLIHQSITGLYETRID LSQLGGD
110.1 Cas9 TadAins 1250 多肽序列	1324	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL

[1311]

		KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFDSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPGSSGSETPGTSESATPESSGSEV EFSHEYWMRHALTLAKRARDEREVPVGA VLVLNNRVIGEGWNRAIGLHDPTAHAEIM ALRQGGLVMQNYRLIDATLYVTFEPCVM CAGAMIHSRIGRVVFGVRNAKTGAAGSL MDVLHYPGMNHRVEITEGILADECAALLC YFFRMPREDNEQKQLFVEQHKHYLDEIIE QISEFSKRVILADANLDKVLSAYNKHRDK PIREQAENIIHLFTLTNLGAPAAFKYFDTTI DRKRYTSTKEVLDATLIHQSITGLYETRID LSQLGGD
110.2 Cas9 TadAins 1250 多肽序列	1325	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK

[1312]

		DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFDSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPGSSGSSGSETPGTSESATPESSG SEVEFSHEYWMRHALTLAKRARDEREVP VGAVLVLNNRVIGEGWNRAIGLHDPTAH AEIMALRQGGLVMQNYRLIDATLYVTFEP CVMCAGAMIHSRIGRVVFGVRNAKTGAA GSLMDVLHYPGMNHRVEITEGILADECAA LLCYFFRMPREDNEQKQLFVEQHKHYLDE IIEQISEFSKRVILADANLDKVLSAYNKHR DKPIREQAENIIHLFTLTNLGAPAAFKYFDT TIDRKRYTSTKEVLDATLIHQSITGLYETRI

[1313]

		DLSQLGGD
110.3 Cas9 TadAins 1250 多肽序列	1326	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFDSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASAGELQKGNELALPSKYVNFLYLASH

[1314]

		YEKLKGSPGSSGSSGSETPGTSESATPESGS SSGSEVEFSHEYWMRHALTLAKRARDERE VPVGAVLVLNNRVIGEGWNRAIGLHDPTA HAEIMALRQGGLVMQNYRLIDATLYVTFE PCVMCAGAMIHSRIGRVVFGVRNAKTGA AGSLMDVLHYPGMNHRVEITEGILADECA ALLCYFFRMPREDNEQKQLFVEQHKHYL DEIIEQISEFSKRVILADANLDKVLSAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYF DTTIDRKRYTSTKEVLDATLIHQSITGLYET RIDLSQLGGD
110.4 Cas9 TadAins 1250 多肽序列	1327	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN

			DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFDSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPGSSGSSGSETPGTSESATPESGS SSGSEVEFSHEYWMRHALTLAKRARDERE VPVGAVLVLNNRVIGEGWNRAIGLHDPTA HAEIMALRQGGLVMQNYRLIDATLYVTFE PCVMCAGAMIHSRIGRVVFGVRNAKTGA AGSLMDVLHYPGMNHRVEITEGILADECA ALLCYFFRMRREDNEQKQLFVEQHKHYL DEIIEQISEFSKRVILADANLDKVLSAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYF DTTIDRKRYTSTKEVLDATLIHQSITGLYET RIDLSQLGGD
[1315]	110.5 Cas9 TadAins 1249 多肽序列	1328	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR

[1316]

		RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFDSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSGSSGSSGSETPGTSESATPESGSS SGSEVEFSHEYWMRHALTLAKRARDERE VPVGAVLVLNNRVIGEGWNRAIGLHDPTA HAEIMALRQGGLVMQNYRLIDATLYVTFE PCVMCAGAMIHSRIGRVVFGVRNAKTGA AGSLMDVLHYPGMNHRVEITEGILADECA ALLCYFFRMRRPEDNEQKQLFVEQHKHYL DEIIEQISEFSKRVILADANLDKVLSAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYF DTTIDRKRYTSTKEVLDATLIHQSITGLYET RIDLSQLGGD
110.5 Cas9 TadAins Δ 59-66 1250 多肽序列	1329	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN

[1317]

		GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFDSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPGSSGSSGSETPGTSESATPESGS SGSEVEFSHEYWMRHALTLAKRARDERE VPVGAVLVLNNRVIGEGWNRAHAEIMAL RQGGLVMQNYRLIDATLYVTFEPCVMCA GAMIHSRIGRVVFGVRNAKTGAAGSLMD VLHYPGMNHRVEITEGILADECAALLCYF FRMPRQVFNAQKKAQSSTDEDNEQKQLF VEQHKHYLDEIIEQISEFSKRVILADANLD KVLSAYNKHRDKPIREQAENIIHLFTLTNL GAPAAFKYFDTTIDRKRYTSTKEVLDATLI HQSITGLYETRIDLSQLGGD
110.6 Cas9	1330	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA

[1318]

TadAins 1251 多肽序列		EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFDSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPEGSSGSSGSETPGTSESATPESG SSSGSEVEFSHEYWMRHALTLAKRARDER EVPVGAVLVLNNRVIGEGWNRAIGLHDPT

[1319]

		AHAEIMALRQGGLVMQNYRLIDATLYVTF EPCVMCAGAMIHSRIGRVVFGVRNAKTG AAGSLMDVLHYPGMNHRVEITEGILADEC AALLCYFFRMRRDNEQKQLFVEQHKHYL DEIIEQISEFSKRVILADANLDKVLSAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYF DTTIDRKRYTSTKEVLDATLIHQSITGLYET RIDLSQLGGD
110.7 Cas9 TadAins 1252 多肽序列	1331	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEIGKATA

			KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFDSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPEDGSSGSSGSETPGTSESATPES GSSSGSEVEFSHEYWMRHALTLAKRARDE REVPVGAVLVLNNRVIGEGWNRAIGLHDP TAHAEIMALRQGGLVMQNYRLIDATLYV TFEPCVMCAGAMIHSRIGRVVFGVRNAKT GAAGSLMDVLHYPGMNHRVEITEGILADE CAALLCYFFRMRRNEQKQLFVEQHKHYL DEIIEQISEFSKRVILADANLDKVLSAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYF DTTIDRKRYTSTKEVLDATLIHQSITGLYET RIDLSQLGGD
[1320]	110.8 Cas9 TadAins Δ 59-66 C 截短的 1250 多肽序列	1332	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV

[1321]

		VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFDSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPGSSGSETPGTSESATPESSGSEV EFSHEYWMRHALTLAKRARDEREVPVGA VLVLNNRVIGEGWNRAHAEIMALRQGGL VMQNYRLIDATLYVTFEPCVMCAGAMIHS RIGRVVFGVRNAKTGAAGSLMDVLHYPG MNHRVEITEGILADECAALLCYFFRMPRQ EDNEQKQLFVEQHKHYLDEIIEQISEFSKR VILADANLDKVLSAYNKHRDKPIREQAEN IIHLFTLTNLGAPAAFKYFDTTIDRKRYTST KEVLDATLIHQSITGLYETRIDLSQLGGD
	111.1 Cas9 TadAins 997 多肽序列	1333	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN

[1322]

		FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALSHEYWMR HALTLAKRARDEREVPVGAVLVLNNRVIG EGWNRAIGLHDPTAHAEIMALRQGGLVM QNYRLIDATLYVTFEPCVMCAGAMIHSRI GRVVFGVRNAKTGAAGSLMDVLHYPGM NHRVEITEGILADECAALLCYFFRMPRQVF NAQKKAQSSTDGSSGSETPGTSESATPESS GIKKYPKLESEFVYGDYKVYDVRKMIAKS EQEIGKATAKYFFYSNIMNFFKTEITLANG EIRKRPLIETNGETGEIVWDKGRDFATVRK VLSMPQVNIVKKTEVQTGGFSKESILPKRN SDKLIARKKDWDPKKYGGFDSPTVAYSVL VVAKVEKGKSKKLKSVKELLGITIMERSSF EKNPIDFLEAKGYKEVKKDLIIKLPKYSLF ELENGRKRMLASAGELQKGNELALPSKY VNFLYLASHYEKLKGSPEDNEQKQLFVEQ HKHYLDEIIEQISEFSKRVILADANLDKVLS AYNKHRDKPIREQAENIIHLFTLTNLGAPA AFKYFDTTIDRKRYTSTKEVLDATLIHQSIT GLYETRIDLSQLGGD
111.2 Cas9 TadAins 997 多肽序列	1334	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN

[1323]

		SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALSHEYWMR HALTLAKRARDEREVPVGAVLVLNNRVIG EGWNRAIGLHDPTAHAEIMALRQGGLVM QNYRLIDATLYVTFEPCVMCAGAMIHSRI GRVVFGVRNAKTGAAGSLMDVLHYPGM NHRVEITEGILADECAALLCYFFRMPRQVF NAQKKAQSSTDGSSGSSGSETPGTSESATP ESSGGSSIKKYPKLESEFVYGDYKVYDVR KMIAKSEQEIGKATAKYFFYSNIMNFFKTE ITLANGEIRKRPLIETNGETGEIVWDKGRD FATVRKVLSMPQVNIVKKTEVQTGGFSKE SILPKRNSDKLIARKKDWDPKKYGGFDSP TVAYSVLVVAKVEKGKSKKLKSVKELLGI TIMERSSFEKNPIDFLEAKGYKEVKKDLIIK LPKYSLFELENGRKRMLASAGELQKGNEL ALPSKYVNFLYLASHYEKLKGSPEDNEQK

[1324]

		QLFVEQHKHYLDEIIEQISEFSKRVILADAN LDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTIDRKRYTSTKEVLDA TLIHQSITGLYETRIDLSQLGGD
112 Δ HNH TadA 多肽序列	1335	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSEVEFSHEY WMRHALTLAKRARDEREVPVGAVLVLNN RVIGEGWNRAIGLHDPTAHAEIMALRQGG LVMQNYRLIDATLYVTFEPCVMCAGAMI HSRIGRVVFGVRNAKTGAAGSLMDVLHY PGMNHRVEITEGILADECAALLCYFFRMP RQVFNAQKKAQSSTDGGLSELDKAGFIKR QLVETRQITKHVAQILDSRMNTKYDENDK LIREVKVITLKSKLVSDFRKDFQFYKVREI NNYHHAHDAYLNAVVGTALIKKYPKLES EFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI

[1325]

		VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFDSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIE QISEFSKRVILADANLDKVLSAYNKHRDK PIREQAENIIHLFTLTNLGAPAAFKYFDTTI DRKRYTSTKEVLDATLIHQSITGLYETRID LSQLGGD
113 N 端单 TadA 螺旋截短的 165 端多肽序列	1336	MSEVEFSHEYWMRHALTLAKRARDEREV PVGAVLVLNNRVIGEGWNRAIGLHDPTAH AEIMALRQGGLVMQNYRLIDATLYVTFEP CVMCAGAMIHSRIGRVVFGVRNAKTGAA GSLMDVLHYPGMNHRVEITEGILADECAA LLCYFFRMPRSGGSSGGSSGSETPGTSESA TPESSGGSSGGSDKKYSIGLAIGTNSVGWA VITDEYKVPSKKFKVLGNTDRHSIKKNLIG ALLFDSGETAEATRLKRTARRRYTRRKNR ICYLQEIFSNEMAKVDDSFFHRLEESFLVE EDKKHERHPIFGNIVDEVAYHEKYPTIYHL RKKLVDSTDKADLRLIYLALAHMIKFRGH FLIEGDLNPDNSDVDKLFIQLVQTYNQLFE ENPINASGVDAKAILSARLSKSRRLENLIA QLPGEKKNGLFGNLIALSLGLTPNFKSNFD LAEDAKLQLSKDTYDDDLDNLLAQIGDQ YADLFLAAKNLSDAILLSDILRVNTEITKA PLSASMIKRYDEHHQDLTLLKALVRQQLP EKYKEIFFDQSKNGYAGYIDGGASQEEFY KFIKPILEKMDGTEELLVKLNREDLLRKQR TFDNGSIPHQIHLGELHAILRRQEDFYPFLK DNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETITPWNFEEVVDKGASAQSFIER MTNFDKNLPNEKVLPKHSLLYEYFTVYNE LTKVKYVTEGMRKPAFLSGEQKKAIVDLL FKTNRKVTVKQLKEDYFKKIECFDSVEISG VEDRFNASLGTYHDLLKIIKDKDFLDNEEN EDILEDIVLTLTLFEDREMIEERLKTYAHLF DDKVMKQLKRRRYTGWGRLSRKLINGIR DKQSGKTILDFLKSDGFANRNFMQLIHDD SLTFKEDIQKAQVSGQGDSLHEHIANLAGS PAIKKGILQTVKVVDELVKVMGRHKPENI VIEMARENQTTQKGQKNSRERMKRIEEGI

		KELGSQILKEHPVENTQLQNEKLYLYYLQ NGRDMYVDQELDINRLSDYDVDHIVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEV VKKMKNYWRQLLNAKLITQRKFDNLTKA ERGGLSELDKAGFIKRQLVETRQITKHVA QILDSRMNTKYDENDKLIREVKVITLKSKL VSDFRKDFQFYKVREINNYHHAHDAYLN AVVGTALIKKYPKLESEFVYGDYKVYDVR KMIAKSEQEIGKATAKYFFYSNIMNFFKTE ITLANGEIRKRPLIETNGETGEIVWDKGRD FATVRKVLSMPQVNIVKKTEVQTGGFSKE SILPKRNSDKLIARKKDWDPKKYGGFDSP TVAYSVLVVAKVEKGKSKKLKSVKELLGI TIMERSSFEKNPIDFLEAKGYKEVKKDLIIK LPKYSLFELENGRKRMLASAGELQKGNEL ALPSKYVNFLYLASHYEKLKGSPEDNEQK QLFVEQHKHYLDEIIEQISEFSKRVILADAN LDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTIDRKRYTSTKEVLDA TLIHQSITGLYETRIDLSQLGGD
[1326]	114 N 端单 TadA 螺旋截短 165 端 Δ 59-65 多肽序列	1337	MSEVEFSHEYWMRHALTLAKRARDEREV PVGAVLVLNNRVIGEGWNRTAHAEIMAL RQGGLVMQNYRLIDATLYVTFEPCVMCA GAMIHSRIGRVVFGVRNAKTGAAGSLMD VLHYPGMNHRVEITEGILADECAALLCYF FRMPRSGGSSGGSSGSETPGTSESATPESSG GSSGGSDKKYSIGLAIGTNSVGWAVITDEY KVPSKKFKVLGNTDRHSIKKNLIGALLFDS GETAEATRLKRTARRRYTRRKNRICYLQEI FSNEMAKVDDSFFHRLEESFLVEEDKKHE RHPIFGNIVDEVAYHEKYPTIYHLRKKLVD STDKADLRLIYLALAHMIKFRGHFLIEGDL NPDNSDVDKLFIQLVQTYNQLFEENPINAS GVDAKAILSARLSKSRRLENLIAQLPGEKK NGLFGNLIALSLGLTPNFKSNFDLAEDAKL QLSKDTYDDDLDNLLAQIGDQYADLFLAA KNLSDAILLSDILRVNTEITKAPLSASMIKR YDEHHQDLTLLKALVRQQLPEKYKEIFFD QSKNGYAGYIDGGASQEEFYKFIKPILEKM DGTEELLVKLNREDLLRKQRTFDNGSIPH QIHLGELHAILRRQEDFYPFLKDNREKIEKI LTFRIPYYVGPLARGNSRFAWMTRKSEETI TPWNFEEVVDKGASAQSFIERMTNFDKNL

[1327]

		PNEKVLPKHSLLYEYFTVYNELTKVKYVT EGMRKPAFLSGEQKKAIVDLLFKTNRKVT VKQLKEDYFKKIECFDSVEISGVEDRFNAS LGTYHDLLKIIKDKDFLDNEENEDILEDIVL TLTLFEDREMIEERLKTYAHLFDDKVMKQ LKRRRYTGWGRLSRKLINGIRDKQSGKTIL DFLKSDGFANRNFMQLIHDDSLTFKEDIQK AQVSGQGDSLHEHIANLAGSPAIKKGILQT VKVVDELVKVMGRHKPENIVIEMARENQ TTQKGQKNSRERMKRIEEGIKELGSQILKE HPVENTQLQNEKLYLYYLQNGRDMYVDQ ELDINRLSDYDVDHIVPQSFLKDDSIDNKV LTRSDKNRGKSDNVPSEEVVKKMKNYWR QLLNAKLITQRKFDNLTKAERGGLSELDK AGFIKRQLVETRQITKHVAQILDSRMNTK YDENDKLIREVKVITLKSKLVSDFRKDFQF YKVREINNYHHAHDAYLNAVVGTALIKK YPKLESEFVYGDYKVYDVRKMIAKSEQEI GKATAKYFFYSNIMNFFKTEITLANGEIRK RPLIETNGETGEIVWDKGRDFATVRKVLS MPQVNIVKKTEVQTGGFSKESILPKRNSDK LIARKKDWDPKKYGGFDSPTVAYSVLVV AKVEKGKSKKLKSVKELLGITIMERSSFEK NPIDFLEAKGYKEVKKDLIIKLPKYSLFELE NGRKRMLASAGELQKGNELALPSKYVNF LYLASHYEKLKGSPEDNEQKQLFVEQHKH YLDEIIEQISEFSKRVILADANLDKVLSAYN KHRDKPIREQAENIIHLFTLTNLGAPAAFK YFDTTIDRKRYTSTKEVLDATLIHQSITGL YETRIDLSQLGGD
115.1 Cas9 TadAins1 004 多肽序列	1338	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE

[1328]

		ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKGS SGSETPGTSESATPESSGSEVEFSHEYWMR HALTLAKRARDEREVPVGAVLVLNNRVIG EGWNRAIGLHDPTAHAEIMALRQGGLVM QNYRLIDATLYVTFEPCVMCAGAMIHSRI GRVVFGVRNAKTGAAGSLMDVLHYPGM NHRVEITEGILADECAALLCYFFRMPRQLE SEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFDSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIE QISEFSKRVILADANLDKVLSAYNKHRDK PIREQAENIIHLFTLTNLGAPAAFKYFDTTI DRKRYTSTKEVLDATLIHQSITGLYETRID LSQLGGD
115.2 Cas9 TadAins1	1339	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE

[1329]

005 多肽序列		MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLG SSGSETPGTSESATPESSGSEVEFSHEYWM RHALTLAKRARDEREVPVGAVLVLNNRVI GEGWNRAIGLHDPTAHAEIMALRQGGLV MQNYRLIDATLYVTFEPCVMCAGAMIHSR IGRVVFGVRNAKTGAAGSLMDVLHYPGM NHRVEITEGILADECAALLCYFFRMPRQES EFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFDSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE

		AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIE QISEFSKRVILADANLDKVLSAYNKHRDK PIREQAENIIHLFTLTNLGAPAAFKYFDTTI DRKRYTSTKEVLDATLIHQSITGLYETRID LSQLGGD
115.3 Cas9 TadAins1006 多肽序列	1340	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE GSSGSETPGTSESATPESSGSEVEFSHEYW MRHALTLAKRARDEREVPVGAVLVLNNR

[1330]

			VIGEGWNRAIGLHDPTAHAEIMALRQGGL VMQNYRLIDATLYVTFEPCVMCAGAMIHS RIGRVVFGVRNAKTGAAGSLMDVLHYPG MNHRVEITEGILADECAALLCYFFRMPRQ SEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFDSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIE QISEFSKRVILADANLDKVLSAYNKHRDK PIREQAENIIHLFTLTNLGAPAAFKYFDTTI DRKRYTSTKEVLDATLIHQSITGLYETRID LSQLGGD
[1331]	115.4 Cas9 TadAins1007 多肽序列	1341	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK

[1332]

		GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SGSSGSETPGTSESATPESSGSEVEFSHEYW MRHALTLAKRARDEREVPVGAVLVLNNR VIGEGWNRAIGLHDPTAHAEIMALRQGGL VMQNYRLIDATLYVTFEPCVMCAGAMIHS RIGRVVFGVRNAKTGAAGSLMDVLHYPG MNHRVEITEGILADECAALLCYFFRMPRQ EFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFDSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIE QISEFSKRVILADANLDKVLSAYNKHRDK PIREQAENIIHLFTLTNLGAPAAFKYFDTTI DRKRYTSTKEVLDATLIHQSITGLYETRID LSQLGGD
116.1 Cas9 TadAinsC 端截短的 2 792 多肽序列	1342	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN

[1333]

		FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGGSSGSETPGT SESATPESSGSEVEFSHEYWMRHALTLAK RARDEREVPVGAVLVLNNRVIGEGWNRAI GLHDPTAHAEIMALRQGGLVMQNYRLID ATLYVTFEPCVMCAGAMIHSRIGRVVFGV RNAKTGAAGSLMDVLHYPGMNHRVEITE GILADECAALLCYFFRMPRQSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFDSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIE QISEFSKRVILADANLDKVLSAYNKHRDK PIREQAENIIHLFTLTNLGAPAAFKYFDTTI DRKRYTSTKEVLDATLIHQSITGLYETRID LSQLGGD
116.2 Cas9 TadAinsC 端截短的 2 791	1343	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN

[1334]

多肽序列		SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSSGSETPGTS ESATPESSGSEVEFSHEYWMRHALTLAKR ARDEREVPVGAVLVLNNRVIGEGWNRAIG LHDPTAHAEIMALRQGGLVMQNYRLIDAT LYVTFEPCVMCAGAMIHSRIGRVVFGVRN AKTGAAGSLMDVLHYPGMNHRVEITEGIL ADECAALLCYFFRMPRQGSQILKEHPVEN TQLQNEKLYLYYLQNGRDMYVDQELDIN RLSDYDVDHIVPQSFLKDDSIDNKVLTRSD KNRGKSDNVPSEEVVKKMKNYWRQLLN AKLITQRKFDNLTKAERGGLSELDKAGFIK RQLVETRQITKHVAQILDSRMNTKYDEND KLIREVKVITLKSKLVSDFRKDFQFYKVRE INNYHHAHDAYLNAVVGTALIKKYPKLES EFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFDSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIE

[1335]

		QISEFSKRVILADANLDKVLSAYNKHRDK PIREQAENIIHLFTLTNLGAPAAFKYFDTTI DRKRYTSTKEVLDATLIHQSITGLYETRID LSQLGGD
116.3 Cas9 TadAins C 端截短 的 2 790 多肽序列	1344	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKEGSSGSETPGTSE SATPESSGSEVEFSHEYWMRHALTLAKRA RDEREVPVGAVLVLNNRVIGEGWNRAIGL HDPTAHAEIMALRQGGLVMQNYRLIDATL YVTFEPCVMCAGAMIHSRIGRVVFGVRNA KTGAAGSLMDVLHYPGMNHRVEITEGILA DECAALLCYFFRMPRQLGSQILKEHPVEN TQLQNEKLYLYYLQNGRDMYVDQELDIN RLSDYDVDHIVPQSFLKDDSIDNKVLTRSD KNRGKSDNVPSEEVVKKMKNYWRQLLN AKLITQRKFDNLTKAERGGLSELDKAGFIK RQLVETRQITKHVAQILDSRMNTKYDEND KLIREVKVITLKSKLVSDFRKDFQFYKVRE

		INNYHHAHDAYLNAVVGTALIKKYPKLES EFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFDSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIE QISEFSKRVILADANLDKVLSAYNKHRDK PIREQAENIIHLFTLTNLGAPAAFKYFDTTI DRKRYTSTKEVLDATLIHQSITGLYETRID LSQLGGD
[1336]	117 Cas9 Δ1017-1069 多肽序列	1345	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS

[1337]

		DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYSSGSEVEFSHEYWMRHA LTLAKRARDEREVPVGAVLVLNNRVIGEG WNRAIGLHDPTAHAEIMALRQGGLVMQN YRLIDATLYVTFEPCVMCAGAMIHSRIGR VVFGVRNAKTGAAGSLMDVLHYPGMNH RVEITEGILADECAALLCYFFRMPRQVFNA QKKAQSSTDGEIVWDKGRDFATVRKVLS MPQVNIVKKTEVQTGGFSKESILPKRNSDK LIARKKDWDPKKYGGFDSPTVAYSVLVV AKVEKGKSKKLKSVKELLGITIMERSSFEK NPIDFLEAKGYKEVKKDLIIKLPKYSLFELE NGRKRMLASAGELQKGNELALPSKYVNF LYLASHYEKLKGSPEDNEQKQLFVEQHKH YLDEIIEQISEFSKRVILADANLDKVLSAYN KHRDKPIREQAENIIHLFTLTNLGAPAAFK YFDTTIDRKRYTSTKEVLDATLIHQSITGL YETRIDLSQLGGD
118 Cas9 TadA-CP 116ins 1067 多肽序列	1346	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL

[1338]

		FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NMNHRVEITEGILADECAALLCYFFRMPR QVFNAQKKAQSSTDGSSGSETPGTSESATP ESSGSEVEFSHEYWMRHALTLAKRARDER EVPVGAVLVLNNRVIGEGWNRAIGLHDPT AHAEIMALRQGGLVMQNYRLIDATLYVTF EPCVMCAGAMIHSRIGRVVFGVRNAKTG AAGSLMDVLHYPGGETGEIVWDKGRDFA TVRKVLSMPQVNIVKKTEVQTGGFSKESIL PKRNSDKLIARKKDWDPKKYGGFDSPTVA YSVLVVAKVEKGKSKKLKSVKELLGITIM ERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK YSLFELENGRKRMLASAGELQKGNELALP SKYVNFLYLASHYEKLKGSPEDNEQKQLF VEQHKHYLDEIIEQISEFSKRVILADANLD KVLSAYNKHRDKPIREQAENIIHLFTLTNL GAPAAFKYFDTTIDRKRYTSTKEVLDATLI HQSITGLYETRIDLSQLGGD
119 Cas9 TadAins 701 多肽序列	1347	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH

[1339]

		HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSGSSGSETPGTSES ATPESSGSEVEFSHEYWMRHALTLAKRAR DEREVPVGAVLVLNNRVIGEGWNRAIGLH DPTAHAEIMALRQGGLVMQNYRLIDATLY VTFEPCVMCAGAMIHSRIGRVVFGVRNAK TGAAGSLMDVLHYPGMNHRVEITEGILAD ECAALLCYFFRMPRQVFNAQKKAQSSTDL TFKEDIQKAQVSGQGDSLHEHIANLAGSP AIKKGILQTVKVVDELVKVMGRHKPENIV IEMARENQTTQKGQKNSRERMKRIEEGIK ELGSQILKEHPVENTQLQNEKLYLYYLQN GRDMYVDQELDINRLSDYDVDHIVPQSFL KDDSIDNKVLTRSDKNRGKSDNVPSEEVV KKMKNYWRQLLNAKLITQRKFDNLTKAE RGGLSELDKAGFIKRQLVETRQITKHVAQI LDSRMNTKYDENDKLIREVKVITLKSKLV SDFRKDFQFYKVREINNYHHAHDAYLNA VVGTALIKKYPKLESEFVYGDYKVYDVRK MIAKSEQEIGKATAKYFFYSNIMNFFKTEI TLANGEIRKRPLIETNGETGEIVWDKGRDF ATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFDSPT VAYSVLVVAKVEKGKSKKLKSVKELLGIT IMERSSFEKNPIDFLEAKGYKEVKKDLIIKL PKYSLFELENGRKRMLASAGELQKGNELA LPSKYVNFLYLASHYEKLKGSPEDNEQKQ LFVEQHKHYLDEIIEQISEFSKRVILADANL DKVLSAYNKHRDKPIREQAENIIHLFTLTN LGAPAAFKYFDTTIDRKRYTSTKEVLDAT LIHQSITGLYETRIDLSQLGGD
120 Cas9	1348	MDKKYSIGLAIGTNSVGWAVITDEYKVPS

[1340]

TadACP1 36ins 1248 多肽序列		KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFDSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSMNHRVEITEGILADECAALLCY FFRMPRQVFNAQKKAQSSTDGSSGSETPG

[1341]

		TSESATPESSGSEVEFSHEYWMRHALTLA KRARDEREVPVGAVLVLNNRVIGEGWNR AIGLHDPTAHAEIMALRQGGLVMQNYRLI DATLYVTFEPCVMCAGAMIHSRIGRVVFG VRNAKTGAAGSLMDVLHYPGPEDNEQKQ LFVEQHKHYLDEIIEQISEFSKRVILADANL DKVLSAYNKHRDKPIREQAENIIHLFTLTN LGAPAAFKYFDTTIDRKRYTSTKEVLDAT LIHQSITGLYETRIDLSQLGGD
121 Cas9 TadACP1 36ins 1052 多肽序列	1349	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE

		SEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLAMNHRVEITEGIL ADECAALLCYFFRMPRQVFNAQKKAQSST DGSSGSETPGTSESATPESSGSEVEFSHEY WMRHALTLAKRARDEREVPVGAVLVLNN RVIGEGWNRAIGLHDPTAHAEIMALRQGG LVMQNYRLIDATLYVTFEPCVMCAGAMI HSRIGRVVFGVRNAKTGAAGSLMDVLHY PGNGEIRKRPLIETNGETGEIVWDKGRDFA TVRKVLSMPQVNIVKKTEVQTGGFSKESIL PKRNSDKLIARKKDWDPKKYGGFDSPTVA YSVLVVAKVEKGKSKKLKSVKELLGITIM ERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK YSLFELENGRKRMLASAGELQKGNELALP SKYVNFLYLASHYEKLKGSPEDNEQKQLF VEQHKHYLDEIIEQISEFSKRVILADANLD KVLSAYNKHRDKPIREQAENIIHLFTLTNL GAPAAFKYFDTTIDRKRYTSTKEVLDATLI HQSITGLYETRIDLSQLGGD
[1342]	122 Cas9 TadACP1 36ins 1041 多肽序列	1350	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV

[1343]

		SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSMNHRVEITEGILADECAALLCYFF RMPRQVFNAQKKAQSSTDGSSGSETPGTS ESATPESSGSEVEFSHEYWMRHALTLAKR ARDEREVPVGAVLVLNNRVIGEGWNRAIG LHDPTAHAEIMALRQGGLVMQNYRLIDAT LYVTFEPCVMCAGAMIHSRIGRVVFGVRN AKTGAAGSLMDVLHYPGNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFAT VRKVLSMPQVNIVKKTEVQTGGFSKESILP KRNSDKLIARKKDWDPKKYGGFDSPTVA YSVLVVAKVEKGKSKKLKSVKELLGITIM ERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK YSLFELENGRKRMLASAGELQKGNELALP SKYVNFLYLASHYEKLKGSPEDNEQKQLF VEQHKHYLDEIIEQISEFSKRVILADANLD KVLSAYNKHRDKPIREQAENIIHLFTLTNL GAPAAFKYFDTTIDRKRYTSTKEVLDATLI HQSITGLYETRIDLSQLGGD
123 Cas9 TadACP1 39ins 1299 多肽序列	1351	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG

[1344]

		ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFDSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIE QISEFSKRVILADANLDKVLSAYNKHRMN HRVEITEGILADECAALLCYFFRMPRQVFN AQKKAQSSTDGSSGSETPGTSESATPESSG SEVEFSHEYWMRHALTLAKRARDEREVP VGAVLVLNNRVIGEGWNRAIGLHDPTAH AEIMALRQGGLVMQNYRLIDATLYVTFEP CVMCAGAMIHSRIGRVVFGVRNAKTGAA GSLMDVLHYPGDKPIREQAENIIHLFTLTN LGAPAAFKYFDTTIDRKRYTSTKEVLDAT LIHQSITGLYETRIDLSQLGGD
124 Cas9 Δ792-872 TadAins 多肽序列	1352	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF

[1345]

		GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSEVEFSHEY WMRHALTLAKRARDEREVPVGAVLVLNN RVIGEGWNRAIGLHDPTAHAEIMALRQGG LVMQNYRLIDATLYVTFEPCVMCAGAMI HSRIGRVVFGVRNAKTGAAGSLMDVLHY PGMNHRVEITEGILADECAALLCYFFRMP RQVFNAQKKAQSSTDEEVVKKMKNYWR QLLNAKLITQRKFDNLTKAERGGLSELDK AGFIKRQLVETRQITKHVAQILDSRMNTK YDENDKLIREVKVITLKSKLVSDFRKDFQF YKVREINNYHHAHDAYLNAVVGTALIKK YPKLESEFVYGDYKVYDVRKMIAKSEQEI GKATAKYFFYSNIMNFFKTEITLANGEIRK RPLIETNGETGEIVWDKGRDFATVRKVLS MPQVNIVKKTEVQTGGFSKESILPKRNSDK LIARKKDWDPKKYGGFDSPTVAYSVLVV AKVEKGKSKKLKSVKELLGITIMERSSFEK NPIDFLEAKGYKEVKKDLIIKLPKYSLFELE NGRKRMLASAGELQKGNELALPSKYVNF LYLASHYEKLKGSPEDNEQKQLFVEQHKH YLDEIIEQISEFSKRVILADANLDKVLSAYN

[1346]

		KHRDKPIREQAENIIHLFTLTNLGAPAAFK YFDTTIDRKRYTSTKEVLDATLIHQSITGL YETRIDLSQLGGD
125 Cas9 Δ792-906 TadAins 多肽序列	1353	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSEVEFSHEY WMRHALTLAKRARDEREVPVGAVLVLNN RVIGEGWNRAIGLHDPTAHAEIMALRQGG LVMQNYRLIDATLYVTFEPCVMCAGAMI HSRIGRVVFGVRNAKTGAAGSLMDVLHY PGMNHRVEITEGILADECAALLCYFFRMP RQVFNAQKKAQSSTDGLSELDKAGFIKRQ LVETRQITKHVAQILDSRMNTKYDENDKL IREVKVITLKSKLVSDFRKDFQFYKVREIN NYHHAHDAYLNAVVGTALIKKYPKLESEF VYGDYKVYDVRKMIAKSEQEIGKATAKY FFYSNIMNFFKTEITLANGEIRKRPLIETNG ETGEIVWDKGRDFATVRKVLSMPQVNIVK KTEVQTGGFSKESILPKRNSDKLIARKKD

[1347]

		WDPKKYGGFDSPTVAYSVLVVAKVEKGK SKKLKSVKELLGITIMERSSFEKNPIDFLEA KGYKEVKKDLIIKLPKYSLFELENGRKRM LASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQ ISEFSKRVILADANLDKVLSAYNKHRDKPI REQAENIIHLFTLTNLGAPAAFKYFDTTIDR KRYTSTKEVLDATLIHQSITGLYETRIDLSQ LGGD
126 TadA CP65ins 1003 多肽序列	1354	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKTA

			HAEIMALRQGGLVMQNYRLIDATLYVTFE PCVMCAGAMIHSRIGRVVFGVRNAKTGA AGSLMDVLHYPGMNHRVEITEGILADECA ALLCYFFRMPRQVFNAQKKAQSSTDGSSG SETPGTSESATPESSGSEVEFSHEYWMRHA LTLAKRARDEREVPVGAVLVLNNRVIGEG WNRAIGLHDPLESEFVYGDYKVYDVRKM IAKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFAT VRKVLSMPQVNIVKKTEVQTGGFSKESILP KRNSDKLIARKKDWDPKKYGGFDSPTVA YSVLVVAKVEKGKSKKLKSVKELLGITIM ERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK YSLFELENGRKRMLASAGELQKGNELALP SKYVNFLYLASHYEKLKGSPEDNEQKQLF VEQHKHYLDEIIEQISEFSKRVILADANLD KVLSAYNKHRDKPIREQAENIIHLFTLTNL GAPAAFKYFDTTIDRKRYTSTKEVLDATLI HQSITGLYETRIDLSQLGGD
[1348]	127 TadA CP65ins 1016 多肽序列	1355	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV

[1349]

		SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVTAHAEIMALRQGGLVMQN YRLIDATLYVTFEPCVMCAGAMIHSRIGR VVFGVRNAKTGAAGSLMDVLHYPGMNH RVEITEGILADECAALLCYFFRMPRQVFNA QKKAQSSTDGSSGSETPGTSESATPESSGS EVEFSHEYWMRHALTLAKRARDEREVPV GAVLVLNNRVIGEGWNRAIGLHDPYDVR KMIAKSEQEIGKATAKYFFYSNIMNFFKTE ITLANGEIRKRPLIETNGETGEIVWDKGRD FATVRKVLSMPQVNIVKKTEVQTGGFSKE SILPKRNSDKLIARKKDWDPKKYGGFDSP TVAYSVLVVAKVEKGKSKKLKSVKELLGI TIMERSSFEKNPIDFLEAKGYKEVKKDLIIK LPKYSLFELENGRKRMLASAGELQKGNEL ALPSKYVNFLYLASHYEKLKGSPEDNEQK QLFVEQHKHYLDEIIEQISEFSKRVILADAN LDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTIDRKRYTSTKEVLDA TLIHQSITGLYETRIDLSQLGGD
128 TadA CP65ins 1022 多肽序列	1356	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG

[1350]

		ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMITAHAEIMALRQ GGLVMQNYRLIDATLYVTFEPCVMCAGA MIHSRIGRVVFGVRNAKTGAAGSLMDVL HYPGMNHRVEITEGILADECAALLCYFFR MPRQVFNAQKKAQSSTDGSSGSETPGTSE SATPESSGSEVEFSHEYWMRHALTLAKRA RDEREVPVGAVLVLNNRVIGEGWNRAIGL HDPAKSEQEIGKATAKYFFYSNIMNFFKTE ITLANGEIRKRPLIETNGETGEIVWDKGRD FATVRKVLSMPQVNIVKKTEVQTGGFSKE SILPKRNSDKLIARKKDWDPKKYGGFDSP TVAYSVLVVAKVEKGKSKKLKSVKELLGI TIMERSSFEKNPIDFLEAKGYKEVKKDLIIK LPKYSLFELENGRKRMLASAGELQKGNEL ALPSKYVNFLYLASHYEKLKGSPEDNEQK QLFVEQHKHYLDEIIEQISEFSKRVILADAN LDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTIDRKRYTSTKEVLDA TLIHQSITGLYETRIDLSQLGGD
129 TadA CP65ins 1029 多肽序列	1357	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF

[1351]

		GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEITAHAE IMALRQGGLVMQNYRLIDATLYVTFEPCV MCAGAMIHSRIGRVVFGVRNAKTGAAGS LMDVLHYPGMNHRVEITEGILADECAALL CYFFRMPRQVFNAQKKAQSSTDGSSGSET PGTSESATPESSGSEVEFSHEYWMRHALTL AKRARDEREVPVGAVLVLNNRVIGEGWN RAIGLHDPGKATAKYFFYSNIMNFFKTEIT LANGEIRKRPLIETNGETGEIVWDKGRDFA TVRKVLSMPQVNIVKKTEVQTGGFSKESIL PKRNSDKLIARKKDWDPKKYGGFDSPTVA YSVLVVAKVEKGKSKKLKSVKELLGITIM ERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK

[1352]

		YSLFELENGRKRMLASAGELQKGNELALP SKYVNFLYLASHYEKLKGSPEDNEQKQLF VEQHKHYLDEIIEQISEFSKRVILADANLD KVLSAYNKHRDKPIREQAENIIHLFTLTNL GAPAAFKYFDTTIDRKRYTSTKEVLDATLI HQSITGLYETRIDLSQLGGD
130 TadA CP65ins 1041 多肽序列	1358	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSTAHAEIMALRQGGLVMQNYRLID ATLYVTFEPCVMCAGAMIHSRIGRVVFGV

		RNAKTGAAGSLMDVLHYPGMNHRVEITE GILADECAALLCYFFRMPRQVFNAQKKAQ SSTDGSSGSETPGTSESATPESSGSEVEFSH EYWMRHALTLAKRARDEREVPVGAVLVL NNRVIGEGWNRAIGLHDPNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFAT VRKVLSMPQVNIVKKTEVQTGGFSKESILP KRNSDKLIARKKDWDPKKYGGFDSPTVA YSVLVVAKVEKGKSKKLKSVKELLGITIM ERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK YSLFELENGRKRMLASAGELQKGNELALP SKYVNFLYLASHYEKLKGSPEDNEQKQLF VEQHKHYLDEIIEQISEFSKRVILADANLD KVLSAYNKHRDKPIREQAENIIHLFTLTNL GAPAAFKYFDTTIDRKRYTSTKEVLDATLI HQSITGLYETRIDLSQLGGD
[1353]	131 TadA CP65ins 1054 多肽序列	1359	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE

[1354]

		NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANTAHAEIMALR QGGLVMQNYRLIDATLYVTFEPCVMCAG AMIHSRIGRVVFGVRNAKTGAAGSLMDV LHYPGMNHRVEITEGILADECAALLCYFFR MPRQVFNAQKKAQSSTDGSSGSETPGTSE SATPESSGSEVEFSHEYWMRHALTLAKRA RDEREVPVGAVLVLNNRVIGEGWNRAIGL HDPGEIRKRPLIETNGETGEIVWDKGRDFA TVRKVLSMPQVNIVKKTEVQTGGFSKESIL PKRNSDKLIARKKDWDPKKYGGFDSPTVA YSVLVVAKVEKGKSKKLKSVKELLGITIM ERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK YSLFELENGRKRMLASAGELQKGNELALP SKYVNFLYLASHYEKLKGSPEDNEQKQLF VEQHKHYLDEIIEQISEFSKRVILADANLD KVLSAYNKHRDKPIREQAENIIHLFTLTNL GAPAAFKYFDTTIDRKRYTSTKEVLDATLI HQSITGLYETRIDLSQLGGD
132 TadA CP65ins 1246 多肽序列	1360	MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN FEEVVDKGASAQSFIERMTNFDKNLPNEK

[1355]

		VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFDSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASAGELQKGNELALPSKYVNFLYLASH YEKLKGTAHAEIMALRQGGLVMQNYRLI DATLYVTFEPCVMCAGAMIHSRIGRVVFG VRNAKTGAAGSLMDVLHYPGMNHRVEIT EGILADECAALLCYFFRMPRQVFNAQKKA QSSTDGSSGSETPGTSESATPESSGSEVEFS HEYWMRHALTLAKRARDEREVPVGAVLV LNNRVIGEGWNRAIGLHDPSPEDNEQKQL FVEQHKHYLDEIIEQISEFSKRVILADANLD KVLSAYNKHRDKPIREQAENIIHLFTLTNL GAPAAFKYFDTTIDRKRYTSTKEVLDATLI HQSITGLYETRIDLSQLGGD
TadA 多肽序列	1363	MGSHMTNDIYFMTLAIEEAKKAAQLGEVP IGAIITKDDEVIARAHNLRETLQQPTAHAE HIAIERAAKVLGSWRLEGCTLYVTLEPCV MCAGTIVMSRIPRVVYGADDPKGGCSGSL MNLLQQSNFNHRAIVDKGVLKEACSTLLT TFFKNLRANKKSTN
TadA 多	1364	MTQDELYMKEAIKEAKKAEEKGEVPIGAV

[1356]

肽序列		LVINGEIIARAHNLRETEQRSIAHAEMLVID EACKALGTWRLEGATLYVTLEPCPMCAG AVVLSRVEKVVFGAFDPKGGCSGTLMNL LQEERFNHQAEVVSGVLEEECGGMLSAFF RELRKKKKAARKNLSE
TadA 多肽序列	1365	MPPAFITGVTSLSDVELDHEYWMRHALTL AKRAWDEREVPVGAVLVHNHRVIGEGW NRPIGRHDPTAHAEIMALRQGGLVLQNYR LLDTTLYVTLEPCVMCAGAMVHSRIGRVV FGARDAKTGAAGSLIDVLHHPGMNHRVEI IEGVLRDECATLLSDFFRMRRQEIKALKKA DRAEGAGPAV
TadA 多肽序列	1366	MDEYWMQVAMQMAEKAEAAGEVPVGA VLVKDGQQIATGYNLSISQHDPTAHAEILC LRSAGKKLENYRLLDATLYITLEPCAMCA GAMVHSRIARVVYGARDEKTGAAGTVVN LLQHPAFNHQVEVTSGVLAEACSAQLSRF FKRRRDEKKALKLAQRAQQGIE
TadA 多肽序列	1367	MDAAKVRSEFDEKMMRYALELADKAEA LGEIPVGAVLVDDARNIIGEGWNLSIVQSD PTAHAEIIALRNGAKNIQNYRLLNSTLYVT LEPCTMCAGAILHSRIKRLVFGASDYKTG AIGSRFHFFDDYKMNHTLEITSGVLAEECS QKLSTFFQKRREEKKIEKALLKSLSDK
TadA 多肽序列	1368	MRTDESEDQDHRMMRLALDAARAAAEA GETPVGAVILDPSTGEVIATAGNGPIAAHD PTAHAEIAAMRAAAAKLGNYRLTDLTLV VTLEPCAMCAGAISHARIGRVVFGADDPK GGAVVHGPKFFAQPTCHWRPEVTGGVLA DESADLLRGFFRARRKAKI
TadA 多肽序列	1369	MSSLKKTPIRDDAYWMGKAIREAAKAAA RDEVPIGAVIVRDGAVIGRGHNLREGSNDP SAHAEMIAIRQAARRSANWRLTGATLYVT LEPCLMCMGAIILARLERVVFGCYDPKGG AAGSLYDLSADPRLNHQVRLSPGVCQEEC GTMLSDFFRDLRRRKKAKATPALFIDERK VPPEP
ecTadA 多肽序列	1370	MSEVEFSHEYWMRHALTLAKRARDEREV PVGAVLVLNNRVIGEGWNRAIGLHDPTAH AEIMALRQGGLVMQNYRLIDATLYVTFEP CVMCAGAMIHSRIGRVVFGVRNAKTGAA GSLMDVLHYPGMNHRVEITEGILADECAA LLCYFFRMPRQVFNAQKKAQSSTD

[1357]

TadA*7.10	8	MSEVEFSHEYWMRHALTLAKRARDEREV PVGAVLVLNNRVIGEGWNRAIGLHDPTAH AEIMALRQGGLVMQNYRLIDATLYVTFEP CVMCAGAMIHSRIGRVVFGVRNAKTGAA GSLMDVLHYPGMNHRVEITEGILADECAA LLCYFFRMPRQVFNAQKKAQSSTD
TadA*8	12	MSEVEFSHEYWMRHALTLAKRARDEREV PVGAVLVLNNRVIGEGWNRAIGLHDPTAH AEIMALRQGGLVMQNYRLIDATLYVTFEP CVMCAGAMIHSRIGRVVFGVRNAKTGAA GSLMDVLHYPGMNHRVEITEGILADECAA LLCTFFRMPRQVFNAQKKAQSSTD
gRNA 支架核苷酸序列	224	GUUUUAGAGCUAGAAAUAGCAAGUUAA AAUAAGGCUAGUCCGUUAUCAACUUGA AAAAGUGGCACCGAGUCGGUGCUUUU
gRNA 支架核苷酸序列	225	GUUUUUGUACUCUCAAGAUUUAAGUAA CUGUACAACGAAACUUACACAGUUACU UAAAUCUUGCAGAAGCUACAAAGAUAA GGCUUCAUGCCGAAAUCAACACCCUGU CAUUUUAUGGCAGGGUG
酿脓链球菌 gRNA 支架核苷酸序列	226	GUUUUAGAGCUAGAAAUAGCAAGUUAA AAUAAGGCUAGUCCGUUAUCAACUUGA AAAAGUGGCACCGAGUCGGUGC
金黄色葡萄球菌 gRNA 支架核苷酸序列	227	GUUUUAGUACUCUGUAAUGAAAAUUAC AGAAUCUACUAAAACAAGGCAAAAUGC CGUGUUUAUCUCGUCAACUUGUUGGCG AGA
BhCas12b gRNA 支架核苷酸序列	228	GUUCUGUCUUUUGGUCAGGACAACCGU CUAGCUAUAAGUGCUGCAGGGUGUGAG AAACUCCUAUUGCUGGACGAUGUCUCU UACGAGGCAUUAGCAC
BvCas12b gRNA 支架核苷酸序列	229	GACCUAUAGGGUCAAUGAAUCUGUGCG UGUGCCAUAAGUAAUUAAAAAUUACCC ACCACAGGAGCACCUGAAAACAGGUGC UUGGCAC
gRNA 支架核苷酸序列	230	GUUUUAGAGCUAGAAAUAGCAAGUUAA AAUAAGGCUAGUCCGUUAUCAACUUGA AAAAGUGGGACCGAGUCGGUGCUUUU

[1358]

gRNA 支架核苷酸序列	3000	GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU
BhCas12b gRNA 支架+引导序列	243	GUUCUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUGCAGGGUGUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUACGAGGCAUUAGCACNNNNNNNNNNNNNNNNNNN
BvCas12b gRNA 支架+引导序列	244	GACCUAUAGGGUCAAUGAAUCUGUGCGUGUGCCAUAAGUAAUUAAAAAUUACCCACCACAGGAGCACCUGAAAACAGGUGCUUGGCACNNNNNNNNNNNNNNNNNNN
AaCas12b gRNA 支架+引导序列	245	GUCUAAAGGACAGAAUUUUUCAACGGGUGUGCCAAUGGCCACUUUCCAGGUGGCAAAGCCCGUUGAACUUCUCAAAAAGAACGAUCUGAGAAGUGGCACNNNNNNNNNNNNNNNNNNN
SpyMacCas9 多肽序列	1307	MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQIYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRGMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGHSLHEQIANLAGSPAIKKGILQTVK

[1359]

		IVDELVKVMGHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFIKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEIQTVGQNGGLFDDNPKSPLEVTPS KLVPLKKELNPKKYGGYQKPTTAYPVLLI TDTKQLIPISVMNKKQFEQNPVKFLRDRG YQQVGKNDFIKLPKYTLVDIGDGIKRLWA SSKEIHKGNQLVVSKKSQILLYHAHHLDS DLSNDYLQNHNQQFDVLFNEIISFSKKCKL GKEHIQKIENVYSNKKNSASIEELAESFIKL LGFTQLGATSPFNFLGVKLNQKQYKGKKD YILPCTEGTLIRQSITGLYETRVDLSKIGED
NLS	83	PKKKRKVEGADKRTADGSEFESPKKKRK V
NLS	84	KRTADGSEFESPKKKRKV
NLS	85	KRPAATKKAGQAKKKK
NLS	86	KKTELQTTNAENKTKKL
NLS	87	KRGINDRNFWRGENGRKTR
NLS	1424	RKSGKIAAIVVKRPRKPKKKRKV
NLS	90	MDSLLMNRRKFLYQFKNVRWAKGRRETY LC
接头	1425	(SGGS)2
pNMG-B 335 ABE8.1_ Y147T_C P5_NGC PAM_单体多肽序列	1426	MSEVEFSHEYWMRHALTLAKRARDEREV PVGAVLVLNNRVIGEGWNRAIGLHDPTAH AEIMALRQGGLVMQNYRLIDATLYVTFEP CVMCAGAMIHSRIGRVVFGVRNAKTGAA GSLMDVLHYPGMNHRVEITEGILADECAA LLCTFFRMPRQVFNAQKKAQSSTDSGGSS GGSSGSETPGTSESATPESSGGSSGGSEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRP LIETNGETGEIVWDKGRDFATVRKVLSMP QVNIVKKTEVQTGGFSKESILPKRNSDKLI ARKKDWDPKKYGGFMQPTVAYSVLVVA

[1360]

		KVEKGKSKKLKSVKELLGITIMERSSFEKN PIDFLEAKGYKEVKKDLIIKLPKYSLFELEN GRKRMLASAKFLQKGNELALPSKYVNFL YLASHYEKLKGSPEDNEQKQLFVEQHKH YLDEIIEQISEFSKRVILADANLDKVLSAYN KHRDKPIREQAENIIHLFTLTNLGAPRAFK YFDTTIARKEYRSTKEVLDATLIHQSITGL YETRIDLSQLGGDGGSGGSGGSGGSGGSG GSGGMDKKYSIGLAIGTNSVGWAVITDEY KVPSKKFKVLGNTDRHSIKKNLIGALLFDS GETAEATRLKRTARRRYTRRKNRICYLQEI FSNEMAKVDDSFFHRLEESFLVEEDKKHE RHPIFGNIVDEVAYHEKYPTIYHLRKKLVD STDKADLRLIYLALAHMIKFRGHFLIEGDL NPDNSDVDKLFIQLVQTYNQLFEENPINAS GVDAKAILSARLSKSRRLENLIAQLPGEKK NGLFGNLIALSLGLTPNFKSNFDLAEDAKL QLSKDTYDDDLDNLLAQIGDQYADLFLAA KNLSDAILLSDILRVNTEITKAPLSASMIKR YDEHHQDLTLLKALVRQQLPEKYKEIFFD QSKNGYAGYIDGGASQEEFYKFIKPILEKM DGTEELLVKLNREDLLRKQRTFDNGSIPH QIHLGELHAILRRQEDFYPFLKDNREKIEKI LTFRIPYYVGPLARGNSRFAWMTRKSEETI TPWNFEEVVDKGASAQSFIERMTNFDKNL PNEKVLPKHSLLYEYFTVYNELTKVKYVT EGMRKPAFLSGEQKKAIVDLLFKTNRKVT VKQLKEDYFKKIECFDSVEISGVEDRFNAS LGTYHDLLKIIKDKDFLDNEENEDILEDIVL TLTLFEDREMIEERLKTYAHLFDDKVMKQ LKRRRYTGWGRLSRKLINGIRDKQSGKTIL DFLKSDGFANRNFMQLIHDDSLTFKEDIQK AQVSGQGDSLHEHIANLAGSPAIKKGILQT VKVVDELVKVMGRHKPENIVIEMARENQ TTQKGQKNSRERMKRIEEGIKELGSQILKE HPVENTQLQNEKLYLYYLQNGRDMYVDQ ELDINRLSDYDVDHIVPQSFLKDDSIDNKV LTRSDKNRGKSDNVPSEEVVKKMKNYWR QLLNAKLITQRKFDNLTKAERGGLSELDK AGFIKRQLVETRQITKHVAQILDSRMNTK YDENDKLIREVKVITLKSKLVSDFRKDFQF YKVREINNYHHAHDAYLNAVVGTALIKK YPKLESEFVYGDYKVYDVRKMIAKSEQEG

[1361]

		ADKRTADGSEFESPKKKRKV
具有NGC PAM CP5多肽序列的pNMG-357_ABE8.14	1427	MSEVEFSHEYWMRHALTLAKRAWDEREV PVGAVLVHNNRVIGEGWNRPIGRHDPTAH AEIMALRQGGLVMQNYRLIDATLYVTLEP CVMCAGAMIHSRIGRVVFGARDAKTGAA GSLMDVLHHPGMNHRVEITEGILADECAA LLSDFFRMRRQEIKAQKKAQSSTDGGSSG GSSGSETPGTSESATPESSGGSSGGSMSEV EFSHEYWMRHALTLAKRARDEREVPVGA VLVLNNRVIGEGWNRAIGLHDPTAHAEIM ALRQGGLVMQNYRLIDATLYVTFEPCVM CAGAMIHSRIGRVVFGVRNAKTGAAGSL MDVLHYPGMNHRVEITEGILADECAALLC TFFRMPRQVFNAQKKAQSSTDSGGSSGGS SGSETPGTSESATPESSGGSSGGSEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFMQPTVAYSVLVVAKVEK GKSKKLKSVKELLGITIMERSSFEKNPIDFL EAKGYKEVKKDLIIKLPKYSLFELENGRKR MLASAKFLQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIE QISEFSKRVILADANLDKVLSAYNKHRDK PIREQAENIIHLFTLTNLGAPRAFKYFDTTI ARKEYRSTKEVLDATLIHQSITGLYETRID LSQLGGDGGSGGSGGSGGSGGSGGSGGM DKKYSIGLAIGTNSVGWAVITDEYKVPSK KFKVLGNTDRHSIKKNLIGALLFDSGETAE ATRLKRTARRRYTRRKNRICYLQEIFSNEM AKVDDSFFHRLEESFLVEEDKKHERHPIFG NIVDEVAYHEKYPTIYHLRKKLVDSTDKA DLRLIYLALAHMIKFRGHFLIEGDLNPDNS DVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTE ELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWN

		FEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQV SGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQK GQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDEN DKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEGADKR TADGSEFESPKKKRKV
[1362]	ABE8.8-m 多肽序列	1428	MSEVEFSHEYWMRHALTLAKRARDEREV PVGAVLVLNNRVIGEGWNRAIGLHDPTAH AEIMALRQGGLVMQNYRLIDATLYVTFEP CVMCAGAMIHSRIGRVVFGVRNAKTGAA GSLMDVLHHPGMNHRVEITEGILADECAA LLCRFFRMPRRVFNAQKKAQSSTDSGGSS GGSSGSETPGTSESATPESSGGSSGGSDKK YSIGLAIGTNSVGWAVITDEYKVPSKKFKV LGNTDRHSIKKNLIGALLFDSGETAEATRL KRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIV DEVAYHEKYPTIYHLRKKLVDSTDKADLR LIYLALAHMIKFRGHFLIEGDLNPDNSDVD KLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRRLENLIAQLPGEKKNGLFGNLI ALSLGLTPNFKSNFDLAEDAKLQLSKDTY DDDLDNLLAQIGDQYADLFLAAKNLSDAI LLSDILRVNTEITKAPLSASMIKRYDEHHQ DLTLLKALVRQQLPEKYKEIFFDQSKNGY AGYIDGGASQEEFYKFIKPILEKMDGTEEL LVKLNREDLLRKQRTFDNGSIPHQIHLGEL HAILRRQEDFYPFLKDNREKIEKILTFRIPY YVGPLARGNSRFAWMTRKSEETITPWNFE

[1363]

		EVVDKGASAQSFIERMTNFDKNLPNEKVL PKHSLLYEYFTVYNELTKVKYVTEGMRKP AFLSGEQKKAIVDLLFKTNRKVTVKQLKE DYFKKIECFDSVEISGVEDRFNASLGTYHD LLKIIKDKDFLDNEENEDILEDIVLTLTLFE DREMIEERLKTYAHLFDDKVMKQLKRRR YTGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGRHKPENIVIEMARENQTTQKG QKNSRERMKRIEEGIKELGSQILKEHPVEN TQLQNEKLYLYYLQNGRDMYVDQELDIN RLSDYDVDHIVPQSFLKDDSIDNKVLTRSD KNRGKSDNVPSEEVVKKMKNYWRQLLN AKLITQRKFDNLTKAERGGLSELDKAGFIK RQLVETRQITKHVAQILDSRMNTKYDEND KLIREVKVITLKSKLVSDFRKDFQFYKVRE INNYHHAHDAYLNAVVGTALIKKYPKLES EFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFDSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIE QISEFSKRVILADANLDKVLSAYNKHRDK PIREQAENIIHLFTLTNLGAPAAFKYFDTTI DRKRYTSTKEVLDATLIHQSITGLYETRID LSQLGGDEGADKRTADGSEFESPKKKRKV
ABE8.8-d 多肽序列	1429	MSEVEFSHEYWMRHALTLAKRAWDEREV PVGAVLVHNNRVIGEGWNRPIGRHDPTAH AEIMALRQGGLVMQNYRLIDATLYVTLEP CVMCAGAMIHSRIGRVVFGARDAKTGAA GSLMDVLHHPGMNHRVEITEGILADECAA LLSDFFRMRRQEIKAQKKAQSSTDSGGSS GGSSGSETPGTSESATPESSGGSSGGSSEVE FSHEYWMRHALTLAKRARDEREVPVGAV LVLNNRVIGEGWNRAIGLHDPTAHAEIMA LRQGGLVMQNYRLIDATLYVTFEPCVMC AGAMIHSRIGRVVFGVRNAKTGAAGSLM DVLHHPGMNHRVEITEGILADECAALLCR

[1364]

		FFRMPRRVFNAQKKAQSSTDSGGSSGGSS GSETPGTSESATPESSGGSSGGSDKKYSIGL AIGTNSVGWAVITDEYKVPSKKFKVLGNT DRHSIKKNLIGALLFDSGETAEATRLKRTA RRRYTRRKNRICYLQEIFSNEMAKVDDSFF HRLEESFLVEEDKKHERHPIFGNIVDEVAY HEKYPTIYHLRKKLVDSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQL VQTYNQLFEENPINASGVDAKAILSARLSK SRRLENLIAQLPGEKKNGLFGNLIALSLGL TPNFKSNFDLAEDAKLQLSKDTYDDDLDN LLAQIGDQYADLFLAAKNLSDAILLSDILR VNTEITKAPLSASMIKRYDEHHQDLTLLK ALVRQQLPEKYKEIFFDQSKNGYAGYIDG GASQEEFYKFIKPILEKMDGTEELLVKLNR EDLLRKQRTFDNGSIPHQIHLGELHAILRR QEDFYPFLKDNREKIEKILTFRIPYYVGPLA RGNSRFAWMTRKSEETITPWNFEEVVDKG ASAQSFIERMTNFDKNLPNEKVLPKHSLLY EYFTVYNELTKVKYVTEGMRKPAFLSGEQ KKAIVDLLFKTNRKVTVKQLKEDYFKKIE CFDSVEISGVEDRFNASLGTYHDLLKIIKD KDFLDNEENEDILEDIVLTLTLFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGR LSRKLINGIRDKQSGKTILDFLKSDGFANR NFMQLIHDDSLTFKEDIQKAQVSGQGDSL HEHIANLAGSPAIKKGILQTVKVVDELVK VMGRHKPENIVIEMARENQTTQKGQKNSR ERMKRIEEGIKELGSQILKEHPVENTQLQN EKLYLYYLQNGRDMYVDQELDINRLSDY DVDHIVPQSFLKDDSIDNKVLTRSDKNRG KSDNVPSEEVVKKMKNYWRQLLNAKLIT QRKFDNLTKAERGGLSELDKAGFIKRQLV ETRQITKHVAQILDSRMNTKYDENDKLIRE VKVITLKSKLVSDFRKDFQFYKVREINNY HHAHDAYLNAVVGTALIKKYPKLESEFVY GDYKVYDVRKMIAKSEQEIGKATAKYFFY SNIMNFFKTEITLANGEIRKRPLIETNGETG EIVWDKGRDFATVRKVLSMPQVNIVKKTE VQTGGFSKESILPKRNSDKLIARKKDWDP KKYGGFDSPTVAYSVLVVAKVEKGKSKK LKSVKELLGITIMERSSFEKNPIDFLEAKGY KEVKKDLIIKLPKYSLFELENGRKRMLASA

[1365]

		GELQKGNELALPSKYVNFLYLASHYEKLK GSPEDNEQKQLFVEQHKHYLDEIIEQISEFS KRVILADANLDKVLSAYNKHRDKPIREQA ENIIHLFTLTNLGAPAAFKYFDTTIDRKRYT STKEVLDATLIHQSITGLYETRIDLSQLGGD EGADKRTADGSEFESPKKKRKV
ABE8.13-m 多肽序列	1430	MSEVEFSHEYWMRHALTLAKRARDEREV PVGAVLVLNNRVIGEGWNRAIGLHDPTAH AEIMALRQGGLVMQNYRLYDATLYVTFE PCVMCAGAMIHSRIGRVVFGVRNAKTGA AGSLMDVLHHPGMNHRVEITEGILADECA ALLCRFFRMPRRVFNAQKKAQSSTDSGGS SGGSSGSETPGTSESATPESSGGSSGGSDK KYSIGLAIGTNSVGWAVITDEYKVPSKKFK VLGNTDRHSIKKNLIGALLFDSGETAEATR LKRTARRRYTRRKNRICYLQEIFSNEMAK VDDSFFHRLEESFLVEEDKKHERHPIFGNI VDEVAYHEKYPTIYHLRKKLVDSTDKADL RLIYLALAHMIKFRGHFLIEGDLNPDNSDV DKLFIQLVQTYNQLFEENPINASGVDAKAI LSARLSKSRRLENLIAQLPGEKKNGLFGNL IALSLGLTPNFKSNFDLAEDAKLQLSKDTY DDDLDNLLAQIGDQYADLFLAAKNLSDAI LLSDILRVNTEITKAPLSASMIKRYDEHHQ DLTLLKALVRQQLPEKYKEIFFDQSKNGY AGYIDGGASQEEFYKFIKPILEKMDGTEEL LVKLNREDLLRKQRTFDNGSIPHQIHLGEL HAILRRQEDFYPFLKDNREKIEKILTFRIPY YVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVL PKHSLLYEYFTVYNELTKVKYVTEGMRKP AFLSGEQKKAIVDLLFKTNRKVTVKQLKE DYFKKIECFDSVEISGVEDRFNASLGTYHD LLKIIKDKDFLDNEENEDILEDIVLTLTLFE DREMIEERLKTYAHLFDDKVMKQLKRRR YTGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGRHKPENIVIEMARENQTTQKG QKNSRERMKRIEEGIKELGSQILKEHPVEN TQLQNEKLYLYYLQNGRDMYVDQELDIN RLSDYDVDHIVPQSFLKDDSIDNKVLTRSD KNRGKSDNVPSEEVVKKMKNYWRQLLN

		AKLITQRKFDNLTKAERGGLSELDKAGFIK RQLVETRQITKHVAQILDSRMNTKYDEND KLIREVKVITLKSKLVSDFRKDFQFYKVRE INNYHHAHDAYLNAVVGTALIKKYPKLES EFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFDSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIE QISEFSKRVILADANLDKVLSAYNKHRDK PIREQAENIIHLFTLTNLGAPAAFKYFDTTI DRKRYTSTKEVLDATLIHQSITGLYETRID LSQLGGDEGADKRTADGSEFESPKKKRKV
[1366]	ABE8.13-d 多肽序列	1431	MSEVEFSHEYWMRHALTLAKRAWDEREV PVGAVLVHNNRVIGEGWNRPIGRHDPTAH AEIMALRQGGLVMQNYRLIDATLYVTLEP CVMCAGAMIHSRIGRVVFGARDAKTGAA GSLMDVLHHPGMNHRVEITEGILADECAA LLSDFFRMRRQEIKAQKKAQSSTDSGGSS GGSSGSETPGTSESATPESSGGSSGGSSEVE FSHEYWMRHALTLAKRARDEREVPVGAV LVLNNRVIGEGWNRAIGLHDPTAHAEIMA LRQGGLVMQNYRLYDATLYVTFEPCVMC AGAMIHSRIGRVVFGVRNAKTGAAGSLM DVLHHPGMNHRVEITEGILADECAALLCR FFRMPRRVFNAQKKAQSSTDSGGSSGGSS GSETPGTSESATPESSGGSSGGSDKKYSIGL AIGTNSVGWAVITDEYKVPSKKFKVLGNT DRHSIKKNLIGALLFDSGETAEATRLKRTA RRRYTRRKNRICYLQEIFSNEMAKVDDSFF HRLEESFLVEEDKKHERHPIFGNIVDEVAY HEKYPTIYHLRKKLVDSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQL VQTYNQLFEENPINASGVDAKAILSARLSK SRRLENLIAQLPGEKKNGLFGNLIALSLGL TPNFKSNFDLAEDAKLQLSKDTYDDDLDN LLAQIGDQYADLFLAAKNLSDAILLSDILR VNTEITKAPLSASMIKRYDEHHQDLTLLK ALVRQQLPEKYKEIFFDQSKNGYAGYIDG

[1367]

		GASQEEFYKFIKPILEKMDGTEELLVKLNR EDLLRKQRTFDNGSIPHQIHLGELHAILRR QEDFYPFLKDNREKIEKILTFRIPYYVGPLA RGNSRFAWMTRKSEETITPWNFEEVVDKG ASAQSFIERMTNFDKNLPNEKVLPKHSLLY EYFTVYNELTKVKYVTEGMRKPAFLSGEQ KKAIVDLLFKTNRKVTVKQLKEDYFKKIE CFDSVEISGVEDRFNASLGTYHDLLKIIKD KDFLDNEENEDILEDIVLTLTLFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGR LSRKLINGIRDKQSGKTILDFLKSDGFANR NFMQLIHDDSLTFKEDIQKAQVSGQGDSL HEHIANLAGSPAIKKGILQTVKVVDELVK VMGRHKPENIVIEMARENQTTQKGQKNSR ERMKRIEEGIKELGSQILKEHPVENTQLQN EKLYLYYLQNGRDMYVDQELDINRLSDY DVDHIVPQSFLKDDSIDNKVLTRSDKNRG KSDNVPSEEVVKKMKNYWRQLLNAKLIT QRKFDNLTKAERGGLSELDKAGFIKRQLV ETRQITKHVAQILDSRMNTKYDENDKLIRE VKVITLKSKLVSDFRKDFQFYKVREINNY HHAHDAYLNAVVGTALIKKYPKLESEFVY GDYKVYDVRKMIAKSEQEIGKATAKYFFY SNIMNFFKTEITLANGEIRKRPLIETNGETG EIVWDKGRDFATVRKVLSMPQVNIVKKTE VQTGGFSKESILPKRNSDKLIARKKDWDP KKYGGFDSPTVAYSVLVVAKVEKGKSKK LKSVKELLGITIMERSSFEKNPIDFLEAKGY KEVKKDLIIKLPKYSLFELENGRKRMLASA GELQKGNELALPSKYVNFLYLASHYEKLK GSPEDNEQKQLFVEQHKHYLDEIIEQISEFS KRVILADANLDKVLSAYNKHRDKPIREQA ENIIHLFTLTNLGAPAAFKYFDTTIDRKRYT STKEVLDATLIHQSITGLYETRIDLSQLGGD EGADKRTADGSEFESPKKKRKV
	ABE8.17-m 多肽序列	1432	MSEVEFSHEYWMRHALTLAKRARDEREV PVGAVLVLNNRVIGEGWNRAIGLHDPTAH AEIMALRQGGLVMQNYRLIDATLYSTFEP CVMCAGAMIHSRIGRVVFGVRNAKTGAA GSLMDVLHYPGMNHRVEITEGILADECAA LLCYFFRMPRRVFNAQKKAQSSTDSGGSS GGSSGSETPGTSESATPESSGGSSGGSDKK YSIGLAIGTNSVGWAVITDEYKVPSKKFKV

[1368]

		LGNTDRHSIKKNLIGALLFDSGETAEATRL KRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIV DEVAYHEKYPTIYHLRKKLVDSTDKADLR LIYLALAHMIKFRGHFLIEGDLNPDNSDVD KLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRRLENLIAQLPGEKKNGLFGNLI ALSLGLTPNFKSNFDLAEDAKLQLSKDTY DDDLDNLLAQIGDQYADLFLAAKNLSDAI LLSDILRVNTEITKAPLSASMIKRYDEHHQ DLTLLKALVRQQLPEKYKEIFFDQSKNGY AGYIDGGASQEEFYKFIKPILEKMDGTEEL LVKLNREDLLRKQRTFDNGSIPHQIHLGEL HAILRRQEDFYPFLKDNREKIEKILTFRIPY YVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVL PKHSLLYEYFTVYNELTKVKYVTEGMRKP AFLSGEQKKAIVDLLFKTNRKVTVKQLKE DYFKKIECFDSVEISGVEDRFNASLGTYHD LLKIIKDKDFLDNEENEDILEDIVLTLTLFE DREMIEERLKTYAHLFDDKVMKQLKRRR YTGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGRHKPENIVIEMARENQTTQKG QKNSRERMKRIEEGIKELGSQILKEHPVEN TQLQNEKLYLYYLQNGRDMYVDQELDIN RLSDYDVDHIVPQSFLKDDSIDNKVLTRSD KNRGKSDNVPSEEVVKKMKNYWRQLLN AKLITQRKFDNLTKAERGGLSELDKAGFIK RQLVETRQITKHVAQILDSRMNTKYDEND KLIREVKVITLKSKLVSDFRKDFQFYKVRE INNYHHAHDAYLNAVVGTALIKKYPKLES EFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFDSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIE QISEFSKRVILADANLDKVLSAYNKHRDK

		PIREQAENIIHLFTLTNLGAPAAFKYFDTTI DRKRYTSTKEVLDATLIHQSITGLYETRID LSQLGGDEGADKRTADGSEFESPKKKRKV
ABE8.17-d 多肽序列	1433	MSEVEFSHEYWMRHALTLAKRAWDEREV PVGAVLVHNNRVIGEGWNRPIGRHDPTAH AEIMALRQGGLVMQNYRLIDATLYVTLEP CVMCAGAMIHSRIGRVVFGARDAKTGAA GSLMDVLHHPGMNHRVEITEGILADECAA LLSDFFRMRRQEIKAQKKAQSSTDSGGSS GGSSGSETPGTSESATPESSGGSSGGSSEVE FSHEYWMRHALTLAKRARDEREVPVGAV LVLNNRVIGEGWNRAIGLHDPTAHAEIMA LRQGGLVMQNYRLIDATLYSTFEPCVMCA GAMIHSRIGRVVFGVRNAKTGAAGSLMD VLHYPGMNHRVEITEGILADECAALLCYF FRMPRRVFNAQKKAQSSTDSGGSSGGSSG SETPGTSESATPESSGGSSGGSDKKYSIGLA IGTNSVGWAVITDEYKVPSKKFKVLGNTD RHSIKKNLIGALLFDSGETAEATRLKRTAR RRYTRRKNRICYLQEIFSNEMAKVDDSFFH RLEESFLVEEDKKHERHPIFGNIVDEVAYH EKYPTIYHLRKKLVDSTDKADLRLIYLALA HMIKFRGHFLIEGDLNPDNSDVDKLFIQLV QTYNQLFEENPINASGVDAKAILSARLSKS RRLENLIAQLPGEKKNGLFGNLIALSLGLT PNFKSNFDLAEDAKLQLSKDTYDDDLDNL LAQIGDQYADLFLAAKNLSDAILLSDILRV NTEITKAPLSASMIKRYDEHHQDLTLLKAL VRQQLPEKYKEIFFDQSKNGYAGYIDGGA SQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQE DFYPFLKDNREKIEKILTFRIPYYVGPLARG NSRFAWMTRKSEETITPWNFEEVVDKGAS AQSFIERMTNFDKNLPNEKVLPKHSLLYE YFTVYNELTKVKYVTEGMRKPAFLSGEQ KKAIVDLLFKTNRKVTVKQLKEDYFKKIE CFDSVEISGVEDRFNASLGTYHDLLKIIKD KDFLDNEENEDILEDIVLTLTLFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGR LSRKLINGIRDKQSGKTILDFLKSDGFANR NFMQLIHDDSLTFKEDIQKAQVSGQGDSL HEHIANLAGSPAIKKGILQTVKVVDELVK VMGRHKPENIVIEMARENQTTQKGQKNSR

[1369]

			ERMKRIEEGIKELGSQILKEHPVENTQLQN EKLYLYYLQNGRDMYVDQELDINRLSDY DVDHIVPQSFLKDDSIDNKVLTRSDKNRG KSDNVPSEEVVKKMKNYWRQLLNAKLIT QRKFDNLTKAERGGLSELDKAGFIKRQLV ETRQITKHVAQILDSRMNTKYDENDKLIRE VKVITLKSKLVSDFRKDFQFYKVREINNY HHAHDAYLNAVVGTALIKKYPKLESEFVY GDYKVYDVRKMIAKSEQEIGKATAKYFFY SNIMNFFKTEITLANGEIRKRPLIETNGETG EIVWDKGRDFATVRKVLSMPQVNIVKKTE VQTGGFSKESILPKRNSDKLIARKKDWDP KKYGGFDSPTVAYSVLVVAKVEKGKSKK LKSVKELLGITIMERSSFEKNPIDFLEAKGY KEVKKDLIIKLPKYSLFELENGRKRMLASA GELQKGNELALPSKYVNFLYLASHYEKLK GSPEDNEQKQLFVEQHKHYLDEIIEQISEFS KRVILADANLDKVLSAYNKHRDKPIREQA ENIIHLFTLTNLGAPAAFKYFDTTIDRKRYT STKEVLDATLIHQSITGLYETRIDLSQLGGD EGADKRTADGSEFESPKKKRKV
[1370]	ABE8.20-m 多肽序列	1434	MSEVEFSHEYWMRHALTLAKRARDEREV PVGAVLVLNNRVIGEGWNRAIGLHDPTAH AEIMALRQGGLVMQNYRLYDATLYSTFEP CVMCAGAMIHSRIGRVVFGVRNAKTGAA GSLMDVLHHPGMNHRVEITEGILADECAA LLCRFFRMPRRVFNAQKKAQSSTDSGGSS GGSSGSETPGTSESATPESSGGSSGGSDKK YSIGLAIGTNSVGWAVITDEYKVPSKKFKV LGNTDRHSIKKNLIGALLFDSGETAEATRL KRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIV DEVAYHEKYPTIYHLRKKLVDSTDKADLR LIYLALAHMIKFRGHFLIEGDLNPDNSDVD KLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRRLENLIAQLPGEKKNGLFGNLI ALSLGLTPNFKSNFDLAEDAKLQLSKDTY DDDLDNLLAQIGDQYADLFLAAKNLSDAI LLSDILRVNTEITKAPLSASMIKRYDEHHQ DLTLLKALVRQQLPEKYKEIFFDQSKNGY AGYIDGGASQEEFYKFIKPILEKMDGTEEL LVKLNREDLLRKQRTFDNGSIPHQIHLGEL HAILRRQEDFYPFLKDNREKIEKILTFRIPY

[1371]

		YVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVL PKHSLLYEYFTVYNELTKVKYVTEGMRKP AFLSGEQKKAIVDLLFKTNRKVTVKQLKE DYFKKIECFDSVEISGVEDRFNASLGTYHD LLKIIKDKDFLDNEENEDILEDIVLTLTLFE DREMIEERLKTYAHLFDDKVMKQLKRRR YTGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGRHKPENIVIEMARENQTTQKG QKNSRERMKRIEEGIKELGSQILKEHPVEN TQLQNEKLYLYYLQNGRDMYVDQELDIN RLSDYDVDHIVPQSFLKDDSIDNKVLTRSD KNRGKSDNVPSEEVVKKMKNYWRQLLN AKLITQRKFDNLTKAERGGLSELDKAGFIK RQLVETRQITKHVAQILDSRMNTKYDEND KLIREVKVITLKSKLVSDFRKDFQFYKVRE INNYHHAHDAYLNAVVGTALIKKYPKLES EFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFDSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIE QISEFSKRVILADANLDKVLSAYNKHRDK PIREQAENIIHLFTLTNLGAPAAFKYFDTTI DRKRYTSTKEVLDATLIHQSITGLYETRID LSQLGGDEGADKRTADGSEFESPKKKRKV
	ABE8.20-d 多肽序列	1435	MSEVEFSHEYWMRHALTLAKRAWDEREV PVGAVLVHNNRVIGEGWNRPIGRHDPTAH AEIMALRQGGLVMQNYRLIDATLYVTLEP CVMCAGAMIHSRIGRVVFGARDAKTGAA GSLMDVLHHPGMNHRVEITEGILADECAA LLSDFFRMRRQEIKAQKKAQSSTDSGGSS GGSSGSETPGTSESATPESSGGSSGGSSEVE FSHEYWMRHALTLAKRARDEREVPVGAV LVLNNRVIGEGWNRAIGLHDPTAHAEIMA LRQGGLVMQNYRLYDATLYSTFEPCVMC AGAMIHSRIGRVVFGVRNAKTGAAGSLM

[1372]

		DVLHHPGMNHRVEITEGILADECAALLCR FFRMPRRVFNAQKKAQSSTDSGGSSGGSS GSETPGTSESATPESSGGSSGGSDKKYSIGL AIGTNSVGWAVITDEYKVPSKKFKVLGNT DRHSIKKNLIGALLFDSGETAEATRLKRTA RRRYTRRKNRICYLQEIFSNEMAKVDDSFF HRLEESFLVEEDKKHERHPIFGNIVDEVAY HEKYPTIYHLRKKLVDSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQL VQTYNQLFEENPINASGVDAKAILSARLSK SRRLENLIAQLPGEKKNGLFGNLIALSLGL TPNFKSNFDLAEDAKLQLSKDTYDDDLDN LLAQIGDQYADLFLAAKNLSDAILLSDILR VNTEITKAPLSASMIKRYDEHHQDLTLLK ALVRQQLPEKYKEIFFDQSKNGYAGYIDG GASQEEFYKFIKPILEKMDGTEELLVKLNR EDLLRKQRTFDNGSIPHQIHLGELHAILRR QEDFYPFLKDNREKIEKILTFRIPYYVGPLA RGNSRFAWMTRKSEETITPWNFEEVVDKG ASAQSFIERMTNFDKNLPNEKVLPKHSLLY EYFTVYNELTKVKYVTEGMRKPAFLSGEQ KKAIVDLLFKTNRKVTVKQLKEDYFKKIE CFDSVEISGVEDRFNASLGTYHDLLKIIKD KDFLDNEENEDILEDIVLTLTLFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGR LSRKLINGIRDKQSGKTILDFLKSDGFANR NFMQLIHDDSLTFKEDIQKAQVSGQGDSL HEHIANLAGSPAIKKGILQTVKVVDELVK VMGRHKPENIVIEMARENQTTQKGQKNSR ERMKRIEEGIKELGSQILKEHPVENTQLQN EKLYLYYLQNGRDMYVDQELDINRLSDY DVDHIVPQSFLKDDSIDNKVLTRSDKNRG KSDNVPSEEVVKKMKNYWRQLLNAKLIT QRKFDNLTKAERGGLSELDKAGFIKRQLV ETRQITKHVAQILDSRMNTKYDENDKLIRE VKVITLKSKLVSDFRKDFQFYKVREINNY HHAHDAYLNAVVGTALIKKYPKLESEFVY GDYKVYDVRKMIAKSEQEIGKATAKYFFY SNIMNFFKTEITLANGEIRKRPLIETNGETG EIVWDKGRDFATVRKVLSMPQVNIVKKTE VQTGGFSKESILPKRNSDKLIARKKDWDP KKYGGFDSPTVAYSVLVVAKVEKGKSKK LKSVKELLGITIMERSSFEKNPIDFLEAKGY

[1373]

		KEVKKDLIIKLPKYSLFELENGRKRMLASA GELQKGNELALPSKYVNFLYLASHYEKLK GSPEDNEQKQLFVEQHKHYLDEIIEQISEFS KRVILADANLDKVLSAYNKHRDKPIREQA ENIIHLFTLTNLGAPAAFKYFDTTIDRKRYT STKEVLDATLIHQSITGLYETRIDLSQLGGD EGADKRTADGSEFESPKKKRKV
01. monoABE8.1_bpNLS + Y147T 多肽序列	1436	MSEVEFSHEYWMRHALTLAKRARDEREV PVGAVLVLNNRVIGEGWNRAIGLHDPTAH AEIMALRQGGLVMQNYRLIDATLYVTFEP CVMCAGAMIHSRIGRVVFGVRNAKTGAA GSLMDVLHYPGMNHRVEITEGILADECAA LLCTFFRMPRQVFNAQKKAQSSTDSGGSS GGSSGSETPGTSESATPESSGGSSGGSDKK YSIGLAIGTNSVGWAVITDEYKVPSKKFKV LGNTDRHSIKKNLIGALLFDSGETAEATRL KRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIV DEVAYHEKYPTIYHLRKKLVDSTDKADLR LIYLALAHMIKFRGHFLIEGDLNPDNSDVD KLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRRLENLIAQLPGEKKNGLFGNLI ALSLGLTPNFKSNFDLAEDAKLQLSKDTY DDDLDNLLAQIGDQYADLFLAAKNLSDAI LLSDILRVNTEITKAPLSASMIKRYDEHHQ DLTLLKALVRQQLPEKYKEIFFDQSKNGY AGYIDGGASQEEFYKFIKPILEKMDGTEEL LVKLNREDLLRKQRTFDNGSIPHQIHLGEL HAILRRQEDFYPFLKDNREKIEKILTFRIPY YVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVL PKHSLLYEYFTVYNELTKVKYVTEGMRKP AFLSGEQKKAIVDLLFKTNRKVTVKQLKE DYFKKIECFDSVEISGVEDRFNASLGTYHD LLKIIKDKDFLDNEENEDILEDIVLTLTLFE DREMIEERLKTYAHLFDDKVMKQLKRRR YTGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGRHKPENIVIEMARENQTTQKG QKNSRERMKRIEEGIKELGSQILKEHPVEN TQLQNEKLYLYYLQNGRDMYVDQELDIN RLSDYDVDHIVPQSFLKDDSIDNKVLTRSD

			KNRGKSDNVPSEEVVKKMKNYWRQLLN AKLITQRKFDNLTKAERGGLSELDKAGFIK RQLVETRQITKHVAQILDSRMNTKYDEND KLIREVKVITLKSKLVSDFRKDFQFYKVRE INNYHHAHDAYLNAVVGTALIKKYPKLES EFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFVSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASARELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIE QISEFSKRVILADANLDKVLSAYNKHRDK PIREQAENIIHLFTLTNLGAPAAFKYFDTTI DRKQYRSTKEVLDATLIHQSITGLYETRID LSQLGGDEGADKRTADGSEFESPKKKRKV
[1374]	02. monoABE8.1_bpNLS + Y147R 多肽序列	1437	MSEVEFSHEYWMRHALTLAKRARDEREV PVGAVLVLNNRVIGEGWNRAIGLHDPTAH AEIMALRQGGLVMQNYRLIDATLYVTFEP CVMCAGAMIHSRIGRVVFGVRNAKTGAA GSLMDVLHYPGMNHRVEITEGILADECAA LLCRFFRMPRQVFNAQKKAQSSTDSGGSS GGSSGSETPGTSESATPESSGGSSGGSDKK YSIGLAIGTNSVGWAVITDEYKVPSKKFKV LGNTDRHSIKKNLIGALLFDSGETAEATRL KRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIV DEVAYHEKYPTIYHLRKKLVDSTDKADLR LIYLALAHMIKFRGHFLIEGDLNPDNSDVD KLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRRLENLIAQLPGEKKNGLFGNLI ALSLGLTPNFKSNFDLAEDAKLQLSKDTY DDDLDNLLAQIGDQYADLFLAAKNLSDAI LLSDILRVNTEITKAPLSASMIKRYDEHHQ DLTLLKALVRQQLPEKYKEIFFDQSKNGY AGYIDGGASQEEFYKFIKPILEKMDGTEEL LVKLNREDLLRKQRTFDNGSIPHQIHLGEL HAILRRQEDFYPFLKDNREKIEKILTFRIPY YVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVL PKHSLLYEYFTVYNELTKVKYVTEGMRKP

[1375]

		AFLSGEQKKAIVDLLFKTNRKVTVKQLKE DYFKKIECFDSVEISGVEDRFNASLGTYHD LLKIIKDKDFLDNEENEDILEDIVLTLTLFE DREMIEERLKTYAHLFDDKVMKQLKRRR YTGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGRHKPENIVIEMARENQTTQKG QKNSRERMKRIEEGIKELGSQILKEHPVEN TQLQNEKLYLYYLQNGRDMYVDQELDIN RLSDYDVDHIVPQSFLKDDSIDNKVLTRSD KNRGKSDNVPSEEVVKKMKNYWRQLLN AKLITQRKFDNLTKAERGGLSELDKAGFIK RQLVETRQITKHVAQILDSRMNTKYDEND KLIREVKVITLKSKLVSDFRKDFQFYKVRE INNYHHAHDAYLNAVVGTALIKKYPKLES EFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFVSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASARELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIE QISEFSKRVILADANLDKVLSAYNKHRDK PIREQAENIIHLFTLTNLGAPAAFKYFDTTI DRKQYRSTKEVLDATLIHQSITGLYETRID LSQLGGDEGADKRTADGSEFESPKKKRKV
03. monoABE8.1_bpNLS + Q154S 多肽序列	1438	MSEVEFSHEYWMRHALTLAKRARDEREV PVGAVLVLNNRVIGEGWNRAIGLHDPTAH AEIMALRQGGLVMQNYRLIDATLYVTFEP CVMCAGAMIHSRIGRVVFGVRNAKTGAA GSLMDVLHYPGMNHRVEITEGILADECAA LLCYFFRMPRSVFNAQKKAQSSTDSGGSS GGSSGSETPGTSESATPESSGGSSGGSDKK YSIGLAIGTNSVGWAVITDEYKVPSKKFKV LGNTDRHSIKKNLIGALLFDSGETAEATRL KRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIV DEVAYHEKYPTIYHLRKKLVDSTDKADLR LIYLALAHMIKFRGHFLIEGDLNPDNSDVD KLFIQLVQTYNQLFEENPINASGVDAKAIL

[1376]

		SARLSKSRRLENLIAQLPGEKKNGLFGNLI ALSLGLTPNFKSNFDLAEDAKLQLSKDTY DDDLDNLLAQIGDQYADLFLAAKNLSDAI LLSDILRVNTEITKAPLSASMIKRYDEHHQ DLTLLKALVRQQLPEKYKEIFFDQSKNGY AGYIDGGASQEEFYKFIKPILEKMDGTEEL LVKLNREDLLRKQRTFDNGSIPHQIHLGEL HAILRRQEDFYPFLKDNREKIEKILTFRIPY YVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVL PKHSLLYEYFTVYNELTKVKYVTEGMRKP AFLSGEQKKAIVDLLFKTNRKVTVKQLKE DYFKKIECFDSVEISGVEDRFNASLGTYHD LLKIIKDKDFLDNEENEDILEDIVLTLTLFE DREMIEERLKTYAHLFDDKVMKQLKRRR YTGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGRHKPENIVIEMARENQTTQKG QKNSRERMKRIEEGIKELGSQILKEHPVEN TQLQNEKLYLYYLQNGRDMYVDQELDIN RLSDYDVDHIVPQSFLKDDSIDNKVLTRSD KNRGKSDNVPSEEVVKKMKNYWRQLLN AKLITQRKFDNLTKAERGGLSELDKAGFIK RQLVETRQITKHVAQILDSRMNTKYDEND KLIREVKVITLKSKLVSDFRKDFQFYKVRE INNYHHAHDAYLNAVVGTALIKKYPKLES EFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFVSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASARELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIE QISEFSKRVILADANLDKVLSAYNKHRDK PIREQAENIIHLFTLTNLGAPAAFKYFDTTI DRKQYRSTKEVLDATLIHQSITGLYETRID LSQLGGDEGADKRTADGSEFESPKKKRKV
04. monoAB E8.1_bpN	1439	MSEVEFSHEYWMRHALTLAKRARDEREV PVGAVLVLNNRVIGEGWNRAIGLHDPTAH AEIMALRQGGLVMQNYRLIDATLYVTFEP

[1377]

LS + Y123H 多肽序列		CVMCAGAMIHSRIGRVVFGVRNAKTGAA GSLMDVLHHPGMNHRVEITEGILADECAA LLCYFFRMPRQVFNAQKKAQSSTDSGGSS GGSSGSETPGTSESATPESSGGSSGGSDKK YSIGLAIGTNSVGWAVITDEYKVPSKKFKV LGNTDRHSIKKNLIGALLFDSGETAEATRL KRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIV DEVAYHEKYPTIYHLRKKLVDSTDKADLR LIYLALAHMIKFRGHFLIEGDLNPDNSDVD KLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRRLENLIAQLPGEKKNGLFGNLI ALSLGLTPNFKSNFDLAEDAKLQLSKDTY DDDLDNLLAQIGDQYADLFLAAKNLSDAI LLSDILRVNTEITKAPLSASMIKRYDEHHQ DLTLLKALVRQQLPEKYKEIFFDQSKNGY AGYIDGGASQEEFYKFIKPILEKMDGTEEL LVKLNREDLLRKQRTFDNGSIPHQIHLGEL HAILRRQEDFYPFLKDNREKIEKILTFRIPY YVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVL PKHSLLYEYFTVYNELTKVKYVTEGMRKP AFLSGEQKKAIVDLLFKTNRKVTVKQLKE DYFKKIECFDSVEISGVEDRFNASLGTYHD LLKIIKDKDFLDNEENEDILEDIVLTLTLFE DREMIEERLKTYAHLFDDKVMKQLKRRR YTGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGRHKPENIVIEMARENQTTQKG QKNSRERMKRIEEGIKELGSQILKEHPVEN TQLQNEKLYLYYLQNGRDMYVDQELDIN RLSDYDVDHIVPQSFLKDDSIDNKVLTRSD KNRGKSDNVPSEEVVKKMKNYWRQLLN AKLITQRKFDNLTKAERGGLSELDKAGFIK RQLVETRQITKHVAQILDSRMNTKYDEND KLIREVKVITLKSKLVSDFRKDFQFYKVRE INNYHHAHDAYLNAVVGTALIKKYPKLES EFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFVSPTVAYSVLVVAKVEKG

[1378]

		KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASARELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIE QISEFSKRVILADANLDKVLSAYNKHRDK PIREQAENIIHLFTLTNLGAPAAFKYFDTTI DRKQYRSTKEVLDATLIHQSITGLYETRID LSQLGGDEGADKRTADGSEFESPKKKRKV
05. monoABE8.1_bpNLS + V82S 多肽序列	1440	MSEVEFSHEYWMRHALTLAKRARDEREV PVGAVLVLNNRVIGEGWNRAIGLHDPTAH AEIMALRQGGLVMQNYRLIDATLYSTFEP CVMCAGAMIHSRIGRVVFGVRNAKTGAA GSLMDVLHYPGMNHRVEITEGILADECAA LLCYFFRMPRQVFNAQKKAQSSTDSGGSS GGSSGSETPGTSESATPESSGGSSGGSDKK YSIGLAIGTNSVGWAVITDEYKVPSKKFKV LGNTDRHSIKKNLIGALLFDSGETAEATRL KRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIV DEVAYHEKYPTIYHLRKKLVDSTDKADLR LIYLALAHMIKFRGHFLIEGDLNPDNSDVD KLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRRLENLIAQLPGEKKNGLFGNLI ALSLGLTPNFKSNFDLAEDAKLQLSKDTY DDDLDNLLAQIGDQYADLFLAAKNLSDAI LLSDILRVNTEITKAPLSASMIKRYDEHHQ DLTLLKALVRQQLPEKYKEIFFDQSKNGY AGYIDGGASQEEFYKFIKPILEKMDGTEEL LVKLNREDLLRKQRTFDNGSIPHQIHLGEL HAILRRQEDFYPFLKDNREKIEKILTFRIPY YVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVL PKHSLLYEYFTVYNELTKVKYVTEGMRKP AFLSGEQKKAIVDLLFKTNRKVTVKQLKE DYFKKIECFDSVEISGVEDRFNASLGTYHD LLKIIKDKDFLDNEENEDILEDIVLTLTLFE DREMIEERLKTYAHLFDDKVMKQLKRRR YTGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGRHKPENIVIEMARENQTTQKG QKNSRERMKRIEEGIKELGSQILKEHPVEN TQLQNEKLYLYYLQNGRDMYVDQELDIN

		RLSDYDVDHIVPQSFLKDDSIDNKVLTRSD KNRGKSDNVPSEEVVKKMKNYWRQLLN AKLITQRKFDNLTKAERGGLSELDKAGFIK RQLVETRQITKHVAQILDSRMNTKYDEND KLIREVKVITLKSKLVSDFRKDFQFYKVRE INNYHHAHDAYLNAVVGTALIKKYPKLES EFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFVSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASARELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIE QISEFSKRVILADANLDKVLSAYNKHRDK PIREQAENIIHLFTLTNLGAPAAFKYFDTTI DRKQYRSTKEVLDATLIHQSITGLYETRID LSQLGGDEGADKRTADGSEFESPKKKRKV
[1379]	06. monoABE8.1_bpNLS + T166R 多肽序列	1441	MSEVEFSHEYWMRHALTLAKRARDEREV PVGAVLVLNNRVIGEGWNRAIGLHDPTAH AEIMALRQGGLVMQNYRLIDATLYVTFEP CVMCAGAMIHSRIGRVVFGVRNAKTGAA GSLMDVLHYPGMNHRVEITEGILADECAA LLCYFFRMPRQVFNAQKKAQSSRDSGGSS GGSSGSETPGTSESATPESSGGSSGGSDKK YSIGLAIGTNSVGWAVITDEYKVPSKKFKV LGNTDRHSIKKNLIGALLFDSGETAEATRL KRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIV DEVAYHEKYPTIYHLRKKLVDSTDKADLR LIYLALAHMIKFRGHFLIEGDLNPDNSDVD KLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRRLENLIAQLPGEKKNGLFGNLI ALSLGLTPNFKSNFDLAEDAKLQLSKDTY DDDLDNLLAQIGDQYADLFLAAKNLSDAI LLSDILRVNTEITKAPLSASMIKRYDEHHQ DLTLLKALVRQQLPEKYKEIFFDQSKNGY AGYIDGGASQEEFYKFIKPILEKMDGTEEL LVKLNREDLLRKQRTFDNGSIPHQIHLGEL HAILRRQEDFYPFLKDNREKIEKILTFRIPY YVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVL

[1380]

		PKHSLLYEYFTVYNELTKVKYVTEGMRKP AFLSGEQKKAIVDLLFKTNRKVTVKQLKE DYFKKIECFDSVEISGVEDRFNASLGTYHD LLKIIKDKDFLDNEENEDILEDIVLTLTLFE DREMIEERLKTYAHLFDDKVMKQLKRRR YTGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGRHKPENIVIEMARENQTTQKG QKNSRERMKRIEEGIKELGSQILKEHPVEN TQLQNEKLYLYYLQNGRDMYVDQELDIN RLSDYDVDHIVPQSFLKDDSIDNKVLTRSD KNRGKSDNVPSEEVVKKMKNYWRQLLN AKLITQRKFDNLTKAERGGLSELDKAGFIK RQLVETRQITKHVAQILDSRMNTKYDEND KLIREVKVITLKSKLVSDFRKDFQFYKVRE INNYHHAHDAYLNAVVGTALIKKYPKLES EFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFVSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASARELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIE QISEFSKRVILADANLDKVLSAYNKHRDK PIREQAENIIHLFTLTNLGAPAAFKYFDTTI DRKQYRSTKEVLDATLIHQSITGLYETRID LSQLGGDEGADKRTADGSEFESPKKKRKV
	07. monoABE8.1_bpNLS + Q154R 多肽序列	1442	MSEVEFSHEYWMRHALTLAKRARDEREV PVGAVLVLNNRVIGEGWNRAIGLHDPTAH AEIMALRQGGLVMQNYRLIDATLYVTFEP CVMCAGAMIHSRIGRVVFGVRNAKTGAA GSLMDVLHYPGMNHRVEITEGILADECAA LLCYFFRMPRRVFNAQKKAQSSTDSGGSS GGSSGSETPGTSESATPESSGGSSGGSDKK YSIGLAIGTNSVGWAVITDEYKVPSKKFKV LGNTDRHSIKKNLIGALLFDSGETAEATRL KRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIV DEVAYHEKYPTIYHLRKKLVDSTDKADLR LIYLALAHMIKFRGHFLIEGDLNPDNSDVD

[1381]

		KLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRRLENLIAQLPGEKKNGLFGNLI ALSLGLTPNFKSNFDLAEDAKLQLSKDTY DDDLDNLLAQIGDQYADLFLAAKNLSDAI LLSDILRVNTEITKAPLSASMIKRYDEHHQ DLTLLKALVRQQLPEKYKEIFFDQSKNGY AGYIDGGASQEEFYKFIKPILEKMDGTEEL LVKLNREDLLRKQRTFDNGSIPHQIHLGEL HAILRRQEDFYPFLKDNREKIEKILTFRIPY YVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVL PKHSLLYEYFTVYNELTKVKYVTEGMRKP AFLSGEQKKAIVDLLFKTNRKVTVKQLKE DYFKKIECFDSVEISGVEDRFNASLGTYHD LLKIIKDKDFLDNEENEDILEDIVLTLTLFE DREMIEERLKTYAHLFDDKVMKQLKRRR YTGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGRHKPENIVIEMARENQTTQKG QKNSRERMKRIEEGIKELGSQILKEHPVEN TQLQNEKLYLYYLQNGRDMYVDQELDIN RLSDYDVDHIVPQSFLKDDSIDNKVLTRSD KNRGKSDNVPSEEVVKKMKNYWRQLLN AKLITQRKFDNLTKAERGGLSELDKAGFIK RQLVETRQITKHVAQILDSRMNTKYDEND KLIREVKVITLKSKLVSDFRKDFQFYKVRE INNYHHAHDAYLNAVVGTALIKKYPKLES EFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFVSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASARELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIE QISEFSKRVILADANLDKVLSAYNKHRDK PIREQAENIIHLFTLTNLGAPAAFKYFDTTI DRKQYRSTKEVLDATLIHQSITGLYETRID LSQLGGDEGADKRTADGSEFESPKKKRKV
08. monoAB	1443	MSEVEFSHEYWMRHALTLAKRARDEREV PVGAVLVLNNRVIGEGWNRAIGLHDPTAH

[1382]

E8.1_bpNLS + Y147R_Q154R_Y123H 多肽序列		AEIMALRQGGLVMQNYRLIDATLYVTFEP CVMCAGAMIHSRIGRVVFGVRNAKTGAA GSLMDVLHHPGMNHRVEITEGILADECAA LLCRFFRMPRRVFNAQKKAQSSTDSGGSS GGSSGSETPGTSESATPESSGGSSGGSDKK YSIGLAIGTNSVGWAVITDEYKVPSKKFKV LGNTDRHSIKKNLIGALLFDSGETAEATRL KRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIV DEVAYHEKYPTIYHLRKKLVDSTDKADLR LIYLALAHMIKFRGHFLIEGDLNPDNSDVD KLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRRLENLIAQLPGEKKNGLFGNLI ALSLGLTPNFKSNFDLAEDAKLQLSKDTY DDDLDNLLAQIGDQYADLFLAAKNLSDAI LLSDILRVNTEITKAPLSASMIKRYDEHHQ DLTLLKALVRQQLPEKYKEIFFDQSKNGY AGYIDGGASQEEFYKFIKPILEKMDGTEEL LVKLNREDLLRKQRTFDNGSIPHQIHLGEL HAILRRQEDFYPFLKDNREKIEKILTFRIPY YVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVL PKHSLLYEYFTVYNELTKVKYVTEGMRKP AFLSGEQKKAIVDLLFKTNRKVTVKQLKE DYFKKIECFDSVEISGVEDRFNASLGTYHD LLKIIKDKDFLDNEENEDILEDIVLTLTLFE DREMIEERLKTYAHLFDDKVMKQLKRRR YTGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGRHKPENIVIEMARENQTTQKG QKNSRERMKRIEEGIKELGSQILKEHPVEN TQLQNEKLYLYYLQNGRDMYVDQELDIN RLSDYDVDHIVPQSFLKDDSIDNKVLTRSD KNRGKSDNVPSEEVVKKMKNYWRQLLN AKLITQRKFDNLTKAERGGLSELDKAGFIK RQLVETRQITKHVAQILDSRMNTKYDEND KLIREVKVITLKSKLVSDFRKDFQFYKVRE INNYHHAHDAYLNAVVGTALIKKYPKLES EFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK

[1383]

		DWDPKKYGGFVSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASARELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIE QISEFSKRVILADANLDKVLSAYNKHRDK PIREQAENIIHLFTLTNLGAPAAFKYFDTTI DRKQYRSTKEVLDATLIHQSITGLYETRID LSQLGGDEGADKRTADGSEFESPKKKRKV
09. monoABE8.1_bpNLS + Y147R_Q154R_I76Y 多肽序列	1444	MSEVEFSHEYWMRHALTLAKRARDEREV PVGAVLVLNNRVIGEGWNRAIGLHDPTAH AEIMALRQGGLVMQNYRLYDATLYVTFE PCVMCAGAMIHSRIGRVVFGVRNAKTGA AGSLMDVLHYPGMNHRVEITEGILADECA ALLCRFFRMPRRVFNAQKKAQSSTDSGGS SGGSSGSETPGTSESATPESSGGSSGGSDK KYSIGLAIGTNSVGWAVITDEYKVPSKKFK VLGNTDRHSIKKNLIGALLFDSGETAEATR LKRTARRRYTRRKNRICYLQEIFSNEMAK VDDSFFHRLEESFLVEEDKKHERHPIFGNI VDEVAYHEKYPTIYHLRKKLVDSTDKADL RLIYLALAHMIKFRGHFLIEGDLNPDNSDV DKLFIQLVQTYNQLFEENPINASGVDAKAI LSARLSKSRRLENLIAQLPGEKKNGLFGNL IALSLGLTPNFKSNFDLAEDAKLQLSKDTY DDDLDNLLAQIGDQYADLFLAAKNLSDAI LLSDILRVNTEITKAPLSASMIKRYDEHHQ DLTLLKALVRQQLPEKYKEIFFDQSKNGY AGYIDGGASQEEFYKFIKPILEKMDGTEEL LVKLNREDLLRKQRTFDNGSIPHQIHLGEL HAILRRQEDFYPFLKDNREKIEKILTFRIPY YVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVL PKHSLLYEYFTVYNELTKVKYVTEGMRKP AFLSGEQKKAIVDLLFKTNRKVTVKQLKE DYFKKIECFDSVEISGVEDRFNASLGTYHD LLKIIKDKDFLDNEENEDILEDIVLTLTLFE DREMIEERLKTYAHLFDDKVMKQLKRRR YTGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGRHKPENIVIEMARENQTTQKG QKNSRERMKRIEEGIKELGSQILKEHPVEN

			TQLQNEKLYLYYLQNGRDMYVDQELDIN RLSDYDVDHIVPQSFLKDDSIDNKVLTRSD KNRGKSDNVPSEEVVKKMKNYWRQLLN AKLITQRKFDNLTKAERGGLSELDKAGFIK RQLVETRQITKHVAQILDSRMNTKYDEND KLIREVKVITLKSKLVSDFRKDFQFYKVRE INNYHHAHDAYLNAVVGTALIKKYPKLES EFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFVSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASARELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIE QISEFSKRVILADANLDKVLSAYNKHRDK PIREQAENIIHLFTLTNLGAPAAFKYFDTTI DRKQYRSTKEVLDATLIHQSITGLYETRID LSQLGGDEGADKRTADGSEFESPKKKRKV
[1384]	10. monoABE8.1_bpNLS + Y147R_Q154R_T166R 多肽序列	1445	MSEVEFSHEYWMRHALTLAKRARDEREV PVGAVLVLNNRVIGEGWNRAIGLHDPTAH AEIMALRQGGLVMQNYRLIDATLYVTFEP CVMCAGAMIHSRIGRVVFGVRNAKTGAA GSLMDVLHYPGMNHRVEITEGILADECAA LLCRFFRMPRRVFNAQKKAQSSRDSGGSS GGSSGSETPGTSESATPESSGGSSGGSDKK YSIGLAIGTNSVGWAVITDEYKVPSKKFKV LGNTDRHSIKKNLIGALLFDSGETAEATRL KRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIV DEVAYHEKYPTIYHLRKKLVDSTDKADLR LIYLALAHMIKFRGHFLIEGDLNPDNSDVD KLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRRLENLIAQLPGEKKNGLFGNLI ALSLGLTPNFKSNFDLAEDAKLQLSKDTY DDDLDNLLAQIGDQYADLFLAAKNLSDAI LLSDILRVNTEITKAPLSASMIKRYDEHHQ DLTLLKALVRQQLPEKYKEIFFDQSKNGY AGYIDGGASQEEFYKFIKPILEKMDGTEEL LVKLNREDLLRKQRTFDNGSIPHQIHLGEL HAILRRQEDFYPFLKDNREKIEKILTFRIPY YVGPLARGNSRFAWMTRKSEETITPWNFE

[1385]

		EVVDKGASAQSFIERMTNFDKNLPNEKVL PKHSLLYEYFTVYNELTKVKYVTEGMRKP AFLSGEQKKAIVDLLFKTNRKVTVKQLKE DYFKKIECFDSVEISGVEDRFNASLGTYHD LLKIIKDKDFLDNEENEDILEDIVLTLTLFE DREMIEERLKTYAHLFDDKVMKQLKRRR YTGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGRHKPENIVIEMARENQTTQKG QKNSRERMKRIEEGIKELGSQILKEHPVEN TQLQNEKLYLYYLQNGRDMYVDQELDIN RLSDYDVDHIVPQSFLKDDSIDNKVLTRSD KNRGKSDNVPSEEVVKKMKNYWRQLLN AKLITQRKFDNLTKAERGGLSELDKAGFIK RQLVETRQITKHVAQILDSRMNTKYDEND KLIREVKVITLKSKLVSDFRKDFQFYKVRE INNYHHAHDAYLNAVVGTALIKKYPKLES EFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFVSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASARELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIE QISEFSKRVILADANLDKVLSAYNKHRDK PIREQAENIIHLFTLTNLGAPAAFKYFDTTI DRKQYRSTKEVLDATLIHQSITGLYETRID LSQLGGDEGADKRTADGSEFESPKKKRKV
11. monoABE8.1_bpNLS + Y147T_Q154R 多肽序列	1446	MSEVEFSHEYWMRHALTLAKRARDEREV PVGAVLVLNNRVIGEGWNRAIGLHDPTAH AEIMALRQGGLVMQNYRLIDATLYVTFEP CVMCAGAMIHSRIGRVVFGVRNAKTGAA GSLMDVLHYPGMNHRVEITEGILADECAA LLCTFFRMPRRVFNAQKKAQSSTDSGGSS GGSSGSETPGTSESATPESSGGSSGGSDKK YSIGLAIGTNSVGWAVITDEYKVPSKKFKV LGNTDRHSIKKNLIGALLFDSGETAEATRL KRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIV DEVAYHEKYPTIYHLRKKLVDSTDKADLR

[1386]

		LIYLALAHMIKFRGHFLIEGDLNPDNSDVD KLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRRLENLIAQLPGEKKNGLFGNLI ALSLGLTPNFKSNFDLAEDAKLQLSKDTY DDDLDNLLAQIGDQYADLFLAAKNLSDAI LLSDILRVNTEITKAPLSASMIKRYDEHHQ DLTLLKALVRQQLPEKYKEIFFDQSKNGY AGYIDGGASQEEFYKFIKPILEKMDGTEEL LVKLNREDLLRKQRTFDNGSIPHQIHLGEL HAILRRQEDFYPFLKDNREKIEKILTFRIPY YVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVL PKHSLLYEYFTVYNELTKVKYVTEGMRKP AFLSGEQKKAIVDLLFKTNRKVTVKQLKE DYFKKIECFDSVEISGVEDRFNASLGTYHD LLKIIKDKDFLDNEENEDILEDIVLTLTLFE DREMIEERLKTYAHLFDDKVMKQLKRRR YTGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGRHKPENIVIEMARENQTTQKG QKNSRERMKRIEEGIKELGSQILKEHPVEN TQLQNEKLYLYYLQNGRDMYVDQELDIN RLSDYDVDHIVPQSFLKDDSIDNKVLTRSD KNRGKSDNVPSEEVVKKMKNYWRQLLN AKLITQRKFDNLTKAERGGLSELDKAGFIK RQLVETRQITKHVAQILDSRMNTKYDEND KLIREVKVITLKSKLVSDFRKDFQFYKVRE INNYHHAHDAYLNAVVGTALIKKYPKLES EFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFVSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASARELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIE QISEFSKRVILADANLDKVLSAYNKHRDK PIREQAENIIHLFTLTNLGAPAAFKYFDTTI DRKQYRSTKEVLDATLIHQSITGLYETRID LSQLGGDEGADKRTADGSEFESPKKKRKV
12.	1447	MSEVEFSHEYWMRHALTLAKRARDEREV

[1387]

monoABE8.1_bpNLS + Y147T_Q154S 多肽序列		PVGAVLVLNNRVIGEGWNRAIGLHDPTAH AEIMALRQGGLVMQNYRLIDATLYVTFEP CVMCAGAMIHSRIGRVVFGVRNAKTGAA GSLMDVLHYPGMNHRVEITEGILADECAA LLCTFFRMPRSVFNAQKKAQSSTDSGGSS GGSSGSETPGTSESATPESSGGSSGGSDKK YSIGLAIGTNSVGWAVITDEYKVPSKKFKV LGNTDRHSIKKNLIGALLFDSGETAEATRL KRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIV DEVAYHEKYPTIYHLRKKLVDSTDKADLR LIYLALAHMIKFRGHFLIEGDLNPDNSDVD KLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRRLENLIAQLPGEKKNGLFGNLI ALSLGLTPNFKSNFDLAEDAKLQLSKDTY DDDLDNLLAQIGDQYADLFLAAKNLSDAI LLSDILRVNTEITKAPLSASMIKRYDEHHQ DLTLLKALVRQQLPEKYKEIFFDQSKNGY AGYIDGGASQEEFYKFIKPILEKMDGTEEL LVKLNREDLLRKQRTFDNGSIPHQIHLGEL HAILRRQEDFYPFLKDNREKIEKILTFRIPY YVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVL PKHSLLYEYFTVYNELTKVKYVTEGMRKP AFLSGEQKKAIVDLLFKTNRKVTVKQLKE DYFKKIECFDSVEISGVEDRFNASLGTYHD LLKIIKDKDFLDNEENEDILEDIVLTLTLFE DREMIEERLKTYAHLFDDKVMKQLKRRR YTGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGRHKPENIVIEMARENQTTQKG QKNSRERMKRIEEGIKELGSQILKEHPVEN TQLQNEKLYLYYLQNGRDMYVDQELDIN RLSDYDVDHIVPQSFLKDDSIDNKVLTRSD KNRGKSDNVPSEEVVKKMKNYWRQLLN AKLITQRKFDNLTKAERGGLSELDKAGFIK RQLVETRQITKHVAQILDSRMNTKYDEND KLIREVKVITLKSKLVSDFRKDFQFYKVRE INNYHHAHDAYLNAVVGTALIKKYPKLES EFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI

[1388]

		VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFVSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASARELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIE QISEFSKRVILADANLDKVLSAYNKHRDK PIREQAENIIHLFTLTNLGAPAAFKYFDTTI DRKQYRSTKEVLDATLIHQSITGLYETRID LSQLGGDEGADKRTADGSEFESPKKKRKV
13. monoABE8.1_bpNLS + H123Y123H_Y147R_Q154R_I76Y 多肽序列	1448	MSEVEFSHEYWMRHALTLAKRARDEREV PVGAVLVLNNRVIGEGWNRAIGLHDPTAH AEIMALRQGGLVMQNYRLYDATLYVTFE PCVMCAGAMIHSRIGRVVFGVRNAKTGA AGSLMDVLHHPGMNHRVEITEGILADECA ALLCRFFRMPRRVFNAQKKAQSSTDSGGS SGGSSGSETPGTSESATPESSGGSSGGSDK KYSIGLAIGTNSVGWAVITDEYKVPSKKFK VLGNTDRHSIKKNLIGALLFDSGETAEATR LKRTARRRYTRRKNRICYLQEIFSNEMAK VDDSFFHRLEESFLVEEDKKHERHPIFGNI VDEVAYHEKYPTIYHLRKKLVDSTDKADL RLIYLALAHMIKFRGHFLIEGDLNPDNSDV DKLFIQLVQTYNQLFEENPINASGVDAKAI LSARLSKSRRLENLIAQLPGEKKNGLFGNL IALSLGLTPNFKSNFDLAEDAKLQLSKDTY DDDLDNLLAQIGDQYADLFLAAKNLSDAI LLSDILRVNTEITKAPLSASMIKRYDEHHQ DLTLLKALVRQQLPEKYKEIFFDQSKNGY AGYIDGGASQEEFYKFIKPILEKMDGTEEL LVKLNREDLLRKQRTFDNGSIPHQIHLGEL HAILRRQEDFYPFLKDNREKIEKILTFRIPY YVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVL PKHSLLYEYFTVYNELTKVKYVTEGMRKP AFLSGEQKKAIVDLLFKTNRKVTVKQLKE DYFKKIECFDSVEISGVEDRFNASLGTYHD LLKIIKDKDFLDNEENEDILEDIVLTLTLFE DREMIEERLKTYAHLFDDKVMKQLKRRR YTGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGRHKPENIVIEMARENQTTQKG

			QKNSRERMKRIEEGIKELGSQILKEHPVEN TQLQNEKLYLYYLQNGRDMYVDQELDIN RLSDYDVDHIVPQSFLKDDSIDNKVLTRSD KNRGKSDNVPSEEVVKKMKNYWRQLLN AKLITQRKFDNLTKAERGGLSELDKAGFIK RQLVETRQITKHVAQILDSRMNTKYDEND KLIREVKVITLKSKLVSDFRKDFQFYKVRE INNYHHAHDAYLNAVVGTALIKKYPKLES EFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFVSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASARELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIE QISEFSKRVILADANLDKVLSAYNKHRDK PIREQAENIIHLFTLTNLGAPAAFKYFDTTI DRKQYRSTKEVLDATLIHQSITGLYETRID LSQLGGDEGADKRTADGSEFESPKKKRKV
[1389]	14. monoABE8.1_bpNLS + V82S + Q154R 多肽序列	1449	MSEVEFSHEYWMRHALTLAKRARDEREV PVGAVLVLNNRVIGEGWNRAIGLHDPTAH AEIMALRQGGLVMQNYRLIDATLYSTFEP CVMCAGAMIHSRIGRVVFGVRNAKTGAA GSLMDVLHYPGMNHRVEITEGILADECAA LLCYFFRMPRRVFNAQKKAQSSTDSGGSS GGSSGSETPGTSESATPESSGGSSGGSDKK YSIGLAIGTNSVGWAVITDEYKVPSKKFKV LGNTDRHSIKKNLIGALLFDSGETAEATRL KRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIV DEVAYHEKYPTIYHLRKKLVDSTDKADLR LIYLALAHMIKFRGHFLIEGDLNPDNSDVD KLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRRLENLIAQLPGEKKNGLFGNLI ALSLGLTPNFKSNFDLAEDAKLQLSKDTY DDDLDNLLAQIGDQYADLFLAAKNLSDAI LLSDILRVNTEITKAPLSASMIKRYDEHHQ DLTLLKALVRQQLPEKYKEIFFDQSKNGY AGYIDGGASQEEFYKFIKPILEKMDGTEEL LVKLNREDLLRKQRTFDNGSIPHQIHLGEL HAILRRQEDFYPFLKDNREKIEKILTFRIPY

[1390]

		YVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVL PKHSLLYEYFTVYNELTKVKYVTEGMRKP AFLSGEQKKAIVDLLFKTNRKVTVKQLKE DYFKKIECFDSVEISGVEDRFNASLGTYHD LLKIIKDKDFLDNEENEDILEDIVLTLTLFE DREMIEERLKTYAHLFDDKVMKQLKRRR YTGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGRHKPENIVIEMARENQTTQKG QKNSRERMKRIEEGIKELGSQILKEHPVEN TQLQNEKLYLYYLQNGRDMYVDQELDIN RLSDYDVDHIVPQSFLKDDSIDNKVLTRSD KNRGKSDNVPSEEVVKKMKNYWRQLLN AKLITQRKFDNLTKAERGGLSELDKAGFIK RQLVETRQITKHVAQILDSRMNTKYDEND KLIREVKVITLKSKLVSDFRKDFQFYKVRE INNYHHAHDAYLNAVVGTALIKKYPKLES EFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIET NGETGEIVWDKGRDFATVRKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFVSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKR MLASARELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIE QISEFSKRVILADANLDKVLSAYNKHRDK PIREQAENIIHLFTLTNLGAPAAFKYFDTTI DRKQYRSTKEVLDATLIHQSITGLYETRID LSQLGGDEGADKRTADGSEFESPKKKRKV
接头	65	PAPAP
接头	66	PAPAPA
接头	67	PAPAPAP
接头	68	PAPAPAPA
接头	69	P(AP)4
接头	70	P(AP)7
接头	71	P(AP)10
N基因(核酸)	4001	atggatgccgacaagattgtattcaaagtcaataatcaggtggtctc tttgaagcctgagattatcgtggatcaatatgagtacaagtaccctg ccatcaaagatttgaaaaagccctgtataaccctaggaaaggctcc

[1391]

		cgatttaaataaagcatacaagtcagttttgtcaggcatgagcgcc gccaaacttaatcctgacgatgtatgttcctatttggcagcggcaat gcagttttttgaggggacatgtccggaagactggaccagctatgg aattgtgattgcacgaaaaggagataagatcaccccaggttctctg gtggagataaaacgtactgatgtagaagggaattgggctctgaca ggaggcatggaactgacaagagaccccactgtccctgagcatgc gtccttagtcggtcttctcttgagtctgtataggttgagcaaaatatcc gggcaaaacactggtaactataagacaaacattgcagacaggata gagcagatttttgagacagcccctttttgttaaaatcgtggaacacca tactctaatgacaactcacaaaatgtgtgctaattggagtactatacc aaacttcagatttttggccggaacctatgacatgtttttctcccggatt gagcatctatattcagcaatcagagtgggcacagttgtcactgctta tgaagactgttcaggactggtatcatttactgggttcataaaacaaat caatctcaccgctagagaggcaatactatatttcttccacaagaact ttgaggaagagataagaagaatgtttgagccagggcaggagaca gctgttcctcactcttatttcatccacttccgttcactaggcttgagtg ggaaatctccttattcatcaaatgctgttggtcacgtgttcaatctcat tcactttgtaggatgctatatgggtcaagtcagatccctaaatgcaa cggttattgctgcatgtgctcctcatgaaatgtctgttctaggggct atctgggagaggaattcttcgggaaagggacatttgaaagaagatt cttcagagatgagaaagaacttcaagaatacgaggcggctgaact gacaaagactgacgtagcactggcagatgatggaactgtcaactc tgacgacgaggactacttttcaggtgaaaccagaagtccggaggc tgtttatactcgaatcatgatgaatggaggtcgactaaagagatctc acatacggagatatgtctcagtcagttccaatcatcaagcccgtcc aaactcattcgccgagtttctaaacaagacatattcgagtgactca
N基因(氨基酸)	4002	MDADKIVFKVNNQVVSLKPEIIVDQYEYK YPAIKDLKKPCITLGKAPDLNKAYKSVLSG MSAAKLNPDDVCSYLAAAMQFFEGTCPE DWTSYGIVIARKGDKITPGSLVEIKRTDVE GNWALTGGMELTRDPTVPEHASLVGLLLS LYRLSKISGQNTGNYKTNIADRIEQIFETAP FVKIVEHHTLMTTHKMCANWSTIPNFRFL AGTYDMFFSRIEHLYSAIRVGTVVTAYED CSGLVSFTGFIKQINLTAREAILYFFHKNFE EEIRRMFEPGQETAVPHSYFIHFRSLGLSG KSPYSSNAVGHVFNLIHFVGCYMGQVRSL NATVIAACAPHEMSVLGGYLGEEFFGKGT FERRFFRDEKELQEYEAAELTKTDVALAD DGTVNSDDEDYFSGETRSPEAVYTRIMMN GGRLKRSHIRRYVSVSSNHQARPNSFAEFL NKTYSSDS
L基因(核	4003	ctcgatcctggagaggtctatgatgaccctattgacccaatcgagtt

[1392]

酸)		agaggctgaacccagaggaacccccattgtccccaacatcttgag gaactctgactacaatctcaactctcctttgatagaagatcctgctag actaatgttagaatggttaaaaacagggaatagaccttatcggatg actctaacagacaattgctccaggtctttcagagttttgaaagattatt tcaagaaggtagatttgggttctctcaaggtgggcggaatggctgc acagtcaatgatttctctctggttatatggtgcccactctgaatccaa caggagccggagatgtataacagacttggcccatttctattccaag tcgtcccccatagagaagctgttgaatctcacgctaggaaatagag ggctgagaatccccccagagggagtgttaagttgccttgagaggg ttgattatgataatgcatttggaaggtatcttgccaacacgtattcctc ttacttgttcttccatgtaatcaccttatacatgaacgccctagactgg gatgaagaaaagaccatcctagcattatggaaagatttaacctcag tggacatcgggaaggacttggtaaagttcaaagaccaaatatggg gactgctgatcgtgacaaaggactttgtttactcccaaagttccaatt gtctttttgacagaaactacacacttatgctaaaagatcttttcttgtct cgcttcaactccttaatggtcttgctctctcccccagagccccgata ctcagatgacttgatatctcaactatgccagctgtacattgctgggg atcaagtcttgtctatgtgtggaaactccggctatgaagtcatcaaa atattggagccatatgtcgtgaatagtttagtccagagagcagaaa agtttaggcctctcattcattccttgggagactttcctgtatttataaaa gacaaggtaagtcaacttgaagagacgttcggtccctgtgcaaga aggttctttagggctctggatcaattcgacaacatacatgacttggtt tttgtgtttggctgttacaggcattgggggcacccatatatagattat cgaaagggtctgtcaaaactatatgatcaggttcaccttaaaaaaat gatagataagtcctaccaggagtgcttagcaagcgacctagccag gaggatccttagatggggttttgataagtactccaagtggtatctgg attcaagattcctagcccgagaccacccttgactccttatatcaaa acccaaacatggccacccaaacatattgtagacttggtgggggat acatggcacaagctcccgatcacgcagatctttgagattcctgaat caatggatccgtcagaaatattggatgacaaatcacattctttcacc agaacgagactagcttcttggctgtcagaaaaccgaggggggcc tgttcctagcgaaaaagttattatcacggccctgtctaagccgcctg tcaatccccgagagtttctgaggtctatagacctcggaggattgcc agatgaagacttgataattggcctcaagccaaaggaacgggaatt gaagattgaaggtcgattctttgctctaatgtcatggaatctaagatt gtattttgtcatcactgaaaaactcttggccaactacatcttgccactt tttgacgcgctgactatgacagacaacctgaacaaggtgtttaaaa agctgatcgacagggtcaccgggcaagggcttttggactattcaa gggtcacatatgcatttcacctggactatgaaaagtggaacaacca tcaaagattagagtcaacagaggatgtattttctgtcctagatcaagt gtttggattgaagagagtgttttctagaacacacgagtttttttcaaaa ggcctggatctattattcagacagatcagacctcatcgggttacgg gaggatcaaatatactgcttagatgcgtccaacggcccaacctgtt

[1393]

		ggaatggccaggatggcgggctagaaggcttacggcagaaggg ctggagtctagtcagcttattgatgatagatagagaatctcaaatca ggaacacaagaaccaaaatactagctcaaggagacaaccaggtt ttatgtccgacatacatgttgtcgccagggctatctcaagaggggct cctctatgaattggagagaatatcaaggaatgcactttcgatataca gagccgtcgaggaaggggcatctaagctagggctgatcatcaag aaagaagagaccatgtgtagttatgacttcctcatctatggaaaaac ccctttgtttagaggtaacatattggtgcctgagtccaaaagatggg ccagagtctcttgcgtctctaatgaccaaatagtcaacctcgccaat ataatgtcgacagtgtccaccaatgcgctaacagtggcacaacact ctcaatctttgatcaaaccgatgagggattttctgctcatgtcagtac aggcagtctttcactacctgctatttagcccaatcttaaagggaaga gtttacaagattctgagcgctgaaggggagagctttctcctagccat gtcaaggataatctatctagatccttctttgggagggatatctggaat gtccctcggaagattccatatacgacagttctcagaccctgtctctg aagggttatccttctggagagagatctggttaagctcccaagagtc ctggattcacgcgttgtgtcaagaggctggaaacccagatcttgga gagagaacactcgagagcttcactcgccttctagaagatccgacc accttaaatatcagaggaggggccagtcctaccattctactcaagg atgcaatcagaaaggctttatatgacgaggtggacaaggtggaaa attcagagtttcgagaggcaatcctgttgtccaagacccatagaga taattttatactcttcttaatatctgttgagcctctgtttcctcgatttctc agtgagctattcagttcgtctttttgggaatccccgagtcaatcattg gattgatacaaaactcccgaacgataagaaggcagtttagaaaga gtctctcaaaaactttagaagaatccttctacaactcagagatccac gggattagtcggatgacccagacacctcagagggttgggggggt gtggccttgctcttcagagagggcagatctacttagggagatctctt ggggaagaaaagtggtaggcacgacagttcctcacccttctgaga tgttgggattacttcccaagtcctctatttcttgcacttgtggagcaac aggaggaggcaatcctagagtttctgtatcagtactcccgtcctttg atcagtcatttttttcacgaggccccctaaagggatacttgggctcgt ccacctctatgtcgacccagctattccatgcatgggaaaaagtcac taatgttcatgtggtgaagagagctctatcgttaaaagaatctataaa ctggttcattactagagattccaacttggctcaagctctaattaggaa cattatgtctctgacaggccctgatttccctctagaggaggcccctg tcttcaaaaggacggggtcagccttgcataggttcaagtctgccag atacagcgaaggagggtattcttctgtctgcccgaacctcctctctc atatttctgttagtacagacaccatgtctgatttgacccaagacggg aagaactacgatttcatgttccagccattgatgctttatgcacagaca tggacatcagagctggtacagagagacacaaggctaagagactc tacgtttcattggcacctccgatgcaacaggtgtgtgagacccattg acgacgtgaccctggagacctctcagatcttcgagtttccggatgt gtcgaaaagaatatccagaatggtttctggggctgtgcctcacttcc

[1394]

		agaggcttcccgatatccgtctgagaccaggagattttgaatctcta agcggtagagaaaagtctcaccatatcggatcagctcaggggctc ttatactcaatcttagtggcaattcacgactcaggatacaatgatgg aaccatcttccctgtcaacatatacggcaaggtttcccctagagact atttgagagggctcgcaaggggagtattgataggatcctcgatttg cttcttgacaagaatgacaaatatcaatattaatagacctcttgaattg gtctcaggggtaatctcatatattctcctgaggctagataaccatcc ctccttgtacataatgctcagagaaccgtctcttagaggagagatat tttctatccctcagaaaatccccgccgcttatccaaccactatgaaa gaaggcaacagatcaatcttgtgttatctccaacatgtgctacgcta tgagcgagagataatcacggcgtctccagagaatgactggctatg gatcttttcagactttagaagtgccaaaatgacgtacctatccctcat tacttaccagtctcatcttctactccagagggttgagagaaacctatc taagagtatgagagataacctgcgacaattgagttctttgatgaggc aggtgctgggcgggcacggagaagataccttagagtcagacga caacattcaacgactgctaaaagactctttacgaaggacaagatgg gtggatcaagaggtgcgccatgcagctagaaccatgactggaga ttacagccccaacaagaaggtgtcccgtaaggtaggatgttcaga atgggtctgctctgctcaacaggttgcagtctctacctcagcaaacc cggcccctgtctcggagcttgacataagggccctctctaagaggtt ccagaacccctttgatctcgggcttgagagtggttcagtgggcaacc ggtgctcattataagcttaagcctattctagatgatctcaatgttttcc catctctctgccttgtagttggggacgggtcaggggggatatcaag ggcagtcctcaacatgtttccagatgccaagcttgtgttcaacagtc ttttagaggtgaatgacctgatggcttccggaacacatccactgcct ccttcagcaatcatgaggggaggaaatgatatcgtctccagagtg atagatcttgactcaatctgggaaaaaccgtccgacttgagaaactt ggcaacctggaaatacttccagtcagtccaaaagcaggtcaacat gtcctatgacctcattatttgcgatgcagaagttactgacattgcatct atcaaccggatcaccctgttaatgtccgattttgcattgtctatagatg gaccactctatttggtcttcaaaacttatgggactatgctagtaaatc caaactacaaggctattcaacacctgtcaagagcgttcccctcggt cacagggtttatcacccaagtaacttcgtctttttcatctgagctctac ctccgattctccaaacgagggaagtttttcagagatgctgagtactt gacctcttccacccttcgagaaatgagccttgtgttattcaattgtag cagccccaagagtgagatgcagagagctcgttccttgaactatca ggatcttgtgagaggatttcctgaagaaatcatatcaaatccttaca atgagatgatcataactctgattgacagtgatgtagaatcttttctagt ccacaagatggttgatgatcttgagttacagaggggaactctgtcta aagtggctatcattatagccatcatgatagttttctccaacagagtctt caacgtttccaaaccccctaactgacccctcgttctatccaccgtctg atcccaaaatcctgaggcacttcaacatatgttgcagtactatgatgt atctatctactgctttaggtgacgtccctagcttcgcaagacttcacg

[1395]

		acctgtataacagacctataacttattacttcagaaagcaagtcattc gagggaacgtttatctatcttggagttggtccaacgacacctcagtg ttcaaaagggtagcctgtaattctagcctgagtctgtcatctcactgg atcaggttgatttacaagatagtgaagactaccagactcgttggca gcatcaaggatctatccagagaagtggaaagacaccttcataggt acaacaggtggatcaccctagaggatatcagatctagatcatccct actagactacagttgcctg
L 基因(氨基酸)	4004	LDPGEVYDDPIDPIELEAEPRGTPIVPNILR NSDYNLNSPLIEDPARLMLEWLKTGNRPY RMTLTDNCSRSFRVLKDYFKKVDLGSLKV GGMAAQSMISLWLYGAHSESNRSRRCITD LAHFYSKSSPIEKLLNLTLGNRGLRIPPEGV LSCLERVDYDNAFGRYLANTYSSYLFFHVI TLYMNALDWDEEKTILALWKDLTSVDIG KDLVKFKDQIWGLLIVTKDFVYSQSSNCL FDRNYTLMLKDLFLSRFNSLMVLLSPPEPR YSDDLISQLCQLYIAGDQVLSMCGNSGYE VIKILEPYVVNSLVQRAEKFRPLIHSLGDFP VFIKDKVSQLEETFGPCARRFFRALDQFDN IHDLVFVFGCYRHWGHPYIDYRKGLSKLY DQVHLKKMIDKSYQECLASDLARRILRWG FDKYSKWYLDSRFLARDHPLTPYIKTQTW PPKHIVDLVGDTWHKLPITQIFEIPESMDPS EILDDKSHSFTRTRLASWLSENRGGPVPSE KVIITALSKPPVNPREFLRSIDLGGLPDEDLI IGLKPKERELKIEGRFFALMSWNLRLYFVI TEKLLANYILPLFDALTMTDNLNKVFKKLI DRVTGQGLLDYSRVTYAFHLDYEKWNNH QRLESTEDVFSVLDQVFGLKRVFSRTHEFF QKAWIYYSDRSDLIGLREDQIYCLDASNGP TCWNGQDGGLEGLRQKGWSLVSLLMIDR ESQIRNTRTKILAQGDNQVLCPTYMLSPGL SQEGLLYELERISRNALSIYRAVEEGASKL GLIIKKEETMCSYDFLIYGKTPLFRGNILVP ESKRWARVSCVSNDQIVNLANIMSTVSTN ALTVAQHSQSLIKPMRDFLLMSVQAVFHY LLFSPILKGRVYKILSAEGESFLLAMSRIIYL DPSLGGISGMSLGRFHIRQFSDPVSEGLSF WREIWLSSQESWIHALCQEAGNPDLGERT LESFTRLLEDPTTLNIRGGASPTILLKDAIR KALYDEVDKVENSEFREAILLSKTHRDNFI LFLISVEPLFPRFLSELFSSSFLGIPESIIGLIQ NSRTIRRQFRKSLSKTLEESFYNSEIHGISR

[1396]

		MTQTPQRVGGVWPCSSERADLLREISWGR KVVGTTVPHPSEMLGLLPKSSISCTCGATG GGNPRVSVSVLPSFDQSFFSRGPLKGYLGS STSMSTQLFHAWEKVTNVHVVKRALSLK ESINWFITRDSNLAQALIRNIMSLTGPDFPL EEAPVFKRTGSALHRFKSARYSEGGYSSV CPNLLSHISVSTDTMSDLTQDGKNYDFMF QPLMLYAQTWTSELVQRDTRLRDSTFHW HLRCNRCVRPIDDVTLETSQIFEFPDVSKRI SRMVSGAVPHFQRLPDIRLRPGDFESLSGR EKSHHIGSAQGLLYSILVAIHDSGYNDGTIF PVNIYGKVSPRDYLRGLARGVLIGSSICFLT RMTNININRPLELVSGVISYILLRLDNHPSL YIMLREPSLRGEIFSIPQKIPAAYPTTMKEG NRSILCYLQHVLRYEREIITASPENDWLWI FSDFRSAKMTYLSLITYQSHLLLQRVERNL SKSMRDNLRQLSSLMRQVLGGHGEDTLES DDNIQRLLKDSLRRTRWVDQEVRHAART MTGDYSPNKKVSRKVGCSEWVCSAQQVA VSTSANPAPVSELDIRALSKRFQNPLISGLR VVQWATGAHYKLKPILDDLNVFPSLCLVV GDGSGGISRAVLNMFPDAKLVFNSLLEVN DLMASGTHPLPPSAIMRGGNDIVSRVIDLD SIWEKPSDLRNLATWKYFQSVQKQVNMS YDLIICDAEVTDIASINRITLLMSDFALSIDG PLYLVFKTYGTMLVNPNYKAIQHLSRAFP SVTGFITQVTSSFSSELYLRFSKRGKFFRDA EYLTSSTLREMSLVLFNCSSPKSEMQRARS LNYQDLVRGFPEEIISNPYNEMIITLIDSDV ESFLVHKMVDDLELQRGTLSKVAIIIAIMIV FSNRVFNVSKPLTDPSFYPPSDPKILRHFNI CCSTMMYLSTALGDVPSFARLHDLYNRPI TYYFRKQVIRGNVYLSWSWSNDTSVFKR VACNSSLSLSSHWIRLIYKIVKTTRLVGSIK DLSREVERHLHRYNRWITLEDIRSRSSLLD YSCL
M 基因 (核酸)	4005	ttctagaagcagagaggaatctttgtcctcttcggacctttgtgtctg aagagacatgtcagaccatagttgacatgctctcgggttcatgttga tacaccagactctgccctggatatgacactgttttgcaatcactctta tttgcaatccgacgaactcagtatcatcatcccaagtgatctcctga gagtattccaactcctccccttcaagagggcccctggaatcagccc actggaagataaaggttctcctcaatttgtatacccagttcaggccc tcagggactggagatcctgacaaagccagtccaataaccactttg

[1397]

		actaacccgatcatcctatgattcccagaatatatctcgtcgaatgat ttcagaatgtgccgcaggatcctgaacgagtaaccattcgggctac acactttaacccttccgttgatacaaaagttcctcatgttcttcttgcct gtaagttctttcagcgggacgtattcagggggtggaagccacaag tcatcgtcatccagaggggctgacgcgggagaggatttttgagtgt cctcgtccctgcggtttttcactatcttacgtaggaggtt
M 基因 (氨基酸)	4006	NLLRKIVKNRRDEDTQKSSPASAPLDDDD LWLPPPEYVPLKELTGKKNMRNFCINGRV KVCSPNGYSFRILRHILKSFDEIYSGNHRMI GLVKVVIGLALSGSPVPEGLNWVYKLRRT FIFQWADSRGPLEGEELEYSQEITWDDDTE FVGLQIRVIAKQCHIQGRVWCINMNPRAC QLWSDMSLQTQRSEEDKDSSLLLE
P 基因(核 酸)	4007	agcaagatctttgtcaatcctagtgctattagagccggtctggccga tcttgagatggctgaagaaactgttgatctgatcaatagaaatatcg aagacaatcaggctcatctccaaggggaacccatagaggtggac aatctccctgaggatatggggcgacttcacctggatgatggaaaat cgcccaaccatggtgagatagccaaggtgggagaaggcaagtat cgagaggactttcagatggatgaaggagaggatcctagcttcctgt tccagtcatacctggaaaatgttggagtccaaatagtcagacaaat gaggtcaggagagagatttctcaagatatggtcacagaccgtaga agagattatatcctatgtcgcggtcaactttcccaaccctccaggaa agtcttcagaggataaatcaacccagactactggccgagagctca agaaggagacaacacccactccttctcagagagaaagccaatcat cgaaagccaggatggcggctcaaattgcttctggccctccagccc ttgaatggtcggctaccaatgaagaggatgatctatcagtggaggc tgagatcgctcaccagattgcagaaagtttctccaaaaaatataagt ttccctctcgatcctcagggatactcttgtataattttgagcaattgaa aatgaaccttgatgatatagttaaagaggcaaaaaatgtaccaggt gtgacccgtttagcccatgacgggtccaaactccccctaagatgtg tactgggatgggtcgctttggccaactctaagaaattccagttgtta gtcgaatccgacaagctgagtaaaatcatgcaagatgacttgaatc gctatacatcttgc
P 基因(氨 基酸)	4008	SKIFVNPSAIRAGLADLEMAEETVDLINRNI EDNQAHLQGEPIEVDNLPEDMGRLHLDDG KSPNHGEIAKVGEGKYREDFQMDEGEDPS FLFQSYLENVGVQIVRQMRSGERFLKIWS QTVEEIISYVAVNFPNPPGKSSEDKSTQTT GRELKKETTPTPSQRESQSSKARMAAQIAS GPPALEWSATNEEDDLSVEAEIAHQIAESF SKKYKFPSRSSGILLYNFEQLKMNLDDIVK EAKNVPGVTRLAHDGSKLPLRCVLGWVA LANSKKFQLLVESDKLSKIMQDDLNRYTS

[1398]

		C
G基因(核酸)	4009	atggttcctcaggctctcctgtttgtacccttctggtttttccattgtgt tttgggaaattccctatttacacgataccagacaagcttggtccctg gagtccgattgacatacatcacctcagctgcccaaacaatttggta gtggaggacgaaggatgcaccaacctgtcagggttctcctacatg gaacttaaagttggatacatcttagccataaaagtgaacgggttcac ttgcacaggcgttgtgacggaggctgaaacctacactaacttcgtt ggttatgtcacaaccacgttcaaaagaaagcatttccgcccaacac cagatgcatgtagagccgcgtacaactggaagatggccggtgac cccagatatgaagagtctctacacaatccgtaccctgactaccgct ggcttcgaactgtaaaaaccaccaaggagtctctcgttatcatatct ccaagtgtggcagatttggacccatatgacagatcccttcactcga gggtcttccctagcgggaagtgctcaggagtagcggtgtcttctac ctactgctccactaaccacgattacaccatttggatgcccgagaatc cgagactagggatgtcttgtgacatttttaccaatagtagagggaa gagagcatccaaagggagtgagacttgcggctttgtagatgaaag aggcctatataagtctttaaaaggagcatgcaaactcaagttatgtg gagttctaggacttagacttatggatggaacatgggtctcgatgca aacatcaaatgaaaccaaatggtgccctcccgataagttggtgaac ctgcacgactttcgctcagacgaaattgagcaccttgttgtagagg agttggtcaggaagagagaggagtgtctggatgcactagagtcca tcatgacaaccaagtcagtgagtttcagacgtctcagtcatttaaga aaacttgtccctgggtttggaaaagcatataccatattcaacaagac cttgatggaagccgatgctcactacaagtcagtcagaacttggaat gagatcctcccttcaaaagggtgtttaagagttgggggggaggtgtc atcctcatgtgaacggggtgtttttcaatggtataatattaggacctg acggcaatgtcttaatcccagagatgcaatcatccctcctccagca acatatggagttgttggaatcctcggttatccccccttgtgcacccct ggcagacccgtctaccgttttcaaggacggtgacgaggctgagg attttgttgaagttcaccttcccgatgtgcacaatcaggtctcaggag ttgacttgggtctcccgaactgggggaagtatgtattactgagtgca ggggccctgactgccttgatgttgataattttcctgatgacatgttgt agaagagtcaatcgatcagaacctacgcaacacaatctcagagg gacagggagggaggtgtcagtcactccccaaagcgggaagatc atatcttcatgggaatcacacaagagtgggggtgagaccagactg
G基因(氨基酸)	4010	MVPQALLFVPLLVFPLCFGKFPIYTIPDKLG PWSPIDIHHLSCPNNLVVEDEGCTNLSGFS YMELKVGYILAIKVNGFTCTGVVTEAETY TNFVGYVTTTFKRKHFRPTPDACRAAYN WKMAGDPRYEESLHNPYPDYRWLRTVKT TKESLVIISPSVADLDPYDRSLHSRVFPSGK CSGVAVSSTYCSTNHDYTIWMPENPRLGM SCDIFTNSRGKRASKGSETCGFVDERGLYK

[1399]

		SLKGACKLKLCGVLGLRLMDGTWVSMQT SNETKWCPPDKLVNLHDFRSDEIEHLVVE ELVRKREECLDALESIMTTKSVSFRRLSHL RKLVPGFGKAYTIFNKTLMEADAHYKSVR TWNEILPSKGCLRVGGRCHPHVNGVFFNG IILGPDGNVLIPEMQSSLLQQHMELLESSVI PLVHPLADPSTVFKDGDEAEDFVEVHLPD VHNQVSGVDLGLPNWGKYVLLSAGALTA LMLIIFLMTCCRRVNRSEPTQHNLRGTGRE VSVTPQSGKIISSWESHKSGGETRL
HEK2-2靶	4011	gaacacaaagcatagactgc

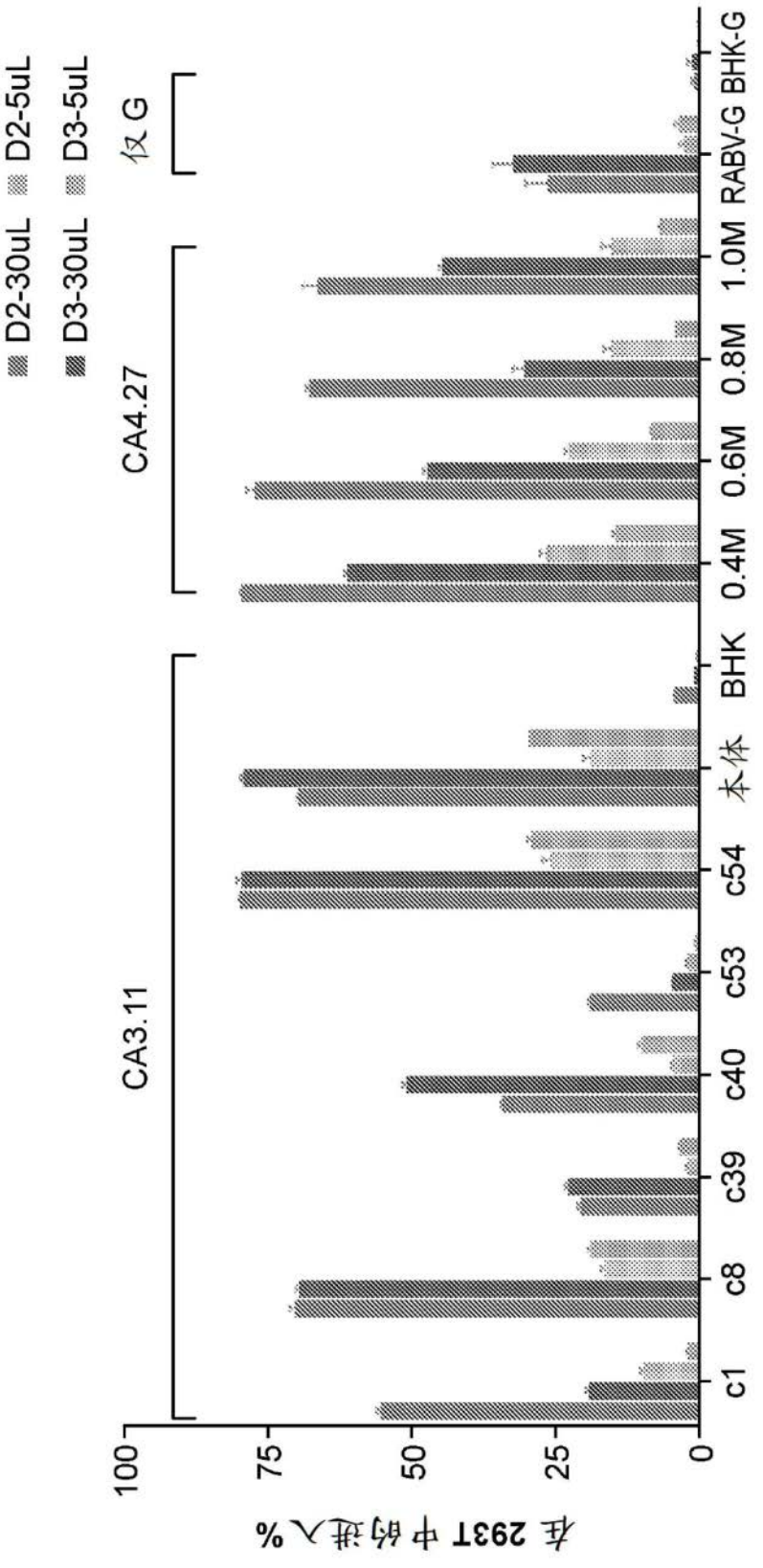

D2-30uL
D2-5uL
D3-30uL
D3-5uL
仅 G
CA4.27
CA3.11
在 293T 中的进入%
100
75
50
25
0
c1
c8
c39
c40
c53
c54
本体
BHK
0.4M
0.6M
0.8M
1.0M
RABV-G
BHK-G

图1

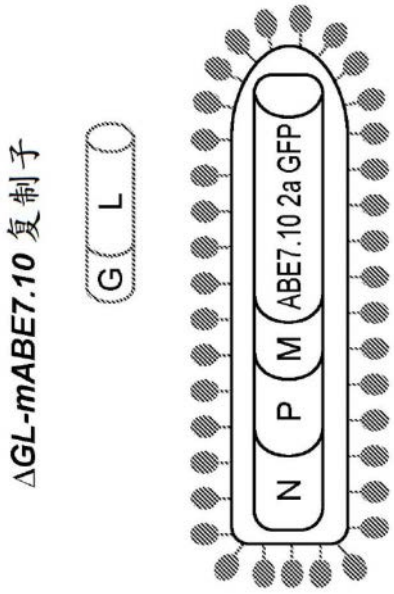


图2A

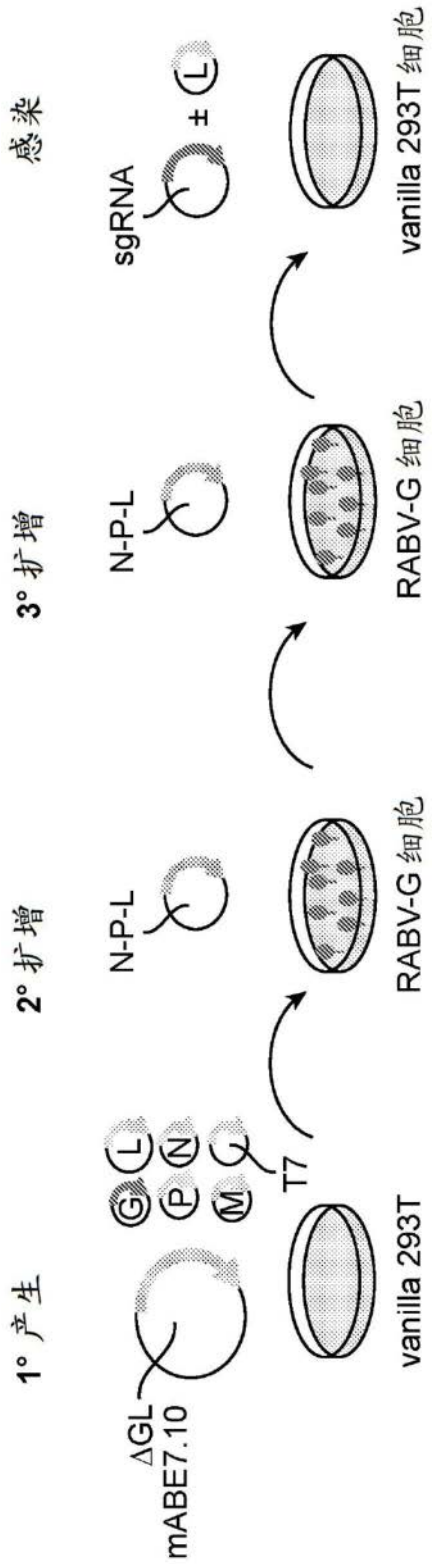

1° 产生
2° 扩增
3° 扩增
感染
ΔGL
mABE7.10
G
L
P
N
M
L
T7
vanilla 293T
RABV-G 细胞
N-P-L
RABV-G 细胞
N-P-L
sgRNA
± L
vanilla 293T 细胞

图2B

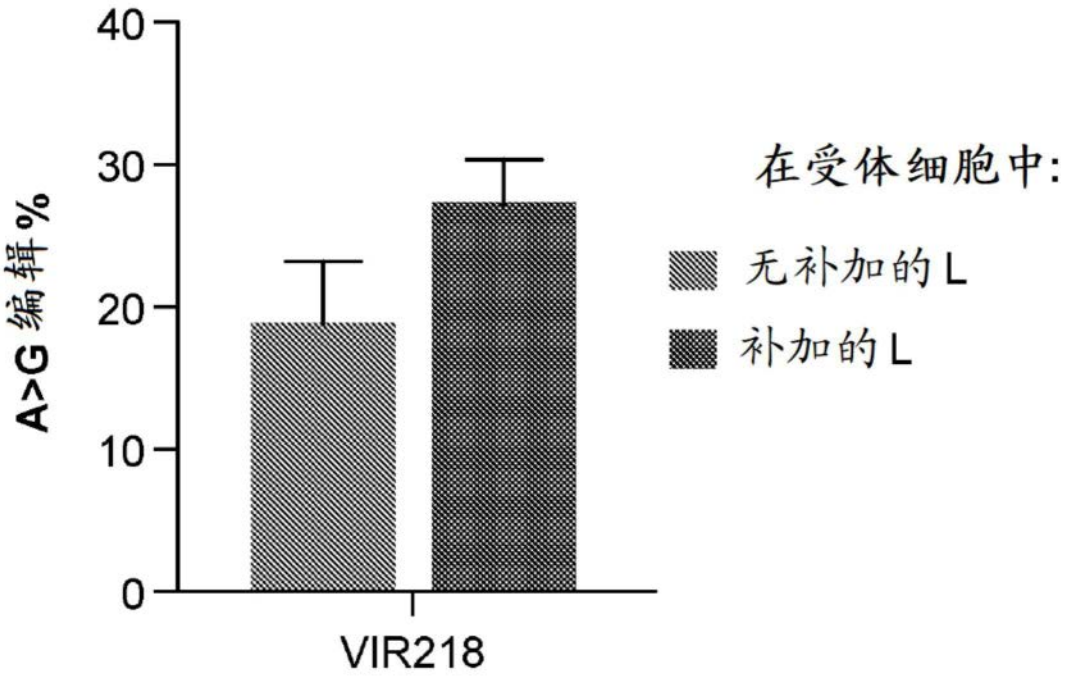


图2C

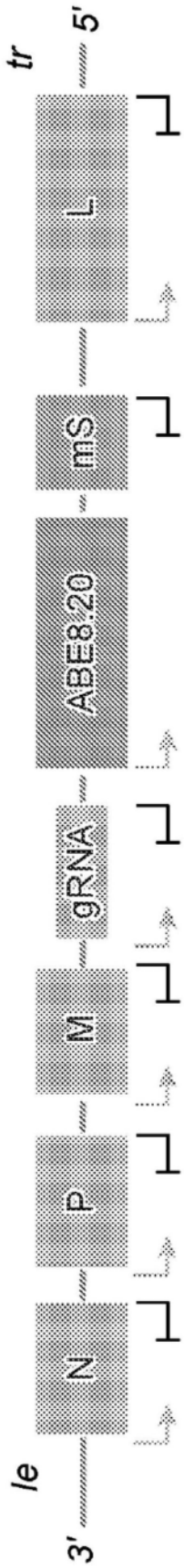


图3A

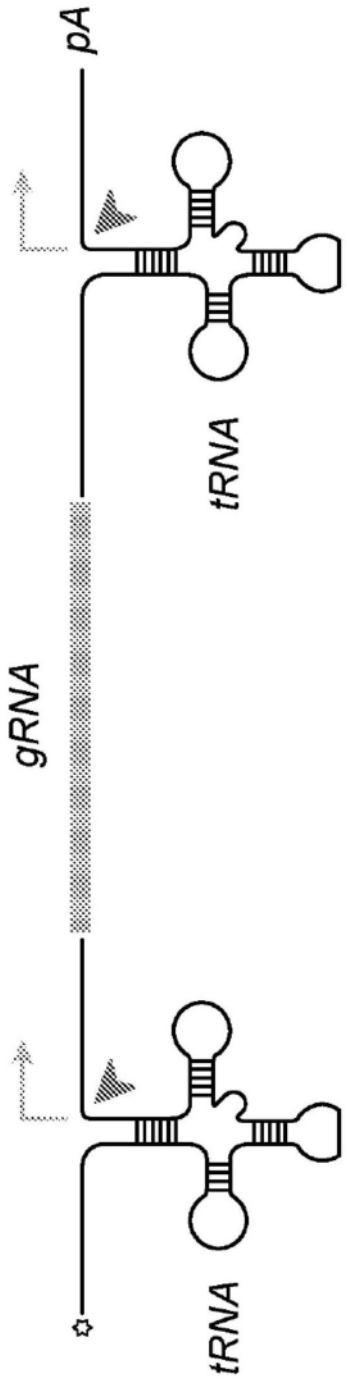


图3B

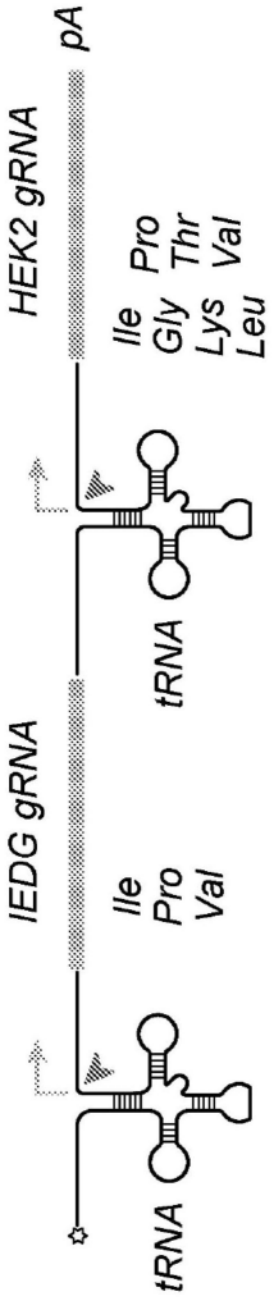

pA
HEK2 gRNA
Ile
Gly
Lys
Leu
Pro
Thr
Val
tRNA
IEDG gRNA
Ile
Pro
Val
tRNA

图3C

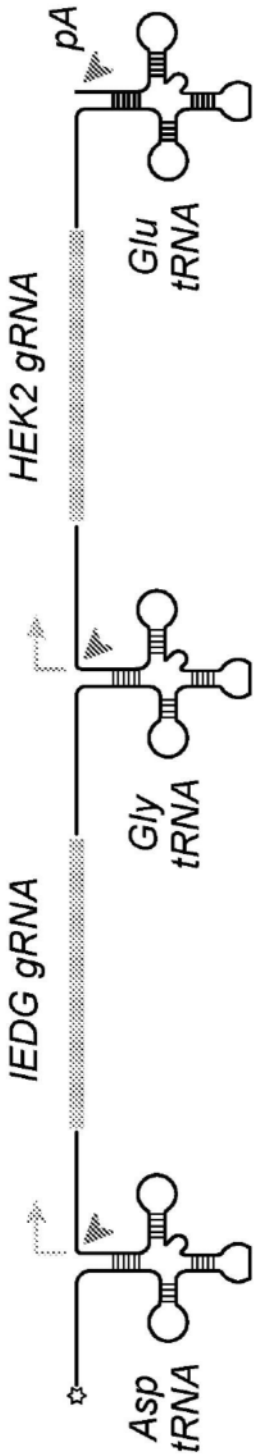


图3D

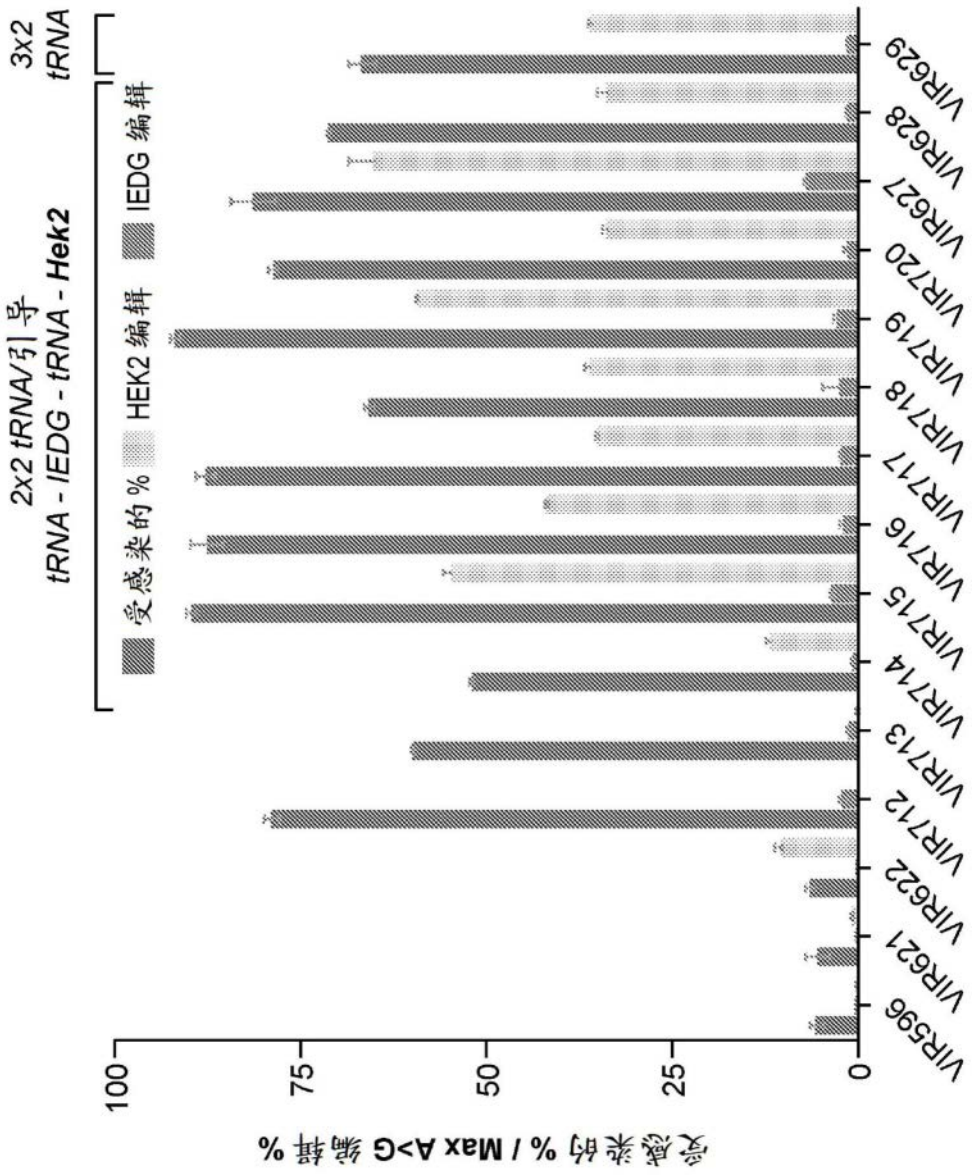

2x2 tRNA/引导
tRNA - IEDG - tRNA - Hek2
受感染的 % HEK2 编辑 IEDG 编辑
3x2
tRNA
受感染的 % / Max A>G 编辑 %
100
75
50
25
0
VIR596
VIR621
VIR622
VIR712
VIR713
VIR714
VIR715
VIR716
VIR717
VIR718
VIR719
VIR720
VIR627
VIR628
VIR629

图3E

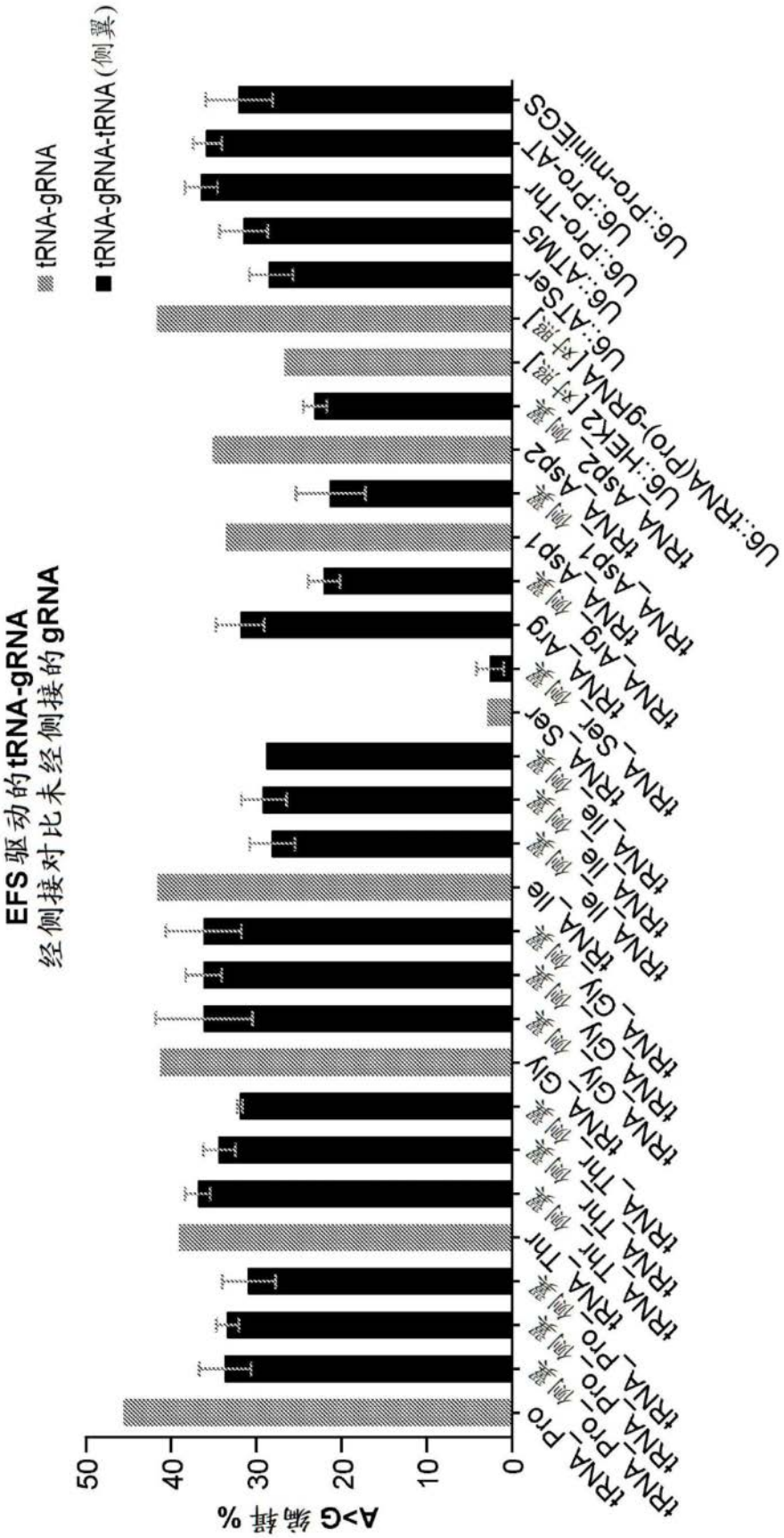

EFS 驱动的tRNA-gRNA
经侧接对比未经侧接的 gRNA
tRNA-gRNA
tRNA-gRNA-tRNA (侧翼)
A>G 编辑 %
0
10
20
30
40
50
tRNA_Pro
tRNA_Pro_侧翼
tRNA_Pro_侧翼
tRNA_Pro_侧翼
tRNA_Thr
tRNA_Thr_侧翼
tRNA_Thr_侧翼
tRNA_Thr_侧翼
tRNA_Gly
tRNA_Gly_侧翼
tRNA_Gly_侧翼
tRNA_Gly_侧翼
tRNA_Ile
tRNA_Ile_侧翼
tRNA_Ile_侧翼
tRNA_Ile_侧翼
tRNA_Ser
tRNA_Ser_侧翼
tRNA_Arg
tRNA_Arg_侧翼
tRNA_Asp1
tRNA_Asp1_侧翼
tRNA_Asp2
tRNA_Asp2_侧翼
U6::HEK2 [对照]
U6::tRNA(Pro)-gRNA [对照]
U6::ATSer [对照]
U6::ATM5
U6::Pro-Thr
U6::Pro-AT
U6::Pro-miniEGS

图4A

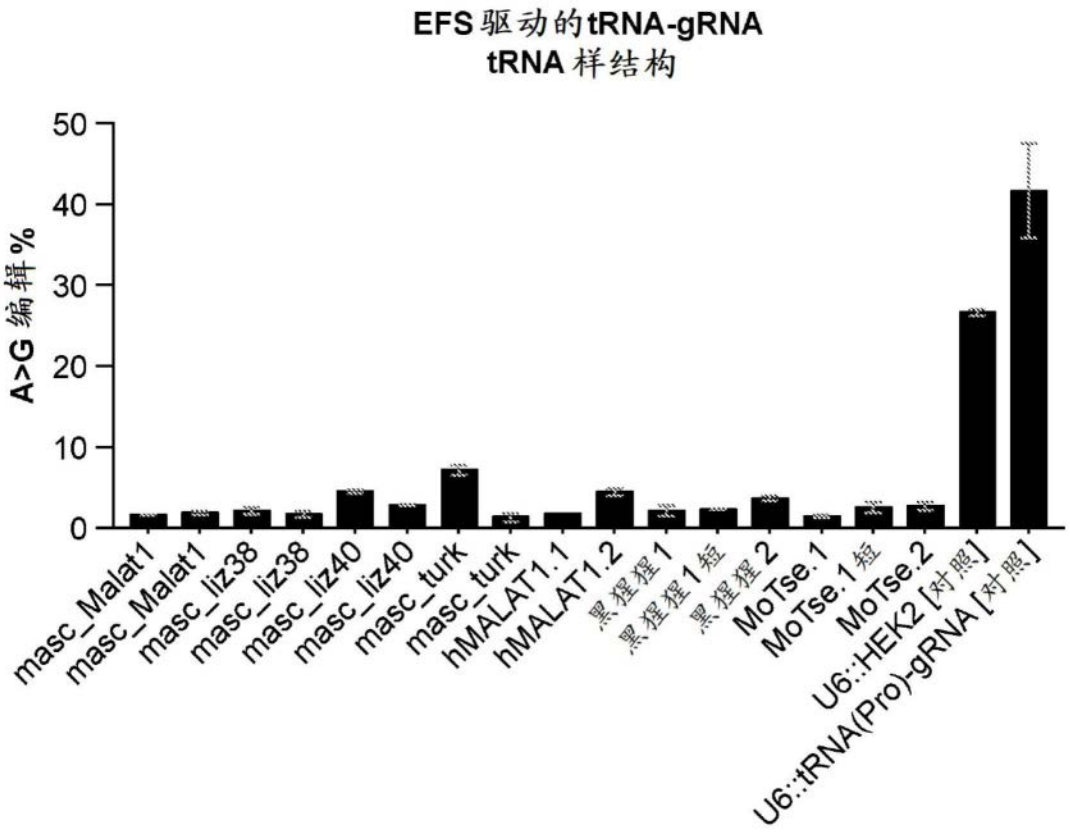

EFS驱动的tRNA-gRNA
tRNA样结构
A>G 编辑 %
50
40
30
20
10
0
masc_Malat1
masc_Malat1
masc_liz38
masc_liz38
masc_liz40
masc_liz40
masc_turk
masc_turk
hMALAT1.1
hMALAT1.2
黑猩猩1
黑猩猩1短
黑猩猩2
MoTse.1
MoTse.1短
MoTse.2
U6::HEK2 [对照]
U6::tRNA(Pro)-gRNA [对照]

图4B

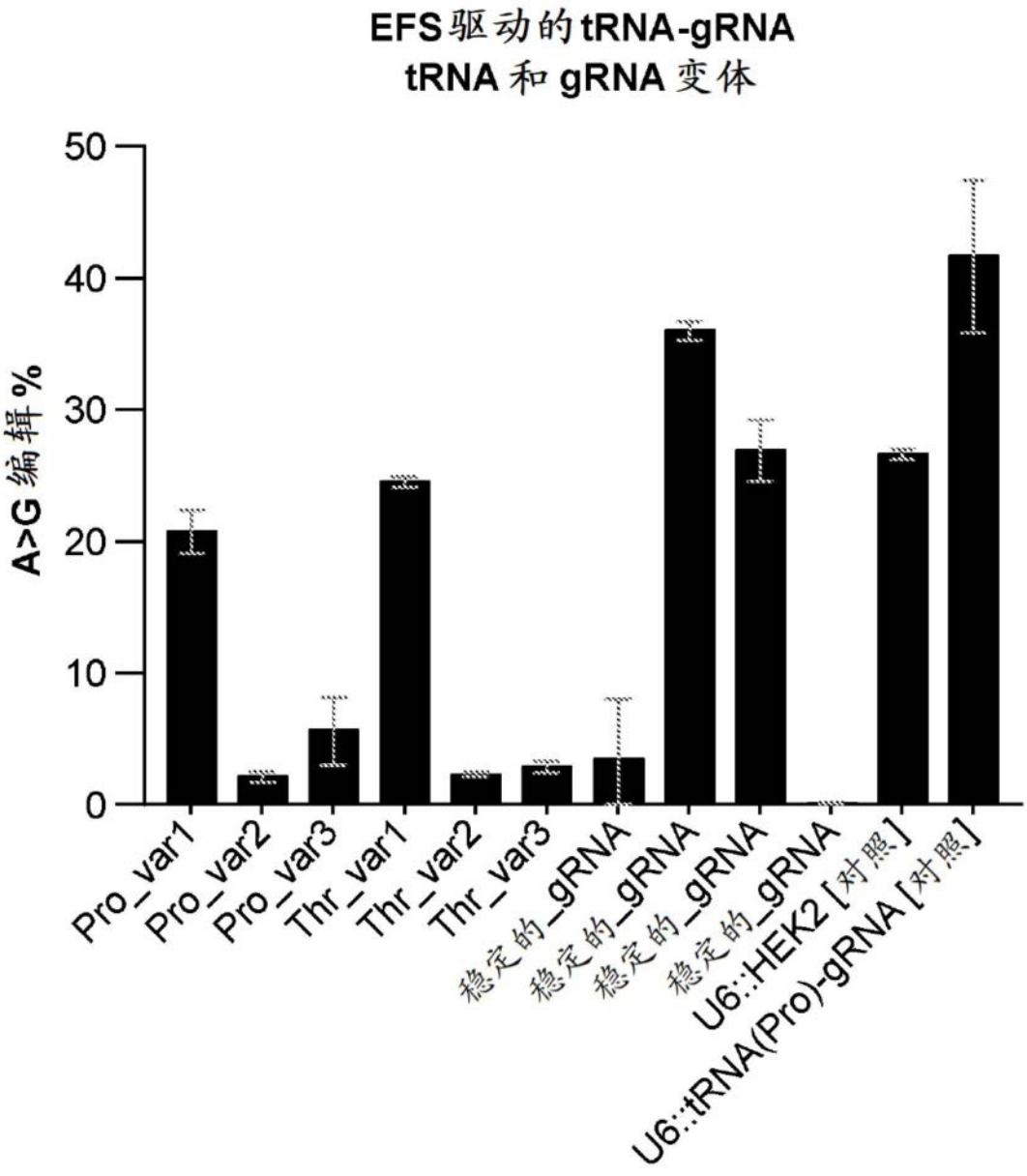

EFS驱动的tRNA-gRNA
tRNA和gRNA变体
A>G编辑%
50
40
30
20
10
0
Pro_var1
Pro_var2
Pro_var3
Thr_var1
Thr_var2
Thr_var3
稳定的_gRNA
稳定的_gRNA
稳定的_gRNA
稳定的_gRNA
U6::HEK2[对照]
U6::tRNA(Pro)-gRNA[对照]

图4C

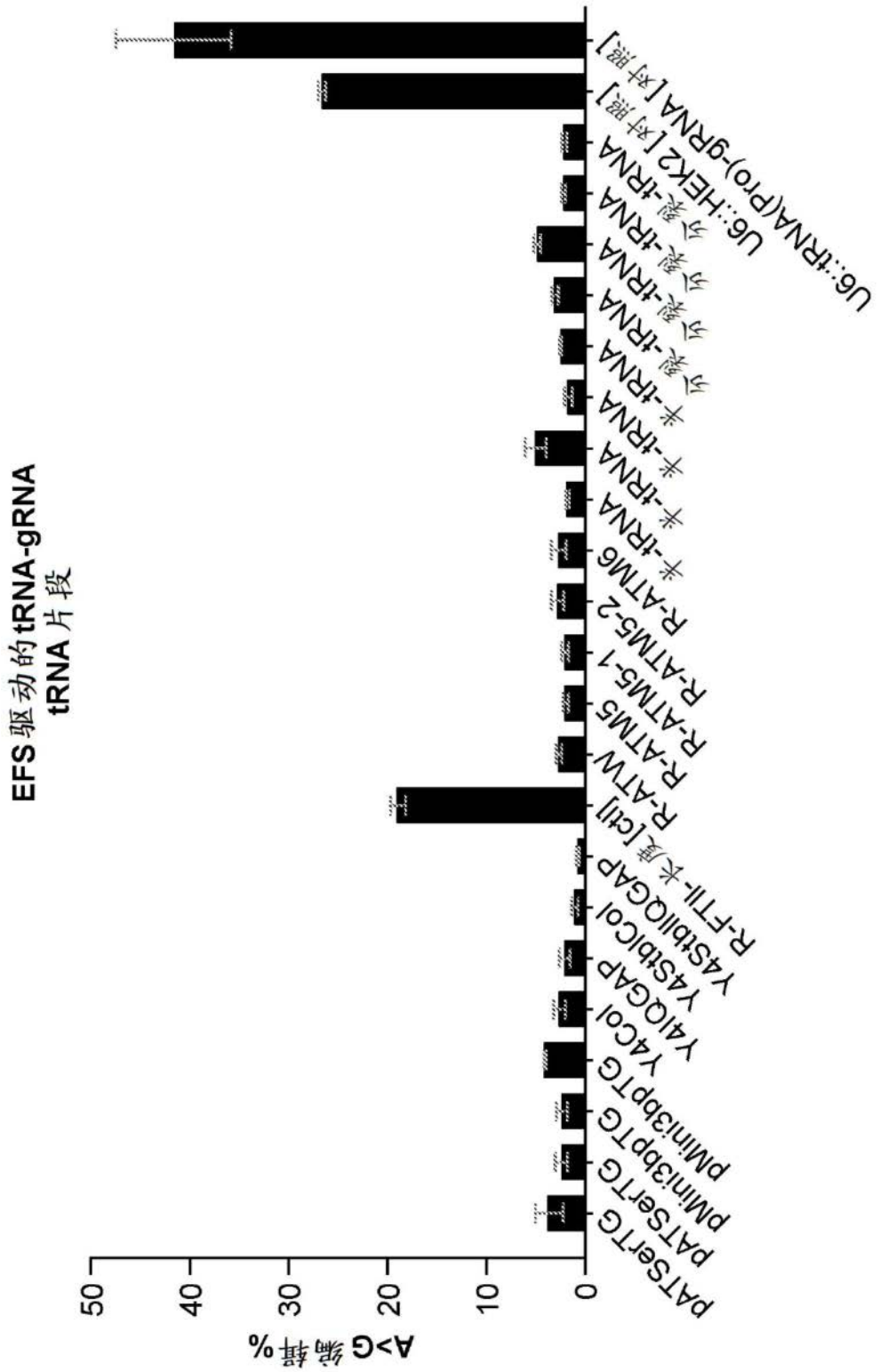

EFS驱动的tRNA-gRNA
tRNA片段
A>G编辑%
0
10
20
30
40
50
pATSerTG
pATSerTG
pMini3bpTG
pMini3bpTG
Y4CoI
Y4IGAP
Y4StbIICoI
R-FTII-末端[cII]
R-ATW
R-ATM5
R-ATM5-1
R-ATM5-2
R-ATM6
末-tRNA
末-tRNA
末-tRNA
内源-tRNA
内源-tRNA
内源-tRNA
内源-tRNA
U6::HEK2[对照]
U6::tRNA(Pro)-gRNA[对照]

图4D

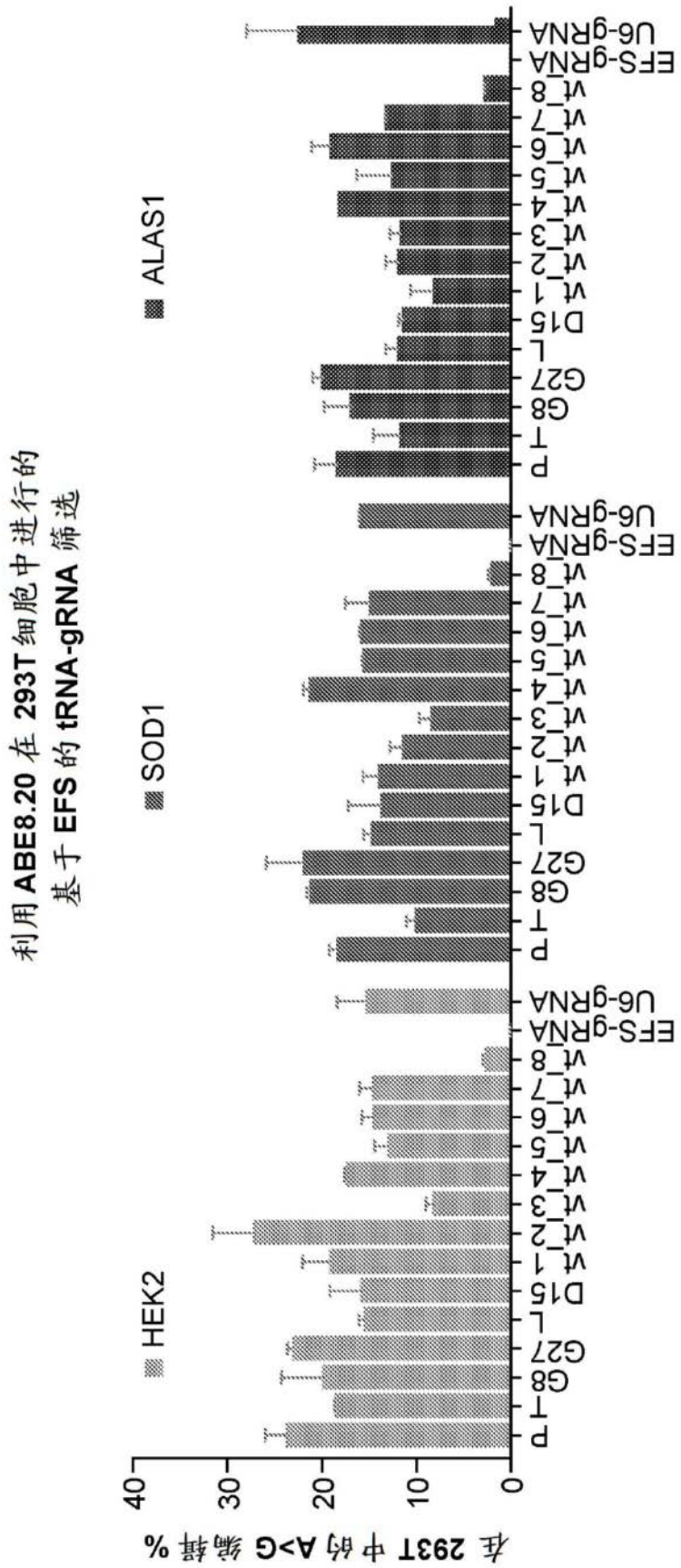

利用 ABE8.20 在 293T 细胞中进行的
基于 EFS 的 tRNA-gRNA 筛选
HEK2
SOD1
ALAS1
在 293T 中的 A>G 编辑 %
0
10
20
30
40
T
G8
G27
L
D15
vt_1
vt_2
vt_3
vt_4
vt_5
vt_6
vt_7
vt_8
EFS-gRNA
U6-gRNA

图5

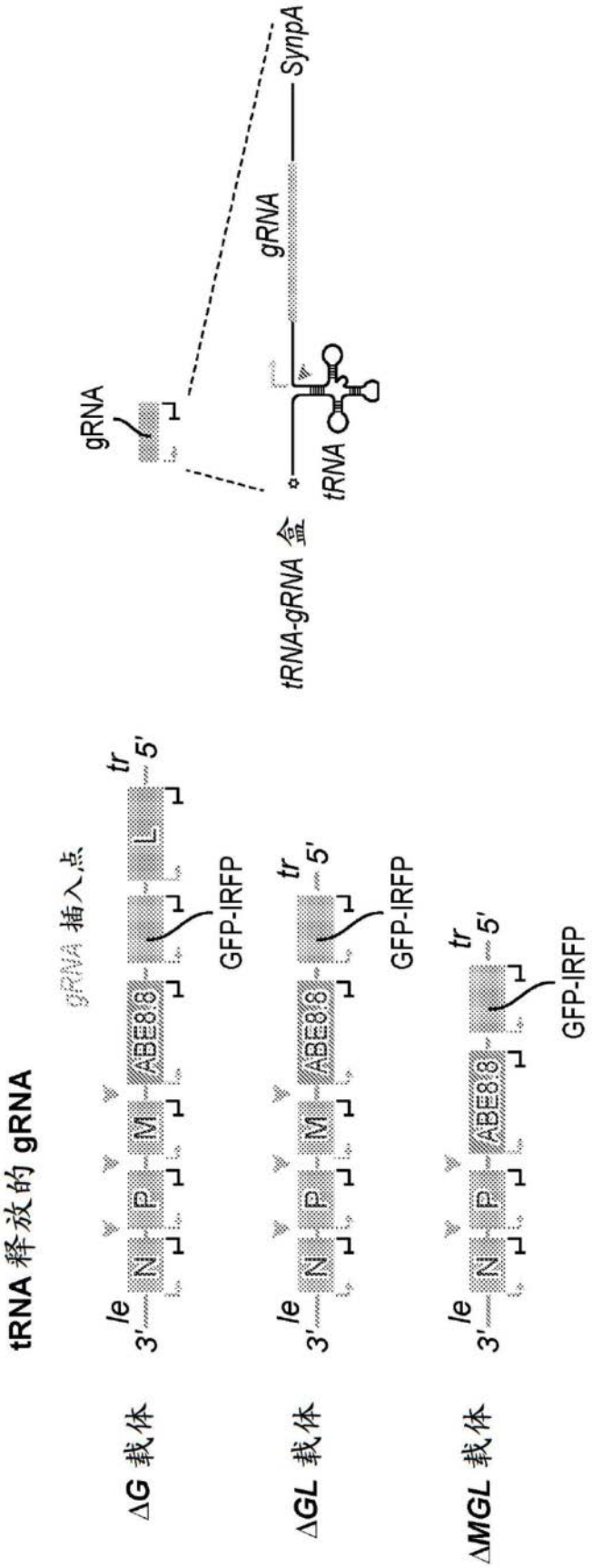

tRNA-gRNA 盒
gRNA
tRNA
gRNA
SynpA
tRNA 释放的 gRNA
gRNA 插入点
ΔG 载体
ΔGL 载体
ΔMGL 载体
le
3'
N
P
M
ABE8.8
L
tr
5'
GFP-IRFP

图6A

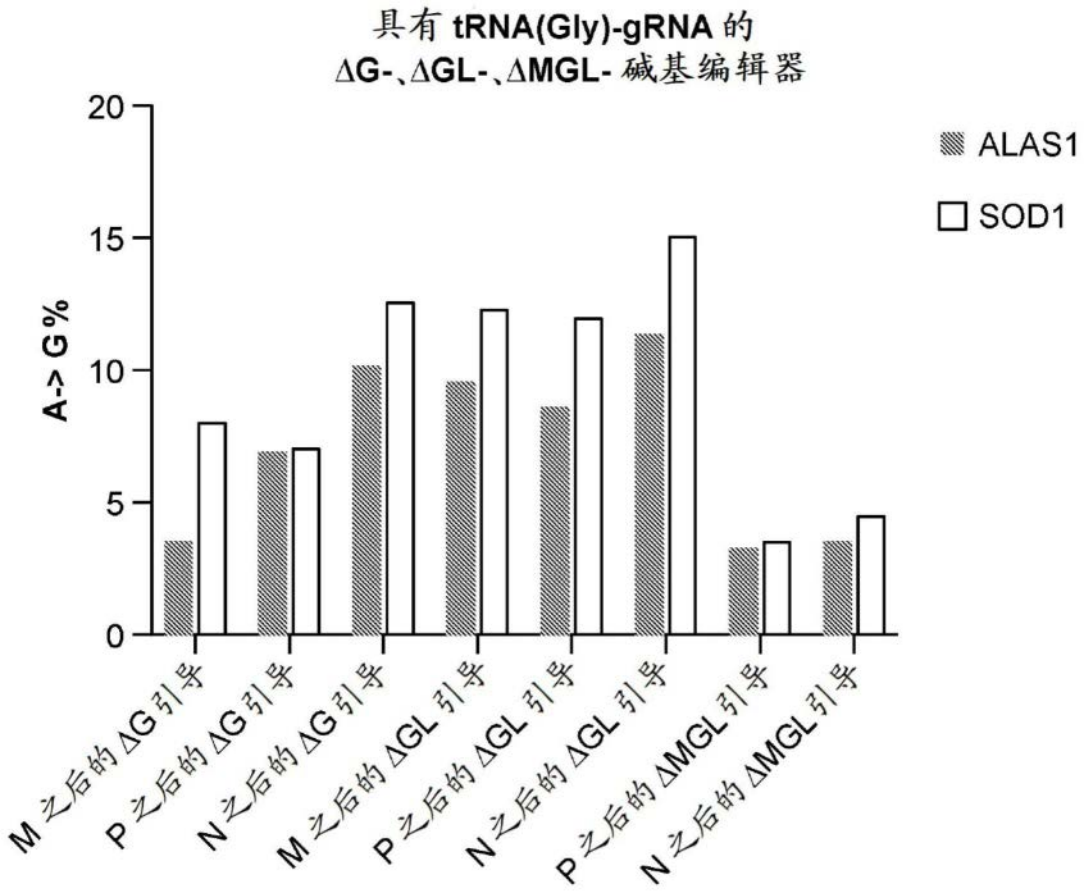

具有 tRNA(Gly)-gRNA 的
ΔG-、ΔGL-、ΔMGL- 碱基编辑器
A-> G %
20
15
10
5
0
ALAS1
SOD1
M 之后的 ΔG 引导
P 之后的 ΔG 引导
N 之后的 ΔG 引导
M 之后的 ΔGL 引导
P 之后的 ΔGL 引导
N 之后的 ΔGL 引导
P 之后的 ΔMGL 引导
N 之后的 ΔMGL 引导

图6B