



(12) 发明专利申请

(10) 申请公布号 CN 118339285 A

(43) 申请公布日 2024.07.12

(21) 申请号 202280079229.1

(22) 申请日 2022.12.09

(66) 本国优先权数据

PCT/CN2021/136761 2021.12.09 CN

(85) PCT国际申请进入国家阶段日

2024.05.29

(86) PCT国际申请的申请数据

PCT/CN2022/137920 2022.12.09

(87) PCT国际申请的公布数据

W02023/104185 EN 2023.06.15

(71) 申请人 北京干细胞与再生医学研究院

地址 100101 北京市朝阳区大屯路甲3号

申请人 中国科学院动物研究所

(72) 发明人 李伟 周琪 陈阳灿

(74) 专利代理机构 北京彩和律师事务所 11688

专利代理师 张红春

(51) Int. Cl.

C12N 9/22 (2006.01)

C12N 15/113 (2010.01)

C12N 15/864 (2006.01)

C12N 5/10 (2006.01)

C12Q 1/6816 (2018.01)

A61K 48/00 (2006.01)

A61K 38/46 (2006.01)

A61P 35/00 (2006.01)

A61P 9/00 (2006.01)

A61P 37/02 (2006.01)

A61P 3/00 (2006.01)

A61P 25/00 (2006.01)

A61P 27/02 (2006.01)

A61P 31/04 (2006.01)

A61P 31/12 (2006.01)

权利要求书5页 说明书77页
序列表(电子公布) 附图17页

(54) 发明名称

工程化Cas12b效应子蛋白及其使用方法

(57) 摘要

本申请提供了工程化Cas12b核酸酶或其衍生物,其包含一个或多个类型的突变,具有提高的活性(例如,基因编辑活性)或消除的核酸酶活性。还提供了工程化Cas12b效应子蛋白、工程化gRNA(例如,sgRNA或tracrRNA)、工程化的CRISPR-Cas12b系统,及其使用方法。

1. 一种工程化Cas12b核酸酶,相对于参考Cas12b核酸酶,其包含一种、两种或三种类型的突变,其中所述突变包含:

(1) 所述参考Cas12b核酸酶中与前间区邻近基序(PAM)相互作用的一个或多个氨基酸残基用带正电荷的氨基酸残基的取代;和/或

(2) 所述参考Cas12b核酸酶中参与打开DNA双链的一个或多个氨基酸残基用具有芳香环的氨基酸残基的取代;和/或

(3) 所述参考Cas12b核酸酶的RuvC结构域中与单链DNA底物相互作用的一个或多个氨基酸残基用带正电荷的氨基酸残基或疏水性氨基酸残基的取代。

2. 根据权利要求1所述的工程化Cas12b核酸酶,其中所述参考Cas12b核酸酶是野生型Cas12b核酸酶。

3. 根据权利要求1或2所述的工程化Cas12b核酸酶,其中所述参考Cas12b核酸酶包含SEQ ID NO:1的氨基酸序列。

4. 根据权利要求1-3中任一项所述的工程化Cas12b核酸酶,其中所述工程化Cas12b核酸酶包含所述参考Cas12b核酸酶中与PAM相互作用的一个或多个氨基酸残基用带正电荷的氨基酸残基的取代。

5. 根据权利要求4所述的工程化Cas12b核酸酶,其中所述与PAM相互作用的一个或多个氨基酸残基在三维结构中距离PAM在10埃以内。

6. 根据权利要求4或5所述的工程化Cas12b核酸酶,其中所述与PAM相互作用的一个或多个氨基酸残基位于以下一个或多个位置:116、123、130、132、144、145、153、173、222、395、400和475;并且其中所述氨基酸残基按照SEQ ID NO:1编号。

7. 根据权利要求4-6中任一项所述的工程化Cas12b核酸酶,其中所述与PAM相互作用的一个或多个氨基酸残基包含以下一个或多个氨基酸残基:D116和E475;并且其中所述氨基酸残基按照SEQ ID NO:1编号。

8. 根据权利要求4-7中任一项所述的工程化Cas12b核酸酶,其中所述带正电荷的氨基酸残基是R或K。

9. 根据权利要求4-8中任一项所述的工程化Cas12b核酸酶,其中所述参考Cas12b核酸酶中与PAM相互作用的一个或多个氨基酸残基用带正电荷的氨基酸残基的取代是下列一个或多个取代:D116R和E475R;其中所述氨基酸残基按照SEQ ID NO:1编号。

10. 根据权利要求1-9中任一项所述的工程化Cas12b核酸酶,其中所述工程化Cas12b核酸酶包含SEQ ID NO:2或3的氨基酸序列。

11. 根据权利要求1-10中任一项所述的工程化Cas12b核酸酶,其中所述工程化Cas12b核酸酶包含所述参考Cas12b核酸酶中参与打开所述DNA双链的一个或多个氨基酸残基用具有芳香环的氨基酸残基的取代。

12. 根据权利要求11所述的工程化Cas12b核酸酶,其中所述参与打开所述DNA双链的一个或多个氨基酸残基与相对于靶链的3'端的PAM中的最后一个碱基对相互作用。

13. 根据权利要求11或12所述的工程化Cas12b核酸酶,其中所述参与打开所述DNA双链的一个或多个氨基酸残基位于以下一个或多个位置:118和119;并且其中所述氨基酸残基按照SEQ ID NO:1编号。

14. 根据权利要求11-13中任一项所述的工程化Cas12b核酸酶,其中所述具有芳香环的

氨基酸残基为Y、F或W。

15. 根据权利要求11-14中任一项所述的工程化Cas12b核酸酶,其中所述参考Cas12b核酸酶中参与打开DNA双链的一个或多个氨基酸残基用具有芳香环的氨基酸残基的取代为Q119Y、Q119F或Q119W;其中所述氨基酸残基按照SEQ ID NO:1编号。

16. 根据权利要求1-3和11-15中任一项所述的工程化Cas12b核酸酶,其中所述工程化Cas12b核酸酶包含SEQ ID NO:4-6中任一个所述的氨基酸序列。

17. 根据权利要求1-16中任一项所述的工程化Cas12b核酸酶,其中所述工程化Cas12b核酸酶包含所述参考Cas12b核酸酶中位于所述RuvC结构域中并与所述单链DNA底物相互作用的一个或多个氨基酸残基用带正电荷的氨基酸残基或疏水性氨基酸残基的取代。

18. 根据权利要求17所述的工程化Cas12b核酸酶,其中所述RuvC结构域中与所述单链DNA底物相互作用的一个或多个氨基酸残基在三维结构中距离所述单链DNA底物10埃以内。

19. 根据权利要求17或18所述的工程化Cas12b核酸酶,其中所述RuvC结构域中与所述单链DNA底物相互作用的一个或多个氨基酸残基位于以下一个或多个位置:300、301、304、329、636、639、647、682、757、758、761、764、768、852、854、856、857、858、860、862、863、865、866、867、869、938、956、957、958、994、1093和1097;并且其中所述氨基酸残基按照SEQ ID NO:1编号。

20. 根据权利要求17-19中任一项所述的工程化Cas12b核酸酶,其中所述工程化Cas12b核酸酶包含以下一个或多个氨基酸残基用带正电荷的氨基酸残基的取代:E636、Q639、T647、Q682、I757、E758、E761、K768、Q854、N857、D858、N865、Q866、I994、Q1093和W1097;并且其中所述氨基酸残基按照SEQ ID NO:1编号。

21. 根据权利要求17-20中任一项所述的工程化Cas12b核酸酶,其中所述带正电荷的氨基酸残基是R或K。

22. 根据权利要求17-21中任一项所述的工程化Cas12b核酸酶,其中所述参考Cas12b核酸酶中位于所述RuvC结构域中并与所述单链DNA底物相互作用的一个或多个氨基酸残基的取代是以下一个或多个取代:E636R、Q639R、T647R、Q682R、I757R、E758R、E761R、Q854R、N857K、D858R、I994R、Q1093R和W1097R;并且其中所述氨基酸残基按照SEQ ID NO:1编号。

23. 根据权利要求1-3和17-22中任一项所述的工程化Cas12b核酸酶,其中所述工程化Cas12b核酸酶包含SEQ ID NO:7-13中任一个的氨基酸序列。

24. 根据权利要求17-19中任一项所述的工程化Cas12b核酸酶,其中所述工程化Cas12b核酸酶包含以下氨基酸残基中的一个或多个用疏水性氨基酸残基的取代:E758、E761、E863、N865、Q866、Q869、Q956和Q1093;并且其中所述氨基酸残基按照SEQ ID NO:1编号。

25. 根据权利要求17-19和24中任一项所述的工程化Cas12b核酸酶,其中所述疏水性氨基酸残基为W、Y、F或M。

26. 根据权利要求17-19、24和25中任一项所述的工程化Cas12b核酸酶,其中所述参考Cas12b核酸酶中位于所述RuvC结构域中并与所述单链DNA底物相互作用的一个或多个氨基酸残基的取代是以下一个或多个取代:N865W、N865Y、Q866M、Q869M、Q1093W和Q1093Y;并且其中所述氨基酸残基按照SEQ ID NO:1编号。

27. 根据权利要求1-3、17-19和24-26中任一项所述的工程化Cas12b核酸酶,其中所述工程化Cas12b核酸酶包含SEQ ID NO:14-19中任一个的氨基酸序列。

28. 权利要求1-3中任一项所述的工程化Cas12b核酸酶,其中所述工程化Cas12b核酸酶包含以下任一种取代或其组合:(1) D116R;(2) E475R;(3) Q119F+E475R;(4) Q119F+E475R+E758R;(5) Q119Y;(6) Q119F;(7) Q119W;(8) I757R;(9) E758R;(10) E761R;(11) K768R;(12) I757R+E758R;(13) I757R+E761R;(14) I757R+K768R;(15) E758R+E761R;(16) E758R+K768R;(17) E761R+K768R;(18) I757R+E758R+E761R;(19) I757R+E758R+K768R;(20) I757R+E761R+K768R;(21) E758R+E761R+K768R;(22) I757R+E758R+E761R+K768R;(23) Q866M;(24) Q869M;(25) Q866M+Q869M;(26) E636R;(27) Q854R;(28) N857K;(29) N865W;(30) N865Y;(31) Q1093W;(32) Q1093Y;和(33) D858R;并且其中所述氨基酸残基按照SEQ ID NO:1编号。

29. 根据权利要求1-3和28中任一项所述的工程化Cas12b核酸酶,其中所述工程化Cas12b核酸酶包含以下任一种取代或其组合:(1) Q866M+Q869M;(2) Q119F+E475R;和(3) Q119F+E475R+E758R;并且其中所述氨基酸残基按照SEQ ID NO:1编号。

30. 根据权利要求1-3、28和29中任一项所述的工程化Cas12b核酸酶,其中所述工程化Cas12b核酸酶包含SEQ ID NO:20-22中任一个的氨基酸序列。

31. 根据权利要求1-30中任一项所述的工程化Cas12b核酸酶,其还包含所述参考Cas12b核酸酶中的一个或多个突变,所述突变增加包含第855-859位氨基酸残基的柔性区的柔性;并且其中所述氨基酸残基按照SEQ ID NO:1编号。

32. 根据权利要求31所述的工程化Cas12b核酸酶,其中增加柔性的所述一个或多个突变包含N856G。

33. 一种工程化Cas12b效应子蛋白,其包含权利要求1-32中任一项所述的工程化Cas12b核酸酶或其功能衍生物。

34. 根据权利要求33所述的工程化Cas12b效应子蛋白,其中所述工程化Cas12b核酸酶或其功能衍生物具有酶促活性。

35. 根据权利要求33或34所述的工程化Cas12b效应子蛋白,其中所述工程化Cas12b效应子蛋白能够:i) 诱导DNA分子中的双链断裂;和/或ii) 诱导DNA分子中的单链断裂。

36. 根据权利要求33所述的工程化Cas12b效应子蛋白,其中所述工程化Cas12b效应子蛋白包含所述工程化Cas12b核酸酶的酶促失活突变体。

37. 根据权利要求36所述的工程化Cas12b效应子蛋白,其中所述工程化Cas12b核酸酶的酶促失活突变体包含选自下组的一个或多个氨基酸残基的取代:D570A、E848A、R785A、E848A、R911A和D977A;并且其中所述氨基酸残基按照SEQ ID NO:1编号。

38. 根据权利要求36或37所述的工程化Cas12b效应子蛋白,其中所述工程化Cas12b核酸酶的酶促失活突变体包含SEQ ID NO:79-81中任一个的氨基酸序列。

39. 根据权利要求33-38中任一项所述的工程化Cas12b效应子蛋白,其中所述工程化Cas12b效应子蛋白还包含与所述工程化Cas12b核酸酶或其功能衍生物融合的功能结构域。

40. 根据权利要求39所述的工程化Cas12b效应子蛋白,其中所述功能结构域选自下组:翻译启动结构域、转录阻遏物结构域、反式激活结构域、表观遗传修饰结构域、核碱基编辑结构域、逆转录酶结构域、报告子结构域和核酸酶结构域。

41. 根据权利要求40所述的工程化Cas12b效应子蛋白,其中所述转录阻遏物结构域是Krüppel相关盒(KRAB)结构域。

42. 根据权利要求41所述的工程化Cas12b效应子蛋白,其中所述KRAB结构域包含SEQ

ID NO:72的氨基酸序列。

43. 一种单指导RNA (sgRNA), 其包含SEQ ID NO:25-53中任一个的序列。

44. 一种工程化的CRISPR-Cas12b系统, 其包含:

(a) 权利要求1-32中任一项所述的工程化Cas12b核酸酶或权利要求33-42中任一项所述的工程化Cas12b效应子蛋白, 或编码其的核酸; 和

(b) 指导RNA (gRNA), 其包含与靶核酸的靶序列互补的指导序列、或编码所述gRNA的核酸,

其中, 所述工程化Cas12b核酸酶或所述工程化Cas12b效应子蛋白和所述gRNA能够形成特异性结合所述靶核酸的CRISPR复合物, 并诱导所述靶核酸的修饰。

45. 根据权利要求44所述的工程化的CRISPR-Cas12b系统, 其中所述gRNA包含crRNA和tracrRNA。

46. 根据权利要求44或45所述的工程化的CRISPR-Cas12b系统, 其中所述工程化的CRISPR-Cas12b系统包含编码多个crRNA的前体gRNA阵列。

47. 根据权利要求44所述的工程化的CRISPR-Cas12b系统, 其中所述gRNA是sgRNA。

48. 根据权利要求47所述的工程化的CRISPR-Cas12b系统, 其中所述sgRNA包含SEQ ID NO:23-53中任一个的序列。

49. 一种工程化的CRISPR-Cas12b系统, 其包含:

(a) Cas12b核酸酶或Cas12b效应子蛋白, 其包含SEQ ID NO:1-22和79-81中任一个的氨基酸序列、或编码其的核酸; 和

(b) gRNA, 其包含与靶核酸的靶序列互补的指导序列, 或编码所述gRNA的核酸, 其中所述gRNA包含工程化支架, 所述支架包含SEQ ID NO:25-53中任一个的序列;

其中, 所述Cas12b核酸酶或所述Cas12b效应子蛋白和所述gRNA能够形成与所述靶核酸特异性结合的CRISPR复合物, 并诱导所述靶核酸的修饰。

50. 根据权利要求49所述的工程化的CRISPR-Cas12b系统, 其中所述gRNA包含crRNA和tracrRNA, 并且其中所述tracrRNA包含工程化支架或其部分。

51. 根据权利要求49或50所述的工程化的CRISPR-Cas12b系统, 其中所述工程化的CRISPR-Cas12b系统包含编码多个crRNA的前体gRNA阵列。

52. 根据权利要求49所述的工程化的CRISPR-Cas12b系统, 其中所述gRNA是sgRNA。

53. 根据权利要求44-52中任一项所述的工程化的CRISPR-Cas12b系统, 其中所述工程化的CRISPR-Cas12b系统包含编码所述工程化Cas12b核酸酶、所述工程化Cas12b效应子蛋白、所述Cas12b核酸酶或所述Cas12b效应子蛋白的一个或多个载体。

54. 根据权利要求53所述的工程化的CRISPR-Cas12b系统, 其中所述一个或多个载体是腺相关病毒 (AAV) 载体。

55. 根据权利要求53或54所述的工程化的CRISPR-Cas12b系统, 其中所述一个或多个载体进一步编码所述gRNA。

56. 一种检测样品中靶核酸的方法, 其包括: (a) 使所述样品与权利要求44-55中任一项所述的工程化的CRISPR-Cas12b系统和标记的检测核酸接触, 其中所述gRNA包含与所述靶核酸的靶序列互补的指导序列, 并且其中所述标记的检测核酸是单链的并且不与所述gRNA的指导序列杂交; 和 (b) 测量由所述工程化Cas12b核酸酶或其效应子蛋白切割所述标记的

检测核酸而产生的可检测信号,由此检测所述靶核酸。

57.一种修饰包含靶序列的靶核酸的方法,其包括使所述靶核酸与权利要求44-55中任一项所述的工程化的CRISPR-Cas12b系统接触。

58.根据权利要求56或57所述的方法,其中所述方法在体外进行。

59.根据权利要求56或57所述的方法,其中所述靶核酸存在于细胞中。

60.根据权利要求59所述的方法,其中所述细胞是细菌细胞、酵母细胞、植物细胞或动物细胞。

61.根据权利要求56或57所述的方法,其中所述方法离体进行。

62.根据权利要求56或57所述的方法,其中所述方法在体内进行。

63.根据权利要求57-62中任一项所述的方法,其中通过所述工程化的CRISPR-Cas12b系统切割所述靶核酸或改变所述靶核酸中的所述靶序列。

64.根据权利要求57-62中任一项所述的方法,其中通过所述工程化的CRISPR-Cas12b系统改变所述靶核酸的表达。

65.根据权利要求56-64中任一项所述的方法,其中所述靶核酸是基因组DNA。

66.根据权利要求56-65中任一项所述的方法,其中所述靶序列与疾病或病况相关。

67.根据权利要求56-66中任一项所述的方法,其中所述工程化的CRISPR-Cas12b系统包含编码多个crRNA的前体指导RNA阵列,并且其中每个crRNA包含不同的指导序列。

68.一种治疗与个体的细胞中的靶核酸相关的疾病或病况的方法,其包括使用权利要求44-55中任一项所述的工程化的CRISPR-Cas12b系统修饰所述个体的细胞中的所述靶核酸,由此治疗所述疾病或病况。

69.根据权利要求68所述的方法,其中所述疾病或病况选自下组:癌症、心血管疾病、遗传性疾病、自身免疫性疾病、代谢疾病、神经退行性疾病、眼部疾病、细菌感染和病毒感染。

70.一种包含修饰的靶核酸的工程化的细胞,其中所述靶核酸已使用权利要求57-67中任一项所述的方法被修饰。

71.一种工程化的非人类动物,其包含一个或多个权利要求70所述的工程化的细胞。

工程化Cas12b效应子蛋白及其使用方法

[0001] 对相关申请的交叉引用

[0002] 本申请要求于2021年12月9日提交的国际专利申请号PCT/CN2021/136761的优先权,其内容通过引用以其整体并入本文。

[0003] 关于电子序列表

[0004] 电子序列表(253112000541SEQLIST.xml;大小:111,583字节;和创建日期:2022年11月22日)的内容通过引用以其整体并入本文。

技术领域

[0005] 本申请通常涉及生物技术领域。更具体地,本申请涉及具有改善的活性(例如,基因编辑活性)或消除的核酸酶活性的工程化Cas12b效应子蛋白和工程化gRNA支架,及其使用方法。

背景技术

[0006] 基因组编辑是基因组研究和多种应用中一项重要且有用的技术。多种系统均可用于基因组编辑,包含成簇的规律散布的短回文重复(CRISPR)-Cas系统、转录激活子样效应核酸酶(TALEN)系统和锌指核酸酶(ZFN)系统。

[0007] CRISPR-Cas系统是一种高效、经济的基因组编辑技术,广泛应用于从酵母、植物到斑马鱼和人类等一系列真核生物(由Van der Oost 2013,Science 339:768-770,and Charpentier and Doudna,2013,Nature 495:50-51综述)。CRISPR-Cas系统通过结合Cas效应子蛋白和CRISPR RNA(crRNA)在古菌和细菌中提供适应性免疫。迄今为止,根据该系统的突出功能和进化模块性,已经对包含六种类型(I-VI型)的CRISPR-Cas系统的两类(第1类和第2类)系统进行了表征。在第2类CRISPR-Cas系统中,II型Cas9系统和V-A/B/E/J型Cas12a/Cas12b/Cas12e/Cas12f/Cas12j系统已被用于基因组编辑,对生物医学研究具有巨大的前景。

[0008] 发明概述

[0009] 当前的CRISPR-Cas系统有多种局限性,包含基因编辑效率有限。本申请提供了用于横跨多种基因座进行有效基因组编辑的改进方法和系统。具体而言,本文提供了具有改善的酶活性的工程化Cas12b核酸酶、工程化Cas12b效应子蛋白、包含工程化支架的工程化gRNA(例如,sgRNA或tracrRNA),和如在基因编辑中使用所述工程化Cas12b效应子蛋白和/或工程化gRNA的方法。在一方面,本申请提供了工程化Cas12b核酸酶,其相对于参考Cas12b核酸酶包含一、二或三种类型的突变,其中所述突变包含:(1)参考Cas12b核酸酶中与前间区邻近基序(PAM)相互作用的一个或多个氨基酸残基用带正电荷的氨基酸残基(例如,R、H、K)的取代;和/或(2)参考Cas12b核酸酶中参与打开DNA双链(dsDNA)的一个或多个氨基酸残基用具有芳香环的氨基酸残基(例如,F、Y、W)的取代;和/或(3)参考Cas12b核酸酶的RuvC结构域中与单链DNA(ssDNA)底物相互作用的一个或多个氨基酸残基用带正电荷的氨基酸残基(例如,R、H、K)或疏水性氨基酸残基(例如,F、Y、W、M)的取代。在一些实施方案中,参考

Cas12b核酸酶是野生型Cas12b核酸酶。在一些实施方案中,参考Cas12b核酸酶是来自嗜酸脂环酸芽孢杆菌(AaCas12b)的Cas12b核酸酶。在一些实施方案中,参考Cas12b核酸酶包含SEQ ID NO:1的氨基酸序列。

[0010] 在根据上述任一种的工程化Cas12b核酸酶的一些实施方案中,所述工程化Cas12b核酸酶包含参考Cas12b核酸酶中与PAM相互作用的一个或多个氨基酸残基用带正电荷氨基酸残基的取代。在一些实施方案中,与PAM相互作用的一个或多个氨基酸残基在三维结构中距离PAM在10埃(例如,9、8、7、6、5、4、3、2、1或更小)以内。与PAM相互作用的一个或多个氨基酸残基位于以下一个或多个位置:116、123、130、132、144、145、153、173、222、395、400和475。在一些实施方案中,与PAM相互作用的一个或多个氨基酸残基包含以下一个或多个氨基酸残基:D116、K123、D130、D132、N144、K145、E153、D173、Q222、D395、N400和E475。在一些实施方案中,与PAM相互作用的一个或多个氨基酸残基包含以下一个或多个氨基酸残基:D116和E475。在一些实施方案中,带正电荷的氨基酸残基是R或K。在一些实施方案中,工程化Cas12b核酸酶包含以下一个或多个取代:D116R和E475R。在一些实施方案中,氨基酸残基按照SEQ ID NO:1编号。在一些实施方案中,工程化Cas12b核酸酶包含SEQ ID NO:2或3的氨基酸序列。

[0011] 在根据上述任一种的工程化Cas12b核酸酶的一些实施方案中,所述工程化Cas12b核酸酶包含参考Cas12b核酸酶中参与打开DNA双链的一个或多个氨基酸残基用具有芳香环的氨基酸残基的取代。在一些实施方案中,参与打开DNA双链的一个或多个氨基酸残基与相对于靶链的3'端的PAM中的最后一个碱基对相互作用。在一些实施方案中,参与打开DNA双链的一个或多个氨基酸残基位于以下一个或多个位置:118和119。在一些实施方案中,具有芳香环的氨基酸残基为Y、F或W。在一些实施方案中,参考Cas12b核酸酶中参与打开DNA双链的一个或多个氨基酸残基用具有芳香环的氨基酸残基的取代之为Q119Y、Q119F或Q119W。在一些实施方案中,氨基酸残基按照SEQ ID NO:1编号。在一些实施方案中,所述工程化Cas12b核酸酶包含SEQ ID NO:4-6中任一个的氨基酸序列。

[0012] 在根据上述任一种的工程化Cas12b核酸酶的一些实施方案中,所述工程化Cas12b核酸酶包含参考Cas12b核酸酶中位于RuvC结构域中并与单链DNA底物相互作用的一个或多个氨基酸残基用带正电荷的氨基酸残基或疏水性氨基酸残基的取代。在一些实施方案中,在RuvC结构域中并与单链DNA底物相互作用的一个或多个氨基酸残基在三维结构中距离所述单链DNA底物10埃(例如,9、8、7、6、5、4、3、2、1或更小)以内。在一些实施方案中,RuvC结构域中并与所述单链DNA底物相互作用的一个或多个氨基酸残基位于以下一个或多个位置:300、301、304、329、636、639、647、682、757、758、761、764、768、852、854、856、857、858、860、862、863、865、866、867、869、938、956、957、958、994、1093和1097。在一些实施方案中,在RuvC结构域中并与所述单链DNA底物相互作用的一个或多个氨基酸残基包含以下一个或多个氨基酸残基:D300、K301、E304、N329、E636、Q639、T647、Q682、I757、E758、E761、E764、K768、E852、Q854、N856、N857、D858、P860、S862、E863、N865、Q866、L867、Q869、E938、E956、G957、E958、I994、Q1093和W1097。在一些实施方案中,工程化Cas12b核酸酶包含以下一个或多个氨基酸残基用带正电荷的氨基酸残基的取代:E636、Q639、T647、Q682、I757、E758、E761、K768、Q854、N857、D858、N865、Q866、I994、Q1093和W1097。在一些实施方案中,带正电荷的氨基酸残基是R或K。在一些实施方案中,工程化Cas12b核酸酶包含以下一个或多个取

代:E636R、Q639R、T647R、Q682R、I757R、E758R、E761R、Q854R、N857K、D858R、I994R、Q1093R和W1097R。在一些实施方案中,所述工程化Cas12b核酸酶包含以下氨基酸残基中的一个或多个用疏水性氨基酸残基的取代:E758、E761、E863、N865、Q866、Q869、Q956和Q1093。在一些实施方案中,所述疏水性氨基酸残基为W、Y、F或M,如W、Y或M。在一些实施方案中,所述工程化Cas12b核酸酶包含以下一个或多个取代:N865W、N865Y、Q866M、Q869M、Q1093W和Q1093Y。在一些实施方案中,所述氨基酸残基按照SEQ ID NO:1编号。在一些实施方案中,所述工程化Cas12b核酸酶包含SEQ ID NO:7-19中任一个的氨基酸序列。

[0013] 在根据上述任一种的工程化Cas12b核酸酶的一些实施方案中,所述工程化Cas12b核酸酶包含以下任一种取代或其组合:(1) D116R; (2) E475R; (3) Q119F+E475R; (4) Q119F+E475R+E758R; (5) Q119Y; (6) Q119F; (7) Q119W; (8) I757R; (9) E758R; (10) E761R; (11) K768R; (12) I757R+E758R; (13) I757R+E761R; (14) I757R+K768R; (15) E758R+E761R; (16) E758R+K768R; (17) E761R+K768R; (18) I757R+E758R+E761R; (19) I757R+E758R+K768R; (20) I757R+E761R+K768R; (21) E758R+E761R+K768R; (22) I757R+E758R+E761R+K768R; (23) Q866M; (24) Q869M; (25) Q866M+Q869M; (26) E636R; (27) Q854R; (28) N857K; (29) N865W; (30) N865Y; (31) Q1093W; (32) Q1093Y; 和 (33) D858R; 并且其中氨基酸残基按照SEQ ID NO:1编号。在一些实施方案中,所述工程化Cas12b核酸酶包含以下任一种取代或其组合:(1) Q866M+Q869M; (2) Q119F+E475R; 和 (3) Q119F+E475R+E758R; 并且其中氨基酸残基按照SEQ ID NO:1编号。在一些实施方案中,所述工程化Cas12b核酸酶包含SEQ ID NO:20-22中任一个的氨基酸序列。

[0014] 在根据上述任一种的工程化Cas12b核酸酶的一些实施方案中,所述工程化Cas12b核酸酶包含与SEQ ID NO:1-22中的任一个具有至少约85% (例如,至少约88%、90%、95%、96%、97%、98%、99%或更高中的任一种) 序列同一性的氨基酸序列。在一些实施方案中,工程化Cas12b核酸酶包含(或由其组成,或基本由其组成) SEQ ID NO:2-22中任一个的氨基酸序列。

[0015] 在根据上述任一种的工程化Cas12b核酸酶的一些实施方案中,所述工程化Cas12b核酸酶还包含参考Cas12b核酸酶中的一个或多个突变,所述突变增加包含氨基酸残基855-859的柔性区的柔性。在一些实施方案中,增加柔性的一个或多个突变包含N856G。在一些实施方案中,氨基酸位置编号参考SEQ ID NO:1。

[0016] 本申请的一方面提供了一种工程化Cas12b核酸酶,其包含以下任一个或多个突变:(1) D116R; (2) E475R; (3) Q119F+E475R; (4) Q119F+E475R+E758R; (5) Q119Y; (6) Q119F; (7) Q119W; (8) I757R; (9) E758R; (10) E761R; (11) K768R; (12) I757R+E758R; (13) I757R+E761R; (14) I757R+K768R; (15) E758R+E761R; (16) E758R+K768R; (17) E761R+K768R; (18) I757R+E758R+E761R; (19) I757R+E758R+K768R; (20) I757R+E761R+K768R; (21) E758R+E761R+K768R; (22) I757R+E758R+E761R+K768R; (23) Q866M; (24) Q869M; (25) Q866M+Q869M; (26) E636R; (27) Q854R; (28) N857K; (29) N865W; (30) N865Y; (31) Q1093W; (32) Q1093Y; 和 (33) D858R; 并且其中所述氨基酸位置编号参考SEQ ID NO:1。在一些实施方案中,所述工程化Cas12b核酸酶包含以下任一种取代或其组合:(1) Q866M+Q869M; (2) Q119F+E475R; 和 (3) Q119F+E475R+E758R; 并且其中氨基酸残基按照SEQ ID NO:1编号。在一些实施方案中,工程化Cas12b核酸酶包含取代Q119F+E475R+E758R; 并且其中氨基酸残基按照SEQ ID NO:1编号。

[0017] 本申请的一方面提供了一种工程化Cas12b核酸酶,其与SEQ ID NO:2-22中的任一个具有至少约85% (例如,至少约88%、90%、95%、96%、97%、98%、99%或更高中的任一个)序列同一性,或包含(或由其组成、或基本由其组成)SEQ ID NO:2-22中任一个的氨基酸序列。

[0018] 本申请的一方面提供了一种工程化Cas12b效应子蛋白,其包含根据上述任一种的工程化Cas12b核酸酶的工程化Cas12b核酸酶或其变体或功能衍生物。在一些实施方案中,工程化Cas12b核酸酶或其功能衍生物具有酶促活性。在一些实施方案中,工程化Cas12b效应子蛋白能够诱导DNA分子中的双链断裂。在一些实施方案中,工程化Cas12b效应子蛋白能够诱导DNA分子中的单链断裂。在一些实施方案中,工程化Cas12b效应子蛋白包含工程化Cas12b核酸酶的酶促失活突变体。在一些实施方案中,工程化Cas12b核酸酶的酶促失活突变体包含选自下组的一个或多个氨基酸残基的取代:D570A、E848A、R785A、E848A、R911A和D977A;并且其中氨基酸残基按照SEQ ID NO:1编号。在一些实施方案中,工程化Cas12b核酸酶的酶促失活突变体包含(或由其组成,或基本由其组成)SEQ ID NO:79-81中任一个的氨基酸序列、或与SEQ ID NO:79-81中任一个具有至少约85% (例如,至少约88%、90%、95%、96%、97%、98%、99%或更多中的任一个)序列同一性的其变体。

[0019] 在根据上述任一种的工程化Cas12b效应子蛋白的一些实施方案中,工程化Cas12b效应子蛋白还包含与工程化Cas12b核酸酶或其功能衍生物融合的功能结构域。在一些实施方案中,所述功能结构域选自下组:翻译启动结构域、转录阻遏物结构域、反式激活结构域、表观遗传修饰结构域、核碱基编辑结构域、逆转录酶结构域、报告子结构域和核酸酶结构域。在一些实施方案中,转录阻遏物结构域是Krüppel相关盒(KRAB)结构域,如包含SEQ ID NO:72的氨基酸序列。

[0020] 在根据上述任一种工程化Cas12b效应子蛋白的一些实施方案中,所述工程化Cas12b效应子蛋白包含第一多肽和第二多肽,所述第一多肽包含工程化Cas核酸酶或其功能衍生物的N末端部分,所述第二多肽包含工程化Cas核酸酶或其功能衍生物的C末端部分,其中所述第一多肽和第二多肽能够在包含指导序列的指导RNA的存在下相互缔合以形成CRISPR复合物,所述CRISPR复合物特异性结合包含与指导序列互补的靶序列的靶核酸。在一些实施方案中,所述工程化Cas12b效应子蛋白包含第一多肽和第二多肽,其中第一多肽包含工程化Cas12b核酸酶或其功能衍生物的N末端氨基酸残基1至X,其中第二多肽包含工程化Cas12b核酸酶或其功能衍生物的第X+1残基至其C末端,其中第一多肽和第二多肽能够在含有指导序列的指导RNA的存在下相互缔合以形成特异性结合靶核酸的CRISPR复合物,其中所述靶核酸包含与指导序列互补的靶序列。在一些实施方案中,所述第一多肽和第二多肽各自包含二聚化结构域。在一些实施方案中,第一二聚化结构域和第二二聚化结构域在诱导剂存在下相互缔合。在一些实施方案中,第一多肽和第二多肽不包含二聚化结构域。

[0021] 本申请的另一方面提供了一种单指导RNA(sgRNA),其包含SEQ ID NO:25-53中任一个的序列。

[0022] 本申请的另一方面提供了一种工程化CRISPR-Cas12b系统,其包含:(a)根据上述任一种工程化Cas12b核酸酶的工程化Cas12b核酸酶、或根据上述任一种工程化Cas12b效应子蛋白的工程化Cas12b效应子蛋白、或编码其的核酸;和(b)包含与靶核酸的靶序列互补的指导序列的指导RNA,或编码所述指导RNA的核酸,其中所述工程化Cas12b核酸酶或工程化

Cas12b效应子蛋白和指导RNA能够形成特异性结合包含靶序列的靶核酸的CRISPR复合物,并诱导对靶核酸的修饰。在一些实施方案中,所述指导RNA包含crRNA和tracrRNA。在一些实施方案中,所述工程化的CRISPR-Cas12b系统包含编码多个crRNA的前体指导RNA阵列。在一些实施方案中,所述指导RNA是单指导RNA (sgRNA)。在一些实施方案中,sgRNA包含SEQ ID NO:23-53中任一个的序列。在一些实施方案中,工程化的CRISPR-Cas12b系统包含编码工程化Cas12b核酸酶或工程化Cas12b效应子蛋白的一个或多个载体。在一些实施方案中,一个或多个载体是腺相关病毒(AAV)载体。在一些实施方案中,AAV载体进一步编码指导RNA。

[0023] 本申请的另一方面提供了一种工程化CRISPR-Cas12b系统,其包含:(a)根据上述任一种工程化Cas12b核酸酶的工程化Cas12b核酸酶、或根据上述任一种工程化Cas12b效应子蛋白的工程化Cas12b效应子蛋白、包含SEQ ID NO:1-22和79-81中任一个的氨基酸序列的Cas12b核酸酶或其效应子蛋白、或编码其的核酸;和(b)包含与靶核酸的靶序列互补的指导序列的gRNA、或编码gRNA的核酸,其中所述gRNA包含工程化支架,所述工程化支架包含SEQ ID NO:25-53中任一个的序列;其中所述Cas12b核酸酶(例如,工程化的)或其效应子蛋白和gRNA能够形成特异性结合靶核酸的CRISPR复合物,并诱导靶核酸的修饰。在一些实施方案中,所述gRNA包含crRNA和tracrRNA,并且其中所述tracrRNA包含工程化支架或其部分。在一些实施方案中,所述工程化CRISPR-Cas12b系统包含编码多个crRNA的前体gRNA阵列。在一些实施方案中,gRNA是sgRNA。在一些实施方案中,所述工程化CRISPR-Cas12b系统包含编码工程化Cas12b核酸酶或其效应子蛋白的一个或多个载体、或Cas12b核酸酶或其效应子蛋白。在一些实施方案中,所述一个或多个载体是AAV载体。在一些实施方案中,所述一个或多个载体进一步编码gRNA。

[0024] 本申请的一方面提供了一种检测样品中的靶核酸的方法,其包括:(a)使样品与根据上述任一个的工程化CRISPR-Cas12b系统和标记的检测核酸接触,其中,所述gRNA包含与靶核酸的靶序列互补的指导序列,并且其中,所述标记的检测核酸是单链的并且不与指导RNA的指导序列杂交;和(b)测量由工程化Cas12b核酸酶或其效应子蛋白切割标记的检测核酸产生的可检测信号,由此检测靶核酸。

[0025] 本申请的一方面提供了一种修饰包含靶序列的靶核酸的方法,其包括将靶核酸与根据上述任一个工程化CRISPR-Cas12b系统的工程化CRISPR-Cas12b系统接触。在一些实施方案中,所述方法是在体外进行的。在一些实施方案中,靶核酸存在于细胞中。在一些实施方案中,细胞是细菌细胞、酵母细胞、植物细胞或动物细胞(例如,哺乳动物细胞)。在一些实施方案中,所述方法是离体进行的。在一些实施方案中,所述方法是在体内进行的。在一些实施方案中,所述靶核酸被裂解。在一些实施方案中,所述靶核酸中的靶序列被工程化CRISPR-Cas12b系统改变。在一些实施方案中,靶核酸的表达被工程化CRISPR-Cas12b系统改变。在一些实施方案中,靶核酸是基因组DNA。在一些实施方案中,靶序列与疾病或病况相关。在一些实施方案中,工程化CRISPR-Cas12b系统包含编码多个crRNA的前体指导RNA阵列,其中每个crRNA包含一个不同的指导序列。

[0026] 本申请的另一方面提供了一种治疗个体的细胞中与靶核酸相关的疾病或病况的方法,包括使用根据上述任一个工程化CRISPR-Cas12b系统的工程化CRISPR-Cas12b系统修饰所述个体细胞中的靶核酸,由此治疗所述疾病或病况。在一些实施方案中,所述疾病或病况选自下组:癌症、心血管疾病、遗传性疾病、自身免疫性疾病、代谢疾病、神经退行性疾病、

眼部疾病、细菌感染和病毒感染。

[0027] 还提供了包含修饰的靶核酸的工程化细胞,其中已使用根据上述修饰靶核酸的任一种方法对靶核酸进行了修饰。还提供了包含其一个或多个工程化细胞的工程化非人动物。

[0028] 还提供了用于上述任一种方法的组合物、试剂盒和制品。

[0029] 应当理解,为了清楚起见,在单独的实施方案的上下文中描述的本公开的某些特征也可以在单个实施方案中组合提供。相反,为简洁起见,在单个实施方案的上下文中描述的本公开的多种特征也可以单独提供或以任何合适的子组合提供。本公开内容具体地涵盖了涉及特定方法步骤、试剂或条件或组合物组分的实施方案的所有组合,并且在本文中公开,就像单独且明确地公开了每一个组合一样。

附图说明

[0030] 图1显示了示例性AaCas12b变体的基因编辑效率(%插入/缺失),其中野生型AaCas12b中与PAM相互作用的氨基酸残基被R取代。与野生型(WT)AaCas12b相比,具有D116R或E475R取代的AaCas12b变体显示出提高的编辑效率。

[0031] 图2显示了示例性AaCas12b变体的基因编辑效率,其中野生型AaCas12b中参与打开DNA双链的氨基酸残基被芳香族氨基酸残基取代。与WT AaCas12b相比,具有Q119Y、Q119F或Q119W取代的AaCas12b变体显示出提高的基因编辑效率。

[0032] 图3显示了示例性AaCas12b变体的基因编辑效率,其中野生型AaCas12b中位于RuvC结构域中并与单链DNA底物相互作用的氨基酸残基被R取代。

[0033] 图4A-4B显示了示例性AaCas12b变体的基因编辑效率,其中野生型AaCas12b中位于RuvC结构域中并与单链DNA相互作用的氨基酸残基被赖氨酸(K)或精氨酸(R)残基取代。图4A显示了基因组位点CCR5-3处的编辑效率,而图4B显示了基因组位点RNF2-5处的编辑效率。与WT AaCas12b相比,具有E636R、I757R、E758R、E761R、Q854R、D858R、E758K、I994R、N857K或D858K取代的AaCas12b变体表现出最大提高的基因编辑效率。

[0034] 图5显示了示例性AaCas12b变体的基因编辑效率,其中野生型AaCas12b中位于RuvC结构域中并与单链DNA底物相互作用的氨基酸残基被疏水性氨基酸残基W、Y、F或M取代。与WT AaCas12b相比,具有N865W、N865Y、Q866M、Q869M、Q1093W或Q1093Y取代的AaCas12b变体显示出最大提高的基因编辑效率。

[0035] 图6显示了与WT AaCas12b相比,具有组合突变的示例性AaCas12b变体的基因编辑效率。

[0036] 图7显示与WT AaCas12b及相应的单突变体相比,AaCas12b变体Q119F+E475R+E758R具有显著提高的基因编辑效率。

[0037] 图8显示了Cas12b蛋白的氨基酸序列比对,包括嗜酸脂环酸芽孢杆菌(*Alicyclobacillus acidiphilus*)Cas12b(AaCas12b)(SEQ ID NO:1),卡克加文脂环酸芽孢杆菌(*Alicyclobacillus kakegawensis*)Cas12b(AkCas12b)(SEQ ID NO:54),大孢脂环酸芽孢杆菌(*Alicyclobacillus macrosporangiidus*)Cas12b(AmCas12b)(SEQ ID NO:55),芽孢杆菌属(*Bacillus* sp.)V3-13 Cas12b(Bs3Cas12b)(SEQ ID NO:56),芽孢杆菌(*Bacillus*)Cas12b(BsCas12b)(SEQ ID NO:57),沉积物莱西氏菌(*Laceyella sediminis*)

Cas12b (LsCas12b) (SEQ ID NO:58), 外村尚芽孢杆菌 (*Bacillus hisashii*) Cas12b (BhCas12b) (SEQ ID NO:59), 和螺旋体门细菌 (*Spirochaetes bacterium*) Cas12b (SbCas12b) (SEQ ID NO:60)。本文描述的基于AaCas12b的取代可以在本文描述的任何Cas12b直系同源物中的相应氨基酸位置做出。

[0038] 图9显示具有工程化支架的sgRNA大大提高了AaCas12b变体Q119F+E475R+E758R的基因编辑效率。具有AaCas12b Aa-sg支架或AacCas12b sgRNA支架 (V0) 的sgRNA作为对照。

[0039] 图10A是编码在CMV启动子控制下的AaCas12b变体Q119F+E475R+E758R+D570A和在U6启动子控制下的sgRNA的示例性构建体的示意图。图10B显示了T7EI测定结果, 作为AaCas12b (Q119F+E475R+E758R) 和AaCas12b (Q119F+E475R+E758R+D570A) 的核酸酶活性的量度。sgRNA1和sgRNA2特异性识别HBG1/2中的靶位点。不靶向任何HBG1/2序列的对照sgRNA作为阴性对照。

[0040] 图11A是编码在CMV启动子控制下的AaCas12b变体Q119F+E475R+E758R+D570A+E848A或Q119F+E475R+E758R+D570A+D977A和在U6启动子控制下的sgRNA的示例性构建体的示意图。图11B显示了T7EI测定结果作为由sgRNA1和sgRNA2介导的AaCas12b (Q119F+E475R+E758R)、AaCas12b (Q119F+E475R+E758R+D570A+E848A) 和AaCas12b (Q119F+E475R+E758R+D570A+D977A) 特异性识别HBG1/2中的靶位点的核酸酶活性的量度。不靶向任何HBG1/2序列的对照sgRNA作为阴性对照。

[0041] 图12A是编码在CMV启动子的控制下与KRAB融合的AaCas12b (Q119F+E475R+E758R+D570A+D977A) 和在U6启动子控制下的sgRNA的示例性构建体的示意图。图12B显示不同sgRNA介导的靶向SCN9A基因不同位点的AaCas12b (Q119F+E475R+E758R+D570A+D977A) - KRAB融合蛋白转染的小鼠N2a细胞中小鼠Nav1.7的mRNA相对水平。无sgRNA转染作为对照。

[0042] 发明详述

[0043] 本申请通过引入相对于参考Cas12b核酸酶的一种、两种或三种类型的突变, 提供了具有增强的酶活性 (如基因编辑活性) 的工程化Cas12b核酸酶。还提供了具有降低或消除的核酸酶活性的工程化Cas12b核酸酶或其效应子蛋白 (如, dCas12b)。还提供了具有工程化支架序列的工程化指导RNA (gRNA), 当与Cas12b核酸酶 (野生型或工程化的) 一起使用时, 可以增加Cas12b酶活性 (例如, 基因编辑活性)。还提供了工程化Cas12b效应子蛋白、使用所述工程化Cas12b核酸酶或工程化Cas12b效应子蛋白的方法、和/或工程化gRNA。

[0044] I、定义

[0045] 除非另有定义, 本文中使用的所有技术和科学术语均具有与本公开内容所属领域的普通技术人员通常理解的相同的含义。

[0046] 如本文所用, 术语“Cas12b蛋白”以其最广义使用, 包含亲本或参考Cas12b蛋白 (例如, 包含SEQ ID NO:1的AaCas12b蛋白)、其衍生物或变体 (例如, 工程化Cas12b、dCas12b或工程化Cas12b效应子蛋白)、和其功能片段, 如寡核苷酸结合片段。

[0047] 如本文所用, “效应子蛋白”是指具有活性的蛋白质, 所述活性例如位点特异性结合活性、单链DNA切割或编辑活性、双链DNA切割或编辑活性、单链RNA切割或编辑活性、或转录调控活性。

[0048] 如本文所用, “指导RNA”和“gRNA”在此可互换使用, 指能够与Cas12b核酸酶或效应子蛋白和靶核酸 (例如, 双螺旋DNA) 形成复合物的RNA。指导RNA可以包含单个RNA分子、或通

过两个或更多个RNA分子中的互补区域杂交而彼此缔合的两个或多个RNA分子。当与双RNA指导的Cas核酸酶(如Cas12b)结合使用时,指导RNA包含crRNA和tracrRNA、或单指导RNA(sgRNA)。“crRNA”或“CRISPR RNA”包含与靶核酸(例如,双螺旋DNA)的靶序列具有足够互补性的指导序列,其指导CRISPR复合物与靶核酸的序列特异性结合。所述“tracrRNA”或“反式激活CRISPR RNA”与crRNA部分互补且碱基配对,可在crRNA成熟过程中发挥作用。“单指导RNA”或“sgRNA”是一种工程化指导RNA,其crRNA和tracrRNA在单个分子中彼此融合。

[0049] 如本文所用,术语“CRISPR阵列”是指包含CRISPR重复序列和间隔区的核酸(例如,DNA)片段,其从第一个CRISPR重复序列的第一个核苷酸开始,到最后一个(末端)CRISPR重复序列的最后一个核苷酸结束。一般而言,CRISPR阵列中的每个间隔区位于两个重复序列之间。如本文所用,术语“CRISPR重复序列”或“CRISPR直接重复序列”或“直接重复序列”是指在CRISPR阵列中表现出很少或没有序列变化的多个短直接重复序列。适当地,V-I直接重复序列可以形成茎环结构。

[0050] 如本文所用,“供体模板核酸”或“供体模板”可互换使用,是指在本文所述的CRISPR酶改变靶核酸后,可被一个或多个细胞蛋白质用来改变靶核酸结构的核酸分子。在一些实例中,供体模板核酸是双链核酸。在一些实例中,供体模板核酸是单链核酸。在一些实例中,供体模板核酸是线性的。在一些实例中,供体模板核酸是环状的(例如,质粒)。在一些实例中,供体模板核酸是外源核酸分子。在一些实例中,供体模板核酸是内源核酸分子(例如,染色体)。

[0051] 术语“核酸”、“多核苷酸”和“核苷酸序列”可互换使用,指任何长度的核苷酸的聚合形式,包含脱氧核糖核苷酸、核糖核苷酸、它们的组合及其类似物。“寡核苷酸”和“寡”可互换使用,指的是长度不超过约50个核苷酸的短多核苷酸。

[0052] 如本文所用,“互补性”是指一个核酸通过传统的沃森-克里克碱基配对与另一个核酸形成氢键的能力。互补性百分比表示核酸分子中能够与第二个核酸形成氢键(即,沃森-克里克碱基配对)的残基的百分比(例如,10个中的约5、6、7、8、9、10个,分别为约50%、60%、70%、80%、90%和100%互补)。“完美互补”是指一个核酸序列的所有连续残基与第二个核酸序列中相同数量的连续残基形成氢键。本文所用的“基本上互补”是指在约40、50、60、70、80、100、150、200、250个或更多个核苷酸的区域内,互补程度至少为约70%、75%、80%、85%、90%、95%、97%、98%、99%或100%中的任一个,或指两个核酸在严格条件下杂交。

[0053] 本文所用的杂交“严格条件”是指与靶序列互补的核酸主要与靶序列杂交而基本不与非靶序列杂交的条件。严格条件通常依赖于序列,并且根据多种因素而变化。一般而言,序列越长,该序列与其靶序列特异性杂交的温度越高。严格条件的非限制性实例在Tijssen(1993),Laboratory Techniques In Biochemistry And Molecular Biology-Hybridization With Nucleic Acid Probes Part I,Second Chapter“Overview of principles of hybridization and the strategy of nucleic acid probe assay,”Elsevier,N,Y.中详细描述。

[0054] “杂交”是指一个或多个多核苷酸反应形成经由核苷酸残基的碱基之间的氢键来稳定的复合物的反应。氢键可以通过Watson Crick碱基配对、Hoogsteen结合或任何其他序列特定方式发生。能够与给定序列杂交的序列被称为给定序列的“互补序列”。

[0055] 关于核酸序列的“序列同一性百分比(%)”定义为在通过允许间隙(如果必要)以实现最大序列同一性百分比来对序列进行比对后,候选序列中与特定核酸序列中的核苷酸相同的核苷酸的百分比。对于肽、多肽或蛋白质序列而言,“序列同一性百分比(%)”是指在通过允许间隙(如有必要)以实现最大序列同一性百分比来对序列进行比对后,候选序列中与特定肽或氨基酸序列中的氨基酸残基相同替换的氨基酸残基的百分比。为了确定氨基酸序列同一性百分比而进行的比对可以通过本领域技术人员已知的多种方式实现,例如,使用可公开获得的计算机软件,如BLAST、BLAST-2、ALIGN或MEGALIGN™(DNASTAR)软件。本领域技术人员可以确定用于测量比对的适当参数,包括在所比较序列的全长上实现最大比对所需的任何算法。

[0056] 术语“多肽”和“肽”在本文中可互换使用,指任意长度的氨基酸聚合物。所述聚合物可以是直链或支链的,它可以包含修饰的氨基酸,也可以被非氨基酸中断。一个蛋白质可具有一个或多个多肽。术语还涵盖已被修饰的氨基酸聚合物;例如,二硫键形成、糖基化、脂化、乙酰化、磷酸化或任何其他操作,如与标记组分缀合。

[0057] 如本文所用,“变体”是指与参考多核苷酸或多肽分别不同但保留基本特性的多核苷酸或多肽。多核苷酸的典型变体在核酸序列上与另一个参考多核苷酸不同。变体的核酸序列的变化可或不可改变参考多核苷酸编码的多肽的氨基酸序列。如下所述,核苷酸变化可能导致参考序列编码的多肽中的氨基酸取代、添加、缺失、融合和截短。多肽的典型变体在氨基酸序列上与另一个参考多肽不同。一般而言,差异是有限的,因此参考多肽和变体的序列总体上非常相似,并且在许多区域是相同的。变体和参考多肽在氨基酸序列上可以通过一个或多个取代、添加、缺失以任意组合而不同。取代或插入的氨基酸残基可以是或可以不是由遗传密码编码的。多核苷酸或多肽的变体可以是天然存在的,如等位基因变体,也可以是未知的天然存在的变体。多核苷酸和多肽的非天然存在的变体可以通过诱变技术、直接合成和本领域技术人员已知的其他重组方法来制成。

[0058] 如本文所用,术语“野生型”具有本领域技术人员通常理解的含义,是指生物体、菌株、基因的典型形式,或者指其在自然界中存在时与突变体或变体相区别的特征。它可以从自然界来源中分离出来,并且不会被刻意修饰。

[0059] 如本文所用,术语“非天然存在的”或“工程化的”可互换使用,并且指的是人工参与。当这些术语用于描述核酸分子或多肽时,其意思是核酸分子或多肽至少基本上脱离了自然界中与其相关的或在自然界中发现与其相关的至少一种其它组分。

[0060] 如本文所用,术语“直系同源物(orthologue)”或“直系同源基因(ortholog)”具有本领域普通技术人员通常理解的含义。作为进一步的指导,本文中所提及的蛋白质的“直系同源物”是指属于不同物种的蛋白质,其与作为其直系同源物的蛋白质具有相同或相似的功能。

[0061] 如本文所用,术语“同一性”用于指两个多肽之间或两个核酸之间的序列匹配。当被比较的两个序列中的某个位置由同一种碱基或氨基酸单体亚基占据时(例如,两个DNA分子中的每个的位置由腺嘌呤占据,或两个多肽中的每个位置由赖氨酸占据,则每个分子在该位置上都是相同的。两个序列之间的“同一性百分比”是两个序列共享的匹配位置数除以要比较的位置数 $\times 100$ 的函数。例如,如果两个序列的10个位置中有6个匹配,则这两个序列具有60%的同一性。例如,DNA序列CTGACT和CAGGTT共享50%的同一性(总共6个位置中有3

个匹配)。一般而言,当两个序列对齐以产生最大同一性时,进行比较。这样的比对可以通过例如Needleman et al. (1970) J.Mol.Biol.48:443-453的方法实现,所述方法可以通过计算机程序(如Align程序(DNAstar, Inc.))方便地进行。也可以使用集成到ALIGN程序(2.0版)中的E.Meyers和W.Miller的算法(Comput. Appl Biosci., 4:11-17 (1988)),使用PAM 120权重残基表。使用间隙长度罚分12和间隙罚分4来确定两个氨基酸序列之间的同一性百分比。此外,还可以使用GCG软件包(可在www.gcg.com上获得)中集成的GAP程序中的Needleman和Wunsch(J MoI Biol.48:444-453(1970))算法,采用Blossum 62矩阵或PAM250矩阵和间隙权重16、14、12、10、8、6或4和长度权重1、2、3、4、5或6来确定两个氨基酸序列之间的同一性百分比。

[0062] 如本文所用,“细胞”不仅指特定的单个细胞,还指该细胞的后代或潜在后代。由于突变或环境影响,某些修饰可发生在后续世代中,因此这样的后代实际上可能与亲本细胞不相同,但仍包括在此处所用术语的范围内。

[0063] 如本文所用,术语“转导”和“转染”包含本领域已知的所有方法,使用传染试剂(如病毒)或其他手段将DNA引入细胞以表达感兴趣的蛋白质或分子。除病毒或病毒样试剂外,还有基于化学的转染方法,如使用磷酸钙、树枝状聚合物、脂质体或阳离子聚合物(例如,DEAE-葡聚糖或聚乙烯亚胺)的方法;非化学方法,如电穿孔、细胞挤压、声孔、光转染、穿刺感染、原生质体融合、质粒递送或转座子;基于粒子的方法,如使用基因枪、磁转染或磁辅助转染、粒子轰击;和杂交方法,如核转染。

[0064] 本文中使用的术语“转染”或“转化”或“转导”是指使外源核酸转移或引入到宿主细胞中的过程。“转染”或“转化”或“转导”的细胞是已被外源核酸转染、转化或转导的细胞。

[0065] 术语“体内”是指在获取细胞的生物体的体内。“离体”或“体外”是指在获取细胞的生物体的体外。

[0066] 如本文所用,“治疗”或“处理”是获得有益或所希望结果(包含临床结果)的方法。对于本申请的目的而言,有益的或所希望的临床结果包括但不限于下列一项或多项:缓解由疾病引起的一个或多个症状、减小疾病的程度、稳定疾病(例如,防止或延缓疾病的恶化)、防止或延缓疾病的扩散(例如,转移)、防止或延缓疾病的复发、降低疾病的复发率、延缓或减缓疾病的进展、改善疾病状态、提供疾病的缓解(部分或全部)、减少治疗疾病所需的一个或多个其他药物的剂量、延缓疾病的进展、提高生活质量、和/或延长生存期。“治疗”还涵盖减少癌症的病理后果。本申请的方法考虑了这些治疗方面的任意一个或多个。

[0067] 如本文所用,术语“有效量”是指足以治疗特定病症、病况或疾病(如改善、减轻、减弱和/或延缓其一个或多个症状)的化合物或组合物的量。如本领域所理解的,“有效量”可以是一个或多个剂量,即可能需要单剂量或多剂量来达到所希望的治疗终点。

[0068] 用于治疗目的时,“受试者”、“个体”或“患者”在本文中可互换使用,是指任何动物,如哺乳动物(包括人类、家养和农场动物,和动物园、运动或宠物动物,如狗、马、猫、仓鼠、豚鼠、兔子、猴子、绵羊、牛等)、鸟、爬行动物、鱼等。在一些实施方案中,个体是人类个体。

[0069] 应当理解,本文所述本申请的实施方案包括“由……组成”和/或“基本由……组成”的实施方案。

[0070] 本文中提及“大约”值或参数包括(并描述)针对该值或参数本身的变化。例如,提

及“大约X”的描述包括“X”的描述。

[0071] 如本文所用,提及“不是”一个值或参数通常意味着并描述“除”一个值或参数之外。例如,该方法不用于治疗X型癌症意味着该方法用于治疗除X型以外的其他类型的癌症。

[0072] 本文所用的术语“约X-Y”与“约X至约Y”具有相同的含义。

[0073] 如本文和所附权利要求书中所用,单数形式“一”、“一个”和“该(the)”包括复数指代物,除非上下文另有明确规定。还应当注意的是,权利要求书可以被撰写成排除任何可选的元素。因此,该声明旨在作为在叙述权利要求元件时使用“仅仅”、“唯一”等专属术语或使用“否定式”限制的先前基础。

[0074] 本文中术语如“A和/或B”中使用的术语“和/或”旨在包括A和B;A或B;A(单独);和B(单独)。同样,本文中术语如“A、B和/或C”中使用的术语“和/或”旨在涵盖下列各实施方案:A、B和C;A、B或C;A或C;A或B;B或C;A和C;A和B;B和C;A(单独);B(单独);和C(单独)。

[0075] 本领域的普通技术人员会理解,尿嘧啶和胸腺嘧啶都可以用‘t’表示,而不是用‘u’表示尿嘧啶和用‘t’表示胸腺嘧啶;在核糖核酸的上下文中,除非另有说明,否则应理解用‘t’表示尿嘧啶。

[0076] II. Cas12b核酸酶和效应子蛋白

[0077] 本申请提供了工程化Cas12b核酸酶和效应子蛋白,其具有改进的活性,如靶结合、双链切割活性,切口酶活性,和/或基因编辑活性。还提供了具有降低或消除的核酸酶活性的工程化Cas12b核酸酶(dCas12b)。在一些实施方案中,提供了一种工程化Cas12b效应子蛋白(例如,Cas12b核酸酶、Cas12b切口酶、Cas12b融合效应子蛋白、或断裂Cas12b效应子蛋白),其包含本文所述的任何一种工程化Cas12b核酸酶或其功能衍生物。

[0078] 工程化Cas12b核酸酶

[0079] 本申请在一方面提供了具有改善的活性(例如,靶结合、双链切割活性、切口酶活性、和/或基因编辑活性)的工程化Cas12b效应子蛋白。

[0080] 在一些实施方案中,提供了一种工程化Cas12b核酸酶,相对于参考Cas12b核酸酶其包含一种、两种或三种类型的突变,其中所述突变包含:(1)参考Cas12b核酸酶中与前间区邻近基序(PAM)相互作用的一个或多个氨基酸残基用带正电荷的氨基酸残基(例如,R、H、K)取代;和/或(2)参考Cas12b核酸酶中参与打开DNA双链(dsDNA)的一个或多个氨基酸残基用具有芳香环的氨基酸残基(例如,F、Y、W)的取代;和/或(3)参考Cas12b核酸酶的RuvC结构域中与单链DNA底物相互作用的一个或多个氨基酸残基用带正电荷的氨基酸残基(例如,R、H、K)或疏水性氨基酸残基(例如,F、Y、W、M)的取代。在一些实施方案中,参考Cas12b核酸酶是天然存在的野生型Cas12b核酸酶。在一些实施方案中,参考Cas12b核酸酶是天然变体Cas12b核酸酶。在一些实施方案中,参考Cas12b核酸酶是来自嗜酸脂环酸芽孢杆菌(AaCas12b)的Cas12b核酸酶。在一些实施方案中,参考Cas12b核酸酶包含SEQ ID NO:1的氨基酸序列。在一些实施方案中,与参考Cas12b核酸酶相比,工程化Cas12b核酸酶具有增加的活性(例如,增加至少约10%、20%、30%、40%、50%、60%、70%、80%、90%、1倍、1.2倍、1.5倍、2倍、5倍、10倍、20倍、50倍、100倍或更高)(例如,靶结合、双链切割活性,切口酶活性,和/或基因编辑活性)。

[0081] 工程化Cas12b核酸酶可能包含下面A-C部分中描述的一个或多个突变。在一些实施方案中,本申请中的一个或多个突变可以与任一已知的Cas12b突变(如下面D部分中描述

的突变)结合,以产生具有改善活性的工程化Cas12b核酸酶。

[0082] 在一些实施方案中,提供了一种工程化Cas12b核酸酶,其包含相对于参考Cas12b核酸酶的一个或多个突变,其中一个或多个突变包含参考Cas12b核酸酶中与PAM相互作用的一个或多个氨基酸残基用带正电荷的氨基酸残基(例如,R或K)的取代。在一些实施方案中,提供了一种工程化Cas12b核酸酶,其包含相对于参考Cas12b核酸酶的一个或多个突变,其中所述一个或多个突变包含参考Cas12b核酸酶中参与打开DNA双链的一个或多个氨基酸残基用具有芳香环的氨基酸残基(例如,W、Y或F)的取代。在一些实施方案中,提供了一种工程化Cas12b核酸酶,其包含相对于参考Cas12b核酸酶的一个或多个突变,其中所述一个或多个突变包含参考Cas12b核酸酶的RuvC结构域中与单链DNA底物相互作用的一个或多个氨基酸残基用带正电荷的氨基酸残基(例如,R或K)的取代。在一些实施方案中,提供了一种工程化Cas12b核酸酶,其包含相对于参考Cas12b核酸酶的一个或多个突变,其中所述一个或多个突变包含参考Cas12b核酸酶的RuvC结构域中与单链DNA底物相互作用的一个或多个氨基酸残基用疏水性氨基酸残基(例如,W、Y、F或M)的取代。在一些实施方案中,参考Cas12b核酸酶包含SEQ ID NO:1的氨基酸序列。

[0083] 在一些实施方案中,提供了一种工程化Cas12b核酸酶,其包含相对于参考Cas12b核酸酶的一个或多个突变,其中所述一个或多个突变包含:1)参考Cas12b核酸酶中与PAM相互作用的一个或多个氨基酸残基用带正电荷的氨基酸残基(例如,R、H、K)的取代,和2)将参考Cas12b核酸酶中参与打开DNA双链的一个或多个氨基酸残基用具有芳香环的氨基酸残基(例如,F、Y、W)的取代。在一些实施方案中,提供了一种工程化Cas12b核酸酶,其包含相对于参考Cas12b核酸酶的一个或多个突变,其中所述一个或多个突变包含:1)参考Cas12b核酸酶中与PAM相互作用的一个或多个氨基酸残基用带正电荷的氨基酸残基(例如,R、H、K)的取代,和2)参考Cas12b核酸酶的RuvC结构域中与单链DNA底物相互作用的一个或多个氨基酸残基用带正电荷的氨基酸残基(例如,R、H、K)或疏水性氨基酸残基(例如,F、Y、W、M)的取代。在一些实施方案中,提供了一种工程化Cas12b核酸酶,其包含相对于参考Cas12b核酸酶的一个或多个突变,其中所述一个或多个突变包含:1)参考Cas12b核酸酶中参与打开DNA双链的一个或多个氨基酸残基用具有芳香环的氨基酸残基(例如,F、Y、W)的取代,和2)参考Cas12b核酸酶的RuvC结构域中与单链DNA底物相互作用的一个或多个氨基酸残基用带正电荷的氨基酸残基(例如,R、H、K)或疏水性氨基酸残基(例如,F、Y、W、M)的取代。在一些实施方案中,参考Cas12b核酸酶包含SEQ ID NO:1的氨基酸序列。

[0084] 在一些实施方案中,提供了一种工程化Cas12b核酸酶,其包含相对于参考Cas12b核酸酶的一个或多个突变,其中所述一个或多个突变包含:1)参考Cas12b核酸酶中与PAM相互作用的一个或多个氨基酸残基用带正电荷的氨基酸残基(例如,R、H、K)的取代,2)参考Cas12b核酸酶中参与打开DNA双链的一个或多个氨基酸残基用具有芳香环的氨基酸残基(例如,F、Y、W)的取代,和3)参考Cas12b核酸酶的RuvC结构域中与单链DNA底物相互作用的一个或多个氨基酸残基用带正电荷的氨基酸残基(例如,R、H、K)的取代。在一些实施方案中,参考Cas12b核酸酶包含SEQ ID NO:1的氨基酸序列。

[0085] 在一些实施方案中,提供了一种工程化Cas12b核酸酶,其包含相对于参考Cas12b核酸酶的一个或多个突变,其中所述一个或多个突变包含:1)参考Cas12b核酸酶中与PAM相互作用的一个或多个氨基酸残基用带正电荷的氨基酸残基(例如,R、H、K)的取代,2)参考

Cas12b核酸酶中参与打开DNA双链的一个或多个氨基酸残基用具有芳香环的氨基酸残基(例如,F、Y、W)的取代,和3)参考Cas12b核酸酶的RuvC结构域中与单链DNA底物相互作用的一个或多个氨基酸残基用疏水性氨基酸残基(例如,F、Y、W、M)的取代。在一些实施方案中,参考Cas12b核酸酶包含SEQ ID NO:1的氨基酸序列。

[0086] 本文描述的突变可以基于参考Cas12b核酸酶的结构进行设计。Yang H., et al. *Cell* 167:1814-1828(2016)和Liu L. et al. *Mol. Cell* 65:310-322(2017)中已经描述了以酸土脂环酸芽孢杆菌(*Alicyclobacillus acidoterrestris*)Cas12b与sgRNA结合作为二元复合物、以及再与靶DNA结合作为三元复合物的晶体结构。简而言之,所述晶体结构显示2个不连续的REC(识别,残基15-386、658-783)和NUC(核酸酶,残基1-14、387-658和784-1129)瓣叶,每个瓣叶(lobe)由几个域组成。所述crRNA(或单指导RNA,sgRNA)结合于两个瓣叶之间的中央通道中。PAM识别是序列特异性的,主要经由与REC1(helical-1)和WED-II(OBD-II)域的相互作用发生。sgRNA-靶DNA异源双链主要以序列独立的方式与REC瓣叶结合。

[0087] 应当理解,其他Cas12b直系同源物,例如BhCas12b(SEQ ID NO:59)、Bs3Cas12b(SEQ ID NO:56)、LsCas12b(SEQ ID NO:58)、SbCas12b(SEQ ID NO:60)、AkCas12b(SEQ ID NO:54)、AmCas12b(SEQ ID NO:55)、BsCas12b(SEQ ID NO:57)和DiCas12b等,具有与AaCas12b(SEQ ID NO:1)和本文所述的其他示例性参考Cas12b蛋白相似的域结构,并且可以基于任一个直系同源物使用与本文所述的示例性的工程化AaCas12b蛋白相对应的断裂位置来设计工程化的Cas12b蛋白。对应位置是指当两条多肽的氨基酸序列彼此对齐时,这两条多肽中的位置是彼此对齐的。参见本申请的图8。此外,Teng F. et al., *Cell Discovery* (2019) 5:23的图S2提供了AaCas12b、AkCas12b、AmCas12b、Bs3Cas12b、BsCas12b、LsCas12b、BhCas12b和SbCas12b的比对,其通过引用以其整体并入本文。

[0088] A.参考Cas12b中与PAM相互作用的一个或多个氨基酸残基用带正电荷的氨基酸残基的取代。

[0089] 在一些实施方案中,工程化Cas12b核酸酶包含参考Cas12b核酸酶中与PAM相互作用的一个或多个氨基酸残基用带正电荷的氨基酸残基(例如,R、H、K)的取代。在一些实施方案中,工程化Cas12b核酸酶包含1、2、3、4、5或6个氨基酸取代。

[0090] 在一些实施方案中,参考Cas12b核酸酶中与PAM相互作用的一个或多个氨基酸残基是三维结构中距离PAM 15(例如,14、13、12、11、10、9、8、7、6、5、4、3、2、1或更小)中的任一个以内)埃以内的氨基酸。在一些实施方案中,参考Cas12b核酸酶中与PAM相互作用的一个或多个氨基酸残基是三维结构中距离PAM 10埃内的氨基酸。在一些实施方案中,参考Cas12b核酸酶中与PAM相互作用的一个或多个氨基酸残基是三维结构中距离PAM 9埃内的氨基酸。在一些实施方案中,与PAM相互作用的一个或多个氨基酸残基位于以下位置中的一个或多个:116、123、130、132、144、145、153、173、222、395、400和475。在一些实施方案中,与PAM相互作用的一个或多个氨基酸残基包含以下氨基酸残基中的一个或多个:D116、K123、D130、D132、N144、K145、E153、D173、Q222、D395、N400和E475。在一些实施方案中,与PAM相互作用的一个或多个氨基酸残基包含以下氨基酸残基中的一个或多个:D116和E475。在一些实施方案中,氨基酸残基按照SEQ ID NO:1编号。

[0091] 在本申请的上下文中,D116是指所引用的氨基酸序列中的第116位氨基酸D(天冬

氨酸)。常用氨基酸的3字母缩写和单字母缩写如下：

[0092]	Ala (A)	Leu (L)	Gln (Q)	Ser (S)
	Arg (R)	Lys (K)	Glu (E)	Thr (T)
	Asn (N)	Met (M)	Gly (G)	Trp (W)
	Asp (D)	Phe (F)	His (H)	Tyr (Y)
	Cys (C)	Pro (P)	Ile (I)	Val (V)

[0093] 如本文所用，“氨基酸位于位置X，其中氨基酸按照SEQ ID NO:1编号”是指当将参考酶Cas12b的氨基酸序列与SEQ ID NO:1基于序列同源性比对时，位于参考酶Cas12b的某一位置的氨基酸残基，其对应于SEQ ID NO:1中的X位置。例如，图8显示了Cas12b直系同源物的氨基酸序列的同源性比对 (SEQ ID NO:1和54-60)。本领域技术人员可以容易地使用已知软件，如Clustal Omega，比较和比对任意参考Cas12b核酸酶的氨基酸序列与SEQ ID NO:1，以确定与SEQ ID NO:1中位置X相对应的氨基酸位置。

[0094] 在一些实施方案中，带正电荷的氨基酸残基为R、H或K。在一些实施方案中，带正电荷的氨基酸残基为R。在一些实施方案中，带正电荷的氨基酸残基为K。

[0095] 在一些实施方案中，参考Cas12b核酸酶中与PAM相互作用的一个或多个氨基酸残基用带正电荷的氨基酸残基的取代是以下一个或多个取代：D116R、K123R、D130R、D132R、N144R、K145R、E153R、D173R、Q222R、D395R、N400R和E475R。在一些实施方案中，参考Cas12b核酸酶中与PAM相互作用的一个或多个氨基酸残基用带正电荷的氨基酸残基的取代是以下一个或多个取代：D116R和E475R。在一些实施方案中，工程化Cas12b核酸酶包含D116R突变。在一些实施方案中，工程化Cas12b核酸酶包含E475R突变。在一些实施方案中，氨基酸残基按照SEQ ID NO:1编号。

[0096] 在一些实施方案中，工程化Cas12b核酸酶包含与SEQ ID NO:2或3的氨基酸序列具有至少约85%序列同一性 (如至少约87%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列同一性的任一个)的氨基酸序列。在一些实施方案中，工程化Cas12b核酸酶包含SEQ ID NO:2或3的氨基酸序列。

[0097] B. 参考Cas12b核酸酶中参与打开DNA双链的一个或多个氨基酸残基用具有芳香环的氨基酸残基的取代

[0098] 在一些实施方案中，工程化Cas12b核酸酶包含参考Cas12b核酸酶中参与打开DNA双链的一个或多个氨基酸残基用具有芳香环的氨基酸残基 (例如，F、Y、W) 的取代。在一些实施方案中，工程化Cas12b核酸酶包含1、2、3、4、5或6个氨基酸残基的取代。

[0099] 在一些实施方案中，参与打开DNA双链的一个或多个氨基酸残基与PAM中相对于靶链3'端的最后一个碱基对相互作用。例如，AaCas12b识别的PAM序列是5'-TTN-3'碱基对。PAM中相对于靶链3'端的最后一个碱基对是PAM序列3'端的N碱基形成的碱基对，其后是靶位点的序列。

[0100] 在一些实施方案中，参与打开DNA双链的一个或多个氨基酸残基位于以下一个或多个位置：118和119，如Q118和Q119。在一些实施方案中，氨基酸残基按照SEQ ID NO:1编号。

[0101] 在一些实施方案中，具有芳香环的氨基酸残基为Y、F或W。在一些实施方案中，参与打开DNA双链的氨基酸残基用F、Y或W取代。在一些实施方案中，工程化Cas12b核酸酶包含以

下任一种:i)Q118Y、Q118F或Q118W;和/或ii)Q119Y、Q119F或Q119W。在一些实施方案中,氨基酸残基按照SEQ ID NO:1编号。

[0102] 在一些实施方案中,所述参考Cas12b核酸酶中参与打开DNA双链的一个或多个氨基酸残基用具有芳香环的氨基酸的取代为Q119Y、Q119F或Q119W。在一些实施方案中,氨基酸残基按照SEQ ID NO:1编号。

[0103] 在一些实施方案中,工程化Cas12b核酸酶包含与SEQ ID NO:4、5或6的氨基酸序列具有至少约85%序列同一性(如至少约88%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列同一性的任一个)的氨基酸序列。在一些实施方案中,工程化Cas12b核酸酶包含SEQ ID NO:4、5或6的氨基酸序列。

[0104] C.参考Cas12b核酸酶的RuvC结构域中与单链DNA底物相互作用的一个或多个氨基酸残基用带正电荷的氨基酸残基或疏水性氨基酸残基的取代。

[0105] 在一些实施方案中,工程化Cas12b核酸酶包含参考Cas12b核酸酶中位于RuvC结构域中并与单链DNA底物相互作用的一个或多个氨基酸残基用具有带正电荷的氨基酸残基(例如,R、H、K)的取代。在一些实施方案中,工程化Cas12b核酸酶包含参考Cas12b核酸酶中位于RuvC结构域中并与单链DNA底物相互作用的一个或多个氨基酸残基用疏水性氨基酸残基(例如,F、Y、W、M)的取代。在一些实施方案中,工程化Cas12b核酸酶包含1、2、3、4、5或6个氨基酸残基的取代。

[0106] 在一些实施方案中,位于RuvC结构域中并与单链DNA底物相互作用的一个或多个氨基酸残基在三维结构中距离单链DNA底物15埃以内(例如,14、13、12、11、10、9、8、7、6、5、4、3、2、1或更小中的任一个以内)。在一些实施方案中,位于RuvC结构域中并与单链DNA底物相互作用的一个或多个氨基酸残基在三维结构中距离单链DNA底物10埃以内。在一些实施方案中,位于RuvC结构域中并与单链DNA底物相互作用的一个或多个氨基酸残基在三维结构中距离单链DNA底物9埃以内。

[0107] RuvC结构域是Cas12b蛋白的活性结构域,负责切割单链DNA或双链DNA。在蛋白的一级序列中,RuvC结构域包含第一RuvC结构域(RuvC-I)、第二RuvC结构域(RuvC-II)和第三RuvC结构域(RuvC-III)。

[0108] 在一些实施方案中,位于RuvC结构域中并与单链DNA底物相互作用的一个或多个氨基酸残基位于以下位置中的一个或多个:300、301、304、329、636、639、647、682、757、758、761、764、768、852、854、856、857、858、860、862、863、865、866、867、869、938、956、957、958、994、1093和1097。在一些实施方案中,位于RuvC结构域中并且与单链DNA底物相互作用的一个或多个氨基酸残基包含以下氨基酸残基中的一个或多个:D300、K301、E304、N329、E636、Q639、T647、Q682、I757、E758、E761、E764、K768、E852、Q854、N856、N857、D858、P860、S862、E863、N865、Q866、L867、Q869、E938、E956、G957、E958、I994、Q1093和W1097。在一些实施方案中,位于RuvC结构域中并与单链DNA底物相互作用的一个或多个氨基酸残基包含以下氨基酸残基中的一个或多个:D300、K301、E636、Q639、T647、Q682、I757、E758、E761、K768、Q854、N857、D858、N865、Q866、Q869、I994、Q1093和W1097。在一些实施方案中,位于RuvC结构域中并与单链DNA底物相互作用的一个或多个氨基酸残基包含以下氨基酸残基中的一个或多个:E636、I757、E758、E761、Q854、N857、D858、N865、Q866、Q869和Q1093。在一些实施方案中,氨基酸残基按照SEQ ID NO:1编号。

[0109] 在一些实施方案中,工程化Cas12b核酸酶包含参考Cas12b核酸酶中位于RuvC结构域中并与单链DNA底物相互作用的一个或多个氨基酸残基用具有带正电荷的氨基酸残基(例如,R、H、K)的取代。在一些实施方案中,带正电荷的氨基酸残基是R。在一些实施方案中,带正电荷的氨基酸残基是K。在一些实施方案中,工程化Cas12b核酸酶包含以下一个或多个取代:D300R,K301R,E304R,N329R,E636R,Q639R,T647R,Q682R,I757R,E758R,E761R,E764R,K768R,E852R,Q854R,N856R,N857R,D858R,P860R,S862R,E863R,N865R,Q866R,L867R,Q869R,E938R,E956R,G957R,E958R,I994R,Q1093R,W1097R,E636K,Q639K,T647K,Q682K,I757K,E758K,E761K,Q854K,N857K,D858K,N865K,Q866K,I994K,Q1093K和W1097K,其中氨基酸残基按照SEQ ID NO:1编号。在一些实施方案中,工程化Cas12b核酸酶包含以下一个或多个取代:D300R,K301R,E636R,Q639R,T647R,Q682R,I757R,E758R,E761R,K768R,Q854R,N857R,D858R,N865R,Q866R,I994R,Q1093R,W1097R,E636K,Q639K,T647K,Q682K,I757K,E758K,E761K,Q854K,N857K,D858K,N865K,I994K,Q1093K和W1097K,其中氨基酸残基按照SEQ ID NO:1编号。在一些实施方案中,工程化Cas12b核酸酶包含以下一个或多个取代:E636R,I757R,E758R,E761R,Q854R,D858R,E636K,I757K,E758K,E761K,Q854K,N857K和D858K,其中氨基酸残基按照SEQ ID NO:1编号。在一些实施方案中,工程化Cas12b核酸酶包含以下一个或多个取代:E636R,I757R,E758R,E761R,Q854R和D858R,其中氨基酸残基按照SEQ ID NO:1编号。在一些实施方案中,工程化Cas12b核酸酶包含以下一个或多个取代:E636K,I757K,E758K,E761K,Q854K,N857K和D858K,其中氨基酸残基按照SEQ ID NO:1编号。

[0110] 在一些实施方案中,参考Cas12b核酸酶中位于RuvC结构域中并与单链DNA底物相互作用的一个或多个氨基酸残基的取代是以下一个或多个取代:E636R,I757R,E758R,E761R,Q854R,N857K和D858R,其中氨基酸残基按照SEQ ID NO:1编号。在一些实施方案中,工程化Cas12b核酸酶包含与SEQ ID NO:7-13中任一个的氨基酸序列具有至少约85%序列同一性(如至少约88%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列同一性中的任一个)的氨基酸序列。在一些实施方案中,工程化Cas12b核酸酶包含SEQ ID NO:7-13中任一个的氨基酸序列。

[0111] 在一些实施方案中,工程化Cas12b核酸酶包含参考Cas12b核酸酶中位于RuvC结构域中并与单链DNA底物相互作用的一个或多个氨基酸残基用疏水性氨基酸残基的取代。在一些实施方案中,所述疏水性氨基酸残基是A、M、L、I、V、C、Y、F或W。在一些实施方案中,所述疏水性氨基酸残基是W、Y、F或M。在一些实施方案中,所述疏水性氨基酸残基是W、Y或M。在一些实施方案中,工程化Cas12b核酸酶包含以下一个或多个取代:i) E758W、E758Y、E758F或E758M, ii) E761W、E761Y、E761F或E761M, iii) E863W、E863Y、E863F或E863M, iv) N865W、N865Y、N865F或N865M, v) Q866W、Q866F、Q866Y或Q866M, vi) Q869W、Q869Y、Q869F或Q869M, vii) E956W、E956Y、E956F或E956M,和viii) Q1093W、Q1093F、Q1093Y或Q1093M;其中氨基酸残基按照SEQ ID NO:1编号。在一些实施方案中,工程化Cas12b核酸酶包含以下一个或多个取代:i) E758W、E758Y或E758M, ii) E761Y, iii) N865W、N865F或N865Y, iv) Q866M, v) Q869M,和vi) Q1093W、Q1093F、Q1093Y或Q1093M;其中氨基酸残基按照SEQ ID NO:1编号。在一些实施方案中,工程化Cas12b核酸酶包含以下一个或多个取代:i) N865W或N865Y, ii) Q866M, iii) Q869M,和iv) Q1093W或Q1093Y;其中氨基酸残基按照SEQ ID NO:1编号。在一些实施方案中,工程化Cas12b核酸酶包含以下一个或多个取代:865W、865Y、866M、869M、1093W和1093Y。在

一些实施方案中,参考Cas12b核酸酶中位于RuvC结构域中并与单链DNA底物相互作用的一个或多个氨基酸残基的取代是以下一个或多个取代:N865W、N865Y、Q866M、Q869M、Q1093W和Q1093Y。在一些实施方案中,工程化Cas12b核酸酶包含Q866M和Q869M取代。在一些实施方案中,氨基酸残基按照SEQ ID NO:1编号。

[0112] 在一些实施方案中,工程化Cas12b核酸酶包含与SEQ ID NO:14-20中任一个的氨基酸序列具有至少约85%序列同一性(如至少约88%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列同一性中的任一个)的氨基酸序列。在一些实施方案中,工程化Cas12b核酸酶包含SEQ ID NO:14-20中任一个的氨基酸序列。

[0113] D. 其他突变

[0114] 上述A-C部分中描述的任何个或多个突变可以与增加Cas12b活性的任何一个或多个已知突变相结合,所述Cas12b活性如靶结合、靶特异性、双链切割活性,切口酶活性,和/或基因编辑活性。示例性突变可见于例如以下文献W02022120520、W02022040909、W02022042557、CN113308451A和CN112195164A,其内容通过引用以其整体并入本文。

[0115] 在一些实施方案中,参考Cas12b蛋白从N末端到C末端包含以下中的一个或多个:第一WED结构域(WED-I)、第一REC结构域(REC1)、第二WED结构域(WED-II)、第一RuvC结构域(RuvC-I)、BH结构域、第二REC结构域(REC2)、第二RuvC结构域(RuvC-II)、第一Nuc结构域(Nuc-I)、第三RuvC结构域(RuvC-III)和第二Nuc结构域(Nuc-II)。在一些实施方案中,其他一个或多个突变(例如,插入、缺失、取代)可以存在于一个或多个这样的结构域中。

[0116] 在一些实施方案中,工程化Cas12b核酸酶还包含一个或多个柔性区突变,所述突变增加了参考Cas12b核酸酶中柔性区的柔性。可以使用本领域已知的任何方法确定参考Cas12b核酸酶中的柔性区。在一些实施方案中,仅基于参考酶的氨基酸序列确定多个柔性区。在一些实施方案中,基于参考酶的结构信息确定多个柔性区,包括例如二级结构、晶体结构、NMR结构等。

[0117] 在一些实施方案中,使用从以下组中选择的程序确定多个柔性区:PredyFlexy、FoldUnfold、PROFbval、Flexserv、FlexPred、DynaMine和Disomine。在一些实施方案中,多个柔性区位于随机卷曲(random crimp)处。在一些实施方案中,多个柔性区位于参考Cas12b核酸酶的DNA和/或RNA相互作用结构域中。在一些实施方案中,柔性区的长度至少为约5(例如,5)个氨基酸。

[0118] 在一些实施方案中,工程化Cas12b核酸酶包含一个或多个突变,所述突变增加对应于氨基酸残基855至859的柔性区的柔性,其中氨基酸残基基于SEQ ID NO:1编号,其中与参考Cas12b核酸酶相比,工程化Cas12b核酸酶具有增加的活性(例如,增加至少约10%、20%、30%、40%、50%、60%、70%、80%、90%、1倍、1.2倍、1.5倍、2倍、5倍、10倍、20倍、50倍、100倍或更高)(例如,靶结合、双链切割活性,切口酶活性,和/或基因编辑活性)。在一些实施方案中,参考Cas12b核酸酶是AaCas12b。在一些实施方案中,参考Cas12b核酸酶包含SEQ ID NO:1的氨基酸序列。在一些实施方案中,一个或多个突变包含在柔性区中插入一个或多个(例如,2个)G残基。在一些实施方案中,把所述一个或多个G残基插入柔性区中柔性氨基酸残基的N末端,其中柔性氨基酸残基选自下组:G、S、N、D、H、M、T、E、Q、K、R、A和P。在一些实施方案中,所述柔性氨基酸残基根据偏好选择:G>S>N>D>H>M>T>E>Q>K>R>A>P。在一些实施方案中,所述一个或多个突变包含用G残基取代柔性区中的疏水性氨基酸残基,其中所

述疏水性氨基酸残基选自下组:L、I、V、C、Y、F和W。在一些实施方案中,增加柔性的一个或多个突变包含N856G。

[0119] E.突变的组合

[0120] 采用本说明书中A-D部分所描述的方法结合示例性序列列表中的多个氨基酸取代所获得的工程化的酶均在本申请的范围内。在一些实施方案中,工程化Cas12b核酸酶包含上面A-D部分中描述的一个或多个突变(例如,取代)。

[0121] 在一些实施方案中,工程化Cas12b核酸酶包含在以下任一氨基酸残基位置的取代或取代的组合:(1) 116;(2) 475;(3) 119和475;(4) 119, 475, 和758;(5) 119;(6) 636;(7) 757;(8) 758;(9) 761;(10) 768;(11) 858;(12) 854;(13) 857;(14) 119, 475, 和758;(15) 768;(16) 757和758;(17) 757和761;(18) 757和768;(19) 758和761;(20) 758和768;(21) 761和768;(22) 757, 758, 和761;(23) 757, 758, 和768;(24) 757, 761和768;(25) 758, 761, 和768;(26) 757, 758, 761, 和768;(27) 865;(28) 866;(29) 869;(30) 1093;和(31) 866和869,其中氨基酸位置编号是相对于SEQ ID NO:1而言的。

[0122] 在一些实施方案中,工程化Cas12b核酸酶包含在以下任一氨基酸残基处的取代或取代的组合:(1) D116;(2) E475;(3) Q119和E475;(4) Q119, E475, 和E758;(5) Q119;(6) E636;(7) I757;(8) E758;(9) E761;(10) K768;(11) D858;(12) Q854;(13) N857;(14) Q119, E475, 和E758;(15) K768;(16) I757和E758;(17) I757和E761;(18) I757和K768;(19) E758和E761;(20) E758和K768;(21) E761和K768;(22) I757, E758, 和E761;(23) I757, E758, 和K768;(24) I757, E761和K768;(25) E758, E761, 和K768;(26) I757, E758, E761, 和K768;(27) N865;(28) Q866;(29) Q869;(30) Q1093;和(31) Q866和Q869;其中氨基酸位置编号是相对于SEQ ID NO:1而言的。在一些实施方案中,工程化Cas12b核酸酶包含以下任一氨基酸残基的取代或取代的组合:(1) Q866+Q869;(2) Q119+E475;和(3) Q119+E475+E758;其中氨基酸残基按照SEQ ID NO:1编号。在一些实施方案中,氨基酸位置D116和/或E475处的取代是用带正电荷的氨基酸残基(如R或K)取代。在一些实施方案中,氨基酸位置Q119处的取代是用具有芳香族侧链的氨基酸残基(如Y、F或W)取代。在一些实施方案中,氨基酸位置E636、I757、E758、E761、K768、Q854、D858和或N857处的取代是用带正电荷的氨基酸残基(如R或K)取代。在一些实施方案中,氨基酸位置N865、Q866、Q869和/或Q1093处的取代是用疏水性氨基酸残基(如W、Y或M)的取代。

[0123] 在一些实施方案中,工程化Cas12b核酸酶包含以下氨基酸残基中的任意一个或多个或其组合:(1) 116R;(2) 475R;(3) 119F和475R;(4) 119F, 475R, 和758R;(5) 119Y;(6) 119F;(7) 119W;(8) 636R;(9) 757R;(10) 758R;(11) 761R;(12) 854R;(13) 857K;(14) 768R;(15) 757R和758R;(16) 757R和761R;(17) 757R和768R;(18) 758R和761R;(19) 758R和768R;(20) 761R和768R;(21) 757R, 758R, 和761R;(22) 757R, 758R, 和768R;(23) 757R, 761R, 和768R;(24) 758R, 761R, 和768R;(25) 757R, 758R, 761R, 和768R;(26) 865W;(27) 865Y;(28) 866M;(29) 869M;(30) 1093W;(31) 1093Y;(32) 866M和869M;和(33) 858R;其中氨基酸位置编号是相对于SEQ ID NO:1而言的。

[0124] 在一些实施方案中,工程化Cas12b核酸酶包含以下任一种取代或其组合:(1) D116R;(2) E475R;(3) Q119F+E475R;(4) Q119F+E475R+E758R;(5) Q119Y;(6) Q119F;(7) Q119W;(8) I757R;(9) E758R;(10) E761R;(11) K768R;(12) I757R+E758R;(13) I757R+E761R;

(14) I757R+K768R; (15) E758R+E761R; (16) E758R+K768R; (17) E761R+K768R; (18) I757R+E758R+E761R; (19) I757R+E758R+K768R; (20) I757R+E761R+K768R; (21) E758R+E761R+K768R; (22) I757R+E758R+E761R+K768R; (23) Q866M; (24) Q869M; (25) Q866M+Q869M; (26) E636R; (27) Q854R; (28) N857K; (29) N865W; (30) N865Y; (31) Q1093W; (32) Q1093Y; 和 (33) D858R; 其中氨基酸位置编号是相对于SEQ ID NO:1而言的。在一些实施方案中,工程化Cas12b核酸酶包含以下任一种取代或其组合: (1) Q866M+Q869M; (2) Q119F+E475R; 和 (3) Q119F+E475R+E758R; 其中氨基酸残基按照SEQ ID NO:1编号。

[0125] 在一些实施方案中,工程化Cas12b核酸酶包含以下一个或多个取代:D116R, K123R, D130R, D132R, N144R, K145R, E153R, D173R, Q222R, D395R, N400R, 和E475R。在一些实施方案中,工程化Cas12b核酸酶包含以下一个或多个取代:Q118Y, Q118F, Q118W, Q119Y, Q119F, 和Q119W。在一些实施方案中,工程化Cas12b核酸酶包含以下一个或多个取代:D300R, K301R, E304R, N329R, E636R, Q639R, T647R, Q682R, I757R, E758R, E761R, E764R, K768R, E852R, Q854R, N856R, N857R, D858R, P860R, S862R, E863R, N865R, Q866R, L867R, Q869R, E938R, E956R, G957R, E958R, I944R, Q1093R, 和W1097R。在一些实施方案中,工程化Cas12b核酸酶包含以下一个或多个取代:E636K, Q639K, T647K, Q682K, I757K, E758K, E761K, Q854K, N857K, D858K, N865K, Q866K, I994K, Q1093K, 和W1097K。在一些实施方案中,工程化Cas12b核酸酶包含以下一个或多个取代:E758W, E758Y, E758F, E758M, E761W, E761Y, E761F, E761M, E863W, E863Y, E863F, E863M, N865W, N865Y, N865F, N865M, Q866W, Q866Y, Q866F, Q866M, Q869W, Q869Y, Q869F, Q869M, E956W, E956Y, E956F, E956M, Q1093W, Q1093Y, Q1093F, 和Q1093M。在一些实施方案中,氨基酸位置编号是相对于SEQ ID NO:1而言的。

[0126] 在一些实施方案中,工程化Cas12b核酸酶包含Q866和Q869处的氨基酸取代。在一些实施方案中,工程化Cas12b核酸酶包含氨基酸取代Q866M和Q869M。在一些实施方案中,氨基酸位置编号是相对于SEQ ID NO:1而言的。在一些实施方案中,工程化Cas12b核酸酶包含与SEQ ID NO:20的氨基酸序列具有至少约85%序列同一性(如至少约88%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列同一性中的任一个)的氨基酸序列。在一些实施方案中,工程化Cas12b核酸酶包含SEQ ID NO:20的氨基酸序列。

[0127] 在一些实施方案中,工程化Cas12b核酸酶包含Q119和E475处的氨基酸取代。在一些实施方案中,工程化Cas12b核酸酶包含氨基酸取代Q119F和E475R。在一些实施方案中,氨基酸位置编号是相对于SEQ ID NO:1而言的。在一些实施方案中,工程化Cas12b核酸酶包含与SEQ ID NO:21的氨基酸序列具有至少约85%序列同一性(如至少约88%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列同一性中的任一个)的氨基酸序列。在一些实施方案中,工程化Cas12b核酸酶包含SEQ ID NO:21的氨基酸序列。

[0128] 在一些实施方案中,工程化Cas12b核酸酶包含Q119、E475和E758处的氨基酸取代。在一些实施方案中,工程化Cas12b核酸酶包含氨基酸取代Q119F、E475R和E758R。在一些实施方案中,氨基酸位置编号是相对于SEQ ID NO:1而言的。在一些实施方案中,工程化Cas12b核酸酶包含与SEQ ID NO:22的氨基酸序列具有至少约85%序列同一性(如至少约88%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列同一性中的任一个)的氨基酸序列。在一些实施方案中,工程化Cas12b核酸酶包含SEQ ID NO:22的氨基酸序列。

[0129] 参考Cas12b核酸酶

[0130] 在一些实施方案中,参考Cas12b核酸酶是AaCas12b,或其直系同源物。在一些实施方案中,参考Cas12b核酸酶是天然存在的Cas12b核酸酶。在一些实施方案中,参考Cas12b核酸酶是野生型Cas12b核酸酶。在一些实施方案中,参考Cas12b核酸酶是工程化Cas12b核酸酶。

[0131] 来自多种生物体的Cas12b核酸酶可作为参考Cas12b核酸酶,以提供本申请的工程化Cas12b核酸酶和效应子蛋白。在一些实施方案中,参考Cas12b核酸酶具有酶活性。在一些实施方案中,参考Cas12b是切割靶双螺旋核酸(例如,双螺旋DNA)的两条链的核酸酶。在一些实施方案中,参考Cas12b是切口酶,其切割靶双螺旋核酸(例如,双螺旋DNA)的单链。在一些实施方案中,参考Cas12b核酸酶是酶促无活性的(例如,dCas12b)。与Cas12b或其功能衍生物具有一定序列同一性(例如,至少约60%、70%、80%、85%、90%、95%、98%或更高)的直系同源物可作为设计本申请的工程化Cas12b核酸酶或效应子蛋白的基础。在一些实施方案中,参考Cas12b核酸酶是突变体Cas12b,但不包含上述A-E部分中描述的任何突变。

[0132] 在一些实施方案中,工程化Cas12b核酸酶基于天然存在的Cas12b核酸酶的功能变体。在一些实施方案中,所述功能变体具有一个或多个突变,如氨基酸取代、插入和/或缺失。例如,与野生型天然存在的Cas12b核酸酶相比,功能变体可以包括1、2、3、4、5、6、7、8、9、10个或更多个氨基酸取代。在一些实施方案中,一个或多个取代是保守取代。在一些实施方案中,功能变体具有天然存在的Cas12b核酸酶的所有结构域。在一些实施方案中,功能变体不具有天然存在的Cas12b核酸酶的一个或多个结构域。

[0133] V-B型CRISPR-Cas12b(也称为C2c1)系统已被鉴定为双RNA指导(即,crRNA和tracrRNA)的DNA内切酶系统,具有与Cas9和Cas12a不同的特征(Shmakov, S. et al. Mol. Cell 60, 385-397 (2015))。首先,报道了,当Cas12b与crRNA/tracrRNA双链体重构时,它在体外在PAM位点远端生成交错端。其次,尽管Cas12b的RuvC结构域与Cas9和Cas12a的RuvC结构域相似,但其假定的Nuc结构域与Cas9的HNH结构域和Cas12a的Nuc结构域在序列或结构上没有相似性。此外,Cas12b蛋白比最广泛使用的SpCas9和Cas12a更小(例如,AacCas12b:1, 129个氨基酸(aa); SpCas9:1, 369个aa; AsCas12a:1, 353个aa; LbCas12a:1, 228个aa),这使得在基因治疗中Cas12b适合于腺相关病毒(AAV)介导的体内递送。与小尺寸的Cas9蛋白(如SaCas9和CjCas9)相比,Cas12b识别更简单的PAM序列(例如,AacCas12b:5'-TTN-3');与SaCas9:5'-NNGRRT-3'、CjCas9:5'-NNNNRYAC-3'相比,这显著增加了基因组中Cas12b的靶向范围。此外,Cas12b的脱靶效应最小,因此可以作为治疗和临床应用的更安全选择。

[0134] 来自多种生物体的Cas12b(C2c1)核酸酶可用作参考Cas12b核酸酶,以提供本申请的工程化Cas12b效应子蛋白。示例性的Cas12b核酸酶已在,例如,Shmakov, S. et al. Mol. Cell 60, 385-397 (2015); Shmakov, S. et al. Nat. Rev. Microbiol. 15, 169-182 (2017); W02016205764, 和W02020087631中进行了描述,其内容通过引用以其整体并入本文。

[0135] 在一些实施方案中,工程化Cas12b效应子蛋白基于参考Cas12b蛋白(例如,Cas12b核酸酶),所述参考Cas12b蛋白选自嗜酸脂环酸芽孢杆菌的Cas12b蛋白(AaCas12b)、来自卡克加文脂环酸芽孢杆菌的Cas12b(AkCas12b)、来自大孢脂环酸芽孢杆菌的Cas12b(AmCas12b)、来自外村尚芽孢杆菌的Cas12b(BhCas12b)、来自芽孢杆菌的BsCas12b、来自芽

孢杆菌的Bs3Cas12b、来自脱硫弧菌 (*Desulfovibrio inopinatus*) 的Cas12b (DiCas12b)、来自沉积物莱西氏菌的Cas12b (LsCas12b)、来螺旋体门细菌的Cas12b (SbCas12b)、来自热生肿块芽孢杆菌 (*Tuberibacillus calidus*) 的Cas12b (TcCas12b), 及其功能衍生物。天然存在的Cas12b蛋白的序列是已知的, 例如在UniProtKB IDs: T0D7A2, A0A6I3SPI6, 和 A0A6I7FUC4中, 其通过引用以其整体并入本文。

[0136] 在一些实施方案中, 参考Cas12b蛋白是来自嗜酸脂环酸芽孢杆菌 (AaCas12b) 的Cas12b核酸酶或其功能衍生物。在一些实施方案中, 工程化Cas12b效应子蛋白基于参考Cas12b蛋白, 该参考Cas12b蛋白包含与SEQ ID NO:1的氨基酸序列具有至少约85% (例如, 至少约88%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%中的任一个) 序列同一性的氨基酸序列。在一些实施方案中, 工程化Cas12b效应子蛋白基于包含SEQ ID NO:1的氨基酸序列的参考Cas12b核酸酶。

[0137] 值得注意的是, 与参考Cas12b蛋白或其片段具有一定序列同一性 (例如, 至少约60%、70%、80%、85%、90%、95%、98%或更高中的任一个) 的直系同源物可用作设计本申请的工程化Cas12b效应子蛋白的基础。本领域技术人员可以根据目的和应用确定适合用于本申请的Cas12b直系同源物或其片段的序列同一性百分比。确定序列同一性值的方法可在Computational Molecular Biology, Lesk, A.M., ed., Oxford University Press, New York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D.W., ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; 和Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991) 中找到。WO2020/087631中描述了多种Cas12b直系同源物, 其内容通过引用以其整体并入本文。在一些实施方案中, 工程化Cas12b效应子蛋白基于参考Cas12b蛋白, 所述参考Cas12b蛋白包含与SEQ ID NO:54-60中任一个的氨基酸序列具有至少约85% (例如, 至少约90%、91%、92%、93%、94%、95%、96%、97%、98%或99%中的任一个) 序列同一性的氨基酸序列。

[0138] 工程化Cas12b的活性

[0139] 在一些实施方案中, 与参考Cas12b核酸酶相比, 工程化Cas12b核酸酶具有增加的活性。在一些实施方案中, 活性是靶DNA结合活性。在一些实施方案中, 活性是位点特异性核酸酶活性。在一些实施方案中, 活性是双链DNA切割活性。在一些实施方案中, 活性是单链DNA切割活性, 包括, 例如, 位点特异性DNA切割活性或非特异性DNA切割活性。在一些实施方案中, 活性是单链RNA切割活性, 如位点特异性RNA切割活性或非特异性RNA切割活性。在一些实施方案中, 活性是在体外测量的。在一些实施方案中, 活性在细胞例如细菌细胞、植物细胞或真核细胞中测量。在一些实施方案中, 活性在哺乳动物细胞 (如啮齿动物细胞或人类细胞) 中测量。在一些实施方案中, 活性在人类细胞 (如293T细胞) 中测量。在一些实施方案中, 活性在小鼠细胞 (如Hepa1-6细胞) 中测量。在一些实施方案中, 与参考Cas12b核酸酶相比, 相对于参考Cas12b核酸酶, 工程化Cas12b核酸酶具有至少约10%、20%、30%、40%、50%、60%、70%、80%、90%、1倍、1.2倍、1.5倍、2倍、3倍、4倍、5倍、10倍、20倍、50倍或更多的更高活性中的任意一个。工程化Cas12b核酸酶的位点特异性核酸酶活性可以使用本领域

已知的方法来测量,包括,例如PCR、测序或凝胶迁移测定,如本文提供的实例中所述。

[0140] 在一些实施方案中,活性是细胞中的基因编辑活性。在一些实施方案中,细胞是细菌细胞、植物细胞或真核细胞。在一些实施方案中,细胞是哺乳动物细胞,如啮齿动物细胞或人类细胞。在一些实施方案中,细胞是293T细胞。在一些实施方案中,所述活性在小鼠细胞(如Hepa1-6细胞)中测量。在一些实施方案中,活性是在细胞中靶基因组位点处的插入/缺失形成活性,如通过工程化Cas12b核酸酶对靶核酸进行位点特异性切割和非同源末端连接(NHEJ)机制进行DNA修复。在一些实施方案中,活性是在细胞中的靶基因组位点插入外源核酸序列,例如,通过工程化Cas12b核酸酶对靶核酸进行位点特异性切割和同源重组(HR)机制进行DNA修复。在一些实施方案中,通过工程化Cas12b核酸酶切割后的同源重组还包括引入供体模板。在一些实施方案中,与参考Cas12b核酸酶相比,工程化Cas12b核酸酶在细胞(例如,人类细胞,如293T细胞或小鼠Hepa1-6细胞)的基因组位点处具有至少约20%(例如,至少约30%、40%、50%、60%、70%、80%、90%、1倍、1.2倍、1.5倍、2倍、3倍、4倍、5倍、10倍、20倍、50倍或更多中的任一个)增加的基因编辑活性(例如,插入/缺失形成)。在一些实施方案中,工程化Cas12b核酸酶能够编辑比参考Cas12b核酸酶更多数量(例如,2、3、4、5、10、20、50、100个或更多个)的基因组位点。在一些实施方案中,工程化Cas12b核酸酶的共有PAM序列与参考Cas12b核酸酶相同。在一些实施方案中,与参考Cas12b核酸酶相比,工程化Cas12b核酸酶识别更多(例如,1、2、3、4、5、10、20、50、100个或更多个)PAM序列。

[0141] 可使用本领域已知的任何方法来确定细胞中工程化Cas12b核酸酶的切割或基因编辑效率,包括,例如T7内切酶1(T7E1)测定、PCR、靶DNA测序(包括,例如,Sanger序列和第二代测序)、缺失追踪插入和缺失(TIDE)测定或通过扩增子分析进行插入/缺失检测(IDAA)测定。参见例如Sentmanat MF et al., "A survey of validation strategies for CRISPR-Cas9 editing," *Scientific Reports*, 2018, 8, article number 888, 其内容通过引用以其整体并入本文。在一些实施方案中,例如,如本文实施例中所述,使用靶向的下一代测序(NGS)来测量细胞中工程化Cas12b核酸酶的基因编辑效率。用于确定工程化Cas12b核酸酶的切割或基因编辑效率的示例性基因组位点包括但不限于CCR5、AAVS、CD34、RNF2、SCN9A、HBG1/2和EMX1。在一些实施方案中,与人类细胞基因组中参考Cas12b核酸酶的平均切割或基因编辑效率相比,工程化Cas12b核酸酶可以切割或编辑至少约1、2、5、10、15、20、25、30、35、40、45、50、55、60、65、100个或更多个基因座。在一些实施方案中,工程化Cas12b核酸酶的切割或基因编辑效率(例如,插入/缺失率)至少约为参考Cas12b核酸酶的10%、20%、30%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、1倍、2倍、5倍、10倍、20倍、50倍或更高。

[0142] 工程化Cas12b效应子蛋白

[0143] 本申请还提供基于本文所述的任一种工程化Cas12b核酸酶、变体(例如,dCas12b)或功能衍生物的工程化Cas12b效应子蛋白。在一些实施方案中,工程化Cas12b效应子蛋白包含(或由其组成,或基本上由其组成)本文所述的任一种工程化Cas12b核酸酶、变体或功能衍生物。在一些实施方案中,工程化Cas12b效应子蛋白包含工程化Cas12b核酸酶的功能衍生物,如下面“功能衍生物”部分中描述的任何一种功能衍生物。

[0144] 在一些实施方案中,工程化Cas12b效应子蛋白具有酶促活性。在一些实施方案中,工程化Cas12b效应子蛋白是切割靶双链核酸(例如,双链DNA)的两条链的核酸酶。在一些实

实施方案中,工程化Cas12b效应子蛋白是切口酶,即切割靶双链核酸(例如,双链DNA)的单链。在一些实施方案中,工程化Cas12b效应子蛋白包含工程化Cas12b核酸酶(dCas12b)的酶促失活突变体。Cas12b核酸酶活性位点中一个或多个氨基酸残基的突变可导致酶促死亡的Cas12b(dCas12b)。例如,AaCas12b(SEQ ID NO:1)的D570A、E848A、R785A、E848A、R911A和/或D977A突变体在人类细胞中显著降低(例如,降低至少约60%、70%、80%、90%、95%或更多中的任一)或没有核酸酶活性。例如,参见Teng F.et al.,Cell Discovery,4,Article number:63(2018),其内容通过引用以其整体并入本文。在一些实施方案中,工程化Cas12b效应子蛋白包含具有对应于AaCas12b的D570A、E848A、R785A、E848A、R911A和D977A的一个或多个突变的工程化Cas12b。在一些实施方案中,将选自下组的一个或多个突变进一步引入包含Q119F+E475R+E758R突变的AaCas12b:D570A、E848A、R785A、E848A、R911A和D977A。在一些实施方案中,工程化Cas12b核酸酶的酶促失活突变体包含SEQ ID NO:79-81中任一个的氨基酸序列。在一些实施方案中,工程化Cas12b效应子蛋白包含具有与AaCas12b的R785A突变相对应的突变的工程化的Cas12b。在一些实施方案中,工程化Cas12b效应子蛋白包含具有与AaCas12b的R911A突变相对应的突变的工程化的Cas12b。在一些实施方案中,工程化Cas12b效应子蛋白包含具有与AaCas12b的D977A突变相对应的突变的工程化的Cas12b。在一些实施方案中,工程化Cas12b效应子蛋白包含具有与AaCas12b的E848A突变相对应的突变的工程化的Cas12b。在一些实施方案中,工程化Cas12b效应子蛋白包含具有与AaCas12b的D570A突变相对应的突变的工程化的Cas12b。在一些实施方案中,工程化Cas12b效应子蛋白包含具有与AaCas12b的D570A+E848A突变或AaCas12b的D570A+D977A突变相对应的突变的工程化的Cas12b。

[0145] 在一些实施方案中,提供了一种工程化Cas12b切口酶。在一些实施方案中,提供了一种工程化的Cas12b融合效应子蛋白,其包含工程化Cas12b核酸酶或其变体或功能衍生物(例如,工程化Cas12b核酸酶的酶促失活突变体,如SEQ ID NO:79-81中的任一个),其融合至功能结构域,如翻译启动结构域、转录阻遏物结构域(例如,Krüppel相关盒(KRAB)结构域)、反式激活结构域、表观遗传修饰结构域、核碱基编辑结构域(例如,胞嘧啶碱基编辑器(CBE)或腺嘌呤碱基编辑器(ABE)结构域)、逆转录酶结构域、报告子结构域(例如,荧光结构域)或核酸酶结构域(例如,ZFN结构域)。在一些实施方案中,提供了一种工程化的Cas12b碱基编辑器,其包含与胞嘧啶脱氨酶结构域或腺苷脱氨酶结构域融合的本文所述的任一种工程化Cas12b核酸酶(例如,SEQ ID NO:79-81中的任一个)的催化无活性变体。在一些实施方案中,提供了一种工程化的Cas12b碱基编辑器,其包含与KRAB结构域或其功能片段(如ZIM3KRAB结构域(SEQ ID NO:72))融合的本文所述的任一种工程化Cas12b核酸酶(例如,SEQ ID NO:79-81中的任一个)的催化无活性变体。在一些实施方案中,提供了一种工程化的Cas12b先导编辑器,其包含与逆转录酶结构域融合的本文所述的任一种工程化Cas12b核酸酶(例如,SEQ ID NO:79-81中的任一个)的催化无活性变体。在一些实施方案中,提供了一种断裂Cas12b效应子蛋白系统。

[0146] 变体/功能衍生物

[0147] 本申请还提供本文描述的任意工程化Cas12b核酸酶的变体和功能衍生物。在一些实施方案中,提供了一种工程化Cas12b效应子蛋白,其包含(或由其组成,或基本由其组成)本文所述的任何工程化Cas12b核酸酶的功能变体。在一些实施方案中,与相应的工程化

Cas12b核酸酶(例如,SEQ ID NO:2-22中的任一个)的氨基酸序列相比,功能变体的氨基酸序列具有至少一个氨基酸残基的差异(例如,具有缺失、插入、取代和/或融合)。在一些实施方案中,功能变体具有一个或多个突变,如氨基酸取代、插入和/或缺失。例如,与工程化Cas12b核酸酶相比,功能变体可包含1、2、3、4、5、6、7、8、9、10个或更多个氨基酸取代中的任一个。在一些实施方案中,一个或多个取代是保守取代。在一些实施方案中,功能变体具有工程化Cas12b核酸酶的所有结构域。在一些实施方案中,功能变体不具有工程化Cas12b核酸酶的一个或多个结构域。

[0148] 对于本文所述的任何Cas12b变体蛋白(例如,切口酶Cas12b蛋白、失活或催化失活的Cas12b(dCas12b)、融合Cas12b),所述Cas12b变体可包含与本文所述的任意Cas12b蛋白序列相同的参数(例如,结构域、序列同一性百分比等)。

[0149] Cas12b功能变体中的示例性突变描述于W02016205764、W02016205749和W02020/087631中,其各自的内容通过引用以其整体并入本文。

[0150] 催化活性

[0151] 在一些实施方案中,功能变体与工程化Cas12b核酸酶的非突变形式相比具有不同的催化活性。在一些实施方案中,突变(例如,氨基酸取代、插入和/或缺失)位于Cas12b效应子蛋白的催化结构域(例如,RuvC结构域)中。在一些实施方案中,变体包含多个催化域中的突变。切割双链靶核酸的一条链但不切割另一条链的Cas12b效应子蛋白在本文中称为“切口酶”(例如,“切口酶Cas”)。在一些实施方案中,工程化Cas12b效应子蛋白包含(或由其组成,或基本由其组成)工程化Cas12b核酸酶的切口酶突变体。基本上没有核酸酶活性的Cas12b蛋白在本文中被称为死亡的Cas12b蛋白(“dCas12b”) (但需要注意的是,在融合Cas12b效应子蛋白的情况下,核酸酶活性可以由异源多肽(融合伴侣)提供,这将在下文中更详细地描述)。在一些实施方案中,当突变的Cas12b的DNA切割活性比其未突变形式低约25%、20%、10%、5%、1%、0.1%、0.01%或更低中的任一个时,Cas12b效应子蛋白被认为基本上缺乏所有DNA切割活性。

[0152] 在一些实施方案中,工程化Cas12b核酸酶是dCas12b。在一些实施方案中,工程化的Cas12b功能变体包含对应于AaCas12b(SEQ ID NO:1)的D570A的突变。在一些实施方案中,工程化的Cas12b功能变体包含对应于AaCas12b的E848A的突变。在一些实施方案中,工程化的Cas12b功能变体包含对应于AaCas12b的R785A的突变。在一些实施方案中,工程化的Cas12b功能变体包含对应于AaCas12b的E848A的突变。在一些实施方案中,工程化的Cas12b功能变体包含对应于AaCas12b的R991A的突变。在一些实施方案中,工程化的Cas12b功能变体包含对应于AaCas12b的D977A的突变。在一些实施方案中,工程化的Cas12b功能变体包含对应于BthCas12b的D573A的突变。在一些实施方案中,AaCas12b的无催化活性或基本无活性的变体(Q119F+E475R+E758R)还包含选自下组的一个或多个取代:D570A、E848A和D977A,其中所述氨基酸位置对应于SEQ ID NO:22。在一些实施方案中,dCas12b包含SEQ ID NO:79-81中任一个的氨基酸序列。

[0153] 断裂Cas12b效应子蛋白

[0154] 本文描述的CRISPR-Cas12b系统可包含任意一对多肽(本文中也称为“断裂Cas12b多肽”),其包含本节中的断裂Cas12b部分。示例性的断裂Cas12b蛋白系统已在例如PCT/CN2020/111057和PCT/CN2021/114339中进行了描述,其各自的内容通过引用以其整体并入

本文。

[0155] 在一些实施方案中,提供了一种断裂Cas12b效应子蛋白,其包含第一多肽和第二多肽,所述第一多肽包含本文所述的任一种工程化Cas12b核酸酶或其变体或功能衍生物(在本节中也称为“亲本Cas12b蛋白”)的N末端部分,所述第二多肽包含工程化Cas12b核酸酶或其变体或功能衍生物的C末端部分,其中在包含指导序列的指导RNA的存在下,所述第一多肽和第二多肽能够彼此缔合以形成CRISPR复合物,所述CRISPR复合物特异性结合包含与指导序列互补的靶序列的靶核酸。在一些实施方案中,所述第一多肽和第二多肽各自包含二聚化结构域。在一些实施方式中,所述第一二聚化结构域和第二二聚化结构域在诱导剂(例如,雷帕霉素)的存在下彼此缔合。在一些实施方案中,所述第一多肽和第二多肽不包含任何二聚化结构域。在一些实施方案中,断裂Cas12b效应子蛋白是自我诱导的。

[0156] 断裂Cas12b部分是基于本文所述的任一种工程化Cas12b核酸酶、或其变体或功能变体而设计的。

[0157] Cas12b蛋白具有多种结构域。在一些实施方案中,亲本Cas12b蛋白从N末端到C末端包含:第一WED结构域(WED-I;也称为OBD-I结构域)、第一REC结构域(REC1)、第二WED结构域(WED-II;也称为OBD-II结构域)、第一RuvC结构域(RuvC-I)、桥螺旋(BH)结构域、第二RuvC结构域(RuvC-II)、第一Nuc结构域(Nuc-I;也称为UK-I结构域)、第三RuvC结构域(RuvC-III)和第二Nuc结构域(Nuc-II;也称为UK-II结构域)。可以使用本领域中已知的方法确定结构域边界,如基于天然存在的Cas12b蛋白的晶体结构(例如,AaCas12b的PDB ID号:5U30、5U31、5U33、5U34和5WQE),和/或与亲本Cas12b蛋白中已知功能结构域的序列同源性。在一些实施方案中,AaCas12b具有以下结构域:WEB-I结构域(氨基酸残基1-14)、REC1结构域(氨基酸残基15-386)、WED-II结构域(氨基酸残基387-518)、RuvC-I结构域(氨基酸残基519-628)、BH结构域(氨基酸残基629-658)、REC2结构域(氨基酸残基659-783)、RuvC-II结构域(氨基酸残基784-900)、Nuc-I结构域(氨基酸残基901-974)、RuvC-III结构域(氨基酸残基975-993)和Nuc-II结构域(氨基酸残基994-1129),其中氨基酸基于SEQ ID NO:1编号。

[0158] 工程化Cas12b核酸酶或其变体或功能衍生物被断裂,因为两个断裂Cas12b部分基本上包含功能性Cas12b。Cas12b可以作为基因组编辑酶(当与靶DNA和指导RNA形成复合物时)发挥作用,如切割双链核酸的单链或双链的核酸酶;或者它可以是催化死亡的Cas12b(dCas12b),其本质上是一种DNA结合蛋白,由于其催化结构域的典型突变,因此具有很少或没有催化活性。参考Cas12b活性位点中一个或多个氨基酸残基的突变可导致催化死亡的Cas12b,如AaCas12b的D570A、E848A、R785A、E848A、R911A和/或D977A突变体。

[0159] 本文所述的断裂Cas12b部分可以通过将工程化Cas12b核酸酶或其变体或功能衍生物(本文中称为“亲本Cas12b蛋白”;如SEQ ID NO:2-22和79-81中的任一个)(例如,全长Cas12b蛋白或其功能变体)在断裂位置分成两半来设计,该断裂位置是亲本Cas12b蛋白的N末端部分与C末端部分分离的点。在一些实施方案中,所述N末端部分包含氨基酸残基1至X,而所述C末端部分包含氨基酸残基X+1至亲本Cas12b蛋白的C末端。在这个实例中,编号是连续的,但这可并不总是必要的,因为可以从任一断裂端的末端修剪氨基酸(或编码它们的核苷酸),和/或还考虑在多肽链的内部区域进行突变(例如,插入、缺失和取代),前提是保留了重建的Cas12b蛋白的足够的DNA结合活性,并且如果需要,保留DNA切口酶或双链切割活

性,例如与亲本Cas12b蛋白相比至少约30%、40%、50%、60%、70%、80%、90%、95%或更高的活性。

[0160] 相对于本文描述的工程化Cas12b核酸酶,还考虑了具有一些N-和/或C-末端截断或缺失、和/或内部突变的断裂Cas12b部分。本领域技术人员可以容易地使用本文所述的示例性断裂Cas12b多肽的信息来基于其他Cas12b蛋白和功能变体设计对应的断裂Cas12b多肽,例如,通过使用标准序列比对工具。

[0161] 断裂位置可以位于柔性区内,如环内。优选地,断裂位置发生在氨基酸序列的中断不会导致结构特征(例如, α 螺旋或 β 折叠)的部分或完全破坏的地方。非结构化区域(不会在晶体结构中显示的区域,因为这些区域不够结构化从而不能在晶体中“冻结”)通常是优选的选择。考虑可以在亲本Cas12b蛋白表面暴露的非结构化区域进行断裂。

[0162] 在一些实施方案中,亲本Cas12b蛋白未在参与与指导RNA和/或靶RNA相互作用的氨基酸残基处或其附近(例如,在约10、8、6、5、4、3、2或1个氨基酸残基内)断裂。例如,AaCas12b蛋白的氨基酸残基4-9、118-122、143-144、442-446、573-574、742-746、753-754、792-796、800-819、835-839、897-900和973-978参与与单指导RNA和/或靶DNA的相互作用,其中基于SEQ ID NO:1编号。

[0163] 在一些实施方案中,亲本Cas12b蛋白在对应于AaCas12b蛋白的氨基酸残基516至793的氨基酸残基内的氨基酸残基处断裂,其中编号基于SEQ ID NO:1而定。在一些实施方案中,亲本Cas12b蛋白在与WED-II结构域和RuvC-I结构域接壤的氨基酸残基处断裂。在一些实施方案中,亲本Cas12b蛋白在对应于AaCas12b蛋白的氨基酸残基516至519的氨基酸残基内的氨基酸残基处断裂,其中编号基于SEQ ID NO:1而定。在一些实施方案中,亲本Cas12b蛋白在与BH结构域和REC2结构域接壤的氨基酸残基处断裂。在一些实施方案中,亲本Cas12b蛋白在对应于AaCas12b蛋白的氨基酸残基621至627的氨基酸残基内的氨基酸残基处断裂,其中编号基于SEQ ID NO:1而定。在一些实施方案中,亲本Cas12b蛋白在与REC2结构域和RuvC-II结构域接壤的氨基酸残基处断裂。在一些实施方案中,亲本Cas12b蛋白在对应于AaCas12b蛋白的氨基酸残基777至793的氨基酸残基内的氨基酸残基处断裂,其中编号基于SEQ ID NO:1而定。在一些实施方案中,亲本Cas12b蛋白在RCE2结构域内断裂。在一些实施方案中,亲本Cas12b蛋白在对应于AaCas12b蛋白的659至664、676至684或702至706氨基酸残基的氨基酸残基内的氨基酸残基处断裂,其中编号基于SEQ ID NO:1而定。

[0164] 在一些实施方案中,亲本Cas12b蛋白在距对应于AaCas12b蛋白的氨基酸残基518的氨基酸残基不超过约20个(例如,不超过约18、16、14、12、10、8、7、6、5、4、3、2或1个中的任一个)氨基酸残基的氨基酸残基处断裂,其中编号基于SEQ ID NO:1而定。在一些实施方案中,亲本Cas12b蛋白在对应于AaCas12b蛋白的氨基酸残基518的氨基酸残基处被断裂,其中编号基于SEQ ID NO:1而定。在一些实施方案中,亲本Cas12b蛋白在距对应于AaCas12b蛋白的氨基酸残基658的氨基酸残基不超过约20个(例如,不超过约18、16、14、12、10、8、7、6、5、4、3、2或1个中的任一个)氨基酸残基的氨基酸残基处断裂,其中编号基于SEQ ID NO:1而定。在一些实施方案中,亲本Cas12b蛋白在对应于AaCas12b蛋白的氨基酸残基658的氨基酸残基处断裂,其中编号基于SEQ ID NO:1而定。在一些实施方案中,亲本Cas12b蛋白在距对应于AaCas12b蛋白的氨基酸残基783的氨基酸残基不超过约20个(例如,不超过约18、16、14、12、10、8、7、6、5、4、3、2或1个中的任一个)氨基酸残基的氨基酸残基处断裂,其中编号基

于SEQ ID NO:1而定。在一些实施方案中,亲本Cas12b蛋白在对应于AaCas12b蛋白的氨基酸残基783的氨基酸残基处断裂,其中编号基于SEQ ID NO:1而定。

[0165] 在一些实施方案中,亲本Cas12b蛋白的N末端部分包含AaCas12b蛋白的WED-I、REC1、WED-II、RuvC-I和BH结构域,并且其中亲本Cas12b蛋白的C末端部分包含AaCas12b蛋白的REC2、RuvC-II、Nuc-I、RuvC-III和Nuc-II结构域。在一些实施方案中,亲本Cas12b蛋白的N末端部分包含亲本Cas12b蛋白的第1至658位氨基酸残基,亲本Cas12b蛋白的C末端部分包含亲本Cas12b蛋白的第659至1129位氨基酸残基,其中氨基酸残基按照SEQ ID NO:1编号。

[0166] 在一些实施方案中,亲本Cas12b蛋白的N末端部分包含亲本Cas12b蛋白的WED-I、REC1、WED-II、RuvC-I、BH和REC2结构域,并且其中亲本Cas12b蛋白的C末端部分包含亲本Cas12b蛋白的RuvC-II、Nuc-I、RuvC-III和Nuc-II结构域。在一些实施方案中,亲本Cas12b蛋白的N末端部分包含亲本Cas12b蛋白的第1至783位氨基酸残基,亲本Cas12b蛋白的C末端部分包含亲本Cas12b蛋白的第784至1129位氨基酸残基,其中氨基酸残基按照SEQ ID NO:1编号。

[0167] 在一些实施方案中,亲本Cas12b蛋白的N末端部分包含亲本Cas12b蛋白的WED-I、REC1、WED-II、RuvC-I和BH结构域,其中亲本Cas12b蛋白的C末端部分包含亲本Cas12b蛋白的RuvC-II、Nuc-I、RuvC-III和Nuc-II结构域,并且其中亲本Cas12b蛋白的REC2结构域在亲本Cas12b蛋白的N末端部分和亲本Cas12b蛋白的C末端部分之间断裂。

[0168] 在一些实施方案中,亲本Cas12b蛋白的N末端部分包含亲本Cas12b蛋白的WED-I、REC1和WED-II结构域,并且其中亲本Cas12b蛋白的C末端部分包含亲本Cas12b蛋白的RuvC-I、BH、REC2、RuvC-II、Nuc-I、RuvC-III和Nuc-II结构域。在一些实施方案中,亲本Cas12b蛋白的N末端部分包含亲本Cas12b蛋白的第1至518位氨基酸残基,亲本Cas12b蛋白的C末端部分包含亲本Cas12b蛋白的第519至1129位氨基酸残基,其中氨基酸残基按照SEQ ID NO:1编号。

[0169] 断裂点通常在计算机上设计并克隆到构建体中。两个断裂Cas12b部分,即N末端和C末端部分,一起形成功能性Cas12b蛋白,其优选包含亲本Cas12b蛋白的至少约70%或更多的氨基酸序列,如亲本Cas12b蛋白的至少约75%、80%、85%、90%、95%、98%、99%或更多的氨基酸序列中的任一个。预计有一些修剪和突变。非功能性域可被完全删除。对于所有断裂Cas12b系统,可将两个断裂Cas12b部分放在一起,并恢复或重建所希望的Cas12b功能。可以使用本领域已知的方法来评估重建的Cas12b蛋白或CRISPR复合物(Cas12b+指导RNA复合物)的活性。例如,可以使用T7核酸内切酶I(T7EI)测定来评估细胞内的核酸酶活性。基因编辑活性也可以通过DNA测序来评估。

[0170] 在一些实施方案中,亲本Cas12b蛋白断裂成两个以上的部分,如3、4、5或6个部分。

[0171] 断裂Cas12b效应子蛋白可各自包含一个或多个二聚化结构域。在一些实施方案中,所述第一多肽包含与第一断裂Cas12b效应子部分融合的第一二聚化结构域,并且所述第二多肽包含与第二断裂Cas12b效应子部分融合的第二二聚化结构域。所述二聚化结构域可经由肽接头(例如,柔性肽接头,如GS接头)或化学键与断裂Cas12b效应部分融合。在一些实施方案中,二聚化结构域与断裂的Cas12b效应子部分的N末端融合。在一些实施方案中,二聚化结构域与断裂Cas12b效应部分的C末端融合。

[0172] 在一些实施方案中,断裂Cas12b效应子蛋白不包含任何二聚化结构域。

[0173] 在一些实施方案中,所述二聚化结构域促进两个断裂Cas12b效应部分的缔合。在一些实施方案中,断裂Cas12b效应部分通过诱导剂诱导缔合或二聚化为功能性Cas12b效应子蛋白。在一些实施方案中,断裂Cas12b效应子蛋白包含可诱导的二聚化结构域。在一些实施方案中,二聚化结构域不是可诱导的二聚化结构域,即,二聚化结构域在没有诱导剂存在的情况下二聚化。

[0174] 诱导剂可以是诱导能量源或除指导RNA之外的诱导分子(例如,sgRNA)。诱导剂经由诱导二聚化结构域的二聚化将两个断裂Cas12b效应子部分重组成功能性Cas12b效应子蛋白起作用。在一些实施方案中,诱导剂通过可诱导的二聚化结构域的诱导缔合作用将两个断裂Cas12b效应子部分结合到一起。在一些实施方案中,如果没有诱导剂,两个断裂Cas12b效应子部分就不会相互缔合以重建为功能性Cas12b效应子蛋白。在一些实施方案中,在没有诱导剂的情况下,两个断裂Cas12b效应子部分可相互缔合以在指导RNA(例如,sgRNA)存在下重组成功能性Cas12b效应子蛋白。

[0175] 本申请的诱导剂可以是热量、超声波、电磁能或化合物。在一些实施方案中,诱导剂是抗生素、小分子、激素、激素衍生物、类固醇或类固醇衍生物。在一些实施方案中,诱导剂是脱落酸(ABA)、多西环素(DOX)、cumate、雷帕霉素、4-羟基他莫昔芬(4OHT)、雌激素或蜕皮激素。在一些实施方案中,断裂Cas12b效应子系统是选自下组的诱导剂控制的系统:基于抗生素的诱导系统、基于电磁能的诱导系统、基于小分子的诱导系统、基于核受体的诱导系统和基于激素的诱导系统。在一些实施方案中,断裂Cas12b效应子系统是诱导物控制的系统,其选自下组:四环素(Tet)/DOX诱导系统、光诱导系统、ABA诱导系统、cumate阻遏物/操纵子系统、4OHT/雌激素诱导系统、基于蜕皮激素的诱导系统和FKBP12/FRAP(FKBP12-雷帕霉素复合物)诱导系统。这样的诱导剂也在本文中和PCT/US2013/051418中进行了讨论,其内容通过引用以其整体并入本文。FRB/FKBP/雷帕霉素系统已在Paulmurugan and Gambhir, *Cancer Res*, August 15, 2005 65;7413; and Crabtree et al., *Chemistry & Biology* 13, 99-107, Jan 2006中描述,其各自的内容均通过引用以其整体并入本文。

[0176] 在一些实施方案中,断裂Cas12b效应子蛋白对是分开的并且无活性的,直到诱导二聚化结构域(例如,FRB和FKBP)的二聚化,这导致功能性Cas12b效应子核酸酶的重新组装。在一些实施方案中,包含可诱导二聚体的第一半(例如,FRB)的第一断裂Cas12b效应子蛋白与包含可诱导二聚体的第二半(例如,FKBP)的第二断裂Cas12b效应子蛋白分开递送和/或分开定位。

[0177] 可用于本文描述的诱导剂控制的断裂Cas12b效应子系统的其他示例性基于FKBP的诱导系统包括但不限于:在FK506存在下与钙调神经磷酸酶A(CNA)二聚化的FKBP、在FKCsA存在下与CyP-Fas二聚化的FKBP、在雷帕霉素存在下与FRB二聚化的FKBP、在香豆霉素存在下与GryB二聚化的GyrB、在赤霉素存在下与GID1二聚化的GAI、或在HaXS存在下与HaloTag二聚化的Snap-tag。

[0178] 还考虑了FKBP家族本身的备选的方案。例如,在FK1012存在下FKBP发生同源二聚化(即,一个FKBP与另一个FKBP二聚)。

[0179] 在一些实施方案中,二聚化结构域是FKBP并且诱导剂是FK1012。在一些实施方案中,二聚化结构域是GryB并且诱导剂是香豆霉素。在一些实施方案中,二聚化结构域是ABA

并且诱导剂是赤霉素。

[0180] 在一些实施方案中,断裂Cas12b效应子部分可被自动诱导(即,自动激活或自我诱导)以在没有诱导剂存在的情况下缔合/二聚化为功能性Cas12b效应子蛋白。不受任何理论或假设的束缚,断裂Cas12b效应子部分的自动诱导可通过结合指导RNA(如sgRNA)来介导。在一些实施方案中,第一多肽和第二多肽不包含二聚化结构域。在一些实施方案中,第一多肽和第二多肽包含二聚化结构域。

[0181] 在一些实施方案中,本文所述的断裂Cas12b效应子系统(包含诱导剂控制的和自动诱导系统)的重建Cas12b效应子蛋白具有亲本Cas12b效应子蛋白的编辑效率的至少约70%(如至少约70%、75%、80%、85%、90%、95%、98%、99%或更高效率、或100%效率中的任一个)的编辑效率。

[0182] 在一些实施方案中,本文所述的诱导剂控制的断裂Cas12b效应系统的重建Cas12b效应子蛋白在不存在诱导剂的情况下(即,由于自动诱导)具有不超过亲本Cas12b效应子蛋白的编辑效率的约50%(如不超过约50%、45%、40%、35%、30%、25%、20%、15%、10%、5%或更低的效率,或0%效率中的任一个)。

[0183] 融合Cas12b效应子蛋白

[0184] 本申请还提供了工程化Cas12b效应子蛋白,其包含额外的蛋白质结构域和/或组分,如接头、核定位/输出序列、功能结构域和/或报告蛋白。

[0185] 在一些实施方案中,工程化Cas12b效应子蛋白是蛋白质复合物,除了工程化Cas12b核酸酶或其变体或功能衍生物的核酸靶向结构域之外,还包含一个或多个异源蛋白质结构域(例如,约或多于约1、2、3、4、5、6、7、8、9、10个或更多个结构域中的任一个)。在一些实施方案中,工程化Cas12b效应子蛋白是融合蛋白,其包含一个或多个异源蛋白质结构域(例如,约或多于约1、2、3、4、5、6、7、8、9、10个或更多个结构域中的任一个),其与工程化Cas12b核酸酶或其变体或功能衍生物融合。

[0186] 在一些实施方案中,本申请的工程化Cas12b效应子蛋白可以包含(例如,经由融合蛋白,如经由一个或多个肽接头,例如,GS肽接头等)或与一个或多个功能结构域缔合(例如,经由多个蛋白质的共表达)。在一些实施方案中,一个或多个功能结构域是酶促结构域。这些功能结构域可以具有多种活性,例如,DNA和/或RNA甲基化酶活性、去甲基化酶活性、转录激活活性、转录抑制活性、转录释放因子活性、组蛋白修饰活性、RNA切割活性、DNA切割活性、核酸结合活性和开关活性(例如,光诱导的)。在一些实施方案中,一个或多个功能结构域是反式激活结构域(即,反式激活结构域)或阻遏物结构域。在一些实施方案中,反式激活结构域或阻遏物结构域可以募集染色质修饰剂。在一些实施方案中,一个或多个功能结构域是组蛋白修饰结构域。在一些实施方案中,一个或多个功能结构域是转座酶结构域、HR(同源重组)机械结构域、重组酶结构域和/或整合酶结构域。在一些实施方案中,功能结构域是Krüppel相关盒(KRAB)、VP64、VP16、Fok1、P65、HSF1、MyoD1、生物素-APEX、APOBEC1、AID、PmCDA1、Tad1和M-MLV逆转录酶。在一些实施方案中,功能结构域选自下组:翻译启动子结构域、转录阻遏物结构域、反式激活结构域、表观遗传修饰结构域、核碱基编辑结构域(例如,CBE或ABE结构域)、逆转录酶结构域、报告子结构域(例如,荧光结构域)和核酸酶结构域。在一些实施方案中,功能结构域是KRAB结构域,如ZIM3的KRAB结构域。在一些实施方案中,KRAB结构域包含SEQ ID NO:72的氨基酸序列。

[0187] 在一些实施方案中,工程化Cas12b效应子蛋白中一个或多个功能结构域的定位允许功能结构域正确的空间取向以影响具有归因功能效应的靶。例如,如果功能结构域是转录激活因子(例如,VP16、VP64或p65),则该转录激活因子被放置在允许其影响靶转录的空间取向上。同样,定位转录阻遏物以影响靶的转录,而定位核酸酶(例如,Fok1)以切割或部分切割靶。在一些实施方案中,功能结构域(例如,KRAB结构域,如包含SEQ ID NO:72)位于工程化Cas12b效应子蛋白(例如,SEQ ID NO:79-81中的任一个,如SEQ ID NO:81)的N末端。在一些实施方案中,功能结构域(例如,KRAB结构域,如包含SEQ ID NO:72)位于工程化Cas12b效应子蛋白(例如,SEQ ID NO:79-81中的任一个,如SEQ ID NO:81)的C末端。在一些实施方案中,工程化Cas12b效应子蛋白包含N末端的第一功能结构域和C末端的第二功能结构域。在一些实施方案中,工程化Cas12b效应子蛋白包含与一个或多个功能结构域(例如,KRAB结构域)融合的本文所述的任一种工程化Cas12b核酸酶的催化失活突变体(例如,SEQ ID NO:79-81中的任一个)。

[0188] 在一些实施方案中,工程化Cas12b效应子蛋白是转录激活因子。在一些实施方案中,工程化Cas12b效应子蛋白包含与反式激活结构域融合的本文所述的任一种工程化Cas12b核酸酶的酶促失活变体(例如,SEQ ID NO:79-81中的任一个)。在一些实施方案中,反式激活结构域选自下组:VP64、p65、HSF1、VP16、MyoD1、HSF1、RTA、SET7/9及其组合。在一些实施方案中,反式激活结构域包含VP64、p65和HSF1。在一些实施方案中,工程化Cas12b效应子蛋白包含两个断裂Cas12b效应子多肽,每个多肽都与一个反式激活结构域融合。在一些实施方案中,工程化Cas12b效应子蛋白还包含一个或多个核定位序列(例如,SEQ ID NO:61、62和82中的任一个)。

[0189] 在一些实施方案中,工程化Cas12b效应子蛋白是转录阻遏物。在一些实施方案中,工程化Cas12b效应子蛋白包含与转录阻遏结构域(例如,KRAB)融合的本文所述的任一种工程化Cas12b核酸酶的酶促失活变体(例如,SEQ ID NO:79-81中的任一个)。在一些实施方案中,所述转录阻遏物结构域选自下组:Krüppel相关盒(KRAB)、EnR、NuE、NcoR、SID、SID4X及其组合。在一些实施方案中,工程化Cas12b效应子蛋白包含两个断裂Cas12b效应多肽,每个多肽都与转录阻遏物结构域融合。在一些实施方案中,所述工程化Cas12b效应子蛋白还包含一个或多个核定位序列(例如,SEQ ID NO:61、62和82中的任一个)。

[0190] 在一些实施方案中,工程化Cas12b效应子蛋白是碱基编辑器,如胞嘧啶编辑器或腺苷编辑器。在一些实施方案中,工程化Cas12b效应子蛋白包含与编辑核碱基编辑结构域(如胞嘧啶碱基编辑(CBE)结构域或腺苷碱基编辑(ABE)结构域)融合的本文所述的任一种工程化Cas12b核酸酶的酶促失活变体(例如,SEQ ID NO:79-81中的任一种)。在一些实施方案中,核碱基编辑结构域是编辑DNA的结构域。在一些实施方案中,核碱基编辑结构域具有脱氨酶活性。在一些实施方案中,核碱基编辑结构域是胞嘧啶脱氨酶结构域。在一些实施方案中,核碱基编辑结构域是腺苷脱氨酶结构域。基于Cas核酸酶的示例性碱基编辑器已在例如W02018/165629A1和W02019/226953A1中描述,其各自的内容通过引用以其整体并入本文。示例性的CBE结构域包括但不限于:活化诱导的胞嘧啶脱氨酶或AID(例如,hAID)、编辑载脂蛋白BmRNA的复合物或APOBEC(例如,大鼠APOBEC1、hAPOBEC3A/B/C/D/E/F/G)和PmCDA1。示例性的ABE结构域包括但不限于TadA、ABE8及其变体(参见,例如,Gaudelli et al.,2017,Nature 551:464-471;和Richter et al.,2020,Nature Biotechnology 38:883-

891;其各自内容通过引用以其整体并入本文)。在一些实施方案中,所述功能结构域是APOBEC1结构域,例如,大鼠APOBEC1结构域。在一些实施方案中,所述功能结构域是TadA结构域。在一些实施方案中,工程化Cas12b效应子蛋白还包含一个或多个核定位序列(例如,SEQ ID NO:61、62和82中的任一个)。

[0191] 在一些实施方案中,所述工程化Cas12b效应子蛋白是先导编辑器。基于Cas9的先导编辑器已在例如A.Anzalone et al.,*Nature*,2019,576(7785):149-157中描述,其内容通过引用以其整体并入本文。在一些实施方案中,工程化Cas12b效应子蛋白包含与逆转录酶结构域融合的本文所述的任一种工程化Cas12b核酸酶的切口酶变体。在一些实施方案中,功能结构域是逆转录酶结构域。在一些实施方案中,逆转录酶结构域是M-MLV逆转录酶,或其变体,例如,具有D200N、T306K、W313F、T330P和L603W中一个或多个突变的M-MLV逆转录酶。在一些实施方案中,提供了包含先导编辑器的工程化的CRISPR/Cas12b系统。在一些实施方案中,工程化的CRISPR/Cas12b系统还包含第二个Cas12b切口酶,例如,基于与先导编辑器相同的工程化Cas12b核酸酶。在一些实施方案中,工程化的CRISPR/Cas12b系统包含先导编辑器指导RNA(pegRNA),其包含引物结合位点和逆转录酶(RT)模板序列。

[0192] 在一些实施方案中,本申请提供了断裂Cas12b效应系统,其具有一个或多个(例如,1、2、3、4、5、6或更多)与一个或两个断裂Cas12b效应子部分缔合(即,结合或融合)的功能结构域。所述功能结构域可以作为第一和/或第二断裂Cas12b效应子蛋白的一部分提供,作为该构建体内的融合。功能结构域通常经由肽接头(如GS接头)与断裂Cas12b效应子蛋白中的其他部分(例如,断裂Cas12b效应部分)融合。该功能结构域可用于基于催化失活的Cas12b效应子来重新调整利用断裂Cas12b效应子系统的功能。

[0193] 在一些实施方案中,工程化Cas12b效应子蛋白包含一个或多个核定位序列(NLS)和/或一个或多个核输出序列(NES)。示例性的NLS序列包含,例如,PKKKRKV(SEQ ID NO:82)、PKKKRKVPG(SEQ ID NO:61)和ASPKKRKV(SEQ ID NO:62)。NLS和/或NES可操作地连接至工程化Cas12b效应子蛋白或工程化Cas12b效应子蛋白中的多肽链的N末端和/或C末端。

[0194] 在一些实施方案中,工程化Cas12b效应子蛋白可编码另外的成分,如报告蛋白。在一些实施方案中,工程化Cas12b效应子蛋白包含荧光蛋白,例如GFP。这样的系统可以允许基因组位点的成像(参见,例如,“Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System”Chen B et al.*Cell* 2013)。在一些实施方案中,工程化Cas12b效应子蛋白是可诱导的断裂Cas效应子系统,该系统可用于成像基因组位点。

[0195] 工程化的CRISPR-Cas12b系统

[0196] 还提供了工程化的CRISPR-Cas12b系统,其包含:(a)本文所述的任一种工程化Cas12b核酸酶或其变体或衍生物(例如,SEQ ID NO:2-22和79-81中的任一个)或工程化Cas12b效应子蛋白(例如,工程化Cas12b核酸酶、切口酶、断裂Cas12b蛋白、转录阻遏物、转录激活物、碱基编辑器或先导编辑器),或编码其的核酸;和(b)包含与靶核酸的靶序列互补的指导序列的指导RNA、或编码指导RNA的一个或多个核酸,其中所述工程化Cas12b核酸酶或工程化Cas12b效应子蛋白和指导RNA能够形成CRISPR复合物,所述复合物特异性结合包含靶序列的靶核酸并诱导靶核酸的修饰。在一些实施方案中,提供了一种工程化的CRISPR-

Cas12b系统,其包含:(a)工程化Cas12b核酸酶或其效应子蛋白,其相对于参考Cas12b核酸酶包含一种、两种或三种类型的突变,其中所述突变包含:(1)参考Cas12b核酸酶中与PAM相互作用的一个或多个氨基酸残基(例如,以下一个或多个位置:116、123、130、132、144、145、153、173、222、395、400和475)用带正电荷的氨基酸残基(例如,R、H、K)的取代;和/或(2)参考Cas12b核酸酶中参与打开DNA双链的一个或多个氨基酸残基(例如,以下一个或多个位置:118和119)用具有芳香环的氨基酸残基(例如,F、Y、W)的取代;和/或(3)参考Cas12b核酸酶的RuvC结构域中与ssDNA底物相互作用的一个或多个氨基酸残基(例如,以下一个或多个位置:300、301、304、329、636、639、647、682、757、758、761、764、768、852、854、856、857、858、860、862、863、865、866、867、869、938、956、957、958、994、1093和1097)用带正电荷的氨基酸残基(例如,R、H、K)或疏水性氨基酸残基(例如,F、Y、W、M)的取代,其中所述参考Cas12b核酸酶包含SEQ ID NO:1的氨基酸序列,或编码工程化Cas12b核酸酶或其效应子蛋白的核酸;和(b)包含与靶核酸的靶序列互补的指导序列的gRNA,或编码所述gRNA的核酸,其中所述工程化Cas12b核酸酶或其效应子蛋白和gRNA能够形成CRISPR复合物,所述复合物特异性结合包含靶序列的靶核酸并诱导所述靶核酸的修饰。在一些实施方案中,所述工程化的CRISPR-Cas12b系统包含编码工程化Cas12b核酸酶或其变体或衍生物或工程化Cas12b效应子蛋白和/或指导RNA的一个或多个核酸。在一些实施方案中,所述gRNA包含crRNA和tracrRNA。在一些实施方案中,工程化的CRISPR-Cas12b系统包含前体指导RNA阵列,其可被例如工程化Cas12b核酸酶或其变体或衍生物或工程化Cas12b效应子蛋白加工成多个crRNA。在一些实施方案中,所述gRNA是sgRNA。在一些实施方案中,所述sgRNA包含SEQ ID NO:23-53中任一个的支架序列。在一些实施方案中,工程化的CRISPR-Cas12b系统包含编码工程化Cas12b核酸酶或其变体或衍生物或工程化Cas12b效应子蛋白和/或指导RNA的一个或多个载体。在一些实施方案中,工程化Cas12b核酸酶或其变体或衍生物或工程化Cas12b效应子蛋白和/或指导RNA由一个或多个载体(例如,腺相关病毒(AAV)载体)编码。在一些实施方案中,工程化的CRISPR-Cas12b系统包含核糖核蛋白(RNP)复合物,所述复合物包含与指导RNA结合的工程化Cas12b核酸酶或其变体或衍生物或工程化Cas12b效应子蛋白。

[0197] 在一些实施方案中,提供了一种工程化的CRISPR-Cas12b系统,其包含:(a)包含SEQ ID NO:1的氨基酸序列的Cas12b核酸酶或其效应子蛋白(例如,切口酶、断裂Cas12b蛋白、转录阻遏物、转录激活物、碱基编辑器或先导编辑器(prime editor)),或本文所述的任一种工程化Cas12b核酸酶或其变体或衍生物(例如,SEQ ID NO:2-22和79-81中的任一种)或工程化Cas12b效应子蛋白(例如,切口酶、断裂Cas12b蛋白、转录阻遏物、转录激活物、碱基编辑器或先导编辑器),或编码其的核酸;和(b)包含与靶核酸的靶序列互补的指导序列的gRNA,或编码gRNA的核酸,其中所述gRNA包含工程化支架,所述支架包含SEQ ID NO:25-53中任一个的序列;其中i)所述Cas12b核酸酶或其效应子蛋白或工程化Cas12b核酸酶或其变体或衍生物或工程化Cas12b效应子蛋白、和ii)所述gRNA能够形成特异性结合靶核酸的CRISPR复合物,并诱导靶核酸的修饰。在一些实施方案中,提供了一种工程化的CRISPR-Cas12b系统,其包含:(a)包含SEQ ID NO:1的氨基酸序列的Cas12b核酸酶或其效应子蛋白(例如,切口酶、断裂Cas12b蛋白、转录阻遏物、转录激活物、碱基编辑器或先导编辑器)、或工程化Cas12b核酸酶或其效应子蛋白,其相对于参考Cas12b核酸酶包含一种、两种或三种类型的突变,其中所述突变包含:(1)参考Cas12b核酸酶中与PAM相互作用的一个或多个氨

氨基酸残基(例如,以下一个或多个位置:116、123、130、132、144、145、153、173、222、395、400和475)用带正电荷的氨基酸残基(例如,R、H、K)的取代;和/或(2)参考Cas12b核酸酶中参与打开DNA双链的一个或多个氨基酸残基(例如,以下一个或多个位置:118和119)用具有芳香环的氨基酸残基(例如,F、Y、W)的取代;和/或(3)参考Cas12b核酸酶的RuvC结构域中与ssDNA底物相互作用的一个或多个氨基酸残基(例如,以下一个或多个位置:300、301、304、329、636、639、647、682、757、758、761、764、768、852、854、856、857、858、860、862、863、865、866、867、869、938、956、957、958、994、1093和1097)用带正电荷的氨基酸残基(例如,R、H、K)或疏水性氨基酸残基(例如,F、Y、W、M)的取代,其中所述参考Cas12b核酸酶包含SEQ ID NO:1的氨基酸序列,或编码Cas12b核酸酶或其效应子蛋白的核酸;和(b)包含与靶核酸的靶序列互补的指导序列的gRNA,或编码所述gRNA的核酸,其中所述gRNA包含工程化支架,所述工程化支架包含SEQ ID NO:25-53中任一个的序列;其中i)所述Cas12b核酸酶或其效应子蛋白或工程化Cas12b核酸酶或其效应子蛋白、和ii)所述gRNA能够形成特异性结合靶核酸的CRISPR复合物,并诱导靶核酸的修饰。在一些实施方案中,提供了一种工程化的CRISPR-Cas12b系统,其包含:(a)包含SEQ ID NO:1-22和79-81中任一个的氨基酸序列的Cas12b核酸酶或Cas12b效应子蛋白,或编码其的核酸;和(b)包含与靶核酸的靶序列互补的指导序列的gRNA,或编码所述gRNA的核酸,其中所述gRNA包含工程化支架,所述工程化支架包含SEQ ID NO:25-53中任一个的序列;其中所述Cas12b核酸酶或Cas12b效应子蛋白和所述gRNA能够形成特异性结合靶核酸的CRISPR复合物,并诱导靶核酸的修饰。在一些实施例中,所述gRNA包含crRNA和tracrRNA,并且其中所述tracrRNA包含工程支架或其部分。在一些实施方案中,工程化的CRISPR-Cas12b系统包含编码多个crRNA的前体gRNA阵列。在一些实施方案中,所述gRNA是sgRNA。在一些实施方案中,工程化的CRISPR-Cas12b系统包含编码工程化Cas12b核酸酶、工程化Cas12b效应子蛋白、Cas12b核酸酶或Cas12b效应子蛋白的一个或多个载体。在一些实施方案中,所述一个或多个载体是AAV载体。在一些实施方案中,所述一个或多个载体进一步编码gRNA。

[0198] PAM

[0199] 在一些实施方案中,工程化Cas12b核酸酶或其变体或衍生物、工程化Cas12b效应子蛋白、Cas12b核酸酶或Cas12b效应子蛋白Cas12b识别包含(或由其组成)5'-TTN-3'序列(其中N为A、T、G或C)的PAM。在一些实施方案中,所述PAM包含或由5'-TTC-3'、5'-TTA-3'、5'-TTT-3'或5'-TTG-3'组成。

[0200] 指导RNA

[0201] 本申请的工程化的CRISPR-Cas12b系统可包含任何合适的指导RNA。指导RNA(gRNA)可包含能够与感兴趣的靶核酸中的靶序列(如细胞中感兴趣的基因组位点)杂交的指导序列(或间隔子)。在一些实施方案中,所述gRNA包含CRISPR RNA(crRNA)序列,所述CRISPR RNA(crRNA)序列包含指导序列的。在一些实施方案中,本文所述的crRNA包含直接重复(DR)序列和间隔子序列。在某些实施方案中,所述crRNA包含(基本由其组成、或由其组成)连接至指导序列或间隔子序列的直接重复序列。在某些实施方案中,所述直接重复序列可以位于指导序列或间隔子序列的上游(即,5')。在其他实施方案中,直接重复序列可以位于指导序列或间隔子序列的下游(即,3')。在一些实施方案中,所述crRNA包含直接重复序列、间隔子序列和直接重复序列(DR-间隔子-DR),这是典型的前体crRNA(pre-crRNA)构型。

在一些实施方案中,crRNA包含截短的直接重复序列和间隔子序列,这是典型的加工或成熟的crRNA。在一些实施方案中,crRNA包含突变的DR序列和间隔子序列。在一些实施方案中,gRNA包含反式激活CRISPR RNA(tracrRNA)序列。在一些实施方案中,tracrRNA与DR序列5'端的crRNA融合。在一些实施方案中,指导RNA是单指导RNA(sgRNA)。在一些实施方案中,gRNA或sgRNA包含tracrRNA和crRNA。在一些实施方案中,sgRNA包含SEQ ID NO:23-53中任一个的序列。在一些实施方案中,tracrRNA包含SEQ ID NO:23-53中任一个的序列或其部分。

[0202] 在一些实施方案中,gRNA包含在参考Cas12b蛋白的CRISPR基因座中天然不存在的非同源crRNA序列和/或tracrRNA序列。例如,AaCas12b、AkCas12b、AmCas12b、BhCas12b、BsCas12b、Bs3Cas12b、LsCas12b和SbCas12b的同源tracrRNA和crRNA序列,以及示例性sgRNA序列已在Teng F. et al., Cell Discovery (2019) 5:23的图S4和图S8中描述,其内容通过引用以其整体并入本文。

[0203] 在一些实施方案中,本文所述的CRISPR-Cas12b系统包含一个或多个gRNA(例如,crRNA、tracrRNA或sgRNA)(例如,1、2、3、4、5、10、15个或更多个),或编码其的核酸。在一些实施方案中,两个或更多个gRNA靶向不同的靶位点,例如,相同靶DNA或基因的2个靶位点,或2个不同靶DNA或基因的2个靶位点。

[0204] 本文所述的gRNA的序列和长度可以进行优化。在一些实施方案中,可以通过识别crRNA的加工形式或通过对crRNA的经验长度研究来确定gRNA的最佳长度。在一些实施方案中,gRNA包含碱基修饰,如在gRNA支架区域中。

[0205] 间隔子不需要完全互补,只要gRNA(例如,crRNA或sgRNA)具有足够的互补性以发挥作用(即,将Cas12b核酸酶(例如,工程化的)或其效应子蛋白指向靶位点)。可以通过引入一个或多个错配(例如,间隔子序列和靶序列之间的1或2个错配,包含沿着间隔子/靶序列的错配的位置)来调整由gRNA介导的Cas12b核酸酶(例如,工程化的)或其效应子蛋白的编辑或切割效率。当错配(如双重错配)位于间隔子的更中心位置(即,不在间隔子的3'或5'端)时,对切割效率具有更大的影响。因此,通过选择间隔子序列上错配的位置,可以调整Cas12b核酸酶(例如,工程化的)或其效应子蛋白的编辑或切割效率。例如,如果希望对靶标序列进行少于100%的编辑或切割(例如,在细胞的群体中),则可以把间隔子序列和靶标序列之间的1或2个错配引入到间隔子序列中。

[0206] 在一些实施方案中,指导序列或间隔子被设计为与靶序列具有至少一个错配,使得指导序列与靶序列之间形成的异源双链体包含与靶A相对的指导序列中的非配对C,或与靶C相对的指导序列中的非配对A,以在靶序列上进行脱氨基化(例如,进行碱基编辑)。在一些实施方案中,除了该A-C或C-A错配之外,当使用合适的比对算法进行最佳比对时,互补程度约为或超过约50%、60%、75%、80%、85%、90%、95%、97.5%、99%或更高。

[0207] 指导序列可具有合适的长度。在一些实施方案中,指导序列或间隔子序列的长度为约10nt至约50nt。在一些实施方案中,指导序列或间隔子序列的长度为至少约16个核苷酸,优选约16至约100个核苷酸,更优选约16至约50个核苷酸(例如,约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、50个核苷酸中的任一项)。在一些实施方案中,所述间隔子为约16至约27个核苷酸,如约17至约24个核苷酸、约18至约24个核苷酸、或约18至约22个核苷酸中的任一项。在

一些实施方案中,指导序列为约18-约35个核苷酸,包含例如,18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34或35个核苷酸中的任一项。

[0208] 在一些实施方案中,指导序列或间隔子序列与靶序列至少约60% (例如,至少约70%、75%、80%、85%、90%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%中的任一个) 互补。在一些实施方案中,间隔子序列与靶核酸(例如,DNA)的靶序列之间存在至少约15个(例如,至少约16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、35、40、45、50个或更多中的任一个) 碱基配对。

[0209] 最佳比对可以使用任何适合的序列比对算法来确定,其非限制性例子包含Smith-Waterman算法、Needleman-Wunsch算法、基于Burrows-Wheeler变换的算法(例如,Burrows Wheeler Aligner)、ClustalW、Clustal X、BLAT、Novoalign(Novocraft Technologies;可在www.novocraft.com获得)、ELAND(Illumina, San Diego, CA)、SOAP(可在soap.genomics.org.cn获得)、和Maq(可在maq.sourceforge.net获得)。可以通过任何合适的测定来评估指导序列(在靶向核酸的指导RNA内) 指导靶向核酸的复合物与靶核酸序列的序列特异性结合的能力。例如,可以将足以形成靶向核酸的复合物的靶向核酸的CRISPR系统的组分(包含待测试的指导序列) 提供给具有相应靶核酸序列的宿主细胞,如通过用编码靶向核酸的复合物组分的载体转染,然后评估靶核酸序列内的优先靶向(例如,切割),如通过本文所述的Surveyor测定。类似地,可通过提供靶核酸序列、靶向核酸的复合物的组分(包含待测试的指导序列和不同于测试指导序列的对照指导序列),并比较测试和对照指导序列反应之间在靶序列处的结合或切割速率,来在试管中评价靶核酸序列的切割。其他测定是可能的,并且将发生在本领域技术人员中。

[0210] 如本文所用,靶核酸与靶序列或靶核酸序列可互换使用,指包含与crRNA或gRNA中的全部或部分间隔子互补的核酸序列的特定核酸。在一些实例中,靶核酸包含基因或基因内的序列。在一些实例中,靶核酸包含非编码区(例如,启动子)。在一些实例中,靶核酸是单链的。在一些实例中,靶核酸是双链的。可以选择靶核酸以靶向任何靶核酸序列,如DNA或RNA序列(例如,mRNA)。

[0211] 靶核酸应该与PAM即CRISPR复合物识别的短序列缔合。根据CRISPR-Cas蛋白的性质,应选择靶序列,使得其在DNA双链中的互补序列(靶序列的互补序列) 位于PAM的上游或下游。在本申请的一个实施方案中,靶序列的互补序列位于PAM的下游或3'。对PAM的确切序列和长度的要求取决于所使用的Cas12b蛋白。

[0212] “tracrRNA”序列或类似术语包含与crRNA序列具有足够互补性以进行杂交的任意多核苷酸序列。在一些实施方案中,当最佳对齐时,tracrRNA序列和crRNA序列之间沿两者中较短序列的长度的互补程度约为或超过约25%、30%、40%、50%、60%、70%、80%、90%、95%、97.5%、99%或更高。在一些实施方案中,tracr序列的长度约为或超过约5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、40、50个或更多个核苷酸。在一些实施方案中,tracr序列和crRNA序列包含在单个转录本内,使得两者之间的杂交产生具有二级结构(如一个或多个发夹结构)的转录本。一般而言,互补的程度是参考指导序列和tracr序列沿着两个序列中较短序列的长度进行的最佳比对。最佳比对可以通过任何合适的比对算法来确定,并且可以进一步考虑二级结构。

[0213] 任何可以介导本文所述的Cas12b蛋白与相应gRNA(例如,crRNA) 结合的gRNA支架

或tracrRNA或DR序列均可用于本申请。在一些实施方案中,gRNA支架或tracrRNA或DR序列包含靠近5'或3'端(紧邻间隔子序列)的茎环结构。“茎环结构”是指具有二级结构的核酸,该二级结构包含已知或预测形成双链(茎)部分的核苷酸的区域,并且该区域在一端通过基本上单链核苷酸的连接区(环)连接。术语“发夹”结构在本文中也用于指茎环结构。这样的结构在本领域中是众所周知的,并且这些术语按照其在本领域中通常所知的含义使用。茎环结构不需要精确的碱基配对。因此,茎可包含一个或多个碱基错配。或者,碱基配对可能是精确的,即不包含任何错配。

[0214] 在一些实施方案中,gRNA支架或tracrRNA或DR是野生型支架或tracrRNA或DR的“功能变体”,如“功能上截短的版本”、“功能上延伸的版本”或“功能上替换的版本”。gRNA支架或tracrRNA或DR的“功能变体”是参考支架或tracrRNA或DR(例如,亲本DR)的5'和/或3'延伸的(功能上延伸的版本)或截短的(功能上截短的版本)变体,或包含相对于参考支架或tracrRNA或DR(例如,亲本DR)的一个或多个核苷酸的一个或多个插入、缺失和/或取代(功能替换版本),同时仍保留所述参考支架或tracrRNA或DR的至少约20%(如至少约30%、40%、50%、60%、60%、70%、80%、90%、95%或更高中的任一个)的功能性,即介导Cas12b核酸酶(例如,工程化的)或其效应子蛋白与相应的sgRNA或crRNA的结合的功能。gRNA支架或tracrRNA或DR功能变体通常保留可用于结合Cas12b核酸酶(例如,工程化的)或其效应子蛋白的茎环样二级结构或其部分。在一些实施方案中,gRNA支架或tracrRNA或DR或其功能变体包含至少2个(例如,2、3、4、5或更多个)可用于结合Cas12b核酸酶(例如,工程化的)或其效应子蛋白的茎环样二级结构或其部分。

[0215] 在一些实施方案中,DR或其功能变体包含至少约16个核苷酸(nt),例如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个或更多个核苷酸。在一些实施方案中,DR包含约20nt至约40nt,如约20nt至约30nt、约22nt至约40nt、约23nt至约38nt、约23nt至约36nt、或约30nt至约40nt。在一些实施方案中,DR包含22nt、23nt或24nt。在一些实施方案中,DR包含35nt、36nt或37nt。在一些实施方案中,sgRNA支架或其功能变体包含约50nt至约180nt,如约70nt至约140nt、或约90nt至约120nt中的任一项。

[0216] 在一些实施方案中,sgRNA包含支架序列,所述支架序列包含位于间隔子序列5'端附近的茎环结构(例如,1、2、3、4或更多个茎环)。在一些实施方案中,茎包含至少约4bp,其包含互补的X和Y序列,尽管也考虑更多(例如5、6、7、8、9、10、11或12个)或更少(例如,3、2个)碱基对的茎。因此,例如可以考虑X₂₋₁₀和Y₂₋₁₀(其中X和Y代表任何互补的核苷酸组)。在一些实施方案中,由X和Y核苷酸构成的茎与环一起会在整体二级结构中形成完整的发夹;并且,这可能是有优势的,并且碱基对的数量可以是形成完整发夹的任何量。在一些实施方案中,只要保留整个指导分子的二级结构,任何互补的X:Y碱基配对序列(例如,就长度而言)都是可以容忍的。在一些实施方案中,连接由X:Y碱基对构成的茎的环可以是相同长度(例如,4或5个核苷酸)或更长的任何序列,只要不中断指导分子的整体二级结构。在一些实施方案中,茎包含含有互补的X和Y序列的约5-7bp,尽管也考虑了更多或更少碱基对的茎。在一些实施方案中,考虑非沃森-克里克碱基配对,其中这样的配对通常会保留该位置的茎环结构。在一些实施方案中,支架序列中所含的茎包含(例如,由其组成)5对相互杂交的互补碱基,并且环长度为6、7、8或9个核苷酸。在一些实施方案中,茎可以包含至少2个、至少3个、至少4个或至少5个碱基对。在一些实施方案中,茎环结构包含长度为5个核苷酸的第

一茎核苷酸链；长度为5个核苷酸的第二茎核苷酸链，其中第一和第二茎核苷酸链可以相互杂交；和排列在所述第一和第二茎核苷酸链之间的环状核苷酸链，其中所述环状核苷酸链包含6、7或8个核苷酸。

[0217] 在一些实施方案中，所述指导分子的天然发夹或茎环结构被延伸或被延伸的茎环替换。在某些情况下已经证明，茎的延伸可以增强指导分子与CRISPR-Cas蛋白的组装 (Chen et al. *Cell*. (2013) ;155 (7) :1479-1491)。在一些实施方案中，茎环的茎延伸至少1、2、3、4、5或更多个互补碱基对 (即对应于在指导分子中添加2、4、6、8、10或更多个核苷酸)。在一些实施方案中，它们位于茎的末端，邻近茎环的环。

[0218] 如本文所用，两种或更多sgRNA或tracrRNA的二级结构基本上相同或基本上无差异是指这些sgRNA或tracrRNA含有长度相差不超过1、2或3个核苷酸的茎和/或环；就核苷酸类型 (A、U、G或C) 而言，这些sgRNA或tracrRNA的核苷酸序列当经序列比对比较时相差不超过1、2、3、4、5、6、7或8个核苷酸。在一些实施方案中，两个或多个sgRNA或tracrRNA的二级结构基本上相同或基本上没有差异是指所述sgRNA或tracrRNA含有最多一对互补碱基不同的茎，和/或最多一个核苷酸长度不同的环，和/或含有长度相同但碱基错配的茎。

[0219] 在一些实施方案中，可以将本申请的任何工程化Cas12b效应子蛋白引导至靶位点的gRNA支架序列包含一个或多个核苷酸变化，所述核苷酸变化选自下组：核苷酸添加、插入、缺失和取代，这些变化不会导致与SEQ ID NO:23-53中的任一个或其功能截短版本中所示的支架序列相比在二级结构上存在实质性差异。在一些实施方案中，gRNA支架包含SEQ ID NO:25-53中的任一个的序列，或其包含至多约10nt (例如，10、9、8、7、6、5、4、3、2或1nt) 差异的变体。

[0220] 在一些实施方案中，指导RNA包含crRNA。在一些实施方案中，工程化的CRISPR-Cas12b系统包含编码多个crRNA的前体指导RNA阵列。在一些实施方案中，Cas12b效应子蛋白切割前体指导RNA阵列以产生多个crRNA。在一些实施方案中，工程化的CRISPR-Cas12b系统包含编码多个crRNA的前体指导RNA阵列，其中每个crRNA包含不同的指导序列。在一些实施方案中，前体指导RNA阵列编码的crRNA与tracrRNA缔合。

[0221] 构建体和载体

[0222] 本文还提供了编码本文所述的任一种工程化Cas12b效应子蛋白 (包含工程化Cas12b核酸酶) 的构建体、载体和表达系统。在一些实施方案中，构建体、载体或表达系统还包含一个或多个gRNA (例如，sgRNA) 或crRNA阵列。

[0223] “载体”是包含分离的核酸并可用于将分离的核酸递送至细胞内部的物质的组合物。本领域已知多种载体，包括但不限于线性多核苷酸、与离子或两亲化合物相关的多核苷酸、质粒和病毒。一般而言，合适的载体含有在至少一个生物体中有功能的复制起点、启动子序列、方便的限制性内切酶位点和一个或多个可选择标记。术语“载体”还应解释为包含非质粒和非病毒化合物，其有助于将核酸转移到细胞中，如，例如，聚赖氨酸化合物、脂质体等。

[0224] 在一些实施方案中，载体是病毒载体。病毒载体的实例包括但不限于腺病毒载体、腺相关病毒载体、慢病毒载体、逆转录病毒载体、痘苗载体、单纯疱疹病毒载体，及其衍生物。在一些实施方案中，载体是噬菌体载体。病毒载体技术在本领域中是众所周知的，例如在 Sambrook et al. (2001, *Molecular Cloning: A Laboratory Manual*, Cold Spring

Harbor Laboratory, New York), 和其他病毒学和分子生物学手册中描述的。

[0225] 已经开发了许多基于病毒的系统用于将基因转移到哺乳动物细胞中。例如, 逆转录病毒为基因传递系统提供了便捷的平台。可以使用本领域已知的技术将异源核酸插入载体并包装在逆转录病毒颗粒中。然后可以分离重组病毒并将其递送至体外或离体的工程化的哺乳动物细胞。本领域已知多种逆转录病毒系统。在一些实施方案中, 使用腺病毒载体。本领域已知多种腺病毒载体。在一些实施方案中, 使用慢病毒载体。在一些实施方案中, 使用自失活慢病毒载体。

[0226] 在某些实施方案中, 载体是腺相关病毒 (AAV) 载体, 例如, AAV2、AAV8或AAV9, 其可以以单剂量施用, 所述单剂量含有至少 1×10^5 个颗粒 (也称为颗粒单位, pu) 的腺病毒或腺相关病毒。在一些实施方案中, 剂量为至少约 1×10^6 个颗粒、至少约 1×10^7 个颗粒、至少约 1×10^8 个颗粒、或至少约 1×10^9 个颗粒的腺相关病毒。递送方法和剂量在例如, WO 2016205764和美国专利号8,454,972中描述, 其各自内容通过引用以其整体并入本文中。

[0227] 在一些实施方案中, 所述载体是重组腺相关病毒 (rAAV) 载体。例如, 在一些实施方案中, 修饰的AAV载体可以用于递送。修饰的AAV载体可以基于几种衣壳类型中的一种或多种, 包括AAV1、AAV2、AAV5、AAV6、AAV8、AAV8.2、AAV9、AAV rh10、修饰的AAV载体 (例如, 修饰的AAV2、修饰的AAV3、修饰的AAV6) 和假型AAV (例如, AAV2/8、AAV2/5和AAV2/6)。可用于产生rAAV颗粒的示例性AAV载体和技术是本领域已知的 (参见, 例如, Aponte-Ubillus et al. (2018) Appl. Microbiol. Biotechnol. 102 (3): 1045-54; Zhong et al. (2012) J. Genet. Syndr. Gene Ther. S1:008; West et al. (1987) Virology 160:38-47 (1987); Tratschin et al. (1985) Mol. Cell. Biol. 5:3251-60); 美国专利号4,797,368和5,173,414; 和国际公开号WO 2015/054653and WO 93/24641, 其各自通过引用并入)。

[0228] 任何一个已知的用于递送Cas9和其他Cas12b蛋白的AAV载体均可用于递送本申请的工程化Cas12b核酸酶或效应子蛋白或系统。

[0229] 将载体引入哺乳动物细胞的方法是本领域已知的。所述载体可以通过物理、化学或生物方法转移到宿主细胞中。

[0230] 将载体引入宿主细胞的物理方法包括磷酸钙沉淀、脂质转染、粒子轰击、显微注射、电穿孔等。生产包含载体和/或外源核酸的细胞的方法是本领域所熟知的。参见, 例如, Sambrook et al. (2001) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, New York。在一些实施方案中, 所述载体通过电穿孔引入细胞中。

[0231] 把异源核酸引入宿主细胞的生物学方法包括使用DNA和RNA载体。病毒载体已成为将基因插入哺乳动物 (例如, 人类) 细胞的最广泛使用的方法。

[0232] 将载体引入宿主细胞的化学方法包括胶体分散系统, 如大分子复合物、纳米胶囊、微球、珠子、和基于脂质的系统, 包括水包油乳液、胶束、混合胶束和脂质体。用作体外递送载体的示例性胶体系统是脂质体 (例如, 人工膜囊泡)。在一些实施方案中, 工程化的CRISPR-Cas12b系统在纳米颗粒中以RNP递送。

[0233] 在一些实施方案中, 编码CRISPR-Cas12b系统或其组分的载体或表达系统包含一个或多个可选择的或可检测的标记, 其提供分离或有效选择含有和/或已被CRISPR-Cas12b系统修饰的细胞的方法, 例如, 在早期阶段和以大规模。

[0234] 报告子基因可以用于鉴定潜在转染的细胞并评价调控序列的功能。一般而言, 报

告子基因是一种在受体生物体或组织中不存在或不被其表达的基因,其编码的多肽的表达通过一些容易检测的特性(例如,酶活性)表现出来。在DNA被引入受体细胞后的适当时间测定报告子基因的表达。合适的报告子基因可包含编码荧光素酶、 β -半乳糖苷酶、氯霉素乙酰转移酶、分泌性碱性磷酸酶的基因,或绿色荧光蛋白基因(例如,Ui-Tei et al.FEBS Letters 479:79-82(2000))。

[0235] 确认宿主细胞中异源核酸的存在的方法包含,例如,本领域技术人员众所周知的分子生物学测定,如DNA印迹(Southern blotting)和RNA印迹(Northern blotting)、RT-PCR和PCR;生化测定,如检测特定肽的存在或不存在,例如,通过免疫学方法(如ELISA和蛋白质印迹)。

[0236] 在一些实施方案中,编码工程化Cas12b核酸酶或效应子蛋白和/或指导RNA的核酸序列可操作地连接到启动子。在一些实施方案中,启动子是相对于使用工程化的CRISPR-Cas12b系统工程化的细胞的内源性启动子。例如,可以使用本领域已知的任何方法将编码工程化Cas12b效应子蛋白的核酸敲入到内源启动子下游的工程化的哺乳动物细胞的基因组中。在一些实施方案中,所述内源启动子是丰富蛋白质(如 β -肌动蛋白)的启动子。在一些实施方案中,内源启动子是诱导型启动子,例如,可由工程化的哺乳动物细胞的内源活化信号诱导。在一些实施方案中,其中所述工程化的哺乳动物细胞是T细胞,所述启动子是T细胞活化依赖性启动子(如IL-2启动子、NFAT启动子或NF κ B启动子)。

[0237] 在一些实施方案中,所述启动子相对于使用工程化的CRISPR-Cas12b系统工程化的细胞来说是异源启动子。已经探索了多种启动子用于哺乳动物细胞中的基因表达,并且本领域已知的任何启动子都可以在本申请中使用。启动子大致可分为组成型启动子或受调控的启动子,如诱导型启动子。

[0238] 在一些实施方案中,编码工程化Cas12b效应子蛋白和/或指导RNA的核酸序列可操作地连接到组成型启动子。组成型启动子允许异源基因(也称为转基因)在宿主细胞中组成型表达。本文考虑的示例性组成型启动子包括但不限于巨细胞病毒(CMV)启动子、人类延伸因子-1 α (hEF1 α)、泛素C启动子(UbiC)、磷酸甘油激酶启动子(PGK)、猿猴病毒40早期启动子(SV40)和与CMV早期增强子(CAG)偶联的鸡 β -肌动蛋白启动子。在一些实施方案中,所述启动子是包含巨细胞病毒(CMV)早期增强子元件、启动子、鸡 β -肌动蛋白基因的第一外显子和第一内含子、和兔 β -珠蛋白基因的剪接受体的CAG启动子。

[0239] 在一些实施方案中,编码工程化的CRISPR-Cas12b蛋白和/或指导RNA的核酸序列可操作地连接到诱导型启动子。诱导型启动子属于调控型启动子类别。诱导型启动子可以由一个或多个条件诱导,如物理条件、微环境或宿主细胞的生理状态、诱导剂(即,诱导药剂),或其组合。在一些实施方案中,诱导条件选自下组:诱导剂、辐射(如电离辐射、光)、温度(如热)、氧化还原状态、肿瘤环境和要通过工程化的CRISPR-Cas12b系统进行工程化的细胞的活化状态。在一些实施方案中,启动子可由小分子诱导剂(如化合物)诱导。在一些实施方案中,所述小分子选自下组:强力霉素、四环素、酒精、金属或类固醇。化学诱导的启动子得到了最广泛的研究。这样的启动子包含其转录活性受小分子化学物质(如强力霉素、四环素、酒精、类固醇、金属和其他化合物)的存在或不存在调节的启动子。具有反式四环素控制的反式活化因子(rtTA)和四环素反应元件启动子(TRE)的强力霉素诱导系统是目前最为成熟的系统。W09429442描述了四环素反应启动子对真核细胞中基因表达的严格控制。

W09601313公开了四环素调节的转录调节剂。此外,Tet技术(如Tet-on系统)已在例如TetSystems.com的网站上进行了描述。任何已知的化学调控启动子均可用于驱动编码本申请的工程化的CRISPR-Cas12b蛋白和/或指导RNA的表达。

[0240] 在一些实施方案中,编码工程化Cas12b核酸酶或效应子蛋白的核酸序列是密码子优化的。在一些实施方案中,表达构建体编码可操作地连接到所述工程化Cas12b核酸酶或效应子蛋白的C末端的标签(例如,10xHis标签)。在一些实施方案中,每个工程化的断裂Cas12b构建体编码荧光蛋白,如GFP或RFP。报告蛋白可用于评估工程化的断裂Cas12b蛋白的共定位和/或二聚化,例如,使用显微镜。编码工程化Cas12b效应子蛋白的核酸序列可以使用编码自我裂解肽的序列与编码另外成分的核酸序列融合,所述自我裂解肽如T2A、P2A、E2A或F2A肽。

[0241] 在一些实施方案中,提供了用于哺乳动物细胞(例如,人类细胞)的表达构建体,其包含编码工程化Cas12b核酸酶或效应子蛋白的核酸序列。在一些实施方案中,所述表达构建体包含插入到pCAG-2A-eGFP载体中的编码工程化Cas12b核酸酶或效应子蛋白的密码子优化序列,使得所述Cas12b蛋白可操作地连接到eGFP。在一些实施方案中,提供第二载体用于在哺乳动物细胞(例如,人类细胞)中表达指导RNA(例如,sgRNA、crRNA或pre-crRNA阵列)。在一些实施方案中,编码指导RNA的序列在pUC19-U6-Aa-sgRNA载体骨架中表达。

[0242] 在一些实施方案中,编码Cas12b蛋白的核酸和编码gRNA的核酸位于不同的载体上。在一些实施方案中,编码Cas12b蛋白的核酸和编码gRNA的核酸位于同一载体上。在一些实施方案中,编码Cas12b蛋白的核酸和编码gRNA的核酸受不同启动子的控制,例如CMV启动子和U6启动子。在一些实施方案中,编码Cas12b蛋白的核酸位于编码gRNA的核酸的上游。在一些实施方案中,编码Cas12b蛋白的核酸位于编码gRNA的核酸的下游。在一些实施方案中,编码Cas12b蛋白的核酸和编码gRNA的核酸与靶核酸接触或同时引入细胞中。在一些实施方案中,编码Cas12b蛋白的核酸和编码gRNA的核酸与靶核酸接触或依次引入细胞中,如在编码gRNA的核酸之前引入编码Cas12b蛋白的核酸,或在编码gRNA的核酸之后引入编码Cas12b蛋白的核酸。在一些实施方案中,细胞已经表达Cas12b蛋白。在一些实施方案中,仅将编码gRNA的核酸引入细胞中。在一些实施方案中,细胞已经表达gRNA。在一些实施方案中,仅将编码Cas12b蛋白的核酸引入细胞。

[0243] III. 使用方法

[0244] 本申请的一方面提供了使用本文所述的任一种工程化Cas12b核酸酶或效应子蛋白或CRISPR-Cas12b系统在体外、离体或体内检测靶核酸或修饰核酸的方法,和使用所述工程化Cas12b核酸酶或效应子蛋白或CRISPR-Cas12b系统治疗或诊断的方法。还提供了本文所述的工程化Cas12b核酸酶或效应子蛋白或CRISPR-Cas12b系统用于检测或修饰细胞中的核酸、和用于治疗或诊断受试者的疾病或病况的用途;和包含任一种工程化Cas12b核酸酶或效应子蛋白或所述工程化的CRISPR-Cas12b系统的一个或多个组分的组合物用于制造用于检测或修饰核酸(例如,在细胞中)以及用于治疗或诊断受试者的疾病或病况的药物。

[0245] 修饰的方法

[0246] 在一些实施方案中,本申请提供了一种修饰包含靶序列的靶核酸的方法,包括将靶核酸与本文所述的任一种工程化的CRISPR-Cas12b系统或其组分接触。例如,当已经存在Cas12b蛋白或编码其的核酸时,仅需进一步提供gRNA或编码其的核酸;当已经存在gRNA或

编码其的核酸时,仅需进一步提供Cas12b蛋白或编码其的核酸。在一些实施方案中,提供了一种修饰包含靶序列的靶核酸的方法,包括将所述靶核酸与CRISPR-Cas12b系统(例如,工程化的、非天然存在的)接触(例如,体外、离体或体内),其中所述CRISPR-Cas12b系统包含:

(a) 工程化Cas12b核酸酶或其效应子蛋白(例如,切口酶、断裂Cas12b蛋白、转录阻遏物、转录激活物、碱基编辑器或先导编辑器),其包含相对于参考Cas12b核酸酶的一种、两种或三种类型的突变,其中所述突变包含:(1) 参考Cas12b核酸酶中与PAM相互作用的一个或多个氨基酸残基(例如,以下一个或多个位置:116、123、130、132、144、145、153、173、222、395、400和475)用带正电荷的氨基酸残基(例如,R、H、K)的取代;和/或(2) 参考Cas12b核酸酶中参与打开DNA双链的一个或多个氨基酸残基(例如,以下一个或多个位置:118和119)用具有芳香环的氨基酸残基(例如,F、Y、W)的取代;和/或(3) 参考Cas12b核酸酶的RuvC结构域中与ssDNA底物相互作用的一个或多个氨基酸残基(例如,以下一个或多个位置:300、301、304、329、636、639、647、682、757、758、761、764、768、852、854、856、857、858、860、862、863、865、866、867、869、938、956、957、958、994、1093和1097)用带正电荷的氨基酸残基(例如,R、H、K)或疏水性氨基酸残基(例如,F、Y、W、M)的取代,其中所述参考Cas12b核酸酶包含SEQ ID NO:1的氨基酸序列,或编码工程化Cas12b核酸酶或其效应子蛋白的核酸;和(b) 包含与靶核酸的靶序列互补的指导序列的gRNA,或编码gRNA的核酸,其中所述指导序列与靶核酸的靶序列的杂交介导工程化Cas12b核酸酶或其效应子蛋白与靶核酸的靶序列的接触,这导致工程化Cas12b核酸酶或其效应子蛋白对靶核酸的修饰。在一些实施方案中,所述gRNA包含支架,所述支架包含SEQ ID NO:23和25-53中任一个的序列。在一些实施方案中,工程化Cas12b核酸酶或其效应子蛋白包含SEQ ID NO:2-22和79-81中任一个的氨基酸序列。在一些实施方案中,提供了一种修饰包含靶序列的靶核酸的方法,包括将靶核酸与CRISPR-Cas12b系统(例如,工程化的、非天然存在的)接触(例如,体外、离体或体内),其中所述CRISPR-Cas12b系统包含:(a) 包含SEQ ID NO:1的氨基酸序列的Cas12b核酸酶或其效应子蛋白(例如,切口酶、断裂Cas12b蛋白、转录阻遏物、转录激活物、碱基编辑器或先导编辑器),或工程化Cas12b核酸酶或其效应子蛋白,相对于参考Cas12b核酸酶包含一种、两种或三种类型的突变,其中所述突变包含:(1) 参考Cas12b核酸酶中与PAM相互作用的一个或多个氨基酸残基(例如,以下一个或多个位置:116、123、130、132、144、145、153、173、222、395、400和475)用带正电荷的氨基酸残基(例如,R、H、K)的取代;和/或(2) 将参考Cas12b核酸酶中参与打开DNA双链的一个或多个氨基酸残基(例如,以下一个或多个位置:118和119)用具有芳香环的氨基酸残基(例如,F、Y、W)的取代;和/或(3) 参考Cas12b核酸酶的RuvC结构域中与ssDNA底物相互作用的一个或多个氨基酸残基(例如,以下一个或多个位置:300、301、304、329、636、639、647、682、757、758、761、764、768、852、854、856、857、858、860、862、863、865、866、867、869、938、956、957、958、994、1093和1097)用带正电荷的氨基酸残基(例如,R、H、K)或疏水性氨基酸残基(例如,F、Y、W、M)的取代,其中所述参考Cas12b核酸酶包含SEQ ID NO:1的氨基酸序列,或编码Cas12b核酸酶(例如,工程化的)或其效应子蛋白的核酸;和(b) 包含与靶核酸的靶序列互补的指导序列的gRNA,或编码gRNA的核酸,其中所述gRNA包含工程化支架,所述工程化支架包含SEQ ID NO:25-53中任一个的序列;其中所述指导序列与靶核酸的靶序列的杂交介导Cas12b核酸酶(例如,工程化的)或其效应子蛋白与靶核酸的靶序列的接触,这导致Cas12b核酸酶(例如,工程化的)或其效应子蛋白对靶核酸的修饰。在一些实施方案中,工程

化Cas12b核酸酶或其效应子蛋白包含SEQ ID NO:2-22和79-81中任一个的序列。在一些实施方案中,该方法还包括提供包含修复/供体核酸的修复/供体模板,其中所述修复/供体核酸能够在靶序列处掺入修饰的靶核酸中(例如,经由同源重组)。在一些实施方案中,靶核酸的修饰将靶核酸中的突变(例如,功能丧失突变)修复为野生型(或无害版本)序列。在一些实施方案中,靶核酸的修饰引入了外源序列。在一些实施方案中,该方法是在体外进行的。在一些实施方案中,靶核酸存在于细胞中。在一些实施方案中,细胞是细菌细胞、酵母细胞、植物细胞或动物细胞(例如,哺乳动物细胞,如人类或小鼠细胞)。在一些实施方案中,该方法是离体进行的。在一些实施方案中,该方法是在体内进行的。在一些实施方案中,靶核酸被工程化的CRISPR-Cas12b系统切割或靶核酸中的靶序列被其改变(例如,碱基编辑)。在一些实施方案中,靶核酸的表达被工程化的CRISPR-Cas12b系统改变。在一些实施方案中,靶核酸是基因组DNA,如细胞内的。在一些实施方案中,靶序列与疾病或病况相关。在一些实施方案中,修饰靶序列的方法治疗与靶序列相关的疾病或病况。在一些实施方案中,工程化的CRISPR-Cas12b系统包含编码多个crRNA的前体指导RNA阵列,其中每个crRNA包含不同的指导序列。

[0247] 在一些实施方案中,本申请提供了一种治疗个体细胞中与靶核酸相关的疾病或病况的方法,包括使用本文所述的任一种修饰靶核酸的方法来修饰个体的细胞中的靶核酸,由此治疗疾病或病况。在一些实施方案中,所述疾病或病况选自下组:癌症、心血管疾病、遗传性疾病、自身免疫性疾病、代谢疾病、神经退行性疾病、眼部疾病、细菌感染和病毒感染。

[0248] 根据CRISPR-Cas12b系统中工程化Cas12b效应子蛋白的类型,本文描述的工程化的CRISPR-Cas12b系统可以以多种方式修饰细胞中的靶核酸。在一些实施方案中,该方法诱导靶核酸的位点特异性切割。在一些实施方案中,该方法切割细胞中的基因组DNA,如细菌细胞、植物细胞或动物细胞(例如,哺乳动物细胞)。在一些实施方案中,该方法通过切割细胞中的基因组DNA来杀死细胞。在一些实施方案中,该方法切割细胞中的病毒核酸。在一些实施方案中,该方法对靶核酸进行碱基编辑,如将有害或疾病相关的突变修复为非疾病相关序列。在一些实施方案中,该方法增强(例如,增加至少约10%、20%、30%、40%、50%、60%、70%、80%、90%、1倍、2倍、5倍、10倍、20倍或更多中的任一项)靶核酸的表达(例如,修复下调表达的有害突变)。在一些实施方案中,该方法降低(例如,降低至少约10%、20%、30%、40%、50%、60%、70%、80%、90%、1倍、2倍、5倍、10倍、20倍或更多中的任一项)靶核酸的表达(例如,修复上调表达的有害突变)。

[0249] 在一些实施方案中,该方法改变(如增加或减少)细胞中靶核酸的表达水平。在一些实施方案中,该方法例如,使用基于与反式激活结构域融合的酶促失活的Cas12b蛋白(例如,SEQ ID NO:79-81中的任一个)的工程化Cas12b效应子蛋白增加细胞中靶核酸的表达水平。在一些实施方案中,该方法例如,使用基于与转录阻遏结构域(如KRAB结构域)融合的酶促失活的Cas12b蛋白(例如,SEQ ID NO:79-81中的任一个)的工程化Cas12b效应子蛋白降低细胞中靶核酸的表达水平。在一些实施方案中,该方法例如,使用基于与表观遗传修饰结构域融合的酶促失活Cas12b蛋白(例如,SEQ ID NO:79-81中的任一个)的工程化Cas12b效应子蛋白把表观遗传修饰引入细胞中的靶核酸。在一些实施方案中,该方法例如,使用基于与胞嘧啶脱氨酶结构域或腺苷脱氨酶结构域(例如,TadA)或其功能片段融合的酶促失活Cas12b蛋白(例如,SEQ ID NO:79-81中的任一个)的工程化Cas12b效应子蛋白把碱基编辑

引入细胞中的靶核酸中。根据工程化Cas12b效应子蛋白所包含的功能域,本文所述的工程化的Cas12b系统可以用于把其他修饰引入靶核酸。

[0250] 在一些实施方案中,该方法改变细胞中靶核酸中的靶序列。在一些实施方案中,该方法向细胞中的靶核酸引入突变。在一些实施方案中,该方法使用细胞中的一个或多个内源性DNA修复途径(如非同源末端连接(NHEJ)或同源性定向重组(HDR))来修复由于CRISPR复合物的序列特异性切割而导致的靶DNA中引起的双链断裂。示例性突变包括但不限于插入、缺失、取代和移码。在一些实施方案中,该方法将供体DNA插入靶基因座。在一些实施方案中,所述供体DNA的插入导致选择标记或报告蛋白被引入细胞。在一些实施方案中,供体DNA的插入导致基因的敲入。在一些实施方案中,供体DNA的插入会导致敲除突变。在一些实施方案中,供体DNA的插入导致取代突变,例如单核苷酸取代。在一些实施方案中,该方法诱导细胞的表型改变。

[0251] 在一些实施方案中,所述工程化的CRISPR-Cas12b系统用作遗传线路(genetic circuit)的一部分,或用于将遗传线路插入到细胞的基因组DNA中。本文所述的诱导剂控制的工程化的断裂Cas12b效应子蛋白可以是特别有用的,作为遗传线路的组分。遗传线路可对基因治疗有用。设计和使用遗传线路的方法和技术是本领域已知的。可进一步参考例如Brophy, Jennifer AN, and Christopher A. Voigt. "Principles of genetic circuit design." *Nature methods* 11.5 (2014): 508.

[0252] 本文所述的工程化的CRISPR-Cas12b系统可用于修饰广泛的靶核酸。在一些实施方案中,靶核酸位于细胞中。在一些实施方案中,靶核酸是基因组DNA。在一些实施方案中,靶核酸是染色体外DNA。在一些实施方案中,靶核酸对于细胞而言是外源的。在一些实施方案中,靶核酸是病毒核酸,如病毒DNA。在一些实施方案中,靶核酸是细胞中的质粒。在一些实施方案中,靶核酸是水平转移的质粒。在一些实施方案中,靶核酸是RNA,如mRNA。

[0253] 在一些实施方案中,靶核酸是分离的核酸,如分离的DNA。在一些实施方案中,靶核酸存在于无细胞环境中。在一些实施方案中,靶核酸是分离的载体,如质粒。在一些实施方案中,靶核酸是分离的线性DNA片段。

[0254] 本文描述的方法适用于任何合适的细胞类型。在一些实施方案中,细胞是细菌、酵母细胞、真菌细胞、藻类细胞、植物细胞或动物细胞(例如,哺乳动物细胞,如人类细胞)。在一些实施方案中,细胞是从天然来源分离的细胞,如组织活检。在一些实施方案中,细胞是从体外培养的细胞系分离的细胞。在一些实施方案中,细胞来自原代细胞系。在一些实施方案中,细胞来自永生化细胞系。在一些实施方案中,细胞是基因工程化的细胞。

[0255] 在一些实施方案中,细胞是来自生物体的动物细胞,包括但不限于猫、狗、小鼠、大鼠、仓鼠、牛、绵羊、山羊、马、驴、猪、鹿、鸡、鸭、鹅、兔和鱼。

[0256] 在一些实施方案中,细胞是来自选自下组的生物体的植物细胞:玉米、小麦、大麦、燕麦、水稻、大豆、油棕、红花、芝麻、烟草、亚麻、棉花、向日葵、珍珠粟、粟、高粱、油菜、大麻、蔬菜作物、饲料作物、工业作物、木本作物和生物质作物。

[0257] 在一些实施方案中,细胞是哺乳动物细胞。在一些实施方案中,细胞是小鼠细胞,如Neuro 2A(N2a)细胞。在一些实施方案中,细胞是人类细胞。在一些实施方案中,人类细胞是人类胚胎肾293T(HEK293T或293T)细胞或HeLa细胞。在一些实施方案中,哺乳动物细胞选自下组:免疫细胞、肝细胞、肿瘤细胞、干细胞、神经元细胞、受精卵、肌肉细胞和皮肤细胞。

[0258] 在一些实施方案中,细胞是选自下组的免疫细胞:细胞毒性T细胞、辅助T细胞、自然杀伤(NK)T细胞、iNK-T细胞、NK-T样细胞、 $\gamma\delta$ T细胞、肿瘤浸润T细胞和树突状细胞(DC)激活的T细胞。在一些实施方案中,该方法产生修饰的免疫细胞,如CAR-T细胞、CAR-NK细胞或TCR-T细胞。

[0259] 在一些实施方案中,所述细胞是胚胎干(ES)细胞、诱导性多能干(iPS)细胞、配子的祖细胞、配子、受精卵、或胚胎中的细胞。

[0260] 本文所述的方法可用于在体内、离体或体外修饰靶细胞,并且可以以改变细胞的方式进行,使得一旦修饰,修饰的细胞的后代或细胞系保留改变的表型。修饰的细胞和后代可以是多细胞生物体如植物或动物的一部分,具有离体或体内应用,如基因组编辑和基因治疗。

[0261] 在一些实施方案中,修饰的方法是离体进行。在一些实施方案中,将工程化的CRISPR-Cas12b系统引入细胞后,修饰的细胞(例如,哺乳动物细胞)离体繁殖。在一些实施方案中,将修饰的细胞培养以繁殖至少约1天、2天、3天、4天、5天、6天、7天、10天、12天或14天中的任一天。在一些实施方案中,培养修饰的细胞不超过约1天、2天、3天、4天、5天、6天、7天、10天、12天或14天中的任一项。在一些实施方案中,进一步评价或筛选修饰的细胞,以选择具有一个或多个所希望的表型或特性的细胞,或通过PCR或测序进行评估或筛选。

[0262] 在一些实施方案中,靶序列是与疾病或病况相关的序列。示例性疾病或病况包括但不限于癌症、血液疾病、心血管疾病、遗传性疾病、自身免疫疾病、代谢疾病、神经系统疾病、神经退行性疾病、眼部疾病、细菌感染和病毒感染。在一些实施例中,所述疾病或病况是移植物抗宿主病(GvHD)或宿主抗移植物(HvG)疾病。在一些实施例中,所述疾病或病况是遗传疾病。在一些实施方案中,所述疾病或病况是单基因疾病或病况。在一些实施方案中,疾病或病况是多基因疾病或病况。

[0263] 在一些实施方案中,与野生型序列相比靶序列具有突变。在一些实施方案中,靶序列具有与疾病或病况相关的单核苷酸多态性(SNP)。

[0264] 在一些实施方案中,插入靶核酸的供体DNA编码选自下组的生物产物:报告蛋白、抗原特异性受体、治疗性蛋白、抗生素抗性蛋白、RNAi分子、细胞因子、激酶、抗原、抗原特异性受体、嵌合受体、细胞因子受体和自杀多肽。在一些实施方案中,供体DNA编码治疗性蛋白质,例如,细胞因子。在一些实施方案中,供体DNA编码可用于基因治疗的治疗性蛋白质。在一些实施方案中,供体DNA编码治疗性抗体。在一些实施方案中,供体DNA编码工程化受体,如嵌合抗原受体(CAR)或工程化的TCR。在一些实施方案中,供体DNA编码治疗性RNA,如小RNA(例如,siRNA、shRNA或miRNA)或长非编码RNA(lincRNA)。

[0265] 本文描述的方法可用于两个或更多个(例如,2、3、4、5、6、8、10个或更多个)不同靶基因座的多重基因编辑或调控。在一些实施方案中,该方法检测或修饰多个靶核酸或靶核酸序列。在一些实施方案中,该方法包括将靶核酸与包含多个(例如,2、3、4、5、6、8、10个或更多个)crRNA序列的指导RNA接触,其中每个所述crRNA包含不同的靶序列。

[0266] 还提供了包含修饰的靶核酸的工程化的细胞,其是使用本文描述的任何一种修饰方法产生的。所述工程化的细胞可用于细胞治疗。可以使用自体或同种异体细胞,利用本文所述的方法制备工程化的细胞,进行细胞治疗。

[0267] 本文所述的方法还可用于生成等基因细胞系(例如,哺乳动物细胞)以研究遗传变

体。

[0268] 还提供了包含本文所述的工程化细胞的工程化的植物或非人类动物。在一些实施方案中,所述工程化的植物或非人类动物是基因组被编辑的植物或非人类动物。所述工程化的非人类动物可用作疾病模型。

[0269] 产生非人类基因组编辑或转基因动物的技术在本领域中是众所周知的,包括但不限于原核显微注射、病毒感染、和胚胎干细胞和诱导多能干(iPS)细胞的转化。可以使用的详细方法包括但不限于在Sundberg和Ichiki(2006,Genetically Engineered Mice Handbook,CRC Press)和Gibson(2004,A Primer Of Genome Science 2nd ed.Sunderland,Mass.:Sinauer)中描述的那些方法。

[0270] 工程化的动物可以是任何合适的物种,包括但不限于,如牛科动物、马科动物、羊科动物、犬科动物、鹿科动物、猫科动物、山羊、猪、灵长类动物和不太常见的哺乳动物,如大象、鹿、斑马或骆驼。

[0271] 治疗的方法

[0272] 还提供了使用本文所述的任一种修饰细胞中的靶核酸的方法进行治疗的方法,和使用本文所述的任一种检测靶核酸的方法进行诊断的方法。

[0273] 在一些实施方案中,本申请提供了一种治疗个体细胞中与靶核酸相关的疾病或病况的方法,包括将靶核酸与本文所述的任一种工程化的CRISPR-Cas12b系统接触,其中指导RNA的指导序列与靶核酸的靶序列互补,其中Cas12b核酸酶(例如,工程化的)或其效应子蛋白(例如,包含SEQ ID NO:1-22和79-81中的任一个)和指导RNA相互关联以结合靶核酸以修饰该靶核酸,由此治疗疾病或病况。在一些实施方案中,将突变(例如,敲除或敲入突变)引入至靶核酸。在一些实施方案中,增强靶核酸的表达。在一些实施方案中,抑制靶核酸的表达。

[0274] 在一些实施方案中,本申请提供了一种治疗个体的疾病或病况的方法,包括向个体施用有效量的本文所述的任一种工程化的CRISPR-Cas12b系统,和编码治疗剂的供体DNA,其中所述指导RNA的指导序列与个体的靶核酸的靶序列互补,其中所述Cas12b核酸酶(例如,工程化的)或其效应子蛋白(例如,包含SEQ ID NO:1-22和79-81中的任一个)和指导RNA相互关联以结合靶核酸并将供体DNA插入靶序列中,由此治疗疾病或病况。

[0275] 在一些实施方案中,本申请提供了一种治疗个体的疾病或病况的方法,包括向个体施用有效量的包含修饰的靶核酸的工程化的细胞,其中所述工程化的细胞是通过将所述细胞与本文所述的任一种工程化的CRISPR-Cas12b系统接触而制备的,其中所述指导RNA的指导序列与靶核酸的靶序列互补,其中所述Cas12b核酸酶(例如,工程化的)或其效应子蛋白(例如,包含SEQ ID NO:1-22和79-81中的任一个)和指导RNA相互关联以结合靶核酸来修饰靶核酸。在一些实施方案中,所述工程化的细胞是免疫细胞。

[0276] 在一些实施例中,提供了一种治疗与个体(例如,人类)细胞中的靶核酸相关的疾病或病况的方法,包括使靶核酸(例如,离体或体内)与个体接触或向个体施用有效量的CRISPR-Cas12b系统(例如,工程化的、非天然存在的),其中所述CRISPR-Cas12b系统包含:(a)工程化Cas12b核酸酶或其效应子蛋白(例如,切口酶、断裂Cas12b蛋白、转录阻遏物、转录激活物、碱基编辑器或先导编辑器),其包含相对于参考Cas12b核酸酶的一种、两种或三种类型的突变,其中所述突变包含:(1)参考Cas12b核酸酶中与PAM相互作用的一个或多个

氨基酸残基(例如,以下一个或多个位置:116、123、130、132、144、145、153、173、222、395、400和475)用带正电荷的氨基酸残基(例如,R、H、K)的取代;和/或(2)参考Cas12b核酸酶中参与打开DNA双链的一个或多个氨基酸残基(例如,以下一个或多个位置:118和119)用具有芳香环的氨基酸残基(例如,F、Y、W)的取代;和/或(3)参考Cas12b核酸酶的RuvC结构域中与ssDNA底物相互作用的一个或多个氨基酸残基(例如,以下一个或多个位置:300、301、304、329、636、639、647、682、757、758、761、764、768、852、854、856、857、858、860、862、863、865、866、867、869、938、956、957、958、994、1093和1097)用带正电荷的氨基酸残基(例如,R、H、K)或疏水性氨基酸残基(例如,F、Y、W、M)的取代,其中所述参考Cas12b核酸酶包含SEQ ID NO:1的氨基酸序列,或编码工程化Cas12b核酸酶或其效应子蛋白的核酸;和(b)包含与靶核酸的靶序列互补的指导序列的gRNA,或编码所述gRNA的核酸,其中所述指导序列与靶核酸的靶序列的杂交介导工程化Cas12b核酸酶或其效应子蛋白与靶核酸的靶序列的接触,这导致工程化Cas12b核酸酶或其效应子蛋白对靶核酸进行修饰,由此治疗疾病或病况。在一些实施方案中,所述gRNA包含支架,所述支架包含SEQ ID NO:23和25-53中任一个的序列。在一些实施方案中,提供了一种治疗与个体(例如,人类)细胞中的靶核酸相关的疾病或病况的方法,该方法包括使靶核酸与个体接触(例如,离体或体内)或向个体施用有效量的CRISPR-Cas12b系统(例如,工程化的、非天然存在的),其中所述CRISPR-Cas12b系统包含:(a)包含SEQ ID NO:1的氨基酸序列的Cas12b核酸酶或其效应子蛋白(例如,切口酶、断裂Cas12b蛋白、转录抑制因子、转录激活因子、碱基编辑器或先导编辑器)、或工程化Cas12b核酸酶或其效应子蛋白,其相对于参考Cas12b核酸酶包含一种、两种或三种类型的突变,其中所述突变包含:(1)参考Cas12b核酸酶中与PAM相互作用的一个或多个氨基酸残基(例如,以下一个或多个位置:116、123、130、132、144、145、153、173、222、395、400和475)用带正电荷的氨基酸残基(例如,R、H、K)的取代;和/或(2)参考Cas12b核酸酶中参与打开DNA双链的一个或多个氨基酸残基(例如,以下一个或多个位置:118和119)用具有芳香环的氨基酸残基(例如,F、Y、W)的取代;和/或(3)参考Cas12b核酸酶的RuvC结构域中与ssDNA底物相互作用的一个或多个氨基酸残基(例如,以下一个或多个位置:300、301、304、329、636、639、647、682、757、758、761、764、768、852、854、856、857、858、860、862、863、865、866、867、869、938、956、957、958、994、1093和1097)用带正电荷的氨基酸残基(例如,R、H、K)或疏水性氨基酸残基(例如,F、Y、W、M)的取代,其中所述参考Cas12b核酸酶包含SEQ ID NO:1的氨基酸序列,或编码Cas12b核酸酶(例如,工程化的)或其效应子蛋白的核酸;和(b)包含与靶核酸的靶序列互补的指导序列的gRNA,或编码所述gRNA的核酸,其中所述gRNA包含工程化支架,所述工程化支架包含SEQ ID NO:25-53中任一个的序列;其中所述指导序列与靶核酸的靶序列的杂交介导Cas12b核酸酶(例如,工程化的)或其效应子蛋白与靶核酸的靶序列的接触,这导致Cas12b核酸酶(例如,工程化的)或其效应子蛋白对靶核酸的修饰,由此治疗疾病或病况。在一些实施方案中,工程化Cas12b核酸酶或其效应子蛋白包含SEQ ID NO:2-22和79-81中任一个的氨基酸序列。在一些实施方案中,该方法还包括使靶核酸(例如,离体或体内)与有效量的修复/供体核酸接触或向个体施用有效量的修复/供体核酸,其中所述修复/供体核酸能够在靶序列处掺入修饰的靶核酸中(例如,经由同源重组)。在一些实施方案中,所述靶核酸的修饰将靶核酸中的突变(例如,功能丧失突变)修复为野生型(或无害版本)序列。在一些实施方案中,所述靶核酸的修饰引入了外源序列。

[0277] 在一些实施方案中,个体是人类。在一些实施方案中,个体是动物,例如,模型动物如啮齿动物(例如,小鼠、大鼠、仓鼠)、宠物(例如,猫、狗、兔子)或农场动物(例如,马、牛、绵羊、山羊、驴、猪)。在一些实施方案中,个体是哺乳动物。

[0278] 在一些实施方案中,疾病或病况与个体(例如,人类)的靶核酸中的异常(例如,致病点突变)相关。在一些实施方案中,疾病或病况是由于CRISPR-Cas12b系统或复合物对靶核酸的修饰(例如,切割、碱基编辑或修复)(例如,修复异常)而治疗的。在一些实施方案中,疾病是由一个或多个靶基因的过表达或错误表达(例如,错义突变、移码突变、无义突变)引起的,其中,所述CRISPR-Cas12b系统或复合物可以靶向所述一个或多个靶基因进行靶向修饰,如切割、碱基编辑、或序列修复(例如,通过进一步引入修复/供体模板,以通过同源重组藉由CRISPR-Cas12b系统或复合物修复切割的靶基因)。

[0279] 在一些实施方案中,所述疾病或病况选自下组:癌症、心血管疾病、遗传性疾病、自身免疫性疾病、代谢疾病、神经退行性疾病、眼部疾病、细菌感染和病毒感染。

[0280] 在一些实施方案中,所述疾病或病况选自下组:转甲状腺素蛋白淀粉样变性(ATTR)(如转甲状腺素蛋白相关野生型淀粉样变性(ATTRwt)、转甲状腺素蛋白相关遗传性淀粉样变性(ATTRm)、家族性淀粉样多发性神经病(FAP,ATTR-PN)或家族性淀粉样心肌病(FAC,ATTR-CM)、囊性纤维化、遗传性血管性水肿(HAE)、糖尿病、进行性假肥大型肌营养不良症、贝克肌营养不良(BMD)、 α -1抗胰蛋白酶缺乏症(AAT缺乏症)、庞贝病、肌强直性营养不良症、亨廷顿氏病、脆性X综合征(FXS)、弗里德赖希共济失调(FRDA)、肌萎缩侧索硬化症(ALS)、额颞叶痴呆(FTD)、遗传性慢性肾病、高脂血症、高胆固醇血症(例如,家族性高胆固醇血症)、莱伯先天性黑蒙(LCA)、镰状细胞病(SCD)和 β -地中海贫血。在一些实施方案中,CRISPR-Cas12b系统或复合物经由脂质纳米颗粒进行包装和递送。在一些实施方案中,脂质纳米颗粒经由静脉注射或输注施用于个体。

[0281] 在一些实施方案中,靶核酸是PCSK9。在一些实施方案中,疾病或病况是心血管疾病。在一些实施方案中,疾病或病况是冠状动脉疾病。在一些实施方案中,该方法降低个体的胆固醇水平。在一些实施方案中,该方法治疗个体的糖尿病。在一些实施方案中,疾病或病况是高胆固醇血症,如家族性高胆固醇血症。

[0282] 在一些实施方案中,靶核酸是HBG1和/或HBG2。在一些实施方案中,疾病或病况是镰状细胞病或 β -地中海贫血。在一些实施方案中,疾病或病况是遗传性胎儿血红蛋白持续存在症(HPFH)、HbS-基因缺失HPFH、或由于点突变导致的HbS-HPFH。

[0283] 在一些实施方案中,靶核酸是C-C趋化因子受体(CCR)5(CCR5),其编码主要的HIV-1协同受体。在一些实施方案中,所述疾病或病况是传染病,例如,AIDS。在一些实施方案中,所述疾病或病况是非传染病,如癌症(例如,乳腺癌或前列腺癌)、动脉粥样硬化、中风或炎症性肠病(IBD)。

[0284] 在一些实施方案中,靶核酸是CD34。在一些实施方案中,疾病或病况是癌症。

[0285] 在一些实施方案中,靶核酸是环指蛋白2(RNF2)。在一些实施方案中,疾病或病况是神经系统疾病,如Luo-Schoch-Yamamoto综合征或非特异性综合征性智力障碍。

[0286] 检测的方法

[0287] 本申请还提供了使用具有提高的活性的任一种工程化Cas12b核酸酶或其效应子蛋白(例如,包含SEQ ID NO:2-22和79-81中的任一种)或CRISPR-Cas12b系统来检测靶核酸

的方法。使用Cas12b效应子蛋白作为检测剂利用了以下发现的优势：V型CRISPR/Cas12蛋白（例如，Cas12a、Cas12b、Cas12c、Cas12d、Cas12e (CasX) 和Cas12i）一旦通过检测靶DNA而激活，便可以杂乱切割非靶向的单链DNA (ssDNA)。使用Cas12b蛋白作为检测剂的方法已在例如US10253365和W02020/056924中描述，其内容均通过引用以其整体并入本文。在一些实施方案中，检测样本中的靶核酸诊断疾病或病况。

[0288] 在一些实施方案中，一旦Cas12b效应子蛋白被指导RNA激活，当样品包含指导RNA杂交的靶DNA时（即，样品包含靶DNA），Cas12b核酸酶或其效应子蛋白就变成杂乱地切割单链核酸（例如，非靶ssDNA或RNA，即，指导RNA的指导序列不杂交的单链核酸）的核酸酶。因此，当靶DNA（双链或单链）存在于样本中（例如，在某些情况下高于阈值量）时，结果是样本中的单链核酸被切割，这可以使用任何方便的检测方法检测（例如，使用标记的单链检测核酸，如DNA或RNA）。Cas12b可以切割ssDNA和ssRNA。

[0289] 在一些实施方案中，提供了一种检测样品中的靶DNA（例如，双链或单链）的方法，包括：(a) 使样品与以下物质接触：(i) 本文所述的任一种工程化Cas12b核酸酶或其效应子蛋白（例如，包含SEQ ID NO:2-22和79-81中的任一个）；(ii) 包含与靶DNA杂交的指导序列的指导RNA；和(iii) 单链的检测核酸（即，“单链检测核酸”）并且不与指导RNA的指导序列杂交；和(b) 测量由工程化Cas12b效应子蛋白切割单链检测核酸产生的可检测信号。在一些实施方案中，提供了一种检测样品中的靶DNA（例如，双链或单链）的方法，包括：(a) 使样品与以下物质接触：(i) 本文所述的任何Cas12b核酸酶（例如，工程化的或野生型）或其效应子蛋白（例如，包含SEQ ID NO:1-22和79-81中的任一个）；(ii) 包含与靶DNA杂交的指导序列的指导RNA，和包含SEQ ID NO:25-53中的任一个的序列的工程化支架；和(iii) 单链的检测核酸（即，“单链检测核酸”）并且不与指导RNA的指导序列杂交；和(b) 测量由工程化Cas12b效应子蛋白切割单链检测核酸产生的可检测信号。在一些实施方案中，提供了一种检测样品中的靶核酸的方法，包括：(a) 将样品与本文描述的任何工程化的CRISPR-Cas12b系统和标记的检测核酸接触，其中所述gRNA包含与靶核酸的靶序列互补的指导序列，并且其中标记的检测核酸是单链的并且不与gRNA的指导序列杂交；和(b) 测量由工程化的CRISPR-Cas12b系统切割标记的检测核酸产生的可检测信号，由此检测所述靶核酸。在一些情况下，单链检测核酸包含荧光发射染料对（例如，荧光发射染料对是荧光能量共振转移 (FRET) 对、猝灭剂/荧光对）。在一些情况下，所述靶DNA是病毒DNA（例如，乳头状瘤病毒、嗜肝DNA病毒、疱疹病毒、腺病毒、痘病毒、细小病毒等）。在一些实施方案中，单链检测核酸是DNA。在一些实施方案中，单链检测核酸是RNA。在一些实施方案中，工程化Cas12b效应子蛋白是工程化Cas12b核酸酶。在一些实施方案中，该方法是在体外进行的。在一些实施方案中，所述靶核酸存在于细胞中，如细菌细胞、酵母细胞、植物细胞或动物细胞。在一些实施方案中，该方法是在离体进行的。在一些实施方案中，该方法是在体内进行的。在一些实施方案中，所述靶核酸是基因组DNA。在一些实施方案中，所述靶序列与疾病或病况相关。

[0290] 本公开的检测样本中的靶DNA（单链或双链）的方法可以高灵敏度地检测靶DNA。在一些情况下，本公开的方法可用于检测包含多个DNA（包含靶DNA和多个非靶DNA）的样本中存在的靶DNA，其中所述靶DNA以一个或多个拷贝（例如，一个或多个拷贝靶DNA每 10^6 个非靶DNA、一个或多个拷贝靶DNA每 10^5 个非靶DNA、一个或多个拷贝靶DNA每 10^4 个非靶DNA、一个或多个拷贝靶DNA每 10^3 个非靶DNA、一个或多个拷贝靶DNA每 10^2 个非靶DNA、一个或多个拷贝靶

DNA每50个非靶DNA、一个或多个拷贝靶DNA每20个非靶DNA、一个或多个拷贝靶DNA每10个非靶DNA、或一个或多个拷贝靶DNA每5个非靶DNA)的靶DNA每 10^7 个非靶DNA的形式存在。

[0291] 在一些实施方案中,本文所述的工程化Cas12b核酸酶或其效应子蛋白(例如,包含SEQ ID NO:2-22中的任一个)可以与参考Cas12b核酸酶(例如,SEQ ID NO:1)相比以更高的灵敏度检测靶DNA。在一些实施方案中,与参考Cas12b核酸酶相比,工程化Cas12b效应子蛋白可以以10%、15%、20%、25%、30%、40%、50%、60%、70%、80%、90%或更高的灵敏度检测靶DNA。

[0292] 递送的方法

[0293] 在一些实施方案中,本文所述的工程化的CRISPR-Cas12b系统或其组分、其核酸分子、或编码或提供其组分的核酸分子,可以通过多种递送系统如质粒或病毒载体(例如,上文“构建体和载体”小节中所述的任一个载体)递送至宿主细胞。在一些实施方案或方法中,所述工程化的CRISPR-Cas12b系统可以通过其他方法传递,如由工程化Cas12b核酸酶或其效应子蛋白及其同源RNA指导物组成的核糖核蛋白复合物的核转染或电穿孔。

[0294] 在一些实施方案中,所述递送是经由纳米颗粒或外泌体进行的。

[0295] 在一些实施方案中,可以使用纳米颗粒或其他直接蛋白质递送方法直接递送配对的Cas12b切口酶复合物,使得包含两种配对的crRNA元件的复合物共同递送。此外,蛋白质可以通过病毒载体递送至细胞或直接地递送至细胞,然后直接递送含有两个配对的间隔区的CRISPR阵列以进行双切口。在一些情况下,为了直接递送RNA,RNA可以与至少一个糖部分缀合,如N-乙酰半乳糖胺(GalNAc)(特别是,三触角N-乙酰半乳糖胺)。在一些实施方案中,CRISPR-Cas12b系统或其组分经由脂质纳米颗粒进行包装和递送。在一些实施方案中,脂质纳米颗粒经由静脉注射或输注施用于个体。

[0296] IV. 试剂盒和制品

[0297] 还提供了组合物、试剂盒、单位剂量和制品,其包含本文所述的任一种工程化Cas12b核酸酶或其效应子蛋白、包含工程化支架的sgRNA(例如,SEQ ID NO:25-53中的任一个)、或工程化的CRISPR-Cas12b系统的一个或多个组分。

[0298] 在一些实施方案中,提供了一种试剂盒,其包含:一个或多个AAV载体,其编码本文所述的任一种工程化Cas12b核酸酶或其效应子蛋白、或工程化的CRISPR-Cas12b系统。在一些实施方案中,所述试剂盒还包含一个或多个指导RNA,如包含工程化支架的sgRNA(例如,SEQ ID NO:25-53中的任一个)。在一些实施方案中,所述试剂盒还包含供体DNA。在一些实施方案中,试剂盒还包含细胞,如人类细胞。

[0299] 所述试剂盒可含有一个或多个另外的组分,如容器、试剂、培养基、细胞因子、缓冲液、抗体等,以允许工程细胞繁殖。试剂盒还可以含有用于组合物的施用的装置。

[0300] 试剂盒还可以包含使用本文所述的工程化的CRISPR-Cas12b系统的说明书,如检测或修饰靶核酸的方法。在一些实施方案中,所述试剂盒包含用于治疗或诊断疾病或病况的说明书。与试剂盒组分的使用相关的说明书通常包含有关预期治疗的剂量、给药时间表和施用的途径的信息。所述容器可以是单位剂量、大包装(例如,多剂量包装)或亚单位剂量。例如,试剂盒可以提供含有足够剂量的本文公开的组合物以为个体提供长期有效的治疗。试剂盒还可以包含组合物的多个单位剂量和使用的说明书,其包装数量足以在药房(例如医院药房和配药药房)中储存和使用。

[0301] 本申请的试剂盒采用合适的包装。合适的包装包括但不限于小瓶、瓶子、罐子、软包装(例如密封的聚酯薄膜或塑料袋)等。试剂盒可任选地提供另外的组件,例如缓冲液和解释信息。因此,本申请还提供制品,其包含小瓶(如密封的小瓶)、瓶子、罐子、软包装等。

[0302] 所述制品可以包含容器和容器上或与容器相关的标签或包装插页。合适的容器包含,例如,瓶子、小瓶、注射器等。所述容器可以由多种材料形成,如玻璃或塑料。一般而言,所述容器容纳可有效治疗本文所述疾病或病况的组合物,并且可以具有无菌接入口(例如容器可以是静脉溶液袋或具有可被皮下注射针刺穿的塞子的小瓶)。标签或包装插页表明该组合物用于治疗个体的特定病况。标签或包装插页会进一步包含向个体施用该组合物的说明书。

[0303] 包装插页是指通常包含在治疗产品商业包装中的说明书,其包含有关这些治疗产品的适应症、用法、剂量、施用、禁忌症和/或有关使用的警告的信息。

[0304] 此外,所述制品还可包含第二容器,所述第二容器包含药学上可接受的缓冲液,如注射用抑菌水(BWFI)、磷酸盐缓冲盐水、林格氏溶液和葡萄糖溶液。它还可包含从商业和用户角度所希望的其他材料,包含其他缓冲液、稀释剂、过滤器、针头和注射器。

[0305] 示例性实施方案

[0306] 实施方案1.一种工程化Cas12b核酸酶,相对于参考Cas12b核酸酶,其包含一种、两种或三种类型的突变,其中所述突变包含:(1)所述参考Cas12b核酸酶中与前间区邻近基序(PAM)相互作用的一个或多个氨基酸残基用带正电荷的氨基酸残基的取代;和/或(2)所述参考Cas12b核酸酶中参与打开DNA双链的一个或多个氨基酸残基用具有芳香环的氨基酸残基的取代;和/或(3)所述参考Cas12b核酸酶的RuvC结构域中与单链DNA底物相互作用的一个或多个氨基酸残基用带正电荷的氨基酸残基或疏水性氨基酸残基的取代。

[0307] 实施方案2.根据实施方案1所述的工程化Cas12b核酸酶,其中所述参考Cas12b核酸酶是野生型Cas12b核酸酶。

[0308] 实施方案3.根据实施方案1或2所述的工程化Cas12b核酸酶,其中所述参考Cas12b核酸酶包含SEQ ID NO:1的氨基酸序列。

[0309] 实施方案4.根据实施方案1-3中任一项所述的工程化Cas12b核酸酶,其包含所述参考Cas12b核酸酶中与PAM相互作用的一个或多个氨基酸残基用带正电荷的氨基酸残基的取代。

[0310] 实施方案5.根据实施方案4所述的工程化Cas12b核酸酶,其中所述与PAM相互作用的一个或多个氨基酸残基在三维结构中距离PAM在9埃以内。

[0311] 实施方案6.根据实施方案4或5所述的工程化Cas12b核酸酶,其中所述与PAM相互作用的一个或多个氨基酸残基位于以下一个或多个位置:116、123、130、132、144、145、153、173、222、395、400和/或475;其中所述氨基酸残基按照SEQ ID NO:1编号。

[0312] 实施方案7.根据实施方案6所述的工程化Cas12b核酸酶,其中所述与PAM相互作用的一个或多个氨基酸残基包含以下一个或多个氨基酸残基:D116,K123,D130,D132,N144,K145,E153,D173,Q222,D395,N400,和/或E475;并且其中所述氨基酸残基按照SEQ ID NO:1编号。

[0313] 实施方案8.根据实施方案7所述的工程化Cas12b核酸酶,其中所述与PAM相互作用的一个或多个氨基酸残基包含以下一个或多个氨基酸残基:D116和E475;其中所述氨基酸

残基按照SEQ ID NO:1编号。

[0314] 实施方案9.根据实施方案4-8中任一项所述的工程化Cas12b核酸酶,其中所述带正电荷的氨基酸残基是R或K。

[0315] 实施方案10.根据实施方案9所述的工程化Cas12b核酸酶,其中所述参考Cas12b核酸酶中与PAM相互作用的一个或多个氨基酸残基用带正电荷的氨基酸残基的取代是下列一个或多个取代:D116R和/或E475R;其中所述氨基酸残基按照SEQ ID NO:1编号。

[0316] 实施方案11.根据实施方案1-10中任一项所述的工程化Cas12b核酸酶,其包含所述参考Cas12b核酸酶中参与打开所述DNA双链的一个或多个氨基酸残基用具有芳香环的氨基酸残基的取代。

[0317] 实施方案12.根据实施方案11所述的工程化Cas12b核酸酶,其中参与打开所述DNA双链的一个或多个氨基酸残基与相对于靶链的3'端的PAM中的最后一个碱基对相互作用。

[0318] 实施方案13.根据实施方案11或12所述的工程化Cas12b核酸酶,其中参与打开所述DNA双链的一个或多个氨基酸残基位于以下一个或多个位置:118和/或119;其中所述氨基酸残基按照SEQ ID NO:1编号。

[0319] 实施方案14.根据实施方案11-13中任一项所述的工程化Cas12b核酸酶,其中所述具有芳香环的氨基酸残基为Y、F或W。

[0320] 实施方案15.根据实施方案14所述的工程化Cas12b核酸酶,其中所述参考Cas12b核酸酶中参与打开DNA双链的一个或多个氨基酸残基用具有芳香环的氨基酸残基的取代为Q119Y、Q119F或Q119W;其中所述氨基酸残基按照SEQ ID NO:1编号。

[0321] 实施方案16.根据实施方案1-16中任一项所述的工程化Cas12b核酸酶,其包含所述参考Cas12b核酸酶中位于所述RuvC结构域中并与所述单链DNA底物相互作用的一个或多个氨基酸残基用带正电荷的氨基酸残基或疏水性氨基酸残基的取代。

[0322] 实施方案17.根据实施方案16所述的工程化Cas12b核酸酶,其中所述RuvC结构域中与所述单链DNA底物相互作用的一个或多个氨基酸残基在三维结构中距离所述单链DNA底物9埃以内。

[0323] 实施方案18.根据实施方案17所述的工程化Cas12b核酸酶,其中所述RuvC结构域中与所述单链DNA底物相互作用的一个或多个氨基酸残基位于以下一个或多个位置:300、301、304、329、636、639、647、682、757、758、761、764、768、852、854、856、857、858、860、862、863、865、866、867、869、938、956、957、958、994、1093和/或1097;其中所述氨基酸残基按照SEQ ID NO:1编号。

[0324] 实施方案19.根据实施方案18所述的工程化Cas12b核酸酶,其中在RuvC结构域中与单链DNA底物相互作用的一个或多个氨基酸残基包含以下氨基酸残基中的一个或者多个:D300、K301、E304、N329、E636、Q639、T647、Q682、I757、E758、E761、E764、K768、E852、Q854、N856、N857、D858、P860、S862、E863、N865、Q866、L867、Q869、E938、E956、G957、E958、I994、Q1093,以和/或W1097;其中所述氨基酸残基按照SEQ ID NO:1编号。

[0325] 实施方案20.根据实施方案19所述的工程化Cas12b核酸酶,其包含以下一个或多个氨基酸残基用带正电荷的氨基酸残基的取代:E636、I757、E758、E761、Q854、N857、N865、Q866、Q869、和/或Q1093;其中所述氨基酸残基按照SEQ ID NO:1编号。

[0326] 实施方案21.根据实施方案20所述的工程化Cas12b核酸酶,其中所述带正电荷的

氨基酸残基是R或K。

[0327] 实施方案22. 根据实施方案21所述的工程化Cas12b核酸酶, 其中所述参考Cas12b核酸酶中位于所述RuvC结构域中并与所述单链DNA底物相互作用的一个或多个氨基酸残基的取代是以下一个或多个取代: E636R、I757R、E758R、E761R、Q854R和/或N857K、; 其中所述氨基酸残基按照SEQ ID NO:1编号。

[0328] 实施方案23. 根据实施方案19所述的工程化Cas12b核酸酶, 其包含以下氨基酸残基中的一个或多个用疏水性氨基酸残基的取代: E758、E761、E863、N865、Q866、Q869、Q956和/或Q1093; 其中所述氨基酸残基按照SEQ ID NO:1编号。

[0329] 实施方案24. 根据实施方案23所述的工程化Cas12b核酸酶, 其中所述疏水性氨基酸残基为W、Y、F或M。

[0330] 实施方案25. 根据实施方案24所述的工程化Cas12b核酸酶, 其中所述参考Cas12b核酸酶中位于所述RuvC结构域中并与所述单链DNA底物相互作用的一个或多个氨基酸残基的取代是以下一个或多个取代: N865W、N865Y、Q866M、Q869M、Q1093W和/或Q1093Y; 其中所述氨基酸残基按照SEQ ID NO:1编号。

[0331] 实施方案26. 根据实施方案1-3中任一项所述的工程化Cas12b核酸酶, 其中所述工程化Cas12b核酸酶包含以下任一种取代或其组合: (1) D116R; (2) E475R; (3) Q119F和E475R; (4) Q119F, E475R, 和E758R; (5) Q119Y; (6) Q119F; (7) Q119W; (8) I757R; (9) E758R; (10) E761R; (11) K768R; (12) I757R和E758R; (13) I757R和E761R; (14) I757R和K768R; (15) E758R和E761R; (16) E758R和K768R; (17) E761R和K768R; (18) I757R, E758R, 和E761R; (19) I757R, E758R, 和K768R; (20) I757R, E761R和K768R; (21) E758R, E761R, 和K768R; (22) I757R, E758R, E761R, 和K768R; (23) Q866M; (24) Q869M; 和(25) Q866M和Q869M; 其中所述氨基酸残基按照SEQ ID NO:1编号。

[0332] 实施方案27. 根据实施方案1-26中任一项所述的工程化Cas12b核酸酶, 其包含与SEQ ID NO:20-22中任一个具有至少85%序列同一性的氨基酸序列。

[0333] 实施方案28. 根据实施方案1-27中任一项所述的工程化Cas12b核酸酶, 其还包含一个或多个突变, 所述突变增加包含氨基酸残基855-859的柔性区的柔性; 其中所述氨基酸位置按照SEQ ID NO:1编号。

[0334] 实施方案29. 根据实施方案28所述的工程化Cas12b核酸酶, 其中增加柔性的一个或多个突变包含N856G。

[0335] 实施方案30. 一种工程化Cas12b核酸酶, 其包含以下突变的任一个或多个: (1) D116R; (2) E475R; (3) Q119F和E475R; (4) Q119F, E475R, 和E758R; (5) Q119Y; (6) Q119F; (7) Q119W; (8) Q119F和E475R; (9) Q119F, E475R和E758R (10) E636R; (11) I757R; (12) E758R; (13) E761R; (14) Q854R; (15) N857K; (16) Q119F, E475R, 和E758R; (17) K768R; (18) I757R和E758R; (19) I757R和E761R; (20) I757R和K768R; (21) E758R和E761R; (22) E758R和K768R; (23) E761R和K768R; (24) I757R, E758R, 和E761R; (25) I757R, E758R, 和K768R; (26) I757R, E761R和K768R; (27) E758R, E761R, 和K768R; (28) I757R, E758R, E761R, 和K768R (29) N865W; (30) N865Y; (31) Q866M; (32) Q869M; (33) Q1093W; (34) Q1093Y; 和/或 (35) Q866M和Q869M; 其中所述氨基酸位置按照SEQ ID NO:1编号。

[0336] 实施方案31. 一种工程化Cas12b核酸酶, 其包含SEQ ID NO:2-22中任一项的氨基

酸序列。

[0337] 实施方案32.一种工程化Cas12b效应子蛋白,其包含实施方案1-31中任一项所述的工程化Cas12b核酸酶或其功能衍生物。

[0338] 实施方案33.根据实施方案32所述的工程化Cas12b效应子蛋白,其中所述工程化Cas12b核酸酶或其功能衍生物具有酶促活性。

[0339] 实施方案34.根据实施方案32或33所述的工程化Cas12b效应子蛋白,其中所述工程化Cas12b效应子蛋白能够诱导DNA分子中的双链断裂。

[0340] 实施方案35.根据实施方案32或33所述的工程化Cas12b效应子蛋白,其中所述工程化Cas12b效应子蛋白能够诱导DNA分子中的单链断裂。

[0341] 实施方案36.根据实施方案32所述的工程化Cas12b效应子蛋白,其中所述工程化Cas12b效应子蛋白包含所述工程化Cas12b核酸酶的酶促失活突变。

[0342] 实施方案37.根据实施方案36所述的工程化Cas12b效应子蛋白,其中所述酶促失活突变包含:D570A、R785A、E848A、R911A和/或D977A。

[0343] 实施方案38.根据实施方案32-37中任一项所述的工程化Cas12b效应子蛋白,其还包含与所述工程化Cas12b核酸酶或其功能衍生物融合的功能结构域。

[0344] 实施方案39.根据实施方案38所述的工程化Cas12b效应子蛋白,其中所述功能结构域选自下组:翻译启动子结构域、转录阻遏物结构域、反式激活结构域、表观遗传修饰结构域、核碱基编辑结构域、逆转录酶结构域、报告子结构域和核酸酶结构域。

[0345] 实施方案40.根据实施方案32-37中任一项所述的工程化Cas12b效应子蛋白,其包含第一多肽和第二多肽,所述第一多肽包含工程化Cas核酸酶或其功能衍生物的N-末端部分,所述第二多肽包含工程化Cas核酸酶及其功能衍生物的C-末端部分,其中所述第一多肽和所述第二多肽能够在包含指导序列的指导RNA的存在下相互缔合,以形成特异性结合包含与所述指导序列互补的靶序列的靶核酸的成簇规律间隔短回文重复序列(CRISPR)复合物。

[0346] 实施方案41.根据实施方案40所述的工程化Cas12b效应子蛋白,其包含第一多肽和第二多肽,其中所述第一多肽包含工程化Cas12b核酸酶或其功能衍生物的N末端氨基酸残基1至X,其中所述第二多肽包含第X+1位残基至所述工程化Cas12b核酸酶或其功能衍生物的C末端,其中所述第一多肽和第二多肽能够在含有指导序列的指导RNA的存在下相互缔合,以形成特异性结合靶核酸的成簇规律间隔短回文重复序列(CRISPR)复合物,其中所述靶核酸包含与指导序列互补的靶序列。

[0347] 实施方案42.根据实施方案40或41所述的工程化Cas12b效应子蛋白,其中所述第一多肽和所述第二多肽各自包含二聚化结构域。

[0348] 实施方案43.根据实施方案42所述的工程化Cas12b效应子蛋白,其中所述第一二聚化结构域和所述第二二聚化结构域在诱导物的存在下彼此缔合。

[0349] 实施方案44.根据实施方案40或41所述的工程化Cas12b效应子蛋白,其中所述第一多肽和第二多肽不包含二聚化结构域。

[0350] 实施方案45.一种工程化的CRISPR-Cas12b系统,其包含:(a)实施方案32-44中任一项所述的工程化Cas12b效应子蛋白,或编码所述工程化Cas12b效应子蛋白的核酸;和(b)指导RNA,其包含与靶序列互补的指导序列,或编码所述指导RNA的核酸,其中,所述工程化

Cas12b效应子蛋白和所述指导RNA能够形成特异性结合包含所述靶序列的靶核酸的CRISPR复合物,并诱导所述靶核酸的修饰。

[0351] 实施方案46.根据实施方案45所述的工程化的CRISPR-Cas12b系统,其中所述gRNA包含crRNA和tracrRNA。

[0352] 实施方案47.根据实施方案45或46所述的工程化的CRISPR-Cas12b系统,其包含编码多个crRNA的前体指导RNA阵列。

[0353] 实施方案48.根据实施方案45-47中任一项所述的工程化的CRISPR-Cas12b系统,其中所述指导RNA是单指导RNA (sgRNA)。

[0354] 实施方案49.根据实施方案45-48中任一项所述的工程化的CRISPR-Cas12b系统,其包含编码所述工程化Cas12b效应子蛋白的一个或多个载体。

[0355] 实施方案50.根据实施方案49所述的工程化的CRISPR-Cas12b系统,其中所述一个或多个载体是腺相关病毒(AAV)载体。

[0356] 实施方案51.根据实施方案50所述的工程化的CRISPR-Cas12b系统,其中所述AAV载体进一步编码所述指导RNA。

[0357] 实施方案52.一种检测样品中靶核酸的方法,其包括:(a)使所述样品与实施方案45-51中任一项所述的工程化的CRISPR-Cas12b系统和标记的检测核酸接触,其中所述标记的检测核酸是单链的并且不与所述指导RNA的指导序列杂交;和(b)测量由所述工程化Cas12b效应子蛋白切割所述标记的检测核酸而产生的可检测信号,由此检测所述靶核酸。

[0358] 实施方案53.一种修饰包含靶序列的靶核酸的方法,其包括使所述靶核酸与实施方案45-51中任一项所述的工程化的CRISPR-Cas12b系统接触。

[0359] 实施方案54.根据实施方案53所述的方法,其中所述方法在体外进行。

[0360] 实施方案55.根据实施方案53所述的方法,其中所述靶核酸存在于细胞中。

[0361] 实施方案56.根据实施方案55所述的方法,其中所述细胞是细菌细胞、酵母细胞、哺乳动物细胞、植物细胞、或动物细胞。

[0362] 实施方案57.根据实施方案53所述的方法,其中所述方法离体进行。

[0363] 实施方案58.根据实施方案53所述的方法,其中所述方法在体内进行。

[0364] 实施方案59.根据实施方案53-58中任一项所述的方法,其中通过所述工程化的CRISPR-Cas12b系统切割所述靶核酸或改变所述靶核酸中的所述靶序列。

[0365] 实施方案60.根据实施方案53-58中任一项所述的方法,其中通过所述工程化的CRISPR-Cas12b系统改变所述靶核酸的表达。

[0366] 实施方案61.根据实施方案53-60中任一项所述的方法,其中所述靶核酸是基因组DNA。

[0367] 实施方案62.根据实施方案53-61中任一项所述的方法,其中所述靶序列与疾病或病况相关。

[0368] 实施方案63.根据实施方案53-62中任一项所述的方法,其中所述工程化的CRISPR-Cas12b系统包含编码多个crRNA的前体指导RNA阵列,其中每个crRNA包含不同的指导序列。

[0369] 实施方案64.一种治疗与个体的细胞中的靶核酸相关的疾病或病况的方法,其包括使用实施方案45-51中任一项所述的工程化的CRISPR-Cas12b系统修饰所述个体的细胞

中的所述靶核酸,由此治疗所述疾病或病况。

[0370] 实施方案65.根据实施方案64所述的方法,其中所述疾病或病况选自下组:癌症、心血管疾病、遗传性疾病、自身免疫性疾病、代谢疾病、神经退行性疾病、眼部疾病、细菌感染和病毒感染。

[0371] 实施方案66.一种包含修饰的靶核酸的工程化的细胞,其中所述靶核酸已使用实施方案53-63中任一项所述的方法被修饰。

[0372] 实施方案67.一种工程化的非人类动物,其包含一个或多个根据实施方案66所述的工程化的细胞。

实施例

[0373] 以下实施例仅用于举例说明本申请,并且因此不应被视为以任何方式限制本申请。提供下列实施例和详细描述意在示例说明,而非限制。

[0374] 方法

[0375] 质粒的构建

[0376] AaCas12b的编码序列经过密码子优化以在人类细胞中表达和合成。通过基于PCR的定点诱变产生编码工程化的AaCas12b蛋白突变体的核酸序列。具体来说,编码参考AaCas12b蛋白的DNA序列以突变位点为中心分为两部分。设计两对引物以分别扩增两部分DNA序列,通过Gibson克隆将其组装成一段DNA片段,并整合到pCAG-2A-eGFP载体中。通过将编码参考AaCas12b蛋白的DNA断裂成多段,然后使用PCR和Gibson克隆进行扩增和组装,构建突变组合。编码所述工程化AaCas12b蛋白的DNA被插入pCAG-2A-eGFP载体的XmaI和NheI位点之间。AaCas12b蛋白变体的突变位点是基于使用常用于本领域的蛋白质结构可视化软件(例如PyMol或Chimera)对AaCas12b的晶体结构进行分析来设计的。AaCas12b的晶体结构可在RCSB PDB数据库中得到,访问编号为6LTU、6LTR、6LU0和6LTP。所述AaCas12b变体使用pCAG-2A-eGFP载体在人类293T细胞中表达。编码sgRNA支架的DNA序列从头合成并经由Gibson克隆组装成pUC19-U6骨架。编码间隔子序列的核酸也被连接到相同的pUC19-U6骨架中。

[0377] 细胞培养、转染和荧光激活细胞分选(FACS)

[0378] HEK293T细胞在含有1%青霉素-链霉素(Gibco)和10%胎牛血清(Gibco)的DMEM(Gibco)中培养。将细胞接种于24孔培养皿(Corning)中,培养16小时,直至细胞汇合度达到70%。使用Lipofectamine 3000(Invitrogen),将600ng的编码AaCas12b蛋白的质粒和不同量的编码sgRNA的质粒转染到24孔培养皿的每个孔中的细胞中。转染68小时后,通过胰蛋白酶-EDTA(0.05%)(Gibco)消化所述HEK293T细胞,并基于GFP信号(表明转染成功)使用MoFlo XDP(Beckman Coulter)进行FACS分选。

[0379] 针对基因组修饰的靶向的深度测序分析

[0380] 将通过FACS分选出的GFP阳性HEK293T细胞用缓冲液L裂解,在55°C下孵育3小时,然后在95°C下孵育10分钟。使用相应的引物对不同基因座上含有靶位点的dsDNA片段进行PCR扩增。针对靶向的深度测序,所述细胞裂解物直接用作模板DNA,进行条形码PCR扩增。将PCR产物纯化后汇集到几个文库中进行高通量测序。通过计算包含插入或缺失的读数的比率,使用CRISPResso2软件分析插入/缺失的频率(%)。在本应用中,使用插入/缺失频率

(%)的指标来比较和分析不同工程化的Cas12b蛋白和/或不同sgRNA支架存在下的基因编辑效率。少于总读数的0.05%的读数的任何数值都会被丢弃。

[0381] 实施例1:参考AaCas12b核酸酶中与PAM相互作用的的一个或多个氨基酸残基用带正电荷的氨基酸残基的取代。

[0382] 按照上述方法设计并表达了与PAM相互作用的氨基酸残基中具有单点突变的工程化AaCas12b酶。简而言之,选择AaCas12b中PAM的9 Å内的10个氨基酸:D116、K123、D130、D132、N144、K145、E153、D173、Q222、D395、N400和E475,并用精氨酸(R)取代每个氨基酸残基。设计针对靶位点CCR5-11(SEQ ID NO:63)、CD34-7(SEQ ID NO:64)和RNF2-1(SEQ ID NO:65)的编码sgRNA的核酸,其从5'到3'包含:编码Aa-sg sgRNA支架序列的DNA(SEQ ID NO:23)一编码间隔子序列的DNA,并将其克隆到pUC19-U6骨架中。使用Lipofectamine 3000(Invitrogen),将编码AaCas12b蛋白的600ng质粒和编码sgRNA的300ng质粒如上述方法转染到24孔培养皿的每个孔中的HEK293T细胞中。野生型AaCas12b(SEQ ID NO:1)作为对照。AaCas12b酶中的氨基酸取代及相应的基因编辑效率如图1和表1所示。与野生型AaCas12b相比,具有氨基酸取代D116R(SEQ ID NO:2)或E475R(SEQ ID NO:3)的AaCas12b变体显示出提高的基因编辑效率。如图1所示,AaCas12b-D116R(SEQ ID NO:2)和AaCas12b-E475R(SEQ ID NO:3)在三个基因组位点上的平均基因编辑效率超过约20%,而参考野生型AaCas12b核酸酶的平均基因编辑效率约为6%。AaCas12b-D116R(SEQ ID NO:2)和AaCas12b-E475R(SEQ ID NO:3)突变体的插入/缺失频率显著高于使用此类测试的其他AaCas12b突变体的插入/缺失频率。与野生型AaCas12b相比,AaCas12b-D395R在CD34-7基因座上实现了更高的基因编辑效率,但在其他测试的基因座上则不然。

[0383] 表1.不同AaCas12b在不同基因座的基因编辑效率

AaCas12b 变体	CCR5-11 的插入/缺失(%)	CD34-7 的插入/缺失(%)	RNF2-1 的插入/缺失(%)
WT (SEQ ID NO: 1)	2.87	9.20	6.60
D116R (SEQ ID NO: 2)	9.09	32.75	24.74
K123R	2.28	3.59	3.73
D130R	0.00	0.00	0.00
D132R	1.83	4.56	2.58
N144R	3.25	8.67	5.33
K145R	2.18	4.68	3.70
E153R	3.00	7.42	7.09
D173R	0.11	0.12	0.15
Q222R	0.00	0.07	0.00
D395R	1.54	22.84	1.37
N400R	0.04	2.96	0.59
E475R (SEQ ID NO: 3)	8.15	51.42	18.70

[0385] 实施例2:参考AaCas12b核酸酶中参与打开DNA双链的一个或多个氨基酸残基用具有芳香环的氨基酸残基的取代

[0386] 按照上述方法设计并表达了工程化的AaCas12b核酸酶,该核酸酶在参与打开DNA双链的氨基酸残基上进行了单个取代。简而言之,氨基酸残基Q118或Q119被芳香族氨基酸残基(例如,Y、F或W)取代。这里使用了与实施例1中相同的sgRNA编码质粒。使用Lipofectamine 3000(Invitrogen),将编码AaCas12b蛋白的600ng质粒和编码sgRNA的300ng质粒如上述方法转染到24孔培养皿的每个孔中的HEK293T细胞中。野生型AaCas12b

(SEQ ID NO:1) 作为对照。AaCas12b酶中的氨基酸取代及相应的基因编辑效率如图2和表2所示。与野生型AaCas12b相比,具有氨基酸取代Q119Y、Q119F或Q119W的AaCas12b在所有测试的基因座均显示出提高的基因编辑效率。AaCas12b-Q119Y、AaCas12b-Q119F和AaCas12b-Q119W突变体的插入/缺失频率显著高于使用此类测试的其他AaCas12b突变体(Q118Y、Q118F、Q118W)的插入/缺失频率。

[0387] 表2. 不同AaCas12b在不同基因座的基因编辑效率

AaCas12b变体	CCR5-11的插入/缺失(%)	CD34-7的插入/缺失(%)	RNF2-1的插入/缺失(%)
WT (SEQ ID NO: 1)	2.87	9.20	6.60
Q118Y	0.00	0.07	0.00
Q118F	0.03	0.09	0.04
Q118W	0.08	0.41	0.29
Q119Y (SEQ ID NO: 4)	5.84	12.77	12.00
Q119F (SEQ ID NO: 5)	6.98	22.28	15.66
Q119W (SEQ ID NO: 6)	9.56	14.95	11.32

[0389] 实施例3:参考AaCas12b核酸酶的RuvC结构域中与单链DNA底物相互作用的一个或多个氨基酸残基用带正电荷的氨基酸残基或疏水性氨基酸残基的取代。

[0390] 按照上述方法设计并表达了与单链DNA底物相互作用的RuvC结构域中的氨基酸残基进行单个氨基酸取代的工程化AaCas12b核酸酶。设计针对靶位点CCR5-3(SEQ ID NO:66)和RNF2-5(SEQ ID NO:67)的编码sgRNA的核酸,其从5'到3'包含:编码Aa-sgRNA支架序列的DNA(SEQ ID NO:23)一编码间隔子序列的DNA,并将其克隆到pUC19-U6骨架中。使用Lipofectamine3000(Invitrogen),将编码AaCas12b蛋白的600ng质粒和编码sgRNA的300ng质粒转染到24孔培养皿的每个孔中的HEK293T细胞中。野生型AaCas12b(SEQ ID NO:1)作为对照。

[0391] 在第一组AaCas12b突变体中,表3中的每个氨基酸残基都被带正电荷的氨基酸残基精氨酸(R)取代。AaCas12b酶中的氨基酸取代和相应的基因编辑效率如图3-4B和表3所示。

[0392] 表3. 不同AaCas12b在不同基因座的基因编辑效率

AaCas12b变体	CCR5-3的插入/缺失(%)	RNF2-5的插入/缺失(%)
WT	11.54	4.63
D300R	16.50	6.38
K301R	18.83	9.84
E304R	7.37	2.79
N329R	3.56	0.32
E636R(SEQ ID NO:7)	34.92	19.54
Q639R	22.75	11.00
T647R	26.25	16.53
Q682R	22.73	12.47
I757R(SEQ ID NO:8)	37.69	24.15
E758R(SEQ ID NO:9)	61.70	49.84

E761R(SEQ ID NO:10)	38.78	27.33
E764R	12.18	5.14
K768R	23.53	12.28
E852R	11.17	10.33
Q854R(SEQ ID NO:11)	34.66	26.88
N856R	2.66	0.79
N857R	14.85	10.22
D858R(SEQ ID NO:12)	46.16	42.42
P860R	1.22	0.13
S862R	1.49	0.30
E863R	1.48	0.17
N865R	17.78	19.48
Q866R	24.89	14.88
L867R	1.72	0.18
Q869R	9.49	2.11
E938R	1.92	1.17
E956R	0.81	0.29
G957R	1.37	1.00
E958R	5.65	2.97
I994R	28.14	27.87
Q1093R	14.68	10.62
W1097R	17.38	26.37

[0394] 在第二组AaCas12b突变体中,表4中的每个氨基酸残基都被带正电荷的氨基酸残基赖氨酸(K)取代。AaCas12b酶中的氨基酸取代和相应的基因编辑效率如表4和图4A-4B所示。

[0395] 如表3-4和图3-4B所示,与野生型AaCas12b相比,具有氨基酸取代D300R、K301R、E636R、Q639R、T647R、Q682R、I757R、E758R、E761R、K768R、Q854R、N857R、D858R、N865R、Q866R、I994R、Q1093R、W1097R、E636K、Q639K、T647K、Q682K、I757K、E758K、E761K、Q854K、N857K、D858K、N865K、I994K、Q1093K或W1097K的AaCas12b突变体在两个测试的基因座均具有提高的基因编辑效率。AaCas12b E636R、I757R、E758R、E761R、Q854R、D858R、E758K、N857K、I994R和D858K突变体的插入/缺失频率显著高于使用在此类中测试的其他AaCas12b突变体(用带正电荷的氨基酸残基取代)的插入/缺失频率。

[0396] 表4. 不同AaCas12b在不同基因座的基因编辑效率

[0397]

AaCas12b变体	CCR5-3的插入/缺失(%)	RNF2-5的插入/缺失(%)
WT	11.54	4.63
E636K	39.58	26.45
Q639K	24.23	12.68
T647K	28.84	17.31
Q682K	18.46	12.10

I757K	22.40	8.53
E758K	50.37	38.27
E761K	28.20	16.90
Q854K	24.24	17.08
N857K (SEQ ID NO:13)	45.97	43.19
D858K	42.27	41.78
N865K	20.81	7.27
Q866K	12.39	2.70
I994K	14.20	8.66
Q1093K	20.60	23.15
W1097K	21.66	33.05

[0398] 在第三组AaCas12b突变体中,以下每个氨基酸残基都被疏水性氨基酸残基(例如, Y、F、M或W)取代:E758、E761、E863、N865、Q866、Q869、E956和Q1093。AaCas12b酶中的氨基酸取代及相应的基因编辑效率如图5和表5所示。与野生型AaCas12b相比,具有氨基酸取代 E758W、E758Y、E758M、E761Y、N865W、N865Y、N865F、Q866M、Q869M、Q1093W、Q1093Y、Q1093F或 Q1093M的AaCas12b突变体在两个测试的基因座均显示出提高的基因编辑效率。AaCas12b N865W、N865Y、Q866M、Q869M、Q1093W和Q1093Y突变体的插入/缺失频率显著高于此类中测试的其他AaCas12b突变体(用疏水性氨基酸残基取代)的插入/缺失频率。

[0399] 表5. 不同AaCas12b在不同基因座的基因编辑效率

AaCas12b变体	CCR5-3的插入/缺失(%)	RNF2-5的插入/缺失(%)
WT	11.54	4.63
E758W	24.31	17.13
E758Y	34.73	17.20
E758F	11.57	3.87

[0400]

	E758M	22.39	9.52
	E761W	5.17	1.87
	E761Y	21.47	11.78
	E761F	8.02	4.06
	E761M	9.86	4.55
	E863W	1.48	0.36
	E863Y	5.58	0.52
	E863F	1.49	0.35
	E863M	2.86	0.48
	N865W (SEQ ID NO: 14)	32.50	23.18
	N865Y (SEQ ID NO: 15)	51.61	55.65
	N865F	14.85	9.85
	N865M	6.25	2.49
	Q866W	2.95	0.27
[0401]	Q866Y	1.24	0.27
	Q866F	4.71	1.47
	Q866M (SEQ ID NO: 16)	50.09	47.76
	Q869W	7.93	0.61
	Q869Y	8.47	1.83
	Q869F	10.70	2.57
	Q869M (SEQ ID NO: 17)	20.65	15.71
	E956W	4.73	1.88
	E956Y	5.57	1.71
	E956F	4.90	2.85
	E956M	2.66	0.59
	Q1093W (SEQ ID NO: 18)	29.32	26.26
	Q1093Y (SEQ ID NO: 19)	23.73	27.12
	Q1093F	15.97	12.65
	Q1093M	18.14	9.92

[0402] 实施例4: 实施例1-3中的突变的组合及其基因编辑效率的表征。

[0403] 实施例1、2和3中筛选的具有所需基因编辑效率的氨基酸取代, 即Q866M、Q869M、I757R、E758R、E761R、K768R和I757R, 以制备具有多个突变的AaCas12b蛋白, 即Q866M+Q869M、I757R+E758R、I757R+E761R、I757R+K768R、E758R+E761R、E758R+K768R、E761R+K768R、I757R+E758R+E761R、I757R+E758R+K768R、I757R+E761R+K768R、E758R+E761R+K768R和I757R+E758R+E761R+K768R。设计针对靶位点CCR5-3 (SEQ ID NO: 66)、CCR5-11 (SEQ ID NO: 63)、CD34-1 (SEQ ID NO: 68) 和RNF2-5 (SEQ ID NO: 67) 的编码sgRNA的核酸, 其从5'到3'包含: 编码Aa-sg sgRNA支架序列的DNA (SEQ ID NO: 23) — 编码间隔子序列的DNA, 并将其克隆到pUC19-U6骨架中。野生型AaCas12b (SEQ ID NO: 1) 作为对照。使用Lipofectamine 3000 (Invitrogen), 将编码上述AaCas12b蛋白的600ng质粒和编码sgRNA的300ng质粒转染到24孔培养皿的每个孔中的HEK293T细胞中。它们的基因编辑效率如图6和表6所示。与野生型AaCas12b相比, 具有氨基酸取代组合的AaCas12b突变体在所有测试的基因座均表现出显著提高的基因编辑效率。某些AaCas12b组合突变体, 如Q866M+Q869M、E758R+E761R、E758R+E768R、I757R+E758R+K768R和E758R+E761R+K768R, 与相应的单突变体相比, 在某些基因座具有提高的基因编辑效率。

[0404] 表6. 不同AaCas12b在不同基因座的基因编辑效率

[0405]

AaCas12b 变体	CCR5-3 的插入/缺失(%)	CCR5-11 的插入/缺失(%)	CD34-1 的插入/缺失(%)	RNF2-5 的插入/缺失(%)
WT	14.52	3.42	2.16	6.17
Q866M (SEQ ID NO: 16)	45.78	14.21	7.98	51.87
Q869M (SEQ ID NO: 17)	17.16	8.03	6.11	20.01
Q866M+Q869M (SEQ ID NO: 20)	56.97	17.48	12.39	70.75
I757R (SEQ ID NO: 8)	38.29	6.51	5.29	23.37
E758R (SEQ ID NO: 9)	54.69	12.97	6.55	44.75
E761R (SEQ ID NO: 10)	41.32	12.22	4.17	26.48
K768R	23.14	2.61	2.76	14.33
I757R+E758R	49.55	14.89	8.48	50.96
I757R+E761R	28.80	11.21	5.94	23.57
I757R+K768R	40.14	7.24	9.45	34.32
E758R+E761R	56.90	13.48	5.64	49.74
E758R+K768R	62.90	7.57	7.85	56.64
E761R+K768R	43.56	10.26	5.86	38.56
I757R+E758R+E761R	15.24	9.45	3.26	19.00
I757R+E758R+K768R	50.13	12.45	8.93	56.24
I757R+E761R+K768R	29.07	9.77	6.83	27.37
E758R+E761R+K768R	55.44	12.94	6.63	60.03
I757R+E758R+E761R+K768R	34.10	8.81	4.56	28.23

[0406] AaCas12b-Q119F+E475R和AaCas12b-Q119F+E475R+E758R是如上所述生成的。这里使用了与实施例1中相同的sgRNA编码质粒。野生型AaCas12b (SEQ ID NO:1) 作为对照。使用 Lipofectamine 3000 (Invitrogen), 将编码上述AaCas12b蛋白的600ng质粒和编码sgRNA的300ng质粒转染到24孔培养皿的每个孔中的HEK293T细胞中。它们的基因编辑效率如图7和表7所示。结果表明, 与野生型AaCas12b相比, AaCas12b-Q119F+E475R和AaCas12b-Q119F+E475R+E758R在所有测试的基因座均显著提高了基因编辑效率。与野生型AaCas12b或具有单个取代的相应AaCas12b变体相比, 在所有测试的基因座 (CCR5-11、CD34-7和RNF2-1), AaCas12b-Q119F+E475R+E758R在基因编辑效率方面表现出最显著的提高。

[0407] 表7. 不同AaCas12b在不同基因座的基因编辑效率

[0408]

AaCas12b 变体	CCR5-11 的插入/缺失(%)	CD34-7 的插入/缺失(%)	RNF2-1 的插入/缺失(%)
WT	3.06	9.71	6.72
Q119F (SEQ ID NO: 5)	7.02	22.02	16.21
E475R (SEQ ID NO: 3)	8.29	52.13	18.87
Q119F+E475R (SEQ ID NO: 21)	12.95	51.89	23.39
E758R (SEQ ID NO: 9)	6.46	10.87	8.31
Q119F+E475R+E758R (SEQ ID NO: 22)	35.68	59.39	42.22

[0409] 实施例5: 使用带有工程化支架的sgRNA增强工程化AaCas12b的基因编辑活性。

[0410] 在此实施例中, 使用实施例4中的AaCas12b突变体 (Q119F+E475R+E758R) 测试了具有工程化支架的多种sgRNA的基因编辑活性。设计针对靶位点CCR5-11 (SEQ ID NO:63) 的编

码sgRNA的核酸,其从5'到3'包含:编码sgRNA支架序列的DNA—编码间隔子序列的DNA,并将其克隆到pUC19-U6骨架中。使用Lipofectamine 3000(Invitrogen),将编码AaCas12b变体蛋白的600ng质粒和编码具有工程化支架的sgRNA的300ng质粒(SEQ ID NO:25-53;基于AacCas12b sgRNA支架V0进行修饰的)、AacCas12b sgRNA支架(SEQ ID NO:24;V0,对照;H.Yang et al.,Cell.2016;167(7):1814-1828.e12)、或AaCas12b Aa-sg支架(SEQ ID NO:23;对照)转染到24孔培养皿每个孔中的HEK293T细胞中。它们的基因编辑效率如图9所示。图9中的数据显示,与AacCas12b sgRNA支架(V0)相比,所有sgRNA工程化支架均显著提高了AaCas12b(Q119F+E475R+E758R)变体的基因编辑效率。与Aa-sg支架相比,所有sgRNA工程化支架(V1和V8除外)也显著提高了AaCas12b(Q119F+E475R+E758R)变体的基因编辑效率。

[0411] 实施例6:具有失活的核酸酶活性的工程化AaCas12b。

[0412] 为了生成失活的AaCas12b蛋白,来自实施例4的AaCas12b(Q119F+E475R+E758R)变体(SEQ ID NO:22)被进一步修饰以在核酸水解结构域中包含另外的单点突变(D570A)(图10A)。使用与上述类似的方法,将共同编码i)在CMV启动子控制下的AaCas12b(Q119F+E475R+E758R)或AaCas12b(Q119F+E475R+E758R+D570A)(SEQ ID NO:79),和ii)在U6启动子控制下的对照sgRNA(不靶向血红蛋白亚基 γ 1/2(HBG1/2)内的任何序列)、sgRNA1(针对HBG1/2的靶序列SEQ ID NO:70)、或sgRNA2(针对HBG1/2的靶序列SEQ ID NO:71)的质粒转染到HEK293细胞中(表8;质粒构建参见图10A)。使用sgRNA支架(V9)(SEQ ID NO:53)构建这些sgRNA。转染后3天,从转染的细胞中提取基因组DNA。进行T7内切酶I(T7EI)错配检测测定以确定切割效率(M.Crispo et al.,PLoS One.2015;10(8):e0136690)。T7EI测定中所用的引物序列列于表9中。

[0413] 如图10B所示,在sgRNA1或sgRNA2指导下切割HBG1/2的两个不同靶位点时,AaCas12b(Q119F+E475R+E758R+D570A)的催化活性与AaCas12b(Q119F+E475R+E758R)相比显著降低。

[0414] 表8.PAM和靶向HBG1/2的sgRNA的靶位点

sgRNA	PAM	靶序列
sgRNA1	TTG	AGATAGTGTGGGAAGGGGC (SEQ ID NO:70)
sgRNA2	TTT	GCATTGAGATAGTGTGGGA (SEQ ID NO:71)

[0416] 表9.T7EI测定中使用的引物序列

SEQ ID NO	引物序列
69	TCCTGCACTGAACTGTTGC
78	TCCTGAGAAGCGACCTGGA

[0418] 为了进一步降低工程化AaCas12b的核酸酶活性,在AaCas12b(Q119F+E475R+E758R+D570A)中引入另外的点突变以生成AaCas12b(Q119F+E475R+E758R+D570A+E848A)(SEQ ID NO:80)或AaCas12b(Q119F+E475R+E758R+D570A+D977A)(SEQ ID NO:81)(图11A)。使用与上述类似的方法,将共同编码i)在CMV启动子控制下的AaCas12b(Q119F+E475R+E758R)、AaCas12b(Q119F+E475R+E758R+D570A+E848A)、或AaCas12b(Q119F+E475R+E758R+D570A+D977A),和ii)在U6启动子控制下的sgRNA1(针对HBG1/2的靶序列SEQ ID NO:70)或sgRNA2(针对HBG1/2的靶序列SEQ ID NO:71)的质粒转染到HEK293细胞中(质粒构建参见图11A)。作为阴性对照,将编码AaCas12b(Q119F+E475R+E758R)、AaCas12b(Q119F+E475R+E758R+

D570A+E848A)、或AaCas12b(Q119F+E475R+E758R+D570A+D977A)的质粒和对照sgRNA(不靶向血红蛋白亚基 γ 1/2(HBG1/2)内的任何序列)类似地转染到没有任何编码sgRNA序列的HEK293细胞中。如图11B所示,AaCas12b(Q119F+E475R+E758R+D570A+E848A)和AaCas12b(Q119F+E475R+E758R+D570A+D977A)两者完全消除了AaCas12b(Q119F+E475R+E758R)的核酸酶活性。

[0419] 实施例7:使用工程化AaCas12b融合蛋白进行转录抑制。

[0420] 来自实施例6的AaCas12b(Q119F+E475R+E758R+D570A+D977A)(SEQ ID NO:81)经过进一步工程化以生成融合蛋白来沉默靶基因的转录。AaCas12b(Q119F+E475R+E758R+D570A+D977A)(两侧翼是两个拷贝的核定位序列NLS)与转录抑制模块ZIM3的Krüppel相关盒(KRAB)结构域(SEQ ID NO:72)融合,这可以招募抑制性染色质修饰因子。KRAB融合在AaCas12b(Q119F+E475R+E758R+D570A+D977A)的C末端或在N末端,融合蛋白分别命名为Cd12bk和Nd12bk。同一质粒还编码在U6启动子控制下特异性识别SCN9A基因(编码电压门控的钠通道1.7Nav1.7)中不同靶位点的sgRNA(图12A;表10)。使用sgRNA支架(V9)(SEQ ID NO:53)来构建这些sgRNA。

[0421] 为了检查Cd12bk和Nd12bk融合蛋白是否可以募集染色质修饰复合物来沉默SCN9A的转录,将编码融合蛋白和sgRNA的质粒转染到Neuro 2A(N2a;小鼠神经嵴衍生的细胞系)细胞中。作为对照,将编码Cd12bk的质粒与对照sgRNA(不靶向SCN9A内的任何序列)类似地转染到N2a细胞中。转染后3天,收集转染的细胞并使用RNA提取试剂盒(Vazyme,货号RC112-01)提取RNA。通过qPCR确定每个样本中的Nav1.7的mRNA水平。数据使用对照sgRNA("Cd12bk-non-target")归一化为Cd12bk。如图12B所示,Cd12bk或Nd12bk与sgRNA-msg6、sgRNA-msg8、sgRNA-msg13或sgRNA-msg18一起均能大大地抑制SCN9A的转录,其中sgRNA-msg8和sgRNA-msg13显示最强抑制。Nd12bk与sgRNA-msg11一起也能大大地抑制SCN9A的转录。这些结果表明,与KRAB融合的dAaCas12b(例如,AaCas12b(Q119F+E475R+E758R+D570A+D977A))可用作真核细胞中靶向的转录调控工具。

[0422] 表10.PAM和靶向SCN9A的sgRNA的靶位点

[0423]

sgRNA	PAM	靶位点
msg6	TTA	GCTGCCCGCCACACTGGCGC(SEQ ID NO:73)
msg8	TTG	GGCGTGGTGATGCTAGGGAT(SEQ ID NO:74)
msg11	TTC	TAGTCTGCTCAGGATGAAGC(SEQ ID NO:75)
msg13	TTC	AATCCTGCCCACTGTGCAGG(SEQ ID NO:76)
msg18	TTC	CCTTGATCAGAATCCGCAG(SEQ ID NO:77)

[0424] 尽管以上结合附图对本申请的实施方案进行了描述,但是本申请不限于上述具体实施方案和应用领域。上述具体实施方案仅是示例说明的、指导性的,而非限制性的。在本说明书的启示下,并且在不脱离本申请权利要求的保护范围的情况下,本领域的普通技术人员还可以做出许多形式,这些形式都属于本申请的保护范围。

[0425] 示例性序列

[0426]

#	序列
SEQ ID NO: 1 野生型 AaCas12b	MAVKSMKVKLRLDNMPEIRAGLWKLHTEVNAGVRYYTEWLSLLRQENLYRRSP NGDGEQECYKTAEECKAELLERLRARQVENGHCGPAGSDDPELLQLARQLYELLV PQAIGAKGDAQIARKFLSPLADKDAVGGGLGIAKAGNKPRWVRMREAGEPGWEE EKAKAEARKSTDRTADVLRALADFLKPLMRVYTDSDMSSVQWKPLRKGQAVR TWDRDMFQQAIERMMSWESWNQRVGEAYAKLVEQKSRFEQKNFVQGEHLVQL VNQLQQDMKEASHGLESKEQTAHYLTGRALRGSDKVFEKWEKLDPDAPFDLYDT EIKNVQRRNTRRFGSHDLFAKLAEPKYQALWREDASFLTRYAVYNSIVRKLNHAK MFATFTLPDATAHPIWTRFDKLGGNLHQYTFLFNEFGEGRHAIRFQKLLTVEDGV AKEVDDVTPISMSAQLDDLLPRDPHELVALYFQDYGAEQHLAGEFGGAKIQYRR DQLNHLHARRGARDVYLNLSVRVQSQSEARGERRPPYAAVFRVLVGDNHRAVHF DKLSDYLAEHPDDGKLGSEGLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNSE GRVPFCFPIEGNENLVAVHERSQQLLKLPGETESKDLRAIREERQRTLRLQRLAYL RLLVRCGSEDEVGRRERSWAKLIEQPM DANQMTPDWREAFEDQLKLSLYGICG DREWTEAVYESVRRVWRHMKGQVRDWRKDVRSGERPKIRGYQKDVVGGNSIEQ IEYLERQYKFLKSWFFGKVSQGVIRAEKGSRFAITLREHIDHAKEDRLKKLADRII MEALGYVYALDDERGGKGVVAKYPPCQLILLEELSEYQFNDRPPSENNQLMQW SHRGVVFQELLNQAQVHDLVGTMYAAFSSRFDARTGAPGIRCRRV PARCAREQNP EPFPWWLNKFVAEHKLDGCPLRADDLIPTGEGEFFVSPFSAEEGDFHIHADLNAA QNLQRRLLWSDFDISQIRLRCDWGEVDGEPVLIPRTTGKRTADSYGNKVFFYTKTGV TYYERERGGKRRKVFQAEELSEEEAELLVEADEAREKSVVLMRDPGSIINRGDWT RQKEFWSMVNQRIEGYLVKQIRSRVRLQESACENTGDI
SEQ ID NO: 2 AaCas12b-D116R	MAVKSMKVKLRLDNMPEIRAGLWKLHTEVNAGVRYYTEWLSLLRQENLYRRSP NGDGEQECYKTAEECKAELLERLRARQVENGHCGPAGSDDPELLQLARQLYELLV PQAIGAKGRAQQAARKFLSPLADKDAVGGGLGIAKAGNKPRWVRMREAGEPGWEE EKAKAEARKSTDRTADVLRALADFLKPLMRVYTDSDMSSVQWKPLRKGQAVR TWDRDMFQQAIERMMSWESWNQRVGEAYAKLVEQKSRFEQKNFVQGEHLVQL VNQLQQDMKEASHGLESKEQTAHYLTGRALRGSDKVFEKWEKLDPDAPFDLYDT EIKNVQRRNTRRFGSHDLFAKLAEPKYQALWREDASFLTRYAVYNSIVRKLNHAK MFATFTLPDATAHPIWTRFDKLGGNLHQYTFLFNEFGEGRHAIRFQKLLTVEDGV AKEVDDVTPISMSAQLDDLLPRDPHELVALYFQDYGAEQHLAGEFGGAKIQYRR DQLNHLHARRGARDVYLNLSVRVQSQSEARGERRPPYAAVFRVLVGDNHRAVHF

[0427]

	DKLSDYLAEHPDDGKLGSEGLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNSE GRVPCFPIEGNENLVAVHERSQQLLKLPGETESKDLRAIREERQRTLRLQRLAYL RLLVRCGSEDEVGRRERSWAKLIEQPMQANQMPDWREAFEDLQKLSLYGICG DREWTEAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYQKDVVGGNSIEQ IEYLERQYKFLKSWFFGKVSQVIRAEKGSRAITLREHIDHAKEDRLKKLADRII MEALGYVYALDDERGGKGVVAKYPPCQLILLEELSEYQFNDRPPSENNQLMQW SHRGVVFQELLNQAQVHDLVGTMYAAFSSRFDARTGAPGIRCRVRCAREQNP EPPFWWLNKFVAEHKLDGCPLRADDLIPTGEGEFFVSPFSAEEGDFHQIHADLNAA QNLQRRLWSDFDISQIRLRCDWGEVDGEPVLIPRTTGKRTADSYGNKVFFYTKTGV TYERERGGKRRKVFQAEELSEEEAELLVEADEAREKSVVLMRDPGSIINRGDWT RQKEFWSMVNQRIEGLYLVKQIRSRVRLQESACENTGDI
SEQ ID NO: 3 AaCas12b-E475R	MAVKSMKVKLRLDNMPEIRAGLWKLHTEVNAAGVRYYTEWLSLLRQENLYRRSP NGDGEQECYKTAEECKAELLERLRARQVENGHCGPAGSDDELLQLARQLYELLV PQAIGAKGDAQIARKFLSPLADKDAVGGLGIAKAGNKPRWVRMREAGEPGWEE EKAKAEARKSTDRTADVLRALADFLKPLMRVYTDSDMSSVQWKPLRKGQAVR TWRDMMFQQAIERMMSWESWNQRVGEAYAKLVEQKSRFEQKNFVQGEHLVQL VNQLQDMKEASHGLESKEQTAHYLTGRALRGSDKVFKEKWEKLPDAPFDLYDT EIKNVQRRNTRRFSGHDLFAKLAEPKYQALWREDAFLTRYAVYNSIVRKLNHAK MFAFTLPDATAHPIWTRFDKLGGNLHQYTFLNEFGEGRHAIRFQKLLTVEDGV AKEVDDVTVPISMSAQLDDLLPRDPHELVALYFQDYGAEQHLAAGFGGAKIQYRR DQLNHLHARRGARDVYLNLSVRVQSQSEARGERRPPYAAVFRVLDGNHRAVHF DKLSDYLAEHPDDGKLGSEGLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNSE GRVPCFPIEGNENLVAVHERSQQLLKLPGETESKDLRAIREERQRTLRLQRLAYL RLLVRCGSEDEVGRRERSWAKLIEQPMQANQMPDWREAFEDLQKLSLYGICG DREWTEAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYQKDVVGGNSIEQ IEYLERQYKFLKSWFFGKVSQVIRAEKGSRAITLREHIDHAKEDRLKKLADRII MEALGYVYALDDERGGKGVVAKYPPCQLILLEELSEYQFNDRPPSENNQLMQW SHRGVVFQELLNQAQVHDLVGTMYAAFSSRFDARTGAPGIRCRVRCAREQNP EPPFWWLNKFVAEHKLDGCPLRADDLIPTGEGEFFVSPFSAEEGDFHQIHADLNAA QNLQRRLWSDFDISQIRLRCDWGEVDGEPVLIPRTTGKRTADSYGNKVFFYTKTGV TYERERGGKRRKVFQAEELSEEEAELLVEADEAREKSVVLMRDPGSIINRGDWT RQKEFWSMVNQRIEGLYLVKQIRSRVRLQESACENTGDI
SEQ ID NO: 4 AaCas12b-Q119Y	MAVKSMKVKLRLDNMPEIRAGLWKLHTEVNAAGVRYYTEWLSLLRQENLYRRSP NGDGEQECYKTAEECKAELLERLRARQVENGHCGPAGSDDELLQLARQLYELLV PQAIGAKGDAQIARKFLSPLADKDAVGGLGIAKAGNKPRWVRMREAGEPGWEE EKAKAEARKSTDRTADVLRALADFLKPLMRVYTDSDMSSVQWKPLRKGQAVR TWRDMMFQQAIERMMSWESWNQRVGEAYAKLVEQKSRFEQKNFVQGEHLVQL VNQLQDMKEASHGLESKEQTAHYLTGRALRGSDKVFKEKWEKLPDAPFDLYDT EIKNVQRRNTRRFSGHDLFAKLAEPKYQALWREDAFLTRYAVYNSIVRKLNHAK MFAFTLPDATAHPIWTRFDKLGGNLHQYTFLNEFGEGRHAIRFQKLLTVEDGV AKEVDDVTVPISMSAQLDDLLPRDPHELVALYFQDYGAEQHLAAGFGGAKIQYRR DQLNHLHARRGARDVYLNLSVRVQSQSEARGERRPPYAAVFRVLDGNHRAVHF DKLSDYLAEHPDDGKLGSEGLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNSE GRVPCFPIEGNENLVAVHERSQQLLKLPGETESKDLRAIREERQRTLRLQRLAYL RLLVRCGSEDEVGRRERSWAKLIEQPMQANQMPDWREAFEDLQKLSLYGICG DREWTEAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYQKDVVGGNSIEQ IEYLERQYKFLKSWFFGKVSQVIRAEKGSRAITLREHIDHAKEDRLKKLADRII MEALGYVYALDDERGGKGVVAKYPPCQLILLEELSEYQFNDRPPSENNQLMQW SHRGVVFQELLNQAQVHDLVGTMYAAFSSRFDARTGAPGIRCRVRCAREQNP EPPFWWLNKFVAEHKLDGCPLRADDLIPTGEGEFFVSPFSAEEGDFHQIHADLNAA QNLQRRLWSDFDISQIRLRCDWGEVDGEPVLIPRTTGKRTADSYGNKVFFYTKTGV TYERERGGKRRKVFQAEELSEEEAELLVEADEAREKSVVLMRDPGSIINRGDWT RQKEFWSMVNQRIEGLYLVKQIRSRVRLQESACENTGDI
SEQ ID NO: 5 AaCas12b-Q119F	MAVKSMKVKLRLDNMPEIRAGLWKLHTEVNAAGVRYYTEWLSLLRQENLYRRSP NGDGEQECYKTAEECKAELLERLRARQVENGHCGPAGSDDELLQLARQLYELLV PQAIGAKGDAQIARKFLSPLADKDAVGGLGIAKAGNKPRWVRMREAGEPGWEE EKAKAEARKSTDRTADVLRALADFLKPLMRVYTDSDMSSVQWKPLRKGQAVR TWRDMMFQQAIERMMSWESWNQRVGEAYAKLVEQKSRFEQKNFVQGEHLVQL VNQLQDMKEASHGLESKEQTAHYLTGRALRGSDKVFKEKWEKLPDAPFDLYDT

[0428]

	EIKNVQRRNTRRFGSHDLFAKLAEPKYQALWREDASFLTRYAVYNSIVRKLNHAK MFATFTLPDATAHPIWTRFDKLGGNLHQYTFLFNEFGEGRHAIRFQKLLTVEDGV AKEVDDVTVPISMSAQLDDLLPRDPHELVALYFQDYGAEQHLAGEFGGAKIQYRR DQLNHLHARRGARDVYLNLSVRVQSQSEARGERRPPYAAVFRVLVGDNHRAFVHF DKLSDYLAEHPDDGKLGSEGLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNSE GRVPPFCPIEGNENLVAVHERSQLLKLPGETESKDLRAIREERQRTLRLQRLAYL RLLVRCGSEDVGRRERSWAKLIEQPM DANQMPDWREAFEDLQKLSLYGICG DREWTEAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYQKDVVGGNSIEQ IEYLERQYKFLKSWSFFGK VSGQVIRA EKGSRFAITLREHIDHAKEDRLKKLADRII MEALGYVYALDDERGGKGVVAKYPPCQLILLEELSEYQFNDRPPSENNQLMQW SHRGVVFQELLNQAQVHDLVGTMYAAFSSRFDARTGAPGIRCRRVPARCAREQN EPFPWWLNKFVAEHKLDGCP LRADDLIPTGEGEFFVSPFSAEEGDFHQIHADLNA QNLQRRLLWSDFDISQIRLRCDWGEVDGEPVLIPRTTGKRTADSYGNKVFYTKTGV TYERERGGKRRKVF AQEELSEEEAELLVEADEAREKSVVLMRDPGSIINRGDWT RQKEFWSMVNQRIEGLVVKQIRSRVRLQESACENTGDI
SEQ ID NO: 6 AaCas12b-Q119W	MAVKSMKVKLRLDNMPEIRAGLWKLHTEV NAGVRYYTEWLSLLRQENLYRRSP NGDGEQECYKTAEECKAELLERLRARQVENGHCGPAGSDDELLQLARQLYELLV PQAIGAKGDAQWIARKFLSPLADKDAVGGGLGIAKAGNKPRWVRMREAGEPGWE EEKAKAEARKSTDRTADVLRALADFG LKPLMRVYTDSDMSSVQWKPLRKGQAV RTWDRDMFQQA IERMMSWESWNQRVGEAYAKLVEQKSRFEQKNFVQGEHLVQ LVNQLQQDMKEASHGLESKEQTAHYLTGRALRGSDKVFEKWEKLDPADPFDLYD TEIKNVQRRNTRRFGSHDLFAKLAEPKYQALWREDASFLTRYAVYNSIVRKLNHAK KMFAFTLPDATAHPIWTRFDKLGGNLHQYTFLFNEFGEGRHAIRFQKLLTVEDG VAKEVDDVTVPISMSAQLDDLLPRDPHELVALYFQDYGAEQHLAGEFGGAKIQYR RDQLNHLHARRGARDVYLNLSVRVQSQSEARGERRPPYAAVFRVLVGDNHRAFVH FDKLSYLAEHPDDGKLGSEGLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNSE EGRVPPFCPIEGNENLVAVHERSQLLKLPGETESKDLRAIREERQRTLRLQRLAYL RLLVRCGSEDVGRRERSWAKLIEQPM DANQMPDWREAFEDLQKLSLYGICG DREWTEAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYQKDVVGGNSIEQ QIEYLERQYKFLKSWSFFGK VSGQVIRA EKGSRFAITLREHIDHAKEDRLKKLADRI IMEALGYVYALDDERGGKGVVAKYPPCQLILLEELSEYQFNDRPPSENNQLMQ WSHRGVVFQELLNQAQVHDLVGTMYAAFSSRFDARTGAPGIRCRRVPARCAREQ NPEFPWWLNKFVAEHKLDGCP LRADDLIPTGEGEFFVSPFSAEEGDFHQIHADLN AAQNLQRRLLWSDFDISQIRLRCDWGEVDGEPVLIPRTTGKRTADSYGNKVFYTKT GVTYERERGGKRRKVF AQEELSEEEAELLVEADEAREKSVVLMRDPGSIINRGD WTRQKEFWSMVNQRIEGLVVKQIRSRVRLQESACENTGDI
SEQ ID NO: 7 AaCas12b-E636R	MAVKSMKVKLRLDNMPEIRAGLWKLHTEV NAGVRYYTEWLSLLRQENLYRRSP NGDGEQECYKTAEECKAELLERLRARQVENGHCGPAGSDDELLQLARQLYELLV PQAIGAKGDAQQIARKFLSPLADKDAVGGGLGIAKAGNKPRWVRMREAGEPGWEE EKAKAEARKSTDRTADVLRALADFG LKPLMRVYTDSDMSSVQWKPLRKGQAVR TWDRDMFQQA IERMMSWESWNQRVGEAYAKLVEQKSRFEQKNFVQGEHLVQ VNQLQQDMKEASHGLESKEQTAHYLTGRALRGSDKVFEKWEKLDPADPFDLYDT EIKNVQRRNTRRFGSHDLFAKLAEPKYQALWREDASFLTRYAVYNSIVRKLNHAK MFATFTLPDATAHPIWTRFDKLGGNLHQYTFLFNEFGEGRHAIRFQKLLTVEDGV AKEVDDVTVPISMSAQLDDLLPRDPHELVALYFQDYGAEQHLAGEFGGAKIQYRR DQLNHLHARRGARDVYLNLSVRVQSQSEARGERRPPYAAVFRVLVGDNHRAFVHF DKLSDYLAEHPDDGKLGSEGLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNSE GRVPPFCPIEGNENLVAVHERSQLLKLPGETESKDLRAIRRERQRTLRLQRLAYL LRLVRCGSEDVGRRERSWAKLIEQPM DANQMPDWREAFEDLQKLSLYGICG DREWTEAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYQKDVVGGNSIEQ QIEYLERQYKFLKSWSFFGK VSGQVIRA EKGSRFAITLREHIDHAKEDRLKKLADRI IMEALGYVYALDDERGGKGVVAKYPPCQLILLEELSEYQFNDRPPSENNQLMQ WSHRGVVFQELLNQAQVHDLVGTMYAAFSSRFDARTGAPGIRCRRVPARCAREQ NPEFPWWLNKFVAEHKLDGCP LRADDLIPTGEGEFFVSPFSAEEGDFHQIHADLN AAQNLQRRLLWSDFDISQIRLRCDWGEVDGEPVLIPRTTGKRTADSYGNKVFYTKT GVTYERERGGKRRKVF AQEELSEEEAELLVEADEAREKSVVLMRDPGSIINRGD WTRQKEFWSMVNQRIEGLVVKQIRSRVRLQESACENTGDI
SEQ ID NO: 8	MAVKSMKVKLRLDNMPEIRAGLWKLHTEV NAGVRYYTEWLSLLRQENLYRRSP NGDGEQECYKTAEECKAELLERLRARQVENGHCGPAGSDDELLQLARQLYELLV

[0429]

AaCas12b-I757R	PQAIGAKGDAQQIARKFLSPLADKDAVGGGLGIAKAGNKPRWVRMREAGEPGWEE EKAKAEARKSTDRTADVLRALADFLKPLMRVYTDSDMSSVQWKPLRKGQAVR TWDRDMFQQAIERMMSWESWNQRVGEAYAKLVEQKSRFEQKNFVQGEHLVQL VNQLQQDMKEASHGLESKEQTAHYLTGRALRGSDKVFKEKWEKLPDAPFDLYDT EIKNVQRRNTRRFSGHDLFAKLAEPKYQALWREDASFLTRYAVYNSIVRKLNHAK MFATFTLPDATAHPIWTRFDKLGGNLHQYTFLFNEFGEGRHAIRFQKLLTVEDGV AKEVDDVTVPISMSAQLDDLLPRDPHELVALYFQDYGAEQHLAGEFGGAKIQYRR DQLNHLHARRGARDVYLNLSVRVQSQSEARGERRPPYAAVFRVLDGNHRAFVHF DKLSDYLAEHPDDGKLGSEGLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNSE GRVPPFCPIEGNENLVAVHERSOLLKLPGETESKDLRAIREERQRTLRLQRTQLAYL RLLVRCGSEDVGRRERSWAKLIEQPMQANQMPDWREAFEDLQKLSLYGICG DREWTEAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYQKDVVGGNSRE QIEYLERQYKFLKSWFFGKVSQVIRAEKGSRAITLREHIDHAKEDRLKKLADRI IMEALGYVYALDDERGGKGVVAKYPPCQLILLEELSEYQFNDRPPSENNQLMQW SHRGVVFQELNQAQVHDLVGTMYAAFSSRFDARTGAPGIRCRVPARCAREQNP NPEFPWWLNKFVAEHKLDGCPLRADDLIPTGEGEFFVSPFSAEEGDFHQIHADLN AAQNLQRRLWSDFDISQIRLRCDWGEVDGEPVLIPTTGKRTADSYGNKVFYTKT GVTYYERERGGKRRKVFQAQEELSEEEAELLVEADEAREKSVVLMRDPGSIINRGD WTRQKEFWSMVNRQIEGYLVKQIRSRVRLQESACENTGDI
SEQ ID NO: 9 AaCas12b-E758R	MAVKSMKVKLRLDNMPEIRAGLWKLHTEVNAQVRYYTEWLSLLRQENLYRRSP NGDGEQECYKTAEECKAELLERLRARQVENGHCGPAGSDDELLQLARQLYELLV PQAIGAKGDAQQIARKFLSPLADKDAVGGGLGIAKAGNKPRWVRMREAGEPGWEE EKAKAEARKSTDRTADVLRALADFLKPLMRVYTDSDMSSVQWKPLRKGQAVR TWDRDMFQQAIERMMSWESWNQRVGEAYAKLVEQKSRFEQKNFVQGEHLVQL VNQLQQDMKEASHGLESKEQTAHYLTGRALRGSDKVFKEKWEKLPDAPFDLYDT EIKNVQRRNTRRFSGHDLFAKLAEPKYQALWREDASFLTRYAVYNSIVRKLNHAK MFATFTLPDATAHPIWTRFDKLGGNLHQYTFLFNEFGEGRHAIRFQKLLTVEDGV AKEVDDVTVPISMSAQLDDLLPRDPHELVALYFQDYGAEQHLAGEFGGAKIQYRR DQLNHLHARRGARDVYLNLSVRVQSQSEARGERRPPYAAVFRVLDGNHRAFVHF DKLSDYLAEHPDDGKLGSEGLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNSE GRVPPFCPIEGNENLVAVHERSOLLKLPGETESKDLRAIREERQRTLRLQRTQLAYL RLLVRCGSEDVGRRERSWAKLIEQPMQANQMPDWREAFEDLQKLSLYGICG DREWTEAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYQKDVVGGNSIRQ IEYLERQYKFLKSWFFGKVSQVIRAEKGSRAITLREHIDHAKEDRLKKLADRII MEALGYVYALDDERGGKGVVAKYPPCQLILLEELSEYQFNDRPPSENNQLMQW SHRGVVFQELNQAQVHDLVGTMYAAFSSRFDARTGAPGIRCRVPARCAREQNP EFPFWLNKFVAEHKLDGCPLRADDLIPTGEGEFFVSPFSAEEGDFHQIHADLNAA QNLQRRLWSDFDISQIRLRCDWGEVDGEPVLIPTTGKRTADSYGNKVFYTKTGV TYERERGGKRRKVFQAQEELSEEEAELLVEADEAREKSVVLMRDPGSIINRGDWT RQKEFWSMVNRQIEGYLVKQIRSRVRLQESACENTGDI
SEQ ID NO: 10 AaCas12b-E761R	MAVKSMKVKLRLDNMPEIRAGLWKLHTEVNAQVRYYTEWLSLLRQENLYRRSP NGDGEQECYKTAEECKAELLERLRARQVENGHCGPAGSDDELLQLARQLYELLV PQAIGAKGDAQQIARKFLSPLADKDAVGGGLGIAKAGNKPRWVRMREAGEPGWEE EKAKAEARKSTDRTADVLRALADFLKPLMRVYTDSDMSSVQWKPLRKGQAVR TWDRDMFQQAIERMMSWESWNQRVGEAYAKLVEQKSRFEQKNFVQGEHLVQL VNQLQQDMKEASHGLESKEQTAHYLTGRALRGSDKVFKEKWEKLPDAPFDLYDT EIKNVQRRNTRRFSGHDLFAKLAEPKYQALWREDASFLTRYAVYNSIVRKLNHAK MFATFTLPDATAHPIWTRFDKLGGNLHQYTFLFNEFGEGRHAIRFQKLLTVEDGV AKEVDDVTVPISMSAQLDDLLPRDPHELVALYFQDYGAEQHLAGEFGGAKIQYRR DQLNHLHARRGARDVYLNLSVRVQSQSEARGERRPPYAAVFRVLDGNHRAFVHF DKLSDYLAEHPDDGKLGSEGLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNSE GRVPPFCPIEGNENLVAVHERSOLLKLPGETESKDLRAIREERQRTLRLQRTQLAYL RLLVRCGSEDVGRRERSWAKLIEQPMQANQMPDWREAFEDLQKLSLYGICG DREWTEAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYQKDVVGGNSIEQ IRYLERQYKFLKSWFFGKVSQVIRAEKGSRAITLREHIDHAKEDRLKKLADRII MEALGYVYALDDERGGKGVVAKYPPCQLILLEELSEYQFNDRPPSENNQLMQW SHRGVVFQELNQAQVHDLVGTMYAAFSSRFDARTGAPGIRCRVPARCAREQNP EFPFWLNKFVAEHKLDGCPLRADDLIPTGEGEFFVSPFSAEEGDFHQIHADLNAA QNLQRRLWSDFDISQIRLRCDWGEVDGEPVLIPTTGKRTADSYGNKVFYTKTGV

[0430]

	TYYERERGGKRRKVFQAQEELSEEEAELLVEADEAREKSVVLMRDPGIIINRGDWT RQKEFWSMVNQRIEGLYLVKQIRSRVRLQESACENTGDI
SEQ ID NO: 11 AaCas12b-Q854R	MAVKSMKVKLRLDNMPEIRAGLWKLHTEVNAGVRYYTEWLSLLRQENLYRRSP NGDGEQECYKTAEECKAELLERLRARQVENGHCGPAGSDDELLQLARQLYELLV PQAIGAKGDAQIARKFLSPLADKDAVGGGLGIAKAGNKPRWVRMREAGEPGWEE EKAKAEARKSTDRTADVLRALADFLGKPLMRVYTDSDMSSVQWKPLRKGQAVR TWDRDMFQQAIERMMSWESWNQRVGEAYAKLVEQKSRFEQKNFVQGEHLVQL VNQLQQDMKEASHGLESKEQTAHYLTGRALRGSDKVFEKWEKLPDAPFDLYDT EIKNVQRRNTRRRFGSHDLFAKLAEPKYQALWREDASFLTRYAVYNSIVRKLNHAK MFATFTLPDATAHPIWTRFDKLGGNLHQYTFLNFEFGEGRHAIRFQKLLTVEDGV AKEVDDVTPISMSAQLDDLLPRDPHELVALYFQDYGAEQHLAGEFGGAKIQYRR DQLNHLHARRGARDVYLNLSVRVQSQSEARGERRPPYAAVFRLVGDNHRAVHF DKLSDYLAEHPDDGKLGSEGLSGLRVMSVDLGLRTSASISVFRVARKDELKPNSE GRVPFCFPIEGNENLVAVHERSQLLKLPGETESKDLRAIREERQRTLRLQRTQLAYL RLLVRCGSEDVGRRERSWAKLIEQPM DANQMTDPWREAFEDLQKLSLYGICG DREWTEAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYQKDVVGGNSIEQ IEYLERQYKFLKSWFFGK VSGQVIRAEKGSRFAITLREHIDHAKEDRLKKLADRII MEALGYVYALDDERGGKGVVAKYPPCQLILLEELSEYRFNNDRPPSENNQLMQW SHRGVQFELLNQAQVHDLVGTMYAAFSSRFDARTGAPGIRCRVPARCAREQNP EPPFWWLNKFVAEHKLDGCPRLADDLIPTGEGEFFVSPFSAEEGDFHQIHADLNAA QNLQRRLLWSDFDISQIRLRCDWGEVDGEPVLIPRTTGKRTADSYGNKVFFYTKTG TYYERERGGKRRKVFQAQEELSEEEAELLVEADEAREKSVVLMRDPGIIINRGDWT RQKEFWSMVNQRIEGLYLVKQIRSRVRLQESACENTGDI
SEQ ID NO: 12 AaCas12b-D858R	MAVKSMKVKLRLDNMPEIRAGLWKLHTEVNAGVRYYTEWLSLLRQENLYRRSP NGDGEQECYKTAEECKAELLERLRARQVENGHCGPAGSDDELLQLARQLYELLV PQAIGAKGDAQIARKFLSPLADKDAVGGGLGIAKAGNKPRWVRMREAGEPGWEE EKAKAEARKSTDRTADVLRALADFLGKPLMRVYTDSDMSSVQWKPLRKGQAVR TWDRDMFQQAIERMMSWESWNQRVGEAYAKLVEQKSRFEQKNFVQGEHLVQL VNQLQQDMKEASHGLESKEQTAHYLTGRALRGSDKVFEKWEKLPDAPFDLYDT EIKNVQRRNTRRRFGSHDLFAKLAEPKYQALWREDASFLTRYAVYNSIVRKLNHAK MFATFTLPDATAHPIWTRFDKLGGNLHQYTFLNFEFGEGRHAIRFQKLLTVEDGV AKEVDDVTPISMSAQLDDLLPRDPHELVALYFQDYGAEQHLAGEFGGAKIQYRR DQLNHLHARRGARDVYLNLSVRVQSQSEARGERRPPYAAVFRLVGDNHRAVHF DKLSDYLAEHPDDGKLGSEGLSGLRVMSVDLGLRTSASISVFRVARKDELKPNSE GRVPFCFPIEGNENLVAVHERSQLLKLPGETESKDLRAIREERQRTLRLQRTQLAYL RLLVRCGSEDVGRRERSWAKLIEQPM DANQMTDPWREAFEDLQKLSLYGICG DREWTEAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYQKDVVGGNSIEQ IEYLERQYKFLKSWFFGK VSGQVIRAEKGSRFAITLREHIDHAKEDRLKKLADRII MEALGYVYALDDERGGKGVVAKYPPCQLILLEELSEYQFNRRPPSENNQLMQW SHRGVQFELLNQAQVHDLVGTMYAAFSSRFDARTGAPGIRCRVPARCAREQNP EPPFWWLNKFVAEHKLDGCPRLADDLIPTGEGEFFVSPFSAEEGDFHQIHADLNAA QNLQRRLLWSDFDISQIRLRCDWGEVDGEPVLIPRTTGKRTADSYGNKVFFYTKTG TYYERERGGKRRKVFQAQEELSEEEAELLVEADEAREKSVVLMRDPGIIINRGDWT RQKEFWSMVNQRIEGLYLVKQIRSRVRLQESACENTGDI
SEQ ID NO: 13 AaCas12b-N857K	MAVKSMKVKLRLDNMPEIRAGLWKLHTEVNAGVRYYTEWLSLLRQENLYRRSP NGDGEQECYKTAEECKAELLERLRARQVENGHCGPAGSDDELLQLARQLYELLV PQAIGAKGDAQIARKFLSPLADKDAVGGGLGIAKAGNKPRWVRMREAGEPGWEE EKAKAEARKSTDRTADVLRALADFLGKPLMRVYTDSDMSSVQWKPLRKGQAVR TWDRDMFQQAIERMMSWESWNQRVGEAYAKLVEQKSRFEQKNFVQGEHLVQL VNQLQQDMKEASHGLESKEQTAHYLTGRALRGSDKVFEKWEKLPDAPFDLYDT EIKNVQRRNTRRRFGSHDLFAKLAEPKYQALWREDASFLTRYAVYNSIVRKLNHAK MFATFTLPDATAHPIWTRFDKLGGNLHQYTFLNFEFGEGRHAIRFQKLLTVEDGV AKEVDDVTPISMSAQLDDLLPRDPHELVALYFQDYGAEQHLAGEFGGAKIQYRR DQLNHLHARRGARDVYLNLSVRVQSQSEARGERRPPYAAVFRLVGDNHRAVHF DKLSDYLAEHPDDGKLGSEGLSGLRVMSVDLGLRTSASISVFRVARKDELKPNSE GRVPFCFPIEGNENLVAVHERSQLLKLPGETESKDLRAIREERQRTLRLQRTQLAYL RLLVRCGSEDVGRRERSWAKLIEQPM DANQMTDPWREAFEDLQKLSLYGICG DREWTEAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYQKDVVGGNSIEQ IEYLERQYKFLKSWFFGK VSGQVIRAEKGSRFAITLREHIDHAKEDRLKKLADRII

[0431]

	MEALGYVYALDDERGGKWKVAKYPPCQLILLEELSEYQFNKDRPPSENNQLMQW SHRGVVFQELLNQAQVHDLLVGTMYAAFSSRFDARTGAPGIRCRRVPARCAREQNP EPFPWWLNKFVAEHKLDGCPLRADDLIPTGEGEFFVSPFSAEEGDFHQIHADLNAA QNLQRRLLWSDFDISQIRLRCDWGEVDGEPVLIPTTGKRTADSYGNKVFYTKTGV TYYERERGGKRRKVFQAQEEELSEEEAELLVEADEAREKSVVLMRDPGIIINRGDWT RQKEFWSMVNQRIEGYLVKQIRSRVRLQESACENTGDI
SEQ ID NO: 14 AaCas12b-N865W	MAVKSMKVKLRLDNMPEIRAGLWKLHTEVNAAGVRYYTEWLSLLRQENLYRRSP NGDGEQECYKTAEECKAELLERLRARQVENGHCGPAGSDELLQLARQLYELLV PQAIGAKGDAQIARKFLSPLADKDAVGGGLGIAKAGNKPRWVRMREAGEPGWEE EKAKAEARKSTDRTADVLRALADFLKPLMRVYTDSDMSSVQWKPLRKGQAVR TWDRDMFQQAIERMMSWESWNQRVGEAYAKLVEQKSRFEQKNFVQGEHLVQL VNQLQQDMKEASHGLESKEQTAHYLTGRALRGSDKVFKEKWEKLDPAFPDLYDT EIKNVQRRNTRRFGSHDLFAKLAEPKYQALWREDASFLTRYAVYNSIVRKLNHAK MFAFTLPPDATAHPIWTRFDKLGGNLHQYTFLNEFGEGRHAIRFQKLLTVEDGV AKEVDDVTVPISMSAQLDDLLPRDPHELVALYFQDYGAEQHLAGEFGGAKIQYRR DQLNHLHARRGARDVYLNLSVRVQSQSEARGERRPPYAAVFRLVGDNRHAFVHF DKLSDYLAEHPDDGKLGSEGLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNSE GRVPFCFPIEGNENLVAVHERSQQLLKLPGETESKDLRAIREERQRTLRLQRLAYL RLLVRCGSEDEVGRRERSWAKLIEQPMANQMTDPWREAFEDQLKLSLYGICG DREWTEAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYQKDVVGGNSIEQ IEYLERQYKFLKSWFFGKVSQVIRAEKGSRFAITLREHIDHAKEDRLKKLADRII MEALGYVYALDDERGGKWKVAKYPPCQLILLEELSEYQFNDRPPSENWQLMQ WSHRGVVFQELLNQAQVHDLLVGTMYAAFSSRFDARTGAPGIRCRRVPARCAREQ NPEFPWWLNKFVAEHKLDGCPLRADDLIPTGEGEFFVSPFSAEEGDFHQIHADLN AAQNLQRRLLWSDFDISQIRLRCDWGEVDGEPVLIPTTGKRTADSYGNKVFYTKT GVTYYERERGGKRRKVFQAQEEELSEEEAELLVEADEAREKSVVLMRDPGIIINRGD WTRQKEFWSMVNQRIEGYLVKQIRSRVRLQESACENTGDI
SEQ ID NO: 15 AaCas12b-N865Y	MAVKSMKVKLRLDNMPEIRAGLWKLHTEVNAAGVRYYTEWLSLLRQENLYRRSP NGDGEQECYKTAEECKAELLERLRARQVENGHCGPAGSDELLQLARQLYELLV PQAIGAKGDAQIARKFLSPLADKDAVGGGLGIAKAGNKPRWVRMREAGEPGWEE EKAKAEARKSTDRTADVLRALADFLKPLMRVYTDSDMSSVQWKPLRKGQAVR TWDRDMFQQAIERMMSWESWNQRVGEAYAKLVEQKSRFEQKNFVQGEHLVQL VNQLQQDMKEASHGLESKEQTAHYLTGRALRGSDKVFKEKWEKLDPAFPDLYDT EIKNVQRRNTRRFGSHDLFAKLAEPKYQALWREDASFLTRYAVYNSIVRKLNHAK MFAFTLPPDATAHPIWTRFDKLGGNLHQYTFLNEFGEGRHAIRFQKLLTVEDGV AKEVDDVTVPISMSAQLDDLLPRDPHELVALYFQDYGAEQHLAGEFGGAKIQYRR DQLNHLHARRGARDVYLNLSVRVQSQSEARGERRPPYAAVFRLVGDNRHAFVHF DKLSDYLAEHPDDGKLGSEGLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNSE GRVPFCFPIEGNENLVAVHERSQQLLKLPGETESKDLRAIREERQRTLRLQRLAYL RLLVRCGSEDEVGRRERSWAKLIEQPMANQMTDPWREAFEDQLKLSLYGICG DREWTEAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYQKDVVGGNSIEQ IEYLERQYKFLKSWFFGKVSQVIRAEKGSRFAITLREHIDHAKEDRLKKLADRII MEALGYVYALDDERGGKWKVAKYPPCQLILLEELSEYQFNDRPPSENYQLMQW SHRGVVFQELLNQAQVHDLLVGTMYAAFSSRFDARTGAPGIRCRRVPARCAREQNP EPFPWWLNKFVAEHKLDGCPLRADDLIPTGEGEFFVSPFSAEEGDFHQIHADLNAA QNLQRRLLWSDFDISQIRLRCDWGEVDGEPVLIPTTGKRTADSYGNKVFYTKTGV TYYERERGGKRRKVFQAQEEELSEEEAELLVEADEAREKSVVLMRDPGIIINRGDWT RQKEFWSMVNQRIEGYLVKQIRSRVRLQESACENTGDI
SEQ ID NO: 16 AaCas12b-Q866M	MAVKSMKVKLRLDNMPEIRAGLWKLHTEVNAAGVRYYTEWLSLLRQENLYRRSP NGDGEQECYKTAEECKAELLERLRARQVENGHCGPAGSDELLQLARQLYELLV PQAIGAKGDAQIARKFLSPLADKDAVGGGLGIAKAGNKPRWVRMREAGEPGWEE EKAKAEARKSTDRTADVLRALADFLKPLMRVYTDSDMSSVQWKPLRKGQAVR TWDRDMFQQAIERMMSWESWNQRVGEAYAKLVEQKSRFEQKNFVQGEHLVQL VNQLQQDMKEASHGLESKEQTAHYLTGRALRGSDKVFKEKWEKLDPAFPDLYDT EIKNVQRRNTRRFGSHDLFAKLAEPKYQALWREDASFLTRYAVYNSIVRKLNHAK MFAFTLPPDATAHPIWTRFDKLGGNLHQYTFLNEFGEGRHAIRFQKLLTVEDGV AKEVDDVTVPISMSAQLDDLLPRDPHELVALYFQDYGAEQHLAGEFGGAKIQYRR DQLNHLHARRGARDVYLNLSVRVQSQSEARGERRPPYAAVFRLVGDNRHAFVHF DKLSDYLAEHPDDGKLGSEGLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNSE

[0432]

	GRVPCFPIEGNENLVAVHERSQQLLKLPGETESKDLRAIREERQRTLRLQRTQLAYL RLLVRCGSEDVGRRERSWAKLIEQPM DANQMPDWREAFEDLQKLSLYGICG DREWTEAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYQKDVVGGNSIEQ IEYLERQYKFLKSWFFGKVSQVIRAEKGSRFITLREHIDHAKEDRLKKLADRII MEALGYVYALDDERGGKGVVAKYPPCQLILLEELSEYQFNDRPPSENNMLMQ WSHRGVFQELLNQAQVHDLVGTMYAAFSSRFDARTGAPGIRCRVPARCAREQ NPEFPWWLNKFVAEHKLDGCLRADDLIPTGEGEFFVSPFSAEEGDFHQIHADLN AAQNLQRRLWSDFDISQIRLRCDWGEVDGEPVLIPRTTGKRTADSYGNKVFYTKT GVTYYERERGGKRRRVFAQEELSEEEAELLVEADEAREKSVVLMRDPGSIINRGD WTRQKEFWSMVNQRIEGLVKQIRSRVRLQESACENTGDI
SEQ ID NO: 17 AaCas12b-Q869M	MAVKSMKVKLRLDNMPEIRAGLWKLHTEVNAGVRYYTEWLSLLRQENLYRRSP NGDGEQECYKTAEECKAELLERLRARQVENGHCGPAGSDDPELLQLARQLYELLV PQAIGAKGDAQIARKFLSPLADKDAVGGLGIAKAGNKPRWVRMREAGEPGWEE EKAKAEARKSTDRTADVLRALADFLKPLMRVYTDSDMSSVQWKPLRKGQAVR TWDRDMFQQAIERMMSWESWNQRVGEAYAKLVEQKSRFEQKNFVQGEHLVQL VNQLQQDMKEASHGLESKEQTAHYLTGRALRGSDKVFEKWEKLPDAPFDLYDT EIKNVQRRNTRRFSGHDLFAKLAEPKYQALWREDASFLTRYAVYNSIVRKLNHAK MFATFTLPDATAHPIWTRFDKLGGNLHQYTFLNEFGEGRHAIRFQKLLTVEDGV AKEVDDVTVPISMSAQLDDLLPRDPHELVALYFQDYGAEQHLAGEFGGAKIQYRR DQLNHLHARRGARDVYLNLSVRVQSQSEARGERRPPYAAVFLRVGDNHRAFVHF DKLSDYLAEHPDDGKLGSEGLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNSE GRVPCFPIEGNENLVAVHERSQQLLKLPGETESKDLRAIREERQRTLRLQRTQLAYL RLLVRCGSEDVGRRERSWAKLIEQPM DANQMPDWREAFEDLQKLSLYGICG DREWTEAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYQKDVVGGNSIEQ IEYLERQYKFLKSWFFGKVSQVIRAEKGSRFITLREHIDHAKEDRLKKLADRII MEALGYVYALDDERGGKGVVAKYPPCQLILLEELSEYQFNDRPPSENNMLMM WSHRGVFQELLNQAQVHDLVGTMYAAFSSRFDARTGAPGIRCRVPARCAREQ NPEFPWWLNKFVAEHKLDGCLRADDLIPTGEGEFFVSPFSAEEGDFHQIHADLN AAQNLQRRLWSDFDISQIRLRCDWGEVDGEPVLIPRTTGKRTADSYGNKVFYTKT GVTYYERERGGKRRRVFAQEELSEEEAELLVEADEAREKSVVLMRDPGSIINRGD WTRQKEFWSMVNQRIEGLVKQIRSRVRLQESACENTGDI
SEQ ID NO: 18 AaCas12b-Q1093 W	MAVKSMKVKLRLDNMPEIRAGLWKLHTEVNAGVRYYTEWLSLLRQENLYRRSP NGDGEQECYKTAEECKAELLERLRARQVENGHCGPAGSDDPELLQLARQLYELLV PQAIGAKGDAQIARKFLSPLADKDAVGGLGIAKAGNKPRWVRMREAGEPGWEE EKAKAEARKSTDRTADVLRALADFLKPLMRVYTDSDMSSVQWKPLRKGQAVR TWDRDMFQQAIERMMSWESWNQRVGEAYAKLVEQKSRFEQKNFVQGEHLVQL VNQLQQDMKEASHGLESKEQTAHYLTGRALRGSDKVFEKWEKLPDAPFDLYDT EIKNVQRRNTRRFSGHDLFAKLAEPKYQALWREDASFLTRYAVYNSIVRKLNHAK MFATFTLPDATAHPIWTRFDKLGGNLHQYTFLNEFGEGRHAIRFQKLLTVEDGV AKEVDDVTVPISMSAQLDDLLPRDPHELVALYFQDYGAEQHLAGEFGGAKIQYRR DQLNHLHARRGARDVYLNLSVRVQSQSEARGERRPPYAAVFLRVGDNHRAFVHF DKLSDYLAEHPDDGKLGSEGLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNSE GRVPCFPIEGNENLVAVHERSQQLLKLPGETESKDLRAIREERQRTLRLQRTQLAYL RLLVRCGSEDVGRRERSWAKLIEQPM DANQMPDWREAFEDLQKLSLYGICG DREWTEAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYQKDVVGGNSIEQ IEYLERQYKFLKSWFFGKVSQVIRAEKGSRFITLREHIDHAKEDRLKKLADRII MEALGYVYALDDERGGKGVVAKYPPCQLILLEELSEYQFNDRPPSENNMLMQ WSHRGVFQELLNQAQVHDLVGTMYAAFSSRFDARTGAPGIRCRVPARCAREQNP EFPWWLNKFVAEHKLDGCLRADDLIPTGEGEFFVSPFSAEEGDFHQIHADLNAA QNLQRRLWSDFDISQIRLRCDWGEVDGEPVLIPRTTGKRTADSYGNKVFYTKTGV TYERERGGKRRRVFAQEELSEEEAELLVEADEAREKSVVLMRDPGSIINRGDWT RWKEFWSMVNQRIEGLVKQIRSRVRLQESACENTGDI
SEQ ID NO: 19 AaCas12b-Q1093Y	MAVKSMKVKLRLDNMPEIRAGLWKLHTEVNAGVRYYTEWLSLLRQENLYRRSP NGDGEQECYKTAEECKAELLERLRARQVENGHCGPAGSDDPELLQLARQLYELLV PQAIGAKGDAQIARKFLSPLADKDAVGGLGIAKAGNKPRWVRMREAGEPGWEE EKAKAEARKSTDRTADVLRALADFLKPLMRVYTDSDMSSVQWKPLRKGQAVR TWDRDMFQQAIERMMSWESWNQRVGEAYAKLVEQKSRFEQKNFVQGEHLVQL VNQLQQDMKEASHGLESKEQTAHYLTGRALRGSDKVFEKWEKLPDAPFDLYDT EIKNVQRRNTRRFSGHDLFAKLAEPKYQALWREDASFLTRYAVYNSIVRKLNHAK

[0433]

	MFATFLPDATAHPIWTRFDKLGGNLHQYTFLNFEFGEGRHAIRFQKLLTVEDGV AKEVDDVTVPISMSAQLDDLLPRDPHELVALYFQDYGAEQHLAGEFGGAKIQYRR DQLNHLHARRGARDVYLNLSVRVQSQSEARGERRPPYAAVFRLVGDNRHAFVHF DKLSDYLAEHPDDGKLGSEGLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNSE GRVPCFPIEGNENLVAVHERSQLLKLPGETESKDLRAIREERQRTLRLQLAYL RLLVRCGSEDVGRRRERSWAKLIEQPM DANQMPDWREAFEDQLKLSLYGICG DREWTEAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYQKDVVGGNSIEQ IEYLERQYKFLKSWSFFGKVSQVIRAEKGSRFAITLREHIDHAKEDRLKKLADRII MEALGYVYALDDERGGKGVVAKYPPCQLILLEELSEYQFNDRPPSENNQLMQW SHRGVVFQELLNQAQVHDLVGTMYAAFSSRFDARTGAPGIRCRRV PARCAREQNP EPFPWWLNKFVAEHKLDGCPLRADDLIPTGEGEFFVSPFSAEEGDFHQIHADLNAA QNLQRRLWSDFDISQIRLRCDWGEVDGEPVLIPRTTGKRTADSYGNKV FYTKTG TYERERGGKRRKVF AQEELSEEEAELLVEADEAREKSVVLMRDPGSIINRGDWT RYKEFWSMVNQRIEYLVKQIRSRVRLQESACENTGDI
SEQ ID NO: 20 Q866M+Q869M	MAVKSMKVKLRLDNMPEIRAGLWKLHTEVNA GVRYYTEWLSLLRQENLYRRSP NGDGEQECYKTAEECKAELLERLRARQVENGHCGPAGSDDELLQLARQLYELLV PQAIGAKGDAQIARKFLSPLADKDAVGGGLGIAKAGNKPRWVRMREAGEPGWEE EKAKAEARKSTDR TADVLRALADFLKPLMRVYTDSDMSSVQWKPLRKGQAVR TWRDRMFQQAIERMMSWESWNQRVGEAYAKLVEQKSRFEQKNFVQGEHLVQL VNQLQQDMKEASHGLESKEQTAHYLTGRALRGSDKVF EKWEKLPDAPFDLYDT EIKNVQRRNTRRFGSHDLFAKLAEPKYQALWREDA SFLTRYAVYNSIVRKLNHAK MFATFLPDATAHPIWTRFDKLGGNLHQYTFLNFEFGEGRHAIRFQKLLTVEDGV AKEVDDVTVPISMSAQLDDLLPRDPHELVALYFQDYGAEQHLAGEFGGAKIQYRR DQLNHLHARRGARDVYLNLSVRVQSQSEARGERRPPYAAVFRLVGDNRHAFVHF DKLSDYLAEHPDDGKLGSEGLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNSE GRVPCFPIEGNENLVAVHERSQLLKLPGETESKDLRAIREERQRTLRLQLAYL RLLVRCGSEDVGRRRERSWAKLIEQPM DANQMPDWREAFEDQLKLSLYGICG DREWTEAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYQKDVVGGNSIEQ IEYLERQYKFLKSWSFFGKVSQVIRAEKGSRFAITLREHIDHAKEDRLKKLADRII MEALGYVYALDDERGGKGVVAKYPPCQLILLEELSEYQFNDRPPSENNMLMM WSHRGVVFQELLNQAQVHDLVGTMYAAFSSRFDARTGAPGIRCRRV PARCAREQ NPEFPWWLNKFVAEHKLDGCPLRADDLIPTGEGEFFVSPFSAEEGDFHQIHADLN AAQNLQRRLWSDFDISQIRLRCDWGEVDGEPVLIPRTTGKRTADSYGNKV FYTKT GVTYERERGGKRRKVF AQEELSEEEAELLVEADEAREKSVVLMRDPGSIINRGD WTRQKEFWSMVNQRIEYLVKQIRSRVRLQESACENTGDI
SEQ ID NO: 21 AaCas12b-Q119F+ E475R	MAVKSMKVKLRLDNMPEIRAGLWKLHTEVNA GVRYYTEWLSLLRQENLYRRSP NGDGEQECYKTAEECKAELLERLRARQVENGHCGPAGSDDELLQLARQLYELLV PQAIGAKGDAQIARKFLSPLADKDAVGGGLGIAKAGNKPRWVRMREAGEPGWEE EKAKAEARKSTDR TADVLRALADFLKPLMRVYTDSDMSSVQWKPLRKGQAVR TWRDRMFQQAIERMMSWESWNQRVGEAYAKLVEQKSRFEQKNFVQGEHLVQL VNQLQQDMKEASHGLESKEQTAHYLTGRALRGSDKVF EKWEKLPDAPFDLYDT EIKNVQRRNTRRFGSHDLFAKLAEPKYQALWREDA SFLTRYAVYNSIVRKLNHAK MFATFLPDATAHPIWTRFDKLGGNLHQYTFLNFEFGEGRHAIRFQKLLTVEDGV AKEVDDVTVPISMSAQLDDLLPRDPHELVALYFQDYGAEQHLAGEFGGAKIQYRR DQLNHLHARRGARDVYLNLSVRVQSQSEARGERRPPYAAVFRLVGDNRHAFVHF DKLSDYLAEHPDDGKLGSEGLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNSE GRVPCFPIEGNENLVAVHERSQLLKLPGETESKDLRAIREERQRTLRLQLAYL RLLVRCGSEDVGRRRERSWAKLIEQPM DANQMPDWREAFEDQLKLSLYGICG DREWTEAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYQKDVVGGNSIEQ IEYLERQYKFLKSWSFFGKVSQVIRAEKGSRFAITLREHIDHAKEDRLKKLADRII MEALGYVYALDDERGGKGVVAKYPPCQLILLEELSEYQFNDRPPSENNQLMQW SHRGVVFQELLNQAQVHDLVGTMYAAFSSRFDARTGAPGIRCRRV PARCAREQNP EPFPWWLNKFVAEHKLDGCPLRADDLIPTGEGEFFVSPFSAEEGDFHQIHADLNAA QNLQRRLWSDFDISQIRLRCDWGEVDGEPVLIPRTTGKRTADSYGNKV FYTKTG TYERERGGKRRKVF AQEELSEEEAELLVEADEAREKSVVLMRDPGSIINRGDWT RQKEFWSMVNQRIEYLVKQIRSRVRLQESACENTGDI
SEQ ID NO: 22 AaCas12b-Q119F+	MAVKSMKVKLRLDNMPEIRAGLWKLHTEVNA GVRYYTEWLSLLRQENLYRRSP NGDGEQECYKTAEECKAELLERLRARQVENGHCGPAGSDDELLQLARQLYELLV PQAIGAKGDAQIARKFLSPLADKDAVGGGLGIAKAGNKPRWVRMREAGEPGWEE

[0434]

E475R+E758R	EKAKAEARKSTDRTADVLRALADFLKPLMRVYTDSDMSSVQWKPLRKGQAVR TWDRDMFQQAIERMMSWESWNQRVGEAYAKLVEQKSRFEQKNFVQGEHLVQL VNQLQQDMKEASHGLESKEQTAHYLTGRALRGSDKVFEKWEKLPDAPFDLYDT EIKNVQRNTRRRFGSHDLFAKLAEPKYQALWREDASFLTRYAVYNSIVRKLNHAK MFATFTLPDATAHPIWTRFDKLGGNLHQYTFLFNEFGEGRHAIRFQKLLTVEDGV AKEVDDVTVPISMSAQLDDLLPRDPHELVALYFQDYGAEQHLAAGRFGGAKIQYRR DQLNHLHARRGARDVYLNLSVRVQSQSEARGERRPPYAAVFRLVGDNRHAFVHF DKLSDYLAEHPDDGKLGSEGLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNSE GRVPPFCPIEGNENLVAVHERSQLLKLPGETESKDLRAIREERQRTLRLQRTQLAYL RLLVRCGSEDVGRRERSWAKLIEQPM DANQMPDWREAFEDLQKLSLYGICG DREWTEAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYQKDVVGGNSIRQ IEYLERQYKFLKSWFFGK VSGQVIRA EKGSRFAITLREHIDHAKEDRLKKLADRII MEALGYVYALDDERGGKGVVAKYPPCQLILLEELSEYQFNDRPPSENNQLMQW SHRGVVFQELLNQAQVHDLLVGTMYAAFSSRFDARTGAPGIRCRVRPARCAREQNP EPPFPWWLNKFVAEHKLDGCPLRADDLIPTGEGEFFVSPFSAEEGDFHQIHADLNAA QNLQRRLWSDFDISQIRLRCDWGEVDGEPVLIPRTTGKRTADSYGNKVIFYTKTV TYERERGGKRRKVFQAEELSEEEAELLVEADEAREKSVVLMRDPGSIINRGDWT RQKEFWSMVNQRIEGYLKQIRSRVRLQESACENTGDI
SEQ ID NO: 79 AaCas12b-Q119F+ E475R+E758R+D5 70A	MAVKSMKVKLRLDNMPEIRAGLWKLHTEV NAGVRYYTEWLSLLRQENLYRRSP NGDGEQECYKTAEECKAELLERLRARQVENGHCGPAGSDDELLQLARQLYELLV PQAIGAKGDAQFIARKFLSPLADKDAVGGLGIAKAGNKPRVWRMREAGEPGWEE EKAKAEARKSTDRTADVLRALADFLKPLMRVYTDSDMSSVQWKPLRKGQAVR TWDRDMFQQAIERMMSWESWNQRVGEAYAKLVEQKSRFEQKNFVQGEHLVQL VNQLQQDMKEASHGLESKEQTAHYLTGRALRGSDKVFEKWEKLPDAPFDLYDT EIKNVQRNTRRRFGSHDLFAKLAEPKYQALWREDASFLTRYAVYNSIVRKLNHAK MFATFTLPDATAHPIWTRFDKLGGNLHQYTFLFNEFGEGRHAIRFQKLLTVEDGV AKEVDDVTVPISMSAQLDDLLPRDPHELVALYFQDYGAEQHLAAGRFGGAKIQYRR DQLNHLHARRGARDVYLNLSVRVQSQSEARGERRPPYAAVFRLVGDNRHAFVHF DKLSDYLAEHPDDGKLGSEGLLSGLRVMSV[A]LGLRTSASISVFRVARKDELKPNSE EGRVPPFCPIEGNENLVAVHERSQLLKLPGETESKDLRAIREERQRTLRLQRTQLAY LRLVRCGSEDVGRRERSWAKLIEQPM DANQMPDWREAFEDLQKLSLYGIC GDREWTEAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYQKDVVGGNSIR QIEYLERQYKFLKSWFFGK VSGQVIRA EKGSRFAITLREHIDHAKEDRLKKLADRI IMEALGYVYALDDERGGKGVVAKYPPCQLILLEELSEYQFNDRPPSENNQLMQ WSHRGVVFQELLNQAQVHDLLVGTMYAAFSSRFDARTGAPGIRCRVRPARCAREQ NPEPPFPWWLNKFVAEHKLDGCPLRADDLIPTGEGEFFVSPFSAEEGDFHQIHADLN AAQNLQRRLWSDFDISQIRLRCDWGEVDGEPVLIPRTTGKRTADSYGNKVIFYTKT GVTYERERGGKRRKVFQAEELSEEEAELLVEADEAREKSVVLMRDPGSIINRGD WTRQKEFWSMVNQRIEGYLKQIRSRVRLQESACENTGDI
SEQ ID NO: 80 AaCas12b-Q119F+ E475R+E758R+D5 70A+E848A	MAVKSMKVKLRLDNMPEIRAGLWKLHTEV NAGVRYYTEWLSLLRQENLYRRSP NGDGEQECYKTAEECKAELLERLRARQVENGHCGPAGSDDELLQLARQLYELLV PQAIGAKGDAQFIARKFLSPLADKDAVGGLGIAKAGNKPRVWRMREAGEPGWEE EKAKAEARKSTDRTADVLRALADFLKPLMRVYTDSDMSSVQWKPLRKGQAVR TWDRDMFQQAIERMMSWESWNQRVGEAYAKLVEQKSRFEQKNFVQGEHLVQL VNQLQQDMKEASHGLESKEQTAHYLTGRALRGSDKVFEKWEKLPDAPFDLYDT EIKNVQRNTRRRFGSHDLFAKLAEPKYQALWREDASFLTRYAVYNSIVRKLNHAK MFATFTLPDATAHPIWTRFDKLGGNLHQYTFLFNEFGEGRHAIRFQKLLTVEDGV AKEVDDVTVPISMSAQLDDLLPRDPHELVALYFQDYGAEQHLAAGRFGGAKIQYRR DQLNHLHARRGARDVYLNLSVRVQSQSEARGERRPPYAAVFRLVGDNRHAFVHF DKLSDYLAEHPDDGKLGSEGLLSGLRVMSV[A]LGLRTSASISVFRVARKDELKPNSE EGRVPPFCPIEGNENLVAVHERSQLLKLPGETESKDLRAIREERQRTLRLQRTQLAY LRLVRCGSEDVGRRERSWAKLIEQPM DANQMPDWREAFEDLQKLSLYGIC GDREWTEAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYQKDVVGGNSIR QIEYLERQYKFLKSWFFGK VSGQVIRA EKGSRFAITLREHIDHAKEDRLKKLADRI IMEALGYVYALDDERGGKGVVAKYPPCQLILL[A]ELSEYQFNDRPPSENNQLMQ WSHRGVVFQELLNQAQVHDLLVGTMYAAFSSRFDARTGAPGIRCRVRPARCAREQ NPEPPFPWWLNKFVAEHKLDGCPLRADDLIPTGEGEFFVSPFSAEEGDFHQIHADLN AAQNLQRRLWSDFDISQIRLRCDWGEVDGEPVLIPRTTGKRTADSYGNKVIFYTKT GVTYERERGGKRRKVFQAEELSEEEAELLVEADEAREKSVVLMRDPGSIINRGD

[0435]

	WTRQKEFWSMVNRQIEGYLVKQIRSRVRLQESACENTGDI
SEQ ID NO: 81 AaCas12b-Q119F+E475R+E758R+D570A+D977A	MAVKSMKVKLRLDNMPEIRAGLWKLHTEVNAQVRYYTEWLSLLRQENLYRRSP NGDGEQECYKTAEECKAELLERLRARQVENGHCGPAGSDDELLQLARQLYELLV PQAIGAKGDAQFIARKFLSPLADKDAVGGGLIAKAGNKPRWVRMREAGEPGWEE EKAKAEARKSTDRTADVLRALADFLKPLMRVYTDSDMSSVQWKPLRKQAVR TWDRDMFQQAIERMMSWESWNQRVGEAYAKLVEQKSRFEQKNFVQGEHLVQL VNQLQQDMKEASHGLESKEQTAHYLTGRALRGSDKVFKEKWEKLDPAFDLYDT EIKNVQRRNTRRFGSHDLFAKLAEPKYQALWREDASFLTRYAVYNSIVRKLNHAK MFATFTLPDATAHPIWTRFDKLGGNLHQYTFLENEFGEGRHAIRFQKLLTVEDGV AKEVDDVTPISMSAQLDDLLPRDPHELVALYFQDYGAEQHLAQRFGGAKIQYRR DQLNHLHARRGARDVYLNLSVRVQSQSEARGERRPPYAAVFRVLDGNHRAVHF DKLSDYLAEHPDDGKLGSEGLLSGLRVMSV ¹ ALGLRTSASISVFRVARKDELKPN EGRVPFCPIEGNENLVAVHERSOLLKLPGETESKDLRAIREERQRTLRLQRTQLAY LRLVRCGSEDEVGRRERSWAKLIEQPMANQMPDWREAFEDLQKLSLYGIC GDREWTEAVYESVRRVVRHMGKQVRDWRKDVRSGERPKIRGYQKDVVGGNSIR QIEYLERQYKFLKSWSFFGKVSQVIRAEEKSRFAITLREHIDHAKEDRLKKLADRI IMEALGYVYALDDERGGKGVVAKYPPCQLILLEELSEYQFNDRPPSENNQLMQ WSHRGVFQELNQAQVHDLVGTMYAAAFSSRFDARTGAPGIRCRVPARCAREQ NPEFPWWLNKFVAEHKLDGCLPRADDLIPTGEGEFFVSPFSAEEGDFHQIHA ² LN AAQNLQRRLLWSDFDISQIRLRCDWGEVDGEPVLIPTTGTKRTADSYGNKVFYTKT GVTYYERERGGKRRKVFQAEELSEEEAELLVEADEAREKSVLMDRDPGSIINRGD WTRQKEFWSMVNRQIEGYLVKQIRSRVRLQESACENTGDI
SEQ ID NO: 23 AaCas12b sgRNA 支架Aa-sg	UCUAAAGGACAGAAUUUUUCAACGGGUGUGCCAAUGGCCACUUCCAGGUG GCAAAGCCCGUUGAACUUCUCAAAAAGAACGCUCGCUCAGUGUUCUGACGU CGGAUCACUGAGCGAGCGAUCUGAGAAGUGGCAC
SEQ ID NO: 24 sgRNA 支架 V0 (AaCas12b 的支 架)	GGUCUAGAGGACAGAAUUUUUCAACGGGUGUGCCAAUGGCCACUUCCAGG UGGCAAAGCCCGUUGAGCUUCUCAAAUCUGAGAAGUGGCAC
SEQ ID NO: 25 sgRNA 支架 V1	GGUCUAGAGGACAGAAUGUGCUC AACGGGUGUGCCAAUGGCCACUUCCAGG UGGCAAAGCCCGUUGAGCUUCUCAAAUCUGAGAAGUGGCAC
SEQ ID NO: 26 sgRNA 支架 V2.1	GGUCUAGAGGACAGAAUGUGCUC AACGGGUGUGCCAAUGGCCACUUCCAGG UGGCAAAGCCCGUUGAGCUUCUCAgAAUCcUGAGAAGUGGCAC
SEQ ID NO: 27 sgRNA 支架 V2.2	GGUCUAGAGGACAGAAUGUGCUC AACGGGUGUGCCAAUGGCCACUUCCAGG UGGCAAAGCCCGUUGAGCUUCUCAgAAUCcUGAGAAGUGGCAC
SEQ ID NO: 28 sgRNA 支架 V2.3	GGUCUAGAGGACAGAAUGUGCUC AACGGGUGUGCCAAUGGCCACUUCCAGG UGGCAAAGCCCGUUGAGCUUCUCAgggAAUCcccUGAGAAGUGGCAC
SEQ ID NO: 29 sgRNA 支架 V2.4	GGUCUAGAGGACAGAAUGUGCUC AACGGGUGUGCCAAUGGCCACUUCCAGG UGGCAAAGCCCGUUGAGCUUCUCAgggcAAUCgcccUGAGAAGUGGCAC
SEQ ID NO: 30 sgRNA 支架 V2.5	GGUCUAGAGGACAGAAUGUGCUC AACGGGUGUGCCAAUGGCCACUUCCAGG UGGCAAAGCCCGUUGAGCUUCUCAgggccAAUCggcccUGAGAAGUGGCAC
SEQ ID NO: 31 sgRNA 支架 V3.1	GGcUCUAGAGGAgCAGAAGUGCUC AACGGGUGUGCCAAUGGCCACUUCCAG GUGGCAAAGCCCGUUGAGCUUCUCAAAUCUGAGAAGUGGCAC
SEQ ID NO: 32 sgRNA 支架 V3.2	GGccUCUAGAGGAaggCAGAAGUGCUC AACGGGUGUGCCAAUGGCCACUUCCA GGUGGCAAAGCCCGUUGAGCUUCUCAAAUCUGAGAAGUGGCAC
SEQ ID NO: 33 sgRNA 支架 V3.3	GGcccUCUAGAGGAaggCAGAAGUGCUC AACGGGUGUGCCAAUGGCCACUUUC CAGGUGGCAAAGCCCGUUGAGCUUCUCAAAUCUGAGAAGUGGCAC
SEQ ID NO: 34 sgRNA 支架 V3.4	GGccccUCUAGAGGAcgggCAGAAGUGCUC AACGGGUGUGCCAAUGGCCACUU CCAGGUGGCAAAGCCCGUUGAGCUUCUCAAAUCUGAGAAGUGGCAC
SEQ ID NO: 35 sgRNA 支架 V3.5	GGccccUCUAGAGGAaccgggCAGAAGUGCUC AACGGGUGUGCCAAUGGCCACU UCCAGGUGGCAAAGCCCGUUGAGCUUCUCAAAUCUGAGAAGUGGCAC
SEQ ID NO: 36 sgRNA 支架 V4.1	GGcUCUAGAGGAgCAGAAGUGCUC AACGGGUGUGCCAAUGGCCACUUCCAG GUGGCAAAGCCCGUUGAGCUUCUCAgggccAAUCggcccUGAGAAGUGGCAC
SEQ ID NO: 37 sgRNA 支架 V4.2	GGccUCUAGAGGAaggCAGAAGUGCUC AACGGGUGUGCCAAUGGCCACUUCCA GGUGGCAAAGCCCGUUGAGCUUCUCAgggccAAUCggcccUGAGAAGUGGCAC
SEQ ID NO: 38	GGcccUCUAGAGGAaggCAGAAGUGCUC AACGGGUGUGCCAAUGGCCACUUUC

[0436]

sgRNA 支架 V4.3	CAGGUGGCAAAGCCCCGUUGAGCUUCUCAgggccAAUCggcccUGAGAAGUGGCAC
SEQ ID NO: 39	GGcccUCUAGAGGAgcgggCAGAAGUGCUC AACGGGUGUGCCAAUGGCCACUUU
sgRNA 支架 V4.4	CCAGGUGGCAAAGCCCCGUUGAGCUUCUCAgggccAAUCggcccUGAGAAGUGGCA C
SEQ ID NO: 40	GGcUCUAGAGGAgCAGAAGUGCUC AACCGGUGUGCCAAUGGCCACUUUCCAG
sgRNA 支架 V5.1	GUGGCAAAGCCCCGUUGAGCUUCUCAggAAUCccUGAGAAGUGGCAC
SEQ ID NO: 41	GGccUCUAGAGGAggCAGAAGUGCUC AACCGGUGUGCCAAUGGCCACUUUCCA
sgRNA 支架 V5.2	GGUGGCAAAGCCCCGUUGAGCUUCUCAggAAUCccUGAGAAGUGGCAC
SEQ ID NO: 42	GGcccUCUAGAGGAgggCAGAAGUGCUC AACCGGUGUGCCAAUGGCCACUUUC
sgRNA 支架 V5.3	CAGGUGGCAAAGCCCCGUUGAGCUUCUCAggAAUCccUGAGAAGUGGCAC
SEQ ID NO: 43	GGcccUCUAGAGGAgcgggCAGAAGUGCUC AACCGGUGUGCCAAUGGCCACUUU
sgRNA 支架 V5.4	CCAGGUGGCAAAGCCCCGUUGAGCUUCUCAggAAUCccUGAGAAGUGGCAC
SEQ ID NO: 44	GGcUCUAGAGGAgCAGAAGUGCUC AACCGGUGUGCCAAUGGCCACUUUCCAG
sgRNA 支架 V6.1	GUGGCAAAGCCCCGUUGAGCUUCUCAgggAAUCcccUGAGAAGUGGCAC
SEQ ID NO: 45	GGccUCUAGAGGAggCAGAAGUGCUC AACCGGUGUGCCAAUGGCCACUUUCCA
sgRNA 支架 V6.2	GGUGGCAAAGCCCCGUUGAGCUUCUCAgggAAUCcccUGAGAAGUGGCAC
SEQ ID NO: 46	GGcccUCUAGAGGAgggCAGAAGUGCUC AACCGGUGUGCCAAUGGCCACUUUC
sgRNA 支架 V6.3	CAGGUGGCAAAGCCCCGUUGAGCUUCUCAgggAAUCcccUGAGAAGUGGCAC
SEQ ID NO: 47	GGcccUCUAGAGGAgcgggCAGAAGUGCUC AACCGGUGUGCCAAUGGCCACUUU
sgRNA 支架 V6.4	CCAGGUGGCAAAGCCCCGUUGAGCUUCUCAgggAAUCcccUGAGAAGUGGCAC
SEQ ID NO: 48	GGcUCUAGAGGAgCAGAAGUGCUC AACCGGUGUGCCAAUGGCCACUUUCCAG
sgRNA 支架 V7.1	GUGGCAAAGCCCCGUUGAGCUUCUCAgggcAAUCgcccUGAGAAGUGGCAC
SEQ ID NO: 49	GGccUCUAGAGGAggCAGAAGUGCUC AACCGGUGUGCCAAUGGCCACUUUCCA
sgRNA 支架 V7.2	GGUGGCAAAGCCCCGUUGAGCUUCUCAgggcAAUCgcccUGAGAAGUGGCAC
SEQ ID NO: 50	GGcccUCUAGAGGAgggCAGAAGUGCUC AACCGGUGUGCCAAUGGCCACUUUC
sgRNA 支架 V7.3	CAGGUGGCAAAGCCCCGUUGAGCUUCUCAgggcAAUCgcccUGAGAAGUGGCAC
SEQ ID NO: 51	GGcccUCUAGAGGAgcgggCAGAAGUGCUC AACCGGUGUGCCAAUGGCCACUUU
sgRNA 支架 V7.4	CCAGGUGGCAAAGCCCCGUUGAGCUUCUCAgggcAAUCgcccUGAGAAGUGGCAC
SEQ ID NO: 52	GUCUAAAAGGACAGAAUUUUUCAACGGGUGUGCCAAUGGCCACUUUCCAGGU
sgRNA 支架 V8	GGCAAAGCCCCGUUGAACUUAAGCGAAGUGGCAC
SEQ ID NO: 53	GGUCGUCUAUAGGACGGCGAGUUUUUCAACGGGUGUGCCAAUGGCCACUUU
sgRNA 支架 V9	CCAGGUGGCAAAGCCCCGUUGAGCUUCAAAGAAGUGGCAC
SEQ ID NO: 54 AkCas12b	MAVKSIVKLRLESECPDILAGMWQLHRATNAGVRYYTEWVSLMRQEILYSRGPD GGQQCYMTAEDCQRELLRRLRNLHNGRQDQPGTDADLLAISRRLYEILVLQSI GKRGDAQQIASSFLSPLVDPNSKGGRGEAKSGRKPAWQKMRDQGDPRWVAARE KYEQRKAVDPSKEILNSLDALGLRPLFAVFTETYSRQVDWVPLGKSGQVVRTWDR DMFQQALERLMSWESWNRVGEYARLFQQKMKFEQHFHAEQSHLVKLARALE ADMRAASQGF EAKRGTAHQITRRALRGADRVFEIWKSIPEEALFSQYDEVIRVQV AEKRRDFGSHDLFAKLAEPKYQPLWRADETFTRYALYNGVLRDLEKARQFATFT LPDACVNPWIWTRFESSQGSNLHKEYFLFDHLGPRHAVRFQRLLVVESEGAKERDS VVVPVAPSGQLDKLVLREEEKSSVALHLHDTARPDGFMAEWAGAKLQYERSTLA RKARRDKQGMRSWRRQPSMLMSAAQMLEDAKQAGDVYLNISVRVKSPSEVRGQ RRPPY AALFRIDDKQRRVTVNYNKL SAYLEEHDPKQIPGAPGLLSGLRVMSVDLG LRTSASISVFRVAKKEEVEALGDGRPPHYPIHGTDLDLAVHERSHLIQMPGETET KQLRKLREERQAVLRPLFAQLALLRLLVRCGAADERIRTRSWQRLTKQGREFTKR LTPSWREALELELTRLEAYCGRVPDDEWSRIVDRTVIALWRRMGKQVRDWRKQV KSGAKVKVKGYQLDVVGGNSLAQIDYLEQQYKFLRRWSFFARASGLVVRADRES HFAVALRQHIEAKRDRLKKLADRILMEALGYVYEASGPREGQWTAQHPPCQLII LEELSA YRFSDDRPPSENSKLMWGHGILEELVNQAQVHDVLVGTVYAAFSSRF DARTGAPGVRCRRV PARFVGATVDDSLPLWLTEFLDKHRLDKNLLRPDDVIPTGE GEFLVSPCGEEAARVRQVHADINAAQNLQRRLWQNF DITELRLRCDVKMGGEGT VLVPRVNNARAKQLFGKKVLVSQDGVTFPERSQTGGKPHSEKQTDLTDKELELIA EAD EARA KSVVLF RDP SGHIGKGHWIRQREFWSLVKQRIESHTAERIRVRGVGSSLD
SEQ ID NO: 55 AmCas12b	MNVAVKSIKVKLMLGHLPEIREGLWHLHEAVNLGVRYYTEWLALLRQGNLYRR GKDGAQECYMTAEQCRQELLVRLRDRQKRNGHTGDPGTDEELLVARRLYELLV

[0437]

	<p>PQSVGKKGQAQMLASGFLSPLADPKSEGGKGTSKSGRKPAWMGMKEAGDSRWV EAKARYEANKAKDPTKQVIASLEMYGLRPLFDVFTETKYKTIKIRWMLPGKHQGVRA WDRDMFQQSLERLMSWESWNERVGAEFARLVDRDRFRREKHTFGQEHLVVALAQ RLEQEMKEASPGFESKSSQAHRITKRALRGADGIIDDWLKLESEGEVDRFDEILRK RQAQNPFRFGSHDLFLKLAEPVFQPLWREDPSFLSRWASYNEVLNKLEDAKQFAT FTLSPSCSNPVWARFENAEGTNIFKYDFLFDHFGKGRHGVRVFRQRMIVMRDGVPT VEGIVVPIAPSRQLDALAPNDAASPIDVFGDPAAPGAFRQFGGAKIQYRRSALV RKGRREEKAYLCGFRLPSQRRGTGPADDAGEVFLNLSLREVSEQSEQAGRRNPPYA AVFHISDQTRRVIVRYGEIERYLAEHPDTGIPGSRGLTSGLRVMSVDLGLRTSAAIS VFRVAHRDELTPDAHGRQPFPPHGMMDHLVALHERSHLIRLPGETESKKVRSIREQ RLDRLNRLRSQMASLRLVVRTGVLDQKRDRNWERLQSSMERGGGERMPDWWDD LFQAQVRYLAQHRDASGEAWGRMVQAAVRTLWRQLAKQVRDWRKEVRRNAD KVKIRGIARDVPGGHSQAQLDYLERQYRFLRSWSAFSVQAGQVVRACDLSRFAVA LREHIDNGKKDRLKKLADRILMEALGYVYVTDGRRAGQWQVAPPCQLVLEEL SEYRFSNDRPPSENSQLMVWSHRGVLEELIHQAQVHDVLTGTPAAAFSSRFDRATG APGIRCRRVPSIPLKDAPSIPIWLSHYLKQTERDAAALRPGELIPTGDGEFLVTPAGR GASGVRVHADINAHNLRRLWENFDLSDIRVRCRRREGKDGTVVLIPRLTNQR VKERYSGVIFTSEGDVSVFTVGDATRRLSSASQGEGLDLSDEEQELLAADDARE RSVVLFDRDPSGFVNGGRWTAQRAFVGMVHNRIETLLAERFSVSGAAEKVRG</p>
<p>SEQ ID NO: 56 Bs3Cas12b</p>	<p>MAIRSIKLMKMTNSGTDSIYLKALWRTHQLINEGIAYYMNLLTYRQEAIGDKTK EAYQAEILINIIRNQQRNNGSSEEHGSDQEILALLRQYELIIPSSIGESGDANQLGK FLYPLVDPNSQSGKGTSNAGRKPRWKRLKEEGNPDWELEKKKDEERKAKDPTVK IFDNLNKYGLLPLFPLFTNIQKDIEWLPLGKRQSVRKWDKDMFIQAIERLLSWESW NRRVADEYKQKKEKTESYYKEHLTGGEWIEKIRKFEKERNMELEKNAFAPNDG YFITSRQIRGWDRVYEWKSLPESASPEELWKVVAEQQNKMSGEFGDPKVFSLA NRENDRDIWRGHSERIYHIAAYNGLQKLSRTEQATFTLPAIEHPLWIRYESPGG TNLNLFKLEEKQKKNYYVTLKIIWPSEEKWIENIEIPLPSIQFNRQIKLKQHVK GKQEIFSDYSSRISLDGVLGGSRIQFNRYIKNHKELLGEGDIGPVFNLVVDVAP LQETRNGRLQSPIGKALKVISSDFSKVIDYKPKELMDWMNTGSASNSFGVASLLEG MRVMSIDMGQRTSASVSIFEVVKELPKDQEQKLFYSINDTELFaihKRSFLLNLPGE VVTKNKQQRQERRKQFVRSQIRMLANVLRLETKKTPDERKKAHKLMEIVQS YDSWTASQKEVWEKELNLLTNMAAFNDEIWKESLVELHHRIEPYVGGIVSKWRK GLSEGRKNLAGISMWNIDELEDTRLLISWKSRSRTPGEANRIETDEPFSSLLQHI QNVKDDRLKQMANLIMTALGFKYDKEEKDRYKRWKETYPACQILFENLNRVLF NLDRSRRENSRLMKWAHRSIPRTVSMQGEMFGLQVGDVRSEYSSRFHAKTGAPGI RCHALTEEDLKAGSNTLKRLEDGFINESELAYLKKGDIIPSQGGELFVTLKRYKK DSDNNELTVIHADINAQNLRKQFWQNSVYRVPCLARMGEDKLYIPKSQTET IKKYFGKGSFVKNNTQEYVYKWEKSEKMKIKTDTTDFLDQLDGFEDISKTIELAQE QKKYLTMFRDPSGYFFNETWRPQKEYWSIVNIIKSLCKKILSNKVEL</p>
<p>SEQ ID NO: 57 BsCas12b</p>	<p>MAIRSIKLLKTHTGPEAQNLKGIWRTHRLNNEGVAAYYMKMLLLFRQESTGERP KEELQEELICHIREQQQRNQADKNQALPLDKALEALRQYELLPSSVGGSGDA QIISRKFLSPLVDPNSEGKGTSKAGAKPTWQKKKEANDPTWEQDYEKWKKRRE EDPTASVITTEEYGIRPIFPLYTNTVTDIAWLPLQSNQFVRTWDRDMLQQAIERLL SWESWNRVQEEYAKLKEKMAQLNEQLEGGQEWISLLEQYEEENRERELRENMTA ANDKYRITKRQMKGWNELYELWSTFPASASHEQYKEALKRVQQRGRFGDAHF FQYLMEEKNRLIWGNPQRIHYFVARNELTKRLEEAQKQATMTLPNARKHPLWV RFDARGGNLQDYLLTAEADKPRSRRFVTFSQLIWPSESGWMEKDDVEVELALSRO FYQQVKLLKNDKQKQIEFKDKGSGSTFNHGLGGAQLQLERGDLEKEEKNFEDG EIGSVYLVNVIDFEPLQEVKNGRVQAPYQVQLLIRRPNEFPKVTTYKSEQLVEWI KASPQHSAGVESLASGFRVMSIDLGLRAAAATSIFSVEESSDKNAADFSYWIEGTP LVAVHQRSYMLRPLGQVEKQVMEKRDERFQLHQRVKFQIRVLAQIMRMANKQ YGDRWDELDSLKQAVEQKKSPLDQTDRTFWEGIVCDLTKVLPNEADWEQAVV QIHRKAEYVGVKAVQAWRKRFAADERKGIAGLSMWNIEELEGLRKLKLLISWSRTR NPQEVNRFERGHSTHQRLLTTHIQNVKEDRLKQLSHAIVMTALGYVYDERKQEW AEYPACQVILFENLSQYRSNDRSTKENSTLMKWAHRSIPKYVHMQAEPYGIQIGD VRAEYSSRFYAKTGTGIRCKKVRGQDLQGRRFENLQKRLVNEQFLTEEQVKQLR PGDIVPDDSGELFMTLTDGSGSKEVFLQADINAHNLRKQFWQRYNELFKVSCR VIVRDEEYLVPKTKSVQAKLKGFLVKKSDTAWKDVYVWDSQAKLKGKTTTFTE ESESPEQLEDFQEIIIEAEEAKGTYRTLFRDPSGVFFPESVWYPQKDFWGEVKKRL</p>

[0438]

	YGKLRERFLTKAR
SEQ ID NO: 58 LsCas12b	MSIRSFKLIKTKSGVNAEELRRGLWRTHQLINDGIAYYMNWLVLRRQEDLFINRE ETNEIEKRSKEEIQGELLERVHKQQQRNQWSGEVDDQTLTQLRHLVEEIVPSVIG KSGNASLKARFFLGPLVDPNNTTKDVS KSGPTPKWKKMKDAGDPNVVQVEYK YMAERQTLVRLLEEMGLIPLFPMYTDDEVGDHWPQASGYTRTWDRDMFQQAIE LLSWESWNRVRERRAQFEKKTDFASRFSESDVQWMNKLREYEAQQEKSLEEN AFAPNEPYALTKKALRGWERVYHWSMRLDSAASEEAYWQEVATCQTAMRGEFG DPAIYQFLAQKENHDIWRGYPERVDFAE LNHLQRELRRAKEDATFTLPDSVDHPL WVRYEAPGGTNIHG YDLVQDTRNLTLLDKFILPDENGSWHEVKKVPFSLAKSK QFHRQVWLQEEQKQKKREVVFYDYSTNPLHLGTLAGAKLQWDRNFLNKRTOQQ IEETGEIGKVVFNISVDVRPAVEVKNRGLQNGLGKALTVLTHPDGKIVTGWKA LEKWWGESGRVSSGLDLSLSEGLRVMSIDLGQRTSATVSVFEITKEAPDNPYKFF QLEGTFLFAVHQRSFLLALPGENPPQKIKQMRERWKERNRIKQQVDQLSAILRLH KKVNEDERIQAIKLLQKVASWQLNEEIIATAWNQALSQLYSKAKKENDLQWNQAI KNAHHQLEPVVGKQISLWRKDLSTGRQGIAGLSLWSIEELEATKLLTRWSKRSR EPGVVKRIERFETFAKQIQHHINQVKENRLKQLANLIVMTALGYKYDQEQKKWIE VYPACQVVLFFENLRSYRFSYERSRRENKMLMEWSHR SIPKLVQM QGELFGLQVAD VYAA YSSRYHGRTGAPGIRCHALTEADLRNETNIIHELIEAGFIKEEHRPYLQQGDL VPWSSGELFATLQKPYDNPRILTLHADINAAQNIQKRFWHPMSMWFVNCESVME GEIVTYVPKNKTVHKKQGKTRFRVVKVEGSDVYEWAKWSKNRNKNFTSSITERKPP SSMILFRDPSGTFFFKEQWVEQKTFWGWKQSMIQA YMKKTIVQRMEE
SEQ ID NO: 59 BhCas12b	MATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIEYEHHEQDPKN PKKVSKAEIQAELWDFVLKMQKCNSTHEVDKDEVFNILRELYEELVPSSVEKKG EANQLSNKFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEKKKWEED KKKDLAKILGKLAEYGLIPLFIPYTDSEPIVKEIKWMEKSRNQSVRRLDKDMFIQ ALERFLSWESWNLKVKEEYKVEKEYKTELEERIKEDIQALKALEQYERKQEQELL RDTLNTNEYRLSKRGLRGWREIIQKWLKMDENEPSEKYLEVFKDYQRKHPREAG DYSVYEFLSKKENHFIWRNHPEYPYLYATFC EIDK KKKDAKQQATFTLADPINHPL WVRFEERSGSNLNRYRILTEQLHTEKLLKKT VQLDRLIYPTESGGWEEKGKVDI VLLPSRQFYNQIFLDIEEKGKHAFTYKDESIFPLKGTGGARVQFDRDHLRRYPH KVESGNVGRIFYNMTVNIPTESPVSKSLKIHRDDFPKVVNFKPKELTEWIKDSKG KKLKS GIESLEIGLRVMSIDLGQRQAAAASIFEVVDQKPDIEGKLFPIKGTELYAV HRASFNIKLPGETLVKSREVLKAREDNLKL MNQKLNFLRNVLHFQQFEDITEREK RVTKWISRQENS DVPLVYQDEL IQI RELMYKPYKDWVAF LKQLHKRLEVEIKEV KHWRKSLSDGRKGLYGISLKNIDEIDRTRKFLLRWSLRPTEPGEVRRLEPGQRFAID QLNHLNALKEDRLK KMANTIIIMHALGYCYDVRKKKWKQAKNPACQIILFEDLSNY NPYEERSRFENSKLMKWSRREIPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPG IRCSVVTKEKLQDNRFKLNQREGRLTLDKIAVLKEGDLYPDKGGEKFISL SKDRK CVTTHADINAAQNLQKRFWTRTHGFYK VYCKAYQVDGQTVYIPESKDQKQKIIIE FGEGYFILKDG VYEWVNAAGLKKIKK GSSKQSSSELVDSILKDDSFDLASELKGKEL MLYRDPSGNVFPSPDKWMAAGVFFGKLERILISKLTNQYISISTIEDSSKQSM
SEQ ID NO: 60 SbCas12b	MSFTISYPFKLIKNKDEAKALLDTHQYMNNEGKYYLEKLLMFRQEKIFIGEDTG KRIYIEETEYKQIEEFYLIKKTELGRNLTLLDEFKTLMRELYICLVSSSMENKKG PNAQQASLNIFPLFDAESKGYILKEENNNISLIHKDYGKILLKRLRDNLIPIFTKFT DIKKITAKLSPTALDRMIFAQAIKLLSYESWCKLMIKERFDKEVKIKELNKCENK QERDKIFEILEKYEERQKTFEQDSGFAKKGKGYITGRMLKGFDEIKEKWLKEKDR SEQNLINILNKYQTDNSKLVGDRNLFEFIKLENQCLWNGDIDY LKIKRINDKNQIW LDRPEMPRFTMPDFK KHLWYRYEDPSNSNFRNYKIEVVKDENYITIPLITERNNE YFEENYTFNLAKLKKLSENITFIPKSKNKEFEFIDSND EEDKDKQKSKQYIKYCD TAKNTSYGKSGGIRLYFNRNELENYKDGKKMDSYTVFTLSIRDYKSLFAKEKLQP QIFNTVDNKITSLKIQKFGNEEQTNFLSYFTQNQITKKDWMDEKTFQNVKELNEG IRVLSVDLGRFFAAVSCFEIMSEIDNNKLFNLDQNHKIIRINDKNYYAKHIYSK TIKLSGEDDDLYKERKINKNYKLSYQERKNKIGIFTRQINKLNQLLKIIRNDEIDKEK FKELIETTKRYVKNTYNDGIIDWNNVDNKILSYENKEDVINLHKELDKKLEIDFKEF IRECRKPIFRSGGLSMQRIDFLEKLNKLRKWWARTQKSAESIVLTPKFGYK LKEHI NELKDNRVKQGVNYILMTALGYIKDNEIKNDSKKKQKEDWVKKNRACQIILMEK LTEYTFAEDRPREENSKLRMWSHRQIFN FLQKASLW GILVGDVFPYTSKCLSD NNAPGIRCHQVTKKDLIDNSWFLKIVVKDDAFCDLIEINKENVKNKSIKINDILPLR GGELFASIKDGLHIVQADINASRNIAKRFLSQINPFRVVLKDKDETFHLKNPNY

[0439]

	LKNYYSILNFVPTNEELTFFKVEENKDIKPTKRIKMDKHEKESTDEGDDYSKNQIALFRDDSGIFFDKSLWVDGKIFWSVVKNKMTKLLRERNKKNKNGSK
SEQ ID NO: 61 NLS	PKKKRKVPG
SEQ ID NO: 62 NLS	ASPKKKRKV
SEQ ID NO: 82 NLS	PKKKRKV
SEQ ID NO: 63 CCR5-11靶位点	CTCTTCTGGGCTCCCTACAA
SEQ ID NO: 64 CD34-7靶位点	AACCACTAGCACTAGCCTTG
SEQ ID NO: 65 RNF2-1靶位点	TACAGGAGGCAATAACAGAT
SEQ ID NO: 66 CCR5-3靶位点	AGGCCAAAGAATTCCTGGAA
SEQ ID NO: 67 RNF2-5靶位点	GGCACATTAATTCCTGTGT
SEQ ID NO: 68 CD34-1靶位点	CAGGTGACAGGCTAGGCTTC
SEQ ID NO: 72 Krüppel 相关盒 (KRAB) 结构域	MNNSQGRVTFEDVTNVNFTQGEWQRLNPEQRNLYRDVMLENYSNLVSVGQGETTKPDVILRLEQGKEPWLEEEVLSGRAEKNGDIGGQIWPKDVKESL

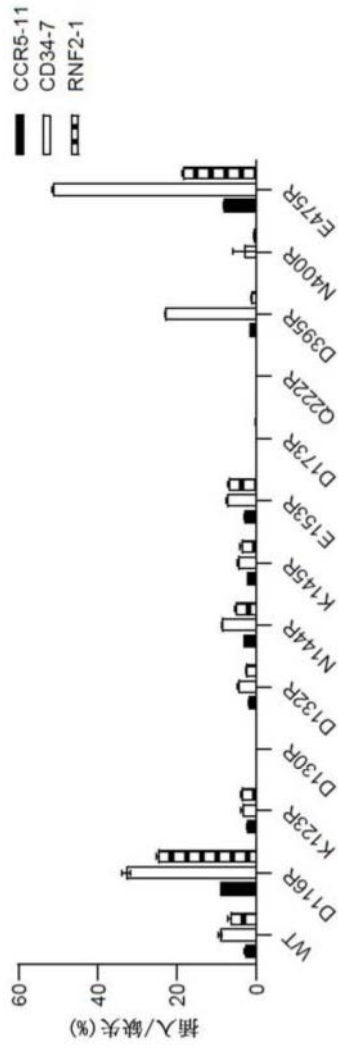


图1

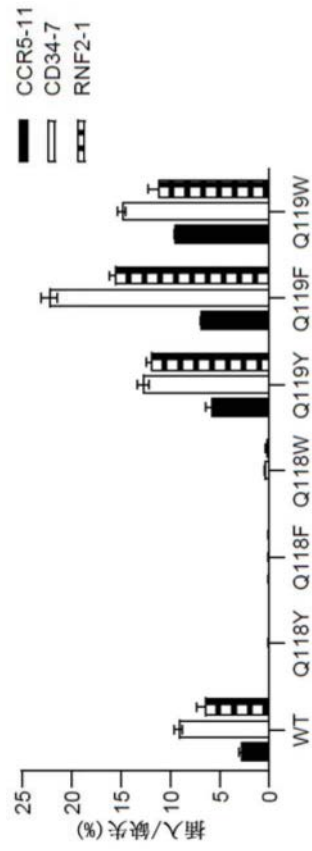


图2

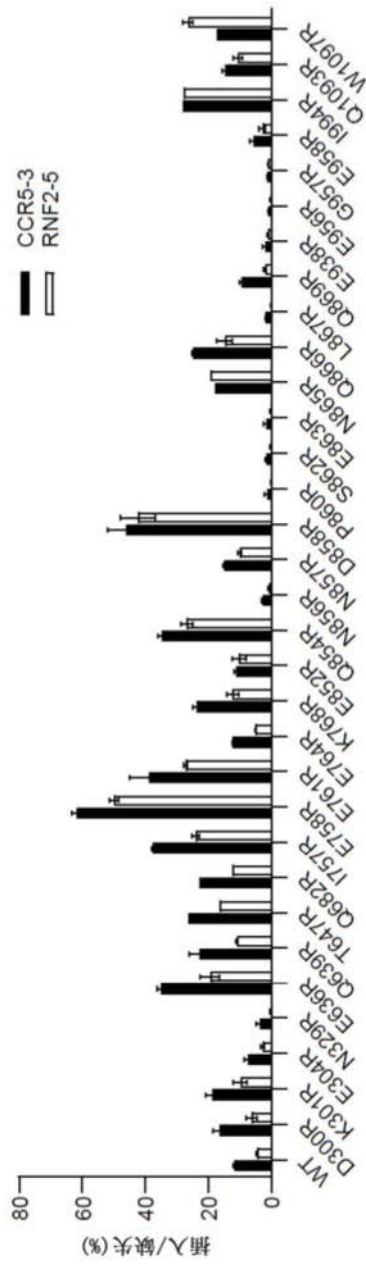


图3

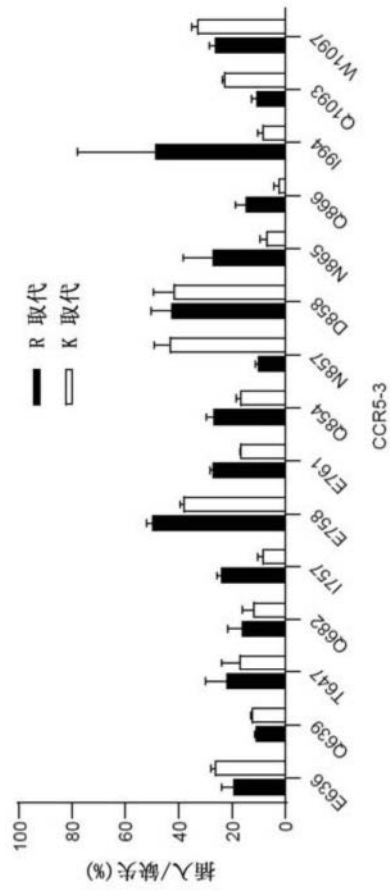


图4A

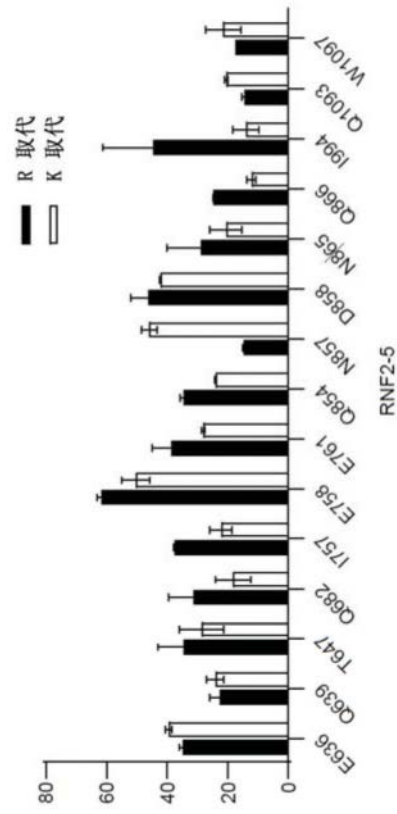


图4B

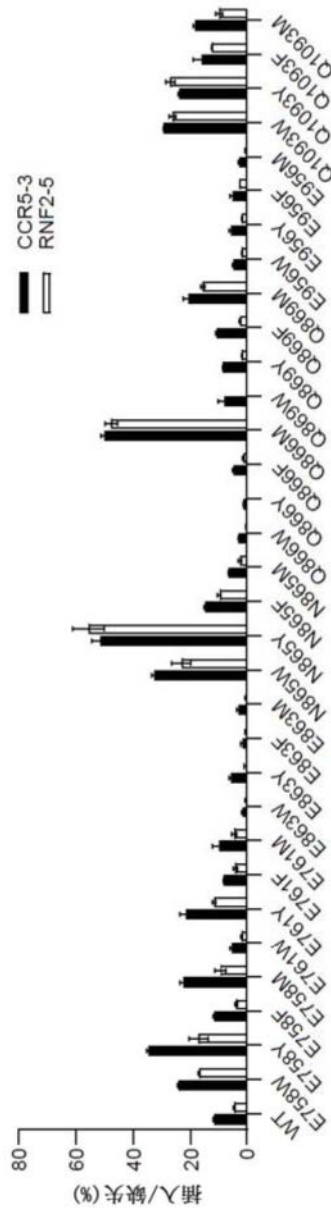


图5

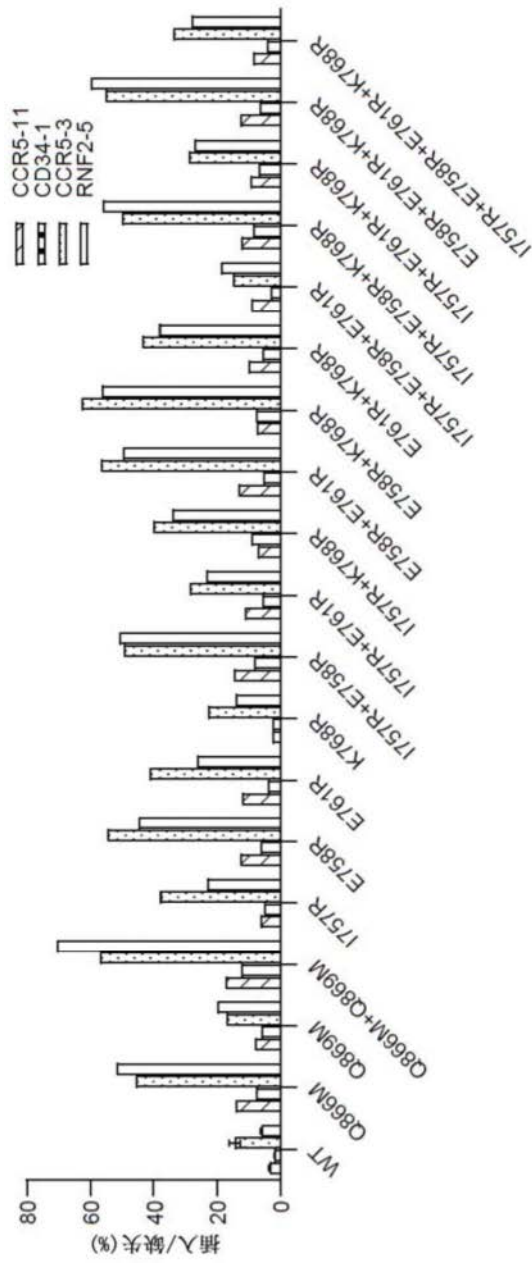


图6

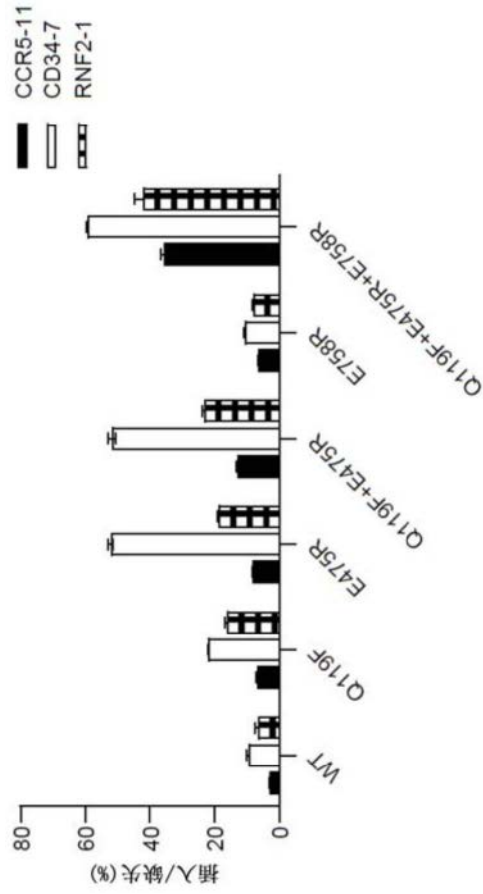


图7

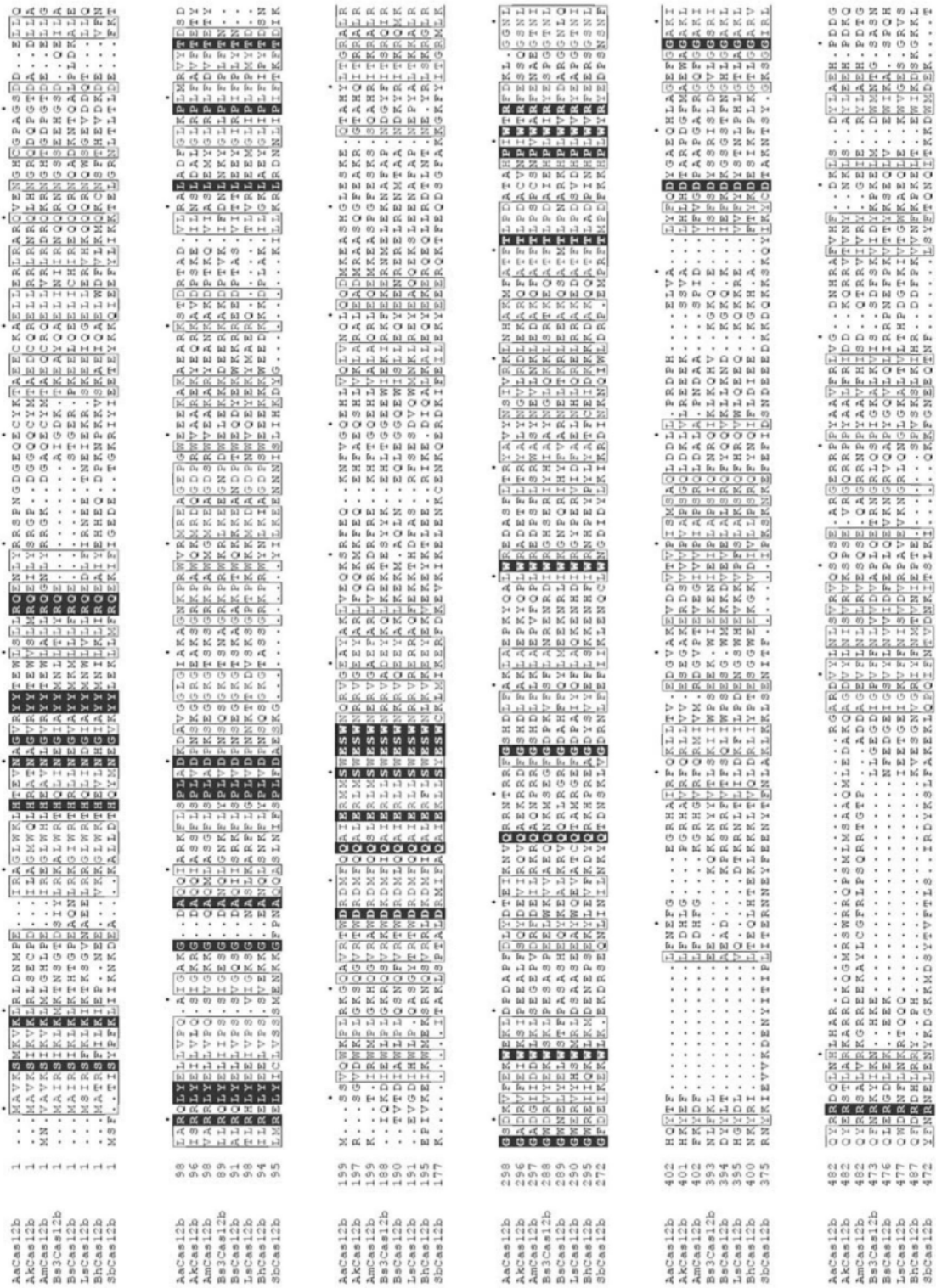


图8

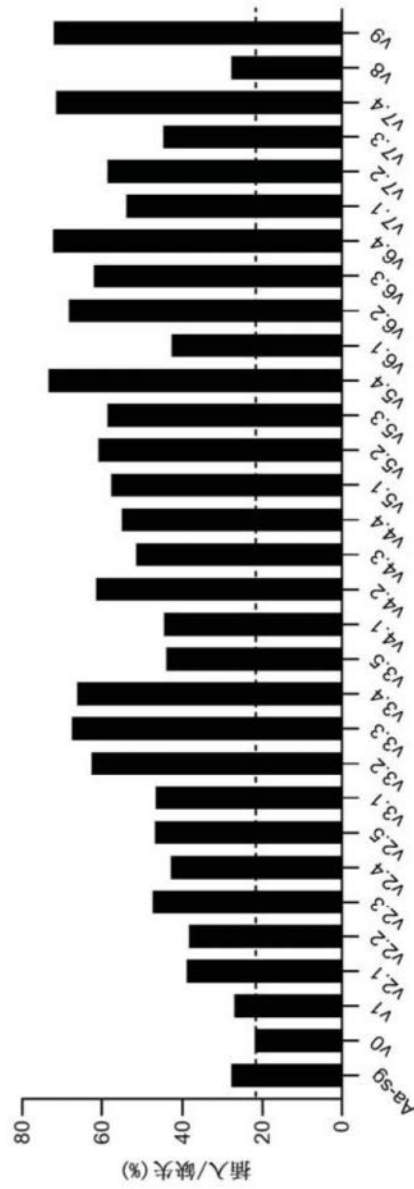


图9



图10A

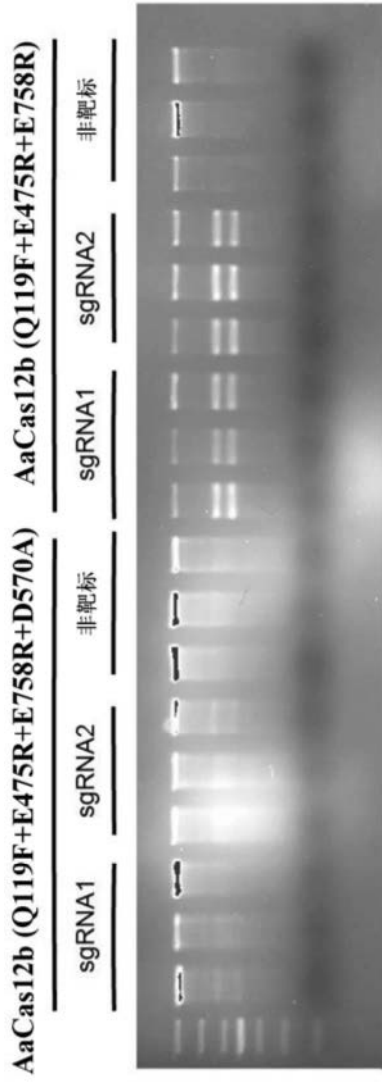


图10B

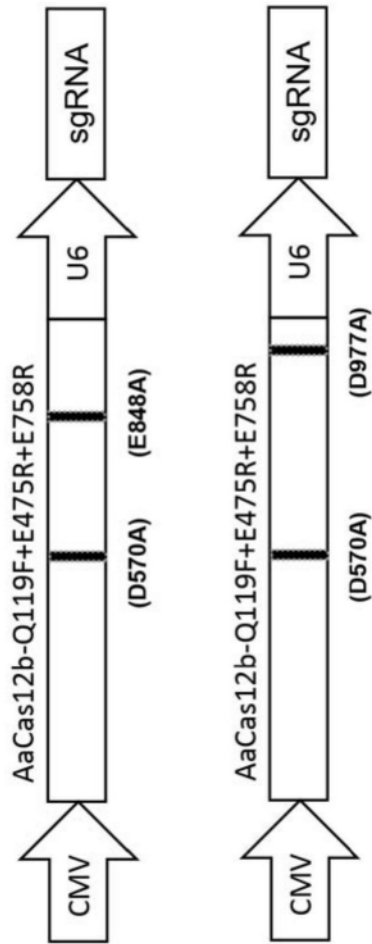


图11A



图11B

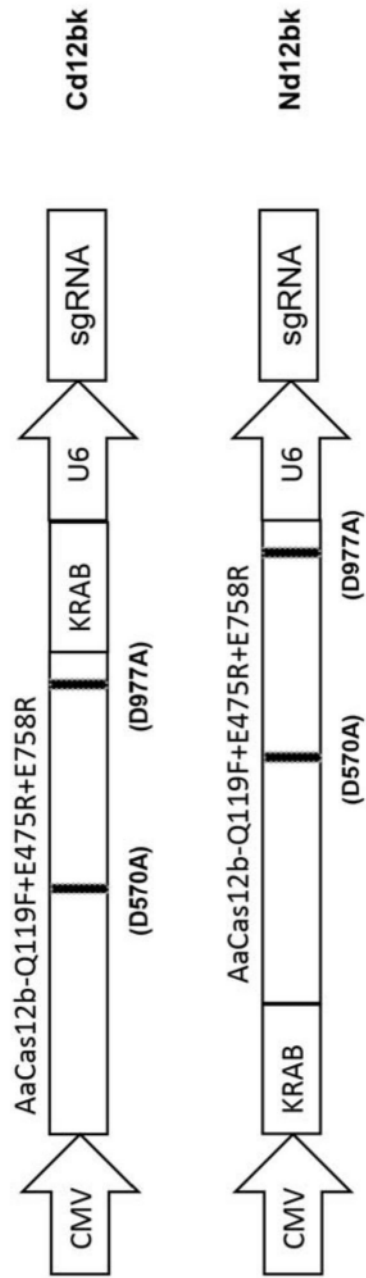


图12A

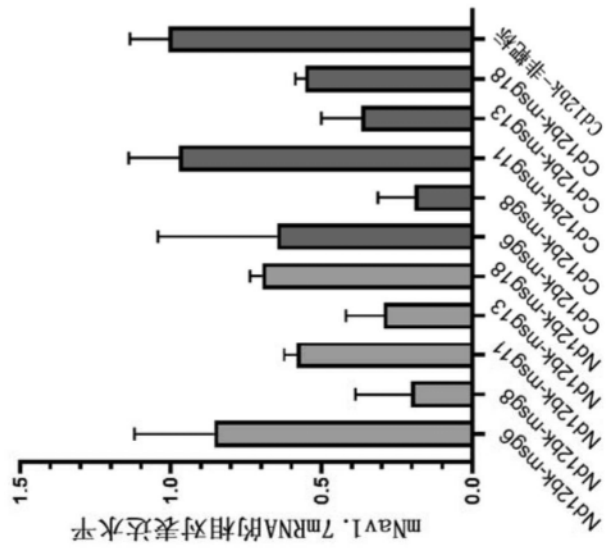


图12B