

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2022年12月1日 (01.12.2022)



(10) 国际公布号
WO 2022/247873 A1

- (51) 国际专利分类号:
C12N 9/22 (2006.01) C12Q 1/68 (2018.01)
C12N 15/113 (2010.01)
- (21) 国际申请号: PCT/CN2022/095072
- (22) 国际申请日: 2022年5月25日 (25.05.2022)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
PCT/CN2021/096477
2021年5月27日 (27.05.2021) CN
202110581290.3 2021年5月27日 (27.05.2021) CN
202111347952.7 2021年11月15日 (15.11.2021) CN
- (71) 申请人: 中国科学院动物研究所 (INSTITUTE OF ZOOLOGY, CHINESE ACADEMY OF SCIENCES) [CN/CN]; 中国北京市朝阳区北辰西路1号院, Beijing 100101 (CN)。北京干细胞与再生医学研究院 (BEIJING INSTITUTE FOR STEM CELL AND REGENERATIVE MEDICINE) [CN/CN]; 中国北京市朝阳区大屯路甲3号, Beijing 100101 (CN)。
- (72) 发明人: 李伟(LI, Wei); 中国北京市朝阳区北辰西路1号院, Beijing 100101 (CN)。周琪(ZHOU, Qi); 中国北京市朝阳区北辰西路1号院, Beijing 100101 (CN)。陈阳灿(CHEN, Yangcan); 中国北京市朝阳区北辰西路1号院, Beijing 100101 (CN)。胡艳萍(HU, Yanping); 中国北京市朝阳区北辰西路1号院, Beijing 100101 (CN)。王鑫阁(WANG, Xingge); 中国北京市朝阳区北辰西路1号院, Beijing 100101 (CN)。骆胜球(LUO, Shengqiu); 中国北京市朝阳区北辰西路1号院, Beijing 100101 (CN)。陈逸(CHEN, Yi); 中国北京市朝阳区北辰西路1号院, Beijing 100101 (CN)。
- (74) 代理人: 北京彩和律师事务所 (BEIJING CAI HE LAW FIRM); 中国北京市海淀区大柳树路17号富海国际港1602室, Beijing 100081 (CN)。
- (81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE,

(54) Title: ENGINEERED CAS12I NUCLEASE, EFFECTOR PROTEIN AND USE THEREOF

(54) 发明名称: 工程化的Cas12i核酸酶、效应蛋白及其用途

(57) Abstract: Provided is an engineered Cas12i nuclease, which comprises one or more of the following mutations based on a reference Cas12i nuclease: (1) replacing one or more amino acids interacting with PAM in the reference Cas12i nuclease with positively charged amino acids; and/or (2) replacing one or more amino acids, involved in opening the double strands of DNA, in the reference Cas12i nuclease with amino acids with aromatic rings; and/or (3) replacing one or more amino acids, which interact with a single-stranded DNA substrate and are located in an RuvC domain in the reference Cas12i nuclease, with positively charged amino acids; and/or (4) replacing one or more amino acids interacting with a DNA-RNA double helix in the reference Cas12i nuclease with positively charged amino acids; and/or (5) replacing one or more polar or positively charged amino acids interacting with the DNA-RNA double helix in the reference Cas12i nuclease with hydrophobic amino acids. The engineered Cas12i nuclease has significantly improved gene editing efficiency and/or significantly reduced off-target phenomenon compared with the reference Cas12i nuclease.

(57) 摘要: 提供了一种工程化的Cas12i核酸酶; 其包含一种或多种以下基于参比Cas12i核酸酶的突变: (1)将参比Cas12i核酸酶中与PAM相互作用的一个或多个氨基酸替换为带正电的氨基酸; 和/或(2)将参比Cas12i核酸酶中参与打开DNA双链的一个或多个氨基酸替换为带芳香环的氨基酸; 和/或(3)将参比Cas12i核酸酶中位于RuvC结构域并与单链DNA底物相互作用的一个或多个氨基酸替换为带正电的氨基酸; 和/或(4)将参比Cas12i核酸酶中与DNA-RNA双螺旋相互作用的一个或多个氨基酸替换为带正电的氨基酸; 和/或(5)将参比Cas12i核酸酶中与DNA-RNA双螺旋相互作用的一个或多个极性带正电荷的氨基酸替换为疏水氨基酸。该工程化的Cas12i核酸酶较参比Cas12i核酸酶具有显著提高的基因编辑效率和/或显著降低的脱靶现象。

WO 2022/247873 A1

PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE,
SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ,
UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

- (84) 指定国(除另有指明, 要求每一种可提供的地区
保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ,
NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM,
AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG,
CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU,
IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT,
RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI,
CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

- 包括国际检索报告(条约第21条(3))。
- 包括说明书序列表部分(细则5.2(a))。

说明书

工程化的 Cas12i 核酸酶、效应蛋白及其用途

5 优先权信息

本申请要求：于 2021 年 11 月 15 日提交的中国专利申请 CN202111347952.7、于 2021 年 5 月 27 日提交的中国专利申请 CN202110581290.3、以及于 2021 年 5 月 27 日提交的国际申请 PCT/CN2021/096477 的优先权，它们的全部内容通过整体引用并入本文。

10 同时提交的序列文件

下列 ASCII 码文本文件的全部内容通过整体引用并入本文：计算机可读格式(CRF)的序列表(FF00614PCT-NSequence6thVersion-20220525.txt, 日期：2022 年 5 月 25 日，大小：218kb)

15 技术领域

本申请属于生物技术领域。更具体地说，本申请涉及具有提高的催化活性(例如基因编辑活性)的 Cas12i 核酸酶、效应蛋白及其用途。

背景技术

20 基因组编辑是在基因组研究中的重要且有用的技术。有多个系统可用于基因组编辑，包括成簇规律间隔短回文重复序列(CRISPR)-Cas 系统、转录激活子样效应因子核酸酶(TALEN)系统以及锌指核酸酶(ZFN)系统。

CRISPR-Cas 系统是一种高效且具有成本效益的基因组编辑技术，可广泛应用于从酵母、植物到斑马鱼和人类的一系列真核生物中(参见综述：Van der Oost 2013, Science 339: 768-770, 以及 Charpentier and Doudna, 2013, Nature 495: 50-51)。CRISPR-Cas 系统通过结合 Cas12i 效应蛋白和 CRISPR RNA(crRNA)在古细菌和细菌中提供适应性免疫。迄今为止，基于该系统的突出的功能上和进化上的模块性，已经对包括六型(I-VI 型)两类(第 1 类和第 2 类)的 CRISPR-Cas 系统进行了表征。在第 2 类 CRISPR-Cas 系统中，II 型 Cas9 系统和 V 型-A/B/E/J Cas12a/Cas12b/Cas12e/Cas12j 系统已被利用来进行基因组编辑，并为
30 生物医学研究提供了广阔的前景。

但是，当前的 CRISPR-Cas 系统具有多种局限性，包括有限的基因编辑效率。因此，

需要改进方法和系统以进行有效的跨多基因座的基因组编辑。

发明概述

本申请提供了下述技术方案：

5 1.一种工程化的 Cas12i 核酸酶；其包含一种或多种基于参比 Cas12i 核酸酶的突变，所述突变选自：

(1)将参比 Cas12i 核酸酶中与 PAM 相互作用的一个或多个氨基酸替换为带正电的氨基酸；

10 (2)将参比 Cas12i 核酸酶中参与打开 DNA 双链的一个或多个氨基酸替换为带芳香环的氨基酸

(3)将参比 Cas12i 核酸酶中位于 RuvC 结构域并与单链 DNA 底物相互作用的一个或多个氨基酸替换为带正电的氨基酸

(4)将参比 Cas12i 核酸酶中与 DNA-RNA 双螺旋相互作用的一个或多个氨基酸替换为带正电的氨基酸；和

15 (5)将参比 Cas12i 核酸酶中与 DNA-RNA 双螺旋相互作用的一个或多个极性 or 带正电荷的氨基酸替换为疏水氨基酸；

优选地，所述参比 Cas12i 核酸酶为氨基酸序列为 SEQ ID NO: 1 的野生型 Cas12i2 核酸酶。

2.如项 1 所述的工程化的 Cas12i 核酸酶，其中，所述与 PAM 相互作用的一个或多个氨基酸是与 PAM 在三维结构上距离在 9 埃以内的氨基酸；优选，所述与 PAM 相互作用的一个或多个氨基酸是下述位置的一个或多个氨基酸：176、178、226、227、229、237、238、264、447 和 563，

进一步优选，所述与 PAM 相互作用的一个或多个氨基酸是下述一个或多个氨基酸：E176、E178、Y226、A227、N229、E237、K238、K264、T447 和 E563，

25 进一步优选，所述与 PAM 相互作用的一个或多个氨基酸是下述一个或多个氨基酸：E176、K238、T447 和 E563，

其中，所述氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

3.如项 1 或 2 所述的工程化的 Cas12i 核酸酶，其中，所述带正电的氨基酸是 R，K 或 H。

30 4.如项 1-3 中任一项所述的工程化的 Cas12i 核酸酶，其中，所述将参比 Cas12i 核酸酶中与 PAM 相互作用的一个或多个氨基酸替换为带正电的氨基酸是指如下替换中的一种

或多种：E176R、K238R、T447R 和 E563R；

优选，所述工程化的 Cas12i 核酸酶包含下述任何一种突变或突变组合：(1)E563R；
(2)E176R、T447R、E176R 和 E563R；(3)K238R 和 E563R；(4)E176R、K238R 和 T447R；
(5)E176R、K238R 和 E563R；(6)E176R、T447R 和 E563R；和(7)E176R、K238R、
5 T447R 和 E563R；

其中，所述氨基酸位置如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

5. 如项 1~4 中任一项所述的工程化的 Cas12i 核酸酶，其中，所述参与打开 DNA 双链的一个或多个氨基酸是与 PAM 中相对于靶标链的 3'端最后一个碱基对相互作用的氨基酸；

10 优选，所述参与打开 DNA 双链的一个或多个氨基酸是下述位置的一个或多个氨基酸：
163 和 164；

进一步优选，所述参与打开 DNA 双链的一个或多个氨基酸是下述一个或多个氨基酸：
Q163 和 N164；进一步优选，所述参与打开 DNA 双链的氨基酸为 N164；

其中，所述氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

15 6. 如项 1-5 中任一项所述的工程化的 Cas12i 核酸酶，其中，所述参与打开 DNA 双链的一个或多个氨基酸替换为带芳香环的氨基酸，所述带芳香环的氨基酸是 F、Y 或 W，
优选，所述带芳香环的氨基酸为 F 或 Y。

7. 如项 1-6 中任一项所述的工程化的 Cas12i 核酸酶，其中，所述将参比 Cas12i 核酸酶中参与打开 DNA 双链的一个或多个氨基酸的替换为带芳香环的氨基酸包括以下一个或多个替换：Q163F、Q163Y、Q163W、和 N164F；优选，所述 Cas12i 核酸酶包含 N164Y
20 或 N164F 突变；进一步优选，所述工程化的 Cas12i 核酸酶包含 N164Y。

8. 如项 1~7 中任一项所述的工程化的 Cas12i 核酸酶，其中，所述位于 RuvC 结构域并与单链 DNA 底物相互作用的一个或多个氨基酸是在三维结构上与单链 DNA 底物距离在 9 埃以内的氨基酸；

25 优选，所述位于 RuvC 结构域并与单链 DNA 底物相互作用的一个或多个氨基酸是下述位置的一个或多个氨基酸：323、327、355、359、360、361、362、388、390、391、
392、393、414、417、418、421、424、425、650、652、653、696、705、708、709、
751、752、755、840、848、851、856、885、897、925、926、928、929、932、1022。

进一步优选，所述位于 RuvC 结构域并与单链 DNA 底物相互作用的一个或多个氨基酸是下述一个或多个氨基酸：E323、L327、V355、G359、G360、K361、D362、L388、
30 N390、N391、F392、K393、Q414、L417、L418、K421、Q424、Q425、S650、E652、

G653、I696、K705、K708、E709、L751、S752、E755、N840、N848、S851、A856、Q885、M897、N925、I926、T928、G929、Y932、A1022；

进一步优选，所述位于 RuvC 结构域并与单链 DNA 底物相互作用的一个或多个氨基酸是是下述一个或多个氨基酸：E323、D362、Q425、N925、I926、N391、Q424 和
5 G929；

其中，所述氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

9.如项 1-8 中任一项所述的工程化的 Cas12i 核酸酶，其中，将参比 Cas12i 核酸酶中位于 RuvC 结构域并与单链 DNA 底物相互作用的一个或多个氨基酸替换为带正电的氨基酸包括替换为 R 或 K，优选，所述带正电的该氨基酸为 R。

10 10.如项 1-9 中任一项所述的工程化的 Cas12i 核酸酶，其中，将参比 Cas12i 核酸酶位于 RuvC 结构域并与单链 DNA 底物相互作用的一个或多个氨基酸替换为带正电的氨基酸是指包括如下替换中的一种或多种：E323R、D362R、N391R、Q424R、Q425R、N925R、I926R 和 G929R；

优选，所述工程化的 Cas12i 核酸酶包含下述任何一种突变或突变组合：(1)E323R；
15 (2)D362R；(3)Q425R；(4)N925R；(5)I926R；(6)E323R 和 D362R；(7)E323R 和 Q425R；(8)E323R 和 I926R；(9)Q425R 和 I926R；(10)D362R 和 I926R；(11)N925R 和 I926R；(12)E323R、D362R 和 Q425R；(13)E323R、D362R 和 I926R；(14)E323R、Q425R 和 I926R；(15)D362R、N925R 和 I926R；和(16) E323R、D362R、Q425R 和 I926R；

其中，所述氨基酸位置如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

20 11.如项 1~10 中任一项所述的工程化的 Cas12i 核酸酶，其中，所述与 DNA-RNA 双螺旋相互作用的一个或多个氨基酸是与 DNA-RNA 双螺旋在三维结构上距离在 9 埃以内的氨基酸；

优选，所述与 DNA-RNA 双螺旋相互作用的一个或多个氨基酸是下述位置的一个或多个氨基酸：116、117、156、159、160、161、247、293、294、297、301、305、306、
25 308、312、313、316、319、320、343、348、349、427、433、438、441、442、679、683、691、782、783、797、800、852、853、855、861、865、957、958；

进一步优选，所述与 DNA-RNA 双螺旋相互作用的一个或多个氨基酸是下述一个或多个氨基酸：G116、E117、A156、T159、S161、T301、I305、K306、T308、N312、F313、D427、K433、V438、N441、Q442、M852、L855、N861、Q865、E160、Q316、
30 E319、Q320、E247、E343、E348、E349、N679、E683、E691、D782、E783、E797、E800、D853、S957、D958、G293、E294 和 N297；G116、E117、A156、T159、E160、

S161、E247、G293、E294、N297、T301、I305、K306、T308、N312、F313、Q316、E319、Q320、E343、E348、E349、D427、K433、V438、N441、Q442、N679、E683、E691、D782、E783、E797、E800、M852、D853、L855、N861、Q865、S957、D958；

进一步优选，所述与 DNA-RNA 双螺旋相互作用的一个或多个氨基酸是下述一个或多个氨基酸：G116、E117、T159、S161、E319、E343 和 D958；其中，所述氨基酸位置如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

12. 如项 1-11 中任一项所述的工程化的 Cas12i 核酸酶，其中，将参比 Cas12i 核酸酶中与 DNA-RNA 双螺旋相互作用的一个或多个氨基酸替换为带正电的氨基酸包括替换为 R 或 K，优选，所述带正电的氨基酸为 R。

10 13. 如项 1-12 中任一项所述的工程化的 Cas12i 核酸酶，其中，将参比 Cas12i 核酸酶中与 DNA-RNA 双螺旋相互作用的一个或多个氨基酸替换为带正电的氨基酸包括替换为以下的一种或多种：G116R、E117R、T159R、S161R、E319R、E343R 和 D958R；

优选，所述工程化的 Cas12i 核酸酶包含 D958R 替换；

其中，所述氨基酸位置如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

15 14. 如项 1~13 中任一项所述的工程化的 Cas12i 核酸酶，所述与 DNA-RNA 双螺旋相互作用的一个或多个极性或非极性带正电荷的氨基酸选自如下的一个或多个位置的氨基酸：357、394、715、719、807、844、848、857、861，即下述一个或多个氨基酸：H357、K394、R715、R719、K807、K844、N848、R857、R861；

其中，所述氨基酸位置如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

20 15. 如项 1-14 中任一项所述的工程化的 Cas12i 核酸酶，其中，将参比 Cas12i 核酸酶中与 DNA-RNA 双螺旋相互作用的一个或多个极性或非极性带正电荷的氨基酸替换为疏水氨基酸包括替换为丙氨酸(A)。

16. 如项 1-15 中任一项所述的工程化的 Cas12i 核酸酶，其包含选自下组的一个或多个突变：H357A、K394A、R715A、R719A、K807A、K844A、N848A、R857A、R861A；

25 优选，包含选自下组的一个或多个突变：K394A、R719A、K844A、R857A；

其中，所述氨基酸位置如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

17. 如项 1-16 中任一项所述的工程化的 Cas12i 核酸酶，其包含 R719A 和 K844A 氨基酸取代，或包含 R857A 和 K844A 氨基酸取代；其中，所述氨基酸位置如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

30 18. 如项 1~17 中任一项所述的工程化的 Cas12i 核酸酶，其还包含一个或多个柔性区突变，所述突变增加了参比 Cas12i 核酸酶中的柔性区的柔性，所述柔性区选自氨基酸残

基 439-443 或氨基酸残基 925-929;

优选, 所述柔性区突变位于 439 和/或 926 位点;

进一步优选, 所述柔性区突变是 L439 和/或 I926 的突变;

其中, 所述氨基酸位置如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

5 19. 如项 18 所述的工程化的 Cas12i 核酸酶, 其中, 所述一个或多个柔性区突变为:
将该柔性区氨基酸替换为 G、和/或在其后插入一个或两个 G;

优选, 所述一个或多个柔性区突变包含 I926G、L439(L+G)或 L439(L+GG);

进一步优选, 所述一个或多个柔性区突变包含 L439(L+G)或 L439(L+GG);

其中, 所述氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

10 20. 一种工程化的 Cas12i 核酸酶;

所述工程化的 Cas12i 核酸酶包含任何以下一组或多组突变:

(1)E563R; (2)E176R 和 T447R; (3)E176R 和 E563R; (4)K238R 和 E563R; (5) E176R、
K238R 和 T447R; (6)E176R、 T447R 和 E563R; (7) E176R、 K238R 和 E563R; (8) E176R、
K238R、 T447R 和 E563R; (9) N164Y; (10) N164F; (11) E323R; (12) D362R; (13) Q425R;
15 (14) N925R; (15) I926R; (16) D958R; (17) E323R 和 D362R; (18) E323R 和 Q425R; (19)
E323R 和 I926R; (20) Q425R 和 I926R; (21) D362R 和 I926R; (22) N925R 和 I926R; (23)
E323R、 D362R 和 Q425R; (24) E323R、 D362R 和 I926R; (25) E323R、 Q425R 和 I926R;
(26) D362R、 N925R 和 I926R; (27) E323R、 D362R、 Q425R 和 I926R; (28) D362R 和
I926G; (29)N925R 和 I926G; (30)D362R、 N925R 和 I926G; (31)I926R 和 L439(L+G);
20 (32)I926R 和 L439(L+GG); (33)E323R、 D362R 和 I926G; (34)R719A 和 K844A; 和
(35)R857A 和 K844A;

优选, 所述工程化的 Cas12i 核酸酶包含任何以下一组或多组突变: (1)E176R、
K238R、 T447R 和 E563R; (2)N164Y; (3)I926R; (4) E323R 和 D362R; (4) I926G; (5)I926R
和 L439(L+G); (6)I926R 和 L439(L+GG); 和(7)D958R;

25 其中, 所述氨基酸位置如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

21. 一种工程化的 Cas12i 核酸酶, 所述工程化的 Cas12i 核酸酶包含以下任一组突变:
(1)E176R、 K238R、 T447R、 E563R 和 N164Y; (2)E176R、 K238R、 T447R、 E563R 和
I926R; (3)N164Y、 E323R 和 D362R; (4)E176R、 K238R、 T447R、 E563R、 E323R 和
D362R; (5)N164Y 和 I926R; (6)E176R、 K238R、 T447R、 E563R、 N164Y 和 I926R;
30 (7)E176R、 K238R、 T447R、 E563R、 N164Y、 E323R 和 D362R; (8)E176R、 K238R、
T447R、 E563R、 N164Y、 I926R、 E323R 和 D362R; (9)E176R、 K238R、 T447R、 E563R、

N164Y、E323R、D362R 和 I926G；(10)E176R、K238R、T447R、E563R、N164Y、E323R、D362R、I926G 和 L439(L+GG)；(11) E176R、K238R、T447R、E563R、N164Y、E323R、D362R、I926G 和 L439(L+G)；(12)E176R、K238R、T447R、E563R、N164Y 和 D958R；(13)E176R、K238R、T447R、E563R、I926R 和 D958R；(14) E176R、K238R、T447R、E563R、E323R、D362R 和 D958R；(15) N164Y、I926R 和 D958R；(16) N164Y、E323R、D362R 和 D958R；(17) E176R、K238R、T447R、E563R、N164Y、I926R 和 D958R；(18)E176R、K238R、T447R、E563R、N164Y、E323R、D362R 和 D958R；(19)E176R、K238R、T447R、E563R、N164Y、I926R、E323R、D362R 和 D958R；(20) E176R、K238R、T447R、E563R、N164Y、E323R、D362R、I926G 和 D958R；(21)E176R、K238R、T447R、E563R、N164Y、E323R、D362R、I926G、L439(L+GG)和 D958R；(22)E176R、K238R、T447R、E563R、N164Y、E323R、D362R、I926G、L439(L+G)和 D958R；(23) E176R、K238R、T447R、E563R、N164Y、E323R、D362R 和 R857A；(24)E176R、K238R、T447R、E563R、N164Y、E323R、D362R 和 N861A；(25)E176R、K238R、T447R、E563R、N164Y、E323R、D362R 和 K807A；(26)E176R、K238R、T447R、E563R、N164Y、E323R、D362R 和 N848A；(27)E176R、K238R、T447R、E563R、N164Y、E323R、D362R 和 R715A；(28)E176R、K238R、T447R、E563R、N164Y、E323R、D362R 和 R719A；(29) E176R、K238R、T447R、E563R、N164Y、E323R、D362R 和 K394A；(30)E176R、K238R、T447R、E563R、N164Y、E323R、D362R 和 H357A；(31)E176R、K238R、T447R、E563R、N164Y、E323R、D362R 和 K844A；(32)E176R、K238R、T447R、E563R、N164Y、E323R、D362R、R719A 和 K844A；或(33)E176R、K238R、T447R、E563R、N164Y、E323R、D362R、R857A 和 K844A；其中，所述氨基酸位置如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

22. 如项 1~21 中任一项所述的工程化的 Cas12i 核酸酶，其包含如 SEQ ID NOs: 2~24 中任一项所示氨基酸序列的工程化 Cas12i 核酸酶，或与如 SEQ ID NOs: 2~24 中任一序列所示的氨基酸序列具有至少 80%同一性的氨基酸序列。

23. 一种工程化的 Cas12i 效应蛋白，其包含项 1~22 中任一项所述的工程化的 Cas12i 核酸酶或其功能衍生物；

任选的，所述工程化的 Cas12i 核酸酶或其功能衍生物具有酶活性，或者所述工程化的 Cas12i 核酸酶或其功能衍生物为酶失活突变体。

24. 如项 23 所述的工程化的 Cas12i 效应蛋白，其中所述 Cas12i 效应蛋白能够诱导 DNA 分子中的双链断裂或单链断裂。

25. 如项 23 所述的工程化的 Cas12i 效应蛋白，其中所述工程化的 Cas12i 核酸酶或其功能衍生物为包含以下一种或多种突变的酶失活突变体：D599A、E833A、S883A、H884A、R900A 和 D1019A；其中，所述氨基酸位置如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

5 26. 如项 23~25 中任一项所述的工程化的 Cas12i 效应蛋白，其还包含与所述工程化的 Cas12i 核酸酶或其功能衍生物融合的功能结构域。

27. 如项 26 所述的工程化的 Cas12i 效应蛋白，其中所述功能结构域选自下组中的一个或多个：翻译起始结构域、转录阻遏结构域、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、报告分子结构域和核酸酶结构域。

10 28. 如项 23-27 中任一项所述的工程化的 Cas12i 效应蛋白，所述工程化的 Cas12i 效应蛋白包含：含有所述工程化的 Cas12i 核酸酶或其功能衍生物的 N 末端部分的第一多肽和含有所述工程化的 Cas12i 核酸酶或其功能衍生物的 C 末端部分的第二多肽，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 的存在下彼此缔合，以形成与靶核酸特异性结合的成簇规律间隔短回文重复序列(CRISPR)复合物，所述靶核酸包含与
15 所述指导序列互补的靶序列；

优选，所述第一多肽包含项 1-22 中任一项所述的工程化的 Cas12i 核酸酶的 N 末端部分氨基酸残基 1 至 X，所述第二多肽包含项 1-22 中任一项所述的工程化的 Cas12i 核酸酶的氨基酸残基 X+1 至所述 Cas12i 核酸酶的 C 末端；

可选的，所述第一多肽和所述第二多肽各自包含二聚化结构域；

20 可选的，所述第一二聚结构域和所述第二二聚结构域在诱导剂存在下彼此缔合。

29. 一种工程化的 CRISPR-Cas12i 系统，包括：

(a)项 1-28 中任一项所述的工程化的 Cas12i 核酸酶、或项 23-28 中任一项所述的工程化的 Cas12i 效应蛋白；以及

25 (b)包含与靶序列互补的指导序列的指导 RNA，或编码所述指导 RNA 的一种或多种核酸，

其中所述工程化的 Cas12i 核酸酶或工程化的 Cas12i 效应蛋白和所述指导 RNA 能够形成 CRISPR 复合物，所述 CRISPR 复合物特异性结合包含所述靶序列的靶核酸并诱导所述靶核酸的修饰；

优选，所述指导 RNA 是包含所述指导序列的 crRNA；

30 进一步优选，所述工程化的 CRISPR-Cas12i 系统包括编码多个 crRNA 的前体指导 RNA 阵列(array)；

或者, 优选所述工程化的 Cas12i 核酸酶或工程化的 Cas12i 效应蛋白是主编辑器, 所述指导 RNA 是引导编辑指导 RNA (pegRNA)。

30. 如项 29 所述的工程化的 CRISPR-Cas12i 系统, 其包含一种或多种编码所述工程化的 Cas12i 核酸酶或工程化的 Cas12i 效应蛋白的载体;

5 优选, 所述一种或多种载体选自下组: 逆转录病毒载体、慢病毒载体、腺病毒载体、腺相关的载体和单纯疱疹载体;

进一步优选, 所述一种或多种载体是腺相关病毒 AAV 载体;

进一步优选, 所述 AAV 载体还编码所述指导 RNA。

31. 一种检测样品中靶核酸的方法, 包括:

10 (a) 使样品与项 29 中的工程化的 CRISPR-Cas12i 系统以及加标签的检测核酸接触, 该检测核酸为单链且不与所述指导 RNA 的指导序列杂交; 以及

(b) 测量通过所述工程化的 Cas12i 核酸酶或工程化的 Cas12i 效应蛋白切割所述加标签的检测核酸而产生的可检测信号, 从而检测所述靶核酸。

15 32. 一种修饰包含靶序列的靶核酸的方法, 包括使所述靶核酸与项 29 或 30 所述的工程化的 CRISPR-Cas12i 系统接触;

优选, 所述方法在体外进行、离体进行或在体内进行;

进一步优选, 所述靶核酸存在于细胞中;

进一步优选, 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;

20 进一步优选, 所述靶核酸是基因组 DNA;

进一步优选, 所述靶序列与疾病或病症相关;

进一步优选, 所述工程化的 CRISPR-Cas12i 系统包括编码多个 crRNA 的前体指导 RNA 阵列, 其中每个 crRNA 包含不同的指导序列。

25 33. 如项 29 或 30 所述的工程化的 CRISPR-Cas12i 系统在制备治疗与个体的细胞中靶核酸相关的疾病或病症的药物中的用途; 优选, 所述疾病或病症选自下组: 癌症、心血管疾病、遗传性疾病、自身免疫疾病、代谢性疾病、神经退行性疾病、眼病、细菌感染和病毒感染。

30 34. 一种治疗与个体的细胞中的靶核酸相关的疾病或病症的方法, 所述方法包含使用项 32 所述的方法来修饰所述个体的细胞中的靶核酸, 从而治疗所述疾病或病症; 优选, 所述疾病或病症选自下组: 癌症、心血管疾病、遗传性疾病、自身免疫疾病、代谢性疾病、神经退行性疾病、眼病、细菌感染和病毒感染。

35. 一种修饰包含靶序列的靶核酸的方法，包括使所述靶核酸与项 29 或 30 中所述的工程化的 CRISPR-Cas12i 系统接触。

36. 包含如项 1-22 中任一项所述的工程化的 Cas12i 核酸酶、如项 23-28 中任一项所述的工程化的 Cas12i 效应蛋白的组合物或试剂盒。

5 37. 工程化的细胞，其包含经修饰的靶核酸，其中靶核酸由如项 32 或 35 所述的方法修饰。

38. 工程化的非人类动物，其包含一种或多种如项 37 所述的工程化的细胞。

本申请的技术方案取得的有益效果

10 本申请工程化的 Cas12i 核酸酶及其效应蛋白具有更高的活性，如切割核酸底物的催化效率及细胞内的基因编辑效率。本申请中的工程化 Cas12i 核酸酶具有较现有常规 Cas 基因编辑工具更卓越的哺乳动物细胞(如人类细胞)内的基因编辑效率；例如本申请中的一些示例性 Cas12i 核酸酶突变体在人类细胞中多个位点(如 62 个位点)测试基因编辑效率，发现有 57 个位点的基因编辑效率超过约 60%，平均基因编辑效率接近 70%。在一些
15 实施例中，本申请工程化的 Cas12i 核酸酶及其效应蛋白还具有以下一个或多个优点：蛋白小(1,054aa)，crRNA 组分简单，PAM 序列简单，并且蛋白自身能加工前体 crRNA。另外，本申请提供在高活性工程化 Cas12i 核酸酶基础上(如 SEQ ID NO.8)进一步人工改造的具有更低的脱靶率和更高的特异性的 Cas12i 核酸酶。这些优点使得本申请高效的工程化的 Cas12i 核酸酶及其效应蛋白非常适用于在体内进行基因编辑或者基因调控。

20

附图说明

图 1a：将参比 Cas12i 核酸酶(如 SEQ ID NO.1 所示的野生型 Cas12i 核酸酶)中与 PAM 相互作用的氨基酸替换为带正电的氨基酸，从而提高基因编辑效率。如图所示，E176R, K238R, T447R, E563R 四种突变体能够显著地提高在人 293T 细胞中的基因编辑效率。
25 率。

图 1b：将图 1a 中能够显著提高基因编辑效率的氨基酸突变(E176R, K238R, T447R, E563R)进行组合，发现组合后的突变体能够展现在人 293T 细胞中更高的基因编辑效率。

图 2：将参比 Cas12i 核酸酶中参与打开 DNA 双链的氨基酸替换为带芳香环的氨基酸从而提高基因编辑的效率。如图所示，Q163F、Q163Y、Q163W、N164F、N164Y 这些
30 突变体能够显著提高在人 293T 细胞中的基因编辑效率。

图 3a、3b、3c：将参比 Cas12i 核酸酶中位于 RuvC 结构域并与单链 DNA 底物相互

作用的氨基酸替换为带正电的氨基酸，从而提高基因编辑的效率。如图 3a、3b、3c 所示，E323R、L327R、V355R、G359R、G360R、D362R、N391R、Q424R、Q425R、N925R、I926R 和 G929R 等突变体能够显著地提高在人 293T 细胞中的基因编辑效率。

图 3d: 将图 3a、3b 中提高效率的点突变进行组合，发现组合后的突变体能够展现在人 293T 细胞中更高的基因编辑效率。

图 3e: 将图 3a、3b 中提高效率的点突变以及根据分子柔性原理改造突变(L439(L+GG), I926G)进行组合，发现组合后的突变体能够展现在人 293T 细胞中更高的基因编辑效率。

图 4: 将参比 Cas12i 酶中与 DNA-RNA 双螺旋相互作用的氨基酸替换为带正电的氨基酸，从而提高基因编辑效率。如图 4 所示，G116R、E117R、T159R、S161R、E319R、E343R、D958R 突变体能够明显地提高在人 293T 细胞中的基因编辑效率。其中，D958R 突变最优。

图 5: 将图 1a~图 3e 中三种改造策略得到的高效率突变体以及根据分子柔性原理改造突变(L439(L+GG)、I926G)进行组合，发现组合后的突变体能够展现在人 293T 细胞中更高的基因编辑效率。组合后能够极大地提高基因编辑效率。选出基因编辑效果最佳的突变体并命名为 CasXX 供后续实验使用。CasXX (SEQ ID NO: 8) 是具有 E176R+K238R+T447R+E563R+N164Y+E323R+D362R 突变组合(基于氨基酸序列 SEQ ID NO: 1 的参比 Cas12i2)的 Cas12i 工程酶。

图 6a: CasXX 在 62 个人基因组位点的基因编辑效率汇总。PAM = NTTN。

图 6b: CasXX 与 AsCas12a, BhCas12b v4 基因编辑效率的比较。

图 6c: CasXX 与 SpCas9, SaCas9, SaCas9-KKH 基因编辑效率的比较。

图 6d: CasXX 在小鼠 Hepa1-6 细胞系的基因编辑效率统计。可以看到，CasXX 在小鼠 Hepa1-6 细胞内 65 个位点展示出强大的基因编辑能力，平均基因编辑效率超过 60%。

图 7 展示了 CasXX 在 64 个人基因位点的基因编辑效率。其中，这 64 个位点包含的 PAM 序列覆盖了所有 NNNN 的组合：NTTN, NTAN, NTCN, NTGN, NATN, NAAN, NACN, NAGN, NCTN, NCAN, NCCN, NCGN, NGTN, NGAN, NGCN, NGGN。如图 7 所示，CasXX 在 NTTN, NTAN, NTCN, NTGN, NATN, NAAN, NCTN, NCAN, NGTN 这些 PAM 序列都展现了高效的基因编辑效率，在这些位点的平均基因编辑效率超过 40%。

图 8 展示了野生型 Cas12i2 蛋白(SEQ ID NO: 1)以及 CasXX 蛋白在体外切割双链 DNA 的结果。野生型 Cas12i2 蛋白仅仅对含有 NTTN PAM 的双链 DNA 具有部分切割的效率，但是对其余的 PAM 几乎没有切割活性。但是，CasXX 蛋白对含有 NTTN, NTAN,

NTCN, NTGN, NATN, NAAN, NACN, NCTN, NCAN, NGTN, NGAN PAM 的双链 DNA 都展现了高效的切割效率。几乎能完全切割含有 NTTN, NTAN, NTCN, NATN, NAAN, NACN, NCTN, NCAN, NGTN PAM 的双链 DNA。

图 9: Cas12i2(SEQ ID NO: 1)与 Cas12i1(SEQ ID NO: 13)的氨基酸序列的同源性比对。阴影标注的氨基酸代表 2 种 Cas12i 蛋白相同的氨基酸, 用白色框标注的氨基酸代表 2 种 Cas12i 蛋白性质相似的氨基酸。

图 10 为使用 GUIDE-Seq 检测 CasXX 在 EMX1-7 靶位点的脱靶效应。序列之后的数字代表每种序列被测到的读数(reads)。第一行序列为参考靶序列(SEQ ID NO:77), 下面的序列分别代表靶序列和脱靶序列以及双链 DNA 标签在靶序列和脱靶序列富集的读数(reads)。该图说明 CasXX 在 EMX1-7 位点存在脱靶效应。

图 11 为在 CasXX 原有序列的基础上添加新的点突变或者删减原有的点突变, 构建新的 HF 突变体。实验结果表明, R857A, N861A, K807A, N848A, R715A, R719A, K394A, H357A, K844A 这些基于 CasXX 序列的单点突变体能够有效地降低在 EMX1-7-OT-1, EMX1-7-OT-2, EMX1-7-OT-3 的插入缺失比例。

图 12 为在 CasXX 原有序列的基础上添加新的点突变或者删减原有的点突变, 构建新的 HF 突变体。实验结果表明, R857A, N861A, K807A, N848A, R715A, R719A, K394A, H357A, K844A 这些基于 CasXX 序列的单点突变体在使用模拟脱靶的 crRNA-RNF2-1-Mis-1/2, crRNA-RNF2-1-Mis-5/6, crRNA-RNF2-1-Mis-17/18 和 crRNA-RNF2-1-Mis-19/20 时能够有效降低插入缺失的比例。

图 13 为使用荧光报告系统筛选提高基因编辑的特异性的突变体(基于 CasXX 的 HF 突变体)的实验流程。将 600ng 编码 Cas 蛋白的质粒, 300ng 编码 crRNA 的质粒和 100ng 编码 mCherry 的质粒, 转染到 24 孔细胞培养皿培养的细胞中。转染 3 天后, 使用流式分析计算每个样品剩余 mCherry 阳性的比例。计算公式如图展示。

图 14 为 CasXX 以及突变体在 293T 细胞中编辑位于质粒上的 mCherry 基因。质粒转染后, 培养三天, 进行流式分析。Cas 蛋白的基因编辑效率越高, 细胞中 mCherry 阳性的比例就会越低。实验流程对应图 13 实验流程示意图; 针对 mCherry-FM、mCherry-Mis-1/2、mCherry-Mis-5/6 以及 mCherry-Mis-19/20 的间隔 (spacer) 序列分别如 SEQ ID NO:37、38、39 和 40 所示。

图 15 为在 CasXX 原有序列的基础上构建新的 HF 突变体组合。实验结果表明, R857A、R719A、K394A、K844A、R719A/K844A 以及 R857A/K844A 这些基于 CasXX 序列的突变体能够有效地降低在 EMX1-7-OT-1, EMX1-7-OT-2, EMX1-7-OT-3 的脱靶插

入缺失比例，并且不牺牲在靶位点的效率。

图 16 为在 CasXX 原有序列的基础上构建新的 HF 突变体组合。实验结果表明，R857A, R719A, K394A, K844A, R719A/K844A 以及 R857A/K844A 这些基于 CasXX 序列的突变体在使用模拟脱靶的 crRNA-RNF2-1-Mis-1/2, crRNA-RNF2-1-Mis-5/6、crRNA-RNF2-1-Mis-17/18 和 crRNA-RNF2-1-Mis-19/20 时能够有效降低插入缺失的比例，并且不牺牲在靶位点的效率。

图 17 为使用 GUIDE-Seq 检测 CasXX 以及 CasXX + K394A 突变体在 CD34-7 靶位点的脱靶效应。序列之后的数字代表每种序列被测到的 reads 数目。第一行序列为参考靶序列(SEQ ID NO:78)，下面的序列分别代表靶序列和脱靶序列以及双链 DNA 标签在靶序列和脱靶序列富集的 reads 数目。该图说明 CasXX 在 CD34-7 位点存在脱靶效应，但是 CasXX + K394A 突变体在该位点没有脱靶效应。

具体实施方式

需要说明的是，在说明书及权利要求当中使用了某些词汇来指称特定组件。本领域技术人员应可以理解，技术人员可能会用不同名词来称呼同一个组件。本说明书及权利要求并不以名词的差异来作为区分组件的方式，而是以组件在功能上的差异来作为区分的准则。如在通篇说明书及权利要求当中所提及的“包含”或“包括”为一开放式用语，故应解释成“包含但不限于”。说明书后续描述为实施本发明的较佳实施方式，然所述描述乃以说明书的一般原则为目的，并非用以限定本发明的范围。本发明的保护范围当视所附权利要求所界定者为准。除非另有定义，本文中使用的所有技术和科学术语具有与本公开所属领域的普通技术人员通常所理解的不同含义。

I. 术语

如本文所用，“效应蛋白”是指具有活性如位点特异性结合活性、单链 DNA 切割活性、双链 DNA 切割活性、单链 RNA 切割活性、DNA 或 RNA 修饰(例如切割，碱基置换、插入、移除)或转录调节活性的蛋白。

如本文所用，“指导 RNA”和“gRNA”在本文中可互换使用，是指能够与 Cas12i 效应蛋白和靶核酸(例如，双链 DNA)形成复合物的 RNA。本文还考虑了可以被加工成多个 crRNA 的前体指导 RNA 阵列。“crRNA”或“CRISPR RNA”包含与靶核酸(例如，双链 DNA)的靶序列具有足够互补性的指导序列，其指导 CRISPR 复合物与靶核酸的靶序列特异性结合。

本文使用的术语“CRISPR 阵列”是指包括 CRISPR 重复和间隔子的核酸(例如 DNA)片段, 其从第一个 CRISPR 重复的第一个核苷酸开始并以最后一个(末端)CRISPR 重复的最后一个核苷酸结束。典型地, CRISPR 阵列中的每个间隔子位于两个重复之间。本文中使用的术语“CRISPR 重复”或“CRISPR 直接重复”或“直接重复”是指多个短的直接重复序列, 其在 CRISPR 阵列中显示出非常小的序列变化或没有序列变化。适当地, 直接重复可以形成茎环结构。

术语“核酸”、“多核苷酸”和“核苷酸序列”可互换使用, 是指任何长度的核苷酸的聚合形式, 包括脱氧核糖核苷酸、核糖核苷酸、其组合及其类似物。“寡核苷酸”和“低聚核苷酸”可互换使用, 是指具有不超过约 50 个核苷酸的短多核苷酸。如本文所用, “互补性”是指核酸通过传统的沃森-克里克(Watson-Crick)碱基配对与另一核酸形成氢键的能力。互补性百分比表示可与第二种核酸形成氢键(即, 沃森-克里克碱基配对)的核酸分子中的残基百分比(例如, 10 分之 5、6、7、8、9、10, 分别互补约 50%、60%、70%、80%、90% 和 100%)。“完全互补”是指核酸序列的所有连续残基与第二核酸序列中相同数目的连续残基形成氢键。如本文所用, “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上, 互补程度为至少约 70%、75%、80%、85%、90%、95%、97%、98%、99% 或 100 中任一个, 或指在严格条件下杂交的两种核酸。

如本文所用, 用于杂交的“严格条件”是指与靶序列具有互补性的核酸主要与靶序列杂交, 而基本上不与非靶序列杂交的条件。严格条件通常是序列依赖性的, 并且取决于许多因素而变化。通常, 序列越长, 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例子详细描述于 Tijssen (1993), *Laboratory Techniques In Biochemistry And Molecular Biology- Hybridization With Nucleic Acid Probes*, 第一部分第二章“Overview of principles of hybridization and the strategy of nucleic acid probe assay,” Elsevier, N,Y。

“杂交”是指一种或多种多核苷酸反应形成复合物的反应, 所述复合物通过核苷酸残基的碱基之间的氢键而稳定化。氢键可通过沃森克里克碱基配对、霍格斯坦(Hoogsteen)结合或以任何其他序列特异性方式发生。能够与给定序列杂交的序列称为所述给定序列的“互补体”。

针对核酸序列的“序列同一性百分比(%)”定义为, 在通过允许空缺(gaps)来比对序列(如有必要)以实现最大的序列同一性百分比后, 候选序列中与特定核酸序列中的核苷酸相同的核苷酸百分比。针对肽、多肽或蛋白质序列的“序列同一性百分比(%)”, 是在通过允许空缺来比对序列(如有必要)以实现最大的序列同源性百分比后, 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

性百分比的目的，比对可以以本领域技术范围内的各种方式来实现，例如，使用诸如 BLAST、BLAST-2、ALIGN 或 MEGALIGN™ (DNASTAR) 软件之类的公众可获得的计算机软件。本领域技术人员可以确定用于测量比对的合适参数，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。

5 术语“多肽”和“肽”在本文可互换使用，是指任何长度的氨基酸的聚合物。所述聚合物可以是直链或支链的，它可以包含经修饰的氨基酸，并且可以被非氨基酸中断。蛋白质可以具有一个或多个多肽。该术语还涵盖已经过修饰的氨基酸聚合物；例如，二硫键的形成、糖基化、脂质化、乙酰化、磷酸化或任何其他操作(诸如与标记组分的缀合)。

如本文所用，“变体”解释为分别不同于参比多核苷酸或多肽但保留必要特性的多核苷酸或多肽。多核苷酸的典型变体与另一参比多核苷酸的核酸序列不同。变体核酸序列的变化可以改变或不改变参比多核苷酸编码的多肽的氨基酸序列。核苷酸变化可导致参比序列编码的多肽中的氨基酸替换、添加、缺失、融合和截短，如下所述。多肽的典型变体与另一参比多肽在氨基酸序列上不同。通常，差异是有限的，使得参比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。变体和参比多肽的氨基酸序列
10 可以通过一个或多个替换、添加、缺失的任何组合而不同。替换或插入的氨基酸残基可以是或可以不是遗传密码编码的氨基酸残基。多核苷酸或多肽的变体可以是天然存在的(诸如等位基因变体)，或者可以是未知天然存在的变体。多核苷酸和多肽的非天然存在的变体可以通过诱变技术，通过直接合成，以及通过本领域技术人员已知的其他重组方法来制备。

20 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，意指当它存在于大自然中时，将其与突变体或变体区分开的、典型形式的生物体、菌株、基因或特征。它可以与自然界中的资源隔离开来，并没有被刻意修饰。

如本文所用，术语“非天然存在”或“工程化的”可互换使用，是指人工参与。当这些术语用于描述核酸分子或多肽时，是指所述核酸分子或多肽至少基本上不含其天然缔合的或天然存在的至少一种其他组分。
25

如本文所用，术语“直系同源物(orthologue/ 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，本文所指的蛋白质的“直系同源物”是指属于不同物种的蛋白质，其执行与作为其直系同源物的蛋白质相同或相似的功能。

如本文所用，术语“同一性”用于表示两个多肽之间或两个核酸之间的序列匹配。当
30 两个比较序列中的一个位置被相同的碱基或氨基酸单体亚基占据时(例如，两个 DNA 分子的每个中的一个位置都被腺嘌呤占据，或者两个多肽的每个中的一个位置被赖氨酸占

据), 那么在那个位置每个分子均相同。这两个序列之间的“同一性百分比”是两个序列共有的匹配位置数除以要比较的位置数 x 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 具有 50% 的同一性(总共 6 个位置中有 3 个匹配)。通常, 当两个序列进行比对以产生最大的同一性时, 进行这种比较。这种比对可以通过例如 Needleman et al. (1970) J. Mol. Biol. 48: 443-453 中的方法来实现, 所述方法可方便地通过计算机程序如比对(Align)程序 (DNASTar, Inc.)来进行。也可以采用 PAM 120 权重残基表, 使用 E. Meyers 和 W. Miller 的算法(Comput. Appl Biosci., 4: 11-17 (1988))集成到 ALIGN 程序(2.0 版)中。空缺长度罚分 12 和空缺罚分 4, 用于确定两个氨基酸序列之间的同一性百分比。此外, 可以使用集成到 GCG 软件包(可从 www.gcg.com 获得)的 GAP 程序中的 Needleman 和 Wunsch(J Mol Biol. 48: 444-453 (1970))算法, 采用 Blossum 62 矩阵或 PAM250 矩阵, 空缺权重为 16、14、12、10、8、6 或 4, 长度权重为 1、2、3、4、5 或 6, 以确定两个氨基酸序列之间的同一性百分比。

本文所用的“细胞”应理解为不仅指特定的单个细胞, 而且指该细胞的后代或潜在后代。因为由于突变或环境影响, 可能在后代中发生某些修饰, 所以此类后代可能事实上与亲本细胞不同, 但仍包括在本文术语的范围内。

如本文所用, 术语“转导”和“转染”包括本领域已知的使用感染剂(如病毒)或其他方式将 DNA 引入细胞中以表达目的蛋白质或分子的方法。除了病毒或类似病毒的试剂外, 还有基于化学的转染方法, 如使用磷酸钙、树状聚合物, 脂质体或阳离子聚合物(例如 DEAE-葡聚糖或聚乙烯亚胺)的转染方法; 非化学方法, 如电穿孔、细胞挤压(cell squeezing)、声致穿孔(sonoporation)、光学转染、穿刺转染(impalefection)、原生质体融合、质粒递送或转座子; 基于颗粒的方法, 如使用基因枪、磁转染或磁体辅助转染、颗粒轰击; 以及杂交方法(诸如核转染)。

如本文所用, 术语“转染的”、“转化的”或“转导的”是指将外源核酸转移或引入宿主细胞的过程。“转染的”、“转化的”或“转导的”细胞是已经用外源核酸转染、转化或转导的细胞。

术语“体内”是指从其中获得细胞的该生物体内。“离体”或“体外”是指从其中获得细胞的该生物体外。

如本文所用, “治疗(treatment/treating)”是用于获得有益的或期望的结果(包括临床结果)的方法。为了本发明的目的, 有益的或期望的临床结果包括但不限于以下的一种或多种: 减轻由疾病引起的一种或多种症状, 减轻疾病的程度, 稳定疾病(例如预防或延缓疾

病的恶化), 预防或延缓疾病的扩散(例如转移), 预防或延缓疾病的复发, 降低疾病的复发率, 延缓或减慢疾病的进展, 改善疾病状态, 提供疾病的(部分或全部)缓解, 减少治疗该疾病所需的一种或多种其他药物的剂量, 延缓疾病的进展, 提高生活质量, 和/或延长生存期。“治疗”还包括减少病症、病况或疾病的病理后果。本发明的方法考虑了这些治疗的方面中的任何一个或多个。

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、缓解、减轻和/或延迟其一种或多种症状)的化合物或组合物的量。如本领域中所理解的, “有效量”可以以一次或多次给药, 即, 可能需要单次给药或多此给药来达到期望的治疗终点。

“受试者”、“个体”或“患者”在本文中可互换使用, 以达到治疗目的, 是指任何归类为哺乳动物的动物, 包括人类、家畜和农场动物, 以及动物园、农场或宠物动物如狗、马、猫、牛等。在一些实施方案中, 所述个体是人类个体。

应理解, 本文所述的本发明的实施方案包括“由...组成”和/或“基本上由...组成”的实施方案。在本文中对“约”值或参数的提及包括(并描述了)针对该值或参数本身的变化。例如, 提及“大约 X”的描述, 包括对“X”的描述。

如本文所用, 对“不”值或参数的提及通常意指并描述了“除...外”值或参数。例如, 所述方法不用于治疗 X 型癌症, 意味着所述方法用于治疗除 X 型以外的癌症。

如本文所用, 术语“大约 X-Y”具有与“大约 X 至大约 Y”相同的含义。

如本文和所附权利要求书中所使用的, 单数形式“一个/一种(a/an)”和“所述”包括复数对象, 除非上下文另外明确指出。还应注意, 权利要求可以被撰写为排除任何可选的要素。因此此陈述旨在作为与权利要求要素的叙述结合使用诸如“只”、“仅”等排他性术语的先行基础, 或使用“否”的限制。

如本文所用, 术语“和/或”在词语诸如“A 和/或 B”中, 旨在既包括 A 和 B; A 或 B; A(单独); 以及 B(单独)。同样地, 如本文所用, 术语“和/或”在词语诸如“A、B 和/或 C”中, 旨在包括以下每个实施方案: 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); 以及 C(单独)。

II. Cas12i 核酸酶及效应蛋白

工程化的 Cas12i 核酸酶

在一些实施方式中, 提供了一种工程化的 Cas12i 核酸酶; 所述工程化的 Cas12i 核酸酶包含以下一种或多种(例如两种、三种、四种或五种)基于参比 Cas12i 核酸酶的突变, 所述突变选自:

(1)将参比 Cas12i 核酸酶中与 PAM 相互作用的氨基酸替换为带正电的氨基酸(如 R、H 或 K);

(2)将参比 Cas12i 核酸酶中参与打开 DNA 双链的氨基酸替换为带芳香环的氨基酸(如 F、Y 或 W);

5 (3)将参比 Cas12i 核酸酶中位于 RuvC 结构域并与单链 DNA 底物相互作用的氨基酸替换为带正电的氨基酸(如 R、H 或 K);

(4)将参比 Cas12i 核酸酶中与 DNA-RNA 双螺旋相互作用的氨基酸替换为带正电的氨基酸(如 R、H 或 K); 和

(5)将参比 Cas12i 核酸酶中与 DNA-RNA 双螺旋相互作用的一个或多个极性或带正电荷的氨基酸替换为疏水氨基酸(如 A、V、I、L、M、F、Y、P、C 或 W)。在一些实施方式中, 所述参比 Cas12i 核酸酶为天然 Cas12i 核酸酶, 比如氨基酸序列如 SEQ ID NO: 1 所示的野生型 Cas12i2 核酸酶。在一些实施方式中, 所述参比 Cas12i 核酸酶为 Cas12i 核酸酶的变体, 比如天然变体。在一些实施方式中, 所述参比 Cas12i 核酸酶为不包含上述
10 (1)-(5)中一种或多种突变的工程化的 Cas12i(例如 Cas12i2 或 Cas12i1)核酸酶。在一些实施方式中, 所述氨基酸位置如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。在一些实施方式中, 所述氨基酸位置如具有不同于 SEQ ID NO: 1 序列的野生型 Cas12i 核酸酶的相对应
15 SEQ ID NO: 1 的氨基酸位置所定义。

如本文所用,“该氨基酸在 X 位置处, 其中所述氨基酸位置如 SEQ ID NO: 1 所示的野生型 Cas12i 核酸酶的相应氨基酸位置所定义”, 或“该氨基酸在 X 位置处, 其中所述氨基酸位置如具有不同于 SEQ ID NO: 1 序列的野生型 Cas12i 核酸酶的相对应 SEQ ID NO: 1
20 的氨基酸位置所定义”的含义是: 该氨基酸残基位于参比酶 Cas12i 的某位置处, 其相当在 SEQ ID NO: 1 的 X 位置处, 而且参比酶 Cas12i 的氨基酸序列与 SEQ ID NO: 1 的氨基酸序列基于序列同源性相互对齐。例如, 图 7 示出了 CAS12i2(SEQ ID NO: 1)与 CAS12i1(SEQ ID NO: 13)的氨基酸序列的同源性比对。本领域人员可以用本领域常用的
25 软件, 如 Clustal Omega, 将任一参比 Cas12i 核酸酶的氨基酸序列与 SEQ ID NO: 1 进行序列同一性比较和对齐(alignment), 进而得到与本申请中所述基于 SEQ ID NO: 1 所定义的氨基酸位点相对应的所述参比 Cas12i 核酸酶中的氨基酸位点。

本申请提供了通过引入氨基酸突变来工程化改造酶的方法, 所述氨基酸突变基于上述五种改造原理的任何一种或多种的组合, 这导致体外和/或体内酶活性(如 DNA 单链或
30 双链切割活性)的增加(比如, 切割基因效率增加了约 100 倍)、可识别 PAM 数目的增加(比如, 从图 8 可以看出, 野生型的酶可以识别一种 PAM, 而基于上述氨基酸突变的

CasXX 酶可以识别至少 11 种 PAM)、和/或降低脱靶率及提高特异性(比如, 从表 18 可以看出, HF-Cas 相对于 CasXX 脱靶效率可以降低约 99%)。所述工程化 Cas12i 核酸酶含有一个或多个如以下 1)-6) 节中所描述的具体突变。在一些实施方式中, 本申请中所述的任何一个或多个突变可以与现有的 Cas12i 突变组合(例如以下 7) 节中所描述的突变), 以提供具有更高活性的工程化 Cas12i 核酸酶。

5 在一些实施方式中, 提供了一种工程化的 Cas12i 核酸酶; 所述工程化的 Cas12i 核酸酶包含将参比 Cas12i 核酸酶中与 PAM 相互作用的一个或多个氨基酸替换为带正电的氨基酸的突变。在一些实施方式中, 提供了一种工程化的 Cas12i 核酸酶; 所述工程化的 Cas12i 核酸酶包含将参比 Cas12i 核酸酶中参与打开 DNA 双链的一个或多个氨基酸替换为带芳香环的氨基酸。在一些实施方式中, 提供了一种工程化的 Cas12i 核酸酶; 所述工程化的 Cas12i 核酸酶包含将参比 Cas12i 核酸酶中位于 RuvC 结构域并与单链 DNA 底物相互作用的一个或多个氨基酸替换为带正电的氨基酸。在一些实施方式中, 提供了一种工程化的 Cas12i 核酸酶; 所述工程化的 Cas12i 核酸酶包含将参比 Cas12i 核酸酶中与 DNA-RNA 双螺旋相互作用的一个或多个氨基酸替换为带正电的氨基酸。在一些实施方式中, 提供了一种工程化的 Cas12i 核酸酶; 所述工程化的 Cas12i 核酸酶包含将参比 Cas12i 核酸酶中与 DNA-RNA 双螺旋相互作用的一个或多个极性

10 或带正电荷的氨基酸替换为疏水氨基酸。

15 在一些实施方式中, 提供了一种工程化的 Cas12i 核酸酶; 所述工程化的 Cas12i 核酸酶包含: 1) 将参比 Cas12i 核酸酶中与 PAM 相互作用的一个或多个氨基酸替换为带正电的氨基酸的突变; 和 2) 将参比 Cas12i 核酸酶中参与打开 DNA 双链的一个或多个氨基酸替换为带芳香环的氨基酸。在一些实施方式中, 提供了一种工程化的 Cas12i 核酸酶; 所述工程化的 Cas12i 核酸酶包含: 1) 将参比 Cas12i 核酸酶中与 PAM 相互作用的一个或多个氨基酸替换为带正电的氨基酸的突变; 和 2) 将参比 Cas12i 核酸酶中位于 RuvC 结构域并与单链 DNA 底物相互作用的一个或多个氨基酸替换为带正电的氨基酸。在一些实施方式中, 提供了一种工程化的 Cas12i 核酸酶; 所述工程化的 Cas12i 核酸酶包含: 1) 将参比 Cas12i 核酸酶中参与打开 DNA 双链的一个或多个氨基酸替换为带芳香环的氨基酸; 和 2) 将参比 Cas12i 核酸酶中位于 RuvC 结构域并与单链 DNA 底物相互作用的一个或多个氨基酸替换为带正电的氨基酸。在一些实施方式中, 所述工程化的 Cas12i 核酸酶还包含将参比 Cas12i 核酸酶中与 DNA-RNA 双螺旋相互作用的一个或多个极性

20 或带正电荷的氨基酸替换为疏水氨基酸。在一些实施方式中, 所述工程化的 Cas12i 核酸酶还包含一个或多个柔性区突变(如替换为 G, 和/或在其后插入一个或两个 G)以增加柔性区的柔性。

在一些实施方式中，经过如此修改的 Cas12i 工程化酶增加了至少约 10% 的柔性，比如增加了至少约 20%、30%、50%、100%、150%、200%、500%、1000% 或更高层次的柔性。

在一些实施方式中，提供了一种工程化的 Cas12i 核酸酶；所述工程化的 Cas12i 核酸酶包含：1) 将参比 Cas12i 核酸酶中与 PAM 相互作用的一个或多个氨基酸替换为带正电的氨基酸的突变；和 2) 将参比 Cas12i 核酸酶中与 DNA-RNA 双螺旋相互作用的一个或多个氨基酸替换为带正电的氨基酸。在一些实施方式中，提供了一种工程化的 Cas12i 核酸酶；所述工程化的 Cas12i 核酸酶包含：1) 将参比 Cas12i 核酸酶中参与打开 DNA 双链的一个或多个氨基酸替换为带芳香环的氨基酸；和 2) 将参比 Cas12i 核酸酶中与 DNA-RNA 双螺旋相互作用的一个或多个氨基酸替换为带正电的氨基酸。在一些实施方式中，提供了一种工程化的 Cas12i 核酸酶；所述工程化的 Cas12i 核酸酶包含：1) 将参比 Cas12i 核酸酶中位于 RuvC 结构域并与单链 DNA 底物相互作用的一个或多个氨基酸替换为带正电的氨基酸；和 2) 将参比 Cas12i 核酸酶中与 DNA-RNA 双螺旋相互作用的一个或多个氨基酸替换为带正电的氨基酸。在一些实施方式中，所述工程化的 Cas12i 核酸酶还包含将参比 Cas12i 核酸酶中与 DNA-RNA 双螺旋相互作用的一个或多个极性或带正电荷的氨基酸替换为疏水氨基酸。在一些实施方式中，所述工程化的 Cas12i 核酸酶还包含一个或多个柔性区突变以增加柔性区的柔性。

在一些实施方式中，提供了一种工程化的 Cas12i 核酸酶；所述工程化的 Cas12i 核酸酶包含：1) 将参比 Cas12i 核酸酶中与 PAM 相互作用的一个或多个氨基酸替换为带正电的氨基酸的突变；2) 将参比 Cas12i 核酸酶中参与打开 DNA 双链的一个或多个氨基酸替换为带芳香环的氨基酸；和 3) 将参比 Cas12i 核酸酶中位于 RuvC 结构域并与单链 DNA 底物相互作用的一个或多个氨基酸替换为带正电的氨基酸。在一些实施方式中，提供了一种工程化的 Cas12i 核酸酶；所述工程化的 Cas12i 核酸酶包含：1) 将参比 Cas12i 核酸酶中与 PAM 相互作用的一个或多个氨基酸替换为带正电的氨基酸的突变；2) 将参比 Cas12i 核酸酶中参与打开 DNA 双链的一个或多个氨基酸替换为带芳香环的氨基酸；和 3) 将参比 Cas12i 核酸酶中与 DNA-RNA 双螺旋相互作用的一个或多个氨基酸替换为带正电的氨基酸。在一些实施方式中，提供了一种工程化的 Cas12i 核酸酶；所述工程化的 Cas12i 核酸酶包含：1) 将参比 Cas12i 核酸酶中与 PAM 相互作用的一个或多个氨基酸替换为带正电的氨基酸的突变；2) 将参比 Cas12i 核酸酶中位于 RuvC 结构域并与单链 DNA 底物相互作用的一个或多个氨基酸替换为带正电的氨基酸；和 3) 将参比 Cas12i 核酸酶中与 DNA-RNA 双螺旋相互作用的一个或多个氨基酸替换为带正电的氨基酸。在一些实施方式中，提供了一种工程化的 Cas12i 核酸酶；所述工程化的 Cas12i 核酸酶包

含：1) 将参比 Cas12i 核酸酶中参与打开 DNA 双链的一个或多个氨基酸替换为带芳香环的氨基酸；2) 将参比 Cas12i 核酸酶中位于 RuvC 结构域并与单链 DNA 底物相互作用的一个或多个氨基酸替换为带正电的氨基酸；和 3) 将参比 Cas12i 核酸酶中与 DNA-RNA 双螺旋相互作用的一个或多个氨基酸替换为带正电的氨基酸。在一些实施方式中，
5 所述工程化的 Cas12i 核酸酶还包含将参比 Cas12i 核酸酶中与 DNA-RNA 双螺旋相互作用的一个或多个极性或带正电荷的氨基酸替换为疏水氨基酸。在一些实施方式中，所述工程化的 Cas12i 核酸酶还包含一个或多个柔性区突变以增加柔性区的柔性。

在一些实施方式中，提供了一种工程化的 Cas12i 核酸酶；所述工程化的 Cas12i 核酸酶包含：1) 将参比 Cas12i 核酸酶中与 PAM 相互作用的一个或多个氨基酸替换为带正电的氨基酸的突变；2) 将参比 Cas12i 核酸酶中参与打开 DNA 双链的一个或多个氨基酸替换为带芳香环的氨基酸；3) 将参比 Cas12i 核酸酶中位于 RuvC 结构域并与单链 DNA 底物相互作用的一个或多个氨基酸替换为带正电的氨基酸；和 4) 将参比 Cas12i 核酸酶中与 DNA-RNA 双螺旋相互作用的一个或多个氨基酸替换为带正电的氨基酸。在一些实施方式中，所述工程化的 Cas12i 核酸酶还包含将参比 Cas12i 核酸酶中与 DNA-RNA 双螺旋相互作用的一个或多个极性或带正电荷的氨基酸替换为疏水氨基酸。在一些实施方式中，提供了一种工程化的 Cas12i 核酸酶；所述工程化的 Cas12i 核酸酶包含：1) 将参比 Cas12i 核酸酶中与 PAM 相互作用的一个或多个氨基酸替换为带正电的氨基酸的突变；2) 将参比 Cas12i 核酸酶中参与打开 DNA 双链的一个或多个氨基酸替换为带芳香环的氨基酸；3) 将参比 Cas12i 核酸酶中位于 RuvC 结构域并与单链 DNA 底物相互作用的一个或多个氨基酸替换为带正电的氨基酸；4) 将参比 Cas12i 核酸酶中与 DNA-RNA 双螺旋相互作用的一个或多个氨基酸替换为带正电的氨基酸；和 5) 将参比 Cas12i 核酸酶中与 DNA-RNA 双螺旋相互作用的一个或多个极性或带正电荷的氨基酸替换为疏水氨基酸。在一些实施方式中，所述工程化的 Cas12i 核酸酶还包含一个或多个柔性区突变以增加柔性区的柔性。

25 在一些实施方式中，所述工程化的 Cas12i 核酸酶包含相对于 SEQ ID NO: 1 所示相应氨基酸位置的突变：N164Y+E176R+K238R+E323R+D362R +T447R+E563R (自此命名为“CasXX”)。在一些实施方式中，所述工程化的 Cas12i 核酸酶包含 SEQ ID NO: 8 的序列。

1) 将参比 Cas12i 核酸酶中与 PAM 相互作用的氨基酸替换为带正电的氨基酸

30 在一些实施方式中，所述工程化的 Cas12i 核酸酶包含一个或多个基于参比 Cas12i 核酸酶(例如 Cas12i2)的突变，所述突变为将参比 Cas12i 核酸酶中与 PAM 相互作用的氨基酸替换为带正电的氨基酸(如 R、H 或 K)。在一些实施方式中，所述工程化的 Cas12i 核酸

酶包含一个、两个、三个、四个、五个、六个、七个、八个或更多所述氨基酸残基的替换。

在一些实施方式中,所述与 PAM 相互作用的氨基酸是与 PAM 在三维结构上距离在 9 埃以内的氨基酸,例如可以为:与 PAM 在三维结构上距离在 9 埃以内的氨基酸、与 PAM 在三维结构上距离在 8 埃以内的氨基酸、与 PAM 在三维结构上距离在 7 埃以内的氨基酸、与 PAM 在三维结构上距离在 6 埃以内的氨基酸、与 PAM 在三维结构上距离在 5 埃以内的氨基酸、与 PAM 在三维结构上距离在 4 埃以内的氨基酸、与 PAM 在三维结构上距离在 3 埃以内的氨基酸、与 PAM 在三维结构上距离在 2 埃以内的氨基酸、或更近的氨基酸。

10 PAM 与氨基酸的在空间结构上的距离由被解析的 Cas 蛋白-RNA-DNA 三维复合物的 3D 结构(PDB 文件)中的原子之间的距离界定,原子之间的相互距离可以通过 PDB 文件识别软件显示。在一些实施方式中,所述 PAM 与氨基酸在空间结构上的距离由氨基酸残基和核苷酸包含的原子之间的最小距离界定。可以用来测量 PAM 与氨基酸在空间结构上的距离的程序,或 PDB 文件识别软件在本领域广为人知,包括但不限于 PyMOL、
15 ChimeraX、Swiss-pdbviewer 等。

在一些实施方式中,所述一个或多个将参比 Cas12i 核酸酶中与 PAM 相互作用的氨基酸替换为带正电的氨基酸的突变是在下述位置的一个或多个氨基酸的突变: 176、178、226、227、229、237、238、264、447 和 563。在一些实施方式中,所述一个或多个将参比 Cas12i 核酸酶中与 PAM 相互作用的氨基酸替换为带正电的氨基酸的突变是在下述一个或多个氨基酸的突变: E176、E178、Y226、A227、N229、E237、K238、K264、T447 和 E563。在一些实施方式中,所述一个或多个将参比 Cas12i 核酸酶中与 PAM 相互作用的氨基酸替换为带正电的氨基酸的突变是在下述一个或多个氨基酸的突变: E176、K238、T447 和 E563。在一些实施方式中,所述将参比 Cas12i 核酸酶中与 PAM 相互作用的氨基酸替换为带正电的氨基酸的突变位于 563 号氨基酸残基,例如 E563。在一些实施方式中,所述氨基酸位置如 SEQ ID NO: 1 所示的野生型 Cas12i 核酸酶的相应氨基酸位置所定义。

在一些实施方式中,所述工程化的 Cas12i 核酸酶包含下述一个或多个氨基酸的突变: E176R、E178R、Y226R、A227R、N229R、E237R、K238R、K264R、T447R 和 E563R,所述氨基酸位置如 SEQ ID NO: 1 所示的野生型 Cas12i 核酸酶的相应氨基酸位置所定义。

30 在本说明书的上下文中, E176 的含义是:在所援引的氨基酸序列中(如相对于 SEQ ID NO: 1), 第 176 号氨基酸 E(谷氨酸);在此,常见的氨基酸及其三字母和单字母缩写列

举性地说明如下： 丙氨酸 Ala A；精氨酸 Arg R；天冬氨酸 Asp D；半胱氨酸 Cys C；谷氨酰胺 Gln Q；谷氨酸 Glu E；组氨酸 His H；异亮氨酸 Ile I；甘氨酸 Gly G；天冬酰胺 Asn N；亮氨酸 Leu L；赖氨酸 Lys K；甲硫氨酸 Met M；苯丙氨酸 Phe F；脯氨酸 Pro P；丝氨酸 Ser S；苏氨酸 Thr T；色氨酸 Trp W；酪氨酸 Tyr Y；缬氨酸 Val V。

5 在一些实施方式中,所述将参比 Cas12i 核酸酶中与 PAM 相互作用的氨基酸替换为带正电的氨基酸的突变为将参比 Cas12i 核酸酶中相应的氨基酸残基替换为 R、H 或 K,例如 R 或 K。在一些实施方式中,所述将参比 Cas12i 核酸酶中与 PAM 相互作用的氨基酸替换为带正电的氨基酸的突变为将参比 Cas12i 核酸酶中相应的氨基酸残基替换为 R。

10 在一些实施方式中,所述工程化的 Cas12i 核酸酶包含以下位点一个或多个氨基酸突变: 176R、238R、447R 和 563R,其中,所述氨基酸位置如 SEQ ID NO: 1 所示的野生型 Cas12i 核酸酶的相应氨基酸位置所定义。在一些实施方式中,所述工程化的 Cas12i 核酸酶包含以下一个或多个基于参比 Cas12i 核酸酶的突变: E176、K238、T447 和 E563;其中,氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。在一些实施方式中,所述工程化的 Cas12i 核酸酶包含以下一个或多个基于参比 Cas12i 核酸酶的突变:
15 K238R、T447R 和 E563R;其中,氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。在一些实施方式中,所述工程化的 Cas12i 核酸酶包含 E563R 突变;其中,氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。在一些实施方式中,为改善基因编辑效率的目的,也可以使用与上述经工程化的 Cas12i 核酸酶(将参比 Cas12i 核酸酶中与 PAM 相互作用的一个或多个氨基酸替换为带正电的氨基酸)具有至少约 85% (例如任一至少约 86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、
20 97%、98%、99%)序列一致性的工程化的 Cas12i 核酸酶。

在本说明书的上下文中, E176R 表示的含义是,在所援引的氨基酸序列中,将第 176 号氨基酸 E(谷氨酸)替换为 R(精氨酸)。

25 在一些实施方式中,所述工程化 Cas12i 核酸酶包含任一个在下述氨基酸残基位置的突变或突变组合: (i)176、238、264、447、563、176+238、176+447、176+563、238+447、238+563、447+563、176+238+447、176+238+563、176+447+563、238+447+563、176+238+447+563;其中,氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。在一些实施方式中,所述将参比 Cas12i 核酸酶中与 PAM 相互作用的一个或多个氨基酸替换为带正电的氨基酸包括替换为 R、H 或 K,例如 R 或 K,优选的,例如 R。在
30 一些实施方式中,所述工程化 Cas12i 核酸酶包含任一个在下述氨基酸残基的突变或突变组合: 在一些实施方式中,所述工程化 Cas12i 核酸酶包含任一个在下述氨基酸残基的突

变或突变组合：E176、K238、E264、T447、E563、E176+K238、E176+T447、E176+E563、K238+T447、K238+E563、T447+E563、E176+K238+T447、E176+K238+E563、E176+T447+E563、K238+T447+E563、和 E176+K238+T447+E563；其中，氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

5 在一些实施方式中，所述工程化 Cas12i 核酸酶包含下述突变/突变组合的任一个：E176R、K238R、E264R、T447R、E563R、E176R+K238R、E176R+T447R、E176R+E563R、K238R+T447R、K238R+E563R、T447R+E563R、E176R+K238R+T447R、E176R+K238R+ E563R、E176R+T447R+E563R、K238R+T447R+E563R、和 E176R+K238R+ T447R+E563R；其中，氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。在一些实施方式中，所述工程化 Cas12i 核酸酶包含下述突变/突变组合的任一个：E563R、E176R+T447R、E176R+E563R、K238R+E563R、E176R+K238R+T447R、E176R+K238R+E563R、E176R+T447R+E563R、和 E176R+K238R+T447R+E563R；其中，氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。为改善基因编辑效率的目的，也可以使用与上述经工程化的 Cas12i 核酸酶(将参比 Cas12i 核酸酶中与 PAM 相互作用的一个或多个氨基酸替换为带正电的氨基酸)具有至少约 85%(例如任一至少约 86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%)序列一致性的工程化的 Cas12i 核酸酶。

20

2) 将参比 Cas12i 核酸酶中参与打开 DNA 双链的氨基酸替换为带芳香环的氨基酸

在一些实施方式中，所述工程化的 Cas12i 核酸酶包含一个或多个基于参比 Cas12i 核酸酶(例如 Cas12i2)的突变，所述突变为将参比 Cas12i 核酸酶中参与打开 DNA 双链的一个或多个氨基酸替换为带芳香环的氨基酸(如 F、Y 或 W)。在一些实施方式中，所述工程化的 Cas12i 核酸酶包含一个、两个、三个、四个、五个、六个、或更多所述氨基酸残基的替换。

其中，所述参与打开 DNA 双链的一个或多个氨基酸是与 PAM 中相对于靶标链的 3' 端最后一个碱基对相互作用的氨基酸。例如，Cas12i2 识别的 PAM 序列是 5'-NTTN-3' 碱基对，其中 PAM 序列中 3' 末端的 N 碱基与靶标链所形成的碱基对就是文本所述的“与 PAM 中相对于靶标链的 3' 端最后一个碱基对，”而该碱基对之后就是靶向位点的序列。

30 在一些实施方式中，所述一个或多个参与打开 DNA 双链的氨基酸位于下述位置：

163 和/或 164; 其中氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。在一些实施方式中, 所述一个或多个参与打开 DNA 双链的氨基酸是下述一个或多个氨基酸: Q163、N164; 其中氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。在一些实施方式中, 所述参与打开 DNA 双链的氨基酸为 N164; 其中氨基酸位置编号如
5 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

在一些实施方式中, 所述参与打开 DNA 双链的氨基酸被替换为 F、Y 或 W。在一些实施方式中, 所述参与打开 DNA 双链的氨基酸被替换为 F。在一些实施方式中, 所述参与打开 DNA 双链的氨基酸被替换为 Y。

10 在一些实施方式中, 所述工程化 Cas12i 核酸酶包含下述任一个或多个氨基酸残基的突变: 163F、163Y、163W、164W、164F 或 164Y; 其中氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。在一些实施方式中, 所述工程化 Cas12i 核酸酶包含下述任一个或多个氨基酸残基的突变: 163F、163Y、163W、164F 或 164Y; 其中氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

15 在一些实施方式中, 所述工程化 Cas12i 核酸酶包含下述任一个突变: Q163 和/或 N164; 其中氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。在一些实施方式中, 所述工程化 Cas12i 核酸酶包含下述任一个突变: Q163F、Q163Y、Q163W、N164W、N164F 或 N164Y; 其中氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。在一些实施方式中, 所述工程化 Cas12i 核酸酶包含下述任一个突变: Q163F、Q163Y、Q163W、N164F 或 N164Y; 其中氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置
20 所定义。在一些实施方式中, 所述工程化 Cas12i 核酸酶包含 N164Y 或 N164F 突变; 其中氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。在一些实施方式中, 所述工程化 Cas12i 核酸酶包含 N164Y 突变; 其中氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。在一些实施例中, 为实现改善基因编辑效率的目的, 也可以使用与上述工程化的 Cas12i 核酸酶(将参比 Cas12i 核酸酶中参与打开 DNA 双链的一个或
25 多个氨基酸替换为带芳香环的氨基酸)具有至少约 85%(例如任一至少约 86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%)序列一致性的工程化的 Cas12i 核酸酶。

30 3) 将参比 Cas12i 核酸酶中位于 RuvC 结构域并与单链 DNA 底物相互作用的氨基酸替换为带正电的氨基酸

在一些实施方式中, 所述工程化的 Cas12i 核酸酶包含一个或多个基于参比 Cas12i 核

酸酶(例如 Cas12i2)的突变, 所述突变为将参比 Cas12i 酶中位于 RuvC 结构域并与单链 DNA 底物相互作用的氨基酸替换为带正电的氨基酸(如 R、H 或 K)。在一些实施方式中, 所述工程化的 Cas12i 酶包含一个、两个、三个、四个、五个、六个或更多所述氨基酸残基的替换。

5 其中, 所述位于 RuvC 结构域并与单链 DNA 底物相互作用的一个或多个氨基酸是在三维结构上与单链 DNA 底物距离在 9 埃以内的氨基酸, 例如可以为: 在三维结构上与单链 DNA 底物距离在 8 埃以内的氨基酸、在三维结构上与单链 DNA 底物距离在 7 埃以内的氨基酸、在三维结构上与单链 DNA 底物距离在 6 埃以内的氨基酸、在三维结构上与单链 DNA 底物距离在 5 埃以内的氨基酸、在三维结构上与单链 DNA 底物距离在 4 埃以内的氨基酸、在三维结构上与单链 DNA 底物距离在 3 埃以内的氨基酸、在三维结构上与单链 DNA 底物距离在 2 埃以内的氨基酸、或更近的氨基酸。

RuvC 结构域是 Cas12i 蛋白中负责切割单链 DNA 或者双链 DNA 的酶活结构域。在蛋白质的一级序列中, Cas12i 的 RuvC 结构域被分为 3 个部分: RuvC-1, RuvC-2 以及 RuvC-3。这 3 个部分在三维结构中相邻近, 一同组成具有酶切活性的催化口袋。Cas12i2 15 的三维晶体结构, 其结构域组成, 及与 DNA 底物相互作用的描述见 Huang X. *et al.*, *Nature Communications*, 11, Article number: 5241 (2020)。Cas12i1 的三维晶体结构, 其结构域组成, 及与 DNA 底物相互作用的描述见 Zhang H. *et al. Nature Structural & Molecular Biology* 27, 1069-1076(2020)。可以通过同源结构比较和建模(homology modeling)通过已知的 Cas12i 三维晶体结构, 得到参比 Cas12i 和底物相互作用的三维结构模型。实施例 3 20 中描述了一种建模方式以得到 Cas12i2 中位于 RuvC 结构域并与单链 DNA 底物距离在 9 埃以内的氨基酸。

在一些实施方式中, RuvC 结构域中的氨基酸与单链 DNA 底物在空间结构上的距离可以由被解析的 Cas 蛋白-RNA-DNA 三维复合物的 3D 结构(PDB 文件)中的原子之间的距离界定, 原子之间的相互距离可以通过 PDB 文件识别软件显示。在一些实施方式中, 所 25 述 RuvC 结构域中的氨基酸与单链 DNA 底物在空间结构上的距离由氨基酸残基和核苷酸包含的原子之间的最小距离界定。可以用来测量 RuvC 结构域中的氨基酸与单链 DNA 底物在空间结构上的距离的程序, 或 PDB 文件识别软件在本领域广为人知, 包括但不限于 PyMOL、ChimeraX、Swiss-pdbviewer 等。

在一些实施方式中, 所述位于 RuvC 结构域并与单链 DNA 底物相互作用的一个或多个 30 氨基酸是下述位置的一个或多个氨基酸: 323、327、355、359、360、361、362、388、390、391、392、393、414、417、418、421、424、425、650、652、653、696、705、

708、709、751、752、755、840、848、851、856、885、897、925、926、928、929、932、和 1022。在一些实施方式中，所述位于 RuvC 结构域并与单链 DNA 底物相互作用的一个或多个氨基酸是下述一个或多个氨基酸：E323、L327、V355、G359、G360、K361、D362、L388、N390、N391、F392、K393、Q414、L417、L418、K421、Q424、Q425、S650、E652、G653、I696、K705、K708、E709、L751、S752、E755、N840、N848、S851、A856、Q885、M897、N925、I926、T928、G929、Y932、A1022。在一些实施方式中，所述位于 RuvC 结构域并与单链 DNA 底物相互作用的一个或多个氨基酸是下述一个或多个氨基酸：E323、D362、L388、N391、L417、Q424、Q425、N925、I926、和 G929。在一些实施方式中，所述氨基酸位置如 SEQ ID NO: 1 所示的野生型 Cas12i 核酸酶的相应氨基酸位置所定义。

在一些实施方式中，所述工程化的 Cas12i 核酸酶包含将参比 Cas12i 核酸酶中位于 RuvC 结构域并与单链 DNA 底物相互作用的一个或多个氨基酸替换为 R、H 或 K(例如 R 或 K)的突变。在一些实施方式中，所述工程化的 Cas12i 核酸酶包含将参比 Cas12i 核酸酶中位于 RuvC 结构域并与单链 DNA 底物相互作用的一个或多个氨基酸替换为 R 的突变。

在一些实施方式中，所述工程化的 Cas12i 核酸酶包含下述一个或多个氨基酸突变或突变组合：N390R、N391R、F392R、L751R、E755R、N840R、N848R、S851R、A856R、Q885R、M897R、I926R、G929R、Y932R、E323R、L327R、V355R、G359R、G360R、K361R、D362R、Q414R、K421R、Q425R、S650R、E652R、K705R、K708R、E709R、S752R、N925R、T928R、E323R+D362R、E323R+Q425R、E323R+I926R、Q425R+I926R、E323R+D362R+Q425R、E323R+D362R+I926R、E323R+Q425R+I926R、E323R+D362R+Q425R+I926R、D362R+I926R、N925R+I926R、D362R+N925R+I926R、D362R+N925R；所述氨基酸位置如 SEQ ID NO: 1 所示的野生型 Cas12i 核酸酶的相应氨基酸位置所定义。

在一些实施方式中，所述工程化 Cas12i 核酸酶包含任一个下述氨基酸残基位置的突变或突变组合：E323、D362、Q425、N925、I926 和 G929；其中，氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。在一些实施方式中，所述工程化 Cas12i 核酸酶包含任一个下述氨基酸残基位置的突变或突变组合：E323、D362、Q425、N925、I926、E323+D362、E323+Q425、E323+I926、D362+Q425、D362+N925、D362+I926、Q425+I926、N925+I926、E323+D362+Q425、E323+D362+I926、E323+Q425+I926、D362+N925+I926、D362+Q425+I926、E323+D362+Q425+I926；其中，氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。在一些实施方式中，所述突变为将所述

位置氨基酸残基替换为 R、H 或 K(如 R)的突变。在一些实施方式中，所述工程化 Cas12i 核酸酶包含下述任何一个氨基酸或氨基酸组合：323R、362R、425R、925R、926R、323R+362R、323R+425R、323R+926R、362R+425R、362R+926R、425R+926R、925R+926R、323R+362R+425R、323R+362R+926R、323R+425R+926R、5 362R+925R+926R、323R+362R+425R+926R、和 362R+425R+926R；其中，氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。在一些实施方式中，所述工程化 Cas12i 核酸酶包含任一个下述突变或突变组合：E323R、D362R、Q424R、Q425R、N925R、I926R、和 G929R；其中，氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。在一些实施方式中，所述工程化 Cas12i 核酸酶包含任一个下述突变或突变组合：E323R、D362R、Q425R、N925R、I926R、E323R+D362R、E323R+Q425R、E323R+I926R、D362R+Q425R、Q425R+I926R、D362R+I926R、N925R+I926R、E323R+D362R+Q425R、E323R+D362R+I926R、E323R+Q425R+I926R、D362R+N925R+I926R、D362R+Q425R+I926R、和 E323R+D362R+Q425R+I926R；其中，氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。在一些实施方式中，所述工程化 Cas12i 核酸酶包含 I926R 突变；其中，氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。在一些实施方式中，所述工程化 Cas12i 核酸酶包含 E323R+D362R 突变；其中，氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。在一些实施例中，为实现改善基因编辑效率的目的，也可以使用与上述工程化的 Cas12i 核酸酶(将参比 Cas12i 核酸酶位于 RuvC 结构域并与单链 DNA 底物相互作用的一个或多个氨基酸替换为带正电的氨基酸)具有至少约 85%(例如任一至少约 86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%)序列一致性的工程化的 Cas12i 核酸酶。

25 4) 将参比 Cas12i 核酸酶中与 DNA-RNA 双螺旋相互作用的一个或多个氨基酸替换为带正电的氨基酸

在一些实施方式中，所述工程化的 Cas12i 核酸酶包含一个或多个基于参比 Cas12i 核酸酶(例如 Cas12i2)的突变，所述突变为将参比 Cas12i 核酸酶中与 DNA-RNA 双螺旋相互作用的一个或多个氨基酸替换为带正电的氨基酸(如 R、H 或 K)。在一些实施方式中，所述工程化的 Cas12i 酶包含一个、两个、三个、四个、五个、六个、或更多所述氨基酸残基的替换。

其中，所述与 DNA-RNA 双螺旋相互作用的一个或多个氨基酸是在三维结构上与

DNA-RNA 双螺旋距离在 9 埃以内的氨基酸，例如可以为：在三维结构上与 DNA-RNA 双螺旋距离在 8 埃以内的氨基酸、在三维结构上与 DNA-RNA 双螺旋距离在 7 埃以内的氨基酸、在三维结构上与 DNA-RNA 双螺旋距离在 6 埃以内的氨基酸、在三维结构上与 DNA-RNA 双螺旋距离在 5 埃以内的氨基酸、在三维结构上与 DNA-RNA 双螺旋距离在 4 埃以内的氨基酸、在三维结构上与 DNA-RNA 双螺旋距离在 3 埃以内的氨基酸、在三维结构上与 DNA-RNA 双螺旋距离在 2 埃以内的氨基酸、或更近的氨基酸。某些 Cas 核酸酶的工作原理如下：Cas 与指导 RNA(如 crRNA)形成复合体，其中 crRNA 和靶向 DNA 相互配对形成的 DNA-RNA 双螺旋，并与 Cas 核酸酶相互作用，打开双链靶向 DNA，并形成 R-loop，得以使 Cas 的酶切活性位点完成对 dsDNA 的切割。Cas12i2 的三维晶体结构，其结构域组成，及与 DNA-RNA 双螺旋相互作用的描述见 Huang X. *et al.*, Nature Communications, 11, Article number: 5241 (2020)。

在一些实施方式中，DNA-RNA 双螺旋和 Cas 氨基酸在空间结构上的距离可以由被解析的 Cas 蛋白-RNA-DNA 三维复合物的 3D 结构(PDB 文件)中的氨基酸残基和核苷酸包含的原子之间的最小距离界定，原子之间的相互距离可以通过 PDB 文件识别软件显示。在一些实施方式中，所述 Cas 氨基酸与 DNA-RNA 双螺旋在空间结构上的距离由原子假设位置的距离界定。可以用来测量 Cas 氨基酸与 DNA-RNA 双螺旋在空间结构上的距离的程序，或 PDB 文件识别软件在本领域广为人知，包括但不限于 PyMOL、ChimeraX、Swiss-pdbviewer 等。

在一些实施方式中，所述与 DNA-RNA 双螺旋相互作用的一个或多个氨基酸是下述位置的一个或多个氨基酸：116、117、156、159、160、161、247、293、294、297、301、305、306、308、312、313、316、319、320、343、348、349、427、433、438、441、442、679、683、691、782、783、797、800、852、853、855、861、865、957、958。在一些实施方式中，所述与 DNA-RNA 双螺旋相互作用的一个或多个氨基酸是下述一个或多个氨基酸：G116、E117、A156、T159、E160、S161、E247、G293、E294、N297、T301、I305、K306、T308、N312、F313、Q316、E319、Q320、E343、E348、E349、D427、K433、V438、N441、Q442、N679、E683、E691、D782、E783、E797、E800、M852、D853、L855、N861、Q865、S957、D958。在一些实施方式中，所述与 DNA-RNA 双螺旋相互作用的一个或多个氨基酸是下述一个或多个氨基酸：G116、E117、T159、S161、E319、E343、或 D958。在一些实施方式中，所述与 DNA-RNA 双螺旋相互作用的氨基酸是 D958。在一些实施方式中，所述氨基酸位置如 SEQ ID NO: 1 所示的野生型 Cas12i 核酸酶的相应氨基酸位置所定义。

在一些实施方式中，所述工程化的 Cas12i 核酸酶包含将参比 Cas12i 核酸酶中与 DNA-RNA 双螺旋相互作用的一个或多个氨基酸替换为 R、H 或 K(例如 R 或 K)的突变。在一些实施方式中，所述工程化的 Cas12i 核酸酶包含将参比 Cas12i 核酸酶中参与 DNA-RNA 双螺旋相互作用的一个或多个氨基酸替换为为 R 的突变。

5 在一些实施方式中，所述工程化的 Cas12i 核酸酶包含下述一个或多个氨基酸突变：G116R、E117R、A156R、T159R、S161R、T301R、I305R、K306R、T308R、N312R、F313R、D427R、K433R、V438R、N441R、Q442R、M852R、L855R、N861R、Q865R、E160R、Q316R、E319R、Q320R、E247R、E343R、E348R、E349R、N679R、E683R、E691R、D782R、E783R、E797R、E800R、D853R、S957R、D958R、G293R、E294R、
10 N297R；所述氨基酸位置如 SEQ ID NO: 1 所示的野生型 Cas12i 核酸酶的相应氨基酸位置所定义。

在一些实施方式中，所述工程化 Cas12i 核酸酶包含任一个下述氨基酸残基位置的突变或突变组合：G116、E117、T159、S161、E319、E343、或 D958；其中，氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。在一些实施方式中，所述突变为将
15 所述位置氨基酸残基替换为 R、H 或 K(如 R)的突变。在一些实施方式中，所述工程化 Cas12i 核酸酶包含下述任何一个位点氨基酸或氨基酸突变组合：116R、117R、159R、161R、319R、343R、或 958R；其中，氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。在一些实施方式中，所述工程化 Cas12i 核酸酶包含任一个下述突变或突变组合：G116R、E117R、T159R、S161R、E319R、E343R、或 D958R；其中，氨基酸
20 位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。在一些实施方式中，所述工程化 Cas12i 核酸酶包含 D958R 突变；其中，氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。在一些实施例，为实现改善基因编辑效率的目的，也可以使用与上述工程化的 Cas12i 核酸酶具有至少约 85%(例如任一至少约 86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%)序列一致性的工程化的
25 Cas12i 核酸酶。

5) 将参比 Cas12i 核酸酶中与 DNA-RNA 双螺旋相互作用的一个或多个极性 or 带正电荷的氨基酸替换为疏水氨基酸

30 在一些实施方式中，所述工程化的 Cas12i 核酸酶包含一个或多个基于参比 Cas12i 核酸酶 (例如 Cas12i2) 的突变，所述突变为将参比 Cas12i 核酸酶中与 DNA-RNA 双螺旋相互作用的一个或多个极性 or 带正电荷的氨基酸替换为疏水氨基酸(如 A、V、I、L、M、F、

Y、P、C 或 W)。所述突变(在此也称为“高特异性”或“HF”突变)可以降低 Cas12i 核酸酶的脱靶率(即提高特异性)。在一些实施方式中, CasXX-HF 相对于 CasXX 脱靶率可以降低至少约 10%、20%、30%、40%、50%、60%、70%、80%、90%、95%、99% 或降低更多。在一些实施方式中, 所述工程化的 Cas12i 酶包含一个、两个、三个、四个、五个、六个
5 或更多所述氨基酸残基的替换。

其中, 所述与 DNA-RNA 双螺旋相互作用的一个或多个氨基酸是在三维结构上与 DNA-RNA 双螺旋距离在 9 埃以内的氨基酸, 例如可以为: 在三维结构上与 DNA-RNA 双螺旋距离在 8 埃以内的氨基酸、在三维结构上与 DNA-RNA 双螺旋距离在 7 埃以内的氨基酸、在三维结构上与 DNA-RNA 双螺旋距离在 6 埃以内的氨基酸、在三维结构上与
10 DNA-RNA 双螺旋距离在 5 埃以内的氨基酸、在三维结构上与 DNA-RNA 双螺旋距离在 4 埃以内的氨基酸、在三维结构上与 DNA-RNA 双螺旋距离在 3 埃以内的氨基酸、在三维结构上与 DNA-RNA 双螺旋距离在 2 埃以内的氨基酸、或更近的氨基酸。Cas12i2 的三维晶体结构, 其结构域组成, 及与 DNA-RNA 双螺旋相互作用的描述见 Huang X. *et al.*, Nature Communications, 11, Article number: 5241 (2020)。

15 在一些实施方式中, 所述与 DNA-RNA 双螺旋相互作用的一个或多个极性 or 带正电荷的氨基酸是下述位置的一个或多个氨基酸: 119、164、297、308、309、312、346、357、394、395、402、441、433、565、715、719、766、782、807、841、844、845、848、857、861 和 865。在一些实施方式中, 所述与 DNA-RNA 双螺旋相互作用的一个或多个极性 or 带正电荷的氨基酸是下述一个或多个氨基酸: Y119、Y164、N297、T308、
20 R309、N312、S346、H357、K394、E395、R402、N441、K433、S565、R715、R719、S766、D782、K807、N841、K844、K845、N848、R857、N861 和 Q865。在一些实施方式中, 所述与 DNA-RNA 双螺旋相互作用的一个或多个极性 or 带正电荷的氨基酸是下述一个或多个氨基酸: R857、R861、K807、N848、R715、R719、K394、H357 和 K844。在一些实施方式中, 所述与 DNA-RNA 双螺旋相互作用的一个或多个极性 or 带正电荷的氨基酸是下述一个或多个氨基酸: R857、R719、K394、和 K844。其中, 如上所述氨基酸位置如 SEQ ID NO: 1 或 8 所示的相应氨基酸位置所定义。
25

在一些实施方式中, 所述工程化的 Cas12i 核酸酶包含下述一个或多个突变: S565A、N297A、Q865A、T308A、R309A、N312A、N441A、R857A、N861A、Y119F、K433A、K807A、N841A、N848A、K845A、D782A、R715A、R719A、S766A、K394A、H357A、
30 K844A、E395A、S346A、R402A、Y164N; 其中, 氨基酸位置编号如 SEQ ID NO: 1 或 8 所示的相应氨基酸位置所定义。

在一些实施方式中，所述工程化的 Cas12i 核酸酶包含将参比 Cas12i 核酸酶中位于 357、394、715、719、807、844、848、857、861 的一个或多个极性 or 带正电荷的氨基酸替换为疏水氨基酸的突变。在一些实施方式中，所述疏水氨基酸选自 A、V、L、I、P 和 F，例如 A、V、L 或 I。在一些实施方式中，所述工程化的 Cas12i 核酸酶包含将参比 Cas12i 核酸酶中位于 H357、K394、R715、R719、K807、K844、N848、R857 和/或 R861 的一个或多个极性 or 带正电荷的氨基酸替换为 A 的突变。其中，所述氨基酸位置编号如 SEQ ID NO: 1 或 8 所示的相应氨基酸位置所定义。在一些实施方式中，所述工程化 Cas12i 核酸酶包含下述任何一个位置氨基酸或氨基酸组合的突变：H357A、K394A、R715A、R719A、K807A、K844A、N848A、R857A、和 R861A；其中，氨基酸位置编号如 SEQ ID NO: 1 或 8 所示的相应氨基酸位置所定义。

在一些实施方式中，所述工程化 Cas12i 核酸酶包含任一个下述氨基酸残基位置的突变或突变组合：857、719、394、或 844；其中，氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。在一些实施方式中，所述突变为将所述位置氨基酸残基替换为疏水氨基酸(如 A)的突变。在一些实施方式中，所述工程化 Cas12i 核酸酶包含下述任何一个位置氨基酸或氨基酸组合的突变：R857、R719、K394、或 K844；其中，氨基酸位置编号如 SEQ ID NO: 1 或 8 所示的相应氨基酸位置所定义。在一些实施方式中，所述工程化 Cas12i 核酸酶包含任一个下述氨基酸残基位置的突变或突变组合：R857、R719、K394、K844、R719+K394、K394+K844、R857+K394、R719+K844、R857+R719、R857+K844、R719+K394+K844、R857+R719+K394、R857+K394+K844、R857+R719+K844、R857+R719+K394+K844；其中，氨基酸位置编号如 SEQ ID NO: 1 或 8 所示的相应氨基酸位置所定义；所述突变为将所述位置氨基酸残基替换为疏水氨基酸(如 A)的突变。在一些实施方式中，所述工程化 Cas12i 核酸酶包含任一个下述氨基酸残基位置的突变或突变组合：R857A、R719A、K394A、K844A、R719A+K394A、K394A+K844A、R857A+K394A、R719A+K844A、R857A+R719A、R857A+K844A、R719A+K394A+K844A、R857A+R719A+K394A、R857A+K394A+K844A、R857A+R719A+K844A、R857A+R719A+K394A+K844A；其中，氨基酸位置编号如 SEQ ID NO: 1 或 8 所示的相应氨基酸位置所定义。在一些实施方式中，所述工程化 Cas12i 核酸酶包含任一个下述突变或突变组合：R857A、R719A、K394A、或 K844A；其中，氨基酸位置编号如 SEQ ID NO: 1 或 8 所示的相应氨基酸位置所定义。在一些实施方式中，所述工程化 Cas12i 核酸酶包含 K844A 突变；其中，氨基酸位置编号如 SEQ ID NO: 1 或 8 所示的相应氨基酸位置所定义。在一些实施方式中，所述工程化 Cas12i 核酸酶包含

R719A 和 K844A 突变；其中，氨基酸位置编号如 SEQ ID NO: 1 或 8 所示的相应氨基酸位置所定义。在一些实施方式中，所述工程化 Cas12i 核酸酶包含 R857A 和 K844A 突变；其中，氨基酸位置编号如 SEQ ID NO: 1 或 8 所示的相应氨基酸位置所定义。在一些实施例中，为实现改善基因编辑效率的目的，也可以使用与上述工程化的 Cas12i 核酸酶(将参
5 比 Cas12i 核酸酶中与 DNA-RNA 双螺旋相互作用的一个或多个极性或带正电荷的氨基酸替换为疏水氨基酸)具有至少约 85%(例如任一至少约 86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%)序列一致性的工程化的 Cas12i 核酸酶。

10 6)其它突变

如未特殊说明，本文所述突变可以包括一个或多个：插入、删除、置换，可以是单个氨基酸或多个氨基酸的突变。

在 1)至 5)节中描述的任一个或者多个突变可以结合任一个或多个增加 Cas12i 活性的已知突变，例如靶标结合、双链切割活性、切口酶活性和/或基因编辑活性。示例行的突
15 变例如可见于下述文献 PCT/CN2020/0134249 和 CN 112195164 A，这些文献通过整体引用而并入本文。在 1)至 5)节中描述的任一个或者多个突变也可以结合任一个或多个降低 Cas12i 活性的已知突变，例如靶标结合、双链切割活性、切口酶活性和/或基因编辑活性。

在一些实施方式中，所述工程化的 Cas12i 核酸酶(如包含上述 1)至 5)节中一个或者多个突变的工程化的 Cas12i 核酸酶)还包含一个或多个柔性区突变，所述突变增加了(比
20 如增加至少约任一 10%、20%、50%、60%、70%、80%、90%、1 倍、1.1 倍、1.2 倍、1.5 倍、2 倍、3 倍、4 倍、5 倍、10 倍、20 倍、50 倍、100 倍或更多)参比 Cas12i 核酸酶(或所述包含(1)-(5)节中任一个或者多个突变的工程化的 Cas12i 核酸酶)中的柔性区的柔性(flexibility)。所述参比 Cas12i 核酸酶中的柔性区可以使用本领域中已知的任何方法来确定。在一些实施方案中，仅基于所述参比 Cas12i 核酸酶的氨基酸序列确定多个柔性区。
25 在一些实施方案中，基于所述参比 Cas12i 核酸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。

此处所述的工程化改造 Cas12i 核酸酶柔性区的方法包括：(a)获得多种工程化的 Cas12i 核酸酶，每种工程化的 Cas12i 核酸酶包含一个或多个突变，所述突变增加了参比 Cas12i 核酸酶的一个或多个柔性区中的柔性区柔性；以及(b)从所述多种工程化的 Cas12i
30 核酸酶中选择一种或多种工程化的 Cas12i 核酸酶，其中所述一种或多种工程化的 Cas12i 核酸酶与所述参比 Cas12i 核酸酶相比具有增加的活性(例如靶标结合、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,所述方法还包括确定所述参比 Cas12i 核酸酶中的一个或多个柔性区。在一些实施方案中,所述方法还包括在真核细胞诸如哺乳动物细胞(例如人细胞)中测量所述工程化的 Cas12i 核酸酶的活性。

在一些实施方案中,使用选自下组的程序确定多个柔性区: PredyFlexy、FoldUnfold、
5 PROFbval、Flexserv、FlexPred、DynaMine 和 Disomine。在一些实施方案中,所述一个或多个柔性区位于无规卷曲处。在一些实施方案中,所述一个或多个柔性区在参比 Cas12i 核酸酶(或所述包含(1)-(5)节中任一个或者多个突变的工程化的 Cas12i 核酸酶)的与 DNA 和/或 RNA 相互作用的结构域中。在一些实施方案中,所述柔性区的长度为至少约 5 个(例如 5 个)氨基酸。

10 在一些实施方案中,所述一个或多个突变包括在柔性区中插入一个或多个(例如 2 个)甘氨酸(G)残基。在一些实施方案中,所述一个或多个 G 残基插入在柔性区中的柔性氨基酸残基的 N 末端,其中所述柔性氨基酸残基选自下组: G、丝氨酸(S)、天冬酰胺(N)、天冬氨酸(D)、组氨酸(H)、蛋氨酸(M)、苏氨酸(T)、谷氨酸(E)、谷氨酰胺(Q)、赖氨酸(K)、精氨酸(R)、丙氨酸(A)和脯氨酸(P)。在一些实施方案中,根据以下优先级选择所述柔性
15 氨基酸残基: 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括用一个或多个 G 残基替换一个或多个非 G 残基。

在一些实施方案中,所述一个或多个突变包括用 G 残基替换柔性区中的疏水氨基酸残基,其中所述疏水氨基酸残基选自下组: 亮氨酸(L)、异亮氨酸(I)、缬氨酸(V)、半胱氨酸(C)、酪氨酸(Y)、苯丙氨酸(F)和色氨酸(W)。

20 在一些实施方案中,所述活性是位点特异性核酸酶活性。在一些实施方案中,所述活性是真核细胞(例如人细胞)中的基因编辑活性。在一些实施方案中,所述基因编辑效率是使用如下方法来测量的: T7 核酸内切酶 1 (T7E1)测定、靶 DNA 的测序、由分解追踪插入缺失(TIDE)测定或通过扩增子分析进行插入缺失检测(IDAA)测定。

在一些实施方案中,所述工程化 Cas12i 核酸酶(例如所述包含(1)-(5)节中任一个或者
25 多个突变的工程化的 Cas12i 核酸酶)包含一个或多个柔性区突变,所述柔性区突变增加了参比 Cas12i 核酸酶(如 Cas12i2 核酸酶、或所述包含(1)-(5)节中任一个或者多个突变的工程化的 Cas12i 核酸酶)中柔性区的柔性,所述柔性区选自对应于以下的区的组: 氨基酸残基 228-232、氨基酸残基 439-443、氨基酸残基 478-482、氨基酸残基 500-504、氨基酸残基 775-779 和氨基酸残基 925-929,其中所述氨基酸残基编号如 SEQ ID NO: 1 所示的相
30 应氨基酸位置所定义。在一些实施方案中,所述柔性区选自氨基酸残基 439-443 或氨基酸残基 925-929,其中所述氨基酸残基编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定

义。在一些实施方案中，所述参比 Cas12i 酶是 Cas12i2 (SEQ ID NO: 1)。在一些实施方案中，所述一个或多个柔性区突变包括在所述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，所述一个或多个 G 残基插入在所述柔性区中柔性氨基酸残基的 N 末端，其中所述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，根据以下优先级选择所述柔性氨基酸残基：
5 G>S>N>D>H>M>T>E>Q>K>R>A>P。在一些实施方案中，所述一个或多个柔性区突变包括用 G 残基替换柔性区中的疏水氨基酸残基，其中所述疏水氨基酸残基选自下组：A、V、I、L、M、F、Y、P、C 或 W；优选的，选自：L、I、V、C、Y、F 和 W。

在一些实施方式中，所述柔性区突变位于 439 和/或 926。在一些实施方式中，它们是下述一个或多个氨基酸：L439、I926。其中，所述氨基酸位置如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。
10

在一些实施方案中，所述工程化的 Cas12i 核酸酶(例如所述包含(1)-(5)节中任一个或者多个突变的工程化的 Cas12i 核酸酶)包含 926G 和/或 439(L+G)突变。在一些实施方案中，所述工程化的 Cas12i 核酸酶包含一个或多个下列柔性区突变：I926G、L439(L+G)和 L439(L+GG)。在一些实施方案中，所述工程化的 Cas12i 核酸酶包含 I926G 突变。在一些
15 实施方案中，所述工程化的 Cas12i 核酸酶包含 L439(L+G)突变。在一些实施方案中，所述工程化的 Cas12i 核酸酶包含 L439(L+GG)突变。其中，所述氨基酸残基编号基于 SEQ ID NO: 1。

在本说明书的上下文中，L439(L+G) 的含义是，在所援引的氨基酸序列中(如 SEQ ID NO: 1)，在第 439 号氨基酸后插入一个甘氨酸(G)，原有第 439 位的 L 序列不变；在本申请的上下文以及附图中，有时也表示为 439G。而 L439(L+GG) 的含义是，在所援引的氨基酸序列中，在第 439 号氨基酸后插入两个甘氨酸(GG)，原有第 439 位的 L 序列不变；在本申请的上下文中以及附图中，有时也表示为 439GG。
20

在一些实施方式中，所述工程化 Cas12i 核酸酶(例如所述包含(1)-(5)节中任一个或者多个突变的工程化的 Cas12i 核酸酶)包含任一个下述氨基酸残基位置的突变或突变组合：
25 926、439、925+926、362+925+926、439+926、323+362+926 (例如 I926、L439、N925+I926、D362+N925+I926、L439+I926、E323+D362+I926)；其中，氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。在一些实施方式中，所述位于氨基酸位置 323、362、925 或 926 的突变为将所述位置氨基酸残基替换为 R、H 或 K(如 R)的突变；
30 其中，氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。在一些实施方式中，所述位于氨基酸位置 439 或 926 的突变为将所述位置氨基酸残基替换为 G 或在所述

氨基酸残基后插入 G 或 GG 的突变;其中,氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

在一些实施方式中,所述工程化 Cas12i 核酸酶(例如所述包含(1)-(5)节中任一个或者多个突变的工程化的 Cas12i 核酸酶)包含任一个下述氨基酸残基或氨基酸残基组合的突变:
5 926G、439(L+GG)、925R+926G、362R+925R+926G、439(L+GG)+926R、或 323R+362R+926G;其中,氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

在一些实施方式中,所述工程化 Cas12i 核酸酶包含任一个下述突变或突变组合:
10 I926G、L439(L+GG)、L439(L+GG)+I926R、N925R+I926G、D362R+N925R+I926G、E323R+D362R+I926G;其中氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

在一些实施方案中,为实现改善基因编辑效率的目的,也可以使用与上述工程化的 Cas12i 核酸酶(例如所述包含(1)-(5)节中任一个或者多个突变的工程化的 Cas12i 核酸酶,和/或包含上述柔性区突变的工程化的 Cas12i 核酸酶)具有至少约 85%(例如任一至少约
15 86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%) 序列一致性的工程化的 Cas12i 核酸酶。

7) 组合突变

利用根据本说明书中 1)-(6)节所描述的突变以及表 1 至表 5,表 9、表 12、表 14、表
20 16-18 中的一个或多个氨基酸替换/插入的组合而得到的工程化的 Cas12i 核酸酶都在本申请的请求保护范围之下。

在一些实施方案中,所述工程化的 Cas12i 核酸酶包含任一个下述氨基酸残基位置的突变或突变组合: 164、176、238、323、357、362、394、439、447、563、715、719、
807、844、848、857、861、925、926、958、176+238+447+563、323+362、
25 176+238+447+563+164、176+238+447+563+926、176+238+447+563+323+362、164+926、164+323+362、176+238+447+563+164+926、176+238+447+563+164+323+362、176+238+447+563+164+926+323+362、176+238+447+563+164+926+323+362、176+238+447+563+164+926+323+362+926+439、719+844、857+844、362+926、925+926、362+925+926、439+926、323+362+926;其中,氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。在一些实施方式中,所述位于氨基酸位置 176、238、323、362、447、563、926、或 958 的突变为将所述位置氨基酸残基替换为 R、H 或 K(如 R)的突变。在一些实施方式中,所述位于氨基酸位置 164
30

的突变为将所述位置氨基酸残基替换为 Y 或 F(如 Y)的突变。在一些实施方式中, 所述位于氨基酸位置 439 或 926 的突变为将所述位置氨基酸残基替换为 G 或在所述氨基酸残基后插入 G 或 GG 的突变。在一些实施方式中, 所述位于氨基酸位置 357、394、715、719、807、844、848、857、或 861 的突变为将所述位置氨基酸残基替换为疏水氨基酸(如 A)。

5 其中, 氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

在一些实施方式中, 所述工程化 Cas12i 核酸酶包含任一个下述氨基酸残基或氨基酸残基组合的突变: N164、E176; K238; E323、D362、T447; E563、I926、I926、D958、L439、R857、N861、K807、N848、R715、R719、K394、H357、K844、

E176+K238+T447+E563、E323+D362、E176+K238+T447+E563+N164、

10 E176+K238+T447+E563+I926、E176+K238+T447+E563+E323+D362、N164+I926、

N164+E323+D362、E176+K238+T447+E563+N164+I926、

E176+K238+T447+E563+N164+E323+D362、

E176+K238+T447+E563+N164+I926+E323+D362、

E176+K238+T447+E563+N164+E323+D362+I926、E176+K238+T447+E563+

15 N164+E323+D362+I926+L439、E176+K238+T447+E563+N164+E323+D362+I926+L439;

E176+K238+T447+E563+N164+D958;

E176+K238+T447+E563+I926+D958;

E176+K238+T447+E563+E323+D362+D958;

N164+I926+D958;

20 N164+E323+D362+D958;

E176+K238+T447+E563+N164+D958;

E176+K238+T447+E563+N164+I926+D958;

E176+K238+T447+E563+N164+E323+D362+D958;

E176+K238+T447+E563+N164+I926+E323+D362+D958;

25 E176+K238+T447+E563+N164+E323+D362+I926+D958;

E176+K238+T447+E563+N164+E323+D362+I926+L439+D958;

E176+K238+T447+E563+N164+E323+D362+I926+L439+D958;

R719+K844;

R857+K844;

30 E176+K238+T447+E563+N164+E323+D362+R857;

E176+K238+T447+E563+N164+E323+D362+N861;

E176+K238+T447+E563+N164+E323+D362+K807;

E176+K238+T447+E563+N164+E323+D362+N848;

E176+K238+T447+E563+N164+E323+D362+R715;

E176+K238+T447+E563+N164+E323+D362+ R719;

5 E176+K238+T447+E563+N164+E323+D362+K394;

E176+K238+T447+E563+N164+E323+D362+H357;

E176+K238+T447+E563+N164+E323+D362+K844;

E176+K238+T447+E563+N164+E323+D362+R719+K844;

E176+K238+T447+E563+N164+E323+D362+R857+K844; 或

10 E176+K238+T447+E563+N164+E323+D362+K394;

其中，氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

在一些实施方式中，所述工程化 Cas12i 核酸酶包含任一个下述突变或突变组合：

N164Y; E176R; K238R; E323R; D362R; T447R; E563R; I926R; D958R; L439(L+G);

L439(L+GG); R857A; N861A; K807A; N848A; R715A; R719A; K394A; H357A; K844A;

15 R719A+ K844A; N164Y+I926R; E323R+D362R; R857A+K844A; R857A+K844A;

N164Y+I926R+D958R; N164Y+E323R+D362R; N164Y+E323R+D362R+D958R;

E176R+K238R+T447R+E563R; E176R+K238R+T447R+E563R+N164Y;

E176R+K238R+T447R+E563R+I926R; E176R+K238R+T447R+E563R+E323R+D362R;

E176R+K238R+T447R+E563R+N164Y+I926R;

20 E176R+K238R+T447R+E563R+N164Y+E323R+D362R;

E176R+K238R+T447R+E563R+N164Y+I926R+E323R+D362R;

E176R+K238R+T447R+E563R+N164Y+E323R+D362R+I926G;

E176R+K238R+T447R+E563R+N164Y+E323R+D362R+I926G+L439(L+GG);

E176R+K238R+T447R+E563R+N164Y+E323R+D362R+I926G+L439(L+G);

25 E176R+K238R+T447R+E563R+N164Y+D958R;

E176R+K238R+T447R+E563R+I926R+D958R;

E176R+K238R+T447R+E563R+E323R+D362R+D958R;

E176R+K238R+T447R+E563R+N164Y+D958R;

E176R+K238R+T447R+E563R+N164Y+I926R+D958R;

30 E176R+K238R+T447R+E563R+N164Y+E323R+D362R+D958R;

E176R+K238R+T447R+E563R+N164Y+I926R+E323R+D362R+D958R;

E176R+K238R+T447R+E563R+N164Y+E323R+D362R+I926G+D958R;
 E176R+K238R+T447R+E563R+N164Y+E323R+D362R+I926G+L439(L+GG)+D958R;
 E176R+K238R+T447R+E563R+N164Y+E323R+D362R+I926G+L439(L+G)+D958R;
 E176R+K238R+T447R+E563R+N164Y+E323R+D362R+R857A;
 5 E176R+K238R+T447R+E563R+N164Y+E323R+D362R+N861A;
 E176R+K238R+T447R+E563R+N164Y+E323R+D362R+K807A;
 E176R+K238R+T447R+E563R+N164Y+E323R+D362R+N848A;
 E176R+K238R+T447R+E563R+N164Y+E323R+D362R+R715A;
 E176R+K238R+T447R+E563R+N164Y+E323R+D362R+ R719A;
 10 E176R+K238R+T447R+E563R+N164Y+E323R+D362R+ K394A;
 E176R+K238R+T447R+E563R+N164Y+E323R+D362R+H357A;
 E176R+K238R+T447R+E563R+N164Y+E323R+D362R+K844A;
 E176R+K238R+T447R+E563R+N164Y+E323R+D362R+ R719A+K844A;
 E176R+K238R+T447R+E563R+N164Y+E323R+D362R+R857A+K844A; 或
 15 E176R+K238R+T447R+E563R+N164Y+E323R+D362R+K394A;

其中，氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

在一些实施方式中，所述工程化 Cas12i 核酸酶包含如下突变组合：

E176R+K238R+T447R+E563R+N164Y+E323R+D362R，其中，氨基酸位置编号如 SEQ ID
 NO: 1 所示的相应氨基酸位置所定义。该突变体以下被命名为 CasXX，其序列编号为
 20 SEQ ID NO: 8。

在一些实施方式中，所述工程化 Cas12i 核酸酶包含任一个下述突变组合：

E176R+K238R+T447R+E563R+N164Y+E323R+D362R+R857A;

E176R+K238R+T447R+E563R+N164Y+E323R+D362R+R719A;

E176R+K238R+T447R+E563R+N164Y+E323R+D362R+K394A;

25 E176R+K238R+T447R+E563R+N164Y+E323R+D362R+K844A;

E176R+K238R+T447R+E563R+N164Y+E323R+D362R+R719A+K844A;

E176R+K238R+T447R+E563R+N164Y+E323R+D362R+R857A+K844A; 其中，氨基
 酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。这些 Cas12i 突变体是在
 CasXX(SEQ ID NO: 8)基础上进行了进一步的优化所得到的优化的突变体。

30 在一些实施方式中，所述工程化 Cas12i 核酸酶包含如下突变组合：

E176R+K238R+T447R+E563R+N164Y+E323R+D362R+K394A，其中，氨基酸位置编号

如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。该突变体以下被命名为“HF-20”，其序列如 SEQ ID NO: 20 所示。

在一些实施方式中，所述工程化 Cas12i 核酸酶包含 E176R、K238R、T447R、E563R、N164Y、E323R、和 D362R 突变。在一些实施方式中，所述工程化 Cas12i 核酸酶包含 E176R、K238R、T447R、E563R、N164Y、和 I926R 突变。在一些实施方式中，所述工程化 Cas12i 核酸酶包含 E176R、K238R、T447R、E563R、E323R、和 D362R 突变。在一些实施方式中，所述工程化 Cas12i 核酸酶包含 E176R、K238R、T447R、E563R、N164Y、E323R、D362R、和 I926G 突变。在一些实施方式中，所述工程化 Cas12i 核酸酶包含 E176R、K238R、T447R、E563R、N164Y、E323R、D362R、I926G、和 L439(L+GG)突变。在一些实施方式中，所述工程化 Cas12i 核酸酶包含 E176R、K238R、T447R、E563R、N164Y、E323R、D362R、I926G、和 L439(L+G)突变。在一些实施方式中，所述工程化 Cas12i 核酸酶包含 E176R、K238R、T447R、E563R、N164Y、和 D958R 突变。在一些实施方式中，所述工程化 Cas12i 核酸酶包含 E176R、K238R、T447R、E563R、N164Y、I926R、和 D958R 突变。在一些实施方式中，所述工程化 Cas12i 核酸酶包含 E176R、K238R、T447R、E563R、N164Y、E323R、D362R、和 D958R 突变。在一些实施方式中，所述工程化 Cas12i 核酸酶包含 E176R、K238R、T447R、E563R、N164Y、E323R、D362R、I926G、和 D958R 突变。在一些实施方式中，所述工程化 Cas12i 核酸酶包含 E176R、K238R、T447R、E563R、N164Y、E323R、D362R、I926G、L439(L+GG)、和 D958R 突变。在一些实施方式中，所述工程化 Cas12i 核酸酶包含 E176R、K238R、T447R、E563R、N164Y、E323R、D362R 和 K844A 突变。在一些实施方式中，所述工程化 Cas12i 核酸酶包含 E176R、K238R、T447R、E563R、N164Y、E323R、D362R、R719A 和 K844A 突变。在一些实施方式中，所述工程化 Cas12i 核酸酶包含 E176R、K238R、T447R、E563R、N164Y、E323R、D362R、R857A 和 K844A 突变。其中，所述氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

在一些实施方案中，为实现改善基因编辑效率的目的，也可以使用与上述工程化的 Cas12i 核酸酶具有至少约 80%(例如任一至少约 81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%)序列一致性的工程化的 Cas12i 核酸酶。

在一些实施方案中，提供了一种工程化的 Cas12i 核酸酶，其包含如 SEQ ID NOs:

2~24 所示的任一氨基酸序列, 或具有与 SEQ ID NOs: 2~24 所示的任一氨基酸序列具有至少约 80%(例如任一至少约 81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%)序列一致性的氨基酸序列。

5 参比 Cas12i 核酸酶

在一些实施方案中, 所述参比 Cas12i 核酸酶为 Cas12i1、Cas12i2, 或其直系同源物。在一些实施方案中, 所述参比 Cas12i 核酸酶为天然 Cas12i1, 或其变体(如自然存在的变体)。在一些实施方案中, 所述参比 Cas12i 核酸酶为天然 Cas12i2(如 SEQ ID NO: 1 所示), 或其变体(如自然存在的变体)。在一些实施方案中, 所述参比 Cas12i 核酸酶为经过工程改造后的 Cas12i 核酸酶(如本发明所述的包含(1)-(5)节中任一个或者多个突变的工程化的 Cas12i 核酸酶)。在一些实施方案中, 所述参比 Cas12i 核酸酶为 CasXX (SEQ ID NO: 8)。

V-I 型 CRISPR-Cas12i 已被鉴定为 RNA-指导的 DNA 核酸内切酶系统。与诸如 Cas12b 或 Cas9 的 CRISPR-Cas 系统不同, 基于 Cas12i 的 CRISPR 系统不需要 tracrRNA 序列。在一些实施方案中, 所述 RNA 指导序列包括 crRNA。通常, 本文所述的 crRNA 包括直接重复序列和间隔区序列。在某些实施方案中, 所述 crRNA 包括与指导序列或间隔区序列连接的直接重复序列、基本由或由其组成。在一些实施方案中, 所述 crRNA 包括直接重复序列、间隔区序列和直接重复序列(DR-间隔区-DR), 其是其他 CRISPR 系统中的前体 crRNA (pre- crRNA)构型的典型特征。在一些实施方案中, 所述 crRNA 包括截短的直接重复序列和间隔区序列, 其是经加工的或成熟的 crRNA 的典型特征。在一些实施方案中, 所述 CRISPR-Cas12i 效应蛋白与 RNA 指导序列形成复合物, 并且所述间隔区序列将复合物引导至与靶核酸的序列特异性结合, 所述靶核酸与间隔区序列互补(如至少 70%互补)。

在一些实施方案中, 本申请的工程化的 Cas12i 是核酸内切酶, 其结合靶序列的特定位点并在指导 RNA 的指导下切割, 并且具有 DNA 和 RNA 内切核酸酶活性。在一些实施方案中, 所述 Cas12i 能够通过加工前体 crRNA 阵列来进行自主的 crRNA 生物发生。自主的前体 crRNA 处理可促进 Cas12i 的递送, 从而实现双切口应用, 因为可以由单个 crRNA 转录物靶向两个单独的基因组位点。然后, Cas12i 蛋白将 CRISPR 阵列加工成两个同源的 crRNA, 从而形成配对的切口复合物。V-I 型(Cas12i)效应蛋白的复用 (Multiplexing)是利用所述效应蛋白的前体 crRNA 处理能力完成的, 其中可以针对具有不同序列的多个靶标在单个 RNA 指导序列上编程。这样, 可以同时操纵多个基因或 DNA 靶标以用于治疗应用。在一些实施方案中, 所述指导 RNA 包含由 CRISPR 阵列表达的前

体 crRNA，所述 CRISPR 阵列由与未加工的 DR 序列交错的靶序列组成，通过所述效应蛋白的内在前体 crRNA 加工而重复以使得能够同时靶向一个、两个或多个位点。

来自多种生物体的 Cas12i 核酸酶可以用作所述参比 Cas12i 核酸酶，以提供本申请的工程化的 Cas12i 核酸酶及效应蛋白。示例性 Cas12i 核酸酶已经描述于例如
5 WO2019/201331A1 和 US2020/0063126A1 中，其通过整体引用而并入本文。在一些实施方案中，所述参比 Cas12i 核酸酶具有酶活性。在一些实施方案中，所述参比 Cas12i 是核酸酶，即切割靶双螺旋核酸(例如，双螺旋 DNA)的两条链。在一些实施方案中，所述参比 Cas12i 是切口酶，即切割靶双螺旋核酸(例如，双螺旋 DNA)的单链。在一些实施方案中，所述参比 Cas12i 核酸酶是酶失活的。在一些实施方案中，所述参比 Cas12i 核酸酶是
10 Cas12i1、Cas12i2 或 Cas12i-Phi。在一些实施方案中，所述参比 Cas12i 核酸酶含有 SEQ ID NO: 1 的序列。与 Cas12i(如 Cas12i2)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%、96%、97%、98%、99%或更多中任一个)的直系同源物可以用作设计本申请的工程化的 Cas12i 核酸酶或效应蛋白的基础。

15 工程化的 Cas12i 核酸酶的功能变体

在一些实施方案中，所述工程化的 Cas12i 核酸酶是基于天然存在的 Cas12i 核酸酶的功能变体(或功能衍生物)。在一些实施方案中，当与相应的工程化的 Cas12i 核酸酶(如包含上述(1)-(7)中任何一种或多种突变的工程化的 Cas12i 核酸酶)的氨基酸序列相比时，功能变体(或功能衍生物)的氨基酸序列具有至少一个氨基酸残基的不同(例如，具有缺失、
20 插入、替换和/或融合)。在一些实施方案中，所述功能变体具有一个或多个突变，如氨基酸替换、插入和缺失。举例来说，与野生型天然存在的 Cas12i 核酸酶相比，或与前述工程化的 Cas12i 核酸酶相比，所述功能变体可包含 1、2、3、4、5、6、7、8、9、10 个或更多个氨基酸中任一个替换。在一些实施方案中，所述一个或多个替换是保守替换。在一些实施方案中，所述功能变体具有天然存在的 Cas12i 核酸酶的所有结构域。在一些实
25 施方案中，所述功能变体不具有天然存在的 Cas12i 核酸酶的一个或多个结构域。在一些实施方案中，所述功能变体具有工程化的 Cas12i 核酸酶的所有结构域。在一些实施方案中，所述功能变体不具有工程化的 Cas12i 核酸酶的一个或多个结构域。在一些实施方式中，Cas12i 核酸酶的功能变体的生物学活性由于其氨基酸的变化而变化，例如由天然核酸酶转变为酶失活突变体。

30 对于本文所述 Cas12i 变体蛋白(例如，切口酶 Cas12i 蛋白、失活或催化失活的 Cas12i (dCas12i))中任一种，所述 Cas12i 变体可包括具有上述相同参数(例如，存在的结

构域、同一性百分比等)的 Cas12i 蛋白序列。

5 在一些实施方式中, 所述工程化的 Cas12i 核酸酶的功能变体具有酶活性(如 DNA 双链或单链切割活性), 或至少具有参比 Cas12i 核酸酶(或其亲本工程化的 Cas12i 核酸酶)约 60%(比如至少约 65%、70%、80%、90%、95%, 96%、97%、98%、99%、或 100%中的任一个)的酶活性。在一些实施方式中, 所述工程化 Cas12i 核酸酶的功能变体的酶活性是其亲本工程化 Cas12i 核酸酶的至少 1.1 倍(比如至少 1.2 倍、1.5 倍、2 倍、3 倍、4 倍、5 倍、10 倍或更多)的酶活性。

10 在一些实施方案中, 所述工程化 Cas12i 核酸酶的功能变体与其工程化的 Cas12i 核酸酶的非功能变体突变形式相比, 具有不同的催化活性。在一些实施方案中, 所述功能变体突变(例如, 氨基酸替换、插入和/或缺失)在 Cas12i 核酸酶的催化结构域(例如, RuvC 结构域)中。在一些实施方案中, 所述工程化 Cas12i 核酸酶的功能变体包含一个或多个催化结构域中的突变。切割双链靶核酸的一条链而不切割另一条链的 Cas12i 核酸酶, 在本文中被称为“切口酶”(例如“Cas12i 切口酶”)。在本文中, 基本上不具有核酸酶活性的 Cas12i 核酸酶称为失活 Cas12i 蛋白(“dCas12i”)(与其融合的异源多肽(融合伴侣)可以提供核酸酶活性, 详见下面“工程化的 Cas12i 效应蛋白”章节)。在一些实施方案中, 相对于其非功能变体的突变形式而言, 当功能变体突变酶的 DNA 切割活性小于约 25%、10%、5%、1%、0.1%、0.01%或更低时, 认为该 Cas12i 核酸酶功能变体基本上缺乏所有的 DNA 切割活性。

20 Cas12i 核酸酶活性位点中一个或多个氨基酸残基的突变会导致降低或失去酶活的 Cas12i(dCas12i), 在本发明中也称为“核酸酶活性缺失的 Cas12i”或“酶失活突变体”。在一些实施方案中, 本文提供的工程化的 Cas12i 核酸酶可以被修饰以具有减少或缺失的核酸酶活性, 例如, 核酸酶活性缺失的 Cas12i 与野生型 Cas12i 核酸酶(或其亲本工程化的 Cas12i 核酸酶)相比, 核酸酶活性(如 DNA 双链或单链切割活性)降低至少约 50%(比如降低至少约 60%、70%、80%、90%、95%, 96%、97%、98%、99%、或 100%中的任一个)。
25 所述 Cas12i 核酸酶活性可以通过几种方法来降低, 例如, 将突变引入所述 Cas12i 核酸酶的一个或多个结构域中: 与 PAM 相互作用的结构域、参与打开 DNA 双链的结构域、RuvC 结构域、与核酸(DNA/RNA)相互作用的区域等。在一些实施方案中, 所述 Cas12i 核酸酶活性的催化残基(如通过任何通用鉴定方法鉴定出的催化残基)可以被不同的氨基酸残基(例如, 甘氨酸或丙氨酸)替换以降低所述核酸酶活性。Cas12i1(如 SEQ ID NO: 13 所示)的此类突变的实例包括 D647A、E894A 和/或 D948A。Cas12i2(如 SEQ ID NO: 1 所示)
30 的此类突变的实例包括 D599A、E833A、S883A、H884A、D886A、R900A 和/或

D1019A。在一些实施方案中，所述工程化的 Cas12i 核酸酶或其功能衍生物(比如包含上述(1)-(7)中任何一种或多种突变的工程化的 Cas12i 核酸酶)包含以下一种或多种的核酸酶活性缺失突变：D599A、E833A、S883A、H884A、R900A 和 D1019A；其中，所述氨基酸位置如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

5

工程化的 Cas12i 核酸酶或其功能变体的特性

在一些实施方案中，所述工程化的 Cas12i 核酸酶(或其功能变体)与所述参比 Cas12i 核酸酶相比具有增加的活性。在一些实施方案中，所述工程化的 Cas12i 核酸酶(或其功能变体)与所述参比 Cas12i 核酸酶相比具有降低的活性。在一些实施方案中，所述活性是靶 DNA 结合活性。在一些实施方案中，所述活性是位点特异性核酸酶活性。在一些实施方案中，所述活性是打开 DNA 双链的活性。在一些实施方案中，所述活性是双链 DNA 切割活性。在一些实施方案中，所述活性是单链 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，所述活性是单链 RNA 切割活性，例如位点特异性 RNA 切割活性或非特异性 RNA 切割活性。在一些实施方案中，所述活性是在体外测量的。在一些实施方案中，所述活性是在细胞中测量的，如细菌细胞、植物细胞或真核细胞。在一些实施方案中，所述活性是在哺乳动物细胞如啮齿动物细胞或人细胞中测量的。在一些实施方案中，所述活性是在人细胞诸如 293T 细胞中测量的。在一些实施方案中，所述活性是在小鼠细胞，例如 Hepa1-6 细胞中测量的。在一些实施方案中，所述工程化的 Cas12i 核酸酶(或其功能变体)具有与参比 Cas12i 核酸酶相比增加至少约 10%、20%、30%、40%、50%、60%、70%、80%、90%、95%、1 倍、1.1 倍、1.5 倍、2 倍、3 倍、4 倍、5 倍、10 倍或更多中任一个的活性(如以上所述一种或多种活性，比如位点特异性核酸酶活性)。在一些实施方案中，所述工程化的 Cas12i 核酸酶(或其功能变体)具有与参比 Cas12i 核酸酶相比降低至少约 10%、20%、30%、40%、50%、60%、70%、80%、90%、95%、1 倍、1.1 倍、1.5 倍、2 倍、3 倍、4 倍、5 倍、10 倍或更多中任一个的活性(如以上所述一种或多种活性，比如位点特异性核酸酶活性)。所述工程化的 Cas12i 核酸酶(或其功能变体)的位点特异性核酸酶活性可以使用本领域中已知的方法来测量，包括例如，凝胶迁移测定，如本文提供的实施例中所所述的基于琼脂糖凝胶电泳的体外切割测定。

在一些实施方案中，所述活性是细胞中的基因编辑活性。在一些实施方案中，所述细胞是细菌细胞、植物细胞或真核细胞。在一些实施方案中，所述细胞是哺乳动物细胞如啮齿动物细胞或人细胞。在一些实施方案中，所述细胞是 293T 细胞。在一些实施方案

30

中, 所述活性是在小鼠细胞, 例如 Hepa1-6 细胞中测量的。在一些实施方案中, 所述活性是在细胞中靶基因组位点的插入缺失形成活性, 例如通过所述工程化的 Cas12i 核酸酶 (或其功能变体) 对靶核酸进行位点特异性切割和通过非同源末端连接(NHEJ)机制进行 DNA 修复。在一些实施方案中, 所述活性是在细胞中靶基因组位点插入外源核酸序列的活性, 例如通过所述工程化的 Cas12i 核酸酶 (或其功能变体) 对靶核酸进行位点特异性切割和通过同源重组(HR; 通过进一步引入一个修复模板)机制进行 DNA 修复。在一些实施方案中, 所述工程化的 Cas12i 核酸酶 (或其功能变体) 与参比 Cas12i 核酸酶相比在细胞 (例如人细胞如 293T 细胞, 或小鼠 Hepa1-6 细胞) 的靶基因组位点处增加至少约 10%、20%、30%、40%、50%、60%、70%、80%、90%、95%、1 倍、1.1 倍、1.5 倍、2 倍、3 倍、4 倍、5 倍、10 倍或更多中任一个的基因编辑 (例如, 切割、插入缺失形成、或修复) 活性。在一些实施方案中, 所述工程化的 Cas12i 核酸酶 (或其功能变体) 与参比 Cas12i 核酸酶相比在细胞 (例如人细胞如 293T 细胞或小鼠 Hepa1-6 细胞) 的多个 (例如 2、3、4、5、6、7、8、9、10 个或更多) 靶基因组位点处增加至少约 10%、20%、30%、40%、50%、60%、70%、80%、90%、95%、1 倍、1.1 倍、1.5 倍、2 倍、3 倍、4 倍、5 倍、10 倍或更多中任一个的基因编辑 (例如, 切割、插入缺失形成、或修复) 活性。在一些实施方案中, 所述工程化的 Cas12i 核酸酶 (或其功能变体) 与参比 Cas12i 核酸酶相比, 能够编辑更多数目 (例如 2、3、4、5、6、7、8、9、10、20、30、40、50、60、70 个或更多) 的基因组位点, 比如能识别更多 (例如 2、3、4、5、6、7、8、9、10、20、30、40、50、60、70 个或更多) 的 PAM 序列。在一些实施方案中, 所述工程化的 Cas12i 核酸酶 (或其功能变体) 的共有 PAM 序列与参比 Cas12i 核酸酶相同。

可使用本领域已知的方法确定体外或细胞中工程化的 Cas12i 核酸酶 (或其功能变体) 的基因编辑效率, 包括例如 T7 核酸内切酶 1 (T7E1) 测定、靶 DNA 的测序 (包括例如, Sanger 序列, 以及二代测序)、由分解追踪插入缺失 (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, 文章编号 888, 其通过整体引用而并入本文。在一些实施方案中, 例如, 如本文实施例中所述, 使用靶向的二代测序 (NGS) 来测量细胞中所述工程化的 Cas12i 核酸酶的基因编辑效率。可用于确定所述工程化的 Cas12i 核酸酶 (或其功能变体) 的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 和 EMX1。在一些实施方案中, 所述工程化的 Cas12i 核酸酶 (或其功能变体) 的基因编辑效率为所述工程化的 Cas12i 核酸酶在至少 5 个、10 个、15 个、20 个、25 个、30 个、35 个、40 个、45 个、50 个、55 个、60 个、65 个或更多位点 (如人类

细胞基因组位点)的平均基因编辑效率。在一些实施方案中,所述工程化的 Cas12i 核酸酶(或其功能变体)的基因编辑效率(例如插入缺失率)达到至少 10%、20%、30%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%或更高。

5 在一些实施方案中,所述工程化的 Cas12i 核酸酶(或其功能变体)与所述参比 Cas12i 核酸酶相比具有增加的靶向特异性,具有降低的脱靶率(例如减少识别的脱靶位点、和/或降低对一个或多个脱靶位点的编辑效率),和/或增加的靶序列编辑效率。在一些实施方案中,所述工程化的 Cas12i 核酸酶(或其功能变体)与所述参比 Cas12i 核酸酶相比,降低至少约 5%(如降低至少约 10%、20%、30%、40%、50%、60%、70%、80%、90%、95%、或 100%中任一个)的脱靶率。在一些实施方案中,所述工程化的 Cas12i 核酸酶(或其功能
10 变体)与所述参比 Cas12i 核酸酶相比,减少至少一个(如至少 2、3、4、5、6、7、8、9、10、20、50 或更多)脱靶位点。在一些实施方案中,所述工程化的 Cas12i 核酸酶(或其功能变体)与所述参比 Cas12i 核酸酶相比,对靶序列的编辑效率相同或相似(比如在 1.1 倍以内),或对靶序列的编辑效率增加(例如是参比 Cas12i 核酸酶至少 1.2 倍、1.5 倍、2 倍、3 倍、5 倍、10 倍或更多)。在一些实施方案中,所述工程化的 Cas12i 核酸酶(或其功能变
15 体)与所述参比 Cas12i 核酸酶相比,降低至少约 5%的脱靶率,且对靶序列的编辑效率相同、相似(比如在 1.1 倍以内)、或增加。

指导 RNA 或 crRNA

20 在一些实施例中,指导 RNA 或 crRNA 从 5'到 3'包括以下或由以下组成:直接重复序列、间隔序列。在一些实施例中,指导 RNA 从 5'到 3'包括以下或由以下组成:直接重复序列、间隔序列、直接重复序列的串联构建的核苷酸序列。在一些实施例中, RNA 指导物包括 crRNA。在一些实施例中,指导 RNA 不包含 tracrRNA。

通常,本文描述的 crRNA 包括直接重复序列和间隔子序列。在某些实施例中, crRNA 包括以下,基本上由以下组成或由以下组成:连接到指导序列或间隔子序列的直接重复序列。在一些实施例中, crRNA 包括直接重复序列、间隔子序列和直接重复序列
25 (DR-间隔子-DR),其是其他 CRISPR 系统中典型的前体 crRNA(前 crRNA)构型。在一些实施例中, crRNA 包括截短的直接重复序列和间隔子序列,其是典型的经加工或成熟的 crRNA。在一些实施例中, CRISPR-Cas 效应子蛋白与 RNA 指导物形成复合物,间隔子序列将该复合物导向与跟间隔子序列互补的靶核酸进行序列特异性结合。

30 在一些实施例中, RNA 指导物包含直接重复。在一些实施例中, RNA 指导物可以形成二级结构,例如,如本文所述的茎环结构。

在一些实施例中,本文描述的 CRISPR 系统包括多个 RNA 指导物(例如, 2, 3, 4, 5,

10, 15 个或更多个)或编码多个 RNA 指导物的多个核酸。在一些实施例中, 本文描述的 CRISPR 系统包括单个 RNA 链或编码单个 RNA 链的核酸, 其中 RNA 指导物串联布置。所述单个 RNA 链可包括相同 RNA 指导物的多个拷贝、不同 RNA 指导物的多个拷贝、或其组合。在一些实施方式中, 每种 RNA 指导物对不同的靶核酸具有特异性。

5 在一些实施例中, 本文描述的 CRISPR 系统包括 RNA 指导物或编码 RNA 指导物的核酸。在一些实施例中, RNA 指导物包含以下或由以下组成: 直接重复序列和能够与靶核酸杂交(例如, 在适当条件下杂交)的间隔子序列。

10 在一些实施方式中, RNA 指导序列可以允许形成 CRISPR 效应子复合物并成功结合靶序列, 同时不允许成功的核酸酶活性(即, 没有核酸酶活性/没有引起插入缺失)的方式进行修饰。这些经修饰的指导物序列被称为“失活指导物”或“失活指导物序列”。这些失活指导物或失活指导物序列对于核酸酶活性而言可以是催化上失活的或构象上失活的。失活指导物序列通常比导致活性 RNA 切割的相应指导物序列短。在一些实施例中, 失活指导物比具有核酸酶活性的相应 RNA 指导物短至少约 5%、10%、20%、30%、40% 或 50%。RNA 指导物的失活指导物序列的长度可以是 13 至 15 个核苷酸(例如, 13、14 或 15 个核苷酸)、15 至 19 个核苷酸、或 17 至 18 个核苷酸(例如, 17 个核苷酸)。在一些实施方式中, 失活指导物 RNA 能够与靶序列杂交, 使得 CRISPR 系统被导向细胞中目的基因组基因座而没有可检测的切割活性。

20 本文描述的 RNA 指导物和 crRNA 的序列和长度可以优化。在一些实施例中, RNA 指导物的优化长度可通过鉴定 crRNA 的加工形式或通过 crRNA 的 RNA 指导物的经验长度研究来确定。在一些实施方式中, RNA 指导序列包含碱基修饰。

在一些实施方式中, 本发明还提供核酸(例如 cDNA)的所有可能的变体, 所述变体可以通过基于可能的密码子选择来选择组合而制备。这些组合是根据应用于编码天然存在变体的多核苷酸的标准三联体遗传密码进行的, 并且所有这些变体都被认为是具体公开的。

25 间隔(spacer)序列

10 在一些实施例中, 间隔序列(或间隔子、指导序列)可与靶核酸(如 DNA)的靶序列互补, 例如至少约 70%(例如至少约 75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99% 或 100%)互补。在一些实施例中, 间隔序列与靶核酸(如 DNA)的靶序列有至少 15 个(如至少 16、17、18、19、20、21、22、23、24、25、26、 30 27、28、29、30、35、40、45、50 或更多个)核苷酸互补。

在本领域已知, 不需要完全的互补性, 前提是有足够的互补性发挥作用。可以通过

引入错配(例如一个或多个错配,例如间隔子序列和靶序列之间(包括沿着间隔子/靶的错配的位置)的1或2个错配)来利用切割效率的调节。错配,例如双错配,位于越中心(即,不在3'或5'端);切割效率受到的影响越大。因此,通过选择沿间隔子序列的错配位置,可以调节切割效率。例如,如果期望靶的小于100%的切割(例如,在细胞群体中),则间隔子序列中可引入间隔子序列和靶序列之间的1或2个错配。

所述指导序列可以具有合适的长度。RNA指导物的间隔子长度可在约11至50个(例如约15至50个)核苷酸的范围内。在一些实施方案中,所述指导序列在约18至约35个核苷酸之间,包括例如18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34或35个核苷酸。在一些实施例中,RNA指导物的间隔子长度为至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸或至少22个核苷酸。在一些实施例中,间隔子长度为15至17个核苷酸、15至23个核苷酸、16至22个核苷酸、17至20个核苷酸、20至24个核苷酸(例如、20、21、22、23或24个核苷酸)、23至25个核苷酸(例如、23、24或25个核苷酸)、24至27个核苷酸、27至30个核苷酸、30至45个核苷酸(例如、30、31、32、33、34、35、40或45个核苷酸)、30或35至40个核苷酸、41至45个核苷酸、45至50个核苷酸,或更长。在一些实施例中,RNA指导物的间隔子长度为31个核苷酸。在一些实施例中,RNA指导物的直接重复长度为至少21个核苷酸,或为21至37个核苷酸(例如,23、24、25、30、35或36个核苷酸)。在一些实施例中,间隔序列包括以下或由以下组成:约15至约34个核苷酸(例如,16、17、18、19、20、21或22个核苷酸)。在一些实施例中,间隔序列长度在17个核苷酸和31个核苷酸之间。在一些实施例中,间隔序列长度在15个核苷酸和24个核苷酸之间。在一些实施例中,RNA指导物的直接重复长度为23或20个核苷酸。

直接重复(DR)序列

直接重复序列能够引导Cas12i蛋白(如任一本发明所述工程化的Cas12i核酸酶、其功能变体或效应蛋白)结合至指导gRNA(或crRNA)以形成靶向靶序列的CRISPR-Cas复合物。任何能够引导本申请中工程化的Cas12i核酸酶或效应蛋白结合至指导gRNA(或crRNA)以形成靶向靶序列的CRISPR-Cas复合物的DR都可以用于本发明,例如US11168324(其全部内容通过整体引用并入本文)中描述的DR序列。

直接重复可以包括两个核苷酸段,它们可以彼此互补,被间插的核苷酸隔开,这样使得直接重复可以杂交形成双链RNA双链体(dsRNA双链体),导致茎环结构,其中两个互补核苷酸段形成茎,并且间插的核苷酸形成环或发夹。例如,形成“环”的中间核苷酸

具有从约 6 个核苷酸到约 8 个核苷酸，或约 7 个核苷酸的长度。在不同的实施例中，茎可以包括至少 2 个、至少 3 个、至少 4 个或 5 个碱基对。

在一些实施方式中，直接重复可以包括长度为约 4 个至约 7 个（如 4、5、6、7 个）核苷酸的两个核苷酸互补段，其间间隔约 5 个至约 9 个（如 5、6、7、8、9 个）核苷酸。

5 本领域技术人员可以模拟已知的直接重复的结构。

直接重复可以包括以下或由以下组成：约 13 至约 23 个核苷酸、约 22 至约 40 个核苷酸、或约 23 至约 38 个核苷酸或约 23 至约 36 个核苷酸。

10 在一些实施例中，直接重复序列包括靠近 3' 末端(紧邻间隔子序列)的茎环结构。在一些实施例中，直接重复序列包括靠近 3' 末端的茎环，其中茎的长度为 5 个核苷酸。在一些实施例中，直接重复序列包括靠近 3' 末端的茎环，其中茎的长度为 5 个核苷酸，并且环的长度为 7 个核苷酸。在一些实施例中，直接重复序列包括靠近 3' 末端的茎环，其中茎的长度为 5 个核苷酸，并且环的长度为 6、7 或 8 个核苷酸。

15 在一些实施例中，直接重复序列包括靠近 3' 末端的序列 5'-CCGUCNNNNNUGACGG-3'(SEQ ID NO: 68)，其中 N 是指任何核碱基。在一些实施例中，直接重复序列包括靠近 3' 末端的序列 5'-GUGCCNNNNNUGGCAC-3'(SEQ ID NO: 69)，其中 N 是指任何核碱基。

20 在一些实施例中，直接重复序列包括靠近 3' 末端的序列 5'-GUGUCN₅₋₆UGACAX₁₋₃(SEQ ID NO: 70 或 71)，其中 N₅₋₆是指任何 5 或 6 个核碱基的连续序列，并且 X₁是指 C 或 T 或 U。在一些实施例中，直接重复序列包括靠近 3' 末端的序列 5'-UCX₃UX₅X₆X₇UUGACGG-3'(SEQ ID NO: 72)，其中 X₃是指 C 或 T 或 U，X₅是指 A 或 T 或 U，X₆是指 A 或 C 或 G，并且 X₇是指 A 或 G。在一些实施例中，直接重复序列包括靠近 3' 末端的序列 5'-CCX₃X₄X₅CX₇UUGGCAC-3'(SEQ ID NO: 73)，其中 X₃是指 C 或 T 或 U，X₄是指 A 或 T 或 U，X₅是指 C 或 T 或 U，并且 X₇是指 A 或 G。在一些实施例中，编码直接重复序列的核苷酸包括以下或由以下组成：与 SEQ ID NO: 25 59(AGAAATCCGTCTTTCATTGACGG) 或 SEQ ID NO: 79(GTTGCAAAACCCAAGAAATCCGTCTTTCATTGACGG)至少约 80% 相同，例如至少约 85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99% 或 100% 相同的核苷酸序列。在一些实施例中，编码直接重复序列的核苷酸包含 SEQ ID NO: 79 的至少 21 个（例如 21、22、23、24、25、26、27、28、29、30、31、32、33、34、35 或 36 个）核苷酸。

30 “茎环结构”是指具有二级结构的核酸，所述二级结构包括已知或预测形成双链(茎部分)的核苷酸区域，所述双链(茎部分)在一侧由主要为单链核苷酸的区域(环部分)连接。术

语“发夹”和“折回”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公知的，并且这些术语与其在本领域中的公知含义一致地使用。如本领域已知的，茎环结构不需要精确的碱基配对。因此，茎可以包括一个或多个碱基错配。可替代地，碱基配对可以是精确的，即不包括任何错配。在一些实施例中，直接重复的茎由 5 个互相杂交的互补核碱基组成，并且环长度是 6、7 或 9 个核苷酸。

在一些实施例中，编码直接重复的序列包含以下或由以下组成：SEQ ID NO: 79 所示序列。在一些实施例中，编码直接重复的序列包含以下或由以下组成：SEQ ID NO: 79 核酸序列的具有起始三个 5'核苷酸的截短的核酸。在一些实施例中，编码直接重复的序列包含以下或由以下组成：SEQ ID NO: 79 核酸序列的具有起始四个 5'核苷酸的截短的核酸。在一些实施例中，编码直接重复的序列包含以下或由以下组成：SEQ ID NO: 79 核酸序列的具有起始五个 5'核苷酸的截短的核酸。在一些实施例中，编码直接重复的序列包含以下或由以下组成：SEQ ID NO: 79 核酸序列的具有起始六个 5'核苷酸的截短的核酸。在一些实施例中，编码直接重复的序列包含以下或由以下组成：SEQ ID NO: 79 核酸序列的具有起始七个 5'核苷酸的截短的核酸。在一些实施例中，编码直接重复的序列包含以下或由以下组成：SEQ ID NO: 79 核酸序列的具有起始八个 5'核苷酸的截短的核酸。在一些实施例中，编码直接重复的序列包含以下或由以下组成：SEQ ID NO: 59 所示序列。

在一些实施方式中，直接重复是 SEQ ID NO: 59 或 79 编码的 RNA 序列的“功能性变体”，例如“功能性截短版本”、“功能性延长版本”、或“功能性替换版本”，比如 SEQ ID NO: 79 的一部分(截短版本)，仍然具有 DR 功能。DR“功能性变体”是参比 DR(如亲本 DR)的 5'和/或 3'端延长(功能性延长版本)或截短(功能性截短版本)，和/或参比 DR 序列中插入、缺失、和/或替换(功能性替换版本)一个或多个核苷酸后，依然具有参比 DR 的至少 20%(如至少约任何 30%、40%、50%、60%、70%、80%、90%、95%、或更高)功能的 DR 序列，即介导 Cas12i 蛋白与相对应的 crRNA 结合的功能。DR 功能性变体一般保留可供 Cas12i 蛋白结合的茎环样二级结构或其部分。在一些实施方式中，DR 或其功能性变体包含一个可供 Cas12i 蛋白结合的茎环样二级结构或其部分。在一些实施方式中，DR 或其功能性变体包含至少两个(比如 2、3、4、5 或更多个)可供 Cas12i 蛋白结合的茎环样二级结构或其部分。

引导编辑指导 prime editing guide RNA (pegRNA)

PEGRNA(相对于标准向导 RNA)改变为包含延伸部分，该延伸部分提供编码单链 DNA 瓣的 DNA 合成模板序列，其与待编辑的靶向内源性 DNA 序列的链同源但包含期望

的一个或多个核苷酸变化，并且在被聚合酶(如，逆转录酶)合成之后掺入靶 DNA 分子中。PEgRNA 已描述于例如 WO2020191246、WO2021226558，其通过整体引用而并入本文。

在不同实施方案中，延伸的向导 RNA 包含(a)向导 RNA 和(b)在向导 RNA 的 5'或 3'端处或在向导 RNA 的分子内位置中的 RNA 延伸。优选地，延伸部分的分子内定位不破坏原间隔区的功能。RNA 延伸可包含(i)包含期望的核苷酸变化的逆转录模板序列，(ii)逆转录引物结合位点，和(iii)任选的接头序列。在不同实施方案中，逆转录模板序列可编码与邻近切口位点的内源性 DNA 序列互补的单链 DNA 瓣，其中单链 DNA 瓣包含期望的核苷酸变化。单链 DNA 瓣可在切口位点置换内源性单链 DNA。在不同实施方案中，掺入靶 DNA 中的期望的核苷酸变化可以是单核苷酸变化(如，转换或颠换)、一个或多个核苷酸的插入、或者一个或多个核苷酸的缺失。

在不同实施方案中，期望的核苷酸变化可以是单核苷酸取代(如，转换或颠换变化)、缺失或插入。例如，期望的核苷酸变化可以是(1)G 至 T 取代，(2)G 至 A 取代，(3)G 至 C 取代，(4)T 至 G 取代，(5)T 至 A 取代，(6)T 至 C 取代，(7)C 至 G 取代，(8)C 至 T 取代，(9)C 至 A 取代，(10)A 至 T 取代，(11)A 至 G 取代，或(12)A 至 C 取代。

PAM

本发明所述的工程化的 Cas12i 核酸酶(或其功能变体)能够识别原间隔区相邻基序(PAM)。在一些实施方式中，所述靶核酸包括 PAM。在一些实施方式中，PAM 位于靶核酸中与引导 RNA 的靶向序列互补的序列的 5'端。在一些实施方式中，PAM 包括核酸序列 5'-TTN-3'、5'-TTH-3'、5'-TTY-3'或 5'-TTC-3'，或者由其组成。在一些实施方式中，适用于本发明所述的工程化的 Cas12i 核酸酶(或其功能变体)的 PAM 包括核酸序列 5'-NNNN-3'(N=A, T, G, C)或者由其组成，比如 5'-NTTN-3'，5'-NTAN-3'，5'-NTCN-3'，5'-NTGN-3'，5'-NATN-3'，5'-NAAN-3'，5'-NACN-3'，5'-NAGN-3'，5'-NCTN-3'，5'-NCAN-3'，5'-NCCN-3'，5'-NCGN-3'，5'-NGTN-3'，5'-NGAN-3'，5'-NGCN-3'，5'-NGGN-3'。在一些实施方式中，所述 PAM 包括核酸序列 5'-TTTN-3'或者由其组成。在一些实施方式中，所述 PAM 包括核酸序列 5'-TTN-3'或者由其组成。

在一些实施方式中，适用于本发明所述的工程化的 Cas12i 核酸酶(或其功能变体)的 PAM 包括核酸序列 5'-TTTA-3'、5'-CTTA-3'、5'-GTTA-3'、5'-ATTA-3'、5'-TTTC-3'、5'-CTTC-3'、5'-GTTC-3'、5'-ATTC-3'、5'-TTTG-3'、5'-CTTG-3'、5'-GTTG-3'、5'-ATTG-3'、5'-TTTT-3'、5'-CTTT-3'、5'-GTTT-3'、5'-ATTT-3'。

工程化的 Cas12i 效应蛋白

本申请提供了工程化的 Cas12i(如 Cas12i2)效应蛋白, 其包含本发明所述任何工程化的 Cas12i 核酸酶或其功能变体(例如 SEQ ID NO: 8 所示 CasXX), 其具有改善的活性, 例如靶标结合、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, 提供了工程化的 Cas12i 效应蛋白(例如, Cas12i 核酸酶、Cas12i 切口酶、Cas12i 融合效应蛋白或分割型(split) Cas12i 效应蛋白), 其包含本文所述工程化的 Cas12i 核酸酶或其功能衍生物(如 dCas12i)中的任一种。在一些实施方案中, 工程化的 Cas12i 效应蛋白包含本发明所述任何工程化的 Cas12i 核酸酶或其功能变体, 或主要由其组成、或由其组成。

还提供了基于本文所述工程化的 Cas12i2 核酸酶或其功能变体(例如 SEQ ID NO: 8 所示 CasXX)中任一种的工程化的 Cas12i 效应蛋白。在一些实施方案中, 所述工程化的 Cas12i 效应蛋白具有酶活性(如 DNA 双链切割活性)。在一些实施方案中, 所述工程化的 Cas12i 效应蛋白有切割靶双螺旋核酸(例如, 双螺旋 DNA)的两条链的核酸酶活性。在一些实施方案中, 所述工程化的 Cas12i 效应蛋白有切口酶活性, 即切割靶双螺旋核酸(例如, 双螺旋 DNA)的单链。在一些实施方案中, 所述工程化的 Cas12i 效应蛋白包含所述工程化的 Cas12i 核酸酶的酶失活突变体。

分割型 Cas12i 效应蛋白

本申请还提供了基于本文所述工程化的 Cas12i 核酸酶或其功能变体(例如 SEQ ID NO: 8 所示 CasXX)中任一种的分割型 Cas12i 效应蛋白。分割型 Cas12i 效应蛋白对于递送可能是有利的。在一些实施方案中, 所述工程化的 Cas12i 效应蛋白被分割成酶的两个部分, 可以将它们重构在一起以提供基本起作用的 Cas12i 效应蛋白。可利用已知方法提供 Cas 效应蛋白, 例如, Cas12 和 Cas9 蛋白的分割形式已描述于例如 WO2016/112242、WO2016/205749 和 PCT/CN 2020 /111057, 其通过整体引用而并入本文。

在一些实施方案中, 提供了分割型 Cas12i 效应蛋白, 其包含第一多肽和第二多肽, 所述第一多肽包含本文所述工程化的 Cas12i 核酸酶或其功能衍生物中任一种的 N 末端部分, 所述第二多肽包含所述工程化的 Cas12i 核酸酶或其功能衍生物中任一种的 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 存在下彼此缔合, 以形成与靶核酸特异性结合的 CRISPR 复合物, 所述靶核酸包含与所述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, 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)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二聚化结构域。在一些实施方案中, 所述分割型 Cas12i 效应蛋白是自诱导的。

本发明所述工程化的 Cas12i 效应蛋白可以以不影响催化结构域的方式进行分割。Cas12i 效应蛋白可以用作核酸酶(包括切口酶)或可以是灭活的酶,其本质上是具有很少或没有催化活性(例如由于其催化结构域中的突变)的 RNA 指导的 DNA 结合蛋白。

5 在一些实施方案中,工程化的 Cas12i 效应蛋白的核酸酶叶和 α -螺旋叶表达为分开多肽。尽管所述核酸酶叶和 α -螺旋叶不自行相互作用,但 RNA 指导序列将它们募集到一个复合物中,所述复合物概括了全长 Cas12i 核酸酶的活性并催化位点特异性的 DNA 切割。在一些实施方案中,经修饰的 RNA 指导序列可以用于通过防止二聚化而消除分割型酶的活性,从而允许诱导型二聚化系统的发展。所述分割型酶描述于例如, Wright, Addison V., et al. "Rational design of a split-Cas9 enzyme complex," Proc. Nat'l. Acad. Sci., 10 112.10 (2015): 2984-2989, 其通过整体引用并入本文。

本文所述分割型 Cas12i 效应蛋白部分可设计为通过将参比工程化的 Cas12i 效应蛋白(例如全长工程化的 Cas12i 核酸酶)在一个分割位置处分开(即,分割)成两半,所述位置是所述参比 Cas12i 效应蛋白的 N 端部分与 C 端部分分开的点。在一些实施方案中,所述 N 末端部分包含所述参比 Cas12i 效应蛋白氨基酸残基 1 至 X,而所述 C 末端部分包含氨基
15 酸残基 X+1 至所述参比 Cas12i 效应蛋白的 C 末端。在该实例中,编号是连续的,但这并非必需的,因为还考虑了氨基酸(或编码它们的核苷酸)可以从分割的末端和/或突变(例如,插入、删除和替换)中任一个在所述多肽链的内部区域修剪而来,条件是重构的工程化 Cas12i 效应蛋白保留了足够的 DNA 结合活性(如果需要)、DNA 切口酶或切割活性,例如
20 与所述参比 Cas12i 效应蛋白相比,具有至少约 40%(如至少约 50%、60%、70%、80%、90%、95%或更高)的活性。

可以通过计算机(in silico)设计分割点并将其克隆到构建体中。在此过程中,可以将突变引入分割型 Cas12i 效应蛋白,并且可以去除非功能结构域。在一些实施方案中,所述分割型 Cas12i 效应蛋白的两个部分或片段(即, N 末端和 C 末端片段)可以形成完整的 Cas12i 效应蛋白,其包含例如完整 Cas12i 效应蛋白序列的至少约 70%(如至少约 80%、
25 90%、95%、96%、97%、98%、99%或更高)。

所述分割型 Cas12i 效应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分割型 Cas12i 效应蛋白部分融合的第一二聚结构域,并且所述第二多肽包含与第二分割型 Cas12i 效应蛋白部分融合的第二二聚结构域。可通过肽接头(例如,柔性肽接头如 GS 接头)或化学键将所述二聚化结构域融合至所述分割型
30 Cas12i 效应蛋白部分。在一些实施方案中,所述二聚化结构域与所述分割型 Cas12i 效应蛋白部分的 N 末端融合。在一些实施方案中,所述二聚化结构域与所述分割型 Cas12i 效

应蛋白部分的 C 末端融合。

在一些实施方案中，该分割型 Cas12i 效应蛋白不包含任何二聚化结构域。

在一些实施方案中，所述二聚化结构域促进两个分割型 Cas12i 效应蛋白部分的缔合。在一些实施方案中，所述分割型 Cas12i 效应蛋白部分被诱导剂诱导而缔合或二聚化为功能性 Cas12i 效应蛋白。在 5 一些实施方案中，所述分割型 Cas12i 效应蛋白包含可诱导的二聚化结构域。在一些实施方案中，所述二聚化结构域不是可诱导的二聚化结构域，即，所述二聚化结构域在诱导剂不存在下进行二聚化。

诱导剂可以是除指导 RNA(例如 crRNA)以外的诱导能量源或诱导分子。所述诱导剂通过诱导的二聚化结构域的二聚化将两个分割型 Cas12i 效应蛋白部分重构为功能性 10 Cas12i 效应蛋白。在一些实施方案中，所述诱导剂通过可诱导的二聚化结构域的诱导缔合的作用将两个分割型 Cas12i 效应蛋白部分聚集在一起。在一些实施方案中，在没有诱导剂的情况下，两个分割型 Cas12i 效应蛋白部分不彼此缔合以重构为功能性 Cas12i 效应蛋白。在一些实施方案中，在没有诱导剂的情况下，两个分开的 Cas12i 效应蛋白部分可以在指导 RNA(例如 crRNA)存在下彼此缔合以重构为功能性 Cas12i 效应蛋白。

15 本申请的诱导剂可以是热、超声、电磁能或化合物。在一些实施方案中，所述诱导剂是抗生素、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，所述诱导剂是脱落酸(ABA)、多西霉素(DOX)、异丙基苯甲酸(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，所述分割型 Cas12i 效应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、 20 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，所述分割型 Cas12i 效应系统是选自下组的诱导剂控制的系统：四环素(Tet)/DOX 诱导系统、光诱导系统、ABA 诱导系统、异丙基苯甲酸(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。这样的诱导剂也讨论于本文和 PCT/US2013/051418，其通过整体引用 25 而并入本文。FRB/FKBP/雷帕霉素系统已描述于 Paulmurugan and Gambhir, Cancer Res, August 15, 2005 65; 7413，以及 Crabtree *et al.*, Chemistry & Biology 13, 99-107, Jan 2006，其通过整体引用而并入本文。

在一些实施方案中，成对的分割型 Cas12i 效应蛋白是分开的并且是无活性的，直到诱导了所述二聚化结构域的二聚化(例如，FRB 和 FKBP)，该二聚化导致功能性 Cas12i 30 效应蛋白核酸酶的再组装。在一些实施方案中，包含诱导型二聚体(例如 FRB)的第一部分的第一分割型 Cas12i 效应蛋白被分开递送，和/或处在与包含诱导型二聚体(例如

FKBP)的第二半部分的第二分割型 Cas12i 效应蛋白分开的位置。

可用于本文所述的诱导剂控制的分割型 Cas12i 效应系统中的其他示例性的基于 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(CNA)进行二聚化的 FKBP；在 FKCsA 存在下与 CyP-Fas 进行二聚化的 FKBP；在雷帕霉素存在下与 FRB 进行二聚化的 FKBP；在香豆霉素存在下与 GryB 进行二聚化的 GyrB；在赤霉素存在下与 GID1 进行二聚化的 GAI；或在 HaXS 存在下与 HaloTag 进行二聚化的 Snap-tag。

也考虑了 FKBP 家族本身内的替代方案。例如，在 FK1012 存在下 FKBP 进行均二聚(即，一个 FKBP 与另一种 FKBP 进行二聚化)。

在一些实施方案中，所述二聚化结构域是 FKBP，而诱导剂是 FK1012。在一些实施方案中，所述二聚化结构域是 GryB，而诱导剂是香豆霉素。在一些实施方案中，所述二聚化结构域是 ABA，而诱导剂是赤霉素。

在一些实施方案中，可以在不存在诱导剂的情况下自动诱导(即，自动激活或自诱导)所述分割型 Cas12i 效应蛋白部分以缔合/二聚化为功能性 Cas12i 效应蛋白。不受任何理论或假设的束缚，可以通过与指导 RNA 如 crRNA 结合来介导所述分割型 Cas12i 效应蛋白部分的自动诱导。在一些实施方案中，所述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，所述第一多肽和第二多肽包含二聚化结构域。

在一些实施方案中，本文所述分割型 Cas12i 效应系统(包括诱导剂控制的和自动诱导的系统)的重构的 Cas12i 效应蛋白，相对于参比 Cas12i 效应蛋白编辑效率，具有至少约 60% (如至少约 70%、75%、80%、85%、90%、95%、98%、99%或更高中任一种)的编辑效率。

在一些实施方案中，本文所述诱导剂控制的分割型 Cas12i 效应系统的重构的 Cas12i 效应蛋白，在没有诱导剂存在下(即由于自动诱导)，相对于参比 Cas12i 效应蛋白编辑效率，其具有不超过约 50%(如不超过约 45%、40%、35%、30%、25%、20%、15%、10%、5%或更低效率中任一个)的编辑效率。

25 **融合 Cas12i 效应蛋白**

本申请还提供了工程化的 Cas12i 效应蛋白，其包含另外的蛋白结构域和/或组分，如接头、核定位/输出序列、功能结构域和/或报告蛋白。

在一些实施方案中，所述工程化的 Cas12i 效应蛋白是包含一个或多个异源蛋白结构域(例如，约或大于约 1、2、3、4、5、6、7、8、9、10 个或更多个结构域)以及任一本发明所述工程化的 Cas12i 核酸酶的核酸靶向结构域或其功能衍生物的蛋白复合物。在一些实施方案中，所述工程化的 Cas12i 效应蛋白是包含与所述工程化的 Cas12i 核酸酶或其功

能变体(例如 SEQ ID NO: 8 所示 CasXX)融合的一个或多个异源蛋白结构域(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)的融合蛋白。

5 在一些实施方案中, 本申请的工程化的 Cas12i 效应蛋白可包含(例如, 通过融合蛋白, 如通过一个或多个肽接头, 例如 GS 肽接头等) 一个或多个功能结构域或缔合(例如, 通过多种蛋白的共表达)于它。在一些实施方案中, 所述一个或多个功能结构域是酶结构域。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、核苷酸脱氨酶活性(如腺苷脱氨酶活性、胞苷脱氨酶活性)、脱甲基酶活性、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(如双链内切酶活性、切口酶活性)、核酸结合活性和开关活性(例如, 光诱导或化学诱导的)。在一些
10 实施方案中, 所述一个或多个功能结构域是转录激活结构域(即, 反式激活结构域)或阻遏物结构域。在一些实施方案中, 所述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, 所述一个或多个功能结构域是转座酶结构域、HR(同源重组)机构结构域、重组酶结构域和/或整合酶结构域。在一些实施方案中, 所述功能结构域是 Krüppel 相关盒(KRAB)、VP64、VP16、Fok1、P65、HSF1、MyoD1、生物素-APEX、APOBEC1、
15 AID、PmCDA1、Tad1 和 M-MLV 逆转录酶。在一些实施方案中, 所述功能结构域选自下组: 翻译起始结构域、转录阻遏结构域、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、报告分子结构域(例如, 荧光结构域)和核酸酶结构域。

20 在一些实施方式中, 所述功能结构域具有修饰靶 DNA 或靶 DNA 相关蛋白的活性, 所述活性选自核酸酶活性(例如, HNH 核酸酶, RuvC 核酸酶, Trex1 核酸酶, Trex2 核酸酶)、甲基化活性、脱甲基化活性、DNA 修复合活性、DNA 损伤活性、脱氨基活性、歧化酶活性、烷基化活性、脱嘌呤活性、氧化活性、嘧啶二聚体形成活性、整合酶活性、转座酶活性、重组酶活性、聚合酶活性、连接酶活性、解旋酶活性、光裂合酶活性、糖基化酶活性、乙酰基转移酶活性、脱乙酰酶活性、激酶活性、磷酸酶活性、泛素连接酶活
25 性、去泛素化活性、腺苷酸化活性、脱腺苷酸化活性、SUMO 化活性、脱 SUMO 化活性、核糖基化活性、脱核糖基化活性、豆蔻酰化活性、脱豆蔻酰化活性、糖基化活性(例如, 来自 O-GlcNAc 转移酶)、脱糖基化活性、转录抑制活性和转录激活活性中的一种或多种。靶 DNA 相关蛋白是指可结合靶 DNA 的蛋白, 或可与可结合靶 DNA 的蛋白相结合的蛋白, 比如组蛋白、转录因子、Mediator 等。

30 在一些实施方案中, 所述工程化的 Cas12i 效应蛋白中一个或多个功能结构域的定位允许功能结构域的正确空间定向以影响具有所赋予功能作用的靶标。例如, 如果所述功

能结构域是转录激活子(例如, VP16、VP64 或 p65), 则将所述转录激活子放置在使其能够影响靶标转录的空间取向上。同样地, 定位转录阻遏物以影响靶标的转录, 并且将核酸酶(例如, Fok1)定位以切割或部分切割所述靶标。在一些实施方案中, 所述功能结构域位于所述工程化的 Cas12i 效应蛋白的 N 末端。在一些实施方案中, 所述功能结构域位于工程化的 Cas12i 效应蛋白的 C 末端。在一些实施方案中, 所述工程化的 Cas12i 效应蛋白在 N 末端包含第一功能结构域, 并在 C 末端包含第二功能结构域。在一些实施方案中, 所述工程化 Cas12i 效应蛋白包含与一个或多个功能结构域融合的本文所述工程化的 Cas12i 核酸酶中任一种的催化失活突变体(dCas12i)。

在一些实施方案中, 所述工程化的 Cas12i 效应蛋白是转录激活子。在一些实施方案中, 所述工程化的 Cas12i 效应蛋白包含与反式激活结构域融合的本文所述工程化的 Cas12i 核酸酶中任一种的酶失活变体。在一些实施方案中, 所述反式激活域选自下组: VP64、p65、HSF1、VP16、MyoD1、HSF1、RTA、SET7/9 及其组合。在一些实施方案中, 所述反式激活结构域包含 VP64、p65 和 HSF1。在一些实施方案中, 所述工程化的 Cas12i 效应蛋白包含两个分割型 Cas12i 效应多肽, 每个都与反式激活结构域融合。

在一些实施方案中, 所述工程化的 Cas12i 效应蛋白是转录阻遏物。在一些实施方案中, 所述工程化的 Cas12i 效应蛋白包含与转录阻遏结构域融合的本文所述工程化的 Cas12i 核酸酶中任一种的酶失活变体。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, 所述工程化的 Cas12i 效应蛋白包含两个分割型 Cas12i 效应多肽, 每个都与转录阻遏结构域融合。

在一些实施方案中, 所述工程化的 Cas12i 效应蛋白是碱基编辑器, 如胞嘧啶编辑器或腺苷编辑器。在一些实施方案中, 所述工程化的 Cas12i 效应蛋白包含与核碱基编辑结构域融合的本文所述任何工程化的 Cas12i 核酸酶中任一种的酶失活变体, 所述核碱基编辑结构域诸如为胞嘧啶碱基编辑(CBE)结构域或腺苷碱基编辑(ABE)域。在一些实施方案中, 所述核碱基编辑结构域是 DNA 编辑结构域。在一些实施方案中, 所述核碱基编辑结构域具有脱氨酶活性。在一些实施方案中, 所述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, 所述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描述于例如, WO2018/165629A1 和 WO2019/226953A1, 其通过整体引用而并入本文。示例性 CBE 结构域包括但不限于: 活化诱导的胞苷脱氨酶或 AID (例如 hAID)、载脂蛋白 B mRNA 编辑复合物或 APOBEC(例如大鼠 APOBEC1、hAPOBEC3 A/B/C/D/E/F/G)和 PmCDA1。示例性 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; and Richter *et al.*, 2020, Nature Biotechnology 38: 883-891)。在一些实施方案中, 所述功能性结构域是 APOBEC1 结构域, 例如大鼠 APOBEC1 结构域。在一些实施方案中, 所述功能性结构域是 TadA 结构域, 例如大肠杆菌(*E. coli*) TadA 结构域。在一些实施方案中所述工程化的 Cas12i 效应蛋白还包含一个或多个核定位序列。

腺苷脱氨酶

如本文所用, 术语“腺苷脱氨酶”或“腺苷脱氨酶蛋白”是指蛋白质, 多肽, 或蛋白质或多肽的一个或多个功能结构域, 其能够催化将腺嘌呤(或分子的腺嘌呤部分)转化为次黄嘌呤(或分子的次黄嘌呤部分)的水解脱氨反应, 如下所示。在一些实施方式中, 含腺嘌呤的分子是腺苷(A), 并且含次黄嘌呤的分子是肌苷(I)。含腺嘌呤的分子可以是脱氧核糖核酸(DNA)或核糖核酸(RNA)。

可与本发明工程化的 Cas12i 核酸酶中任一种的酶失活变体结合使用的腺苷脱氨酶包括但不限于称为作用于 RNA 的腺苷脱氨酶的酶家族成员(ADAR), 称为作用于 tRNA 的腺苷脱氨酶的酶家族成员(ADAT), 以及其他含腺苷脱氨酶结构域(ADAD)的家族成员。腺苷脱氨酶能够靶向 RNA/DNA 和 RNA 双链体中的腺嘌呤。实际上, Zheng 等人 (Nucleic Acids Res. 2017, 45(6): 3369-3377)证实 ADAR 可对 RNA/DNA 和 RNA/RNA 双链体进行腺苷至肌苷的编辑反应。在一些实施方式中, 腺苷脱氨酶可以被修饰以增加其编辑 RNA 双链体的 RNA/DNA 异源双链体中的 DNA 的能力。

在一些实施方式中, 腺苷脱氨酶源自一种或多种后生动物物种, 包括但不限于哺乳动物、鸟类、青蛙、鱿鱼、鱼、蝇和蠕虫。在一些实施方式中, 腺苷脱氨酶是人类、鱿鱼或果蝇腺苷脱氨酶。

在一些实施方式中, 腺苷脱氨酶是人类 ADAR, 包括 hADAR1、hADAR2、hADAR3。在一些实施方式中, 腺苷脱氨酶是秀丽隐杆线虫(*Caenorhabditis elegans*) ADAR 蛋白, 包括 ADR-1 和 ADR-2。在一些实施方式中, 腺苷脱氨酶是果蝇 ADAR 蛋白, 包括 dAdar。在一些实施方式中, 腺苷脱氨酶是鱿鱼(长翼鱿鱼(*Loligo pealeii*)) ADAR 蛋白, 包括 sqADAR2a 和 sqADAR2b。在一些实施方式中, 腺苷脱氨酶是人类 ADAT 蛋白。在一些实施方式中, 腺苷脱氨酶是果蝇 ADAT 蛋白。在一些实施方式中, 腺苷脱氨酶是人类 ADAD 蛋白, 包括 TENR (hADAD1)和 TENRL (hADAD2)。在一些实施方式中, 腺苷脱氨酶是 TadA8e。

在一些实施方式中, 腺苷脱氨酶是 TadA 蛋白, 例如大肠杆菌 TadA。参见 Kim 等人, Biochemistry 45:6407-6416 (2006); Wolf 等人, EMBO J. 21:3841-3851 (2002)。在一些实

实施方式中，腺苷脱氨酶是小鼠 ADA。参见 Grunebaum 等人，Curr. Opin. Allergy Clin. Immunol. 13:630-638 (2013)。在一些实施方式中，腺苷脱氨酶是人类 ADAT2。参见 Fukui 等人，J. Nucleic Acids 2010:260512 (2010)。在一些实施方式中，脱氨酶(例如腺苷或胞苷脱氨酶)是以下文献中描述的那些中的一种或多种：Cox 等人，Science. 2017 年 11 月 24 日; 358(6366): 1019-1027; Komore 等人，Nature. 2016 年 5 月 19 日; 533(7603):420-4; 以及 Gaudelli 等人，Nature. 2017 年 11 月 23 日; 551(7681):464-471。

在一些实施方式中，腺苷脱氨酶蛋白包含一个或多个脱氨酶结构域。不希望受特定理论的束缚，预期脱氨酶结构域用于识别双链核酸底物中所含的一个或多个靶腺苷(A)残基并将其转化为肌苷(I)残基。

10 胞苷脱氨酶

在一些实施方式中，脱氨酶是胞苷脱氨酶。如本文所用，术语“胞苷脱氨酶”或“胞苷脱氨酶蛋白”是指蛋白质、多肽或者蛋白质或多肽的一个或多个功能结构域，其能够催化将胞嘧啶(或分子的胞嘧啶部分)转化为尿嘧啶(或分子的尿嘧啶部分)的水解脱氨基反应。在一些实施方式中，含胞嘧啶的分子是胞苷(C)，并且含尿嘧啶的分子是尿苷(U)。所述含胞嘧啶的分子可以是脱氧核糖核酸(DNA)或核糖核酸(RNA)。

可与本发明工程化的 Cas12i 核酸酶中任一种的酶失活变体结合使用的胞苷脱氨酶包括但不限于被称为载脂蛋白 B mRNA 编辑复合物(APOBEC)家族脱氨酶的酶家族的成员，激活诱导的脱氨酶(AID)，或胞苷脱氨酶 1 (CDA1)。在一些实施方式中，APOBEC1 脱氨酶、APOBEC2 脱氨酶、APOBEC3A 脱氨酶、APOBEC3B 脱氨酶、APOBEC3C 脱氨酶和 APOBEC3D 脱氨酶、APOBEC3E 脱氨酶、APOBEC3F 脱氨酶、APOBEC3G 脱氨酶、APOBEC3H 脱氨酶或 APOBEC4 脱氨酶中的脱氨酶。

在一些实施方式中，胞苷脱氨酶能够靶向 DNA 单链中的胞嘧啶。在一些实施方式中，胞苷脱氨酶可在存在于结合组分外部的单链上进行编辑。在一些实施方式中，胞苷脱氨酶可在局部化泡，例如由靶标编辑位点处但指导序列错配形成的局部化泡处编辑。在一些实施方式中，胞苷脱氨酶可包含有助于聚焦活性的突变，例如 Kim 等人，Nature Biotechnology (2017) 35(4):371-377 (doi:10.1038/nbt.3803 中所述的那些)。

在一些实施方式中，胞苷脱氨酶源自一种或多种后生动物物种，包括但不限于哺乳动物、鸟类、青蛙、鱿鱼、鱼、蝇和蠕虫。在一些实施方式中，胞苷脱氨酶是人类、灵长类、牛、狗、大鼠或小鼠胞苷脱氨酶。

30 在一些实施方式中，胞苷脱氨酶是人类 APOBEC，包括 hAPOBEC1 或 hAPOBEC3。在一些实施方式中，胞苷脱氨酶是人类 AID。

在一些实施方式中，胞苷脱氨酶蛋白包含一个或多个脱氨酶结构域。不希望受理论的束缚，预期脱氨酶结构域用于识别 RNA 双链体的单链泡中所含的一个或多个靶胞嘧啶(C)残基并将其转化为尿嘧啶(U)残基。

5 在一些实施方案中，所述工程化的 Cas12i 效应蛋白是主编辑器。基于 Cas9 的主编辑器描述于例如，A. Anzalone *et al.*, Nature, 2019, 576 (7785): 149-157，其通过整体引用而并入本文。在一些实施方案中，所述工程化的 Cas12i 效应蛋白包含与逆转录酶结构域融合的本文所述工程化的 Cas12i 核酸酶中任一种的切口酶变体。在一些实施方案中，所述功能结构域是逆转录酶结构域。在一些实施方案中，所述逆转录酶结构域是 M-MLV 逆转录酶或其变体，例如具有 D200N、T306K、W313F、T330P 和 L603W 的一个或多个突
10 变的 M-MLV 逆转录酶。在一些实施方案中，提供了包含所述主编辑器的工程化 CRISPR/Cas12i 系统。在一些实施方案中，所述工程化的 CRISPR/Cas12i 系统还包含第二 Cas12i 切口酶，例如基于与主编辑器的相同的工程化的 Cas12i 核酸酶。在一些实施方案中，所述工程化的 CRISPR/Cas12i 系统包含引导编辑指导 RNA (pegRNA)，其包含引物结合位点和逆转录酶(RT)模板序列。

15 在一些实施方案中，本申请提供了具有与分割型 Cas12i 效应蛋白部分之一或两者结合(即，结合或融合)的一个或多个(例如，1、2、3、4、5、6 个或更多个)功能结构域的分割型 Cas12i 效应系统。所述功能结构域可以作为所第一和/或第二分割型 Cas12i 效应蛋白的一部分提供，作为该构建体内的融合物。所述功能结构域通常通过肽接头(诸如 GS 接头)与分割型 Cas12i 效应蛋白中其他部分融合(例如，分割型 Cas12i 效应蛋白部分)。这
20 些功能结构域可用于基于催化失活 Cas12i 效应蛋白重新改换该分割型 Cas12i 效应系统的功能。

在一些实施方案中，所述工程化的 Cas12i 效应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 序列包括例如 PKKKRKVPG (SEQ ID NO: 66) 和 ASPKKKRKV (SEQ ID NO: 67)。NLS 和/或 NES 可以可操作地连接至所述工
25 程化的 Cas12i 效应蛋白的 N 末端和/或 C 末端或所述工程化的 Cas12i 效应蛋白中的多肽链。在一些实施方式中，NLS 和/或 NES 可以连接至本发明所述任何工程化的 Cas12i 核酸酶或其功能变体的 N 末端和/或 C 末端。

30 在一些实施方案中，所述工程化的 Cas12i 效应蛋白可以编码另外的组分，例如报告蛋白。在一些实施方案中，所述工程化的 Cas12i 效应蛋白包含荧光蛋白，例如 GFP。这样的系统可以允许对基因组位点进行成像(例如，参见“Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System” Chen B *et al.* Cell 2013)。在一

些实施方案中所述工程化的 Cas12i 效应蛋白是可用于对基因组位点成像的可诱导分割型 Cas12i 效应系统。

在一些实施方案中，提供了一种工程化的 Cas12i 效应蛋白，其中所述效应蛋白能够诱导 DNA 分子中的双链断裂或单链断裂。

- 5 在一些实施方案中，提供了一种工程化的 Cas12i 效应蛋白，其中所述工程化的 Cas12i 核酸酶的功能衍生物为酶失活突变体，例如含有 D599A、E833A、S883A、H884A、D886A、R900A 和/或 D1019A 的 Cas12i2 核酸酶失活突变体(所述氨基酸位置如 SEQ ID NO: 1 所示的相应氨基酸位置所定义)，和含有 D647A、E894A 和/或 D948A 的 Cas12i1 核酸酶失活突变体(所述氨基酸位置如 SEQ ID NO: 13 所示的相应氨基酸位置所定
- 10 义)。已知的 Cas12i2 核酸酶的酶失活突变体，如在 US10808245B2 和 Huang X. *et al.*, Nature Communications, 11, Article number: 5241 (2020)中描述的 Cas12i2 核酸酶的任何酶失活突变都可以和本申请中的突变组合在一起以提供工程化的 Cas12i 核酸酶的功能衍生物及其相应的效应蛋白。

15 工程化的 CRISPR-Cas12i 系统

在一些实施方式中，提供了一种工程化的 CRISPR-Cas12i 系统，包括：(a)本申请所述的任一种工程化的 Cas12i 效应蛋白(如工程化 Cas12i 核酸酶或其功能变体，例如 SEQ ID NO: 8 所示 CasXX)；以及(b)包含与靶序列互补的指导序列的指导 RNA，或编码所述指导 RNA 的一种或多种核酸；

- 20 其中所述工程化的 Cas12i 效应蛋白和所述指导 RNA 能够形成 CRISPR 复合物，所述 CRISPR 复合物特异性结合包含所述靶序列的靶核酸并诱导所述靶核酸的修饰(如双链或单链切割、碱基编辑等)。在本说明书的上下文中，术语“修饰”涵盖核酸酶对于核酸双链或单链上的靶位点处的切割、碱基编辑、替换、修复等。

- 在一些实施方式中，所述工程化的 CRISPR-Cas12i 系统包括：(a)本文所述工程化的
- 25 Cas12i 效应蛋白中任一种(例如，任一所述工程化的 Cas12i 核酸酶或其功能变体，或基于该工程化的 Cas12i 核酸酶或其功能变体的切口酶、分割型 Cas12i、转录阻遏物、转录激活子、碱基编辑器或主编辑器)；以及(b)包含与靶序列互补的指导序列的指导 RNA，或编码所述指导 RNA 的一种或多种核酸；其中所述工程化的 Cas12i 效应蛋白和所述指导
- 30 RNA 能够形成 CRISPR 复合物，所述 CRISPR 复合物特异性结合于包含所述靶序列的靶核酸并诱导所述靶核酸的修饰(如双链或单链切割、碱基编辑等)。在一些实施方案中，所述工程化的 CRISPR-Cas12i 系统包含编码所述工程化的 Cas12i 效应蛋白(如工程化 Cas12i

核酸酶或其功能变体, 例如 SEQ ID NO: 8 所示 CasXX)和/或所述指导 RNA 的一种或多种核酸。在一些实施方案中, 所述工程化的 CRISPR-Cas12i 系统包含前体指导 RNA 阵列, 其可以例如通过所述工程化的 Cas12i 效应蛋白加工成多个 crRNA。在一些实施方案中, 所述工程化的 CRISPR-Cas12i 系统包含一种或多种编码所述工程化的 Cas12i 效应蛋白和/或所述指导 RNA 的载体。在一些实施方案中, 所述工程化的 CRISPR-Cas12i 系统包含核糖核蛋白(RNP)复合物, 其包含结合至所述指导 RNA 的所述工程化的 Cas12i 效应蛋白。

本申请的工程化的 CRISPR-Cas12i 系统可包含任何合适的指导 RNA。指导 RNA(gRNA)可以包含能够与目标靶核酸(诸如细胞中的目标基因组位点)中的靶序列杂交的指导序列。在一些实施方案中, 所述 gRNA 包含有所述指导序列的 CRISPR RNA(crRNA)序列。

通常, 本文所述的 crRNA 包括直接重复序列和间隔区序列。在某些实施方案中, 所述 crRNA 包括与指导序列或间隔区序列连接的直接重复序列、基本由或由它组成。在一些实施方案中, 所述 crRNA 包括直接重复序列、间隔区序列和直接重复序列(DR-间隔区序列-DR), 其是前体 crRNA (pre-crRNA)构型的典型特征。在一些实施方案中, 所述 crRNA 包括截短的直接重复序列和间隔区序列, 其是经加工的或成熟的 crRNA 的典型特征。在一些实施方案中, 所述 CRISPR-Cas12i 效应蛋白与 RNA 指导序列形成复合物, 并且所述间隔区序列将所述复合物引导至与靶核酸进行序列特异性结合, 所述靶核酸与间隔区序列互补(如至少 70%互补)。

在一些实施方案中, 所述指导 RNA 是包含指导序列的 crRNA。在一些实施方案中, 所述工程化的 CRISPR-Cas12i 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 Cas12i 效应蛋白切割所述前体指导 RNA 阵列以产生多个 crRNA。在一些实施方案中, 所述工程化的 CRISPR-Cas12i 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。

构建体和载体

本文还提供了编码本文所述工程化的 Cas12i 效应蛋白(例如工程化 Cas12i 核酸酶或其功能变体)中任一种的构建体、载体和表达系统。在一些实施方案中, 所述构建体、载体或表达系统还包含一种或多种 gRNA 或 crRNA 阵列。

“载体”是包含分离的核酸并且可以用于将所述分离的核酸递送至细胞内部的物质组合物。许多载体是本领域已知的, 包括但不限于: 线性多核苷酸、与离子或两亲性化合物缔合的多核苷酸、质粒和病毒。通常, 合适的载体包含在至少一种生物中起作用的复制起点、启动子序列、方便的限制性核酸内切酶位点和一种或多种选择性标记物。术语

“载体”也应被解释为包括非质粒和非病毒化合物，其促进核酸转移到细胞中，诸如例如，聚赖氨酸化合物、脂质体等。

5 在一些实施方案中，所述载体是病毒载体。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、牛痘载体、单纯疱疹病毒载体及其衍生物。在一些实施方案中，所述载体是噬菌体载体。病毒载体技术是本领域众所周知的，并且例如描述于 Sambrook 等人(2001, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, New York), 以及其他病毒学和分子生物学手册。

10 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，逆转录病毒为基因递送系统提供了便利的平台。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，并在体外或离体递送至所述工程化的哺乳动物细胞。许多逆转录病毒系统是本领域已知的。在一些实施方案中，使用腺病毒载体。许多腺病毒载体是本领域已知的。在一些实施方案中，使用慢病毒载体。在一些实施方案中，使用自灭活的慢病毒载体。

15 在某些实施方案中，所述载体是腺相关病毒(AAV)载体，例如 AAV2、AAV8 或 AAV9，其可以按包含至少 1×10^5 个颗粒(也称为颗粒单位，pu)的单次给药施用腺病毒或腺相关病毒。在一些实施方案中，所述给药量是至少约 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。递送方法和给药量描述于例如 WO 2016205764 和美国专利 No.8,454,972，其通过整体引用而并入本文。

20 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，可以使用经修饰的 AAV 载体递送。经修饰的 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，经修饰的 AAV 载体(例如经修饰的 AAV2，经修饰的 AAV3，经修饰的 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 Nos.4,797,368 和 5,173,414; 国际公开第 WO2015/054653 号和第 WO93/24641 号，其各自通过引用并入本文)。

30 用于递送 Cas9 和其他 Cas 蛋白的任何已知的 AAV 载体都可以用于递送本申请的工程化的 Cas12i 系统。

在一些实施方案中，rAAV 构建体可经肠内向受试者施用。在一些实施方案中，

rAAV 构建体可经肠胃外向受试者施用。在一些实施方案中，rAAV 颗粒可经皮下、眼内、玻璃体内、视网膜下、静脉内(IV)、脑室内、肌内、鞘内(IT)、脑池内、腹膜内、经由吸入、局部或通过直接注射到一种或多种细胞、组织或器官。在一些实施方案中，rAAV 颗粒可通过注射到肝动脉或门静脉中向受试者施用。

5 将载体引入哺乳动物细胞的方法是本领域已知的。可以通过物理、化学或生物学方法将载体转移到宿主细胞中。

用于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。产生包含载体和/或外源核酸的细胞的方法是本领域众所周知的。参见，例如 Sambrook *et al.* (2001) *Molecular Cloning: A Laboratory Manual*, Cold Spring

10 Harbor Laboratory, New York。在一些实施方案中，通过电穿孔将所述载体引入所述细胞。

用于将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细胞的最广泛使用的方法。

用于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，所述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
15 施方案中，所述工程化的 CRISPR-Cas12i 系统以 RNP 的形式在纳米颗粒中递送。

在一些实施方案中，编码所述 CRISPR-Cas12i 系统或其组分的载体或表达系统包含一种或多种可选择或可检测的标记物，该标记物提供了分离或有效选择含有和/或已被 CRISPR-Cas12i 系统(例如在早期和大规模)修饰的细胞的手段。

20 报告基因可用于鉴定潜在转染的细胞和评估调节序列的功能。通常，报告基因是在受体生物体或组织中不存在或不表达的基因，其编码的多肽的表达是通过某些易于检测的性质(例如酶活性)来证明。在将 DNA 引入受体细胞后的合适时间测定报告基因的表达。合适的报告基因可以包括编码荧光素酶、 β -半乳糖苷酶、氯霉素乙酰转移酶、分泌的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))。

25 确认宿主细胞中异源核酸的存在的方法包括例如本领域技术人员众所周知的分子生物学测定，诸如 Southern 和 Northern 印迹、RT-PCR 和 PCR；生化测定，例如通过免疫学方法(如 ELISA 和 Western 印迹)检测特定肽的存在或不存在。

30 在一些实施方案中，编码所述工程化的 Cas12i 效应蛋白和/或所述指导 RNA 的核酸序列与启动子可操作地连接。在一些实施方案中，所述启动子是相对于使用所述工程化的 CRISPR-Cas12i 系统进行工程化改造的细胞的内源启动子。例如，可以使用本领域已知的任何方法将编码所述工程化的 Cas12i 效应蛋白的核酸敲入工程化的哺乳动物细胞的

基因组中位于内源性启动子下游处。在一些实施方案中，所述内源启动子是丰富蛋白(如β-肌动蛋白)的启动子。在一些实施方案中，所述内源启动子是可诱导的启动子，例如，可通过工程化的哺乳动物细胞的内源激活信号来诱导。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(诸如IL-2启动子、NFAT启动子或NFκB启动子)。

在一些实施方案中，所述启动子是相对于使用所述工程化的CRISPR-Cas12i系统进行工程化改造的细胞的异源启动子。已经探索了多种启动子以在哺乳动物细胞中表达基因，并且本领域中已知的任何启动子都可以用于本申请中。启动子可大致分为组成型启动子或受调节的启动子，例如诱导型启动子。

在一些实施方案中，编码所述工程化的Cas12i效应蛋白和/或所述指导RNA的核酸序列与组成型启动子可操作地连接。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、人延伸因子-1α(hEF1α)、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)以及鸡β-肌动蛋白启动子与CMV早期增强子(CAG)耦联。在一些实施方案中，所述启动子是CAG启动子，其包含巨细胞病毒(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，以及兔β-珠蛋白基因的剪接受体。

在一些实施方案中，编码所述工程化的CRISPR-Cas12i效应蛋白和/或所述指导RNA的核酸序列可操作地连接至诱导型启动子。诱导型启动子属于受调控的启动子类型。所述诱导型启动子可以通过一种或多种条件如物理条件、微环境或宿主细胞的生理状态、诱导物(即，诱导剂)或其组合来诱导。在一些实施方案中，所述诱导条件选自下组：诱导剂、辐照(如电离辐射、光)、温度(如热)、氧化还原状态、肿瘤环境和待通过工程化的CRISPR-Cas12i系统进行工程化改造的细胞的活化状态。在一些实施方案中，所述启动子可被小分子诱导剂诸如化合物诱导。在一些实施方案中，所述小分子选自下组：强力霉素、四环素、醇、金属或类固醇。化学诱导的启动子已被最广泛地研究。这样的启动子包括其转录活性受小分子化学物质例如强力霉素、四环素、醇、类固醇、金属和其他化合物的存在或不存在的启动子。具有逆四环素受控反式激活因子(rtTA)和四环素响应元件启动子(TRE)的强力霉素诱导系统是目前最成熟的系统。WO9429442描述了通过四环素响应性启动子严格控制真核细胞中基因表达。WO9601313公开了四环素调节的转录调节剂。另外，诸如Tet-on系统之类的Tet技术已描述于例如TetSystems.com网站。在本申请中，任何已知的化学调节的启动子均可用于驱动编码所述工程化的CRISPR-Cas12i蛋白和/或所述指导RNA的表达。

在一些实施方案中，编码所述工程化的 Cas12i 效应蛋白的核酸序列是经过密码子优化的。

在一些实施方案中，提供了表达构建体，其包含经密码子优化的序列，编码所述工程化的 Cas12i 效应蛋白的该序列连接到 BPK2104-ccdB 载体上。在一些实施方案中，所述表达构建体编码与所述工程化的 Cas12i 效应蛋白的 C 末端可操作连接的标签(例如 10×His 标签)。

在一些实施方案中，每种工程化的分割型 Cas12i 构建体编码荧光蛋白诸如 GFP 或 RFP。所述报告蛋白可以用于评估所述工程化的 Cas12i 蛋白的共定位和/或二聚化，例如使用显微镜。可以使用编码自切割肽诸如 T2A、P2A、E2A 或 F2A 肽的序列，将编码工程化的 Cas12i 效应蛋白的核酸序列与编码另外的组分的核酸序列融合。

在一些实施方案中，提供了用于哺乳动物细胞(例如人细胞)的表达构建体，所述构建体包含编码所述工程化的 Cas12i 效应蛋白的核酸序列。在一些实施方案中，所述表达构建体包含经密码子优化的序列，所述序列编码插入 pCAG-2A-eGFP 载体中的所述工程化的 Cas12i 效应蛋白，从而使 Cas12i 蛋白可操作地连接至 eGFP。在一些实施方案中，提供第二载体用于在哺乳动物细胞(例如人细胞)中表达指导 RNA(例如 crRNA 或前体 crRNA 阵列)。在一些实施方案中，编码所述指导 RNA 的序列在 pUC19-U6-i2-cr RNA 载体主链中表达。

在一些实施方式中，将 CRISPR-Cas12i 系统的一个或多个元件表达的一个或多个载体引入宿主细胞，使得 CRISPR-Cas12i 系统的元件的表达引导核酸靶向复合物在一个或多个靶位点的形成。例如，Cas12i 核酸靶向效应酶和核酸靶向指导 RNA 可各自可操作地连接至分开的载体上的分开的调控元件。可将核酸靶向系统的 RNA 递送至转基因 Cas12i 核酸靶向效应蛋白动物或哺乳动物，例如，组成性或诱导性或条件性表达核酸靶向效应蛋白的动物或哺乳动物；或以其他方式表达核酸靶向效应蛋白或者具有含有核酸靶向效应蛋白的细胞的动物或哺乳动物，例如通过事先向其施用编码并表达体内核酸靶向效应蛋白的一个或多个载体。或者，可将由相同或不同调控元件表达的两个或更多个元件组合在单个载体中，而一个或多个额外载体提供不包含在第一载体中的核酸靶向系统的任何组分。组合在单个载体中的核酸靶向系统元件可以任何合适的方向排列，例如一个元件位于相对于第二元件的(“上游”) 5'或相对于第二元件的(“下游”) 3'。一个元件的编码序列可位于第二元件的编码序列的相同或相反链上，并以相同或相反的方向定向。在一些实施方式中，单个启动子驱动编码 Cas12i 核酸靶向效应蛋白和核酸靶向指导 RNA 的转录物的表达，所述转录物嵌入一个或多个内含子序列内(例如，各自在不同的内含子中，

两个或更多个在至少一个内含子中,或全部在单个内含子中)。在一些实施方式中,核酸靶向效应蛋白和核酸靶向指导 RNA 可以可操作地连接至同一启动子并从同一启动子表达。用于表达核酸靶向系统的一个或多个元件的递送媒介物、载体、粒子、纳米粒子、制剂及其组分如前述文件例如 WO 2014/093622 (PCT/US2013/074667)中所使用。在一些实施方式中,载体包含一个或多个插入位点,例如限制性核酸内切酶识别序列(也称为“克隆位点”)。在一些实施方式中,一个或多个插入位点(例如,约或大于约 1、2、3、4、5、6、7、8、9、10 个或更多个插入位点)位于一个或多个载体的一个或多个序列元件的上游和/或下游。当使用多个不同的指导序列时,单个表达构建体可用于将核酸靶向活性靶向细胞内的多个不同的相应靶序列。例如,单个载体可包含约或大于约 1、2、3、4、5、6、7、8、9、10、15、20 个或更多个指导序列。在一些实施方式中,可提供约或大于约 1、2、3、4、5、6、7、8、9、10 个或更多个这样的含指导序列的载体,并任选地递送至细胞。在一些实施方式中,载体包含与编码核酸靶向效应蛋白的酶编码序列可操作地连接的调控元件。Cas12i 核酸靶向效应蛋白或者一种或多种核酸靶向指导 RNA 可分开递送;并且有利地,这些中的至少一种经由粒子复合物递送。可在核酸靶向指导 RNA 之前递送核酸靶向效应蛋白 mRNA,以留出时间表达 Cas12i 核酸靶向效应蛋白。核酸靶向效应蛋白 mRNA 可在施用核酸靶向指导 RNA 之前 1-12 小时(优选约 2-6 小时)施用。或者,核酸靶向效应蛋白 mRNA 和核酸靶向指导 RNA 可一起施用。有利地,可在初次施用核酸靶向效应蛋白 mRNA + 指导 RNA 后 1-12 小时(优选约 2-6 小时)施用指导 RNA 的第二加强剂量。核酸靶向效应蛋白 mRNA 和/或指导 RNA 的其他施用可能对实现最有效的基因组修饰水平有用。

在一些实施方式中,提供了一种 CRISPR-Cas12i 系统,其包含:(1)任一所述的工程化的 Cas12i 效应蛋白(如任一所述工程化的 Cas12i 核酸酶或其功能变体)或编码任一所述工程化的 Cas12i 效应蛋白的多核苷酸;和(2) crRNA 或编码所述 crRNA 的多核苷酸,所述 crRNA 包含:(i)能够与靶 DNA 的靶序列杂交的间隔序列,和(ii)连接至所述间隔序列的能够引导所述工程化的 Cas12i 效应蛋白结合至所述 crRNA 以形成靶向所述靶序列的 CRISPR-Cas12i 复合物的直接重复序列。

在一些实施方式中,提供了一种 CRISPR-Cas12i 系统,其包括一个或多个载体,所述一个或多个载体包含:(1)第一调控元件,所述第一调控元件可操作地连接至编码任一所述工程化的 Cas12i 效应蛋白(如任一所述工程化的 Cas12i 核酸酶或其功能变体)的核苷酸序列;和(2)第二调控元件,所述第二调控元件可操作地连接至编码 crRNA 的多核苷酸,所述 crRNA 包含:(i)能够与靶 DNA 的靶序列杂交的间隔序列,和(ii)连接至所述间隔

列的能够引导所述工程化的 Cas12i 效应蛋白结合至所述 crRNA 以形成靶向所述靶序列的 CRISPR-Cas12i 复合物的直接重复序列；其中所述第一调控元件和所述第二调控元件位于所述 CRISPR-Cas12i 系统的相同或不同载体上。

在某些实施方式中，所述第一调控元件和所述第二调控元件位于所述 CRISPR-Cas12i 系统的不同载体上。在某些实施方式中，所述第一调控元件和所述第二调控元件位于所述 CRISPR-Cas12i 系统的相同载体上。在某些实施方式中，所述第一调控元件和所述编码工程化的 Cas12i 效应蛋白的核苷酸序列在所述第二调控元件和所述编码 crRNA 的多核苷酸的上游。在某些实施方式中，所述第一调控元件和所述编码工程化的 Cas12i 效应蛋白的核苷酸序列在所述第二调控元件和所述编码 crRNA 的多核苷酸的下游。在某些实施方式中，所述第一调控元件和所述第二调控元件相同。在某些实施方式中，所述第一调控元件和所述第二调控元件不同。

在一些实施方式中，提供了一种 CRISPR-Cas12i 系统，其包括一个载体，所述载体包含：(1)编码任一所述工程化的 Cas12i 效应蛋白(如任一所述工程化的 Cas12i 核酸酶或其功能变体)的多核苷酸；(2)编码 crRNA 的多核苷酸，所述 crRNA 包含：(i)能够与靶 DNA 的靶序列杂交的间隔序列，和(ii)连接至所述间隔序列的能够引导所述工程化的 Cas12i 效应蛋白结合至所述 crRNA 以形成靶向所述靶序列的 CRISPR-Cas12i 复合物的直接重复序列；和(3)调控元件，所述调控元件可操作地连接至编码所述工程化的 Cas12i 效应蛋白的多核苷酸和编码所述 crRNA 的多核苷酸。

在某些实施方式中，所述载体从 5'到 3'包含所述调控元件、编码所述工程化的 Cas12i 效应蛋白的多核苷酸和编码所述 crRNA 的多核苷酸。在某些实施方式中，所述载体从 5'到 3'包含所述调控元件、编码所述 crRNA 的多核苷酸和编码所述工程化的 Cas12i 效应蛋白的多核苷酸。在某些实施方式中，编码所述 crRNA 的多核苷酸和编码所述工程化的 Cas12i 效应蛋白的多核苷酸由连接序列连接，例如编码 P2A、T2A、E2A、F2A、BmCPV 2A、BmIFV 2A、(GS)_n (SEQ ID NO: 74)、(GGGS)_n (SEQ ID NO: 75)和 (GGGG)_n (SEQ ID NO: 76)中任一项的多核苷酸序列(其中 n 是至少为 1 的整数)，或任一 IRES、SV40、CMV、UBC、EF1 α 、PGK 和 CAGG 的多核苷酸序列，或其任何组合。

在一些实施方式中，本发明所述 CRISPR-Cas12i 系统的组分可以各种形式递送，例如 DNA/RNA 或 RNA/RNA 或蛋白质 RNA 的组合。例如，工程化的 Cas12i 效应蛋白(如任一所述工程化的 Cas12i 核酸酶或其功能变体)可作为编码 DNA 的多核苷酸或编码 RNA 的多核苷酸或作为蛋白质被递送。所述指导物可作为 DNA 编码多核苷酸或 RNA 被递送。可以使用混合的递送形式。

在一些方面，本发明提供了包括将一个或多个多核苷酸例如如本文所述的一个或多个载体、其一个或多个转录物和/或从其转录的一个或多个蛋白质递送至宿主细胞的方法。

III. 使用方法

5 本申请提供了使用本文所述工程化的 Cas12i 效应蛋白(如任一所述工程化的 Cas12i 核酸酶或其功能变体)或 CRISPR-Cas12i 系统中任一种在体外、离体或体内检测靶核酸或修饰核酸的方法，以及使用所述工程化的 Cas12i 效应蛋白或 CRISPR-Cas12i 系统进行治疗(如基因编辑)或诊断的方法。还提供了本文所述工程化的 Cas12i 效应蛋白或 CRISPR-Cas12i 系统用于检测或修饰细胞中的核酸，以及用于治疗或诊断受试者的疾病或病况的
10 用途；以及包含所述工程化的 Cas12i 效应蛋白中任一种或所述工程化的 CRISPR-Cas12i 系统的一种或多种组分的组合物在制备用于检测或修饰细胞中的核酸以及用于治疗或诊断受试者的疾病或病况的药物中的用途。

检测样品中靶核酸的方法

本申请还提供了使用具有改善的活性的所述工程化的 Cas12i 效应蛋白或 CRISPR-
15 Cas12i 系统中任一种来检测靶核酸的方法。使用 Cas12i 效应蛋白作为检测试剂可利用以下发现：一旦通过检测到靶 DNA 而被激活，V 型 CRISPR/Cas 蛋白(例如，Cas12i)可混杂地切割非靶向单链 DNA(ssDNA 或 RNA，即指导 RNA 的指导序列不与之杂交的单链核酸)。因此，当样品中存在靶 DNA(双链或单链)时(例如，在某些情况下超过阈值量)，结果是样品中单链核酸的切割，这可使用任何方便的检测方法进行检测(例如，使用加标
20 签的单链检测核酸如 DNA 或 RNA)。Cas12i 可以切割 ssDNA 和 ssRNA。例如使用 Cas 蛋白作为检测试剂的方法描述于 US10253365 和 WO2020/056924，其通过整体引用而并入本文。

在一些实施方案中，提供了检测样品中的靶 DNA(例如，双链或单链)的方法，其包括：(a)使所述样品与以下项接触：(i)本文所述工程化的 Cas12i 效应蛋白(如任一所述工程
25 化的 Cas12i 核酸酶或其功能变体)中任一种；(ii)指导 RNA，其包含与所述靶 DNA 杂交的指导序列；和(iii)检测核酸，其为单链的(即“单链检测核酸”)并且不与所述指导 RNA 的指导序列杂交；以及(b)测量通过所述工程化的 Cas12i 效应蛋白切割所述单链检测核酸而产生的可检测信号。在某些情况下，所述单链检测核酸包括发射荧光的染料对(例如，发射荧光的染料对是荧光共振能量转移(FRET)对、猝灭剂/荧光对)。在某些情况下，所述
30 靶 DNA 是病毒 DNA(例如，乳头状病毒、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，所述单链检测核酸是 DNA。在一些实施方案中，所

述单链检测核酸是 RNA。

本公开的用于检测样品中的靶 DNA(单链或双链)的方法, 可以以高灵敏度检测靶 DNA。在一些情况下, 本公开的方法可以用于检测存在于包含多个 DNA(包括所述靶 DNA 和多个非靶 DNA)的样品中的靶 DNA, 其中所述靶 DNA 以每 10^7 个非靶 DNA 中的一个或更多个拷贝存在(例如, 每 10^6 个非靶 DNA 中的一个或更多个拷贝, 每 10^5 个非靶 DNA 中的一个或更多个拷贝, 每 10^4 个非靶 DNA 中的一个或更多个拷贝, 每 10^3 个非靶 DNA 中的一个或更多个拷贝, 每 10^2 个非靶 DNA 中的一个或更多个拷贝, 每 50 个非靶 DNA 中的一个或更多个拷贝, 每 20 个非靶 DNA 中的一个或更多个拷贝, 每 10 个非靶 DNA 中的一个或更多个拷贝或每 5 个非靶 DNA 中的一个或更多个拷贝)。在一些实施方案中, 与所述参比 Cas12i 核酸酶相比, 本文所述工程化的 Cas12i 效应蛋白可以以更高的灵敏度检测靶 DNA。在一些实施方案中, 与所述参比 Cas12i 核酸酶相比, 所述工程化的 Cas12i 效应蛋白可以以 10%、15%、20%、25%、30%、40%、50%、60%、70%、80%、90%、95% 或更高的灵敏度检测靶 DNA。

修饰的方法

在一些实施方案中, 本申请提供了修饰包含靶序列的靶核酸的方法, 所述方法包括使所述靶核酸与本文所述工程化的 CRISPR-Cas12i 系统中任一种接触。在一些实施方案中, 所述方法在体外进行。在一些实施方案中, 所述靶核酸存在于细胞中。在一些实施方案中, 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。在一些实施方案中, 所述方法离体进行。在一些实施方案中, 所述方法在体内进行。靶核酸修饰包括但不限于靶核酸单链切割、双链切割、碱基替换、碱基插入、碱基删除、突变(如致病突变)序列修复等。

在一些实施方案中, 通过所述工程化的 CRISPR-Cas12i 系统切割所述靶核酸或改变所述靶核酸中的靶序列。在一些实施方案中, 通过所述工程化的 CRISPR-Cas12i 系统改变所述靶核酸的表达。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, 例如基于靶序列的错误表达(如过表达、不表达、或表达致病 RNA 或蛋白)。在一些实施方案中, 所述工程化的 CRISPR-Cas12i 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, 本申请提供了处理与个体细胞中的靶核酸相关的疾病或病况的方法, 其包括使用本文所述方法中任一种来修饰所述个体细胞中的所述靶核酸, 从而处理所述疾病或病况。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(例如基因缺陷性疾病, 如镰刀型红细胞贫血(SCD)或 β -地中海贫血(TDT))、

自身免疫疾病、代谢性疾病、神经退行性疾病、眼病、细菌感染和病毒感染。

本文所述工程化的 CRISPR-Cas12i 系统可以用多种方式修饰细胞中的靶核酸，这取决于所述 CRISPR-Cas12i 系统中工程化的 Cas12i 效应蛋白的类型。在一些实施方案中，所述方法诱导靶核酸中的位点特异性切割。在一些实施方案中，所述方法在细胞如细菌
5 细胞、植物细胞或动物细胞(例如哺乳动物细胞)中切割基因组 DNA。在一些实施方案中，所述方法通过切割细胞中的基因组 DNA 来杀死细胞。在一些实施方案中所述方法在细胞中切割病毒核酸。

在一些实施方案中，所述方法改变(如增加或减少)细胞中所述靶核酸的表达水平。在一些实施方案中，所述方法例如基于融合至反式激活结构域的无酶活性的 Cas12i 蛋白，
10 使用工程化的 Cas12i 效应蛋白来提高细胞中所述靶核酸的表达水平。在一些实施方案中，所述方法例如基于融合至转录阻遏结构域的无酶活性的 Cas12i 蛋白，使用工程化的 Cas12i 效应蛋白来降低细胞中所述靶核酸的表达水平。在一些实施方案中，所述方法例如基于融合至表观遗传修饰结构域的无酶活性的 Cas12i 蛋白，使用工程化的 Cas12i 效应蛋白将表观遗传修饰引入细胞中的所述靶核酸或靶核酸相关蛋白(如结合至靶核酸或其临
15 近的蛋白，例如转录因子或组蛋白)。本文所述工程化的 Cas12i 系统可用于将其他修饰引入所述靶核酸，这取决于所述工程化的 Cas12i 效应蛋白所包含的功能结构域。

在一些实施方案中，所述方法改变细胞中所述靶核酸中的靶序列。在一些实施方案中，所述方法将突变引入细胞中的所述靶核酸，比如将致病突变突变为非致病序列(例如，使用与包含腺苷脱氨酶或胞苷脱氨酶融合的任一本发明所述 dCas12i)。在一些实施方案
20 中，所述方法使用一种或多种内源性 DNA 修复途径，如非同源末端连接(NHEJ)或同源定向重组(HDR)，在细胞中修复靶 DNA 中诱导的双链断裂，作为由 CRISPR 复合物进行序列特异性切割的结果。示例性的突变包括但不限于：插入、缺失、替换和移码。在一些实施方案中，使用 pegRNA 来同时进行靶序列切割和提供 DNA 修复模板。在一些实施方案中，所述方法在靶位点处插入供体 DNA。在一些实施方案中，供体 DNA 的插入导
25 致将选择性标记物或报告蛋白引入所述细胞。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 的插入导致敲除突变。在一些实施方案中，供体 DNA 的插入导致替换突变诸如单核苷酸替换。在一些实施方案中，所述方法诱导细胞的表型变化。

在一些实施方案中，所述工程化的 CRISPR-Cas12i 系统用于基因电路(genetic circuit)
30 的一部分，或用于将基因电路插入细胞的基因组 DNA 中。本文所述的诱导剂控制的工程化的分割型 Cas12i 效应蛋白尤其可用作基因电路的组分。基因电路可用于基因疗法。设

计和使用基因电路的方法和技术是本领域已知的。可以进一步参考例如 Brophy, Jennifer AN, and Christopher A. Voigt. "Principles of genetic circuit design." *Nature methods* 11.5 (2014): 508。

本文所述工程化的 CRISPR-Cas12i 系统可用于修饰多种靶核酸。在一些实施方案中，所述靶核酸在细胞中。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，所述靶核酸是染色体外 DNA。在一些实施方案中，所述靶核酸对于细胞是外源的。在一些实施方案中，所述靶核酸是病毒核酸诸如病毒 DNA。在一些实施方案中，所述靶核酸是细胞中的质粒。在一些实施方案中，所述靶核酸是经水平转移(horizontally transferred)的质粒。在一些实施方案中，所述靶核酸是 RNA。

在一些实施方案中，所述靶核酸是经分离的核酸诸如经分离的 DNA。在一些实施方案中，所述靶核酸存在于无细胞的环境中。在一些实施方案中，所述靶核酸是经分离的载体诸如质粒。在一些实施方案中，所述靶核酸是经分离的线性 DNA 片段。

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，所述细胞是细菌、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，如人细胞)。在一些实施方案中，所述细胞是自然来源的诸如由组织活检分离出的细胞。在一些实施方案中，所述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，所述细胞来自原代细胞系。在一些实施方案中，所述细胞来自永生化细胞系。在一些实施方案中，所述细胞是基因工程化的细胞。

在一些实施方案中，所述细胞是选自下组的生物的动物细胞：牛、绵羊、山羊、马、猪、鹿、鸡、鸭、鹅、兔和鱼。

在一些实施方案中，所述细胞是选自下组的生物的植物细胞：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、珍珠小米、粟米、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。

在一些实施方案中，所述细胞是哺乳动物细胞。在一些实施方案中，所述细胞是人细胞。在一些实施方案中，所述人细胞是人胚胎肾 293T (HEK293T 或 293T) 细胞或 HeLa 细胞。在一些实施方案中，所述细胞是人胚肾(HEK293T)细胞。在一些实施方案中，所述细胞是小鼠 Hepa1-6 细胞。在一些实施方案中，所述哺乳动物细胞选自下组：免疫细胞、肝细胞、肿瘤细胞、干细胞、血液细胞、神经细胞、合子、肌肉细胞(如心肌细胞)和皮肤细胞。

在一些实施方案中，所述细胞是选自下组的免疫细胞：细胞毒性 T 细胞、辅助 T 细胞、天然杀伤(NK)T 细胞、iNK-T 细胞、NK-T 样细胞、 $\gamma\delta$ T 细胞、肿瘤浸润性 T 细胞和

树突状细胞(DC)激活的 T 细胞。在一些实施方案中,所述方法产生经修饰的免疫细胞,诸如 CAR-T 细胞或 TCR-T 细胞。

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、合子或胚胎中的细胞。

5 本文所述的方法可以用于在体内、离体或体外修饰靶细胞,并且可以用改变所述细胞的方式进行,以使得一旦被修饰,所述经修饰的细胞的后代或细胞系就保留经改变的表型。所述经修饰的细胞和后代可以是多细胞生物的一部分,诸如具有离体或体内应用(如基因组编辑和基因疗法)的植物或动物。

10 在一些实施方案中,所述方法离体进行。在一些实施方案中,将所述工程化的 CRISPR-Cas12i 系统引入细胞后,所述经修饰的细胞(例如,哺乳动物细胞)离体繁殖。在一些实施方案中,将所述经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 天中任一个。在一些实施方案中,将所述经修饰的细胞培养不超过约 1 天、2 天、3 天、4 天、5 天、6 天、7 天、10 天、12 天或 14 天中任一个。在一些实施方案中,进一步评价或筛选所述经修饰的细胞以选择具有一种或多种所需表型或特性的细胞。

15 在一些实施方案中,所述靶序列是与疾病或病况相关的序列。示例性的疾病或病况包括但不限于:癌症、心血管疾病、遗传性疾病、自身免疫性疾病、代谢性疾病、神经退行性疾病、眼病、细菌感染和病毒感染。在一些实施方案中,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,20 所述疾病或病况是多基因疾病或病况。

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

25 在一些实施方案中,插入所述靶核酸中的供体 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、细胞因子受体和自杀多肽。在一些实施方案中,所述供体 DNA 编码治疗性蛋白。在一些实施方案中,所述供体 DNA 编码可用于基因疗法的治疗性蛋白。在一些实施方案中,所述供体 DNA 编码治疗性抗体。在一些实施方案中,所述供体 DNA 编码工程化的受体,诸如嵌合抗原受体(CAR)或工程化的 TCR。在一些实施方案中,所述供体 DNA 编码治疗性 RNA,诸如小 RNA(例如, siRNA、shRNA 或 miRNA)或长的非编码 RNA (lincRNA)。

30 本文所述的方法可以用于在两个或更多个(例如 2、3、4、5、6、8、10 个或更多)不

同的靶位点处进行多重基因编辑或调节。在一些实施方案中，所述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，所述方法包括使靶核酸与包含多个(例如 2、3、4、5、6、8、10 个或更多) crRNA 序列的指导 RNA 接触，其中每个 crRNA 均包含不同的靶序列。

5 还提供了包含经修饰的靶核酸的工程化的细胞，所述细胞使用本文所述方法中任一种来产生。所述工程化的细胞可以用于细胞疗法。可以使用本文所述用于细胞疗法的方法将自体或同种异体细胞用于制备工程化的细胞。

本文所述的方法还可用于生成细胞(例如哺乳动物细胞)的等基因系以研究遗传变体。

10 还提供了包含本文所述工程化的细胞的工程化非人类动物。在一些实施方案中，所述工程化的非人类动物是经过基因组编辑的非人类动物。所述工程化的非人类动物可以用作疾病模型。

产生非人类基因组编辑或转基因动物的技术是本领域众所周知的，包括但不限于：原核显微注射、病毒感染、对胚胎干细胞和诱导性多能干(iPS)细胞的转化。可以使用的详细方法包括但不限于 Sundberg 和 Ichiki 描述的方法(2006, Genetically Engineered Mice Handbook, CRC Press)和 Gibson 描述的方法(2004, A Primer Of Genome Science 2nd ed. Sunderland, Mass.: Sinauer)。

所述工程化的动物可以是任何合适的物种，包括但不限于：牛、马、羊、犬、鹿、猫科动物，山羊、猪、灵长类，以及了解较少的哺乳动物诸如大象、鹿、斑马或骆驼。

治疗方法

20 在一些实施方式中，提供了前述工程化的 CRISPR-Cas12i 系统在制备治疗与个体的细胞中靶核酸相关的疾病或病症的药物中的用途。在一些实施方式中，提供了前述工程化的 CRISPR-Cas12i 系统在治疗与个体的细胞中靶核酸相关的疾病或病症的方法。

在一些实施方式中，本发明提供了一种治疗有需要的受试者(如人)的疾病的方法，所述方法包括向所述受试者施用(如静脉注射或滴注)本发明的 CRISPR-Cas12i 系统，其包含：
25 (1)任一所述工程化的 Cas12i 效应蛋白(如任一所述工程化的 Cas12i 核酸酶或其功能变体)，或编码所述工程化的 Cas12i 效应蛋白的多核苷酸；和(2)crRNA，或编码所述 crRNA 的多核苷酸，所述 crRNA 包含：(i)能够和与该疾病相关的靶核酸上的靶序列杂交的间隔序列，和(ii)连接至所述间隔序列的能够引导所述工程化的 Cas12i 效应蛋白结合至所述 crRNA 以形成靶向所述靶序列的 CRISPR-Cas12i 复合物的直接重复序列；其中所述间隔序列和
30 所述靶序列的杂交介导所述工程化的 Cas12i 效应蛋白与所述靶序列接触，导致所述工程化的 Cas12i 效应蛋白对所述靶序列进行修饰(例如切割或碱基编辑)，从而治疗所述受试

者的疾病。其中，所述 CRISPR-Cas12i 系统的任一组分可以同时递送，依次递送、或以 DNA/RNA、DNA/DNA、RNA/RNA、蛋白/RNA、或蛋白/DNA 任一形式递送。

进一步提供了使用根据本文所述的修饰细胞中的靶核酸的方法中的任一种的治疗方法。在一些实施方案中，本申请提供了治疗与个体细胞中的靶核酸相关的疾病或病况的方法，其包括使所述靶核酸与本文所述工程化的 CRISPR-Cas12i 系统中任一种接触，其中所述指导 RNA 的指导序列与所述靶核酸的靶序列互补，其中所述工程化的 Cas12i 效应蛋白和所述指导 RNA 彼此缔合以结合至所述靶核酸以修饰(如切割或碱基替换)所述靶核酸，从而使所述疾病或病况得到治疗。在一些实施方案中，将突变(例如，敲除或敲入突变)引入所述靶核酸。在一些实施方案中，所述靶核酸的表达被增强。在一些实施方案中，所述靶核酸的表达被抑制。在一些实施方案中，本申请提供了治疗个体的疾病或病况的方法，其包括向所述个体施用有效量的本文所述工程化的 CRISPR-Cas12i 系统中任一种以及编码治疗剂的供体 DNA(如非致病的天然序列用作修复模板)，其中所述指导 RNA 的指导序列与所述个体的靶核酸的靶序列互补，其中所述工程化的 Cas12i 效应蛋白和所述指导 RNA 彼此结合以结合至所述靶核酸并将供体 DNA 插入所述靶序列中，从而使所述疾病或病况得到治疗。

在一些实施方案中，本申请提供了治疗个体的疾病或病况的方法，其包括向所述个体施用有效量的包含经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使所述细胞与本文所述工程化的 CRISPR-Cas12i 系统中任一种接触而制备的，其中所述指导 RNA 的指导序列与所述靶核酸的靶序列互补，其中所述工程化的 Cas12i 效应蛋白和所述指导 RNA 相互缔合以结合至靶核酸以修饰所述靶核酸。在一些实施方案中，所述工程化的细胞是免疫细胞。在一些实施方案中，所述工程化的细胞是干细胞(如造血干细胞、神经干细胞)。在一些实施方案中，所述工程化的细胞是神经细胞。在一些实施方案中，所述个体是人。在一些实施方案中，所述个体是动物，例如模型动物诸如啮齿动物、宠物或农场动物。在一些实施方案中，所述个体是哺乳动物，如猫、狗、兔、仓鼠等。

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、代谢性疾病、神经退行性疾病、眼病、细菌感染和病毒感染。在一些实施方案中，所述靶核酸是 PCSK9。在一些实施方案中，所述疾病或病况是心血管疾病。在一些实施方案中，所述疾病或病况是冠状动脉疾病。在一些实施方案中，所述方法降低个体的胆固醇水平。在一些实施方案中，所述方法治疗个体的糖尿病。

本发明的 CRISPR-Cas12i 系统可以用于治疗的病症或疾病包括但不限于，囊性纤维化、遗传性血管水肿、糖尿病、进行性假肥大性肌营养不良、贝克肌营养不良、 α -1-抗胰

蛋白酶缺乏、庞贝病、强直性肌营养不良、亨廷顿病、脆性 X 综合征、弗里德赖希共济失调、肌萎缩侧索硬化、额颞叶痴呆、遗传性慢性肾脏病、高脂血症、高胆固醇血症、莱伯氏先天性黑蒙、镰状细胞病或 β 地中海贫血等。在某些实施方式中，所述病症或疾病为转甲状腺素蛋白淀粉样变性，例如野生型转甲状腺素蛋白淀粉样变性(transthyretin-related wild-type amyloidosis, ATTRwt)、遗传性转甲状腺素蛋白淀粉样变性(transthyretin-related hereditary amyloidosis, ATTRh)、家族性淀粉样多发性神经病(familial amyloid polyneuropathy, FAP, ATTR-PN)、或家族性淀粉样变心肌病(familial amyloid cardiomyopathy, FAC, ATTR-CM)。在某些实施方式中，所述病症或疾病为 TTR 基因突变或异常表达(如高表达)引起的转甲状腺素蛋白不稳定。在某些实施方式中，所述病症或疾病为 TTR 基因突变或异常表达(如高表达)引起的其他病症或疾病，或衍生病症或疾病。

递送方法

在一些实施方案中，本文所述工程化的 CRISPR-Cas12i 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通过多种递送系统诸如质粒或病毒递送至宿主细胞(例如，上面“构建体和载体”小节中描述的载体中的任一种)。在一些实施方案或方法中，所述工程化的 CRISPR-Cas12i 系统可以通过其他方法递送，如由所述工程化的 Cas12i 效应蛋白及其一个或多个同源 RNA 指导序列组成的核糖核蛋白复合物的核转染或电穿孔。

在一些实施方案中，通过纳米颗粒或外泌体来递送。

在一些实施方案中，成对的 Cas12i 切口酶复合物可以使用纳米颗粒或其他直接蛋白递送方法直接递送，使得包含两个成对的 crRNA 元件的复合物被共同递送。此外，蛋白可以通过病毒载体或直接地递送至细胞，然后直接递送包含针对双切口的两个成对间隔区的 CRISPR 阵列。在某些情况下，对于直接 RNA 递送，可以将 RNA 缀合于至少一个糖部分诸如 N-乙酰基半乳糖胺(GalNAc)(特别是三触角 GalNAc)。

在一些实施方案或方法中，所述工程化的 CRISPR-Cas12i 系统可以使用任何适合所治疗疾病的递送方法来递送，例如通过静脉注射或滴注方式进行递送，或患病部位局部递送(如肿瘤内递送)。施用本文所述工程化的 CRISPR-Cas12i 系统的合适途径包括但不限于：局部、皮下、经皮、皮内、病灶内、关节内、腹膜内、膀胱内、经粘膜、牙龈、牙内、耳蜗内、经鼓膜、器官内、硬膜外、鞘内、肌肉内、静脉内、血管内、骨内、眼周、瘤内、脑内和脑室内施用。在一些实施方案中，本文所述工程化的 CRISPR-Cas12i 系统通过注射、通过导管、通过栓剂或通过植入物向受试者施用，植入物为多孔、无孔或凝胶状材料，包括膜如唾液酸膜，或纤维。

IV. 试剂盒和制品

还提供了包含本文所述工程化的 Cas12i 核酸酶或其功能变体、工程化的 Cas12i 效应蛋白或工程化的 CRISPR-Cas12i 系统中任一种的一种或多种组分的组合物、试剂盒、单位药剂和制品。

5 在一些实施方案中，提供了试剂盒，其包含：一个或多个 AAV 载体，其编码本文所述工程化的 Cas12i 核酸酶或其功能变体、工程化 Cas12i 效应蛋白或工程化 CRISPR-Cas12i 系统中任一种。在一些实施方案中，所述试剂盒还包含一种或多种指导 RNA(或其编码 DNA 或载体)。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，所述试剂盒还包含细胞诸如人细胞。

10 所述试剂盒可包含一种或多种额外组分，诸如容器、试剂、培养基、细胞因子、缓冲液、抗体等，以允许工程化的细胞的繁殖。所述试剂盒还可包含用于施用组合物的装置。

所述试剂盒还可包含使用本文所述工程化的 CRISPR-Cas12i 系统的说明书，诸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处理或诊断疾病或病况的说明书。与所述试剂盒组分使用有关的说明书通常包括针对该刻意处理的给药量、15 给药时间表和给药途径的信息。所述容器可以是单位剂量、散装包装(例如多剂量包装)或亚单位剂量。例如，可以提供这样的试剂盒(其包含足够剂量的本文所公开的组合物)以在延长的时期内提供对个体的有效处理。试剂盒还可以包括多个单位剂量的组合物和使用说明书，其包装数量足以在药房(例如医院药房和复方药房)中存储和使用。

20 本发明的试剂盒处于合适的包装中。合适的包装包括但不限于：小瓶、瓶子、广口瓶、软包装(例如密封的聚酯薄膜或塑料袋)等。试剂盒可以任选地提供额外的组分，如缓冲液和解释性信息。因此，本申请还提供了一种制品，其包括小瓶(如密封的小瓶)、瓶子、广口瓶、柔性包装等。

所述制品可以包括容器以及在容器上或与容器粘合的标签或包装插页。合适的容器25 包括例如瓶子、小瓶、注射器等。所述容器可以由多种材料(诸如玻璃或塑料)形成。通常，所述容器容纳有效处理本文所述的疾病或病症的组合物，并且可以具有无菌入口(例如，容器可以是静脉注射溶液袋或具有可通过皮下注射针刺穿的塞子的小瓶)。所述标签或包装插页表明该组合物用于处理个体中的特定病况。所述标签或包装插页将进一步包括用于将组合物向个体施用的说明书。

30 包装插页是指通常包含在治疗产品的商业包装中的说明，所述说明包含关于使用此类治疗产品的适应症、用法、给药量、施用、禁忌症和/或警告的信息。

此外, 所述制品还可包括第二容器, 其包含药学上可接受的缓冲液, 如抑菌注射用水(BWFI)、磷酸盐缓冲盐水、林格氏溶液和葡萄糖溶液。从商业和用户的角度来看, 它还可以包括其他材料, 包括其他缓冲液、稀释剂、过滤器、针头和注射器。必要时, 还可包括增溶剂和局部麻醉剂(如, 利多卡因), 以缓解注射部位的疼痛。

- 5 通常, 成分以单位剂型单独提供或混合在一起提供, 例如, 作为显示活性剂量的密封容器如安瓿或小袋中的干燥冻干粉或无水浓缩物。当药物通过输液施用, 可用装有无菌药用级水或盐水的输液瓶进行配药。当药物组合物通过注射施用, 可提供无菌注射用水或盐水的安瓿, 以便在施用前可混合成分。

10 实施例部分

下面将参照附图更详细地描述本发明的具体实施例。虽然附图中显示了本发明的具体实施例, 然而应当理解, 可以以各种形式实现本发明而不应被这里阐述的实施例所限制。相反, 提供这些实施例是为了能够更透彻地理解本发明, 并且能够将本发明的范围完整的传达给本领域的技术人员。

- 15 **实施例 1: 将参比 Cas12i2 酶中与 PAM 相互作用的氨基酸替换为带正电的氨基酸, 并验证其基因编辑效率。**

质粒构建

- Cas12i2 的编码序列经过密码子优化(人类)并合成。通过基于 PCR 的定点诱变产生 Cas 蛋白的变体。具体的方法是以突变的位点为中心将 Cas12i2 蛋白的 DNA 序列设计分成两部分, 设计两对引物分别扩增这两部分 DNA 序列, 同时引物上引入需要突变的序列, 最后通过 Gibson 克隆的方式将两个片段装载到 pCAG-2A-eGFP 载体(带有青霉素抗性)上。突变体的组合则通过将 Cas12i2 蛋白的 DNA 拆分成多段, 使用 PCR、Gibson clone 实现构建。突变体的位置确定使用本领域常用的蛋白质结构可视化软件(例如 PyMol、Chimera 等软件都可以选用)分析 Cas12i2 的结构信息而获得的。Cas12i2 的结构信息参照 PDB: 6LTU, 6LTR, 6LU0, 6LTP)。Cas12i2 效应蛋白通过 pCAG-2A-eGFP 载体在人 293T 细胞中表达。将编码 Cas12i2 蛋白的 DNA 插入 XmaI 和 NheI 之间。通过将含有靶序列的退火寡核苷酸连接到 BamHI 消化的 pUC19-U6-i2-crRNA 骨架中构建了在 293T 中表达 Cas12i2 的 crRNA 的载体。编码 DR 的核苷酸序列为 SEQ ID NO: 59。

细胞培养、转染以及荧光激活细胞分选法(FACS)

- 30 将 HEK293T 细胞在含 1% 青霉素-链霉素(Gibco)和 10% 胎牛血清(Gibco)的 DMEM(Gibco)中培养。将细胞接种在 24 孔-细胞培养皿(Corning)中 16 小时, 直到细胞密

度达到 70%。通过使用 Lipofectamine 3000 (Invitrogen), 将 600ng 编码 Cas12i2 蛋白的质粒和 300ng 编码 crRNA 的质粒转染到每个 24 孔-细胞培养皿中。转染约 72h 后, 用胰蛋白酶-EDTA (0.05%) (Gibco) 消化待荧光激活细胞分选(FACS) 的 HEK293T 细胞。使用具有 GFP 通道的 MoFlo XDP (Beckman Coulter) 进行细胞分选。

5 靶向深度测序分析以进行基因组修饰

将 FACS 分选的 GFP-阳性 HEK293T 细胞用缓冲液 L 裂解, 并在 55°C 下孵育 3 小时, 然后在 95°C 下孵育 10 分钟。使用相应的引物对在不同基因组位点中包含靶位点的 dsDNA 片段进行 PCR 扩增。对于靶向的深度测序, 直接使用细胞裂解液作为模板, 通过条形码(barcode) PCR 直接扩增靶位点。纯化 PCR 产物并归集到几个文库中以进行高通量测序。通过计算包含插入或缺失的读取(reads)的比率, 采用 CRISPResso2 软件分析插入缺失的频率(%)。在本申请中, 统一使用插入缺失的频率(%)这一指标来进行基因编辑效率的对比和分析。数量少于完整读取的 0.05% 的读取被废弃。

实施例 1-A 选出了四个具有单个氨基酸替换的工程化 Cas12i2

按照实施例 1 描述的方法分别表达氨基酸序列具有单个突变的工程化 Cas12i2 酶, 优选的氨基酸替换方式及其相应的基因编辑效率的见于图 1a、图 1b 和表 1。我们首先选取了 Cas12i2 距离 PAM DNA 9Å 以内的 10 个氨基酸: E176、E178、Y226、A227、N229、E237、K238、K264、T447、E563, 进行精氨酸(R)的点突变测试。如图 1a 和表 1 所示, 通过在 293T 细胞中比较这些 Cas12i 突变体与野生型 Cas12i2 (SEQ ID NO: 1) 在 2 个基因组位点: CCR5-3、RNF2-7 的基因编辑效率, 带有下述氨基酸替换的 Cas12i 突变体: 20 E176R、K238R、T447R 和 E563R 能够有效地提高基因编辑效率(图 1a 中显示, E176R、K238R、T447R 和 E563R 四种单氨基酸替换都获得了高于约 10%(此为针对 CCR5-3 的参比酶的基因编辑效率)以及高于约 12%(此为针对 RNF2-7 的参比酶的基因编辑效率)插入缺失率, 明显优于其他的单个氨基酸替换方案), 其余 6 种突变体则对基因编辑效率的提高没有帮助, 甚至降低。在此所使用的针对 CCR5-3 和针对 RNF2-7 的 crRNA 间隔序列的编码序列分别如 SEQ ID NO: 60 和 61 所示。

实施例 1-B 比较同时具有多个优选的氨基酸替换的工程化 Cas12i2

按照实施例 1 描述的方法分别表达其氨基酸序列同时具有两个以上优选的氨基酸替换的工程化 Cas12i2 酶, 其组合方式及其基因编辑效率的对比见于表 1 和图 1b。我们将由实施例 1-A 筛选得到的 E176R、K238R、T447R 和 E563R 这 4 个能够提高效率的突变体中的点突变进行组合。如图 1b 和表 1 所示, 通过在 293T 细胞中比较这些突变体与野生型 Cas12i2 在 3 个基因组位点: CCR5-3、CCR5-5、RNF2-7 的基因编辑效率, 我们发现 30

点突变组合之后能够获得效率更进一步提升的突变体。尤其是当把4种突变组合到一起(E176R+K238R+T447R+E563R)时能够获得效率最优的组合突变。在此所使用的针对CCR5-3、针对RNF2-7以及针对CCR5-5的crRNA间隔序列的编码序列分别如SEQ ID NO: 60、61和62所示。

5 表1 实施例1的结果(基因编辑效率)总结

单个氨基酸替换方式以及组合替换方式	基因编辑效率(CCR5-3处的插入缺失率)	基因编辑效率(RNF2-7处的插入缺失率)
参比酶(Cas12i2; SEQ ID NO: 1)	10.44	11.70
E178R	0.15	0.00
Y226R	0.19	0.00
A227R	0.35	0.29
N229R	8.68	5.37
E237R	0.26	0.07
K264R	15.83	19.35
E176R	24.07	29.16
K238R	15.83	19.35
T447R	28.61	43.71
E563R	55.49	68.30
E176R+K238R	28.76	34.76
E176R+T447R	49.47	59.64
E176R+E563R	74.78	75.94
K238R+T447R	37.52	50.41
K238R+E563R	64.08	73.38
E176R+K238R+T447R	56.42	63.16
E176R+K238R+E563R	73.17	74.50
E176R+T447R+E563R	68.39	73.89
E176R+K238R+T447R+E563R	73.45	86.29

实施例2：将参比 Cas12i2 酶中参与打开 DNA 双链的氨基酸替换为带芳香环的氨基酸，并分别验证它们的基因编辑的效率

质粒构建

10 通过基于PCR的定点诱变产生Cas12i2蛋白的变体，具体的方法是以突变的位点为中心将Cas12i2蛋白的DNA序列设计分成两部分，设计两对引物分别扩增这两部分DNA序列，同时引物上引入需要突变的序列，最后通过Gibson clone的方式将两个片段装载到pCAG-2A-eGFP载体上。氨基酸替换位置的确定可以使用常见的蛋白质结构可视化软件(例如，可以采用PyMol、Chimera等软件)分析Cas12i2的结构信息得到。Cas12i2

15 的结构信息参照PDB: 6LTU, 6LTR, 6LU0, 6LTP)。Cas12i2效应蛋白通过pCAG-2A-eGFP载体在人293T细胞中表达。将编码Cas12i2蛋白的DNA插入XmaI和NheI之间。通过将含有靶序列的退火寡核苷酸连接到BamI消化的pUC19-U6-i2-crRNA骨架中构建了在293T中表达Cas12i2 crRNA的载体。编码DR的核苷酸序列为SEQ ID NO: 59。

细胞培养、转染以及荧光激活细胞分选法(FACS)

将 HEK293T 细胞在含 1% 青霉素-链霉素 (Gibco) 和 10% 胎牛血清 (Gibco) 的 DMEM(Gibco) 中培养。将细胞接种在 24 孔-细胞培养皿(Corning) 中 16 小时, 直到细胞密度达到 70%。通过使用 Lipofectamine 3000 (Invitrogen), 将 600ng 编码 Cas12i2 蛋白的质粒和 300ng 编码 crRNA 的质粒转染到每个 24 孔-细胞培养皿中。转染 68 后, 用胰蛋白酶-EDTA (0.05%) (Gibco) 消化待荧光激活细胞分选(FACS) 的 HEK293T 细胞。使用具有 GFP 通道的 MoFlo XDP (Beckman Coulter) 进行细胞分选。

靶向深度测序分析以进行基因组修饰

将 FACS 分选的 GFP-阳性 HEK293T 细胞用缓冲液 L 裂解, 并在 55°C 下孵育 3 小时, 然后在 95°C 下孵育 10 分钟。使用相应的引物对在不同基因组位点中包含靶位点的 dsDNA 片段进行 PCR 扩增。对于靶向的深度测序, 直接使用细胞裂解液作为模板, 通过条形码(barcode) PCR 直接扩增靶位点。纯化 PCR 产物并归集到几个文库中以进行高通量测序。通过计算包含插入或缺失的读取(reads) 的比率, 采用 CRISPResso2 软件分析插入缺失的频率(%). 数量少于完整读取的 0.05% 的读取被废弃。

首先我们选取了 Cas12i2 酶中参与打开 DNA 双链的氨基酸: Q163 和 N164 进行带有芳香环的氨基酸(Y, F, W) 的点突变测试。从图 2 和表 2 可以看出, 通过在 293T 细胞中比较这些突变体与野生型 Cas12i2 在 3 个基因组位点: CCR5-3, CCR5-5, RNF2-7 的基因编辑效率, 我们发现, 有 5 个突变体: Q163W, Q163Y, Q163F, N164Y, N164F 能够有效地提高基因编辑效率(至少在一个基因组位点)。特别地, N164Y 以及 N164F 在 3 个基因组位点都展示了极佳的基因编辑效率。N164W 相比于参比酶没有改善基因编辑效率的效果。

表 2 实施例 2 的结果(基因编辑效率)总结

单个氨基酸替换方式	基因编辑效率(%)(CCR5-3 处的插入缺失率)	基因编辑效率(%)(RNF2-7 处的插入缺失率)	基因编辑效率(%)(CCR5-5 处的插入缺失率)
Q163W	22.07	16.97	10.64
Q163Y	12.51	9.70	5.47
Q163F	17.42	12.22	7.09
N164W	8.36	7.45	3.97
N164Y	29.72	43.77	21.66
N164F	28.89	45.48	22.99
参比酶(Cas12i2)	10.44	11.70	9.79

实施例 3: 将参比 Cas12i2 酶中位于 RuvC 结构域并与单链 DNA 底物相互作用的氨基酸替换为带正电的氨基酸, 并验证其基因编辑的效率

25 质粒构建

通过基于 PCR 的定点诱变产生 Cas12i2 蛋白的变体, 具体的方法是以突变的位点为中心将 Cas12i2 蛋白的 DNA 序列设计分成两部分, 设计两对引物分别扩增这两部分 DNA 序列, 同时引物上引入需要突变的序列, 最后通过 Gibson clone 的方式将两个片段装载到 pCAG-2A-eGFP 载体上。突变体的组合则通过将 Cas12i2 蛋白的 DNA 拆分成多段, 5 使用 PCR、Gibson clone 实现构建。突变体的位置确定是用常用的蛋白质结构可视化软件(例如, PyMol、Chimera 等软件均可采用)分析 Cas12i2 的结构信息得到的。Cas12i2 的结构信息参照 PDB ID: 6LTU, 6LTR, 6LU0, 6LTP。这些 Cas12i2 结构中显示的 ssDNA 底物只有 5nt。为了得到更长的 ssDNA 与 Cas12i2 相互作用的信息, 我们将 Cas12i1 的结构(PDB ID: 6W5C, 6W62 和 6W64; Zhang H. *et al. Nature Structural & Molecular Biology* 27, 10 1069-1076(2020)) 与 Cas12i2 的结构做了同源比对, 以便将该 Cas12i1 结构中的 ssDNA 底物(9nt)放入了 Cas12i2 的 RuvC 催化口袋, 通过该模型进一步寻找 9A 之内的氨基酸。Cas12i2 效应蛋白通过 pCAG-2A-eGFP 载体在人 293T 细胞中表达。将编码 Cas12i2 蛋白的 DNA 插入 XmaI 和 NheI 之间。通过将含有靶序列的退火寡核苷酸连接到 BsaI 消化的 pUC19-U6-i2-crRNA 骨架中构建了在 293T 中表达 Cas12i2 crRNA 的载体。编码 DR 的核苷酸序列为 SEQ ID NO: 59。针对 CCR5-3 和针对 RNF2-7 的 crRNA 间隔序列的编码序列 15 分别如 SEQ ID NO: 60 和 61 所示。

细胞培养、转染以及荧光激活细胞分选法(FACS)

将 HEK293T 细胞在含 1% 青霉素-链霉素(Gibco)和 10% 胎牛血清(Gibco)的 DMEM(Gibco)中培养。将细胞接种在 24 孔-细胞培养皿(Corning)中 16 小时, 直到细胞密度 20 达到 70%。通过使用 Lipofectamine 3000 (Invitrogen), 将 600ng 编码 Cas12i2 蛋白的质粒和 300ng 编码 crRNA 的质粒转染到每个 24 孔-细胞培养皿中。转染 68h 后, 用胰蛋白酶-EDTA (0.05%) (Gibco)消化待荧光激活细胞分选(FACS)的 HEK293T 细胞。使用具有 GFP 通道的 MoFlo XDP (Beckman Coulter) 进行细胞分选。

靶向深度测序分析以进行基因组修饰

25 将 FACS 分选的 GFP-阳性 HEK293T 细胞用缓冲液 L 裂解, 并在 55°C 下孵育 3 小时, 然后在 95°C 下孵育 10 分钟。使用相应的引物对在不同基因组位点中包含靶位点的 dsDNA 片段进行 PCR 扩增。对于靶向的深度测序, 直接使用细胞裂解液作为模板, 通过条形码(barcode) PCR 直接扩增靶位点。纯化 PCR 产物并归集到几个文库中以进行高通量测序。通过计算包含插入或缺失的读取(reads)的比率, 采用 CRISPResso2 软件分析插入 30 缺失的频率(%)。数量少于完整读取的 0.05% 的读取被废弃。我们将参比 Cas12i2 酶中位于 RuvC 结构域并与单链 DNA 底物相互作用的氨基酸替换为

带正电的氨基酸。在 293T 细胞中比较这些突变体与野生型 Cas12i2 在 CCR5-3 和/或 RNF2-7 基因组位点的基因编辑效率。如图 3a 和表 3 所示, N391R、I926R、G929R 能够有效地提高基因编辑效率(至少在一个基因组位点)。其中 I926R 在 2 个基因组位点都展示了极佳的基因编辑效率。如图 3b、3c 和表 3 所示, 许多突变体能够有效地提高基因编辑效率(至少在一个基因组位点)。其中基因编辑效率提高的单个氨基酸替换方案的排名为: D362R>E323R>Q425R>N925R>其他效率提高的突变体。

我们将图 3a、3b、3c 筛选得到的 E323R、D362R、Q425R 和 I926R 这 4 个能够提高效率的突变体中的点突变进行组合。如图 3d 和表 3 所示, 通过在 293T 细胞中比较这些突变体与野生型 Cas12i2 在 2 个基因组位点: CCR5-3, RNF2-7 的基因编辑效率, 我们发现点突变组合之后能够获得基因编辑效率更进一步提升的突变体(至少在一个位点)。

我们将图 3a、3b、3c、3d 筛选得到的部分能够提高效率的突变体中的点突变或组合与柔性区突变 I926G、L439(L+GG)进行组合。如图 3e 和表 3 所示, 通过在 293T 细胞中比较这些突变体与野生型 Cas12i2 在 RNF2-7 位点的基因编辑效率, 我们发现点突变组合之后能够获得效率更进一步提升的突变体, 例如 N925R+I926G、I926R+L439(L+GG)。

由此可见, 进一步引入柔性区突变可以增加 Cas12i 突变体的基因编辑效率。

表 3 实施例 3 的结果(基因编辑效率)总结

单个氨基酸替换方式 以及组合替换方式	基因编辑效率%(CCR5-3 处 的插入缺失率)	基因编辑效率%(RNF2-7 处的 插入缺失率)
参比酶(Cas12i2)	10.44	11.70
N390R	1.64	9.13
N391R	16.92	28.91
F392R	0.20	0.71
L751R	0.08	0.74
E755R	1.95	6.84
N840R	8.29	19.68
N848R	0.09	0.13
S851R	0.75	9.40
A856R	0.10	0.86
Q885R	0.13	0.48
M897R	0.26	1.15
I926R	66.51	73.60
G929R	32.35	40.67
Y932R	0.06	0.44
E323R	46.44	55.50
L327R	28.57	40.21
V355R	21.62	29.88
G359R	15.50	25.12
G360R	19.92	29.03
K361R	16.79	15.15
D362R	53.15	62.88
Q414R	14.77	22.41

单个氨基酸替换方式 以及组合替换方式	基因编辑效率%(CCR5-3 处 的插入缺失率)	基因编辑效率%(RNF2-7 处的 插入缺失率)
K421R	12.37	13.41
Q425R	32.51	45.73
S650R	1.48	3.81
E652R	2.05	3.80
K705R	15.11	16.78
K708R	9.98	15.64
E709R	0.65	5.68
S752R	4.26	6.64
N925R	30.39	42.01
T928R	3.38	3.46
E323R+D362R	73.83	84.37
E323R+Q425R	45.80	64.08
E323R+I926R	80.58	86.10
Q425R+I926R	70.05	81.84
E323R+D362R+Q425R	60.76	74.79
E323R+D362R+I926R	78.08	83.50
E323R+Q425R+I926R	73.44	83.19
E323R+D362R+Q425R+I926R	80.51	86.62
N925R		33.13
I926G		43.68
L439(L+GG)*		31.22
D362R+I926R		63.87
N925R+I926R		49.84
N925R+I926G		62.42
D362R+N925R+I926R		72.99
D362R+N925R+I926G		55.06
I926R+L439(L+GG)		80.53
E323R+D362R+I926G		81.93

*L439(L+GG)的含义是，在 439 号氨基酸后插入两个甘氨酸。

实施例 4：将参比 Cas12i2 酶中与 DNA-RNA 双螺旋相互作用的氨基酸替换为带正电的氨基酸，并验证其基因编辑的效率

5 质粒构建

通过基于 PCR 的定点诱变产生 Cas12i2 蛋白的变体，具体的方法是以突变的位点为中心将 Cas12i2 蛋白的 DNA 序列设计分成两部分，设计两对引物分别扩增这两部分 DNA 序列，同时引物上引入需要突变的序列，最后通过 Gibson clone 的方式将两个片段装载到 pCAG-2A-eGFP 载体上。突变体的组合则通过将 Cas12i2 蛋白的 DNA 拆分成多段，使用 PCR、Gibson clone 实现构建。突变体的位置确定是用常用的蛋白质结构可视化软件 (例如，PyMol、Chimera 等软件均可采用) 分析 Cas12i2 的结构信息得到的。Cas12i2 的结构信息参照 PDB: 6LTU, 6LTR, 6LU0, 6LTP)。Cas12i2 效应蛋白通过 pCAG-2A-eGFP 载体在人 293T 细胞中表达。将编码 Cas12i2 蛋白的 DNA 插入 XmaI 和 NheI 之间。通过将含

有靶序列的退火寡核苷酸连接到 *Bsa*I 消化的 pUC19-U6-i2-crRNA 骨架中构建了在 293T 中表达 Cas12i2 crRNA 的载体。编码 DR 的核苷酸序列为 SEQ ID NO: 59。针对 CCR5-3 和针对 RNF2-7 的 crRNA 间隔序列的编码序列分别如 SEQ ID NO: 60 和 61 所示。

细胞培养、转染以及荧光激活细胞分选法(FACS)

- 5 将 HEK293T 细胞在含 1% 青霉素-链霉素 (Gibco) 和 10% 胎牛血清 (Gibco) 的 DMEM(Gibco) 中培养。将细胞接种在 24 孔-细胞培养皿(Corning) 中 16 小时, 直到细胞密度达到 70%。通过使用 Lipofectamine 3000 (Invitrogen), 将 600ng 编码 Cas12i2 蛋白的质粒和 300ng 编码 crRNA 的质粒转染到每个 24 孔-细胞培养皿中。转染 68h 后, 用胰蛋白酶-EDTA (0.05%) (Gibco) 消化待荧光激活细胞分选(FACS) 的 HEK293T 细胞。使用具有
- 10 GFP 通道的 MoFlo XDP (Beckman Coulter) 进行细胞分选。

靶向深度测序分析以进行基因组修饰

- 将 FACS 分选的 GFP-阳性 HEK293T 细胞用缓冲液 L 裂解, 并在 55°C 下孵育 3 小时, 然后在 95°C 下孵育 10 分钟。使用相应的引物对在不同基因组位点中包含靶位点的 dsDNA 片段进行 PCR 扩增。对于靶向的深度测序, 直接使用细胞裂解液作为模板, 通过
- 15 条形码(barcode) PCR 直接扩增靶位点。纯化 PCR 产物并归集到几个文库中以进行高通量测序。通过计算包含插入或缺失的读取(reads) 的比率, 采用 CRISPResso2 软件分析插入缺失的频率(%). 数量少于完整读取的 0.05% 的读取被废弃。

- 图 4 和表 4 总结了在 293T 细胞中比较本实施例中的 Cas12i2 突变体与野生型 Cas12i2 在 2 个基因组位点: CCR5-3、RNF2-7 的基因编辑效率。我们发现, 有 7 个突变体: G116R、E117R、T159R、S161R、E319R、E343R、和 D958R 能够有效地提高基因
- 20 编辑效率(至少在一个基因组位点)。其中 D958R 在 2 个基因组位点都展示了极佳的基因编辑效率。

表 4 实施例 4 的结果(基因编辑效率)总结

单个氨基酸替换方式以及组合替换方式	基因编辑效率(CCR5-3 处的插入缺失率)	基因编辑效率(RNF2-7 处的插入缺失率)
参比酶(Cas12i2)	10.44	11.70
G116R	19.14	23.95
E117R	19.79	25.01
A156R	6.15	3.64
T159R	22.59	21.89
S161R	21.81	23.64
T301R	0.75	0.18
I305R	0.84	0.45
K306R	9.35	4.30
T308R	11.23	6.48
N312R	12.93	11.03
F313R	19.45	12.58

单个氨基酸替换方式以及组合替换方式	基因编辑效率(CCR5-3 处的插入缺失率)	基因编辑效率(RNF2-7 处的插入缺失率)
D427R	14.11	16.28
K433R	12.72	11.85
V438R	10.81	7.03
N441R	12.75	11.96
Q442R	5.22	1.34
M852R	1.33	1.77
L855R	2.03	1.70
N861R	5.82	12.65
Q865R	9.23	6.63
E160R	13.15	17.70
Q316R	12.70	15.86
E319R	24.62	--
Q320R	18.85	20.11
E247R	8.56	11.63
E343R	18.20	24.24
E348R	14.83	20.37
E349R	7.71	11.39
N679R	8.98	12.43
E683R	16.67	18.17
E691R	10.93	14.98
D782R	9.67	21.61
E783R	9.78	13.25
E797R	12.27	20.91
E800R	0.71	1.53
D853R	0.50	0.00
S957R	13.41	17.61
D958R	29.00	36.82
G293R	14.32	19.40
E294R	13.07	16.98
N297R	9.02	10.90

实施例 5：将实施例 1-4 中筛选得到的部分提高基因编辑效率 Cas12i2 工程化改造氨基酸突变进行组合，并验证它们的基因编辑的效率。

质粒构建

- 5 突变体的组合则通过将 Cas12i2 蛋白的 DNA 拆分成多段，使用 PCR、Gibson clone 实现构建。突变体的位置确定是使用常用蛋白质结构可视化软件(例如，PyMol、Chimera 等软件都可以采用)分析 Cas12i2 的结构信息得到的。Cas12i2 的结构信息参照 PDB: 6LTU, 6LTR, 6LU0, 6LTP)。Cas12i2 效应蛋白通过 pCAG-2A-eGFP 载体在人 293T 细胞中表达。将编码 Cas12i2 蛋白的 DNA 插入 XmaI 和 NheI 之间。通过将含有靶序列的退火寡核苷酸
- 10 连接到 BasI 消化的 pUC19-U6-i2-crRNA 骨架中构建了在 293T 中表达 Cas12i2 crRNA 的载体。编码 DR 的核苷酸序列为 SEQ ID NO:59。

细胞培养、转染以及荧光激活细胞分选法(FACS)

将 HEK293T 细胞在含 1% 青霉素-链霉素(Gibco)和 10% 胎牛血清(Gibco)的 DMEM(Gibco)中培养。将细胞接种在 24 孔-细胞培养皿(Corning)中 16 小时,直到细胞密度达到 70%。通过使用 Lipofectamine 3000 (Invitrogen),将 600ng 编码 Cas12i2 蛋白的质粒和 300ng 编码 crRNA 的质粒转染到每个 24 孔-细胞培养皿中。转染约 68h 后,用胰蛋白酶-EDTA (0.05%) (Gibco)消化待荧光激活细胞分选(FACS)的 HEK293T 细胞。使用具有 GFP 通道的 MoFlo XDP (Beckman Coulter) 进行细胞分选。

靶向深度测序分析以进行基因组修饰

将 FACS 分选的 GFP-阳性 HEK293T 细胞用缓冲液 L 裂解,并在 55°C 下孵育 3 小时,然后在 95°C 下孵育 10 分钟。使用相应的引物对在不同基因组位点中包含靶位点的 dsDNA 片段进行 PCR 扩增。对于靶向的深度测序,直接使用细胞裂解液作为模板,通过条形码(barcode) PCR 直接扩增靶位点。纯化 PCR 产物并归集到几个文库中以进行高通量测序。通过计算包含插入或缺失的读取(reads)的比率,采用 CRISPResso2 软件分析插入缺失的频率(%).数量少于完整读取的 0.05%的读取被废弃。

我们将实施例 1-4 中筛选得到的氨基酸突变或者氨基酸突变组合:
 E176R+K238R+T447R+E563R, N164Y, E323R+D362R, I926R, E323R+D362R+I926R,
 E323R+D362R+I926G, E323R+D362R+I926G+L439(L+G),
 E323R+D362R+I926G+L439(L+GG)进行进一步的组合。如图 5 和表 5 所示,通过在 293T 细胞中比较这些突变体与野生型 Cas12i 作用于 5 个基因组位点: CCR5-3, CCR5-5, CD34-8, CD34-9, RNF2-14 的基因编辑效率,我们发现不同种类的点突变组合之后能够获得效率更进一步提升的突变体。同时,将我们认为效率最优的突变体 (E176R+K238R+T447R+E563R+N164Y+E323R+D362R)命名为 CasXX。在此,所使用的针对 CD34-8、CD34-9 以及 RNF2-14 的 crRNA 间隔序列的编码序列分别如 SEQ ID NO: 63、64 和 65 所示。

表 5 实施例 5 的结果(基因编辑效率)总结(仅以 RNF2-14 处体现的基因编辑效率数据为例)

突变种类(编号如“工程化的 Cas12i 核酸酶”章节所示)	组合替换方式	氨基酸序列编号 SEQ ID NO:	基因编辑效率(%) (RNF2-14 处插入缺失率)
/	参比酶(Cas12i2)	1	1.06
(1)	E176R+K238R+T447R+E563R		33.73
(2)	N164Y		10.08
(3)/(4)	I926R		12.07
(3)/(4)	E323R+D362R		14.48
(1)+(2)	E176R+K238R+T447R+E563R+N164Y	2	70.28
(1)+(3)/(4)	E176R+K238R+T447R+E563R+I926R	3	74.31
(1)+(3)/(4)	E176R+K238R+T447R+E563R+E323R+D362R	4	81.65
(2)+(3)/(4)	N164Y+I926R	5	46.89

(2)+(3)/(4)	N164Y+E323R+D362R	6	49.15
(1)+(2)+(3)/(4)	E176R+K238R+T447R+E563R+N164Y+I926R	7	86.96
(1)+(2)+(3)/(4)	E176R+K238R+T447R+E563R+N164Y+E323R+D362R (命名为 CasXX)	8	81.07
(1)+(2)+(3)/(4)	E176R+K238R+T447R+E563R+N164Y+I926R+E323R+D362R	9	76.54
(1)+(2)+(3)/(4)+(6)	E176R+K238R+T447R+E563R+N164Y+E323R+D362R+I926G	10	86.74
(1)+(2)+(3)/(4)+(6)	E176R+K238R+T447R+E563R+N164Y+E323R+D362R+I926G+L439(L+GG)	11	83.76
(1)+(2)+(3)/(4)+(6)	E176R+K238R+T447R+E563R+N164Y+E323R+D362R+I926G+L439(L+G)*	12	84.85

*L439(L+G)的含义是，在 439 号氨基酸后插入一个甘氨酸；在本说明书的上下文以及附图中，有时也表示为 439G。

此外，我们还构建了以下突变组合：E176R+K238R+T447R+E563R+N164Y+D958R；

E176R+K238R+T447R+E563R+I926R+D958R；

- 5 E176R+K238R+T447R+E563R+E323R+D362R+D958R；
N164Y+I926R+D958R； N164Y+E323R+D362R+D958R；
E176R+K238R+T447R+E563R+N164Y+I926R+D958R；
E176R+K238R+T447R+E563R+N164Y+E323R+D362R+D958R；
E176R+K238R+T447R+E563R+N164Y+I926R+E323R+D362R+D958R；
- 10 E176R+K238R+T447R+E563R+N164Y+E323R+D362R+I926G+D958R；
E176R+K238R+T447R+E563R+N164Y+E323R+D362R+I926G+L439(L+GG)+D958R；和
E176R+K238R+T447R+E563R+N164Y+E323R+D362R+I926G+L439(L+G)+D958R。可以通过 T7 核酸内切酶 1 (T7E1)测定和靶向深度测序检测其基因编辑效率。

15 实施例 6：CasXX 与常规基因编辑工具的比较验证其基因编辑效率。

质粒构建

- 为比较 CasXX 和其他 Cas 的基因编辑活性，AsCas12a, BhCas12b v4, SpCas9, SaCas9, SaCas9-KKH 的编码序列经过密码子优化(人类)并合成。Cas 效应蛋白通过 pCAG-2A-eGFP 载体在人 293T 细胞中表达。将编码 Cas 蛋白的 DNA 插入 XmaI 和 NheI 之间。通过
- 20 将含有靶序列的退火寡核苷酸连接到 BasI 消化的 pUC19-U6-i2-crRNA 骨架中，构建了在 293T 中表达 AsCas12a, BhCas12b v4, SpCas9, SaCas9, SaCas9-KKH 和 Cas12i2 的 sgRNA 或 crRNA 的载体。编码 CasXX 的 DR 的核苷酸序列为 SEQ ID NO: 59。

细胞培养、转染以及荧光激活细胞分选法(FACS)

将 HEK293T 细胞在含 1% 青霉素-链霉素(Gibco)和 10% 胎牛血清(Gibco)的

DMEM(Gibco)中培养。将细胞接种在 24 孔-细胞培养皿(Corning)中 16 小时,直到细胞密度达到 70%。通过使用 Lipofectamine 3000 (Invitrogen),将 600ng 编码 Cas12i2 蛋白的质粒和 300ng 编码 crRNA 的质粒转染到每个 24 孔-细胞培养皿中。转染约 68h 后,用胰蛋白酶-EDTA (0.05%) (Gibco)消化待荧光激活细胞分选(FACS)的 HEK293T 细胞。使用具有 GFP 通道的 MoFlo XDP (Beckman Coulter) 进行细胞分选。

靶向深度测序分析以进行基因组修饰

将 FACS 分选的 GFP-阳性 HEK293T 细胞用缓冲液 L 裂解,并在 55°C 下孵育 3 小时,然后在 95°C 下孵育 10 分钟。使用相应的引物对在不同基因组位点中包含靶位点的 dsDNA 片段进行 PCR 扩增。对于靶向的深度测序,直接使用细胞裂解液作为模板,通过条形码(barcode) PCR 直接扩增靶位点。纯化 PCR 产物并归集到几个文库中以进行高通量测序。通过计算包含插入或缺失的读取(reads)的比率,采用 CRISPResso2 软件分析插入缺失的频率(%)。数量少于完整读取的 0.05% 的读取被废弃。

我们首先测试了 CasXX 在包含不同 PAM 序列的 62 个人基因组位点的基因编辑效率。设计的间隔序列为 20 个核苷酸。如图 6a 所示,CasXX 展现了极其强大的基因编辑能力,平均基因编辑效率超过 60%,并且几乎在测试的所有位点的基因编辑效率都超过了 50%。并且,对于任意的 NTTN PAM(N=A, T, G, C)都展现了很高的基因编辑效率。

为了进一步展示工程化改造的 CasXX 的基因编辑能力,我们将 CasXX 与 AsCas12a 在 TTTN PAM 位点进行比较,以及与 BhCas12b v4 在 TTN PAM 位点进行比较。如图 6b 所示,CasXX 在带有两种 PAM 的靶位点均展现了更高的平均基因编辑效率。

我们还将 CasXX 与 SpCas9、SaCas9、SaCas9-KKH 在相同位点进行比较。如图 6c 所示,CasXX 在测试位点均展现了更高的平均基因编辑效率。

为测试 CasXX 在体内的基因编辑活性,编码 CasXX 的 pCAG-2A-eGFP 载体以及编码 crRNA 的 pUC19-U6-i2-crRNA 载体用脂质体转染法转染至小鼠 Hepa1-6 肝癌细胞系。其中,分别对 65 个内源基因位点设计了相应 crRNA。设计的 spacer 序列为 20 个核苷酸。插入缺失频率由 PCR 扩增测序所得,和外源基因编辑分析方法类似。

如图 6d 所示,CasXX 对小鼠 Hepa1-6 细胞系的 65 个内源基因位点展示出强大的基因编辑能力,平均基因编辑效率超过 60%。

实施例 7: 使用 CasXX 在含有不同 PAM 的基因组位点进行基因编辑。

质粒构建

将编码 CasXX 的 DNA 序列装载到 pCAG-2A-EGFP 质粒中,从而构建了表达

CasXX 质粒。通过将含有靶序列的退火寡核苷酸连接到 BsaI 消化的 pUC19-U6-crRNA 骨架中构建了用于在 HEK293T 中表达 crRNA 的载体。其中，设计了 64 个不同的 crRNA 靶向 64 个人内源位点。所述内源靶核酸的 5' 端有如图 7 所示不同 PAM 5'-NNNN-3'(N=A、T、G 或 C)，以检测 CasXX 对不同 PAM 的识别能力。其中，这 64 个位点包含的 PAM 序列覆盖了所有 NNNN 的组合：NTTN, NTAN, NTCN, NTGN, NATN, NAAN, NACN, NAGN, NCTN, NCAN, NCCN, NCGN, NGTN, NGAN, NGCN, NGGN。编码 DR 的核苷酸序列为 SEQ ID NO: 59。设计的间隔序列为 20 个核苷酸。

细胞培养、转染以及荧光激活细胞分选法(FACS)

将 HEK293T 细胞在含 1%青霉素-链霉素(Gibco)和 10%胎牛血清(Gibco)的 DMEM(Gibco)中培养。将细胞接种在 24-细胞培养皿(Corning)中 16 小时，直到细胞密度达到 70%。通过使用 Lipofectamine 3000 (Invitrogen)，将 600ng 编码 Cas 蛋白的质粒和 300ng 编码 crRNA 的质粒转染到 24 孔细胞培养皿培养的细胞中。转染 72h 后，用胰蛋白酶-EDTA (0.05%) (Gibco)消化细胞，然后进行 GFP 荧光激活细胞分选(FACS)。

靶向深度测序分析以进行基因组修饰

将 FACS 分选的 GFP-阳性 293FT 细胞用缓冲液 L 裂解，并在 55°C 下孵育 3 小时，然后在 95°C 下孵育 10 分钟。使用相应的引物对在不同基因组位点中包含靶位点的 dsDNA 片段进行 PCR 扩增。对于靶向的深度测序，直接使用细胞裂解液作为模板，通过条形码(barcode) PCR 直接扩增靶位点。纯化 PCR 产物并归集到几个文库中以进行高通量测序。通过计算包含插入或缺失的读取(reads)的比率，采用 CRISPResso2 软件分析插入缺失的频率(%)。数量少于完整读取的 0.05%的读取被废弃。实验结果如图 7 所示。

如图 7 所示，CasXX 在 5' 端带有 NTTN, NTAN, NTCN, NTGN, NATN, NAAN, NCTN, NCAN, NGTN 这些 PAM 的靶位点都展现了高效的基因编辑效率，在这些位点的平均基因编辑效率超过 40%。

实施例 8：使用 CasXX 在体外切割含有不同 PAM 的双链 DNA

构建突变体的表达质粒

我们将编码野生型 Cas12i2 (SEQ ID NO: 1)以及 CasXX (SEQ ID NO: 8)的 DNA 序列装载到 BPK2014 质粒(带有氯霉素抗性)中，从而构建了原核表达 Cas12i2 以及 CasXX 蛋白的质粒。

蛋白质纯化

将 BPK2014 原核表达质粒转化到大肠杆菌菌株 BL21 (λ DE3) (TransGen Biotech) 中，

并将转化后的菌液涂布在含有氯霉素的固体 LB 上。将三个克隆挑取到 5 ml 液体 LB 中并培养过夜。然后将细菌转移到 3 L 液体 LB 中继续培养，直至 OD600 达到 0.6~0.8。然后用 IPTG (0.5 mM) 在 16°C 下诱导 20 小时。超速离心，收获表达 Cas12i 的细菌，重悬于裂解缓冲液(50 mM Tris-HCl、pH7.5、300 mM NaCl)中并通过超声破碎。离心后，首先

5 用 Ni 柱纯化上清液中的 Cas12i 蛋白。简而言之，在与上清液一起温育后，Ni 柱依次用添加了 0 mM、20 mM 和 50 mM 咪唑的裂解缓冲液洗涤。然后，Cas12i 蛋白被添加 500 mM 咪唑的裂解缓冲液洗脱。然后将收集的样品加载到离子交换柱(CM Sepharose Fast Flow, GE)中。野生型 Cas12i2 和 CasXX 蛋白用储存缓冲液(20 mM Tris-HCl、300 mM NaCl、1 mM TCEP、10% 甘油、pH7.5)洗脱。蛋白质通过过滤器灭菌并储存在-80°C。

10 crRNA 体外转录

编码 DR 的核苷酸序列为 SEQ ID NO: 59。合成含有 T7 启动子序列的寡核苷酸(命名为 T7-F)和含有 crRNA 和 T7 启动子互补序列的寡核苷酸(命名为 T7-12i-crRNA-R)，并在 1x NEBufferTM2 (NEB) 中退火。这些寡核苷酸的序列列于表 6。将退火产物用作模板，通过 HiScribeTM T7 Quick High Yield RNA Synthesis Kit (NEB) 生产 crRNA。使用

15 Monarch® RNA Cleanup Kit (NEB) 纯化转录的 crRNA。

表 6 crRNA 引物序列

SEQ ID NO: 41	T7-F	TAATACGACTCACTATAGG
SEQ ID NO: 42	T7-12i-crRNA-R	AAAAATTTGTTTGGATATGTTGAACCGTCAATGAAAGACG GATTTCTCCCTATAGTGAGTCGTATTA

Cas12i 蛋白体外酶切

为了制备用于体外切割的线性 dsDNA 底物，首先将含有相同原型间隔区和不同 PAM 的靶标克隆到由 EcoR1 和 HindIII 处理的 pUC19 (带有青霉素抗性)中。5'带有 PAM

20 的靶标序列列于表 7。然后将携带靶标的 pUC19 质粒通过 SacI 线性化，并使用 DNA Clean & Concentrator (Zymo Research) 进行纯化。对于体外切割实验，首先将 400 nM Cas12i 蛋白与 2 μM crRNA 在 37°C 下孵育 15 分钟。接下来，在含有 1x NEBufferTM3.1 (NEB) 的 10 μl 反应体系中，将 Cas12i-crRNA RNP 与 150 ng 线性化靶 DNA 在 37°C 下反应 40 分钟。然后，用 50 mM EDTA 终止反应，用 RNase cocktail(Invitrogen)在 37°C 下消

25 化 RNA 15 分钟。最后，用蛋白酶 (NEB) 在 37°C 下处理样品 15 分钟。通过在 1.2% 琼脂糖凝胶上电泳分离反应产物。

如图 8 所示，野生型 Cas12i2 蛋白仅仅对含有 5'-NTTN-3' PAM 的双链 DNA 具有部分切割的效率，但是对其余的 PAM 几乎没有切割活性。但是，CasXX 蛋白对含有 NTTN, NTAN, NTCN, NTGN, NATN, NAAN, NACN, NCTN, NCAN, NGTN, NGAN PAM 的双链

30 DNA 都展现了高效的切割效率。CasXX 几乎能完全切割含有 NTTN, NTAN, NTCN,

NATN, NAAN, NACN, NCTN, NCAN, NGTN PAM 的双链 DNA。因此, 本发明所提供的工程化的 Cas12i 核酸酶能用于更广泛的基因编辑或治疗。

表 7: 5' PAM – 靶标序列

SEQ ID NO: 43	NTTN-靶标	CTTCTTCAACATATCCAAACAAAT
SEQ ID NO: 44	NTAN-靶标	CTACTTCAACATATCCAAACAAAT
SEQ ID NO: 45	NTCN-靶标	CTCCTTCAACATATCCAAACAAAT
SEQ ID NO: 46	NTGN-靶标	CTGCTTCAACATATCCAAACAAAT
SEQ ID NO: 47	NATN-靶标	CATCTTCAACATATCCAAACAAAT
SEQ ID NO: 48	NAAN-靶标	CAACTTCAACATATCCAAACAAAT
SEQ ID NO: 49	NACN-靶标	CACCTTCAACATATCCAAACAAAT
SEQ ID NO: 50	NAGN-靶标	CAGCTTCAACATATCCAAACAAAT
SEQ ID NO: 51	NCTN-靶标	CCTCTTCAACATATCCAAACAAAT
SEQ ID NO: 52	NCAN-靶标	CCACTTCAACATATCCAAACAAAT
SEQ ID NO: 53	NCCN-靶标	CCCCTTCAACATATCCAAACAAAT
SEQ ID NO: 54	NCGN-靶标	CCGCTTCAACATATCCAAACAAAT
SEQ ID NO: 55	NGTN-靶标	CGTCTTCAACATATCCAAACAAAT
SEQ ID NO: 56	NGAN-靶标	CGACTTCAACATATCCAAACAAAT
SEQ ID NO: 57	NGCN-靶标	CGCCTTCAACATATCCAAACAAAT
SEQ ID NO: 58	NGGN-靶标	CGGCTTCAACATATCCAAACAAAT

5 **实施例 9: 使用 GUIDE-Seq 检测 CasXX 的脱靶效应**

质粒构建

我们将编码 CasXX 的 DNA 序列装载到 pCAG-2A-EGFP 质粒中, 从而构建了表达 CasXX 质粒。通过将含有靶序列的退火寡核苷酸连接到 *BasI* 消化的 pUC19-U6-crRNA 骨架中构建了在 293T(人肾上皮细胞系) 中表达 Cas 蛋白 crRNA 的载体。其中 crRNA 包含能靶向 EMX1-7 内源位点的间隔序列(spacer; SEQ ID NO: 77)。编码 DR 的核苷酸序列为 SEQ ID NO: 59。

细胞培养、转染以及荧光激活细胞分选法(FACS)

15 将 HEK293T 细胞在含 1%青霉素-链霉素(Gibco)和 10%胎牛血清(Gibco)的 DMEM(Gibco)中培养。将细胞接种在 24-细胞培养皿(Corning)中 16 小时, 直到细胞密度达到 70%。通过使用 Lipofectamine 3000 (Invitrogen), 将 600ng 编码 Cas 蛋白的质粒、300ng 编码 crRNA 的质粒和 10pmol 的退火的双链 DNA 标签(序列见表 8), 转染到 24 孔细胞培养皿培养的细胞中。转染 72h 后, 用胰蛋白酶-EDTA (0.05%) (Gibco)消化细胞, 然后进行 GFP 荧光激活细胞分选(FACS)。分选出的是成功表达 Cas 酶的细胞。

表 8 双链 DNA 标签的引物序列 (P 代表磷酸修饰, *代表硫代磷酸二酯键修饰)

SEQ ID NO: 25	GUIDE-F1	5'- P-G*T*TTAATTGAGTTGTCATATGTTAATAACGGT*A*T -3'
SEQ ID NO: 26	GUIDE-R1	5'- P-A*T*ACCGTTATTAACATATGACAACTCAATTA*A*C -3'

20 全基因组范围检测 CasXX 的脱靶效应

使用 E.Z.N.A.® MicroElute Genomic DNA Kit (Omega)试剂盒提取 FACS 分选的 GFP-阳性 293T 细胞用的基因组。纯化好的基因组使用 Qubit 定量。使用 Covaris S220 仪器，按照仪器推荐的程序将基因组打断到 500 bp 左右。然后使用 VAHTS Universal Pro DNA Library Prep Kit for Illumina(Vazyme)试剂盒进行 DNA 建库。建库流程参照参考文献 (Tsai, SQ *et al.* “GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases,” *Nat Biotechnol.* 2015;33(2):187-197; 其内容通过引用整体并入本文)。将建库产物进行高通量测序。分析测序结果，寻找潜在的脱靶位点。我们对 CasXX 在 EMX1-7 位点的脱靶效应进行了系统的分析，具体的分析结果见附图 10。第一行序列为参考靶序列，下面的序列分别代表靶序列和脱靶序列以及双链 DNA 标签在靶序列和脱靶序列富集的 reads 数目。

从较高的脱靶数目可以看出，CasXX 酶在哺乳动物细胞中的编辑特异性并非很理想，还需要进一步优化。

实施例 10：在 CasXX 序列(SEQ ID NO: 8)基础上引入新的氨基酸突变，检验该突变体特异性提高的特征(选择了两个靶向位点进行突变体筛选实验，即 EMX1-7 位点和 RNF2-1 位点)

筛选实验准备：构建突变体的表达质粒

在 CasXX(包含基于 SEQ ID NO: 1 的 N164Y+E176R+K238R+E323R+D362R+T447R+E563R 突变)的基础上，我们拟进一步引入新的氨基酸点突变来提高 CasXX 基因编辑的特异性，降低脱靶效应。我们基于 CasXX 的序列(SEQ ID NO: 8)，使用带有突变碱基的引物对 CasXX 的 DNA 序列进行 PCR，使用 NEBuilder® HiFi DNA Assembly Master Mix (NEB)试剂盒将纯化的 PCR 产物装载到 pCAG-2A-EGFP 质粒中，从而构建了表达引入相关的点突变的 CasXX 质粒。我们一共获得了 26 种基于 CasXX 序列、包含单个氨基酸突变的突变体，分别命名为 CasXX-HF-1 到 HF-26。具体的突变方式见表 9。其中，CasXX-HF-26 将 CasXX 中的 N164Y 点突变恢复成野生型 Cas12i2 的原始氨基酸 N，即 Y164N（即删减了 N164Y 这个突变）。

表 9：CasXX-HF 突变体

在参比酶 CasXX(SEQ ID NO: 8)基础上进行点突变：		
CasXX-HF 突变体编号	突变方式	SEQ ID NO:
HF-1	S565A	
HF-2	N297A	
HF-3	Q865A	
HF-4	T308A	

HF-5	R309A	
HF-6	N312A	
HF-7	N441A	
HF-8	R857A	14
HF-9	N861A	15
HF-10	Y119F	
HF-11	K433A	
HF-12	K807A	16
HF-13	N841A	
HF-14	N848A	17
HF-15	K845A	
HF-16	D782A	
HF-17	R715A	18
HF-18	R719A	19
HF-19	S766A	
HF-20	K394A	20
HF-21	H357A	21
HF-22	K844A	22
HF-23	E395A	
HF-24	S346A	
HF-25	R402A	
HF-26	Y164N	

构建靶向 EMX1-7 位点的 crRNA 的表达质粒

通过将含有靶序列的退火寡核苷酸使用 T4 ligase(NEB) 连接到 BasI (NEB) 消化处理的 pUC19-U6-crRNA 骨架中, 构建了在 293T 中表达 Cas 蛋白 crRNA 的载体。所述 crRNA 设计为靶向 EMX1-7 位点。具体序列见表 10。编码 DR 的核苷酸序列为 SEQ ID NO: 59。

表 10: 编码靶向 EMX1-7 的 spacer 的 DNA 序列

序列号	crRNA 靶向位点名称	spacer 编码序列
SEQ ID NO: 27	EMX1-7	TGGTTGCCACCCCTAGTCAT

细胞培养、转染以及荧光激活细胞分选法(FACS)

将 HEK293T 细胞在含 1% 青霉素-链霉素(Gibco) 和 10% 胎牛血清(Gibco) 的 DMEM(Gibco) 中培养。将细胞接种在 24-细胞培养皿(Corning) 中 16 小时, 直到细胞密度达到 70%。通过使用 Lipofectamine 3000 (Invitrogen), 将 600ng 编码 Cas 蛋白的质粒和 300ng 编码 crRNA 的质粒转染到 24 孔细胞培养皿培养的细胞中。转染 72h 后, 用胰蛋白酶-EDTA (0.05%) (Gibco) 消化细胞, 然后进行 GFP 荧光激活细胞分选(FACS)。

使用 T7E1 酶切方法检测 Cas 蛋白在 EMX1-7 靶位点以及脱靶位点的编辑效率

FACS 分选的 GFP 阳性 HEK293T 细胞用 40 μ L 缓冲液 L(bimake)裂解, 在 55 $^{\circ}$ C 下
 5 孵育 3 小时, 然后在 95 $^{\circ}$ C 下孵育 10 分钟。使用针对 EMX1-7 靶位点、EMX1-7 脱靶位点
 1、EMX1-7 脱靶位点 2、和 EMX1-7 脱靶位点 3(如图 10)的引物对在不同基因组位点中含
 有靶位点或者脱靶位点的 dsDNA 片段进行 PCR 扩增, 靶位点或者脱靶位点的序列见表
 11。之后, 使用 10 μ L PCR 产物, 进行再退火程序以形成异源双链 dsDNA。然后, 将混
 合物用 1/10 体积的 NEBufferTM 2.1 和 0.2 μ L T7 endonuclease I (NEB)在 37 $^{\circ}$ C 下处理 50 分
 10 钟。通过~2.5%琼脂糖凝胶电泳分析消化产物。根据条带的灰度值计算插入缺失的比例
 (Indel, %)。每个 Cas12i 突变体在靶位点或者脱靶位点的插入缺失的比例值见表 12, 对应
 图 11。实验结果表明, R857A, N861A, K807A, N848A, R715A, R719A, K394A, H357A,
 K844A 这些基于 CasXX 序列(SEQ ID NO: 8)的单点突变体(对应 SEQ ID NO: 14-22 的工程
 化 Cas12i 核酸酶, 见表 9)能够有效地降低在脱靶位点 EMX1-7-OT-1, EMX1-7-OT-2,
 EMX1-7-OT-3 的插入缺失比例, 具有更高的特异性。在实验结果中, 我们把脱靶位点处
 插入缺失率的降低和靶位点处编辑效率基本不变(或升高)定义为编辑特异性的提高。

15 表 11 EXM1-7 靶位点和脱靶位点序列

SEQ ID NO:	名称	缩写	序列
28	靶位点	EMX1-7	TGGTTGCCCACCCTAGTCAT (和 SEQ ID NO: 27 相同)
29	脱靶位点 1	EMX1-7-OT-1	TGTTTGCCCACCCTAGTCCT
30	脱靶位点 2	EMX1-7-OT-2	TGATTGCCCACCCTACTCCT
31	脱靶位点 3	EMX1-7-OT-3	TGGTCACCCACCCTAGTCTG

表 12 CasXX-HF1 至 CasXX-HF-26 在 EMX1-7 靶序列和脱靶序列编辑效率比较

CasXX-HF 突 变体编号	突变方式	在 EMX1-7 插入缺失的 比例值	在 EMX1-7- OT-1 插入缺失 的比例值	在 EMX1-7- OT-2 插入缺失 的比例值	在 EMX1-7-OT- 3 插入缺失的比 例值
HF-1	S565A	35.89	18.06	13.24	13.55
HF-2	N297A	39.41	19.88	16.24	11.69
HF-3	Q865A	42.77	23.78	10.50	12.67
HF-4	T308A	40.61	17.94	0.00	9.27
HF-5	R309A	43.83	27.64	18.95	14.92
HF-6	N312A	42.11	9.50	18.94	0.00
HF-7	N441A	36.82	14.41	10.99	9.03
HF-8	R857A	30.08	1.11	0.00	0.00
HF-9	N861A	34.97	10.35	0.00	1.52
HF-10	Y119F	38.00	22.47	13.99	14.17
HF-11	K433A	41.34	19.06	0.00	10.86
HF-12	K807A	5.20	18.89	0.00	1.68
HF-13	N841A	39.46	18.37	13.08	12.86
HF-14	N848A	30.98	0.00	0.00	0.00

HF-15	K845A	39.17	19.51	0.00	13.42
HF-16	D782A	37.33	11.45	0.00	5.19
HF-17	R715A	35.07	5.21	0.00	1.45
HF-18	R719A	41.08	18.94	0.00	8.16
HF-19	S766A	38.50	22.90	0.00	8.32
HF-20	K394A	37.30	0.00	0.00	0.00
HF-21	H357A	39.24	7.70	0.00	2.39
HF-22	K844A	39.22	3.89	0.00	0.00
HF-23	E395A	35.70	19.60	0.00	8.15
HF-24	S346A	37.03	18.33	13.00	9.37
HF-25	R402A	34.83	15.83	0.00	7.72
HF-26	Y164N	39.14	12.56	4.81	11.95
CasXX	参比	42.05	16.41	11.66	9.76

构建靶向 RNF2-1 位点的 crRNA 的表达质粒

- 通过将含有靶序列的退火寡核苷酸使用 T4 ligase(NEB) 连接到 BasI (NEB) 消化处理的 pUC19-U6-crRNA 骨架中, 构建了在 HEK293T 中表达 Cas 蛋白 crRNA 的载体。最终的 crRNA 都靶向同一个 RNF2-1 位点, 但是 spacer 序列有所不同。编码这些 spacer 的具体序列见表 13。其中 RNF2-1-FM 代表 crRNA 中的 spacer 序列与 RNF2-1 位点完全匹配。
- 5 RNF2-1-Mis-1/2 代表 crRNA 中的 spacer 序列与 RNF2-1 位点在第 1 个和第 2 个碱基位置不匹配, 在其余位置匹配。RNF2-1-Mis-5/6 代表 crRNA 中的 spacer 序列与 RNF2-1 位点在第 5 个和第 6 个碱基位置不匹配, 在其余位置匹配。RNF2-1-Mis-17/1 代表 crRNA 中的 spacer 序列与 RNF2-1 位点在第 17 个和第 18 个碱基位置不匹配, 在其余位置匹配。
- 10 RNF2-1-Mis-19/20 代表 crRNA 中的 spacer 序列与 RNF2-1 位点在第 19 个和第 20 个碱基位置不匹配, 在其余位置匹配。在此, 设置碱基不匹配的目的是为了模拟脱靶效应。编码 DR 的核苷酸序列为 SEQ ID NO: 59。

表 13 靶向 RNF2-1 的间隔序列的 DNA 编码序列

SEQ ID NO:	crRNA 名称	spacer 编码序列
32	crRNA-RNF2-1-FM	TACAGGAGGCAATAACAGAT
33	crRNA-RNF2-1-Mis-1/2	GCCAGGAGGCAATAACAGAT
34	crRNA-RNF2-1-Mis-5/6	TACATTAGGCAATAACAGAT
35	crRNA-RNF2-1-Mis-17/18	TACAGGAGGCAATAACCTAT
36	crRNA-RNF2-1-Mis-19/20	TACAGGAGGCAATAACAGCG

细胞培养、转染以及荧光激活细胞分选法(FACS)

- 15 将 HEK293T 细胞在含 1% 青霉素-链霉素(Gibco)和 10% 胎牛血清(Gibco)的 DMEM(Gibco)中培养。将细胞接种在 24-细胞培养皿(Corning)中 16 小时, 直到细胞密度达到 70%。通过使用 Lipofectamine 3000 (Invitrogen), 将 600ng 编码 Cas 蛋白的质粒和 300ng 编码 crRNA 的质粒, 转染到 24 孔细胞培养皿培养的细胞中。转染 72h 后, 用胰蛋白酶-EDTA (0.05%) (Gibco) 消化细胞, 然后进行 GFP 荧光激活细胞分选(FACS)。分选出
- 20 的是成功表达 Cas 酶的细胞。

使用 T7E1 酶切法检测 Cas12i 突变体在 RNF2-1 以及脱靶位点的编辑效率

FACS 分选的 GFP 阳性 HEK293T 细胞用 40 μ L 缓冲液 L(bimake)裂解, 在 55 $^{\circ}$ C 下孵育 3 小时, 然后在 95 $^{\circ}$ C 下孵育 10 分钟。使用 RNF2-1 引物对 RNF2-1 位点 dsDNA 片段进行 PCR 扩增。之后, 使用 10 μ L PCR 产物, 进行再退火程序以形成异源双链 dsDNA。然后, 将混合物用 1/10 体积的 NEBufferTM 2.1 和 0.2 μ L T7 endonuclease I (NEB) 在 37 $^{\circ}$ C 下处理 50 分钟。通过~2.5%琼脂糖凝胶电泳分析消化产物。根据条带的灰度值计算插入缺失的比例(Indel, %)。每个 Cas12i 突变体在使用不同 crRNA 时在靶位点的插入缺失的比例值见表 14, 对应图 12。实验结果表明, R857A, N861A, K807A, N848A, R715A, R719A, K394A, H357A, K844A 这些基于 CasXX 序列的单点突变体(对应 SEQ ID NOs: 14-22 的工程化 Cas12i 核酸酶, 见表 9)即便在使用模拟脱靶的 crRNA-RNF2-1-Mis-1/2, crRNA-RNF2-1-Mis-5/6, crRNA-RNF2-1-Mis-17/18 和 crRNA-RNF2-1-Mis-19/20 时依然能够有效降低插入缺失的比例。在实验结果中, 我们把脱靶位点处插入缺失率的降低和靶位点处编辑效率基本不变(或提高)定义为编辑特异性的提高。

表 14: 不同 CasXX-HF 突变体在使用靶向靶序列和模拟脱靶 spacer 序列的 crRNA

对 RNF2-1 位点的编辑效率

CasXX-HF 突变体编号	突变方式	On-target spacer (%)	Off-target spacer-1 (%)	Off-target spacer-2 (%)	Off-target spacer-3 (%)	Off-target spacer-4 (%)
HF-1	S565A	33.68	26.30	18.91	39.35	42.04
HF-2	N297A	41.56	29.55	21.42	42.23	43.42
HF-3	Q865A	43.64	28.30	21.79	43.02	43.62
HF-4	T308A	29.13	0.00	0.00	35.76	39.49
HF-5	R309A	32.22	30.67	0.00	26.92	38.92
HF-6	N312A	32.69	0.00	0.00	42.18	43.69
HF-7	N441A	38.05	12.86	10.40	17.67	31.53
HF-8	R857A	39.15	3.18	0.00	0.00	3.89
HF-9	N861A	44.45	11.79	3.42	3.41	41.61
HF-10	Y119F	35.25	11.46	9.31	15.33	32.11
HF-11	K433A	31.50	9.80	0.00	16.26	33.46
HF-12	K807A	30.46	12.48	0.00	0.00	33.29
HF-13	N841A	37.30	17.92	13.45	23.58	43.25
HF-14	N848A	23.74	0.00	0.00	0.00	32.51
HF-15	K845A	38.94	12.87	5.20	4.17	40.51
HF-16	D782A	41.08	12.28	12.86	4.56	29.88
HF-17	R715A	39.99	5.00	0.00	0.00	23.18
HF-18	R719A	39.31	5.07	0.00	0.00	2.81
HF-19	S766A	39.49	18.53	9.09	7.07	38.14
HF-20	K394A	35.86	0.00	0.00	0.00	3.20
HF-21	H357A	35.84	5.96	0.00	0.00	19.12
HF-22	K844A	37.23	2.25	0.00	0.00	15.07
HF-23	E395A	37.68	12.80	5.46	0.00	33.70
HF-24	S346A	34.71	15.28	8.65	8.03	40.87
HF-25	R402A	37.54	12.97	3.39	0.00	32.88
HF-26	Y164N	32.33	10.00	9.50	11.73	36.56

CasXX	参比	39.67	17.02	10.42	12.10	37.56
-------	----	-------	-------	-------	-------	-------

实施例 11: 在 CasXX 序列基础上引入新的氨基酸突变的突变体

针对在 CasXX 序列基础上引入新的氨基酸突变的突变体, 在荧光报告系统中筛选特异性提高的那些突变体, 本实施例和实施例 10 的区别在于体现编辑效率的分析手段由琼脂糖凝胶电泳改为荧光。

构建靶向 mCherry 基因的 crRNA 的表达质粒

通过将含有靶序列的退火寡核苷酸使用 T4 ligase(NEB)连接到 BasI (NEB)消化处理的 pUC19-U6-crRNA 骨架中, 构建了在 HEK293T 中表达 Cas 蛋白 crRNA 的载体。最终的 crRNA 都靶向同一个 mCherry 位点, 但是 spacer 序列有所不同。编码 spacer 的具体序列见表 15。其中 mCherry-FM 代表 crRNA 中的 spacer 序列与 mCherry 位点完全匹配。crRNA-mCherry-Mis-1/2 代表 crRNA 中的 spacer 序列与 crRNA-mCherry 位点在第 1 个和第 2 个碱基位置不匹配, 在其余位置匹配。crRNA-mCherry-Mis-5/6 代表 crRNA 中的 spacer 序列与 crRNA-mCherry 位点在第 5 个和第 6 个碱基位置不匹配, 在其余位置匹配。crRNA-mCherry-Mis-19/20 代表 crRNA 中的 spacer 序列与 crRNA-mCherry 位点在第 19 个和第 20 个碱基位置不匹配, 在其余位置匹配。编码 DR 的核苷酸序列为 SEQ ID NO: 59。

表 15: 靶向 mCherry 及模拟脱靶的 spacer 编码序列

SEQ ID NO:	crRNA 靶向位点名称	Spacer 编码序列
37	crRNA-mCherry-FM	ACCTTG TAGATGAACTCGCC
38	crRNA-mCherry-Mis-1/2	CACTTG TAGATGAACTCGCC
39	crRNA-mCherry-Mis-5/6	ACCTGTTAGATGAACTCGCC
40	crRNA-mCherry-Mis-19/20	ACCTTG TAGATGAACTCGAA

细胞培养、转染

将 HEK293T 细胞在含 1%青霉素-链霉素(Gibco)和 10%胎牛血清(Gibco)的 DMEM(Gibco)中培养。将细胞接种在 24-细胞培养皿(Corning)中 16 小时, 直到细胞密度达到 70%。通过使用 Lipofectamine 3000 (Invitrogen), 将 600ng 编码 Cas 蛋白的质粒, 300ng 编码 crRNA 的质粒和 100ng 编码 mCherry 的质粒, 转染到 24 孔细胞培养皿培养的细胞中。

使用流式分析的方法检测 Cas 蛋白在 mCherry 位点的编辑效率

转染 72h 后, 用胰蛋白酶-EDTA (0.05%) (Gibco)消化细胞, 然后用流式细胞仪分析, 计算 GFP, mCherry 荧光的比例。mCherry 的比例越低, 意味着 Cas 蛋白编辑 mCherry 基因的效率越高。流程示意图以及 mCherry 编辑效率公式见附图 13。

实验结果表明, R857A, R719A, K394A, K844A(分别对应 SEQ ID NO: 14、19、20、22)这些基于 CasXX 序列的单点突变体在使用模拟脱靶的 crRNA-mCherry-Mis-1/2,

crRNA-mCherry-Mis-5/6 和 crRNA-mCherry-Mis-19/20 时编辑 mCherry 基因的效率明显降低, 见附图 14(在脱靶位点 mCherry%越高, 脱靶效率越低)。在此, 我们把脱靶位点处插入缺失率的降低和靶位点处编辑效率基本不变(或升高)定义为编辑特异性的提高。

5 **实施例 12: 在 CasXX 序列基础上引入氨基酸组合突变, 在内源基因(组合)位点筛选特异性提高的突变体**

筛选实验准备: 构建突变体的表达质粒

10 基于实施例 10 和 11, 我们选取 R857A, R719A, K394A, K844A 这几个点突变进行进一步组合探究。我们基于 CasXX 的序列, 使用带有突变碱基的引物对 CasXX 的 DNA 序列进行 PCR, 我们使用 NEBuilder® HiFi DNA Assembly Master Mix (NEB)试剂盒将纯化的 PCR 产物装载到 pCAG-2A-EGFP 质粒中, 从而构建了带有组合突变的 CasXX 质粒。我们一共获得了 11 种基于 CasXX 序列(SEQ ID NO: 8)、包含组合突变的突变体命名及具体的突变方式见表 16。

表 16: CasXX-HF 突变体

在参比酶 CasXX(SEQ ID NO: 8)基础上进行点突变:		
CasXX-HF 突变体编号	进一步突变方式	SEQ ID NO:
HF-8	R857A	14
HF-18	R719A	19
HF-20	K394A	20
HF-22	K844A	22
HF-18+20	R719A/K394A	
HF-20+22	K394A/K844A	
HF-8+20	R857A/K394A	
HF-18+22	R719A/K844A	23
HF-8+18	R857A/R719A	
HF-8+22	R857A/K844A	24
HF-18+20+22	R719A/K394A/K844A	
HF-8+18+20	R857A/R719A/K394A	
HF-8+20+22	R857A/K394A/K844A	
HF-8+18+22	R857A/R719A/K844A	
HF-8+18+20+22	R857A/R719A/K394A/K844A	

15

构建靶向 EMX1-7 位点以及 RNF2-1 的 crRNA 的表达质粒

通过将含有靶序列的退火寡核苷酸使用 T4 ligase(NEB)连接到 BasI (NEB)消化处理的 pUC19-U6-crRNA 骨架中, 构建了在 293T 中表达 Cas 蛋白 crRNA 的载体。crRNA 如

实施例 10 所述进行设计, 靶向 EMX1-7 或 RNF2-1, 或模拟 RNF2-1 脱靶。Spacer 编码序列见表 10 和 13。

细胞培养、转染以及荧光激活细胞分选法(FACS)

5 将 HEK293T 细胞在含 1%青霉素-链霉素(Gibco)和 10%胎牛血清(Gibco)的 DMEM(Gibco)中培养。将细胞接种在 24-细胞培养皿(Corning)中 16 小时, 直到细胞密度达到 70%。通过使用 Lipofectamine 3000 (Invitrogen), 将 600ng 编码 Cas 蛋白的质粒和 300ng 编码 crRNA 的质粒, 转染到 24 孔细胞培养皿培养的细胞中。转染 72h 后, 用胰蛋白酶-EDTA (0.05%) (Gibco)消化细胞, 然后进行荧光激活细胞分选(FACS)。分选出的是成功表达 Cas 酶的细胞。

10 使用 T7E1 酶切方法检测 Cas 蛋白在靶位点以及脱靶位点的编辑效率

FACS 分选的古FP 阳性 HEK293T 细胞用 40 μ L 缓冲液 L(bimake)裂解, 在 55°C 下孵育 3 小时, 然后在 95°C 下孵育 10 分钟。使用实施例 10 所述相应的引物对在不同基因组位点中含有靶位点或者脱靶位点的 dsDNA 片段进行 PCR 扩增, 靶位点或者脱靶位点的序列见表 11 和 13。之后, 使用 10 μ L PCR 产物, 进行再退火程序以形成异源双链 dsDNA。然后, 将混合物用 1/10 体积的 NEBuffer™ 2.1 和 0.2 μ L T7 endonuclease I (NEB) 15 在 37°C 下处理 50 分钟。通过~2.5%琼脂糖凝胶电泳分析消化产物。根据条带的灰度值计算插入缺失的比例(Indel, %)。实验结果表明, R857A, R719A, K394A, K844A, R719A/K844A (SEQ ID NO: 23)以及 R857A/K844A (SEQ ID NO: 24)这些基于 CasXX 序列的突变体能够有效地降低在 EMX1-7-OT-1, EMX1-7-OT-2, EMX1-7-OT-3 脱靶位点的插入缺失比例, 并且不牺牲在靶位点的编辑效率, 见表 17, 对应图 15。同时, R857A, 20 R719A, K394A, K844A, R719A/K844A 以及 R857A/K844A 这些基于 CasXX 序列的突变体在使用模拟脱靶的 crRNA-RNF2-1-Mis-1/2, crRNA-RNF2-1-Mis-5/6, crRNA-RNF2-1-Mis-17/18 和 crRNA-RNF2-1-Mis-19/20 时能够有效降低插入缺失的比例, 并且不牺牲在靶位点的效率, 见表 18, 对应图 16。在此, 我们把脱靶位点处插入缺失率的降低和靶位点处编辑效率基本不变(或升高)定义为编辑特异性的提高。

表 17: 不同 CasXX-HF 突变体对 EMX1-7 的靶向和脱靶编辑效率

CasXX-HF 突变体	突变方式	在 EMX1-7 on-target 基因编辑效率	在 EMX1-7 off-target-1 基因编辑效率	在 EMX1-7 off-target-2 基因编辑效率	在 EMX1-7 off-target-3 基因编辑效率
HF-8	R857A	30.24	1.00	0.00	1.73
HF-18	R719A	36.78	5.71	0.00	3.22
HF-20	K394A	32.58	0.00	0.00	0.00
HF-22	K844A	32.34	2.91	0.00	0.00
HF-18+20	R719A/K394A	25.05	0.00	0.00	0.00

HF-20+22	K394A/K844A	18.64	0.00	0.00	0.00
HF-8+20	R857A/K394A	18.42	0.00	0.00	0.00
HF-18+22	R719A/K844A	32.71	0.00	0.00	0.00
HF-8+18	R857A/R719A	30.33	0.00	0.00	0.00
HF-8+22	R857A/K844A	22.81	0.00	0.00	0.00
HF-18+20+22	R719A/K394A/K844A	0.00	0.00	0.00	0.00
HF-8+18+20	R857A/R719A/K394A	0.00	0.00	0.00	0.00
HF-8+20+22	R857/K394A/K844A	12.61	0.00	0.00	0.00
HF-8+18+22	R857A/R719A/K844A	0.00	0.00	0.00	0.00
HF-8+18+20+22	R857A/R719A/K394A/K844A	0.00	0.00	0.00	0.00
CasXX	参比	35.73	14.54	9.91	16.29

表 18: 不同 CasXX-HF 突变体对 RNF2-1 的靶向和脱靶编辑效率

CasXX-HF 突变体	突变方式	在 RNF2-1 on-target 基因编辑效率	Mismatch-1,2	Mismatch-5,6	Mismatch-17,18	Mismatch-19,20
HF-8	R857A	40.45	8.50	0.00	0.00	29.50
HF-18	R719A	42.46	12.62	0.00	0.00	41.69
HF-20	K394A	40.26	0.00	0.00	0.00	7.22
HF-22	K844A	40.34	0.00	0.00	0.00	26.86
HF-18+20	R719A/K394A	27.46	0.00	0.00	0.00	2.40
HF-20+22	K394A/K844A	16.79	0.00	0.00	0.00	10.00
HF-8+20	R857A/K394A	6.93	0.00	0.00	0.00	0.00
HF-18+22	R719A/K844A	42.26	0.00	0.00	0.00	19.31
HF-8+18	R857A/R719A	32.49	0.00	0.00	0.00	19.08
HF-8+22	R857A/K844A	23.54	0.00	0.00	0.00	0.00
HF-18+20+22	R719A/K394A/K844A	0.00	0.00	0.00	0.00	0.00
HF-8+18+20	R857A/R719A/K394A	0.00	0.00	0.00	0.00	0.00
HF-8+20+22	R857A/K394A/K844A	7.51	0.00	0.00	0.00	0.00
HF-8+18+22	R857A/R719A/K844A	0.00	0.00	0.00	0.00	0.00
HF-8+18+20+22	R857A/R719A/K394A/K844A	0.00	0.00	0.00	0.00	0.00
CasXX	参比	41.50	25.08	23.71	24.80	45.69

本说明书中的 SEQ ID NO: 23 和 SEQ ID NO: 24 所示突变体(对应于表 16 中标出序列号的突变体)具有明显更高的特异性。

5

实施例 13: 使用 GUIDE-Seq 检测 CasXX 以及 CasXX + K394A 突变体的脱靶效应。
质粒构建

将编码 CasXX (SEQ ID NO: 8)以及 CasXX + K394A 突变体(SEQ ID NO: 20)的 DNA 序列装载到 pCAG-2A-EGFP 质粒中, 从而构建了表达 CasXX 或(CasXX + K394A)的质粒。通过将含有靶序列的退火寡核苷酸连接到 *BasI* 消化的 pUC19-U6-crRNA 骨架中构建了在 HEK293T 中表达 Cas12i 蛋白 crRNA 的载体。其中 crRNA 设计为包含能靶向 CD34-7 内源位点的间隔序列。编码 DR 的核苷酸序列为 SEQ ID NO: 59。编码 spacer 的核苷酸序列包含 SEQ ID NO: 78。

细胞培养、转染以及荧光激活细胞分选法(FACS)

将 HEK293T 细胞在含 1%青霉素-链霉素(Gibco)和 10%胎牛血清(Gibco)的 DMEM(Gibco)中培养。将细胞接种在 24-细胞培养皿(Corning)中 16 小时, 直到细胞密度达到 70%。通过使用 Lipofectamine 3000 (Invitrogen), 将 600ng 编码 CasXX 或(CasXX + K394A)蛋白的质粒、300ng 编码 crRNA 的质粒和 10pmol 的退火的双链 DNA 标签(序列见表 8), 转染到 24 孔细胞培养皿培养的细胞中。转染 72h 后, 用胰蛋白酶-EDTA (0.05%) (Gibco)消化细胞, 然后进行 GFP 荧光激活细胞分选。

全基因组范围检测 Cas12i2 突变体的脱靶效应

使用 E.Z.N.A.® MicroElute Genomic DNA Kit (Omega)试剂盒提取 FACS 分选的古FP-阳性 HEK293T 细胞用的基因组。纯化好的基因组使用用 Qubit 定量。使用 Covaris S220 仪器, 按照仪器推荐的程序将基因组打断到 500 bp 左右。然后使用 VAHTS Universal Pro DNA Library Prep Kit for Illumina(Vazyme)试剂盒进行 DNA 建库。建库流程参照参考文献(Tsai, SQ *et al.* “GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases,” *Nat Biotechnol.* 2015;33(2):187-197)。将建库产物进行高通量测序。分析测序结果, 寻找潜在的脱靶位点。我们对 CasXX 和(CasXX + K394A)在 CD34-7 内源位点的脱靶效应进行了系统的分析, 具体的分析结果见图 17。第一行序列为参考靶序列, 下面的序列分别代表靶序列和脱靶序列以及双链 DNA 标签在靶序列和脱靶序列富集的 reads 数目。图 17 说明 CasXX 在 CD34-7 位点存在脱靶效应, 但是 CasXX + K394A 突变体在该位点没有脱靶效应。因此, 在 CasXX 序列的基础上额外引入 K394A 点突变能够大幅度提高 CasXX 的保真度, 使得 CasXX 具有高的特异性。

尽管以上结合附图对本发明的实施方案进行了描述, 但本发明并不局限于上述的具体实施方案和应用领域, 上述的具体实施方案仅仅是示意性的、指导性的, 而不是限制性的。本领域的普通技术人员在本说明书的启示下和在不脱离本发明权利要求所保护的范围的情况下, 还可以做出很多种的形式, 这些均属于本发明保护之列。

序列表

编号	氨基酸序列
SEQ ID NO: 1	MSSAIKSYKSVLRPNERKNQLLKSTIQCLEDEGSAFFKMLQGLFGGITPEIVRFSTEQEKQQQDIALWCAVNWFRPVSQDS LTHTIASDNLVEKFEEYYGGTASDAIKQYFASIGESYYWDCRQYYDLCRELGVEVSDLTHDLEILCREKCLAVATESN QNNISISVLFGTGEKEDRSVKLRITKKILEAISNLKEIPKNVAPIQEILNVAKATKETFRQVYAGNLGAPSTLEKFIKDGQK EFDLKKLQTDLKKVIRGKSKERDWCCQEELRSYVEQNTIQYDLWAWGEMFNKAHTALKIKSTRNYNFAKQRLQEFKQEIQ SLNNLLVVKLNDFFDSEFFSGEETY TICVHHLGGKDL SKLYKAWEDDPADPENAIIVLCDDLKNNFKKEPIRNILRYIFTI RQEC SAQDILAAAKYNQQLDRYKSQKANPSVLGNQGF TWRNAVILPEKAQRNDRPNSDLRIWL YLKL RHPDGRWKKH HIPFYDTRFFQEIY AAGNSPVDTCQFRTPRFGYHLPKLDQTAIRVNKKHVKA AKTEARIRLAIQGGTLPVSNLKITEISATI NSKGQVRIPVKFDVGRQKGT LQIGDRFCGYDQNQTASHAYSLEWVVKEGQYHKELGCFVRFISSGDIVSITENRGNQFDQ LSYEGLAYPQYADWRKKASKFVSLWQITKKNKKEIVTVEAKEKFD AICKYQPRLYKFNKEYAYLLRDIVRGKSLVELQ QIRQEIFRFIEQDCGVTRLGSLSLSTLETVKAVKGIISYFSTALNASKNPNISDEQRKEFDPPEL FALLEKLELIRTRKKKQKV ERIANSLIQTCLENNIKFIRGEGDLSTNNATKKKANSRSMDWLARGVFNKIRQLAPMHNTLFGCGSLY TSHQDPLVHRN PDKAMKCRWAAIPVKDIGDWLRLKLSQNLRAKNIGTGEYYHQGVKEFLSHYELQDLEEEELKWRSDRKSNI PCWVLQN RLAEKLGKNEAVVYIPVRGGRIYFATHKVATGAVSIVFDQKQVWVCNADHVAAAANIALTVKGIGEQSSDEENPDGSR IKL QLTS
SEQ ID NO: 2	MSSAIKSYKSVLRPNERKNQLLKSTIQCLEDEGSAFFKMLQGLFGGITPEIVRFSTEQEKQQQDIALWCAVNWFRPVSQDS LTHTIASDNLVEKFEEYYGGTASDAIKQYFASIGESYYWDCRQYYDLCRELGVEVSDLTHDLEILCREKCLAVATESN QYNSISVLFGTGRKEDRSVKLRITKKILEAISNLKEIPKNVAPIQEILNVAKATKETFRQVYAGNLGAPSTLERFIKDGQK EFDLKKLQTDLKKVIRGKSKERDWCCQEELRSYVEQNTIQYDLWAWGEMFNKAHTALKIKSTRNYNFAKQRLQEFKQEIQ SLNNLLVVKLNDFFDSEFFSGEETY TICVHHLGGKDL SKLYKAWEDDPADPENAIIVLCDDLKNNFKKEPIRNILRYIFTI RQEC SAQDILAAAKYNQQLDRYKSQKANPSVLGNQGF TWRNAVILPEKAQRNDRPNSDLRIWL YLKL RHPDGRWKKH HIPFYDTRFFQEIY AAGNSPVDTCQFRTPRFGYHLPKLDQTAIRVNKKHVKA AKTEARIRLAIQGGTLPVSNLKITEISATI NSKGQVRIPVKFDVGRQKGT LQIGDRFCGYDQNQTASHAYSLEWVVKEGQYHKELGCFVRFISSGDIVSITENRGNQFDQ LSYEGLAYPQYADWRKKASKFVSLWQITKKNKKEIVTVEAKEKFD AICKYQPRLYKFNKEYAYLLRDIVRGKSLVELQ QIRQEIFRFIEQDCGVTRLGSLSLSTLETVKAVKGIISYFSTALNASKNPNISDEQRKEFDPPEL FALLEKLELIRTRKKKQKV ERIANSLIQTCLENNIKFIRGEGDLSTNNATKKKANSRSMDWLARGVFNKIRQLAPMHNTLFGCGSLY TSHQDPLVHRN PDKAMKCRWAAIPVKDIGDWLRLKLSQNLRAKNIGTGEYYHQGVKEFLSHYELQDLEEEELKWRSDRKSNI PCWVLQN RLAEKLGKNEAVVYIPVRGGRIYFATHKVATGAVSIVFDQKQVWVCNADHVAAAANIALTVKGIGEQSSDEENPDGSR IKL QLTS
SEQ ID NO: 3	MSSAIKSYKSVLRPNERKNQLLKSTIQCLEDEGSAFFKMLQGLFGGITPEIVRFSTEQEKQQQDIALWCAVNWFRPVSQDS LTHTIASDNLVEKFEEYYGGTASDAIKQYFASIGESYYWDCRQYYDLCRELGVEVSDLTHDLEILCREKCLAVATESN QNNISISVLFGTGRKEDRSVKLRITKKILEAISNLKEIPKNVAPIQEILNVAKATKETFRQVYAGNLGAPSTLERFIKDGQK EFDLKKLQTDLKKVIRGKSKERDWCCQEELRSYVEQNTIQYDLWAWGEMFNKAHTALKIKSTRNYNFAKQRLQEFKQEIQ SLNNLLVVKLNDFFDSEFFSGEETY TICVHHLGGKDL SKLYKAWEDDPADPENAIIVLCDDLKNNFKKEPIRNILRYIFTI RQEC SAQDILAAAKYNQQLDRYKSQKANPSVLGNQGF TWRNAVILPEKAQRNDRPNSDLRIWL YLKL RHPDGRWKKH HIPFYDTRFFQEIY AAGNSPVDTCQFRTPRFGYHLPKLDQTAIRVNKKHVKA AKTEARIRLAIQGGTLPVSNLKITEISATI NSKGQVRIPVKFDVGRQKGT LQIGDRFCGYDQNQTASHAYSLEWVVKEGQYHKELGCFVRFISSGDIVSITENRGNQFDQ LSYEGLAYPQYADWRKKASKFVSLWQITKKNKKEIVTVEAKEKFD AICKYQPRLYKFNKEYAYLLRDIVRGKSLVELQ QIRQEIFRFIEQDCGVTRLGSLSLSTLETVKAVKGIISYFSTALNASKNPNISDEQRKEFDPPEL FALLEKLELIRTRKKKQKV ERIANSLIQTCLENNIKFIRGEGDLSTNNATKKKANSRSMDWLARGVFNKIRQLAPMHNTLFGCGSLY TSHQDPLVHRN PDKAMKCRWAAIPVKDIGDWLRLKLSQNLRAKNIGTGEYYHQGVKEFLSHYELQDLEEEELKWRSDRKSNI PCWVLQN RLAEKLGKNEAVVYIPVRGGRIYFATHKVATGAVSIVFDQKQVWVCNADHVAAAANIALTVKGIGEQSSDEENPDGSR IKL QLTS
SEQ ID NO: 4	MSSAIKSYKSVLRPNERKNQLLKSTIQCLEDEGSAFFKMLQGLFGGITPEIVRFSTEQEKQQQDIALWCAVNWFRPVSQDS LTHTIASDNLVEKFEEYYGGTASDAIKQYFASIGESYYWDCRQYYDLCRELGVEVSDLTHDLEILCREKCLAVATESN QNNISISVLFGTGRKEDRSVKLRITKKILEAISNLKEIPKNVAPIQEILNVAKATKETFRQVYAGNLGAPSTLERFIKDGQK EFDLKKLQTDLKKVIRGKSKERDWCCQEELRSYVEQNTIQYDLWAWGEMFNKAHTALKIKSTRNYNFAKQRLQEFKQEIQ SLNNLLVVKLNDFFDSEFFSGEETY TICVHHLGGKDL SKLYKAWEDDPADPENAIIVLCDDLKNNFKKEPIRNILRYIFTI RQEC SAQDILAAAKYNQQLDRYKSQKANPSVLGNQGF TWRNAVILPEKAQRNDRPNSDLRIWL YLKL RHPDGRWKKH HIPFYDTRFFQEIY AAGNSPVDTCQFRTPRFGYHLPKLDQTAIRVNKKHVKA AKTEARIRLAIQGGTLPVSNLKITEISATI NSKGQVRIPVKFDVGRQKGT LQIGDRFCGYDQNQTASHAYSLEWVVKEGQYHKELGCFVRFISSGDIVSITENRGNQFDQ LSYEGLAYPQYADWRKKASKFVSLWQITKKNKKEIVTVEAKEKFD AICKYQPRLYKFNKEYAYLLRDIVRGKSLVELQ QIRQEIFRFIEQDCGVTRLGSLSLSTLETVKAVKGIISYFSTALNASKNPNISDEQRKEFDPPEL FALLEKLELIRTRKKKQKV ERIANSLIQTCLENNIKFIRGEGDLSTNNATKKKANSRSMDWLARGVFNKIRQLAPMHNTLFGCGSLY TSHQDPLVHRN PDKAMKCRWAAIPVKDIGDWLRLKLSQNLRAKNIGTGEYYHQGVKEFLSHYELQDLEEEELKWRSDRKSNI PCWVLQN RLAEKLGKNEAVVYIPVRGGRIYFATHKVATGAVSIVFDQKQVWVCNADHVAAAANIALTVKGIGEQSSDEENPDGSR IKL QLTS
SEQ ID NO: 5	MSSAIKSYKSVLRPNERKNQLLKSTIQCLEDEGSAFFKMLQGLFGGITPEIVRFSTEQEKQQQDIALWCAVNWFRPVSQDS LTHTIASDNLVEKFEEYYGGTASDAIKQYFASIGESYYWDCRQYYDLCRELGVEVSDLTHDLEILCREKCLAVATESN QYNSISVLFGTGEKEDRSVKLRITKKILEAISNLKEIPKNVAPIQEILNVAKATKETFRQVYAGNLGAPSTLEKFIKDGQK EFDLKKLQTDLKKVIRGKSKERDWCCQEELRSYVEQNTIQYDLWAWGEMFNKAHTALKIKSTRNYNFAKQRLQEFKQEIQ SLNNLLVVKLNDFFDSEFFSGEETY TICVHHLGGKDL SKLYKAWEDDPADPENAIIVLCDDLKNNFKKEPIRNILRYIFTI RQEC SAQDILAAAKYNQQLDRYKSQKANPSVLGNQGF TWRNAVILPEKAQRNDRPNSDLRIWL YLKL RHPDGRWKKH HIPFYDTRFFQEIY AAGNSPVDTCQFRTPRFGYHLPKLDQTAIRVNKKHVKA AKTEARIRLAIQGGTLPVSNLKITEISATI NSKGQVRIPVKFDVGRQKGT LQIGDRFCGYDQNQTASHAYSLEWVVKEGQYHKELGCFVRFISSGDIVSITENRGNQFDQ LSYEGLAYPQYADWRKKASKFVSLWQITKKNKKEIVTVEAKEKFD AICKYQPRLYKFNKEYAYLLRDIVRGKSLVELQ QIRQEIFRFIEQDCGVTRLGSLSLSTLETVKAVKGIISYFSTALNASKNPNISDEQRKEFDPPEL FALLEKLELIRTRKKKQKV ERIANSLIQTCLENNIKFIRGEGDLSTNNATKKKANSRSMDWLARGVFNKIRQLAPMHNTLFGCGSLY TSHQDPLVHRN PDKAMKCRWAAIPVKDIGDWLRLKLSQNLRAKNIGTGEYYHQGVKEFLSHYELQDLEEEELKWRSDRKSNI PCWVLQN RLAEKLGKNEAVVYIPVRGGRIYFATHKVATGAVSIVFDQKQVWVCNADHVAAAANIALTVKGIGEQSSDEENPDGSR IKL QLTS
SEQ ID NO: 6	MSSAIKSYKSVLRPNERKNQLLKSTIQCLEDEGSAFFKMLQGLFGGITPEIVRFSTEQEKQQQDIALWCAVNWFRPVSQDS LTHTIASDNLVEKFEEYYGGTASDAIKQYFASIGESYYWDCRQYYDLCRELGVEVSDLTHDLEILCREKCLAVATESN

	QYNSIISVLFGTGEKEDRSVKLRITKKILEAISNLKEIPKNVAPIQEILNVAKATKETFRQVYAGNLGAPSTLEKFIKADGQK EFDLKKLQTDLKKVIRGKSKERDWCCQEELRSYVEQNTIQYDLWAWGEMFNKAHTALKIKSTRNYNFAKQRLEQFKRIQ SLNLLVVKLNDFFDSEFFSGEETY TICVHHLGGKRLSKLYKAWEDDPADPENAIIVL CDDLKNFCKEPIRNILRYIFTI RQEC SAQDILAAAKYNQQLDRYKSQKANPSVLGNQGF TWNAVILPEKAQRNDRPNSDLRIWL YLKL RHPDGRWKKH HIPFYDTRFFQEYI AAGNSPVDTCQFRTPRFGYHLPKLDQTAIRVNKKHVKA AKTEARIRLAIQGGTLPVSNLKITRISATI NSKGQVRIPVKFDVGRQKGT LQIGDRFCGYDQNTASHAYSLEWVVEGQYHKELGCFVRFISSGDIVSITENRGNQFDQ LSYEGLAYPQYADWRKKASKFVSLWQITKKNKKEIVTVEAKEKFD AICKYQPRLYKFNKEYAYLLRDIVRGKSLVELQ QIRQEIFRFEIQDCGVTRLGSLSLSTLETVKAVKGIYSYFSTALNASKNPNISDEQRKEFDPPEL FALLEKLELIRTRKKKQKV ERIANSLIQTCLENNIKFIRGEGDLSTNNATKKKANSRSMDWLARGVFNKIRQLAPMHNITLFGCGSLY TSHQDPLVHRN PDKAMKCRWAAIPVKDIGDWLRLKLSQNLRAKNIGTGEYHQGVKEFLSHYELQDLEEEELKWRSDRKSNI PCWVLQN RLAEKLGKNEAVVYIPVRGGRIYFATHKVATGAVSIVFDQKQVWVCNADHVAAAANIALTVKGIGEQSSDEENPDGSR IKL QLTS
SEQ ID NO. 7	MSSAIKSYKSVLRPNERKNQLLSTIQCLELDSAFFKMLQGLFGGITPEIVRFSTEQEKQQQDIALWCAVNWFRPVSQDS LTHTIASDNLVEKFEEY YGGTASDAIKQYFSA SIGESY YWDCRQYYDLCREL GVEVSDLTHDLEILCREKCLAVATESN QYNSIISVLFGTGRKEDRSVKLRITKKILEAISNLKEIPKNVAPIQEILNVAKATKETFRQVYAGNLGAPSTLERFIKADGQK EFDLKKLQTDLKKVIRGKSKERDWCCQEELRSYVEQNTIQYDLWAWGEMFNKAHTALKIKSTRNYNFAKQRLEQFKRIQ SLNLLVVKLNDFFDSEFFSGEETY TICVHHLGGKRLSKLYKAWEDDPADPENAIIVL CDDLKNFCKEPIRNILRYIFTI RQEC SAQDILAAAKYNQQLDRYKSQKANPSVLGNQGF TWNAVILPEKAQRNDRPNSDLRIWL YLKL RHPDGRWKKH HIPFYDTRFFQEYI AAGNSPVDTCQFRTPRFGYHLPKLDQTAIRVNKKHVKA AKTEARIRLAIQGGTLPVSNLKITRISATI NSKGQVRIPVKFDVGRQKGT LQIGDRFCGYDQNTASHAYSLEWVVEGQYHKELGCFVRFISSGDIVSITENRGNQFDQ LSYEGLAYPQYADWRKKASKFVSLWQITKKNKKEIVTVEAKEKFD AICKYQPRLYKFNKEYAYLLRDIVRGKSLVELQ QIRQEIFRFEIQDCGVTRLGSLSLSTLETVKAVKGIYSYFSTALNASKNPNISDEQRKEFDPPEL FALLEKLELIRTRKKKQKV ERIANSLIQTCLENNIKFIRGEGDLSTNNATKKKANSRSMDWLARGVFNKIRQLAPMHNITLFGCGSLY TSHQDPLVHRN PDKAMKCRWAAIPVKDIGDWLRLKLSQNLRAKNIGTGEYHQGVKEFLSHYELQDLEEEELKWRSDRKSNI PCWVLQN RLAEKLGKNEAVVYIPVRGGRIYFATHKVATGAVSIVFDQKQVWVCNADHVAAAANIALTVKGIGEQSSDEENPDGSR IKL QLTS
SEQ ID NO. 8	MSSAIKSYKSVLRPNERKNQLLSTIQCLELDSAFFKMLQGLFGGITPEIVRFSTEQEKQQQDIALWCAVNWFRPVSQDS LTHTIASDNLVEKFEEY YGGTASDAIKQYFSA SIGESY YWDCRQYYDLCREL GVEVSDLTHDLEILCREKCLAVATESN QYNSIISVLFGTGRKEDRSVKLRITKKILEAISNLKEIPKNVAPIQEILNVAKATKETFRQVYAGNLGAPSTLERFIKADGQK EFDLKKLQTDLKKVIRGKSKERDWCCQEELRSYVEQNTIQYDLWAWGEMFNKAHTALKIKSTRNYNFAKQRLEQFKRIQ SLNLLVVKLNDFFDSEFFSGEETY TICVHHLGGKRLSKLYKAWEDDPADPENAIIVL CDDLKNFCKEPIRNILRYIFTI RQEC SAQDILAAAKYNQQLDRYKSQKANPSVLGNQGF TWNAVILPEKAQRNDRPNSDLRIWL YLKL RHPDGRWKKH HIPFYDTRFFQEYI AAGNSPVDTCQFRTPRFGYHLPKLDQTAIRVNKKHVKA AKTEARIRLAIQGGTLPVSNLKITRISATI NSKGQVRIPVKFDVGRQKGT LQIGDRFCGYDQNTASHAYSLEWVVEGQYHKELGCFVRFISSGDIVSITENRGNQFDQ LSYEGLAYPQYADWRKKASKFVSLWQITKKNKKEIVTVEAKEKFD AICKYQPRLYKFNKEYAYLLRDIVRGKSLVELQ QIRQEIFRFEIQDCGVTRLGSLSLSTLETVKAVKGIYSYFSTALNASKNPNISDEQRKEFDPPEL FALLEKLELIRTRKKKQKV ERIANSLIQTCLENNIKFIRGEGDLSTNNATKKKANSRSMDWLARGVFNKIRQLAPMHNITLFGCGSLY TSHQDPLVHRN PDKAMKCRWAAIPVKDIGDWLRLKLSQNLRAKNIGTGEYHQGVKEFLSHYELQDLEEEELKWRSDRKSNI PCWVLQN RLAEKLGKNEAVVYIPVRGGRIYFATHKVATGAVSIVFDQKQVWVCNADHVAAAANIALTVKGIGEQSSDEENPDGSR IKL QLTS
SEQ ID NO. 9	MSSAIKSYKSVLRPNERKNQLLSTIQCLELDSAFFKMLQGLFGGITPEIVRFSTEQEKQQQDIALWCAVNWFRPVSQDS LTHTIASDNLVEKFEEY YGGTASDAIKQYFSA SIGESY YWDCRQYYDLCREL GVEVSDLTHDLEILCREKCLAVATESN QYNSIISVLFGTGRKEDRSVKLRITKKILEAISNLKEIPKNVAPIQEILNVAKATKETFRQVYAGNLGAPSTLERFIKADGQK EFDLKKLQTDLKKVIRGKSKERDWCCQEELRSYVEQNTIQYDLWAWGEMFNKAHTALKIKSTRNYNFAKQRLEQFKRIQ SLNLLVVKLNDFFDSEFFSGEETY TICVHHLGGKRLSKLYKAWEDDPADPENAIIVL CDDLKNFCKEPIRNILRYIFTI RQEC SAQDILAAAKYNQQLDRYKSQKANPSVLGNQGF TWNAVILPEKAQRNDRPNSDLRIWL YLKL RHPDGRWKKH HIPFYDTRFFQEYI AAGNSPVDTCQFRTPRFGYHLPKLDQTAIRVNKKHVKA AKTEARIRLAIQGGTLPVSNLKITRISATI NSKGQVRIPVKFDVGRQKGT LQIGDRFCGYDQNTASHAYSLEWVVEGQYHKELGCFVRFISSGDIVSITENRGNQFDQ LSYEGLAYPQYADWRKKASKFVSLWQITKKNKKEIVTVEAKEKFD AICKYQPRLYKFNKEYAYLLRDIVRGKSLVELQ QIRQEIFRFEIQDCGVTRLGSLSLSTLETVKAVKGIYSYFSTALNASKNPNISDEQRKEFDPPEL FALLEKLELIRTRKKKQKV ERIANSLIQTCLENNIKFIRGEGDLSTNNATKKKANSRSMDWLARGVFNKIRQLAPMHNITLFGCGSLY TSHQDPLVHRN PDKAMKCRWAAIPVKDIGDWLRLKLSQNLRAKNIGTGEYHQGVKEFLSHYELQDLEEEELKWRSDRKSNI PCWVLQN RLAEKLGKNEAVVYIPVRGGRIYFATHKVATGAVSIVFDQKQVWVCNADHVAAAANIALTVKGIGEQSSDEENPDGSR IKL QLTS
SEQ ID NO. 10	MSSAIKSYKSVLRPNERKNQLLSTIQCLELDSAFFKMLQGLFGGITPEIVRFSTEQEKQQQDIALWCAVNWFRPVSQDS LTHTIASDNLVEKFEEY YGGTASDAIKQYFSA SIGESY YWDCRQYYDLCREL GVEVSDLTHDLEILCREKCLAVATESN QYNSIISVLFGTGRKEDRSVKLRITKKILEAISNLKEIPKNVAPIQEILNVAKATKETFRQVYAGNLGAPSTLERFIKADGQK EFDLKKLQTDLKKVIRGKSKERDWCCQEELRSYVEQNTIQYDLWAWGEMFNKAHTALKIKSTRNYNFAKQRLEQFKRIQ SLNLLVVKLNDFFDSEFFSGEETY TICVHHLGGKRLSKLYKAWEDDPADPENAIIVL CDDLKNFCKEPIRNILRYIFTI RQEC SAQDILAAAKYNQQLDRYKSQKANPSVLGNQGF TWNAVILPEKAQRNDRPNSDLRIWL YLKL RHPDGRWKKH HIPFYDTRFFQEYI AAGNSPVDTCQFRTPRFGYHLPKLDQTAIRVNKKHVKA AKTEARIRLAIQGGTLPVSNLKITRISATI NSKGQVRIPVKFDVGRQKGT LQIGDRFCGYDQNTASHAYSLEWVVEGQYHKELGCFVRFISSGDIVSITENRGNQFDQ LSYEGLAYPQYADWRKKASKFVSLWQITKKNKKEIVTVEAKEKFD AICKYQPRLYKFNKEYAYLLRDIVRGKSLVELQ QIRQEIFRFEIQDCGVTRLGSLSLSTLETVKAVKGIYSYFSTALNASKNPNISDEQRKEFDPPEL FALLEKLELIRTRKKKQKV ERIANSLIQTCLENNIKFIRGEGDLSTNNATKKKANSRSMDWLARGVFNKIRQLAPMHNITLFGCGSLY TSHQDPLVHRN PDKAMKCRWAAIPVKDIGDWLRLKLSQNLRAKNIGTGEYHQGVKEFLSHYELQDLEEEELKWRSDRKSNI PCWVLQN RLAEKLGKNEAVVYIPVRGGRIYFATHKVATGAVSIVFDQKQVWVCNADHVAAAANIALTVKGIGEQSSDEENPDGSR IKL QLTS
SEQ ID NO. 11	MSSAIKSYKSVLRPNERKNQLLSTIQCLELDSAFFKMLQGLFGGITPEIVRFSTEQEKQQQDIALWCAVNWFRPVSQDS LTHTIASDNLVEKFEEY YGGTASDAIKQYFSA SIGESY YWDCRQYYDLCREL GVEVSDLTHDLEILCREKCLAVATESN QYNSIISVLFGTGRKEDRSVKLRITKKILEAISNLKEIPKNVAPIQEILNVAKATKETFRQVYAGNLGAPSTLERFIKADGQK EFDLKKLQTDLKKVIRGKSKERDWCCQEELRSYVEQNTIQYDLWAWGEMFNKAHTALKIKSTRNYNFAKQRLEQFKRIQ SLNLLVVKLNDFFDSEFFSGEETY TICVHHLGGKRLSKLYKAWEDDPADPENAIIVL CDDLKNFCKEPIRNILRYIFTI RQEC SAQDILAAAKYNQQLDRYKSQKANPSVLGNQGF TWNAVILPEKAQRNDRPNSDLRIWL YLKL RHPDGRWKKH HIPFYDTRFFQEYI AAGNSPVDTCQFRTPRFGYHLPKLDQTAIRVNKKHVKA AKTEARIRLAIQGGTLPVSNLKITRISATI NSKGQVRIPVKFDVGRQKGT LQIGDRFCGYDQNTASHAYSLEWVVEGQYHKELGCFVRFISSGDIVSITENRGNQFDQ LSYEGLAYPQYADWRKKASKFVSLWQITKKNKKEIVTVEAKEKFD AICKYQPRLYKFNKEYAYLLRDIVRGKSLVELQ QIRQEIFRFEIQDCGVTRLGSLSLSTLETVKAVKGIYSYFSTALNASKNPNISDEQRKEFDPPEL FALLEKLELIRTRKKKQKV ERIANSLIQTCLENNIKFIRGEGDLSTNNATKKKANSRSMDWLARGVFNKIRQLAPMHNITLFGCGSLY TSHQDPLVHRN PDKAMKCRWAAIPVKDIGDWLRLKLSQNLRAKNIGTGEYHQGVKEFLSHYELQDLEEEELKWRSDRKSNI PCWVLQN RLAEKLGKNEAVVYIPVRGGRIYFATHKVATGAVSIVFDQKQVWVCNADHVAAAANIALTVKGIGEQSSDEENPDGSR IKL QLTS

	DQLSYEGLAYPQYADWRKKASKFVSLWQITKKNKKKEIVTVEAKEKFDACKYQPRLYKFNKEYAYLLRDIVRGKSLVELQQRIRQEIFRFEIQDCGVTRLGSLSLSTLETVKAVKGIISYFSTALNASKNNPISDEQRKEFDPPELFALEKLELIRTRKKKQKVERIANSLIQTCLENNIKFIRGEGDLSTTNNATKKKANSRSMDWLARGVFNKIRQLAPMHNTLFGCGSLYTSHQDPLVHRNPDKAMKCRWAAIPVKDIGDWVLRKLSQNLRAKNIGTGEYYHQGVKEFLSHYELQDLEEEELKWRSDRKSNIPCWVLQNRLAEKLGNKEAVVYIPVRGGRIYFATHKVATGAVSIVFDQKQVWVCNADHVAAAANIALTVKGIGEQSSDEENPDGSRIKLQTS
SEQ ID NO. 12	MSSAIKSYKSVLRPNERKNQLLKSTIQCLEDEGSAFFKMLQGLFGGITPEIVRFSTEQEQQQDIALWCAVNWFRPVSQDSLTHTIASDNLVEKFEEYGGTASDAIKQYFSAISIGESYYWDCRQYYDLCRELGVEVSDLTHDLEILCREKCLAVATESNQYNSIISVLFVGTGRKEDRSVKLRITKKILEAISNLKEIPKNVAPIQEIILNVAKATKETFRQVYAGNLGAPSTLERFIAKDGGKEFDLKKLQTDLKKVIRGKSKERDWCCQEELRSYVEQNTIQYDLWAWGEMFNKAHTALKIKSTRNYNFAKQRLQEQFKRIQSLNLLVVKLNDFFDSEFFSGEETYTICVHHLGGKRLSKLYKAWEDDPADPENAIIVL CDDLKNNFKKEPIRNILRYIFTIRQECQAQDILAAAKYNQQLDRYKSQKANPSVLGNQGGFTWRNAVILPEKAQRNDRPNSLDLRIWLYLKL RHPDGRWKKHHIPFYDTRFFQEIYAAGNSPVDTCQFRTPRFGYHLPKLDQTAIRVNNKHHVKAATEARIRLAIQGGTLPVSNLKITRISATINSKQVVRIPVKFDVGRQKGTQIGDRFCGYDQNTASHAYSLEWVVEKQYHKLGCVFRFISGGDIVSITENRGNQFDQLSYEGLAYPQYADWRKKASKFVSLWQITKKNKKKEIVTVEAKEKFDACKYQPRLYKFNKEYAYLLRDIVRGKSLVELQQRIRQEIFRFEIQDCGVTRLGSLSLSTLETVKAVKGIISYFSTALNASKNNPISDEQRKEFDPPELFALEKLELIRTRKKKQKVERIANSLIQTCLENNIKFIRGEGDLSTTNNATKKKANSRSMDWLARGVFNKIRQLAPMHNTLFGCGSLYTSHQDPLVHRNPDKAMKCRWAAIPVKDIGDWVLRKLSQNLRAKNIGTGEYYHQGVKEFLSHYELQDLEEEELKWRSDRKSNIPCWVLQNRLAEKLGNKEAVVYIPVRGGRIYFATHKVATGAVSIVFDQKQVWVCNADHVAAAANIALTVKGIGEQSSDEENPDGSRIKLQTS
SEQ ID NO. 13	MSNKEKNASETRKAYTTKMIPRSHDRMKLGNFMDYLMGDTPIFFELWNQFGGGIDRDIISGTANKDKISDLDLAVNWFKVMPINSKPGVSPSNLANLFQYSGSEPDIAQEQYFASNFDTEKHQWKMVRVEYERLLAELQLSRSDMHDLKLMYKEKCIGLSLSTAHYITSMVFGTGAKNNRQTKHQFYSKVIQLEESTQINSVEQLASILKAGDCDSYRKLIRCSRKGATPSILKIVQDYELGTNHDDDEVNVPISLANLKEKLGREFEYCEWCKMEKIKAFASKVGPYYLGSYSAMLENALSPIKGMSTTKNCKFVLLKQIDAKNDIKYENEPFGKIEGFFDSPYFESDITNVKVVWVLPHPHIGESNIKTLEWDLNAIHSKYEEDIASLTKKKEKRIKVVYQGDVQCTINTYCEEVGEAKTPLVQLLRYLYSRKDDIAVDKIIDGITFLSKKHKVEKQKINPVIQKYPSPFNFGNSKLLGKIISPKDKLKHNLKCNRNQVDNYIWEIKVLNTKTMREKHHYALSSSTRFLEEYYPATSENPPDALAARFRKTNGYEPALSAEQIEQIRSAAPVGLRKKVKKRQMRLEAARQONLLPRYTGWKDFNINICKRGNFVEVTLATKVKKKKEKNYKVVLYGDANIVRKNNTYAAIEAHANGDGVIDYNDLPVKPIESGFVTVESQVRDKSYDQLSYNGVKLLYCKPHVESRRSFLEKYRNGTMKDNDRGNNIQIDFMKDFEAIADDETSLYFNMKYCKLLQSSIRNHSQAKEYREEIFELRDGKLSVLLKSSLSNLSFVMEKVAKSLIGTYFGHLLKPKNSKSDVKAPPITDEDKQKADPEMFLRLALEEKRLNKVSKKEVIANKIVAKALELRDKYFVPLIKGENISDITTKGKKSSTNSFLMDWLRAGVANVKEMVMHMQGLEFVEVNPNTSHQDPPFVHNKPNFTFRARYSRCTPSELTEKNRKEILSFLSDKPSKRPTNAYYNEGAMAFATYGLKKNVDLVGSLEKFKQIMANILHQRSEDLQFPSPRGGMFYLATYKLDADATSVNWNKGQFVWCNADLVAAYNVGLVDIQKDFK
SEQ ID NO. 14 CasXX - R857A	MSSAIKSYKSVLRPNERKNQLLKSTIQCLEDEGSAFFKMLQGLFGGITPEIVRFSTEQEQQQDIALWCAVNWFRPVSQDSLTHTIASDNLVEKFEEYGGTASDAIKQYFSAISIGESYYWDCRQYYDLCRELGVEVSDLTHDLEILCREKCLAVATESNQYNSIISVLFVGTGRKEDRSVKLRITKKILEAISNLKEIPKNVAPIQEIILNVAKATKETFRQVYAGNLGAPSTLERFIAKDGGKEFDLKKLQTDLKKVIRGKSKERDWCCQEELRSYVEQNTIQYDLWAWGEMFNKAHTALKIKSTRNYNFAKQRLQEQFKRIQSLNLLVVKLNDFFDSEFFSGEETYTICVHHLGGKRLSKLYKAWEDDPADPENAIIVL CDDLKNNFKKEPIRNILRYIFTIRQECQAQDILAAAKYNQQLDRYKSQKANPSVLGNQGGFTWRNAVILPEKAQRNDRPNSLDLRIWLYLKL RHPDGRWKKHHIPFYDTRFFQEIYAAGNSPVDTCQFRTPRFGYHLPKLDQTAIRVNNKHHVKAATEARIRLAIQGGTLPVSNLKITRISATINSKQVVRIPVKFDVGRQKGTQIGDRFCGYDQNTASHAYSLEWVVEKQYHKLGCVFRFISGGDIVSITENRGNQFDQLSYEGLAYPQYADWRKKASKFVSLWQITKKNKKKEIVTVEAKEKFDACKYQPRLYKFNKEYAYLLRDIVRGKSLVELQQRIRQEIFRFEIQDCGVTRLGSLSLSTLETVKAVKGIISYFSTALNASKNNPISDEQRKEFDPPELFALEKLELIRTRKKKQKVERIANSLIQTCLENNIKFIRGEGDLSTTNNATKKKANSRSMDWLARGVFNKIRQLAPMHNTLFGCGSLYTSHQDPLVHRNPDKAMKCRWAAIPVKDIGDWVLRKLSQNLRAKNIGTGEYYHQGVKEFLSHYELQDLEEEELKWRSDRKSNIPCWVLQNRLAEKLGNKEAVVYIPVRGGRIYFATHKVATGAVSIVFDQKQVWVCNADHVAAAANIALTVKGIGEQSSDEENPDGSRIKLQTS
SEQ ID NO. 15 CasXX - N861A	MSSAIKSYKSVLRPNERKNQLLKSTIQCLEDEGSAFFKMLQGLFGGITPEIVRFSTEQEQQQDIALWCAVNWFRPVSQDSLTHTIASDNLVEKFEEYGGTASDAIKQYFSAISIGESYYWDCRQYYDLCRELGVEVSDLTHDLEILCREKCLAVATESNQYNSIISVLFVGTGRKEDRSVKLRITKKILEAISNLKEIPKNVAPIQEIILNVAKATKETFRQVYAGNLGAPSTLERFIAKDGGKEFDLKKLQTDLKKVIRGKSKERDWCCQEELRSYVEQNTIQYDLWAWGEMFNKAHTALKIKSTRNYNFAKQRLQEQFKRIQSLNLLVVKLNDFFDSEFFSGEETYTICVHHLGGKRLSKLYKAWEDDPADPENAIIVL CDDLKNNFKKEPIRNILRYIFTIRQECQAQDILAAAKYNQQLDRYKSQKANPSVLGNQGGFTWRNAVILPEKAQRNDRPNSLDLRIWLYLKL RHPDGRWKKHHIPFYDTRFFQEIYAAGNSPVDTCQFRTPRFGYHLPKLDQTAIRVNNKHHVKAATEARIRLAIQGGTLPVSNLKITRISATINSKQVVRIPVKFDVGRQKGTQIGDRFCGYDQNTASHAYSLEWVVEKQYHKLGCVFRFISGGDIVSITENRGNQFDQLSYEGLAYPQYADWRKKASKFVSLWQITKKNKKKEIVTVEAKEKFDACKYQPRLYKFNKEYAYLLRDIVRGKSLVELQQRIRQEIFRFEIQDCGVTRLGSLSLSTLETVKAVKGIISYFSTALNASKNNPISDEQRKEFDPPELFALEKLELIRTRKKKQKVERIANSLIQTCLENNIKFIRGEGDLSTTNNATKKKANSRSMDWLARGVFNKIRQLAPMHNTLFGCGSLYTSHQDPLVHRNPDKAMKCRWAAIPVKDIGDWVLRKLSQNLRAKNIGTGEYYHQGVKEFLSHYELQDLEEEELKWRSDRKSNIPCWVLQNRLAEKLGNKEAVVYIPVRGGRIYFATHKVATGAVSIVFDQKQVWVCNADHVAAAANIALTVKGIGEQSSDEENPDGSRIKLQTS
SEQ ID NO. 16 CasXX - K807A	MSSAIKSYKSVLRPNERKNQLLKSTIQCLEDEGSAFFKMLQGLFGGITPEIVRFSTEQEQQQDIALWCAVNWFRPVSQDSLTHTIASDNLVEKFEEYGGTASDAIKQYFSAISIGESYYWDCRQYYDLCRELGVEVSDLTHDLEILCREKCLAVATESNQYNSIISVLFVGTGRKEDRSVKLRITKKILEAISNLKEIPKNVAPIQEIILNVAKATKETFRQVYAGNLGAPSTLERFIAKDGGKEFDLKKLQTDLKKVIRGKSKERDWCCQEELRSYVEQNTIQYDLWAWGEMFNKAHTALKIKSTRNYNFAKQRLQEQFKRIQSLNLLVVKLNDFFDSEFFSGEETYTICVHHLGGKRLSKLYKAWEDDPADPENAIIVL CDDLKNNFKKEPIRNILRYIFTIRQECQAQDILAAAKYNQQLDRYKSQKANPSVLGNQGGFTWRNAVILPEKAQRNDRPNSLDLRIWLYLKL RHPDGRWKKHHIPFYDTRFFQEIYAAGNSPVDTCQFRTPRFGYHLPKLDQTAIRVNNKHHVKAATEARIRLAIQGGTLPVSNLKITRISATINSKQVVRIPVKFDVGRQKGTQIGDRFCGYDQNTASHAYSLEWVVEKQYHKLGCVFRFISGGDIVSITENRGNQFDQLSYEGLAYPQYADWRKKASKFVSLWQITKKNKKKEIVTVEAKEKFDACKYQPRLYKFNKEYAYLLRDIVRGKSLVELQQRIRQEIFRFEIQDCGVTRLGSLSLSTLETVKAVKGIISYFSTALNASKNNPISDEQRKEFDPPELFALEKLELIRTRKKKQKVERIANSLIQTCLENNIKFIRGEGDLSTTNNATKKKANSRSMDWLARGVFNKIRQLAPMHNTLFGCGSLYTSHQDPLVHRNPDKAMKCRWAAIPVKDIGDWVLRKLSQNLRAKNIGTGEYYHQGVKEFLSHYELQDLEEEELKWRSDRKSNIPCWVLQNRLAEKLGNKEAVVYIPVRGGRIYFATHKVATGAVSIVFDQKQVWVCNADHVAAAANIALTVKGIGEQSSDEENPDGSRIKLQTS

<p>SEQ ID NO. 17 CasXX - N848A</p>	<p>MSSAIKSYKSVLRPNERKNQLLKSTIQCLEDGSAFFKMLQGLFGGITPEIVRFSTEQEKQQQDIALWCAVNWFRPVSQDS LHTIASDNLVEKFEEYYGGTASDAIKQYFSAIGESYYWDCRQYYDLCRELGVEVSDLTHDLEILCREKCLAVATESN QYNSISVLFVGTGRKEDRSVKLRITKKILEAISNLKEIPKNVAPIQEILNVAKATKETFRQVYAGNLGAPSTLERFIADGQK EFDLKKLQTDLKKVIRGKSKERDWCCQEELRSYVEQNTIQYDLWAWGEMFNKAHTALKIKSTRNYNFAKQRLQEFKRIQ SLNLLVVKLNDFFDSEFFSGEETY TICVHHLGGKRLSKLYKAWEDDPADPENAIIVLCDDLKNNFKKEPIRNILRYIFTI RQEC SAQDILAAAKYNQQLDRYKSQKANPSVLGNQGFTWRNAVILPEKAQRNDRPNSLDLRIWL YLKL RHPDGRWKKH HIPFYDTRFFQEY AAGNSPVDTCQFRTPRFGYHLPKLDQTAIRVNKKHVKA AKTEARIRLAIQGGTLPVSNLKITRISATI NSKGQVRIPVKFDVGRQKGTQIGDRFCGYDQNTASHAYSLWEVVKEGQYHKELGCFVRFISSGDIVSITENRGNQFDQ LSYEGLAYPQYADWRKKASKFVSLWQITKKNKKEIVTVEAKEKFD AICKYQPRLYKFNKEYAYLLRDIVRGKSLVELQ QIRQEIFRFEIQDCGVTRLGSLSLSTLETVKAVKGIYSYFSTALNASKNPNISDEQRKEFDPPEL FALLEKLELIRTRKKKQKV ERIANSLIQTCLENNIKFIRGEGDLSTTNNATKKAASRSMDWLAGVFNKIRQLAPMHNTLFGCGSLYTSHQDPLVHRN PDKAMKCRWAAIPVKDIGDWLRLKLSQNLRAKNIGTGEYHQGVKEFLSHYELQDLEELLKWRSDRKSNI PCWVLQN RLAEKLGKNEAVVYIPVRGGRIYFATHKVATGAVSIVFDQKQVWVCNADHVAAAANIALTVKGIGEQSSDEENPDGSRIKL QLTS</p>
<p>SEQ ID NO. 18 CasXX - R715A</p>	<p>MSSAIKSYKSVLRPNERKNQLLKSTIQCLEDGSAFFKMLQGLFGGITPEIVRFSTEQEKQQQDIALWCAVNWFRPVSQDS LHTIASDNLVEKFEEYYGGTASDAIKQYFSAIGESYYWDCRQYYDLCRELGVEVSDLTHDLEILCREKCLAVATESN QYNSISVLFVGTGRKEDRSVKLRITKKILEAISNLKEIPKNVAPIQEILNVAKATKETFRQVYAGNLGAPSTLERFIADGQK EFDLKKLQTDLKKVIRGKSKERDWCCQEELRSYVEQNTIQYDLWAWGEMFNKAHTALKIKSTRNYNFAKQRLQEFKRIQ SLNLLVVKLNDFFDSEFFSGEETY TICVHHLGGKRLSKLYKAWEDDPADPENAIIVLCDDLKNNFKKEPIRNILRYIFTI RQEC SAQDILAAAKYNQQLDRYKSQKANPSVLGNQGFTWRNAVILPEKAQRNDRPNSLDLRIWL YLKL RHPDGRWKKH HIPFYDTRFFQEY AAGNSPVDTCQFRTPRFGYHLPKLDQTAIRVNKKHVKA AKTEARIRLAIQGGTLPVSNLKITRISATI NSKGQVRIPVKFDVGRQKGTQIGDRFCGYDQNTASHAYSLWEVVKEGQYHKELGCFVRFISSGDIVSITENRGNQFDQ LSYEGLAYPQYADWRKKASKFVSLWQITKKNKKEIVTVEAKEKFD AICKYQPRLYKFNKEYAYLLRDIVRGKSLVELQ QIRQEIFRFEIQDCGVTRLGSLSLSTLETVKAVKGIYSYFSTALNASKNPNISDEQRKEFDPPEL FALLEKLELIRTRKKKQKV ERIANSLIQTCLENNIKFIRGEGDLSTTNNATKKAASRSMDWLAGVFNKIRQLAPMHNTLFGCGSLYTSHQDPLVHRN PDKAMKCRWAAIPVKDIGDWLRLKLSQNLRAKNIGTGEYHQGVKEFLSHYELQDLEELLKWRSDRKSNI PCWVLQN RLAEKLGKNEAVVYIPVRGGRIYFATHKVATGAVSIVFDQKQVWVCNADHVAAAANIALTVKGIGEQSSDEENPDGSRIKL QLTS</p>
<p>SEQ ID NO. 19 CasXX - R719A</p>	<p>MSSAIKSYKSVLRPNERKNQLLKSTIQCLEDGSAFFKMLQGLFGGITPEIVRFSTEQEKQQQDIALWCAVNWFRPVSQDS LHTIASDNLVEKFEEYYGGTASDAIKQYFSAIGESYYWDCRQYYDLCRELGVEVSDLTHDLEILCREKCLAVATESN QYNSISVLFVGTGRKEDRSVKLRITKKILEAISNLKEIPKNVAPIQEILNVAKATKETFRQVYAGNLGAPSTLERFIADGQK EFDLKKLQTDLKKVIRGKSKERDWCCQEELRSYVEQNTIQYDLWAWGEMFNKAHTALKIKSTRNYNFAKQRLQEFKRIQ SLNLLVVKLNDFFDSEFFSGEETY TICVHHLGGKRLSKLYKAWEDDPADPENAIIVLCDDLKNNFKKEPIRNILRYIFTI RQEC SAQDILAAAKYNQQLDRYKSQKANPSVLGNQGFTWRNAVILPEKAQRNDRPNSLDLRIWL YLKL RHPDGRWKKH HIPFYDTRFFQEY AAGNSPVDTCQFRTPRFGYHLPKLDQTAIRVNKKHVKA AKTEARIRLAIQGGTLPVSNLKITRISATI NSKGQVRIPVKFDVGRQKGTQIGDRFCGYDQNTASHAYSLWEVVKEGQYHKELGCFVRFISSGDIVSITENRGNQFDQ LSYEGLAYPQYADWRKKASKFVSLWQITKKNKKEIVTVEAKEKFD AICKYQPRLYKFNKEYAYLLRDIVRGKSLVELQ QIRQEIFRFEIQDCGVTRLGSLSLSTLETVKAVKGIYSYFSTALNASKNPNISDEQRKEFDPPEL FALLEKLELIRTRKKKQKV ERIANSLIQTCLENNIKFIRGEGDLSTTNNATKKAASRSMDWLAGVFNKIRQLAPMHNTLFGCGSLYTSHQDPLVHRN PDKAMKCRWAAIPVKDIGDWLRLKLSQNLRAKNIGTGEYHQGVKEFLSHYELQDLEELLKWRSDRKSNI PCWVLQN RLAEKLGKNEAVVYIPVRGGRIYFATHKVATGAVSIVFDQKQVWVCNADHVAAAANIALTVKGIGEQSSDEENPDGSRIKL QLTS</p>
<p>SEQ ID NO. 20 CasXX - K394A</p>	<p>MSSAIKSYKSVLRPNERKNQLLKSTIQCLEDGSAFFKMLQGLFGGITPEIVRFSTEQEKQQQDIALWCAVNWFRPVSQDS LHTIASDNLVEKFEEYYGGTASDAIKQYFSAIGESYYWDCRQYYDLCRELGVEVSDLTHDLEILCREKCLAVATESN QYNSISVLFVGTGRKEDRSVKLRITKKILEAISNLKEIPKNVAPIQEILNVAKATKETFRQVYAGNLGAPSTLERFIADGQK EFDLKKLQTDLKKVIRGKSKERDWCCQEELRSYVEQNTIQYDLWAWGEMFNKAHTALKIKSTRNYNFAKQRLQEFKRIQ SLNLLVVKLNDFFDSEFFSGEETY TICVHHLGGKRLSKLYKAWEDDPADPENAIIVLCDDLKNNFKKEPIRNILRYIFTI RQEC SAQDILAAAKYNQQLDRYKSQKANPSVLGNQGFTWRNAVILPEKAQRNDRPNSLDLRIWL YLKL RHPDGRWKKH HIPFYDTRFFQEY AAGNSPVDTCQFRTPRFGYHLPKLDQTAIRVNKKHVKA AKTEARIRLAIQGGTLPVSNLKITRISATI NSKGQVRIPVKFDVGRQKGTQIGDRFCGYDQNTASHAYSLWEVVKEGQYHKELGCFVRFISSGDIVSITENRGNQFDQ LSYEGLAYPQYADWRKKASKFVSLWQITKKNKKEIVTVEAKEKFD AICKYQPRLYKFNKEYAYLLRDIVRGKSLVELQ QIRQEIFRFEIQDCGVTRLGSLSLSTLETVKAVKGIYSYFSTALNASKNPNISDEQRKEFDPPEL FALLEKLELIRTRKKKQKV ERIANSLIQTCLENNIKFIRGEGDLSTTNNATKKAASRSMDWLAGVFNKIRQLAPMHNTLFGCGSLYTSHQDPLVHRN PDKAMKCRWAAIPVKDIGDWLRLKLSQNLRAKNIGTGEYHQGVKEFLSHYELQDLEELLKWRSDRKSNI PCWVLQN RLAEKLGKNEAVVYIPVRGGRIYFATHKVATGAVSIVFDQKQVWVCNADHVAAAANIALTVKGIGEQSSDEENPDGSRIKL QLTS</p>
<p>SEQ ID NO. 21 CasXX - H357A</p>	<p>MSSAIKSYKSVLRPNERKNQLLKSTIQCLEDGSAFFKMLQGLFGGITPEIVRFSTEQEKQQQDIALWCAVNWFRPVSQDS LHTIASDNLVEKFEEYYGGTASDAIKQYFSAIGESYYWDCRQYYDLCRELGVEVSDLTHDLEILCREKCLAVATESN QYNSISVLFVGTGRKEDRSVKLRITKKILEAISNLKEIPKNVAPIQEILNVAKATKETFRQVYAGNLGAPSTLERFIADGQK EFDLKKLQTDLKKVIRGKSKERDWCCQEELRSYVEQNTIQYDLWAWGEMFNKAHTALKIKSTRNYNFAKQRLQEFKRIQ SLNLLVVKLNDFFDSEFFSGEETY TICVHHLGGKRLSKLYKAWEDDPADPENAIIVLCDDLKNNFKKEPIRNILRYIFTI RQEC SAQDILAAAKYNQQLDRYKSQKANPSVLGNQGFTWRNAVILPEKAQRNDRPNSLDLRIWL YLKL RHPDGRWKKH HIPFYDTRFFQEY AAGNSPVDTCQFRTPRFGYHLPKLDQTAIRVNKKHVKA AKTEARIRLAIQGGTLPVSNLKITRISATI NSKGQVRIPVKFDVGRQKGTQIGDRFCGYDQNTASHAYSLWEVVKEGQYHKELGCFVRFISSGDIVSITENRGNQFDQ LSYEGLAYPQYADWRKKASKFVSLWQITKKNKKEIVTVEAKEKFD AICKYQPRLYKFNKEYAYLLRDIVRGKSLVELQ QIRQEIFRFEIQDCGVTRLGSLSLSTLETVKAVKGIYSYFSTALNASKNPNISDEQRKEFDPPEL FALLEKLELIRTRKKKQKV ERIANSLIQTCLENNIKFIRGEGDLSTTNNATKKAASRSMDWLAGVFNKIRQLAPMHNTLFGCGSLYTSHQDPLVHRN PDKAMKCRWAAIPVKDIGDWLRLKLSQNLRAKNIGTGEYHQGVKEFLSHYELQDLEELLKWRSDRKSNI PCWVLQN RLAEKLGKNEAVVYIPVRGGRIYFATHKVATGAVSIVFDQKQVWVCNADHVAAAANIALTVKGIGEQSSDEENPDGSRIKL QLTS</p>
<p>SEQ ID NO. 22 CasXX - K844A</p>	<p>MSSAIKSYKSVLRPNERKNQLLKSTIQCLEDGSAFFKMLQGLFGGITPEIVRFSTEQEKQQQDIALWCAVNWFRPVSQDS LHTIASDNLVEKFEEYYGGTASDAIKQYFSAIGESYYWDCRQYYDLCRELGVEVSDLTHDLEILCREKCLAVATESN QYNSISVLFVGTGRKEDRSVKLRITKKILEAISNLKEIPKNVAPIQEILNVAKATKETFRQVYAGNLGAPSTLERFIADGQK EFDLKKLQTDLKKVIRGKSKERDWCCQEELRSYVEQNTIQYDLWAWGEMFNKAHTALKIKSTRNYNFAKQRLQEFKRIQ SLNLLVVKLNDFFDSEFFSGEETY TICVHHLGGKRLSKLYKAWEDDPADPENAIIVLCDDLKNNFKKEPIRNILRYIFTI RQEC SAQDILAAAKYNQQLDRYKSQKANPSVLGNQGFTWRNAVILPEKAQRNDRPNSLDLRIWL YLKL RHPDGRWKKH</p>

	<p>HIPFYDTRFFQEIYAAGNSPVDTCQFRTPRFGYHLPKLTQDTAIRVNKKHVKAakteARIRLAIQQGTLFVSNLKITRISATI NSKGQVRIPVKFDVGRQKGTQIGDRFCGYDQNQTASHAYSLEVVKEGQYHKELGCFVRFISSGDIVSITENRGNQFDQ LSYEGLAYPQYADWRKKASKFVSLWQITKKNKKKEIVTVEAKEKFDaICKYQPRLYKFNKEYAYLLRDIVRGKSLVELQ QIRQEIFRFIEQDCGVTRLGSLSLSTLETVKAVKGIISYFSTALNASKNNPISDEQRKEFDPELFALLEKLELIRTRKKKQKV ERIANSLIQTCLENNIKFIRGEGDLSTNNATAKKANSRSMDWLAGVFNKIRQLAPMHNITLFGCGSLYTSHQDPLVHRN PDKAMKCRWAAIPVKDIGDWVLRKLSQNLRAKNIGTGEYYHQGVKEFLSHYELQDLEEEELKWRSDRKSNIPCWVLQN RLAEKLGKNEAVVYIPVRGGRIYFATHKVATGAVSIVFDQKQVWVCNADHVAAAANIALTVKGIGEQSSDEENPDGSRIKL QLTS</p>
<p>SEQ ID NO. 23 CasXX - R719A + K844A</p>	<p>MSSAIKSYKSVLRPNERKNQLLKSTIQCLEDGSAFFKMLQGLFGGITPEIVRFSTEQEKQQQDIALWCAVNWFRPVSQDS LTHTIASDNLVEKFEEYYGGTASDAIKQYFSASIGESYYWDCRQYYDLCRELGVEVSDLTHDLEILCREKCLAVATESN QYNSIISVLFGTGRKEDRSVKLRITKKILEAISNLKEIPKNVAPIQEILNVAKATKETFRQVYAGNLGAPSTLERFIAKDGQK EFDLKKLQTDLKKVIRGKSKERDWCCQEELRSYVEQNTIQYDLWAWGEMFNKAHTALKIKSTRNYNFAKQRLQEQFKRIQ SLNLLVVKLNDFFDSEFFSGEETY TICVHHLGGKRLSKLYKAWEDDPADPENAIIVL CDDLKNNFKKEPIRNILRYIFTI RQEC SAQDILAAAKYNQQLDRYKSQKANPSVLGNQGGFTWRNAVILPEKAQRNDRPNSDLRIWL YLKL RHPDGRWKKH HIPFYDTRFFQEIYAAGNSPVDTCQFRTPRFGYHLPKLTQDTAIRVNKKHVKAakteARIRLAIQQGTLFVSNLKITRISATI NSKGQVRIPVKFDVGRQKGTQIGDRFCGYDQNQTASHAYSLEVVKEGQYHKELGCFVRFISSGDIVSITENRGNQFDQ LSYEGLAYPQYADWRKKASKFVSLWQITKKNKKKEIVTVEAKEKFDaICKYQPRLYKFNKEYAYLLRDIVRGKSLVELQ QIRQEIFRFIEQDCGVTRLGSLSLSTLETVKAVKGIISYFSTALNASKNNPISDEQRKEFDPELFALLEKLELIRTRKKKQKV ERIANSLIQTCLENNIKFIRGEGDLSTNNATAKKANSRSMDWLAGVFNKIRQLAPMHNITLFGCGSLYTSHQDPLVHRN PDKAMKCRWAAIPVKDIGDWVLRKLSQNLRAKNIGTGEYYHQGVKEFLSHYELQDLEEEELKWRSDRKSNIPCWVLQN RLAEKLGKNEAVVYIPVRGGRIYFATHKVATGAVSIVFDQKQVWVCNADHVAAAANIALTVKGIGEQSSDEENPDGSRIKL QLTS</p>
<p>SEQ ID NO: 24 CasXX - R857A + K844A</p>	<p>MSSAIKSYKSVLRPNERKNQLLKSTIQCLEDGSAFFKMLQGLFGGITPEIVRFSTEQEKQQQDIALWCAVNWFRPVSQDS LTHTIASDNLVEKFEEYYGGTASDAIKQYFSASIGESYYWDCRQYYDLCRELGVEVSDLTHDLEILCREKCLAVATESN QYNSIISVLFGTGRKEDRSVKLRITKKILEAISNLKEIPKNVAPIQEILNVAKATKETFRQVYAGNLGAPSTLERFIAKDGQK EFDLKKLQTDLKKVIRGKSKERDWCCQEELRSYVEQNTIQYDLWAWGEMFNKAHTALKIKSTRNYNFAKQRLQEQFKRIQ SLNLLVVKLNDFFDSEFFSGEETY TICVHHLGGKRLSKLYKAWEDDPADPENAIIVL CDDLKNNFKKEPIRNILRYIFTI RQEC SAQDILAAAKYNQQLDRYKSQKANPSVLGNQGGFTWRNAVILPEKAQRNDRPNSDLRIWL YLKL RHPDGRWKKH HIPFYDTRFFQEIYAAGNSPVDTCQFRTPRFGYHLPKLTQDTAIRVNKKHVKAakteARIRLAIQQGTLFVSNLKITRISATI NSKGQVRIPVKFDVGRQKGTQIGDRFCGYDQNQTASHAYSLEVVKEGQYHKELGCFVRFISSGDIVSITENRGNQFDQ LSYEGLAYPQYADWRKKASKFVSLWQITKKNKKKEIVTVEAKEKFDaICKYQPRLYKFNKEYAYLLRDIVRGKSLVELQ QIRQEIFRFIEQDCGVTRLGSLSLSTLETVKAVKGIISYFSTALNASKNNPISDEQRKEFDPELFALLEKLELIRTRKKKQKV ERIANSLIQTCLENNIKFIRGEGDLSTNNATAKKANSRSMDWLAGVFNKIRQLAPMHNITLFGCGSLYTSHQDPLVHRN PDKAMKCRWAAIPVKDIGDWVLRKLSQNLRAKNIGTGEYYHQGVKEFLSHYELQDLEEEELKWRSDRKSNIPCWVLQN RLAEKLGKNEAVVYIPVRGGRIYFATHKVATGAVSIVFDQKQVWVCNADHVAAAANIALTVKGIGEQSSDEENPDGSRIKL QLTS</p>

权利要求书

1.一种工程化的 Cas12i 核酸酶；其包含一种或多种基于参比 Cas12i 核酸酶的突变，所述突变选自：

5 (1)将参比 Cas12i 核酸酶中与 PAM 相互作用的一个或多个氨基酸替换为带正电的氨基酸；

(2)将参比 Cas12i 核酸酶中参与打开 DNA 双链的一个或多个氨基酸替换为带芳香环的氨基酸

10 (3)将参比 Cas12i 核酸酶中位于 RuvC 结构域并与单链 DNA 底物相互作用的一个或多个氨基酸替换为带正电的氨基酸

(4)将参比 Cas12i 核酸酶中与 DNA-RNA 双螺旋相互作用的一个或多个氨基酸替换为带正电的氨基酸；和

(5)将参比 Cas12i 核酸酶中与 DNA-RNA 双螺旋相互作用的一个或多个极性或带正电荷的氨基酸替换为疏水氨基酸；

15 优选地，所述参比 Cas12i 核酸酶为氨基酸序列为 SEQ ID NO: 1 的野生型 Cas12i2 核酸酶。

2.如权利要求 1 所述的工程化的 Cas12i 核酸酶，其中，所述与 PAM 相互作用的一个或多个氨基酸是与 PAM 在三维结构上距离在 9 埃以内的氨基酸，优选，所述与 PAM 相互作用的一个或多个氨基酸是下述位置的一个或多个氨基酸：176、178、226、227、229、
20 237、238、264、447 和 563，

进一步优选，所述与 PAM 相互作用的一个或多个氨基酸是下述一个或多个氨基酸：E176、E178、Y226、A227、N229、E237、K238、K264、T447 和 E563，

进一步优选，所述与 PAM 相互作用的一个或多个氨基酸是下述一个或多个氨基酸：E176、K238、T447 和 E563，

25 其中，所述氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

3.如权利要求 1 或 2 所述的工程化的 Cas12i 核酸酶，其中，所述带正电的氨基酸是 R，K 或 H。

4.如权利要求 1-3 中任一项所述的工程化的 Cas12i 核酸酶，其中，所述将参比 Cas12i 核酸酶中与 PAM 相互作用的一个或多个氨基酸替换为带正电的氨基酸是指如下替换中的一种或多种：E176R、K238R、T447R 和 E563R；
30

优选，所述工程化的 Cas12i 核酸酶包含下述任何一种突变或突变组合：(1)E563R；(2)E176R、T447R、E176R 和 E563R；(3)K238R 和 E563R；(4)E176R、K238R 和 T447R；(5)E176R、K238R 和 E563R；(6)E176R、T447R 和 E563R；和(7)E176R、K238R、T447R 和 E563R；

5 其中，所述氨基酸位置如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

5. 如权利要求 1~4 中任一项所述的工程化的 Cas12i 核酸酶，其中，所述参与打开 DNA 双链的一个或多个氨基酸是与 PAM 中相对于靶标链的 3'端最后一个碱基对相互作用的氨基酸；

10 优选，所述参与打开 DNA 双链的一个或多个氨基酸是下述位置的一个或多个氨基酸：163 和 164；

进一步优选，所述参与打开 DNA 双链的一个或多个氨基酸是下述一个或多个氨基酸：Q163 和 N164；进一步优选，所述参与打开 DNA 双链的氨基酸为 N164；

其中，所述氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

15 6. 如权利要求 1-5 中任一项所述的工程化的 Cas12i 核酸酶，其中，所述参与打开 DNA 双链的一个或多个氨基酸替换为带芳香环的氨基酸，所述带芳香环的氨基酸是 F、Y 或 W，优选，所述带芳香环的氨基酸为 F 或 Y。

7. 如权利要求 1-6 中任一项所述的工程化的 Cas12i 核酸酶，其中，所述将参比 Cas12i 核酸酶中参与打开 DNA 双链的一个或多个氨基酸的替换为带芳香环的氨基酸包括以下一个或多个替换：Q163F、Q163Y、Q163W、和 N164F；优选，所述 Cas12i 核酸酶
20 包含 N164Y 或 N164F 突变；进一步优选，所述工程化的 Cas12i 核酸酶包含 N164Y。

8. 如权利要求 1~7 中任一项所述的工程化的 Cas12i 核酸酶，其中，所述位于 RuvC 结构域并与单链 DNA 底物相互作用的一个或多个氨基酸是在三维结构上与单链 DNA 底物距离在 9 埃以内的氨基酸；

25 优选，所述位于 RuvC 结构域并与单链 DNA 底物相互作用的一个或多个氨基酸是下述位置的一个或多个氨基酸：323、327、355、359、360、361、362、388、390、391、392、393、414、417、418、421、424、425、650、652、653、696、705、708、709、751、752、755、840、848、851、856、885、897、925、926、928、929、932、1022；

30 进一步优选，所述位于 RuvC 结构域并与单链 DNA 底物相互作用的一个或多个氨基酸是下述一个或多个氨基酸：E323、L327、V355、G359、G360、K361、D362、L388、N390、N391、F392、K393、Q414、L417、L418、K421、Q424、Q425、S650、E652、

G653、I696、K705、K708、E709、L751、S752、E755、N840、N848、S851、A856、Q885、M897、N925、I926、T928、G929、Y932、A1022；

进一步优选，所述位于 RuvC 结构域并与单链 DNA 底物相互作用的一个或多个氨基酸是是下述一个或多个氨基酸：E323、D362、Q425、N925、I926、N391、Q424 和 G929；

5 其中，所述氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

9.如权利要求 1-8 中任一项所述的工程化的 Cas12i 核酸酶，其中，将参比 Cas12i 核酸酶中位于 RuvC 结构域并与单链 DNA 底物相互作用的一个或多个氨基酸替换为带正电的氨基酸包括替换为 R 或 K，优选，所述带正电的该氨基酸为 R。

10.如权利要求 1-9 中任一项所述的工程化的 Cas12i 核酸酶，其中，将参比 Cas12i 核酸酶位于 RuvC 结构域并与单链 DNA 底物相互作用的一个或多个氨基酸替换为带正电的氨基酸是指包括如下替换中的一种或多种：E323R、D362R、N391R、Q424R、Q425R、N925R、I926R 和 G929R；

15 优选，所述工程化的 Cas12i 核酸酶包含下述任何一种突变或突变组合：(1)E323R；(2)D362R；(3)Q425R；(4)N925R；(5)I926R；(6)E323R 和 D362R；(7)E323R 和 Q425R；(8)E323R 和 I926R；(9)Q425R 和 I926R；(10)D362R 和 I926R；(11)N925R 和 I926R；(12)E323R、D362R 和 Q425R；(13)E323R、D362R 和 I926R；(14)E323R、Q425R 和 I926R；(15)D362R、N925R 和 I926R；和(16) E323R、D362R、Q425R 和 I926R；

其中，所述氨基酸位置如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

20 11.如权利要求 1~10 中任一项所述的工程化的 Cas12i 核酸酶，其中，所述与 DNA-RNA 双螺旋相互作用的一个或多个氨基酸是与 DNA-RNA 双螺旋在三维结构上距离在 9 埃以内的氨基酸；

25 优选，所述与 DNA-RNA 双螺旋相互作用的一个或多个氨基酸是下述位置的一个或多个氨基酸：116、117、156、159、160、161、247、293、294、297、301、305、306、308、312、313、316、319、320、343、348、349、427、433、438、441、442、679、683、691、782、783、797、800、852、853、855、861、865、957、958；

30 进一步优选，所述与 DNA-RNA 双螺旋相互作用的一个或多个氨基酸是下述一个或多个氨基酸：G116、E117、A156、T159、E160、S161、E247、G293、E294、N297、T301、I305、K306、T308、N312、F313、Q316、E319、Q320、E343、E348、E349、D427、K433、V438、N441、Q442、N679、E683、E691、D782、E783、E797、E800、M852、D853、L855、N861、Q865、S957、D958；

进一步优选，所述与 DNA-RNA 双螺旋相互作用的一个或多个氨基酸是下述一个或多个氨基酸：G116、E117、T159、S161、E319、E343 和 D958；其中，所述氨基酸位置如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

5 12. 如权利要求 1-11 中任一项所述的工程化的 Cas12i 核酸酶，其中，将参比 Cas12i 核酸酶中与 DNA-RNA 双螺旋相互作用的一个或多个氨基酸替换为带正电的氨基酸包括替换为 R 或 K，优选，所述带正电的氨基酸为 R。

13. 如权利要求 1-12 中任一项所述的工程化的 Cas12i 核酸酶，其中，将参比 Cas12i 核酸酶中与 DNA-RNA 双螺旋相互作用的一个或多个氨基酸替换为带正电的氨基酸包括替换为以下的一种或多种：G116R、E117R、T159R、S161R、E319R、E343R 和 D958R；
10 优选，所述工程化的 Cas12i 核酸酶包含 D958R 替换；

其中，所述氨基酸位置如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

14. 如权利要求 1~13 中任一项所述的工程化的 Cas12i 核酸酶，所述与 DNA-RNA 双螺旋相互作用的一个或多个极性或带正电荷的氨基酸选自如下的一个或多个位置的氨基酸：357、394、715、719、807、844、848、857、861，即下述一个或多个氨基酸：H357、
15 K394、R715、R719、K807、K844、N848、R857、R861；

其中，所述氨基酸位置如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

15. 如权利要求 1-14 中任一项所述的工程化的 Cas12i 核酸酶，其中，将参比 Cas12i 核酸酶中与 DNA-RNA 双螺旋相互作用的一个或多个极性或带正电荷的氨基酸替换为疏水氨基酸包括替换为丙氨酸(A)。

20 16. 如权利要求 1-15 中任一项所述的工程化的 Cas12i 核酸酶，其包含选自下组的一个或多个突变：H357A、K394A、R715A、R719A、K807A、K844A、N848A、R857A、R861A；

优选，包含选自下组的一个或多个突变：K394A、R719A、K844A、R857A；

其中，所述氨基酸位置如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

25 17. 如权利要求 1-16 中任一项所述的工程化的 Cas12i 核酸酶，其包含 R719A 和 K844A 氨基酸取代，或包含 R857A 和 K844A 氨基酸取代；其中，所述氨基酸位置如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

18. 如权利要求 1~17 中任一项所述的工程化的 Cas12i 核酸酶，其还包含一个或多个柔性区突变，所述突变增加了参比 Cas12i 核酸酶中的柔性区的柔性，所述柔性区选自氨基酸残基 439-443 或氨基酸残基 925-929；
30

优选, 所述柔性区突变位于 439 和/或 926 位点;

进一步优选, 所述柔性区突变是 L439 和/或 I926 的突变;

其中, 所述氨基酸位置如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

19. 如权利要求 18 所述的工程化的 Cas12i 核酸酶, 其中, 所述一个或多个柔性区突变为: 将该柔性区氨基酸替换为 G、和/或在其后插入一个或两个 G;

优选, 所述一个或多个柔性区突变包含 I926G、L439(L+G)或 L439(L+GG);

进一步优选, 所述一个或多个柔性区突变包含 L439(L+G)或 L439(L+GG);

其中, 所述氨基酸位置编号如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

20. 一种工程化的 Cas12i 核酸酶;

10 所述工程化的 Cas12i 核酸酶包含任何以下一组或多组突变:

(1)E563R; (2)E176R 和 T447R; (3)E176R 和 E563R; (4)K238R 和 E563R; (5) E176R、K238R 和 T447R; (6)E176R、 T447R 和 E563R; (7) E176R、 K238R 和 E563R; (8) E176R、K238R、 T447R 和 E563R; (9) N164Y; (10) N164F; (11) E323R; (12) D362R; (13) Q425R; (14) N925R; (15) I926R; (16) D958R; (17) E323R 和 D362R; (18) E323R 和 Q425R; (19) E323R 和 I926R; (20) Q425R 和 I926R; (21) D362R 和 I926R; (22) N925R 和 I926R; (23) E323R、 D362R 和 Q425R; (24) E323R、 D362R 和 I926R; (25) E323R、 Q425R 和 I926R; (26) D362R、 N925R 和 I926R; (27) E323R、 D362R、 Q425R 和 I926R; (28) D362R 和 I926G; (29)N925R 和 I926G; (30)D362R、 N925R 和 I926G; (31)I926R 和 L439(L+G); (32)I926R 和 L439(L+GG); (33)E323R、 D362R 和 I926G; (34)R719A 和 K844A; 和
20 (35)R857A 和 K844A;

优选, 所述工程化的 Cas12i 核酸酶包含任何以下一组或多组突变: (1)E176R、K238R、 T447R 和 E563R; (2)N164Y; (3)I926R; (4) E323R 和 D362R; (4) I926G; (5)I926R 和 L439(L+G); (6)I926R 和 L439(L+GG); 和(7)D958R;

其中, 所述氨基酸位置如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

21. 一种工程化的 Cas12i 核酸酶, 所述工程化的 Cas12i 核酸酶包含以下任一组突变:
25 (1)E176R、K238R、T447R、E563R 和 N164Y; (2)E176R、K238R、T447R、E563R 和 I926R; (3)N164Y、E323R 和 D362R; (4)E176R、K238R、T447R、E563R、E323R 和 D362R; (5)N164Y 和 I926R; (6)E176R、K238R、T447R、E563R、N164Y 和 I926R; (7)E176R、K238R、T447R、E563R、N164Y、E323R 和 D362R; (8)E176R、K238R、
30 T447R、E563R、N164Y、I926R、E323R 和 D362R; (9)E176R、K238R、T447R、E563R、

N164Y、E323R、D362R 和 I926G；(10)E176R、K238R、T447R、E563R、N164Y、E323R、D362R、I926G 和 L439(L+GG)；(11)E176R、K238R、T447R、E563R、N164Y、E323R、D362R、I926G 和 L439(L+G)；(12)E176R、K238R、T447R、E563R、N164Y 和 D958R；(13)E176R、K238R、T447R、E563R、I926R 和 D958R；(14)E176R、K238R、T447R、E563R、E323R、D362R 和 D958R；(15)N164Y、I926R 和 D958R；(16)N164Y、E323R、D362R 和 D958R；(17) E176R、K238R、T447R、E563R、N164Y、I926R 和 D958R；(18)E176R、K238R、T447R、E563R、N164Y、E323R、D362R 和 D958R；(19)E176R、K238R、T447R、E563R、N164Y、I926R、E323R、D362R 和 D958R；(20)E176R、K238R、T447R、E563R、N164Y、E323R、D362R、I926G 和 D958R；(21)E176R、K238R、T447R、E563R、N164Y、E323R、D362R、I926G、L439(L+GG)和 D958R；(22)E176R、K238R、T447R、E563R、N164Y、E323R、D362R、I926G、L439(L+G)和 D958R；(23) E176R、K238R、T447R、E563R、N164Y、E323R、D362R 和 R857A；(24)E176R、K238R、T447R、E563R、N164Y、E323R、D362R 和 N861A；(25)E176R、K238R、T447R、E563R、N164Y、E323R、D362R 和 K807A；(26)E176R、K238R、T447R、E563R、N164Y、E323R、D362R 和 N848A；(27)E176R、K238R、T447R、E563R、N164Y、E323R、D362R 和 R715A；(28)E176R、K238R、T447R、E563R、N164Y、E323R、D362R 和 R719A；(29) E176R、K238R、T447R、E563R、N164Y、E323R、D362R 和 K394A；(30)E176R、K238R、T447R、E563R、N164Y、E323R、D362R 和 H357A；(31)E176R、K238R、T447R、E563R、N164Y、E323R、D362R 和 K844A；(32)E176R、K238R、T447R、E563R、N164Y、E323R、D362R、R719A 和 K844A；或(33)E176R、K238R、T447R、E563R、N164Y、E323R、D362R、R857A 和 K844A；其中，所述氨基酸位置如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

22. 如权利要求 1~21 中任一项所述的工程化的 Cas12i 核酸酶，其包含如 SEQ ID NOs: 2~24 中任一项所示氨基酸序列的工程化 Cas12i 核酸酶，或与如 SEQ ID NOs: 2~24 中任一序列所示的氨基酸序列具有至少 80%同一性的氨基酸序列。

23. 一种工程化的 Cas12i 效应蛋白，其包含权利要求 1~22 中任一项所述的工程化的 Cas12i 核酸酶或其功能衍生物；

任选的，所述工程化的 Cas12i 核酸酶或其功能衍生物具有酶活性，或者所述工程化的 Cas12i 核酸酶或其功能衍生物为酶失活突变体。

24. 如权利要求 23 所述的工程化的 Cas12i 效应蛋白，其中所述 Cas12i 效应蛋白能够

诱导 DNA 分子中的双链断裂或单链断裂。

25. 如权利要求 23 所述的工程化的 Cas12i 效应蛋白，其中所述工程化的 Cas12i 核酸酶或其功能衍生物为包含以下一种或多种突变的酶失活突变体：D599A、E833A、S883A、H884A、R900A 和 D1019A；其中，所述氨基酸位置如 SEQ ID NO: 1 所示的相应氨基酸位置所定义。

26. 如权利要求 23~25 中任一项所述的工程化的 Cas12i 效应蛋白，其还包含与所述工程化的 Cas12i 核酸酶或其功能衍生物融合的功能结构域。

27. 如权利要求 26 所述的工程化的 Cas12i 效应蛋白，其中所述功能结构域选自下组中的一个或多个：翻译起始结构域、转录阻遏结构域、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、报告分子结构域和核酸酶结构域。

28. 如权利要求 23-27 中任一项所述的工程化的 Cas12i 效应蛋白，所述工程化的 Cas12i 效应蛋白包含：含有所述工程化的 Cas12i 核酸酶或其功能衍生物的 N 末端部分的第一多肽和含有所述工程化的 Cas12i 核酸酶或其功能衍生物的 C 末端部分的第二多肽，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 的存在下彼此缔合，以形成与靶核酸特异性结合的成簇规律间隔短回文重复序列(CRISPR)复合物，所述靶核酸包含与所述指导序列互补的靶序列；

优选，所述第一多肽包含权利要求 1-22 中任一项所述的工程化的 Cas12i 核酸酶的 N 末端部分氨基酸残基 1 至 X，所述第二多肽包含权利要求 1-22 中任一项所述的工程化的 Cas12i 核酸酶的氨基酸残基 X+1 至所述 Cas12i 核酸酶的 C 末端；

可选的，所述第一多肽和所述第二多肽各自包含二聚化结构域；

可选的，所述第一多肽的二聚化结构域和所述第二多肽的二聚化结构域在诱导剂存在下彼此缔合。

29. 一种工程化的 CRISPR-Cas12i 系统，包括：

(a) 权利要求 1-28 中任一项所述的工程化的 Cas12i 核酸酶、或权利要求 23-28 中任一项所述的工程化的 Cas12i 效应蛋白；以及

(b) 包含与靶序列互补的指导序列的指导 RNA，或编码所述指导 RNA 的一种或多种核酸，

其中所述工程化的 Cas12i 核酸酶或工程化的 Cas12i 效应蛋白和所述指导 RNA 能够形成 CRISPR 复合物，所述 CRISPR 复合物特异性结合包含所述靶序列的靶核酸并诱导所述靶核酸的修饰；

优选，所述指导 RNA 是包含所述指导序列的 crRNA；

进一步优选，所述工程化的 CRISPR-Cas12i 系统包括编码多个 crRNA 的前体指导 RNA 阵列(array)；

5 或者，优选所述工程化的 Cas12i 核酸酶或工程化的 Cas12i 效应蛋白是主编辑器，所述指导 RNA 是引导编辑指导 RNA (pegRNA)。

30.如权利要求 29 所述的工程化的 CRISPR-Cas12i 系统，其包含一种或多种编码所述工程化的 Cas12i 核酸酶或工程化的 Cas12i 效应蛋白的载体；

优选，所述一种或多种载体选自下组：逆转录病毒载体、慢病毒载体、腺病毒载体、腺相关的载体和单纯疱疹载体；

10 进一步优选，所述一种或多种载体是腺相关病毒 AAV 载体；

进一步优选，所述 AAV 载体还编码所述指导 RNA。

31.一种检测样品中靶核酸的方法，包括：

(a)使样品与权利要求 29 中的工程化的 CRISPR-Cas12i 系统以及加标签的检测核酸接触，该检测核酸为单链且不与所述指导 RNA 的指导序列杂交；以及

15 (b)测量通过所述工程化的 Cas12i 核酸酶或工程化的 Cas12i 效应蛋白切割所述加标签的检测核酸而产生的可检测信号，从而检测所述靶核酸。

32.一种修饰包含靶序列的靶核酸的方法，包括使所述靶核酸与权利要求 29 或 30 所述的工程化的 CRISPR-Cas12i 系统接触；

优选，所述方法在体外进行、离体进行或在体内进行；

20 进一步优选，所述靶核酸存在于细胞中；

进一步优选，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；

进一步优选，所述靶核酸是基因组 DNA；

进一步优选，所述靶序列与疾病或病症相关；

25 进一步优选，所述工程化的 CRISPR-Cas12i 系统包括编码多个 crRNA 的前体指导 RNA 阵列，其中每个 crRNA 包含不同的指导序列。

33.如权利要求 29 或 30 所述的工程化的 CRISPR-Cas12i 系统在制备治疗与个体的细胞中靶核酸相关的疾病或病症的药物中的用途；优选，所述疾病或病症选自下组：癌症、心血管疾病、遗传性疾病、自身免疫疾病、代谢性疾病、神经退行性疾病、眼病、细菌感染和病毒感染。

30 34.一种治疗与个体的细胞中的靶核酸相关的疾病或病症的方法，所述方法包含使用

权利要求 32 所述的方法来修饰所述个体的细胞中的靶核酸，从而治疗所述疾病或病症；
优选，所述疾病或病症选自下组：癌症、心血管疾病、遗传性疾病、自身免疫疾病、代谢
性疾病、神经退行性疾病、眼病、细菌感染和病毒感染。

5 35. 一种修饰包含靶序列的靶核酸的方法，包括使所述靶核酸与权利要求 29 或 30 中
所述的工程化的 CRISPR-Cas12i 系统接触。

36. 包含如权利要求 1-22 中任一项所述的工程化的 Cas12i 核酸酶、如权利要求 23-28
中任一项所述的工程化的 Cas12i 效应蛋白的组合物或试剂盒。

37. 工程化的细胞，其包含经修饰的靶核酸，其中靶核酸由如权利要求 32 或 35 所述
的方法修饰。

10 38. 工程化的非人类动物，其包含一种或多种如权利要求 37 所述的工程化的细胞。

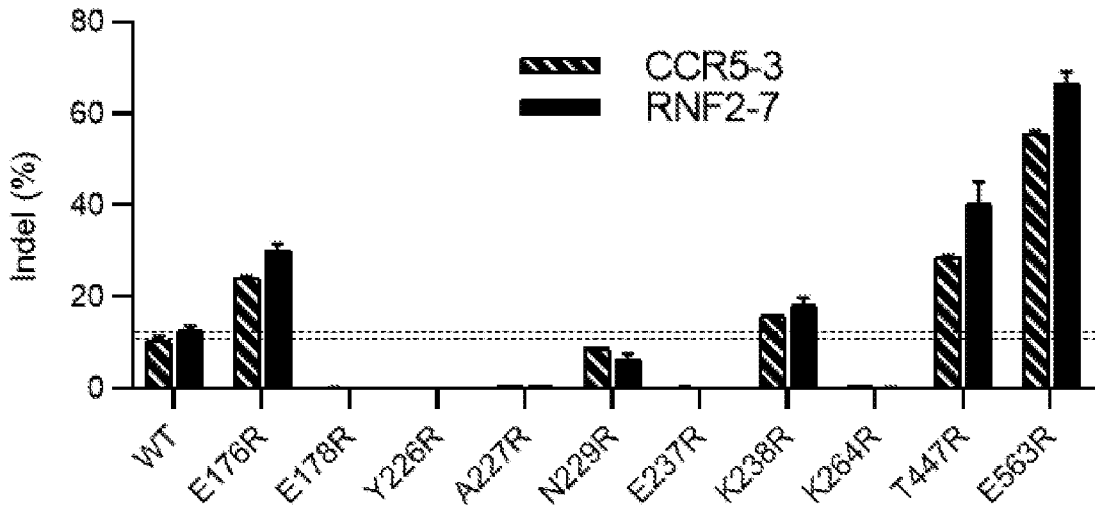


图 1a

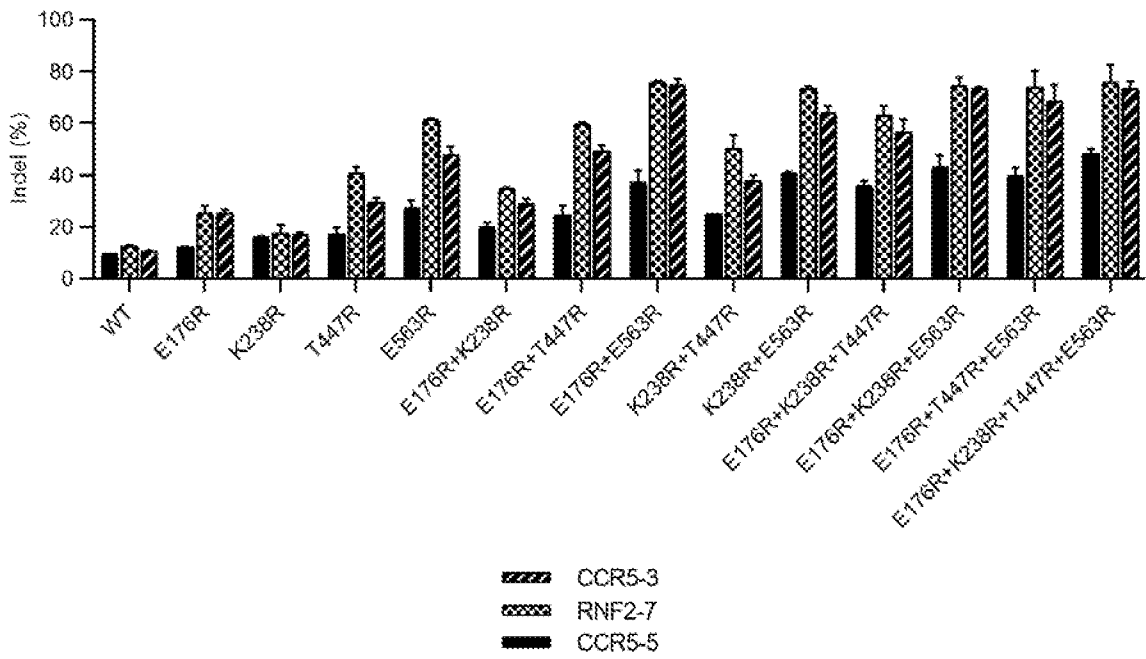


图 1b

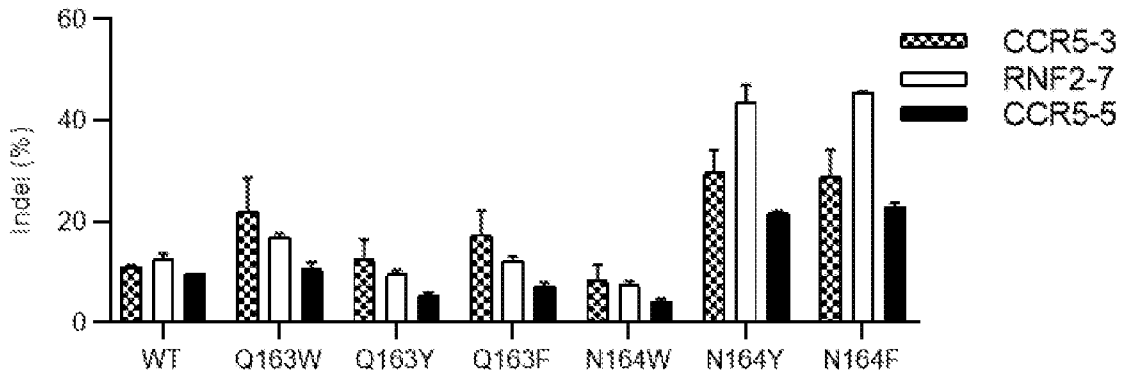


图 2

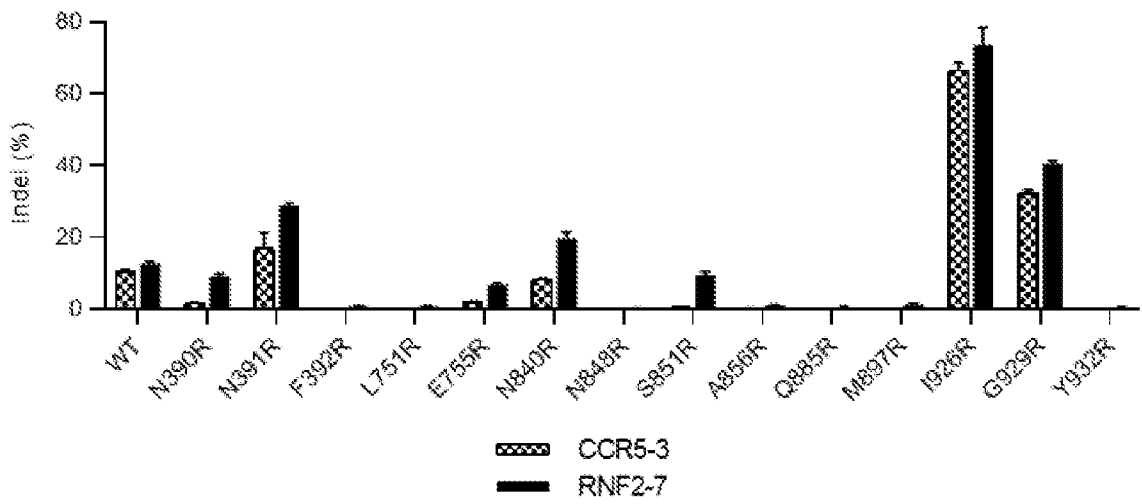
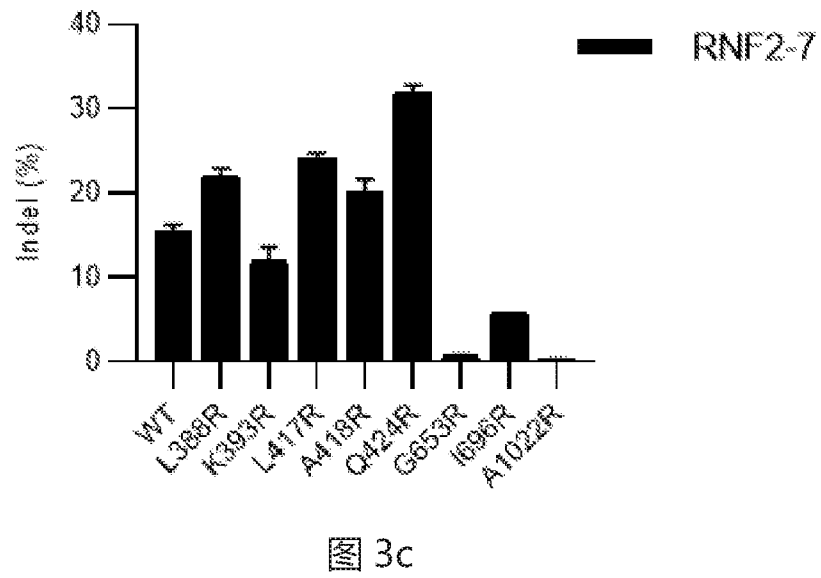
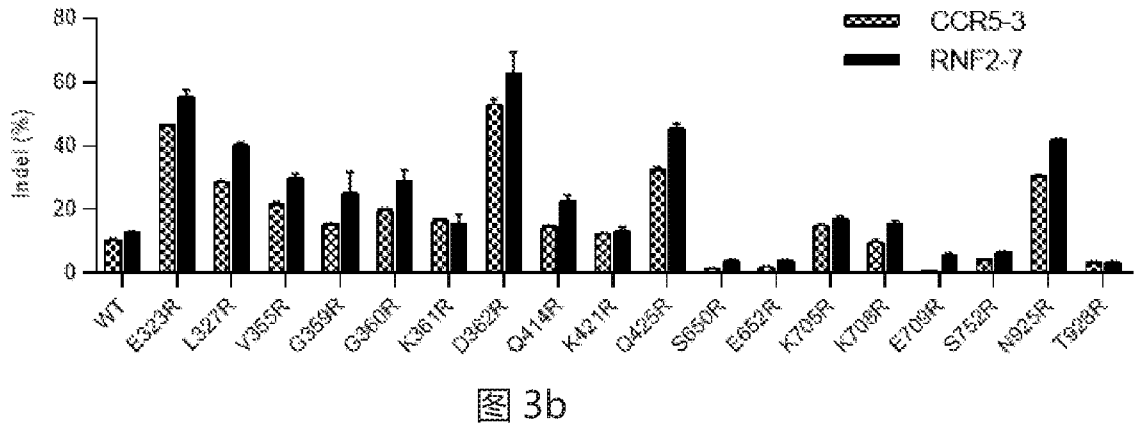
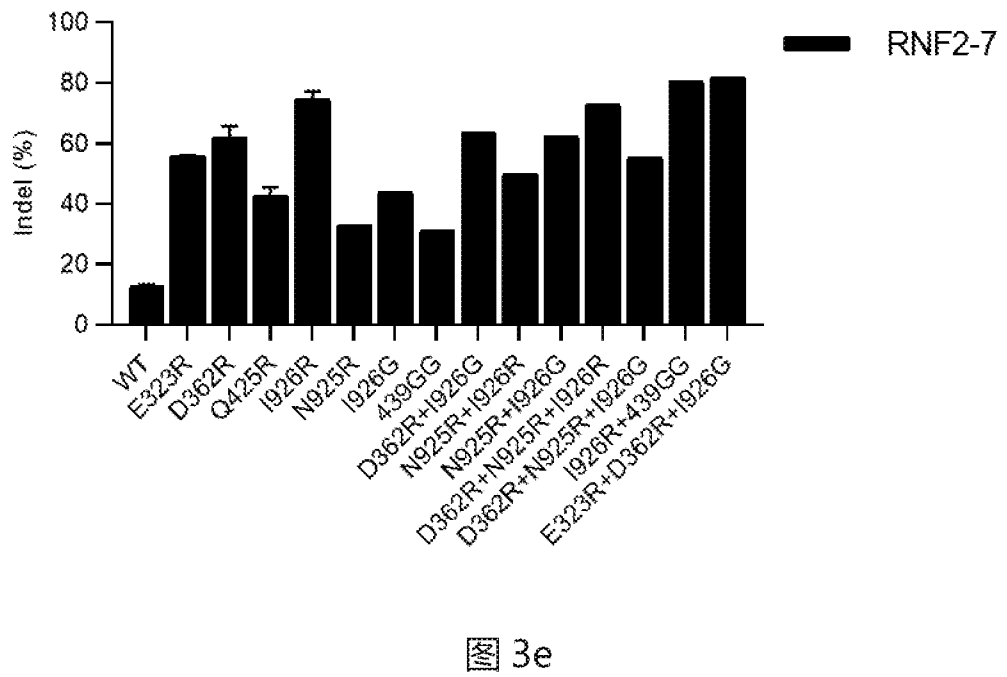
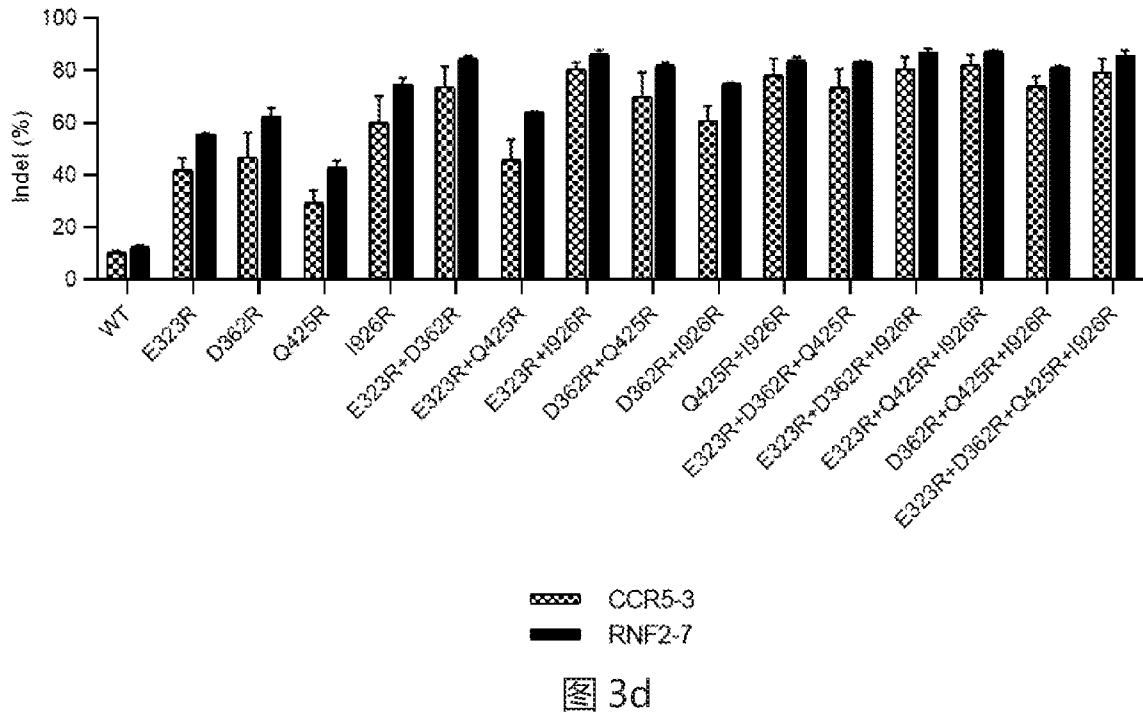
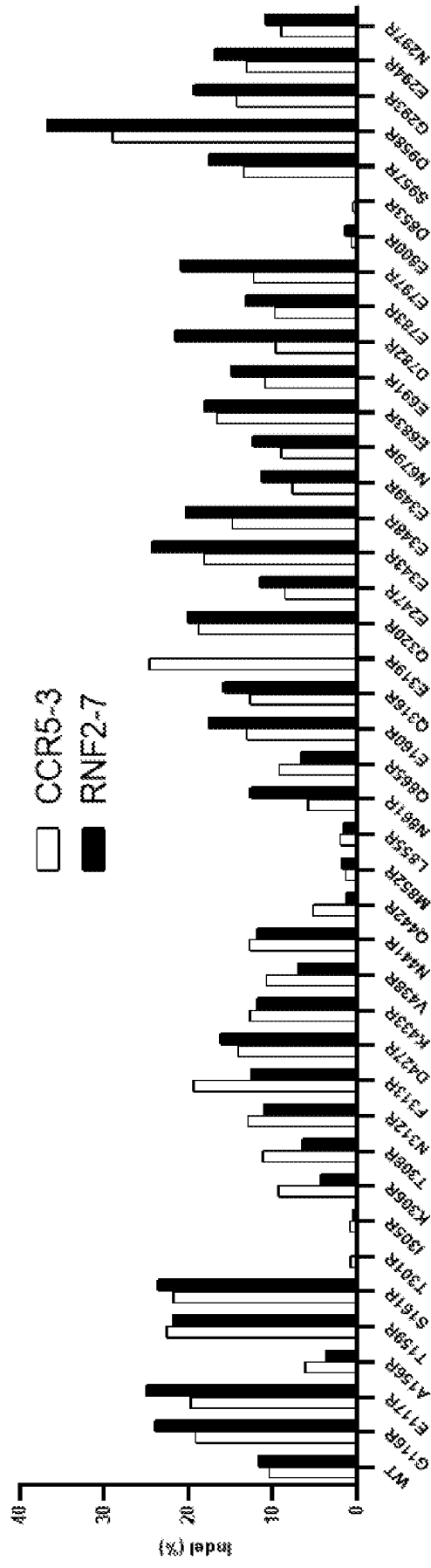


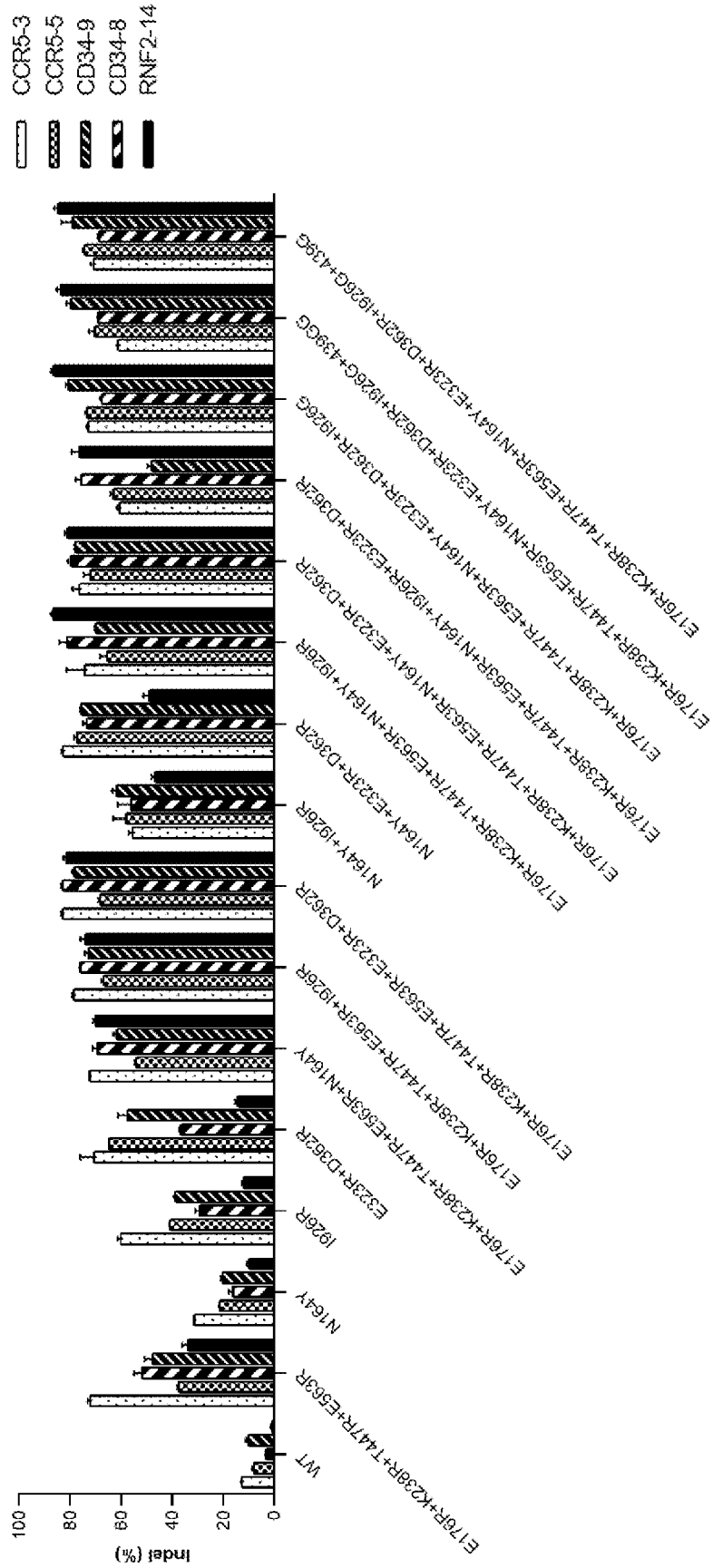
图 3a







4



5

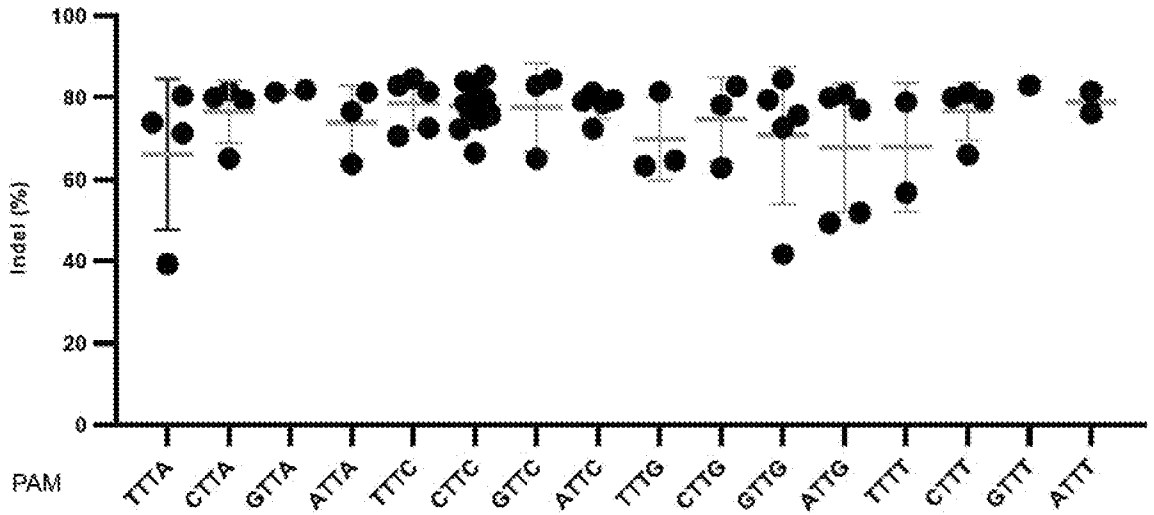
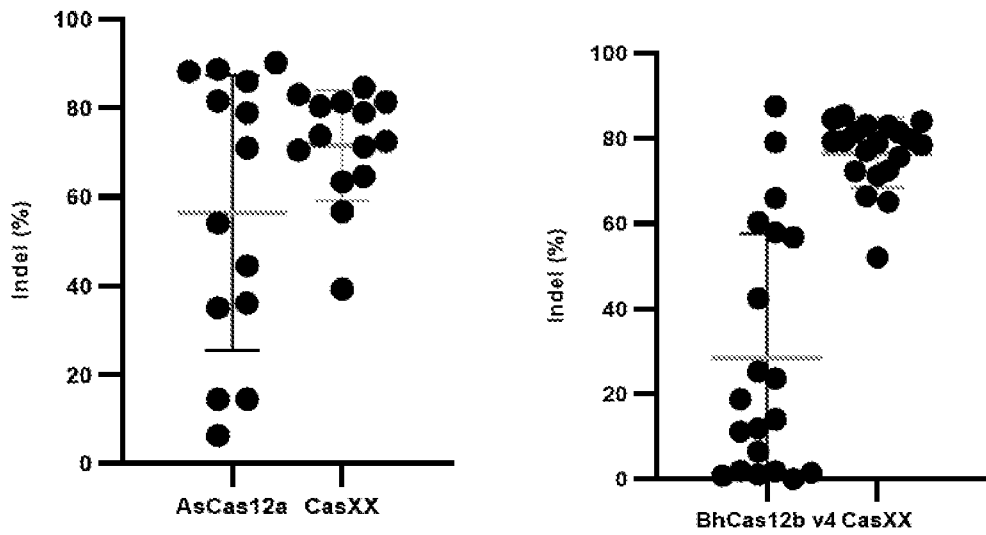
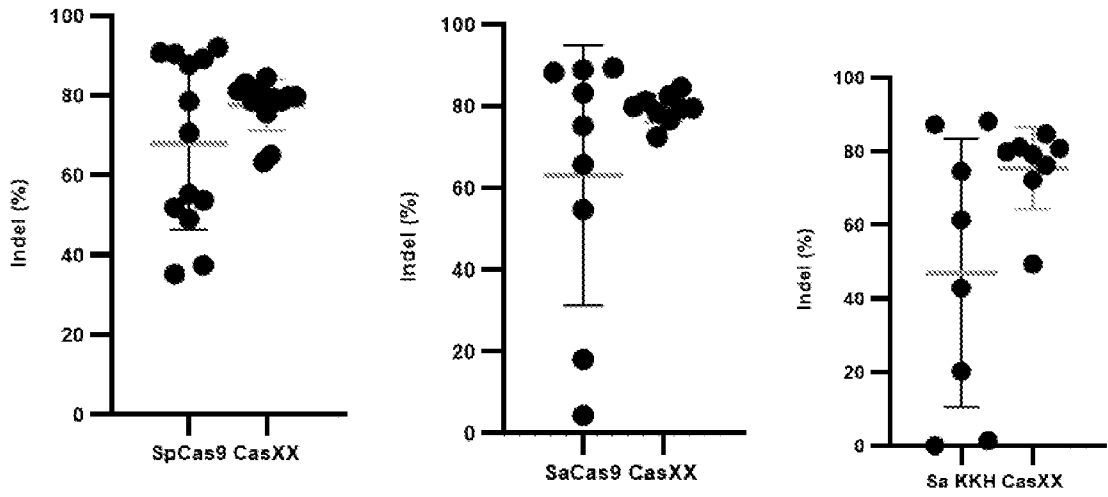


图 6a



CasXX与AsCas12a/BhCas12b v4比较

图 6b



CasXX与SpCas9/SaCas9/SaCas9-KKH比较

图 6c

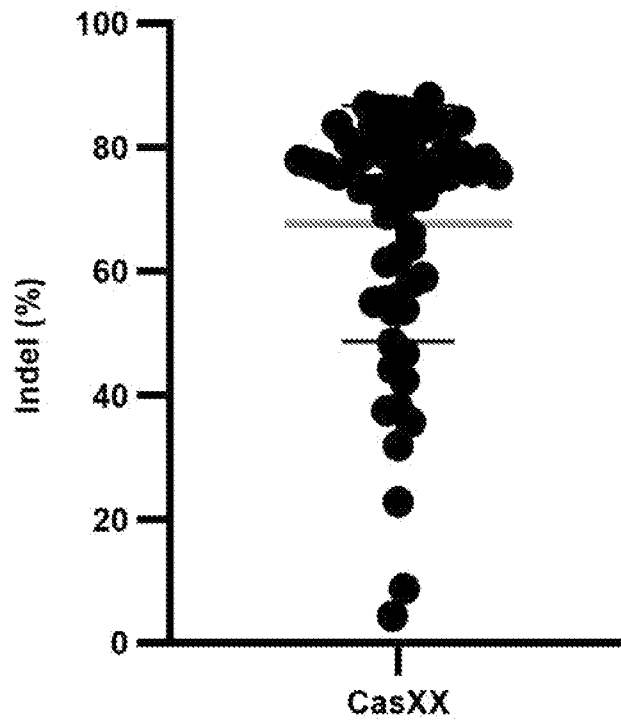


图 6d

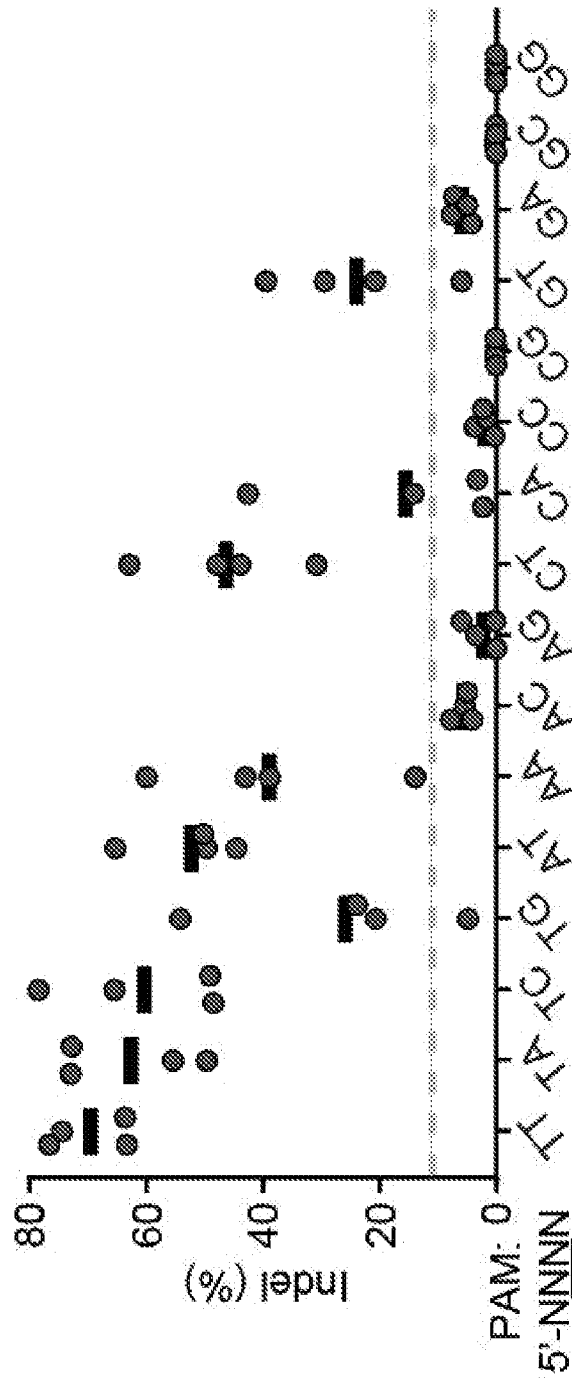


图7

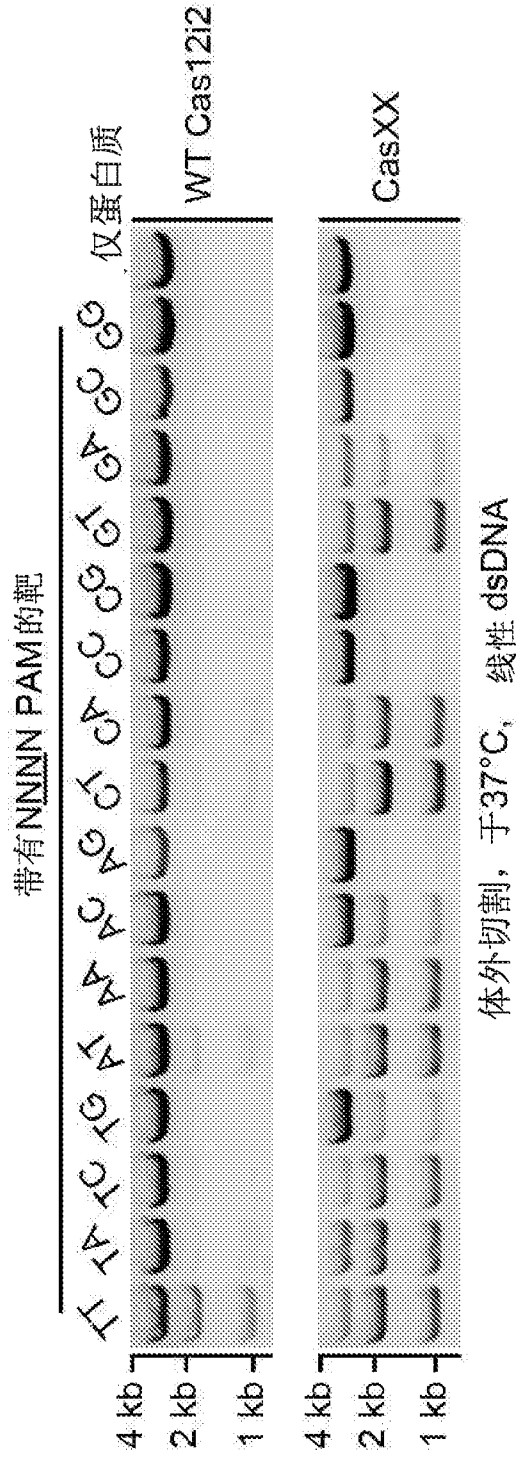


图8

Cas1211 1 MSNKKKMAETFA...
 Cas1212 1 ...
 Cas1211 86 ...
 Cas1212 80 ...
 Cas1211 171 ...
 Cas1212 166 ...
 Cas1211 257 ...
 Cas1212 249 ...
 Cas1211 343 ...
 Cas1212 335 ...
 Cas1211 429 ...
 Cas1212 397 ...
 Cas1211 518 ...
 Cas1212 475 ...
 Cas1211 602 ...
 Cas1212 553 ...
 Cas1211 688 ...
 Cas1212 639 ...
 Cas1211 775 ...
 Cas1212 723 ...
 Cas1211 861 ...
 Cas1212 804 ...
 Cas1211 948 ...
 Cas1212 896 ...
 Cas1211 1022 ...
 Cas1212 971 ...

EMX1-7

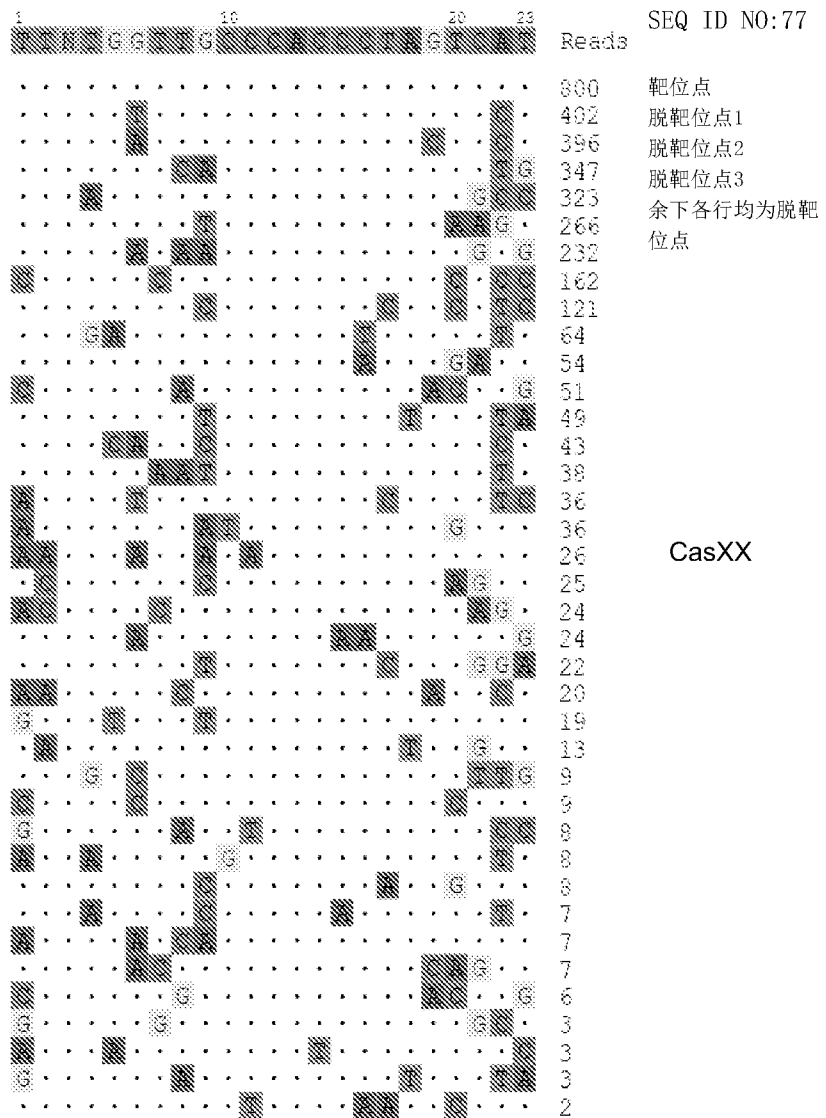


图10

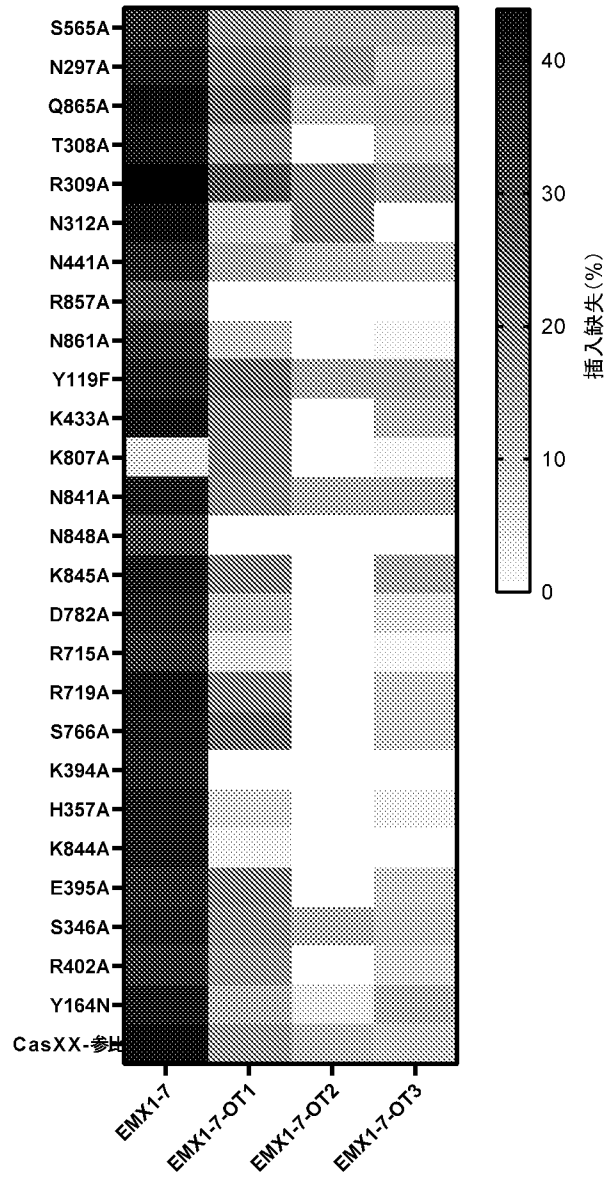


图11

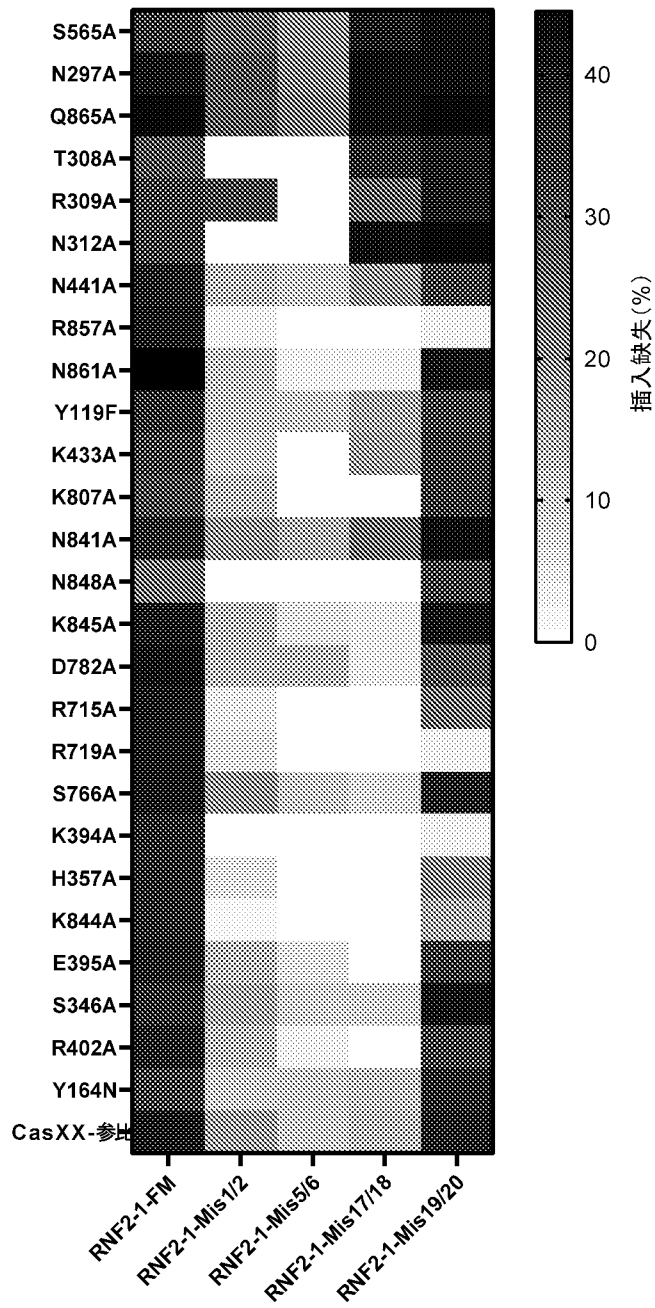


图12

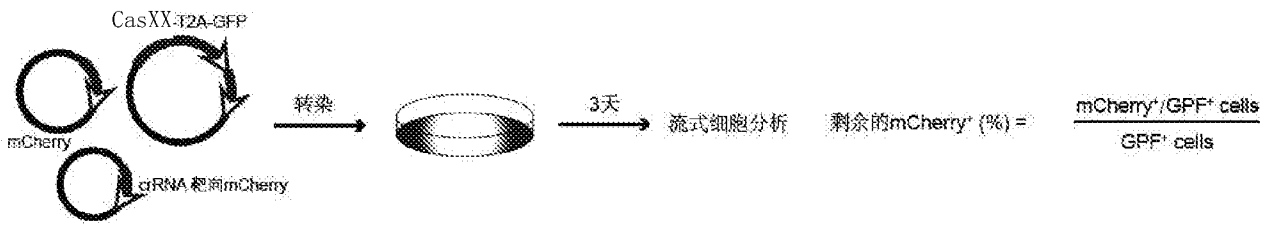


图13

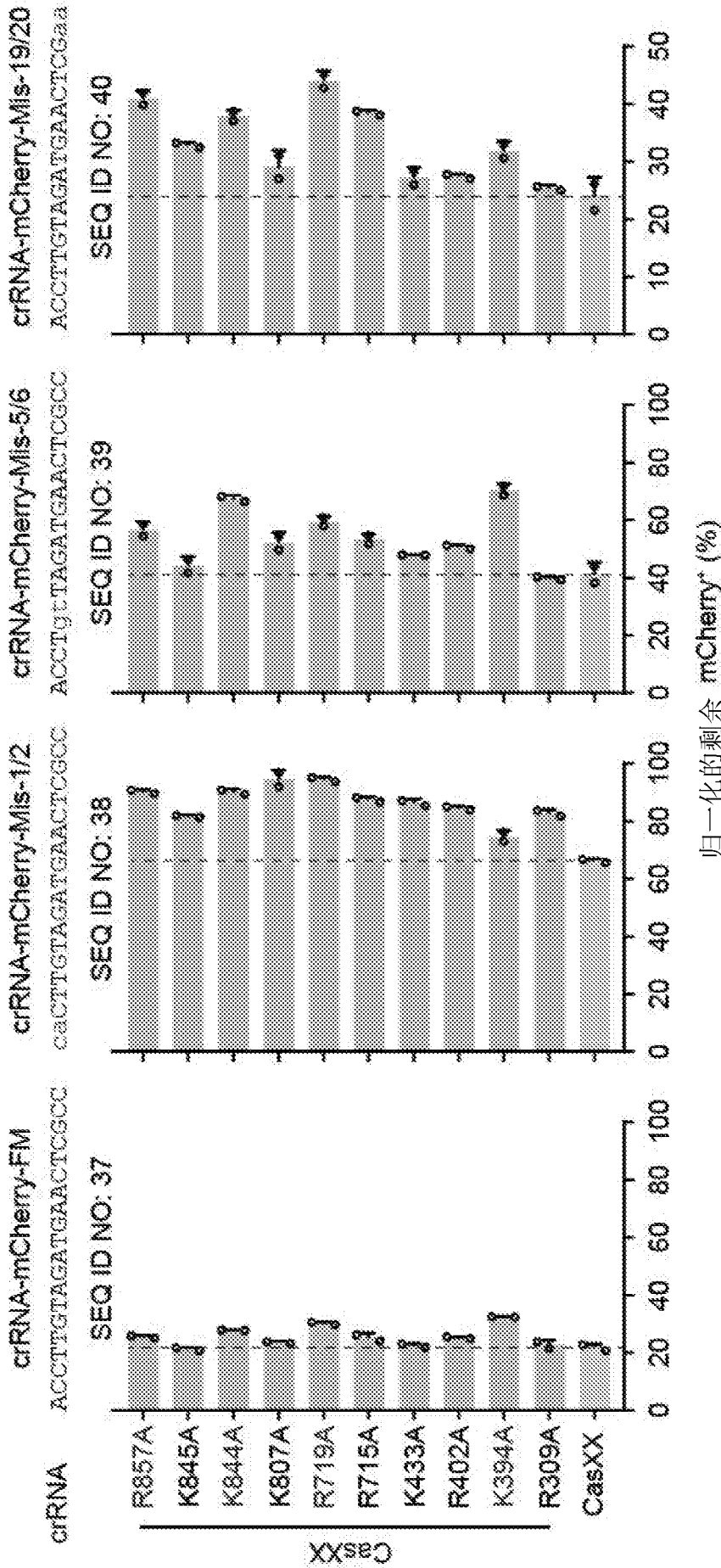


图14

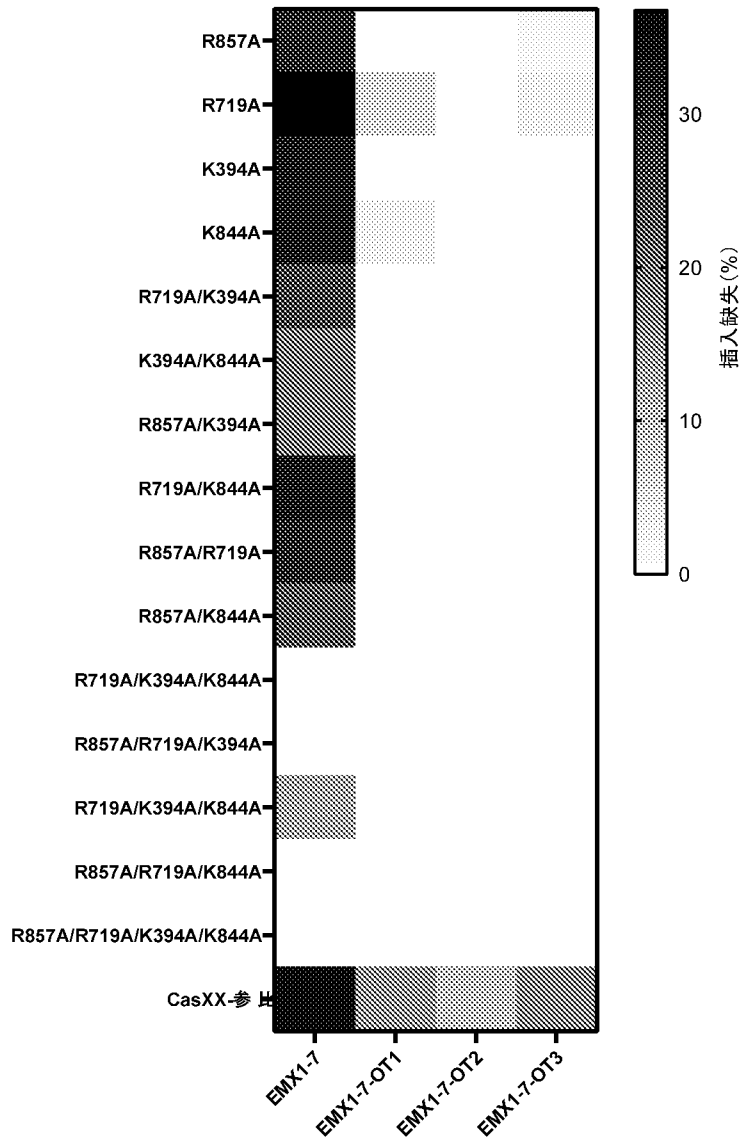


图15

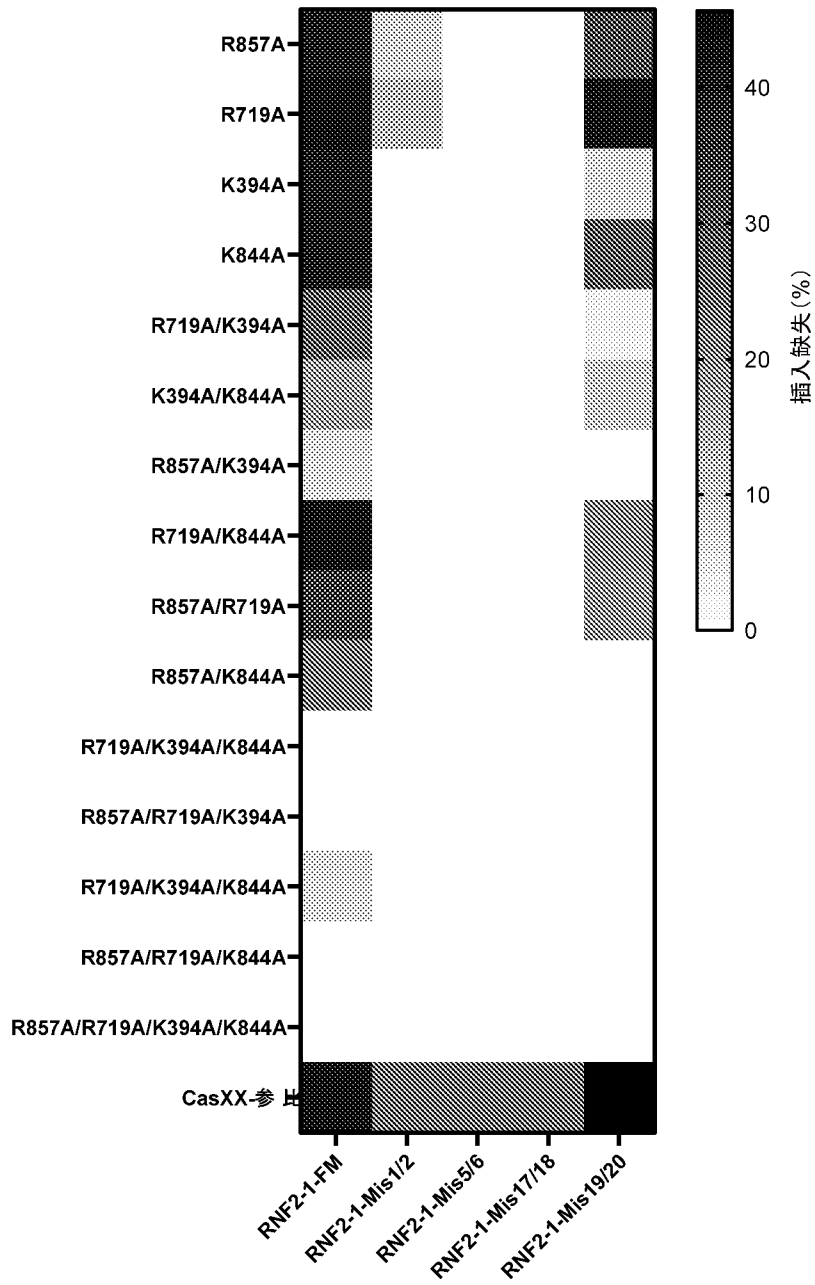


图16

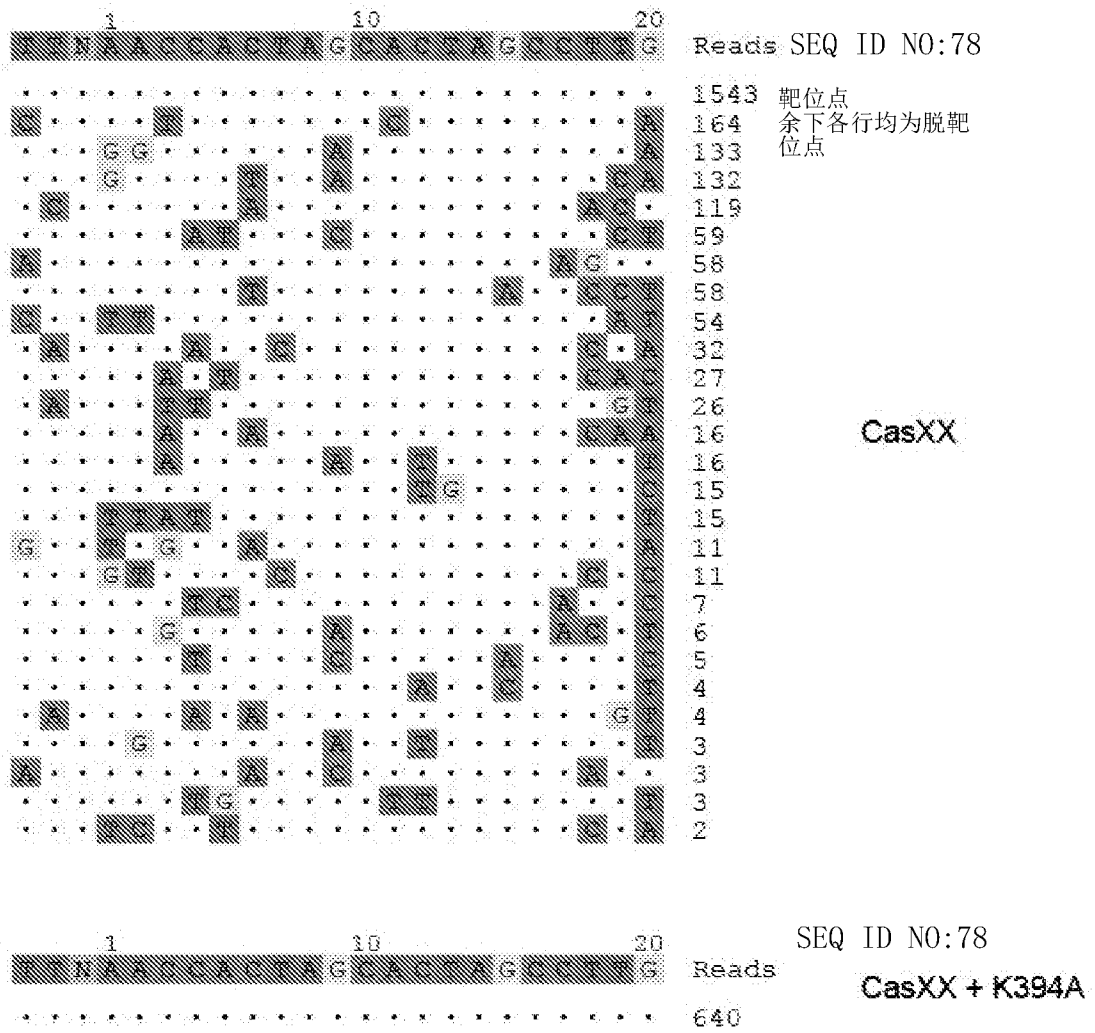


图17

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2022/095072

A. CLASSIFICATION OF SUBJECT MATTER		
C12N 9/22(2006.01)i; C12N 15/113(2010.01)i; C12Q 1/68(2018.01)i		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols)		
C12N; C12Q		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)		
CNTXT, WPABSC, ENTXTC, ENTXT, USTXT, WOTXT, CJFD, VEN, CNKI, 万方, WANFANG, 百度, BAIDU, PUBMED, ISI Web of Knowledge, NCBI GenBank, EBI-EMBL, 读秀, DUXIU: SEQ ID Nos: 1-79, Cas12, Cas12i, Cas12i2, Cas, 核酸酶, PAM, 正电, 氨基酸, 芳香环, 解旋, 打开, DNA, RuvC, 双螺旋, CRISPR, 效应蛋白, 融合, 二聚体, 缔结, 分割, 柔性, 突变, 替换, 增加, 插入, 修饰, 治疗, 疾病, 检测, Nuclease, Positively Charged, Amino Acid, Aromatic Ring, Unwind, DNA, Duplex, Effector, Fusion, Dimer, Associate, Split, Flexible, Mutation, Substitution, Addition+, Insert+, modif+, treatment, disease?, detect +.		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 113151215 A (INSTITUTE OF ZOOLOGY, CHINESE ACADEMY OF SCIENCES et al.) 23 July 2021 (2021-07-23) claims 1-28, and description, paragraphs 5-291	1-38
Y	CN 112195164 A (INSTITUTE OF ZOOLOGY, CHINESE ACADEMY OF SCIENCES et al.) 08 January 2021 (2021-01-08) claims 1-23, and description, paragraphs 224-419	1-3, 5-6, 8-9, 11-12, 14-15, 18-19, 23-38
Y	CN 112041444 A (ARBOR BIOTECHNOLOGIES INC.) 04 December 2020 (2020-12-04) description, paragraphs 209-231	1-3, 5-6, 8-9, 11-12, 14-15, 18-19, 23-38
A	CN 112004932 A (CHINA AGRICULTURAL UNIVERSITY) 27 November 2020 (2020-11-27) entire document	1-38
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search		Date of mailing of the international search report
10 August 2022		23 August 2022
Name and mailing address of the ISA/CN		Authorized officer
China National Intellectual Property Administration (ISA/CN) No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088, China		
Facsimile No. (86-10)62019451		Telephone No.

C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	韦涛等 (WEI, Tao et al.). "CRISPR/Cas基因编辑系统研究进展 (Research Progress of CRISPR/Cas Gene Editing System)" <i>生命的化学 (Chemistry of Life)</i> , Vol. 39, No. 1, 31 December 2019 (2019-12-31), pp. 36-45	1-38
A	ZHANG, Heng et al. "Mechanisms for Target Recognition and Cleavage by the Cas12i RNA-guided Endonuclease" <i>Nature Structural & Molecular Biology</i> , Vol. 27, No. 11, 07 September 2020 (2020-09-07), pp. 1069-1076	1-38
A	WO 2021087246 A1 (INARI AGRICULTURE, INC.) 06 May 2021 (2021-05-06) entire document	1-38

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
 - a. forming part of the international application as filed:
 - in the form of an Annex C/ST.25 text file.
 - on paper or in the form of an image file.
 - b. furnished together with the international application under PCT Rule 13ter.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.
 - c. furnished subsequent to the international filing date for the purposes of international search only:
 - in the form of an Annex C/ST.25 text file (Rule 13ter.1(a)).
 - on paper or in the form of an image file (Rule 13ter.1(b) and Administrative Instructions, Section 713).
2. In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.
3. Additional comments:

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claims Nos.: 32(部分)、34、35(部分)
because they relate to subject matter not required to be searched by this Authority, namely:
 - [1] The subject matter set forth in claims 32, 35 (relating to an in vivo portion), and 34 relates to a method for treating a human or animal body. Therefore, no international search is conducted according to PCT Rule 39.1(iv). A search was conducted on the basis of the method for modifying a cell or nucleic acid in vitro by engineered CRISPR-Cas12i of claims 32 and 34-35 or a use thereof in the preparation of a drug for treating cancer.
2. Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2022/095072

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	113151215	A	23 July 2021	None			
CN	112195164	A	08 January 2021	None			
CN	112041444	A	04 December 2020	JP	2021516970	A	15 July 2021
				WO	2019178427	A1	19 September 2019
				US	2022098579	A1	31 March 2022
				EP	3765615	A1	20 January 2021
				DE	202019005567	U1	16 February 2021
				CA	3093334	A1	19 September 2019
				US	2021123047	A1	29 April 2021
				AU	2019236210	A1	10 September 2020
				US	2020063126	A1	27 February 2020
				US	2020407715	A1	31 December 2020
				US	2020407716	A1	31 December 2020
CN	112004932	A	27 November 2020	WO	2019201331	A1	24 October 2019
WO	2021087246	A1	06 May 2021	None			

国际检索报告

国际申请号

PCT/CN2022/095072

<p>A. 主题的分类</p> <p>C12N 9/22(2006.01)i; C12N 15/113(2010.01)i; C12Q 1/68(2018.01)i</p> <p>按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类</p>																				
<p>B. 检索领域</p> <p>检索的最低限度文献(标明分类系统和分类号)</p> <p>C12N; C12Q</p> <p>包含在检索领域中的除最低限度文献以外的检索文献</p> <p>在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))</p> <p>CNXTX, WPABSC, ENTXTX, ENTXT, USTXT, WOTXT, CJFD, VEN, CNKI, 万方, 百度, PUBMED, ISI Web of Knowledge, NCBI GenBank, EBI-EMBL, 读秀:SEQ ID Nos: 1-79, Cas12, Cas12i, Cas12i2, Cas, 核酸酶, PAM, 正电, 氨基酸, 芳香环, 解旋, 打开, DNA, RuvC, 双螺旋, CRISPR, 效应蛋白, 融合, 二聚体, 缔结, 分割, 柔性, 突变, 替换, 增加, 插入, 修饰, 治疗, 疾病, 检测, Nuclease, Positively Charged, Amino Acid, Aromatic Ring, Unwind, DNA, Duplex, Effector, Fusion, Dimer, Associate, Split, Flexible, Mutation, Substitution, Addition+, Insert+, modif+, treatment, disease?, detect+.</p>																				
<p>C. 相关文件</p> <table border="1"> <thead> <tr> <th>类型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>PX</td> <td>CN 113151215 A (中国科学院动物研究所等) 2021年7月23日 (2021 - 07 - 23) 权利要求1-28, 说明书第5-291段</td> <td>1-38</td> </tr> <tr> <td>Y</td> <td>CN 112195164 A (中国科学院动物研究所等) 2021年1月8日 (2021 - 01 - 08) 权利要求1-23, 说明书第224-419段</td> <td>1-3, 5-6, 8-9, 11-12, 14-15, 18-19, 23-38</td> </tr> <tr> <td>Y</td> <td>CN 112041444 A (阿伯生物技术公司) 2020年12月4日 (2020 - 12 - 04) 说明书第209-231段</td> <td>1-3, 5-6, 8-9, 11-12, 14-15, 18-19, 23-38</td> </tr> <tr> <td>A</td> <td>CN 112004932 A (中国农业大学) 2020年11月27日 (2020 - 11 - 27) 全文</td> <td>1-38</td> </tr> <tr> <td>A</td> <td>韦涛等. "CRISPR/Cas基因编辑系统研究进展" 生命的化学, 第39卷, 第1期, 2019年12月31日 (2019 - 12 - 31), 第36-45页</td> <td>1-38</td> </tr> </tbody> </table>			类型*	引用文件, 必要时, 指明相关段落	相关的权利要求	PX	CN 113151215 A (中国科学院动物研究所等) 2021年7月23日 (2021 - 07 - 23) 权利要求1-28, 说明书第5-291段	1-38	Y	CN 112195164 A (中国科学院动物研究所等) 2021年1月8日 (2021 - 01 - 08) 权利要求1-23, 说明书第224-419段	1-3, 5-6, 8-9, 11-12, 14-15, 18-19, 23-38	Y	CN 112041444 A (阿伯生物技术公司) 2020年12月4日 (2020 - 12 - 04) 说明书第209-231段	1-3, 5-6, 8-9, 11-12, 14-15, 18-19, 23-38	A	CN 112004932 A (中国农业大学) 2020年11月27日 (2020 - 11 - 27) 全文	1-38	A	韦涛等. "CRISPR/Cas基因编辑系统研究进展" 生命的化学, 第39卷, 第1期, 2019年12月31日 (2019 - 12 - 31), 第36-45页	1-38
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求																		
PX	CN 113151215 A (中国科学院动物研究所等) 2021年7月23日 (2021 - 07 - 23) 权利要求1-28, 说明书第5-291段	1-38																		
Y	CN 112195164 A (中国科学院动物研究所等) 2021年1月8日 (2021 - 01 - 08) 权利要求1-23, 说明书第224-419段	1-3, 5-6, 8-9, 11-12, 14-15, 18-19, 23-38																		
Y	CN 112041444 A (阿伯生物技术公司) 2020年12月4日 (2020 - 12 - 04) 说明书第209-231段	1-3, 5-6, 8-9, 11-12, 14-15, 18-19, 23-38																		
A	CN 112004932 A (中国农业大学) 2020年11月27日 (2020 - 11 - 27) 全文	1-38																		
A	韦涛等. "CRISPR/Cas基因编辑系统研究进展" 生命的化学, 第39卷, 第1期, 2019年12月31日 (2019 - 12 - 31), 第36-45页	1-38																		
<p><input checked="" type="checkbox"/> 其余文件在C栏的续页中列出。 <input checked="" type="checkbox"/> 见同族专利附件。</p> <p>* 引用文件的具体类型: "A" 认为不特别相关的表示了现有技术一般状态的文件 "E" 在国际申请日的当天或之后公布的在先申请或专利 "L" 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) "O" 涉及口头公开、使用、展览或其他方式公开的文件 "P" 公布日先于国际申请日但迟于所要求的优先权日的文件 "T" 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 "X" 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 "Y" 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 "&" 同族专利的文件</p>																				
<p>国际检索实际完成的日期</p> <p>2022年8月10日</p>		<p>国际检索报告邮寄日期</p> <p>2022年8月23日</p>																		
<p>ISA/CN的名称和邮寄地址</p> <p>中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088</p> <p>传真号 (86-10)62019451</p>		<p>授权官员</p> <p>沈晶晶</p> <p>电话号码 86-(10)-53961944</p>																		

C. 相关文件		
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求
A	ZHANG, H. 等. "Mechanisms for target recognition and cleavage by the Cas12i RNA-guided endonuclease" Nature Structural & Molecular Biology, 第27卷, 第11期, 2020年9月7日 (2020 - 09 - 07), 第1069-1076页	1-38
A	WO 2021087246 A1 (INARI AGRICULTURE, INC.) 2021年5月6日 (2021 - 05 - 06) 全文	1-38

第1栏 核苷酸和/或氨基酸序列(续第1页第1. c项)

1. 关于国际申请中所公开的任何核苷酸和/或氨基酸序列,国际检索是基于下列序列列表进行的:
- a. 作为国际申请的一部分提交的:
- 附件C/ST. 25文本文件形式
 - 纸件或图形文件形式
- b. 根据细则13之三. 1(a) 仅为国际检索目的以附件C/ST. 25文本文件形式与国际申请同时提交的:
- c. 仅为国际检索目的在国际申请日之后提交的:
- 附件C/ST. 25文本文件形式(细则13之三. 1(a))
 - 纸件或图形文件形式(细则13之三. 1(b)和行政规程第713段)
2. 另外,在提交/提供了多个版本或副本的序列列表的情况下,提供了关于随后提交的或附加的副本中的信息与申请时提交的作为申请一部分的序列列表的信息相同或未超出申请时提交的申请中的信息范围(如适用)的所需声明。
3. 补充意见:

第II栏 某些权利要求被认为是不能检索的意见(续第1页第2项)

根据条约第17条(2)(a), 对某些权利要求未做国际检索报告的理由如下:

1. 权利要求: 32(部分)、34、35(部分)
因为它们涉及不要求本单位进行检索的主题, 即:
[1] 权利要求32与35(涉及体内部分)、34的保护主题涉及对人体或者动物体进行治疗的方法, 因此依据PCT实施细则39.1(iv)不进行国际检索。检索是基于权利要求32、34-35的工程化CRISPR-Cas12i 修饰体外细胞或核酸的方法或其在制备治疗癌症的药物用途作出。
2. 权利要求:
因为它们涉及国际申请中不符合规定的要求的部分, 以致不能进行任何有意义的国际检索, 具体地说:
3. 权利要求:
因为它们是从属权利要求, 并且没有按照细则6.4(a)第2句和第3句的要求撰写。

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2022/095072

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	113151215	A	2021年7月23日	无			
CN	112195164	A	2021年1月8日	无			
CN	112041444	A	2020年12月4日	JP	2021516970	A	2021年7月15日
				WO	2019178427	A1	2019年9月19日
				US	2022098579	A1	2022年3月31日
				EP	3765615	A1	2021年1月20日
				DE	202019005567	U1	2021年2月16日
				CA	3093334	A1	2019年9月19日
				US	2021123047	A1	2021年4月29日
				AU	2019236210	A1	2020年9月10日
				US	2020063126	A1	2020年2月27日
				US	2020407715	A1	2020年12月31日
				US	2020407716	A1	2020年12月31日
CN	112004932	A	2020年11月27日	WO	2019201331	A1	2019年10月24日
WO	2021087246	A1	2021年5月6日	无			