



(12)发明专利申请

(10)申请公布号 CN 106957831 A

(43)申请公布日 2017.07.18

(21)申请号 201710343397.8

(22)申请日 2017.05.16

(71)申请人 上海交通大学

地址 200240 上海市闵行区东川路800号

(72)发明人 吴强 李金环 寿佳

(74)专利代理机构 上海光华专利事务所 31219

代理人 李艳 许亦琳

(51) Int. Cl.

C12N 9/22(2006.01)

C12N 15/55(2006.01)

C12N 15/63(2006.01)

C12N 15/10(2006.01)

权利要求书2页 说明书28页

序列表19页 附图7页

(54)发明名称

一种Cas9核酸酶K918A及其用途

(57)摘要

本发明属于生物技术领域,具体涉及一种Cas9核酸酶及用途。本发明的Cas9核酸酶(Cas9-K918A),具有Cas9核酸酶活性,适用于CRISPR/Cas9系统,所述Cas9核酸酶(Cas9-K918A)是将野生型Cas9核酸酶第918位赖氨酸突变成丙氨酸获得。采用所述Cas9核酸酶(Cas9-K918A)对DNA双链进行切割可产生突出断裂末端,以补平连接的方式可加入与突出断裂末端互补的碱基,能够实现基因组DNA片段的特定位置的精准编辑。

1. 一种Cas9核酸酶,具有Cas9核酸酶活性,适用于CRISPR/Cas9系统,所述Cas9核酸酶是将野生型Cas9核酸酶第918位赖氨酸突变成丙氨酸获得。

2. 根据权利要求1所述的Cas9核酸酶,其特征在于,所述Cas9核酸酶与野生型Cas9核酸酶相比,对目的基因组DNA片段进行切割时产生的突出断裂末端与钝断裂末端的比例不同。

3. 根据权利要求1所述的Cas9核酸酶,其特征在于,所述野生型Cas9核酸酶为SpCas9。

4. 根据权利要求3所述的Cas9核酸酶,其特征在于,所述野生型Cas9核酸酶的氨基酸序列如SEQ ID NO.7所示。

5. 根据权利要求1所述的Cas9核酸酶,其特征在于,所述Cas9核酸酶含有如SEQ ID NO.9所示的氨基酸序列。

6. 一种多核苷酸,其编码如权利要求1~5任一项所述的Cas9核酸酶。

7. 一种表达载体,其含有如权利要求6所述的多核苷酸。

8. 一种宿主细胞,其被如权利要求7所述的表达载体所转化。

9. 一种制备如权利要求1~5任一项所述的Cas9核酸酶的方法,包括步骤:构建含有如权利要求1~5任一项所述Cas9核酸酶编码多核苷酸序列的表达载体,然后将所述表达载体转化至宿主细胞中诱导表达,从表达产物中分离获得所述的Cas9核酸酶。

10. 如权利要求1~5任一项所述的Cas9核酸酶、或如权利要求6所述的多核苷酸或如权利要求7所述的表达载体用于基因组DNA片段编辑或用于制备基因组DNA片段编辑工具的用途。

11. 根据权利要求10所述的用途,其特征在于,所述编辑包括单位点编辑和多位点编辑,所述多位点编辑的编辑位点数为两个及以上。

12. 根据权利要求10所述的用途,其特征在于,所述编辑的方式包括突变、删除、反转或倒位、重复、易位或插入。

13. 一种DNA基因组片段编辑工具,为CRISPR/Cas9系统,所述CRISPR/Cas9系统包括如权利要求1~5任一项所述的Cas9核酸酶、或如权利要求6所述的多核苷酸或如权利要求7所述的表达载体。

14. 根据权利要求13所述的一种DNA片段编辑工具,其特征在于,所述CRISPR/Cas9系统还包括针对目的DNA片段的一个或多个sgRNA,所述多个是指两个及以上。

15. 一种基因组DNA片段编辑方法,采用如权利要求1~4任一项所述Cas9核酸酶以及与之配合的一个或多个sgRNA,利用CRISPR/Cas9系统对待编辑的基因组DNA片段进行编辑。

16. 根据权利要求15所述的基因组DNA片段编辑方法,其特征在于,所述编辑包括单位点编辑和多位点编辑,所述多位点编辑的编辑位点数为两个及以上。

17. 根据权利要求15所述的基因组DNA片段编辑方法,所述编辑的方式包括突变、删除、反转或倒位、重复、易位或插入。

18. 根据权利要求15所述的基因组DNA片段编辑方法,其特征在于,将如权利要求6所述的含有Cas9核酸酶编码多核苷酸的表达载体以及与之配合的一个或多个sgRNA一同转入细胞中,对待编辑的基因组DNA片段进行编辑。

19. 一种基因组DNA片段单位点编辑方法,利用CRISPR/Cas9系统,采用如权利要求1~5任一项所述的Cas9核酸酶对DNA双链进行切割产生突出断裂末端,通过细胞自身修复系统,以补平连接的方式加入与突出断裂末端互补的碱基。

20. 根据权利要求19所述的基因组DNA片段单位点编辑方法,其特征在于,所述方法可以改变单位点编辑时碱基突变的特征。

一种Cas9核酸酶K918A及其用途

技术领域

[0001] 本发明属于生物技术领域,具体涉及一种Cas9核酸酶K918A及用途。

背景技术

[0002] 自从人类基因组计划(Human Genome Project)和DNA元件百科全书(Encyclopedia of DNA Elements)项目的完成,科学家们分析和鉴定了大量的基因组中的基因和DNA调控元件[1,2]。在基因表达调控中起重要作用的DNA调控元件包括启动子、增强子、沉默子和绝缘子等。然而,多数调控元件的功能并没有得到实验的验证和阐明[2-8]。探索基因和DNA调控元件的功能,可以通过遗传学DNA片段编辑来进行研究。

[0003] 早期的基因编辑和基因功能修饰是通过基因转座和转基因实现的[9-14]。伴随测序技术的发展反向遗传学被应用于对基因组进行特定的突变[15,16]。特别是依赖于同源重组的基因打靶小鼠迅速地被应用到科学研究中[15,17,18]。此外,在小鼠和斑马鱼中DNA片段的反转和重复被应用于去研究特定的基因组结构变化[19-24]。

[0004] 近几年,源于细菌和古菌的II型成簇规律间隔短回文重复系统[Clustered regularly interspaced short palindromic repeats(CRISPR)/CRISPR-associated nuclease 9(Cas9),CRISPR/Cas9]是新兴基因组编辑技术[25-27],由于它设计简单和操作方便,迅速地被应用到真核基因组编辑。我们利用CRISPR/Cas9系统在人细胞系和小鼠中实现了DNA片段遗传编辑(删除、反转和重复)[28]。通过Cas9和两个sgRNAs在基因组中进行两个位点靶向断裂后在CtIP等蛋白参与的修复系统作用下可以实现DNA片段的删除、反转(倒位)、重复、易位和插入(如果提供供体)等[29-32]。通过对DNA片段编辑的遗传操作,能够用来研究原钙粘蛋白和珠蛋白的基因表达调控和三维基因组结构[28,31-33]。

[0005] 目前可以通过CRISPR/Cas9系统实现DNA片段的编辑,但是对于深入研究特定DNA区段的精准功能,有效地实现DNA片段的精准遗传编辑的Cas9核酸酶还有待发现。

发明内容

[0006] 为了克服现有技术中所存在的问题,本发明的目的在于提供一种Cas9核酸酶及用途。

[0007] 为了实现上述目的以及其他相关目的,本发明采用如下技术方案:

[0008] 本发明的第一方面,提供一种Cas9核酸酶(Cas9-K918A),适用于CRISPR/Cas9系统,所述Cas9核酸酶(Cas9-K918A)是将野生型Cas9核酸酶第918位赖氨酸突变成丙氨酸获得。

[0009] 优选地,与野生型Cas9核酸酶相比,所述Cas9核酸酶(Cas9-K918A)对目的基因组DNA片段进行切割时产生的突出断裂末端与钝断裂末端的比例不同。

[0010] 优选地,所述野生型Cas9核酸酶为SpCas9。

[0011] 进一步地,所述野生型Cas9核酸酶的氨基酸序列如SEQ ID NO.7所示。

[0012] 优选地,所述Cas9核酸酶(Cas9-K918A)含有如SEQ ID NO.9所示的氨基酸序列。

- [0013] 优选地,所述Cas9核酸酶 (Cas9-K918A) 的氨基酸序列如SEQ ID NO.9所示。
- [0014] 本发明的第二方面,提供一种多核苷酸,其编码所述Cas9核酸酶 (Cas9-K918A) 。
- [0015] 本发明的第三方面,提供一种表达载体,其含有前述多核苷酸。
- [0016] 本发明的第四方面,提供一种宿主细胞,其被前述表达载体所转化。
- [0017] 本发明的第五方面,提供一种制备所述Cas9核酸酶 (Cas9-K918A) 的方法,包括步骤:构建含有Cas9核酸酶 (Cas9-K918A) 编码多核苷酸的表达载体,然后将所述表达载体转化至宿主细胞中诱导表达,从表达产物中分离获得所述的Cas9核酸酶 (Cas9-K918A) 。
- [0018] 本发明的第六方面,提供前述Cas9核酸酶 (Cas9-K918A) 或其编码多核苷酸或含有所述编码多核苷酸的表达载体用于基因组DNA片段编辑或用于制备基因组DNA片段编辑工具的用途。
- [0019] 优选地,所述编辑包括单位点编辑和多位点编辑。所述多位点编辑的编辑位点数为两个及以上。
- [0020] 优选地,所述编辑的方式包括突变、删除、反转或倒位、重复、易位或插入。
- [0021] 本发明的第七方面,提供一种基因组DNA片段编辑工具,所述基因组DNA片段编辑工具为CRISPR/Cas9系统,所述CRISPR/Cas9系统包括前述Cas9核酸酶 (Cas9-K918A) 或其编码多核苷酸或含有所述编码多核苷酸的表达载体。
- [0022] 优选地,所述CRISPR/Cas9系统包括前述Cas9-K918A和针对目的DNA片段的一个或多个sgRNA。所述多个是指两个及以上。
- [0023] 本发明的第八方面,提供一种基因组DNA片段编辑方法,采用前述Cas9核酸酶 (Cas9-K918A) 以及与之配合的一个或多个sgRNA,利用CRISPR/Cas9系统对待编辑的基因组DNA片段进行编辑。
- [0024] 优选地,所述编辑包括单位点编辑和多位点编辑。所述多位点编辑的编辑位点数为两个及以上。
- [0025] 优选地,所述编辑的方式包括突变、删除、反转或倒位、重复、易位或插入。
- [0026] 优选地,将含有前述Cas9核酸酶 (Cas9-K918A) 编码多核苷酸的表达载体以及与之配合的一个或多个sgRNA一同转入细胞中,对待编辑的基因组DNA片段进行编辑。
- [0027] 本发明的第九方面,提供一种基因组DNA片段单位点编辑方法,利用CRISPR/Cas9系统,采用如权利要求1所述的Cas9核酸酶 (Cas9-K918A) 对DNA双链进行切割产生突出断裂末端,通过细胞自身修复系统,以补平连接的方式加入与突出断裂末端互补的碱基。
- [0028] 所述基因组DNA片段单位点编辑方法可以改变单位点编辑时碱基突变的特征。
- [0029] 与现有技术相比,本发明具有如下有益效果:
- [0030] 本发明的Cas9核酸酶 (Cas9-K918A) ,适用于CRISPR/Cas9系统,所述Cas9核酸酶 (Cas9-K918A) 含有如SEQ ID NO.9所示的氨基酸序列,所述Cas9核酸酶 (Cas9-K918A) 与野生型Cas9核酸酶相比,对目的基因组DNA片段进行切割时产生的突出断裂末端及钝断裂末端的比例相对于野生型Cas9核酸酶不同。采用所述Cas9核酸酶 (Cas9-K918A) 对DNA双链进行切割可产生突出断裂末端,通过细胞自身修复系统以补平连接的方式可加入与突出断裂末端互补的碱基,能够实现对基因组DNA片段的特定位置加入特定碱基的精准编辑。

附图说明

- [0031] 图1A: Cas9在两个sgRNAs介导下对DNA双链进行切割产生四个断裂末端, 这些断裂末端在细胞修复系统的作用下产生DNA片段删除、反转和重复。
- [0032] 图1B: 针对HS51位点的DNA片段删除、反转和重复情况。
- [0033] 图1C: DNA片段删除接头处存在“G”的加入。
- [0034] 图1D: DNA片段重复接头处存在“T”的加入。
- [0035] 图1E: DNA片段下游反转接头处存在“A”、“G”和“AG”的加入。
- [0036] 图1F: 针对这两个特定序列的sgRNAs, Cas9切割方式比例特征。
- [0037] 图2A: Cas9核酸酶结构示意图。
- [0038] 图2B: β -globin RE2位点进行DNA片段编辑的两个sgRNAs的示意图。
- [0039] 图2C: 通过检测DNA片段重复接头连接情况统计出各Cas9核酸酶在sgRNA1的介导下对基因组DNA片段进行切割时所产生的各种切割末端的占比。
- [0040] 图2D: 针对上游sgRNA1, Cas9以及Cas9突变体对目的DNA片段的切割情况。
- [0041] 图2E: 通过检测DNA片段删除接头连接情况统计出各Cas9核酸酶在sgRNA2的介导下对基因组DNA片段进行切割时所产生的各种切割末端的占比。
- [0042] 图2F: 针对下游sgRNA2, Cas9以及Cas9突变体对目的DNA片段的切割情况。
- [0043] 图2G: Cas9以及Cas9突变体在DNA片段反转一侧接头处碱基加入的实际和预测比例。
- [0044] 图3A: 在STM位点, 针对上游sgRNA1, Cas9以及Cas9突变体对目的DNA片段的切割情况。
- [0045] 图3B: 在STM位点, 针对下游sgRNA2, Cas9以及Cas9突变体对目的DNA片段的切割情况。

具体实施方式

[0046] 一、Cas9核酸酶

[0047] 本发明的Cas9核酸酶 (Cas9-K918A), 具有Cas9核酸酶活性, 适用于CRISPR/Cas9系统, 所述Cas9核酸酶 (Cas9-K918A) 是将野生型Cas9核酸酶第918位赖氨酸突变成丙氨酸获得。所述Cas9核酸酶 (Cas9-K918A) 与野生型Cas9核酸酶相比, 对目的基因组DNA片段进行切割时产生的突出断裂末端与钝断裂末端的比例不同。进一步地, 所述野生型Cas9核酸酶为SpCas9。进一步地, 所述野生型Cas9核酸酶的氨基酸序列如SEQ ID NO.7所示。

[0048] 进一步地, 所述Cas9核酸酶 (Cas9-K918A) 含有如SEQ ID NO.9所示的氨基酸序列。本发明的一些实施方式中, 例举了所述Cas9-K918A的氨基酸序列如SEQ ID NO.9所示。

[0049] 二、编码Cas9核酸酶的多核苷酸

[0050] 编码所述Cas9核酸酶 (Cas9-K918A) 的多核苷酸, 可以是DNA形式或RNA形式。DNA形式包括cDNA、基因组DNA或人工合成的DNA。DNA可以是单链的或是双链的。

[0051] 编码所述Cas9核酸酶 (Cas9-K918A) 的多核苷酸, 可以通过本领域技术人员熟知的任何适当的技术制备。所述技术见于本领域的一般描述, 如《分子克隆实验指南》(J. 萨姆布鲁克等, 科学出版社, 1995)。包括但不限于重组DNA技术、化学合成等方法。

[0052] 本发明的一些实施方式中, 例举了编码所述Cas9核酸酶 (Cas9-K918A) 的多核苷酸如SEQ ID NO.10所示。

[0053] 三、表达载体

[0054] 所述表达载体含有编码所述Cas9核酸酶 (Cas9-K918A) 的多核苷酸。本领域的技术人员熟知的方法能用于构建所述表达载体。这些方法包括重组DNA技术、DNA合成技术等。可将所述Cas9核酸酶 (Cas9-K918A) 的DNA有效连接到载体中的多克隆位点上,以指导mRNA合成进而表达蛋白。

[0055] 四、宿主细胞

[0056] 所述宿主细胞被表达所述Cas9核酸酶 (Cas9-K918A) 的表达载体所转化。宿主细胞可以是原核细胞,如细菌细胞;或是低等真核细胞,如酵母细胞;或是高等真核细胞,如哺乳动物细胞。代表性例子有:大肠杆菌,链霉菌属;鼠伤寒沙门菌、李斯特细菌;真菌细胞如酵母;植物细胞;果蝇S2或Sf9的昆虫细胞;CHO、COS. 293细胞、或Bowes黑素瘤细胞的动物细胞等。

[0057] 五、制备Cas9核酸酶 (Cas9-K918A) 的方法

[0058] 制备前述Cas9核酸酶 (Cas9-K918A) 的方法,包括步骤:构建含有Cas9核酸酶 (Cas9-K918A) 编码多核苷酸序列的表达载体,然后将所述表达载体转化至宿主细胞中诱导表达,从表达产物中分离获得所述的Cas9核酸酶 (Cas9-K918A)。

[0059] 本领域技术人员可根据Cas9核酸酶 (Cas9-K918A) 的性质来选择合适的表达载体和宿主细胞。

[0060] 六、Cas9核酸酶 (Cas9-K918A) 或其编码多核苷酸或含有所述编码多核苷酸的表达载体的用途

[0061] 本发明的Cas9核酸酶 (Cas9-K918A) 或其编码多核苷酸或含有所述编码多核苷酸的表达载体可用于基因组DNA片段编辑或用于制备基因组DNA片段编辑工具。

[0062] 进一步地,所述编辑包括单位点编辑和多位点编辑。所述多位点编辑的编辑位点数为两个及以上。所述编辑的方式包括突变、删除、反转或倒位、重复、易位或插入。

[0063] 七、基因组DNA片段编辑工具

[0064] 本发明的基因组DNA片段编辑工具可以是CRISPR/Cas9系统,所述CRISPR/Cas9系统包括前述Cas9核酸酶 (Cas9-K918A) 或其编码多核苷酸或含有所述编码多核苷酸的表达载体。进一步地,所述CRISPR/Cas9系统还包括针对目的DNA片段的一个或多个sgRNA。所述sgRNA为针对目的DNA片段所设计,在sgRNA (Single-guide RNA) 的介导下,Cas9-K918A能够在PAM (Protospacer adjacent motif) 位点上游对DNA双链进行切割,形成DNA双链断裂,通过细胞自身修复系统,完成DNA片段的精准编辑。针对目的基因的sgRNA可以是一个或两个及以上。当sgRNA是一个的时候,可以实现对目的DNA片段的单位点编辑,当sgRNA是两个及以上的时候,可以实现对目的DNA片段的多位点编辑。

[0065] 八、基因组DNA片段编辑方法

[0066] 本发明的基因组DNA片段编辑方法,采用前述Cas9核酸酶 (Cas9-K918A) 以及与之配合的一个或多个sgRNA,利用CRISPR/Cas9系统对待编辑的基因组DNA片段进行编辑。所述编辑包括单位点编辑和多位点编辑。所述多位点编辑的编辑位点数为两个及以上。当sgRNA是一个的时候,可以实现对目的DNA片段的单位点编辑,当sgRNA是两个及以上的时候,可以实现对目的DNA片段的多位点编辑。进一步地,可将前述Cas9核酸酶 (Cas9-K918A) 编码多核苷酸的表达载体以及与之配合的一个或多个sgRNA一同转入细胞中,对待编辑的基因组DNA

片段进行编辑。

[0067] 九、基因组DNA片段单位点编辑方法

[0068] 利用CRISPR/Cas9系统,采用本发明的Cas9核酸酶(Cas9-K918A)对DNA双链进行切割产生突出断裂末端,以补平连接的方式加入与突出断裂末端互补的碱基,可实现对基因组DNA片段的单位点编辑。所述基因组DNA片段单位点编辑方法可以改变单位点编辑时碱基突变的特征。

[0069] 所述补平连接是指:所述突出断裂末端会先通过碱基互补配对加入与突出的末端互补的碱基补平为钝末端之后再连接。

[0070] 如本发明的一些实施方式中所列举的,Cas9核酸酶K918A在sgRNA1的介导下,对基因组DNA片段(β -globin RE2位点)进行切割时,所产生的突出断裂末端U4,在细胞修复系统的作用下,所述突出断裂末端U4会先通过碱基互补配对加入与突出的末端C互补的碱基G补平为钝末端后再与连接接头连接。

[0071] Cas9核酸酶K918A在sgRNA2的介导下,对基因组DNA片段(β -globin RE2位点)进行切割时,所产生的突出断裂末端D4,在细胞修复系统的作用下,所述突出断裂末端D4会先通过碱基互补配对加入与突出的末端A互补的碱基T补平为钝末端后再与连接接头连接。

[0072] 说明:

[0073] 在本发明中,Cas9可作为Cas9核酸酶的简称使用,意思与Cas9核酸酶相同。在本发明中,Cas9-K918A、K918A、K918A突变体之间可替换使用,意思均为名称为K918A的Cas9核酸酶。

[0074] 在进一步描述本发明具体实施方式之前,应理解,本发明的保护范围不局限于下述特定的具体实施方案;还应当理解,本发明实施例中使用的术语是为了描述特定的具体实施方案,而不是为了限制本发明的保护范围。下列实施例中未注明具体条件的试验方法,通常按照常规条件,或者按照各制造商所建议的条件。

[0075] 当实施例给出数值范围时,应理解,除非本发明另有说明,每个数值范围的两个端点以及两个端点之间任何一个数值均可选用。除非另外定义,本发明中使用的所有技术和科学术语与本技术领域技术人员通常理解的意义相同。除实施例中使用的具体方法、设备、材料外,根据本技术领域的技术人员对现有技术的掌握及本发明的记载,还可以使用与本发明实施例中所述的方法、设备、材料相似或等同的现有技术的任何方法、设备和材料来实现本发明。

[0076] 除非另外说明,本发明中所公开的实验方法、检测方法、制备方法均采用本技术领域常规的分子生物学、生物化学、染色质结构和分析、分析化学、细胞培养、重组DNA技术及相关领域的常规技术。这些技术在现有文献中已有完善说明,具体可参见Sambrook等MOLECULAR CLONING:A LABORATORY MANUAL,Second edition,Cold Spring Harbor Laboratory Press,1989and Third edition,2001;Ausubel等,CURRENT PROTOCOLS IN MOLECULAR BIOLOGY,John Wiley&Sons,New York,1987and periodic updates;the series METHODS IN ENZYMOLOGY,Academic Press,San Diego;Wolffe,CHROMATIN STRUCTURE AND FUNCTION,Third edition,Academic Press,San Diego,1998;METHODS IN ENZYMOLOGY,Vol.304,Chromatin(P.M.Wassarman and A.P.Wolffe,eds.),Academic Press,San Diego,1999;和METHODS IN MOLECULAR BIOLOGY,Vol.119,Chromatin

Protocols (P.B.Becker, ed.) Humana Press, Totowa, 1999等。

[0077] 实施例1研究DNA片段编辑接头的连接情况发现Cas9切割新机制

[0078] 针对HS51位点, 构建针对HS51位点的sgRNAs质粒:

[0079] (1) 购买引物

[0080] 从上海桑尼生物科技有限公司购买分别针对HS51位点和的sgRNAs靶向序列的有5' 悬挂端“ACCG”和“AAAC”可以互补配对的正反向脱氧寡核苷酸;

[0081] 针对上述HS51位点的sgRNAs靶向序列:

[0082] HS51 RE1sgRNA1:GCCACACATCCAAGGCTGAC (SEQ ID NO.1)

[0083] HS51 RE1sgRNA2:GAGATTTGGGGCGTCAGGAAG (SEQ ID NO.2)

[0084] (2) 获得互补配对的带有悬挂端的双链DNA

[0085] 1) 用ddH₂O将脱氧寡核苷酸溶解至100μM, 并稀释至20μM;

[0086] 2) 将正反脱氧寡核苷酸加入如下反应体系:

正脱氧寡核苷酸 2 μl

反脱氧寡核苷酸 2 μl

[0087] NEB buffer2 (10x) 2 μl

ddH₂O 14 μl

Total volume 20 μl

[0088] 反应条件: 95℃水浴, 5min, 然后打开水浴锅盖子温度降至60℃左右, 盖上盖子冷却至室温。

[0089] (3) 酶切pGL3-U6-sgRNA-PGK-Puro vector

[0090] 1) 用BsaI限制性内切酶酶切载体质粒, 反应体系如下:

NEB buffer4 (10x) 2 μl

BSA (10x) 2 μl

质粒 ~700 ng

[0091] BsaI 1 μl

ddH₂O 12 μl

Total volume 20 μl

[0092] 反应条件: 37℃, 1.5小时;

[0093] 2) 胶回收纯化DNA酶切片段, 按照胶回收试剂盒 (Axygen) 说明纯化。

[0094] (4) 连接酶切后的载体与带有悬挂端的双链DNA

[0095] 连接体系如下:

- | | | |
|--------|----------------------------|--------------|
| | 带有悬挂端的双链 DNA | 0.5 μ l |
| | BsaI 酶切载体 | 0.5 μ l |
| [0096] | 10x NEB T4 ligation buffer | 0.5 μ l |
| | NEB T4 ligase | 0.25 μ l |
| | ddH ₂ O | 3.25 μ l |
| | Total volume | 5 μ l |
- [0097] 反应条件:室温反应1.5小时;
- [0098] (5) 转化连接产物
- [0099] 用Stb13感受态转化连接产物,在含氨苄抗生素 (Amp, 100mg/L) LB平板培养过夜, 37°C。
- [0100] (6) 挑取单克隆测序
- [0101] 1) 从氨苄抗生素LB平板上挑取单菌落, LB (Amp, 100mg/L) 液体培养过夜;
- [0102] 2) 质粒提取,按照质粒小抽试剂盒 (Axygen) 说明提取;
- [0103] 3) 提取后的质粒送上海桑尼生物科技有限公司测序。
- [0104] (7) 测序成功质粒进行中抽
- [0105] 1) 测序成功的质粒用Stb13感受态重新转化,在含Amp (100mg/L) 的LB平板培养过夜;
- [0106] 2) 上午挑取单菌落在2ml LB (Amp, 100mg/L) 液体培养基中培养8小时,然后转接到200ml LB (Amp, 100mg/L) 液体培养基中培养过夜;
- [0107] 3) 收集细菌,按照质粒中抽试剂盒 (Qiagen) 说明提取质粒。
- [0108] 2. 人源化Cas9质粒制备
- [0109] 1) 人源化Cas9质粒从北京大学席建中实验室获得;
- [0110] 2) 用Stb13感受态重新转化,在LB平板 (Amp, 100mg/L) 培养过夜;
- [0111] 3) 上午挑取单菌落在2ml LB (Amp, 100mg/L) 液体培养基中培养8小时,然后转接到200ml LB (Amp, 100mg/L) 液体培养基中培养过夜,进行质粒中抽。
- [0112] 3. 用Lipofectamine 2000进行细胞转染
- [0113] 1) HEK293T细胞培养在培养瓶中,在37°C,含有5%CO₂细胞培养箱中培养,待其长至培养瓶80~90%。
- [0114] 2) 将长好的细胞在12孔板中用DMEM完全无抗培养基 (加入10%胎牛血清,无青链霉素双抗) 进行铺板,过夜培养。
- [0115] 3) 待12孔板中的细胞长至80~90%时,将准备好的人源化Cas9质粒 (800ng) 和针对HS51位点的sgRNAs质粒 (各600ng) 通过Lipofectamine 2000进行细胞转染,每个样品各两个重复。
- [0116] 4) 转染后两天,收集细胞,用基因组提取试剂盒 (Wizard[®] Genomic DNA Purification kit, Promega) 提取基因组。
- [0117] 4. 制备高通量测序文库
- [0118] 在DNA片段预期删除、反转和重复接头的精准连接位点上游大约30bp处设计引物,

然后将引物5'端加上带有barcode的Illumina的测序接头,下游引物可以设计在远离拼接位点一些的位置并加上Illumina的测序接头,进行PCR扩增,然后使用罗氏PCR纯化试剂盒(Product No.:11732676001)进行纯化,DNA产物溶解在10mM Tris-HCL buffer (PH=8.5),等量混合后形成库,进行高通量测序。

[0119] 高通量引物:

[0120] HiSeq-hHs51-aF:

[0121]

ATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGCAAGGAGATCCGTGTCTGTC
(SEQ ID NO.3)

[0122] HiSeq-hs51-aRa:

[0123]

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTAAGGATGTTGTGGAAGGCG
AGCAG (SEQ ID NO.4)

[0124] HiSeq-hs51-bFa:

[0125]

CAAGCAGAAGACGGCATAACGAGATGGACGGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTTTACATGACAG
CTTCCGGTAG (SEQ ID NO.5)

[0126] HiSeq-hHs51-bR:

[0127]

CAAGCAGAAGACGGCATAACGAGATTTGACTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTTTTGGCTAACA
ACATAGTGCTTC (SEQ ID NO.6)。

[0128] 5.高通量测序数据处理

[0129] 高通量测序完成后,使用Linux程序将样品的测序结果从文库中通过barcode分出来,保存在各自的文件夹,然后进行BWA-MEM比对,比对后的序列通过Varscan2程序(V2.3.9)分析DNA片段的插入和删除突变,Varscan2程序参数如下:

Min coverage: 8

Min reads2: 2

[0130] Min var freq: 0.01

Min avg qual: 15

P-value thresh: 0.01。

[0131] 本发明通过研究DNA片段编辑的末端连接情况发现Cas9切割新机制。

[0132] 如图1A所示,采用两个sgRNAs形成的sgRNA组合及Cas9核酸酶对基因组DNA片段进行编辑时,Cas9核酸酶在两个sgRNAs介导下对基因组DNA双链进行切割产生四个断裂末端(DSB),这些断裂末端(DSB)在细胞修复系统(例如,MRN/CtIP)的作用下产生DNA片段删除、反转和重复等DNA片段编辑。

[0133] 如图1B所示,针对基因组DNA片段HS51RE1(HS51位点),我们采用sgRNA1和sgRNA2形成的sgRNA组合及Cas9核酸酶对其进行编辑。而后,我们检测到了DNA片段删除、反转和重复,再利用高通量测序技术检测DNA片段删除、反转和重复连接接头的情况,除了与预期相

符的精准连接 (Joined precisely) 外, DNA 片段删除连接接头、反转下游连接接头和重复连接接头处都存在一定比例的碱基加入 (Insertion)。

[0134] 如图1C所示, 利用高通量测序技术检测DNA片段删除连接接头的情况, 与预期相符的精准连接 (Joined precisely) 比例占79.23%, 删除接头处还存在“G”碱基的加入 (Insertion, 与预期的精准连接相比), 其比例占11.13%。

[0135] 与预期的精准连接相比, 推测DNA片段删除连接接头处加入的“G”碱基是来源于模版DNA (HS51 RE1, HS51位点) 的PAM上游3bp附近 (具体为PAM上游4bp处) 的碱基。因此, 推测Cas9核酸酶对与sgRNA互补的DNA链进行切割时, 是在PAM上游3bp处进行切割; 而Cas9核酸酶对与sgRNA非互补的DNA链进行切割时, 可在PAM上游3bp处更远的4bp处进行切割。根据DNA片段删除连接接头处存在“G”碱基的加入 (与预期的精准连接相比), 推测Cas9核酸酶在sgRNA2介导下对基因组DNA片段进行切割时, 有钝末端切割和突出末端切割, 进而产生不同断裂末端。当Cas9核酸酶在sgRNA2介导下对基因组DNA片段进行了钝末端切割时, 也就是Cas9核酸酶对与sgRNA互补的DNA链及非互补的DNA链进行切割时均是在PAM上游3bp处进行切割, 产生了钝断裂末端“E3”。钝断裂末端“E3”在细胞修复系统的作用下产生DNA片段删除时, 不会导致DNA片段删除连接接头处“G”碱基的加入, 而是产生与预期相符的精准连接 (Joined precisely)。当Cas9核酸酶在sgRNA2介导下对基因组DNA片段进行了突出末端切割时, 也就是Cas9核酸酶对与sgRNA互补的DNA链进行切割时是在PAM上游3bp处进行切割, 而对与sgRNA非互补的DNA链进行切割时是在PAM上游4bp处进行切割, 从而产生了5' 突出断裂末端“E4”。5' 突出断裂末端“E4”在细胞修复系统的作用下产生DNA片段删除时, 会导致DNA片段删除连接接头处“G”碱基的加入。

[0136] 因此, 我们认为: 在Cas9核酸酶的切割下, 产生的断裂末端中, 钝断裂末端E3的比例=预期相符的精准连接 (Joined precisely) 的比例=79.23%。突出断裂末端E4的比例=“G”碱基的加入比例=11.13%。

[0137] 但是, 我们观察到, 除了与预期相符的精准连接 (Joined precisely) 以及DNA片段删除连接接头处存在“G”碱基的加入这两大类情况以外, 还有一类随机的碱基删除 (Small deletion)。我们认为这类随机的碱基删除 (Small deletion) 是各断裂末端 (钝断裂末端E3和突出断裂末端E4) 在细胞修复系统的作用下随机产生的, 各断裂末端以均等的概率来产生碱基删除 (Small deletion), 各断裂末端在细胞修复系统的作用下所产生的碱基删除 (Small deletion) 的数量与各断裂末端的数量成正比。

[0138] 基于随机碱基删除现象的存在, 我们认为, 经过测序获得的各断裂末端的实测比例与其真实比例存在差距, 需要进行修正还原, 即以各种断裂末端的实测比例之和为基准, 计算各断裂末端的比例, 以此作为该断裂末端的占比。即对Cas9核酸酶的切割所产生的各断裂末端的比例进行标准化计算, 钝断裂末端E3的比例为87.7%【计算方法为: $79.23\% \div (79.23\% + 11.13\%)$ 】。突出断裂末端E4的比例为12.3%【计算方法为: $11.13\% \div (79.23\% + 11.13\%)$ 】。亦即, Cas9核酸酶在sgRNA2的介导下对基因组DNA片段进行的切割方式中, 钝末端切割的比例为87.7%, 突出末端切割的比例为12.3%。

[0139] 如图1D所示, 利用高通量测序技术检测DNA片段重复连接接头的情况, 与预期相符的精准连接 (Joined precisely) 的比例占8.96%, 连接接头处存在“T”碱基的加入 (Insertion, 与预期的精准连接相比) 的比例占82.92%。

[0140] 与预期的精准连接相比,推测DNA片段重复连接接头处加入的“T”碱基是来源于模版DNA(HS51RE1,HS51位点)上的PAM上游3bp附近(具体为PAM上游4bp处)的碱基。因此,推测Cas9核酸酶对与sgRNA互补的DNA链进行切割时,是在PAM上游3bp处进行切割;而Cas9核酸酶对与sgRNA非互补的DNA链进行切割时,可在PAM上游3bp处更远的4bp处进行切割。根据DNA片段重复连接接头处检测到存在“T”碱基的加入(与预期的精准连接相比),推测Cas9核酸酶在sgRNA1介导下对基因组DNA片段进行切割时,有钝末端切割和突出末端切割,进而产生不同断裂末端。当Cas9核酸酶在sgRNA1介导下对基因组DNA片段进行了钝末端切割时,也就是Cas9核酸酶对与sgRNA互补的DNA链及非互补的DNA链进行切割时均是在PAM上游3bp处进行切割,产生了钝断裂末端“C3”。钝断裂末端“C3”在细胞修复系统的作用下产生DNA片段重复时,不会导致DNA片段重复连接接头处“T”碱基的加入,而是产生与预期相符的精准连接(Joined precisely)。当Cas9核酸酶在sgRNA1介导下对基因组DNA片段进行了突出末端切割时,也就是Cas9核酸酶对与sgRNA互补的DNA链进行切割时是在PAM上游3bp处进行切割,而对与sgRNA非互补的DNA链进行切割时是在PAM上游4bp处进行切割,从而产生了5'突出断裂末端“C4”。5'突出断裂末端“C4”在细胞修复系统的作用下产生DNA片段重复时,会导致DNA片段重复连接接头处“T”碱基的加入。

[0141] 因此,我们认为:在Cas9核酸酶的切割下,产生的断裂末端中,钝断裂末端C3的比例=预期相符的精准连接(Joined precisely)的比例=8.96%。突出断裂末端C4的比例=“T”碱基的加入比例=82.92%。

[0142] 但是,我们观察到,除了与预期相符的精准连接(Joined precisely)以及DNA片段重复连接接头处存在“T”碱基的加入这两大类情况以外,还有一类随机的碱基删除(Small deletion)。我们认为这类随机的碱基删除(Small deletion)是各断裂末端(钝断裂末端C3和突出断裂末端C4)在细胞修复系统的作用下随机产生的,各断裂末端以均等的概率来产生碱基删除(Small deletion),各断裂末端在细胞修复系统的作用下所产生的碱基删除(Small deletion)的数量与各断裂末端的数量成正比。

[0143] 基于随机碱基删除现象的存在,我们认为,经过测序获得的各断裂末端的实测比例与其真实比例存在差距,需要进行修正还原,即以各种断裂末端的实测比例之和为基准,计算各断裂末端的比例,以此作为该断裂末端的占比。即对Cas9核酸酶的切割所产生的各断裂末端的比例进行标准化计算,钝断裂末端C3的比例为9.75%【计算方法为:8.96%÷(8.96%+82.92%)】。突出断裂末端C4的比例为90.25%【计算方法为:82.92%÷(8.96%+82.92%)】。亦即,Cas9核酸酶在sgRNA1的介导下对基因组DNA片段进行的切割方式中,钝末端切割的比例为9.75%,突出末端切割的比例为90.25%。

[0144] 如图1E所示,根据Cas9核酸酶在sgRNA1和sgRNA2的介导下分别对基因组DNA片段进行切割的方式比例,预测产生的断裂末端的序列,进而推算出DNA片段反转下游连接接头处的碱基加入情况及比例。

[0145] 当Cas9核酸酶在sgRNA1的介导下对基因组DNA片段进行突出末端切割,产生突出断裂末端“C4”,Cas9核酸酶在sgRNA2的介导下对基因组DNA片段进行钝末端切割,产生钝断裂末端“E3”,则在细胞修复系统的作用下,DNA片段反转下游接头处会出现“A”碱基的加入,且发生的比例为79.14%【计算方法为:“C4”突出断裂末端占比(90.25%)×“E3”钝断裂末端占比(87.7%)=79.14%】,与实验检测到的DNA片段反转下游接头处“A”碱基加入比例

71.94%相近。

[0146] 当Cas9核酸酶在sgRNA1的介导下对基因组DNA片段进行钝末端切割,产生钝断裂末端“C3”,Cas9核酸酶在sgRNA2的介导下对基因组DNA片段进行突出末端切割,产生突出断裂末端“E4”,则在细胞修复系统的作用下,DNA片段反转下游接头处会出现“G”碱基的加入,且发生的比例为1.19%【计算方法为:“C3”钝断裂末端占比(9.75%)x“E4”突出断裂末端占比(12.3%)=1.19%】,与实验检测到的DNA片段反转下游接头处“G”碱基加入比例8.54%相近。

[0147] 当Cas9核酸酶在sgRNA1的介导下对基因组DNA片段进行突出末端切割,产生突出断裂末端“C4”,Cas9核酸酶在sgRNA2的介导下对基因组DNA片段进行突出末端切割,产生突出断裂末端“E4”,则在细胞修复系统的作用下,DNA片段反转下游接头处会出现“AG”碱基的加入,且发生的比例为11%【计算方法为:“C4”突出断裂末端占比(90.25%)x“E4”突出断裂末端占比(12.3%)=11%】,与实验检测到的DNA片段反转下游接头处“AG”碱基加入比例3.66%相近。

[0148] 当Cas9核酸酶在sgRNA1的介导下对基因组DNA片段进行钝末端切割,产生钝断裂末端“C3”,Cas9核酸酶在sgRNA2的介导下对基因组DNA片段进行钝末端切割,产生钝断裂末端“E3”,则在细胞修复系统的作用下,DNA片段反转下游接头精准连接,且发生的比例为8.55%【计算方法为:“C3”钝断裂末端占比(9.75%)x“E3”钝断裂末端占比(87.7%)=8.55%】,与实验检测到的DNA片段反转下游接头精准连接比例6.67%相近。

[0149] 综上所述,图1E的实验结果进一步证实了:Cas9核酸酶对与sgRNA非互补的DNA链进行切割时,可在PAM上游3bp处到更远碱基处进行切割。Cas9核酸酶在sgRNA介导下对基因组DNA片段进行切割时,有钝末端切割和突出末端切割,进而产生不同断裂末端。这些断裂末端在细胞修复系统的作用下,产生与预期相符的精准DNA片段编辑(特定碱基的精准编辑)或者与预期不符的基因编辑(随机的碱基删除)。

[0150] 如图1F所示,sgRNA组合中,sgRNA的设计不同(靶序列不同),Cas9核酸酶在sgRNA的介导下对基因组DNA片段进行切割方式比例不同,产生的断裂末端比例不同。具体地,Cas9核酸酶在sgRNA1的介导下对基因组DNA片段进行切割时,钝末端切割方式的占比高于突出末端切割方式占比,产生的钝断裂末端占比高于5'突出断裂末端占比。然而Cas9核酸酶在sgRNA2的介导下对基因组DNA片段进行切割时,突出末端切割方式的占比高于钝末端切割方式占比,产生的5'突出断裂末端占比也高于钝断裂末端占比。

[0151] 由于发现Cas9核酸酶在sgRNA介导下对基因组DNA片段进行切割的方式有钝末端切割和突出末端切割,当Cas9核酸酶在sgRNA介导下对基因组DNA片段进行突出末端切割,产生突出断裂末端时,按照补平连接的方式可加入与突出断裂末端互补的碱基,从而实现基因组DNA片段特定位置的碱基加入。

[0152] 实施例2突变SpCas9获得切割方式改变的特定Cas9实现精准的DNA片段编辑

[0153] 1. 构建Cas9突变体

[0154] 1) 使用NEB突变试剂盒(Q5Site-Directed Mutagenesis Kit,#E0554S)构建Cas9突变体,首先进行PCR扩增,反应如下:

	Q5 Hot Start High-Fidelity 2X Master Mix	10 μ l
	10 μ M Forward Primer	1 μ l
	10 μ M Reverse Primer	1 μ l
[0155]	Cas9 plasmid (10 ng)	0.38 μ l
	Nuclease-free water	7.62 μ l
	Total volume	20 μ l

反应条件: 98°C, 30 sec;

	98°C, 10sec	} 25 个循环
[0156]	60°C, 20sec	
	72°C, 6.5min	
	72°C, 3min	

[0157] 2) KLD (Kinase, Ligase&DpnI) 处理, 反应如下:

	PCR 产物	0.5 μ l
	2X KLD Reaction Buffer	2.5 μ l
[0158]	10X KLD Enzyme Mix	0.5 μ l
	Nuclease-free Water	1.5 μ l
	Total volume	5 μ l

[0159] 反应条件: 室温10分钟

[0160] 3) 将2) 中的反应产物全部用于感受态细菌Stb13 (50 μ l) 的转化, 在含氨苄抗生素 (Amp, 100mg/L) LB平板培养过夜, 37°C。挑取单克隆, 质粒提取后送测序。

[0161] SpCas9 (Cas9WT) 的氨基酸序列如SEQ ID NO.7所示, 具体为:

```

MAPKKKRKVG IHGVPAAMDK KYSIGLDIGT NSVGWAVITD EYKVPSKKEK VLGNTDRHSI
KKNLIGALLF DSGETAETR LKRTARRRYT RRKNRICYLQ EIFSNEMAKV DDSFFHRLEE
SFLVEEDKKH ERHPIFGNIV DEVAYHEKYP TIYHLRKKLV DSTDKADLRL IYLALAHMIK
FRGHFLIEGD LNPDNSDVDK LFIQLVQTYN QLFEEPNINA SGVDAKAILS ARLSKSRRL
[0162] NLIAQLPGEK KNGLFGNLIA LSLGLTPNFK SNFDLAEDAK LQLSKDITYDD DLNLLAQIG
DQYADLFLAA KNLSDAILLS DILRVNTEIT KAPLSASMIK RYDEHHQDLT LLKALVRQQL
PEKYKEIFFD QSKNGYAGYI DGGASQEEFY KFIKPILEKM DGTEELLVKL NREDLLRKQR
TFDNGSIPHQ IHLGELHAIL RRQEDFYFPL KDNREKIEKI LTFRIPYYVG PLARGNSRFA
WMTRKSEETI TPWNFEEVVD KGASAQSFIE RMTNFDKNLP NEKVLPKHSL LYEYFTVYNE

```

LTKVKYVTEG MRKPAFLSGE QKKAIVDLLF KTNRKVTVKQ LKEDYFKKIE CFDSVEISGV
 EDRFNASLGT YHDLLKIIKD KDFLDNEENE DILEDIVLTL TLFEDREMIE ERLKTYAHLF
 DDKVMKQLKR RRYTGWGRLS RKLINGIRDK QSGKTILDFL KSDGFANRNF MQLIHDDSLT
 FKEDIQKAQV SGQGDSLHEH IANLAGSPAI KKGILQTVKV VDELVKVMGR HKPENIVIEM
 ARENQTTQKG QKNSRERMKR IEEDIKELGS QILKEHPVEN TQLQNEKLYL YYLQNGRDMY
 VDQELDINRL SDYDVDHIVP QSFLKDDSID NKVLTRSDKN RGKSDNVPSE EVVKMKMKNYW
 RQLLNAKLIT QRKFDNLTKA ERGGLSELDK AGFIKRQLVE TRQITKHVAQ ILDSRMNTKY
 [0163] DENDKLIREV KVITLKSCLV SDFRKDFQFY KVVREINNYHH AHDAYLNAVV GTALIKKYPK
 LESEFVYGDY KVDVDRKMLA KSEQEIGKAT AKYFFYSNIM NFFKTEITLA NGEIRKRPLI
 ETNGETGEIV WDKGRDFATV RKLVSMPQVN IVKKTEVQTG GFSKESILPK RNSDKLIARK
 KDWDPKKYGG FDSPTVAYSV LVVAKVEKGG SKKLKSVKEL LGITIMERSS FEKNPIDFLE
 AKGYKEVKKD LIIKLPKYSL FELENGRKRK LASAGELQKG NELALPSKYV NFLYLASHYE
 KLKGSPEDE QKQLFVEQHK HYLDEIEQI SEFSKRVILA DANLDKVLISA YNKHRDKPIR
 EQAENIIHLF TLTNLGAPAA FKYEDTTIDR KRYTSTKEVL DATLIHQSIIT GLYETRIDL
 QLGGDKRPAA TKKAGQAKKKK。

[0164] SpCas9 (Cas9WT) 的编码核苷酸序列如SEQ ID NO.8所示, 具体为:

atggccccaa agaagaagcg gaaggteggt atccacgggtg tcccagcagc catggacaag
 aagtaactcca ttgggctcga tatcggcaca aacagcgctcg gctgggcccgt cattacggac
 gagtacaagg tgccgagcaa aaaattcaaa gttctgggca ataccgateg ccacagcata
 aagaagaacc tcattggcgc cctcctgttc gactccgggg agaeggccga agccacgcgg
 ctcaaaaagaa cagcacggcg cagatatacc cgcagaaaaga atcgggatctg ctacctgcag
 gagatcttta gtaatgagat ggctaagggtg gatgactctt tcttccatag gctggaggag
 tcttttttg tggaggagga taaaaagcac gagcgcacc caatcttttg caatctctg
 [0165] gacgagggtg cgtaccatga aaagtaccca accatatabc atctgaggaa gaagcttcta
 gacagtactg ataaggctga cttggcgttg atctatctcg cgtggcgcga tatgatcaaa
 tttcggggac acttctctat cgagggggac ctgaaccag acaacagcga tctcgacaaa
 ctcttttatcc aactgggtca gacttacaat cagctttctg aagagaacce gatcaacgca
 tccggagttg acgcccaga aatcctgagc gctaggctgt ccaaatcccg gcggctcgaa
 aacotcatcg cacagctccc tggggagaag aagaacggcc tgtttggtaa tottatcgcc
 ctgtcaactcg ggctgaaccc caactttaa totaactctg acctggccga agatgccaag

[0166]

cttcaactga gcaaagacac ctacgatgat gatctcgaca atctgctggc ccagatcggc
gaccagtaacg cagacctttt tttggcgga aagaacctgt cagacgccat totgctgagt
gatattctgc gagtgaacac ggagatcacc aaagctccgc tgagcgctag tatgatcaag
cgctatgatg agcaccacca agacttgact ttgctgaagg cccttgctcag acagcaactg
cctgagaagt acaaggaaat tttcttcgat cagtctaaaa atggctaccg cggatacatt
gacggcggag caagccagga ggaattttac aaatttatta agccatctt ggaaaaaatg
gacggcacog aggagctgct ggtaaagctt aacagagaag atctggtgcg caaacagcgc
actttcgaca atggaagcat cccccaccag attcacctgg gcgaaactgca cgtataactc
aggcggcaag aggatttota cccctttttg aaagataaca gggaaaagat tgagaaaatc
ctcacatttc ggatacccta ctatgtagge cccctcgccc ggggaaattc cagattcgcg
tggtatgactc gcaaatacaga agagaccate actccctgga acttcgagga agtcgtggat
aagggggcct ctgccagtc cttcatcgaa aggatgacta actttgataa aaatctgct
aacgaaaagg tgcctcctaa acactctctg ctgtacgagt acttcacagt ttataacgag
ctcaccaagg tcaaatacgt cacagaaggg atgagaaaag cagcattcct gtctggagag
cagaagaaag ctatogtga cctcctcttc aagacgaacc ggaaagtac cgtgaaacag
ctcaaagaag actatttcaa aaagattgaa tgtttcgaet ctgttgaaat cagcggagtg
gaggatogct tcaaagcact cctgggaacg tatcaagatc cctgaaaaat cattaaagac
aaggacttcc tggacaatga ggagaacgag gacattcttg aggacattgt cctcaccctt
acgttgtttg aagatagggg gatgattgaa gaacgottga aaacttacgc toatctcttc
gacgacaaag tcatgaaaca gctcaagagg cgcgatata caggatgggg gcggctgtca
agaaaactga tcaatgggat ccgagacaag cagagtggaa agacaatcct ggattttctt
aagtcogatg gatttgocaa ccggaacttc atgcagttga tccatgatga ctctctcacc
tttaaggagg acatccagaa agcacaagtt tctggccagg gggacagtct toacgagcac
atcgttaate ttgcaggtag cccagctatc aaaaagggaa tactgcagac cgttaaggtc
gtggatgaac tctcaaaagt aatgggaagg cataagccc agaatatcgt tatcgagatg
gcccagagaga accaaaactac ccagaaggga cagaagaaca gtagggaaag gatgaagagg
attgaagagg gtataaaaga actggggctc caaatcetta aggaacaccc agttgaaaac
accagcttc agaatgagaa gctctacctg tactacctgc agaacggcag ggacatgtac
gtggatcagg aactggacat caatcggctc tccgactacg acgtggatca tatcgtgcc
cagctctttc tcaaagatga ttctattgat aataaagtg tgaacaagatc cgataaaaaat

agaggggaaga gtgataacgt ccctcagaa gaagttgtca agaaaatgaa aaattattgg
 eggcagctgc tgaacgocaa actgatcaca caacgggaagt tcgataatct gactaaggct
 gaacgaggtg gcctgtctga gttggataaa gcaggcttca tcaaaaggca gcttgttgag
 acacgccaga tcaccaagca cgtggcccaa attctcgatt cacgcatgaa caaccaagtac
 gatgaaaatg acaaactgat tcgagaggtg aaagttatta ctctgaagtc taagctggtc
 tcagatttca gaaaggactt tcagttttat aagggtgagag agatcaacaa ttaccaccat
 gcgcatgatg cctacctgaa tgcagtggtg ggcactgcac ttatcaaaaa atatcccaag
 cttgaatctg aatttgttta cggagactat aaagtgtacg atgttaggaa aatgatcgca
 aagtctgagc aggaaatagg caaggccacc gctaagtact tcttttacag caatattatg
 aattttttca agaccgagat tacactggcc aatggagaga ttcggaagcg accacttato
 gaaacaaaacg gagaaacagg agaaatcgtg tgggacaagg gtagggattt cgcgacagtc
 cgggaaggtec tgtccatgcc gcaggtgaac atcgttaaaa agaccgaagt acagaccgga
 ggctttotcca aggaaagtat cctcccgaaa aggaacagcg acaagctgat cgcacgcaaa
 [0167] aaagattggg accccaagaa ataocggcga ttcgattctc ctacagtcgc ttacagtgta
 ctggttgtgg ccaaagtgga gaaagggag tctaaaaaac tcaaaagcgt caaggaactg
 ctgggcatca caatcatgga gcgatcaage ttcgaaaaaa aceccatcga ctttctcgag
 gcgaaaggat ataaagaggt caaaaaagac ctcateatta agcttcccaa gtactctctc
 tttgagcttg aaaacggccg gaaacgaatg ctogctagtg cgggcgagct gcagaaaggt
 aacgagctgg cactgccctc taaatcgtt aatttcttgt atctggccag ccactatgaa
 aagctcaaag ggtctccoga agataatgag cagaagcagc tgttcgtgga acaacacaaa
 cactaccttg atgagateat cgagcaaata agcgaattct ccaaaagagt gatcctcgcc
 gacgctaacc tcgataaggt gctttctgct tacaataage acagggataa gcccatcagg
 gacgaggcag aaaacattat coacttgttt actctgacca acttgggcgc gcctgcagcc
 ttcaagtact tcgacaccac catagacaga aagcggtaaa cctctacaaa ggaggtcctg
 gacgccacac tgattcatca gtcaattacg gggctctatg aaacaagaat cgacctctc
 cagctcggtg gagacaagcg tcctgctgct actaagaaag ctggccaagc taagaaaaag
 aaataa.

[0168] 如图2A所示, Cas9核酸酶, 含有RuvC和HNH功能域, RuvC功能域负责切割与sgRNA非互补的DNA链, HNH功能域负责切割与sgRNA互补的DNA链。

[0169] 本发明要求保护的Cas9核酸酶突变体命名为Cas9-K918A (将SpCas9核酸酶第918位赖氨酸突变成丙氨酸获得),

[0170] Cas9-K918A的氨基酸序列SEQ ID NO.9所示,具体为:

MAPKKKRKVG IHGVPAAMDK KYSIGLDIGT NSVGVAVITD EYKVPSKKFK VLGNTDRHSI
 KKNLIGALLF DSGETAEATR LKRTARRRYT RRKNRICYLQ EIFSNEMAKV DDSFFHRLEE
 SFLVEEDKKH ERHPIFGNIV DEVAYHEKYP TIYHLRKKLV DSTDKADLRL IYLALAHMIK
 FRGHFLIEGD LNPDNSDVVK LFIQLVQTYN QLFEENPINA SGVDAKAILS ARLSKSRRLE
 NLIAQLPGEK KNGFLGNLIA LSLGLTPNFK SNFDLAEDAK LQLSKDITYDD DLDNLLAQIG
 DQYADLFLAA KNLSDAILLS DILRVNTEIT KAPLSASMIK RYDEHHQDLT LLKALVRQQL
 PEKYKEIFFD QSKNGYAGYI DGGASQEEFY KFIKPILEKM DGTEELLVKL NREDLLRKQR
 TFDNGSIPHQ IHLGELHAIL RRQEDFYFPL KDNREKIEKI LTFRIPYYVG PLARGNSRFA
 WMTRKSEETI TPWNFEEVVD KGASAQSFIE RMTNFDKNLP NEKVLPKHSL LYEYFTVYNE
 LTKVKYVTEG MRKPAFLSGE QKKAIVDLLF KTNRKVTVKQ LKEDYFKKIE CFDSVEISGV
 EDRFNASLGT YHDLLKIIKD KDFLDNEENE DILEDIVLTL TLFEDREMIE ERLKTYAHLF
 DDKVMKQLKR RRYTGWRSL RKLINGIRDK QSGKTILDFL KSDGFANRNF MQLIHDDSLT

[0171]

FKEDIQKAQV SGQGDSLHEH IANLAGSPAI KKGILQTVKV VDELVKVMGR HKPENIVIEM
 ARENQTTQKG QKNSRERMKR IEEGIKELGS QILKEHPVEN TQLQNEKLYL YYLQNGRDMY
 VDQELDINRL SDYDVDHIVP QSFLKDDSID NKVLTRSDKN RGKSDNVPSE EVVKKMKNYW
 RQLLNAKLIT QRKFDNLTKA ERGGLSELDK AGFIARQLVE TRQITKHVAQ ILDSRMNTKY
 DENDKLIREV KVITLKSCLV SDFRKDFQFY KVREINNYHH AHDAYLNAVV GTALIKKYPK
 LESEFVYGDY KVDVVRKMAI KSEQEIGKAT AKYFFYSNIM NFFKTEITLA NGEIRKRPLI
 ETNGETGEIV WDKGRDEFATV RKLVSMPQVN IVKKTEVQTG GFSKESILPK RNSDKLIARK
 KDWDPKKYGG FDSPTVAYSV LVVAKVEK GK SKKLKSVKEL LGITIMERSS FEKNPIDFLE
 AKGYKEVKKD LIIKLPKYSL FELENGRKRK LASAGELQKG NELALPSKYV NFLYLASHYE
 KLKGSPEDE QKQLFVEQHK HYLDEIIEQI SEFSKRVILA DANLTKVLSA YNKHRDKPIR
 EQAENIIHLF TLTNLGAPAA FKYFDTTIDR KRYTSTKEVL DATLIHQISIT GLYETRIDL
 QLGGDKRPAA TKKAGQAKKK K。

[0172] Cas9-K918A的编码核苷酸序列如SEQ ID NO.10所示,具体为:

atggccccaa agaagaagcg gaaggtcggg atccacgggtg tcccagcagc catggacaag

[0173]

aagtactcca ttgggctcga tatcggcaca aacagcgtcg gctgggcccgt cattacggac
 gagtacaagg tgccgagcaa aaaattcaaa gttctgggca ataccgatcg ccacagcata

[0174]

aagaagaacc tcattggcgc cctcctgttc gaotccgggg agacggccga agccacgcgg
ctcaaaagaa cagcacggcg cagatatacc cgcagaaaga atcggatctg ctacctgcag
gagatcttta gtaatgagat ggctaagggtg gatgactctt tcttccatag gctggaggag
tcctttttgg tggaggagga taaaaagcac gagcgcacc caatctttgg caatatcgtg
gacgaggtgg cgtaccatga aaagtaccca accatatac atctgaggaa gaagcttgta
gacagtactg ataaggetga cttgcgggtg atctatctcg cgtggcgea tatgatcaaa
tttcggggac acttctcat cgagggggac ctgaaocccag acaacagcga tgtogacaaa
ctctttatcc aactggttca gacttacaat cagctttctg aagagaacc gatcaacgca
tcgggagttg acgcccaggc aatcctgagc gctaggctgt ccaaatcccg gggctcga
aacctcatcg cacagctccc tggggagaag aagaacggcc tgtttggtaa tcttatcgcc
ctgtcaactcg ggctgacccc caactttaaa totaacttgc acctggccga agatgccaa
cttcaactga gcaaagacac ctacgatgat gatctogaca atctgctggc ccagatcggc
gaccagtaog cagacctttt ttggcggca aagaacctgt cagacgccat tctgctgagt
gatattctgc gagtgaacac ggagatcacc aaagctccgc tgagcgtag tatgatcaag
cgtatgatg agcaccacca agacttgact ttgctgaagg ccttctgtag acagcaactg
cctgagaagt acaaggaaat tctcttcgat cagctaaaa atggctacgc cggatacatt
gacggcggag caagccaggg ggaattttac aaatttatta agccatctt ggaaaaaatg
gacggcaocg aggagctgct ggtaaagctt aacagagaag atctgctgag caaacagcgc
actttcgaca atggaagcat cccccaccag attcacttgg gcgaactgca cgtatactc
aggcggcaag aggatttcta cccctttttg aaagataaca gggaaaagat tgagaaaato
ctcacatttc ggatacccta ctatgtagge cccctcgecc ggggaaatto cagattcgcg
tggatgactc gcaaateaga agagaccate actccttggg acttcgagga agtegtgat
aagggggcct ctgccagtc ctctatcgaa aggatgacta accttgataa aaatctgcct
aacgaaaagg tgcttccctaa aactctctg ctgtaocgagt acctcacagt ttataacgag
ctcaccaagg tcaaatcgt cacagaaggg atgagaaage cagcattcct gtctggagag
cagaagaaag ctatcgtgga cctcctcttc aagacgaacc ggaaagttac cgtgaaacag
ctcaaaagag actatttcaa aaagattgaa tgtttcgtact ctggtgaaat cagcggagtg
gaggatcgt tcaacgcate cctgggaacg tatcaogate tctgaaaat cattaaagac
aaggacttcc tggacaatga ggagaacgag gacattcttg aggacattgt cctcaccctt
acgttgtttg aagataggga gatgattgaa gaaogcttga aaacttaocg tcatctctc

[0175]

gacgacaaaag tcatgaaaca gctcaagagg cgccgatata caggatgggg ggggctgtea
agaaaactga tcaatgggat ccgagacaag cagagtggaa agacaatcct ggattttott
aagtccgatg gatttgccaa ccggaacttc atgcagttga tccatgatga ctctctcacc
tttaaggagg acatccagaa agcacaagtt totggccagg gggacagtct tcacgagcac
ategetaate ttgcaggtag ccagctate aaaaagggaa tactgcagac cgttaaggtc
gtggatgaac togtcaaagt aatgggaagg cataagcccg agaatactgt tatcgagatg
gcccgagaga accaaactac ccagaagggg cagaagaaca gtagggaaag gatgaagagg
attgaagagg gtataaaaga actggggctc caaatcotta aggaacaccc agttgaaaac
accagcttc agaatgagaa gctctacctg tactacctgc agaacggcag ggacatgtac
gtggatcagg aactggacat caatcggctc tccgactacg acgtggatca tatcgtgcc
cagtccttttc tcaaagatga tctattgat aataaagtgt tgacaagatc cgataaaaat
agaggggaaga gtgataacgt cccctcagaa gaagttgtca agaaaatgaa aaattattgg
cggcagctgc tgaacgcocaa actgatcaca caacggaggt togataatct gactaaggct
gaacgagggtg gctgtctga gctggataaa gcaggcttca togccaggca gcttgttgag
acacgccaga tcaccaagea cgtggcccaa attctcgatt caccgatgaa caccaagtac
gatgaaaatg acaaaactgat togagagggt aaagttatta ctctgaagtc taagctggtc
tcagatttca gaaaggactt tcagttttat aagggtgagag agatcaacaa ttaccaccat
ggcctatgat cctacctgaa tgcagtggtg ggcactgcac ttatcaaaaa atatcccaag
cttgaatctg aatttgttta cggagactat aaagtgtacg atgttaggaa aatgatogca
aagtctgagc aggaaatagg caaggccacc gctaagtact tcttttacag caatattatg
aattttttca agaccgagat tacactggcc aatggagaga ttcggaagcg accacttate
gaaacaaacg gagaaacagg agaaatcgtg tgggacaagg gtagggattt cggacagtc
cggaaaggtc tgtccatgcc gcagggtgac atcgttaaaa agaccgaagt acagaccgga
ggcttctcca aggaaagtat cctcccgaaa aggaacagcg acaagctgat cgcacgcaaa
aaagattggg accccaagaa atacggcgga ttcgattctc ctacagtgc ttacagtgta
ctggttgtgg ccaaagtgga gaaagggag tctaaaaaac tcaaaagcgt caaggaactg
ctgggcatca caatcatgga gcgatcaagc ttcgaaaaaa accccatcga etttctcgag
gcgaaaggat ataaagaggt caaaaaagac ctcatcatta agcttcccaa gtactctctc
cttgagcttg aaaacggccg gaaacgaatg ctctctagtg cggcgagct gcagaaaggt
aacgagctgg cactgcctc taaatacgtt aatttcttgt atctggccag ccactatgaa

aagctcaaaag ggtctcccga agataatgag cagaagcagc tgttcgtgga acaacacaaa
 cactaccttg atgagatcat cgagcaaata agcgaattct ccaaagaggt gatcctcgcc
 gagcetaacc tcgataaggt gctttctgct tacaataagc acagggataa gcccatcagg
 gagcaggcag aaaacattat ccacttggtt actctgacca acttgggcgc gcctgcagcc
 [0176] ttcaagtact tcgacaccac catagacaga aagcggtaaa cctctacaaa ggagggtcctg
 gagcgcacac tgattcatca gtcaattacg gggctctatg aaacaagaat cgaactctct
 cagctcgggtg gagacaagcg tctgtctgct actaagaaaag ctgggtcaagc taagaaaaag
 aaataa.

[0177] 此外,以对SpCas9进行随机突变获得的突变体K775A、R778A、E779A、K918P作为对照,这些对照突变体与本发明的Cas9-K918A的序列均不同。

[0178] 2. Cas9核酸酶突变体进行DNA片段编辑

[0179] (1) 针对 β -globin RE2 (RRM21位点), 构建RRM21位点(β -globin RE2)的sgRNAs。

[0180] 所述sgRNAs靶向序列:

[0181] β -globin RE2 sgRNA1:ACCCAATGACCTCAGGCTGT (SEQ ID NO.11);

[0182] β -globin RE2 sgRNA2:TCACTTGTTAGCGGCATCTG (SEQ ID NO.12);

[0183] 从上海桑尼生物科技有限公司购买针对 β -globin RE2 (RRM21位点)的sgRNAs靶向序列的有5'悬挂端“ACCG”和“AAAC”可以互补配对的正反向脱氧寡核苷酸。

[0184] (2) 获得互补配对的带有悬挂端的双链DNA

[0185] 1) 用ddH₂O将脱氧寡核苷酸溶解至100 μ M,并稀释至20 μ M;

[0186] 2) 将正反脱氧寡核苷酸加入如下反应体系:

正脱氧寡核苷酸 2 μ l

反脱氧寡核苷酸 2 μ l

[0187] NEB buffer2 (10x) 2 μ l

ddH₂O 14 μ l

Total volume 20 μ l

[0188] 反应条件:95 $^{\circ}$ C水浴,5min,然后打开水浴锅盖子温度降至60 $^{\circ}$ C左右,盖上盖子冷却至室温。

[0189] (3) 酶切pGL3-U6-sgRNA-PGK-Puro vector

[0190] 1) 用BsaI限制性内切酶酶切载体质粒,反应体系如下:

[0191] NEB buffer4 (10x) 2 μ l

- | | | |
|--------|--------------------|------------|
| | BSA (10x) | 2 μ l |
| | 质粒 | ~700 ng |
| [0192] | BsaI | 1 μ l |
| | ddH ₂ O | 12 μ l |
| | Total volume | 20 μ l |
- [0193] 反应条件:37℃,1.5小时;
- [0194] 2) 胶回收纯化DNA酶切片段,按照胶回收试剂盒(Axygen)说明纯化。
- [0195] (4) 连接酶切后的载体与带有悬挂端的双链DNA
- [0196] 连接体系如下:
- | | | |
|--------|----------------------------|--------------|
| | 带有悬挂端的双链 DNA | 0.5 μ l |
| | BsaI 酶切载体 | 0.5 μ l |
| | 10x NEB T4 ligation buffer | 0.5 μ l |
| [0197] | NEB T4 ligase | 0.25 μ l |
| | ddH ₂ O | 3.25 μ l |
| | Total volume | 5 μ l |
- [0198] 反应条件:室温反应1.5小时;
- [0199] (5) 转化连接产物
- [0200] 用Stb13感受态转化连接产物,在含氨苄抗生素(Amp,100mg/L) LB平板培养过夜,37℃。
- [0201] (6) 挑取单克隆测序
- [0202] 1) 从氨苄抗生素LB平板上挑取单菌落, LB(Amp,100mg/L) 液体培养过夜;
- [0203] 2) 质粒提取,按照质粒小抽试剂盒(Axygen)说明提取;
- [0204] 3) 提取后的质粒送上海桑尼生物科技有限公司测序。
- [0205] (7) 测序成功质粒进行中抽
- [0206] 1) 测序成功的质粒用Stb13感受态重新转化,在含Amp(100mg/L)的LB平板培养过夜;
- [0207] 2) 上午挑取单菌落在2ml LB(Amp,100mg/L)液体培养基中培养8小时,然后转接到200ml LB(Amp,100mg/L)液体培养基中培养过夜;
- [0208] 3) 收集细菌,按照质粒中抽试剂盒(Qiagen)说明提取质粒。
- [0209] (8) 用Lipofectamine 2000进行细胞转染
- [0210] 1) HEK293T细胞培养在培养瓶中,在37℃,含有5%CO₂细胞培养箱中培养,待其长至培养瓶80~90%,将长好的细胞在12孔板中用DMEM完全无抗培养基进行铺板,过夜培养;
- [0211] 2) 待12孔板中的细胞长至80~90%时,将制备好的Cas9和Cas9突变体质粒(800ng)与针对RRM21位点的sgRNAs质粒(各600ng)通过Lipofectamine 2000进行细胞转染,每个样品各两个重复。
- [0212] 3) 转染后两天,收集细胞,用基因组提取试剂盒(Wizard[®] Genomic DNA

Purification kit, Promega) 提取基因组。

[0213] (9) 制备高通量测序文库

[0214] 在DNA片段预期删除、反转和重复接头的精准连接位点上游大约30bp处设计引物,然后将引物5'端加上带有barcode的Illumina的测序接头,下游引物可以设计在远离拼接位点一些的位置并加上Illumina的测序接头,进行PCR扩增,然后使用罗氏PCR纯化试剂盒(Product No.:11732676001)进行纯化,DNA产物溶解在10mM Tris-HCL buffer (PH=8.5),等量混合后形成库,进行高通量测序。

[0215] Cas9突变引物:

[0216] Cas9-K918A-F:AGGCTTCATCgccAGGCAGCTTG (SEQ ID NO.13);

[0217] Cas9-K918A-R:GCTTTATCCAACCTCAGAC (SEQ ID NO.14);

[0218] (10) 高通量测序数据处理

[0219] 高通量测序完成后,使用Linux程序将样品的测序结果从文库中通过barcode分出来,保存在各自的文件夹,然后进行BWA-MEM比对,比对后的序列通过Varscan2程序(V2.3.9)分析DNA片段的插入和删除突变,Varscan2程序参数如下:

Min coverage: 8

Min reads2: 2

[0220] Min var freq: 0.01

Min avg qual: 15

P-value thresh: 0.01。

[0221] 针对 β -globin RE2位点,利用高通量测序引物进行PCR扩增DNA片段删除、反转和重复,建库进行高通量测序。

[0222] 高通量引物:

[0223] Hi seq-RRM-1F3:

[0224]

AATGATACGGCGACCACCGAGATCTACTCTTTCCCTACACGACGCTCTCCGATCTATATGGCATCCTAGCCTTAGAACTAG (SEQ ID NO.15)

[0225] Hi seq-RRM-1R2:

[0226]

AATGATACGGCGACCACCGAGATCTACTCTTTCCCTACACGACGCTCTCCGATCTTACGACGCAGGAGCCGTATCATG (SEQ ID NO.16)

[0227] Hi seq-RRM-3F2:

[0228]

CAAGCAGAAGACGGCATAACGAGATAAGCTAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTATAGCAATGAAATCTTGAAGGAGTG (SEQ ID NO.17)

[0229] Hi seq-RRM-3R2:

[0230]

CAAGCAGAAGACGGCATAACGAGATTCAAGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCACAGCCCTGCTCTATTACG (SEQ ID NO.18)。

[0231] 参照上述实施例1的方法,采用两个sgRNAs形成的sgRNA组合及Cas9核酸酶对基因组DNA片段进行编辑后,可利用高通量测序技术检测DNA片段删除和重复的连接接头碱基加入情况,进而计算出Cas9核酸酶在各sgRNA介导下对基因组DNA片段进行切割时,钝末端切割方式和突出末端切割方式的占比。

[0232] 具体地,野生型SpCas9核酸酶(简称Cas9WT,WT)(图2A)和K918A在sgRNA组合中各sgRNA介导下对基因组DNA片段 β -globin RE2位点进行编辑的两个sgRNAs的示意图如图2B。

[0233] 如图2C所示,利用高通量测序技术检测DNA片段重复连接接头的情况,除了与预期相符的精准连接(Joined precisely)以外,还存在与预期的精准连接相比,连接接头处加入了“C”碱基和“GC”碱基的情况。选用不同的Cas9核酸酶时,检测到的与预期相符的精准连接(Joined precisely)、“+C”碱基、“+GC”碱基的占比不同。以选用K918A这个Cas9核酸酶为例,检测到与预期相符的精准连接(Joined precisely)的占比为86.28%，“+C”碱基的占比为2.25%,DNA片段重复连接接头处加入的“+GC”碱基的占比为0.00%。

[0234] 鉴于DNA片段重复连接接头处检测到存在“C”碱基的加入(与预期的精准连接相比),我们推测DNA片段重复连接接头处加入的“C”碱基是来源于模版DNA(β -globin RE2位点)上的PAM(AGG)上游4bp处的碱基。并且,进一步推测K918A这个Cas9核酸酶在sgRNA1的介导下对基因组DNA片段(β -globin RE2位点)进行切割时,对与sgRNA互补的DNA链进行切割时,是在PAM上游3bp处进行切割,而对与sgRNA非互补DNA链进行切割时,则是在PAM(AGG)上游4bp处进行突出末端切割,从而产生了突出断裂末端U4。突出断裂末端U4在细胞修复系统的作用下产生DNA片段重复时,导致了DNA片段重复连接接头处“C”碱基的加入。

[0235] 同理,鉴于DNA片段重复连接接头处检测到存在“GC”碱基的加入(与预期的精准连接相比),我们推测DNA片段重复连接接头处加入的“GC”碱基是来源于模版DNA(β -globin RE2位点)上的PAM(AGG)上游4bp处和5bp的碱基。进一步推测K918A这个Cas9核酸酶在sgRNA1的介导下对基因组DNA片段(β -globin RE2位点)进行切割时,对与sgRNA互补的DNA链进行切割时,是在PAM上游3bp处进行切割,而对与sgRNA非互补DNA链进行切割时,是在PAM(AGG)上游5bp处进行突出末端切割,从而产生了突出断裂末端U5。突出断裂末端U5在细胞修复系统的作用下产生DNA片段重复时,导致了DNA片段重复连接接头处“GC”碱基的加入。

[0236] 而当K918A这个Cas9核酸酶在sgRNA1的介导下对基因组DNA片段(β -globin RE2位点)进行切割时,对与sgRNA互补的DNA链进行切割时,是在PAM上游3bp处进行切割,对与sgRNA非互补DNA链进行切割时,是在PAM(AGG)上游3bp处进行钝末端切割,从而产生了钝断裂末端U3。钝断裂末端U3在细胞修复系统的作用下产生DNA片段重复时,不会导致DNA片段重复连接接头处碱基的加入,而是产生与预期相符的精准连接(Joined precisely)。

[0237] 因此,我们认为:在Cas9核酸酶K918A的切割下,产生的断裂末端中,钝断裂末端U3的占比=预期相符的精准连接(Joined precisely)的比例=86.28%。突出断裂末端U4的比例=“C”碱基的加入比例=2.25%。突出断裂末端U5的比例=“GC”碱基的加入比例=0.00%。

[0238] 但是,我们观察到,除了与预期相符的精准连接(Joined precisely)、“C”碱基的加入、以及“GC”碱基的加入这三大类情况以外,还有一类随机的碱基删除(Small deletion)。我们认为这类随机的碱基删除(Small deletion)是各断裂末端(钝断裂末端

U3/突出断裂末端U4/突出断裂末端U5)在细胞修复系统的作用下随机产生的,各断裂末端以均等的概率来产生碱基删除 (Small deletion),各断裂末端在细胞修复系统的作用下所产生的碱基删除 (Small deletion)的数量与各断裂末端的数量成正比。

[0239] 基于随机碱基删除现象的存在,我们认为,经过测序获得的各断裂末端的实测比例与其真实比例存在差距,需要进行修正还原,即以各种断裂末端的实测比例之和为基准,计算各断裂末端的比例,以此作为该断裂末端的占比。即对Cas9核酸酶K918A的切割所产生的各断裂末端的占比进行标准化计算,钝断裂末端U3的占比为97.46%【计算方法为:86.28% ÷ (86.28%+2.25%+0.00%)】。突出断裂末端U4的比例为2.54%【计算方法为:2.25% ÷ (86.28%+2.25%+0.00%)】。突出断裂末端U5的比例为0.00%【计算方法为:0.00% ÷ (86.28%+2.25%+0.00%)】。

[0240] 亦即,Cas9核酸酶K918A在sgRNA1的介导下对基因组DNA片段进行的切割方式中,U3钝末端切割的比例为97.46%,U4突出末端切割的比例为2.54%,U5突出末端切割的比例为0.00%。

[0241] 参照上述方法,再计算野生型Cas9核酸酶(简称Cas9WT,WT)在sgRNA1的介导下对基因组DNA片段进行的切割方式中,U3钝末端切割的占比X1、U4突出末端切割X2、U5突出末端切割的占比X3。结果,如图2D和下表2-1所示:

[0242] 表2-1

[0243]

选用的 Cas9 核酸酶	sgRNA1		
	U3 (Precise) 占比 X1	U4 (+C) 占比 X2	U5 (+GC) 占比 X3
Cas9WT	94.30%	5.70%	0.00%
K918A	97.46%	2.54%	0.00%

[0244] 可见,在sgRNA1的介导下,相比于SpCas9核酸酶(Cas9WT),K918A这个Cas9核酸酶突变体对与sgRNA1非互补的DNA链进行切割时,在PAM上游4bp处进行切割的比例明显减少(U4),在PAM上游3bp处进行切割的比例增多(U3)。

[0245] 如图2E所示,利用高通量测序技术检测DNA片段删除连接接头的情况,除了与预期相符的精准连接(Joined precisely)以外,还存在与预期的精准连接相比,删除连接接头处加入了“T”碱基、“AT”碱基、“CAT”碱基的情况。选用不同的Cas9核酸酶时,检测到的与预期相符的精准连接(Joined precisely)、“+T”碱基、“+AT”碱基、“+CAT”碱基的占比不同。以选用K918A这个Cas9核酸酶为例,检测到与预期相符的精准连接(Joined precisely)的占比为17.57%，“+T”碱基的占比为27.48%，“+AT”碱基的占比为34.98%，“+CAT”碱基的占比为2.04%。

[0246] 鉴于DNA片段删除连接接头处检测到存在“T”碱基的加入(与预期的精准连接相比),我们推测DNA片段删除连接接头处加入的“T”碱基是来源于模版DNA(β -globin RE2位点)上的PAM(TGG)上游4bp处的碱基。并且,进一步推测K918A这个Cas9核酸酶在sgRNA2的介导下对基因组DNA片段(β -globin RE2位点)进行切割时,对与sgRNA互补的DNA链进行切割时,是在PAM上游3bp处进行切割,而对与sgRNA非互补DNA链进行切割时,则是在PAM(TGG)上游4bp处进行突出末端切割,从而产生了突出断裂末端D4。突出断裂末端D4在细胞修复系统

的作用下产生DNA片段删除时,导致了DNA片段删除连接接头处“T”碱基的加入。

[0247] 同理,鉴于DNA片段删除连接接头处检测到存在“AT”碱基的加入(与预期的精准连接相比),我们推测DNA片段删除连接接头处加入的“AT”碱基是来源于模版DNA(β -globin RE2位点)上的PAM(TGG)上游4bp和5bp处的碱基。进一步推测K918A这个Cas9核酸酶在sgRNA2的介导下对基因组DNA片段(β -globin RE2位点)进行切割时,对与sgRNA互补的DNA链进行切割时,是在PAM上游3bp处进行切割,而对与sgRNA非互补DNA链进行切割时,是在PAM(TGG)上游5bp处进行突出末端切割,从而产生了突出断裂末端D5。突出断裂末端D5在细胞修复系统的作用下产生DNA片段删除时,导致了DNA片段删除连接接头处“AT”碱基的加入。

[0248] 同理,鉴于DNA片段删除连接接头处检测到存在“CAT”碱基的加入(与预期的精准连接相比),我们推测DNA片段删除连接接头处加入的“CAT”碱基是来源于模版DNA(β -globin RE2位点)上的PAM(TGG)上游4bp、5bp、6bp处的碱基。进一步推测K918A这个Cas9核酸酶在sgRNA2的介导下对基因组DNA片段(β -globin RE2位点)进行切割时,对与sgRNA互补的DNA链进行切割时,是在PAM上游3bp处进行切割,而对与sgRNA非互补DNA链进行切割时,是在PAM(TGG)上游6bp处进行突出末端切割,从而产生了突出断裂末端D6。突出断裂末端D6在细胞修复系统的作用下产生DNA片段删除时,导致了DNA片段删除连接接头处“CAT”碱基的加入。

[0249] 而当K918A这个Cas9核酸酶在sgRNA2的介导下对基因组DNA片段(β -globin RE2位点)进行切割时,对与sgRNA互补的DNA链进行切割时,是在PAM上游3bp处进行切割,对与sgRNA非互补DNA链进行切割时,是在PAM(TGG)上游3bp处进行钝末端切割,从而产生了钝断裂末端D3。钝断裂末端D3在细胞修复系统的作用下产生DNA片段删除时,不会导致DNA片段删除连接接头处碱基的加入,而是产生与预期相符的精准连接(Joined precisely)。

[0250] 因此,我们认为:在Cas9核酸酶K918A的切割下,产生的断裂末端中,钝断裂末端D3的占比=预期相符的精准连接(Joined precisely)的占比=17.57%。突出断裂末端D4的占比=“T”碱基的加入占比=27.48%。突出断裂末端D5的占比=“AT”碱基的加入占比=34.98%。突出断裂末端D6的占比=“CAT”碱基的加入占比=2.04%。

[0251] 但是,我们观察到,除了与预期相符的精准连接(Joined precisely)、DNA片段删除连接接头处加入了“T”碱基、“+AT”碱基、“+CAT”碱基这四大类情况以外,还有一类随机的碱基删除(Small deletion)。我们认为这类随机的碱基删除(Small deletion)是各断裂末端(钝断裂末端D3/突出断裂末端D4/突出断裂末端D5/突出断裂末端D6)在细胞修复系统的作用下随机产生的,各断裂末端以均等的概率来产生碱基删除(Small deletion),各断裂末端在细胞修复系统的作用下所产生的碱基删除(Small deletion)的数量与各断裂末端的数量成正比。

[0252] 基于随机碱基删除现象的存在,我们认为,经过测序获得的各断裂末端的实测比例与其真实比例存在差距,需要进行修正还原,即以各种断裂末端的实测比例之和为基准,计算各断裂末端的比例,以此作为该断裂末端的占比。即对Cas9核酸酶K918A的切割所产生的各断裂末端的占比进行标准化计算,钝断裂末端D3的占比为21.41%

[0253] 【计算方法为: $17.57\% \div (17.57\% + 27.48\% + 34.98\% + 2.04\%)$ 】。

[0254] 突出断裂末端D4的比例为33.48%

[0255] 【计算方法为： $27.48\% \div (17.57\% + 27.48\% + 34.98\% + 2.04\%)$ 】。

[0256] 突出断裂末端D5的比例为42.62%

[0257] 【计算方法为： $34.98\% \div (17.57\% + 27.48\% + 34.98\% + 2.04\%)$ 】。

[0258] 突出断裂末端D6的比例为2.48%

[0259] 【计算方法为： $2.04\% \div (17.57\% + 27.48\% + 34.98\% + 2.04\%)$ 】。

[0260] 亦即,Cas9核酸酶K918A在sgRNA2的介导下对基因组DNA片段进行的切割方式中,D3钝末端切割的占比为21.41%,D4突出末端切割的占比为33.48%,D5突出末端切割的占比为42.62%,D6突出末端切割的占比为2.48%。

[0261] 参照上述方法,计算出野生型Cas9核酸酶在sgRNA2的介导下对基因组DNA片段进行的切割方式中,D3钝末端切割的占比Y1、D4突出末端切割的占比Y2、D5突出末端切割的占比Y3、D6突出末端切割的占比Y4。结果如图2F和表2-2所示:

[0262] 表2-2

[0263]

选用的 Cas9 核酸酶	sgRNA2			
	D3 (Precise) 占比 Y1	D4 (+T) 占比 Y2	D5 (+AT) 占比 Y3	D6 (+CAT) 占比 Y4
Cas9WT	28.74%	24.50%	42.57%	4.19%
K918A	21.41%	33.48%	42.62%	2.48%

[0264] 可见,在sgRNA2的介导下,相比于SpCas9核酸酶(Cas9WT),K918A突变体对基因组DNA片段中与sgRNA2非互补的DNA链进行切割时,在PAM上游4bp处进行切割的比例明显提高(D4)。

[0265] 参照实施例1的方法,根据Cas9核酸酶在sgRNA1和sgRNA2的介导下分别对基因组DNA片段进行切割的方式比例,预测产生的断裂末端的序列,进而推算出DNA片段反转下游连接接头处的碱基加入情况及比例。结果如图2G所示,推算结果与实验检测到的碱基加入比例相近。更进一步证实了Cas9核酸酶在sgRNA组合的介导下,可以在PAM上游3bp处到更远碱基处切割非互补DNA链。

[0266] 此外,还将本发明的Cas9核酸酶突变体Cas9-K918A及对照突变体K775A、R778A、E779A、K918P与针对STM位点(β -globin RE1)的两个sgRNAs一起转染人胚肾HEK293T细胞,转染48小时后收集基因组DNA,利用高通量测序引物进行PCR扩增DNA片段删除、反转和重复,建库进行高通量测序。根据DNA片段删除和重复的连接接头碱基加入情况,计算这些突变体在两个sgRNAs介导下的切割方式比例。

[0267] 针对STM位点(β -globin RE1)的sgRNAs靶向序列:

[0268] β -globin RE1sgRNA1:GATTGTTGTTGCCTTGGAGTG (SEQ ID NO.19);

[0269] β -globin RE1sgRNA2:GCTGGTCCCCTGGTAACCTGG (SEQ ID NO.20);

[0270] 正反向脱氧寡核苷酸:

[0271] β -globin RE1sgRNA1F:accgATTGTTGTTGCCTTGGAGTG (SEQ ID NO.21);

[0272] β -globin RE1sgRNA1R:aaacCACTCCAAGGCAACAACAAT (SEQ ID NO.22);

[0273] β -globin RE1sgRNA2F:accgCTGGTCCCCTGGTAACCTGG (SEQ ID NO.23);

[0274] β -globin RE1sgRNA2R:aaacCCAGGTTACCAGGGGACCAG (SEQ ID NO.24);

[0275] 高通量引物:

[0276] HiSeq-hSTM-aF1:

[0277]

AATGATACGGCGACCACCGAGATCTACACTCTTCCCTACACGACGCTCTCCGATCTTGCTTAGAGCCAGGACTAA
TTGC; (SEQ ID NO.25)

[0278] HiSeq-hSTM-aR2:

[0279]

AATGATACGGCGACCACCGAGATCTACACTCTTCCCTACACGACGCTCTCCGATCTTGGGTGTAGAAATGAGCAA
ATAAGT (SEQ ID NO.26)

[0280] HiSeq-hSTM-2F:

[0281]

CAAGCAGAAGACGGCATAACGAGATGATCGTGTGACTGGAGTTCAGACGTGTGCTCTCCGATCTAGATTGAGTTCTG
TTTGTTCATCTAC (SEQ ID NO.27)

[0282] HiSeq-hSTM-2R:

[0283]

CAAGCAGAAGACGGCATAACGAGATAGTCAAGTGACTGGAGTTCAGACGTGTGCTCTCCGATCTCAGCTCTGCCTGA
AAGGAGTC (SEQ ID NO.28)。

[0284] 如图3A和3B所示,与野生型SpCas9核酸酶(简称Cas9WT)相比,对照突变体K775A、R778A、E779A和K918P在sgRNA1和sgRNA2的介导下对基因组DNA片段进行切割的方式没有明显的改变;而Cas9核酸酶突变体Cas9-K918A与野生型SpCas9核酸酶(简称Cas9WT)相比,在sgRNA1和sgRNA2的介导下对基因组DNA片段进行切割的方式发生了明显的改变。

[0285] 综上所述,本发明的Cas9核酸酶(Cas9-K918A)与野生型Cas9核酸酶相比,对目的基因组DNA片段进行切割时产生的突出断裂末端与钝断裂末端的比例不同。采用本发明的Cas9核酸酶(Cas9-K918A)可以实现对目的基因组DNA片段特定位置的切割并产生突出末端,以补平连接的方式可加入与突出断裂末端互补的碱基,进而可实现特定位置的精准DNA片段编辑。

[0286] 本申请的参考文献如下:

[0287] 1.Stamatoyannopoulos,JA.(2012).What does our genome encode?Genome Res,22:1602-1611.

[0288] 2.The ENCODE Project Consortium.(2012).An integrated encyclopedia of DNA elements in the human genome.Nature,489:57-74.

[0289] 3.Banerji,J,L Olson,and W Schaffner.(1983).A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes.Cell,33:729-740.

[0290] 4.Zhang,T,P Haws,and Q Wu.(2004).Multiple variable first exons:a mechanism for cell-and tissue-specific gene regulation.Genome Res,14:79-89.

[0291] 5.Neph,S,et al.(2012).An expansive human regulatory lexicon encoded in transcription factor footprints.Nature,489:83-90.

- [0292] 6. Shen, Y, et al. (2012). A map of the cis-regulatory sequences in the mouse genome. *Nature*, 488:116-120.
- [0293] 7. Thurman, RE, et al. (2012). The accessible chromatin landscape of the human genome. *Nature*, 489:75-82.
- [0294] 8. de Laat, W and D Duboule. (2013). Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature*, 502:499-506.
- [0295] 9. McClintock, B. (1950). The origin and behavior of mutable loci in maize. *Proc Natl Acad Sci U S A*, 36:344-355.
- [0296] 10. McClintock, B. (1984). The significance of responses of the genome to challenge. *Science*, 226:792-801.
- [0297] 11. Brinster, RL, et al. (1981). Somatic expression of herpes thymidine kinase in mice following injection of a fusion gene into eggs. *Cell*, 27:223-231.
- [0298] 12. Harbers, K, D Jahner, and R Jaenisch. (1981). Microinjection of cloned retroviral genomes into mouse zygotes: integration and expression in the animal. *Nature*, 293:540-542.
- [0299] 13. Gordon, JW, et al. (1980). Genetic transformation of mouse embryos by microinjection of purified DNA. *Proc Natl Acad Sci U S A*, 77:7380-7384.
- [0300] 14. Palmiter, RD, et al. (1982). Dramatic growth of mice that develop from eggs microinjected with metallothionein-growth hormone fusion genes. *Nature*, 300:611-615.
- [0301] 15. Capecchi, MR. (2005). Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. *Nat Rev Genet*, 6:507-512.
- [0302] 16. Carroll, D. (2014). Genome engineering with targetable nucleases. *Annu Rev Biochem*, 83:409-439.
- [0303] 17. Smithies, O, et al. (1985). Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. *Nature*, 317:230-234.
- [0304] 18. Thomas, KR and MR Capecchi. (1986). Introduction of homologous DNA sequences into mammalian cells induces mutations in the cognate gene. *Nature*, 324:34-38.
- [0305] 19. Zheng, B, et al. (2000). Engineering mouse chromosomes with Cre-loxP: range, efficiency, and somatic applications. *Mol Cell Biol*, 20:648-655.
- [0306] 20. Wu, S, et al. (2007). Toward simpler and faster genome-wide mutagenesis in mice. *Nat Genet*, 39:922-930.
- [0307] 21. Gupta, A, et al. (2013). Targeted chromosomal deletions and inversions in zebrafish. *Genome Res*, 23:1008-1017.
- [0308] 22. Xiao, A, et al. (2013). Chromosomal deletions and inversions mediated by TALENs and CRISPR/Cas in zebrafish. *Nucleic Acids Res*, 41:e141.
- [0309] 23. Kraft, K, et al. (2015). Deletions, Inversions, Duplications: Engineering

of Structural Variants using CRISPR/Cas in Mice. *Cell Rep*,10:833-839.

[0310] 24. Wu, S, et al. (2008). A protocol for constructing gene targeting vectors: generating knockout mice for the cadherin family and beyond. *Nature Protocol*,3:1056-1076.

[0311] 25. Jinek, M, et al. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*,337:816-821.

[0312] 26. Cong, L, et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science*,339:819-823.

[0313] 27. Mali, P, et al. (2013). RNA-guided human genome engineering via Cas9. *Science*,339:823-826.

[0314] 28. Li, J, et al. (2015). Efficient inversions and duplications of mammalian regulatory DNA elements and gene clusters by CRISPR/Cas9. *J Mol Cell Biol*,7:284-298.

[0315] 29. Sartori, AA, et al. (2007). Human CtIP promotes DNA end resection. *Nature*,450:509-514.

[0316] 30. Anand, R, et al. (2016). Phosphorylated CtIP Functions as a Co-factor of the MRE11-RAD50-NBS1 Endonuclease in DNA End Resection. *Mol Cell*,64:940-950.

[0317] 31. Li, J, J Shou, and Q Wu. (2015). DNA fragment editing of genomes by CRISPR/Cas9. *Hereditas*,37:992-1002.

[0318] 32. Huang, H and Q Wu. (2016). CRISPR Double Cutting through the Labyrinthine Architecture of 3D Genomes. *J Genet Genomics*,43:273-288.

[0319] 33. Guo, Y, et al. (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell*,162:900-910.

[0320] 以上所述,仅为本发明的较佳实施例,并非对本发明任何形式上和实质上的限制,应当指出,对于本技术领域的普通技术人员,在不脱离本发明方法的前提下,还将可以做出若干改进和补充,这些改进和补充也应视为本发明的保护范围。凡熟悉本专业的技术人员,在不脱离本发明的精神和范围的情况下,当可利用以上所揭示的技术内容而做出的些许更动、修饰与演变的等同变化,均为本发明的等效实施例;同时,凡依据本发明的实质技术对上述实施例所作的任何等同变化的更动、修饰与演变,均仍属于本发明的技术方案的范围

SEQUENCE LISTING

<110> 上海交通大学

<120> 一种Cas9核酸酶 K918A及其用途

<130> 171287

<160> 28

<170> PatentIn version 3.3

<210> 1

<211> 20

<212> DNA

<213> Artificial

<220>

<223> HS51 RE1sgRNA1

<400> 1

gccacacatc caaggctgac 20

<210> 2

<211> 21

<212> DNA

<213> Artificial

<220>

<223> HS51 RE1sgRNA2

<400> 2

gagatttggg gcgtcaggaa g 21

<210> 3

<211> 77

<212> DNA

<213> Artificial

<220>

<223> Hiseq-hHs51-aF

<400> 3

atgatacggc gaccaccgag atctacaacte tttcctaca cgacgetett cgatctgca 60

aggagatccg tgtcgtc 77

<210> 4

<211> 82

<212> DNA

<213> Artificial

<220>

<223> Hiseq-hs51-aRa

<400> 4

aatgatacgg cgaccaccga gatctacact ctttcctac acgacgtctt tccgatctaa 60
ggatgttgtg gaaggcgagc ag 82

<210> 5

<211> 87

<212> DNA

<213> Artificial

<220>

<223> Hiseq-hs51-bFa

<400> 5

caagcagaag acggcatacg agatggacgg gtgactggag ttcagacgtg tgctcttccg 60
atctcttttac atgacagctt ccggtag 87

<210> 6

<211> 89

<212> DNA

<213> Artificial

<220>

<223> Hiseq-hHs51-bR

<400> 6

caagcagaag acggcatacg agatttgact gtgactggag ttcagacgtg tgctcttccg 60
atcttttttg gctaacaaca tagtgettc 89

<210> 7

<211> 1401

<212> PRT

<213> Artificial

<220>

<223> SpCas9

<400> 7

Met Ala Pro Lys Lys Lys Arg Lys Val Gly Ile His Gly Val Pro Ala

1 5 10 15

Ala Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser

20 25 30

Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys

35 40 45

Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu

50 55 60

Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg

65 70 75 80

Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile

85 90 95

Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp			
100	105	110	
Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys			
115	120	125	
Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala			
130	135	140	
Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val			
145	150	155	160
Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala			
165	170	175	
His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn			
180	185	190	
Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr			
195	200	205	
Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp			
210	215	220	
Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu			
225	230	235	240
Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly			
245	250	255	
Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn			
260	265	270	
Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr			
275	280	285	
Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala			
290	295	300	
Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser			
305	310	315	320
Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala			
325	330	335	
Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu			
340	345	350	
Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe			
355	360	365	
Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala			
370	375	380	
Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met			
385	390	395	400
Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu			

	405	410	415
Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His			
	420	425	430
Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro			
	435	440	445
Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg			
	450	455	460
Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala			
465	470	475	480
Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu			
	485	490	495
Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met			
	500	505	510
Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His			
	515	520	525
Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val			
	530	535	540
Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu			
545	550	555	560
Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val			
	565	570	575
Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe			
	580	585	590
Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu			
	595	600	605
Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu			
	610	615	620
Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu			
625	630	635	640
Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr			
	645	650	655
Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg			
	660	665	670
Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg			
	675	680	685
Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly			
	690	695	700
Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr			
705	710	715	720

Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser
 725 730 735
 Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys
 740 745 750
 Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met
 755 760 765
 Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn
 770 775 780
 Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg
 785 790 795 800
 Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His
 805 810 815
 Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr
 820 825 830
 Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn
 835 840 845
 Arg Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu
 850 855 860
 Lys Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn
 865 870 875 880
 Arg Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met
 885 890 895
 Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg
 900 905 910
 Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu
 915 920 925
 Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile
 930 935 940
 Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr
 945 950 955 960
 Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys
 965 970 975
 Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val
 980 985 990
 Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala
 995 1000 1005
 Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser
 1010 1015 1020
 Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met

1025	1030	1035
Ile Ala Lys Ser Glu Gln	Glu Ile Gly Lys Ala	Thr Ala Lys Tyr
1040	1045	1050
Phe Phe Tyr Ser Asn Ile	Met Asn Phe Phe Lys	Thr Glu Ile Thr
1055	1060	1065
Leu Ala Asn Gly Glu Ile	Arg Lys Arg Pro Leu	Ile Glu Thr Asn
1070	1075	1080
Gly Glu Thr Gly Glu Ile	Val Trp Asp Lys Gly	Arg Asp Phe Ala
1085	1090	1095
Thr Val Arg Lys Val Leu	Ser Met Pro Gln Val	Asn Ile Val Lys
1100	1105	1110
Lys Thr Glu Val Gln Thr	Gly Gly Phe Ser Lys	Glu Ser Ile Leu
1115	1120	1125
Pro Lys Arg Asn Ser Asp	Lys Leu Ile Ala Arg	Lys Lys Asp Trp
1130	1135	1140
Asp Pro Lys Lys Tyr Gly	Gly Phe Asp Ser Pro	Thr Val Ala Tyr
1145	1150	1155
Ser Val Leu Val Val Ala	Lys Val Glu Lys Gly	Lys Ser Lys Lys
1160	1165	1170
Leu Lys Ser Val Lys Glu	Leu Leu Gly Ile Thr	Ile Met Glu Arg
1175	1180	1185
Ser Ser Phe Glu Lys Asn	Pro Ile Asp Phe Leu	Glu Ala Lys Gly
1190	1195	1200
Tyr Lys Glu Val Lys Lys	Asp Leu Ile Ile Lys	Leu Pro Lys Tyr
1205	1210	1215
Ser Leu Phe Glu Leu Glu	Asn Gly Arg Lys Arg	Met Leu Ala Ser
1220	1225	1230
Ala Gly Glu Leu Gln Lys	Gly Asn Glu Leu Ala	Leu Pro Ser Lys
1235	1240	1245
Tyr Val Asn Phe Leu Tyr	Leu Ala Ser His Tyr	Glu Lys Leu Lys
1250	1255	1260
Gly Ser Pro Glu Asp Asn	Glu Gln Lys Gln Leu	Phe Val Glu Gln
1265	1270	1275
His Lys His Tyr Leu Asp	Glu Ile Ile Glu Gln	Ile Ser Glu Phe
1280	1285	1290
Ser Lys Arg Val Ile Leu	Ala Asp Ala Asn Leu	Asp Lys Val Leu
1295	1300	1305
Ser Ala Tyr Asn Lys His	Arg Asp Lys Pro Ile	Arg Glu Gln Ala
1310	1315	1320

Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro
 1325 1330 1335
 Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr
 1340 1345 1350
 Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser
 1355 1360 1365
 Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly
 1370 1375 1380
 Gly Asp Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys
 1385 1390 1395
 Lys Lys Lys
 1400

<210> 8

<211> 4206

<212> DNA

<213> Artificial

<220>

<223> SpCas9

<400> 8

atggcccaaa agaagaagcg gaaggtcggg atccacgggtg tcccagcagc catggacaag 60
 aagtactcca ttgggctcga tatcggcaca aacagcgtcg gctgggcccgt cattacggac 120
 gagtacaagg tgccgagcaa aaaattcaaa gttctgggca ataccgatcg ccacagcata 180
 aagaagaacc tcattggcgc cctcctgttc gactccgggg agacggccga agccacgcgg 240
 ctcaaaagaa cagcacggcg cagatatacc cgcagaaaga atcggatctg ctacctgcag 300
 gagatcttta gtaatgagat ggctaagggtg gatgactctt tcttccatag gctggaggag 360
 tcctttttgg tggaggagga taaaagcac gagcgcacc caatctttgg caatatcgtg 420
 gacgaggtgg cgtaccatga aaagtacca accatatac atctgaggaa gaagcttgta 480
 gacagtactg ataaggctga cttgcggttg atctatctcg cgctggcgca tatgatcaaa 540
 tttcggggac acttcctcat cgagggggac ctgaaccag acaacagcga tgtcgacaaa 600
 ctctttatcc aactggttca gacttacaat cagcttttcg aagagaacc gatcaacgca 660
 tccggagttg acgcaaaage aatcctgagc gctaggtgtg ccaaatccc gcggtcga 720
 aacctcatcg cacagctccc tggggagaag aagaacggcc tgtttggtaa tcttatcgcc 780
 ctgtcactcg ggctgacccc caactttaa tetaacttcg acctggccga agatgccaag 840
 cttcaactga gcaaagacac ctacgatgat gatctcgaca atctgctggc ccagatcggc 900
 gaccagtacg cagacctttt tttggcgga aagaacctgt cagacgcat tctgctgagt 960
 gatattctgc gagtgaacac ggagatecacc aaagetccgc tgagcgtag tatgatcaag 1020
 cgctatgatg agcaccacca agaettgact ttgctgaagg cccttgtag acagcaactg 1080
 cctgagaagt acaaggaaat tttcttcgat cagtctaaaa atggctacgc cggatacatt 1140
 gacggcggag caagccagga ggaattttac aaatttatta agcccatctt gaaaaaatg 1200

gacggcaccg aggagctgct ggtaaagctt aacagagaag atctgttgcg caaacagcgc 1260
actttcgaca atggaagcat cccccaccag attcacctgg gcgaactgca cgctatactc 1320
aggcggcaag aggatttcta cccctttttg aaagataaca gggaaaagat tgagaaaatc 1380
ctcacatttc ggatacccta ctatgtaggc cccctcgccc ggggaaattc cagattcgcg 1440
tggatgactc gcaaatcaga agagaccatc actccctgga acttcgagga agtcgtggat 1500
aagggggcct ctgcccagtc cttcatcgaa aggatgacta actttgataa aaatctgcct 1560
aacgaaaagg tgcttcctaa aactctctg ctgtacgagt acttcacagt ttataacgag 1620
ctcaccaagg tcaaatcagt cacagaaggg atgagaaagc cagcattcct gtctggagag 1680
cagaagaaag ctatcgtgga cctcctcttc aagacgaacc ggaaagtac cgtgaaacag 1740
ctcaaagaag actatttcaa aaagattgaa tgtttcgact ctgttgaaat cagcggagtg 1800
gaggatcgct tcaacgcac cctgggaacg tatcacgac tectgaaaat cattaagac 1860
aaggacttcc tggacaatga ggagaacgag gacattcttg aggacattgt cctcaccctt 1920
acgttgtttg aagatagga gatgattgaa gaacgcttga aaacttacgc tcctctcttc 1980
gacgacaaag tcatgaaaca gctcaagagg cgccgatata caggatgggg gcggctgtca 2040
agaaaactga tcaatgggat ccgagacaag cagagtggaa agacaatcct ggattttctt 2100
aagtccgatg gatttgccaa ccggaacttc atgcagttga tccatgatga ctctctcacc 2160
ttaaaggagg acatccagaa agcacaagtt tctggccagg gggacagtct tcacgagcac 2220
atcgctaate ttgcaggtag cccagctatc aaaaaggaa tactgcagac cgttaaggtc 2280
gtggatgaac tcgtcaaagt aatgggaagg cataagccc agaatatcgt tatcgagatg 2340
gcccagagaga accaaactac ccagaaggga cagaagaaca gtagggaaag gatgaagagg 2400
attgaagagg gtataaaaga actgggtcc caaatcetta agaacacce agttgaaaac 2460
accagcttc agaatgagaa gctctacctg tactacctgc agaacggcag ggacatgtac 2520
gtggatcagg aactggacat caatcggtc tccgactacg acgtggatca tatcgtgccc 2580
cagctttttc tcaaagatga ttctattgat aataaagtgt tgacaagatc cgataaaaat 2640
agaggggaaga gtgataacgt cccctcagaa gaagttgtca agaaaatgaa aaattattgg 2700
cggcagctgc tgaacgcca actgatcaca caacggaagt tcgataatct gactaaggct 2760
gaacgaggtg gcctgtctga gttggataaa gcaggcttca tcaaaaggca gcttgttgag 2820
acacgccaga tcaccaagca cgtggcccaa attctcgatt cacgcatgaa caccaagtac 2880
gatgaaaatg acaaactgat tcgagaggtg aaagttatta ctctgaagtc taagctggtc 2940
tcagatttca gaaaggactt tcagttttat aaggtgagag agatecaaca ttaccacat 3000
gcgcatgatg cctacctgaa tgcagtggta ggcactgcac ttatcaaaaa atatcccaag 3060
cttgaatctg aatttgttta cggagactat aaagtgtacg atgttaggaa aatgatcgca 3120
aagtctgagc aggaaaatagg caaggccacc gctaagtact tcttttacag caatattatg 3180
aattttttca agaccgagat tacaactggc aatggagaga ttcggaagcg accacttate 3240
gaaacaaacg gagaaacagg agaaatcgtg tgggacaagg gtagggattt cgcgacagtc 3300
cggaaggtec tgtccatgcc gcagtgtaac atcgttaaaa agaccgaagt acagaccgga 3360
ggcttctcca aggaaagtat cctcccgaaa aggaacagcg acaagctgat cgcacgcaa 3420
aaagattggg accccaagaa atacggcgga ttcgattctc ctacagtcgc ttacagtgta 3480
ctggttggtg ccaaagtgga gaaagggaag tctaaaaaac tcaaaagcgt caaggaactg 3540

ctgggcatca caatcatgga gcgatcaagc ttcgaaaaa accecatcga ctttctcgag 3600
 gcgaaaggat ataaagaggt caaaaaagac ctcatcatta agcttcccaa gtactctctc 3660
 tttgagcttg aaaacggccg gaaacgaatg ctcgctagtg cgggagct gcagaaaggt 3720
 aacgagctgg cactgccctc taaatacgtt aatttctgt atctggccag ccactatgaa 3780
 aagctcaaag ggtctcccga agataatgag cagaagcagc tgttcgtgga acaacacaaa 3840
 cactaccttg atgagatcat cgagcaaata agcgaattct ccaaagagt gatcctcgcc 3900
 gacgctaacc tcgataaggt gtttctgct tacaataagc acagggataa gcccatcagg 3960
 gagcaggcag aaaacattat ccacttggtt actctgacca acttgggagc gcctgcagcc 4020
 ttcaagtact tcgacaccac catagacaga aagcgttaca cctctacaaa ggaggtcctg 4080
 gacgccacac tgattcatca gtcaattacg gggtctatg aaacaagaat cgacctctct 4140
 cagctcggtg gagacaagcg tctgtctgct actaagaaag ctggtcaagc taagaaaaag 4200
 aaataa 4206

<210> 9

<211> 1401

<212> PRT

<213> Artificial

<220>

<223> Cas9- K918A

<400> 9

Met	Ala	Pro	Lys	Lys	Lys	Arg	Lys	Val	Gly	Ile	His	Gly	Val	Pro	Ala
1			5					10					15		
Ala	Met	Asp	Lys	Lys	Tyr	Ser	Ile	Gly	Leu	Asp	Ile	Gly	Thr	Asn	Ser
			20					25					30		
Val	Gly	Trp	Ala	Val	Ile	Thr	Asp	Glu	Tyr	Lys	Val	Pro	Ser	Lys	Lys
			35				40					45			
Phe	Lys	Val	Leu	Gly	Asn	Thr	Asp	Arg	His	Ser	Ile	Lys	Lys	Asn	Leu
			50				55				60				
Ile	Gly	Ala	Leu	Leu	Phe	Asp	Ser	Gly	Glu	Thr	Ala	Glu	Ala	Thr	Arg
65					70					75				80	
Leu	Lys	Arg	Thr	Ala	Arg	Arg	Arg	Tyr	Thr	Arg	Arg	Lys	Asn	Arg	Ile
					85					90				95	
Cys	Tyr	Leu	Gln	Glu	Ile	Phe	Ser	Asn	Glu	Met	Ala	Lys	Val	Asp	Asp
					100					105				110	
Ser	Phe	Phe	His	Arg	Leu	Glu	Glu	Ser	Phe	Leu	Val	Glu	Glu	Asp	Lys
					115					120				125	
Lys	His	Glu	Arg	His	Pro	Ile	Phe	Gly	Asn	Ile	Val	Asp	Glu	Val	Ala
					130					135				140	
Tyr	His	Glu	Lys	Tyr	Pro	Thr	Ile	Tyr	His	Leu	Arg	Lys	Lys	Leu	Val
145					150					155				160	

Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala
 165 170 175
 His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn
 180 185 190
 Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr
 195 200 205
 Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp
 210 215 220
 Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu
 225 230 235 240
 Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly
 245 250 255
 Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn
 260 265 270
 Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr
 275 280 285
 Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala
 290 295 300
 Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser
 305 310 315 320
 Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala
 325 330 335
 Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu
 340 345 350
 Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe
 355 360 365
 Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala
 370 375 380
 Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met
 385 390 395 400
 Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu
 405 410 415
 Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His
 420 425 430
 Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro
 435 440 445
 Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg
 450 455 460
 Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala

465	470	475	480
Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu			
	485	490	495
Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met			
	500	505	510
Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His			
	515	520	525
Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val			
	530	535	540
Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu			
545	550	555	560
Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val			
	565	570	575
Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe			
	580	585	590
Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu			
	595	600	605
Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu			
	610	615	620
Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu			
625	630	635	640
Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr			
	645	650	655
Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg			
	660	665	670
Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg			
	675	680	685
Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly			
	690	695	700
Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr			
705	710	715	720
Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser			
	725	730	735
Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys			
	740	745	750
Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met			
	755	760	765
Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn			
	770	775	780

Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg																			
785					790					795									800
Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His																			
					805					810									815
Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr																			
					820					825									830
Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn																			
					835					840									845
Arg Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu																			
					850					855									860
Lys Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn																			
865					870					875									880
Arg Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met																			
					885					890									895
Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg																			
					900					905									910
Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu																			
					915					920									925
Asp Lys Ala Gly Phe Ile Ala Arg Gln Leu Val Glu Thr Arg Gln Ile																			
					930					935									940
Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr																			
945					950					955									960
Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys																			
					965					970									975
Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val																			
					980					985									990
Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala																			
					995					1000									1005
Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser																			
					1010					1015									1020
Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met																			
					1025					1030									1035
Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr																			
					1040					1045									1050
Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr																			
					1055					1060									1065
Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn																			
					1070					1075									1080
Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala																			

1085	1090	1095
Thr Val Arg Lys Val Leu	Ser Met Pro Gln Val	Asn Ile Val Lys
1100	1105	1110
Lys Thr Glu Val Gln Thr	Gly Gly Phe Ser Lys	Glu Ser Ile Leu
1115	1120	1125
Pro Lys Arg Asn Ser Asp	Lys Leu Ile Ala Arg	Lys Lys Asp Trp
1130	1135	1140
Asp Pro Lys Lys Tyr Gly	Gly Phe Asp Ser Pro	Thr Val Ala Tyr
1145	1150	1155
Ser Val Leu Val Val Ala	Lys Val Glu Lys Gly	Lys Ser Lys Lys
1160	1165	1170
Leu Lys Ser Val Lys Glu	Leu Leu Gly Ile Thr	Ile Met Glu Arg
1175	1180	1185
Ser Ser Phe Glu Lys Asn	Pro Ile Asp Phe Leu	Glu Ala Lys Gly
1190	1195	1200
Tyr Lys Glu Val Lys Lys	Asp Leu Ile Ile Lys	Leu Pro Lys Tyr
1205	1210	1215
Ser Leu Phe Glu Leu Glu	Asn Gly Arg Lys Arg	Met Leu Ala Ser
1220	1225	1230
Ala Gly Glu Leu Gln Lys	Gly Asn Glu Leu Ala	Leu Pro Ser Lys
1235	1240	1245
Tyr Val Asn Phe Leu Tyr	Leu Ala Ser His Tyr	Glu Lys Leu Lys
1250	1255	1260
Gly Ser Pro Glu Asp Asn	Glu Gln Lys Gln Leu	Phe Val Glu Gln
1265	1270	1275
His Lys His Tyr Leu Asp	Glu Ile Ile Glu Gln	Ile Ser Glu Phe
1280	1285	1290
Ser Lys Arg Val Ile Leu	Ala Asp Ala Asn Leu	Asp Lys Val Leu
1295	1300	1305
Ser Ala Tyr Asn Lys His	Arg Asp Lys Pro Ile	Arg Glu Gln Ala
1310	1315	1320
Glu Asn Ile Ile His Leu	Phe Thr Leu Thr Asn	Leu Gly Ala Pro
1325	1330	1335
Ala Ala Phe Lys Tyr Phe	Asp Thr Thr Ile Asp	Arg Lys Arg Tyr
1340	1345	1350
Thr Ser Thr Lys Glu Val	Leu Asp Ala Thr Leu	Ile His Gln Ser
1355	1360	1365
Ile Thr Gly Leu Tyr Glu	Thr Arg Ile Asp Leu	Ser Gln Leu Gly
1370	1375	1380

Gly Asp Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys
 1385 1390 1395
 Lys Lys Lys
 1400
 <210> 10
 <211> 4206
 <212> DNA
 <213> Artificial
 <220>
 <223> Cas9- K918A
 <400> 10
 atggcccca agaagaagcg gaaggtcggg atccacgggt tcccagcagc catggacaag 60
 aagtactcca ttgggtcga tatecggeaca aacagcgtcg gctgggcegt cattacggac 120
 gagtacaagg tgccgagcaa aaaattcaaa gttctgggca ataccgatcg ccacagcata 180
 aagaagaacc tcattggcgc cctcctgttc gactccgggg agacggccga agccacgcgg 240
 ctcaaaagaa cagcacggcg cagatatacc cgcagaaaga atcggatctg ctacctgcag 300
 gagatcttta gtaatgagat ggctaagggt gatgactctt tcttccatag gctggaggag 360
 tcctttttgg tggaggagga taaaaagcac gagegccacc caatctttgg caatatcgtg 420
 gacgaggtgg cgtaccatga aaagtacca accatatac atctgaggaa gaagcttgta 480
 gacagtactg ataaggctga cttgcggtt atctatctcg cgctggcgca tatgatcaaa 540
 tttcggggac acttctcat cgagggggac ctgaaccag acaacagcga tgcgacaaa 600
 ctctttatcc aactggttca gacttacaat cagcttttcg aagagaacc gatcaacgca 660
 tccggagttg acgccaagc aatcctgagc gctaggtgt ccaaatccc gcggctcga 720
 aacctcatcg cacagctccc tggggagaag aagaacggcc tgtttggtaa tcttatcgcc 780
 ctgtcactcg ggctgacccc caactttaa tctaacttcg acctggccga agatgccaa 840
 cttcaactga gcaaagacac ctacgatgat gatctcgaca atctgctggc ccagatcggc 900
 gaccagtacg cagacctttt tttggcggca aagaacctgt cagacgcat tctgctgagt 960
 gatattctgc gactgaacac ggagatcacc aaagctccgc tgagcgctag tatgatcaag 1020
 cgctatgatg agcaccacca agacttgact ttgctgaagg cccttgctag acagcaactg 1080
 cctgagaagt acaaggaaat tttcttcgat cagtctaaaa atggctacgc cggatacatt 1140
 gacggcggag caagccagga ggaattttac aaatttatta agccatctt ggaaaaaatg 1200
 gacggcaccg aggagctgct ggtaaagctt aacagagaag atctgttgcg caaacagcgc 1260
 actttcgaca atggaagcat cccccaccag attcaactgg ggaactgca cgctatactc 1320
 aggcgcaag aggatttcta cccctttttg aaagataaca gggaaaagat tgagaaaatc 1380
 ctcacatttc ggatacccta ctatgtaggc cccctcgccc ggggaaatc cagattecgc 1440
 tggatgactc gcaaatcaga agagaccatc actccctgga acttcgagga agtcgtggat 1500
 aagggggcct ctgccagtc cttcatcgaa aggatgacta actttgataa aaatctgctt 1560
 aacgaaaagg tgettcctaa aactctctg ctgtacgagt acttcacagt ttataacgag 1620
 ctcaccaagg tcaaatcgt cacagaaggg atgagaaagc cagcattcct gtctggagag 1680

cagaagaaag ctatcgtgga cctcctcttc aagacgaacc ggaaagtac cgtgaaacag 1740
ctcaaagaag actatttcaa aaagattgaa tgtttcgact ctgttgaaat cagcggagtg 1800
gaggatcgct tcaacgcac cctgggaacg taccacgatc tcctgaaaat cattaagac 1860
aaggacttcc tggacaatga ggagaacgag gacattcttg aggacattgt cctcaccctt 1920
acgttgtttg aagataggga gatgattgaa gaacgcttga aaacttacgc tcatctcttc 1980
gacgacaaag tcatgaaaca gctcaagagg cgccgatata caggatgggg gcggctgtca 2040
agaaaactga tcaatgggat ccgagacaag cagagtggaa agacaatcct ggattttctt 2100
aagtccgatg gatttgccaa ccggaacttc atgcagtga tccatgatga ctctctcacc 2160
ttaaaggagg acatccagaa agcacaagtt tctggccagg gggacagtct tcacgagcac 2220
atcgctaate ttgcaggtag ccagctatc aaaaaggaa tactgcagac cgtaaggctc 2280
gtggatgaac tcgtcaaagt aatgggaagg cataagcccg agaatatcgt tatcgagatg 2340
gcccagagaga accaaaactac ccagaaggga cagaagaaca gtagggaaag gatgaagagg 2400
attgaagagg gtataaaaaga actggggctc caaatcetta agaacaccc agttgaaaac 2460
accagcttc agaatgagaa gctctacctg tactacctgc agaacggcag ggacatgtac 2520
gtggatcagg aactggacat caatcgctc tccgactacg acgtggatca tategtgccc 2580
cagtcttttc tcaaagatga ttctattgat aataaagtgt tgacaagatc cgataaaaat 2640
agaggggaaga gtgataacgt ccctcagaa gaagttgtca agaaaatgaa aaattattgg 2700
cggcagctgc tgaacgcaa actgatcaca caacggaagt tcgataatct gactaaggct 2760
gaacgaggtg gcctgtctga gttggataaa gcaggettca tcgccaggca gcttgttgag 2820
acacgccaga tcaccaagca cgtggcccaa attctcgatt cacgcatgaa caccaagtac 2880
gatgaaaatg acaaactgat tcgagaggtg aaagttatta ctctgaagtc taagctggctc 2940
tcagatttca gaaaggactt tcagttttat aaggtgagag agatcaaca ttaccacat 3000
gcgcatgatg cctacctgaa tgcagtggta ggactgcac ttatcaaaa atatcccaag 3060
cttgaatctg aatttgttta cggagactat aaagtgtacg atgttaggaa aatgatcgca 3120
aagtctgagc aggaaaatagg caaggccacc gctaagtact tcttttacag caatattatg 3180
aattttttca agaccgagat tacactggcc aatggagaga ttcggaagcg accacttatc 3240
gaaacaaacg gagaaacagg agaaatcgtg tgggacaagg gtagggattt cgcgacagtc 3300
cggaaggctc tgtccatgcc gcaggtgaac atcgttaaaa agaccgaagt acagaccgga 3360
ggcttctcca aggaaaatg cctcccgaag agaacagcg acaagctgat cgcacgcaa 3420
aaagattggg accccaagaa atacggcgga ttcgattctc ctacagtcgc ttacagtgtg 3480
ctggttgtgg ccaaagtgga gaaagggaa tctaaaaaac tcaaagcgt caaggaactg 3540
ctgggcatca caatcatgga gcgatcaagc ttcgaaaaaa accccatcga ctttctcgag 3600
gcgaaaggat ataaagaggt caaaaaagac ctcatcatta agcttccca gtactctctc 3660
tttgagcttg aaaaacggccg gaaacgaatg ctctctagtg cgggcgagct gcagaaagg 3720
aacgagctgg cactgccctc taaatacgtt aatttctgt atctggccag ccactatgaa 3780
aagctcaaag ggtctcccga agataatgag cagaagcagc tgttcgtgga acaacacaaa 3840
cactacctg atgagatcat cgagcaata agcgaattct ccaaagagt gatcctcgcc 3900
gacgtaacc tcgataaggt gtttctctg tacaataagc acagggataa gcccatcagg 3960
gagcaggcag aaaacattat ccacttgttt actctgacca acttgggcgc gcctgcagcc 4020

ttcaagtact tcgacaccac catagacaga aagecgtaca cctctacaaa ggaggtcctg 4080
gagccacac tgattcatca gtcaattacg gggtctatg aaacaagaat cgacctctct 4140
cagctcgggtg gagacaagcg tcctgctgct actaagaaag ctggtcaagc taagaaaaag 4200
aaataa 4206

<210> 11

<211> 20

<212> DNA

<213> Artificial

<220>

<223> β -globin RE2sgRNA1

<400> 11

accaatgac ctcaggctgt 20

<210> 12

<211> 20

<212> DNA

<213> Artificial

<220>

<223> β -globin RE2sgRNA2

<400> 12

tcacttgta gcggcatctg 20

<210> 13

<211> 23

<212> DNA

<213> Artificial

<220>

<223> Cas9- K918A-F

<400> 13

aggcttcac gccaggcagc ttg 23

<210> 14

<211> 18

<212> DNA

<213> Artificial

<220>

<223> Cas9- K918A-R

<400> 14

gctttatcca actcagac 18

<210> 15

<211> 86

<212> DNA

<213> Artificial
<220>
<223> Hiseq-RRM-1F3
<400> 15
aatgatacgg cgaccaccga gatctaacct ctttcctac acgacgctct tccgatctat 60
atggcatcct agccttaaga aactag 86
<210> 16
<211> 81
<212> DNA
<213> Artificial
<220>
<223> Hiseq-RRM-1R2
<400> 16
aatgatacgg cgaccaccga gatctaacct ctttcctac acgacgctct tccgatctta 60
cgacgcagga gccgtatcat g 81
<210> 17
<211> 89
<212> DNA
<213> Artificial
<220>
<223> Hiseq-RRM-3F2
<400> 17
caagcagaag acggcatacg agataagcta gtgactggag ttcagacgtg tgctcttccg 60
atctatagca atgaaatctt gaaggagtg 89
<210> 18
<211> 85
<212> DNA
<213> Artificial
<220>
<223> Hiseq-RRM-3R2
<400> 18
caagcagaag acggcatacg agattcaagt gtgactggag ttcagacgtg tgctcttccg 60
atctgcacag ccctgeteta ttacg 85
<210> 19
<211> 21
<212> DNA
<213> Artificial
<220>
<223> β -globin RE1sgRNA1

<400> 19
gattgttggt gccttgagt g 21
<210> 20
<211> 21
<212> DNA
<213> Artificial
<220>
<223> β -globin RE1sgRNA2
<400> 20
gctggcccc tgtaacctg g 21
<210> 21
<211> 24
<212> DNA
<213> Artificial
<220>
<223> β -globin RE1sgRNA1F
<400> 21
accgattgtt gttgccttg agtg 24
<210> 22
<211> 24
<212> DNA
<213> Artificial
<220>
<223> β -globin RE1sgRNA1R
<400> 22
aaaccactcc aaggcaaca caat 24
<210> 23
<211> 24
<212> DNA
<213> Artificial
<220>
<223> β -globin RE1sgRNA2F
<400> 23
accgctggtc ccctggtaac ctgg 24
<210> 24
<211> 24
<212> DNA
<213> Artificial
<220>

<223> β -globin RE1sgRNA2R
<400> 24
aaaccaggt taccagggga ccag 24
<210> 25
<211> 81
<212> DNA
<213> Artificial
<220>
<223> Hiseq-hSTM-aF1
<400> 25
aatgatacgg cgaccaccga gatctacact ctttcctac acgacgetct tccgatcttg 60
cttagagcca ggactaattg c 81
<210> 26
<211> 83
<212> DNA
<213> Artificial
<220>
<223> Hiseq-hSTM-aR2
<400> 26
aatgatacgg cgaccaccga gatctacact ctttcctac acgacgetct tccgatcttg 60
ggtgtagaaa tgagcaaata agt 83
<210> 27
<211> 91
<212> DNA
<213> Artificial
<220>
<223> Hiseq-hSTM-2F
<400> 27
caagcagaag acggcatacg agatgatcgt gtgactggag ttcagacgtg tgctcttccg 60
atctagattg agttctgttt gtttcacteta c 91
<210> 28
<211> 85
<212> DNA
<213> Artificial
<220>
<223> Hiseq-hSTM-2R
<400> 28
caagcagaag acggcatacg agatagtcaa gtgactggag ttcagacgtg tgctcttccg 60
atctcagctc tgctgaaag gagtc 85

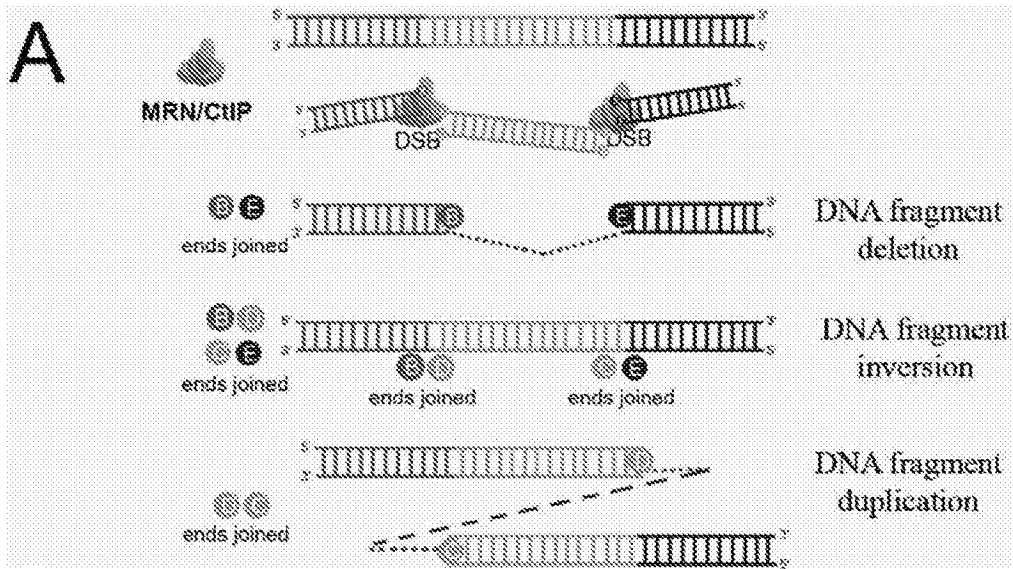


图1A

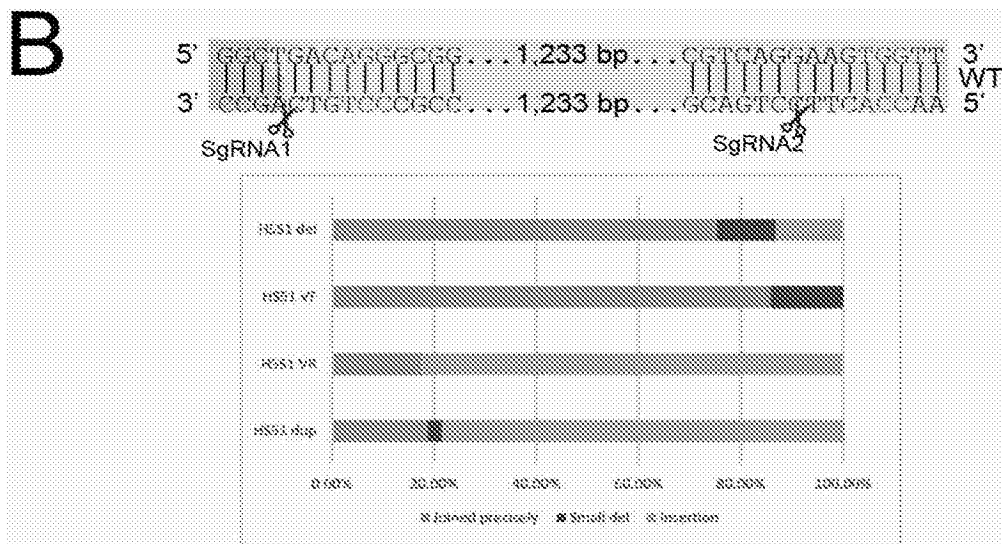


图1B

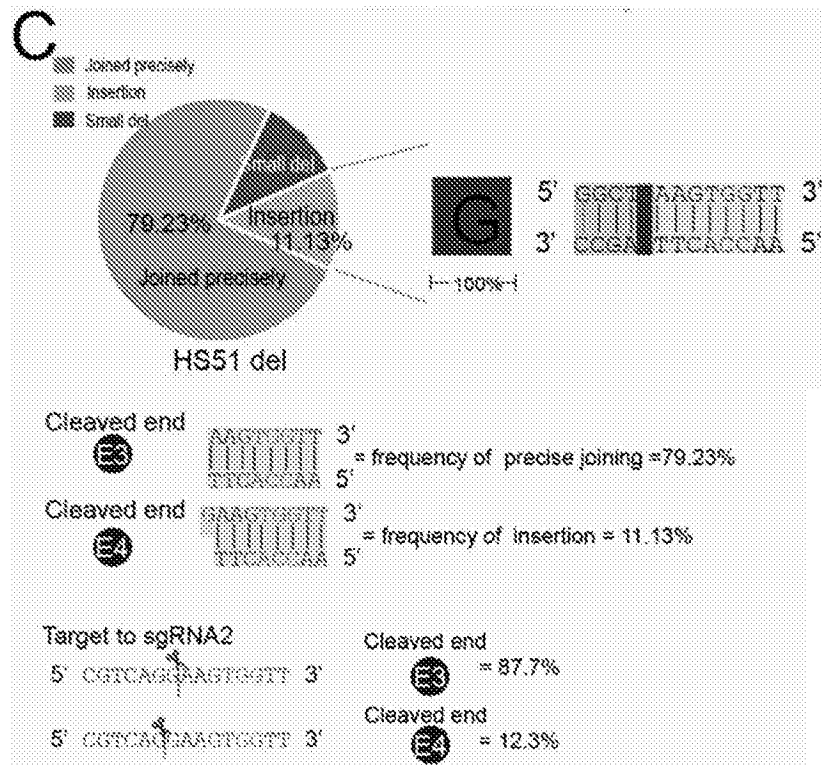


图1C

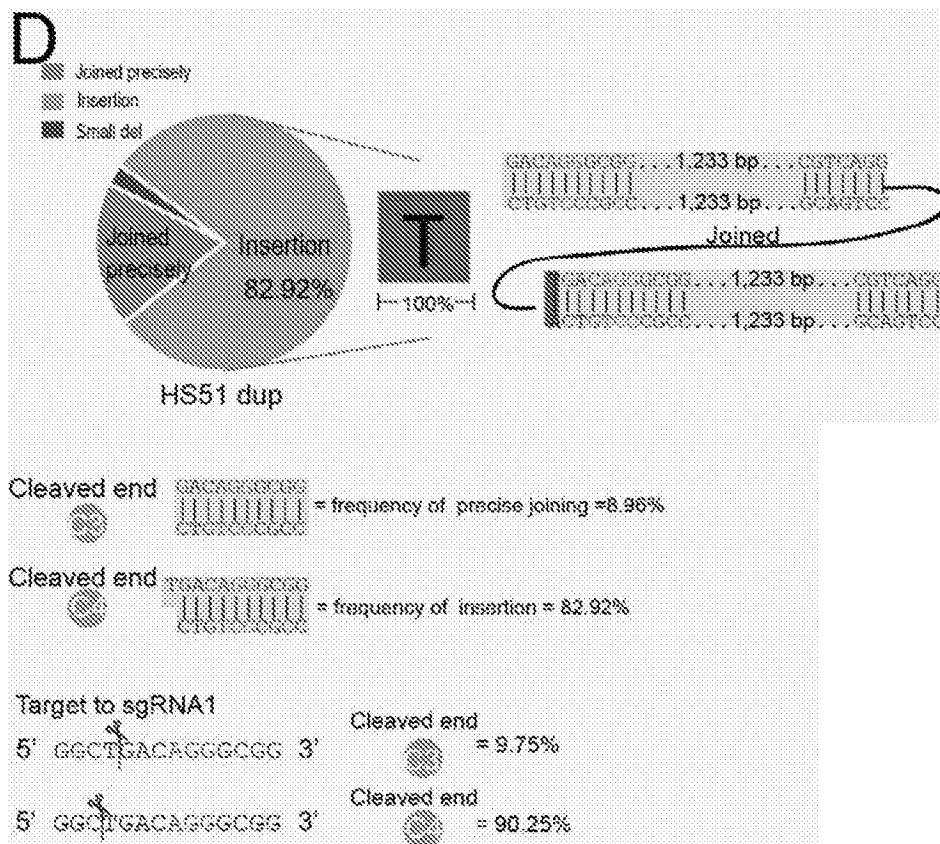


图1D

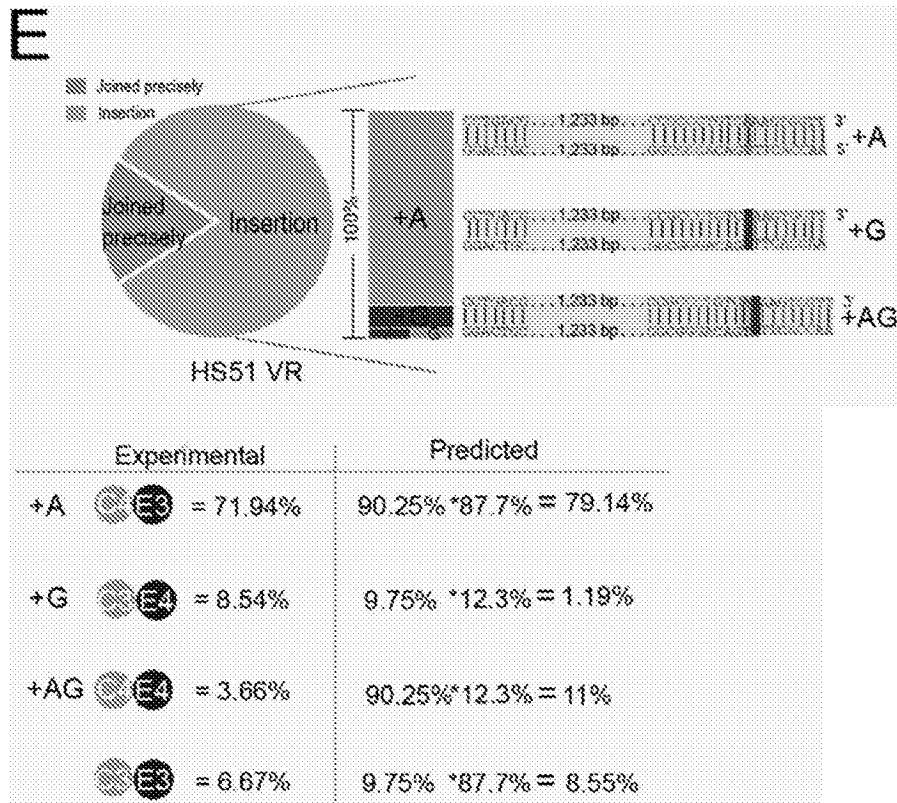


图1E

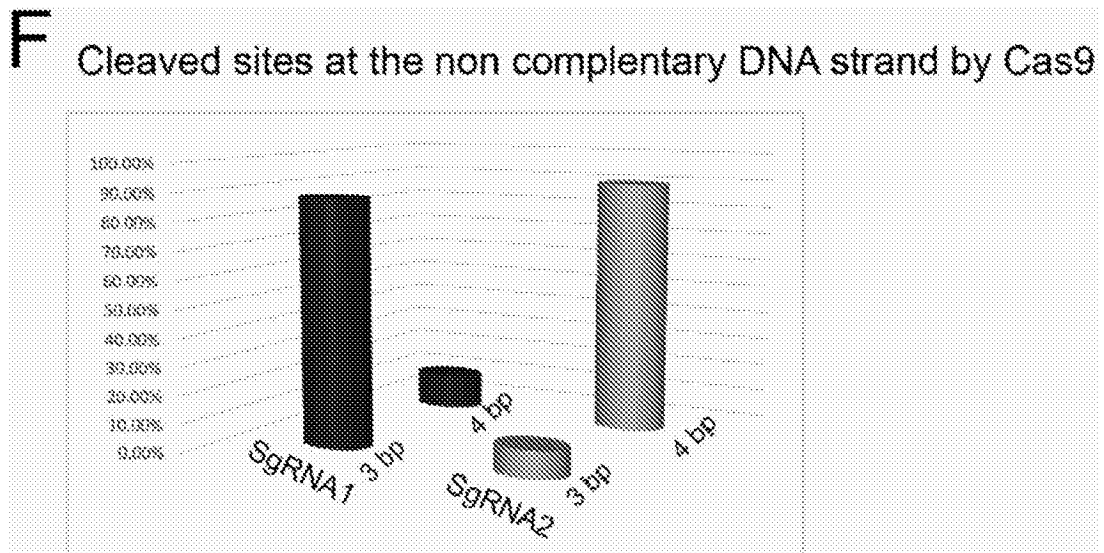


图1F

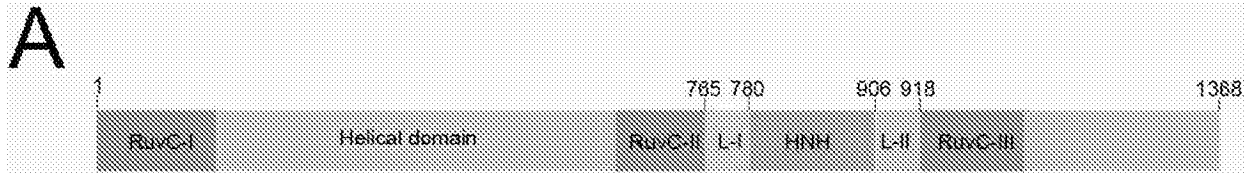


图2A

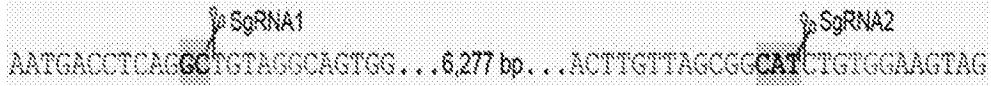


图2B

Insertion at the duplication junction

		Cas9	
		WT	K918A
		1	2
SgRNA1	+	82.76	86.28
	+C	5.00	2.25
	+GC	0.00	0.00
	+	0.00	0.00

图2C

β -globin RE2

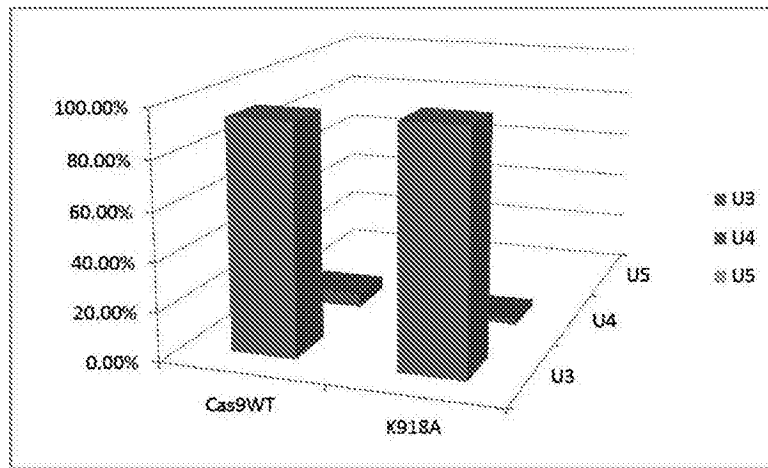


图2D

Insertion at the deletion junction

		Cas9	
		WT	K918A
SgRNA2	Inserted nt	1	2
	None	22.57	17.57
	+T	19.24	27.48
	+AT	33.43	34.98
	+CAT	3.29	2.04
	None		

图2E

β -globin RE2

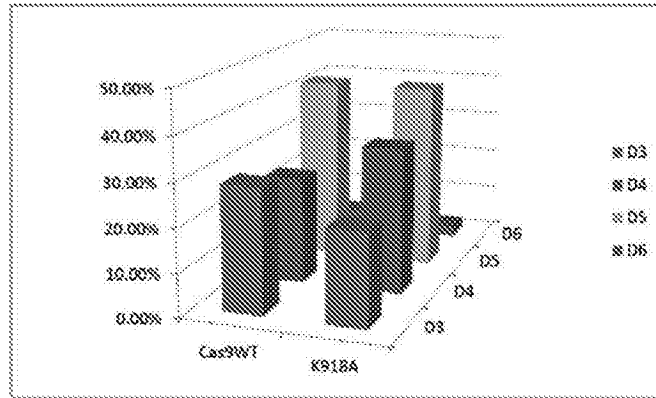


图2F

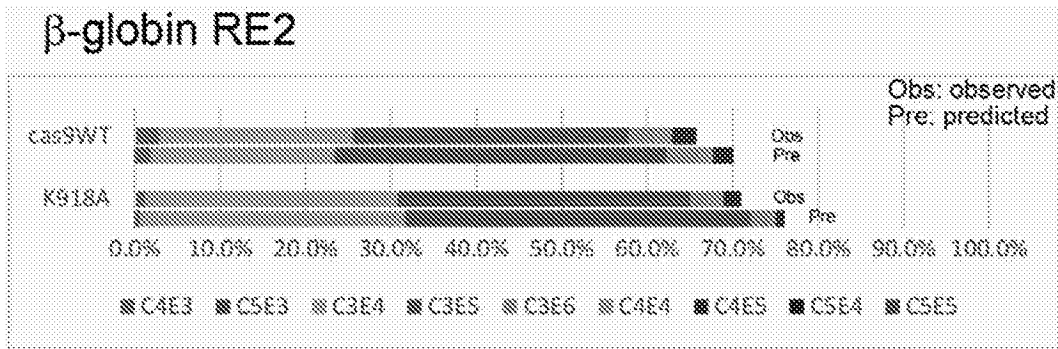


图2G

A STM

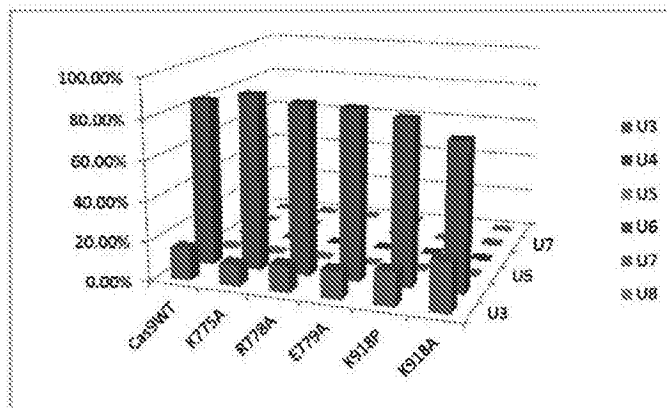


图3A

B STM

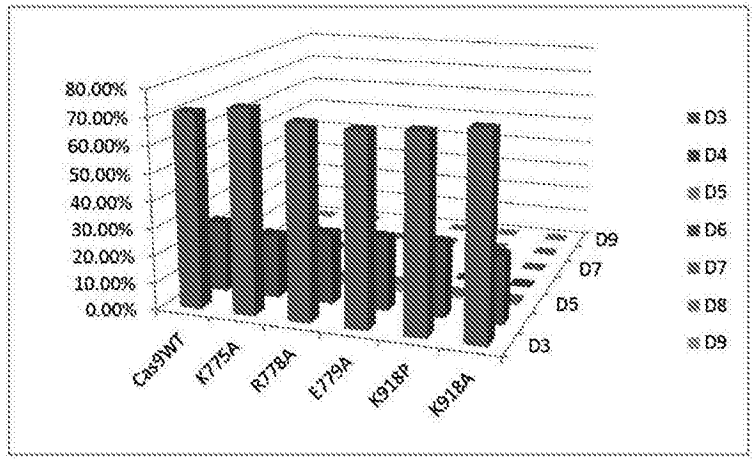


图3B