



(12) 发明专利

(10) 授权公告号 CN 106957830 B

(45) 授权公告日 2020.12.25

(21) 申请号 201710343377.0

(22) 申请日 2017.05.16

(65) 同一申请的已公布的文献号

申请公布号 CN 106957830 A

(43) 申请公布日 2017.07.18

(73) 专利权人 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(普通

合伙) 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)

(56) 对比文件

Jia Shou et al., "Precise and Predictable DNA Fragment Editing Reveals Principles of Cas9-mediated Nucleotide Insertion".《bioRxiv》.2017,第1-55页尤其是摘要、Fig. 4、S5、Fig. 7、Fig. S2、Fig. S3、第3页最后一段.

审查员 汪豪杰

权利要求书1页 说明书28页

序列表19页 附图8页

(54) 发明名称

一种Cas9核酸酶 Δ F916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- Δ F916),具有Cas9核酸酶活性,适用于CRISPR/Cas9系统,所述Cas9核酸酶(Cas9- Δ F916)是将野生型Cas9核酸酶第916位的苯丙氨酸删除后获得。采用所述Cas9核酸酶(Cas9- Δ F916)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核酸酶用于制备CRISPR/Cas9系统的用途,所述Cas9核酸酶的氨基酸序列如SEQ ID NO.9所示,所述CRISPR/Cas9系统具有针对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功能,所述CRISPR/Cas9系统可用于基因组DNA片段编辑。

2. 根据权利要求1所述的用途,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2所述的用途,其特征在于,所述野生型Cas9核酸酶为SpCas9。

4. 根据权利要求3所述的用途,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所述的用途,其特征在于,所述编辑包括单位点编辑或多位点编辑,所述多位点编辑的编辑位点数为两个及以上。

6. 根据权利要求1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。

7. 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中任一项所述的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。

8. 根据权利要求7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 Δ F916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 Δ F916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简单,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- Δ F916),适用于CRISPR/Cas9系统,所述Cas9核酸酶(Cas9- Δ F916)是将野生型Cas9核酸酶第916位的苯丙氨酸删除后获得。

[0009] 优选地,与野生型Cas9核酸酶相比,所述Cas9核酸酶(Cas9- Δ F916)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- Δ F916)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- Δ F916)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- Δ F916)。
- [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- Δ F916)的方法,包括步骤:构建含有Cas9核酸酶 (Cas9- Δ F916)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- Δ F916)。
- [0018] 本发明的第六方面,提供前述Cas9核酸酶 (Cas9- Δ F916)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- Δ F916)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- Δ F916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- Δ F916)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- Δ F916)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- Δ F916)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- Δ F916),适用于CRISPR/Cas9系统,所述Cas9核酸酶 (Cas9- Δ F916)含有如SEQ ID NO.9所示的氨基酸序列,所述Cas9核酸酶 (Cas9- Δ F916)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- Δ F916)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- Δ F916), 具有Cas9核酸酶活性, 适用于CRISPR/Cas9系统, 所述Cas9核酸酶 (Cas9- Δ F916) 是将野生型Cas9核酸酶第916位苯丙氨酸删除后获得。所述Cas9核酸酶 (Cas9- Δ F916) 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- Δ F916) 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- Δ F916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- Δ F916) 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- Δ F916) 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- Δ F916) 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- Δ F916) 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- Δ F916) 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- Δ F916) 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- Δ F916) 的方法

[0058] 制备前述Cas9核酸酶 (Cas9- Δ F916) 的方法,包括步骤:构建含有Cas9核酸酶 (Cas9- Δ F916) 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- Δ F916)。

[0059] 本领域技术人员可根据Cas9核酸酶 (Cas9- Δ F916) 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- Δ F916) 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- Δ F916) 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、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- Δ F916) 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- Δ F916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- Δ F916) 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- Δ F916) 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- Δ F916)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 Δ F916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 Δ F916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- Δ F916、 Δ F916、 Δ F916突变体之间可替换使用,意思均为名称为 Δ F916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:GAGATTTGGGCGTCAGGAAG (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℃。
- [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℃,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,将制备好的人源化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] ATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGCAAGGAGATCGTGTCGTC (SEQ ID NO.3)

[0122] Hiseq-hs51-aRa:

[0123] AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTAAGGATGTTGTGGAAGGCGAGCAG (SEQ ID NO.4)

[0124] Hiseq-hs51-bFa:

[0125] CAAGCAGAAGACGGCATACGAGATGGACGGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTTTACATGACAGCTTCCGGTAG (SEQ ID NO.5)

[0126] Hiseq-hHs51-bR:

[0127] CAAGCAGAAGACGGCATACGAGATTTGACTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTTTTTGGCTAACAACATAGTGCTTC (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:

[0130] Min coverage:8

[0131] Min reads2:2

[0132] Min var freq:0.01

[0133] Min avg qual:15

[0134] P-value thresh:0.01。

[0135] 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。

[0136] 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。

[0137] 如图1B所示,针对基因组DNA片段HS51 RE1 (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)。

[0138] 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%。

[0139] 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。

[0140] 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%。

[0141] 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。

[0142] 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%÷(79.23%+11.13%)】。突出断裂末端E4的比例为12.3%【计算方法为:11.13%÷(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%。

[0143] 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%。

[0144] 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(HS51 RE1,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。

[0145] 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%。

[0146] 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。

[0147] 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%。

[0148] 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。

[0149] 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。

[0150] 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%)×“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。

[0151] 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。

[0152] 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。

[0153] 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)。

[0154] 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。

[0155] 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。

[0156] 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1. 构建Cas9突变体

[0157] 1) 使用NEB突变试剂盒(Q5 Site-Directed Mutagenesis Kit, #E0554S) 构建Cas9突变体，首先进行PCR扩增，反应如下：

[0158]	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
	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℃, 30 sec;

[0159]	98℃, 10sec	} 25 个循环
	60℃, 20sec	
	72℃, 6.5min	
	72℃, 3min	

[0160] 2) KLD (Kinase, Ligase&DpnI) 处理, 反应如下:

[0161]	PCR 产物	0.5 μ l
	2X KLD Reaction Buffer	2.5 μ l
	10X KLD Enzyme Mix	0.5 μ l
	Nuclease-free Water	1.5 μ l
	Total volume	5 μ l

[0162] 反应条件: 室温10分钟

[0163] 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℃。挑取单克隆, 质粒提取后送测序。

[0164] SpCas9 (Cas9WT) 的氨基酸序列如SEQ ID NO.7所示, 具体为:

```

MAPKKKRKVG IHGVPAAMDK KYSIGLDIGT NSVGWAVITD EYKVPSKKFK VLGNTDRHSI
KKNLIGALLF DSGETAETR LKRTARRRYT RRKNRICYLQ EIFSNEMAKV DDSFFHRLEE
SFLVEEDKKH ERHPIFGNIV DEVAYHEKYP TIYHLRKKLV DSTDKADLRL IYLALAHMIK
FRGHFLIEGD LNPDNSDVK LFIQLVQTYN QLFEENPINA SGVDAKAILS ARLSKSRRL
[0165] NLIAQLPGEK KNGLFGNLIA LSLGLTPNFK SNFDLAEDAK LQLSKDTYDD DLDNLLAQIG
DQYADLFLAA KNLSDAILLS DILRVNTEIT KAPLSASMIK RYDEHHQDLT LLKALVRQQL
PEKYKEIFFD QSKNGYAGYI DGGASQEEFY KFIKPILEKM DGTEELLVKL NREDLLRKQR
TFDNGSIPHQ IHLGELHAIL RRQEDFYFPL KDNREKIEKI LTFRIPYYVG PLARGNSRFA
WMTRKSEETI TPWNFEVVD KGASAQSFIE RMTNFDKNLP NEKVLPKHSL LYEYFTVYNE

```

LTKVKYVTEG MRKPAFLSGE QKKAIVDLLF KTNRKVTVKQ LKEDYFKKIE CFDSVEISGV
EDRFNASLGT YHDLKIIKD KDFLDNEENE DILEDIVLTL TLFEDREMIE ERLKTYAHLF
DDKVMKQLKR RRYTGWGRLS RKLINGIRDK QSGKTILDFL KSDGFANRNF MQLIHDDSLT
FKEDIQKAQV SGQGDSLHEH IANLAGSPAI KKGILQTVKV VDELVKVMGR HKPENIVIEM
ARENQTTQKG QKNSRERMKR IEEGIKELGS QILKEHPVEN TQLQNEKLYL YYLQNGRDMY
VDQELDINRL SDYDVDHIVP QSFLKDDSID NKVLTRSDKN RGKSDNVPSE EVVKMKMKNYW
RQLLNAKLIT QRKFDNLTKA ERGGLSELDK AGFIKRLVE TRQITKHVAQ ILDSRMNTKY
[0166] DENDKLIREV KVITLKSCLV SDFRKDFQFY KVREINNYHH AHDAYLNAVV GTALIKKYPK
LESEFVYGDY KVIDVRKMIA KSEQEIGKAT AKYFFYSNIM NFFKTEITLA NGEIRKRPLI
ETNGETGEIV WDKGRDFATV RKVLSMPQVN IVKKTEVQTG GFSKESILPK RNSDKLIARK
KDWDPKKYGG FDSPTVAYSV LVVAKVEKGK SKKLKSVKEL LGITIMERSS FEKNPIDFLE
AKGYKEVKKD LIIKLPKYSL FELENGRKRM LASAGELQKG NELALPSKYV NFLYLASHYE
KLKGSPEDNE QKQLFVEQHK HYLDEIIEQI SEFSKRVILA DANLDKVLISA YNKHRDKPIR
EQAENIIHLF TLTNLGAPAA FKYFDTTIDR KRYTSTKEVL DATLIHQSIIT GLYETRIDL
QLGGDKRPAA TKKAGQAKKKK。

[0167] SpCas9 (Cas9WT) 的编码核苷酸序列如SEQ ID NO.8所示,具体为:

atggccccc aaagaagcg gaaggctcgt atccacggtg tcccagcagc catggacaag
aagtactcca ttgggctcga tatcggcaca aacagcgctc gctgggcccgt cattacggac
gagtacaagg tgccgagcaa aaaattcaaa gttctgggca ataccgatcg ccacagcata
aagaagaacc tcattggcgc cctcctgttc gactccgggg agacggccga agccacgcgg
ctcaaaagaa cagcacggcg cagatatacc cgcagaaaga atcggatctg ctacctgcag
gagatcttta gtaatgagat ggctaagggt gatgactctt tcttccatag gctggaggag
tccttttttg tggaggagga taaaaagcac gagcgccacc caatcttttg caatatcgtg
[0168] gacgaggtgg cgtaccatga aaagtaccca accatatatc atctgaggaa gaagcttgta
gacagtactg ataaggctga cttgcggttg atctatctcg cgctggcgca tatgatcaaa
tttcggggac acttcctcat cgagggggac ctgaaccag acaacagcga tgtcgacaaa
ctctttatcc aactggttca gacttacaat cagcttttcg aagagaaccc gatcaacgca
tccggagttg acgccaagc aatcctgagc gctaggctgt ccaaatcccg gcggtcga
aacctcatcg cacagctccc tggggagaag aagaacggcc tgtttggtaa tcttatcgcc
ctgtcactcg ggctgacccc caactttaaa tctaacttcg acctggccga agatgccaa

[0169]

cttcaactga gcaaagacac ctacgatgat gatctcgaca atctgctggc ccagatcggc
gaccagtacg cagacctttt ttggcgga aagaacctgt cagacgccat tctgctgagt
gatattctgc gagtgaacac ggagatcacc aaagctccgc tgagcgctag tatgatcaag
cgctatgatg agcaccacca agacttgact ttgctgaagg cccttgctcag acagcaactg
cctgagaagt acaaggaaat tttcttcgat cagtctaaaa atggctacgc cggatacatt
gacggcggag caagccagga ggaattttac aaatttatta agcccatctt ggaaaaaatg
gacggcaccg aggagctgct ggtaaagctt aacagagaag atctgttgcg caaacagcgc
acttttcgaca atggaagcat cccccaccag attcacctgg gcgaactgca cgctatactc
aggcggcaag aggatcttcta cccctttttg aaagataaca gggaaaagat tgagaaaatc
ctcacatttc ggatacccta ctatgtaggc cccctcgccc ggggaaattc cagattcgcg
tggtatgactc gcaaatacaga agagaccatc actccctgga acttcgagga agtcgtggat
aagggggcct ctgcccagtc cttcatcgaa aggatgacta actttgataa aaatctgcct
aacgaaaagg tgcttcctaa aactctctg ctgtacgagt acttcacagt ttataacgag
ctcaccaagg tcaaatacgt cacagaaggg atgagaaaagc cagcattcct gtctggagag
cagaagaaaag ctatcgtgga cctcctcttc aagacgaacc ggaaagtac cgtgaaacag
ctcaaagaag actatttcaa aaagattgaa tgtttcgact ctgttgaaat cagcggagtg
gaggatcgct tcaacgcac cctgggaacg tatcacgac tctgaaaat cattaaagac
aaggacttcc tggacaatga ggagaacgag gacattcttg aggacattgt cctcaccctt
acgttgtttg aagataggga gatgattgaa gaacgcttga aaacttacgc tcatctcttc
gacgacaaaag tcatgaaaca gctcaagagg cgccgatata caggatgggg gggctgtca
agaaaactga tcaatgggat ccgagacaag cagagtggaa agacaatcct ggattttctt
aagtccgatg gatttgccaa ccggaacttc atgcagttga tccatgatga ctctctcacc
tttaaggagg acatccagaa agcacaagtt tctggccagg gggacagtct tcacgagcac
atcgctaate ttgcaggtag ccagctatc aaaaaggga tactgcagac cgtaaggtc
gtggatgaac tcgtcaaagt aatgggaagg cataagccc agaatatcgt tatcgagatg
gcccgagaga accaaactac ccagaaggga cagaagaaca gtagggaaaag gatgaagagg
attgaagagg gtataaaaga actgggtcc caaatcctta aggaacaccc agttgaaaac
accagcttc agaatagaga gctctacctg tactacctgc agaacggcag ggacatgtac
gtggatcagg aactggacat caatcggctc tccgactacg acgtggatca tatcgtgccc
cagtcttttc tcaaagatga ttctattgat aataaagtgt tgacaagatc cgataaaaat

agaggggaaga gtgataacgt cccctcagaa gaagttgtca agaaaatgaa aaattattgg
cggcagctgc tgaacgccaa actgatcaca caacggaagt tcgataatct gactaaggct
gaacgaggtg gcctgtctga gttggataaa gcaggcttca tcaaaaggca gcttggtgag
acacgccaga tcaccaagca cgtggcccaa attctcgatt cacgcatgaa caccaagtac
gatgaaaatg acaaactgat tcgagaggtg aaagttatta ctctgaagtc taagctggtc
tcagatttca gaaaggactt tcagttttat aagggtgagag agatcaacaa ttaccaccat
gcgcatgatg cctacctgaa tgcagtggta ggcactgcac ttatcaaaaa atatcccaag
cttgaatctg aatttgttta cggagactat aaagtgtacg atgttaggaa aatgatcgca
aagtctgagc aggaaatagg caaggccacc gctaagtact tcttttacag caatattatg
aattttttca agaccgagat tacactggcc aatggagaga ttcggaagcg accacttattc
gaaacaaacg gagaaacagg agaaatcgtg tgggacaagg gtagggattt cgcgacagtc
cggaaggtcc tgtccatgcc gcaggtgaac atcgttaaaa agaccgaagt acagaccgga
ggcttctcca aggaaagtat cctcccgaaa aggaacagcg acaagctgat cgcacgcaaa
[0170] aaagattggg accccaagaa atacggcgga ttcgattctc ctacagtcgc ttacagtgtg
ctggttgtgg ccaaagtgga gaaagggaag tctaaaaaac tcaaaagcgt caagggaactg
ctgggcatca caatcatgga gcgatcaagc ttcgaaaaaa accccatcga ctttctcgag
gcgaaaggat ataaagaggt caaaaaagac ctcatcatta agcttcccaa gtactctctc
tttgagcttg aaaacggccg gaaacgaatg ctgctagtgc cgggcgagct gcagaaaggc
aacgagctgg cactgccctc taaatacggtt aatttcttgt atctggccag ccactatgaa
aagctcaaag ggtctcccga agataatgag cagaagcagc tgttcgtgga acaacacaaa
cactaccttg atgagatcat cgagcaaata agcgaattct ccaaaagagt gatcctcgcc
gacgctaacc tcgataaggt gctttctgct tacaataagc acagggataa gcccatcagg
gagcaggcag aaaacattat ccacttgttt actctgacca acttgggcgc gcctgcagcc
ttcaagtact tcgacaccac catagacaga aagcggtaga cctctacaaa ggaggtcctg
gacgccacac tgattcatca gtcaattacg gggctctatg aaacaagaat cgacctctct
cagctcgggtg gagacaagcg tcctgctgct actaagaaag ctggtcaagc taagaaaaag
aaataa。

[0171] 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。

[0172] 本发明要求保护的Cas9核酸酶突变体命名为Cas9-ΔF916(将SpCas9核酸酶第916位苯丙氨酸删除后获得),

[0173] Cas9-ΔF916的氨基酸序列SEQ ID NO.9所示,具体为:

[0174]

MAPKKKRKVG IHGVPAAMDK KYSIGLDIGT NSVGWAVITD EYKVPSKKFK VLGNTRHSHI
 KKNLIGALLF DSGETAETR LKRTARRRYT RRKNRICYLQ EIFSNEMAKV DDSFFHRLEE
 SFLVEEDKKH ERHPIFGNIV DEVAYHEKYP TIYHLRKKLV DSTDKADLRL IYLALAHMIK
 FRGHFLIEGD LNPDNSDVK LFIQLVQTYN QLFEENPINA SGVDAKAILS ARLSKSRRL
 NLIAQLPGEK KNGLFGNLIA LSLGLTPNFK SNFDLAEDAK LQLSKDTYDD DLDNLLAQIG
 DQYADFLAA KNLSDAILLS DILRVNTEIT KAPLSASMIK RYDEHHQDLT LLKALVRQQL
 PEKYKEIFFD QSKNGYAGYI DGGASQEEFY KFIKPILEKM DGTEELLVKL NREDLLRKQR
 TFDNGSIPHQ IHLGELHAIL RRQEDFYFPL KDNREKIEKI LTFRIPYYVG PLARGNSRFA
 WMTRKSEETI TPWNFEVVD KGASAQSFIE RMTNFDKNLP NEKVLPKHSL LYEYFTVYNE
 LTKVKYVTEG MRKPAFLSGE QKKAIVDLLF KTNRKVTVKQ LKEDYFKKIE CFDSVEISGV
 EDRFNASLGT YHDLKIIKD KDFLDNEENE DILEDIVLTL TLFEDREMIE ERLKTYAHLF
 DDKVMKQLKR RRYTGWGRSL RKLINGIRDK QSGKTILDFL KSDGFANRNF MQLIHDDSLT
 FKEDIQKAQV SGQGDSLHEH IANLAGSPAI KKGILQTVKV VDELVKVMGR HKPENIVIE
 ARENQTTQKG QKNSRERMKR IEEGIKELGS QILKEHPVEN TQLQNEKLYL YYLQNGRDMY
 VDQELDINRL SDYDVHIVP QSFLKDDSID NKVLTRSDKN RGKSDNVPSE EVVKMKKNYW
 RQLLNAKLIT QRKFDNLTKA ERGGLSELDK AGIKRQLVET RQITKHVAQI LDSRMNTKYD
 ENDKLIREVK VITLKSCLVS DFRKDFQFYK VREINNYHHA HDAYLNAVVG TALIKKYPKL
 ESEFVYGDYK VYDVRKMIK SEQEIGKATA KYFFYSNIMN FFKTEITLAN GEIRKRPLIE
 TNGETGEIVW DKGRDFATVR KVLSPQVNI VKKTEVQTGG FSKESILPKR NSDKLIARKK
 DWDPKKYGGF DSPTVAYSVL VVAKVEKGKS KKLKSVKELL GITIMERSSE EKNPIDFLEA
 KGYKEVKKDL IIKLPKYSLF ELENKRKRL ASAGELQKGN ELALPSKYVN FLYLASHYEK
 LKGSPEDNEQ KQLFVEQHKH YLDEIIEQIS EFSKRVLAD ANLDKVL SAY NKHRDKPIRE
 QAENIIHLFT LTNLGAPAAF KYFDTTIDRK RYTSTKEVLD ATLIHQ SITG LYETRIDL SQ
 LGGDKRPAAT KKAGQAKKKK。

[0175] Cas9-ΔF916的编码核苷酸序列如SEQ ID NO.10所示,具体为:

atggccccaagaagaagcggaaggtcggatccacgggtgccagcagccatggacaag

[0176] aagtactcca ttgggctcga tatcggcaca aacagcgctcg gctgggccgt cattacggac
 gagtacaagg tgccgagcaa aaaattcaaa gttctgggca ataccgatcg ccacagcata

[0177]

aagaagaacc tcattggcgc cctcctgttc gactccgggg agacggccga agccacgcgg
ctcaaaagaa cagcacggcg cagatatacc cgcagaaaga atcggatctg ctacctgcag
gagatcttta gtaatgagat ggctaagggtg gatgactctt tcttccatag gctggaggag
tccttttttg tggaggagga taaaaagcac gagcgccacc caatcttttg caatatcgtg
gacgaggtgg cgtaccatga aaagtaccca accatatatc atctgaggaa gaagcttgta
gacagtactg ataaggctga ctgcggttg atctatctcg cgctggcgca tatgatcaaa
tttcggggac acttcctcat cgagggggac ctgaaccag acaacagcga tgtcgacaaa
ctctttatcc aactgggtca gacttacaat cagcttttcg aagagaaccc gatcaacgca
tccggagttg acgcaaagc aatcctgagc gctaggctgt ccaaattccg gcggctcgaa
aacctcatcg cacagctccc tggggagaag aagaacggcc tgtttggtta tcttatcgcc
ctgtcactcg ggctgacccc caactttaaa tctaacttcg acctggccga agatgccaaag
cttcaactga gcaaagacac ctacgatgat gatctcgaca atctgctggc ccagatcggc
gaccagtacg cagacctttt tttggcgga aagaacctgt cagacgcat tctgctgagt
gatattctgc gagtgaacac ggagatcacc aaagctccgc tgagcgctag tatgatcaag
cgctatgatg agcaccacca agacttgact ttgctgaagg cccttgctcag acagcaactg
cctgagaagt acaaggaaat tttcttcgat cagtctaaaa atggctacgc cggatacatt
gacggcggag caagccagga ggaattttac aaatttatta agcccatctt ggaaaaaatg
gacggcaccg aggagctgct ggtaaagctt aacagagaag atctgttgcg caaacagcgc
actttcgaca atggaagcat cccccaccag attcacctgg gcgaactgca cgctatactc
aggcggcaag aggatttcta cccctttttg aaagataaca gggaaaagat tgagaaaatc
ctcacatttc ggatacccta ctatgtaggc cccctcgccc ggggaaattc cagattcgcg
tggatgactc gcaaatacaga agagaccatc actccctgga acttcgagga agtcgtggat
aagggggcct ctgcccagtc ctcatcgaa aggatgacta actttgataa aaatctgcct
aacgaaaagg tgcttcctaa acactctctg ctgtacgagt acttcacagt ttataacgag
ctcaccaagg tcaaatacgt cacagaaggg atgagaaagc cagcattcct gtctggagag
cagaagaaag ctatcgtgga cctcctcttc aagacgaacc ggaaagtac cgtgaaacag
ctcaaagaag actatttcaa aaagattgaa tgtttcgact ctgttgaaat cagcggagtg
gaggatcgct tcaacgcac cctgggaacg tatcacgac tctgaaaat cattaagac
aaggacttcc tggacaatga ggagaacgag gacattcttg aggacattgt cctcaccctt
acgttgtttg aagataggga gatgattgaa gaacgcttga aaacttacgc tcatctcttc

[0178]

gacgacaaag tcatgaaaca gctcaagagg cgccgatata caggatgggg gcggctgtca
agaaaactga tcaatgggat ccgagacaag cagagtggaa agacaatcct ggatttttctt
aagtccgatg gatttgccaa ccggaacttc atgcagttga tccatgatga ctctctcacc
tttaaggagg acatccagaa agcacaagtt tctggccagg gggacagtct tcacgagcac
atcgctaato ttgcaggtag ccagctatc aaaaaggga tactgcagac cgtaagggtc
gtggatgaac tcgtcaaagt aatgggaagg cataagcccg agaatatcgt tatcgagatg
gcccagagaga accaaactac ccagaaggga cagaagaaca gtagggaaag gatgaaggag
attgaaggagg gtataaaaga actgggggtcc caaatcctta aggaacaccc agttgaaaac
accagcttc agaatgagaa gctctacctg tactacctgc agaacggcag ggacatgtac
gtggatcagg aactggacat caatcggtc tccgactacg acgtggatca tatcgtgccc
cagtcctttt tcaaagatga ttctattgat aataaagtgt tgacaagatc cgataaaaaat
agagggaaga gtgataacgt cccctcagaa gaagttgtca agaaaatgaa aaattattgg
cggcagctgc tgaacgcaa actgatcaca caacggaagt tcgataatct gactaaggct
gaacgagggtg gcctgtctga gttggataaa gcaggcatca aaaggcagct tgttgagaca
cgccagatca ccaagcacgt ggcccaaatt ctcgattcac gcatgaacac caagtacgat
gaaaatgaca aactgattcg agagggtgaaa gttattactc tgaagtctaa gctggtctca
gatttcagaa aggactttca gttttataag gtgagagaga tcaacaatta ccaccatgcy
catgatgcct acctgaatgc agtggtaggc actgcactta tcaaaaaata tcccaagctt
gaatctgaat ttgtttacgg agactataaa gtgtacgatg ttaggaaaat gatcgcaaag
tctgagcagg aaataggcaa ggccaccgct aagtacttct ttacagcaa tattatgaat
tttttcaaga ccgagattac actggccaat ggagagattc ggaagcgacc acttatcgaa
acaaacggag aaacaggaga aatcgtgtgg gacaagggtg gggatttcgc gacagtccgg
aaggctcctgt ccatgccgca ggtgaacatc gttaaaaaga ccgaagtaca gaccggaggc
ttctccaagg aaagtatcct cccgaaaagg aacagcgaca agctgatcgc acgcaaaaaa
gattgggacc ccaagaaata cggcggattc gattctccta cagtcgctta cagtgtactg
gttgtggcca aagtggagaa agggaaagtct aaaaaactca aaagcgtcaa ggaactgctg
ggcatcacia tcatggagcg atcaagcttc gaaaaaaacc ccatcgactt tctcgaggcg
aaaggatata aagagggtcaa aaaagacctc atcattaagc ttcccaagta ctctctctt
gagcttgaaa acggccggaa acgaatgctc gctagtgcgg gcgagctgca gaaaggtaac
gagctggcac tgccctctaa atacgttaat ttcttgtatc tggccagcca ctatgaaaag

ctcaaaggggt ctcccgaaga taatgagcag aagcagctgt tcgtggaaca acacaaacac
 taccttgatg agatcatcga gcaaataagc gaattctcca aaagagtgat cctcgccgac
 gctaacctcg ataagggtgt ttctgcttac aataagcaca gggataagcc catcagggag
 caggcagaaa acattatcca cttgtttact ctgaccaact tgggcgcgcc tgcagccttc
 [0179] aagtacttcg acaccacat agacagaaag cggtacacct ctacaaagga ggtcctggac
 gccacactga ttcacagtc aattacgggg ctctatgaaa caagaatcga cctctctcag
 ctcggtggag acaagcgtcc tgctgctact aagaaagctg gtcaagctaa gaaaaagaaa
 taa。

[0180] 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- Δ F916的序列均不同。

[0181] 2.Cas9核酸酶突变体进行DNA片段编辑

[0182] (1) 针对 β -globin RE2 (RRM21位点),构建RRM21位点(β -globin RE2)的sgRNAs。

[0183] 所述sgRNAs靶向序列:

[0184] β -globin RE2 sgRNA1:ACCCAATGACCTCAGGCTGT (SEQ ID NO.11);

[0185] β -globin RE2 sgRNA2:TCACTTGTTAGCGGCATCTG (SEQ ID NO.12);

[0186] 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。

[0187] (2) 获得互补配对的带有悬挂端的双链DNA

[0188] 1) 用ddH₂O将脱氧寡核苷酸溶解至100 μ M,并稀释至20 μ M;

[0189] 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
[0190] NEB buffer2 (10x)	2 μ l
ddH ₂ O	14 μ l
Total volume	20 μ l

[0191] 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。

[0192] (3) 酶切pGL3-U6-sgRNA-PGK-Puro vector

[0193] 1) 用BsaI限制性内切酶酶切载体质粒,反应体系如下:

[0194] NEB buffer4 (10x)	2 μ l
BSA (10x)	2 μ l
质粒	~700 ng
[0195] BsaI	1 μ l
ddH ₂ O	12 μ l
Total volume	20 μ l

[0196] 反应条件:37℃,1.5小时;

[0197] 2) 胶回收纯化DNA酶切片段,按照胶回收试剂盒(Axygen)说明纯化。

[0198] (4) 连接酶切后的载体与带有悬挂端的双链DNA

[0199] 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
[0200]	NEB T4 ligase	0.25 μ l
	ddH ₂ O	3.25 μ l
	Total volume	5 μ l

[0201] 反应条件:室温反应1.5小时;

[0202] (5) 转化连接产物

[0203] 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℃。

[0204] (6) 挑取单克隆测序

[0205] 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;

[0206] 2) 质粒提取,按照质粒小抽试剂盒(Axygen)说明提取;

[0207] 3) 提取后的质粒送上海桑尼生物科技有限公司测序。

[0208] (7) 测序成功质粒进行中抽

[0209] 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;

[0210] 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;

[0211] 3) 收集细菌,按照质粒中抽试剂盒(Qiagen)说明提取质粒。

[0212] (8) 用Lipofectamine 2000进行细胞转染

[0213] 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;

[0214] 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。

[0215] 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。

[0216] (9) 制备高通量测序文库

[0217] 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。

[0218] Cas9突变引物:

[0219] Cas9- Δ F916-F:ATCAAAAGGCAGCTTGTTG (SEQ ID NO.13);

[0220] Cas9- Δ F916-R:GCCTGCTTTATCCAATC (SEQ ID NO.14);

[0221] (10) 高通量测序数据处理

[0222] 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:

[0223] Min coverage:8

[0224] Min reads:2

[0225] Min var freq:0.01

[0226] Min avg qual:15

[0227] P-value thresh:0.01。

[0228] 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。

[0229] 高通量引物:

[0230] HiSeq-RRM-1F3:

[0231] AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTATATGGCATCTAGCCTTAAGAACTAG (SEQ ID NO.15)

[0232] HiSeq-RRM-1R2:

[0233] AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTACGACGCAGGAGCCGTATCATG (SEQ ID NO.16)

[0234] HiSeq-RRM-3F2:

[0235] CAAGCAGAAGACGGCATACGAGATAAGCTAGTGAAGTTCAGACGTGTGCTCTTCCGATCTATAGCAATGAAATCTTGAAGGAGTG (SEQ ID NO.17)

[0236] HiSeq-RRM-3R2:

[0237] CAAGCAGAAGACGGCATACGAGATTCAAGTGTGAAGTTCAGACGTGTGCTCTTCCGATCTGCACAGCCCTGCTCTATTACG (SEQ ID NO.18)。

[0238] 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。

[0239] 具体地,野生型SpCas9核酸酶(简称Cas9WT,WT)(图2A)和 Δ F916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。

[0240] 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 Δ F916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17.25%，“+C”碱基的占

比为33.46%，+“GC”碱基的占比为11.14%。

[0241] 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 Δ F916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。

[0242] 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 Δ F916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。

[0243] 而当 Δ F916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)。

[0244] 因此，我们认为：在Cas9核酸酶 Δ F916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=17.25%。突出断裂末端U4的比例=“C”碱基的加入比例=33.46%。突出断裂末端U5的比例=“GC”碱基的加入比例=11.14%。

[0245] 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。

[0246] 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 Δ F916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27.89%【计算方法为： $17.25\% \div (17.25\% + 33.46\% + 11.14\%)$ 】。突出断裂末端U4的比例为54.10%【计算方法为： $33.46\% \div (17.25\% + 33.46\% + 11.14\%)$ 】。突出断裂末端U5的比例为18.01%【计算方法为： $11.14\% \div (17.25\% + 33.46\% + 11.14\%)$ 】。

[0247] 亦即，Cas9核酸酶 Δ F916在sgRNA1的介导下对基因组DNA片段进行切割的方式中，

U3钝末端切割的比例为27.89%，U4突出末端切割的比例为54.10%，U5突出末端切割的比例为18.01%。

[0248] 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：

[0249] 表2-1

[0250]	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%
	ΔF916	27.89%	54.10%	18.01%

[0251] 可见，在sgRNA1的介导下，相比于SpCas9核酸酶(Cas9WT)，ΔF916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提高(U4)，在PAM上游3bp处进行切割的比例减少(U3)。

[0252] 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。

[0253] 以选用ΔF916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10.96%，“+T”碱基的占比为22.89%，“+AT”碱基的占比为13.34%，“+CAT”碱基的占比为0.83%。

[0254] 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ΔF916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。

[0255] 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。

[0256] 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ΔF916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。

[0257] 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 Δ F916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“CAT”碱基的加入。

[0258] 而当 Δ F916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)。

[0259] 因此,我们认为:在Cas9核酸酶 Δ F916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=10.96%。突出断裂末端D4的占比=“T”碱基的加入占比=22.89%。突出断裂末端D5的占比=“AT”碱基的加入占比=13.34%。突出断裂末端D6的占比=“CAT”碱基的加入占比=0.83%。

[0260] 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。

[0261] 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 Δ F916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22.82%

[0262] 【计算方法为: $10.96\% \div (10.96\% + 22.89\% + 13.34\% + 0.83\%)$ 】。

[0263] 突出断裂末端D4的比例为47.67%

[0264] 【计算方法为: $22.89\% \div (10.96\% + 22.89\% + 13.34\% + 0.83\%)$ 】。

[0265] 突出断裂末端D5的比例为27.78%

[0266] 【计算方法为: $13.34\% \div (10.96\% + 22.89\% + 13.34\% + 0.83\%)$ 】。

[0267] 突出断裂末端D6的比例为1.73%

[0268] 【计算方法为: $0.83\% \div (10.96\% + 22.89\% + 13.34\% + 0.83\%)$ 】。

[0269] 亦即,Cas9核酸酶 Δ F916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22.82%,D4突出末端切割的占比为47.67%,D5突出末端切割的占比为27.78%,D6突出末端切割的占比为1.73%。

[0270] 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:

[0271]	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%
	Δ F916	22.82%	47.67%	27.78%	1.73%

[0272] 可见,在sgRNA2的介导下,相比于SpCas9核酸酶(Cas9WT), Δ F916突变体对基因组DNA片段中与sgRNA2非互补的DNA链进行切割时, Δ F916突变体在PAM上游4bp处进行切割的比例明显提高(D4)。

[0273] 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。

[0274] 此外,还将本发明的Cas9核酸酶突变体Cas9- Δ F916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。

[0275] 针对STM位点(β -globin RE1)的sgRNAs靶向序列:

[0276] β -globin RE1sgRNA1:GATTGTTGTTGCCTTGGAGTG (SEQ ID NO.19);

[0277] β -globin RE1sgRNA2:GCTGGTCCCCTGGTAACCTGG (SEQ ID NO.20);

[0278] 正反向脱氧寡核苷酸:

[0279] β -globin RE1sgRNA1F:accgATTGTTGTTGCCTTGGAGTG (SEQ ID NO.21);

[0280] β -globin RE1sgRNA1R:aaacCACTCCAAGGCAACAACAAT (SEQ ID NO.22);

[0281] β -globin RE1sgRNA2F:accgCTGGTCCCCTGGTAACCTGG (SEQ ID NO.23);

[0282] β -globin RE1sgRNA2R:aaacCCAGGTTACCAGGGGACCAG (SEQ ID NO.24);

[0283] 高通量引物:

[0284] HiSeq-hSTM-aF1:

[0285] AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTGCTTAGAGC
CAGGACTAATTGC; (SEQ ID NO.25)

[0286] HiSeq-hSTM-aR2:

[0287] AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTGGGTGTAGA
AATGAGCAAATAAGT (SEQ ID NO.26)

[0288] HiSeq-hSTM-2F:

[0289] CAAGCAGAAGACGGCATACGAGATGATCGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAGAT
TGAGTTCTGTTTGTTCATCTAC (SEQ ID NO.27)

[0290] HiSeq-hSTM-2R:

[0291] CAAGCAGAAGACGGCATACGAGATAGTCAAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAGC

TCTGCCTGAAAGGAGTC (SEQ ID NO.28)。

[0292] 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- Δ F916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。

[0293] 综上所述,本发明的Cas9核酸酶(Cas9- Δ F916)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- Δ F916)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,进而可实现特定位置的精准DNA片段编辑。

[0294] 本申请的参考文献如下:

[0295] 1.Stamatoyannopoulos,JA.(2012).What does our genome encode?Genome Res,22:1602-1611.

[0296] 2.The ENCODE Project Consortium.(2012).An integrated encyclopedia of DNA elements in the human genome.Nature,489:57-74.

[0297] 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.

[0298] 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.

[0299] 5.Neph,S,et al.(2012).An expansive human regulatory lexicon encoded in transcription factor footprints.Nature,489:83-90.

[0300] 6.Shen,Y,et al.(2012).A map of the cis-regulatory sequences in the mouse genome.Nature,488:116-120.

[0301] 7.Thurman,RE,et al.(2012).The accessible chromatin landscape of the human genome.Nature,489:75-82.

[0302] 8.de Laat,W and D Duboule.(2013).Topology of mammalian developmental enhancers and their regulatory landscapes.Nature,502:499-506.

[0303] 9.McClintock,B.(1950).The origin and behavior of mutable loci in maize.Proc Natl Acad Sci U S A,36:344-355.

[0304] 10.McClintock,B.(1984).The significance of responses of the genome to challenge.Science,226:792-801.

[0305] 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.

[0306] 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.

[0307] 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.

[0308] 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.

[0309] 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.

[0310] 16. Carroll, D. (2014). Genome engineering with targetable nucleases. *Annu Rev Biochem*, 83:409-439.

[0311] 17. Smithies, O, et al. (1985). Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. *Nature*, 317:230-234.

[0312] 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.

[0313] 19. Zheng, B, et al. (2000). Engineering mouse chromosomes with Cre-loxP: range, efficiency, and somatic applications. *Mol Cell Biol*, 20:648-655.

[0314] 20. Wu, S, et al. (2007). Toward simpler and faster genome-wide mutagenesis in mice. *Nat Genet*, 39:922-930.

[0315] 21. Gupta, A, et al. (2013). Targeted chromosomal deletions and inversions in zebrafish. *Genome Res*, 23:1008-1017.

[0316] 22. Xiao, A, et al. (2013). Chromosomal deletions and inversions mediated by TALENs and CRISPR/Cas in zebrafish. *Nucleic Acids Res*, 41:e141.

[0317] 23. Kraft, K, et al. (2015). Deletions, Inversions, Duplications: Engineering of Structural Variants using CRISPR/Cas in Mice. *Cell Rep*, 10:833-839.

[0318] 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.

[0319] 25. Jinek, M, et al. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337:816-821.

[0320] 26. Cong, L, et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science*, 339:819-823.

[0321] 27. Mali, P, et al. (2013). RNA-guided human genome engineering via Cas9. *Science*, 339:823-826.

[0322] 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.

[0323] 29. Sartori, AA, et al. (2007). Human CtIP promotes DNA end resection. *Nature*, 450:509-514.

[0324] 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.

[0325] 31.Li,J,J Shou,and Q Wu.(2015).DNA fragment editing of genomes by CRISPR/Cas9.Hereditas,37:992-1002.

[0326] 32.Huang,H and Q Wu.(2016).CRISPR Double Cutting through the Labyrinthine Architecture of 3D Genomes.J Genet Genomics,43:273-288.

[0327] 33.Guo,Y,et al.(2015).CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function.Cell,162:900-910.

[0328] 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

- [0001] SEQUENCE LISTING
- [0002] <110> 上海交通大学
- [0003] <120> 一种Cas9核酸酶 Δ F916及其用途
- [0004] <130> 171285
- [0005] <160> 28
- [0006] <170> PatentIn version 3.3
- [0007] <210> 1
- [0008] <211> 20
- [0009] <212> DNA
- [0010] <213> Artificial
- [0011] <220>
- [0012] <223> HS51 RE1sgRNA1
- [0013] <400> 1
- [0014] gccacacatc caaggctgac 20
- [0015] <210> 2
- [0016] <211> 21
- [0017] <212> DNA
- [0018] <213> Artificial
- [0019] <220>
- [0020] <223> HS51 RE1sgRNA2
- [0021] <400> 2
- [0022] gagatttggg gcgtcaggaa g 21
- [0023] <210> 3
- [0024] <211> 77
- [0025] <212> DNA
- [0026] <213> Artificial
- [0027] <220>
- [0028] <223> Hiseq-hHs51-aF
- [0029] <400> 3
- [0030] atgatacggc gaccaccgag atctacactc tttccctaca cgacgctctt ccgatctgca 60
- [0031] aggagatccg tgtcgtc 77
- [0032] <210> 4
- [0033] <211> 82
- [0034] <212> DNA
- [0035] <213> Artificial
- [0036] <220>
- [0037] <223> Hiseq-hs51-aRa
- [0038] <400> 4

[0039] aatgatacgg cgaccaccga gatctacact ctttccctac acgacgctct tccgatctaa 60
 [0040] ggatgttgtg gaaggcgagc ag 82
 [0041] <210> 5
 [0042] <211> 87
 [0043] <212> DNA
 [0044] <213> Artificial
 [0045] <220>
 [0046] <223> Hiseq-hs51-bFa
 [0047] <400> 5
 [0048] caagcagaag acggcatacg agatggacgg gtgactggag ttcagacgtg tgctcttccg 60
 [0049] atctcttttac atgacagctt ccggtag 87
 [0050] <210> 6
 [0051] <211> 89
 [0052] <212> DNA
 [0053] <213> Artificial
 [0054] <220>
 [0055] <223> Hiseq-hHs51-bR
 [0056] <400> 6
 [0057] caagcagaag acggcatacg agatttgact gtgactggag ttcagacgtg tgctcttccg 60
 [0058] atcttttttg gctaacaaca tagtgcttc 89
 [0059] <210> 7
 [0060] <211> 1401
 [0061] <212> PRT
 [0062] <213> Artificial
 [0063] <220>
 [0064] <223> SpCas9
 [0065] <400> 7
 [0066] Met Ala Pro Lys Lys Lys Arg Lys Val Gly Ile His Gly Val Pro Ala
 [0067] 1 5 10 15
 [0068] Ala Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser
 [0069] 20 25 30
 [0070] Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys
 [0071] 35 40 45
 [0072] Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu
 [0073] 50 55 60
 [0074] Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg
 [0075] 65 70 75 80
 [0076] Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile
 [0077] 85 90 95

[0078]	Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp
[0079]	100 105 110
[0080]	Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys
[0081]	115 120 125
[0082]	Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala
[0083]	130 135 140
[0084]	Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val
[0085]	145 150 155 160
[0086]	Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala
[0087]	165 170 175
[0088]	His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn
[0089]	180 185 190
[0090]	Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr
[0091]	195 200 205
[0092]	Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp
[0093]	210 215 220
[0094]	Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu
[0095]	225 230 235 240
[0096]	Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly
[0097]	245 250 255
[0098]	Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn
[0099]	260 265 270
[0100]	Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr
[0101]	275 280 285
[0102]	Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala
[0103]	290 295 300
[0104]	Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser
[0105]	305 310 315 320
[0106]	Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala
[0107]	325 330 335
[0108]	Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu
[0109]	340 345 350
[0110]	Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe
[0111]	355 360 365
[0112]	Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala
[0113]	370 375 380
[0114]	Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met
[0115]	385 390 395 400
[0116]	Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu

[0117]	405																410				415				
[0118]	Arg	Lys	Gln	Arg	Thr	Phe	Asp	Asn	Gly	Ser	Ile	Pro	His	Gln	Ile	His									
[0119]	420																425				430				
[0120]	Leu	Gly	Glu	Leu	His	Ala	Ile	Leu	Arg	Arg	Gln	Glu	Asp	Phe	Tyr	Pro									
[0121]	435																440				445				
[0122]	Phe	Leu	Lys	Asp	Asn	Arg	Glu	Lys	Ile	Glu	Lys	Ile	Leu	Thr	Phe	Arg									
[0123]	450																455				460				
[0124]	Ile	Pro	Tyr	Tyr	Val	Gly	Pro	Leu	Ala	Arg	Gly	Asn	Ser	Arg	Phe	Ala									
[0125]	465	470																475				480			
[0126]	Trp	Met	Thr	Arg	Lys	Ser	Glu	Glu	Thr	Ile	Thr	Pro	Trp	Asn	Phe	Glu									
[0127]	485																490				495				
[0128]	Glu	Val	Val	Asp	Lys	Gly	Ala	Ser	Ala	Gln	Ser	Phe	Ile	Glu	Arg	Met									
[0129]	500																505				510				
[0130]	Thr	Asn	Phe	Asp	Lys	Asn	Leu	Pro	Asn	Glu	Lys	Val	Leu	Pro	Lys	His									
[0131]	515																520				525				
[0132]	Ser	Leu	Leu	Tyr	Glu	Tyr	Phe	Thr	Val	Tyr	Asn	Glu	Leu	Thr	Lys	Val									
[0133]	530																535				540				
[0134]	Lys	Tyr	Val	Thr	Glu	Gly	Met	Arg	Lys	Pro	Ala	Phe	Leu	Ser	Gly	Glu									
[0135]	545	550																555				560			
[0136]	Gln	Lys	Lys	Ala	Ile	Val	Asp	Leu	Leu	Phe	Lys	Thr	Asn	Arg	Lys	Val									
[0137]	565																570				575				
[0138]	Thr	Val	Lys	Gln	Leu	Lys	Glu	Asp	Tyr	Phe	Lys	Lys	Ile	Glu	Cys	Phe									
[0139]	580																585				590				
[0140]	Asp	Ser	Val	Glu	Ile	Ser	Gly	Val	Glu	Asp	Arg	Phe	Asn	Ala	Ser	Leu									
[0141]	595																600				605				
[0142]	Gly	Thr	Tyr	His	Asp	Leu	Leu	Lys	Ile	Ile	Lys	Asp	Lys	Asp	Phe	Leu									
[0143]	610																615				620				
[0144]	Asp	Asn	Glu	Glu	Asn	Glu	Asp	Ile	Leu	Glu	Asp	Ile	Val	Leu	Thr	Leu									
[0145]	625	630																635				640			
[0146]	Thr	Leu	Phe	Glu	Asp	Arg	Glu	Met	Ile	Glu	Glu	Arg	Leu	Lys	Thr	Tyr									
[0147]	645																650				655				
[0148]	Ala	His	Leu	Phe	Asp	Asp	Lys	Val	Met	Lys	Gln	Leu	Lys	Arg	Arg	Arg									
[0149]	660																665				670				
[0150]	Tyr	Thr	Gly	Trp	Gly	Arg	Leu	Ser	Arg	Lys	Leu	Ile	Asn	Gly	Ile	Arg									
[0151]	675																680				685				
[0152]	Asp	Lys	Gln	Ser	Gly	Lys	Thr	Ile	Leu	Asp	Phe	Leu	Lys	Ser	Asp	Gly									
[0153]	690																695				700				
[0154]	Phe	Ala	Asn	Arg	Asn	Phe	Met	Gln	Leu	Ile	His	Asp	Asp	Ser	Leu	Thr									
[0155]	705	710																715				720			

[0156]	Phe	Lys	Glu	Asp	Ile	Gln	Lys	Ala	Gln	Val	Ser	Gly	Gln	Gly	Asp	Ser
[0157]						725				730					735	
[0158]	Leu	His	Glu	His	Ile	Ala	Asn	Leu	Ala	Gly	Ser	Pro	Ala	Ile	Lys	Lys
[0159]						740				745					750	
[0160]	Gly	Ile	Leu	Gln	Thr	Val	Lys	Val	Val	Asp	Glu	Leu	Val	Lys	Val	Met
[0161]						755				760					765	
[0162]	Gly	Arg	His	Lys	Pro	Glu	Asn	Ile	Val	Ile	Glu	Met	Ala	Arg	Glu	Asn
[0163]						770				775					780	
[0164]	Gln	Thr	Thr	Gln	Lys	Gly	Gln	Lys	Asn	Ser	Arg	Glu	Arg	Met	Lys	Arg
[0165]	785					790				795					800	
[0166]	Ile	Glu	Glu	Gly	Ile	Lys	Glu	Leu	Gly	Ser	Gln	Ile	Leu	Lys	Glu	His
[0167]						805				810					815	
[0168]	Pro	Val	Glu	Asn	Thr	Gln	Leu	Gln	Asn	Glu	Lys	Leu	Tyr	Leu	Tyr	Tyr
[0169]						820				825					830	
[0170]	Leu	Gln	Asn	Gly	Arg	Asp	Met	Tyr	Val	Asp	Gln	Glu	Leu	Asp	Ile	Asn
[0171]						835				840					845	
[0172]	Arg	Leu	Ser	Asp	Tyr	Asp	Val	Asp	His	Ile	Val	Pro	Gln	Ser	Phe	Leu
[0173]						850				855					860	
[0174]	Lys	Asp	Asp	Ser	Ile	Asp	Asn	Lys	Val	Leu	Thr	Arg	Ser	Asp	Lys	Asn
[0175]	865					870				875					880	
[0176]	Arg	Gly	Lys	Ser	Asp	Asn	Val	Pro	Ser	Glu	Glu	Val	Val	Lys	Lys	Met
[0177]						885				890					895	
[0178]	Lys	Asn	Tyr	Trp	Arg	Gln	Leu	Leu	Asn	Ala	Lys	Leu	Ile	Thr	Gln	Arg
[0179]						900				905					910	
[0180]	Lys	Phe	Asp	Asn	Leu	Thr	Lys	Ala	Glu	Arg	Gly	Gly	Leu	Ser	Glu	Leu
[0181]						915				920					925	
[0182]	Asp	Lys	Ala	Gly	Phe	Ile	Lys	Arg	Gln	Leu	Val	Glu	Thr	Arg	Gln	Ile
[0183]						930				935					940	
[0184]	Thr	Lys	His	Val	Ala	Gln	Ile	Leu	Asp	Ser	Arg	Met	Asn	Thr	Lys	Tyr
[0185]	945					950				955					960	
[0186]	Asp	Glu	Asn	Asp	Lys	Leu	Ile	Arg	Glu	Val	Lys	Val	Ile	Thr	Leu	Lys
[0187]						965				970					975	
[0188]	Ser	Lys	Leu	Val	Ser	Asp	Phe	Arg	Lys	Asp	Phe	Gln	Phe	Tyr	Lys	Val
[0189]						980				985					990	
[0190]	Arg	Glu	Ile	Asn	Asn	Tyr	His	His	Ala	His	Asp	Ala	Tyr	Leu	Asn	Ala
[0191]						995				1000					1005	
[0192]	Val	Val	Gly	Thr	Ala	Leu	Ile	Lys	Lys	Tyr	Pro	Lys	Leu	Glu	Ser	
[0193]						1010				1015					1020	
[0194]	Glu	Phe	Val	Tyr	Gly	Asp	Tyr	Lys	Val	Tyr	Asp	Val	Arg	Lys	Met	

[0195]	1025	1030	1035
[0196]	Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr		
[0197]	1040	1045	1050
[0198]	Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr		
[0199]	1055	1060	1065
[0200]	Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn		
[0201]	1070	1075	1080
[0202]	Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala		
[0203]	1085	1090	1095
[0204]	Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys		
[0205]	1100	1105	1110
[0206]	Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu		
[0207]	1115	1120	1125
[0208]	Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp		
[0209]	1130	1135	1140
[0210]	Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr		
[0211]	1145	1150	1155
[0212]	Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys		
[0213]	1160	1165	1170
[0214]	Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg		
[0215]	1175	1180	1185
[0216]	Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly		
[0217]	1190	1195	1200
[0218]	Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr		
[0219]	1205	1210	1215
[0220]	Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser		
[0221]	1220	1225	1230
[0222]	Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys		
[0223]	1235	1240	1245
[0224]	Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys		
[0225]	1250	1255	1260
[0226]	Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln		
[0227]	1265	1270	1275
[0228]	His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe		
[0229]	1280	1285	1290
[0230]	Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu		
[0231]	1295	1300	1305
[0232]	Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala		
[0233]	1310	1315	1320

[0234]	Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro	
[0235]	1325	1330 1335
[0236]	Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr	
[0237]	1340	1345 1350
[0238]	Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser	
[0239]	1355	1360 1365
[0240]	Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly	
[0241]	1370	1375 1380
[0242]	Gly Asp Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys	
[0243]	1385	1390 1395
[0244]	Lys Lys Lys	
[0245]	1400	
[0246]	<210> 8	
[0247]	<211> 4206	
[0248]	<212> DNA	
[0249]	<213> Artificial	
[0250]	<220>	
[0251]	<223> SpCas9	
[0252]	<400> 8	
[0253]	atggcccca	agaagaagcg gaaggtcggg atccacgggt tcccagcagc catggacaag 60
[0254]	aagtactcca	ttgggctcga tatcggcaca aacagcgtcg gctgggccgt cattacggac 120
[0255]	gagtacaagg	tgccgagcaa aaaattcaaa gttctgggca ataccgatcg ccacagcata 180
[0256]	aagaagaacc	tcattggcgc cctcctgttc gactccgggg agacggccga agccacgcgg 240
[0257]	ctcaaagaa	cagcacggcg cagatatacc cgcagaaaga atcggatctg ctacctgcag 300
[0258]	gagatcttta	gtaatgagat ggctaagggtg gatgactctt tcttccatag gctggaggag 360
[0259]	tccttttttg	tggaggagga taaaagcac gagcgccacc caatcttttg caatatcgtg 420
[0260]	gacgaggtgg	cgtaccatga aaagtaccca accatatatc atctgaggaa gaagcttgta 480
[0261]	gacagtactg	ataaggtga cttgcggttg atctatctcg cgctggcgca tatgatcaaa 540
[0262]	tttcggggac	acttcctcat cgagggggac ctgaaccag acaacagcga tgtcgacaaa 600
[0263]	ctctttatcc	aactggttca gacttacaat cagcttttcg aagagaaccg gatcaacgca 660
[0264]	tccggagttg	acgccaagc aatcctgagc gctaggctgt ccaaatcccg gcggtctgaa 720
[0265]	aacctcatcg	cacagctccc tggggagaag aagaacggcc tgtttggtaa tcttatcgcc 780
[0266]	ctgtcactcg	ggctgacccc caactttaaa tctaacttcg acctggccga agatgccaag 840
[0267]	cttcaactga	gcaaagacac ctacgatgat gatctcgaca atctgctggc ccagatcggc 900
[0268]	gaccagtacg	cagacctttt tttggcggca aagaacctgt cagacgcat tctgctgagt 960
[0269]	gatattctgc	gagtgaacac ggagatcacc aaagctccgc tgagcgtag tatgatcaag 1020
[0270]	cgctatgatg	agcaccacca agacttgact ttgctgaagg cccttgctcag acagcaactg 1080
[0271]	cctgagaagt	acaaggaaat tttcttcgat cagtctaaaa atggctacgc cggatacatt 1140
[0272]	gacggcggag	caagccagga ggaattttac aaatttatta agcccatctt ggaaaaaatg 1200

[0273]	gacggcaccg	aggagctgct	ggtaaagctt	aacagagaag	atctgttgcg	caaacagcgc	1260
[0274]	actttcgaca	atggaagcat	ccccaccag	attcacctgg	gcgaactgca	cgctatactc	1320
[0275]	aggcggcaag	aggattttcta	cccctttttg	aaagataaca	gggaaaagat	tgagaaaatc	1380
[0276]	ctcacatttc	ggatacccta	ctatgttaggc	cccctcgccc	ggggaaattc	cagattcgcg	1440
[0277]	tggatgactc	gcaaatcaga	agagaccatc	actccctgga	acttcgagga	agtcgtggat	1500
[0278]	aagggggcct	ctgcccagtc	cttcacgaa	aggatgacta	actttgataa	aaatctgcct	1560
[0279]	aacgaaaagg	tgcttcctaa	acactctctg	ctgtacgagt	acttcacagt	ttataacgag	1620
[0280]	ctcaccaagg	tcaaatacgt	cacagaagg	atgagaaagc	cagcattcct	gtctggagag	1680
[0281]	cagaagaaag	ctatcgtgga	cctcctcttc	aagacgaacc	ggaaaattac	cgtgaaacag	1740
[0282]	ctcaaagaag	actattttcaa	aaagattgaa	tgtttcgact	ctgttgaaat	cagcggagtg	1800
[0283]	gaggatcgct	tcaacgcata	cctgggaacg	tatcacgata	tcctgaaaat	cattaaagac	1860
[0284]	aaggacttcc	tggacaatga	ggagaacgag	gacattcttg	aggacattgt	cctcacctt	1920
[0285]	acgttgtttg	aagataggga	gatgattgaa	gaacgcttga	aaacttacgc	tcattctctc	1980
[0286]	gacgacaaag	tcatgaaaca	gctcaagagg	cgccgatata	caggatgggg	gcggctgtca	2040
[0287]	agaaaactga	tcaatgggat	ccgagacaag	cagagtggaa	agacaatcct	ggattttctt	2100
[0288]	aagtccgatg	gatttgccaa	ccggaacttc	atgcagttga	tccatgatga	ctctctcacc	2160
[0289]	tttaaggagg	acatccagaa	agcacaagtt	tctggccagg	gggacagtct	tcacgagcac	2220
[0290]	atcgctaata	ttgcaggtag	cccagctatc	aaaaaggga	tactgcagac	cgttaaggtc	2280
[0291]	gtggatgaac	tcgtcaaagt	aatgggaagg	cataagcccc	agaatatcgt	tatcgagatg	2340
[0292]	gcccagagag	accaaactac	ccagaaggga	cagaagaaca	gtagggaaag	gatgaagagg	2400
[0293]	attgaagagg	gtataaaaga	actgggggtc	caaatcctta	aggaacaccc	agttgaaaac	2460
[0294]	accagcttc	agaatgagaa	gctctacctg	tactacctgc	agaacggcag	ggacatgtac	2520
[0295]	gtggatcagg	aactggacat	caatcggtc	tccgactacg	acgtggatca	tatcgtgccc	2580
[0296]	cagtcttttc	tcaaagatga	ttctattgat	aataaagtgt	tgacaagatc	cgataaaaat	2640
[0297]	agagggaaga	gtgataacgt	cccctcagaa	gaagttgtca	agaaaatgaa	aaattatttg	2700
[0298]	cggcagctgc	tgaacgccaa	actgatcaca	caacggaagt	tcgataatct	gactaaggct	2760
[0299]	gaacgaggtg	gcctgtctga	gttgataaaa	gcaggcttca	tcaaaaggca	gcttggttag	2820
[0300]	acacgccaga	tcaccaagca	cgtggcccaa	attctcgatt	cacgcatgaa	caccaagtac	2880
[0301]	gatgaaaatg	acaaactgat	tcgagaggtg	aaagttatta	ctctgaagtc	taagctggtc	2940
[0302]	tcagatttca	gaaaggactt	tcagttttat	aaggtgagag	agatcaacaa	ttaccaccat	3000
[0303]	gcgcatgatg	cctacctgaa	tgcagtggta	ggcactgcac	ttatcaaaaa	atatcccaag	3060
[0304]	cttgaatctg	aatttgttta	cggagactat	aaagtgtacg	atgttaggaa	aatgatcgca	3120
[0305]	aagtctgagc	aggaaatagg	caaggccacc	gctaagtact	tcttttacag	caatattatg	3180
[0306]	aattttttca	agaccgagat	tacactggcc	aatggagaga	ttcggaagcg	accacttatc	3240
[0307]	gaaacaaacg	gagaaacagg	agaaatcgtg	tgggacaagg	gtagggattt	cgcgacagtc	3300
[0308]	cggaaggtcc	tgtccatgcc	gcaggtgaac	atcgttaaaa	agaccgaagt	acagaccgga	3360
[0309]	ggcttctcca	aggaaagtat	cctcccga	aggaacagcg	acaagctgat	cgcacgcaaa	3420
[0310]	aaagattggg	acccaagaa	atacggcgga	ttcgattctc	ctacagtcgc	ttacagtgtg	3480
[0311]	ctggttgtgg	ccaaagtgga	gaaagggaag	tctaaaaaac	tcaaaagcgt	caaggaactg	3540

```

[0312] ctgggcatca caatcatgga gcgatcaagc ttcgaaaaaa accccatcga ctttctcgag 3600
[0313] gcgaaaggat ataaagaggt caaaaaagac ctcatcatta agcttcccaa gtactctctc 3660
[0314] tttgagcttg aaaacggccg gaaacgaatg ctcgctagtg cgggcgagct gcagaaaggt 3720
[0315] aacgagctgg cactgccctc taaatacggtt aatttcttgt atctggccag ccactatgaa 3780
[0316] aagctcaaag ggtctcccga agataatgag cagaagcagc tgttcgtgga acaacacaaa 3840
[0317] cactaccttg atgagatcat cgagcaaata agcgaattct ccaaaagagt gatcctcgcc 3900
[0318] gacgctaacc tcgataaggt gctttctgct tacaataagc acagggataa gcccatcagg 3960
[0319] gacgaggcag aaaacattat ccacttgttt actctgacca acttgggcgc gcctgcagcc 4020
[0320] ttcaagtact tcgacaccac catagacaga aagcgggtaca cctctacaaa ggaggtcctg 4080
[0321] gacgccacac tgattcatca gtcaattacg gggctctatg aaacaagaat cgacctctct 4140
[0322] cagctcgggtg gagacaagcg tcctgctgct actaagaaag ctggtcaagc taagaaaaag 4200
[0323] aaataa 4206
[0324] <210> 9
[0325] <211> 1400
[0326] <212> PRT
[0327] <213> Artificial
[0328] <220>
[0329] <223> Cas9- ΔF916
[0330] <400> 9
[0331] Met Ala Pro Lys Lys Lys Arg Lys Val Gly Ile His Gly Val Pro Ala
[0332] 1 5 10 15
[0333] Ala Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser
[0334] 20 25 30
[0335] Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys
[0336] 35 40 45
[0337] Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu
[0338] 50 55 60
[0339] Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg
[0340] 65 70 75 80
[0341] Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile
[0342] 85 90 95
[0343] Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp
[0344] 100 105 110
[0345] Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys
[0346] 115 120 125
[0347] Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala
[0348] 130 135 140
[0349] Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val
[0350] 145 150 155 160

```

[0351]	Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala		
[0352]		165	170
[0353]	His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn		175
[0354]		180	185
[0355]	Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr		190
[0356]		195	200
[0357]	Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp		205
[0358]		210	215
[0359]	Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu		220
[0360]		225	230
[0361]	Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly		235
[0362]		245	250
[0363]	Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn		240
[0364]		260	265
[0365]	Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr		255
[0366]		275	280
[0367]	Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala		285
[0368]		290	295
[0369]	Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser		300
[0370]		305	310
[0371]	Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala		315
[0372]		325	330
[0373]	Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu		335
[0374]		340	345
[0375]	Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe		350
[0376]		355	360
[0377]	Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala		365
[0378]		370	375
[0379]	Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met		380
[0380]		385	390
[0381]	Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu		395
[0382]		405	410
[0383]	Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His		415
[0384]		420	425
[0385]	Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro		430
[0386]		435	440
[0387]	Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg		445
[0388]		450	455
[0389]	Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala		460

[0390]	465	470	475	480
[0391]	Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu			
[0392]		485	490	495
[0393]	Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met			
[0394]		500	505	510
[0395]	Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His			
[0396]		515	520	525
[0397]	Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val			
[0398]		530	535	540
[0399]	Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu			
[0400]	545	550	555	560
[0401]	Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val			
[0402]		565	570	575
[0403]	Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe			
[0404]		580	585	590
[0405]	Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu			
[0406]		595	600	605
[0407]	Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu			
[0408]		610	615	620
[0409]	Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu			
[0410]	625	630	635	640
[0411]	Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr			
[0412]		645	650	655
[0413]	Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg			
[0414]		660	665	670
[0415]	Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg			
[0416]		675	680	685
[0417]	Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly			
[0418]		690	695	700
[0419]	Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr			
[0420]	705	710	715	720
[0421]	Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser			
[0422]		725	730	735
[0423]	Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys			
[0424]		740	745	750
[0425]	Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met			
[0426]		755	760	765
[0427]	Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn			
[0428]		770	775	780

[0429]	Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg
[0430]	785 790 795 800
[0431]	Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His
[0432]	805 810 815
[0433]	Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr
[0434]	820 825 830
[0435]	Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn
[0436]	835 840 845
[0437]	Arg Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu
[0438]	850 855 860
[0439]	Lys Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn
[0440]	865 870 875 880
[0441]	Arg Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met
[0442]	885 890 895
[0443]	Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg
[0444]	900 905 910
[0445]	Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu
[0446]	915 920 925
[0447]	Asp Lys Ala Gly Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
[0448]	930 935 940
[0449]	Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
[0450]	945 950 955 960
[0451]	Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
[0452]	965 970 975
[0453]	Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
[0454]	980 985 990
[0455]	Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
[0456]	995 1000 1005
[0457]	Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu
[0458]	1010 1015 1020
[0459]	Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile
[0460]	1025 1030 1035
[0461]	Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe
[0462]	1040 1045 1050
[0463]	Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu
[0464]	1055 1060 1065
[0465]	Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly
[0466]	1070 1075 1080
[0467]	Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr

[0468]	1085	1090	1095
[0469]	Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys		
[0470]	1100	1105	1110
[0471]	Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro		
[0472]	1115	1120	1125
[0473]	Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp		
[0474]	1130	1135	1140
[0475]	Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser		
[0476]	1145	1150	1155
[0477]	Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu		
[0478]	1160	1165	1170
[0479]	Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser		
[0480]	1175	1180	1185
[0481]	Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr		
[0482]	1190	1195	1200
[0483]	Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser		
[0484]	1205	1210	1215
[0485]	Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala		
[0486]	1220	1225	1230
[0487]	Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr		
[0488]	1235	1240	1245
[0489]	Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly		
[0490]	1250	1255	1260
[0491]	Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His		
[0492]	1265	1270	1275
[0493]	Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser		
[0494]	1280	1285	1290
[0495]	Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser		
[0496]	1295	1300	1305
[0497]	Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu		
[0498]	1310	1315	1320
[0499]	Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala		
[0500]	1325	1330	1335
[0501]	Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr		
[0502]	1340	1345	1350
[0503]	Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile		
[0504]	1355	1360	1365
[0505]	Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly		
[0506]	1370	1375	1380

[0507]	Asp Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys	
[0508]	1385	1390 1395
[0509]	Lys Lys	
[0510]	1400	
[0511]	<210> 10	
[0512]	<211> 4203	
[0513]	<212> DNA	
[0514]	<213> Artificial	
[0515]	<220>	
[0516]	<223> Cas9- ΔF916	
[0517]	<400> 10	
[0518]	atggcccca	agaagaagcg gaaggtcggt atccacggtg tcccagcagc catggacaag 60
[0519]	aagtactcca	ttgggctcga tateggcaca aacagcgctg gctgggccgt cattacggac 120
[0520]	gagtacaagg	tgccgagcaa aaaattcaaa gttctgggca ataccgatcg ccacagcata 180
[0521]	aagaagaacc	tcattggcgc cctcctgttc gactccgggg agacggccga agccacgcgg 240
[0522]	ctcaaaagaa	cagcacggcg cagatatacc cgcagaaaga atcggatctg ctacctgcag 300
[0523]	gagatcttta	gtaatgagat ggctaaggtg gatgactctt tcttccatag gctggaggag 360
[0524]	tccttttttg	tggaggagga taaaagcac gagcgccacc caatcttttg caatatcgtg 420
[0525]	gacgaggttg	cgtaccatga aaagtaccca accatatatc atctgaggaa gaagcttgta 480
[0526]	gacagtactg	ataaggtgta cttgcggttg atctatctcg cgctggcgca tatgatcaaa 540
[0527]	tttcggggac	acttcctcat cgagggggac ctgaaccag acaacagcga tgtcgacaaa 600
[0528]	ctctttatcc	aactggttca gacttacaat cagcttttcg aagagaaccc gatcaacgca 660
[0529]	tccggagttg	acgccaaagc aatcctgagc gctaggctgt ccaaattccc gcggtctgaa 720
[0530]	aacctcatcg	cacagctccc tggggagaag aagaacggcc tgtttggtaa tcttatcgcc 780
[0531]	ctgtcactcg	ggctgacccc caactttaaa tctaacttcg acctggccga agatgccaag 840
[0532]	cttcaactga	gcaaagacac ctacgatgat gatctcgaca atctgctggc ccagatcggc 900
[0533]	gaccagtacg	cagacctttt tttggcggca aagaacctgt cagacgcat tctgctgagt 960
[0534]	gatattctgc	gagtgaacac ggagatcacc aaagctccgc tgagcgctag tatgatcaag 1020
[0535]	cgctatgatg	agcaccacca agacttgact ttgctgaagg cccttgctcag acagcaactg 1080
[0536]	cctgagaagt	acaaggaaat tttcttcgat cagtctaaaa atggctacgc cggatacatt 1140
[0537]	gacggcggag	caagccagga ggaattttac aaatttatta agcccatctt ggaaaaaatg 1200
[0538]	gacggcaccg	aggagctgct ggtaaagctt aacagagaag atctgttgcg caaacagcgc 1260
[0539]	actttcgaca	atggaagcat cccccaccag attcacctgg gcgaactgca cgctatactc 1320
[0540]	aggcggcaag	aggatttcta cccctttttg aaagataaca gggaaaagat tgagaaaatc 1380
[0541]	ctcacatttc	ggatacccta ctatgtaggc cccctcgccc ggggaaattc cagattcgcg 1440
[0542]	tggatgactc	gcaaatcaga agagaccatc actccctgga acttcgagga agtcgtggat 1500
[0543]	aagggggcct	ctgcccagtc cttcatcgaa aggatgacta actttgataa aaatctgcct 1560
[0544]	aacgaaaagg	tgcttcctaa acactctctg ctgtacgagt acttcacagt ttataacgag 1620
[0545]	ctcaccaagg	tcaaatacgt cacagaaggg atgagaaagc cagcattcct gtctggagag 1680

[0546]	cagaagaaag	ctatcgtgga	cctcctcttc	aagacgaacc	ggaaagttag	cgtgaaacag	1740
[0547]	ctcaaagaag	actattttcaa	aaagattgaa	tgtttcgact	ctgttgaaat	cagcggagtg	1800
[0548]	gaggatcgct	tcaacgcac	cctgggaacg	tatcacgac	tcctgaaaat	cattaaagac	1860
[0549]	aaggacttcc	tggacaatga	ggagaacgag	gacattcttg	aggacattgt	cctcaccctt	1920
[0550]	acgttggttg	aagataggga	gatgattgaa	gaacgcttga	aaacttacgc	tcctctcttc	1980
[0551]	gacgacaaag	tcatgaaaca	gctcaagagg	cgccgatata	caggatgggg	gaggctgtca	2040
[0552]	agaaaactga	tcaatgggat	ccgagacaag	cagagtggaa	agacaatcct	ggattttctt	2100
[0553]	aagtccgatg	gatttgccaa	ccggaacttc	atgcagttga	tccatgatga	ctctctcacc	2160
[0554]	tttaaggagg	acatccagaa	agcacaagtt	tctggccagg	gggacagtct	tcacgagcac	2220
[0555]	atcgctaate	ttgcaggtag	cccagctatc	aaaaagggaa	tactgcagac	cgttaagggtc	2280
[0556]	gtggatgaac	tcgtcaaagt	aatgggaagg	cataagcccc	agaatatcgt	tatcgagatg	2340
[0557]	gcccagagaga	accaaactac	ccagaaggga	cagaagaaca	gtagggaag	gatgaagagg	2400
[0558]	attgaagagg	gtataaaaga	actgggggtc	caaatcctta	aggaacaccc	agttgaaaac	2460
[0559]	accagcttc	agaatgagaa	gctctacctg	tactacctgc	agaacggcag	ggacatgtac	2520
[0560]	gtggatcagg	aactggacat	caatcggtc	tcgactacg	acgtggatca	tatcgtgccc	2580
[0561]	cagtcttttc	tcaaagatga	ttctattgat	aataaagtgt	tgacaagatc	cgataaaaat	2640
[0562]	agagggaaga	gtgataacgt	cccctcagaa	gaagttgtca	agaaaatgaa	aaattatttg	2700
[0563]	cggcagctgc	tgaacgcaa	actgatcaca	caacggaagt	tcgataatct	gactaaggct	2760
[0564]	gaacgaggtg	gctgtctga	gttgataaaa	gcaggcatca	aaaggcagct	tgttgagaca	2820
[0565]	cgccagatca	ccaagcacgt	ggcccaaatt	ctcgattcac	gcatgaacac	caagtacgat	2880
[0566]	gaaaatgaca	aactgattcg	agaggtgaaa	gttattactc	tgaagtctaa	gctggtctca	2940
[0567]	gatttcagaa	aggactttca	gttttataag	gtgagagaga	tcaacaatta	ccaccatgcg	3000
[0568]	catgatgcct	acctgaatgc	agtggtaggc	actgcactta	tcaaaaaata	tcccaagctt	3060
[0569]	gaatctgaat	ttgtttacgg	agactataaa	gtgtacgatg	ttaggaaaat	gatcgcaaag	3120
[0570]	tctgagcagg	aaataggcaa	ggccaccgct	aagtacttct	tttacagcaa	tattatgaat	3180
[0571]	tttttcaaga	ccgagattac	actggccaat	ggagagattc	ggaagcgacc	acttatcgaa	3240
[0572]	acaaacggag	aaacaggaga	aatcgtgttg	gacaagggtg	gggatttcgc	gacagtccgg	3300
[0573]	aaggtcctgt	ccatgccgca	ggtgaacatc	gttaaaaaga	ccgaagtaca	gaccggaggc	3360
[0574]	ttctccaagg	aaagtatcct	cccgaagg	aacagcgaca	agctgatcgc	acgcaaaaaa	3420
[0575]	gattgggacc	ccaagaaata	cggcggattc	gatttctcta	cagtcgctta	cagtgtactg	3480
[0576]	gttgtggcca	aagtggagaa	agggaagtct	aaaaaactca	aaagcgtcaa	ggaactgctg	3540
[0577]	ggcatcaca	tcatggagcg	atcaagcttc	gaaaaaaacc	ccatcgactt	tctcgaggcg	3600
[0578]	aaaggatata	aagaggtcaa	aaaagacctc	atcattaagc	ttcccaagta	ctctctcttt	3660
[0579]	gagcttgaaa	acggccggaa	acgaatgctc	gctagtgcgg	gcgagctgca	gaaaggtaac	3720
[0580]	gagctggcac	tgccctctaa	atacgttaat	ttcttgatc	tggccagcca	ctatgaaaag	3780
[0581]	ctcaaagggt	ctcccgaaga	taatgagcag	aagcagctgt	tcgtggaaca	acacaaacac	3840
[0582]	taccttgatg	agatcatcga	gcaaataagc	gaattctcca	aaagagtgat	cctcgccgac	3900
[0583]	gctaacctcg	ataaggtgct	ttctgcttac	aataagcaca	gggataagcc	catcaggagg	3960
[0584]	caggcagaaa	acattatcca	cttgtttact	ctgaccaact	tgggcgcgcc	tgcagccttc	4020

[0585]	aagtacttcg acaccacccat agacagaaag cggtagacct ctacaaagga ggtcctggac	4080
[0586]	gccacactga ttcacagtc aattacgggg ctctatgaaa caagaatcga cctctctcag	4140
[0587]	ctcgggtggag acaagcgtcc tgctgctact aagaaagctg gtcaagctaa gaaaaagaaa	4200
[0588]	taa	4203
[0589]	<210> 11	
[0590]	<211> 20	
[0591]	<212> DNA	
[0592]	<213> Artificial	
[0593]	<220>	
[0594]	<223> β -globin RE2sgRNA1	
[0595]	<400> 11	
[0596]	acccaatgac ctcaggctgt	20
[0597]	<210> 12	
[0598]	<211> 20	
[0599]	<212> DNA	
[0600]	<213> Artificial	
[0601]	<220>	
[0602]	<223> β -globin RE2sgRNA2	
[0603]	<400> 12	
[0604]	tcacttggtta gcggcatctg	20
[0605]	<210> 13	
[0606]	<211> 19	
[0607]	<212> DNA	
[0608]	<213> Artificial	
[0609]	<220>	
[0610]	<223> Cas9- Δ F916-F	
[0611]	<400> 13	
[0612]	atcaaaaggc agcttggtg	19
[0613]	<210> 14	
[0614]	<211> 18	
[0615]	<212> DNA	
[0616]	<213> Artificial	
[0617]	<220>	
[0618]	<223> Cas9- Δ F916-R	
[0619]	<400> 14	
[0620]	gcctgcttta tccaactc	18
[0621]	<210> 15	
[0622]	<211> 86	
[0623]	<212> DNA	

[0624] <213> Artificial
[0625] <220>
[0626] <223> Hiseq-RRM-1F3
[0627] <400> 15
[0628] aatgatacgg cgaccaccga gatctacact ctttcacctac acgacgctct tccgatctat 60
[0629] atggcatcct agccttaaga aactag 86
[0630] <210> 16
[0631] <211> 81
[0632] <212> DNA
[0633] <213> Artificial
[0634] <220>
[0635] <223> Hiseq-RRM-1R2
[0636] <400> 16
[0637] aatgatacgg cgaccaccga gatctacact ctttcacctac acgacgctct tccgatctta 60
[0638] cgacgcagga gccgtatcat g 81
[0639] <210> 17
[0640] <211> 89
[0641] <212> DNA
[0642] <213> Artificial
[0643] <220>
[0644] <223> Hiseq-RRM-3F2
[0645] <400> 17
[0646] caagcagaag acggcatacg agataagcta gtgactggag ttcagacgtg tgctcttccg 60
[0647] atctatagca atgaaatctt gaaggagtg 89
[0648] <210> 18
[0649] <211> 85
[0650] <212> DNA
[0651] <213> Artificial
[0652] <220>
[0653] <223> Hiseq-RRM-3R2
[0654] <400> 18
[0655] caagcagaag acggcatacg agattcaagt gtgactggag ttcagacgtg tgctcttccg 60
[0656] atctgcacag ccttgcctta ttacg 85
[0657] <210> 19
[0658] <211> 21
[0659] <212> DNA
[0660] <213> Artificial
[0661] <220>
[0662] <223> β -globin RE1sgRNA1

[0663] <400> 19
[0664] gattgttggt gccttgagat g 21
[0665] <210> 20
[0666] <211> 21
[0667] <212> DNA
[0668] <213> Artificial
[0669] <220>
[0670] <223> β -globin RE1sgRNA2
[0671] <400> 20
[0672] gctggtcccc tggtaacctg g 21
[0673] <210> 21
[0674] <211> 24
[0675] <212> DNA
[0676] <213> Artificial
[0677] <220>
[0678] <223> β -globin RE1sgRNA1F
[0679] <400> 21
[0680] accgattggt gttgccttgg agtg 24
[0681] <210> 22
[0682] <211> 24
[0683] <212> DNA
[0684] <213> Artificial
[0685] <220>
[0686] <223> β -globin relsgrnalr
[0687] <400> 22
[0688] aaaccactcc aaggcaaca caat 24
[0689] <210> 23
[0690] <211> 24
[0691] <212> DNA
[0692] <213> Artificial
[0693] <220>
[0694] <223> β -globin RE1sgRNA2F
[0695] <400> 23
[0696] accgctggte ccctggtaac ctgg 24
[0697] <210> 24
[0698] <211> 24
[0699] <212> DNA
[0700] <213> Artificial
[0701] <220>

[0702] <223> β -globin RE1sgRNA2R
[0703] <400> 24
[0704] aaaccaggt taccagggga ccag 24
[0705] <210> 25
[0706] <211> 81
[0707] <212> DNA
[0708] <213> Artificial
[0709] <220>
[0710] <223> Hiseq-hSTM-aF1
[0711] <400> 25
[0712] aatgatacgg cgaccaccga gatctacact ctttccctac acgacgtctt tccgatcttg 60
[0713] cttagagcca ggactaattg c 81
[0714] <210> 26
[0715] <211> 83
[0716] <212> DNA
[0717] <213> Artificial
[0718] <220>
[0719] <223> Hiseq-hSTM-aR2
[0720] <400> 26
[0721] aatgatacgg cgaccaccga gatctacact ctttccctac acgacgtctt tccgatcttg 60
[0722] ggtgtagaaa tgagcaaata agt 83
[0723] <210> 27
[0724] <211> 91
[0725] <212> DNA
[0726] <213> Artificial
[0727] <220>
[0728] <223> Hiseq-hSTM-2F
[0729] <400> 27
[0730] caagcagaag acggcatacg agatgatcgt gtgactggag ttcagacgtg tgctcttccg 60
[0731] atctagattg agttctgttt gtttcatcta c 91
[0732] <210> 28
[0733] <211> 85
[0734] <212> DNA
[0735] <213> Artificial
[0736] <220>
[0737] <223> Hiseq-hSTM-2R
[0738] <400> 28
[0739] caagcagaag acggcatacg agatagtcaa gtgactggag ttcagacgtg tgctcttccg 60
[0740] atctcagctc tgcctgaaag gagtc 85

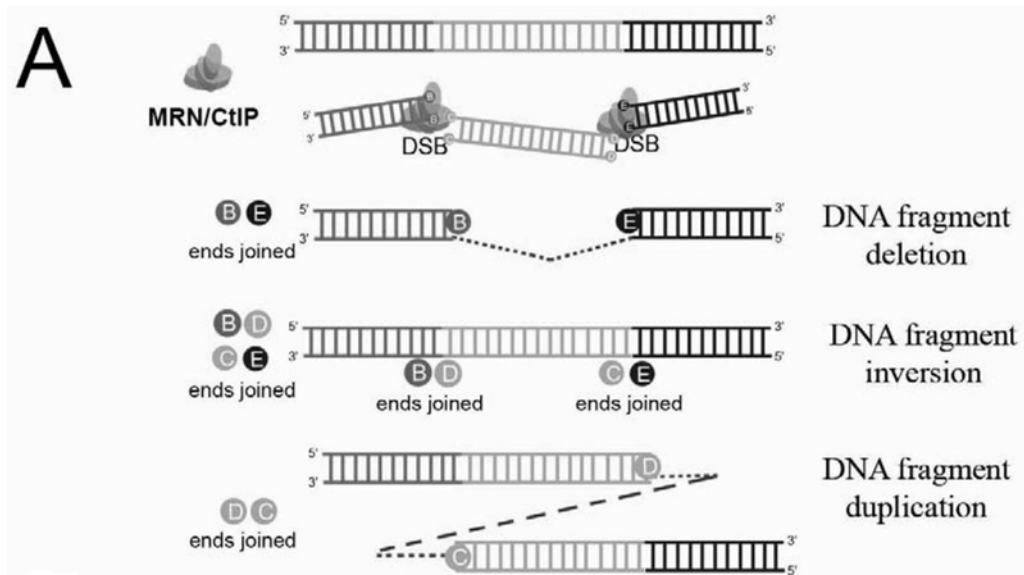


图1A

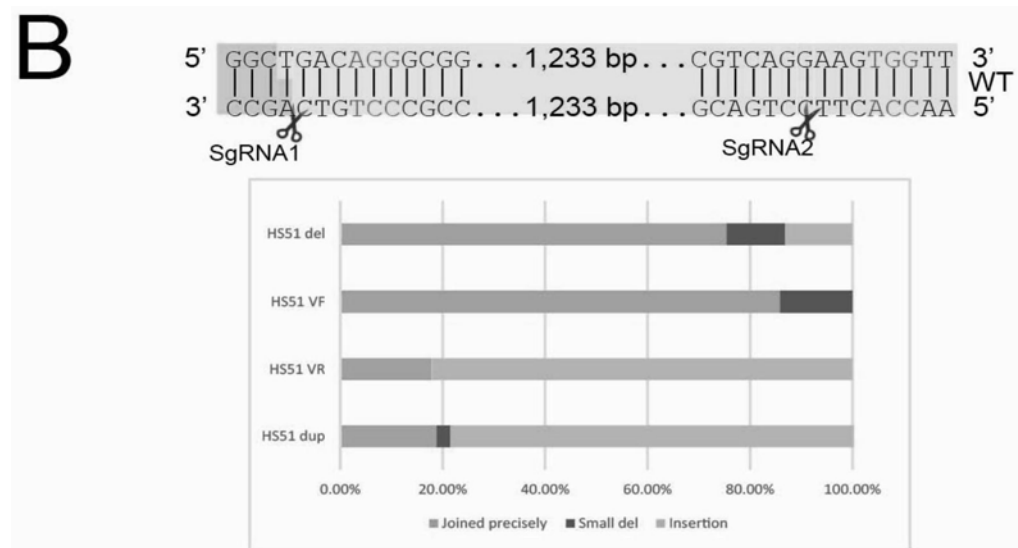


图1B

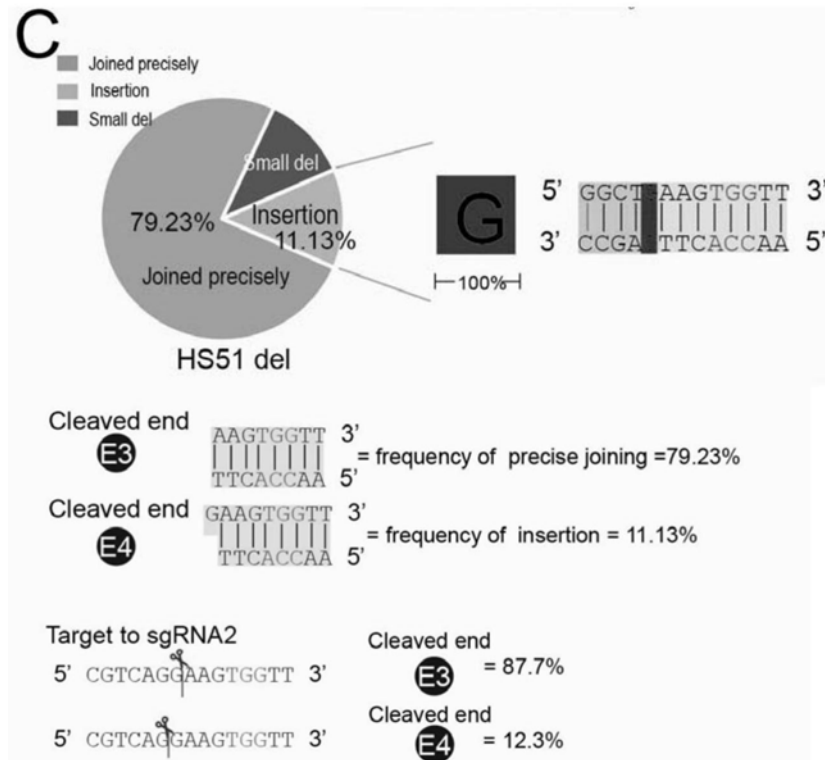


图1C

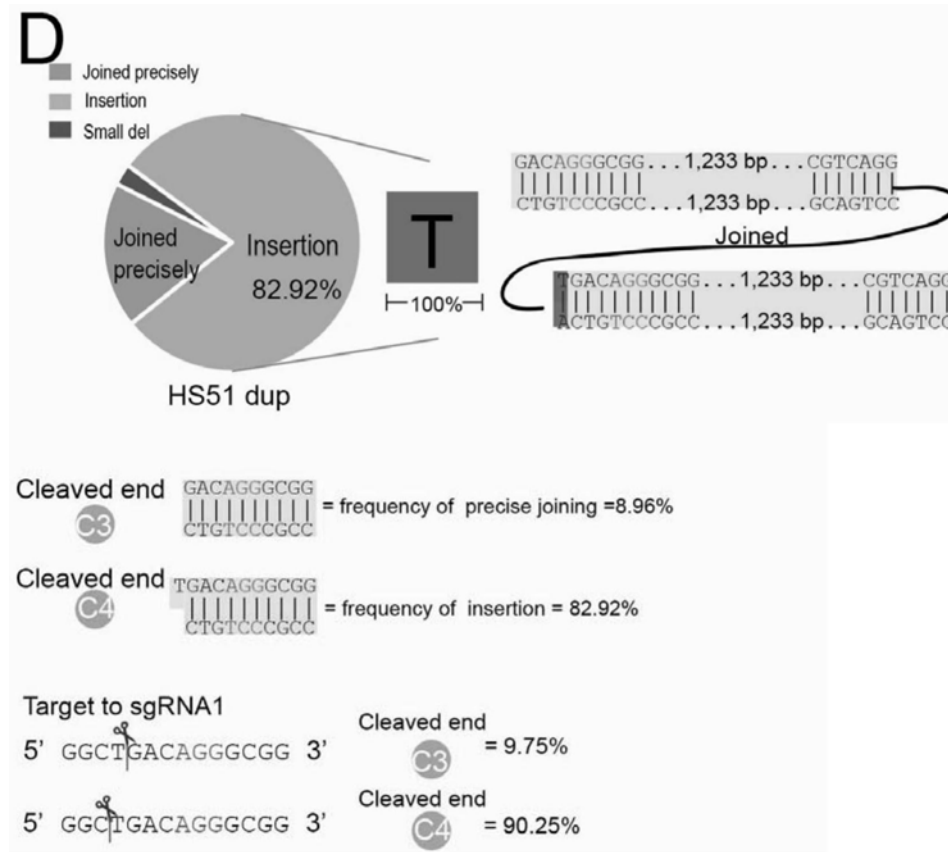


图1D

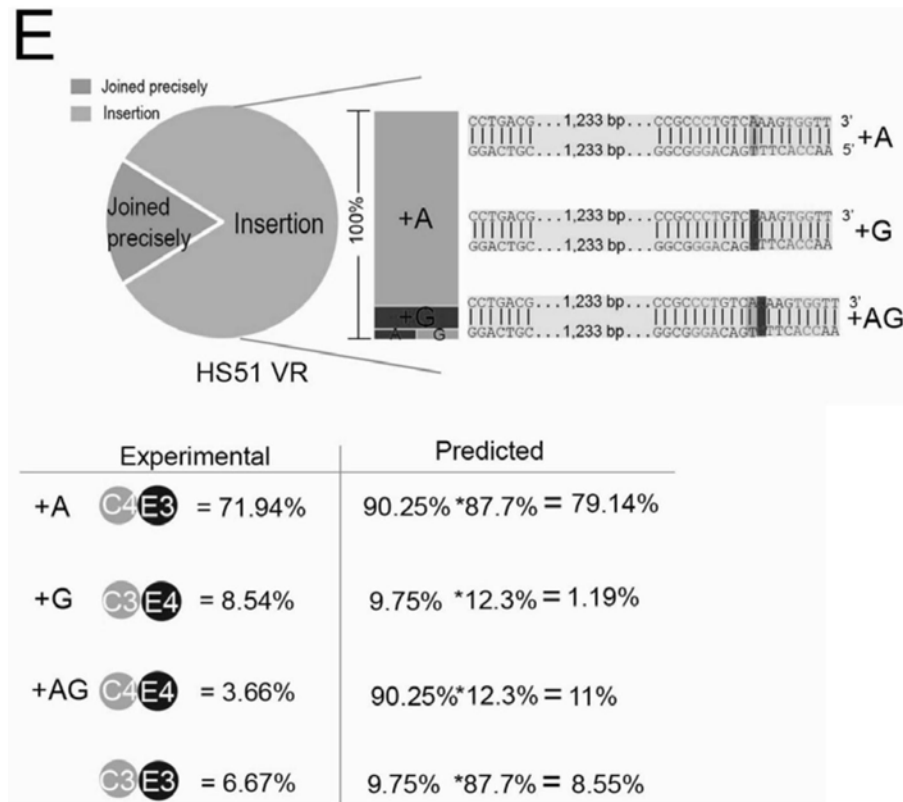


图1E

F Cleaved sites at the non complementary DNA strand by Cas9

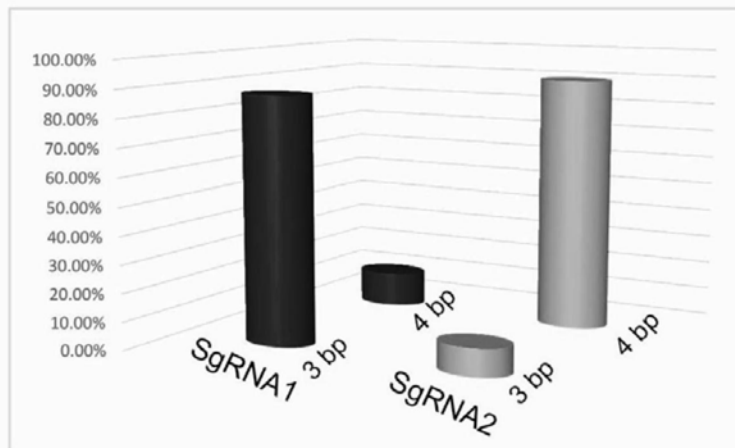


图1F

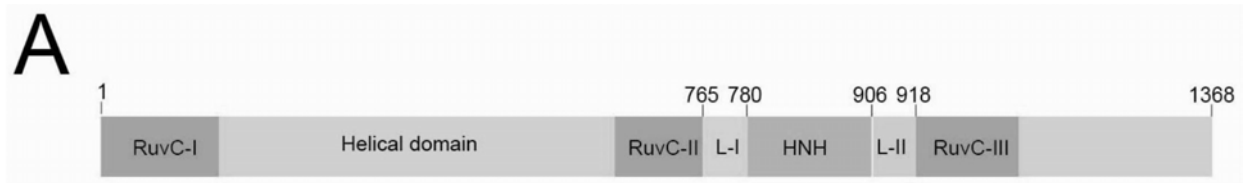


图2A

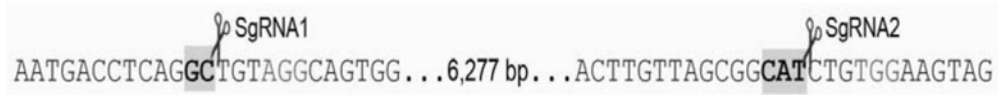


图2B

Insertion at the duplication junction

		Cas9	
		WT	Δ F916
		1	2
SgRNA1	NGG NCC	82.76	17.25
	+C		
	C NGG NCC	5.00	33.46
	+GC		
	GC NGG NCC	0.00	11.14

图2C

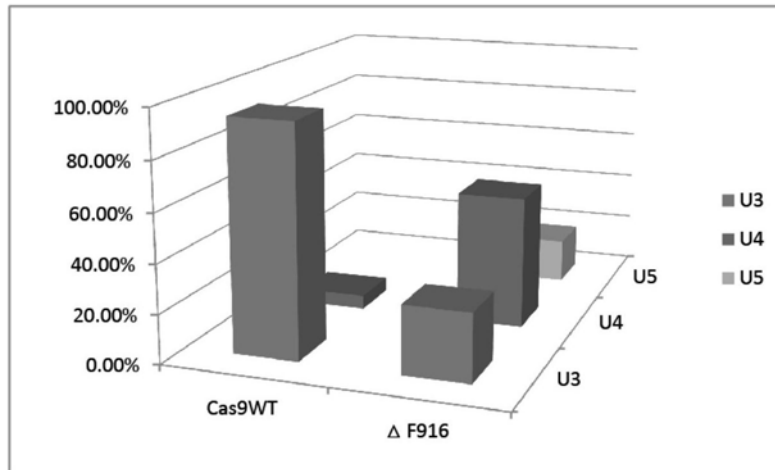
β -globin RE2

图2D

Insertion at the deletion junction

		Cas9	
		WT	Δ F916
		1	2
SgRNA2	NGG NCC	22.57	10.96
	+T		
	T NGG NCC	19.24	22.89
	+AT		
	AT NGG NCC	33.43	13.34
	+CAT		
	CAT NGG NCC	3.29	0.83

图2E

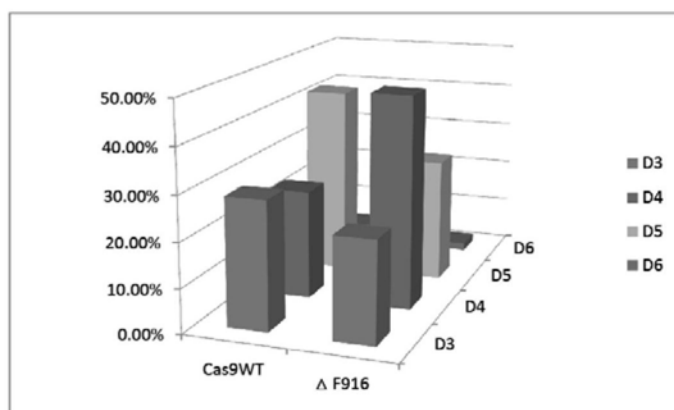
β -globin RE2

图2F

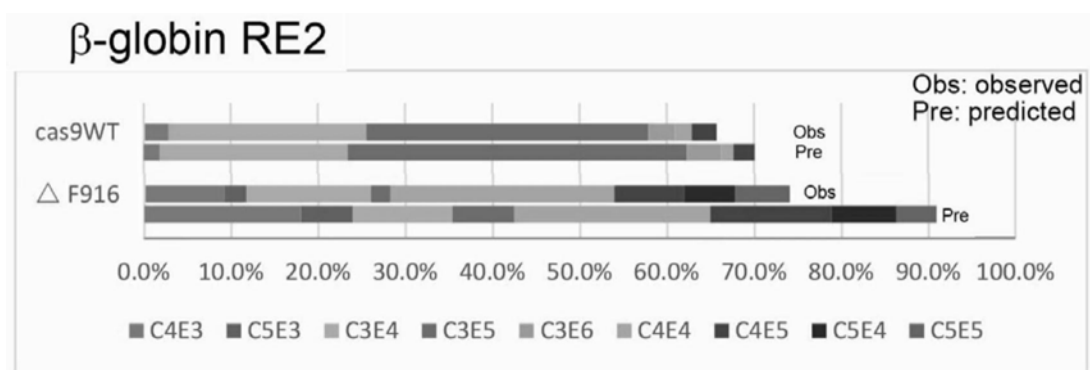


图2G

A STM

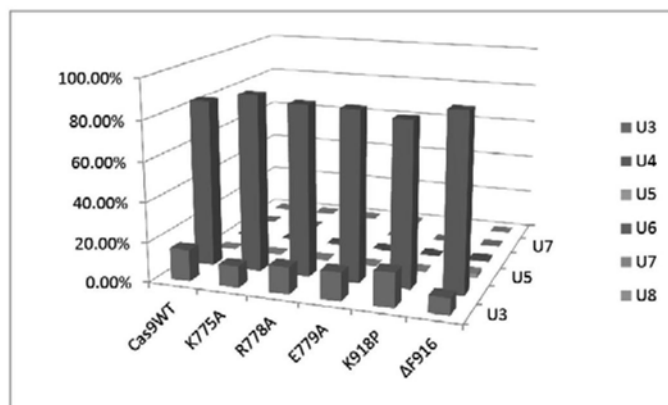


图3A

B STM

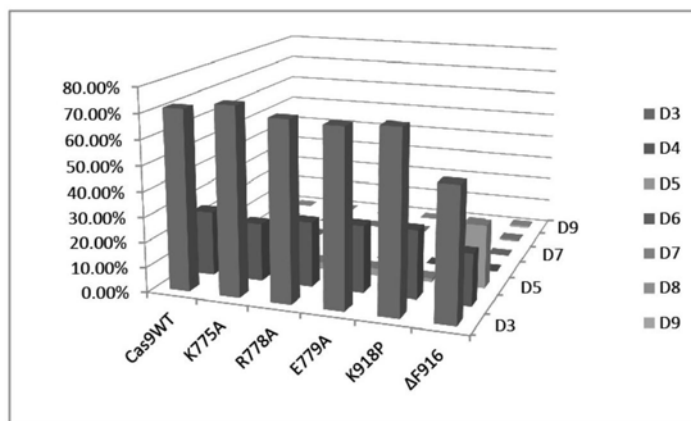


图3B