



(12)发明专利申请

(10)申请公布号 CN 110643600 A

(43)申请公布日 2020.01.03

(21)申请号 201910950683.X

(51)Int.Cl.

(22)申请日 2013.10.23

C12N 15/10(2006.01)

C12N 9/22(2006.01)

(30)优先权数据

C12N 15/113(2010.01)

61/717,324 2012.10.23 US

C12N 15/63(2006.01)

61/803,599 2013.03.20 US

C12N 15/82(2006.01)

61/837,481 2013.06.20 US

C12Q 1/683(2018.01)

(62)分案原申请数据

201380066348.4 2013.10.23

(71)申请人 基因工具股份有限公司

地址 韩国首尔

(72)发明人 金进秀 曹承于 金素贞

J·M·金 金爽中

(74)专利代理机构 北京市中咨律师事务所

11247

代理人 陈迎春 黄革生

权利要求书4页 说明书42页

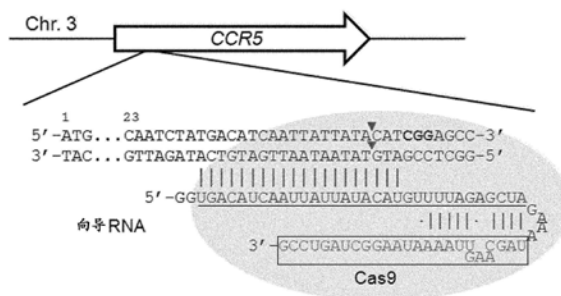
序列表42页 附图62页

(54)发明名称

用于切割靶DNA的系统及其用途

(57)摘要

本发明涉及真核细胞或生物体中的靶向基因组编辑。更具体地,本发明涉及用于在真核细胞或生物体中切割靶DNA的组合物及其用途,所述组合物包含特异于靶DNA的向导RNA和Cas蛋白质编码核酸或Cas蛋白质。



1. 用于在真核细胞或真核生物中切割靶核酸的II型CRISPR/Cas系统,所述CRISPR/Cas系统包含:

- (i) Cas9多肽或编码Cas9多肽的核酸,和
- (ii) 特异于靶核酸的向导RNA或编码向导RNA的核酸,

其中该向导RNA是包含与靶核酸杂交的crRNA (CRISPR RNA) 和tracrRNA (反式激活crRNA) 的dualRNA,或者包含crRNA和tracrRNA的至少部分并与靶核酸杂交的单链向导RNA (sgRNA)。

2. 包含权利要求1的系统的分离的真核细胞或真核生物。

3. 一种体外切割真核细胞或真核生物中的靶核酸的方法,包括用权利要求1的系统转染包含靶核酸的真核细胞的步骤。

4. 一种制备真核细胞或真核生物的方法,包括用权利要求1的系统转染包含靶核酸的真核细胞或真核生物的步骤。

5. 制备基因组修饰的动物的方法,包括:将权利要求1的系统引入动物的胚胎;和将胚胎转入假孕代母的输卵管以产生基因组修饰的动物,其中所述动物和假孕代母不是人类。

6. 制备基因组修饰的植物的方法,包括:

- (1) 将权利要求1的系统转染进入包含靶核酸的真核细胞或真核生物的原生质体,和
- (2) 再生原生质体以产生基因组修饰的植物。

7. 权利要求3或权利要求4的方法制备的真核细胞或真核生物。

8. 改变至少一种基因产物的表达的方法,包括向包含和表达具有靶序列并编码基因产物的DNA分子的真核细胞引入包含载体的成簇的规律间隔的短回文重复序列 (CRISPR) — CRISPR相关 (Cas) (CRISPR-Cas) 系统,该载体包含:

- a) 编码与靶序列杂交的CRISPR-Cas系统向导RNA的核酸,和
- b) 编码Type-II Cas9蛋白质的核酸,

其中向导RNA靶向靶序列且Cas9蛋白质切割DNA分子,

由此该至少一种基因产物的表达被改变;和,

其中该Cas9蛋白质和该向导RNA不天然存在在一起。

9. II型CRISPR-Cas系统,其包含Cas9蛋白质和至少一种靶向真核细胞中的DNA分子靶序列并与该靶序列杂交的向导RNA,

其中该DNA分子编码至少一种基因产物,该真核细胞表达该至少一种基因产物,而Cas9蛋白质切割该DNA分子,

由此该至少一种基因产物的表达被改变;

其中该Cas9蛋白质和该向导RNA不天然存在在一起;和,

其中该向导RNA包含tracr序列与融合的向导序列。

10. 制备用于在真核细胞或真核生物中切割靶核酸的重组Cas9多肽的方法,包括:

- (i) 使用人类密码子使用表重构并使用寡核苷酸合成编码Cas9多肽的核酸序列,
- (ii) 使用重叠的寡核苷酸和聚合酶组装DNA片段,并克隆到T-载体,
- (iii) 通过重叠PCR,使用DNA片段组装全长Cas9序列,

(iv) 所述Cas9编码DNA片段亚克隆到p3s,在此载体中,包含HA表位和核定位信号 (NLS) 的肽标签加至Cas9的C末端,通过蛋白质印迹,使用抗HA抗体,确认细胞中Cas9蛋白质的表

达和核定位，

(v) 将Cas9盒亚克隆到pET28-b(+), 并转化到BL21 (DE3), 使用IPTG诱导Cas9表达。

11. 制备用于在真核细胞或真核生物中切割靶核酸的向导RNA的方法, 包括:

(i) 筛选靶核酸的基因组序列, 检查外显子中NGG基序的存在;

(ii) 通过两个部分重叠的寡核苷酸延伸产生一个或多个DNA模板;

(iii) 使用模板DNA转录向导RNA。

12. 编辑真核细胞基因组的方法, 其包括:

向真核细胞中引入组合物, 该组合物包含:

(A) 重组Cas9蛋白质或多肽或编码Cas9蛋白质或多肽的核酸, 其中所述Cas9蛋白质或多肽与核定位信号(NLS)连接, 和

(B) 合成的向导RNA或编码合成的向导RNA的核酸,

其中该合成的向导RNA包含CRISPR RNA (crRNA) 和反式激活crRNA (tracrRNA), 该crRNA包含i) 必要部分, 其能与tracrRNA的至少一部分杂交, 和ii) 与靶DNA互补的部分,

其中该向导RNA是单链向导RNA (sgRNA), 其中CRISPR RNA (crRNA) 的必要部分与反式激活crRNA (tracrRNA) 融合;

其中该Cas9蛋白质或多肽能够与合成的向导RNA复合形成活性核酸内切酶, 和在真核细胞基因组中诱导DNA切割,

其中该方法诱导真核细胞的内源靶位点处的组合物介导的突变, 该突变是以下至少一种真核细胞基因组中一个或多个核苷酸的插入, 和

真核细胞基因组中一个或多个核苷酸的缺失。

13. 用于在真核细胞中编辑基因组的组合物, 其包含:

(A) 重组Cas9蛋白质或多肽或编码Cas9蛋白质或多肽的核酸, 其中所述Cas9蛋白质或多肽与核定位信号(NLS)连接, 和

(B) 合成的向导RNA或编码合成的向导RNA的核酸,

其中该合成的向导RNA包含CRISPR RNA (crRNA) 和反式激活crRNA (tracrRNA), 该crRNA包含i) 必要部分, 其能与tracrRNA的至少一部分杂交, 和ii) 与靶DNA互补的部分,

其中该向导RNA是单链向导RNA (sgRNA), 其中CRISPR RNA (crRNA) 的必要部分与反式激活crRNA (tracrRNA) 融合;

其中该Cas9蛋白质或多肽能够与合成的向导RNA复合形成活性核酸内切酶, 并在真核细胞基因组中诱导DNA切割。

14. 基因组编码的真核细胞, 其具有由权利要求13的组合物介导的突变,

其中该突变是以下至少一种

在真核细胞基因组中的一个或多个核苷酸插入, 和

在真核细胞基因组中的一个或多个核苷酸缺失。

15. 与至少两个靶点互补结合的向导RNA, 其中所述两个靶点包括第一靶点和第二靶点, 其中所述第一靶点与第二靶点具有小于100%的同源性, 且其中所述第一靶点具有匹配部分和与第二靶点相比有至少一个不同核苷酸的错配部分, 该向导RNA包含:

crRNA (CRISPR RNA) 的互补部分,

其中该互补部分包含第一核苷酸和第二核苷酸,

其中第一核苷酸与第一靶点的第一对应核苷酸错配,该第一对应核苷酸在与该向导RNA的第一核苷酸对应的位置,其中第一靶点的第一对应核苷酸位于该第一靶点的匹配部分,

其中第二核苷酸与第一靶点的第二对应核苷酸匹配,该第二对应核苷酸在与该向导RNA的第二核苷酸对应的位置,其中第一靶点的第二对应核苷酸位于该第一靶点的错配部分,

其中该第二核苷酸与第二靶点的第三对应核苷酸错配,该第三对应核苷酸在与该向导RNA的第二核苷酸对应的位置,

其中该向导RNA包含5'末端和3'末端,和

其中该第一核苷酸位于向导RNA中自第二核苷酸的5'末端或3'末端的4个核苷酸的序列之内。

16. 用于在核酸样本中检测单核苷酸多态性 (SNP) 的方法,所述方法包括:

制备包含第一核苷酸和第二核苷酸且能够与至少两个靶点互补结合的向导RNA,

其中该两个靶点包括第一靶点和第二靶点,

其中该第一靶点与第二靶点具有小于100%的同源性,

其中该第一靶点具有匹配部分和a与第二靶点相比包含至少一个不同核苷酸的错配部分,

其中该一个不同核苷酸是单核苷酸多态性 (SNP),

其中该向导RNA包含5'末端和3'末端,

其中第一核苷酸与第一靶点的第一对应核苷酸错配,该第一对应核苷酸在与该向导RNA的第一核苷酸对应的位置,其中第一靶点的第一对应核苷酸位于该第一靶点的匹配部分,

其中第二核苷酸与第一靶点的第二对应核苷酸匹配,该第二对应核苷酸在与该向导RNA的第二核苷酸对应的位置,其中第一靶点的第二对应核苷酸位于该第一靶点的错配部分,

其中该第二核苷酸与第二靶点的第三对应核苷酸错配,该第三对应核苷酸在与该向导RNA的第二核苷酸对应的位置,

其中该第二对应核苷酸是SNP,

其中该第二核苷酸位于向导RNA自5'末端的第8至第20序列之间,

其中该第一核苷酸位于向导RNA中自第二核苷酸的5'末端或3'末端的4个核苷酸的序列之内;

用检测组合物处理该核酸样本,

其中该核酸样本包含第一靶点、第二靶点或两者,

其中该检测组合物包含制备的向导RNA和Cas蛋白质;和

识别核酸样本中的切割,其中核酸样本中的切割通过该组合在第一靶点诱导。

17. 设计能与至少两个靶点互补结合的向导RNA的方法,该向导RNA包含至少一个与两个靶点错配的核苷酸,所述两个靶点包括第一靶点和第二靶点,该方法包括:

选择包括单核苷酸多态性 (SNP) 的第一靶点,

其中该第一靶点具有至少一个与第二靶点不同的核苷酸,和该至少一个不同核苷酸是

SNP,

其中该第一靶点包括18bp至23bp序列的互补部分,

其中该第一靶点位于PAM序列的5'末端或3'末端,

其中该SNP位于自PAM序列的5'末端或3'末端第1至第12序列之间;

确定与第一靶点错配的向导RNA,

其中该向导RNA是与第二靶点错配,

其中该向导RNA包含5'末端和3'末端,

其中该向导RNA包含第一核苷酸和第二核苷酸,

其中第一核苷酸与第一靶点的第一对应核苷酸错配,该第一对应核苷酸在与该向导RNA的第一核苷酸对应的位置,其中第一靶点的第一对应核苷酸位于该第一靶点的匹配部分,

其中第二核苷酸与第一靶点的第二对应核苷酸匹配,该第二对应核苷酸在与该向导RNA的第二核苷酸对应的位置,其中第一靶点的第二对应核苷酸位于该第一靶点的错配部分,

其中该第二核苷酸与第二靶点的第三对应核苷酸错配,该第三对应核苷酸在与该向导RNA的第二核苷酸对应的位置,

其中该第二对应核苷酸是SNP,

其中该第二核苷酸位于向导RNA自5'末端的第8至第20序列之间,

其中该第一核苷酸位于向导RNA中自第二核苷酸的5'末端或3'末端的4个核苷酸的序列之内。

18. 用于检测单核苷酸多态性(SNP)的组合物,其包含:

能与至少两个靶点互补结合的向导RNA,或者编码该向导RNA的核酸序列,

其中该两个靶点包括第一靶点和第二靶点,

其中该第一靶点与第二靶点具有小于100%的同源性,

其中该第一靶点具有至少一个与第二靶点不同的核苷酸,

其中该向导RNA包含5'末端和3'末端,

其中该向导RNA包含第一核苷酸和第二核苷酸,

其中第一核苷酸与第一靶点的第一对应核苷酸错配,该第一对应核苷酸在与该向导RNA的第一核苷酸对应的位置,其中第一靶点的第一对应核苷酸位于该第一靶点的匹配部分,

其中第二核苷酸与第一靶点的第二对应核苷酸匹配,该第二对应核苷酸在与该向导RNA的第二核苷酸对应的位置,其中第一靶点的第二对应核苷酸位于该第一靶点的错配部分,

其中该第二核苷酸与第二靶点的第三对应核苷酸错配,该第三对应核苷酸在与该向导RNA的第二核苷酸对应的位置,

其中该第二对应核苷酸是SNP,

其中该第一核苷酸位于向导RNA中自第二核苷酸的5'末端或3'末端的4个核苷酸的序列之内;和

Cas蛋白质或Cas蛋白质的核酸序列。

## 用于切割靶DNA的系统及其用途

[0001] 本申请是申请日为2013年10月23日的、发明名称为“包含特异于靶 DNA的向导RNA和CAS蛋白质编码核酸或CAS蛋白质的用于切割靶 DNA的组合物及其用途”的中国专利申请CN201380066348.4 (PCT/ KR2013/009488)的分案申请。

### 技术领域

[0002] 本发明涉及真核细胞或生物体中的靶向基因组编辑。更具体地说,本发明涉及一种用于在真核细胞或生物体中切割靶DNA的组合物及其用途,所述组合物包括特异于靶DNA的向导RNA和Cas蛋白质编码核酸或Cas 蛋白质。

### 背景技术

[0003] CRISPR (成簇的规律间隔的短回文重复序列) 是含有多个短同向重复的基因座,其被发现存在于约40%测序细菌的基因组中和90%测序古细菌的基因组中。CRISPR作为原核的免疫系统发挥功能,其赋予对外来遗传元件例如质粒和噬菌体的抵抗性。CRISPR系统提供了一种获得性免疫形式。外源DNA的短片段 (称为间隔区) 整合在CRISPR重复序列之间的基因组中,作为过去暴露的记忆。然后CRISPR间隔区以类似于真核生物中 RNAi的方式用于识别和沉默外来遗传元件。

[0004] Cas9, II型CRISPR/Cas系统中一种重要的蛋白质成分,当与称为 CRISPR RNA (crRNA) 和反式激活crRNA (tracrRNA) 的两个RNA复合时,形成活性核酸内切酶,从而切断入侵噬菌体或质粒中的外源遗传元件,以保护宿主细胞。crRNA从宿主基因组中的CRISPR元件转录,其中该 CRISPR元件之前自外源入侵物捕获。最近, Jinek等 (1) 证明,通过融合 crRNA和tracrRNA的必要部分产生的单链嵌合RNA可以取代Cas9/RNA 复合体中的两个RNA以形成功能性核酸内切酶。

[0005] CRISPR/Cas系统相对于锌指和转录激活因子样效应物DNA结合蛋白提供了优势——因为在核苷酸结合CRISPR-Cas蛋白中的位点特异性由RNA分子调控而不是DNA结合蛋白调控 (这在设计和合成上是更具挑战性的)。

[0006] 然而,到现在为止,尚未开发出使用基于CRISPR/Cas系统的RNA向导核酸内切酶 (RGEN) 的基因组编辑方法。

[0007] 同时,限制性片段长度多态性 (RFLP) 是最古老,最方便,和最便宜的基因分型方法之一,其仍然广泛应用于分子生物学和遗传学,但其往往受限于缺乏适当的限制性内切酶识别位点。

[0008] 可以通过各种方法检测由工程化核酸酶诱导的突变,其中包括错配敏感的T7核酸内切酶I (T7E1) 或Surveyor核酸酶测定法,RFLP,荧光PCR 产物的毛细管电泳,双脱氧测序和深度测序。T7E1和Surveyor测定法广泛使用,但很繁琐。此外,这些酶倾向于低估突变频率,这是因为突变序列可彼此形成同源双链,从而不能从野生型细胞中区分纯合双等位基因突变体克隆。RFLP没有这些限制,因而是首选的方法。实际上,RFLP是检测细胞和动物中由工程化核酸酶介导的突变的最早方法之一。然而,不幸的是,RFLP受限于适当限制性位点

的可得性。在所关注的靶位点有可能没有限制性位点。

## 发明内容

### [0009] 技术问题

[0010] 到现在为止,尚未开发使用基于CRISPR/Cas系统的RNA向导核酸内切酶(RGEN)进行基因组编辑和基因分型的方法。

[0011] 在这种情况下,本发明人进行了大量努力来开发基于CRISPR/Cas系统的基因组编辑方法,最终建立了一个可程序化的RNA向导核酸内切酶,该RNA向导核酸内切酶可以在真核细胞和生物体中以靶向方式切割DNA。

[0012] 另外,本发明人进行了大量努力,开发一种新的在RFLP分析中利用RNA向导核酸内切酶(RGEN)的方法。其利用RGEN,对癌症中发现的以及细胞和生物体中由工程化核酸酶(包括RGEN自身)诱导的频发突变进行基因分型,从而完成了本发明。

### [0013] 技术方案

[0014] 本发明的一个目的是提供一种在真核细胞或生物体中切割靶DNA的组合物,其包括特异于靶DNA的向导RNA或编码向导RNA的DNA和Cas蛋白质编码核酸或Cas蛋白质。

[0015] 本发明的另一个目的是提供一种在真核细胞或生物体中诱导靶向诱变的组合物,其包括特异于靶DNA的向导RNA或编码向导RNA的DNA和Cas蛋白质编码核酸或Cas蛋白质。

[0016] 本发明的另一个目的是提供一种在真核细胞或生物体中切割靶DNA的试剂盒,其包括特异于靶DNA的向导RNA或编码向导RNA的DNA和Cas蛋白质编码核酸或Cas蛋白质。

[0017] 本发明的另一个目的是提供一种在真核细胞或生物体中诱导靶向诱变的试剂盒,其包括特异于靶DNA的向导RNA或编码向导RNA的DNA和Cas蛋白质编码核酸或Cas蛋白质。

[0018] 本发明的再一目的是提供一种制备含有Cas蛋白质和向导RNA的真核细胞或生物体的方法,所述方法包括用Cas蛋白质编码核酸或Cas蛋白质以及向导RNA或编码向导RNA的DNA共转染或顺序转染真核细胞或生物体的步骤。

[0019] 本发明的另一个目的是提供一种真核细胞或生物体,其含有特异于靶DNA的向导RNA或编码向导RNA的DNA和Cas蛋白质编码核酸或Cas蛋白质。

[0020] 本发明的另一个目的是提供一种在真核细胞或生物体中切割靶DNA的方法,所述方法包括步骤:用组合物转染含有靶DNA的真核细胞或生物体,所述组合物含有特异于靶DNA的向导RNA或编码向导RNA的DNA和Cas蛋白质编码核酸或Cas蛋白质。

[0021] 本发明的另一个目的是提供一种在真核细胞或生物体中诱导靶向诱变的方法,所述方法包括步骤:用组合物处理真核细胞或生物体,其中所述组合物含有特异于靶DNA的向导RNA或编码向导RNA的DNA和Cas蛋白质编码核酸或Cas蛋白质。

[0022] 本发明的再一个目的是提供胚胎、基因组修饰的动物或基因组修饰的植物,其包括由组合物编辑的基因组,所述组合物含有特异于靶DNA的向导RNA或编码向导RNA的DNA和Cas蛋白质编码核酸或Cas蛋白质。

[0023] 本发明的另一个目的是提供一种制备基因组修饰的动物的方法,所述方法包括步骤:将含有特异于靶DNA的向导RNA或编码向导RNA的DNA和Cas蛋白质编码核酸或Cas蛋白质的组合物,引入动物胚胎中;和将胚胎转移到假孕代母的输卵管中,以产生基因组修饰的动物。

[0024] 本发明的另一个目的是提供一种组合物,其用于在分离的生物样品中基因分型突变或变异,所述组合物包含特异于靶DNA序列的向导RNA和 Cas蛋白质。

[0025] 本发明的另一个目的是提供一种使用RNA向导核酸内切酶 (RGEN) 对细胞中由工程化的核酸酶诱导的突变或天然存在的突变或变异进行基因分型的方法,其中所述RGEN包含特异于靶DNA的向导RNA和Cas蛋白质。

[0026] 本发明的另一个目的是提供对细胞中由工程化的核酸酶诱导的突变或天然存在的突变或变异进行基因分型的试剂盒,所述试剂盒含有RNA向导核酸内切酶 (RGEN),其中所述RGEN含有特异于靶DNA的向导RNA和 Cas蛋白质。

[0027] 本发明的一个目的是提供在真核细胞或生物体中切割靶DNA的组合物,所述组合物含有特异于靶DNA的向导RNA或编码向导RNA的DNA 和Cas蛋白质编码核酸或Cas蛋白质。

[0028] 本发明的另一个目的是提供在真核细胞或生物体中诱导靶向诱变的组合物,所述组合物含有特异于靶DNA的向导RNA或编码向导RNA的DNA 和Cas蛋白质编码核酸或Cas蛋白质。

[0029] 本发明的另一个目的是提供在真核细胞或生物体中切割靶DNA的试剂盒,所述试剂盒含有特异于靶DNA的向导RNA或编码向导RNA的DNA 和Cas蛋白质编码核酸或Cas蛋白质。

[0030] 本发明的另一个目的是提供在真核细胞或生物体中诱导靶向诱变的试剂盒,所述试剂盒含有特异于靶DNA的向导RNA或编码向导RNA的DNA 和Cas蛋白质编码核酸或Cas蛋白质。

[0031] 本发明的另一目的是提供制备含有Cas蛋白质和向导RNA的真核细胞或生物体的方法,所述方法包括用Cas蛋白质编码核酸或Cas蛋白质以及向导RNA或编码向导RNA的DNA共转染或顺序转染真核细胞或生物体的步骤。

[0032] 本发明的另一个目的是提供一种真核细胞或生物体,其含有特异于靶 DNA的向导RNA或编码向导RNA的DNA和Cas蛋白质编码核酸或Cas 蛋白质。

[0033] 本发明的另一个目的是提供一种在真核细胞或生物体中切割靶DNA 的方法,所述方法包括用组合物转染含有靶DNA的真核细胞或生物体的步骤,所述组合物含有特异于靶DNA的向导RNA或编码向导RNA的DNA 和Cas蛋白质编码核酸或Cas蛋白质。

[0034] 本发明的另一个目的是提供一种在真核细胞或生物体中诱导靶向诱变的方法,所述方法包括用组合物处理真核细胞或生物体的步骤,所述组合物含有特异于靶DNA的向导RNA或编码向导RNA的DNA和Cas蛋白质编码核酸或Cas蛋白质。

[0035] 本发明的另一个目的是提供胚胎、基因组修饰的动物、或基因组修饰的植物,其包含由组合物编辑的基因组,所述组合物含有特异于靶DNA 的向导RNA或编码向导RNA的DNA和Cas蛋白质编码核酸或Cas蛋白质。

[0036] 本发明的另一个目的是提供一种制备基因组修饰的动物的方法,所述方法包括步骤:将含有特异于靶DNA的向导RNA或编码向导RNA的DNA 和Cas蛋白质编码核酸或Cas蛋白质的组合物,引入动物胚胎中;和将胚胎转移到假孕代母的输卵管,以产生基因组修饰的动物。

[0037] 本发明的另一个目的是提供在分离的生物样品中基因分型突变或变异的组合物,所述组合物含有特异于靶DNA序列的向导RNA和Cas蛋白质。

[0038] 本发明的另一个目的是提供在分离的生物样品中对病原微生物的核酸序列进行基因分型的组合物,所述组合物含有特异于靶DNA序列的向导 RNA和Cas蛋白质。

[0039] 本发明的另一个目的是提供一种在分离的生物样品中基因分型突变或变异的试剂盒,所述试剂盒含有组合物,特别地含有RNA向导核酸内切酶 (RGEN),其中,所述RGEN包含特异于靶DNA的向导RNA和Cas蛋白质。

[0040] 本发明的另一个目的是提供在分离的生物样品中基因分型突变或变异的方法,该方法使用组合物,特别地所述组合物包含RNA向导核酸内切酶 (RGEN),其中,所述RGEN包含特异于靶DNA的向导RNA和Cas蛋白质。

[0041] 有益效果

[0042] 包含特异于靶DNA的向导RNA和Cas蛋白质编码核酸或Cas蛋白质的用于在真核细胞或生物体中切割靶DNA或诱导靶向诱变的本发明组合物、包含所述组合物的本发明试剂盒、以及诱导靶向诱变的本发明方法,提供了新的方便的基因组编辑工具。另外,因为可以设计定制RGEN以靶向任何DNA序列,所以几乎任何单核苷酸多态性或小的插入/缺失(indel) 均可以通过RGEN介导的RFLP进行分析,因此,本发明的组合物和方法可以用于检测和切割天然存在的变异和突变。

[0043] 附图简述

[0044] 图1显示Cas9催化的体外质粒DNA切割。(a) 靶DNA和嵌合RNA 序列的示意图。红色三角指示切割位点。Cas9识别的PAM序列以粗体显示。向导RNA中源自crRNA和tracrRNA的序列分别以框和下划线显示。(b) Cas9体外切割质粒DNA。完整的环状质粒或ApaLI消化的质粒与 Cas9和向导RNA孵育。

[0045] 图2显示Cas9在附加体(episomal) 靶位点诱导的诱变。(a) 使用 RFP-GFP报告分子的基于细胞测定法的概略示意图。因为GFP序列在读框外融合至RFP序列,故GFP不从该报告分子表达。只有当两个序列之间的靶位点被位点特异性核酸酶切割后,RFP-GFP融合蛋白才表达。(b) 转染了Cas9的细胞的流式细胞术。显示表达RFP-GFP融合蛋白的细胞的百分比。

[0046] 图3显示在内源染色体位点上由RGEN驱动的突变。(a) CCR5基因座。(b) C4BPB基因座。(顶部) 使用T7E1测定法检测RGEN驱动的突变。箭头指示由T7E1切割的DNA条带的预期位置。通过测量条带强度,计算突变频率(Indel (%))。(底部) CCR5和C4BPB野生型(WT) 和突变克隆的DNA序列。互补于向导RNA的靶序列区域示于框中(in box)。PAM序列以粗体显示。三角形指示切割位点。对应于微同源性 (microhomologies) 的碱基加下划线。右手列显示插入或缺失碱基的数目。

[0047] 图4显示不能检测到RGEN驱动的脱靶(off-target) 突变。(a) 标靶 (On-target) 序列和潜在的脱靶序列。在硅片上搜索人类基因组,以寻找潜在的脱靶位点。鉴定了四个位点,其中每个携带3个碱基的与CCR5标靶位点的错配。错配碱基加下划线。(b) 使用T7E1测定法调查这些位点是否在转染Cas9/RNA复合体的细胞中发生突变。在这些位点没有检测到突变。N/A(不适用),基因间位点。(c) Cas9没有诱导脱靶相关的染色体缺失。在人细胞中表达CCR5特异的RGEN和ZFN。使用PCR检测在这些细胞中15-kb染色体缺失的诱导。

[0048] 图5显示小鼠中RGEN诱导的Foxn1基因打靶。(a) 描绘了特异于小鼠Foxn1基因外显子2的sgRNA的示意图。外显子2中的PAM表示为红色,sgRNA中互补于外显子2的序列加下划线。三角形指示切割位点。(b) 代表性的T7E1测定,表明通过胞质内注射而递送至一细胞阶

段的小鼠胚胎中的Cas9 mRNA+Foxn1特异性sgRNA的基因打靶效率。数字指示由最高剂量产生的独立建立者小鼠。箭头指示由T7E1切割的条带。(c) b中鉴定的三个Foxn1突变建立者小鼠中观察到的突变等位基因的DNA序列。发生数显示在括号中。(d)对Foxn1建立者小鼠#108和野生型 FVB/NTac杂交产生的F1后代进行的PCR基因分型。注意,存在于Foxn1 建立者小鼠#108中的突变等位基因在后代中分离。

[0049] 图6显示通过胞质内注射Cas9 mRNA和Foxn1-sgRNA在小鼠胚胎中的Foxn1基因打靶。(a)代表性T7E1测定的结果,其监测注射最高剂量后的突变率。箭头指示由T7E1切割的条带。(b)T7E1检测结果总结。显示在胞质内注射所示剂量RGEN后获得的体外培养胚胎中的突变体分数。(c)从T7E1阳性突变胚胎子集鉴定的Foxn1突变等位基因的DNA 序列。野生型等位基因的靶序列表示在框内。

[0050] 图7显示使用重组Cas9蛋白质:Foxn1-sgRNA复合体在小鼠胚胎中的 Foxn1基因打靶。(a)和(b)是代表性T7E1测定结果和其总结。胚胎进行原核(a)或胞质内注射(b)后体外培养。红色数字表示T7E1阳性突变建立者小鼠。(c)从体外培养的胚胎鉴定的Foxn1突变等位基因的 DNA序列,该胚胎通过以最高剂量原核注射重组Cas9蛋白质:Foxn1-sgRNA复合体获得。野生型等位基因的靶序列表示在框内。

[0051] 图8显示在Foxn1突变建立者#12中发现的突变等位基因的种系传播。(a) fPCR分析。(b)PCR基因分型野生型FVB/NTac、建立者小鼠和他们的F1后代。

[0052] 图9显示通过杂交Prkdc突变建立者小鼠产生的胚胎的基因型。Prkdc 突变建立者小鼠♂25和♀15杂交并分离E13.5胚胎。(a) fPCR分析野生型,建立者小鼠♂25,建立者小鼠♀15。需要注意的是,由于fPCR分析的技术局限性,这些结果显示出与突变等位基因的精确序列有微小差别;例如,从序列分析,在建立者小鼠♂25和♀15中分别鉴定了 $\Delta 269/\Delta 61/WT$ 和 $\Delta 5+1/+7/+12/WT$ 。(b)产生的胚胎的基因型。

[0053] 图10显示Cas9蛋白质/sgrRNA复合体诱导的靶向突变。

[0054] 图11表示重组Cas9蛋白质在拟南芥原生质体中诱导的突变。

[0055] 图12表示重组Cas9蛋白质在拟南芥BRI1基因中诱导的突变序列。

[0056] 图13显示T7E1测定,其表明在293细胞中通过Cas9-mal-9R4L和 sgRNA/C9R4LC复合体处理破坏内源性CCR5基因。

[0057] 图14(a,b)显示Fu等(2013)报道的RGEN在靶位点和脱靶位点上的突变频率。T7E1测定分析了来自K562细胞的基因组DNA,所述K562 细胞顺序转染了20 $\mu$ g Cas9编码质粒和分别为60 $\mu$ g和120 $\mu$ g的体外转录的GX19 crRNA和tracrRNA( $1 \times 10^6$ 个细胞)(R),或共转染了1 $\mu$ g Cas9编码质粒和1 $\mu$ g GX19 sgRNA表达质粒( $2 \times 10^5$ 个细胞)(D)。

[0058] 图15(a,b)显示向导RNA结构的比较。使用T7E1测定法,测量了靶和脱靶位点上的Fu等(2013)报道的RGEN的突变频率。K562细胞共转染了Cas9编码质粒和编码GX19 sgRNA或GGX20 sgRNA的质粒。脱靶位点(OT1-3等)如Fu等(2013)中标记。

[0059] 图16显示Cas9切口酶在体外DNA切割。(a)Cas9核酸酶和配对的 Cas9切口酶的示意图概略。PAM序列和切割位点显示在框中。(b)人类 AAVS1基因座中的靶位点。每个靶位点的位置以三角形表示。(c)DNA 切割反应的示意图概略。FAM染料(示于框内)连接至DNA底物的两个5'末端。(d)使用荧光毛细管电泳分析DSB和SSB。荧光标记的DNA底物在电泳前与Cas9核酸酶或切口酶孵育。

[0060] 图17显示Cas9核酸酶和切口酶的行为比较。(a)与Cas9核酸酶(WT)、切口酶(D10A)、和配对切口酶相关的靶突变频率。显示产生5'突出端或3'突出端的配对切口酶。(b)Cas9核酸酶和配对切口酶的脱靶效应分析。分析了三个sgRNA的总共7个潜在的脱靶位点。

[0061] 图18显示在其他内源性人基因座测试的配对Cas9切口酶。(a,c)在人CCR5和BRCA2基因座上的sgRNA靶位点。PAM序列显示为红色。(b,d)通过T7E1测定法检测每个靶位点上的基因组编辑活性。产生5'突出端的两个缺口的修复比产生3'突出端的两个缺口的修复导致了频繁得多的indel形成。

[0062] 图19显示配对Cas9切口酶介导同源重组。(a)检测同源重组的策略。供体DNA包括两个同源臂之间的XbaI限制性酶切位点,而内源性靶位点缺乏这个限制性酶切位点。PCR测定法用来检测已经进行同源重组的序列。为了防止污染供体DNA的扩增,使用对基因组DNA特异的引物。(b)同源重组的效率。只有发生了同源重组的区域的扩增子可以用XbaI消化;切割条带的强度被用来测量该方法的效率。

[0063] 图20显示由配对Cas9切口酶诱导的DNA拼接。(a)人AAVS1基因座中配对切口酶的靶位点。示出了AS2位点和其他各位点之间的距离。箭头指示PCR引物。(b)用PCR检测基因组缺失。星号表示缺失特异的PCR产物。(c)使用AS2和L1 sgRNA获得的缺失特异性PCR产物的DNA序列。靶位点PAM序列显示在框内,sgRNA匹配序列以大写字母表示。完整的sgRNA匹配序列带下划线。(d)配对Cas9切口酶介导的染色体缺失的示意模型。新合成的DNA链示于框中。

[0064] 图21显示配对Cas9切口酶不诱导易位。(a)靶位点和脱靶位点之间染色体易位的示意图概略。(b)PCR扩增以检测染色体易位。(c)由Cas9核酸酶而不是由切口酶对诱导的易位。

[0065] 图22显示T7E1和RFLP测定法的概念图。(a)二倍体细胞中工程化的核酸酶处理后四个可能的情况中测定切割反应的比较:(A)野生型,(B)单等位基因突变,(C)不同的双等位基因突变(杂合),和(D)相同的双等位基因突变(纯合)。黑线代表源自每个等位基因的PCR产物;短虚线框和点框指示由NHEJ产生的插入/缺失突变。(b)由电泳解析的T7E1和RGEN消化物的预期结果。

[0066] 图23示出含有C4BPB靶位点(带有indel)的线性化质粒的体外切割测定。各质粒底物的DNA序列(上图)。PAM序列加下划线。插入碱基显示在框中。箭头(下图)指示用野生型特异的RGEN切割的DNA条带在电泳后的预期位置。

[0067] 图24显示通过RGEN介导的RFLP对细胞中由工程化的核酸酶诱导的突变进行的基因分型。(a)C4BPB突变体K562细胞克隆的基因型。(b)比较错配敏感性T7E1测定法与RGEN介导的RFLP分析。黑色箭头指示通过T7E1酶或RGEN处理的切割产物。

[0068] 图25显示通过RGEN-RFLP技术基因分型RGEN诱导的突变。(a)使用RGEN-RFLP和T7E1测定法分析C4BPB破坏的克隆。箭头指示RGEN或T7E1切割的DNA条带的预期位置。(b)定量比较RGEN-RFLP分析与T7E1测定法。以不同比例混合来自野生型和C4BPB破坏的K562细胞的基因组DNA样品,并进行PCR扩增。(c)用RFLP和T7E1测定法对HeLa细胞中HLA-B基因中RGEN诱导的突变进行的基因分型。

[0069] 图26显示通过RGEN介导的RFLP对生物体中由工程化的核酸酶诱导的突变进行的

基因分型。(a) Pibf1突变体建立者小鼠的基因型。(b) 比较错配敏感性T7E1测定法与RGEN介导的RFLP分析。黑色箭头指示 T7E1酶或RGEN处理的切割产物。

[0070] 图27显示对ZFN诱导的突变进行RGEN介导的基因分型。ZFN靶位点示于框中。黑色箭头指示由T7E1切割的DNA条带。

[0071] 图28显示在人HLA-B基因区域中的多态性位点。包围RGEN靶位点的序列,是来自HeLa细胞的PCR扩增子的序列。多态性位置显示在框中。RGEN靶位点和PAM序列分别示于虚线框和粗体框中。引物序列加下划线。

[0072] 图29显示通过RGEN-RFLP分析致癌突变的基因分型。(a) 通过RGEN 检测HCT116细胞中人CTNNB1基因中的频发突变(TCT的c.133-135缺失)。HeLa细胞用作阴性对照。(b) 用包含错配向导RNA的RGEN在 A549癌细胞系中基因分型KRAS取代突变(c.34G>A)。错配的核苷酸示于框中。HeLa细胞用作阴性对照。箭头表示RGEN切割的DNA条带。示出通过Sanger测序证实的DNA序列。

[0073] 图30显示通过RGEN-RFLP分析在HEK293T细胞中基因分型CCR5 delta32等位基因。(a) 细胞系的RGEN-RFLP测定。K562, SKBR3和 HeLa细胞被用作野生型对照。箭头指示RGEN切割的DNA条带。(b) 野生型和delta32 CCR5等位基因的DNA序列。在RFLP分析中使用的 RGEN的标靶和脱靶位点加下划线。两个位点之间的单核苷酸错配显示在框中。PAM序列加下划线。(c) 使用野生型特异的RGEN在体外切割携带WT或de132 CCR5等位基因的质粒。(d) 确认在CCR5基因座上存在 CCR5-delta32特异的RGEN的脱靶位点。使用各种量的de132特异的RGEN 对带有标靶序列或脱靶序列的质粒进行体外切割测定。

[0074] 图31显示KRAS点突变(c.34G>A)的基因分型。(a) 癌细胞系中 KRAS突变(c.34G>A)的RGEN-RFLP分析。用具有特异于野生型序列或突变序列的完美匹配的crRNA的RGEN,消化来自HeLa细胞(用作野生型对照)或A549细胞(点突变纯合)的PCR产物。在这些细胞中KRAS基因型通过Sanger测序证实。(b) 使用带有完美匹配的crRNA或减弱的单碱基错配的crRNA的RGEN消化质粒(其携带野生型或突变的KRAS序列)。被选择用于基因分型的减弱的crRNA在凝胶上方的框中标出。

[0075] 图32显示PIK3CA点突变(c.3140A>G)基因分型。(a) 癌细胞系中PIK3CA突变(c.3140A>G)的RGEN-RFLP分析。用具有特异于野生型序列或突变序列的完美匹配的crRNA的RGEN,消化来自HeLa细胞(用作野生型对照)或HCT116细胞(杂合点突变)的PCR产物。在这些细胞中 PIK3CA基因型通过Sanger测序证实。(b) 使用带有完美匹配的crRNA 或减弱的单碱基错配的crRNA的RGEN,消化携带野生型或突变PIK3CA 序列的质粒。被选择用于基因分型的减弱的crRNA在凝胶上方的框中标出。

[0076] 图33显示在癌细胞系中频发点突变的基因分型。RGEN-RFLP测定在 (a) IDH(c.394c>T), (b) PIK3CA(c.3140A>G), (c) NRAS(c.181C> A), 和(d) BRAF基因(c.1799T>A)中的频发致癌点突变。显示用Sanger 测序证实的每个细胞系的基因型。错配核苷酸示于框中。黑色箭头指示由 RGEN切割的DNA条带。

[0077] 优选实施方式

[0078] 根据本发明的一个方面,本发明提供了用于在真核细胞或生物体中切割靶DNA的组合物,其包括特异于靶DNA的向导RNA或编码该向导RNA 的DNA、和Cas蛋白质编码核酸或Cas蛋白质。另外,本发明提供该组合物用于在真核细胞或生物体中切割靶DNA的用途,所述

组合物包括特异于靶DNA的向导RNA或编码该向导RNA的DNA、和Cas蛋白质编码核酸或Cas蛋白质。

[0079] 在本发明中,该组合物也被称为RNA向导核酸内切酶(RGEN)组合物。

[0080] ZFN和TALEN能够在哺乳动物细胞,模式生物,植物和家畜中实现靶向诱变,但用各核酸酶得到的突变频率彼此大不相同。此外,一些ZFN和TALEN不显示任何基因组编辑活性。DNA甲基化可能会限制这些工程化的核酸酶与靶位点结合。此外,产生定制核酸酶在技术上是富于挑战且费时的。

[0081] 本发明人已开发了基于Cas蛋白质的新的RNA向导核酸内切酶组合物,以克服ZFN和TALEN的缺点。

[0082] 在本发明之前,Cas蛋白质的核酸内切酶活性是已知的。然而,由于真核基因组的复杂性,并不知道Cas蛋白质的核酸内切酶活性是否会在真核细胞中起作用。此外,到现在为止,还没有开发在真核细胞或生物体中切割靶DNA的包含Cas蛋白质或Cas蛋白编码核酸和特异于靶DNA的向导RNA的组合物。

[0083] 相比ZFN和TALEN,基于Cas蛋白质的本发明RGEN组合物可以更容易地定制,这是因为:为产生新的基因组编辑核酸酶,可以仅替换合成的向导RNA成分。不涉及亚克隆步骤来产生定制RNA向导核酸内切酶。此外,与一对TALEN基因(~6kbp)相比,Cas基因相对小的尺寸(例如,Cas9为4.2kbp)在一些应用(如病毒介导的基因递送)中为该RNA向导核酸内切酶组合物提供了优点。另外,该RNA向导核酸内切酶没有脱靶效应,因此不引起不需要的突变、缺失、倒位和重复。这些特性使本发明的RNA向导核酸内切酶组合物成为在真核细胞和生物体中进行基因组工程化的一种可扩展的通用方便工具。另外,RGEN可以设计成靶向任何DNA序列,几乎任何单核苷酸多态性或小的插入/缺失(indel)均可以通过RGEN介导的RFLP进行分析。RGEN的特异性由RNA成分和Cas9蛋白质决定,其中所述RNA成分与长度不超过20个碱基对(bp)的靶DNA序列杂交,所述Cas9蛋白质识别原间隔区邻近基序(proto-spacer-adjacent motif (PAM))。通过替换RNA成分可以容易地重新编程RGEN。因此,RGEN提供了一个将简单可靠的RFLP分析用于各种序列变异的平台。

[0084] 靶DNA可以是内源性DNA或人工DNA,优选内源性DNA。

[0085] 如本文所用,术语“Cas蛋白”指CRISPR/CAS系统中的必要蛋白质成分,当与称为CRISPR RNA(crRNA)和反式激活crRNA(tracrRNA)的2个RNA复合时Cas蛋白形成活性核酸内切酶或切口酶。

[0086] Cas基因和蛋白质的信息可从美国国家生物技术信息中心(NCBI)的GenBank中获得,没有限制。

[0087] 编码Cas蛋白质的CRISPR相关(cas)基因通常与CRISPR重复序列-间隔区阵列相关。已经描述了多于四十个不同的Cas蛋白质家族。在这些蛋白质家族中,Cas1似乎遍布各种不同的CRISPR/CAS系统中。有三种类型CRISPR-Cas系统。其中,涉及Cas9蛋白质和crRNA和tracrRNA的II型CRISPR/Cas系统是代表性的,是众所周知的。cas基因和重复结构的特定组合被用来定义8种CRISPR亚型(Ecoli,Ypest,Nmeni,Dvulg,Tneap,Hmari,Aperrn和Mtube)。

[0088] Cas蛋白质可以连接到蛋白质转导结构域。蛋白质转导结构域可以是聚精氨酸或源自HIV的TAT蛋白,但不限于此。

[0089] 本发明组合物可包含蛋白质形式或编码Cas蛋白质的核酸形式的Cas 成分。

[0090] 在本发明中,Cas蛋白质可以是任何Cas蛋白质,只要当其向与向导RNA 复合时具有核酸内切酶或切口酶活性即可。

[0091] 优选地,Cas蛋白质是Cas9蛋白质或其变体。

[0092] Cas9蛋白质的变体可以是Cas9的突变体形式,其中催化性天冬氨酸残基改变为任何其它氨基酸。优选地,所述其它氨基酸可以是丙氨酸,但不限于此。

[0093] 此外,Cas蛋白质可以是生物体如链球菌属物种(*Streptococcus* sp.),优选化脓性链球菌(*Streptococcus pyogenes*)中分离的蛋白质、或重组蛋白质,但并不限于此。

[0094] 源自化脓性链球菌的Cas蛋白质可识别NGG三核苷酸。Cas蛋白质可包括SEQ ID NO:109的氨基酸序列,但不限于此。

[0095] 术语“重组”,当用于例如细胞、核酸、蛋白质或载体时,表示该细胞、核酸、蛋白质或载体已通过引入异源核酸或蛋白质、或改变天然核酸或蛋白质而被修饰,或该细胞源自此修饰的细胞。因此,例如,可以通过使用人密码子表重构Cas蛋白质编码序列而产生重组Cas蛋白质。

[0096] 就本发明而言,Cas蛋白质编码核酸可以是载体形式,如包含在启动子如CMV或CAG下的Cas编码序列的质粒。当Cas蛋白质是Cas9时,Cas9编码序列可源自链球菌属,优选源自化脓性链球菌。例如,Cas9编码核酸可以包含SEQ ID NO:1的核苷酸序列。此外,Cas9编码核酸可包含与SEQ ID NO:1的序列具有至少50%同源性的核苷酸序列,优选与SEQ ID NO:1具有至少60,70,80,90,95,97,98,或99%的同源性,但不限于此。Cas9编码核酸可以包含SEQ ID NOs.108、110,106或107的核苷酸序列。

[0097] 如本文所用,术语“向导RNA”指的是特异于靶DNA的RNA,其可以与Cas蛋白质形成复合体并把Cas蛋白质带至靶DNA。

[0098] 在本发明中,向导RNA可以由两个RNA,即CRISPR RNA (crRNA) 和反式激活crRNA (tracrRNA) 组成,或者向导RNA可以通过融合crRNA 和tracrRNA的必要部分而产生的单链RNA (sgRNA)。

[0099] 向导RNA可以是包含crRNA和tracrRNA的双重RNA (dualRNA)。

[0100] 如果向导RNA包含crRNA和tracrRNA的必要部分和互补于靶的部分,则任何向导RNA都可以在本发明中使用。

[0101] 所述crRNA可与靶DNA杂交。

[0102] RGEN可由Cas蛋白质和dualRNA (不变的tracrRNA和靶特异性 crRNA) 组成,或由Cas蛋白质和sgRNA (不变的tracrRNA和靶特异性 crRNA的必要部分的融合物) 组成,并且可以通过替换crRNA而容易地重新编程。

[0103] 向导RNA可以在单链向导RNA或dualRNA的crRNA的5'末端还包含一个或多个附加的核苷酸。

[0104] 优选,向导RNA可以在单链向导RNA或dualRNA的crRNA的5'末端还包含2个附加的鸟嘌呤核苷酸。

[0105] 向导RNA可以以编码该向导RNA的RNA或DNA的形式转移到细胞或生物体中。向导RNA可以是分离的RNA、并入病毒载体的RNA的形式、或者在载体中编码。优选地,载体可以是病毒载体、质粒载体、或农杆菌载体,但不限于此。

[0106] 编码向导RNA的DNA可以是包含编码向导RNA序列的载体。例如,可以通过用分离的向导RNA或包含编码向导RNA的序列和启动子的质粒 DNA转染细胞或生物体,将向导RNA转染到细胞或生物体。

[0107] 可替代地,可以使用病毒介导的基因递送,将向导RNA转移到细胞或生物体。

[0108] 当向导RNA以分离RNA的形式转染细胞或生物体时,可使用本领域中已知的任何体外转录系统通过体外转录来制备向导RNA。向导RNA优选以分离RNA的形式,而不是以包含向导RNA的编码序列的质粒的形式,转移到细胞。如本文所用,术语“分离RNA”可与“裸露的RNA”互换使用。因为不需要克隆步骤,这能节省成本和时间。然而,不排除使用质粒DNA或病毒介导的基因传递来转染向导RNA。

[0109] 由于向导RNA对靶的特异性和Cas蛋白质的核酸内切酶或切口酶活性,本发明的包含Cas蛋白质或Cas蛋白质编码核酸和向导RNA的RGEN 组合物可特异地切割靶DNA。

[0110] 如本文所用,术语“切割”是指核苷酸分子共价骨架的断裂。

[0111] 在本发明中,向导RNA可以制备为特异于任何待切割的靶。因此,本发明的RGEN组合物可以通过操作或基因分型向导RNA的靶特异性部分,切割任何靶DNA。

[0112] 向导RNA和Cas蛋白质可以以对子(pair)发挥作用。如本文所用,术语“配对Cas切口酶”可以指以对子发挥作用的向导RNA和Cas蛋白质。该对子(pair)包括两个向导RNA。向导RNA和Cas蛋白质可以以对子发挥作用,在不同DNA链上诱导两个缺口。这两个缺口可以分开至少100个bps,但不限于此。

[0113] 在实施例中,本发明人证实,配对Cas切口酶允许在人类细胞中实现靶向诱变和高达1-kbp的染色体片段的大缺失。重要的是,配对切口酶不在脱靶位点诱导indel,而其相应的核酸酶在脱靶位点处诱导突变。此外,不同于核酸酶,配对切口酶不促进与脱靶DNA切割相关的不需要的易位。原则上,配对切口酶使Cas9介导的诱变的特异性加倍,可以扩大RNA向导酶在需要精确基因组编辑的应用(如基因和细胞治疗)中的用途。

[0114] 在本发明中,组合物可用于体外对真核细胞或生物体的基因组作基因分型。

[0115] 在一个具体的实施方案中,向导RNA可以包含SEQ ID.No.1的核苷酸序列,其中的核苷酸位置3~22的部分为靶特异性部分,因此,该部分的序列可以根据靶而改变。

[0116] 如本文所使用的,没有限制地,真核细胞或生物体可以是酵母、真菌、原生动物、植物、高等植物、和昆虫、或两栖动物细胞、或哺乳动物细胞例如CHO、HeLa、HEK293和COS-1,例如,培养的细胞(体外)、移植细胞和原代细胞培养物(体外和离体)、和体内细胞,以及本领域中通常使用的哺乳动物细胞,包括人类细胞。

[0117] 在一个具体的实施方案中,发现Cas9蛋白质/单链向导RNA可以在体外和在哺乳动物细胞中产生位点特异性DNA双链断裂,其自发修复以高频率诱导靶向基因组突变。

[0118] 此外,发现,可通过向一细胞阶段的胚胎注射Cas9蛋白质/向导RNA 复合体或Cas9 mRNA/向导RNA诱导基因敲除小鼠,并可以通过Cas9/向导RNA系统生成种系可传递的突变。

[0119] 使用Cas蛋白质而不是编码Cas蛋白质的核酸诱导靶向诱变是有利的,这是因为不向生物体中引入外源DNA。因此,包括Cas蛋白质和向导RNA 的组合物可以用于开发治疗剂或增值农作物、家畜、家禽、鱼、宠物等。

[0120] 根据本发明的另一个方面,本发明提供用于在真核细胞或生物体中诱导定向诱变的组合物,其包括特异于靶DNA的向导RNA或编码该向导RNA的DNA和Cas蛋白质编码核酸或

Cas蛋白质。另外,本发明提供组合物用于在真核细胞或生物体中诱导靶向诱变的用途,所述组合物包含特异于靶DNA的向导RNA或编码该向导RNA的DNA和Cas蛋白质编码核酸或Cas蛋白质。

[0121] 向导RNA、Cas蛋白质编码核酸或Cas蛋白质在以上描述。

[0122] 根据本发明的另一个方面,本发明提供用于在真核细胞或生物体中切割靶DNA或诱导靶向诱变的试剂盒,其包括特异于靶DNA的向导RNA 或编码该向导RNA的DNA和Cas蛋白质编码核酸或Cas蛋白质。

[0123] 向导RNA,Cas蛋白质编码核酸或Cas蛋白质在以上描述。

[0124] 所述试剂盒可包括作为分开成分或作为一种组合物的向导RNA和Cas 蛋白质编码核酸或Cas蛋白质。

[0125] 本发明的试剂盒可以包含将向导RNA和Cas成分转移至细胞或生物体所必需的一些其它成分。例如,试剂盒可包括注射缓冲液如DEPC处理的注射缓冲液、以及分析靶DNA突变所必需的物质,但不限于此。

[0126] 根据另一个方面,本发明提供制备包含Cas蛋白质和向导RNA的真核细胞或生物体的方法,所述方法包括用Cas蛋白质编码核酸或Cas蛋白质和向导RNA或编码所述向导RNA的DNA共转染或顺序转染真核细胞或生物体的步骤。

[0127] 向导RNA,Cas蛋白质编码核酸或Cas蛋白质在以上描述。

[0128] 在本发明中,Cas蛋白质编码核酸或Cas蛋白质和向导RNA或编码该向导RNA的DNA,可以通过本领域中已知的各种方法转移到细胞中,如微注射、电穿孔、DEAE-葡聚糖处理、脂转染、纳米颗粒介导的转染、蛋白质转导结构域介导的转导、病毒介导的基因递送、和PEG介导的原生质体转染,等等,但不限于此。另外,Cas蛋白质编码核酸或Cas蛋白质和向导RNA可通过本领域中已知的各种施用基因或蛋白质的方法(如注射)而转移到生物体。Cas蛋白质编码核酸或Cas蛋白质可以以与向导RNA复合体的形式,或单独地,被转移到细胞中。融合至蛋白质转导结构域(如Tat)的Cas蛋白质也可以被有效地递送到细胞中。

[0129] 优选地,真核细胞或生物体被Cas9蛋白质和向导RNA共转染或顺序转染。

[0130] 顺序转染可通过如下进行:用Cas蛋白质编码核酸进行第一转染,接着用裸向导RNA进行第二转染。优选地,第二转染在3,6,12,18,24 小时之后,但并不限于此。

[0131] 根据另一个方面,本发明提供真核细胞或生物体,其包含特异于靶 DNA的向导RNA或编码该向导RNA的DNA和Cas蛋白质编码核酸或 Cas蛋白质。

[0132] 可以通过将组合物转移至细胞或生物体中制备真核细胞或生物体,所述组合物包含特异于靶DNA的向导RNA或编码该向导RNA的DNA和 Cas蛋白质编码核酸或Cas蛋白质。

[0133] 真核细胞可以是酵母、真菌、原生动物、高等植物、昆虫、或两栖动物细胞或哺乳动物细胞,例如CHO、HeLa、HEK293和COS-1,例如,培养的细胞(体外)、移植细胞和原代细胞培养物(在体外和离体)、和体内细胞,也可以是在本领域中通常使用的哺乳动物细胞,包括人类细胞,等等,没有限制。进一步地生物体可以是酵母、真菌、原生动物、植物、高等植物、昆虫、两栖动物、或哺乳动物。

[0134] 根据本发明的另一个方面,本发明提供一种用于在真核细胞或生物体中切割靶DNA或诱导靶向诱变的方法,所述方法包括用组合物处理含有靶DNA的细胞或生物体的步骤,所述组合物包含特异于靶DNA的向导 RNA或编码该向导RNA的DNA和Cas蛋白质编码核酸

或Cas蛋白质。

[0135] 用组合物处理细胞或生物体的步骤可通过将本发明的组合物转移到细胞或生物体中进行,所述组合物包含特异于靶DNA的向导RNA或编码该向导RNA的DNA和Cas蛋白质编码核酸或Cas蛋白。

[0136] 如上述,这种转移可以通过微注射、转染、电穿孔等等来进行。

[0137] 根据本发明的另一个方面,本发明提供包含通过本发明RGEN组合物编辑的基因组的胚胎,所述RGEN组合物包含特异于靶DNA的向导RNA 或编码该向导RNA的DNA和Cas蛋白质编码核酸或Cas蛋白质。

[0138] 任何胚胎均可以在本发明中使用,对于本发明,胚胎可以是小鼠胚胎。胚胎可通过如下产生:向4~7周雌性小鼠注射PMSG(孕母马血清促性腺激素)和hCG(人绒毛膜促性腺激素),可以将超排卵雌性小鼠与雄性小鼠交配,并可从输卵管收集受精的胚胎。

[0139] 引入胚胎的本发明RGEN组合物可通过Cas蛋白质作用来切割与向导 RNA互补的靶DNA,引起靶DNA中的突变。因此,引入了本发明RGEN 组合物的胚胎具有编辑的基因组。

[0140] 在一个具体实施方案中,已发现,本发明RGEN组合物可导致小鼠胚胎突变,所述突变可以被传递给后代。

[0141] 将RGEN组合物引入胚胎的方法可以是本领域已知的任何方法,如微注射、干细胞插入、逆转录病毒插入等等。优选,可使用微注射技术。

[0142] 根据另一个方面,本发明提供通过将胚胎转移至动物输卵管而获得的基因组修饰动物,其中所述胚胎包含由本发明RGEN组合物编辑的基因组。

[0143] 在本发明中,术语“基因组修饰动物”是指其基因组在胚胎阶段已被本发明RGEN组合物修饰的动物,动物的种类没有限制。

[0144] 基因组修饰的动物具有通过基于本发明RGEN组合物的靶向诱变引起的突变。所述突变可以是缺失、插入、易位、倒位中的任一种。突变的位点取决于RGEN组合物的向导RNA序列。

[0145] 具有基因突变的基因组修饰动物可用于确定该基因的功能。

[0146] 根据本发明的另一个方面,本发明提供了制备基因组修饰动物的方法,所述方法包括将本发明RGEN组合物(其包含特异于靶DNA的向导RNA 或编码该向导RNA的DNA和Cas蛋白质编码核酸或Cas蛋白质)引入动物胚胎的步骤;和将胚胎转移到假孕代母输卵管的步骤,以产生基因组修饰动物。

[0147] 引入本发明RGEN组合物的步骤可通过本领域中已知的任何方法来完成,例如微注射、干细胞插入、逆转录病毒插入等等。

[0148] 根据本发明的另一个方面,本发明提供从基因组修饰的原生质体再生的植物,所述原生质体通过用于产生含有RGEN组合物的真核细胞的方法制备。

[0149] 根据本发明的另一个方面,本发明提供用于在分离的生物样品中基因分型突变或变异的组合物,所述组合物含有特异于靶DNA序列的向导 RNA和Cas蛋白。另外,本发明提供用于在分离的生物样品中基因分型病原微生物的核酸序列的组合物,所述组合物包含特异于靶DNA序列的向导RNA和Cas蛋白质。

[0150] 向导RNA、Cas蛋白质编码核酸或Cas蛋白质在以上描述。

[0151] 本文所用术语“基因分型”指的是“限制性片段长度多态性(RFLP)分析”。

[0152] RFLP可用于1) 检测由工程化的核酸酶诱导的细胞或生物体中的 indel, 2) 对细胞或生物体中天然存在的突变或变异进行基因分型, 或3) 对感染的病原微生物(包括病毒或细菌等)的DNA进行基因分型。

[0153] 可通过工程化的核酸酶在细胞中诱导突变或变异。

[0154] 工程化的核酸酶可以是锌指核酸酶(ZFN)、转录激活因子样效应物核酸酶(TALEN)、或RGEN, 但不限于此。

[0155] 如本文所用术语“生物样品”包括进行分析的样品, 如组织、细胞、全血、SEMM、血浆、唾液、痰、脑脊液或尿液, 但并不限于此。

[0156] 突变或变异可以是天然存在的突变或变异。

[0157] 突变或变异由病原微生物引起。即, 突变或变异由于病原微生物的感染而发生, 当检测病原微生物时, 该生物样品被鉴定为受感染。

[0158] 病原微生物可以是病毒或细菌, 但不限于此。

[0159] 工程化核酸酶诱导的突变可通过各种方法检测, 其中包括错配敏感的 Surveyor 或T7内切核酸酶I (T7E1) 测定法、RFLP分析、荧光PCR、DNA 熔解分析和Sanger和深度测序(deep sequencing)。T7E1和Surveyor测定法广泛使用, 但往往低估突变频率, 因为这些测定法可以检测异源双链(由突变体和野生型序列杂交形成或由两个不同突变体序列杂交形成); 但它们无法检测由两个相同突变体序列杂交形成的同源双链。因此, 这些测定法不能从野生型细胞中区分纯合双等位基因突变体克隆, 也不能从杂合单等位基因突变体中区分杂合双等位基因突变体(图22)。此外, 核酸酶靶位点附近的序列多态性可以造成混淆的结果, 因为所述酶可切割由这些不同野生型等位基因杂交形成的异源双链。RFLP分析没有这些限制, 并因此是首选的方法。实际上, RFLP分析是最早用于检测工程化核酸酶介导的突变的方法之一。然而不幸的是, 其受到适当限制性位点可得性的限制。

[0160] 根据本发明的另一个方面, 本发明提供用于在分离的生物样品中基因分型突变或变异的试剂盒, 所述试剂盒包括用于在分离的生物样品中基因分型突变或变异的组合物。另外, 本发明提供用于在分离的生物样品中基因分型病原微生物中的核酸序列的试剂盒, 所述试剂盒包括特异于靶DNA 序列的向导RNA和Cas蛋白质。

[0161] 向导RNA, Cas蛋白质编码核酸或Cas蛋白质在以上描述。

[0162] 根据本发明的另一个方面, 本发明提供了在分离的生物样品中基因分型突变或变异的方法, 该方法使用组合物用于在分离的生物样品中基因分型突变或变异。另外, 本发明提供在分离的生物样品中基因分型病原微生物中的核酸序列的方法, 其包括特异于靶DNA 序列的向导RNA和Cas蛋白质。

[0163] 向导RNA, Cas蛋白质编码核酸或Cas蛋白质在以上描述。

[0164] 发明具体实施方式

[0165] 以下, 参照实施例对本发明进行更详细地描述。然而, 这些实施例仅用于举例说明的目的, 且本发明不意图受这些实施例的限制。

[0166] 实施例1: 基因组编辑测定

[0167] 1-1. Cas9蛋白质的DNA切割活性

[0168] 首先, 测试了在体外存在或不存在嵌合的向导RNA时源自化脓性链球菌的Cas9的DNA切割活性。

[0169] 为此,使用在大肠杆菌中表达并纯化的重组Cas9蛋白质切割预先消化的或环状质粒DNA,所述质粒DNA包含23个碱基对(bp)的人CCR5 靶序列。Cas9靶序列由互补于crRNA或嵌合向导RNA的20-bp DNA序列和Cas9本身识别的三核苷酸(5'-NGG-3')原间隔区邻近基序(PAM)组成(图1A)。

[0170] 具体地,源自化脓性链球菌菌株M1 GAS (NC\_002737.1)的Cas9编码序列(4104bp)使用人类密码子使用表重构,并使用寡核苷酸合成。首先,使用重叠的~35mer寡核苷酸和Phusion聚合酶(New England Biolabs) 组装1-kb DNA片段,并克隆到T-载体(SolGent)。通过重叠PCR,使用四个1-kbp的DNA片段,组装全长Cas9序列。所述Cas9编码DNA片段亚克隆到p3s,所述p3s来自pcDNA3.1 (Invitrogen)。在此载体中,包含 HA表位和核定位信号(NLS)的肽标签(NH<sub>2</sub>-GGSGPPKKRKYVYPYDV PDYA-COOH,SEQ ID NO:2)加至Cas9的C末端。通过蛋白印迹,使用抗HA抗体(Santa Cruz),确认HEK 293T细胞中Cas9蛋白质的表达和核定位。

[0171] 然后,将Cas9盒亚克隆到pET28-b(+),并转化到BL21 (DE3)。使用0.5mM IPTG在25℃下诱导Cas9表达4小时。使用Ni-NTA琼脂糖树脂(Qiagen)纯化在C末端含有His6标签的Cas9蛋白质,并在20mM HEPES (pH值7.5),150mM KCl,1mM DTT和10%甘油中透析(1)。纯化的Cas9(50nM)与超螺旋或预先消化的质粒DNA(300ng)及嵌合RNA(50nM)在20微升反应体积中在NEB缓冲液3中在37℃下孵育1小时。用0.8%琼脂糖凝胶电泳,分析消化的DNA。

[0172] 只有在合成的RNA存在时Cas9才在预期位置有效地切割质粒DNA,但不切割缺乏靶序列的对照质粒(图1B)。

[0173] 1-2.Cas9/向导RNA复合体在人类细胞中切割DNA

[0174] RFP-GFP报告分子用来研究Cas9/向导RNA复合体是否能在哺乳动物细胞中切割在RFP和GFP序列之间引入的靶序列。

[0175] 在该报告分子中,将GFP序列与RFP序列在读框外融合(2)。只有当靶序列被位点特异性核酸酶切割,通过双链断裂(DSB)的易错非同源末端连接(NHEJ)修复在靶序列周围造成移码小插入或缺失(indel),活性GFP才表达(图2)。

[0176] 在本研究中使用的RFP-GFP报告分子质粒如前所述构建(2)。合成对应于靶位点的寡核苷酸(表1)(Macrogen)并退火。将退火的寡核苷酸连接至用EcoRI和BamHI消化的报告分子载体。

[0177] 用Cas9编码质粒(0.8μg)和RFP-GFP报告分子质粒(0.2μg)在24孔板中使用Lipofectamine 2000(Invitrogen)共转染HEK 293T细胞。

[0178] 同时,如下制备体外转录的嵌合RNA。根据制造商的手册,使用MEGAscript T7试剂盒(Ambion),使用Runoff反应,体外转录RNA。通过退火两个互补的单链DNA或通过PCR扩增,生成用于RNA体外转录的模板(表1)。转录的RNA在8%变性尿素PAGE凝胶中分离。切出含RNA的凝胶切片并转移至探针洗脱缓冲液。回收RNA于无核酸酶的水中,然后用苯酚:氯仿提取,氯仿提取和乙醇沉淀。由光谱法定量纯化的RNA。

[0179] 在转染后第12小时,将体外转录制备的嵌合RNA(1μg)用Lipofectamine 2000转染。

[0180] 在转染后第3天,对转染的细胞进行流式细胞分析,对同时表达RFP和GFP的细胞进行计数。

[0181] 结果发现,只有当首先用Cas9质粒转染、然后12小时后用向导RNA转染细胞,才获

得了表达GFP的细胞(图2),表明RGEN可以在培养的人细胞中识别并切割靶DNA序列。由此,表达GFP的细胞通过Cas9质粒和向导RNA的顺序转染获得而不是通过共转染获得。

[0182] 表1

基因		序列 (5' 至 3')	SEQ ID NO.
用于构建报告分子质粒的寡核苷酸			
CCR5	F	AATTCATGACATCAATTATTATACATCGGAGGAG	3
	R	GATCCTCCTCCGATGTATAATAATTGATGTCATG	4
用于T7E1测定法中的引物			
CCR5	F1	CTCCATGGTGCTATAGAGCA	5
	F2	GAGCCAAGCTCTCCATCTAGT	6

	R	GCCCTGTCAAGAGTTGACAC	7
C4BPB	F1	TATTTGGCTGGTTGAAAGGG	8
	R1	AAAGTCATGAAATAAACACACCCA	9
	F2	CTGCATTGATATGGTAGTACCATG	10
	R2	GCTGTTCATTGCAATGGAATG	11
用于扩增脱靶位点的引物			
ADCY5	F1	GCTCCACCTTAGTGCTCTG	12
	R1	GGTGGCAGGAACCTGTATGT	13
	F2	GTCATTGGCCAGAGATGTGGA	14
	R2	GTCCCATGACAGGCGTGTAT	15
KCNJ6	F	GCCTGGCCAAGTTTCAGTTA	16
[0184]	R1	TGGAGCCATTGGTTTGCATC	17
	R2	CCAGAACTAAGCCGTTTCTGAC	18
CNTNAP2	F1	ATCACCGACAACCAGTTTCC	19
	F2	TGCAGTGCAGACTCTTTCCA	20
	R	AAGGACACAGGGCAACTGAA	21
N/A Chr. 5	F1	TGTGGAACGAGTGGTGACAG	22
	R1	GCTGGATTAGGAGGCAGGATTC	23
	F2	GTGCTGAGAACGCTTCATAGAG	24
	R2	GGACCAAACCACATTCTTCTCAC	25
用于检测染色体缺失的引物			
缺失	F	CCACATCTCGTTCTCGGTTT	26
	R	TCACAAGCCCACAGATATTT	27

[0185] 1-3. 通过RGEN靶向破坏哺乳动物细胞中的内源基因

[0186] 为了测试RGEN是否可用于靶向破坏哺乳动物细胞中的内源基因,使用T7内切核酸酶I (T7E1) (一种错配敏感的内切核酸酶,其特异性识别并切割由野生型和突变体DNA序列杂交形成的异源双链) 分析了分离自转染细胞的基因组DNA (3)。

[0187] 为了使用RGEN在哺乳动物细胞中引入DSB,根据生产商的方案,使用4D-Nucleofector, SF细胞系4D-Nucleofector X试剂盒, Program FF-120 (Lonza), 用20 $\mu$ g Cas9编码质粒转染 $2 \times 10^6$  K562细胞。对于本实验, K562 (ATCC, CCL-243) 细胞在含有10% FBS和青霉素/链霉素混合物 (分别为100U/ml和100 $\mu$ g/ml) 的RPMI-1640上培养。

[0188] 24小时后,将10-40 $\mu$ g体外转录的嵌合RNA核感染至 $1 \times 10^6$ K562细胞。如在实施例1-2中描述的制备体外转录的嵌合RNA。

[0189] RNA转染后两天收集细胞,分离基因组DNA。包括靶位点的区域使用表1所述的引物进行PCR扩增。如前所述(3),对扩增子进行T7E1 测定。为了测序分析,纯化对应于基因组修饰的PCR产物并使用T-平端 PCR克隆试剂盒(SolGent)克隆至T-平端载体。克隆产物用M13引物测序。

[0190] 结果发现,只有用Cas9编码质粒、然后用向导RNA顺序转染细胞时,才诱导了突变(图3)。从相对DNA条带强度估计的突变频率(图3A中 indel(%))是RNA剂量依赖的,范围从1.3%到5.1%。PCR扩增子的 DNA测序分析证实了在内源性位点处诱导了RGEN介导的突变。在靶位点观察到易错NHEJ特征性的indel和微同源性。通过直接测序检测的突变频率为7.3%(=7突变克隆/96个克隆),与使用锌指核酸酶(ZFN)或转录激活因子样效应物核酸酶(TALENS)所获得的突变频率相媲美。

[0191] 需要顺序转染Cas9质粒和向导RNA以诱导细胞中的突变。但是,当质粒编码向导RNA时,顺序转染则不是必要的,细胞可用Cas9质粒和向导RNA编码质粒共转染。

[0192] 已成功地开发了ZFN和TALEN来破坏人CCR5基因(3-6),该基因编码G蛋白偶联趋化因子受体——HIV感染的一个主要的共受体。目前 CCR5特异性ZFN在美国正处于艾滋病治疗的临床研究阶段(7)。然而,这些ZFN和TALEN具有脱靶效应,在序列与靶序列同源的位点上诱导局部突变(6,8-10)并诱导基因组重排(其起因于对靶和脱靶位点上诱导的两个并行DSB的修复)(11-12)。与这些CCR5特异性工程化核酸酶相关的最显著的脱靶位点位于CCR2基因座中,CCR2是CCR5的密切同源物,位于CCR5上游15kbp。为避免CCR2基因上的脱靶突变和不需要的CCR5靶和CCR2脱靶位点之间15-kbp染色体片段的缺失、倒位,和复制,本发明人有意选择我们的CCR5特异性RGEN的靶位点来识别 CCR5序列中的区域,该区域与CCR2序列没有明显的同源性。

[0193] 本发明人研究了CCR5特异性RGEN是否有脱靶效应。为此,通过鉴定与预定23-bp靶序列最同源的位点,搜索了人类基因组中潜在的脱靶位点。正如预期的那样,在CCR2基因中没有发现这样的位点。但发现了四个位点,其中每一个均携带与靶位点的3个碱基错配(图4A)。T7E1 分析表明,在这些位点中没有检测到突变(检测灵敏度, $\sim 0.5\%$ ),这表明RGEN极高的特异性(图4B)。此外,用PCR检测了细胞中染色体缺失的诱导,所述细胞分开转染了编码ZFN的质粒和CCR5特异性RGEN。ZFN诱导缺失,而RGEN不诱导缺失(图4C)。

[0194] 接着,通过用设计以靶向人C4BPB基因的新合成RNA代替CCR5特异性向导RNA来重新编程RGEN,所述人C4BPB基因编码C4b结合蛋白(一种转录因子)的 $\beta$ 链。该RGEN在K562细胞中以高频率诱导了染色体靶位点上的突变(图3B)。由T7E1测定法和直接测序测得的突变频率分别为14%和8.3%(=4突变克隆/48克隆)。在四个突变序列中,两个克隆精确地在切割位点含有单碱基或两碱基插入,该模式也在CCR5靶位点上观察到。这些结果表明,RGEN在细胞中于预期位置切割染色体靶 DNA。

[0195] 实施例2:蛋白质性RGEN介导的基因组编辑

[0196] RGEN可以以多种不同形式递送到细胞。RGEN由Cas9蛋白、crRNA 和tracrRNA组成。两个RNA可以融合形成单链向导RNA(sgRNA)。在启动子如CMV或CAG下编码Cas9的质粒可以被转染到细胞中。crRNA, tracrRNA,或sgRNA也可以使用编码这些RNA的质粒在细胞中表

达。但是,使用质粒经常导致质粒全部或部分整合在宿主基因组中。掺入质粒 DNA中的细菌序列可能在体内引起不希望的免疫应答。用于细胞治疗的质粒转染的细胞或源自DNA转染细胞的动物和植物在大多数发达国家在市场批准之前必须经过一个昂贵且冗长的管理程序。此外,质粒DNA能在转染后数天存在于细胞中,从而加重RGEN的脱靶效应。

[0197] 这里,我们使用与体外转录的向导RNA复合的重组Cas9蛋白质,以诱导人细胞内源基因的靶向破坏。融合六个组氨酸标签的重组Cas9蛋白质在大肠杆菌中表达,并使用标准镍离子亲和层析和凝胶过滤从大肠杆菌纯化。纯化的重组Cas9蛋白在存储缓冲液(20mM HEPES pH 7.5,150mM KCl,1mM DTT和10%甘油)中浓缩。Cas9蛋白质/sgRNA复合体通过核转染直接导入K562细胞:使用4D-Nucleofector,SF细胞系4D-Nucleofector X试剂盒,Program FF-120(Lonza),根据制造商的方案,使用100 $\mu$ l溶液中与100ug(29 $\mu$ M)体外转录的sgRNA(或40ug crRNA和80ug tracrRNA)混合的22.5-225(1.4-14 $\mu$ M)Cas9蛋白质,转染 $1 \times 10^6$ 个K562细胞。核转染(nucleofection)后,将细胞置于6孔板的生长培养基中温育48小时。当使用1/5按比例下调的方案转染 $2 \times 10^5$ 个K562细胞时,使用4.5-45  $\mu$ g Cas9蛋白质和6-60ug体外转录的sgRNA(或8 $\mu$ g crRNA和16 $\mu$ g tracrRNA)的混合物,在20 $\mu$ l溶液中进行核转染。然后将核转染的细胞置于48孔板的生长培养基中。48小时后,收集细胞并分离基因组DNA。PCR扩增跨靶位点的基因组DNA区域,并进行T7E1测定。

[0198] 如图10所示,Cas9蛋白质/sgRNA复合体在CCR5基因座上诱导了靶向突变,诱变频率以sgRNA或Cas9蛋白质剂量依赖性方式从4.8至38%,该频率与Cas9质粒转染所获得的频率相当(45%)。Cas9蛋白质 /crRNA/tracrRNA复合体能够以9.4%频率诱导突变。Cas9蛋白质单独不能诱导突变。当使用1/5按比例下调的Cas9蛋白质和sgRNA剂量转染 $2 \times 10^5$ 个细胞时,CCR5基因座上的突变频率以剂量依赖性方式为2.7至57%,大于共转染Cas9质粒和sgRNA质粒获得的频率(32%)。

[0199] 我们还测试了靶向ABCC11基因的Cas9蛋白质/sgRNA复合体,发现该复合体以35%的频率诱导Indel,这表明该方法的普遍适用性。

[0200] [表2]向导RNA序列

靶	RNA 类型	RNA序列(5'至3')	长度	SEQ ID NO
[0201] CCR5	sgRNA	GGUGACAUCAAUUUAUACAUGUUU UAGAGCUAGAAAUAGCAAGUUAAAAU AAGGCUAGUCCGUUAUCAACUUGAAAA AGUGGCACCGAGUCGGUGCUUUUUUU	104bp	28
	crRNA	GGUGACAUCAAUUUAUACAUGUUU UAGAGCUAUGCUGUUUUG	44bp	29
	tracrRNA	GGAACCAUUCAAAACAGCAUAGCAAGU UAAAAUAAGGCUAGUCCGUUAUCAACU UGAAAAAGUGGCACCGAGUCGGUGCUU UUUUU	86bp	30

[0202] 实施例3:小鼠中RNA指导的基因组编辑

[0203] 为了检验在原核 (PN) 阶段小鼠胚胎中RGEN的基因靶向潜力,使用了叉头框N1 (Foxn1) 基因 (其对胸腺发育和角质细胞分化是重要的 (Nehls 等,1996)) 和蛋白激酶的DNA激活的催化多肽 (Prkdc) 基因 (其编码对 DNA DSB修复和重组重要的酶 (Taccioli等, 1998))。

[0204] 为了评估Foxn1-RGEN的基因组编辑活性,我们将Cas9 mRNA (10 ng/ $\mu$ l溶液) 与各种剂量的sgRNA注射到PN-阶段小鼠胚胎的细胞质中 (图 5a), 并使用从体外培养的胚胎获得的基因组DNA进行了T7内切核酸酶 (T7E1) 测定 (Kim等人,2009年) (图6a)。

[0205] 可替代地,我们直接将RGEN以重组Cas9蛋白质 (0.3至30ng/ $\mu$ l) 与2倍摩尔过量的Foxn1特异性sgRNA (0.14至14ng/ $\mu$ l) 的复合物形式注射到一细胞小鼠胚胎的细胞质或原核中,并使用体外培养的胚胎分析了 Foxn1基因中的突变 (图7)。

[0206] 具体地说,分别使用mMESSAGE mMACHINE T7 Ultra试剂盒 (Ambion) 和MEGAscript T7试剂盒 (Ambion),根据制造商的说明,从线性DNA模板体外合成了Cas9 mRNA和sgRNA,并用合适量的焦碳酸二乙酯 (DEPC, Sigma) 处理的注射缓冲液 (0.25mM EDTA, 10mM Tris, pH 7.4) 进行稀释。使用表3所列的寡核苷酸产生用于sgRNA合成的模板。重组Cas9蛋白质从ToolGen, Inc. 获得。

[0207] [表3]

[0208] 表3

RNA 名称	方向	序列(5'至 3')	SEQ ID NO
Foxn1 #1 sgRNA	F	<u>GAAATTAATACGACTCACTATAGGCA</u> GTCTGACGTCACTTCCGTTTTAGAGC TAGAAATAGCAAGTTAAAATAAGGCT AGTCCG	31
Foxn1 #2 sgRNA	F	<u>GAAATTAATACGACTCACTATAGGACT</u> TCCAGGCTCCACCCGACGTTTTAGAGC TAGAAATAGCAAGTTAAAATAAGGCT AGTCCG	32
Foxn1 #3 sgRNA	F	<u>GAAATTAATACGACTCACTATAGGCCA</u> GGCTCCACCCGACTGGAGTTTTAGAGC TAGAAATAGCAAGTTAAAATAAGGCT AGTCCG	33
Foxn1 #4 sgRNA	F	<u>GAAATTAATACGACTCACTATAGGACT</u> GGAGGGCGAACCCCAAGTTTTAGAGC TAGAAATAGCAAGTTAAAATAAGGCT AGTCCG	34
Foxn1 #5 sgRNA	F	<u>GAAATTAATACGACTCACTATAGGACC</u> CCAAGGGGACCTCATGCGTTTTAGAGC TAGAAATAGCAAGTTAAAATAAGGCT AGTCCG	35
Prkdc #1 sgRNA	F	<u>GAAATTAATACGACTCACTATAGGTTA</u> <u>GTTTTTCCAGAGACTTGTTTTAGAGCT</u> <u>AGAAATAGCAAGTTAAAATAAGGCTA</u> <u>GTCCG</u>	36
Prkdc #2 sgRNA	F	<u>GAAATTAATACGACTCACTATAGGTTG</u> <u>GTTTGCTTGTTTATCGTTTTAGAGCT</u> <u>AGAAATAGCAAGTTAAAATAAGGCTA</u> <u>GTCCG</u>	37
Prkdc #3 sgRNA	F	<u>GAAATTAATACGACTCACTATAGGCAC</u> AAGCAAACCAAGTCTCGTTTTAGAGCT AGAAATAGCAAGTTAAAATAAGGCTA GTCCG	38

[0209]

[0210]	Prkdc #4 sgRNA	F	<u>GAAATTAATACGACTCACTATAGGCCT</u> CAATGCTAAGCGACTTCGTTTTAGAGCT AGAAATAGCAAGTTAAAATAAGGCTA GTCCG	39
--------	-------------------	---	---	----

[0211] 所有动物实验均按照韩国食品和药物管理局 (KFDA) 的指导方针进行。实验方案被延世大学实验动物研究中心的动物护理和使用委员会 (IACUC) 审查并批准 (许可证编号: 2013-0099)。所有小鼠维持在延世实验动物研究中心的无特定病原体的设施中。FVB/NTac (Taconic) 和ICR 小鼠品系分别被用作胚胎供体和代母。雌性FVB/NTac小鼠 (7-8周龄) 通过以48小时间隔腹膜内注射5IU孕马血清促性腺激素 (PMSG, Sigma) 和5IU人绒毛膜促性腺激素 (hCG, Sigma) 而超排卵。超排卵雌性小鼠与FVB/Ntac雄性小鼠种畜交配, 从输卵管收集受精胚胎。

[0212] 利用压电驱动的微操作机 (Prime Tech) 将M2培养基 (Sigma) 中的 Cas9 mRNA和sgRNA注射到具有清楚辨认的原核的受精卵的细胞质中。

[0213] 在注射重组Cas9蛋白质的情况下, 所述重组Cas9蛋白质:Foxn1-sgRNA复合体用DEPC处理的注射缓冲液 (0.25mM EDTA, 10 mM Tris, pH 7.4) 稀释, 并用TransferMan NK2微操作机和FemtoJet微注射器 (Eppendorf) 注射到雄性原核中。

[0214] 将操作的胚胎移入假孕代母的输卵管中以产生活的动物, 或在体外培养用于进一步分析。

[0215] 为了筛选具有RGEN诱发的突变的F0小鼠和体外培养的小鼠胚胎, 如前所述 (Cho等, 2013), 使用来自尾活检物的基因组DNA样品和全胚胎裂解物进行T7E1测定法。

[0216] 简言之, PCR扩增包含RGEN靶位点的基因组区域、解链、并再退火以形成异源双链DNA, 用T7内切核酸酶1 (New England Biolabs) 处理, 然后通过琼脂糖凝胶电泳分析。通过bowtie 0.12.9搜索鉴定了潜在的脱靶位点, 这些位点也类似地通过T7E1测定法监测。在这些测定中使用的引物对列于表4和5。

[0217] [表4]

[0218] 用于T7E1测定中的引物

[0219]

基因	方向	序列(5'至3')	SEQ ID NO
Foxn1	F1	GTCTGTCTATCATCTCTTCCCTTCTCTCC	40
	F2	TCCCTAATCCGATGGCTAGCTCCAG	41
	R1	ACGAGCAGCTGAAGTTAGCATGC	42
	R2	CTACTCAATGCTCTTAGAGCTACCAGGCTTG C	43
Prkdc	F	GACTGTTGTGGGGAGGGCCG	44
	F2	GGGAGGGCCGAAAGTCTTATTTG	45
	R1	CCTGAAGACTGAAGTTGGCAGAAGTGAG	46
	R2	CTTTAGGGCTTCTTCTCTACAATCACG	47

[0220] [表5]

[0221] 用于扩增脱靶位点的引物

[0222]

基因	符号	方向	序列(5'至3')	SEQ ID NO
Foxn1	off 1	F	CTCGGTGTGTAGCCCTGAC	48
		R	AGACTGGCCTGGAACCTCACA G	49
	off 2	F	CACTAAAGCCTGTCAGGAAG CCG	50
		R	CTGTGGAGAGCACACAGCAG C	51
	off 3	F	GCTGCGACCTGAGACCATG	52
		R	CTTCAATGGCTTCCTGCTTAG GCTAC	53
	off 4	F	GGTTCAGATGAGGCCATCCTT TC	54
		R	CCTGATCTGCAGGCTTAACCC TTG	55

[0223]

Prkdc	off 1	F	CTCACCTGCACATCACATGTG G	56
		R	GGCATCCACCCTATGGGGTC	57
	off 2	F	GCCTTGACCTAGAGCTTAAAG AGCC	58
		R	GGTCTTGTTAGCAGGAAGGAC ACTG	59
	off 3	F	AAAACCTGCTTGATGGGATA TGTGGG	60
		R	CTCTCACTGGTTATCTGTGCT CCTTC	61
	off 4	F	GGATCAATAGGTGGTGGGGG ATG	62
		R	GTGAATGACACAATGTGACA GCTTCAG	63
	off 5	F	CACAAGACAGACCTCTCAAC ATTCAGTC	64
		R	GTGCATGCATATAATCCATTC TGATTGCTCTC	65
	off 6	F1	GGGAGGCAGAGGCAGGT	66
		F2	GGATCTCTGTGAGTTTGAGGC CA	67
		R1	GCTCCAGAACTCACTCTTAGG CTC	68

[0224] T7E1测定法鉴定的突变体建立者小鼠通过fPCR进一步分析。基因组 DNA的合适区域如前所述 (Sung等, 2013) 进行测序。对于F1后代的常规PCR基因分型, 以下引物对用于野生型和突变体等位基因:

[0225] 5'-CTACTCCCTCCGAGTCTGA-3' (SEQ ID NO:69) 和

[0226] 5'-CCAGGCCTAGGTTCCAGGTA-3' (SEQ ID NO:70) 用于Foxn1基因,

[0227] 5'-CCCCAGCATTGCAGATTTCC-3' (SEQ ID NO:71) 和

[0228] 5'-AGGGCTTCTTCTCTACAATCACG-3' (SEQ ID NO:72) 用于Prkdc基因。

[0229] 在注射Cas9 mRNA的情况下, 突变体分数 (突变体胚胎数/总胚胎数) 是剂量依赖性的, 范围从33% (1ng/ $\mu$ l sgRNA) 至91% (100ng/ $\mu$ l) (图6b)。序列分析证实了Foxn1基因中的突变; 大多数突变是小缺失 (图6c), 让人联想起ZFN和TALENs诱导的那些 (Kim等, 2013)。

[0230] 在注射Cas9蛋白质的情况下, 这些注射剂量和方法极小地影响体外小鼠胚胎的生存和发育: 在两个实验中70%以上注射RGEN的胚胎正常孵化。再次, 用Cas9蛋白质注射所得突变体的分数是剂量依赖性的, 通过原核注射在最高剂量下达到多达88%, 通过胞质内注射达到多达71% (图7a和 7b)。与Cas9 mRNA加sgRNA诱导的突变模式类似 (图6c), Cas9蛋白质-sgRNA复合体诱导的突变大多是小缺失 (图7c)。这些结果清楚地表明, RGEN在小鼠胚胎中具有高的基因靶向活性。

[0231] 受到RGEN诱发的高突变频率和低毒性的鼓舞,我们将小鼠胚胎移至假孕代母的输卵管中产生了活的动物。

[0232] 值得注意的是,出生率非常高,从58%至73%,并且没有受到增加的 Foxn1-sgRNA剂量的影响(表6)。

[0233] [表6]

[0234] 在FVB/NTac小鼠中RGEN介导的基因靶向

靶基因	Cas9 mRNA + sgRNA (ng/ $\mu$ l)	注射的胚胎	转移的胚胎 (%)	总新生小鼠 (%)	活的新生小鼠* (%)	建立者小鼠+ (%)
Foxn1	10 + 1	76	62 (82)	45 (73)	31 (50)	12 (39)
	10 + 10	104	90 (87)	52 (58)	58 (64)	33 (57)
	10 + 100	100	90 (90)	62 (69)	58 (64)	54 (93)
	总共	280	242 (86)	159 (66)	147 (61)	99 (67)
Prkdc	50 + 50	73	58 (79)	35 (60)	33 (57)	11 (33)
	50 + 100	79	59 (75)	22 (37)	21 (36)	7 (33)
	50 + 250	94	73 (78)	37 (51)	37 (51)	21 (57)
	总共	246	190 (77)	94 (49)	91 (48)	39 (43)

[0236] 在147只新生小鼠中,我们获得99只突变建立者小鼠。与在培养的胚胎中观察到的结果一致(图6c),突变体分数与Foxn1-sgRNA剂量成比例,最长达93%(100ng/ $\mu$ l Foxn1-sgRNA)(表6和7,图5b)。

[0237] [表7]

[0238] 从T7E1阳性突变体建立者小鼠子集鉴定的Foxn1突变体等位基因的DNA 序列

[0239]

ACTTCCAGGCTCCACCCGACTGGAGGGCGAACC CCAAGGGGACCTCATGCAGG	del+ins	#	建立者 小鼠
ACTTCCAGGC-----AACCCCAAGGGGAC CTCATGCAGG	Δ19	1	20
ACTTCCAGGC-----GAACCCCAAGGGGA CCTCATGCAGG	Δ18	1	115
ACTTCCAGGCTCC-----	Δ60	1	19
ACTTCCAGGCTCC-----	Δ44	1	108
ACTTCCAGGCTCC-----CAAGGGGACC TCATGCAGG	Δ21	1	64
ACTTCCAGGCTCC-----TTAGGAGGCGAACCC CAAGGGGACCTCA	Δ12+6	1	126
ACTTCCAGGCTCCACC-----TCAT GCAGG	Δ28	1	5
ACTTCCAGGCTCCACCC-----CCAAGG GACCTCATG	Δ21+4	1	61
ACTTCCAGGCTCCACCC-----AAGGGGAC CTCATGCAGG	Δ18	2	95, 29
ACTTCCAGGCTCCACCC-----CAAGGGGA CCTCATGCAGG	Δ17	7	12, 14, 27, 66, 108, 114, 126
ACTTCCAGGCTCCACCC-----ACCCAAGGG GACCTCATGCAG	Δ15+1	1	32
ACTTCCAGGCTCCACCC-----CACCCAAGG GGACCTCATGCA	Δ15+2	1	124
ACTTCCAGGCTCCACCC-----ACCCCAAGGG GACCTCATGCAGG	Δ13	1	32
ACTTCCAGGCTCCACCC-----GGCGAACCCCAAG GGGACCTCATGCAGG	Δ8	1	110
ACTTCCAGGCTCCACCCT-----GGGGAC CTCATGCAGG	Δ20+1	1	29

[0240]

ACTTCCAGGCTCCACCCG-----AACCCCAAGG GGACCTCATGCAGG	$\Delta 11$	1	111
ACTTCCAGGCTCCACCCGA-----ACCT CATGCAGG	$\Delta 22$	1	79
ACTTCCAGGCTCCACCCGA-----GGGGAC CTCATGCAGG	$\Delta 18$	2	13, 127
ACTTCCAGGCTCCACCCCA-----AGGGGA CCTCATGCAGG	$\Delta 17$	1	24
ACTTCCAGGCTCCACCCGA-----ACCCCAAGG GGACCTCATGCAGG	$\Delta 11$	5	14, 53, 58, 69, 124
ACTTCCAGGCTCCACCCGA-----GACCCCAAGG GGACCTCATGCAGG	$\Delta 10$	1	14
ACTTCCAGGCTCCACCCGA----GGGCGAACCCCA AGGGGACCTCATGCAGG	$\Delta 5$	3	53, 79, 115
ACTTCCAGGCTCCACCCGAC-----CTC ATGCAGG	$\Delta 23$	1	108
ACTTCCAGGCTCCACCCGAC-----CCCCAAGG GGACCTCATGCAGG	$\Delta 11$	1	3
ACTTCCAGGCTCCACCCGAC-----GAAGGGCC CCAAGGGGACCTCA	$\Delta 11+6$	1	66
ACTTCCAGGCTCCACCCGAC-----GAACCCCAAG GGGACCTCATGCAGG	$\Delta 8$	2	3, 66
ACTTCCAGGCTCCACCCGAC----GGCGAACCCCA AGGGGACCTCATGCAGG	$\Delta 5$	1	27
ACTTCCAGGCTCCACCCGAC--GTGCTTGAGGGCG AACCCCAAGGGGACCTCA	$\Delta 2+6$	2	5
ACTTCCAGGCTCCACCCGACT-----CACTATCTTC TGGGCTCCTCCATGTC	$\Delta 6+25$	2	21, 114
ACTTCCAGGCTCCACCCGACT----TGGCGAACCCC AAGGGGACCTCATGCAG	$\Delta 4+1$	1	53
ACTTCCAGGCTCCACCCGACT--TGCAGGGCGAAC CCCAAGGGGACCTCATGC	$\Delta 2+3$	1	126
ACTTCCAGGCTCCACCCGACTTGGAGGGCGAAC CCCAAGGGGACCTCATGCAG	+1	15	3, 5, 12, 19, 29, 55, 56,

			61, 66, 68, 81, 108, 111, 124, 127	
[0241]	ACTTCCAGGCTCCACCCGACTTTGGAGGGCGAAC CCCAAGGGGACCTCATGCA	+2	2	79, 120
	ACTTCCAGGCTCCACCCGACTGTTGGAGGGCGA ACCCCAAGGGGACCTCATGC	+3	1	55
	ACTTCCAGGCTCCACCCGACTGGAG(+455)GGCG AACCCCAAGGGGACCTCC	+455	1	13

[0242] 为了产生Prkdc靶向小鼠,我们采用了浓度高5倍的Cas9 mRNA (50 ng/ $\mu$ l) 和递增剂量的Prkdc-sgRNA (50、100和250ng/ $\mu$ l)。再次,出生率是非常高的,范围从51%至60%,足以产生足够数量的新生小鼠用于分析(表6)。在最大剂量的Prkdc-sgRNA时突变体分数为57% (37只新生小鼠中21只突变体建立者小鼠)。由RGEN获得的这些出生率比我们以前研究报道的由TALEN获得的出生率高约2至10倍(Sung等,2013)。这些结果表明,RGEN是有效的基因靶向试剂,具有极小的毒性。

[0243] 为了测试突变体等位基因的种系传递,我们使Foxn1突变体建立者小鼠#108(具有四个不同等位基因的嵌合体(图5c和表8))与野生型小鼠杂交,并监测F1后代的基因型。

[0244] [表8]

[0245] Foxn1突变体小鼠的基因型

建立者编号	sgRNA (ng/ml)	基因分型总结	检测到的等位基因
58*	1	未确定	$\Delta 11$
19	100	双等位基因	$\Delta 60/+1$
20	100	双等位基因	$\Delta 67/\Delta 19$
13	100	双等位基因	$\Delta 18/+455$
32	10	双等位基因(杂合子)	$\Delta 13/\Delta 15+1$
115	10	双等位基因(杂合子)	$\Delta 18/\Delta 5$
111	10	双等位基因(杂合子)	$\Delta 11/+1$
110	10	双等位基因(纯合子)	$\Delta 8/\Delta 8$
120	10	双等位基因(纯合子)	$+2/+2$
81	100	杂合子	$+1/WT$
69	100	纯合子	$\Delta 11/\Delta 11$
55	1	嵌合体	$\Delta 18/\Delta 1/+1/+3$
56	1	嵌合体	$\Delta 127/\Delta 41/\Delta 2/+1$
[0246] 127	1	嵌合体	$\Delta 18/+1/WT$
53	1	嵌合体	$\Delta 11/\Delta 5/\Delta 4+1/WT$
27	10	嵌合体	$\Delta 17/\Delta 5/WT$
29	10	嵌合体	$\Delta 18/\Delta 20+1/+1$
95	10	嵌合体	$\Delta 18/\Delta 14/\Delta 8/\Delta 4$
108	10	嵌合体	$+1/\Delta 17/\Delta 23/\Delta 44$
114	10	嵌合体	$\Delta 17/\Delta 8/\Delta 6+25$
124	10	嵌合体	$\Delta 11/\Delta 15+2/+1$
126	10	嵌合体	$\Delta 17/\Delta 2+3/\Delta 12+6$
12	100	嵌合体	$\Delta 30/\Delta 28/\Delta 17/+1$
5	100	嵌合体	$\Delta 28/\Delta 11/\Delta 2+6/+1$
14	100	嵌合体	$\Delta 17/\Delta 11/\Delta 10$
21	100	嵌合体	$\Delta 127/\Delta 41/\Delta 2/\Delta 6+25$
24	100	嵌合体	$\Delta 17/+1/WT$
64	100	嵌合体	$\Delta 31/\Delta 21/+1/WT$

[0247]	68	100	嵌合体	<u><math>\Delta 17/\Delta 11/+1/WT</math></u>
	79	100	嵌合体	<u><math>\Delta 22/\Delta 5/+2/WT</math></u>
	61	100	嵌合体	<u><math>\Delta 21+4/\Delta 6/+1/+9</math></u>
	66**	100	嵌合体	<u><math>\Delta 17/\Delta 8/\Delta 11+6/+1/WT</math></u>
	3	100	嵌合体	<u><math>\Delta 11/\Delta 8/+1</math></u>

[0248] 下划线的等位基因进行了测序。

[0249] 红色的等位基因通过测序检测,但没有进行fPCR。

[0250] \*只测序了一个克隆。

[0251] \*\*没有通过fPCR检测。

[0252] 如预期的,所有的后代都是具有野生型等位基因和一个突变体等位基因的杂合突变体(图5d)。我们还证实了在独立的Foxn1(图8)和Prkdc(图9)建立者小鼠中的种系传递。根据我们所知,这些结果提供了RGEN诱导的突变体等位基因在动物中稳定地传递到F1后代的第一证据。

[0253] 实施例4:在植物中RNA引导的基因组编辑

[0254] 4-1. Cas9蛋白质的生产

[0255] 源自化脓性链球菌菌株M1 GAS (NC\_002737.1)的Cas9编码序列(4104bps)克隆到pET28-b(+)质粒。在蛋白质N末端包括核靶向序列(NLS),以确保将蛋白质定位至细胞核。含有Cas9 ORF的pET28-b(+)质粒转化到BL21(DE3)中。然后在18°C用0.2mM IPTG诱导Cas9进行16小时,并使用Ni-NTA琼脂糖珠(Qiagen)按照制造商的说明纯化。使用Ultrasel 100K(Millipore)浓缩纯化的Cas9蛋白质。

[0256] 4-2. 向导RNA的生产

[0257] 筛选编码BRI1的拟南芥基因的基因组序列,检查外显子中NGG基序的存在,所述基序,即所谓的原间隔区邻近基序(PAM),是Cas9靶向所必需的。为了破坏拟南芥中的BRI1基因,我们在包含NGG基序的外显子中鉴定了2个RGEN靶位点。用模板DNA体外生成sgRNA。每个模板DNA通过两个部分重叠的寡核苷酸(Macrogen,表X1)和Phusion聚合酶(Thermo Scientific),通过延伸而产生,其中使用了以下条件:-98°C 30sec {98°C 10sec, 54°C 20sec, 72°C 2min} x20, 72°C 5min。

[0258] [表9]

[0259] 制备用于体外转录的模板DNA的寡核苷酸

寡核苷酸	序列(5'-3')	SEQ ID NO
<b>BRI1 靶 1</b> (正向)	GAAATTAATACGACTCACTATAGGTTTGAA AGATGGAAGCGCGGGTTTTAGAGCTAGAA ATAGCAAGTTAAAATAAGGCTAGTCCG	73
[0260] <b>BRI1 靶 2</b> (正向)	GAAATTAATACGACTCACTATAGGTGAAAC TAAACTGGTCCACAGTTTTAGAGCTAGAAA TAGCAAGTTAAAATAAGGCTAGTCCG	74
<b>通用</b> (反向)	AAAAAAGCACCGACTCGGTGCCACTTTTTTC AAGTTGATAACGGACTAGCCTTATTTAAC TTGC	75

[0261] 纯化延伸的DNA,并用作模板,使用MEGAscript T7试剂盒(Life Technologies)体外生产向导RNA。然后通过苯酚/氯仿提取和乙醇沉淀,纯化向导RNA。为了制备Cas9/sgRNA复合体,10 $\mu$ l纯化的Cas9蛋白质(12 $\mu$ g/ $\mu$ l)和两个sgRNA(11 $\mu$ g/ $\mu$ l)各4 $\mu$ l混合于20 $\mu$ l NEB3缓冲液(New England Biolabs)中,并在37 $^{\circ}$ C下孵育10分钟。

[0262] 4-3.Cas9/sgRNA复合体转染原生质体

[0263] 在培养皿中无菌生长的4周龄拟南芥幼苗的叶子,在酶溶液(1%纤维素R10,0.5%离析酶R10,450mM甘露醇,20mM MES pH 5.7和CPW盐)中,25 $^{\circ}$ C暗处以40rpm振荡,消化8~16小时。过滤酶/原生质体溶液,并在100 $\times$ g下离心3~5分钟。在显微镜(X100)下使用血细胞计数器计数细胞后,将原生质体再悬浮在CPW溶液中。最后,原生质体以1 $\times$ 10<sup>6</sup>/ml再悬浮于MMG溶液(4mM HEPES pH 5.7,400mM甘露醇和15mM MgCl<sub>2</sub>)中。为了用Cas9/sgRNA复合体转染原生质体,将200 $\mu$ l原生质体悬浮液(200,000原生质体)与3.3或10 $\mu$ l Cas9/sgRNA复合体[Cas9蛋白质(6  $\mu$ g/ $\mu$ l)和两个sgRNA(各2.2 $\mu$ g/ $\mu$ l)]和200 $\mu$ l 40%聚乙二醇转染缓冲液(40%PEG4000,200mM甘露糖醇和100mM CaCl<sub>2</sub>)在2ml管中轻轻混合。在室温下孵育5~20分钟后,通过加入含有W5溶液(2mM MES pH 5.7, 154mM NaCl,125mM CaCl<sub>2</sub>和5mM KCl)的洗涤缓冲液,终止转染。然后通过100 $\times$ g下离心5分钟收集原生质体,用1ml W5溶液洗涤,在100  $\times$ g下再离心5分钟。将原生质体的密度调节至1 $\times$ 10<sup>5</sup>/ml并将其培养在含400mM葡萄糖的修饰的KM 8p液体培养基中。

[0264] 4-4.拟南芥原生质体和植物中突变的检测

[0265] 转染后24小时或72小时后,收集原生质体并分离基因组DNA。对跨越两个靶位点的基因组DNA区域进行PCR扩增,并进行T7E1测定。如图11所示,rgen以高频率诱导indel,频率从50%至70%。令人惊奇的是,在转染后24小时诱导了突变。显然,Cas9蛋白质在转染后立即发挥功能。纯化PCR产物,并克隆到T平端PCR克隆试剂盒(Solgent)。纯化质粒并用M13F引物进行Sanger测序。一个突变体序列在一个位点具有7-bp缺失(图12)。其他三个突变体序列在两个rgen位点之间具有~220-bp的DNA片段缺失。

[0266] 实施例5:使用细胞穿透肽或蛋白质转导结构域转导Cas9蛋白质

[0267] 5-1.His-Cas9编码质粒的构建

[0268] 使用先前描述的Cas9质粒{Cho,2013#166}作为模板,通过PCR扩增,制备在C末端

具有半胱氨酸的Cas9,并克隆到pET28-(a)载体(Novagen, Merck Millipore,Germany)中,在N-末端含有His-标签。

#### [0269] 5-2.细胞培养

[0270] 293T(人胚肾细胞系)和HeLa(人卵巢癌细胞系)生长在补充有10 %FBS和1%青霉素和链霉素的DMEM(GIBCO-BRL Rockville)中。

#### [0271] 5-3.Cas9蛋白质的表达和纯化

[0272] 为了表达Cas9蛋白质,用编码Cas9的pET28-(a)载体转染大肠杆菌BL21细胞,铺板至含有50 $\mu$ g/mL卡那霉素的Luria-Bertani(LB)琼脂培养基(Amresco,Solon,OH)。第二天,挑取单菌落并在含有50 $\mu$ g/mL卡那霉素的LB肉汤中在37 $^{\circ}$ C过夜培养。第二天,将0.10D 600的起子培养物接种到含有50 $\mu$ g/mL卡那霉素的Luria肉汤中,在37 $^{\circ}$ C培养2小时,直到OD 600达到0.6-0.8。为了诱导Cas9蛋白质表达,加入异丙基 $\beta$ -D-硫代半乳糖苷(IPTG)(Promega, Madison,WI)至终浓度0.5mM后,将细胞在30 $^{\circ}$ C过夜培养。

[0273] 通过在4000rpm离心15-20分钟收集细胞,再悬浮于裂解缓冲液(20mM Tris-Cl pH8.0,300mM NaCl,20mM咪唑,1X蛋白酶抑制剂混合物,1mg/ml 溶菌酶),并通过超声处理裂解(40%占空(duty),10秒脉冲,30秒间歇,放置于冰上10分钟)。在4 $^{\circ}$ C在15000rpm下离心20分钟后分离作为上清液的可溶级分。使用含Ni-NTA琼脂糖树脂的柱(QIAGEN)和AKTA prime 仪器(AKTA prime,GE Healthcare,UK),在4 $^{\circ}$ C下纯化Cas9蛋白质。在层析步骤期间,可溶性蛋白质级分以1mL/min的流速上样到Ni-NTA琼脂糖树脂柱(GE Healthcare,UK)。将柱用洗涤缓冲液(20mM Tris-Cl pH8.0,300mM NaCl,20mM咪唑,1X蛋白酶抑制剂混合物)洗涤,结合的蛋白质以0.5ml/min的流速用洗脱缓冲液(20mM Tris-Cl pH 8.0,300mM NaCl,250mM咪唑,1X蛋白酶抑制剂混合物)洗脱。将合并的洗脱级分浓缩并相对于存储缓冲液(50mM Tris-HCl,pH8.0,200mMKCl,0.1mM EDTA,1mM DTT,0.5mM PMSF,20%甘油)透析。通过Bradford测定法(Biorad, Hercules,CA)进行蛋白质浓度定量,使用牛血清白蛋白作为对照,通过SDS-PAGE分析纯度。

#### [0274] 5-4.Cas9与9R4L的缀合

[0275] 使用转子在室温下2小时并在4 $^{\circ}$ C下过夜,轻轻混合以浓度1mg/mL 稀释在PBS中的1mg Cas9蛋白质和25 $\mu$ L DW50中的50 $\mu$ g马来酰亚胺-9R4L肽(Peptron,Korea)。为了除去未结合的马来酰亚胺-9R4L,将样品用50kDa分子量截断膜对DPBS(pH 7.4)在4 $^{\circ}$ C下进行24小时透析。从透析膜收集Cas9-9R4L蛋白质并用Bradford测定法测定蛋白质量。

#### [0276] 5-5.sgRNA-9R4L的制备

[0277] sgRNA(1 $\mu$ g)轻轻加入到100 $\mu$ l DPBS(pH 7.4)中的各种量C9R4LC 肽(从1至40的重量比)中。将该混合物在室温下孵育30分钟,使用无RNA酶的去离子水稀释10倍。用动态光散射(Zetasizer-nano analyzer ZS; Malvern instruments,Worcestershire,UK),测定所形成的纳米粒子的流体动力学直径和z电位。

#### [0278] 5-6.Cas9蛋白质和sgRNA处理

[0279] 如下所述用Cas9-9R4L和sgRNA-C9R4LC处理细胞:将1 $\mu$ g sgRNA 和15 $\mu$ g C9R4LC肽加入到250mL OPTIMEM培养基中,并在室温下孵育30分钟。在接种后24小时,用OPTIMEM培养基洗涤细胞并用sgRNA-C9R4LC复合体在37 $^{\circ}$ C下处理4小时。将细胞用OPTIMEM培养基再次洗涤并用Cas9-9R4L在37 $^{\circ}$ C下处理2小时。处理后,培养基用含血清的完全培养基取代,并

在下一次处理前在37℃下孵育24小时。连续3天以相同的程序实施Cas9和sgRNA的多次处理。

[0280] 5-7. Cas9-9R4L和sgRNA-9R4L能编辑培养的哺乳动物细胞中的内源基因而不使用额外的递送工具

[0281] 为了确定Cas9-9R4L和sgRNA-9R4L在不使用额外的递送工具时是否可以编辑培养的哺乳动物细胞中的内源基因,我们用Cas9-9R4L和靶向CCR5基因的sgRNA-9R4L处理了293个细胞,并分析了基因组DNA。T7E1测定表明,在以Cas9-9R4L和sgRNA-9R4L两者处理的细胞中9%的CCR5基因被破坏,而在对照细胞(包括未经处理的细胞,用Cas9-9R或sgRNA-9R4L处理的细胞,或用未修饰的Cas9和sgRNA两者处理的细胞(图13))中没有观察到CCR5基因的破坏,这表明,用Cas9-9R4L蛋白质和与9R4L缀合的sgRNA(而不是未修饰的Cas9和sgRNA)处理可以导致哺乳动物细胞中高效的基因组编辑。

[0282] 实施例6:根据向导RNA结构控制脱靶突变

[0283] 最近,三个小组报道了RGEN在人细胞中具有脱靶效应。出乎我们意料的是,RGEN在与靶位点相差3至5个核苷酸的脱靶位点有效地诱导了突变。然而,我们注意到,我们所使用的RGEN和其他人使用的RGEN有一些差异。首先,我们用dualRNA——其是crRNA加tracrRNA,而不是由crRNA和tracrRNA的必要部分组成的单向导RNA(sgRNA)。第二,我们用合成的crRNA而非crRNA编码质粒转染K562细胞(而不是HeLa细胞)。用crRNA编码质粒转染HeLa细胞。其他小组使用sgRNA编码质粒。第三,我们的向导RNA在5'末端有两个附加的鸟嘌呤核苷酸,这对于T7聚合酶的体外有效转录是必需的。在其他人使用的sgRNA中没有包括这样的额外核苷酸。因此,我们的向导RNA的RNA序列可以表示为5'-GGX<sub>20</sub>,而5'-GX<sub>19</sub>代表其他人使用的序列,其中X<sub>20</sub>或GX<sub>19</sub>对应于20-bp靶序列。第一个鸟嘌呤核苷酸是细胞中由RNA聚合酶转录所需的。为了测试脱靶RGEN效应是否可归因于这些差异,我们选择了在人类细胞中以高频率诱导脱靶突变的4个RGEN(13)。首先,我们在K562细胞中比较了使用体外转录dualRNA的我们的方法与转染sgRNA编码质粒的方法,通过T7E1测定法测定了在靶和脱靶位点的突变频率。三个RGEN显示了在靶和脱靶位点可比较的突变频率,与向导RNA组成无关。有趣的是,当使用合成的dualRNA时,一个RGEN(VEFGA位点1)在一个验证的脱靶位点上没有诱导Indel,该脱靶位点与靶位点有三个核苷酸不同(称为OT1-11,图14)。但合成的dualRNA没有辨别出其它验证的脱靶位点(OT1-3),所述脱靶位点与靶位点有两个核苷酸不同。

[0284] 下一步,通过比较5'-GGX<sub>20</sub>(或5'-GGGX<sub>19</sub>)sgRNA与5'-GX<sub>19</sub>sgRNA,我们测试了在sgRNA的5'末端添加两个鸟嘌呤核苷酸是否可以使RGEN更特异。四个复合Cas9的GX<sub>19</sub>sgRNA同等有效地在靶和脱靶位点诱导了Indel,容忍多达四个碱基错配。与之形成鲜明对比,GGX<sub>20</sub>sgRNA有效区别脱靶位点。事实上,当我们使用四个GGX<sub>20</sub>sgRNA时,T7E1测定在七个验证的脱靶位点的六个中几乎没有检测到RGEN诱导的indel(图15)。然而,我们注意到,两个GGX<sub>20</sub>sgRNA(VEGFA位点1和3)在靶位点的活性比对应的GX<sub>19</sub>sgRNA的活性低。这些结果表明,或许通过改变向导RNA的稳定性、浓度或二级结构,在5'端的额外核苷酸可以影响在靶和脱靶位点的突变频率。

[0285] 这些结果表明,三个因素——使用合成的向导RNA而不是向导RNA编码质粒,使用dualRNA而非sgRNA,以及使用GGX<sub>20</sub>sgRNA而非GX<sub>19</sub>sgRNA——在脱靶位点的辨别上有累积

效应。

#### [0286] 实施例7:配对的Cas9切口酶

[0287] 原则上,单链断裂(SSB)不能被易错NHEJ修复,但仍引发高保真同源定向修复(HDR)或碱基切除修复。但通过HDR的切口酶诱导的定向诱变比核酸酶诱导的诱变效率低得多。我们推理,配对Cas9切口酶可以产生复合DSB,其触发通过NHEJ或HDR的DNA修复,导致有效的诱变(图16A)。此外,配对切口酶使基于Cas9的基因组编辑的特异性增倍。

[0288] 我们首先通过荧光毛细管电泳在体外测试了几种设计为靶向AAVS1 基因座中靶位点的Cas9核酸酶和切口酶(图16B)。与Cas9核酸酶(其切割DNA底物的两条链)不同,Cas9切口酶由向导RNA和Cas9的突变体形式组成,在所述Cas9突变体中催化性天冬氨酸残基改变为丙氨酸(D10A Cas9),该Cas9切口酶仅切割一条链,产生位点特异性缺口(图16C,D)。然而,有趣的是,一些切口酶(在图17A中的AS1,AS2,AS3,和S6) 在人类细胞中在靶位点诱导了Indel,这表明缺口可以体内转化为DSB,尽管低效。在相对的DNA链上产生两个相邻缺口的配对Cas9切口酶以从14 %至91%的频率产生Indel,相当于配对核酸酶(图17A)的效应。在三个基因组基因座上,产生5'突出端的两个切口的修复,比产生3'突出端的两个切口的修复,以更高频率地导致Indel形成(图17A和图18)。此外,配对切口酶比单一切口酶能够更高效地通过同源定向修复进行靶向基因组编辑(图19)。

[0289] 接下来,我们使用深度测序测定了脱靶位点上配对切口酶与核酸酶的突变频率。与三个sgRNA复合的Cas9核酸酶在六个位点诱导了脱靶突变,所述位点与其相应的标靶位点相差一个或两个核苷酸,突变频率范围从0.5 %至10%(图17B)。与此相反,配对Cas9切口酶在六个脱靶位点的任何一个上均未产生0.1%检测限以上的Indel。S2 Off-1位点(其与其标靶位点在PAM的第一位置(即,NGG中的N)上相差一个核苷酸)可以被认为是另一个标靶位点。正如所料,与S2 sgRNA复合的Cas9核酸酶在该位点和标靶位点上具有相同的效率。与之形成鲜明对比,与S2和AS2 sgRNA 复合的D10A Cas9以270倍的因数区分该位点和标靶位点。该配对切口酶还分别以160倍和990倍的因数区分AS2脱靶位点(图17B中的Off-1和Off-9)和靶位点。

#### [0290] 实施例8:配对Cas9切口酶诱导的染色体DNA剪接

[0291] 已经报道,由工程化的核酸酶ZFN和TALEN产生的两个并行DSB 能促进介于中间的染色体片段的大缺失。我们测试了由配对Cas9切口酶诱导的两个SSB是否也能在人细胞中产生缺失。我们使用PCR检测缺失事件,发现7个配对切口酶与配对Cas9核酸酶以同样效率诱导了高达1.1-kbp 染色体片段的缺失(图20A,B)。PCR产物的DNA序列证实了该缺失事件(图20C)。有趣的是,在七个缺失特异性PCR扩增子的两个中sgRNA 匹配序列保持完整(20C中下划线所示)。与此相反,Cas9核酸酶对没有产生含有完整靶位点的序列。这一发现表明,远离的两个缺口没有转化为两个分开的DSB以促进居间染色体片段的缺失。此外,因为解链温度非常高,不可能相隔100bp以上的两个切口在生理条件下造成具有大突出端的复合DSB。

[0292] 我们提出,两个远离的缺口通过头对头方向的链置换而被修复,导致在中间形成DSB,其通过NHEJ的修复引起小的缺失(图20D)。因为在这个过程中两个靶位点保持完整,切口酶能再次诱导SSB,反复触发循环,直到靶位点缺失。这一机制解释了为什么产生5'突出端的两个并列缺口但不是产生3'突出端的两个并列缺口在三个基因座有效诱导了Indel。

[0293] 然后我们调查了Cas9核酸酶和切口酶是否能诱发由于标靶和脱靶 DNA切割的NHEJ修复而导致的不希望的染色体易位(图21A)。使用PCR,我们能够检测到Cas9核酸酶诱导的易位(图21B,C)。使用分离自转染了编码AS2+S3 Cas9切口酶对的质粒的细胞的基因组DNA,没有扩增出这样的PCR产物。这一结果与以下事实一致,即,不同于其相应的核酸酶,AS2和S3切口酶在脱靶位点不产生Indel(图17B)。

[0294] 这些结果表明,配对Cas9切口酶允许在人类细胞中的靶向诱变和高达1-kbp染色体片段的大缺失。重要的是,配对切口酶在脱靶位点没有引起Indel,而在该脱靶位点处其相应的核酸酶诱发突变。此外,与核酸酶不同,配对切口酶不促进与脱靶DNA切割相关的不想要的易位。原则上,配对切口酶使Cas9介导的诱变的特异性加倍,这将扩大RNA向导酶在需要精确基因组编辑应用(如基因和细胞治疗)中的效用。这种方法的一个附加说明是,需要两个高活性sgRNA构成有效的切口酶对,这限制了可靶向的位点。如该研究和其他研究所示,并非所有sgRNA都具有同样活性。当单克隆而不是细胞群被用于进一步的研究或应用时,代表基因组中独特序列的向导RNA的选择以及优化的向导RNA的使用将足以避免与Cas9核酸酶相关的脱靶突变。我们提出,Cas9核酸酶和配对切口酶均为有利于细胞和生物体中精确基因组编辑的强大选项。

[0295] 实施例9:用CRISPR/Cas9衍生的RNA向导核酸内切酶的基因分型

[0296] 接着,我们推定RGEN可代替常规的限制酶用于限制性片段长度多态性(RFLP)分析中。当由核酸酶引起的DSB由易错非同源末端连接(NHEJ)系统修复时,包括RGEN的工程化核酸酶将在靶位点诱导Indel。被设计识别靶序列的RGEN不能切割带有indel的突变序列,但能有效地切割野生型靶序列。

[0297] 9-1.RGEN成分

[0298] 使用MEGAscript T7试剂盒(Ambion),根据制造商的说明书,通过体外转录制备crRNA和tracrRNA。转录的RNA在8%变性尿素PAGE凝胶上分离。切出含RNA的凝胶片,并转移到洗脱缓冲液。将RNA回收至无核酸酶水中,然后用苯酚:氯仿提取,氯仿提取和乙醇沉淀。用光谱法定量纯化的RNA。通过退火序列如下所示的寡核苷酸和其互补寡核苷酸,制备crRNA的模板:5'-GAAATTAATACGACTCACTATAGG X<sub>20</sub>GTTTTAGAGCTA TGCTGTTTTG-3'(SEQ ID NO:76),其中X<sub>20</sub>是靶序列。使用Phusion聚合酶(New England Biolabs),通过正向和反向寡核苷酸的延伸,合成了tracrRNA模板:

[0299] (5'-GAAATTAATACGACTCACTATAGGAACCATTCAAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCG-3'(SEQ ID NO:77)和

[0300] 5'-AAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTAACTTGCTATG-3'(SEQ ID NO:78))。

[0301] 9-2.重组Cas9蛋白质纯化

[0302] 在我们以前的实施例中使用的Cas9 DNA构建体(其编码在C末端融合His6标签的Cas9)插入至pET-28a表达载体。重组Cas9蛋白质表达于大肠杆菌菌株BL21(DE3),用1mM IPTG诱导后在25°C培养在LB培养基中4小时。收获细胞,并再悬浮于包含20mM Tris pH 8.0,500mM NaCl,5mM 咪唑和1mM PMSF的缓冲液中。将细胞在液氮中冷冻,在4°C解冻,并超声处理。离心后,将裂解物中的Cas9蛋白质结合至Ni-NTA琼脂糖树脂(Qiagen)上,用含有20mM Tris pH8.0,500mM NaCl和20mM咪唑的缓冲液洗涤,并用包含20mM Tris pH 8.0,

500mM NaCl和250mM咪唑的缓冲液洗脱。纯化的Cas9蛋白质对20mM HEPES (pH 7.5), 150mM KCl, 1mM DTT和10%甘油进行透析,并通过SDS-PAGE分析。

[0303] 9-3.T7内切核酸酶I测定

[0304] T7E1测定如下进行。简要地说,用基因组DNA扩增的PCR产物在 95°C变性,在16°C退火,并在37°C与5单位T7内切核酸酶I (New England BioLabs) 孵育20分钟。将反应产物用2%至2.5%的琼脂糖凝胶电泳分离。

[0305] 9-4.RGEN-RFLP分析

[0306] PCR产物 (100-150ng) 在37°C下与优化浓度(表10)的Cas9蛋白质、tracrRNA, crRNA在10 $\mu$ l NEB缓冲液3 (1X) 中孵育60分钟。切割反应后,加入RNA酶A (4 $\mu$ g), 并将反应混合物在37°C孵育30分钟,以除去RNA。用含有30%甘油,1.2%SDS和100mM EDTA的6X终止溶液缓冲液终止反应。产物用1-2.5%琼脂糖凝胶电泳分离,并用EtBr染色以显现。

[0307] [表10]

[0308] RFLP测定中RGEN成分的浓度

[0309]

靶名称	Cas9 (ng/ $\mu$ l)	crRNA (ng/ $\mu$ l)	tracrRNA (ng/ $\mu$ l)
<i>C4BPB</i>	100	25	60
<i>PIBF</i> -NGG-RGEN	100	25	60
<i>HLA-B</i>	1.2	0.3	0.7
<i>CCR5</i> -ZFN	100	25	60
<i>CTNNB1</i> 野生型特异的	30	10	20
<i>CTNNB1</i> 突变体特异的	30	10	20
<i>CCR5</i> WT 特异的	100	25	60
<i>CCR5</i> $\Delta$ 32 特异的	10	2.5	6
<i>KRAS</i> WT 特异的(wt)	30	10	20
<i>KRAS</i> 突变体特异的(m8)	30	10	20
<i>KRAS</i> WT 特异的 (m6)	30	10	20
<i>KRAS</i> 突变体特异的 (m6,8)	30	10	20
<i>PIK3CA</i> WT 特异的 (wt)	100	25	60
<i>PIK3CA</i> 突变体特异的(m4)	30	10	20
<i>PIK3CA</i> WT 特异的 (m7)	100	25	60
<i>PIK3CA</i> 突变体特异的(m4,7)	30	10	20
<i>BRAF</i> WT 特异的	30	10	20
<i>BRAF</i> 突变体特异的	100	25	60
<i>NRAS</i> WT 特异的	100	25	60
<i>NRAS</i> 突变体特异的	30	10	20
<i>IDH</i> WT 特异的	30	10	20
<i>IDH</i> 突变体特异的	30	10	20
<i>PIBF</i> -NAG-RGEN	30	10	60

[0310] [表11]

[0311] 引物

基因(位点)	方向	序列(5'至3')	SEQ ID NO
CCR5(RGEN)	F1	CTCCATGGTGCTATAGAGCA	79
	F2	GAGCCAAGCTCTCCATCTAGT	80
	R	GCCCTGTCAAGAGTTGACAC	81
CCR5(ZFN)	F	GCACAGGGTGGAACAAGATGGA	82
	R	GCCAGGTACCTATCGATTGTCAGG	83
CCR5(del32)	F	GAGCCAAGCTCTCCATCTAGT	84
	R	ACTCTGACTG GGTACCAGC	85
C4BPB	F1	TATTTGGCTGGTTGAAAGGG	86
	R1	AAAGTCATGAAATAAACACACCCA	87
	F2	CTGCATTGATATGGTAGTACCATG	88
	R2	GCTGTTCAATTGCAATGGAATG	89
CTNNB1	F	ATGGAGTTGGACATGGCCATGG	90
	R	ACTCACTATCCACAGTTCAGCATTAC C	91
KRAS	F	TGGAGATAGCTGTCAGCAACTTT	92
	R	CAACAA AGCAAAG- GTAAAGTTGGTAATAG	93
PIK3CA	F	GGTTTCAGGAGATGTGTTACAAGGC	94
	R	GATTGTGCAATTCCTATGCAATCGGTC	95
NRAS	F	CACTGGGTACTIONAATCTGTAGCCTC	96
	R	GGTCCAAGTCATTCCCAGTAGC	97
IDH1	F	CATCACTGCAGTTGTAGGTTATAACTATC C	98
	R	TTGAAAACCACAGATCTGGTTGAACC	99
BRAF	F	GGAGTGCCAAGAGAATATCTGG	100
	R	CTGAAACTGGTTTCAAATATTCGTTTTA AGG	101

[0312]

[0313]	<i>PIBF</i>	<i>F</i>	<i>GCTCTGTATGCCCTGTAGTAGG</i>	<i>102</i>
		<i>R</i>	<i>TTTGCATCTGACCTTACCTTTG</i>	<i>103</i>

[0314] 9-5. 质粒切割测定

[0315] 限制性内切酶处理的线性化质粒(100ng)在37℃下与Cas9蛋白质(0.1 μg)、tracrRNA(60ng)和crRNA(25ng)在10μg NEB3缓冲液(1X)中孵育60分钟。用含有30%甘油,1.2%SDS和100mM EDTA的6X终止溶液终止反应。产物用1%琼脂糖凝胶电泳分离,并用EtBr染色而显现。

[0316] 9-6. RFLP策略

[0317] 具有所需DNA特异性的新RGEN可以通过替换crRNA而容易地创建;一旦获得重组Cas9蛋白质,则不需要从头纯化定制蛋白质。当核酸酶引起的DSB通过易错非同源末端连接(NHEJ)修复时,包括RGEN的工程化核酸酶在靶位点诱发小的插入或缺失(indel)。设计成识别靶序列的RGEN有效切割野生型序列,但不切割具有indel的突变序列(图22)。

[0318] 我们首先测试RGEN能否差异地切割包含野生型或修饰C4BPB靶序列(其在切割位点具有1至3个碱基Indel)的质粒。具有这些Indel的六个质粒中没有一个被靶特异性crRNA、tracrRNA和重组Cas9蛋白质组成的C4BPB特异性RGEN5切割(图23)。与此相反,具有完整靶序列的质粒被该RGEN有效切割。

[0319] 9-7. 使用RGEN介导的RFLP检测由同一RGEN诱导的突变

[0320] 接下来,为了测试RGEN介导的RFLP检测由同一RGEN诱导的突变的可行性,我们利用了通过RGEN靶向C4BPB基因建立的基因修饰的K562人癌细胞克隆(表12)。

[0321] [表12]

[0322] 在这项研究中使用的RGEN的靶序列

基因	靶序列	SEQ ID NO
人C4BPB	<u>AATGACCACTACATCCTCAAGGG</u>	104
小鼠Pibf1	<u>AGATGATGTCTCATCATCAGAGG</u>	105

[0324] 本研究中使用的C4BPB突变体克隆具有范围从94bp缺失至67bp插入的各种突变(图24A)。重要的是,发生在突变体克隆中的所有突变均导致了RGEN靶位点的丧失。在分析的6个C4BPB克隆中,4个克隆具有野生型和突变体等位基因(+/-),2个克隆仅具有突变体等位基因(-/-)。

[0325] 用靶特异性crRNA、tracrRNA和从大肠杆菌表达和纯化的重组蛋白Cas9组成的RGEN,完全消化从野生型K562基因组DNA扩增的跨RGEN靶位点的PCR产物(图24B/第1道)。当使用RGEN对C4BPB突变体克隆进行RFLP分析时,含有野生型和突变体等位基因的+/-克隆的PCR扩增子被部分消化,不含有野生型等位基因的-/-克隆的扩增子根本不被消化,不产生对应于野生型序列的切割产物(图24B)。甚至在靶位点的单碱基插入也阻碍C4BPB RGEN对扩增的突变体等位基因的消化(#12和#28克隆),表明RGEN介导的RFLP的高特异性。我们对PCR扩增子平行地进行错配敏感的T7E1测定(图24B)。值得注意的是,T7E1测定无法区分-/-克隆和+/-克隆。更糟的是,T7E1测定不能区分含有相同突变序列的纯合突变体克隆和野生型克隆,原因是相同突变序列的退火将形成同源双链体。因此,RGEN介导的RFLP比常规的错配敏感的核酸酶测定法在分析工程化核酸酶(包括ZFN,TALEN和RGEN)诱导的突变体

克隆方面更优越。

[0326] 9-8. RGEN-RFLP分析的定量测定

[0327] 我们还调查了RGEN-RFLP分析是否是一个定量的方法。从C4BPB无效克隆和野生型细胞中分离的基因组DNA样品以各种比例混合,并用于PCR扩增。对PCR产物平行进行RGEN基因分型和T7E1测定(图25b)。正如所料,由RGEN切割的DNA与野生型对突变的比率成比例。与此相反,T7E1测定结果与从该比率推断的突变频率具有很差的相关性,特别是在高突变%——互补突变序列可相互杂交形成同源双链的情形——是不准确的。

[0328] 9-9. 使用RGEN介导的RFLP基因分型分析突变体建立者小鼠

[0329] 我们还应用RGEN介导的RFLP基因分型(简称RGEN基因分型),进行了突变体建立者小鼠的分析,该小鼠通过向小鼠一细胞胚胎中注射TALEN而建立(图26A)。我们设计并使用了识别Pibf1基因中的TALEN靶位点的RGEN(表10)。从野生型小鼠和突变体小鼠中分离基因组DNA,在PCR扩增后进行RGEN基因分型。RGEN基因分型成功地检测了各种突变,从1至27-bp的缺失(图26B)。不同于T7E1测定法,RGEN基因分型能够差异检测+/-与-/-建立者。

[0330] 9-10. 使用RGEN检测CCR5特异性ZFN在人细胞中诱导的突变

[0331] 此外,我们使用RGEN检测CCR5特异性ZFN(代表另一类工程化的核酸酶)在人细胞中诱导的突变(图27)。这些结果表明,RGEN可检测由RGEN自身外的其它核酸酶诱导的突变。事实上,我们预期RGEN可被设计为检测由大多数(即使不是全部的话)工程化核酸酶诱导的突变。在RGEN基因分型测定的设计中唯一的限制是:要求在Cas9蛋白质识别的PAM序列中有GG或AG(互补链上为CC或CT)二核苷酸,其平均每4bp出现一次。预期,在crRNA和PAM核苷酸中在几个碱基的种子区域内的任何位置处诱导的Indel都将破坏RGEN催化的DNA切割。事实上,我们在大多数(98%)ZFN和TALEN位点鉴定到至少一个RGEN位点。

[0332] 9-11. 使用RGEN检测多态性或变异

[0333] 接下来,我们设计并测试了靶向高度多态性基因座HLA-B(其编码人白细胞抗原B(又名MHC I类蛋白))的新RGEN(图28)。用RGEN质粒转染HeLa细胞,对基因组DNA平行进行T7E1和RGEN-RFLP分析。T7E1产生了假阳性条带,这些条带由于靶位点附近的序列多态性而导致(图25c)。但是,正如预期的,用于基因破坏的相同RGEN完全切割来自野生型细胞的PCR产物,但部分切割来自RGEN转染细胞的PCR产物,表明在靶位点存在RGEN诱导的Indel。这一结果表明,RGEN-RFLP分析比T7E1测定法有着明显的优势,特别是当不知道在所感兴趣的细胞中靶基因是否具有多态性或变异时。

[0334] 9-12. 通过RGEN-RFLP分析检测在癌症中的频发突变和天然存在的多态性

[0335] RGEN-RFLP分析的应用超出基因分型工程化核酸酶诱导的突变。我们试图利用RGEN基因分型检测癌症中的频发突变和天然存在的多态性。我们选择了人类结直肠癌细胞系HCT116,其在编码β-联蛋白的致癌基因CTNNB1上带有功能获得型3bp缺失。从HCT116基因组DNA扩增的PCR产物被野生型特异性和突变体特异性的RGEN部分切割,与HCT116细胞中的杂合基因型相符(图29A)。与之形成鲜明对比,从仅带有野生型等位基因的HeLa细胞的DNA扩增的PCR产物完全被野生型特异性RGEN消化,而根本不被突变特异性RGEN切割。

[0336] 我们还注意到,HEK293细胞在CCR5基因中具有32bp缺失(de132),CCR5基因编码重要的HIV感染共受体:纯合的de132 CCR5携带者对HIV感染具免疫性。我们设计了一个特异于de132等位基因的RGEN和特异于野生型等位基因的另一RGEN。正如预期的,野生型特异

性RGEN完全切割来自K562, SKBR3或HeLa细胞(用作野生型对照)的PCR产物,但是部分切割来自HEK293细胞的PCR产物(图30a),证实了在HEK293 细胞中存在不可切割的de132等位基因。然而,出乎意料的是,de132特异性RGEN同样有效地切割来自野生型细胞的PCR产物和来自HEK293细胞的PCR产物。有趣的是,该RGEN在靶位点的紧下游具有一个带有单碱基错配的脱靶位点(图30)。这些结果表明,RGEN可以用来检测天然存在的Indel,但由于其脱靶效应而不能区分具有单核苷酸多态性或点突变的序列。

[0337] 为了使用RGEN基因分型致癌的单核苷酸变异,我们通过采用单碱基错配向导RNA代替完全匹配的RNA来减弱RGEN活性。含有特异于野生型序列或突变体序列的完全匹配的向导RNA的RGEN切割两个序列(图 31a和32a)。与此相反,含有单碱基错配的向导RNA的RGEN区分了两个序列,从而能够对人癌细胞系中KRAS、PIK3CA和IDH1基因中的三个频发致癌点突变进行基因分型(图29b和图33a,b)。此外,使用识别 NAG PAM序列的RGEN,我们能够检测BRAF和NRAS基因中的点突变(图33c,d)。我们相信,我们能够使用RGEN-RFLP对几乎所有(如果不是全部的话)的人和其他基因组中的突变或多态性进行基因分型。

[0338] 上述数据提出,RGEN可以提供使用简单且稳健的RFLP分析用于各种序列变异的平台。在重编程靶序列方面具有高度灵活性,RGEN可用于检测各种遗传变异(单核苷酸变异,小的插入/缺失,结构变异),如疾病相关的复发突变、药物反应相关的患者基因型、以及由工程化的核酸酶在细胞中诱导的突变。这里,我们使用RGEN基因分型检测了由工程化的核酸酶在细胞和动物中诱导的突变。原则上,也可以使用RGEN特异性地检测和切割天然存在的变体和突变。

[0339] 根据上面的描述,本领域技术人员应当理解,在实施本发明时可使用本文所述实施方案的各种变体方案,而不偏离权利要求中限定的本发明的技术构思或者基本特征。在这方面,上述实施例仅用于举例说明的目的,并且本发明不旨在受这些实施例的限制。本发明的范围应被理解为包括在权利要求或者其等同构思的含义和范围上衍生的所有修改或修饰形式。

[0340] 本发明包括如下实施方案:

[0341] 1. 一种在真核细胞或生物体中切割靶DNA的组合物,其包含特异于靶DNA的向导RNA或编码向导RNA的DNA、和Cas蛋白质编码核酸或Cas蛋白质。

[0342] 2. 根据实施方案1所述的组合物,其中所述靶DNA是内源性靶DNA。

[0343] 3. 根据实施方案1所述的组合物,其中所述向导RNA是含有crRNA和 tracrRNA的dualRNA。

[0344] 4. 根据实施方案1所述的组合物,其中所述向导RNA是单链向导RNA(sgRNA)。

[0345] 5. 根据实施方案4所述的组合物,其中所述单链向导RNA含有crRNA和 tracrRNA的部分。

[0346] 6. 根据实施方案1所述的组合物,其中所述向导RNA在单链向导RNA或dualRNA的crRNA的5'末端还包含一个或多个附加的核苷酸。

[0347] 7. 根据实施方案1所述的组合物,其中所述向导RNA在单链向导RNA或 dualRNA的crRNA的5'末端还包含2个附加的鸟嘌呤核苷酸。

[0348] 8. 根据实施方案1所述的组合物,其在真核细胞或生物体中诱导靶向诱变。

[0349] 9. 根据实施方案1所述的组合物,其用于对真核细胞或生物体中的基因组进行体

外基因分型。

[0350] 10. 根据实施方案1所述的组合物,其中向导RNA和Cas蛋白质可以以对子的形式发挥作用,其中该对子包含在不同的链上诱导两个缺口的两个向导RNA。

[0351] 11. 根据实施方案1所述的组合物,其中向导RNA可以是分离的RNA的形式,或者在载体中编码,其中载体是病毒载体、质粒载体或农杆菌载体。

[0352] 12. 根据实施方案1所述的组合物,其包含特异于靶DNA的向导RNA和Cas9 蛋白质。

[0353] 13. 根据实施方案1所述的组合物,其用于离体或体内切割真核细胞或生物体中的靶DNA。

[0354] 14. 根据实施方案1所述的组合物,其中Cas蛋白质编码核酸或Cas蛋白质源自链球菌属细菌。

[0355] 15. 根据实施方案14所述的组合物,其中链球菌属细菌是化脓性链球菌。

[0356] 16. 根据实施方案15所述的组合物,其中Cas蛋白质识别NGG三核苷酸。

[0357] 17. 根据实施方案1所述的组合物,其中Cas蛋白质是Cas9蛋白质或其变体。

[0358] 18. 根据实施方案1所述的组合物,其中Cas蛋白质连接至蛋白质转导结构域。

[0359] 19. 根据实施方案17所述的组合物,其中Cas9蛋白质变体是Cas9的突变体形式,其中催化性天冬氨酸残基转变为任何其它氨基酸。

[0360] 20. 根据实施方案19所述的组合物,其中氨基酸是丙氨酸。

[0361] 21. 根据实施方案1所述的组合物,其中Cas蛋白质编码核酸包含SEQ ID NO.1 的核苷酸序列或与SEQ ID NO.1具有至少50%同源性的核苷酸序列。

[0362] 22. 根据实施方案1至22任一项所述的组合物用于在真核细胞或生物体中靶向诱变的用途。

[0363] 23. 用于在真核细胞或生物体中切割靶DNA的试剂盒,其含有根据实施方案1 至21任一项所述的组合物。

[0364] 24. 一种制备含有Cas蛋白质和向导RNA的真核细胞或生物体的方法,所述方法包括用Cas蛋白质编码核酸或Cas蛋白质以及向导RNA或编码向导RNA的 DNA共转染或顺序转染真核细胞或生物体的步骤。

[0365] 25. 根据实施方案24的方法,其中所述向导RNA是含有crRNA和tracrRNA 的 dualRNA。

[0366] 26. 根据实施方案24所述的方法,其中所述向导RNA是单链向导RNA。

[0367] 27. 根据实施方案26所述的方法,其中所述单链向导RNA含有crRNA和 tracrRNA的部分。

[0368] 28. 根据实施方案24所述的方法,其中所述向导RNA在单链向导RNA或 dualRNA的crRNA的5'末端还包含一个或多个附加的核苷酸。

[0369] 29. 根据实施方案24所述的方法,其中所述向导RNA在单链向导RNA或 dualRNA的crRNA的5'末端还包含2个附加的鸟嘌呤核苷酸。

[0370] 30. 根据实施方案24所述的方法,其中真核细胞或生物体共转染或顺序转染 Cas9蛋白质和向导RNA。

[0371] 31. 根据实施方案24所述的方法,其中顺序转染通过如下进行:用Cas蛋白质编码核酸进行第一转染,接着用裸向导RNA进行第二转染。

- [0372] 32. 根据实施方案24所述的方法,其中Cas蛋白质是Cas9蛋白质或其变体。
- [0373] 33. 根据实施方案24所述的方法,其中Cas蛋白质编码核酸或Cas蛋白质源自链球菌属细菌。
- [0374] 34. 根据实施方案33所述的方法,其中链球菌属细菌是化脓性链球菌。
- [0375] 35. 根据实施方案32所述的方法,其中Cas9蛋白质变体是Cas9的突变体形式,其中催化性天冬氨酸残基转变为任何其它氨基酸。
- [0376] 36. 根据实施方案35所述的方法,其中氨基酸是丙氨酸。
- [0377] 37. 根据实施方案24所述的方法,其中向导RNA和Cas蛋白质以对子的形式发挥作用,其中该对子含有在不同的DNA链上诱导两个缺口的两个向导RNA。
- [0378] 38. 根据实施方案37所述的方法,其中两个缺口至少分开100bp。
- [0379] 39. 根据实施方案24所述的方法,其中转染通过选自以下的方法进行:微注射、电穿孔、DEAE-葡聚糖处理、脂转染、纳米颗粒介导的转染、蛋白质转导结构域介导的转导、病毒介导的基因递送和PEG介导的原生质体转染。
- [0380] 40. 一种含有Cas蛋白质和向导RNA的真核细胞或生物体,其通过根据实施方案24至39任一项所述的方法制备。
- [0381] 41. 一种在真核细胞或生物体中切割靶DNA的方法,所述方法包括用根据实施方案1至21任一项的组合物转染含有靶DNA的真核细胞或生物体的步骤。
- [0382] 42. 根据实施方案41所述的方法,其中真核生物体是哺乳动物或植物。
- [0383] 43. 根据实施方案41所述的方法,其中向导RNA和Cas蛋白质以对子的形式发挥作用,其中该对子含有在不同的DNA链上诱导两个缺口的两个向导RNA。
- [0384] 44. 根据实施方案43所述的方法,其中两个缺口至少分开100bp。
- [0385] 45. 根据实施方案41所述的方法,其中转染是共转染或顺序转染。
- [0386] 46. 根据实施方案45所述的方法,其中顺序转染可通过如下进行:用Cas蛋白质编码核酸进行第一转染,接着用裸向导RNA进行第二转染。
- [0387] 47. 根据实施方案41所述的方法,其还包括分析切割模式的步骤,其中所述模式指示基因组中突变或变异的检测。
- [0388] 48. 一种胚胎,其含有通过根据实施方案1至21任一项的组合物编辑的基因组。
- [0389] 49. 一种基因组修饰的动物,其通过将实施方案48的胚胎转移到输卵管中获得。
- [0390] 50. 一种植物,其由实施方案24至39任一项所述的方法制备的基因组修饰的原生质体再生。
- [0391] 51. 一种制备基因组修饰动物的方法,所述方法包括将根据实施方案1至21任一项所述的组合物引入动物胚胎的步骤;和将胚胎转移到假孕代母输卵管中以产生基因组修饰动物的步骤。
- [0392] 52. 一种用于在分离的生物样品中基因分型突变或变异的组合物,所述组合物含有特异于靶DNA序列的向导RNA和Cas蛋白。
- [0393] 53. 根据实施方案52所述的组合物,其中突变或变异由工程化的核酸酶在细胞中诱导。
- [0394] 54. 根据实施方案52所述的组合物,其中突变或变异是天然存在的突变或变异。
- [0395] 55. 一种用于在分离的生物样品中基因分型病原微生物中的核酸序列的组合物,

所述组合物包含特异于靶DNA序列的向导RNA和Cas蛋白质。

[0396] 56. 一种用于在分离的生物样品中基因分型突变或变异的试剂盒,其包含实施方案52至55任一项所述的组合物。

[0397] 57. 一种用于在分离的生物样品中基因分型突变或变异的方法,其使用实施方案

[0398] 52至55任一项所述的组合物。

[0399] [参考文献]

[0400] 1.M.Jinek等,Science 337,816(Aug 17,2012) .

[0401] 2.H.Kim,E.Um,S.R.Cho,C.Jung,J.S.Kim,Nat Methods 8,941(Nov,2011) .

[0402] 3.H.J.Kim,H.J.Lee,H.Kim,S.W.Cho,J.S.Kim,Genome Res 19,1279(Jul, 2009) .

[0403] 4.E.E.Perez等,Nat Biotechnol 26,808(Jul,2008) .

[0404] 5.J.C.Miller等,Nat Biotechnol 29,143(Feb,2011) .

[0405] 6.C.Mussolino等,Nucleic Acids Res 39,9283(Nov,2011) .

[0406] 7.J.Cohen,Science 332,784(May 13,2011) .

[0407] 8.V.Pattanayak,C.L.Ramirez,J.K.Joung,D.R.Liu,Nat Methods 8,765(Sep, 2011) .

[0408] 9.R.Gabriel等,Nat Biotechnol 29,816(Sep,2011) .

[0409] 10.E.Kim等,Genome Res, (Apr 20,2012) .

[0410] 11.H.J.Lee,J.Kweon,E.Kim,S.Kim,J.S.Kim,Genome Res 22,539(Mar, 2012) .

[0411] 12.H.J.Lee,E.Kim,J.S.Kim,Genome Res 20,81(Jan,2010) .

[0412] 13.Fu Y,Foden JA,Khayter C,Maeder ML,Reyon D,Joung JK,Sander JD. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells.Nat Biotech advance online publication(2013)

## 序列表

<110> 基因工具股份有限公司 (TOOLGEN INCORPORATED)

<120> 包含特异于靶DNA的向导RNA和CAS蛋白质编码核酸或CAS蛋白质的用于切割靶DNA的组合物及其用途

<130> OPA13141PCT

<150> US 61/717,324

<151> 2012-10-23

<150> US 61/803,599

<151> 2013-03-20

<150> US 61/837,481

<151> 2013-06-20

<160> 111

<170> KopatentIn 2.0

<210> 1

<211> 4107

<212> DNA

<213> 人工序列

<220>

<223> Cas9-编码序列

<400> 1

```

atggacaaga agtacagcat cggcctggac atcggtagca acagcgtggg ctgggccgtg      60
atcaccgacg agtacaaggt gccagcaag aagttcaagg tgctgggcaa caccgaccgc      120
cacagcatca agaagaacct gatcggcgcc ctgctgttcg acagcggcga gaccgccgag      180
gccacccgcc tgaagcgcac cgcccggcgc cgctacacce gccgaagaa ccgcatctgc      240
tacctgcagg agatcttcag caacgagatg gccaaagtgg acgacagctt cttccaccgc      300
ctggaggaga gcttcctggt ggaggaggac aagaagcacg agcgccacce catcttcggc      360
aacatcgtgg acgaggtggc ctaccacgag aagtacccca ccatctacca cctgcgcaag      420
aagctggtgg acagcaccga caaggccgac ctgcgcctga tctacctggc cctggcccac      480
atgatcaagt tccgcggcca cttctgatc gagggcgacc tgaaccccga caacagcgac      540
gtggacaagc tgttcatcca gctggtgcag acctacaacc agctgttcga ggagaacccc      600
atcaacgcca gcggcgtgga cgccaaggcc atcctgagcg cccgcctgag caagagccgc      660
cgcctggaga acctgatcgc ccagctgccc ggcgagaaga agaacggcct gttcggcaac      720
ctgatcgcgc tgagcctggg cctgaccccc aacttcaaga gcaacttga cctggccgag      780
gacgccaagc tgcagctgag caaggacacc tacgacgacg acctggacaa cctgctggcc      840
cagatcggcg accagtacgc cgacctgttc ctggccgcca agaacctgag cgacgccatc      900
ctgctgagcg acatcctgcg cgtgaacacc gagatcacca agccccctc gacgcccagc      960
atgatcaagc gctacgacga gcaccaccag gacctgacce tgctgaaggc cctggtgcgc      1020
cagcagctgc ccgagaagta caaggagatc ttcttcgacc agagcaagaa cggctacgcc      1080

```

ggctacatcg	acggcggcgc	cagccaggag	gagttctaca	agttcatcaa	gcccacctcg	1140
gagaagatgg	acggcaccga	ggagctgctg	gtgaagctga	accgcgagga	cctgctgcgc	1200
aagcagcgca	ccttcgacaa	cggcagcatc	ccccaccaga	tccacctggg	cgagctgcac	1260
gccatcctgc	gccgccagga	ggacttctac	cccttcctga	aggacaaccg	cgagaagatc	1320
gagaagatcc	tgaccttccg	catcccctac	tacgtggggc	ccctggcccc	cggcaacagc	1380
cgcttcgcct	ggatgaccgg	caagagcgag	gagaccatca	ccccctggaa	cttcgaggag	1440
gtggtggaca	agggcgccag	cgcccagagc	ttcatcgagc	gcatgaccaa	cttcgacaag	1500
aacctgcccc	acgagaaggt	gctgccccag	cacagcctgc	tgtacgagta	cttcaccgtg	1560
tacaacgagc	tgaccaaggt	gaagtacgtg	accgagggca	tgcgcaagcc	cgcttctctg	1620
agcggcgagc	agaagaaggc	catcgtggac	ctgctgttca	agaccaaccg	caaggtgacc	1680
gtgaagcagc	tgaaggagga	ctacttcaag	aagatcgagt	gcttcgacag	cgtggagatc	1740
agcggcgtgg	aggaccgctt	caacgccagc	ctgggcacct	accacgacct	gctgaagatc	1800
atcaaggaca	aggacttctt	ggacaacgag	gagaacgagg	acatcctgga	ggacatcgtg	1860
ctgaccctga	ccctgttcga	ggaccgcgag	atgatcgagg	agcgcctgaa	gacctacgcc	1920
cacctgttcg	acgacaaggt	gatgaagcag	ctgaagcgcc	gccgctacac	cggctggggc	1980
cgcttgagcc	gcaagcttat	caacggcatc	cgcgacaagc	agagcggcaa	gaccatcctg	2040
gacttctctga	agagcgacgg	cttcgccaac	cgcaacttca	tgcagctgat	ccacgacgac	2100
agcctgacct	tcaaggagga	catccagaag	gcccaggtga	gcgccaggg	cgacagcctg	2160
cacgagcaca	tcgccaacct	ggccggcagc	cccgccatca	agaaggcat	cctgcagacc	2220
gtgaaggtgg	tggacgagct	ggtgaaggtg	atgggcccgc	acaagcccga	gaacatcgtg	2280
atcgagatgg	cccgcgagaa	ccagaccacc	cagaaggggc	agaagaacag	ccgcgagcgc	2340
atgaagcgca	tcgaggaggg	catcaaggag	ctgggcagcc	agatcctgaa	ggagcacccc	2400
gtggagaaca	cccagctgca	gaacgagaag	ctgtacctgt	actacctgca	gaacggcccgc	2460
gacatgtacg	tggaccagga	gctggacatc	aaccgcctga	gcgactacga	cgtggaccac	2520
atcgtgcccc	agagcttctt	gaaggacgac	agcatcgaca	acaaggtgct	gacccgcagc	2580
gacaagaacc	gcggaagag	cgacaacgtg	cccagcgagg	agtggtgtaa	gaagatgaag	2640
aactactggc	gccagctgct	gaacgccaag	ctgatcacc	agcgcaagtt	cgacaacctg	2700
accaaggccg	agcgcggcgg	cctgagcgag	ctggacaagg	ccggcttcat	caagcgccag	2760
ctggtggaga	cccgccagat	caccaagcac	gtggcccaga	tcttgacag	ccgcatgaac	2820
accaagtacg	acgagaacga	caagctgac	cgcgaggtga	aggtgatcac	cctgaagagc	2880
aagctggtga	gcgacttccg	caaggacttc	cagttctaca	aggtgcgca	gatcaacaac	2940
taccaccacg	cccacgacgc	ctacctgaac	gccgtggtgg	gcaccgcctt	gatcaagaag	3000
tacccaagc	tggagagcga	gttcgtgtac	ggcgactaca	aggtgtacga	cgtgcgcaag	3060
atgatcgcca	agagcgagca	ggagatcggc	aaggccaccg	ccaagtactt	cttctacagc	3120
aacatcatga	acttcttcaa	gaccgagatc	accctggcca	acggcgagat	ccgcaagcgc	3180
cccctgatcg	agaccaacgg	cgagaccggc	gagatcgtgt	gggacaaggg	ccgcgacttc	3240
gccaccgtgc	gcaaggtgct	gagcatgccc	caggtgaaca	tcgtgaagaa	gaccgaggtg	3300
cagaccggcg	gcttcagcaa	ggagagcatc	ctgccccagc	gcaacagcga	caagctgac	3360
gcccgcaaga	aggactggga	cccccaagaag	tacggcggct	tcgacagccc	caccgtggcc	3420

tacagcgtgc tgggtggtggc caaggtggag aagggaaga gcaagaagct gaagagcgtg 3480  
aaggagctgc tgggcatcac catcatggag cgcagcagct tcgagaagaa ccccatcgac 3540  
ttcctggagg ccaagggcta caaggaggtg aagaaggacc tgatcatcaa gctgcccagg 3600  
tacagcctgt tcgagctgga gaacggccgc aagcgcagtc tggccagcgc cggcgagctg 3660  
cagaagggca acgagctggc cctgcccagc aagtacgtga acttcctgta cctggccagc 3720  
cactacgaga agctgaaggg cagccccgag gacaacgagc agaagcagct gttcgtggag 3780  
cagcacaagc actacctgga cgagatcatc gagcagatca gcgagttcag caagcgcgtg 3840  
atcctggccg acgccaacct ggacaaggtg ctgagcgcct acaacaagca ccgcgacaag 3900  
cccatccgcg agcaggccga gaacatcatc cacctgttca ccctgaccaa cctgggccc 3960  
cccgccgct tcaagtactt cgacaccacc atcgaccgca agcgctacac cagcaccaag 4020  
gaggtgctgg acgccacct gatccaccag agcatcaccg gtctgtacga gacccgcac 4080  
gacctgagcc agctggggcgg cgactaa 4107

<210> 2

<211> 21

<212> PRT

<213> 人工序列

<220>

<223> 肽标签

<400> 2

Gly Gly Ser Gly Pro Pro Lys Lys Lys Arg Lys Val Tyr Pro Tyr Asp

1                    5                    10                    15

Val Pro Asp Tyr Ala

20

<210> 3

<211> 34

<212> DNA

<213> 人工序列

<220>

<223> F引物用于CCR5

<400> 3

aattcatgac atcaattatt atacatcgga ggag 34

<210> 4

<211> 34

<212> DNA

<213> 人工序列

<220>

<223> R引物用于CCR5

<400> 4

gatcctcctc cgatgtataa taattgatgt catg 34

<210>	5	
<211>	20	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	F1引物用于CCR5	
<400>	5	
	ctccatgggtg ctatagagca	20
<210>	6	
<211>	21	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	F2引物用于CCR5	
<400>	6	
	gagccaagct ctccatctag t	21
<210>	7	
<211>	20	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	R引物用于CCR5	
<400>	7	
	gccctgtcaa gagttgacac	20
<210>	8	
<211>	20	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	F1引物用于C4BPB	
<400>	8	
	tatthggctg gttgaaaggg	20
<210>	9	
<211>	24	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	R1引物用于C4BPB	
<400>	9	

aaagtcatga aataaacaca ccca	24
<210> 10	
<211> 24	
<212> DNA	
<213> 人工序列	
<220>	
<223> F2引物用于C4BPB	
<400> 10	
ctgcattgat atggtagtac catg	24
<210> 11	
<211> 21	
<212> DNA	
<213> 人工序列	
<220>	
<223> R2引物用于C4BPB	
<400> 11	
gctgttcatt gcaatggaat g	21
<210> 12	
<211> 20	
<212> DNA	
<213> 人工序列	
<220>	
<223> F1引物用于ADCY5	
<400> 12	
gctcccacct tagtgctctg	20
<210> 13	
<211> 20	
<212> DNA	
<213> 人工序列	
<220>	
<223> R1引物用于ADCY5	
<400> 13	
ggtggcagga acctgtatgt	20
<210> 14	
<211> 21	
<212> DNA	
<213> 人工序列	
<220>	
<223> F2引物用于ADCY5	

<400>	14	
	gtcattggcc agagatgtgg a	21
<210>	15	
<211>	20	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	R2引物用于ADCY5	
<400>	15	
	gtcccatgac aggcgtgtat	20
<210>	16	
<211>	20	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	F引物用于KCNJ6	
<400>	16	
	gcctggccaa gtttcagtta	20
<210>	17	
<211>	20	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	R1引物用于KCNJ6	
<400>	17	
	tggagccatt ggtttgcac	20
<210>	18	
<211>	22	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	R2引物用于KCNJ6	
<400>	18	
	ccagaactaa gccgtttctg ac	22
<210>	19	
<211>	20	
<212>	DNA	
<213>	人工序列	
<220>		

<223>	F1引物用于CNTNAP2	
<400>	19	
	atcaccgaca accagtttcc	20
<210>	20	
<211>	20	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	F2引物用于CNTNAP2	
<400>	20	
	tgcagtgcag actctttcca	20
<210>	21	
<211>	20	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	R引物用于CNTNAP2	
<400>	21	
	aaggacacag ggcaactgaa	20
<210>	22	
<211>	20	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	F1引物用于N/A Chr. 5	
<400>	22	
	tgtggaacga gtggtgacag	20
<210>	23	
<211>	22	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	R1引物用于N/A Chr. 5	
<400>	23	
	gctggattag gaggcaggat tc	22
<210>	24	
<211>	22	
<212>	DNA	
<213>	人工序列	

<220>		
<223>	F2引物用于N/A Chr. 5	
<400>	24	
	gtgctgagaa cgcttcatag ag	22
<210>	25	
<211>	23	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	R2引物用于N/A Chr. 5	
<400>	25	
	ggaccaaacc acatttettet cac	23
<210>	26	
<211>	20	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	F引物用于缺失	
<400>	26	
	ccacatctcg ttctcggtt	20
<210>	27	
<211>	20	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	R引物用于缺失	
<400>	27	
	tcacaagccc acagatattt	20
<210>	28	
<211>	105	
<212>	RNA	
<213>	人工序列	
<220>		
<223>	sgRNA用于CCR5	
<400>	28	
	ggugacauca auuauuauac auguuuuaga gcuagaaaua gcaaguuaaa auaaggcuag	60
	uccguuauca acuugaaaaa guggcaccga gucggugcuu uuuuu	105
<210>	29	
<211>	44	

<212>	RNA	
<213>	人工序列	
<220>		
<223>	crRNA用于CCR5	
<400>	29	
	ggugacauca auuauuauac auguuuuaga gcuaugcugu uuug	44
<210>	30	
<211>	86	
<212>	RNA	
<213>	人工序列	
<220>		
<223>	tracrRNA 用于CCR5	
<400>	30	
	ggaaccuuuc aaaacagcau agcaaguuaa aauaaggcua guccguuauc aacuugaaaa	60
	aguggcaccg agucggugcu uuuuuu	86
<210>	31	
<211>	86	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	Foxn1 #1 sgRNA	
<400>	31	
	gaaattaata cgactcacta taggcagtct gacgtcacac ttccgtttta gagctagaaa	60
	tagcaagtta aaataaggct agtccg	86
<210>	32	
<211>	86	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	Foxn1 #2 sgRNA	
<400>	32	
	gaaattaata cgactcacta taggacttec aggetccacc cgacgtttta gagctagaaa	60
	tagcaagtta aaataaggct agtccg	86
<210>	33	
<211>	86	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	Foxn1 #3 sgRNA	

<400>	33		
		gaaattaata cgactcacta taggccaggc tccacccgac tggagtttta gagctagaaa	60
		tagcaagtta aaataaggct agtccg	86
<210>	34		
<211>	86		
<212>	DNA		
<213>	人工序列		
<220>			
<223>	Foxn1 #4 sgRNA		
<400>	34		
		gaaattaata cgactcacta taggactgga gggcgaacc caaggtttta gagctagaaa	60
		tagcaagtta aaataaggct agtccg	86
<210>	35		
<211>	86		
<212>	DNA		
<213>	人工序列		
<220>			
<223>	Foxn1 #5 sgRNA		
<400>	35		
		gaaattaata cgactcacta taggaccca agggacctc atgcgtttta gagctagaaa	60
		tagcaagtta aaataaggct agtccg	86
<210>	36		
<211>	86		
<212>	DNA		
<213>	人工序列		
<220>			
<223>	Prkdc #1 sgRNA		
<400>	36		
		gaaattaata cgactcacta taggttagtt tttccagag acttgtttta gagctagaaa	60
		tagcaagtta aaataaggct agtccg	86
<210>	37		
<211>	86		
<212>	DNA		
<213>	人工序列		
<220>			
<223>	Prkdc #2 sgRNA		
<400>	37		
		gaaattaata cgactcacta taggttggtt tgcttggtt tatcgtttta gagctagaaa	60
		tagcaagtta aaataaggct agtccg	86

<210>	38	
<211>	86	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	Prkdc #3 sgRNA	
<400>	38	
	gaaattaata cgactcacta taggcacaag caaaccaaag tctcgtttta gagctagaaa	60
	tagcaagtta aaataaggct agtccg	86
<210>	39	
<211>	86	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	Prkdc #4 sgRNA	
<400>	39	
	gaaattaata cgactcacta taggcctcaa tgctaagcga cttcgtttta gagctagaaa	60
	tagcaagtta aaataaggct agtccg	86
<210>	40	
<211>	29	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	F1引物用于Foxn1	
<400>	40	
	gtctgtctat catctcttcc cttctctcc	29
<210>	41	
<211>	25	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	F2引物用于Foxn1	
<400>	41	
	tccctaatacc gatggctagc tccag	25
<210>	42	
<211>	23	
<212>	DNA	
<213>	人工序列	
<220>		

<223>	R1引物用于Foxn1	
<400>	42	
	acgagcagct gaagttagca tgc	23
<210>	43	
<211>	32	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	R2引物用于Foxn1	
<400>	43	
	ctactcaatg ctcttagagc taccaggett gc	32
<210>	44	
<211>	20	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	F引物用于Prkdc	
<400>	44	
	gactgttggtg gggagggccg	20
<210>	45	
<211>	24	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	F2引物用于Prkdc	
<400>	45	
	gggagggccg aaagtcttat ttg	24
<210>	46	
<211>	28	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	R1引物用于Prkdc	
<400>	46	
	cctgaagact gaagttggca gaagtgag	28
<210>	47	
<211>	27	
<212>	DNA	
<213>	人工序列	

<220>		
<223>	R2引物用于Prkdc	
<400>	47	
	ctttagggct tcttctctac aatcacg	27
<210>	48	
<211>	38	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	F引物用于Foxn1	
<400>	48	
	ctcgggtgtg agccctgacc tcggtgtgta gcctgac	38
<210>	49	
<211>	21	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	R引物用于Foxn1	
<400>	49	
	agactggcct ggaactcaca g	21
<210>	50	
<211>	23	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	F引物用于Foxn1	
<400>	50	
	cactaaagcc tgtcaggaag ccg	23
<210>	51	
<211>	21	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	R引物用于Foxn1	
<400>	51	
	ctgtggagag cacacagcag c	21
<210>	52	
<211>	19	
<212>	DNA	

<213>	人工序列	
<220>		
<223>	F引物用于Foxn1	
<400>	52	
	gctgcgacct gagaccatg	19
<210>	53	
<211>	26	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	R引物用于Foxn1	
<400>	53	
	cttcaatggc ttctgetta ggetac	26
<210>	54	
<211>	23	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	F引物用于Foxn1	
<400>	54	
	ggttcagatg aggccatcct ttc	23
<210>	55	
<211>	24	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	R引物用于Foxn1	
<400>	55	
	cctgatctgc aggettaacc cttg	24
<210>	56	
<211>	22	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	F引物用于Prkdc	
<400>	56	
	ctcacctgca catcacatgt gg	22
<210>	57	
<211>	20	

<212>	DNA	
<213>	人工序列	
<220>		
<223>	R引物用于Prkdc	
<400>	57	
	ggcatccacc ctatggggtc	20
<210>	58	
<211>	25	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	F引物用于Prkdc	
<400>	58	
	gccttgacct agagcttaaa gagcc	25
<210>	59	
<211>	25	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	R引物用于Prkdc	
<400>	59	
	ggtcttgтта gcaggaagga cactg	25
<210>	60	
<211>	27	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	F引物用于Prkdc	
<400>	60	
	aaaactctgc ttgatgggat atgtggg	27
<210>	61	
<211>	26	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	R引物用于Prkdc	
<400>	61	
	ctctcactgg ttatctgtgc tccttc	26
<210>	62	

<211>	23	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	F引物用于Prkdc	
<400>	62	
	ggatcaatag gtggtggggg atg	23
<210>	63	
<211>	27	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	R引物用于Prkdc	
<400>	63	
	gtgaatgaca caatgtgaca gcttcag	27
<210>	64	
<211>	28	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	F引物用于Prkdc	
<400>	64	
	cacaagacag acctctcaac attcagtc	28
<210>	65	
<211>	32	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	R引物用于Prkdc	
<400>	65	
	gtgcatgcat ataatccatt ctgattgetc tc	32
<210>	66	
<211>	17	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	F1引物用于Prkdc	
<400>	66	
	gggaggcaga ggcaggt	17

<210>	67	
<211>	23	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	F2引物用于Prkdc	
<400>	67	
	ggatctctgt gagtttgagg cca	23
<210>	68	
<211>	24	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	R1引物用于Prkdc	
<400>	68	
	gctccagaac tcactcttag gctc	24
<210>	69	
<211>	20	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物用于Foxn1	
<400>	69	
	ctactccctc cgcagtctga	20
<210>	70	
<211>	20	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物用于Foxn1	
<400>	70	
	ccaggcctag gttccaggta	20
<210>	71	
<211>	20	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物用于Prkdc	
<400>	71	

ccccagcatt gcagatttcc	20
<210> 72	
<211> 23	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物用于Prkdc	
<400> 72	
agggcttctt ctctacaatc acg	23
<210> 73	
<211> 86	
<212> DNA	
<213> 人工序列	
<220>	
<223> BRI1靶1	
<400> 73	
gaaattaata cgactcacta taggtttgaa agatggaagc gcgggtttta gagctagaaa	60
tagcaagtta aaataaggct agtccg	86
<210> 74	
<211> 86	
<212> DNA	
<213> 人工序列	
<220>	
<223> BRI1靶2	
<400> 74	
gaaattaata cgactcacta taggtgaaac taaactggtc cacagtttta gagctagaaa	60
tagcaagtta aaataaggct agtccg	86
<210> 75	
<211> 64	
<212> DNA	
<213> 人工序列	
<220>	
<223> 通用	
<400> 75	
aaaaaagcac cgactcggcg ccaacttttc aagttgataa cggactagcc ttattttaac	60
ttgc	64
<210> 76	
<211> 65	
<212> DNA	

<213>	人工序列	
<220>		
<223>	用于crRNA的模板	
<400>	76	
	gaaattaata cgactcacta taggnnnnnn nnnnnnnnnn nnnngtttta gagctatgct	60
	gtttt	65
<210>	77	
<211>	67	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	tracrRNA	
<400>	77	
	gaaattaata cgactcacta taggaacct tcaaacagc atagcaagtt aaaataaggc	60
	tagtccg	67
<210>	78	
<211>	69	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	tracrRNA	
<400>	78	
	aaaaaaagca ccgactcgt gccactttt caagtgata acgactagc cttattttaa	60
	cttgctatg	69
<210>	79	
<211>	20	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	79	
	ctccatggtg ctatagagca	20
<210>	80	
<211>	21	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	80	

gagccaagct ctccatctag t	21
<210> 81	
<211> 20	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物	
<400> 81	
gccctgtcaa gagttgacac	20
<210> 82	
<211> 22	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物	
<400> 82	
gcacagggtg gaacaagatg ga	22
<210> 83	
<211> 24	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物	
<400> 83	
gccaggtacc tatcgattgt cagg	24
<210> 84	
<211> 21	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物	
<400> 84	
gagccaagct ctccatctag t	21
<210> 85	
<211> 20	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物	

<400>	85	
	actctgactg ggtcaccagc	20
<210>	86	
<211>	20	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	86	
	tatttggctg gttgaaagg	20
<210>	87	
<211>	24	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	87	
	aaagtcatga aataaacaca ccca	24
<210>	88	
<211>	24	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	88	
	ctgcattgat atggtagtac catg	24
<210>	89	
<211>	21	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	89	
	gctgttcatt gcaatggaat g	21
<210>	90	
<211>	22	
<212>	DNA	
<213>	人工序列	
<220>		

<223>	引物	
<400>	90	
	atggagttgg acatggccat gg	22
<210>	91	
<211>	28	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	91	
	actcactatc cacagttcag catttacc	28
<210>	92	
<211>	23	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	92	
	tggagatagc tgtcagcaac ttt	23
<210>	93	
<211>	29	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	93	
	caacaaagca aaggtaaagt tggtaatag	29
<210>	94	
<211>	25	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	94	
	ggtttcagga gatgtgttac aaggc	25
<210>	95	
<211>	27	
<212>	DNA	
<213>	人工序列	

<220>		
<223>	引物	
<400>	95	
	gattgtgcaa ttcctatgca atcggtc	27
<210>	96	
<211>	25	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	96	
	cactgggtac ttaatctgta gcctc	25
<210>	97	
<211>	23	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	97	
	ggttccaagt cattcccagt agc	23
<210>	98	
<211>	30	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	98	
	catcactgca gttgtaggtt ataactatcc	30
<210>	99	
<211>	26	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	99	
	ttgaaaacca cagatctggt tgaacc	26
<210>	100	
<211>	22	
<212>	DNA	

<213>	人工序列	
<220>		
<223>	引物	
<400>	100	
	ggagtgccaa gagaatatct gg	22
<210>	101	
<211>	32	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	101	
	ctgaaactgg tttcaaaata ttcgttttaa gg	32
<210>	102	
<211>	22	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	102	
	gctctgtatg ccctgtagta gg	22
<210>	103	
<211>	22	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	103	
	tttgcacatctg accttacctt tg	22
<210>	104	
<211>	23	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	RGEN的靶序列	
<400>	104	
	aatgaccact acatcctcaa ggg	23
<210>	105	
<211>	23	

<212>	DNA	
<213>	人工序列	
<220>		
<223>	RGEN的靶序列	
<400>	105	
	agatgatgtc tcatcatcag agg	23
<210>	106	
<211>	4170	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	Cas9-编码序列在p3s-Cas9HC中 (人源化的, C-末端加标签, 人细胞实验)	
<400>	106	
	atggacaaga agtacagcat cggcctggac atcgtacca acagcgtggg ctgggccgtg	60
	atcaccgacg agtacaaggt gccagcaag aagttcaagg tgctgggcaa caccgaccgc	120
	cacagcatca agaagaacct gatcggcgcc ctgctgttcg acagcggcga gaccgccgag	180
	gccacccgcc tgaagcgcac cgcccgccgc cgctacacc gccgaagaa ccgcatctgc	240
	tacctgcagg agatcttcag caacgagatg gccaaaggtgg acgacagctt cttccaccgc	300
	ctggaggaga gcttcctggt ggaggaggac aagaagcacg agcgccacc catcttcggc	360
	aacatcgtgg acgagggtgg ctaccacgag aagtaccca ccatctacca cctgcgcaag	420
	aagctggtgg acagcaccga caaggccgac ctgcgcctga tctacctggc cctggccacc	480
	atgatcaagt tccgcggcca cttcctgatc gagggcgacc tgaaccccga caacagcgac	540
	gtggacaagc tgttcatcca gctggtgcag acctacaacc agctgttcga ggagaacccc	600
	atcaacgcca gcggcgtgga cgccaaggcc atcctgagcg cccgcctgag caagagccgc	660
	cgcctggaga acctgatcgc ccagctgccc ggcgagaaga agaacggcct gttcggcaac	720
	ctgatcgccc tgagcctggg cctgaccccc aacttcaaga gcaacttcga cctggccgag	780
	gacgccaagc tgcagctgag caaggacacc tacgacgacg acctggacaa cctgctggcc	840
	cagatcggcg accagtacgc cgacctgttc ctggccgcca agaacctgag cgacgccatc	900
	ctgctgagcg acatcctgcg cgtgaacacc gagatcaca aggccccct gagcgccagc	960
	atgatcaagc gctacgacga gcaccaccag gacctgacc tgctgaaggc cctggtgcgc	1020
	cagcagctgc ccgagaagta caaggagatc ttcttcgacc agagcaagaa cggctacgcc	1080
	ggctacatcg acggcggcgc cagccaggag gagttctaca agttcatcaa gcccatcctg	1140
	gagaagatgg acggcaccga ggagctgctg gtgaagctga accgcgagga cctgctgcgc	1200
	aagcagcgca ccttcgacaa cggcagcatc ccccaccaga tccacctggg cgagctgcac	1260
	gccatcctgc gccgccagga ggactttctac cccttctga aggacaaccg cgagaagatc	1320
	gagaagatcc tgaccttcg catcccctac tacgtgggcc ccctggcccc cggaacagc	1380
	cgcttcgcct ggatgaccg caagagcgag gagacatca cccctggaa cttcgaggag	1440
	gtggtggaca agggcgccag cgcccagagc ttcctcgagc gcatgaccaa cttcgacaag	1500

aacctgcccc	acgagaaggt	gctgccccaa	cacagcctgc	tgtacgagta	cttcaccgtg	1560
tacaacgagc	tgaccaaggt	gaagtacgtg	accgagggca	tgcgcaagcc	cgcttcctg	1620
agcggcgagc	agaagaaggc	catcgtggac	ctgctgttca	agaccaaccg	caaggtgacc	1680
gtgaagcagc	tgaaggagga	ctacttcaag	aagatcgagt	gcttcgacag	cgtaggagtc	1740
agcggcgtgg	aggaccgctt	caacgccagc	ctgggcacct	accacgacct	gctgaagatc	1800
atcaaggaca	aggacttcct	ggacaacgag	gagaacgagg	acatcctgga	ggacatcgtg	1860
ctgacctga	ccctgttcga	ggaccgag	atgatcgagg	agcgcctgaa	gacctagcc	1920
cacctgttcg	acgacaaggt	gatgaagcag	ctgaagcgcc	gccgctacac	cggtggggc	1980
cgctgagcc	gcaagcttat	caacggcatc	cgcgacaagc	agagcggcaa	gaccatcctg	2040
gacttcctga	agagcgacgg	cttcgccaac	cgcaacttca	tgcagctgat	ccacgacgac	2100
agcctgacct	tcaaggagga	catccagaag	gcccaggtga	gcgccagg	cgacagcctg	2160
cacgagcaca	tcgccaacct	ggccggcagc	cccgccatca	agaaggcat	cctgcagacc	2220
gtgaaggtgg	tggacgagct	ggtgaaggtg	atgggccgcc	acaagcccga	gaacatcgtg	2280
atcgagatgg	cccgcgagaa	ccagaccacc	cagaagggcc	agaagaacag	ccgcgagcgc	2340
atgaagcgca	tcgaggaggg	catcaaggag	ctgggcagcc	agatcctgaa	ggagcacc	2400
gtggagaaca	cccagctgca	gaacgagaag	ctgtacctgt	actacctgca	gaacggccgc	2460
gacatgtacg	tggaccagga	gctggacatc	aaccgcctga	gcgactacga	cgtaggacc	2520
atcgtgcccc	agagcttcct	gaaggacgac	agcatcgaca	acaaggtgct	gaccgcagc	2580
gacaagaacc	gcggaagag	cgacaacgtg	cccagcgagg	aggtggtgaa	gaagatgaag	2640
aactactggc	gccagctgct	gaacgccaag	ctgatcacc	agcgcaagtt	cgacaacctg	2700
accaaggccg	agcgcggcgg	cctgagcgag	ctggacaagg	ccggcttcat	caagcgccag	2760
ctggtggaga	cccgccagat	caccaagcac	gtggcccaga	tcctggacag	ccgcatgaac	2820
accaagtacg	acgagaacga	caagctgatc	cgcgaggtga	aggtgatcac	cctgaagagc	2880
aagctggtga	gcgacttccg	caaggacttc	cagttctaca	aggtgcgca	gatcaacaac	2940
taccaccacg	cccacgacgc	ctacctgaac	gccgtggtgg	gcaccgccct	gatcaagaag	3000
tacccaagc	tggagagcga	gttcgtgtac	ggcgactaca	aggtgtacga	cgtagcgaag	3060
atgatcgcca	agagcgagca	ggagatcggc	aaggccaccg	ccaagtactt	cttctacagc	3120
aacatcatga	acttcttcaa	gaccgagatc	accctggcca	acggcgagat	ccgcaagcgc	3180
cccctgatcg	agaccaacgg	cgagaccggc	gagatcgtgt	gggacaagg	ccgcgacttc	3240
gccaccgtgc	gcaaggtgct	gagcatgccc	caggtgaaca	tcgtgaagaa	gaccgaggtg	3300
cagaccggcg	gcttcagcaa	ggagagcatc	ctgccaagc	gcaacagcga	caagctgatc	3360
gcccgaaga	aggactggga	ccccagaag	tacggcggt	tcgacagccc	caccgtggcc	3420
tacagcgtgc	tggtggtggc	caaggtggag	aagggaaga	gcaagaagct	gaagagcgtg	3480
aaggagctgc	tgggcatcac	catcatggag	cgcagcagct	tcgagaagaa	ccccatcgac	3540
ttcctggagg	ccaagggcta	caaggaggtg	aagaaggacc	tgatcatcaa	gctgccccag	3600
tacagcctgt	tcgagctgga	gaacggccgc	aagcgcagtc	tggccagcgc	cggcgagctg	3660
cagaagggca	acgagctggc	cctgcccagc	aagtacgtga	acttctgta	cctggccagc	3720
cactacgaga	agctgaagg	cagccccgag	gacaacgagc	agaagcagct	gttcgtggag	3780
cagcacaagc	actacctgga	cgagatcatc	gagcagatca	gcgagttcag	caagcgcgtg	3840

atcctggccg acgccaacct ggacaaggtg ctgagcgcct acaacaagca ccgcgacaag	3900
cccatccgcg agcaggccga gaacatcatc cacctgttca ccctgaccaa cctgggcgcc	3960
cccgccgcct tcaagtactt cgacaccacc atcgaccgca agcgcctacac cagcaccaag	4020
gaggtgctgg acgccaccct gatccaccag agcatcaccg gtctgtacga gacccgcac	4080
gacctgagcc agctgggcgg cgacggcggc tccggacctc caaagaaaa gagaaaagta	4140
taccctacg acgtgcccga ctacgcctaa	4170
<210> 107	
<211> 4194	
<212> DNA	
<213> 人工序列	
<220>	
<223> Cas9 编码序列在p3s-Cas9HN中 (人源化的密码子, N-端加标签 (下划线), 人细胞实验)	
<400> 107	
atggtgtacc cctacgacgt gcccgactac gccgaattgc ctccaaaaa gaagagaaag	60
gtagggatcc gaattcccgg ggaaaaaccg gacaagaagt acagcatcgg cctggacatc	120
ggtaccaaca gcgtgggctg ggccgtgatc accgacgagt acaaggtgcc cagcaagaag	180
ttcaagggtg tgggcaacac cgaccgccac agcatcaaga agaacctgat cggcgcctg	240
ctgttcgaca gcggcgagac cgccgaggcc accgcctga agcgcaccgc ccgccccgc	300
tacaccgcc gcaagaaccg catctgctac ctgcaggaga tcttcagcaa cgagatggcc	360
aaggtggacg acagcttctt ccaccgcctg gaggagagct tctgtgtgga ggaggacaag	420
aagcacgagc gccaccccat cttcggaac atcgtggacg aggtggccta ccacgagaag	480
tacccacca tctaccacct gcgcaagaag ctggtggaca gcaccgaaa ggccgacctg	540
cgctgatct acctggccct ggcccacatg atcaagtcc gcggccactt cctgatcgag	600
ggcgacctga accccgaaa cagcgacgtg gacaagctgt tcatccagct ggtgcagacc	660
tacaaccagc tgttcgagga gaacccatc aacgccagcg gcgtggacgc caaggccatc	720
ctgagcgccc gcctgagcaa gagccgccgc ctggagaacc tgatcgcca gctgcccggc	780
gagaagaaga acggcctggt cggcaacctg atcgccctga gcctgggcct gaccccaac	840
ttcaagagca acttcgacct ggccgaggac gccaaactgc agctgagcaa ggacacctac	900
gacgacgacc tggacaacct gctggcccag atcggcgacc agtacgccga cctgttctctg	960
gccccaaga acctgagcga cgccatctg ctgagcgaca tctgcgctg gaacaccgag	1020
atcaccaagg cccccctgag cgccagcatg atcaagcgt acgacgagca ccaccaggac	1080
ctgaccctgc tgaaggccct ggtgcgccag cagctgcccg agaagtaca ggagatcttc	1140
ttcgaccaga gcaagaacgg ctacgccggc tacatcgac gcggcgccag ccaggaggag	1200
ttctacaagt tcatcaagcc catcctggag aagatggac gcaccgagga gctgctggtg	1260
aagctgaacc gcgaggacct gctgcgcaag cagcgcacct tcgacaacgg cagcateccc	1320
caccagatcc acctgggcga getgcacgcc atcctgcgcc gccaggagga cttctacccc	1380
ttcctgaagg acaaccgcga gaagatcgag aagatcctga ccttcgcat ccctactac	1440
gtgggcccc tggcccgcgg caacagccgc ttcgctgga tgaccgcaa gagcgaggag	1500

accatcacc	cctggaactt	cgaggaggtg	gtggacaagg	gcccagcgc	ccagagcttc	1560
atcgagcgc	tgaccaactt	cgacaagaac	ctgccaacg	agaaggtgct	gccaagcac	1620
agcctgctgt	acgagtactt	caccgtgtac	aacgagctga	ccaagtgaa	gtacgtgacc	1680
gagggcatgc	gcaagcccgc	cttcctgagc	ggcgagcaga	agaaggccat	cgtggacctg	1740
ctgttcaaga	ccaaccgcaa	ggtgaccgtg	aagcagctga	aggaggacta	cttcaagaag	1800
atcgagtgt	tcgacagcgt	ggagatcagc	ggcgtggagg	accgcttcaa	cgccagcctg	1860
ggcacctacc	acgacctgct	gaagatcatc	aaggacaagg	acttcctgga	caacgaggag	1920
aacgaggaca	tcctggagga	catcgtgctg	accctgacct	tgttcgagga	ccgagagatg	1980
atcgaggagc	gcctgaagac	ctacgccac	ctgttcgacg	acaaggtgat	gaagcagctg	2040
aagcggccgc	gctacaccgg	ctggggccgc	ctgagccgca	agcttatcaa	cggcatccgc	2100
gacaagcaga	gcggaagac	catcctggac	ttcctgaaga	gagcagcgtt	cgccaaccgc	2160
aacttcatgc	agctgatcca	cgacgacagc	ctgaccttca	aggaggacat	ccagaaggcc	2220
caggtgagcg	gccagggcga	cagcctgcac	gagcacatcg	ccaacctggc	cgccagcccc	2280
gcatcaaga	aggcctcct	gcagaccgtg	aaggtggtgg	acgagctggt	gaaggtgatg	2340
ggccgccaca	agcccagaaa	catcgtgatc	gagatggccc	gagagaacca	gaccaccag	2400
aagggccaga	agaacagccg	cgagcgcgat	aagcgcgatc	aggaggcat	caaggagctg	2460
ggcagccaga	tcctgaagga	gcaccccgtg	gagaacaccc	agctgcagaa	cgagaagctg	2520
tacctgtact	acctgcagaa	cggccgcgac	atgtacgtgg	accaggagct	ggacatcaac	2580
cgctgagcg	actacgacgt	ggaccacatc	gtgcccaga	gcttcctgaa	ggacgacagc	2640
atcgacaaca	aggtgctgac	ccgagcgcac	aagaaccgcg	gcaagagcga	caacgtgccc	2700
agcgaggagg	tggtgaagaa	gatgaagaac	tactggcgcc	agctgctgaa	cgccaagctg	2760
atcaccagc	gcaagtcca	caacctgacc	aaggccgagc	gcccggcct	gagcagctg	2820
gacaaggccg	gcttcatcaa	gcgccagctg	gtggagacct	gccagatcac	caagcacgtg	2880
gcccagatcc	tgacagccg	catgaacacc	aagtacgacg	agaacgaaa	gctgatccgc	2940
gaggtgaagg	tgatcacctt	gaagagcaag	ctggtgagcg	acttccgaa	ggacttccag	3000
ttctacaagg	tgcgcgagat	caacaactac	caccacgccc	acgacgccta	cctgaacgcc	3060
gtggtgggca	ccgccctgat	caagaagtac	ccaagctgg	agagcgagtt	cgtgtacggc	3120
gactacaagg	gtacgacgt	gcgcaagatg	atgccaaga	gagcagcaga	gatcggcaag	3180
gccaccgcca	agtacttctt	ctacagcaac	atcatgaact	tcttcaagac	cgagatcacc	3240
ctggccaacg	gcgagatccg	caagcgcgcc	ctgatcgaga	ccaacggcga	gaccggcgag	3300
atcgtgtggg	acaagggccg	cgacttcgcc	accgtgcgca	aggtgctgag	catgccccag	3360
gtgaacatcg	tgaagaagac	cgaggtgcag	accggcgctt	tcagcaagga	gagcatcctg	3420
ccaagcgc	acagcgaaa	gctgatcgcc	cgcaagaagg	actgggacct	caagaagtac	3480
ggcggcttcg	acagccccac	cgtggcctac	agcgtgctgg	tggtggccaa	ggtggagaag	3540
ggcaagagca	agaagctgaa	gagcgtgaag	gagctgctgg	gcatcacat	catggagcgc	3600
agcagcttcg	agaagaacct	catcgaattc	ctggaggcca	agggttaca	ggaggtgaag	3660
aaggacctga	tcacaaagct	gcccagctac	agcctgttcg	agctggagaa	cgcccgcaag	3720
cgcatgctgg	ccagcggcg	cgagctgcag	aaggcgaacg	agctggcctt	gcccagcaag	3780
tacgtgaact	tcctgtacct	ggccagccac	tacgagaagc	tgaaggcgag	ccccgaggac	3840

aacgagcaga agcagctggt cgtggagcag cacaagcact acctggacga gatcatcgag	3900
cagatcagcg agttcagcaa gcgcgtgatc ctggccgacg ccaacctgga caaggtgctg	3960
agcgcctaca acaagcaccg cgacaagccc atccgcgagc aggccgagaa catcatccac	4020
ctgttcaccc tgaccaacct gggcgcccc gccgccttca agtacttcca caccaccatc	4080
gaccgcaagc gctacaccag caccaaggag gtgctggacg ccaccctgat ccaccagagc	4140
atcaccggtc tgtacgagac ccgcatcgac ctgagccagc tggcgggcga ctaa	4194
<210> 108	
<211> 4107	
<212> DNA	
<213> 人工序列	
<220>	
<223> 化脓性链球菌的Cas9-编码序列	
<400> 108	
atggataaga aatactcaat aggettagat atcggcacaa atagcgtcgg atgggcggtg	60
atcactgatg aatataaggt tccgtctaaa aagttcaagg ttctgggaaa tacagaccgc	120
cacagtatca aaaaaaatct tataggggct cttttatttg acagtggaga gacagcggaa	180
gcgactcgtc tcaaacggac agctcgtaga aggtatacac gtcggaagaa tcgtatttgt	240
tatctacagg agattttttc aaatgagatg gcgaaagtag atgatagttt ctttcatcga	300
cttgaagagt cttttttggt ggaagaagac aagaagcatg aacgtcatcc tatttttggg	360
aatatagtag atgaagttgc ttatcatgag aaatatccaa ctatctatca tctgcgaaaa	420
aaattggtag attctactga taaagcggat ttgcgcttaa tctatttggc cttagcgc	480
atgattaagt ttcgtggtca ttttttgatt gagggagatt taaatcctga taatagtgat	540
gtggacaaac tatttatcca gttggtacaa acctacaatc aattatttga agaaaacct	600
attaacgcaa gtggagtaga tgctaaagcg attccttctg cacgattgag taaatcaaga	660
cgattagaaa atctcattgc tcagctcccc ggtgagaaga aaaatggctt atttgggaat	720
ctcattgctt tgtcattggg tttgaccctc aattttaaat caaattttga tttggcagaa	780
gatgctaaat tacagctttc aaaagatact tacgatgatg atttagataa tttattggcg	840
caaattggag atcaatatgc tgatttgtt ttggcagcta agaatttatc agatgctatt	900
ttactttcag atatcctaag agtaaatact gaataacta aggctcccct atcagcttca	960
atgattaaac gctacgatga acatcatcaa gacttgactc ttttaaaagc tttagttcga	1020
caacaacttc cagaaaagta taaagaaatc tttttgatc aatcaaaaaa cggatatgca	1080
ggttatattg atgggggagc tagccaagaa gaattttata aatttatcaa accaatttta	1140
gaaaaaatgg atggtactga ggaattattg gtgaaactaa atcgtgaaga tttgctgcgc	1200
aagcaacgga cctttgacaa cggetctatt ccccatcaaa ttcacttggg tgagctgcat	1260
gctattttga gaagacaaga agacttttat ccatttttaa aagacaatcg tgagaagatt	1320
gaaaaaatct tgacttttcg aattccttat tatgttggtc cattggcgcg tggcaatagt	1380
cgttttgcat ggatgactcg gaagtctgaa gaaacaatta ccccatggaa ttttgaagaa	1440
gttgtcgata aagggtcttc agctcaatca tttattgaac gcatgacaaa ctttgataaa	1500
aatcttccaa atgaaaaagt actaccaaaa catagtttgc tttatgagta ttttacggtt	1560

tataacgaat	tgacaaaggt	caaatatgtt	actgaaggaa	tgcgaaaacc	agcattttctt	1620
tcaggatgaac	agaagaaagc	cattgttgat	ttactcttca	aaacaaatcg	aaaagtaacc	1680
gttaagcaat	taaaagaaga	ttatttcaaa	aaaatagaat	gttttgatag	tgttgaaatt	1740
tcaggagttg	aagatagatt	taatgcttca	ttaggtacct	accatgattt	gctaaaaatt	1800
attaaagata	aagatTTTTT	ggataatgaa	gaaaatgaag	atatcttaga	ggatattggt	1860
ttaacattga	ccttatttga	agatagggag	atgattgagg	aaagacttaa	aacatattgct	1920
cacctctttg	atgataaggt	gatgaaacag	cttaaactgc	gccgttatac	tggttgggga	1980
cgtttgtctc	gaaaattgat	taatggtatt	agggataagc	aatctggcaa	aacaatatta	2040
gatttttttga	aatcagatgg	ttttgccaat	cgcaatttta	tgacagctgat	ccatgatgat	2100
agtttgacat	ttaaagaaga	cattcaaaaa	gcacaagtgt	ctggacaagg	cgatagttta	2160
catgaacata	ttgcaaattt	agctggtagc	ctgctatta	aaaaaggtat	tttacagact	2220
gtaaaagttg	ttgatgaatt	ggtcaaagta	atggggcggc	ataagccaga	aaatatcggt	2280
attgaaatgg	cacgtgaaaa	tcagacaact	caaaagggcc	agaaaaattc	gcgagagcgt	2340
atgaaacgaa	tcgaagaagg	tatcaaagaa	ttaggaagtc	agattcttaa	agagcatcct	2400
gttgaaaata	ctcaattgca	aaatgaaaag	ctctatctct	attatctcca	aaatggaaga	2460
gacatgtatg	tggaccaaga	attagatatt	aatcgtttaa	gtgattatga	tgctgatcac	2520
attgttccac	aaagtctcct	taaagacgat	tcaatagaca	ataaggtctt	aacgcgttct	2580
gataaaaaatc	gtggtaaadc	ggataacggt	ccaagtgaag	aagtagtcaa	aaagatgaaa	2640
aactattgga	gacaacttct	aaacgccaag	ttaatcactc	aacgtaagtt	tgataattta	2700
acgaaagctg	aacgtggagg	tttgagtga	cttgataaag	ctggttttat	caaacgccaa	2760
ttggttgaaa	ctcgccaaat	cactaagcat	gtggcacaaa	ttttggatag	tcgcatgaat	2820
actaaatacg	atgaaaatga	taaacttatt	cgagaggtta	aagtgattac	cttaaaatct	2880
aaattagttt	ctgacttccg	aaaagatttc	caattctata	aagtacgtga	gattaacaat	2940
taccatcatg	cccatgatgc	gtatctaaat	gccgtcgttg	gaactgcttt	gattaagaaa	3000
tatccaaaac	ttgaatcggg	gtttgtctat	ggtgattata	aagtttatga	tgttcgtaaa	3060
atgattgcta	agtctgagca	agaaataggc	aaagcaaccg	caaaatattt	cttttactct	3120
aatatcatga	acttcttcaa	aacagaaatt	acacttgcaa	atggagagat	tcgcaaacgc	3180
cctctaatacg	aaactaatgg	ggaaactgga	gaaattgtct	gggataaagg	gcgagatttt	3240
gccacagtgc	gcaaagtatt	gtccatgcc	caagtcaata	ttgtcaagaa	aacagaagta	3300
cagacaggcg	gattctccaa	ggagtcaatt	ttaccaaaaa	gaaattcgga	caagcttatt	3360
gctcgtaaaa	aagactggga	tccaaaaaaa	tatggtggtt	ttgatagtcc	aacggtagct	3420
tattcagtcc	tagtggttgc	taaggtgga	aaagggaaat	cgaagaagtt	aaaatccggt	3480
aaagagttac	tagggatcac	aattatggaa	agaagttcct	ttgaaaaaaa	tccgattgac	3540
tttttagaag	ctaaaggata	taaggaagtt	aaaaaagact	taatcattaa	actacctaaa	3600
tatagtcttt	ttgagttaga	aaacggctgc	aaacggatgc	tggttagtgc	cggagaatta	3660
caaaaaggaa	atgagctggc	tctgccaagc	aaatatgtga	atTTTTTata	tttagctagt	3720
cattatgaaa	agttgaaggg	tagtccagaa	gataacgaac	aaaacaatt	gtttgtggag	3780
cagcataagc	attatttaga	tgagattatt	gagcaaatca	gtgaattttc	taagcgtggt	3840
atTTtagcag	atgccaattt	agataaagtt	cttagtgcat	ataacaaca	tagagacaaa	3900

ccaatacgtg aacaagcaga aaatattatt catttattta cgttgacgaa tcttgagct	3960
cccgctgctt ttaaataattt tgatacaaca attgatcgta aacgatatac gtctacaaaa	4020
gaagtttttag atgccactct tatccatcaa tccatcactg gtctttatga aacacgcatt	4080
gatttgagtc agctaggagg tgactaa	4107
<210> 109	
<211> 1368	
<212> PRT	
<213> 人工序列	
<220>	
<223> 来自化脓性链球菌的Cas9的氨基酸序列	
<400> 109	
Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val	
1 5 10 15	
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe	
20 25 30	
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile	
35 40 45	
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu	
50 55 60	
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys	
65 70 75 80	
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser	
85 90 95	
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys	
100 105 110	
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr	
115 120 125	
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp	
130 135 140	
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His	
145 150 155 160	
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro	
165 170 175	
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr	
180 185 190	
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala	
195 200 205	
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn	
210 215 220	



530	535	540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr		
545	550	555
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp		
	565	570
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly		
	580	585
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp		
	595	600
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr		
610	615	620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala		
625	630	635
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr		
	645	650
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp		
	660	665
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe		
675	680	685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe		
690	695	700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu		
705	710	715
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly		
	725	730
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly		
	740	745
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln		
	755	760
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile		
770	775	780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro		
785	790	795
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu		
	805	810
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg		
	820	825
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys		
835	840	845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg  
 850 855 860  
 Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys  
 865 870 875 880  
 Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys  
 885 890 895  
 Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp  
 900 905 910  
 Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr  
 915 920 925  
 Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp  
 930 935 940  
 Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser  
 945 950 955 960  
 Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg  
 965 970 975  
 Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val  
 980 985 990  
 Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe  
 995 1000 1005  
 Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys  
 1010 1015 1020  
 Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser  
 1025 1030 1035 1040  
 Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu  
 1045 1050 1055  
 Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile  
 1060 1065 1070  
 Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser  
 1075 1080 1085  
 Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly  
 1090 1095 1100  
 Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile  
 1105 1110 1115 1120  
 Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser  
 1125 1130 1135  
 Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly  
 1140 1145 1150  
 Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile

	1155	1160	1165	
	Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala			
	1170	1175	1180	
	Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys			
	1185	1190	1195	1200
	Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser			
	1205	1210	1215	
	Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr			
	1220	1225	1230	
	Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser			
	1235	1240	1245	
	Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His			
	1250	1255	1260	
	Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val			
	1265	1270	1275	1280
	Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys			
	1285	1290	1295	
	His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu			
	1300	1305	1310	
	Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp			
	1315	1320	1325	
	Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp			
	1330	1335	1340	
	Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile			
	1345	1350	1355	1360
	Asp Leu Ser Gln Leu Gly Gly Asp			
	1365			
<210>	110			
<211>	4221			
<212>	DNA			
<213>	人工序列			
<220>				
<223>	Cas9-编码序列在pET-Cas9N3T中,用于在大肠杆菌中产生 重组Cas9蛋白(人源化的密码子; 六His标签和细胞核定位信号在N末端)			
<400>	110			
	atgggcagca gccatcatca tcatcatcat gtgtaccct acgacgtgcc cgactagcc			60
	gaattgcctc caaaaaagaa gagaaaggta gggatcgaga acctgtactt ccagggcgac			120
	aagaagtaca gcatcggcct ggacatcggg accaacagcg tgggctgggc cgtgatcacc			180

gacgagtaca	aggtgcccag	caagaagttc	aaggtgctgg	gcaacaccga	ccgccacagc	240
atcaagaaga	acctgatcgg	cgccctgctg	ttgacagcgc	gcgagaccgc	cgaggccacc	300
cgcctgaagc	gcaccgcccg	ccgccgctac	accgcgcgca	agaaccgcat	ctgctacctg	360
caggagatct	tcagcaacga	gatggccaag	gtggacgaca	gcttcttcca	ccgcctggag	420
gagagcttcc	tgggtggagga	ggacaagaag	cacgagcgcc	accccatctt	cggcaacatc	480
gtggacgagg	tggcctacca	cgagaagtac	cccaccatct	accacctgcg	caagaagctg	540
gtggacagca	ccgacaaggc	cgacctgctc	ctgatctacc	tggccctggc	ccacatgatc	600
aagttccgcg	gccacttctt	gatcgagggc	gacctgaacc	ccgacaacag	cgacgtggac	660
aagctgttca	tccagctggt	gcagacctac	aaccagctgt	tcgaggagaa	ccccatcaac	720
gccagcggcg	tggacgcaa	ggccatctct	agcgcgccgc	tgagcaagag	ccgccgcctg	780
gagaacctga	tcgcccagct	gcccggcgag	aagaagaacg	gcctgttcgg	caacctgatc	840
gccctgagcc	tgggctgac	ccccaaactt	aagagcaact	tcgacctggc	cgaggacgcc	900
aagctgcagc	tgagcaagga	cacctacgac	gacgacctgg	acaacctgct	ggcccagatc	960
ggcgaccagt	acgccgacct	gttcttgccc	gccaagaacc	tgagcgacgc	catcttgctg	1020
agcgacatcc	tgcgctgtaa	caccgagatc	accaaggccc	ccctgagcgc	cagcatgatc	1080
aagcgctacg	acgagcacca	ccaggacctg	accctgctga	aggccctggt	gcgccagcag	1140
ctgcccgaga	agtacaagga	gatcttcttc	gaccagagca	agaacggcta	cgccggctac	1200
atcgacggcg	gcgccagcca	ggaggagttc	tacaagttca	tcaagcccat	cctggagaag	1260
atggacggca	ccgaggagct	gctggtgaa	ctgaaccgcg	aggacctgct	gcgcaagcag	1320
cgcaccttcg	acaacggcag	catccccac	cagatccacc	tgggcgagct	gcacgccatc	1380
ctgcgccgcc	aggaggactt	ctacccttc	ctgaaggaca	accgcgagaa	gatcgagaag	1440
atcctgacct	tccgcatccc	ctactacgtg	ggccccctgg	cccgcggcaa	cagccgcttc	1500
gcctggatga	cccgcaagag	cgaggagacc	atccccctt	ggaacttcga	ggaggtggtg	1560
gacaagggcg	ccagcgccca	gagcttcac	gagcgcatga	ccaacttcga	caagaacctg	1620
cccaacgaga	aggtgctgcc	caagcacagc	ctgctgtacg	agtacttcac	cgtgtacaac	1680
gagctgacca	aggtgaagta	cgtgaccgag	ggcatgcgca	agcccgcctt	cctgagcggc	1740
gagcagaaga	aggccatcgt	ggacctgctg	ttcaagacca	accgcaaggt	gaccgtgaag	1800
cagctgaagg	aggactactt	caagaagatc	gagtgccttc	acagcgtgga	gatcagcggc	1860
gtggaggacc	gcttcaacgc	cagcctgggc	acctaccagc	acctgctgaa	gatcatcaag	1920
gacaaggact	tcctggacaa	cgaggagaac	gaggacatcc	tggaggacat	cgtgctgacc	1980
ctgaccctgt	tcgaggaccg	cgagatgatc	gaggagcgc	tgaagacct	cgcccacctg	2040
ttcgacgaca	aggtgatgaa	gcagctgaag	cgccgccgct	acaccggctg	gggccgcctg	2100
agccgcaagc	ttatcaacgg	catccgcgac	aagcagagcg	gcaagaccat	cctggacttc	2160
ctgaagagcg	acggcttcgc	caaccgcaac	ttcatgcagc	tgatccacga	cgacagcctg	2220
accttcaagg	aggacatcca	gaaggcccag	gtgagcggcc	agggcgacag	cctgcacgag	2280
cacatcgcca	acctggccgg	cagccccgcc	atcaagaagg	gcatcctgca	gaccgtgaag	2340
gtggtggacg	agctggtgaa	ggtgatgggc	cgccacaagc	ccgagaacat	cgtgatcgag	2400
atggcccgcg	agaaccagac	caccagaag	ggccagaaga	acagccgca	gcgcatgaag	2460
cgcatcgagg	aggcatcaa	ggagctgggc	agccagatcc	tgaaggagca	ccccgtggag	2520

aacacccagc tgcagaacga gaagctgtac ctgtactacc tgcagaacgg ccgcgacatg	2580
tacgtggacc aggagctgga catcaaccgc ctgagcgact acgacgtgga ccacatcgtg	2640
ccccagagct tcctgaagga cgacagcatc gacaacaagg tgctgacccg cagcgacaag	2700
aaccgcggca agagcgacaa cgtgcccagc gaggaggtgg tgaagaagat gaagaactac	2760
tggcgcccagc tgctgaacgc caagctgata acccagcgcga agttcgacaa cctgaccaag	2820
gccgagcgcg gcgccctgag cgagctggac aaggccggct tcatcaagcg ccagctgggtg	2880
gagacccgcc agatcaccaa gcacgtggcc cagatcctgg acagccgcat gaacaccaag	2940
tacgacgaga acgacaagct gatccgcgag gtgaaggtga tcaccctgaa gagcaagctg	3000
gtgagcgact tccgcaagga cttccagttc tacaaggtgc gcgagatcaa caactaccac	3060
cacgccccag acgcctacct gaacgccgtg gtgggcaccg ccctgatcaa gaagtacccc	3120
aagctggaga gcgagttcgt gtacggcgac tacaaggtgt acgacgtgcg caagatgata	3180
gccaagagcg agcaggagat cggcaaggcc accgccaagt actttcttcta cagcaacatc	3240
atgaacttct tcaagaccga gatcacctg gccaacggcg agatccgcaa gcgccccctg	3300
atcgagacca acggcgagac cggcgagatc gtgtgggaca agggccgcga cttcgccacc	3360
gtgcgcaagg tgctgagcat gccccaggtg aacatcgtga agaagaccga ggtgcagacc	3420
ggcggcttca gcaaggagag catcctgccc aagcgaaca gcgacaagct gatcgccccg	3480
aagaaggact gggaccccaa gaagtacggc ggcttcgaca gccccaccgt ggcctacagc	3540
gtgctgggtg tggccaaggt ggagaagggc aagagcaaga agctgaagag cgtgaaggag	3600
ctgctgggca tcaccatcat ggagcgcagc agcttcgaga agaaccctat cgacttctg	3660
gaggccaagg gctacaagga ggtgaagaag gacctgatca tcaagctgcc caagtacagc	3720
ctgttcgagc tggagaacgg ccgcaagcgc atgctggcca gcgccggcga gctgcagaag	3780
ggcaacgagc tggccctgcc cagcaagtac gtgaacttcc tgtacctggc cagccactac	3840
gagaagctga agggcagccc cgaggacaac gagcagaagc agctgttcgt ggagcagcac	3900
aagcactacc tggacgagat catcgagcag atcagcgagt tcagcaagcg cgtgatcctg	3960
gccgacgcca acctggacaa ggtgctgagc gcctacaaca agcaccgcga caagccctac	4020
cgcgagcagg ccgagaacat catccacctg ttcacctga ccaacctggg cgccccgcc	4080
gccttcaagt acttcgacac caccatcgac cgcaagcgt acaccagcac caaggagtg	4140
ctggacgcca ccctgatcca ccagagcatc accggtctgt acgagaccg catcgacctg	4200
agccagctgg gcgcgacta a	4221

<210> 111

<211> 1406

<212> PRT

<213> 人工序列

<220>

<223> Cas9 (pET-Cas9N3T) 的氨基酸序列

<400> 111

Met Gly Ser Ser His His His His His Val Tyr Pro Tyr Asp Val

1                    5                    10                    15

Pro Asp Tyr Ala Glu Leu Pro Pro Lys Lys Lys Arg Lys Val Gly Ile



Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys  
340 345 350

Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln  
355 360 365

Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys  
370 375 380

Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr  
385 390 395 400

Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro  
405 410 415

Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn  
420 425 430

Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile  
435 440 445

Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln  
450 455 460

Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys  
465 470 475 480

Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly  
485 490 495

Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr  
500 505 510

Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser  
515 520 525

Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys  
530 535 540

Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn  
545 550 555 560

Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala  
565 570 575

Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys  
580 585 590

Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys  
595 600 605

Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg  
610 615 620

Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys  
625 630 635 640

Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp

	645	650	655
Ile Val Leu Thr	Leu Thr Leu Phe	Glu Asp Arg Glu Met	Ile Glu Glu
	660	665	670
Arg Leu Lys Thr Tyr Ala His	Leu Phe Asp Asp Lys Val	Met Lys Gln	
	675	680	685
Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg	Leu Ser Arg Lys Leu		
	690	695	700
Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr	Ile Leu Asp Phe		
705	710	715	720
Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln	Leu Ile His		
	725	730	735
Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser			
	740	745	750
Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser			
	755	760	765
Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu			
	770	775	780
Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu			
785	790	795	800
Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg			
	805	810	815
Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln			
	820	825	830
Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys			
	835	840	845
Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln			
	850	855	860
Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His Ile Val			
865	870	875	880
Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu Thr			
	885	890	895
Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu Glu			
	900	905	910
Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys			
	915	920	925
Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly			
	930	935	940
Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val			
945	950	955	960

Glu Thr Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg		
	965	970 975
Met Asn Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys		
	980	985 990
Val Ile Thr Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe		
	995	1000 1005
Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp		
	1010	1015 1020
Ala Tyr Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro		
	1025	1030 1035 1040
Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val		
	1045	1050 1055
Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala		
	1060	1065 1070
Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile		
	1075	1080 1085
Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn		
	1090	1095 1100
Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr		
	1105	1110 1115 1120
Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr		
	1125	1130 1135
Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg		
	1140	1145 1150
Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys		
	1155	1160 1165
Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val		
	1170	1175 1180
Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu		
	1185	1190 1195 1200
Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro		
	1205	1210 1215
Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu		
	1220	1225 1230
Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg		
	1235	1240 1245
Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu		
	1250	1255 1260
Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr		

1265	1270	1275	1280
Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe			
	1285	1290	1295
Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser			
	1300	1305	1310
Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val			
	1315	1320	1325
Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala			
	1330	1335	1340
Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala			
1345	1350	1355	1360
Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser			
	1365	1370	1375
Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly			
	1380	1385	1390
Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp			
	1395	1400	1405

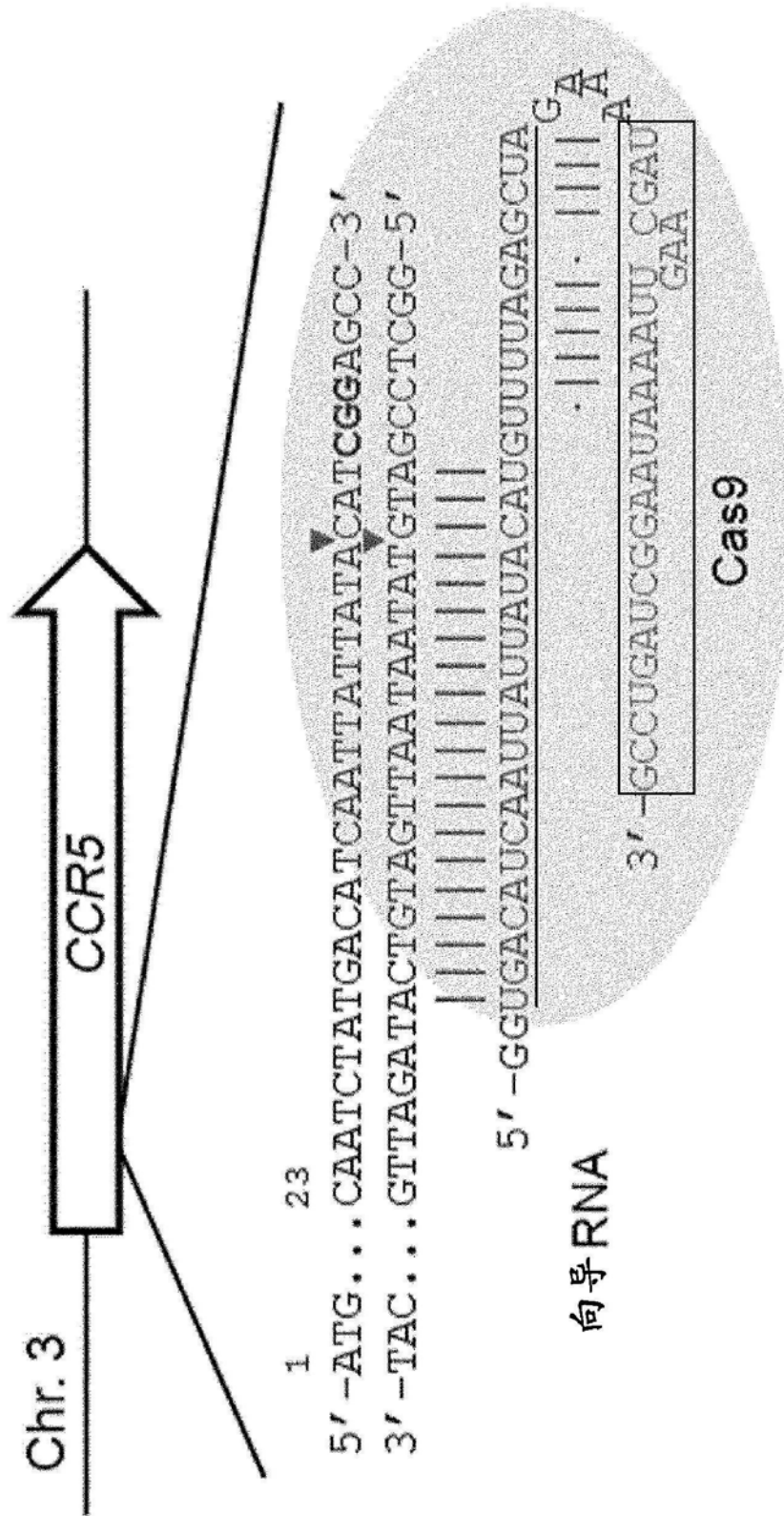


图1a

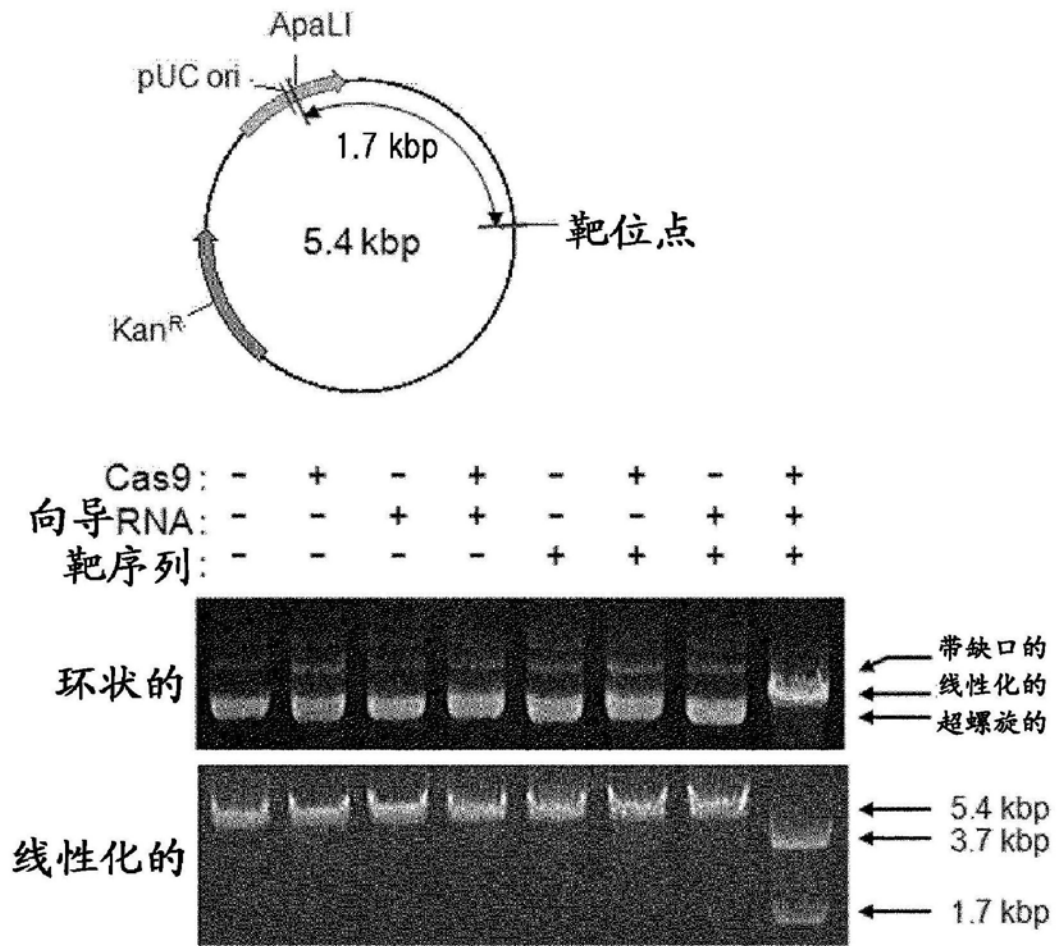


图1b

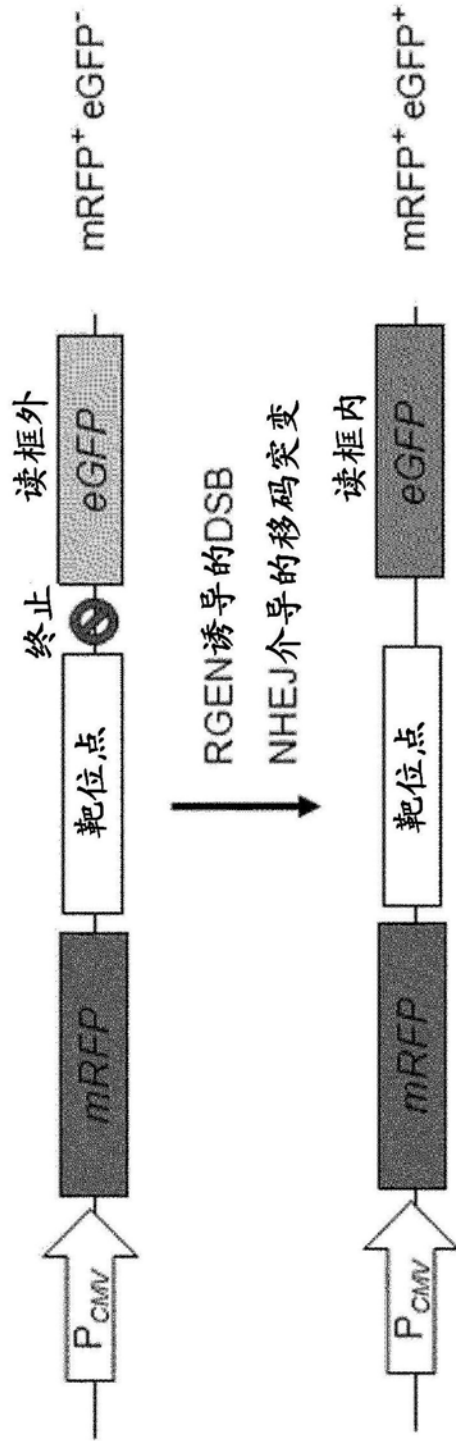


图2a

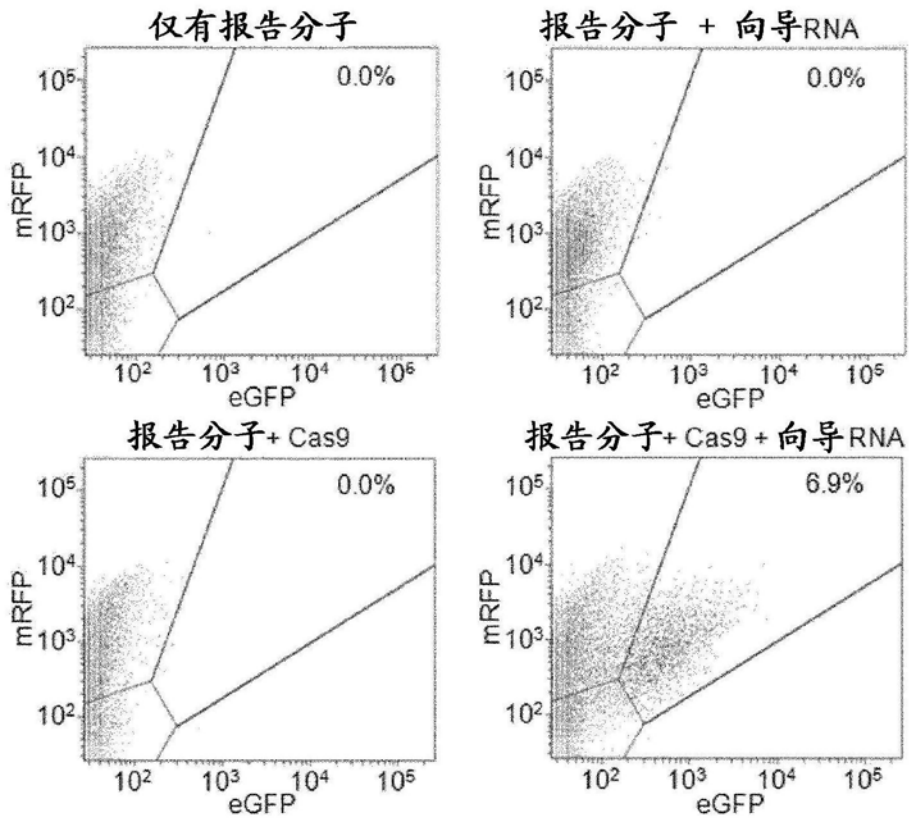


图2b

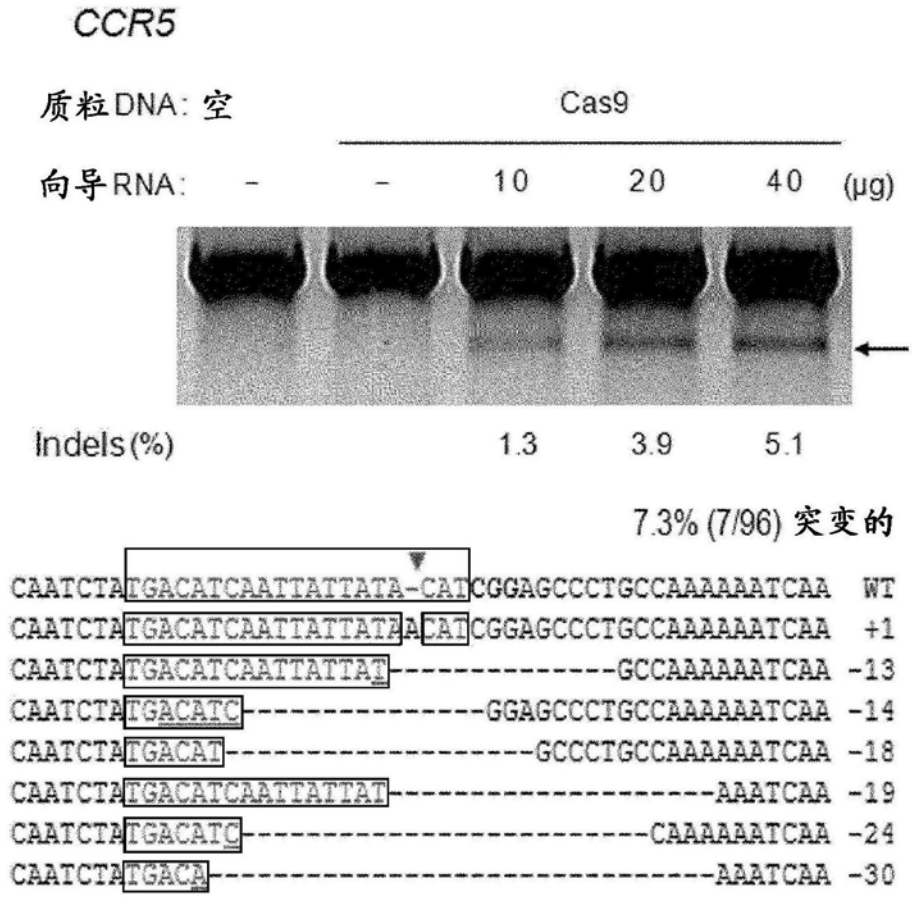


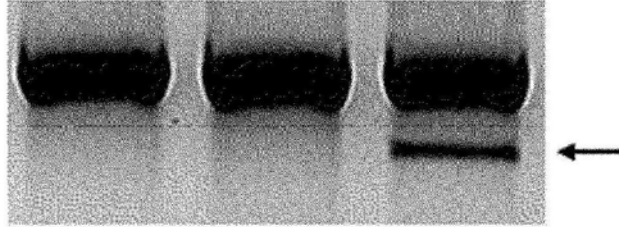
图3a

**C4BPB**

质粒 DNA: 空

Cas9

向导 RNA: - - +



Indels (%)

14

8.3% (4/48) 突变的

```

TATGTGC AATGACCACTACATCCT--CAAGGGCAGCAATCGGAGCCAG WT
TATGTGC AATGACCACTACATCCTT-CAAGGGCAGCAATCGGAGCCAG +1
TATGTGC AATGACCACTACATCCTCTCAAGGGCAGCAATCGGAGCCAG +2
TATGTGC AATG-----GAGCCAG -30
TATGTGC AATGAC-----180 bp----- -180
    
```

图3b

```

TGACATCAATTATTATACATCGG CCR5
TGACATCAATTATTATAgATgGa ADCY5
TGACATCAcTTATTATgCATgGG KCNJ6
TGACATaAATTATTcTACATgGG CNTNAP2
TGAaATCAATTATcATAgATCGG Chr. 5 N/A
    
```

图4a

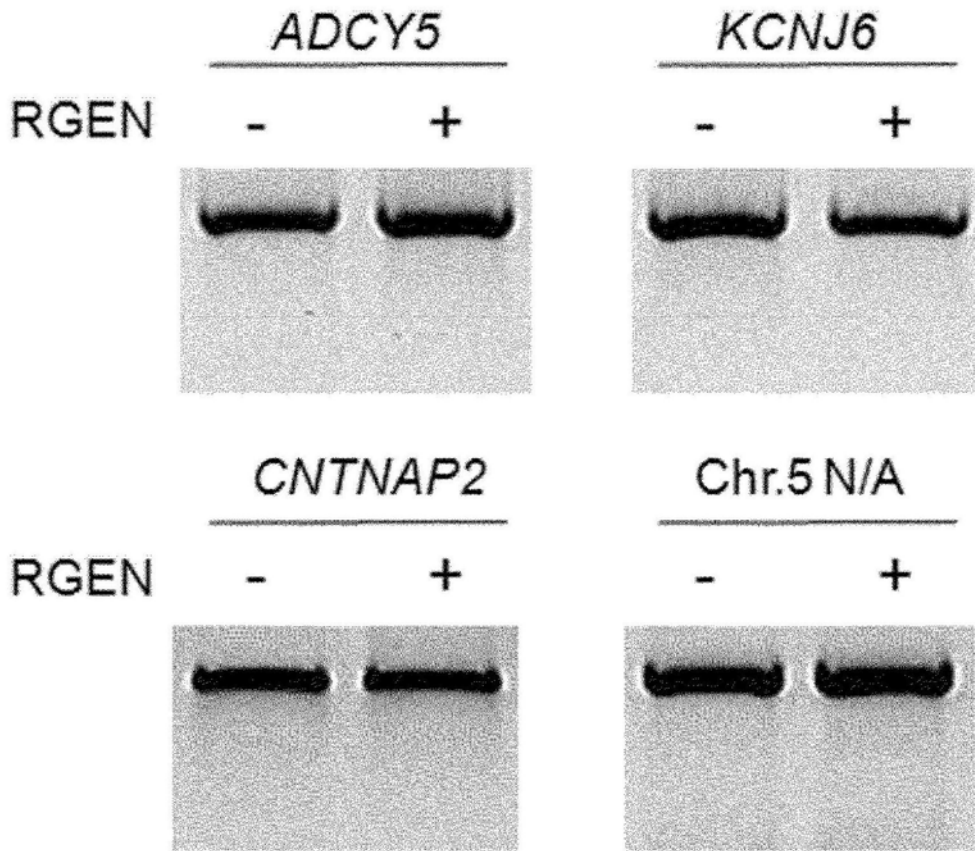


图4b

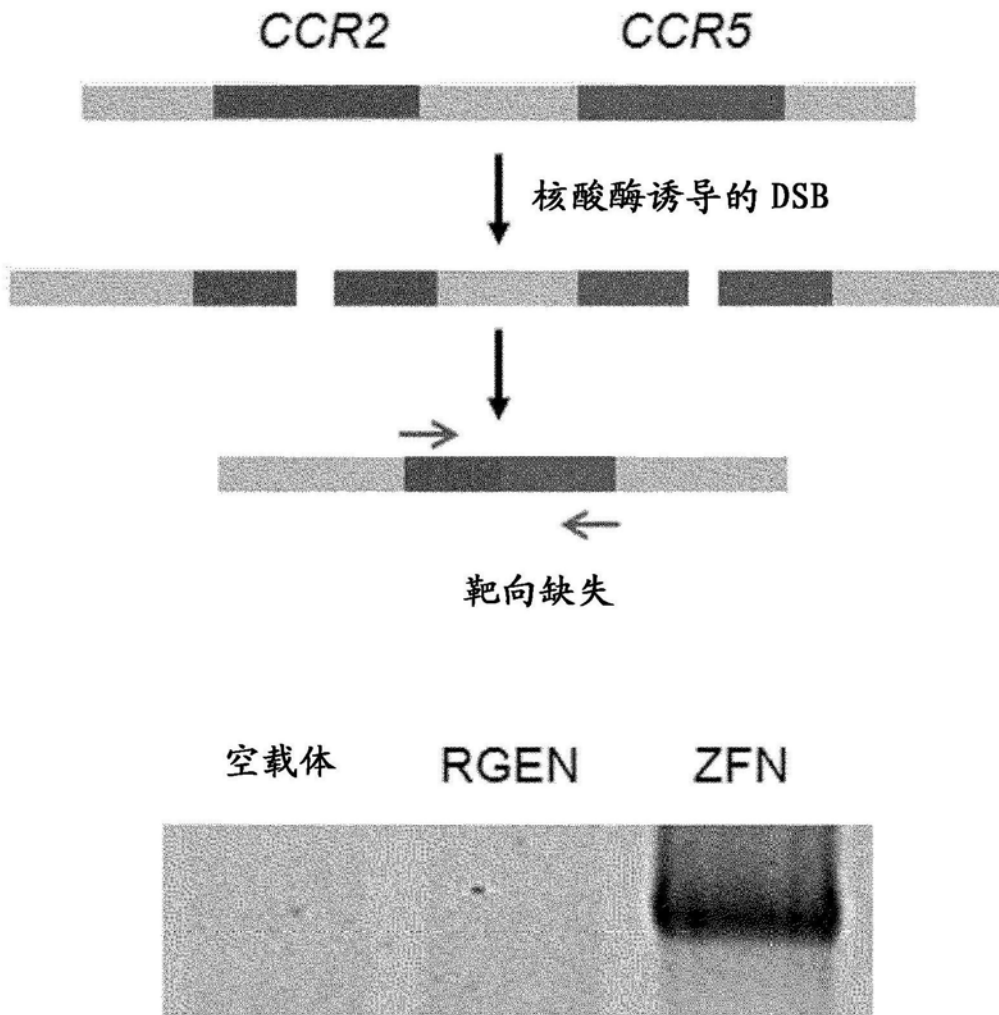


图4c

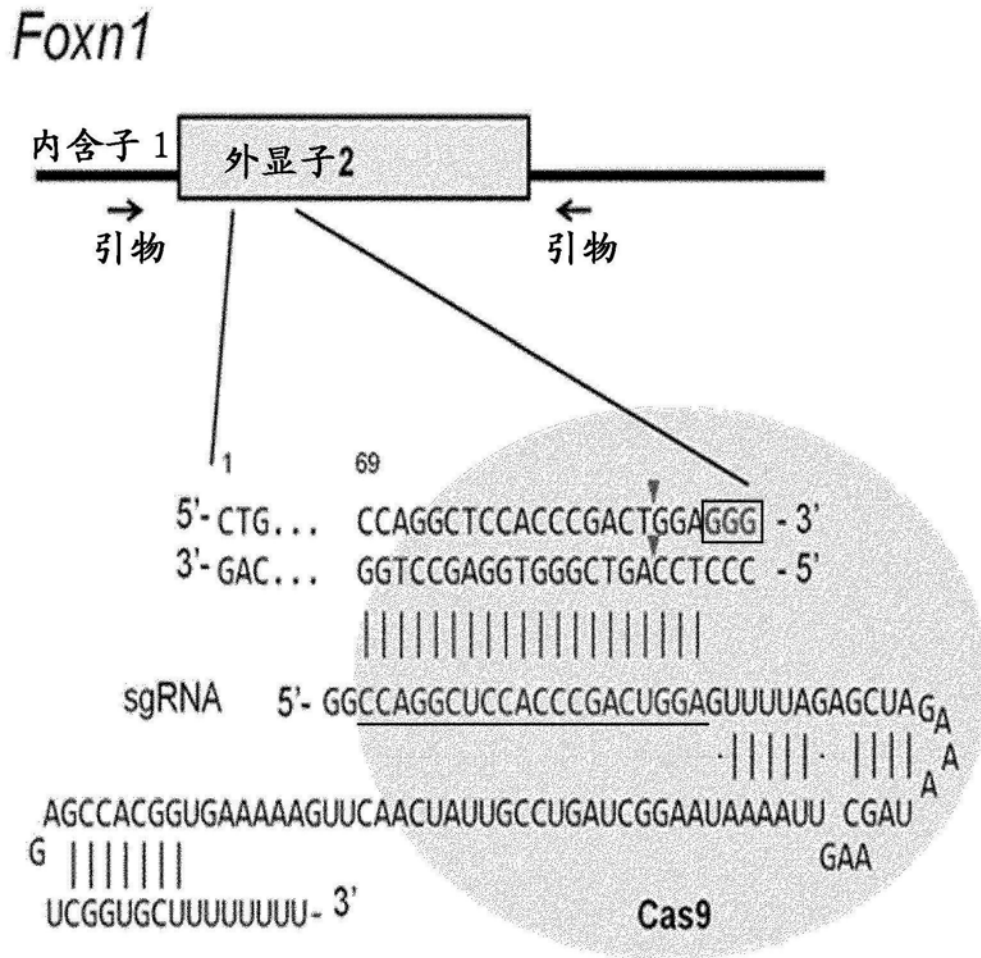


图5a

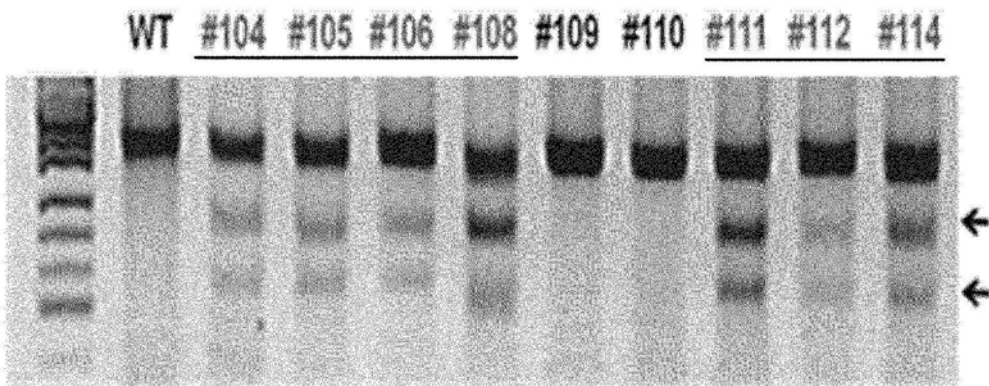


图5b

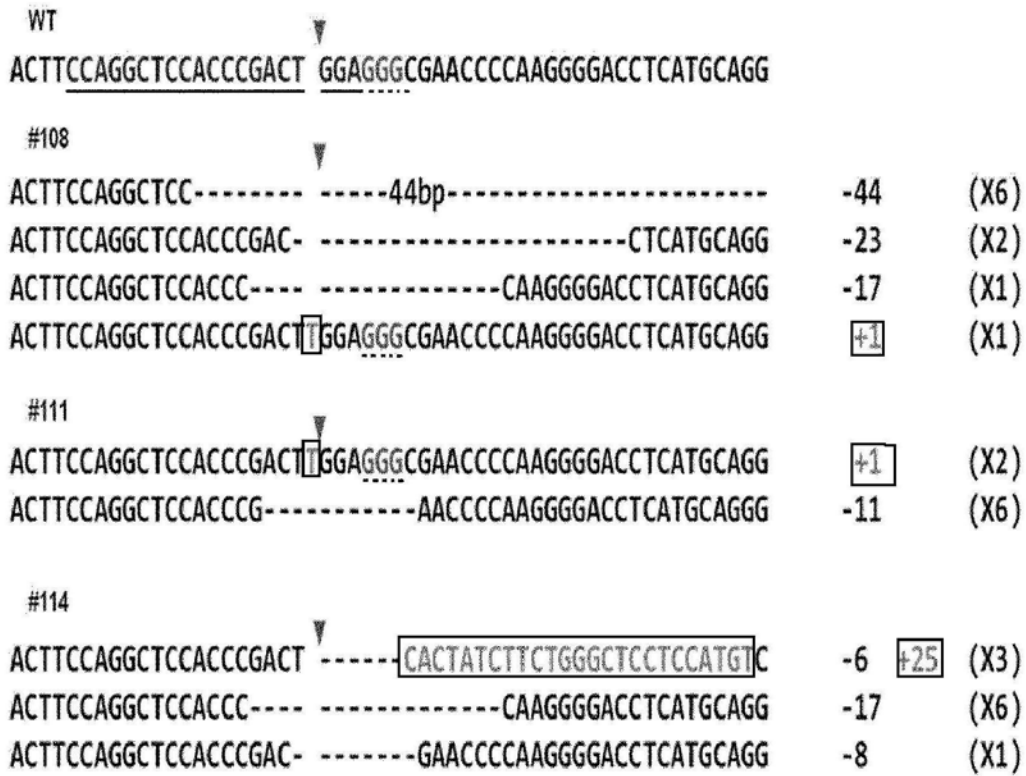


图5c

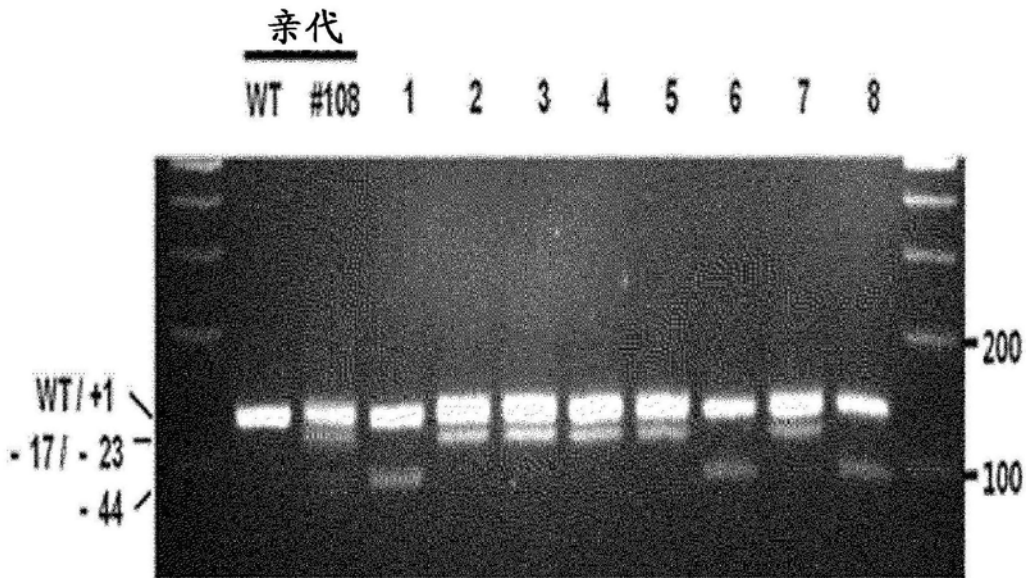


图5d

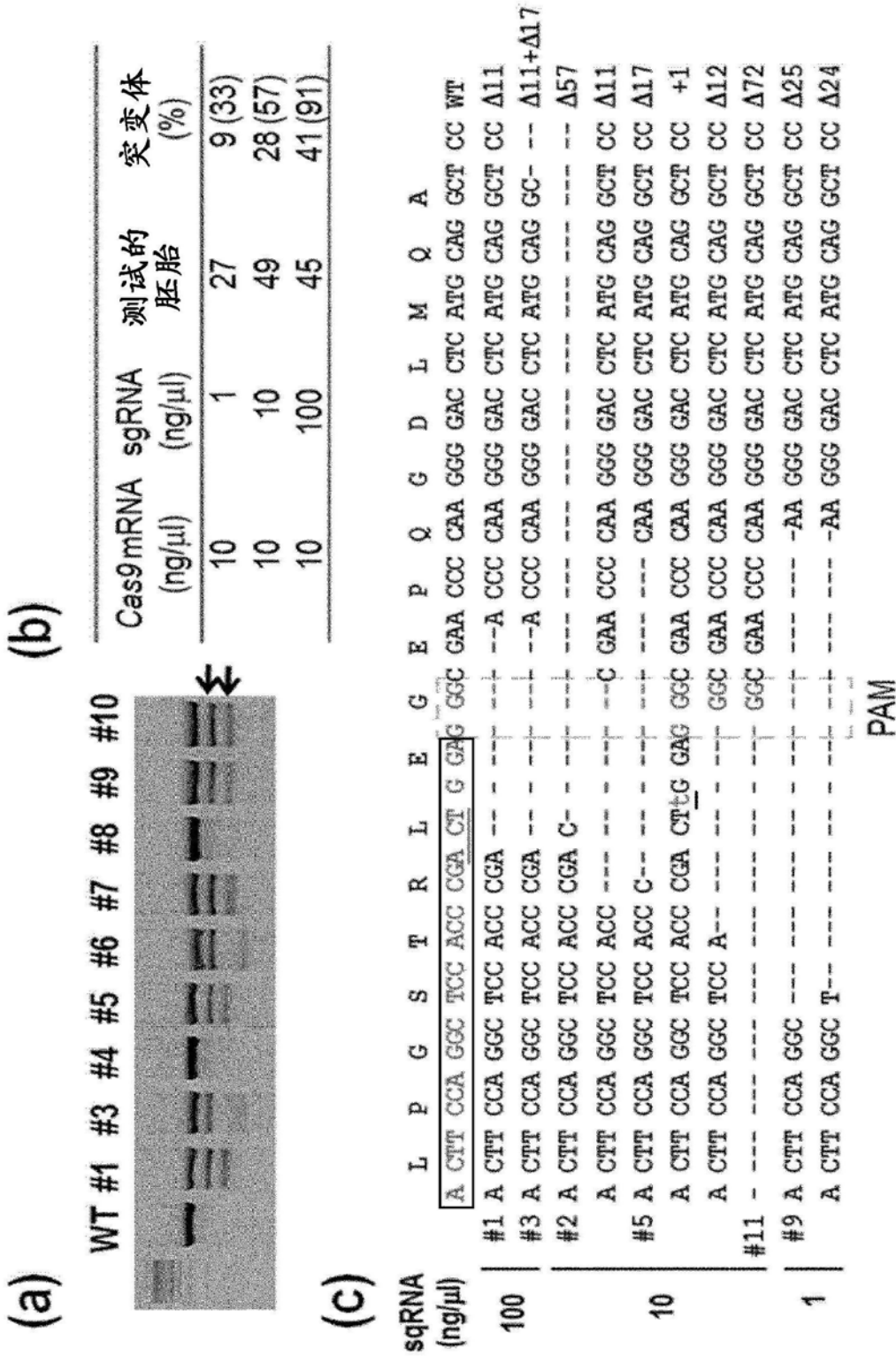


图6

原核注射

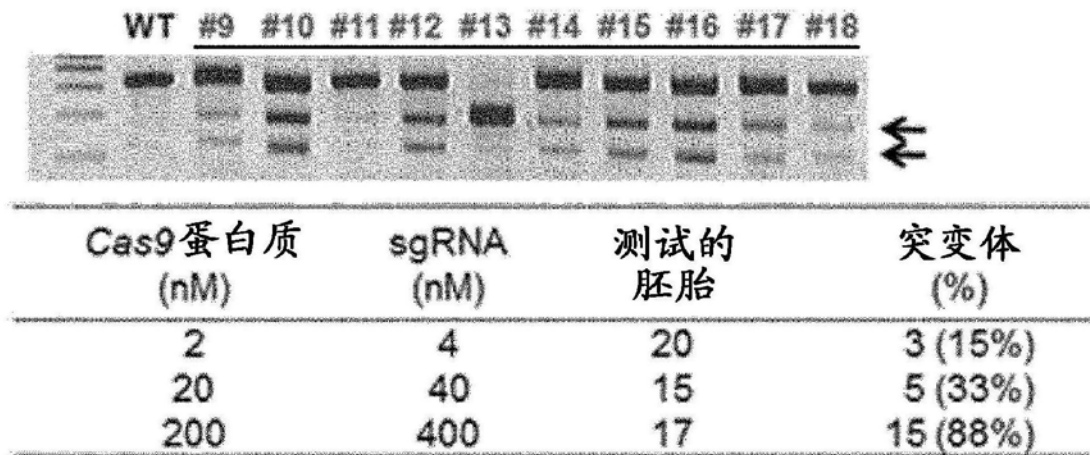


图7a

胞质内注射

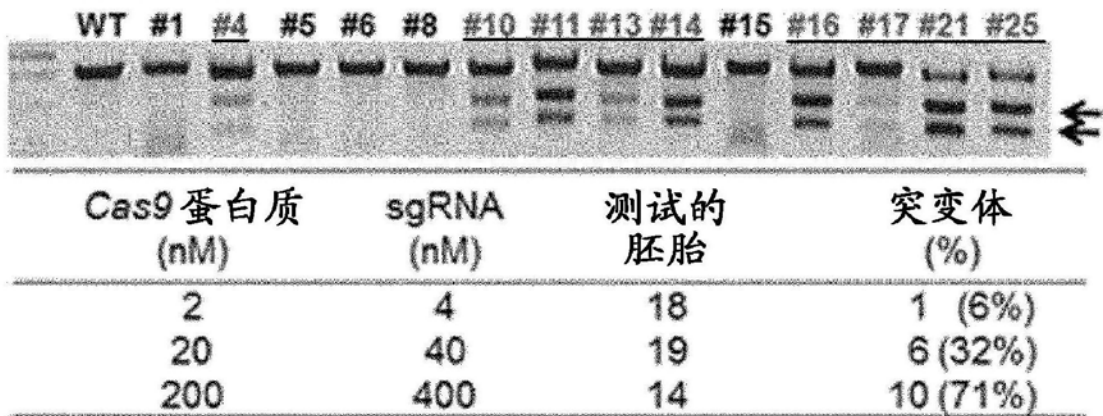


图7b

序列	Indels	胚胎数
ACTTCCAGGCTCCACCCGACTGGA	WT	
ACTTCCAGGCTCCACCC-----AAGGGACCTCATGCAG	Δ18	2
ACTTCCAGGCTCCAC-----AAGGGACCTCATGCAG	Δ20	1
ACTTCCAGGCTCCACCC-----AAGGGACCTCATGCC	Δ19	1
ACTTCCAGGCTCCACCC-----CAAGGGACCTCATGCAG	Δ17	1
ACTTCCAGGCTCCACCCGA-----ACCCAAAGGGACCTCATGCAG	Δ11	3
ACTTCCAGGCTCCACCCGAA--GGAGGGGAACCCCAAGGGACCTCATGCAG	Δ3+1	1
ACTTCCAGGCTCCACCCGACT--AGGGGAACCCCAAGGGACCTCATGCAG	Δ2	1
ACTTCCAGGCTCCACCCGACTGGAGGGGAACCCCAAGGGACCTCATGCAG	+1	1
ACTTCCAGGCTCCACCCGACTTGGAGGGGAACCCCAAGGGACCTCATGCAG	+1	10
ACTTCCAGGCTCCACCCGA-----GGGAACCCCAAGGGACCTCATGCAG	Δ6	1
ACTTCCAGGCTCCACCCGA-----GGGGAAACCCCAAGGGACCTCATGCAG	Δ5	2
ACTTCCAGGCTCCACCC-----TCATGCAG	Δ28	1
-----AGGGGAACCCCAAGGGACCTCATGCAG	Δ126	1
总计		26

图7c

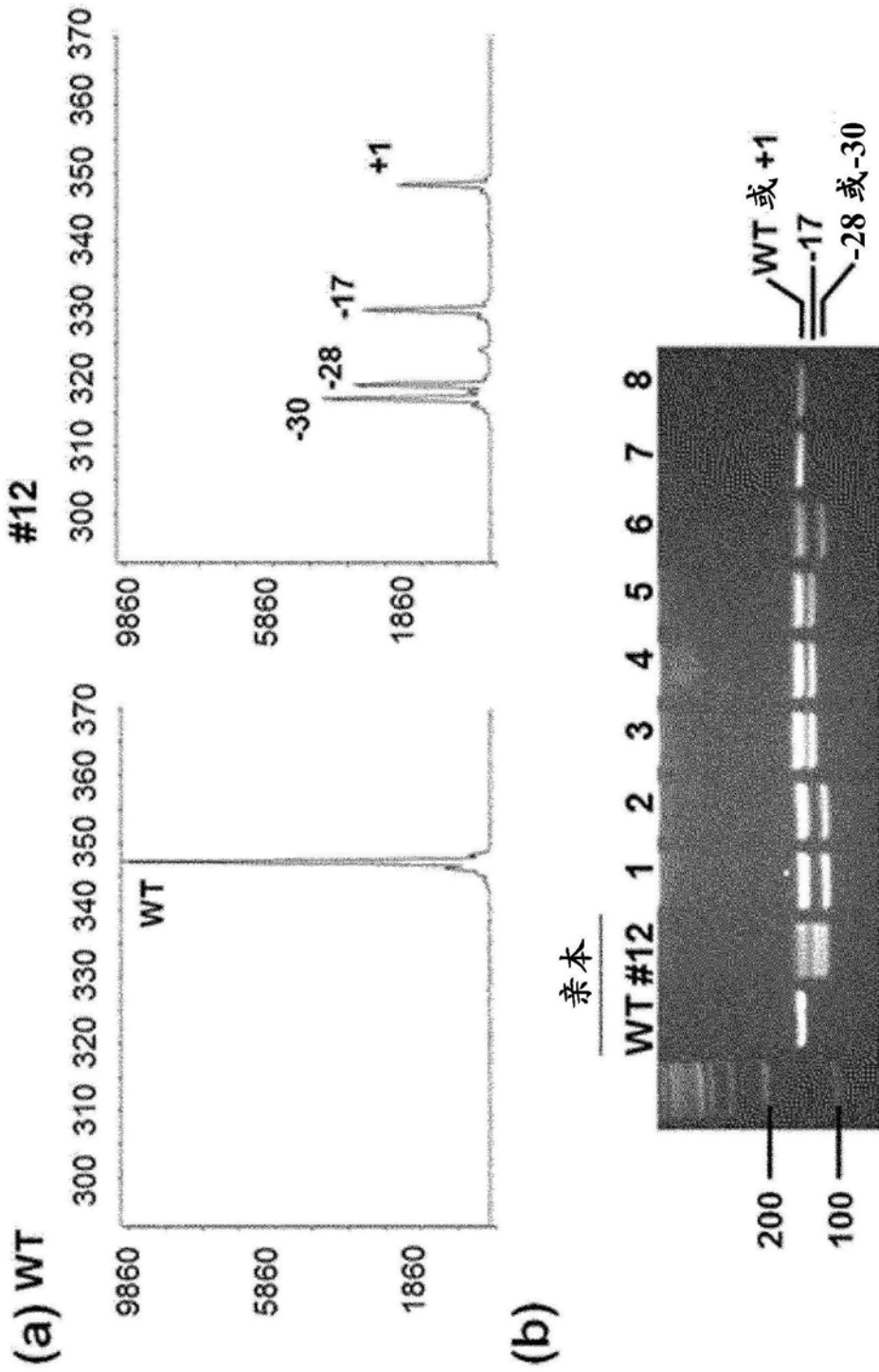


图8

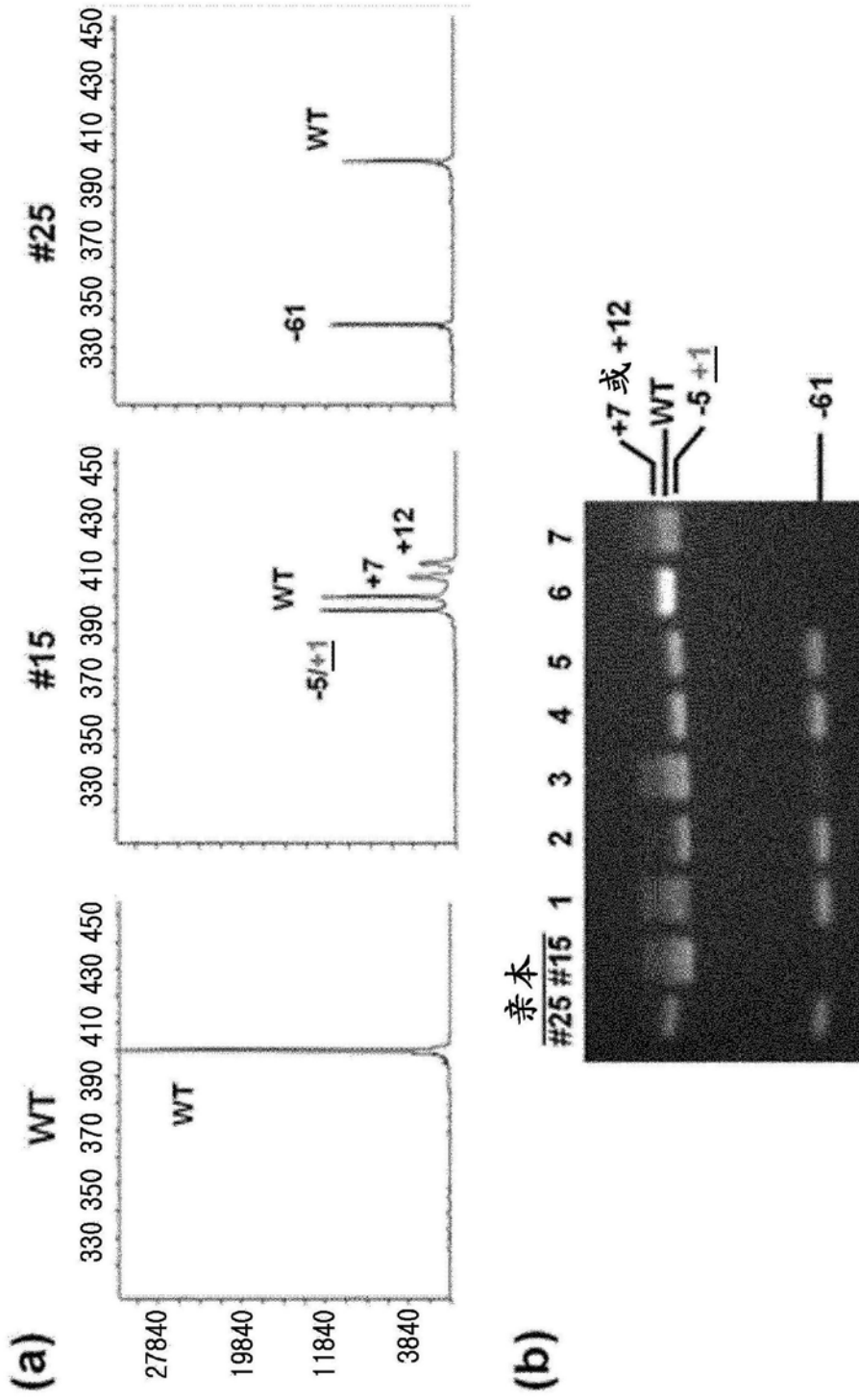


图9

**CCR5#4**

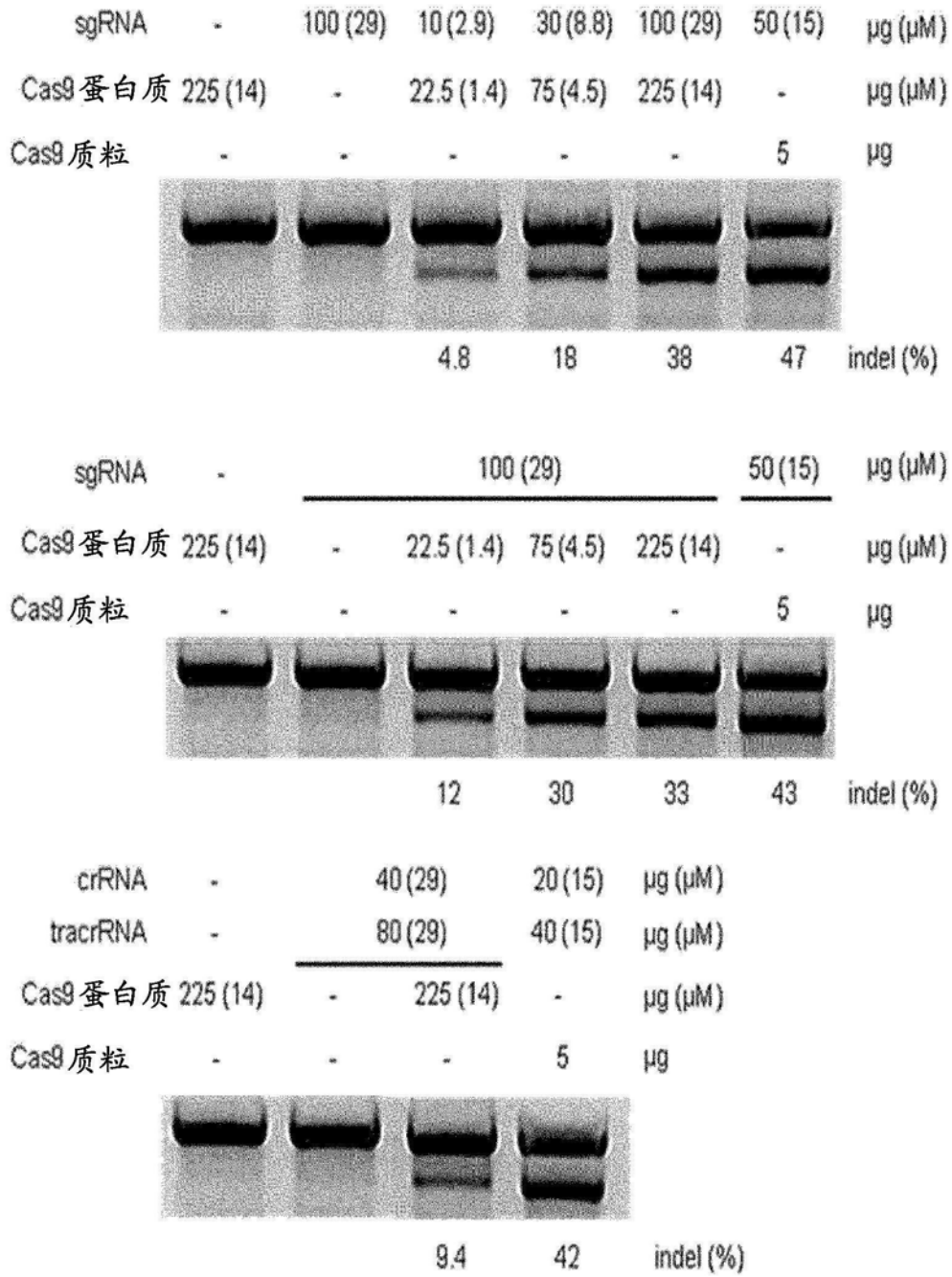


图10a







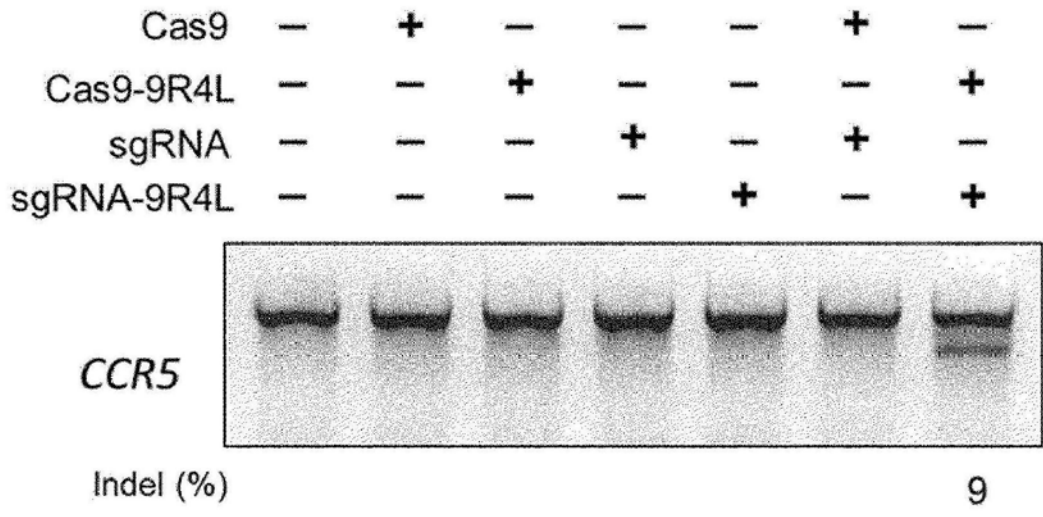


图13

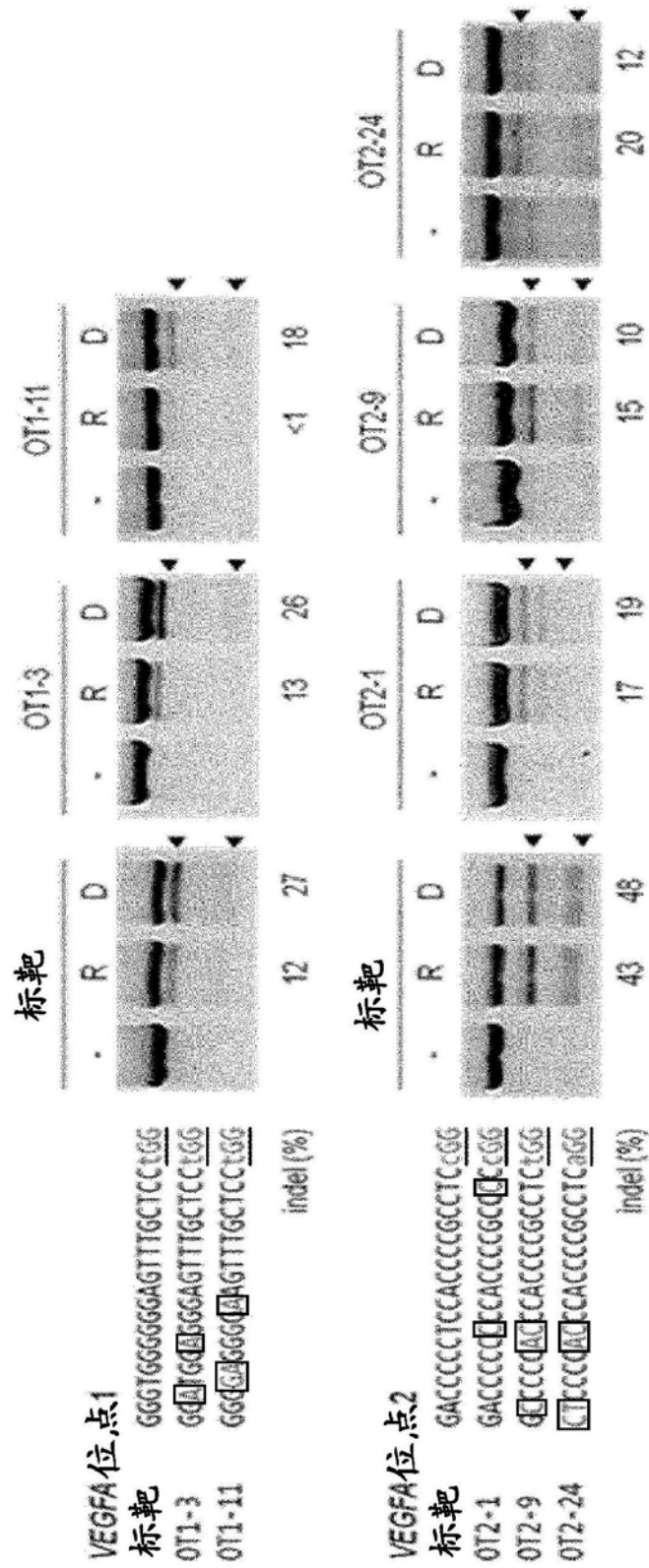


图14a

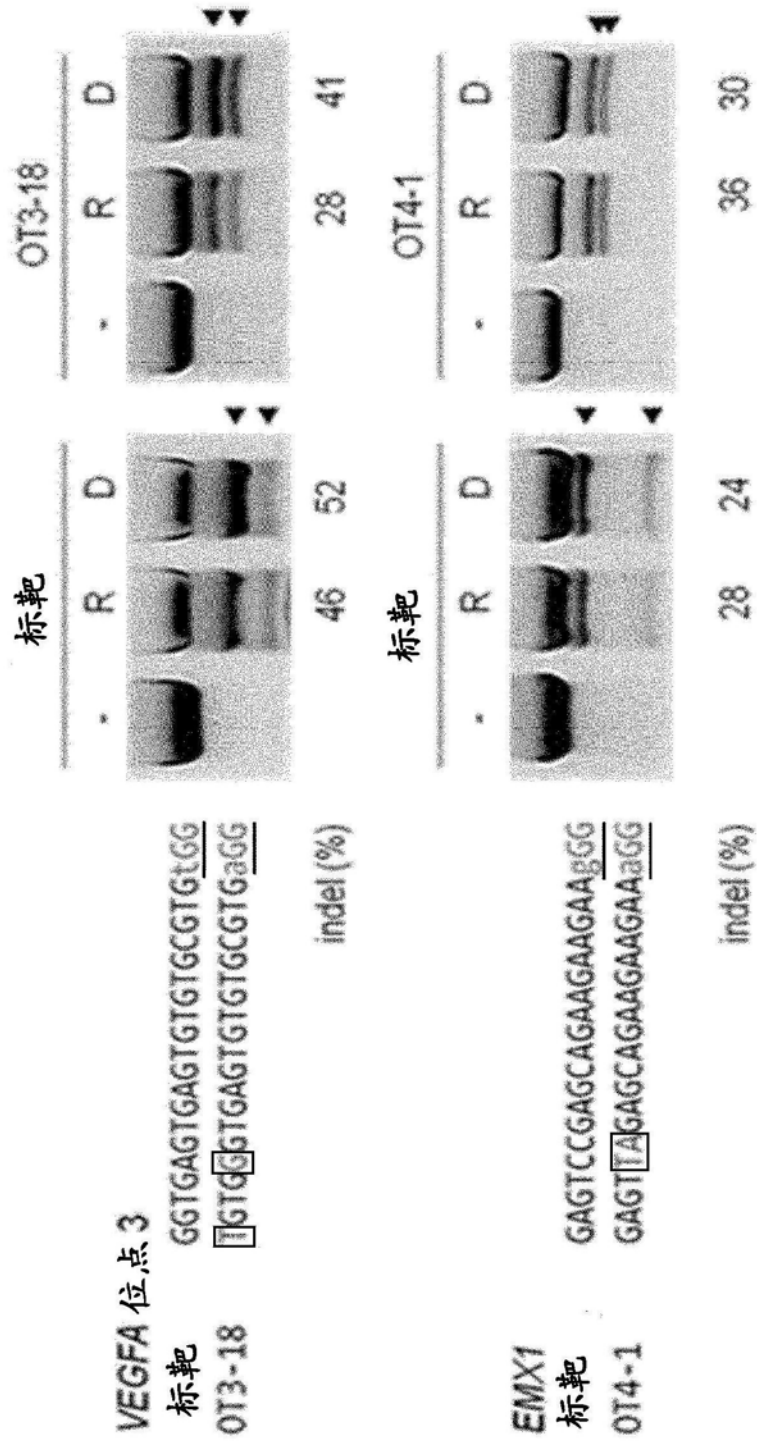


图14b



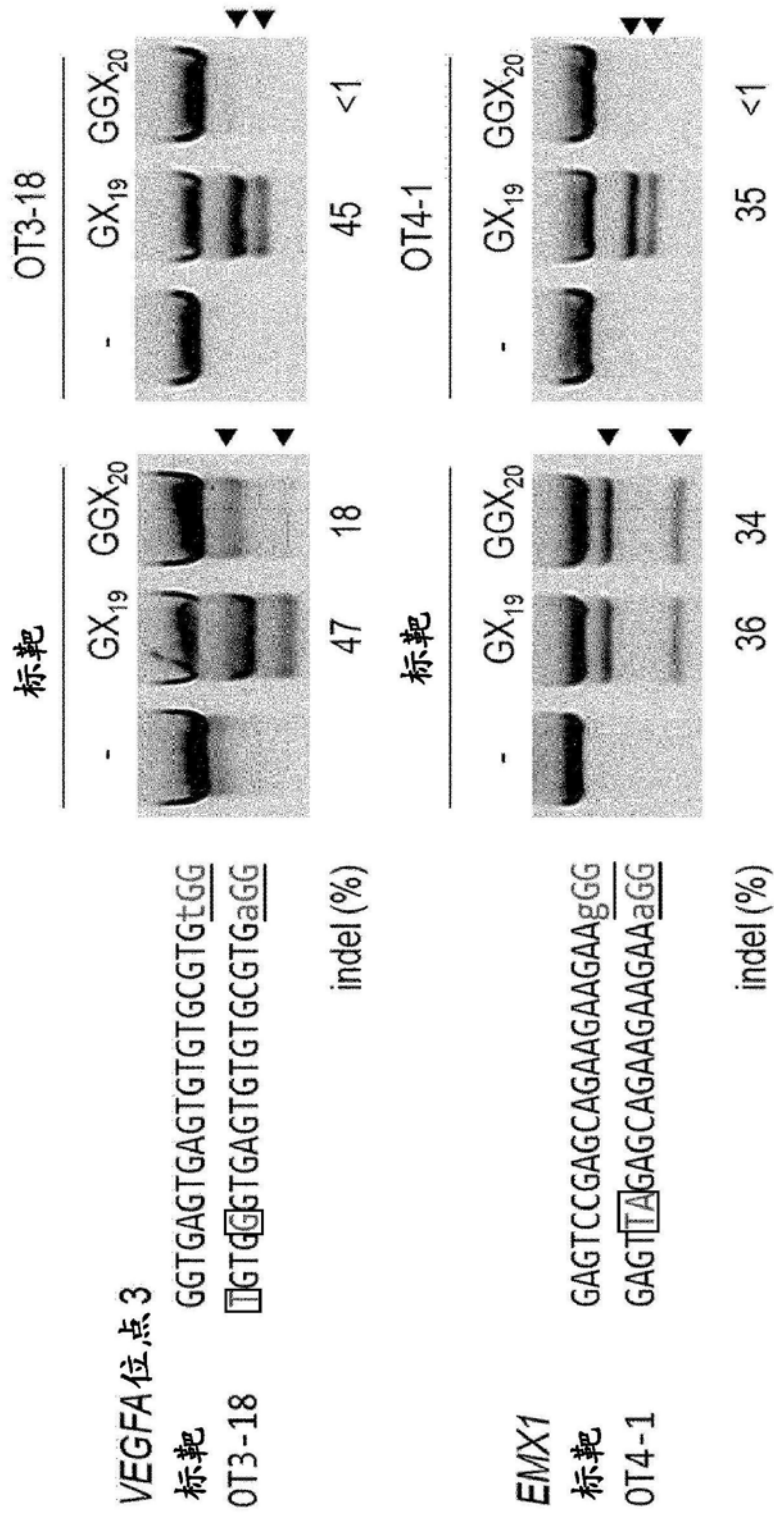


图15b

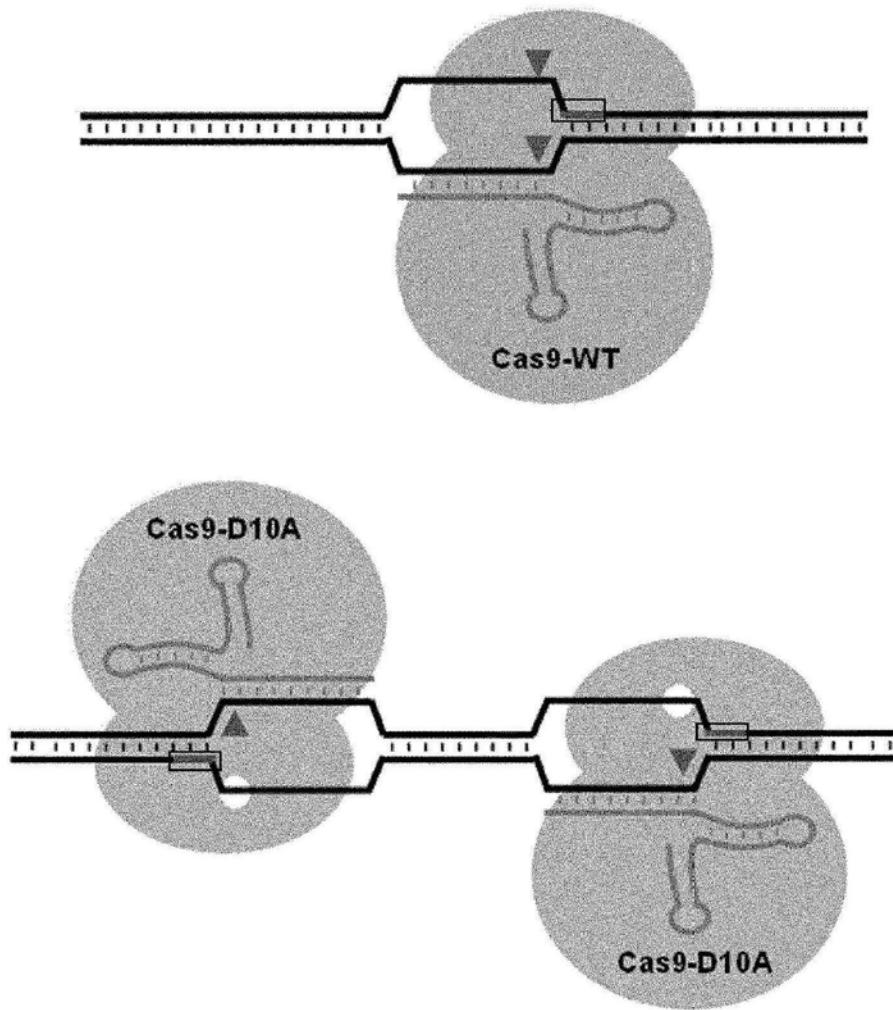


图16a

AAVS1基因座

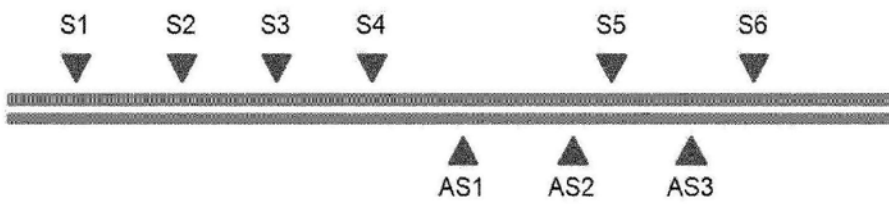


图16b

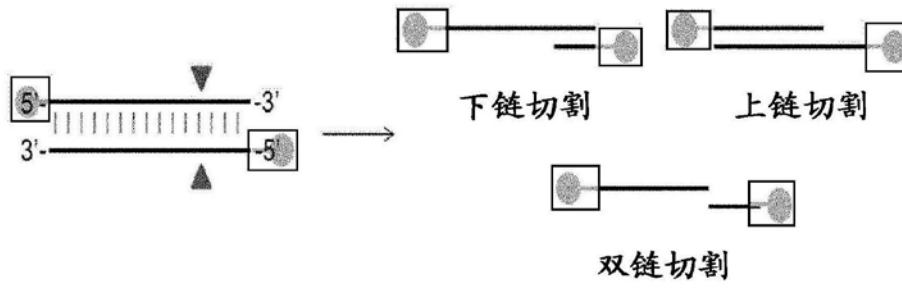


图16c

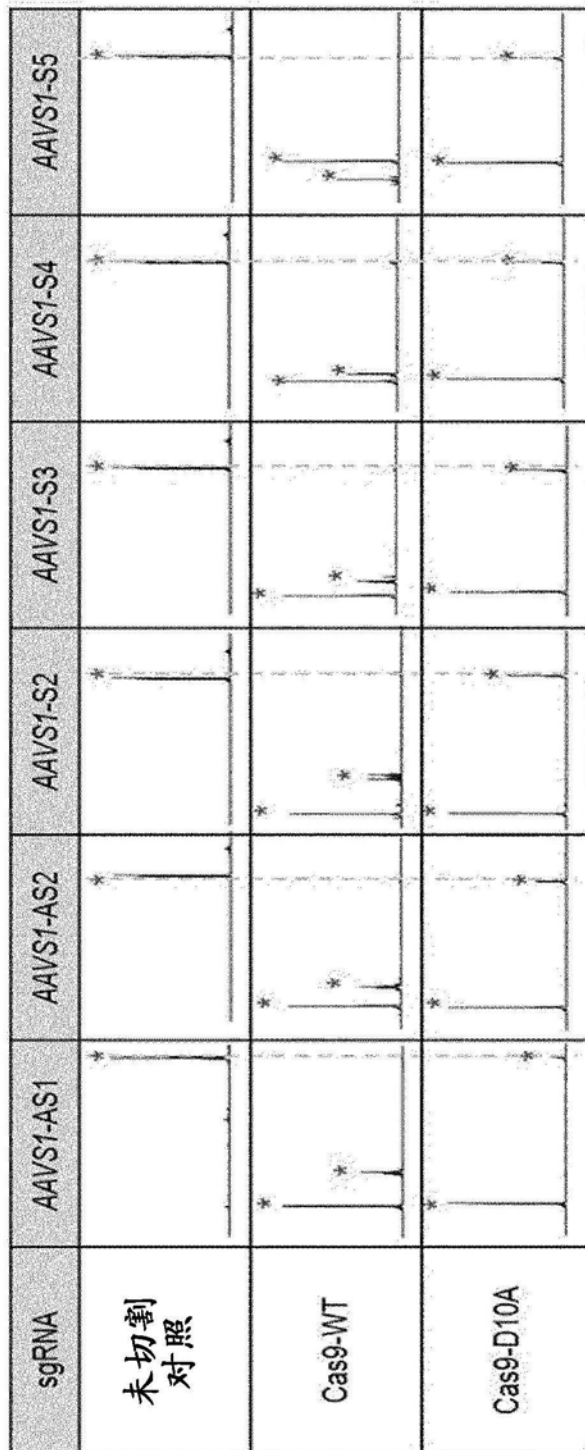


图16d

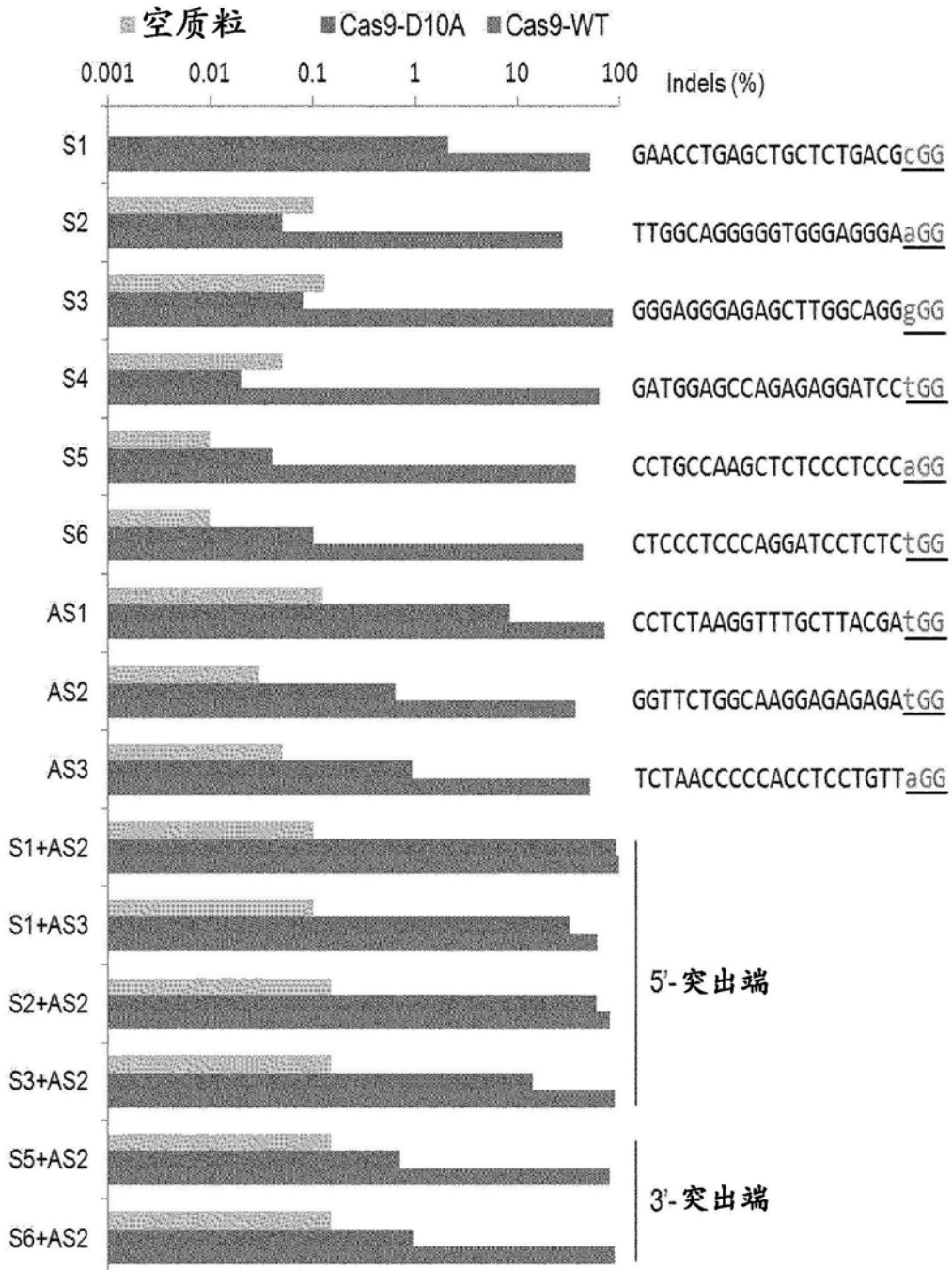


图17a

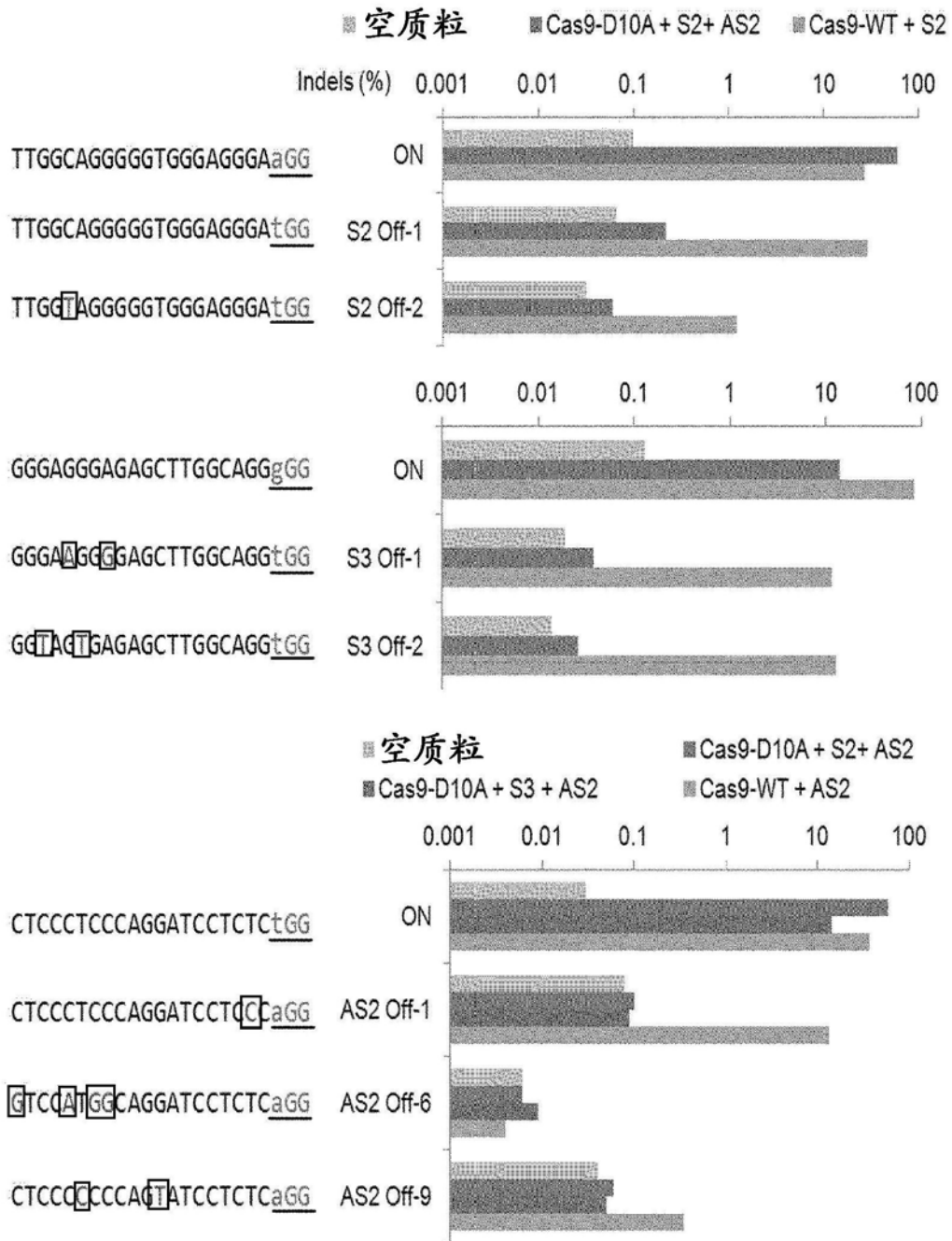


图17b

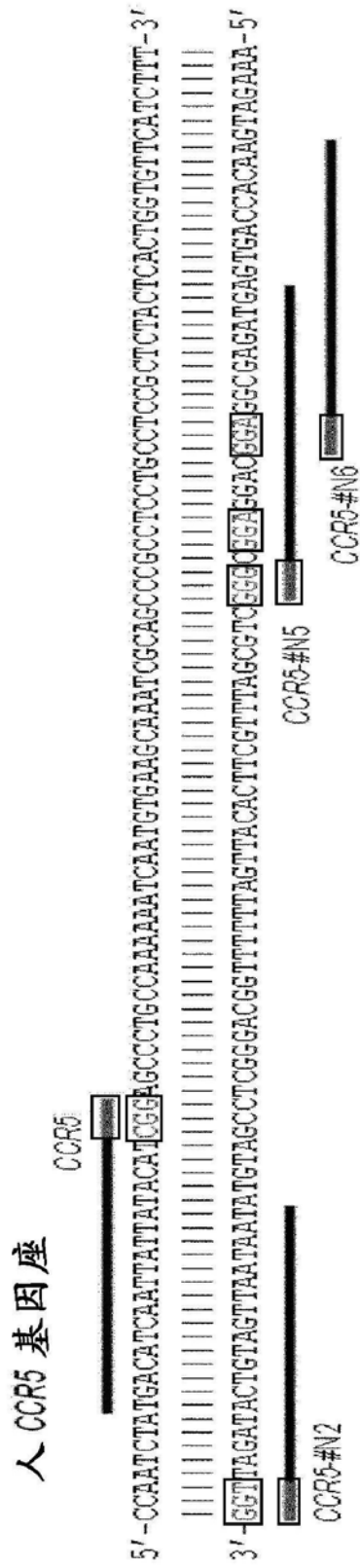


图18a

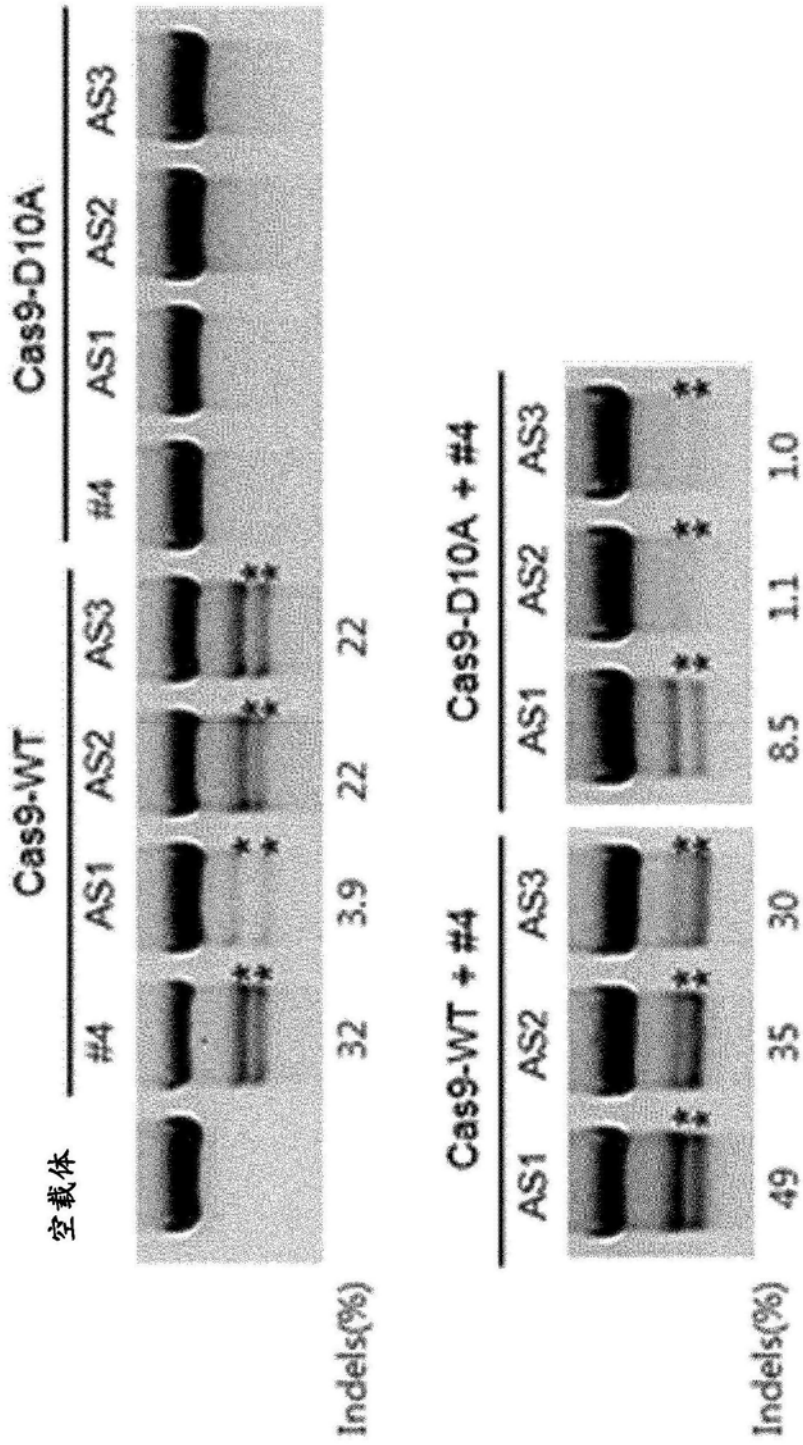


图18b



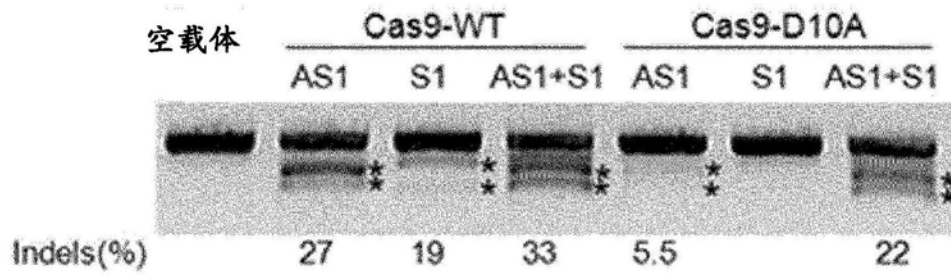


图18d

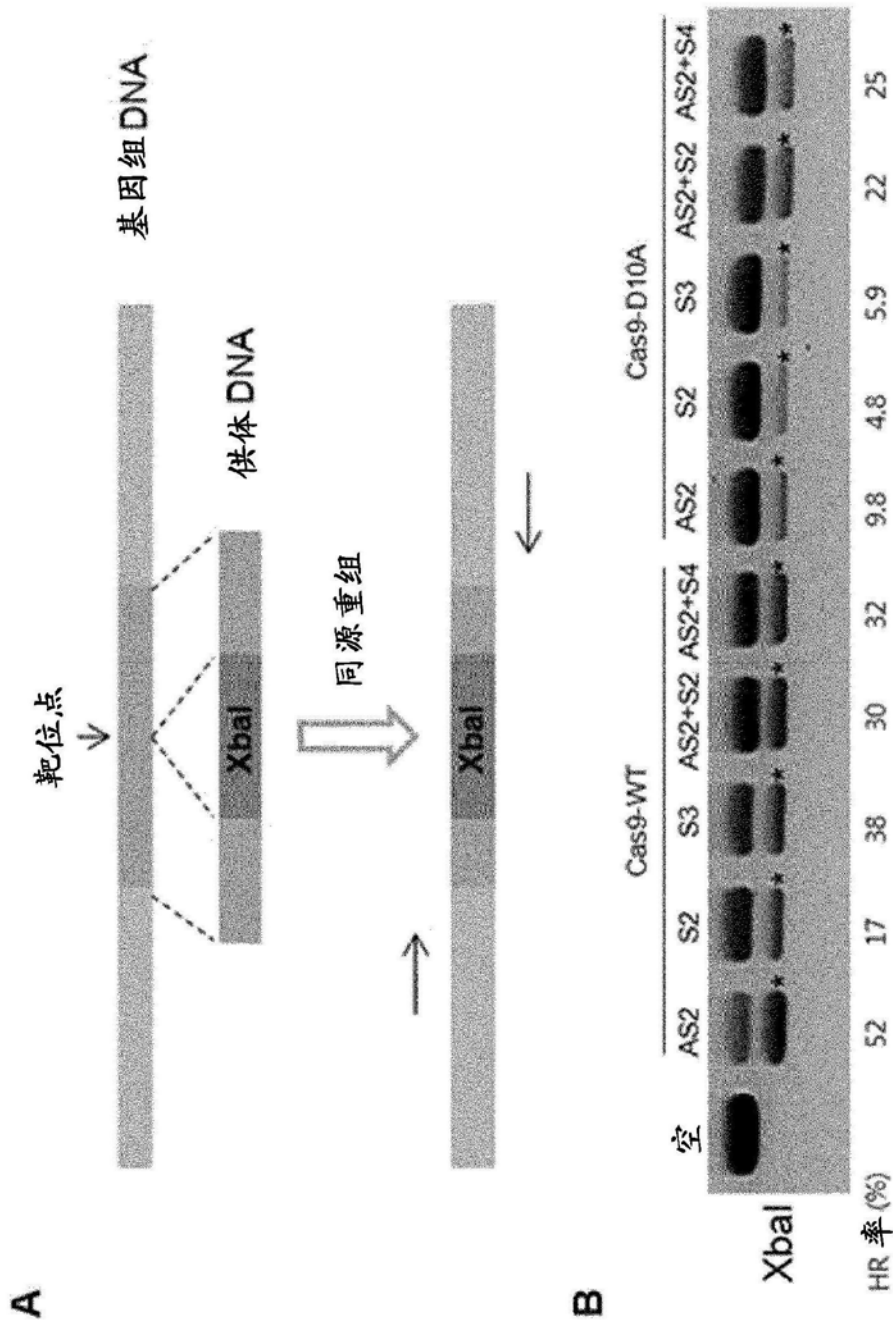


图19



图20a

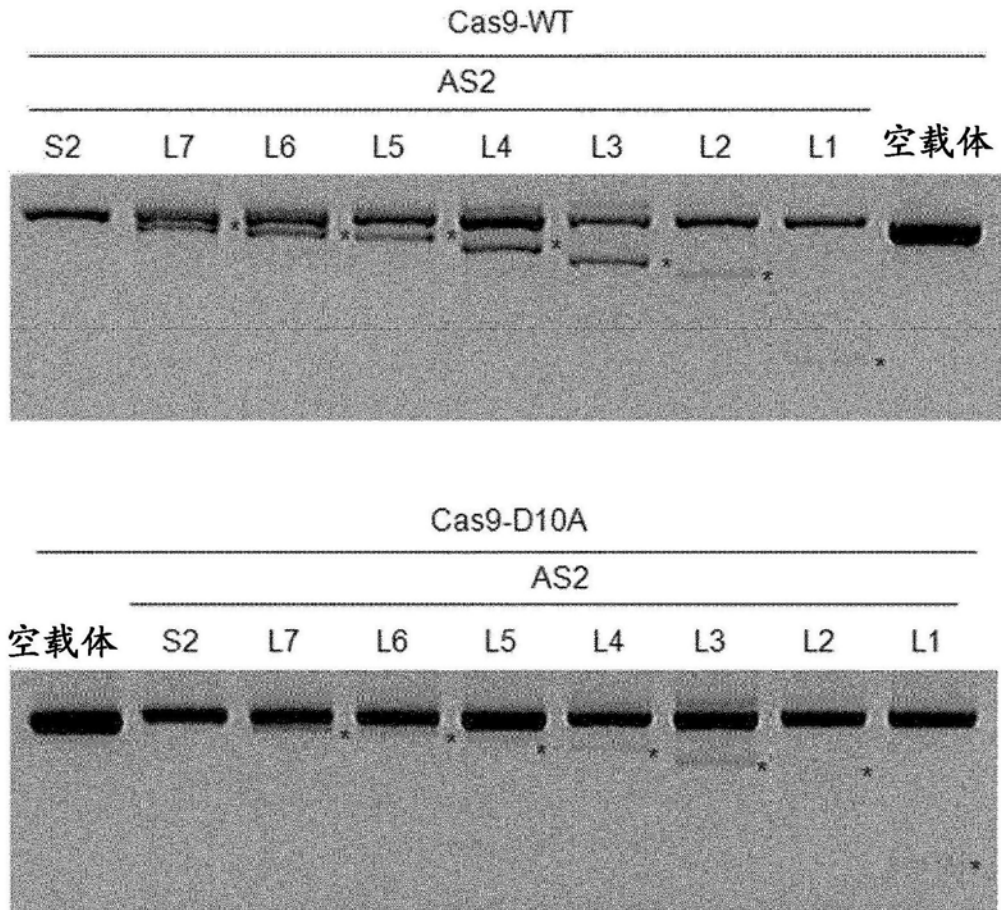


图20b

**Cas9-WT AS2 + L1 (~1050bp 缺失)**

```

ggcggggaatcaagagtcacccagagacagtgaccAACCCATccctgttt...//...agctctCCCTCCAGGATCCCTCTCTGGgtccatcgttaagcaaaccttagagggttctggccaaggagagagatg WT
ggcggggaatcaagagtcacccag------TGACCAACCATCctt-----gtaaagcaaaccttagagggttctggccaaggagagagatg
ggcggggaatcaagagtcacccag------CGTTGGgtccatcgttaagcaaaccttagagggttctggccaaggagagagatg-----gaa
ggcggggaatcaagagtcacccAGA-----CCTCTTTGGgtccatcgttaagcaaaccttagagggttctggccaaggagagagatg
ggcggggaatcaagagtcaccc-----taacag
ggcggggaatcaaga-----cgTTGGgtccatcgttaagcaaaccttagagggttctggccaaggagagagatg
ggc-----tccatcgttaagcaaaccttagagggttctggccaaggagagagatg
ggcggggaatcaagagtcacccAGA-----CTCTTTGGgtccatcgttaagcaaaccttagagggttctggccaaggagagagatg

```

**Cas9-D10A AS2 + L1 (~1050bp 缺失)**

```

ggcggggaatcaagagtcacccagagacagtgaccAACCCATccctgttt...//...agctctCCCTCCAGGATCCCTCTCTGGgtccatcgttaagcaaaccttagagggttctggccaaggagagagatg WT
ggcggggaatcaagagtcacccagagacagtgaccAACCCATC-----gtaaagcaaaccttagagggttctggccaaggagagagatg
gg-----tccatcgttaagcaaaccttagagggttctggccaaggagagagatg
ggcggggaatcaagagtcacccA-----tccatcgttaagcaaaccttagagggttctggccaaggagagagatg
ggcggggaatcaagagtcacccA-----tccatcgttaagcaaaccttagagggttctggccaaggagagagatg
gg-----CTTTGGgtccatcgttaagcaaaccttagagggttctggccaaggagagagatg
ggcggggaatcaagagtcacccagagacagtgaccAACCCATCcc-----tccatcgttaagcaaaccttagagggttctggccaaggagagagatg
ggcggggaatcaagagtc-----tcgtaagcaaaccttagagggttctggccaaggagagagatg x2

```

图20c

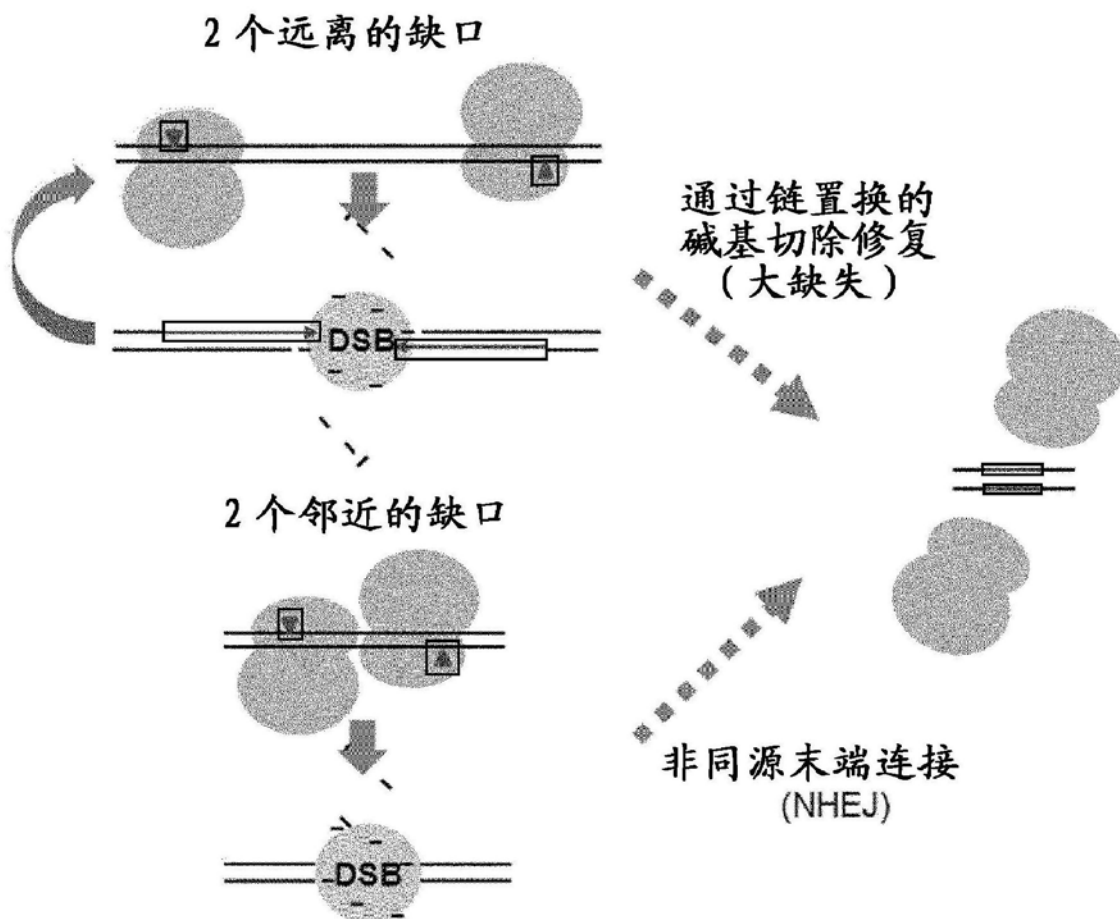


图20d

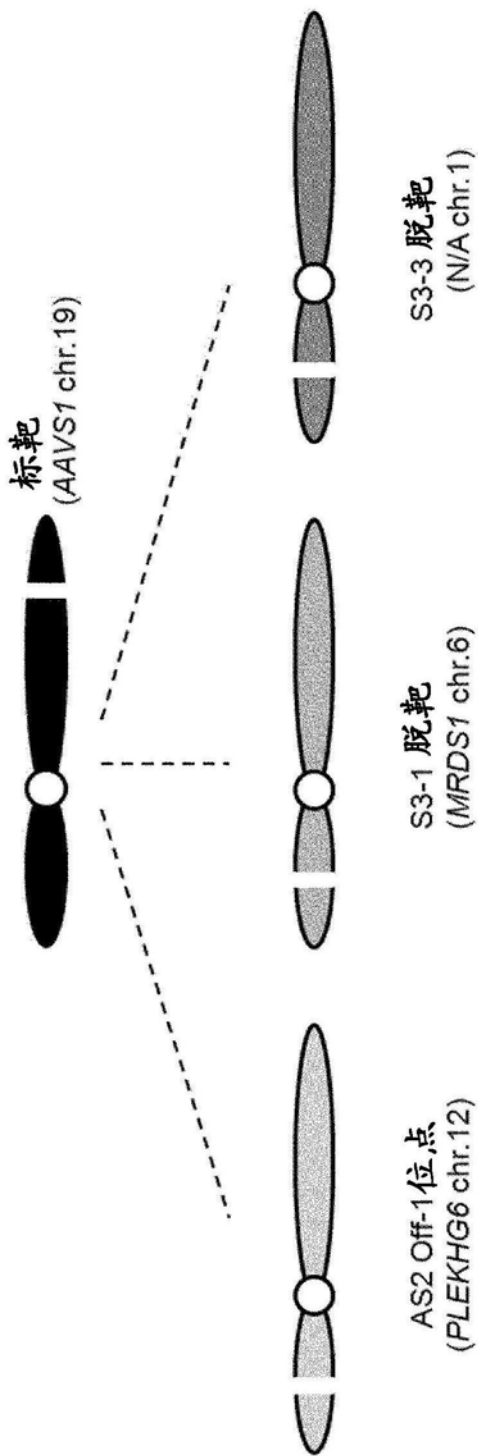


图21a

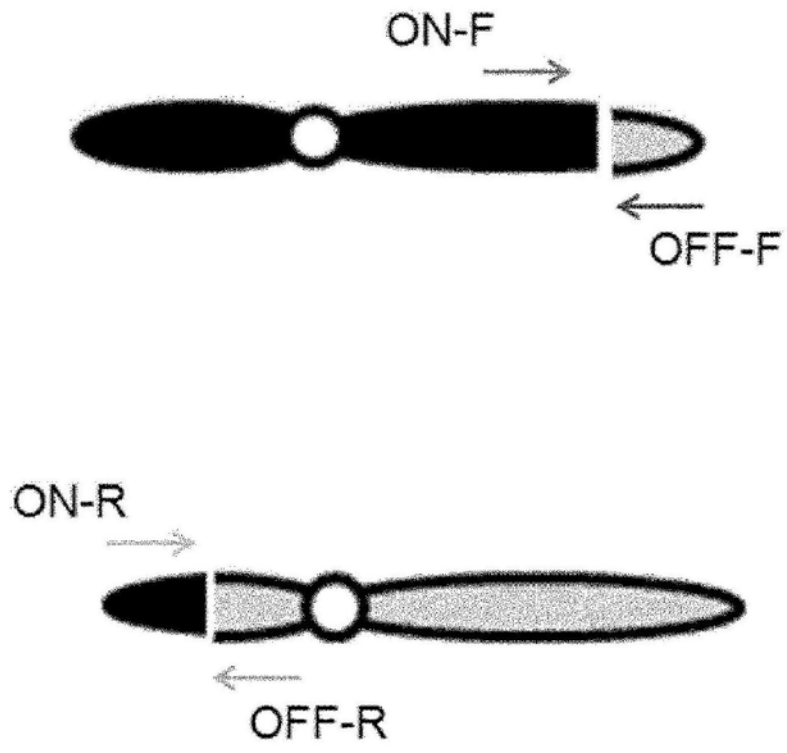


图21b

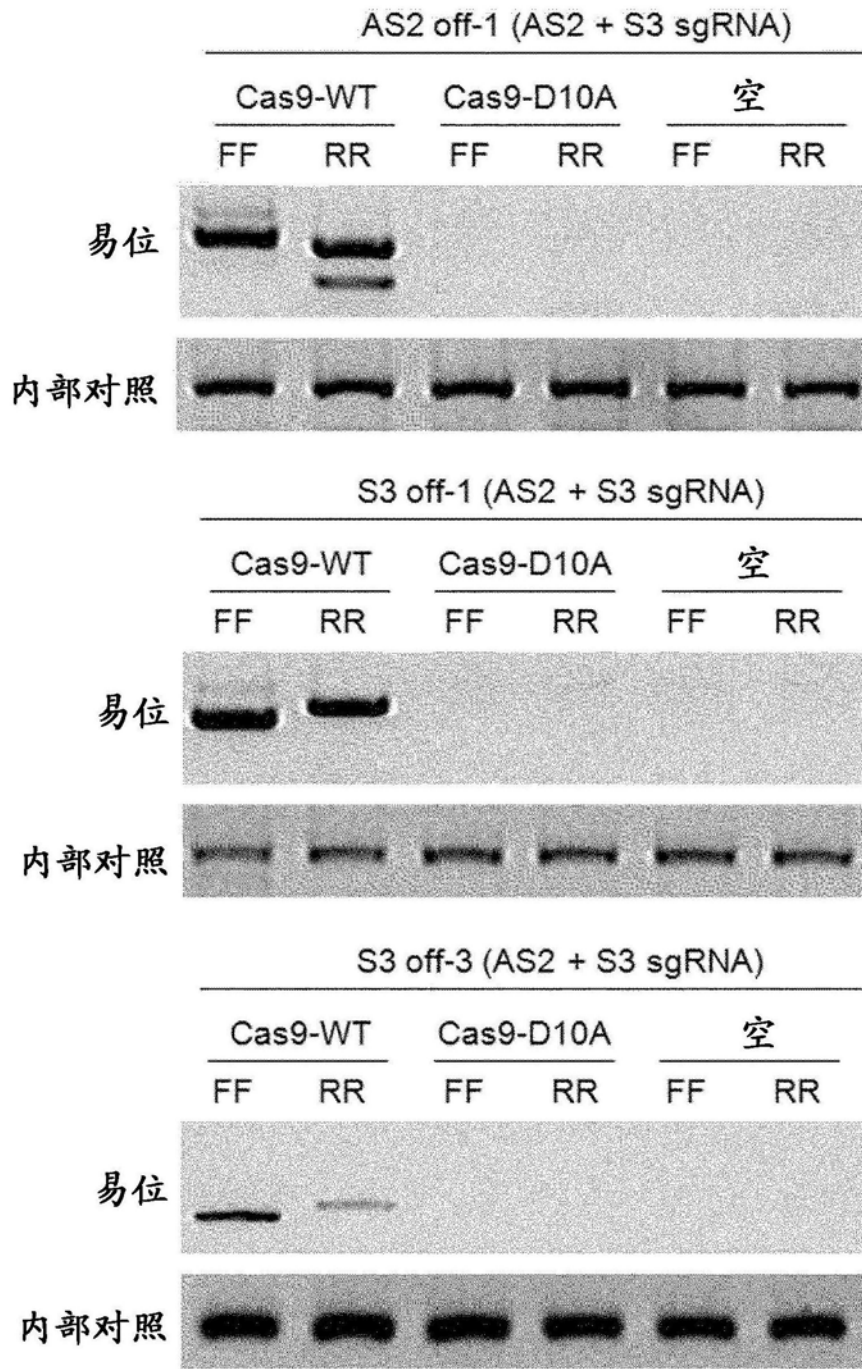


图21c

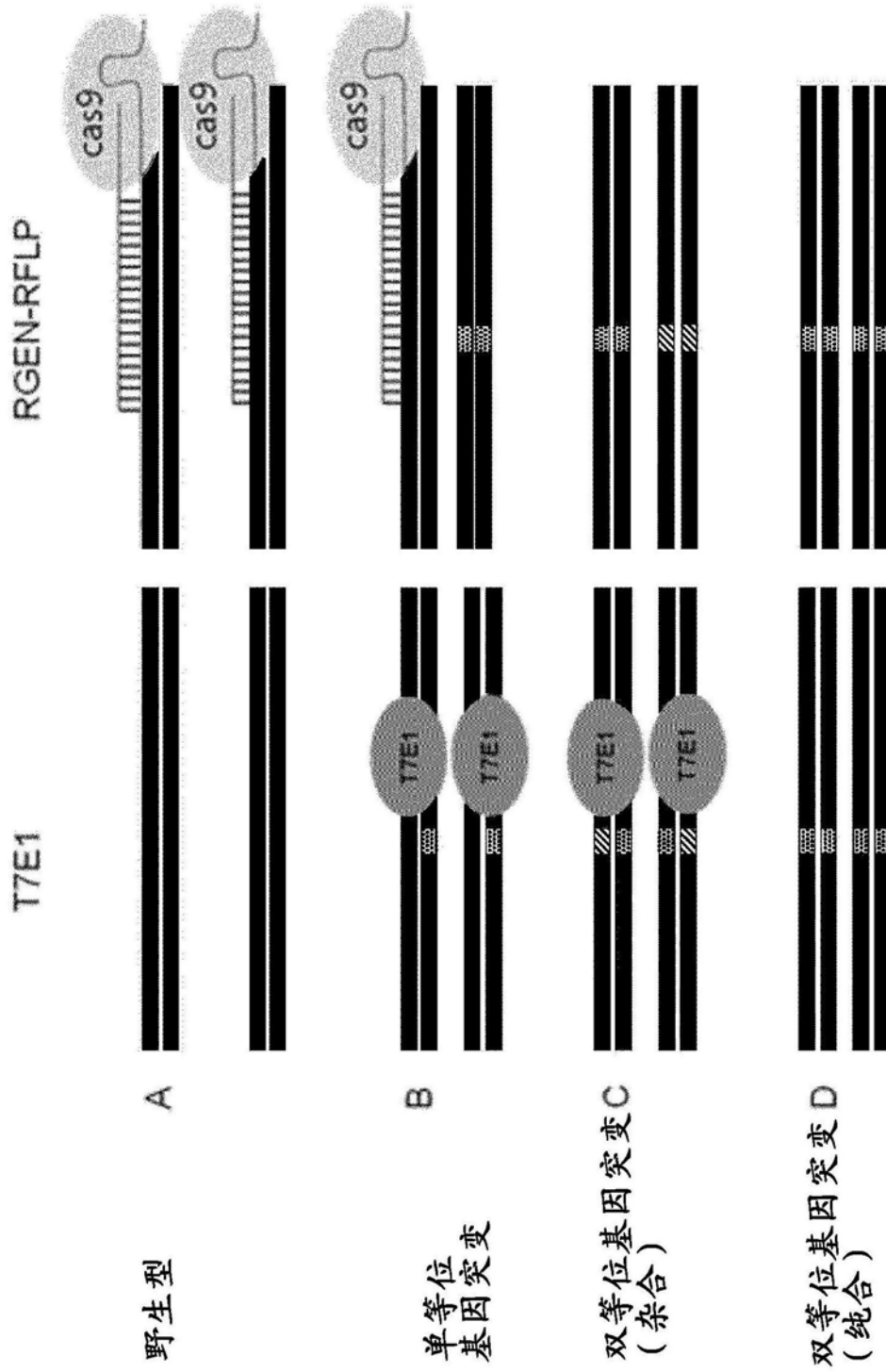


图22a

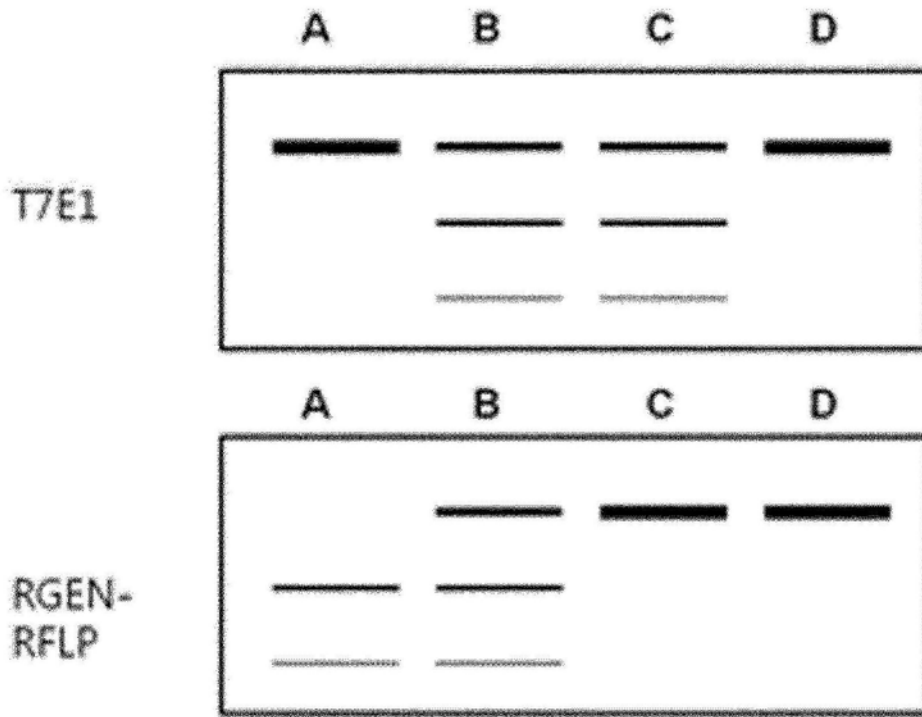


图22b

质粒靶序列

AATGACCACTACATCCT---CAAGGG WT  
 AATGACCACTACATCCTT---CAAGGG I1  
 AATGACCACTACATCCTTT---CAAGGG I2  
 AATGACCACTACATCCTTTTCAAGGG I3  
 AATGACCACTACATCCT----AAGGG D1  
 AATGACCACTACATCCT-----AGGG D2  
 AATGACCACTACATCCT-----GGG D3

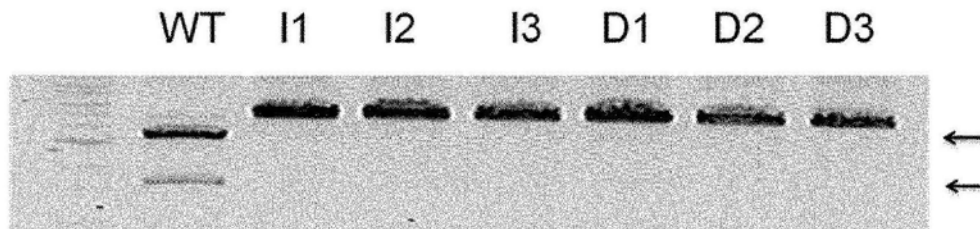


图23

```

#1 (+/-)
TATGTGCAATGACCACTACATCCT---CAAAGGGCAGCAATCGGAG WT
TATGTGCAATGACCACTACATCCTCCTCAAAGGGCAGCAATCGGAG +3

#2 (+/-)
TATGTGCAATGACCACTACATCCTCAAAGGGCAGCAATCGGAG WT
TATGTGCAATGACCACTACATC-----AATCGGAG -12

#5 (+/-)
TATGTGCAATGACCACTACATCCTCAAAGGGCAGCAATCGGAG WT
TATGTGCAATGACCACTACAT-----CAGCAATCGGAG -9

#6 (+/-)
TATGTGCAATGACCACTACATCCTCAAAGGGCAGCAATCGGAG WT
TATGTGCAATGACCACTACATCC-----AGCAATCGGAG -8

#12 (-/-)
-----CAGCAATCGG -36
TATGTGCAATGACCACTACATCCT-----CAAAGGGCAGCAATCGG +1
TATGTGCAATGACCACTACATCCT-----CAAAGGGCAGCAATCGG +1
TATGTGCAATGACCACTACATCCT/67bp/CAAAGGGCAGCAATCGG +67

#28 (-/-)
TATGTGCAATGACCACTACATCCTCAAAGGGCAGCAATCGG +1
TATGTGCAATGACCACTACAT---T-----GGCAGCAATCGG -7,+1
TATGTGCAATGACCACTACAT----- -94
    
```

图24a

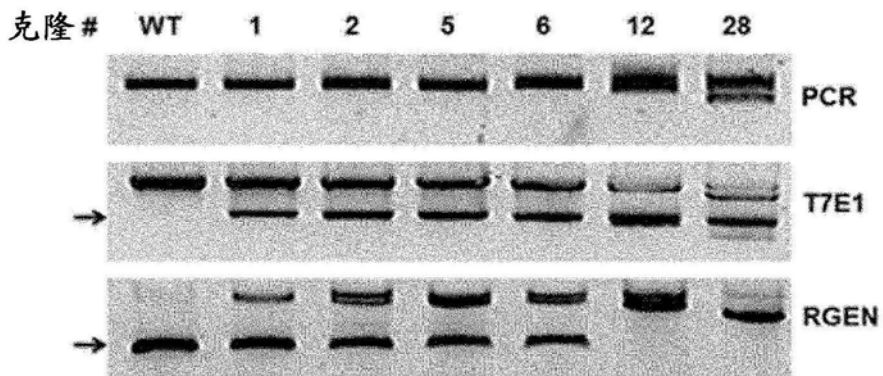


图24b

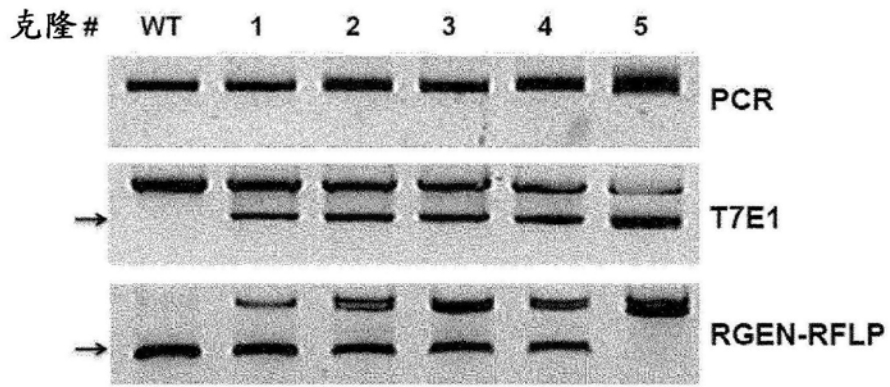


图25a

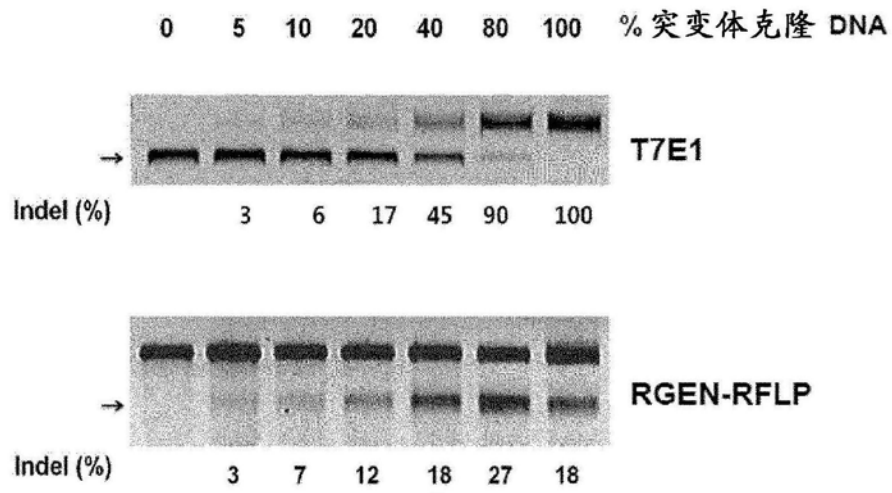


图25b

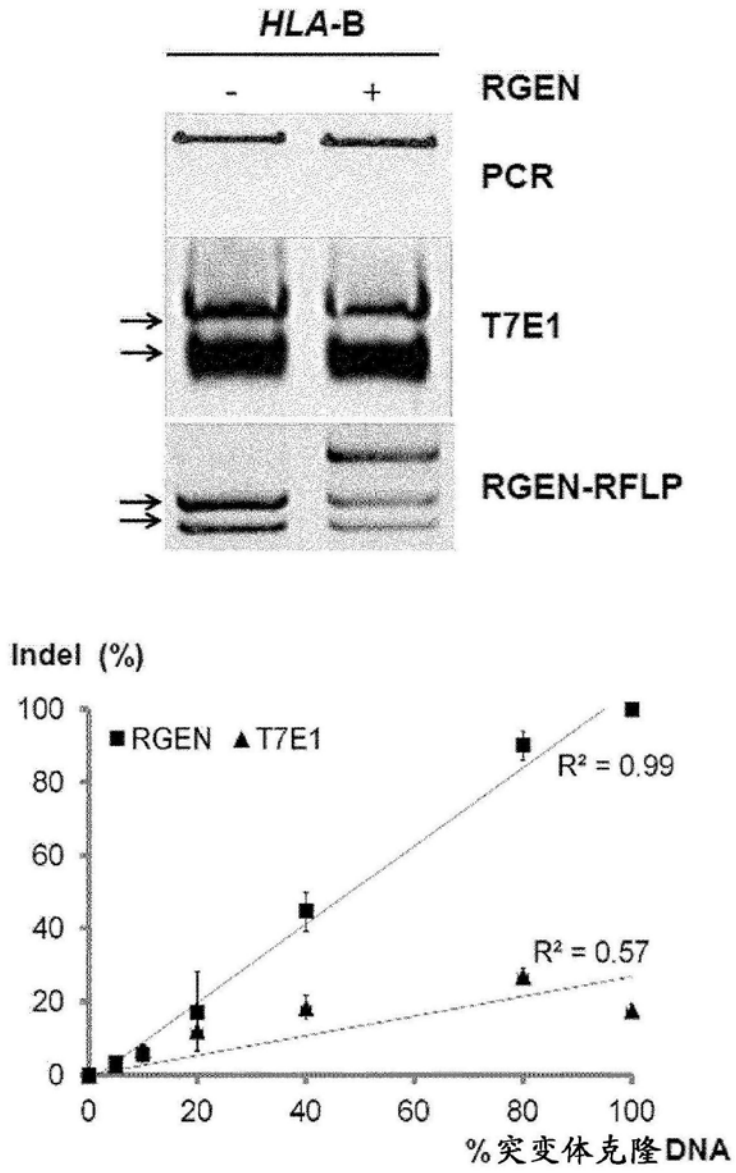


图25c

	野生型																	
	T	CAT	ACA	GAT	GAT	GTC	TCA	TCA	GAG	GAG	CGA	GAA	GGT	AAA	GTC	AAA	ATC	A
#1	T	CAT	ACA	GAT	GAT	GGT	GGT	GGT	AAA	GTC	AAA	ATC	A	Δ21 bp	(-/-)			
	T	CAT	ACA	GGT	GAT	G--	---	---	---	---	---	---	---	---	---	---	---	---
#3	T	CAT	ACA	GAT	GAT	GTC	TCA	TCA	GAG	---	CGA	GAA	GGT	AAA	GTC	AAA	ATC	A
	T	CAT	ACA	GAT	GAT	GTC	TCA	TCA	G--	---	CGA	GAA	GGT	AAA	GTC	AAA	ATC	A
#4	T	CAT	ACA	GAT	GAT	GTC	TCA	TCA	G-G	GAG	CGA	GAA	GGT	AAA	GTC	AAA	ATC	A
	T	CAT	ACA	GAT	GAT	GTC	TCA	TCA	GAG	GAG	CGA	GAA	GGT	AAA	GTC	AAA	ATC	A
#5	T	CAT	ACA	GAT	GAT	GTC	TCA	TC-	---	-AG	CGA	GAA	GGT	AAA	GTC	AAA	ATC	A
	T	CAT	ACA	GAT	GAT	GTC	TCA	TCA	GAG	GAG	CGA	GAA	GGT	AAA	GTC	AAA	ATC	A
#6	T	CAT	ACA	GAT	GAT	GTC	TC-	---	---	-G	CGA	GAA	GGT	AAA	GTC	AAA	ATC	A
	T	CAT	ACA	GAT	GAT	G--	---	---	---	---	---	---	---	---	---	---	---	---
#8	T	CAT	ACA	GAT	GAT	GTC	TAC	AGA	---	---	---	---	---	---	---	---	---	---
	T	CAT	ACA	GAT	GAT	GTC	TCA	TCA	G--	GAG	CGA	GAA	GGT	AAA	GTC	AAA	ATC	A
#11	T	CAT	ACA	GAT	GAT	G--	---	---	---	---	---	---	---	---	---	---	---	---
	T	CAT	ACA	GAT	GAT	GTC	AAA	ATC	A	Δ24 bp	(-/-)							

图26a

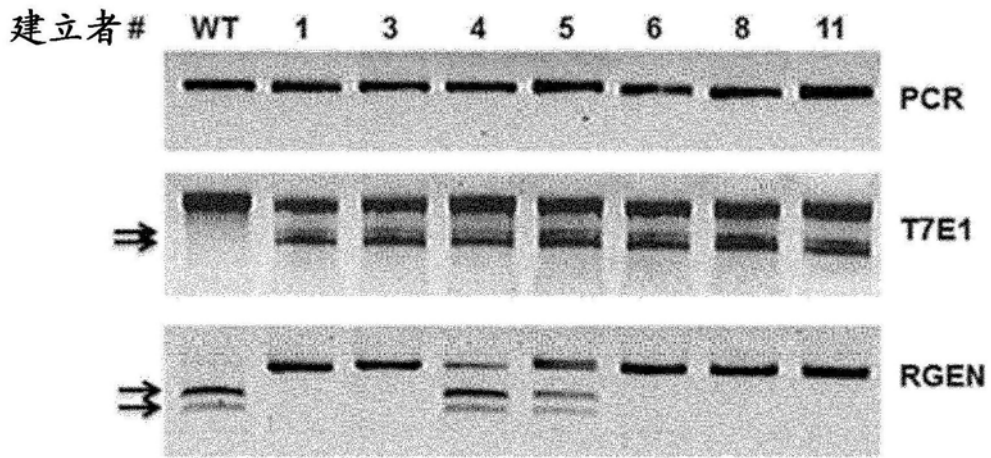


图26b

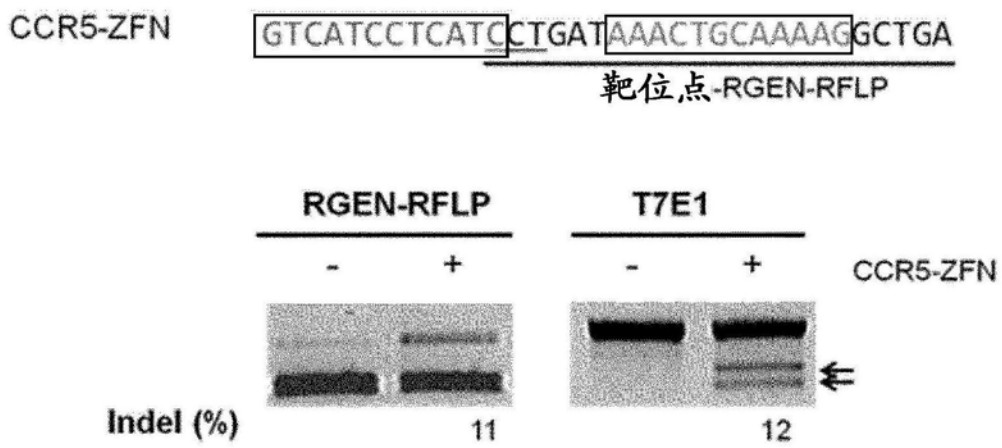


图27

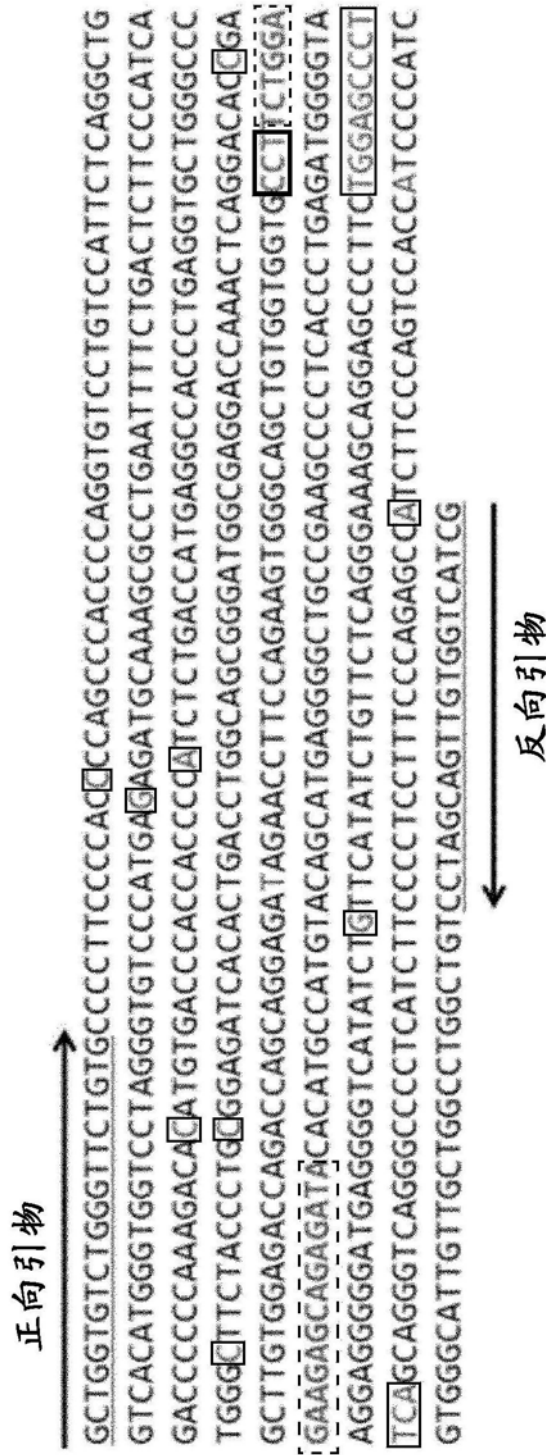
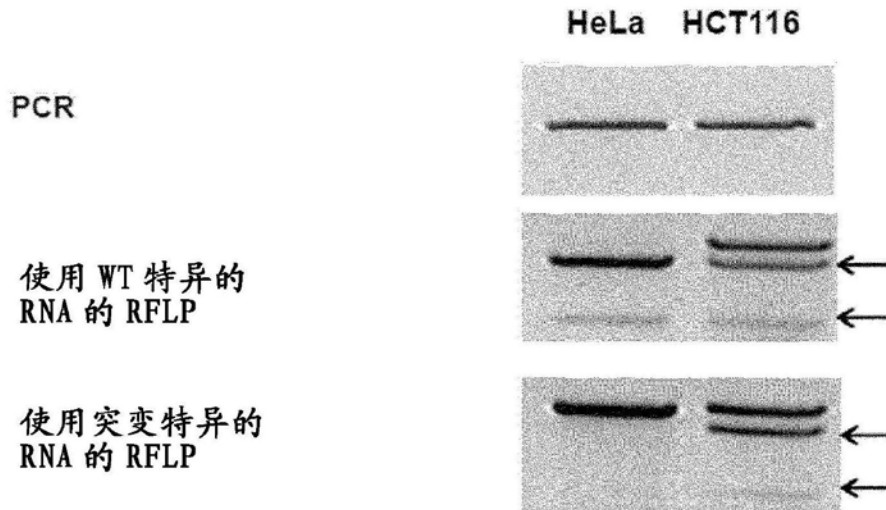


图28



**HeLa**

ACTACCACAGCTCCTTCTCTGAGTGG 野生型

**HCT116**

ACTACCACAGCTCCTTCTCTGAGTGG 野生型

ACTACCACAGCTCCT---CTGAGTGG c.133-135 del TCT

图29a

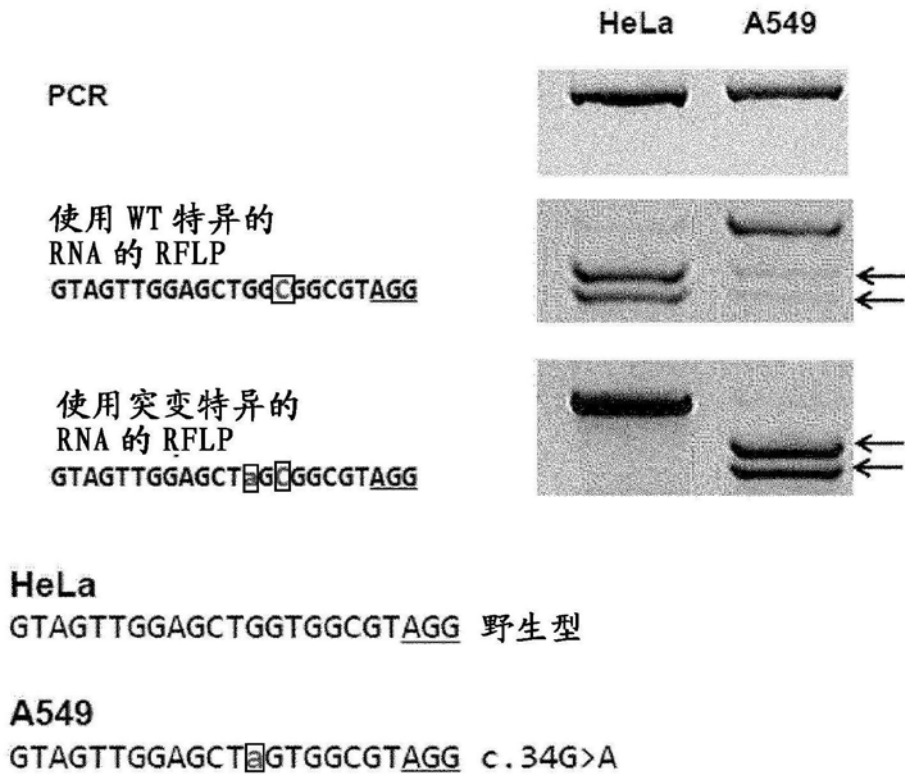


图29b

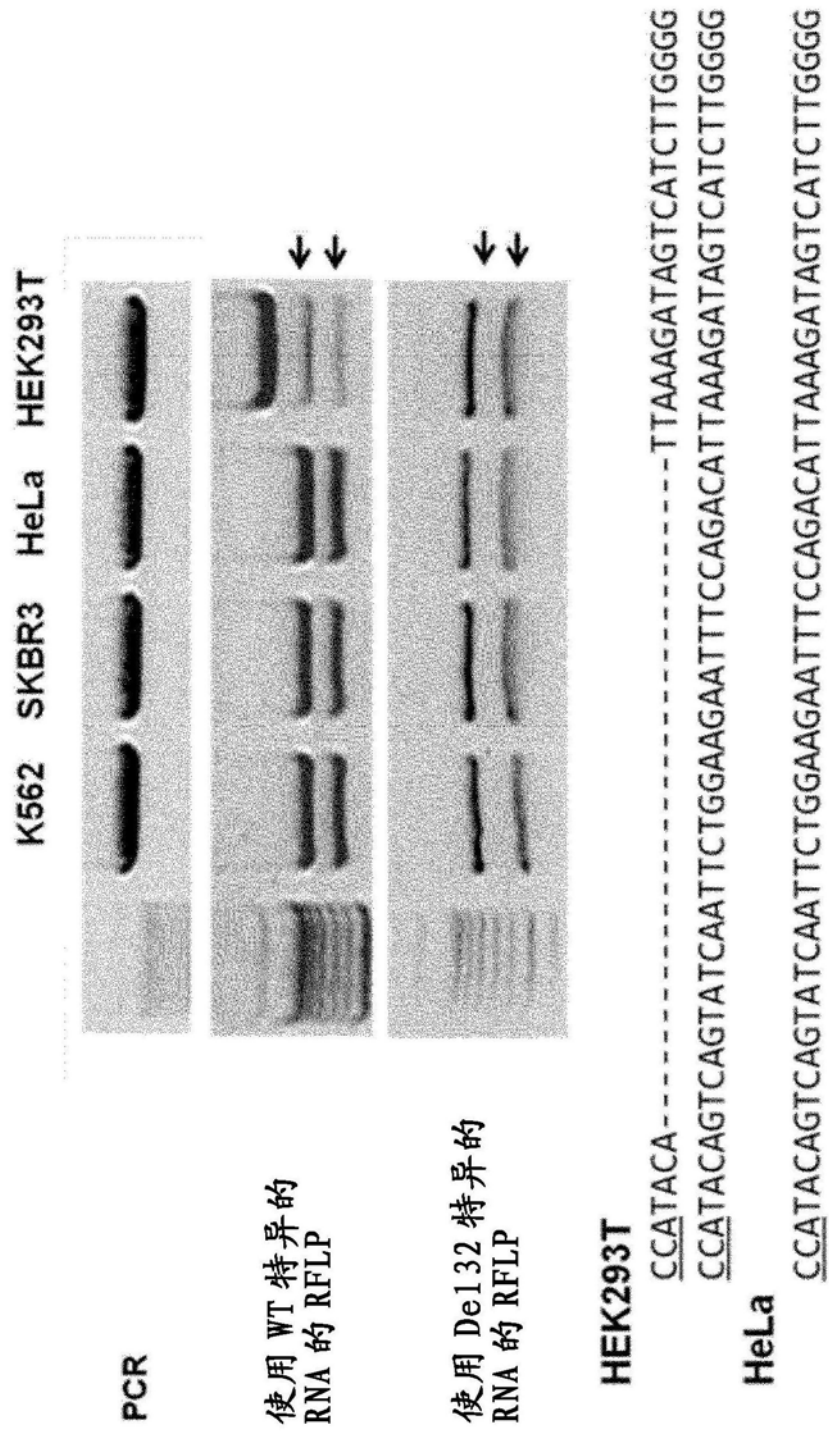


图30a

CCR5 WT  
ccatacagtcagtcataatctggaagaatttccagacattaaagatagtcatct  
CCR5 WT 靶位点                      CCR5 Δ32 脱靶位点

CCR5 Δ32  
ccataca-----ttaagatagtcatct  
CCR5 Δ32 靶位点

图30b



图30c

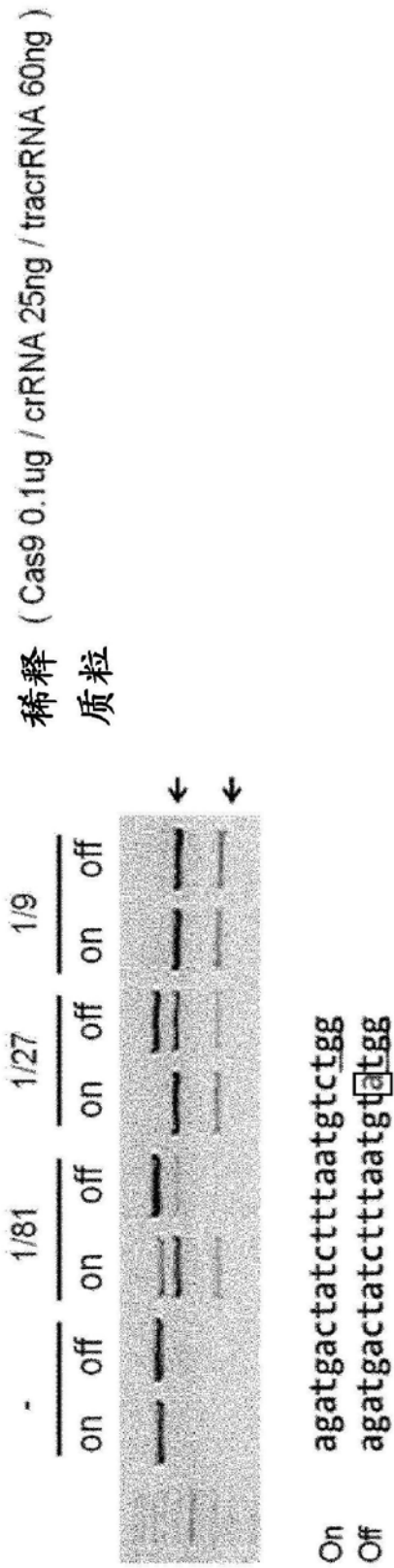


图30d

**KRAS****HeLa**

GTAGTTGGAGCTGGTGGCGTAGG 野生型

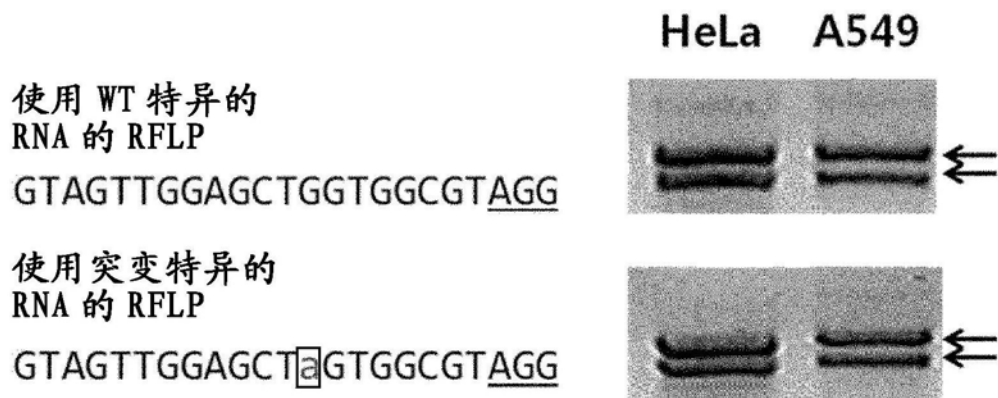
**A549**GTAGTTGGAGCT $\square$ aGTGGCGTAGG c.34G>A

图31a

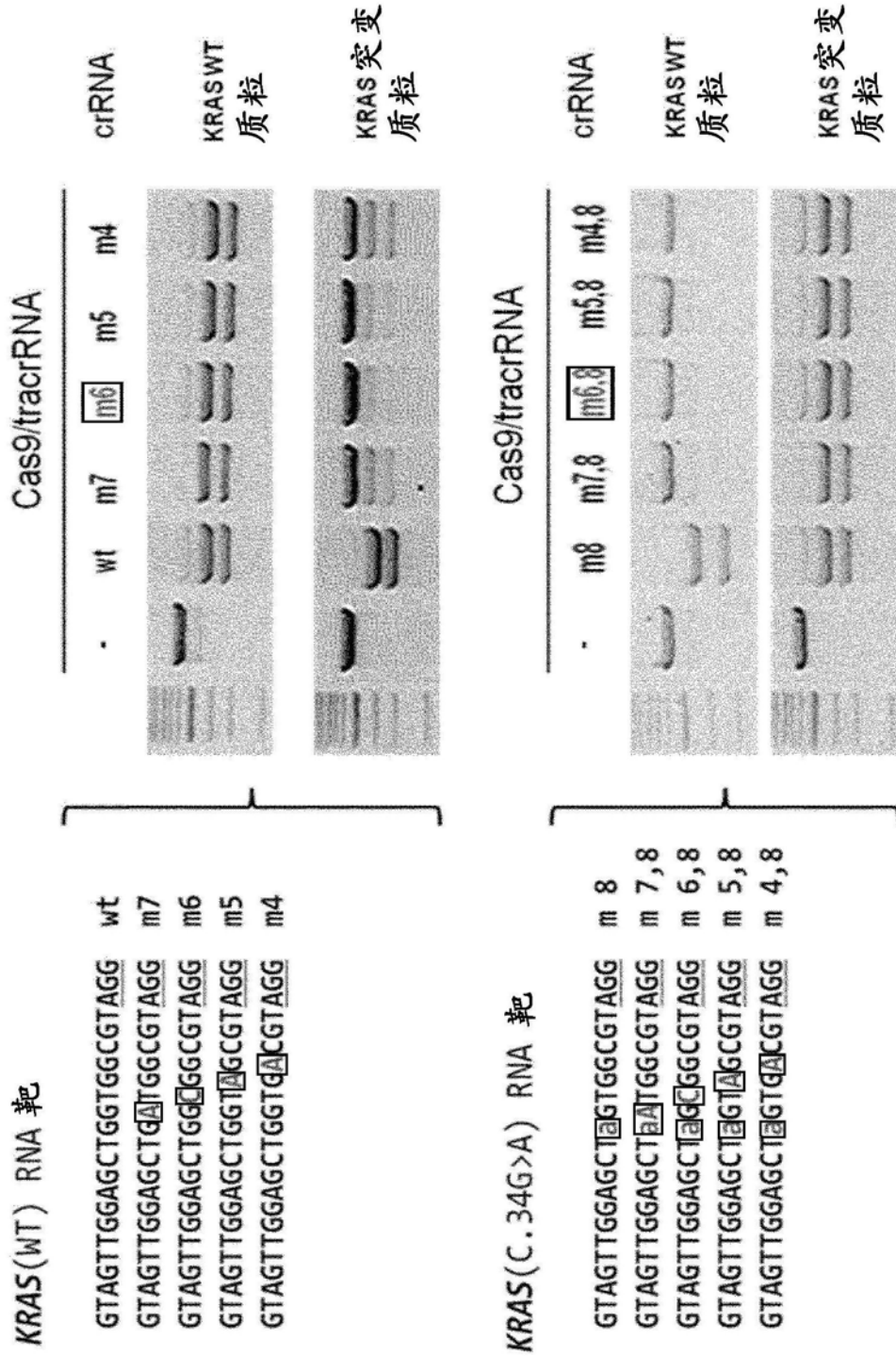


图31b

**PIK3CA****HeLa**

CAAATGAATGATGCACATCATGG 野生型

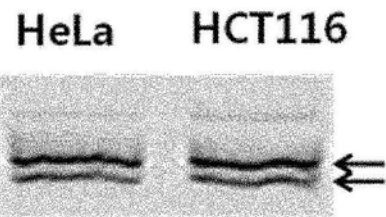
**HCT116**

CAAATGAATGATGCACATCATGG 野生型

CAAATGAATGATGCACgTCATGG C.3140A&gt;G

使用 WT 特异的  
RNA 的 RFLP

CAAATGAATGATGCACATCATGG

使用突变特异的  
RNA 的 RFLP

CAAATGAATGATGCACgTCATGG

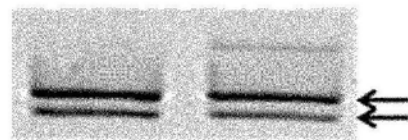


图32a

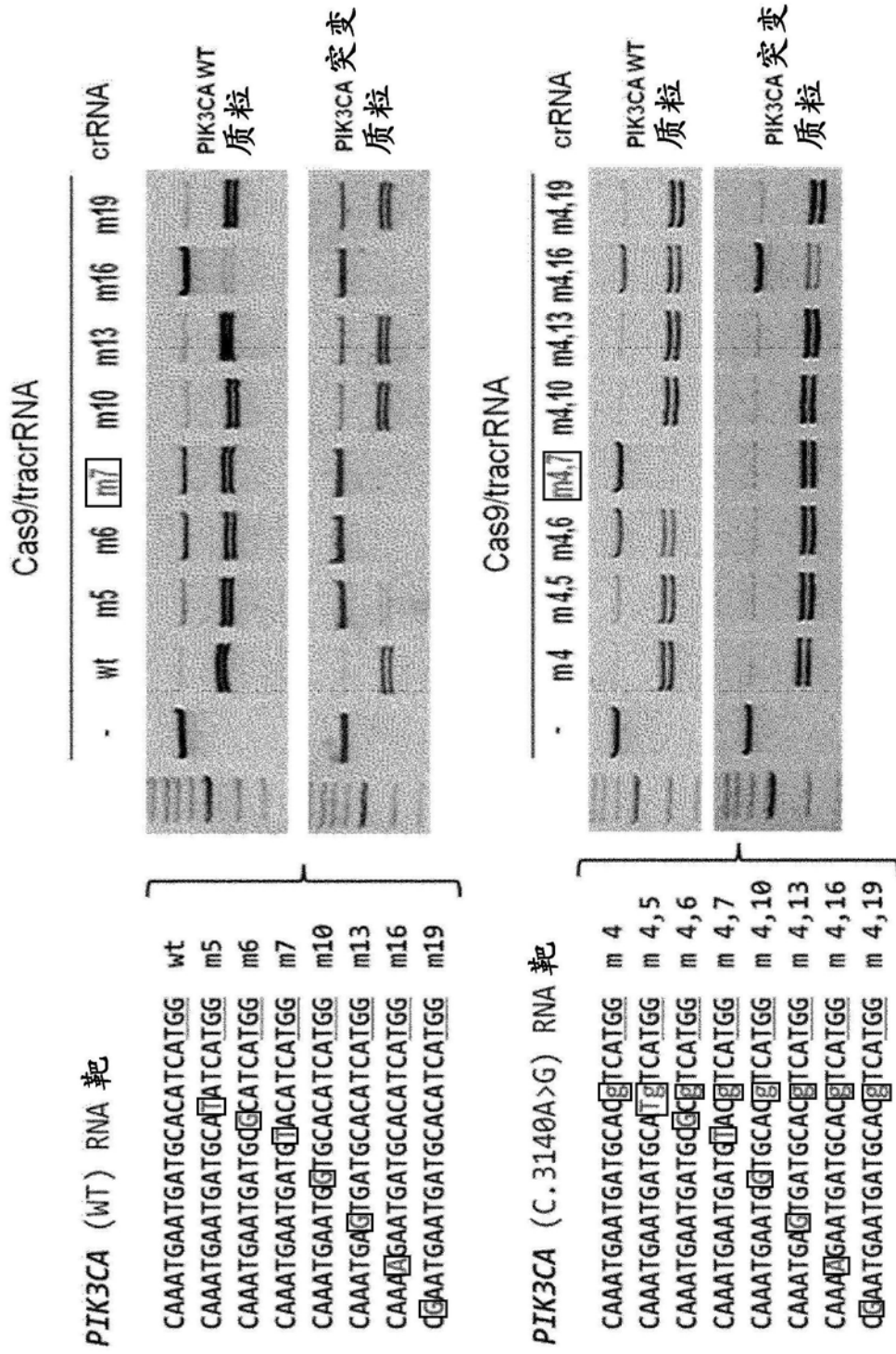


图32b

# IDH1

## HeLa

ATCATAGGTCGTCATGCTTATGG 野生型

## HT1080

ATCATAGGTCGTCATGCTTATGG 野生型

ATCATAGGTtGTCATGCTTATGG c.394C>T

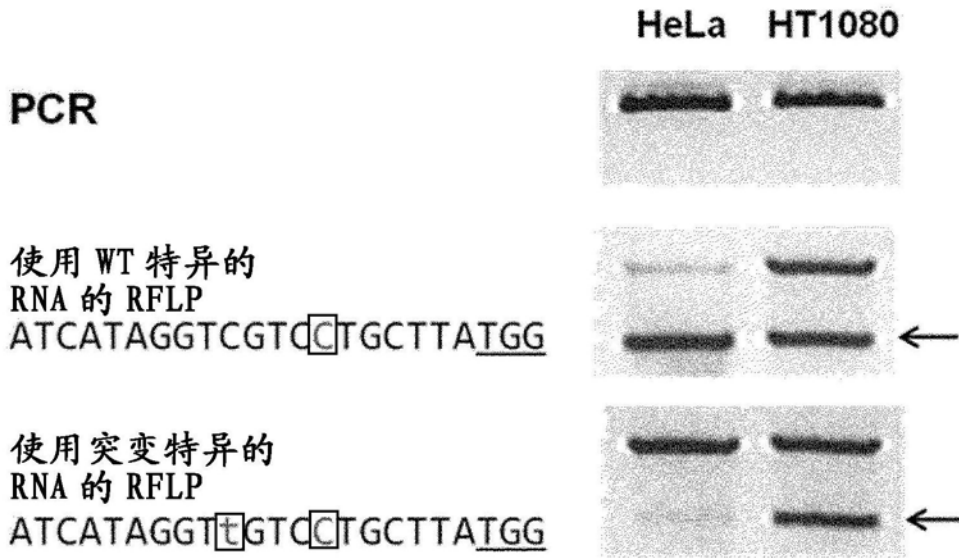


图33a

# PIK3CA

## HeLa

CAAATGAATGATGCACATCATGG 野生型

## HCT116

CAAATGAATGATGCACATCATGG 野生型

CAAATGAATGATGCACgTCATGG C.3140A>G

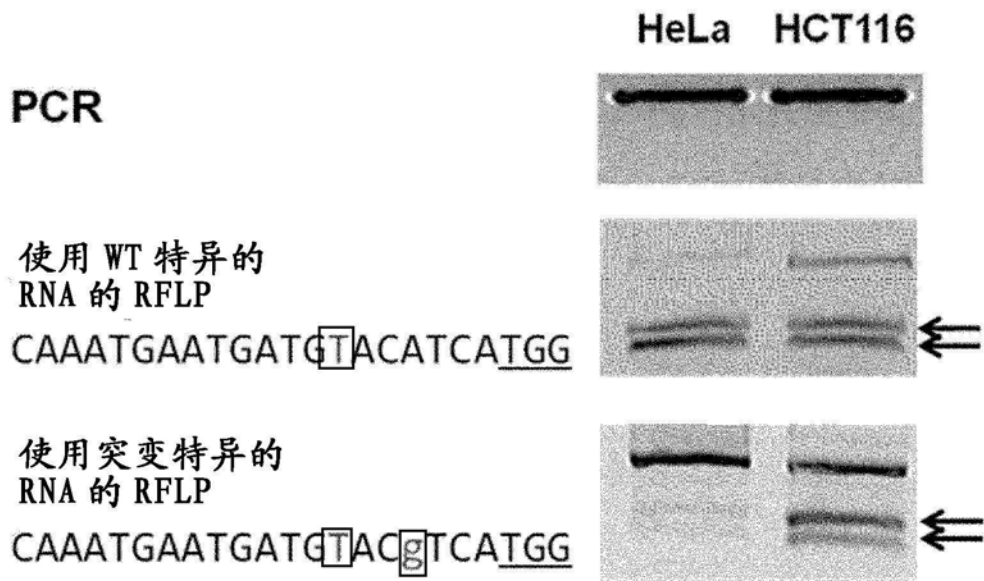


图33b

**NRAS****HeLa**

CTGGACAAGAAGAGTACAGTGCC 野生型

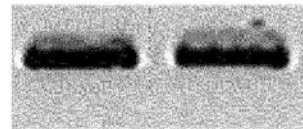
**HT1080**

CTGGACAAGAAGAGTACAGTGCC 野生型

CTGGAAAAAGAAGAGTACAGTGCC c.181C>A

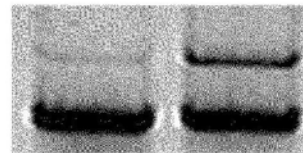
**PCR**

HeLa HT1080



使用 WT 特异的  
RNA 的 RFLP

CTGGACAAGAAGAGTACAGTGCC



使用突变特异的  
RNA 的 RFLP

CTGGAAAAAGAAGAGTACAGTGCC

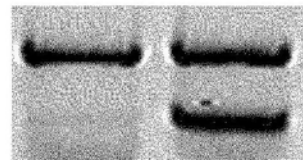


图33c

**BRAF****HeLa**

ACTCCATCGAGATTTCACTGTAG 野生型

**HT29**

ACTCCATCGAGATTTCACTGTAG 野生型

ACTCCATCGAGATTTCTCTGTAG (c.1799T&gt;A)

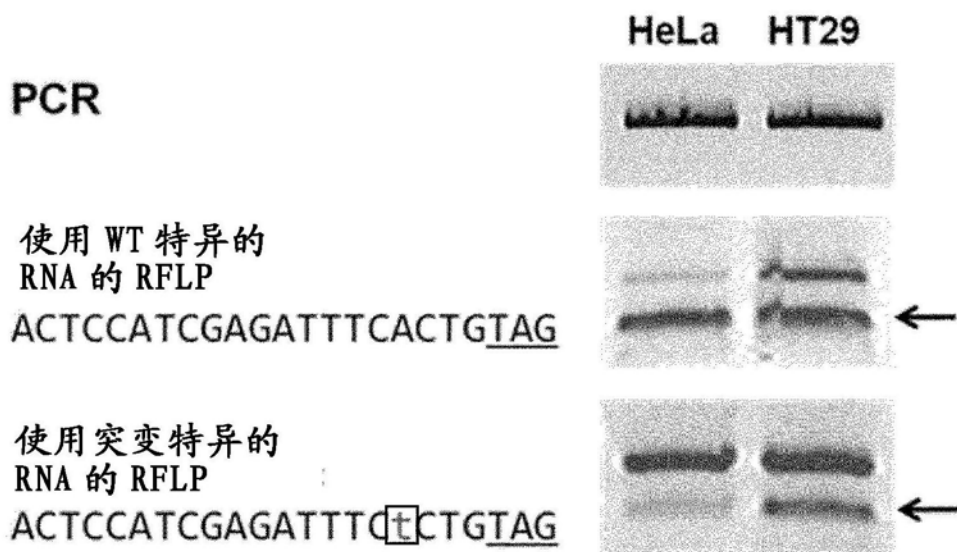


图33d