(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 114634950 A
(43)申请公布日 2022.06.17

(21)申请号 202111233527.5

(22)申请日 2013.12.12

(30)优先权数据
61/736,527 2012.12.12 US
61/748,427 2013.01.02 US
61/768,959 2013.02.25 US
61/791,409 2013.03.15 US
61/835,931 2013.06.17 US

(62)分案原申请数据
201380072868.6 2013.12.12

(71)申请人 布罗德研究所有限公司
地址 美国马萨诸塞州
申请人 麻省理工学院 洛克菲勒大学
哈佛大学校长及研究员协会

(72)发明人 张锋 D·O·比卡德 丛乐
D·B·T·科克斯 P·徐 W·蒋
林帅亮 L·马拉菲尼
R·J·普拉特 然霏
N·E·桑加纳

(74)专利代理机构 北京市金杜律师事务所 11256
专利代理师 陈文平

(51)Int.Cl.
C12N 15/864(2006.01)
C12N 9/22(2006.01)
C12N 15/113(2010.01)
C07K 19/00(2006.01)
A61K 48/00(2006.01)

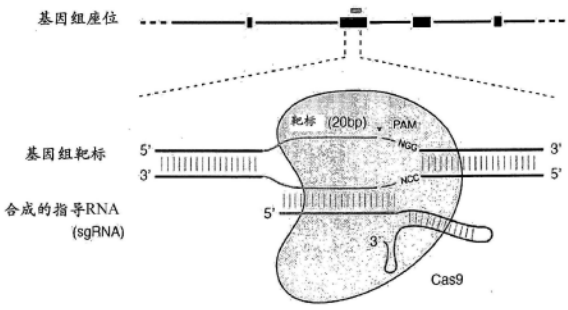
权利要求书2页 说明书128页
序列表240页 附图111页

(54)发明名称

用于序列操纵的CRISPR-CAS组分系统、方法以及组合物

(57)摘要

本发明提供了用于操纵序列和/或靶序列的活性的系统、方法、和组合物。提供了多种载体和载体系统、以及用于设计和使用此类载体的方法，其中一些载体和载体系统编码CRISPR复合物的一个或多个组分。还提供了引导CRISPR复合物在真核细胞中形成的方法以及用于通过利用CRISPR/Cas系统引入精确突变来筛选特异性细胞的方法。



CN 114634950 A

1.一种工程化的、非天然存在的CRISPR-Cas系统，包含：

a)包含连接到一个或多个核定位信号(NLS)的Cas9和一个或多个异源蛋白结构域的蛋白质，或编码该蛋白质的多核苷酸；

b)CRISPR-Cas系统嵌合指导，其包含能够杂交到真核细胞的细胞核中邻近于原型间隔子相邻基序(PAM)的靶序列并且引导CRISPR-Cas复合物与该靶序列的序列特异性结合的指导序列，能够杂交到tracr序列上的tracr配对序列，并且该tracr序列包含至少30个核苷酸的长度。

2.一种工程化的、非天然存在的CRISPR-Cas载体系统，该载体系统包含一个或多个载体，该一个或多个载体包含

a)一个第一调节元件，该第一调节元件可操作地连接到编码包含连接到一个或多个核定位信号(NLS)的Cas9和一个或多个异源蛋白结构域的蛋白质的核苷酸序列上；

b)一个第二调节元件，该第二调节元件可操作地连接到编码CRISPR-Cas系统嵌合指导的核苷酸序列上，所述CRISPR-Cas系统嵌合指导包含能够杂交到真核细胞的细胞核中邻近于原型间隔子相邻基序(PAM)的靶序列并且引导CRISPR-Cas复合物与该靶序列的序列特异性结合的指导序列，能够杂交到tracr序列上的tracr配对序列，并且该tracr序列包含至少30个核苷酸的长度；

其中组分a)和b)位于该系统的相同或不同载体上。

3.如权利要求2所述的系统，其中载体是病毒载体。

4.如权利要求3所述的系统，其中病毒载体是腺相关载体。

5.如权利要求1或2所述的系统，其中tracr序列包含至少40个核苷酸的长度。

6.如权利要求1或2所述的系统，其中tracr序列包含至少50个核苷酸的长度。

7.如权利要求1或2所述的系统，其中指导序列包含少15-30个核苷酸的长度。

8.如权利要求1或2所述的系统，其中指导序列包含至少20个核苷酸的长度。

9.如权利要求1或2所述的系统，其中该蛋白质包含至少一个连接至Cas9的SV40病毒大型T-抗原NLS(PKKKRKV)。

10.如权利要求1或2所述的系统，其中该蛋白质包含至少两个连接至Cas9的NLS。

11.如权利要求1或2所述的系统，其中该蛋白质包含至少一个N-末端NLS和至少一个C-末端NLS。

12.如权利要求1或2所述的系统，其中该异源蛋白质结构域具有下列活性的一者或多者：甲基酶活性、脱甲基酶活性、转录激活活性、转录阻遏活性、转录释放因子活性、组蛋白修饰活性、RNA切割活性和核酸结合活性。

13.如权利要求1或2所述的系统，其中该Cas9是化脓链球菌Cas9，并且PAM是NGG；或者该Cas9是嗜热链球菌Cas9，并且PAM是NNAGAAW。

14.如权利要求1或2所述的系统，其中该Cas9是Cas9切口酶，其包含在RuvC或HNH结构域中的突变并且缺乏切割一条DNA链的能力。

15.如权利要求14所述的系统，其中该Cas9包含对应于化脓链球菌Cas9的D10A、H840A、N854A或N863A的突变。

16.如权利要求1或2所述的系统，其中该Cas9是催化钝化的Cas9，其包含在RuvC结构域中的突变和在HNH结构域中的突变并且实质上缺乏所有DNA切割活性。

17.如权利要求16所述的系统，其中该Cas9包含对应于化脓链球菌Cas9的D10A的突变和对应于化脓链球菌Cas9的H840A、N854A或N863A的突变。

18.如权利要求1或2所述的系统，其中该系统还包含模板多核苷酸，其中该模板多核苷酸与靶序列的至少5个核苷酸重叠。

19.如权利要求1所述的系统，其中编码蛋白质的多核苷酸进行了密码子优化以便在哺乳动物或人细胞中表达，并且其中编码蛋白质的多核苷酸是mRNA。

20.如权利要求1或2所述的系统在制备用于基因疗法或基因编辑的药物中的用途。

21.如权利要求1或2所述的系统用于离体或体内修饰真核细胞的用途。

22.一种真核细胞，包含如权利要求1或2所述的系统。

23.一种非人生物体，包含如权利要求22所述的真核细胞，其中该非人生物体不是人或植物品种。

24.源自于化脓链球菌的Cas9酶。

25.源自于嗜热链球菌的Cas9酶。

26.一种试剂盒，包含Cas9酶。

27.一种融合蛋白，包含RNAse、mCherry和核定位信号(NLS)。

28.一种组合物，包含选自如下的一个或多个PCR引物：

AAAACCACCCTTCTCTCTGGC、GGAGATTGGAGACACGGAGAG、CTGGAAAGCCAATGCCTGAC和GGCAGCAAACTCCTTGTCCT。

用于序列操纵的CRISPR-CAS组分系统、方法以及组合物

本申请是申请日为2013年12月12日和发明名称为“用于序列操纵的 CRISPR-Cas组分系统、方法以及组合物”的201380072868.6号发明专利申请 的分案申请。

相关应用和引用参考

[0001] 本申请要求分别具有博德参考号BI-2011/008/WSGR案卷号44063- 701.101、BI-2011/008/WSGR案卷号44063-701.102，博德参考号BI- 2011/008/VP案卷号44790.01.2003、BI-2011/008/VP案卷号44790.02.2003以 及BI-2011/008/VP案卷号44790.03.2003的美国临时专利申请61/736,527， 61/748,427,61/768,959,61/791,409以及61/835,931的优先权，所有标题都为 “用于序列操纵的系统、方法以及组合物”，分别提交于2012年12月12 日、2013年1月2日、2013年2月25日、2013年3月15日和2013年6月 17日。

[0002] 参考了美国临时专利申请61/758,468;61/769,046;61/802,174； 61/806,375;61/814,263;61/819,803以及61/828,130，这些临时专利申请各自 的标题为“用于序列操纵的系统、方法以及组合物的工程化以及优化”，分 别提交于2013年1月30日;2013年2月25;2013年3月15;2013年3月 28;2013年4月20;2013年5月6日以及2013年5月28日。还参考了美国 临时专利申请61/835,936、61/836,127、61/836,101、61/836,080、61/836,123 以及61/835,973，这些临时专利申请各自提交于2013年6月17日。还参考了 美国临时专利申请61/842,322以及美国专利申请14/054,414，这些临时专利申 请各自具有博德参考号BI-2011/008A，标题为“用于改变基因产物的表达的 CRISPR-Cas系统及方法”，分别提交于2013年7月2日和2013年10月15 日。

[0003] 前述申请、以及在其中或在它们的审查程序期间引用的所有文献 (“申请引用文献”)以及在这些申请引用文献中引用或参考的所有文献、以及在本文中引用或参考的所有文献(“本文引用的文献”)、在本文中引 用的文献中引用或参考的所有文献，连同针对在本文中提及或通过引用结合 在本文中的任何文献中的任何产品的任何制造商的说明书、说明、产品规 格、和产品表，特此通过引用并入本文，并且可以在本发明的实践中采用。更具体地说，所有参考的文献均通过引用并入本文，其程度如同每个单独的 文献被确切地并单独地指明通过引用而并入本文。

发明领域

[0004] 本发明总体上涉及用于控制涉及序列靶向的基因表达的系统、方 法、和组合物，该序列靶向例如可使用涉及规律间隔成簇短回文重复序列 (CRISPR)及其组分的载体系统的、基因组微扰或基因编辑。

关于联邦资助研究的声明

[0005] 本发明是在美国国立卫生研究院授予的NIH先锋奖(Pioneer Award)DP1MH100706的政府支持下完成的。美国政府享有本发明的某些权 利。

发明背景

[0006] 在基因组测序技术和分析方法中的最新进展显著加速了对与范围广 泛的生物功能和疾病相关联的遗传因子进行编目和映射的能力。精确的基因 组靶向技术对于通过允

许个体遗传元件的选择性干扰而使得因果性遗传变异 的统性逆向工程成为可能、以及推进合成生物学、生物技术应用、和医学应 用是需要的。虽然基因组编辑技术,如设计师锌指、转录激活子样效应因子 (TALE)、或归巢大范围核酸酶(homing meganuclease)对于产生靶向的基 因组干扰是可得的,但是仍然需要负担得起的、易于建立的、可扩展的、并 且便于靶向真核基因组内的多个位置的新的基因组工程技术。

发明概述

[0007] 对于具有一系列广泛应用的序列靶向的替代性且稳健的系统和技术 存在着迫切需要。本发明着手解决这种需要并且提供了相关优点。CRISPR/Cas或CRISPR-Cas系统(两个术语在本申请通篇可互换地使用)不 要求产生靶向特异性序列的定制蛋白,而是通过一个短RNA分子识别一个特 异性DNA靶标,可以对单个Cas酶进行程序设计,换句话说可以使用所述短 RNA分子将Cas酶募集到一个特异性DNA靶标。将该CRISPR-Cas系统添加 到基因组测序技术和分析方法的组库中可以显著使该方法论简化并且提高对 与范围广泛的生物功能和疾病相关联的遗传因子进行编目和映射的能力。为 了有效而无有害作用地利用CRISPR-Cas系统用于基因组编辑,了解这些基因 组工程工具的工程化和优化方面是至关重要的,这些方面是请求保护的本发 明的方面。

[0008] 在一个方面,本发明提供了包括一个或多个载体的载体系统。在一 些实施例中,该系统包括:(a)一个第一调节元件,该第一调节元件可操作 地连接到一个tracr配对序列以及用于在该tracr配对序列的上游插入一个或多 个指导序列的一个或多个插入位点,其中当表达时,该指导序列引导CRISPR 复合物在真核细胞中与一个靶序列的序列特异性结合,其中该CRISPR复合 物包括与以下各项复合的一种CRISPR酶:(1)杂交到该靶序列的指导序 列,以及(2)杂交到该tracr序列的tracr配对序列;以及(b)一个第二调节 元件,该第二调节元件可操作地连接到包括一个核定位序列的、编码所述 CRISPR酶的酶编码序列;其中组分(a)以及(b)位于该系统的相同或不同 载体上。在一些实施例中,组分(a)进一步包括在该第一调节元件的控制之 下在tracr配对序列的下游的tracr序列。在一些实施例中,组分(a)进一步 包括可操作地连接到该第一调节元件的两个或更多个指导序列,其中当表达 时,该两个或更多个指导序列中的每个引导CRISPR复合物在真核细胞中与 一个不同靶序列的序列特异性结合。在一些实施例中,该系统包括在一个第 三调节元件(诸如一个聚合酶III启动子)控制之下的tracr序列。在一些实施 例中,当进行最佳比对时,沿着该tracr配对序列的长度,该tracr序列展现至 少50%、60%、70%、80%、90%、95%、或99%序列互补性。确定最佳比对 在本领域的普通技术人员的能力范围内。例如,存在公开和可商购的比对算 法和程序,诸如但不限于ClustalW、matlab中的史密斯-沃特曼算法(Smith-Waterman)、Bowtie、Geneious、Biopython以及SeqMan。在一些实施例 中,该CRISPR复合物包括一个或多个核定位序列,该一个或多个核定位序 列具有足够强度来在真核细胞的细胞核中驱动所述CRISPR复合物以可检测 到的量积聚。不希望受理论限制,认为核定位序列对于真核生物中的CRISPR 复合物活性不是必要的,但包括此类序列增强该系统的活性,尤其对于靶向 细胞核中的核酸分子而言。在一些实施例中,该CRISPR酶是II型CRISPR 系统酶。在一些实施例中,该CRISPR酶是Cas9酶。在一些实施例中,该 Cas9酶是肺炎链球菌、化脓链球菌、或嗜热链球菌Cas9,并且可包括源自于 这些生物体的突变的Cas9。该酶可以是一种Cas9同系物或直向同源物。在一 些实施例中,该CRISPR酶是密码子优化的以便在真核

细胞中表达。在一些 实施例中，该CRISPR酶引导在该靶序列位置处的一条或两条链的切割。在 一些实施例中，该CRISPR酶缺少DNA链切割活性。在一些实施例中，该第 一调节元件是一种聚合酶III启动子。在一些实施例中，该第二调节元件是一 种聚合酶II启动子。在一些实施例中，该指导序列在长度上是至少15个、16 个、17个、18个、19个、20个、25个核苷酸，或10-30个之间、或15-25个 之间、或15-20个之间的核苷酸。通常，并且贯穿本说明书，术语“载体”是 指一种核酸分子，它能够运送与其连接的另一种核酸分子。载体包括但不限 于，单链、双链、或部分双链的核酸分子；包括一个或多个自由端、无自由 端(例如环状的)的核酸分子；包括DNA、RNA、或两者的核酸分子；以及 本领域已知的其他多种多样的多核苷酸。一种类型的载体是“质粒”，其是 指其中可以例如通过标准分子克隆技术插入另外的DNA片段的环状双链 DNA环。另一种类型的载体是病毒载体，其中病毒衍生的DNA或RNA序列 存在于用于包装病毒(例如，逆转录病毒、复制缺陷型逆转录病毒、腺病 毒、复制缺陷型腺病毒、以及腺相关病毒)的载体中。病毒载体还包含由用 于转染到一种宿主细胞中的病毒携带的多核苷酸。某些载体(例如，具有细 菌复制起点的细菌载体和附加型哺乳动物载体)能够在它们被导入的宿主细 胞中自主复制。其他载体(例如，非附加型哺乳动物载体)在引入宿主细胞 后整合到该宿主细胞的基因组中，并且由此与该宿主基因组一起复制。而 且，某些载体能够指导它们可操作连接的基因的表达。这样的载体在此被称 为“表达载体”。在重组DNA技术中使用的普通表达栽体通常是质粒形式。

[0009] 重组表达载体可包含处于适合于在宿主细胞中的核酸表达的形式的 本发明的核酸，这意味着这些重组表达载体包含基于待用于表达的宿主细胞 而选择的一种或多种调节元件，所述调节元件可操作地连接至待表达的核酸 序列。在重组表达载体内，“可操作地连接”旨在表示感兴趣的核苷酸序列 以一种允许该核苷酸序列的表达的方式被连接至该一个或多个调节元件(例 如，处于一种体外转录/翻译系统中或当该载体被引入到宿主细胞中时，处于 该宿主细胞中)。

[0010] 术语“调节元件”旨在包括启动子、增强子、内部核糖体进入位点 (IRES)、和其他表达控制元件(例如转录终止信号，如多聚腺苷酸化信号 和多聚U序列)。这样的调节序列例如描述于戈德尔(Goeddel)，《基因表 达技术：酶学方法》(GENE EXPRESSION TECHNOLOGY:METHODS IN ENZYMOLOGY)185，学术出版社(Academic Press)，圣地亚哥(San Diego)，加利福尼亚州(1990)中。调节元件包括指导一个核苷酸序列在许 多类型的宿主细胞中的组成型表达的那些序列以及指导该核苷酸序列只在某 些宿主细胞中表达的那些序列(例如，组织特异型调节序列)。组织特异型 启动子可主要指导在感兴趣的期望组织中的表达，所述组织例如肌肉、神经 元、骨、皮肤、血液、特定的器官(例如肝脏、胰腺)、或特殊的细胞类型 (例如淋巴细胞)。调节元件还可以时序依赖性方式(如以细胞周期依赖性 或发育阶段依赖性方式)指导表达，该方式可以是或者可以不是组织或细胞 类型特异性的。在一些实施例中，一个载体包含一个或多个pol III启动子 (例如1、2、3、4、5、或更多个pol III启动子)、一个或多个pol II启动子 (例如1、2、3、4、5、或更多个pol II启动子)、一个或多个pol I启动子 (例如1、2、3、4、5、或更多个pol I启动子)、或其组合。pol III启动子的实例包括但不限于U6和H1启动子。pol II启动子的实例包括但不限于逆转录 劳斯肉瘤病毒(RSV)LTR启动子(任选地具有RSV增强子)、巨细胞病毒 (CMV)启动子(任选地具有CMV增强子)[参见，例如，波沙特 (Boshart)等人，《细胞》(Cell)41:521-530(1985)]、SV40启动

子、二 氢叶酸还原酶启动子、β-肌动蛋白启动子、磷酸甘油激酶(PGK)启动子、 和EF1α启动子。还被术语“调节元件”涵盖的是增强子元件,如WPRE; CMV增强子;在HTLV-I的LTR中的R-U5’片段《分子细胞生物学》(Mol.Cell.Biol.,第8(1)卷,第466-472页,1988);SV40增强子;以及 在兔β-珠蛋白的外显子2与3之间的内含子序列(《美国国家科学院院刊》(Proc.Natl.Acad.Sci.USA.),第78(3)卷,第1527-31页,1981)。本领 域技术人员将理解,表达载体的设计可取决于诸如待转化的宿主细胞的选 择、所希望的表达水平等因素。一种载体可以被引入到宿主细胞中而由此产 生转录物、蛋白质、或肽,包括由如本文所述的核酸编码的融合蛋白或肽 (例如,规律间隔成簇短回文重复序列(CRISPR)转录物、蛋白质、酶、其 突变体形式、其融合蛋白,等等)。

[0011] 有利的载体包括慢病毒以及腺相关病毒,并且也可选择此类型的载 体以靶向具体类型的细胞。

[0012] 在一个方面,本发明提供了包括一个调节元件的载体,该调节元件 可操作地连接到编码CRISPR酶的酶编码序列上,该CRISPR酶包含一个或多 个核定位序列。在一些实施例中,所述调节元件驱动真核细胞中CRISPR酶 的转录,使得所述CRISPR酶以可检测到的量在该真核细胞的细胞核中积 聚。在一些实施例中,该调节元件是一种聚合酶II启动子。在一些实施例 中,该CRISPR酶是II型CRISPR系统酶。在一些实施例中,该CRISPR酶 是Cas9酶。在一些实施例中,该Cas9酶是肺炎链球菌、化脓链球菌或嗜热 链球菌Cas9,并且可包括源自于这些生物体的突变的Cas9。在一些实施例 中,该CRISPR酶是密码子优化的以便在真核细胞中表达。在一些实施例 中,该CRISPR酶引导在该靶序列位置处的一条或两条链的切割。在一些实 施例中,该CRISPR酶缺少DNA链切割活性。

[0013] 在一个方面,本发明提供了一种CRISPR酶,该CRISPR酶包括一 个或多个核定位序列,该一个或多个核定位序列具有足够强度来在真核细胞 的细胞核中驱动所述CRISPR酶以可检测到的量积聚。在一些实施例中,该 CRISPR酶是II型CRISPR系统酶。在一些实施例中,该CRISPR酶是Cas9 酶。在一些实施例中,该Cas9酶是肺炎链球菌、化脓链球菌或嗜热链球菌 Cas9,并且可包括源自于这些生物体的突变的Cas9。该酶可以是一种Cas9同 系物或直向同源物。在一些实施例中,该CRISPR酶缺少切割它结合至其上 的靶序列的一条或多条链的能力。

[0014] 在一个方面,本发明提供了一种真核宿主细胞,该真核宿主细胞包 括:(a)一个第一调节元件,该第一调节元件可操作地连接到一个tracr配对 序列以及用于在该tracr配对序列的上游插入一个或多个指导序列的一个或多 个插入位点,其中当表达时,该指导序列引导CRISPR复合物在真核细胞中 与一个靶序列的序列特异性结合,其中该CRISPR复合物包括与以下各项复 合的一种CRISPR酶:(1)杂交到该靶序列的指导序列,以及(2)杂交到该tracr序列的tracr配对序列;和/或(b)一个第二调节元件,该第二调节元 件可操作地连接到包括一个核定位序列的、编码所述CRISPR酶的酶编码序 列。在一些实施例中,该宿主细胞包括组分(a)以及(b)。在一些实施例 中,组分(a)、组分(b)、或组分(a)和(b)稳定地整合到该宿主真核 细胞的一个基因组中。在一些实施例中,组分(a)进一步包括在该第一调节 元件的控制之下在tracr配对序列的下游的tracr序列。在一些实施例中,组分 (a)进一步包括可操作地连接到该第一调节元件的两个或更多个指导序列, 其中当表达时,该两个或更多个指导序列中的每个引导CRISPR复合物在真 核细胞中与一个不同靶序列的序列特

异性结合。在一些实施例中,该真核宿 主细胞进一步包括一种第三调节元件,诸如一种聚合酶III启动子,该第三调 节元件可操作地连接到所述tracr序列。在一些实施例中,当进行最佳比对 时,沿着该tracr配对序列的长度,该tracr序列展现至少50%、60%、70%、80%、90%、95%、或99%序列互补性。在一些实施例中,该CRISPR酶包括 一个或多个核定位序列,该一个或多个核定位序列具有足够强度来在真核细 胞的细胞核中驱动所述CRISPR酶以可检测到的量积聚。在一些实施例中, 该CRISPR酶是II型CRISPR系统酶。在一些实施例中,该CRISPR酶是 Cas9酶。在一些实施例中,该Cas9酶是肺炎链球菌、化脓链球菌或嗜热链球 菌Cas9,并且可包括源自于这些生物体的突变的Cas9。该酶可以是一种Cas9同系物或直向同源物。在一些实施例中,该CRISPR酶是密码子优化的以便 在真核细胞中表达。在一些实施例中,该CRISPR酶引导在该靶序列位置处 的一条或两条链的切割。在一些实施例中,该CRISPR酶缺少DNA链切割活 性。在一些实施例中,该第一调节元件是一种聚合酶III启动子。在一些实施 例中,该第二调节元件是一种聚合酶II启动子。在一些实施例中,该指导序 列在长度上是至少15个、16个、17个、18个、19个、20个、25个核苷酸,或10-30个之间、或15-25个之间、或15-20个之间的核苷酸。在一个方面, 本发明提供一种非人的真核生物;优选地一种多细胞真核生物,包括根据所 述实施例中任一个的一种真核宿主细胞。在其他方面,本发明提供一种真核 生物;优选地一种多细胞真核生物,包括根据所述实施例中任一个的一种真 核宿主细胞。在这些方面的一些实施例中,该生物体可以是一种动物;例如 一种哺乳动物。另外,该生物体可以是一种节肢动物,诸如一种昆虫。该生 物体也可以是一种植物。进一步,该生物体可以是一种真菌。

[0015] 在一个方面,本发明提供了一种试剂盒,该试剂盒包括在此所述的 组分中的一种或多种。在一些实施例中,该试剂盒包括一种载体系统以及用 于使用该试剂盒的说明书。在一些实施例中,该载体系统包括:(a)一个第 一调节元件,该第一调节元件可操作地连接到一个tracr配对序列以及用于在 该tracr配对序列的上游插入一个或多个指导序列的一个或多个插入位点,其 中当表达时,该指导序列引导CRISPR复合物在真核细胞中与一个靶序列的 序列特异性结合,其中该CRISPR复合物包括与以下各项复合的一种CRISPR 酶:(1)杂交到该靶序列的指导序列,以及(2)杂交到该tracr序列的tracr 配对序列;和/或(b)一个第二调节元件,该第二调节元件可操作地连接到包 括一个核定位序列的编码所述CRISPR酶的酶编码序列。在一些实施例中, 该试剂盒包括位于该系统的相同或不同载体上的组分(a)和(b)。在一些 实施例中,组分(a)进一步包括在该第一调节元件的控制之下在tracr配对序 列的下游的tracr序列。在一些实施例中,组分(a)进一步包括可操作地连接到该第一调节元件的两个或更多个指导序列,其中当表达时,该两个或更多 个指导序列中的每个引导CRISPR复合物在真核细胞中与一个不同靶序列的 序列特异性结合。在一些实施例中,该系统进一步包括一种第三调节元件, 诸如一种聚合酶III启动子,该第三调节元件可操作地连接到所述tracr序列。 在一些实施例中,当进行最佳比对时,沿着该tracr配对序列的长度,该tracr 序列展现至少50%、60%、70%、80%、90%、95%、或99%序列互补性。在 一些实施例中,该CRISPR酶包括一个或多个核定位序列,该一个或多个核 定位序列具有足够强度来在真核细胞的细胞核中驱动所述CRISPR酶以可检 测到的量积聚。在一些实施例中,该CRISPR酶是II型CRISPR系统酶。在 一些实施例中,该CRISPR酶是Cas9酶。在一些实施例中,该Cas9酶是肺 炎链球菌、化脓链球菌或嗜热链球菌Cas9,并且可包括源自于

这些生物体的 突变的Cas9。该酶可以是一种Cas9同系物或直向同源物。在一些实施例中，该CRISPR酶是密码子优化的以便在真核细胞中表达。在一些实施例中，该 CRISPR酶引导在该靶序列位置处的一条或两条链的切割。在一些实施例中， 该CRISPR酶缺少DNA链切割活性。在一些实施例中，该第一调节元件是一 种聚合酶III启动子。在一些实施例中，该第二调节元件是一种聚合酶II启动 子。在一些实施例中，该指导序列在长度上是至少15个、16个、17个、18 个、19个、20个、25个核苷酸，或10-30个之间、或15-25个之间、或15-20 个之间的核苷酸。

[0016] 在一个方面，本发明提供了修饰在一种真核细胞中的靶多核苷酸的 方法。在一些实施例中，该方法包括允许一种CRISPR复合物结合到该靶多 核苷酸上以实施所述靶多核苷酸的切割，由此修饰该靶多核苷酸，其中该 CRISPR复合物包括与杂交到在所述靶多核苷酸内的一个靶序列上的指导序列 复合的CRISPR酶，其中所述指导序列连接到一个tracr配对序列上，该tracr 配对序列进而杂交到一个tracr序列上。在一些实施例中，所述切割包括通过 所述CRISPR酶切割在该靶序列位置的一条或两条链。在一些实施例中，所 述切割导致靶基因的转录降低。在一些实施例中，该方法进一步包括通过与 外源模板多核苷酸同源重组修复所述切割的靶多核苷酸，其中所述修复导致 一种突变，包括所述靶多核苷酸的一个或多个核苷酸的插入、缺失、或取 代。在一些实施例中，所述突变导致在从包含该靶序列的基因表达的蛋白质 中的一个或多个氨基酸改变。在一些实施例中，该方法进一步包括将一个或 多个载体递送到所述真核细胞，其中该一个或多个载体驱动以下一项或多项 的表达：该CRISPR酶、连接到该tracr配对序列上的指导序列、以及该tracr 序列。在一些实施例中，所述载体被递送到受试者内的真核细胞中。在一些 实施例中，所述修饰发生在细胞培养物中的所述真核细胞中。在一些实施例 中，该方法进一步包括在所述修饰之前从受试者中分离所述真核细胞。在一 些实施例中，该方法进一步包括使所述真核细胞和/或从中衍生的细胞返回到 所述受试者中。

[0017] 在一个方面，本发明提供了修饰一种多核苷酸在真核细胞中的表达 的方法。在一些实施例中，该方法包括允许一种CRISPR复合物结合到该多 核苷酸上，使得所述结合导致所述多核苷酸表达增加或减少；其中该CRISPR 复合物包含与杂交到在所述靶多核苷酸内的一个靶序列上的指导序列复合的 CRISPR酶，其中所述指导序列连接到一个tracr配对序列上，该tracr配对序 列进而杂交到一个tracr序列上。在一些实施例中，该方法进一步包括将一个 或多个载体递送到所述真核细胞，其中该一个或多个载体驱动以下一项或多 项的表达：该CRISPR酶、连接到该tracr配对序列上的指导序列、以及该 tracr序列。

[0018] 在一个方面，本发明提供了一种产生包含突变的疾病基因的模型真 核细胞的方法。在一些实施例中，疾病基因是与患有或产生一种疾病的风险 的增加相关联的任何基因。在一些实施例中，该方法包括(a)将一个或多个 载体引入真核细胞中，其中该一个或多个载体驱动以下一项或多项的表达： CRISPR酶、连接到tracr配对序列上的指导序列、以及tracr序列；以及(b) 允许一种CRISPR复合物结合到一个靶多核苷酸上以实施在所述疾病基因内 的该靶多核苷酸的切割，其中该CRISPR复合物包含与(1)杂交到该靶多核 苷酸内的该靶序列上的指导序列、和(2)杂交到该tracr序列上的该tracr配 对序列复合的CRISPR酶，由此产生包含突变的疾病基因的模式真核细胞。 在一些实施例中，所述切割包括通过所述CRISPR酶切割在该靶序列位置的 一条或两条链。在一些实施例中，所述切割导致靶基

因的转录降低。在一些 实施例中，该方法进一步包括通过与外源模板多核苷酸同源重组修复所述切 割的靶多核苷酸，其中所述修复导致一种突变，包括所述靶多核苷酸的一个 或多个核苷酸的插入、缺失、或取代。在一些实施例中，所述突变导致在从 包含该靶序列的基因表达的蛋白质中的一个或多个氨基酸改变。

[0019] 在一个方面，本发明提供了一种用于研发生物活性剂的方法，该生 物活性剂调制与疾病基因相关的细胞信号传导事件。在一些实施例中，疾病 基因是与患有或产生一种疾病的风险的增加相关联的任何基因。在一些实施 例中，该方法包括(a)使一种测试化合物与所述实施例中的任一项的一种模 式细胞接触；并且(b)检测读数变化，该变化指示与所述疾病基因的所述突 变关联的细胞信号传导事件的减少或增加，从而开发调制与所述疾病基因关 联的所述细胞信号传导事件的所述生物活性剂。

[0020] 在一个方面，本发明提供包括在tracr配对序列上游的指导序列的 一种重组多核苷酸，其中当表达时该指导序列引导一种CRISPR复合物与存 在于真核细胞中的一个对应的靶序列的序列特异性结合。在一些实施例中， 该靶序列是存在于真核细胞中的病毒序列。在一些实施例中，该靶序列是原 癌基因或癌基因。

[0021] 在一个方面，本发明提供了一种通过在一个或多个原核细胞中的基 因中引入一个或多个突变来选择一个或多个原核细胞的方法，该方法包括：将一个或多个载体引入一个或多个原核细胞中，其中该一个或多个载体驱动 以下一项或多项的表达：CRISPR酶、连接到tracr配对序列上的指导序列、 tracr序列、以及编辑模板；其中该编辑模板包含消除CRISPR酶切的一个或 多个突变；允许该编辑模板与在有待选择的一个或多个细胞中的靶多核苷酸 同源重组；允许CRISPR复合物结合到靶多核苷酸上以实施在所述基因内的 靶多核苷酸的切割，其中该CRISPR复合物包含与(1)杂交到该靶多核苷酸 内的靶序列上的指导序列、和(2)杂交到该tracr序列上的tracr配对序列复 合的CRISPR酶，其中该CRISPR复合物到该靶多核苷酸上的结合诱导细胞死 亡，由此允许选择其中已经引入一个或多个突变的一个或多个原核细胞。在 一个优选的实施例中，该CRISPR酶是Cas9。在本发明的另一方面，该有待 选择的细胞可以是真核细胞。本发明的方面允许选择特异细胞，而不需要选 择标记或可能包括反选择系统的两步法。

[0022] 因此，本发明的目的在于，在本发明中不涵盖任何先前已知的产 品、制造该产品的过程、或使用该产品的方法，使得申请人保留和特此披露 放弃任何先前已知的产品、过程、或方法的权利。进一步指出的是，在本发 明的范围之内，本发明并不旨在涵盖任何产品、过程、或该产品的制造或使 用该产品的方法，其不符合USPTO(35U.S.C.§112，第一段)或EPO(EPC 的第83条)的书面说明和可实施性要求，使得申请人保留和特此公开放弃任 何先前描述的产品、制造该产品的过程、或使用该产品的方法的权利。

[0023] 应当指出的是，在本披露中并且尤其是在权利要求书和/或段落 中，术语如“包含(comprises)”、“包含(comprised)”、“包含 (comprising)”等等可具有在美国专利法中属于它的含义；例如，它们可以 表示“包括(includes)”、“包括(included)”、“包括(including)”等 等；并且术语如“基本上由……组成(consisting essentially of)”和“基本上 由……组成(consists essentially of)”具有在美国专利法中归于它们的含义，例如，它们允许未被清楚叙述的要素，但是排除在现有技术中发现或者影响 本发明的基本或新颖特征的要素。这些和其他实施例披露于以下详细说明 中，或者据其显而易见并且由

其涵盖。

附图简要说明

[0024] 本发明的新颖特征在所附权利要求书中具体提出。通过参考提出说 明性实施例的以下详细说明,将获得对本发明的特征和优点的更好理解,在 这些实施例中利用了本发明的原理,并且在这些附图中:

[0025] 图1显示了该CRISPR系统的示意模型。来自化脓链球菌(黄色) 的Cas9核酸酶经由合成的指导RNA(sgRNA)而被靶向基因组DNA,该指 导RNA包含一个20个nt的指导序列(蓝色)和一个支架(红色)。具有 DNA靶(蓝色)的指导序列碱基对,直接地在必要的5'-NGG原型间隔子相 邻基序(PAM;红紫色)的上游,并且Cas9在该PAM(红色三角形)的上 游约3bp介导了双链断裂(DSB)。

[0026] 图2A-F显示了一个示例性CRISPR系统、一个可能的作用机制、 一个在真核细胞中的表达的示例性适应性改变、以及评估核定位和CRISPR 活性的试验的结果。

[0027] 图3展示了一个用于在真核细胞中表达CRISPR系统元件的示例性 表达盒、示例性指导序列的预测结构以及如在真核和原核细胞中测量的 CRISPR系统活性。

[0028] 图4A-D显示了针对示例性靶标的SpCas9特异性的评估结果。

[0029] 图5A-G显示了一个示例性载体系统以及它在指导真核细胞中的同 源重组中的使用结果。

[0030] 图6提供了原型间隔子序列的表,并且总结了基于示例性脓链球菌 和嗜热链球菌CRISPR系统设计的原型间隔子靶标的修饰效率结果,这些 CRISPR系统具有针对人类和小鼠基因组中的座位的相应PAM。用Cas9以及 前-crRNA/tracrRNA或嵌合RNA转染细胞,在转染72小时之后进行分析。基 于来自指示细胞系的Surveyor分析结果,计算了indel(插入缺失)百分比 (对于所有原型间隔子靶标N=3,误差为S.E.M.,N.D.表示使用Surveyor测 定未检测到,并且N.T.表示在本研究中未检测)。

[0031] 图7A-C显示了对于Cas9介导的基因靶向的不同tracrRNA转录本 的比较。

[0032] 图8显示了对于双链断裂诱导的微插入和微缺失的检测的测量员核 酸酶试验的示意图。

[0033] 图9A-B显示了用于CRISPR系统元件在真核细胞中的表达的示例 性双顺反子表达载体。

[0034] 图10显示一种细菌性质粒转化干扰测定、其中使用的表达盒和质 粒、以及其中使用的细胞转化率。

[0035] 图11A-C显示了在人类基因组中在邻近的化脓链球菌SF370座位1 PAM(NGG)(图10A)和嗜热链球菌LMD9座位2PAM(NNAGAAW) (图10B)之间的距离、以及对于每个PAM就染色体而言的距离(Chr) (图10C)的直方图。

[0036] 图12A-C显示了一个示例性CRISPR系统、一个在真核细胞中的表 达的示例性适应性改变、以及评估CRISPR活性的试验的结果。

[0037] 图13A-C显示了用于靶向哺乳动物细胞中的基因组座位的CRISPR 系统的示例性操纵。

[0038] 图14A-B显示了在哺乳动物中的crRNA加工的Northern印迹分析 的结果。

[0039] 图15显示了在人PVALB和小鼠Th座位中的原型间隔子的示例性 筛选。

[0040] 图16显示了在人EMX1座位中的嗜热链球菌CRISPR系统的示例 性原型间隔子和相应PAM序列靶标。

[0041] 图17提供了用于Surveyor、RFLP、基因组测序、以及Northern印 迹分析的引物和探针序列的表。

[0042] 图18A-C显示了具有嵌合RNA的CRISPR系统的示例性操纵以及 在真核细胞中的系统活性的SURVEYOR分析的结果。

[0043] 图19A-B显示了在真核细胞中的CRISPR系统活性的SURVEYOR 分析的结果的图示。

[0044] 图20显示了使用UCSC基因组浏览器进行的在人类基因组中的一 些化脓链球菌Cas9靶位点的示例性可视化。

[0045] 图21显示包括指导序列、tracr配对序列、以及tracr序列的示例性 嵌合RNA的预测的二级结构。

[0046] 图22显示了用于CRISPR系统元件在真核细胞中的表达的示例性 双顺反子表达载体。

[0047] 图23显示针对内源靶的Cas9核酸酶活性可被用于基因组编辑。(a)使用该CRISPR系统的基因组编辑的概念。该CRISPR靶向构建体引导 一种染色体座位的切割,并且与和该靶标重组的一个编辑模板共转化以防止 切割。从CRISPR攻击幸存的卡那霉素抗性转化体包含通过该编辑模板引入 的修饰。tracr,反式激活CRISPR RNA;aphA-3,卡那霉素抗性基因。(b) 在无编辑模板、R6野生型srtA或R6370.1编辑模板情况下在$R6^{8232.5}$细胞中进 行的crR6M DNA的转化。R6 srtA或$R6^{370.1}$的重组通过Cas9防止切割。转化 率计算为集落形成单位(cfu)/μg的crR6M DNA;显示了来自至少三个独立 实验的平均值与标准差。在每个转化中在8个克隆上进行PCR分析。“Un.” 表明菌株$R6^{8232.5}$的未编辑的srtA座位;“Ed.”显示编辑模板。$R6^{8232.5}$和 $R6^{370.1}$靶标通过EaeI的限制酶切而区分开。

[0048] 图24显示对消除Cas9切割的PAM以及种子序列的分析。(a)具 有随机化PAM序列或随机化种子序列的PCR产物在crR6细胞中转化。这些 细胞表达装载有crRNA(靶向$R6^{8232.5}$细胞的染色体区域(以粉红色凸显)) 的Cas9,该染色体区域自该R6基因组缺失。合并超过2×105种氯霉素抗性 转化体(携带无活性的PAM或种子序列),用于对该靶区进行扩增以及深度 测序。(b)在crR6细胞中随机PAM构建体的转化之后读数数目的相对比例 (相较于R6转化体中的读数的数目)。显示了每个3-核苷酸PAM序列的相 对丰度。以红色示出代表性严重不足的序列(NGG);代表性部分地不足的 序列以橘色示出(NAG)(c)在crR6细胞中随机种子序列构建体的转化之 后读数数目的相对比例(相较于R6转化体中的读数的数目)。显示了针对该 原型间隔子的前20个核苷酸的每个位置的每个核苷酸的相对丰度。高丰度表 明缺少由Cas9进行的切割,即一种CRISPR钝化突变。灰色线显示野生型序 列的水平。虚线代表如下水平,在此水平之上一种突变显著破坏切割(参见 实例5中的“深度测序数据分析”部分)

[0049] 图25显示在肺炎链球菌中使用CRISPR系统引入单个和多个突 变。(a)野生型和编辑的(绿色核苷酸;加下划线的氨基酸残基)bgaA的核 苷酸以及氨基酸序列。显示了该原型间隔子、PAM以及限制酶切位点。 (b)在存在一种编辑模板或对照的情况下用靶向构建体转化的细胞的转化 率。(c)针对每个编辑试验的8个转化体进行PCR分析,随后用BtgZI (R→A)以及TseI(NE→AA)进行消化。bgaA的缺失显示为一种较小的 PCR产物。(d)进行米勒

(Miller)测定以测量野生型和编辑的菌株的β-半 乳糖苷酶的活性。(e)对于单步骤、双向缺失,该靶向构建体包含两个间隔 子(在此情况下匹配srtA以及bgaA)并且用两种不同的编辑模板共转化。 (f)PCR分析8个转化体以检测srtA以及bgaA座位的缺失。6/8的转化体包含两个基因的缺失。

[0050] 图26提供使用该CRISPR系统的编辑下潜在的机制。(a)将一个 终止密码子引入红霉素抗性基因ermAM以产生菌株JEN53。该野生型序列可 以通过以下修复:用该CRISPR::ermAM(终止子)构建体靶向该终止密码 子,并且使用ermAM野生型序列作为编辑模板。(b)突变体以及野生型 ermAM序列。(c)从总或卡那霉素抗性(kan^R)cfu计算的红霉素抗性(erm^R)cfu的分数。(d)获得该CRISPR构建体以及该编辑模板两者的总 细胞的分数。靶向构建体的CRISPR的共转化产生更多转化体(t检验,p= 0.011)。在所有情况下,针对三个独立实验,这些值显示平均±s.d.。

[0051] 图27说明在大肠杆菌中用该CRISPR系统编辑基因组。(a)携带 靶向待编辑的基因的CRISPR阵列(pCRISPR)的一个卡那霉素抗性质粒可 在包含一个氯霉素抗性质粒(包含cas9以及tracr(pCas9))连同带有指定 该突变的一个寡核苷酸的HME63重组工程菌株中被转化。(b)将赋予链霉 素抗性的一个K42T突变引入rpsL基因(c)从总或卡那霉素抗性(kan^R)cfu 计算的链霉素抗性($strep^R$)cfu的分数。(d)获得该pCRISPR质粒以及编辑 寡核苷酸两者的总细胞的分数。靶向质粒的pCRISPR的共转化产生更多转化 体(t检验,p=0.004)。在所有情况下,针对三个独立实验,这些值显示平 均±s.d.。

[0052] 图28示出了crR6基因组DNA的转化导致靶定的座位的编辑(a) 肺炎链球菌R6的IS1167元件被化脓链球菌SF370的CRISPR01座位替代以 产生crR6菌株。此基因座编码Cas9核酸酶,即具有六个间隔子的一个 CRISPR阵列,crRNA生源说所需要的tracrRNA,以及Cas1、Cas2和Csn2 (对于靶向来说非必要的蛋白)。菌株crR6M包含没有cas1、cas2以及csn2的一个最低限度的功能性的CRISPR系统。aphA-3基因编码卡那霉素抗性。来自链球菌的细菌噬菌体Φ8232.5以及Φ370.1的原型间隔子融合到一个氯霉 素抗性基因(cat)并且整合到菌株R6的srtA基因,以产生菌株R68232.5以 及R6370.1。(b)左小图:crR6以及crR6M基因组DNA在$R6^{8232.5}$和$R^{6370.1}$中的转化。作为细胞感受态的一个对照,也转化了一个链霉素抗性基因。右 小图:具有crR6基因组DNA的的$8R6^{8232.5}$转化体的PCR分析。扩增srtA座 位的引物被用于PCR。7/8的基因分型的菌落通过来自crR6基因组DNA的 WT座位替代R68232.5 srtA座位。

[0053] 图29提供在此研究中获得的编辑的细胞的DNA序列的色谱图。在 所有情况下,指明了野生型和突变的原型间隔子以及PAM序列(或其反向互 补体)。当相关时,提供了由该原型间隔子编码的氨基酸序列。对于每个编 辑试验,对如下所有菌株进行测序,对这些菌株而言PCR以及限制酶切分析 证实引入了所希望的修饰。显示了代表性色谱图。(a)将一种PAM突变引 入至该$R6^{8232.5}$靶中的色谱图(图24a)。(b)将R>A以及NE>AA突变引 入至β-半乳糖苷酶(bgaA)中的色谱图(图25c)。(c)将6664bp缺失引 入bgaA ORF内的色谱图(图25c以及25f)。虚线表示缺失限。(d)将729 bp缺失引入srtA ORF内的色谱图(图25f)。虚线表示缺失限。(e)ermAM 内一个早熟终止密码子的产生的色谱图(图33)。(f)大肠杆菌中的rpsL 编辑(图27)。

[0054] 图30示出针对包含不同PAM的随机肺炎链球菌靶的CRISPR免 疫。(a)肺炎链球菌

R6基因组上10个随机的靶的位置。所选择的靶具有不 同PAM并且在两条链上。(b)将对应于这些靶的间隔子以最低限度 CRISPR阵列在质粒pLZ12上克隆并且转化至菌株crR6Rc中,该菌株提供反 向加工以及靶向机构。(c)菌株R6以及crR6Rc中不同质粒的转化率。在 crR6Rc中转化pDB99-108(T1-T10)未重新获得菌落。虚线表示该测定的检 出限。

[0055] 图31提供针对靶向的基因组编辑的一种通用方案。为了促进靶向 的基因组编辑,crR6M被进一步工程化为包含tracrRNA、Cas9以及仅一个重 复的CRISPR阵列,随后是卡那霉素抗性标记物(aphA-3),以产生菌株 crR6Rk。来自此菌株的DNA被用作PCR的模板,引物被设计为引入一个新 的间隔子(用N指定的绿框)。使用吉布森(Gibson)法拼接左和右PCR以 建立靶向构建体。然后将靶向以及编辑构建体两者转化至菌株crR6Rc中,该 菌株crR6Rc是一种等效于crR6Rk的菌株,但具有的卡那霉素抗性标记物被 一种氯霉素抗性标记物(cat)替代。约90%的卡那霉素抗性转化体包含所希 望的突变。

[0056] 图32说明PAM之间的距离分布。NGG以及CCN被认为是有效的 PAM。显示了针对该肺炎链球菌R6基因组连同具有相同长度和相同GC-含量 (39.7%)的一个随机序列的数据。虚线代表R6基因组中的PAM之间的平 均距离(12)。

[0057] 图33示出了使用基因组DNA作为靶向构建体进行的ermAM座位 的CRISPR介导的编辑。为了使用基因组DNA作为靶向构建体,有必要避免 CRISPR自身免疫,并且因此必须使用针对不存在于该染色体中的一个序列 (在此情况下,是ermAM红霉素抗性基因)的一个间隔子。(a)野生型以 及突变(用红字标明的)ermAM基因的核苷酸以及氨基酸序列。显示了该原型间隔子以及PAM序列。(b)使用基因组DNA进行的ermAM座位的 CRISPR介导的编辑的示意图。通过PCR以及吉布森拼接(Gibson Assembly)制备携带一个ermAM-靶向间隔子(蓝框)的一个构建体,并且转 化至菌株crR6Rc中,以产生菌株JEN37。然后使用JEN37的基因组DNA作为靶向构建体,并且用该编辑模板共转化至JEN38中,该JEN38是这样一种 菌株,其中该srtA基因被ermAM的一个野生型拷贝替代。卡那霉素抗性转化 体包含编辑的基因型(JEN43)。(c)共转化靶向以及编辑或对照模板之后 获得的卡那霉素抗性细胞的数目。在存在对照模板的情况下,获得5.4×10^3 cfu/ml,并且当使用该编辑模板时获得4.3×10^5cfu/ml。这个差异表明约99% 的编辑率[$(4.3\times10^5-5.4\times10^3)/4.3\times10^5$]。(d)为了检查编辑的细胞的存 在,将七个卡那霉素抗性克隆以及JEN38在有(erm+)或没有(erm-)红霉 素的琼脂板上划线。仅阳性对照显示对红霉素的抗性。还通过DNA测序核实 这些转化体中的一个的ermAM mut基因型(图29e)。

[0058] 图34示出了通过CRISPR介导的基因组编辑顺序引入突变。(a) 通过CRISPR介导的基因组编辑顺序引入突变的示意图。首先,R6被工程化 以产生crR6Rk。将crR6Rk用融合至cat的一种srtA-靶向构建体共转化以进行 对编辑的细胞的氯霉素筛选,连同针对ΔsrtA框内缺失用编辑构建体进行共转 化。菌株crR6ΔsrtA是通过氯霉素筛选产生的。随后,将该ΔsrtA菌株用融合 至aphA-3的一种bgaA-靶向构建体共转化以进行对编辑的细胞的卡那霉素筛 选,并且用包含ΔbgaA框内缺失的一个编辑构建体进行共转化。最终,工程 化的CRISPR座位可以通过以下步骤从该染色体清除,首先共转化包含野生 型IS1167座位的R6 DNA以及携带一个bgaA原型间隔子(pDB97)的质粒, 并且针对大观霉素进行筛选。(b)针对8个氯霉素(Cam)抗性转化体进行 PCR分析以检测srtA座位的缺失。(c)通过Miller测定测量的β-半乳糖苷酶 活性。在肺炎链球菌中,此酶通过分选酶A被锚定至细胞壁。该srtA基因的

缺失导致β-半乳糖苷酶释放至上清液中。Δ bgaA突变体不显示活性。(d)针 对8个大观霉素(Spec)抗性转化体进行PCR分析以检测野生型IS1167对 CRISPR座位的替代。

[0059] 图35示出了肺炎链球菌中CRISPR的背景突变频率。(a)JEN53 中CRISPR::Ø或CRISPR::erm(终止子)靶向构建体的转化,使用或不使用 ermAM编辑模板。CRISPR::Ø以及CRISPR::erm(终止子)之间的kan^R CFU 的差异表明Cas9切割杀死未编辑的细胞。以3×10^{-3}的频率观察到在不存在 编辑模板的情况下逃脱CRISPR干扰的突变体。(b)逃脱者的CRISPR座位 的PCR分析显示7/8具有间隔子缺失。(c)逃脱者#2在cas9中携带点突 变。

[0060] 图36示出了使用pCas9,化脓链球菌CRISPR座位1的必需元件在 大肠杆菌中重组。该质粒包含tracrRNA、Cas9、连同驱动该crRNA阵列的一 个前导子序列。这些pCRISPR质粒仅包含前导子以及阵列。使用退火的寡核 苷酸,间隔子可被插入至该crRNA阵列中BsaI位点之间。寡核苷酸设计显示 在底部。pCas9携带氯霉素抗性(CmR)并且是基于低拷贝pACYC184质粒 骨架的。pCRISPR是基于高拷贝数pZE21质粒。需要两个质粒,因为使用此 生物体作为克隆宿主,如果Cas9也存在的话(它将杀死该宿主),不可构建 包含靶向大肠杆菌染色体的间隔子的pCRISPR质粒。

[0061] 图37示出了大肠杆菌MG1655中CRISPR引导的编辑。将携带赋 予链霉素抗性并且废除CRISPR免疫两者的点突变的一个寡核苷酸(W542) 和一种靶向rpsL(pCRISPR::rpsL)的质粒或一个对照质粒(pCRISPR::Ø)一 起共转化至包含pCas9的野生型大肠杆菌菌株MG1655中。在包含链霉素或 卡那霉素的培养基上筛选转化体。虚线表明该转化测定的检出限。

[0062] 图38示出了大肠杆菌HME63中CRISPR的背景突变频率。(a) pCRISPR::Ø或pCRISPR::rpsL质粒转化至HME63感受态细胞中。以2.6×10^{-4}的频率观察到逃脱CRISPR干扰的突变体。(b)逃脱者的CRISPR阵列的 扩增显示8/8缺失间隔子。

[0063] 图39A-D显示了揭示五个Cas9家族的系统发生分析的环状描绘，这五个家族包括三组大的Cas9(约1400个氨基酸)和两组小的Cas9(约 1100个氨基酸)。

[0064] 图40A-F显示了揭示五个Cas9家族的系统发生分析的线性描绘，这五个家族包括三组大的Cas9(约1400个氨基酸)和两组小的Cas9(约 1100个氨基酸)。

[0065] 图41A-M显示了在突变点位于SpCas9基因内的情况下的序列。

[0066] 图42显示如下的一个示意性构建体,其中转录激活结构域 (VP64)融合至Cas9,催化结构域中有两个突变(D10以及H840)。

[0067] 图43A-D显示了经由同源重组进行的基因组编辑。(a)SpCas9切 口酶的示意图,具有在RuvC I催化结构域中的D10A突变。(b)使用有义或 反义单链寡核苷酸作为修复模板的代表在人EMX1座位的同源重组(HR)的 示意图。以上红色箭头表示sgRNA切割位点;用于基因分型的PCR引物(表 J和K)表示为右图中的箭头。(c)由HR修饰的区域的序列。d,对于野生型(wt)和切口酶(D10A)SpCas9介导的在EMX1靶标1座位的indel的 SURVEYOR分析(n＝3)。箭头指示预期片段大小的位置。

[0068] 图44A-B显示了用于SpCas9的单个载体设计。

[0069] 图45显示NLS-Csn1构建体NLS-Csn1、Csn1、Csn1-NLS、NLS- Csn1-NLS、NLS-Csn1-GFP-NLS以及UnTFN的切割的量化。

[0070] 图46显示NLS-Cas9、Cas9、Cas9-NLS以及NLS-Cas9-NLS的索 引频率。

[0071] 图47显示一种凝胶，该凝胶表明具有切口酶突变的SpCas9(单独 地)不诱导双链断裂。

[0072] 图48显示在此试验中用作同源重组(HR)模板的寡DNA的一种 设计以及由Cas9蛋白以及HR模板的不同组合诱导的HR率的比较。

[0073] 图49A显示了条件型Cas9、Rosa26靶向载体图谱。

[0074] 图49B显示了组成型Cas9、Rosa26靶向载体图谱。

[0075] 图50A-H显示存在于图49A-B的载体图谱中的每个元件的序列。

[0076] 图51显示了在组成型和条件型Cas9构建体中的重要元件的示意 图。

[0077] 图52显示组成型和条件型Cas9构建体的表达的功能性确认。

[0078] 图53显示通过Surveyor确认Cas9核酸酶活性。

[0079] 图54显示Cas9核酸酶活性的量化。

[0080] 图55显示构建体设计以及同源重组(HR)策略。

[0081] 图56显示组成型(右)和条件型(左)构建体在两个不同凝胶暴 露时间(顶行持续3min并且底行持续1min)的基因组PCR基因分型结果。

[0082] 图57显示了在mESC中的Cas9激活。

[0083] 图58显示用以使用切口酶版本的Cas9连同两个指导RNA通过 NHEJ介导基因敲除的策略的示意图。

[0084] 图59显示DNA双链断裂(DSB)修复如何促进基因编辑。在易错 非同源末端连接(NHEJ)途径中，DSB的末端由内源DNA修复机构加工并 重新连接在一起，可导致在连接位点的随机indel(indel)突变。发生在基因 编码区内的indel突变可产生移码和提前终止密码子，导致基因敲除。可替代 地，处于质粒或单链寡脱氧核苷酸(ssODN)形式的修复模板可以被提供为 利用同源定向修复(HDR)途径，该途径允许高保真和精确编辑。

[0085] 图60显示实验的时间轴以及概要。针对试剂设计、构建、确认、 以及细胞系扩增的步骤。在计算机中(in silico)通过在线设计工具(在网站 genome-engineering.org/tools可获得)设计针对每个靶标的自定义sgRNA(浅 蓝色条)连同基因分型引物。然后将sgRNA表达载体克隆至包含Cas9 (PX330)的一个质粒中并且通过DNA测序进行证实。然后将完整的质粒 (pCRISPR)以及促进同源定向修复的任选的修复模板转染至细胞中并且针 对介导靶向切割的能力进行测定。最终，转染的细胞可以克隆性扩增以得到 具有定义的突变的同基因细胞系。

[0086] 图61A-C显示靶筛选以及试剂制备。(a)化脓链球菌Cas9，20- bp靶标(用蓝色凸显)后面必须有5’-NGG，其可发生在基因组DNA的任一 条链。我们建议使用此方案中所述的在线工具以辅助靶标筛选 (www.genome-engineering.org/tools)。(b)Cas9表达质粒(PX165)以及 PCR-扩增的U6-驱动sgRNA表达盒的共转染的示意图。使用包含U6启动子 的PCR模板以及固定的正向引物(U6 Fwd)，sgRNA-编码DNA可附加到该 U6反向引物(U6 Rev)上并且合成为延伸的DNA寡聚物(来自IDT的 Ultramer寡聚物)。注意，U6 Rev中的指导序列(蓝色N)是5’-NGG侧翼 靶序列的反向互补体。(c)指导序列寡聚物无疤(scarless)克隆至包含 Cas9以及sgRNA支架(PX330)的一个质粒中的示意图。指导寡聚物(蓝色 N)包含突出端以连接至PS330上的成对BbsI位点上，顶部和底部链定向匹 配该基因组靶标的那些(即，顶部寡聚物是基因组DNA中在5’-NGG之前的 20-bp序列)。用BbsI消化PX330允许II型限制

酶切位点(蓝色轮廓线)被 退火的寡聚物的同向插入替代。值得注意地是一个额外的G被放置在该指导 序列的第一个碱基之前。申请人已经发现在该指导序列的前面的一个额外的G不会不利地影响靶向率。当所选择的20-nt指导序列不以鸟嘌呤起始时,额 外的鸟嘌呤将确保该sgRNA被该U6启动子有效地转录,其优选是在该转录 物的第一碱基中的一个鸟嘌呤。

[0087] 图62A-D显示针对多元NHEJ的预期结果。(a)用以确定indel百 分数的SURVEYOR测定的示意图。首先,通过PCR扩增来自Cas9-靶细胞的 异质群体的基因组DNA。然后将扩增子缓慢重退火以产生异源双链体。将这 些重退火的异源双链体通过SURVEYOR核酸酶切割,而同源双链体保持完 整。基于切割的DNA的分数计算Cas9介导的切割率(%indel),如通过凝胶条带的累积强度确定的。(b)两个sgRNA(橙色以及蓝色条)被设计以靶 向人GRIN2B以及DYRK1A座位。SURVEYOR凝胶显示在转染细胞中两个 基因座处的修饰。着色的箭头表明针对每个基因座预测的片段尺寸。(c)一 对sgRNA(浅蓝以及绿色条)被设计以切除人EMX1座位中的外显子(深蓝 色)。靶序列以及PAM(红色)以各自的颜色示出,并且红色三角形表明切割位点。预测的接点如下示出。通过PCR(外Fwd,外Rev)测定从用 sgRNA 3、4、或两者转染的细胞群体分离的个体克隆,反映了约270-bp的缺 失。显示了没有修饰(12/23)、单等位基因修饰(10/23)以及双等位基因 (1/23)修饰的代表性克隆。内Fwd以及内Rev引物被用以针对倒位事件进 行筛选。(d)带有EMX1外显子缺失的克隆系的量化。两对sgRNA(3.1、 3.2左侧翼sgRNA;4.1、4.2右侧翼sgRNA)被用以介导一个EMX1外显子 周围可变尺寸的缺失。转染细胞被克隆性地分离并且扩增以针对缺失以及倒 位事件进行基因分型分析。筛选了105个克隆,其中分别有51(49%)以及 11(10%)携带杂合的以及纯合的缺失。由于接点可以是可变的,给出了大 概的缺失尺寸。

[0088] 图63A-C显示应用ssODN以及靶向载体以介导HR,HEK293FT 以及HUES9细胞中Cas9的野生型以及切口酶突变体两者效率范围是从1.0%- 27%。

[0089] 图64显示在哺乳动物细胞中用于快速且有效的CRISPR靶向的一 种基于PCR的方法的示意图。使用一个U6-特异性正向引物以及携带该U6 启动子的部分的反向互补体、具有指导序列的sgRNA(+85)支架、以及用 于转录终止的7个T核苷酸的一个反向引物,将包含人RNA聚合酶III启动 子U6的一个质粒进行PCR扩增。纯化所得到的PCR产物并且用携带CBh启 动子驱动的Cas9的一种质粒进行共递送。

[0090] 图65显示针对每个gRNA以及各自的对照,来自转基因组学公司 (Transgenomics)SURVEYOR突变检测试剂盒的结果。一个阳性 SURVEYOR结果是对应于基因组的PCR的一条大的条带以及两个较小的条 带,这两个较小的条带是使得双链在一个突变位点断裂的SURVEYOR核酸 酶的产物。在小鼠细胞系即Neuro-N2a中通过用hSpCas9进行的脂质体瞬时共转染验证每个gRNA。转染后72小时,使用来自Epicentre的QuickExtract DNA纯化基因组DNA。进行PCR以扩增感兴趣的基因座。

[0091] 图66显示针对38个活的幼兽(道1-38)、1个死的幼兽(道39) 以及用于比较的1个野生型幼兽(道40)的Surveyor结果。将幼兽1-19用 gRNA Chd8.2进行注射,并且将幼兽20-38用gRNA Chd8.3进行注射。该38 个活的幼兽中13个针对一种突变是阳性的。这个死的幼兽也具有一种突变。 在野生型样品中未检测到突变。基因组PCR测序与SURVEYOR测定发现是 一致的。

[0092] 图67显示不同Cas9 NLS构建体的一种设计。所有Cas9都是人-密 码子-优化版本的Sp Cas9。在N-末端或C-末端,NLS序列被连接至该cas9基 因。将具有不同NLS设计的所有Cas9变体克隆至包含EF1a启动子的一种骨 架载体中,所以它由EF1a启动子驱动。在相同的载体中,存在靶向人EMX1 座位的由U6启动子驱动的一种嵌合RNA,一起形成一个二组分系统。

[0093] 图68显示由具有不同NLS设计的Cas9变体诱导的基因组切割的 效率。该百分数表明被每个构建体切割的人EMX1基因组DNA的部分。所有 实验来自3个生物重复。n=3,误差表示为S.E.M.。

[0094] 图69A显示具有转录激活活性的CRISPR-TF(转录因子)的一种 设计。该嵌合RNA由U6启动子表达,而可操作地连接到三个NLS以及 VP64功能域的一种人-密码子-优化的双突变版本的Cas9蛋白(hSpCas9m) 由一种EF1a启动子表达。该双突变,即D10A以及H840A,导致当由嵌合 RNA指导时,cas9蛋白不能够引入任何切割但维持其结合到靶DNA的能 力。

[0095] 图69B显示用CRISPR-TF系统(嵌合RNA以及Cas9-NLS-VP64 融合蛋白)对人SOX2基因进行的转录激活。将293FT细胞用如下质粒转 染,所述质粒具有两个组分:(1)U6-驱动的不同嵌合RNA,靶向人SOX2 基因组座位内或周围的20-bp序列,以及(2)EF1a-驱动的hSpCas9m(双重 突变型)-NLS-VP64融合蛋白。转染后96小时,收获293FT细胞,并且通过 使用一种qRT-PCR测定引入mRNA表达以测量激活水平。将所有表达水平针 对对照组(灰色条)标准化,对照组表示来自用该CRISPR-TF骨架质粒(不 含嵌合RNA)转染的细胞的结果。用于检测SOX2 mRNA的qRT-PCR探针 是Taqman人基因表达测定用探针(生命技术公司(Life Technologies))。所有实验表示来自3个生物重复的数据,n=3,误差条显示为s.e.m.。

[0096] 图70描绘了针对SpCas9的NLS构造优化。

[0097] 图71显示针对NGGNN序列的QQ绘图。

[0098] 图72显示具有拟合正态分布(黑色线)以及.99分位数(虚线)的 数据密度的直方图。

[0099] 图73A-C显示dgRNA::cas9**表达的bgaA的RNA指导的阻遏。a. 该Cas9蛋白结合到该tracrRNA,并且结合到通过RNA酶III加工的前体 CRISPR RNA以形成crRNA。该crRNA引导Cas9与该bgaA启动子的结合并 且阻遏转录。b.表示了用以将Cas9**引导至该bgaA启动子的靶标。假定的- 35、-10连同该bgaA起始密码子是加粗的。c.通过米勒测定在不存在靶向的情 况下并且针对四种不同的靶标测量的β半乳糖苷酶活性。

[0100] 图74A-E显示Cas9**介导的阻遏的表征。a.gfpmut2基因及其启动 子(包括-35以及-10信号)是伴随着该研究使用的不同靶位点的位置一起示 出的。b.靶向编码链时的相对荧光。c.靶向非编码链时的相对荧光。d.对从 T5、T10、B10或不含靶标的对照菌株提取的RNA,用探针B477以及B478 进行Northern印迹。e.在B1、T5以及B10的crRNA的5'端中,增加数目的 突变的影响。

[0101] 本文的这些图仅仅是出于说明性的目的,并且不一定是按比例绘制 的。

本发明的详细说明

[0102] 术语“多核苷酸”、“核苷酸”、“核苷酸序列”、“核酸”和 “寡核苷酸”可互换地使用。它们是指具有任何长度的核苷酸的聚合形式, 是脱氧核糖核苷酸或核糖核苷酸、或其类似物。多核苷酸可具有任何三维结 构,并且可以执行已知或未知的任何功能。以下是多核苷

酸的非限制性实 例:基因或基因片段的编码区或非编码区、根据连接分析定义的多个座位(一个座位)、外显子、内含子、信使RNA(mRNA)、转运RNA、核糖体 RNA、短干扰RNA(siRNA)、短发夹RNA(shRNA)、micro-RNA (miRNA)、核酶、cDNA、重组多核苷酸、分支多核苷酸、质粒、载体、任 何序列的分离的DNA、任何序列的分离的RNA、核酸探针、和引物。多核苷 酸可以包含一个或多个经修饰的核苷酸,如甲基化的核苷酸和核苷酸类似 物。如果存在,可以在聚合物组装之前或之后进行核苷酸结构的修饰。核苷 酸的序列可以被非核苷酸组分中断。多核苷酸可以在聚合后,如通过与标记 的组分缀合来进一步修饰。

[0103] 在本发明的多个方面,术"嵌合RNA"、"嵌合指导RNA"、 "指导RNA"、"单个指导RNA"以及"合成指导RNA"是可互换地使用 的并且是指包括指导序列、tracr序列以及tracr配对序列的多核苷酸序列。术 语"指导序列"是指在指定靶位点的指导RNA内约20bp的序列,并且可与 术语"指导"或"间隔子"互换地使用。术语"tracr配对序列"也可与术语 "(一个或多个)同向重复"互换地使用。

[0104] 如本文所用的术语"野生型"是本领域技术人员所理解的术语,并 且表示生物、菌株、基因的典型形式或者当它在自然界存在时区别于突变体 或变体形式的特征。

[0105] 如本文所用的术语"变体"应当被理解为表示具有衍生自在自然界 中存在的模式的性质的展示。

[0106] 术语"非天然存在的"或"工程化的"可互换地使用并且表面人工 的参与。这些术语,当指核酸分子或多肽时,表示该核酸分子或多肽至少基 本上从它们在自然界中或如发现于自然界中的与其结合的至少另一种组分游 离出来。

[0107] "互补性"是指核酸与另一个核酸序列借助于传统的沃森-克里克 碱基配对或其他非传统类型形成一个或多个氢键的能力。互补百分比表示一 个核酸分子中可与一个第二核酸序列形成氢键(例如,沃森-克里克碱基配 对)的残基的百分比(例如,10个之中有5、6、7、8、9、10个即为50%、 60%、70%、80%、90%、和100%互补)。"完全互补"表示一个核酸序列 的所有连续残基与一个第二核酸序列中的相同数目的连续残基形成氢键。如 本文使用的"基本上互补"是指在一个具有8、9、10、11、12、13、14、 15、16、17、18、19、20、21、22、23、24、25、30、35、40、45、50个或 更多个核苷酸的区域上至少为60%、65%、70%、75%、80%、85%、90%、 95%、97%、98%、99%、或100%的互补程度,或者是指在严格条件下杂交 的两个核酸。

[0108] 如本文使用的对于杂交的"严格条件"是指与靶序列具有互补性的 一个核酸主要地与该靶序列杂交并且基本上不杂交到非靶序列上的条件。严 格条件通常是序列依赖性的,并且取决于许多因素而变化。一般而言,该序 列越长,则该序列特异性地杂交到其靶序列上的温度就越高。严格条件的非 限制性实例描述于蒂森(Tijssen)(1993)的《生物化学和分子生物学中的 实验室技术-核酸探针杂交》(Laboratory Techniques In Biochemistry And Molecular Biology-Hybridization With Nucleic Acid Probes),第I部分,第二 章,"杂交原理概述和核酸探针分析策略"("Overview of principles of hybridization and the strategy of nucleic acid probe assay"),爱思唯尔(Elsevier),纽约。

[0109] "杂交"是指其中一个或多个多核苷酸反应形成一种复合物的反 应,该复合物经由这些核苷酸残基之间的碱基的氢键键合而稳定化。氢键键 合可以借助于沃森-克里克碱

基配对、Hoogstein结合或以任何其他序列特异性 方式而发生。该复合物可包含形成一个双链体的两条链、形成多链复合物的 三条或多条链、单个自我杂交链、或这些的任何组合。杂交反应可以构成一 个更广泛的过程(如PCR的开始、或经由一种酶的多核苷酸的切割)中的一 个步骤。能够与一个给定序列杂交的序列被称为该给定序列的“互补物”。

[0110] 如本文使用的“表达”是指藉此从DNA模板转录成多核苷酸(如 转录成mRNA或其他RNA转录物)的过程和/或转录的mRNA随后藉此翻译 成肽、多肽或蛋白质的过程。转录物和编码的多肽可以总称为“基因产 物”。如果多核苷酸来源于基因组DNA,表达可以包括真核细胞中mRNA的 剪接。

[0111] 术语“多肽”、“肽”和“蛋白质”在本文可互换地使用,是指具 有任何长度的氨基酸的聚合物。该聚合物可以是可以是直链或支链的,它可 以包含修饰的氨基酸,并且它可以被非氨基酸中断。这些术语还涵盖已经被 修饰的氨基酸聚合物;这些修饰例如二硫键形成、糖基化、脂化 (lipidation)、乙酰化、磷酸化、或任何其他操纵,如与标记组分的结合。如本文使用的术语“氨基酸”包括天然的和/或非天然的或者合成的氨基酸, 包括甘氨酸以及D和L旋光异构体、以及氨基酸类似物和肽模拟物。

[0112] 术语“受试者”、“个体”和“患者”在本文中是可互换地使用 的,指的是脊椎动物,优选地是哺乳动物,更加优选地是人类。哺乳动物包 括但不限于鼠类、猴、人、农畜、体育用动物和宠物。也包括体内获得或体 外培养的一种生物实体的组织、细胞及其子代。

[0113] 术语“治疗剂(therapeutic agent)”、“可用于治疗的试剂 (therapeutic capable agent)”或“处理剂(treatment agent)”是可互换地使 用的,并且是指在给予受试者时赋予某种有益影响的一种分子或化合物。该 有益影响包括诊断确定的实现;改进一种疾病、症状、障碍、或病理学病 况;减少或预防一种疾病、症状、障碍或病况的发作;并且总体上对抗一种 疾病、症状、障碍或病理学病况。

[0114] 如此处使用的,“治疗(treatment)”或“进行治疗(treating)” 或“减轻”或“改善”是可互换地使用的。这些术语是指如下途径,该途径 用于获得有益或希望的结果,包括但不限于一种治疗益处和/或一种预防益 处。治疗益处意指治疗中的一种或多种疾病、病况、或症状上的任何治疗上 相关的改进或对其的影响。对于预防益处,该组合物可给予至处于发展一种 具体的疾病、病况、或症状的风险的受试者,或给予至报告了一种疾病的一 个或多个生理学症状的受试者,尽管该疾病、病况、或症状可能还没有体现 出来。

[0115] 术语“有效量”或“治疗有效量”是指一种药剂的足以实现有益或 希望的结果的量。治疗有效量可依赖于正治疗的受试者和疾病病状、受试者 的重量和年龄、疾病病况的严重度、给药方式等中一项或多个而改变,并可 以由本领域普通技术人员容易地确定。该术语也适用通过此处描述的显像方 法中的任一项提供一种检测用图像的一个剂量。具体剂量可依赖于以下中一 个或多个而变化:所选择的具体药剂、所遵循的给药方案、是否与其他化合 物组合给予、给予时间、待显像的组织、以及携带它的物理递送系统。

[0116] 除非另有说明,本发明的实践采用免疫学、生物化学、化学、分子 生物学、微生物学、细胞生物学、基因组学和重组DNA的常规技术,这些在 本领域的技能之内。参见萨姆布鲁克(Sambrook)、弗里奇(Fritsch)和马 尼亚蒂斯(Maniatis),《分子克隆:实验室手册》(MOLECULAR CLONING:A LABORATORY MANUAL),第2次编辑(1989);《当代分 子生物学实验手册》(CURRENT PROTOCOLS IN MOLECULAR BIOLOGY)(F.M.奥苏贝尔(F.M.Ausubel)等人编

辑,(1987));《酶 学方法》(METHODS IN ENZYMOLOGY)系列(学术出版公司):《PCR 2:实用方法》(PCR 2:A PRACTICAL APPROACH)(M.J.麦克弗森(M.J. MacPherson)、B.D.黑姆斯(B.D.Hames)和G.R.泰勒(G.R.Taylor)编辑 (1995))、哈洛(Harlow)和拉内(Lane)编辑(1988)《抗体:实验室 手册》(ANTIBODIES,A LABORATORY MANUAL),以及《动物细胞培 养》(ANIMAL CELL CULTURE)(R.I.弗雷谢尼(R.I.Freshney)编辑 (1987))。

[0117] 本发明的若干方面涉及包括一个或多个载体的载体系统,或载体本 身。载体可以被设计为用于在原核或真核细胞中表达CRISPR转录物(例如 核酸转录物、蛋白质、或酶)。例如,CRISPR转录物可表达于例如大肠杆菌 的细菌细胞、昆虫细胞(使用杆状病毒表达载体)、酵母细胞、或哺乳动物 细胞中。适合的宿主细胞进一步讨论于Goeddel(戈德尔),《基因表达技 术:酶学方法》(GENE EXPRESSION TECHNOLOGY:METHODS IN ENZYMOLOGY)185,学术出版社(Academic Press),圣地亚哥(San Diego),加利福尼亚州(Calif.)(1990年)中。可替代地,重组表达载体 可在体外例如使用T7启动子调节序列和T7聚合酶来转录和翻译。

[0118] 载体可以被引入原核生物中并且在其中增殖。在一些实施例中,原 核生物用来扩增有待引入真核细胞中的载体的多个拷贝,或者作为有待引入 真核细胞中的载体的产生中的中间载体(例如,扩增质粒作为病毒载体包装 系统的一部分)。在一些实施例中,原核生物用来扩增一个载体的多个拷贝 并且表达一种或多种核酸,如提供用于递送到宿主细胞或宿主生物中的一种 或多种蛋白质的来源。蛋白质在原核生物中的表达最经常在大肠杆菌中用含 有指导融合或非融合蛋白的表达的组成型或诱导型启动子的载体来进行。融合载体将多个氨基酸添加到在其中编码的蛋白质上,如该重组蛋白的氨基端 上。这样的融合载体可以用于一个或多个目的,如:(i)增加重组蛋白的表 达;(ii)增加重组蛋白的溶解性;以及(iii)通过在亲和纯化中充当配体来辅助 重组蛋白的纯化。通常,在融合表达载体中,将蛋白切割位点引入至融合部 分与重组蛋白的接合处以使得能够在纯化融合蛋白之后将重组蛋白与融合部 分分离。这类酶以及它们的同源识别序列包括因子Xa、凝血酶以及肠激酶。示例性融合表达载体包括pGEX(发玛西亚生物技术有限公司(Pharmacia Biotech Inc);史密斯和约翰逊,1988.《基因》(Gene)67:31-40)、pMAL (纽英伦生物技术公司(New England Biolabs),贝弗利(Beverly),马萨 诸塞州(Mass.))以及pRIT5(发玛西亚公司,皮斯卡塔韦(Piscataway), 新泽西州(N.J.)),它们分别将谷胱甘肽S-转移酶(GST)、麦芽糖E结合 蛋白或蛋白A融合至靶重组蛋白。

[0119] 适合的诱导型非融合大肠杆菌表达载体的实例包括pTrc(Amrann (阿姆兰)等人,(1988年)《基因》(Gene)69:301-315)以及pET 11d (Studier(斯图迪尔)等人,《基因表达技术:酶学方法》(GENE EXPRESSION TECHNOLOGY:METHODS IN ENZYMOLOGY)185,学术出 版社(Academic Press),圣地亚哥(San Diego),加利福尼亚州(Calif.) (1990年)60-89。

[0120] 在一些实施例中,载体是酵母表达载体。用于在酵母酿酒中表达的 载体的实例包括pYepSec1(巴尔戴利(Baldari)等人,1987.《欧洲分子生物 学学会杂志》(EMBO J)6:229-234)、pMFa(库尔坚(Kurjan)和赫斯库 伍特兹(Herskowitz),1982.《细胞》(Cell)30:933-943)、pJRY88(舒尔 茨(Schultz)等人,1987.《基因》(Gene)54:113-123)、pYES2(Invitrogen Corporation(英杰公司),San Diego(圣地亚哥),Calif(加利 福尼亚州))、以及picZ(Invitrogen Corp(英杰公司),San Diego(圣地亚 哥),Calif(加利福尼亚州))。

[0121] 在一些实施例中,使用杆状病毒载体,载体驱动昆虫细胞中的蛋白 质表达。可用

于在培养的昆虫细胞(例如,SF9细胞)中表达蛋白质的杆状 病毒载体包括pAc系列(史密斯(Smith)等人,1983.《分子细胞生物学》 (Mol.Cell.Biol.)3:2156-2165)和pVL系列(拉克瑙(Lucklow)和萨莫斯 (Summers),1989.《病毒学》(Virology)170:31-39)。

[0122] 在一些实施例中,使用哺乳动物表达载体,载体能够驱动一种或多 种序列在哺乳动物细胞中表达。哺乳动物表达载体的实例包括pCDM8(锡德 (Seed),1987.《自然》(Nature)329:840)和pMT2PC(考夫曼 (Kaufman)等人,1987.《欧洲分子生物学学会杂志》(EMBO J.)6:187- 195)。当用于哺乳动物细胞中时,表达载体的控制功能典型地由一个或多个 调节元件提供。例如,常用的启动子来源于多瘤、腺病毒2、巨细胞病毒、猴 病毒40、以及本文所述和本领域已知的其他病毒。对于用于原核细胞和真核 细胞两者的其他适合的表达系统参见例如萨姆布鲁克(Sambrook)等人的 《分子克隆实验指南》(Molecular Cloning:A Laboratory Manual.)(第2 版)的第16和第17章,冷泉港实验室(Cold Spring Harbor Laboratory),冷 泉港实验室出版社(Cold Spring Harbor Laboratory Press),冷泉港(Cold Spring Harbor),纽约,1989。

[0123] 在一些实施例中,重组哺乳动物表达载体能够指导核酸优先在特定 细胞类型中表达(例如,使用组织特异型调节元件来表达核酸)。组织特异 型调节元件是本领域中已知的。适合的组织特异型启动子的非限制性实例包 括白蛋白启动子(肝脏特异性的;平克特(Pinkert)等人,1987.《基因发 育》(Genes Dev.)1:268-277),淋巴特异性启动子(卡里曼(Calame)和 伊顿(Eaton),1988.《免疫学进展》(Adv.Immunol)43:235-275),特别 是T细胞受体的启动子(维纳图(Winoto)和巴尔迪莫(Baltimore),1989.《欧洲分子生物学学会杂志》(EMBO J.)8:729-733)和免疫球蛋白(巴纳 吉(Baneiji)等人,1983.《细胞》(Cell)33:729-740;奎恩(Queen)和巴 尔迪莫(Baltimore),1983.《细胞》(Cell)33:741-748),神经元特异性启 动子(例如,神经丝启动子;伯恩(Byrne)和鲁德尔(Ruddle),1989.《美 国科学院院刊》(Proc.Natl.Acad.Sci USA)86:5473-5477),胰腺特异性启 动子(艾德兰德(Edlund)等人,1985.《科学》(Science)230:912-916), 以及乳腺特异性启动子(例如,乳清启动子;美国专利号4,873,316和欧洲申 请公开号264,166)。还涵盖发育型调节启动子,例如,鼠科动物同源框蛋白 (hox)启动子凯赛尔((Kessel)和格鲁斯(Gruss),1990.《科学》(Science)249:374-379)和α-甲胎蛋白启动子(卡姆皮斯(Campes)和蒂尔 曼(Tilghman),1989.《基因发育》(Genes Dev.)3:537-546)。

[0124] 在一些实施例中,调节元件可操作地连接至CRISPR系统的一个或 多个元件,从而驱动该CRISPR系统的该一个或多个元件的表达。一般而 言,CRISPR(规律间隔成簇短回文重复序列),也称为SPIDR(SPacer间隔 开的同向重复),构成通常对于特定细菌物种而言特异性的DNA基因座的家 族。该CRISPR座位包含在大肠杆菌中被识别的间隔开的短序列重复(SSR) 的一个不同类(石野(Ishino)等人,《细菌学杂志》(J.Bacteriol.),169: 5429-5433[1987];和中田(Nakata)等人,《细菌学杂志》(J. Bacteriol.),171:3553-3556[1989])、以及相关基因。类似的间隔开的SSR 已经鉴定于地中海富盐菌(Haloferax mediterranei)、化脓链球菌、鱼腥藻 属、和结核分枝杆菌中(参见,格鲁恩(Groenen)等人,《分子微生物学》(Mol.Microbiol.),10:1057-1065[1993];霍(Hoe)等人,《新发感染性 疾病》(Emerg.Infect.Dis.),5:254-263[1999];马斯波尔(Masepohl)等 人,《生物化学与生物物理学学报》(Biochim.Biophys.Acta)1307:26-30 [1996];以及莫吉卡(Mojica)等人,《分子

微生物学》,17:85-93 [1995])。这些CRISPR座位典型地不同于其他SSR的重复结构,这些重复已 被称为规律间隔的短重复(SRSR)(詹森(Janssen)等人,《OMICS:整合 生物学杂志》(OMICS J.Integ.Biol.),6:23-33[2002];以及莫吉卡 (Mojica)等人,《分子微生物学》(Mol.Microbiol.),36:244-246 [2000])。一般而言,这些重复是以簇存在的短元件,其被具有基本上恒定长 度的独特间插序列规律地间隔开(莫吉卡(Mojica)等人,[2000],同上)。虽然重复序列在菌株之间是高度保守的,许多间隔开的重复和这些间隔区的 序列一般在菌株与菌株之间不同(冯·埃姆登(van Embden)等人,《细菌学 杂志》(J.Bacteriol.),182:2393-2401[2000])。已经在40种以上的原核生 物中鉴定出CRISPR座位(参见,例如,詹森(Janssen)等人,《分子微生 物学》(Mol.Microbiol.),43:1565-1575[2002];以及莫吉卡(Mojica)等 人,[2005]),包括但不限于:气火菌属(Aeropyrum)、热棒菌属 (Pyrobaculum)、硫化叶菌属(Sulfolobus)、古球菌属(Archaeoglobus)、 盐盒菌属(Halocarcula)、甲烷杆菌属(Methanobacterium)、甲烷球菌属 (Methanococcus)、甲烷八叠球菌属(Methanosarcina)、甲烷火菌属 (Methanopyrus)、焦球菌属(Pyrococcus)、嗜酸菌属(Picrophilus)、热 原体属(Thermoplasma)、棒杆菌属(Corynebacterium)、分枝杆菌属 (Mycobacterium)、链霉菌属(Streptomyces)、产水菌属(Aquifex)、紫 单胞菌属(Porphyromonas)、绿菌属(Chlorobium)、栖热菌属 (Thermus)、芽孢杆菌属(Bacillus)、利斯特菌属(Listeria)、葡萄球菌 属(Staphylococcus)、梭菌属(Clostridium)、好热厌氧杆菌属 (Thermoanaerobacter)、支原体属(Mycoplasma)、梭杆菌属 (Fusobacterium)、固氮弓菌属(Azarcus)、色杆菌属 (Chromobacterium)、奈瑟球菌属(Neisseria)、亚硝化单胞菌属 (Nitrosomonas)、脱硫弧菌属(Desulfovibrio)、地杆菌属(Geobacter)、 粘球菌属(Myxococcus)、弯曲杆菌属(Campylobacter)、类杆菌属 (Wolinella)、不动杆菌属(Acinetobacter)、欧文菌属(Erwinia)、埃希菌 属(Escherichia)、军团菌属(Legionella)、甲基球菌属 (Methylococcus)、巴斯德菌属(Pasteurella)、发光细菌属(Photobacterium)、沙门菌属(Salmonella)、黄单胞菌属 (Xanthomonas)、耶尔森菌属(Yersinia)、密螺旋体属(Treponema)、和 栖热袍菌属(Thermotoga)。

[0125] 一般而言,“CRISPR系统”统称为转录物和涉及CRISPR相关 (“Cas”)基因的表达或指导其活性的其他元件,包括编码Cas基因的序 列、tracr(反式激活CRISPR)序列(例如tracrRNA或活性部分 tracrRNA)、tracr配对序列(涵盖“同向重复”和在内源CRISPR系统背景 下的tracrRNA加工的部分同向重复)、指导序列(在内源CRISPR系统背景 下也称为“spacer”)、或来自CRISPR座位的其他序列和转录物。在一些实 施例中,CRISPR系统的一个或多个元件来源于I型、II型、或III型CRISPR 系统。在一些实施例中,CRISPR系统的一个或多个元件来源于包含内源 CRISPR系统的特殊生物,如化脓链球菌。一般而言,CRISPR系统的特征为 促进在靶序列的位点处的CRISPR复合物(在内源CRISPR系统的背景下也称 为原型间隔子)的形成的元件。在CRISPR复合物形成的背景下,“靶序 列”是指一个指导序列被设计为对其具有互补性的序列,其中在靶序列与指 导序列之间的杂交促进CRISPR复合物的形成。完全互补性不是必需的,条 件是存在足够互补性以引起杂交并且促进一种CRISPR复合物的形成。一个 靶序列可以包含任何多核苷酸,如DNA或RNA多核苷酸。在一些实施例中,靶序列位于细胞的细胞核或细胞质中。在一些实施例中,该靶序列可位 于真核细胞的一个细胞器例如线粒体或叶绿体内。可被用于重组到包括该靶 序列的靶基因座中的序列

或模板被称为“编辑模板”或“编辑多核苷酸”或 “编辑序列”。在本发明的方面,外源的模板多核苷酸可被称为编辑模板。 在本发明的一个方面,该重组是同源重组。

[0126] 典型地,在内源CRISPR系统的背景下,CRISPR复合物(包含杂 交到靶序列上并且与一种或多种Cas蛋白复合的指导序列)的形成导致在该 靶序列中或其附近(例如在1、2、3、4、5、6、7、8、9、10、20、50、或更 多个碱基对之内)的一条链或两条链的切割。不希望受到理论的束缚,该 tracr序列(其可以包含或其组成为野生型tracr序列的全部或部分(例如野生 型tracr序列的约或多于约20、26、32、45、48、54、63、67、85个、或更多 个核苷酸))也可以形成CRISPR复合物的一部分,如通过沿着该tracr序列 的至少一部分杂交到与该指导序列可操作地连接的tracr配对序列的全部或部 分上。在一些实施例中,该tracr序列与一个tracr配对序列具有足够的互补性 以进行杂交,并参与一种CRISPR复合物的形成。至于该靶序列,据信不需 要完全互补性,只要足以具备功能性。在一些实施例中,当进行最佳比对 时,沿着该tracr配对序列的长度,该tracr序列具有至少50%、60%、70%、 80%、90%、95%、或99%序列互补性。在一些实施例中,将驱动CRISPR系 统的一个或多个元件的表达的一个或多个载体引入到宿主细胞中,使得该 CRISPR系统的这些元件的表达在一个或多个靶位点指导CRISPR复合物的形 成。例如,Cas酶、连接到tracr配对序列上的指导序列、以及tracr序列可以 各自可操作地连接到在分开的载体上的分开的调节元件上。可替代地,从相 同或不同调节元件表达的二个或更多个元件可以结合在单个载体中,其中提 供该CRISPR系统的任何组分的一个或多个另外的载体不包括在第一载体 中。结合在单个载体中的CRISPR系统元件可以按照任何适合的取向排列, 如一个元件相对于第二元件位于5’(“上游”)或3’(“下游”)。一个元 件的编码序列可以位于第二元件的编码序列的同一条链或相对链上,并且取 向为相同或相对的方向。在一些实施例中,单个启动子驱动编码CRISPR酶 的转录物以及该指导序列、tracr配对序列(任选地可操作地连接到该指导序 列上)、和嵌入在一个或多个内含子序列之内(例如,各自在不同的内含子 中,两个或更多个在至少一个内含子中,或者全部都在单个内含子中)的 tracr序列中的一者或多者的表达。在一些实施例中,该CRISPR酶、指导序 列、tracr配对序列、和tracr序列可操作地连接到相同的启动子上并且从其表 达。

[0127] 在一些实施例中,一个载体包含一个或多个插入位点,如限制性核 酸内切酶识别序列(也称为“克隆位点”)。在一些实施例中,一个或多个 插入位点(例如,约或多于约1、2、3、4、5、6、7、8、9、10个、或更多 个插入位点)位于一个或多个载体的一个或多个序列元件的上游和/或下游。 在一些实施例中,一个载体包含在tracr配对序列的上游、并且任选地在可操 作地连接到该tracr配对序列上的调节元件的下游的插入位点,使得在将指导 序列插入到该插入位点中之后并且在表达时该指导序列指导CRISPR复合物 与真核细胞中的靶序列的序列特异性结合。在一些实施例中,一个载体包含 一个或多个插入位点,每个插入位点位于两个tracr配对序列之间,从而允许 在每个位点插入指导序列。在这样一种安排中,两个或更多个指导序列可以 包含两个或更多个拷贝的单指导序列、两个或更多个不同的指导序列、或这 些的组合。当使用多个不同的指导序列时,可以使用单个表达构建体使CRISPR活性靶向到细胞内的多个不同的相应靶序列。例如,单个载体可以包 含约或多于约1、2、3、4、5、6、7、8、9、10、15、20个、或更多个指导 序列。在一些实施例中,可以提供约或多于约1、2、3、4、5、6、7、8、9、 10个、或更多个这样的含有靶序列的载体,并且任选地将其递送

到细胞中。

[0128] 在一些实施例中,一个载体包含可操作地连接到编码CRISPR酶 (如Cas蛋白)的酶编码序列上的调节元件。Cas蛋白的非限制性实例包括:Cas1、Cas1B、Cas2、Cas3、Cas4、Cas5、Cas6、Cas7、Cas8,Cas9(也称为 Csn1和Csx12)、Cas10、Csy1、Csy2、Csy3、Cse1、Cse2、Csc1、Csc2、 Csa5、Csn2、Csm2、Csm3、Csm4、Csm5、Csm6、Cmr1、Cmr3、Cmr4、 Cmr5、Cmr6、Csb1、Csb2、Csb3、Csx17、Csx14、Csx10、Csx16、CsaX、 Csx3、Csx1、Csx15、Csf1、Csf2、Csf3、Csf4、其同系物、或其修饰形式。这些酶是已知的;例如,化脓链球菌Cas9蛋白的氨基酸序列可见于SwissProt 数据库登录号Q99ZW2下。在一些实施例中,未修饰的CRISPR酶如Cas9具 有DNA切割活性。在一些实施例中,该CRISPR酶是Cas9,并且可以是来自 化脓链球菌或肺炎链球菌的Cas9。在一些实施例中,该CRISPR酶指导在靶 序列位置处(例如在该靶序列之内和/或在该靶序列的互补物之内)的一条链 或两条链的切割。在一些实施例中,该CRISPR酶指导距离靶序列的第一个 或最后一个核苷酸约1、2、3、4、5、6、7、8、9、10、15、20、25、50、 100、200、500个、或更多个碱基对之内的一条链或两条链的切割。在一些实 施例中,一个载体编码相对于相应的野生型酶被突变的CRISPR酶,使得该 突变的CRISPR酶缺乏切割含有一个靶序列的靶多核苷酸的一条链和两条链 的能力。例如,在来自化脓链球菌的Cas9的RuvC I催化结构域中的天冬氨 酸到丙氨酸取代(D10A)将Cas9从切割两条链的核酸酶转化成切口酶(切 割单条链)。致使Cas9为切口酶的其他突变实例包括而不限于H840A、N854A、和N863A。在一些实施例中,一种Cas9切口酶可与一个或多个指导 序列组合使用,例如,两个指导序列,其分别靶向该DNA靶标的有义以及反 义链。此组合允许两条链被切口并且用以诱导NHEJ。申请人已证明(数据未 显示)两种切口酶靶标(即,在相同位置靶向DNA的不同链的sgRNA)在 诱导诱变的NHEJ方面的效力。一种单切口酶(具有单个sgRNA的Cas9- D10A)不能够诱导NHEJ并且建立indel(indel),但申请人已经显示,双切 口酶(Cas9-D10A以及在相同位置靶向不同链的两个sgRNA)在人胚胎干细 胞(hESC)中可以这样做。该效率在hESC中是约50%的核酸酶(即,不具 有D10突变的常规Cas9)。

[0129] 作为一个另外的实例,Cas9的两个或更多个催化结构域(RuvC I、RuvC II、和RuvC III)可以被突变为产生实质上缺乏所有DNA切割活性的突 变的Cas9。在一些实施例中,D10A突变与H840A、N854A、或N863A突变 的一者或多者相组合,以便产生实质上缺乏所有DNA切割活性的Cas9酶。 在一些实施例中,当该突变的CRISPR酶的DNA切割活性比它的非突变形式 低约25%、10%、5%、1%、0.1%、0.01%、或更少时,则该酶被考虑为实质 上缺乏所有DNA切割活性。其他突变可能是有用的;其中Cas9或其他 CRISPR酶是来自除化脓链球菌外的物种,可产生相应氨基酸的突变以实现类 似效果。

[0130] 在一些实施例中,编码CRISPR酶的酶编码序列经密码子优化,以 便在特定的细胞如真核细胞中表达。这些真核细胞可以是特定生物的那些或 来源于特定生物,如哺乳动物,包括但不限于人、小鼠、大鼠、兔、狗、或 非人类灵长动物。一般而言,密码子优化是指通过用在宿主细胞的基因中更 频繁地或者最频繁地使用的密码子代替天然序列的至少一个密码子(例如约 或多于约1、2、3、4、5、10、15、20、25、50个、或更多个密码子同时维 持该天然氨基酸序列而修饰一个核酸序列以便增强在感兴趣宿主细胞中的表 达的方法。不同的物种对于具有特定氨基酸的某些密码子展示出特定的偏 好。密码子偏好性(在生物之间的密码子使用的差异)经常与信使RNA (mRNA)的翻译效率相关,而该翻译效率则被认为依赖

于(除其他之外) 被翻译的密码子的性质和特定的转运RNA(tRNA)分子的可用性。细胞内选定的tRNA的优势一般反映了最频繁用于肽合成的密码子。因此,可以将基 因定制为基于密码子优化在给定生物中的最佳基因表达。密码子利用率表可 以容易地获得,例如在密码子使用数据库("Codon Usage Database")中, 并且这些表可以通过不同的方式调整适用。参见,中村Y.(Nakamura Y.)等 人,"从国际DNA序列数据库中制表的密码子使用:2000年的状态" (Codon usage tabulated from the international DNA sequence databases: status for the year 2000)《核酸研究》(Nucl.Acids Res.)28:292(2000年)。用于 密码子优化特定的数列以便在特定的宿主细胞中表达的计算机算法也是可得 的,如基因制造(Gene Forge)(Aptagen公司;雅各布斯(Jacobus), PA),也是可得的。在一些实施例中,在编码CRISPR酶的序列中的一个或 多个密码子(例如1、2、3、4、5、10、15、20、25、50个、或更多个、或 所有密码子)相应于对于特定氨基酸最频繁使用的密码子。

[0131] 在一些实施例中,一种载体编码包含一个或多个核定位序列 (NLS)如约或多于约1、2、3、4、5、6、7、8、9、10个、或更多个NLS 的CRISPR酶。在一些实施例中,该CRISPR酶包含在或接近于氨基端的约或 多于约1、2、3、4、5、6、7、8、9、10个、或更多个NLS,在或接近于羧 基端约或多于约1、2、3、4、5、6、7、8、9、10个、或更多个NLS,或这 些的组合(例如在氨基端的一个或多个NLS以及在羧基端的一个或多个 NLS)。当存在多于一个NLS时,每一个可以被选择为不依赖于其他NLS, 使得在多于一个拷贝中可以存在单个NLS和/或与在一个或多个拷贝中存在一 个或多个其他NLS相组合。在本发明的一个优选实施例中,该CRISPR酶包 含至多6个NLS。在一些实施例中,当NLS的最近的氨基酸是在从N端或C 端沿着该多肽链的约1、2、3、4、5、10、15、20、25、30、40、50个、或 更多个氨基酸之内时,NLS可以被视为接近该N端或C端。典型地,NLS由 暴露于蛋白表面上的带正电的赖氨酸或精氨酸的一个或多个短序列组成,但 其他类型的NLS是已知的。NLS的非限制性实例包括来源于以下项的NLS序 列:SV40病毒大T抗原的NLS,其具有氨基酸序列PKKKRKV;来自核质蛋 白的NLS(例如,具有序列KRPAATKKAGQAKKKK的核质蛋白二分 NLS);c-myc NLS,其具有氨基酸序列PAAKRVKLD或RQRRNELKRSP; hRNPA1 M9 NLS,其具有序列 NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY;来自输入蛋白-α 的IBB结构域的序列 RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV;肌瘤T蛋 白的序列VSRKRPRP和PPKKARED;人p53的序列POPKKKPL;小鼠c-abl IV的序列SALIKKKKKMAP;流感病毒NS1的序列DRLRR和PKQKKRK; 肝炎病毒δ抗原的序列RKLKKKIKKL;小鼠Mx1蛋白的序列 REKKKFLKRR;人聚(ADP-核糖)聚合酶的序列 KRKGDEVDGVDEVAKKKSKK;以及类固醇激素受体(人)糖皮质激素的 序列RKCLQAGMNLEARKTKK。

[0132] 一般而言,该一个或多个NLS具有足够的强度,以便在真核细胞 的核中驱动该CRISPR复合物以可检测到的量积聚。一般而言,核定位活性 的强度可以驱动在该CRISPR酶中的NLS的数目、所使用的一个或多个特定 的NLS、或这些因素的组合。可以通过任何适合的技术进行细胞核中积聚的 检测。例如,一种检测标记可以融合到CRISPR酶上,使得细胞内的位置可 以被可视化,如与检测细胞核的位置的手段(例如,对于细胞核特异的染 料,如DAPI)相结合。可检测标记物的实例包括荧光蛋白(例如绿色荧光蛋 白,或GFP;RFP;CFP),以及表位标签(HA标签,flag标签,SNAP标 签)。还可以将细胞核从细胞中分离出来,然后可以通过任何适合的用于检 测蛋白质的方法分析其内容物,如免疫组织化学、western印迹或酶活性测 定。如通过测定CRISPR复合物形成的作用(例如,测定在靶序列处

的DNA 切割或突变、或测定由于CRISPR复合物形成和/或CRISPR酶活性的影响而 改变的基因表达活性),与没有暴露于CRISPR酶或复合物、或暴露于缺乏 该一个或多个NLS的CRISPR酶的对照相比较,还可以间接地确定细胞核中 的积聚。

[0133] 一般而言,指导序列是与靶多核苷酸序列具有足够互补性以便与该 靶序列杂交并且指导CRISPR复合物与该靶序列的序列特异性结合的任何多 核苷酸序列。在一些实施例中,当使用适合的比对算法进行最佳比对是,在 指导序列与其相应的靶序列之间的互补程度是约或多于约50%、60%、75%、 80%、85%、90%、95%、97.5%、99%、或更多。可以使用用于比对序列的任 何适合的算法来确定最佳比对,其非限制性实例包括史密斯-沃特曼(Smith- Waterman)算法、尼德曼-翁施(Needleman-Wunsch)算法、基于伯罗斯-惠 勒变换(Burrows-Wheeler Transform)的算法(例如伯罗斯-惠勒比对工具 (Burrows Wheeler Aligner))、ClustalW、Clustal X、BLAT、Novoalign (Novocraft技术公司)、ELAND(亿明达公司(Illumina),圣地亚哥,加 利福尼亚州)、SOAP(在soap.genomics.org.cn可获得)、以及Maq(在 maq.sourceforge.net可获得)。在一些实施例中,一个指导序列在长度上为约或多于约5、10、11、12、13、14、15、16、17、18、19、20、21、22、23、 24、25、26、27、28、29、30、35、40、45、50、75个、或更多个核苷酸。 在一些实施例中,一个指导序列在长度上为少于约75、50、45、40、35、 30、25、20、15、12个、或更少的核苷酸。可以通过任何适合的测定法来评 估指导序列指导CRISPR复合物与靶序列的序列特异性结合的能力。例如, 足以形成CRISPR复合物的CRISPR系统的组分,包括有待测试的指导序列在 内,可以例如通过用编码该CRISPR序列的组分的载体进行转染而被提供到 具有相应靶序列的宿主细胞中,随后通过如本文所述的测量员测定来评估在 该靶序列之内的优先切割。类似地,通过提供该靶序列、包括有待测试的指 导序列在内的CRISPR复合物的组分、和不同于该测试指导序列的对照指导序列,并且比较在该测试指导序列与该对照指导序列反应之间的靶序列处的 结合或切割率,可以在一个试管中评估靶多核苷酸序列的切割。其他测定法 是可能的,并且将由本领域技术人员想到。

[0134] 一个指导序列可以被选择为靶向任何靶序列。在一些实施例中,该 靶序列是在细胞的基因组内的序列。示例性靶序列包括在靶基因组中为独特 的那些。例如,对于化脓链球菌Cas9,在基因组中的独特靶序列可以包括形 式MMMMMMMMNNNNNNNNNNNNXGG的Cas9靶位点,其中 NNNNNNNNNNNNXGG(N是A、G、T、或C;并且X可以是任何物)在 该基因组中单次出现。在基因组中的独特靶序列可以包括形式 MMMMMMMMMNNNNNNNNNNNNXGG的化脓链球菌Cas9靶位点,其中 NNNNNNNNNNNXGG(N是A、G、T、或C;并且X可以是任何物)在该 基因组中单次出现。对于嗜热链球菌CRISPR1 Cas9,在基因组中的独特靶序 列可以包括形式MMMMMMMMNNNNNNNNNNNNXXAGAAW的Cas9靶 位点,其中NNNNNNNNNNNNXXAGAAW(N是A、G、T、或C;X可以 是任何物;并且W是A或T)在该基因组中单次出现。在基因组中的独特靶 序列可以包括形式MMMMMMMMMNNNNNNNNNNNNXXAGAAW的嗜热链 球菌CRISPR1 Cas9靶位点,其中NNNNNNNNNNNXXAGAAW(N是A、 G、T、或C;X可以是任何物;并且W是A或T)在该基因组中单次出现。 对于化脓链球菌Cas9,在基因组中的独特靶序列可以包括形式MMMMMMMMNNNNNNNNNNNNXGGXG的Cas9靶位点,其中 NNNNNNNNNNNNXGGXG(N是A、G、T、或C;并且X可以是任何物) 在该基因组中单次出现。在基因组中的独特靶序列可以包括形式MMMMMMMMMNNNNNNNNNNNNXGGXG的化脓链球菌Cas9靶位点,其 中NNNNNNNNNNNXGGXG(N是A、G、

T、或C;并且X可以是任何 物)在该基因组中单次出现。在这些序列的每一个中,“M”可以是A、G、 T、或C,并且在序列鉴定中不必考虑为是独特的。

[0135] 在一些实施例中,选择指导序列以减少在该指导序列内的二级结构 水平。可以通过任何适合的多核苷酸折叠算法来确定二级结构。一些算法是 基于计算最小吉布斯(Gibbs)自由能。一种这样的算法的实例是mFold,正 如祖克(Zuker)和施蒂格勒(Stiegler)《核酸研究》(Nucleic Acids Res.)9 (1981),133-148)所描述。折叠算法的另一个实例是使用由维也纳大学(University of Vienna)的理论化学研究所(Institute for Theoretical Chemistry)研发的在线网络服务器RNAfold(参见例如A.R.格鲁伯(A.R. Gruber)等人,2008,《细胞》(Cell)106(1):23-24;以及PA卡尔(PA Carr)和GM丘奇(GM Church),2009,《自然生物技术》(Nature Biotechnology)27(12):1151-62)。另外的算法可以发现于通过引用结合在此 的美国申请序列号TBA(代理人案号44790.11.2022;博德参考号BI- 2013/004A)中。

[0136] 一般而言,tracr配对序列包括与tracr序列具有足够互补性以促进 下列一项或多项的任何序列:(1)在含有相应tracr序列的细胞中的侧翼为 tracr配对序列的指导序列的切除;以及(2)在靶序列处的CRISPR复合物的 形成,其中该CRISPR复合物包含杂交到tracr序列上的tracr配对序列。通 常,互补程度是就tracr配对序列与tracr序列沿着这两个序列的较短者的长度 的最佳比对而言。可以通过任何适合的比对算法来确定最佳比对,并且可以 可以进一步对二级结构做出解释,如在该tracr序列或tracr配对序列之内的自我互补性。在一些实施例中,在进行最佳比对时,在该tracr序列与tracr配对 序列之间沿着这两者的较短者的长度的互补程度是约或多于约25%、30%、 40%、50%、60%、70%、80%、90%、95%、97.5%、99%、或更高。图12B 以及13B中提供了一个tracr序列以及一个tracr配对序列之间的最佳比对的实 例说明。在一些实施例中,该tracr序列在长度上为约或多于约5、6、7、8、 9、10、11、12、13、14、15、16、17、18、19、20、25、30、40、50个、或 更多个核苷酸。在一些实施例中,该tracr序列和tracr配对序列被包含在单个 转录物中,使得在这两者之间的杂交产生具有二级结构(如发夹)的转录 物。用于在发夹结构中使用的优选环形成序列在长度上为四个核苷酸,并且 最优选地具有序列GAAA。然而,可以使用更长或更短的环序列,正如可替 代的序列。这些序列优选地包括三联体(例如,AAA)、和另外的核苷酸 (例如C或G)。环形成序列的实例包括CAAA和AAAG。在本发明的一个 实施例中,该转录物或转录的多核苷酸序列具有至少两个或更多个发夹。在 优选的实施例中,该转录物具有两个、三个、四个或五个发夹。在本发明的 一个另外的实施例中,该转录物具有至多五个发夹。在一些实施例中,该单 个转录物进一步包括一种转录终止序列;优选地这是一个polyT序列,例如 六个T的核苷酸。图13B中的下面部分中提供了这样的发夹结构的实例说 明,其中最终“N”的序列5’的部分以及环的上游对应于该tracr配对序列, 并且该环的序列3’的部分对应于该tracr序列。另外的包含指导序列、tracr配 对序列、和tracr序列的单个多核苷酸的非限制性实例如下(列出为5’到 3’),其中“N”代表指导序列的碱基,小写字体的第一区代表tracr配对序 列,且小写字体的第二区代表tracr序列,并且该最后的poly-T序列代表转录 终止子:(1) NNNNNNNNNNNNNNNNNNNgtttttgtactctcaagatttaGAAAtaaatcttgcagaagctaca aagataaggcttcatgccgaaatcaacaccctgtcatttatggcagggtgttttcgttatttaaTTTTTT;(2) NNNNNNNNNNNNNNNNNNNgtttttgtactctcaGAAAtgcagaagctacaaagataaggctt

catgccgaaatcaacaccctgtcattttatggcagggtgttttcgttatttaaTTTTTT;(3) NNNNNNNNNNNNNNNNNNNgttttgtactctcaGAAAtgcagaagctacaaagataaggctt catgccgaaatcaacaccctgtcattttatggcagggtgtTTTTTT;(4) NNNNNNNNNNNNNNNNNNNgttttagagctaGAAAtagcaagttaaaataaggctagtccgtt atcaacttgaaaaagtggcaccgagtcggtgcTTTTTT;(5) NNNNNNNNNNNNNNNNNNNgttttagagctaGAAATAGcaagttaaaataaggctagtccg ttatcaacttgaaaaagtgTTTTTTT;以及(6) NNNNNNNNNNNNNNNNNNNgttttagagctagAAATAGcaagttaaaataaggctagtccgttatcaTTTTTTTT。在一些实施例中,序列(1)到(3)与来自嗜热链球菌 CRISPR1的Cas9结合使用。在一些实施例中,序列(4)到(6)与来自化脓 链球菌CRISPR1的Cas9结合使用。在一些实施例中,该tracr序列是一个与 包含该tracr配对序列的转录物分开的转录物(例如在图13B的顶部部分中所 示的)。

[0137] 在一些实施例中,还提供了重组模板。重组模板可以是如本文所述 的另一个载体的组分,其被包含在一个分开的载体中,或者被提供为一个分 开的多核苷酸。在一些实施例中,重组模板被设计为用作在同源重组中的模 板,如在被作为CRISPR复合物的一部分的CRISPR酶切开或切割的靶序列之 内或在其附近。模板多核苷酸可以具有任何适合的长度,如约或多于约10、 15、20、25、50、75、100、150、200、500、1000个、或更多个核苷酸长 度。在一些实施例中,该模板多核苷酸与包含该靶序列的多核苷酸的一部分 互补。当进行最佳比对时,模板多核苷酸能够与靶序列的一个或多个核苷酸 (例如约或多于约1、5、10、15、20、25、30、35、40、45、50、60、70、 80、90、100个、或更多个核苷酸)重叠。在一些实施例中,当一个模板序列 与包含靶序列的多核苷酸进行最佳比对时,该模板多核苷酸的最近的核苷酸在距离该靶序列的约1、5、10、15、20、25、50、75、100、200、300、 400、500、1000、5000、10000个、或更多个核苷酸之内。

[0138] 在一些实施例中,该CRISPR酶是包含一个或多个异源蛋白结构域 (例如除了该CRISPR酶之外的约或多于约1、2、3、4、5、6、7、8、9、10 个、或更多个结构域)的融合蛋白的一部分。CRISPR酶融合蛋白可以包含任 何其他蛋白质,以及任选地在任何两个结构域之间的连接序列。可以融合到 CRISPR酶上的蛋白质结构域的实例包括但不限于,表位标签、报告基因序 列、以及具有下列活性的一者或多者的蛋白质结构域:甲基酶活性、脱甲基 酶活性、转录激活活性、转录阻遏活性、转录释放因子活性、组蛋白修饰活 性、RNA切割活性和核酸结合活性。表位标签的非限制性实例包括组氨酸 (His)标签、V5标签、FLAG标签、流感病毒血凝素(HA)标签、Myc标 签、VSV-G标签、和硫氧还蛋白(Trx)标签。报告基因的实例包括,但不限 于,谷胱甘肽-S-转移酶(GST)、辣根过氧化物酶(HRP)、氯霉素乙酰转 移酶(CAT)、β-半乳糖苷酶、β-葡糖醛酸糖苷酶、萤光素酶、绿色荧光蛋白 (GFP)、HcRed、DsRed、青荧光蛋白(CFP)、黄色荧光蛋白(YFP)、 以包括蓝色荧光蛋白(BFP)的自发荧光蛋白。CRISPR酶可以融合到编码一 种蛋白质或蛋白质片段的基因序列上,所述蛋白质或蛋白质片段结合DNA分 子或结合其他细胞分子,其包括,但不限于,麦芽糖结合蛋白(MBP)、S- tag、Lex A DNA结合结构域(DBD)融合物、GAL4 DNA结合结构域融合 物、以及单纯疱疹病毒(HSV)BP16蛋白融合物。可以形成包含CRISPR酶 的融合蛋白的一部分的另外的结构域描述于US 20110059502中,通过引用将 其并入本文。在一些实施例中,使用标记的CRISPR酶来鉴定靶序列的位 置。

[0139] 在一些方面,本发明提供了以下方法,这些方法包括向宿主细胞递 送一个或多个

多核苷酸,如或如在此描述的一个或多个载体、其一个或多个 转录本和/或一个或自其转录的蛋白。在一些方面,本发明进一步提供了通过 这样的方法产生的细胞以及包括这样的细胞或由这样的细胞产生的生物体 (例如动物、植物、或真菌)。在一些实施例中,将与(并且任选地与其复 合)指导序列组合的CRISPR酶递送至细胞。可以使用常规的病毒和非病毒基的基因转移方法将核酸引入哺乳动物细胞或靶组织中。可以使用这样的方 法向培养物中或宿主生物中的细胞给予编码CRISPR系统的组分的核酸。非 病毒载体递送系统包括DNA质粒、RNA(例如在此描述的载体的转录本)、裸核酸以及与递送赋形剂(如脂质体)复合的核酸。病毒载体递送系统包括 DNA和RNA病毒,在被递送至细胞后它们具有游离型或整合型基因组。关 于基因疗法程序的综述,参见安德森(Anderson),《科学》(Science)256: 808-813(1992);纳贝尔(Nabel)&费尔格纳(Felgner),TIBTECH 11: 211-217(1993);三谷(Mitani)&卡斯基(Caskey),TIBTECH 11:162- 166(1993);狄龙(Dillon),TIBTECH 11: 167-175(1993);米勒 (Miller),《自然》(Nature)357:455-460(1992);范·布朗特(Van Brunt),《生物技术》(Biotechnology)6(10):1149-1154(1988);维涅 (Vigne),《恢复神经学和神经科学》(Restorative Neurology and Neuroscience)8:35-36(1995);克雷默(Kremer)&佩里科德特 (Perricaudet),《英国医学公报》(British Medical Bulletin)51(1):31-44 (1995);哈塔塔(Haddada)等人,在《微生物学和免疫学当前主题》 (Current Topics in Microbiology and Immunology)中多尔夫勒(Doerfler)和 博姆(Böhm)(编辑)(1995);以及余(Yu)等人,《基因疗法》(Gene Therapy)1:13-26(1994)。

[0140] 核酸的非病毒递送方法包括脂转染、核转染、显微注射、基因枪、 病毒颗粒、脂质体、免疫脂质体、聚阳离子或脂质:核酸共轭物、裸DNA、人 工病毒体以及DNA的试剂增强性摄取。脂转染描述于例如美国专利号 5,049,386、4,946,787和4,897,355中并且脂转染试剂是市售的(例如, Transfectam™和Lipofectin™)。适于多核苷酸的有效的受体识别脂转染的阳 离子和中性脂质包括Felgner(费尔格纳),WO 91/17424;WO 91/16024的 那些。递送可以针对细胞(例如体外或离体给予)或靶组织(例如体内给 予)。

[0141] 脂质:核酸复合物(包括靶向的脂质体,如免疫脂质复合物)的制 备是本领域的技术人员熟知的(参见例如,克丽丝特尔(Crystal),《科 学》(Science)270:404-410(1995);布莱泽(Blaese)等人,《癌症基因 疗法》(Cancer Gene Ther.)2:291-297(1995);贝尔(Behr)等人,《生 物共轭化学》(Bioconjugate Chem.)5:382-389(1994);雷米(Remy)等人,《生物共轭化学》5:647-654(1994);高(Gao)等人,《基因疗法》 (Gene Therapy)2:710-722(1995);艾哈迈德(Ahmad)等人,《癌症研 究》(Cancer Res.)52:4817-4820(1992);美国专利号4,186,183、 4,217,344、4,235,871、4,261,975、4,485,054、4,501,728、4,774,085、4,837,028以及4,946,787)。

[0142] 使用RNA或DNA病毒基的系统递送核酸利用将病毒靶向体内的特 定细胞并将病毒有效负荷(payload)运至细胞核中的高度进化的过程。可以 将病毒载体直接给予至患者(体内)或可以使用它们在体外处理细胞,并且 任选地,可以将修饰的细胞给予至患者(离体)。常规的病毒基的系统可以 包括用于基因转移的逆转录病毒载体、慢病毒载体、腺病毒载体、腺相关病 毒载体以及单纯疱疹病毒载体。用逆转录病毒、慢病毒和腺相关病毒基因转 移方法整合进宿主基因组中是可能的,这通常导致插入转基因的长期表达。另外,已经在不同细胞类型和靶组织中观察到高转导效率。

[0143] 可以通过掺入外源包膜蛋白,扩展靶细胞的潜在靶群而改变逆转录 病毒的向性。慢病毒载体是能够转导或感染非分裂细胞并典型地产生较高病 毒效价的逆转录病毒载体。因此,逆转录病毒基因转移系统的选择将依赖于 靶组织。逆转录病毒载体由顺式作用长末端重复组成,这些长末端重复具有 包装多达6-10kb的外源序列的能力。最低量的顺式作用LTR对于载体的复 制和包装而言是足够的,然后使用这些载体将治疗基因整合进靶细胞中,以 提供永久的转基因表达。广泛使用的逆转录病毒载体包括基于鼠白血病病毒(MuLV)、长臂猿白血病病毒(GaLV)、猴免疫缺陷病毒(SIV)、人类 免疫缺陷病毒(HIV)及其组合的那些(参见例如,布赫谢尔(Buchscher) 等人,《病毒学杂志》(J.Virol.)66:2731-2739(1992);约翰(Johann) 等人,《病毒学杂志》66:1635-1640(1992);佐姆内尔费尔特(Sommnerfelt)等人,《病毒学》(Virol.)176:58-59(1990);威尔逊 (Wilson)等人,《病毒学杂志》63:2374-2378(1989);米勒(Miller)等 人,《病毒学杂志》65:2220-2224(1991);PCT/US 94/05700)。在瞬时表 达是优选的应用中,可以使用腺病毒基系统。腺病毒基载体在许多细胞类型 中能够具有非常高的转导效并且无需细胞分裂。用这样的载体,已经获得了较高的效价和表达水平。可以在相对简单的系统中大量地产生此载体。还可 以使用腺相关病毒("AAV")载体转导具有靶核酸的细胞,例如,在体外 产生核酸和肽,以及用于体内和离体基因疗法程序(参见例如,韦斯特 (West)等人,《病毒学》(Virology)160:38-47(1987);美国专利号 4,797,368;WO 93/24641;科丁(Kotin),《人类基因疗法》(Human Gene Therapy)5:793-801(1994);缪斯科斯卡(Muzyczka),《临床研究杂 志》(J.Clin.Invest.)94:1351(1994))。重组AAV载体的构建描述于多 个出版物中,包括美国专利号5,173,414;特拉特斯金(Tratschin)等人,《分子与细胞生物学》(Mol.Cell.Biol.)5:3251-3260(1985);特拉特斯金 等人,《分子与细胞生物学》4:2072-2081(1984);埃尔莫奈特(Hermonat)&缪斯科斯卡(Muzyczka),《美国国家科学院院刊》 (PNAS)81:6466-6470(1984);以及萨穆尔斯基(Samulski)等人,《病 毒学杂志》(J.Virol.)63:03822-3828(1989)。

[0144] 典型地使用包装细胞形成能够感染宿主细胞的病毒粒子。这样的细 胞包括包装腺病毒的293细胞和包含逆转录病毒的ψ2细胞或PA317细胞。在 基因疗法中使用的病毒载体通常由将核酸载体包装进病毒粒子的细胞系产 生。这些载体典型地包含包装并且随后整合进宿主所需的最低量序列,其他 病毒序列由用于有待表达的一个或多个多核苷酸的表达盒替换。失去的病毒 功能典型地由包装细胞系反式地提供。例如,在基因疗法中使用的AAV载体 典型地仅具有来自AAV基因组的ITR序列,这些序列为包装并整合进宿主基 因组所需。将病毒DNA包装进以下细胞系中,该细胞系包含编码辅助质粒的 其他AAV基因,即rep和cap,但缺少ITR序列。该细胞系还可以被作为辅 助者的腺病毒感染。该辅助病毒促进AAV载体的复制和从辅助质粒表达 AAV基因。由于缺少ITR序列,未以显著的量包装该辅助质粒。可以通过例 如与AAV相比腺病毒更加敏感的热处理减少腺病毒的污染。用于将核酸递送 至细胞的另外的方法是本领域的技术人员已知的。参见例如通过引用结合在 此的US 20030087817。

[0145] 在一些实施例中,用一个或多个在此描述的载体瞬时地或非瞬时地 转染宿主细胞。在一些实施例中,当细胞天然地出现在受试者体内时将其转 染。在一些实施例中,被转染的细胞取自受试者。在一些实施例中,该细胞 来源于取自受试者的细胞,如细胞系。用于

组织培养的多种多样的细胞系在 本领域是已知的。细胞系的实例包括但不限于,C8161、CCRF-CEM、 MOLT、mIMCD-3、NHDF、HeLa-S3、Huh1、Huh4、Huh7、HUVEC、 HASMC、HEKn、HEKa、MiaPaCell、Panc1、PC-3、TF1、CTLL-2、C1R、 Rat6、CV1、RPTE、A10、T24、J82、A375、ARH-77、Calu1、SW480、 SW620、SKOV3、SK-UT、CaCo2、P388D1、SEM-K2、WEHI-231、HB56、 TIB55、Jurkat、J45.01、LRMB、Bcl-1、BC-3、IC21、DLD2、Raw264.7、 NRK、NRK-52E、MRC5、MEF、HepG2、海拉B、海拉T4、COS、COS- 1、COS-6、COS-M6A、BS-C-1猴肾上皮细胞、BALB/3T3小鼠胚胎成纤维细 胞、3T3 Swiss、3T3-L1、132-d5人类胎儿成纤维细胞;10.1小鼠成纤维细 胞、293-T、3T3、721、9L、A2780、A2780ADR、A2780cis、A172、A20、 A253、A431、A-549、ALC、B16、B35、BCP-1细胞、BEAS-2B、bEnd.3、 BHK-21、BR 293、BxPC3、C3H-10T1/2、C6/36、Cal-27、CHO、CHO-7、 CHO-IR、CHO-K1、CHO-K2、CHO-T、CHO Dhfr-/-、COR-L23、COR- L23/CPR、COR-L23/5010、COR-L23/R23、COS-7、COV-434、CML T1、 CMT、CT26、D17、DH82、DU145、DuCaP、EL4、EM2、EM3、 EMT6/AR1、EMT6/AR10.0、FM3、H1299、H69、HB54、HB55、HCA2、 HEK-293、HeLa、Hepa1c1c7、HL-60、HMEC、HT-29、Jurkat、JY细胞、 K562细胞、Ku812、KCL22、KG1、KYO1、LNCap、Ma-Mel 1-48、MC- 38、MCF-7、MCF-10A、MDA-MB-231、MDA-MB-468、MDA-MB-435、 MDCK II、MDCK II、MOR/0.2R、MONO-MAC 6、MTD-1A、MyEnd、NCI- H69/CPR、NCI-H69/LX10、NCI-H69/LX20、NCI-H69/LX4、NIH-3T3、 NALM-1、NW-145、OPCN/OPCT细胞系、Peer、PNT-1A/PNT 2、RenCa、 RIN-5F、RMA/RMAS、Saos-2细胞、Sf-9、SkBr3、T2、T-47D、T84、THP1 细胞系、U373、U87、U937、VCaP、维洛(Vero)细胞、WM39、WT-49、 X63、YAC-1、YAR及其转基因变种。细胞系可获得自本领域的技术人员已 知的多种来源(参见例如,美国典型培养物保藏中心(American Type Culture Collection)(ATCC)(马纳萨斯(Manassus),弗吉尼亚州))。在一些实 施例中,使用用一个或多个在此描述的载体转染的细胞建立新的细胞系,该 新的细胞系包括一个或多个载体来源的序列。在一些实施例中,使用用如在 此描述的CRISPR系统的组分转染(如通过用一个或多个载体进行瞬时转染 或用RNA进行转染)且通过CRISPR复合物的活性修饰的细胞建立新的细胞 系,该新的细胞系包括以下细胞,这些细胞包含吸水但是缺少任何其他外源 序列。在一些实施例中,在评估一种或多种测试化合物中使用用一个或多个 在此描述的载体瞬时或非瞬时转染的细胞或来源于这样的细胞的细胞系。

[0146] 在一些实施例中,使用一个或多个在此描述的载体产生非人转基因 动物或转基因植物。在一些实施例中,该转基因动物是一种哺乳动物,如小 鼠、大鼠或兔。在某些实施例中,该生物体或受试者是植物。在某些实施例 中,该生物体或受试者或植物是藻类。用于产生转基因植物和动物的方法在 本领域是已知的,并且通常以如在此描述的细胞转染方法开始。

[0147] 在一个方面,本发明提供了用于修饰在一种真核细胞中的靶多核苷 酸的方法。在一些实施例中,该方法包括允许一种CRISPR复合物结合到该 靶多核苷酸上以实施所述靶多核苷酸的切割,由此修饰该靶多核苷酸,其中 该CRISPR复合物包括与杂交到在所述靶多核苷酸内的一个靶序列上的指导 序列复合的CRISPR酶,其中所述指导序列连接到一个tracr配对序列上,该 tracr配对序列进而杂交到一个tracr序列上。

[0148] 在一个方面,本发明提供了修饰一种多核苷酸在真核细胞中的表达 的方法。在一些实施例中,该方法包括允许一种CRISPR复合物结合到该多 核苷酸上,使得所述结合导致所述多核苷酸表达增加或减少;其中该CRISPR 复合物包含与杂交到在所述靶多核苷酸内

的一个靶序列上的指导序列复合的 CRISPR酶,其中所述指导序列连接到一个tracr配对序列上,该tracr配对序 列进而杂交到一个tracr序列上。

[0149] 随着作物基因组学的最新进展,使用CRISPR-Cas系统进行有效且 符合成本效益的基因编辑和操纵的能力将允许快速筛选并比较单个和多元遗 传操作,以转化这样的基因组,以便改善生产并增强性状。在此方面,参考 美国专利和出版物:美国专利号6,603,061-农杆菌介导的植物转化法 (Agrobacterium-Mediated Plant Transformation Method);美国专利号 7,868,149-植物基因组序列及其用途(Plant Genome Sequences and Uses Thereof)以及US 2009/0100536-转具有增强的农艺性状的基因植物(Transgenic Plants with Enhanced Agronomic Traits),将每者的所有内容和披 露通过引用以其全文结合在此。在本发明的实践中,莫雷尔(Morrell)等人“作物基因组学:进展与应用(Crop genomics:advances and applications)”《遗传学自然评论》(Nat Rev Genet.)2011年12月29日;13(2):85-96的内 容和披露也通过引用以其全文结合在此。在本发明的一个有利实施例中,该 CRISPR/Cas9系统被用以工程化微藻类(实例15)。因此,加上必要的变 更,此处对动物细胞的提及也可适用植物细胞,除非另外是显然的。

[0150] 在一个方面,本发明提供了修饰真核细胞中的靶多核苷酸的方法, 这些方法可以在体内、离体或在体外进行。在一些实施例中,该方法包括从 人或非人动物或植物(包括微藻)取样细胞或细胞群,并且修饰该细胞或这 些细胞。培养可以发生在离体的任何阶段。该细胞或这些细胞甚至可以被重 新引入该非人动物或植物(包括微藻)中。

[0151] 在植物中,病原体常常是宿主特异性的。例如,番茄尖镰孢菌番茄 专化型(Fusarium oxysporum f.sp.lycopersici)引起番茄枯萎病,而且只攻击 番茄,并且香石竹尖镰孢禾柄锈菌小麦专化型(F.oxysporum f.dianthii Puccinia graminis f.sp.tritici)只攻击小麦。植物具有抵抗大多数病原体的现 有的和诱导性的防御。跨植物代的突变和重组事件导致产生敏感性的遗传变 异性,特别是当病原体以比植物更高的频率繁殖时。在植物中可以存在非宿 主抗性,例如,该宿主和病原体是不相容的。还可以存在水平抗性,例如典 型地由许多基因控制的针对所有种的病原体的部分抗性,以及垂直抗性,例 如典型地由很少的基因控制的针对病原体的某些种而不是其他种的完全抗 性。在基因对基因水平中,植物和病原体一起演化,并且在一者中的遗传变 化使在另一者中的变化平衡。因此,使用自然变异,育种者针对产率、质 量、均匀性、耐性、抗性将最有用的基因进行结合。抗性基因的来源包括天 然或外来品种、传家宝品种(Heirloom Varieties)、近缘野生植物、以及诱导 的突变,例如用诱变剂处理植物材料。利用本发明,向植物育种者提供了一 种新的诱导突变的工具。因此,本领域的技术人员可以分析抗性基因的来源 的基因组,并且在具有所希望的特征或性状的品种方面,采用本发明来诱导 抗性基因的发生,这样具有比诱变剂更好的精确性,并且因此加速并改良植 物育种计划。

[0152] 在一个方面,本发明提供了以下试剂盒,这些试剂盒包含披露于以 上方法和组合物中的元件中的任何一个或多个。在一些实施例中,该试剂盒 包括一种载体系统以及用于使用该试剂盒的说明书。在一些实施例中,该载 体系统包括:(a)一个第一调节元件,该第一调节元件可操作地连接到一个 tracr配对序列以及用于在该tracr配对序列的上游插入一个指导序列的一个或 多个插入位点,其中当表达时,该指导序列引导CRISPR复合物在真核细胞 中与一个靶序列的序列特异性结合,其中该CRISPR复合物包括与以下各项 复合的

一种CRISPR酶:(1)杂交到该靶序列的指导序列,以及(2)杂交 到该tracr序列的tracr配对序列;和/或(b)一个第二调节元件,该第二调节 元件可操作地连接到包括一个核定位序列的编码所述CRISPR酶的酶编码序 列。元件可以单独地或组合地提供,并且可以被提供于任何适合的容器中, 如小瓶、瓶子或管。在一些实施例中,该试剂盒包括一种或多种语言,例如 多于一种语言的说明书。

[0153] 在一些实施例中,试剂盒包括一种或多种用于在利用在此描述的元 件中的一种或多种的方法中使用的试剂。试剂可以被提供于任何适合的容器 中。例如,试剂盒可以提供一种或多种反应或存储缓冲液。可以按在具体测 定中可用的形式或按在使用之前需要添加一种或多种其他组分的形式(例如 按浓缩或冻干形式)提供试剂。缓冲液可以是任何缓冲液,包括但不限于碳 酸钠缓冲液、碳酸氢钠缓冲液、硼酸盐缓冲液、Tris缓冲液、MOPS缓冲液、 HEPES缓冲液及其组合。在一些实施例中,该缓冲液是碱性的。在一些实施 例中,该缓冲液具有从约7至约10的pH。在一些实施例中,该试剂盒包括 一个或多个寡核苷酸,该一个或多个寡核苷酸对应于一个用于插入进载体中 的指导序列,以便可操作地连接该指导序列和一个调节元件。在一些实施例 中,该试剂盒包括一个同源重组模板多核苷酸。

[0154] 在一个方面,本发明提供了用于使用CRISPR系统的一个或多个元 件的方法。本发明的CRISPR复合物提供了用于修饰靶多核苷酸的有效手 段。本发明的CRISPR复合物具有多种多样的实用性,包括修饰(例如,缺 失、插入、转位、失活、激活)多种细胞类型中的靶多核苷酸。正因为如 此,本发明的CRISPR复合物在例如基因疗法、药物筛选、疾病诊断以及预后中具有广阔的应用谱。示例性CRISPR复合物包括与一个指导序列复合的 CRISPR酶,该指导序列与靶多核苷酸内的靶序列杂交。该指导序列与tracr 配对序列连接,该tracr配对序列进而与一个tracr序列杂交。

[0155] CRISPR复合物的靶多核苷酸可以是对真核细胞而言内源或外源的 任何多核苷酸。例如,该靶多核苷酸可以是一种驻留在真核细胞的细胞核中 的多核苷酸。该靶多核苷酸可以是一个编码基因产物(例如,蛋白质)的序 列或一个非编码序列(例如,调节多核苷酸或无用DNA)。不希不希望被理 论所束缚,据信该靶序列应该与PAM(原型间隔子相邻基序)相关;也就是 说,由CRISPR复合物识别的短序列相关。对PAM的精确序列和长度要求取决于使用的CRISPR酶而不同,但是PAM典型地是临近原型间隔子(也就是 说,靶序列)的2-5个碱基对序列。在以下实例部分中给出PAM序列的实 例,并且熟练人员将能够鉴定与给定的CRISPR酶一起使用的另外的PAM序 列。

[0156] CRISPR复合物的靶多核苷酸可以包括多个疾病相关基因和多核苷 酸以及信号传导生化途径相基因和多核苷酸,如分别提交于2012年12月12 日和2013年1月2日、标题均为“用于序列操纵的系统方法和组合物” (SYSTEMS METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION)的分别具有博德参考号BI-2011/008/WSGR案卷号44063-701.101和BI-2011/008/WSGR案卷号44063-701.102的美国临时专利申请 61/736,527和61/748,427中所列举,将所有这些申请的内容通过引用以其全文 结合在此。

[0157] 靶多核苷酸的实例包括与信号传导生化途径相关的序列,例如信号 传导生化途径相关基因或多核苷酸。靶多核苷酸的实例包括疾病相关基因或 多核苷酸。“疾病相关”基因或多核苷酸是指与非疾病对照的组织或细胞相 比,在来源于疾病影响的组织的细胞中以异常水平或以异常形式产生转录或 翻译产物的任何基因或多核苷酸。在改变的表达与

疾病的出现和/或进展相关 的情况下,它可以是一个以异常高的水平被表达的基因;它可以是一个以异 常低的水平被表达的基因。疾病相关基因还指具有一个或多个突变或直接负 责或与一个或多个负责疾病的病因学的基因连锁不平衡的遗传变异的基因。转录的或翻译的产物可以是已知的或未知的,并且可以处于正常或异常水 平。

[0158] 疾病相关基因和多核苷酸的实例可获得自约翰斯·霍普金斯大学的 麦考斯克-纳森遗传医学研究所(McKusick-Nathans Institute of Genetic Medicine,Johns Hopkins University)(巴尔的摩,马里兰州)和国立医学图书 馆的国家生物技术信息中心(National Center for Biotechnology Information, National Library of Medicine)(贝塞斯达,马里兰州)。

[0159] 疾病相关基因和多核苷酸的实例列于表A和B中。在万维网上可 获得的疾病具体信息可获得自约翰斯·霍普金斯大学的麦考斯克-纳森遗传医学 研究所(McKusick-Nathans Institute of Genetic Medicine,Johns Hopkins University)(巴尔的摩,马里兰州)和国立医学图书馆的国家生物技术信息 中心(National Center for Biotechnology Information,National Library of Medicine)(贝塞斯达,马里兰州)。信号传导生化途径相关基因和多核苷酸 的实例列于表C中。

[0160] 这些基因和途径的突变可以导致产生不当的蛋白质或以不当的量影 响功能的蛋白质。通过引用而特此结合来自于2012年12月12日提交的美国 临时申请61/736,527和2013年2月2日提交的61/748,427的基因、疾病和蛋 白质的另外的实例。这样的基因、蛋白质和途径可以是CRISPR复合物的靶 多核苷酸。

表A

疾病/障碍	一个或多个基因
瘤形成	PTEN; ATM; ATR; EGFR; ERBB2; ERBB3; ERBB4;
	Notch1; Notch2; Notch3; Notch4; AKT; AKT2; AKT3; HIF;
	HIF1a; HIF3a; Met; HRG; Bcl2; PPARα; PPAR
	γ; WT1 (维耳姆斯瘤 (Wilms Tumor)); FGF 受体家族
	成员 (5 个成员: 1、2、3、4、5); CDKN2a; APC; RB
	(视网膜母细胞瘤); MEN1; VHL; BRCA1; BRCA2; AR
	(雄激素受体); TSG101; IGF; IGF 受体; Igf1 (4
	种变体); Igf2 (3 种变体); Igf 1 受体; Igf 2 受体;
	Bax; Bcl2; 半胱天冬酶家族 (9 个成员:

	1、2、3、4、6、7、8、9、12）；Kras；Apc
年龄相关性黄斑	Abcr；Ccl2；Cc2；cp（铜蓝蛋白）；Timp3；组织蛋白酶 D；
变性	Vldlr；Ccr2
精神分裂症	神经调节蛋白 1（Nrg1）；Erb4（神经调节蛋白受体）；
	复蛋白（Complexin）1（Cplx1）；Tph1 色氨酸羟化酶；Tph2
	色氨酸羟化酶 2；轴突蛋白（Neurexin）1；GSK3；GSK3a；
	GSK3b
障碍	5-HTT（Slc6a4）；COMT；DRD（Drd1a）；SLC6A3；DAOA；
	DTNBP1；Dao（Dao1）
三核苷酸重复	HTT（亨廷顿氏 Dx）；SBMA/SMAX1/AR（肯尼迪氏
障碍	Dx）；FXN/X25（弗里德里希共济失调）；ATX3（马查多-
	约瑟夫氏 Dx）；ATXN1 和 ATXN2（脊髓小脑
	共济失调）；DMPK（强直性肌营养不良）；Atrophin-1 和 Atn1
	（DRPLA Dx）；CBP（Creb-BP -整体不稳定性）；VLDLR
	（阿尔茨海默病）；Atxn7；Atxn10
脆性 X 综合征	FMR2；FXR1；FXR2；mGLUR5
分泌酶相关	APH-1（α和β）；早老素（Psen1）；呆蛋白
障碍	（Ncstn）；PEN-2
其他	Nos1；Parp1；Nat1；Nat2
朊病毒相关障碍	Prp
ALS	SOD1；ALS2；STEX；FUS；TARDBP；VEGF（VEGF-a；
	VEGF-b；VEGF-c）
药物成瘾	Prkce（酒精）；Drd2；Drd4；ABAT（酒精）；GRIA2；
	Grm5；Grin1；Htr1b；Grin2a；Drd3；Pdyn；Gria1（酒精）
自闭症	Mecp2；BZRAP1；MDGA2；Sema5A；轴突蛋白 1；脆性 X
	（FMR2（AFF2）；FXR1；FXR2；Mglur5）
阿尔茨海默病	E1；CHIP；UCH；UBB；Tau；LRP；PICALM；簇集蛋白；PS1；
	SORL1；CR1；Vldlr；Uba1；Uba3；CHIP28（Aqp1，
	水通道蛋白 1）；Uchl1；Uchl3；APP
炎症	IL-10；IL-1（IL-1a；IL-1b）；IL-13；IL-17（IL-17a（CTLA8）；IL-
	17b；IL-17c；IL-17d；IL-17f）；II-23；Cx3cr1；ptpn22；TNFa；

	IBD 的 NOD2/CARD15；IL-6；IL-12（IL-12a；IL-12b）；
	CTLA4；Cx3cl1
帕金森病	x-突触核蛋白；DJ-1；LRRK2；帕金蛋白（Parkin）；PINK1

表B：

血液和凝血疾病以及障碍	贫血（CDAN1、CDA1、RPS19、DBA、PKLR、PK1、NT5C3、UMPH1、PSN1、RHAG、RH50A、NRAMP2、SPTB、ALAS2、ANH1、ASB、ABCB7、ABC7、ASAT）；裸淋巴细胞综合征（TAPBP、TPSN、TAP2、ABCB3、PSF2、RING11、MHC2TA、C2TA、RFX5、RFXAP、RFX5），出血障碍（TBXA2R、P2RX1、P2X1）；因子 H 和因子 H 样 1（HF1、CFH、HUS）；因子 V 和因子 VIII（MCFD2）；因子 VII 缺乏（F7）；因子 X 缺乏（F10）；因子 XI 缺乏（F11）；因子 XII 缺乏（F12、HAF）；因子 XIIIA 缺乏（F13A1、F13A）；因子 XIIIB 缺乏（F13B）；范科尼贫血（FANCA、FACA、FA1、FA、FAA、FAAP95、FAAP90、FLJ34064、FANCB、FANCC、FACC、BRCA2、FANCD1、FANCD2、FANCD、FACD、FAD、FANCE、FACE、FANCF、XRCC9、FANCG、BRIP1、BACH1、FANCJ、PHF9、FANCL、FANCM、KIAA1596）；噬血细胞淋巴组织细胞增生障碍（PRF1、HPLH2、UNC13D、MUNC13-4、HPLH3、HLH3、FHL3）；血友病 A（F8、F8C、HEMA）；血友病 B（F9、HEMB），出血性障碍（PI、ATT、F5）；白细胞缺乏和障碍（ITGB2、CD18、LCAMB、LAD、EIF2B1、EIF2BA、EIF2B2、EIF2B3、EIF2B5、LVWM、CACH、CLE、EIF2B4）；镰状细胞贫血（HBB）；地中海贫血（HBA2、HBB、HBD、LCRB、HBA1）。
细胞调节异常和肿瘤疾病以及障碍	B 细胞非霍奇金淋巴瘤（BCL7A、BCL7）；白血病（TAL1、TCL5、SCL、TAL2、FLT3、NBS1、NBS、ZNFN1A1、IK1、LYF1、HOXD4、HOX4B、BCR、CML、PHL、ALL、ARNT、KRAS2、RASK2、GMPS、AF10、ARHGEF12、LARG、KIAA0382、CALM、CLTH、CEBPA、CEBP、CHIC2、BTL、FLT3、KIT、PBT、LPP、NPM1、NUP214、D9S46E、CAN、CAIN、RUNX1、CBFA2、AML1、WHSC1L1、NSD3、FLT3、AF1Q、NPM1、NUMA1、

	ZNF145、PLZF、PML、MYL、STAT5B、AF10、CALM、CLTH、ARL11、ARLTS1、P2RX7、P2X7、BCR、CML、PHL、ALL、GRAF、NF1、VRNF、WSS、NFNS、PTPN11、PTP2C、SHP2、NS1、BCL2、CCND1、PRAD1、BCL1、TCRA、GATA1、GF1、ERYF1、NFE1、ABL1、NQO1、DIA4、NMOR1、NUP214、D9S46E、CAN、CAIN）。
炎症和免疫相关疾病以及障碍	AIDS（KIR3DL1、NKAT3、NKB1、AMB11、KIR3DS1、IFNG、CXCL12、SDF1）；自身免疫性淋巴增生综合征（TNFRSF6、APT1、FAS、CD95、ALPS1A）；联合免疫缺陷，（IL2RG、SCIDX1、SCIDX、IMD4）；HIV-1（CCL5、SCYA5、D17S136E、TCP228），HIV 易感性或感染（IL10、CSIF、CMKBR2、CCR2、CMKBR5、CCCKR5（CCR5））；免疫缺陷（CD3E、CD3G、AICDA、AID、HIGM2、TNFRSF5、CD40、UNG、DGU、HIGM4、TNFSF5、CD40LG、HIGM1、IGM、FOXP3、IPEX、AIID、XPID、PIDX、TNFRSF14B、TACI）；炎症（IL-10、IL-1（IL-1a、IL-1b）、IL-13、IL-17（IL-17a（CTLA8）、IL-17b、IL-17c、IL-17d、IL-17f）、II-23、Cx3cr1、ptpn22、TNFa、IBD 的 NOD2/CARD15、IL-6、IL-12（IL-12a、IL-12b）、CTLA4、Cx3cl1）；重症联合免疫缺陷（SCID）（JAK3、JAKL、DCLRE1C、ARTEMIS、SCIDA、RAG1、RAG2、ADA、PTPRC、CD45、LCA、IL7R、CD3D、T3D、IL2RG、SCIDX1、SCIDX、IMD4）。
代谢、肝、肾和蛋白疾病以及障碍	淀粉样蛋白神经病变（TTR、PALB）；淀粉样变性（APOA1、APP、AAA、CVAP、AD1、GSN、FGA、LYZ、TTR、PALB）；肝硬化（KRT18、KRT8、CIRH1A、NAIC、TEX292、KIAA1988）；囊性纤维化（CFTR、ABCC7、CF、MRP7）；糖原贮积病（SLC2A2、GLUT2、G6PC、G6PT、G6PT1、GAA、LAMP2、LAMPB、AGL、GDE、GBE1、GYS2、PYGL、PFKM）；肝腺瘤，142330（TCF1、HNF1A、MODY3），肝衰竭，早期发病以及神经障碍（SCOD1、SCO1），肝脂酶缺乏（LIPC）、肝母细胞癌、癌症（cancer）和癌（carcinomas）（CTNNB1、PDGFRL、PDGRL、PRLTS、AXIN1、AXIN、CTNNB1、TP53、P53、LFS1、IGF2R、

	MPRI、MET、CASP8、MCH5；髓质囊肾病（UMOD、HNFJ、FJHN、MCKD2、ADMCKD2）；苯丙酮尿症（PAH、PKU1、QDPR、DHPR、PTS）；多囊肾和肝病（FCYT、PKHD1、ARPKD、PKD1、PKD2、PKD4、PKDTS、PRKCSH、G19P1、PCLD、SEC63）。
肌肉/骨骼疾病以及障碍	贝克肌营养不良（DMD、BMD、MYF6），杜氏肌营养不良（Duchenne Muscular Dystrophy）（DMD、BMD）；埃默里-德赖富斯（Emery-Dreifuss）肌营养不良（LMNA、LMN1、EMD2、FPLD、CMD1A、HGPS、LGMD1B、LMNA、LMN1、EMD2、FPLD、CMD1A）；面肩肱型肌营养不良（FSHMD1A、FSHD1A）；肌营养不良（FKRP、MDC1C、LGMD2I、LAMA2、LAMM、LARGE、KIAA0609、MDC1D、FCMD、TTID、MYOT、CAPN3、CANP3、DYSF、LGMD2B、SGCG、LGMD2C、DMDA1、SCG3、SGCA、ADL、DAG2、LGMD2D、DMDA2、SGCB、LGMD2E、SGCD、SGD、LGMD2F、CMD1L、TCAP、LGMD2G、CMD1N、TRIM32、HT2A、LGMD2H、FKRP、MDC1C、LGMD2I、TTN、CMD1G、TMD、LGMD2J、POMT1、CAV3、LGMD1C、SEPN1、SELN、RSMD1、PLEC1、PLTN、EBS1）；骨硬化病（LRP5、BMND1、LRP7、LR3、OPPG、VBCH2、CLCN7、CLC7、OPTA2、OSTM1、GL、TCIRG1、TIRC7、OC116、OPTB1）；肌萎缩（VAPB、VAPC、ALS8、SMN1、SMA1、SMA2、SMA3、SMA4、BSCL2、SPG17、GARS、SMAD1、CMT2D、HEXB、IGHMBP2、SMUBP2、CATF1、SMARD1）。
神经和神经元疾病以及障碍	ALS（SOD1、ALS2、STEX、FUS、TARDBP、VEGF（VEGF-a、VEGF-b、VEGF-c）；阿尔茨海默病（APP、AAA、CVAP、AD1、APOE、AD2、PSEN2、AD4、STM2、APBB2、FE65L1、NOS3、PLAU、URK、ACE、DCP1、ACE1、MPO、PACIP1、PAXIP1L、PTIP、A2M、BLMH、BMH、PSEN1、AD3）；自闭症（Mecp2、BZRAP1、MDGA2、Sema5A、轴突蛋白 1、GLO1、MECP2、RTT、PPMX、MRX16、MRX79、NLGN3、NLGN4、KIAA1260、AUTSX2）；脆性 X 综合征（FMR2、FXR1、FXR2、mGLUR5）；亨廷顿病和疾病样障碍

	（HD、IT15、PRNP、PRIP、JPH3、JP3、HDL2、TBP、SCA17）；帕金森病（NR4A2、NURR1、NOT、TINUR、SNCAIP、TBP、SCA17、SNCA、NACP、PARK1、PARK4、DJ1、PARK7、LRRK2、PARK8、PINK1、PARK6、UCHL1、PARK5、SNCA、NACP、PARK1、PARK4、PRKN、PARK2、PDJ、DBH、NDUFV2）；雷特综合征（Rett syndrome）（MECP2、RTT、PPMX、MRX16、MRX79、CDKL5、STK9、MECP2、RTT、PPMX、MRX16、MRX79、x-突触核蛋白、DJ-1）；精神分裂症（神经调节蛋白 1（Nrg1）、Erb4（神经调节蛋白受体）、复蛋白 1（Cplx1）、Tph1 色氨酸羟化酶、Tph2、色氨酸羟化酶 2、轴突蛋白 1、GSK3、GSK3a、GSK3b、5-HTT（Slc6a4）、COMT、DRD（Drd1a）、SLC6A3、DAOA、DTNBP1、Dao（Dao1））；分泌酶相关障碍（APH-1（α 和 β）、早老素（Psen1）、呆蛋白，（Ncstn）、PEN-2、Nos1、Parp1、Nat1、Nat2）；三核苷酸重复障碍（HTT（亨廷顿氏 Dx）、SBMA/SMAX1/AR（肯尼迪氏 Dx）、FXN/X25（弗里德里希共济失调）、ATX3（马查多-约瑟夫氏 Dx）、ATXN1 和 ATXN2（脊髓小脑共济失调）、DMPK（强直性肌营养不良）、Atrophin-1 和 Atn1（DRPLA Dx）、CBP（Creb-BP-整体不稳定性）、VLDLR（阿尔茨海默病）、Atxn7、Atxn10）。
眼部疾病以及障碍	年龄相关性黄斑变性（Abcr、Ccl2、Cc2、cp（铜蓝蛋白）、Timp3、组织蛋白酶 D、Vldlr、Ccr2）；白内障（CRYAA、CRYA1、CRYBB2、CRYB2、PITX3、BFSP2、CP49、CP47、CRYAA、CRYA1、PAX6、AN2、MGDA、CRYBA1、CRYB1、CRYGC、CRYG3、CCL、LIM2、MP19、CRYGD、CRYG4、BFSP2、CP49、CP47、HSF4、CTM、HSF4、CTM、MIP、AQP0、CRYAB、CRYA2、CTPP2、CRYBB1、CRYGD、CRYG4、CRYBB2、CRYB2、CRYGC、CRYG3、CCL、CRYAA、CRYA1、GJA8、CX50、CAE1、GJA3、CX46、CZP3、CAE3、CCM1、CAM、KRIT1）；角膜混浊和营养不良（APOA1、TGFBI、CSD2、CDGG1、CSD、BIGH3、CDG2、TACSTD2、TROP2、M1S1、VSX1、RINX、PPCD、PPD、KTCN、COL8A2、FECD、PPCD2、PIP5K3、CFD）；角膜先天性扁平（Cornea plana congenital）（KERA、

	CNA2）；青光眼（MYOC、TIGR、GLC1A、JOAG、GPOA、OPTN、GLC1E、FIP2、HYPL、NRP、CYP1B1、GLC3A、OPA1、NTG、NPG、CYP1B1、GLC3A）；莱伯（Leber）先天性黑朦（CRB1、RP12、CRX、CORD2、CRD、RPGRIP1、LCA6、CORD9、RPE65、RP20、AIPL1、LCA4、GUCY2D、GUC2D、LCA1、CORD6、RDH12、LCA3）；黄斑营养不良（ELOVL4、ADMD、STGD2、STGD3、RDS、RP7、PRPH2、PRPH、AVMD、AOFMD、VMD2）。

表C：

细胞功能	基因
PI3K/AKT信号传导	PRKCE；ITGAM；ITGA5；IRAK1；PRKAA2；EIF2AK2；
	PTEN；EIF4E；PRKCZ；GRK6；MAPK1；TSC1；PLK1；
	AKT2；IKBKB；PIK3CA；CDK8；CDKN1B；NFKB2；BCL2；
	PIK3CB；PPP2R1A；MAPK8；BCL2L1；MAPK3；TSC2；
	ITGA1；KRAS；EIF4EBP1；RELA；PRKCD；NOS3；
	PRKAA1；MAPK9；CDK2；PPP2CA；PIM1；ITGB7；
	YWHAZ；ILK；TP53；RAF1；IKBKG；RELB；DYRK1A；
	CDKN1A；ITGB1；MAP2K2；JAK1；AKT1；JAK2；PIK3R1；
	CHUK；PDPK1；PPP2R5C；CTNNB1；MAP2K1；NFKB1；
	PAK3；ITGB3；CCND1；GSK3A；FRAP1；SFN；ITGA2；
	TTK；CSNK1A1；BRAF；GSK3B；AKT3；FOXO1；SGK；
	HSP90AA1；RPS6KB1
ERK/MAPK信号传导	PRKCE；ITGAM；ITGA5；HSPB1；IRAK1；PRKAA2；
	EIF2AK2；RAC1；RAP1A；TLN1；EIF4E；ELK1；GRK6；
	MAPK1；RAC2；PLK1；AKT2；PIK3CA；CDK8；CREB1；
	PRKCI；PTK2；FOS；RPS6KA4；PIK3CB；PPP2R1A；
	PIK3C3；MAPK8；MAPK3；ITGA1；ETS1；KRAS；MYCN；
	EIF4EBP1；PPARG；PRKCD；PRKAA1；MAPK9；SRC；
	CDK2；PPP2CA；PIM1；PIK3C2A；ITGB7；YWHAZ；
	PPP1CC；KSR1；PXN；RAF1；FYN；DYRK1A；ITGB1；
	MAP2K2；PAK4；PIK3R1；STAT3；PPP2R5C；MAP2K1；
	PAK3；ITGB3；ESR1；ITGA2；MYC；TTK；CSNK1A1；
	CRKL；BRAF；ATF4；PRKCA；SRF；STAT1；SGK

糖皮质激素受体	RAC1；TAF4B；EP300；SMAD2；TRAF6；PCAF；ELK1；
信号传导	MAPK1；SMAD3；AKT2；IKBKB；NCOR2；UBE2I；
	PIK3CA；CREB1；FOS；HSPA5；NFKB2；BCL2；
	MAP3K14；STAT5B；PIK3CB；PIK3C3；MAPK8；BCL2L1；
	MAPK3；TSC22D3；MAPK10；NRIP1；KRAS；MAPK13；
	RELA；STAT5A；MAPK9；NOS2A；PBX1；NR3C1；
	PIK3C2A；CDKN1C；TRAF2；SERPINE1；NCOA3；
	MAPK14；TNF；RAF1；IKBKG；MAP3K7；CREBBP；
	CDKN1A；MAP2K2；JAK1；IL8；NCOA2；AKT1；JAK2；
	PIK3R1；CHUK；STAT3；MAP2K1；NFKB1；TGFBR1；
	ESR1；SMAD4；CEBPB；JUN；AR；AKT3；CCL2；MMP1；
	STAT1；IL6；HSP90AA1
轴突指导信号传导	PRKCE；ITGAM；ROCK1；ITGA5；CXCR4；ADAM12；
	IGF1；RAC1；RAP1A；EIF4E；PRKCZ；NRP1；NTRK2；
	ARHGEF7；SMO；ROCK2；MAPK1；PGF；RAC2；
	PTPN11；GNAS；AKT2；PIK3CA；ERBB2；PRKCI；PTK2；
	CFL1；GNAQ；PIK3CB；CXCL12；PIK3C3；WNT11；
	PRKD1；GNB2L1；ABL1；MAPK3；ITGA1；KRAS；RHOA；
	PRKCD；PIK3C2A；ITGB7；GLI2；PXN；VASP；RAF1；
	FYN；ITGB1；MAP2K2；PAK4；ADAM17；AKT1；PIK3R1；
	GLI1；WNT5A；ADAM10；MAP2K1；PAK3；ITGB3；
	CDC42；VEGFA；ITGA2；EPHA8；CRKL；RND1；GSK3B；
	AKT3；PRKCA
肝配蛋白受体信号传导	PRKCE；ITGAM；ROCK1；ITGA5；CXCR4；IRAK1；
	PRKAA2；EIF2AK2；RAC1；RAP1A；GRK6；ROCK2；
	MAPK1；PGF；RAC2；PTPN11；GNAS；PLK1；AKT2；
	DOK1；CDK8；CREB1；PTK2；CFL1；GNAQ；MAP3K14；
	CXCL12；MAPK8；GNB2L1；ABL1；MAPK3；ITGA1；
	KRAS；RHOA；PRKCD；PRKAA1；MAPK9；SRC；CDK2；
	PIM1；ITGB7；PXN；RAF1；FYN；DYRK1A；ITGB1；
	MAP2K2；PAK4；AKT1；JAK2；STAT3；ADAM10；
	MAP2K1；PAK3；ITGB3；CDC42；VEGFA；ITGA2；

	EPHA8; TTK; CSNK1A1; CRKL; BRAF; PTPN13; ATF4;
	AKT3; SGK
肌动蛋白细胞骨架	ACTN4; PRKCE; ITGAM; ROCK1; ITGA5; IRAK1;
信号传导	PRKAA2; EIF2AK2; RAC1; INS; ARHGEF7; GRK6;
	ROCK2; MAPK1; RAC2; PLK1; AKT2; PIK3CA; CDK8;
	PTK2; CFL1; PIK3CB; MYH9; DIAPH1; PIK3C3; MAPK8;
	F2R; MAPK3; SLC9A1; ITGA1; KRAS; RHOA; PRKCD;
	PRKAA1; MAPK9; CDK2; PIM1; PIK3C2A; ITGB7;
	PPP1CC; PXN; VIL2; RAF1; GSN; DYRK1A; ITGB1;
	MAP2K2; PAK4; PIP5K1A; PIK3R1; MAP2K1; PAK3;
	ITGB3; CDC42; APC; ITGA2; TTK; CSNK1A1; CRKL;
	BRAF; VAV3; SGK
亨廷顿病	PRKCE; IGF1; EP300; RCOR1; PRKCZ; HDAC4; TGM2;
信号传导	MAPK1; CAPNS1; AKT2; EGFR; NCOR2; SP1; CAPN2;
	PIK3CA; HDAC5; CREB1; PRKCI; HSPA5; REST;
	GNAQ; PIK3CB; PIK3C3; MAPK8; IGF1R; PRKD1;
	GNB2L1; BCL2L1; CAPN1; MAPK3; CASP8; HDAC2;
	HDAC7A; PRKCD; HDAC11; MAPK9; HDAC9; PIK3C2A;
	HDAC3; TP53; CASP9; CREBBP; AKT1; PIK3R1;
	PDPK1; CASP1; APAF1; FRAP1; CASP2; JUN; BAX;
	ATF4; AKT3; PRKCA; CLTC; SGK; HDAC6; CASP3
凋亡信号传导	PRKCE; ROCK1; BID; IRAK1; PRKAA2; EIF2AK2; BAK1;
	BIRC4; GRK6; MAPK1; CAPNS1; PLK1; AKT2; IKBKB;
	CAPN2; CDK8; FAS; NFKB2; BCL2; MAP3K14; MAPK8;
	BCL2L1; CAPN1; MAPK3; CASP8; KRAS; RELA;
	PRKCD; PRKAA1; MAPK9; CDK2; PIM1; TP53; TNF;
	RAF1; IKBKG; RELB; CASP9; DYRK1A; MAP2K2;
	CHUK; APAF1; MAP2K1; NFKB1; PAK3; LMNA; CASP2;
	BIRC2; TTK; CSNK1A1; BRAF; BAX; PRKCA; SGK;
	CASP3; BIRC3; PARP1
B 细胞受体信号传导	RAC1; PTEN; LYN; ELK1; MAPK1; RAC2; PTPN11;
	AKT2; IKBKB; PIK3CA; CREB1; SYK; NFKB2; CAMK2A;

	MAP3K14; PIK3CB; PIK3C3; MAPK8; BCL2L1; ABL1;
	MAPK3; ETS1; KRAS; MAPK13; RELA; PTPN6; MAPK9;
	EGR1; PIK3C2A; BTK; MAPK14; RAF1; IKBKG; RELB;
	MAP3K7; MAP2K2; AKT1; PIK3R1; CHUK; MAP2K1;
	NFKB1; CDC42; GSK3A; FRAP1; BCL6; BCL10; JUN;
	GSK3B; ATF4; AKT3; VAV3; RPS6KB1
白细胞外渗	ACTN4; CD44; PRKCE; ITGAM; ROCK1; CXCR4; CYBA;
信号传导	RAC1; RAP1A; PRKCZ; ROCK2; RAC2; PTPN11;
	MMP14; PIK3CA; PRKCI; PTK2; PIK3CB; CXCL12;
	PIK3C3; MAPK8; PRKD1; ABL1; MAPK10; CYBB;
	MAPK13; RHOA; PRKCD; MAPK9; SRC; PIK3C2A; BTK;
	MAPK14; NOX1; PXN; VIL2; VASP; ITGB1; MAP2K2;
	CTNND1; PIK3R1; CTNNB1; CLDN1; CDC42; F11R; ITK;
	CRKL; VAV3; CTTN; PRKCA; MMP1; MMP9
整合素信号传导	ACTN4; ITGAM; ROCK1; ITGA5; RAC1; PTEN; RAP1A;
	TLN1; ARHGEF7; MAPK1; RAC2; CAPNS1; AKT2;
	CAPN2; PIK3CA; PTK2; PIK3CB; PIK3C3; MAPK8;
	CAV1; CAPN1; ABL1; MAPK3; ITGA1; KRAS; RHOA;
	SRC; PIK3C2A; ITGB7; PPP1CC; ILK; PXN; VASP;
	RAF1; FYN; ITGB1; MAP2K2; PAK4; AKT1; PIK3R1;
	TNK2; MAP2K1; PAK3; ITGB3; CDC42; RND3; ITGA2;
	CRKL; BRAF; GSK3B; AKT3
急性期反应	IRAK1; SOD2; MYD88; TRAF6; ELK1; MAPK1; PTPN11;
信号传导	AKT2; IKBKB; PIK3CA; FOS; NFKB2; MAP3K14;
	PIK3CB; MAPK8; RIPK1; MAPK3; IL6ST; KRAS;
	MAPK13; IL6R; RELA; SOCS1; MAPK9; FTL; NR3C1;
	TRAF2; SERPINE1; MAPK14; TNF; RAF1; PDK1;
	IKBKG; RELB; MAP3K7; MAP2K2; AKT1; JAK2; PIK3R1;
	CHUK; STAT3; MAP2K1; NFKB1; FRAP1; CEBPB; JUN;
	AKT3; IL1R1; IL6
PTEN 信号传导	ITGAM; ITGA5; RAC1; PTEN; PRKCZ; BCL2L11;
	MAPK1; RAC2; AKT2; EGFR; IKBKB; CBL; PIK3CA;
	CDKN1B; PTK2; NFKB2; BCL2; PIK3CB; BCL2L1;

	MAPK3; ITGA1; KRAS; ITGB7; ILK; PDGFRB; INSR;
	RAF1; IKBKG; CASP9; CDKN1A; ITGB1; MAP2K2;
	AKT1; PIK3R1; CHUK; PDGFRA; PDPK1; MAP2K1;
	NFKB1; ITGB3; CDC42; CCND1; GSK3A; ITGA2;
	GSK3B; AKT3; FOXO1; CASP3; RPS6KB1
p53 信号传导	PTEN; EP300; BBC3; PCAF; FASN; BRCA1; GADD45A;
	BIRC5; AKT2; PIK3CA; CHEK1; TP53INP1; BCL2;
	PIK3CB; PIK3C3; MAPK8; THBS1; ATR; BCL2L1; E2F1;
	PMAIP1; CHEK2; TNFRSF10B; TP73; RB1; HDAC9;
	CDK2; PIK3C2A; MAPK14; TP53; LRDD; CDKN1A;
	HIPK2; AKT1; PIK3R1; RRM2B; APAF1; CTNNB1;
	SIRT1; CCND1; PRKDC; ATM; SFN; CDKN2A; JUN;
	SNAI2; GSK3B; BAX; AKT3
芳烃受体	HSPB1; EP300; FASN; TGM2; RXRA; MAPK1; NQO1;
信号传导	NCOR2; SP1; ARNT; CDKN1B; FOS; CHEK1;
	SMARCA4; NFKB2; MAPK8; ALDH1A1; ATR; E2F1;
	MAPK3; NRIP1; CHEK2; RELA; TP73; GSTP1; RB1;
	SRC; CDK2; AHR; NFE2L2; NCOA3; TP53; TNF;
	CDKN1A; NCOA2; APAF1; NFKB1; CCND1; ATM; ESR1;
	CDKN2A; MYC; JUN; ESR2; BAX; IL6; CYP1B1;
	HSP90AA1
异生物质代谢	PRKCE; EP300; PRKCZ; RXRA; MAPK1; NQO1;
信号传导	NCOR2; PIK3CA; ARNT; PRKCI; NFKB2; CAMK2A;
	PIK3CB; PPP2R1A; PIK3C3; MAPK8; PRKD1;
	ALDH1A1; MAPK3; NRIP1; KRAS; MAPK13; PRKCD;
	GSTP1; MAPK9; NOS2A; ABCB1; AHR; PPP2CA; FTL;
	NFE2L2; PIK3C2A; PPARGC1A; MAPK14; TNF; RAF1;
	CREBBP; MAP2K2; PIK3R1; PPP2R5C; MAP2K1;
	NFKB1; KEAP1; PRKCA; EIF2AK3; IL6; CYP1B1;
	HSP90AA1
SAPK/JNK 信号传导	PRKCE; IRAK1; PRKAA2; EIF2AK2; RAC1; ELK1;
	GRK6; MAPK1; GADD45A; RAC2; PLK1; AKT2; PIK3CA;
	FADD; CDK8; PIK3CB; PIK3C3; MAPK8; RIPK1;

	GNB2L1; IRS1; MAPK3; MAPK10; DAXX; KRAS;
	PRKCD; PRKAA1; MAPK9; CDK2; PIM1; PIK3C2A;
	TRAF2; TP53; LCK; MAP3K7; DYRK1A; MAP2K2;
	PIK3R1; MAP2K1; PAK3; CDC42; JUN; TTK; CSNK1A1;
	CRKL; BRAF; SGK
PPAr/RXR 信号传导	PRKAA2; EP300; INS; SMAD2; TRAF6; PPARA; FASN;
	RXRA; MAPK1; SMAD3; GNAS; IKBKB; NCOR2;
	ABCA1; GNAQ; NFKB2; MAP3K14; STAT5B; MAPK8;
	IRS1; MAPK3; KRAS; RELA; PRKAA1; PPARGC1A;
	NCOA3; MAPK14; INSR; RAF1; IKBKG; RELB; MAP3K7;
	CREBBP; MAP2K2; JAK2; CHUK; MAP2K1; NFKB1;
	TGFBR1; SMAD4; JUN; IL1R1; PRKCA; IL6; HSP90AA1;
	ADIPOQ
NF-KB 信号传导	IRAK1; EIF2AK2; EP300; INS; MYD88; PRKCZ; TRAF6;
	TBK1; AKT2; EGFR; IKBKB; PIK3CA; BTRC; NFKB2;
	MAP3K14; PIK3CB; PIK3C3; MAPK8; RIPK1; HDAC2;
	KRAS; RELA; PIK3C2A; TRAF2; TLR4; PDGFRB; TNF;
	INSR; LCK; IKBKG; RELB; MAP3K7; CREBBP; AKT1;
	PIK3R1; CHUK; PDGFRA; NFKB1; TLR2; BCL10;
	GSK3B; AKT3; TNFAIP3; IL1R1
神经调节蛋白信号传导	ERBB4; PRKCE; ITGAM; ITGA5; PTEN; PRKCZ; ELK1;
	MAPK1; PTPN11; AKT2; EGFR; ERBB2; PRKCI;
	CDKN1B; STAT5B; PRKD1; MAPK3; ITGA1; KRAS;
	PRKCD; STAT5A; SRC; ITGB7; RAF1; ITGB1; MAP2K2;
	ADAM17; AKT1; PIK3R1; PDPK1; MAP2K1; ITGB3;
	EREG; FRAP1; PSEN1; ITGA2; MYC; NRG1; CRKL;
	AKT3; PRKCA; HSP90AA1; RPS6KB1
Wnt & β 连环蛋白信号传导	CD44; EP300; LRP6; DVL3; CSNK1E; GJA1; SMO;
	AKT2; PIN1; CDH1; BTRC; GNAQ; MARK2; PPP2R1A;
	WNT11; SRC; DKK1; PPP2CA; SOX6; SFRP2; ILK;
	LEF1; SOX9; TP53; MAP3K7; CREBBP; TCF7L2; AKT1;
	PPP2R5C; WNT5A; LRP5; CTNNB1; TGFBR1; CCND1;

	GSK3A；DVL1；APC；CDKN2A；MYC；CSNK1A1；GSK3B；
	AKT3；SOX2
胰岛素受体信号传导	PTEN；INS；EIF4E；PTPN1；PRKCZ；MAPK1；TSC1；
	PTPN11；AKT2；CBL；PIK3CA；PRKCI；PIK3CB；PIK3C3；
	MAPK8；IRS1；MAPK3；TSC2；KRAS；EIF4EBP1；
	SLC2A4；PIK3C2A；PPP1CC；INSR；RAF1；FYN；
	MAP2K2；JAK1；AKT1；JAK2；PIK3R1；PDPK1；MAP2K1；
	GSK3A；FRAP1；CRKL；GSK3B；AKT3；FOXO1；SGK；
	RPS6KB1
IL-6 信号传导	HSPB1；TRAF6；MAPKAPK2；ELK1；MAPK1；PTPN11；
	IKBKB；FOS；NFKB2；MAP3K14；MAPK8；MAPK3；
	MAPK10；IL6ST；KRAS；MAPK13；IL6R；RELA；SOCS1；
	MAPK9；ABCB1；TRAF2；MAPK14；TNF；RAF1；IKBKG；
	RELB；MAP3K7；MAP2K2；IL8；JAK2；CHUK；STAT3；
	MAP2K1；NFKB1；CEBPB；JUN；IL1R1；SRF；IL6
肝胆汁淤积	PRKCE；IRAK1；INS；MYD88；PRKCZ；TRAF6；PPARA；
	RXRA；IKBKB；PRKCI；NFKB2；MAP3K14；MAPK8；
	PRKD1；MAPK10；RELA；PRKCD；MAPK9；ABCB1；
	TRAF2；TLR4；TNF；INSR；IKBKG；RELB；MAP3K7；IL8；
	CHUK；NR1H2；TJP2；NFKB1；ESR1；SREBF1；FGFR4；
	JUN；IL1R1；PRKCA；IL6
IGF-1 信号传导	IGF1；PRKCZ；ELK1；MAPK1；PTPN11；NEDD4；AKT2；
	PIK3CA；PRKCI；PTK2；FOS；PIK3CB；PIK3C3；MAPK8；
	IGF1R；IRS1；MAPK3；IGFBP7；KRAS；PIK3C2A；
	YWHAZ；PXN；RAF1；CASP9；MAP2K2；AKT1；PIK3R1；
	PDPK1；MAP2K1；IGFBP2；SFN；JUN；CYR61；AKT3；
	FOXO1；SRF；CTGF；RPS6KB1
NRF2 介导的氧化	PRKCE；EP300；SOD2；PRKCZ；MAPK1；SQSTM1；
应激反应	NQO1；PIK3CA；PRKCI；FOS；PIK3CB；PIK3C3；MAPK8；
	PRKD1；MAPK3；KRAS；PRKCD；GSTP1；MAPK9；FTL；
	NFE2L2；PIK3C2A；MAPK14；RAF1；MAP3K7；CREBBP；

	MAP2K2; AKT1; PIK3R1; MAP2K1; PPIB; JUN; KEAP1;
	GSK3B; ATF4; PRKCA; EIF2AK3; HSP90AA1
肝纤维化/肝	EDN1; IGF1; KDR; FLT1; SMAD2; FGFR1; MET; PGF;
星形细胞激活	SMAD3; EGFR; FAS; CSF1; NFKB2; BCL2; MYH9;
	IGF1R; IL6R; RELA; TLR4; PDGFRB; TNF; RELB; IL8;
	PDGFRA; NFKB1; TGFBR1; SMAD4; VEGFA; BAX;
	IL1R1; CCL2; HGF; MMP1; STAT1; IL6; CTGF; MMP9
PPAR 信号传导	EP300; INS; TRAF6; PPARA; RXRA; MAPK1; IKBKB;
	NCOR2; FOS; NFKB2; MAP3K14; STAT5B; MAPK3;
	NRIP1; KRAS; PPARG; RELA; STAT5A; TRAF2;
	PPARGC1A; PDGFRB; TNF; INSR; RAF1; IKBKG;
	RELB; MAP3K7; CREBBP; MAP2K2; CHUK; PDGFRA;
	MAP2K1; NFKB1; JUN; IL1R1; HSP90AA1
FcεRI信号传导	PRKCE; RAC1; PRKCZ; LYN; MAPK1; RAC2; PTPN11;
	AKT2; PIK3CA; SYK; PRKCI; PIK3CB; PIK3C3; MAPK8;
	PRKD1; MAPK3; MAPK10; KRAS; MAPK13; PRKCD;
	MAPK9; PIK3C2A; BTK; MAPK14; TNF; RAF1; FYN;
	MAP2K2; AKT1; PIK3R1; PDPK1; MAP2K1; AKT3;
	VAV3; PRKCA
G-蛋白偶联	PRKCE; RAP1A; RGS16; MAPK1; GNAS; AKT2; IKBKB;
受体信号传导	PIK3CA; CREB1; GNAQ; NFKB2; CAMK2A; PIK3CB;
	PIK3C3; MAPK3; KRAS; RELA; SRC; PIK3C2A; RAF1;
	IKBKG; RELB; FYN; MAP2K2; AKT1; PIK3R1; CHUK;
	PDPK1; STAT3; MAP2K1; NFKB1; BRAF; ATF4; AKT3;
	PRKCA
磷酸肌醇	PRKCE; IRAK1; PRKAA2; EIF2AK2; PTEN; GRK6;
代谢	MAPK1; PLK1; AKT2; PIK3CA; CDK8; PIK3CB; PIK3C3;
	MAPK8; MAPK3; PRKCD; PRKAA1; MAPK9; CDK2;
	PIM1; PIK3C2A; DYRK1A; MAP2K2; PIP5K1A; PIK3R1;
	MAP2K1; PAK3; ATM; TTK; CSNK1A1; BRAF; SGK

PDGF 信号传导	EIF2AK2；ELK1；ABL2；MAPK1；PIK3CA；FOS；PIK3CB；
	PIK3C3；MAPK8；CAV1；ABL1；MAPK3；KRAS；SRC；
	PIK3C2A；PDGFRB；RAF1；MAP2K2；JAK1；JAK2；
	PIK3R1；PDGFRA；STAT3；SPHK1；MAP2K1；MYC；
	JUN；CRKL；PRKCA；SRF；STAT1；SPHK2
VEGF 信号传导	ACTN4；ROCK1；KDR；FLT1；ROCK2；MAPK1；PGF；
	AKT2；PIK3CA；ARNT；PTK2；BCL2；PIK3CB；PIK3C3；
	BCL2L1；MAPK3；KRAS；HIF1A；NOS3；PIK3C2A；PXN；
	RAF1；MAP2K2；ELAVL1；AKT1；PIK3R1；MAP2K1；SFN；
	VEGFA；AKT3；FOXO1；PRKCA
自然杀伤细胞信号传导	PRKCE；RAC1；PRKCZ；MAPK1；RAC2；PTPN11；
	KIR2DL3；AKT2；PIK3CA；SYK；PRKCI；PIK3CB；
	PIK3C3；PRKD1；MAPK3；KRAS；PRKCD；PTPN6；
	PIK3C2A；LCK；RAF1；FYN；MAP2K2；PAK4；AKT1；
	PIK3R1；MAP2K1；PAK3；AKT3；VAV3；PRKCA
细胞周期：G1/S	HDAC4；SMAD3；SUV39H1；HDAC5；CDKN1B；BTRC；
检查点调节	ATR；ABL1；E2F1；HDAC2；HDAC7A；RB1；HDAC11；
	HDAC9；CDK2；E2F2；HDAC3；TP53；CDKN1A；CCND1；
	E2F4；ATM；RBL2；SMAD4；CDKN2A；MYC；NRG1；
	GSK3B；RBL1；HDAC6
T 细胞受体信号传导	RAC1；ELK1；MAPK1；IKBKB；CBL；PIK3CA；FOS；
	NFKB2；PIK3CB；PIK3C3；MAPK8；MAPK3；KRAS；
	RELA；PIK3C2A；BTK；LCK；RAF1；IKBKG；RELB；FYN；
	MAP2K2；PIK3R1；CHUK；MAP2K1；NFKB1；ITK；BCL10；
	JUN；VAV3
死亡受体信号传导	CRADD；HSPB1；BID；BIRC4；TBK1；IKBKB；FADD；
	FAS；NFKB2；BCL2；MAP3K14；MAPK8；RIPK1；CASP8；
	DAXX；TNFRSF10B；RELA；TRAF2；TNF；IKBKG；RELB；
	CASP9；CHUK；APAF1；NFKB1；CASP2；BIRC2；CASP3；
	BIRC3

FGF 信号传导	RAC1; FGFR1; MET; MAPKAPK2; MAPK1; PTPN11;
	AKT2; PIK3CA; CREB1; PIK3CB; PIK3C3; MAPK8;
	MAPK3; MAPK13; PTPN6; PIK3C2A; MAPK14; RAF1;
	AKT1; PIK3R1; STAT3; MAP2K1; FGFR4; CRKL; ATF4;
	AKT3; PRKCA; HGF
GM-CSF 信号传导	LYN; ELK1; MAPK1; PTPN11; AKT2; PIK3CA; CAMK2A;
	STAT5B; PIK3CB; PIK3C3; GNB2L1; BCL2L1; MAPK3;
	ETS1; KRAS; RUNX1; PIM1; PIK3C2A; RAF1; MAP2K2;
	AKT1; JAK2; PIK3R1; STAT3; MAP2K1; CCND1; AKT3;
	STAT1
肌萎缩性脊髓侧索	BID; IGF1; RAC1; BIRC4; PGF; CAPNS1; CAPN2;
硬化症信号传导	PIK3CA; BCL2; PIK3CB; PIK3C3; BCL2L1; CAPN1;
	PIK3C2A; TP53; CASP9; PIK3R1; RAB5A; CASP1;
	APAF1; VEGFA; BIRC2; BAX; AKT3; CASP3; BIRC3
JAK/Stat 信号传导	PTPN1; MAPK1; PTPN11; AKT2; PIK3CA; STAT5B;
	PIK3CB; PIK3C3; MAPK3; KRAS; SOCS1; STAT5A;
	PTPN6; PIK3C2A; RAF1; CDKN1A; MAP2K2; JAK1;
	AKT1; JAK2; PIK3R1; STAT3; MAP2K1; FRAP1; AKT3;
	STAT1
烟酸和烟酰胺	PRKCE; IRAK1; PRKAA2; EIF2AK2; GRK6; MAPK1;
代谢	PLK1; AKT2; CDK8; MAPK8; MAPK3; PRKCD; PRKAA1;
	PBEF1; MAPK9; CDK2; PIM1; DYRK1A; MAP2K2;
	MAP2K1; PAK3; NT5E; TTK; CSNK1A1; BRAF; SGK
趋化因子信号传导	CXCR4; ROCK2; MAPK1; PTK2; FOS; CFL1; GNAQ;
	CAMK2A; CXCL12; MAPK8; MAPK3; KRAS; MAPK13;
	RHOA; CCR3; SRC; PPP1CC; MAPK14; NOX1; RAF1;
	MAP2K2; MAP2K1; JUN; CCL2; PRKCA
IL-2 信号传导	ELK1; MAPK1; PTPN11; AKT2; PIK3CA; SYK; FOS;

	STAT5B; PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS;
	SOCS1; STAT5A; PIK3C2A; LCK; RAF1; MAP2K2;
	JAK1; AKT1; PIK3R1; MAP2K1; JUN; AKT3
突触长时程	PRKCE; IGF1; PRKCZ; PRDX6; LYN; MAPK1; GNAS;
抑制	PRKCI; GNAQ; PPP2R1A; IGF1R; PRKD1; MAPK3;
	KRAS; GRN; PRKCD; NOS3; NOS2A; PPP2CA;
	YWHAZ; RAF1; MAP2K2; PPP2R5C; MAP2K1; PRKCA
雌激素受体	TAF4B; EP300; CARM1; PCAF; MAPK1; NCOR2;
信号传导	SMARCA4; MAPK3; NRIP1; KRAS; SRC; NR3C1;
	HDAC3; PPARGC1A; RBM9; NCOA3; RAF1; CREBBP;
	MAP2K2; NCOA2; MAP2K1; PRKDC; ESR1; ESR2
蛋白质泛素化	TRAF6; SMURF1; BIRC4; BRCA1; UCHL1; NEDD4;
途径	CBL; UBE2I; BTRC; HSPA5; USP7; USP10; FBXW7;
	USP9X; STUB1; USP22; B2M; BIRC2; PARK2; USP8;
	USP1; VHL; HSP90AA1; BIRC3
IL-10 信号传导	TRAF6; CCR1; ELK1; IKBKB; SP1; FOS; NFKB2;
	MAP3K14; MAPK8; MAPK13; RELA; MAPK14; TNF;
	IKBKG; RELB; MAP3K7; JAK1; CHUK; STAT3; NFKB1;
	JUN; IL1R1; IL6
VDR/RXR 激活	PRKCE; EP300; PRKCZ; RXRA; GADD45A; HES1;
	NCOR2; SP1; PRKCI; CDKN1B; PRKD1; PRKCD;
	RUNX2; KLF4; YY1; NCOA3; CDKN1A; NCOA2; SPP1;
	LRP5; CEBPB; FOXO1; PRKCA
TGF-β 信号传导	EP300; SMAD2; SMURF1; MAPK1; SMAD3; SMAD1;
	FOS; MAPK8; MAPK3; KRAS; MAPK9; RUNX2;
	SERPINE1; RAF1; MAP3K7; CREBBP; MAP2K2;
	MAP2K1; TGFBR1; SMAD4; JUN; SMAD5
Toll 样受体信号传导	IRAK1; EIF2AK2; MYD88; TRAF6; PPARA; ELK1;
	IKBKB; FOS; NFKB2; MAP3K14; MAPK8; MAPK13;
	RELA; TLR4; MAPK14; IKBKG; RELB; MAP3K7; CHUK;

	NFKB1；TLR2；JUN
p38 MAPK信号传导	HSPB1；IRAK1；TRAF6；MAPKAPK2；ELK1；FADD；FAS；
	CREB1；DDIT3；RPS6KA4；DAXX；MAPK13；TRAF2；
	MAPK14；TNF；MAP3K7；TGFBR1；MYC；ATF4；IL1R1；
	SRF；STAT1
神经营养蛋白/TRK信号传导	NTRK2；MAPK1；PTPN11；PIK3CA；CREB1；FOS；
	PIK3CB；PIK3C3；MAPK8；MAPK3；KRAS；PIK3C2A；
	RAF1；MAP2K2；AKT1；PIK3R1；PDPK1；MAP2K1；
	CDC42；JUN；ATF4
FXR/RXR激活	INS；PPARA；FASN；RXRA；AKT2；SDC1；MAPK8；
	APOB；MAPK10；PPARG；MTTP；MAPK9；PPARGC1A；
	TNF；CREBBP；AKT1；SREBF1；FGFR4；AKT3；FOXO1
突触长时程	PRKCE；RAP1A；EP300；PRKCZ；MAPK1；CREB1；
增强	PRKCI；GNAQ；CAMK2A；PRKD1；MAPK3；KRAS；
	PRKCD；PPP1CC；RAF1；CREBBP；MAP2K2；MAP2K1；
	ATF4；PRKCA
钙信号传导	RAP1A；EP300；HDAC4；MAPK1；HDAC5；CREB1；
	CAMK2A；MYH9；MAPK3；HDAC2；HDAC7A；HDAC11；
	HDAC9；HDAC3；CREBBP；CALR；CAMKK2；ATF4；
	HDAC6
EGF信号传导	ELK1；MAPK1；EGFR；PIK3CA；FOS；PIK3CB；PIK3C3；
	MAPK8；MAPK3；PIK3C2A；RAF1；JAK1；PIK3R1；
	STAT3；MAP2K1；JUN；PRKCA；SRF；STAT1
心血管系统	EDN1；PTEN；EP300；NQO1；UBE2I；CREB1；ARNT；
中的缺氧信号传导	HIF1A；SLC2A4；NOS3；TP53；LDHA；AKT1；ATM；
	VEGFA；JUN；ATF4；VHL；HSP90AA1
LPS/IL-1介导的	IRAK1；MYD88；TRAF6；PPARA；RXRA；ABCA1；
RXR功能的抑制	MAPK8；ALDH1A1；GSTP1；MAPK9；ABCB1；TRAF2；

	TLR4; TNF; MAP3K7; NR1H2; SREBF1; JUN; IL1R1
LXR/RXR 激活	FASN; RXRA; NCOR2; ABCA1; NFKB2; IRF3; RELA;
	NOS2A; TLR4; TNF; RELB; LDLR; NR1H2; NFKB1;
	SREBF1; IL1R1; CCL2; IL6; MMP9
淀粉样蛋白处理	PRKCE; CSNK1E; MAPK1; CAPNS1; AKT2; CAPN2;
	CAPN1; MAPK3; MAPK13; MAPT; MAPK14; AKT1;
	PSEN1; CSNK1A1; GSK3B; AKT3; APP
IL-4 信号传导	AKT2; PIK3CA; PIK3CB; PIK3C3; IRS1; KRAS; SOCS1;
	PTPN6; NR3C1; PIK3C2A; JAK1; AKT1; JAK2; PIK3R1;
	FRAP1; AKT3; RPS6KB1
细胞周期: G2/M DNA	EP300; PCAF; BRCA1; GADD45A; PLK1; BTRC;
损伤检查点	CHEK1; ATR; CHEK2; YWHAZ; TP53; CDKN1A;
调节	PRKDC; ATM; SFN; CDKN2A
心血管系统	KDR; FLT1; PGF; AKT2; PIK3CA; PIK3CB; PIK3C3;
中的一氧化氮信号传导	CAV1; PRKCD; NOS3; PIK3C2A; AKT1; PIK3R1;
	VEGFA; AKT3; HSP90AA1
嘌呤代谢	NME2; SMARCA4; MYH9; RRM2; ADAR; EIF2AK4;
	PKM2; ENTPD1; RAD51; RRM2B; TJP2; RAD51C;
	NT5E; POLD1; NME1
cAMP 介导的信号传导	RAP1A; MAPK1; GNAS; CREB1; CAMK2A; MAPK3;
	SRC; RAF1; MAP2K2; STAT3; MAP2K1; BRAF; ATF4
线粒体功能异常	SOD2; MAPK8; CASP8; MAPK10; MAPK9; CASP9;
	PARK7; PSEN1; PARK2; APP; CASP3
Notch 信号传导	HES1; JAG1; NUMB; NOTCH4; ADAM17; NOTCH2;
	PSEN1; NOTCH3; NOTCH1; DLL4
内质网	HSPA5; MAPK8; XBP1; TRAF2; ATF6; CASP9; ATF4;
应激途径	EIF2AK3; CASP3
嘧啶代谢	NME2; AICDA; RRM2; EIF2AK4; ENTPD1; RRM2B;

	NT5E；POLD1；NME1
帕金森病信号传导	UCHL1；MAPK8；MAPK13；MAPK14；CASP9；PARK7；
	PARK2；CASP3
心脏&β肾上腺素能信号传导	GNAS；GNAQ；PPP2R1A；GNB2L1；PPP2CA；PPP1CC；PPP2R5C
糖酵解/糖异生	HK2；GCK；GPI；ALDH1A1；PKM2；LDHA；HK1
干扰素信号传导	IRF1；SOCS1；JAK1；JAK2；IFITM1；STAT1；IFIT3
音猬因子信号传导	ARRB2；SMO；GLI2；DYRK1A；GLI1；GSK3B；DYRK1B
甘油磷脂代谢	PLD1；GRN；GPAM；YWHAZ；SPHK1；SPHK2
磷脂降解	PRDX6；PLD1；GRN；YWHAZ；SPHK1；SPHK2
色氨酸代谢	SIAH2；PRMT5；NEDD4；ALDH1A1；CYP1B1；SIAH1
赖氨酸降解	SUV39H1；EHMT2；NSD1；SETD7；PPP2R5C
核苷酸切除修复途径	ERCC5；ERCC4；XPA；XPC；ERCC1
淀粉和蔗糖代谢	UCHL1；HK2；GCK；GPI；HK1
氨基糖代谢	NQO1；HK2；GCK；HK1
花生四烯酸代谢	PRDX6；GRN；YWHAZ；CYP1B1
昼夜节律信号传导	CSNK1E；CREB1；ATF4；NR1D1
凝血系统	BDKRB1；F2R；SERPINE1；F3
多巴胺受体信号传导	PPP2R1A；PPP2CA；PPP1CC；PPP2R5C
谷胱甘肽代谢	IDH2；GSTP1；ANPEP；IDH1
甘油脂代谢	ALDH1A1；GPAM；SPHK1；SPHK2
亚油酸代谢	PRDX6；GRN；YWHAZ；CYP1B1

甲硫氨酸代谢	DNMT1；DNMT3B；AHCY；DNMT3A
丙酮酸代谢	GLO1；ALDH1A1；PKM2；LDHA
精氨酸和脯氨酸	ALDH1A1；NOS3；NOS2A
代谢	
类花生酸信号传导	PRDX6；GRN；YWHAZ
果糖和甘露糖	HK2；GCK；HK1
代谢	
半乳糖代谢	HK2；GCK；HK1
芪、香豆素和	PRDX6；PRDX1；TYR
木质素生物合成	
抗原呈递	CALR；B2M
途径	
类固醇的生物合成	NQO1；DHCR7
丁酸代谢	ALDH1A1；NLGN1
柠檬酸循环	IDH2；IDH1
脂肪酸代谢	ALDH1A1；CYP1B1
甘油磷脂	PRDX6；CHKA
代谢	
组氨酸代谢	PRMT5；ALDH1A1
肌醇代谢	ERO1L；APEX1
细胞色素p450	GSTP1；CYP1B1
对外源物的代谢	
甲烷代谢	PRDX6；PRDX1
苯丙氨酸代谢	PRDX6；PRDX1
丙酸代谢	ALDH1A1；LDHA
含硒氨基酸	PRMT5；AHCY

代谢	
鞘脂代谢	SPHK1；SPHK2
氨基膦酸	PRMT5
代谢	
雄激素和雌激素	PRMT5
代谢	
抗坏血酸和Aldarate	ALDH1A1
代谢	
胆汁酸生物合成	ALDH1A1
半胱氨酸代谢	LDHA
脂肪酸生物合成	FASN
谷氨酸受体	GNB2L1
信号传导	
NRF2 介导的氧化	PRDX1
应激反应	
磷酸戊糖	GPI
途径	
戊糖和葡糖醛酸	UCHL1
相互转化	
视黄醇代谢	ALDH1A1
核黄素代谢	TYR
酪氨酸代谢	PRMT5，TYR
泛醌生物合成	PRMT5
缬氨酸、亮氨酸和	ALDH1A1
异亮氨酸降解	
甘氨酸、丝	CHKA

氨酸和	
苏氨酸代谢	
赖氨酸降解	ALDH1A1
疼痛/味觉	TRPM5; TRPA1
疼痛	TRPM7; TRPC5; TRPC6; TRPC1; Cnr1; cnr2; Grk2;
	Trpa1; Pomc; Cgrp; Crf; Pka; Era; Nr2b; TRPM5; Prkaca;
	Prkacb; Prkar1a; Prkar2a
线粒体功能	AIF; CytC; SMAC (Diablo) ; Aifm-1; Aifm-2
发育神经学	BMP-4; 脊索素 (Chordin) (Chrd) ; 头发生素 (Nog) ; WNT (Wnt2;
	Wnt2b; Wnt3a; Wnt4; Wnt5a; Wnt6; Wnt7b; Wnt8b;
	Wnt9a; Wnt9b; Wnt10a; Wnt10b; Wnt16) ; β-连环蛋白;
	Dkk-1; 卷曲相关蛋白; Otx-2; Gbx2; FGF-8;
	颤蛋白 (Reelin) ; Dab1; unc-86 (Pou4f1 或 Brn3a) ; Numb; Reln

[0161] 本发明的实施例还涉及与敲除基因、扩增基因以及修复与DNA重 复不稳定性和神经障碍相关的具体突变有关的方法和组合物(罗伯特D.•威尔 斯(Robert D.Wells)、芦沢哲夫(Tetsuo Ashizawa),遗传不稳定性与神经 疾病(Genetic Instabilities and Neurological Diseases),第二版,学术出版社 (Academic Press),2011年10月13日-《医学》(Medical))。已经发现 串联重复序列的特定方面对超过二十种人类疾病负责(重复不稳定新见:RNA•DNA杂交体的作用(New insights into repeat instability:role of RNA•DNA hybrids),麦基弗EI(McIvor EI)、波拉克U(Polak U)、纳皮 尔拉拉M(Napierala M),《RNA生物学》(RNA Biol.)2010年9月-10 月;7(5):551-8)。可以利用CRISPR-Cas系统修正基因组不稳定性缺陷。

[0162] 本发明的一个另外的方面涉及利用CRISPR-Cas系统修正EMP2A 和EMP2B基因缺陷,这些基因已经被鉴定为与拉福拉病(Lafora disease)相 关。拉福拉病是一种由可以作为青年期的癫痫发作而开始的进行性肌阵挛性 癫痫表征的常染色体隐性病症。该疾病的少数病可以由尚未鉴定的基因的突 变引起。该疾病引起发作、肌肉痉挛、行走困难、痴呆,并且最终引起死 亡。当前没有疗法被证明有效对抗疾病进展。与癫痫相关的其他遗传异常还 可以靶向CRISPR-Cas系统并且基础遗传学进一步描述于由朱利亚诺阿文济 尼(Giuliano Avanzini)、杰弗里L.诺贝尔斯(Jeffrey L.Noebels)编辑的《癫痫与遗传癫痫遗传学》(Genetics of Epilepsy and Genetic Epilepsies),马 里亚尼儿科神经学基金会(Mariani Foundation Paediatric Neurology):20;2009)中。

[0163] 在本发明的又另一个方面,该CRISPR-Cas系统可以用来矫正几种 基因突变引起的眼部缺陷,其进一步描述于《眼的遗传疾病》(Genetic Diseases of the Eye),第二次编辑,由伊莱亚斯(Elias)I.特拉布勒西 (Traboulsi)编辑,牛津大学出版社,2012年。

[0164] 本发明的若干另外的方面涉及修正与范围广泛的遗传性疾病相关的 缺陷,这些遗传性疾病在专题小节遗传性障碍(Genetic Disorders)下被进一 步描述于国立卫生研

究院的网站(网址为 health.nih.gov/topic/GeneticDisorders)。遗传性脑病可以包括但不限于,肾上 腺脑白质营养不良、胼胝体发育不全、艾卡尔迪综合征(Aicardi Syndrome)、阿尔佩斯病(Alpers'Disease)、阿尔茨海默病、巴特综合征 (Barth Syndrome)、巴藤病(Batten Disease)、CADASIL、小脑变性、费 波瑞病(Fabry's Disease)、格斯特曼-施特劳斯纳病(Gerstmann-Straussler- Scheinker Disease)、亨廷顿病以及其他三联体重复障碍、莱氏病(Leigh's Disease)、莱施-奈恩综合征、门克斯病、线粒体肌病以及NINDS空洞脑(Colpocephaly)。这些疾病在小节遗传性脑部障碍(Genetic Brain Disorders)下被进一步描述于国立卫生研究院的网站。

[0165] 在一些实施例中,该病症可以是瘤形成。在一些实施例中,在该病 症是瘤形成的情况下,有待被靶向的基因是列于表A中的那些基因中的任一 种(在这种情况下是PTEN等)。在一些实施例中,该病症可以是年龄相关 性黄斑变性。在一些实施例中,该病症可以是一种精神分裂症障碍。在一些 实施例中,该病症可以是一种三核苷酸重复障碍。在一些实施例中,该病症 可以是脆性X综合征。在一些实施例中,该病症可以是一种分泌酶相关障碍。在一些实施例中,该病症可以是一种朊病毒相关障碍。在一些实施例 中,该病症可以是ALS。在一些实施例中,该病症可以是一种药物成瘾。在 一些实施例中,该病症可以是自闭症。在一些实施例中,该病症可以是阿尔 茨海默病。在一些实施例中,该病症可以是炎症。在一些实施例中,该病症 可以是帕金森病。

[0166] 与帕金森病相关的蛋白质的实例包括但不限于α-突触核蛋白、DJ- 1、LRRK2、PINK1、帕金蛋白、UCHL1、Synphilin-1以及NURR1。

[0167] 成瘾相关的蛋白质的实例可以包括例如ABAT。

[0168] 炎症相关蛋白的实例可以包括例如由Ccr2基因编码的单核细胞趋 化蛋白-1(MCP1)、由Ccr5基因编码的C-C趋化因子受体类型5 (CCR5)、由Fcgr2b基因编码的IgG受体IIB(FCGR2b,亦称CD32)或由 Fcer1g基因编码的FcεR1g(FCER1g)蛋白。

[0169] 心血管疾病相关蛋白的实例可以包括例如IL1B(白介素1,β)、XDH(黄嘌呤脱氢酶)、TP53(肿瘤蛋白p53)、PTGIS(前列腺素I2(前 列腺环素)合酶)、MB(肌红蛋白)、IL4(白介素4)、ANGPT1(血管生 成素1)、ABCG8(ATP-结合盒,亚家族G(WHITE),成员8)或CTSK (组织蛋白酶K)。

[0170] 阿尔茨海默病相关蛋白的实例可以包括例如由VLDLR基因编码的 极低密度脂蛋白受体蛋白(VLDLR)、由UBA1基因编码的泛素样改性剂激 活酶1(UBA1)或由UBA3基因编码的NEDD8-激活酶E1催化亚基蛋白 (UBE1C)。

[0171] 自闭症谱系障碍相关蛋白的实例可以包括例如由BZRAP1基因编码 的苯二氮卓受体(外周)相关蛋白1(BZRAP1)、由AFF2基因编码的 AF4/FMR2家族成员2蛋白(AFF2)(亦称MFR2)、由FXR1基因编码的 脆性X智力迟钝常染色体同系物1蛋白(FXR1)或由FXR2基因编码的脆性 X智力迟钝常染色体同系物2蛋白(FXR2)。

[0172] 黄斑变性相关蛋白的实例可以包括例如由ABCR基因编码的ATP- 结合盒,亚家族A(ABC1)成员4蛋白(ABCA4)、由APOE基因编码的 载脂蛋白E蛋白(APOE)或由CCL2基因编码的趋化因子(C-C基序)配体 2蛋白(CCL2)。

[0173] 精神分裂症相关蛋白的实例可以包括NRG1、ErbB4、CPLX1、 TPH1、TPH2、NRXN1、GSK3A、BDNF、DISC1、GSK3B及其组合。

[0174] 涉及肿瘤抑制的蛋白质的实例可以包括例如ATM(共济失调性毛 细血管扩张突变的)、ATR(共济失调性毛细血管扩张和Rad3相关的)、EGFR(表皮生长因子受体)、ERBB2(v-erb-b2红白血病病毒癌基因同系物 2)、ERBB3(v-erb-b2红白血病病毒癌基因同系物3)、ERBB4(v-erb-b2红 白血病病毒癌基因同系物4)、Notch 1、Notch 2、Notch 3或Notch 4。

[0175] 与分泌酶障碍相关的蛋白质的实例可以包括例如PSENEN(早老素 增强子2同系物(秀丽隐杆线虫))、CTSB(组织蛋白酶B)、PSEN1(早 老素1)、APP(淀粉样蛋白β(A4)前体蛋白)、APH1B(前咽缺陷1同系 物B(秀丽隐杆线虫))、PSEN2(早老素2(阿尔茨海默病4))或BACE1 (β-位点APP-切割酶1)。

[0176] 与肌萎缩性侧索硬化相关的蛋白质的实例可以包括SOD1(超氧化 物歧化酶1)、ALS2(肌萎缩性侧索硬化2)、FUS(融合在肉瘤中)、TARDBP(TAR DNA结合蛋白)、VAGFA(血管内皮生长因子A)、VAGFB(血管内皮生长因子B)以及VAGFC(血管内皮生长因子C)及其 任何组合。

[0177] 与朊病毒病相关的蛋白质的实例可以包括SOD1(超氧化物歧化酶 1)、ALS2(肌萎缩性侧索硬化2)、FUS(融合在肉瘤中)、TARDBP (TAR DNA结合蛋白)、VAGFA(血管内皮生长因子A)、VAGFB(血管 内皮生长因子B)以及VAGFC(血管内皮生长因子C)及其任何组合。

[0178] 与朊病毒病症中的神经退行性疾病相关的蛋白质的实例可以包括例 如A2M(α-2-巨球蛋白)、AATF(凋亡拮抗转录因子)、ACPP(前列腺的 酸性磷酸酶)、ACTA2(肌动蛋白α2平滑肌主动脉)、ADAM22(ADAM 金属肽酶结构域)、ADORA3(腺苷A3受体)或ADRA1D(α-1D肾上腺素 能受体(Alpha-1D adrenergic receptor或Alpha-1D adrenoreceptor))。

[0179] 与免疫缺陷相关的蛋白质的实例可以包括例如A2M[α-2-巨球蛋 白];AANAT[芳烷基胺N-乙酰转移酶];ABCA1[ATP-结合盒,亚家族A (ABC1),成员1];ABCA2[ATP-结合盒,亚家族A(ABC1),成员2]; 或ABCA3[ATP-结合盒,亚家族A(ABC1),成员3]。

[0180] 与三核苷酸重复障碍相关的蛋白质的实例包括例如AR(雄激素受 体)、FMR1(脆性X智力迟钝1)、HTT(亨廷顿蛋白)或DMPK(强直性 肌营养不良蛋白激酶)、FXN(弗氏共济失调蛋白(frataxin))、ATXN2 (共济失调蛋白2)。

[0181] 与神经传递障碍相关的蛋白质的实例包括例如SST(生长抑素)、NOS1(一氧化氮合酶1(神经元的))、ADRA2A(肾上腺素能的,α-2A-, 受体)、ADRA2C(肾上腺素能的,α-2C-,受体)、TACR1(速激肽受体 1)或HTR2c(5-羟色胺(血清素)受体2C)。

[0182] 神经发育相关序列的实例包括例如A2BP1[共济失调蛋白2-结合蛋 白1]、AADAT[氨基己二酸氨基转移酶],AANAT[芳烷基胺N-乙酰转移 酶]、ABAT[4-氨基丁酸氨基转移酶]、ABCA1[ATP-结合盒,亚家族A (ABC1),成员1]或ABCA13[ATP-结合盒,亚家族A(ABC1),成员 13]。

[0183] 可用本发明系统治疗的优选病症的另外的实例可以选自:艾卡迪- 古铁雷斯综合征(Aicardi-Goutières Syndrome);亚历山大病;阿伦-赫恩登- 达德利综合征(Allan-Herndon-Dudley Syndrome);POLG相关障碍;α-甘露 糖苷贮积症(II和III型);阿尔斯特伦综合征(Alström Syndrome);安格 曼;综合征;共济失调-毛细血管扩张;神经元蜡样-脂褐质沉积;β-地中海贫 血;双侧视神经萎缩和1型(婴儿)视神经萎缩;视网膜母细胞瘤(双侧的);卡纳万病(Canavan Disease);脑-眼-面-骨骼综合征 (Cerebrooculofacioskeletal Syndrome)1[COFS1];脑腱黄瘤病;科尼利亚迪 兰吉综合征(Cornelia de Lange

Syndrome);MAPT相关障碍;遗传性朊病毒 病;德拉韦综合征(Dravet Syndrome);早期发病家族性阿尔茨海默病;弗 里德赖希共济失调[FRDA];多发性畸形;岩藻糖苷贮积症;福山(Fukuyama)先天性肌营养不良;半乳糖唾液酸贮积症;戈谢病(Gaucher Disease);有机酸血症;噬血细胞淋巴组织细胞增生症;早衰症;粘脂贮积 症II;婴儿游离唾液酸贮积病;PLA2G6相关神经变性;耶韦尔和朗格-尼尔 森综合征(Jervell and Lange-Nielsen Syndrome);结合性大疱性表皮松解; 亨廷顿病;克拉伯病(Krabbe Disease)(婴儿的);线粒体DNA相关雷吉 综合征(Mitochondrial DNA-Associated Leigh Syndrome)和NARP;莱施-奈 恩综合征;LIS1相关无脑回;洛氏综合征(Lowe Syndrome);槭糖尿病; MECP2复制综合征;ATP7A相关铜转运障碍;LAMA2相关肌营养不良;芳 基硫酸酯酶A缺乏;I、II或III型粘多糖贮积症;过氧化物酶体生物发生障 碍,齐薇格谱系综合征(Zellweger Syndrome Spectrum);伴随脑铁累积障碍 的神经变性;酸性鞘磷脂酶缺乏;C型尼曼-皮克病;甘氨酸脑病;ARX相关 障碍;尿素循环障碍;COL1A1/2相关成骨不全;线粒体DNA缺失综合征; PLP1相关障碍;佩里综合征(Perry Syndrome);费伦-麦克德米德综合征 (Phelan-McDermid Syndrome);II型糖原贮积症(蓬佩病(Pompe Disease))(婴儿的);MAPT相关障碍;MECP2相关障碍;1型肢近端型 点状软骨发育不全(Rhizomelic Chondrodysplasia Punctata Type 1);罗伯茨 综合征(Roberts Syndrome);桑德霍夫病(Sandhoff Disease);辛德勒病(Schindler Disease)-1型;腺苷脱氨酶缺乏;史-伦-奥三氏综合征;脊髓性 肌萎缩;婴儿发病脊髓小脑性共济失调;己糖胺酶A缺乏;1型致死性发育 不良;胶原VI型相关障碍;I型乌谢尔综合征(Usher Syndrome);先天性肌 营养不良;沃尔夫-赫奇霍恩综合征(Wolf-Hirschhorn Syndrome);溶酶体酸 脂酶缺乏;以及着色性干皮病。

[0184] 如将是显而易见的,设想的是可以使用本发明系统靶向任何感兴趣 的多核苷酸序列。使用本发明系统进行治疗可能是有用的病症或疾病的一些 实例被包括在上表中并且当前与那些病症相关的基因的实例也被提供于此。然而,示例的基因并不是穷尽性的。

实例

[0185] 以下实例出于说明本发明的不同实施例的目的而给出并且并不意在 以任何方式限制本发明。本发明实例连同在此描述的方法目前代表优选的实 施例,是示例性的,并且并不旨在限制本发明的范围。涵盖在如由权利要求 书的范围所定义的本发明的精神内的在此的变化以及其他用途是本领域普通 技术人员可以想到的。

实例1:真核细胞的细胞核中的CRISPR复合物活性

[0186] 一个示例性II型CRISPR系统是来自化脓链球菌SF370的II型 CRISPR座位,该座位包含四个基因Cas9、Cas1、Cas2和Csn1的聚簇以及两 个非编码RNA元件tracrRNA和由非重复序列的短段(间隔子,每个约30 bp)间隔开的重复序列(同向重复)的特征性阵列。在此系统中,以四个连 续步骤产生靶向的DNA双链断裂(DSB)(图2A)。第一步,从CRISPR 座位转录两个非编码RNA、前-crRNA阵列和tracrRNA。第二步,将 tracrRNA杂交到前-crRNA的同向重复上,然后将其加工成含有单独的间隔子 序列的成熟crRNA。第三步,该成熟crRNA:tracrRNA复合物经由在crRNA 的间隔子区与原型间隔子DNA之间形成异源双链体而将Cas9指向由原型间 隔子和对应的PAM组成的DNA靶标。最后,Cas9介导PAM上游的靶DNA 的切割,以在原型间隔子内产生一个DSB(图2A)。此实例描述了一种用于 将此RNA可编程的核酸酶系统适配成指导真核细胞的细胞核中的CRISPR复 合物活性的示例性过程。

[0187] 细胞培养和转染

[0188] 在37℃下,伴随5%CO_2孵育,将人类胚肾(HEK)细胞系HEK 293FT(生命技术公司)维持在补充有10%胎牛血清(海可隆公司 (HyClone))、2mM GlutaMAX(生命技术公司)、100U/mL青霉素及 100μg/mL链霉素的杜尔贝科改良伊格尔培养基(Dulbecco's modified Eagle's Medium)(DMEM)中。在37℃下,在5%CO_2下,用补充有5%胎牛血清 (海可隆公司)、2mM GlutaMAX(生命技术公司)、100U/mL青霉素及 100μg/mL链霉素的DMEM维持小鼠neuro2A(N2A)细胞系(ATCC)。

[0189] 在转染前一天,将HEK 293FT或N2A细胞以200,000个细胞/孔的 密度接种到24孔板(康宁公司(Corning))中。遵循制造商推荐的方案,使 用Lipofectamine 2000(生命技术公司)转染细胞。对于24孔板的每个孔,使 用总共800ng的质粒。

[0190] 基因组修饰的Surveyor测定和测序分析

[0191] 如上所述,用质粒DNA转染HEK 293FT或N2A细胞。转染后, 在基因组DNA提取之前,将细胞在37℃下孵育72小时。遵循制造商的方 案,使用QuickExtract DNA提取试剂盒(Epicentre)提取基因组DNA。简言 之,将细胞重悬于QuickExtract溶液中并在65℃下孵育15分钟并且在98℃ 下孵育10分钟。将提取的基因组DNA立即加工或存储在-20℃下。

[0192] 对于每个基因而言,PCR扩增围绕CRISPR靶位点的基因组区域, 并且遵循制造商的方案,使用QiaQuick Spin柱(凯杰公司(Qiagen))纯化 产物。将总共400ng的纯化的PCR产物与2μl 10X Taq聚合酶PCR缓冲液 (Enzymatics公司)和超纯水混合,至终体积为20μl,并使其经受重退火过 程,以使得可以形成异源双链体:95℃持续10min,以-2℃/s 95℃降至 85℃,以-0.25℃/s 85℃降至25℃,并且在25℃保持1分钟。重退火后,遵 循制造商推荐的方案,将产物用Surveyor核酸酶和Surveyor增强子S(转基 因组学公司(Transgenomics))处理,并且在4%-20%Novex TBE聚丙烯酰 胺凝胶(生命技术公司)上进行分析。将凝胶用SYBR Gold DNA染色剂(生 命技术公司)染色30分钟并用Gel Doc凝胶成像系统(伯乐公司(Bio- rad))成像。定量是基于作为切割的DNA的部分的量度的相对条带强度。图8提供了此Surveyor测定的一个示意图。

[0193] 用于检测同源重组的限制性片段长度多态性测定

[0194] 将HEK 293FT和N2A细胞用质粒DNA转染,并且如上所述,在 基因组DNA提取之前,将其在37℃下孵育72小时。使用引物在同源重组 (HR)模板的同源臂外PCR扩增靶基因组区域。将PCR产物分离在1%琼脂 糖凝胶上并用MinElute GelExtraction试剂盒(凯杰公司)提取。将纯化产物 用HindIII(费尔芒斯公司(Fermentas))消化并在6%Novex TBE聚丙烯酰 胺凝胶(生命技术公司)上进行分析。

[0195] RNA二级结构预测和分析

[0196] 使用质心结构预测算法,使用由维也纳大(University of Vienna) 的理论化学研究所(Institute for Theoretical Chemistry)研发的在线网络服务 器RNAfold进行RNA二级结构预测(参见例如A.R.格鲁伯(A.R.Gruber) 等人,2008,《细胞》(Cell)106(1):23-24;以及PA卡尔(PA Carr)和 GM丘奇(GM Church),2009,《自然生物技术》(Nature Biotechnology) 27(12):1151-62)。

[0197] 细菌性质粒转化干扰测定

[0198] 使用pCRISPR质粒在大肠杆菌中重组足以用于CRISPR活性的化 脓链球菌CRISPR

座位1的元件(示意性地说明于图10A中)。pCRISPR包 含tracrRNA、SpCas9、以及驱动crRNA阵列的一个前导子序列。使用退火的 寡核苷酸,间隔子(也称为“指导序列”)被插入至该crRNA阵列中BsaI位 点之间,如所示的。在干扰测定中使用的激发质粒是通过如下构建的,将原 型间隔子序列(也称为“靶序列”)连同一个邻近的CRISPR基序序列 (PAM)插入至pUC19中(参见图10B)。该激发质粒包含氨苄青霉素抗 性。图10C提供该干扰测定的示意性表示。将已携带pCRISPR以及适合的间 隔子的化学感受态的大肠杆菌菌株用包含相应的原型间隔子-PAM序列的激发 质粒进行转化。pUC19被用以评估每个携带pCRISPR的感受态菌株的转化 率。CRISPR活性导致携带该原型间隔子pPSP质粒的切割(排除另外由缺少 该原型间隔子的pUC19赋予的氨苄青霉素抗性)。图10D示出在图4C中所 示的测定中使用的每个携带pCRISPR的大肠杆菌菌株的感受态。

[0199] RNA纯化

[0200] 将HEK 293FT细胞如上所述地维持和转染。通过胰酶消化收获细 胞,随后在磷酸盐缓冲生理盐水(PBS)中洗涤。遵循制造商的方案,用TRI 试剂(西格玛公司)提取总细胞RNA。使用Naonodrop(赛默飞世尔科技公 司(Thermo Scientific))对提取的总RNA进行定量并标准化至同一浓度。

[0201] 哺乳动物细胞中的crRNA和tracrRNA表达的Northern印迹分析

[0202] 将RNA与等体积的2X加样缓冲液(Ambion)混合,加热至95℃ 持续5min,在冰上冷冻1min,并且然后在将凝胶预跑胶至少30分钟后,加 样到8%变性聚丙烯酰胺凝胶(SequaGel,国立诊断公司(National Diagnostics))上。将样品以40W极限电泳1.5小时。之后,在室温下,将 RNA转移到半干转移设备(伯乐公司(Bio-rad))中的300mA的Hybond N+膜(通用电气医疗公司(GE Healthcare))上,持续1.5小时。使用 Stratagene UV Crosslinker the Stratalinker(Stratagene)上的自动交联按钮将 RNA交联到该膜上。在42℃下,伴随旋转,将该膜在ULTRAhyb-Oligo杂交 缓冲液(Ambion)中预杂交30min,并且然后添加探针并杂交过夜。探针订 购自IDT并将其用具有T4多核苷酸激酶(新英格兰生物实验室(New England Biolabs)))的[γ-^{32}P]ATP(珀金埃尔默公司)标记。将该膜用预加温 的(42℃)2xSSC、0.5%SDS洗涤一次,持续1min,随后在42℃下洗涤两 次,每次持续30分钟。将该膜在室温下向荧光屏暴露一小时或过夜并且然后 用感光成像仪(Typhoon)扫描。

[0203] 细菌CRISPR系统构建和评价

[0204] 从具有用于吉布森拼接(Gibson Assembly)的侧翼同源臂的化脓链 球菌SF370基因组DNA PCR扩增CRISPR座位元件(包括tracrRNA、Cas9 和前导子)。将两个BsaI类型IIS位点引入在两个同向重复之间,以促进间 隔子的方便插入(图9)。使用Gibson Assembly Master Mix(NEB)将PCR 产物克隆进EcoRV-消化的pACYC184中,在tet启动子的下游。省略其他内 源CRISPR系统元件,Csn2的最后50bp除外。将编码具有互补突出端的间 隔子的寡聚物(整合DNA技术公司(Integrated DNA Technology))克隆进 BsaI-消化的载体pDC000(NEB)中并且然后用T7连接酶(Enzymatics公 司)连接,以产生pCRISPR质粒。通过将携带兼容的突出端的杂交的寡核苷 酸(整合DNA技术公司)连接至BamHI消化的pUC19中来产生包含间隔子 (携带有PAM序列(此处也称为“CRISPR基序序列”))的激发质粒。所 有构建体的克隆都是在大肠杆菌菌株JM109(Zymo研究公司)中进行。

[0205] 根据制造商说明,使用Z-感受态大肠杆菌转化试剂盒以及缓冲液 套组(Zymo研究

公司,T3001)使携带pCRISPR的细胞变为感受态。在该转 化测定中,将50uL等分试样的、携带pCRISPR的感受态细胞在冰上解冻并 且用1ng的间隔子质粒或pUC19在冰上转化30分钟,随后在42℃下热激45 秒并且在冰上热激2分钟。接着,添加250ul SOC(英杰公司(Invitrogen)),随后在37℃摇动孵育1小时,并且将100uL的SOC后生 成物铺板于双筛选板(12.5ug/ml氯霉素,100ug/ml氨苄西林)上。为了获 得cfu/ng的DNA,总菌落数目乘以3。

[0206] 为了改进CRISPR组分在哺乳动物细胞中的表达,来自化脓链球菌(Streptococcus pyogenes/S.pyogenes)的SF370基因座1的两个基因是密码子 优化的Cas9(SpCas9)以及RNA酶III(SpRNA酶III)。为了促进核定位, 在SpCas9以及SpRNA酶III两者的氨基(N)-或羧基(C)-末端包括一个核 定位信号(NLS)(图2B)。为了促进蛋白表达的可视化,在两种蛋白的N- 或C-末端也包括一种荧光蛋白标记物(图2B)。也产生了具有附接至N-以 及C-末端(2xNLS-SpCas9)两者的NLS的一个版本的SpCas9。将包含NLS 融合的SpCas9以及SpRNA酶III的构建体转染至293FT人胚肾(HEK)细胞 中,并且发现该NLS与SpCas9以及SpRNA酶III的相对定位影响其核定位 效率。而该C-末端NLS足以将SpRNA酶III靶向该细胞核,在此系统中,单 拷贝的这些具体的NLS与SpCas9的N-或C-末端的附接不能够实现充分的核 定位。在此实例中,该C-末端NLS是核质蛋白 (KRPAATKKAGQAKKKK)的C-末端NLS,并且该C-末端NLS是SV40 大型T-抗原(PKKKRKV)的C-末端NLS。在测试的该版本的SpCas9中, 仅2xNLS-SpCas9展现核定位(图2B)。

[0207] 来自化脓链球菌SF370的CRISPR座位的tracrRNA具有两个转录 起始位点,导致两种转录物,89-核苷酸(nt)以及171nt,这两种转录物随后 被加工成相同的75nt成熟tracrRNA。选择较短的89nt tracrRNA用于在哺乳 动物细胞中表达(图7A中所示出的表达构建体,其中功能性是如显示于图 7B中的Surveyor测定的结果所确定的)。将转录起始位点标为+1,并且还 指示了转录终止子和通过Northern印迹探测的序列。还通过Northern印迹确 认了加工的tracrRNA的表达。图7C显示了提取自293FT细胞的总RNA的 Northern印迹分析的结果,用携带长或短tracrRNA以及SpCas9和DR- EMX1(1)-DR的U6表达构建体转染这些细胞。左图和右图分别是来自用或没 用SpRNA酶III转染的293FT细胞。U6指示用靶向人类U6 snRNA的探针进 行印迹的加样对照。短tracrRNA表达构建体的转染产生丰富水平的加工形式 的tracrRNA(约75bp)。在Northern印迹上检测极低量的长tracrRNA。

[0208] 为了促进转录精确起始,选择RNA聚合酶III基的U6启动子,以 驱动tracrRNA的表达(图2C)。类似地,研发U6启动子基的构建体,以表 达包含侧翼为两个同向重复(DR,还被术语“tracr配对序列”所涵盖;图 2C)的单个间隔子的前-crRNA阵列。将起始间隔子设计成靶向人EMX1座位 (图2C)中的33-碱基对(bp)靶位点(满足Cas9的NGG识别基序的30-bp 原型间隔子加3-bp CRISPR基序(PAM)序列),该座位是大脑皮层的发育 中的一个关键基因。

[0209] 为了测试CRISPR系统(SpCas9、SpRNA酶III、tracrRNA以及前- crRNA)在哺乳动物细胞中的异源表达是否可以实现哺乳动物染色体的靶向 切割,用CRISPR组分的组合转染HEK 293FT细胞。由于哺乳动物细胞核中 的DSB通过导致形成indel的非同源末端连接(NHEJ)途径而部分地修复, 使用Surveyor测定检测靶EMX1座位处的潜在的切割活性(图8)(参见例 如格斯钦(Guschin)等人,2010,《分子生物学方法》(Methods Mol Biol) 649:247)。所有四种CRISPR组分的共转染能够在原型间隔子中诱导高达 5.0%切割(参见图2D)。所有CRISPR组分(减去SpRNA酶III)的共转染 在原型间隔子中也诱导高达4.7%

indel,表明可能存在能够辅助crRNA成熟 的内源性哺乳动物RNA酶,如例如相关的切丁酶(Dicer)和Drosha酶。除 去剩余的三种组分中的任何组分消除CRISPR系统的基因组切割活性(图 2D)。包含已证实切割活性的靶座位的扩增子的桑格测序:在43个测序的克 隆中,发现5个突变的等位基因(11.6%)。使用多种指导序列的类似实验产 生了高达29%的indel百分比(参见图4-7、12和13)。这些结果定义了一种 在哺乳动物细胞中的有效的CRISPR介导的基因组修饰的三组分系统。为了 优化切割效率,申请人还测试了tracrRNA的不同亚型是否影响切割效率并且 发现,在实例系统中,仅仅短(89-bp)转录本形式能够介导人类EMX1基因 组座位的切割(图7B)。

[0210] 图14提供了哺乳动物细胞中的crRNA加工的另外的Northern印迹 分析。图14A示出了一个示意图,显示了侧翼为两个同向重复(DR- EMX1(1)-DR)的单个间隔子的表达载体。靶向人EMX1座位原型间隔子1 (参见图6)的30bp间隔子和同向重复序列显示于图14A下方的序列中。直 线指示其反向互补系列用于产生供EMX1(1)crRNA检测用的Northern印迹探 针的区域。图14B显示了提取自293FT细胞的总RNA的Northern印迹分 析,用携带DR-EMX1(1)-DR的U6表达构建体转染这些细胞。左图和右图分 别是来自使用或未使用SpRNA酶III转染的293FT细胞。DR-EMX1(1)-DR仅 在存在SpCas9和短tracrRNA的情况下被加工为成熟crRNA并且不依赖于 SpRNA酶III的存在。检测的来自转染的293FT总RNA的成熟crRNA是约 33bp并且比来自化脓链球菌的39-42bp成熟crRNA短。这些结果证明, CRISPR系统可以被移植进真核细胞中并被重新编程,以促进内源哺乳动物靶 多核苷酸的切割。

[0211] 图2示出了描述于此实例中的细菌CRISPR系统。图2A示出了一 个示意图,显示了来自化脓链球菌SF370的CRISPR座位和通过此系统 CRISPR介导的DNA切割的建议机制。加工自同向重复-间隔子阵列的成熟 crRNA指导Cas9到以下基因组靶标,这些靶标由互补原型间隔子和原型间隔 子相邻基序(PAM)组成。靶标-间隔子碱基配对后,Cas9介导靶DNA中的 双链断裂。图2B示出了具有使得导入到哺乳动物细胞核中成为可能的核定位 信号(NLS)的化脓链球菌Cas9(SpCas9)和RNA酶III(SpRNA酶III)的 工程化。图2C示出了由组成型EF1a启动子驱动的SpCas9和SpRNA酶III以 及由RNA Pol3启动子U6驱动的tracrRNA和前-crRNA阵列(DR-间隔子- DR)的哺乳动物表达,上述启动子促进转录精确起始和终止。将来自具有令 人满意的PAM序列的人EMX1座位的原型间隔子用作前-crRNA阵列中的间 隔子。图2D示出了SpCas9介导的次要插入和缺失的Surveyor核酸酶测定。 使用和不使用SpRNA酶III、tracrRNA以及携带EMX1-靶间隔子的前-crRNA 阵列表达SpCas9。图2E示出了靶座位与EMX1-靶向crRNA之间的碱基配对 的一个示意表示以及一个显示了与SpCas9切割位点邻近的微小缺失的示例性 色谱图。图2F示出了从43个克隆扩增子的测序分析鉴定出的突变的等位基 因,这些扩增子显示出多种微小插入和缺失。破折号指示缺失的碱基,并且 非比对或错配碱基指示插入或突变。比例尺 =10μm。

[0212] 为了进一步简化该三组分系统,改编嵌合的crRNA-tracrRNA杂交 体设计,其中成熟crRNA(包括一个指导序列)经由茎-环而融合到部分 tracrRNA上,以模拟天然的crRNA:tracrRNA双链体(图3A)。为了增加共 递送效率,产生双顺反子表达载体,以驱动嵌合RNA和SpCas9在转染细胞 中的共表达(图3A和8)。平行地,使用该双顺反子载体表达前-crRNA(DR-指导序列-DR)与SpCas9,以诱导用分开表达的tracrRNA加工为 crRNA(比较图13B的顶图和底图)。图9提供了前-crRNA阵列(图9A) 或嵌合的crRNA(由图9B中的指导序列插入位

点的下游和EF1α启动子的上 游的短线表示)与hSpCas9的双顺反子表达载体的示意图,显示了各种元件 的位置和指导序列插入点。图9B中的围绕指导序列插入位点的位置的展开的 序列还显示了部分DR序列(GTTTAGAGCTA)和部分tracrRNA序列(TAGCAAGTTAAAATAAGGCTAGTCCGTTTTT)。可以使用退火的寡核 苷酸将指导序列插入在BbsI位点之间。寡核苷酸的序列设计显示在图9中的 示意图的下方,其中指示了适当的连接适配子。WPRE表示土拨鼠肝炎病毒 转录后调节元件。通过靶向与上述相同的EMX1座位而测试嵌合RNA介导的 切割效率。使用扩增子的Surveyor测定和桑格测序两者,申请人确认嵌合RNA设计以大约4.7%的修饰率促进人EMX1座位的切割(图4)。

[0213] 通过设计靶向人EMX1和PVALB以及小鼠Th座位中的多个位点的 嵌合RNA,通过靶向人类和小鼠细胞两者中的另外的基因组座位而测试 CRISPR介导的切割在真核细胞中的普遍性。图15示出了人类PVALB(图15A)和小鼠Th(图15B)座位中的一些另外的靶向原型间隔子的筛选。提 供了各自的最后的外显子内的基因座位位和三个原型间隔子的位置的示意 图。加下划线的序列包括30bp的原型间隔子序列和对应于PAM序列的3'端 处的3bp。分别在DNA序列的上方和下方指示了正义链和反义链上的原型间 隔子。人类PVALB和小鼠Th座位分别达到了6.3%和0.75%的修饰率,证明 了CRISPR系统跨多种生物在修饰不同座位方面的广阔适用性(图3B和 6)。虽然对于每个座位而言,使用嵌合的构建体仅检测到三分之一的间隔子 的切割,但是当使用共表达的前-crRNA安排时,所有靶序列都被切割,伴随达到27%的有效的indel产生(图6)。

[0214] 图13提供了可以被重新编程为靶向哺乳动物细胞中的多个基因组 座位的SpCas9的另外的图解。图13A提供了人EMX1座位的示意图,显示 了五个原型间隔子的位置,由加下划线的序列指示。图13B提供了前- crRNA/trcrRNA复合物的示意图,显示了前-crRNA和tracrRNA的同向重复区 之间的杂交(顶部),以及嵌合RNA设计的示意图,包括一个20bp指导序 列和由部分同向重复组成的tracr配对和tracr序列以及以发夹结构杂交的tracrRNA序列(底部)。比较人EMX1座位中的五个原型间隔子处的Cas9介 导的切割的效率的Surveyor测定的结果示于图13C中。使用加工的前- crRNA/tracrRNA复合物(crRNA)或嵌合的RNA(chiRNA)靶向每个原型 间隔子。

[0215] 由于RNA的二级结构对于分子间相互作用可以是决定性的,使用 基于最小自由能和玻尔兹曼(Boltzmann)加权结构集合的结构预测算法比较 基因组靶向实验(图3B)中使用的所有指导序列的假定的二级结构(参见例 如格鲁贝尔(Gruber)等人,2008,《核酸研究》(Nucleic Acids Research),36:W70)。分析揭示到在大多数情况下,在嵌合的crRNA背景 下有效的指导序列基本上不含二级结构基序,而无效的指导序列更可能形成 可以阻止与靶原型间隔子DNA进行碱基配对的内部二级结构。因此,可能的 是,当使用嵌合的crRNA时,间隔子二级结构的变异性可能影响CRISPR介 导的干扰的效率。

[0216] 图3示例了实例表达载体。图3A提供了用于驱动一种合成crRNA- tracrRNA嵌合体(嵌合RNA)连同SpCas9的表达的一种双顺反子载体的示 意图。该嵌合指导RNA包含相应于基因组的靶位点中的原型间隔子的一个 20-bp指导序列。图3B提供显示靶向人EMX1、PVALB、以及小鼠Th基因座 的指导序列,连同其预测的二级结构的示意图。表明在每个靶位点的修饰效 率低于该RNA二级结构图(EMX1,n=216个扩增子测序读数;PVALB,n= 224个读数;Th,n=265个读数)。折叠算法产生一个输出,其中每个碱基 根据其假定该预测二级

结构的可能性着色,如通过在图3B中以灰色标度复现 的彩虹标度(rainbow scale)表明的。SpCas9的另外的载体设计显示于图44 中,示出了掺入连接到指导寡核苷酸的插入位点的U6启动子和连接到 SpCas9编码序列上的Cbh启动子的单个表达载体。显示于图44b中的载体包 括一个连接到H1启动子上的tracrRNA编码序列。

[0217] 为了测试包含二级结构的间隔子是否能够在CRISPR天然操作的原 核细胞中起作用,在异源表达化脓链球菌SF370 CRISPR基因座1的大肠杆菌 菌株中测试具有原型间隔子的质粒的转化干扰(图10)。将该CRISPR座位 克隆至一种低拷贝大肠杆菌表达载体中,并且该crRNA阵列被侧翼为一对 DR(pCRISPR)的一个单间隔子替代。将包含不同pCRISPR质粒的大肠杆菌 菌株用包含相应的原型间隔子以及PAM序列的激发质粒进行转化(图10C)。在细菌测定中,所有的间隔子均促进有效的CRISPR干扰(图 4C)。这些结果表明,可能存在影响哺乳动物细胞中的CRISPR活性效率的 另外的因素。

[0218] 为了研究CRISPR介导的切割的特异性,使用一系列具有单点突变 的EMX1-靶向嵌合的crRNA分析指导序列中的单核苷酸突变对哺乳动物基因 组中的原型间隔子切割的影响(图4A)。图4B示出了比较当与不同突变的 嵌合的RNA配对时Cas9的切割效率的Surveyor核酸酶测定的结果。PAM的 5’的高达12-bp的单碱基错配基本上消除了由SpCas9进行的基因组切割,而 在离上游位置更远处具有突变的间隔子保留针对原始原型间隔子靶标的活性 (图4B)。除PAM之外,SpCas9在间隔子的最后12-bp内具有单碱基特异 性。此外,CRISPR能够像一对靶向同一EMX1原型间隔子的TALE核酸酶 (TALEN)一样有效地介导基因组切割。图4C提供了一个示意图,显示了 靶向EMX1的TALEN的设计并且图4D显示了比较TALEN和Cas9的效率的 Surveyor凝胶(n=3)。

[0219] 已经通过易错NHEJ机制建立了一套用于在哺乳动物细胞中实现 CRISPR介导的基因编辑的组分,测试CRISPR刺激同源重组(HR)的能 力,同源重组是一种用于使得基因组中的编辑精确的高保真基因修复途径。野生型SpCas9能够介导位点特异的DSB,这些DSB可以通过NHEJ和HR 两者进行修复。另外,将SpCas9的RuvC I催化结构域中的天冬氨酸至丙氨酸取代(D10A)工程化,以将核酸酶转化为切口酶(SpCas9n;示于图5A 中)(参见例如萨普拉诺萨克斯(Sapranauskas)等人,2011,《核酸研究》(Nucleic Acids Research),39:9275;加索纳斯(Gasiunas)等人,2012,《美国国家科学院院刊》(Proc.Natl.Acad.Sci.USA),109:E2579),这样使 得带切口的基因组DNA经历高保真同源定向修复(HDR)。Surveyor测定确 认到,SpCas9n在EMX1原型间隔子靶标处不产生indel。如图5B所示, EMX1-靶向嵌合的crRNA与SpCas9的共表达在靶位点中产生indel,而与 SpCas9n的共表达不产生indel(n=3)。此外,327个扩增子的测序未检测到 由SpCas9n诱导的任何indel。选择相同的座位,以通过用靶向EMX1、hSpCas9或hSpCas9n的嵌合RNA以及用于在原型间隔子附近引入一对限制 酶切位点(HindIII和NheI)的HR模板共转染HEK 293FT细胞而测试CRISPR介导的HR。图5C提供了HR策略的示意图,包含重组点的相对位置 和引物退火序列(箭头)。SpCas9和SpCas9n的确催化将HR模板整合进 EMX1座位中。PCR扩增靶区域,随后用HindIII进行限制酶切消化揭示了对 应于期望片段尺寸的切割产物(显示于图5D中的限制性片段长度多态性分析 中的箭头),其中SpCas9和SpCas9n介导类似水平的HR效率。申请人使用 基因组扩增子的桑格测序进一步证实了HR(图5E)。这些结果证明了 CRISPR促进哺乳动物基因组中的靶向基因插入的实用性。鉴于14-bp(来自 间隔子的12-bp和来自PAM的2-bp)特异地靶向

野生型SpCas9,切口酶的可 获得性可以显著减少脱靶修饰的可能性,因为单链断裂不是易错NHEJ途径 的底物。

[0220] 构建模拟具有阵列的间隔子的CRISPR座位的天然构造的表达构建 体(图2A),以测试多元序列靶向的可能性。使用编码一对EMX1-和 PVALB-靶向间隔子的单CRISPR阵列,在两个座位处都检测到了有效的切割 (图5F,显示了crRNA阵列的示意设计和显示出切割的有效修饰的Surveyor 印迹两者)。使用针对由119bp间隔开的EMX1内的两个靶标的间隔子通过 共存DSB还检测到了较大基因组区的靶向缺失,并且检测到了1.6%缺失效率 (182个扩增子中的3个;图5G)。这证明,CRISPR系统可以介导单个基因 组中的多元编辑。

实例2:CRISPR系统修饰和替代方案

[0221] 使用RNA编程序列特异的DNA切割的能力定义了一个新类别的针 对多种研究和工业应用的基因组工程化工具。可以进一步改善CRISPR系统 的若干方面,以增加CRISPR靶向的效率和通用性。优化的Cas9活性可以依 赖于处于比存在于哺乳动物细胞核中的游离Mg^{2+}高的水平的游离Mg^{2+}的可获 得性(参见例如季聂克(Jinek)等人,2012,《科学》(Science),337: 816),并且对恰好位于原型间隔子的下游的NGG基序的偏好限制靶向人类 基因组中的平均每12-bp的能力(图11,评价了人类染色体序列的正链和负 链两者)。这些约束中的一些可以通过探索CRISPR座位跨微生物宏基因组 的多样性而克服(参见例如马卡洛娃(Makarova)等人,2011,《自然微生 物学综述》(Nat Rev Microbiol),9:467)。可以通过类似于实例1中描述 的方法将其他CRISPR座位移植进哺乳动物细胞环境中。例如,图12示出了 来自嗜热链球菌LMD-9的CRISPR 1的用于在哺乳动物细胞中进行异源表达 的II型CRISPR系统的适应,以实现CRISPR介导的基因组编辑。图12A提 供了来自嗜热链球菌LMD-9的CRISPR 1的示意图。图12B示出了嗜热链球 菌CRISPR系统的表达系统的设计。使用组成型EF1α启动子表达人类密码子 优化的hStCas9。使用促进转录精确起始的U6启动子表达tracrRNA和crRNA 的成熟版本。示出了来自成熟crRNA和tracrRNA的序列。crRNA序列中的由小写字母“a”指示的单个碱基用于除去作为RNA polIII转录终止子的 polyU序列。图12C提供显示靶向人EMX1座位的指导序列,连同其预测的 二级结构的示意图。表明在每个靶位点的修饰效率低于这些RNA二级结构。 产生这些结构的算法将每个碱基根据其假定该预测二级结构的可能性着色, 如通过在图12C中以灰色标度复现的彩虹标度表明的。显示了使用Surveyor 测定的靶座位中的hStCas9介导的切割的结果。RNA指导间隔子1和2分别 诱导了14%和6.4%。在这两个原型间隔子位点处跨生物复制物的切割活性的 统计分析还提供于图6中的。图16提供了嗜热链球菌CRISPR系统的另外的 原型间隔子和在人EMX1座位中的对应的PAM序列靶标示意图。两个原型间 隔子序列被凸显并且通过相对于对应的凸显序列为3’加下划线而指示满足 NNAGAAW基序的对应的PAM序列。两个原型间隔子均靶向反义链。

实例3:样品靶序列选择算法

[0222] 设计一个软件程序,以基于指定的CRISPR酶的希望的指导序列长 度和CRISPR基序序列(PAM),在输入DNA序列的两条链上鉴定候选 CRISPR靶序列。例如,可以通过在输入序列和该输入序列的反向互补体两者 上检索5’-N_x-NGG-3’而鉴定来自化脓链球菌的Cas9的具有PAM序列NGG 的靶位点。同样地,可以通过在输入序列和该输入序列的反向互补体两者上 检索5’-Nx-NNAGAAW-3’而鉴定化脓链球菌CRISPR1的Cas9的具有PAM 序列NNAGAAW的靶位点。同样地,可以通过在输入序列和该输入序列的反 向互补体两者上检索5’-N_x-

NGGNG-3’而鉴定化脓链球菌CRISPR3的Cas9的 具有PAM序列NGGNG的靶位点。可以通过程序固定或由使用者指定N_x中 的值“x”,如20。

[0223] 由于在DNA靶位点的基因组中出现多次可以导致非特异的基因组 编辑,因此在鉴定所有潜在的位点后,基于序列在相关参比基因组中出现的 次数,该程序将其滤出。对于其序列特异性是通过‘种子(seed)’序列(如 来自PAM序列的11-12bp 5’,包括PAM序列本身)确定的那些CRISPR 酶,过滤步骤可以基于该种子序列。因此,为了避免另外的基因组座位处的 编辑,基于种子:PAM序列在相关基因组中出现的次数过滤结果。可以允许 使用者选择种子序列的长度。处于通过过滤器的目的,还可以允许使用者指 定种子:PAM序列在基因组中出现的次数。默认筛选独特序列。通过改变种 子序列的长度和该序列在基因组中出现的次数两者而改变过滤水平。该程序 可以另外或可替代地通过提供鉴定的一个或多个靶序列的反向互补体而提供 与报告的一个或多个靶序列互补的指导序列的序列。

[0224] 用于优化序列筛选的方法和算法的另外的细节可以发现于通过引用 结合在此的美国申请序列号61/836,080(代理人案号44790.11.2022)中。

实例4:多种嵌合的crRNA-tracrRNA杂交体的评价

[0225] 此实例描述了针对具有掺入不同长度的野生型tracrRNA序列的 tracr序列的嵌合的RNA(chiRNAs;在单个转录本中包括一个指导序列、一 个tracr配对序列以及一个tracr序列)而获得的结果。图18a示出了嵌合 RNA和Cas9的双顺反子表达载体的示意图。Cas9由CBh启动子驱动并且嵌 合的RNA由U6启动子驱动。嵌合的指导RNA由接合到tracr序列的20bp指 导序列(N)组成(从下链的第一个“U”到该转录本的结尾),其在如指示 的各点被截短。指导和tracr序列由tracr-配对序列GUUUUAGAGCUA,随后 是环序列GAAA隔开。人EMX1和PVALB座位处的Cas9介导的indel的 SURVEYOR测定的结果分别示于图18b和18c中。箭头指示预期的 SURVEYOR片段。ChiRNA由其“+n”标志指示,并且crRNA是指指导序列 和tracr序列被表达为分开的转录本的杂交体RNA。一式三份地进行的这些结 果的定量示于图19a和19b的直方图中,分别对应于图18b和18c(“N.D.” 指示未检测到indel)。原型间隔子ID及其对应的基因组靶标、原型间隔子、PAM序列以及链位置提供于表D中。将指导序列设计成与整个原型间隔子序 列互补(在杂交体系统中的分开的转录本的情况下),或仅为部分加下划线 (在嵌合的RNA的情况下)。

表D:

原型间隔子ID	基因组靶标	原型间隔子序列(5’至3’)	PAM	链
1	*EMX1*	GGACATCGATGTCACCTCCAATGACTAGGG	TGG	+
2	*EMX1*	CATTGGAGGTGACATCGATGTCCTCCCCAT	TGG	-
3	*EMX1*	GGAAGGGCCTGAGTCCGAGCAGAAGAAGAA	GGG	+
4	*PVALB*	GGTGGCGAGAGGGGCCGAGATTGGGTGTTC	AGG	+
5	*PVALB*	ATGCAGGAGGGTGGCGAGAGGGGCCGAGAT	TGG	+

[0226] 细胞培养和转染

[0227] 在37℃下,伴随5%CO_2孵育,将人类胚肾(HEK)细胞系293FT (生命技术公司)维持在补充有10%胎牛血清(海可隆公司(HyClone))、2mM GlutaMAX(生命技术公司)、100U/mL青霉素及100μg/mL链霉素的 杜尔贝科改良伊格尔培养基(Dulbecco’s modified Eagle’s

Medium) (DMEM)中。在转染前24小时,将293FT细胞以150,000个细胞/孔的密度 接种到24孔板(康宁公司(Corning))上。遵循制造商推荐的方案,使用 Lipofectamine 2000(生命技术公司)转染细胞。对于24孔板的每个孔,使用 总共500ng的质粒。

[0228] 基因组修饰的SURVEYOR测定

[0229] 如上所述,用质粒DNA转染293FT细胞。转染后,在基因组DNA 提取之前,将细胞在37℃下孵育72小时。遵循制造商的方案,使用 QuickExtract DNA提取溶液(Epicentre)提取基因组DNA。简言之,将丸状 的细胞重悬于QuickExtract溶液中并在65℃下孵育15分钟并且在98℃下孵 育10分钟。对于每个基因而言,PCR扩增(引物列于表E中)在CRISPR靶 位点侧翼的基因组区域,并且遵循制造商的方案,使用QiaQuick Spin柱(凯 杰公司(Qiagen))纯化产物。将总共400ng的纯化的PCR产物与2μl 10X Taq DNA聚合酶PCR缓冲液(Enzymatics公司)和超纯水混合,至终体积为 20μl,并使其经受重退火过程,以使得可以形成异源双链体:95℃持续10 min,以-2℃/s从95℃降至85℃,以-0.25℃/s从85℃降至25℃,并且在 25℃保持1分钟。重退火后,遵循制造商推荐的方案,将产物用 SURVEYOR核酸酶和SURVEYOR增强子S(转基因组学公司 (Transgenomics))处理,并且在4%-20%Novex TBE聚丙烯酰胺凝胶(生 命技术公司)上进行分析。将凝胶用SYBR Gold DNA染色剂(生命技术公 司)染色30分钟并用Gel Doc凝胶成像系统(伯乐公司(Bio-rad))成像。基于相对条带强度进行量化。

表E:

引物名称	基因组靶标	引物序列(5’至3’)
Sp-EMX1-F	EMX1	AAAACCACCCTTCTCTCTGGC
Sp-EMX1-R	EMX1	GGAGATTGGAGACACGGAGAG
Sp-PVALB-F	PVALB	CTGGAAAGCCAATGCCTGAC
Sp-PVALB-R	PVALB	GGCAGCAAACTCCTTGTCCT

[0230] 独特的CRISPR靶位点的计算机鉴定

[0231] 为了鉴定针对人、小鼠、大鼠、斑马鱼、果蝇、以及秀丽隐杆线虫 基因组中的化脓链球菌SF370 Cas9(SpCas9)酶的独特的靶位点,我们开发 了一个软件包以扫描一个DNA序列的两条链并且鉴定所有可能的SpCas9靶 位点。对于此实例,每个SpCas9靶位点操作上定义为一个20bp序列,随后 定义一个NGG原型间隔子相邻基序(PAM)序列,并且我们鉴定了所有染色 体上满足此5’-N_{20}-NGG-3’定义的所有序列。为了预防非特异性基因组编辑, 在鉴定所有潜在的位点后,基于序列在相关参比基因组中出现的次数,过滤 所有靶位点。为了利用一个‘种子’序列(其可以是,例如,从该PAM序列 5’开始大约11-12bp序列)赋予的Cas9活性的序列特异性,5’- NNNNNNNNNN-NGG-3’序列被选择为在该相关基因组中是独特的。从UCSC 基因组浏览器(人类基因组hg19,小鼠基因组mm9,大鼠基因组rn5,斑马 鱼基因组danRer7,果蝇基因组dm4以及秀丽隐杆线虫基因组ce10)下载所 有基因组序列。使用UCSC基因组浏览器信息能够获得全面检索结果。一些 靶位点在人类基因组中的示例性可视化提供于图21中。

[0232] 初始,靶向人HEK 293FT细胞中的EMX1座位内的三个位点。使 用SURVEYOR核酸酶测定评估每种chiRNA的基因组修饰效率,该测定检测 由DNA双链断裂(DSB)及其通过非同源末端连接(NHEJ)DNA损伤修复 途径进行的随后修复造成的突变。Co被指定为chiRNA(+n)

的构建体指示，野 生型tracrRNA的直达第+n个核苷酸被包括在嵌合的RNA构建体中，其中48、54、67和85的值用于n。嵌合的RNA在所有三个EMX1靶位点处都包 含长片段的野生型tracrRNA(chiRNA(+67)和chiRNA(+85))介导的DNA切 割，其中chiRNA(+85)尤其展示出比以分开的转录本形式表达指导序列和 tracr序列的对应的crRNA/tracrRNA杂交体显著更高水平的DNA切割(图 18b和19a)。还使用chiRNA靶向PVALB座位中的两个位点，其不产生使用杂交体系统(作为分开的转录本而表达的指导序列和tracr序列)的可检测 的切割。chiRNA(+67)和chiRNA(+85)能够在两个PVALB原型间隔子处介导 显著的切割(图18c和19b)。

[0233] 对于EMX1和PVALB座位中的所有五个靶标，伴随增加的tracr 序列长度观察到一致增加的基因组修饰效率。不希望受任何理论的束缚，由 tracrRNA的3’端形成的二级结构可以在增强CRISPR复合物形成率方面发挥 一定作用。图21中提供了在此实例中使用的嵌合RNA中每个的预测二级结 构的图解。使用最小自由能以及配分函数算法，使用RNA折叠(http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi)预测该二级结构。每个碱基的 假色(以灰度复现)表明成对的可能性。因为具有更长的tracr序列的 chiRNA能够切割未被天然CRISPR crRNA/tracrRNA杂交体切割的靶标，可能 的是，嵌合RNA可比其天然杂交体配对物更有效地装载到Cas9上。为了促 进Cas9在真核细胞以及生物体中应用于位点特异性基因组编辑，在人、小 鼠、大鼠、斑马鱼、秀丽隐杆线虫、以及果蝇基因组中计算机鉴定该化脓链 球菌Cas9的所有预测的独特的靶位点。嵌合RNA可以针对来自其他微生物 的Cas9酶进行设计以扩展CRISPR RNA可编程的核酸酶的靶空间。

[0234] 图22示出了用于表达包括野生型tracr RNA序列的高达+85核苷酸 以及带有核定位序列的SpCas9的嵌合RNA的示例性双顺反子表达载体的示 意图。SpCas9是从一种CBh启动子表达的并被bGH polyA信号(bGH pA) 终止。紧接着该示意图所示的扩展的序列对应于该指导序列插入位点周围的 区域，并且从5’至3’包括U6启动子(第一遮蔽区)的3’-部分、BbsI切割位 点(箭头)、部分同向重复(tracr配对序列GTTTTAGAGCTA，加下划 线)、环序列GAAA、以及+85tracr序列(环序列后加下划线的序列)。在 该指导序列插入位点之后示出了一个示例性的指导序列插入片段，针对所选 择的靶标，指导序列的核苷酸由“N”表示。

[0235] 以上实例中所述的序列如下(多核苷酸序列是5’至3’)：

[0236] U6-短tracrRNA(化脓链球菌SF370)：

[0237] GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACA AGGCTGTTAGAGAGATAATTGGAATTAATTTGACTGTAAACACAAAGAT ATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTT GCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAAC TTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAA ACACC**GGAACCATTCAAAACAGCATAGCAAGTTAAAATAAGGCTAGT CCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC**TTTTTTT

(粗 体＝tracrRNA序列;下划线＝终止序列)

[0238] U6-长tracrRNA(化脓链球菌SF370)：

[0239] GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACA AGGCTGTTAGAGAGATAATTG GAATTAATTTGACTGTAAACACAAAGAT ATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGT TT GCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAAC TTGAAAGTATTTCGATTTCTTG

GCTTTATATATCTTGTGGAAAGGACGAA ACACCGGTAGTATTAAGTATTGTTTTATGGCTGATAAATTTCTTTGAATT TCTCCTTGATTATTTGTTATAAAAGTTATAAAATAATCTTGTTGGAACCA TTCAAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTG AAAAAGTGGCACCGAGTCGGTGCTTTTTTT

[0240] U6-DR-BbsI骨架-DR(化脓链球菌SF370)：

[0241] GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACA AGGCTGTTAGAGAGATAATTGGAATTAATTTGACTGTAAACACAAAGAT ATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTT GCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAAC TTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAA ACACCGGGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACGGGTCTT CGAGAAGACGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAAC

[0242] U6-嵌合RNA-BbsI骨架(化脓链球菌SF370)

[0243] GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACA AGGCTGTTAGAGAGATAATTGGAATTAATTTGACTGTAAACACAAAGAT ATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTT GCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAAC TTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAA ACACCGGGTCTTCGAGAAGACCTGTTTTAGAGCTAGAAATAGCAAGTTA AAATAAGGCTAGTCCG

[0244] NLS-SpCas9-EGFP：

[0245] MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAADKKY SIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGE TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEE DKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKL QLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAP LSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGA SQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETI TPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNEL TKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFED REMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTI LDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGS PAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDIN RLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKN YWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVA QILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHH AHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKAT AKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRK VLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDS PTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLA SHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVL SAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSITGLYETRIDLSQLGGDAAAVSKGEELFTGVVPILVELDGDVNG HKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIEL KGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSV QLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYK

[0246] SpCas9-EGFP-NLS：

[0247] MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLI GALLFDSGETAEATRLKRT ARRRYTRRKNRICYLQEIFSNEMAKVDDSFFH RLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKL VDSTDKADL RLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS GVDAKAILSARL SKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFL AAKNLSDAILLSDILR VNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNG YAGYIDG GASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI PHQIHLGELHAILRRQEDFYPFLKDNREKI EKILTFRIPYYVGPLARGNSRFA WMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLL YEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNA SLGTYHDLLKIIKDKDFLDNEENEDILE DIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSR KLIN GIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL HEHIANLAGSPAIKKGIL QTVKVVDELVKVMGRHKPENIVIEMARENQTTQ KGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYL YYLQNGRDM YVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSE EVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVE TRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKL VSDFRKDFQFYK VREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAK SEQEIGKATAKY FFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESIL PKRNSDKLIARKKDWDP KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID FLEAK GYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSK YVNFLYLASHYEKLKGSPEDNEQKQLFVE QHKHYLDEIIEQISEFSKRVILA DANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR YTSTKEVLDATLIHQSITGLYETRIDLSQLGGDAAAVSKGEELFTGVVPILVE LDGDVNGHKFSVSGEGEGDATY GKLTLKFICTTGKLPVPWPTLVTTLTYGV QCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKF EGDT LVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRH NIEDGSVQLADHYQQNTPI GDGPVLLPDNHYLSTQSALSKDPNEKRDHMVL LEFVTAAGITLGMDELYKKRPAATKKAGQAKKKK

[0248] NLS-SpCas9-EGFP-NLS：

[0249] MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAADKKY SIGLDIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHSIKKNLIGALLFDSGE TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEE SFLVEE DKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH MIKFRGHFLIEGDLNPD NSDVDKLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNF KSNFDLAEDAKL QLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAP LSASMIKRYD EHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGA SQEEFYKFIKPILEKMDGTEELLVKLNREDLLRK QRTFDNGSIPHQIHLGELH AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETI TP WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNEL TKVKYVTEGMRKPAFLSGEQKKAIVD LLFKTNRKVTVKQLKEDYFKKIECF DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTL FED REMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTI LDFLKSDGFANRNFMQLIHD DSLTFKEDIQKAQVSGQGDSLHEHIANLAGS PAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQ KNSRER MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDIN RLSDYDVDHIVPQSFLK DDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKN YWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVE TRQITKHVA QILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHH AHDAYLNAVVGTAL IKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKAT AKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI

VWDKGRDFATVRK VLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDS PTVAYSVLVV AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK EVKKDLIIKLPKYSLFELENGRKRMLASAGELQ KGNELALPSKYVNFLYLA SHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVL SAYN KHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSITGLYETRIDLSQLGGDAA AVSKGEELFTGVVPILVELDGDVNG HKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYP DHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIEL KGIDFKEDGNILGHKLEYNYNSHNV YIMADKQKNGIKVNFKIRHNIEDGSV QLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTA AGITLGMDELYKKRPAATKKAGQAKKKK

[0250] NLS-SpCas9-NLS：

[0251] MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAADKKY SIGLDIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHSIKKNLIGALLFDSGE TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEE SFLVEE DKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH MIKFRGHFLIEGDLNPD NSDVDKLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNF KSNFDLAEDAKL QLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAP LSASMIKRYD EHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGA SQEEFYKFIKPILEKMDGTEELLVKLNREDLLRK QRTFDNGSIPHQIHLGELH AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETI TP WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNEL TKVKYVTEGMRKPAFLSGEQKKAIVD LLFKTNRKVTVKQLKEDYFKKIECF DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTL FED REMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTI LDFLKSDGFANRNFMQLIHD DSLTFKEDIQKAQVSGQGDSLHEHIANLAGS PAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQ KNSRER MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDIN RLSDYDVDHIVPQSFLK DDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKN YWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVE TRQITKHVA QILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHH AHDAYLNAVVGTAL IKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKAT AKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI VWDKGRDFATVRK VLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDS PTVAYSVLVV AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK EVKKDLIIKLPKYSLFELENGRKRMLASAGELQ KGNELALPSKYVNFLYLA SHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVL SAYN KHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSITGLYETRIDLSQLGGDKR PAATKKAGQAKKKK

[0252] NLS-mCherry-SpRNA酶3：

[0253] MFLFLSLTSFLSSSRTLVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEF EIEGEGEGRPYEGTQTAKL KVTKGGPLPFAWDILSPQFMYGSKAYVKHPAD IPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYK VKLRGTNF PSDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLKDGGHYDAEVK TTYKAKKPVQLPGAYNV NIKLDITSHNEDYTIVEQYERAEGRHSTGGMDEL YKGSKQLEELLSTSFDIQFNDLTLLETAFTHTSYANEHRLL NVSHNERLEFL GDAVLQLIISEYLFAKYPKKTEGDMSKLRSMIVREESLAGFSRFCSFDAYIKL GKGEEKSGGR RRDTILGDLFEAFLGALLLDKGIDAVRRFLKQVMIPQVEKG NFERVKDYKTCLQEFLQTKGDVAIDYQVISEKGP AHAKQFEVSIVVNGAVL SKGLGKSKKLAEQDAAKNALAQLSEV

[0254] SpRNA酶3-mCherry-NLS：

[0255] MKQLEELLSTSFDIQFNDLTLLETAFTHTSYANEHRLLNVSHNERLEFL GDAVLQLIISEYLFAKYP

KKTEGDMSKLRSMIVREESLAGFSRFCSFDAYIKL GKGEEKSGGRRRDTILGDLFEAFLGALLLDKGIDAVRRFL KQVMIPQVEKG NFERVKDYKTCLQEFLQTKGDVAIDYQVISEKGPAHAKQFEVSIVVNGAVL SKGLGKSKKLAE QDAAKNALAQLSEVGSVSKGEEDNMAIIKEFMRFKVHM EGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLP FAWDILSPQFMYGS KAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIY KVKLRGTNF PSDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLKD GGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSH NEDYTIVEQYERAEG RHSTGGMDELYKKRPAATKKAGQAKKKK

[0256] NLS-SpCas9n-NLS(D10A切口酶突变是小写字母)：

[0257] MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAADKKY SIGLaIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHSIKKNLIGALLFDSGET AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEE SFLVEED KKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMI KFRGHFLIEGDLNPD NSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSA RLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNF KSNFDLAEDAKLQL SKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYD EHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQ EEFYKFIKPILEKMDGTEELLVKLNREDLLRK QRTFDNGSIPHQIHLGELHAI LRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTK VKYVTEGMRKPAFLSGEQKKAIVD LLFKTNRKVTVKQLKEDYFKKIECFDS VEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDRE MIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILD FLKSDGFANRNFMQLIHD DSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAI KKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQ KNSRERMKR IEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSD YDVDHIVPQSFLK DDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQ LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVE TRQITKHVAQILDS RMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA YLNAVVGTAL IKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFF YSNIMNFFKTEITLANGEIRKRPLIETNGETGEI VWDKGRDFATVRKVLSMP QVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAY SVLVV AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK DLIIKLPKYSLFELENGRKRMLASAGELQ KGNELALPSKYVNFLYLASHYE KLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYN KHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLI HQSITGLYETRIDLSQLGGDKR PAATKKAGQAKKKK

[0258] hEMX1-HR模板-HindII-NheI：

[0259] GAATGCTGCCCTCAGACCCGCTTCCTCCCTGTCCTTGTCTGTCCAAG GAGAATGAGGTCTCACTGGT GGATTTCGGACTACCCTGAGGAGCTGGCA CCTGAGGGACAAGGCCCCCCACCTGCCCAGCTCCAGCCTCTGATGA GGG GTGGGAGAGAGCTACATGAGGTTGCTAAGAAAGCCTCCCCTGAAGGAG ACCACACAGTGTGTGAGGTTGGA GTCTCTAGCAGCGGGTTCTGTGCCCC CAGGGATAGTCTGGCTGTCCAGGCACTGCTCTTGATATAAACACCACCT CCTAGTTATGAAACCATGCCCATTCTGCCTCTCTGTATGGAAAAGAGCA TGGGGCTGGCCCGTGGGGTGGTGTCC ACTTTAGGCCCTGTGGGAGATCA TGGGAACCCACGCAGTGGGTCATAGGCTCTCTCATTTACTACTCACATCC A CTCTGTGAAGAAGCGATTATGATCTCTCCTCTAGAAACTCGTAGAGTC CCATGTCTGCCGGCTTCCAGAGCCTGC ACTCCTCCACCTTGGCTTGGCTT TGCTGGGGCTAGAGGAGCTAGGATGCACAGCAGCTCTGTGACCCTTTGT TT GAGAGGAACAGGAAAACCACCCTTCTCTCTGGCCCACTGTGTCCTCTT CCTGCCCTGCCATCCCCTTCTGTGAAT GTTAGACCCATGGGAGCAGCTG GTCAGAGGGGACCCCGGCCTGGGGCCCCTAACCCTATGTAGCCTCAGTC TTC CCATCAGGCTCTCAGCTCAGCCTGAGTGTTGAGGCCCCAGTGGCTG CTCTGGGGGCCTCCTGAGTTTCTCATCTG

TGCCCCTCCCTCCCTGGCCCA GGTGAAGGTGTGGTTCCAGAACCGGAGGACAAAGTACAAACGGCAGAA GCTGG AGGAGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGGGCTCCCA TCACATCAACCGGTGGCGCATTGCCACGAAGC AGGCCAATGGGGAGGA CATCGATGTCACCTCCAATGACaagcttgctagcGGTGGGCAACCACAAACCC ACGAG GGCAGAGTGCTGCTTGCTGCTGGCCAGGCCCCTGCGTGGGCCCA AGCTGGACTCTGGCCACTCCCTGGCCAGGCT TTGGGGAGGCCTGGAGTC ATGGCCCCACAGGGCTTGAAGCCCGGGGCCGCCATTGACAGAGGGACA AGCAATGG GCTGGCTGAGGCCTGGGACCACTTGGCCTTCTCCTCGGAGA GCCTGCCTGCCTGGGCGGGCCCGCCCGCCACCGC AGCCTCCCAGCTGCT CTCCGTGTCTCCAATCTCCCTTTTGTTTTGATGCATTTCTGTTTTAATTTA TTTTCCAG GCACCACTGTAGTTTAGTGATCCCCAGTGTCCCCCTTCCCTA TGGGAATAATAAAAGTCTCTCTCTTAATGACAC GGGCATCCAGCTCCAG CCCCAGAGCCTGGGGTGGTAGATTCCGGCTCTGAGGGCCAGTGGGGGCT GGTAGAGCA AACGCGTTCAGGGCCTGGGAGCCTGGGGTGGGGTACTGG TGGAGGGGGTCAAGGGTAATTCATTAACTCCTCTCT TTTGTTGGGGGAC CCTGGTCTCTACCTCCAGCTCCACAGCAGGAGAAACAGGCTAGACATAG GGAAGGGCCATC CTGTATCTTGAGGGAGGACAGGCCCAGGTCTTTCTTA ACGTATTGAGAGGTGGGAATCAGGCCCAGGTAGTTCAA TGGGAGAGGG AGAGTGCTTCCCTCTGCCTAGAGACTCTGGTGGCTTCTCCAGTTGAGGA GAAACCAGAGGAAAG GGGAGGATTGGGGTCTGGGGGAGGGAACACCAT TCACAAAGGCTGACGGTTCCAGTCCGAAGTCGTGGGCCCACC AGGATGC TCACCTGTCCTTGGAGAACCGCTGGGCAGGTTGAGACTGCAGAGACAGG GCTTAAGGCTGAGCCTGC AACCAGTCCCCAGTGACTCAGGGCCTCCTCA GCCCAAGAAAGAGCAACGTGCCAGGGCCCGCTGAGCTCTTGTGT TCACC TG

[0260] NLS-StCsn1-NLS：

[0261] MKRPAATKKAGQAKKKKSDLVLGLDIGIGSVGVGILNKVTGEIIHKNS RIFPAAQAENNLVRRTNRQ GRRLARRKKHRRVRLNRLFEESGLITDFTKISI NLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLDDASD DGNSSVGDY AQIVKENSKQLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHRLINVFPTS AYRSEALRILQTQ QEFNPQITDEFINRYLEILTGKRKYYHGPGNEKSRTDYG RYRTSGETLDNIFGILIGKCTFYPDEFRAAKASYTA QEFNLLNDLNNLTVPTE TKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVADIKGYRIDKSG KAEIHT FEAYRKMKTLETLDIEQMDRETLDKLAYVLTLNTEREGIQEALEH EFADGSFSQKQVDELVQFRKANSSIFGKGW HNFSVKLMMELIPELYETSEE QMTILTRLGKQKTTSSSNKTKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAIK EYGDFDNIVIEMARETNEDDEKKAIQKIQKANKDEKDAAMLKAANQYNGK AELPHSVFHGHKQLATKIRLWHQQG ERCLYTGKTISIHDLINNSNQFEVDHI LPLSITFDDSLANKVLVYATANQEKGQRTPYQALDSMDDAWSFRELKA FV RESKTLSNKKKEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHF RAHKIDTKVSVVRGQFTSQL RRHWGIEKTRDTYHHHAVDALIIAASSQLNL WKKQKNTLVSYSEDQLLDIETGELISDDEYKESVFKAPYQHFVD TLKSKEF EDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKADETYVLGKIKDIYTQD GYDAFMKIYKKDKSK FLMYRHDPQTFEKVIEPILENYPNKQINEKGKEVPC NPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLGNHI DITPKDSNNKVVL QSVSPWRADVYFNKTTGKYEILGLKYADLQFEKGTGTYKISQEKYNDIKKK EGVDSDSEFK FTLYKNDLLLVKDTETKEQQLFRFLSRTMPKQKHYVELKPY DKQKFEGGEALIKVLGNVANSGQCKKGLGKSNIS IYKVRTDVLGNQHIIKN EGDKPKLDFKRPAATKKAGQAKKKK

[0262] U6-St_tracrRNA(7-97)：

[0263] GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACA AGGCTGTTAGAGAGATAATTG GAATTAATTTGACTGTAAACACAAAGAT ATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGT TT GCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAAC TTGAAAGTATTTCGATTTCTTG GCTTTATATATCTTGTGGAAAGGACGAA ACACCGTTACTTAAATCTTGCAGAAGCTACAAAGATAAGGCTTCATG

CC GAAATCAACACCCTGTCATTTTATGGCAGGGTGTTTTCGTTATTTAA

[0264] U6-DR-间隔子-DR(化脓链球菌SF370)

[0265] gagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattggaattaatttga
ctgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttt
taaaatggactatcatatgcttaccgtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgaaacacc
gggttttagagctatgctgttttgaatggtcccaaaacNNNNNNNNNNNNNNNNNNNNNNNN

NNNNNgttttagagctatgctgttttgaatggtcccaaaacTTTTTTT (小写字母下划线=同向 重复;N=指导序列;粗体=终止子)

[0266] 嵌合RNA,包含+48tracr RNA(化脓链球菌SF370)

[0267] gagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattggaattaatttga
ctgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttt
taaaatggactatcatatgcttaccgtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgaaacacc
NNNNNNNNNNNNNNNNNNNgttttagagctagaaatagcaagttaaaataaggctagtccgTT

TTTTT (N=指导序列;第一下划线=tracr配对序列;第二下划线=tracr序列;粗体=终止子)

[0268] 嵌合RNA,包含+54tracr RNA(化脓链球菌SF370)

[0269] gagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattggaattaatttga
ctgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttt
taaaatggactatcatatgcttaccgtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgaaacacc
NNNNNNNNNNNNNNNNNNNgttttagagctagaaatagcaagttaaaataaggctagtccgttatc

aTTTTTTTT (N=指导序列;第一下划线=tracr配对序列;第二下划线= tracr序列;粗体=终止子)

[0270] 嵌合RNA,包含+67tracr RNA(化脓链球菌SF370)

[0271] gagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattggaattaatttga
ctgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttt
taaaatggactatcatatgcttaccgtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgaaacacc
NNNNNNNNNNNNNNNNNNNgttttagagctagaaatagcaagttaaaataaggctagtccgttatc

aacttgaaaaagtgTTTTTTT (N=指导序列;第一下划线=tracr配对序列;第二 下划线=tracr序列;粗体=终止子)

[0272] 嵌合RNA,包含+85tracr RNA(化脓链球菌SF370)

[0273] gagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattggaattaatttga
ctgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttt
taaaatggactatcatatgcttaccgtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgaaacacc
NNNNNNNNNNNNNNNNNNNgttttagagctagaaatagcaagttaaaataaggctagtccgttatc

aacttgaaaaagtggcaccgagtcggtgcTTTTTTT (N=指导序列;第一下划线=tracr 配对序列;第二下划线=tracr序列;粗体=终止子)

[0274] CBh-NLS-SpCas9-NLS

[0275] CGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAAC GACCCCCGCCCATTGACGTCA ATAATGACGTATGTTCCCATAGTAACGC CAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAA AC TGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCT ATTGACGTCAATGACGGTAAATG GCCCGCCTGGCATTATGCCCAGTACA TGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATC

GCTATTACCATGGTCGAGGTGAGCCCCACGTTCTGCTTCACTCTCCCCAT CTCCCCCCCCTCCCCACCCCCAATT
TTGTATTTATTTATTTTTTAATTATT TTGTGCAGCGATGGGGGCGGGGGGGGGGGGGGGCGCGCGCCAGGCGG
GGCGGGGCGGGGCGAGGGGCGGGGCGGGGCGAGGCGGAGAGGTGCGG CGGCAGCCAATCAGAGCGGCGCGCTCCG
AAAGTTTCCTTTTATGGCGAG GCGGCGGCGGCGGCGGCCCTATAAAAAGCGAAGCGCGCGGCGGGCGGG AGTCG
CTGCGACGCTGCCTTCGCCCCGTGCCCCGCTCCGCCGCCGCCTCG CGCCGCCCGCCCCGGCTCTGACTGACCGCG
TTACTCCCACAGGTGAGCG GGCGGGACGGCCCTTCTCCTCCGGGCTGTAATTAGCTGAGCAAGAGGTA AGGGTT
TAAGGGATGGTTGGTTGGTGGGGTATTAATGTTTAATTACCTG GAGCACCTGCCTGAAATCACTTTTTTTCAGGT
TGGaccggtgccaccATGGACT ATAAGGACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGAT
GACGATAAGATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGACAAGAAGTACAGCA
TCGGCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATT
CAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAA
ACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGC
AAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGA
AGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCC
ACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGG
CCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCT
GTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACGCCAAG
GCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGA
ATGGCCTGTTCGGCAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGA
GGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAG
TACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCG
AGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAA
AGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGC
TACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCG
AGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCA
CCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGG
GAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCG
CCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGC
CCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTG
CTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCT
TCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCT
GAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCC
TCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACA
TTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGC
CCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAG
CTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACA
GAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCA
GGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTG
AAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGA

ACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGG
CAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAG
AATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGC
CTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGA
CAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACC
CAGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGA
GACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGA
CGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGAT
TTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAA
CCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAA
GATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTT
TTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGG
AGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAA
AAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCC
AGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGG
CCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAG
CAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAG
CTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGG
GAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTC
CCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATC
AGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGG
ATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTT
CAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCAC
CAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGCGACTTTCTTTTTCTTAGCTTGA
CCAGCTTTCTTAGTAGCAGCAGGACGCTTTAA(下划线＝NLS- hSpCas9-NLS)

[0276] 针对嗜热链球菌LMD-9CRISPR1 Cas9的实例嵌合RNA(具有 NNAGAAW的PAM)

[0277] NNNNNNNNNNNNNNNNNNgtttttgtactctcaagatttaGAAAtaaatcttgcagaa gctacaaagataaggcttcatgccgaaatcaacaccctgtcattttatggcagggtgttttcgttatttaaTTTTTT (N＝指导序列;第一下划线＝tracr配对序列;第二下划线＝tracr序列;粗 体＝终止子)

[0278] 针对嗜热链球菌LMD-9CRISPR1 Cas9的实例嵌合RNA(具有 NNAGAAW的PAM)

[0279] NNNNNNNNNNNNNNNNNNNNgtttttgtactctcaGAAAtgcagaagctacaaagataa ggcttcatgccgaaatcaacaccctgtcattttatggcagggtgttttcgttatttaaTTTTTT (N＝指导序 列;第一下划线＝tracr配对序列;第二下划线＝tracr序列;粗体＝终止子)

[0280] 针对嗜热链球菌LMD-9CRISPR1 Cas9的实例嵌合RNA(具有 NNAGAAW的PAM)

[0281] NNNNNNNNNNNNNNNNNNNNgtttttgtactctcaGAAAtgcagaagctacaaagataa ggcttcatgccgaaatcaacaccctgtcattttatggcagggtgtTTTTTT (N＝指导序列;第一下 划线＝tracr配对序列;第二下划线＝tracr序列;粗体＝终止子)

[0282] 针对嗜热链球菌LMD-9CRISPR1 Cas9的实例嵌合RNA(具有 NNAGAAW的PAM)

[0283] NNNNNNNNNNNNNNNNNNgttattgtactctcaagatttaGAAAtaaatcttgcagaagctacaaagataaggcttcatgccgaaatcaacaccctgtcattttatggcagggtgttttcgttatttaaTTTTTT (N＝指导序列;第一下划线＝tracr配对序列;第二下划线＝tracr序列;粗 体＝终止子)

[0284] 针对嗜热链球菌LMD-9CRISPR1 Cas9的实例嵌合RNA(具有 NNAGAAW的PAM)

[0285] NNNNNNNNNNNNNNNNNNgttattgtactctcaGAAAtgcagaagctacaaagataaggcttcatgccgaaatcaacaccctgtcattttatggcagggtgttttcgttatttaaTTTTTT (N＝指导序 列;第一下划线＝tracr配对序列;第二下划线＝tracr序列;粗体＝终止子)

[0286] 针对嗜热链球菌LMD-9CRISPR1 Cas9的实例嵌合RNA(具有 NNAGAAW的PAM)

[0287] NNNNNNNNNNNNNNNNNNgttattgtactctcaGAAAtgcagaagctacaaagataaggcttcatgccgaaatcaacaccctgtcattttatggcagggtgtTTTTTT (N＝指导序列;第一下划线＝tracr配对序列;第二下划线＝tracr序列;粗体＝终止子)

[0288] 针对嗜热链球菌LMD-9CRISPR1 Cas9的实例嵌合RNA(具有 NNAGAAW的PAM)

[0289] NNNNNNNNNNNNNNNNNNgttattgtactctcaagatttaGAAAtaaatcttgcagaagctacaatgataaggcttcatgccgaaatcaacaccctgtcattttatggcagggtgttttcgttatttaaTTTTTT (N＝指导序列;第一下划线＝tracr配对序列;第二下划线＝tracr序列;粗 体＝终止子)

[0290] 针对嗜热链球菌LMD-9CRISPR1 Cas9的实例嵌合RNA(具有 NNAGAAW的PAM)

[0291] NNNNNNNNNNNNNNNNNNgttattgtactctcaGAAAtgcagaagctacaatgataaggcttcatgccgaaatcaacaccctgtcattttatggcagggtgttttcgttatttaaTTTTTT (N＝指导序 列;第一下划线＝tracr配对序列;第二下划线＝tracr序列;粗体＝终止子)

[0292] 针对嗜热链球菌LMD-9CRISPR1 Cas9的实例嵌合RNA(具有 NNAGAAW的PAM)

[0293] NNNNNNNNNNNNNNNNNNgttattgtactctcaGAAAtgcagaagctacaatgataaggcttcatgccgaaatcaacaccctgtcattttatggcagggtgtTTTTTT (N＝指导序列;第一下 划线＝tracr配对序列;第二下划线＝tracr序列;粗体＝终止子)

[0294] 针对嗜热链球菌LMD-9CRISPR3 Cas9的实例嵌合RNA(具有 NGGNG的PAM)

[0295] NNNNNNNNNNNNNNNNNNgttttagagctgtgGAAAcacagcgagttaaaataaggcttagtccgtactcaacttgaaaaggtggcaccgattcggtgtTTTTTT (N＝指导序列;第一加 下划线的＝tracr配对序列;第二加下划线的＝tracr序列;加粗的＝终止子)

[0296] 来自嗜热链球菌LMD-9CRISPR3基因座的密码子优化版本的Cas9 (在5'以及3'两端都具有一个NLS)

[0297] ATGAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAA GAAAAAGACCAAGCCCTACAGCATCGGCCTGGACATCGGCACCAATAG CGTGGGCTGGGCCGTGACCACCGACAACTACAAGGTGCCCAGCAAGAA AATGAAGGTGCTGGGCAACACCTCCAAGAAGTACATCAAGAAAAACCT GCTGGGCGTGCTGCTGTTCGACAGCGGCATTACAGCCGAGGGCAGACGG CTGAAGAGAACCGCCAGACGGCGGTACACCCGGCGGAGAAACAGAATC CTGTATCTGCAAGAGATCTTCAGCACCGAGATGGCTACCCTGGACGACG CCTTCTTCCAGCGGCTGGACGACAGCTTCCT

GGTGCCCGACGACAAGCG GGACAGCAAGTACCCCATCTTCGGCAACCTGGTGGAAGAGAAGGCCTA CCACGACG
AGTTCCCCACCATCTACCACCTGAGAAAGTACCTGGCCGAC AGCACCAAGAAGGCCGACCTGAGACTGGTGTATC
TGGCCCTGGCCCACA TGATCAAGTACCGGGGCCACTTCCTGATCGAGGGCGAGTTCAACAGCAA GAACAACGAC
ATCCAGAAGAACTTCCAGGACTTCCTGGACACCTACAAC GCCATCTTCGAGAGCGACCTGTCCCTGGAAAACAGC
AAGCAGCTGGAA GAGATCGTGAAGGACAAGATCAGCAAGCTGGAAAAGAAGGACCGCATC CTGAAGCTGTTCCC
CGGCGAGAAGAACAGCGGAATCTTCAGCGAGTTTC TGAAGCTGATCGTGGGCAACCAGGCCGACTTCAGAAAGTG
CTTCAACCT GGACGAGAAAGCCAGCCTGCACTTCAGCAAAGAGAGCTACGACGAGGA CCTGGAAACCCTGCTGG
GATATATCGGCGACGACTACAGCGACGTGTTC CTGAAGGCCAAGAAGCTGTACGACGCTATCCTGCTGAGCGGCT
TCCTGA CCGTGACCGACAACGAGACAGAGGCCCCACTGAGCAGCGCCATGATTA AGCGGTACAACGAGCACAAA
GAGGATCTGGCTCTGCTGAAAGAGTACA TCCGGAACATCAGCCTGAAAACCTACAATGAGGTGTTCAAGGACGAC
AC CAAGAACGGCTACGCCGGCTACATCGACGGCAAGACCAACCAGGAAGA TTTCTATGTGTACCTGAAGAAGCT
GCTGGCCGAGTTCGAGGGGGCCGAC TACTTTCTGGAAAAAATCGACCGCGAGGATTTCCTGCGGAAGCAGCGGA
CCTTCGACAACGGCAGCATCCCCTACCAGATCCATCTGCAGGAAATGCG GGCCATCCTGGACAAGCAGGCCAAGT
TCTACCCATTCCTGGCCAAGAAC AAAGAGCGGATCGAGAAGATCCTGACCTTCCGCATCCCTTACTACGTGG GC
CCCCTGGCCAGAGGCAACAGCGATTTTGCCTGGTCCATCCGGAAGCG CAATGAGAAGATCACCCCCTGGAACTTC
GAGGACGTGATCGACAAAGA GTCCAGCGCCGAGGCCTTCATCAACCGGATGACCAGCTTCGACCTGTAC CTGCC
CGAGGAAAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGACAT TCAATGTGTATAACGAGCTGACCAAAGTGCG
GTTTATCGCCGAGTCTAT GCGGGACTACCAGTTCCTGGACTCCAAGCAGAAAAAGGACATCGTGCG GCTGTACT
TCAAGGACAAGCGGAAAGTGACCGATAAGGACATCATCGA GTACCTGCACGCCATCTACGGCTACGATGGCATCG
AGCTGAAGGGCATC GAGAAGCAGTTCAACTCCAGCCTGAGCACATACCACGACCTGCTGAACA TTATCAACGAC
AAAGAATTTCTGGACGACTCCAGCAACGAGGCCATCAT CGAAGAGATCATCCACACCCTGACCATCTTTGAGGAC
CGCGAGATGATC AAGCAGCGGCTGAGCAAGTTCGAGAACATCTTCGACAAGAGCGTGCTG AAAAAGCTGAGCAG
ACGGCACTACACCGGCTGGGGCAAGCTGAGCGCC AAGCTGATCAACGGCATCCGGGACGAGAAGTCCGGCAACAC
AATCCTG GACTACCTGATCGACGACGGCATCAGCAACCGGAACTTCATGCAGCTGA TCCACGACGACGCCCTGA
GCTTCAAGAAGAAGATCCAGAAGGCCCAGA TCATCGGGGACGAGGACAAGGGCAACATCAAAGAAGTCGTGAAGT
CCC TGCCCGGCAGCCCCGCCATCAAGAAGGGAATCCTGCAGAGCATCAAGA TCGTGGACGAGCTCGTGAAAGTG
ATGGGCGGCAGAAAGCCCGAGAGCA TCGTGGTGGAAATGGCTAGAGAGAACCAGTACACCAATCAGGGCAAGA G
CAACAGCCAGCAGAGACTGAAGAGACTGGAAAAGTCCCTGAAAGAGC TGGGCAGCAAGATTCTGAAAGAGAATAT
CCCTGCCAAGCTGTCCAAGAT CGACAACAACGCCCTGCAGAACGACCGGCTGTACCTGTACTACCTGCAG AATG
GCAAGGACATGTATACAGGCGACGACCTGGATATCGACCGCCTG AGCAACTACGACATCGACCATATTATCCCCC
AGGCCTTCCTGAAAGACA ACAGCATTGACAACAAAGTGCTGGTGTCCTCCGCCAGCAACCGCGGCAA GTCCGAT
GATGTGCCCAGCCTGGAAGTCGTGAAAAAGAGAAAGACCTTC TGGTATCAGCTGCTGAAAAGCAAGCTGATTAGC
CAGAGGAAGTTCGACA ACCTGACCAAGGCCGAGAGAGGCGGCCTGAGCCCTGAAGATAAGGCCG GCTTCATCCA
GAGACAGCTGGTGGAAACCCGGCAGATCACCAAGCACGT GGCCAGACTGCTGGATGAGAAGTTTAACAACAAGAA
GGACGAGAACAA CCGGGCCGTGCGGACCGTGAAGATCATCACCCTGAAGTCCACCCTGGTG TCCCAGTTCCGGA
AGGACTTCGAGCTGTATAAAGTGCGCGAGATCAATG ACTTTCACCACGCCCACGACGCCTACCTGAATGCCGTGG
TGGCTTCCGC CCTGCTGAAGAAGTACCCTAAGCTGGAACCCGAGTTCGTGTACGGCGAC TACCCCAAGTACAAC
TCCTTCAGAGAGCGGAAGTCCGCCACCGAGAAG GTGTACTTCTACTCCAACATCATGAATATCTTTAAGAAGTCC
ATCTCCCT GGCCGATGGCAGAGTGATCGAGCGGCCCCTGATCGAAGTGAACGAAGA GACAGGCGAGAGCGTGTG

GAACAAAGAAAGCGACCTGGCCACCGTGCG GCGGGTGCTGAGTTATCCTCAAGTGAATGTCGTGAAGAAGGTGGA AGA ACAGAACCACGGCCTGGATCGGGGCAAGCCCAAGGGCCTGTTCAACGC CAACCTGTCCAGCAAGCCTAAGC CCAACTCCAACGAGAATCTCGTGGGG GCCAAAGAGTACCTGGACCCTAAGAAGTACGGCGGATACGCCGGCATC TCCAATAGCTTCACCGTGCTCGTGAAGGGCACAATCGAGAAGGGCGCTA AGAAAAAGATCACAAACGTGCTGGAA TTTCAGGGGATCTCTATCCTGGA CCGGATCAACTACCGGAAGGATAAGCTGAACTTTCTGCTGGAAAAAGGC TA CAAGGACATTGAGCTGATTATCGAGCTGCCTAAGTACTCCCTGTTCG AACTGAGCGACGGCTCCAGACGGATGCT GGCCTCCATCCTGTCCACCAA CAACAAGCGGGGCGAGATCCACAAGGGAAACCAGATCTTCCTGAGCCA GAAAT TTGTGAAACTGCTGTACCACGCCAAGCGGATCTCCAACACCATC AATGAGAACCACCGGAAATACGTGGAAAACC ACAAGAAAGAGTTTGAG GAACTGTTCTACTACATCCTGGAGTTCAACGAGAACTATGTGGGAGCCA AGAAGAAC GGCAAACTGCTGAACTCCGCCTTCCAGAGCTGGCAGAACC ACAGCATCGACGAGCTGTGCAGCTCCTTCATCGGC CCTACCGGCAGCGA GCGGAAGGGACTGTTTGAGCTGACCTCCAGAGGCTCTGCCGCCGACTTT GAGTTCCTGGG AGTGAAGATCCCCCGGTACAGAGACTACACCCCCTCTA GTCTGCTGAAGGACGCCACCCTGATCCACCAGAGCGT GACCGGCCTGTA CGAAACCCGGATCGACCTGGCTAAGCTGGGCGAGGGAAAGCGTCCTGC TGCTACTAAGAAAG CTGGTCAAGCTAAGAAAAAGAAATAA

实例5:使用CRISPR-Cas系统的细菌基因组的RNA指导的编辑。

[0298] 申请人使用CRISPR-相关的内切核酸酶Cas9以在肺炎链球菌以及 大肠杆菌的基因组中引入精确的突变。该途径依赖于在靶位点处的Cas9-指导 的切割以杀死未突变的细胞,并且回避对选择性标记或反选择系统的需要。Cas9特异性是通过改变短CRISPR RNA(crRNA)的序列的重新编程的,以 使得在编辑模板上发生单以及多核苷酸改变。同时使用两个crRNA使得多元 诱变成为可能。在肺炎链球菌中,将近100%的从Cas9切割幸存的细胞包含 所希望的突变,并且当在大肠杆菌中与重组工程组合使用时是65%。申请人 全面分析了Cas9靶标需求以定义可靶定序列的范围,并且显示不满足这些需 求的编辑位点的策略,表明了细菌基因组工程的这种技术的多样性。

[0299] 理解基因功能依赖于以受控的方式改变细胞内的DNA序列的可能 性。真核生物中的位点特异性诱变是使用序列特异性核酸酶实现的,这些核 酸酶促进包含感兴趣的突变的一种模板DNA的同源重组。锌指核酸酶 (ZFN)、转录激活子样效应因子核酸酶(TALEN)以及归巢大范围核酸酶 (homing meganuclease)可以被编程以切割特异性位置中的基因组,但这些 途径需要针对每个靶序列的新的酶的工程化。在原核生物体中,诱变方法在 该编辑的基因座中引入一种选择标记或需要一个包括反选择系统的两步法。 最近,噬菌体重组蛋白已被用于重组工程,这是促进线性DNA或寡核苷酸的 同源重组的一种技术。然而,因为不存在对突变的筛选,重组工程化效率可 能是相对低的(从针对点突变的0.1%-10%低至针对更大的修饰的10^{-5}-10^{-6}),在许多情况下需要筛选大数目的菌落。因此,仍存在对可承受的、容易 使用并且有效的新的技术的需求,以用于真核以及原核生物体两者中的基因工程化。

[0300] 对原核生物的CRISPR(规律间隔成簇短回文重复序列)适应性免 疫系统的最近工作已导致鉴定序列特异性被小RNA程序化的核酸酶。CRISPR基因座是由被‘间隔子’序列分离的一系列重复序列构成的,这些重 复序列匹配细菌噬菌体以及其他可动遗传因子的基因组。该重复-间隔子阵列 被转录为一个长前体并且在重复序列内被加工以产生小crRNA,该小crRNA 指定被CRISPR系统切割的靶序列(也被称为原型间隔子)。对切割而言必 要地

是恰好在该靶区域的下游的序列基序的存在，该序列基序被称为原型间 隔子相邻基序(PAM)。CRISPR-相关(cas)基因通常在该重复-间隔子阵列 侧翼并且编码对crRNA生源说以及靶向负责的酶机构。Cas9是一种dsDNA 内切核酸酶，使用一种crRNA指导以指定切割位点。该crRNA指导向Cas9 上的装载发生在加工该crRNA前体过程中并且需要反义于该前体的一个小 RNA、tracrRNA、以及RNA酶III。与用ZFN或TALEN编辑的基因组相 比，改变Cas9靶特异性不需要蛋白质工程化而只需要短crRNA指导的设计。

[0301] 申请人最近揭示在肺炎链球菌中，引入靶向一个染色体基因座的一 种CRISPR系统导致杀死这些转化细胞。观察到偶然的幸存者在靶区域包含 突变，表明可使用针对内源靶标的Cas9 dsDNA内切核酸酶活性用于基因组编 辑。申请人揭示，无标记物突变可以通过将在该基因组中重组并且将消除 Cas9靶识别的一个模板DNA片段的转化来引入。具有若干不同crRNA的 Cas9的特异性的引导允许在相同的时间引入多个突变。申请人也详细表征了Cas9靶向的序列需求，并且揭示该途径可以与重组工程组合，用于在大肠杆 菌中进行基因组编辑。

[0302] 结果：通过一个染色体靶标的Cas9切割进行基因组编辑

[0303] 肺炎链球菌菌株crR6包含一个基于Cas9的CRISPR系统，该系统 切割存在于细菌噬菌体Φ8232.5中的一个靶序列。此靶标被整合至一种第二菌 株$R6^{8232.5}$的srtA染色体基因座中。在PAM区域包含一种突变的一个改变的 靶序列被整合至一种第三菌株$R6^{370.1}$的srtA基因座中，导致此菌株对 CRISPR切割‘免疫’(图28a)。申请人用来自crR6细胞的基因组DNA转 化$R6^{8232.5}$以及$R6^{370.1}$细胞，期望成功转化$R6^{8232.5}$细胞会导致该靶基因座的切 割以及细胞死亡。与此期望相反，申请人分离$R6^{8232.5}$转化体，尽管其比 $R6^{370.1}$转化体具有少大约10倍的效率(图28b)。八个$R6^{8232.5}$转化体的遗传 分析(图28)揭示大多数是双重组事件的产物，通过将Φ8232.5靶标用crR6 基因组的野生型srtA基因座替换，该双重组事件消除Cas9靶向的毒性，该消 除不包含Cas9识别所需要的原型间隔子。这些结果是如下的证据，并行引入 靶向一个基因组座位的一个CRISPR系统(靶向构建体)连同用以重组至靶 定的基因座中的一个模板(编辑模板)导致靶定的基因组编辑(图23a)。

[0304] 为了建立用于基因组编辑的一个简化系统，通过缺失cas1、cas2以 及csn2，即已显示对于CRISPR靶向来说可有可无的的基因，申请人修饰菌 株crR6中的CRISPR座位，产生菌株crR6M(图28a)。此菌株保留crR6的 相同特性(图28b)。为了增加基于Cas9的编辑的效率并且表明所选择的一 个模板DNA可用以控制引入的突变，申请人用野生型srtA基因或突变体 $R6^{370.1}$靶标的PCR产物共转化$R6^{8232.5}$细胞，这两个靶标中任一个应该对Cas9 的切割具有抗性。这导致转化频率相较于单独的基因组crR6 DNA的5至10 倍增加(图23b)。编辑的效率也实质性增加，8/8所测试的转化体包含一个 野生型srtA拷贝，并且7/8包含存在于该$R6^{370.1}$靶标中的PAM突变(图23b 以及图29a)。一起考虑，这些结果显示Cas9辅助基因组编辑的可能性。

[0305] 对Cas9靶需求的分析：为了在该基因组中引入特异性改变，必须 使用携带消除Cas9介导的切割的突变的一个编辑模板，从而防止细胞死亡。当寻求靶标缺或其被另一个序列(基因插入)替代时这是容易实现的。当目 标是产生基因融合或产生单核苷酸突变时，Cas9核酸酶活性的消除仅在通过 在改变PAM或原型间隔子的编辑模板中引入突变时有可能。为了确定 CRISPR介导的编辑的约束，申请人全面分析了废止CRISPR靶向的PAM以 及原

型间隔子突变。

[0306] 前述研究提出,化脓链球菌Cas9需要恰好在该原型间隔子的下游 的一个NGG PAM。然而,因为迄今为止仅描述了非常有限数目的PAM钝化 突变,申请人进行了一个系统性分析,从而发现在消除CRISPR切割的原型 间隔子之后的所有5-核苷酸序列。申请人使用随机化寡核苷酸产生在转化至 crR6或R6细胞中的一种异质PCR产物中的所有可能的1,024个PAM序列。 预期携带功能性PAM的构建体被crR6(而不是R6细胞)中的Cas9识别并 破坏(图24a)。超过2×10^5个菌落汇集在一起以提取DNA,用于用作共扩 增所有靶标的模板。对PCR产物进行深度测序并且发现包含所有1,024个序 列,覆盖范围从5至42,472个读数(参见“深度测序数据分析”部分)。将 每个PAM的功能通过其在crR6样品中与R6样品中的读数的相对比例进行估 计。PAM的前三个碱基的分析(经最后两个碱基平均)清楚地显示,NGG型在crR6转化体中是非代表性的(图24b)。此外,后两个碱基对NGG PAM 没有可检测到的影响(参见“深度测序数据分析”部分),表明NGGNN序 列足以许可Cas9活性。针对NAG PAM序列观察到部分靶向(图24b)。另 外,该NNGGN型部分地钝化CRISPR靶向(表G),表明该NGG基序当移 位1bp时仍可通过Cas9以降低的效率被识别。这些数据阐明Cas9靶识别的 分子机制,并且它们揭示NGG(或互补链上的CCN)序列足以靶向Cas9, 并且在该编辑模板中应避免NGG至NAG或NNGGN突变。由于这些三核苷 酸序列的高频率(每隔8bp一次),这意味着该基因组的几乎任何位置都可 以被编辑。确实,申请人测试了十个随机选取的携带不同PAM的靶标并且发 现所有都是功能性的(图30)。

[0307] 破坏Cas9介导的切割的另一个方式是在该编辑模板的原型间隔子 区域引入突变。已知‘种子序列’(恰邻近于该PAM的8至10个原型间隔 子核苷酸)内的点突变可消除CRISPR核酸酶进行的切割。然而,此区域的 精确长度是未知的,并且不清楚该种子中是否突变为任何核苷酸均可破坏 Cas9靶识别。申请人遵循上述相同的深度测序途径以随机化整个涉及到与该 crRNA接触的碱基对中的原型间隔子,并且确定所有破坏靶向的序列。存在 于$R6^{8232.5}$细胞中的spc1靶标中的20个匹配核苷酸(14)的每个位置(图23a)被随机化并且转化至crR6以及R6细胞中(图24a)。与一个种子序列 的存在一致,仅恰好在该PAM的上游的12个核苷酸的突变废止Cas9进行的 切割(图24c)。然而,不同突变展现显著不同的影响。该种子(12至7)的 末端(从该PAM开始)位置耐受大部分突变并且仅一个具体的碱基取代废止 靶向。相比之下,在近端位置(6至1,除3之外)突变为任何核苷酸消除 Cas9活性,尽管每个具体取代处于不同的水平。在位置3,仅两个取代影响 CRISPR活性并且是以不同强度。申请人推断,尽管种子序列突变可防止 CRISPR靶向,但存在关于可以在种子的每个位置发生的核苷酸变化的限制。 此外,这些限制最可能针对不同间隔子序列而变化。因此申请人认为该PAM 序列上的突变,如果可能,应该是优选的编辑策略。可替代地,该种子序列 上的多个突变可被引入以防止Cas9核酸酶活性。

[0308] 肺炎链球菌中Cas9介导的基因组编辑:为了研发一种用于靶定的 基因组编辑的快速且有效的方法,申请人工程化菌株crR6Rk,这是间隔子可 容易通过PCR引入其中的一种菌株(图33)。申请人决定编辑肺炎链球菌的 β-半乳糖苷酶(bgaA)基因,其活性可以容易地测量。申请人在此酶的活性 位点中引入氨基酸的丙氨酸取代:R481A(R→A)以及N563A,E564A (NE→AA)突变。为了说明不同编辑策略,申请人设计该PAM序列以及该 原型间隔子种子两者的突变。在两种情况下,使用具有与邻近于TGG PAM 序列的β-半乳糖苷酶基因的

一个区域互补的crRNA相同的靶向构建体(互补 链中的CCA,图26)。该R→A编辑模板在该原型间隔子种子序列上产生一 个三核苷酸错配(CGT至GCA,也引入一个BtgZI限制酶切位点)。在该 NE→AA编辑模板中,申请人同时引入一个同义突变,该同义突变产生一个 钝化PAM(TGG至TTG)连同在该原型间隔子区域的218nt下游的突变 (AAT GAA至GCT GCA,也产生一个TseI限制酶切位点)。后一种编辑策 略表明使用一个远端PAM在如下位置制造突变的可能性,在该位置处可能难 以选择一个适合的靶标。例如,尽管肺炎链球菌R6基因组(其具有39.7% GC含量)平均每隔12bp包含一个PAM基序,一些PAM基序被分离成高达 194bp(图33)。此外申请人设计了6,664bp的一种ΔbgaA框内缺失。在所 有三种情况下,靶向以及编辑模板的共转化比用一种包含野生型bgaA序列的 对照编辑模板共转化产生多10倍的卡那霉素抗性细胞(图25b)。申请人对 24个转化体(每个编辑试验为8个)进行基因分型,并且发现除一个外全部 结合所希望的改变(图25c)。DNA测序也证实在该靶区不仅存在引入的突变而且也不存在继发性突变(图29b,c)。最终,申请人测量β-半乳糖苷酶 活性以证实所有编辑的细胞都展现所期望的表型(图25d)。

[0309] Cas9介导的编辑也可用以产生多个突变用以研究生物通路。针对 锚定表面蛋白至革兰氏阳性菌的包膜的分选酶依赖性通路,申请人决定说明 这一点。申请人通过一个氯霉素抗性靶向构建体以及一个ΔsrtA编辑模板的共 转化而引入一个分选酶缺失(图33a,b),随后使用替代前述那个的一种卡 那霉素抗性靶向构建体引入一个ΔbgaA缺失。在肺炎链球菌中,β-半乳糖苷 酶通过分选酶共价连接至该细胞壁。因此,srtA的缺失导致表面蛋白被释放 至上清液中,而该双缺失不具有可检测到的β-半乳糖苷酶活性(图34c)。这 样的顺序筛选可根据需要重复许多次,直到产生多个突变。

[0310] 这两个突变也可在相同的时间引入。申请人设计一种包含两个间隔 子、一个匹配srtA以及另一个匹配bgaA的靶向构建体,并且用两个编辑模板 在相同的时间进行共转化(图25e)。转化体的遗传分析显示,在6/8的情况 下编辑发生(图25f)。值得注意地,剩下的两个克隆各自包含一个ΔsrtA或 一个ΔbgaA缺失,表明使用Cas9进行组合诱变的可能性。最终,为了消除该 CRISPR序列,申请人引入包含该bgaA靶标以及一个大观霉素抗性基因连同 来自野生型菌株R6的基因组DNA的一个质粒。保留该质粒的大观霉素抗性 转化体消除了该CRISPR序列(图34a,d)。

[0311] 编辑的机制和效率:为了理解使用Cas9进行的基因组编辑的潜在 机制,申请人设计一个试验,其中编辑效率是独立于Cas9切割而测量的。申 请人在该srtA基因座中整合该ermAM红霉素抗性基因,并且使用Cas9介导 的编辑引入一个早熟终止密码子(图33)。所得的菌株(JEN53)包含一个 ermAM(终止子)等位基因并且是红霉素敏感的。此菌株可被用于评估如下 效率,在此效率该ermAM基因通过测量在有或没有使用Cas9切割的情况下 恢复抗生素抗性的细胞的分数得以修复。JEN53是用一个恢复该野生型等位 基因的编辑模板转化的,连同一个靶向该ermAM(终止子)等位基因的卡那 霉素抗性CRISPR构建体(CRISPR::ermAM(终止子))或一个不具间隔子 的对照构建体(CRISPR::Ø)(图26a,b)。在不存在卡那霉素筛选的情况 下,编辑的菌落的分数近似于10^{-2}(红霉素抗性cfu/总cfu)(图26c),表 示在没有针对未编辑的细胞进行Cas9介导的筛选的情况下重组的基线频率。然而,如果应用卡那霉素筛选并且该对照CRISPR构建体被共转化,编辑的 菌落的分数增加至约10^{-1}(卡那霉素-以及红霉素抗性cfu/卡那霉素抗性cfu) (图26c)。此结果显示对针对独立于该基因

组的Cas9切割在该ermAM基因 座中重组的该CRISPR座位共筛选的重组的筛选，表明一个亚群的细胞更容 易转化和/或重组。该CRISPR::ermAM(终止子)构建体的转化随后是卡那霉素筛选导致红霉素抗性、编辑的细胞的分数增加至99％(图26c)。为了确定 此增加是否是由未编辑的细胞的杀死引起的，申请人比较了用该 CRISPR::ermAM(终止子)或CRISPR::Ø构建体共转化JEN53细胞之后获得 的这些卡那霉素抗性菌落形成单位(cfu)。

[0312] 申请人在该ermAM(终止子)构建体的转化之后对卡那霉素抗性 菌落进行的计数少了5.3倍($2.5\times10^4/4.7\times10^3$，图35a)，该结果表明Cas9 靶向一个染色体基因座确实导致未编辑的细胞的杀死。最终，因为已知在该 细菌染色体中引入dsDNA断裂引发增加受损的DNA的重组率的修复机制， 申请人研究了Cas9进行的切割是否诱导该编辑模板的重组。申请人用该 CRISPR::erm(终止)构建体共转化之后比用该CRISPR::Ø构建体的情况对菌落的计数多了2.2倍(图26d)，表明存在对重组的适度诱导。一起考虑，这 些结果显示可转化细胞的共筛选、通过Cas9介导的切割对重组的诱导以及针 对未编辑的细胞的筛选各自贡献肺炎链球菌中高效率的基因组编辑。

[0313] 因为Cas9对该基因组的切割会杀死未编辑的细胞，技术人员将不 期望恢复包含Cas9盒而无该编辑模板的接受卡那霉素抗性的任何细胞。然 而，在不存在该编辑模板的情况下，申请人在该CRISPR::ermAM(终止子) 构建体的转化之后恢复许多卡那霉素抗性菌落(图35a)。这些‘逃 脱’CRISPR诱导的死亡的细胞产生确定该方法的限制的一个背景。此背景频 率可被计算为在此试验中CRISPR::ermAM(终止子)/CRISPR::Øcfu，2.6×10^{-3}($7.1\times10^1/2.7\times10^4$)的比率，意味着如果该编辑模板的重组频率少于此 值，在此背景上，CRISPR筛选可能不会有效地恢复希望的突变体。为了理解 这些细胞的起源，申请人对8个背景菌落进行基因分型并且发现7个包含靶 向间隔子的缺失(图35b)并且一个在Cas9中包含一个假定钝化突变(图35c)。

[0314] 用Cas9在大肠杆菌中进行基因组编辑：仅可能在高重组发生的生 物体中通过CRISPR-Cas系统的染色体整合激活Cas9靶向。为了研发一个适 用于其他微生物的更通用的方法，申请人决定在大肠杆菌中使用一个基于质 粒的CRISPR-Cas系统进行基因组编辑。构建了两个质粒：携带该 tracrRNA、Cas9以及一个氯霉素抗性盒的一个pCas9质粒(图36)，以及携 带该CRISPR间隔子的阵列的一个pCRISPR卡那霉素抗性质粒。为了测量独 立于CRISPR筛选的编辑效率，申请人寻求在该rpsL基因中引入一个A至C 颠换，这赋予链霉素抗性。申请人构建包含一个间隔子的一个pCRISPR::rpsL 质粒，其将指导该野生型而不是突变的rpsL等位基因的Cas9切割(图27b)。首先将该pCas9质粒引入至大肠杆菌MG1655中，并且将所得的菌株 用该pCRISPR::rpsL质粒以及W542(包含该A至C突变的一个编辑寡核苷酸)共转化。在该pCRISPR::rpsL质粒的转化之后仅恢复了链霉素抗性菌落， 表明Cas9切割诱导该寡核苷酸的重组(图37)。然而，链霉素抗性菌落的数 目比卡那霉素抗性菌落的数目低两个量级，推测其是逃脱Cas9的切割的细 胞。因此，在这些条件下，Cas9的切割促进突变引入，但效率不足以在‘逃 脱者’的背景上选择突变细胞。

[0315] 为了改进大肠杆菌中的基因组编辑效率，申请人应用其CRISPR系 统进行重组工程化，使用Cas9诱导的细胞死亡以选择所希望的突变。该 pCas9质粒被引入至工程化菌株HME63(31)中，该菌株包含□-红色噬菌体 的Gam、Exo以及β功能。将所得的菌株用该pCRISPR::rpsL质粒(或一种 pCRISPR::Ø对照)以及该W542寡核苷酸共转化(图27a)。工

程化效率是 5.3×10^{-5},计算为当使用该对照质粒时变为链霉素抗性的总细胞的分数(图27c)。相比之下,用该pCRISPR::rpsL质粒转化增加突变细胞的百分数至 65%±14%(图27c以及29f)。申请人观察到,在该pCRISPR::rpsL质粒的 转化之后相比于该对照质粒,cfu的数目减少约三个数量级($4.8\times10^{5}/5.3\times 10^{2}$,图38a),表明起因于未编辑的细胞的CRISPR诱导的死亡的筛选。为 了测量Cas9切割被钝化的比率(申请人的方法的一个重要参数),申请人用 pCRISPR::rpsL或不具有该W542编辑寡核苷酸的对照质粒转化细胞(图38a)。CRISPR‘逃脱者’的背景(测量为pCRISPR::rpsL/pCRISPR::Øcfu的 比率),是2.5×10^{-4}($1.2\times10^{2}/4.8\times10^{5}$)。对这些逃脱者中的八个进行基因 分型揭示,在所有情况下,存在该靶间隔子的缺失(图38b)。此背景高于该rpsL突变的重组工程效率,5.3×10^{-5},表明为了获得65%的编辑的细胞, Cas9切割必须诱导寡核苷酸重组。为了证实这一点,申请人比较了在 pCRISPR::rpsL或pCRISPR::Ø的转化之后卡那霉素以及链霉素抗性cfu的数 目(图27d)。和肺炎链球菌的情况一样,申请人观察到对重组的适度诱导, 约6.7倍($2.0\times10^{-4}/3.0\times10^{-5}$)。一起考虑,这些结果表明,该CRISPR系统 提供用于通过重组工程化引入选择性突变的方法。

[0316] 申请人揭示,CRISPR-Cas系统可被用于在细菌中靶向通过共引入 杀死野生型细胞的一个靶向构建体以及消除CRISPR切割并引入所希望的突 变两者的一个编辑模板而编辑的基因组。可产生不同类型的突变(插入、缺 失或无疤单核苷酸置换)。多个突变可在相同的时间引入。使用该CRISPR 系统进行的编辑的特异性以及多样性依赖于该Cas9内切核酸酶的若干独特的 特性:(i)其靶特异性可被一个小RNA程序化,不需要酶工程,(ii)靶特异性是非常高的,由与低可能性的非靶识别相互作用的20bp RNA-DNA相互 作用确定,(iii)几乎可靶定任何序列,仅需要存在一个相邻的NGG序列, (iv)几乎该NGG序列中的任何突变,连同该原型间隔子的种子序列中的突 变都消除了靶向。

[0317] 申请人显示,使用该CRISPR系统进行的基因组工程化不仅在高度 引起重组的细菌诸如肺炎链球菌,也在大肠杆菌中起作用。大肠杆菌中的结 果表明,该方法可适用于质粒可被引入的其他微生物。在大肠杆菌中,该途 径补充了诱变寡核苷酸的重组工程化。为了在重组工程是不可能的微生物中 使用此方法,可通过在一个质粒上提供该编辑模板使用该宿主同源重组机 构。此外,因为积累的证据表明CRISPR介导的染色体的切割在许多细菌以 及古生菌中导致至细胞死亡,能够预想使用内源CRISPR-Cas系统用于编辑目 的。

[0318] 在肺炎链球菌以及大肠杆菌两者中,申请人观察到,尽管可转化细 胞的共筛选以及通过Cas9切割在该靶位点对重组的小诱导促进编辑,向编辑 贡献最多的机制是针对未编辑的细胞进行的筛选。因此,该方法的主要限制 是逃脱CRISPR诱导的细胞死亡并且缺少所希望的突变的细胞的背景的存 在。申请人显示,这些‘逃脱者’主要通过该靶向间隔子的缺失出现,大概在侧 翼于该靶向间隔子的重复序列重组之后。进一步改进可聚集于仍可支持功能 性crRNA的生源说但与彼此充分不同以消除重组的侧翼序列的工程化。可替 代地,可探索嵌合的crRNA的直接转化。在大肠杆菌的具体情况下,该 CRISPR-Cas系统的构建是不可能的,如果此生物体也被用作克隆宿主的话。 通过将Cas9以及该tracrRNA放置于一个不同于CRISPR阵列的质粒,申请人 解决了这个问题。可诱导系统的工程化也可避免此限制。

[0319] 尽管新的DNA合成技术提供成本有效地建立具有高通量的任何序 列的能力,在活

细胞中整合合成DNA以建立功能性基因组仍是一个挑战。最 近,显示共筛选MAGE策略通过筛选具有在给定的基因座处或周围实现重组 的增加的可能性的一个亚群的细胞来改进重组工程的突变效率。在此方法 中,引入可选择的突变,用以增加产生附近的不可选择的突变的机会。与此 策略提供的间接筛选相反,使用该CRISPR系统使得能够直接筛选所希望的突变并且以高效率恢复它。这些技术加上遗传工程师的工具箱,并且连同 DNA合成,它们可实质性地推进破译基因功能和操控生物体两方面的能力, 以用于生物技术的目的。两个其他研究也涉及哺乳动物基因组的CRISPR相 关的工程。预期这些crRNA指导的基因组编辑技术可广泛有用于基础和医药 科学。

[0320] 菌株和培养条件。肺炎链球菌菌株R6是由亚历山大托马斯 (Alexander Tomasz)博士提供。在前述研究中出现了菌株crR6。肺炎链球菌 的液体培养物生长于THYE培养基(30g/l Todd-Hewitt琼脂,5g/l酵母提取 物)。将细胞铺板于补充有5%去纤维蛋白的绵羊血液的胰胨豆胨琼脂 (TSA)上。当适合时,如下添加抗生素:卡那霉素(400μg/ml),氯霉素(5μg/ml),红霉素(1μg/ml)链霉素(100μg/ml)或大观霉素(100 μg/ml)。使用如之前所述的米勒测定(Miller assay)来测量β-半乳糖苷酶活 性。

[0321] 分别通过杰夫罗伯茨(Jeff Roberts)以及唐纳德考特(Donald Court)提供大肠杆菌菌株MG1655以及HME63(源自于MG1655,Δ(argF- lac)U169λcI857Δcro-bioA galK tyr 145UAG mutS<>amp)(31)。大肠杆 菌的液体培养物生长于LB培养基(Difco)中。当适合时,如下添加抗生 素:氯霉素(25μg/ml),卡那霉素(25μg/ml)以及链霉素(50μg/ml)。

[0322] 肺炎链球菌转化。如在前(23)所述制备感受态细胞。对于所有基 因组编辑转化,将细胞在冰上缓慢解冻并且重悬于10体积的M2培养基(补 充有100ng/ml的感受态刺激肽CSP1(40)),并且随后添加编辑构建体 (编辑构建体是以0.7ng/μl至2.5μg/ul之间的最终浓度添加到细胞中的)。 在添加2μl的靶向构建体之前将细胞在37℃孵育20min,并且然后在37℃ 孵育40min。将连续稀释的细胞铺板于适合的培养基以确定集落形成单位 (cfu)计数。

[0323] 大肠杆菌λ-红重组工程。针对所有重组工程实验使用菌株 HME63。制备重组工程细胞并且根据之前公开的方案(6)进行处理。简而言 之,使由从板上获得的单菌落接种的2ml过夜培养物(LB培养基)生长于 30℃。将该过夜培养物稀释100倍并伴随摇动(200rpm)生长于30℃直至 该OD_{600}是从0.4-0.5(大约3小时)。对于λ-红诱导,将该培养物转移至 42℃水浴以在200rpm摇动15min。恰在诱导之后,使该培养物旋流于冰-水 浆中并且在冰上冷冻5-10min。然后洗涤细胞并且根据该方案等分。为了电 转化,将50μl的细胞与1mM的无盐寡核苷酸(IDT)或100-150ng的质粒 DNA(通过QIAprep Spin Miniprep试剂盒,凯杰公司(Qiagen)制备)混 合。将细胞使用1mm基因脉冲发生器小池(Gene Pulser cuvette)(伯乐公司 (Bio-rad))在1.8kV进行电穿孔并且立即重悬于1ml的室温LB培养基 中。在铺板于具有适合的抗生素抗性的LB琼脂之前,将细胞在30℃恢复1- 2小时,并且在32℃孵育过夜。

[0324] 肺炎链球菌基因组DNA的制备。为了转化目的,使用Wizard基因 组DNA纯化试剂盒提取肺炎链球菌基因组DNA,遵循由制造商(普洛麦格 (Promega))提供的说明。为了基因分型的目的,使700ul的过夜肺炎链球 菌培养物沉淀,重悬于60ul的溶菌酶溶液(2mg/ml)中并且在37℃孵育30 min。使用QIAprep Spin Miniprep试剂盒(凯杰公司(Qiagen))提取基因组 DNA。

[0325] 菌株构建。在此研究中使用的所有引物提供在表G中。为了产生肺 炎链球菌crR6M,制备一种中间菌株,LAM226。在此菌株中,邻近于肺炎 链球菌crR6菌株的CRISPR阵列的aphA-3基因(提供卡那霉素抗性)被一个 cat基因(提供氯霉素抗性)替代。简而言之,分别使用引物L448/L444以及 L447/L481扩增crR6基因组DNA。使用引物L445/L446从质粒pC194扩增该 cat基因。每种PCR产物都是凝胶纯化的并且通过基因拼接(SOEing)PCR 用引物L448/L481融合所有三种。将所得的PCR产物转化至感受态肺炎链球 菌crR6细胞中,并且筛选氯霉素抗性转化体。为了产生肺炎链球菌crR6M, 通过PCR分别使用引物L409/L488以及L448/L481扩增肺炎链球菌crR6基因 组DNA。每种PCR产物都是凝胶纯化的并且通过基因拼接(SOEing)PCR 用引物L409/L481融合它们。将所得的PCR产物转化至感受态肺炎链球菌LAM226细胞中,并且筛选卡那霉素抗性转化体。

[0326] 为了产生肺炎链球菌crR6Rc,通过PCR使用引物L430/W286扩增 肺炎链球菌crR6M基因组DNA,并且通过PCR使用引物W288/L481扩增肺 炎链球菌LAM226基因组DNA。每种PCR产物都是凝胶纯化的并且通过基 因拼接(SOEing)PCR用引物L430/L481融合它们。将所得的PCR产物转化 至感受态肺炎链球菌crR6M细胞中,并且筛选氯霉素抗性转化体。

[0327] 为了产生肺炎链球菌crR6Rk,通过PCR分别使用引物L430/W286 以及W287/L481扩增肺炎链球菌crR6M基因组DNA。每种PCR产物都是凝 胶纯化的并且通过基因拼接(SOEing)PCR用引物L430/L481融合它们。将 所得的PCR产物转化至感受态肺炎链球菌crR6Rc细胞中,并且筛选卡那霉 素抗性转化体。

[0328] 为了产生JEN37,通过PCR分别使用引物L430/W356以及W357/L481扩增肺炎链球菌crR6Rk基因组DNA。每种PCR产物都是凝胶纯 化的并且通过基因拼接(SOEing)PCR用引物L430/L481融合它们。将所得 的PCR产物转化至感受态肺炎链球菌crR6Rc细胞中,并且筛选卡那霉素抗 性转化体。

[0329] 为了产生JEN38,分别使用引物L422/L461以及L459/L426扩增 R6基因组DNA。[43]使用引物L457/L458从质粒pFW15扩增ermAM基因(指 定红霉素抗性)。每种PCR产物都是凝胶纯化的并且通过基因拼接 (SOEing)PCR用引物L422/L426融合所有三种。将所得的PCR产物转化至 感受态肺炎链球菌crR6Rc细胞中,并且筛选红霉素抗性转化体。

[0330] 肺炎链球菌JEN53是以两个步骤产生的。首先如在图33中所示 的,构建JEN43。JEN53是通过将JEN25的基因组DNA转化至感受态JEN43 细胞中并且在氯霉素以及红霉素两者上进行筛选产生的。

[0331] 为了产生肺炎链球菌JEN62,通过PCR分别使用引物W256/W365 以及W366/L403扩增肺炎链球菌crR6Rk基因组DNA。将每种PCR产物纯化 并且通过吉布森拼接连接。将拼接产物转化至感受态肺炎链球菌crR6Rc细胞 中,并且筛选卡那霉素抗性转化体。

[0332] 质粒构建。pDB97是通过磷酸化作用以及对寡核苷酸B296/B297退 火,随后在由EcoRI/BamHI消化的pLZ12spec中连接构建的。申请人完全测 序了pLZ12spec并且将其序列保存于基因库(登记号:KC112384)。

[0333] pDB98是在克隆该CRISPR之后获得的,连同一个重复-间隔子-重 复单位一起将前导子序列克隆至pLZ12spec中。这是通过用引物B298/B320 以及B299/B321扩增crR6Rc DNA,随后进行两种产物的基因拼接PCR并且 以限制酶切位点BamHI/EcoRI克隆到pLZ12spec中实现的。以此方式, pDB98中的间隔子序列被工程化以包含两个处于相反方向

的BsaI限制酶切位 点,这允许无疤克隆新的间隔子。

[0334] pDB99至pDB108是通过退火寡核苷酸B300/B301(pDB99)、B302/B303(pDB100)、B304/B305(pDB101)、B306/B307(pDB102)、B308/B309(pDB103)、B310/B311(pDB104)、B312/B313(pDB105)、B314/B315(pDB106)、B315/B317(pDB107)、B318/B319(pDB108),随后在pDB98切口通过BsaI连接来构建的。

[0335] 该pCas9质粒是如下构建的。从具有供吉布森拼接(Gibson Assembly)的侧翼同源臂的化脓链球菌SF370基因组DNA扩增必需的 CRISPR元件。用寡核苷酸HC008以及HC010扩增该tracrRNA以及Cas9。HC011/HC014以及HC015/HC009扩增前导子以及CRISPR序列,以将两个 BsaI类型IIS位点引入在两个同向重复之间,以促进间隔子的方便插入。

[0336] pCRISPR是通过在pZE21-MCS1中通过用寡聚物B298+B299扩增 亚克隆pCas9 CRISPR阵列并用EcoRI以及BamHI限制来构建的。rpsL靶向 间隔子是如下克隆的,通过退火寡聚物B352+B353并且在该BsaI切口 pCRISPR进行克隆,给出pCRISPR::rpsL。

[0337] 靶向以及编辑构建体的产生。用于基因组编辑的靶向构建体是通过 左PCR和右PCR的吉布森拼接制备的(表G)。当可适用时编辑构建体是通 过基因拼接PCR融合PCR产物A(PCR A)、PCR产物B(PCR B)以及 PCR产物C(PCR C)制备的(表G)。CRISPR::Ø以及CRISPR::ermAM (终止子)靶向构建体是分别通过JEN62以及crR6基因组DNA(具有寡聚 物L409以及L481)的PCR扩增产生的。

[0338] 具有随机化的PAM或原型间隔子的靶标的产生。该间隔子1靶标 后的5个核苷酸通过用引物W377/L426扩增R6$^{8232.5}$基因组DNA被随机化。然后此PCR产物与用引物L422/W376从相同的模板扩增的cat基因以及srtA 上游区域拼接。80ng的拼接DNA被用以转化菌株R6以及crR6。用于该随 机化靶标的样品是使用以下引物制备的:B280-B290/L426以随机化该靶标的 碱基1-10以及B269-B278/L426以随机化碱基10-20。引物L422/B268以及 L422/B279被用以扩增分别待与最初以及最后10个PCR产物拼接的cat基因 以及srtA上游区域。将这些拼接构建体汇集在一起并且在R6以及crR6中转 化30ng。转化之后,将细胞铺板于氯霉素筛选板上。对于每个样品,超过2 $\times 10^5$个细胞一起被汇集在1ml的THYE中,并且用普洛麦格(Promega) Wizard试剂盒提取基因组DNA。引物B250/B251被用以扩增该靶区域。标记PCR产物并且使用300个循环在一个亿明达(Illumina)MiSeq双末端泳道上 跑胶。

[0339] 深度测序数据分析

[0340] 随机化PAM:对于随机化PAM试验,针对crR6获得3,429,406个 读数并且针对R6获得3,253,998个读数。期望它们中仅一半对应于该PAM- 靶标而另一半将对PCR产物的另一端进行测序。1,623,008个crR6读数以及 1,537,131个R6读数含有一个无误差靶序列。这些读数中每个可能的PAM的 发生示于补充文件中。为了评估PAM的功能性,计算crR6样品中与R6样品 中的相对比例并且表示为r_{ijklm},其中I、j、k、l、m是这4个可能碱基中的一 个。构建以下统计模型:

[0341] $\log(r_{ijklm}) = \mu + b2_i + b3_j + b4_k + b2b3_{i,j} + b3b4_{j,k} + \varepsilon_{ijklm}$,

[0342] 其中ε是残差,b2是PAM的第2个碱基的作用,b3是第三个碱基 的作用,b4是第四个碱基的作用,b2b3是第二和第三碱基之间的相互作用, b3b4是第三和第四碱基之间的相互作用。进行方差分析:

[0343] Anova表

	Df	平方和	均方	F值	Pr(>F)	
b3	3	151.693	50.564	601.8450	< 2.2e-16	***
b2	3	90.521	30.174	359.1454	< 2.2e-16	***
b4	3	1.881	0.627	7.4623	6.070e-05	***
b3:b2	9	228.940	25.438	302.7738	< 2.2e-16	***
b3:b4	9	3.010	0.334	3.9809	5.227e-05	***
残差	996	83.680	0.084			

[0344] 当添加到此模型时,b1或b5看起来不显著并且除包含的那些外的 其他相互作用也可舍弃。模型选择是使用R中的anova方法或多或少通过完 整模型的连续比较进行的。Tukey显著性检验(Tukey's honest significance test)被用以确定各影响之间成对发生的差异是否显著。

[0345] NGGNN型与所有其他型显著不同,并且具有最强影响(参见下 表)。

[0346] 为了显示位置1、4或5不影响该NGGNN型,申请人仅着眼于这 些序列。它们的影响看起来是正态分布的(参见图71中的QQ绘图),并且 使用R中的anova方法进行的模型比较显示空(null)模型是最好的一个, 即,不存在b1、b4以及b5的显著作用。

[0347] 针对该NGGNN序列使用R中的anova方法进行的模型比较

```
模型 1: 比率对数 ~ 1
模型 2: 比率对数 ~ b1 + b4 + b5
 Res.Df    RSS Df 平方和      F Pr(>F)
1     63 14.579
2     54 11.295  9  3.2836 1.7443 0.1013
```

[0348] NAGNN以及NNGGN型的部分干扰

[0349] NAGNN型与所有其他型显著不同但具有比NGGNN小得多的影响 (参见以下的Tukey显著性检验)。

[0350] 最终,NTGGN以及NCGGN型是类似的并且显示比NTGHN以及 NCGHN(其中H是A、T或C)型显著更多的CRISPR干扰,如通过邦费罗 尼(bonferroni)调节的成对学生检验显示的。

[0351] 使用t检验以汇集的SD成对比较b4对NYGNN序列的影响

数据 : b4

	A	C	G
C	1.00	-	-
G	9.2e-05	2.4e-06	-
T	0.31	1.00	1.2e-06

[0352] 一起考虑,这些结果允许推断,通常NNGGN型在NGGGN的情况 下产生完全干扰,或在NAGGN、NTGGN或NCGGN的情况下产生部分干 扰。

[0353] 平均值的Tukey多重比较:95%多重比较(family-wise)置信水平

$b2:b3

	diff	lwr	upr	p adj
G:G-A:A	-2.76475	-2.94075	-2.58875	<1E -07
G:G-C:A	-2.79911	-2.97511	-2.62311	<1E -07
G:G-T:A	-2.7809	-2.9569	-2.6049	<1E -07
G:G-A:C	-2.81643	-2.99244	-2.64043	<1E -07
G:G-C:C	-2.77903	-2.95504	-2.60303	<1E -07
G:G-G:C	-2.64867	-2.82468	-2.47267	<1E -07
G:G-T:C	-2.79718	-2.97319	-2.62118	<1E -07
G:G-A:G	-2.67068	-2.84668	-2.49468	<1E -07
G:G-C:G	-2.73525	-2.91125	-2.55925	<1E -07
G:G-T:G	-2.7976	-2.62159	-2.9736	<1E -07
G:G-A:T	-2.76727	-2.59127	-2.94328	<1E -07
G:G-C:T	-2.84114	-2.66513	-3.01714	<1E -07
G:G-G:T	-2.76409	-2.58809	-2.94009	<1E -07
G:G-T:T	-2.76781	-2.59181	-2.94381	<1E -07
G:G-G:A	-2.13964	-2.31565	-1.96364	<1E -07
G:A-A:A	-0.62511	-0.80111	-0.4491	<1E -07
G:A-C:A	-0.65947	-0.83547	-0.48346	<1E -07
G:A-T:A	-0.64126	-0.46525	-0.81726	<1E -07
G:A-A:C	-0.67679	-0.50078	-0.85279	<1E -07
G:A-C:C	-0.63939	-0.46339	-0.81539	<1E -07
G:A-G:C	-0.50903	-0.33303	-0.68503	<1E -07
G:A-T:C	-0.65754	-0.48154	-0.83354	<1E -07
G:A-A:G	-0.53104	-0.35503	-0.70704	<1E -07
G:A-C:G	-0.59561	-0.4196	-0.77161	<1E -07
G:A-T:G	-0.65795	-0.48195	-0.83396	<1E -07
G:A-A:T	-0.62763	-0.45163	-0.80363	<1E -07
G:A-C:T	-0.70149	-0.52549	-0.8775	<1E -07
G:A-G:T	-0.62445	-0.44844	-0.80045	<1E -07
G:A-T:T	-0.62817	-0.45216	-0.80417	<1E -07

$b3:b4

	diff	lwr	upr	p adj
G:G-G:A	-0.33532	-0.51133	-0.15932	<1E-07
G:G-G:C	-0.18118	-0.35719	-0.00518	0.036087
G:G-G:T	-0.31626	-0.14026	-0.49226	<1E-07

[0354] 随机化靶标

[0355] 对于随机化靶标试验,针对crR6获得540,726个读数并且针对R6 获得753,570个读数。如前所述,仅预期这些读数中一半对该PCR产物的感 兴趣的端进行测序。在过滤之后,对于携带一个靶标的无误差或具有单个点 突变的读数针对crR6以及R6分别保留217,656个和353,141个。计算了该 crR6样品中每个突变体与该R6样品中的相对比例(图24c)。该种子序列 (距离该PAM 13-20个碱基)外的所有突变显示完全干扰。那些序列被用作 一个参考以确定该种子序列内的其他突变是否可以被认为是显著破坏干扰。使用MASS R软件包的fitdistr函数对这些序列拟合正态分布。在图24c中, 该拟合的分布的0.99分位点显示为一条虚线。图72显示具有拟合正态分布 (黑色线)以及.99分位数(虚线)的数据密度的直方图。

[0356] 表F.crR6/R6样品中PAM序列的相对丰度(经碱基1和5进行平 均)。

第2位置	第3位置 A		第3位置 C		第3位置 G		第3位置 T		第4位置
A	AAA	1.04	ACA	1.12	AGA	0.73	ATA	1.10	A
	AAC	1.07	ACC	1.04	AGC	0.66	ATC	0.97	C
	AAG	1.00	ACG	1.09	AGG	0.61	ATG	1.07	G
	AAT	0.98	ACT	1.02	AGT	0.65	ATT	1.01	T
C	CAA	1.05	CCA	1.05	CGA	0.99	CTA	1.07	A
	CAC	1.04	CCC	1.02	CGC	1.08	CTC	1.04	C
	CAG	1.08	CCG	1.08	CGG	0.81	CTG	1.05	G
	CAT	1.13	CCT	1.05	CGT	1.07	CTT	1.08	T
G	GAA	0.97	GCA	1.05	GGA	0.68	GTA	0.99	A
	GAC	0.92	GCC	1.00	GGC	0.65	GTC	1.15	C
	GAG	0.96	GCG	0.98	GGG	0.67	GTG	0.98	G
	GAT	0.98	GCT	0.99	GGT	0.66	GTT	1.05	T
T	TAA	1.08	TCA	1.15	TGA	1.05	TTA	1.14	A
	TAC	1.00	TCC	1.08	TGC	1.08	TTC	1.05	C
	TAG	1.02	TCG	1.11	TGG	0.77	TTG	1.01	G
	TAT	1.01	TCT	1.12	TGT	1.21	TTT	1.02	T

[0357] 表G.这项研究中使用的引物。

引物	序列 5'-3'
B217	TCCTAGCAGGAATTCTGATATTACTGTCACGTTTTAGAGCTATGCTGTTTTGA
B218	GTGACAGTAATATCAGAAATCCTGCTAGGAGTTTTGGGACCATTCAAAACAGC
B229	GGGTTTCAAGTCTTTGTAGCAAGAG
B230	GCCAATGAACGGGAACCCTTGGTC
B250	NNNNGACGAGGCAATGGCTGAAATC
B251	NNNNTTATTTGGCTCATATTTGCTG
B255	CTTTACACCAATCGCTGCAACAGAC
B256	CAAAATTTCTAGTCTTCTTTGCCTTTCCCCATAAAACCCTCCTTA
B257	AGGGTTTTATGGGGAAAGGCAAAGAAGACTAGAAATTTTGATACC
B258	CTTACGGTGCATAAAGTCAATTTCC
B269	TGGCTCGATTTCAGCCATTGC
B270	CTTTGACGAGGCAATGGCTGAAATCGAGCCAANAAAGCGCAAG
B271	CTTTGACGAGGCAATGGCTGAAATCGAGCCAAANAAGCGCAAG
B272	CTTTGACGAGGCAATGGCTGAAATCGAGCCAAAANAGCGCAAG
B273	CTTTGACGAGGCAATGGCTGAAATCGAGCCAAAAANGCGCAAG
B274	CTTTGACGAGGCAATGGCTGAAATCGAGCCAAAAAANCGCAAG
B275	CTTTGACGAGGCAATGGCTGAAATCGAGCCAAAAAAGNGCAAG
B276	CTTTGACGAGGCAATGGCTGAAATCGAGCCAAAAAAGCNCAAGAAG
B277	CTTTGACGAGGCAATGGCTGAAATCGAGCCAAAAAAGCGNAAGAAG
B278	CTTTGACGAGGCAATGGCTGAAATCGAGCCAAAAAAGCGCNAGAAG
B279	GCGCTTTTTTGGCTCGATTTCAG
B280	CAATGGCTGAAATCGAGCCAAAAAAGCGCANGAAGAAATC
B281	CAATGGCTGAAATCGAGCCAAAAAAGCGCAANAAGAAATC
B282	CAATGGCTGAAATCGAGCCAAAAAAGCGCAAGNAGAAATC
B283	CAATGGCTGAAATCGAGCCAAAAAAGCGCAAGANGAAATC
B284	CAATGGCTGAAATCGAGCCAAAAAAGCGCAAGAANAAATC
B285	CAATGGCTGAAATCGAGCCAAAAAAGCGCAAGAAGNAATCAACC
B286	CAATGGCTGAAATCGAGCCAAAAAAGCGCAAGAAGANATCAACC
B287	CAATGGCTGAAATCGAGCCAAAAAAGCGCAAGAAGAANTCAACC
B288	CAATGGCTGAAATCGAGCCAAAAAAGCGCAAGAAGAAANCAACC
B289	CAATGGCTGAAATCGAGCCAAAAAAGCGCAAGAAGAAATNAACCAGC
B290	CAATGGCTGAAATCGAGCCAAAAAAGCGCAAGAAGAAATCNACCAGC
B296	gaatccTCCATCCGTACAACCCACAACCCTGg
B297	aattcCAGGGTTGTGGGTTGTACGGATGGAg
B298	CATCGATCCTATTTCTTAATAACTAAAAATATGG
B299	CATGAATTCAACTCAACAAGTCTCAGTGTGCTG
B300	AAACATTTTTTCTCCATTTAGGAAAAAGGATGCTG
B301	AAAACAGCATCCTTTTTCCTAAATGGAGAAAAAAT

B302	AAACCTTAAATCAGTCACAAATAGCAGCAAAATTG
B303	AAAACAATTTTGCTGCTATTTGTGACTGATTTAAG
B304	AAACTTTTCATCATACGACCAATCTGCTTTATTTG
B305	AAAACAAATAAAGCAGATTGGTCGTATGATGAAAA
B306	AAACTCGTCCAGAAGTTATCGTAAAAGAAATCGAG
B307	AAAACTCGATTTCTTTTACGATAACTTCTGGACGA
B308	AAACAATCTCTCCAAGGTTTCCTTAAAAATCTCTG
B309	AAAACAGAGATTTTTAAGGAAACCTTGGAGAGATT
B310	AAACGCCATCGTCAGGAAGAAGCTATGCTTGAGTG
B311	AAAACACTCAAGCATAGCTTCTTCCTGACGATGGC
B312	AAACATCTCTATACTTATTGAAATTTCTTTGTATG
B313	AAAACATACAAAGAAATTTCAATAAGTATAGAGAT
B314	AAACTAGCTGTGATAGTCCGCAAAACCAGCCTTCG
B315	AAAACGAAGGCTGGTTTTGCGGACTATCACAGCTA
B316	AAACATCGGAAGGTCGAGCAAGTAATTATCTTTTG
B317	AAAACAAAAGATAATTACTTGCTCGACCTTCCGAT
B318	AAACAAGATGGTATCGCAAAGTAAGTGACAATAAG
B319	AAAACTTATTGTCACTTACTTTGCGATACCATCTT
B320	GAGACCTTTGAGCTTCCGAGACTGGTCTCAGTTTTGGGACCATTCAAAACAG
B321	TGAGACCAGTCTCGGAAGCTCAAAGGTCTCGTTTTAGAGCTATGCTGTTTTG
B352	aaacTACTTTACGCAGCGCGGAGTTCGGTTTTTTg
B353	aaaacAAAAAACCGAACTCCGCGCTGCGTAAAGTA
HC008_SP	ATGCCGGTACTGCCGGGCCTCTTGCGGGATTACGAAATCATCCTG
HC009_SP	GTGACTGGCGATGCTGTCGGAATGGACGATCACACTACTCTTCTT
HC010_SP	TTAAGAAATAATCTTCATCTAAAATATACTTCAGTCACCTCCTAGCTGAC
HC011_SP	ATTGATTTGAGTCAGCTAGGAGGTGACTGAAGTATATTTTAGATGAAG
HC014_SP	GAGACCTTTGAGCTTCCGAGACTGGTCTCAGTTTTGGGACCATTCAAAACAGCATAGCTCTAAAACCTCGTAGACTA TTTTTGTC
HC015_SP	GAGACCAGTCTCGGAAGCTCAAAGGTCTCGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACTTCAGCACACTG AGACTTG
L403	AGTCATCCCAGCAACAAATGG
L409	CGTGGTAAATCGGATAACGTTCCAAGTGAAG
L422	Tgctcttcttcacaaacaaggg
L426	AAGCCAAAGTTTGGCACCACC
L430	GTAGCTTATTCAGTCCTAGTGG
L444	CGTTTGTTGAACTAATGGGTGCAAATTACGAATCTTCTCCTGACG
L445	CGTCAGGAGAAGATTCGTAATTTGCACCCATTAGTTCAACAAACG
L446	GATATTATGGAGCCTATTTTTGTGGGTTTTTAGGCATAAAACTATATG
L447	CATATAGTTTTATGCCTAAAAACCcACAAAAATAGGCTCCATAATATC
L448	ATTATTTCTTAATAACTAAAAATATGG
L457	CGTgtacaattgctagcgtacggc
L458	GCACCGGTGATCACTAGTCCTAGG

L459	cctaggactagtgatcaccggtGCAAATATGAGCCAAATAAATATAT
L461	GCCGTACGCTAGCAATTGTACACGTTTGTTGAACTAATGGGTGC
L481	TTCAAATTTTCCCATTTGATTCTCC
L488	CCATATTTTTAGTTATTAAGAAATAATACCAGCCATCAGTCACCTCC
W256	AGACGATTCAATAGACAATAAGG
W286	GTTTTGGGACCATTCAAAACAGCATAGCTCTAAAACCTCGTAGAC
W287	GCTATGCTGTTTTGAATGGTCCCAAAACcattattttaacacacgaggtg
W288	GCTATGCTGTTTTGAATGGTCCCAAAACGCACCCATTAGTTCAACAAACG
W326	AATTCTTTTCTTCATCATCGGTC
W327	AAGAAAGAATGAAGATTGTTCATG
W341	GGTACTAATCAAAATAGTGAGGAGG
W354	GTTTTTCAAAATCTGCGGTTGCG
W355	AAAAATTGAAAAAATGGTGGAAACAC
W356	ATTTCGTAAACGGTATCGGTTTCTTTTAAAGTTTTGGGACCATTCAAAACAGC
W357	TTTAAAAGAAACCGATACCGTTTACGAAATGTTTTAGAGCTATGCTGTTTTGA
W365	AAACGGTATCGGTTTCTTTTAAATTCAATTGTTTTGGGACCATTCAAAACAGC
W366	AATTGAATTTAAAAGAAACCGATACCGTTTGTTTTAGAGCTATGCTGTTTTGA
W370	GTTCCTTAAACCAAAACGGTATCGGTTTCTTTTAAATTC
W371	GAAACCGATACCGTTTTGGTTTAAGGAACAGGTAAAGGGCATTTAAC
W376	CGATTTCAGCCATTGCCTCGTC
W377	GCCTTTGACGAGGCAATGGCTGAAATCGNNNNNAAAAAGCGCAAGAAGAAATCAAC
W391	TCCGTACAACCCACAACCCTGCTAGTGAGCGTTTTGGGACCATTCAAAACAGC
W392	GCTCACTAGCAGGGTTGTGGGTTGTACGGAGTTTTAGAGCTATGCTGTTTTGA
W393	TTGTTGCCACTCTTCCTTCTTTC
W397	CAGGGTTGTGGGTTGTTGCGATGGAGTTAACTCCCATCTCC
W398	GGGAGTTAACTCCATCGCAACAACCCACAACCCTGCTAGTG
W403	GTGGTATCTATCGTGATGTGAC
W404	TTACCGAAACGGAATTTATCTGC
W405	AAAGCTAGAGTTCCGCAATTGG
W431	GTGGGTTGTACGGATTGAGTTAACTCCCATCTCCTTC
W432	GATGGGAGTTAACTCAATCCGTACAACCCACAACCCTG
W433	GCTTCACCTATTGCAGCACCAATTGACCACATGAAGATAG
W434	GTGGTCAATTGGTGCTGCAATAGGTGAAGCTAATGGTGATG
W463	CTGATTTGTATTAATTTTGAGACATTATGCTTCACCTTC
W464	GCATAATGTCTCAAAATTAATACAAATCAGTGAAATCATG
W465	GTTTTGGGACCATTCAAAACAGCATAGCTCTAAAACGTGACAGTAATATCAG
W466	GTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACGCTCACTAGCAGGGTTG
W542	ATACTTTACGCAGCGCGGAGTTCGGTTTTgTAGGAGTGGTAGTATATACACGAGTACAT

[0358] 表H.在此研究中使用的靶向以及编辑构建体的设计

	靶向构建体				
编辑	模板 DNA	左 PCR	右 PCR	间隔子序列	PAM
bgaA R>A	crR6Rk	W256/W391	W392/L403	GCTCACTAGCAGGGTTGTGGGTTGTACGGA	TGG
bgaA NE>AA	crR6Rk	W256/W391	W392/L403	GCTCACTAGCAGGGTTGTGGGTTGTACGGA	TGG
ΔbgaA	crR6Rk	W256/W391	W392/L403	GCTCACTAGCAGGGTTGTGGGTTGTACGGA	TGG
ΔsrtA	crR6Rc	W256/B218	B217/L403	TCCTAGCAGGATTTCTGATATTACTGTCAC	TGG
ermB 终止子	crR6Rk	W256/W356	W357/L403	TTTAAAAGAAACCGATACCGTTTACGAAAT	TGG
ΔsrtA ΔbgaA	JEN51 (用于左 PCR) 以及 JEN52 (用于右 PCR)	W256/W465	W466/W403	与针对 ΔsrtA 和 ΔbgaA 使用的相同	TGG

编辑	编辑构建体 模板 DNA	PCR A	PCR B	PCR C	基因拼接 PCR	所得菌株的名称	用于验证编辑的基因型的引物
bgaA R>A	R6	W403/W397	W398/W404	N/A	W403/W404	JEN56	W403/W404
bgaA NE>AA	R6	W403/W431	W432/W433	W434/W404	W403/W404	JEN60	W403/W404
ΔbgaA	R6	B255/B256	B257/B258	N/A	B255/B258	JEN52	W393/W405
ΔsrtA	R6	B230/W463	W464/B229	N/A	B230/B229	JEN51	W422/W426
ermB 终止子 ΔsrtA ΔbgaA	JEN38	L422/W370	W371/L426	N/A	L422/L426	JEN43	L457/L458
	与针对 ΔsrtA 和 ΔbgaA 使用的相同					JEN64	与针对 ΔsrtA 和 ΔbgaA 使用的相同

实例6:对化脓链球菌Cas9(称为SpCas9)的指导RNA的优化。

[0359] 申请人使该tracrRNA以及同向重复序列突变,或使该嵌合指导 RNA突变以增强细胞中的RNA。

[0360] 该优化是基于以下观察,在该tracrRNA以及指导RNA中存在胸腺 嘧啶(T)的延伸,这能够通过pol 3启动子导致早期转录终止。因此申请人 生成以下优化序列。优化的tracrRNA以及对应的优化的同向重复是成对显示 的。

[0361] 优化的tracrRNA 1(突变加下划线):

[0362] GGAACCATTCAtAACAGCATAGCAAGTTAtAATAAGGCTAGTCCGTT ATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTT

[0363] 优化的同向重复1(突变加下划线):

[0364] GTTaTAGAGCTATGCTGTTaTGAATGGTCCCAAAAC

[0365] 优化的tracrRNA 2(突变加下划线):

[0366] GGAACCATTCAAtACAGCATAGCAAGTTAAtATAAGGCTAGTCCGTT ATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTT

[0367] 优化的同向重复2(突变加下划线):

[0368] GTaTTAGAGCTATGCTGTaTTGAATGGTCCCAAAAC

[0369] 为了真核细胞中的最佳活性,申请人还优化了嵌合指导RNA。

[0370] 原始的指导RNA:

[0371] NNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAGT TAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGG TGCTTTTTTT

[0372] 优化的嵌合指导RNA序列1:

[0373] NNNNNNNNNNNNNNNNNNNNGTATTAGAGCTAGAAATAGCAAGT TAATATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGG TGCTTTTTTT

[0374] 优化的嵌合指导RNA序列2:

[0375] NNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTATGCTGTTTTGGAA ACAAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGA AAAAGTGGCACCGAGTCGGTGCTTTTTTT

[0376] 优化的嵌合指导RNA序列3:

[0377] NNNNNNNNNNNNNNNNNNNNGTATTAGAGCTATGCTGTATTGGAA ACAATACAGCATAGCAAGTTAATATAAGGCTAGTCCGTTATCAACTTGA AAAAGTGGCACCGAGTCGGTGCTTTTTTT

[0378] 申请人揭示,优化的嵌合指导RNA更好地起作用,如在图3中表 明的。通过用Cas9以及一个U6-指导RNA DNA盒共转染293FT细胞进行该 试验,以表达以上所示的四个RNA形式之一。该指导RNA的靶标在人 EMX1座位中是同一个靶位点:“GTCACCTCCAATGACTAGGG”

实例7:嗜热链球菌LMD-9CRISPR1 Cas9(称为St1Cas9)的优化。

[0379] 申请人设计如图4中所示的指导嵌合RNA。

[0380] 通过断裂多胸腺嘧啶(T)进行延伸,该St1Cas9指导RNA可经受 如针对SpCas9指导RNA进行的相同类型的优化。

实例8:Cas9多样性以及突变

[0381] CRISPR-Cas系统是一种由跨细菌和古生菌的相异物种利用的针对 侵入型外源DNA的获得性免疫机制。II型CRISPR-Cas9系统由一套编码负责 将外源DNA“捕获(acquisition)”进CRISPR座位中的蛋白质的基因以及一 套编码负责“执行(execution)”DNA切割机制的蛋白质的基因组成;这些 包括DNA核酸酶(Cas9)、非编码反式激活cr-RNA(tracrRNA)以及侧翼 为同向重复的外源DNA衍生的间隔子阵列(crRNA)。当被Cas9成熟后,tracRNA和crRNA双链体指导Cas9核酸酶到由间隔子指导序列规定的靶 DNA序列,并且在靶DNA中介导短序列基序附近的DNA中的双链断裂,该 导短序列基序为切割所需并且对于每种CRISPR-Cas系统而言是特异的。II型 CRISPR-Cas系统发现在整个细菌界中并且在用于靶标切割的Cas9蛋白序列 和尺寸、tracrRNA和crRNA同向重复序列、这些元件的基因组组织以及基序 要求方面高度相异。一个物种可以具有多种相异的CRISPR-Cas系统。

[0382] 申请人评价了来自基于与已知Cas9的序列同源性和与已知亚结构 域直向同源的结构鉴定的细菌种类的207个假定的Cas9,这些亚结构域包括 HNH内切核酸酶结构域和RuvC内切核酸酶结构域[信息来自尤金库宁 (Eugene Koonin)和基拉马卡洛娃(Kira Makarova)]。基于此套的蛋白质序 列保守的系统发生分析揭示了五个Cas9家族,包括三组大的Cas9(约1400 个氨基酸)和两组小的Cas9(约1100个氨基酸)(参见图39和40A-F)。

[0383] 在此实例中,申请人显示以下突变可以将SpCas9转化为切口酶:D10A、E762A、H840A、N854A、N863A、D986A。

[0384] 申请人提供了显示出序列突变点位于SpCas9基因内何处的序列 (图41)。申请人还揭示,这些切口酶仍能够介导同源重组(图2中指出的 测定)。此外,申请人揭示,具有这些突变的SpCas9(单独地)不诱导双链 断裂(图47)。

实例9:对RNA指导的Cas9核酸酶的DNA靶向特异性的补充

[0385] 细胞培养和转染

[0386] 在37℃下,伴随5%CO_2孵育,将人类胚肾(HEK)细胞系293FT (生命技术公司)维持在补充有10%胎牛血清(海可隆公司(HyClone))、2mM GlutaMAX(生命技术公司)、100U/mL青霉素及100μg/mL链霉素的 杜尔贝科改良伊格尔培养基(Dulbecco's modified Eagle's Medium) (DMEM)中。

[0387] 在转染之前24小时将293FT细胞种到6孔板、24孔板、或96孔板 (康宁公司(Corning))上。遵循制造商推荐的方案,使用Lipofectamine 2000(生命技术公司)以80%-90%融合转染细胞。对于6孔板的每个孔,使 用总共1ug的Cas9+sgRNA质粒。对于24孔板的每个孔,使用总共500ng Cas9+sgRNA质粒,除非另外指明。对于96孔板的每个孔,以与U6-sgRNA PCR产物1:1摩尔比率使用65ng的Cas9质粒。

[0388] 人胚胎干细胞系HUES9(哈佛干细胞研究所中心(Harvard Stem Cell Institute core))维持在补充有100ug/ml Normocin(InvivoGen)的 mTesR培养基(干细胞技术(Stemcell Technologies))中GelTrex(生命技 术公司)上的无进料器的条件中。遵循制造商方案用Amaxa P3原代细胞4-D 核转染仪试剂盒(Lonza)转染HUES9细胞。

[0389] 用于基因组修饰的SURVEYOR核酸酶测定

[0390] 如上所述,用质粒DNA转染293FT细胞。转染后,在基因组DNA 提取之前,将细胞在37℃下孵育72小时。遵循制造商的方案,使用 QuickExtract DNA提取溶液(Epicentre)提取基因组DNA。简言之,将丸状 的细胞重悬于QuickExtract溶液中并在65℃下孵育15分钟并且在98℃下孵 育10分钟。

[0391] 对于每个基因而言,PCR扩增(引物列于表J和K中)在CRISPR 靶位点侧翼的基因组区域,并且遵循制造商的方案,使用QiaQuick Spin柱 (凯杰公司(Qiagen))纯化产物。将总共400ng的纯化的PCR产物与2μl 10X Taq DNA聚合酶PCR缓冲液(Enzymatics公司)和超纯水混合,至终体 积为20μl,并使其经受重退火过程,以使得可以形成异源双链体:95℃持续10min,以-2℃/s从95℃降至85℃,以-0.25℃/s从85℃降至25℃,并且在 25℃保持1分钟。重退火后,遵循制造商推荐的方案,将产物用 SURVEYOR核酸酶和SURVEYOR增强子S(转基因组学公司(Transgenomics))处理,并且在4%-20%Novex TBE聚丙烯酰胺凝胶(生 命技术公司)上进行分析。将凝胶用SYBR Gold DNA染色剂(生命技术公 司)染色30分钟并用Gel Doc凝胶成像系统(伯乐公司(Bio-rad))成像。基于相对条带强度进行量化。

[0392] 人细胞中的tracrRNA表达的Northern印迹分析

[0393] 如前所述进行Northern印迹。1简言之,在装载于8%变性多丙烯 酰胺凝胶(SequaGel,国立诊断公司(National Diagnostics))上之前,将 RNA加热至95℃持续5min。之后,将RNA转移至一种预杂交的Hybond N+膜(通用电气医疗公司(GE Healthcare))并且与Stratagene UV交联剂 (Stratagene)交联。将探针用具有T4多核苷酸激酶(新英格兰生物实验室 (New England Biolabs)))的[γ-32P]ATP(珀金埃尔默公司)标记。在洗涤 后,将膜暴露于荧光屏一小时并且用感光成像仪(Typhoon)扫描。

[0394] 亚硫酸氢盐测序以评估DNA甲基化状态

[0395] 如上所述,用Cas9转染HEK 293FT细胞。将基因组DNA以 DNeasy血液&组织试剂盒(凯杰公司(Qiagen))分离并且用EZ DNA甲基 化闪电试剂盒(Methylation-Lightning Kit)(Zymo研究公司)进行亚硫酸氢 盐转化。使用KAPA2G Robust HotStart DNA聚合酶(KAPA生物系统)进行 亚硫酸氢盐PCR,其中使用亚硫酸氢盐引物探索者(Bisulfite Primer Seeker) (Zymo研究公司,表J以及K)设计引物。将所得的PCR扩增子进行凝胶纯化,用EcoRI以及HindIII消化,并且在转化之前连接至pUC19骨架中。然后 将个体克隆进行桑格测序以评估DNA甲基化状态。

[0396] 体外转录以及切割测定

[0397] 如上所述,用Cas9转染HEK 293FT细胞。然后用补充有蛋白酶抑 制剂混合液(罗氏公司(Roche))的一种裂解缓冲液(20mM HEPES,100 mM KCl,5mM MgCl2,1mM DTT,5%甘油,0.1%Triton X-100)制备全细 胞裂解液。遵循制造商建议的方案,使用自定义的寡聚物(实例10)以及 HiScribe T7体外转录试剂盒(NEB)体外转录T7驱动的sgRNA。为了制备 甲基化靶位点,将pUC19质粒通过M.SssI甲基化并且然后通过NheI线性 化。如下进行体外切割测定:对于一种20uL切割反应,将10uL的细胞裂解 液用2uL切割缓冲液(100mM HEPES,500mM KCl,25mM MgCl2,5mM DTT,25%甘油)、体外转录的RNA、以及300ng pUC19质粒DNA孵育。

[0398] 深度测序以评估靶向特异性

[0399] 在基因组DNA提取之前,将铺板于96孔板的HEK 293FT细胞用 Cas9质粒DNA以及单个指导RNA(sgRNA)PCR盒转染72小时(图 72)。通过一种融合PCR方法以将亿明达P5适配子连同独特的样品特异性 条形码附接至该靶扩增子,针对每个基因的侧翼CRISPR靶位点的基因组区 域进行扩增(图74,(实例10)(图73中示意性描述)。遵循制造商建议 的方案,将PCR产物使用EconoSpin96孔滤板(时代生命科学公司(Epoch Life Sciences))纯化。

[0400] 将带条形码并纯化的DNA样品通过Quant-iT PicoGreen dsDNA测 定试剂盒或Qubit 2.0荧光计(生命技术公司)定量并且以等摩尔比率汇集。然后用亿明达MiSeq个人测序仪(生命技术公司)对测序文库进行深度测 序。

[0401] 测序数据分析以及indel检测

[0402] 将MiSeq读数通过至少23的平均Phred质量(Q得分)连同与条 形码以及扩增子正向引物匹配的完美序列过滤。通过首先针对包括靶位点 (总共120bp)的上游以及下游的50个核苷酸的扩增子序列进行史密斯-沃特 曼(Smith-Waterman)比对,分析来自中靶(on-target)以及脱靶(off- target)基因座的读数。同时,针对从靶位点(总共30bp)的上游5个核苷酸 到下游5个核苷酸的indel来分析比对。如果其比对的部分落到该MiSeq读数 本身外,或如果匹配的碱基对包括少于其总长的85%,则所分析的靶区域被 舍弃。

[0403] 针对每个样品的阴性对照提供了将indel的包含或排除作为假定的 切割事件的一个量规。针对每个样品,仅在如果其质量得分超过μ-σ的时候 对一个indel进行计数,其中μ是相应于该样品的该阴性对照的平均质量得 分,并且σ是其标准差。这产生针对阴性对照及其对应样品两者的整个靶区域 indel率。使用该阴性对照的每靶区域每读数(per-target-region-per-read)误差 率,q,该样品的观察到的indel计数n,及其读数计数R(即具有真实indel的 靶区域的读数分数的极大似然估计)p是通过应用一个二项误差模型导出的, 如下所示。

[0404] 使一个具有不正确地计数为具有至少1个indel的靶区域的样品中 的读数的(未知)数目是E,我们可将其书写为(不对真实indel的数目做任 何假定)

$$\mathrm{Prob}(E|p) = \binom{R(1-p)}{E} q^{E}(1-q)^{R(1-p)-E}$$

[0405] 因为R(1-p)是具有无真实indel的靶区域的读数的数目。同时,因 为观察到具有indel的读数的数目是n,n=E+Rp,换句话说,有误差但无真 实indel的靶区域的读数的数目加上靶区域正确地具有indel的读数的数目。然后我们可重写以上等式为

$$\mathrm{Prob}(E|p) = \mathrm{Prob}(n = E + Rp|p) = \binom{R(1-p)}{n-Rp} q^{n-Rp}(1-q)^{E-n}$$

[0406] 使具有真实indelp的靶区域的频率的所有值与先验同等可能, Prob(n|p)∝Prob(p|n)。因此将对具有真实indel的靶区域的频率的极大似然估计(MLE)设为最大化Prob(n|p)的p的值。在数字上对此进行评价。

[0407] 为了对测序文库本身中的真实indel读数频率设置误差界限,考虑 到针对真实indel靶区域Rp的MLE估计,以及读数的数目R,针对每个样品计 算威尔逊(Wilson)得分间隔(2)。明确地,将下限l以及上限u计算为

$$l = \left(Rp + \frac{z^2}{2} - z\sqrt{Rp(1-p) + z^2/4}\right)/(R + z^2)$$

$$u=\left(Rp+\frac{z^2}{2}+z\sqrt{Rp(1-p)+z^2/4}\right)/(R+z^2)$$

[0408] 其中z,方差1的正态分布中需要的置信度的标准得分被设置为 1.96,意味着95%的置信度。

[0409] 相对Cas9以及sgRNA表达的qRT-PCR分析

[0410] 将铺板于24孔板的293FT细胞如上进行转染。转染后72小时,用 miRNeasy微试剂盒(凯杰公司(Qiagen))收获总RNA。用qScript Flex cDNA试剂盒(VWR)以及自定义的第一链合成引物(表J以及K)进行针 对sgRNA的反向链合成。用快速SYBR绿色预混合液(Fast SYBR Green Master Mix)(生命技术公司)以及自定义的引物(表J以及K),使用 GAPDH作为内源对照进行qPCR分析。通过ΔΔCT法计算相对定量。

[0411] 表I|靶位点序列。针对具有必备的PAM的化脓链球菌II型 CRISPR系统,测试靶位点。针对每个靶标,用Cas9以及crRNA-tracrRNA或 嵌合sgRNA转染细胞。

靶位点ID	基因组靶标	靶位点序列(5'至3')	PAM	链
1	*EMX1*	GTCACCTCCAATGACTAGGG	TGG	+
2	*EMX1*	GACATCGATGTCCTCCCCAT	TGG	-
3	*EMX1*	GAGTCCGAGCAGAAGAAGAA	GGG	+
6	*EMX1*	GCGCCACCGGTTGATGTGAT	GGG	-
10	*EMX1*	GGGGCACAGATGAGAAACTC	AGG	-
11	*EMX1*	GTACAAACGGCAGAAGCTGG	AGG	+
12	*EMX1*	GGCAGAAGCTGGAGGAGGAA	GGG	+
13	*EMX1*	GGAGCCCTTCTTCTTCTGCT	CGG	-
14	*EMX1*	GGGCAACCACAAACCCACGA	GGG	+
15	*EMX1*	GCTCCCATCACATCAACCGG	TGG	+
16	*EMX1*	GTGGCGCATTGCCACGAAGC	AGG	+
17	*EMX1*	GGCAGAGTGCTGCTTGCTGC	TGG	+
18	*EMX1*	GCCCCTGCGTGGGCCCAAGC	TGG	+
19	*EMX1*	GAGTGGCCAGAGTCCAGCTT	GGG	-
20	*EMX1*	GGCCTCCCCAAAGCCTGGCC	AGG	-
4	*PVALB*	GGGGCCGAGATTGGGTGTTC	AGG	+
5	*PVALB*	GTGGCGAGAGGGGCCGAGAT	TGG	+
1	*SERPINB5*	GAGTGCCGCCGAGGCGGGGC	GGG	+
2	*SERPINB5*	GGAGTGCCGCCGAGGCGGGG	CGG	+
3	*SERPINB5*	GGAGAGGAGTGCCGCCGAGG	CGG	+

[0412] 表J|引物序列

SURVEYOR测定

引物名称	基因组靶标	引物序列(5’至3’)
Sp-EMX1-F1	EMX1	AAAACCACCCTTCTCTCTGGC
Sp-EMX1-R1	EMX1	GGAGATTGGAGACACGGAGAG
Sp-EMX1-F2	EMX1	CCATCCCCTTCTGTGAATGT
Sp-EMX1-R2	EMX1	GGAGATTGGAGACACGGAGA
Sp-PVALB-F	PVALB	CTGGAAAGCCAATGCCTGAC
Sp-PVALB-R	PVALB	GGCAGCAAACTCCTTGTCCT

qRT-PCR,用于Cas9以及sgRNA表达

引物名称	引物序列（5’至3’）
sgRNA 反向链合成	AAGCACCGACTCGGTGCCAC
EMX1.1 sgRNA qPCR F	TCACCTCCAATGACTAGGGG
EMX1.1 sgRNA qPCR R	CAAGTTGATAACGGACTAGCCT
EMX1.3 sgRNA qPCR F	AGTCCGAGCAGAAGAAGAAGTTT
EMX1.3 sgRNA qPCR R	TTTCAAGTTGATAACGGACTAGCCT
Cas9 qPCR F	AAACAGCAGATTCGCCTGGA
Cas9 qPCR R	TCATCCGCTCGATGAAGCTC
GAPDH qPCR F	TCCAAAATCAAGTGGGGCGA
GAPDH qPCR R	TGATGACCCTTTTGGCTCCC

亚硫酸氢盐PCR和测序

引物名称	引物序列（5’至3’）
亚硫酸氢盐 PCR F（*SERPINB5* 座位）	GAGGAATTCTTTTTTTGTTYGAATATGTTGGAGGTTTTTTGGAAG
亚硫酸氢盐 PCR R（*SERPINB5* 座位）	GAGAAGCTTAAATAAAAAACRACAATACTCAACCCAACAACC
pUC19 测序	CAGGAAACAGCTATGAC

[0413] 表K|测试sgRNA构造的引物的序列。引物杂交至该U6启动子的 反向链,除非另外表明。U6引发位点是斜体的,该指导序列被表示为一段 N,该同向重复序列是加粗凸显的,并且该tracrRNA序列是加了下划线的。

每个sgRNA构造的二级结构示于图43中。

引物名称	引物序列（5’至3’）
U6-正向	GCCTCTA*GAGGTACCTGAGGGCCTATTTCCCATGATTCC*
I：sgRNA（DR +12，tracrRNA +85）	ACCTCTAGAAAAAAAGCACCGACTCGGTGCCACTTTT TCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCT ATTTC**TAGCTCTAAAAC**NNNNNNNNNNNNNNNNN NN*GGTGTTTCGTCCTTTCCACAAG*
II：sgRNA（DR +12，tracrRNA +85）mut2	ACCTCTAGAAAAAAAGCACCGACTCGGTGCCACTTTT TCAAGTTGATAACGGACTAGCCTTATATTAACTTGCT ATTTC**TAGCTCTAATAC**NNNNNNNNNNNNNNNNN NN*GGTGTTTCGTCCTTTCCACAAG*
III：sgRNA（DR +22，tracrRNA +85）	ACCTCTAGAAAAAAAGCACCGACTCGGTGCCACTTTT TCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCT ATGCTGTTTTGTTTC**CAAAACAGCATAGCTCTAAAA** **C**NNNNNNNNNNNNNNNNNN*GGTGTTTCGTCCTTTC CACAAG*
IV：sgRNA（DR +22，tracrRNA +85）mut4	ACCTCTAGAAAAAAAGCACCGACTCGGTGCCACTTTT TCAAGTTGATAACGGACTAGCCTTATATTAACTTGCT ATGCTGTATTGTTTC**CAATACAGCATAGCTCTAATA** **C**NNNNNNNNNNNNNNNNNN*GGTGTTTCGTCCTTTC CACAAG*

[0414] 表L|用以测试Cas9的PAM特异性的具有交替的PAM的靶位点 发现用于PAM特异性测试的所有靶位点在人EMX1座位内。

靶位点序列（5’至3’）	PAM
AGGCCCCAGTGGCTGCTCT	NAA
ACATCAACCGGTGGCGCAT	NAT

AAGGTGTGGTTCCAGAACC	NAC
CCATCACATCAACCGGTGG	NAG
AAACGGCAGAAGCTGGAGG	NTA
GGCAGAAGCTGGAGGAGGA	NTT
GGTGTGGTTCCAGAACCGG	NTC
AACCGGAGGACAAAGTACA	NTG
TTCCAGAACCGGAGGACAA	NCA
GTGTGGTTCCAGAACCGGA	NCT
TCCAGAACCGGAGGACAAA	NCC
CAGAAGCTGGAGGAGGAAG	NCG
CATCAACCGGTGGCGCATT	NGA
GCAGAAGCTGGAGGAGGAA	NGT
CCTCCCTCCCTGGCCCAGG	NGC
TCATCTGTGCCCCTCCCTC	NAA
GGGAGGACATCGATGTCAC	NAT
CAAACGGCAGAAGCTGGAG	NAC
GGGTGGGCAACCACAAACC	NAG
GGTGGGCAACCACAAACCC	NTA
GGCTCCCATCACATCAACC	NTT
GAAGGGCCTGAGTCCGAGC	NTC
CAACCGGTGGCGCATTGCC	NTG
AGGAGGAAGGGCCTGAGTC	NCA
AGCTGGAGGAGGAAGGGCC	NCT
GCATTGCCACGAAGCAGGC	NCC
ATTGCCACGAAGCAGGCCA	NCG
AGAACCGGAGGACAAAGTA	NGA
TCAACCGGTGGCGCATTGC	NGT
GAAGCTGGAGGAGGAAGGG	NGC

实例10:补充序列

[0415] 所有序列均处于5’至3’方向。对于U6转录,加下划线的一串T充 当转录终止子。

[0416] >U6-短tracrRNA(化脓链球菌SF370)

[0417] gagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattggaattaatttga
ctgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttt
taaaatggactatcatatgcttaccgtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgaaacacc
GGAACCATTCAAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGT

TATCAACTTGAAAAAGTGGCACCGAGTCGGTGC<u>TTTTTTT</u>

[0418] (tracrRNA序列被加粗)

[0419] >U6-DR-指导序列-DR(化脓链球菌SF370)

gagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattggaattaatttga
ctgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttt

[0420] taaaatggactatcatatgcttaccgtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgaaacacc
gggttttagagctatgctgttttgaatggtcccaaaac**NNNNNNNNNNNNNNNNNNNNNNNNN**
NNNNNgttttagagctatgctgttttgaatggtcccaaaacTTTTTTT

[0421] (同向重复序列是以灰色凸显的并且指导序列是加粗的N)

[0422] >sgRNA,包含+48tracrRNA(化脓链球菌SF370)

gagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattggaattaatttga
ctgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttt

[0423] taaaatggactatcatatgcttaccgtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgaaacacc
NNNNNNNNNNNNNNNNNNNgttttagagctagaaa**tagcaagttaaaataaggctagtccg**T
TTTTTT

[0424] (指导序列是加粗的N并且tracrRNA片段被加粗)

[0425] >sgRNA,包含+54tracrRNA(化脓链球菌SF370)

gagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattggaattaatttga
ctgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttt

[0426] taaaatggactatcatatgcttaccgtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgaaacacc
NNNNNNNNNNNNNNNNNNNgttttagagctagaaa**tagcaagttaaaataaggctagtccgtt**
atcaTTTTTTTT

[0427] (指导序列是加粗的N并且tracrRNA片段被加粗)

[0428] >sgRNA,包含+67tracrRNA(化脓链球菌SF370)

gagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattggaattaatttga
ctgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttt

[0429] taaaatggactatcatatgcttaccgtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgaaacacc
NNNNNNNNNNNNNNNNNNNgttttagagctagaaa**tagcaagttaaaataaggctagtccgtt**
atcaacttgaaaaagtgTTTTTTT

[0430] (指导序列是加粗的N并且tracrRNA片段被加粗)

[0431] >sgRNA,包含+85tracrRNA(化脓链球菌SF370)

gagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattggaattaatttga
ctgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttt

[0432] taaaatggactatcatatgcttaccgtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgaaacacc
NNNNNNNNNNNNNNNNNNNgttttagagctagaaa**tagcaagttaaaataaggctagtccgtt**
atcaacttgaaaaagtggcaccgagtcggtgcTTTTTTT

[0433] (指导序列是加粗的N并且tracrRNA片段被加粗)

[0434] >CBh-NLS-SpCas9-NLS

CGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAAC
GACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGC
CAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAAC
TGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCT
ATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACA
TGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATC
GCTATTACCATGGTCGAGGTGAGCCCCACGTTCTGCTTCACTCTCCCCAT
CTCCCCCCCCTCCCCACCCCCAATTTTGTATTTATTTATTTTTTAATTATT
TTGTGCAGCGATGGGGGCGGGGGGGGGGGGGGGCGCGCGCCAGGCGG
GGCGGGGCGGGGCGAGGGGCGGGGCGGGGCGAGGCGGAGAGGTGCGG
CGGCAGCCAATCAGAGCGGCGCGCTCCGAAAGTTTCCTTTTATGGCGAG
GCGGCGGCGGCGGCGGCCCTATAAAAAGCGAAGCGCGCGGCGGGCGGG
AGTCGCTGCGACGCTGCCTTCGCCCCGTGCCCCGCTCCGCCGCCGCCTCG
CGCCGCCCGCCCCGGCTCTGACTGACCGCGTTACTCCCACAGGTGAGCG
GGCGGGACGGCCCTTCTCCTCCGGGCTGTAATTAGCTGAGCAAGAGGTA
AGGGTTTAAGGGATGGTTGGTTGGTGGGGTATTAATGTTTAATTACCTG
GAGCACCTGCCTGAAATCACTTTTTTTCAGGTTGGaccggtgccaccATGGAC
TATAAGGACCACGACGGAGACTACAAGGATCATGATATTGATTACA
AAGACGATGACGATAAGATGGCCCCAAAGAAGAAGCGGAAGGTCG
GTATCCACGGAGTCCCAGCAGCCGACAAGAAGTACAGCATCGGCCT
GGACATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAG
TACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACC
[0435] GGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAG
CGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAG
AAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATC
TTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGAC
TGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCA
CCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAG
TACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCG
ACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGAT
CAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGAC
AACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACA
ACCAGCTGTTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACGC
CAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAA
AATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCG
GCAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAG
CAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGAC
ACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACC
AGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCAT
CCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCC
CCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGG
ACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAA
GTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGC
TACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCA
AGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAA

GCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAAC
GGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTC
TGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGA
AAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGC
CCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGA
GCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAA
GGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGAT
AAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGT
ACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGT
GACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCAGAA
AAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACC
GTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCG
ACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCT
GGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTC
CTGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGA
CCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAA
AACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAG
CGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATC
AACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCC
TGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCA
CGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTG
TCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCG
GCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGT
GGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACAT
CGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACA
GAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAA
AGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACC
CAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGC
GGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGA
CTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGAC
TCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCA
AGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAA
CTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAG
TTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGG
ATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGAT
CACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAG
TACGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCC
TGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTAC
AAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACC
TGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCT
GGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGG
AAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCA
AGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATT
ACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAA
ACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTG
CCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAA

AAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTG
CCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGG
GACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATT
CTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACT
GAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGC
AGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACA
AAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCT
GTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGC
GAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGA
ACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCC
CGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCAC
TACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAG
TGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAA
CAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATC
CACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGT
ACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGA
GGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTAC
GAGACACGGATCGACCTGTCTCAGCTGGGAGGCGACTTTCTTTTTC
TTAGCTTGACCAGCTTTCTTAGTAGCAGCAGGACGCTTTAA

[0436] (NLS-hSpCas9-NLS是以加粗凸显的)

[0437] >测序扩增子,用于EMX1指导1.1、1.14、1.17

[0438] CCAATGGGGAGGACATCGATGTCACCTCCAATGACTAGGGTGGGC AACCACAAACCCACGAGGGCAG AGTGCTGCTTGCTGCTGGCCAGGCCCC TGCGTGGGCCCAAGCTGGACTCTGGCCAC

[0439] >测序扩增子,用于EMX1指导1.2、1.16

[0440] CGAGCAGAAGAAGAAGGGCTCCCATCACATCAACCGGTGGCGCAT TGCCACGAAGCAGGCCAATGGG GAGGACATCGATGTCACCTCCAATGA CTAGGGTGGGCAACCACAAACCCACGAG

[0441] >测序扩增子,用于EMX1指导1.3、1.13、1.15

[0442] GGAGGACAAAGTACAAACGGCAGAAGCTGGAGGAGGAAGGGCCT GAGTCCGAGCAGAAGAAGAAGGG CTCCCATCACATCAACCGGTGGCGC ATTGCCACGAAGCAGGCCAATGGGGAGGACATCGAT

[0443] >测序扩增子,用于EMX1指导1.6

[0444] AGAAGCTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAG GGCTCCCATCACATCAACCGGTG GCGCATTGCCACGAAGCAGGCCAATG GGGAGGACATCGATGTCACCTCCAATGACTAGGGTGG

[0445] >测序扩增子,用于EMX1指导1.10

[0446] CCTCAGTCTTCCCATCAGGCTCTCAGCTCAGCCTGAGTGTTGAGGC CCCAGTGGCTGCTCTGGGGGC CTCCTGAGTTTCTCATCTGTGCCCCTCCC TCCCTGGCCCAGGTGAAGGTGTGGTTCCA

[0447] >测序扩增子,用于EMX1指导1.11、1.12

[0448] TCATCTGTGCCCCTCCCTCCCTGGCCCAGGTGAAGGTGTGGTTCCA GAACCGGAGGACAAAGTACAA ACGGCAGAAGCTGGAGGAGGAAGGGC CTGAGTCCGAGCAGAAGAAGAAGGGCTCCCATCACA

[0449] >测序扩增子,用于EMX1指导1.18、1.19

[0450] CTCCAATGACTAGGGTGGGCAACCACAAACCCACGAGGGCAGAGT GCTGCTTGCTGCTGGCCAGGCC CCTGCGTGGGCCCAAGCTGGACTCTGG CCACTCCCTGGCCAGGCTTTGGGGAGGCCTGGAGT

[0451] >测序扩增子,用于EMX1指导1.20

[0452] CTGCTTGCTGCTGGCCAGGCCCCTGCGTGGGCCCAAGCTGGACTCT GGCCACTCCCTGGCCAGGCTT

TGGGGAGGCCTGGAGTCATGGCCCCACA GGGCTTGAAGCCCGGGGCCGCCATTGACAGAG

[0453] >T7启动子F引物,用于靶链退火

[0454] GAAATTAATACGACTCACTATAGGG

[0455] >包含pUC19靶位点1的寡聚物,用于甲基化(T7反向)

[0456] AAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACT AGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACAACGACGAGCGTGA CACCACCCTATAGTGAGTCGTATTAATTTC

[0457] >包含pUC19靶位点2的寡聚物,用于甲基化(T7反向)

[0458] AAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACT AGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGCAACAATTAATAG ACTGGACCTATAGTGAGTCGTATTAATTTC

实例11:寡聚物介导的Cas9诱导的同源重组

[0459] 寡聚物同源重组测试是跨不同Cas9变体和不同的HR模板的效率 的比较(寡聚物对质粒)。

[0460] 使用293FT细胞。SpCas9=野生型Cas9并且SpCas9n=切口酶 Cas9(D10A)。该嵌合RNA靶标是与在实例5、9以及10中相同的EMX1 原型间隔子靶标1并且使用PAGE纯化通过IDT合成寡聚物。

[0461] 图44描绘了在此试验中用作同源重组(HR)模板的寡DNA的一 种设计。长寡聚物包含与EMX1座位以及HindIII限制酶切位点100bp同源 性。用以下各项共转染293FT细胞:第一,包含一种靶向人EMX1座位以及 野生型cas9蛋白的嵌合RNA的一种质粒,以及第二,作为HR模板的寡 DNA。样品来自用Lipofectamine 2000转染96小时后收集的293FT细胞。将 所有产物用一种EMX1 HR引物扩增,凝胶纯化,随后用HindIII消化以检测 HR模板整合至人类基因组中的效率。

[0462] 图45以及46描绘了由Cas9蛋白以及HR模板的不同组合诱导的 HR效率的比较。使用的Cas9构建体是野生型Cas9或切口酶版本的Cas9 (Cas9n)。使用的HR模板是:反义寡DNA(上图中的反义寡聚物),或有 义寡DNA(上图中的有义寡聚物),或质粒HR模板(上图中的HR模 板)。该有义/反-有义定义是,具有相应于转录的mRNA的序列的有效转录 的链被定义为基因组的有义链。HR效率显示为HindIII消化条带对所有基因 组PCR扩增产物的百分数(底数)。

实例12:孤独症小鼠

[0463] 最近的大规模测序项目已产生大量的与疾病相关联的基因。基因的 发现仅是理解该基因做了什么以及它如何导致病态表型的开始。研究候选基 因的当前技术以及途径是缓慢且费力的。基因靶向以及遗传敲除的黄金标准 需要在财政和研发人员两项上在时间以及资源上进行大量投资。申请人着手 利用hSpCas9核酸酶以靶向许多基因并且以相较于任何其他技术更高的效率 和更低的周转时间如此进行。因为hSpCas9的高效率,申请人可将RNA注射 至小鼠合子中,并且立即获得基因组修饰的动物而不需要在mESC中进行任何初步的基因靶向。

[0464] 克罗莫结构域螺旋酶DNA结合蛋白8(CHD8)是一种参与早期脊 椎发育以及形态发生的关键基因。缺少CHD8的小鼠在胚胎发育过程中死 亡。该CHD8基因的突变已与人的自闭症谱系障碍相关联。这种关联是在《自然》(Nature)中以三篇不同文章同时公开的。相同的三项研究鉴定了种 类繁多的与自闭症谱系障碍相关联的基因。申请人的目标是针对在所

有文章 中发现的四个基因Chd8、Katnal2、Kctd13、以及Scn2a，建立基因敲除小 鼠。此外，申请人选择了与自闭症谱系障碍、精神分裂症、以及ADHD相关 联的两个其他基因、GIT1、CACNA1C、以及CACNB2。并且最终，作为一 个阳性对照，申请人决定靶向MeCP2。

[0465] 针对每个基因，申请人设计很可能敲除该基因的三个gRNA。在 hSpCas9核酸酶之后发生的敲除引起双链断裂并且易错DNA修复通路、非同 源末端连接校正该断裂，建立一个突变。最可能的结果是会敲除该基因的一 种移码突变。该靶向策略涉及在该基因的外显子中发现原型间隔序列，该基 因具有一个PAM序列，NGG，并且在基因组中是独特的。在第一外显子中优 选原型间隔序列，其最有害于该基因。

[0466] 在小鼠细胞系即Neuro-N2a中通过用hSpCas9进行的脂质体瞬时共 转染验证每个gRNA。转染后72小时，使用来自Epicentre的QuickExtract DNA纯化基因组DNA。进行PCR以扩增感兴趣的基因座。随后，使用来自 转基因组学公司(Transgenomics)的SURVEYOR突变检测试剂盒。针对每 个gRNA以及各自的对照的SURVEYOR结果示于图A1中。一个阳性SURVEYOR结果是对应于基因组的PCR的一条大的条带以及两个较小的条 带，这两个较小的条带是使得双链在一个突变位点断裂的SURVEYOR核酸 酶的产物。还针对每个gRNA测定每个gRNA的平均切割效率。所选择的用 于注射的gRNA是最高效的gRNA，其在该基因组内是最独特的。

[0467] 将RNA(hSpCas9+gRNA RNA)注射至一种合子的前核中并且之 后移植至代孕母亲中。允许母亲妊娠期满并且出生后10天对幼兽通过剪尾取 样。提取DNA并且用作PCR的模板，然后通过SURVEYOR进行处理。此 外，将PCR产物送去测序。将在SURVEYOR测定或PCR测序中检测为阳性 的动物的基因组PCR产物克隆至一种pUC19载体中并且测序以确定来自每个等位基因的假定的突变。

[0468] 迄今为止，来自Chd8靶向试验的小鼠幼兽已被完全处理为达到等 位基因测序的程度。图A2显示针对38个活的幼兽(道1-38)、1个死的幼 兽(道39)以及用于比较的1个野生型幼兽(道40)的Surveyor结果。将幼 兽1-19用gRNA Chd8.2进行注射，并且将幼兽20-38用gRNA Chd8.3进行注 射。该38个活的幼兽中13个针对一种突变是阳性的。这个死的幼兽也具有 一种突变。在野生型样品中未检测到突变。基因组PCR测序与SURVEYOR 测定发现是一致的。

实例13:CRISPR/Cas介导的转录调制

[0469] 图67描绘具有转录激活活性的CRISPR-TF(转录因子)的一种设 计。该嵌合RNA由U6启动子表达，而可操作地连接到三个NLS以及VP64 功能域的一种人-密码子-优化的双突变版本的Cas9蛋白(hSpCas9m)由一种 EF1a启动子表达。该双突变，即D10A以及H840A，导致当由嵌合RNA指 导时，cas9蛋白不能够引入任何切割但维持其结合到靶DNA的能力。

[0470] 图68描绘用CRISPR-TF系统(嵌合RNA以及Cas9-NLS-VP64融 合蛋白)对人SOX2基因进行的转录激活。将293FT细胞用如下质粒转染，所述质粒具有两个组分：(1)U6-驱动的不同嵌合RNA，靶向人SOX2基因 组座位内或周围的20-bp序列，以及(2)EF1a-驱动的hSpCas9m(双重突变 型)-NLS-VP64融合蛋白。转染后96小时，收获293FT细胞，并且通过使用 一种qRT-PCR测定引入mRNA表达以测量激活水平。将所有表达水平针对对 照组(灰色条)标准化，对照组表示来自用该CRISPR-TF骨架质粒(不含嵌 合RNA)转染的细胞的结果。用于检测SOX2 mRNA的qRT-PCR探针是 Taqman人基因表达测定用探针(生命技术公司(Life

Technologies))。所有 实验表示来自3个生物重复的数据,n=3,误差条显示为s.e.m.。

实例14:NLS:Cas9 NLS

[0471] 将293FT细胞用包含两种组分的质粒转染:(1)EF1a启动子,驱 动具有不同NLS设计的Cas9(野生型人-密码子优化Sp Cas9)的表达,(2)U6启动子,驱动靶向人EMX1座位的相同的嵌合RNA。

[0472] 遵循制造商方案,在转染后72h时间点收集细胞,并且然后用50 μl的QuickExtract基因组DNA提取溶液进行提取。将靶EMX1基因组DNA 进行PCR扩增,并且然后用1%琼脂糖凝胶进行凝胶纯化。遵循制造商方 案,将基因组PCR产物重退火并且经受Surveyor测定。不同构建体的基因组 切割效率使用SDS-PAGE在一种4%-12%TBE-PAGE凝胶(生命技术公司) 上进行测量,分析并且用ImageLab(伯乐公司(Bio-rad))软件定量,全部 操作遵循制造商方案。

[0473] 图69描绘了不同Cas9 NLS构建体的一种设计。所有Cas9都是人- 密码子-优化版本的Sp Cas9。在N-末端或C-末端,NLS序列被连接至该cas9 基因。将具有不同NLS设计的所有Cas9变体克隆至包含EF1a启动子的一种 骨架载体中,所以它由EF1a启动子驱动。在相同的载体中,存在靶向人 EMX1座位的由U6启动子驱动的一种嵌合RNA,一起形成一个二组分系 统。

[0474] 表M.Cas9 NLS设计测试结果。通过surveyor测定对不同cas9-nls 构建体的基因组切割进行定量。

通过 surveyor 测定测量的基因组切割百分数	生物学重复 1 (%)	生物学重复 2 (%)	生物学重复 3 (%)	平均 (%)	误差 (S.E.M.,平均值的标准差)。
Cas9 (无 NLS)	2.50	3.30	2.73	2.84	0.24
具有 N-末端 NLS 的 Cas9	7.61	6.29	5.46	6.45	0.63
具有 C-末端 NLS 的 Cas9	5.75	4.86	4.70	5.10	0.33
具有双 (N-末端以及 C-末端) NLS 的 Cas9	9.08	9.85	7.78	8.90	0.60

[0475] 图70描绘了由具有不同NLS设计的Cas9变体诱导的基因组切割 的效率。该百分数表明被每个构建体切割的人EMX1基因组DNA的部分。所 有实验来自3个生物重复。n=3,误差表示为S.E.M.。

实例15:使用Cas9工程化微藻

[0476] 递送Cas9的方法

[0477] 方法1:申请人使用在组成型启动子(如Hsp70A-Rbc S2或β2-微 管蛋白)的控制之下表达Cas9的载体递送Cas9和指导RNA。

[0478] 方法2:申请人使用在组成型启动子(如Hsp70A-Rbc S2或β2-微 管蛋白)的控制之下表达Cas9和T7聚合酶的载体递送Cas9和T7聚合酶。将使用包含驱动指导RNA的T7启动子的载体递送该指导RNA。

[0479] 方法3:申请人将Cas9 mRNA和体外转录的指导RNA递送到藻类 细胞中。RNA可以在体外转录。Cas9 mRNA将由Cas9的编码区以及来自的 Cop1的3'UTR组成,以保证Cas9 mRNA的稳定。

[0480] 用于同源重组,申请人提供了一个另外的同源定向修复模板。

[0481] 在位于Cop1的3'UTR之后的β-2微管蛋白启动子的控制之下驱动 Cas9的表达的盒的序列。

[0482] TCTTTCTTGCGCTATGACACTTCCAGCAAAAGGTAGGGCGGGCTGC GAGACGGCTTCCCGGCGCTGC ATGCAACACCGATGATGCTTCGACCCCC CGAAGCTCCTTCGGGGCTGCATGGGCGCTCCGATGCCGCTCCAGGGC GA GCGCTGTTTAAATAGCCAGGCCCCCGATTGCAAAGACATTATAGCGAGC TACCAAAGCCATATTCAAACACC TAGATCACTACCACTTCTACACAGGC CACTCGAGCTTGTGATCGCACTCCGCTAAGGGGGCGCCTCTTCCTCTTCG TTTCAGTCACAACCCGCAAACATGTACCCATACGATGTTCCAGATTACG CTTCGCCGAAGAAAAAGCGCAAGGTC GAAGCGTCCGACAAGAAGTACA GCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTGATCAC CGA CGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACAC CGACCGGCACAGCATCAAGAAGAACCTGAT CGGAGCCCTGCTGTTCGAC AGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGA AGATACA CCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCA GCAACGAGATGGCCAAGGTGGACGACAGCTTCT TCCACAGACTGGAAG AGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTT CGGCAACATC GTGGACGAGGTGGCCTACCACGAGAAGTACCCCACCATC TACCACCTGAGAAAGAAACTGGTGGACAGCACCGAC AAGGCCGACCTG CGGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACT TCCTGATCGAGGG CGACCTGAACCCCGACAACAGCGACGTGGACAAGC TGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGA GGAAAACCC CATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTG AGCAAGAGCAGACGGC TGGAAAATCTGATCGCCCAGCTGCCCGGCGAG AAGAAGAATGGCCTGTTCGGCAACCTGATTGCCCTGAGCCTGG GCCTGA CCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCA GCTGAGCAAGGACACCTAC GACGACGACCTGGACAACCTGCTGGCCCA GATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTG TCC GACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCA AGGCCCCCCTGAGCGCCTCTAT GATCAAGAGATACGACGAGCACCACCA GGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAG TACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACA TTGACGGCGGAGCCAGCCAGGAAGAG TTCTACAAGTTCATCAAGCCCAT CCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAG AGA GGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCC CCACCAGATCCACCTGGGAGAGCTGCACGC CATTCTGCGGCGGCAGGAA GATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCC TGACCT TCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAG CAGATTCGCCTGGATGACCAGAAAGAGCGAGG AAACCATCACCCCCTG GAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATC GAGCGGATG ACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTG CCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTG

TATAACGAGCTGA CCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGA GCGGCGAGCAGAA
AAAGGCCATCGTGGACCTGCTGTTCAAGACCAACC GGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAA
GAAAATCG AGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGC CTCCCTGGGCACATACC
ACGATCTGCTGAAAATTATCAAGGACAAGGAC TTCCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCG
TGCTGA CCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAA CCTATGCCCACCTGTTCGAC
GACAAAGTGATGAAGCAGCTGAAGCGGCG GAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGC
AT CCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGAC GGCTTCGCCAACAGAAACTTCAT
GCAGCTGATCCACGACGACAGCCTGA CCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATA
GCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAA GGGCATCCTGCAGACAGTGAAGGTGG
TGGACGAGCTCGTGAAAGTGAT GGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAA CCAG
ACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGC GGATCGAAGAGGGCATCAAAGAGCTGGGCAGC
CAGATCCTGAAAGAAC ACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTA CCTGCAGAA
TGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAA CCGGCTGTCCGACTACGATGTGGACCATATCGTGCC
TCAGAGCTTTCTG AAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAAC CGGGGCAAGAGCG
ACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATG AAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTA
CCCAGAGA AAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTG GATAAGGCCGGCTTCATC
AAGAGACAGCTGGTGGAAACCCGGCAGATC ACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAG
TAC GACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAG TCCAAGCTGGTGTCCGATTTCCG
GAAGGATTTCCAGTTTTACAAAGTGC GCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGT
CGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTC GTGTACGGCGACTACAAGGTGTACGA
CGTGCGGAAGATGATCGCCAAG AGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCA ACA
TCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGAT CCGGAAGCGGCCTCTGATCGAGACAAACG
GCGAAACCGGGGAGATCGT GTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATG CCCCAA
GTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTC AGCAAAGAGTCTATCCTGCCCAAGAGGAACAGC
GATAAGCTGATCGCC AGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCC ACCGTGGCCTA
TTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGT CCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGAT
CACCATCATGG AAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGG GCTACAAAGAAGTGA
AAAAGGACCTGATCATCAAGCTGCCTAAGTACTC CCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCT
CTGCCGGC GAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACT TCCTGTACCTGGCCAGC
CACTATGAGAAGCTGAAGGGCTCCCCCGAGGA TAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTAC
CTGGAC GAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCG ACGCTAATCTGGACAAAGT
GCTGTCCGCCTACAACAAGCACCGGGATAA GCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCT
GACC AATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACC GGAAGAGGTACACCAGCACCA
AAGAGGTGCTGGACGCCACCCTGATCC ACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCT
GGGAGGCGACAGCCCCAAGAAGAAGAGAAAGGTGGAGGCCAGCTAAG GATCCGGCAAGACTGGCCCCGCTTGGCA
ACGCAACAGTGAGCCCCTCCC TAGTGTGTTTGGGGATGTGACTATGTATTCGTGTGTTGGCCAACGGGTCA ACC
CGAACAGATTGATACCCGCCTTGGCATTTCCTGTCAGAATGTAACG TCAGTTGATGGTACT

[0483] 在位于Cop1的3'UTR之后的β-2微管蛋白启动子的控制之下驱动 T7聚合酶的表达的盒的序列：

[0484] TCTTTCTTGCGCTATGACACTTCCAGCAAAAGGTAGGGCGGGCTGC GAGACGGCTTCCCGGCGCTGC ATGCAACACCGATGATGCTTCGACCCCC CGAAGCTCCTTCGGGGCTGCATGGGCGCTCCGATGCCGCTCCAGGGC GA GCGCTGTTTAAATAGCCAGGCCCCCGATTGCAAAGACATTATAGCGAGC TACCAAAGCCATATTCAAACACC TAGATCACTACCACTTCTACACAGGC CACTCGAGCTTGTGATCGCACTCCGCTAAGGGGGCGCCTCTTCCTCTTCG TTTCAGTCACAACCCGCAAACatgcctaagaagaagaggaaggttaacacgattaacatcgctaag aacgacttc tctgacatcgaactggctgctatcccgttcaacactctggctgaccattacggtgagcgtttagctcgcgaa cag ttggcccttgagcatgagtcttacgagatgggtgaagcacgcttccgcaagatgtttgagcgtcaacttaaagctgg tgaggttgcggataacgctgccgccaagcctctcatcactaccctactccctaagatgattgcacgcatcaacgac tggtt tgaggaagtgaaagctaagcgcggcaagcgcccgacagccttccagttcctgcaagaaatcaagccggaa gccgta gcgtacatcaccattaagaccactctggcttgcctaaccagtgctgacaatacaaccgttcaggctgta gcaagcgcaat cggtcgggccattgaggacgaggctcgcttcggtcgtatccgtgaccttgaagctaagcacttc aagaaaaacgttgag gaacaactcaacaagcgcgtagggcacgtctacaagaaagcatttatgcaagttgtcgag gctgacatgctctctaagg gtctactcggtggcgaggcgtggtcttcgtggcataaggaagactctattcatgta ggagtacgctgcatcgagatgctc attgagtcaaccggaatggttagcttacaccgccaaaatgctggcgtagta ggtcaagactctgagactatcgaactcgc acctgaatacgctgaggctatcgcaacccgtgcaggtgcgctggct ggcatctctccgatgttccaaccttgcgtagttc ctcctaagccgtggactggcattactggtggtggctattgg gctaacggtcgtcgtcctctggcgctggtgcgtactcaca gtaagaaagcactgatgcgctacgaagacgtttac atgcctgaggtgtacaaagcgattaacattgcgcaaaacaccgc atggaaaatcaacaagaaagtcctagcggtc gccaacgtaatcaccaagtggaagcattgtccggtcgaggacatccc tgcgattgagcgtgaagaactcccgatg aaaccggaagacatcgacatgaatcctgaggctctcaccgcgtggaaacg tgctgccgctgctgtgtaccgcaag gacaaggctcgcaagtctcgccgtatcagccttgagttcatgcttgagcaagcca ataagtttgctaaccataag gccatctggttcccttacaacatggactggcgcggtcgtgtttacgctgtgtcaatgttcaac ccgcaaggtaac gatatgaccaaaggactgcttacgctggcgaaaggtaaaccaatcggtaaggaaggttactactgg ctgaaaatc cacggtgcaaactgtgcgggtgtcgacaaggttccgttccctgagcgcatcaagttcattgaggaaaacc acgag aacatcatggcttgcgctaagtctccactggagaacacttggtgggctgagcaagattctccgttctgcttccttg cgttctgctttgagtacgctggggtacagcaccacggcctgagctataactgctcccttccgctggcgtttgacgg gtctt gctctggcatccagcacttctccgcgatgctccgagatgaggtaggtggtcgcgcggttaacttgcttcc tagtgaaacc gttcaggacatctacgggattgttgctaagaaagtcaacgagattctacaagcagacgcaatcaa tgggaccgataacg aagtagttaccgtgaccgatgagaacactggtgaaatctctgagaaagtcaagctgggcac taaggcactggctggtca atggctggcttacggtgttactcgcagtgtgactaagcgttcagtcatgacgctggc ttacgggtccaaagagttcggctt ccgtcaacaagtgctggaagataccattcagccagctattgattccggcaa gggtctgatgttcactcagccgaatcagg ctgctggatacatggctaagctgatttgggaatctgtgagcgtgac ggtggtagctgcggttgaagcaatgaactggctt aagtctgctgctaagctgctggctgctgaggtcaaagataa gaagactggagagattcttcgcaagcgttgcgctgtgca ttgggtaactcctgatggtttccctgtgtggcagga atacaagaagcctattcagacgcgcttgaacctgatgttcctcggt cagttccgcttacagcctaccattaacac caacaaagatagcgagattgatgcacacaaacaggagtctggtatcgctcc taactttgtacacagccaagacgg tagccaccttcgtaagactgtagtgtgggcacacgagaagtacggaatcgaatctt ttgcactgattcacgactc cttcggtacgattccggctgacgctgcgaacctgttcaaagcagtgcgcgaaactatggttg acacatatgagtc ttgtgatgtactggctgatttctacgaccagttcgctgaccagttgcacgagtctcaattggacaaaat gccagc

acttccggctaaaggtaacttgaacctccgtgacatcttagagtcggacttcgcgttcgcgtaaGGATCC GGCAA GACTGGCCCCGCTTGGCAACGCAACAGTGAGCCCCTCCCTAGTG TGTTTGGGGATGTGACTATGTATTCGTGTGT TGGCCAACGGGTCAACCC GAACAGATTGATACCCGCCTTGGCATTTCCTGTCAGAATGTAACGTCAG TTGATGGTACT

[0485] 由T7启动子驱动的指导RNA的序列(T7启动子,N表示靶向序 列):

[0486] gaaatTAATACGACTCACTATANNNNNNNNNNNNNNNNNNNgttttag agctaGAAAtagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgctttttt

[0487] 基因递送:

[0488] 来自衣藻资源中心(Chlamydomonas Resource Center)的莱茵衣藻 株CC-124和CC-125将被用于电穿孔。电穿孔方案遵循来自GeneArt Chlamydomonas Engineering试剂盒的标准推荐方案。

[0489] 而且,申请人产生组成性地表达Cas9的莱茵衣藻系。这可以通过 使用pChlamy1(使用PvuI线性化)并选择潮霉素抗性菌落而完成。以下是包 含Cas9的pChlamy1的序列。以此方式实现基因敲除,简单地需要递送指导 RNA的RNA。对于同源重组,申请人递送了指导RNA以及线性化同源重组 模板。

[0490] pChlamy1-Cas9:

[0491] TGCGGTATTTCACACCGCATCAGGTGGCACTTTTCGGGGAAATGTG CGCGGAACCCCTATTTGTTTA TTTTTCTAAATACATTCAAATATGTATCC GCTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATT TAA AAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTG ACAGTTACCAATGCTTAATCAG TGAGGCACCTATCTCAGCGATCTGTCT ATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTAC GA TACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGA CCCACGCTCACCGGCTCCAGATT TATCAGCAATAAACCAGCCAGCCGGA AGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGT CTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAG TTTGCGCAACGTTGTTGCCATTGCTA CAGGCATCGTGGTGTCACGCTCGT CGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTT ACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCC GATCGTTGTCAGAAGTAAGTTGGCC GCAGTGTTATCACTCATGGTTATG GCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTC TGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGG CGACCGAGTTGCTCTTGCCCGGCGTC AATACGGGATAATACCGCGCCAC ATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCG AA AACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCC ACTCGTGCACCCAACTGATCTTCAGCAT CTTTTACTTTCACCAGCGTTTC TGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAA GGGCG ACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTAT TGAAGCATTTATCAGGGTTATTGTCTCATG ACCAAAATCCCTTAACGTGA GTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCT TCTTG AGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAA ACCACCGCTACCAGCGGTGGTTTGTTTGCC GGATCAAGAGCTACCAACT CTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTG TTCTTC TAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCA CCGCCTACATACCTCGCTCTGCTAATCCTGT TACCAGTGGCTGTTGCCAG TGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCG GATAAG GCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCC AGCTTGGAGCGAACGACCTACACCGAACTGAGA TACCTACAGCGTGAG CTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTAT CCGGTAAGCGG CAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCA GGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTT

TCGCCACCTCTG ACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGG AAAAACGCCAGCA
ACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCC TTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATT
CTGTGGATAACCG TATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACC GAGCGCAGCGAG
TCAGTGAGCGAGGAAGCGGTCGCTGAGGCTTGACAT GATTGGTGCGTATGTTTGTATGAAGCTACAGGACTGATT
TGGCGGGCTA TGAGGGCGGGGGAAGCTCTGGAAGGGCCGCGATGGGGCGCGCGGCGTC CAGAAGGCGCCATACG
GCCCGCTGGCGGCACCCATCCGGTATAAAAGC CCGCGACCCCGAACGGTGACCTCCACTTTCAGCGACAAACGAG
CACTTA TACATACGCGACTATTCTGCCGCTATACATAACCACTCAGCTAGCTTAA GATCCCATCAAGCTTGCAT
GCCGGGCGCGCCAGAAGGAGCGCAGCCAA ACCAGGATGATGTTTGATGGGGTATTTGAGCACTTGCAACCCTTAT
CCG GAAGCCCCCTGGCCCACAAAGGCTAGGCGCCAATGCAAGCAGTTCGCAT GCAGCCCCTGGAGCGGTGCCCT
CCTGATAAACCGGCCAGGGGGCCTATG TTCTTTACTTTTTTACAAGAGAAGTCACTCAACATCTTAAAATGGCCA
GG TGAGTCGACGAGCAAGCCCGGCGGATCAGGCAGCGTGCTTGCAGATTTG ACTTGCAACGCCCGCATTGTGTC
GACGAAGGCTTTTGGCTCCTCTGTCGC TGTCTCAAGCAGCATCTAACCCTGCGTCGCCGTTTCCATTTGCAGGAG
AT TCGAGGTACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAG AAAAAGCGCAAGGTCGAAGCGTC
CGACAAGAAGTACAGCATCGGCCTG GACATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACA
AGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACA GCATCAAGAAGAACCTGATCGGAGCCC
TGCTGTTCGACAGCGGCGAAA CAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCA GACGG
AAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGAT GGCCAAGGTGGACGACAGCTTCTTCCACAGA
CTGGAAGAGTCCTTCCTG GTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATC GTGGACGA
GGTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTGA GAAAGAAACTGGTGGACAGCACCGACAAGGCCGA
CCTGCGGCTGATCT ATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGA GGGCGACCTGA
ACCCCGACAACAGCGACGTGGACAAGCTGTTCATCCA GCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACC
CCATCAACGCC AGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGC AGACGGCTGGAAAAT
CTGATCGCCCAGCTGCCCGGCGAGAAGAAGAAT GGCCTGTTCGGCAACCTGATTGCCCTGAGCCTGGGCCTGACC
CCCAACT TCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCA AGGACACCTACGACGACGA
CCTGGACAACCTGCTGGCCCAGATCGGCG ACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGC
CAT CCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCC CTGAGCGCCTCTATGATCAAGA
GATACGACGAGCACCACCAGGACCTGA CCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGA
GATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGC GGAGCCAGCCAGGAAGAGTTCTACAA
GTTCATCAAGCCCATCCTGGAAA AGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACC TGC
TGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGAT CCACCTGGGAGAGCTGCACGCCATTCTGC
GGCGGCAGGAAGATTTTTAC CCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCC GCATC
CCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGC CTGGATGACCAGAAAGAGCGAGGAAACCATC
ACCCCCTGGAACTTCGA GGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGAT GACCAACTT
CGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCA CAGCCTGCTGTACGAGTACTTCACCGTGTATAACGA
GCTGACCAAAGTG AAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAG CAGAAAAAGGCCA
TCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTG ACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCG
AGTGCTTC GACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGG GCACATACCACGATCTG
CTGAAAATTATCAAGGACAAGGACTTCCTGGA CAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACC
CTGAC ACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGC CCACCTGTTCGACGACAAAGT

GATGAAGCAGCTGAAGCGGCGGAGATA CACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGA
CAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTC GCCAACAGAAACTTCATGCAGCTGAT
CCACGACGACAGCCTGACCTTTA AAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGC ACG
AGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCAT CCTGCAGACAGTGAAGGTGGTGGACGAGC
TCGTGAAAGTGATGGGCCG GCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGAC CACCCAG
AAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCG AAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTG
AAAGAACACCCCG TGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCA GAATGGGCGGGA
TATGTACGTGGACCAGGAACTGGACATCAACCGGCT GTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTT
TCTGAAGGAC GACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGC AAGAGCGACAACGTGC
CCTCCGAAGAGGTCGTGAAGAAGATGAAGAAC TACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAA
AGTTCG ACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGG CCGGCTTCATCAAGAGACAG
CTGGTGGAAACCCGGCAGATCACAAAGC ACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGA
ATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCT GGTGTCCGATTTCCGGAAGGATTTCC
AGTTTTACAAAGTGCGCGAGATC AACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAA CC
GCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGG CGACTACAAGGTGTACGACGTGCGGAAG
ATGATCGCCAAGAGCGAGCA GGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATG AACTT
TTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGC GGCCTCTGATCGAGACAAACGGCGAAACCGG
GGAGATCGTGTGGGATA AGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGT GAATATCG
TGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGA GTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGA
TCGCCAGAAAGAA GGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCC TATTCTGTGCTG
GTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAA CTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATG
GAAAGAAGC AGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAG AAGTGAAAAAGGACCT
GATCATCAAGCTGCCTAAGTACTCCCTGTTCGA GCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGA
ACTGCA GAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTAC CTGGCCAGCCACTATGAGA
AGCTGAAGGGCTCCCCCGAGGATAATGAG CAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGA
TC ATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTA ATCTGGACAAAGTGCTGTCCGCC
TACAACAAGCACCGGGATAAGCCCAT CAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTG
GGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGA GGTACACCAGCACCAAAGAGGTGCTG
GACGCCACCCTGATCCACCAGA GCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGG CGA
CAGCCCCAAGAAGAAGAGAAAGGTGGAGGCCAGCTAACATATGAT TCGAATGTCTTTCTTGCGCTATGACACTTC
CAGCAAAAGGTAGGGCGGG CTGCGAGACGGCTTCCCGGCGCTGCATGCAACACCGATGATGCTTCGAC CCCCCG
AAGCTCCTTCGGGGCTGCATGGGCGCTCCGATGCCGCTCCAGG GCGAGCGCTGTTTAAATAGCCAGGCCCCCGAT
TGCAAAGACATTATAGC GAGCTACCAAAGCCATATTCAAACACCTAGATCACTACCACTTCTACAC AGGCCACT
CGAGCTTGTGATCGCACTCCGCTAAGGGGGCGCCTCTTCCT CTTCGTTTCAGTCACAACCCGCAAACATGACACA
AGAATCCCTGTTACTT CTCGACCGTATTGATTCGGATGATTCCTACGCGAGCCTGCGGAACGACC AGGAATTCT
GGGAGGTGAGTCGACGAGCAAGCCCGGCGGATCAGGCAG CGTGCTTGCAGATTTGACTTGCAACGCCCGCATTGT
GTCGACGAAGGCT TTTGGCTCCTCTGTCGCTGTCTCAAGCAGCATCTAACCCTGCGTCGCCGT TTCCATTTGCA
GCCGCTGGCCCGCCGAGCCCTGGAGGAGCTCGGGCTGC CGGTGCCGCCGGTGCTGCGGGTGCCCGGCGAGAGCAC
CAACCCCGTACT GGTCGGCGAGCCCGGCCCGGTGATCAAGCTGTTCGGCGAGCACTGGTGC GGTCCGGAGAGCC

TCGCGTCGGAGTCGGAGGCGTACGCGGTCCTGGCG GACGCCCCGGTGCCGGTGCCCCGCCTCCTCGGCCGCGGCG AGCTGCGGC CCGGCACCGGAGCCTGGCCGTGGCCCTACCTGGTGATGAGCCGGATGAC CGGCACCACCTGGCGG TCCGCGATGGACGGCACGACCGACCGGAACGC GCTGCTCGCCCTGGCCCGCGAACTCGGCCGGGTGCTCGGCCGG CTGCAC AGGGTGCCGCTGACCGGGAACACCGTGCTCACCCCCCATTCCGAGGTCT TCCCGGAACTGCTGCGGGA ACGCCGCGCGGCGACCGTCGAGGACCACC GCGGGTGGGGCTACCTCTCGCCCCGGCTGCTGGACCGCCTGGAGGA CTG GCTGCCGGACGTGGACACGCTGCTGGCCGGCCGCGAACCCCGGTTCGTC CACGGCGACCTGCACGGGACCA ACATCTTCGTGGACCTGGCCGCGACCG AGGTCACCGGGATCGTCGACTTCACCGACGTCTATGCGGGAGACTCCCG CTACAGCCTGGTGCAACTGCATCTCAACGCCTTCCGGGGCGACCGCGAG ATCCTGGCCGCGCTGCTCGACGGGGC GCAGTGGAAGCGGACCGAGGAC TTCGCCCGCGAACTGCTCGCCTTCACCTTCCTGCACGACTTCGAGGTGTT CG AGGAGACCCCGCTGGATCTCTCCGGCTTCACCGATCCGGAGGAACTG GCGCAGTTCCTCTGGGGGCCGCCGGACA CCGCCCCCGGCGCCTGATAAG GATCCGGCAAGACTGGCCCCGCTTGGCAACGCAACAGTGAGCCCCTCCC TAGT GTGTTTGGGGATGTGACTATGTATTCGTGTGTTGGCCAACGGGTCA ACCCGAACAGATTGATACCCGCCTTGGCA TTTCCTGTCAGAATGTAACG TCAGTTGATGGTACT

[0492] 对于所有修饰的莱茵衣藻细胞而言,申请人使用了PCR、SURVEYOR核酸酶测定和DNA测序,以验证成功的修饰。

实例16:在细菌中使用Cas9作为转录阻遏子

[0493] 人工控制转录的能力对于研究基因功能以及具有希望的特性的合成 基因网络的构建两者是至关重要的。申请人在此处描述了使用RNA指导的 Cas9蛋白作为一种可编程的转录阻遏子。

[0494] 申请人先前已表明如何使用化脓链球菌SF370的Cas9蛋白在肺炎 链球菌中引导基因组编辑。在此研究中,申请人工程化包含一个最低限度CRISPR系统的crR6Rk菌株,该系统由以下各项组成:cas9、tracrRNA以及 一个重复。将D10A-H840突变引入至此菌株中的cas9中,给出菌株 crR6Rk**。靶向bgaAβ-半乳糖苷酶基因启动子的不同位置的四个间隔子被克 隆到之前所述的pDB98质粒携带的CRISPR阵列中。申请人观察到β-半乳糖 苷酶活性的X至Y倍降低依赖于该靶向位置,表明Cas9作为一种可编程的阻 遏子的可能性(图73)。

[0495] 为了实现大肠杆菌中的Cas9**阻遏,构建一种绿色荧光蛋白 (GFP)报告基因质粒(pDB127)以从组成型启动子表达gfpmut2基因。设 计该启动子以在两条链上携带若干NPPPAM,以测量结合在不同位置的 Cas9**的影响。申请人将D10A-H840突变引入至pCas9中,针对新间隔子的 简单克隆设计携带该tracrRNA、cas9以及一个最低限度CRISPR阵列的、所描述的质粒。设计二十二个不同的间隔子以靶向gfpmut2启动子以及开放阅读 框的不同区域。在靶向重叠或邻近于-35以及-10启动子元件以及邻近于该夏 因-达尔加诺序列(Shine-Dalgarno sequence)的区域后观察到荧光的大约20 倍减少。在两条链上的靶标显示类似的阻遏水平。这些结果表明,Cas9**与 该启动子区域的任何位置的结合防止了转录起始,这大概是通过RNAP结合 的空间抑制实现的。

[0496] 为了确定Cas9**是否可防止转录延伸,申请人将其定向到gpfmut2 的阅读框。在靶向的编码和非编码链两者上观察到萦光的减少,表明Cas9结 合实际足够强以代表对运行RNAP的障碍。然而,而当该编码链是靶标时观 察到表达的40%减少,对于非编码链观察到20倍降低(图74b-74c,将T9、 T10以及T11与B9、B10以及B11进行比较)。为了直接确定Cas9**结合对 转录的影响,申请人从携带该T5、T10、B10或一个不靶向pDB127的对照构 建

体的菌株提取RNA,并且使用在该B10以及T10靶位点之前(B477)或之 后(B510)结合的一种探针使其经受Northern印迹分析。与申请人的萦光方 法一致,当Cas9**被定向至该启动子区域(T5靶标)时未检测到gfpmut2转 录并且在该T10区域的靶向之后观察到转录。有意思地是,用该B477探针观 察到更小的转录物。此条带对应于期望尺寸的会被Cas9**中断的一种转录 物,并且是由结合到该编码链的dgRNA::Cas9**引起的转录终止的直接显 示。令人惊讶地,当靶向该非编码链(B10)时申请人未检测到转录物。由于 结合到该B10区域的Cas9**不可能干扰转录起始,此结果表明该mRNA退 化。DgRNA::Cas9显示体外结合ssRNA。申请人推测,结合可通过宿主核酸 酶引发mRNA的降解。的确,核糖体停顿可诱导在大肠杆菌中翻译的mRNA 的切割。

[0497] 一些应用需要精确协调基因表达而不是完全阻遏。申请人寻求通过 引入将弱化crRNA/靶相互作用的错配,实现中间阻遏水平。申请人建立基于B1、T5以及B10构建体的一系列的间隔子,这些构建体在crRNA的5'端具 有增加数目的突变。B1以及T5中高达8个突变不影响该阻遏水平,并且针 对额外的突变观察到荧光的渐进增加。

[0498] 以crRNA及其靶标之间仅8nt匹配观察到的阻遏增加了使用 Cas9**作为转录调节子的脱靶作用的问题。由于对于Cas9结合也需要良好的 PAM(NGG),匹配以获得某一水平的呼吸作用的核苷酸的数目是10。10nt 匹配每隔约1Mbp随机发生一次,并且因此甚至很可能在小的细菌性基因组 中发现此类位点。然而,为了有效地阻遏转录,此类位点需要在基因的启动 子区域中,其使得更少可能脱靶。申请人也显示基因表达可受影响,如果靶 向一种基因的非编码链的话。为了让此发生,随机的靶标必须处于正确取 向,但此类事件相对更可能发生。事实上,在此研究的进程中,申请人不能 够构建在pCas9**上设计的间隔子之一。申请人之后发现此间隔子在必需的 murC基因中显示邻接良好的PAM的12bp匹配。此类脱靶可容易地通过所 设计的间隔子的系统爆炸避免。

[0499] 在以下编号的段落中进一步说明本发明的多个方面:

1.一种包括一个或多个载体的载体系统,其中该系统包括

a.一个第一调节元件,该第一调节元件可操作地连接到一个traer配 对序列以及用于在该traer配对序列的上游插入一个指导序列的一个或多个插 入位点,其中当表达时,该指导序列引导CRISPR复合物在真核细胞中与一 个靶序列的序列特异性结合,其中该CRISPR复合物包括与以下各项复合的 一种CRISPR酶:(1)杂交到该靶序列的指导序列,以及(2)杂交到该 traer序列的traer配对序列;和

b.一个第二调节元件,该第二调节元件可操作地连接到编码所述 CRISPR酶的酶编码序列上,该CRISPR酶包含一个核定位序列;

其中组分(a)和(b)位于该系统的相同或不同载体上。

2.如段落1所述的载体系统,其中组分(a)进一步包括在该第一调 节元件的控制之下在traer配对序列的下游的traer序列。

3.如段落1所述的载体系统,其中组分(a)进一步包括可操作地连 接到该第一调节元件的两个或更多个指导序列,其中当表达时,该两个或更 多个指导序列中的每个引导CRISPR复合物在真核细胞中与一个不同靶序列 的序列特异性结合。

4.如段落1所述的载体系统,其中该系统包括在一个第三调节元件控 制之下的traer序列。

5.如段落1所述的载体系统,其中当进行最佳比对时,沿着该traer 配对序列的长度,该traer序列展现至少50%序列互补性。

6.如段落1所述的载体系统,其中该CRISPR酶包括一个或多个核定 位序列,该一个或多个核定位序列具有足够强度来在真核细胞的细胞核中驱 动所述CRISPR酶以可检测到的量积聚。

7.如段落1所述的载体系统,其中该CRISPR酶是一种II型CRISPR 系统酶。

8.如段落1所述的载体系统,其中该CRISPR酶是一种Cas9酶。

9.如段落1所述的载体系统,其中该CRISPR酶进行了密码子优化以 便在真核细胞中表达。

10.如段落1所述的载体系统,其中该CRISPR酶引导在该靶序列位置 处的一条或两条链的切割。

11.如段落1所述的载体系统,其中该CRISPR酶缺少DNA链切割活 性。

12.如段落1所述的载体系统,其中该第一调节元件是一种聚合酶III 启动子。

13.如段落1所述的载体系统,其中该第二调节元件是一种聚合酶II启 动子。

14.如段落4所述的载体系统,其中该第三调节元件是一种聚合酶III 启动子。

15.如段落1所述的载体系统,其中该指导序列在长度上是至少15个 核苷酸。

16.如段落1所述的载体系统,其中在最佳地折叠时,该指导序列的少 于50%的核苷酸参与自我互补碱基配对。

17.一种载体,包括一个调节元件,该调节元件可操作地连接到编码 CRISPR酶的酶编码序列上,该CRISPR酶包含一个或多个核定位序列,其中 所述调节元件驱动真核细胞中CRISPR酶的转录,使得所述CRISPR酶以可检 测到的量在该真核细胞的细胞核中积聚。

18.如段落17所述的载体,其中所述调节元件是一种聚合酶II启动 子。

19.如段落17所述的载体,其中所述CRISPR酶是一种II型CRISPR 系统酶。

20.如段落17所述的载体,其中所述CRISPR酶是一种Cas9酶。

21.如段落17所述的载体,其中所述CRISPR酶缺少切割它结合至其 上的靶序列的一条或多条链的能力。

22.一种CRISPR酶,该酶包括一个或多个核定位序列,该一个或多个 核定位序列具有足够强度来在真核细胞的细胞核中驱动所述CRISPR酶以可 检测到的量积聚。

23.如段落22所述的CRISPR酶,其中所述CRISPR酶是一种II型CRISPR系统酶。

24.如段落22所述的CRISPR酶,其中所述CRISPR酶是一种Cas9 酶。

25.如段落22所述的CRISPR酶,其中所述CRISPR酶缺少切割它结 合至其上的靶序列的一条或多条链的能力。

26.一种真核宿主细胞,包括:

a.一个第一调节元件,该第一调节元件可操作地连接到一个traer配 对序列以及用于在该traer配对序列的上游插入一个指导序列的一个或多个插 入位点,其中当表达时,该指导序列引导CRISPR复合物在真核细胞中与一 个靶序列的序列特异性结合,其中该CRISPR复合物包括与以下各项复合的 一种CRISPR酶:(1)杂交到该靶序列的指导序列,以及(2)杂交到该 traer序列的traer配对序列;和/或

b.一个第二调节元件,该第二调节元件可操作地连接到编码所述 CRISPR酶的酶

编码序列上,该CRISPR酶包含一个核定位序列。

27.如段落26所述的真核宿主细胞,其中所述宿主细胞包括组分(a) 以及(b)。

28.如段落26所述的真核宿主细胞,其中组分(a)、组分(b)、或 组分(a)和(b)稳定地整合到该宿主真核细胞的一个基因组中。

29.如段落26所述的真核宿主细胞,其中组分(a)进一步包括在该第 一调节元件的控制之下在traer配对序列的下游的traer序列。

30.如段落26所述的真核宿主细胞,其中组分(a)进一步包括可操作 地连接到该第一调节元件的两个或更多个指导序列,其中当表达时,该两个 或更多个指导序列中的每个引导CRISPR复合物在真核细胞中与一个不同靶 序列的序列特异性结合。

31.如段落26所述的真核宿主细胞,进一步包括可操作地连接到所述 traer序列的一个第三调节元件。

32.如段落26所述的真核宿主细胞,其中当进行最佳比对时,沿着该 traer配对序列的长度,该traer序列展现至少50%序列互补性。

33.如段落26所述的真核宿主细胞,其中该CRISPR酶包括一个或多 个核定位序列,该一个或多个核定位序列具有足够强度来在真核细胞的细胞 核中驱动所述CRISPR酶以可检测到的量积聚。

34.如段落26所述的真核宿主细胞,其中该CRISPR酶是一种II型 CRISPR系统酶。

35.如段落26所述的真核宿主细胞,其中该CRISPR酶是一种Cas9 酶。

36.如段落26所述的真核宿主细胞,其中该CRISPR酶进行了密码子 优化以便在真核细胞中表达。

37.如段落26所述的真核宿主细胞,其中该CRISPR酶引导在该靶序 列位置处的一条或两条链的切割。

38.如段落26所述的真核宿主细胞,其中该CRISPR酶缺少DNA链切 割活性。

39.如段落26所述的真核宿主细胞,其中该第一调节元件是一种聚合 酶III启动子。

40.如段落26所述的真核宿主细胞,其中该第二调节元件是一种聚合 酶II启动子。

41.如段落31所述的真核宿主细胞,其中该第三调节元件是一种聚合 酶III启动子。

42.如段落26所述的真核宿主细胞,其中该指导序列在长度上是至少 15个核苷酸。

43.如段落26所述的真核宿主细胞,其中在最佳地折叠时,该指导序 列的少于50%的核苷酸参与自我互补碱基配对。

44.一种非人的动物,该动物包括段落26-43中任一项所述的真核宿主 细胞。

45.一种试剂盒,包括一种载体系统以及用于使用所述试剂盒的说明 书,该载体系统包括:

a.一个第一调节元件,该第一调节元件可操作地连接到一个traer配 对序列以及用于在该traer配对序列的上游插入一个指导序列的一个或多个插 入位点,其中当表达时,该指导序列引导CRISPR复合物在真核细胞中与一 个靶序列的序列特异性结合,其中该

CRISPR复合物包括与以下各项复合的 一种CRISPR酶:(1)杂交到该靶序列的指导序列,以及(2)杂交到该 traer序列的traer配对序列;和/或

b.一个第二调节元件,该第二调节元件可操作地连接到编码所述 CRISPR酶的酶编码序列上,该CRISPR酶包含一个核定位序列。

46.如段落45所述的试剂盒,其中所述试剂盒包括位于该系统的相同 或不同载体上的组分(a)和(b)。

47.如段落45所述的试剂盒,其中组分(a)进一步包括在该第一调节 元件的控制之下在traer配对序列的下游的traer序列。

48.如段落45所述的试剂盒,其中组分(a)进一步包括可操作地连接 到该第一调节元件的两个或更多个指导序列,其中当表达时,该两个或更多 个指导序列中的每个引导CRISPR复合物在真核细胞中与一个不同靶序列的 序列特异性结合。

49.如段落45所述的试剂盒,其中该系统包括在一个第三调节元件控 制之下的traer序列。

50.如段落45所述的试剂盒,其中当进行最佳比对时,沿着该traer配 对序列的长度,该traer序列展现至少50%序列互补性。

51.如段落45所述的试剂盒,其中该CRISPR酶包括一个或多个核定 位序列,该一个或多个核定位序列具有足够强度来在真核细胞的细胞核中驱 动所述CRISPR酶以可检测到的量积聚。

52.如段落45所述的试剂盒,其中该CRISPR酶是一种II型CRISPR 系统酶。

53.如段落45所述的试剂盒,其中该CRISPR酶是一种Cas9酶。

54.如段落45所述的试剂盒,其中该CRISPR酶进行了密码子优化以 便在真核细胞中表达。

55.如段落45所述的试剂盒,其中该CRISPR酶引导在该靶序列位置 处的一条或两条链的切割。

56.如段落45所述的试剂盒,其中该CRISPR酶缺少DNA链切割活 性。

57.如段落45所述的试剂盒,其中该第一调节元件是一种聚合酶III启 动子。

58.如段落45所述的试剂盒,其中该第二调节元件是一种聚合酶II启 动子。

59.如段落49所述的试剂盒,其中该第三调节元件是一种聚合酶III启 动子。

60.如段落45所述的试剂盒,其中该指导序列在长度上是至少15个核 苷酸。

61.如段落45所述的试剂盒,其中在最佳地折叠时,该指导序列的少 于50%的核苷酸参与自我互补碱基配对。

62.一种计算机系统,用于在通过一种CRISPR复合物进行靶向的真核 细胞中的核酸序列内选择候选靶序列,该系统包括:

a.一个存储单元,被配置成接受和/或储存所述核酸序列;以及

b.一个或多个单独或组合的处理器,被编码为(i)将一个CRISPR基 序序列定位于所述核酸序列内,以及(ii)选择邻近于所述定位的CRISPR基 序序列的一个序列作为该CRISPR复合物结合至其上的候选靶序列。

63.如段落62所述的计算机系统,其中所述定位步骤包括鉴定与所述 靶序列距离少于约500个核苷酸定位的一个CRISPR基序序列。

64.如段落62所述的计算机系统,其中所述候选靶序列在长度上是至 少10个核苷酸。

65.如段落62所述的计算机系统,其中在该候选靶序列的3'端的核苷 酸被定位成离该CRISPR基序序列上游不超过约10个核苷酸。

66.如段落62所述的计算机系统,其中该真核细胞中的核酸序列对该 真核基因组来说是内源的。

67.如权利要求62所述的计算机系统,其中该真核细胞中的核酸序列 对该真核基因组来说是外源的。

68.一种计算机可读介质,包括如下编码,所述编码在通过一个或多个 处理器执行时,实施在通过一种CRISPR复合物进行靶向的真核细胞中的核 酸序列内选择候选靶序列的方法,所述方法包括:(a)将一个CRISPR基序 序列定位于所述核酸序列内,以及(b)选择邻近于所述定位的CRISPR基序 序列的一个序列作为该CRISPR复合物结合至其上的候选靶序列。

69.如段落68所述的计算机可读介质,其中所述定位包括与所述靶序 列距离少于约500个核苷酸来定位一个CRISPR基序序列。

70.如段落68所述的计算机可读物,其中所述候选靶序列在长度上是 至少10个核苷酸。

71.如段落68所述的计算机可读物,其中在该候选靶序列的3'端的核 苷酸被定位成离该CRISPR基序序列上游不超过约10个核苷酸。

72.如段落68所述的计算机可读物,其中该真核细胞中的核酸序列对 该真核基因组来说是内源的。

73.如段落68所述的计算机可读物,其中该真核细胞中的核酸序列对 该真核基因组来说是外源的。

74.在真核细胞中修饰靶多核苷酸的一种方法,该方法包括允许一种 CRISPR复合物结合到该靶多核苷酸上以实现所述靶多核苷酸的切割,由此修 饰该靶多核苷酸,其中该CRISPR复合物包括与杂交到所述靶多核苷酸内的 一个靶序列上的指导序列复合的CRISPR酶,其中所述指导序列连接到一个 traer配对序列上,该traer配对序列进而杂交到一个traer序列上。

75.如段落74所述的方法,其中所述切割包括通过所述CRISPR酶切 割在该靶序列位置处的一条或两条链。

76.如段落74所述的方法,其中所述切割导致靶基因的转录降低。

77.如段落74所述的方法,进一步包括通过与外源模板多核苷酸同源 重组修复所述切割的靶多核苷酸,其中所述修复导致一个突变,该突变包括 所述靶多核苷酸的一个或多个核苷酸的插入、缺失、或取代。

78.如段落77所述的方法,其中所述突变在由包含该靶序列的基因表 达的蛋白质中导致一个或多个氨基酸改变。

79.如段落74所述的方法,进一步包括将一个或多个载体递送到所述 真核细胞,其中该一个或多个载体驱动以下一项或多项的表达:CRISPR酶、连接到traer配对序列上的指导序列、以及traer序列。

80.如段落79所述的方法,其中所述载体被递送到受试者内的真核细 胞中。

81.如段落74所述的方法,其中所述修饰发生在细胞培养物中的所述 真核细胞中。

82.如段落74所述的方法,进一步包括在所述修饰之前从受试者中分 离所述真核细胞。

83.如段落82所述的方法,进一步包括使所述真核细胞和/或从中衍生 的细胞返回到所述受试者中。

84.一种修饰多核苷酸在真核细胞中的表达的方法,该方法包括允许一 种CRISPR复合物结合到该多核苷酸上,使得所述结合导致所述多核苷酸表 达增加或减少;其中该CRISPR复合物包含与杂交到所述靶多核苷酸内的一 个靶序列上的指导序列复合的CRISPR酶,其中所述指导序列连接到一个 traer配对序列上,该traer配对序列进而杂交到一个traer序列上。

85.如段落74所述的方法,进一步包括将一个或多个载体递送到所述 真核细胞,其中该一个或多个载体驱动以下一项或多项的表达:CRISPR酶、连接到traer配对序列上的指导序列、以及traer序列。

86.产生包括一个突变疾病基因的模式真核细胞的方法,该方法包括:

a.将一个或多个载体引入一个真核细胞中,其中该一个或多个载体驱 动以下一项或多项的表达:CRISPR酶、连接到traer配对序列上的指导序 列、以及traer序列;并且

b.允许一种CRISPR复合物结合到一个靶多核苷酸上以实施在所述疾 病基因内的该靶多核苷酸的切割,其中该CRISPR复合物包含与(1)杂交到 该靶多核苷酸内的靶序列上的指导序列、和(2)杂交到该traer上的该traer 配对序列复合的CRISPR酶,由此产生包含突变的疾病基因的模式真核细 胞。

87.如段落86所述的方法,其中所述切割包括通过所述CRISPR酶切 割在该靶序列位置处的一条或两条链。

88.如段落86所述的方法,其中所述切割导致靶基因的转录降低。

89.如段落86所述的方法,进一步包括通过与外源模板多核苷酸同源 重组修复所述切割的靶多核苷酸,其中所述修复导致一个突变,该突变包括 所述靶多核苷酸的一个或多个核苷酸的插入、缺失、或取代。

90.如段落89所述的方法,其中所述突变在由包含该靶序列的基因表 达的蛋白质中导致一个或多个氨基酸改变。

91.一种研发生物活性剂的方法,该生物活性剂调制与疾病基因相关的 细胞信号传导事件,该方法包括:

a.使一种测试化合物与段落86-90中任一项所述的一种模式细胞接 触;并且

b.检测读数变化,该变化指示与所述疾病基因的所述突变关联的细胞 信号传导事件的减少或增加,从而开发调制与所述疾病基因关联的所述细胞 信号传导事件的所述生物活性剂。

92.一种重组多核苷酸,包括在traer配对序列上游的指导序列,其中 该指导序列当表达时引导一种CRISPR复合物与存在于真核细胞中的一个对 应的靶序列的序列特异性结合。

93.如段落89所述的重组多核苷酸,其中该靶序列是存在于真核细胞 中的病毒序列。

94.如段落89所述的重组多核苷酸,其中该靶序列是原癌基因或癌基 因。

[0500] 虽然已经在此示出并描述了本发明的优选实施例,但是对本领域的 普通技术人员而言应该显而易见是这样的实施例仅以举例方式提供。在不偏 离本发明的情况下众多变化、改变和取代现在将被本领域的普通技术人员想 到。应该理解的是,在实践本发明中可以采用在此描述的本发明的实施例的 不同替代方案。旨在按照以下权利要求限定本发明的范围,并且在这些权利 要求范围内的方法和结构及其等同物也包括在本发明的范围内。

[0501] 参考文献:

1.乌尔诺夫,F.D.(Urnov,F.D.)、雷巴尔,E.J.(Rebar,E.J.)、福尔摩斯,M.C.(Holmes,M.C.)、张,H.S.(Zhang,H.S.)&格雷戈里, P.D.(Gregory,P.D.)用工程化的锌指核酸酶进行的基因组编辑(Genome editing with engineered zinc finger nucleases).《遗传学自然评论》(Nat Rev Genet)11,636-646(2010)。

2.保达诺夫,A.J.(Bogdanove,A.J.)&沃伊塔斯,D.F.(Voytas, D.F.)TAL效应物:用于DNA靶向的可定制蛋白(TAL effectors: customizable proteins for DNA targeting).《科学》(Science)333,1843-1846 (2011)。

3.斯托达德,B.L.(Stoddard,B.L.)归巢内切核酸酶结构以及功能 (Homing endonuclease structure and function).《生物物理学季评》(Q.Rev. Biophys.)38,49-95(2005)。

4.贝,T.(Bae,T.)&施内温德,O.(Schneewind,O.)用诱导型反 选择在金黄色葡萄球菌中进行的等位基因置换(Allelic replacement in Staphylococcus aureus with inducible counter-selection).《质粒》(Plasmid) 55,58-63(2006)。

5.孙,C.K.(Sung,C.K.),李,H.(Li,H.),克拉维瑞斯,J.P. (Claverys,J.P.)&莫里森,D.A.(Morrison,D.A.)一种rpsL盒,janus, 用于通过阴性筛选在肺炎链球菌中进行基因置换(An rpsL cassette,janus,for gene replacement through negative selection in Streptococcus pneumoniae).《应 用与环境微生物学》(Appl.Environ.Microbiol).67,5190-5196(2001)。

6.沙兰,S.K.(Sharan,S.K.),托马森,L.C.(Thomason,L.C.), 库兹涅佐夫,S.G.(Kuznetsov,S.G.)&考特,D.L.(Court,D.L.)重组工 程:基于同源重组的基因工程方法(Recombineering:a homologous recombination-based method of genetic engineering).《自然方案》(Nat. Protoc.)4,206-223(2009)。

7.季聂克,M.(Jinek,M.)等人获得性细菌免疫中的可编程的双 RNA指导的DNA内切核酸酶(A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity).《科学》(Science)337,816- 821(2012)。

8.德沃,H.(Deveau,H.)、加尔诺,J.E.(Garneau,J.E.)&莫万 诺,S.(Moineau,S.)CRISPR-Cas系统及其在噬菌体细菌相互作用中的作用 (CRISPR-Cas system and its role in phage-bacteria interactions).《微生物学年 评》(Annu Rev Microbiol)64,475-493(2010)。

9.霍瓦特,P.(Horvath,P.)&巴郎镐,R.(Barrangou,R.) CRISPR-Cas,细菌和古生菌的免疫系统(CRISPR-Cas,the immune system of bacteria and archaea).《科学》(Science)327,167-170(2010)。

10.特恩斯,M.P.(Terns,M.P.)&特恩斯,R.M.(Terns,R.M.)基 于CRISPR的获得性免疫系统(CRISPR-based adaptive immune systems).《微 生物学现行观点》(Curr.Opin.Microbiol.)14,321-327(2011)。

11.范德奥斯特,J.,(van der Oost,J.,)捷雷,M.M.(Jore, M.M.),韦斯特拉,E.R.(Westra,E.R.),伦德格伦,M.(Lundgren, M.)&布龙斯,S.J.(Brouns,S.J.)在原核生物中基于CRISPR的获得性以 及可遗传性免疫(CRISPR-based adaptive and heritable immunity in prokaryotes).《生物化学科学趋势》(Trends.Biochem.Sci.)34,401-407(2009)。

12.布龙斯,S.J.(Brouns,S.J.)等人,在原核生物中小CRISPR RNA 指导抗病毒防御(Small CRISPR RNAs guide antiviral defense in prokaryotes).《科学》(Science)321,960-964(2008)。

13.卡特,J.(Carter,J.),王,R.(Wang,R.),李,H.(Li,H.), 特恩斯,R.M.(Terns,R.M.)&特恩斯,M.P.(Terns,M.P.)Cas6是在原核 生物中产生指导RNA用于入侵物防御的内切核糖核酸酶(Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes).《基因发育》(Genes Dev.)22,3489-3496(2008)。

14.德尔特科娃,E.(Deltcheva,E.)等人通过反式编码的小RNA和 宿主因子RNase III熟化CRISPR RNA(CRISPR RNA maturation by trans- encoded small RNA and host factor RNase III).《自然》(Nature)471,602- 607(2011)。

15.哈托姆-阿斯兰,A.(Hatoum-Aslan,A.),曼伊夫,I.(Maniv, I.)&马拉菲尼,L.A.(Marraffini,L.A.)成熟的规律间隔成簇短回文重复序 列RNA(crRNA)长度是通过在前体加工位点锚定的标尺机构测量的 (Mature clustered,regularly interspaced,short palindromic repeats RNA(crRNA) length is measured by a ruler mechanism anchored at the precursor processing site).《美国国家科学院院刊》(Proc.Natl.Acad.Sci.USA.)108,21218-21222 (2011)。

16.华维茨,R.E.(Haurwitz,R.E.),吉耐克,M.(Jinek,M.), M.,威登赫夫特,B.(Wiedenheft,B.),周,K.(Zhou,K.)&杜德那, J.A.(Doudna,J.A.)通过CRISPR内切核酸酶进行序列-以及结构-特异性 RNA加工(Sequence-and structure-specific RNA processing by a CRISPR endonuclease).《科学》(Science)329,1355-1358(2010)。

17.德沃,H.(Deveau,H.)等人,在嗜热链球菌中对CRISPR编码的 抗性的噬菌体响应(Phage response to CRISPR-encoded resistance in Streptococcus thermophilus).《细菌学杂志》(J.Bacteriol.)190,1390-1400 (2008)。

18.加斯诺斯,G.(Gasiunas,G.)、巴郎镐,R.(Barrangou,R.)、 霍瓦特,P.(Horvath,P.)&西克斯尼斯(Siksnys,V.)Cas9-crRNA核糖核 蛋白复合体在细菌中针对获得性免疫介导特异的DNA切割(Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria).《美国国家科学院院刊》(Proc.Natl.Acad.Sci.USA.),E2579- 2586(2012)。

19.马卡洛夫,K.S.(Makarova,K.S.),阿韦德,L.(Aravind, L.),沃尔夫,Y.I.(Wolf,Y.I.)&库宁,E.V.(Koonin,E.V.)Cas蛋白家 族的统一以及CRISPR-Cas系统的起源和进化的简单情境(Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems).《生物学导向》(Biol.Direct.)6,38(2011)。

20.巴郎镐,R.(Barrangou,R.)RNA介导的可编程DNA切割 (RNA-mediated programmable DNA cleavage).《自然生物技术》(Nat. Biotechnol.)30,836-838(2012)。

21.布龙斯,S.J.(Brouns,S.J.)分子生物学.免疫学瑞士军刀 (Molecular biology.A Swiss army knife of immunity).《科学》(Science) 337,808-809(2012)。

22.卡罗尔,D.(Carroll,D.)基因靶向的一种CRISPR途径(A CRISPR Approach to Gene Targeting).《分子疗法》(Mol.Ther.)20,1658- 1660(2012)。

23.碧卡尔德,D.(Bikard,D.),哈托姆-阿斯兰,A.(Hatoum- Aslan,A.),穆奇达,D.(Mucida,D.)&马拉菲尼,L.A.(Marraffini, L.A.)CRISPR干扰可防止在体内细菌感染过程中的天然转化和毒力获取 (CRISPR interference can prevent natural transformation and virulence acquisition during in vivo bacterial infection).《细胞宿主微生物》(Cell Host Microbe)12,177-186(2012)。

24.萨普拉诺斯卡斯(Sapranauskas,R.)等人嗜热链球菌CRISPR-Cas 系统在大肠杆菌中提供免疫力(The Streptococcus thermophilus CRISPR-Cas system provides immunity in Escherichia coli).《核酸研究》(Nucleic Acids Res)(2011)。

25.塞梅诺娃,E.(Semenova,E.)等人规律间隔成簇短回文重复序列 (CRISPR)RNA的干扰是由种子序列支配的(Interference by clustered regularly interspaced short palindromic repeat(CRISPR)RNA is governed by a seed sequence).《美国国家科学院院刊》(Proc.Natl.Acad.Sci.USA.) (2011)。

26.威登赫夫特,B.(Wiedenheft,B.)等人来自细菌性免疫系统的 RNA指导复合体通过种子序列相互作用增强靶识别(RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions).《美国国家科学院院刊》(Proc.Natl.Acad.Sci.USA.) (2011)。

27.扎娜,D.(Zahner,D.)&哈肯贝克,R.(Hakenbeck,R.)肺炎链 球菌β-半乳糖苷酶是一种表面蛋白(The Streptococcus pneumoniae beta- galactosidase is a surface protein).《细菌学杂志》(J.Bacteriol.)182,5919- 5921(2000)。

28.马拉夫尼L.A.(Marraffini,L.A.),迪登特,A.C.(Dedent, A.C.)&施内温德,O.(Schneewind,O.)将蛋白锚定到革兰氏阳性菌的包膜 的分选酶和技术(Sortases and the art of anchoring proteins to the envelopes of gram-positive bacteria).《微生物学分子生物学综述》 (Microbiol.Mol.Biol.Rev.)70,192-221(2006)。

29.莫塔梅迪,M.R.(Motamedi,M.R.),兹格提,S.K.(Szigety, S.K.)&罗森伯格,S.M.(Rosenberg,S.M.)大肠杆菌中的双链断裂修复重 组:体内DNA复制机制的物理证据(Double-strand-break repair recombination in Escherichia coli:physical evidence for a DNA replication mechanism in vivo).《基因发育》(Genes Dev.)13,2889-2903(1999)。

30.霍萨卡,T.(Hosaka,T.)等人核糖体蛋白S12中的新的突变 K87E在大肠杆菌的后期生长阶段过程中增强蛋白合成活性(The novel mutation K87E in ribosomal protein S12 enhances protein synthesis activity during the late growth phase in Escherichia coli).《分子遗传学基因组学》(Mol. Genet.Genomics)271,317-324(2004)。

31.科斯坦蒂诺,N.(Costantino,N.)&考特,D.L.(Court,D.L.) 在错配修复突变体中增强水平的λ红介导的重组体(Enhanced levels of lambda Red-mediated recombinants in mismatch repair mutants).《美国国家科学院院 刊》(Proc.Natl.Acad.Sci.USA)100,15748-15753(2003)。

32.埃德加,R.(Edgar,R.)&奇姆龙,U.(Qimron,U.)大肠杆菌 CRISPR系统保护溶素原免受λ溶源化作用以及原噬菌体诱导(The Escherichia coli CRISPR system protects from lambda lysogenization,lysogens, and prophage induction).《细菌学杂志》(J.Bacteriol.)192,6291-6294 (2010)。

33.马拉夫尼L.A.(Marraffini,L.A.)&松塞摩尔,E.J.(Sontheimer, E.J.)在CRISPR RNA指导的免疫过程中的自我辨别对非自我辨别(Self versus non-self discrimination during CRISPR RNA-directed immunity).《自 然》(Nature)463,568-571(2010)。

34.费舍尔,S.(Fischer,S.)等人古生菌免疫系统可检测多个原型间 隔子相邻基序(PAM)以靶向入侵物DNA(An archaeal immune system can detect multiple Protospacer Adjacent Motifs(PAMs)to target invader DNA).《生 物化学杂志》(J.Biol.Chem.)287,33351-33363(2012)。

35.歌德博格斯道蒂尔,S.(Gudbergsdottir,S.)等人当用载体传播的 病毒和质粒基因以及原型间隔子激发时,硫化裂片菌CRISPR/Cas以及 CRISPR/Cmr系统的动态性能(Dynamic properties of the Sulfolobus CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers).《分子微生物学》(Mol.Microbiol.)79, 35-49(2011)。

36.王,H.H.(Wang,H.H.)等人通过共筛选MAGE进行的基因组规 模启动子工程(Genome-scale promoter engineering by coselection MAGE).《自然方法》(Nat Methods)9,591-593(2012)。

37.丛,L.(Cong,L.)等人使用CRISPR-Cas系统的多元基因组工程 化(Multiplex genome engineering using CRISPR-Cas systems).《科学-正在评 论中》(Science In press)(2013)。

38.玛丽,P.(Mali,P.)等人经由Cas9的RNA指导的人类基因组工 程(RNA-guided human genome engineering via Cas9).《科学-正在评论中》(Science In press)(2013)。

39.霍斯金斯,J.(Hoskins,J.)等人细菌肺炎链球菌菌株R6的基因组 (Genome of the bacterium Streptococcus pneumoniae strain R6).《细菌学杂 志》(J.Bacteriol.)183,5709-5717(2001)。

40.哈维尔斯泰因,L.S.(Havarstein,L.S.),库马拉斯瓦米,G. (Coomaraswamy,G.)&莫里森,D.A.(Morrison,D.A.)未修饰的十七肽 信息素诱导针对肺炎链球菌中的遗传

转化的感受态(An unmodified heptadecapeptide pheromone induces competence for genetic transformation in Streptococcus pneumoniae).《美国国家科学院院刊》(Proc.Natl.Acad.Sci.USA.)92,11140-11144(1995)。

41.堀之内,S.(Horinouchi,S.)&维斯布卢姆,B.(Weisblum,B.) 指定诱导型氯霉素抗性的一种质粒pC194的核苷酸序列以及功能性图谱 (Nucleotide sequence and functional map of pC194,a plasmid that specifies inducible chloramphenicol resistance).《细菌学杂志》(J.Bacteriol.)150, 815-825(1982)。

42.荷顿,R.M.(Horton,R.M.)DNA的体外重组和诱变:基因拼接 连同专用的基因(In Vitro Recombination and Mutagenesis of DNA:SOEing Together Tailor-Made Genes.).《分子生物学方法》(Methods Mol.Biol.) 15,251-261(1993)。

43.波德比尔斯基,A.(Podbielski,A.),斯佩莱博格,B. (Spellerberg,B.),沃伊施尼克,M.(Woischnik,M.),波尔(Pohl, B.)&鲁提肯,R.(Lutticken,R.)用于A组链球菌(GAS)中的基因失活以 及表达分析的新系列的质粒载体(Novel series of plasmid vectors for gene inactivation and expression analysis in group A streptococci (GAS)).《基因》 (Gene)177:137-147(1996)。

44.胡斯曼,L.K.(Husmann,L.K.),斯科特,J.R.(Scott,J.R.), 林达尔,G.(Lindahl,G.)&斯坦伯格,L.(Stenberg,L.)Arp蛋白,即M 蛋白家族的一员的表达不足以抑制化脓链球菌的吞噬作用(Expression of the Arp protein,a member of the M protein family,is not sufficient to inhibit phagocytosis of Streptococcus pyogenes).《传染与免疫》(Infection and immunity)63,345-348(1995)。

45.吉布森,D.G.(Gibson,D.G.)等人数十万碱基的DNA分子的酶 拼接(Enzymatic assembly of DNA molecules up to several hundred kilobases),《自然方法》(Nat Methods):6,343-345(2009)。

优选的实施方式:

1.一种非天然存在的或工程化的组合物,包含一种载体系统,该载体系统 包含一个或多个载体,该一个或多个载体包含

I.一个第一调节元件,该第一调节元件可操作地连接到CRISPR-Cas系统 嵌合RNA(chiRNA)多核苷酸序列上,其中该多核苷酸序列包含

(a)一个指导序列,该指导序列能够杂交到真核细胞中的靶序列上,

(b)一个tracr配对序列,和

(c)一个tracr序列,以及

II.一个第二调节元件,该第二调节元件可操作地连接到编码CRISPR酶 的酶编码序列上,在该CRISPR酶的一个末端附近包含至少一个或多个核定 位序列(NLS),

其中(a)、(b)和(c)以5'到3'方向排列,

其中组分I和II位于该系统的相同或不同载体上,

其中在转录时,该tracr配对序列杂交到该tracr序列上,并且该指导序 列引导CRISPR复合物与该靶序列的序列特异性结合,

其中该CRISPR复合物包含与以下各项复合的CRISPR酶:(1)杂交到 该靶序列上的指导序列、和(2)杂交到该tracr序列上的tracr配对序列,并 且

其中该嵌合RNA多核苷酸序列包括两个或更多个发夹。

2.如项目1所述的组合物,其中使用多个chiRNA多核苷酸序列以提 供一种多元系统。

3.一种多元CRISPR酶系统,其中该系统包含一个载体系统,该载体 系统包含一个或多个载体,该一个或多个载体包含

I.一个第一调节元件,该第一调节元件可操作地连接到CRISPR-Cas系统 嵌合RNA(chiRNA)多核苷酸序列上,其中该多核苷酸序列包含

(a)一个指导序列,该指导序列能够杂交到真核细胞中的靶序列上,

(b)一个tracr配对序列,和

(c)一个tracr序列,以及

II.一个第二调节元件,该第二调节元件可操作地连接到编码CRISPR酶 的酶编码序列上,在该CRISPR酶的一个末端附近包含至少一个或多个核定 位序列(NLS),

其中(a)、(b)和(c)以5'到3'方向排列,

其中组分I和II位于该系统的相同或不同载体上,

其中在转录时,该tracr配对序列杂交到该tracr序列上,并且该指导序 列引导CRISPR复合物与该靶序列的序列特异性结合,

其中该CRISPR复合物包含与以下各项复合的CRISPR酶:(1)杂交到 该靶序列上的指导序列、和(2)杂交到该tracr序列上的tracr配对序列,

其中该chiRNA多核苷酸序列包括两个或更多个发夹,并且

其中在该多元系统中使用多个chiRNA多核苷酸序列。

4.如项目1、2或3所述的组合物或系统,其中该第一调节元件是一 种聚合酶III启动子。

5.如前述项目中任一项所述的组合物或系统,其中该第二调节元件是 一种聚合酶II启动子。

6.如前述项目中任一项所述的组合物或系统,其中该CRISPR酶包括 一个或多个NLS,该一个或多个NLS具有足够强度来在真核细胞的细胞核中 驱动所述CRISPR酶以可检测到的量积聚。

7.如前述项目中任一项所述的组合物或系统,其中当进行最佳比对 时,沿着该tracr配对序列的长度,该tracr序列展现至少50%序列互补性。

8.如前述项目中任一项所述的组合物或系统,其中该CRISPR酶是一 种II型CRISPR系统酶。

9.如前述项目中任一项所述的组合物或系统,其中该CRISPR酶是一 种Cas9酶。

10.如前述项目中任一项所述的组合物或系统,其中该CRISPR酶进行 了密码子优化以便在真核细胞中表达。

11.如前述项目中任一项所述的组合物或系统,其中该指导序列在长度 上是至少15个核苷酸。

12.如前述项目中任一项所述的组合物或系统,其中该嵌合RNA多核 苷酸序列包括两个、三个、四个或五个发夹。

13.一种非天然存在的或工程化的组合物,包含一种载体系统,该载体 系统包含

一个或多个载体,该一个或多个载体包含

I.一个第一调节元件,该第一调节元件可操作地连接至

(a)一个指导序列,该指导序列能够杂交到真核细胞中的靶序列上,和

(b)一个tracr配对序列,

II.一个第二调节元件,该第二调节元件可操作地连接到编码CRISPR酶 的酶编码序列上,在该CRISPR酶的一个末端附近包含至少一个或多个核定 位序列(NLS),以及

III.一个第三调节元件,该第三调节元件可操作地连接至一个tracr序 列,

其中组分I、II和III位于该系统的相同或不同载体上,

其中在转录时,该tracr配对序列杂交到该tracr序列上,并且该指导序 列引导CRISPR复合物与该靶序列的序列特异性结合,并且

其中该CRISPR复合物包含与以下各项复合的CRISPR酶:(1)杂交到 该靶序列上的指导序列、和(2)杂交到该tracr序列上的tracr配对序列。

14.如项目13所述的组合物,其中使用多个指导序列以及单个tracr序 列以提供一种多元系统。

15.一种多元CRISPR酶系统,其中该系统包含一个载体系统,该载体 系统包含一个或多个载体,该一个或多个载体包含

I.一个第一调节元件,该第一调节元件可操作地连接至

(a)一个指导序列,该指导序列能够杂交到真核细胞中的靶序列上,和

(b)一个tracr配对序列,

II.一个第二调节元件,该第二调节元件可操作地连接到编码CRISPR酶 的酶编码序列上,在该CRISPR酶的一个末端附近包含至少一个或多个核定 位序列(NLS),以及

III.一个第三调节元件,该第三调节元件可操作地连接至一个tracr序 列,

其中组分I、II和III位于该系统的相同或不同载体上,

其中在转录时,该tracr配对序列杂交到该tracr序列上,并且该指导序 列引导CRISPR复合物与该靶序列的序列特异性结合,

其中该CRISPR复合物包含与以下各项复合的CRISPR酶:(1)杂交到 该靶序列上的指导序列、和(2)杂交到该tracr序列上的tracr配对序列,并 且

其中在该多元系统中使用多个指导序列以及单个tracr序列。

16.如项目13至15中任一项所述的组合物或系统,其中该第一调节元 件是一种聚合酶III启动子。

17.如项目13至16中任一项所述的组合物或系统,其中该第二调节元 件是一种聚合酶II启动子。

18.如项目13至17中任一项所述的组合物或系统,其中该第三调节元 件是一种聚合酶III启动子。

19.如项目13至18中任一项所述的组合物或系统,其中该CRISPR酶 包括一个或多个NLS,该一个或多个NLS具有足够强度来在真核细胞的细胞 核中驱动所述CRISPR酶以可检测到的量积聚。

20.如项目13至19中任一项所述的组合物或系统,其中当进行最佳比 对时,沿着该tracr配对序列的长度,该tracr序列展现至少50%序列互补性。

21.如项目13至20中任一项所述的组合物或系统,其中该CRISPR酶 是一种II型CRISPR系统酶。

22.如项目13至21中任一项所述的组合物或系统,其中该CRISPR酶 是一种Cas9酶。

23.如项目13至22中任一项所述的组合物或系统,其中该CRISPR酶 进行了密码子优化以便在真核细胞中表达。

24.如项目13至23中任一项所述的组合物或系统,其中该指导序列在 长度上是至少15个核苷酸。

25.一种真核宿主细胞,包含如前述项目中任一项所述的组合物或系 统。

26.一种生物体,包含如项目25所述的真核宿主细胞。

27.一种非人生物体,包含如项目25所述的真核宿主细胞。

28.一种试剂盒,包含如项目1至24中任一项所述的组合物以及用于 使用所述试剂盒的说明书。

29.一种改变感兴趣的基因组座位在真核细胞中的表达的方法,该方法 包括

使该基因组座位与如项目1至24中任一项所述的组合物接触,并且

确定该基因组座位的表达是否已发生改变。

30.如项目29所述的方法,其中基于一种CRISPR基序序列的存在, 该指导序列引导该CRISPR复合物与靶序列的序列特异性结合。

31.如项目30所述的方法,其中该CRISPR基序序列是NAG。

32.通过在一个或多个原核细胞中引入一个或多个基因突变来筛选一个 或多个原核细胞的方法,该方法包括:

将一个或多个载体引入该一个或多个原核细胞中,其中该一 个或多个载体驱动以下一项或多项的表达:CRISPR酶、连接到tracr配对序 列上的指导序列、tracr序列、以及编辑模板;

其中该编辑模板包含消除CRISPR酶切的一个或多个突 变;

允许该编辑模板与靶多核苷酸在有待筛选的一个或多个细胞 中进行同源重组;

允许CRISPR复合物结合到靶多核苷酸上,以实现靶多核苷 酸在所述基因内的切割,其中该CRISPR复合物包含与以下各项复合的 CRISPR酶:(1)杂交到该靶多核苷酸内的靶序列上的指导序列、和(2)杂 交到该tracr序列上的tracr配对序列,

其中该CRISPR复合物与该靶多核苷酸的结合诱导细胞死 亡,

由此允许对其中已经引入一个或多个突变的一个或多个原核 细胞进行筛选。

33.如项目32所述的方法,其中该CRISPR酶是一种II型CRISPR系 统酶。

34.如项目33所述的方法,其中该CRISPR酶是Cas9。

<110> 布罗德研究所有限公司

<120> 用于序列操纵的 CRISPR-CAS 组分系统、方法以及组合物

<130> 44790.99.2003

<140> PCT/US2013/074611
<141> 2013-12-12

<150> 61/835,931
<151> 2013-06-17

<150> 61/791,409
<151> 2013-03-15

<150> 61/768,959
<151> 2013-02-25

<150> 61/748,427
<151> 2013-01-02

<150> 61/736,527
<151> 2012-12-12

[0001] <160> 529

<170> PatentIn version 3.5

<210> 1
<211> 7
<212> PRT
<213> 猴病毒 40

<400> 1
Pro Lys Lys Lys Arg Lys Val
1 5

<210> 2
<211> 16
<212> PRT
<213> 未知

<220>
<221> 来源
<223> /注释="未知的描述:
核质蛋白二分 NLS 序列"

<400> 2

Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys
1 5 10 15

<210> 3
<211> 9
<212> PRT
<213> 未知

<220>
<221> 来源
<223> /注释="未知的描述:
C-myc NLS 序列"

<400> 3
Pro Ala Ala Lys Arg Val Lys Leu Asp
1 5

<210> 4
<211> 11
<212> PRT
<213> 未知

[0002]

<220>
<221> 来源
<223> /注释="未知的描述:
C-myc NLS 序列"

<400> 4
Arg Gln Arg Arg Asn Glu Leu Lys Arg Ser Pro
1 5 10

<210> 5
<211> 38
<212> PRT
<213> 人

<400> 5
Asn Gln Ser Ser Asn Phe Gly Pro Met Lys Gly Gly Asn Phe Gly Gly
1 5 10 15

Arg Ser Ser Gly Pro Tyr Gly Gly Gly Gly Gln Tyr Phe Ala Lys Pro
20 25 30

Arg Asn Gln Gly Gly Tyr
35

<210> 6
<211> 42
<212> PRT
<213> 未知

<220>
<221> 来源
<223> /注释="未知的描述:
来自输入蛋白-α 序列的 IBB 结构域"

<400> 6
Arg Met Arg Ile Glx Phe Lys Asn Lys Gly Lys Asp Thr Ala Glu Leu
1 5 10 15

Arg Arg Arg Arg Val Glu Val Ser Val Glu Leu Arg Lys Ala Lys Lys
20 25 30

Asp Glu Gln Ile Leu Lys Arg Arg Asn Val
35 40

[0003]

<210> 7
<211> 8
<212> PRT
<213> 未知

<220>
<221> 来源
<223> /注释="未知的描述:
Myoma T 蛋白序列"

<400> 7
Val Ser Arg Lys Arg Pro Arg Pro
1 5

<210> 8
<211> 8
<212> PRT
<213> 未知

<220>
<221> 来源
<223> /注释="未知的描述:
Myoma T 蛋白序列"

<400> 8
Pro Pro Lys Lys Ala Arg Glu Asp

1 5

<210> 9
<211> 8
<212> PRT
<213> 人

<400> 9
Pro Gln Pro Lys Lys Lys Pro Leu
1 5

<210> 10
<211> 12
<212> PRT
<213> 小鼠

<400> 10
Ser Ala Leu Ile Lys Lys Lys Lys Lys Met Ala Pro
1 5 10

<210> 11
<211> 5
[0004] <212> PRT
<213> 流感病毒

<400> 11
Asp Arg Leu Arg Arg
1 5

<210> 12
<211> 7
<212> PRT
<213> 流感病毒

<400> 12
Pro Lys Gln Lys Lys Arg Lys
1 5

<210> 13
<211> 10
<212> PRT
<213> 肝炎病毒δ

<400> 13
Arg Lys Leu Lys Lys Lys Ile Lys Lys Leu
1 5 10

<210> 14
<211> 10
<212> PRT
<213> 小鼠

<400> 14
Arg Glu Lys Lys Lys Phe Leu Lys Arg Arg
1 5 10

<210> 15
<211> 20
<212> PRT
<213> 人

<400> 15
Lys Arg Lys Gly Asp Glu Val Asp Gly Val Asp Glu Val Ala Lys Lys
1 5 10 15

Lys Ser Lys Lys
20

[0005]

<210> 16
<211> 17
<212> PRT
<213> 人

<400> 16
Arg Lys Cys Leu Gln Ala Gly Met Asn Leu Glu Ala Arg Lys Thr Lys
1 5 10 15

Lys

<210> 17
<211> 27
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<220>
<221> 修饰_碱基
<222> (1)..(20)
<223> a, c, t 或 g

<220>
<221> 修饰_碱基
<222> (21)..(22)
<223> a, c, t, g, 未知或其它

<400> 17
nnnnnnnnnn nnnnnnnnnn nnagaaw 27

<210> 18
<211> 19
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成寡核苷酸"

[0006]

<220>
<221> 修饰_碱基
<222> (1)..(12)
<223> a, c, t 或 g

<220>
<221> 修饰_碱基
<222> (13)..(14)
<223> a, c, t, g, 未知或其它

<400> 18
nnnnnnnnnn nnnnagaaw 19

<210> 19
<211> 27
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成寡核苷酸"

<220>

<220>
<221> 修饰_碱基
<222> (1)..(20)
<223> a, c, t 或 g

<220>
<221> 修饰_碱基
<222> (21)..(22)
<223> a, c, t, g, 未知或其它

<400> 19
nnnnnnnnnn nnnnnnnnnn nnagaaw 27

<210> 20
<211> 18
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成寡核苷酸"

<220>
[0007] <221> 修饰_碱基
<222> (1)..(11)
<223> a, c, t 或 g

<220>
<221> 修饰_碱基
<222> (12)..(13)
<223> a, c, t, g, 未知或其它

<400> 20
nnnnnnnnnn nnnagaaw 18

<210> 21
<211> 137
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成多核苷酸"

<220>
<221> 修饰_碱基

<222> (1)..(20)
<223> a, c, t, g, 未知或其它

<400> 21
nnnnnnnnnn nnnnnnnnnn gtttttgtac tctcaagatt tagaaataaa tcttgcagaa 60

gctacaaaga taaggcttca tgccgaaatc aacaccctgt cattttatgg cagggtgttt 120

tcgttattta atttttt 137

<210> 22
<211> 123
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<220>
<221> 修饰_碱基
<222> (1)..(20)
[0008] <223> a, c, t, g, 未知或其它

<400> 22
nnnnnnnnnn nnnnnnnnnn gtttttgtac tctcagaaat gcagaagcta caaagataag 60

gcttcatgcc gaaatcaaca ccctgtcatt ttatggcagg gtgttttcgt tatttaattt 120

ttt 123

<210> 23
<211> 110
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<220>
<221> 修饰_碱基
<222> (1)..(20)
<223> a, c, t, g, 未知或其它

<400> 23
nnnnnnnnnn nnnnnnnnnn gtttttgtac tctcagaaat gcagaagcta caaagataag 60

gcttcatgcc gaaatcaaca ccctgtcatt ttatggcagg gtgtttttt 110

<210> 24
<211> 102
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<220>
<221> 修饰_碱基
<222> (1)..(20)
<223> a, c, t, g, 未知或其它

<400> 24
nnnnnnnnnn nnnnnnnnnn gttttagagc tagaaatagc aagttaaaat aaggctagtc 60

[0009] cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt tt 102

<210> 25
<211> 88
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<220>
<221> 修饰_碱基
<222> (1)..(20)
<223> a, c, t, g, 未知或其它

<400> 25
nnnnnnnnnn nnnnnnnnnn gttttagagc tagaaatagc aagttaaaat aaggctagtc 60

cgttatcaac ttgaaaaagt gttttttt 88

<210> 26

<211> 76
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<220>
<221> 修饰_碱基
<222> (1)..(20)
<223> a, c, t, g, 未知或其它

<400> 26
nnnnnnnnnn nnnnnnnnnn gttttagagc tagaaatagc aagttaaaat aaggctagtc 60

cgttatcatt tttttt 76

<210> 27
<211> 12
<212> DNA
<213> 人工序列

[0010]

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 27
gttttagagc ta 12

<210> 28
<211> 31
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 28
tagcaagtta aaataaggct agtccgtttt t 31

<210> 29
<211> 27

<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成寡核苷酸"

<220>
<221> 修饰_碱基
<222> (1)..(22)
<223> a, c, t, g, 未知或其它

<400> 29
nnnnnnnnnn nnnnnnnnnn nnagaaw 27

<210> 30
<211> 12
<212> RNA
<213> 人工序列

<220>
<221> 来源
[0011] <223> /注释="人工序列的描述：合成寡核苷酸"

<400> 30
guuuuagagc ua 12

<210> 31
<211> 33
<212> DNA
<213> 人

<400> 31
ggacatcgat gtcacctcca atgactaggg tgg 33

<210> 32
<211> 33
<212> DNA
<213> 人

<400> 32
cattggaggt gacatcgatg tcctccccat tgg 33

<210> 33

<210> 33
<211> 33
<212> DNA
<213> 人

<400> 33
ggaagggcct gagtccgagc agaagaagaa ggg 33

<210> 34
<211> 33
<212> DNA
<213> 人

<400> 34
ggtggcgaga ggggccgaga ttgggtgttc agg 33

<210> 35
<211> 33
<212> DNA
<213> 人

<400> 35
atgcaggagg gtggcgagag gggccgagat tgg 33

[0012]

<210> 36
<211> 21
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成引物"

<400> 36
aaaaccaccc ttctctctgg c 21

<210> 37
<211> 21
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成引物"

<400> 37

ggagattgga gacacggaga g 21

<210> 38
<211> 20
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 38
ctggaaagcc aatgcctgac 20

<210> 39
<211> 20
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
[0013] 引物"

<400> 39
ggcagcaaac tccttgtcct 20

<210> 40
<211> 335
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<400> 40
gagggcctat ttcccatgat tccttcatat ttgcatatac gatacaaggc tgttagagag 60

ataattggaa ttaatttgac tgtaaacaca aagatattag tacaaaatac gtgacgtaga 120

aagtaataat ttcttgggta gtttgcagtt ttaaaattat gttttaaaat ggactatcat 180

atgcttaccg taacttgaaa gtatttcgat ttcttggctt tatatatctt gtggaaagga 240

cgaaacaccg gaaccattca aaacagcata gcaagttaaa ataaggctag tccgttatca 300

acttgaaaaa gtggcaccga gtcggtgctt ttttt 335

<210> 41
<211> 423
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<400> 41
gagggcctat ttcccatgat tccttcatat ttgcatatac gatacaaggc tgttagagag 60

ataattggaa ttaatttgac tgtaaacaca aagatattag tacaaaatac gtgacgtaga 120

aagtaataat ttcttgggta gtttgcagtt ttaaaattat gttttaaaat ggactatcat 180

atgcttaccg taacttgaaa gtatttcgat ttcttggctt tatatatctt gtggaaagga 240

cgaaacaccg gtagtattaa gtattgtttt atggctgata aatttctttg aatttctcct 300

[0014] tgattatttg ttataaaagt tataaaataa tcttgttgga accattcaaa acagcatagc 360

aagttaaaat aaggctagtc cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt 420

ttt 423

<210> 42
<211> 339
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<400> 42
gagggcctat ttcccatgat tccttcatat ttgcatatac gatacaaggc tgttagagag 60

ataattggaa ttaatttgac tgtaaacaca aagatattag tacaaaatac gtgacgtaga 120

aagtaataat ttcttgggta gtttgcagtt ttaaaattat gttttaaaat ggactatcat 180

atgcttaccg taacttgaaa gtatttcgat ttcttggctt tatatatctt gtggaaagga 240

cgaaacaccg ggttttagag ctatgctgtt ttgaatggtc ccaaaacggg tcttcgagaa 300

gacgttttag agctatgctg ttttgaatgg tcccaaaac 339

<210> 43
<211> 309
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<400> 43
gagggcctat ttcccatgat tccttcatat ttgcatatac gatacaaggc tgttagagag 60

ataattggaa ttaatttgac tgtaaacaca aagatattag tacaaaatac gtgacgtaga 120

aagtaataat ttcttgggta gtttgcagtt ttaaaattat gttttaaaat ggactatcat 180

atgcttaccg taacttgaaa gtatttcgat ttcttggctt tatatatctt gtggaaagga 240

[0015] cgaaacaccg ggtcttcgag aagacctgtt ttagagctag aaatagcaag ttaaaataag 300

gctagtccg 309

<210> 44
<211> 1648
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多肽"

<400> 44
Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp
1 5 10 15

Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys Arg Lys Val
20 25 30

Gly Ile His Gly Val Pro Ala Ala Asp Lys Lys Tyr Ser Ile Gly Leu
35 40 45

Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr
50 55 60

Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His
65 70 75 80

Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu
85 90 95

Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr
100 105 110

Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu
115 120 125

Met Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe
130 135 140

[0016]

Leu Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn
145 150 155 160

Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His
165 170 175

Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu
180 185 190

Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu
195 200 205

Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe
210 215 220

Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile
225 230 235 240

Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser
245 250 255

Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys
260 265 270

Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr
275 280 285

Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln
290 295 300

Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln
305 310 315 320

Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser
325 330 335

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

[0017] Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His
355 360 365

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

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

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

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

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

Ile Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg

450 455 460

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

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

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

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

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

Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr
[0018] 545 550 555 560

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

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

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

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

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

Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu
645 650 655

Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu
660 665 670

Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys
675 680 685

Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys
690 695 700

Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp
705 710 715 720

Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile
725 730 735

His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val
740 745 750

[0019]

Ser Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly
755 760 765

Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp
770 775 780

Glu Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile
785 790 795 800

Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser
805 810 815

Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser
820 825 830

Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu
835 840 845

Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp
850 855 860

Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His Ile
865 870 875 880

Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu
885 890 895

Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu
900 905 910

Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala
915 920 925

Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg
930 935 940

Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu
945 950 955 960

[0020]

Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser
965 970 975

Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val
980 985 990

Lys Val Ile Thr Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp
995 1000 1005

Phe Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His Ala
1010 1015 1020

His Asp Ala Tyr Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys
1025 1030 1035

Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys
1040 1045 1050

Val Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile
1055 1060 1065

Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn
1070 1075 1080

Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys
1085 1090 1095

Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp
1100 1105 1110

Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met
1115 1120 1125

Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly
1130 1135 1140

Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu
1145 1150 1155

[0021] Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe
1160 1165 1170

Asp Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val
1175 1180 1185

Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu
1190 1195 1200

Gly Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile
1205 1210 1215

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
1235 1240 1245

Arg Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn

1250 1255 1260

Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala
1265 1270 1275

Ser His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln
1280 1285 1290

Lys Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile
1295 1300 1305

Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp
1310 1315 1320

Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp
1325 1330 1335

Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr
[0022] 1340 1345 1350

Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr
1355 1360 1365

Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp
1370 1375 1380

Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg
1385 1390 1395

Ile Asp Leu Ser Gln Leu Gly Gly Asp Ala Ala Ala Val Ser Lys
1400 1405 1410

Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu Val Glu Leu
1415 1420 1425

Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly Glu Gly
1430 1435 1440

Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile Cys
1445 1450 1455

Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr
1460 1465 1470

Leu Thr Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met
1475 1480 1485

Lys Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val
1490 1495 1500

Gln Glu Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr
1505 1510 1515

Arg Ala Glu Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile
1520 1525 1530

[0023]

Glu Leu Lys Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly
1535 1540 1545

His Lys Leu Glu Tyr Asn Tyr Asn Ser His Asn Val Tyr Ile Met
1550 1555 1560

Ala Asp Lys Gln Lys Asn Gly Ile Lys Val Asn Phe Lys Ile Arg
1565 1570 1575

His Asn Ile Glu Asp Gly Ser Val Gln Leu Ala Asp His Tyr Gln
1580 1585 1590

Gln Asn Thr Pro Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asn
1595 1600 1605

His Tyr Leu Ser Thr Gln Ser Ala Leu Ser Lys Asp Pro Asn Glu
1610 1615 1620

Lys Arg Asp His Met Val Leu Leu Glu Phe Val Thr Ala Ala Gly
1625 1630 1635

Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys
1640 1645

<210> 45
<211> 1625
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成多肽"

<400> 45
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
[0024] 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

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240
```

[0025]

```
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350
```

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430

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

[0026] Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

```
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

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                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

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                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750
```

[0027]

```
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

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                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845
```

[0028]

```
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
1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1040 1045 1050

[0029]

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys
1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys
1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly
1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val
1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
1235 1240 1245

[0030] Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys
1250 1255 1260

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys
1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala
1280 1285 1290

Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
1295 1300 1305

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala
1310 1315 1320

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr

1340 1345 1350

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

Ala Ala Ala Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val
1370 1375 1380

Pro Ile Leu Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe
1385 1390 1395

Ser Val Ser Gly Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu
1400 1405 1410

Thr Leu Lys Phe Ile Cys Thr Thr Gly Lys Leu Pro Val Pro Trp
1415 1420 1425

Pro Thr Leu Val Thr Thr Leu Thr Tyr Gly Val Gln Cys Phe Ser
[0031] 1430 1435 1440

Arg Tyr Pro Asp His Met Lys Gln His Asp Phe Phe Lys Ser Ala
1445 1450 1455

Met Pro Glu Gly Tyr Val Gln Glu Arg Thr Ile Phe Phe Lys Asp
1460 1465 1470

Asp Gly Asn Tyr Lys Thr Arg Ala Glu Val Lys Phe Glu Gly Asp
1475 1480 1485

Thr Leu Val Asn Arg Ile Glu Leu Lys Gly Ile Asp Phe Lys Glu
1490 1495 1500

Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr Asn Tyr Asn Ser
1505 1510 1515

His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn Gly Ile Lys
1520 1525 1530

Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser Val Gln
1535 1540 1545

Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly Pro
1550 1555 1560

Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu
1565 1570 1575

Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu
1580 1585 1590

Phe Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr
1595 1600 1605

Lys Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys
1610 1615 1620

[0032] Lys Lys
1625

<210> 46
<211> 1664
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多肽"

<400> 46
Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp
1 5 10 15

Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys Arg Lys Val
20 25 30

Gly Ile His Gly Val Pro Ala Ala Asp Lys Lys Tyr Ser Ile Gly Leu
35 40 45

Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr
50 55 60

Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His
65 70 75 80

Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu
85 90 95

Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr
100 105 110

Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu
115 120 125

Met Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe
130 135 140

[0033]

Leu Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn
145 150 155 160

Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His
165 170 175

Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu
180 185 190

Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu
195 200 205

Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe
210 215 220

Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile
225 230 235 240

Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser
245 250 255

Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys
260 265 270

Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr
275 280 285

Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln
290 295 300

Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln
305 310 315 320

Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser
325 330 335

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

[0034]

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

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

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

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

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

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

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

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

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

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

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

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

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

[0035] Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro
565 570 575

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

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

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

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

Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu
645 650 655

Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu

660 665 670

Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys
675 680 685

Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys
690 695 700

Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp
705 710 715 720

Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile
725 730 735

His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val
740 745 750

Ser Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly
[0036] 755 760 765

Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp
770 775 780

Glu Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile
785 790 795 800

Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser
805 810 815

Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser
820 825 830

Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu
835 840 845

Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp
850 855 860

```
Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His Ile
865                 870                 875                 880

Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu
                885                 890                 895

Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu
            900                 905                 910

Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala
        915                 920                 925

Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg
    930                 935                 940

Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu
945                 950                 955                 960
```

[0037]

```
Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser
                965                 970                 975

Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val
            980                 985                 990

Lys Val Ile Thr Leu Lys Ser Lys  Leu Val Ser Asp Phe  Arg Lys Asp
        995                 1000                1005

Phe Gln  Phe Tyr Lys Val Arg  Glu Ile Asn Asn Tyr  His His Ala
    1010                1015                1020

His Asp  Ala Tyr Leu Asn Ala  Val Val Gly Thr Ala  Leu Ile Lys
    1025                1030                1035

Lys Tyr  Pro Lys Leu Glu Ser  Glu Phe Val Tyr Gly  Asp Tyr Lys
    1040                1045                1050

Val Tyr  Asp Val Arg Lys Met  Ile Ala Lys Ser Glu  Gln Glu Ile
    1055                1060                1065
```

Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn
1070 1075 1080

Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys
1085 1090 1095

Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp
1100 1105 1110

Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met
1115 1120 1125

Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly
1130 1135 1140

Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu
1145 1150 1155

[0038]

Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe
1160 1165 1170

Asp Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val
1175 1180 1185

Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu
1190 1195 1200

Gly Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile
1205 1210 1215

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
1235 1240 1245

Arg Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn
1250 1255 1260

Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala
1265 1270 1275

Ser His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln
1280 1285 1290

Lys Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile
1295 1300 1305

Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp
1310 1315 1320

Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp
1325 1330 1335

Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr
1340 1345 1350

[0039] Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr
1355 1360 1365

Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp
1370 1375 1380

Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg
1385 1390 1395

Ile Asp Leu Ser Gln Leu Gly Gly Asp Ala Ala Ala Val Ser Lys
1400 1405 1410

Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu Val Glu Leu
1415 1420 1425

Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly Glu Gly
1430 1435 1440

Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile Cys

1445 1450 1455

Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr
1460 1465 1470

Leu Thr Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met
1475 1480 1485

Lys Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val
1490 1495 1500

Gln Glu Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr
1505 1510 1515

Arg Ala Glu Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile
1520 1525 1530

Glu Leu Lys Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly
[0040] 1535 1540 1545

His Lys Leu Glu Tyr Asn Tyr Asn Ser His Asn Val Tyr Ile Met
1550 1555 1560

Ala Asp Lys Gln Lys Asn Gly Ile Lys Val Asn Phe Lys Ile Arg
1565 1570 1575

His Asn Ile Glu Asp Gly Ser Val Gln Leu Ala Asp His Tyr Gln
1580 1585 1590

Gln Asn Thr Pro Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asn
1595 1600 1605

His Tyr Leu Ser Thr Gln Ser Ala Leu Ser Lys Asp Pro Asn Glu
1610 1615 1620

Lys Arg Asp His Met Val Leu Leu Glu Phe Val Thr Ala Ala Gly
1625 1630 1635

Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys Lys Arg Pro Ala Ala
1640 1645 1650

Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys
1655 1660

<210> 47
<211> 1423
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多肽"

<400> 47
Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp
1 5 10 15

Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys Arg Lys Val
[0041] 20 25 30

Gly Ile His Gly Val Pro Ala Ala Asp Lys Lys Tyr Ser Ile Gly Leu
35 40 45

Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr
50 55 60

Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His
65 70 75 80

Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu
85 90 95

Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr
100 105 110

Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu
115 120 125

Met Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe
130 135 140

Leu Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn
145 150 155 160

Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His
165 170 175

Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu
180 185 190

Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu
195 200 205

Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe
210 215 220

[0042]

Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile
225 230 235 240

Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser
245 250 255

Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys
260 265 270

Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr
275 280 285

Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln
290 295 300

Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln
305 310 315 320

Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser
325 330 335

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

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

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

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

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

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

[0043]

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

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

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

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

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

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

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

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

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

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

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

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

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

[0044] Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu
645 650 655

Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu
660 665 670

Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys
675 680 685

Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys
690 695 700

Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp
705 710 715 720

Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile
725 730 735

His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val

740 745 750

Ser Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly
755 760 765

Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp
770 775 780

Glu Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile
785 790 795 800

Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser
805 810 815

Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser
820 825 830

Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu
[0045] 835 840 845

Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp
850 855 860

Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His Ile
865 870 875 880

Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu
885 890 895

Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu
900 905 910

Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala
915 920 925

Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg
930 935 940

Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu
945 950 955 960

Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser
965 970 975

Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val
980 985 990

Lys Val Ile Thr Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp
995 1000 1005

Phe Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His Ala
1010 1015 1020

His Asp Ala Tyr Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys
1025 1030 1035

[0046] Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys
1040 1045 1050

Val Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile
1055 1060 1065

Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn
1070 1075 1080

Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys
1085 1090 1095

Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp
1100 1105 1110

Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met
1115 1120 1125

Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly
1130 1135 1140

Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu
1145 1150 1155

Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe
1160 1165 1170

Asp Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val
1175 1180 1185

Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu
1190 1195 1200

Gly Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile
1205 1210 1215

Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu
1220 1225 1230

[0047]

Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly
1235 1240 1245

Arg Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn
1250 1255 1260

Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala
1265 1270 1275

Ser His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln
1280 1285 1290

Lys Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile
1295 1300 1305

Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp
1310 1315 1320

Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp
1325 1330 1335

Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr
1340 1345 1350

Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr
1355 1360 1365

Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp
1370 1375 1380

Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg
1385 1390 1395

Ile Asp Leu Ser Gln Leu Gly Gly Asp Lys Arg Pro Ala Ala Thr
1400 1405 1410

Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys
1415 1420

[0048] <210> 48
<211> 483
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成多肽"

<400> 48
Met Phe Leu Phe Leu Ser Leu Thr Ser Phe Leu Ser Ser Ser Arg Thr
1 5 10 15

Leu Val Ser Lys Gly Glu Glu Asp Asn Met Ala Ile Ile Lys Glu Phe
20 25 30

Met Arg Phe Lys Val His Met Glu Gly Ser Val Asn Gly His Glu Phe
35 40 45

Glu Ile Glu Gly Glu Gly Glu Gly Arg Pro Tyr Glu Gly Thr Gln Thr
50 55 60

Ala Lys Leu Lys Val Thr Lys Gly Gly Pro Leu Pro Phe Ala Trp Asp
65 70 75 80

Ile Leu Ser Pro Gln Phe Met Tyr Gly Ser Lys Ala Tyr Val Lys His
85 90 95

Pro Ala Asp Ile Pro Asp Tyr Leu Lys Leu Ser Phe Pro Glu Gly Phe
100 105 110

Lys Trp Glu Arg Val Met Asn Phe Glu Asp Gly Gly Val Val Thr Val
115 120 125

Thr Gln Asp Ser Ser Leu Gln Asp Gly Glu Phe Ile Tyr Lys Val Lys
130 135 140

Leu Arg Gly Thr Asn Phe Pro Ser Asp Gly Pro Val Met Gln Lys Lys
145 150 155 160

[0049] Thr Met Gly Trp Glu Ala Ser Ser Glu Arg Met Tyr Pro Glu Asp Gly
165 170 175

Ala Leu Lys Gly Glu Ile Lys Gln Arg Leu Lys Leu Lys Asp Gly Gly
180 185 190

His Tyr Asp Ala Glu Val Lys Thr Thr Tyr Lys Ala Lys Lys Pro Val
195 200 205

Gln Leu Pro Gly Ala Tyr Asn Val Asn Ile Lys Leu Asp Ile Thr Ser
210 215 220

His Asn Glu Asp Tyr Thr Ile Val Glu Gln Tyr Glu Arg Ala Glu Gly
225 230 235 240

Arg His Ser Thr Gly Gly Met Asp Glu Leu Tyr Lys Gly Ser Lys Gln
245 250 255

Leu Glu Glu Leu Leu Ser Thr Ser Phe Asp Ile Gln Phe Asn Asp Leu

260 265 270

Thr Leu Leu Glu Thr Ala Phe Thr His Thr Ser Tyr Ala Asn Glu His
275 280 285

Arg Leu Leu Asn Val Ser His Asn Glu Arg Leu Glu Phe Leu Gly Asp
290 295 300

Ala Val Leu Gln Leu Ile Ile Ser Glu Tyr Leu Phe Ala Lys Tyr Pro
305 310 315 320

Lys Lys Thr Glu Gly Asp Met Ser Lys Leu Arg Ser Met Ile Val Arg
325 330 335

Glu Glu Ser Leu Ala Gly Phe Ser Arg Phe Cys Ser Phe Asp Ala Tyr
340 345 350

Ile Lys Leu Gly Lys Gly Glu Glu Lys Ser Gly Gly Arg Arg Arg Asp
[0050] 355 360 365

Thr Ile Leu Gly Asp Leu Phe Glu Ala Phe Leu Gly Ala Leu Leu Leu
370 375 380

Asp Lys Gly Ile Asp Ala Val Arg Arg Phe Leu Lys Gln Val Met Ile
385 390 395 400

Pro Gln Val Glu Lys Gly Asn Phe Glu Arg Val Lys Asp Tyr Lys Thr
405 410 415

Cys Leu Gln Glu Phe Leu Gln Thr Lys Gly Asp Val Ala Ile Asp Tyr
420 425 430

Gln Val Ile Ser Glu Lys Gly Pro Ala His Ala Lys Gln Phe Glu Val
435 440 445

Ser Ile Val Val Asn Gly Ala Val Leu Ser Lys Gly Leu Gly Lys Ser
450 455 460

Lys Lys Leu Ala Glu Gln Asp Ala Ala Lys Asn Ala Leu Ala Gln Leu
465 470 475 480

Ser Glu Val

<210> 49
<211> 483
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多肽"

<400> 49
Met Lys Gln Leu Glu Glu Leu Leu Ser Thr Ser Phe Asp Ile Gln Phe
1 5 10 15

Asn Asp Leu Thr Leu Leu Glu Thr Ala Phe Thr His Thr Ser Tyr Ala
[0051] 20 25 30

Asn Glu His Arg Leu Leu Asn Val Ser His Asn Glu Arg Leu Glu Phe
35 40 45

Leu Gly Asp Ala Val Leu Gln Leu Ile Ile Ser Glu Tyr Leu Phe Ala
50 55 60

Lys Tyr Pro Lys Lys Thr Glu Gly Asp Met Ser Lys Leu Arg Ser Met
65 70 75 80

Ile Val Arg Glu Glu Ser Leu Ala Gly Phe Ser Arg Phe Cys Ser Phe
85 90 95

Asp Ala Tyr Ile Lys Leu Gly Lys Gly Glu Glu Lys Ser Gly Gly Arg
100 105 110

Arg Arg Asp Thr Ile Leu Gly Asp Leu Phe Glu Ala Phe Leu Gly Ala
115 120 125

Leu Leu Leu Asp Lys Gly Ile Asp Ala Val Arg Arg Phe Leu Lys Gln
130 135 140

Val Met Ile Pro Gln Val Glu Lys Gly Asn Phe Glu Arg Val Lys Asp
145 150 155 160

Tyr Lys Thr Cys Leu Gln Glu Phe Leu Gln Thr Lys Gly Asp Val Ala
165 170 175

Ile Asp Tyr Gln Val Ile Ser Glu Lys Gly Pro Ala His Ala Lys Gln
180 185 190

Phe Glu Val Ser Ile Val Val Asn Gly Ala Val Leu Ser Lys Gly Leu
195 200 205

Gly Lys Ser Lys Lys Leu Ala Glu Gln Asp Ala Ala Lys Asn Ala Leu
210 215 220

[0052]

Ala Gln Leu Ser Glu Val Gly Ser Val Ser Lys Gly Glu Glu Asp Asn
225 230 235 240

Met Ala Ile Ile Lys Glu Phe Met Arg Phe Lys Val His Met Glu Gly
245 250 255

Ser Val Asn Gly His Glu Phe Glu Ile Glu Gly Glu Gly Glu Gly Arg
260 265 270

Pro Tyr Glu Gly Thr Gln Thr Ala Lys Leu Lys Val Thr Lys Gly Gly
275 280 285

Pro Leu Pro Phe Ala Trp Asp Ile Leu Ser Pro Gln Phe Met Tyr Gly
290 295 300

Ser Lys Ala Tyr Val Lys His Pro Ala Asp Ile Pro Asp Tyr Leu Lys
305 310 315 320

Leu Ser Phe Pro Glu Gly Phe Lys Trp Glu Arg Val Met Asn Phe Glu
325 330 335

Asp Gly Gly Val Val Thr Val Thr Gln Asp Ser Ser Leu Gln Asp Gly
340 345 350

Glu Phe Ile Tyr Lys Val Lys Leu Arg Gly Thr Asn Phe Pro Ser Asp
355 360 365

Gly Pro Val Met Gln Lys Lys Thr Met Gly Trp Glu Ala Ser Ser Glu
370 375 380

Arg Met Tyr Pro Glu Asp Gly Ala Leu Lys Gly Glu Ile Lys Gln Arg
385 390 395 400

Leu Lys Leu Lys Asp Gly Gly His Tyr Asp Ala Glu Val Lys Thr Thr
405 410 415

Tyr Lys Ala Lys Lys Pro Val Gln Leu Pro Gly Ala Tyr Asn Val Asn
420 425 430

[0053] Ile Lys Leu Asp Ile Thr Ser His Asn Glu Asp Tyr Thr Ile Val Glu
435 440 445

Gln Tyr Glu Arg Ala Glu Gly Arg His Ser Thr Gly Gly Met Asp Glu
450 455 460

Leu Tyr Lys Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys
465 470 475 480

Lys Lys Lys

<210> 50
<211> 1423
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多肽"

<400> 50

```
Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp
1               5                  10                  15

Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys Arg Lys Val
            20                  25                  30

Gly Ile His Gly Val Pro Ala Ala Asp Lys Lys Tyr Ser Ile Gly Leu
        35                  40                  45

Ala Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr
    50                  55                  60

Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His
65                  70                  75                  80

Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu
                85                  90                  95
```

[0054]

```
Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr
            100                 105                 110

Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu
        115                 120                 125

Met Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe
    130                 135                 140

Leu Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn
145                 150                 155                 160

Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His
                165                 170                 175

Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu
            180                 185                 190

Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu
        195                 200                 205
```

Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe
210 215 220

Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile
225 230 235 240

Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser
245 250 255

Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys
260 265 270

Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr
275 280 285

Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln
290 295 300

[0055] Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln
305 310 315 320

Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser
325 330 335

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

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

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

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

Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys

405 410 415

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

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

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

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

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

Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile
[0056] 500 505 510

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

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

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

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

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

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

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

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

Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu
645 650 655

Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu
660 665 670

Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys
675 680 685

Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys
690 695 700

[0057]

Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp
705 710 715 720

Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile
725 730 735

His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val
740 745 750

Ser Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly
755 760 765

Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp
770 775 780

Glu Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile
785 790 795 800

Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser
805 810 815

Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser
820 825 830

Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu
835 840 845

Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp
850 855 860

Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His Ile
865 870 875 880

Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu
885 890 895

Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu
900 905 910

[0058]

Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala
915 920 925

Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg
930 935 940

Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu
945 950 955 960

Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser
965 970 975

Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val
980 985 990

Lys Val Ile Thr Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp
995 1000 1005

Phe Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His Ala
1010 1015 1020

His Asp Ala Tyr Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys
1025 1030 1035

Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys
1040 1045 1050

Val Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile
1055 1060 1065

Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn
1070 1075 1080

Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys
1085 1090 1095

Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp
1100 1105 1110

[0059] Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met
1115 1120 1125

Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly
1130 1135 1140

Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu
1145 1150 1155

Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe
1160 1165 1170

Asp Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val
1175 1180 1185

Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu
1190 1195 1200

Gly Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile

1205 1210 1215

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
1235 1240 1245

Arg Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn
1250 1255 1260

Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala
1265 1270 1275

Ser His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln
1280 1285 1290

Lys Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile
[0060] 1295 1300 1305

Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp
1310 1315 1320

Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp
1325 1330 1335

Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr
1340 1345 1350

Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr
1355 1360 1365

Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp
1370 1375 1380

Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg
1385 1390 1395

Ile Asp Leu Ser Gln Leu Gly Gly Asp Lys Arg Pro Ala Ala Thr
1400 1405 1410

Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys
1415 1420

<210> 51
<211> 2012
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成多核苷酸"

<400> 51

[0061]

gaatgctgcc ctcagacccg cttcctccct gtccttgtct gtccaaggag aatgaggtct 60

cactggtgga tttcggacta ccctgaggag ctggcacctg agggacaagg ccccccacct 120

gcccagctcc agcctctgat gaggggtggg agagagctac atgaggttgc taagaaagcc 180

tcccctgaag gagaccacac agtgtgtgag gttggagtct ctagcagcgg gttctgtgcc 240

cccagggata gtctggctgt ccaggcactg ctcttgatat aaacaccacc tcctagttat 300

gaaaccatgc ccattctgcc tctctgtatg gaaaagagca tggggctggc ccgtggggtg 360

gtgtccactt taggccctgt gggagatcat gggaacccac gcagtgggtc ataggctctc 420

tcatttacta ctcacatcca ctctgtgaag aagcgattat gatctctcct ctagaaactc 480

gtagagtccc atgtctgccg gcttccagag cctgcactcc tccaccttgg cttggctttg 540

ctggggctag aggagctagg atgcacagca gctctgtgac cctttgtttg agaggaacag 600

gaaaaccacc cttctctctg gcccactgtg tcctcttcct gccctgccat ccccttctgt 660

gaatgttaga cccatgggag cagctggtca gaggggaccc cggcctgggg cccctaaccc 720

tatgtagcct cagtcttccc atcaggctct cagctcagcc tgagtgttga ggccccagtg 780

gctgctctgg gggcctcctg agtttctcat ctgtgcccct ccctccctgg cccaggtgaa 840

ggtgtggttc cagaaccgga ggacaaagta caaacggcag aagctggagg aggaagggcc 900

tgagtccgag cagaagaaga agggctccca tcacatcaac cggtggcgca ttgccacgaa 960

gcaggccaat ggggaggaca tcgatgtcac ctccaatgac aagcttgcta gcggtgggca 1020

accacaaacc cacgagggca gagtgctgct tgctgctggc caggcccctg cgtgggccca 1080

agctggactc tggccactcc ctggccaggc tttggggagg cctggagtca tggccccaca 1140

gggcttgaag cccggggccg ccattgacag agggacaagc aatgggctgg ctgaggcctg 1200

ggaccacttg gccttctcct cggagagcct gcctgcctgg gcgggcccgc ccgccaccgc 1260

agcctcccag ctgctctccg tgtctccaat ctccctttg ttttgatgca tttctgtttt 1320

aatttatttt ccaggcacca ctgtagttta gtgatcccca gtgtcccctc tccctatggg 1380

aataataaaa gtctctctct taatgacacg ggcatccagc tccagcccca gagcctgggg 1440

tggtagattc cggctctgag ggccagtggg ggctggtaga gcaaacgcgt tcagggcctg 1500

ggagcctggg gtggggtact ggtggagggg gtcaagggta attcattaac tcctctcttt 1560

tgttggggga ccctggtctc tacctccagc tccacagcag gagaaacagg ctagacatag 1620

ggaagggcca tcctgtatct tgagggagga caggcccagg tctttcttaa cgtattgaga 1680

[0062] ggtgggaatc aggcccaggt agttcaatgg gagagggaga gtgcttccct ctgcctagag 1740

actctggtgg cttctccagt tgaggagaaa ccagaggaaa ggggaggatt ggggtctggg 1800

ggagggaaca ccattcacaa aggctgacgg ttccagtccg aagtcgtggg cccaccagga 1860

tgctcacctg tccttggaga accgctgggc aggttgagac tgcagagaca gggcttaagg 1920

ctgagcctgc aaccagtccc cagtgactca gggcctcctc agcccaagaa agagcaacgt 1980

gccagggccc gctgagctct tgtgttcacc tg 2012

<210> 52
<211> 1153
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多肽"

<400> 52
Met Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys
1 5 10 15

Lys Ser Asp Leu Val Leu Gly Leu Asp Ile Gly Ile Gly Ser Val Gly
20 25 30

Val Gly Ile Leu Asn Lys Val Thr Gly Glu Ile Ile His Lys Asn Ser
35 40 45

Arg Ile Phe Pro Ala Ala Gln Ala Glu Asn Asn Leu Val Arg Arg Thr
50 55 60

Asn Arg Gln Gly Arg Arg Leu Ala Arg Arg Lys Lys His Arg Arg Val
65 70 75 80

Arg Leu Asn Arg Leu Phe Glu Glu Ser Gly Leu Ile Thr Asp Phe Thr
85 90 95

Lys Ile Ser Ile Asn Leu Asn Pro Tyr Gln Leu Arg Val Lys Gly Leu
100 105 110

[0063]

Thr Asp Glu Leu Ser Asn Glu Glu Leu Phe Ile Ala Leu Lys Asn Met
115 120 125

Val Lys His Arg Gly Ile Ser Tyr Leu Asp Asp Ala Ser Asp Asp Gly
130 135 140

Asn Ser Ser Val Gly Asp Tyr Ala Gln Ile Val Lys Glu Asn Ser Lys
145 150 155 160

Gln Leu Glu Thr Lys Thr Pro Gly Gln Ile Gln Leu Glu Arg Tyr Gln
165 170 175

Thr Tyr Gly Gln Leu Arg Gly Asp Phe Thr Val Glu Lys Asp Gly Lys
180 185 190

Lys His Arg Leu Ile Asn Val Phe Pro Thr Ser Ala Tyr Arg Ser Glu
195 200 205

Ala Leu Arg Ile Leu Gln Thr Gln Gln Glu Phe Asn Pro Gln Ile Thr
210 215 220

Asp Glu Phe Ile Asn Arg Tyr Leu Glu Ile Leu Thr Gly Lys Arg Lys
225 230 235 240

Tyr Tyr His Gly Pro Gly Asn Glu Lys Ser Arg Thr Asp Tyr Gly Arg
245 250 255

Tyr Arg Thr Ser Gly Glu Thr Leu Asp Asn Ile Phe Gly Ile Leu Ile
260 265 270

Gly Lys Cys Thr Phe Tyr Pro Asp Glu Phe Arg Ala Ala Lys Ala Ser
275 280 285

Tyr Thr Ala Gln Glu Phe Asn Leu Leu Asn Asp Leu Asn Asn Leu Thr
290 295 300

Val Pro Thr Glu Thr Lys Lys Leu Ser Lys Glu Gln Lys Asn Gln Ile
305 310 315 320

[0064] Ile Asn Tyr Val Lys Asn Glu Lys Ala Met Gly Pro Ala Lys Leu Phe
325 330 335

Lys Tyr Ile Ala Lys Leu Leu Ser Cys Asp Val Ala Asp Ile Lys Gly
340 345 350

Tyr Arg Ile Asp Lys Ser Gly Lys Ala Glu Ile His Thr Phe Glu Ala
355 360 365

Tyr Arg Lys Met Lys Thr Leu Glu Thr Leu Asp Ile Glu Gln Met Asp
370 375 380

Arg Glu Thr Leu Asp Lys Leu Ala Tyr Val Leu Thr Leu Asn Thr Glu
385 390 395 400

Arg Glu Gly Ile Gln Glu Ala Leu Glu His Glu Phe Ala Asp Gly Ser
405 410 415

Phe Ser Gln Lys Gln Val Asp Glu Leu Val Gln Phe Arg Lys Ala Asn

420 425 430

Ser Ser Ile Phe Gly Lys Gly Trp His Asn Phe Ser Val Lys Leu Met
435 440 445

Met Glu Leu Ile Pro Glu Leu Tyr Glu Thr Ser Glu Glu Gln Met Thr
450 455 460

Ile Leu Thr Arg Leu Gly Lys Gln Lys Thr Thr Ser Ser Ser Asn Lys
465 470 475 480

Thr Lys Tyr Ile Asp Glu Lys Leu Leu Thr Glu Glu Ile Tyr Asn Pro
485 490 495

Val Val Ala Lys Ser Val Arg Gln Ala Ile Lys Ile Val Asn Ala Ala
500 505 510

Ile Lys Glu Tyr Gly Asp Phe Asp Asn Ile Val Ile Glu Met Ala Arg
[0065] 515 520 525

Glu Thr Asn Glu Asp Asp Glu Lys Lys Ala Ile Gln Lys Ile Gln Lys
530 535 540

Ala Asn Lys Asp Glu Lys Asp Ala Ala Met Leu Lys Ala Ala Asn Gln
545 550 555 560

Tyr Asn Gly Lys Ala Glu Leu Pro His Ser Val Phe His Gly His Lys
565 570 575

Gln Leu Ala Thr Lys Ile Arg Leu Trp His Gln Gln Gly Glu Arg Cys
580 585 590

Leu Tyr Thr Gly Lys Thr Ile Ser Ile His Asp Leu Ile Asn Asn Ser
595 600 605

Asn Gln Phe Glu Val Asp His Ile Leu Pro Leu Ser Ile Thr Phe Asp
610 615 620

Asp Ser Leu Ala Asn Lys Val Leu Val Tyr Ala Thr Ala Asn Gln Glu
625 630 635 640

Lys Gly Gln Arg Thr Pro Tyr Gln Ala Leu Asp Ser Met Asp Asp Ala
645 650 655

Trp Ser Phe Arg Glu Leu Lys Ala Phe Val Arg Glu Ser Lys Thr Leu
660 665 670

Ser Asn Lys Lys Lys Glu Tyr Leu Leu Thr Glu Glu Asp Ile Ser Lys
675 680 685

Phe Asp Val Arg Lys Lys Phe Ile Glu Arg Asn Leu Val Asp Thr Arg
690 695 700

Tyr Ala Ser Arg Val Val Leu Asn Ala Leu Gln Glu His Phe Arg Ala
705 710 715 720

[0066] His Lys Ile Asp Thr Lys Val Ser Val Val Arg Gly Gln Phe Thr Ser
725 730 735

Gln Leu Arg Arg His Trp Gly Ile Glu Lys Thr Arg Asp Thr Tyr His
740 745 750

His His Ala Val Asp Ala Leu Ile Ile Ala Ala Ser Ser Gln Leu Asn
755 760 765

Leu Trp Lys Lys Gln Lys Asn Thr Leu Val Ser Tyr Ser Glu Asp Gln
770 775 780

Leu Leu Asp Ile Glu Thr Gly Glu Leu Ile Ser Asp Asp Glu Tyr Lys
785 790 795 800

Glu Ser Val Phe Lys Ala Pro Tyr Gln His Phe Val Asp Thr Leu Lys
805 810 815

Ser Lys Glu Phe Glu Asp Ser Ile Leu Phe Ser Tyr Gln Val Asp Ser
820 825 830

```
Lys Phe Asn Arg Lys Ile Ser Asp Ala Thr Ile Tyr Ala Thr Arg Gln
        835                 840                 845

Ala Lys Val Gly Lys Asp Lys Ala Asp Glu Thr Tyr Val Leu Gly Lys
    850                 855                 860

Ile Lys Asp Ile Tyr Thr Gln Asp Gly Tyr Asp Ala Phe Met Lys Ile
865                 870                 875                 880

Tyr Lys Lys Asp Lys Ser Lys Phe Leu Met Tyr Arg His Asp Pro Gln
                885                 890                 895

Thr Phe Glu Lys Val Ile Glu Pro Ile Leu Glu Asn Tyr Pro Asn Lys
            900                 905                 910

Gln Ile Asn Glu Lys Gly Lys Glu Val Pro Cys Asn Pro Phe Leu Lys
        915                 920                 925
```

[0067]

```
Tyr Lys Glu Glu His Gly Tyr Ile Arg Lys Tyr Ser Lys Lys Gly Asn
    930                 935                 940

Gly Pro Glu Ile Lys Ser Leu Lys Tyr Tyr Asp Ser Lys Leu Gly Asn
945                 950                 955                 960

His Ile Asp Ile Thr Pro Lys Asp Ser Asn Asn Lys Val Val Leu Gln
                965                 970                 975

Ser Val Ser Pro Trp Arg Ala Asp Val Tyr Phe Asn Lys Thr Thr Gly
            980                 985                 990

Lys Tyr Glu Ile Leu Gly Leu Lys  Tyr Ala Asp Leu Gln  Phe Glu Lys
        995                 1000                1005

Gly Thr  Gly Thr Tyr Lys Ile  Ser Gln Glu Lys Tyr  Asn Asp Ile
    1010                1015                1020

Lys Lys  Lys Glu Gly Val Asp  Ser Asp Ser Glu Phe  Lys Phe Thr
    1025                1030                1035
```

Leu Tyr Lys Asn Asp Leu Leu Leu Val Lys Asp Thr Glu Thr Lys
1040 1045 1050

Glu Gln Gln Leu Phe Arg Phe Leu Ser Arg Thr Met Pro Lys Gln
1055 1060 1065

Lys His Tyr Val Glu Leu Lys Pro Tyr Asp Lys Gln Lys Phe Glu
1070 1075 1080

Gly Gly Glu Ala Leu Ile Lys Val Leu Gly Asn Val Ala Asn Ser
1085 1090 1095

Gly Gln Cys Lys Lys Gly Leu Gly Lys Ser Asn Ile Ser Ile Tyr
1100 1105 1110

Lys Val Arg Thr Asp Val Leu Gly Asn Gln His Ile Ile Lys Asn
1115 1120 1125

[0068]

Glu Gly Asp Lys Pro Lys Leu Asp Phe Lys Arg Pro Ala Ala Thr
1130 1135 1140

Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys
1145 1150

<210> 53
<211> 340
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<400> 53
gagggcctat ttcccatgat tccttcatat ttgcatatac gatacaaggc tgttagagag 60

ataattggaa ttaatttgac tgtaaacaca aagatattag tacaaaatac gtgacgtaga 120

aagtaataat ttcttgggta gtttgcagtt ttaaaattat gttttaaaat ggactatcat 180

atgcttaccg taacttgaaa gtatttcgat ttcttggctt tatatatctt gtggaaagga 240

cgaaacaccg ttacttaaat cttgcagaag ctacaaagat aaggcttcat gccgaaatca 300

acaccctgtc attttatggc agggtgtttt cgttatttaa 340

<210> 54
<211> 360
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<220>
<221> 修饰_碱基
<222> (288)..(317)
<223> a, c, t, g, 未知或其它

<400> 54
gagggcctat ttcccatgat tccttcatat ttgcatatac gatacaaggc tgttagagag 60

[0069]

ataattggaa ttaatttgac tgtaaacaca aagatattag tacaaaatac gtgacgtaga 120

aagtaataat ttcttgggta gtttgcagtt ttaaaattat gttttaaaat ggactatcat 180

atgcttaccg taacttgaaa gtatttcgat ttcttggctt tatatatctt gtggaaagga 240

cgaaacaccg ggttttagag ctatgctgtt ttgaatggtc ccaaaacnnn nnnnnnnnnn 300

nnnnnnnnnn nnnnnnngtt ttagagctat gctgttttga atggtcccaa aacttttttt 360

<210> 55
<211> 318
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<220>
<221> 修饰_碱基
<222> (250)..(269)

<223> a, c, t, g, 未知或其它

<400> 55
gagggcctat ttcccatgat tccttcatat ttgcatatac gatacaaggc tgttagagag 60

ataattggaa ttaatttgac tgtaaacaca aagatattag tacaaaatac gtgacgtaga 120

aagtaataat ttcttgggta gtttgcagtt ttaaaattat gttttaaaat ggactatcat 180

atgcttaccg taacttgaaa gtatttcgat ttcttggctt tatatatctt gtggaaagga 240

cgaaacaccn nnnnnnnnnn nnnnnnnnng ttttagagct agaaatagca agttaaaata 300

aggctagtcc gttttttt 318

<210> 56
<211> 325
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

[0070]

<220>
<221> 修饰_碱基
<222> (250)..(269)
<223> a, c, t, g, 未知或其它

<400> 56
gagggcctat ttcccatgat tccttcatat ttgcatatac gatacaaggc tgttagagag 60

ataattggaa ttaatttgac tgtaaacaca aagatattag tacaaaatac gtgacgtaga 120

aagtaataat ttcttgggta gtttgcagtt ttaaaattat gttttaaaat ggactatcat 180

atgcttaccg taacttgaaa gtatttcgat ttcttggctt tatatatctt gtggaaagga 240

cgaaacaccn nnnnnnnnnn nnnnnnnnng ttttagagct agaaatagca agttaaaata 300

aggctagtcc gttatcattt ttttt 325

<210> 57
<211> 337
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<220>
<221> 修饰_碱基
<222> (250)..(269)
<223> a, c, t, g, 未知或其它

<400> 57
gagggcctat ttcccatgat tccttcatat ttgcatatac gatacaaggc tgttagagag 60

ataattggaa ttaatttgac tgtaaacaca aagatattag tacaaaatac gtgacgtaga 120

aagtaataat ttcttgggta gtttgcagtt ttaaaattat gttttaaaat ggactatcat 180

atgcttaccg taacttgaaa gtatttcgat ttcttggctt tatatatctt gtggaaagga 240

cgaaacaccn nnnnnnnnnn nnnnnnnnng ttttagagct agaaatagca agttaaaata 300

aggctagtcc gttatcaact tgaaaaagtg ttttttt 337

[0071] <210> 58
<211> 352
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<220>
<221> 修饰_碱基
<222> (250)..(269)
<223> a, c, t, g, 未知或其它

<400> 58
gagggcctat ttcccatgat tccttcatat ttgcatatac gatacaaggc tgttagagag 60

ataattggaa ttaatttgac tgtaaacaca aagatattag tacaaaatac gtgacgtaga 120

aagtaataat ttcttgggta gtttgcagtt ttaaaattat gttttaaaat ggactatcat 180

atgcttaccg taacttgaaa gtatttcgat ttcttggctt tatatatctt gtggaaagga 240

cgaaacaccn nnnnnnnnnn nnnnnnnnng ttttagagct agaaatagca agttaaaata 300

aggctagtcc gttatcaact tgaaaaagtg gcaccgagtc ggtgcttttt tt 352

<210> 59
<211> 5101
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<400> 59
cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 60

gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 120

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 180

aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 240

catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 300

[0072] catggtcgag gtgagcccca cgttctgctt cactctcccc atctcccccc cctccccacc 360

cccaattttg tatttattta ttttttaatt attttgtgca gcgatggggg cgggggggg 420

ggggggcgc gcgccaggcg gggcggggcg gggcgagggg cggggcgggg cgaggcggag 480

aggtgcggcg gcagccaatc agagcggcgc gctccgaaag tttcctttta tggcgaggcg 540

gcggcggcgg cggccctata aaaagcgaag cgcgcggcgg gcgggagtcg ctgcgacgct 600

gccttcgccc cgtgccccgc tccgccgccg cctcgcgccg cccgccccgg ctctgactga 660

ccgcgttact cccacaggtg agcgggcggg acggcccttc tcctccgggc tgtaattagc 720

tgagcaagag gtaagggttt aagggatggt tggttggtgg ggtattaatg tttaattacc 780

tggagcacct gcctgaaatc actttttttc aggttggacc ggtgccacca tggactataa 840

ggaccacgac ggagactaca aggatcatga tattgattac aaagacgatg acgataagat 900

ggccccaaag aagaagcgga aggtcggtat ccacggagtc ccagcagccg acaagaagta 960

cagcatcggc ctggacatcg gcaccaactc tgtgggctgg gccgtgatca ccgacgagta 1020

caaggtgccc agcaagaaat tcaaggtgct gggcaacacc gaccggcaca gcatcaagaa 1080

gaacctgatc ggagccctgc tgttcgacag cggcgaaaca gccgaggcca cccggctgaa 1140

gagaaccgcc agaagaagat acaccagacg gaagaaccgg atctgctatc tgcaagagat 1200

cttcagcaac gagatggcca aggtggacga cagcttcttc cacagactgg aagagtcctt 1260

cctggtggaa gaggataaga agcacgagcg gcaccccatc ttcggcaaca tcgtggacga 1320

ggtggcctac cacgagaagt accccaccat ctaccacctg agaaagaaac tggtggacag 1380

caccgacaag gccgacctgc ggctgatcta tctggccctg gcccacatga tcaagttccg 1440

gggccacttc ctgatcgagg gcgacctgaa ccccgacaac agcgacgtgg acaagctgtt 1500

catccagctg gtgcagacct acaaccagct gttcgaggaa aaccccatca acgccagcgg 1560

cgtggacgcc aaggccatcc tgtctgccag actgagcaag agcagacggc tggaaaatct 1620

gatcgcccag ctgcccggcg agaagaagaa tggcctgttc ggcaacctga ttgccctgag 1680

cctgggcctg acccccaact tcaagagcaa cttcgacctg gccgaggatg ccaaactgca 1740

gctgagcaag gacacctacg acgacgacct ggacaacctg ctggcccaga tcggcgacca 1800

gtacgccgac ctgtttctgg ccgccaagaa cctgtccgac gccatcctgc tgagcgacat 1860

[0073]

cctgagagtg aacaccgaga tcaccaaggc ccccctgagc gcctctatga tcaagagata 1920

cgacgagcac caccaggacc tgaccctgct gaaagctctc gtgcggcagc agctgcctga 1980

gaagtacaaa gagattttct tcgaccagag caagaacggc tacgccggct acattgacgg 2040

cggagccagc caggaagagt tctacaagtt catcaagccc atcctggaaa agatggacgg 2100

caccgaggaa ctgctcgtga agctgaacag agaggacctg ctgcggaagc agcggacctt 2160

cgacaacggc agcatccccc accagatcca cctgggagag ctgcacgcca ttctgcggcg 2220

gcaggaagat ttttacccat tcctgaagga caaccgggaa aagatcgaga agatcctgac 2280

cttccgcatc ccctactacg tgggccctct ggccagggga aacagcagat tcgcctggat 2340

gaccagaaag agcgaggaaa ccatcacccc ctggaacttc gaggaagtgg tggacaaggg 2400

cgcttccgcc cagagcttca tcgagcggat gaccaacttc gataagaacc tgcccaacga 2460

gaaggtgctg cccaagcaca gcctgctgta cgagtacttc accgtgtata acgagctgac 2520

caaagtgaaa tacgtgaccg agggaatgag aaagcccgcc ttcctgagcg gcgagcagaa 2580

aaaggccatc gtggacctgc tgttcaagac caaccggaaa gtgaccgtga agcagctgaa 2640

agaggactac ttcaagaaaa tcgagtgctt cgactccgtg gaaatctccg gcgtggaaga 2700

tcggttcaac gcctccctgg gcacatacca cgatctgctg aaaattatca aggacaagga 2760

cttcctggac aatgaggaaa acgaggacat tctggaagat atcgtgctga ccctgacact 2820

gtttgaggac agagagatga tcgaggaacg gctgaaaacc tatgcccacc tgttcgacga 2880

caaagtgatg aagcagctga agcggcggag atacaccggc tggggcaggc tgagccggaa 2940

gctgatcaac ggcatccggg acaagcagtc cggcaagaca atcctggatt tcctgaagtc 3000

cgacggcttc gccaacagaa acttcatgca gctgatccac gacgacagcc tgacctttaa 3060

agaggacatc cagaaagccc aggtgtccgg ccagggcgat agcctgcacg agcacattgc 3120

caatctggcc ggcagccccg ccattaagaa gggcatcctg cagacagtga aggtggtgga 3180

cgagctcgtg aaagtgatgg gccggcacaa gcccgagaac atcgtgatcg aaatggccag 3240

agagaaccag accacccaga agggacagaa gaacagccgc gagagaatga agcggatcga 3300

agagggcatc aaagagctgg gcagccagat cctgaaagaa caccccgtgg aaaacaccca 3360

[0074] gctgcagaac gagaagctgt acctgtacta cctgcagaat gggcgggata tgtacgtgga 3420

ccaggaactg gacatcaacc ggctgtccga ctacgatgtg gaccatatcg tgcctcagag 3480

ctttctgaag gacgactcca tcgacaacaa ggtgctgacc agaagcgaca agaaccgggg 3540

caagagcgac aacgtgccct ccgaagaggt cgtgaagaag atgaagaact actggcggca 3600

gctgctgaac gccaagctga ttacccagag aaagttcgac aatctgacca aggccgagag 3660

aggcggcctg agcgaactgg ataaggccgg cttcatcaag agacagctgg tggaaacccg 3720

gcagatcaca aagcacgtgg cacagatcct ggactcccgg atgaacacta agtacgacga 3780

gaatgacaag ctgatccggg aagtgaaagt gatcaccctg aagtccaagc tggtgtccga 3840

tttccggaag gatttccagt tttacaaagt gcgcgagatc aacaactacc accacgccca 3900

cgacgcctac ctgaacgccg tcgtgggaac cgccctgatc aaaaagtacc ctaagctgga 3960

aagcgagttc gtgtacggcg actacaaggt gtacgacgtg cggaagatga tcgccaagag 4020

cgagcaggaa atcggcaagg ctaccgccaa gtacttcttc tacagcaaca tcatgaactt 4080

tttcaagacc gagattaccc tggccaacgg cgagatccgg aagcggcctc tgatcgagac 4140

[0075]

aaacggcgaa accggggaga tcgtgtggga taagggccgg gattttgcca ccgtgcggaa 4200

agtgctgagc atgccccaag tgaatatcgt gaaaaagacc gaggtgcaga caggcggctt 4260

cagcaaagag tctatcctgc ccaagaggaa cagcgataag ctgatcgcca gaaagaagga 4320

ctgggaccct aagaagtacg gcggcttcga cagccccacc gtggcctatt ctgtgctggt 4380

ggtggccaaa gtggaaaagg gcaagtccaa gaaactgaag agtgtgaaag agctgctggg 4440

gatcaccatc atggaaagaa gcagcttcga gaagaatccc atcgactttc tggaagccaa 4500

gggctacaaa gaagtgaaaa aggacctgat catcaagctg cctaagtact ccctgttcga 4560

gctggaaaac ggccggaaga gaatgctggc ctctgccggc gaactgcaga agggaaacga 4620

actggccctg ccctccaaat atgtgaactt cctgtacctg gccagccact atgagaagct 4680

gaagggctcc cccgaggata atgagcagaa acagctgttt gtggaacagc acaagcacta 4740

cctggacgag atcatcgagc agatcagcga gttctccaag agagtgatcc tggccgacgc 4800

taatctggac aaagtgctgt ccgcctacaa caagcaccgg gataagccca tcagagagca 4860

ggccgagaat atcatccacc tgtttaccct gaccaatctg ggagcccctg ccgccttcaa 4920

gtactttgac accaccatcg accggaagag gtacaccagc accaaagagg tgctggacgc 4980

caccctgatc caccagagca tcaccggcct gtacgagaca cggatcgacc tgtctcagct 5040

gggaggcgac tttcttttte ttagcttgac cagctttctt agtagcagca ggacgcttta 5100

a 5101

<210> 60
<211> 137
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<220>
<221> 修饰_碱基
<222> (1)..(20)
<223> a, c, t, g, 未知或其它

<400> 60
nnnnnnnnnn nnnnnnnnnn gttattgtac tctcaagatt tagaaataaa tcttgcagaa 60

gctacaaaga taaggcttca tgccgaaatc aacaccctgt cattttatgg cagggtgttt 120

tcgttattta atttttt 137

<210> 61
<211> 123
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<220>
<221> 修饰_碱基
<222> (1)..(20)
<223> a, c, t, g, 未知或其它

[0076]
<400> 61
nnnnnnnnnn nnnnnnnnnn gttattgtac tctcagaaat gcagaagcta caaagataag 60

gcttcatgcc gaaatcaaca ccctgtcatt ttatggcagg gtgttttcgt tatttaattt 120

ttt 123

<210> 62
<211> 110
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<220>
<221> 修饰_碱基
<222> (1)..(20)
<223> a, c, t, g, 未知或其它

<400> 62
nnnnnnnnnn nnnnnnnnnn gttattgtac tctcagaaat gcagaagcta caaagataag 60

gcttcatgcc gaaatcaaca ccctgtcatt ttatggcagg gtgttttttt 110

<210> 63
<211> 137
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<220>
<221> 修饰_碱基
<222> (1)..(20)
<223> a, c, t, g, 未知或其它

<400> 63
nnnnnnnnnn nnnnnnnnnn gttattgtac tctcaagatt tagaaataaa tcttgcagaa 60

gctacaatga taaggcttca tgccgaaatc aacaccctgt cattttatgg cagggtgttt 120

[0077] tcgttattta atttttt 137

<210> 64
<211> 123
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<220>
<221> 修饰_碱基
<222> (1)..(20)
<223> a, c, t, g, 未知或其它

<400> 64
nnnnnnnnnn nnnnnnnnnn gttattgtac tctcagaaat gcagaagcta caatgataag 60

gcttcatgcc gaaatcaaca ccctgtcatt ttatggcagg gtgttttcgt tatttaattt 120

ttt 123

<210> 65
<211> 110
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<220>
<221> 修饰_碱基
<222> (1)..(20)
<223> a, c, t, g, 未知或其它

<400> 65
nnnnnnnnnn nnnnnnnnnn gttattgtac tctcagaaat gcagaagcta caatgataag 60

gcttcatgcc gaaatcaaca ccctgtcatt ttatggcagg gtgttttttt 110

<210> 66
<211> 107
<212> DNA
[0078] <213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<220>
<221> 修饰_碱基
<222> (1)..(20)
<223> a, c, t, g, 未知或其它

<400> 66
nnnnnnnnnn nnnnnnnnnn gttttagagc tgtggaaaca cagcgagtta aaataaggct 60

tagtccgtac tcaacttgaa aaggtggcac cgattcggtg ttttttt 107

<210> 67
<211> 4263
<212> DNA
<213> 人工序列

<220>
<221> 来源

<223> /注释="人工序列的描述：合成
多核苷酸"

<400> 67

atgaaaaggc cggcggccac gaaaaaggcc ggccaggcaa aaaagaaaaa gaccaagccc 60

tacagcatcg gcctggacat cggcaccaat agcgtgggct gggccgtgac caccgacaac 120

tacaaggtgc ccagcaagaa aatgaaggtg ctgggcaaca cctccaagaa gtacatcaag 180

aaaaacctgc tgggcgtgct gctgttcgac agcggcatta cagccgaggg cagacggctg 240

aagagaaccg ccagacggcg gtacacccgg cggagaaaca gaatcctgta tctgcaagag 300

atcttcagca ccgagatggc taccctggac gacgccttct tccagcggct ggacgacagc 360

ttcctggtgc ccgacgacaa gcgggacagc aagtacccca tcttcggcaa cctggtggaa 420

gagaaggcct accacgacga gttccccacc atctaccacc tgagaaagta cctggccgac 480

agcaccaaga aggccgacct gagactggtg tatctggccc tggcccacat gatcaagtac 540

cggggccact tcctgatcga gggcgagttc aacagcaaga acaacgacat ccagaagaac 600

ttccaggact tcctggacac ctacaacgcc atcttcgaga gcgacctgtc cctggaaaac 660

[0079]

agcaagcagc tggaagagat cgtgaaggac aagatcagca agctggaaaa gaaggaccgc 720

atcctgaagc tgttccccgg cgagaagaac agcggaatct tcagcgagtt tctgaagctg 780

atcgtgggca accaggccga cttcagaaag tgcttcaacc tggacgagaa agccagcctg 840

cacttcagca aagagagcta cgacgaggac ctggaaaccc tgctgggata tatcggcgac 900

gactacagcg acgtgttcct gaaggccaag aagctgtacg acgctatcct gctgagcggc 960

ttcctgaccg tgaccgacaa cgagacagag gccccactga gcagcgccat gattaagcgg 1020

tacaacgagc acaaagagga tctggctctg ctgaaagagt acatccggaa catcagcctg 1080

aaaacctaca atgaggtgtt caaggacgac accaagaacg gctacgccgg ctacatcgac 1140

ggcaagacca accaggaaga tttctatgtg tacctgaaga agctgctggc cgagttcgag 1200

ggggccgact actttctgga aaaaatcgac cgcgaggatt tcctgcggaa gcagcggacc 1260

ttcgacaacg gcagcatccc ctaccagatc catctgcagg aaatgcgggc catcctggac 1320

aagcaggcca agttctaccc attcctggcc aagaacaaag agcggatcga gaagatcctg 1380

accttccgca tcccttacta cgtgggcccc ctggccagag gcaacagcga ttttgcctgg 1440

tccatccgga agcgcaatga gaagatcacc ccctggaact tcgaggacgt gatcgacaaa 1500

gagtccagcg ccgaggcctt catcaaccgg atgaccagct tcgacctgta cctgcccgag 1560

gaaaaggtgc tgcccaagca cagcctgctg tacgagacat tcaatgtgta taacgagctg 1620

accaaagtgc ggtttatcgc cgagtctatg cgggactacc agttcctgga ctccaagcag 1680

aaaaaggaca tcgtgcggct gtacttcaag gacaagcgga aagtgaccga taaggacatc 1740

atcgagtacc tgcacgccat ctacggctac gatggcatcg agctgaaggg catcgagaag 1800

cagttcaact ccagcctgag cacataccac gacctgctga acattatcaa cgacaaagaa 1860

tttctggacg actccagcaa cgaggccatc atcgaagaga tcatccacac cctgaccatc 1920

tttgaggacc gcgagatgat caagcagcgg ctgagcaagt tcgagaacat cttcgacaag 1980

agcgtgctga aaaagctgag cagacggcac tacaccggct ggggcaagct gagcgccaag 2040

ctgatcaacg gcatccggga cgagaagtcc ggcaacacaa tcctggacta cctgatcgac 2100

gacggcatca gcaaccggaa cttcatgcag ctgatccacg acgacgccct gagcttcaag 2160

[0080] aagaagatcc agaaggccca gatcatcggg gacgaggaca agggcaacat caaagaagtc 2220

gtgaagtccc tgcccggcag ccccgccatc aagaagggaa tcctgcagag catcaagatc 2280

gtggacgagc tcgtgaaagt gatgggcggc agaaagcccg agagcatcgt ggtggaaatg 2340

gctagagaga accagtacac caatcagggc aagagcaaca gccagcagag actgaagaga 2400

ctggaaaagt ccctgaaaga gctgggcagc aagattctga aagagaatat ccctgccaag 2460

ctgtccaaga tcgacaacaa cgccctgcag aacgaccggc tgtacctgta ctacctgcag 2520

aatggcaagg acatgtatac aggcgacgac ctggatatcg accgcctgag caactacgac 2580

atcgaccata ttatccccca ggccttcctg aaagacaaca gcattgacaa caaagtgctg 2640

gtgtcctccg ccagcaaccg cggcaagtcc gatgatgtgc ccagcctgga agtcgtgaaa 2700

aagagaaaga ccttctggta tcagctgctg aaaagcaagc tgattagcca gaggaagttc 2760

gacaacctga ccaaggccga gagaggcggc ctgagccctg aagataaggc cggcttcatc 2820

cagagacagc tggtggaaac ccggcagatc accaagcacg tggccagact gctggatgag 2880

aagtttaaca acaagaagga cgagaacaac cgggccgtgc ggaccgtgaa gatcatcacc 2940

ctgaagtcca ccctggtgtc ccagttccgg aaggacttcg agctgtataa agtgcgcgag 3000

atcaatgact ttcaccacgc ccacgacgcc tacctgaatg ccgtggtggc ttccgccctg 3060

ctgaagaagt accctaagct ggaacccgag ttcgtgtacg gcgactaccc caagtacaac 3120

tccttcagag agcggaagtc cgccaccgag aaggtgtact tctactccaa catcatgaat 3180

atctttaaga agtccatctc cctggccgat ggcagagtga tcgagcggcc cctgatcgaa 3240

gtgaacgaag agacaggcga gagcgtgtgg aacaaagaaa gcgacctggc caccgtgcgg 3300

cgggtgctga gttatcctca agtgaatgtc gtgaagaagg tggaagaaca gaaccacggc 3360

ctggatcggg gcaagcccaa gggcctgttc aacgccaacc tgtccagcaa gcctaagccc 3420

aactccaacg agaatctcgt gggggccaaa gagtacctgg accctaagaa gtacggcgga 3480

tacgccggca tctccaatag cttcaccgtg ctcgtgaagg gcacaatcga gaagggcgct 3540

aagaaaaaga tcacaaacgt gctggaattt caggggatct ctatcctgga ccggatcaac 3600

taccggaagg ataagctgaa ctttctgctg gaaaaaggct acaaggacat tgagctgatt 3660

atcgagctgc ctaagtactc cctgttcgaa ctgagcgacg gctccagacg gatgctggcc 3720

[0081]

tccatcctgt ccaccaacaa caagcggggc gagatccaca agggaaacca gatcttcctg 3780

agccagaaat ttgtgaaact gctgtaccac gccaagcgga tctccaacac catcaatgag 3840

aaccaccgga aatacgtgga aaaccacaag aaagagtttg aggaactgtt ctactacatc 3900

ctggagttca acgagaacta tgtgggagcc aagaagaacg gcaaactgct gaactccgcc 3960

ttccagagct ggcagaacca cagcatcgac gagctgtgca gctccttcat cggccctacc 4020

ggcagcgagc ggaagggact gtttgagctg acctccagag gctctgccgc cgactttgag 4080

ttcctgggag tgaagatccc ccggtacaga gactacaccc cctctagtct gctgaaggac 4140

gccaccctga tccaccagag cgtgaccggc ctgtacgaaa cccggatcga cctggctaag 4200

ctgggcgagg gaaagcgtcc tgctgctact aagaaagctg gtcaagctaa gaaaaagaaa 4260

taa 4263

<210> 68
<211> 53
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 68
tcctagcagg atttctgata ttactgtcac gttttagagc tatgctgttt tga 53

<210> 69
<211> 53
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 69
gtgacagtaa tatcagaaat cctgctagga gttttgggac cattcaaaac agc 53

<210> 70
[0082] <211> 25
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 70
gggtttcaag tctttgtagc aagag 25

<210> 71
<211> 24
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 71
gccaatgaac gggaaccctt ggtc 24

<210> 72
<211> 25
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成引物"

<220>
<221> 修饰_碱基
<222> (1)..(4)
<223> a, c, t, g, 未知或其它

<400> 72
nnnngacgag gcaatggctg aaatc 25

<210> 73
<211> 25
<212> DNA
<213> 人工序列

[0083]

<220>
<221> 来源
<223> /注释="人工序列的描述：合成引物"

<220>
<221> 修饰_碱基
<222> (1)..(4)
<223> a, c, t, g, 未知或其它

<400> 73
nnnnttattt ggctcatatt tgctg 25

<210> 74
<211> 25
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成引物"

<400> 74
ctttacacca atcgctgcaa cagac 25

<210> 75
<211> 45
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 75
caaaatttct agtcttcttt gcctttcccc ataaaaccct cctta 45

<210> 76
<211> 45
<212> DNA
<213> 人工序列

[0084]

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 76
agggttttat ggggaaaggc aaagaagact agaaattttg atacc 45

<210> 77
<211> 25
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 77
cttacggtgc ataaagtcaa tttcc 25

<210> 78
<211> 21
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 78
tggctcgatt tcagccattg c 21

<210> 79
<211> 43
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<220>
<221> 修饰_碱基
<222> (33)..(33)
<223> a, c, t, g, 未知或其它

[0085] <400> 79
ctttgacgag gcaatggctg aaatcgagcc aanaaagcgc aag 43

<210> 80
<211> 43
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<220>
<221> 修饰_碱基
<222> (34)..(34)
<223> a, c, t, g, 未知或其它

<400> 80
ctttgacgag gcaatggctg aaatcgagcc aaanaagcgc aag 43

<210> 81
<211> 43

<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<220>
<221> 修饰_碱基
<222> (35)..(35)
<223> a, c, t, g, 未知或其它

<400> 81
ctttgacgag gcaatggctg aaatcgagcc aaaanagcgc aag 43

<210> 82
<211> 43
<212> DNA
<213> 人工序列

<220>
<221> 来源
[0086] <223> /注释="人工序列的描述：合成
引物"

<220>
<221> 修饰_碱基
<222> (36)..(36)
<223> a, c, t, g, 未知或其它

<400> 82
ctttgacgag gcaatggctg aaatcgagcc aaaaangcgc aag 43

<210> 83
<211> 43
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<220>
<221> 修饰_碱基

<222> (37)..(37)
<223> a, c, t, g, 未知或其它

<400> 83
ctttgacgag gcaatggctg aaatcgagcc aaaaaancgc aag 43

<210> 84
<211> 43
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<220>
<221> 修饰_碱基
<222> (38)..(38)
<223> a, c, t, g, 未知或其它

<400> 84
ctttgacgag gcaatggctg aaatcgagcc aaaaaagngc aag 43

[0087]

<210> 85
<211> 46
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<220>
<221> 修饰_碱基
<222> (39)..(39)
<223> a, c, t, g, 未知或其它

<400> 85
ctttgacgag gcaatggctg aaatcgagcc aaaaaagcnc aagaag 46

<210> 86
<211> 46
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<220>
<221> 修饰_碱基
<222> (40)..(40)
<223> a, c, t, g, 未知或其它

<400> 86
ctttgacgag gcaatggctg aaatcgagcc aaaaaagcgn aagaag 46

<210> 87
<211> 46
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

[0088]

<220>
<221> 修饰_碱基
<222> (41)..(41)
<223> a, c, t, g, 未知或其它

<400> 87
ctttgacgag gcaatggctg aaatcgagcc aaaaaagcgc nagaag 46

<210> 88
<211> 23
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 88
gcgctttttt ggctcgattt cag 23

<210> 89

<210> 40
<211> DNA
<212> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<220>
<221> 修饰_碱基
<222> (31)..(31)
<223> a, c, t, g, 未知或其它

<400> 89
caatggctga aatcgagcca aaaaagcgca ngaagaaatc 40

<210> 90
<211> 40
<212> DNA
<213> 人工序列

<220>
[0089] <221> 来源
<223> /注释="人工序列的描述：合成
引物"

<220>
<221> 修饰_碱基
<222> (32)..(32)
<223> a, c, t, g, 未知或其它

<400> 90
caatggctga aatcgagcca aaaaagcgca anaagaaatc 40

<210> 91
<211> 40
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<220>

<220>
<221> 修饰_碱基
<222> (33)..(33)
<223> a, c, t, g, 未知或其它

<400> 91
caatggctga aatcgagcca aaaaagcgca agnagaaatc 40

<210> 92
<211> 40
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述: 合成
引物"

<220>
<221> 修饰_碱基
<222> (34)..(34)
<223> a, c, t, g, 未知或其它

<400> 92
[0090] caatggctga aatcgagcca aaaaagcgca agangaaatc 40

<210> 93
<211> 40
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述: 合成
引物"

<220>
<221> 修饰_碱基
<222> (35)..(35)
<223> a, c, t, g, 未知或其它

<400> 93
caatggctga aatcgagcca aaaaagcgca agaanaaatc 40

<210> 94
<211> 44
<212> DNA

<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<220>
<221> 修饰_碱基
<222> (36)..(36)
<223> a, c, t, g, 未知或其它

<400> 94
caatggctga aatcgagcca aaaaagcgca agaagnaatc aacc 44

<210> 95
<211> 44
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
[0091] 引物"

<220>
<221> 修饰_碱基
<222> (37)..(37)
<223> a, c, t, g, 未知或其它

<400> 95
caatggctga aatcgagcca aaaaagcgca agaaganatc aacc 44

<210> 96
<211> 44
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<220>
<221> 修饰_碱基
<222> (38)..(38)

<223> a, c, t, g, 未知或其它

<400> 96
caatggctga aatcgagcca aaaaagcgca agaagaantc aacc 44

<210> 97
<211> 44
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述: 合成
引物"

<220>
<221> 修饰_碱基
<222> (39)..(39)
<223> a, c, t, g, 未知或其它

<400> 97
caatggctga aatcgagcca aaaaagcgca agaagaaanc aacc 44

[0092]

<210> 98
<211> 47
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述: 合成
引物"

<220>
<221> 修饰_碱基
<222> (40)..(40)
<223> a, c, t, g, 未知或其它

<400> 98
caatggctga aatcgagcca aaaaagcgca agaagaaatn aaccagc 47

<210> 99
<211> 47
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<220>
<221> 修饰_碱基
<222> (41)..(41)
<223> a, c, t, g, 未知或其它

<400> 99
caatggctga aatcgagcca aaaaagcgca agaagaaatc naccagc 47

<210> 100
<211> 31
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

[0093] <400> 100
gatcctccat ccgtacaacc cacaaccctg g 31

<210> 101
<211> 31
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 101
aattccaggg ttgtgggttg tacggatgga g 31

<210> 102
<211> 34
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成

引物″

<400> 102
catggatcct atttcttaat aactaaaaat atgg 34

<210> 103
<211> 33
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释=″人工序列的描述：合成
引物″

<400> 103
catgaattca actcaacaag tctcagtgtg ctg 33

<210> 104
<211> 35
<212> DNA
<213> 人工序列

[0094] <220>
<221> 来源
<223> /注释=″人工序列的描述：合成
引物″

<400> 104
aaacattttt tctccattta ggaaaaagga tgctg 35

<210> 105
<211> 35
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释=″人工序列的描述：合成
引物″

<400> 105
aaaacagcat ccttttcct aaatggagaa aaaat 35

<210> 106
<211> 35
<212> DNA

<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 106
aaaccttaaa tcagtcacaa atagcagcaa aattg 35

<210> 107
<211> 35
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 107
aaaacaattt tgctgctatt tgtgactgat ttaag 35

[0095] <210> 108
<211> 35
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 108
aaacttttca tcatacgacc aatctgcttt atttg 35

<210> 109
<211> 35
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 109
aaaacaaata aagcagattg gtcgtatgat gaaaa 35

<210> 110
<211> 35
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 110
aaactcgtcc agaagttatc gtaaaagaaa tcgag 35

<210> 111
<211> 35
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

[0096] <400> 111
aaaactcgat ttcttttacg ataacttctg gacga 35

<210> 112
<211> 35
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 112
aaacaatctc tccaaggttt ccttaaaaat ctctg 35

<210> 113
<211> 35
<212> DNA
<213> 人工序列

<220>
<221> 来源

<223> /注释="人工序列的描述：合成
引物"

<400> 113
aaaacagaga tttttaagga aaccttggag agatt 35

<210> 114
<211> 35
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 114
aaacgccatc gtcaggaaga agctatgctt gagtg 35

<210> 115
<211> 35
<212> DNA
<213> 人工序列

[0097]

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 115
aaaacactca agcatagctt cttcctgacg atggc 35

<210> 116
<211> 35
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 116
aaacatctct atacttattg aaatttcttt gtatg 35

<210> 117
<211> 35

<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 117
aaaacataca aagaaatttc aataagtata gagat 35

<210> 118
<211> 35
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 118
aaactagctg tgatagtccg caaaaccagc cttcg 35

[0098]

<210> 119
<211> 35
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 119
aaaacgaagg ctggttttgc ggactatcac agcta 35

<210> 120
<211> 35
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 120

aaacatcgga aggtcgagca agtaattatc ttttg 35

<210> 121
<211> 35
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 121
aaaacaaaag ataattactt gctcgacctt ccgat 35

<210> 122
<211> 35
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
[0099] 引物"

<400> 122
aaacaagatg gtatcgcaaa gtaagtgaca ataag 35

<210> 123
<211> 35
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 123
aaaacttatt gtcacttact ttgcgatacc atctt 35

<210> 124
<211> 52
<212> DNA
<213> 人工序列

<220>

<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 124
gagacctttg agcttccgag actggtctca gttttgggac cattcaaaac ag 52

<210> 125
<211> 52
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 125
tgagaccagt ctcggaagct caaaggtctc gttttagagc tatgctgttt tg 52

<210> 126
<211> 35
<212> DNA
[0100] <213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 126
aaactacttt acgcagcgcg gagttcggtt ttttg 35

<210> 127
<211> 35
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 127
aaaacaaaaa accgaactcc gcgctgcgta aagta 35

<210> 128

<211> 45
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 128
atgccggtac tgccgggcct cttgcgggat tacgaaatca tcctg 45

<210> 129
<211> 45
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 129
[0101] gtgactggcg atgctgtcgg aatggacgat cacactactc ttctt 45

<210> 130
<211> 50
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 130
ttaagaaata atcttcatct aaaatatact tcagtcacct cctagctgac 50

<210> 131
<211> 48
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 131
attgatttga gtcagctagg aggtgactga agtatatttt agatgaag 48

<210> 132
<211> 85
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 132
gagacctttg agcttccgag actggtctca gttttgggac cattcaaaac agcatagctc 60

taaaacctcg tagactattt ttgtc 85

<210> 133
<211> 84
<212> DNA
<213> 人工序列

[0102] <220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 133
gagaccagtc tcggaagctc aaaggtctcg ttttagagct atgctgtttt gaatggtccc 60

aaaacttcag cacactgaga cttg 84

<210> 134
<211> 21
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 134
agtcatccca gcaacaaatg g 21

<210> 135

<211> 31
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 135
cgtggtaaat cggataacgt tccaagtgaa g 31

<210> 136
<211> 22
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 136
[0103] tgctcttctt cacaaacaag gg 22

<210> 137
<211> 21
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 137
aagccaaagt ttggcaccac c 21

<210> 138
<211> 22
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 138
gtagcttatt cagtcctagt gg 22

<210> 139
<211> 45
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 139
cgtttgttga actaatgggt gcaaattacg aatcttctcc tgacg 45

<210> 140
<211> 45
<212> DNA
<213> 人工序列

<220>
<221> 来源
[0104] <223> /注释="人工序列的描述：合成
引物"

<400> 140
cgtcaggaga agattcgtaa tttgcaccca ttagttcaac aaacg 45

<210> 141
<211> 48
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 141
gatattatgg agcctatttt tgtgggtttt taggcataaa actatatg 48

<210> 142
<211> 48
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 142
catatagttt tatgcctaaa aacccacaaa aataggctcc ataatatc 48

<210> 143
<211> 27
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 143
attatttctt aataactaaa aatatgg 27

[0105]
<210> 144
<211> 24
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 144
cgtgtacaat tgctagcgta cggc 24

<210> 145
<211> 24
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 145
gcaccggtga tcactagtcc tagg 24

<210> 146
<211> 47
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 146
cctaggacta gtgatcaccg gtgcaaatat gagccaaata aatatat 47

<210> 147
<211> 44
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 147
[0106] gccgtacgct agcaattgta cacgtttgtt gaactaatgg gtgc 44

<210> 148
<211> 25
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 148
ttcaaatttt cccatttgat tctcc 25

<210> 149
<211> 47
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 149
ccatattttt agttattaag aaataatacc agccatcagt cacctcc 47

<210> 150
<211> 23
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 150
agacgattca atagacaata agg 23

<210> 151
<211> 45
<212> DNA
<213> 人工序列

[0107]

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 151
gttttgggac cattcaaaac agcatagctc taaaacctcg tagac 45

<210> 152
<211> 50
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 152
gctatgctgt tttgaatggt cccaaaacca ttattttaac acacgaggtg 50

<210> 153
<211> 50
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 153
gctatgctgt tttgaatggt cccaaaacgc acccattagt tcaacaaacg 50

<210> 154
<211> 23
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 154
aattcttttc ttcatcatcg gtc 23

[0108]
<210> 155
<211> 24
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 155
aagaaagaat gaagattgtt catg 24

<210> 156
<211> 25
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 156
ggtactaatc aaaatagtga ggagg 25

<210> 157
<211> 23
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 157
gtttttcaaa atctgcggtt gcg 23

<210> 158
<211> 26
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

[0109]

<400> 158
aaaaattgaa aaaatggtgg aaacac 26

<210> 159
<211> 53
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 159
atttcgtaaa cggtatcggt ttcttttaaa gttttgggac cattcaaaac agc 53

<210> 160
<211> 53
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成

引物″

<400> 160
tttaaaagaa accgataccg tttacgaaat gttttagagc tatgctgttt tga 53

<210> 161
<211> 53
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释=″人工序列的描述：合成
引物″

<400> 161
aaacggtatc ggtttctttt aaattcaatt gttttgggac cattcaaaac agc 53

<210> 162
<211> 53
<212> DNA
<213> 人工序列

[0110] <220>
<221> 来源
<223> /注释=″人工序列的描述：合成
引物″

<400> 162
aattgaattt aaaagaaacc gataccgttt gttttagagc tatgctgttt tga 53

<210> 163
<211> 39
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释=″人工序列的描述：合成
引物″

<400> 163
gttccttaaa ccaaaacggt atcggtttct tttaaattc 39

<210> 164
<211> 47
<212> DNA

<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 164
gaaaccgata ccgttttggt ttaaggaaca ggtaaagggc atttaac 47

<210> 165
<211> 22
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 165
cgatttcagc cattgcctcg tc 22

[0111] <210> 166
<211> 56
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<220>
<221> 修饰_碱基
<222> (29)..(33)
<223> a, c, t, g, 未知或其它

<400> 166
gcctttgacg aggcaatggc tgaaatcgnn nnnaaaaagc gcaagaagaa atcaac 56

<210> 167
<211> 53
<212> DNA
<213> 人工序列

<220>

<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 167
tccgtacaac ccacaaccct gctagtgagc gttttgggac cattcaaaac agc 53

<210> 168
<211> 53
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 168
gctcactagc agggttgtgg gttgtacgga gttttagagc tatgctgttt tga 53

<210> 169
<211> 23
<212> DNA
[0112] <213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 169
ttgttgccac tcttccttct ttc 23

<210> 170
<211> 41
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 170
cagggttgtg ggttgttgcg atggagttaa ctcccatctc c 41

<210> 171

<211> 41
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 171
gggagttaac tccatcgcaa caacccacaa ccctgctagt g 41

<210> 172
<211> 22
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 172
[0113] gtggtatcta tcgtgatgtg ac 22

<210> 173
<211> 23
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 173
ttaccgaaac ggaatttatc tgc 23

<210> 174
<211> 22
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 174
aaagctagag ttccgcaatt gg 22

<210> 175
<211> 37
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 175
gtgggttgta cggattgagt taactcccat ctccttc 37

<210> 176
<211> 38
<212> DNA
<213> 人工序列

<220>
<221> 来源
[0114] <223> /注释="人工序列的描述：合成
引物"

<400> 176
gatgggagtt aactcaatcc gtacaaccca caaccctg 38

<210> 177
<211> 40
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 177
gcttcaccta ttgcagcacc aattgaccac atgaagatag 40

<210> 178
<211> 41
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 178
gtggtcaatt ggtgctgcaa taggtgaagc taatggtgat g 41

<210> 179
<211> 39
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 179
ctgatttgta ttaattttga gacattatgc ttcaccttc 39

[0115]
<210> 180
<211> 40
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 180
gcataatgtc tcaaaattaa tacaaatcag tgaaatcatg 40

<210> 181
<211> 52
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 181
gttttgggac cattcaaaac agcatagctc taaaacgtga cagtaatatc ag 52

<210> 182
<211> 53
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 182
gttttagagc tatgctgttt tgaatggtcc caaaacgctc actagcaggg ttg 53

<210> 183
<211> 59
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 183
[0116] atactttacg cagcgcggag ttcggttttg taggagtggt agtatataca cgagtacat 59

<210> 184
<211> 33
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 184
gctcactagc agggttgtgg gttgtacgga tgg 33

<210> 185
<211> 33
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 185
tcctagcagg atttctgata ttactgtcac tgg 33

<210> 186
<211> 33
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成寡核苷酸"

<400> 186
tttaaaagaa accgataccg tttacgaaat tgg 33

<210> 187
<211> 84
<212> DNA
<213> 人工序列

[0117]

<220>
<221> 来源
<223> /注释="人工序列的描述：合成寡核苷酸"

<400> 187
ggaaccattc ataacagcat agcaagttat aataaggcta gtccgttatc aacttgaaaa 60

agtggcaccg agtcggtgct tttt 84

<210> 188
<211> 36
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成寡核苷酸"

<400> 188
gttatagagc tatgctgtta tgaatggtcc caaaac 36

<210> 189
<211> 84

<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 189
ggaaccattc aatacagcat agcaagttaa tataaggcta gtccgttatc aacttgaaaa 60

agtggcaccg agtcggtgct tttt 84

<210> 190
<211> 36
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 190
[0118] gtattagagc tatgctgtat tgaatggtcc caaaac 36

<210> 191
<211> 103
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<220>
<221> 修饰_碱基
<222> (1)..(20)
<223> a, c, t, g, 未知或其它

<400> 191
nnnnnnnnnn nnnnnnnnnn gttttagagc tagaaatagc aagttaaaat aaggctagtc 60

cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt ttt 103

<210> 192

<211> 103
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<220>
<221> 修饰_碱基
<222> (1)..(20)
<223> a, c, t, g, 未知或其它

<400> 192
nnnnnnnnnn nnnnnnnnnn gtattagagc tagaaatagc aagttaatat aaggctagtc 60

cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt ttt 103

<210> 193
<211> 123
<212> DNA
<213> 人工序列

[0119]

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<220>
<221> 修饰_碱基
<222> (1)..(20)
<223> a, c, t, g, 未知或其它

<400> 193
nnnnnnnnnn nnnnnnnnnn gttttagagc tatgctgttt tggaaacaaa acagcatagc 60

aagttaaaat aaggctagtc cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt 120

ttt 123

<210> 194
<211> 123
<212> DNA
<213> 人工序列

<220>

<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<220>
<221> 修饰_碱基
<222> (1)..(20)
<223> a, c, t, g, 未知或其它

<400> 194
nnnnnnnnnn nnnnnnnnnn gtattagagc tatgctgtat tggaaacaat acagcatagc 60

aagttaatat aaggctagtc cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt 120

ttt 123

<210> 195
<211> 20
<212> DNA
<213> 人

<400> 195
gtcacctcca atgactaggg 20

[0120]

<210> 196
<211> 23
<212> DNA
<213> 人

<400> 196
gacatcgatg tcctccccat tgg 23

<210> 197
<211> 23
<212> DNA
<213> 人

<400> 197
gagtccgagc agaagaagaa ggg 23

<210> 198
<211> 23
<212> DNA
<213> 人

<400> 198

gcgccaccgg ttgatgtgat ggg 23

<210> 199
<211> 23
<212> DNA
<213> 人

<400> 199
ggggcacaga tgagaaactc agg 23

<210> 200
<211> 23
<212> DNA
<213> 人

<400> 200
gtacaaacgg cagaagctgg agg 23

<210> 201
<211> 23
<212> DNA
<213> 人

[0121]

<400> 201
ggcagaagct ggaggaggaa ggg 23

<210> 202
<211> 23
<212> DNA
<213> 人

<400> 202
ggagcccttc ttcttctgct cgg 23

<210> 203
<211> 23
<212> DNA
<213> 人

<400> 203
gggcaaccac aaacccacga ggg 23

<210> 204
<211> 23
<212> DNA

<213> 人

<400> 204
gctcccatca catcaaccgg tgg 23

<210> 205
<211> 23
<212> DNA
<213> 人

<400> 205
gtggcgcatt gccacgaagc agg 23

<210> 206
<211> 23
<212> DNA
<213> 人

<400> 206
ggcagagtgc tgcttgctgc tgg 23

[0122] <210> 207
<211> 23
<212> DNA
<213> 人

<400> 207
gcccctgcgt gggcccaagc tgg 23

<210> 208
<211> 23
<212> DNA
<213> 人

<400> 208
gagtggccag agtccagctt ggg 23

<210> 209
<211> 23
<212> DNA
<213> 人

<400> 209
ggcctcccca aagcctggcc agg 23

<210> 210
<211> 23
<212> DNA
<213> 人

<400> 210
ggggccgaga ttgggtgttc agg 23

<210> 211
<211> 23
<212> DNA
<213> 人

<400> 211
gtggcgagag gggccgagat tgg 23

<210> 212
<211> 23
<212> DNA
<213> 人

<400> 212
gagtgccgcc gaggcggggc ggg 23

[0123]

<210> 213
<211> 23
<212> DNA
<213> 人

<400> 213
ggagtgccgc cgaggcgggg cgg 23

<210> 214
<211> 23
<212> DNA
<213> 人

<400> 214
ggagaggagt gccgccgagg cgg 23

<210> 215
<211> 20
<212> DNA
<213> 人工序列

<220>

<220>
<221> 来源
<223> /注释="人工序列的描述：合成引物"

<400> 215
ccatcccctt ctgtgaatgt 20

<210> 216
<211> 20
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成引物"

<400> 216
ggagattgga gacacggaga 20

<210> 217
<211> 20
<212> DNA
[0124] <213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成引物"

<400> 217
aagcaccgac tcggtgccac 20

<210> 218
<211> 20
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成引物"

<400> 218
tcacctccaa tgactagggg 20

<210> 219

<211> 22
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 219
caagttgata acggactagc ct 22

<210> 220
<211> 23
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 220
[0125] agtccgagca gaagaagaag ttt 23

<210> 221
<211> 25
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 221
tttcaagttg ataacggact agcct 25

<210> 222
<211> 20
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 222
aaacagcaga ttcgcctgga 20

<210> 223
<211> 20
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成引物"

<400> 223
tcatccgctc gatgaagctc 20

<210> 224
<211> 20
<212> DNA
<213> 人工序列

[0126]

<220>
<221> 来源
<223> /注释="人工序列的描述：合成引物"

<400> 224
tccaaaatca agtggggcga 20

<210> 225
<211> 20
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成引物"

<400> 225
tgatgaccct tttggctccc 20

<210> 226
<211> 45
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 226
gaggaattct ttttttgtty gaatatgttg gaggtttttt ggaag 45

<210> 227
<211> 42
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 227
gagaagctta aataaaaaac racaatactc aacccaacaa cc 42

[0127]
<210> 228
<211> 17
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 228
caggaaacag ctatgac 17

<210> 229
<211> 39
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 229
gcctctagag gtacctgagg gcctatttcc catgattcc 39

<210> 230
<211> 133
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<220>
<221> 修饰_碱基
<222> (92)..(111)
<223> a, c, t, g, 未知或其它

<400> 230
acctctagaa aaaaagcacc gactcggtgc cacttttca agttgataac ggactagcct 60

tattttaact tgctatttct agctctaaaa cnnnnnnnnn nnnnnnnnnn nggtgtttcg 120

tcctttccac aag 133

[0128]

<210> 231
<211> 133
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<220>
<221> 修饰_碱基
<222> (92)..(111)
<223> a, c, t, g, 未知或其它

<400> 231
acctctagaa aaaaagcacc gactcggtgc cacttttca agttgataac ggactagcct 60

tatattaact tgctatttct agctctaata cnnnnnnnnn nnnnnnnnnn nggtgtttcg 120

tcctttccac aag 133

<210> 232
<211> 153
<212> DNA

<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<220>
<221> 修饰_碱基
<222> (112)..(131)
<223> a, c, t, g, 未知或其它

<400> 232
acctctagaa aaaaagcacc gactcggtgc cacttttca agttgataac ggactagcct 60

tattttaact tgctatgctg ttttgtttcc aaaacagcat agctctaaaa cnnnnnnnnn 120

nnnnnnnnnn nggtgtttcg tcctttccac aag 153

<210> 233
<211> 153
<212> DNA
[0129] <213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<220>
<221> 修饰_碱基
<222> (112)..(131)
<223> a, c, t, g, 未知或其它

<400> 233
acctctagaa aaaaagcacc gactcggtgc cacttttca agttgataac ggactagcct 60

tatattaact tgctatgctg tattgtttcc aatacagcat agctctaata cnnnnnnnnn 120

nnnnnnnnnn nggtgtttcg tcctttccac aag 153

<210> 234
<211> 22
<212> DNA
<213> 人

<220>
<221> 修饰_碱基
<222> (20)..(20)
<223> a, c, t, g, 未知或其它

<400> 234
aggccccagt ggctgctctn aa 22

<210> 235
<211> 22
<212> DNA
<213> 人

<220>
<221> 修饰_碱基
<222> (20)..(20)
<223> a, c, t, g, 未知或其它

<400> 235
acatcaaccg gtggcgcatn at 22

[0130]

<210> 236
<211> 22
<212> DNA
<213> 人

<220>
<221> 修饰_碱基
<222> (20)..(20)
<223> a, c, t, g, 未知或其它

<400> 236
aaggtgtggt tccagaaccn ac 22

<210> 237
<211> 22
<212> DNA
<213> 人

<220>
<221> 修饰_碱基
<222> (20)..(20)
<223> a, c, t, g, 未知或其它

<400> 237

ccatcacatc aaccggtggn ag 22

<210> 238
<211> 22
<212> DNA
<213> 人

<220>
<221> 修饰_碱基
<222> (20)..(20)
<223> a, c, t, g, 未知或其它

<400> 238
aaacggcaga agctggaggn ta 22

<210> 239
<211> 22
<212> DNA
<213> 人

<220>
[0131] <221> 修饰_碱基
<222> (20)..(20)
<223> a, c, t, g, 未知或其它

<400> 239
ggcagaagct ggaggaggan tt 22

<210> 240
<211> 22
<212> DNA
<213> 人

<220>
<221> 修饰_碱基
<222> (20)..(20)
<223> a, c, t, g, 未知或其它

<400> 240
ggtgtggttc cagaaccggn tc 22

<210> 241
<211> 22
<212> DNA

<213> 人

<220>
<221> 修饰_碱基
<222> (20)..(20)
<223> a, c, t, g, 未知或其它

<400> 241
aaccggagga caaagtacan tg 22

<210> 242
<211> 22
<212> DNA
<213> 人

<220>
<221> 修饰_碱基
<222> (20)..(20)
<223> a, c, t, g, 未知或其它

<400> 242
ttccagaacc ggaggacaan ca 22

[0132]

<210> 243
<211> 22
<212> DNA
<213> 人

<220>
<221> 修饰_碱基
<222> (20)..(20)
<223> a, c, t, g, 未知或其它

<400> 243
gtgtggttcc agaaccggan ct 22

<210> 244
<211> 22
<212> DNA
<213> 人

<220>
<221> 修饰_碱基
<222> (20)..(20)

<223> a, c, t, g, 未知或其它

<400> 244
tccagaaccg gaggacaaan cc 22

<210> 245
<211> 22
<212> DNA
<213> 人

<220>
<221> 修饰_碱基
<222> (20)..(20)
<223> a, c, t, g, 未知或其它

<400> 245
cagaagctgg aggaggaagn cg 22

<210> 246
<211> 22
<212> DNA
[0133] <213> 人

<220>
<221> 修饰_碱基
<222> (20)..(20)
<223> a, c, t, g, 未知或其它

<400> 246
catcaaccgg tggcgcattn ga 22

<210> 247
<211> 22
<212> DNA
<213> 人

<220>
<221> 修饰_碱基
<222> (20)..(20)
<223> a, c, t, g, 未知或其它

<400> 247
gcagaagctg gaggaggaan gt 22

<210> 248
<211> 22
<212> DNA
<213> 人

<220>
<221> 修饰_碱基
<222> (20)..(20)
<223> a, c, t, g, 未知或其它

<400> 248
cctccctccc tggcccaggn gc 22

<210> 249
<211> 22
<212> DNA
<213> 人

<220>
<221> 修饰_碱基
<222> (20)..(20)
[0134] <223> a, c, t, g, 未知或其它

<400> 249
tcatctgtgc ccctccctcn aa 22

<210> 250
<211> 22
<212> DNA
<213> 人

<220>
<221> 修饰_碱基
<222> (20)..(20)
<223> a, c, t, g, 未知或其它

<400> 250
gggaggacat cgatgtcacn at 22

<210> 251
<211> 22
<212> DNA
<213> 人

<220>
<221> 修饰_碱基
<222> (20)..(20)
<223> a, c, t, g, 未知或其它

<400> 251
caaacggcag aagctggagn ac 22

<210> 252
<211> 22
<212> DNA
<213> 人

<220>
<221> 修饰_碱基
<222> (20)..(20)
<223> a, c, t, g, 未知或其它

<400> 252
gggtgggcaa ccacaaaccn ag 22

[0135]
<210> 253
<211> 22
<212> DNA
<213> 人

<220>
<221> 修饰_碱基
<222> (20)..(20)
<223> a, c, t, g, 未知或其它

<400> 253
ggtgggcaac cacaaacccn ta 22

<210> 254
<211> 22
<212> DNA
<213> 人

<220>
<221> 修饰_碱基
<222> (20)..(20)
<223> a, c, t, g, 未知或其它

<400> 254

ggctcccatc acatcaaccn tt 22

<210> 255
<211> 22
<212> DNA
<213> 人

<220>
<221> 修饰_碱基
<222> (20)..(20)
<223> a, c, t, g, 未知或其它

<400> 255
gaagggcctg agtccgagcn tc 22

<210> 256
<211> 22
<212> DNA
<213> 人

<220>
[0136] <221> 修饰_碱基
<222> (20)..(20)
<223> a, c, t, g, 未知或其它

<400> 256
caaccggtgg cgcattgccn tg 22

<210> 257
<211> 22
<212> DNA
<213> 人

<220>
<221> 修饰_碱基
<222> (20)..(20)
<223> a, c, t, g, 未知或其它

<400> 257
aggaggaagg gcctgagtcn ca 22

<210> 258
<211> 22
<212> DNA

<213> 人

<220>
<221> 修饰_碱基
<222> (20)..(20)
<223> a, c, t, g, 未知或其它

<400> 258
agctggagga ggaagggccn ct 22

<210> 259
<211> 22
<212> DNA
<213> 人

<220>
<221> 修饰_碱基
<222> (20)..(20)
<223> a, c, t, g, 未知或其它

<400> 259
gcattgccac gaagcaggcn cc 22

[0137]

<210> 260
<211> 22
<212> DNA
<213> 人

<220>
<221> 修饰_碱基
<222> (20)..(20)
<223> a, c, t, g, 未知或其它

<400> 260
attgccacga agcaggccan cg 22

<210> 261
<211> 22
<212> DNA
<213> 人

<220>
<221> 修饰_碱基
<222> (20)..(20)

<223> a, c, t, g, 未知或其它

<400> 261
agaaccggag gacaaagtan ga 22

<210> 262
<211> 22
<212> DNA
<213> 人

<220>
<221> 修饰_碱基
<222> (20)..(20)
<223> a, c, t, g, 未知或其它

<400> 262
tcaaccggtg gcgcattgcn gt 22

<210> 263
<211> 22
<212> DNA
<213> 人

[0138]

<220>
<221> 修饰_碱基
<222> (20)..(20)
<223> a, c, t, g, 未知或其它

<400> 263
gaagctggag gaggaagggn gc 22

<210> 264
<211> 123
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<400> 264
ccaatgggga ggacatcgat gtcacctcca atgactaggg tgggcaacca caaacccacg 60

agggcagagt gctgcttgct gctggccagg cccctgcgtg ggcccaagct ggactctggc 120

cac 123

<210> 265
<211> 121
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<400> 265
cgagcagaag aagaagggct cccatcacat caaccggtgg cgcattgcca cgaagcaggc 60

caatggggag gacatcgatg tcacctccaa tgactagggt gggcaaccac aaacccacga 120

g 121

<210> 266
<211> 128
<212> DNA
<213> 人工序列

[0139]

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<400> 266
ggaggacaaa gtacaaacgg cagaagctgg aggaggaagg gcctgagtcc gagcagaaga 60

agaagggctc ccatcacatc aaccggtggc gcattgccac gaagcaggcc aatggggagg 120

acatcgat 128

<210> 267
<211> 130
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<400> 267
agaagctgga ggaggaaggg cctgagtccg agcagaagaa gaagggctcc catcacatca 60

accggtggcg cattgccacg aagcaggcca atggggagga catcgatgtc acctccaatg 120

actagggtgg 130

<210> 268
<211> 125
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<400> 268
cctcagtctt cccatcaggc tctcagctca gcctgagtgt tgaggcccca gtggctgctc 60

tgggggcctc ctgagtttct catctgtgcc cctccctccc tggcccaggt gaaggtgtgg 120

ttcca 125

[0140]

<210> 269
<211> 129
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<400> 269
tcatctgtgc ccctccctcc ctggcccagg tgaaggtgtg gttccagaac cggaggacaa 60

agtacaaacg gcagaagctg gaggaggaag ggcctgagtc cgagcagaag aagaagggct 120

cccatcaca 129

<210> 270
<211> 129
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<400> 270
ctccaatgac tagggtgggc aaccacaaac ccacgagggc agagtgctgc ttgctgctgg 60

ccaggcccct gcgtgggccc aagctggact ctggccactc cctggccagg ctttggggag 120

gcctggagt 129

<210> 271
<211> 127
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<400> 271
ctgcttgctg ctggccaggc ccctgcgtgg gcccaagctg gactctggcc actccctggc 60

caggctttgg ggaggcctgg agtcatggcc ccacagggct tgaagcccgg ggccgccatt 120

gacagag 127

[0141]

<210> 272
<211> 25
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 272
gaaattaata cgactcacta taggg 25

<210> 273
<211> 126
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<400> 273
aaaaaagcac cgactcggtg ccacttttc aagttgataa cggactagcc ttattttaac 60

ttgctatttc tagctctaaa acaacgacga gcgtgacacc accctatagt gagtcgtatt 120

aatttc 126

<210> 274
<211> 126
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<400> 274
aaaaaagcac cgactcggtg ccacttttc aagttgataa cggactagcc ttattttaac 60

ttgctatttc tagctctaaa acgcaacaat taatagactg gacctatagt gagtcgtatt 120

aatttc 126

[0142]

<210> 275
<211> 4677
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<400> 275
tctttcttgc gctatgacac ttccagcaaa aggtagggcg ggctgcgaga cggcttcccg 60

gcgctgcatg caacaccgat gatgcttcga ccccccgaag ctccttcggg gctgcatggg 120

cgctccgatg ccgctccagg gcgagcgctg tttaaatagc caggcccccg attgcaaaga 180

cattatagcg agctaccaaa gccatattca aacacctaga tcactaccac ttctacacag 240

gccactcgag cttgtgatcg cactccgcta agggggcgcc tcttcctctt cgtttcagtc 300

acaacccgca aacatgtacc catacgatgt tccagattac gcttcgccga agaaaaagcg 360

caaggtcgaa gcgtccgaca agaagtacag catcggcctg gacatcggca ccaactctgt 420

gggctgggcc gtgatcaccg acgagtacaa ggtgcccagc aagaaattca aggtgctggg 480

caacaccgac cggcacagca tcaagaagaa cctgatcgga gccctgctgt tcgacagcgg 540

cgaaacagcc gaggccaccc ggctgaagag aaccgccaga agaagataca ccagacggaa 600

gaaccggatc tgctatctgc aagagatctt cagcaacgag atggccaagg tggacgacag 660

cttcttccac agactggaag agtccttcct ggtggaagag gataagaagc acgagcggca 720

ccccatcttc ggcaacatcg tggacgaggt ggcctaccac gagaagtacc ccaccatcta 780

ccacctgaga aagaaactgg tggacagcac cgacaaggcc gacctgcggc tgatctatct 840

ggccctggcc cacatgatca agttccgggg ccacttcctg atcgagggcg acctgaaccc 900

cgacaacagc gacgtggaca agctgttcat ccagctggtg cagacctaca accagctgtt 960

cgaggaaaac cccatcaacg ccagcggcgt ggacgccaag gccatcctgt ctgccagact 1020

gagcaagagc agacggctgg aaaatctgat cgcccagctg cccggcgaga agaagaatgg 1080

cctgttcggc aacctgattg ccctgagcct gggcctgacc cccaacttca agagcaactt 1140

cgacctggcc gaggatgcca aactgcagct gagcaaggac acctacgacg acgacctgga 1200

[0143]

caacctgctg gcccagatcg gcgaccagta cgccgacctg tttctggccg ccaagaacct 1260

gtccgacgcc atcctgctga gcgacatcct gagagtgaac accgagatca ccaaggcccc 1320

cctgagcgcc tctatgatca agagatacga cgagcaccac caggacctga ccctgctgaa 1380

agctctcgtg cggcagcagc tgcctgagaa gtacaaagag attttcttcg accagagcaa 1440

gaacggctac gccggctaca ttgacggcgg agccagccag gaagagttct acaagttcat 1500

caagcccatc ctggaaaaga tggacggcac cgaggaactg ctcgtgaagc tgaacagaga 1560

ggacctgctg cggaagcagc ggaccttcga caacggcagc atccccccacc agatccacct 1620

gggagagctg cacgccattc tgcggcggca ggaagatttt tacccattcc tgaaggacaa 1680

ccgggaaaag atcgagaaga tcctgacctt ccgcatcccc tactacgtgg gccctctggc 1740

caggggaaac agcagattcg cctggatgac cagaaagagc gaggaaacca tcaccccctg 1800

gaacttcgag gaagtggtgg acaagggcgc ttccgcccag agcttcatcg agcggatgac 1860

caacttcgat aagaacctgc ccaacgagaa ggtgctgccc aagcacagcc tgctgtacga 1920

gtacttcacc gtgtataacg agctgaccaa agtgaaatac gtgaccgagg gaatgagaaa 1980

gcccgccttc ctgagcggcg agcagaaaaa ggccatcgtg gacctgctgt tcaagaccaa 2040

ccggaaagtg accgtgaagc agctgaaaga ggactacttc aagaaaatcg agtgcttcga 2100

ctccgtggaa atctccggcg tggaagatcg gttcaacgcc tccctgggca cataccacga 2160

tctgctgaaa attatcaagg acaaggactt cctggacaat gaggaaaacg aggacattct 2220

ggaagatatc gtgctgaccc tgacactgtt tgaggacaga gagatgatcg aggaacggct 2280

gaaaacctat gcccacctgt tcgacgacaa agtgatgaag cagctgaagc ggcggagata 2340

caccggctgg ggcaggctga gccggaagct gatcaacggc atccgggaca agcagtccgg 2400

caagacaatc ctggatttcc tgaagtccga cggcttcgcc aacagaaact tcatgcagct 2460

gatccacgac gacagcctga cctttaaaga ggacatccag aaagcccagg tgtccggcca 2520

gggcgatagc ctgcacgagc acattgccaa tctggccggc agccccgcca ttaagaaggg 2580

catcctgcag acagtgaagg tggtggacga gctcgtgaaa gtgatgggcc ggcacaagcc 2640

cgagaacatc gtgatcgaaa tggccagaga gaaccagacc acccagaagg gacagaagaa 2700

[0144] cagccgcgag agaatgaagc ggatcgaaga gggcatcaaa gagctgggca gccagatcct 2760

gaaagaacac cccgtggaaa acacccagct gcagaacgag aagctgtacc tgtactacct 2820

gcagaatggg cgggatatgt acgtggacca ggaactggac atcaaccggc tgtccgacta 2880

cgatgtggac catatcgtgc ctcagagctt tctgaaggac gactccatcg acaacaaggt 2940

gctgaccaga agcgacaaga accggggcaa gagcgacaac gtgccctccg aagaggtcgt 3000

gaagaagatg aagaactact ggcggcagct gctgaacgcc aagctgatta cccagagaaa 3060

gttcgacaat ctgaccaagg ccgagagagg cggcctgagc gaactggata aggccggctt 3120

catcaagaga cagctggtgg aaacccggca gatcacaaag cacgtggcac agatcctgga 3180

ctcccggatg aacactaagt acgacgagaa tgacaagctg atccgggaag tgaaagtgat 3240

caccctgaag tccaagctgg tgtccgattt ccggaaggat ttccagtttt acaaagtgcg 3300

cgagatcaac aactaccacc acgcccacga cgcctacctg aacgccgtcg tgggaaccgc 3360

cctgatcaaa aagtacccta agctggaaag cgagttcgtg tacggcgact acaaggtgta 3420

cgacgtgcgg aagatgatcg ccaagagcga gcaggaaatc ggcaaggcta ccgccaagta 3480

cttcttctac agcaacatca tgaacttttt caagaccgag attaccctgg ccaacggcga 3540

gatccggaag cggcctctga tcgagacaaa cggcgaaacc ggggagatcg tgtgggataa 3600

gggccgggat tttgccaccg tgcggaaagt gctgagcatg ccccaagtga atatcgtgaa 3660

aaagaccgag gtgcagacag gcggcttcag caaagagtct atcctgccca agaggaacag 3720

cgataagctg atcgccagaa agaaggactg ggaccctaag aagtacggcg gcttcgacag 3780

ccccaccgtg gcctattctg tgctggtggt ggccaaagtg gaaaagggca agtccaagaa 3840

actgaagagt gtgaaagagc tgctggggat caccatcatg gaaagaagca gcttcgagaa 3900

gaatcccatc gactttctgg aagccaaggg ctacaaagaa gtgaaaaagg acctgatcat 3960

caagctgcct aagtactccc tgttcgagct ggaaaacggc cggaagagaa tgctggcctc 4020

tgccggcgaa ctgcagaagg gaaacgaact ggccctgccc tccaaatatg tgaacttcct 4080

gtacctggcc agccactatg agaagctgaa gggctccccc gaggataatg agcagaaaca 4140

gctgtttgtg gaacagcaca agcactacct ggacgagatc atcgagcaga tcagcgagtt 4200

[0145] ctccaagaga gtgatcctgg ccgacgctaa tctggacaaa gtgctgtccg cctacaacaa 4260

gcaccgggat aagcccatca gagagcaggc cgagaatatc atccacctgt ttaccctgac 4320

caatctggga gcccctgccg ccttcaagta ctttgacacc accatcgacc ggaagaggta 4380

caccagcacc aaagaggtgc tggacgccac cctgatccac cagagcatca ccggcctgta 4440

cgagacacgg atcgacctgt ctcagctggg aggcgacagc cccaagaaga agagaaaggt 4500

ggaggccagc taaggatccg gcaagactgg ccccgcttgg caacgcaaca gtgagcccct 4560

ccctagtgtg tttggggatg tgactatgta ttcgtgtgtt ggccaacggg tcaacccgaa 4620

cagattgata cccgccttgg catttcctgt cagaatgtaa cgtcagttga tggtact 4677

<210> 276
<211> 3150
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<400> 276

tctttcttgc gctatgacac ttccagcaaa aggtagggcg ggctgcgaga cggcttcccg 60

gcgctgcatg caacaccgat gatgcttcga ccccccgaag ctccttcggg gctgcatggg 120

cgctccgatg ccgctccagg gcgagcgctg tttaaatagc caggcccccg attgcaaaga 180

cattatagcg agctaccaaa gccatattca aacacctaga tcactaccac ttctacacag 240

gccactcgag cttgtgatcg cactccgcta agggggcgcc tcttcctctt cgtttcagtc 300

acaacccgca aacatgccta agaagaagag gaaggttaac acgattaaca tcgctaagaa 360

cgacttctct gacatcgaac tggctgctat cccgttcaac actctggctg accattacgg 420

tgagcgttta gctcgcgaac agttggccct tgagcatgag tcttacgaga tgggtgaagc 480

acgcttccgc aagatgtttg agcgtcaact taaagctggt gaggttgcgg ataacgctgc 540

cgccaagcct ctcatcacta ccctactccc taagatgatt gcacgcatca acgactggtt 600

tgaggaagtg aaagctaagc gcggcaagcg cccgacagcc ttccagttcc tgcaagaaat 660

caagccggaa gccgtagcgt acatcaccat taagaccact ctggcttgcc taaccagtgc 720

[0146] tgacaataca accgttcagg ctgtagcaag cgcaatcggt cgggccattg aggacgaggc 780

tcgcttcggt cgtatccgtg accttgaagc taagcacttc aagaaaaacg ttgaggaaca 840

actcaacaag cgcgtagggc acgtctacaa gaaagcattt atgcaagttg tcgaggctga 900

catgctctct aagggtctac tcggtggcga ggcgtggtct tcgtggcata aggaagactc 960

tattcatgta ggagtacgct gcatcgagat gctcattgag tcaaccggaa tggttagctt 1020

acaccgccaa aatgctggcg tagtaggtca agactctgag actatcgaac tcgcacctga 1080

atacgctgag gctatcgcaa cccgtgcagg tgcgctggct ggcatctctc cgatgttcca 1140

accttgcgta gttcctccta agccgtggac tggcattact ggtggtggct attgggctaa 1200

cggtcgtcgt cctctggcgc tggtgcgtac tcacagtaag aaagcactga tgcgctacga 1260

agacgtttac atgcctgagg tgtacaaagc gattaacatt gcgcaaaaca ccgcatggaa 1320

aatcaacaag aaagtcctag cggtcgccaa cgtaatcacc aagtggaagc attgtccggt 1380

cgaggacatc cctgcgattg agcgtgaaga actcccgatg aaaccggaag acatcgacat 1440

gaatcctgag gctctcaccg cgtggaaacg tgctgccgct gctgtgtacc gcaaggacaa 1500

ggctcgcaag tctcgccgta tcagccttga gttcatgctt gagcaagcca ataagtttgc 1560

taaccataag gccatctggt tcccttacaa catggactgg cgcggtcgtg tttacgctgt 1620

gtcaatgttc aacccgcaag gtaacgatat gaccaaagga ctgcttacgc tggcgaaagg 1680

taaaccaatc ggtaaggaag gttactactg gctgaaaatc cacggtgcaa actgtgcggg 1740

tgtcgacaag gttccgttcc ctgagcgcat caagttcatt gaggaaaacc acgagaacat 1800

catggcttgc gctaagtctc cactggagaa cacttggtgg gctgagcaag attctccgtt 1860

ctgcttcctt gcgttctgct ttgagtacgc tggggtacag caccacggcc tgagctataa 1920

ctgctccctt ccgctggcgt ttgacgggtc ttgctctggc atccagcact tctccgcgat 1980

gctccgagat gaggtaggtg gtcgcgcggt taacttgctt cctagtgaaa ccgttcagga 2040

catctacggg attgttgcta agaaagtcaa cgagattcta caagcagacg caatcaatgg 2100

gaccgataac gaagtagtta ccgtgaccga tgagaacact ggtgaaatct ctgagaaagt 2160

caagctgggc actaaggcac tggctggtca atggctggct tacggtgtta ctcgcagtgt 2220

gactaagcgt tcagtcatga cgctggctta cgggtccaaa gagttcggct tccgtcaaca 2280

[0147]

agtgctggaa gataccattc agccagctat tgattccggc aagggtctga tgttcactca 2340

gccgaatcag gctgctggat acatggctaa gctgatttgg gaatctgtga gcgtgacggt 2400

ggtagctgcg gttgaagcaa tgaactggct taagtctgct gctaagctgc tggctgctga 2460

ggtcaaagat aagaagactg gagagattct tcgcaagcgt tgcgctgtgc attgggtaac 2520

tcctgatggt ttccctgtgt ggcaggaata caagaagcct attcagacgc gcttgaacct 2580

gatgttcctc ggtcagttcc gcttacagcc taccattaac accaacaaag atagcgagat 2640

tgatgcacac aaacaggagt ctggtatcgc tcctaacttt gtacacagcc aagacggtag 2700

ccaccttcgt aagactgtag tgtgggcaca cgagaagtac ggaatcgaat cttttgcact 2760

gattcacgac tccttcggta cgattccggc tgacgctgcg aacctgttca aagcagtgcg 2820

cgaaactatg gttgacacat atgagtcttg tgatgtactg gctgatttct acgaccagtt 2880

cgctgaccag ttgcacgagt ctcaattgga caaaatgcca gcacttccgg ctaaaggtaa 2940

cttgaacctc cgtgacatct tagagtcgga cttcgcgttc gcgtaaggat ccggcaagac 3000

tggccccgct tggcaacgca acagtgagcc cctccctagt gtgtttgggg atgtgactat 3060

gtattcgtgt gttggccaac gggtcaaccc gaacagattg atacccgcct tggcatttcc 3120

tgtcagaatg taacgtcagt tgatggtact 3150

<210> 277
<211> 125
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<220>
<221> 修饰_碱基
<222> (23)..(42)
<223> a, c, t, g, 未知或其它

<400> 277
gaaattaata cgactcacta tannnnnnnn nnnnnnnnnn nngttttaga gctagaaata 60

[0148] gcaagttaaa ataaggctag tccgttatca acttgaaaaa gtggcaccga gtcggtgctt 120

ttttt 125

<210> 278
<211> 8452
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<400> 278
tgcggtattt cacaccgcat caggtggcac ttttcgggga aatgtgcgcg gaacccctat 60

ttgtttattt ttctaaatac attcaaatat gtatccgctc atgagattat caaaaaggat 120

cttcacctag atccttttaa attaaaaatg aagttttaaa tcaatctaaa gtatatatga 180

gtaaacttgg tctgacagtt accaatgctt aatcagtgag gcacctatct cagcgatctg 240

tctatttcgt tcatccatag ttgcctgact ccccgtcgtg tagataacta cgatacggga 300

gggcttacca tctggcccca gtgctgcaat gataccgcga gacccacgct caccggctcc 360

agatttatca gcaataaacc agccagccgg aagggccgag cgcagaagtg gtcctgcaac 420

tttatccgcc tccatccagt ctattaattg ttgccgggaa gctagagtaa gtagttcgcc 480

agttaatagt ttgcgcaacg ttgttgccat tgctacaggc atcgtggtgt cacgctcgtc 540

gtttggtatg gcttcattca gctccggttc ccaacgatca aggcgagtta catgatcccc 600

catgttgtgc aaaaaagcgg ttagctcctt cggtcctccg atcgttgtca gaagtaagtt 660

ggccgcagtg ttatcactca tggttatggc agcactgcat aattctctta ctgtcatgcc 720

atccgtaaga tgcttttctg tgactggtga gtactcaacc aagtcattct gagaatagtg 780

tatgcggcga ccgagttgct cttgcccggc gtcaatacgg gataataccg cgccacatag 840

cagaacttta aaagtgctca tcattggaaa acgttcttcg gggcgaaaac tctcaaggat 900

cttaccgctg ttgagatcca gttcgatgta acccactcgt gcacccaact gatcttcagc 960

atcttttact ttcaccagcg tttctgggtg agcaaaaaca ggaaggcaaa atgccgcaaa 1020

aaagggaata agggcgacac ggaaatgttg aatactcata ctcttccttt ttcaatatta 1080

[0149]

ttgaagcatt tatcagggtt attgtctcat gaccaaaatc ccttaacgtg agttttcgtt 1140

ccactgagcg tcagaccccg tagaaaagat caaaggatct tcttgagatc ctttttttct 1200

gcgcgtaatc tgctgcttgc aaacaaaaaa accaccgcta ccagcggtgg tttgtttgcc 1260

ggatcaagag ctaccaactc tttttccgaa ggtaactggc ttcagcagag cgcagatacc 1320

aaatactgtt cttctagtgt agccgtagtt aggccaccac ttcaagaact ctgtagcacc 1380

gcctacatac ctcgctctgc taatcctgtt accagtggct gttgccagtg gcgataagtc 1440

gtgtcttacc gggttggact caagacgata gttaccggat aaggcgcagc ggtcgggctg 1500

aacgggggggt tcgtgcacac agcccagctt ggagcgaacg acctacaccg aactgagata 1560

cctacagcgt gagctatgag aaagcgccac gcttcccgaa gggagaaagg cggacaggta 1620

tccggtaagc ggcagggtcg gaacaggaga gcgcacgagg gagcttccag ggggaaacgc 1680

ctggtatctt tatagtcctg tcgggtttcg ccacctctga cttgagcgtc gatttttgtg 1740

atgctcgtca gggggggcgga gcctatggaa aaacgccagc aacgcggcct ttttacggtt 1800

cctggccttt tgctggcctt ttgctcacat gttctttcct gcgttatccc ctgattctgt 1860

ggataaccgt attaccgcct ttgagtgagc tgataccgct cgccgcagcc gaacgaccga 1920

gcgcagcgag tcagtgagcg aggaagcggt cgctgaggct tgacatgatt ggtgcgtatg 1980

tttgtatgaa gctacaggac tgatttggcg ggctatgagg gcgggggaag ctctggaagg 2040

gccgcgatgg ggcgcgcggc gtccagaagg cgccatacgg cccgctggcg gcacccatcc 2100

ggtataaaag cccgcgaccc cgaacggtga cctccacttt cagcgacaaa cgagcactta 2160

tacatacgcg actattctgc cgctatacat aaccactcag ctagcttaag atcccatcaa 2220

gcttgcatgc cgggcgcgcc agaaggagcg cagccaaacc aggatgatgt ttgatggggt 2280

atttgagcac ttgcaaccct tatccggaag ccccctggcc cacaaaggct aggcgccaat 2340

gcaagcagtt cgcatgcagc ccctggagcg gtgccctcct gataaaccgg ccaggggccc 2400

tatgttcttt actttttac aagagaagtc actcaacatc ttaaaatggc caggtgagtc 2460

gacgagcaag cccggcggat caggcagcgt gcttgcagat ttgacttgca acgcccgcat 2520

tgtgtcgacg aaggcttttg gctcctctgt cgctgtctca agcagcatct aaccctgcgt 2580

[0150] cgccgtttcc atttgcagga gattcgaggt accatgtacc catacgatgt tccagattac 2640

gcttcgccga agaaaaagcg caaggtcgaa gcgtccgaca agaagtacag catcggcctg 2700

gacatcggca ccaactctgt gggctgggcc gtgatcaccg acgagtacaa ggtgcccagc 2760

aagaaattca aggtgctggg caacaccgac cggcacagca tcaagaagaa cctgatcgga 2820

gccctgctgt tcgacagcgg cgaaacagcc gaggccaccc ggctgaagag aaccgccaga 2880

agaagataca ccagacggaa gaaccggatc tgctatctgc aagagatctt cagcaacgag 2940

atggccaagg tggacgacag cttcttccac agactggaag agtccttcct ggtggaagag 3000

gataagaagc acgagcggca ccccatcttc ggcaacatcg tggacgaggt ggcctaccac 3060

gagaagtacc ccaccatcta ccacctgaga aagaaactgg tggacagcac cgacaaggcc 3120

gacctgcggc tgatctatct ggccctggcc cacatgatca agttccgggg ccacttcctg 3180

atcgagggcg acctgaaccc cgacaacagc gacgtggaca agctgttcat ccagctggtg 3240

cagacctaca accagctgtt cgaggaaaac cccatcaacg ccagcggcgt ggacgccaag 3300

gccatcctgt ctgccagact gagcaagagc agacggctgg aaaatctgat cgcccagctg 3360

cccggcgaga agaagaatgg cctgttcggc aacctgattg ccctgagcct gggcctgacc 3420

cccaacttca agagcaactt cgacctggcc gaggatgcca aactgcagct gagcaaggac 3480

acctacgacg acgacctgga caacctgctg gcccagatcg gcgaccagta cgccgacctg 3540

tttctggccg ccaagaacct gtccgacgcc atcctgctga gcgacatcct gagagtgaac 3600

accgagatca ccaaggcccc cctgagcgcc tctatgatca agagatacga cgagcaccac 3660

caggacctga ccctgctgaa agctctcgtg cggcagcagc tgcctgagaa gtacaaagag 3720

attttcttcg accagagcaa gaacggctac gccggctaca ttgacggcgg agccagccag 3780

gaagagttct acaagttcat caagcccatc ctggaaaaga tggacggcac cgaggaactg 3840

ctcgtgaagc tgaacagaga ggacctgctg cggaagcagc ggaccttcga caacggcagc 3900

atcccccacc agatccacct gggagagctg cacgccattc tgcggcggca ggaagatttt 3960

tacccattcc tgaaggacaa ccgggaaaag atcgagaaga tcctgacctt ccgcatcccc 4020

tactacgtgg gccctctggc caggggaaac agcagattcg cctggatgac cagaaagagc 4080

gaggaaacca tcacccccctg gaacttcgag gaagtggtgg acaagggcgc ttccgcccag 4140

[0151]

agcttcatcg agcggatgac caacttcgat aagaacctgc ccaacgagaa ggtgctgccc 4200

aagcacagcc tgctgtacga gtacttcacc gtgtataacg agctgaccaa agtgaaatac 4260

gtgaccgagg gaatgagaaa gcccgccttc ctgagcggcg agcagaaaaa ggccatcgtg 4320

gacctgctgt tcaagaccaa ccggaaagtg accgtgaagc agctgaaaga ggactacttc 4380

aagaaaatcg agtgcttcga ctccgtggaa atctccggcg tggaagatcg gttcaacgcc 4440

tccctgggca cataccacga tctgctgaaa attatcaagg acaaggactt cctggacaat 4500

gaggaaaacg aggacattct ggaagatatc gtgctgaccc tgacactgtt tgaggacaga 4560

gagatgatcg aggaacggct gaaaacctat gcccacctgt tcgacgacaa agtgatgaag 4620

cagctgaagc ggcggagata caccggctgg ggcaggctga gccggaagct gatcaacggc 4680

atccgggaca agcagtccgg caagacaatc ctggatttcc tgaagtccga cggcttcgcc 4740

aacagaaact tcatgcagct gatccacgac gacagcctga cctttaaaga ggacatccag 4800

aaagcccagg tgtccggcca gggcgatagc ctgcacgagc acattgccaa tctggccggc 4860

agccccgcca ttaagaaggg catcctgcag acagtgaagg tggtggacga gctcgtgaaa 4920

gtgatgggcc ggcacaagcc cgagaacatc gtgatcgaaa tggccagaga gaaccagacc 4980

acccagaagg gacagaagaa cagccgcgag agaatgaagc ggatcgaaga gggcatcaaa 5040

gagctgggca gccagatcct gaaagaacac cccgtggaaa acacccagct gcagaacgag 5100

aagctgtacc tgtactacct gcagaatggg cgggatatgt acgtggacca ggaactggac 5160

atcaaccggc tgtccgacta cgatgtggac catatcgtgc ctcagagctt tctgaaggac 5220

gactccatcg acaacaaggt gctgaccaga agcgacaaga accggggcaa gagcgacaac 5280

gtgccctccg aagaggtcgt gaagaagatg aagaactact ggcggcagct gctgaacgcc 5340

aagctgatta cccagagaaa gttcgacaat ctgaccaagg ccgagagagg cggcctgagc 5400

gaactggata aggccggctt catcaagaga cagctggtgg aaacccggca gatcacaaag 5460

cacgtggcac agatcctgga ctcccggatg aacactaagt acgacgagaa tgacaagctg 5520

atccgggaag tgaaagtgat caccctgaag tccaagctgg tgtccgattt ccggaaggat 5580

ttccagtttt acaaagtgcg cgagatcaac aactaccacc acgcccacga cgcctacctg 5640

[0152] aacgccgtcg tgggaaccgc cctgatcaaa aagtacccta agctggaaag cgagttcgtg 5700

tacggcgact acaaggtgta cgacgtgcgg aagatgatcg ccaagagcga gcaggaaatc 5760

ggcaaggcta ccgccaagta cttcttctac agcaacatca tgaacttttt caagaccgag 5820

attaccctgg ccaacggcga gatccggaag cggcctctga tcgagacaaa cggcgaaacc 5880

ggggagatcg tgtgggataa gggccgggat tttgccaccg tgcggaaagt gctgagcatg 5940

ccccaagtga atatcgtgaa aaagaccgag gtgcagacag gcggcttcag caaagagtct 6000

atcctgccca agaggaacag cgataagctg atcgccagaa agaaggactg ggaccctaag 6060

aagtacggcg gcttcgacag ccccaccgtg gcctattctg tgctggtggt ggccaaagtg 6120

gaaaagggca agtccaagaa actgaagagt gtgaaagagc tgctggggat caccatcatg 6180

gaaagaagca gcttcgagaa gaatcccatc gactttctgg aagccaaggg ctacaaagaa 6240

gtgaaaaagg acctgatcat caagctgcct aagtactccc tgttcgagct ggaaaacggc 6300

cggaagagaa tgctggcctc tgccggcgaa ctgcagaagg gaaacgaact ggccctgccc 6360

tccaaatatg tgaacttcct gtacctggcc agccactatg agaagctgaa gggctccccc 6420

gaggataatg agcagaaaca gctgtttgtg gaacagcaca agcactacct ggacgagatc 6480

atcgagcaga tcagcgagtt ctccaagaga gtgatcctgg ccgacgctaa tctggacaaa 6540

gtgctgtccg cctacaacaa gcaccgggat aagcccatca gagagcaggc cgagaatatc 6600

atccacctgt ttaccctgac caatctggga gcccctgccg ccttcaagta ctttgacacc 6660

accatcgacc ggaagaggta caccagcacc aaagaggtgc tggacgccac cctgatccac 6720

cagagcatca ccggcctgta cgagacacgg atcgacctgt ctcagctggg aggcgacagc 6780

cccaagaaga agagaaaggt ggaggccagc taacatatga ttcgaatgtc tttcttgcgc 6840

tatgacactt ccagcaaaag gtagggcggg ctgcgagacg gcttcccggc gctgcatgca 6900

acaccgatga tgcttcgacc ccccgaagct ccttcggggc tgcatgggcg ctccgatgcc 6960

gctccagggc gagcgctgtt taaatagcca ggcccccgat tgcaaagaca ttatagcgag 7020

ctaccaaagc catattcaaa cacctagatc actaccactt ctacacaggc cactcgagct 7080

tgtgatcgca ctccgctaag ggggcgcctc ttcctcttcg tttcagtcac aacccgcaaa 7140

catgacacaa gaatccctgt tacttctcga ccgtattgat tcggatgatt cctacgcgag 7200

[0153]

cctgcggaac gaccaggaat tctgggaggt gagtcgacga gcaagcccgg cggatcaggc 7260

agcgtgcttg cagatttgac ttgcaacgcc cgcattgtgt cgacgaaggc ttttggctcc 7320

tctgtcgctg tctcaagcag catctaaccc tgcgtcgccg tttccatttg cagccgctgg 7380

cccgccgagc cctggaggag ctcgggctgc cggtgccgcc ggtgctgcgg gtgcccggcg 7440

agagcaccaa ccccgtactg gtcggcgagc ccggcccggt gatcaagctg ttcggcgagc 7500

actggtgcgg tccggagagc ctcgcgtcgg agtcggaggc gtacgcggtc ctggcggacg 7560

ccccggtgcc ggtgccccgc ctcctcggcc gcggcgagct gcggcccggc accggagcct 7620

ggccgtggcc ctacctggtg atgagccgga tgaccggcac cacctggcgg tccgcgatgg 7680

acggcacgac cgaccggaac gcgctgctcg ccctggcccg cgaactcggc cgggtgctcg 7740

gccggctgca cagggtgccg ctgaccggga acaccgtgct caccccccat tccgaggtct 7800

tcccggaact gctgcgggaa cgccgcgcgg cgaccgtcga ggaccaccgc gggtggggct 7860

acctctcgcc ccggctgctg gaccgcctgg aggactggct gccggacgtg gacacgctgc 7920

tggccggccg cgaaccccgg ttcgtccacg gcgacctgca cgggaccaac atcttcgtgg 7980

acctggccgc gaccgaggtc accgggatcg tcgacttcac cgacgtctat gcgggagact 8040

cccgctacag cctggtgcaa ctgcatctca acgccttccg gggcgaccgc gagatcctgg 8100

ccgcgctgct cgacggggcg cagtggaagc ggaccgagga cttcgcccgc gaactgctcg 8160

ccttcacctt cctgcacgac ttcgaggtgt tcgaggagac cccgctggat ctctccggct 8220

tcaccgatcc ggaggaactg gcgcagttcc tctgggggcc gccggacacc gcccccggcg 8280

cctgataagg atccggcaag actggccccg cttggcaacg caacagtgag cccctcccta 8340

gtgtgtttgg ggatgtgact atgtattcgt gtgttggcca acgggtcaac ccgaacagat 8400

tgatacccgc cttggcattt cctgtcagaa tgtaacgtca gttgatggta ct 8452

<210> 279
<211> 102
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成多核苷酸"

[0154]

<400> 279
gttttagagc tatgctgttt tgaatggtcc caaaacggaa gggcctgagt ccgagcagaa 60

gaagaagttt tagagctatg ctgttttgaa tggtcccaaa ac 102

<210> 280
<211> 100
<212> DNA
<213> 人

<400> 280
cggaggacaa agtacaaacg gcagaagctg gaggaggaag ggcctgagtc cgagcagaag 60

aagaagggct cccatcacat caaccggtgg cgcattgcca 100

<210> 281
<211> 50
<212> DNA
<213> 人

<400> 281

agctggagga ggaagggcct gagtccgagc agaagaagaa gggctcccac 50

<210> 282
<211> 30
<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成寡核苷酸"

<400> 282
gaguccgagc agaagaagaa guuuuagagc 30

<210> 283
<211> 49
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成寡核苷酸"

[0155]

<400> 283
agctggagga ggaagggcct gagtccgagc agaagagaag ggctcccat 49

<210> 284
<211> 53
<212> DNA
<213> 人

<400> 284
ctggaggagg aagggcctga gtccgagcag aagaagaagg gctcccatca cat 53

<210> 285
<211> 52
<212> DNA
<213> 人

<400> 285
ctggaggagg aagggcctga gtccgagcag aagagaaggg ctcccatcac at 52

<210> 286
<211> 54

<212> DNA
<213> 人

<400> 286
ctggaggagg aagggcctga gtccgagcag aagaaagaag ggctcccatc acat 54

<210> 287
<211> 50
<212> DNA
<213> 人

<400> 287
ctggaggagg aagggcctga gtccgagcag aagaagggct cccatcacat 50

<210> 288
<211> 47
<212> DNA
<213> 人

<400> 288
ctggaggagg aagggcctga gcccgagcag aagggctccc atcacat 47

[0156]

<210> 289
<211> 66
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<220>
<221> 来源
<223> /注释="组合 DNA/RNA 分子的描述：合成
寡核苷酸"

<220>
<221> 修饰_碱基
<222> (1)..(20)
<223> a, c, t, g, 未知或其它

<400> 289
nnnnnnnnnn nnnnnnnnnn guuuuagagc uagaaauagc aaguuaaaau aaggctagtc 60

cguuuu 66

<210> 290
<211> 20
<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成寡核苷酸"

<400> 290
gaguccgagc agaagaagaa 20

<210> 291
<211> 20
<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成寡核苷酸"

[0157] <400> 291
gacaucgaug uccuccccau 20

<210> 292
<211> 20
<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成寡核苷酸"

<400> 292
gucaccucca augacuaggg 20

<210> 293
<211> 20
<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成

寡核苷酸″

<400> 293
auuggguguu cagggcagag 20

<210> 294
<211> 20
<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释=″人工序列的描述：合成寡核苷酸″

<400> 294
guggcgagag gggccgagau 20

<210> 295
<211> 20
<212> RNA
<213> 人工序列

[0158]
<220>
<221> 来源
<223> /注释=″人工序列的描述：合成寡核苷酸″

<400> 295
ggggccgaga uuggguguuc 20

<210> 296
<211> 20
<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释=″人工序列的描述：合成寡核苷酸″

<400> 296
gugccauuag cuaaaugcau 20

<210> 297
<211> 20
<212> RNA

<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 297
guaccaccca caggugccag 20

<210> 298
<211> 20
<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 298
gaaagccucu gggccaggaa 20

[0159]

<210> 299
<211> 48
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 299
ctggaggagg aagggcctga gtccgagcag aagaagaagg gctcccat 48

<210> 300
<211> 20
<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 300
gaguccgagc agaagaagau 20

<210> 301
<211> 20
<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成寡核苷酸"

<400> 301
gaguccgagc agaagaagua 20

<210> 302
<211> 20
<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成寡核苷酸"

[0160] <400> 302
gaguccgagc agaagaacaa 20

<210> 303
<211> 20
<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成寡核苷酸"

<400> 303
gaguccgagc agaagaugaa 20

<210> 304
<211> 20
<212> RNA
<213> 人工序列

<220>
<221> 来源

<223> /注释="人工序列的描述：合成寡核苷酸"

<400> 304
gaguccgagc agaaguagaa 20

<210> 305
<211> 20
<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成寡核苷酸"

<400> 305
gaguccgagc agaugaagaa 20

<210> 306
<211> 20
<212> RNA
<213> 人工序列

[0161]

<220>
<221> 来源
<223> /注释="人工序列的描述：合成寡核苷酸"

<400> 306
gaguccgagc acaagaagaa 20

<210> 307
<211> 20
<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成寡核苷酸"

<400> 307
gaguccgagg agaagaagaa 20

<210> 308
<211> 20

<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成寡核苷酸"

<400> 308
gaguccgugc agaagaagaa 20

<210> 309
<211> 20
<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成寡核苷酸"

<400> 309
gagucggagc agaagaagaa 20

[0162]

<210> 310
<211> 20
<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成寡核苷酸"

<400> 310
gagaccgagc agaagaagaa 20

<210> 311
<211> 24
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成寡核苷酸"

<400> 311

aatgacaagc ttgctagcgg tggg 24

<210> 312
<211> 39
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成寡核苷酸"

<400> 312
aaaacggaag ggcctgagtc cgagcagaag aagaagttt 39

<210> 313
<211> 39
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成寡核苷酸"

[0163]

<400> 313
aaacaggggc cgagattggg tgttcagggc agaggtttt 39

<210> 314
<211> 38
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成寡核苷酸"

<400> 314
aaaacggaag ggcctgagtc cgagcagaag aagaagtt 38

<210> 315
<211> 40
<212> DNA
<213> 人工序列

<220>

<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 315
aacggaggga ggggcacaga tgagaaactc agggttttag 40

<210> 316
<211> 38
<212> DNA
<213> 人

<400> 316
agcccttctt cttctgctcg gactcaggcc cttcctcc 38

<210> 317
<211> 40
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
[0164] 寡核苷酸"

<400> 317
cagggaggga ggggcacaga tgagaaactc aggaggcccc 40

<210> 318
<211> 80
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 318
ggcaatgcgc caccggttga tgtgatggga gcccttctag gaggccccca gagcagccac 60

tggggcctca acactcaggc 80

<210> 319
<211> 23
<212> DNA
<213> 人

<400> 319
gtcacctcca atgactaggg tgg 23

<210> 320
<211> 25
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成寡核苷酸"

<220>
<221> 修饰_碱基
<222> (6)..(25)
<223> a, c, t, g, 未知或其它

<400> 320
caccgnnnnn nnnnnnnnnn nnnnn 25

[0165]

<210> 321
<211> 25
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成寡核苷酸"

<220>
<221> 修饰_碱基
<222> (5)..(24)
<223> a, c, t, g, 未知或其它

<400> 321
aaacnnnnnn nnnnnnnnnn nnnnc 25

<210> 322
<211> 33
<212> DNA
<213> 人

<400> 322

catcgatgtc ctccccattg gcctgcttcg tgg 33

<210> 323
<211> 33
<212> DNA
<213> 人

<400> 323
ttcgtggcaa tgcgccaccg gttgatgtga tgg 33

<210> 324
<211> 33
<212> DNA
<213> 人

<400> 324
tcgtggcaat gcgccaccgg ttgatgtgat ggg 33

<210> 325
<211> 33
<212> DNA
<213> 人

[0166]

<400> 325
tccagcttct gccgtttgta ctttgtcctc cgg 33

<210> 326
<211> 33
<212> DNA
<213> 人

<400> 326
ggagggaggg gcacagatga gaaactcagg agg 33

<210> 327
<211> 33
<212> DNA
<213> 人

<400> 327
aggggccgag attgggtgtt cagggcagag agg 33

<210> 328
<211> 54
<212> DNA

<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 328
aacaccgggt cttcgagaag acctgtttta gagctagaaa tagcaagtta aaat 54

<210> 329
<211> 54
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 329
caaaacgggt cttcgagaag acgttttaga gctatgctgt tttgaatggt ccca 54

[0167]

<210> 330
<211> 33
<212> DNA
<213> 小鼠

<400> 330
caagcactga gtgccattag ctaaatgcat agg 33

<210> 331
<211> 33
<212> DNA
<213> 小鼠

<400> 331
aatgcatagg gtaccaccca caggtgccag ggg 33

<210> 332
<211> 33
<212> DNA
<213> 小鼠

<400> 332
acacacatgg gaaagcctct gggccaggaa agg 33

<210> 333
<211> 37
<212> DNA
<213> 人

<400> 333
ggaggaggta gtatacagaa acacagagaa gtagaat 37

<210> 334
<211> 37
<212> DNA
<213> 人

<400> 334
agaatgtaga ggagtcacag aaactcagca ctagaaa 37

<210> 335
<211> 98
<212> DNA
<213> 人工序列

<220>
<221> 来源
[0168] <223> /注释="人工序列的描述：合成寡核苷酸"

<400> 335
ggacgaaaca ccggaaccat tcaaaacagc atagcaagtt aaaataaggc tagtccgtta 60

tcaacttgaa aaagtggcac cgagtcggtg cttttttt 98

<210> 336
<211> 186
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成多核苷酸"

<400> 336
ggacgaaaca ccggtagtat taagtattgt tttatggctg ataaatttct ttgaatttct 60

ccttgattat ttgttataaa agttataaaa taatcttgtt ggaaccattc aaaacagcat 120

agcaagttaa aataaggcta gtccgttatc aacttgaaaa agtggcaccg agtcggtgct 180

tttttt 186

<210> 337
<211> 95
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成寡核苷酸"

<400> 337
gggttttaga gctatgctgt tttgaatggt cccaaaacgg gtcttcgaga agacgtttta 60

gagctatgct gttttgaatg gtcccaaaac ttttt 95

<210> 338
<211> 36
<212> DNA
<213> 人工序列

[0169]

<220>
<221> 来源
<223> /注释="人工序列的描述：合成寡核苷酸"

<220>
<221> 修饰_碱基
<222> (5)..(34)
<223> a, c, t, g, 未知或其它

<400> 338
aaacnnnnnn nnnnnnnnnn nnnnnnnnnn nnnngt 36

<210> 339
<211> 36
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成寡核苷酸"

<220>
<221> 修饰_碱基
<222> (7)..(36)
<223> a, c, t, g, 未知或其它

<400> 339
taaaacnnnn nnnnnnnnnn nnnnnnnnnn nnnnnn 36

<210> 340
<211> 84
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述: 合成
寡核苷酸"

<400> 340
gtggaaagga cgaaacaccg ggtcttcgag aagacctgtt ttagagctag aaatagcaag 60

ttaaaataag gctagtccgt tttt 84

[0170] <210> 341
<211> 24
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述: 合成
寡核苷酸"

<220>
<221> 修饰_碱基
<222> (6)..(24)
<223> a, c, t, g, 未知或其它

<400> 341
caccgnnnnn nnnnnnnnnn nnnn 24

<210> 342
<211> 24
<212> DNA
<213> 人工序列

<220>

<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<220>
<221> 修饰_碱基
<222> (5)..(23)
<223> a, c, t, g, 未知或其它

<400> 342
aaacnnnnnn nnnnnnnnnn nnnc 24

<210> 343
<211> 88
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 343
[0171] gttttagagc tatgctgttt tgaatggtcc caaaactgag accaaaggtc tcgttttaga 60

gctatgctgt tttgaatggt cccaaaac 88

<210> 344
<211> 35
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 344
aaacggaagg gcctgagtcc gagcagaaga agaag 35

<210> 345
<211> 35
<212> DNA
<213> 人工序列

<220>
<221> 来源

<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 345
aaaacttctt cttctgctcg gactcaggcc cttcc 35

<210> 346
<211> 46
<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<220>
<221> 修饰_碱基
<222> (1)..(19)
<223> a, c, u, g, 未知或其它

<400> 346
[0172] nnnnnnnnnn nnnnnnnnng uuauuguacu cucaagauuu auuuuu 46

<210> 347
<211> 91
<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 347
guuacuuaaa ucuugcagaa gcuacaaaga uaaggcuuca ugccgaaauc aacacccugu 60

cauuuuaugg caggguguuu ucguuauuua a 91

<210> 348
<211> 70
<212> DNA
<213> 人

<400> 348
ttttctagtg ctgagtttct gtgactcctc tacattctac ttctctgtgt ttctgtatac 60

tacctcctcc 70

<210> 349
<211> 122
<212> DNA
<213> 人

<400> 349
ggaggaaggg cctgagtccg agcagaagaa gaagggctcc catcacatca accggtggcg 60

cattgccacg aagcaggcca atggggagga catcgatgtc acctccaatg actagggtgg 120

gc 122

<210> 350
<211> 48
<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述: 合成
寡核苷酸"

[0173]

<220>
<221> 修饰_碱基
<222> (3)..(32)
<223> a, c, u, g, 未知或其它

<400> 350
acnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnguuuuaga gcuaugcu 48

<210> 351
<211> 67
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述: 合成
寡核苷酸"

<220>
<221> 来源
<223> /注释="组合 DNA/RNA 分子的描述 : 合成
寡核苷酸"

<400> 351
agcauagcaa guuaaaauaa ggctaguccg uuaucaacuu gaaaaagugg caccgagucg 60

gugcuuu 67

<210> 352
<211> 62
<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<220>
<221> 修饰_碱基
<222> (1)..(20)
<223> a, c, u, g, 未知或其它

<400> 352
nnnnnnnnnn nnnnnnnnnn guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

[0174] cg 62

<210> 353
<211> 73
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 353
tgaatggtcc caaaacggaa gggcctgagt ccgagcagaa gaagaagttt tagagctatg 60

ctgttttgaa tgg 73

<210> 354
<211> 69
<212> DNA
<213> 人

<400> 354
ctggtcttcc acctctctgc cctgaacacc caatctcggc ccctctcgcc accctcctgc 60

atttctgtt 69

<210> 355
<211> 138
<212> DNA
<213> 小鼠

<400> 355
acccaagcac tgagtgccat tagctaaatg catagggtac cacccacagg tgccaggggc 60

ctttcccaaa gttcccagcc ccttctccaa cctttcctgg cccagaggct ttcccatgtg 120

tgtggctgga ccctttga 138

<210> 356
<211> 20
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成引物"

[0175]

<400> 356
gtgctttgca gaggcctacc 20

<210> 357
<211> 20
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成引物"

<400> 357
cctggagcgc atgcagtagt 20

<210> 358
<211> 22
<212> DNA
<213> 人工序列

<220>

<220>
<221> 来源
<223> /注释="人工序列的描述：合成引物"

<400> 358
accttctgtg tttccaccat tc 22

<210> 359
<211> 20
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成引物"

<400> 359
ttggggagtg cacagacttc 20

<210> 360
<211> 20
<212> DNA
[0176] <213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成引物"

<400> 360
ggctccctgg gttcaaagta 20

<210> 361
<211> 21
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成引物"

<400> 361
agaggggtct ggatgtcgta a 21

<210> 362

<211> 30
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
探针"

<400> 362
tagctctaaa acttcttctt ctgctcggac 30

<210> 363
<211> 30
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
探针"

<400> 363
ctagccttat tttaacttgc tatgctgttt 30

[0177]

<210> 364
<211> 99
<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<220>
<221> 修饰_碱基
<222> (1)..(20)
<223> a, c, u, g, 未知或其它

<400> 364
nnnnnnnnnn nnnnnnnnnn guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuaucaac uugaaaaagu ggcaccgagu cggugcuuu 99

<210> 365
<211> 12

<210> DNA
<213> 人

<400> 365
tagcgggtaa gc 12

<210> 366
<211> 12
<212> DNA
<213> 人

<400> 366
tcggtgacat gt 12

<210> 367
<211> 12
<212> DNA
<213> 人

<400> 367
actccccgta gg 12

[0178]

<210> 368
<211> 12
<212> DNA
<213> 人

<400> 368
actgcgtgtt aa 12

<210> 369
<211> 12
<212> DNA
<213> 人

<400> 369
acgtcgcctg at 12

<210> 370
<211> 12
<212> DNA
<213> 人

<400> 370
taggtcgacc ag 12

<210> 371
<211> 12
<212> DNA
<213> 人

<400> 371
ggcgttaatg at 12

<210> 372
<211> 12
<212> DNA
<213> 人

<400> 372
tgtcgcatgt ta 12

<210> 373
<211> 12
<212> DNA
<213> 人

<400> 373
[0179] atggaaacgc at 12

<210> 374
<211> 12
<212> DNA
<213> 人

<400> 374
gccgaattcc tc 12

<210> 375
<211> 12
<212> DNA
<213> 人

<400> 375
gcatggtacg ga 12

<210> 376
<211> 12
<212> DNA
<213> 人

<400> 376
cggtactctt ac 12

<210> 377
<211> 12
<212> DNA
<213> 人

<400> 377
gcctgtgccg ta 12

<210> 378
<211> 12
<212> DNA
<213> 人

<400> 378
tacggtaagt cg 12

<210> 379
<211> 12
<212> DNA
[0180] <213> 人

<400> 379
cacgaaatta cc 12

<210> 380
<211> 12
<212> DNA
<213> 人

<400> 380
aaccaagata cg 12

<210> 381
<211> 12
<212> DNA
<213> 人

<400> 381
gagtcgatac gc 12

<210> 382
<211> 12

<212> DNA
<213> 人

<400> 382
gtctcacgat cg 12

<210> 383
<211> 12
<212> DNA
<213> 人

<400> 383
tcgtcgggtg ca 12

<210> 384
<211> 12
<212> DNA
<213> 人

<400> 384
actccgtagt ga 12

[0181]

<210> 385
<211> 12
<212> DNA
<213> 人

<400> 385
caggacgtcc gt 12

<210> 386
<211> 12
<212> DNA
<213> 人

<400> 386
tcgtatccct ac 12

<210> 387
<211> 12
<212> DNA
<213> 人

<400> 387
tttcaaggcc gg 12

<210> 388
<211> 12
<212> DNA
<213> 人

<400> 388
cgccggtgga at 12

<210> 389
<211> 12
<212> DNA
<213> 人

<400> 389
gaacccgtcc ta 12

<210> 390
<211> 12
<212> DNA
<213> 人

<400> 390
[0182] gattcatcag cg 12

<210> 391
<211> 12
<212> DNA
<213> 人

<400> 391
acaccggtct tc 12

<210> 392
<211> 12
<212> DNA
<213> 人

<400> 392
atcgtgccct aa 12

<210> 393
<211> 12
<212> DNA
<213> 人

<400> 393
gcgtcaatgt tc 12

<210> 394
<211> 12
<212> DNA
<213> 人

<400> 394
ctccgtatct cg 12

<210> 395
<211> 12
<212> DNA
<213> 人

<400> 395
ccgattcctt cg 12

<210> 396
<211> 12
<212> DNA
[0183] <213> 人

<400> 396
tgcgcctcca gt 12

<210> 397
<211> 12
<212> DNA
<213> 人

<400> 397
taacgtcgga gc 12

<210> 398
<211> 12
<212> DNA
<213> 人

<400> 398
aaggtcgccc at 12

<210> 399
<211> 12

<212> DNA
<213> 人

<400> 399
gtcggggact at 12

<210> 400
<211> 12
<212> DNA
<213> 人

<400> 400
ttcgagcgat tt 12

<210> 401
<211> 12
<212> DNA
<213> 人

<400> 401
tgagtcgtcg ag 12

[0184]

<210> 402
<211> 12
<212> DNA
<213> 人

<400> 402
tttacgcaga gg 12

<210> 403
<211> 12
<212> DNA
<213> 人

<400> 403
aggaagtatc gc 12

<210> 404
<211> 12
<212> DNA
<213> 人

<400> 404
actcgatacc at 12

<210> 405
<211> 12
<212> DNA
<213> 人

<400> 405
cgctacatag ca 12

<210> 406
<211> 12
<212> DNA
<213> 人

<400> 406
ttcataaccg gc 12

<210> 407
<211> 12
<212> DNA
<213> 人

<400> 407
[0185] ccaaacggtt aa 12

<210> 408
<211> 12
<212> DNA
<213> 人

<400> 408
cgattccttc gt 12

<210> 409
<211> 12
<212> DNA
<213> 人

<400> 409
cgtcatgaat aa 12

<210> 410
<211> 12
<212> DNA
<213> 人

<400> 410
agtggcgatg ac 12

<210> 411
<211> 12
<212> DNA
<213> 人

<400> 411
cccctacggc ac 12

<210> 412
<211> 12
<212> DNA
<213> 人

<400> 412
gccaacccgc ac 12

<210> 413
<211> 12
<212> DNA
[0186] <213> 人

<400> 413
tgggacaccg gt 12

<210> 414
<211> 12
<212> DNA
<213> 人

<400> 414
ttgactgcgg cg 12

<210> 415
<211> 12
<212> DNA
<213> 人

<400> 415
actatgcgta gg 12

<210> 416
<211> 12

<212> DNA
<213> 人

<400> 416
tcacccaaag cg 12

<210> 417
<211> 12
<212> DNA
<213> 人

<400> 417
gcaggacgtc cg 12

<210> 418
<211> 12
<212> DNA
<213> 人

<400> 418
acaccgaaaa cg 12

[0187]

<210> 419
<211> 12
<212> DNA
<213> 人

<400> 419
cggtgtattg ag 12

<210> 420
<211> 12
<212> DNA
<213> 人

<400> 420
cacgaggtat gc 12

<210> 421
<211> 12
<212> DNA
<213> 人

<400> 421
taaagcgacc cg 12

<210> 422
<211> 12
<212> DNA
<213> 人

<400> 422
cttagtcggc ca 12

<210> 423
<211> 12
<212> DNA
<213> 人

<400> 423
cgaaaacgtg gc 12

<210> 424
<211> 12
<212> DNA
<213> 人

<400> 424
[0188] cgtgccctga ac 12

<210> 425
<211> 12
<212> DNA
<213> 人

<400> 425
tttaccatcg aa 12

<210> 426
<211> 12
<212> DNA
<213> 人

<400> 426
cgtagccatg tt 12

<210> 427
<211> 12
<212> DNA
<213> 人

<400> 427
cccaaacggt ta 12

<210> 428
<211> 12
<212> DNA
<213> 人

<400> 428
gcgttatcag aa 12

<210> 429
<211> 12
<212> DNA
<213> 人

<400> 429
tcgatggtaa ac 12

<210> 430
<211> 12
<212> DNA
[0189] <213> 人

<400> 430
cgacttttg ca 12

<210> 431
<211> 12
<212> DNA
<213> 人

<400> 431
tcgacgactc ac 12

<210> 432
<211> 12
<212> DNA
<213> 人

<400> 432
acgcgtcaga ta 12

<210> 433
<211> 12

<212> DNA
<213> 人

<400> 433
cgtacggcac ag 12

<210> 434
<211> 12
<212> DNA
<213> 人

<400> 434
ctatgccgtg ca 12

<210> 435
<211> 12
<212> DNA
<213> 人

<400> 435
cgcgtcagat at 12

[0190]

<210> 436
<211> 12
<212> DNA
<213> 人

<400> 436
aagatcggta gc 12

<210> 437
<211> 12
<212> DNA
<213> 人

<400> 437
cttcgcaagg ag 12

<210> 438
<211> 12
<212> DNA
<213> 人

<400> 438
gtcgtggact ac 12

<210> 439
<211> 12
<212> DNA
<213> 人

<400> 439
ggtcgtcatc aa 12

<210> 440
<211> 12
<212> DNA
<213> 人

<400> 440
gttaacagcg tg 12

<210> 441
<211> 12
<212> DNA
<213> 人

<400> 441
[0191] tagctaaccg tt 12

<210> 442
<211> 12
<212> DNA
<213> 人

<400> 442
agtaaaggcg ct 12

<210> 443
<211> 12
<212> DNA
<213> 人

<400> 443
ggtaatttcg tg 12

<210> 444
<211> 69
<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 444
gucaccucca augacuaggg guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuuuuuu 69

<210> 445
<211> 69
<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 445
gacaucgaug uccuccccau guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuuuuuu 69

[0192]

<210> 446
<211> 69
<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 446
gaguccgagc agaagaagaa guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuuuuuu 69

<210> 447
<211> 69
<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成

寡核苷酸″

<400> 447
ggggccgaga uuggguguuc guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuuuuuu 69

<210> 448
<211> 69
<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释=″人工序列的描述：合成
寡核苷酸″

<400> 448
guggcgagag gggccgagau guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuuuuuu 69

<210> 449
[0193] <211> 76
<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释=″人工序列的描述：合成
寡核苷酸″

<400> 449
gucaccucca augacuaggg guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuaucauu uuuuuu 76

<210> 450
<211> 76
<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释=″人工序列的描述：合成
寡核苷酸″

<400> 450

gacaucgaug uccucccccau guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuaucauu uuuuuu 76

<210> 451
<211> 76
<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 451
gaguccgagc agaagaagaa guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuaucauu uuuuuu 76

<210> 452
<211> 76
<212> RNA
<213> 人工序列

[0194]

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 452
ggggccgaga uuggguguuc guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuaucauu uuuuuu 76

<210> 453
<211> 76
<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 453
guggcgagag gggccgagau guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuaucauu uuuuuu 76

<210> 454
<211> 88
<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成寡核苷酸"

<400> 454
gucaccucca augacuaggg guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuaucaac uugaaaaagu guuuuuuu 88

<210> 455
<211> 88
<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成寡核苷酸"

[0195]

<400> 455
gacaucgaug uccuccccau guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuaucaac uugaaaaagu guuuuuuu 88

<210> 456
<211> 88
<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成寡核苷酸"

<400> 456
gaguccgagc agaagaagaa guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuaucaac uugaaaaagu guuuuuuu 88

<210> 457

<211> 88
<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 457
ggggccgaga uuggguguuc guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuaucaac uugaaaaagu guuuuuuu 88

<210> 458
<211> 88
<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

[0196] <400> 458
guggcgagag gggccgagau guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuaucaac uugaaaaagu guuuuuuu 88

<210> 459
<211> 103
<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 459
gucaccucca augacuaggg guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu uuu 103

<210> 460
<211> 103
<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 460
gacaucgaug uccucccccau guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu uuu 103

<210> 461
<211> 103
<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 461
gaguccgagc agaagaagaa guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

[0197] cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu uuu 103

<210> 462
<211> 103
<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 462
ggggccgaga uuggguguuc guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu uuu 103

<210> 463
<211> 103
<212> RNA
<213> 人工序列

<220>
<221> 来源

<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 463
guggcgagag gggccgagau guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu uuu 103

<210> 464
<211> 120
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<400> 464
gtggaaagga cgaaacaccg ggtcttcgag aagacctgtt ttagagctag aaatagcaag 60

ttaaaataag gctagtccgt tatcaacttg aaaaagtggc accgagtcgg tgctttttt 120

[0198] <210> 465
<211> 40
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 465
tcggtgcgct ggttgatttc ttcttgcgct tttttggctt 40

<210> 466
<211> 26
<212> RNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 466
gauuucuucu ugcgcuuuuu guuuua 26

<210> 467
<211> 26
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成寡核苷酸"

<220>
<221> 修饰_碱基
<222> (22)..(26)
<223> a, c, t, g, 未知或其它

<400> 467
tgatttcttc ttgcgctttt tnnnnn 26

<210> 468
<211> 26
<212> DNA
<213> 人工序列

[0199]

<220>
<221> 来源
<223> /注释="人工序列的描述：合成寡核苷酸"

<220>
<221> 修饰_碱基
<222> (21)..(21)
<223> a, c, t, g, 未知或其它

<400> 468
tgatttcttc ttgcgctttt ntggct 26

<210> 469
<211> 26
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成寡核苷酸"

<220>
<221> 修饰_碱基
<222> (2)..(2)
<223> a, c, t, g, 未知或其它

<400> 469
tnatttcttc ttgcgctttt ttggct 26

<210> 470
<211> 23
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 470
gatttcttct tgcgcttttt tgg 23

[0200]

<210> 471
<211> 34
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<220>
<221> CDS
<222> (1)..(33)

<400> 471
tcc atc cgt aca acc cac aac cct gct agt gag c 34
Ser Ile Arg Thr Thr His Asn Pro Ala Ser Glu
1 5 10

<210> 472
<211> 11
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
肽"

<400> 472
Ser Ile Arg Thr Thr His Asn Pro Ala Ser Glu
1 5 10

<210> 473
<211> 34
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<220>
<221> CDS
<222> (1)..(33)

[0201]

<400> 473
tcc atc gca aca acc cac aac cct gct agt gag c 34
Ser Ile Ala Thr Thr His Asn Pro Ala Ser Glu
1 5 10

<210> 474
<211> 11
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
肽"

<400> 474
Ser Ile Ala Thr Thr His Asn Pro Ala Ser Glu
1 5 10

<210> 475
<211> 34
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<220>
<221> CDS
<222> (1)..(33)

<400> 475
tca atc cgt aca acc cac aac cct gct agt gag c 34
Ser Ile Arg Thr Thr His Asn Pro Ala Ser Glu
1 5 10

<210> 476
<211> 42
<212> DNA
<213> 人

<220>
<221> CDS
<222> (1)..(36)

[0202]

<400> 476
caa ttg aat tta aaa gaa acc gat acc gtt ttg gtt taagga 42
Gln Leu Asn Leu Lys Glu Thr Asp Thr Val Leu Val
1 5 10

<210> 477
<211> 12
<212> PRT
<213> 人

<400> 477
Gln Leu Asn Leu Lys Glu Thr Asp Thr Val Leu Val
1 5 10

<210> 478
<211> 42
<212> DNA
<213> 人

<220>
<221> CDS
<222> (1)..(42)

<400> 478
caa ttg aat tta aaa gaa acc gat acc gtt tac gaa att gga 42
Gln Leu Asn Leu Lys Glu Thr Asp Thr Val Tyr Glu Ile Gly
1 5 10

<210> 479
<211> 14
<212> PRT
<213> 人

<400> 479
Gln Leu Asn Leu Lys Glu Thr Asp Thr Val Tyr Glu Ile Gly
1 5 10

<210> 480
<211> 34
<212> DNA
<213> 人

<220>
<221> CDS
<222> (2)..(34)

[0203]

<400> 480
t cct aaa aaa ccg aac tcc gcg ctg cgt aaa gta 34
Pro Lys Lys Pro Asn Ser Ala Leu Arg Lys Val
1 5 10

<210> 481
<211> 11
<212> PRT
<213> 人

<400> 481
Pro Lys Lys Pro Asn Ser Ala Leu Arg Lys Val
1 5 10

<210> 482
<211> 34
<212> DNA
<213> 人

<220>
<221> CDS

<222> (2)..(34)

<400> 482
t cct aca aaa ccg aac tcc gcg ctg cgt aaa gta 34
Pro Thr Lys Pro Asn Ser Ala Leu Arg Lys Val
1 5 10

<210> 483
<211> 11
<212> PRT
<213> 人

<400> 483
Pro Thr Lys Pro Asn Ser Ala Leu Arg Lys Val
1 5 10

<210> 484
<211> 33
<212> DNA
<213> 人

<400> 484
tgcgctggtt gatttcttct tgcgcttttt tgg 33

[0204]

<210> 485
<211> 33
<212> DNA
<213> 人

<400> 485
tacgctggtt gatttcttct tgcgcttttt ttg 33

<210> 486
<211> 27
<212> DNA
<213> 人

<400> 486
ggagggtttt atggggaaag gccattg 27

<210> 487
<211> 29
<212> DNA
<213> 人

<400> 487

gtaaaaaaga agactagaaa ttttgatac 29

<210> 488
<211> 46
<212> DNA
<213> 人

<400> 488
ggagggtttt atggggaaag gcaaagaaga ctagaaattt tgatac 46

<210> 489
<211> 27
<212> DNA
<213> 人

<400> 489
aggtgaagca taatgtctca aaaaata 27

<210> 490
<211> 29
<212> DNA
<213> 人

[0205]

<400> 490
attttattaa tacaaatcag tgaaatcat 29

<210> 491
<211> 46
<212> DNA
<213> 人

<400> 491
aggtgaagca taatgtctca aaattaatac aaatcagtga aatcat 46

<210> 492
<211> 36
<212> DNA
<213> 人

<220>
<221> CDS
<222> (1)..(36)

<400> 492
aat tta aaa gaa acc gat acc gtt tac gaa att gga 36

Asn Leu Lys Glu Thr Asp Thr Val Tyr Glu Ile Gly
1 5 10

<210> 493
<211> 12
<212> PRT
<213> 人

<400> 493
Asn Leu Lys Glu Thr Asp Thr Val Tyr Glu Ile Gly
1 5 10

<210> 494
<211> 36
<212> DNA
<213> 人

<220>
<221> CDS
<222> (1)..(30)

<400> 494
[0206] aat tta aaa gaa acc gat acc gtt ttg gtt taagga 36
Asn Leu Lys Glu Thr Asp Thr Val Leu Val
1 5 10

<210> 495
<211> 10
<212> PRT
<213> 人

<400> 495
Asn Leu Lys Glu Thr Asp Thr Val Leu Val
1 5 10

<210> 496
<211> 36
<212> DNA
<213> 人

<220>
<221> CDS
<222> (1)..(36)

<400> 496

tgg gat cca aaa aaa tat ggt ggt ttt gat agt cca 36
Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro
1 5 10

<210> 497
<211> 12
<212> PRT
<213> 人

<400> 497
Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro
1 5 10

<210> 498
<211> 36
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成寡核苷酸"

[0207]

<220>
<221> CDS
<222> (1)..(36)

<400> 498
tgg gat cca aaa aaa tat tgt ggt ttt gat agt cca 36
Trp Asp Pro Lys Lys Tyr Cys Gly Phe Asp Ser Pro
1 5 10

<210> 499
<211> 12
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成肽"

<400> 499
Trp Asp Pro Lys Lys Tyr Cys Gly Phe Asp Ser Pro
1 5 10

<210> 500
<211> 35
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 500
aaactacttt acgcagcgcg gagttcggtt ttttg 35

<210> 501
<211> 4104
<212> DNA
<213> 人

<220>
<221> CDS
<222> (1)..(4104)

<400> 501
[0208]
```
atg gac aag aag tac agc atc ggc ctg gac atc ggc acc aac tct gtg  48
Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1               5                  10                  15

ggc tgg gcc gtg atc acc gac gag tac aag gtg ccc agc aag aaa ttc  96
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20                  25                  30

aag gtg ctg ggc aac acc gac cgg cac agc atc aag aag aac ctg atc 144
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35                  40                  45

gga gcc ctg ctg ttc gac agc ggc gaa aca gcc gag gcc acc cgg ctg 192
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50                  55                  60

aag aga acc gcc aga aga aga tac acc aga cgg aag aac cgg atc tgc 240
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65                  70                  75                  80

tat ctg caa gag atc ttc agc aac gag atg gcc aag gtg gac gac agc 288
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85                  90                  95

ttc ttc cac aga ctg gaa gag tcc ttc ctg gtg gaa gag gat aag aag 336
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
```

```
                100                 105                 110

cac gag cgg cac ccc atc ttc ggc aac atc gtg gac gag gtg gcc tac        384
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

cac gag aag tac ccc acc atc tac cac ctg aga aag aaa ctg gtg gac        432
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

agc acc gac aag gcc gac ctg cgg ctg atc tat ctg gcc ctg gcc cac        480
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

atg atc aag ttc cgg ggc cac ttc ctg atc gag ggc gac ctg aac ccc        528
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

gac aac agc gac gtg gac aag ctg ttc atc cag ctg gtg cag acc tac        576
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

aac cag ctg ttc gag gaa aac ccc atc aac gcc agc ggc gtg gac gcc        624
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205
```

[0209]

```
aag gcc atc ctg tct gcc aga ctg agc aag agc aga cgg ctg gaa aat        672
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

ctg atc gcc cag ctg ccc ggc gag aag aag aat ggc ctg ttc ggc aac        720
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

ctg att gcc ctg agc ctg ggc ctg acc ccc aac ttc aag agc aac ttc        768
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

gac ctg gcc gag gat gcc aaa ctg cag ctg agc aag gac acc tac gac        816
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

gac gac ctg gac aac ctg ctg gcc cag atc ggc gac cag tac gcc gac        864
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

ctg ttt ctg gcc gcc aag aac ctg tcc gac gcc atc ctg ctg agc gac        912
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

atc ctg aga gtg aac acc gag atc acc aag gcc ccc ctg agc gcc tct        960
```

```
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

atg atc aag aga tac gac gag cac cac cag gac ctg acc ctg ctg aaa      1008
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

gct ctc gtg cgg cag cag ctg cct gag aag tac aaa gag att ttc ttc      1056
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

gac cag agc aag aac ggc tac gcc ggc tac att gac ggc gga gcc agc      1104
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

cag gaa gag ttc tac aag ttc atc aag ccc atc ctg gaa aag atg gac      1152
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

ggc acc gag gaa ctg ctc gtg aag ctg aac aga gag gac ctg ctg cgg      1200
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

aag cag cgg acc ttc gac aac ggc agc atc ccc cac cag atc cac ctg      1248
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

gga gag ctg cac gcc att ctg cgg cgg cag gaa gat ttt tac cca ttc      1296
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

ctg aag gac aac cgg gaa aag atc gag aag atc ctg acc ttc cgc atc      1344
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

ccc tac tac gtg ggc cct ctg gcc agg gga aac agc aga ttc gcc tgg      1392
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

atg acc aga aag agc gag gaa acc atc acc ccc tgg aac ttc gag gaa      1440
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

gtg gtg gac aag ggc gct tcc gcc cag agc ttc atc gag cgg atg acc      1488
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

aac ttc gat aag aac ctg ccc aac gag aag gtg ctg ccc aag cac agc      1536
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510
```

[0210]

```
ctg ctg tac gag tac ttc acc gtg tat aac gag ctg acc aaa gtg aaa      1584
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

tac gtg acc gag gga atg aga aag ccc gcc ttc ctg agc ggc gag cag      1632
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

aaa aag gcc atc gtg gac ctg ctg ttc aag acc aac cgg aaa gtg acc      1680
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

gtg aag cag ctg aaa gag gac tac ttc aag aaa atc gag tgc ttc gac      1728
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

tcc gtg gaa atc tcc ggc gtg gaa gat cgg ttc aac gcc tcc ctg ggc      1776
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

aca tac cac gat ctg ctg aaa att atc aag gac aag gac ttc ctg gac      1824
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

aat gag gaa aac gag gac att ctg gaa gat atc gtg ctg acc ctg aca      1872
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

ctg ttt gag gac aga gag atg atc gag gaa cgg ctg aaa acc tat gcc      1920
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

cac ctg ttc gac gac aaa gtg atg aag cag ctg aag cgg cgg aga tac      1968
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

acc ggc tgg ggc agg ctg agc cgg aag ctg atc aac ggc atc cgg gac      2016
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

aag cag tcc ggc aag aca atc ctg gat ttc ctg aag tcc gac ggc ttc      2064
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

gcc aac aga aac ttc atg cag ctg atc cac gac gac agc ctg acc ttt      2112
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

aaa gag gac atc cag aaa gcc cag gtg tcc ggc cag ggc gat agc ctg      2160
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720
```

[0211]

```
cac gag cac att gcc aat ctg gcc ggc agc ccc gcc att aag aag ggc        2208
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

atc ctg cag aca gtg aag gtg gtg gac gag ctc gtg aaa gtg atg ggc        2256
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

cgg cac aag ccc gag aac atc gtg atc gcc atg gcc aga gag aac cag        2304
Arg His Lys Pro Glu Asn Ile Val Ile Ala Met Ala Arg Glu Asn Gln
        755                 760                 765

acc acc cag aag gga cag aag aac agc cgc gag aga atg aag cgg atc        2352
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

gaa gag ggc atc aaa gag ctg ggc agc cag atc ctg aaa gaa cac ccc        2400
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

gtg gaa aac acc cag ctg cag aac gag aag ctg tac ctg tac tac ctg        2448
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815
```

[0212]

```
cag aat ggg cgg gat atg tac gtg gac cag gaa ctg gac atc aac cgg        2496
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

ctg tcc gac tac gat gtg gac gcc atc gtg cct cag agc ttt ctg aag        2544
Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

gac gac tcc atc gac gcc aag gtg ctg acc aga agc gac aag gcc cgg        2592
Asp Asp Ser Ile Asp Ala Lys Val Leu Thr Arg Ser Asp Lys Ala Arg
    850                 855                 860

ggc aag agc gac aac gtg ccc tcc gaa gag gtc gtg aag aag atg aag        2640
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

aac tac tgg cgg cag ctg ctg aac gcc aag ctg att acc cag aga aag        2688
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

ttc gac aat ctg acc aag gcc gag aga ggc ggc ctg agc gaa ctg gat        2736
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

aag gcc ggc ttc atc aag aga cag ctg gtg gaa acc cgg cag atc aca        2784
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
```

```
          915                 920                 925

aag cac gtg gca cag atc ctg gac tcc cgg atg aac act aag tac gac      2832
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

gag aat gac aag ctg atc cgg gaa gtg aaa gtg atc acc ctg aag tcc      2880
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

aag ctg gtg tcc gat ttc cgg aag gat ttc cag ttt tac aaa gtg cgc      2928
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

gag atc aac aac tac cac cac gcc cac gcc gcc tac ctg aac gcc gtc      2976
Glu Ile Asn Asn Tyr His His Ala His Ala Ala Tyr Leu Asn Ala Val
            980                 985                 990

gtg gga acc gcc ctg atc aaa aag  tac cct aag ctg gaa  agc gag ttc    3024
Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005

gtg tac  ggc gac tac aag gtg  tac gac gtg cgg aag  atg atc gcc       3069
Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020
```

[0213]

```
aag agc  gag cag gaa atc ggc  aag gct acc gcc aag  tac ttc ttc       3114
Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035

tac agc  aac atc atg aac ttt  ttc aag acc gag att  acc ctg gcc       3159
Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
    1040                1045                1050

aac ggc  gag atc cgg aag cgg  cct ctg atc gag aca  aac ggc gaa       3204
Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile Glu Thr  Asn Gly Glu
    1055                1060                1065

acc ggg  gag atc gtg tgg gat  aag ggc cgg gat ttt  gcc acc gtg       3249
Thr Gly  Glu Ile Val Trp Asp  Lys Gly Arg Asp Phe  Ala Thr Val
    1070                1075                1080

cgg aaa  gtg ctg agc atg ccc  caa gtg aat atc gtg  aaa aag acc       3294
Arg Lys  Val Leu Ser Met Pro  Gln Val Asn Ile Val  Lys Lys Thr
    1085                1090                1095

gag gtg  cag aca ggc ggc ttc  agc aaa gag tct atc  ctg ccc aag       3339
Glu Val  Gln Thr Gly Gly Phe  Ser Lys Glu Ser Ile  Leu Pro Lys
    1100                1105                1110

agg aac  agc gat aag ctg atc  gcc aga aag aag gac  tgg gac cct       3384
```

```
Arg Asn  Ser Asp Lys Leu Ile  Ala Arg Lys Lys Asp  Trp Asp Pro
    1115                1120                1125

aag aag  tac ggc ggc ttc gac  agc ccc acc gtg gcc  tat tct gtg      3429
Lys Lys  Tyr Gly Gly Phe Asp  Ser Pro Thr Val Ala  Tyr Ser Val
    1130                1135                1140

ctg gtg  gtg gcc aaa gtg gaa  aag ggc aag tcc aag  aaa ctg aag      3474
Leu Val  Val Ala Lys Val Glu  Lys Gly Lys Ser Lys  Lys Leu Lys
    1145                1150                1155

agt gtg  aaa gag ctg ctg ggg  atc acc atc atg gaa  aga agc agc      3519
Ser Val  Lys Glu Leu Leu Gly  Ile Thr Ile Met Glu  Arg Ser Ser
    1160                1165                1170

ttc gag  aag aat ccc atc gac  ttt ctg gaa gcc aag  ggc tac aaa      3564
Phe Glu  Lys Asn Pro Ile Asp  Phe Leu Glu Ala Lys  Gly Tyr Lys
    1175                1180                1185

gaa gtg  aaa aag gac ctg atc  atc aag ctg cct aag  tac tcc ctg      3609
Glu Val  Lys Lys Asp Leu Ile  Ile Lys Leu Pro Lys  Tyr Ser Leu
    1190                1195                1200

ttc gag  ctg gaa aac ggc cgg  aag aga atg ctg gcc  tct gcc ggc      3654
Phe Glu  Leu Glu Asn Gly Arg  Lys Arg Met Leu Ala  Ser Ala Gly
    1205                1210                1215

gaa ctg  cag aag gga aac gaa  ctg gcc ctg ccc tcc  aaa tat gtg      3699
Glu Leu  Gln Lys Gly Asn Glu  Leu Ala Leu Pro Ser  Lys Tyr Val
    1220                1225                1230

aac ttc  ctg tac ctg gcc agc  cac tat gag aag ctg  aag ggc tcc      3744
Asn Phe  Leu Tyr Leu Ala Ser  His Tyr Glu Lys Leu  Lys Gly Ser
    1235                1240                1245

ccc gag  gat aat gag cag aaa  cag ctg ttt gtg gaa  cag cac aag      3789
Pro Glu  Asp Asn Glu Gln Lys  Gln Leu Phe Val Glu  Gln His Lys
    1250                1255                1260

cac tac  ctg gac gag atc atc  gag cag atc agc gag  ttc tcc aag      3834
His Tyr  Leu Asp Glu Ile Ile  Glu Gln Ile Ser Glu  Phe Ser Lys
    1265                1270                1275

aga gtg  atc ctg gcc gac gct  aat ctg gac aaa gtg  ctg tcc gcc      3879
Arg Val  Ile Leu Ala Asp Ala  Asn Leu Asp Lys Val  Leu Ser Ala
    1280                1285                1290

tac aac  aag cac cgg gat aag  ccc atc aga gag cag  gcc gag aat      3924
Tyr Asn  Lys His Arg Asp Lys  Pro Ile Arg Glu Gln  Ala Glu Asn
    1295                1300                1305
```

[0214]

```
atc atc  cac ctg ttt acc ctg  acc aat ctg gga gcc  cct gcc gcc        3969
Ile Ile  His Leu Phe Thr Leu  Thr Asn Leu Gly Ala  Pro Ala Ala
    1310                 1315                 1320

ttc aag  tac ttt gac acc acc  atc gac cgg aag agg  tac acc agc        4014
Phe Lys  Tyr Phe Asp Thr Thr  Ile Asp Arg Lys Arg  Tyr Thr Ser
    1325                 1330                 1335

acc aaa  gag gtg ctg gac gcc  acc ctg atc cac cag  agc atc acc        4059
Thr Lys  Glu Val Leu Asp Ala  Thr Leu Ile His Gln  Ser Ile Thr
    1340                 1345                 1350

ggc ctg  tac gag aca cgg atc  gac ctg tct cag ctg  gga ggc gac        4104
Gly Leu  Tyr Glu Thr Arg Ile  Asp Leu Ser Gln Leu  Gly Gly Asp
    1355                 1360                 1365
```

<210> 502
<211> 1368
<212> PRT
<213> 人

<400> 502

```
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
```

[0215]

```
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
```

[0216]

```
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320
```

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415

[0217] Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430

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

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620

[0218] Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670

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
705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Ala Met Ala Arg Glu Asn Gln
755 760 765

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
785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
[0219] 820 825 830

Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys
835 840 845

Asp Asp Ser Ile Asp Ala Lys Val Leu Thr Arg Ser Asp Lys Ala 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 Ala 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
    1010                1015                1020
```

[0220]

```
Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035

Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
    1040                1045                1050

Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile Glu Thr  Asn Gly Glu
    1055                1060                1065

Thr Gly  Glu Ile Val Trp Asp  Lys Gly Arg Asp Phe  Ala Thr Val
    1070                1075                1080

Arg Lys  Val Leu Ser Met Pro  Gln Val Asn Ile Val  Lys Lys Thr
    1085                1090                1095

Glu Val  Gln Thr Gly Gly Phe  Ser Lys Glu Ser Ile  Leu Pro Lys
    1100                1105                1110

Arg Asn  Ser Asp Lys Leu Ile  Ala Arg Lys Lys Asp  Trp Asp Pro
    1115                1120                1125
```

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys
1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly
1205 1210 1215

[0221]

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val
1220 1225 1230

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
1250 1255 1260

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys
1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala
1280 1285 1290

Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
1295 1300 1305

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala
1310 1315 1320

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
1340 1345 1350

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 503
<211> 15
<212> DNA
<213> 人

<400> 503
cagaagaaga agggc 15

<210> 504
<211> 51
<212> DNA
[0222] <213> 人

<400> 504
ccaatgggga ggacatcgat gtcacctcca atgactaggg tggtgggcaa c 51

<210> 505
<211> 15
<212> DNA
<213> 人

<400> 505
ctctggccac tccct 15

<210> 506
<211> 52
<212> DNA
<213> 人

<400> 506
acatcgatgt cacctccaat gacaagcttg ctagcggtgg gcaaccacaa ac 52

<210> 507

<211> 1733
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<400> 507
ccgtttaaac aattctgcag gaatctagtt attaatagta atcaattacg gggtcattag 60

ttcatagccc atatatggag ttccgcgtta cataacttac ggtaaatggc ccgcctggct 120

gaccgcccaa cgacccccgc ccattgacgt caataatgac gtatgttccc atagtaacgc 180

caatagggac tttccattga cgtcaatggg tggagtattt acggtaaact gcccacttgg 240

cagtacatca agtgtatcat atgccaagta cgcccccctat tgacgtcaat gacggtaaat 300

ggcccgcctg gcattatgcc cagtacatga ccttatggga ctttcctact tggcagtaca 360

tctacgtatt agtcatcgct attaccatgg tcgaggtgag ccccacgttc tgcttcactc 420

tccccatctc ccccccctcc ccacccccaa ttttgtattt atttattttt taattatttt 480

[0223]

gtgcagcgat gggggcgggg ggggggggggg ggcgcgcgcc aggcggggcg gggcggggcg 540

aggggcgggg cggggcgagg cggagaggtg cggcggcagc caatcagagc ggcgcgctcc 600

gaaagtttcc ttttatggcg aggcggcggc ggcggcggcc ctataaaaag cgaagcgcgc 660

ggcgggcgga agtcgctgcg cgctgccttc gccccgtgcc ccgctccgcc gccgcctcgc 720

gccgcccgcc ccggctctga ctgaccgcgt tactcccaca ggtgagcggg cgggacggcc 780

cttctcctcc gggctgtaat tagcgcttgg tttaatgacg gcttgtttct tttctgtggc 840

tgcgtgaaag ccttgagggg ctccgggagg gccctttgtg cgggggggagc ggctcggggg 900

gtgcgtgcgt gtgtgtgtgc gtggggagcg ccgcgtgcgg ctccgcgctg cccggcggct 960

gtgagcgctg cgggcgcggc gcggggcttt gtgcgctccg cagtgtgcgc gaggggagcg 1020

cggccggggg cggtgccccg cggtgcgggg ggggctgcga ggggaacaaa ggctgcgtgc 1080

ggggtgtgtg cgtgggggggg tgagcagggg gtgtgggcgc gtcggtcggg ctgcaacccc 1140

ccctgcaccc ccctccccga gttgctgagc acggcccggc ttcgggtgcg gggctccgta 1200

cggggcgtgg cgcggggctc gccgtgccgg gcgggggggtg gcggcaggtg ggggtgccgg 1260

gcggggcggg gccgcctcgg gccggggagg gctcgggggа ggggcgcggc ggcccccgga 1320

gcgccggcgg ctgtcgaggc gcggcgagcc gcagccattg ccttttatgg taatcgtgcg 1380

agagggcgca gggacttcct ttgtcccaaa tctgtgcgga gccgaaatct gggaggcgcc 1440

gccgcacccc ctctagcggg cgcggggcga agcggtgcgg cgccggcagg aaggaaatgg 1500

gcggggaggg ccttcgtgcg tcgccgcgcc gccgtcccct tctccctctc cagcctcggg 1560

gctgtccgcg gggggacggc tgccttcggg ggggacgggg cagggcgggg ttcggcttct 1620

ggcgtgtgac cggcggctct agagcctctg ctaaccatgt tcatgccttc ttctttttcc 1680

tacagctcct gggcaacgtg ctggttattg tgctgtctca tcattttggc aaa 1733

<210> 508
<211> 4269
<212> DNA
<213> 人工序列

<220>
<221> 来源
[0224] <223> /注释="人工序列的描述：合成
多核苷酸"

<400> 508
atggactata aggaccacga cggagactac aaggatcatg atattgatta caaagacgat 60

gacgataaga tggccccaaa gaagaagcgg aaggtcggta tccacggagt cccagcagcc 120

gacaagaagt acagcatcgg cctggacatc ggcaccaact ctgtgggctg ggccgtgatc 180

accgacgagt acaaggtgcc cagcaagaaa ttcaaggtgc tgggcaacac cgaccggcac 240

agcatcaaga agaacctgat cggagccctg ctgttcgaca gcggcgaaac agccgaggcc 300

acccggctga agagaaccgc cagaagaaga tacaccagac ggaagaaccg gatctgctat 360

ctgcaagaga tcttcagcaa cgagatggcc aaggtggacg acagcttctt ccacagactg 420

gaagagtcct tcctggtgga agaggataag aagcacgagc ggcaccccat cttcggcaac 480

atcgtggacg aggtggccta ccacgagaag taccccacca tctaccacct gagaaagaaa 540

ctggtggaca gcaccgacaa ggccgacctg cggctgatct atctggccct ggcccacatg 600

atcaagttcc ggggccactt cctgatcgag ggcgacctga accccgacaa cagcgacgtg 660

gacaagctgt tcatccagct ggtgcagacc tacaaccagc tgttcgagga aaaccccatc 720

aacgccagcg gcgtggacgc caaggccatc ctgtctgcca gactgagcaa gagcagacgg 780

ctggaaaatc tgatcgccca gctgcccggc gagaagaaga atggcctgtt cggaaacctg 840

attgccctga gcctgggcct gacccccaac ttcaagagca acttcgacct ggccgaggat 900

gccaaactgc agctgagcaa ggacacctac gacgacgacc tggacaacct gctggcccag 960

atcggcgacc agtacgccga cctgtttctg gccgccaaga acctgtccga cgccatcctg 1020

ctgagcgaca tcctgagagt gaacaccgag atcaccaagg ccccccctgag cgcctctatg 1080

atcaagagat acgacgagca ccaccaggac ctgaccctgc tgaaagctct cgtgcggcag 1140

cagctgcctg agaagtacaa agagattttc ttcgaccaga gcaagaacgg ctacgccggc 1200

tacattgacg gcggagccag ccaggaagag ttctacaagt tcatcaagcc catcctggaa 1260

aagatggacg gcaccgagga actgctcgtg aagctgaaca gagaggacct gctgcggaag 1320

cagcggacct tcgacaacgg cagcatcccc caccagatcc acctgggaga gctgcacgcc 1380

attctgcggc ggcaggaaga tttttaccca ttcctgaagg acaaccggga aaagatcgag 1440

[0225]

aagatcctga ccttccgcat cccctactac gtgggccctc tggccagggg aaacagcaga 1500

ttcgcctgga tgaccagaaa gagcgaggaa accatcaccc cctggaactt cgaggaagtg 1560

gtggacaagg gcgcttccgc ccagagcttc atcgagcgga tgaccaactt cgataagaac 1620

ctgcccaacg agaaggtgct gcccaagcac agcctgctgt acgagtactt caccgtgtat 1680

aacgagctga ccaaagtgaa atacgtgacc gagggaatga gaaagcccgc cttcctgagc 1740

ggcgagcaga aaaaggccat cgtggacctg ctgttcaaga ccaaccggaa agtgaccgtg 1800

aagcagctga aagaggacta cttcaagaaa atcgagtgct tcgactccgt ggaaatctcc 1860

ggcgtggaag atcggttcaa cgcctccctg ggcacatacc acgatctgct gaaaattatc 1920

aaggacaagg acttcctgga caatgaggaa aacgaggaca ttctggaaga tatcgtgctg 1980

accctgacac tgtttgagga cagagagatg atcgaggaac ggctgaaaac ctatgcccac 2040

ctgttcgacg acaaagtgat gaagcagctg aagcggcgga gatacaccgg ctggggcagg 2100

ctgagccgga agctgatcaa cggcatccgg gacaagcagt ccggcaagac aatcctggat 2160

ttcctgaagt ccgacggctt cgccaacaga aacttcatgc agctgatcca cgacgacagc 2220

ctgaccttta aagaggacat ccagaaagcc caggtgtccg gccagggcga tagcctgcac 2280

gagcacattg ccaatctggc cggcagcccc gccattaaga agggcatcct gcagacagtg 2340

aaggtggtgg acgagctcgt gaaagtgatg ggccggcaca agcccgagaa catcgtgatc 2400

gaaatggcca gagagaacca gaccacccag aagggacaga agaacagccg cgagagaatg 2460

aagcggatcg aagagggcat caaagagctg ggcagccaga tcctgaaaga acaccccgtg 2520

gaaaacaccc agctgcagaa cgagaagctg tacctgtact acctgcagaa tgggcgggat 2580

atgtacgtgg accaggaact ggacatcaac cggctgtccg actacgatgt ggaccatatc 2640

gtgcctcaga gctttctgaa ggacgactcc atcgacaaca aggtgctgac cagaagcgac 2700

aagaaccggg gcaagagcga caacgtgccc tccgaagagg tcgtgaagaa gatgaagaac 2760

tactggcggc agctgctgaa cgccaagctg attacccaga gaaagttcga caatctgacc 2820

aaggccgaga gaggcggcct gagcgaactg gataaggccg gcttcatcaa gagacagctg 2880

gtggaaaccc ggcagatcac aaagcacgtg gcacagatcc tggactcccg gatgaacact 2940

[0226] aagtacgacg agaatgacaa gctgatccgg gaagtgaaag tgatcaccct gaagtccaag 3000

ctggtgtccg atttccggaa ggatttccag ttttacaaag tgcgcgagat caacaactac 3060

caccacgccc acgacgccta cctgaacgcc gtcgtgggaa ccgccctgat caaaaagtac 3120

cctaagctgg aaagcgagtt cgtgtacggc gactacaagg tgtacgacgt gcggaagatg 3180

atcgccaaga gcgagcagga aatcggcaag gctaccgcca agtacttctt ctacagcaac 3240

atcatgaact ttttcaagac cgagattacc ctggccaacg gcgagatccg gaagcggcct 3300

ctgatcgaga caaacggcga aaccggggag atcgtgtggg ataagggccg ggattttgcc 3360

accgtgcgga aagtgctgag catgccccaa gtgaatatcg tgaaaaagac cgaggtgcag 3420

acaggcggct tcagcaaaga gtctatcctg cccaagagga acagcgataa gctgatcgcc 3480

agaaagaagg actgggaccc taagaagtac ggcggcttcg acagccccac cgtggcctat 3540

tctgtgctgg tggtggccaa agtggaaaag ggcaagtcca agaaactgaa gagtgtgaaa 3600

gagctgctgg ggatcaccat catggaaaga agcagcttcg agaagaatcc catcgacttt 3660

ctggaagcca agggctacaa agaagtgaaa aaggacctga tcatcaagct gcctaagtac 3720

tccctgttcg agctggaaaa cggccggaag agaatgctgg cctctgccgg cgaactgcag 3780

aagggaaacg aactggccct gccctccaaa tatgtgaact tcctgtacct ggccagccac 3840

tatgagaagc tgaagggctc ccccgaggat aatgagcaga aacagctgtt tgtggaacag 3900

cacaagcact acctggacga gatcatcgag cagatcagcg agttctccaa gagagtgatc 3960

ctggccgacg ctaatctgga caaagtgctg tccgcctaca acaagcaccg ggataagccc 4020

atcagagagc aggccgagaa tatcatccac ctgtttaccc tgaccaatct gggagccccct 4080

gccgccttca agtactttga caccaccatc gaccggaaga ggtacaccag caccaaagag 4140

gtgctggacg ccaccctgat ccaccagagc atcaccggcc tgtacgagac acggatcgac 4200

ctgtctcagc tgggaggcga caaaaggccg gcggccacga aaaaggccgg ccaggcaaaa 4260

aagaaaaag 4269

<210> 509
<211> 780
<212> DNA
<213> 人工序列

[0227]

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<400> 509
ggaagcggag ccactaactt ctccctgttg aaacaagcag gggatgtcga agagaatccc 60

gggccagtga gcaagggcga ggagctgttc accggggtgg tgcccatcct ggtcgagctg 120

gacggcgacg taaacggcca caagttcagc gtgtccggcg agggcgaggg cgatgccacc 180

tacggcaagc tgaccctgaa gttcatctgc accaccggca agctgcccgt gccctggccc 240

accctcgtga ccaccctgac ctacggcgtg cagtgcttca gccgctaccc cgaccacatg 300

aagcagcacg acttcttcaa gtccgccatg cccgaaggct acgtccagga gcgcaccatc 360

ttcttcaagg acgacggcaa ctacaagacc cgcgccgagg tgaagttcga gggcgacacc 420

ctggtgaacc gcatcgagct gaagggcatc gacttcaagg aggacggcaa catcctgggg 480

cacaagctgg agtacaacta caacagccac aacgtctata tcatggccga caagcagaag 540

aacggcatca aggtgaactt caagatccgc cacaacatcg aggacggcag cgtgcagctc 600

gccgaccact accagcagaa cacccccatc ggcgacggcc ccgtgctgct gcccgacaac 660

cactacctga gcacccagtc cgccctgagc aaagacccca acgagaagcg cgatcacatg 720

gtcctgctgg agttcgtgac cgccgccggg atcactctcg gcatggacga gctgtacaag 780

<210> 510
<211> 597
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<400> 510
cgataatcaa cctctggatt acaaaatttg tgaaagattg actggtattc ttaactatgt 60

tgctcctttt acgctatgtg gatacgctgc tttaatgcct ttgtatcatg ctattgcttc 120

ccgtatggct ttcattttct cctccttgta taaatcctgg ttgctgtctc tttatgagga 180

[0228] gttgtggccc gttgtcaggc aacgtggcgt ggtgtgcact gtgtttgctg acgcaacccc 240

cactggttgg ggcattgcca ccacctgtca gctcctttcc gggactttcg ctttcccccc 300

ccctattgcc acggcggaac tcatcgccgc ctgccttgcc cgctgctgga caggggctcg 360

gctgttgggc actgacaatt ccgtggtgtt gtcggggaaa tcatcgtcct ttccttggct 420

gctcgcctgt gttgccacct ggattctgcg cgggacgtcc ttctgctacg tcccttcggc 480

cctcaatcca gcggaccttc cttcccgcgg cctgctgccg gctctgcggc ctcttccgcg 540

tcttcgcctt cgccctcaga cgagtcggat ctccctttgg gccgcctccc cgcatcg 597

<210> 511
<211> 210
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<400> 511

cgacctcgac tgtgccttct agttgccagc catctgttgt ttgcccctcc cccgtgcctt 60

ccttgaccct ggaaggtgcc actcccactg tcctttccta ataaaatgag gaaattgcat 120

cgcattgtct gagtaggtgt cattctattc tgggggggtgg ggtggggcag gacagcaagg 180

gggaggattg ggaagacaat ggcaggcatg 210

<210> 512
<211> 906
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<220>
<221> 修饰_碱基
<222> (109)..(109)
<223> a, c, t, g, 未知或其它

[0229]

<220>
<221> 修饰_碱基
<222> (135)..(135)
<223> a, c, t, g, 未知或其它

<400> 512
ataacttcgt ataatgtatg ctatacgaag ttattcgcga tgaataaatg aaagcttgca 60

gatctgcgac tctagaggat ctgcgactct agaggatcat aatcagccnt accacatttt 120

gtagaggttt tactngcttt aaaaaacctc ccacacctcc ccctgaacct gaaacataaa 180

atgaatgcaa ttgttgttgt taacttgttt attgcagctt ataatggtta caaataaagc 240

aatagcatca caaatttcac aaataaagca ttttttttcac tgcattctag ttgtggtttg 300

tccaaactca tcaatgtatc ttatcatgtc tggatctgcg actctagagg atcataatca 360

gccataccac atttgtagag gttttacttg ctttaaaaaa cctcccacac ctccccctga 420

acctgaaaca taaaatgaat gcaattgttg ttgttaactt gtttattgca gcttataatg 480

gttacaaata aagcaatagc atcacaaatt tcacaaataa agcatttttt tcactgcatt 540

ctagttgtgg tttgtccaaa ctcatcaatg tatcttatca tgtctggatc tgcgactcta 600

gaggatcata atcagccata ccacatttgt agaggtttta cttgctttaa aaaacctccc 660

acacctcccc ctgaacctga aacataaaat gaatgcaatt gttgttgtta acttgtttat 720

tgcagcttat aatggttaca aataaagcaa tagcatcaca aatttcacaa ataaagcatt 780

tttttcactg cattctagtt gtggtttgtc caaactcatc aatgtatctt atcatgtctg 840

gatccccatc aagctgatcc ggaacccttа atataacttc gtataatgta tgctatacga 900

agttat 906

<210> 513
<211> 1079
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<400> 513
caggccctcc gagcgtggtg gagccgttct gtgagacagc cgggtacgag tcgtgacgct 60

[0230]

ggaaggggca agcgggtggt gggcaggaat gcggtccgcc ctgcagcaac cggaggggga 120

gggagaaggg agcggaaaag tctccaccgg acgcggccat ggctcggggg gggggggca 180

gcggaggagc gcttccggcc gacgtctcgt cgctgattgg cttcttttcc tcccgccgtg 240

tgtgaaaaca caaatggcgt gttttggttg gcgtaaggcg cctgtcagtt aacggcagcc 300

ggagtgcgca gccgccggca gcctcgctct gcccactggg tggggcggga ggtaggtggg 360

gtgaggcgag ctggacgtgc gggcgcggtc ggcctctggc ggggcggggg aggggaggga 420

gggtcagcga aagtagctcg cgcgcgagcg gccgcccacc ctccccttcc tctgggggag 480

tcgttttacc cgccgccggc cgggcctcgt cgtctgattg gctctcgggg cccagaaaac 540

tggcccttgc cattggctcg tgttcgtgca agttgagtcc atccgccggc cagcgggggc 600

ggcgaggagg cgctcccagg ttccggccct cccctcggcc ccgcgccgca gagtctggcc 660

gcgcgcccct gcgcaacgtg gcaggaagcg cgcgctgggg gcggggacgg gcagtagggc 720

tgagcggctg cggggcgggt gcaagcacgt ttccgacttg agttgcctca agaggggcgt 780

gctgagccag acctccatcg cgcactccgg ggagtggagg gaaggagcga gggctcagtt 840

gggctgtttt ggaggcagga agcacttgct ctcccaaagt cgctctgagt tgttatcagt 900

aagggagctg cagtggagta ggcggggaga aggccgcacc cttctccgga gggggggaggg 960

gagtgttgca ataccttttct gggagttctc tgctgcctcc tggcttctga ggaccgccct 1020

gggcctggga gaatcccttc cccctcttcc ctcgtgatct gcaactccag tctttctag 1079

<210> 514
<211> 4336
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<400> 514
agatgggcgg gagtcttctg ggcaggctta aaggctaacc tggtgtgtgg gcgttgtcct 60

gcaggggaat tgaacaggtg taaaattgga gggacaagac ttcccacaga ttttcggttt 120

[0231] tgtcgggaag ttttttaata ggggcaaata aggaaaatgg gaggataggt agtcatctgg 180

ggttttatgc agcaaaacta caggttatta ttgcttgtga tccgcctcgg agtattttcc 240

atcgaggtag attaaagaca tgctcacccg agttttatac tctcctgctt gagatcctta 300

ctacagtatg aaattacagt gtcgcgagtt agactatgta agcagaattt taatcatttt 360

taaagagccc agtacttcat atccatttct cccgctcctt ctgcagcctt atcaaaaggt 420

attttagaac actcatttta gccccatttt catttattat actggcttat ccaaccccta 480

gacagagcat tggcattttc cctttcctga tcttagaagt ctgatgactc atgaaaccag 540

acagattagt tacatacacc acaaatcgag gctgtagctg gggcctcaac actgcagttc 600

ttttataact ccttagtaca ctttttgttg atcctttgcc ttgatcctta attttcagtg 660

tctatcacct ctcccgtcag gtggtgttcc acatttgggc ctattctcag tccagggagt 720

tttacaacaa tagatgtatt gagaatccaa cctaaagctt aactttccac tcccatgaat 780

gcctctctcc tttttctcca tttataaact gagctattaa ccattaatgg tttccaggtg 840

gatgtctcct cccccaatat tacctgatgt atcttacata ttgccaggct gatattttaa 900

gacattaaaa ggtatatttc attattgagc cacatggtat tgattactgc ttactaaaat 960

tttgtcattg tacacatctg taaaaggtgg ttccttttgg aatgcaaagt tcaggtgttt 1020

gttgtctttc ctgacctaag gtcttgtgag cttgtatttt ttctatttaa gcagtgcttt 1080

ctcttggact ggcttgactc atggcattct acacgttatt gctggtctaa atgtgatttt 1140

gccaagcttc ttcaggacct ataattttgc ttgacttgta gccaaacaca agtaaaatga 1200

ttaagcaaca aatgtatttg tgaagcttgg tttttaggtt gttgtgttgt gtgtgcttgt 1260

gctctataat aatactatcc aggggctgga gaggtggctc ggagttcaag agcacagact 1320

gctcttccag aagtcctgag ttcaattccc agcaaccaca tggtggctca caaccatctg 1380

taatgggatc tgatgccctc ttctggtgtg tctgaagacc acaagtgtat tcacattaaa 1440

taaataaatc ctccttcttc ttctttttt ttttttaaa gagaatactg tctccagtag 1500

aatttactga agtaatgaaa tactttgtgt ttgttccaat atggtagcca ataatcaaat 1560

tactctttaa gcactggaaa tgttaccaag gaactaattt ttatttgaag tgtaactgtg 1620

gacagaggag ccataactgc agacttgtgg gatacagaag accaatgcag actttaatgt 1680

[0232]

cttttctctt acactaagca ataaagaaat aaaaattgaa cttctagtat cctatttgtt 1740

taaactgcta gctttactta acttttgtgc ttcatctata caaagctgaa agctaagtct 1800

gcagccatta ctaaacatga aagcaagtaa tgataatttt ggatttcaaa aatgtagggc 1860

cagagtttag ccagccagtg gtggtgcttg cctttatgcc tttaatccca gcactctgga 1920

ggcagagaca ggcagatctc tgagtttgag cccagcctgg tctacacatc aagttctatc 1980

taggatagcc aggaatacac acagaaaccc tgttggggag gggggctctg agatttcata 2040

aaattataat tgaagcattc cctaatgagc cactatggat gtggctaaat ccgtctacct 2100

ttctgatgag atttgggtat tatttttct gtctctgctg ttggttgggt cttttgacac 2160

tgtgggcttt ctttaaagcc tccttcctgc catgtggtct cttgtttgct actaacttcc 2220

catggcttaa atggcatggc tttttgcctt ctaagggcag ctgctgagat ttgcagcctg 2280

atttccaggg tggggttggg aaatctttca aacactaaaa ttgtccttta attttttttt 2340

taaaaaatgg gttatataat aaacctcata aaatagttat gaggagtgag gtggactaat 2400

attaaatgag tccctcccct ataaaagagc tattaaggct ttttgtctta tacttaactt 2460

tttttttaaa tgtggtatct ttagaaccaa gggtcttaga gttttagtat acagaaactg 2520

ttgcatcgct taatcagatt ttctagtttc aaatccagag aatccaaatt cttcacagcc 2580

aaagtcaaat taagaatttc tgacttttaa tgttaatttg cttactgtga atataaaaat 2640

gatagctttt cctgaggcag ggtctcacta tgtatctctg cctgatctgc aacaagatat 2700

gtagactaaa gttctgcctg cttttgtctc ctgaatacta aggttaaaat gtagtaatac 2760

ttttggaact tgcaggtcag attcttttat aggggacaca ctaagggagc ttgggtgata 2820

gttggtaaat gtgtttaagt gatgaaaact tgaattatta tcaccgcaac ctacttttta 2880

aaaaaaaaag ccaggcctgt tagagcatgc ttaagggatc cctaggactt gctgagcaca 2940

caagagtagt tacttggcag gctcctggtg agagcatatt tcaaaaaaca aggcagacaa 3000

ccaagaaact acagttaagg ttacctgtct ttaaaccatc tgcatataca cagggatatt 3060

aaaatattcc aaataatatt tcattcaagt tttcccccat caaattggga catggatttc 3120

tccggtgaat aggcagagtt ggaaactaaa caaatgttgg ttttgtgatt tgtgaaattg 3180

[0233] ttttcaagtg atagttaaag cccatgagat acagaacaaa gctgctattt cgaggtctct 3240

tggtttatac tcagaagcac ttctttgggt ttccctgcac tatcctgatc atgtgctagg 3300

cctaccttag gctgattgtt gttcaaataa acttaagttt cctgtcaggt gatgtcatat 3360

gatttcatat atcaaggcaa aacatgttat atatgttaaa catttgtact taatgtgaaa 3420

gttaggtctt tgtgggtttg atttttaatt ttcaaaacct gagctaaata agtcattttt 3480

acatgtctta catttggtgg aattgtataa ttgtggtttg caggcaagac tctctgacct 3540

agtaacccta cctatagagc actttgctgg gtcacaagtc taggagtcaa gcatttcacc 3600

ttgaagttga gacgttttgt tagtgtatac tagtttatat gttggaggac atgtttatcc 3660

agaagatatt caggactatt tttgactggg ctaaggaatt gattctgatt agcactgtta 3720

gtgagcattg agtggccttt aggcttgaat tggagtcact tgtatatctc aaataatgct 3780

ggccttttt aaaagccctt gttctttatc accctgtttt ctacataatt tttgttcaaa 3840

gaaatacttg tttggatctc cttttgacaa caatagcatg ttttcaagcc atatttttt 3900

tcctttttt ttttttttt ggtttttcga gacagggttt ctctgtatag ccctggctgt 3960

cctggaactc actttgtaga ccaggctggc ctcgaactca gaaatccgcc tgcctctgcc 4020

tcctgagtgc cgggattaaa ggcgtgcacc accacgcctg gctaagttgg atattttgtt 4080

atataactat aaccaatact aactccactg ggtggatttt taattcagtc agtagtctta 4140

agtggtcttt attggccctt cattaaaatc tactgttcac tctaacagag gctgttggta 4200

ctagtggcac ttaagcaact tcctacggat atactagcag attaagggtc agggatagaa 4260

actagtctag cgttttgtat acctaccagc tttatactac cttgttctga tagaaatatt 4320

tcaggacatc tagctt 4336

<210> 515
<211> 1846
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

[0234] <400> 515
aattctaccg ggtaggggag gcgcttttcc caaggcagtc tggagcatgc gctttagcag 60

ccccgctggg cacttggcgc tacacaagtg gcctctggcc tcgcacacat tccacatcca 120

ccggtaggcg ccaaccggct ccgttctttg gtggcccctt cgcgccacct tctactcctc 180

ccctagtcag gaagttcccc cccgccccgc agctcgcgtc gtgcaggacg tgacaaatgg 240

aagtagcacg tctcactagt ctcgtgcaga tggacagcac cgctgagcaa tggaagcggg 300

taggcctttg gggcagcggc caatagcagc tttgctcctt cgctttctgg gctcagaggc 360

tgggaagggg tgggtccggg ggcgggctca ggggcgggct caggggcggg gcgggcgccc 420

gaaggtcctc cggaggcccg gcattctgca cgcttcaaaa gcgcacgtct gccgcgctgt 480

tctcctcttc ctcatctccg ggcctttcga cctgcaatcg ccgctagcga agttcctatt 540

ctctagaaag tataggaact tcgccaccat gggatcggcc attgaacaag atggattgca 600

cgcaggttct ccggccgctt gggtggagag gctattcggc tatgactggg cacaacagac 660

aatcggctgc tctgatgccg ccgtgttccg gctgtcagcg caggggcgcc cggttctttt 720

tgtcaagacc gacctgtccg gtgccctgaa tgaactgcag gacgaggcag cgcggctatc 780

gtggctggcc acgacgggcg ttccttgcgc agctgtgctc gacgttgtca ctgaagcggg 840

aagggactgg ctgctattgg gcgaagtgcc ggggcaggat ctcctgtcat ctcaccttgc 900

tcctgccgag aaagtatcca tcatggctga tgcaatgcgg cggctgcata cgcttgatcc 960

ggctacctgc ccattcgacc accaagcgaa acatcgcatc gagcgagcac gtactcggat 1020

ggaagccggt cttgtcgatc aggatgatct ggacgaagag catcaggggc tcgcgccagc 1080

cgaactgttc gccaggctca aggcgcgcat gcccgacggc gatgatctcg tcgtgaccca 1140

tggcgatgcc tgcttgccga atatcatggt ggaaaatggc cgcttttctg gattcatcga 1200

ctgtggccgg ctgggtgtgg cggaccgcta tcaggacata gcgttggcta cccgtgatat 1260

tgctgaagag cttggcggcg aatgggctga ccgcttcctc gtgctttacg gtatcgccgc 1320

tcccgattcg cagcgcatcg ccttctatcg ccttcttgac gagttcttct gaggggatcc 1380

gctgtaagtc tgcagaaatt gatgatctat taaacaataa agatgtccac taaaatggaa 1440

gtttttcctg tcatactttg ttaagaaggg tgagaacaga gtacctacat tttgaatgga 1500

[0235] aggattggag ctacgggggt gggggtgggg tgggattaga taaatgcctg ctctttactg 1560

aaggctcttt actattgctt tatgataatg tttcatagtt ggatatcata atttaaacaa 1620

gcaaaaccaa attaagggcc agctcattcc tcccactcat gatctataga tctatagatc 1680

tctcgtggga tcattgtttt tctcttgatt cccactttgt ggttctaagt actgtggttt 1740

ccaaatgtgt cagtttcata gcctgaagaa cgagatcagc agcctctgtt ccacatacac 1800

ttcattctca gtattgtttt gccaagttct aattccatca gaaagc 1846

<210> 516
<211> 1519
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<400> 516
taccgggtag gggaggcgct tttcccaagg cagtctggag catgcgcttt agcagccccg 60

ctgggcactt ggcgctacac aagtggcctc tggcctcgca cacattccac atccaccggt 120

aggcgccaac cggctccgtt ctttggtggc cccttcgcgc caccttctac tcctccccta 180

gtcaggaagt tccccccccgc cccgcagctc gcgtcgtgca ggacgtgaca aatggaagta 240

gcacgtctca ctagtctcgt gcagatggac agcaccgctg agcaatggaa gcgggtaggc 300

ctttggggca gcggccaata gcagctttgc tccttcgctt tctgggctca gaggctggga 360

aggggtgggt ccgggggcgg gctcaggggc gggctcaggg gcggggcggg cgcccgaagg 420

tcctccggag gcccggcatt ctgcacgctt caaaagcgca cgtctgccgc gctgttctcc 480

tcttcctcat ctccgggcct ttcgacctgc aggtcctcgc catggatcct gatgatgttg 540

ttgattcttc taaatctttt gtgatggaaa acttttcttc gtaccacggg actaaacctg 600

gttatgtaga ttccattcaa aaaggtatac aaaagccaaa atctggtaca caaggaaatt 660

atgacgatga ttggaaaggg ttttatagta ccgacaataa atacgacgct gcgggatact 720

ctgtagataa tgaaaacccg ctctctggaa aagctggagg cgtggtcaaa gtgacgtatc 780

[0236] caggactgac gaaggttctc gcactaaaag tggataatgc cgaaactatt aagaaagagt 840

taggtttaag tctcactgaa ccgttgatgg agcaagtcgg aacggaagag tttatcaaaa 900

ggttcggtga tggtgcttcg cgtgtagtgc tcagccttcc cttcgctgag gggagttcta 960

gcgttgaata tattaataac tgggaacagg cgaaagcgtt aagcgtagaa cttgagatta 1020

attttgaaac ccgtggaaaa cgtggccaag atgcgatgta tgagtatatg gctcaagcct 1080

gtgcaggaaa tcgtgtcagg cgatctcttt gtgaaggaac cttacttctg tggtgtgaca 1140

taattggaca aactacctac agagatttaa agctctaagg taaatataaa atttttaagt 1200

gtataatgtg ttaaactact gattctaatt gtttgtgtat tttagattcc aacctatgga 1260

actgatgaat gggagcagtg gtggaatgca gatcctagag ctcgctgatc agcctcgact 1320

gtgccttcta gttgccagcc atctgttgtt tgcccctccc ccgtgccttc cttgaccctg 1380

gaaggtgcca ctcccactgt cctttcctaa taaaatgagg aaattgcatc gcattgtctg 1440

agtaggtgtc attctattct gggggggtggg gtggggcagg acagcaaggg ggaggattgg 1500

gaagacaata gcaggcatg 1519

<210> 517
<211> 23
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
引物"

<400> 517
gagggcctat ttcccatgat tcc 23

<210> 518
<211> 22
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 518
[0237] cttgtggaaa ggacgaaaca cc 22

<210> 519
<211> 45
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<220>
<221> 修饰_碱基
<222> (4)..(23)
<223> a, c, t, g, 未知或其它

<400> 519
aacnnnnnnn nnnnnnnnnn nnnggtgttt cgtcctttcc acaag 45

<210> 520
<211> 28
<212> DNA

<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 520
cctgagtgtt gaggccccag tggctgct 28

<210> 521
<211> 37
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 521
acgagggcag agtgctgctt gctgctggcc aggcccc 37

[0238]

<210> 522
<211> 68
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 522
catcaggctc tcagctcagc ctgagtgttg aggccctgct ggccaggccc ctgcgtgggc 60

ccaagctg 68

<210> 523
<211> 27
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
寡核苷酸"

<400> 523
gagggcctat ttcccatgat tccttca 27

<210> 524
<211> 125
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<220>
<221> 修饰_碱基
<222> (84)..(103)
<223> a, c, t, g, 未知或其它

<400> 524
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacnnnnnnn nnnnnnnnnn nnnggtgttt cgtcctttcc 120

[0239] acaag 125

<210> 525
<211> 111
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成
多核苷酸"

<220>
<221> 修饰_碱基
<222> (9)..(28)
<223> a, c, t, g, 未知或其它

<400> 525
gaaacaccnn nnnnnnnnnn nnnnnnnngt tttagagcta gaaatagcaa gttaaaataa 60

ggctagtccg ttatcaactt gaaaaagtgg caccgagtcg gtgctttttt t 111

<210> 526

<211> 20
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成寡核苷酸"

<400> 526
agctgtttta ctggtcggct 20

<210> 527
<211> 20
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成寡核苷酸"

<400> 527
aatggataca cctggtcgaa 20

[0240]

<210> 528
<211> 20
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注释="人工序列的描述：合成寡核苷酸"

<400> 528
caatggatac acctggtcga 20

<210> 529
<211> 68
<212> DNA
<213> 人

<400> 529
accatgtata ccacttgggc tttggcagta gctaactgca ctaaatataa tataaggagg 60

gttttatg 68

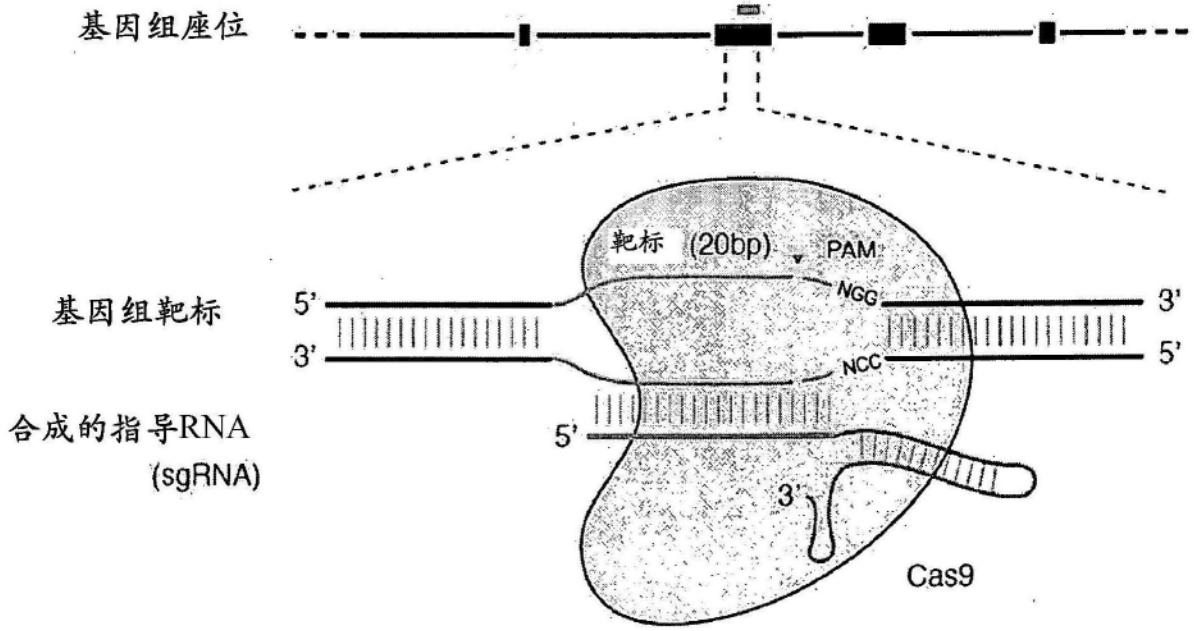


图1

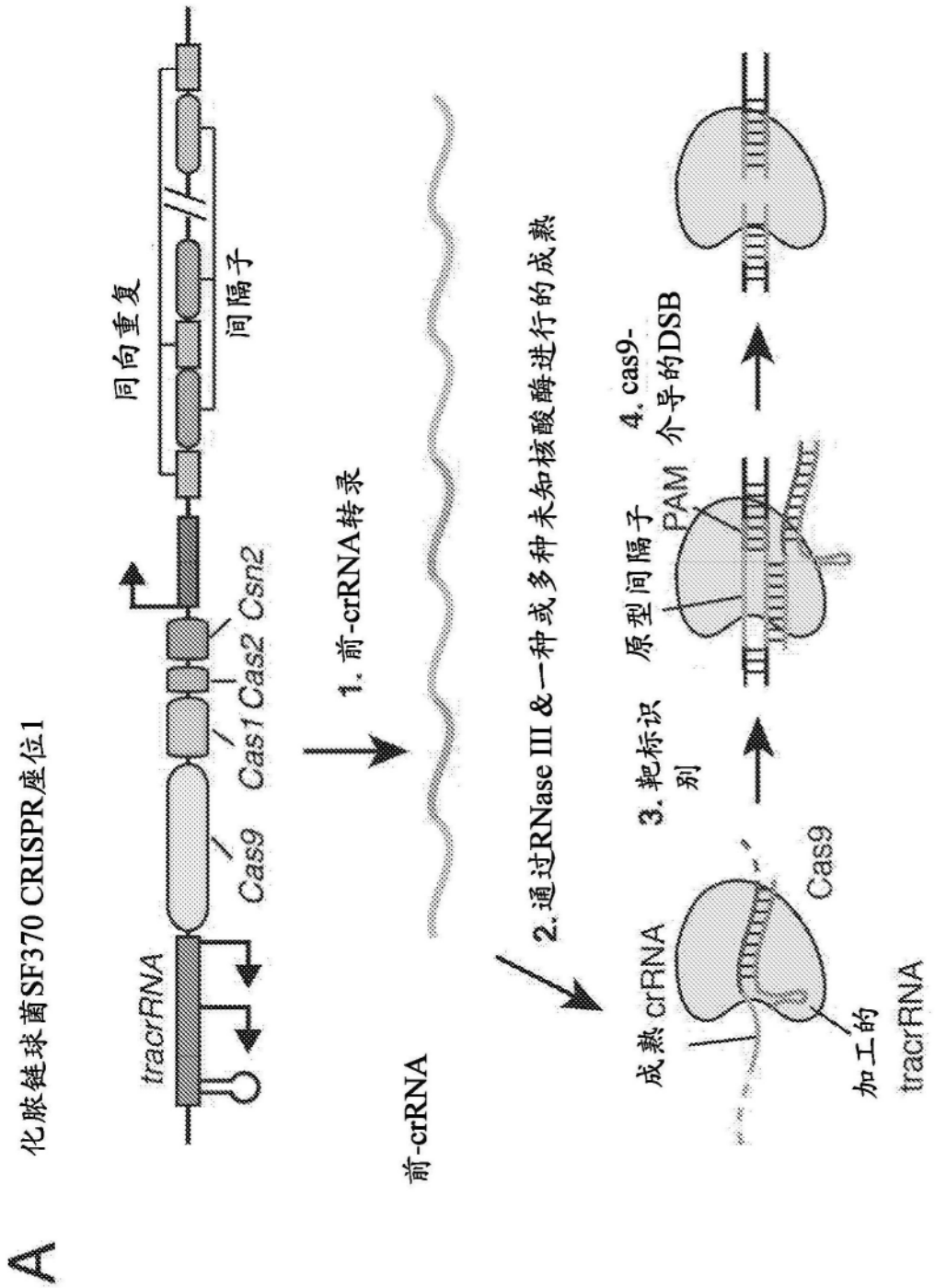

A
化脓链球菌SF370 CRISPR座位1
同向重复
间隔子
tracrRNA
Cas9
Cas1
Cas2
Csn2
1. 前-crRNA转录
前-crRNA
2. 通过RNase III &一种或多种未知核酸酶进行的成熟
成熟 crRNA
3. 靶标识别
原型间隔子
PAM
4. cas9-介导的DSB
Cas9
加工的
tracrRNA

图2A

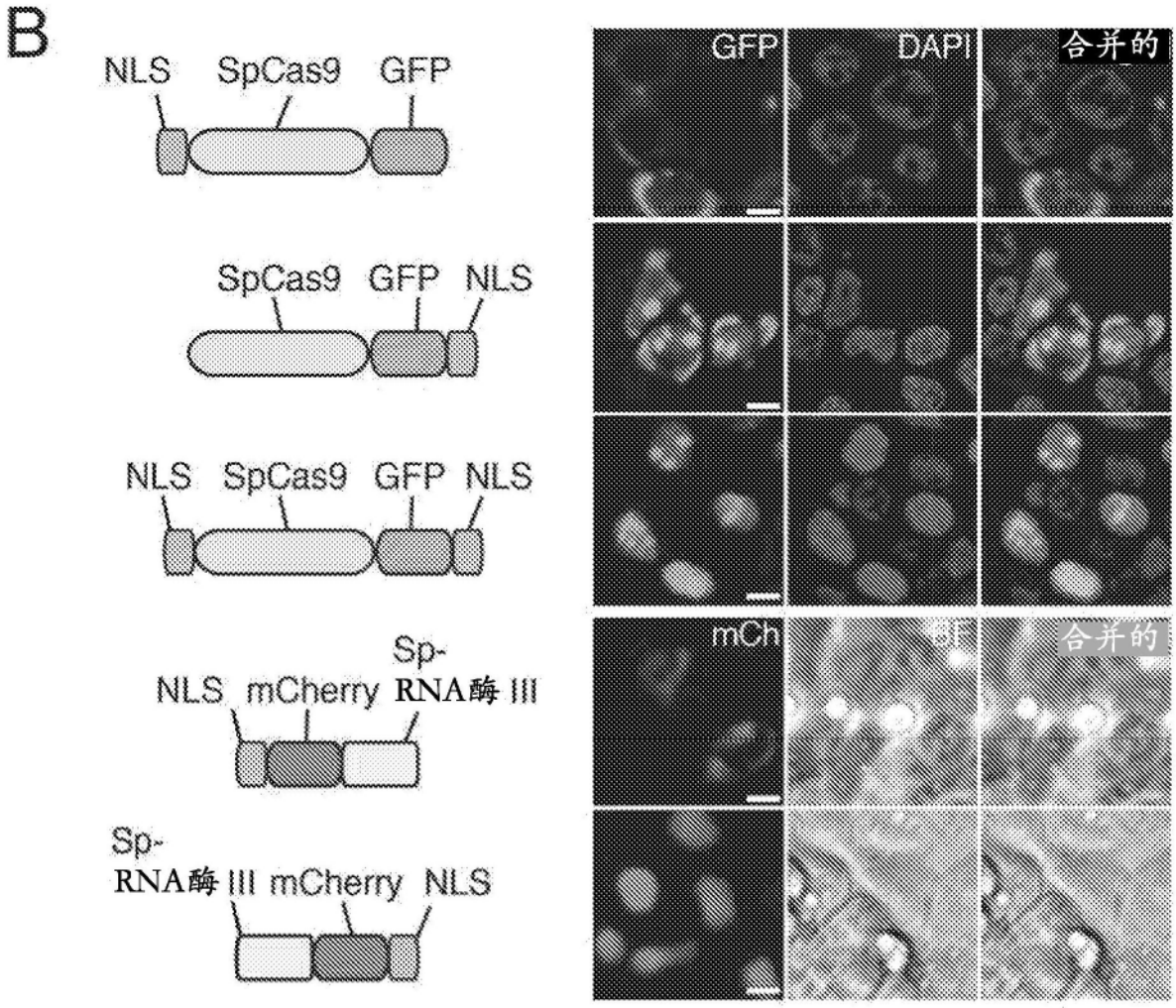

B
NLS SpCas9 GFP
SpCas9 GFP NLS
NLS SpCas9 GFP NLS
NLS mCherry Sp-RNA酶 III
Sp-RNA酶 III mCherry NLS
GFP
DAPI
合并的
mCh
BF
合并的

图2B

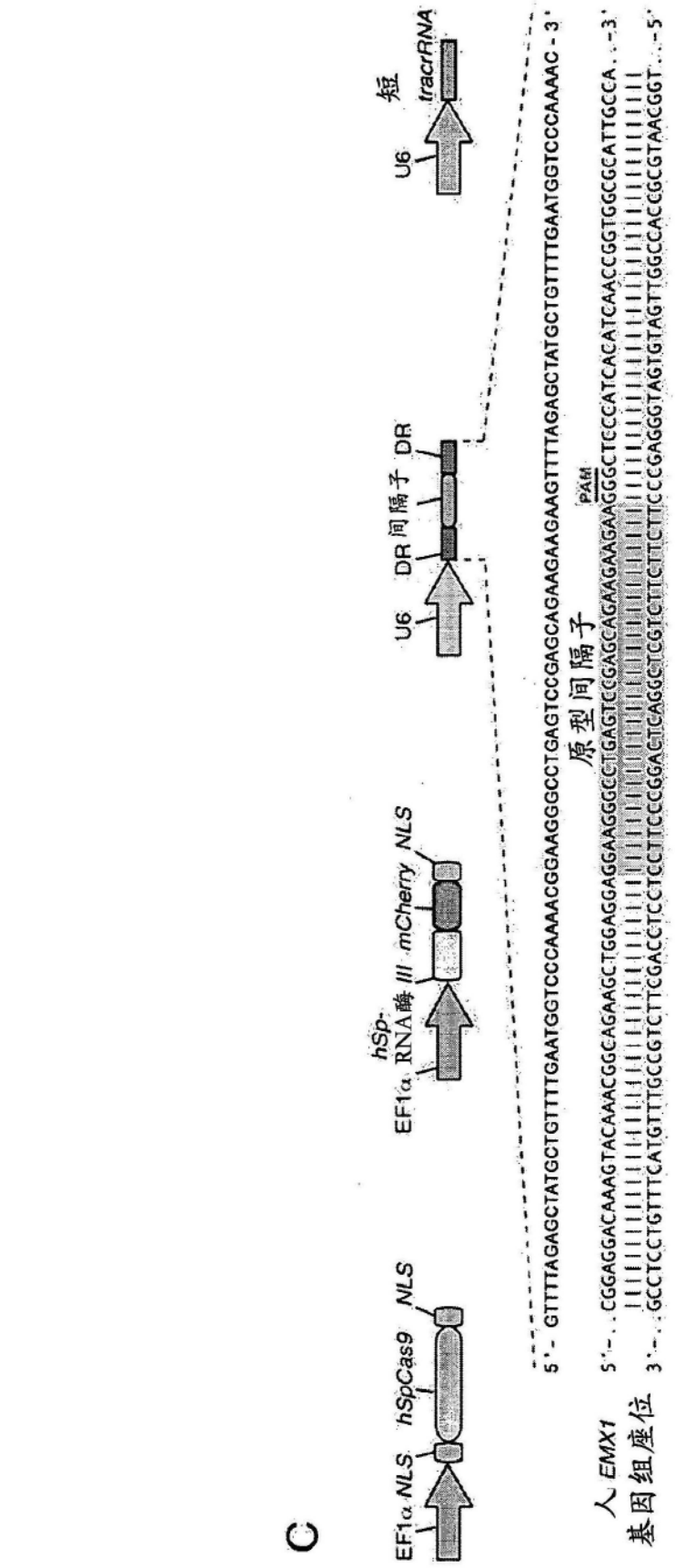

C
EF1α NLS
hSpCas9
NLS
hSp.
EF1α RNA酶 III
mCherry
NLS
U6
DR 间隔子 DR
U6
短
tracrRNA
5'- GTTTTAGAGCTATGCTGTTTTTGAATGGTCCCAAAACGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAAC - 3'
原型间隔子
PAM
人 EMX1
基因组座位
5'- . . CGGAGGACAAAGTACAAACGGCAGAAGCTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGGGCTCCCATCACATCAACCGGTGGCGCATTGCCA . . -3'
3'- . . GCCTCCTGTTTCATGTTTGCCGTCTTCGACCTCCTCCTTCCCGGACTCAGGCTCGTCTTCTTCTTCCCGAGGGTAGTGTAGTTGGCCACCGCGTAACGGT . . -5'

图2C

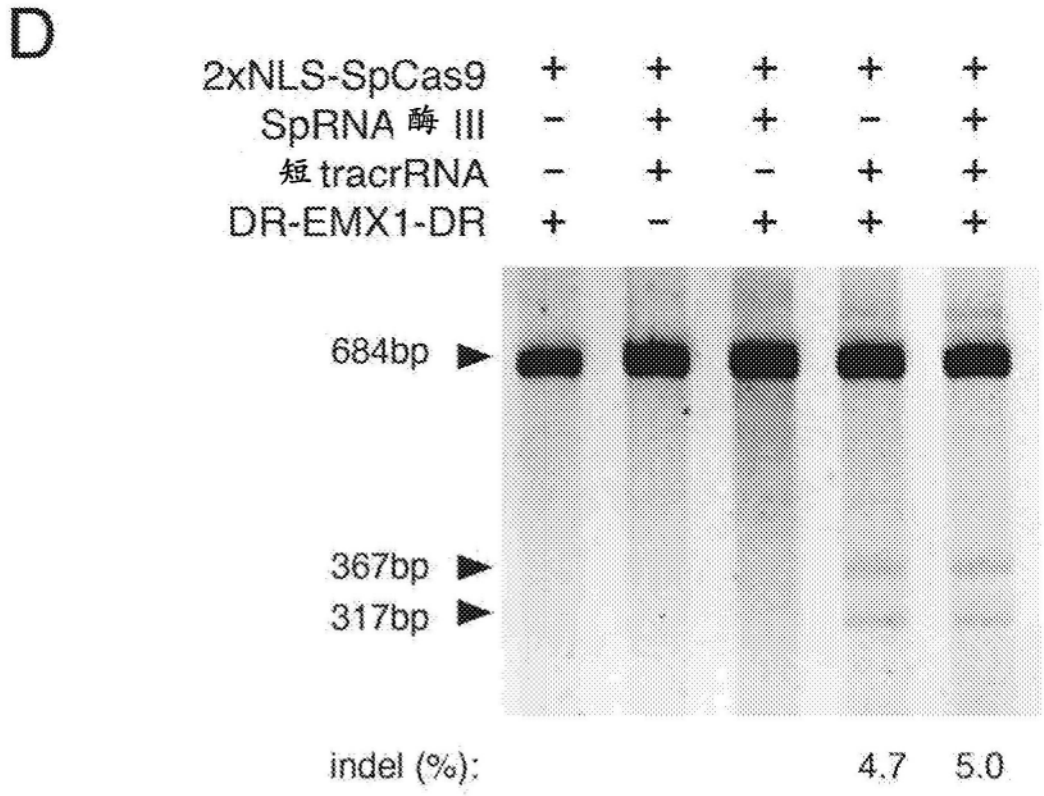

D
2xNLS-SpCas9 + + + + +
SpRNA 酶 III − + + − +
短 tracrRNA − + − + +
DR-EMX1-DR + − + + +
684bp
367bp
317bp
indel (%): 4.7 5.0

图2D

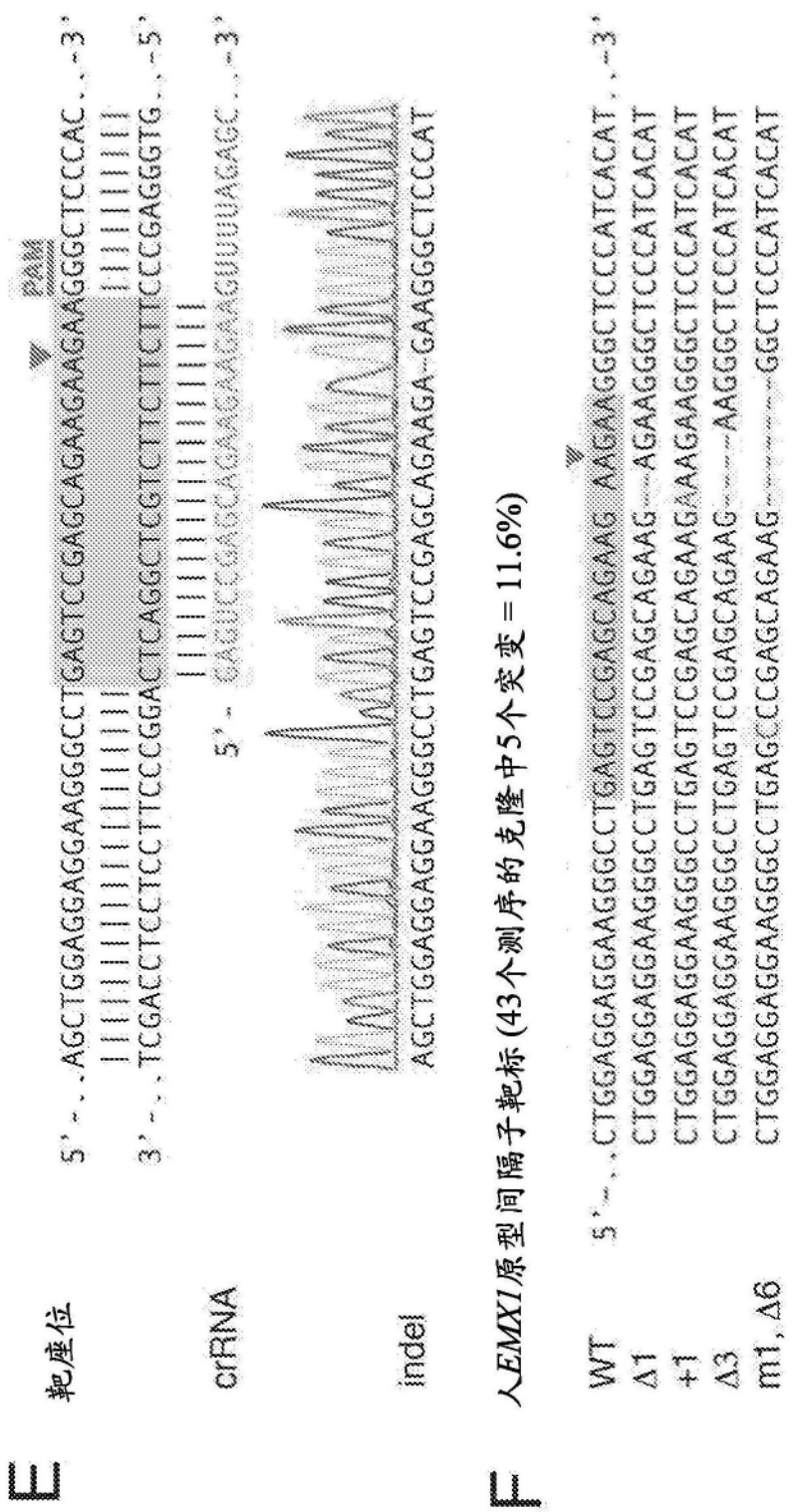

E
靶座位
PAM
5'-..AGCTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGAAGGGCTCCCAC..-3'
3'-..TCGACCTCCTCCTTCCCGGACTCAGGCTCGTCTTCTTCCCGAGGGTG..-5'
crRNA
5'- GAGUCCGAGCAGAAGAAGUUUUAGAGC..-3'
indel
AGCTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGA-GAAGGGCTCCCAT
F
人EMX1原型间隔子靶标 (43个测序的克隆中5个突变 = 11.6%)
WT 5'-..CTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAG AAGAAGGGCTCCCATCACAT..-3'
Δ1 CTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAG---AGAAGGGCTCCCATCACAT
+1 CTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGGGCTCCCATCACAT
Δ3 CTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAG---AAGGGCTCCCATCACAT
m1, Δ6 CTGGAGGAGGAAGGGCCTGAGCCCGAGCAGAAG--------GGCTCCCATCACAT

图2E图2F

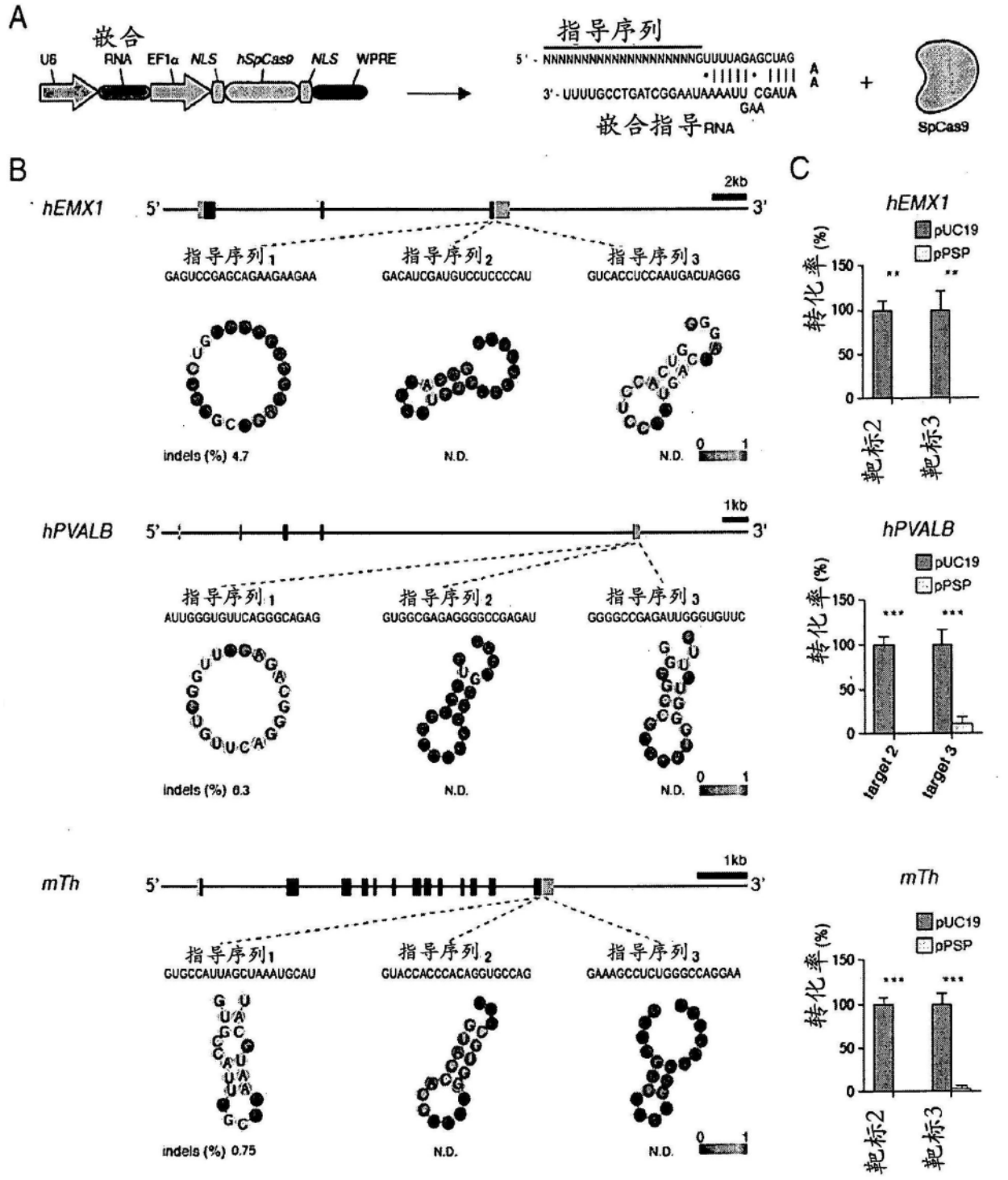

A
嵌合
U6
RNA
EF1α
NLS
hSpCas9
NLS
WPRE
指导序列
5' - NNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAG
3' - UUUUGCCTGATCGGAAUAAAAUU CGAUA
GAA
嵌合指导RNA
SpCas9
B
hEMX1
2kb
5'
3'
指导序列1
GAGUCCGAGCAGAAGAAGAA
指导序列2
GACAUCGAUGUCCUCCCCAU
指导序列3
GUCACCUCCAAUGACUAGGG
indels (%) 4.7
N.D.
N.D.
hPVALB
1kb
指导序列1
AUUGGGUGUUCAGGGCAGAG
指导序列2
GUGGCGAGAGGGGCCGAGAU
指导序列3
GGGGCCGAGAUUGGGUGUUC
indels (%) 0.3
N.D.
N.D.
mTh
1kb
指导序列1
GUGCCAUUAGCUAAAUGCAU
指导序列2
GUACCACCCACAGGUGCCAG
指导序列3
GAAAGCCUCUGGGCCAGGAA
indels (%) 0.75
N.D.
N.D.
C
hEMX1
转化率 (%)
pUC19
pPSP
靶标2
靶标3
hPVALB
target 2
target 3
mTh
靶标2
靶标3

图3

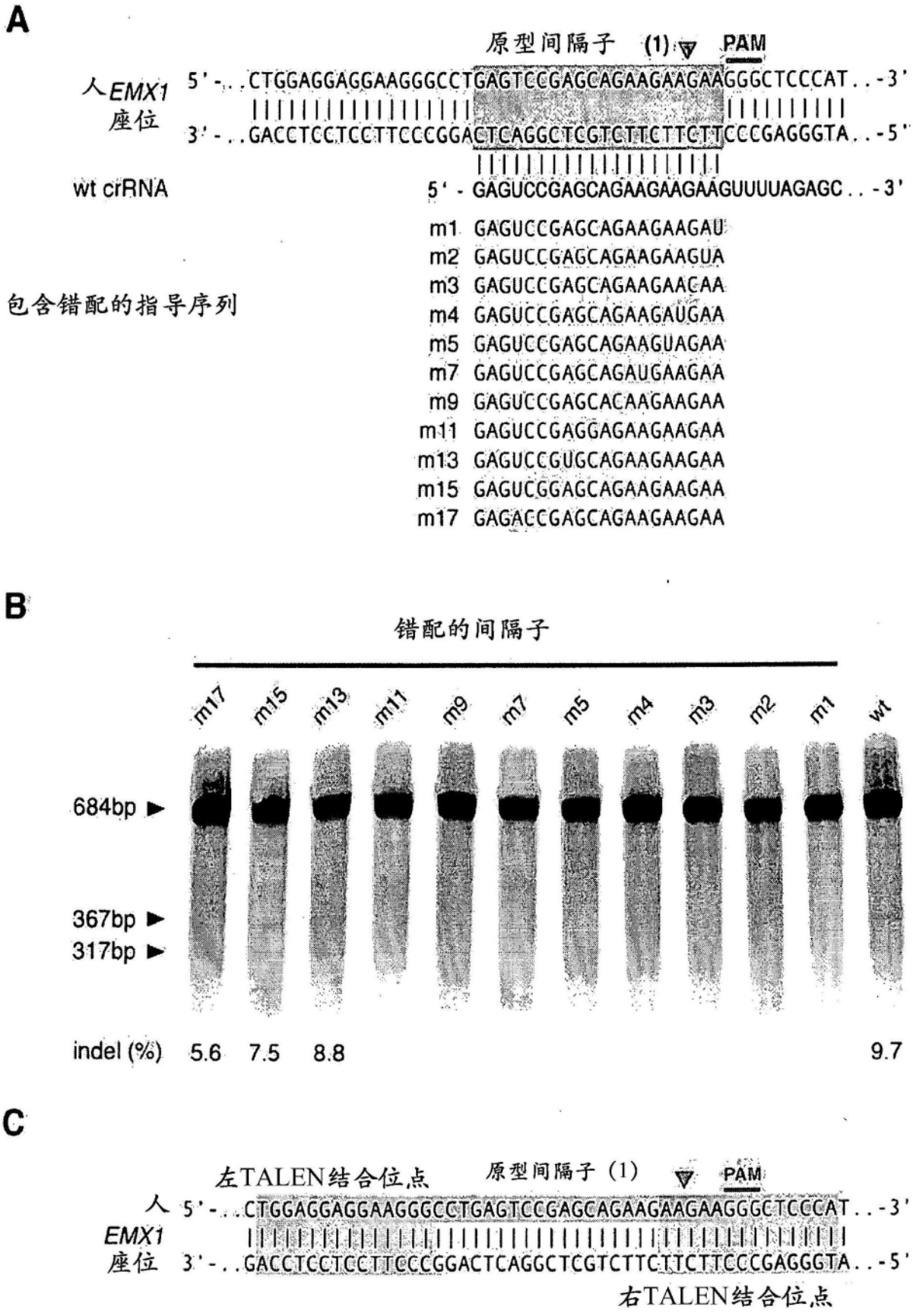

A
原型间隔子
(1)
PAM
人EMX1座位
5'-..CTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGGGCTCCCAT..-3'
3'-..GACCTCCTCCTTCCCGGACTCAGGCTCGTCTTCTTCTTCCCGAGGGTA..-5'
wt crRNA
5'-GAGUCCGAGCAGAAGAAGAAGUUUUAGAGC..-3'
包含错配的指导序列
m1 GAGUCCGAGCAGAAGAAGAU
m2 GAGUCCGAGCAGAAGAAGUA
m3 GAGUCCGAGCAGAAGAACAA
m4 GAGUCCGAGCAGAAGAUGAA
m5 GAGUCCGAGCAGAAGUAGAA
m7 GAGUCCGAGCAGAUGAAGAA
m9 GAGUCCGAGCACAAGAAGAA
m11 GAGUCCGAGGAGAAGAAGAA
m13 GAGUCCGUGCAGAAGAAGAA
m15 GAGUCGGAGCAGAAGAAGAA
m17 GAGACCGAGCAGAAGAAGAA
B
错配的间隔子
m17
m15
m13
m11
m9
m7
m5
m4
m3
m2
m1
wt
684bp
367bp
317bp
indel (%)
5.6
7.5
8.8
9.7
C
左TALEN结合位点
原型间隔子 (1)
PAM
人EMX1座位
5'-..CTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGGGCTCCCAT..-3'
3'-..GACCTCCTCCTTCCCGGACTCAGGCTCGTCTTCTTCTTCCCGAGGGTA..-5'
右TALEN结合位点

图4A图4B图4C

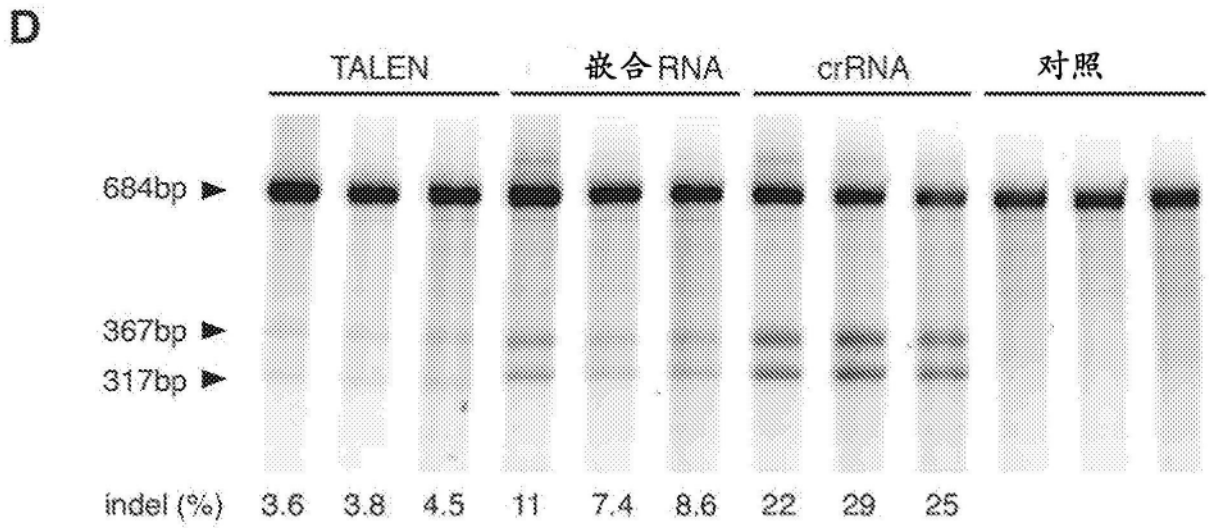

D
TALEN
嵌合RNA
crRNA
对照
684bp
367bp
317bp
indel (%)
3.6
3.8
4.5
11
7.4
8.6
22
29
25

图4D

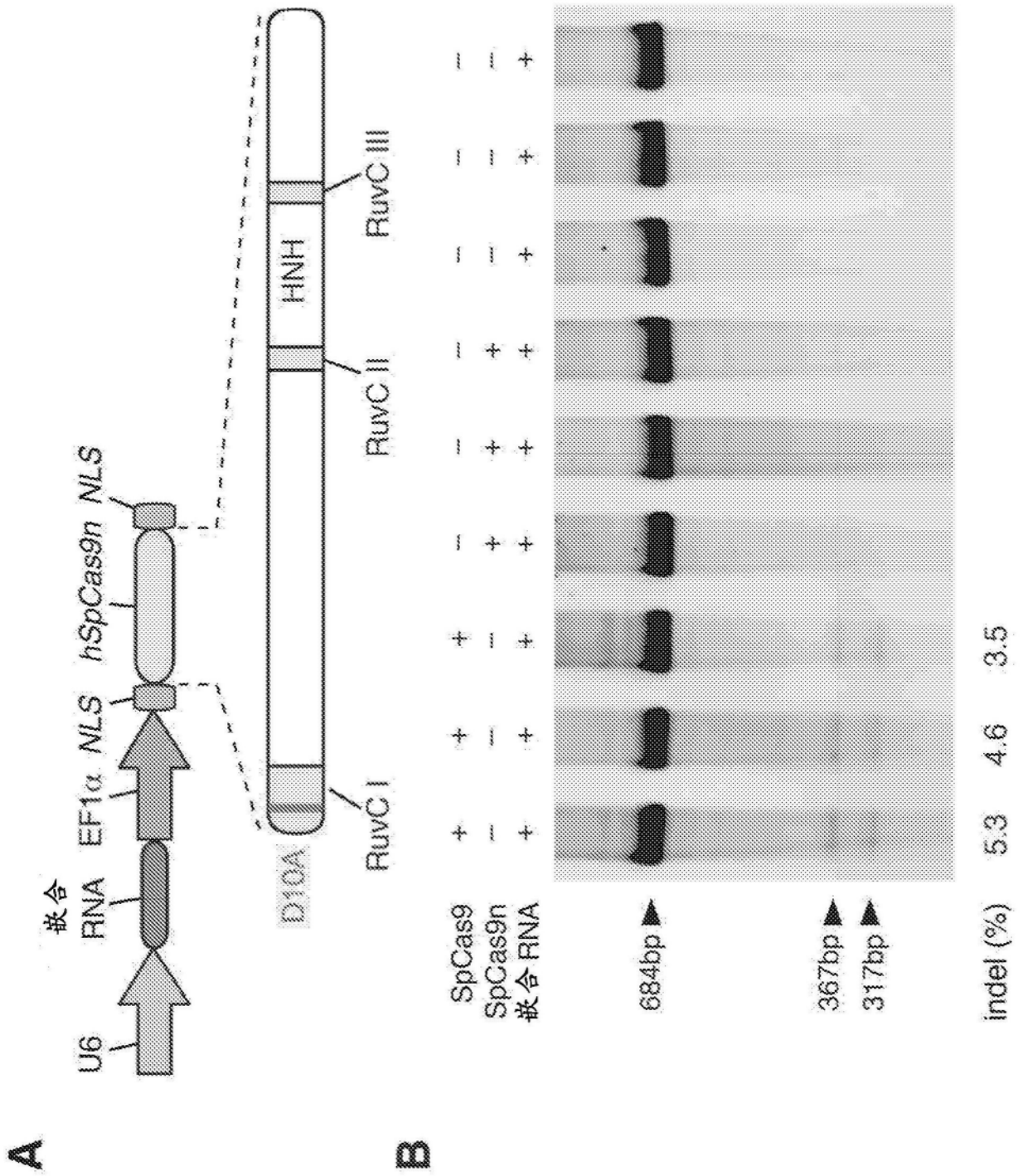

A
U6
嵌合 RNA
EF1α
NLS
hSpCas9n
NLS
D10A
RuvC I
RuvC II
HNH
RuvC III
B
SpCas9
SpCas9n
嵌合 RNA
684bp
367bp
317bp
indel (%)
5.3
4.6
3.5

图5A图5B

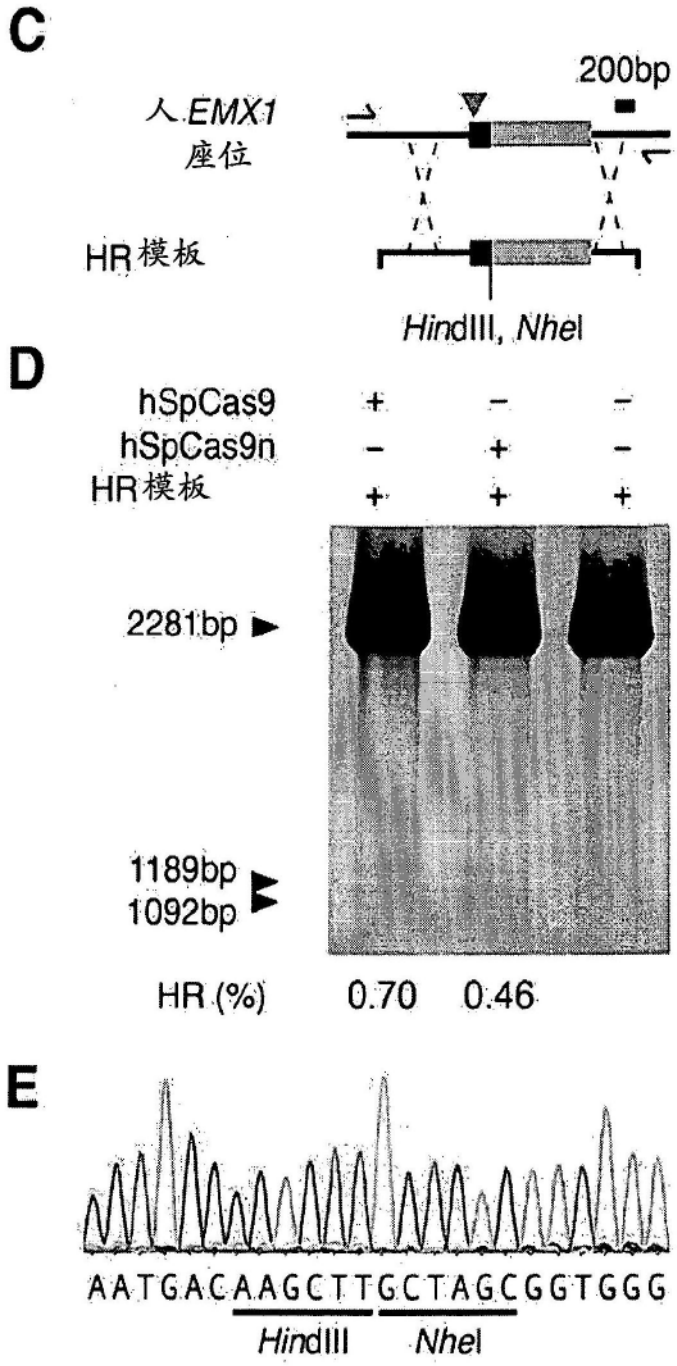

C
200bp
人 EMX1
座位
HR 模板
HindIII, NheI
D
hSpCas9 + − −
hSpCas9n − + −
HR 模板 + + +
2281bp
1189bp
1092bp
HR (%) 0.70 0.46
E
AATGACAAGCTTGCTAGCGGTGGG
HindIII
NheI

图5C图5D图5E

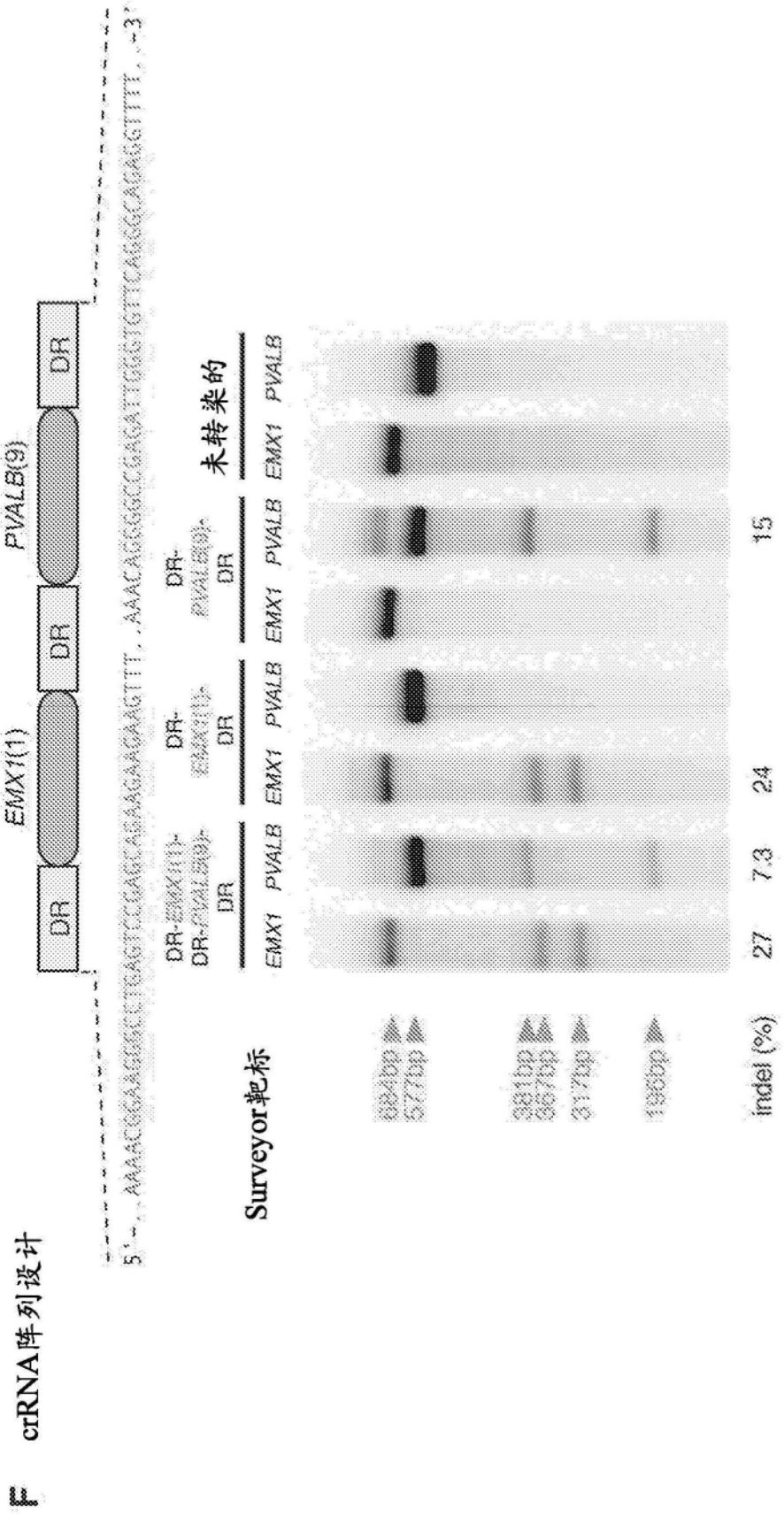

F
crRNA阵列设计
EMX1(1)
PVALB(9)
DR
DR
DR
DR
5'-...AAAACGGAAGGGGCTTGAGTCCCAGCAGAAGAAGAAGAAGTTT...AAACAGGGGCCGAGATTGGGTGTTCAGGGCAGAAGTTTT...-3'
Surveyor靶标
DR-EMX1(1)-DR-PVALB(9)-DR
DR-EMX1(1)-DR
DR-PVALB(9)-DR
未转染的
EMX1
PVALB
EMX1
PVALB
EMX1
PVALB
EMX1
PVALB
884bp
577bp
381bp
667bp
317bp
186bp
indel (%)
27
7.3
24
15

图5F

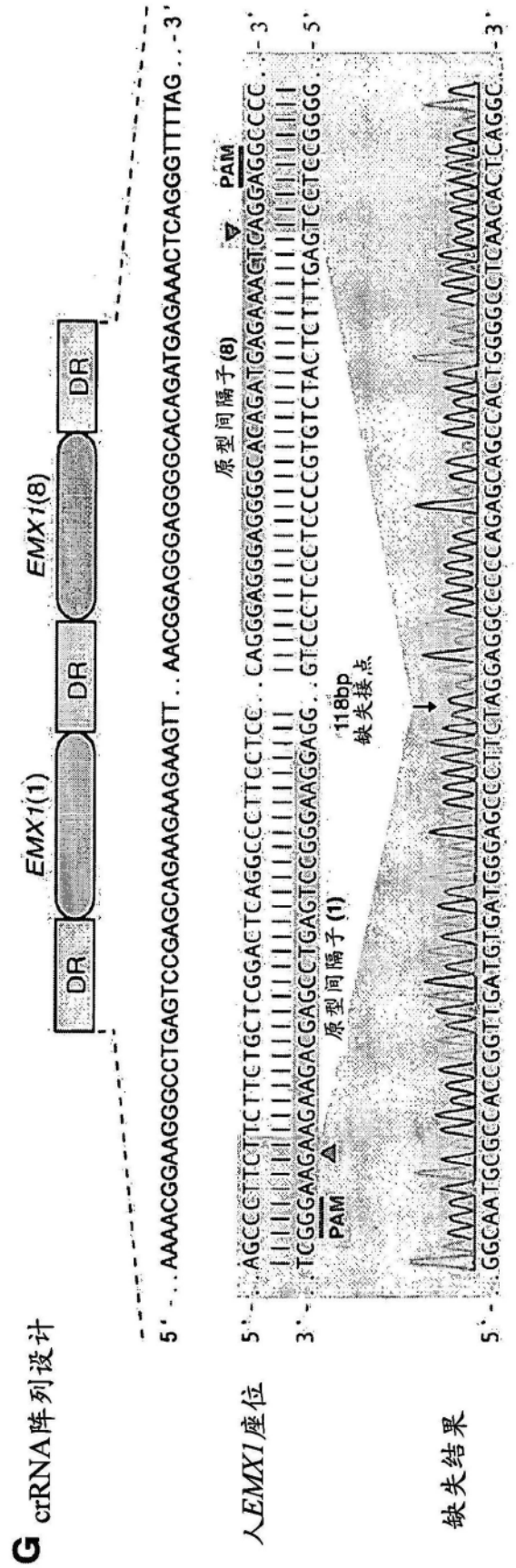
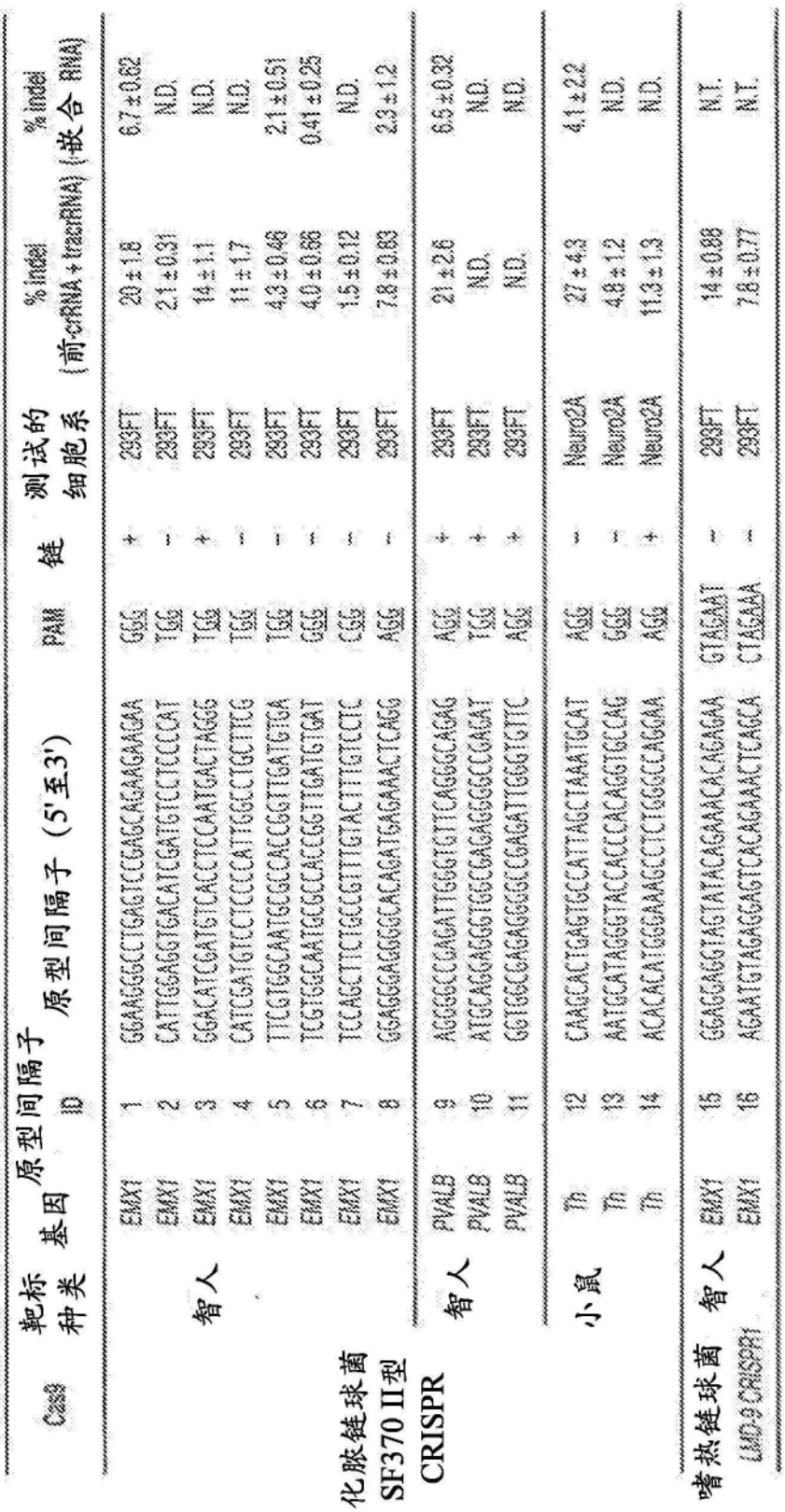

G
crRNA阵列设计
EMX1(1)
EMX1(8)
DR
DR
DR
5' - . . AAAACGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGTT . . AACGGAGGGAGGGGCACAGATGAGAAACTCAGGGTTTTAG . . - 3'
人EMX1座位
原型间隔子(8)
PAM
5' - AGCCCTTCTTCTTCTGCTCGGACTCAGGCCCTTCCTCC . . CAGGGAGGGAGGGGCACAGATGAGAAACTCAGGAGGCCCC - 3'
3' - TCGGGAAGAAGAAGACGAGCCTGAGTCCGGGAAGGAGG . . GTCCCTCCCTCCCCGTGTCTACTCTTTGAGTCCTCCGGGG - 5'
PAM
原型间隔子(1)
118bp
缺失接点
缺失结果
5' - GGCAATGCGCCACCGGTTGATGTGATGGGAGCCCTTCTAGGAGGCCCCCAGAGCAGCCACTGGGGCCTCAACACTCAGGC - 3'

图5G



Cas9	靶标种类	基因	原型间隔子ID	原型间隔子（5'至3'）	PAM	链	测试的细胞系	%indel（前-crRNA+tracrRNA）	%indel（嵌合RNA）
化脓链球菌SF370 II型CRISPR	智人	*EMX1*	1	GGAAGGCCTGAGTCCGAGCAGAAGAAGAA	GGG	+	293FT	20±1.8	6.7±0.62
		EMX1	2	CATTGGAGGTGACATCGATGTCCTTCCCAT	TGG	-	293FT	2.1±0.31	N.D.
		EMX1	3	GGACATCGATGTCACCTCCAATGACTAGGG	TGG	+	293FT	14±1.1	N.D.
		EMX1	4	CATCGATGTCCTCCCCATTGGCCTGCTTCG	TGG	-	293FT	11±1.7	N.D.
		EMX1	5	TTCGTGGCAATGGCCACCGGTTGATGTGA	TGG	-	293FT	4.3±0.48	2.1±0.51
		EMX1	6	TCGTGCCAATGGCCACCGGTTGATGTGAT	GGG	-	293FT	4.0±0.86	0.41±0.25
		EMX1	7	TCCAGCTTCTGCCGTTTGTACTTGTCCTC	CGG	-	293FT	1.5±0.12	N.D.
		EMX1	8	GGAGGAGGGCACAGATGAGAAACTCAGG	AGG	-	293FT	7.8±0.83	2.3±1.2
	智人	*PVALB*	9	AGGGCCCAGATTGGGTGTTCAGGCAGAG	AGG	+	293FT	21±2.6	6.5±0.32
		PVALB	10	ATGCAGGAGGTGGCGACAGGGGCCGAGAT	TGG	+	293FT	N.D.	N.D.
		PVALB	11	GGTGGCCAGAGGGGCCGAGATTGGGTGTTC	AGG	+	293FT	N.D.	N.D.
	小鼠	*Th*	12	CAAGCACTGAGTGCCATTAGCTAAATGCAT	AGG	-	Neuro2A	27±4.3	4.1±2.2
		Th	13	AATGCATAGGGTACCACCCACAGGTGCCAG	GGG	-	Neuro2A	4.8±1.2	N.D.
		Th	14	ACACAATGGGAAAGCCCTCTGGGCCAGGAA	AGG	+	Neuro2A	11.3±1.3	N.D.
嗜热链球菌*LMD-9 CRISPR1*	智人	*EMX1*	15	GGAGGAGGTAGTATACAGAAACACAGAGAA	GTAGAAT	-	293FT	14±0.88	N.T.
		EMX1	16	AGAATGTAGAGGAGTCACAGAAACTCAGCA	CTAGAAA	-	293FT	7.8±0.77	N.T.

图6

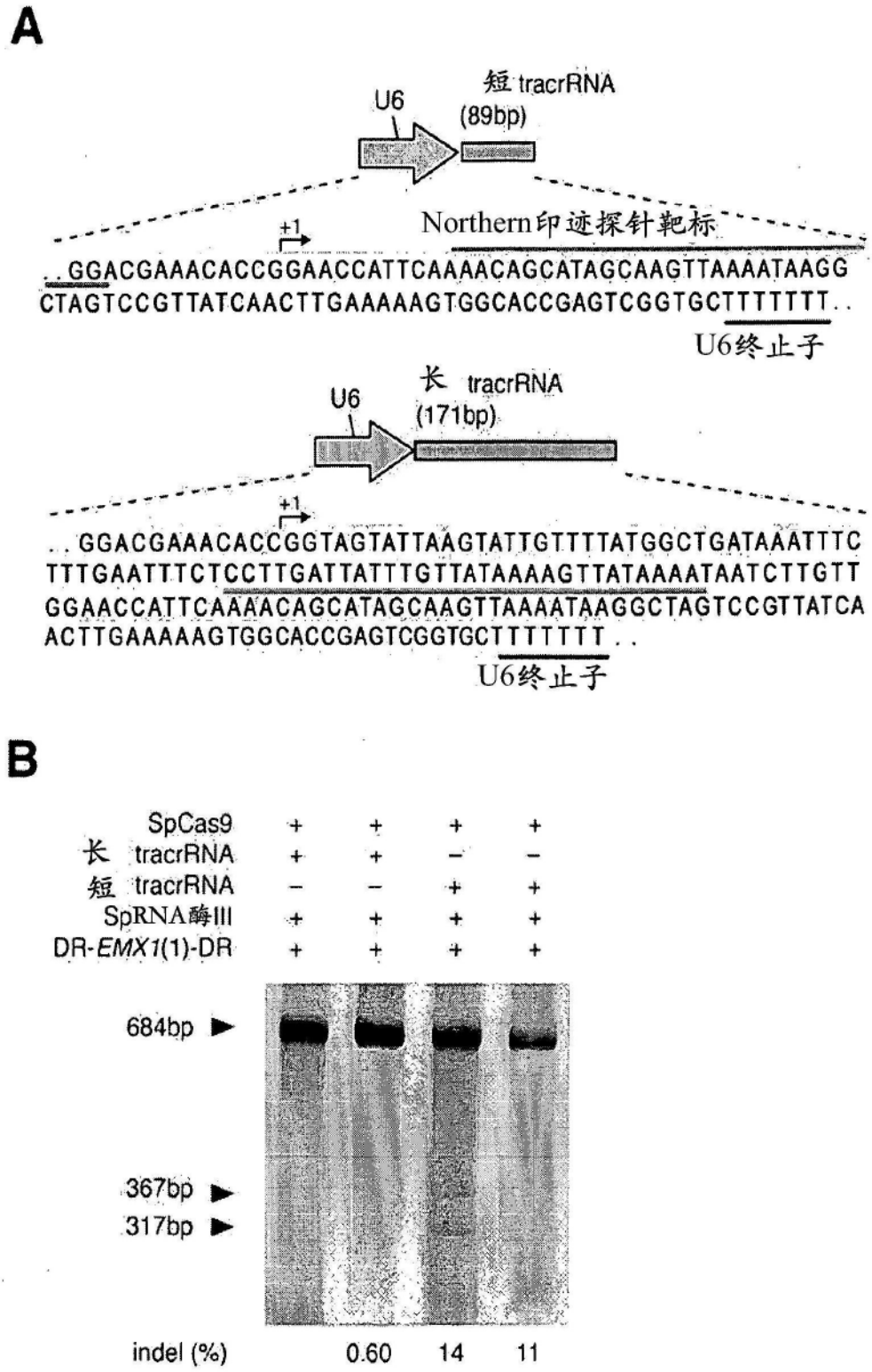

A
U6
短 tracrRNA
(89bp)
+1
Northern印迹探针靶标
..GGACGAAACACCGGAACCATTCAAAACAGCATAGCAAGTTAAAATAAGG
CTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT..
U6终止子
U6
长 tracrRNA
(171bp)
+1
..GGACGAAACACCGGTAGTATTAAGTATTGTTTTATGGCTGATAAATTTC
TTTGAATTTCTCCTTGATTATTTGTTATAAAAGTTATAAAATAATCTTGTT
GGAACCATTCAAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCA
ACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT..
U6终止子
B
SpCas9 + + + +
长 tracrRNA + + − −
短 tracrRNA − − + +
SpRNA酶III + + + +
DR-EMX1(1)-DR + + + +
684bp
367bp
317bp
indel (%) 0.60 14 11

图7A图7B

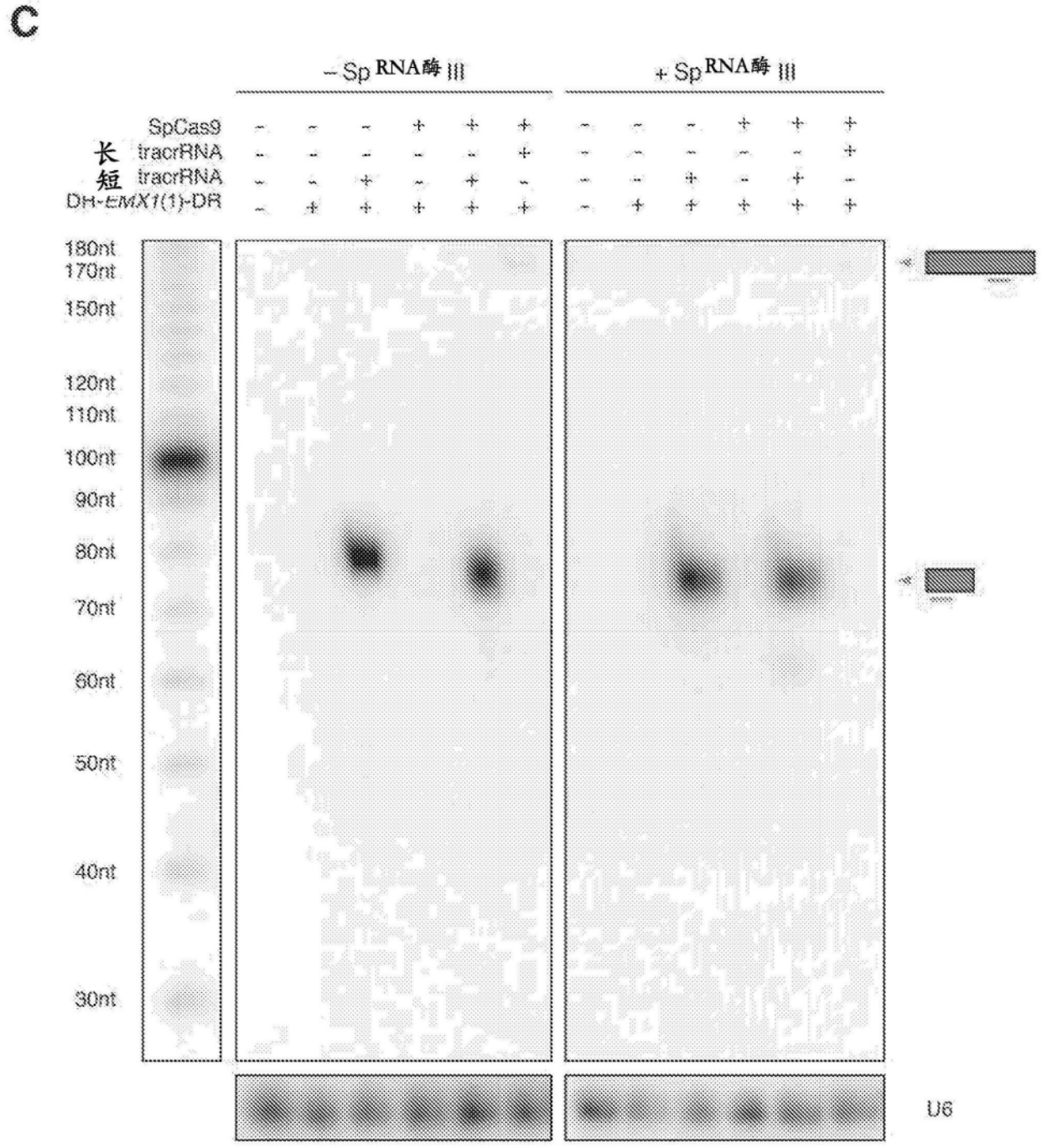

C
– Sp RNA酶 III
+ Sp RNA酶 III
SpCas9
长 tracrRNA
短 tracrRNA
DR-EMX1(1)-DR
180nt
170nt
150nt
120nt
110nt
100nt
90nt
80nt
70nt
60nt
50nt
40nt
30nt
U6

图7C

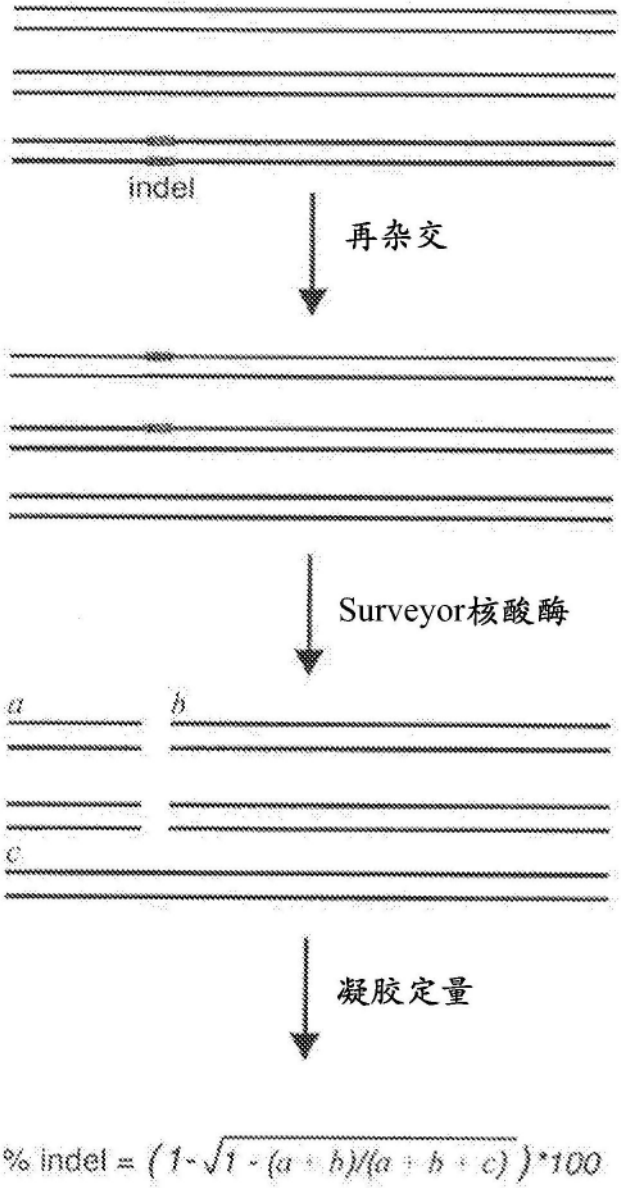


图8

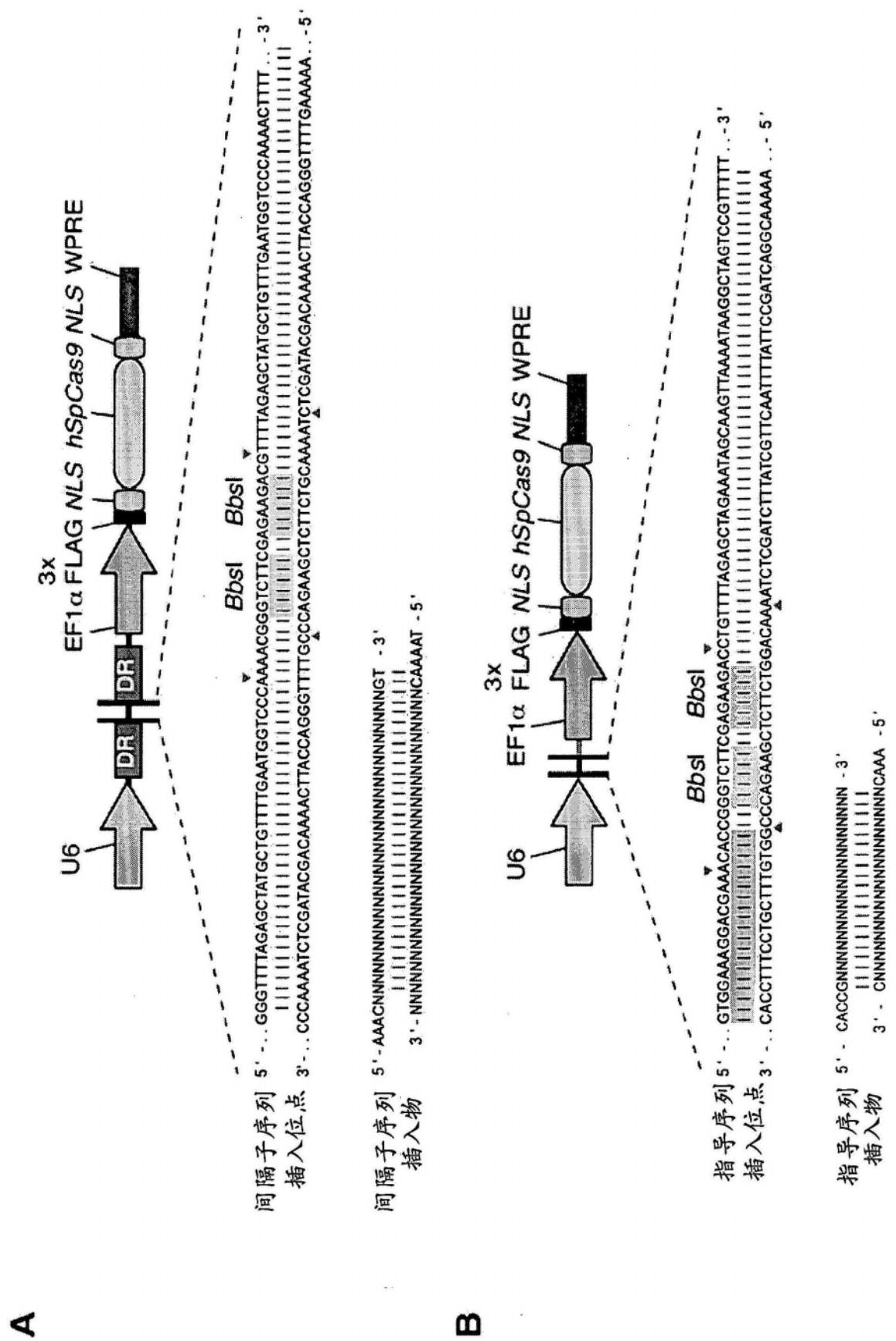

A
U6
DR
DR
3x
EF1α FLAG NLS hSpCas9 NLS WPRE
BbsI
BbsI
间隔子序列 5'-...GGGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACGGGTCTTCGAGAAGACGTTTTAGAGCTATGCTGTTTGAATGGTCCCAAAACTTTT...-3'
插入位点 3'-...CCCAAAATCTCGATACGACAAAACTTACCAGGGTTTTGCCCAGAAGCTCTTCTGCAAAATCTCGATACGACAAAACTTACCAGGGTTTTGAAAAA...-5'
间隔子序列 5'-AAACNNNNNNNNNNNNNNNNNNNNNNNNNNNNGT-3'
插入物 3'-NNNNNNNNNNNNNNNNNNNNNNNNNNNNCAAAT-5'
B
U6
3x
EF1α FLAG NLS hSpCas9 NLS WPRE
BbsI
BbsI
指导序列 5'-...GTGGAAAGGACGAAACACCGGGTCTTCGAGAAGACCTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTTTT...-3'
插入位点 3'-...CACCTTTCCTGCTTTGTGGCCCAGAAGCTCTTCTGGACAAAATCTCGATCTTTATCGTTCAATTTTATTCCGATCAGGCAAAAA...-5'
指导序列 5'-CACCGNNNNNNNNNNNNNNNNNNN-3'
插入物 3'-CNNNNNNNNNNNNNNNNNNNCAAA-5'

图9A图9B

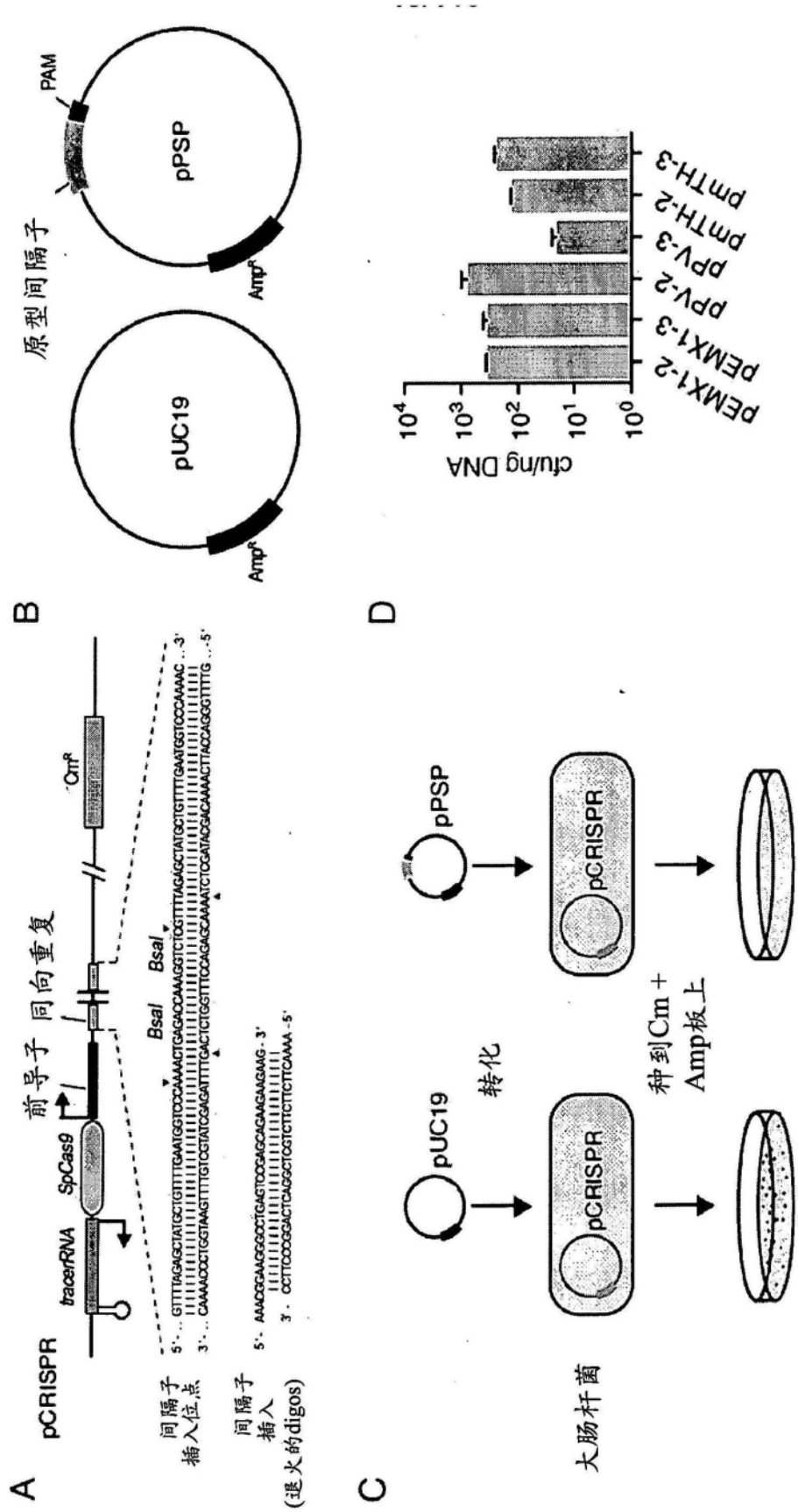
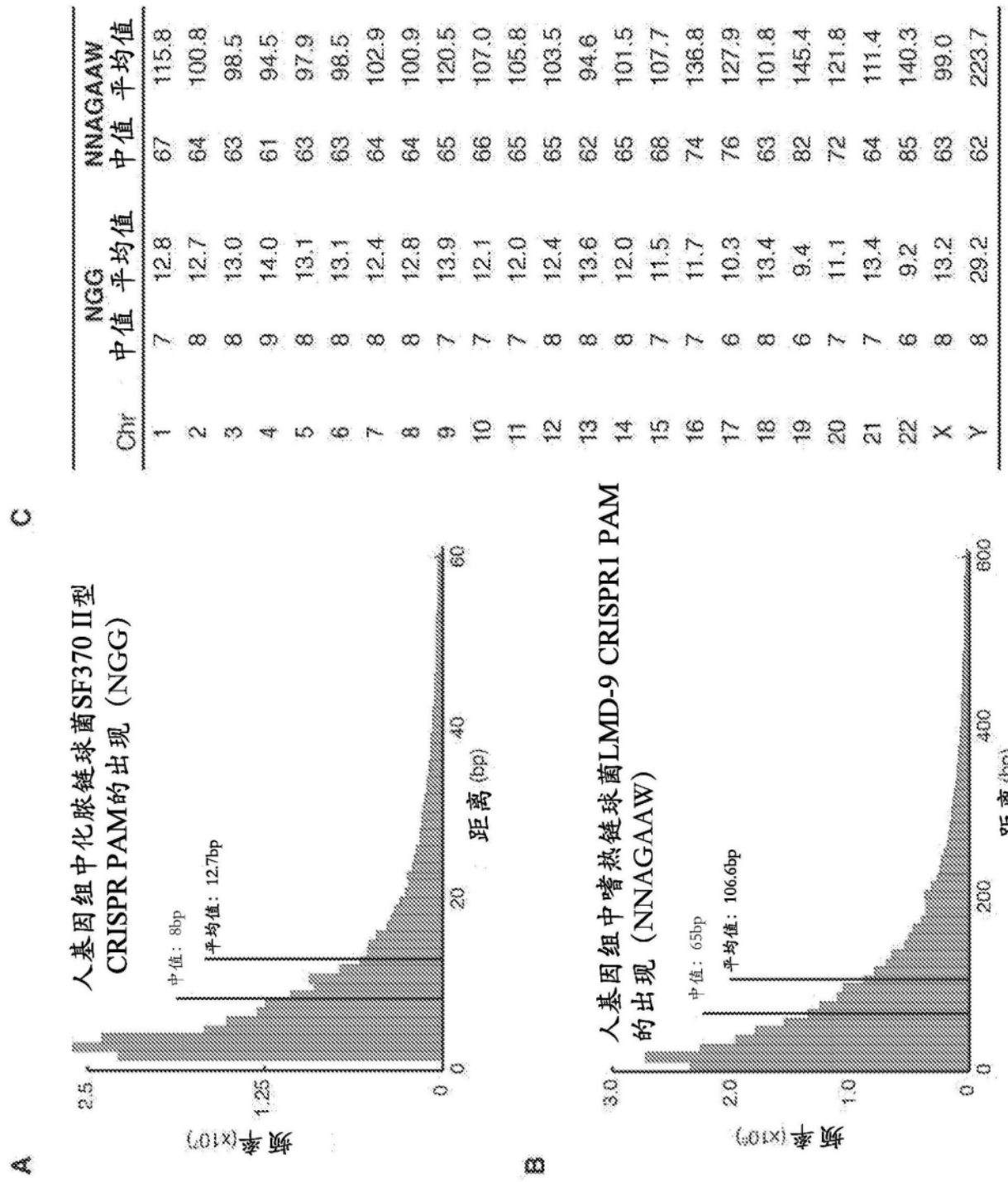

A
pCRISPR
tracrRNA
SpCas9
前导子
同向重复
Cm^R
BsaI
BsaI
间隔子
插入位点
间隔子
插入
(退火的digos)
B
原型间隔子
PAM
pPSP
pUC19
Amp^R
Amp^R
C
pUC19
转化
pCRISPR
大肠杆菌
pPSP
pCRISPR
种到Cm +
Amp板上
D
cfu/ng DNA
10^4
10^3
10^2
10^1
10^0
pEMX1-2
pEMX1-3
pPV-2
pPV-3
pMTH-2
pMTH-3

图10



C

Chr	NGG 中值	NGG 平均值	NNAGAAW 中值	NNAGAAW 平均值
1	7	12.8	67	115.8
2	8	12.7	64	100.8
3	8	13.0	63	98.5
4	9	14.0	61	94.5
5	8	13.1	63	97.9
6	8	13.1	63	98.5
7	8	12.4	64	102.9
8	8	12.8	64	100.9
9	7	13.9	65	120.5
10	7	12.1	66	107.0
11	7	12.0	65	105.8
12	8	12.4	65	103.5
13	8	13.6	62	94.6
14	8	12.0	65	101.5
15	7	11.5	68	107.7
16	7	11.7	74	136.8
17	6	10.3	76	127.9
18	8	13.4	63	101.8
19	6	9.4	82	145.4
20	7	11.1	72	121.8
21	7	13.4	64	111.4
22	6	9.2	85	140.3
X	8	13.2	63	99.0
Y	8	29.2	62	223.7

图11A图11B图11C

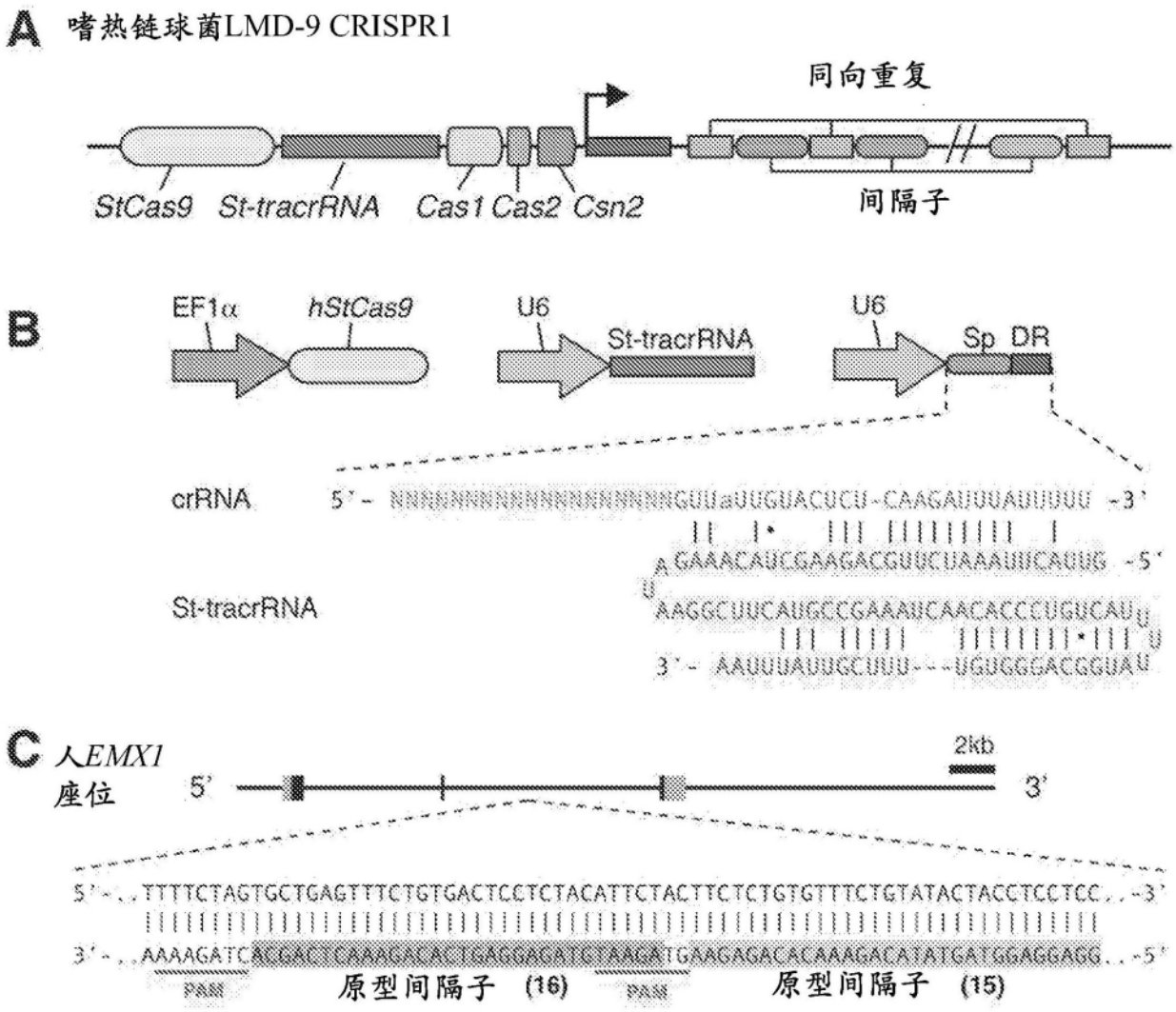


图12A图12B图12C

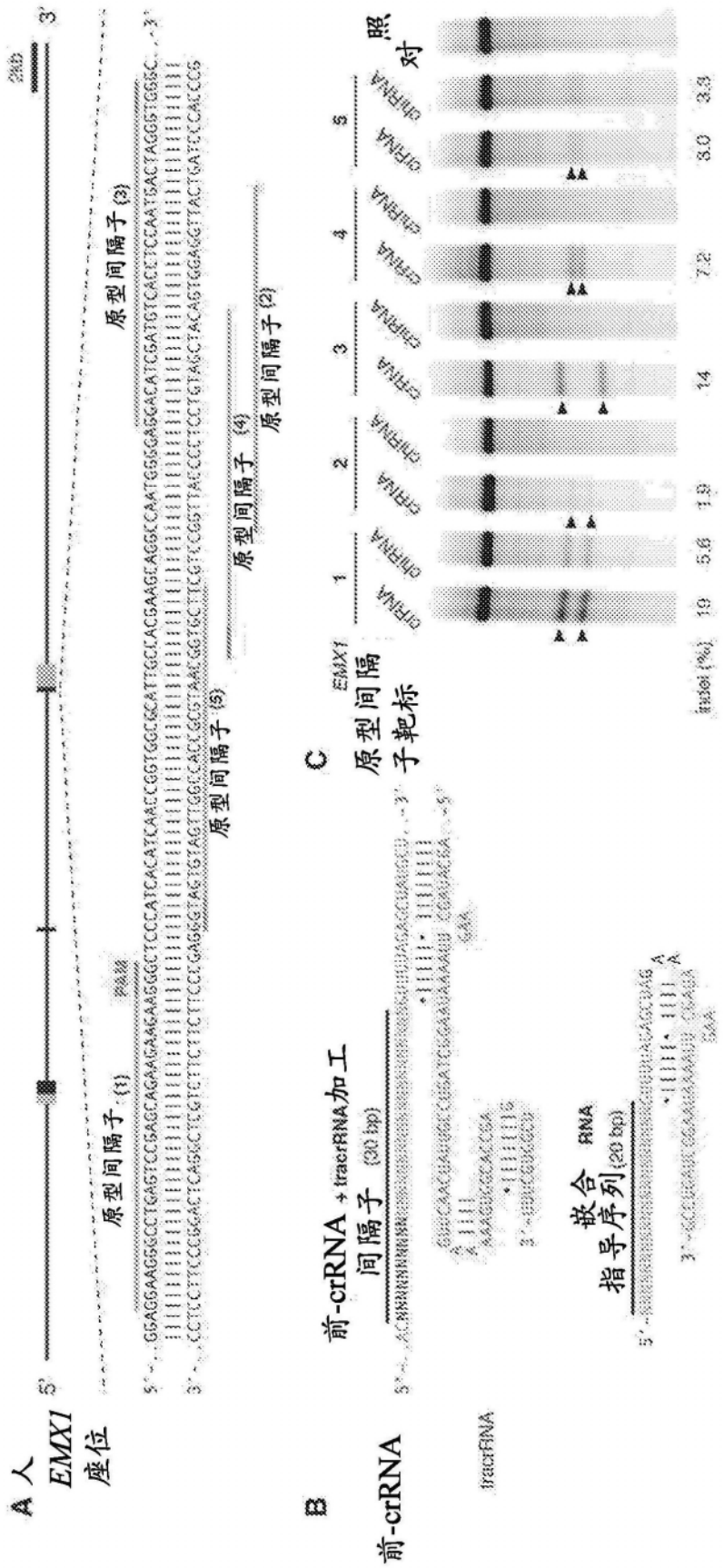


图13A图13B图13C

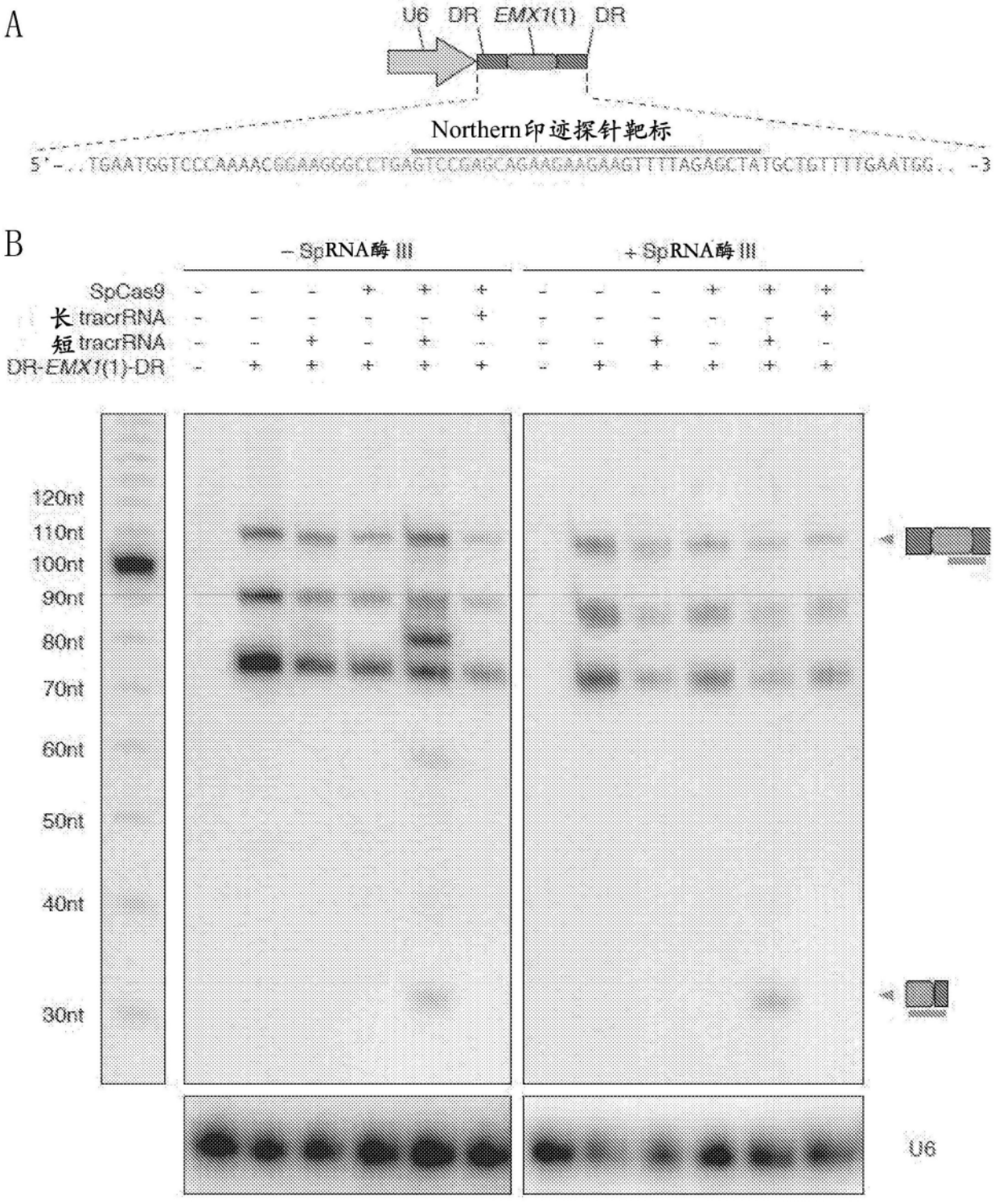

A
U6 DR EMX1(1) DR
Northern印迹探针靶标
5'-...TGAATGGTCCCAAAACGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGTTTTAGAGCTATGCTGTTTTGAATGG... -3
B
– SpRNA酶 III
+ SpRNA酶 III
SpCas9
长 tracrRNA
短 tracrRNA
DR-EMX1(1)-DR
120nt
110nt
100nt
90nt
80nt
70nt
60nt
50nt
40nt
30nt
U6

图14A图14B

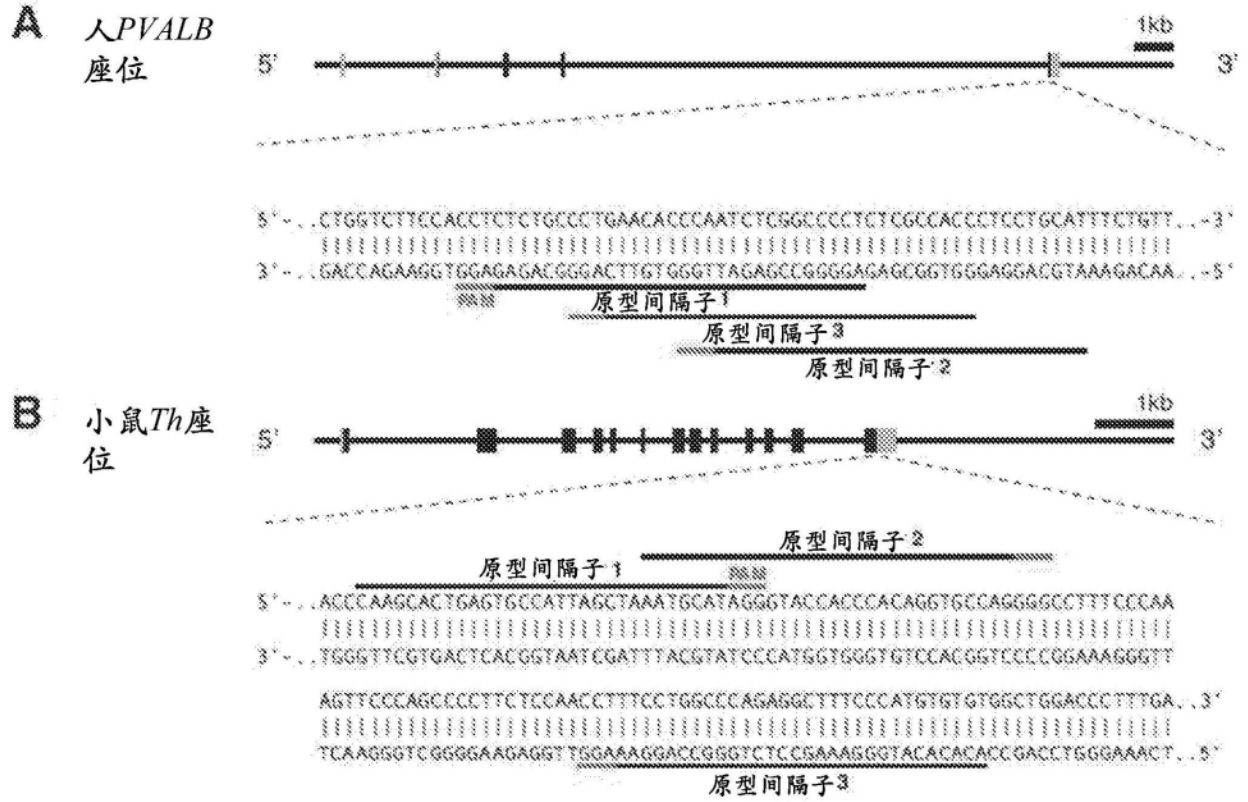

A
人PVALB
座位
1kb
5'
3'
5'-..CTGGTCTTCCACCTCTCTGCCCTGAACACCCAATCTCGGCCCCTCTCGCCACCCTCCTGCATTTCTGTT..-3'
3'-..GACCAGAAGGTGGAGAGACGGGACTTGTGGGTTAGAGCCGGGGAGAGCGGTGGGAGGACGTAAAGACAA..-5'
PAM
原型间隔子1
原型间隔子3
原型间隔子2
B
小鼠Th座
位
1kb
5'
3'
原型间隔子2
原型间隔子1
PAM
5'-..ACCCAAGCACTGAGTGCCATTAGCTAAATGCATAGGGTACCACCCACAGGTGCCAGGGGCCTTTCCCAA
3'-..TGGGTTCGTGACTCACGGTAATCGATTTACGTATCCCATGGTGGGTGTCCACGGTCCCCGGAAAGGGTT
AGTTCCCAGCCCCTTCTCCAACCTTTCCTGGCCCAGAGGCTTTCCCATGTGTGTGGCTGGACCCTTTGA..3'
TCAAGGGTCGGGGAAGAGGTTGGAAAGGACCGGGTCTCCGAAAGGGTACACACACCGACCTGGGAAACT..5'
原型间隔子3

图15

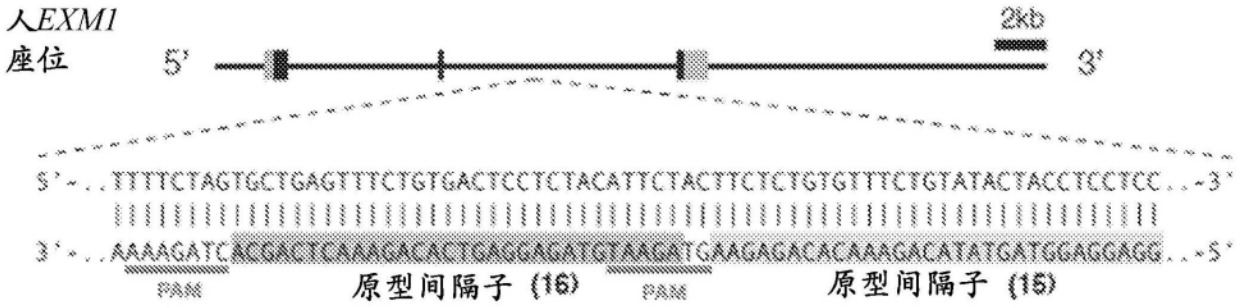

人EXM1
座位
2kb
5'
3'
5'-..TTTTCTAGTGCTGAGTTTCTGTGACTCCTCTACATTCTACTTCTCTGTGTTTCTGTATACTACCTCCTCC..-3'
3'-..AAAAGATCACGACTCAAAGACACTGAGGAGATGTAAGATGAAGAGACACAAAGACATATGATGGAGGAGG..-5'
PAM
原型间隔子 (16)
PAM
原型间隔子 (15)

图16

引物名称	测定	基因组靶标	引物序列
Sp-EMX1-F	SURVEYOR测定,测序	*EMX1*	AAAACCACCCTTCTCTCTGGC
Sp-EMX1-R	SURVEYOR测定,测序	*EMX1*	GGAGATTGGAGACACGGAGAG
Sp-PVALB-F	SURVEYOR测定,测序	*PVALB*	CTGGAAAGCCAATGCCTGAC
Sp-PVALB-R	SURVEYOR测定,测序	*PVALB*	GGCAGCAAACTCCTTGTCCT
Sp-Th-F	SURVEYOR测定,测序	*Th*	GTGCTTTGCAGAGGCCTACC
Sp-Th-R	SURVEYOR测定,测序	*Th*	CCTGGAGCGCATGCAGTAGT
St-EMX1-F	SURVEYOR测定,测序	*EMX1*	ACCTTCTGTGTTTCCACCATTC
St-EMX1-R	SURVEYOR测定,测序	*EMX1*	TTGGGGAGTGCACAGACTTC
Sp-EMX1-RFLP-F	RFLP,测序	*EMX1*	GGCTCCCTGGGTTCAAAGTA
Sp-EMX1-RFLP-R	RFLP,测序	*EMX1*	AGAGGGGTCTGGATGTCGTAA
Pb_EMX1_sp1	Northern印迹探针	不适用	TAGCTCTAAAACTTCTTCTTCTGCTCGGAC
Pb_tracrRNA	Northern印迹探针	不适用	CTAGCCTTATTTTAACTTGCTATGCTGTTT

图17

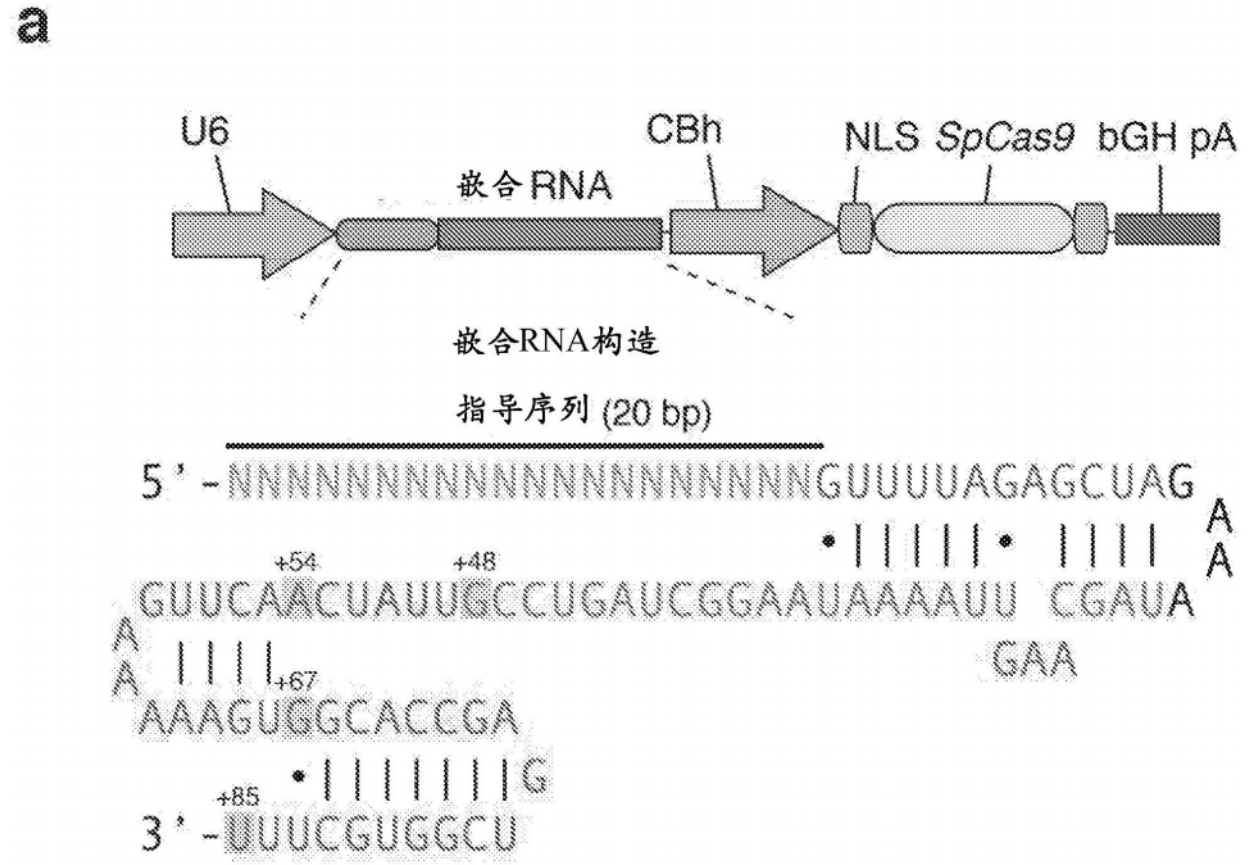

a
U6
CBh
NLS SpCas9 bGH pA
嵌合RNA
嵌合RNA构造
指导序列 (20 bp)
5'-NNNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAG
A
A
+54
+48
GUUCAACUAUUGCCUGAUCGGAAUAAAAUU CGAUA
A
A
GAA
+67
AAAGUGGCACCGA
+85
3'-UUUCGUGGCU

图18A

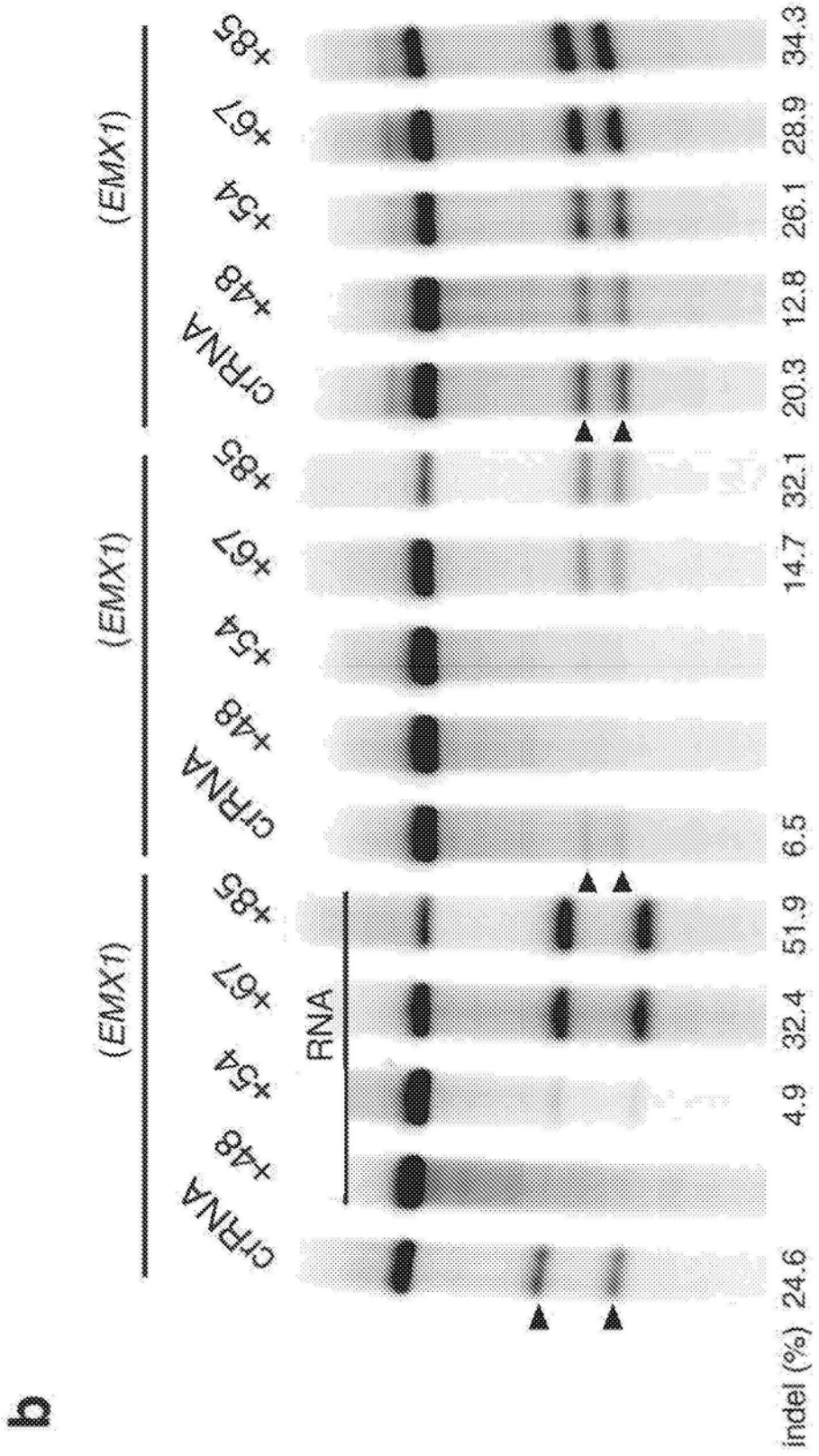

b
(EMX1)
+85
+67
+54
+48
crRNA
RNA
indel (%)
24.6
4.9
32.4
51.9
6.5
14.7
32.1
20.3
12.8
26.1
28.9
34.3

图18B

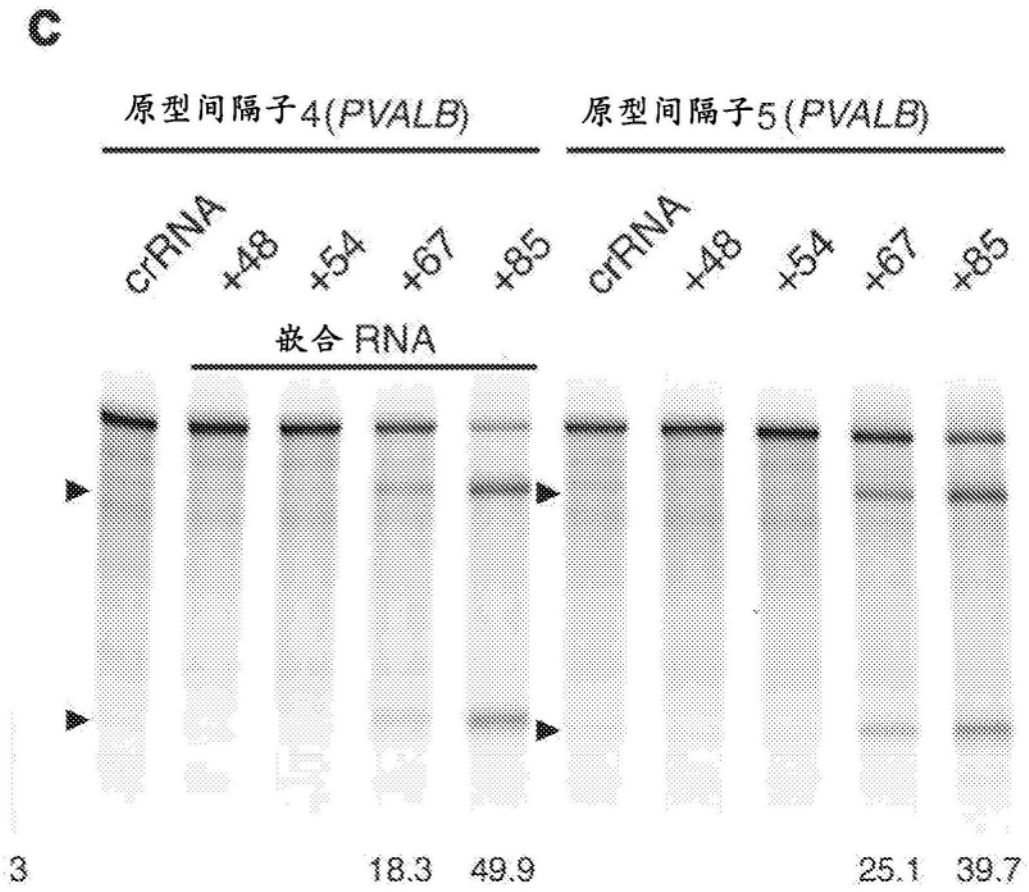

C
原型间隔子4(PVALB)
原型间隔子5(PVALB)
crRNA
+48
+54
+67
+85
crRNA
+48
+54
+67
+85
嵌合 RNA
3
18.3
49.9
25.1
39.7

图18C

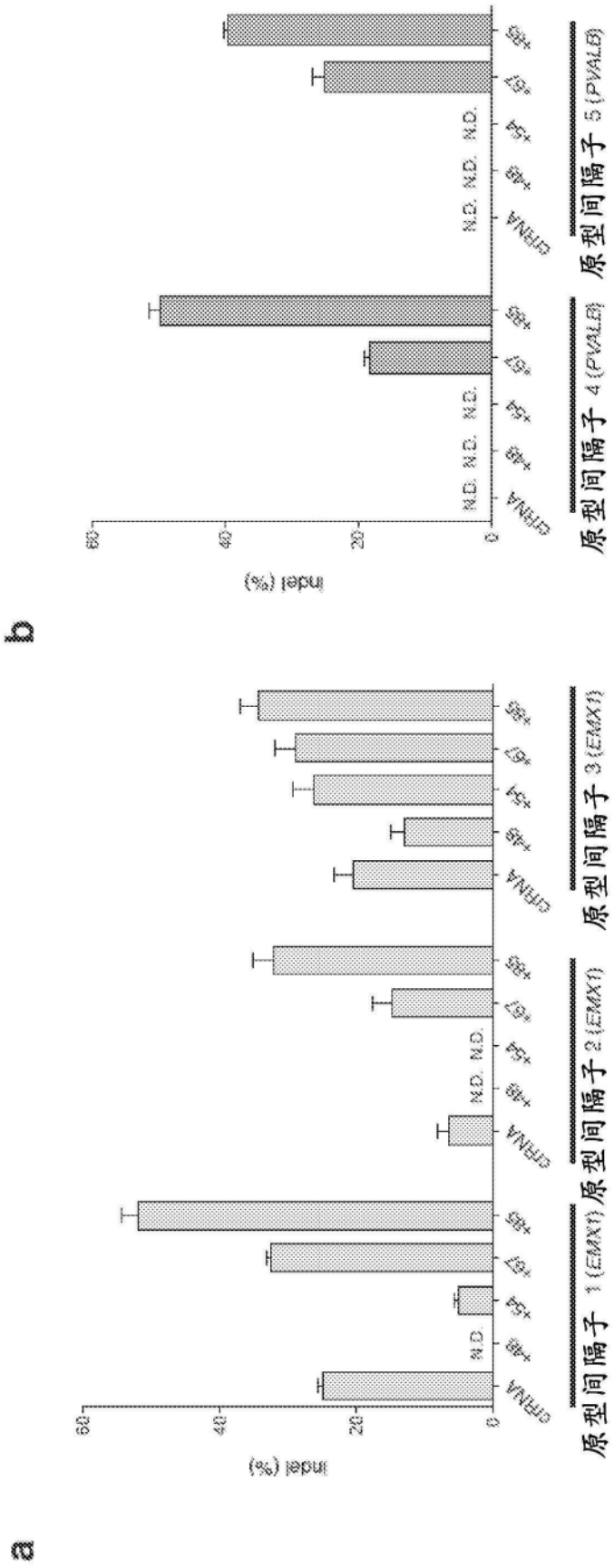

a
b
Indel (%)
0
20
40
60
crRNA
N.D.
N.D. N.D.
原型间隔子 1 (EMX1)
原型间隔子 2 (EMX1)
原型间隔子 3 (EMX1)
原型间隔子 4 (PVALB)
原型间隔子 5 (PVALB)

图19A图19B

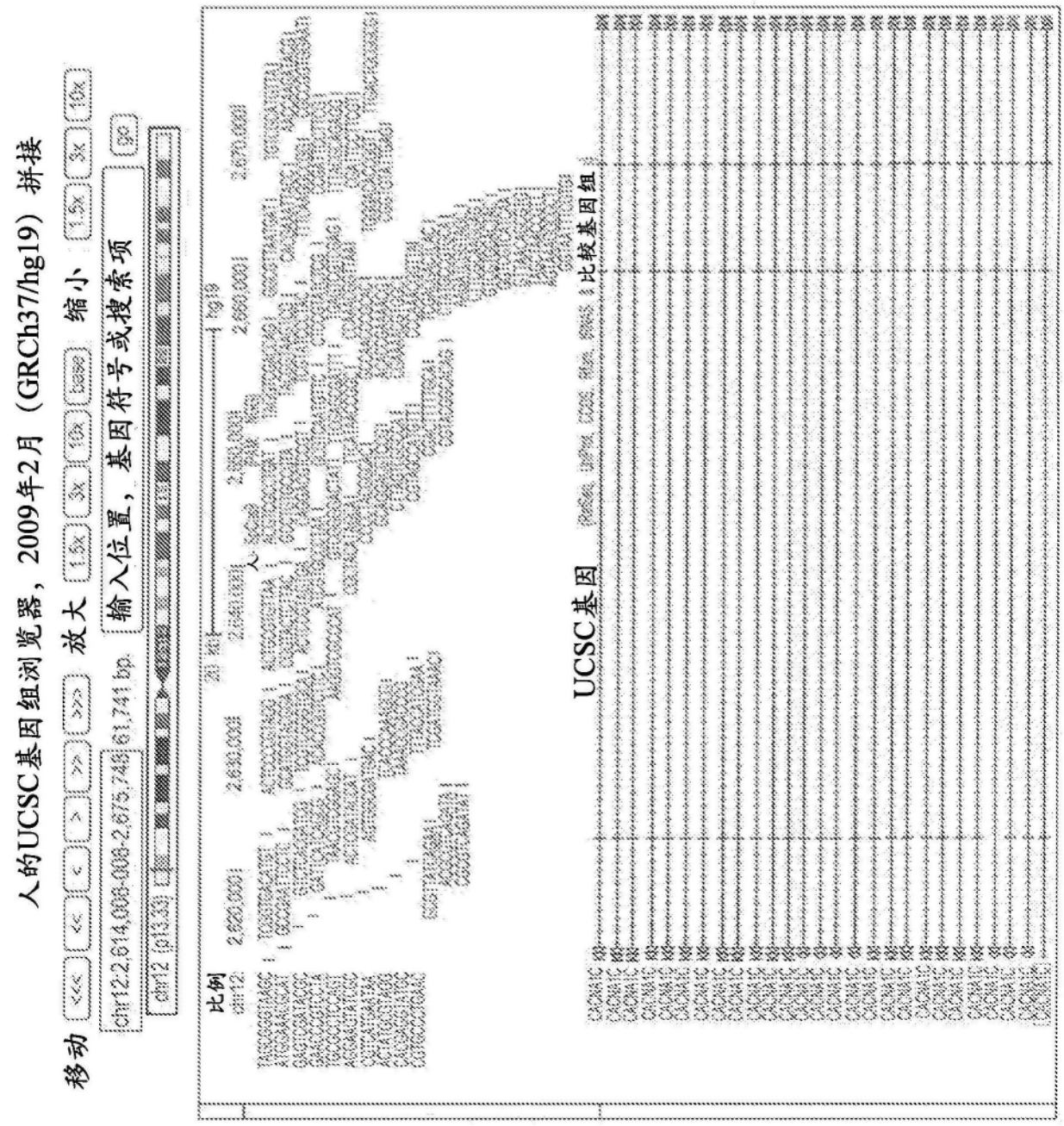

人的UCSC基因组浏览器，2009年2月（GRCh37/hg19）拼接
移动
放大
缩小
输入位置，基因符号或搜索项
chr12:2,614,003-2,675,744
61,741 bp.
chr12 (p13.33)
比例
UCSC基因
比较基因组

图20

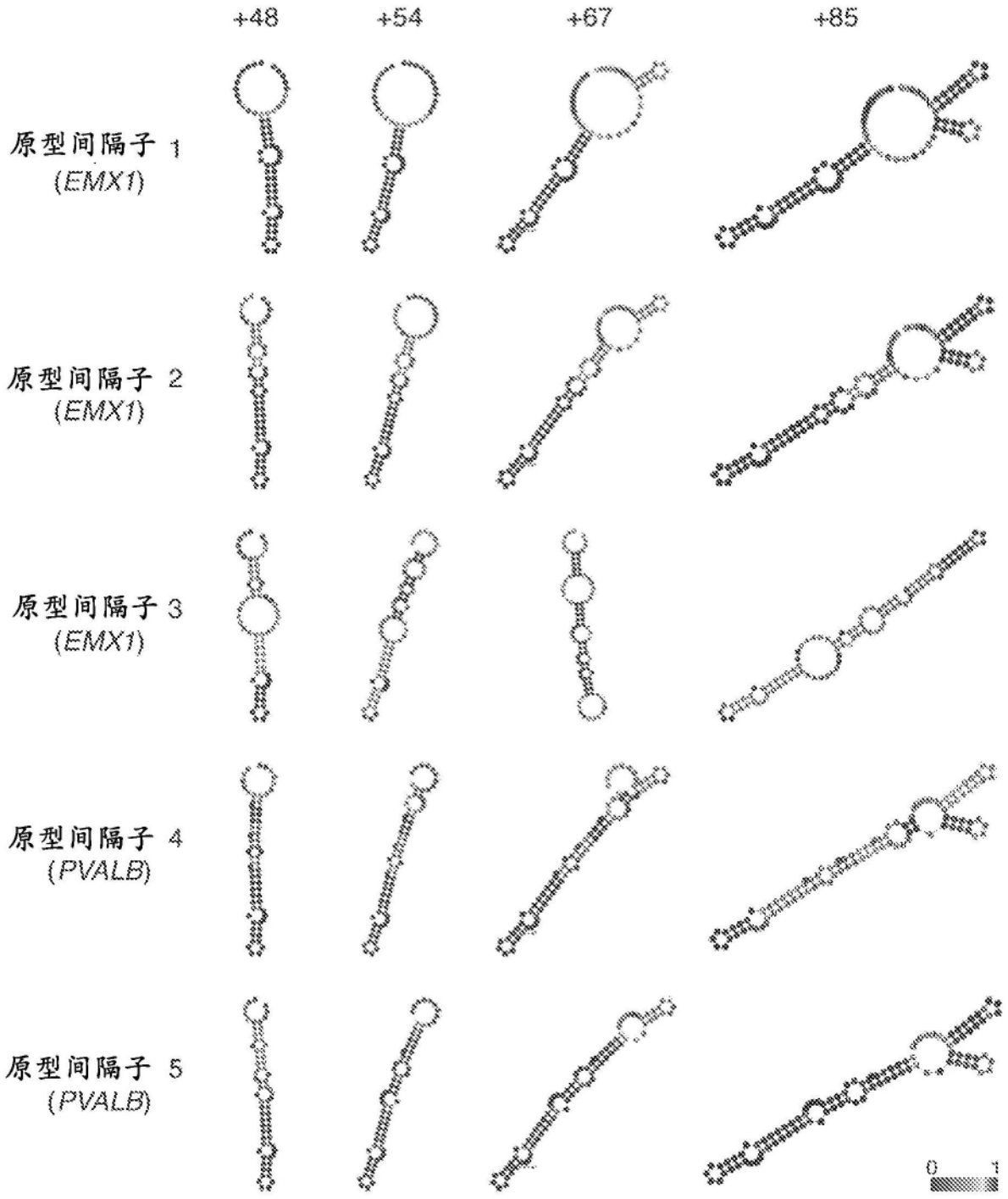

+48
+54
+67
+85
原型间隔子 1
(EMX1)
原型间隔子 2
(EMX1)
原型间隔子 3
(EMX1)
原型间隔子 4
(PVALB)
原型间隔子 5
(PVALB)
0
1

图21

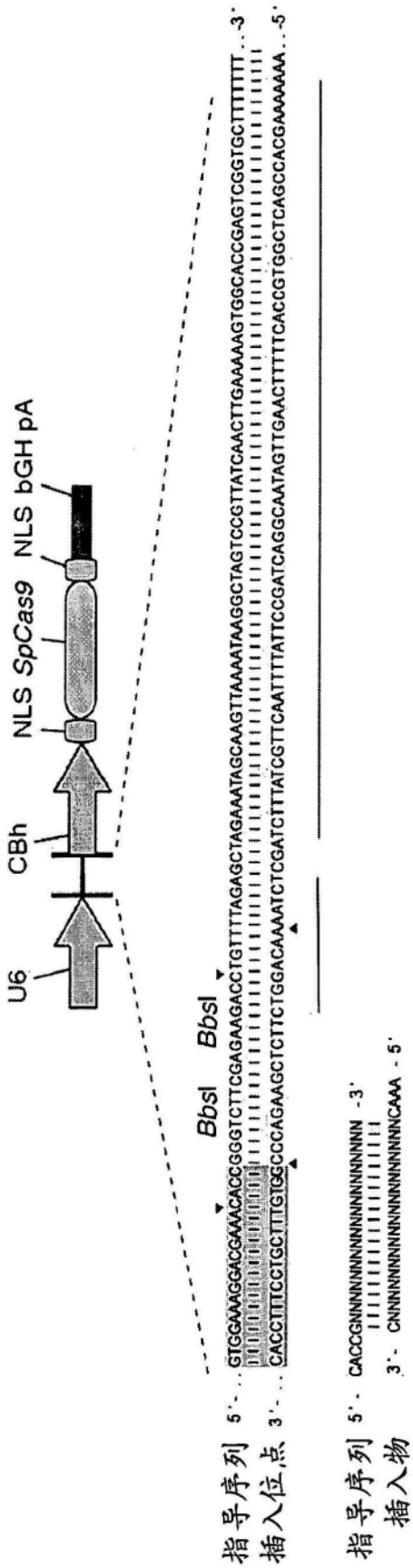
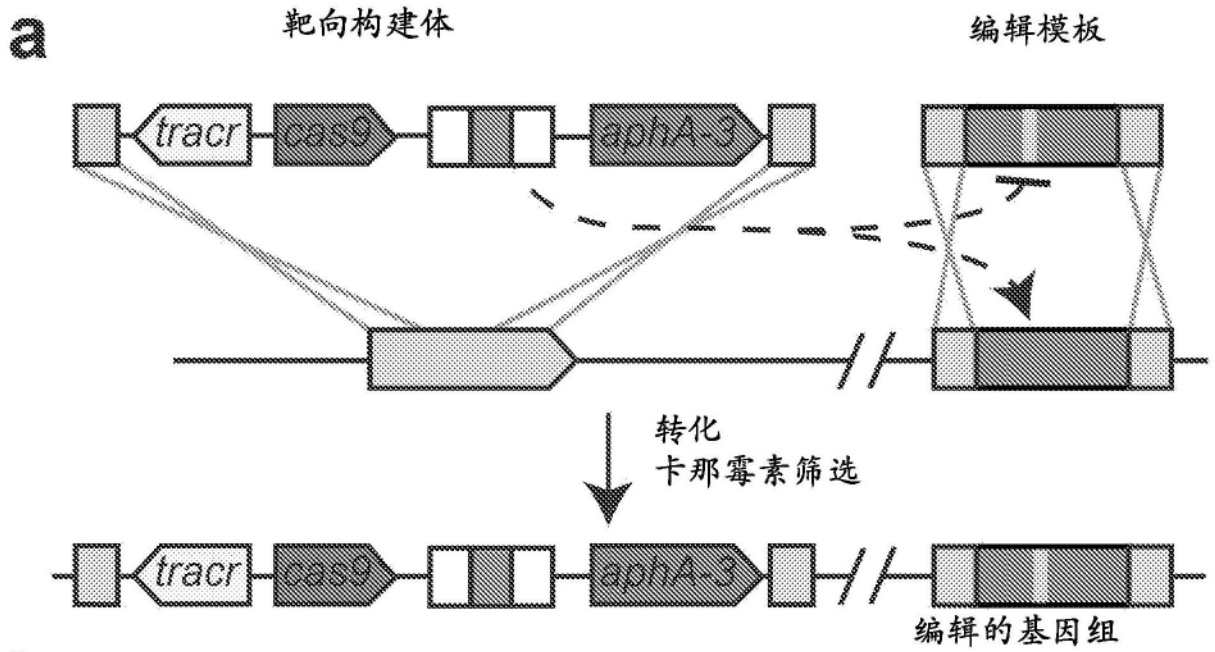

U6
CBh
NLS SpCas9 NLS bGH pA
BbsI BbsI
指导序列 5'- ...GTGGAAAGGACGAAACACCGGGTCTTCGAGAAGACCTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT...-3'
插入位点 3'- ...CACCTTTCCTGCTTTGTGGCCCAGAAGCTCTTCTGGACAAAATCTCGATCTTTATTCGTTCAATTTTATTCCGATCAGGCAATAGTTGAACTTTTTCACCGTGGCTCAGCCACGAAAAAAA...-5'
指导序列 5'- CACCGNNNNNNNNNNNNNNNNNNN -3'
插入物 3'- CNNNNNNNNNNNNNNNNNNNCAAA - 5'

图22



b

编辑构建体	cfu/ug of crR6M DNA
无	2.6 ± 0.9
R6(*srtA*)	14 ± 1.8
$R6^{370.1}$	25 ± 2.6

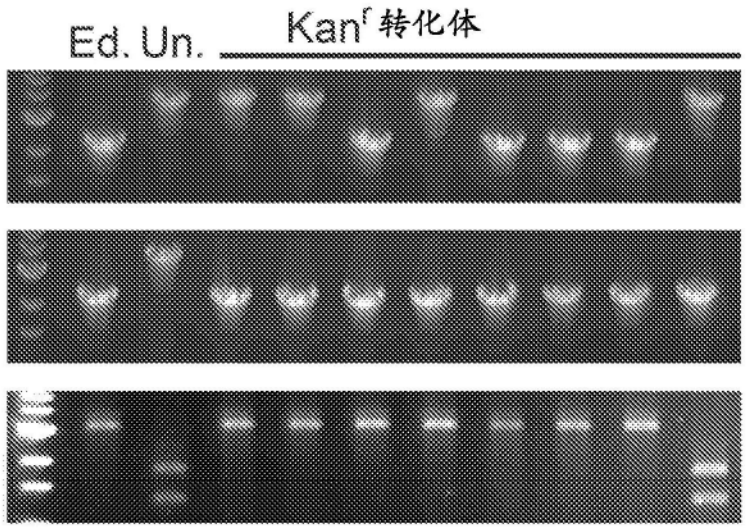
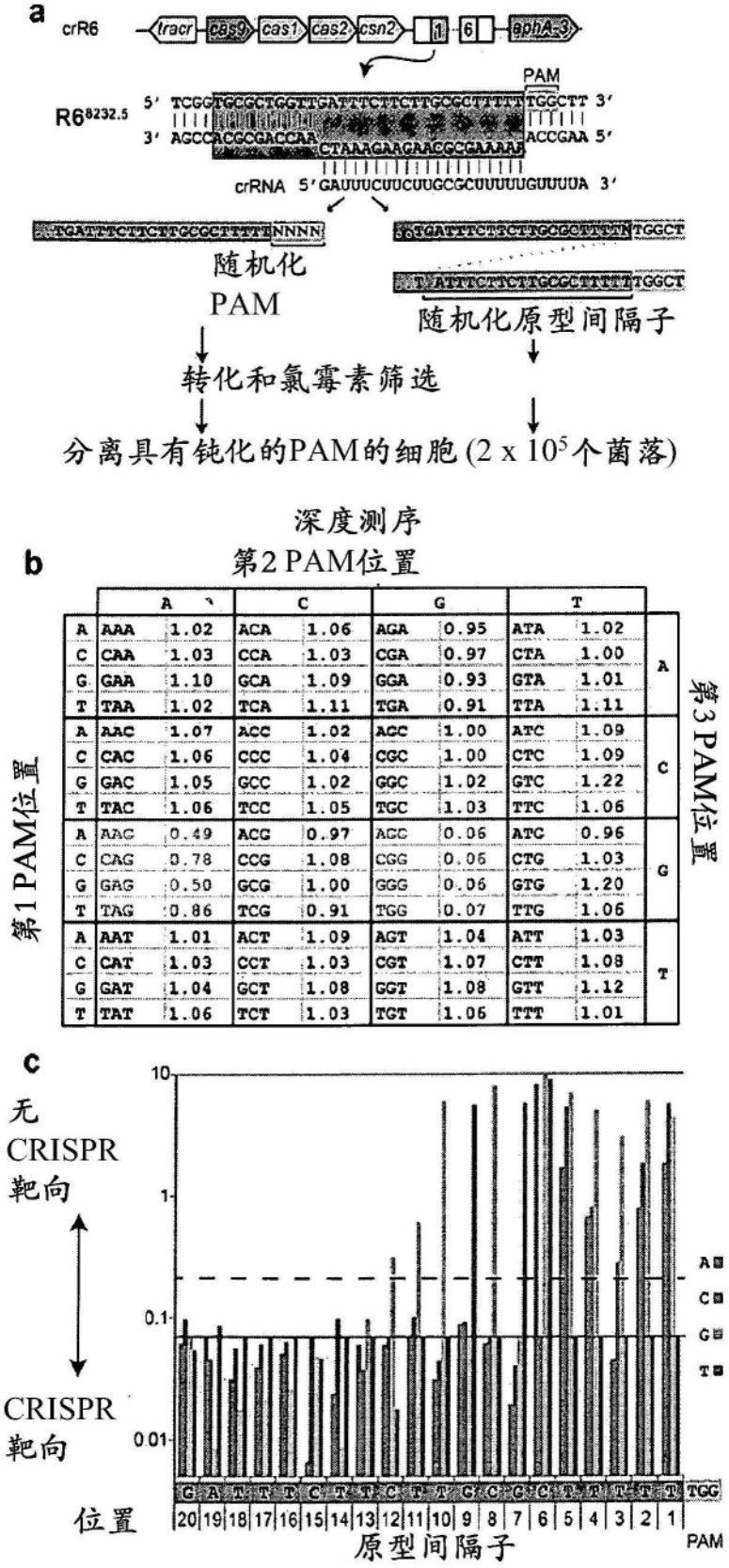


图23



	A		C		G		T		
A	AAA	1.02	ACA	1.06	AGA	0.95	ATA	1.02	A
C	CAA	1.03	CCA	1.03	CGA	0.97	CTA	1.00	
G	GAA	1.10	GCA	1.09	GGA	0.93	GTA	1.01	
T	TAA	1.02	TCA	1.11	TGA	0.91	TTA	1.11	
A	AAC	1.07	ACC	1.02	AGC	1.00	ATC	1.09	C
C	CAC	1.06	CCC	1.04	CGC	1.00	CTC	1.09	
G	GAC	1.05	GCC	1.02	GGC	1.02	GTC	1.22	
T	TAC	1.06	TCC	1.05	TGC	1.03	TTC	1.06	
A	AAG	0.49	ACG	0.97	AGG	0.06	ATG	0.96	G
C	CAG	0.78	CCG	1.08	CGG	0.06	CTG	1.03	
G	GAG	0.50	GCG	1.00	GGG	0.06	GTG	1.20	
T	TAG	0.86	TCG	0.91	TGG	0.07	TTG	1.06	
A	AAT	1.01	ACT	1.09	AGT	1.04	ATT	1.03	T
C	CAT	1.03	CCT	1.03	CGT	1.07	CTT	1.08	
G	GAT	1.04	GCT	1.08	GGT	1.08	GTT	1.12	
T	TAT	1.06	TCT	1.03	TGT	1.05	TTT	1.01	

图24

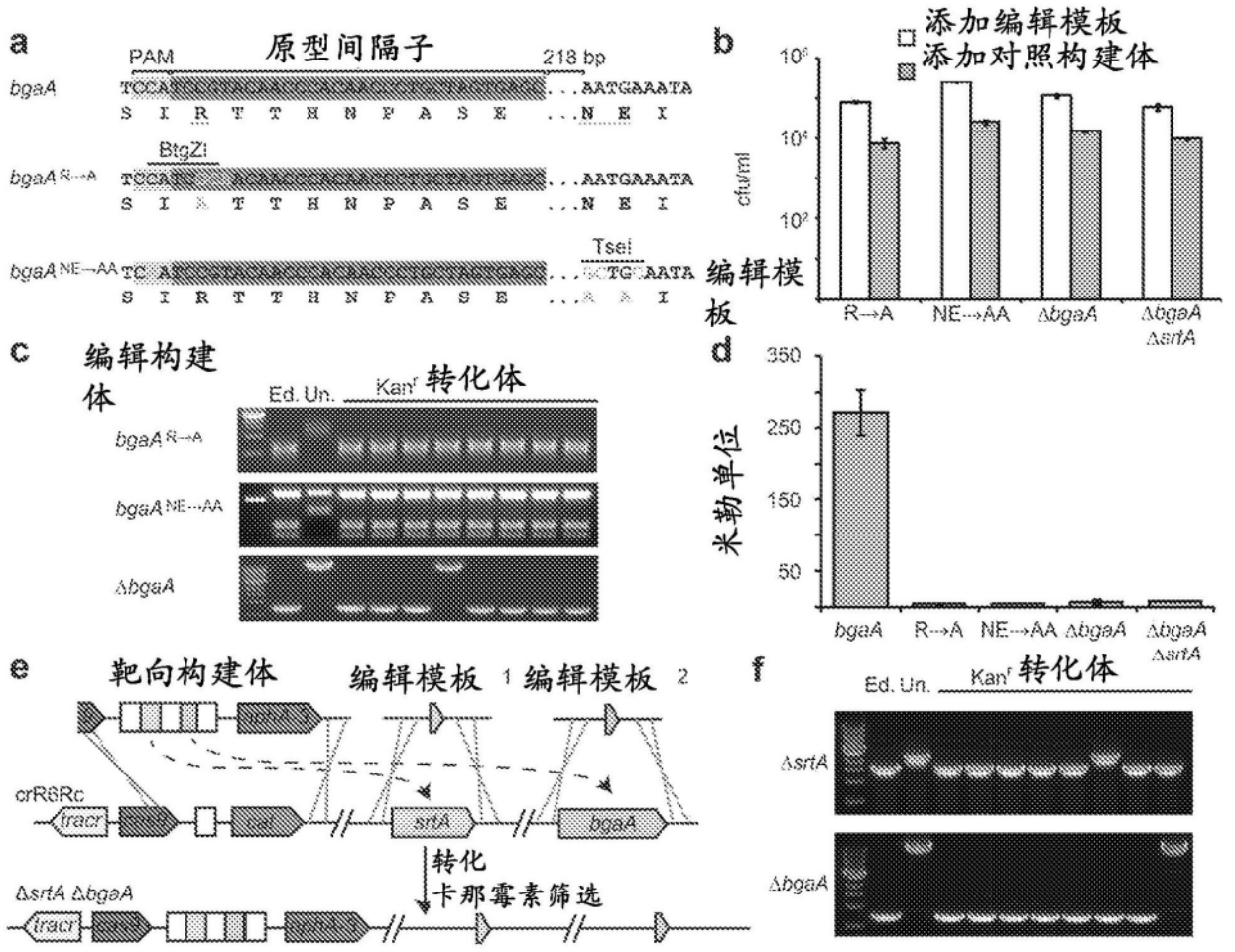

a
原型间隔子
PAM
218 bp
bgaA
BtgZI
TseI
b
添加编辑模板
添加对照构建体
cfu/ml
编辑模板
R→A
NE→AA
ΔbgaA
ΔbgaA ΔsrtA
c
编辑构建体
Ed. Un.
Kan^r 转化体
ΔbgaA
d
米勒单位
bgaA
e
靶向构建体
编辑模板 1
编辑模板 2
crRBRc
tracr
srtA
转化
卡那霉素筛选
ΔsrtA ΔbgaA
f
ΔsrtA

图25

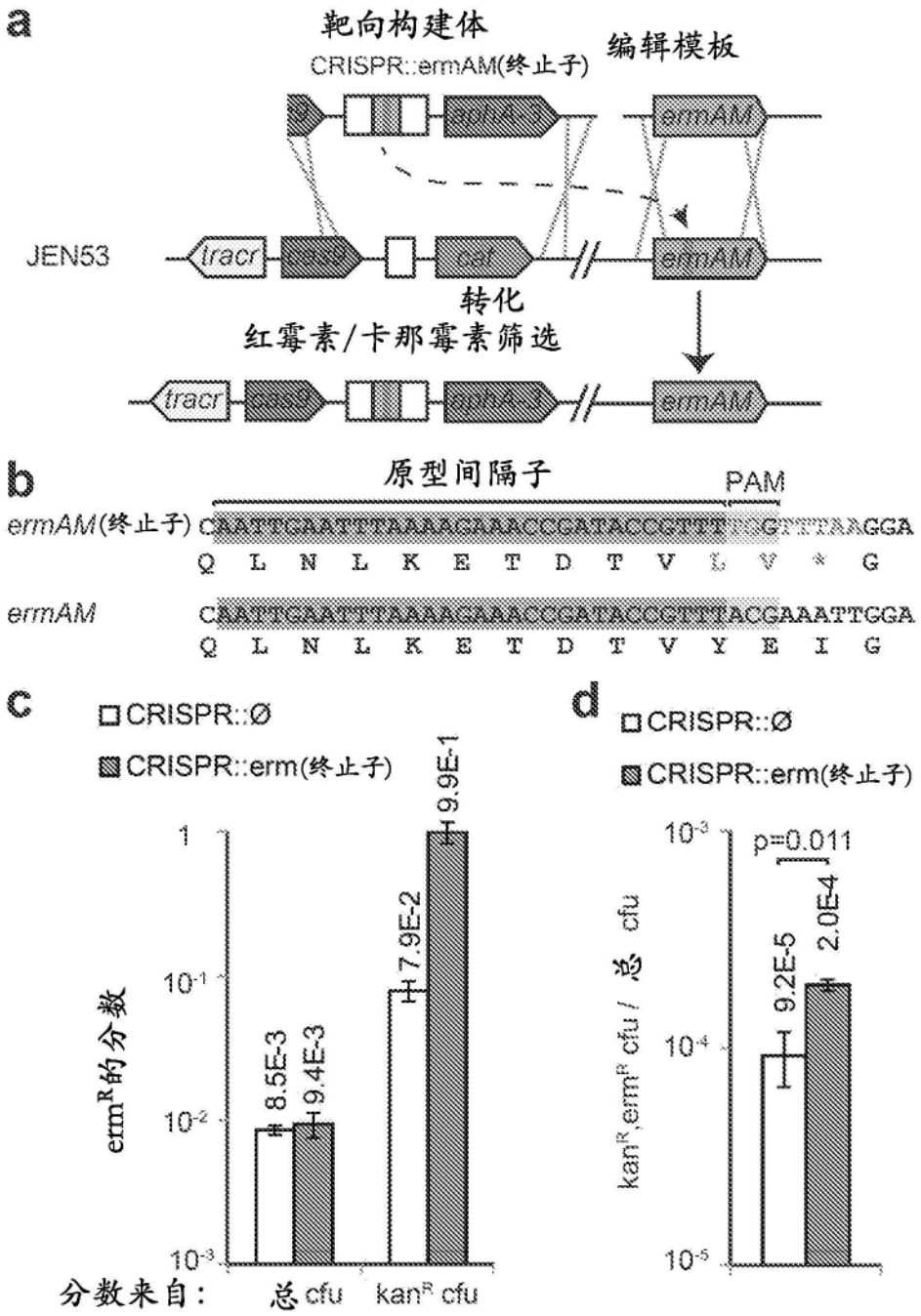

a
靶向构建体
编辑模板
CRISPR::ermAM(终止子)
JEN53
tracr
cas9
cat
ermAM
转化
红霉素/卡那霉素筛选
aphA-3
b
原型间隔子
PAM
ermAM(终止子) CAATTCAATTTAAAAGAAACCGATACCGTTT GTTTAAGGA
Q L N L K E T D T V L V * G
ermAM CAATTCAATTTAAAAGAAACCGATACCGTTTACGAAATTGGA
Q L N L K E T D T V Y E I G
c
CRISPR::Ø
CRISPR::erm(终止子)
ermR的分数
8.5E-3
9.4E-3
7.9E-2
9.9E-1
分数来自:
总 cfu
kanR cfu
d
CRISPR::Ø
CRISPR::erm(终止子)
kanR,ermR cfu / 总 cfu
p=0.011
9.2E-5
2.0E-4

图26

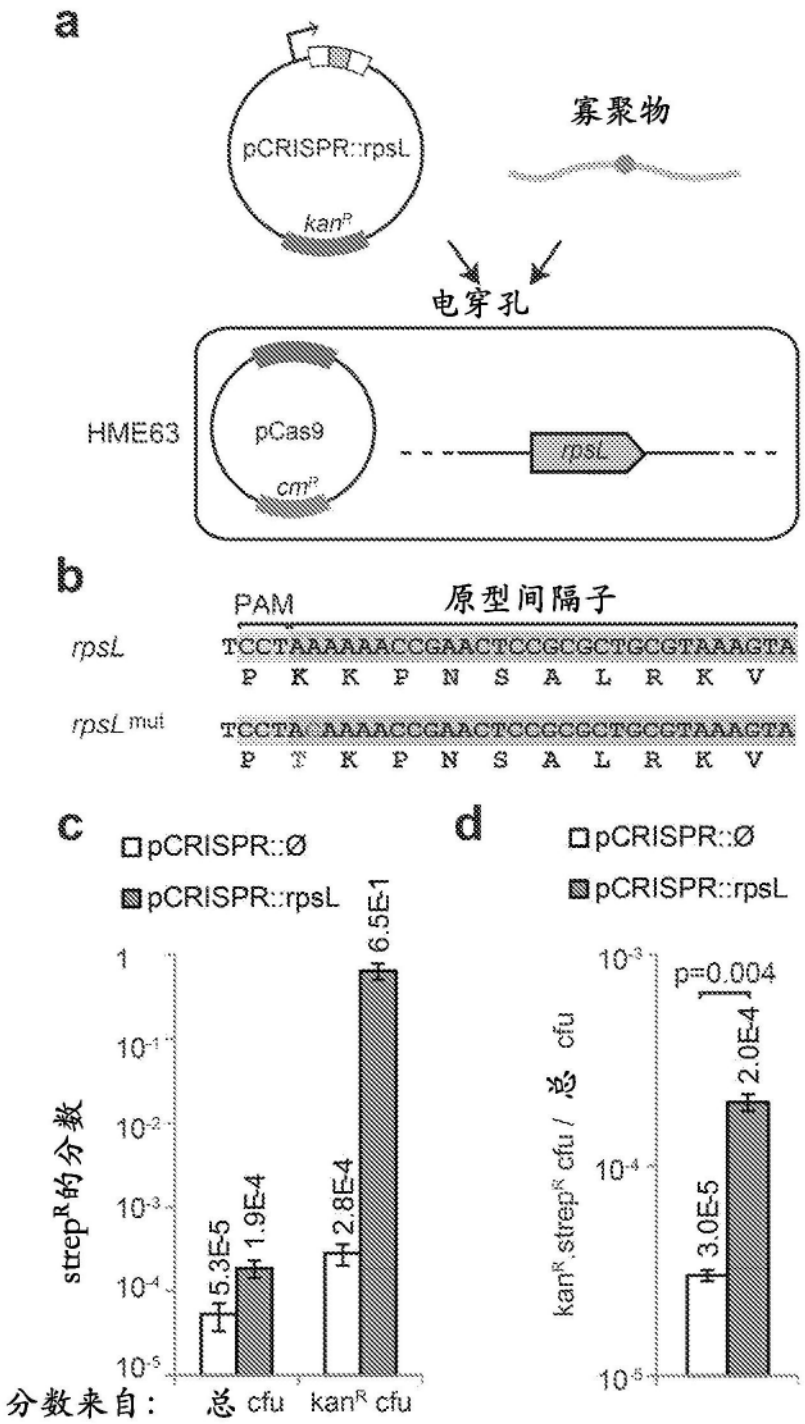

a
寡聚物
pCRISPR::rpsL
kanR
电穿孔
HME63
pCas9
cmR
rpsL
b
PAM
原型间隔子
rpsL
TCCTAAAAAACCGAACTCCGCGCTGCGTAAAGTA
P K K P N S A L R K V
rpsL mut
TCCTA AAAACCGAACTCCGCGCTGCGTAAAGTA
P T K P N S A L R K V
c
pCRISPR::Ø
pCRISPR::rpsL
strepR的分数
1
10-1
10-2
10-3
10-4
10-5
5.3E-5
1.9E-4
2.8E-4
6.5E-1
分数来自：
总 cfu
kanR cfu
d
pCRISPR::Ø
pCRISPR::rpsL
kanR strepR cfu / 总 cfu
10-3
10-4
10-5
p=0.004
3.0E-5
2.0E-4

图27

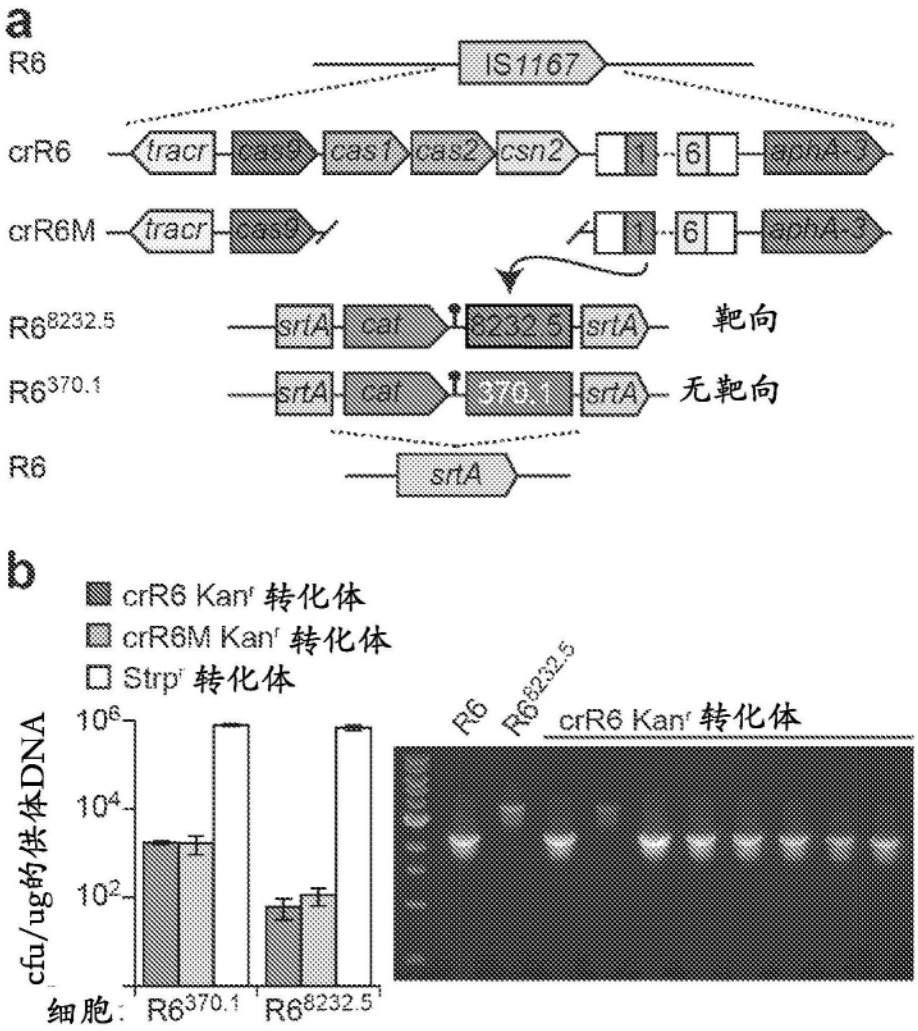

a
R6
IS1167
crR6
tracr
cas9
cas1
cas2
csn2
1
6
aphA-3
crR6M
tracr
cas9
1
6
aphA-3
R6^8232.5
srtA
cat
8232.5
srtA
靶向
R6^370.1
srtA
cat
370.1
srtA
无靶向
R6
srtA
b
crR6 Kan^r 转化体
crR6M Kan^r 转化体
Strp^r 转化体
cfu/ug的供体DNA
10^6
10^4
10^2
细胞:
R6^370.1
R6^8232.5
R6
R6^8232.5
crR6 Kan^r 转化体

图28

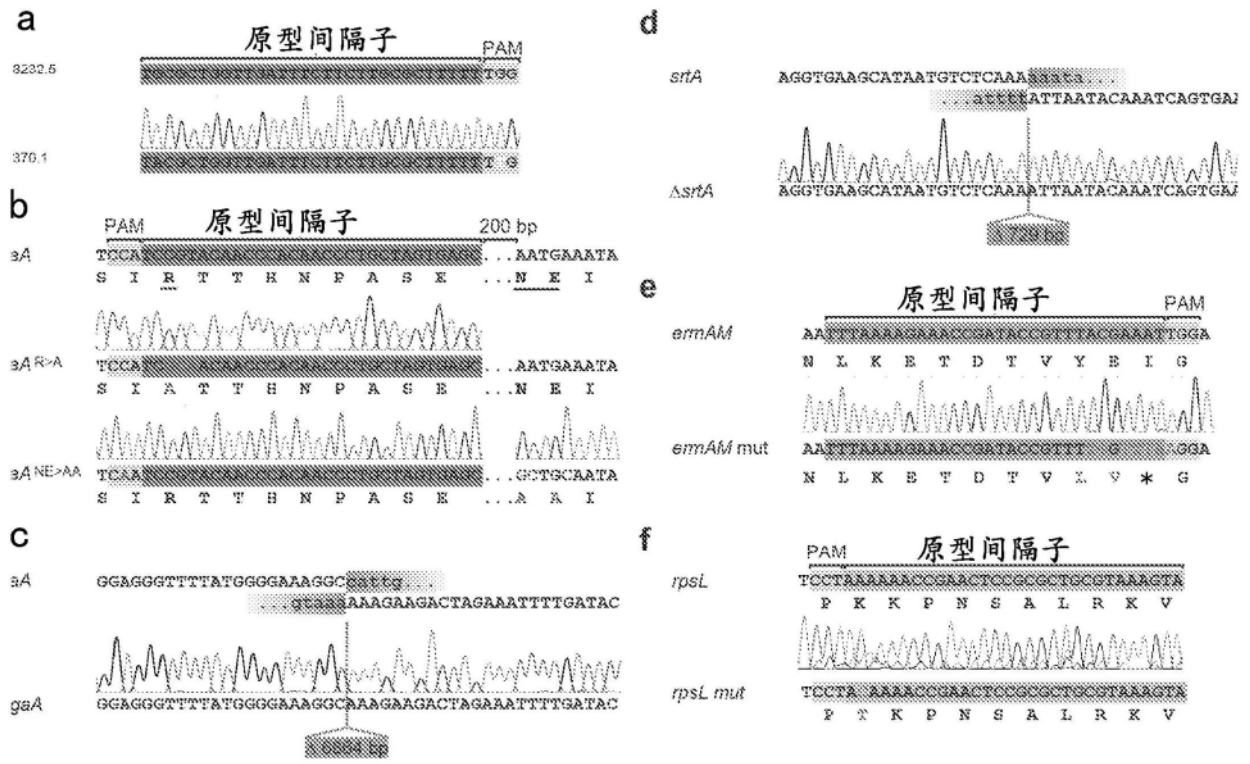

a
原型间隔子
PAM
3232.5
370.1
b
PAM
原型间隔子
200 bp
gA
gA R>A
gA NE>AA
c
gA
gsA
d
srtA
ΔsrtA
e
原型间隔子
PAM
ermAM
ermAM mut
f
PAM
原型间隔子
rpsL
rpsL mut

图29

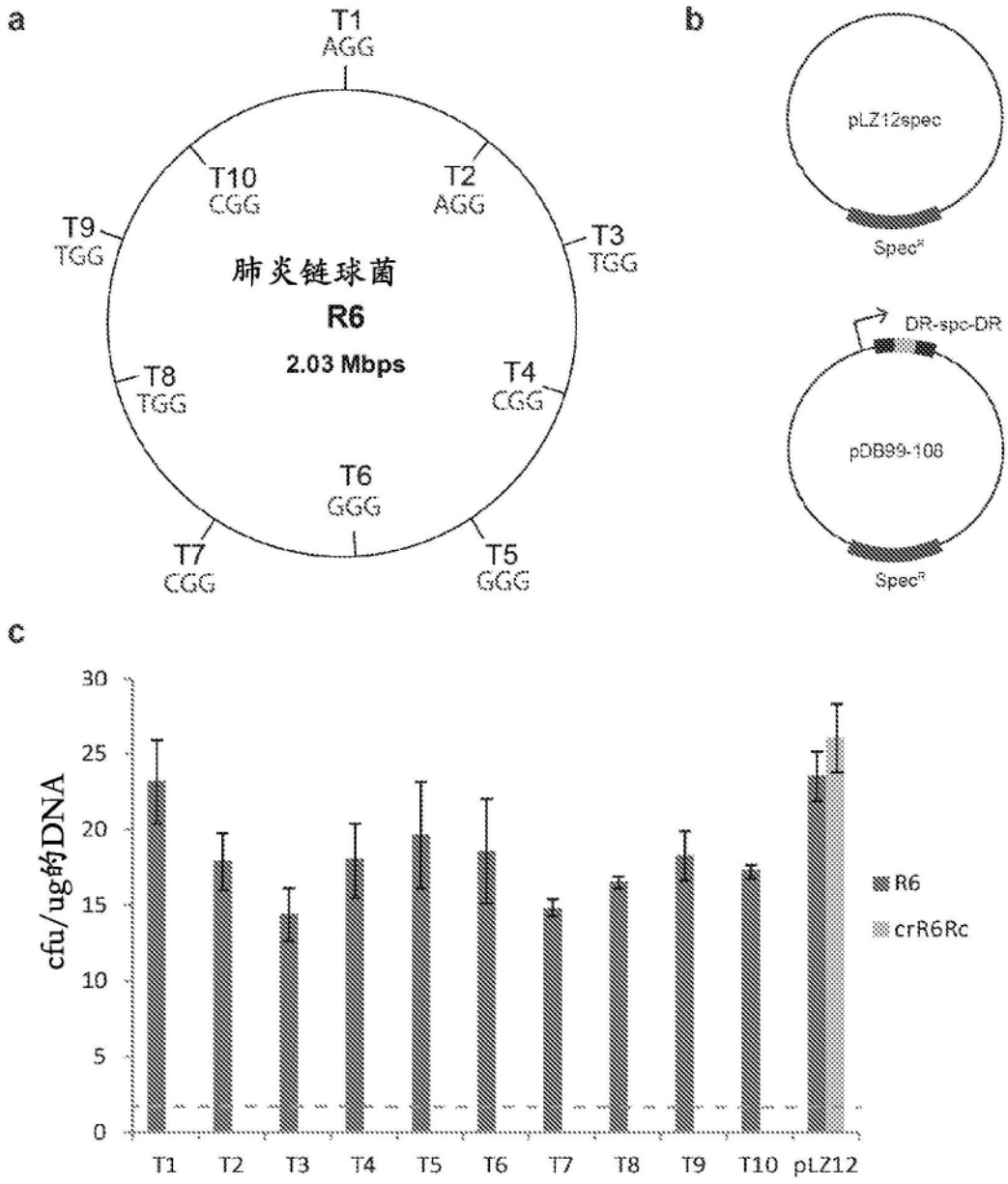

a
T1
AGG
T2
AGG
T3
TGG
T4
CGG
T5
GGG
T6
GGG
T7
CGG
T8
TGG
T9
TGG
T10
CGG
肺炎链球菌
R6
2.03 Mbps
b
pLZ12spec
Spec^R
DR-spc-DR
pDB99-108
Spec^R
c
cfu/ug的DNA
30
25
20
15
10
5
0
T1
T2
T3
T4
T5
T6
T7
T8
T9
T10
pLZ12
R6
crR6Rc

图30

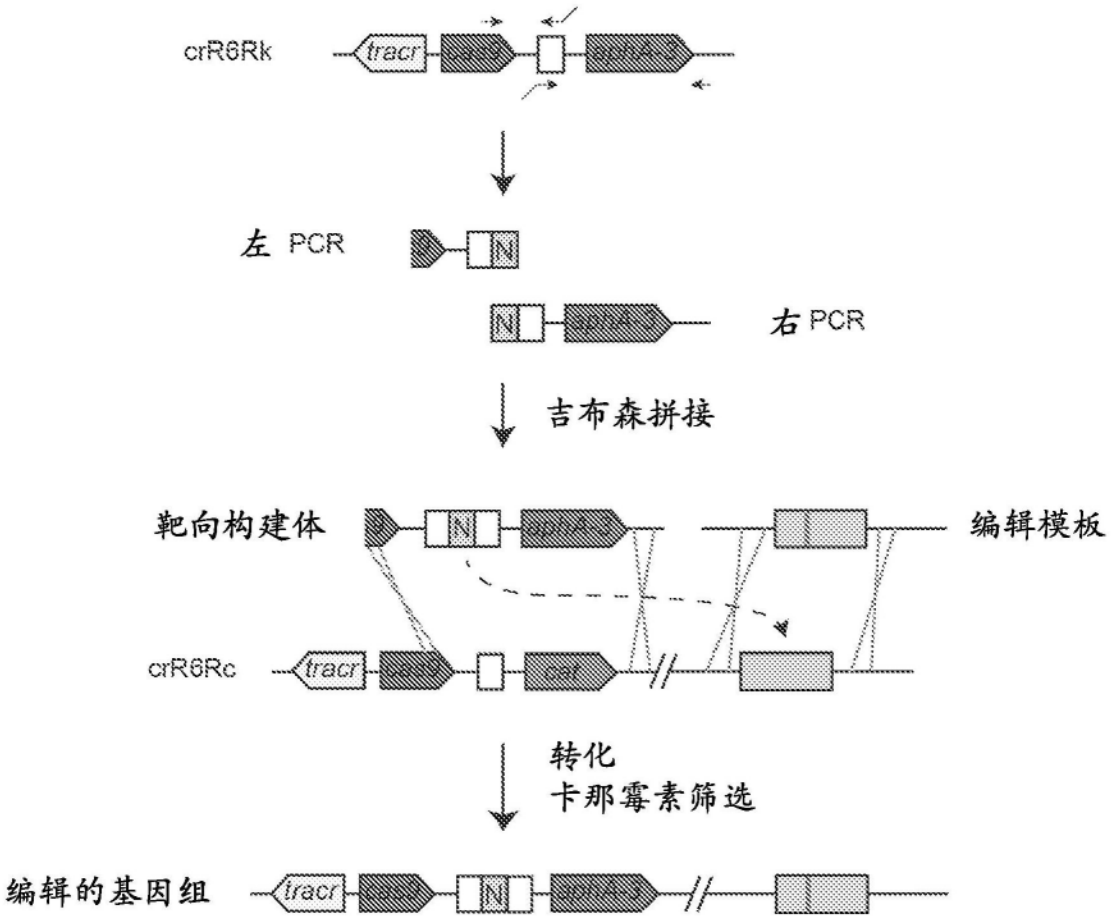

crR6Rk
tracr
左 PCR
N
右 PCR
吉布森拼接
靶向构建体
编辑模板
crR6Rc
cat
转化
卡那霉素筛选
编辑的基因组

图31

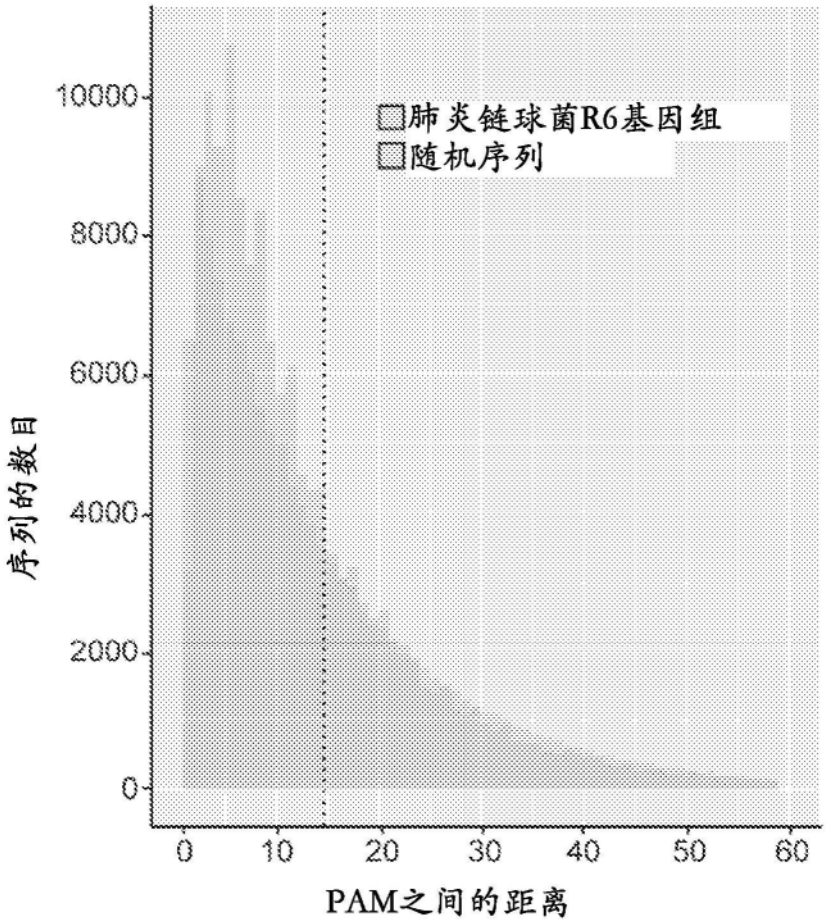

肺炎链球菌R6基因组
随机序列
序列的数目
10000
8000
6000
4000
2000
0
0
10
20
30
40
50
60
PAM之间的距离

图32

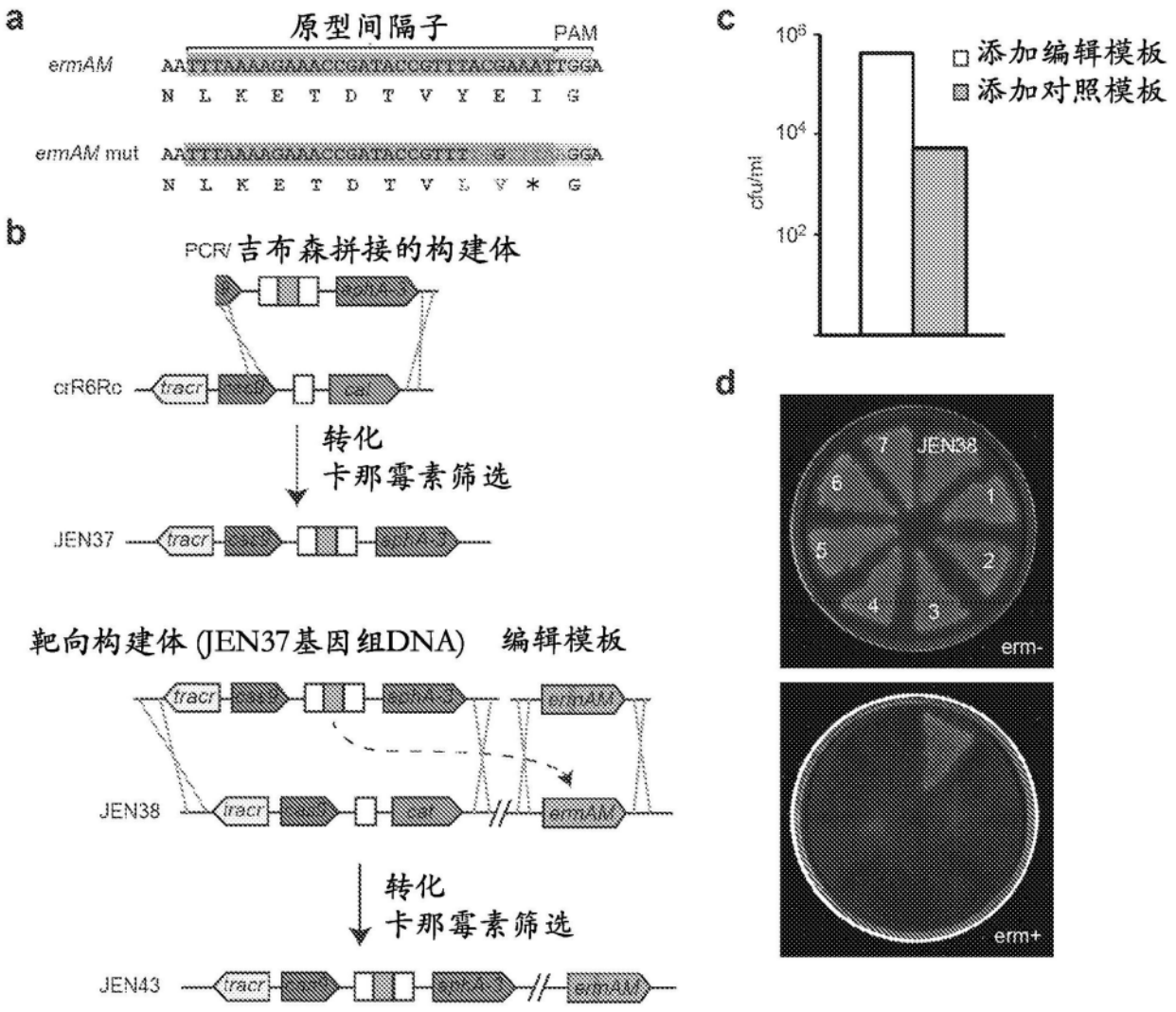

a
原型间隔子
PAM
ermAM
N L K E T D T V Y E I G
ermAM mut
N L K E T D T V L V * G
b
PCR/吉布森拼接的构建体
pCR6Rc
tracr
cat
转化
卡那霉素筛选
JEN37
靶向构建体 (JEN37基因组DNA) 编辑模板
ermAM
JEN38
JEN43
c
添加编辑模板
添加对照模板
cfu/ml
d
JEN38
erm-
erm+

图33

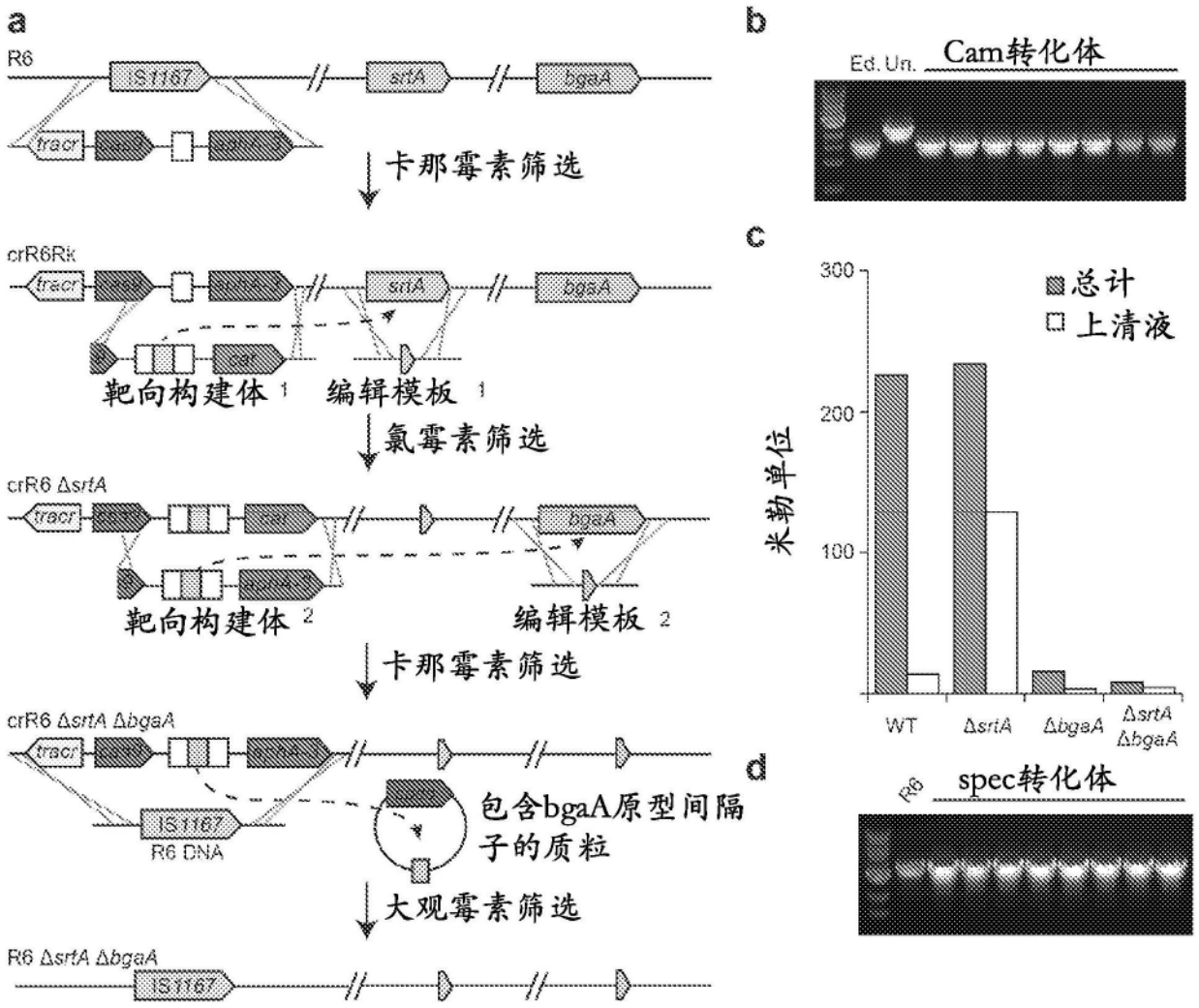

a
R6
IS1167
srtA
bgaA
tracr
卡那霉素筛选
crR6Rk
靶向构建体 1
编辑模板 1
氯霉素筛选
crR6 ΔsrtA
靶向构建体 2
编辑模板 2
卡那霉素筛选
crR6 ΔsrtA ΔbgaA
IS1167
R6 DNA
包含bgaA原型间隔子的质粒
大观霉素筛选
R6 ΔsrtA ΔbgaA
b
Ed. Un.
Cam转化体
c
总计
上清液
米勒单位
300
200
100
WT
ΔsrtA
ΔbgaA
ΔsrtA ΔbgaA
d
R6
spec转化体

图34

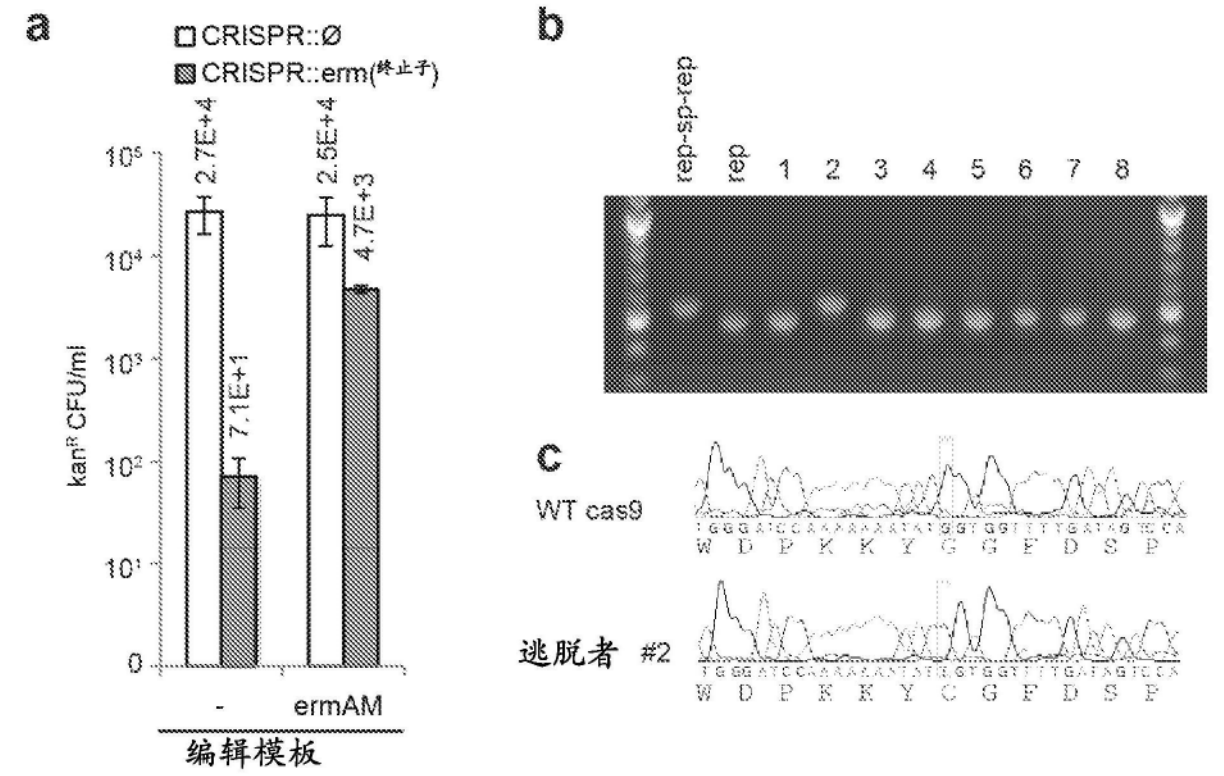

a
CRISPR::Ø
CRISPR::erm(终止子)
2.7E+4
7.1E+1
2.5E+4
4.7E+3
kanR CFU/ml
-
ermAM
编辑模板
b
rep-sp-rep
rep
1 2 3 4 5 6 7 8
c
WT cas9
逃脱者 #2

图35

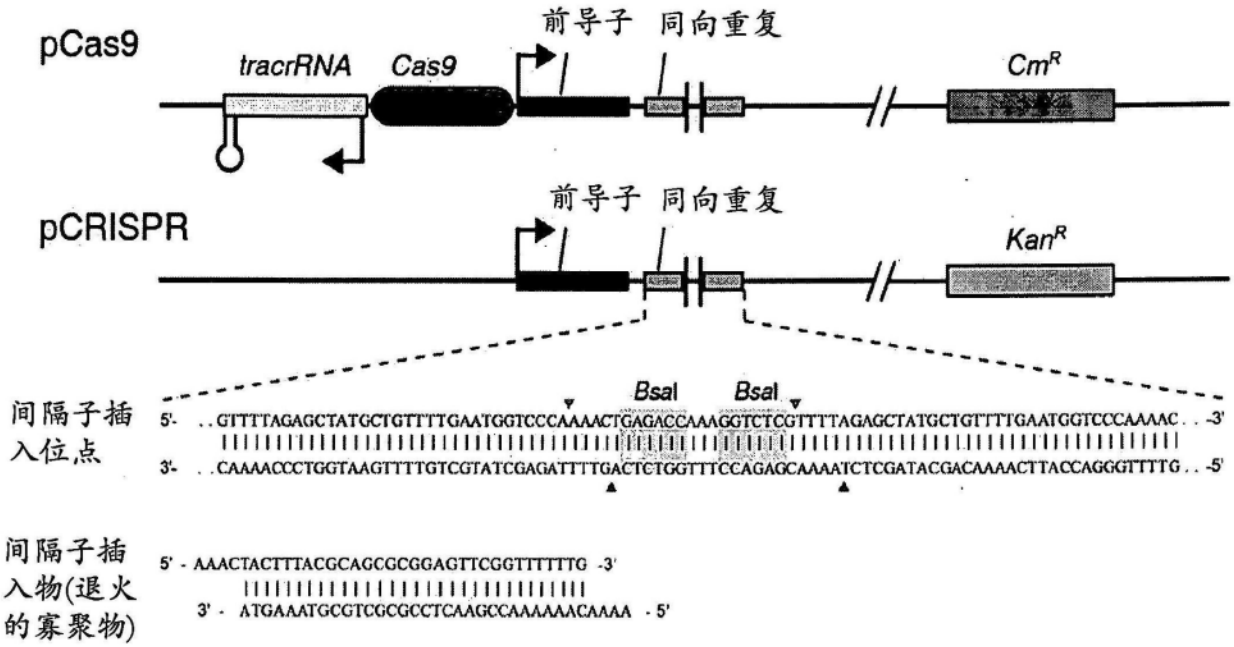

pCas9
tracrRNA
Cas9
前导子 同向重复
CmR
pCRISPR
前导子 同向重复
KanR
BsaI
BsaI
间隔子插入位点
5'- ..GTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACTGAGACCAAAGGTCTCGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAAC..-3'
3'- ..CAAAACCCTGGTAAGTTTTGTCGTATCGAGATTTTGACTCTGGTTTCCAGAGCAAAATCTCGATACGACAAAACTTACCAGGGTTTTG..-5'
间隔子插入物(退火的寡聚物)
5' - AAACTACTTTACGCAGCGCGGAGTTCGGTTTTTTG -3'
3' - ATGAAATGCGTCGCGCCTCAAGCCAAAAAACAAAA - 5'

图36

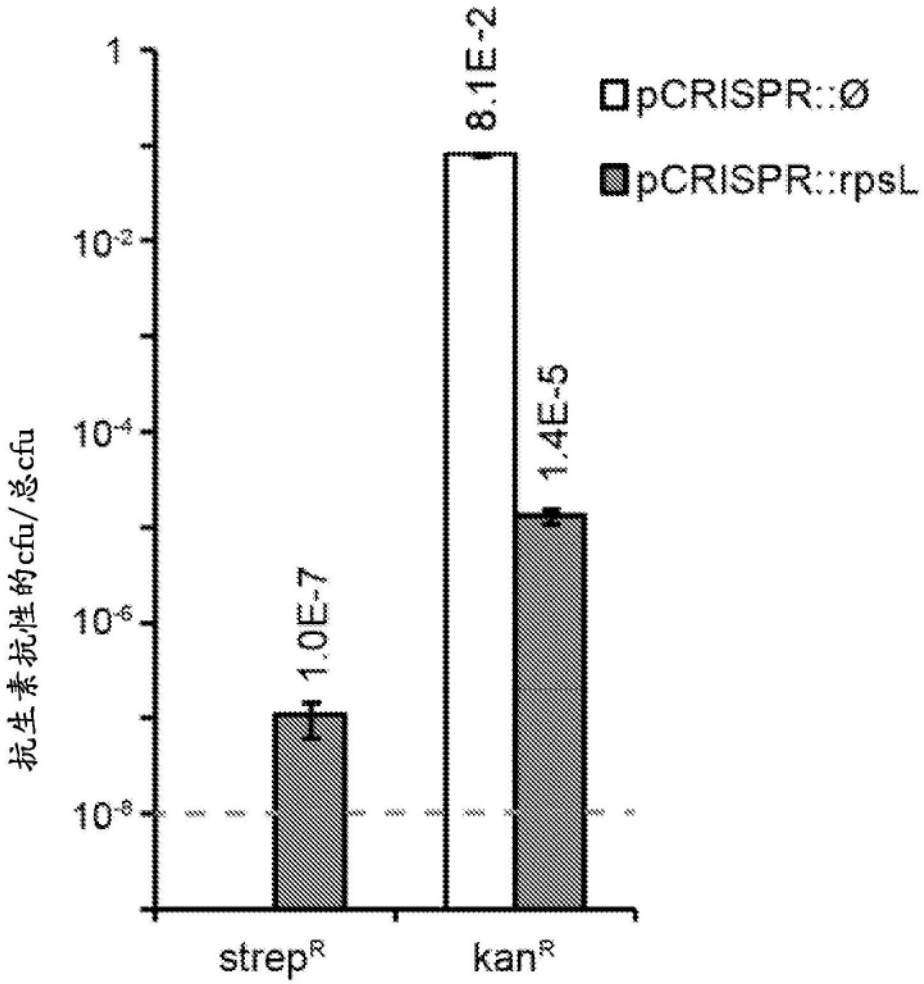

1
10^{-2}
10^{-4}
10^{-6}
10^{-8}
抗生素抗性的cfu/总cfu
8.1E-2
1.4E-5
1.0E-7
pCRISPR::Ø
pCRISPR::rpsL
strep^R
kan^R

图37

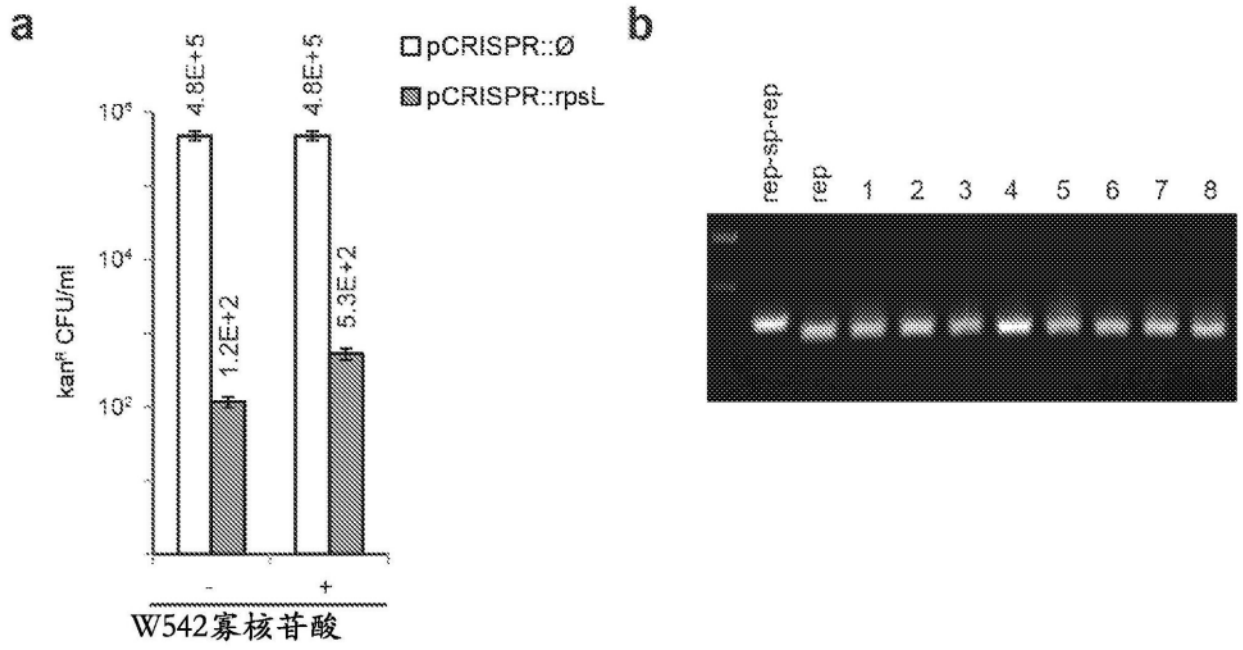

a
pCRISPR::Ø
pCRISPR::rpsL
4.8E+5
4.8E+5
1.2E+2
5.3E+2
kan^R CFU/ml
-
+
W542寡核苷酸
b
rep-sp-rep
rep
1
2
3
4
5
6
7
8

图38

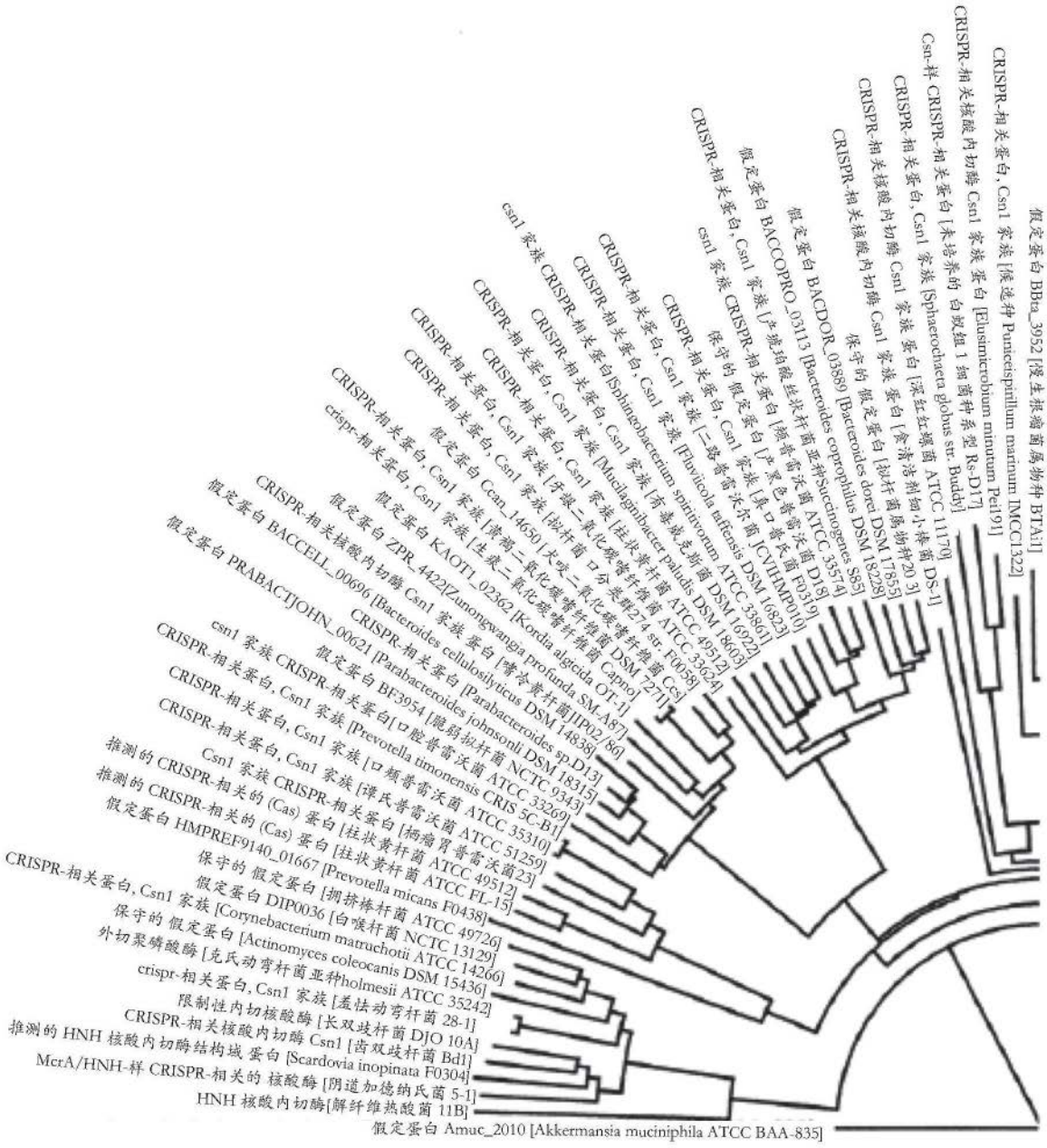

假定蛋白 BBta_3952 [慢生根瘤菌属物种 BTAi1]
CRISPR-相关蛋白, Csn1 家族 [候选种 Puniceispirillum marinum IMCC1322]
CRISPR-相关核酸内切酶 Csn1 家族 蛋白 [Elusimicrobium minutum Pei191]
Csn-样 CRISPR-相关蛋白 [未培养的 白蚁组 1 细菌种系型 Rs-D17]
CRISPR-相关蛋白, Csn1 家族 [Sphaerochaeta globus str. Buddy]
CRISPR-相关核酸内切酶 Csn1 家族 蛋白 [深红螺菌 ATCC 11170]
CRISPR-相关核酸内切酶 Csn1 家族 蛋白 [食清洁剂细小棒菌 DS-1]
保守的 假定蛋白 [拟杆菌属物种20 3]
假定蛋白 BACDOR_03889 [Bacteroides coprophilus DSM 18228]
假定蛋白 BACCOPRO_03113 [Bacteroides dorei DSM 17855]
CRISPR-相关蛋白, Csn1 家族 [解琥珀酸丝状杆菌亚种Succinogenes S85]
CRISPR-相关蛋白, Csn1 家族 [产琥珀酸丝状杆菌亚种Succinogenes ATCC 33574]
保守的 假定蛋白 [产黑色普雷沃菌 F0319]
CRISPR-相关蛋白, Csn1 家族 [二路普雷沃菌 JCVIHMP010]
csn1 家族 CRISPR-相关蛋白 [梅奇普雷沃菌 DSM 16823]
CRISPR-相关蛋白, Csn1 家族 [Fluviicola taffensis DSM 18603]
CRISPR-相关蛋白, Csn1 家族 [布氏威克斯菌 ATCC 33624]
CRISPR-相关蛋白, Csn1 家族 [Sphingobacterium spiritivorum ATCC 33861]
CRISPR-相关蛋白, Csn1 家族 [Mucilaginibacter paludis DSM 16922]
CRISPR-相关蛋白, Csn1 家族 [拉米莱杆菌 ATCC 49512]
CRISPR-相关蛋白, Csn1 家族 [Solitalea canadensis DSM 3403]
csn1 家族 CRISPR-相关蛋白, Csn1 家族 [Sphingobacterium sp. F0058]
CRISPR-相关蛋白, Csn1 家族 [矛盾二氧化碳嗜纤维菌 ATCC 49512]
CRISPR-相关蛋白, Csn1 家族 [拟杆菌 D 分类群274 str. F0036]
CRISPR-相关蛋白, Csn1 家族 [矛盾二氧化碳嗜纤维菌 DSM 7271]
假定蛋白 Ccan_14650 [犬咬二氧化碳嗜纤维菌 Capno]
CRISPR-相关蛋白, Csn1 家族 [生殖二氧化碳嗜纤维菌 DSM 7271]
crispr-相关蛋白, Csn1 家族 [Cnuongwangia profunda SM-A87]
假定蛋白 KAOT1_02362 [Kordia algicida OT-1]
假定蛋白 ZPR_4422[Zunongwangia profunda SM-A87]
CRISPR-相关核酸内切酶 Csn1 家族 蛋白 [嗜冷黄杆菌HIP02/86]
假定蛋白 BACCELL_00696 [Bacteroides cellulosilyticus DSM 14838]
CRISPR-相关蛋白 [Parabacteroides johnsonli DSM 18315]
假定蛋白 PRABACTJOHN_00621 [Parabacteroides johnsonli DSM 18315]
csn1 家族 CRISPR-相关蛋白 BF3954 [脆弱拟杆菌 NCTC 9343]
CRISPR-相关蛋白, Csn1 家族 [Prevotella timonensis CRIS 5C-B1]
CRISPR-相关蛋白, Csn1 家族 [口腔普雷沃菌 ATCC 33269]
推测的 CRISPR-相关的 (Cas) 蛋白 [谭氏普雷沃菌 ATCC 35310]
Csn1 家族 CRISPR-相关蛋白 [口频普雷沃菌 ATCC 51259]
推测的 CRISPR-相关的 (Cas) 蛋白 [柄瘤胃普雷沃菌23]
假定蛋白 HMPREF9140_01667 [Prevotella micans F0438]
保守的 假定蛋白 [柱状黄杆菌 ATCC 49512]
CRISPR-相关蛋白, Csn1 家族 [Prevotella micans ATCC FL-15]
假定蛋白 DIP0036 [白喉杆菌 NCTC 13129]
保守的 假定蛋白 [拥挤棒杆菌 ATCC 49726]
外切聚磷酸酶 [Corynebacterium matruchotii ATCC 14266]
crispr-相关蛋白, Csn1 家族 [克氏动弯杆菌亚种holmesii ATCC 14266]
限制性内切核酸酶 [孟怯动弯杆菌 ATCC 35242]
CRISPR-相关核酸内切酶 Csn1 [长双歧杆菌 28-1]
推测的 HNH 核酸内切酶结构域 蛋白 [齿双歧杆菌 DJO 10A]
McrA/HNH-样 CRISPR-相关的 核酸酶 [Scardovia inopinata Bd1]
HNH 核酸内切酶[阴道加德纳氏菌 F0304]
假定蛋白 Amuc_2010 [解纤维热酸菌 5-1]
[Akkermansia muciniphila ATCC BAA-835] 11B]

图39A

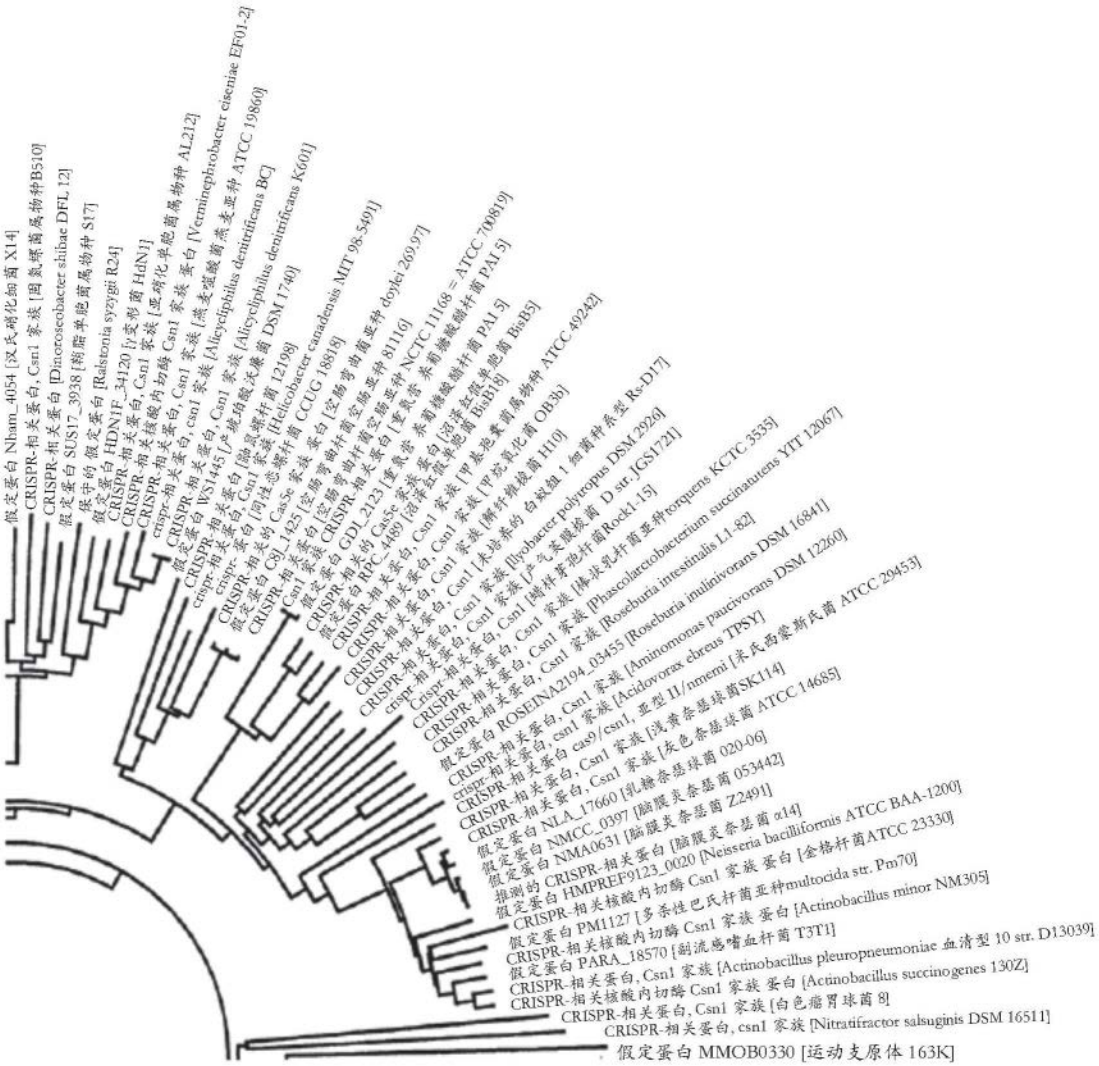

假定蛋白 Nham_4054 [汉氏硝化细菌 X14]
CRISPR-相关蛋白, Csn1 家族 [固氮螺菌属物种BS10]
CRISPR-相关蛋白 [Dinoroseobacter shibae DFL 12]
假定蛋白 SUS17_3938 [梭脂单胞菌属物种 S17]
保守的 假定蛋白 [Ralstonia syzygii R24]
假定蛋白 HDN1F_34120 [少变形菌 HdN1]
CRISPR-相关蛋白, Csn1 家族 [恶碎化单胞菌属物种 AL212]
CRISPR-相关核酸内切酶 Csn1 家族 [燕麦强酸菌燕麦亚种 ATCC 19860]
crispr-相关蛋白, csn1 家族 [Verminiphotobacter eiseniae EF01-2]
假定蛋白 WS1445 [产琥珀酸沃氏菌 DSM 1740]
CRISPR-相关蛋白, Csn1 家族 [Alicycliphilus denitrificans BC]
crispr-相关蛋白 [间柱恋螺杆菌 CasSe 家族 [Helicobacter canadensis MIT 98-5491]
CRISPR-相关蛋白 C9J_1425 [空肠弯曲杆菌空肠亚种 81116]
Csn1 家族 [空肠弯曲杆菌 CCUG 18818]
假定蛋白 GDJ_2123 [芝麻鞘单胞菌 NCTC 11168 = ATCC 700819]
CRISPR-相关蛋白 [芝麻鞘单胞菌 BisB5]
CRISPR-相关的 CasSe 家族蛋白 [沼泽红假单胞菌 BisB18]
CRISPR-相关蛋白, Csn1 家族 [甲基泡囊菌属物种 ATCC 49242]
CRISPR-相关蛋白, Csn1 家族 [嗜纤维极菌 H10]
CRISPR-相关蛋白, Csn1 家族 [产气荚膜梭菌 D str. JGS1721]
CRISPR-相关蛋白, Csn1 家族 [Hydrobacter polymorphus DSM 2926]
CRISPR-相关蛋白, Csn1 家族 [Phascolarctobacterium succinatutens YIT 12067]
CRISPR-相关蛋白, csn1 家族 [Roseburia intestinalis L1-82]
CRISPR-相关蛋白, Csn1 家族 [Aminomonas paucivorans DSM 12260]
CRISPR-相关蛋白, Csn1 家族 [Acidovorax ebreus TPSY]
CRISPR-相关蛋白, Csn1 家族 [浅黄杂基球菌SK114]
CRISPR-相关蛋白, Csn1 家族 [灰色杂基球菌 020-06]
CRISPR-相关蛋白, Csn1 家族 [乳糖杂基球菌 053442]
假定蛋白 NMA0631 [脑膜炎奈瑟菌 α14]
假定蛋白 HMPREF9123_0020 [Neisseria bacilliformis ATCC BAA-1200]
CRISPR-相关蛋白 [金格杆菌ATCC 23330]
CRISPR-相关核酸内切酶 Csn1 家族蛋白 [Actinobacillus minor NM305]
假定蛋白 PM1127 [多杀性巴氏杆菌亚种multocida str. Pm70]
CRISPR-相关蛋白, Csn1 家族 [Actinobacillus pleuropneumoniae 血清型 10 str. D13039]
假定蛋白 PARA_18570 [副流感嗜血杆菌 T3T1]
CRISPR-相关蛋白, Csn1 家族 [白色瘤胃球菌 8]
CRISPR-相关核酸内切酶 Csn1 家族蛋白 [Actinobacillus succinogenes 130Z]
CRISPR-相关蛋白, csn1 家族 [Nitratifractor salsuginis DSM 16511]
假定蛋白 MMOB0330 [运动支原体 163K]

图39B

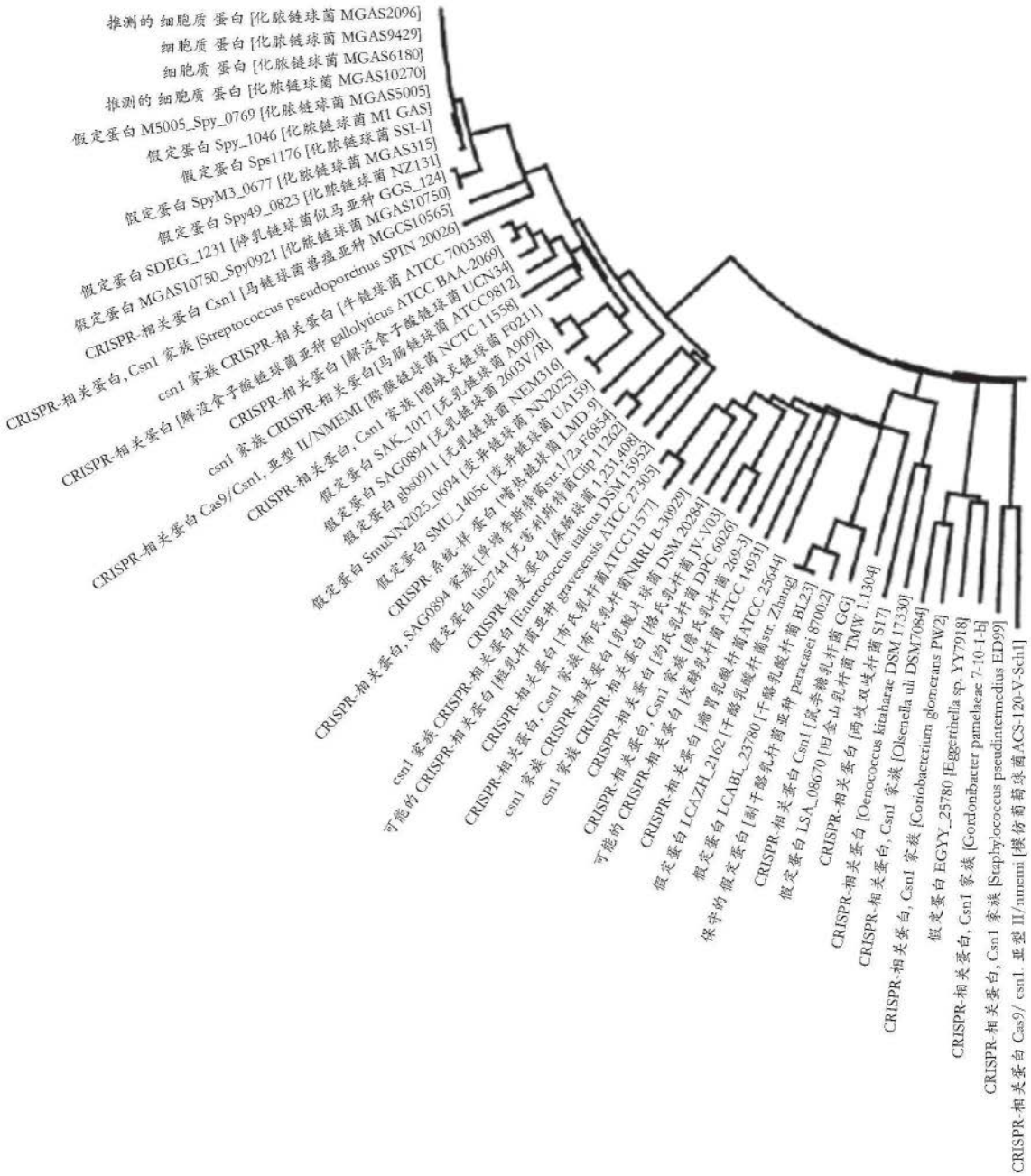

推测的 细胞质 蛋白 [化脓链球菌 MGAS2096]
细胞质 蛋白 [化脓链球菌 MGAS9429]
细胞质 蛋白 [化脓链球菌 MGAS6180]
推测的 细胞质 蛋白 [化脓链球菌 MGAS10270]
假定蛋白 M5005_Spy_0769 [化脓链球菌 MGAS5005]
假定蛋白 Spy_1046 [化脓链球菌 M1 GAS]
假定蛋白 Sps1176 [化脓链球菌 SSI-1]
假定蛋白 SpyM3_0677 [化脓链球菌 MGAS315]
假定蛋白 Spy49_0823 [化脓链球菌 NZ131]
假定蛋白 SDEG_1231 [停乳链球菌似马亚种 GGS_124]
假定蛋白 MGAS10750_Spy0921 [化脓链球菌 MGAS10750]
CRISPR-相关蛋白 Csn1 [马链球菌兽瘟亚种 MGCS10565]
CRISPR-相关蛋白, Csn1 家族 [Streptococcus pseudoporcinus SPIN 20026]
csn1 家族 CRISPR-相关蛋白 [解没食子酸链球菌亚种 gallolyticus ATCC 700338]
CRISPR-相关蛋白 [解没食子酸链球菌 ATCC BAA-2069]
CRISPR-相关蛋白, Csn1 家族 [解没食子酸链球菌 UCN34]
CRISPR-相关蛋白 Cas9/Csn1, 亚型 II/NMEMI [脑膜炎奈瑟菌 ATCC9812]
csn1 家族 CRISPR-相关蛋白, Csn1 家族 [吲哚发链球菌 NCTC 11558]
CRISPR-相关蛋白 [无乳链球菌 A909]
假定蛋白 SAK_1017 [无乳链球菌 F0211]
假定蛋白 SAG0894 [无乳链球菌 2603V/R]
假定蛋白 gbs0911 [无乳链球菌 NEM316]
假定蛋白 SmuNN2025_0694 [变异链球菌 NN2025]
假定蛋白 SMU_1405c [变异链球菌 UA159]
CRISPR-系统-相 蛋白 [坤溶李斯特菌 str.1/2a F6854]
CRISPR-相关蛋白, Csn1 家族 [坤溶李斯特菌 LMD-9]
CRISPR-相关蛋白 [Enterococcus italicus DSM 15952]
csn1 家族 CRISPR-相关蛋白 [肠球菌 1,231,408]
CRISPR-相关蛋白 [屎肠球菌 gravesensis ATCC 27305]
CRISPR-相关蛋白, Csn1 家族 [屎肠球菌 NRRL B-30929]
可能的 CRISPR-相关蛋白 [乳酸片球菌 DSM 20284]
CRISPR-相关蛋白 [格氏乳杆菌 JV-V03]
假定蛋白 LCAZH_2162 [约氏乳杆菌 DPC 6026]
假定蛋白 LCABL_23780 [浅酵乳杆菌 269-3]
保守的 假定蛋白 [副干酪乳杆菌 ATCC 14931]
CRISPR-相关蛋白 Csn1 [干酪乳杆菌亚种 paracasei str. Zhang]
假定蛋白 LSA_08670 [旧金山乳杆菌 BL23]
CRISPR-相关蛋白 [鼠李糖乳杆菌 8700:2]
CRISPR-相关蛋白 [Oenococcus kitaharae DSM 17330]
CRISPR-相关蛋白, Csn1 家族 [Olsenella uli DSM7084]
假定蛋白 EGYY_25780 [Eggerthella sp. YY7918]
CRISPR-相关蛋白, Csn1 家族 [Gordonibacter pamelaeae 7-10-1-b]
CRISPR-相关蛋白, Csn1 家族 [Staphylococcus pseudintermedius ED99]
CRISPR-相关蛋白 Cas9/ csn1 亚型 II/nmemi [模仿葡萄球菌 ACS-120-V-Sch1]

图39C

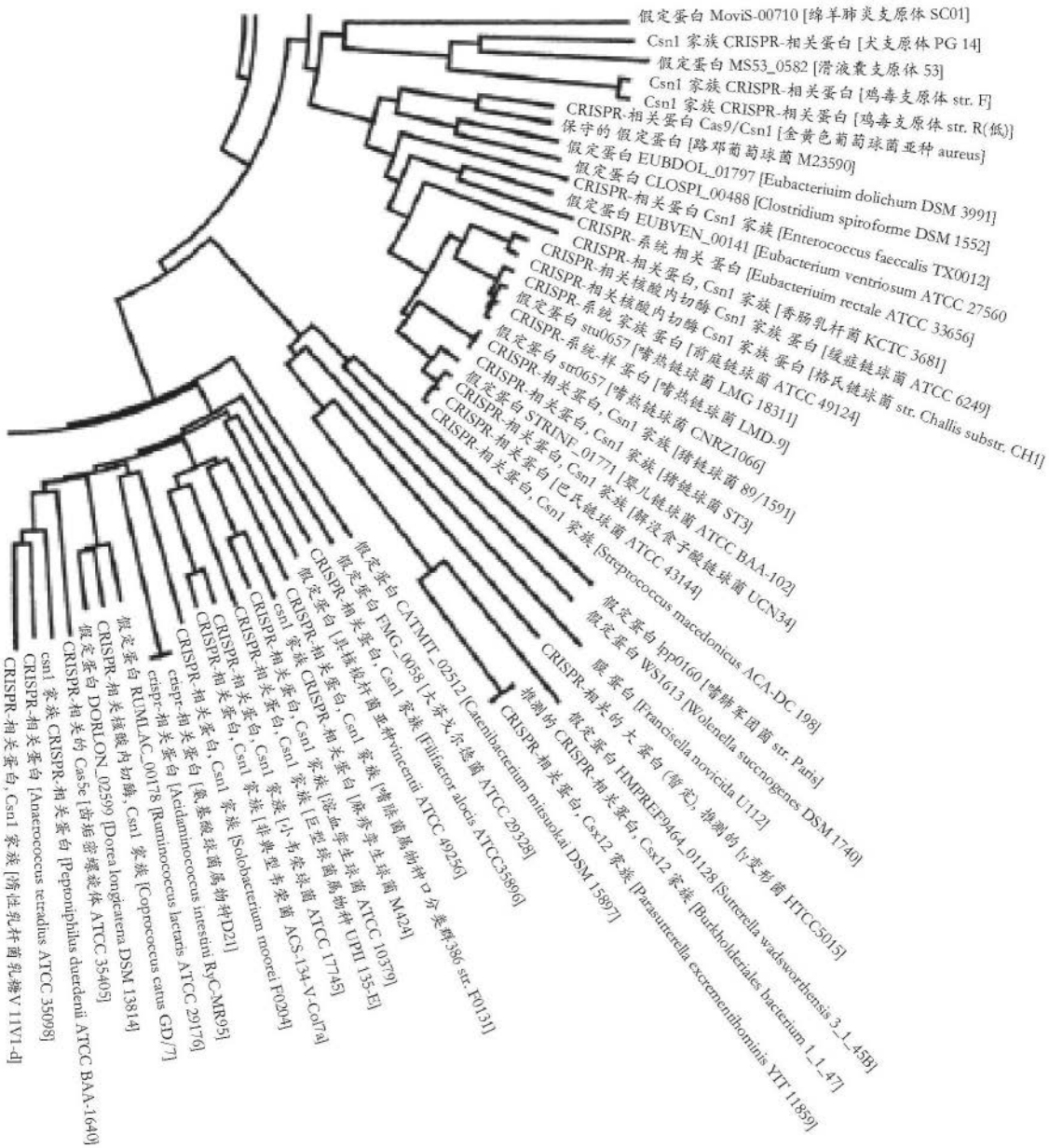

假定蛋白 MoviS-00710 [绵羊肺炎支原体 SC01]
Csn1 家族 CRISPR-相关蛋白 [犬支原体 PG 14]
假定蛋白 MS53_0582 [滑液囊支原体 53]
Csn1 家族 CRISPR-相关蛋白 [鸡毒支原体 str. F]
Csn1 家族 CRISPR-相关蛋白 [鸡毒支原体 str. R(低)]
CRISPR-相关蛋白 Cas9/Csn1 [金黄色葡萄球菌亚种 aureus]
保守的 假定蛋白 [路邓葡萄球菌 M23590]
假定蛋白 EUBDOL_01797 [Eubacteriuim dolichum DSM 3991]
假定蛋白 CLOSPI_00488 [Clostridium spiroforme DSM 1552]
CRISPR-相关蛋白 Csn1 家族 [Enterococcus faeccalis TX0012]
假定蛋白 EUBVEN_00141 [Eubacterium ventriosum ATCC 27560
CRISPR-系统 相关 蛋白 [Eubacteriuim rectale ATCC 33656]
CRISPR-相关蛋白, Csn1 家族 [香肠乳杆菌 KCTC 3681]
CRISPR-相关核酸内切酶 Csn1 家族 蛋白 [缓症链球菌 ATCC 6249]
CRISPR-系统 家族 蛋白 [前庭链球菌 ATCC 49124]
假定蛋白 stu0657 [嗜热链球菌 LMG 18311]
CRISPR-系统-样 蛋白 [嗜热链球菌 LMD-9]
假定蛋白 str0657 [嗜热链球菌 CNRZ1066]
CRISPR-相关蛋白, Csn1 家族 [猪链球菌 89/1591]
假定蛋白 STRINF_01771 [婴儿链球菌 ATCC BAA-102]
CRISPR-相关蛋白, Csn1 家族 [解没食子酸链球菌 UCN34]
CRISPR-相关蛋白, Csn1 家族 [巴氏链球菌 ATCC 43144]
假定蛋白 CATMIT_02512 [Catenibacterium mitsuokai DSM 15897]
假定蛋白 FMG_0058 [大杀戈尔德菌 ATCC 49256]
CRISPR-相关蛋白, Csn1 家族 [Filifactor alocis ATCC35896]
假定蛋白 [具核梭杆菌亚种vincentii ATCC 49256]
CRISPR-相关蛋白, Csn1 家族 [嗜麻黄生球菌 M424]
csn1 家族 CRISPR-相关蛋白 [麻黄生球菌 ATCC 10379]
CRISPR-相关蛋白, Csn1 家族 [溶血球菌属物种 UPII 135-E]
CRISPR-相关蛋白, Csn1 家族 [巨型球菌属物种 ATCC 17745]
CRISPR-相关蛋白, Csn1 家族 [小韦荣球菌 ACS-134-V-Col7a]
CRISPR-相关蛋白, Csn1 家族 [非典型芽孢菌 ACS-134-V-Col7a]
CRISPR-相关蛋白, Csn1 家族 [Solobacterium moorei F0204]
CRISPR-相关蛋白 [氨基酸球菌属物种D21]
CRISPR-相关蛋白, Csn1 家族 [Acidaminococcus intestini RyC-MR95]
crispr-相关蛋白 [Acidaminococcus lactaris ATCC 29176]
假定蛋白 RUMLAC_00178 [Ruminococcus catus GD/7]
CRISPR-相关核酸内切酶, Csn1 家族 [Coprococcus catus GD/7]
假定蛋白 DORLON_02599 [Dorea longicatena DSM 13814]
CRISPR-相关的 Cas5e [齿垢密螺旋体 ATCC 35405]
CRISPR-相关蛋白 [Peptoniphilus duerdenii ATCC BAA-1640]
csn1 家族 CRISPR-相关蛋白 [Anaerococcus tetradius ATCC 35098]
CRISPR-相关蛋白, Csn1 家族 [消化乳杆菌乳糖 V 11V1-d]
假定蛋白 lpp0160 [嗜肺军团菌 str. Paris]
假定蛋白 WS1613 [Wolenella succnogenes DSM 1740]
酸 蛋白 [Francisella novicida U112]
假定蛋白 HMPREF9464_01128 [Sutterella wadsworthensis 3_1_45B]
CRISPR-相关的 大 蛋白 (部分), 推测的 [γ变形菌 HTCC5015]
推测的 CRISPR-相关蛋白, Csx12 家族 [Burkholderiales bacterium 1_1_47]
CRISPR-相关蛋白, Csx12 家族 [Parasutterella excrementihominis YIT 11859]

图39D

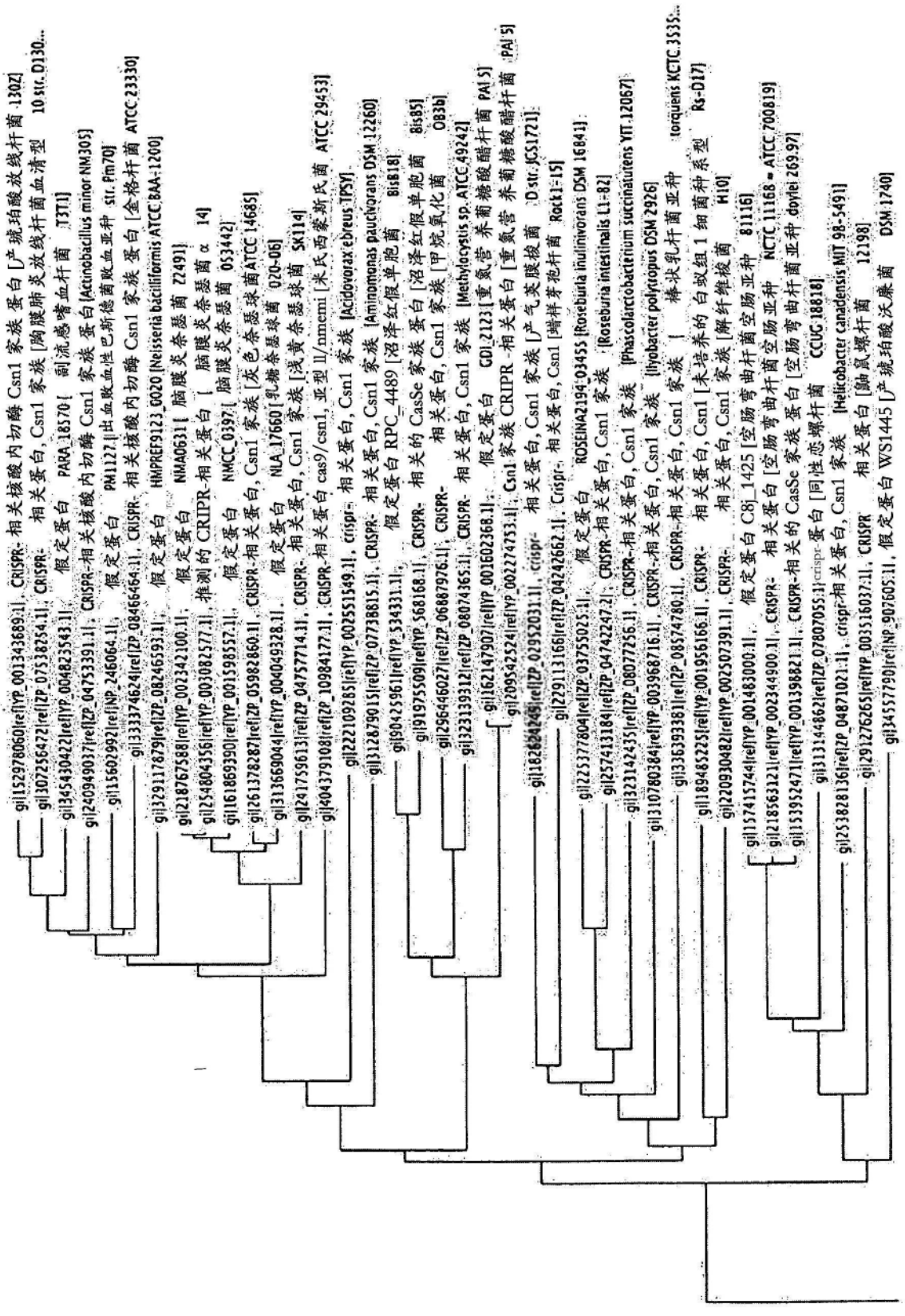


图40A

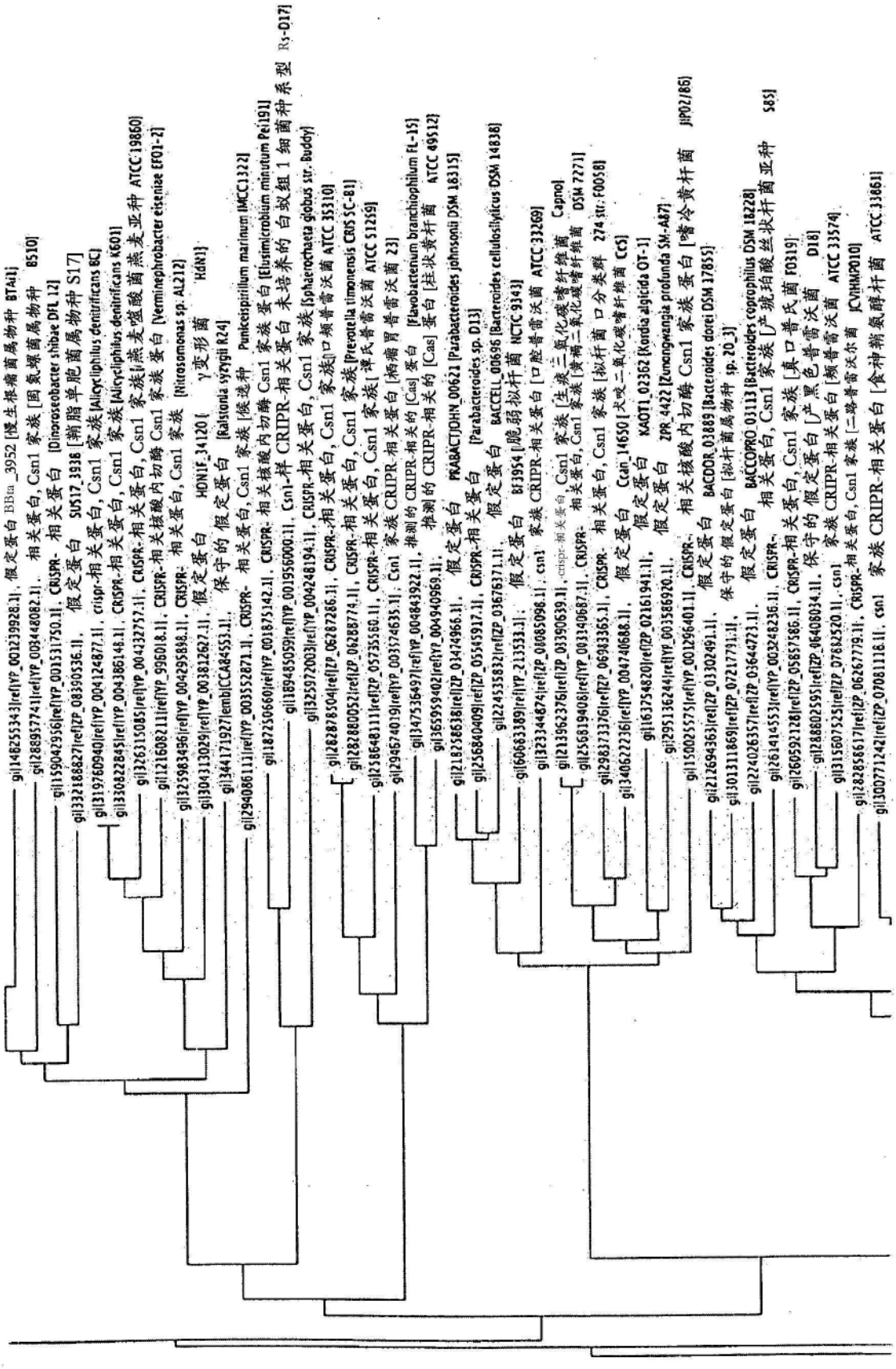


图40B

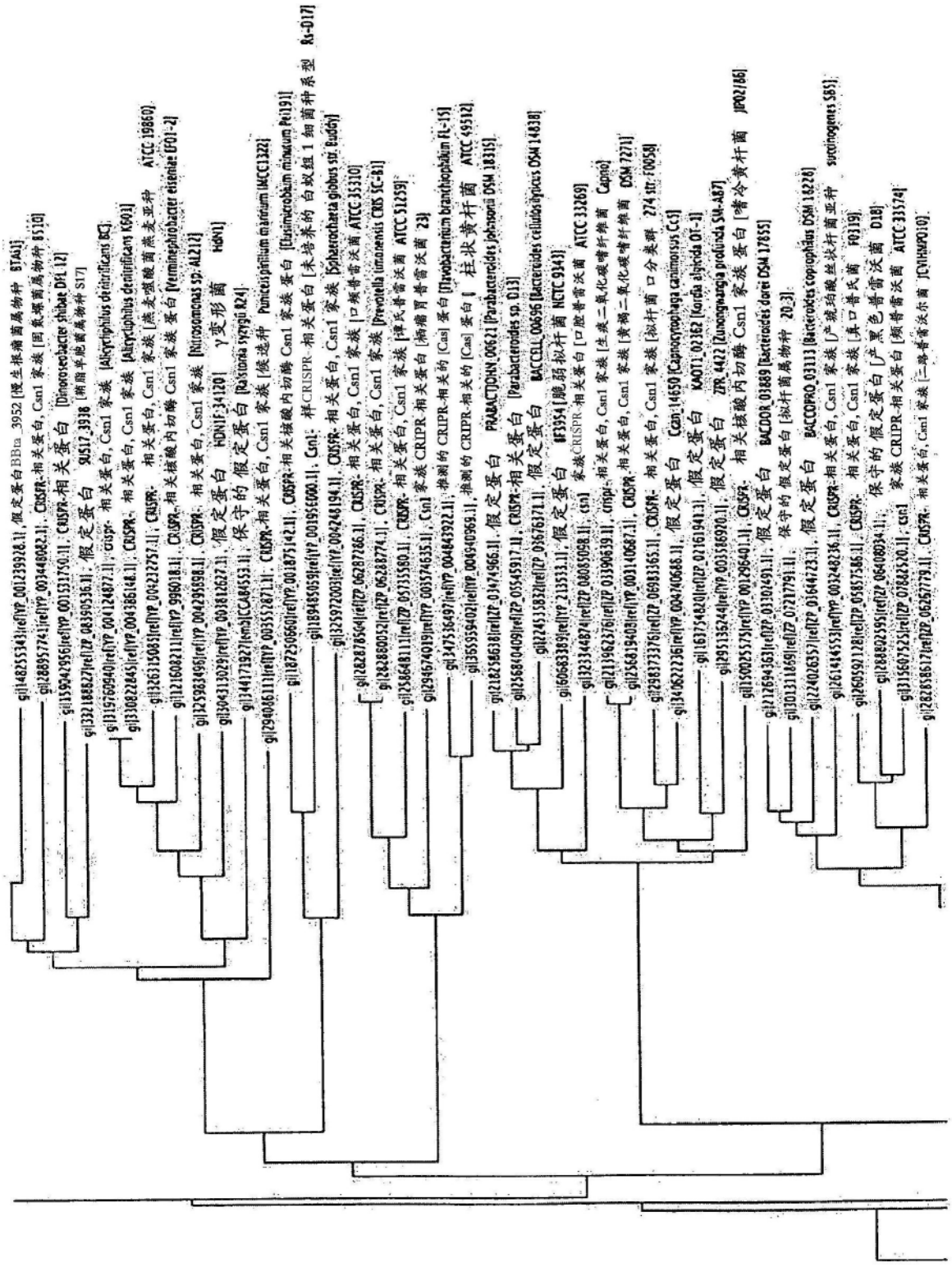


图40C

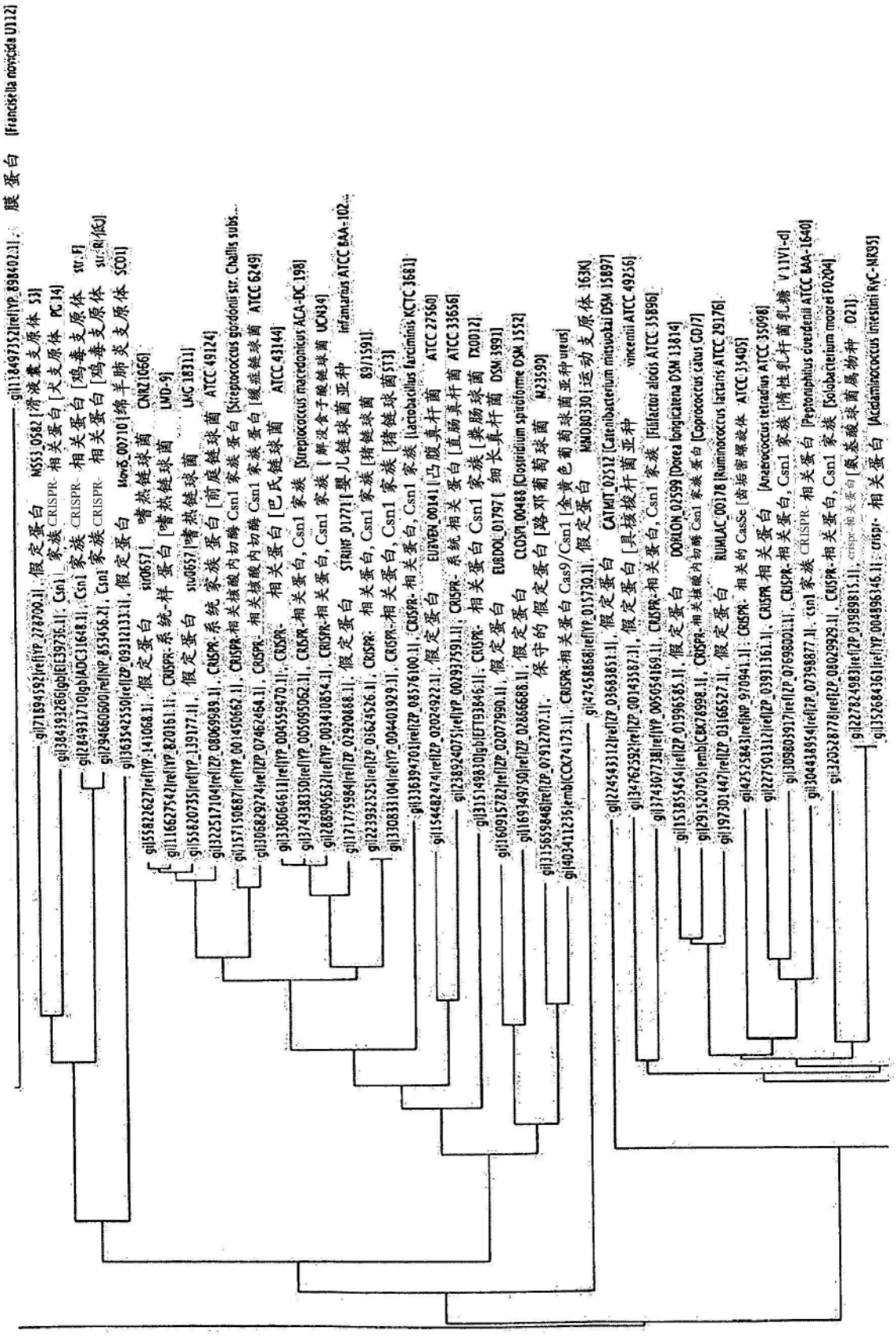


图40D

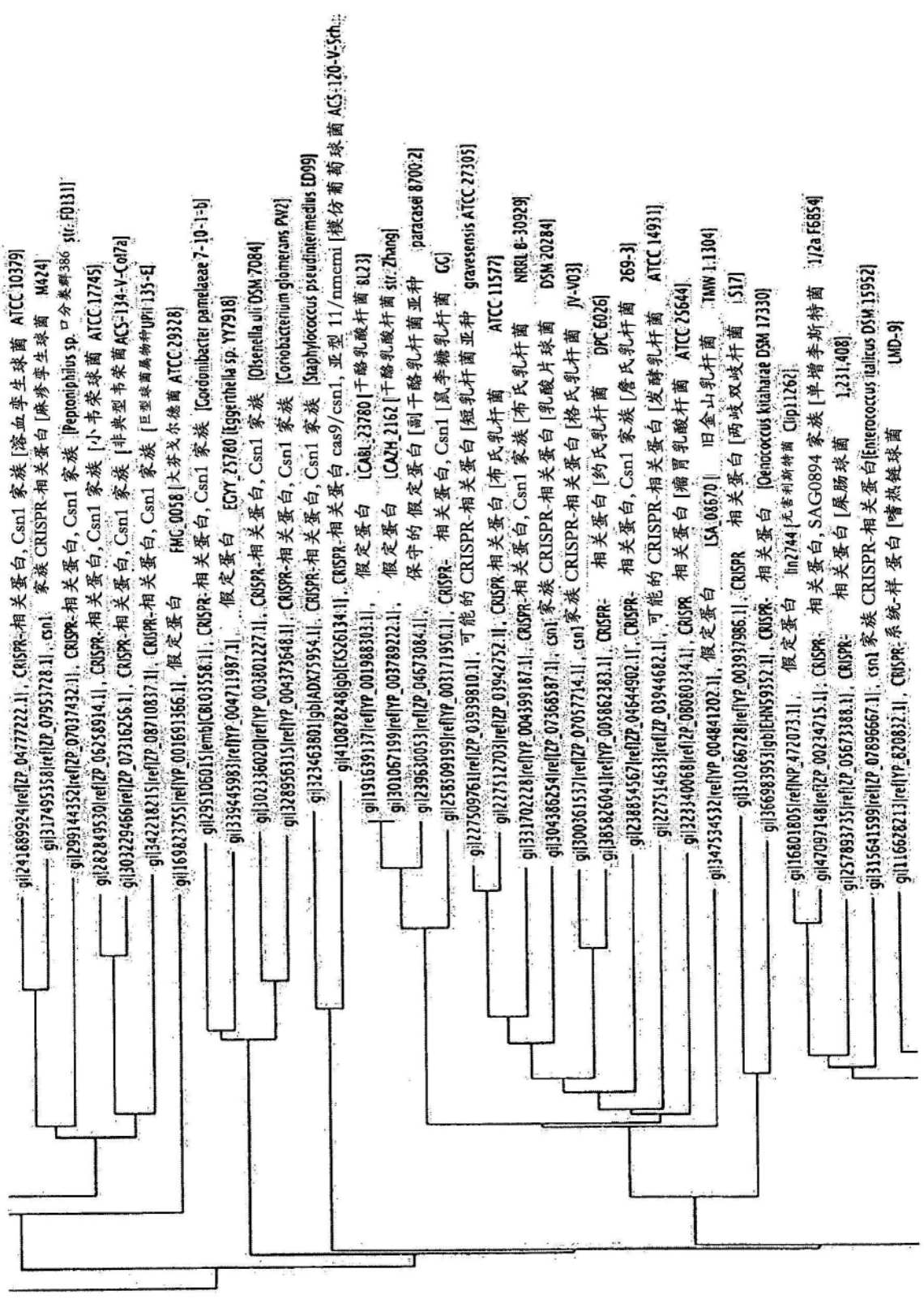


图40E

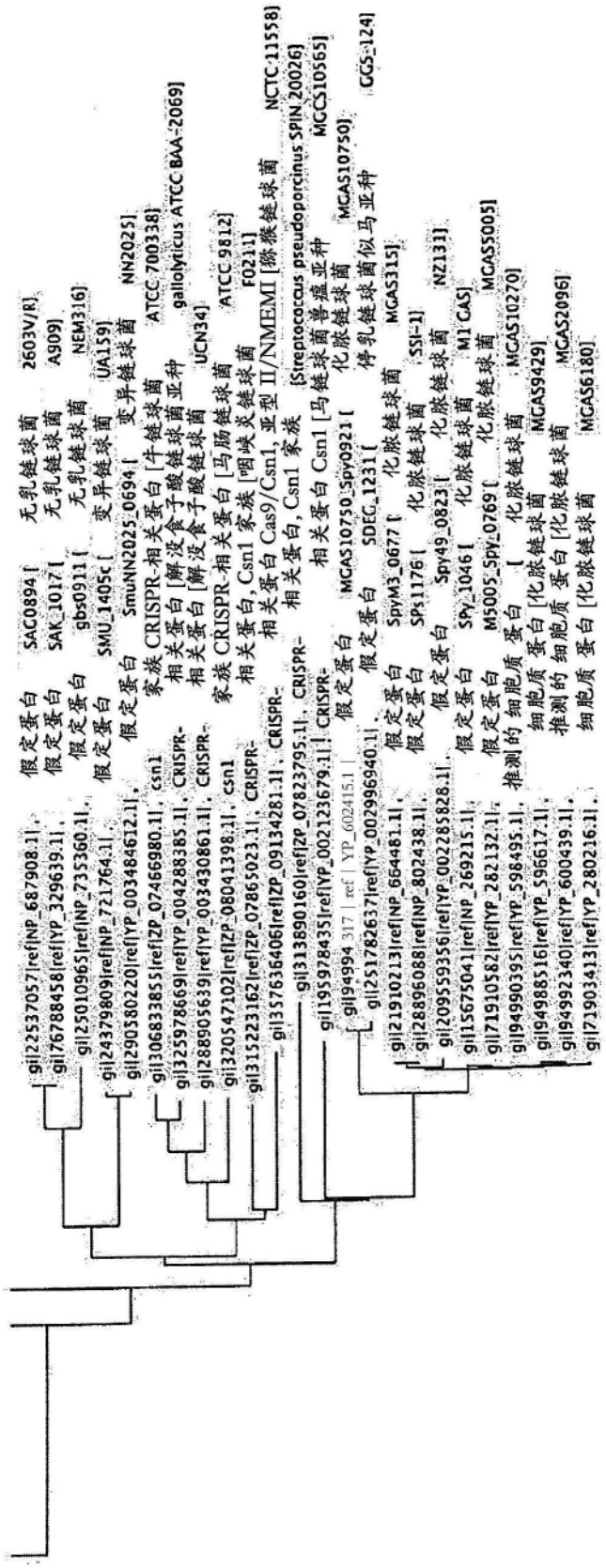

gi|22537057|ref|NP_687908.1|, 假定蛋白 SAG0894 [无乳链球菌 2603V/R]
gi|76788458|ref|YP_329639.1|, 假定蛋白 SAK_1017 [无乳链球菌 A909]
gi|25010965|ref|NP_735360.1|, 假定蛋白 gbs0911 [无乳链球菌 NEM316]
gi|24379809|ref|NP_721764.1|, 假定蛋白 SMU_1405c [变异链球菌 UA159]
gi|29058220|ref|YP_003484612.1|, 假定蛋白 SmuNN2025_0694 [变异链球菌 NN2025]
gi|30683385|ref|ZP_07466980.1|, csn1 家族 CRISPR-相关蛋白 [牛链球菌 ATCC 700338]
gi|32597866|ref|YP_004288385.1|, CRISPR-相关蛋白 [解没食子酸链球菌亚种 gallolyticus ATCC BAA-2069]
gi|28895639|ref|YP_003430861.1|, CRISPR-相关蛋白 [解没食子酸链球菌 UCN34]
gi|32054702|ref|ZP_08041398.1|, csn1 家族 CRISPR-相关蛋白 [马肠链球菌 ATCC 9812]
gi|31522316|ref|ZP_07865023.1|, CRISPR-相关蛋白,Csn1 家族 [咽峡炎链球菌 F0211]
gi|35763640|ref|ZP_09134281.1|, CRISPR-相关蛋白 Cas9/Csn1, 亚型 II/NMEMI [猕猴链球菌 NCTC 11558]
gi|31389016|ref|ZP_07823795.1|, CRISPR-相关蛋白, Csn1 家族 [Streptococcus pseudoporcinus SPIN 20026]
gi|19597843|ref|YP_002123679.1|, CRISPR-相关蛋白 Csn1 [马链球菌兽瘟亚种 MGCS10565]
gi|94994 317 | ref | YP_602415.1 | 假定蛋白 MGAS10750_Spy0921 [化脓链球菌 MGAS10750]
gi|25178263 7|ref|YP_002996940.1|, 假定蛋白 SDEG_1231 [停乳链球菌似马亚种 CGS_124]
gi|21910213|ref|NP_664481.1|, 假定蛋白 SpyM3_0677 [化脓链球菌 MGAS315]
gi|28896088|ref|NP_802438.1|, 假定蛋白 SP11176 [化脓链球菌 SSI-1]
gi|20955935 6|ref|YP_002285828.1|, 假定蛋白 Spy49_0823 [化脓链球菌 NZ131]
gi|15675041|ref|NP_269215.1|, 假定蛋白 SPy_1046 [化脓链球菌 M1 GAS]
gi|71910582|ref|YP_282132.1|, 假定蛋白 M5005_Spy_0769 [化脓链球菌 MGAS5005]
gi|94990395|ref|YP_596495.1|, 推测的细胞质蛋白 [化脓链球菌 MGAS10270]
gi|94988516|ref|YP_596617.1|, 细胞质蛋白 [化脓链球菌 MGAS9429]
gi|94992340|ref|YP_600439.1|, 推测的细胞质蛋白 [化脓链球菌 MGAS2096]
gi|71903413|ref|YP_280216.1|, 细胞质蛋白 [化脓链球菌 MGAS6180]

图40F

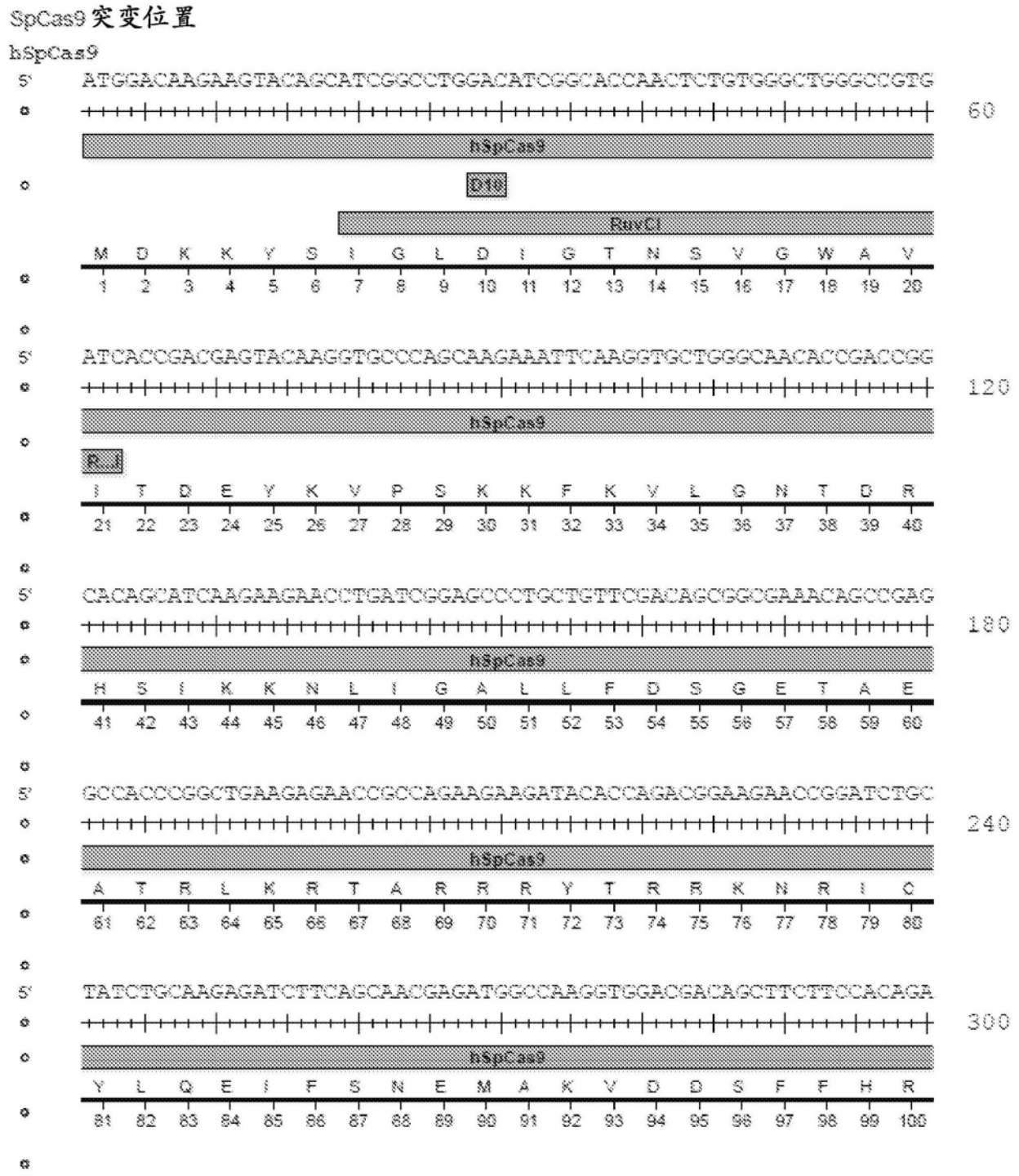

SpCas9突变位置
hSpCas9
5' ATGGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTG
60
hSpCas9
D10
RuvC I
M D K K Y S I G L D I G T N S V G W A V
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
5' ATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGG
120
hSpCas9
R...I
I T D E Y K V P S K K F K V L G N T D R
21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
5' CACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAG
180
hSpCas9
H S I K K N L I G A L L F D S G E T A E
41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
5' GCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGC
240
hSpCas9
A T R L K R T A R R R Y T R R K N R I C
61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80
5' TATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGA
300
hSpCas9
Y L Q E I F S N E M A K V D D S F F H R
81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

图41A

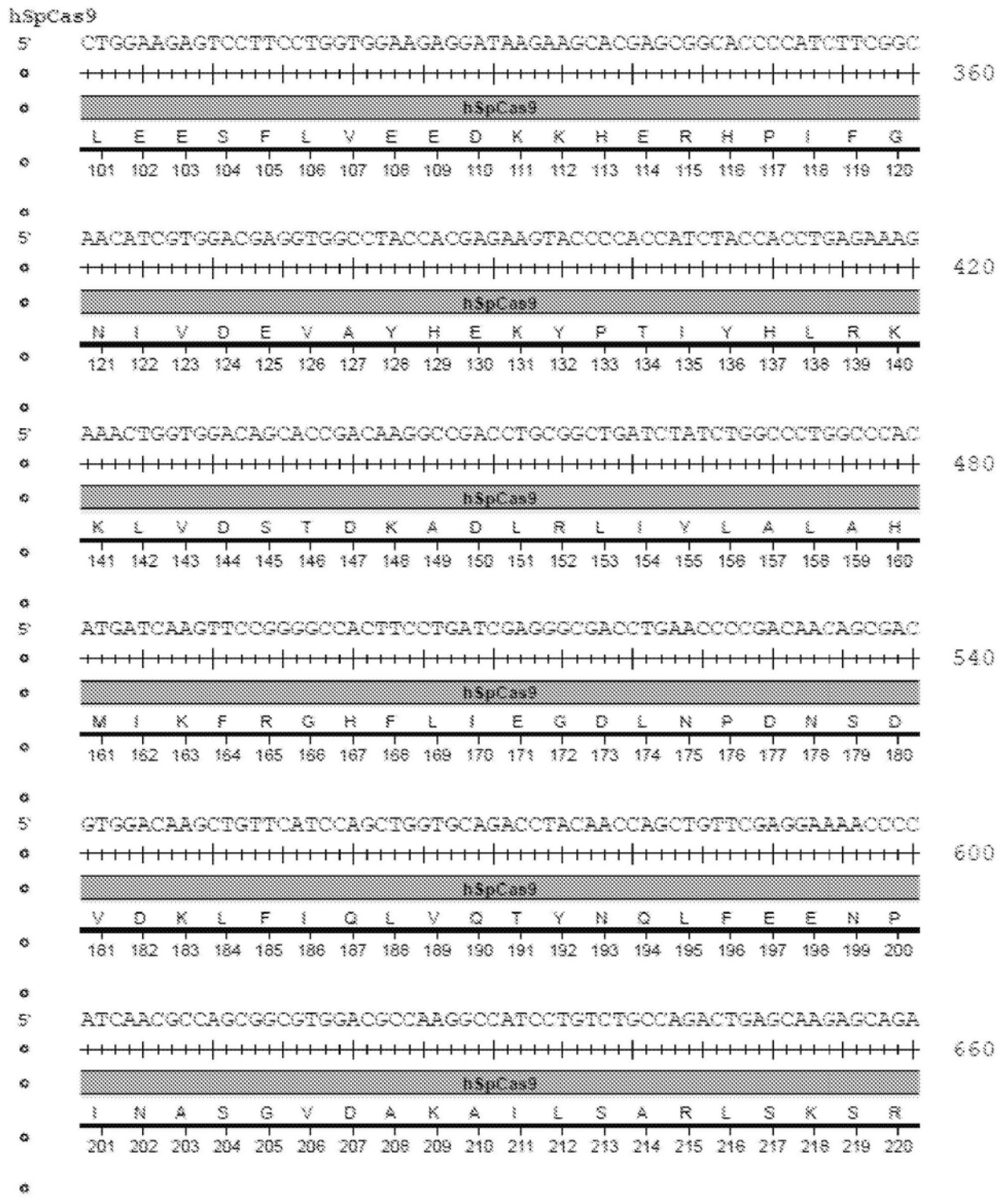

hSpCas9
5' CTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGC
360
hSpCas9
L E E S F L V E E D K K H E R H P I F G
101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120
5' AACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAG
420
hSpCas9
N I V D E V A Y H E K Y P T I Y H L R K
121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140
5' AAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCAC
480
hSpCas9
K L V D S T D K A D L R L I Y L A L A H
141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160
5' ATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGAC
540
hSpCas9
M I K F R G H F L I E G D L N P D N S D
161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180
5' GTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCC
600
hSpCas9
V D K L F I Q L V Q T Y N Q L F E E N P
181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200
5' ATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGA
660
hSpCas9
I N A S G V D A K A I L S A R L S K S R
201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220

图41B

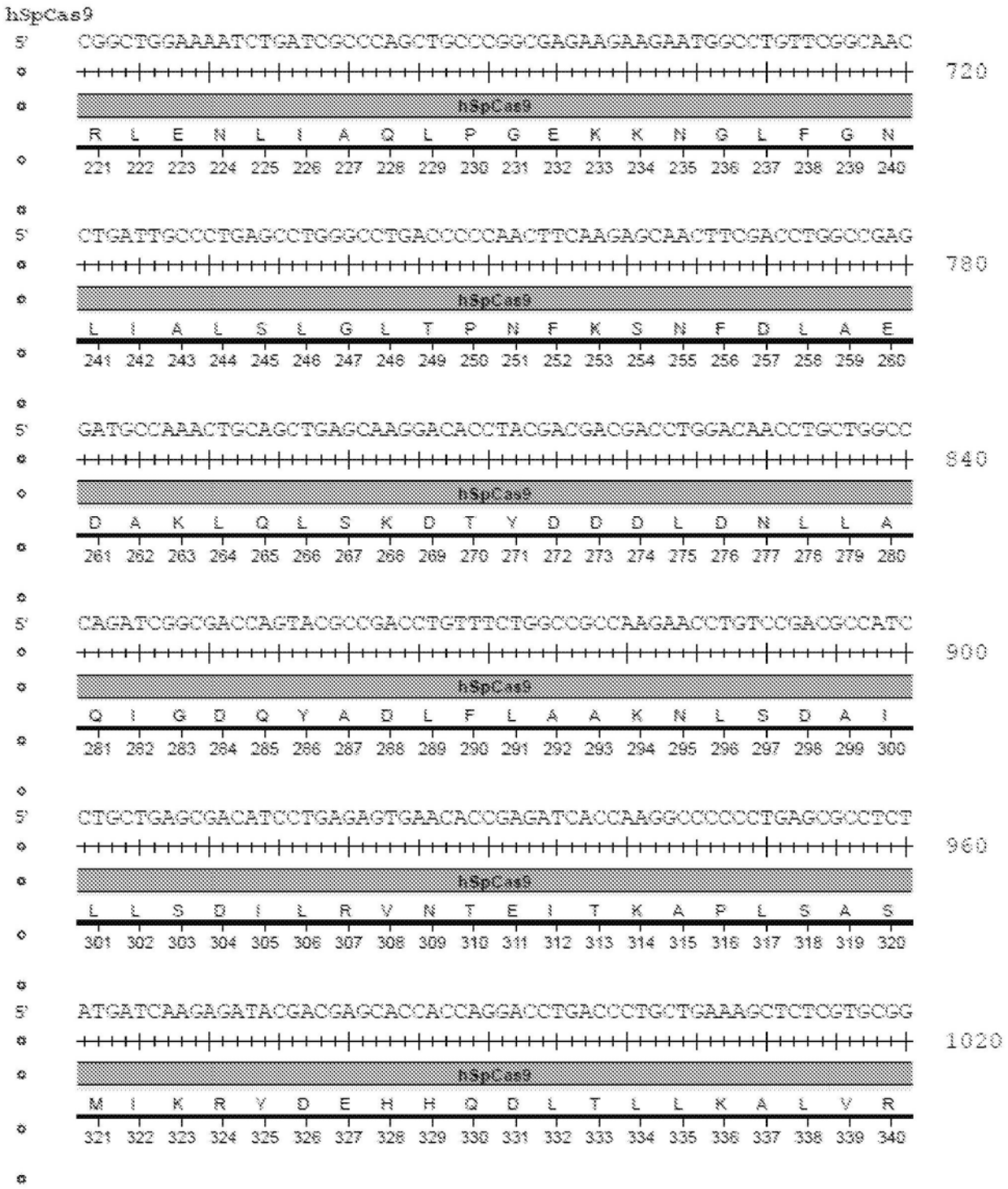

hSpCas9
5' CGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGCAAC
720
hSpCas9
R L E N L I A Q L P G E K K N G L F G N
221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240
5' CTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAG
780
hSpCas9
L I A L S L G L T P N F K S N F D L A E
241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260
5' GATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCC
840
hSpCas9
D A K L Q L S K D T Y D D D L D N L L A
261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280
5' CAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATC
900
hSpCas9
Q I G D Q Y A D L F L A A K N L S D A I
281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300
5' CTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCT
960
hSpCas9
L L S D I L R V N T E I T K A P L S A S
301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320
5' ATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGG
1020
hSpCas9
M I K R Y D E H H Q D L T L L K A L V R
321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340

图41C

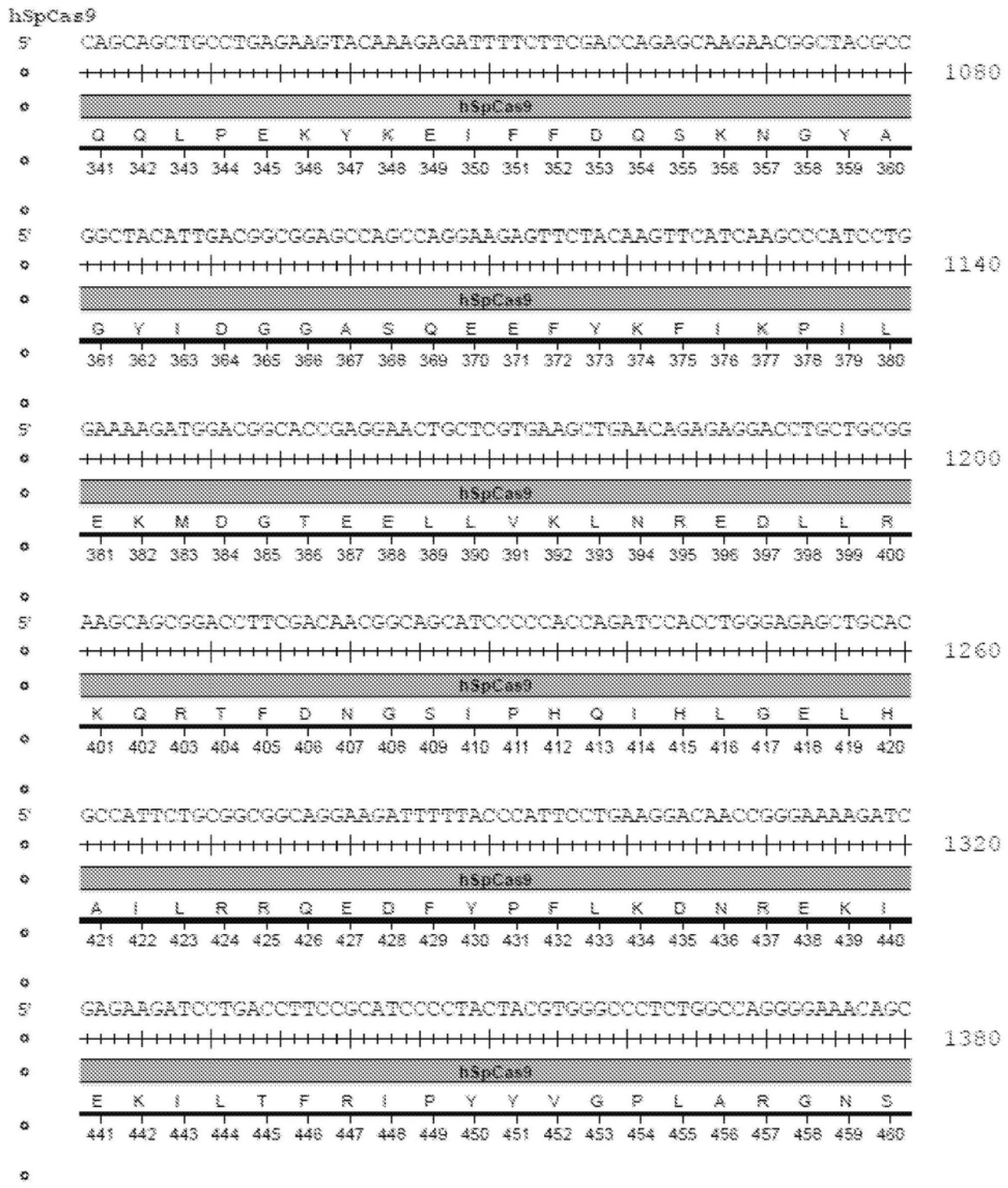

hSpCas9
5' CAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCC
1080
hSpCas9
Q Q L P E K Y K E I F F D Q S K N G Y A
341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360
5' GGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTG
1140
hSpCas9
G Y I D G G A S Q E E F Y K F I K P I L
361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380
5' GAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGG
1200
hSpCas9
E K M D G T E E L L V K L N R E D L L R
381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400
5' AAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCAC
1260
hSpCas9
K Q R T F D N G S I P H Q I H L G E L H
401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420
5' GCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATC
1320
hSpCas9
A I L R R Q E D F Y P F L K D N R E K I
421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440
5' GAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGC
1380
hSpCas9
E K I L T F R I P Y Y V G P L A R G N S
441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460

图41D

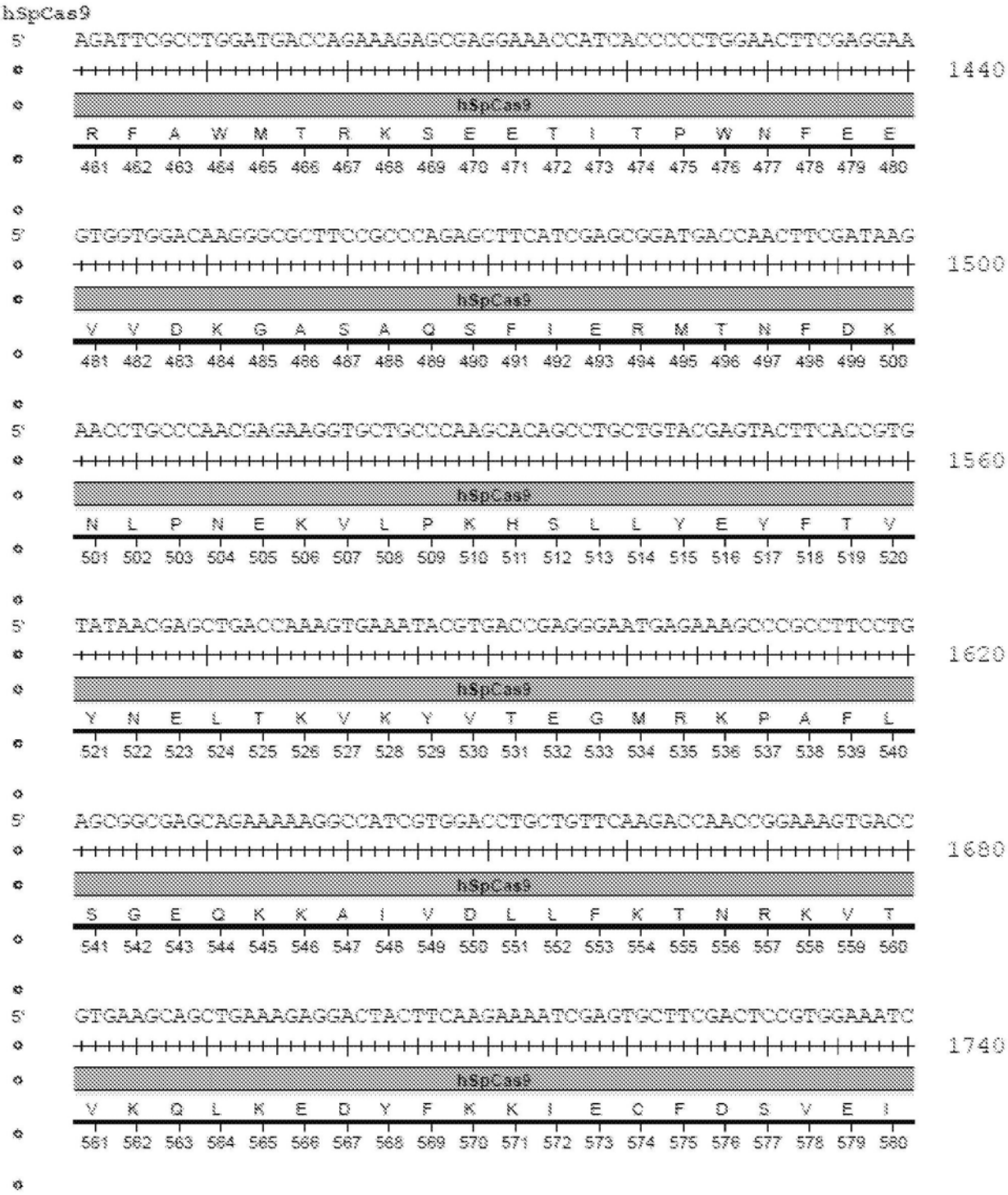

hSpCas9
5' AGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAA
1440
hSpCas9
R F A W M T R K S E E T I T P W N F E E
461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480
5' GTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAG
1500
hSpCas9
V V D K G A S A Q S F I E R M T N F D K
481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500
5' AACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTG
1560
hSpCas9
N L P N E K V L P K H S L L Y E Y F T V
501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520
5' TATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTG
1620
hSpCas9
Y N E L T K V K Y V T E G M R K P A F L
521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540
5' AGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACC
1680
hSpCas9
S G E Q K K A I V D L L F K T N R K V T
541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560
5' GTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATC
1740
hSpCas9
V K Q L K E D Y F K K I E C F D S V E I
561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580

图41E

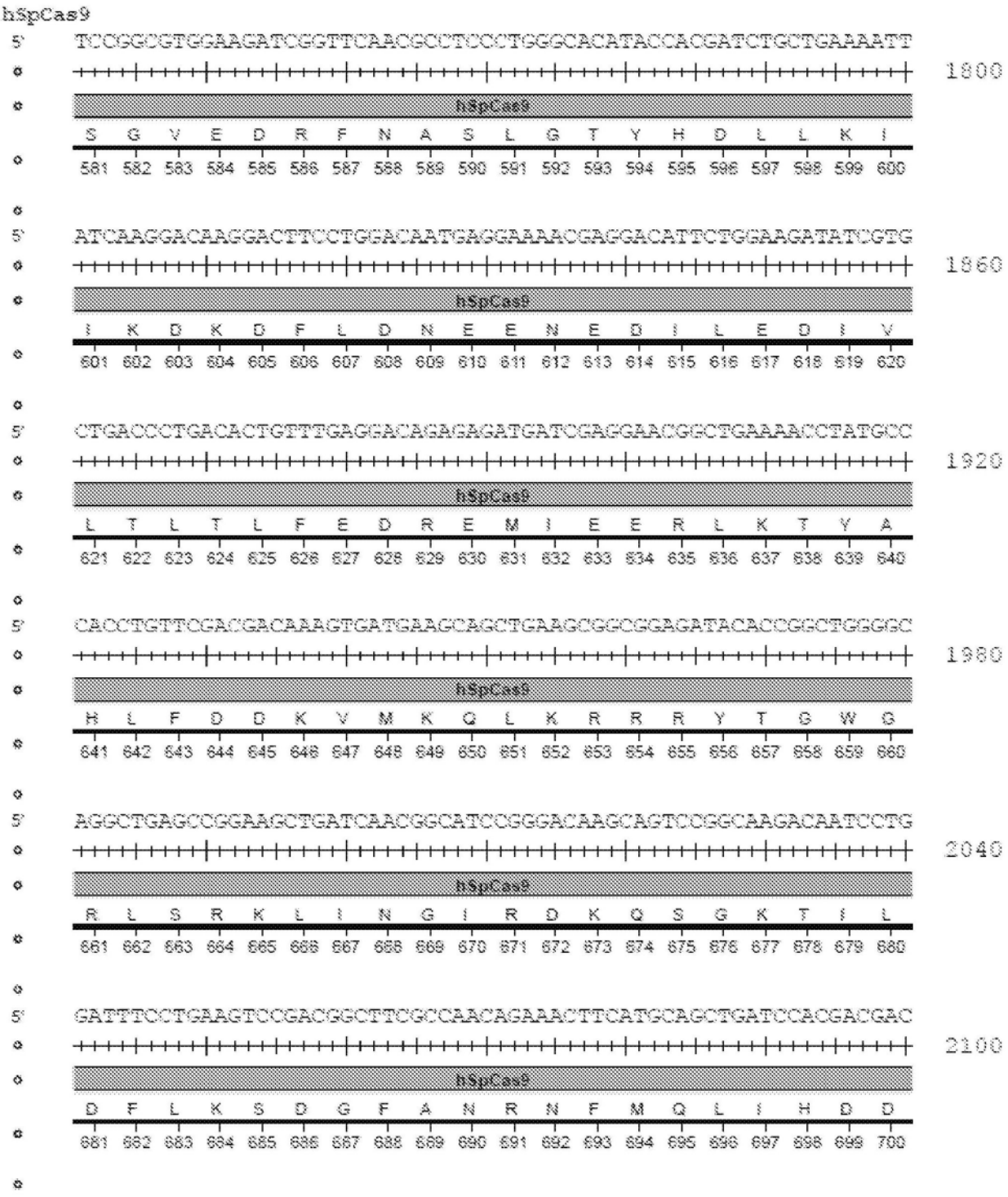

hSpCas9
5' TCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATT
1800
hSpCas9
S G V E D R F N A S L G T Y H D L L K I
581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600
5' ATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTG
1860
hSpCas9
I K D K D F L D N E E N E D I L E D I V
601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620
5' CTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCC
1920
hSpCas9
L T L T L F E D R E M I E E R L K T Y A
621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640
5' CACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGC
1980
hSpCas9
H L F D D K V M K Q L K R R R Y T G W G
641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660
5' AGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTG
2040
hSpCas9
R L S R K L I N G I R D K Q S G K T I L
661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680
5' GATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGAC
2100
hSpCas9
D F L K S D G F A N R N F M Q L I H D D
681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700

图41F

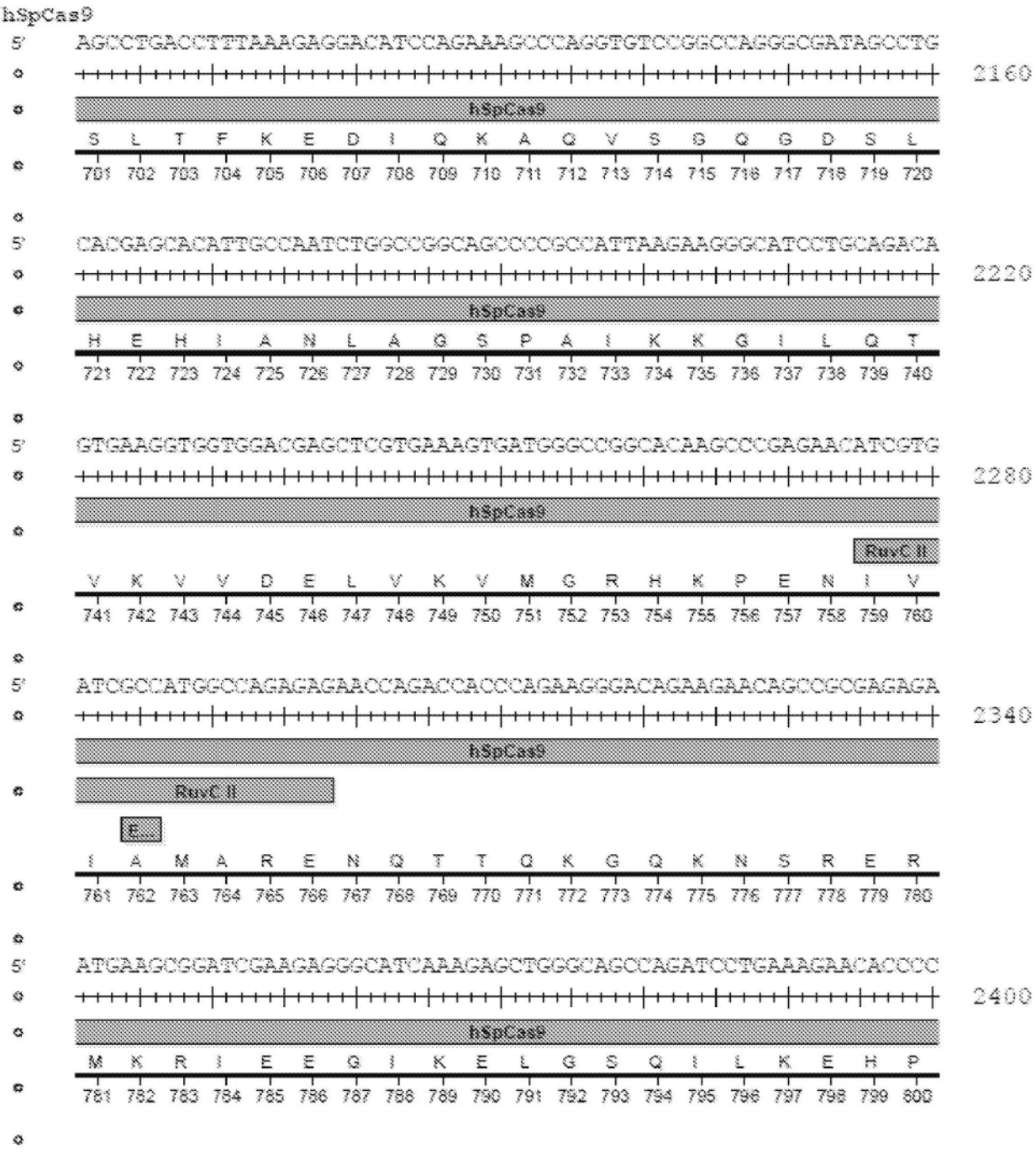

hSpCas9
5' AGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTG
2160
hSpCas9
S L T F K E D I Q K A Q V S G Q G D S L
701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720
5' CACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACA
2220
hSpCas9
H E H I A N L A G S P A I K K G I L Q T
721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740
5' GTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTG
2280
hSpCas9
RuvC II
V K V V D E L V K V M G R H K P E N I V
741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760
5' ATCGCCATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGA
2340
hSpCas9
RuvC II
E...
I A M A R E N Q T T Q K G Q K N S R E R
761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780
5' ATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCC
2400
hSpCas9
M K R I E E G I K E L G S Q I L K E H P
781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800

图41G

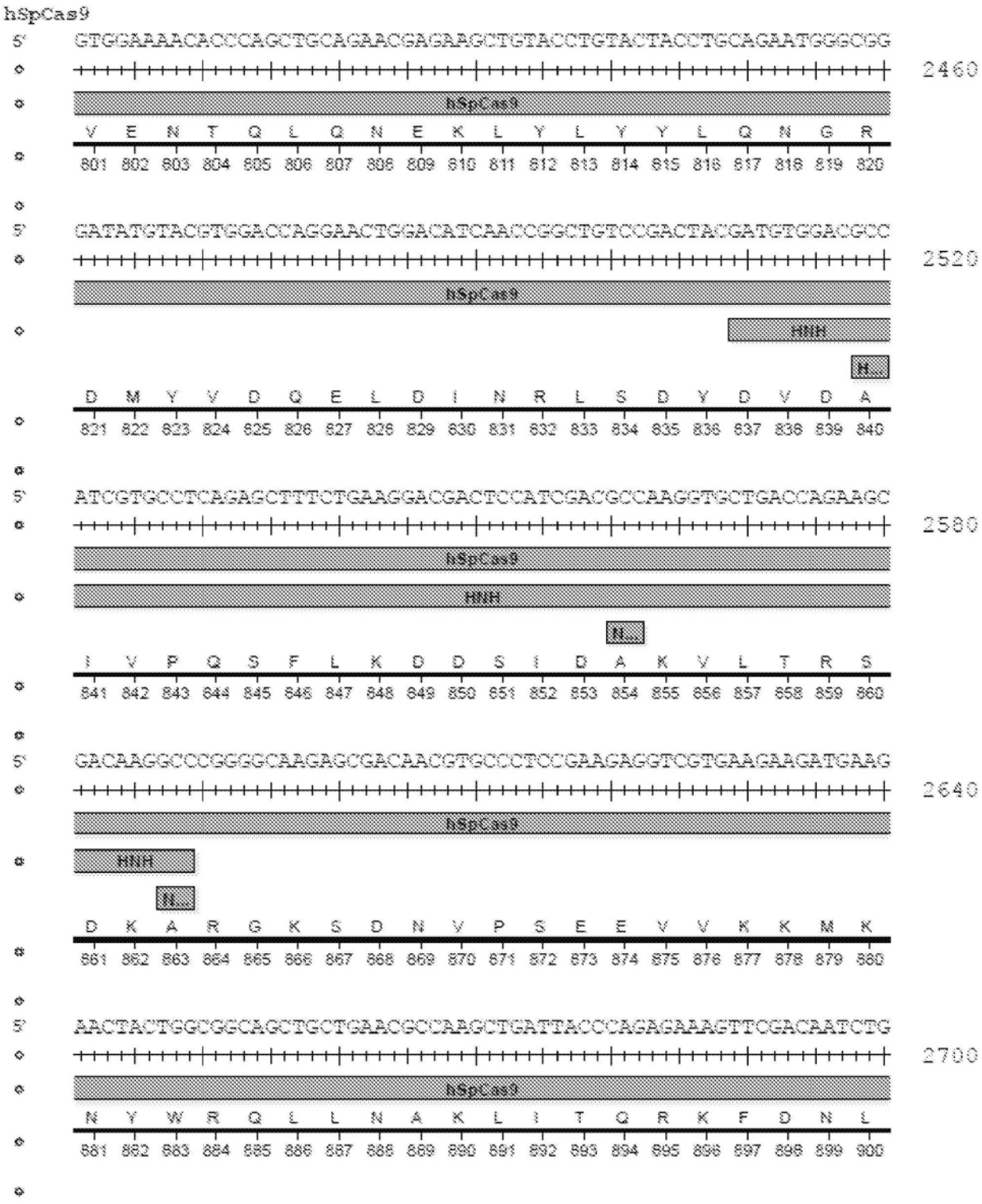

hSpCas9
5' GTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGG
2460
hSpCas9
V E N T Q L Q N E K L Y L Y Y L Q N G R
801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820
5' GATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACGCC
2520
hSpCas9
HNH
H...
D M Y V D Q E L D I N R L S D Y D V D A
821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840
5' ATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACGCCAAGGTGCTGACCAGAAGC
2580
hSpCas9
HNH
N...
I V P Q S F L K D D S I D A K V L T R S
841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860
5' GACAAGGCCCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAG
2640
hSpCas9
HNH
N...
D K A R G K S D N V P S E E V V K K M K
861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880
5' AACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTG
2700
hSpCas9
N Y W R Q L L N A K L I T Q R K F D N L
881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900

图41H

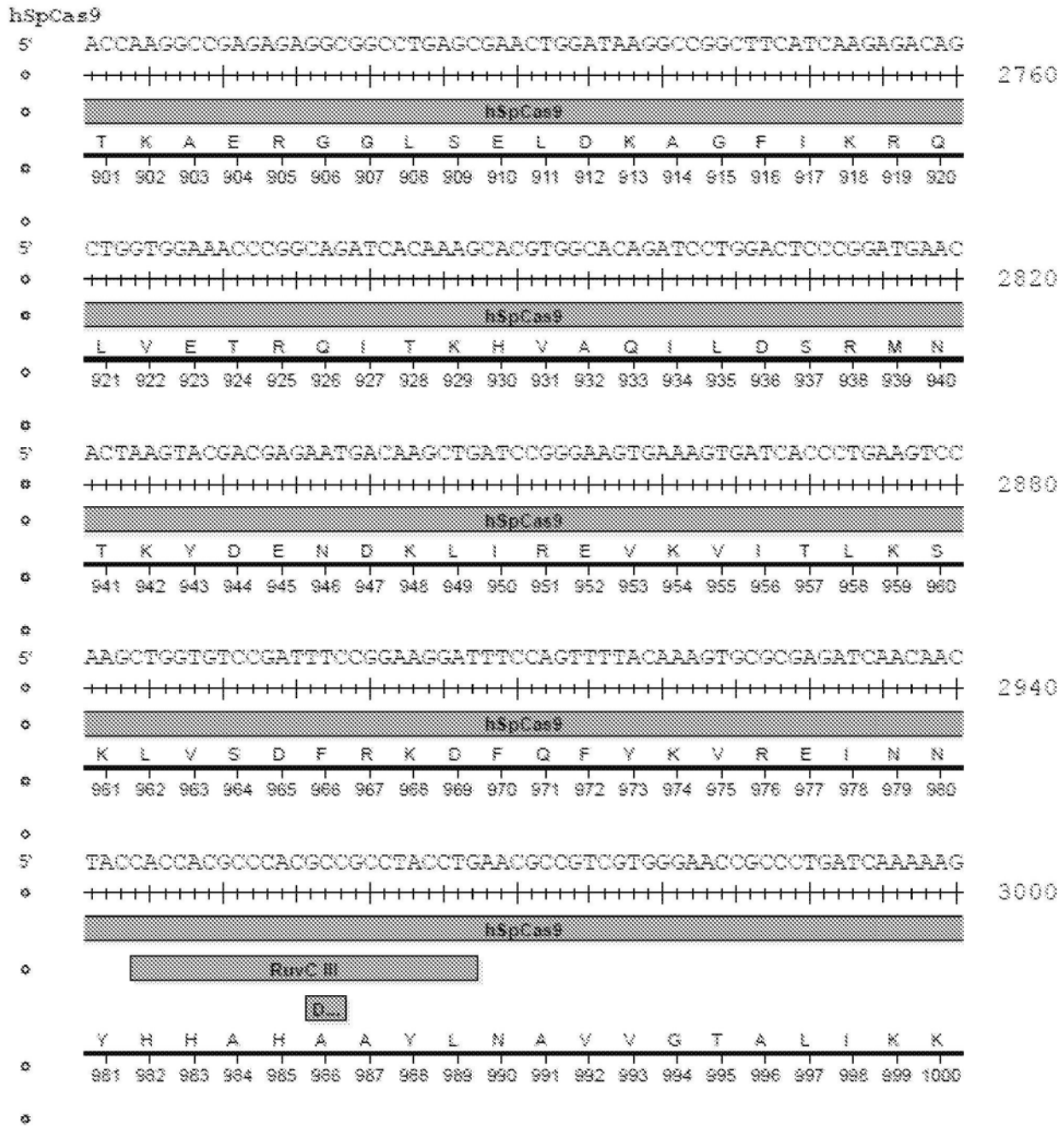

hSpCas9
5'
ACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAG
2760
hSpCas9
T K A E R G G L S E L D K A G F I K R Q
901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920
5'
CTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAAC
2820
hSpCas9
L V E T R Q I T K H V A Q I L D S R M N
921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940
5'
ACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCC
2880
hSpCas9
T K Y D E N D K L I R E V K V I T L K S
941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960
5'
AAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAAC
2940
hSpCas9
K L V S D F R K D F Q F Y K V R E I N N
961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980
5'
TACCACCACGCCCACGCCGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAG
3000
hSpCas9
RuvC III
D...
Y H H A H A A Y L N A V V G T A L I K K
981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

图41I

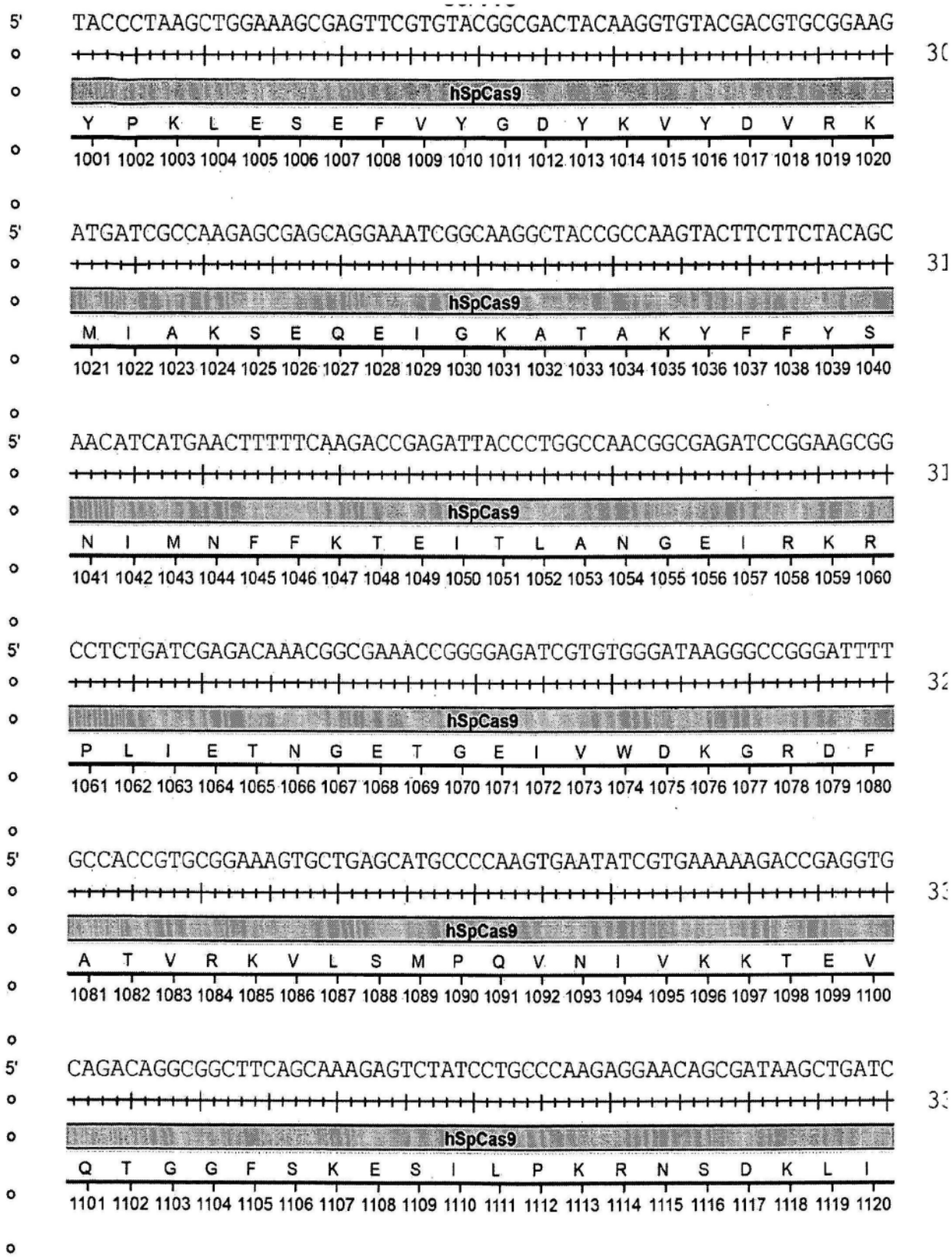

5' TACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAG
hSpCas9
Y P K L E S E F V Y G D Y K V Y D V R K
1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020
5' ATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGC
hSpCas9
M I A K S E Q E I G K A T A K Y F F Y S
1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040
5' AACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGG
hSpCas9
N I M N F F K T E I T L A N G E I R K R
1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1060
5' CCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTT
hSpCas9
P L I E T N G E T G E I V W D K G R D F
1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072 1073 1074 1075 1076 1077 1078 1079 1080
5' GCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTG
hSpCas9
A T V R K V L S M P Q V N I V K K T E V
1081 1082 1083 1084 1085 1086 1087 1088 1089 1090 1091 1092 1093 1094 1095 1096 1097 1098 1099 1100
5' CAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATC
hSpCas9
Q T G G F S K E S I L P K R N S D K L I
1101 1102 1103 1104 1105 1106 1107 1108 1109 1110 1111 1112 1113 1114 1115 1116 1117 1118 1119 1120

图41J

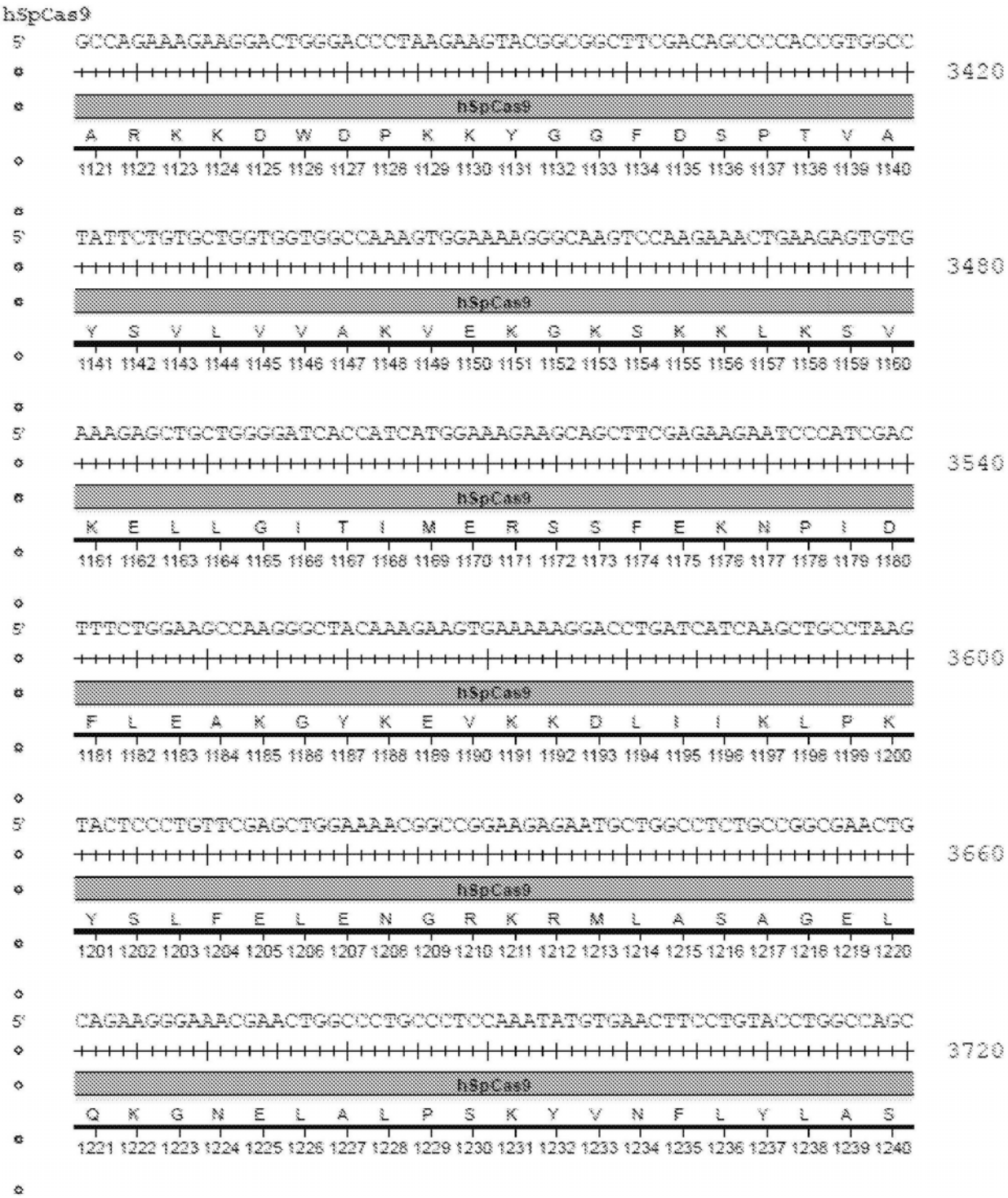

hSpCas9
5' GCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCC
3420
hSpCas9
A R K K D W D P K K Y G G F D S P T V A
1121 1122 1123 1124 1125 1126 1127 1128 1129 1130 1131 1132 1133 1134 1135 1136 1137 1138 1139 1140
5' TATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTG
3480
hSpCas9
Y S V L V V A K V E K G K S K K L K S V
1141 1142 1143 1144 1145 1146 1147 1148 1149 1150 1151 1152 1153 1154 1155 1156 1157 1158 1159 1160
5' AAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGAC
3540
hSpCas9
K E L L G I T I M E R S S F E K N P I D
1161 1162 1163 1164 1165 1166 1167 1168 1169 1170 1171 1172 1173 1174 1175 1176 1177 1178 1179 1180
5' TTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAG
3600
hSpCas9
F L E A K G Y K E V K K D L I I K L P K
1181 1182 1183 1184 1185 1186 1187 1188 1189 1190 1191 1192 1193 1194 1195 1196 1197 1198 1199 1200
5' TACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTG
3660
hSpCas9
Y S L F E L E N G R K R M L A S A G E L
1201 1202 1203 1204 1205 1206 1207 1208 1209 1210 1211 1212 1213 1214 1215 1216 1217 1218 1219 1220
5' CAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGC
3720
hSpCas9
Q K G N E L A L P S K Y V N F L Y L A S
1221 1222 1223 1224 1225 1226 1227 1228 1229 1230 1231 1232 1233 1234 1235 1236 1237 1238 1239 1240

图41K

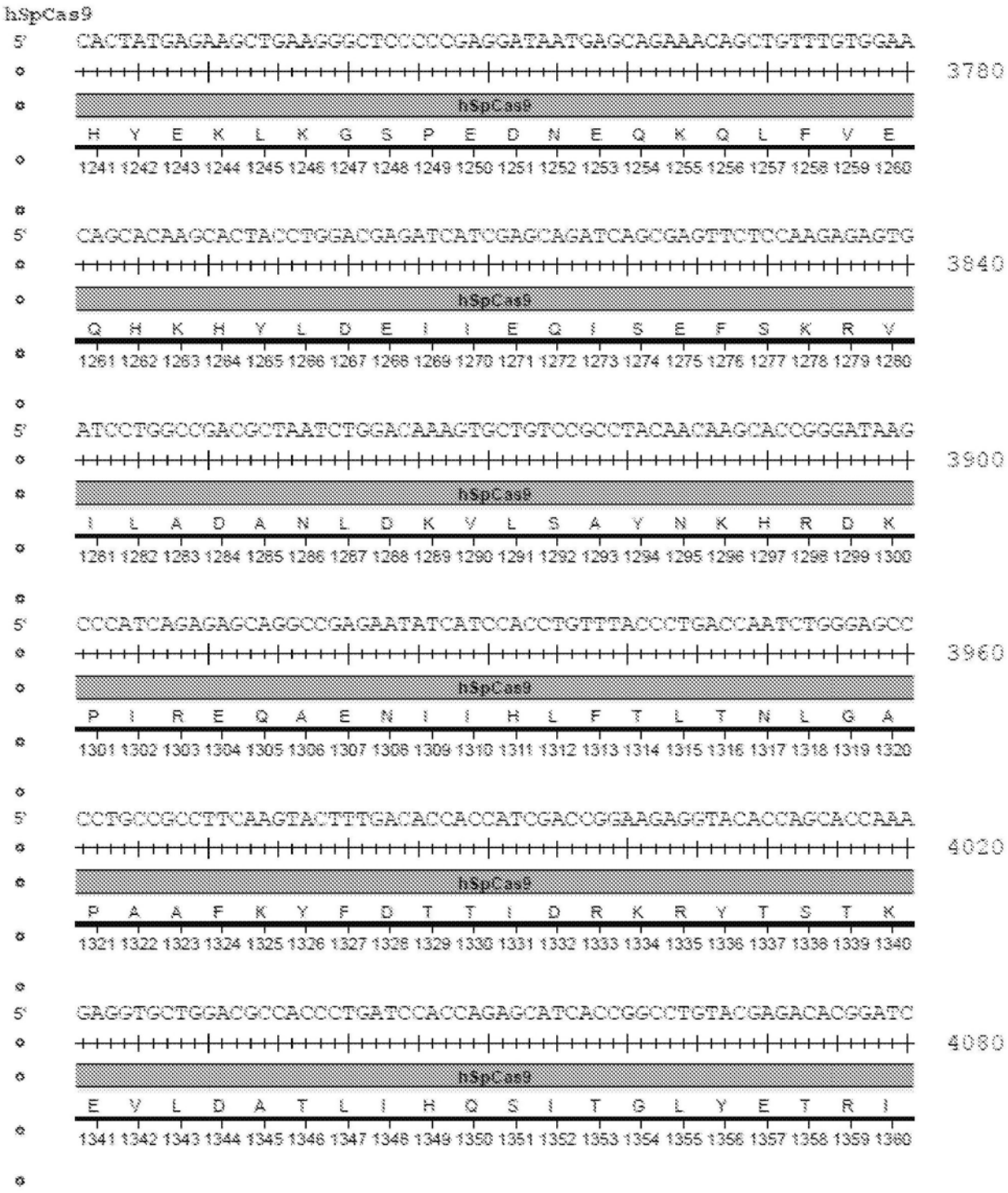

hSpCas9
5' CACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAA
3780
hSpCas9
H Y E K L K G S P E D N E Q K Q L F V E
1241 1242 1243 1244 1245 1246 1247 1248 1249 1250 1251 1252 1253 1254 1255 1256 1257 1258 1259 1260
5' CAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTG
3840
hSpCas9
Q H K H Y L D E I I E Q I S E F S K R V
1261 1262 1263 1264 1265 1266 1267 1268 1269 1270 1271 1272 1273 1274 1275 1276 1277 1278 1279 1280
5' ATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAG
3900
hSpCas9
I L A D A N L D K V L S A Y N K H R D K
1281 1282 1283 1284 1285 1286 1287 1288 1289 1290 1291 1292 1293 1294 1295 1296 1297 1298 1299 1300
5' CCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCC
3960
hSpCas9
P I R E Q A E N I I H L F T L T N L G A
1301 1302 1303 1304 1305 1306 1307 1308 1309 1310 1311 1312 1313 1314 1315 1316 1317 1318 1319 1320
5' CCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAA
4020
hSpCas9
P A A F K Y F D T T I D R K R Y T S T K
1321 1322 1323 1324 1325 1326 1327 1328 1329 1330 1331 1332 1333 1334 1335 1336 1337 1338 1339 1340
5' GAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATC
4080
hSpCas9
E V L D A T L I H Q S I T G L Y E T R I
1341 1342 1343 1344 1345 1346 1347 1348 1349 1350 1351 1352 1353 1354 1355 1356 1357 1358 1359 1360

图41L

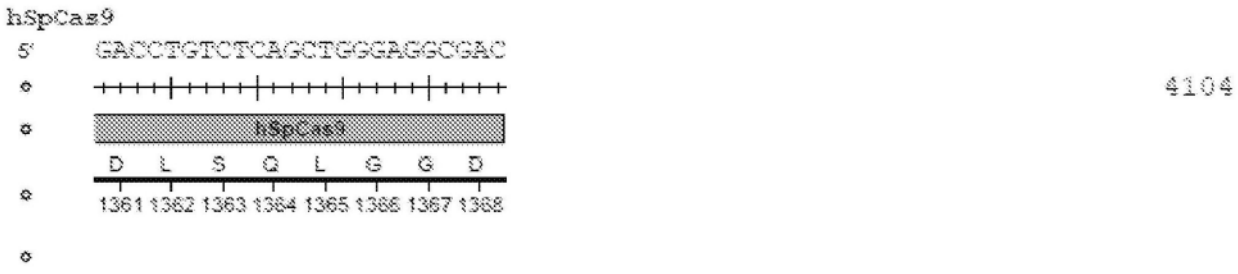


图41M

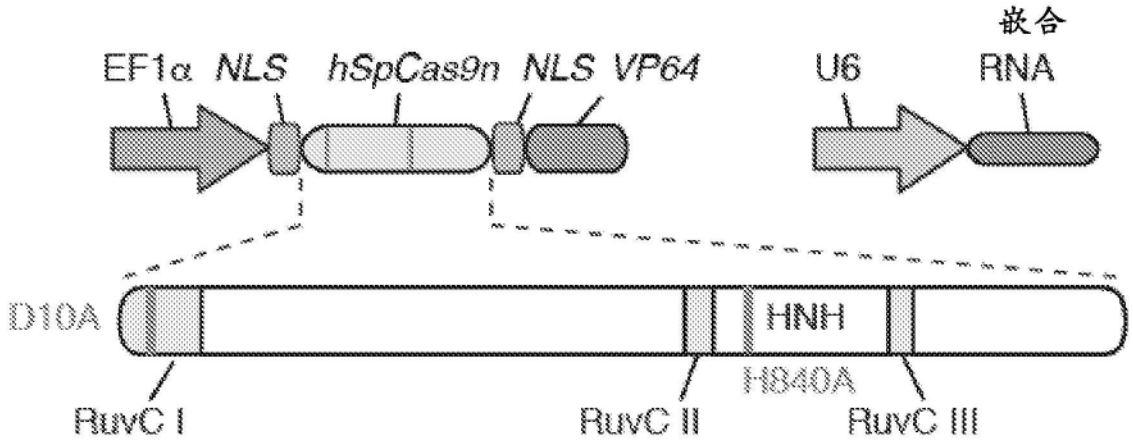


图42

a

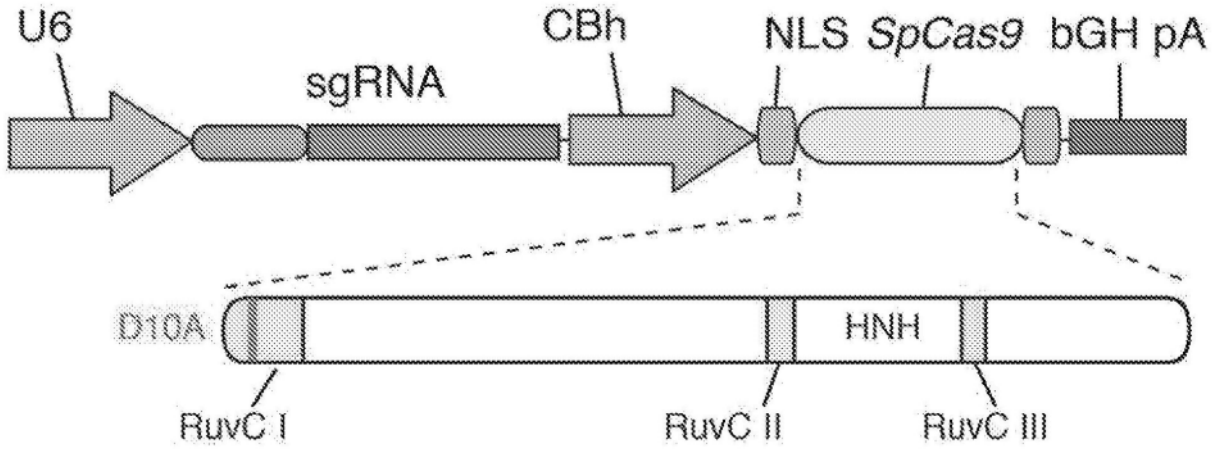


图43A

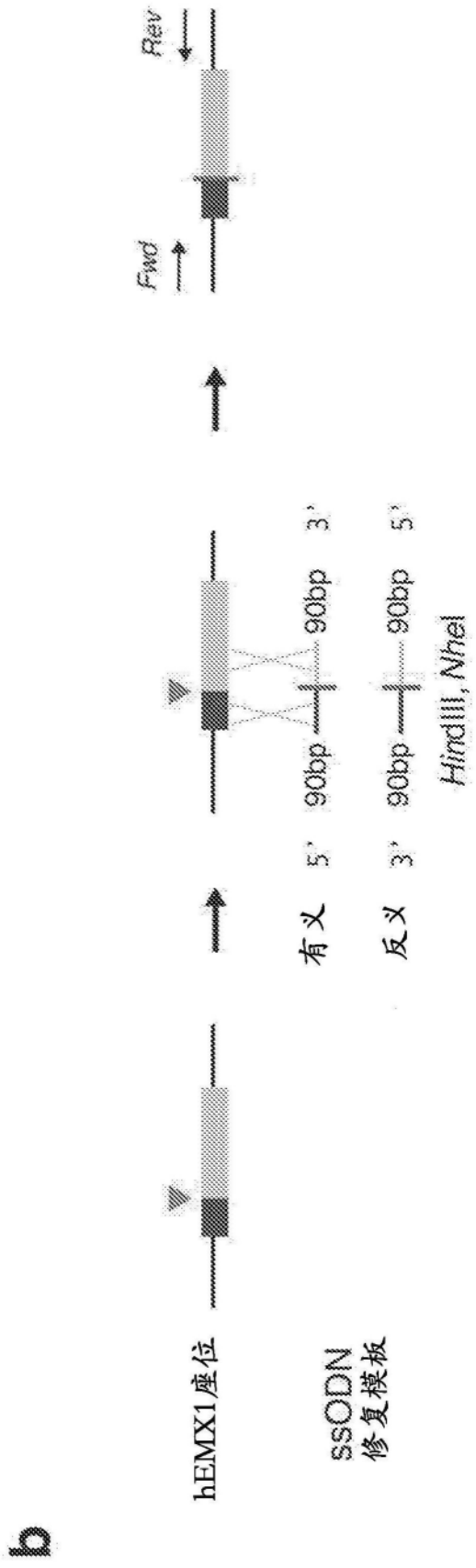

b
hEMX1座位
ssODN
修复模板
有义 5' 90bp 90bp 3'
反义 3' 90bp 90bp 5'
HindIII, NheI
Fwd
Rev

图43B

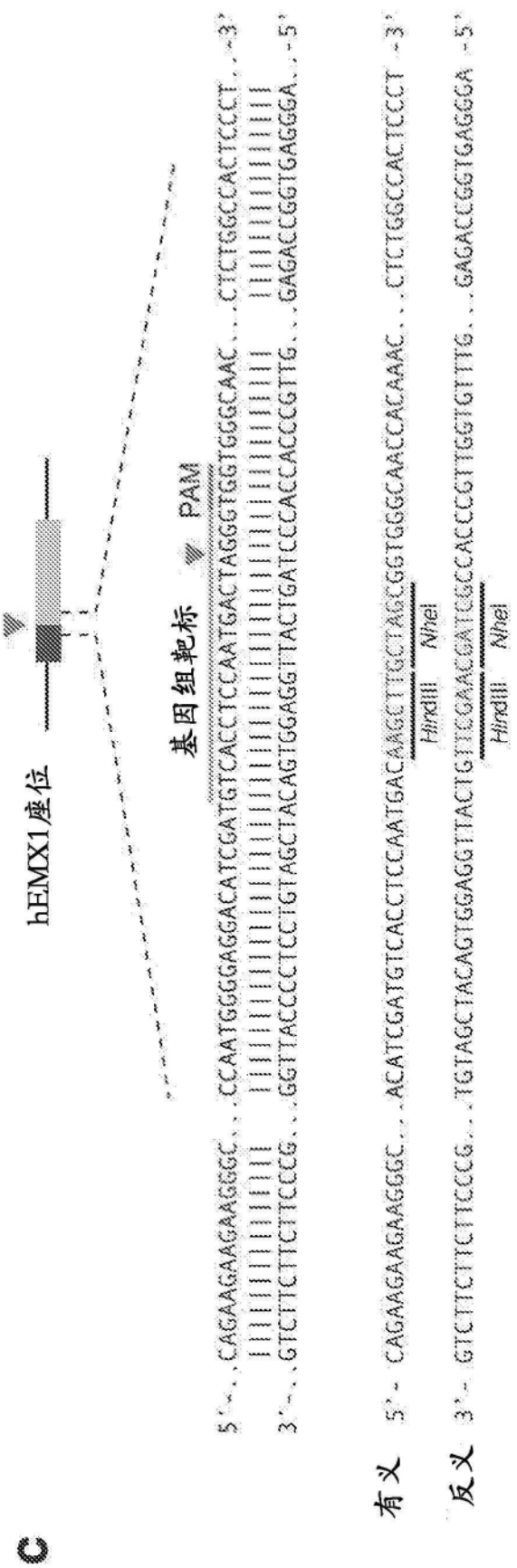

C
hEMX1座位
基因组靶标
PAM
5'-..CAGAAGAAGAAGGGC...CCAATGGGGAGGACATCGATGTCACCTCCAATGACTAGGGTGGTGGGCAAC...CTCTGGCCACTCCCT..-3'
3'-..GTCTTCTTCTTCCCG...GGTTACCCCTCCTGTAGCTACAGTGGAGGTTACTGATCCCACCACCCGTTG...GAGACCGGTGAGGGA..-5'
有义 5'- CAGAAGAAGAAGGGC...ACATCGATGTCACCTCCAATGACAAGCTTGCTAGCGGTGGGCAACCACAAAC...CTCTGGCCACTCCCT -3'
HindIII NheI
反义 3'- GTCTTCTTCTTCCCG...TGTAGCTACAGTGGAGGTTACTGTTCGAACGATCGCCACCCGTTGGTGTTTG...GAGACCGGTGAGGGA -5'
HindIII NheI

图43C

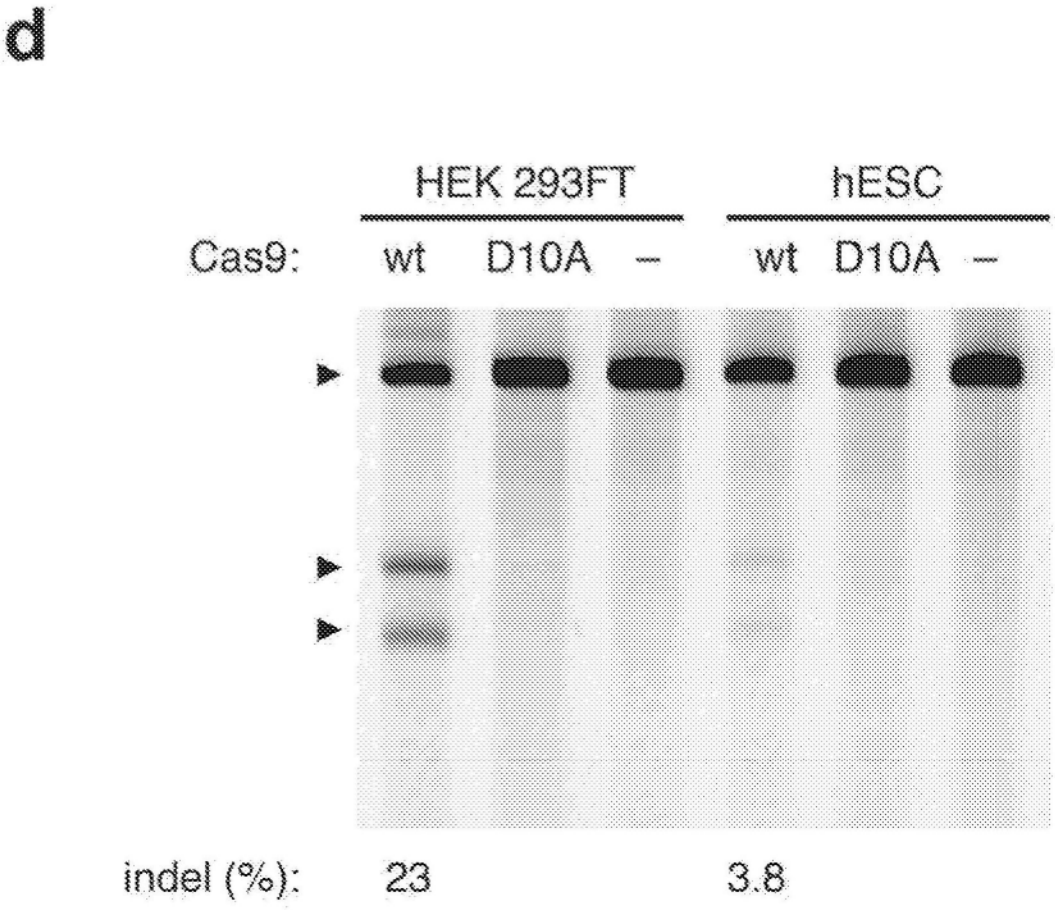


图43D

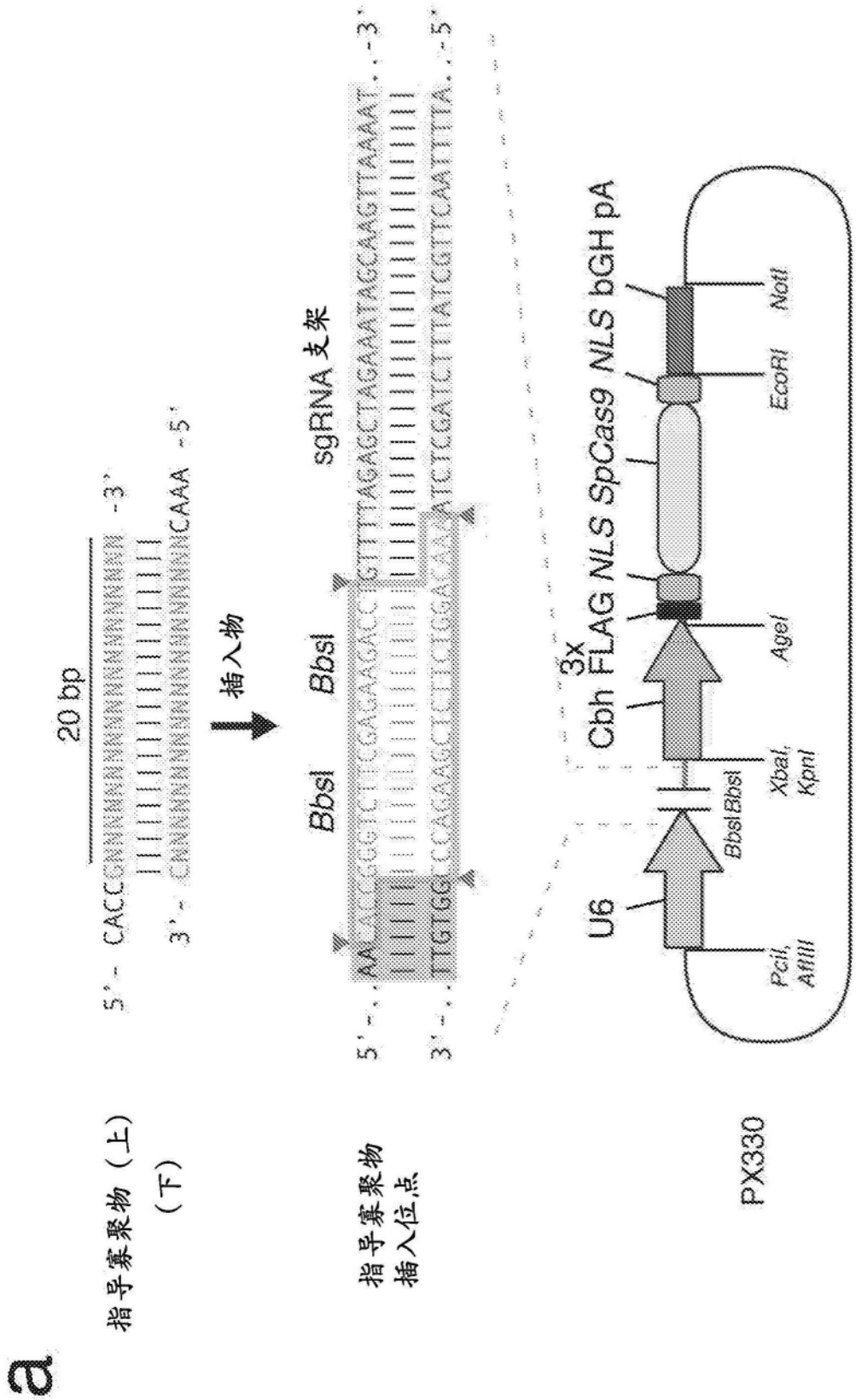

a
指导寡聚物（上）
（下）
20 bp
5'- CACCNNNNNNNNNNNNNNNNNNNN -3'
3'- CNNNNNNNNNNNNNNNNNNNNCAAA -5'
插入物
BbsI
BbsI
sgRNA 支架
5'-..AACACCGGGTCTTCGAGAAGACCGTTTTAGAGCTAGAAATAGCAAGTTAAAAT..-3'
3'-..TTGTGGCCCAGAAGCTCTTCTGGCAAAATCTCGATCTTTATCGTTCAATTTTA..-5'
指导寡聚物
插入位点
U6
3x
Cbh
FLAG NLS SpCas9 NLS bGH pA
PciI, AflIII
BbsI BbsI
XbaI, KpnI
AgeI
EcoRI
NotI
PX330

图44A

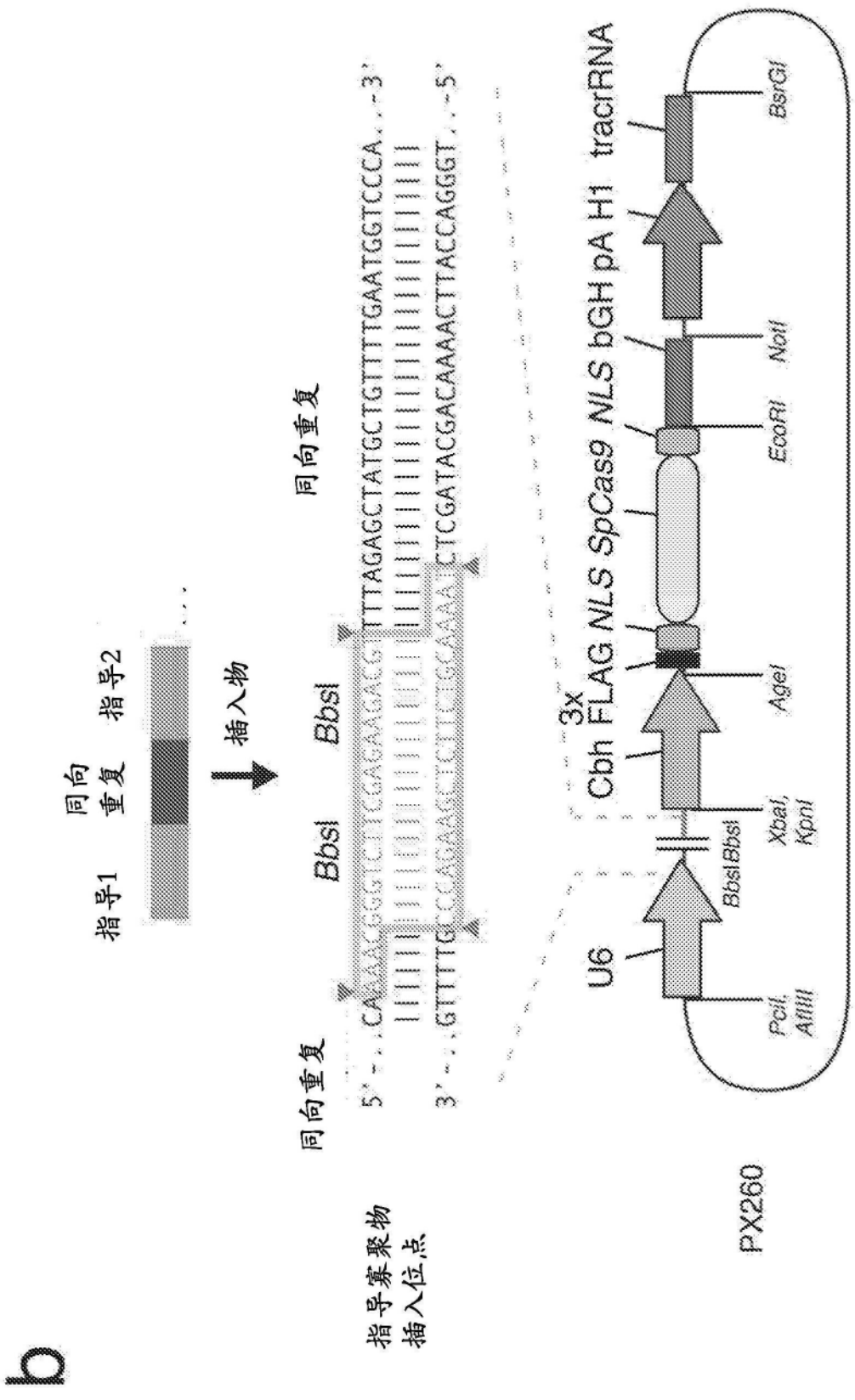

b
指导1
同向
重复
指导2
插入物
同向重复
BbsI
BbsI
同向重复
5'-...CACCGGTCTTCGAGAAGACGTTTTAGAGCTATGCTGTTTTGAATGGTCCCA...-3'
3'-...GTGGCCAGAAGCTCTTCTGCAAAATCTCGATACGACAAAACTTACCAGGGT...-5'
指导寡聚物
插入位点
U6
3x
Cbh FLAG NLS SpCas9 NLS bGH pA H1 tracrRNA
PX260
PciI, AflIII
BbsI BbsI
XbaI, KpnI
AgeI
EcoRI
NotI
BsrGI

图44B

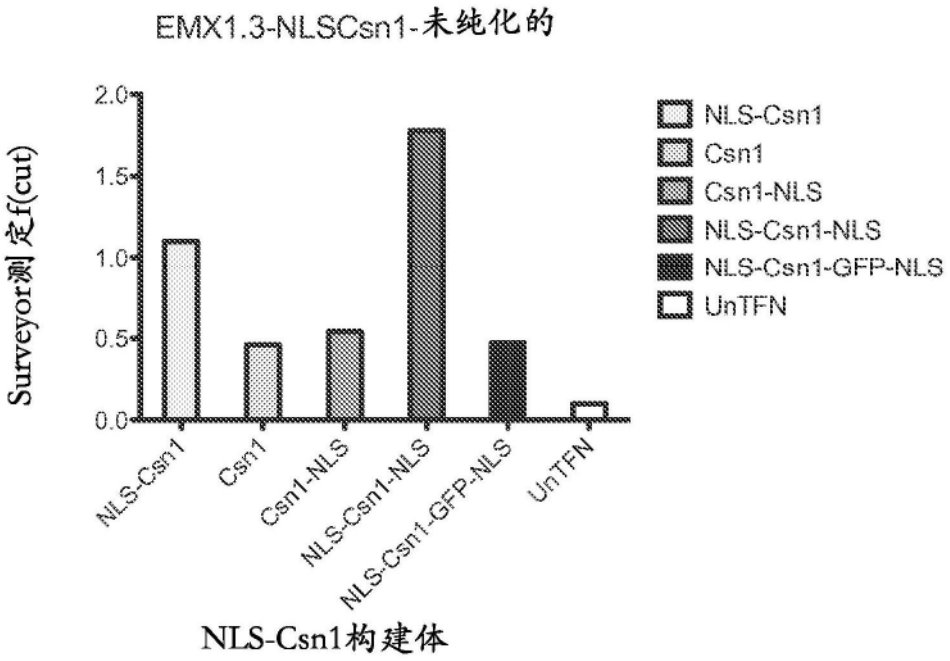

EMX1.3-NLSCsn1-未纯化的
Surveyor测定f(cut)
2.0
1.5
1.0
0.5
0.0
NLS-Csn1
Csn1
Csn1-NLS
NLS-Csn1-NLS
NLS-Csn1-GFP-NLS
UnTFN
NLS-Csn1构建体

图45

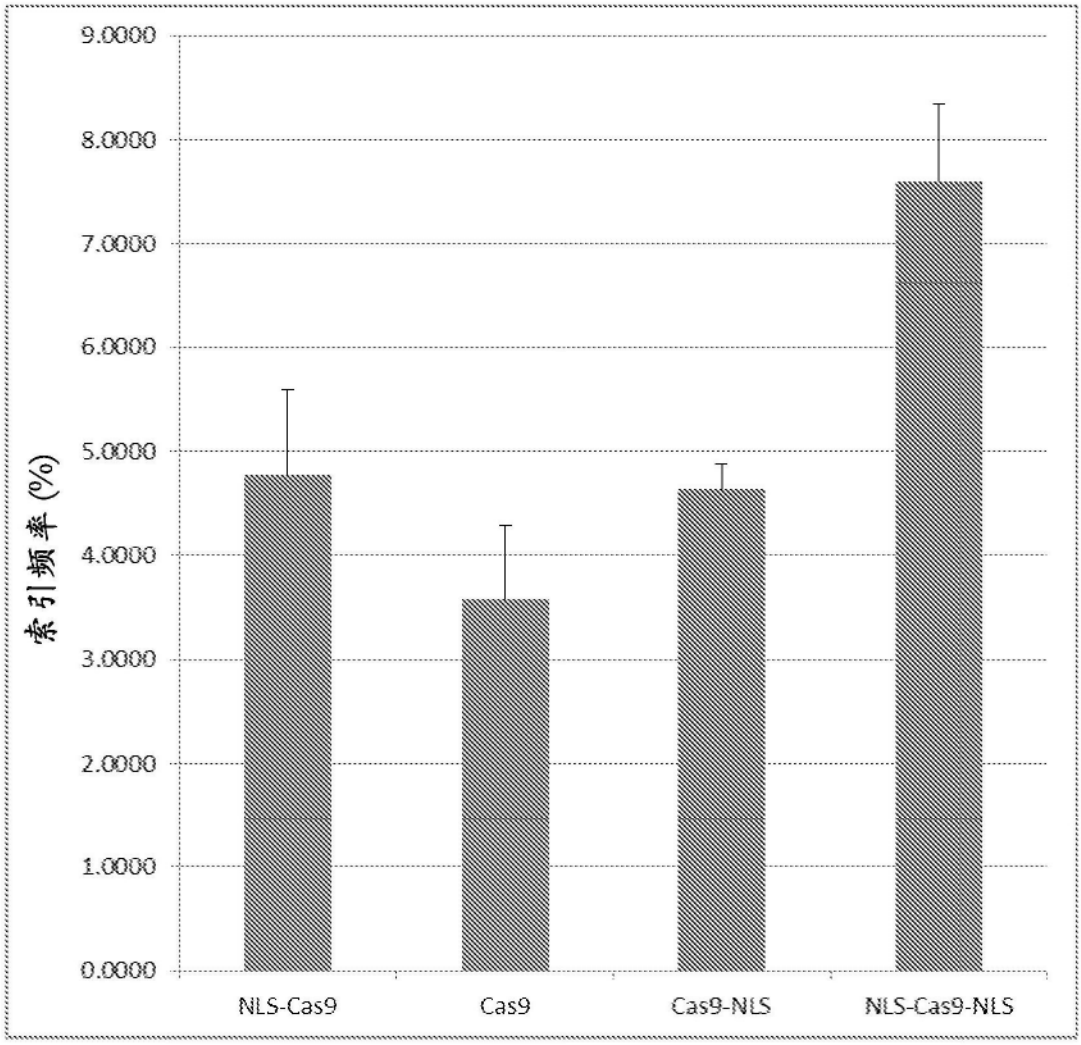

9.0000
8.0000
7.0000
6.0000
5.0000
4.0000
3.0000
2.0000
1.0000
0.0000
索引频率 (%)
NLS-Cas9
Cas9
Cas9-NLS
NLS-Cas9-NLS

图46

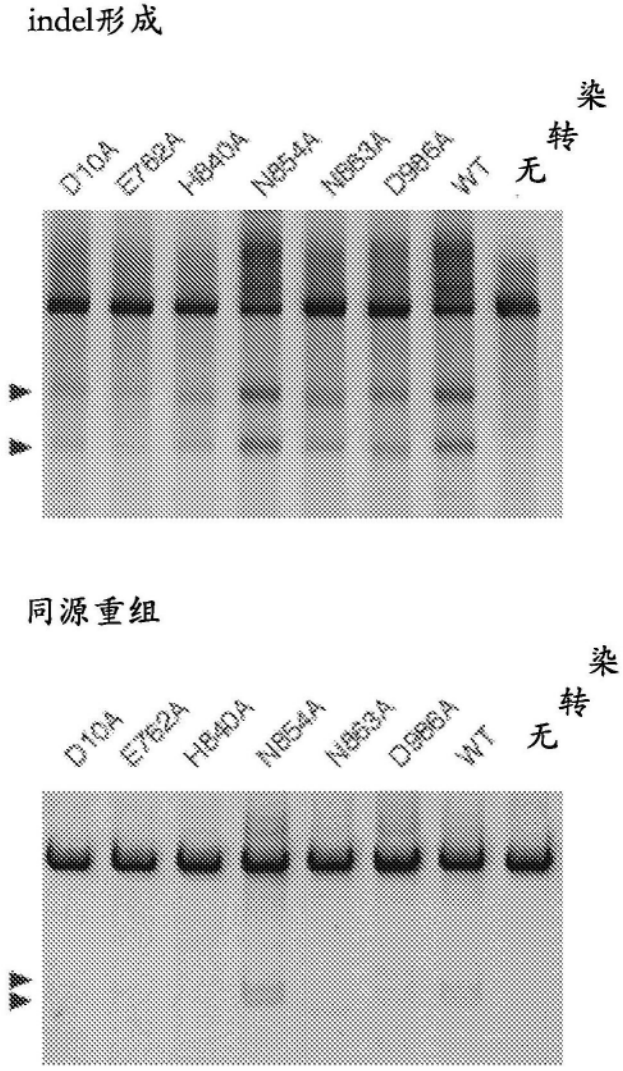

indel形成
染
转
无
D10A
E762A
H840A
N854A
N863A
D986A
WT
同源重组
染
转
无
D10A
E762A
H840A
N854A
N863A
D986A
WT

图47

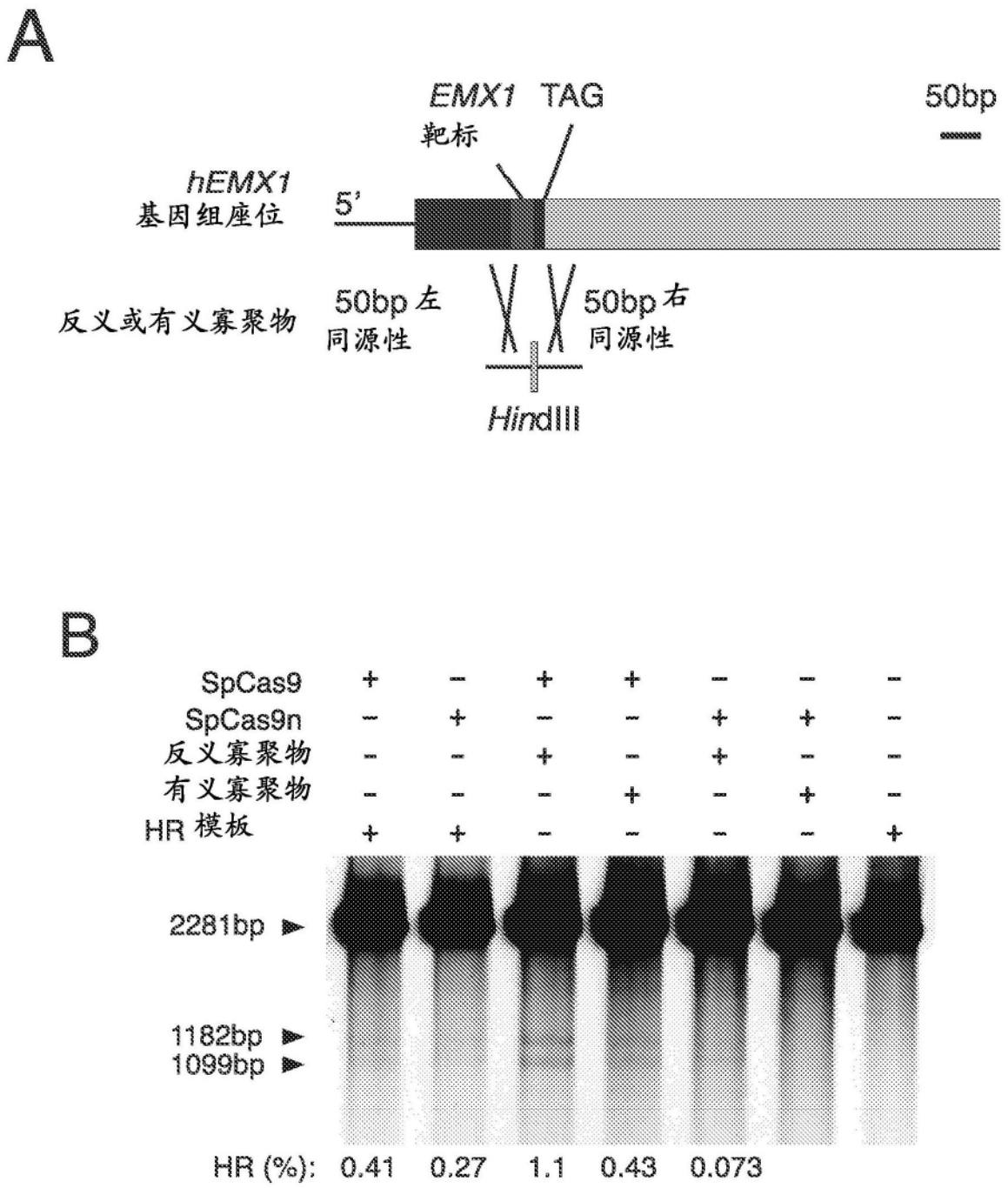

A
EMX1 TAG
靶标
50bp
hEMX1
基因组座位
5'
反义或有义寡聚物
50bp 左
同源性
50bp 右
同源性
HindIII
B
SpCas9
SpCas9n
反义寡聚物
有义寡聚物
HR 模板
2281bp
1182bp
1099bp
HR (%): 0.41 0.27 1.1 0.43 0.073

图48

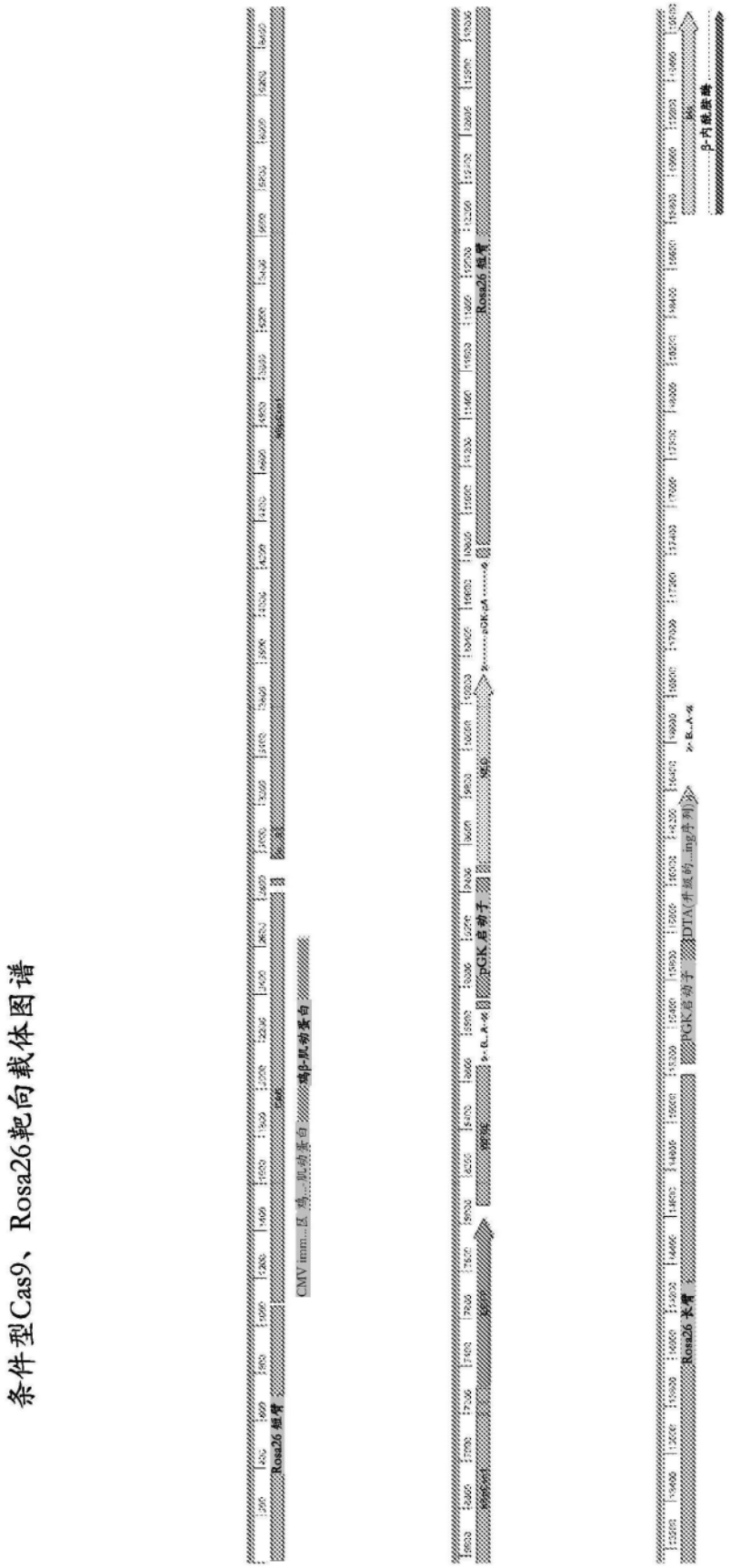

条件型Cas9、Rosa26靶向载体图谱
Rosa26 臂
CMV imm...区
鸡...肌动蛋白
绿荧光蛋白
pGK 启动子
Rosa26 臂
Rosa26 臂
PGK 启动子
DTA(白喉毒...ng序列)
β-内酰胺酶

图49A

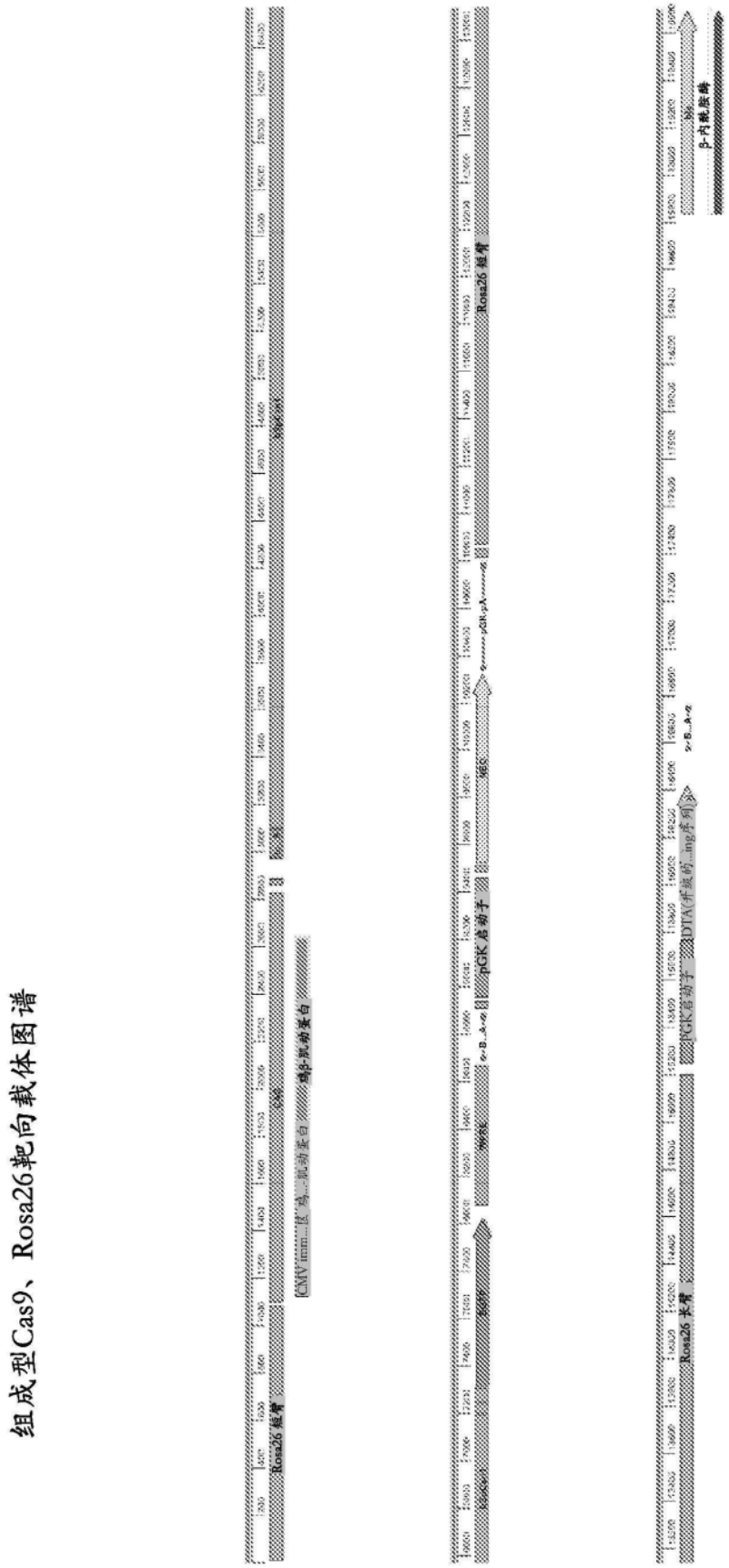

组成型Cas9、Rosa26靶向载体图谱
Rosa26 臂
CMV
Rosa26 臂
pGK 启动子
Rosa26 臂
Rosa26 臂
β-内酰胺酶

图49B

pCAG

ccgtttaaacaattctgcaggaatctagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataactt
acggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggga
ctttccattgacgtcaatgggtggagtatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgcccccctattgac
gtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgc
tattaccatggtcgaggtgagccccacgttctgcttcactctcccatctccccccctcccacccccaattttgtatttatttattttttaattatttt
gtgcagcgatgggggcgggggggggggggggcgcgcgccaggcggggcggggcggggcgaggggcggggcggggcgaggc
ggagaggtgcggcggcagccaatcagagcggcgcgctccgaaagtttccttttatggcgaggcggcggcggcggcggccctataaaaa
gcgaagcgcgcggcgggcggaAGTCGCTGCGcgctgccttcgccccgtgccccgctccgccgccgcctcgcgccgcccgcc
ccggctctgactgaccgcgttactcccacaggtgagcgggcgggacggcccttctcctccgggctgtaattagcgcttggtttaatgacggc
ttgtttcttttctgtggctgcgtgaaagccttgaggggctccgggagggccctttgtgcgggggggagcggctcgggggggtgcgtgcgtgtgt
gtgtgcgtggggagcgccgcgtgcggctccgcgctgcccggcggctgtgagcgctgcgggcgcggcgcggggctttgtgcgctccgc
agtgtgcgcgaggggagcgcggccgggggcggtgccccgcggtgcgggggggggctgcgaggggaacaaaggctgcgtgcggggt
gtgtgcgtgggggggtgagcagggggtgtgggcgcgtcggtcgggctgcaaccccccctgcaccccccctccccgagttgctgagcacg
gcccggcttcgggtgcggggctccgtacggggcgtggcgcggggctcgccgtgccgggcgggggggtggcggcaggtgggggtgcc
gggcggggcggggccgcctcgggccggggagggctcgggggaggggcgcggcggccccccggagcgccggcggctgtcgaggcg
cggcgagccgcagccattgccttttatggtaatcgtgcgagagggcgcagggacttcctttgtcccaaatctgtgcggagccgaaatctgg
gaggcgccgccgcacccccctctagcgggcgcggggcgaagcggtgcggcgccggcaggaaggaaatgggcggggagggccttcgt
gcgtcgccgcgccgccgtcccctctcctctccagcctcggggctgtccgcgggggggacggctgccttcgggggggacggggcaggg
cggggttcggcttctggcgtgtgaccggcggctctagagcctctgctaaccatgttcatgccttcttctttttcctacagctcctgggcaacgtg
ctggttattgtgctgtctcatcattttggcaaa

NLS-Cas9-NLS

ATGGACTATAAGGACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGA
CGATGACGATAAGATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCC
CAGCAGCCGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGC
TGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGG
CAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACA
GCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACAC
CAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCA
AGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGAT
AAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCA
CGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACA
AGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCC
ACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTC

图50A

ATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGCCAG
CGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGG
AAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGaAACCTG
ATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAG
GATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCT
GGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCG
ACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCC
CTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCT
GAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACC
AGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTC
TACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGT
GAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGC
ATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGA
TTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCG
CATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGA
CCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAA
GGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGC
CCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTAT
AACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCT
GAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAA
GTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTC
CGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACG
ATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGAC
ATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGA
GGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGA
AGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGCATC
CGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGC
CAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACA
TCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAAT
CTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGA
CGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGG
CCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAA
GCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCC
GTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGG
GCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGAT

图50B

GTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTG
CTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGG
TCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACC
CAGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGG
ATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCAC
GTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCT
GATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGA
AGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGAC
GCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGA
AAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCA
AGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATC
ATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCC
TCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATT
TTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACC
GAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGA
TAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACA
GCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCC
AAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCA
GCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAA
AAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCG
GAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGC
CCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCT
CCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTG
GACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGC
TAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAG
AGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCG
CCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAG
GTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGAT
CGACCTGTCTCAGCTGGGAGGCGACAAAAGGCCGGCGGCCACGAAAAAGGCCGGC
CAGGCAAAAAAGAAAAAG

P2A-EGFP

ggaagcggagccactaacttctccctgttgaaacaagcaggggatgtcgaagagaatccccgggccaGTGAGCAAGGGCGA
GGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACG
GCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTG
ACCCTGAAGTTCATCT

图50C

GCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACG
GCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGT
CCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGC
AACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCAT
CGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGG
AGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGC
ATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGC
CGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACA
ACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGAT
CACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAG
CTGTACAAG

WPRE

Cgataatcaacctctggattacaaaatttgtgaaagattgactggtattcttaactatgttgctccttttacgctatgtggatacgctgctttaatgc
ctttgtatcatgctattgcttcccgtatggctttcattttctcctccttgtataaatcctggttgctgtctctttatgaggagttgtggcccgttgtcagg
caacgtggcgtggtgtgcactgtgtttgctgacgcaacccccactggttggggcattgccaccacctgtcagctcctttccgggactttcgctt
tccccctccctattgccacggcggaactcatcgccgcctgccttgcccgctgctggacagggctcggctgttgggcactgacaattccgtg
gtgttgtcggggaaatcatcgtcctttccttggctgctcgcctgtgttgccacctggattctgcgcgggacgtccttctgctacgtcccttcggc
cctcaatccagcggaccttccttcccgcggcctgctgccggctctgcggcctcttccgcgtcttcgccttcgccctcagacgagtcggatctc
cctttgggccgcctccccgcatcg

bGHpolyA

cgacCTCGACtgtgccttctagttgccagccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccactgt
ccttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctgggggggtggggtggggcaggacagcaagggggg
aggattgggaagacaatGgcaggcatg

loxP-SV40polyAx3-loxP

ataacttcgtataatgtatgctatacgaagttattcgcgatgaataaatgaaagcttgcagatctgcgactctagaggatctgcgactctagagg
atcataatcagccntaccacattttgtagaggttttactngctttaaaaaacctcccacacctcccccctgaacctgaaacataaaatgaatgcaa
ttgttgttgttaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagt
tgtggtttgtccaaactcatcaatgtatcttatcatgtctggatctgcgactctagaggatcataatcagccataccacatttgtagaggttttactt
gctttaaaaaacctcccacacctcccccctgaacctgaaacataaaatgaatgcaattgttgttgttaacttgtttattgcagcttataatggttaca
aataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtct
ggatctgcgactctagaggatcataatcagccataccacatttgtagaggttttacttgctttaaaaaacctcccacacctcccccctgaacctga
aacataaaatgaatgcaattgttgttgttaacttgtttattgcagcttataatggttacaaata

图50D

aagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctggat
ccccatcaagctgatccggaacccttaatataacttcgtataatgtatgctatacgaagttat

Rosa26 短同源臂

caggccctccgagcgtggtggagccgttctgtgagacagccgggtacgagtcgtgacgctggaaggggcaagcgggtggtgggcagg
aatgcggtccgccctgcagcaaccggaggggggagggagaagggagcggaaaagtctccaccggacgcggccatggctcgggggggg
gggggggcagcggaggaGcgcttccggccgacgtctcgtcgctgattggcttCttttcctcccgccgtgtgtgaaaacacaaatggcgtgtt
ttggttggcgtaaggcgcctgtcagttaacggcagccggagtgcgcagccgccggcagcctcgctctgcccactgggtgggggcgggag
gtaggtggggtgaggcgagctgGacgtgcgggcgcggtcggcctctggcggggcgggggagggggagggagggtcagcgaaagta
gctcgcgcgcgagcggccgcccaccctccccttcctctgggggagtcgttttacccgccgccggccgggcctcgtcgtctgattggctctc
ggggcccagaaaactggcccttgccattggctcgtgttcgtgcaagttgagtccatccgccggccagcgggggcggcgaggaggcgctc
ccaggttccggccctcccctcggccccgcgccgcagagtctggccgcgcgcccctgcgcaacgtggcaggaagcgcgcgctgggggc
ggggacgggcagtagggctgagcggctgcgggggcgggtgcaagcacgtttccgacttgagttgcctcaagaggggcgtgctgagccag
acctccatcgcgcactccggggagtggagggaaggagcgagggctcagttgggctgttttggaggcaggaagcacttgctctcccaaagt
cgctctgagttgttatcagtaagggagctgcagtggagtaggcggggagaaggccgcacccttctccggaggggggagggggagtgttgc
aatacctttctgggagttctctgctgcctcctggcttctgaggaccgccctgggcctgggagaatccctttcccccctcttccctcgtgatctgcaa
ctccagtctttctag

Rosa26 长同源臂

agatgggcgggagtcttctgggcaggcttaaaggctaacctggtgtgtgggcgttgtcctgcaggggaattgaacaggtgtaaaattggag
ggacaagacttcccacagattttcggttttgtcgggaagtttttaatagggggcaaataaggaaaatgggaggataggtagtcatctgggggtttt
atgcagcaaaactacaggttattattgcttgtgatccgcctcggagtattttccatcgaggtagattaaagacatgctcacccgagttttatactct
cctgcttGAGATCCTTACTACAGTATGAAATTACAGTGTCGCGAGTTAGACTATGTAAGC
AGAATTTTAATCATttttaaagagcccagtacttcatatccatttctcccgctccttctgcagccttatcaaaaggtaTtttagaaaca
ctcattttagccccattttcatttattatactggcttatccaaccccctagacagagcattggcatttttccctttcctgatcttagaagtctgatgactca
tgaaaccagacagattagttacatacaccacaaatcgaggctgtagctggggcctcaacactgcagttcttttataactcctagtacacttttg
ttgatcctttgccttgatccttaattttcagtgtctatcacctctcccgtcaggtggtgttccacatttgggcctattctcagtccagggagttttaca
acaatagatgtattgagaatccaacctaaagcttaactttccactcccatgaatgcctctctcctttttctccattTATAAACTGAGCT
ATTAACCATTAATGGTTTCCAGGTGGATGTCTCCTCCCCCAATATTACCTGATGTATC
TTACATATTGCCAGGCTGATATTTTAAGACATTAAAAGGTATATTTCATTATTGAGCC
ACATGGTATTGATTACTGCTtactaaaattttgtcattgtacacatctgtaaaaggtggttccttttggaatgcaaagttcaggt
gtttgttgtctttcctgacctaaggtcttgtgagcttgtatttttttctatttaagcagtgctttctcttggactggcttgactcatggcattctacacgtta
ttgctggtctaaatgtgattttgccaagcttcttcaggacctataattttgcttgacttgtagccaaacacaagtaaaatgattaagcaacaaatgt
atttgtgaagcttggtttttaggttgttgtgttgtgtgtgcttgtgctctataataatactatccaggggctggagaggtggctcggagttcaagag
cacagactgctcttccagaagtc

图50E

ctgagttcaattcccagcaaccacatggtggctcacaaccatctgtaatgggatctgatgccctcttctggtgtgtctgaagaccacaagtgta
ttcacattaaataaataaaTCCTCCTTCTTCTTCTTTTTTTTTTTTTtAAAGAGAATACTGTCTCCAG
TAGAAtTTACTGAAGTAATGAAATACTTTGTGTTTGTTCCAATATGGTAGCCAATAAT
CAAATtACTCTTTaAGCACTGGAAATGTtACCAAGGAACTAaTTTTtATTTgAAGTGTaA
CTGTGGACAGAGGAGCCATAACTGCAGACTTGTGGGATACAGAAGACCAATGCAGA
CtTTAATGTCTTTTCTCTTACACTAAGCAATAAAGAAATAAAAATTGAACTTCTAGTA
TCCTATTTGTTtAAACTGCTAGCTTTACtTAACTTTTGTGCTTCATCTATACAAAGCTG
AAAGCTAAGTCTGCAGCCATTACTAAACATGAAAGCAAGTAATGATAATTTTGGATT
TCAAAAATGTAGGGCCAGAGTTTAGCCAGCCAGTGGTGGTGCTTGCCTTTATGCCtTT
AATCCCAGCACTCTGGAGGCAGAGACAGGCAGATCTCTGAGTTTGAGCCCAGCCTG
GtCTACACATCAAGTTCTATCTAGGATAGCCAGGAATACACACAGAAACCCTGTTGG
GGAGGGGGGCTCTGAGATTTCATAAAATTATAATTGAAGCATTCCCTAATGAGCCAC
TATGGATGTGGCTAAATCCGTCTACCTTTCTGATGAGATTTGGGTATTATTTTTTCTG
TCTCTGCTGTTGGTTGGGTCTTTTGACACTGTGGGCTTTCTTtAAAGCCTCCTTCCTGC
CATGTGGTCTCTTGTTTGCTACTAACTTCCCATGGCTTAAATGGCATGGCTTTTTGCC
TTCTAAGGGCAGCTGCTGAGATTTGCAGCCTGATTTCCAGGGTGGGGTTGGGAAATC
TTTCAAACACTAAAATTGTCCTTTAAtTTTTTTTTTAAAAAATGGGTTATATAATAAA
CCTCATAAAATAGTTATGAGGAGTGAGGTGGACTAATATTAAaTGAGTCCCTCCCCT
ATAAAAGAGCTATTAAGGCTTTTTGTCTTATACTtAACTTTTTTTTTAAATGTGGTATC
TTTAGAACCAAGGGTCTTAGAGTTTTAGTATACAGAAACTGTTGCATCGCTTAATCA
GATTTTCTAGTTTCAAATCCAGAGAATCCAAATTCTTCACAGCCAAAGTCAAATTAA
GAATTTCTGACTTTtAATGTTAaTTTGCtTACTGTGAATATaAAAATGATAGCTTTTCCT
GAGGCAGGGTCTCACTATGTATCTCTGCCTGATCTGCAACAAGATATGTAGACTAAA
GTTCTGCCTGCTTTTGTCTCCTGAATACTAAGGTTAAAATGTAGTAATACTTTTGGAA
CTTGCAGGTCAGATTCTTTTATAGGGGACACACTAAGGGAGCTTGGGTGATAGTTGG
TAAAtgtgtttaagtgatgaaaacttgaattattatcaccgcaacctacttttaaaaaaaaaagccaggcctgttagagcatgctTaaggg
atccctaggacttgctgagcacacaAGAGTAGtTACTTGGCAGGCTCCTGGTGAGAGCATATTTCAA
AAAACAAGGCAGACAACCAAGAAACTACAGTtAAGGTTACCTGTCTTTaAACCATCT
GCATATACACAGGGATATTAAAATATTCCAAATAATATTTCATTCAAGTTTTCCCCC
ATCAAATTGGGACATGGATTTCTCCGGTGAATAGGCAGAGTTGGAAACTAAACAAA
TGTTGGTTTTGTGATTTGTGAAATTGTTTTCAAGTGATAGTTAAAGCCCATGAGATAC
AGAACAAAGCTGCTATTTCGAGGTCTCTTGGTTtATACTCAGAAGCACTTCTTTGGGT
TTCCCTGCACTATCCTGATCATGTGCTAGGCCTACCTTAGGCTGATTGTTGTTCAAAT
aAACTTAAGTTTCCTGTCAGGTGATGTCATATGATTTCATATATCAAGGCAAAACATG
TTATATATGTTAAACATTTGTACTTAATGTGAAAGTTAGGTCTTTGTGGGTT

图50F

TGATTTTtAAtTTTCAAAACCTGAGCTAAATAAGTCATTTTtACATGTCTTACATTTGGT
GgAATTGTATaATTGTGGTTTGCAGGCAAGACTCTCTGACCTAGTAACCCTaCCTATA
GAGCACTTTGCTGGGTCACAAGTCTAGGAGTCAAGCATTTCACCTTGAAGTTGAGAC
GTTTTGTTAGTGTATACTAGTTtATATGTTGGAGGACATGTTTATCCAGAAGATATTC
AGGACTATTTTTGACTGGGCTAAGGAATTGATTCTGATTAGCACTGTTAGTGAGCAT
TGAGTGGCCTTTAGGCTTGAATTggagtcacttgtatatctcaaataatgctggccttttttaaaagcccttgttctttatca
ccctgttttctacataatttttgttcaaagaaatacttgtttggaTCTCCTTTTGACAACAATAGCATGTTTTCAAG
CCATATTTTTTTTCCTTTTTTTTTTTTTTTTGGTTTTTCGAGACAGGGTTTCTCTGTAT
AGCCCTGGCTGTCCTGGAACTCACTTTGTAGACCAGGCTGGCCTCGAACTCAGAAAT
CCGCCTGCCTCTGCCTCCTGAGTGCCGGGATTAAAGGCGTGCACCACCACGCCTGGC
TAAGTTGGATATTTTGTtATATAACTATAACCAATACTAACTCCACTGGGTGGATTTT
TAATTCAGTCAGTAGTCTTAAGTGGTCTTTATTGGCCCTTcATTAAAATCTACTGTTC
ACTCTAACAGAGGCTGTTGGtACTAGTGGCACtTAAGCAACTTCCTACGGATATACTA
GCAGAtTAAGGGTCAGGGATAGAAACTAGTCTAGCGTTTTGTATACCTACCAGCTTtA
TACTACCTTGTTCTGATAGAAATATTTcAGGACATCTAGCTT

pPGK-Neo-pPGK-polyA

aattctaccgggtaggggaggcgcttttcccaaggcagtctggagcatgcgctttagcagccccgctgggcacttggcgctacacaagtgg
cctctggcctcgcacacattccacatccaccggtaggcgccaaccggctccgttctttggtggccccttcgcgccaccttctactcctcccct
agtcaggaagttcccccccgccccgcagctcgcgtcgtgcaggacgtgacaaatggaagtagcacgtctcactagtctcgtgcagatgga
cagcaccgctgagcaatggaagcgggtaggcctttggggcagcggccaatagcagctttgctccttcgctttctgggctcagaggctggg
aaggggtgggtccgggggcgggctcaggggcgggctcaggggcggggcgggcgcccgaaggtcctccggaggcccggcattctgc
acgcttcaaaagcgcacgtctgccgcgctgttctcctcttcctcatctccgggcctttcgacctgcaatcgccgctagcgaagttcctattctct
agaaagtataggaacttcgccaccatgggatcggccattgaacaagatggattgcacgcaggttctccggccgcttgggtggagaggctat
tcggctatgactgggcacaacagacaatcggctgctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttcttttttgtcaag
accgacctgtccggtgccctgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgt
gctcgacgttgtcactgaagcgggaagggactggctgctattgggcgaagtgccggggcaggatctcctgtcatctcaccttgctcctgcc
gagaaagtatccatcatggctgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgaccaccaagcgaaacatcgca
tcgagcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactg
ttcgccaggctcaaggcgcgcatgcccgacggcgatgatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtggaaaatg
gccgcttttctggattcatcgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtgatattgctgaagag
cttggcggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcgccttctatcgccttcttgacgagt
tcttctgaggggatccgctgtaagtctgcagaaattgatgatctattaaacaataaagatgtccactaaaatggaagttttcctgtcatactttgtt
aagaagggtgagaacagagtacctacattttgaatggaaggattggagctacgggggtgggggtggggtgggattagataaatgcctgct
ctttactgaaggctctttactattgctttatgataatgtttcatagttg

图50G

gatatcataatttaaacaagcaaaaccaaattaagggccagctcattcctcccactcatgatctatagatctatagatctctcgtgggatcattgt
ttttctcttgattcccactttgtggttctaagtactgtggtttccaaatgtgtcagtttcatagcctgaagaacgagatcagcagcctctgttccaca
tacacttcattctcagtattgttttgccaagttctaattccatcagaaagc

pPGK-DTA
TACCGGGTAGGGGAGGCGCTTTTCCcAAGGCAGTCTGgAGCATGCGCtTTAGCAGCCC
CGCTgGGCACTTGGCGCTACACAAGTGGCCTCTGGCCTCGCACACATTCCACATCCA
CCGGTAGGCGCCAACCGGCTCCGTTCTTTGGTGGCCCCTTCGCGCCACCTTCTACTCC
TCCCCTAGTCAGGAAGTTCCCCCCCGCCCCGCAGCTCGCGTCGTGcAGGACGTGACA
AATGGAAGTAGCACGTCTCACTAGTCTCGTgCAGATGGACAGCACCGCTGAGCAATG
GAAGCGGGTAGGCCTTTGGGGCAGCGGCCAATAGCAGCTTTGCTCCTTCGCTTTCTG
GGCTCAGAGGCTGGGAAGGGGTGGGTCCGGGGGCGGGCTCAGGGGCGGGCTCAGG
GGCGGGGCGGGCGCCCGAAGGTCCTCCGGAGGCCCGGCATTCTGCACGCTTCAAAA
GCGCACGTCTGCCGCGCTGTTCTCCTCTTCCTCATCTCCGGGCCTTTCGACCTGCAGG
TCCTCGCCATggatcctgatgatgttgttGattcttctaaAtcttttgtGatggaaaactttttcttcgtaccacgggactaaacctggtt
atgtagattccattcaaaaaggtatacaaaagccaaaatctggtacacaaggaaattatgacgatgattggaaagggttttatagtaccgacaa
taaatacgacgctgcgggatactctgtagataatgaaaacccgctctctggaaaagctggaggcgtggtcaaagtgacgtatccaggactg
acgaaggttctcgcactaaaagtggataatgccgaaactattaagaaagagttaggtttaagtctcactgaaccgttgatggagcaagtcgga
acggaagagtttatcaaaaggttcggtgatggtgcttcgcgtgtagtgctcagccttcccttcgctgaggggagttctagcgttgaatatattaa
taactgggaacaggcgaaagcgttaagcgtagaacttgagattaattttgaaacccgtggaaaacgtggccaagatgcgatgtatgagtata
tggctcaagcctgtgcaggaaatcgtgtcaggcgatctctttgtgaaggaaccttacttctgtggtgtgacataattggacaaactacctacag
agatttaaagctctaaggtaaatataaaatttttaagtgtataatgtgttaaactactgattctaattgtttgtgtattttagattccaacctatggaact
gatgaatgggagcagtggtggaatgcagatcctagagctcgctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgcccct
ccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcat
tctattctgggggggtggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatg

图50H

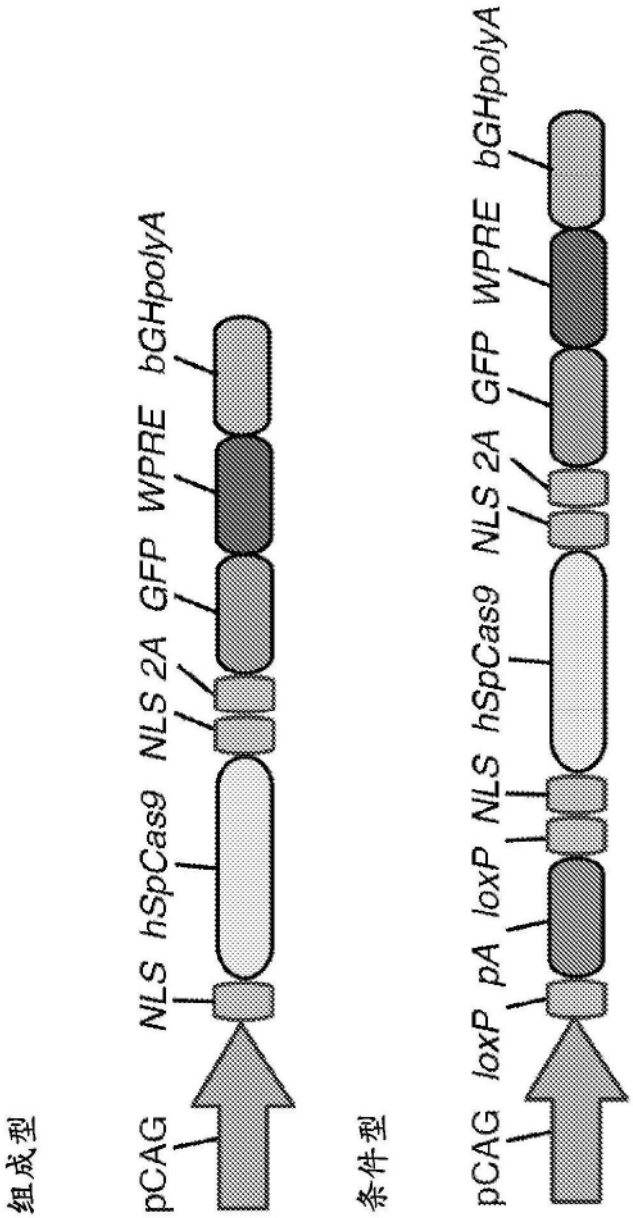

组成型
pCAG
NLS
hSpCas9
NLS
2A
GFP
WPRE
bGHpolyA
条件型
pCAG
loxP
pA
loxP
NLS
hSpCas9
NLS
2A
GFP
WPRE
bGHpolyA

图51

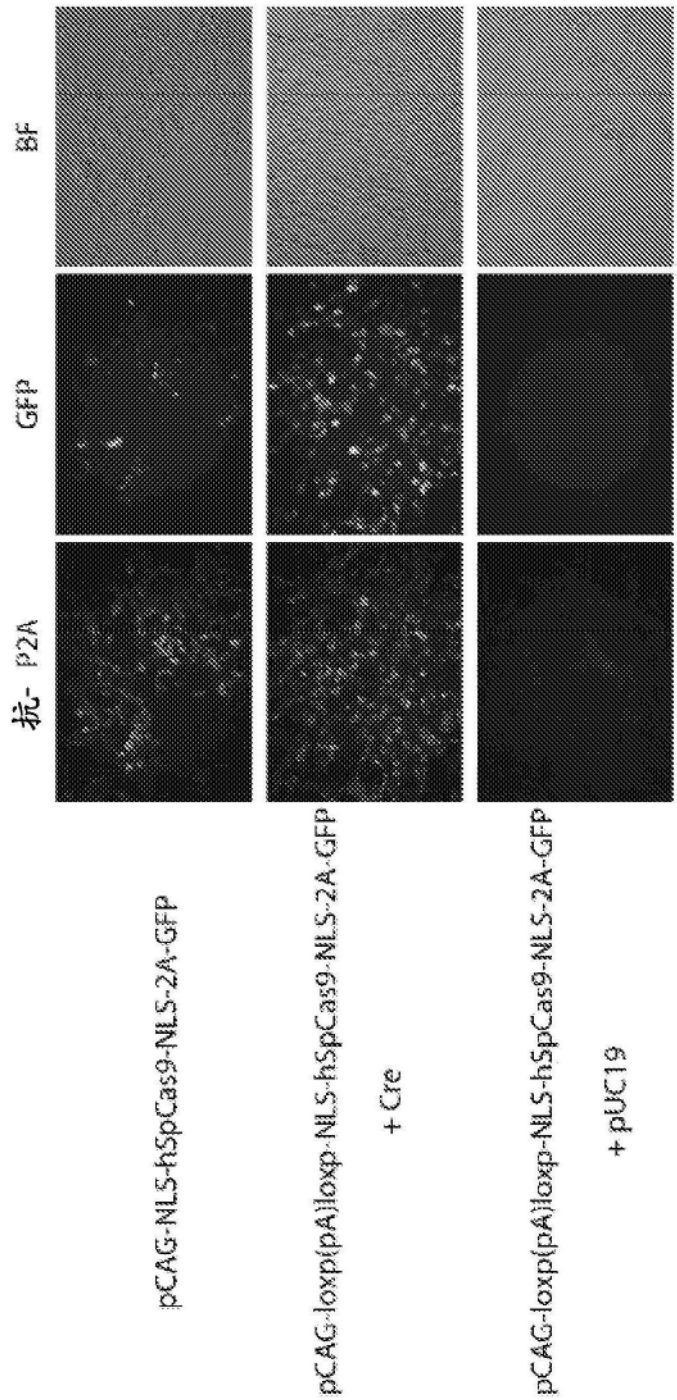

抗- P2A
GFP
BF
pCAG-NLS-hSpCas9-NLS-2A-GFP
pCAG-loxp(pA)loxp-NLS-hSpCas9-NLS-2A-GFP
+ Cre
pCAG-loxp(pA)loxp-NLS-hSpCas9-NLS-2A-GFP
+ pUC19

图52

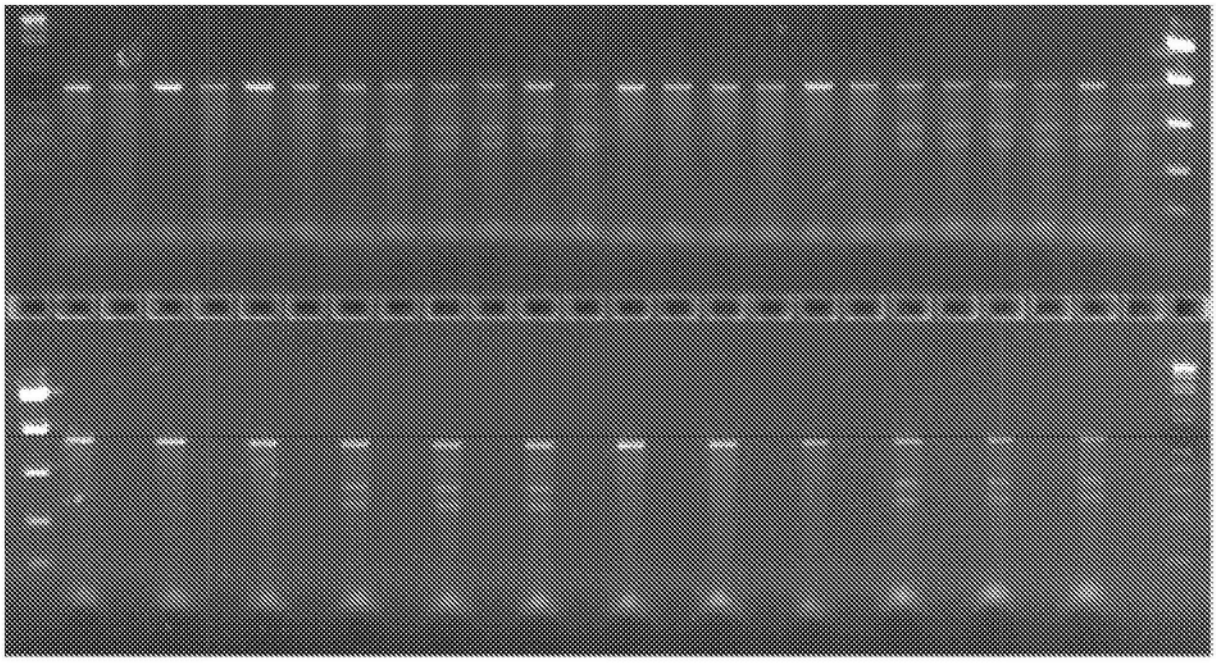


图53

			平均值	标准差
pCAG-loxp(pA)loxp-NLS-hSpCas9-NLS-2A-GFP	克隆	1	32.1	7.1
	克隆	2	27.3	3.5
	克隆	3	35.9	1.4
	克隆	4	39.0	4.7
pCAG-NLS-hSpCas9-NLS-2A-GFP	克隆	1	26.9	1.3
	克隆	2	33.1	2.7

图54

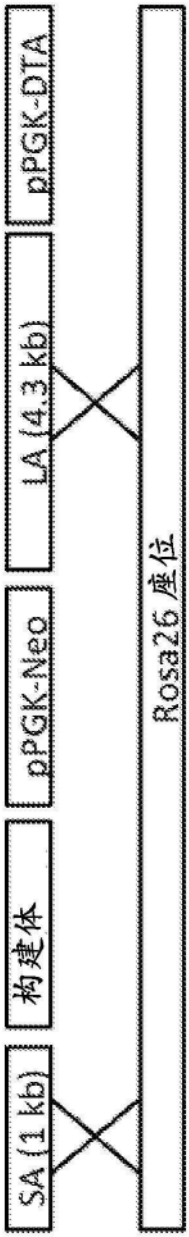


图55

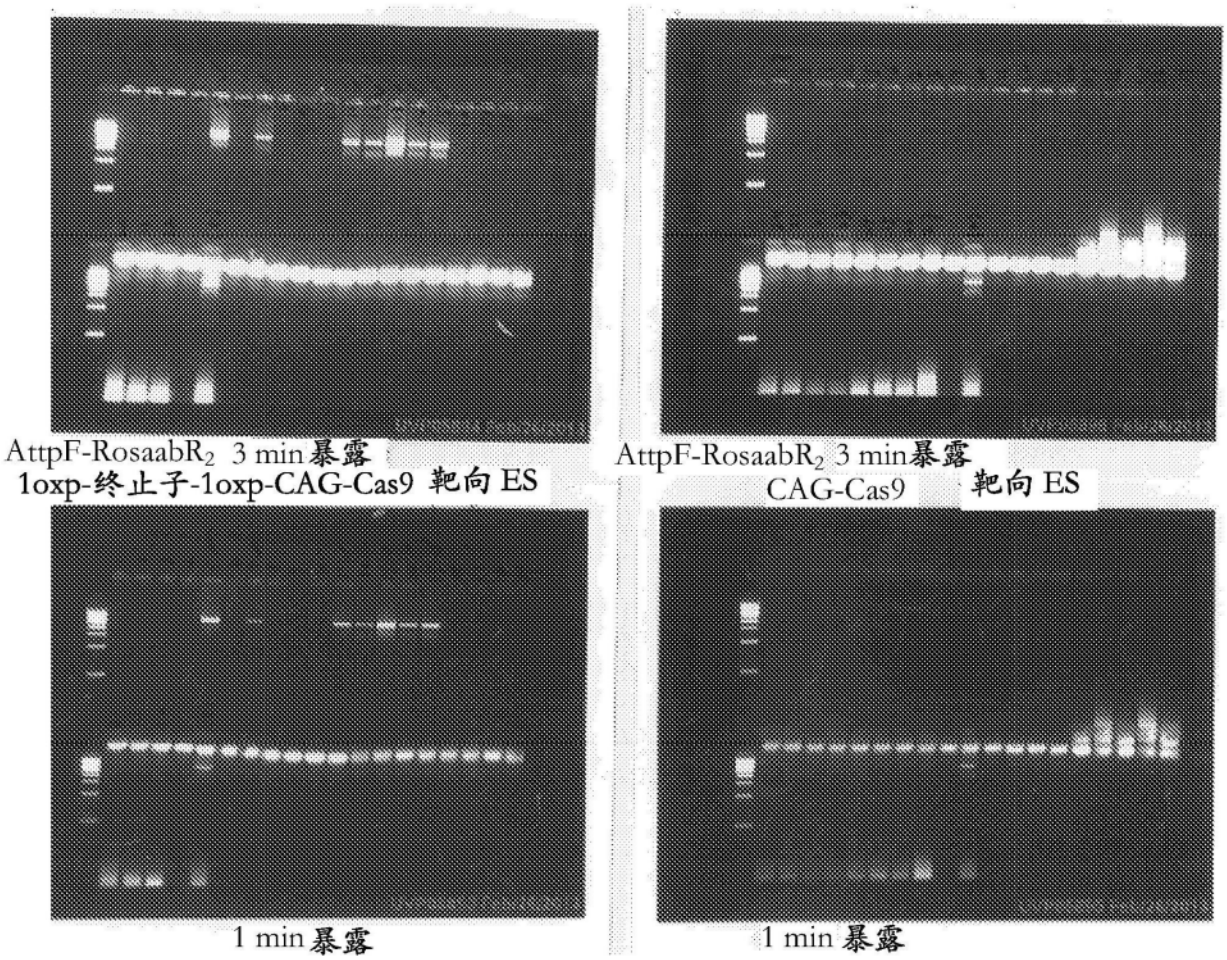

AttpF-RosaabR2 3 min暴露
1oxp-终止子-1oxp-CAG-Cas9 靶向 ES
AttpF-RosaabR2 3 min暴露
CAG-Cas9 靶向 ES
1 min暴露
1 min暴露

图56

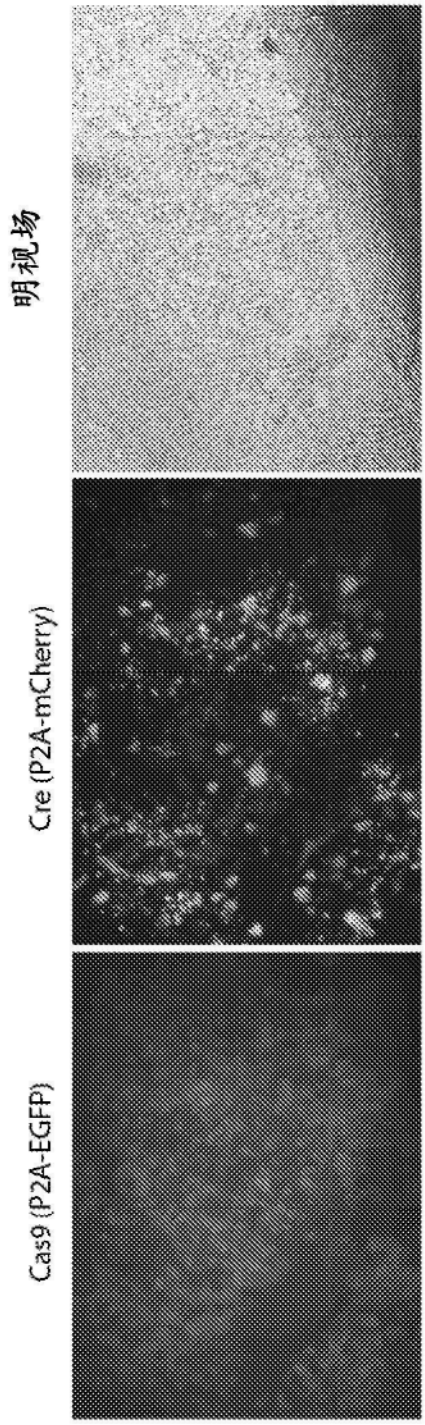


图57

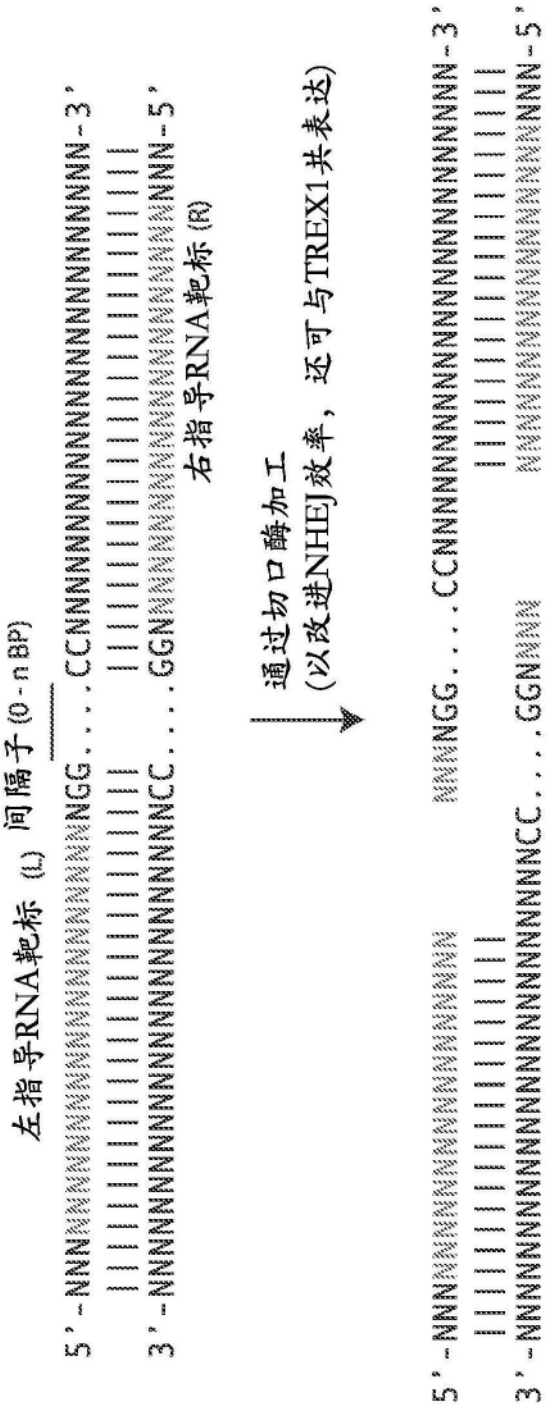

左指导RNA靶标 (L) 间隔子 (0-n BP)
5'-NNNNNNNNNNNNNNNNNNNNNGG.....CCNNNNNNNNNNNNNNNNNNNNNNN-3'
3'-NNNNNNNNNNNNNNNNNNNNNCC.....GGNNNNNNNNNNNNNNNNNNNNNNN-5'
右指导RNA靶标 (R)
通过切口酶加工
(以改进NHEJ效率，还可与TREX1共表达)
5'-NNNNNNNNNNNNNNNNNNN NNNNGG.....CCNNNNNNNNNNNNNNNNNNNNNNN-3'
3'-NNNNNNNNNNNNNNNNNNNNNCC.....GGNNNNNN NNNNNNNNNNNNNNNNNNNNNNN-5'

图58

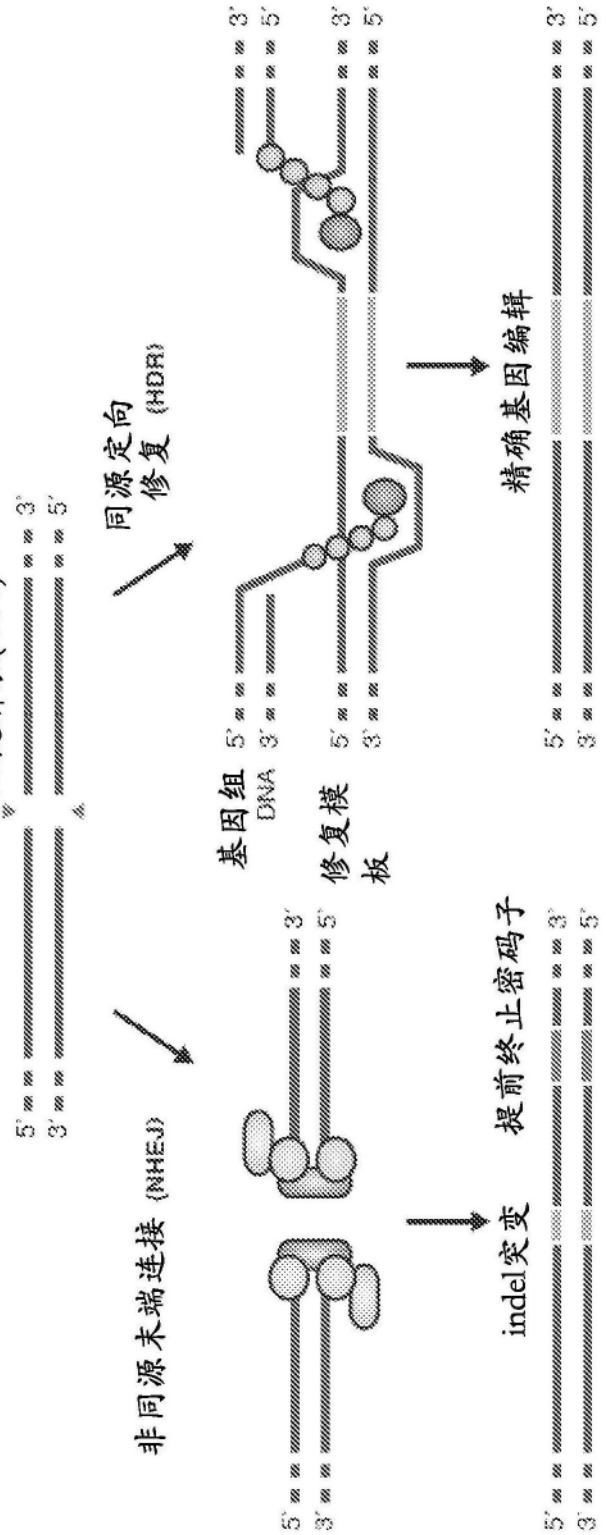

双链断裂(DSB)
同源定向
修复 (HDR)
非同源末端连接 (NHEJ)
基因组
DNA
修复模板
indel突变
提前终止密码子
精确基因编辑

图59

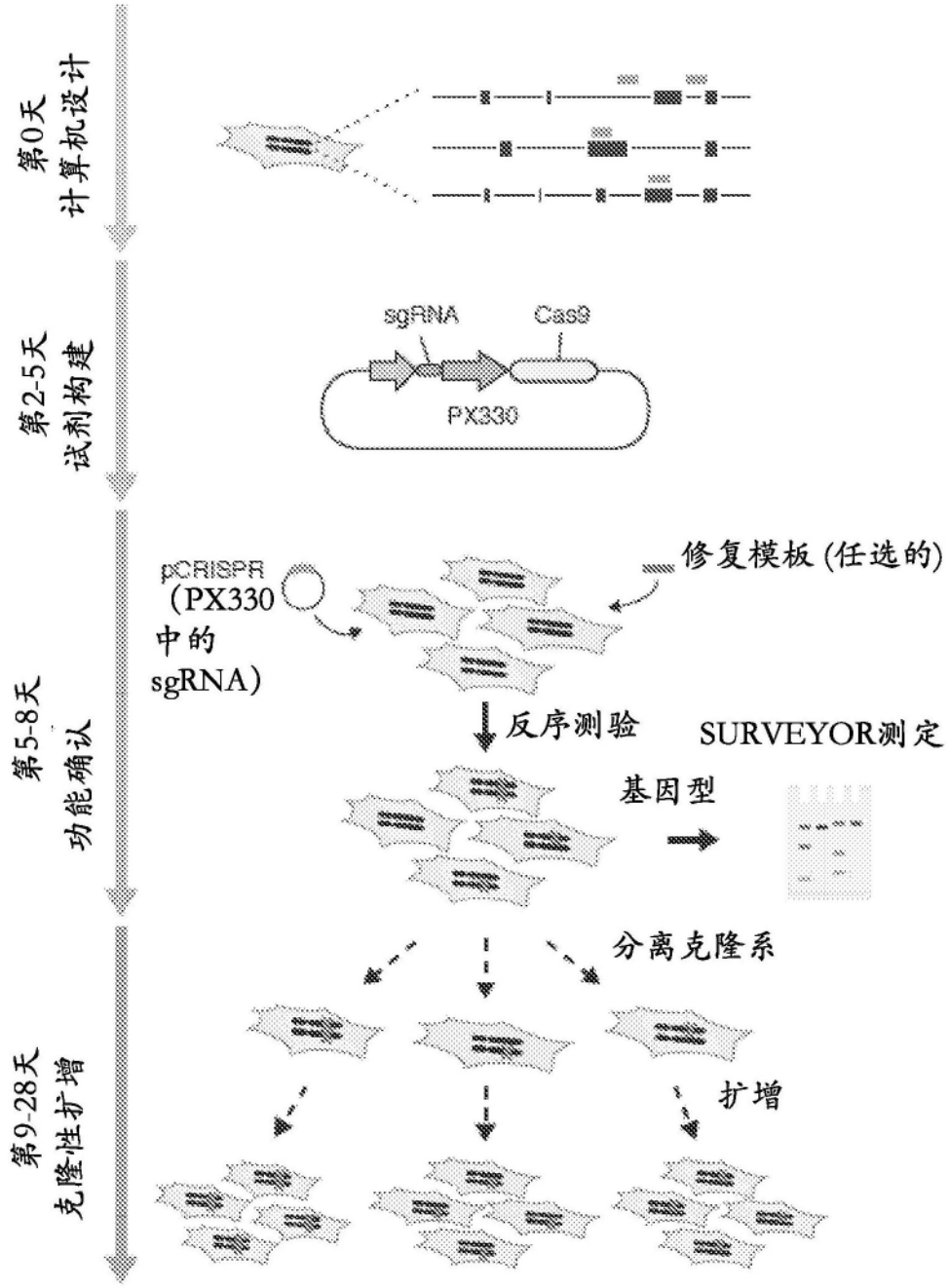

第0天
计算机设计
第2-5天
试剂构建
sgRNA
Cas9
PX330
第5-8天
功能确认
CRISPR
(PX330
中的
sgRNA)
修复模板(任选的)
反序测验
SURVEYOR测定
基因型
分离克隆系
第9-28天
克隆性扩增
扩增

图60

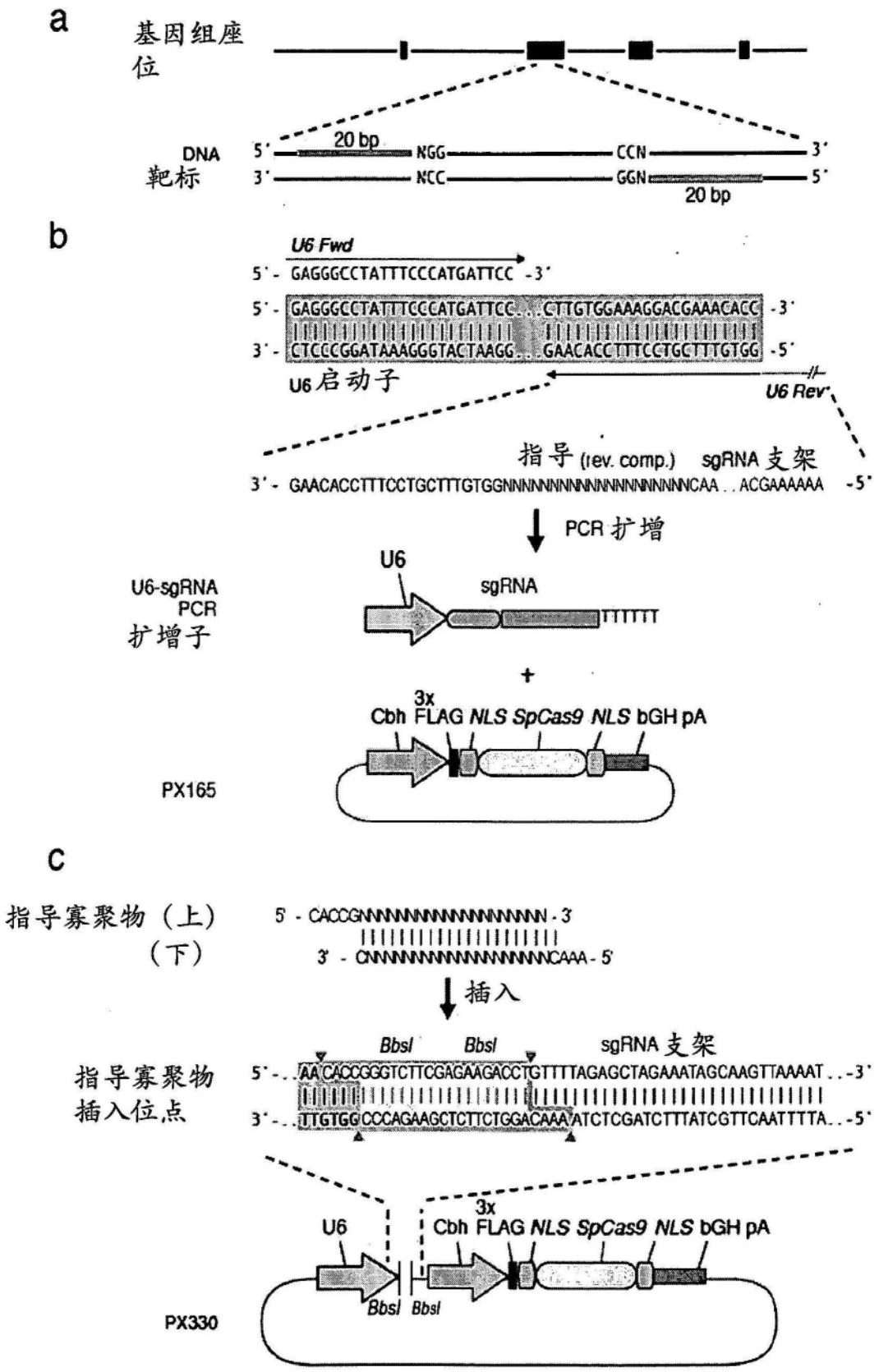
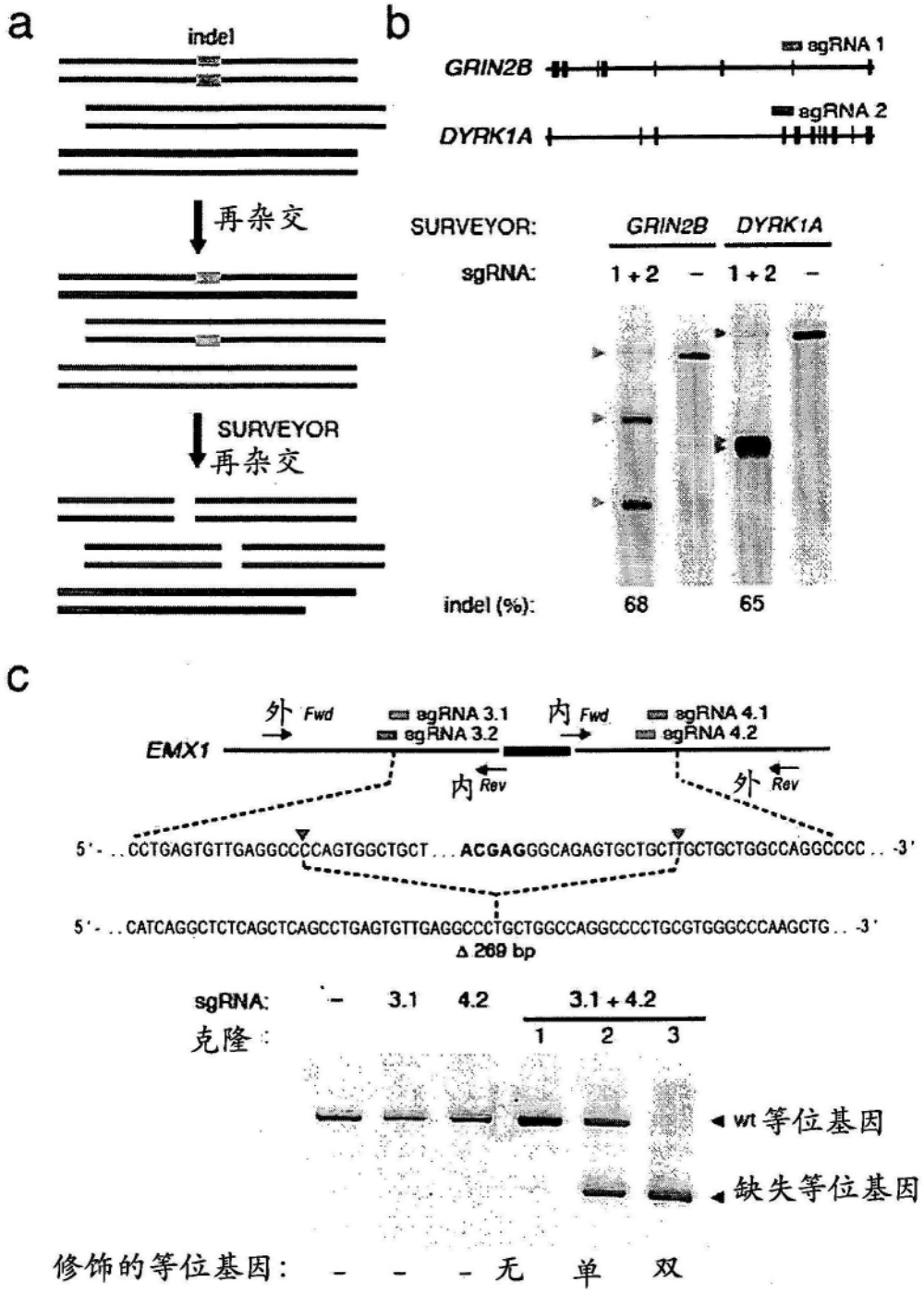

a
基因组座位
DNA
靶标
5'
3'
20 bp
NGG
CCN
NCC
GGN
20 bp
b
U6 Fwd
5'- GAGGGCCTATTTCCCATGATTCC -3'
5'- GAGGGCCTATTTCCCATGATTCC...CTTGTGGAAAGGACGAAACACC -3'
3'- CTCCCGGATAAAGGGTACTAAGG...GAACACCTTTCCTGCTTTGTGG -5'
U6启动子
U6 Rev
指导(rev. comp.) sgRNA支架
3'- GAACACCTTTCCTGCTTTGTGGNNNNNNNNNNNNNNNNNNNCAA..ACGAAAAAA -5'
PCR扩增
U6-sgRNA PCR扩增子
U6
sgRNA
TTTTTT
+
3x
Cbh FLAG NLS SpCas9 NLS bGH pA
PX165
c
指导寡聚物（上）
（下）
5'- CACCGNNNNNNNNNNNNNNNNNN-3'
3'- CNNNNNNNNNNNNNNNNNNCAAA-5'
插入
指导寡聚物插入位点
BbsI
BbsI
sgRNA支架
5'...AACACCGGGTCTTCGAGAAGACCTGTTTTAGAGCTAGAAATAGCAAGTTAAAAT...-3'
3'...TTGTGGCCCAGAAGCTCTTCTGGACAAAATCTCGATCTTTATCGTTCAATTTTA...-5'
3x
Cbh FLAG NLS SpCas9 NLS bGH pA
U6
BbsI
BbsI
PX330

图61



d

sgRNAs	大约缺失大小 (bp)	克隆的数目 +/+	Δ/+	Δ/Δ	倒位	总数
3.1 + 4.1	282	12	10	1	0	23
3.2 + 4.1	237	20	16	2	0	38
3.1 + 4.2	425	4	14	5	0	23
3.2 + 4.2	259	6	11	4	0	21

图62

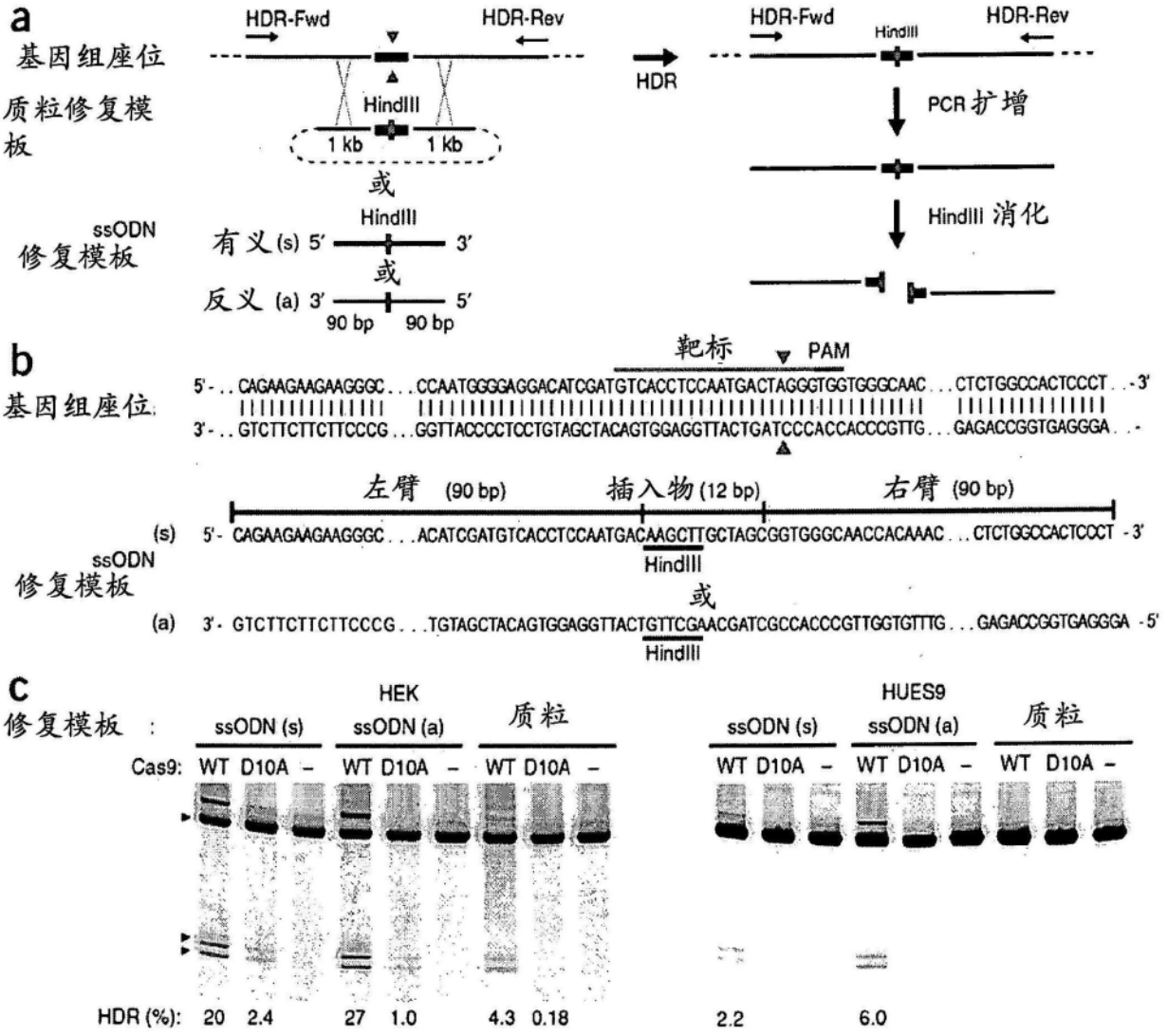


图63A图63B图63C

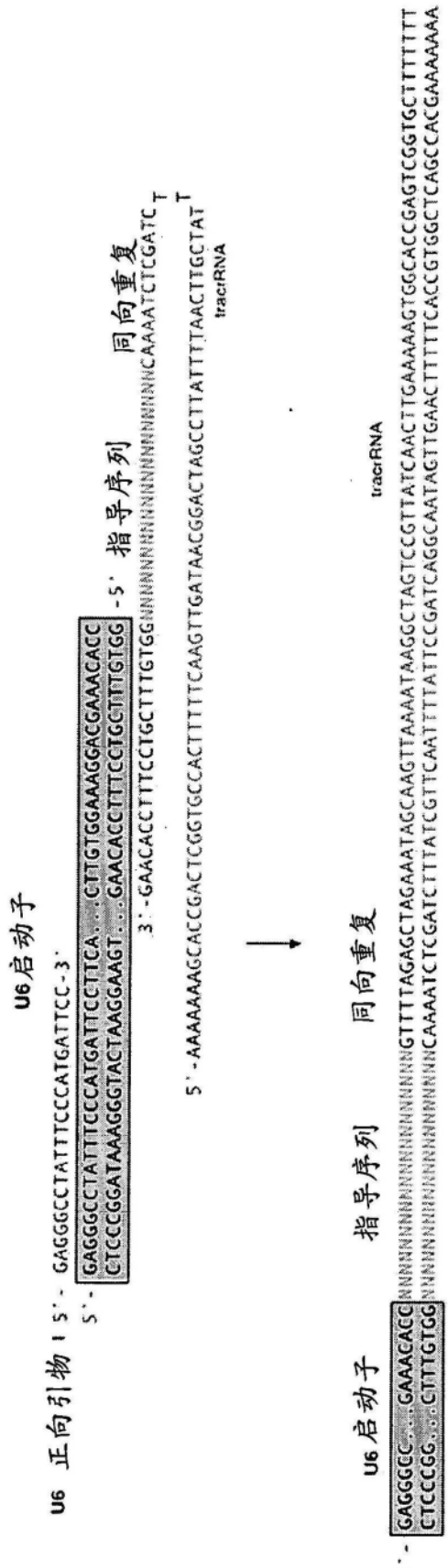

U6启动子
U6 正向引物
5'- GAGGGCCTATTTCCCATGATTCC-3'
5'- GAGGGCCTATTTCCCATGATTCCTTCA...CTTGTGGAAAGGACGAAACACC
CTCCCGGATAAAGGGTACTAAGGAAGT...GAACACCTTTCCTGCTTTGTGG -5' 指导序列 同向重复
3'-GAACACCTTTCCTGCTTTGTGGNNNNNNNNNNNNNNNNNNNNCAAAATCTCGATC T
5'-AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTAT T
tracrRNA
U6启动子 指导序列 同向重复
tracrRNA
5'- GAGGGCC....GAAACACC NNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT
CTCCCGG...CTTTGTGG NNNNNNNNNNNNNNNNNNNNCAAAATCTCGATCTTTATCGTTCAATTTTTATTCCGATCAGGCAATAGTTGAACTTTTTCACCGTGGCTCAGCCACGAAAAAAA
PCR或凝胶纯化
U6
嵌合RNA
TTTTTTT
+
CBh
NLS SpCas9 bGH pA

图64

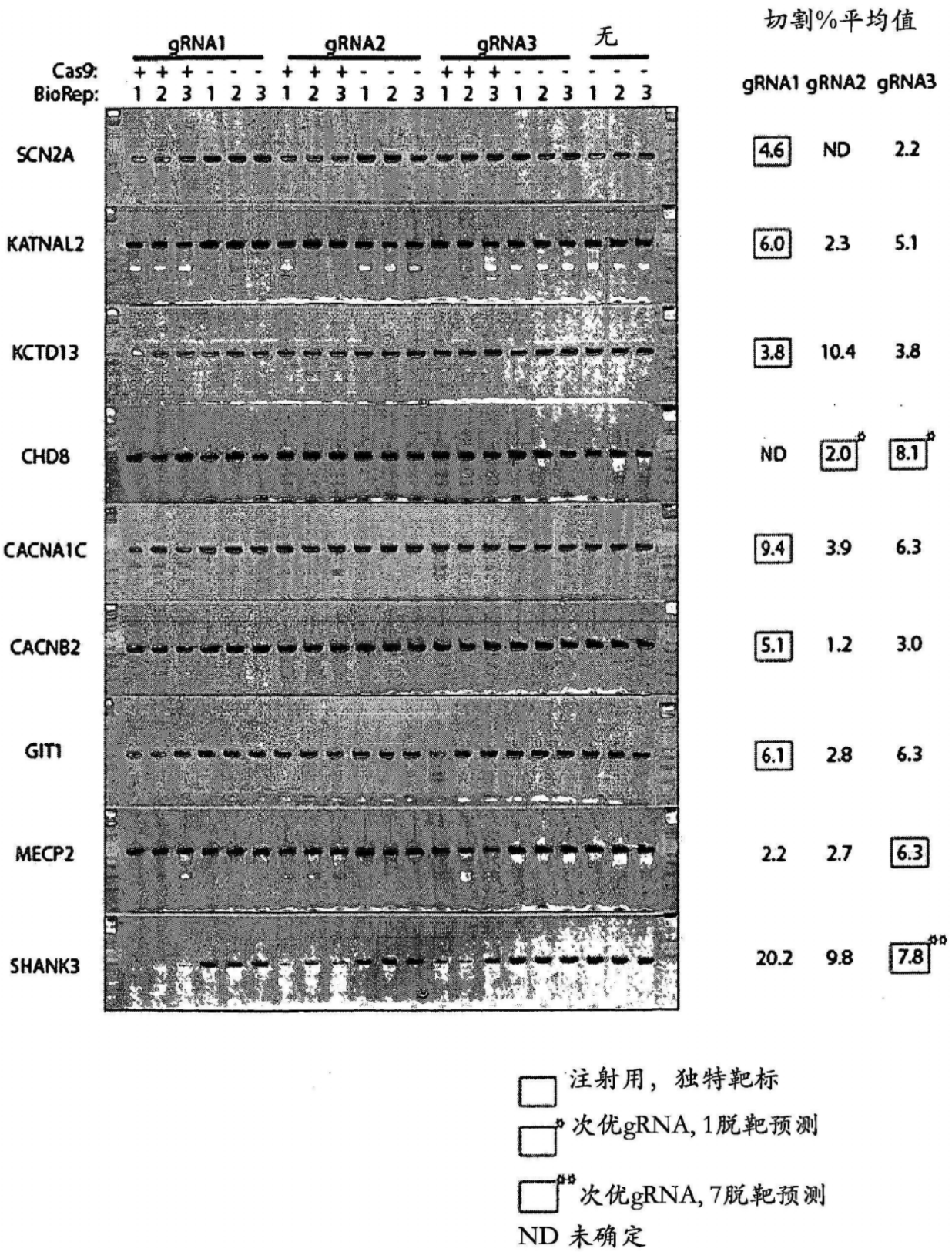
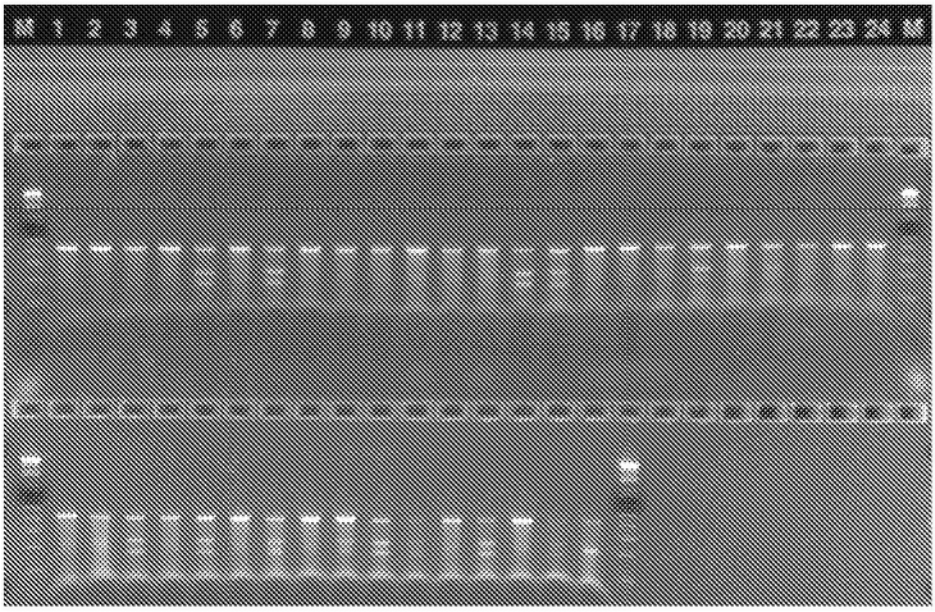

gRNA1
gRNA2
gRNA3
无
Cas9:
BioRep:
SCN2A
KATNAL2
KCTD13
CHD8
CACNA1C
CACNB2
GIT1
MECP2
SHANK3
切割%平均值
gRNA1 gRNA2 gRNA3
4.6 ND 2.2
6.0 2.3 5.1
3.8 10.4 3.8
ND 2.0* 8.1*
9.4 3.9 6.3
5.1 1.2 3.0
6.1 2.8 6.3
2.2 2.7 6.3
20.2 9.8 7.8**
注射用，独特靶标
次优gRNA, 1脱靶预测
次优gRNA, 7脱靶预测
ND 未确定

图65



用于靶向Chd8的gRNA序列

Chd8.1 - agctgttttactggtcggct
Chd8.2 - aatggatacacctggtcgaa
Chd8.3 - caatggatacacctggtcga

图66

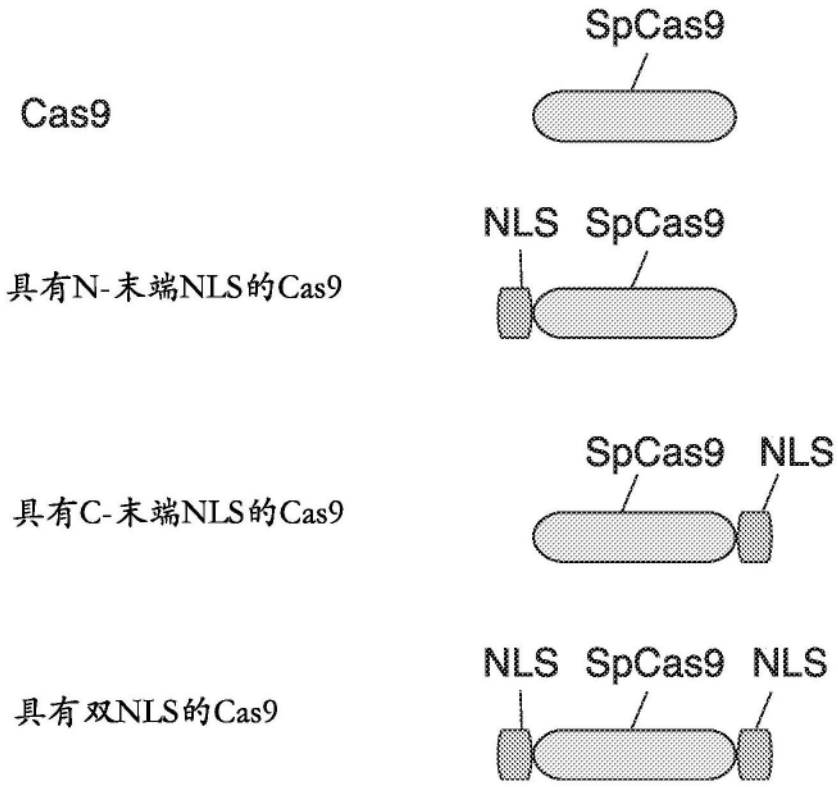

SpCas9
Cas9
NLS SpCas9
具有N-末端NLS的Cas9
SpCas9 NLS
具有C-末端NLS的Cas9
NLS SpCas9 NLS
具有双NLS的Cas9

图. NLS1

图67

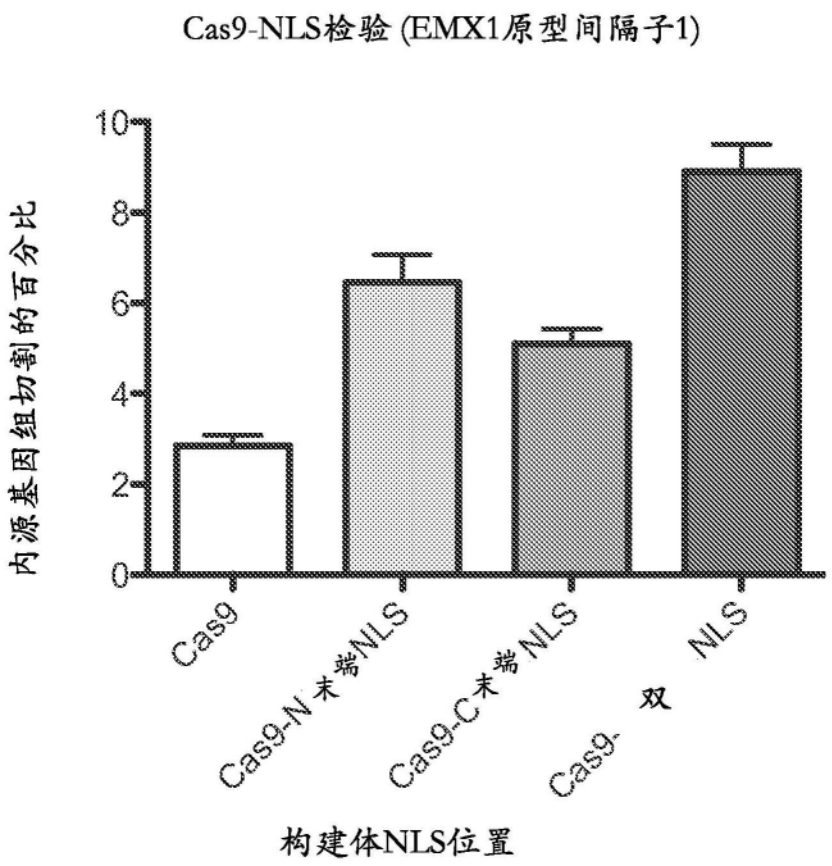
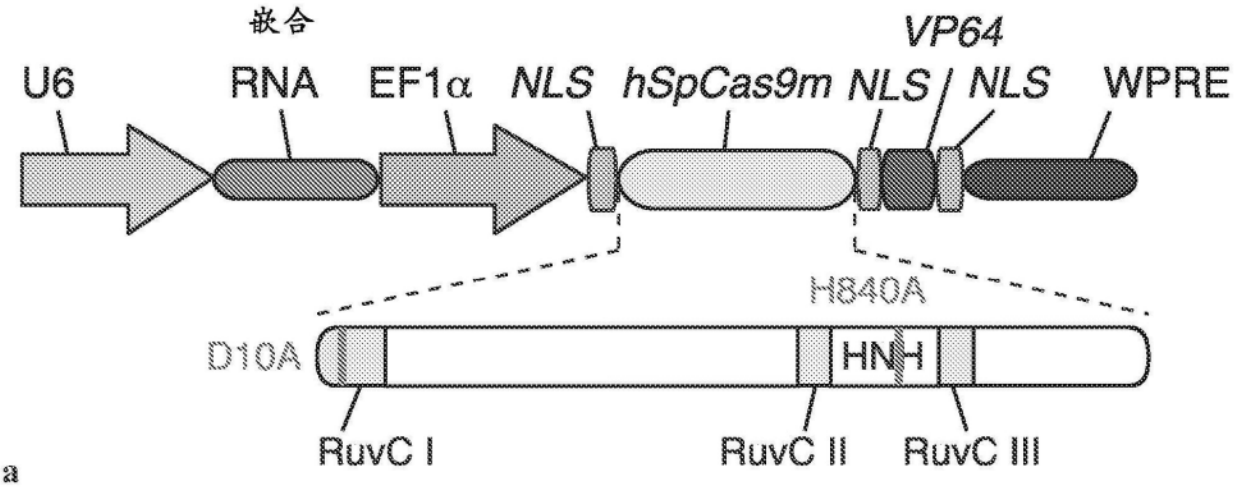
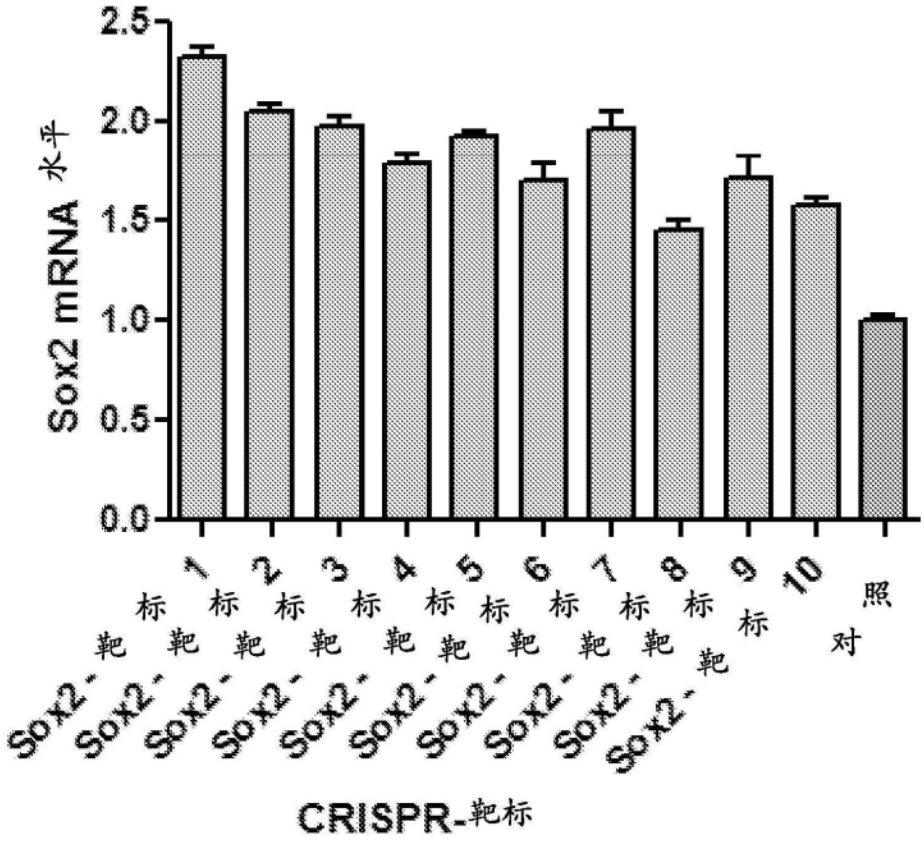

Cas9-NLS检验 (EMX1原型间隔子1)
内源基因组切割的百分比
10
8
6
4
2
0
Cas9
Cas9-N末端NLS
Cas9-C末端NLS
Cas9-双NLS
构建体NLS位置

图68


嵌合
VP64
U6
RNA
EF1α
NLS
hSpCas9m
NLS
NLS
WPRE
H840A
D10A
HNH
RuvC I
RuvC II
RuvC III

a


Sox2 mRNA 水平
2.5
2.0
1.5
1.0
0.5
0.0
1
2
3
4
5
6
7
8
9
10
照
标
靶
对
Sox2-
CRISPR-靶标

b

图69

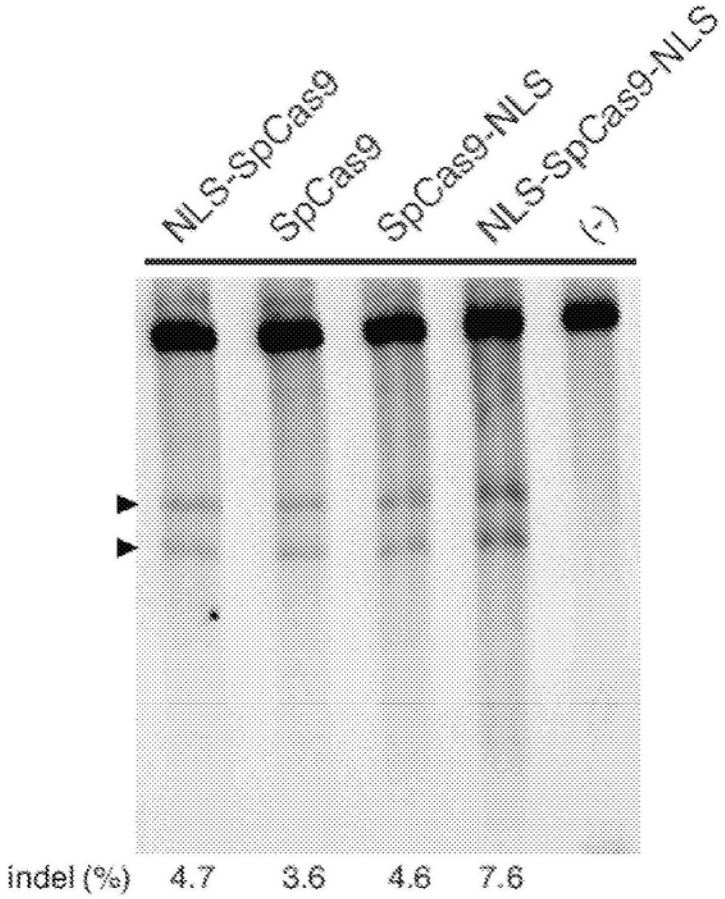

SpCas9的NLS构造优化
NLS-SpCas9
SpCas9
SpCas9-NLS
NLS-SpCas9-NLS
(-)
indel (%) 4.7 3.6 4.6 7.6

图70

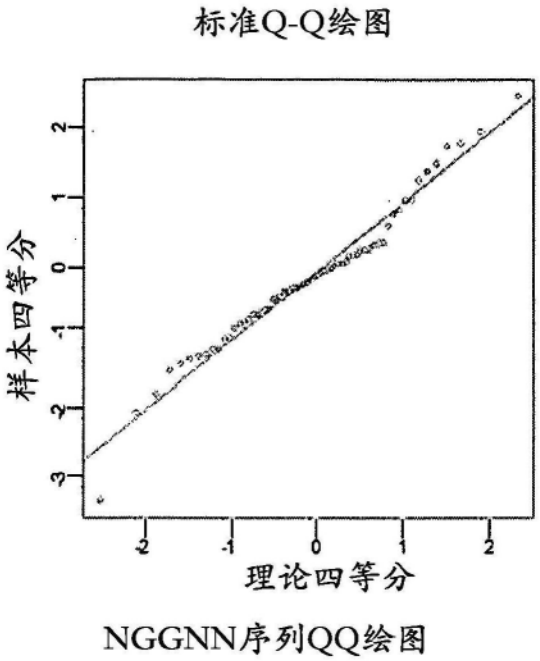


图71

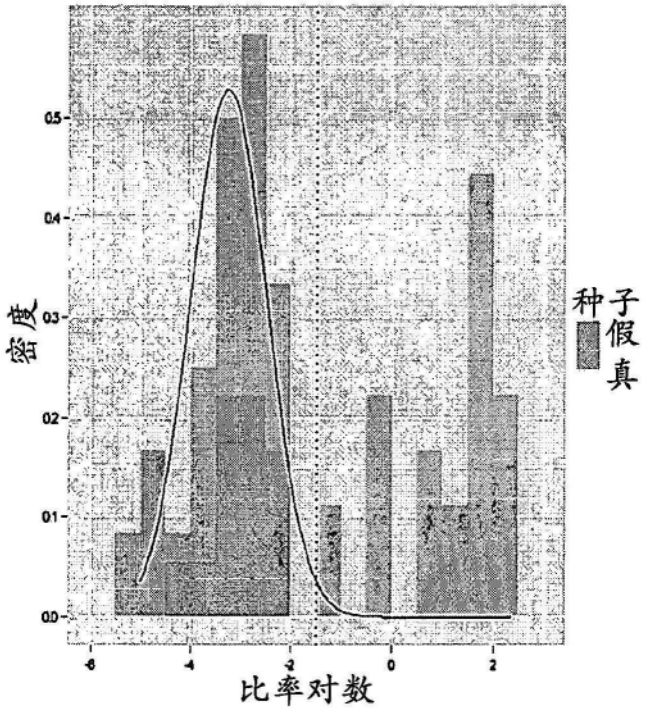


图72

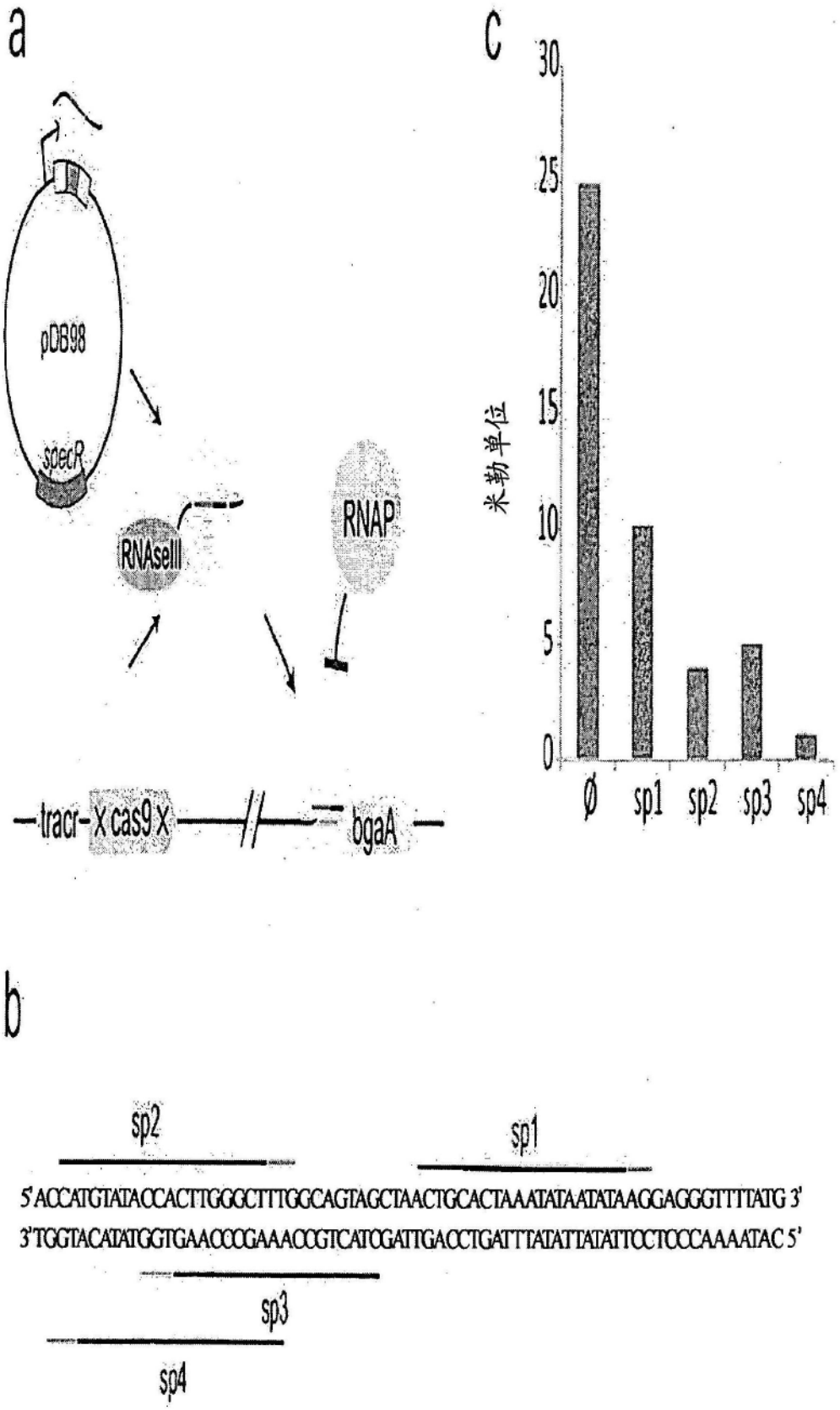

a
pDB98
specR
RNAseIII
RNAP
tracr
X cas9 X
bgaA
c
米勒单位
30
25
20
15
10
5
0
Ø
sp1
sp2
sp3
sp4
b
sp2
sp1
5'ACCATGTATACCACTTGGGCTTTGGCAGTAGCTAACTGCACTAAATATAATATAAGGAGGGTTTTATG 3'
3'TGGTACATATGGTGAACCCGAAACCGTCATCGATTGACCTGATTTATATTATATTCCTCCCAAAATAC 5'
sp3
sp4

图73

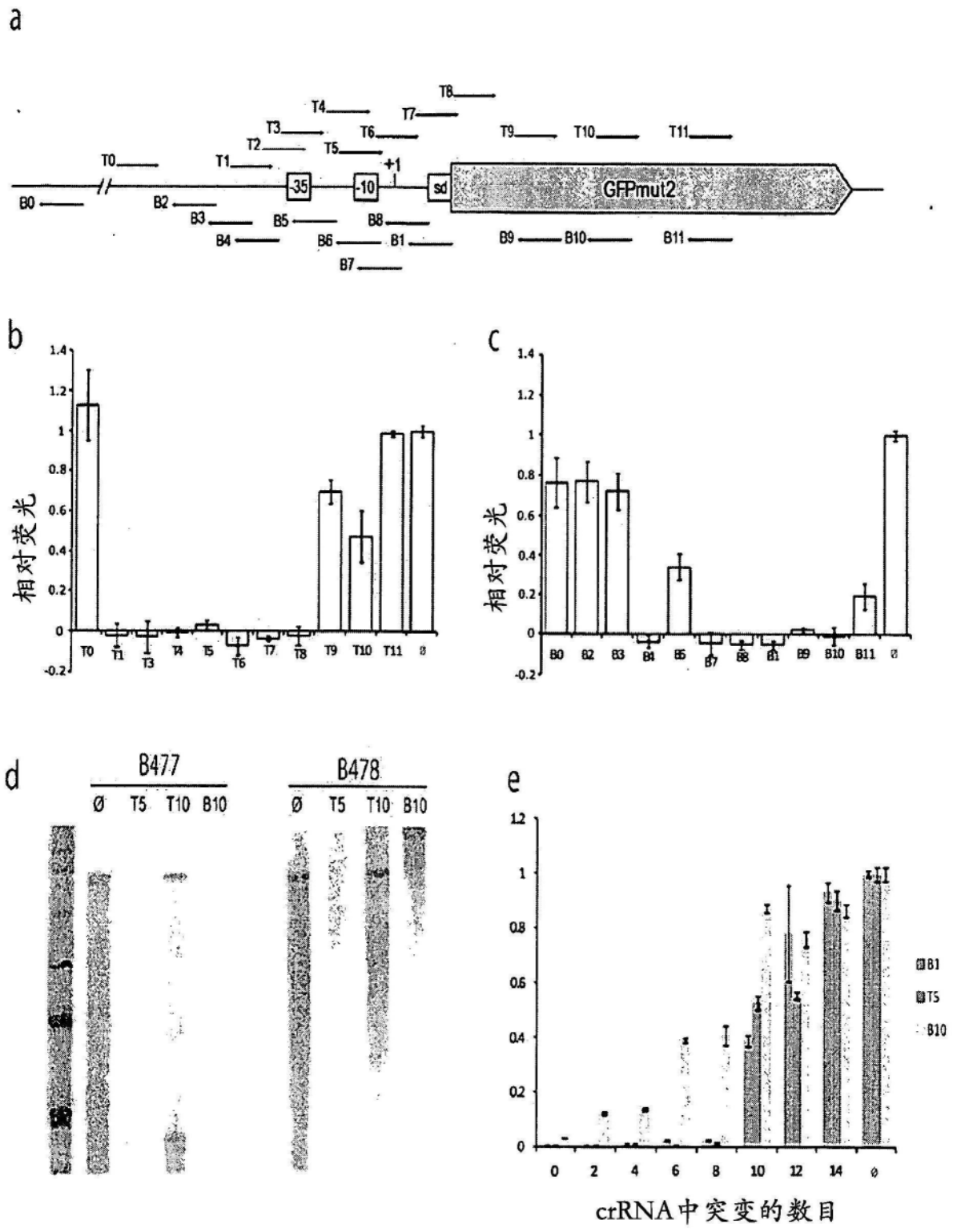

a
T8
T4
T7
T3
T6
T9
T10
T11
T2
T5
T0
T1
+1
-35
-10
sd
GFPmut2
B0
B2
B3
B5
B8
B4
B6
B1
B9
B10
B11
B7
b
相对荧光
1.4
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
T0
T1
T3
T4
T5
T6
T7
T8
T9
T10
T11
Ø
c
相对荧光
1.4
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
B0
B2
B3
B4
B5
B7
B8
B1
B9
B10
B11
Ø
d
B477
Ø
T5
T10
B10
B478
Ø
T5
T10
B10
e
1.2
1
0.8
0.6
0.4
0.2
0
0
2
4
6
8
10
12
14
Ø
B1
T5
B10
crRNA中突变的数目

图74