(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 119752887 A
(43)申请公布日 2025.04.04

(21)申请号 202411628310.8

(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
61/758,468 2013.01.30 US
61/769,046 2013.02.25 US
61/791,409 2013.03.15 US
61/802,174 2013.03.15 US
61/806,375 2013.03.28 US
61/814,263 2013.04.20 US
61/819,803 2013.05.06 US
61/828,130 2013.05.28 US
61/835,931 2013.06.17 US
61/836,127 2013.06.17 US

(62)分案原申请数据
201380070567.X 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
哈佛大学校长及研究员协会

(72)发明人 F·张 L·丛 P·苏 F·蓝

(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 王北南

(51)Int.Cl.
C12N 15/113(2010.01)
C12N 9/22(2006.01)
C07K 19/00(2006.01)
C12N 15/11(2006.01)
A61K 48/00(2006.01)
A61K 38/46(2006.01)

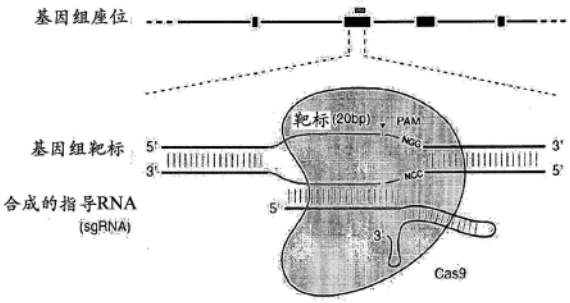
权利要求书2页 说明书79页 附图44页

(54)发明名称

用于序列操纵的系统、方法和优化的指导组合物的工程化

(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。还提供了指导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 119752887 A

1.一种组合物,其包含CRISPR-Cas系统嵌合RNA(chiRNA),其中所述chiRNA从5至3'包含:

(a)一种指导序列,其能够杂交到真核细胞细胞核中邻近原型间隔子邻近基序(PAM)的靶序列上,

(b)一种tracr配对序列,其能够杂交到tracr序列上,和

(c)一种tracr序列,其在长度上具有至少30个核苷酸,

其中所述chiRNA能够与Cas9蛋白形成CRISPR-Cas复合物并引导CRISPR-Cas复合物与真核细胞中邻近PAM的靶序列的序列特异性结合。

2.一种Cas9酶,其源自化脓链球菌或嗜热链球菌。

3.一种试剂盒,其包括Cas9酶。

4.一种融合蛋白,其包含RNAse III,mCherry和核定位信号(NLS)。

5.一种组合物,其包含选自以下的一种或多种PCR引物:AAAACCACCCTTCTCTCTGGC、GGAGATTGGAGACACGGAGAG、CTGGAAAGCCAATGCCTGAC和GGCAGCAAACTCCTTGTCCT。

6.根据权利要求1所述的组合物,其中所述tracr序列的长度为至少40个核苷酸。

7.根据权利要求1所述的组合物,其中所述tracr序列的长度为至少50个核苷酸。

8.根据权利要求1所述的组合物,其中所述指导序列的长度为15-25个核苷酸。

9.根据权利要求1所述的组合物,其中所述指导序列的长度为至少20个核苷酸。

10.根据权利要求1所述的组合物,其中chiRNA包含选自以下的RNA序列:

(a)NNNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAGAAAU AGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAG UGGCACCGAGUCGGUGCUUU;

(b)NNNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAGAAAU AGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAG UG;

(c)NNNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAGAAAU AGCAAGUUAAAAUAAGGCUAGUCCGUUAUCA。

11.根据权利要求1所述的组合物,其中chiRNA由选自以下的DNA序列编码:

(a)NNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATA GCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTG GCACCGAGTCGGTGCTTTTTT;

(b)NNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATA GCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTG TTTTTTT;

(c)NNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATA GCAAGTTAAAATAAGGCTAGTCCGTTATCATTTTTTTT。

12.根据权利要求1所述的组合物,其中所述chiRNA包含甲基化的核苷酸或核苷酸类似物。

13.根据权利要求1所述的组合物,其中所述chiRNA包含选自以下的化学修饰:2'-O-甲基类似物、2'-脱氧类似物、2'-氟代类似物、2-氨基嘌呤、5-溴代-尿苷、假尿苷、肌苷、7-甲基鸟苷、硫代磷酸骨架或其组合。

14.根据权利要求1和6-13中任一项所述的组合物,其还包含Cas9蛋白或编码Cas9蛋白的多核苷酸。

15.根据权利要求14所述的组合物,其中Cas9与一个或多个NLS融合,所述NLS独立地选自PKKKRKV、KRPAATKKAGQAKKKK、PAAKRVKLD、RQRRNELKRSP、NQSSNFGPMKGGNFGGRSSGPYGGGGQ

YFAKPRNQGGY、RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV、VSRKRPRP、PPKKARED、POPKKKPL、SALIKKKKKMAP、DRLRR、PKQKKRK、RKLKKKIKKL、REKKKFLKRR、KRKGDEVDGVDEVAKKKSKK和RKCLQAGMNLEARKTKK。

16.根据权利要求14所述的组合物,其中所述Cas9蛋白是化脓链球菌Cas9。

17.根据权利要求14所述的组合物,其中所述Cas9蛋白是嗜热链球菌Cas9。

18.根据权利要求1和6-17中任一项所述的组合物在制造用于治疗遗传疾病或病症的药物中的用途。

19.根据权利要求1和6-17中任一项所述的组合物在离体或体外修饰真核细胞中的用途。

20.一种真核细胞,其包含根据权利要求1和6-17中任一项所述的组合物。

21.一种非人类生物,其包含根据权利要求20所述的真核细胞,其中所述非人类生物不是动物或植物品种。

用于序列操纵的系统、方法和优化的指导组合物的工程化

[0001] 本申请是申请日为2013年12月12日、申请号为202110406914.8、发明名称为“用于序列操纵的系统、方法和优化的指导组合物的工程化”的发明专利申请的分案申请。该申请是申请日为2013年12月12日、申请号为201380070567.X、发明名称为“用于序列操纵的系统、方法和优化的指导组合物的工程化”的发明专利申请的分案申请。

[0002] 相关应用和引用参考

[0003] 本申请要求于2013年6月17日提交的标题为“用于序列操纵的系统、方法和优化的组合物的工程化(Engineering Of Systems,Methods And Optimized Compositions For Sequence Manipulation)”的美国临时专利申请61/836,127的优先权。本申请还要求美国临时专利申请61/758,468;61/769,046;61/802,174;61/806,375;61/814,263;61/819,803以及61/828,130的优先权,每个的标题为对用于序列操纵的系统、方法以及组合物进行的工程化和优化,分别提交于2013年1月30日;2013年2月25日;2013年3月15日;2013年3月28日;2013年4月20日;2013年5月6日以及2013年5月28日。还要求分别于2012年12月12日和2013年1月2日提交的标题均为“用于序列操纵的系统、方法与组合物(SYSTEMS METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION)”的美国临时专利申请61/736,527和61/748,427的优先权。还要求分别于2013年3月15日和2013年6月17日提交的标题均为BI-2011/008/44790.02.2003和BI-2011/008/44790.03.2003的美国临时专利申请61/791,409和61/835,931的优先权。

[0004] 还参考了美国临时专利申请61/835,936、61/836,101、61/836,080、61/836,123以及61/835,973,每个提交于2013年6月17日。

[0005] 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或(“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

[0006] 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。

[0007] 关于联邦资助研究的声明

[0008] 本发明是在美国国立卫生研究院颁发的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

[0010] 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

[0012] 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。

[0013] 在一个方面,本发明提供了包括一种或多种载体的载体系统。在一些实施例中,该系统包括:(a)一种第一调节元件,该第一调节元件可操作地连接到一种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复合物在细胞例如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)一种第二调节元件,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,该CRISPR酶包括一个核定位序列;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,该系统包括在一种第三调节元件(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载体通常是质粒形式。

[0014] 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)。

[0015] 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,成簇的规律间隔的短回文重复(CRISPR)转录物、蛋白质、酶、其突变体形式、其融合蛋白,等等)。

[0016] 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。

[0017] 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。

[0018] 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。

[0019] 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)一种第一调节元件,该第一调节元件可操作地连接到一种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)一种第二调节元件,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,该CRISPR酶包括一个核定位序列。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。

[0020] 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)一种第一调节元件,该第一调节元件可操作地连接到一种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)一种第二调节元件,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,该CRISPR酶包括一个核定位序列。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。

[0021] 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种tracr配对序列上,该tracr配对序列进而杂交到一种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种或多种载体递送到所述真核细胞,其中该一种或多种载体驱动下列一者或多者的表达: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。

[0022] 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,这样使得所述结合导致所述多核苷酸的表达增加或降低;其中该CRISPR复合物包括与杂交到在所述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种tracr配对序列上,该tracr配对序列进而杂交到一种tracr序列上。在一些实施例中,该方法进一步包括将一种或多种载体递送到所述真核细胞,其中该一种或多种载体驱动下列一者或多者的表达:该CRISPR酶、连接到该tracr配对序列上的指导序列、以及该tracr序列。

[0023] 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式真核细胞的方法。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)将一种或多种载体引入真核细胞中,其中该一种或多种载体驱动下列一者或多者的表达: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。

[0024] 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。

[0025] 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。

[0026] 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:将一种或多种载体引入一个或多个原核细胞中,其中该一种或多种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。

[0027] 在一些方面,本发明提供了一种非天然存在的或工程化的组合物,该组合物包括一个CRISPR-Cas系统嵌合RNA(chiRNA)多核苷酸序列,其中该多核苷酸序列包括(a)一个能够杂交到真核细胞中的靶序列上的指导序列、(b)一种tracr配对序列以及(c)一种tracr序列,其中(a)、(b)和(c)以5’到3’方向排列,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复合物包括与(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配对序列复合的CRISPR酶,

[0028] 或者

[0029] 一种CRISPR酶系统,其中该系统由一种载体系统编码,该载体系统包括一种或多种载体,该一种或多种载体包括I.一种第一调节元件,该第一调节元件可操作地连接到一个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括(a)一个或多个能够杂交到真核细胞中的一个或多个靶序列上的指导序列、(b)一种tracr配对序列和(c)一个或多个tracr序列,以及II.一种第二调节元件,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,该CRISPR酶包括至少一个或多个核定位序列,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复合物包括与(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配对序列复合的CRISPR酶,或一种多元CRISPR酶系统,其中该系统由一种载体系统编码,该载体系统包括一种或多种载体,该一种或多种载体包括I.一种第一调节元件,该第一调节元件可操作地连接到(a)一个或多个能够杂交到细胞中的靶序列上的指导序列、和(b)至少一个或多个tracr配对序列上、II.一种第二调节元件,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、以及III.一种第三调节元件,该第三调节元件可操作地连接到一种tracr序列上,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复合物包括与(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配对序列复合的CRISPR酶,并且其中在该多元系统中,使用了多个指导序列和一个单个的tracr序列;并且其中该指导序列、tracr序列和tracr配对序列中的一个或多个被修饰成改进稳定性。

[0030] 在本发明的方面中,该修饰包括工程化的二级结构。例如,该修饰可以包括减少该tracr配对序列与该tracr序列之间的杂交区域。例如,该修饰还可以包括通过一个人工环融合该tracr配对序列和该tracr序列。该修饰可以包括具有40与120bp之间的长度的tracr序列。在本发明的实施例中,该tracr序列在该tracr的40bp与全长之间。在某些实施例中,tracRNA的长度至少包括野生型tracRNA的核苷酸1-67并且在一些实施例中,至少包括核苷酸1-85。在一些实施例中,可以至少使用对应于野生型化脓链球菌Cas9 tracRNA的核苷酸1-67或1-85的核苷酸。在该CRISPR系统使用除Cas9之外或除SpCas9之外的酶的情况下,则可以存在相关野生型tracRNA中的对应的核苷酸。在一些实施例中,tracRNA的长度包括不超过野生型tracRNA的核苷酸1-67或1-85。该修饰可以包括序列优化。在某些方面,序列优化可以包括减少polyT序列在该tracr和/或tracr配对序列中的发生率。序列优化可以结合对该tracr配对序列与该tracr序列之间的杂交区域的减少;例如,减少tracr序列长度。

[0031] 在本发明的一个方面,提供了CRISPR-Cas系统或CRISPR酶系统,其中该修饰包括减少该tracr和/或tracr配对序列中的polyT序列。在本在本发明的一些方面,存在于相关野生型序列的poly-T序列中的一个或多个T(也就是说,一串超过3、4、5、6或更多个连续T碱基;在一些实施例中,一串不超过10、9、8、7、6个连续T碱基)可以被非T核苷酸(例如,A)取代,这样使得将该串断裂为较小的T段,其中每段具有4个或少于4个(例如,3个或2个)连续的T。除A之外的碱基可以用于取代,例如C或G、或非天然存在的核苷酸或经修饰的核苷酸。如果该T串涉及发夹(或茎环)的形成,则有利的是改变非T碱基的互补碱基,以与非T核苷酸互补。例如,如果非T碱基是A,则可以将其互补体换为T,例如以保留或辅助保留二级结构。例如,5’-TTTTT可以被变为5’-TTTAT并且可以将互补5’-AAAAA改变成5’-ATAAA。

[0032] 在一个方面,本发明提供了CRISPR-Cas系统或CRISPR酶系统,其中该修饰包括添加一个polyT终止子序列。在一个方面,本发明提供了CRISPR-Cas系统或CRISPR酶系统,其中该修饰包括在tracr和/或tracr配对序列中添加一个polyT终止子序列。在一个方面,本发明提供了CRISPR-Cas系统或CRISPR酶系统,其中该修饰包括在该指导序列中添加一个polyT终止子序列。该polyT终止子序列可以包括5个或超过5个连续T碱基。

[0033] 在一个方面,本发明提供了CRISPR-Cas系统或CRISPR酶系统,其中该修饰包括改变环和/或发夹。在一个方面,本发明提供了CRISPR-Cas系统或CRISPR酶系统,其中该修饰包括在该指导序列中提供最低两个发夹。在一个方面,本发明提供了CRISPR-Cas系统或CRISPR酶系统,其中该修饰包括提供通过tracr与tracr配对(同向重复)序列之间的互补形成的发夹。在一个方面,本发明提供了CRISPR-Cas系统或CRISPR酶系统,其中该修饰包括在该tracrRNA序列的3'末端处或朝向该tracrRNA序列的3'末端提供一个或多个另外的发夹。例如,可以通过以下方式形成一个发夹:在该tracRNA序列内提供由环连接的自互补序列,这样使得在自折叠时形成一个发夹。在一个方面,本发明提供了CRISPR-Cas系统或CRISPR酶系统,其中该修饰包括提供另外的添加到该指导序列的3'的发夹。在一个方面,本发明提供了CRISPR-Cas系统或CRISPR酶系统,其中该修饰包括延长该指导序列的5'末端。在一个方面,本发明提供了CRISPR-Cas系统或CRISPR酶系统,其中该修饰包括在该指导序列的5'末端中提供一个或多个发夹。在一个方面,本发明提供了CRISPR-Cas系统或CRISPR酶系统,其中该修饰包括将序列(5'-AGGACGAAGTCCTAA)附加到该指导序列的5'末端上。其他适于形成发夹的序列应该是技术人员已知的,并且可以用于本发明的某些方面。在本发明的一些方面,提供了至少2、3、4、5或更多个另外的发夹。在本发明的一些方面,提供了不超过10、9、8、7、6个另外的发夹。在一个方面,本发明提供了CRISPR-Cas系统或CRISPR酶系统,其中该修饰包括两个发夹。在一个方面,本发明提供了CRISPR-Cas系统或CRISPR酶系统,其中该修饰包括三个发夹。在一个方面,本发明提供了CRISPR-Cas系统或CRISPR酶系统,其中该修饰包括至多五个发夹。

[0034] 在一个方面,本发明提供了CRISPR-Cas系统或CRISPR酶系统,其中该修饰包括在该多核苷酸序列中提供交联或提供一个或多个经修饰的核苷酸。可以在tracr序列、tracr配对序列和/或指导序列中任何或所有序列中,和/或在酶编码序列中,和/或在载体序列中,提供经修饰的核苷酸和/或交联。修饰可以包括包含至少一个非天然存在的核苷酸、或一个经修饰的核苷酸、或其类似物。经修饰的核苷酸可以是在核糖、磷酸和/或碱基部分处被修饰。经修饰的核苷酸可以包括2'-O-甲基类似物、2'-脱氧类似物或2'-氟代类似物。可以修饰核酸骨架,例如,可以使用硫代磷酸骨架。使用锁核酸(LNA)或桥联核酸(BNA)也是可以的。经修饰的碱基的另外的实例包括但不限于2-氨基嘌呤、5-溴代-尿苷、假尿苷、肌苷、7-甲基鸟苷。

[0035] 应该理解的是,可以分开地或与给定的CRISPR-Cas系统或CRISPR酶系统组合地提供以上修饰中的任何或所有修饰。这样的一种系统可以包括所述修饰中的一种、两种、三种、四种、五种或更多种。

[0036] 在一个方面,本发明提供了CRISPR-Cas系统或CRISPR酶系统,其中该CRISPR酶是一种II型CRISPR系统酶,例如Cas9酶。在一个方面,本发明提供了CRISPR-Cas系统或CRISPR酶系统,其中该CRISPR酶由少于一千个氨基酸或少于四千个氨基酸组成。在一个方面,本发明提供了CRISPR-Cas系统或CRISPR酶系统,其中该Cas9酶是StCas9或St1Cas9,或该Cas9酶是一种来自选自下组的生物的Cas9酶,该组由以下各项组成:链球菌属、弯曲杆菌属、Nitratifractor、葡萄球菌属、细小棒菌属(Parvibaculum)、罗氏菌属(Roseburia)、奈瑟菌属、葡糖醋杆菌属、固氮螺菌属、Sphaerochaeta、乳杆菌属、真杆菌属或棒状杆菌属。在一个方面,本发明提供了CRISPR-Cas系统或CRISPR酶系统,其中该CRISPR酶是一种引导在该靶

序列的位置处的两条链的切割的核酸酶。

[0037] 在一个方面,本发明提供了CRISPR-Cas系统或CRISPR酶系统,其中该第一调节元件是一种聚合酶III启动子。在一个方面,本发明提供了CRISPR-Cas系统或CRISPR酶系统,其中该第二调节元件是一种聚合酶II启动子。

[0038] 在一个方面,本发明提供了CRISPR-Cas系统或CRISPR酶系统,其中该指导序列包括至少十五个核苷酸。

[0039] 在一个方面,本发明提供了CRISPR-Cas系统或CRISPR酶系统,其中该修饰包括优化的tracr序列和/或优化的指导序列RNA,和/或tracr序列和/或一个或多个tracr配对序列的共折叠结构,和/或稳定tracr序列的二级结构,和/或具有减少的碱基配对区域的tracr序列,和/或与tracr序列融合的RNA元件;和/或,在该多元系统中,存在两个包含tracer并且包含多个指导物的RNA或一个包含多个嵌合体的RNA。

[0040] 在本发明的方面,根据诱变研究的结果进一步优化该嵌合RNA构造。在具有两个或更多个发夹的嵌合RNA中,用于稳定发夹的近端同向重复中的突变可以导致CRISPR复合物活性的消除。用于缩短或稳定发夹的远端同向重复中的突变对CRISPR复合物活性没有影响。近端与远端重复之间的凸出区域中的序列随机化可以显著降低CRISPR复合物活性。发夹之间的接头区域中的单碱基对变化或序列随机化可以导致完全失去CRISPR复合物活性。在该指导序列后的第一个发夹之后的远端发夹的发夹稳定化可以导致CRISPR复合物活性的维持或改进。因此,在本发明的优选实施例中,可以通过产生一个较小的嵌合RNA而进一步优化该嵌合RNA构造,该较小的嵌合RNA对于治疗性递送选择以及其他用途而言可以是有益的,并且这可以通过改变远端同向重复以便缩短或稳定该发夹而实现。在本发明的另外优选的实施例中,可以通过稳定远端发夹中的一个或多个而进一步优化该嵌合RNA构造。发夹的稳定可以包括修饰适于形成发夹的序列。在本发明的一些方面,提供了至少2、3、4、5或更多个另外的发夹。在本发明的一些方面,提供了不超过10、9、8、7、6个另外的发夹。在本发明的一些方面,稳定可以是交联以及其他修饰。修饰可以包括包含至少一个非天然存在的核苷酸、或一个经修饰的核苷酸、或其类似物。经修饰的核苷酸可以是在核糖、磷酸和/或碱基部分处被修饰。经修饰的核苷酸可以包括2'-O-甲基类似物、2'-脱氧类似物或2'-氟代类似物。可以修饰核酸骨架,例如,可以使用硫代磷酸骨架。使用锁核酸(LNA)或桥联核酸(BNA)也是可以的。修饰的碱基的另外的实例包括但不限于2-氨基嘌呤、5-溴代-尿苷、假尿苷、肌苷、7-甲基鸟苷。

[0041] 在一个方面,本发明提供了CRISPR-Cas系统或CRISPR酶系统,其中该CRISPR酶是经密码子优化的以表达于真核细胞中。

[0042] 因此,在本发明的一些方面,在本发明的构建体(例如,嵌合构建体)中所需的tracRNA的长度不一定是固定的,并且在本发明的一些方面,它可以是在40与120bp之间,并且在本发明的一些方面,是长达tracr的全长,例如,在本发明的一些方面,直到tracr的3'末端,如由细菌基因组中的转录终止信号所打断的。在某些实施例中,tracRNA的长度至少包括野生型tracRNA的核苷酸1-67并且在一些实施例中,至少包括核苷酸1-85。在一些实施例中,可以至少使用对应于野生型化脓链球菌Cas9 tracRNA的核苷酸1-67或1-85的核苷酸。在该CRISPR系统使用除Cas9之外或除SpCas9之外的酶的情况下,则可以存在相关野生型tracRNA中的对应的核苷酸。在在一些实施例中,tracRNA的长度包括不超过野生型

tracRNA的核苷酸1-67或1-85。相对于序列优化(例如,减少polyT序列),例如关于该tracr配对(同向重复)或tracrRNA内部的T串,在本发明的一些方面,存在于相关野生型序列的poly-T序列中的一个或多个T(也就是说,一串超过3、4、5、6或更多个连续T碱基;在一些实施例中,一串不超过10、9、8、7、6个连续T碱基)可以被非T核苷酸(例如,A)取代,这样使得将该串断裂为较小的T段,其中每段具有4个或少于4个(例如,3个或2个)连续的T。如果该T串涉及发夹(或茎环)的形成,则有利的是改变非T碱基的互补碱基,以与非T核苷酸互补。例如,如果非T碱基是A,则可以将其互补体换为T,例如以保留或辅助保留二级结构。例如,5'-TTTTT可以被变为5'-TTTAT并且可以将互补5'-AAAAA改变成5'-ATAAA。关于polyT终止子序列在tracr+tracr配对转录本中的存在,例如,polyT终止子(TTTTT或更多),在本发明的一些方面,有利的是将这样的终止子序列添加到转录本的末尾,无论该转录本处于两个RNA(tracr和tracr配对)还是单个指导RNA形式。就tracr和tracr配对转录本中的环和发夹而言,在本发明的一些方面,有利的是最低两个发夹存在于该嵌合指导RNA中。第一个发夹可以是通过tracr与tracr配对(同向重复)序列之间的互补形成的发夹。第二个发夹可以在该tracrRNA序列的3'末端处,并且这可以提供二级结构以用于与Cas9相互作用。可以将另外的发夹添加到该指导RNA的3',例如,在本发明的一些方面,用于增加该指导RNA的稳定性。另外,在本发明的一些方面,可以延长该指导RNA的5'末端。在本发明的一些方面,可以将5'末端中的20bp考虑为指导序列。可以延长5'部分。可以在5'部分中提供一个或多个发夹,例如,在本发明的一些方面,这也可以改进该指导RNA的稳定性。在本发明的一些方面,可以通过将序列(5'-AGGACGAAGTCCTAA)附加到该指导序列的5'末端上而提供特定的发夹,并且在本发明的一些方面,这可以有助于改进稳定性。其他适于形成发夹的序列应该是技术人员已知的,并且可以用于本发明的某些方面。在本发明的一些方面,提供了至少2、3、4、5或更多个另外的发夹。在本发明的一些方面,提供了不超过10、9、8、7、6个另外的发夹。上述内容还提供了涉及指导序列中的二级结构的本发明的方面。在本发明的一些方面,可以存在交联以及其他修饰,例如,以改进稳定性。修饰可以包括包含至少一个非天然存在的核苷酸、或一个经修饰的核苷酸、或其类似物。经修饰的核苷酸可以是在核糖、磷酸和/或碱基部分处被修饰。经修饰的核苷酸可以包括2'-O-甲基类似物、2'-脱氧类似物或2'-氟代类似物。可以修饰核酸骨架,例如,可以使用硫代磷酸骨架。使用锁核酸(LNA)或桥联核酸(BNA)也是可以的。修饰的碱基的另外的实例包括但不限于2-氨基嘌呤、5-溴代-尿苷、假尿苷、肌苷、7-甲基鸟苷。这样的修饰或交联可以存在于该指导序列或邻近该指导序列的其他序列中。

[0043] 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。

[0044] 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。

[0045] 本发明还涉及以下项目:

[0046] 1.一种非天然存在的或工程化的组合物,该组合物包含:

[0047] A)一种CRISPR-Cas系统嵌合RNA(chiRNA)多核苷酸序列,其中该多核苷酸序列包含

[0048] (a)一种指导序列,该指导序列能够杂交到真核细胞中的靶序列上,

[0049] (b)一种tracr配对序列,和

[0050] (c)一种tracr序列

[0051] 其中(a)、(b)和(c)以5’到3’方向排列,

[0052] 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,

[0053] 其中该CRISPR复合物包含与(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配对序列复合的CRISPR酶,

[0054] 或者

[0055] B)一种CRISPR酶系统,其中该系统由一种载体系统编码,该载体系统包括一种或多种载体,该一种或多种载体包括

[0056] I.一种第一调节元件,该第一调节元件可操作地连接到一种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:

[0057] (a)一个或多个指导序列,该一个或多个指导序列能够杂交到真核细胞中的一个或多个靶序列上,

[0058] (b)一种tracr配对序列,和

[0059] (c)一个或多个tracr序列,以及

[0060] II.一种第二调节元件,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,该CRISPR酶包含至少一个或多个核定位序列,

[0061] 其中(a)、(b)和(c)以5’到3’方向排列,

[0062] 其中组分I和II位于该系统的相同或不同载体上,

[0063] 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,

[0064] 其中该CRISPR复合物包含与(1)杂交到该靶序列上的该指导序列、和(2)杂交到该tracr序列上的该tracr配对序列复合的CRISPR酶,

[0065] 或者

[0066] C)一种多元CRISPR酶系统,其中该系统由一种载体系统编码,该载体系统包括一种或多种载体,该一种或多种载体包括

[0067] I.一种第一调节元件,该第一调节元件可操作地连接到一种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:

[0068] (a)一个或多个指导序列,该一个或多个指导序列能够杂交到真核细胞中的一个或多个靶序列上,

[0069] (b)一种tracr配对序列,和

[0070] (c)一个或多个tracr序列,以及

[0071] II.一种第二调节元件,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,该CRISPR酶包含至少一个或多个核定位序列,

[0072] 其中(a)、(b)和(c)以5'到3'方向排列,

[0073] 其中组分I和II位于该系统的相同或不同载体上,

[0074] 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,

[0075] 其中该CRISPR复合物包含与(1)杂交到该靶序列上的该指导序列、和(2)杂交到该tracr序列上的该tracr配对序列复合的CRISPR酶,

[0076] 其中在该多元系统中,使用多个chiRNA多核苷酸序列

[0077] 或者

[0078] D)一种多元CRISPR酶系统,其中该系统由一种载体系统编码,该载体系统包括一种或多种载体,该一种或多种载体包括

[0079] I.一种第一调节元件,该第一调节元件可操作地连接至

[0080] (a)能够杂交到细胞中的靶序列上的一个或多个指导序列,以及

[0081] (b)至少一个或多个tracr配对序列,

[0082] II.一种第二调节元件,该第二调节元件可操作地连接至编码CRISPR酶的酶编码序列,以及

[0083] III.一种第三调节元件,该第三调节元件可操作地连接至一种tracr序列,

[0084] 其中组分I、II和III位于该系统的相同或不同载体上,

[0085] 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,

[0086] 其中该CRISPR复合物包含与(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配对序列复合的CRISPR酶,并且

[0087] 其中在该多元系统中,使用多个指导序列以及单个tracr序列;

[0088] 并且

[0089] 其中,在A)的多核苷酸序列中,或在B)、C)或D)的系统中,该指导序列、tracr序列和tracr配对序列中的一种或多种被修饰以改进稳定性。

[0090] 2.如项目1所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该修饰包括一个工程化的二级结构。

[0091] 3.如项目1或项目2所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该修饰包括减少该tracr配对序列与该tracr序列之间的杂交区域。

[0092] 4.如任一前述项目所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该修饰包括通过一个人工环融合该tracr配对序列和该tracr序列。

[0093] 5.如任一前述项目所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该修饰包括具有40与120bp之间的长度的tracr序列。

[0094] 6.如任一前述项目所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该tracr序列是在该tracr的40bp与全长之间。

[0095] 7.如任一前述项目所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该tracr序列至少包括对应的野生型tracRNA的核苷酸1-67。

[0096] 8.如任一前述项目所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该tracr序列至少包括对应的野生型tracRNA的核苷酸1-85。

[0097] 9.如任一前述项目所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该tracr序列包括对应于野生型化脓链球菌Cas9 tracRNA的核苷酸1-67的核苷酸。

[0098] 10.如任一前述项目所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该tracr序列包括对应于野生型化脓链球菌Cas9 tracRNA的核苷酸1-85的核苷酸。

[0099] 11.如项目9所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该tracr序列基本上由对应于野生型化脓链球菌Cas9 tracRNA的核苷酸1-67的核苷酸组成。

[0100] 12.如项目10所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该tracr序列基本上由对应于野生型化脓链球菌Cas9 tracRNA的核苷酸1-85的核苷酸组成。

[0101] 13.如任一前述项目所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该修饰包括序列优化。

[0102] 14.如项目13所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该修饰包括减少该tracr和/或tracr配对序列中的polyT序列。

[0103] 15.如项目14所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中存在于相关野生型序列的poly-T序列中的一个或多个T已经被一个非T核苷酸取代。

[0104] 16.如项目13、14或15所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该经修饰的序列不包括任何具有多于4个连续T的polyT序列。

[0105] 17.如任一前述项目所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该修饰包括添加一个polyT终止子序列。

[0106] 18.如项目17所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该修饰包括在tracr和/或tracr配对序列中添加一个polyT终止子序列。

[0107] 19.如项目17或18所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该修饰包括在该指导序列中添加一个polyT终止子序列。

[0108] 20.如任一前述项目所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该修饰包括改变环和/或发夹。

[0109] 21.如项目20所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该修饰包括在该指导序列中提供最低两个发夹。

[0110] 22.如项目20或21所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该修饰包括提供一个通过该tracr与tracr配对序列之间的互补而形成的发夹。

[0111] 23.如项目20至22所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该修饰包括在该tracrRNA序列的3'末端提供一个或多个另外的发夹。

[0112] 24.如项目20至23所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该修饰包括提供一个或多个添加到该指导序列的3'的另外的发夹。

[0113] 25.如任一前述项目所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该修饰包括延长该指导序列的5'末端。

[0114] 26.如项目25所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该修饰包括在

该指导序列的5'末端中提供一个或多个发夹。

[0115] 27.如项目25或26所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该修饰包括将序列(5'-AGGACGAAGTCCTAA)附加到该指导序列的5'末端。

[0116] 28.如任一前述项目所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该修饰包括在该多核苷酸序列中提供交联或提供一个或多个经修饰的核苷酸。

[0117] 29.如项目28所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中在该tracr序列、tracr配对序列和/或指导序列中的任何或所有序列中,和/或在该酶编码序列中,和/或在载体序列中,提供经修饰的核苷酸。

[0118] 30.如项目28或29所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中提供经修饰的核苷酸包括包含至少一个非天然存在的核苷酸、或一个经修饰的核苷酸、或其类似物。

[0119] 31.如项目30所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中这些经修饰的核苷酸是在核糖、磷酸和/或碱基部分处被修饰。

[0120] 32.如项目30所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该经修饰的核苷酸选自下组,该组由以下各项组成:2'-O-甲基类似物、2'-脱氧类似物或2'-氟代类似物。

[0121] 33.如项目30所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该经修饰的核苷酸选自下组,该组由以下各项组成:2-氨基嘌呤、5-溴代-尿苷、假尿苷、肌苷、7-甲基鸟苷。

[0122] 34.如任一前述项目所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该修饰包括两个发夹。

[0123] 35.如任一前述项目所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该修饰包括三个发夹。

[0124] 36.如任一前述项目所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该修饰包括至多五个发夹。

[0125] 37.如任一前述项目所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该CRISPR酶是一种II型CRISPR系统酶。

[0126] 38.如任一前述项目所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该CRISPR酶是Cas9酶。

[0127] 39.如任一前述项目所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该CRISPR酶由少于一千个氨基酸组成。

[0128] 40.如项目1至38中任一项所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该CRISPR酶由少于四千个氨基酸组成。

[0129] 41.如任一前述项目所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该Cas9酶是StCas9或StlCas9。

[0130] 42.如任一前述项目所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该Cas9酶是一种来自选自下组的生物的Cas9酶,该组由以下各项组成:链球菌属、弯曲杆菌属、Nitratifractor、葡萄球菌属、细小棒菌属、罗氏菌属、奈瑟菌属、葡糖醋杆菌属、固氮螺菌属、Sphaerochaeta、乳杆菌属、真杆菌属或棒状杆菌属。

[0131] 43.如任一前述项目所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该CRISPR酶是一种引导在该靶序列的位置处的两条链的切割的核酸酶。

[0132] 44.如任一前述项目所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该第一调节元件是聚合酶III启动子。

[0133] 45.如任一前述项目所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该第二调节元件是聚合酶II启动子。

[0134] 46.如任一前述项目所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该指导序列包括至少十五个核苷酸。

[0135] 47.如任一前述项目所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该修饰包括优化的tracr序列和/或优化的指导序列RNA,和/或tracr序列和/或一个或多个tracr配对序列的共折叠结构,和/或稳定tracr序列的二级结构,和/或具有减少的碱基配对区域的tracr序列,和/或与tracr序列融合的RNA元件;和/或,在该多元系统中,存在两个包含tracer并且包含多个指导物的RNA或一个包含多个嵌合体的RNA。

[0136] 48.如项目1-47中任一项所述的CRISPR-Cas系统chiRNA或CRISPR酶系统,其中该CRISPR酶是经密码子优化的以表达于真核细胞中。

[0137] 49.如任一前述项目所述的组合物,其中该组合物包括CRISPR-Cas系统嵌合RNA(chiRNA)多核苷酸序列。

[0138] 50.如项目49所述的组合物,其中该组合物进一步包括一个编码CRISPR酶的多核苷酸序列,该CRISPR酶包括至少一个或多个核定位序列。

[0139] 51.如任一前述项目所述的CRISPR酶系统。

[0140] 52.如任一前述项目所述的多元CRISPR酶系统。

[0141] 53.如项目49或50所述的组合物、如项目51所述的CRISPR酶系统或如项目52所述的多元CRISPR酶系统的转录或翻译产物。

附图简要说明

[0143] 本发明的新颖特征在所附权利要求书中具体提出。通过参考对说明性实施例进行叙述的以下详细说明,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:

[0144] 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)。与DNA靶(蓝色)的指导序列碱基对,直接地在必要的5'-NGG原型间隔子邻近基序(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)。

[0145] 图2A-F展示了一个示例性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。

[0146] 图3A-C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。

[0147] 图4A-D展示了针对示例性靶标的SpCas9特异性的评估结果。

[0148] 图5A-G展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。

[0149] 图6A-C展示了对于Cas9介导的基因靶向的不同tracrRNA转录本的比较。

[0150] 图7A-D展示了一个示例性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活性的测试的结果。

[0151] 图8A-C展示了用于在哺乳动物细胞中的基因组座位的靶向的CRISPR系统的示例性操纵。

[0152] 图9A-B展示了在哺乳动物中的crRNA加工的Northern印迹分析的结果。

[0153] 图10A-C展示了嵌合RNA的示意表示和真核细胞中针对CRISPR系统活性的SURVEYOR测定的结果。

[0154] 图11A-B展示了在真核细胞中针对CRISPR系统活性的SURVEYOR测定的结果的图示。

[0155] 图12示出了示例性嵌合RNA的预测二级结构,这些嵌合RNA包括指导序列、tracr配对序列和tracr序列。

[0156] 图13是Cas基因的系统发生树。

[0157] 图14A-F显示了揭示五个Cas9家族的系统发生分析,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)。

[0158] 图15显示了一个图表,描绘了不同的优化的指导RNA的功能。

[0159] 图16显示了不同的指导嵌合RNA的序列和结构。

[0160] 图17显示了tracrRNA和同向重复的共折叠结构。

[0161] 图18A和B显示了来自St1Cas9嵌合指导RNA体外优化的数据。

[0162] 图19A-B显示了SpCas9细胞裂解物对未甲基化的或甲基化的靶标的切割。

[0163] 图20A-G显示了SpCas9介导的哺乳动物基因组编辑的指导RNA构造的优化。(a)用于所有后续实验的、U6启动子驱动的单个指导RNA(sgRNA)和CBh启动子驱动的人类密码子优化的化脓链球菌Cas9(hSpCas9)的双顺反子表达载体(PX330)的示意图。该sgRNA由一个20-nt指导序列(蓝色)和支架(红色)组成,其在如所指示的不同位置被截短。(b)人类EMX1和PVALB座位处的SpCas9介导的插入缺失(indel)的SURVEYOR测定。箭头指示预期的SURVEYOR片段(n=3)。(c)四个sgRNA截短构造的Northern印迹分析,其中U1作为加样对照。(d)SpCas9的野生型(wt)或切口酶突变体(D10A)两者均促进HindIII位点插入进人EMX1基因中。将按相对于基因组序列有义或反义方向定向的单链寡核苷酸(ssODN)用作同源重组模板。(e)人SERPINB5座位的示意图。sgRNA和PAM由序列上的彩色条指示;甲基胞嘧啶(Me)被高亮(粉色)并且相对于转录起始位点(TSS,+1)而编号。(f)通过16个克隆的亚硫酸氢盐测序所测定的SERPINB5的甲基化状态。实心圆,甲基化的CpG;空心圆,未甲基化的CpG。(g)被三个靶向SERPINB5的甲基化区域的sgRNA修饰的效率,通过深度测序测定(n=2)。误差条指示威尔逊(Wilson)区间(在线方法)。

[0164] 图21A-B显示了CRISPR-Cas sgRNA构造的进一步优化。(a)四个另外的sgRNA构造I-IV的示意图。每个构造都由一个连接到同向重复(DR,灰色)上的20-nt指导序列(蓝色)组成,该同向重复杂交到tracrRNA(红色)上。该DR-tracrRNA杂交体如所指示的在+12或+22处被一个人工GAAA茎环截短。根据先前针对tracrRNA报道的转录起始位点为tracrRNA截短位置编号。sgRNA构造II和IV在其poly-U束内携带突变,这些束可以充当提前转录终止子。(b)在人EMX1座位处针对靶位点1-3,SpCas9介导的indel的SURVEYOR测定。箭头指示预期的SURVEYOR片段(n=3)。

[0165] 图22展示了一些靶位点在人类基因组中的可视化。

[0166] 图23A-B显示了(A)sgRNA的示意图和(B)SaCas9的五种sgRNA变体的针对具有最高切割效率的优化的截短构造的SURVEYOR分析。

[0167] 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。

[0168] 本发明的详细说明

[0169] 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。

[0170] 在本发明的多个方面,术语“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。

[0171] 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。

[0172] 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。

[0173] 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。

[0174] “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。

[0175] 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。

[0176] “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”。

[0177] 如本文使用的,相对于CRISPR系统的组分,“稳定化”或“增加稳定性”涉及固定或稳定分子的结构。这可以通过以下方式完成:引入一个或多个突变(包括单个或多个碱基对改变),增加发夹的数目,交联,断开具体的核苷酸段以及其他修饰。修饰可以包括包含至少一个非天然存在的核苷酸、或一个经修饰的核苷酸、或其类似物。经修饰的核苷酸可以是在核糖、磷酸和/或碱基部分处被修饰。经修饰的核苷酸可以包括2’-O-甲基类似物、2’-脱氧类似物或2’-氟代类似物。可以修饰核酸骨架,例如,可以使用硫代磷酸骨架。使用锁核酸(LNA)或桥联核酸(BNA)也是可以的。修饰的碱基的另外的实例包括但不限于2-氨基嘌呤、5-溴代-尿苷、假尿苷、肌苷、7-甲基鸟苷。这些修饰可以应用于该CRISPR系统的任何组分。在一个优选实施例中,可以对RNA组分(例如该指导RNA或嵌合多核苷酸序列)做出这些修饰。

[0178] 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。

[0179] 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。

[0180] 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。在一些实施例中,受试者可以是无脊椎动物,例如昆虫或线虫;而在其他实施例中,受试者可以是植物或真菌。

[0181] 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。

[0182] 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。

[0183] 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。

[0184] 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))。

[0185] 本发明的若干方面涉及包括一种或多种载体的载体系统,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。

[0186] 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。

[0187] 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。

[0188] 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))。

[0189] 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)。

[0190] 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。

[0191] 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)。

[0192] 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),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)。

[0193] 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。全互补性不是必需的,其条件是存在引起杂交和促进形成CRISPR复合物的足够的互补性。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。

[0194] 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tracr配对序列具有足够的互补性以杂交并且参与形成CRISPR复合物。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种或多种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。

[0195] 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)位于一种或多种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引导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。

[0196] 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。此组合允许两个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,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)。

[0197] 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。

[0198] 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。

[0199] 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。

[0200] 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。

[0201] 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,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,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引导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Surveyor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。

[0202] 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MMMMMMMMNNNNNNNNNNNNXGG的化脓链球菌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。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MMMMMMMMMNNNNNNNNNNNXXAGAAW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MMMMMMMMMNNNNNNNNNNNXGGXG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。

[0203] 在一些实施例中,指导序列被选择为降低在该指导序列内的二级结构程度。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(博德参考号BI-2012/08444790.11.2022)中。

[0204] 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7B以及13B中提供了一种tracr序列以及一种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)NNNNNNNNNNNNNNNNNNNgttttgtactctcaagatttaGAAAtaaatcttgcagaagctacaaagataaggcttcatgccgaaatcaacaccctgtcattttatggcagggtgttttcgttatttaaTTTTTT;(2)NNNNNNNNNNNNNNNNNNNgttttgtactctcaGAAAtgcagaagctacaaagataaggcttcatgccgaaatcaacaccctgtcattttatggcagggtgttttcgttatttaaTTTTTT;(3)NNNNNNNNNNNNNNNNNNNgttttgtactctcaGAAAtgcagaagctacaaagataaggcttcatgccgaaatcaacaccctgtcattttatggcagggtgtTTTTTT;(4)NNNNNNNNNNNNNNNNNNNgttttagagctaGAAAtagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgcTTTTTT;(5)NNNNNNNNNNNNNNNNNNNgttttagagctaGAAATAGcaagttaaaataaggctagtccgttatcaacttgaaaaagtgTTTTTTT;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)。

[0205] 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。

[0206] 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、转录阻抑活性、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US20110059502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。

[0207] 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种或多种多核苷酸,如或如在此描述的一种或多种载体、其一种或多种转录本和/或一种或多种自其转录的蛋白。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);哈嗒嗒(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)。

[0208] 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)。

[0209] 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)。

[0210] 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。

[0211] 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/US94/05700)。

[0212] 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)。

[0213] 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US20030087817。

[0214] 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、Hep G2、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,使用用一种或多种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种或多种载体来源的序列。在一些实施例中,使用用如在此描述的CRISPR系统的组分转染(如通过用一种或多种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。

[0215] 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。还提供了转基因动物,正如转基因植物一样,尤其是作物和藻类。转基因动物或植物可用于在除了提供疾病模型之外的应用中。通过表达例如比通常在野生型中可见的更高的蛋白质、碳水化合物、营养素或维生素水平,这些应用可包括食物或饲料生产。在这方面,转基因植物,尤其是豆类和块茎类,以及动物,尤其是哺乳动物,如家畜(奶牛、绵羊、山羊和猪),而且还有禽类和食用昆虫,是优选的。

[0216] 转基因藻类或其他植物,如油菜,可以在植物油或生物燃料例如像醇类(尤其是甲醇和乙醇)的生产中特别有用。这些可以被工程化为表达或过表达高水平的油或醇类,以供在油或生物燃料行业中使用。

[0217] 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种tracr配对序列上,该tracr配对序列进而杂交到一种tracr序列上。

[0218] 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,这样使得所述结合导致所述多核苷酸的表达增加或降低;其中该CRISPR复合物包括与杂交到在所述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种tracr配对序列上,该tracr配对序列进而杂交到一种tracr序列上。

[0219] 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选择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US2009/0100536-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14)。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。

[0220] 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。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。

[0221] 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)一种第一调节元件,该第一调节元件可操作地连接到一种tracr配对序列以及用于在该tracr配对序列的上游插入一种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)一种第二调节元件,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。

[0222] 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。

[0223] 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复合物包括与一种指导序列复合的CRISPR酶,该指导序列与靶多核苷酸内的靶序列杂交。该指导序列与一种tracr配对序列连接,该tracr配对序列进而与一种tracr序列杂交。

[0224] 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)。不希望被理论所束缚,据信该靶序列应该与PAM(原型间隔子邻近基序)相关;也就是说,与由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。

[0225] 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。

[0226] 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。

[0227] 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)。

[0228] 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。

[0229] 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美国临时申请61/736,527和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。

[0230] 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（Neurexin）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）；

[0231]

	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

[0232]

[0233] 表B：

[0234]

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））；

[0235]

	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、

[0236]

	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）。

[0237] 表C：

[0238]

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；

[0239]

	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；

[0240]

	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
急性期应答	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；

[0241]

	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；

[0242]

	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	CD44；EP300；LRP6；DVL3；CSNK1E；GJA1；SMO；
信号传导	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

[0243]

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
Fc ε RI 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；

[0244]

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；

[0245]

	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；

[0246]

	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；
中的一氧化氮信号传导	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

[0247]

心脏&β肾上腺素能	GNAS；GNAQ；PPP2R1A；GNB2L1；PPP2CA；PPP1CC；
信号传导	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	PLD1；GRN；GPAM；YWHAZ；SPHK1；SPHK2
代谢	
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	ERCC5；ERCC4；XPA；XPC；ERCC1
途径	
淀粉和蔗糖	UCHL1；HK2；GCK；GPI；HK1
代谢	
氨基糖代谢	NQO1；HK2；GCK；HK1
花生四烯酸	PRDX6；GRN；YWHAZ；CYP1B1
代谢	
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	PPP2R1A；PPP2CA；PPP1CC；PPP2R5C
信号传导	
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

[0248]

细胞色素 p450	GSTP1；CYP1B1
对异生物质的代谢	
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、丝氨酸和	CHKA
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；

[0249]

	颤蛋白（Reelin）；Dab1；unc-86（Pou4f1 或 Brn3a）；Numb；Reln

[0250] 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。

[0251] 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。

[0252] 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。

[0253] 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。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。

[0254] 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。

[0255] 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。

[0256] 成瘾相关的蛋白质的实例可以包括例如ABAT。

[0257] 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。

[0258] 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)。

[0259] 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)。

[0260] 与自闭症谱系障碍相关的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)。

[0261] 与黄斑变性相关的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)。

[0262] 与精神分裂症相关的蛋白的实例可以包括NRG1、ErbB4、CPLX1、TPH1、TPH2、NRXN1、GSK3A、BDNF、DISC1、GSK3B及其组合。

[0263] 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、Notch2、Notch 3或Notch 4。

[0264] 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)。

[0265] 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。

[0266] 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。

[0267] 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))。

[0268] 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]。

[0269] 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)。

[0270] 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)。

[0271] 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]。

[0272] 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。

[0273] 蛋白质治疗剂的长期给药可以引出针对该特殊蛋白质的不可接受的免疫反应。蛋白质药物的免疫原性可以归因于少数免疫显性辅助性T淋巴细胞(HTL)表位。降低包含在这些蛋白质内的这些HTL表位的MHC结合亲和力可以产生具有更低免疫原性的药物(腾格里S(Tangri S)等人“具有降低免疫原性的合理工程化的治疗蛋白”(“Rationally engineered therapeutic proteins with reduced immunogenicity”《免疫学杂志》(J Immunol.)2005年3月15日;174(6):3187-96。)在本发明中,可以具体地遵循首先由腾格里(Tangri)等人针对促红细胞生成素提出并且随后开发的方法来降低该CRISPR酶的免疫原性。因此,定向演化或合理设计可以用来降低在宿主物种(人类或其他物种)中的CRISPR酶(例如Cas9)的免疫原性。

[0274] 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,这样具有比先前的诱变剂更好的精确性,并且因此加速并改良植物育种计划。

[0275] 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。

[0276] 本申请包含以下项目:

[0277] 1.一种工程化的、非天然存在的CRISPR-Cas系统,其包括:

[0278] (a)包含Cas9切口酶的蛋白质或者编码所述蛋白质的多核苷酸,所述Cas9切口酶包含在催化结构域中的至少一个突变并且缺乏切割一条DNA链的能力,其中所述Cas9切口酶连接至至少两个核定位信号(NLS);

[0279] (b)CRISPR-Cas系统嵌合指导序列,其包含能够与真核细胞的核中与原型间隔子邻近基序(PAM)相邻的靶序列杂交并指导CRISPR-Cas复合物与所述靶序列序列特异性结合的指导序列,能够与tracr序列杂交的tracr配对序列,和长度为至少30个核苷酸的tracr序列。

[0280] 2.一种工程化的、非天然存在的CRISPR-Cas载体系统,其包括一个或多个载体,所述载体包含:

[0281] (a)可操作地连接至编码包含Cas9切口酶的蛋白质的核苷酸序列的第一调节元件,所述Cas9切口酶包含在催化结构域中的至少一个突变并且缺乏切割一条DNA链的能力,其中所述Cas9切口酶连接至至少两个核定位信号(NLS);

[0282] (b)可操作地连接至编码CRISPR-Cas系统嵌合指导序列的核苷酸序列的第二调节元件,所述CRISPR-Cas系统嵌合指导序列包含能够与真核细胞的核中与原型间隔子邻近基序(PAM)相邻的靶序列杂交并指导CRISPR-Cas复合物与所述靶序列序列特异性结合的指导序列,能够与tracr序列杂交的tracr配对序列,和长度为至少30个核苷酸的tracr序列;

[0283] 其中组分(a)和(b)位于所述系统的相同或不同载体上。

[0284] 3.根据项目2所述的系统,其中所述载体是病毒载体。

[0285] 4.根据项目3所述的系统,其中所述病毒载体是腺相关载体。

[0286] 5.根据项目1或2所述的系统,其中所述tracr序列的长度为至少40个核苷酸。

[0287] 6.根据项目1或2所述的系统,其中所述tracr序列的长度为至少50个核苷酸。

[0288] 7.根据项目1或2所述的系统,其中所述指导序列的长度为15-30个核苷酸。

[0289] 8.根据项目1或2所述的系统,其中所述指导序列的长度为至少20个核苷酸。

[0290] 9.根据项目1或2所述的系统,其中所述Cas9包含RuvC结构域中的至少一个突变。

[0291] 10.根据项目1或2所述的系统,其中所述Cas9包含与化脓链球菌Cas9的D10A,H840A,N854A或N863A相对应的突变。

[0292] 11.根据项目1或2所述的系统,其中所述Cas9是化脓链球菌Cas9,且PAM是NGG;或者其中所述Cas9是嗜热链球菌Cas9,且PAM是NAGAAW。

[0293] 12.根据项目1或2所述的系统,其中所述蛋白质包含连接至所述Cas9的至少一个异源蛋白结构域。

[0294] 13.根据项目12所述的系统,其中所述异源蛋白结构域具有以下活性中的一种或多种:甲基酶活性,脱甲基酶活性,转录激活活性,转录抑制活性,转录释放因子活性,组蛋白修饰活性,RNA切割活性和核酸结合活性。

[0295] 14.根据项目12所述的系统,其中所述异源蛋白结构域是GFP。

[0296] 15.根据项目1或2所述的系统,其中所述蛋白质包含连接至所述Cas9的至少一个SV40病毒大T抗原NLS(PKKKRKV)。

[0297] 16.根据项目1或2所述的系统,其中所述蛋白质包含至少一个N末端NLS和至少一个C末端NLS。

[0298] 17.根据项目1所述的系统,其中编码所述蛋白质的所述多核苷酸经密码子优化以在哺乳动物细胞或人细胞中表达。

[0299] 18.根据项目1所述的系统,其中编码所述蛋白质的所述多核苷酸是mRNA。

[0300] 19.根据项目1所述的系统,其中所述系统进一步包括模板多核苷酸,其中所述模板多核苷酸与所述靶序列的至少5个核苷酸重叠。

[0301] 20.根据项目1或2所述的系统在制造用于基因治疗的药物中的用途。

[0302] 21.根据项目1或2所述的系统用于离体或体外修饰真核细胞的用途。

[0303] 22.一种真核细胞,其包含根据项目1或2所述的系统。

[0304] 23.一种非人生物体,其包含项目22所述的真核细胞,其中所述非人生物体不是动物或植物品种。

[0305] 24.一种Cas9酶,其源自化脓链球菌。

[0306] 25.一种Cas9酶,其源自嗜热链球菌。

[0307] 26.一种试剂盒,其包括Cas9酶。

[0308] 27.一种融合蛋白,其包含RNAse III,mCherry和核定位信号(NLS)。

[0309] 28.一种组合物,其包含选自以下的一种或多种PCR引物:AAAACCACCCTTCTCTCTGGC、GGAGATTGGAGACACGGAGAG、CTGGAAAGCCAATGCCTGAC、和GGCAGCAAACTCCTTGTCCT。

[0310] 实例

[0311] 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。

[0312] 实例1:真核细胞的细胞核中的CRISPR复合物活性

[0313] 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30bp)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。

[0314] 为了改进CRISPR组分在哺乳动物细胞中的表达,可以将来自化脓链球菌(S.pyogenes)的SF370座位的两个基因进行密码子优化,即Cas9(SpCas9)和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)。

[0315] 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6中所示出的表达构

建体,其中功能性是如显示于图6B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。

[0316] 为了促进转录精确起始,选择基于RNA聚合酶III的U6启动子,以驱动tracrRNA的表达(图2C)。类似地,研发基于U6启动子的构建体,以表达由单个间隔子组成的前-crRNA阵列,该单个间隔子侧翼为两个同向重复(DR,还被术语“tracr配对序列”所涵盖;图2C)。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。

[0317] 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)途径而部分地配对,使用Surveyor测定检测靶EMX1座位处的潜在的切割活性(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%插入缺失,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4-8、10和11)。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。

[0318] 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9提供了哺乳动物细胞中的crRNA加工的另外的Northern印迹分析。图9A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。靶向人类EMX1座位原型间隔子1的30bp间隔子和同向重复序列显示于图9A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9B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用或没用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。

[0319] 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1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。

[0320] 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(包括一种指导序列)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)。

[0321] 可以使用退火的寡核苷酸将指导序列插入在BbsI位点之间。分别在DNA序列的上方和下方指示了有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。虽然对于每个座位而言,使用嵌合的构建体仅伴随三分之一的间隔子检测切割,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4和5)。

[0322] 图5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5A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5B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5C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。

[0323] 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。

[0324] 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3A中,包括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。

[0325] 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3C)。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3C)。在细菌测定中,所有的间隔子均促进有效的CRISPR干扰(图3C)。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。

[0326] 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)。

[0327] 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(Sapranausaks)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的共表达不产生插入缺失(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。

[0328] 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。

[0329] 实例2:CRISPR系统修饰和替代方案

[0330] 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。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。表明在每个靶位点的修饰效率低于这些RNA二级结构。产生结构的算法根据其假设预测的二级结构的可能性为每个碱基标颜色。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7中的。

[0331] 实例3:样品靶序列选择算法

[0332] 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。

[0333] 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。

[0334] 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TBA(博德参考号BI-2012/084 44790.11.2022)中。

[0335] 实例4:多种嵌合的crRNA-tracrRNA杂交体的评价

[0336] 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指导序列、tracr配对序列以及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。人类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10c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11a和11b的直方图中,分别对应于图10b和10c(“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、原型间隔子序列、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)。

[0337] 表D:

[0338]

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	-

[0339]

3	*EMX1*	GGAAGGGCCTGAGTCCGAGCAGAAGAAGAA	GGG	+
4	*PVALB*	GGTGGCGAGAGGGGCCGAGATTGGGTGTTC	AGG	+
5	*PVALB*	ATGCAGGAGGGTGGCGAGAGGGGCCGAGAT	TGG	+

[0340] 细胞培养和转染

[0341] 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。

[0342] 基因组修饰的SURVEYOR测定

[0343] 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(Bio-rad)成像。基于相对条带强度进行量化。

[0344] 表E:

[0345]

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

[0346] 独特的CRISPR靶位点的计算机鉴定

[0347] 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22中。

[0348] 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。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10b和10a)。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10c和10b)。

[0349] 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。

[0350] 图11和21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。

[0351] 以上实例中所述的序列如下(多核苷酸序列是5'至3'):

[0352] U6-短tracrRNA(化脓链球菌SF370):

[0353] GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCG**GGAACCATTCAAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGT**

GGCACCGAGTCGGTGCTTTTTTT(粗体=tracrRNA序列;下划线=终止序列)U6-长tracrRNA(化脓链球菌SF370):

[0354] GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGGTAGTATTAAGTATTGTTTTATGGCTGATAAATTTCTTTGAATTTCTCCTTGATTATTTGTTATAAAAGTTATAAAATAATCTTGTTGGAACCATTCAAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT

[0355] U6-DR-BbsI骨架-DR(化脓链球菌SF370):

[0356] GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTAT

ATATCTTGTGGAAAGGACGAAACACCGGGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACGGGTCTTCGAGAA
GACGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAAC

[0357] U6-嵌合RNA-BbsI骨架(化脓链球菌SF370)

[0358] GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGGA
ATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGC
AGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTAT
ATATCTTGTGGAAAGGACGAAACACCGGGTCTTCGAGAAGACCTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAG
GCTAGTCCG

[0359] NLS-SpCas9-EGFP:

[0360] MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAADKKYSIGLDIGTNSVGWAVITDEYKVPSK
KFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLV
EEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKL
FIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED
AKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL
VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIH
LGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFI
ERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYF
KKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDK
VMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEH
IANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVK
KMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVK
VITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK
ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKE
SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAK
GYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQH
KHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTST
KEVLDATLIHQSITGLYETRIDLSQLGGDAAAVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLK
FICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNR
IELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNH
YLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYK

[0361] SpCas9-EGFP-NLS:

[0362] MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT

EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKP
IREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDAAAVSKGEEL
FTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFF
KSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGI
KVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYK
KRPAATKKAGQAKKKK

[0363] NLS-SpCas9-EGFP-NLS：

[0364] MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAADKKYSIGLDIGTNSVGWAVITDEYKVPSK
KFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLV
EEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKL
FIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED
AKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL
VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIH
LGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFI
ERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYF
KKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDK
VMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEH
IANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVK
KMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVK
VITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK
ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKE
SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAK
GYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQH
KHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTST
KEVLDATLIHQSITGLYETRIDLSQLGGDAAAVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLK
FICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNR
IELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNH
YLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKKRPAATKKAGQAKKKK

[0365] NLS-SpCas9-NLS：

[0366] MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAADKKYSIGLDIGTNSVGWAVITDEYKVPSK
KFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLV
EEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKL
FIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED
AKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL
VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIH
LGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFI
ERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYF
KKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDK
VMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEH
IANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVK
KMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVK
VITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK
ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKE
SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAK
GYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQH
KHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTST
KEVLDATLIHQSITGLYETRIDLSQLGGDKRPAATKKAGQAKKKK

[0367] NLS-mCherry-SpRNA酶3：

[0368] MFLFLSLTSFLSSSRTLVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKV
TKGGPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLR
GTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLKDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDIT
SHNEDYTIVEQYERAEGRHSTGGMDELYKGSKQLEELLSTSFDIQFNDLTLLETAFTHTSYANEHRLLNVSHNERLE
FLGDAVLQLIISEYLFAKYPKKTEGDMSKLRSMIVREESLAGFSRFCSFDAYIKLGKGEEKSGGRRRDTILGDLFEA
FLGALLLDKGIDAVRRFLKQVMIPQVEKGNFERVKDYKTCLQEFLQTKGDVAIDYQVISEKGPAHAKQFEVSIVVNG
AVLSKGLGKSKKLAEQDAAKNALAQLSEV

[0369] SpRNA酶3-mCherry-NLS：

[0370] MKQLEELLSTSFDIQFNDLTLLETAFTHTSYANEHRLLNVSHNERLEFLGDAVLQLIISEYLFAKYPKK
TEGDMSKLRSMIVREESLAGFSRFCSFDAYIKLGKGEEKSGGRRRDTILGDLFEAFLGALLLDKGIDAVRRFLKQVM
IPQVEKGNFERVKDYKTCLQEFLQTKGDVAIDYQVISEKGPAHAKQFEVSIVVNGAVLSKGLGKSKKLAEQDAAKNA
LAQLSEVGSVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQ
FMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTM
GWEASSERMYPEDGALKGEIKQRLKLKDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQYERAE
GRHSTGGMDELYKKRPAATKKAGQAKKKK

[0371] NLS-SpCas9n-NLS(D10A切口酶突变是小写字母)：

[0372] MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAADKKYSIGLaIGTNSVGWAVITDEYKVPS
KKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL
VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDK

LFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAE
DAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKA
LVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQI
HLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSF
IERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDY
FKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDD
KVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE
HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEH
PVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVV
KKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREV
KVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIG
KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSK
ESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEA
KGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ
HKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTS
TKEVLDATLIHQSITGLYETRIDLSQLGGDKRPAATKKAGQAKKKK

[0373] hEMX1-HR模板-HindII-NheI：

[0374] GAATGCTGCCCTCAGACCCGCTTCCTCCCTGTCCTTGTCTGTCCAAGGAGAATGAGGTCTCACTGGTG
GATTTCGGACTACCCTGAGGAGCTGGCACCTGAGGGACAAGGCCCCCCACCTGCCCAGCTCCAGCCTCTGATGAGG
GGTGGGAGAGAGCTACATGAGGTTGCTAAGAAAGCCTCCCCTGAAGGAGACCACACAGTGTGTGAGGTTGGAGTCT
CTAGCAGCGGGTTCTGTGCCCCCAGGGATAGTCTGGCTGTCCAGGCACTGCTCTTGATATAAACACCACCTCCTAG
TTATGAAACCATGCCCATTCTGCCTCTCTGTATGGAAAAGAGCATGGGGCTGGCCCGTGGGGTGGTGTCCACTTTA
GGCCCTGTGGGAGATCATGGGAACCCACGCAGTGGGTCATAGGCTCTCTCATTTACTACTCACATCCACTCTGTGA
AGAAGCGATTATGATCTCTCCTCTAGAAACTCGTAGAGTCCCATGTCTGCCGGCTTCCAGAGCCTGCACTCCTCCA
CCTTGGCTTGGCTTTGCTGGGGCTAGAGGAGCTAGGATGCACAGCAGCTCTGTGACCCTTTGTTTGAGAGGAACAG
GAAAACCACCCTTCTCTCTGGCCCACTGTGTCCTCTTCCTGCCCTGCCATCCCCTTCTGTGAATGTTAGACCCATG
GGAGCAGCTGGTCAGAGGGGACCCCGGCCTGGGGCCCCTAACCCTATGTAGCCTCAGTCTTCCCATCAGGCTCTCA
GCTCAGCCTGAGTGTTGAGGCCCCAGTGGCTGCTCTGGGGGCCTCCTGAGTTTCTCATCTGTGCCCCTCCCTCCCT
GGCCCAGGTGAAGGTGTGGTTCCAGAACCGGAGGACAAAGTACAAACGGCAGAAGCTGGAGGAGGAAGGGCCTGAG
TCCGAGCAGAAGAAGAAGGGCTCCCATCACATCAACCGGTGGCGCATTGCCACGAAGCAGGCCAATGGGGAGGACA
TCGATGTCACCTCCAATGACaagcttgctagcGGTGGGCAACCACAAACCCACGAGGGCAGAGTGCTGCTTGCTGC
TGGCCAGGCCCCTGCGTGGGCCCAAGCTGGACTCTGGCCACTCCCTGGCCAGGCTTTGGGGAGGCCTGGAGTCATGG
CCCCACAGGGCTTGAAGCCCGGGGCCGCCATTGACAGAGGGACAAGCAATGGGCTGGCTGAGGCCTGGGACCACTTG
GCCTTCTCCTCGGAGAGCCTGCCTGCCTGGGCGGGCCCGCCCGCCACCGCAGCCTCCCAGCTGCTCTCCGTGTCTCC
AATCTCCCTTTTGTTTTGATGCATTTCTGTTTTAATTTATTTTCCAGGCACCACTGTAGTTTAGTGATCCCCAGTGT
CCCCCTTCCCTATGGGAATAATAAAAGTCTCTCTCTTAATGACACGGGCATCCAGCTCCAGCCCCAGAGCCTGGGGT
GGTAGATTCCGGCTCTGAGGGCCAGTGGGGGCTGGTAGAGCAAACGCGTTCAGGGCCTGGGAGCCTGGGGTGGGGTA
CTGGTGGAGGGGGTCAAGGGTAATTCATTAACTCCTCTCTTTTGTTGGGGGACCCTGGTCTCTACCTCCAGCTCCAC
AGCAGGAGAAACAGGCTAGACATAGGGAAGGGCCATCCTGTATCTTGAGGGAGGACAGGCCCAGGTCTTTCTTAACG

TATTGAGAGGTGGGAATCAGGCCCAGGTAGTTCAATGGGAGAGGGAGAGTGCTTCCCTCTGCCTAGAGACTCTGGTG
GCTTCTCCAGTTGAGGAGAAACCAGAGGAAAGGGGAGGATTGGGGTCTGGGGGAGGGAACACCATTCACAAAGGCTG
ACGGTTCCAGTCCGAAGTCGTGGGCCCACCAGGATGCTCACCTGTCCTTGGAGAACCGCTGGGCAGGTTGAGACTGC
AGAGACAGGGCTTAAGGCTGAGCCTGCAACCAGTCCCCAGTGACTCAGGGCCTCCTCAGCCCAAGAAAGAGCAACGT
GCCAGGGCCCGCTGAGCTCTTGTGTTCACCTG

[0375] NLS-StCsn1-NLS:

[0376] MKRPAATKKAGQAKKKKSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGR
RLARRKKHRRVRLNRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLDDASDDGNS
SVGDYAQIVKENSKQLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQI
TDEFINRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEFRAAKASYTAQEFNLLNDL
NNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLET
LDIEQMDRETLDKLAYVLTLNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIFGKGWHNFSVKLMMELIPE
LYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETN
EDDEKKAIQKIQKANKDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLINNS
NQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQALDSMDDAWSFRELKAFVRESKTLSNKKKEYLLTEED
ISKFDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYHHHAVDALIIA
ASSQLNLWKKQKNTLVSYSEDQLLDIETGELISDDEYKESVFKAPYQHFVDTLKSKEFEDSILFSYQVDSKFNRKIS
DATIYATRQAKVGKDKADETYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNKQINE
KGKEVPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLGNHIDITPKDSNNKVVLQSVSPWRADVYFNKTTGK
YEILGLKYADLQFEKGTGTYKISQEKYNDIKKKEGVDSDSEFKFTLYKNDLLLVKDTETKEQQLFRFLSRTMPKQKH
YVELKPYDKQKFEGGEALIKVLGNVANSGQCKKGLGKSNISIYKVRTDVLGNQHIIKNEGDKPKLDFKRPAATKKAG
QAKKKK

[0377] U6-St_tracrRNA(7-97):

[0378] GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGGA
ATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGC
AGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTAT
ATATCTTGTGGAAAGGACGAAACACCGTTACTTAAATCTTGCAGAAGCTACAAAGATAAGGCTTCATGCCGAAATCA
ACACCCTGTCATTTTATGGCAGGGTGTTTTCGTTATTTAA

[0379] U6-DR-间隔子-DR(化脓链球菌SF370)

[0380] gagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattggaattaatttgactgtaaacac
aaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttttaaaatggactatcat
atgcttaccgtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgaaacaccgggttttagagctatgctgtt
ttgaatggtcccaaaacNNNNNNNNNNNNNNNNNNNNNNNNNNNNgttttagagctatgctgt

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

[0381] 嵌合RNA,包含+48 tracr RNA(化脓链球菌SF370)

[0382] gagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattggaattaatttgactgtaaacac
aaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttttaaaatggactatcat
atgcttaccgtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgaaacaccNNNNNNNNNNN

NNNNNNNNNgttttagagctagaaatagcaagttaaaataaggctagtccgTTTTTTT (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)

[0383] 嵌合RNA,包含+54 tracr RNA(化脓链球菌SF370)

[0384] gagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattggaattaatttgactgtaaacac aaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttttaaaatggactatcat atgcttaccgtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgaaacaccNNNNNNNNNN

NNNNNNNNNgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaTTTTTTTT (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)

[0385] 嵌合RNA,包含+67 tracr RNA(化脓链球菌SF370)

[0386] gagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattggaattaatttgactgtaaacac aaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttttaaaatggactatcat atgcttaccgtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgaaacaccNNNNNNNNNN NNNNNNNNNgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtgTTTTTT

T (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)

[0387] 嵌合RNA,包含+85 tracr RNA(化脓链球菌SF370)

[0388] gagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattggaattaatttgactgtaaacac aaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttttaaaatggactatcat atgcttaccgtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgaaacaccNNNNNNNNNN NNNNNNNNNgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagt

cggtgcTTTTTTT (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)

[0389] CBh-NLS-SpCas9-NLS

[0390] CGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAA TAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAAC TGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCC GCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTA TTACCATGGTCGAGGTGAGCCCCACGTTCTGCTTCACTCTCCCCATCTCCCCCCCCTCCCCACCCCCAATTTTGTA TTTATTTATTTTTTAATTATTTTGTGCAGCGATGGGGGCGGGGGGGGGGGGGGGCGCGCGCCAGGCGGGGCGGGG CGGGGCGAGGGGCGGGGCGGGGCGAGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGCGCGCTCCGAAAGTTTC CTTTTATGGCGAGGCGGCGGCGGCGGCGGCCCTATAAAAAGCGAAGCGCGCGGCGGGCGGGAGTCGCTGCGACGCT GCCTTCGCCCCGTGCCCCGCTCCGCCGCCGCCTCGCGCCGCCCGCCCCGGCTCTGACTGACCGCGTTACTCCCACA GGTGAGCGGGCGGGACGGCCCTTCTCCTCCGGGCTGTAATTAGCTGAGCAAGAGGTAAGGGTTTAAGGGATGGTTG GTTGGTGGGGTATTAATGTTTAATTACCTGGAGCACCTGCCTGAAATCACTTTTTTTCAGGTTGGaccggtgccac cATGGACTATAAGGACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAGATGGCC CCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGACAAGAAGTACAGCATCGGCCTGGACATCG GCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAA CACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACC CGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCA ACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCA CGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTG AGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATCAAGT

TCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGT
GCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCC
AGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGCA
ACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCA
GCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTT
CTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCC
CCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCA
GCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGA
GCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGA
AGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCACCTGGG
AGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAG
ATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAA
AGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGA
GCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTC
ACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAGC
AGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTT
CAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATAC
CACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCG
TGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGA
CAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGCATC
CGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGCAGC
TGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCA
CGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAG
CTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGA
AGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAA
AGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATG
TACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGA
AGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGA
AGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGAC
AATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAA
CCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCT
GATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAA
GTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAA
AGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAG
CGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATT
ACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATA
AGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCA
GACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGG
GACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGG

GCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAA
TCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCC
CTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCC
TGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGA
GCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAG
AGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAG
AGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACAC
CACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGC
CTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGCGACTTTCTTTTTCTTAGCTTGACCAGCTTTCTTAGTA
GCAGCAGGACGCTTTAA(下划线=NLS-hSpCas9-NLS)

[0391] 针对嗜热链球菌LMD-9 CRISPR1 Cas9的实例嵌合RNA(具有NNAGAAW的PAM)

NNNNNNNNNNNNNNNNNNgtttttgtactctcaagatttaGAAAtaaatcttgcagaagctacaaagat aaggcttcatgccgaaatcaacaccctgtcattttatggcagggtgttttcgttatttaaTTTTTT (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)

[0392] 针对嗜热链球菌LMD-9 CRISPR1 Cas9的实例嵌合RNA(具有NNAGAAW的PAM)

NNNNNNNNNNNNNNNNNNgtttttgtactctcaGAAAtgcagaagctacaaagataaggcttcatgc cgaaatcaacaccctgtcattttatggcagggtgttttcgttatttaaTTTTTT (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)

[0393] 针对嗜热链球菌LMD-9 CRISPR1 Cas9的实例嵌合RNA(具有NNAGAAW的PAM)

NNNNNNNNNNNNNNNNNNgtttttgtactctcaGAAAtgcagaagctacaaagataaggcttcatgc cgaaatcaacaccctgtcattttatggcagggtgtTTTTTT (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)

[0394] 针对嗜热链球菌LMD-9 CRISPR1 Cas9的实例嵌合RNA(具有NNAGAAW的PAM)

NNNNNNNNNNNNNNNNNNgttattgtactctcaagatttaGAAAtaaatcttgcagaagctacaaaga taaggcttcatgccgaaatcaacaccctgtcattttatggcagggtgttttcgttatttaaTTTTTT (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)

[0395] 针对嗜热链球菌LMD-9 CRISPR1 Cas9的实例嵌合RNA(具有NNAGAAW的PAM)

NNNNNNNNNNNNNNNNNNgttattgtactctcaGAAAtgcagaagctacaaagataaggcttcatgc cgaaatcaacaccctgtcattttatggcagggtgttttcgttatttaaTTTTTT (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)

[0396] 针对嗜热链球菌LMD-9 CRISPR1 Cas9的实例嵌合RNA(具有NNAGAAW的PAM)

NNNNNNNNNNNNNNNNNNgttattgtactctcaGAAAtgcagaagctacaaagataaggcttcatgc cgaaatcaacaccctgtcattttatggcagggtgtTTTTTT (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)

[0397] 针对嗜热链球菌LMD-9 CRISPR1 Cas9的实例嵌合RNA(具有NNAGAAW的PAM)

NNNNNNNNNNNNNNNNNNgttattgtactctcaagatttaGAAAtaaatcttgcagaagctacaatgat aaggcttcatgccgaaatcaacaccctgtcattttatggcagggtgttttcgttatttaaTTTTTT (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)

[0398] 针对嗜热链球菌LMD-9 CRISPR1 Cas9的实例嵌合RNA(具有NNAGAAW的PAM)

NNNNNNNNNNNNNNNNNNgttattgtactctcaGAAAtgcagaagctacaatgataaggcttcatgc cgaaatcaacaccctgtcattttatggcagggtgttttcgttatttaaTTTTTT (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)

[0399] 针对嗜热链球菌LMD-9 CRISPR1 Cas9的实例嵌合RNA(具有NNAGAAW的PAM)

NNNNNNNNNNNNNNNNNNgttattgtactctcaGAAAtgcagaagctacaatgataaggcttcatgc cgaaatcaacaccctgtcattttatggcagggtgtTTTTTT (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)

[0400] 针对嗜热链球菌LMD-9 CRISPR3 Cas9的实例嵌合RNA(具有NGGNG的PAM)

NNNNNNNNNNNNNNNNNNgttttagagctgtgGAAAcacagcgagttaaaataaggcttagtccgta ctcaacttgaaaaggtggcaccgattcggtgtTTTTTT (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)

[0401] 来自嗜热链球菌LMD-9 CRISPR3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)

[0402] ATGAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGACCAAGCCCTACAGCATC
GGCCTGGACATCGGCACCAATAGCGTGGGCTGGGCCGTGACCACCGACAACTACAAGGTGCCCAGCAAGAAAATGAA
GGTGCTGGGCAACACCTCCAAGAAGTACATCAAGAAAAACCTGCTGGGCGTGCTGCTGTTCGACAGCGGCATTACAG
CCGAGGGCAGACGGCTGAAGAGAACCGCCAGACGGCGGTACACCCGGCGGAGAAACAGAATCCTGTATCTGCAAGAG
ATCTTCAGCACCGAGATGGCTACCCTGGACGACGCCTTCTTCCAGCGGCTGGACGACAGCTTCCTGGTGCCCGACGA
CAAGCGGGACAGCAAGTACCCCATCTTCGGCAACCTGGTGGAAGAGAAGGCCTACCACGACGAGTTCCCCACCATCT
ACCACCTGAGAAAGTACCTGGCCGACAGCACCAAGAAGGCCGACCTGAGACTGGTGTATCTGGCCCTGGCCCACATG
ATCAAGTACCGGGGCCACTTCCTGATCGAGGGCGAGTTCAACAGCAAGAACAACGACATCCAGAAGAACTTCCAGGA
CTTCCTGGACACCTACAACGCCATCTTCGAGAGCGACCTGTCCCTGGAAAACAGCAAGCAGCTGGAAGAGATCGTGA
AGGACAAGATCAGCAAGCTGGAAAAGAAGGACCGCATCCTGAAGCTGTTCCCCGGCGAGAAGAACAGCGGAATCTTC
AGCGAGTTTCTGAAGCTGATCGTGGGCAACCAGGCCGACTTCAGAAAGTGCTTCAACCTGGACGAGAAAGCCAGCCT
GCACTTCAGCAAAGAGAGCTACGACGAGGACCTGGAAACCCTGCTGGGATATATCGGCGACGACTACAGCGACGTGT
TCCTGAAGGCCAAGAAGCTGTACGACGCTATCCTGCTGAGCGGCTTCCTGACCGTGACCGACAACGAGACAGAGGCC
CCACTGAGCAGCGCCATGATTAAGCGGTACAACGAGCACAAAGAGGATCTGGCTCTGCTGAAAGAGTACATCCGGAA
CATCAGCCTGAAAACCTACAATGAGGTGTTCAAGGACGACACCAAGAACGGCTACGCCGGCTACATCGACGGCAAGA
CCAACCAGGAAGATTTCTATGTGTACCTGAAGAAGCTGCTGGCCGAGTTCGAGGGGGCCGACTACTTTCTGGAAAAA
ATCGACCGCGAGGATTTCCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCTACCAGATCCATCTGCAGGA
AATGCGGGCCATCCTGGACAAGCAGGCCAAGTTCTACCCATTCCTGGCCAAGAACAAAGAGCGGATCGAGAAGATCC
TGACCTTCCGCATCCCTTACTACGTGGGCCCCCTGGCCAGAGGCAACAGCGATTTTGCCTGGTCCATCCGGAAGCGC
AATGAGAAGATCACCCCCTGGAACTTCGAGGACGTGATCGACAAAGAGTCCAGCGCCGAGGCCTTCATCAACCGGAT
GACCAGCTTCGACCTGTACCTGCCCGAGGAAAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGACATTCAATGTGT
ATAACGAGCTGACCAAAGTGCGGTTTATCGCCGAGTCTATGCGGGACTACCAGTTCCTGGACTCCAAGCAGAAAAAG
GACATCGTGCGGCTGTACTTCAAGGACAAGCGGAAAGTGACCGATAAGGACATCATCGAGTACCTGCACGCCATCTA

CGGCTACGATGGCATCGAGCTGAAGGGCATCGAGAAGCAGTTCAACTCCAGCCTGAGCACATACCACGACCTGCTGA
ACATTATCAACGACAAAGAATTTCTGGACGACTCCAGCAACGAGGCCATCATCGAAGAGATCATCCACACCCTGACC
ATCTTTGAGGACCGCGAGATGATCAAGCAGCGGCTGAGCAAGTTCGAGAACATCTTCGACAAGAGCGTGCTGAAAAA
GCTGAGCAGACGGCACTACACCGGCTGGGGCAAGCTGAGCGCCAAGCTGATCAACGGCATCCGGGACGAGAAGTCCG
GCAACACAATCCTGGACTACCTGATCGACGACGGCATCAGCAACCGGAACTTCATGCAGCTGATCCACGACGACGCC
CTGAGCTTCAAGAAGAAGATCCAGAAGGCCCAGATCATCGGGGACGAGGACAAGGGCAACATCAAAGAAGTCGTGAA
GTCCCTGCCCGGCAGCCCCGCCATCAAGAAGGGAATCCTGCAGAGCATCAAGATCGTGGACGAGCTCGTGAAAGTGA
TGGGCGGCAGAAAGCCCGAGAGCATCGTGGTGGAAATGGCTAGAGAGAACCAGTACACCAATCAGGGCAAGAGCAAC
AGCCAGCAGAGACTGAAGAGACTGGAAAAGTCCCTGAAAGAGCTGGGCAGCAAGATTCTGAAAGAGAATATCCCTGC
CAAGCTGTCCAAGATCGACAACAACGCCCTGCAGAACGACCGGCTGTACCTGTACTACCTGCAGAATGGCAAGGACA
TGTATACAGGCGACGACCTGGATATCGACCGCCTGAGCAACTACGACATCGACCATATTATCCCCCAGGCCTTCCTG
AAAGACAACAGCATTGACAACAAAGTGCTGGTGTCCTCCGCCAGCAACCGCGGCAAGTCCGATGATGTGCCCAGCCT
GGAAGTCGTGAAAAAGAGAAAGACCTTCTGGTATCAGCTGCTGAAAAGCAAGCTGATTAGCCAGAGGAAGTTCGACA
ACCTGACCAAGGCCGAGAGAGGCGGCCTGAGCCCTGAAGATAAGGCCGGCTTCATCCAGAGACAGCTGGTGGAAACC
CGGCAGATCACCAAGCACGTGGCCAGACTGCTGGATGAGAAGTTTAACAACAAGAAGGACGAGAACAACCGGGCCGT
GCGGACCGTGAAGATCATCACCCTGAAGTCCACCCTGGTGTCCCAGTTCCGGAAGGACTTCGAGCTGTATAAAGTGC
GCGAGATCAATGACTTTCACCACGCCCACGACGCCTACCTGAATGCCGTGGTGGCTTCCGCCCTGCTGAAGAAGTAC
CCTAAGCTGGAACCCGAGTTCGTGTACGGCGACTACCCCAAGTACAACTCCTTCAGAGAGCGGAAGTCCGCCACCGA
GAAGGTGTACTTCTACTCCAACATCATGAATATCTTTAAGAAGTCCATCTCCCTGGCCGATGGCAGAGTGATCGAGC
GGCCCCTGATCGAAGTGAACGAAGAGACAGGCGAGAGCGTGTGGAACAAAGAAAGCGACCTGGCCACCGTGCGGCGG
GTGCTGAGTTATCCTCAAGTGAATGTCGTGAAGAAGGTGGAAGAACAGAACCACGGCCTGGATCGGGGCAAGCCCAA
GGGCCTGTTCAACGCCAACCTGTCCAGCAAGCCTAAGCCCAACTCCAACGAGAATCTCGTGGGGGCCAAAGAGTACC
TGGACCCTAAGAAGTACGGCGGATACGCCGGCATCTCCAATAGCTTCACCGTGCTCGTGAAGGGCACAATCGAGAAG
GGCGCTAAGAAAAAGATCACAAACGTGCTGGAATTTCAGGGGATCTCTATCCTGGACCGGATCAACTACCGGAAGGA
TAAGCTGAACTTTCTGCTGGAAAAAGGCTACAAGGACATTGAGCTGATTATCGAGCTGCCTAAGTACTCCCTGTTCG
AACTGAGCGACGGCTCCAGACGGATGCTGGCCTCCATCCTGTCCACCAACAACAAGCGGGGCGAGATCCACAAGGGA
AACCAGATCTTCCTGAGCCAGAAATTTGTGAAACTGCTGTACCACGCCAAGCGGATCTCCAACACCATCAATGAGAA
CCACCGGAAATACGTGGAAAACCACAAGAAAGAGTTTGAGGAACTGTTCTACTACATCCTGGAGTTCAACGAGAACT
ATGTGGGAGCCAAGAAGAACGGCAAACTGCTGAACTCCGCCTTCCAGAGCTGGCAGAACCACAGCATCGACGAGCTG
TGCAGCTCCTTCATCGGCCCTACCGGCAGCGAGCGGAAGGGACTGTTTGAGCTGACCTCCAGAGGCTCTGCCGCCGA
CTTTGAGTTCCTGGGAGTGAAGATCCCCCGGTACAGAGACTACACCCCCTCTAGTCTGCTGAAGGACGCCACCCTGA
TCCACCAGAGCGTGACCGGCCTGTACGAAACCCGGATCGACCTGGCTAAGCTGGGCGAGGGAAAGCGTCCTGCTGCT
ACTAAGAAAGCTGGTCAAGCTAAGAAAAAGAAATAA

[0403] 实例5:对化脓链球菌Cas9(称为SpCas9)的指导RNA的优化。

[0404] 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。

[0405] 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。优化的tracrRNA 1(突变加下划线):

[0406] GGAACCATTCAtAACAGCATAGCAAGTTAtAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTT

[0407] 优化的同向重复1(突变加下划线):

[0408] GTTaTAGAGCTATGCTGTTaTGAATGGTCCCAAAAC

[0409] 优化的tracrRNA 2(突变加下划线):

[0410] GGAACCATTCAAtACAGCATAGCAAGTTAAtATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTT

[0411] 优化的同向重复2(突变加下划线):

[0412] GTaTTAGAGCTATGCTGTaTTGAATGGTCCCAAAAC

[0413] 为了真核细胞中的最佳活性,申请人还优化了嵌合指导RNA。

[0414] 原始的指导RNA:

[0415] NNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT

[0416] 优化的嵌合指导RNA序列1:

[0417] NNNNNNNNNNNNNNNNNNNNGTATTAGAGCTAGAAATAGCAAGTTAATATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT

[0418] 优化的嵌合指导RNA序列2:

[0419] NNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTATGCTGTTTTGGAAACAAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT

[0420] 优化的嵌合指导RNA序列3:

[0421] NNNNNNNNNNNNNNNNNNNNGTATTAGAGCTATGCTGTATTGGAAACAATACAGCATAGCAAGTTAATATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT

[0422] 申请人揭示,优化的嵌合指导RNA更好地起作用,如在图9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”

[0423] 实例6:嗜热链球菌LMD-9CRISPR1 Cas9(称为St1Cas9)的优化。

[0424] 申请人设计如图12中所示的指导嵌合RNA。

[0425] 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。

[0426] 实例7:Cas9系统用于体内应用的改进

[0427] 申请人执行了具有小分子量的Cas9的宏基因组检索。大多数Cas9同系物相当大。例如,SpCas9的长度为1368aa左右,这对于容易包装进供递送的病毒载体中而言太大。一些序列可能已经被错误注释并且因此每个长度的准确频率不一定是精确的。尽管如此,但是它提供了对Cas9蛋白分布的模糊认识并且表明存在较短的Cas9同系物。

[0428] 通过计算分析,申请人发现在细菌菌株弯曲杆菌属中,存在两个具有少于1000个氨基酸的Cas9蛋白。下文呈现了来自空肠弯曲杆菌的一种Cas9的序列。以此长度,CjCas9可以被容易地包装进AAV、慢病毒、腺病毒以及其他用于稳健地递送进原代细胞和递送进动物模型体内的病毒载体中。

[0429] >空肠弯曲杆菌Cas9(CjCas9)

[0430] MARILAFDIGISSIGWAFSENDELKDCGVRIFTKVENPKTGESLALPRRLARSARKRLARRKARLNHL KHLIANEFKLNYEDYQSFDESLAKAYKGSLISPYELRFRALNELLSKQDFARVILHIAKRRGYDDIKNSDDKEKGA ILKAIKQNEEKLANYQSVGEYLYKEYFQKFKENSKEFTNVRNKKESYERCIAQSFLKDELKLIFKKQREFGFSFSK KFEEEVLSVAFYKRALKDFSHLVGNCSFFTDEKRAPKNSPLAFMFVALTRIINLLNNLKNTEGILYTKDDLNALLN EVLKNGTLTYKQTKKLLGLSDDYEFKGEKGTYFIEFKKYKEFIKALGEHNLSQDDLNEIAKDITLIKDEIKLKKAL AKYDLNQNQIDSLSKLEFKDHLNISFKALKLVTPLMLEGKKYDEACNELNLKVAINEDKKDFLPAFNETYYKDEVT NPVVLRAIKEYRKVLNALLKKYGKVHKINIELAREVGKNHSQRAKIEKEQNENYKAKKDAELECEKLGLKINSKNI LKLRLFKEQKEFCAYSGEKIKISDLQDEKMLEIDHIYPYSRSFDDSYMNKVLVFTKQNQEKLNQTPFEAFGNDSAK WQKIEVLAKNLPTKKQKRILDKNYKDKEQKNFKDRNLNDTRYIARLVLNYTKDYLDFLPLSDDENTKLNDTQKGSK VHVEAKSGMLTSALRHTWGFSAKDRNNHLHHAIDAVIIAYANNSIVKAFSDFKKEQESNSAELYAKKISELDYKNK RKFFEPFSGFRQKVLDKIDEIFVSKPERKKPSGALHEETFRKEEEFYQSYGGKEGVLKALELGKIRKVNGKIVKNG DMFRVDIFKHKKTNKFYAVPIYTMDFALKVLPNKAVARSKKGEIKDWILMDENYEFCFSLYKDSLILIQTKDMQEP EFVYYNAFTSSTVSLIVSKHDNKFETLSKNQKILFKNANEKEVIAKSIGIQNLKVFEKYIVSALGEVTKAEFRQRE DFKK。

[0431] 此CjCas9的假定的tracrRNA元件是:

[0432] TATAATCTCATAAGAAATTTAAAAAGGGACTAAAATAAAGAGTTTGCGGGACTCTGCGGGGTTACAATC CCCTAAAACCGCTTTTAAAATT

[0433] 同向重复序列是:

[0434] ATTTTACCATAAAGAAATTTAAAAAGGGACTAAAAC

[0435] tracrRNA和同向重复的共折叠结构提供于图6中。

[0436] CjCas9的嵌合的指导RNA的一个实例是:

[0437] NNNNNNNNNNNNNNNNNNNGUUUUAGUCCCGAAAGGGACUAAAAUAAAGAGUUUGCGGGACUCUGCGG GGUUACAAUCCCCUAAAACCGCUUUU

[0438] 申请人还已经使用体外方法优化了Cas9指导RNA。图18显示了来自St1Cas9嵌合指导RNA体外优化的数据。

[0439] 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。

[0440] 实例8:Sa sgRNA优化

[0441] 申请人针对具有最高切割效率的优化的截短构造设计了SaCas9的五种sgRNA变体。另外,在sgRNA旁边测试了天然同向重复:tracr双链体系统。将具有指定长度的指导物用SaCas9共转染,并测试其在HEK 293FT细胞中的活性。将总共100ng sgRNA U6-PCR扩增子(或50ng的同向重复和50ng的tracrRNA)以及400ng的SaCas9质粒共转染进200,000个Hepa1-6小鼠肝细胞中,并且在转染后72小时收获DNA,用于进行SURVEYOR分析。结果显示于图23中。

[0442] 参考文献:

[0443] 1.Urnov,F.D.,Rebar,E.J.,Holmes,M.C.,Zhang,H.S.&Gregory,P.D.Genome

editing with engineered zinc finger nucleases.《遗传学自然评论》(Nat Rev Genet.)11,636-646(2010)。

[0444] 2.博格丹奥韦(Bogdanove),A.J.&沃伊塔斯(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)。

[0445] 3.斯托达德(Stoddard),B.L.复位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)。

[0446] 4.贝(Bae),T.&施内温德(Schneewind),O.用可诱导的逆筛选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)。

[0447] 5.孙(Sun)g,C.K.,Li,H.,克拉维瑞斯(Claverys),J.P.&莫里森(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)。

[0448] 6.沙兰(Sharan),S.K.,托马森(Thomason),L.C.,库兹涅佐夫(Kuznetsov),S.G.&考特(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)。

[0449] 7.季聂克(Jinek),M.等人在自适应细菌免疫中的可编程双重-RNA-引导DNA核酸内切酶(A programmabledual-RNA-guided DNA endonuclease in adaptive bacterial immunity).《科学》(Science)337,816-821(2012)。

[0450] 8.德沃(Deveau),H.,加尔诺(Garneau),J.E.&莫万诺(Moineau),S.CRISPR-Cas系统及其在噬菌体-细菌相互作用中的角色(CRISPR/Cas system and its role in phage-bacteria interactions).微生物学年鉴(Annu.Rev.Microbiol.)64,475-493(2010)。

[0451] 9.霍瓦特(Horvath),P.&巴尔朗勾(Barrangou),R.CRISPR-Cas,细菌和古细菌的免疫系统(CRISPR/Cas,the immune system of bacteria and archaea).《科学》(Science)327,167-170(2010)。

[0452] 10.特恩斯(Terns),M.P.&特恩斯(Terns),R.M.基于CRISPR的适应性免疫系统(CRISPR-based adaptive immune systems).微生物学现行观点(Curr.Opin.Microbiol.)14,321-327(2011)。

[0453] 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的适应性以及可继承免疫(CRISPR-based adaptive and heritable immunity in prokaryotes).《生物化学科学趋势》(Trends.Biochem.Sci.)34,401-407(2009)。

[0454] 12.布龙斯,S.J.(Brouns,S.J.)等人在原核生物中小CRISPR RNA引导抗病毒防御(Small CRISPR RNAs guide antiviral defense in prokaryotes).《科学》(Science)321,960-964(2008)。

[0455] 13.卡特,J.(Carte,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引导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).《基因与发育》(Genes Dev.)22,3489-3496(2008)。

[0456] 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RNA酶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)。

[0457] 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.)成熟成簇的、整齐间隔的、短回文重复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).《美国科学院院刊》(Proc.Natl.Acad.Sci.U.S.A.)108,21218-21222(2011)。

[0458] 16.华维茨,R.E.(Haurwitz,R.E.),吉耐克,M.(Jinek,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)。

[0459] 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)。

[0460] 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).《美国科学院院刊》(Proc.Natl.Acad.Sci.U.S.A.)(2012)。

[0461] 19.马卡洛娃,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).《生物定向》(Biol.Direct.)6,38(2011)。

[0462] 20.巴尔朗勾,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)。

[0463] 21.布龙斯,S.J.(Brouns,S.J.)《分子生物学》(Molecular biology).免疫的瑞士军刀(A Swiss army knife of immunity).《科学》(Science)337,808-809(2012)。

[0464] 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)。

[0465] 23.必考尔德,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)。

[0466] 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)。

[0467] 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是通过种子序列给出的(Interference by clustered regularly interspaced short palindromic repeat(CRISPR)RNA is governed by a seed sequence).《美国科学院院刊》(Proc.Natl.Acad.Sci.U.S.A.)(2011)。

[0468] 26.威登赫夫特,B.(Wiedenheft,B.)等人.来自细菌性免疫系统的RNA引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).《美国科学院院刊》(Proc.Natl.Acad.Sci.U.S.A.)(2011)。

[0469] 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 asurface protein).《细菌学杂志》(J.Bacteriol.)182,5919-5921(2000)。

[0470] 28.马拉菲尼,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被膜的分选酶和技术(Sortases and the art of anchoring proteins to the envelopes of gram-positive bacteria).《微生物学与分子生物学综述》(Microbiol.Mol.Biol.Rev.)70,192-221(2006)。

[0471] 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 mechanismin vivo).《基因与发育》(Genes Dev).13,2889-2903(1999)。

[0472] 30.霍萨卡,T.(Hosaka,T.)等人核糖体蛋白S12中的新颖突变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)。

[0473] 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).《美国科学院院刊》(Proc.Natl.Acad.Sci.U.S.A.)100,15748-15753(2003)。

[0474] 32.埃德加,R.(Edgar,R.)&奇姆龙,U.(Qimron,U.)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)。

[0475] 33.马拉菲尼,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)。

[0476] 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邻近基序(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)。

[0477] 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前间隔序列激发时,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)。

[0478] 36.王,H.H.(Wang,H.H.)等人通过共选择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)。

[0479] 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》(Science)印刷中(2013)。

[0480] 38.玛丽,P.(Mali,P.)等人通过Cas9进行的RNA引导的人类基因组工程(RNA-Guided Human Genome Engineering via Cas9).《科学》(Science)印刷中(2013)。

[0481] 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)。

[0482] 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).《美国科学院院刊》(Proc.Natl.Acad.Sci.U.S.A.)92,11140-11144(1995)。

[0483] 41.堀之内,S.(Horinouchi,S.)&维斯布卢姆,B.(Weisblum,B.)指定可诱导氯霉素抗性的一种质粒pC194的核苷酸序列以及功能性图(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)。

[0484] 42.荷顿(Horton),R.M.DNA的体外重组和诱变:基因拼接一起专用的基因.分子生物学方法(Methods Mol.Biol.)15,251-261(1993)。

[0485] 43.波德比尔斯基(Podbielski),A.,斯佩莱博格(Spellerberg),B.,沃伊施尼克(Woischnik),M.,波尔(Pohl),B.&鲁提肯(Lutticken),R.用于组A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)。

[0486] 44.胡斯曼(Husmann),L.K.,斯科特(Scott),J.R.,林达尔(Lindahl),G.&斯坦伯格(Stenberg),L.Arp蛋白即M蛋白家族的一员的表达不足够抑制化脓链球菌的噬菌体细胞

增多(Expression of the Arp protein,a member of the M protein family,is not sufficient to inhibit phagocytosis of Streptococcus pyogenes).传染以及免疫(Infection and immunity)63,345-348(1995)。

[0487] 45.吉布森(Gibson),D.G.等人几百千碱基的DNA分子的酶拼接(Enzymatic assembly of DNA molecules up to several hundred kilobases).自然:方法6(Nat Methods 6),343-345(2009)。

[0488] 46.腾格里S(Tangri S)等人,“具有减少的免疫原性的合理工程化的治疗性蛋白质(Rationally engineered therapeutic proteins with reduced immunogenicity)”,《免疫学杂志》(J Immunol.)2005年5月15日;174(6):3187-96。

[0489] ****

[0490] 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。

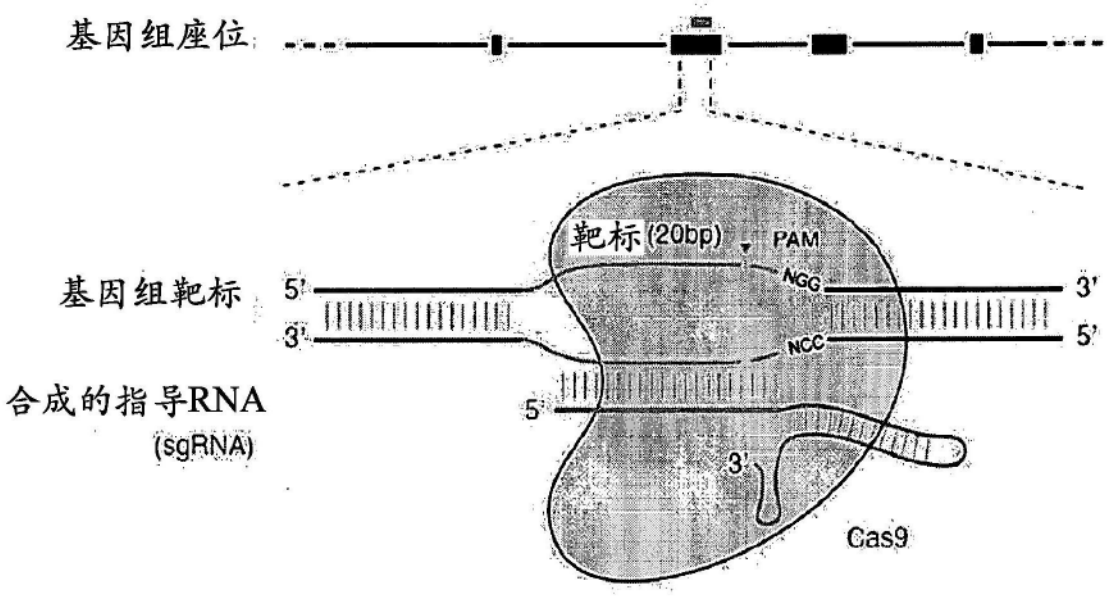


图1

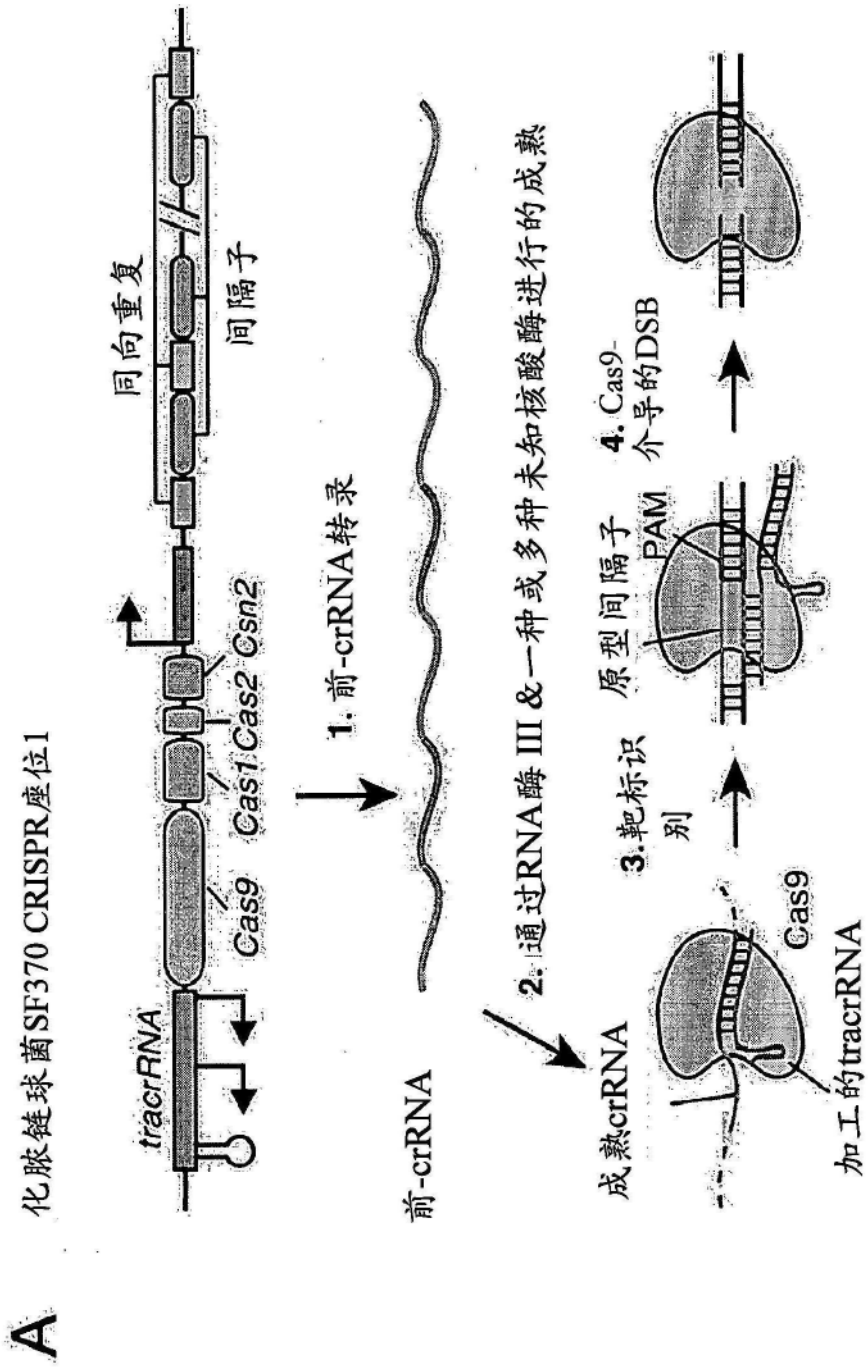

A
化脓链球菌SF370 CRISPR座位1
同向重复
间隔子
tracrRNA
Cas9
Cas1
Cas2
Csn2
1. 前-crRNA转录
前-crRNA
2. 通过RNA酶 III &一种或多种未知核酸酶进行的成熟
成熟crRNA
3.靶标识别
原型间隔子
PAM
4. Cas9介导的DSB
Cas9
加工的tracrRNA

图2A

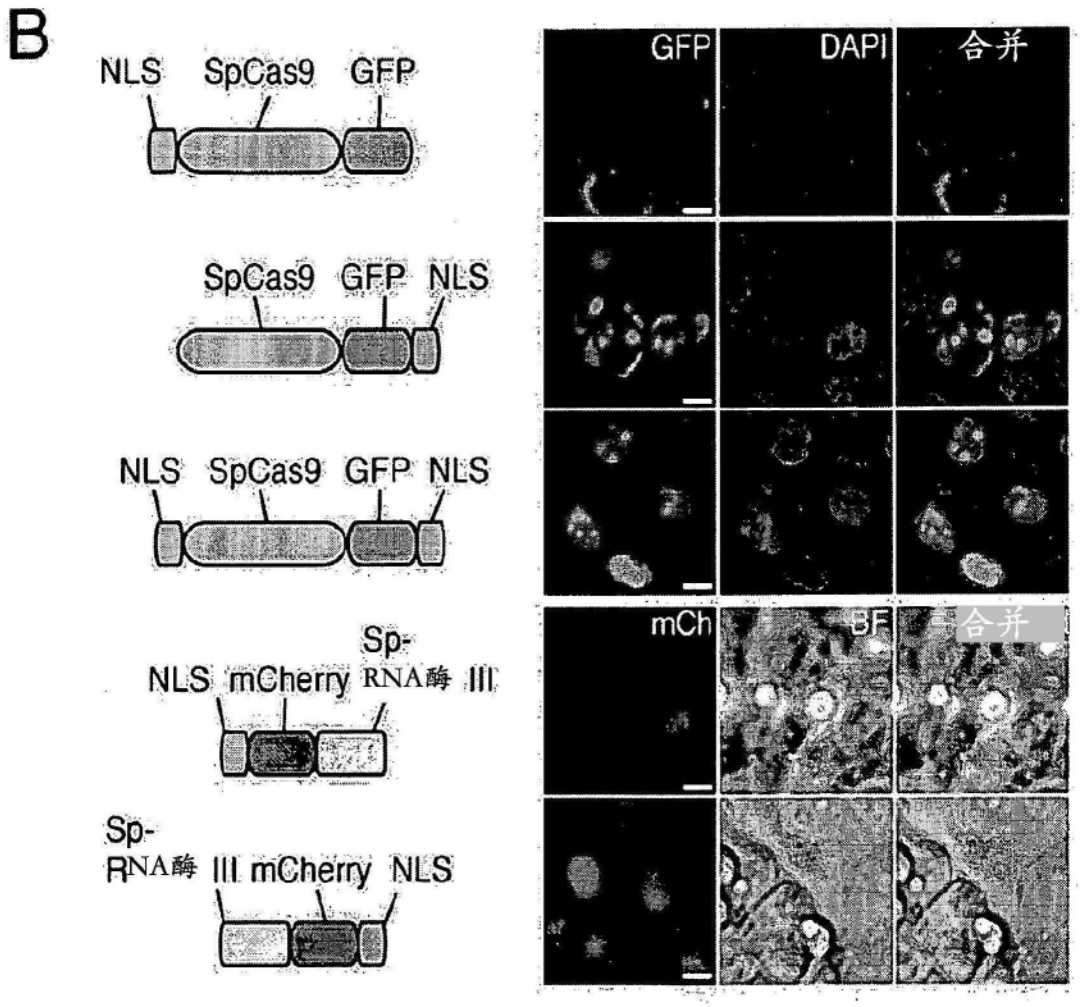

B
NLS SpCas9 GFP
SpCas9 GFP NLS
NLS SpCas9 GFP NLS
NLS mCherry Sp-RNA酶 III
Sp-RNA酶 III mCherry NLS
GFP
DAPI
合并
mCh
BF
合并

图2B

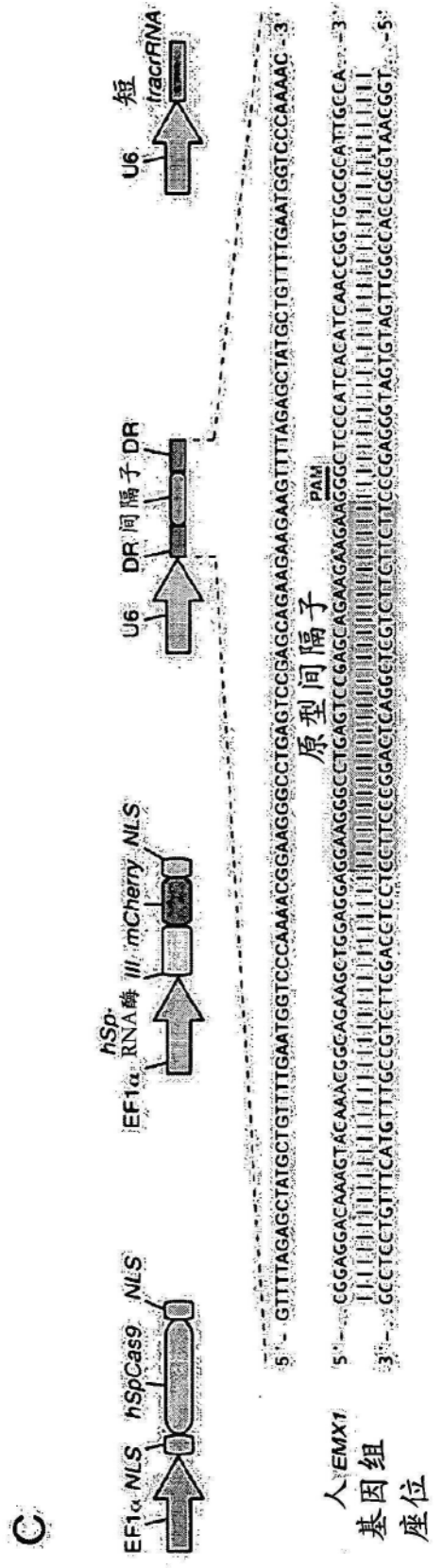

C
EF1α
NLS
hSpCas9
NLS
EF1α
hSp.
RNA酶
III
mCherry
NLS
U6
DR
间隔子
DR
U6
短
tracrRNA
人EMX1
基因组
座位
原型间隔子
PAM

图2C

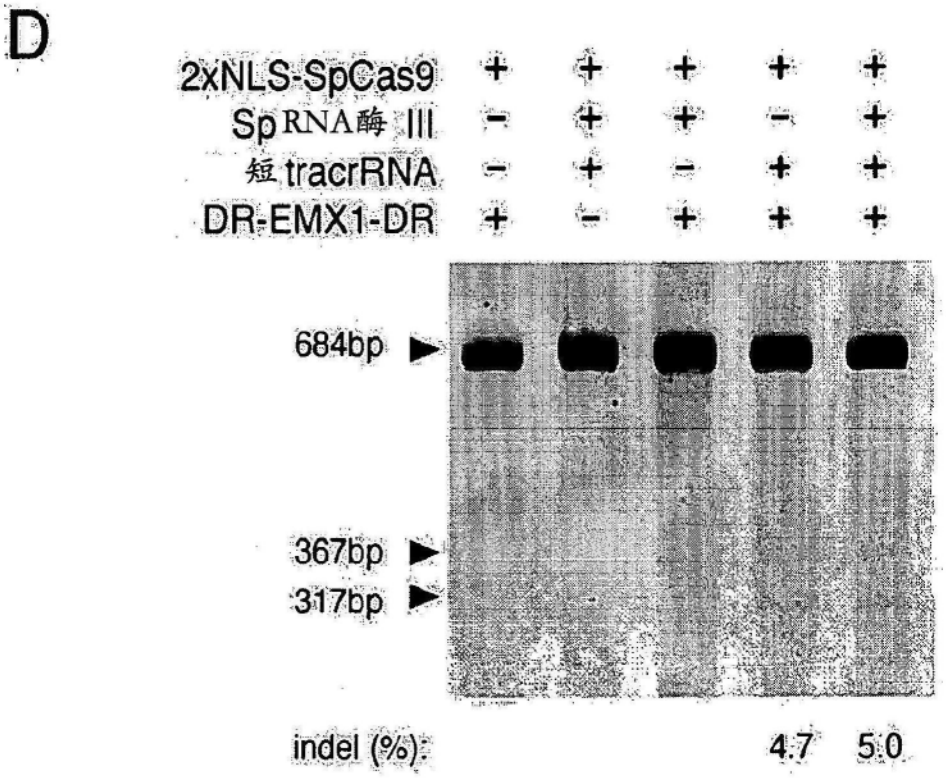

D
2xNLS-SpCas9 + + + + +
SpRNA酶 III - + + - +
短tracrRNA - + - + +
DR-EMX1-DR + - + + +
684bp
367bp
317bp
indel (%): 4.7 5.0

图2D

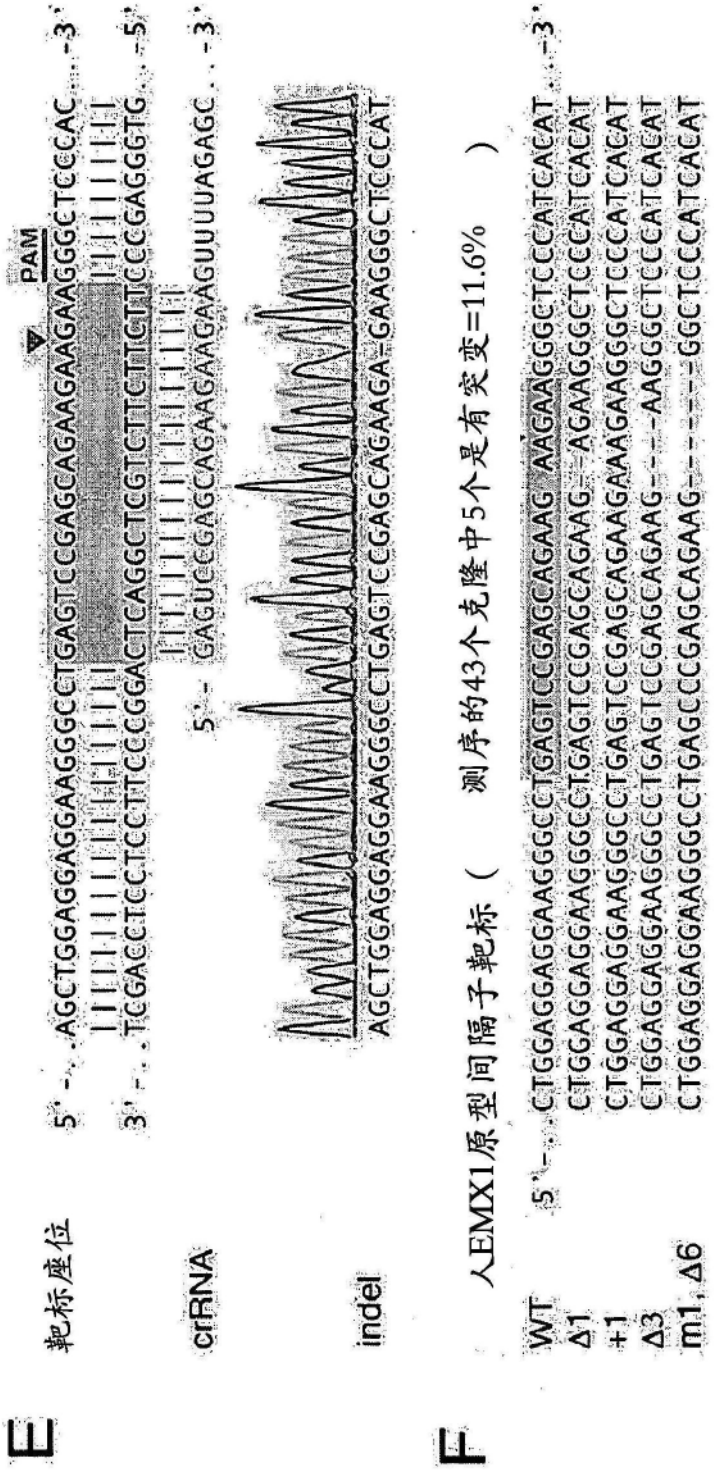

E
靶标座位
PAM
5' - . . AGCTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGGGCTCCCAC . . - 3'
3' - . . TCGACCTCCTCCTTCCCGGACTCAGGCTCGTCTTCTTCTTCCCGAGGGTG . . - 5'
crRNA
5' - GAGUCCGAGCAGAAGAAGAAGUUUUAGAGC . . - 3'
indel
AGCTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGA-GAAGGGCTCCCAT
F
人EMX1原型间隔子靶标 (测序的43个克隆中5个是有突变=11.6%)
WT 5' - . . CTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAG AAGAAGGGCTCCCATCACAT . . - 3'
Δ1 CTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAG-AGAAGGGCTCCCATCACAT
+1 CTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGGGCTCCCATCACAT
Δ3 CTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAG- - - -AAGGGCTCCCATCACAT
m1, Δ6 CTGGAGGAGGAAGGGCCTGAGCCCGAGCAGAAG- - - - - -GGCTCCCATCACAT

图2E-F

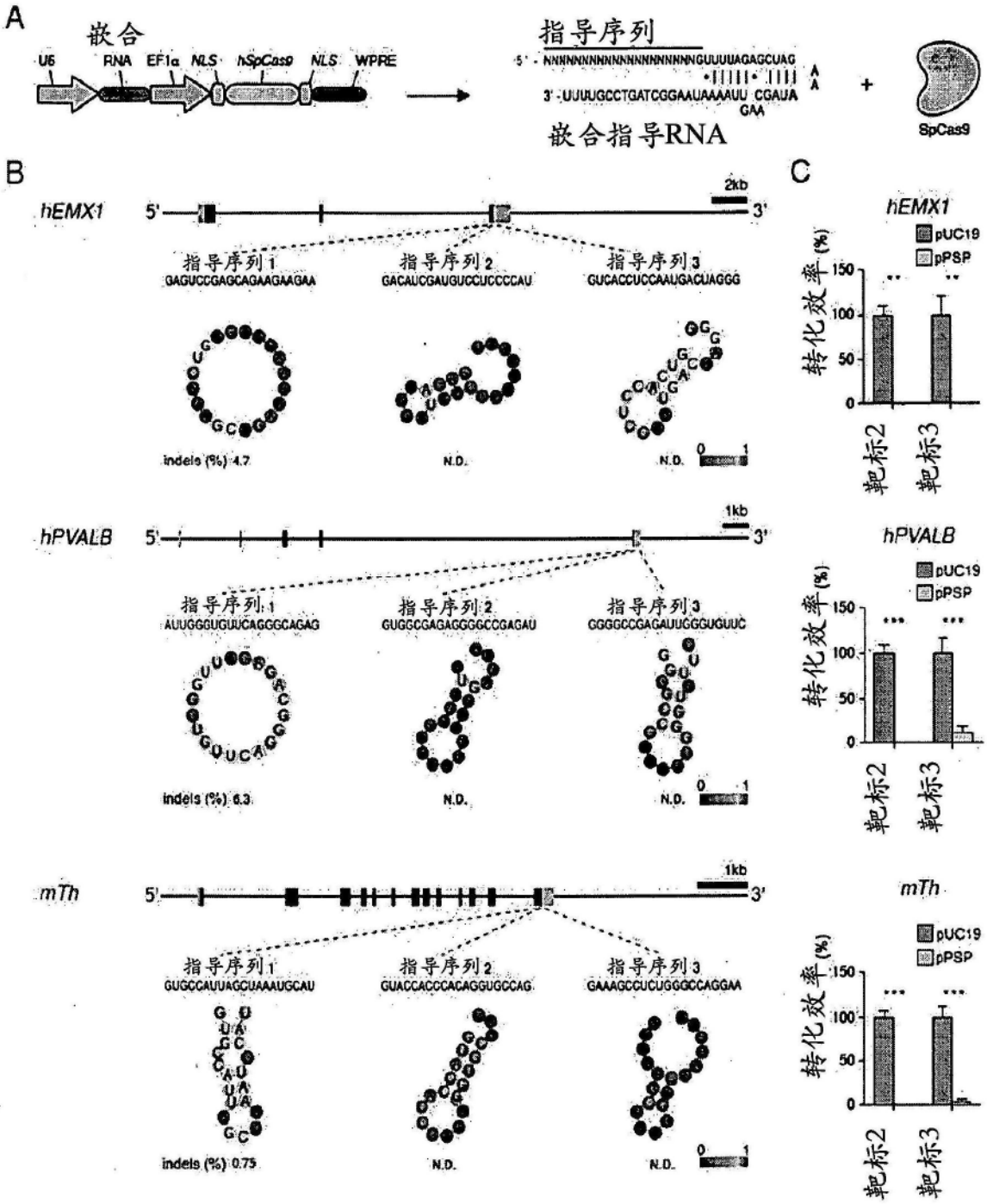

A
嵌合
U6
RNA
EF1α
NLS
hSpCas9
NLS
WPRE
指导序列
5' - NNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAG
3' -UUUUGCCTGATCGGAAUAAAAUU CGAUA
GAA
嵌合指导RNA
SpCas9
B
hEMX1
2kb
5'
3'
指导序列1
GAGUCCGAGCAGAAGAAGAA
指导序列2
GACAUCGAUGUCCUCCCCAU
指导序列3
GUCACCUCCAAUGACUAGGG
indels (%) 4.7
N.D.
N.D.
hPVALB
1kb
指导序列1
AUUGGGUGUUCAGGGCAGAG
指导序列2
GUGGCGAGAGGGGCCGAGAU
指导序列3
GGGGCCGAGAUUGGGUGUUC
indels (%) 6.3
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
pUC19
pPSP
转化效率(%)
150
100
50
0
靶标2
靶标3
hPVALB
mTh

图3A-C

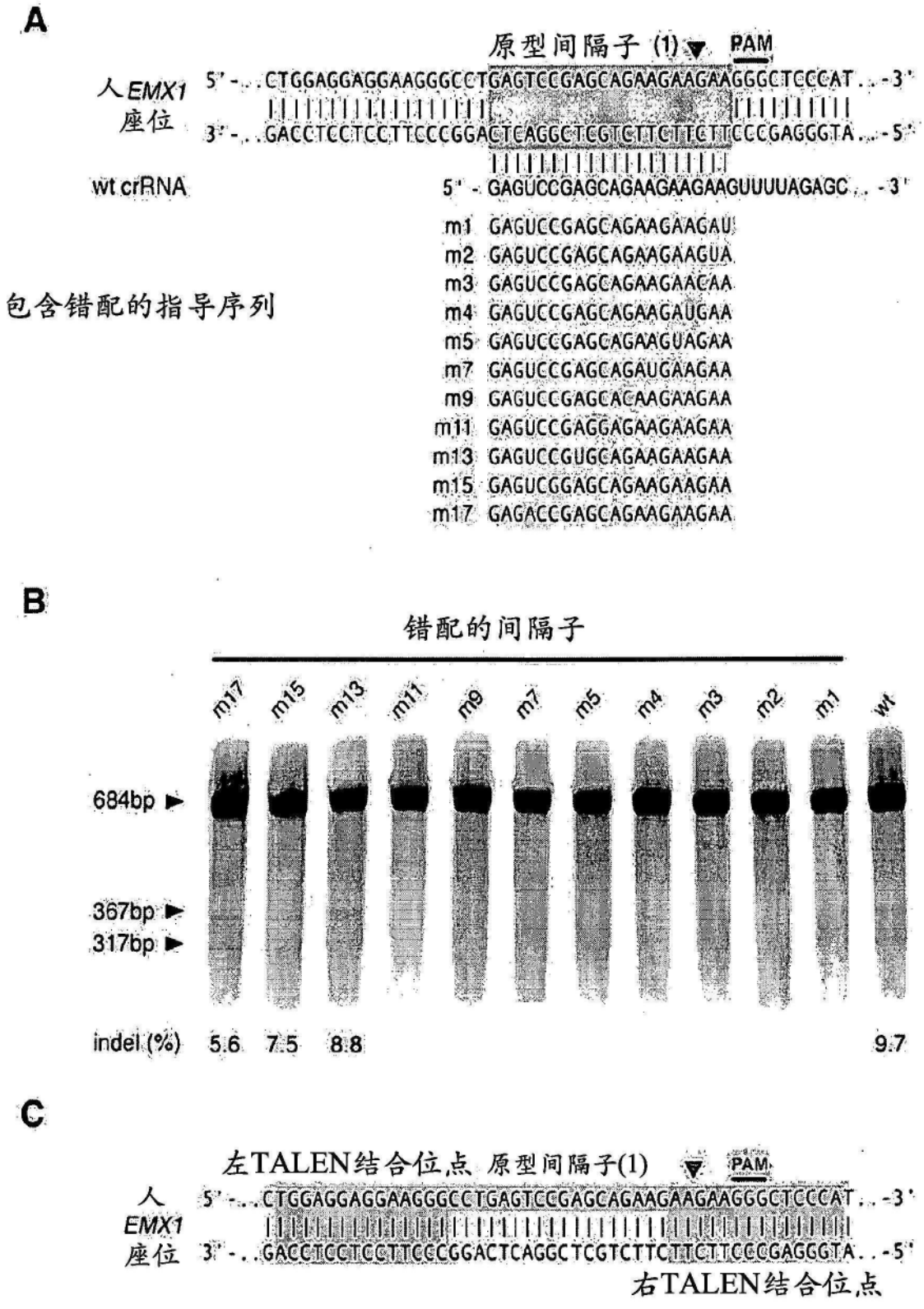

A
原型间隔子 (1)
PAM
人EMX1座位
5' - . . CTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGGGCTCCCAT . . - 3'
3' - . . GACCTCCTCCTTCCCGGACTCAGGCTCGTCTTCTTCTTCCCGAGGGTA . . - 5'
wt crRNA
5' - GAGUCCGAGCAGAAGAAGUUUUAGAGC . . - 3'
包含错配的指导序列
m1 GAGUCCGAGCAGAAGAAGAU
m2 GAGUCCGAGCAGAAGAAGUA
m3 GAGUCCGAGCAGAAGAACAA
m4 GAGUCCGAGCAGAAGAUGAA
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-C

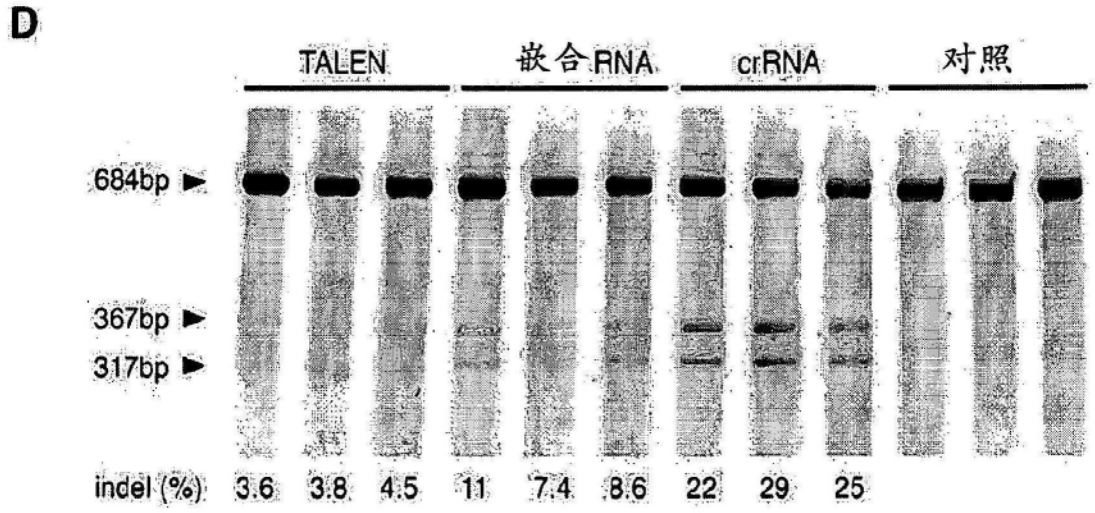

D
TALEN
嵌合RNA
crRNA
对照
684bp
367bp
317bp
indel (%)
3.6
3.8
4.5
11
7.4
8.6
22
29
25

图4D

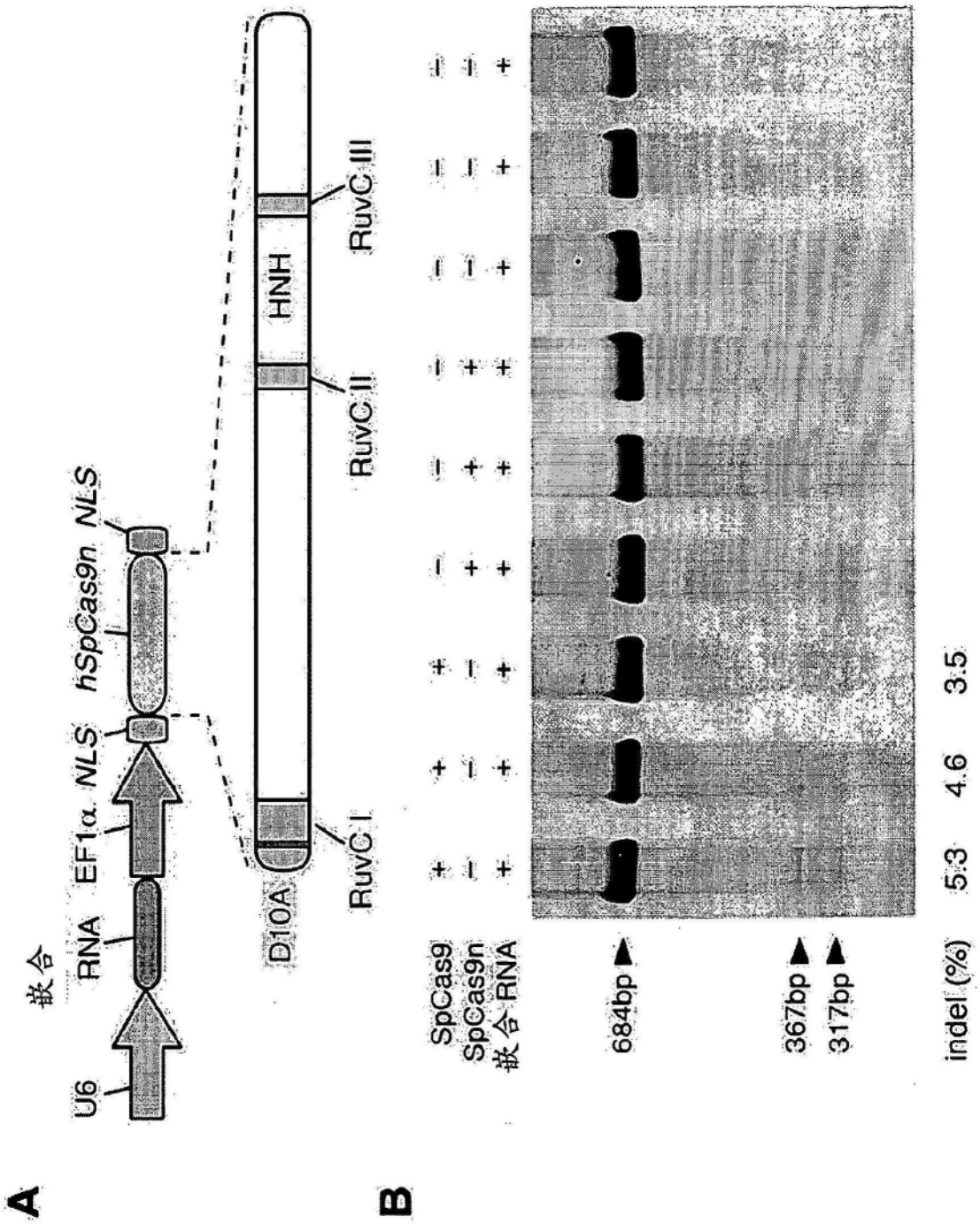

A
U6
嵌合RNA
EF1α
NLS
hSpCas9n
NLS
D10A
RuvC I
RuvC II
HNH
RuvC III
B
SpCas9
SpCas9n
嵌合RNA
684bp
367bp
317bp
indel (%)
5.3
4.6
3.5

图5A-B

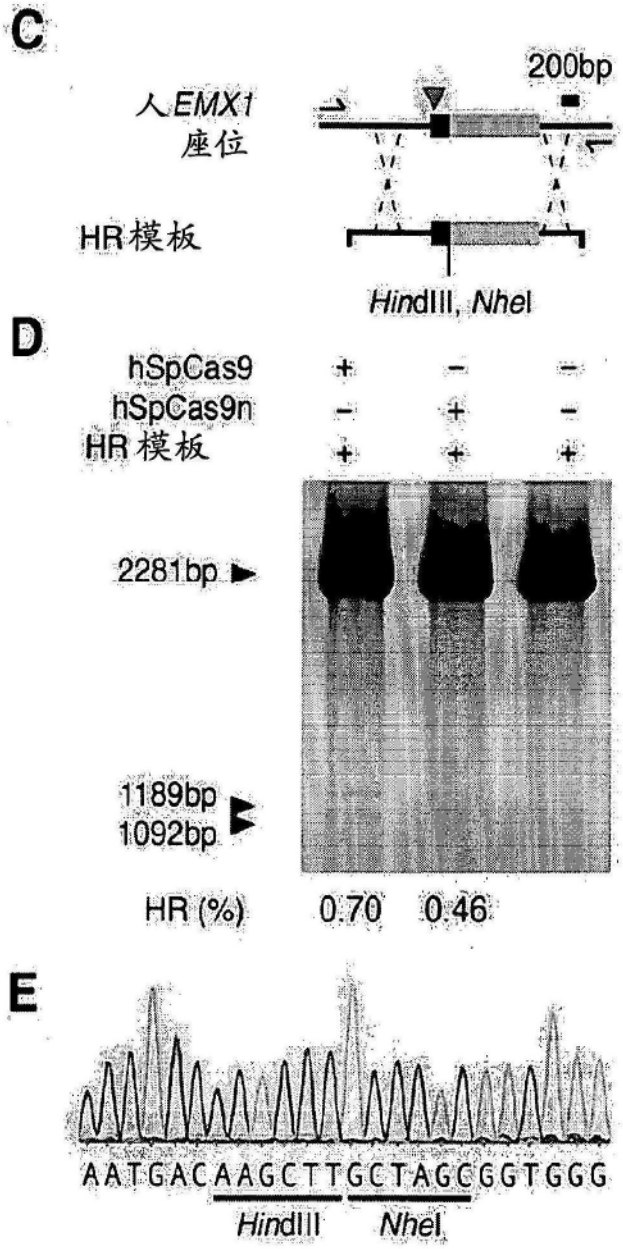

C
200bp
人EMX1
座位
HR模板
HindIII, NheI
D
hSpCas9
hSpCas9n
HR 模板
2281bp
1189bp
1092bp
HR (%)
0.70
0.46
E
AATGACAAGCTTGCTAGCGGTGGG
HindIII
NheI

图5C-E

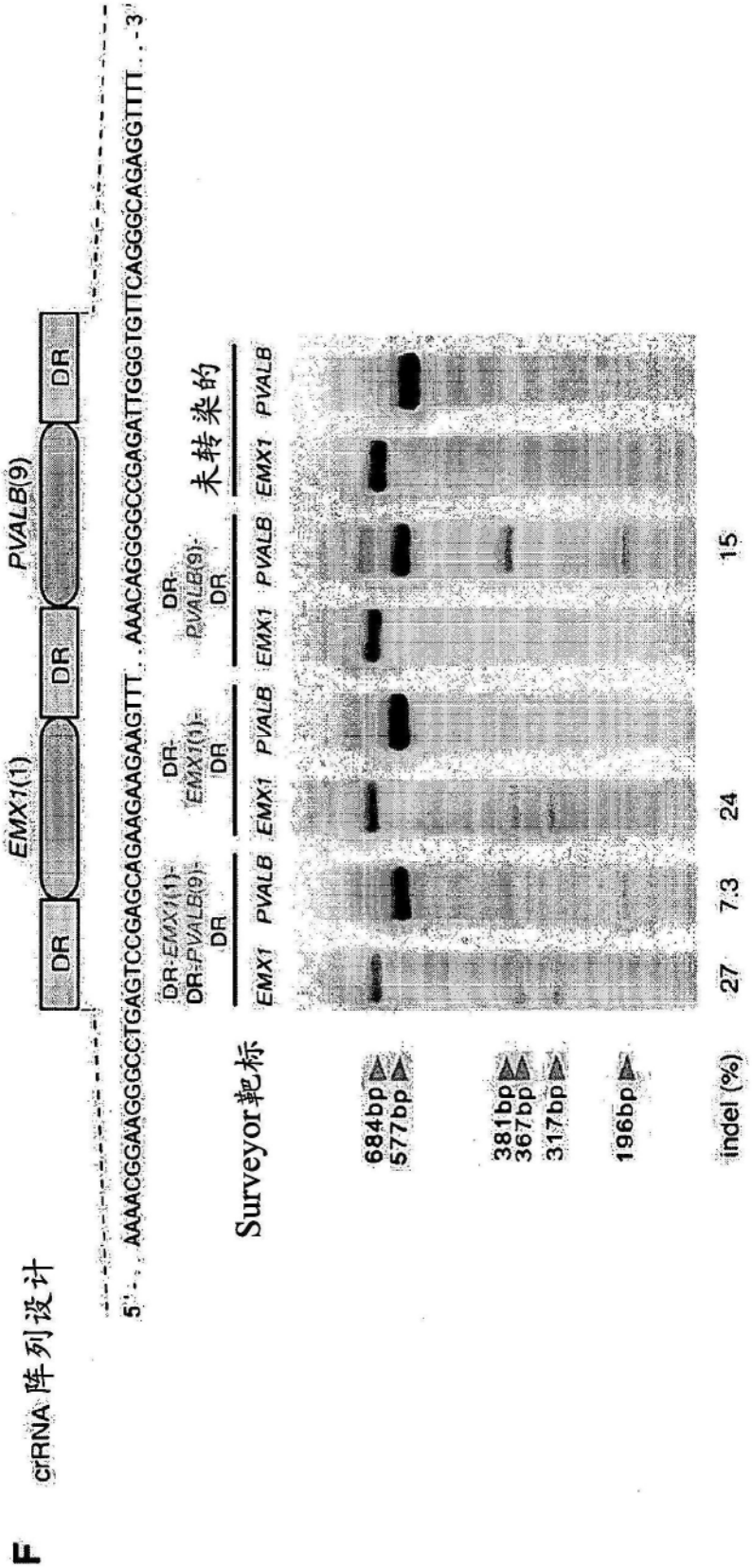

F
crRNA 阵列设计
EMX1(1)
PVALB(9)
DR
DR
DR
DR
5'- AAAACGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGTTT AAACAGGGGCCGAGATTGGGTGTTCAGGGCAGAGGTTTT -3'
Surveyor靶标
DR:EMX1(1)-DR:PVALB(9)-DR
DR-EMX1(1)-DR
DR-PVALB(9)-DR
未转染的
EMX1 PVALB EMX1 PVALB EMX1 PVALB EMX1 PVALB
684bp
577bp
381bp
367bp
317bp
196bp
indel (%) 27 7.3 24 15

图5F

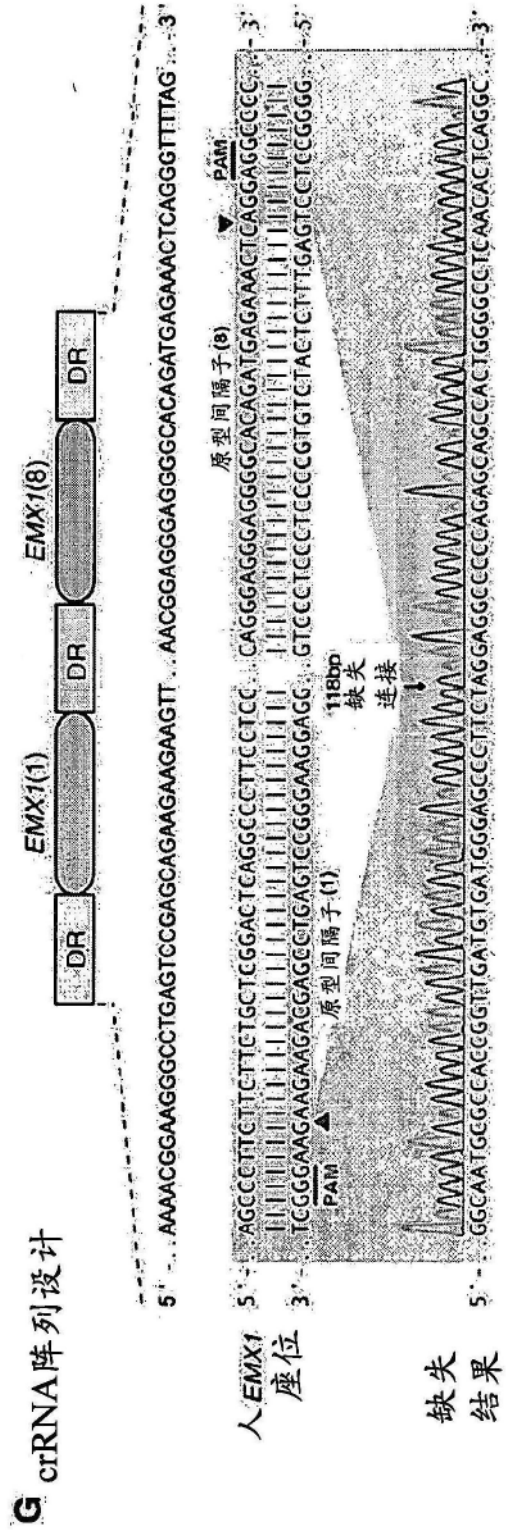

G
crRNA阵列设计
DR
EMX1(1)
DR
EMX1(8)
DR
5' ...AAAACGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGTT...AACGGAGGGAGGGGCACAGATGAGAAACTCAGGGTTTTAG...3'
原型间隔子(8)
PAM
人EMX1座位
5' AGCCCTTCTTCTTCTGCTCGGACTCAGGCCCTTCCTCC CAGGGAGGGAGGGGCACAGATGAGAAACTCAGGAGGCCCC 3'
3' TCGGGAAGAAGAAGACGAGCCTGAGTCCGGGAAGGAGG GTCCCTCCCTCCCCGTGTCTACTCTTTGAGTCCTCCGGGG 5'
PAM
原型间隔子(1)
118bp
缺失
连接
缺失结果
5' GGCAATGCGCCACCGGTTGATGTGATGGGAGCCCTTCTAGGAGGCCCCCAGAGCAGCCACTGGGGCCTCAACACTCAGGC 3'

图5G

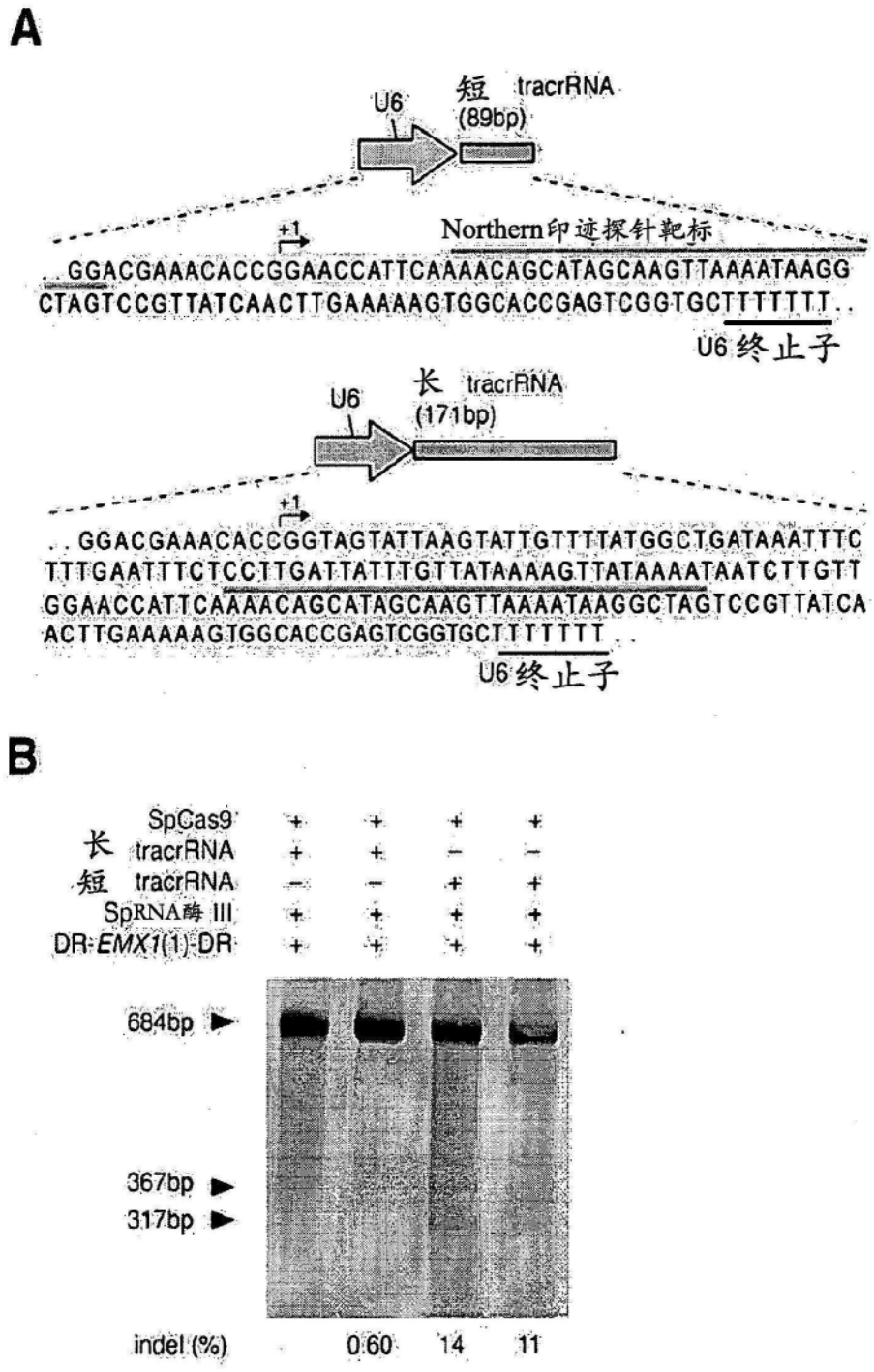

A
U6
短 tracrRNA
(89bp)
+1
Northern印迹探针靶标
..GGACGAAACACCGGAACCATTCAAAACAGCATAGCAAGTTAAAATAAGG
CTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT..
U6终止子
U6
长 tracrRNA
(171bp)
+1
..GGACGAAACACCGGTAGTATTAAGTATTGTTTTATGGCTGATAAATTTC
TTTGAATTTCTCCTTGATTATTTGTTATAAAAGTTATAAAATAATCTTGTT
GGAACCATTCAAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCA
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6A-B

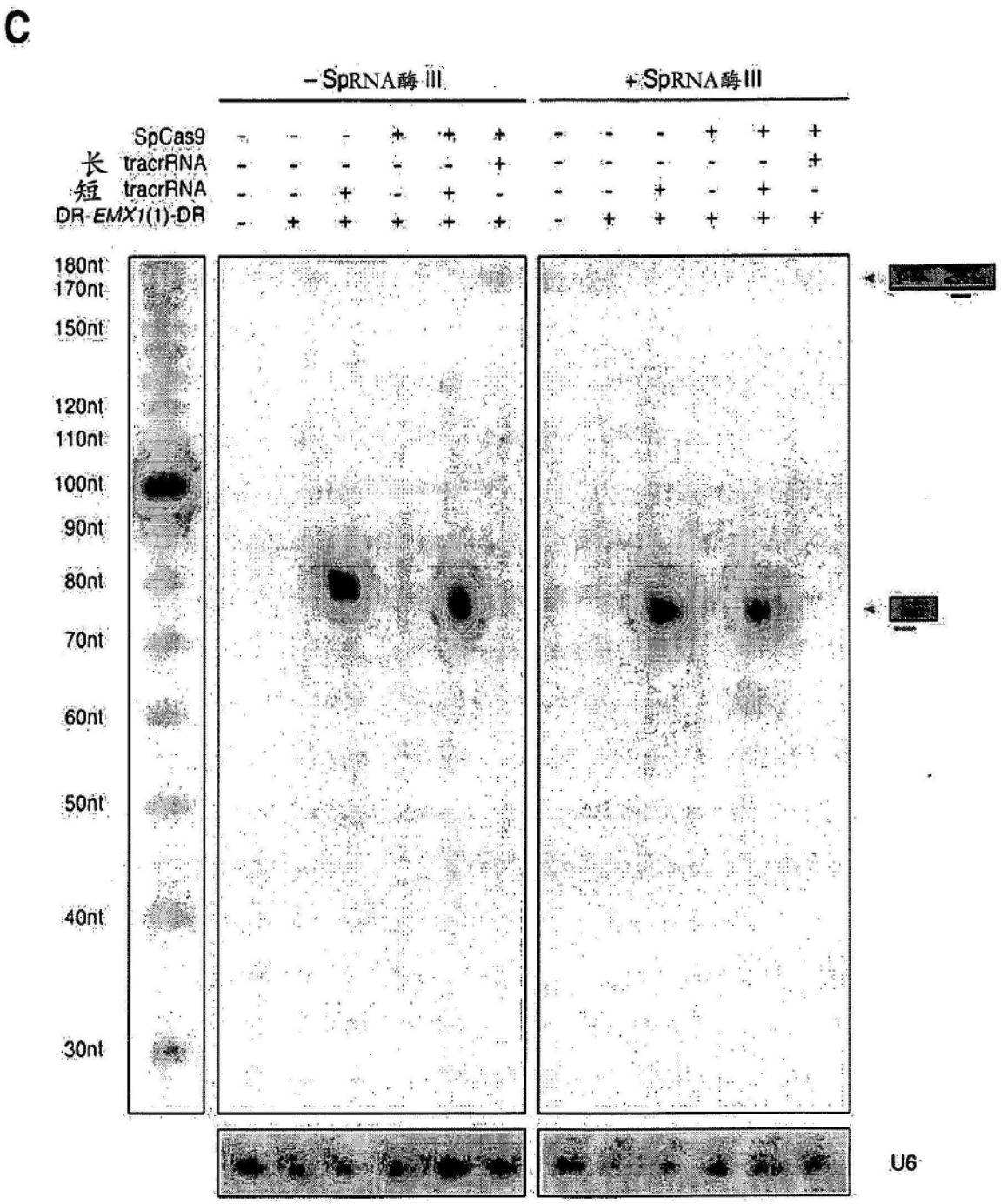

C
-SpRNA酶 III
+SpRNA酶III
SpCas9
长 tracrRNA
短 tracrRNA
DR-EMX1(1)-DR
180nt
170nt
150nt
120nt
110nt
100nt
90nt
80nt
70nt
60nt
50nt
40nt
30nt
U6

图6C

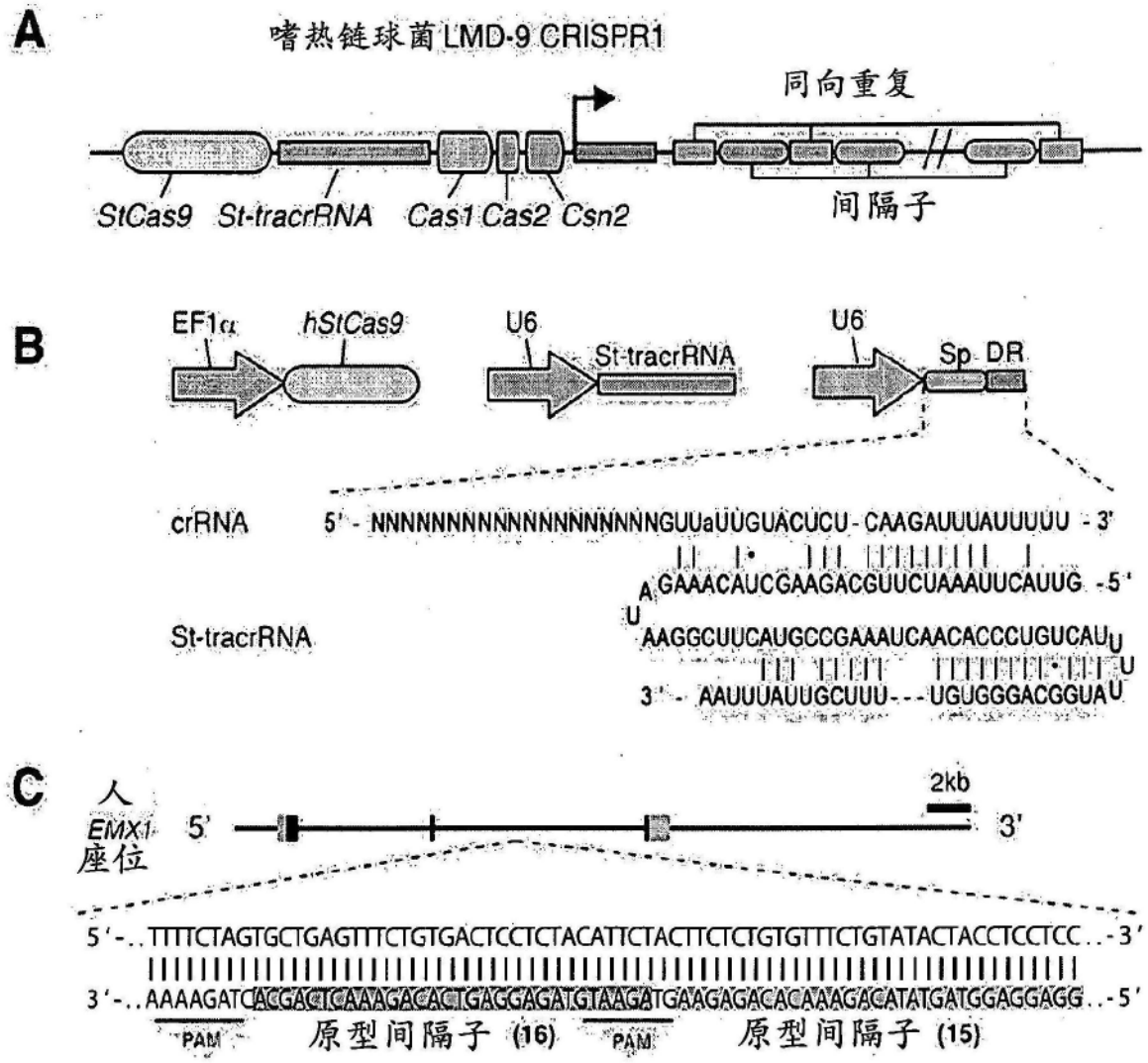

A
嗜热链球菌 LMD-9 CRISPR1
同向重复
StCas9
St-tracrRNA
Cas1 Cas2 Csn2
间隔子
B
EF1α
hStCas9
U6
St-tracrRNA
U6
Sp DR
crRNA
5' - NNNNNNNNNNNNNNNNNNGUUaUUGUACUCU - CAAGAUUUAUUUUU - 3'
A GAAACAUCGAAGACGUUCUAAAUUCAUUG -5'
St-tracrRNA
U AAGGCUUCAUGCCGAAAUCAACACCCUGUCAU U
3' - AAUUUAUUGCUUU - - - UGUGGGACGGUA U
C
人
EMX1
座位
5'
2kb
3'
5'-..TTTTCTAGTGCTGAGTTTCTGTGACTCCTCTACATTCTACTTCTCTGTGTTTCTGTATACTACCTCCTCC..-3'
3'-..AAAAGATCACGACTCAAAGACACTGAGGAGATGTAAGATGAAGAGACACAAAGACATATGATGGAGGAGG..-5'
PAM
原型间隔子 (16)
PAM
原型间隔子 (15)

图7A-C

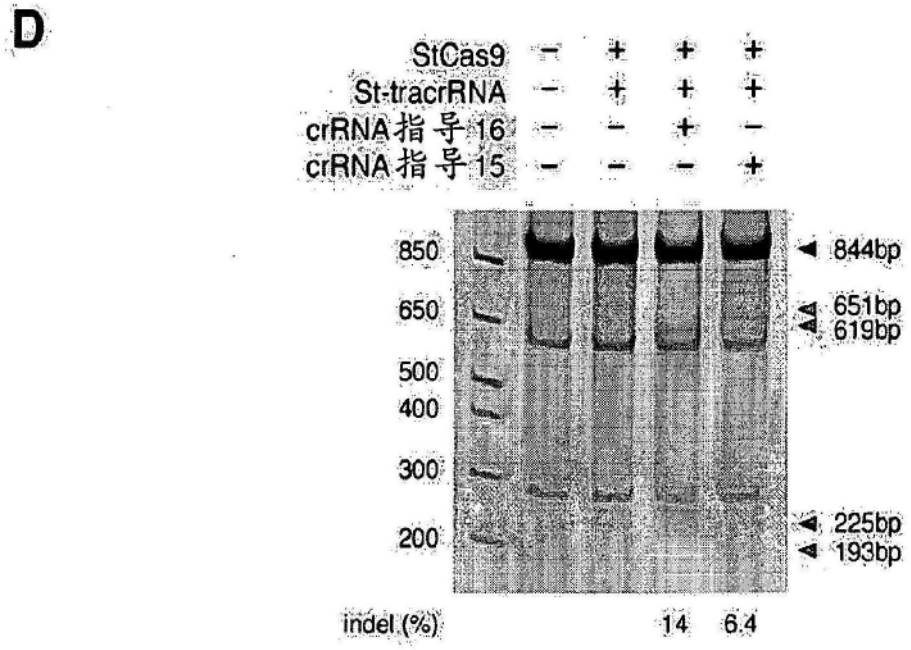

D
StCas9 − + + +
St-tracrRNA − + + +
crRNA指导16 − − + −
crRNA指导15 − − − +
850
650
500
400
300
200
844bp
651bp
619bp
225bp
193bp
indel (%) 14 6.4

图7D

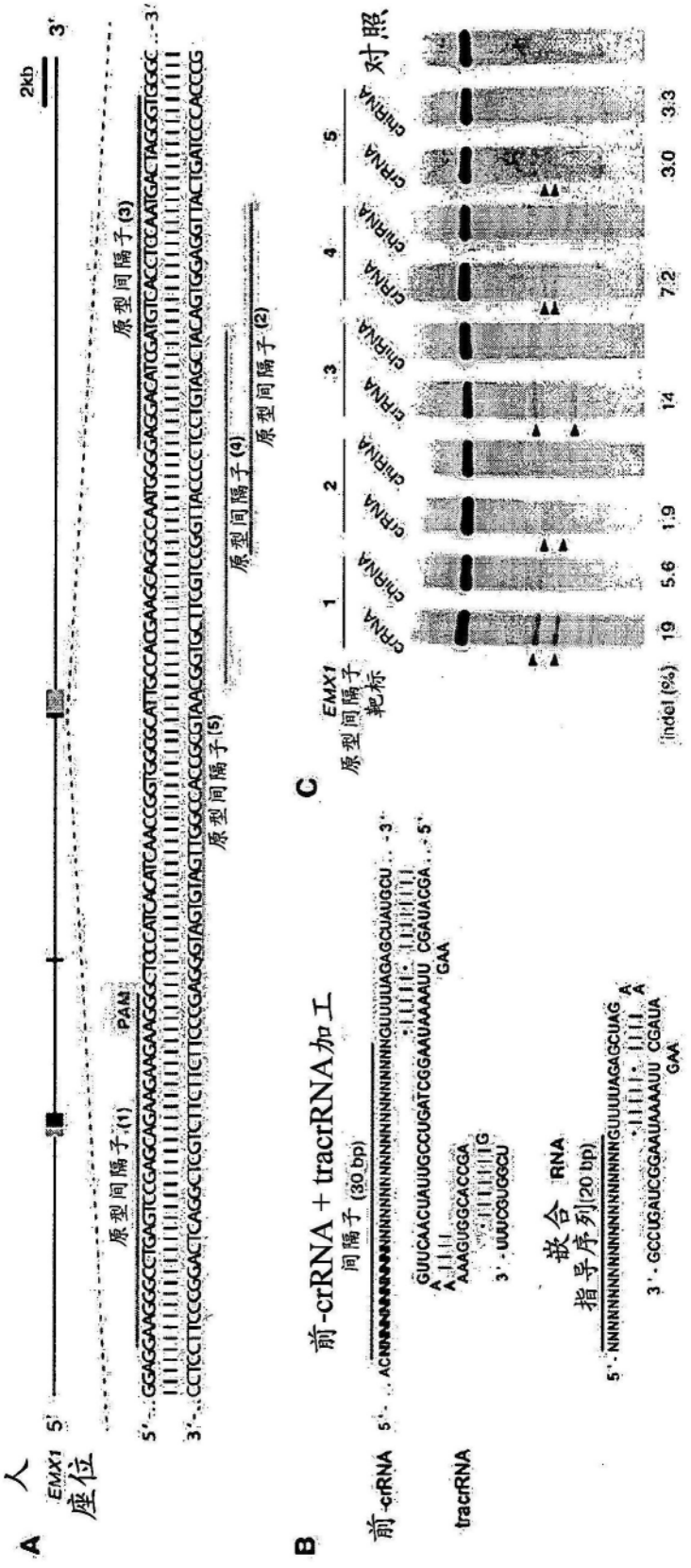

A
人
EMX1
座位
2kb
5'
3'
原型间隔子(1)
PAM
原型间隔子(3)
原型间隔子(5)
原型间隔子(4)
原型间隔子(2)
B
前-crRNA + tracrRNA加工
间隔子 (30 bp)
前-crRNA
tracrRNA
嵌合 RNA
指导序列(20 bp)
C
EMX1
原型间隔子
靶标
crRNA
chiRNA
对照
indel (%)
19
5.6
1.9
14
7.2
3.0
3.3

图8A-C

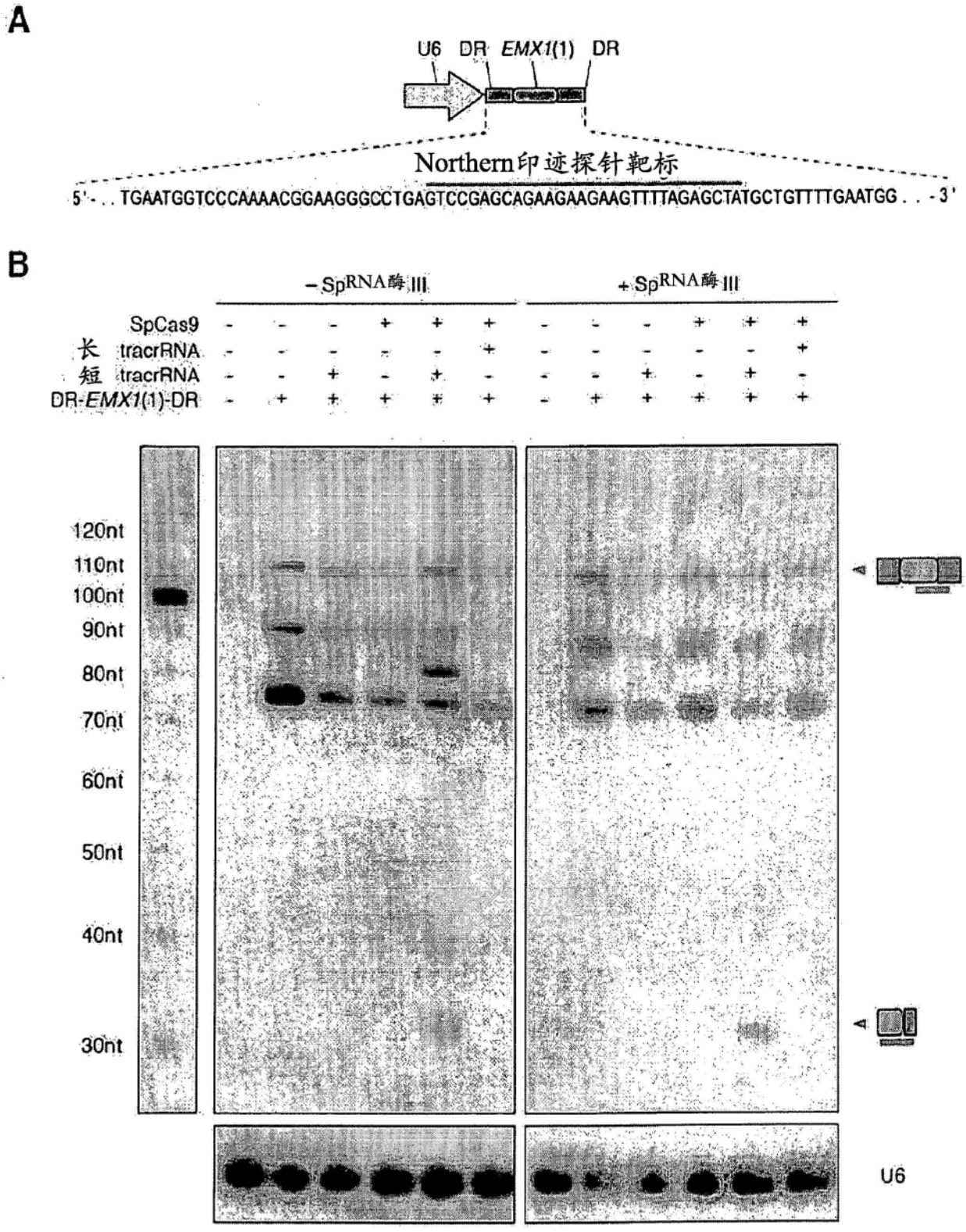

A
U6
DR
EMX1(1)
DR
Northern印迹探针靶标
5'- . . TGAATGGTCCCAAAACGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGTTTTAGAGCTATGCTGTTTTGAATGG . . -3'
B
-SpRNA酶III
+SpRNA酶III
SpCas9
长 tracrRNA
短 tracrRNA
DR-EMX1(1)-DR
120nt
110nt
100nt
90nt
80nt
70nt
60nt
50nt
40nt
30nt
U6

图9A-B

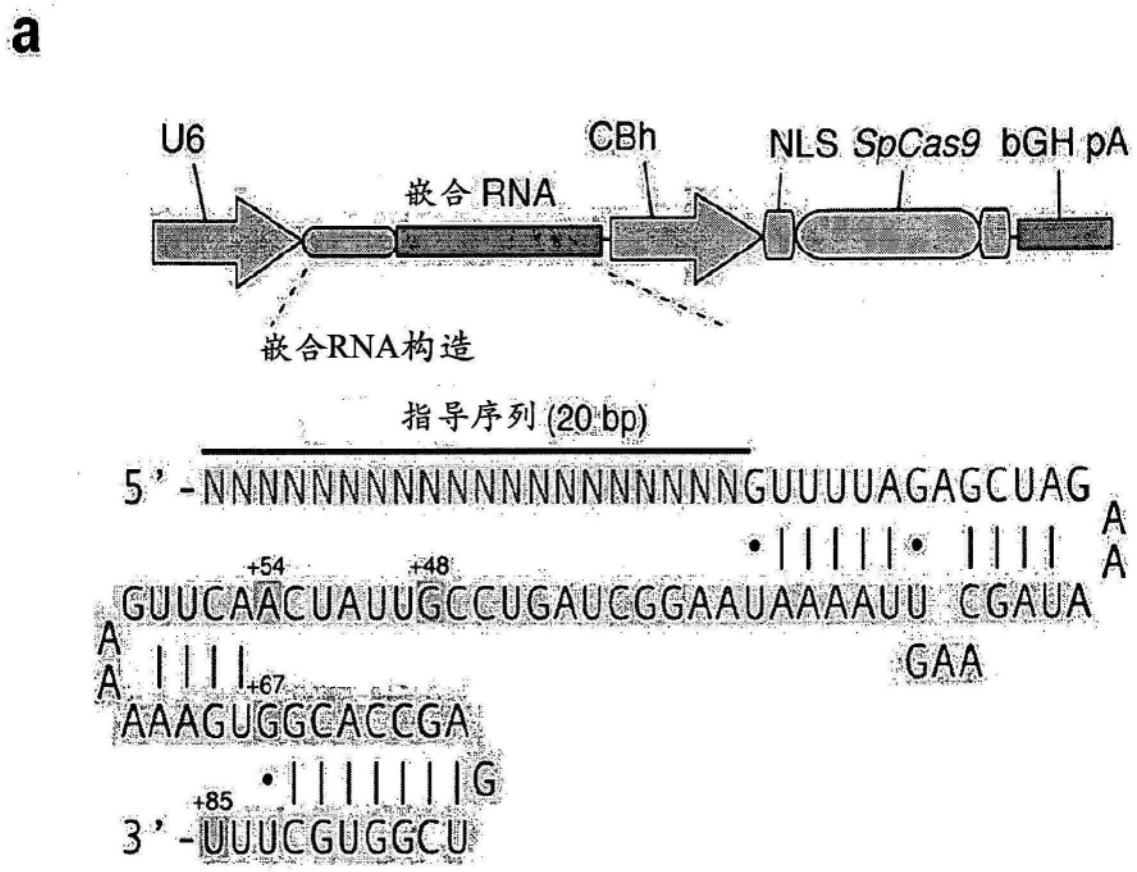

a
U6
嵌合 RNA
CBh
NLS SpCas9 bGH pA
嵌合RNA构造
指导序列 (20 bp)
5'-NNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAG
A
A
+54
+48
GUUCAACUAUUGCCUGAUCGGAAUAAAAUU CGAUA
A
A
GAA
+67
AAAGUGGCACCGA
G
+85
3'-UUUCGUGGCU

图10A

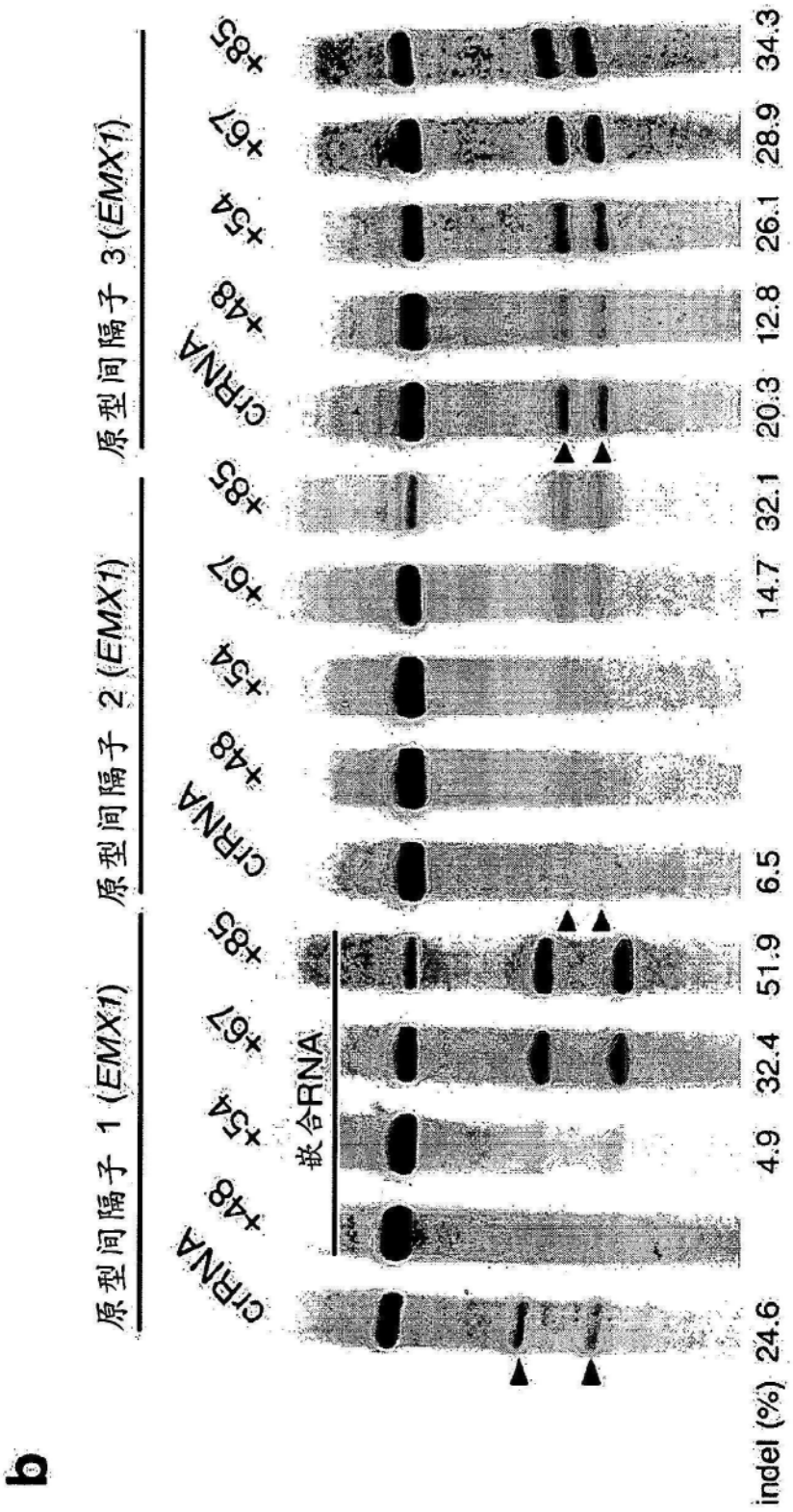

b
原型间隔子 1 (EMX1)
原型间隔子 2 (EMX1)
原型间隔子 3 (EMX1)
gRNA
+48
+54
+67
+85
嵌合RNA
indel (%)
24.6
4.9
32.4
51.9
6.5
14.7
32.1
20.3
12.8
26.1
28.9
34.3

图10B

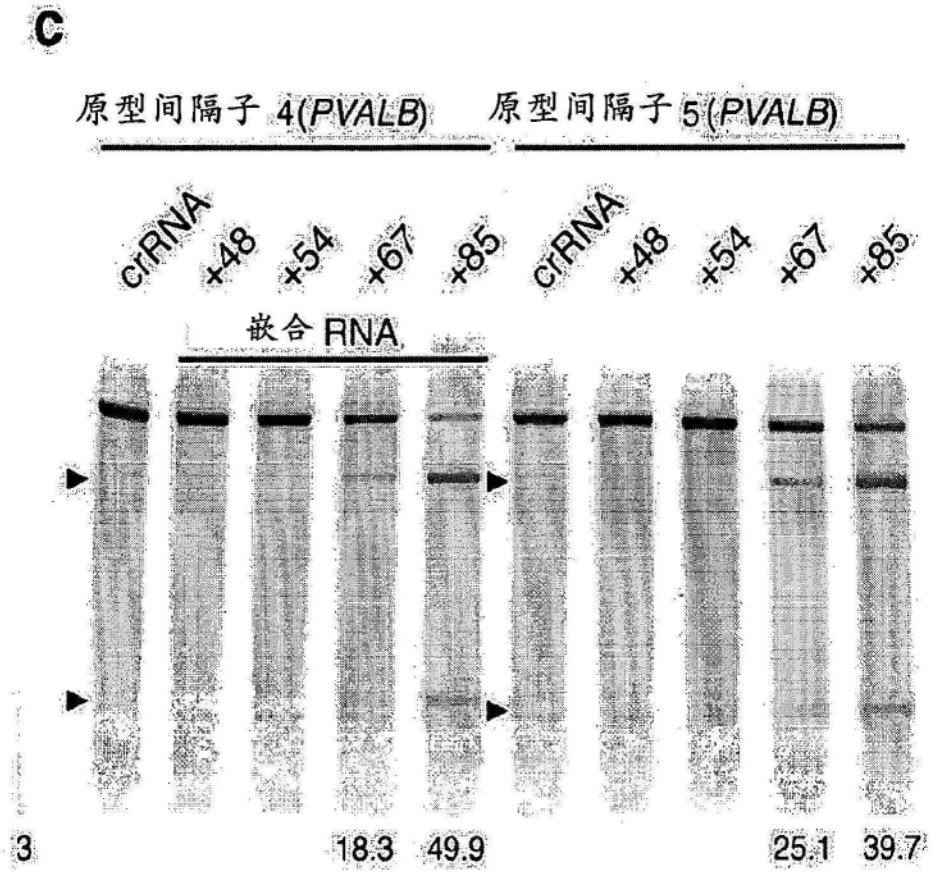

C
原型间隔子 4(PVALB)
原型间隔子 5(PVALB)
crRNA
+48
+54
+67
+85
crRNA
+48
+54
+67
+85
嵌合 RNA
3
18.3
49.9
25.1
39.7

图10C

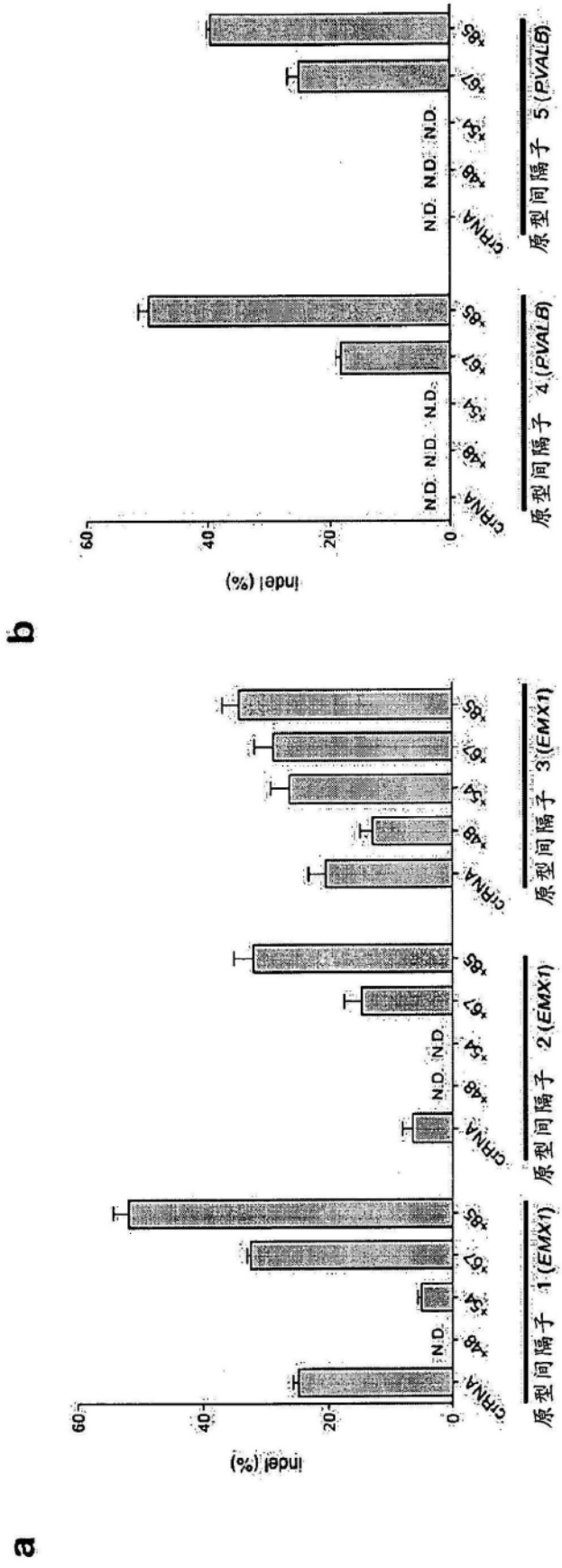

a
b
Indel (%)
0
20
40
60
crRNA
48
54
67
85
N.D.
原型间隔子 1 (EMX1)
原型间隔子 2 (EMX1)
原型间隔子 3 (EMX1)
原型间隔子 4 (PVALB)
原型间隔子 5 (PVALB)

图11A-B

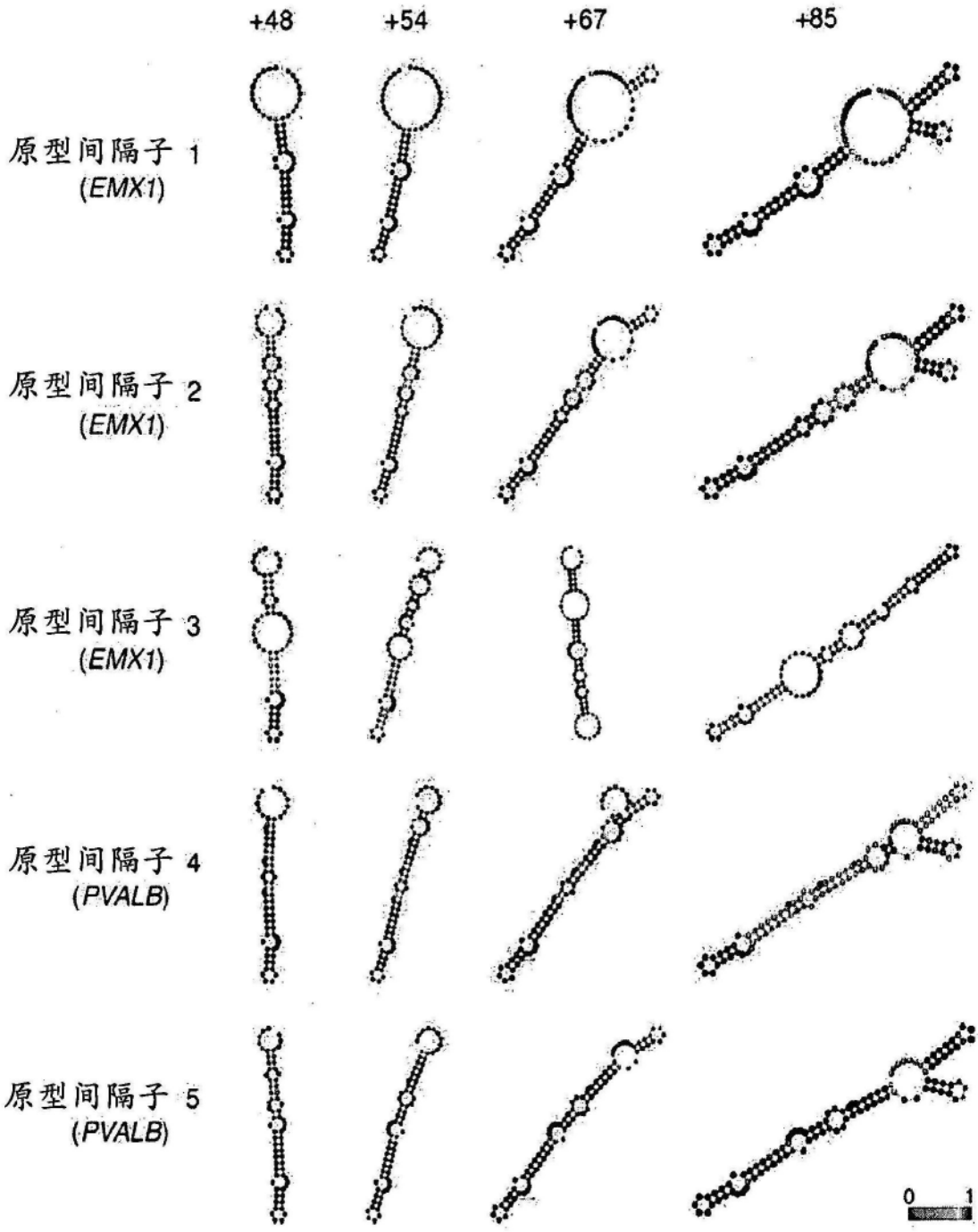

+48
+54
+67
+85
原型间隔子 1
(EMX1)
原型间隔子 2
(EMX1)
原型间隔子 3
(EMX1)
原型间隔子 4
(PVALB)
原型间隔子 5
(PVALB)
0
1

图12

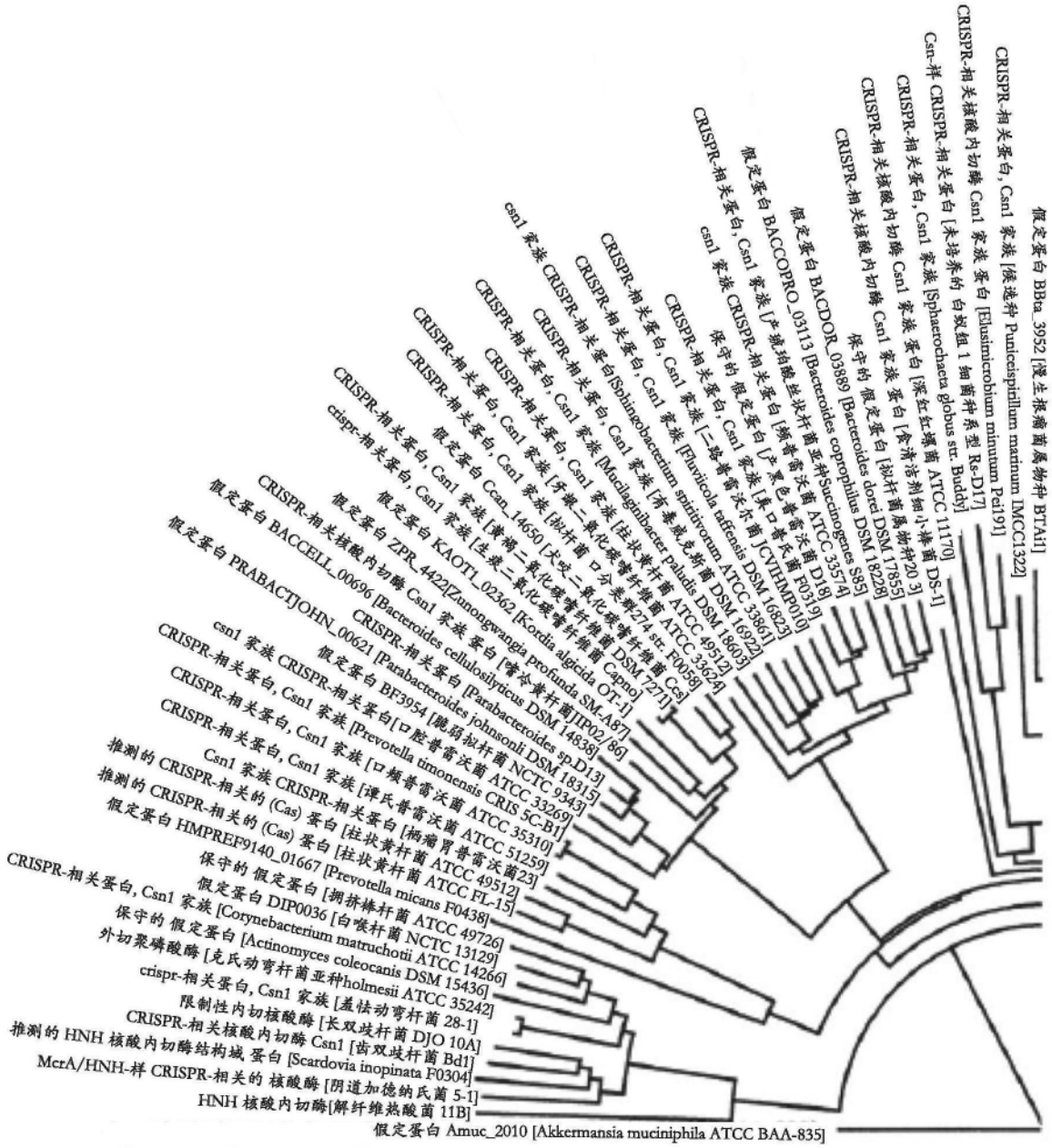

假定蛋白 BBta_3952 [慢生根瘤菌属物种 BTAi1]
CRISPR-相关蛋白, Csn1 家族 [候选种 Punicispirillum marinum IMCC1322]
CRISPR-相关核酸内切酶 Csn1 家族 蛋白 [Elusimicrobium minutum Pei191]
Csn-样 CRISPR-相关蛋白 [未培养的 白蚁组 1 细菌种系型 Rs-D17]
CRISPR-相关蛋白, Csn1 家族 [Sphaerochaeta globus str. Buddy]
CRISPR-相关核酸内切酶 Csn1 家族 蛋白 [深红红螺菌 ATCC 11170]
CRISPR-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 [斯普雷沃菌 ATCC 33574]
csn1 家族 CRISPR-相关蛋白 [卢氏拟杆菌 D18]
保守的 假定蛋白 [卢氏色普雷沃菌 F0319]
CRISPR-相关蛋白, Csn1 家族 [真口普氏菌 JCVIHMP010]
CRISPR-相关蛋白, Csn1 家族 [二级普雷沃尔菌 JCVIHMP022]
CRISPR-相关蛋白, Csn1 家族 [Fluviicola taffensis DSM 16823]
CRISPR-相关蛋白, Csn1 家族 [布莱威克斯菌 DSM 16922]
CRISPR-相关蛋白(Sphinorbacterium spititvorum ATCC 33861]
CRISPR-相关蛋白, Csn1 家族 [Mucilaginibacter paludis DSM 18603]
CRISPR-相关蛋白, Csn1 家族 [拉状黄杆菌 ATCC 49512]
csn1 家族 CRISPR-相关蛋白 [Sphingobacterium spiritivorum ATCC 33624]
CRISPR-相关蛋白, Csn1 家族 [牙龈二氧化碳嗜纤维菌 ATCC 43718]
CRISPR-相关蛋白, Csn1 家族 [拟杆菌 口分类群274 str. F0058]
CRISPR-相关蛋白, Csn1 家族 [犬咬二氧化碳嗜纤维菌 DSM 7271]
CRISPR-相关蛋白, Csn1 家族 [唐纳二氧化碳嗜纤维菌 Capno]
假定蛋白 Ccan_14650 [犬咬二氧化碳嗜纤维菌 Ccs]
crispr-相关蛋白, Csn1 家族 [生菜二氧化碳嗜纤维菌 DSM 7271]
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 [噬冷黄杆菌JIP02/86]
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 sp.D13]
csn1 家族 CRISPR-相关蛋白 BF3954 [脆弱拟杆菌 NCTC 9343]
CRISPR-相关蛋白, Csn1 家族 [口腔普雷沃菌 ATCC 33269]
CRISPR-相关蛋白, Csn1 家族 [Prevotella timonensis CRIS 5C-B1]
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 [栖瘤胃普雷沃菌 ATCC 51259]
推测的 CRISPR-相关的 (Cas) 蛋白 [柱状黄杆菌 ATCC 49512]
假定蛋白 HMPREF9140_01667 [Prevotella micans F0438]
保守的 假定蛋白 [拥挤棒杆菌 ATCC FL-15]
CRISPR-相关蛋白, Csn1 家族 [白喉杆菌 NCTC 13129]
假定蛋白 DIP0036 [Corynebacterium matruchotii ATCC 14266]
保守的 假定蛋白 [Actinomyces coleocanis DSM 15436]
外切聚磷酸酶 [克氏动弯杆菌亚种holmesii ATCC 35242]
crispr-相关蛋白, Csn1 家族 [盖怯动弯杆菌 28-1]
限制性内切核酸酶 [长双歧杆菌 DJO 10A]
CRISPR-相关核酸内切酶 Csn1 [齿双歧杆菌 Bd1]
推测的 HNH 核酸内切酶结构域 蛋白 [Scardovia inopinata F0304]
McrA/HNH-样 CRISPR-相关的 核酸酶 [阴道加德纳氏菌 5-1]
HNH 核酸内切酶[解纤维热酸菌 11B]
假定蛋白 Amuc_2010 [Akkermansia muciniphila ATCC BAA-835]

图13A

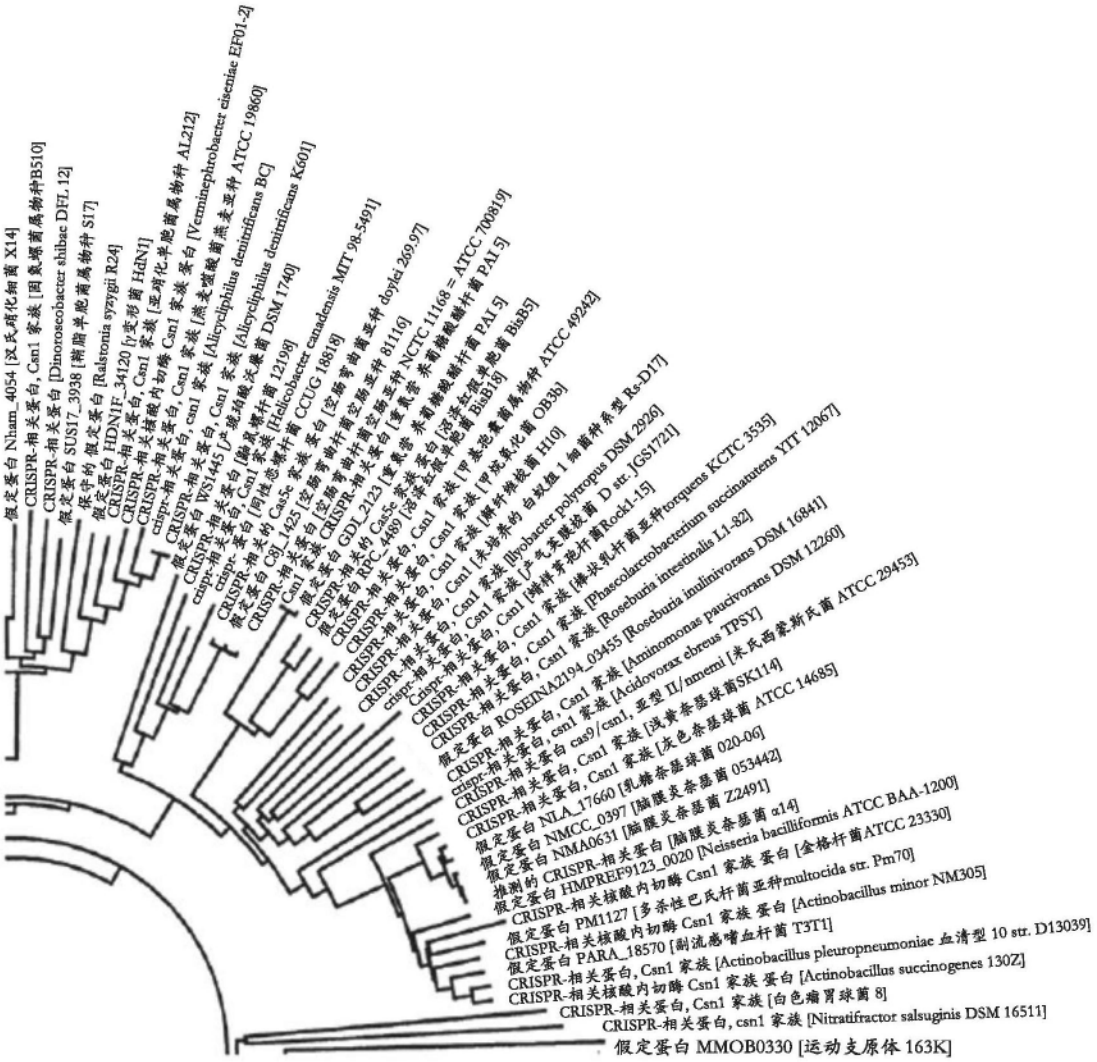

假定蛋白 Nham_4054 [汉氏硝化细菌 X14]
CRISPR-相关蛋白, Csn1 家族 [固氮螺菌属物种B510]
CRISPR-相关蛋白 [Dinoroseobacter shibae DFL 12]
假定蛋白 SUS17_3938 [南甜羊肥菌属物种 S17]
保守的 假定蛋白 [Ralstonia syzygii R24]
假定蛋白 HDN1F_34120 [*克利菌 HdN1]
CRISPR-相关核酸内切酶 Csn1 家族 [亚硝化单胞菌属物种 AL212]
crispr-相关蛋白, csn1 家族 [Alicycliphilus denitrificans BC]
CRISPR-相关蛋白, Csn1 家族 [Alicycliphilus denitrificans K601]
假定蛋白 WS1445 [产*脱硝沃康菌 DSM 1740]
crispr-相关蛋白, Csn1 家族 [脱氮螺杆菌 12198]
CRISPR-相关的 Cas5c 家族 蛋白 [Helicobacter canadensis MIT 98-5491]
假定蛋白 GDL_2123 [空肠弯曲杆菌空肠亚种 81116]
CRISPR-相关蛋白, Csn1 家族 [*嗜胆汁弯曲杆菌亚种 doylei 269.97]
假定蛋白 RFC_4489 [沼泽红假单胞菌 BisB18]
CRISPR-相关蛋白, Csn1 家族 [甲基泥囊菌属物种 ATCC 49242]
CRISPR-相关蛋白, Csn1 家族 [*缺养的 白*红假 1 细菌种系型 Rs-D17]
CRISPR-相关蛋白, Csn1 家族 [Hyphobacter polyropus DSM 2926]
CRISPR-相关蛋白, Csn1 家族 [产气*腐极菌 D str. JGS1721]
CRISPR-相关蛋白, Csn1 家族 [*酵母孢杆菌亚种torquens KCTC 3535]
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 [Roseburia inulinivorans DSM 16841]
CRISPR-相关蛋白, Csn1 家族 [Aminomonas paucivorans DSM 12260]
假定蛋白 ROSEINA2194_03455 [Roseburia inulinivorans DSM 16841]
CRISPR-相关蛋白, csn1 家族 [Acidovorax cbreus TPSY]
假定蛋白 cas9/csn1, 亚型 II/nmemi [关氏西雷斯氏菌 ATCC 29453]
crispr-相关蛋白, Csn1 家族 [浅黄*基球菌SK114]
CRISPR-相关蛋白, Csn1 家族 [灰色*基球菌 020-06]
CRISPR-相关蛋白 NLA_17660 [乳*基球菌 053442]
假定蛋白 NMCC_0397 [脑膜炎*基菌 Z2491]
假定蛋白 NMA0631 [脑膜炎*基菌 α14]
假定蛋白 CRISPR-相关蛋白 [Neisseria bacilliformis ATCC BAA-1200]
推测的 CRISPR-相关蛋白 Csn1 家族 蛋白 [金格杆菌ATCC 23330]
假定蛋白 HMPREF9123_0020 [Neisseria bacilliformis ATCC BAA-1200]
CRISPR-相关核酸内切酶 Csn1 家族 蛋白 [Actinobacillus sp. Pm70]
假定蛋白 PM1127 [多杀性巴氏杆菌亚种multocida str. Pm70]
CRISPR-相关核酸内切酶 Csn1 家族 蛋白 [Actinobacillus minor NM305]
假定蛋白 PARA_18570 [副流感*血杆菌 T3T1]
CRISPR-相关蛋白, Csn1 家族 [Actinobacillus pleuropneumoniae 血清型 10 str. D13039]
CRISPR-相关核酸内切酶 Csn1 家族 蛋白 [Actinobacillus succinogenes 130Z]
CRISPR-相关蛋白, Csn1 家族 [白色*胃球菌 8]
CRISPR-相关蛋白, csn1 家族 [Nitratifractor salsuginis DSM 16511]
假定蛋白 MMOB0330 [运动支原体 163K]

图13B

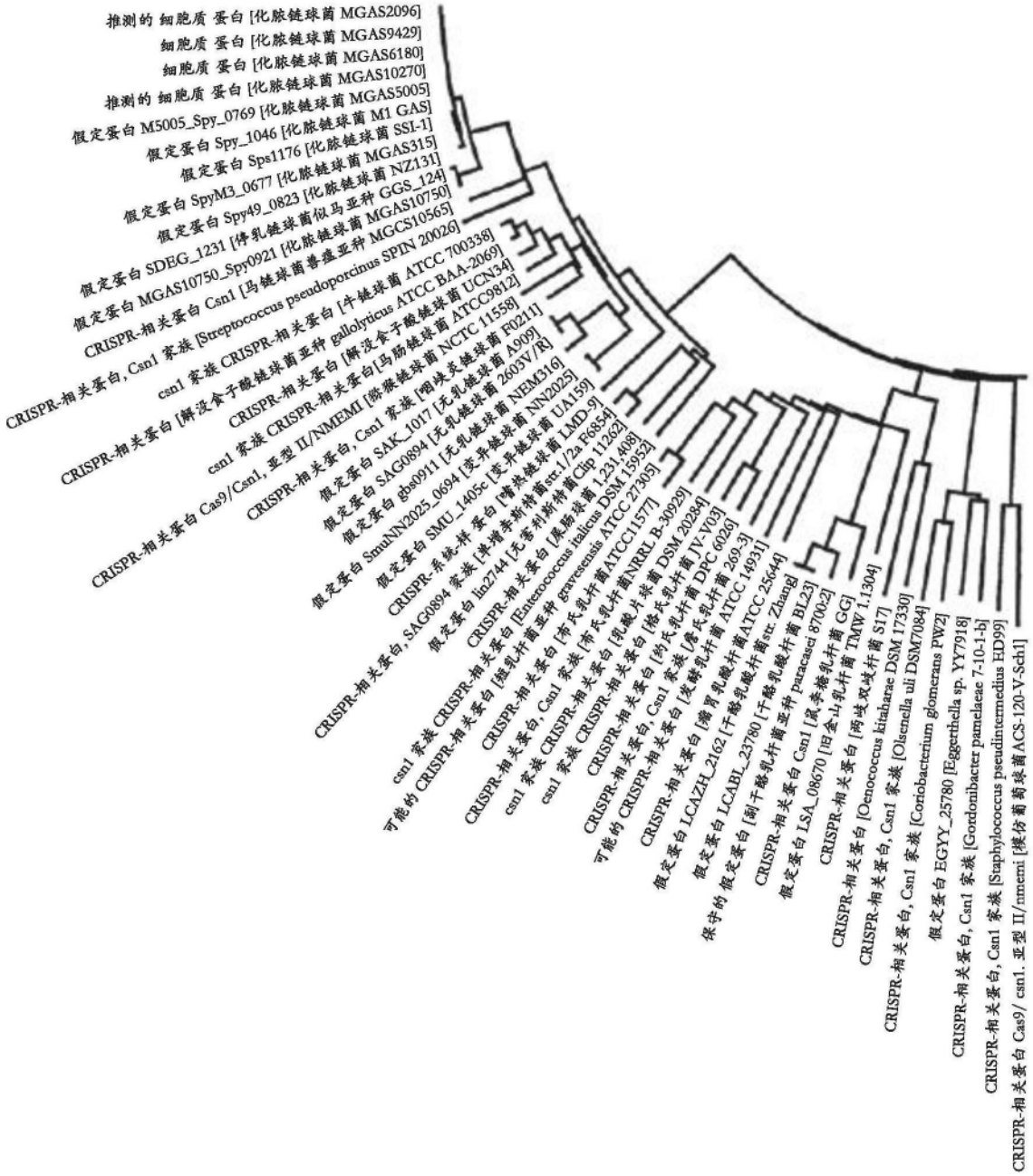

推测的 细胞质 蛋白 [化脓链球菌 MGAS2096]
细胞质 蛋白 [化脓链球菌 MGAS9429]
细胞质 蛋白 [化脓链球菌 MGAS6180]
推测的 细胞质 蛋白 [化脓链球菌 MGAS10270]
假定蛋白 M5005_Spy_0769 [化脓链球菌 MGAS5005]
假定蛋白 Spy_1046 [化脓链球菌 M1 GAS]
假定蛋白 Sps1176 [化脓链球菌 SSI-1]
假定蛋白 SpyM3_0677 [化脓链球菌 MGAS315]
假定蛋白 Spy49_0823 [化脓链球菌 NZ131]
假定蛋白 SDEG_1231 [停乳链球菌似马亚种 GGS_124]
假定蛋白 MGAS10750_Spy0921 [化脓链球菌 MGAS10750]
CRISPR-相关蛋白 Csn1 [马链球菌兽瘟亚种 MGCS10565]
csn1 家族 CRISPR-相关蛋白 [Streptococcus pseudoporcinus SPIN 20026]
CRISPR-相关蛋白, Csn1 家族 [解没食子酸链球菌 ATCC 700338]
CRISPR-相关蛋白 [解没食子酸链球菌 UCN34]
CRISPR-相关蛋白 Cas9/Csn1, 亚型 II/NMEMI [脓肿链球菌 ATCC BAA-2069]
csn1 家族 CRISPR-相关蛋白 [解没食子酸链球菌 ATCC9812]
CRISPR-相关蛋白, Csn1 家族 [肺炎链球菌 NCTC 11558]
假定蛋白 SAK_1017 [无乳链球菌 A909]
假定蛋白 SAG0894 [无乳链球菌 NEM316]
假定蛋白 gbs0911 [无乳链球菌 2603V/R]
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-神 蛋白 [嗜热链球菌str.1/2a F6854]
csn1 家族 CRISPR-相关蛋白 [嗜热链球菌 LMD-9]
CRISPR-相关蛋白, Csn1 家族 [肺炎乳杆菌Clip 11262]
CRISPR-相关蛋白 [Enterococcus italicus DSM 15952]
假定蛋白 lin2744 [无害利斯特菌 1,231,408]
CRISPR-相关蛋白 [屎肠球菌 ATCC 27305]
CRISPR-相关蛋白, SAG0894 家族 [单增李斯特菌 ATCC11577]
csn1 家族 CRISPR-相关蛋白 [屎肠球菌NRRL B-30929]
CRISPR-相关蛋白 [短乳杆菌亚种 gravesensis DSM 20284]
可能的 CRISPR-相关蛋白 [发酵乳杆菌 JV-V03]
CRISPR-相关蛋白, Csn1 家族 [肺炎乳杆菌 DPC 6026]
CRISPR-相关蛋白 [弯曲乳杆菌 269-3]
CRISPR-相关蛋白 [嗜胃乳酸杆菌ATCC 14931]
假定蛋白 LCAZH_2162 [干酪乳酸杆菌str. Zhang]
假定蛋白 LCABL_23780 [干酪乳杆菌亚种 paracasei 8700-2]
保守的 假定蛋白 [副干酪乳杆菌 BL23]
CRISPR-相关蛋白 Csn1 [保 李酵乳杆菌 GG]
假定蛋白 LSA_08670 [旧金山乳杆菌 TMW 1.1304]
CRISPR-相关蛋白 [Oenococcus kitaharae DSM 17330]
CRISPR-相关蛋白, Csn1 家族 [Olsenella uli DSM7084]
CRISPR-相关蛋白, Csn1 家族 [Coriobacterium glomerans PW2]
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13C


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]
假定蛋白 stu0657 [嗜热链球菌 LMG 18311]
CRISPR-系统 家族 蛋白 [前庭链球菌 ATCC 49124]
CRISPR-系统-样 蛋白 [嗜热链球菌 CNRZ1066]
假定蛋白 str0657 [嗜热链球菌 LMD-9]
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 [脾泼食子酸链球菌 UCN34]
CRISPR-相关蛋白, Csn1 家族 [巴氏链球菌 ATCC 43144]
假定蛋白 lpp0160 [嗜肺军团菌 str. Paris]
假定蛋白 WS1613 [Wolenella succnogenes DSM 1740]
CRISPR-相关的 大 蛋白 (暂无), 推测的 [产琥珀酸 HTCC5015]
假定蛋白 HMPREF0464_01128 [Sutterella wadsworthensis 3_1_45B]
CRISPR-相关蛋白, Csx12 家族 [Burkholderiales bacterium 1_1_47]
推测的 CRISPR-相关蛋白, Csx12 家族 [Parasutterella excrementihominis YIT 11859]
CRISPR-相关蛋白, Csn1 家族 [Catenibacterium mitsuokai DSM 15897]
假定蛋白 FMG_0058 [大杀戈尔德菌 ATCC 29328]
CRISPR-相关蛋白, Csn1 家族 [Filifactor alocis ATCC35896]
假定蛋白 [真核微杆菌亚种 vincentii ATCC 49256]
CRISPR-相关蛋白, Csn1 家族 [麻疹亲生球菌 ATCC 10379]
CRISPR-相关蛋白, Csn1 家族 [溶血亲生球菌 M424]
csn1 家族 CRISPR-相关蛋白, Csn1 家族 [嗜酸菌属物种 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 单核球菌 ACS-134-V-Col7a]
CRISPR-相关蛋白, Csn1 家族 [非典型 单核菌 ATCC 51204]
CRISPR-相关蛋白, Csn1 家族 [Solobacterium moorei F0204]
CRISPR-相关蛋白, Csn1 家族 [梭菌属球菌属物种D21]
crispr-相关蛋白 [Acidaminococcus intestini RyC-MR95]
假定蛋白 RUMLAC_00178 [Ruminococcus lactaris ATCC 29176]
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 Cas5c [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 [嗜盐乳杆菌乳酸V 11V1-d]

图13D

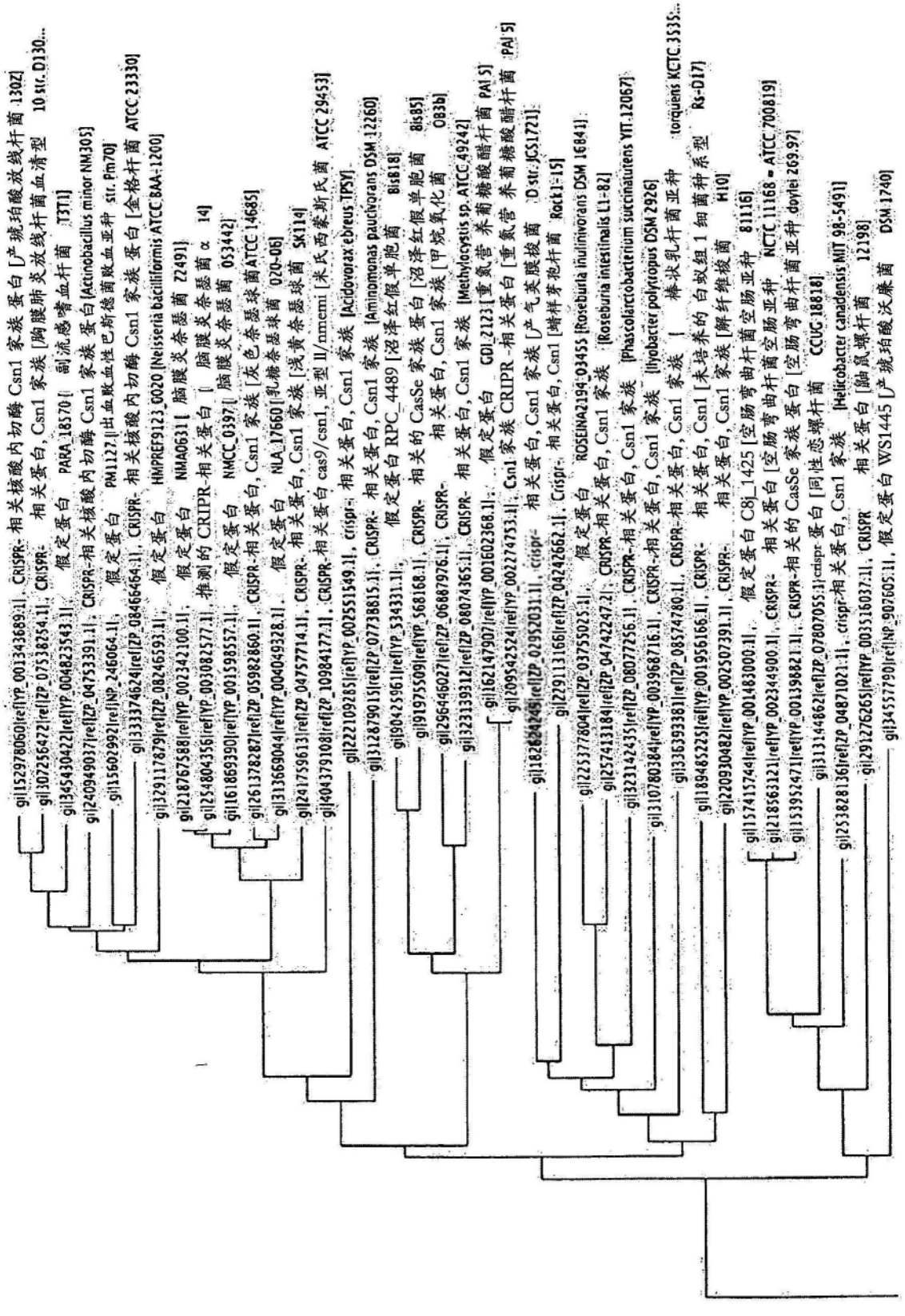


图14A

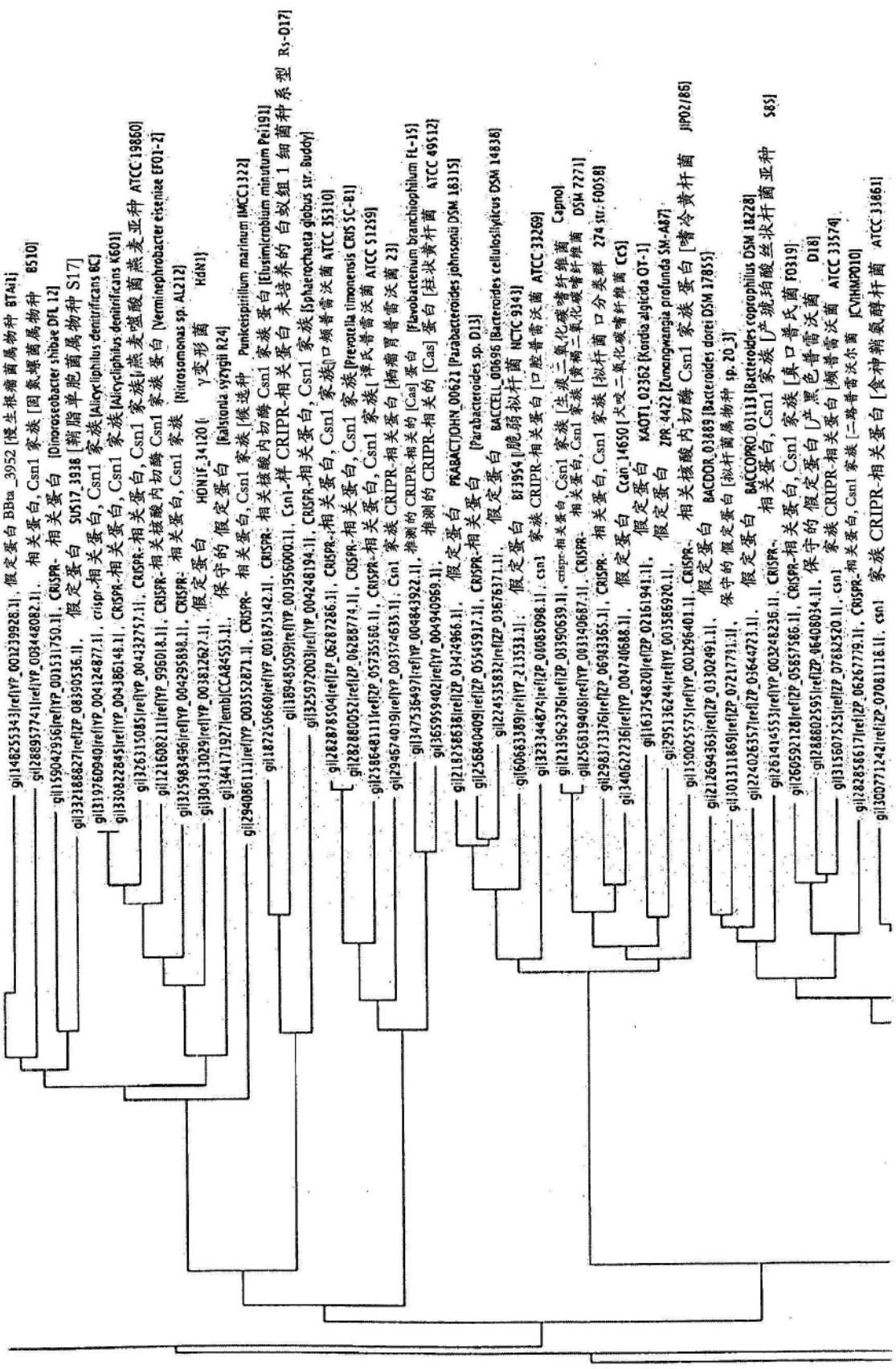

gil|1462553431|ref|YP_001239928.1|. 假定蛋白 BBta_3952 [慢生根瘤菌属物种 BTAi1]
gil|288957741|ref|YP_003448082.1|. 相关蛋白, Csn1 家族 [固氮螺菌属物种 B510]
gil|159047956|ref|YP_001131750.1|. CRISPR- 相关蛋白 [Dinoroseobacter shibae DFL 12]
gil|332188827|ref|ZP_08390536.1|. 假定蛋白 SUS17_3938 [精脂单胞菌属物种 S17]
gil|319769940|ref|YP_004124877.1|. crispr-相关蛋白, Csn1 家族 [Alicycliphilus denitrificans BC]
gil|330822845|ref|YP_004386148.1|. CRISPR-相关蛋白, Csn1 家族 [Alicycliphilus denitrificans K601]
gil|326311085|ref|YP_004232757.1|. CRISPR-相关蛋白, Csn1 家族 [燕麦噬酸菌燕麦亚种 ATCC 19860]
gil|121608211|ref|YP_996018.1|. CRISPR-相关核酸内切酶 Csn1 家族蛋白 [Verminephrobacter eiseniae EF01-2]
gil|325959856|ref|YP_004295898.1|. CRISPR- 相关蛋白, Csn1 家族 [Nitrosomonas sp. AL212]
gil|304313029|ref|YP_003812627.1|. 假定蛋白 HDN1F_34120 [γ变形菌 HdN1]
gil|344171927|emb|CCA84553.1|. 保守的假定蛋白 [Ralstonia syzygii R24]
gil|294086111|ref|YP_003552871.1|. CRISPR- 相关蛋白, Csn1 家族 [候选种 Puniceispirillum marinum IMCC1322]
gil|187250660|ref|YP_001875142.1|. CRISPR- 相关核酸内切酶 Csn1 家族蛋白 [Elusimicrobium minutum Pei191]
gil|189485059|ref|YP_001956000.1|. Csn1-样 CRIPR-相关蛋白 未培养的白蚁组 1 细菌种系型 Rs-D17]
gil|325972003|ref|YP_004248194.1|. CRISPR-相关蛋白, Csn1 家族 [Sphaerochaeta globus str. Buddy]
gil|282878304|ref|ZP_06287286.1|. CRISPR-相关蛋白, Csn1 家族 [口频普雷沃菌 ATCC 35310]
gil|282880052|ref|ZP_06288774.1|. CRISPR-相关蛋白, Csn1 家族 [Prevotella timonensis CRIS 5C-B1]
gil|258648111|ref|ZP_05735560.1|. CRISPR-相关蛋白, Csn1 家族 [拜氏普雷沃菌 ATCC 51129]
gil|294674019|ref|YP_003574635.1|. Csn1 家族 CRIPR-相关蛋白 [肠瘤胃普雷沃菌 23]
gil|347536497|ref|YP_004843922.1|. 推测的 CRIPR-相关的 [Cas] 蛋白 [Flavobacterium branchiophilum FL-15]
gil|365959402|ref|YP_004940969.1|. 推测的 CRIPR-相关的 [Cas] 蛋白 [柱状黄杆菌 ATCC 49512]
gil|215258638|ref|ZP_03474966.1|. 假定蛋白 PRABACTJOHN_00621 [Parabacteroides johnsonii DSM 18315]
gil|256840409|ref|ZP_05545917.1|. CRISPR-相关蛋白 [Parabacteroides sp. D13]
gil|224535832|ref|ZP_03676371.1|. 假定蛋白 BACCELL_00695 [Bacteroides cellulosilyticus DSM 14838]
gil|606683389|ref|YP_213533.1|. 假定蛋白 BF3954 [脆弱拟杆菌 NCTC 9343]
gil|323344874|ref|ZP_08085998.1|. csn1 家族 CRIPR-相关蛋白 [口腔普雷沃菌 ATCC 33269]
gil|213962376|ref|ZP_03390639.1|. crispr-相关蛋白, Csn1 家族 [生澳三氧化碳嗜纤维菌 Capno]
gil|256619408|ref|YP_003140687.1|. CRISPR- 相关蛋白, Csn1 家族 [黄褐三氧化碳嗜纤维菌 DSM 7271]
gil|298373376|ref|ZP_06983365.1|. CRISPR- 相关蛋白, Csn1 家族 [拟杆菌 口分类群 274 str. F0058]
gil|340622236|ref|YP_004740688.1|. 假定蛋白 Ccan_14650 [犬咬二氧化碳嗜纤维菌 CC5]
gil|163754820|ref|ZP_02161941.1|. 假定蛋白 KAOT1_02362 [Kordia algicida OT-1]
gil|295136244|ref|YP_003586920.1|. 假定蛋白 ZMR_4422 [Zymomonas mobilis subsp. mobilis str. ZM4 = ATCC 31821]
gil|115802557|ref|YP_001296401.1|. CRISPR- 相关核酸内切酶 Csn1 家族蛋白 [嗜冷黄杆菌 JIP02/86]
gil|212694363|ref|ZP_03302491.1|. 假定蛋白 BACDOR_03889 [Bacteroides dorei DSM 17855]
gil|301311869|ref|ZP_07217791.1|. 保守的假定蛋白 [拟杆菌属物种 sp. 20_3]
gil|224026357|ref|ZP_03644723.1|. 假定蛋白 BACCOPRO_03113 [Bacteroides coprophilus DSM 18228]
gil|261414553|ref|YP_003248236.1|. CRISPR- 相关蛋白, Csn1 家族 [产琥珀酸丝状杆菌亚种 S85]
gil|260592128|ref|ZP_05857366.1|. CRISPR-相关蛋白, Csn1 家族 [真口普氏菌 F0319]
gil|288802595|ref|ZP_06408034.1|. 保守的假定蛋白 [产黑色普雷沃菌 D18]
gil|315607525|ref|ZP_07882520.1|. csn1 家族 CRIPR-相关蛋白 [频普雷沃菌 ATCC 33574]
gil|282856617|ref|ZP_06267778.1|. CRISPR-相关蛋白, Csn1 家族 [二路普雷沃尔菌 JCVIHMP010]
gil|300771242|ref|ZP_07081116.1|. csn1 家族 CRIPR-相关蛋白 [食神鞘氨醇杆菌 ATCC 33661]

图14B

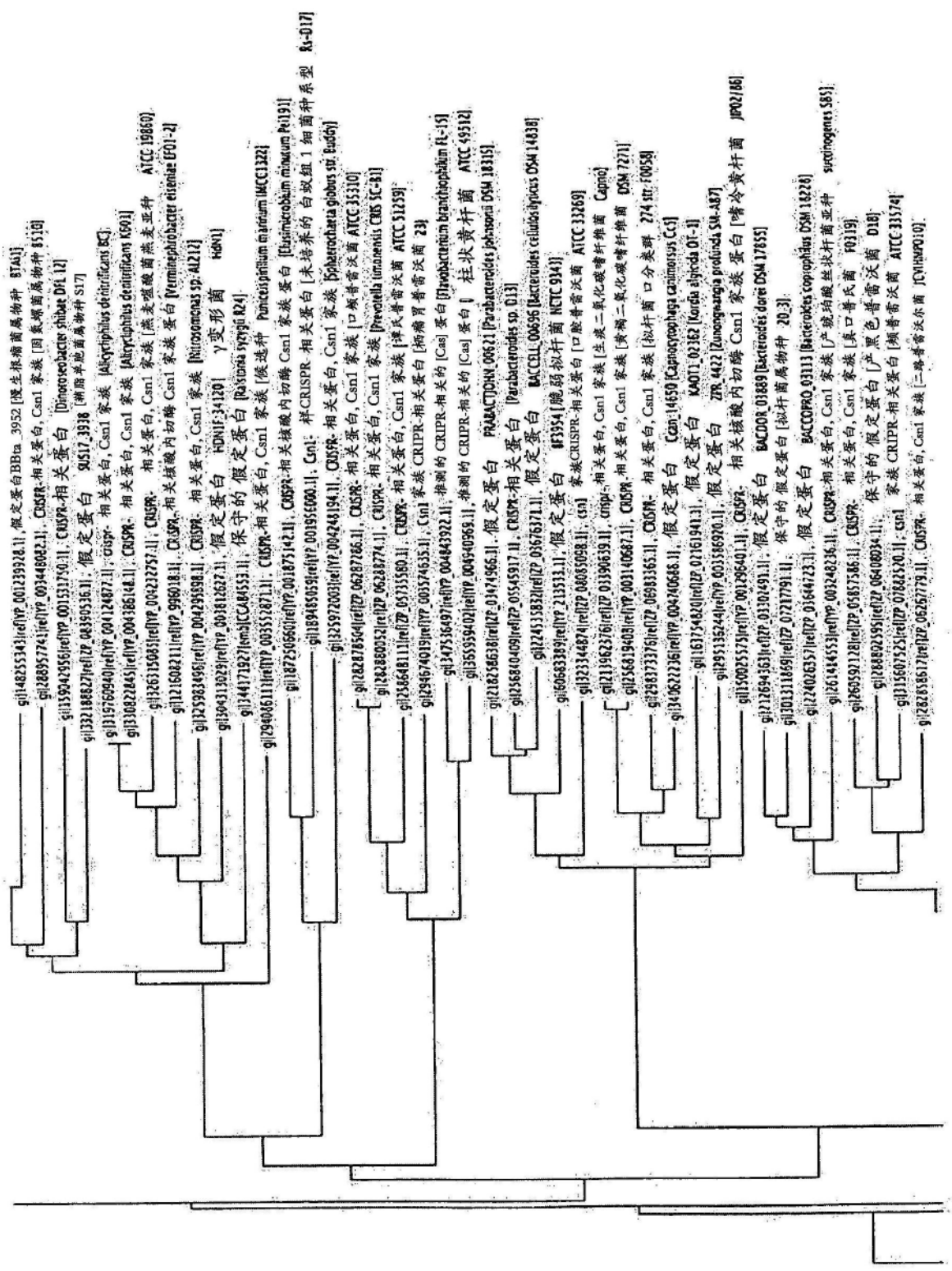


图14C

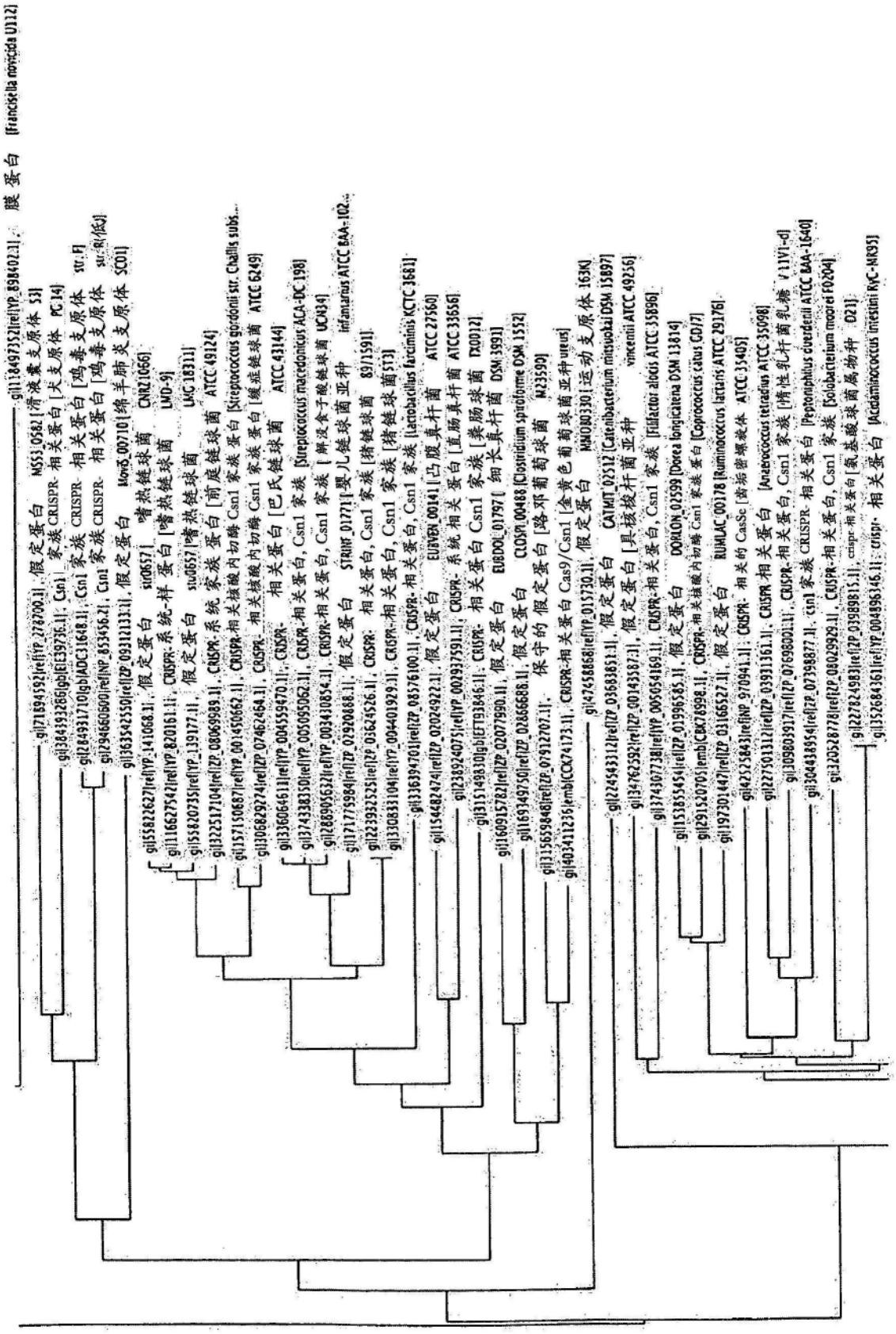


图14D

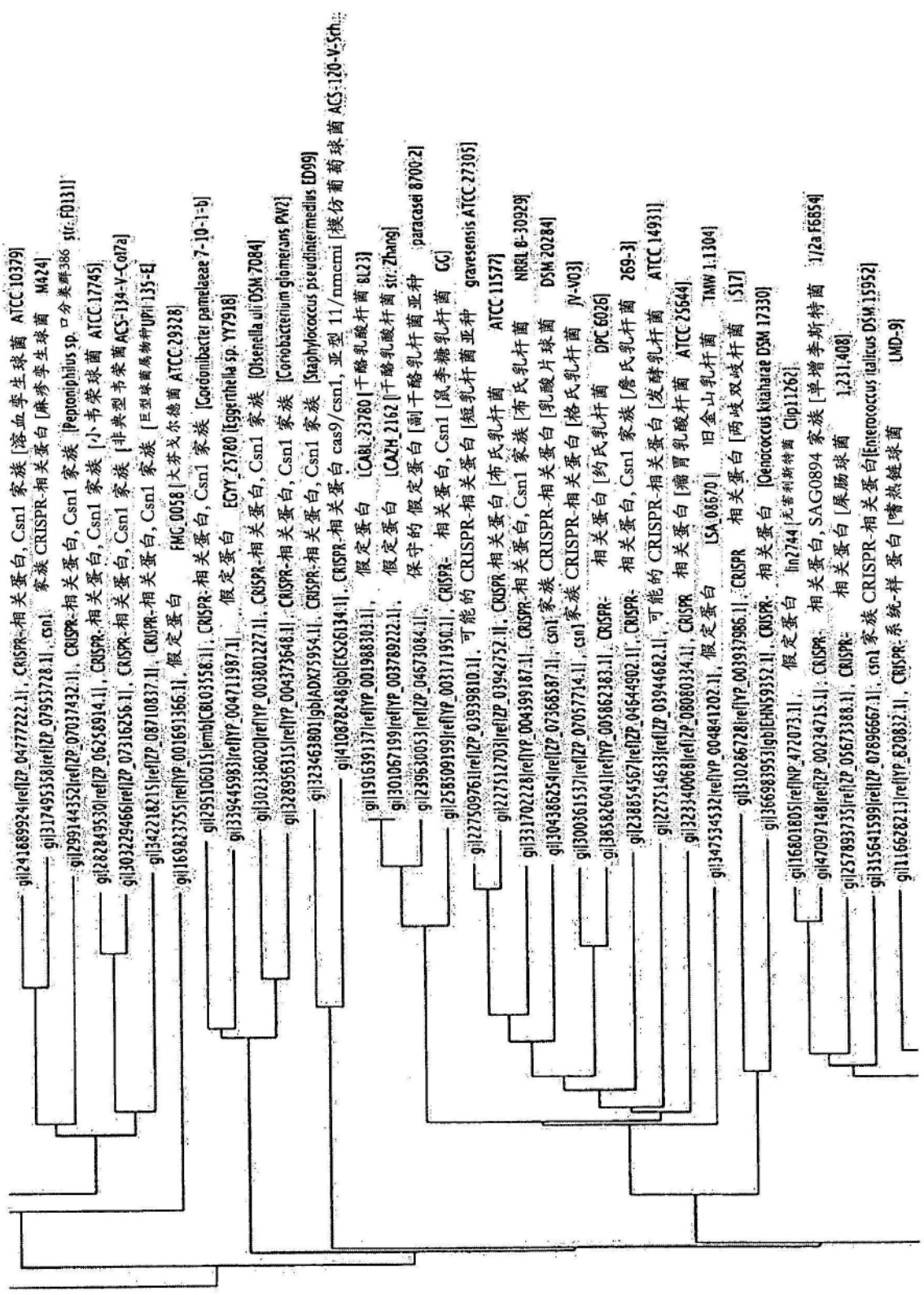


图14E

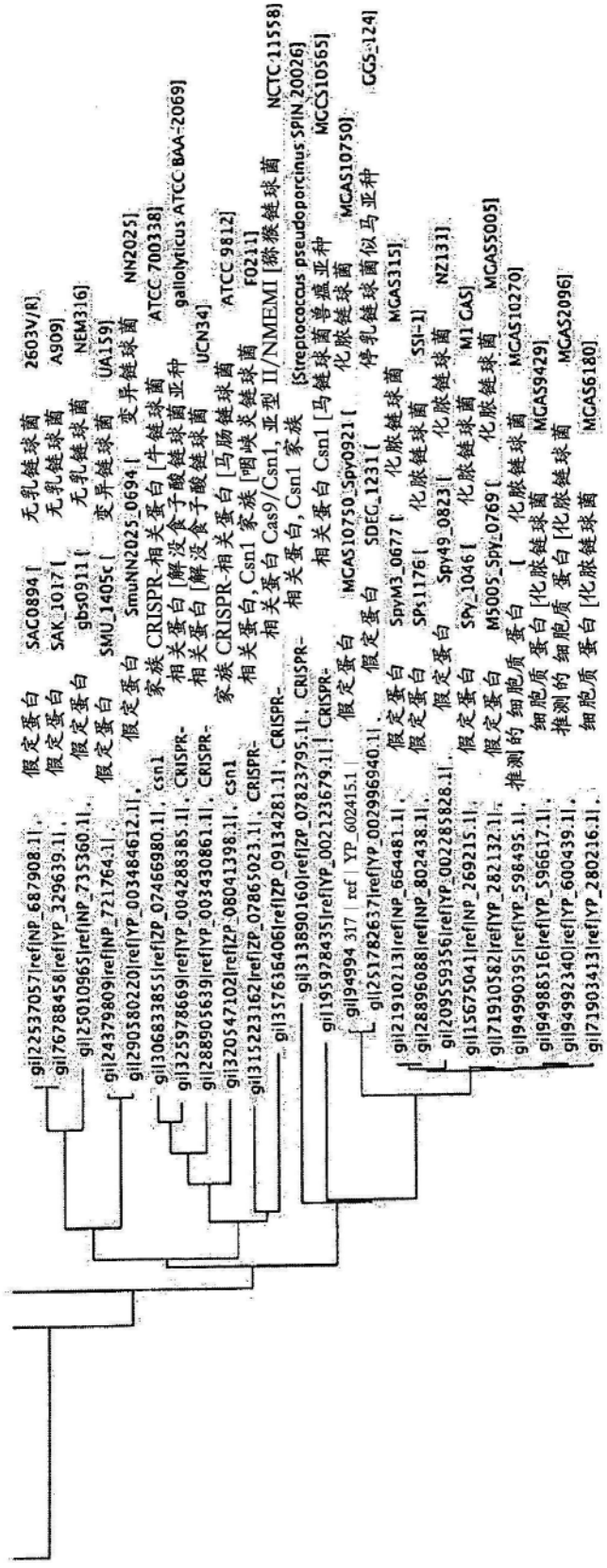

gi|22537057|ref|NP_687908.1|
gi|76788456|ref|YP_329639.1|
gi|25010965|ref|NP_735360.1|
gi|24379809|ref|NP_721764.1|
gi|29058022O|ref|YP_003484612.1|
gi|30683385S|ref|ZP_07466980.1|
gi|32597866g|ref|YP_004288385.1|
gi|28890563g|ref|YP_003430861.1|
gi|32054710Z|ref|ZP_08041398.1|
gi|31522316Z|ref|ZP_07865023.1|
gi|35763640G|ref|ZP_09134281.1|
gi|31389016O|ref|ZP_07823795.1|
gi|19597843S|ref|YP_002123679.1|
gi|94994 317 | ref | YP_602415.1 |
gi|25178263 7|ref|YP_002996940.1|
gi|21910213|ref|NP_664481.1|
gi|28896088|ref|NP_802438.1|
gi|20955935G|ref|YP_002285828.1|
gi|15675041|ref|NP_269215.1|
gi|71910582|ref|YP_282132.1|
gi|19990395|ref|YP_598495.1|
gi|19498851G|ref|YP_596617.1|
gi|19499234O|ref|YP_600439.1|
gi|71903413|ref|YP_280216.1|
假定蛋白
假定蛋白
假定蛋白
假定蛋白
假定蛋白 家族 CRISPR-相关蛋白 [牛链球菌
csn1 相关蛋白 [解没食子酸链球菌亚种
CRISPR- 相关蛋白 [解没食子酸链球菌
CRISPR- 家族 CRISPR-相关蛋白 [马肠链球菌
csn1 相关蛋白,Csn1 家族 [咽峡炎链球菌
CRISPR- 相关蛋白 Cas9/Csn1, 亚型 II/NMEM [猕猴链球菌
CRISPR- 相关蛋白, Csn1 家族 [Streptococcus pseudoporcinus SPIN 20026]
CRISPR- 相关蛋白 Csn1 [马链球菌兽瘟亚种
假定蛋白 MGAS10750_Spy0921 [化脓链球菌 MGAS10750]
假定蛋白 SDEG_1231 [停乳链球菌似马亚种 GGS_124]
假定蛋白 SpyM3_0677 [化脓链球菌 MGAS315]
假定蛋白 SPs1176 [化脓链球菌 SSI-1]
假定蛋白 Spy49_0823 [化脓链球菌 NZ131]
假定蛋白 SPy_1046 [化脓链球菌 M1 GAS]
假定蛋白 M5005_Spy_0769 [化脓链球菌 MGAS5005]
推测的细胞质蛋白 [化脓链球菌 MGAS10270]
细胞质蛋白 [化脓链球菌 MGAS9429]
推测的细胞质蛋白 [化脓链球菌 MGAS2096]
细胞质蛋白 [化脓链球菌 MGAS6180]
2603V/R]
A909]
NEM316]
UA159]
变异链球菌 NN2025]
ATCC 700338]
gallolyticus ATCC BAA-2069]
UCN34]
ATCC 9812]
FO211]
NCTC 11558]
MGCS10565]

图14F

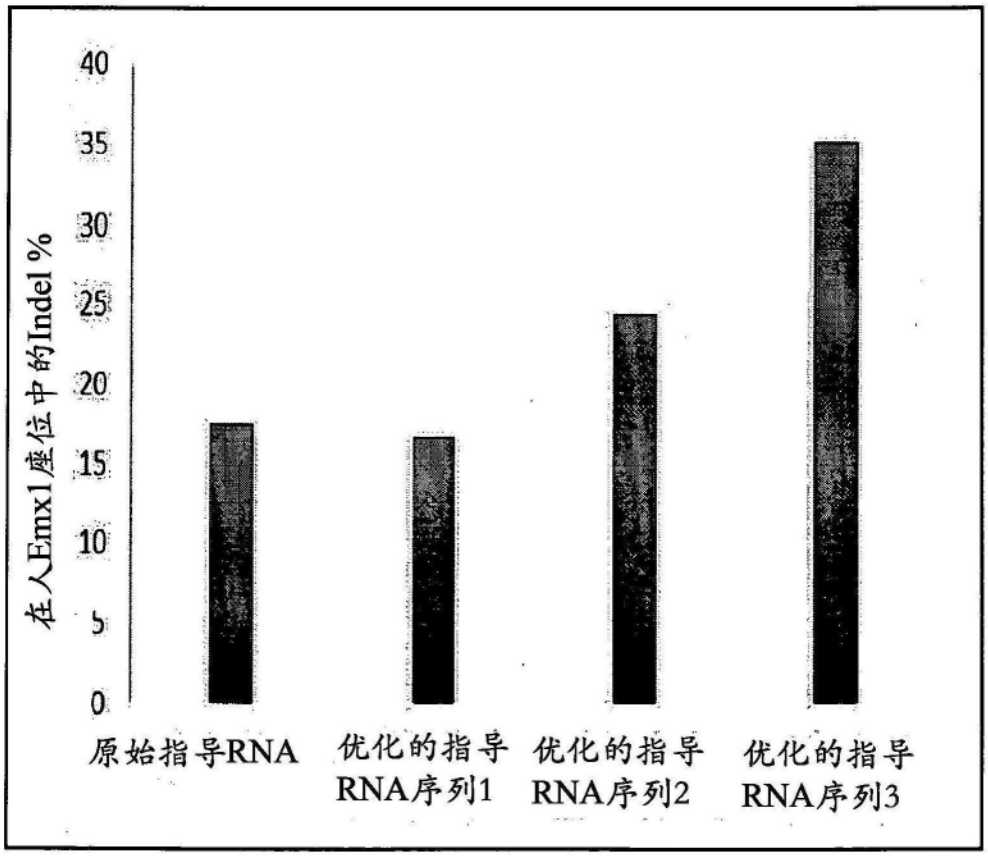
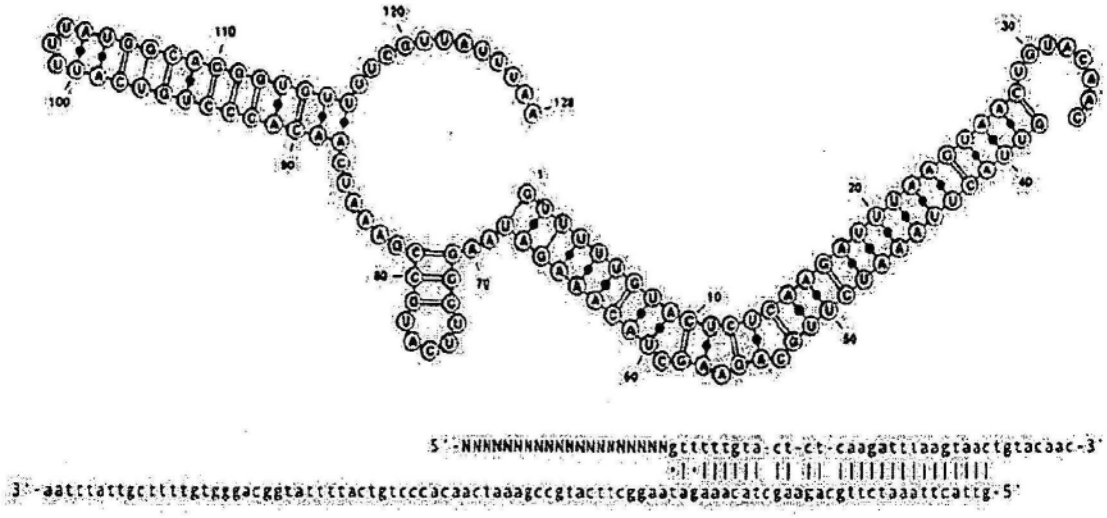

在人Emx1座位中的Indel %
40
35
30
25
20
15
10
5
0
原始指导RNA
优化的指导RNA序列1
优化的指导RNA序列2
优化的指导RNA序列3

图15

带有同向重复的tracrRNA的嗜热链球菌CRISPR1 RNA共折叠物



嵌合指导RNA设计

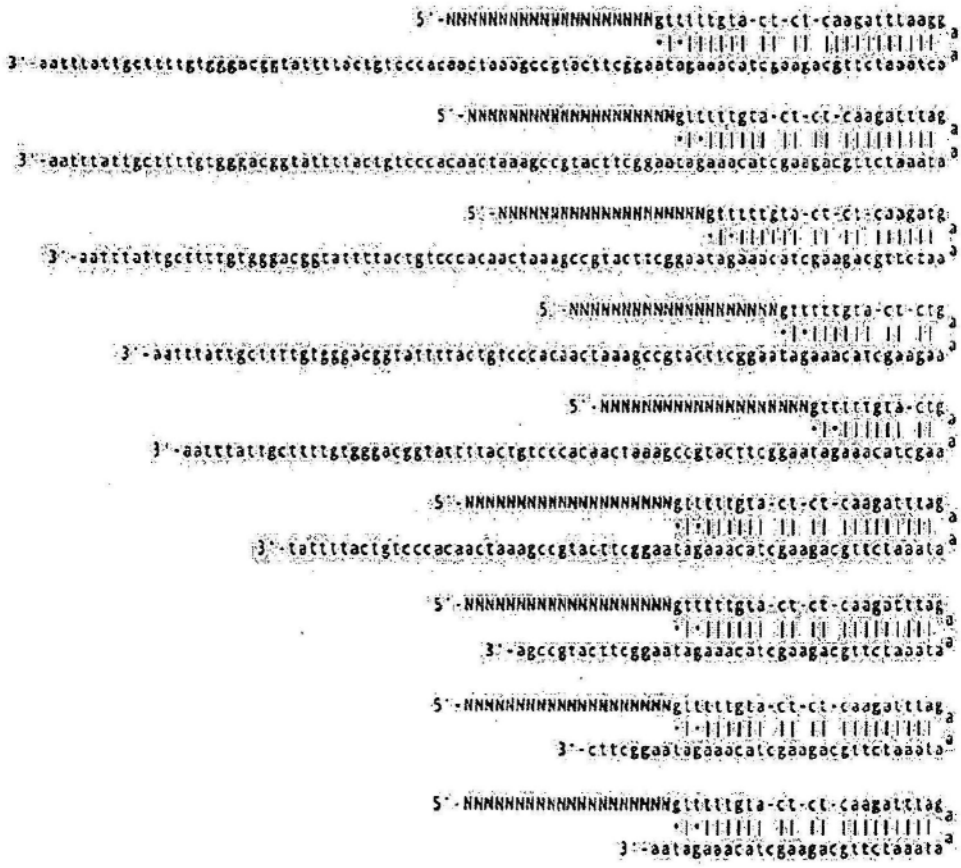


图16

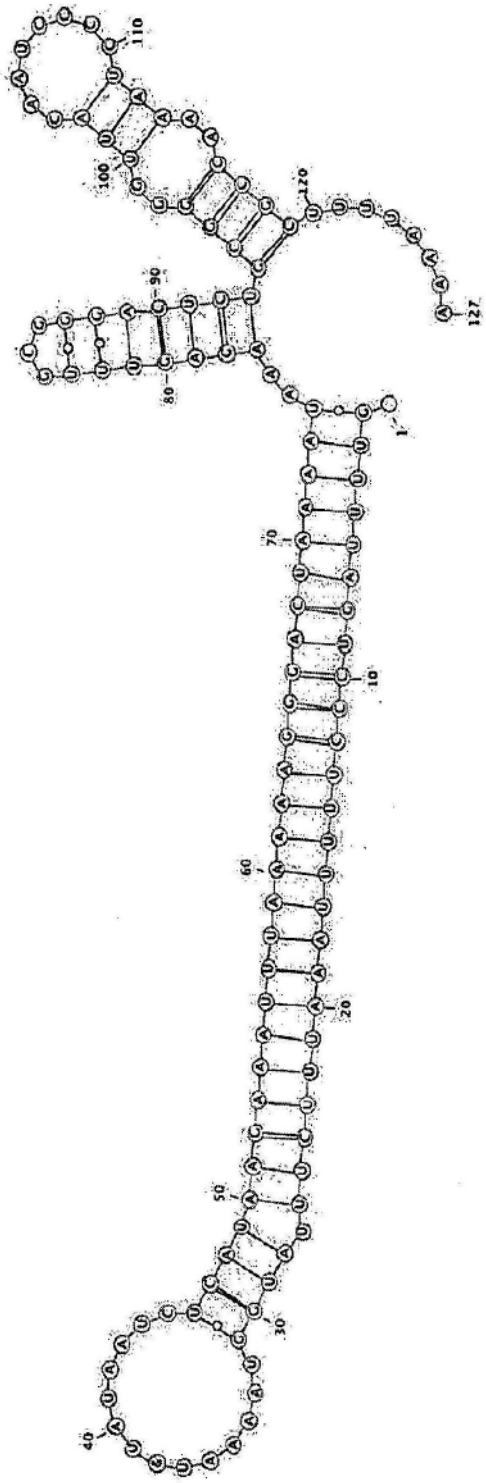
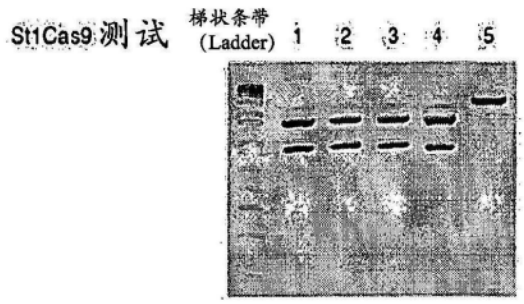


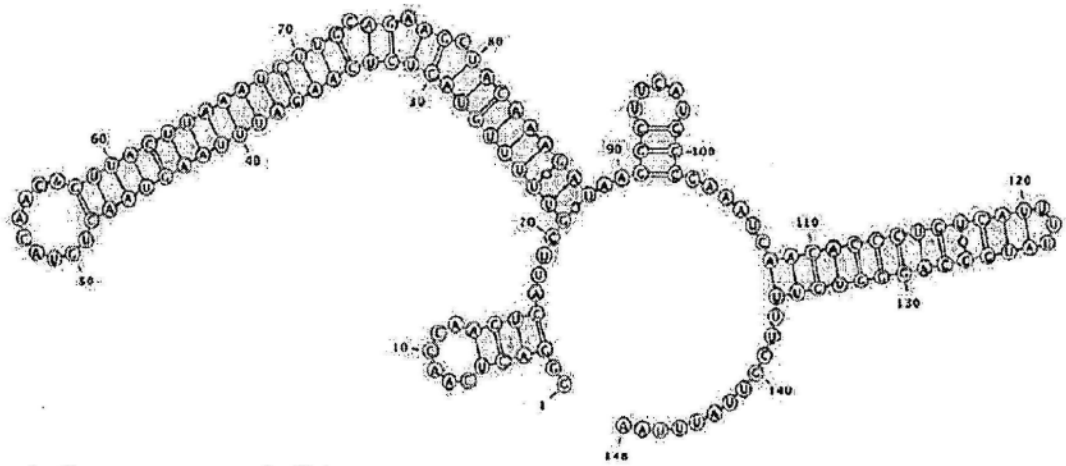
图17



泳道1：全tracrRNA + 全间隔子DR

全 tracrRNA: GUUACUUAAAUCUUGCAGAAGCUACAAAGAUAAGGCUUCAUGCCGAAAUCAACACCCUGUCAUUUUAUGGCAGGGUGUUUUCGUUAUUUAA

全间隔子DR :GGGACUCAACCAAGUCAUUCGUUUUUGUACUCUCAAGAUUUAAGUAACUGUACAAC



泳道2：成熟tracrRNA + 成熟间隔子DR

成熟 tracrRNA: CUUGCAGAAGCUACAAAGAUAAGGCUUCAUGCCGAAAUCAACACCCUGUCAUUUUAUGGCAGGGUGUUUU

成熟间隔子DR :GGGACUCAACCAAGUCAUUCGUUUUUGUACUCUCAAGAUUUA

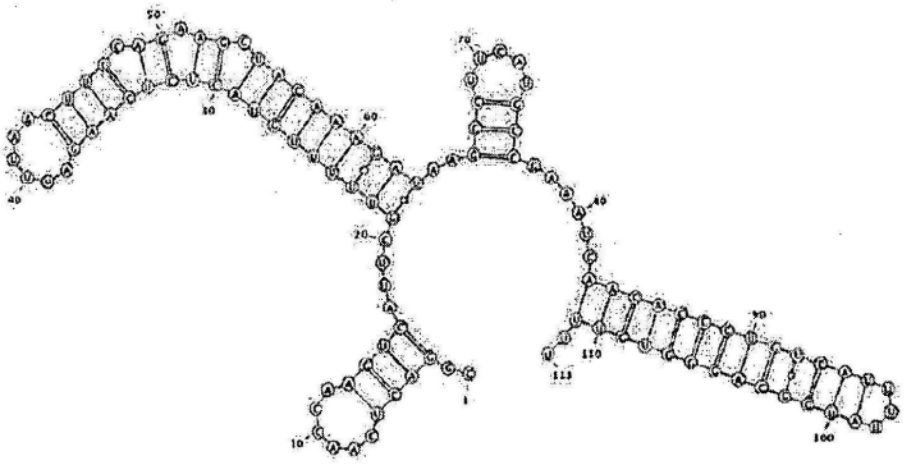
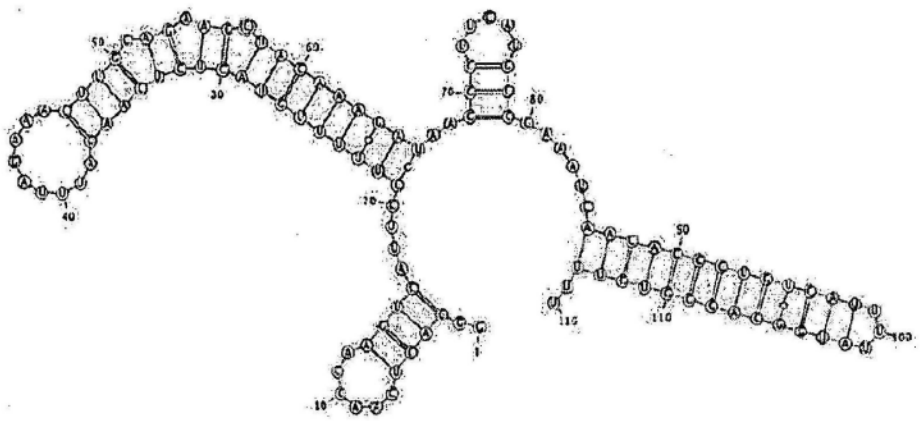


图18A

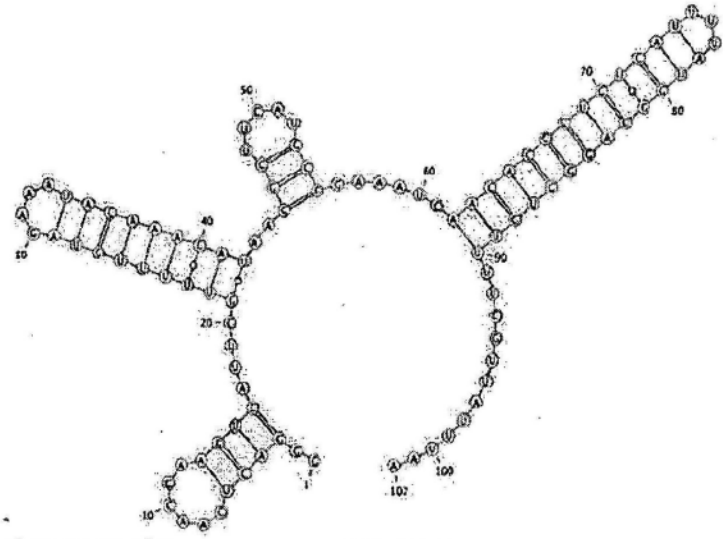
泳道3：嵌合指导RNA设计1 (1-22 DR, 12-81 tracrRNA)

序列 GGGACUCAACCAAGUCAUUCGUUUUUGUACUCUCAAGAUUUAGAAACUUGCAGAAGCUACAAAGAUAAGGCUUCAUGCCGAAAUCAACACCCUGUCAUUUUAUGGCAGGGUGUUUU



泳道4：嵌合指导RNA设计2 (1-9 DR, 20-91 tracrRNA)

序列 GGGACUCAACCAAGUCAUUCGUUUUUGUAGAAAUACAAAGAUAAGGCUUCAUGCCGAAAUCAACACCCUGUCAUUUUAUGGCAGGGUGUUUUCGUUAUUUAA



泳道5：嵌合指导RNA设计3 (1-9 DR, 20-46 tracrRNA)

序列 GGGACUCAACCAAGUCAUUCGUUUUUGUAGAAAUACAAAGAUAAGGCUUCAUGCCGA

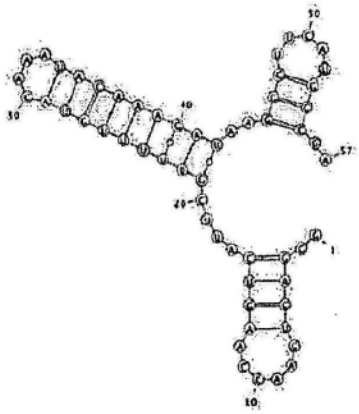


图18B

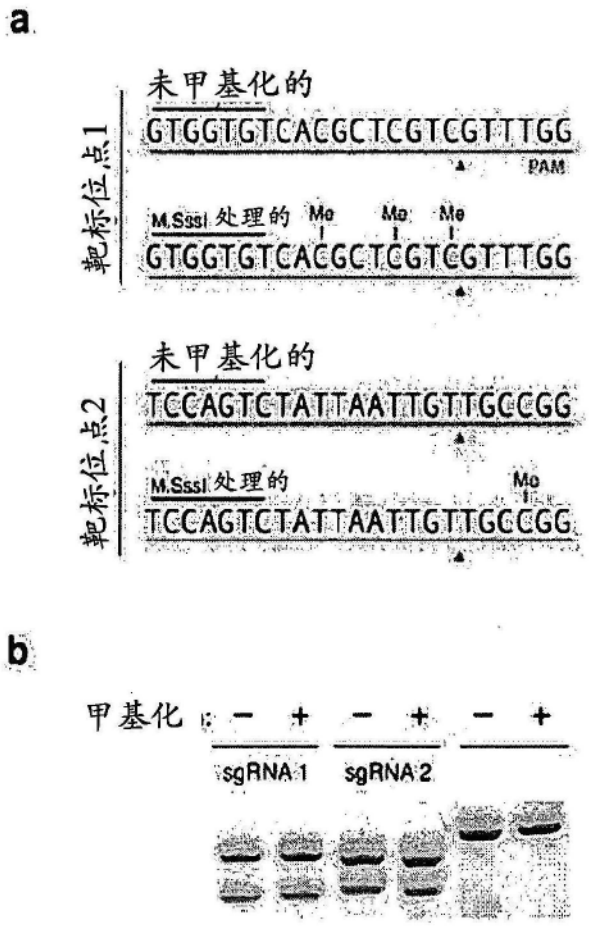

a
靶标位点1
未甲基化的
GTGGTGTCACGCTCGTCGTTTGG
PAM
M.SssI 处理的
Me
Me
Me
GTGGTGTCACGCTCGTCGTTTGG
靶标位点2
未甲基化的
TCCAGTCTATTAATTGTTGCCGG
M.SssI 处理的
Me
TCCAGTCTATTAATTGTTGCCGG
b
甲基化
−
+
−
+
−
+
sgRNA 1
sgRNA 2

图19A-B

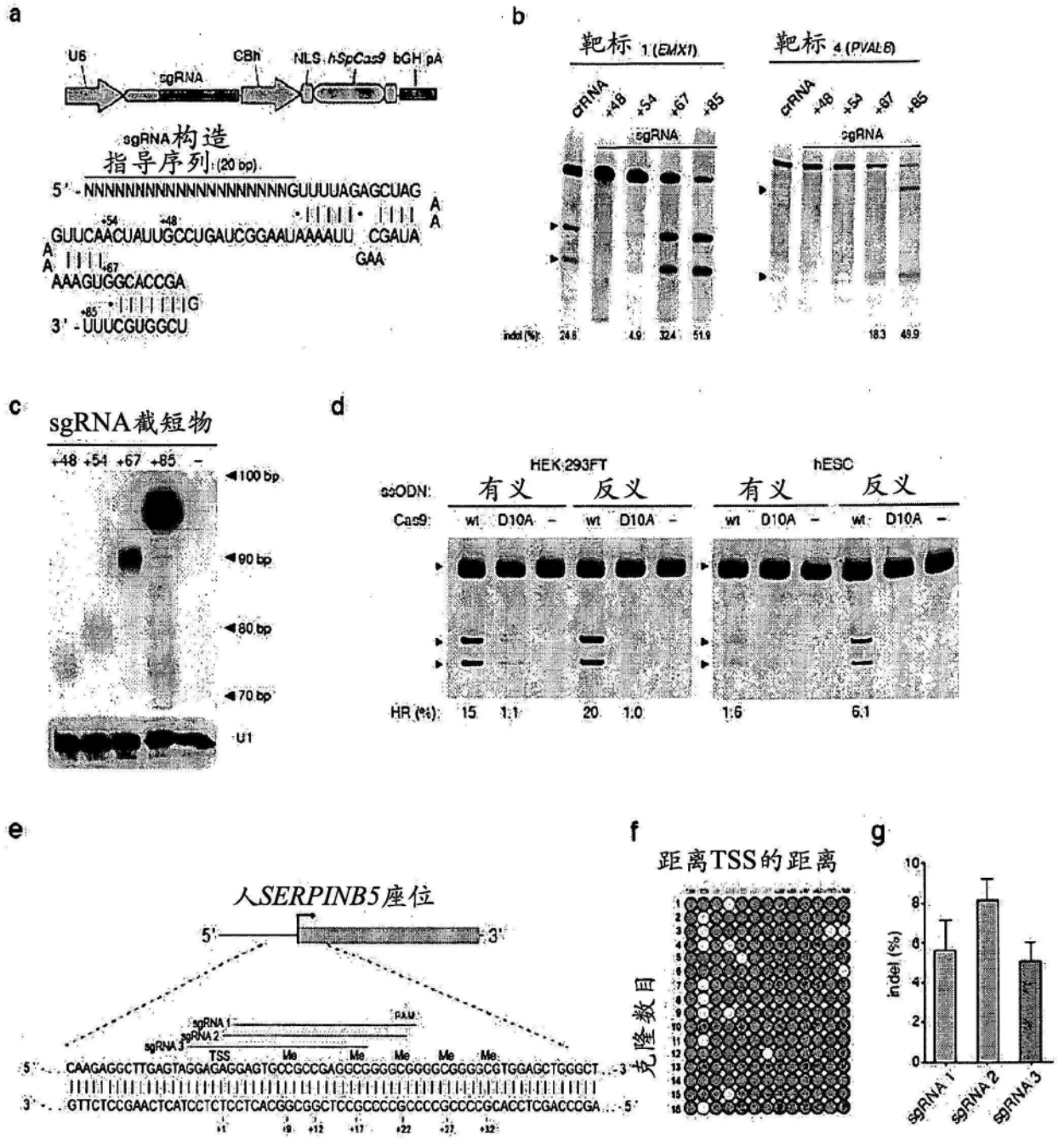

a
U6
CBh
NLS
hSpCas9
bGH pA
sgRNA
sgRNA构造
指导序列(20 bp)
5' - NNNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAG
GUUCAACUAUUGCCUGAUCGGAAUAAAAUU CGAUA
AAAGUGGCACCGA
3' - UUUCGUGGCU
b
靶标 1 (EMX1)
靶标 4 (PVALB)
cRNA
+48
+54
+67
+85
sgRNA
indel (%):
24.8
4.9
32.4
51.9
18.3
49.9
c
sgRNA截短物
+48 +54 +67 +85 –
100 bp
90 bp
80 bp
70 bp
U1
d
HEK 293FT
hESC
ssODN:
有义
反义
Cas9:
wt
D10A
–
HR (%):
15
1.1
20
1.0
1.6
6.1
e
人SERPINB5座位
5'
3'
sgRNA 1
sgRNA 2
sgRNA 3
PAM
TSS
Me
5' ... CAAGAGGCTTGAGTAGGAGAGGAGTGCCGCCGAGGCGGGGCGGGGCGGGGCGTGGAGCTGGGCT ... 3'
3' ... GTTCTCCGAACTCATCCTCTCCTCACGGCGGCTCCGCCCCGCCCCGCCCCGCACCTCGACCCGA ... 5'
f
距离TSS的距离
克隆数目
g
indel (%)
sgRNA 1
sgRNA 2
sgRNA 3

图20A-G

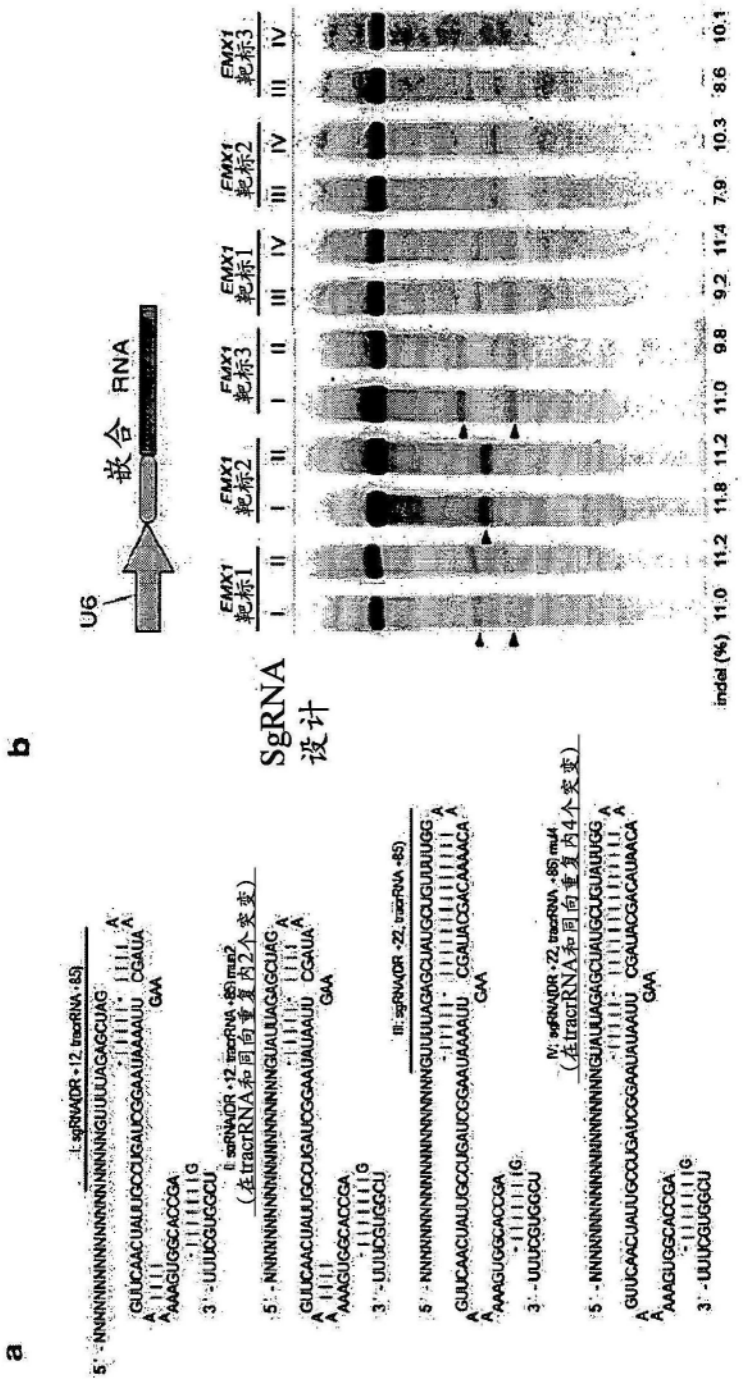

a
b
SgRNA
设计
U6
嵌合 RNA
EMX1
靶标1
EMX1
靶标2
EMX1
靶标3
EMX1
靶标1
EMX1
靶标2
EMX1
靶标3
I
II
I
II
I
II
III
IV
III
IV
III
IV
indel (%)
11.0
11.2
11.8
11.2
11.0
9.8
9.2
11.4
7.9
10.3
8.6
10.1

图21A-B

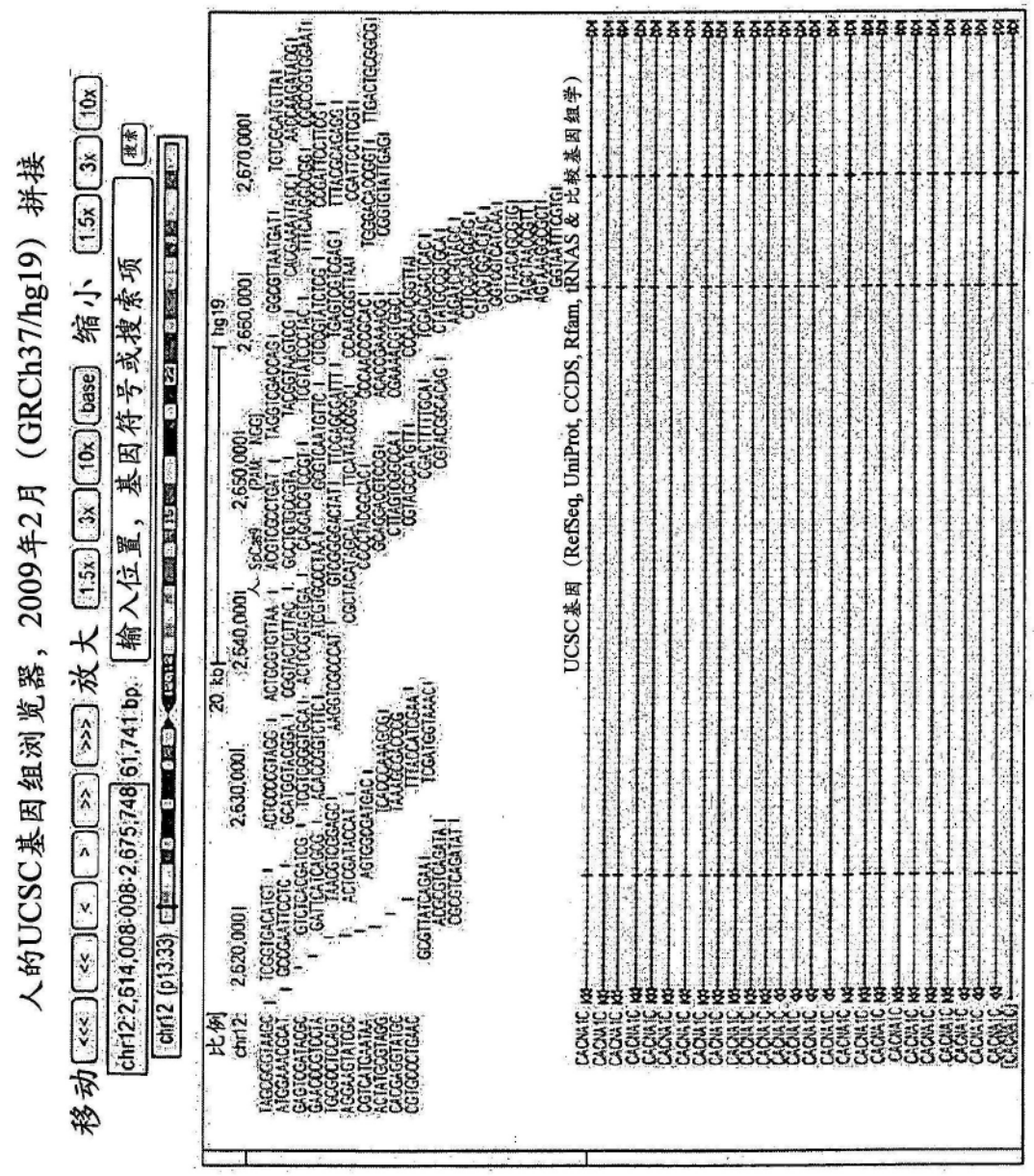

人的UCSC基因组浏览器，2009年2月（GRCh37/hg19）拼接
移动 <<< << < > >> >>> 放大 1.5x 3x 10x base 缩小 1.5x 3x 10x
chr12:2,614,008-008-2,675,748 61,741 bp.
输入位置，基因符号或搜索项
搜索
chr12 (p13.33)
比例
chr12:
2,620,000
2,630,000
2,640,000
20 kb
hg19
2,650,000
2,660,000
2,670,000
UCSC基因 (RefSeq, UniProt, CCDS, Rfam, tRNAS & 比较基因组学)

图22

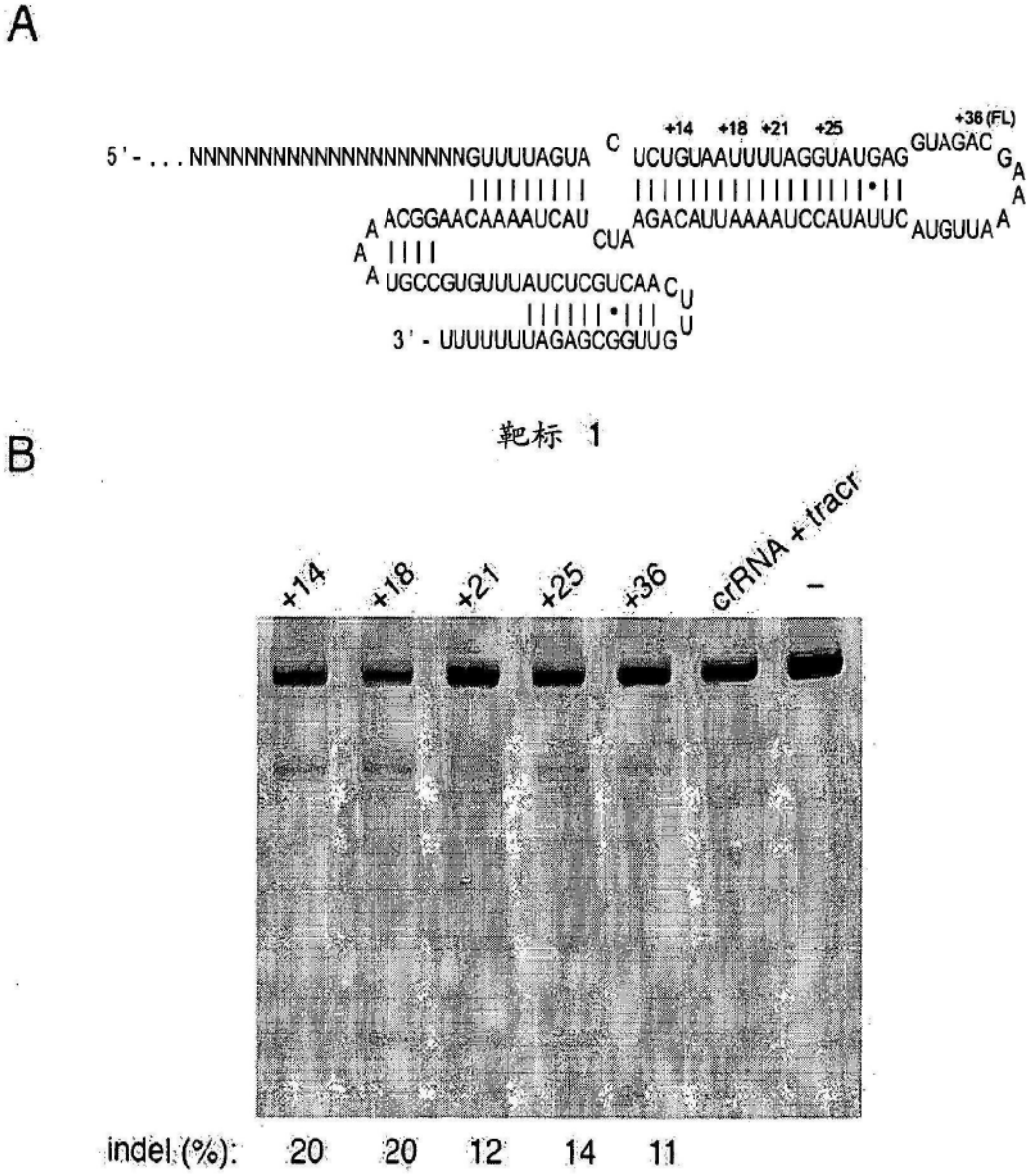

A
+14 +18 +21 +25 +36(FL)
5'-...NNNNNNNNNNNNNNNNNNNNGUUUUAGUA C UCUGUAAUUUUAGGUAUGAG GUAGAC G A
A ACGGAACAAAAUCAU CUA AGACAUUAAAAUCCAUAUUC AUGUUA A A
A UGCCGUGUUUAUCUCGUCAA C U
3'-UUUUUUUAGAGCGGUUG U
B
靶标 1
+14 +18 +21 +25 +36 crRNA + tracr –
indel (%): 20 20 12 14 11

图23A-B