



(21) 申请号 202310155019.2

C12N 9/22 (2006.01)

(22) 申请日 2017.06.08

(30) 优先权数据

62/347,553 2016.06.08 US

15/493,129 2017.04.20 US

(62) 分案原申请数据

201780049055.3 2017.06.08

(71) 申请人 安捷伦科技有限公司

地址 美国加利福尼亚州

(72) 发明人 D·J·德林杰 D·E·赖安

S·罗伊 J·R·桑普森

(74) 专利代理机构 北京坤瑞律师事务所 11494

专利代理师 封新琴

(51) Int. Cl.

C12N 15/113 (2010.01)

权利要求书1页 说明书119页

序列表(电子公布) 附图25页

(54) 发明名称

使用化学修饰的引导RNA的高特异性基因组  
编辑

(57) 摘要

本发明涉及具有化学修饰的引导RNA及其在CRISPR-Cas系统中的用途。所述化学修饰的引导RNA对靶多核苷酸序列具有增强的特异性。本发明还涉及使用化学修饰的引导RNA切割或切刻多核苷酸、以及用于高特异性基因组编辑的方法。

1. 一种合成的引导RNA,其包含:

(a) crRNA区段,其包含(i)能够与靶多核苷酸杂交的引导序列,其中所述靶多核苷酸包含邻近PAM位点的靶序列,(ii)茎序列;和

(b) tracrRNA区段,其包含与所述茎序列部分或完全互补的核苷酸序列,其中所述引导序列由20-N个核苷酸组成,其中N是-10至6之间的整数;其中所述引导序列进一步包含位于所述引导序列中从位置4-N至20-N的任何位置处的至少一个修饰。

2. 权利要求1的合成引导RNA,其中所述引导RNA是单引导RNA(sgRNA)。

3. 一种合成的crRNA,其包含能够与靶多核苷酸杂交的引导序列,所述靶多核苷酸包含邻近PAM位点的靶序列,其中所述引导序列由20-N个核苷酸组成,其中N是-10和6之间的整数;其中引导序列包括位于所述引导序列中从位置4-N至20-N的任何位置处的至少一个修饰。

4. 前述权利要求中任一项的合成的引导RNA或crRNA,其中所述至少一个修饰位于位置4-N,5-N,7-N,9-N,10-N,11-N,14-N,或16-N处。

5. 前述权利要求中任一项的合成的引导RNA或crRNA,其中所述至少一个修饰位于引导序列的位置5-N或11-N,或其组合处。

6. 前述权利要求中任一项的合成的引导RNA或crRNA,其中所述至少一个修饰是核苷酸间键修饰,且其中位置15-N不包含2'-O-甲基-3'-磷酸基乙酸(MP)或2'-O-甲基-3'-硫代磷酸基乙酸(MSP)。

7. 前述权利要求中任一项的合成的引导RNA或crRNA,其中所述至少一个修饰选自下组:硫代磷酸核苷酸间键、手性硫代磷酸核苷酸间键、二硫代磷酸核苷酸间键、硼烷磷酸核苷酸间键、C<sub>1-4</sub>烷基磷酸核苷酸间键、磷酸基羧酸核苷酸间键、磷酸基羧酸酯核苷酸间键、硫代磷酸基羧酸核苷酸间键、和硫代磷酸基羧酸酯核苷酸间键。

8. 前述权利要求中任一项的合成的引导RNA或crRNA,其中所述至少一个修饰选自下组:磷酸基羧酸核苷酸间键、硫代磷酸基羧酸核苷酸间键、修饰的碱基、赋予C3'-内糖皱褶构型的2'修饰、或其组合。

9. 前述权利要求中任一项的合成的引导RNA或crRNA,其中所述至少一个修饰是磷酸基乙酸核苷酸间键(P)或硫代磷酸基乙酸核苷酸间键(SP)。

10. 权利要求8的合成的引导RNA或crRNA,其中所述2'修饰选自2'-O-甲基、2'-氟和2'-O-(2-甲氧基乙基)。

## 使用化学修饰的引导RNA的高特异性基因组编辑

[0001] 本申请是2017年6月8日提交的申请号为201780049055.3、发明名称为“使用化学修饰的引导RNA的高特异性基因组编辑”的发明专利申请的分案申请。

### 发明领域

[0002] 本发明涉及分子生物学领域。具体而言，本发明涉及成簇规律间隔短回文重复序列(CRISPR)技术。

### 发明背景

[0003] 天然的原核CRISPR-Cas系统包括一系列具有恒定长度的介入可变序列的短重复序列(即，成簇规律间隔短回文重复序列或“CRISPR”)以及CRISPR相关(“Cas”)蛋白。转录的CRISPR系列的RNA被一部分Cas蛋白加工成小的引导RNA，引导RNA通常具有两个组分，如下文所述的。至少存在六种不同的系统：I型、II型、III型、IV型、V型和VI型。将RNA加工为成熟crRNA的过程中涉及的酶在这六种系统中是不同的。在天然原核II型系统中，引导RNA(“gRNA”)包括两种短的非编码RNA，称为CRISPR RNA(“crRNA”)和反式作用RNA(“tracrRNA”)。在一个示例性的系统中，gRNA与Cas核酸酶形成复合物。gRNA:Cas核酸酶复合物结合具有原型间隔序列毗邻基序(“PAM”)和原型间隔序列(protospacer)的靶多核苷酸序列，原型间隔序列是与gRNA的一部分互补的序列。gRNA:Cas核酸酶复合物通过识别和结合靶多核苷酸而诱导靶多核苷酸的切割。天然CRISPR-Cas系统作为原核生物中的免疫系统而发挥功能，其中gRNA:Cas核酸酶复合物以类似于真核生物中的RNAi的方式识别外源遗传元件并使其沉默，由此赋予对外源遗传元件(如质粒和噬菌体)的抗性。

[0004] 已经证明，单引导RNA(“sgRNA”)，其中crRNA和tracrRNA共价连接，可以代替天然存在的crRNA与tracrRNA之间形成的复合物。

[0005] 与开发gRNA(包括sgRNA)相关的考虑因素包括特异性、稳定性和功能性。特异性是指特定的gRNA:Cas核酸酶复合物结合、切刻和/或切割期望的靶标序列、而较少发生或不发生结合、切刻和/或切割在序列和/或位置上不同于期望的靶标的多核苷酸的能力。因此，特异性是指将gRNA:Cas核酸酶复合物的脱靶效应最小化。需要提供这样的gRNA(包括sgRNA)，其具有期望的对靶多核苷酸的结合亲和力，同时具有降低的脱靶效应，但仍然具有期望的gRNA功能性。还需要更好的途径来制造和使用具有增强的特异性、同时具有期望的对靶多核苷酸的结合亲和力和/或降低的脱靶结合的gRNA(包括sgRNA)。

### 附图简述

[0006] 图1是示例性CRISPR-Cas系统的说明。复合物由单引导RNA和Cas蛋白形成，并且所述复合物识别并结合靶多核苷酸。该Cas核酸酶是酿脓链球菌Cas9核酸酶。酿脓链球菌Cas9核酸酶识别PAM序列(此处，PAM序列是3核苷酸序列NGG，其中N是A、G、C或T，但已知存在其他的PAM序列，如NAG等)。sgRNA包括引导序列、crRNA序列或区段、以及tracrRNA序列或区段。sgRNA的引导序列与DNA靶标的PAM序列的上游直接杂交。

[0007] 图2A显示了包含crRNA区段203和tracrRNA区段205的示例性引导RNA 201。图2B显示了包含crRNA区段209和tracrRNA区段211的示例性单引导RNA 207，其中crRNA区段和

tracrRNA区段通过环213连接。

[0008] 图3显示了可以包括在引导RNA中的各种化学修饰的一些的结构;当然,图3并不意味着是限制性的,可以采用本文所述的许多其他修饰。

[0009] 图4是显示当寡核苷酸双链体通过加热分离成单独的链时紫外光(UV)吸收如何增加的曲线图。S形曲线反映了双链体解离成分开的链,S形曲线的中心是双链体的 $T_m$ 。这个曲线指示在生理盐浓度下20个核苷酸的RNA/DNA双链体的解链温度约为50℃。

[0010] 图5A展示了单引导RNA(sgRNA)或两片段双引导RNA(“dgRNA”),后者中crRNA区段和tracrRNA区段形成杂交双链体,并显示了引导RNA的延伸区、锁定区、取样区、种子区(通常为10聚体)、双引导茎或单引导茎环、以及tracrRNA区。图5B展示了在引导序列的种子部分和互补DNA序列之间初始形成种子双链体之后,核苷酸的结合如何依次通过取样区和锁定区形成具有解链温度的RNA/DNA双链体。图5C展示了种子区核苷酸中的化学修饰可以降低各个碱基对的结合能,同时保持高水平的协同性,由此延伸取样区并降低RNA/DNA双链体的解链温度。图5C还展示了取样区中的化学修饰。

[0011] 图6A展示了实验性crRNA多核苷酸,其具有20个核苷酸的引导序列并且通过在引导序列中的不同位置掺入各种类型的化学修饰而被修饰。图6B显示了包含未修饰的crRNA的RNA/DNA双链体的解链曲线。图6C至6F显示了包含化学修饰的crRNA的RNA/DNA双链体的解链曲线,所述crRNA在不同的引导序列中在核苷酸6至9处包含不同类型的修饰。

[0012] 图7是显示与互补DNA链杂交的引导序列(在crRNA中)的20个碱基对的双链体,在引导序列渐进的5'截短或5'延伸(如x轴上负整数或正整数指示)之后,解链温度变化的图。

[0013] 图8A显示了在靶向人CLTA基因中的“CLTA1”序列的gRNA中进行的化学修饰对靶多核苷酸的体外切割[称为“中靶”(“ON”)]的影响、以及另行测定的对两个不同的脱靶多核苷酸[称为“脱靶1”(“OFF1”)和“脱靶3”(“OFF3”)]的切割的影响,分别表示CLTA1中靶(ON-target)、CLTA1脱靶1-靶(OFF1-target)和CLTA1脱靶3-靶(OFF3-target)。

[0014] 图8B是基于图8A中的切割结果,通过计算每个受测的合成sgRNA切割的靶多核苷酸与切割的脱靶多核苷酸的比值而导出的。还计算了将每个比值乘以每个测定的引导RNA的对应中靶切割百分比而获得的“特异性评分”。加阴影的比值和特异性评分的值比图8B中所示的其他值更大,值得注意。

[0015] 图9A显示了在靶向人CLTA基因中的“CLTA4”序列的gRNA中不同位置处的2'-O-甲基-3'-PACE(“MP”)修饰对称为“中”(“ON”)的靶多核苷酸的体外切割的影响、以及另行测定的对称为“脱靶1”(“OFF1”)、“脱靶2”(“OFF2”)和“脱靶3”(“OFF3”)的三个不同的脱靶多核苷酸的切割的影响,分别表示为CLTA4中靶(ON-target)、CLTA4脱靶1-靶(OFF1-target)、CLTA4脱靶2-靶(OFF2-target)和CLTA4脱靶3-靶(OFF3-target)。

[0016] 图9B是基于图9A中的切割结果,通过计算每个受测的合成sgRNA切割的靶多核苷酸与切割的脱靶多核苷酸的比值而导出的。如图8B所述计算特异性评分,并将大于或等于1.5的评分用阴影表示。每个脱靶多核苷酸的三个最高评分由较暗的阴影指示。

[0017] 图10显示了在靶向人IL2RG基因中的序列的gRNA中的不同位置处的MP修饰对被称为“中靶”(“ON”)的靶多核苷酸的体外切割的影响、以及另行测定的对称为“脱靶3”(“OFF3”)的脱靶多核苷酸的切割的影响,在本图中分别表示为IL2RG中靶(ON-target)和IL2RG脱靶3-靶(OFF3-target)。针对每个受测的合成sgRNA,计算切割的靶多核苷酸与切割



的脱靶多核苷酸的比值。如图8B所述计算特异性评分,并将大于2.0的评分用阴影表示。三个最高评分由较暗的阴影指示。

[0018] 图11A显示了在靶向人HBB基因中的序列的gRNA中的不同位置处的MP修饰对称为“中靶”(“ON”)的靶多核苷酸的体外切割、以及另行测定的称为“脱靶1”(“OFF1”)的脱靶多核苷酸的切割的影响,在本图中分别表示为HBB中靶(ON-target)和HBB脱靶1-靶(OFF1-target)。针对测定的每个合成sgRNA计算切割的靶多核苷酸与切割的脱靶多核苷酸的比值。如图8B所述计算特异性评分,并将大于2.0的评分用阴影表示。三个最高评分由较暗的阴影指示。

[0019] 图11B显示了用合成sgRNA和表达Cas9的质粒转染的人K562细胞中,靶向人HBB基因中序列的gRNA中的不同类型的修饰对称为“中靶”(“ON”)的靶基因组基因座的切割与称为“脱靶1”(“OFF1”)、“脱靶2”(“OFF2”)和“脱靶3”(“OFF3”)的三个不同的脱靶基因组基因座的同时切割的影响,在本图中分别表示为内源性HBB中靶(ON-target)、HBB脱靶1-靶(OFF1-target)、HBB脱靶2-靶(OFF2-target)和HBB脱靶3-靶(OFF3-target)位点。“未修饰”(“Unmodif”)表示未经化学修饰的sgRNA。“3xM”表示将2'-O-甲基(“M”)核苷酸在sgRNA链的最前三个和最后三个核苷酸,即分别在其5'和3'末端掺入。类似地,“3xMS”表示将2'-O-甲基-3'-硫代磷酸(“MS”)核苷酸同样地在sgRNA的5'和3'末端掺入,而“3xMSP”表示将2'-O-甲基-3'-硫代PACE(“MSP”)核苷酸同样地在sgRNA的5'和3'末端掺入。测定了所有sgRNA对转染细胞中的指示基因座的编辑情况。

[0020] 图12A显示了与图11A中所示的条目1—17相同的结果,其中所述条目按照特异性评分从最高到最低排列。条目18—64显示了在靶向人HBB基因中序列的gRNA中的不同位置处的其他MP修饰对于称为“中靶”(“ON”)的靶多核苷酸的体外切割的影响、以及对另行测定的称为“脱靶1”(“OFF1”)的脱靶多核苷酸的切割的影响,分别表示为HBB中靶(ON-target)和HBB脱靶1-靶(OFF1-target)。“1xMP”表示在5'和3'末端的末端核苷酸都已经用MP修饰。计算每个受测的合成sgRNA切割的靶多核苷酸与切割的脱靶多核苷酸的比值。如图8B所述计算特异性评分。最高评分用阴影表示。

[0021] 图12B显示了在转染细胞中,靶向人HBB基因中序列的gRNA中不同位置处的MP修饰对称为“中靶”(“ON”)的靶基因组基因座的切割、以及称为“脱靶1”(“OFF1”)的脱靶基因组基因座的同时切割的影响,在本图中分别表示内源性HBB中靶(ON-target)和HBB脱靶1-靶(OFF1-target)位点。用合成sgRNA和重组Cas9蛋白的复合物转染培养细胞,在转染后48小时如下确定任一个或两个位点处产生的切割百分比:将纯化的基因组DNA样品分割,分别对中靶基因座和脱靶基因座进行PCR扩增,然后对合并的扩增子进行下一代测序,并且根据被评估的中靶切割位点或脱靶切割位点附近的插入缺失的存在与不存在将序列读段进行生物信息学分区。将转染细胞的每个样品中产生的插入缺失相对于用缓冲液而非sgRNA:Cas9复合物处理的模拟转染细胞的对照样品归一化。对于单独转染的每个sgRNA,计算出显示靶位点插入缺失的序列读段的数目与显示脱靶位点插入缺失的读段的数目的比值。如图8B所述计算特异性评分。“1xMP”表示在5'和3'末端的末端核苷酸都已经用MP修饰。条目1-21显示了通过转染和培养K562细胞获得的结果,而条目22-42显示了通过转染和培养诱导的多能干细胞(也称为iPS细胞或iPSC)获得的结果。将条目1-12根据特异性评分从最高到最低排列。同样,将条目13-19、条目22-33和条目34-40按照分组的特异性评分排列。将大于2.0

的特异性评分用阴影表示。

[0022] 图12C显示了图12B中的结果的另一种组织方式,根据测得的比值组织,以分组的最高到最低比值排序。

[0023] 图13是图9A、9B、10和11A中的结果的比较。图13显示了通过对化学修饰的引导RNA用于切割靶多核苷酸的Cas系统时的靶特异性的研究得出的几种趋势。这些趋势所支持的概念当Cas系统中也存在脱靶多核苷酸时特别有用。

[0024] 图14显示了IL2RG sgRNA和VEGFA sgRNA中的各种类型的修饰在K562细胞中的影响。

图15总结了来自人K562细胞中HBB基因的同源定向修复(HDR)的这些实验的结果。

图16总结了来自第一次评估的结果,其测量了HBB中靶位点处、以及16个修饰的或未修饰的sgRNA的类似的脱靶位点处的插入缺失形成。

图17总结了来自两次评估的扩增的DNA基因座(PCR)或捕获的DNA基因座(SureSelect)的深度测序分析的结果。

#### 发明详述

[0025] 本发明至少部分地基于这样的意外发现,即CRISPR-Cas系统容许对gRNA的某些化学修饰,并且降低Cas:gRNA复合的脱靶效应而基本上不损害Cas:gRNA的靶多核苷酸结合、切刻和/或切割效力。

[0026] 本发明提供了合成引导RNA,其包含至少一个增强特异性的修饰。在某些实施方案中,所述至少一个增强特异性的修饰使合成引导RNA与靶多核苷酸之间的至少一个核苷酸对的结合减弱或加强、和/或使合成引导RNA与至少一个脱靶多核苷酸之间的至少一个核苷酸对的结合减弱,使得脱靶减弱中的至少一者大于中靶减弱(如果存在的话)。该合成引导RNA具有gRNA功能性。增强特异性的修饰可以位于引导序列中,例如,位于锁定区、取样区和/或种子区中。在某些实施方案中,增强特异性的修饰降低由gRNA与靶多核苷酸序列和脱靶多核苷酸序列形成的双链体的解链温度,或升高gRNA/靶双链体的解链温度并降低至少一个gRNA/脱靶双链体的解链温度。本发明还提供了包含这些合成引导RNA的gRNA:Cas蛋白复合物,使用所述合成引导RNA切割、切刻或结合靶多核苷酸的方法,以及包含所述合成引导RNA的集合、文库、试剂盒和阵列。本发明还提供了制备合成引导RNA的方法。

#### I. 定义

[0027] 如本文中使用的,术语“引导RNA”通常是指能与Cas蛋白结合并有助于将Cas蛋白靶向到靶多核苷酸(例如DNA或mRNA分子)内的特定位置的RNA分子(或一组RNA分子的总称)。引导RNA可包含crRNA区段和tracrRNA区段。如本文中使用的,术语“crRNA”或“crRNA区段”是指RNA分子或其部分,其包括靶向多核苷酸的引导序列、茎序列(更清楚地说:茎序列包括了茎形成序列,茎形成序列在单引导RNA中与tracrRNA的相应部分形成茎)、以及任选的5'-突出端序列。如本文中使用的,术语“tracrRNA”或“tracrRNA区段”是指这样的RNA分子或其部分,其包括蛋白结合区段(例如,能够与CRISPR相关蛋白例如Cas9等相互作用的蛋白结合区段)。tracrRNA还包括与crRNA部分或完全杂交的区段。术语“引导RNA”涵盖单引导RNA(sgRNA),其中crRNA区段和tracrRNA区段位于同一RNA分子或链中。术语“引导RNA”还总体上涵盖两个或更多RNA分子的组,其中crRNA区段和tracrRNA区段位于不同的RNA分子中。术语“引导RNA”还涵盖如下所述RNA分子或合适的分子区段的组:它/它们结合Cas9以外的

Cas蛋白(例如,Cpf1蛋白),并且在单链RNA或分段的RNA链内具有引导序列,该引导序列包含引导RNA的功能,包括与Cas蛋白结合以形成能够结合、切刻和/或切割靶多核苷酸中的互补序列(或“靶序列”)的gRNA:Cas蛋白复合物。

[0028] 术语“引导序列”是引导RNA中的连续核苷酸序列,其与靶多核苷酸中的靶序列具有部分或完全互补性,并且可以通过由Cas蛋白促进的碱基配对与靶序列杂交。如图1中所示的实例展示,靶序列邻近PAM位点(PAM序列,以及其在另一条链上的互补序列,一起构成PAM位点)。紧邻PAM序列(cas9的NGG)上游的是与靶序列互补的序列(粗体,图1中最下面的链)。与引导序列杂交的靶序列紧邻PAM序列的互补序列(cas9的CCN)的下游。引导序列的核苷酸1(在5'的第一个核苷酸)与靶序列的最后一个核苷酸互补,而引导序列的最后一个核苷酸(图1中引导序列的核苷酸20)与靶序列的第一个核苷酸互补,所述靶序列与PAM位点相邻(紧邻PAM序列的互补序列的下游)。在其他实例诸如Cpf1中,与引导序列杂交的靶序列的位置可以在PAM序列的互补序列的上游。

[0029] 引导序列可以短至约10个核苷酸,长达约30个核苷酸。典型的引导序列的长度为15、16、17、18、19、20、21、22、23和24个核苷酸。合成引导序列的长度通常为20个核苷酸,但可以更长或更短。当引导序列短于20个核苷酸时,与20个核苷酸的引导序列相比,它通常具有5'端的缺失。举例来说,引导序列可以由与靶序列互补的20个核苷酸组成。换言之,除了DNA和RNA之间的A/U差异之外,引导序列与PAM序列上游的20个核苷酸是相同的。如果该引导序列在5'端的3个核苷酸被截短,则20个核苷酸的引导序列的核苷酸4现在变为17聚体中的核苷酸1,20个核苷酸的引导序列的核苷酸5现在变为17聚体中的核苷酸2,等等。新位置是原始位置减去3。类似地,引导序列可以与靶位点的超过20个核苷酸杂交,多出来的核苷酸位于引导序列的5'端,因为引导序列的3'端与靠近PAM位点的靶互补。再举例来说,在22个核苷酸的引导序列中,20聚体中的原来的核苷酸1现在变为核苷酸3,20聚体中原来的核苷酸2变为核苷酸4等。新位置是原始位置加2或减2(-2)。因此,引导序列由从5'端开始计数的核苷酸1至核苷酸“20减N”(20-N)组成,其中N是在-10至10之间的正或负整数(任选地在-10至6之间)。引导序列内的给定核苷酸位置将记为“位置数减N”(数字-N)。例如,在位置5的核苷酸将被记为“5-N”(5减去N),以指示位置5从参考位置的移位,所述参考位置获自20个核苷酸的引导序列,所述移位发生在引导序列的5'端被截短或延伸N个核苷酸时发生。核苷酸位置是正整数。因此,为负数或零的任何(数字-N)位置都没有实际意义,应该被忽略。引导序列可以位于构成gRNA的一条链或多条链内的任何位置。典型的引导序列位于gRNA链的5'端或3'端处或附近。

[0030] 术语“支架”是指包含在天然生物物种中基本相同或高度保守的序列的引导RNA分子部分。支架包括tracrRNA区段和除crRNA区段的3'端处或附近的靶向多核苷酸的引导序列(重复部分)之外的crRNA区段的部分。

[0031] 术语“核酸”、“多核苷酸”或“寡核苷酸”是指DNA分子、RNA分子或其类似物。如本文中使用的,术语“核酸”、“多核苷酸”和“寡核苷酸”包括但不限于DNA分子,如cDNA、基因组DNA、质粒或载体DNA或合成DNA,以及RNA分子,如引导RNA、信使RNA或合成RNA。此外,如本文中使用的,术语“核酸”和“多核苷酸”包括单链和双链形式。本领域的标准惯例在于,通常将寡核苷酸、多核苷酸、RNA分子、DNA分子的不同链和包含2个或更多个核苷酸的各种核酸从它们的5'端编号,并且这个惯例在全文中使用,包括与这样的分子共价连接的5'延伸或“突

出端”的情况。

[0032] 如本文中使用的,“修饰”或“化学修饰”是指不同于在四种最常见的天然核糖核苷酸:腺苷、鸟苷、胞苷和尿苷核糖核苷酸中所发现的化学模块或化学结构的部分。因此,术语“修饰”是指在腺苷、鸟苷、胞苷或尿苷核糖核苷酸的最常见的天然分子结构之中或之上的分子变化。术语“修饰”可以指在核碱基之中或之上、在糖之中或之上、和/或在核苷酸间磷酸二酯键中的变化。术语“修饰”可以指在自然界中发生的核糖核苷酸中的化学结构变化,例如在天然转运RNA(tRNA)中发生的化学修饰,例如但不限于2'-O-甲基、1-甲基腺苷、2-甲基腺苷、1-甲基鸟苷、7-甲基鸟苷、2-硫胞嘧啶、5-甲基胞嘧啶、5-甲酰基胞嘧啶、假尿苷、二氢尿苷、核糖胸苷等。术语“修饰”可以指通常在自然界中不存在的化学修饰,例如但不限于2'-氟、2'-O-甲氧基乙基、'-O-苯基等。术语“相同的修饰”是指糖在糖之中或之上、或在核苷酸间键中的相同类型的化学修饰;例如,2'-O-甲基可以附接到腺苷、鸟苷、胞苷和/或尿苷核糖核苷酸的2'位置,这些修饰可以称为相同类型的修饰或“相同的修饰”。反过来,对于核碱基的修饰,仅有当经修饰的核碱基由相同的分子结构组成时,才称为“相同的修饰”。举例说明区别:1-甲基鸟苷和7-甲基鸟苷两者都具有最常见的天然鸟苷的甲基修饰,但它们不是“相同的修饰”,因为经修饰的核碱基具有不同的分子结构。另举一例,RNA链可包含三个经修饰的核苷酸,例如各自经2'-O-甲基-3'-膦酰基乙酸修饰的鸟苷和两个胞苷,并且可将这样的核苷酸准确地描述为“以相同的方式被修饰”或“用相同的修饰加以修饰”。相反,不同的RNA链可能包含5-甲基胞苷以及缺乏5-甲基取代基的胞苷,两种胞苷核苷酸都可以用2'-O-甲基-3'-膦酰基乙酸修饰,但是这两种胞苷核苷酸是包含不同的修饰的,并且不会称为“以相同的方式修饰”。

[0033] 在寡核苷酸或多核苷酸语境中的术语“修饰”包括但不限于(a)末端修饰,例如5'端修饰或3'端修饰,(b)核碱基(或“碱基”)修饰,包括置换或去除碱基,(c)糖修饰,包括2'、3'和/或4'位置处的修饰,和(d)主链修饰,包括磷酸二酯键的修饰或置换。术语“修饰的核苷酸”通常是指具有对碱基、糖和磷酸二酯键或主链部分中的一者或多者的化学结构的修饰的核苷酸,包括核苷酸磷酸酯。对引导RNA的化学修饰公开在2015年12月3日提交的美国专利申请号14/757,204中,将其全部内容通过提述并入本文。

[0034] 术语“xA”、“xG”、“xC”、“xT”、“xU”或“x(A、G、C、T、U)”以及“yA”、“yG”、“yC”、“yT”、“yU”或“y(A、G、C、T、U)”是指核苷酸、核碱基或核碱基类似物,如由Krueger等人描述于“Synthesis and Properties of Size-Expanded DNAs:Toward Designed,Functional Genetic Systems”,(2007)Acc.Chem.Res.40,141-50,将其内容通过提述完整并入本文。

[0035] 术语“非结构化核酸”或“UNA”是指如美国专利7,371,580中描述的核苷酸、核碱基或核碱基类似物,将其内容通过提述完整并入本文。非结构化核酸或UNA修饰也称为“假互补”核苷酸、核碱基或核碱基类似物(参见例如Lahoud等人(1991)Nucl.Acids Res.36:10,3409-19)。

[0036] 术语“PACE”和“硫代PACE”分别指含有膦酰基乙酸或硫代膦酰基乙酸基团的核苷酸间磷酸二酯键类似物。这些修饰属于一大类化合物,其包含膦酰基羧酸模块、膦酰基羧酸酯模块、硫代膦酰基羧酸模块和硫代膦酰基羧酸酯模块。这些键可分别通过通式 $P(CR^1R^2)_nCOOR$ 和 $(S)-P(CR^1R^2)_nCOOR$ 来描述,其中n是0到6的整数,且 $R^1$ 和 $R^2$ 各自独立地选自H、烷基和取代的烷基。Yamada,Dellinger等人在“Synthesis and Biochemical Evaluation of

Phosphonoformate Oligodeoxyribonucleotides” (2006) J. Am. Chem. Soc. 128:15,5251-61 中描述了这些修饰中的一些,将其内容通过提述完整并入本文。

[0037] 在本公开的一些地方,特别是在公开合成引导RNA的结构和序列以及关于这样的合成引导RNA的实验结果的图中,使用某些缩写来指示某些修饰。“M”在本文中用于指示2'-O-甲基修饰;“S”在本文中用于指示3'-硫代磷酸核苷酸间键修饰;“P”在本文中用于指示3'-磷酸基乙酸(或PACE)核苷酸间键修饰;“MS”在本文中用于指示2'-O-甲基-3'-硫代磷酸核苷酸间键修饰;“MP”在本文中用于指示2'-O-甲基-3'-磷酸基乙酸(或2'-O-甲基-3'-PACE)核苷酸间键修饰;并且“MSP”在本文中用于指示2'-O-甲基-3'-硫代磷酸基乙酸核苷酸间键修饰。

[0038] “糖皱褶”(sugar pucker)是指由于空间力导致5元糖环的一个或两个原子不在平面上,而得到的非平面的糖环。在呋喃核糖中,平面C1'-O4'-C4'是固定的。内皱褶(Endo-pucker)表示C2'或C3'离开这个平面转向O5'方向。外皱褶(Exo-pucker)描述相反方向的移位。C2'-内和C3'-内天然处于平衡状态,但化学修饰可以将糖驱动到优选的皱褶状态。在RNA中,C3'-内构象占优势。DNA可以调整并且能够呈现两种构象。

[0039] 术语“种子区”是指引导序列的与靶核酸序列互补的区域,其引发所述引导序列与所述靶核酸序列的杂交。在一些情况下,种子区在蛋白质、肽或蛋白质复合物辅助下形成准稳定的双链体。一般而言,术语“gRNA引导序列中的种子区”由20个核苷酸的引导序列中的核苷酸11至20(从引导序列的5'端数起)组成,但该区域可以更短或更长,取决于核苷酸序列和对这个区域中RNA核苷酸的化学修饰或相关肽、蛋白质或蛋白质复合物的修饰。

[0040] 术语“取样区”是指邻近种子区的区域,并且这些核苷酸的结合进行直到双链体的结合能等于正在发生结合的温度时为止。一般而言,术语“gRNA中的取样区”由20个核苷酸的引导序列中的核苷酸5至10(从所述引导序列的5'端数起)组成,但另有说明的除外,例如当一个或多个修饰使得取样区在功能上被延伸时,由此可涵盖引导序列中的核苷酸1至10、或者2至10、或者3至10、或者4至10、或者1至11、或者2至11、或者3至11、或者4至11、或者5至11、或者1至12、或者2至12、或者3至12。

[0041] 术语“锁定区”是指邻近取样区的区域,其中形成的双链体的结合能高于正在发生结合的温度。一般而言,术语“gRNA中的锁定区”由20个核苷酸的引导序列中的核苷酸1至4(从所述引导序列的5'端计数)组成,但另有说明的除外,例如当一个或多个修饰使得锁定区在功能上缩短至引导序列的5'端核苷酸1、或者核苷酸1至2、或者核苷酸1至3时。如果有一个或多个核苷酸与5'端共价连接,使得引导序列从典型的20个核苷酸延伸至21个核苷酸、或者至22个核苷酸、或者至23个核苷酸、或者至24个核苷酸、或者至25个核苷酸或更多,则锁定区可能延伸超过CRISPR-Cas9系统的引导序列的20个核苷酸的典型长度。

[0042] 如本文中使用的,术语“靶多核苷酸”或“靶”是指含有靶核酸序列的多核苷酸。靶多核苷酸可以是单链或双链的,在某些实施方案中是双链DNA。在某些实施方案中,靶多核苷酸是单链RNA。如本文中使用的“靶核酸序列”或“靶序列”表示人们希望使用CRISPR系统来结合、切刻或切割的特异性序列或其互补序列。在某些实施方案中,可以选择两个或更多个靶序列来在同一个反应中结合、切刻或切割,例如以替换用于同源重组目的的两个特定靶序列之间的序列。或者,当要同时结合和富集多个靶时,两个或更多个靶序列也是有用的。因此,在使用两个或更多个靶序列的情况下,取决于目的,对应的靶多核苷酸可以在或

可以不在同一基因中。

[0043] “脱靶多核苷酸”或“脱靶”是指这样的多核苷酸，其含有与预期靶核酸部分同源的核酸序列。脱靶多核苷酸可以是单链或双链的，在某些实施方案中是双链DNA。如本文中使用的，“脱靶核酸序列”或“脱靶序列”表示这样的特定序列或其互补序列，人们不希望使用CRISPR系统结合、切刻或切割它，而且它与靶核酸序列具有实质的序列同一性，但并不完全相同。例如，当脱靶核酸序列与靶核酸序列具有至少约60%、至少约75%、至少约85%、至少约90%、至少约90%、至少约90%-95%、至少约97%、或更多的核苷酸(或氨基酸)序列同一性时，则其具有实质的序列同一性。

[0044] 术语“HBB多核苷酸”、VEGFA多核苷酸、“IL2RG多核苷酸”、“CLTA1”多核苷酸和“CLTA4多核苷酸”是指分别包含基因HBB、VEGFA、IL2RG、CLTA1或CLTA4的至少一部分的任何多核苷酸。这样的多核苷酸包括天然存在的、重组的或合成的多核苷酸。这样的多核苷酸可包括在与这样的基因相关的基因组中的基因座处发现的多核苷酸序列，因此包括这样的基因的等位基因和变体。

[0045] 术语“特异性”是指引导RNA能够多好地区分中靶多核苷酸与一个或多个脱靶多核苷酸。引导RNA的特异性可以如下确定，例如，通过计算中靶切割、结合或切刻百分比以及脱靶切割、结合或切刻百分比；计算中靶：脱靶比；和/或从可比较的中靶和脱靶百分比得到的特异性评分(参见本公开的实施例)。术语“中靶百分比”是指在某一项测定内靶多核苷酸的切割、切刻或结合的百分比；例如，如果引导RNA导致该测定中存在的90%的靶多核苷酸的切割、切刻或结合，则其具有90%的中靶百分比。术语“中靶：脱靶比”是指某一测定的每个引导RNA的中靶百分比和脱靶百分比的比值；例如，具有80%中靶百分比和8%脱靶百分比的引导RNA的中靶：脱靶比为10。术语“特异性评分”是指通过将中靶：脱靶比乘以其对应的测定的每个引导RNA的中靶百分比而获得的数字；举例来说，具有80%中靶：百分比和8%脱靶百分比的引导RNA产生的中靶：脱靶比为10，并且特异性评分为8。在一些测定中，使用切割作为替代对象来评估结合或切刻；例如，在通过测序来定量靶位点处的插入缺失形成以评估切割的测定中，这样的测定也可用于评估gRNA的结合或切刻活性。

[0046] 术语“连续的增强特异性的修饰”是指在引导RNA中彼此邻近的两个或更多个增强特异性的修饰。引导RNA可包含1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23个或24个连续的增强特异性的修饰。在广泛使用的CRISPR-Cas9系统中，引导RNA包含引导序列，其包含从所述引导序列的5'端数起的核苷酸1至20。一个或多个连续的增强特异性的修饰可包括在例如核苷酸1和2处的修饰、核苷酸1至3处的修饰、核苷酸1至4处的修饰、核苷酸1至5处的修饰、核苷酸1至6处的修饰、核苷酸1至7处的修饰、核苷酸1至8处的修饰、核苷酸1至9处的修饰、核苷酸1至10处的修饰、核苷酸2和3处的修饰、核苷酸2至4处的修饰、核苷酸2至5处的修饰、核苷酸2至6处的修饰、核苷酸2至7处的修饰、核苷酸2至8处的修饰、核苷酸2至9处的修饰、核苷酸2至10处的修饰、核苷酸3和4处的修饰、核苷酸3至5处的修饰、核苷酸3至6处的修饰、核苷酸3至7处的修饰、核苷酸3至8处的修饰、核苷酸3至9处的修饰、核苷酸3至10处的修饰，等等。

[0047] 术语“杂交(hybridization)”或“杂交(hybridizing)”是指完全或部分互补的多核苷酸链在适合的杂交条件下聚到一起以形成双链结构或区域的过程，其中两个组成链通过氢键连接。如本文中使用的，术语“部分杂交”包括其中双链结构或区域含有一个或多个

凸起或错配的情况。虽然氢键一般在腺嘌呤和胸腺嘧啶或腺嘌呤和尿嘧啶 (A和T或A和U) 之间或者在胞嘧啶和鸟嘌呤 (C和G) 之间形成,但是其他的非经典碱基对也可以形成氢键 (参见,例如Adams等人,“The Biochemistry of the Nucleic Acids,”11th ed.,1992)。可以预期的是,修饰的核苷酸可以非经典途径形成允许或促进杂交的氢键。

[0048] 术语“切割”是指使多核苷酸的磷酸二酯主链中的共价磷酸二酯键断裂。术语“切割”涵盖单链断裂和双链断裂。双链切割可以作为两个不同的单链切割事件的结果而发生。切割可导致平末端或粘性末端(staggered ends)的产生。

[0049] 术语“CRISPR相关蛋白”或“Cas蛋白”是指野生型Cas蛋白、其片段、或其突变体或变体。术语“Cas突变体”或“Cas变体”是指野生型Cas蛋白的蛋白质或多肽衍生物,例如具有一个或多个点突变、插入、缺失、截短的蛋白质、融合蛋白或其组合。在某些实施方案中,“Cas突变体”或“Cas变体”基本上保留了Cas蛋白的核酸酶活性。在某些实施方案中,“Cas突变体”或“Cas变体”被突变,使得一个或两个核酸酶结构域无活性。在某些实施方案中,“Cas突变体”或“Cas变体”具有核酸酶活性。在某些实施方案中,“Cas突变体”或“Cas变体”缺少其野生型对应物的一些或全部核酸酶活性。术语“CRISPR相关蛋白”或“Cas蛋白”还包括各种原核生物物种的野生型Cpf1蛋白 (并命名为来自普雷沃菌属(Prevotella)和弗朗西丝菌属(Francisella) 1的成簇的规律间隔的短回文重复序列核糖核蛋白或CRISPR/Cpf1核糖核蛋白)、其片段或其突变体或变体。

[0050] 术语Cas蛋白的“核酸酶结构域”是指该蛋白质内具有DNA切割催化活性的多肽序列或结构域。Cas9通常催化PAM序列上游的双链断裂。核酸酶结构域可包含在单个多肽链中,或者,切割活性可以由两个(或更多个)多肽联合产生。单个核酸酶结构域可由给定多肽内多于一个分离的氨基酸段组成。这些结构域的实例包括RuvC样基序 (SEQ ID NO:1中的氨基酸7-22、759-766和982-989) 和HNH基序 (氨基酸837-863);参见Gasiunas等人 (2012) Proc.Natl.Acad.Sci.USA 109:39,E2579-E2586和W02013176772。

[0051] 具有“gRNA功能性”的合成引导RNA是具有天然存在的引导RNA的一种或多种功能的合成引导RNA,所述一种或多种功能例如:与Cas蛋白的缔合,或由与Cas蛋白缔合的引导RNA所执行的功能。在某些实施方案中,功能性包括结合靶多核苷酸。在某些实施方案中,功能性包括将Cas蛋白或gRNA:Cas蛋白复合物靶向至靶多核苷酸。在某些实施方案中,功能性包括切刻靶多核苷酸。在某些实施方案中,功能性包括切割靶多核苷酸。在某些实施方案中,功能性包括与Cas蛋白关联或结合。例如,Cas蛋白可以被工程改造为与一种或多种蛋白质或其部分融合的“死”Cas蛋白 (dCas),所述蛋白质例如转录因子增强子或阻抑物脱氨酶蛋白等,使得一种或多种功能由融合蛋白或其部分执行。在某些实施方案中,功能性是具有Cas蛋白的CRISPR-Cas系统中的引导RNA的任何其他已知功能,所述CRISPR-Cas系统包括具有工程改造的Cas蛋白的人工CRISPR-Cas系统。在某些实施方案中,功能性是天然引导RNA的任何其他功能。合成引导RNA可具有比天然存在的引导RNA更大或更小程度的gRNA功能性。在某些实施方案中,与相似天然存在的引导RNA相比,合成引导RNA可具有关于一种特性的更大功能性和关于另一种特性的更小的功能性。

[0052] 具有单链“切刻”活性的Cas蛋白是指包括Cas突变体或Cas变体在内的Cas蛋白,其与野生型Cas蛋白相比具有降低的切割dsDNA的两条链之一的能力。例如,在某些实施方案中,具有单链切刻活性的Cas蛋白具有降低RuvC结构域(或HNH结构域)功能的突变(例如,氨



基酸取代),因此降低了切割靶DNA的一条链的能力。这样的变体的实例包括酿脓链球菌Cas9中的D10A、H839A/H840A和/或N863A取代,并且还包括在其他物种的Cas9酶中的等同位点处的相同或相似的取代。

[0053] 具有“结合”活性或“结合”靶多核苷酸的Cas蛋白,是指与引导RNA形成复合物的Cas蛋白,并且当在这样的复合物中时,经由引导RNA的碱基与另一个多核苷酸之间的氢键键合形成碱基对,引导RNA与另一个多核苷酸(如靶多核苷酸序列)杂交。氢键键合可以借助于沃森-克里克碱基配对或以任何其他序列特异性方式发生。该杂合体可包含形成双链体的两条链、形成多链三链体的三条或更多条链、或这些的任何组合。

[0054] “CRISPR功能”表示任何可以通过CRISPR系统实现的功能或效应,包括但不限于基因编辑、DNA切割、DNA切刻、DNA结合、基因表达调控、CRISPR激活(CRISPRa)、CRISPR干扰(CRISPRi)、和可以通过将cas蛋白与另一种效应子连接而实现的任何其他功能,由此实现对cas蛋白所识别的靶序列的效应子功能。例如,不含核酸酶的cas蛋白可以与转录因子、脱氨酶、甲基酶等融合。所得到的融合蛋白,在靶标的引导RNA存在下,可用于调节靶标的转录,使靶标脱氨、或甲基化。

[0055] 如本文中使用的,术语序列的“部分”或“片段”是指比完整序列小的序列的任何部分(例如,核苷酸亚序列或氨基酸亚序列)。多核苷酸的部分的长度可以是任何长度,例如长度为至少5、10、15、20、25、30、40、50、75、100、150、200、300或500个或更多个核苷酸。引导序列的一部分可以是引导序列的约50%、40%、30%、20%、10%,例如引导序列的三分之一或更短,例如7、6、5、4、3或2个核苷酸长度。

[0056] 在分子背景下的术语“衍生自”是指使用亲本分子或来自该亲本分子的信息分离或制造的分子。例如,Cas9单突变体切口酶和Cas9双突变体无效核酸酶(也称为失活的Cas9、“死Cas9”或dCas9)衍生自野生型Cas9蛋白。

[0057] 在两个或更多个多核苷酸(或两个或更多个多肽)的背景下,术语“实质上相同”是指当使用序列比较算法或通过目测比较或比对最大对应性时,具有至少约60%、至少约70%、至少约80%、至少约90%、约90%-95%、至少约95%、至少约98%、至少约99%或更多核苷酸(或氨基酸)序列同一性的序列或亚序列。优选地,多核苷酸之间的“实质同一性”存在于多核苷酸的至少约20个核苷酸长度、至少约50个核苷酸长度、至少约100个核苷酸长度、至少约200个核苷酸长度、至少约300个核苷酸长度、至少约500个核苷酸长度的区域上,或在多核苷酸的整个长度上。优选地,多肽之间的“实质同一性”存在于多肽的至少约50个氨基酸残基长度、至少约100个氨基酸残基长度的区域上,或在多肽的整个长度上。

[0058] 如本文中公开的,提供了许多数值范围。应当理解的是,还具体考虑了在该范围的上限和下限之间的距离下限单位的十分之一的每个中间值。还具体考虑了陈述范围所涵盖的每个较小范围或中间值。术语“约”通常指的是指示数值的正或负10%。例如,“约10%”可以指示9%到11%的范围,并且“约20”可以表示18-22。“约”的其他含义可以从上下文显而易见,比如四舍五入,因此,例如“约1”还可以表示从0.5到1.4。

## II. CRISPR介导的序列特异性结合和/或切割或切刻

[0059] 图1显示了CRISPR-Cas9介导的序列特异性DNA切割的图。引导RNA被描绘为sgRNA,其具有位于5'结构域内的示例性的20个核苷酸(或20-nt;核苷酸通常缩写为“nt”)的引导序列(其他引导序列的长度可以是例如约15个至约30个核苷酸)、定位于内部的碱基配对的



茎、以及3'结构域。引导序列与DNA靶标中的示例性20-nt靶序列互补。所述茎对应于crRNA中的一段重复序列,并且与tracrRNA中的序列互补。引导RNA的3'结构域对应于结合Cas9核酸酶的tracrRNA的3'结构域。Cas9:gRNA复合物直接结合并切割在被Cas9识别的PAM序列的上游的靶DNA序列或原型间隔序列。在图1中,例示了3-nt PAM序列;然而,包括4-nt、5-nt的其他PAM序列甚至更长的PAM序列是已知的。

[0060] 用于CRISPR-Cas基因组编辑的引导RNA是在RNA-蛋白质复合物中发挥功能的,其中RNA既用作蛋白质的支架,又用于双链体DNA靶标的序列识别。该复合物识别基因组DNA是首先扫描核苷酸PAM序列。一旦识别了PAM序列,RNA-蛋白质复合物就试图在基因组DNA靶标和引导RNA的引导序列之间形成RNA/DNA双链体。这个双链体首先通过Cas蛋白介导的gRNA的“种子区”的碱基配对来引发,其中种子区被认为长度为约十个核苷酸。在结合种子区之后,通过引导RNA 5'端上的剩余核苷酸的杂交形成稳定的RNA/DNA双链体;这通常产生20个核苷酸的DNA/RNA双链体,接下来去进行蛋白质复合物对靶DNA的双链切割。

[0061] CRISPR-Cas RNA-蛋白质复合物能够用于基因组编辑的一个重要方面是序列特异性。CRISPR-Cas RNA-蛋白质复合物切割基因组DNA,作为通过修复或重组使基因失活或修饰基因的过程的第一步。在这个过程中,基因组DNA在不想要的“脱靶”位点处的切割可能产生不希望的后果,例如在基因组中的其他地方产生序列突变。目前,这些脱靶切割事件要么通过筛选技术检测到,要么通过育种技术去除,要么被忽略。CRISPR-Cas RNA-蛋白复合物在原核生物中进化为适应性免疫系统;改进这些复合物的序列特异性将是推动CRISPR-Cas RNA-蛋白质复合物作为真核基因组学工具广泛应用的一项显著进步和创新。

[0062] 序列特异性首先是gRNA:Cas蛋白复合物能够扫描、检测和结合生物整个基因组内的靶位点或连续核苷酸串。为了做到这一点,靶序列需要足够长,使得其连续核苷酸的串或序列在感兴趣的生物整个基因组中仅出现一次,并且定位于期望的基因组编辑位点。赋予基因组内独特性所必需的连续核苷酸或多核苷酸串的长度由该特定多核苷酸的“信息含量”决定。对于大多数真核细胞和生物,靶多核苷酸的长度必须在18至22个核苷酸的范围内,从而具有在整个基因组中独特的足够信息含量(J.Mol.Biol. (1986),188,415-431)。通常,靶多核苷酸越长,信息含量越多,其序列在基因组中仅出现一次的可能性越高;19个核苷酸的序列比18个核苷酸的序列具有更多的信息含量,20个核苷酸的序列比19个核苷酸的序列具有更多的信息含量,以此类推。然而,一条独特的20个核苷酸的序列可能与基因组中别处的一条不同的20-nt序列中除1个核苷酸之外的全部核苷酸的序列匹配,这条含有单个错配的序列即构成脱靶位点。引导序列与脱靶序列(相对于与20个核苷酸的靶序列)的相对结合受引导序列与中靶序列以及脱靶序列之间各自的结合能和动力学平衡的控制。

[0063] 寡核苷酸的差异结合能受到DNA和RNA结合的协作作用的控制,并且可以该协作作用最大化。已经通过两种方式确定了杂交过程中DNA和RNA结合的协作性。首先,当寡核苷酸开始结合其单独的核苷酸亚基时,核苷酸亚基与互补核碱基的结合对随后寡核苷酸序列中下一个相邻核苷酸的结合具有正影响。同时,单独的核苷酸的去结合对下一个相邻核苷酸的结合具有负影响。当尝试进行错配碱基对时,该错配的去结合(unbinding)对相邻核苷酸对的结合具有负影响,同样,当匹配的核苷酸对成功结合时,其对相邻核苷酸对的结合具有正影响。从总能量的角度来看,当寡核苷酸作为多个单独的核苷酸亚基开始结合时——这些单独的核苷酸亚基在多个渐进的步骤中结合和去结合——相对于这些步骤彼此独立

发生的系统,中间状态在统计学上代表不足。换言之,除了结合或去结合之外,存在有限的自由度和有限数量的动力学状态。从分子角度来看,核苷酸串有一些刚性,一旦核苷酸结合,除了导致结合的能量构象之外,相邻核苷酸可以采用的能量构象非常有限。Z.A.Shabarova (1988) Bioorg.Khim.14:12,1656-62在大概30年前首次证明了保持DNA和RNA协作结合需要分子刚性。用两条寡核苷酸,每条长度为7个核苷酸,构建一条长度为14个核苷酸的DNA,这些核苷酸是利用化学连接作用介由1,3-二氨基丙烷或1,3-丙二醇键共价连接的。将柔性寡核苷酸与互补的14个核苷酸的DNA寡核苷酸结合,并测量结合能。在没有天然DNA主链刚性的情况下,柔性的14个核苷酸的DNA单链结合的结合能显著较低,就好像它是两条独立的缺乏协作性的7个核苷酸的DNA寡核苷酸一样。观察到的这个现象,称为协作性(cooperativity),与非协作性系统中相比,可以更高程度地区分的匹配与错配,并且是增加核苷酸序列特异性的一个重要因素。晶体结构已经显示,Cas蛋白将种子区的10个核苷酸预排序成一个A型螺旋的单链部分。将RNA预排序成A型螺旋可限制gRNA的引导序列可采用的动力学状态的数量,从而增加种子区中DNA/RNA杂交的协作性。

[0064] 一串核苷酸与互补核苷酸串的总结合能通常由解链温度( $T_m$ )定义。解链温度是将两个结合的核苷酸串或链(即,通过碱基配对或杂交结合)解离至它们50%结合和50%未结合的程度所需的温度。解链温度通过一种称为增色效应(hyperchromicity)的现象来测量。当两个结合的寡核苷酸链正在热作用下分开时,UV吸收增加。寡核苷酸的热变性引起双螺旋结构解开形成单链寡核苷酸。当溶液中的两个结合的寡核苷酸被加热至高于其解链温度时,双链体解开形成两条单链,它们比双链体吸收更多的光。当将UV吸光度作为温度的函数绘制曲线时,在双链体开始解离的点处获得S形曲线,并且将该S形曲线的中心定义为 $T_m$ 。

[0065] 图4是表明当寡核苷酸双链体通过加热分离成分别的链时UV吸收如何增加的曲线图。S形曲线反映了双链体解离成分别的链,并且S形曲线的中心是双链体的 $T_m$ 。这个曲线指示在生理盐浓度下20个核苷酸的DNA/RNA双链体的解链温度约为50℃。

[0066] 寡核苷酸双链体在解链温度下——此时仅形成50%的双链体——具有最大的匹配对错配的特异性。在此温度下,如果双链体的结合和去结合具有高度协作性或显示陡峭的S形曲线,则脱靶多核苷酸上的单个错配将阻断寡核苷酸的杂交。如果在37℃进行基因编辑实验,那么使用与其靶标的结合 $T_m$ 为37℃的引导RNA将会获得匹配对错配的最佳区分。这里的问题在于,这个方面仅基于单链核酸的热力学,而具有37℃ $T_m$ 的引导RNA仅能部分地结合其双链靶标,导致总的活性或基因编辑低。尽管如此,在监测活性的同时,逐渐降低引导RNA与其靶标的总 $T_m$ ,应该会增加特异性。Yanfang等人(2014) Nat. Biotechnol. 32, 279-284证明了这一原理,他们将引导RNA从20个核苷酸截短至17个核苷酸长,声称在某些脱靶位点的特异性增加了5,000倍,而没有牺牲中靶基因组编辑效率。在生理盐条件下, RNA/DNA双链体中RNA的截短使该双链体的 $T_m$ 降低约2℃/碱基对。将gRNA中的引导序列截短至17个核苷酸将使结合能降低约6℃,从而产生约42℃的双链体 $T_m$ 。然而,当引导序列从20个核苷酸截短至17个核苷酸时,丢失了相当多信息含量,使得它可以更容易地在整个基因组中发现更多的与17-nt引导序列相同或相似的脱靶位点。更有用的方法是通过化学修饰降低20个核苷酸的RNA/DNA双链体的结合能,同时保持结合与去结合的协作性。

[0067] 用于CRISPR-Cas基因组编辑或用于靶多核苷酸切割或切刻的引导RNA作为单引导RNA或两片段双引导RNA存在,其中两个片段称为crRNA(成簇重复序列RNA)和tracrRNA(反

式激活成簇重复序列RNA)。参见以上图2A和2B。图5A还展示了单引导RNA或两片段双引导RNA501(其中crRNA区段和tracrRNA区段形成杂交的双链体)。从左到右(即,从引导RNA的5'端到3'端),图5A概括性地显示了引导RNA上的延伸部分503(有时称为突出端)、取样和锁定区505、Cas蛋白结合性种子区507(通常为10-nt部分)、双引导茎509或单引导茎环511和tracrRNA区513。Cas9蛋白可以结合这些gRNA区域的全部,也许503除外。引导序列包括锁定区、取样区和种子区。对于单引导RNA和双引导RNA,都是由引导RNA 5'端的约20个核苷酸来与靶DNA结合并形成稳定双链体的。这种相对于与其他相似序列的位点竞争性杂交的杂交结合决定了gRNA对靶标的总体特异性,从而决定了利用gRNA:Cas蛋白质复合物的基因组编辑或基因失活的特异性。

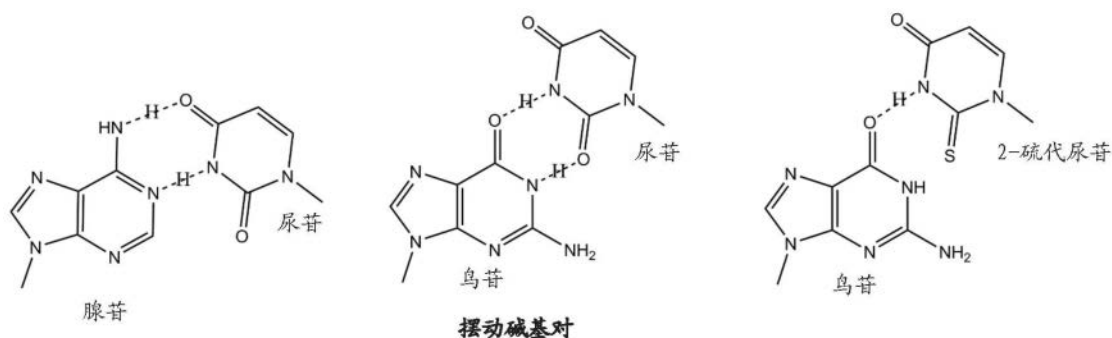
[0068] 在引导序列的3'端的引发下,Cas9介导形成种子RNA/DNA双链体,引导RNA藉此与其靶多核苷酸结合。一旦形成初始种子双链体,gRNA会像拉链一样继续向其5'端杂交。

[0069] 图5B展示了在gRNA501和基因组DNA 515初始形成种子双链体517之后,核苷酸的结合是如何顺序进行通过取样区的。取样区是指邻近种子区的区域,这些核苷酸持续杂交结合,直到双链体的结合能大致等于正在发生结合的温度时为止。在这个区域中的结合和去结合的速率受到结合RNA与未结合RNA(例如,501对519或501对521)之间更大的平衡的控制,该平衡也受到Cas9蛋白的相互作用控制。在引导RNA在生理盐浓度和37℃与基因组DNA形成RNA/DNA双链体的情况下,基于15至16个核苷酸的RNA/DNA双链体的 $T_m$ 大约为37℃的事实,典型的取样区跨越紧邻着10个核苷酸的种子区的5'的5至6个核苷酸。一旦结合的核苷酸的数量超过 $T_m$ 阈值,结合就以连续的方式继续通过锁定区,直到引导序列的20个核苷酸结合,并且如果引导序列的5'端上存在与靶标互补的延伸部分或5'突出端,结合可以进一步进行。一旦结合到达锁定区,在结合RNA与未结合RNA(例如,501对523)之间的平衡就会发生变化,使得其从靶多核苷酸去结合或释放受到结合RNA与未结合RNA的总体平衡的负面影响,该总体平衡现在显著地偏向结合或杂交状态的方向。Slaymaker等人(2016), "Rationally engineered Cas9 nucleases with improved specificity", *Science* 351, 84-8显示了结合RNA与未结合RNA的较大平衡对特异性的影响。Slaymaker等人报道,对Cas9蛋白加以工程改造,将Cas9蛋白的核酸结合沟(或nt沟)中的带正电荷的氨基酸转化为带中性电荷的丙氨酸残基,以减少脱靶插入缺失的形成。这些变化可降低引导RNA对于双链靶DNA的Cas9介导的总亲和力,并迫使亲和力更多地依赖于RNA/DNA杂交,包括碱基对识别。

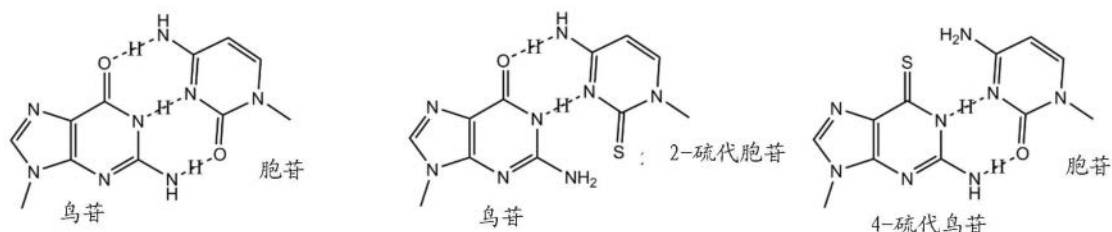
[0070] 图5C展示了在gRNA 525的种子区和取样区中如何利用可降低单个碱基对的结合能但仍保持高水平协作性的核苷酸化学修饰使取样区延伸超过5个或6个核苷酸。该效应可以显著增加引导RNA 525与基因组DNA 515结合的特异性,因为这些化学修饰增加了实现与正在发生结合时的温度相当的杂交结合能所需的核苷酸数。在类似的效应中,使结合RNA与未结合RNA(525对527、531或533)的更大平衡推移所需的核苷酸总数增加,经计算,这导致序列特异性的总体增加。

[0071] 非常令人惊讶和出乎意料的发现是,通过引导序列部分中的三种不同基序中的任何一种:杂环核碱基、糖和核苷酸间磷酸酯键,核苷酸修饰可用于降低gRNA:Cas蛋白复合物对部分互补的脱靶多核苷酸的活性。在具有化学修饰的gRNA中,修饰不显著增加非特异性结合或不显著降低杂交的协作性是重要的,因为非特异性结合的显著增加和/或杂交的协作性的显著降低都可能促进Cas蛋白的脱靶结合、切刻或切割,这是不可取的。

[0072] 在引导序列中的核碱基中,通过碱基配对可及的原子数来降低驱动杂交的氢键键合潜力是有可能的。然而,降低氢键键合潜力还可能导致替代碱基配对,从而增加对脱靶多核苷酸序列的识别。为了避免这种情况,利用引导序列中的高特异性碱基配对是重要的。修饰核碱基以促进高特异性的一个实例是在引导序列中置入2-硫代尿苷来代替尿苷。尿苷核苷酸通常通过两个氢键与腺苷核苷酸结合,但也能够与鸟苷核苷酸结合而产生相当弱的碱基对,通常称为G-U摆动对(wobble pair)。如果使用2-硫代尿苷代替尿苷,则2-硫代尿苷仅能形成不太稳定的G-U摆动对(如果有的话),因为尿嘧啶碱基的C2位上的硫取代基不能用作氢键受体。



[0073] 同样的策略可用于通过将潜在的氢键数从三个减少到两个,例如通过使用2-硫代胞苷或4-硫代鸟苷,以降低典型的鸟苷/胞苷碱基对的氢键键合潜力。在这里说明了它们在修饰的G-C碱基对中形成氢键的可能性降低。



[0074] 还可以使用核糖核苷酸的糖模块修饰来降低引导序列对互补DNA链的亲和力。这可以通过两种方式来完成,第一种是改变糖皱褶,第二种是通过空间拥挤使RNA/DNA双链体变形。通常,糖皱褶被描述为处于两种状态的任一种:2'-内(南,DNA样)糖皱褶,或3'-内(北, RNA样)糖皱褶。2'-内糖皱褶被认为对于碱基配对具有去稳定作用,并且这被认为是由糖苷键的扭转角的变化引起的,从而防止形成最高亲和力的碱基堆积几何形状。可产生2'-内构象的RNA修饰的实例有脱氧核糖、2'-脱氧-2'-氟阿拉伯呋喃糖基、2'-脱氧-2'-氟呋喃核糖基、2'-O-苯基、2'-苯硫基、2'-S-苯硫基、2'-甲基、2'-乙基、2'-丙基、2'-烯丙基、2'-烯丙基苯基、2'-甲基羟基、2'-甲氧基甲基、2'-O-氨基甲酸酯、2'-O-乙基氨基、2'-O-烯丙基氨基、2'-O-丙基氨基和2'-O-取代的苯基取代基。

[0075] 核苷酸间键修饰还可以降低核苷酸的结合亲和力,同时保持碱基对杂交的协作性。这些修饰的实例是膦酰基乙酸、硫代膦酰基乙酸、膦酰基羧酸酯、硫代膦酰基羧酸酯、膦酰基丙酸酯、膦酰基硫代丙酸酯、甲基膦酸酯、甲基硫代膦酸酯和硼烷膦酸。

[0076] 在某些实施方案中,可以添加增加或降低整个引导RNA的结合能的糖修饰或核碱基修饰,以调变或进一步调整对引导RNA掺入其他修饰的结合能。例如,与未修饰的引导RNA相比,掺入2'-O-甲基-硫代膦酰基乙酸(MSP)将引导RNA的结合能降低 $\approx 1.5$ 度。如果将2'-O-

甲基核苷酸掺入引导序列的其他地方,将使总结合能增加 $\approx 0.2$ 度,并且与未修饰的引导RNA相比,所得的引导RNA将具有 $\approx 1.3$ 度的总结合能降低。在某些实施方案中,所述糖修饰包括2'-O-C<sub>1-3</sub>烷基-0-C<sub>1-3</sub>烷基,比如也称为2'-O-(2-甲氧基乙基)或2'-MOE的2'-甲氧基乙氧基(2'-O-CH<sub>2</sub>CH<sub>2</sub>OCH<sub>3</sub>)。在某些实施方案中,所述糖修饰包括2'-卤素,如2'-F、2'-Br、2'-Cl或2'-I。在某些实施方案中,所述糖修饰包括2'-NH<sub>2</sub>。在某些实施方案中,所述糖修饰包括2'-H(例如2'-脱氧核苷酸)。在某些实施方案中,所述糖修饰包括2'-阿拉伯糖基或2'-F-阿拉伯糖基。在某些实施方案中,所述糖修饰包括2'-LNA或2'-ULNA。在某些实施方案中,所述糖包含4'-硫代核糖基。

[0077] 在某些实施方案中,可以将增加或降低整个引导RNA的结合能的核苷酸糖修饰或核碱基修饰掺入到同一个核苷酸上,其中磷酸二酯键被修饰以调变修饰的核苷酸的结合能。例如,可以将3'-磷酸基羧酸酯键与糖修饰如2'-O-甲基、2'-F或2'-O-(2-甲氧基乙基)一起使用。在某些实施方案中,所述糖包含2'-O-C<sub>1-3</sub>烷基-0-C<sub>1-3</sub>烷基,例如2'-甲氧基乙氧基(2'-O-CH<sub>2</sub>CH<sub>2</sub>OCH<sub>3</sub>),又名2'-O-(2-甲氧基乙基)或2'-MOE。在某些实施方案中,所述糖包含2'-卤素,如2'-F、2'-Br、2'-Cl或2'-I。在某些实施方案中,所述糖包含2'-NH<sub>2</sub>。在某些实施方案中,所述糖包含2'-H(例如2'-脱氧核苷酸)。在某些实施方案中,所述糖包含2'-阿拉伯糖基或2'-F-阿拉伯糖基。在某些实施方案中,所述糖包含2'-LNA或2'-ULNA。在某些实施方案中,所述糖包含4'-硫代核糖基。

### III. 引导RNA

[0078] 在至少一个方面,本发明包括经化学修饰的、具有引导RNA的功能性的引导RNA。所述化学修饰的引导RNA包含至少一个增强特异性的修饰,并且可以包含具有比特异性增强更多的功能或具有不同功能的其他化学修饰。

[0079] 包含除四种经典核糖核苷酸(即A、C、G和U)以外的任何核苷酸(无论是非天然的还是天然的(例如,假尿苷、肌苷或脱氧核苷酸))的引导RNA是化学修饰的引导RNA。同样,包含除天然磷酸二酯核苷酸间键以外的任何主链或核苷酸间键的引导RNA具有化学修饰,因此是化学修饰的引导RNA。在某些实施方案中,保留的功能性包括结合Cas蛋白。在某些实施方案中,保留的功能性包括结合靶多核苷酸。在某些实施方案中,保留的功能性包括将Cas蛋白或gRNA:Cas蛋白复合物靶向至靶多核苷酸。在某些实施方案中,保留的功能性包括通过gRNA:Cas蛋白复合物切割靶多核苷酸。在某些实施方案中,保留的功能性包括通过gRNA:Cas蛋白复合物切割靶多核苷酸。在某些实施方案中,保留的功能性是具有Cas蛋白的CRISPR-Cas系统中的引导RNA的任何其他已知功能,所述CRISPR-Cas系统包括具有工程改造的Cas蛋白的人工CRISPR-Cas系统。在某些实施方案中,保留的功能性是天然引导RNA的任何其他功能。

#### A. 示例性修饰

[0080] 在某些实施方案中,增强特异性的修饰是脱氧核糖核苷酸、2'-脱氧-2'-氟阿拉伯呋喃糖基核苷酸、2'-脱氧-2'-氟呋喃糖基核苷酸、具有2'-O-苯基、'-S-苯硫基、2'-甲基、2'-乙基、2'-丙基、2'-烯丙基、2'-烯丙基苯基、2'-甲基羟基、2'-甲氧基甲基、2'-O-氨基甲酸酯、2'-O-乙基氨基、2'-O-烯丙基氨基、2'-O-丙基氨基或2'-O-取代的苯基的糖,或其组合。在某些实施方案中,增强特异性的修饰是磷酸基乙酸、硫代磷酸基乙酸、磷酸基丙酸酯、磷酸基硫代丙酸酯、甲基磷酸酯、甲基硫代磷酸酯或硼烷磷酸;或任何前述的组合。



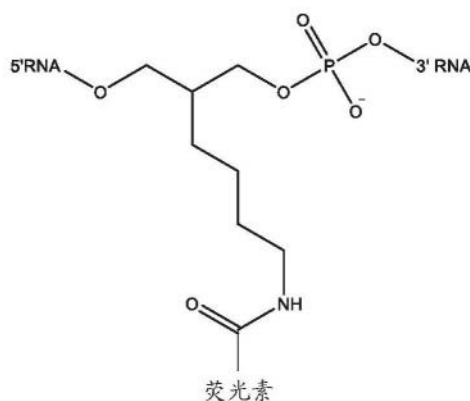
[0081] 在某些实施方案中,掺入引导RNA中的核苷酸糖修饰选自下组:脱氧核糖、2'-脱氧-2'-氟阿拉伯呋喃糖基、2'-脱氧-2'-氟呋喃核糖基、以及具有2'-O-苯基、2'-S-苯--硫基、2'-甲基、2'-乙基、2'-丙基、2'-烯丙基、2'-烯丙基苯基、2'-甲基羟基、2'-甲氧基甲基、2'-O-氨基甲酸酯、2'-O-乙基氨基、2'-O-烯丙基氨基、2'-O-丙基氨基和2'-O-取代的苯基的糖。在某些实施方案中,掺入引导RNA中的核苷酸间键修饰选自下组:硫代磷酸“P(S)” (P(S))、磷酰基羧酸酯( $P(CH_2)_nCOOR$ )如磷酰基乙酸“PACE”( $P(CH_2COO^-)$ )、硫代磷酰基羧酸酯((S) $P(CH_2)_nCOOR$ )如硫代磷酰基乙酸“thioPACE”((S) $P(CH_2COO^-)$ )、烷基膦酸( $P(C_{1-3}\text{烷基})$ )如甲基膦酸酯- $P(CH_3)$ 、硼烷膦酸( $P(BH_3)$ )和二硫代磷酸( $P(S)_2$ )。在某些实施方案中,掺入引导RNA的核苷酸间键修饰选自下组:磷酰基乙酸、硫代磷酰基乙酸、磷酰基丙酸酯、磷酰基硫代丙酸酯、甲基膦酸酯、甲基硫代磷酸和硼烷膦酸。

[0082] 在某些实施方案中,掺入引导RNA中的核碱基(“碱基”)修饰选自下组:2-硫尿嘧啶(“2-硫代U”)、2-硫胞嘧啶(“2-硫代C”)、4-硫尿嘧啶(“4-硫代U”)、6-硫鸟嘌呤(“6-硫代U”)、2-氨基嘌呤、假尿嘧啶、次黄嘌呤、7-脱氮鸟嘌呤、7-脱氮-8-氮杂鸟嘌呤、7-脱氮腺嘌呤、7-脱氮-8-氮杂腺嘌呤、5-甲基胞嘧啶(“5-甲基C”)、5-甲基尿嘧啶(“5-甲基U”)、5-羟甲基胞嘧啶、5-羟甲基尿嘧啶、5,6-脱氢尿嘧啶、5-乙炔基胞嘧啶、5-氨基烯丙基尿嘧啶(“5-氨基烯丙基U”)、5-氨基烯丙基胞嘧啶(“5-氨基烯丙基C”)、脱碱基核苷酸、非结构化核酸(“UNA”)、异鸟嘌呤(“isoG”)、异胞嘧啶(“isoC”) [如描述于“Enzymatic Incorporation of a NewBase pair intoDNA and RNA Extends the GeneticAlphabet.”Piccirilli, J.A.;Krauch,T.;Moroney,S.E.;Benner,S.A.(1990)Nature,343,33]、5-甲基-2-嘧啶(如描述于Rappaport,H.P.(1993)Biochemistry,32,3047)、x(A,G,C,T,U)和Y(A,G,C,T,U)。

[0083] 在某些实施方案中,在核苷酸糖、核碱基、磷酸二酯键和/或核苷酸磷酸酯上引入一个或多个同位素修饰。这样的修饰包括包含用作示踪剂的一个或多个 $^{15}N$ 、 $^{13}C$ 、 $^{14}C$ 、氘、 $^3H$ 、 $^{32}P$ 、 $^{125}I$ 、 $^{131}I$ 原子或其他原子或元素的核苷酸。

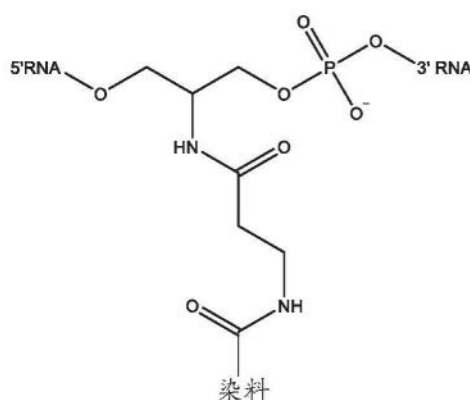
#### [0001] 荧光素

[0084] 在某些实施方案中,掺入引导RNA中的“末端”修饰选自下组:PEG(聚乙二醇)、烃接头(包括:杂原子(O,S,N)取代的烃间隔物;卤代烃间隔物;含酮基-、羧基-、酰胺基-、亚硫酰基-、氨基甲酰基-、硫代氨基甲酰基(thionocarbamaoyl)的烃间隔物);精胺接头;染料,包括与接头连接的荧光染料(例如荧光素、罗丹明、花青),所述接头例如像6--己基、猝灭剂(例如dabcyl,BHQ)和其他标记物(例如生物素、地高辛配基、吖啶、链霉亲和素、亲和素、肽和/或蛋白质)。在某些实施方案中,“末端”修饰包括引导RNA与另一种分子的缀合(或连接),所述另一种分子包括寡核苷酸(包括脱氧核苷酸和/或核糖核苷酸)、肽、蛋白质、糖、寡糖、类固醇、脂质、叶酸、维生素和/或其他分子。在某些实施方案中,掺入引导RNA中的“末端”修饰通过接头(例如2-(4-丁基酰胺基荧光素)丙烷-1,3-二醇双(磷酸二酯)接头)(在下文描绘)定位在引导RNA序列的内部,所述接头作为磷酸二酯键掺入并且可以掺入到引导RNA中的两个核苷酸之间的任何位置。



2-(4-butylamidofluorescein)propane-1,3-diol bis(phosphodiester) linker  
2-(4-丁基酰胺基荧光素)丙烷-1,3-二醇双(磷酸二酯)接头

[0085] 作为说明,其他接头包括例如但不限于:



2-(3-染料-酰胺基)丙酰胺基)丙烷-1,3-二醇双(磷酸二酯)接头

[0086] 在某些实施方案中,末端修饰包括末端官能团,比如胺、硫醇(或巯基)、羟基、羧基、羰基、亚硫酰基、硫代羰基、氨基甲酰基、硫代氨基甲酰基、磷酰基(phosphoryl)、烯、炔、卤素或官能团封端的接头,其中任一者可以随后与期望的模块缀合,所述模块例如荧光染料或非荧光标记物或标签或任何其他分子,例如像寡核苷酸(包括脱氧核苷酸和/或核糖核苷酸,包括适体)、氨基酸、肽、蛋白质、糖、寡糖、类固醇、脂质、叶酸、维生素。缀合采用本领域熟知的标准化学,包括但不限于通过N-羟基琥珀酰亚胺、异硫氰酸酯、DCC(或DCI)和/或任何其他标准方法偶联。

[0087] 在某些实施方案中,标记物或染料与gRNA中的经修饰的核苷酸附接或缀合。使用所谓的“点击”化学或所谓的“方酸”(“squarate”)缀合化学,可以完成荧光染料或其他模块如非荧光标记物或标签(例如生物素、亲和素、链霉亲和素或含有同位素标记物如 $^{15}\text{N}$ 、 $^{13}\text{C}$ 、 $^{14}\text{C}$ 、氘、 $^3\text{H}$ 、 $^{32}\text{P}$ 、 $^{125}\text{I}$ 等等的模块)或任何其他分子,例如像寡核苷酸(包括脱氧核苷酸和/或核糖核苷酸,包括适体)、氨基酸、肽、蛋白质、糖、寡糖、类固醇、脂质、叶酸、维生素或其他分子的缀合。“点击”化学是指炔模块与叠氮化物模块的[3+2]环加成,导致在两个模块之间的三唑连接,如以下方案所示:





8、9、10、11、12、13、14、15、16、17、18、19个或20个增强特异性的修饰核苷酸,并且在gRNA的其他部分或区段中包含至多达100个其他的修饰核苷酸。在其他实施方案中,,修饰的gRNA在引导序列部分上包含5'延伸部分或突出端,并且所述延伸部分的长度为1至20个核苷酸,包含1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19个或20个增强特异性的修饰核苷酸(在gRNA的引导序列部分中的1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19个或20个增强特异性的修饰核苷酸的基础上附加的),并且任选地在gRNA的其他部分中包含另外最多达100个修饰核苷酸。在某些实施方案中,gRNA中的所有核苷酸都是修饰的。在某些实施方案中,所有的修饰都是相同的。在某些实施方案中,所有修饰的核苷酸都具有相同的修饰类型。在某些实施方案中,修饰的gRNA包含不同修饰的核苷酸的组合。在某些实施方案中,修饰的gRNA包含两个或更多个修饰的核苷酸。在某些实施方案中,修饰的gRNA包含三个或更多个修饰的核苷酸。在某些实施方案中,修饰的核苷酸是连续排列的。在某些实施方案中,修饰的gRNA包含修饰的核苷酸的至少一个连续段。在某些实施方案中,修饰的gRNA包含具有至少2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49个或50个修饰的核苷酸的连续段。每个修饰的核苷酸可以独立地包含一个或多个类型的修饰。在某些实施方案中,在修饰的gRNA的序列中,修饰的核苷酸不是连续的,或者一些修饰的核苷酸是连续的,并非所有修饰的核苷酸都是连续的。

[0093] 在某些实施方案中,化学修饰在引导RNA的5'部分内。当引导RNA是双引导RNA时,5'部分内的化学修饰是指引导RNA的crRNA区段的5'部分内的修饰,而不是tracrRNA区段的5'部分内的修饰。当引导RNA是单引导RNA时,它具有一个5'部分,位于crRNA区段中。在某些实施方案中,修饰在引导RNA的5'部分的前五(5)个核苷酸内。在某些实施方案中,修饰在引导RNA的5'部分的前三(3)个核苷酸内。在某些实施方案中,修饰在引导RNA的3'部分内。在某些实施方案中,修饰在引导RNA的3'部分的最后五(5)个核苷酸内。在某些实施方案中,修饰在引导RNA的3'部分的最后三(3)个核苷酸内。在某些实施方案中,修饰在引导RNA的内部区域(即,在5'端和3'端之间)内。

[0094] 在某些实施方案中,化学修饰掺入到引导RNA的5'部分或3'部分中,特别是在5'部分的前5个或10个核苷酸内,或在3'部分的最后5个或10个核苷酸内,以便例如保护RNA免受核酸酶降解或为了其他目的。在一些其他的实施方案中,修饰掺入到引导RNA的5'部分和3'部分两者中,特别是在5'部分的前5个或10个核苷酸内以及在3'部分的最后5个或10个核苷酸内,以便例如保护RNA免受核酸酶降解或为了其他目的。在某些实施方案中,在引导RNA的5'部分和3'部分两者中都存在多于一个类型的修饰。在某些实施方案中,修饰位于引导RNA的5'端、3'端和内部序列内。在某些实施方案中,引导RNA包含在引导RNA的5'或3'部分中的40个或更少、或者20个或更少、或者15个或更少、或者10个或更少、或者5个或更少、或者3个或更少的脱氧核糖核苷酸残基。在引导RNA是双引导RNA的情况下,每个RNA分子在5'端、3'端或两者处都可以包含修饰。在某些实施方案中,末端(5'、3'或两者)处的连续核苷酸(如2、3、4、5个更多个连续核苷酸)被修饰。

[0095] 一般而言,引导序列由20-N个核苷酸组成,其中N是-10至10之间的整数(任选地在-10至6之间)。N可以选自-10、-9、-8、-7、-6、-5、-4、-3、-2、-1、0、1、2、3、4、5、6、7、8、9和10。例如,引导序列可以是20个核苷酸长(N=0)、19个核苷酸长(N=1)、21个核苷酸长(N=-

1),等等。在某些实施方案中,引导序列包含在核苷酸位置(从引导序列的5'端开始)4-N、5-N、7-N、9-N或11-N或其组合的至少一个增强特异性的修饰。下面描述了少数实例。

[0096] 在某些实施方案中,引导序列由从所述引导序列的5'端开始计数的核苷酸1至20组成,并且在选自位置4、5、7、9、10和11的至少一个核苷酸处包含化学修饰。在某些实施方案中,化学修饰在引导序列的核苷酸11处。在某些实施方案中,化学修饰在引导序列的核苷酸5处。在某些实施方案中,化学修饰在引导序列的核苷酸7处。在某些实施方案中,化学修饰在引导序列的核苷酸10处。在某些实施方案中,化学修饰在引导序列的核苷酸9处。在某些实施方案中,化学修饰在引导序列的核苷酸4处。在某些实施方案中,该引导RNA包含至少一个末端修饰。在某些实施方案中,引导RNA在引导序列上游包含5'延伸部分或突出端。

[0097] 在某些实施方案中,引导序列由从所述引导序列的5'端开始计数的核苷酸1至19组成,并且在选自位置3、4、6、8、9和10的至少一个核苷酸处包含化学修饰。在某些实施方案中,化学修饰在引导序列的核苷酸4处。在某些实施方案中,化学修饰在引导序列的核苷酸6处。在某些实施方案中,化学修饰在引导序列的核苷酸8处。在某些实施方案中,化学修饰在引导序列的核苷酸9处。在某些实施方案中,化学修饰在引导序列的核苷酸10处。在某些实施方案中,化学修饰在引导序列的核苷酸3处。在某些实施方案中,该引导RNA包含至少一个末端修饰。在某些实施方案中,引导RNA在引导序列上游包含5'延伸部分或突出端。

[0098] 在某些实施方案中,引导序列由从所述引导序列的5'端开始计数的核苷酸1至18组成,并且在选自位置2、3、5、7、8和9的至少一个核苷酸处包含化学修饰。在某些实施方案中,化学修饰在引导序列的核苷酸3处。在某些实施方案中,化学修饰在引导序列的核苷酸5处。在某些实施方案中,化学修饰在引导序列的核苷酸7处。在某些实施方案中,化学修饰在引导序列的核苷酸8处。在某些实施方案中,化学修饰在引导序列的核苷酸9处。在某些实施方案中,化学修饰在引导序列的核苷酸2处。在某些实施方案中,该引导RNA包含至少一个末端修饰。在某些实施方案中,引导RNA在引导序列上游包含5'延伸部分或突出端。

[0099] 在某些实施方案中,引导序列由从所述引导序列的5'端开始计数的核苷酸1至17组成,并且在选自位置1、2、4、6、7和8的至少一个核苷酸处包含化学修饰。在某些实施方案中,化学修饰在引导序列的核苷酸2处。在某些实施方案中,化学修饰在引导序列的核苷酸4处。在某些实施方案中,化学修饰在引导序列的核苷酸6处。在某些实施方案中,化学修饰在引导序列的核苷酸7处。在某些实施方案中,化学修饰在引导序列的核苷酸8处。在某些实施方案中,化学修饰在引导序列的核苷酸1处。在某些实施方案中,该引导RNA包含至少一个末端修饰。在某些实施方案中,引导RNA在引导序列上游包含5'延伸部分或突出端。

[0100] 在某些实施方案中,引导序列由从所述引导序列的5'端开始计数的核苷酸1至16组成,并且包含在选自位置1、3、5、6和7的至少一个核苷酸处的化学修饰。在某些实施方案中,化学修饰在引导序列的核苷酸1处。在某些实施方案中,化学修饰在引导序列的核苷酸3处。在某些实施方案中,化学修饰在引导序列的核苷酸5处。在某些实施方案中,化学修饰在引导序列的核苷酸6处。在某些实施方案中,化学修饰在引导序列的核苷酸7处。在某些实施方案中,该引导RNA包含至少一个末端修饰。在某些实施方案中,引导RNA在引导序列上游包含5'延伸部分或突出端。

[0100] 在某些实施方案中,引导序列由从所述引导序列的5'端开始计数的核苷酸1至15组成,并且在选自位置2、4、5和6的至少一个核苷酸处包含化学修饰。在某些实施方案中,化

学修饰在引导序列的核苷酸2处。在某些实施方案中,化学修饰在引导序列的核苷酸4处。在某些实施方案中,化学修饰在引导序列的核苷酸5处。在某些实施方案中,化学修饰在引导序列的核苷酸6处。在某些实施方案中,该引导RNA包含至少一个末端修饰。在某些实施方案中,引导RNA在引导序列上游包含5'延伸部分或突出端。

[0101] 在某些实施方案中,引导序列由从所述引导序列的5'端开始计数的核苷酸1至21组成,并且在核苷酸5、6、8、10、11和12中至少一个处包含化学修饰。在某些实施方案中,化学修饰在引导序列的核苷酸12处。在某些实施方案中,化学修饰在引导序列的核苷酸6处。在某些实施方案中,化学修饰在引导序列的核苷酸8处。在某些实施方案中,化学修饰在引导序列的核苷酸11处。在某些实施方案中,化学修饰在引导序列的核苷酸10处。在某些实施方案中,化学修饰在引导序列的核苷酸5处。在某些实施方案中,该引导RNA包含至少一个末端修饰。在某些实施方案中,引导RNA在引导序列上游包含5'延伸部分或突出端。

[0102] 在某些实施方案中,引导序列由从所述引导序列的5'端开始计数的核苷酸1至22组成,并且在核苷酸6、7、9、11、12和13中至少一个处包含化学修饰。在某些实施方案中,化学修饰在引导序列的核苷酸13处。在某些实施方案中,化学修饰在引导序列的核苷酸7处。在某些实施方案中,化学修饰在引导序列的核苷酸9处。在某些实施方案中,化学修饰在引导序列的核苷酸12处。在某些实施方案中,化学修饰在引导序列的核苷酸11处。在某些实施方案中,化学修饰在引导序列的核苷酸6处。在某些实施方案中,该引导RNA包含至少一个末端修饰。在某些实施方案中,引导RNA在引导序列上游包含5'延伸部分或突出端。

[0103] 在某些实施方案中,引导序列由从所述引导序列的5'端计数的核苷酸1至20-N组成,其中N是-10至10之间的正整数或负整数(任选地在-10至6之间),并且在4-N至20-N的任何核苷酸处包含至少一个化学修饰。在某些实施方案中,引导序列包含在选自核苷酸4-N至20-N的至少两个核苷酸处的修饰。在某些实施方案中,引导序列包含在核苷酸4-N、5-N、7-N、9-N、10-N或11-N处的至少一个修饰,以及在选自4-N至20-N(但没有15-N)的核苷酸处的至少另一个修饰。在某些实施方案中,选自4-N至20-N的核苷酸是5-N、6-N、7-N、8-N、9-N、10-N、16-N或17-N。在某些实施方案中,引导RNA进一步包含至少一个末端修饰。在某些实施方案中,引导RNA在引导序列上游包含5'延伸部分或突出端。

[0104] 在某些实施方案中,引导序列由从所述引导序列的5'端开始计数的核苷酸1至20组成,并且包含在选自位置5、6、7、8、9、10、16和17的核苷酸处的至少两个化学修饰。在某些实施方案中,化学修饰在引导序列的核苷酸6和核苷酸10处。在某些实施方案中,化学修饰在引导序列的核苷酸5和核苷酸17处。在某些实施方案中,化学修饰在引导序列的核苷酸6和核苷酸7处。在某些实施方案中,化学修饰在引导序列的核苷酸10和核苷酸17处。在某些实施方案中,化学修饰在引导序列的核苷酸5和核苷酸16处。在某些实施方案中,化学修饰在引导序列的核苷酸10和核苷酸16处。在某些实施方案中,化学修饰在引导序列的核苷酸5和核苷酸9处。在某些实施方案中,化学修饰在引导序列的核苷酸9和核苷酸16处。在某些实施方案中,化学修饰在引导序列的核苷酸8和核苷酸17处。在某些实施方案中,该引导RNA包含至少一个末端修饰。在某些实施方案中,引导RNA在引导序列上游包含5'延伸部分或突出端。

[0105] 在某些实施方案中,引导序列由从所述引导序列的5'端开始计数的核苷酸1至19组成,并且包含在选自位置4、5、6、7、8、9、15和16的核苷酸处的至少两个化学修饰。在某些

实施方案中,化学修饰在引导序列的核苷酸5和核苷酸9处。在某些实施方案中,化学修饰在引导序列的核苷酸4和核苷酸16处。在某些实施方案中,化学修饰在引导序列的核苷酸5和核苷酸6处。在某些实施方案中,化学修饰在引导序列的核苷酸9和核苷酸16处。在某些实施方案中,化学修饰在引导序列的核苷酸4和核苷酸15处。在某些实施方案中,化学修饰在引导序列的核苷酸9和核苷酸15处。在某些实施方案中,化学修饰在引导序列的核苷酸4和核苷酸8处。在某些实施方案中,化学修饰在引导序列的核苷酸8和核苷酸15处。在某些实施方案中,化学修饰在引导序列的核苷酸7和核苷酸16处。

[0106] 在某些实施方案中,引导序列由从所述引导序列的5'端开始计数的核苷酸1至18组成,并且包含在选自位置3、4、5、6、7、8、14和15的核苷酸处的至少两个化学修饰。在某些实施方案中,化学修饰在引导序列的核苷酸4和核苷酸8处。在某些实施方案中,化学修饰在引导序列的核苷酸3和核苷酸15处。在某些实施方案中,化学修饰在引导序列的核苷酸4和核苷酸5处。在某些实施方案中,化学修饰在引导序列的核苷酸8和核苷酸15处。在某些实施方案中,化学修饰在引导序列的核苷酸3和核苷酸14处。在某些实施方案中,化学修饰在引导序列的核苷酸8和核苷酸14处。在某些实施方案中,化学修饰在引导序列的核苷酸3和核苷酸7处。在某些实施方案中,化学修饰在引导序列的核苷酸7和核苷酸14处。在某些实施方案中,化学修饰在引导序列的核苷酸6和核苷酸15处。

[0107] 在某些实施方案中,引导序列由从所述引导序列的5'端开始计数的核苷酸1至17组成,并且包含在选自位置2、3、4、5、6、7、13和14的核苷酸处的至少两个化学修饰。在某些实施方案中,化学修饰在引导序列的核苷酸3和核苷酸7处。在某些实施方案中,化学修饰在引导序列的核苷酸2和核苷酸14处。在某些实施方案中,化学修饰在引导序列的核苷酸3和核苷酸4处。在某些实施方案中,化学修饰在引导序列的核苷酸7和核苷酸14处。在某些实施方案中,化学修饰在引导序列的核苷酸2和核苷酸13处。在某些实施方案中,化学修饰在引导序列的核苷酸7和核苷酸13处。在某些实施方案中,化学修饰在引导序列的核苷酸2和核苷酸6处。在某些实施方案中,化学修饰在引导序列的核苷酸6和核苷酸13处。在某些实施方案中,化学修饰在引导序列的核苷酸5和核苷酸14处。

[0108] 在某些实施方案中,引导序列由从所述引导序列的5'端开始计数的核苷酸1至16组成,并且包含在选自位置1、2、3、4、5、6、12和13的核苷酸处的至少两个化学修饰。在某些实施方案中,化学修饰在引导序列的核苷酸2和核苷酸6处。在某些实施方案中,化学修饰在引导序列的核苷酸1和核苷酸13处。在某些实施方案中,化学修饰在引导序列的核苷酸2和核苷酸3处。在某些实施方案中,化学修饰在引导序列的核苷酸6和核苷酸13处。在某些实施方案中,化学修饰在引导序列的核苷酸1和核苷酸12处。在某些实施方案中,化学修饰在引导序列的核苷酸6和核苷酸12处。在某些实施方案中,化学修饰在引导序列的核苷酸1和核苷酸5处。在某些实施方案中,化学修饰在引导序列的核苷酸5和核苷酸12处。在某些实施方案中,化学修饰在引导序列的核苷酸4和核苷酸13处。

[0109] 在某些实施方案中,引导序列由从所述引导序列的5'端开始计数的核苷酸1至15组成,并且包含在选自位置1、2、3、4、5、11和12的核苷酸处的至少两个化学修饰。在某些实施方案中,化学修饰在引导序列的核苷酸1和核苷酸5处。在某些实施方案中,化学修饰在引导序列的核苷酸1和核苷酸2处。在某些实施方案中,化学修饰在引导序列的核苷酸5和核苷酸12处。在某些实施方案中,化学修饰在引导序列的核苷酸5和核苷酸11处。在某些实施方

案中,化学修饰在引导序列的核苷酸4和核苷酸11处。在某些实施方案中,化学修饰在引导序列的核苷酸3和核苷酸12处。

[0110] 在某些实施方案中,引导序列由从所述引导序列的5'端开始计数的核苷酸1至14组成,并且包含在选自位置1、2、3、4、10和11的核苷酸处的至少两个化学修饰。在某些实施方案中,化学修饰在引导序列的核苷酸4和核苷酸11处。在某些实施方案中,化学修饰在引导序列的核苷酸4和核苷酸10处。在某些实施方案中,化学修饰在引导序列的核苷酸3和核苷酸10处。在某些实施方案中,化学修饰在引导序列的核苷酸2和核苷酸11处。

[0111] 在某些实施方案中,化学修饰包括末端修饰,如5'端修饰或3'端修饰。末端修饰的实例包括但不限于磷酸化(作为天然磷酸酯或多磷酸酯或作为修饰的磷酸酯基团,例如烷基磷酸、磷酸基羧酸酯、磷酸基乙酸、硼烷磷酸、硫代磷酸、二硫代磷酸等)、生物素化、缀合或缀合分子、接头、染料、标记物、标签、官能团(例如但不限于5'-氨基、5'-硫代、5'-酰胺基、5'-羧基等)、反向连接(inverted linkage)、或烃模块,所述模块可包括醚、聚乙二醇(PEG)、酯、羟基、芳基、卤素、磷酸二酯、双环、杂环或其他有机官能团。在某些实施方案中,末端修饰包括二甲氧基三苯甲基。

[0112] 在某些实施方案中,化学修饰包括修饰的碱基。如本文中使用的,“未修饰”碱基包括嘌呤碱基腺嘌呤(A)和鸟嘌呤(G),以及嘧啶碱基胸腺嘧啶(T)、胞嘧啶(C)和尿嘧啶(U)。修饰碱基的实例包括但不限于合成和天然碱基,如2-硫代U、2-硫代C、4-硫代U、6-硫代G、2-氨基A、2-氨基嘌呤、假尿嘧啶、次黄嘌呤、7-脱氮鸟嘌呤、7-脱氮-8-氮杂鸟嘌呤、7-脱氮腺嘌呤、7-脱氮-8-氮杂腺嘌呤、5-甲基C、5-甲基U、5-羟甲基胞嘧啶、5-羟甲基尿嘧啶、5,6-脱氢尿嘧啶、5-丙炔基胞嘧啶、5-丙炔基尿嘧啶、5-乙炔基胞嘧啶、5-乙炔基尿嘧啶、5-烯丙基U、5-烯丙基C、5-氨基烯丙基-尿嘧啶和5-氨基烯丙基-胞嘧啶。在某些实施方案中,修饰包括脱碱基核苷酸。在某些实施方案中,修饰包括非标准嘌呤或嘧啶结构,如Z或P、异鸟嘌呤(isoG)、异胞嘧啶(isoC)、UNA、5-甲基嘧啶、x(A、G、C、T、U)或y(A、G、C、T、U)。在某些实施方案中,修饰的gRNA包含1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49个或50个修饰的碱基。在其他实施方案中,修饰的gRNA包含至少55、60、65、70、75、80、85、90、95、100、110、120、130个或140个修饰的碱基。在某些实施方案中,gRNA中的所有碱基都被修饰。

[0113] 在某些实施方案中,修饰包括经修饰的糖。经修饰的糖的实例包括但不限于具有2'位修饰或4'位修饰的糖。例如,在某些实施方案中,所述糖包含2'-O-C<sub>1-4</sub>烷基,如2'-O-甲基(2'-OMe)。在某些实施方案中,所述糖包含2'-O-C<sub>13</sub>烷基-0-C<sub>13</sub>烷基,如2'-甲氧基乙氧基(2'-O-CH<sub>2</sub>CH<sub>2</sub>OCH<sub>3</sub>),又称2'-O-(2-甲氧基乙基)或2'-MOE。在某些实施方案中,所述糖包含2'-卤素,如2'-F、2'-Br、2'-Cl或2'-I。在某些实施方案中,所述糖包含2'-NH<sub>2</sub>。在某些实施方案中,所述糖包含2'-H(例如2'-脱氧核苷酸)。在某些实施方案中,所述糖包含2'-阿拉伯糖基或2'-F-阿拉伯糖基。在某些实施方案中,所述糖包含2'-LNA或2'-ULNA。在某些实施方案中,所述糖包含4'-硫代核糖基。在某些实施方案中,修饰的gRNA包含1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49个或50个修饰的糖。在其他实施方案中,修饰的gRNA包含至少55、60、65、70、75、80、85、90、95、100、110、120、130个或140个修饰

的糖。在某些实施方案中，gRNA中的所有糖都被修饰。

[0114] 在某些实施方案中，所述修饰包括修饰的主链(即，天然磷酸二酯以外的核苷酸间键)。修饰的核苷酸间键的实例包括但不限于硫代磷酸核苷酸间键、手性硫代磷酸核苷酸间键、二硫代磷酸核苷酸间键、硼烷磷酸核苷酸间键、 $C_{1-4}$ 烷基磷酸核苷酸间键如甲基磷酸核苷酸间键、硼烷磷酸核苷酸间键、磷酰基羧酸核苷酸间键如磷酰基乙酸核苷酸间键、磷酰基羧酸酯核苷酸间键如磷酰基乙酸核苷酸间键、硫代磷酰基羧酸核苷酸间键如硫代磷酰基乙酸核苷酸间键、硫代磷酰基羧酸酯核苷酸间键如硫代磷酰基乙酸核苷酸间键。还包括各种盐、混合盐和游离酸形式。在某些实施方案中，修饰的gRNA包含1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49个或50个修饰的核苷酸间键。在其他实施方案中，修饰的gRNA包含至少55、60、65、70、75、80、85、90、95、100、110、120、130个或140个修饰的核苷酸间键。在某些实施方案中，gRNA中的所有核苷酸间键都是被修饰的。

[0115] 在某些实施方案中，所述修饰为或包含2'-O- $C_{14}$ 烷基、2'-H、2'-O- $C_{13}$ 烷基-O- $C_{13}$ 烷基、2'-F、2'-NH<sub>2</sub>、2'-阿拉伯糖基、2'-F-阿拉伯糖基、2'-LNA、2'-ULNA、4'-硫代核糖基、2-硫代U、2-硫代C、4-硫代U、6-硫代G、2-氨基A、2-氨基嘌呤、假尿嘧啶、次黄嘌呤、7-脱氮鸟嘌呤、7-脱氮-8-氮杂鸟嘌呤、7-脱氮腺嘌呤、7-脱氮-8-氮杂腺嘌呤、5-MeC、5-MeU、5-羟甲基胞嘧啶、5-羟甲基尿嘧啶、5,6-脱氢尿嘧啶、5-丙炔基胞嘧啶、5-丙炔基尿嘧啶、5-乙炔基胞嘧啶、5-乙炔基尿嘧啶、5-烯丙基U、5-烯丙基C、5-氨基烯丙基-尿嘧啶、5-氨基烯丙基-胞嘧啶、脱碱基核苷酸、Z核苷酸、P核苷酸、UNA、异胞嘧啶(isoC)、异鸟嘌呤(isoG)、5-甲基-嘧啶、X(A、G、C、T、U)核苷酸、y(A、G、C、T、U)核苷酸、3'-硫代磷酸基团、3'-磷酰基乙酸基团、3'-磷酰基乙酸基团、3'-硫代磷酰基乙酸基团、3'-硫代磷酰基乙酸基团、3'-甲基磷酸基团、3'-硼烷磷酸基团或3'-二硫代磷酸基团或其组合。

[0116] 在某些实施方案中，修饰的核苷酸包含2'-O-甲基-3'-磷酰基乙酸。在某些实施方案中，修饰的核苷酸包含2'-O-甲基-3'-硫代磷酸。在某些实施方案中，修饰的核苷酸包含2'-O-甲基-3'-硫代磷酰基乙酸。在某些实施方案中，修饰的核苷酸包含2'-O-甲基-3'-磷酰基羧酸酯。在某些实施方案中，修饰的核苷酸包含2'-脱氧-3'-磷酰基乙酸。在某些实施方案中，修饰的核苷酸包含2'-脱氧-3'-硫代磷酸。在某些实施方案中，修饰的核苷酸包含2'-脱氧-3'-硫代磷酰基乙酸。在某些实施方案中，修饰的核苷酸包含2'-脱氧-3'-磷酰基羧酸酯。在某些实施方案中，修饰的核苷酸包含2'-卤代-3'-硫代磷酸。在某些实施方案中，修饰的核苷酸包含2'-卤代-3'-磷酰基乙酸。在某些实施方案中，修饰的核苷酸包含2'-卤代-3'-硫代磷酰基乙酸。在某些实施方案中，修饰的核苷酸包含2'-卤代-3'-磷酰基羧酸酯。在某些实施方案中，修饰的核苷酸包含2'-氟-3'-硫代磷酸。在某些实施方案中，修饰的核苷酸包含2'-氟-3'-磷酰基乙酸。在某些实施方案中，修饰的核苷酸包含2'-氟-3'-硫代磷酰基乙酸。在某些实施方案中，修饰的核苷酸包含2'-氟-3'-磷酰基羧酸酯。在某些实施方案中，修饰的核苷酸包含Z碱基。在某些实施方案中，修饰的核苷酸包含P碱基。

[0117] 在某些实施方案中，引导RNA包含由式(I)表示的寡核苷酸：



[0118] 其中W表示寡核苷酸的一个核苷酸或一段核苷酸，其包含至少一个增强稳定性的修饰，Y表示寡核苷酸的未修饰部分。

[0119] 在某些实施方案中,W在引导RNA的5'部分内。在某些实施方案中,W至少部分地在引导RNA的5'部分的前五(5)个核苷酸内。在某些实施方案中,W至少部分地在引导RNA的5'部分的前四(4)个核苷酸内。在某些实施方案中,W至少部分地在引导RNA的5'部分的前三(3)个核苷酸内。在某些实施方案中,W至少部分地在引导序列的核苷酸4至20内,或者在引导序列的核苷酸5至20内,或者在引导序列的核苷酸10至20内,或者在引导序列的核苷酸13至20内,或者在引导序列的核苷酸13—14或16—19内,或者在引导序列的核苷酸13-14或16-18内。

[0120] 在某些实施方案中,W包含2'-O-C<sub>14</sub>烷基、2'-H、2'-O-C<sub>13</sub>烷基-O-C<sub>13</sub>烷基、2'-F、2'-NH<sub>2</sub>、2'-阿拉伯糖基、2'-F-阿拉伯糖基、2'-LNA、2'-ULNA、4'-硫代核糖基、2-硫代U、2-硫代C、4-硫代U、6-硫代G、2-氨基A、2-氨基嘌呤、假尿嘧啶、次黄嘌呤、7-脱氮鸟嘌呤、7-脱氮-8-氮杂鸟嘌呤、7-脱氮腺嘌呤、7-脱氮-8-氮杂腺嘌呤、5-MeC、5-MeU、5-羟甲基胞嘧啶、5-羟甲基尿嘧啶、5,6-脱氢尿嘧啶、5-丙炔基胞嘧啶、5-丙炔基尿嘧啶、5-乙炔基胞嘧啶、5-乙炔基尿嘧啶、5-烯丙基U、5-烯丙基C、5-氨基烯丙基-尿嘧啶、5-氨基烯丙基-胞嘧啶、脱碱基核苷酸、Z核苷酸、P核苷酸、UNA、异胞嘧啶(isoC)、异鸟嘌呤(isoG)、5-甲基-嘧啶、x(A、G、C、T、U)、y(A、G、C、T、U)、硫代磷酸核苷酸间键、磷酸基乙酸核苷酸间键、磷酸基乙酸核苷酸间键、硫代磷酸基乙酸核苷酸间键、硫代磷酸基乙酸核苷酸间键、甲基磷酸核苷酸间键、硼烷磷酸核苷酸间键、二硫代磷酸核苷酸间键、或其组合。

[0121] 在某些实施方案中,W包含在同一个核苷酸上的2'-O-甲基和3'-磷酸基乙酸基团。在某些实施方案中,W包含在同一个核苷酸上的2'-O-甲基和3'-硫代磷酸基团。在某些实施方案中,W包含在同一个核苷酸上的2'-O-甲基和3'-硫代磷酸基乙酸基团。在某些实施方案中,W包含在同一个核苷酸上的2'-F和3'-硫代磷酸基团。在某些实施方案中,W包含在同一个核苷酸上的2'-F和3'-磷酸基乙酸基团。在某些实施方案中,W包含在同一个核苷酸上的2'-F和3'-硫代磷酸基乙酸基团。

[0122] 在某些实施方案中,W包含1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19个或20个修饰的核苷酸。在某些实施方案中,每个修饰的核苷酸包含相同的修饰。在某些实施方案中,W包含不同修饰的核苷酸的组合。在某些实施方案中,W包含两个或更多个修饰的核苷酸。在某些实施方案中,W包含三个或更多个修饰的核苷酸。在某些实施方案中,由于一个或多个未修饰的核苷酸可能会介于中间,修饰的核苷酸在序列中不是连续排列的或者至少是不完全连续排列的。在某些实施方案中,修饰的核苷酸是连续排列的。在某些实施方案中,W包含修饰的核苷酸的至少一个连续段。在某些实施方案中,W包含至少三(3)个修饰的核苷酸的连续段。在某些实施方案中,W包含至少四(4)个修饰的核苷酸的连续段。在某些实施方案中,W包含至少五(5)个修饰的核苷酸的连续段。

[0123] 在某些实施方案中,修饰是改变稳定性的修饰。稳定性是指gRNA抵抗酶(如核酸酶)的降解及细胞内和细胞外环境中存在的其他物质的降解的能力。在某些实施方案中,相对于没有修饰的引导RNA,修饰增加了引导RNA的核酸酶抗性,因此它增强了引导RNA的稳定性。在某些实施方案中,改变稳定性的修饰是增强稳定性的修饰。例如,在某些实施方案中,增强稳定性的修饰包括2'-O-甲基或2'-O-C<sub>14</sub>烷基核苷酸。在某些实施方案中,增强稳定性的修饰包括2'-卤代核苷酸,例如2'-F、2'-Br、2'-Cl或2'-I。在某些实施方案中,增强稳定性的修饰包括2'-MOE或2'-O-C<sub>13</sub>烷基-O-C<sub>13</sub>烷基。在某些实施方案中,增强稳定性的修饰包



括2'-NH<sub>2</sub>核苷酸。在某些实施方案中,增强稳定性的修饰包括2'-H(或2'-脱氧)核苷酸。在某些实施方案中,增强稳定性的修饰包括2'-阿拉伯糖基或2'-F-阿拉伯糖基。在某些实施方案中,增强稳定性的修饰包括4'-硫代核糖基糖模块。在某些实施方案中,增强稳定性的修饰包括3'-硫代磷酸基团。在某些实施方案中,增强稳定性的修饰包括3'-磷酸基乙酸基团。在某些实施方案中,增强稳定性的修饰包括含有3'-硫代磷酸基乙酸基团的核苷酸。在某些实施方案中,增强稳定性的修饰包括含有3'-甲基磷酸基团的核苷酸。在某些实施方案中,增强稳定性的修饰包括含有3'-硼烷磷酸基团的核苷酸。在某些实施方案中,增强稳定性的修饰包括含有3'-二硫代磷酸基团的核苷酸。在某些实施方案中,增强稳定性的修饰包括解锁核酸("ULNA")核苷酸。

[0124] 在某些实施方案中,增强稳定性的修饰包括在同一个核苷酸上的2'-O-甲基和3'-硫代磷酸基团。在某些实施方案中,增强稳定性的修饰包括在同一个核苷酸上的2'-O-甲基和3'-磷酸基乙酸基团。在某些实施方案中,增强稳定性的修饰包括在同一个核苷酸上的2'-O-甲基和3'-硫代磷酸基乙酸基团。在某些实施方案中,增强稳定性的修饰包括在同一个核苷酸上的2'-氟代和3'-硫代磷酸基团。在某些实施方案中,增强稳定性的修饰包括在同一个核苷酸上的2'-氟代和3'-磷酸基乙酸基团。在某些实施方案中,增强稳定性的修饰包括在同一个核苷酸上的2'-氟代和3'-硫代磷酸基乙酸基团。

[0125] 在某些实施方案中,修饰是改变特异性的修饰。在一些实施方案中,可以通过增强中靶结合和/或切割、或减少脱靶结合和/或切割、或这两者的组合来实现特异性增强。在一些其他的实施方案中,可以例如通过减少中靶结合和/或切割、或增加脱靶结合和/或切割、或这两者的组合来实现特异性降低。

[0126] 在某些实施方案中,改变特异性的修饰包括2'-O-甲基。在某些实施方案中,改变特异性的修饰包括2'-卤代,如2'-氟。

[0127] 在某些实施方案中,改变特异性的修饰包括2-硫尿嘧啶碱基(2-硫代U)。在某些实施方案中,改变特异性的修饰包括2-硫代C。在某些实施方案中,改变特异性的修饰包括4-硫代U。在某些实施方案中,改变特异性的修饰包括6-硫代G。在某些实施方案中,改变特异性的修饰包括2-氨基A。在某些实施方案中,改变特异性的修饰包括2-氨基嘌呤。在某些实施方案中,改变特异性的修饰包括假尿嘧啶。在某些实施方案中,改变特异性的修饰包括次黄嘌呤。在某些实施方案中,改变特异性的修饰包括7-脱氮鸟嘌呤。在某些实施方案中,改变特异性的修饰包括7-脱氮-8-氮杂鸟嘌呤。在某些实施方案中,改变特异性的修饰包括7-脱氮腺嘌呤。在某些实施方案中,改变特异性的修饰包括7-脱氮-8-氮杂腺嘌呤。在某些实施方案中,改变特异性的修饰包括5-甲基C。在某些实施方案中,改变特异性的修饰包括5-甲基U。在某些实施方案中,改变特异性的修饰包括5-羟甲基胞嘧啶。在某些实施方案中,改变特异性的修饰包括5-羟甲基尿嘧啶。在某些实施方案中,改变特异性的修饰包括5,6-脱氢尿嘧啶。在某些实施方案中,改变特异性的修饰包括5-丙炔基胞嘧啶。在某些实施方案中,改变特异性的修饰包括5-丙炔基尿嘧啶。在某些实施方案中,改变特异性的修饰包括5-乙炔基胞嘧啶。在某些实施方案中,改变特异性的修饰包括5-乙炔基尿嘧啶。在某些实施方案中,改变特异性的修饰包括5-烯丙基U。在某些实施方案中,改变特异性的修饰包括5-烯丙基C。在某些实施方案中,改变特异性的修饰包括5-氨基烯丙基U。在某些实施方案中,改变特异性的修饰包括5-氨基烯丙基C。在某些实施方案中,改变特异性的修饰包括脱碱基核苷



酸。在某些实施方案中,改变特异性的修饰包括Z碱基。在某些实施方案中,改变特异性的修饰包括P碱基。在某些实施方案中,改变特异性的修饰包括UNA碱基。在某些实施方案中,改变特异性的修饰包括isoC。在某些实施方案中,改变特异性的修饰包括isoG。在某些实施方案中,改变特异性的修饰包括5-甲基-嘧啶。在某些实施方案中,改变特异性的修饰包括x (A、G、C、T、U)。在某些实施方案中,改变特异性的修饰包括y (A、G、C、T、U)。

[0128] 在某些实施方案中,改变特异性的修饰包括硫代磷酸核苷酸间键。在某些实施方案中,改变特异性的修饰包括膦酰基乙酸核苷酸间键。在某些实施方案中,改变特异性的修饰包括硫代膦酰基乙酸核苷酸间键。在某些实施方案中,改变特异性的修饰包括甲基膦酸核苷酸间键。在某些实施方案中,改变特异性的修饰包括硼烷磷酸核苷酸间键。在某些实施方案中,改变特异性的修饰包括二硫代磷酸核苷酸间键。在某些实施方案中,改变特异性的修饰包括ULNA。在某些实施方案中,改变特异性的修饰包括LNA。

[0129] 在某些实施方案中,相对于没有修饰的引导RNA,修饰通过例如改变引导RNA的解链温度( $T_m$ )而改变RNA碱基配对。在某些实施方案中,相对于没有修饰的引导RNA,修饰降低引导RNA的 $T_m$ 。在某些实施方案中,相对于没有修饰的引导RNA,修饰升高引导RNA的 $T_m$ 。

[0130] 在某些实施方案中,gRNA包含能够与靶多核苷酸杂交的引导序列,并且所述引导序列包含一个或多个修饰,所述修饰相对于没有修饰的相似双链体通过改变gRNA:靶多核苷酸双链体的解链温度( $T_m$ )而改变所述引导序列与所述靶多核苷酸的碱基配对。在某些实施方案中,相对于没有修饰的相似双链体,所述修饰降低gRNA:靶多核苷酸双链体的 $T_m$ 。

[0131] 在某些实施方案中,改变特异性的修饰降低了碱基配对相互作用的 $T_m$ 。在某些实施方案中,增强特异性的修饰使包含引导RNA和靶多核苷酸的第一DNA/RNA双链体的 $T_m$ 降低至少约1℃、或至少约2℃、至少约3℃、至少约4℃、至少约5℃、和/或至多约6℃,或至多约8℃、或至多约10℃、或至多约13℃,例如将 $T_m$ 从降低约1℃至降低约13℃,或降低约1℃至降低约6℃。在某些实施方案中,增强特异性的修饰使包含引导RNA和脱靶多核苷酸的第二DNA/RNA双链体的 $T_m$ 降低至少约1℃、或至少约2℃、至少约3℃、至少约4℃、至少约5℃、和/或至多约6℃,或至多约8℃、或至多约10℃、或至多约13℃,例如将 $T_m$ 从降低约1℃至降低约13℃,或降低约1℃至降低约6℃。

[0132] 在某些实施方案中,相对于没有修饰的引导RNA,合成引导RNA包含改变引导RNA转染效率的化学修饰。在某些实施方案中,相对于没有修饰的引导RNA,修饰增加了引导RNA的转染效率。在某些实施方案中,相对于没有修饰的引导RNA,修饰降低了引导RNA的转染效率。在某些实施方案中,修饰中和了磷酸盐上的阴离子电荷以允许被动扩散进入细胞。在某些实施方案中,中和电荷的修饰包括膦酰基乙酸烷基酯核苷酸间键,如膦酰基乙酸甲酯核苷酸间键。与开发gRNA相关的进一步考虑因素包括可转染性和免疫刺激特性。在某些实施方案中,合成引导RNA包含这样的化学修饰,所述化学修饰促进到细胞特别是真核细胞的细胞核中的有效且可滴定的可转染性,并降低转染细胞中的免疫刺激特性。在某些实施方案中,合成引导RNA包含这样的化学修饰,所述化学修饰促进有效递送到预期的细胞、组织、体液或生物体中并在其中维持足够的持续时间以发挥期望的gRNA功能性。在某些实施方案中,相对于没有修饰的引导RNA,合成引导RNA包含改变引导RNA的免疫刺激效应的化学修饰。

[0133] 在某些实施方案中,相对于没有修饰的引导RNA,合成引导RNA包含增强引导RNA的

稳定性和特异性两者的化学修饰。在某些实施方案中,相对于没有修饰的引导RNA,修饰增强了引导RNA的稳定性和转染效率两者。在某些实施方案中,相对于没有修饰的引导RNA,修饰增强了引导RNA的特异性和转染效率两者。在某些实施方案中,相对于没有修饰的引导RNA,修饰增强了引导RNA的总体效能。

[0134] 在某些实施方案中,具有本申请的化学修饰的引导RNA具有大于1的特异性评分。在某些实施方案中,具有本申请的化学修饰的引导RNA具有至少约1.1的特异性评分。因此,在某些实施方案中,具有本申请的化学修饰的引导RNA具有至少约1.1、至少约1.5、至少约2、至少约3、至少约4、至少约5、至少约6、至少约7、至少约8、至少约9、至少约10、至少约11、至少约12、至少约13、至少约14、至少约15、至少约20、至少约25、至少约30、至少约35、至少约40、至少约45、至少约50、或至少约55的特异性评分。在某些实施方案中,具有本申请的化学修饰的引导RNA具有约2至约60的特异性评分。在某些实施方案中,具有本申请的化学修饰的引导RNA具有约10至约60的特异性评分。

[0135] 在某些实施方案中,具有本申请的化学修饰的引导RNA具有至少约1%的中靶切割。在某些实施方案中,具有本申请的化学修饰的引导RNA具有至少约5%的中靶切割。在某些实施方案中,具有本申请的化学修饰的引导RNA具有至少约10%的中靶切割。在某些实施方案中,包含具有本申请化学修饰的引导RNA的gRNA:Cas蛋白复合物具有至少约30%的中靶切割。因此,在某些实施方案中,具有本申请的化学修饰的引导RNA具有至少约30%、至少约35%、至少约40%、至少约45%、至少约50%、至少约55%、至少约60%、至少约65%、至少约70%、至少约75%、至少约80%、至少约85%、或至少约90%、至少约95%、或至少约99%的中靶切割。在某些实施方案中,具有本申请的化学修饰的引导RNA具有约25%至约99.9%的中靶切割。在某些实施方案中,具有本申请的化学修饰的引导RNA具有约50%至约99.9%的中靶切割。

[0136] 在某些实施方案中,具有本申请的化学修饰的引导RNA具有大于1的中靶:脱靶比。在某些实施方案中,具有本申请的化学修饰的引导RNA具有至少约1.1:1的中靶:脱靶比。因此,在某些实施方案中,具有本申请的化学修饰的引导RNA具有至少约1.1:1、至少约1.5:1、至少约2:1、至少约3:1、至少约4:1、至少约5:1、至少约6:1、至少约7:1、至少约8:1、至少约9:1、至少约10:1、至少约15:1、至少约20:1、至少约25:1、至少约30:1、至少约35:1、至少约40:1、至少约45:1、至少约50:1、至少约60:1、至少约70:1、至少约80:1、至少约90:1、至少约95:1、或至少约99:1的中靶:脱靶比。在某些实施方案中,具有本申请的化学修饰的引导RNA具有至少约1.5:1至约99.9:1的中靶:脱靶比。在某些实施方案中,具有本申请的化学修饰的引导RNA具有至少约10:1至约99.9:1的中靶:脱靶比。

### C. 具有组合修饰的引导RNA

[0137] 在一个方面,本技术提供了具有两个或更多个修饰的组合的引导RNA。在某些实施方案中,两个修饰在同一个核苷酸(例如,一个核苷酸包含2'-O-甲基和3'-磷酸基乙酸模块)上。在其他实施方案中,两个修饰在两个不同的核苷酸(例如,一个核苷酸具有2'-O-甲基,另一个具有3'-磷酸基乙酸模块)上。

[0138] 在某些实施方案中,引导RNA中的每个修饰都是相同的。在某些实施方案中,引导RNA中的至少一个修饰不同于引导RNA中的至少一个其他的修饰。在某些实施方案中,引导RNA包含不同类型修饰的组合,并且所述组合中的至少一个类型存在于引导RNA的多个位置

中。在某些实施方案中,所述组合中的至少一个类型在引导RNA中出现1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19次或20次。

[0139] 在某些实施方案中,所述组合中的至少一个类型的修饰出现在两个或更多个修饰的核苷酸中。在某些实施方案中,所述组合中的至少一个类型的修饰出现在三个或更多个修饰的核苷酸中。在某些实施方案中,由于一个或多个未修饰的核苷酸可能会介于中间,修饰的核苷酸在序列中不是连续排列的或者至少是不完全连续排列的。在某些实施方案中,修饰的核苷酸是连续排列的。在某些实施方案中,引导RNA包含相同类型的修饰核苷酸的连续段。在某些实施方案中,所述段具有至少2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39个或40个修饰的核苷酸。

[0140] 在某些实施方案中,所述组合中的至少一个修饰包括修饰的糖。在某些实施方案中,修饰的gRNA包含1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39个或40个修饰的糖。在其他实施方案中,修饰的gRNA包含至少55、60、65、70、75、80、85、90、95、100、110、120、130个或140个修饰的糖。在某些实施方案中,gRNA中的所有糖都是修饰的。

[0141] 在某些实施方案中,所述组合中的至少一个修饰包括修饰的主链(即,除天然磷酸二酯以外的核苷酸间键)。在某些实施方案中,修饰的gRNA包含1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39个或40个修饰的核苷酸间键。在其他实施方案中,修饰的gRNA包含至少55、60、65、70、75、80、85、90、95、100、110、120、130个或140个修饰的核苷酸间键。在某些实施方案中,gRNA中的所有核苷酸间键都是修饰的。

[0142] 在某些实施方案中,引导RNA包含连续的修饰。在某些实施方案中,引导序列包含1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19个或20个连续的化学修饰。在某些实施方案中,所述化学修饰位于核苷酸1与2、1至3、1至4、1至5、1至6、1至7、1至8、1至9、1至10、2与3、2至4、2至5、2至6、2至7、2至8、2至9或2至10处。

[0143] 在某些实施方案中,所述组合中的至少一个修饰包括2'-O-甲基、2'-氟、2'-氨基、2'-脱氧、2'-阿拉伯糖基、2'-F-阿拉伯糖基、2-硫尿嘧啶、2-氨基腺嘌呤、5-甲基胞嘧啶、5-氨基烯丙基尿嘧啶、Z碱基、3'-硫代磷酸、3'-磷酸基乙酸、3'-磷酸基乙酸、3'-硫代磷酸基乙酸、3'-硫代磷酸基乙酸、3'-甲基磷酸酯、3'-硼烷磷酸、3'-二硫代磷酸,或其组合。在某些实施方案中,所述组合中的至少一个修饰包括2'-O-甲基、2'-脱氧、Z碱基、硫代磷酸核苷酸间键、磷酸基乙酸核苷酸间键、硫代磷酸基乙酸核苷酸间键或其组合。在某些实施方案中,所述组合中的至少一个修饰包括2'-F、2-硫代U、4-硫代U、2-氨基A、5-甲基C、5-甲基U、5-氨基烯丙基U,或其组合。在某些实施方案中,所述组合中的至少一个修饰是“末端”修饰,如末端磷酸酯、PEG、末端胺、末端接头如烃接头、取代的烃接头、方酸接头、三唑接头、内部接头如2-(4-丁基酰胺基荧光素)丙烷-1,3-二醇双(磷酸二酯)接头、与染料缀合的接头、与非荧光标记物缀合的接头、与标签缀合的接头或与固体支持物例如珠子或微阵列缀合的接头。在某些实施方案中,所述组合中的至少两个修饰包括2'-O-甲基核苷酸和硫代磷酸核苷酸间键、2'-O-甲基核苷酸和磷酸基乙酸核苷酸间键、或2'-O-甲基核苷酸和硫代磷酸基乙酸核苷酸间键。在某些实施方案中,所述组合中的至少两个修饰包括2'-O-甲基核苷酸和磷

酰基羧酸核苷酸间键、2'-O-甲基核苷酸和膦酰基羧酸酯核苷酸间键、2'-O-甲基核苷酸和硫代膦酰基羧酸核苷酸间键、2'-O-甲基核苷酸和硫代膦酰基羧酸酯核苷酸间键,或其组合。在其他实施方案中,所述组合中的修饰进一步包含2-硫尿嘧啶、2-硫胞嘧啶、4-硫尿嘧啶、6-硫鸟嘌呤、2-氨基腺嘌呤、2-氨基嘌呤、假尿嘧啶、肌苷、7-脱氮鸟嘌呤、7-脱氮-8-氮杂鸟嘌呤、7-脱氮腺嘌呤、7-脱氮-8-氮杂腺嘌呤、5-甲基胞嘧啶、5-甲基尿嘧啶、5-羟甲基胞嘧啶、5-羟甲基尿嘧啶、5,6-脱氢尿嘧啶、5-丙炔基胞嘧啶、5-丙炔基尿嘧啶、5-乙炔基胞嘧啶、5-乙炔基尿嘧啶、5-烯丙基尿嘧啶、5-烯丙基胞嘧啶、5-氨基烯丙基-尿嘧啶、5-氨基烯丙基-胞嘧啶或脱碱基核苷酸。

[0144] 在某些实施方案中,所述组合中的至少一个修饰包括2'-O-甲基-3'-硫代磷酸。在某些实施方案中,所述组合中的至少一个修饰包括2'-O-甲基-3'-膦酰基乙酸。在某些实施方案中,所述组合中的至少一个修饰包括2'-O-甲基-3'-硫代膦酰基乙酸。在某些实施方案中,所述组合中的至少一个修饰包括2'-卤代-3'-硫代磷酸。在某些实施方案中,所述组合中的至少一个修饰包括2'-卤代-3'-膦酰基乙酸。在某些实施方案中,所述组合中的至少一个修饰包括2'-卤代-3'-硫代膦酰基乙酸。在某些实施方案中,所述组合中的至少一个修饰包括2'-氟-3'-硫代磷酸。在某些实施方案中,所述组合中的至少一个修饰包括2'-氟-3'-膦酰基乙酸。在某些实施方案中,所述组合中的至少一个修饰包括2'-氟-3'-硫代膦酰基乙酸。至少两个或三个修饰的可能组合分别在图6和图7中表示,并且通过提述将其并入本文。

[0145] 在某些实施方案中,引导RNA包含由式(III)或式(IV)表示的寡核苷酸:

W-Y-Q (III);或

Y-W-X-Q (IV)

[0146] 其中Q和W各自独立地表示寡核苷酸的一个核苷酸或一段核苷酸,其包含至少一个增强特异性的修饰,并且Y和X各自独立地表示寡核苷酸的未修饰部分。

[0147] 在某些实施方案中,W在引导RNA的5'部分内。在某些实施方案中,W至少部分地在引导RNA的5'部分的前五(5)个核苷酸内。在某些实施方案中,W至少部分地在引导RNA的5'部分的前三(3)个核苷酸内。在某些实施方案中,W在引导RNA的内部区域(即,在5'端和3'端之间)内。在某些实施方案中,W至少部分地在引导序列的核苷酸4至20内,或者在引导序列的在核苷酸5至20内,或者在引导序列的核苷酸10至20内,或者在引导序列的核苷酸13至20内,或者在引导序列的核苷酸13-14或16-19内,或者在引导序列的核苷酸13-14或16-18内。

[0148] 在某些实施方案中,相对于没有修饰的引导RNA,所述组合中的至少一个修饰增强引导RNA的稳定性和特异性。在某些实施方案中,相对于没有修饰的引导RNA,所述组合中的至少一个修饰增强引导RNA的稳定性和转染效率。在某些实施方案中,相对于没有修饰的引导RNA,所述组合中的至少一个修饰增强引导RNA的特异性和转染效率。

[0149] 在某些实施方案中,所述组合中的至少一个修饰改变引导RNA的二级结构。这个修饰改变引导RNA中任何RNA/RNA内部双链体的碱基配对。这些修饰中的一些增加RNA/RNA结构的碱基配对,或者替代地,增加RNA/RNA双链体的 $T_m$ ,而其他修饰降低一个或多个RNA/RNA双链体的碱基配对(或 $T_m$ )。这样的修饰包括碱基修饰的核苷酸,特别是UNA核苷酸,如2-硫代尿苷和2-氨基腺苷对、Z/P核苷酸对、isoC/isoG对、6-硫代G/5-甲基嘧啶对、以及具有在如前论述的糖或核苷酸间键上的修饰的核苷酸。

[0150] 在某些实施方案中,所述组合包括增加核酸酶抗性(即稳定性)的至少一个修饰或

一组修饰以及增加特异性(即减少脱靶效应)的至少一个修饰或一组修饰。在某些实施方案中,所述组合包括增加核酸酶抗性(即稳定性)的至少一个修饰或一组修饰,以及提高引导RNA中的一些碱基配对的T<sub>m</sub>的至少一个修饰或一组修饰。在某些实施方案中,所述组合包括增加核酸酶抗性(即稳定性)的至少一个修饰,或一组修饰以及降低引导RNA的一些碱基配对的T<sub>m</sub>的至少一个修饰或一组修饰。在某些实施方案中,所述组合包括增加核酸酶抗性(即稳定性)的至少一个修饰或一组修饰、增加引导RNA中的一些碱基配对的T<sub>m</sub>的至少一个修饰或一组修饰、以及降低引导RNA中别处的一些碱基配对的T<sub>m</sub>的至少一个修饰或一组修饰。在某些实施方案中,所述组合包括增加核酸酶抗性(即稳定性)的至少一个修饰或一组修饰以及增加引导RNA与Cas蛋白结合的至少一个修饰或一组修饰。在某些实施方案中,所述组合包括增加核酸酶抗性(即稳定性)的至少一个修饰或一组修饰以及降低引导RNA与Cas蛋白结合的至少一个修饰或一组修饰。在某些实施方案中,引导RNA包含不同类型的修饰的组合。

#### D. 引导RNA结构

[0151] 在某些实施方案中,引导RNA能够与CRISPR相关蛋白形成复合物。在某些实施方案中,CRISPR相关蛋白由CRISPR-Cas II型系统提供或源自CRISPR-Cas II型系统,该系统具有RNA引导的多核苷酸结合和/或核酸酶活性。在某些实施方案中,CRISPR相关蛋白是Cas9、Cas9突变体或Cas9变体。在某些实施方案中,CRISPR相关蛋白是来自酿脓链球菌的Cas9核酸酶。在某些实施方案中,CRISPR相关蛋白是来自嗜热链球菌的Cas9核酸酶。在某些实施方案中,CRISPR相关蛋白是来自金黄色葡萄球菌的Cas9核酸酶。在某些实施方案中,合成引导RNA或合成引导RNA:CRISPR相关蛋白复合物维持天然引导RNA或不具有修饰核苷酸的复合物的功能性。在某些实施方案中,所述功能性包括结合靶多核苷酸。在某些实施方案中,所述功能性包括切割靶多核苷酸。在某些实施方案中,所述功能性包括切割靶多核苷酸。在某些实施方案中,靶多核苷酸在体外核酸中。在某些实施方案中,靶多核苷酸在体内或体外的细胞的基因组内(如在培养的细胞或从生物分离的细胞中)。在某些实施方案中,靶多核苷酸是DNA中的原型间隔序列。

[0152] 在某些实施方案中,crRNA区段包含25至80个核苷酸。在某些实施方案中,crRNA区段包含能够与靶序列杂交的引导序列。在某些实施方案中,引导序列与靶序列或其部分互补。在某些实施方案中,引导序列包含15至30个核苷酸。在某些实施方案中,crRNA区段包含茎序列。在某些实施方案中,茎序列包含10至50个核苷酸。在某些实施方案中,crRNA区段包含5'-突出端序列。在某些实施方案中,5'-突出端序列包含1至10个核苷酸、或者1至5个核苷酸、或者1、2或3个核苷酸。在某些实施方案中,crRNA包含(i)能够与靶序列杂交的引导序列和(ii)茎序列。在某些实施方案中,crRNA包含(i)5'-突出端序列,(ii)能够与靶序列杂交的引导序列,和(iii)茎序列。在某些crRNA区段包含茎序列的实施方案中,tracrRNA区段包含与crRNA区段的茎序列部分或完全互补的核苷酸序列。在某些实施方案中,tracrRNA区段包含至少一个另外的双链体结构。

[0153] 在某些实施方案中,引导RNA是单引导RNA,其中crRNA区段和tracrRNA区段通过环L连接。在某些实施方案中,环L包含1、2、3、4、5、6、7、8、9个或10个核苷酸。在某些实施方案中,环L包含GNRA的核苷酸序列,其中N表示A、C、G或U,并且R表示A或G。在某些实施方案中,环L包含GAAA的核苷酸序列。在某些实施方案中,引导RNA包含多于一个环。

[0154] 引导RNA包含5'部分(即,5'半部分)和3'部分(即,3'半部分)。在某些实施方案中,crRNA区段位于tracrRNA区段的5'(即,上游)。在某些实施方案中,tracrRNA区段相对于crRNA区段位于5'。

[0155] 在某些实施方案中,引导RNA包含至少两个分别的RNA链,例如crRNA链和另外的tracrRNA链。参见例如图2A。在某些实施方案中,每条链都是包含一个或多个修饰的合成链。在某些实施方案中,两条链中的至少一条是包含一个或多个修饰的合成链。在某些实施方案中,两条链一起发挥引导靶多核苷酸被Cas蛋白(如Cas9)结合、切刻或切割的作用。在某些实施方案中,crRNA序列和tracrRNA序列在分别的链上,并通过两个互补序列彼此杂交形成茎或双链体。

[0156] 在某些实施方案中,引导RNA是包含crRNA序列和tracrRNA序列的单引导RNA。参见例如图2B。在某些实施方案中,crRNA序列和tracrRNA序列通过环序列或“环”连接。在某些实施方案中,单引导RNA包含5'部分和3'部分,其中crRNA序列位于tracrRNA序列的上游。

[0157] 在某些实施方案中,两个RNA片段的总长度可以为约50-220(例如,约55-200、60-190、60-180、60-170、60-160、60-150、60-140、60-130和60-120)个核苷酸长度,如约60、70、80、90、100、110、120、130、140、150、160、170、180、190、200个或220个核苷酸长度。类似地,单引导RNA(例如,图2B)可以为约50-220(例如,约55-200、60-190、60-180、60-170、60-160、60-150、60-140、60-130和60-120)个核苷酸长度,如约60、70、80、90、100、110、120、130、140、150、160、170、180、190、200个或220个核苷酸长度。

[0158] 如图2A和2B中所示,合成引导RNA包含:(i) crRNA序列,其包含(a)能够与核酸中的靶序列杂交的引导序列(例如,区段 $G_1-G_n$ ,其中每个G表示引导序列中的核苷酸), (b) 第一茎序列(例如,区段 $X_1-X_n$ ,其中每个X表示第一茎序列中的核苷酸),其能够部分或完全地与第二茎序列杂交,以及任选地(c) 5'-突出端序列(例如,区段 $O_1-O_n$ ,其中每个O表示突出端序列中的核苷酸);和(ii) tracrRNA序列,其包含所述第二茎序列(例如,区段 $Y_1-Y_n$ ,其中每个Y表示第二茎序列中的核苷酸)。tracrRNA序列进一步包含区段 $T_1-T_n$ ,其中每个T表示tracrRNA序列中的核苷酸。图2A中所示的合成引导RNA包括一个或多个修饰。同样,图2B中所示的合成引导RNA包括一个或多个修饰。在某些实施方案中,修饰位于沿着crRNA、tracrRNA或包含crRNA区段、tracrRNA区段和任选的环的单引导RNA的长度的任何点。在某些实施方案中,在图2A和2B中所示的合成引导RNA中由O、G、X、Y或T表示的任何核苷酸可以是修饰的核苷酸。图2B中所示的引导RNA表示单引导RNA(sgRNA),其中crRNA区段和tracrRNA区段通过具有序列GNRA的环连接,其中N表示A、C、G或U,并且R表示A或G。

[0159] 在某些实施方案中,引导RNA的crRNA区段的长度为25-70(例如,30-60、35-50或40-45)个核苷酸。在某些实施方案中,引导序列的长度为12-30(例如16-25、17-20或15-18)个核苷酸。在一些实施方案中,crRNA的5'部分不与靶序列杂交或仅部分杂交。例如,在crRNA区段上可以存在5'-突出端。

[0160] 在某些实施方案中,单引导RNA包含中央部分,所述中央部分包含crRNA区段的茎序列、tracrRNA区段的茎序列以及任选地使crRNA区段与tracrRNA区段共价连接的环。在某些实施方案中,单引导RNA的中央区段的长度为8-60(例如10-55、10-50或20-40)个核苷酸。

[0161] 在某些实施方案中,引导RNA的tracrRNA区段的长度为10-130(例如10-125、10-100、10-75、10-50或10-25)个核苷酸。在某些实施方案中,除了中央区段中的任何发夹或双

链体结构之外, tracrRNA区段还包含一个或多个发夹或双链体结构。

#### E. 引导RNA的合成

[0162] 在某些实施方案中, 通过使用合成有机化学领域的化学合成来产生引导RNA, 包括单引导RNA。包含除了四种主要核糖核苷酸(即A、C、G和U)以外的任何核苷酸(无论是非天然的还是天然的, 如假尿苷、肌苷或脱氧核苷酸)的引导RNA具有化学修饰或取代, 所述修饰或取代位于在化学/结构上不同于RNA中的四种主要核苷酸中任何一种的核苷酸处。

[0163] 可以使用本领域熟知的方法(如TBDMS化学、TOM化学、ACE化学等)化学合成本文所述的合成引导RNA。例如, 可以使用TC化学通过Dellinger等人(2011) J. Am. Chem. Soc. 133, 11540中、在美国专利8, 202, 983和美国专利申请2010/0076183A1中描述的方法合成所述合成引导RNA, 将其内容通过提述完整并入本文。“TC化学”是指通过硫代氨基甲酸酯保护基团使用2'-羟基模块上保护的RNA单体核苷酸前体的组合物和方法, 以合成未修饰的RNA或包含一个或多个修饰的核苷酸的修饰的RNA。使用TC—RNA化学法化学合成相对长的RNA(长达200聚体或更多)的能力允许人们产生具有能够胜过由四种主要核糖核苷酸(A、C、G和U)所实现的特殊特征的引导RNA。还可以使用本领域已知的方法(包括体外转录和基于细胞的表达)制造本文描述的一些合成引导RNA。例如, 可将2'-氟NTP掺入到通过基于细胞的表达产生的合成引导RNA中。

[0164] 引导RNA的合成也可通过RNA序列的化学合成或酶促合成来完成, 所述RNA序列随后通过酶连接在一起, 或通过化学连接加以化学连接, 包括但不限于溴化氰化学法, 如R. Kumar等人(2007) J. Am. Chem. Soc. 129, 6859-64中发表的“点击”化学, 或如K. Hill在名称为“Compositions and methods for conjugating oligonucleotides.”的W02013176844中所述的方酸(squarate)缀合化学。

[0165] 在某些实施方案中, 提供了用于制备合成引导RNA的方法。所述方法包括: 选择基因组中的靶多核苷酸; 鉴定基因组中的一个或多个脱靶多核苷酸; 鉴定一个或多个共有的核苷酸残基, 其中所述共有的核苷酸残基存在于靶多核苷酸和脱靶多核苷酸中; 和设计合成引导RNA, 其中引导序列包括增强特异性的修饰。所述方法可包括合成设计的引导RNA。在某些实施方案中, 通过如下方式鉴定脱靶多核苷酸: 预测脱靶位点及其严重性的算法, 如在<http://www.rgenome.net/Cas-OFFinder>; <https://cm.jefferson.edu/Off-Spotter>; 或<http://crispr.mit.edu>中发现的那些; 或用于鉴定和量化在实际情况下脱靶位点激活的其他技术, 正如在Tsai等人(2015) Nat. Biotechnol. 33, 187-97; Ran等人(2015) Nature 520, 186-91; Frock等人(2015) Nat. Biotechnol. 33, 179-86中所公开。在某些实施方案中, 所述方法进一步包括鉴定靶多核苷酸和脱靶多核苷酸的序列之间的至少一个区别位置, 其中在所述至少一个区别位置上靶多核苷酸和脱靶多核苷酸具有不同的核苷酸残基, 并且在合成引导RNA中纳入与靶多核苷酸中所述至少一个区别位置处的核苷酸匹配(即互补)的核苷酸。

[0166] 如下文进一步描述的, 本文中公开的引导RNA(包括那些包含修饰的核苷酸和/或修饰的核苷酸间键的引导RNA)可用于在体外或体内(如在全细胞测定中, 在完整细胞中或在整个生物体中)执行各种CRISPR介导的功能(包括但不限于编辑基因、调节基因表达、切割靶序列、和与靶序列结合)。对于体外或体内应用, 可以按本领域已知的任何方式将RNA递送到细胞或整个生物体中。

### 文库和阵列

[0167] 在一个方面,本发明提供了多个引导RNA的组或文库。在某些实施方案中,所述文库含有两个或更多个本文中公开的引导RNA。文库可以含有约10个至约 $10^7$ 个个体成员,例如约10个至约 $10^2$ 个、约 $10^2$ 个至约 $10^3$ 个、约 $10^3$ 个至约 $10^5$ 个、约 $10^5$ 个至约 $10^7$ 个成员。文库的个体成员至少在引导序列(即gRNA的DNA靶向区段)中不同于文库的其他成员。另一方面,在某些实施方案中,文库的每个个体成员与文库的所有其他成员的tracrRNA区段可以包含相同或基本相同的核苷酸序列。以这种方式,文库可以包含靶向一个或多个多核苷酸中的不同多核苷酸或不同序列的成员。

[0168] 在某些实施方案中,文库包含至少 $10^2$ 个独特引导序列。在某些实施方案中,文库包含至少 $10^3$ 个独特引导序列。在某些实施方案中,文库包含至少 $10^4$ 个独特引导序列。在某些实施方案中,文库包含至少 $10^5$ 个独特引导序列。在某些实施方案中,文库包含至少 $10^6$ 个独特引导序列。在某些实施方案中,文库包含至少 $10^7$ 个独特引导序列。在某些实施方案中,文库靶向至少10个不同的多核苷酸。在某些实施方案中,文库靶向至少 $10^2$ 个不同的多核苷酸。在某些实施方案中,文库靶向至少 $10^3$ 个不同的多核苷酸。在某些实施方案中,文库靶向至少 $10^4$ 个不同的多核苷酸。在某些实施方案中,文库靶向至少 $10^5$ 个不同的多核苷酸。在某些实施方案中,文库靶向至少 $10^6$ 个不同的多核苷酸。在某些实施方案中,文库靶向至少 $10^7$ 个不同的多核苷酸。

[0169] 在某些实施方案中,文库包含具有相同的序列和相同的修饰的引导RNA的集合,所述相同的序列和相同的修饰位于一个逐步移动的窗口中,此窗口在文库中的全部成员的序列之间移动。在某些实施方案中,这些窗口合起来覆盖RNA的整个长度。

[0170] 在某些实施方案中,文库允许进行高通量、多靶标基因组操作和分析。在某些实施方案中,只有引导RNA的DNA靶向区段是变化的,Cas蛋白结合区段是相同的。在某些实施方案中,文库的第一部分包含具有识别、结合和指导某种特定Cas蛋白的Cas结合区段的引导RNA,而文库的第二部分包含不同的Cas结合区段,其识别、结合和指导不同的Cas蛋白(例如来自不同物种的Cas蛋白),由此允许文库与两种或更多种正交的Cas蛋白共同起作用。在某些实施方案中,第一正交Cas蛋白的诱导表达利用与第一正交Cas蛋白相互作用的文库部分。在某些实施方案中,第一和第二正交Cas蛋白的诱导表达分别利用与第一和第二正交Cas蛋白相互作用的文库部分。在某些实施方案中,第一和第二正交Cas蛋白的诱导表达在不同时间发生。因此,可以通过专门操作或修饰文库中指定的多个靶标来进行大规模的基因编辑或基因调控。

[0171] 在一些实施方案中,文库是“阵列排列”的文库,即可寻址的排列中不同特征或特征池的集合。例如,可以选择性地切割阵列的特征并将其转移到微量滴定板,使得板中的每个孔包含已知特征或已知的特征池。在一些其他的实施方案中,以48孔或96孔微量滴定板格式或在384孔板中合成文库。

[0172] 在某些实施方案中,本发明的引导RNA的合成可以在具有化学实体可以结合的表面的固体支持物上进行。在一些实施方案中,被合成的合成引导RNA直接或间接附接至相同的固体支持物上,并可形成阵列的一部分。“阵列”是已知单体序列的单独分子的集合,每个分子以空间定义和物理寻址的方式排列,使得每个序列的位置是已知的。在本文中可互换使用的“阵列”或“微阵列”包括可寻址区的任何一维、二维或基本二维(以及三维)排列,所



述可寻址区带有特定化学模块或多个模块(如配体,例如与该区域相关联的生物聚合物,如多核苷酸或寡核苷酸序列(核酸)、多肽(例如蛋白质)、碳水化合物、脂质等)。当阵列具有不同模块(例如,不同的多核苷酸序列)的多个区域时,阵列是“可寻址的”,使得在阵列上特定预定位置(即,“地址”)处的区域(即,阵列的“特征”)将检测特定的靶标或靶标类别(虽然特征可能附带地检测到该特征的非靶标)。阵列特征通常(但不一定)通过间隔空间分隔。可包含在阵列上的特征的数目主要由基底的表面积、特征的尺寸和特征之间的间隔决定。阵列可具有多达每 $\text{cm}^2$ 数十万个或更多个特征,如2500至200,000个特征/ $\text{cm}^2$ 的密度。这些特征可以共价键合也可以不共价键合到基底上。

[0173] 适合的固体支持物可具有各种形式和组成,并且衍生自天然存在的材料、已被合成修饰的天然存在的材料或合成材料。适合的支持材料的实例包括但不限于硅石、硅和氧化硅、特氟隆、玻璃、多糖如琼脂糖(例如来自Pharmacia的 **Sepharose®**) 和右旋糖酐(例如也来自Pharmacia的 **Sephadex®** 和 **Sephacyl®**)、聚丙烯酰胺、聚苯乙烯、聚乙烯醇、甲基丙烯酸羟乙酯和甲基丙烯酸甲酯的共聚物等。在一些实施方案中,固体支持物是多个珠子。

[0174] 要在基底表面上合成的引导RNA的初始单体可以与接头结合,接头又与表面亲水基团例如硅石基底上存在的表面羟基模块结合。在一些实施方案中,使用通用接头。在一些其他的实施方案中,初始单体直接与表面羟基模块、表面胺或其他反应性官能团反应。或者,可以首先根据本发明合成引导RNA,并在合成后通过本领域已知的任何方法将其附接至固体基底。因此,本发明可以用于制备引导RNA的阵列,其中寡核苷酸在阵列上合成,或在合成后附接至阵列基底。随后,可以任选地和选择性地将引导RNA或引导RNA的一个池或多个池从阵列基底上切割下来,作为一个文库或多个文库使用。

#### F. crRNA

[0175] 本发明还提供了各种crRNA,其包含如本文所述的针对引导RNA所述的化学修饰。crRNA可以在多区段引导RNA(如双引导)中起作用。因此,在某些实施方案中,本发明提供了合成crRNA,其包含(i)能够与靶多核苷酸杂交的引导序列,所述靶多核苷酸包含邻近PAM位点的靶序列,和(ii)茎序列;其中引导序列由20-N个核苷酸组成,其中N是-10至10之间的整数(任选地在-10至6之间);其中引导序列包含至少一个修饰,并且所述crRNA相比于没有修饰的相应crRNA产生对靶多核苷酸的更高的特异性或更高的gRNA功能性。这里,当crRNA包含在引导RNA或gRNA:Cas蛋白复合物中时,crRNA相比于没有修饰的相应crRNA产生对靶多核苷酸更高的特异性或更高的gRNA功能性,并且所述引导RNA或gRNA:Cas蛋白相比于其中crRNA缺乏修饰的相应引导RNA或gRNA:Cas蛋白复合物具有对靶多核苷酸的更高的特异性或更高的gRNA功能性。

[0176] 感兴趣的修饰在本公开的其他地方描述,包括但不限于在引导序列的核苷酸4-N至20-N处、或在选自4-N、5-N、7-N、9-N、10-N和11-N的至少一个核苷酸处的至少一个修饰。本文中描述了特定修饰和靶多核苷酸的各种实施方案。

#### IV. Cas蛋白

[0177] 如上所述,功能性CRISPR-Cas系统还需要提供期望的活性(如靶结合或靶切刻/切割)的蛋白质组分(例如,Cas蛋白,其可以是Cas核酸酶)。在某些实施方案中,期望的活性是靶结合。在某些实施方案中,期望的活性是靶切刻或靶切割。在某些实施方案中,期望的活

性还包括与本文中公开的Cas蛋白共价融合的多肽所提供的功能。在某些实施方案中,期望的活性还包括与本文中公开的核酸酶缺陷型Cas蛋白共价融合的多肽所提供的功能。这种期望的活性的实例包括如下所述的转录调节活性(激活或抑制)、表观遗传修饰活性或靶标可视化/鉴定活性。可以将Cas蛋白作为纯化或非纯化的(i) Cas蛋白或(ii) 编码用于表达Cas蛋白的mRNA或(iii) 编码用于表达所述蛋白的线性或环状DNA而引入体外或体内系统。提供Cas蛋白的这3种方法在本领域中都是熟知的,并且在本文中提及Cas蛋白或Cas蛋白的用途时都可互换地暗示了这些方法。在某些实施方案中,从mRNA或DNA以组成型方式表达Cas蛋白。在某些实施方案中,从mRNA或DNA表达Cas蛋白是可诱导的或被诱导的。

[0178] 在某些实施方案中,Cas蛋白是化学合成的(参见例如,Creighton,“Proteins: Structures and Molecular Principles,”W.H.Freeman&Co.,NY,1983),或通过本文所述的重组DNA技术产生。关于另外的指南,技术人员可以查阅Frederick M.Ausubel等人,“Current Protocols in Molecular Biology,”John Wiley&Sons,2003;和Sambrook等人,“Molecular Cloning, A Laboratory Manual,”Cold Spring Harbor Press,Cold Spring Harbor, NY,2001)。

[0179] 在某些实施方案中,提供纯化或分离形式的Cas蛋白。在某些实施方案中,提供约80%、约90%、约95%或约99%纯度的Cas蛋白。在某些实施方案中,将Cas蛋白作为组合物的一部分提供。在某些实施方案中,将Cas蛋白提供在水性组合物中,其适合用作用于RNA引导的核酸酶反应的组合物或包含在其中。本领域技术人员清楚地知道可以包含在这样的核酸酶反应组合物中的各种物质。

[0180] 在某些实施方案中,将Cas蛋白作为重组多肽提供。在某些实例中,将重组多肽制备为融合蛋白。例如,在某些实施方案中,编码Cas蛋白的核酸与编码融合配偶体的另一核酸连接,所述融合配偶体例如谷胱甘肽-S-转移酶(GST)、6x-His表位标签或M13基因3蛋白。可以使用适合的宿主细胞来表达融合蛋白。在某些实施方案中,通过本领域已知的方法分离融合蛋白。在某些实施方案中,可以例如通过酶消化进一步处理融合蛋白,以去除所述融合配偶体并获得Cas蛋白。或者,可使用本领域已知的宿主细胞系统或体外翻译-转录系统以重组技术制备Cas蛋白:gRNA复合物。这样的系统和技术的细节可以见于例如W02014144761、W02014144592、W02013176772、US20140273226和US20140273233中,将其内容通过提述完整并入本文。

#### 野生型Cas蛋白

[0181] 在某些实施方案中,Cas蛋白包括衍生自具有RNA引导的多核苷酸结合活性和/或核酸酶活性的CRISPR-Cas I型、II型或III型系统的蛋白质。适合的Cas蛋白的非限制性实例包括Cas3、Cas4、Cas5、Cas5e(或CasD)、Cas6、Cas6e、Cas6f、Cas7、Cas8a1、Cas8a2、Cas8b、Cas8c、Cas9、Cas10、Cas10d、CasF、CasG、CasH、Csy1、Csy2、Csy3、Cse1(或CasA)、Cse2(或CasB)、Cse3(或CasE)、Cse4(或CasC)、Csc1、Csc2、Csa5、Csn2、Csm2、Csm3、Csm4、Csm5、Csm6、Cmr1、Cmr3、Cmr4、Cmr5、Cmr6、Csb1、Csb2、Csb3、Csx17、Csx14、Csx10、Csx16、CsaX、Csx3、Csz1、Csx15、Csf1、Csf2、Csf3、Csf4、和Cu1966。参见例如W02014144761、W02014144592、W02013176772、US20140273226和US20140273233,将其内容通过提述完整并入本文。

[0182] 在某些实施方案中,Cas蛋白衍生自II型CRISPR-Cas系统。在某些实施方案中,Cas蛋白是或衍生自Cas9蛋白。在某些实施方案中,Cas蛋白是或衍生自细菌Cas9蛋白,包括

W02014144761中鉴定的那些。在某些实施方案中,Cas蛋白是或衍生自链球菌属物种(*Streptococcus* sp.)或葡萄球菌属物种(*Staphylococcus* sp.)Cas9蛋白。在某些实施方案中,Cas蛋白是或衍生自嗜热链球菌(*Streptococcus thermophilus*)Cas9蛋白。在某些实施方案中,Cas蛋白是或衍生自酿脓链球菌(*Streptococcus pyogenes*)Cas9蛋白。在某些实施方案中,Cas蛋白是或衍生自金黄色葡萄球菌(*Staphylococcus aureus*)Cas9蛋白。在某些实施方案中,Cas蛋白是或衍生自嗜热链球菌(*Streptococcus thermophilus*)Cas9蛋白。

[0183] 在某些实施方案中,野生型Cas蛋白是Cas9蛋白。在某些实施方案中,野生型Cas9蛋白是来自酿脓链球菌的Cas9蛋白(SEQ ID NO:115)。在某些实施方案中,蛋白质或多肽可以包含SEQ ID NO:115的片段、由SEQ ID NO:115的片段组成或实质上由SEQ ID NO:115的片段组成。

[0184] 通常,Cas蛋白包含与引导RNA相互作用的至少一个RNA结合结构域。在某些实施方案中,修饰Cas蛋白以增加核酸结合亲和力和/或特异性、改变酶活性、和/或改变所述蛋白的另一种特性。例如,可将Cas蛋白的核酸酶(即,DNA酶、RNA酶)结构域修饰、突变、缺失或失活。或者,可以截短Cas蛋白以去除对所述蛋白的功能而言非必需的结构域。在某些实施方案中,将Cas蛋白截短或修饰以优化效应子结构域的活性。在某些实施方案中,Cas蛋白包括核定位序列(NLS),其将带NLS标签的Cas蛋白引入活细胞的核中。在某些实施方案中,Cas蛋白包括两个或更多个修饰。

#### 突变体Cas蛋白

[0185] 在一些实施方案中,Cas蛋白可以是野生型Cas蛋白(如Cas9)的突变体或其片段。在其他实施方案中,Cas蛋白可以衍生自突变体Cas蛋白。例如,可以修饰Cas9蛋白的氨基酸序列以改变所述蛋白的一种或多种特性(例如,核酸酶活性、亲和力、稳定性等)。或者,可以从Cas9蛋白中去除不参与RNA引导的切割的所述蛋白的结构域,使得修饰的Cas9蛋白小于野生型Cas9蛋白。例如,减小Cas9编码序列的大小可使其适配于原本不能容纳野生型序列的转染载体,如AAV载体及其他。在一些实施方案中,本系统利用来自酿脓链球菌的Cas9蛋白,其如在细菌中所编码或经密码子优化以便在真核细胞中表达。以下显示的是野生型酿脓链球菌Cas9蛋白序列的氨基酸序列(SEQ ID NO:115,其可在[www.uniprot.org/uniprot/Q99ZW2](http://www.uniprot.org/uniprot/Q99ZW2)获得)。

```
MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFD  
SGETAETRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVE  
EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMI
```

KFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEEENPINASGVDAKAILSARL  
SKSRLENLIAQLPGEKKNGLFGNLIALLSLGLTPNFKSNFDLAEDAKLQLSKDT  
YDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKR  
YDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP  
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPF  
LKDNREKIEKILTFRIPYVVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGA  
SAQSFIERMTNFDKNLPNEKVLPKHSLLEYFTVYNELTKVKYVTEGMRKPAFL  
SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY  
HDLLKIIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEERLKTYAHLFDDKVM  
KQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDFANRNFMQLIHDDSL  
TFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDELVKVMGRHKP  
ENIVIE**M**ARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEK  
LYLYYLQNGRDMYVDQELDINRLSDYDVD**H**IVPQSFLKDDSIDNKVLTRSDK**N**R  
GKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK  
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQF  
YKVREINNYH**HA**H**D**AYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSE  
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF  
ATVRKVLSPQVNIIVKKTEVQTTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF  
DSPTVAYSVLVVAKEKGKSKKLKSVKELLGITIMERSSSFENPIDFLEAKGYK  
EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY  
EKLKGSPEDEQKQLFVEQHKHYLDEIIIEQISEFSKRVLADANLDKVL SAYNK  
HRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ  
SITGLYETRIDLSQLGGD

[0186] Cas9蛋白通常具有至少两个核酸酶(例如DNA酶)结构域。例如,Cas9蛋白可以具有RuvC样核酸酶结构域和HNH样核酸酶结构域。RuvC和HNH结构域一起运作,在靶位点切断两条链,从而在靶多核苷酸中产生双链断裂。(Jinek等人,Science 337:816-21)。在某些实施方案中,将突变型Cas9蛋白修饰为仅含有一个功能性核酸酶结构域(RuvC样或HNH样核酸酶结构域)。例如,在某些实施方案中,将突变型Cas9蛋白修饰为使得核酸酶结构域之一缺失或突变,以至于它不再起作用(即,核酸酶活性不存在)。在其中一个核酸酶结构域无活性的一些实施方案中,突变体能够将切口引入双链多核苷酸中(这种蛋白称为“切口酶”),但不能切割双链多核苷酸。例如,RuvC样结构域中天冬氨酸到丙氨酸(D10A)的转换将Cas9衍生的蛋白转化成切口酶。同样,在HNH结构域中组氨酸到丙氨酸(H840A)的转换将Cas9衍生的蛋白转化成切口酶。同样,在HNH结构域中天冬酰胺到丙氨酸(N863A)的转换将Cas9衍生的蛋白转化成切口酶。

[0187] 在某些实施方案中,RuvC样核酸酶结构域和HNH样核酸酶结构域两者都被修饰或

消除,使得突变体Cas9蛋白不能切刻或切割靶多核苷酸。在某些实施方案中,Cas9衍生的蛋白的所有核酸酶结构域都被修饰或消除,使得Cas9衍生的蛋白缺乏所有的核酸酶活性。在某些实施方案中,相对于对应的野生型,缺乏一些或所有的核酸酶活性的Cas9蛋白仍然或多或少地维持了靶识别活性。

[0188] 在任何上述实施方案中,可以使用熟知的方法,如定点诱变、PCR介导的诱变和总基因合成,以及本领域已知的其他方法,通过一个或多个缺失突变、插入突变和/或取代突变使任何或所有的核酸酶结构域失活。

[0189] 在某些实施方案中,“Cas突变体”或“Cas变体”与SEQ ID NO:115至少50% (例如在50%至100%之间的任何数值,包括例如50%、60%、70%、80%、90%、95%、98%和99%) 相同。在某些实施方案中,“Cas突变体”或“Cas变体”与RNA分子(例如sgRNA)结合。在某些实施方案中,“Cas突变体”或“Cas变体”介由所述RNA分子靶向特定的多核苷酸序列。

### 融合蛋白

[0190] 在某些实施方案中,Cas蛋白与另一种与Cas蛋白异源的蛋白质或多肽融合以产生融合蛋白。在某些实施方案中,异源序列包括一个或多个效应子结构域,如切割结构域、转录激活结构域、转录阻抑物结构域或表观遗传修饰结构域。另外的效应子结构域的实例包括核定位信号、细胞穿透或转位结构域或标志物结构域。在某些实施方案中,效应子结构域位于融合蛋白的N末端、C末端或内部位置。在某些实施方案中,融合蛋白的Cas蛋白是或衍生自Cas9蛋白。在某些实施方案中,融合蛋白的Cas蛋白是或衍生自其中所有核酸酶结构域已被失活或缺失的修饰的或突变的Cas蛋白。在某些实施方案中,融合蛋白的Cas蛋白是或衍生自缺乏核酸酶活性的修饰的或突变的Cas蛋白。在某些实施方案中,Cas蛋白的RuvC和/或HNH结构域被修饰或突变,使得它们不再具有核酸酶活性。

### 切割结构域

[0191] 在某些实施方案中,融合蛋白的效应子结构域是切割结构域。如本文中使用的,“切割结构域”是指可切割DNA的结构域。切割结构域可从任何核酸内切酶或核酸外切酶获得。可衍生切割结构域的核酸内切酶的非限制性实例包括限制性核酸内切酶和归巢核酸内切酶。参见,例如NewEngland Biolabs Catalog或Belfort等人(1997) *Nucleic Acids Res.* 25,3379-88。切割DNA的其他酶是已知的(例如51核酸酶;绿豆核酸酶;胰腺DNA酶I;微球菌核酸酶;酵母H0核酸内切酶)。还参见Linn等人(编辑) “Nucleases,” ColdSpringHarbor Laboratory Press,1993。可将一种或多种这些酶(或其功能性片段)用作切割结构域的来源。

[0192] 在某些实施方案中,切割结构域可以衍生自II-S型核酸内切酶。II-S型核酸内切酶特异性地在通常距离核酸内切酶的DNA识别位点几个碱基对的位点处切割DNA,因此具有可分离的识别和切割结构域。这些酶通常是单体,所述单体瞬时结合形成二聚体以在交错位置切割DNA的每条链。适合的II-S型核酸内切酶的非限制性实例包括BfiI、BpmI、BsaI、BsgI、BsmBI、BsmI、BspMI、FokI、MbolI和SapI。在某些实施方案中,融合蛋白的切割结构域是FokI切割结构域或其片段或衍生物。参见Miller等人(2007) *Nat. Biotechnol.* 25,778-85; Szczpek等人(2007) *Nat. Biotechnol.* 25,786-93; Doyon等人(2011) *Nat. Methods*, 8,74-81。

### 转录激活结构域

[0193] 在某些实施方案中,融合蛋白的效应子结构域是转录激活结构域。一般而言,转录激活结构域与转录控制元件和/或转录调节蛋白(即转录因子, RNA聚合酶等)相互作用以增加和/或激活基因的转录。在某些实施方案中,转录激活结构域是单纯疱疹病毒VP16激活结构域、VP64(其为VP16的四聚体衍生物)、NF $\kappa$ B p65激活结构域、p53激活结构域1和2、CREB(cAMP反应元件结合蛋白)激活结构域、E2A激活结构域或NFAT(活化T细胞的核因子)激活结构域。在某些实施方案中,转录激活结构域是Gal4、Gcn4、MLL、Rtg3、Gln3、Oaf1、Pip2、Pdr1、Pdr3、Pho4或Leu3。转录激活结构域可以是野生型,或者它可以是原始转录激活结构域的修饰或截短版本。

#### 转录阻抑物结构域

[0194] 在某些实施方案中,融合蛋白的效应子结构域是转录阻抑物结构域。一般而言,转录阻抑物结构域与转录控制元件和/或转录调节蛋白(即转录因子, RNA聚合酶等)相互作用以降低和/或阻止基因的转录。在某些实施方案中,转录阻抑物结构域是诱导型cAMP早期阻抑物因子(ICER)结构域、Kruppel相关盒A(KRAB-A)阻抑物结构域、YY1富含甘氨酸的阻抑物结构域、Sp1样阻抑物、E(sp1)阻抑物、I $\kappa$ B阻抑物或MeCP2。

#### 表观遗传修饰结构域

[0195] 在某些实施方案中,融合蛋白的效应子结构域是表观遗传修饰结构域。一般而言,表观遗传修饰结构域通过修饰组蛋白结构和/或染色体结构来改变基因表达。在某些实施方案中,表观遗传修饰结构域是组蛋白乙酰转移酶结构域、组蛋白去乙酰化酶结构域、组蛋白甲基转移酶结构域、组蛋白去甲基化酶结构域、DNA甲基转移酶结构域或DNA去甲基化酶结构域。

#### 另外的结构域

[0196] 在某些实施方案中,融合蛋白进一步包含至少一个另外的结构域。适合的另外的结构域的非限制性实例包括核定位信号(NLS)、细胞穿透或转位结构域和标志物结构域。NLS通常包含一段碱性氨基酸。参见例如Lange等人(2007) J. Biol. Chem. 282, 5101-5。例如,在某些实施方案中,NLS是单分型(monopartite)序列,如PKKKRKV(SEQ ID NO:116)或PKKKRRV(SEQ ID NO:117)。在某些实施方案中,NLS是双分型(bipartite)序列。在某些实施方案中,NLS是KRPAATKKAGQAKKKK(SEQ ID NO:118)。

[0197] 在某些实施方案中,融合蛋白包含至少一个细胞穿透结构域。在某些实施方案中,细胞穿透结构域是衍生自HIV-1 TAT蛋白的细胞穿透肽序列。例如,TAT细胞穿透序列可以是GRKKRRQRRRPPQPKKKRKV(SEQ ID NO:119)。在某些实施方案中,细胞穿透结构域是TLM(PLSSIFSRIGDPPKKKKRKV;SEQ ID NO:120),其为衍生自人乙型肝炎病毒的细胞穿透肽序列。在某些实施方案中,细胞穿透结构域是MPG(GALFLGWLGAAGSTMGAPKKKKRKV;SEQ ID NO:121或GALFLGFLGAAGSTMGAWSQPKKKRKV;SEQ ID NO:122)。在某些实施方案中,细胞穿透结构域是Pep-1(KETWWETWWTEWSQPKKKRKV;SEQ ID NO:123)、VP22(来自单纯疱疹病毒的细胞穿透肽)或聚精氨酸肽序列。

[0198] 在某些实施方案中,融合蛋白包含至少一个标志物结构域。标志物结构域的非限制性实例包括荧光蛋白、纯化标签和表位标签。在某些实施方案中,标志物结构域是荧光蛋白。适合的荧光蛋白的非限制性实例包括绿色荧光蛋白(例如GFP、GFP-2、tagGFP、turboGFP、EGFP、祖母绿(Emerald)、Azami Green、单体Azami Green、CopGFP、AceGFP、



ZsGreen1)、黄色荧光蛋白(例如YFP、EYFP、Citrine、Venus、YPet、PhiYFP、ZsYellow1)、蓝色荧光蛋白(例如EBFP、EBFP2、Azurite、mKalamal、GFPuv、Sapphire、T-sapphire)、青色荧光蛋白(例如ECFP、Cerulean、CyPet、AmCyan1、Midoriishi-Cyan)、红色荧光蛋白(mKate、mKate2、mPlum、DsRed单体、mCherry、mRFP1、DsRed-Express、DsRed2、DsRed-单体、HcRed-串联、HcRed1、AsRed2、eqFP611、mRaspberry、mStrawberry、Jred)、橙色荧光蛋白(mOrange、mK0、Kusabira-Orange、单体Kusabira-Orange、mTangerine、tdTomato)和任何其他适合的荧光蛋白。在某些实施方案中,标志物结构域是纯化标签和/或表位标签。示例性标签包括但不限于谷胱甘肽-S-转移酶(GST)、甲壳素结合蛋白(CBP)、麦芽糖结合蛋白、硫氧还蛋白(TRX)、聚(NANP)、串联亲和纯化(TAP)标签、myc、AcV5、AU1、AU5、E、ECS、E2、FLAG、HA、nus、Softag 1、Softag 3、Strep、SBP、Glu-Glu、HSV、KT3、S、S1、T7、V5、VSV-G、6xHis、生物素羧基载体蛋白(BCCP)和钙调蛋白。

#### V. 用途和方法

[0199] 在一个方面,本发明提供了用Cas蛋白切割靶多核苷酸的方法。所述方法包括使所述靶多核苷酸与(i)本文所述的引导RNA或一组引导RNA分子和(ii)Cas蛋白接触。在某些实施方案中,所述方法导致靶多核苷酸中的双链断裂。在某些实施方案中,Cas蛋白是具有单链切口活性的Cas蛋白。在某些实施方案中,所述方法导致靶多核苷酸中的单链断裂。在某些实施方案中,将包含引导RNA和具有单链切口活性的Cas蛋白的复合物用于序列靶向的单链DNA切割,即切割。

[0200] 在一个方面,本发明提供了用Cas蛋白切割两个或更多个靶多核苷酸的方法。所述方法包括使所述靶多核苷酸与(i)本文所述的一组引导RNA分子和(ii)Cas蛋白接触。在某些实施方案中,所述方法导致靶多核苷酸中的双链断裂。在某些实施方案中,Cas蛋白是具有单链切口活性的Cas蛋白。在某些实施方案中,所述方法导致靶多核苷酸中的单链断裂。在某些实施方案中,将包含引导RNA和具有单链切口活性的Cas蛋白的复合物用于序列靶向的单链DNA切割,即切割。

[0201] 在一个方面,本发明提供了用Cas蛋白结合靶多核苷酸的方法。所述方法包括使所述靶多核苷酸与(i)本文所述的引导RNA或一组引导RNA分子和(ii)Cas蛋白接触,从而导致靶多核苷酸与Cas蛋白的结合。在某些实施方案中,Cas蛋白是Cas变体。在某些实施方案中,相对于对应物野生型Cas蛋白,Cas变体缺乏一些或所有的核酸酶活性。

[0202] 在一个方面,本发明提供了使两个或更多个靶多核苷酸与Cas蛋白结合的方法。所述方法包括使所述靶多核苷酸与(i)本文所述的一组RNA分子和(ii)Cas蛋白接触,从而导致靶多核苷酸与Cas蛋白的结合。在某些实施方案中,Cas蛋白是Cas变体。在某些实施方案中,相对于对应的野生型Cas蛋白,Cas变体缺乏一些或所有的核酸酶活性。

[0203] 在一个方面,本发明提供了将Cas蛋白靶向至靶多核苷酸的方法。所述方法包括使Cas蛋白与本文所述的引导RNA或一组引导RNA分子接触。在某些实施方案中,所述方法导致引导RNA:Cas蛋白复合物的形成。在某些实施方案中,Cas蛋白是野生型Cas9蛋白。在某些实施方案中,Cas蛋白是Cas9蛋白的突变体或变体。在某些实施方案中,Cas蛋白是具有单链切口活性的Cas蛋白。在某些实施方案中,Cas蛋白是缺乏核酸酶活性的Cas蛋白(例如,Cas蛋白的核酸酶缺陷型突变体)。在某些实施方案中,Cas蛋白是融合蛋白(例如,包含(i)Cas蛋白和(ii)异源多肽的融合蛋白)的一部分。



[0204] 在一个方面,本发明提供了用于将Cas蛋白靶向至两个或更多个靶多核苷酸的方法。所述方法包括使Cas蛋白与本文所述的一组引导RNA分子接触。在某些实施方案中,所述方法导致引导RNA:Cas蛋白复合物的形成。在某些实施方案中,Cas蛋白是野生型Cas9蛋白。在某些实施方案中,Cas蛋白是Cas9蛋白的突变体或变体。在某些实施方案中,Cas蛋白是具有单链切口活性的Cas蛋白。在某些实施方案中,Cas蛋白是缺乏核酸酶活性的Cas蛋白(例如,Cas蛋白的核酸酶缺陷型突变体)。在某些实施方案中,Cas蛋白是融合蛋白(例如,包含(i)Cas蛋白和(ii)异源多肽的融合蛋白)的一部分。

[0205] 在第一个方面,本发明提供了选择合成引导RNA的方法。所述方法涉及使MP修饰在gRNA的整个引导序列部分上“步移”,以鉴定引导序列中的哪个或哪些位置可由于MP修饰的定位而增强特异性。利用中靶与脱靶切割比值、在靶位点处的切割百分比和在一个或多个脱靶位点处的切割百分比、和/或特异性评分,可以对每个测试位置评估特异性增强的幅度,从而确定哪个或哪些修饰的位置改变gRNA的特异性以及改变的程度。单个MP在整个引导序列上的渐进步移还可以鉴定由于测试位置中的一个或多个化学修饰组合所致的特异性的潜在协作改善的位置。

[0206] 在一个实施方案中,所述方法包括提供至少第一合成引导RNA和第二合成引导RNA,两者都包含相同的以下区段的序列:(a) crRNA区段,其包含(i)能够与靶多核苷酸杂交的引导序列、(ii)茎序列;(b) tracrRNA区段,其包含与所述茎序列部分或完全互补的核苷酸序列,其中所述第一合成引导RNA包含在所述引导序列的第一位置处的MP修饰,并且所述第二合成引导RNA包含在所述引导序列的第二位置处的MP修饰;形成包含Cas蛋白和所述第一合成引导RNA的第一gRNA:Cas蛋白复合物,使所述靶多核苷酸与所述第一gRNA:Cas蛋白复合物接触,并切割、切刻或结合所述靶多核苷酸;形成包含Cas蛋白和所述第二合成引导RNA的第二gRNA:Cas蛋白复合物,使所述靶多核苷酸与所述第二gRNA:Cas蛋白复合物接触,并切割、切刻或结合所述靶多核苷酸;确定所述第一gRNA:Cas蛋白复合物和所述第二gRNA:Cas蛋白复合物在所述靶多核苷酸的切割、切刻或结合中的特异性;鉴定哪个引导RNA对所述靶多核苷酸具有更好的特异性。在某些实施方案中,将第一和第二gRNA:Cas蛋白复合物在竞争性测定中一起测试,例如通过用不同的荧光团标记第一和第二gRNA。在某些实施方案中,在平行或依次测定的等效或分割样品中,各自测试第一和第二gRNA:Cas蛋白复合物。

[0207] 在一个方面,本发明提供了用于分析通过使用引导RNA(例如包含特异性增强修饰的引导RNA)进行的CRISPR功能的特异性的方法。该方法包括:鉴定样品中的靶序列和一个或多个脱靶序列,其中靶序列包含在靶多核苷酸中,并且脱靶序列包含在一个或多个脱靶多核苷酸中;使用引导RNA执行CRISPR功能,通过使用寡核苷酸诱饵文库(该文库设计为与靶标和脱靶多核苷酸杂交)捕获靶标和脱靶多核苷酸,分离和分析捕获的多核苷酸以评估靶标和脱靶序列是否已经由于CRISPR功能而被改变。靶序列和非靶序列之间的相对变化程度指示由引导RNA介导的CRISPR功能的特异性。在一些实施方案中,通过测序分析捕获的多核苷酸。

[0208] 多核苷酸诱饵包括设计用于与靶多核苷酸杂交的诱饵和设计用于与脱靶多核苷酸杂交的诱饵。预期诱饵可以直接与靶(或脱靶)序列杂交,或者可以与靶(或非靶)序列附近的序列杂交,例如位于靶(或脱靶)序列的5'或/和3',距靶(或脱靶)序列2000,1500,1000,900,800,700,600,500,400,300,200,100,90,80,70,60,50,40,30或20个碱基对以内

的序列。一些诱饵可以部分地与靶(或脱靶)序列杂交,并且部分在靶(或脱靶)序列之外杂交。在一些实施方案中,诱饵文库包含与靶(或脱靶)序列杂交的诱饵和在靶(或脱靶)序列附近杂交的诱饵。在一些实施方案中,文库仅包含在靶(或脱靶)序列附近杂交的诱饵,不包含与靶(或脱靶)序列杂交的诱饵,或反之。在一些实施方案中,诱饵与以靶序列为中心的约1000bp的区域杂交,以及以每个脱靶序列为中心的约1000bp的区域杂交。在一些实施方案中,诱饵是层叠的(tiled)。

[0209] 在某些实施方案中,增强特异性的修饰包括2'-O-甲基-3'-磷酸基乙酸(MP)、2'-O-甲基-3'-硫代磷酸基乙酸(MSP)、2'-脱氧-3'-磷酸基乙酸(DP)、2'-脱氧-3'-硫代磷酸基乙酸(DSP),或其组合。在某些实施方案中,化学修饰包括赋予C3'-内糖皱褶的2'修饰和磷酸基乙酸或硫代磷酸基乙酸键修饰。在某些实施方案中,2'修饰选自2'-F和2'-O-(2-甲氧基乙基)。在某些实施方案中,第一和第二合成引导RNA包含在引导序列部分中的不同核苷酸位置处的增强特异性的修饰。在某些实施方案中,根据中靶切割活性、脱靶切割活性、中靶:脱靶比、特异性评分或其组合确定特异性。在某些实施方案中,所述方法包括提供包含在引导序列部分中的不同核苷酸位置处的增强特异性的修饰的第一至第二十合成引导RNA,使用每个合成引导RNA形成gRNA:Cas蛋白复合物,使靶多核苷酸与所述gRNA:Cas蛋白复合物接触,切割、切刻或结合所述靶多核苷酸并测量每个合成引导RNA的特异性,并且鉴定提供最大特异性增强的一个或多个修饰位置。在某些实施方案中,gRNA进一步包含增强稳定性的末端修饰。在某些实施方案中,增强稳定性的末端修饰包括在gRNA的5'端和/或3'端的2'-O-甲基-3'-磷酸基乙酸(MP)、2'-O-甲基-3'-硫代磷酸基乙酸(MSP)、2'-O-甲基-3'-硫代磷酸(MS)、2'-脱氧-3'-磷酸基乙酸(DP)、2'-脱氧-3'-硫代磷酸基乙酸(DSP)、2'-氟-3'-磷酸基乙酸(FP)、2'-氟-3'-硫代磷酸基乙酸(FSP)、2'-氟-3'-硫代磷酸(FS)或其组合。

[0210] 在某些实施方案中,通过转染将引导RNA引入到细胞中。用于RNA转染的技术在本领域中是已知的,包括电穿孔和脂质转染。RNA转染的有效技术主要取决于细胞类型。参见,例如Lujambio等人(西班牙国家癌症中心)Cancer Res.2007年2月,其描述了HTC-116结肠癌细胞的转染,并且使用用于转染可商业获得的、修饰的miRNA或前体miRNA的Oligofectamine(Invitrogen)。还参见Cho等人(Seoul National Univ.)Nat.Biotechnol.,2013年3月,其描述了K562细胞的转染,并且使用用于转染转录的sgRNA(约60个nt长)的4D Nucleofection™(Lonza)电穿孔。用于转染RNA的技术也是本领域已知的。例如,已经将治疗性RNA递送到用侵袭素(Invasin)蛋白包被的非致病性大肠杆菌中(以促进摄取到表达 $\beta$ -1整联蛋白的细胞中),其中大肠杆菌被编码以表达李斯特菌溶血素(lysteriolysin)成孔蛋白以允许shRNA从大肠杆菌转入细胞质。还参见Cho等人,(Seoul National Univ.)Nat.Biotechnol.Mar.2013。

[0211] 在某些实施方案中,将引导RNA引入或递送到细胞中。可用于递送引导RNA的技术包括利用可生物降解聚合物、脂质体或纳米颗粒进行包封的技术。这样的聚合物、脂质体和纳米颗粒可以经静脉内递送。在某些实施方案中,对于体内递送,可以将引导RNA注射到组织部位中或全身施用。体内递送也可通过 $\beta$ -葡聚糖递送系统来实现,如美国专利No.5,032,401和5,607,677以及美国公开号2005/0281781中所述的那些,将其内容通过提述完整并入本文。在某些实施方案中,将引导RNA或含有引导RNA的递送载体靶向至特定的组织或身体

隔壁。例如,在某些实施方案中,为了将外源RNA靶向其他组织,将合成载体用针对受体摄取的细胞特异性配体或适体进行装饰,所述装饰例如将RNA包裹在用PEG包被的环糊精纳米颗粒中,并且用人转铁蛋白功能化,以便通过在肿瘤细胞中高表达的转铁蛋白受体进行摄取。其他方法描述于下文或是本领域已知的。

[0212] 本发明已经在人细胞中进行了测试,如描述于Hendel等人(2015) Nat.Biotechnol.33:9,985-9(将其完整并入本申请中)。在引用的文献中,将修饰的引导RNA引入K562细胞、人原代T细胞以及CD34+造血干细胞和祖细胞(HSPC)中。与未修饰的引导RNA相比,修饰的引导RNA显著增强了人细胞(包括人原代T细胞和CD34+HSPC)中的基因组编辑效率。

[0213] 其他用途的实例包括如下所述的基因组编辑和基因表达调控。

#### 基因组编辑

[0214] 在一个方面,本发明提供了用于在体内或体外修饰DNA序列的基因组编辑方法(“体外”包括但不限于无细胞系统、细胞裂解物、细胞的分离组分、和活体外的细胞)。DNA序列可以包括染色体序列、游离型序列、质粒、线粒体DNA序列或功能性基因间序列,如增强子序列或针对非编码RNA的DNA序列。所述方法包括使所述DNA序列与(i)本文所述的引导RNA或一组引导RNA分子和(ii)Cas蛋白接触。在某些实施方案中,使DNA序列接触细胞外部。在某些实施方案中,DNA序列位于细胞内的基因组中并在体外或体内接触。在某些实施方案中,细胞在生物体或组织内。在某些实施方案中,所述细胞是人细胞、非人哺乳动物细胞、干细胞、非哺乳动物脊椎动物细胞、无脊椎动物细胞、植物细胞、单细胞生物或胚胎。在某些实施方案中,引导RNA有助于将Cas蛋白靶向至DNA序列中的靶向位点。在某些实施方案中,Cas蛋白在靶向位点处切割DNA序列的至少一条链。在某些实施方案中,Cas蛋白在靶向位点处切割DNA序列的两条链。

[0215] 在某些实施方案中,所述方法进一步包括将Cas蛋白引入到细胞或另一个系统中。在某些实施方案中,将Cas蛋白作为纯化或非纯化蛋白引入。在某些实施方案中,通过编码Cas蛋白的mRNA引入Cas蛋白。在某些实施方案中,通过编码Cas蛋白的线性或环状DNA引入Cas蛋白。在某些实施方案中,所述细胞或系统包含Cas蛋白或编码Cas蛋白的核酸。

[0216] 在某些实施方案中,可以通过易错非同源末端连接(“NHEJ”)修复过程来修复双链断裂。在某些实施方案中,可以通过同源定向修复(HDR)过程来修复双链断裂,使得供体多核苷酸中的供体序列可以整合到靶向的DNA序列中或与其交换。

[0217] 在某些实施方案中,所述方法还包括将至少一个供体多核苷酸引入到细胞或系统中。在某些实施方案中,供体多核苷酸包含至少一个与DNA序列中的靶向位点任一侧上的序列具有基本序列同一性的同源序列。在某些实施方案中,供体多核苷酸包括这样的供体序列,其可通过同源定向修复如同源重组整合到DNA序列中或与其交换。

[0218] 在某些实施方案中,供体多核苷酸包含上游同源序列和下游同源序列,其各自与分别位于DNA序列中靶向位点的上游和下游的序列具有实质序列同一性。这些序列相似性允许例如供体多核苷酸与靶向的DNA序列之间的同源重组,使得供体序列可以整合到靶向的DNA序列中(或与其交换)。

[0219] 在某些实施方案中,DNA序列中的靶位点跨越或邻近于可能引起疾患或与疾患相关的突变,例如点突变、易位或倒位。在某些实施方案中,所述方法包括通过将至少一个供

体多核苷酸引入细胞或系统中来纠正所述突变,所述供体多核苷酸包含(i)所述突变的对应的野生型和(ii)至少一个与DNA序列中的靶向位点一侧上的序列具有实质序列同一性的同源序列。在某些实施方案中,供体多核苷酸包含与DNA序列中的靶向位点两侧上的序列具有实质序列同一性的同源序列。

[0220] 在某些实施方案中,供体多核苷酸包括外源序列,所述外源序列可通过同源定向修复过程(如同源重组)整合到靶向的DNA序列中或与其交换。在某些实施方案中,外源序列包含蛋白质编码基因,蛋白质编码基因任选地与外源启动子控制序列可操作连接。因此,在某些实施方案中,在整合外源序列后,细胞可以表达由整合的基因编码的蛋白质。在某些实施方案中,将外源序列整合到靶向的DNA序列中,使得其在受体细胞或系统中的表达受到外源启动子控制序列的调控。将外源基因整合到靶向的DNA序列中称为“敲入”。在其他实施方案中,外源序列可以是转录控制序列、另一个表达控制序列、RNA编码序列,等等。

[0221] 在某些实施方案中,供体多核苷酸包含与靶向位点处或附近的DNA序列的一部分基本相同、但包含至少一个核苷酸改变的序列。例如,在某些实施方案中,供体序列包含在靶向位点处或附近的修饰或突变版本的DNA序列,使得在与靶向位点整合或交换后,在靶向位点处产生的序列包含至少一个核苷酸变化。在某些实施方案中,所述至少一个核苷酸变化是一个或多个核苷酸的插入、一个或多个核苷酸的缺失、一个或多个核苷酸的取代或其组合。作为修饰序列的整合的结果,细胞可以从靶向的DNA序列产生修饰的基因产物。

[0222] 在某些实施方案中,所述方法用于多重应用中。在某些实施方案中,所述方法包括将引导RNA文库引入细胞或系统中。在某些实施方案中,文库包含至少10个独特的引导序列。在某些实施方案中,文库包含至少100个独特的引导序列。在某些实施方案中,文库包含至少1,000个独特的引导序列。在某些实施方案中,文库包含至少10,000个独特的引导序列。在某些实施方案中,文库包含至少100,000个独特的引导序列。在某些实施方案中,文库包含至少1,000,000个独特的引导序列。在某些实施方案中,文库靶向至少10个不同的多核苷酸或者一个或多个多核苷酸内的至少10个不同的序列。在某些实施方案中,文库靶向至少100个不同的多核苷酸或者一个或多个多核苷酸内的至少100个不同的序列。在某些实施方案中,文库靶向至少1,000个不同的多核苷酸或者一个或多个多核苷酸内的至少1,000个不同的序列。在某些实施方案中,文库靶向至少10,000个不同的多核苷酸或者一个或多个多核苷酸内的至少10,000个不同的序列。在某些实施方案中,文库靶向至少100,000个不同的多核苷酸或者一个或多个多核苷酸内的至少100,000个不同的序列。在某些实施方案中,文库靶向至少1,000,000个不同的多核苷酸或者一个或多个多核苷酸内的至少1,000,000个不同的序列。人和哺乳动物细胞中的基因组编辑

[0223] 本发明实施方案可用于基因组编辑的方法中,以修饰哺乳动物细胞中的靶多核苷酸(例如DNA序列)。

[0224] 在某些实施方案中,DNA序列是染色体序列。在某些实施方案中,DNA序列是蛋白质编码序列。在某些实施方案中,DNA序列是功能性基因间序列,如增强子序列或非编码序列。在某些实施方案中,DNA是人基因的一部分。在一些这样的实施方案中,人基因是网格蛋白轻链(CLT A1)基因、人白细胞介素2受体 $\gamma$ (IL2RG)基因、人细胞毒性T淋巴细胞相关蛋白4(CLT A4)基因、人血管内皮生长因子A基因(VEGFA)、或包含引起镰状细胞贫血和地中海贫血的突变的人血红蛋白 $\beta$ (HBB)基因。因此,在某些实施方案中,靶多核苷酸是HBB多核苷酸、

VEGFA多核苷酸、IL2RG多核苷酸、CLTA1多核苷酸或CLTA4多核苷酸。

[0225] 在某些实施方案中,合成引导RNA包含能够与HBB、IL2RG、CLTA1、VEGFA或CLTA4多核苷酸杂交的引导序列。在某些实施方案中,所述引导序列由从所述引导序列的5'端计数的核苷酸1至20-N组成,N是-10至0之间的整数,并且引导序列在核苷酸4-N、5-N、7-N、9-N、10-N或11-N处包含至少一个增强特异性的修饰。在某些实施方案中,能够与上述靶多核苷酸之一杂交的引导序列在核苷酸11-N处具有化学修饰。在某些实施方案中,能够与上述靶多核苷酸之一杂交的引导序列在核苷酸5-N处具有化学修饰。在某些实施方案中,能够与上述靶多核苷酸之一杂交的引导序列在核苷酸7-N处具有化学修饰。在某些实施方案中,能够与上述靶多核苷酸之一杂交的引导序列在核苷酸10-N处具有化学修饰。在某些实施方案中,能够与上述靶多核苷酸之一杂交的引导序列在核苷酸9-N处具有化学修饰。在某些实施方案中,能够与上述靶多核苷酸之一杂交的引导序列在核苷酸4-N处具有化学修饰。在某些实施方案中,N等于零。

[0226] 在某些实施方案中,所述引导序列由从所述引导序列的5'端计数的核苷酸1至19、以及在核苷酸3、4、6、8、9或10中至少一个核苷酸处的至少一个化学修饰组成。在某些实施方案中,所述引导序列由从所述引导序列的5'端计数的核苷酸1至18、以及在核苷酸2、3、5、7、8或9中至少一个核苷酸处的至少一个化学修饰组成。在某些实施方案中,所述引导序列由从所述引导序列的5'端计数的核苷酸1至17、以及在核苷酸1、2、4、6、7或8中至少一个核苷酸处的至少一个化学修饰组成。在某些实施方案中,所述引导序列由从所述引导序列的5'端计数的核苷酸1至16、以及在核苷酸1、3、5、6或7中至少一个核苷酸处的至少一个化学修饰组成。在某些实施方案中,所述引导序列由从所述引导序列的5'端计数的核苷酸1至15、以及在核苷酸2、4、5或6中至少一个核苷酸处的至少一个化学修饰组成。在某些实施方案中,所述引导序列由从所述引导序列的5'端计数的核苷酸1至14、以及在核苷酸1、3、4或5中至少一个核苷酸处的至少一个化学修饰组成。在某些实施方案中,所述化学修饰包括2'-O-甲基-3'-磷酸基乙酸(MP)、2'-O-甲基-3'-硫代磷酸基乙酸(MSP)、2'-脱氧-3'-磷酸基乙酸(DP)、2'-脱氧-3'-硫代磷酸基乙酸(DSP),或其组合。在某些实施方案中,所述化学修饰包括赋予C3'-内糖皱褶的2'修饰和磷酸基乙酸或硫代磷酸基乙酸键修饰。在某些实施方案中,2'修饰选自2'-F和2'-O-(2-甲氧基乙基)。

[0227] 在某些实施方案中,哺乳动物细胞是人类细胞。在一些这样的实施方案中,人类细胞是原代人类细胞。在进一步的实施方案中,原代人类细胞是人类原代T细胞。可以刺激或不刺激人类原代T细胞。在某些实施方案中,人类细胞是干/祖细胞,如CD34+造血干细胞和祖细胞(HSPC)。在某些实施方案中,人类细胞来自培养的细胞系,例如可以商业获得。示例性细胞系包括K562细胞,其为人髓性白血病细胞系。

[0228] 在某些实施方案中,细胞在活生物体内。在某些其他的实施方案中,细胞在活生物体之外。

[0229] 所述方法包括使所述DNA序列与(i)本文所述的引导RNA或一组引导RNA分子和(ii)Cas蛋白接触。

[0230] 在某些实施方案中,所述方法进一步包括将引导RNA引入或递送到细胞中。在一些这样的实施方案中,通过转染将引导RNA引入到细胞中。用于RNA转染的技术在本领域中是已知的,包括电穿孔和脂质体转染。在其他实施方案中,通过核转染将引导RNA引入细胞(更

具体地,细胞核)中。用于核转染的技术是本领域已知的,并且可以利用核转染装置,如Lonza Nucleofector 2b或Lonza 4D-Nucleofector和相关试剂。

[0231] 在某些实施方案中,所述方法进一步包括将Cas蛋白引入或递送到细胞中。在一些这样的实施方案中,作为纯化或非纯化蛋白引入Cas蛋白。在其他实施方案中,通过编码Cas蛋白的mRNA引入Cas蛋白。在一些这样的实施方案中,通过转染将编码Cas蛋白的mRNA引入细胞中。在其他实施方案中,通过核转染将编码Cas蛋白的mRNA引入细胞(更具体地,细胞核)中。

[0232] 在某些实施方案中,所述方法使用基于核糖核蛋白(RNP)的递送,使得Cas蛋白与引导RNA一起以复合物的形式引入细胞中。例如,Cas9蛋白可以与Cas9:gRNA复合物中的引导RNA复合,其允许共同递送gRNA和Cas蛋白。例如,Cas:gRNA复合物可以通过核转染进入细胞中。

[0233] 在某些实施方案中,所述方法使用全RNA递送平台。例如,在一些这样的实施方案中,将引导RNA和编码Cas蛋白的mRNA同时或实质上同时(例如,通过共转染或共核转染)引入细胞中。在某些实施方案中,与Cas mRNA和未修饰的gRNA的共同递送相比,Cas mRNA和修饰的gRNA的共同递送导致更高的编辑频率。具体而言,与未修饰的gRNA相比,在5'和3'端两者的三个末端核苷酸处掺入有2'-O-甲基-3'-硫代磷酸("MS")、2'-O-甲基-3'-PACE("MP")或2'-O-甲基-3'-硫代PACE(MSP)的gRNA可提供更高的编辑频率。

[0234] 在某些实施方案中,将引导RNA和编码Cas蛋白的mRNA依次引入细胞中;即,将引导RNA和编码Cas蛋白的mRNA在不同的时间引入细胞中。每种试剂引入之间相隔的时段可在几分钟(或更少)至几小时或几天的范围内。例如,在一些这样的实施方案中,首先递送gRNA,然后在4、8、12或24小时后递送Cas mRNA。在其他这样的实施方案中,首先递送Cas mRNA,然后在4、8、12或24小时后递送gRNA。在一些具体实施方案中,首先递送修饰的gRNA,随后递送Cas mRNA,导致与递送未修饰的gRNA之后递送Cas mRNA相比更高的编辑频率。

[0235] 在某些实施方案中,将gRNA与编码Cas蛋白的DNA质粒一起引入细胞中。在一些这样的实施方案中,通过核转染将gRNA和编码Cas蛋白的DNA质粒引入细胞中。在一些具体实施方案中,相比于基于DNA质粒的递送系统,基于RNP的递送平台或全RNA递送平台在原代细胞中提供更低的细胞毒性。

[0236] 在某些实施方案中,所述方法在人类细胞(包括人类原代T细胞和CD34+HSPC)中提供显著增强的基因组编辑效率。

[0237] 在某些实施方案中,相对于未修饰的gRNA,修饰的gRNA可增加插入或缺失(插入缺失)的频率,这可能指示诱变的NHEJ和基因破坏。具体而言,与未修饰的gRNA相比,在5'和3'端两者的三个末端核苷酸处掺入有2'-O-甲基-3'-硫代磷酸("MS")、2'-O-甲基-3'-PACE("MP")或2'-O-甲基-3'-硫代PACE(MSP)的修饰的gRNA增加了插入缺失的频率。

[0238] 在某些实施方案中,与未修饰的gRNA和Cas mRNA的共递送相比,将修饰的gRNA和Cas mRNA共递送至人类原代T细胞可增加插入缺失的频率。具体而言,与未修饰的gRNA相比,在5'和3'端两者的三个末端核苷酸处掺入有2'-O-甲基-3'-硫代磷酸("MS")、2'-O-甲基-3'-PACE("MP")或2'-O-甲基-3'-硫代PACE(MSP)的修饰的gRNA增加了人类原代T细胞中插入缺失的频率。

[0239] 在某些实施方案中,相对于未修饰的gRNA,修饰的gRNA可改善gRNA稳定性。作为一

个实例,相对于未修饰的gRNA,在5'和3'端两者的三个末端核苷酸处掺入有2'-O-甲基("M")的gRNA可适度地改善对核酸酶的稳定性,并且还可改善碱基配对热稳定性。作为另一个实例,与未修饰的gRNA相比,在5'和3'端两者的三个末端核苷酸处掺入有2'-O-甲基-3'-硫代磷酸("MS")、2'-O-甲基-3'-PACE("MP")或2'-O-甲基-3'-硫代PACE(MSP)的gRNA大幅度地改善了针对核酸酶的稳定性。预期gRNA末端修饰可增强针对核酸外切酶的细胞内稳定性,这使得将Cas mRNA和gRNA共同递送或依次递送到细胞或细胞裂解物中时基因组编辑的效率能够增加。在某些实施方案中,末端的增强稳定性的修饰也可以充当增强特异性的修饰,如果引导序列包含相同的末端的话。在某些实施方案中,包含2'-O-甲基-3'-磷酸基乙酸(MP)、2'-O-甲基-3'-硫代磷酸基乙酸(MSP)、2'-O-甲基-3'-硫代磷酸(MS)、2'-脱氧-3'-磷酸基乙酸(DP)、2'-脱氧-3'-硫代磷酸基乙酸(DSP)、2'-氟-3'-磷酸基乙酸(FP)、2'-氟-3'-硫代磷酸基乙酸(FSP)、2'-氟-3'-硫代磷酸(FS)、2'-O-(2-甲氧基乙基)-3'-磷酸基乙酸、2'-O-(2-甲氧基乙基)-硫代磷酸基乙酸、2'-O-(2-甲氧基乙基)-3'-硫代磷酸或其组合的末端修饰可增加本发明方法的稳定性和特异性。在某些实施方案中,修饰的gRNA刺激基因打靶,其进而允许通过例如同源重组或NHEJ进行基因编辑。具体而言,与未修饰的gRNA相比,在5'和3'端两者的三个末端核苷酸处掺入有2'-O-甲基-3'-硫代磷酸("MS")、2'-O-甲基-3'-PACE("MP")或2'-O-甲基-3'-硫代PACE(MSP)的gRNA刺激了更高水平的同源重组。

[0240] 在某些实施方案中,修饰的gRNA保留了高特异性。在某些实施方案中,与未修饰的gRNA相比,用修饰的gRNA提高了中靶对脱靶的插入缺失频率的比值。在某些实施方案中,与基于DNA质粒的递送系统相比,在具有Cas蛋白的RNP复合物中递送的修饰的gRNA提供了显著更好的中靶对脱靶比。

#### 基因表达调控

[0241] 在某些实施方案中,本文所述的引导RNA用于调节感兴趣的基因的转录或表达。例如,在某些实施方案中,使用包含Cas蛋白(例如,核酸酶缺陷型Cas9)和转录激活因子多肽的融合蛋白来增加基因的转录。类似地,在某些实施方案中,使用包含Cas蛋白(例如,核酸酶缺陷型Cas9)和阻抑物多肽的融合蛋白通过干扰基因的转录来敲低基因表达。

[0242] 在至少一个方面,本发明提供了一种体内或体外调控感兴趣基因的表达的方法。所述方法包括向细胞或另一个系统中引入(i)本文所述的合成引导RNA、和(ii)融合蛋白。在某些实施方案中,融合蛋白包含Cas蛋白和效应子结构域,如转录激活结构域、转录阻抑物结构域或表观遗传修饰结构域。在某些实施方案中,融合蛋白包含突变的Cas蛋白,例如作为无效核酸酶的Cas9蛋白。在某些实施方案中,Cas蛋白包含一个或多个突变,如D10A、H840A和/或N863A。

[0243] 在某些实施方案中,将融合蛋白作为纯化或非纯化蛋白引入细胞或系统中。在某些实施方案中,通过编码融合蛋白的mRNA将融合蛋白引入细胞或系统中。在某些实施方案中,通过编码融合蛋白的线性或环状DNA将融合蛋白引入细胞或系统中。

[0244] 在某些实施方案中,引导RNA有助于将融合蛋白导向至包含染色体序列、游离型序列、质粒、线粒体DNA序列或功能性基因间序列(如增强子或非编码RNA的DNA序列)的特异性靶多核苷酸。在某些实施方案中,效应子结构域调控靶多核苷酸中的序列的表达。可以设计用于调控基因表达的引导RNA,以靶向任何期望的内源基因或编码功能性RNA的序列。可以在内源基因的转录起始位点附近、或者在内源基因的翻译起始位点附近选择基因组靶序



列。在某些实施方案中,靶序列位于传统上称为基因的“启动子近端”区的DNA区域中。在某些实施方案中,靶序列位于转录起始位点上游约1,000个碱基对至转录起始位点下游约1,000个碱基对的区域中。在某些实施方案中,靶序列远离基因的转录起始位点(例如,在另一个染色体上)。

[0245] 在某些实施方案中,所述方法用于多重应用中。在某些实施方案中,所述方法包括将引导RNA文库引入细胞或系统中。在某些实施方案中,文库包含至少10个、至少100个、至少1,000个、至少10,000个、至少100,000个、或至少1,000,000个独特的引导序列。在某些实施方案中,文库靶向至少10个不同的多核苷酸或者一个或多个多核苷酸内的至少10个不同的序列。在某些实施方案中,文库靶向至少100个不同的多核苷酸或者一个或多个多核苷酸内的至少100个不同的序列。在某些实施方案中,文库靶向至少1,000个不同的多核苷酸或者一个或多个多核苷酸内的至少1,000个不同的序列。在某些实施方案中,文库靶向至少10,000个不同的多核苷酸或者一个或多个多核苷酸内的至少10,000个不同的序列。在某些实施方案中,文库靶向至少100,000个不同的多核苷酸或者一个或多个多核苷酸内的至少100,000个不同的序列。在某些实施方案中,文库靶向至少1,000,000个不同的多核苷酸或者一个或多个多核苷酸内的至少1,000,000个不同的序列。

#### 试剂盒

[0246] 在一个方面,本发明提供了含有用于执行上述方法的试剂的试剂盒,所述方法包括产生gRNA:Cas蛋白复合物和/或支持其结合、切刻或切割靶多核苷酸的活性。在某些实施方案中,用于本文中公开的方法的一种或多种反应组分,例如一个或多个引导RNA和Cas蛋白,可以供应为试剂盒的形式以供使用。在某些实施方案中,试剂盒包含Cas蛋白或编码Cas蛋白的核酸、以及本文所述的一个或多个引导RNA、或引导RNA的组或文库。在某些实施方案中,试剂盒包含一种或多种其他反应组分。在某些实施方案中,将适量的一种或多种反应组分提供在一个或多个容器中或保持在基底上。

[0247] 在某些实施方案中,本发明提供了用于选择合成引导RNA的试剂盒,所述合成引导RNA包括至少两个合成引导RNA,除了在引导序列中的不同位置处的不同修饰或修改外,所述合成引导RNA是相同的。每个引导RNA包含(a) crRNA区段,其包含(i)能够与靶多核苷酸杂交的引导序列、(ii)茎序列;和(b) tracrRNA区段,其包含与所述茎序列部分或完全互补的核苷酸序列,其中所述引导序列包含从所述引导序列的5'端计数的核苷酸1至20-N(N是-10至10之间的整数,任选地在-10至6之间),以及在引导序列的核苷酸处的至少一个增强特异性的修饰;其中所述至少两个合成引导RNA由于在引导序列中具有至少一个不同的增强特异性的修饰或由于具有在至少一个不同位置处的增强特异性的修饰而彼此不同。试剂盒还包含Cas蛋白或编码Cas蛋白的多核苷酸。在某些实施方案中,试剂盒中的每个合成引导RNA包含在不同核苷酸处的增强特异性的修饰。在某些实施方案中,试剂盒包含一系列合成引导RNA,每个引导RNA具有在引导序列中的不同核苷酸位置的修饰。在某些实施方案中,试剂盒具有与引导序列中的核苷酸数相同数量的不同引导RNA。在某些实施方案中,Cas蛋白是Cas9。在某些实施方案中,增强特异性的修饰包括2'-O-甲基-3'-磷酸基乙酸(MP)、2'-O-甲基-3'-硫代磷酸基乙酸(MSP)、2'-脱氧-3'-磷酸基乙酸(DP)、或2'-脱氧-3'-硫代磷酸基乙酸(DSP),或其组合。在某些实施方案中,增强特异性的修饰包括赋予C3'-内糖皱褶的2'修饰和磷酸基乙酸或硫代磷酸基乙酸键修饰。在某些实施方案中,2'修饰选自2'-F和2'-O-

(2-甲氧基乙基)。在某些实施方案中,引导RNA是合成的单引导RNA。

[0248] 试剂盒的附加组分的实例包括但不限于:一种或多种不同的聚合酶、一种或多种宿主细胞、用于将外来核酸引入宿主细胞的一种或多种试剂、用于检测引导RNA和/或Cas mRNA或蛋白质的表达或用于验证靶核酸的状态的一种或多种试剂(例如探针或PCR引物)、以及用于反应的缓冲液、转染试剂或培养基(处于1倍或更多倍的浓缩形式)。在某些实施方案中,试剂盒包括一种或多种以下组分:生物化学和物理支持物;终止、修饰和/或消化试剂;渗透剂(osmolyte);和用于反应、转染和/或检测的装置。

[0249] 所使用的反应组分可以以多种形式提供。例如,可以将组分(例如酶、RNA、探针和/或引物)悬浮在水溶液中或结合到珠上,或为冷冻干燥粉或冻干粉或沉淀物的形式。在后一种情况下,这些组分在复原时形成完全的组分混合物,以供在测定中使用。本发明试剂盒可以提供在适合的温度下。例如,为了将包含蛋白质组分或其复合物的试剂盒储存在液体中,优选将它们提供并保持在0℃以下,优选为约-20℃,可能在含有甘油或其他适合的防冻剂的抗冻溶液中。

[0250] 试剂盒或系统可含有足以用于至少一种测定法的量的本文所述组分的任何组合。在一些应用中,可以将一种或多种反应组分以预先测量的单次使用量提供在单独的一般为一次性的管或等同的容器中。通过这样的布置,可以通过将靶核酸或含有靶核酸的样品或细胞直接添加到单独的管中进行RNA引导的核酸酶反应。提供在试剂盒中的组分的量可以是任何适当的量,且可能取决于产品所针对的市场。其中提供组分的容器(一个或多个)可以是能够容纳所提供形式的任何常规容器,例如微量离心管、微量滴定板、安瓿、瓶子、整体试验装置,如流体装置、药筒、侧流装置、或其他类似装置。

[0251] 试剂盒还可以包括用于容纳容器或容器组合的包装材料。用于这样的试剂盒和系统的典型包装材料包括处于任何构造的容纳反应组分或检测探针的固体基质(例如,玻璃、塑料、纸、箔材、微粒,等),这些构造有多种(例如,管形瓶、微孔滴定板、微阵列,等)。试剂盒可进一步包括以有形形式记录的用于使用组分的说明。

## 实施例

### 实施例1

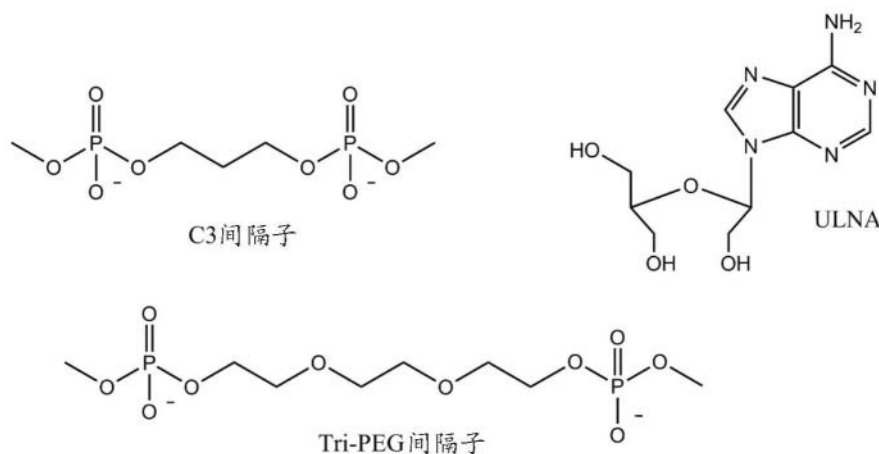
[0252] 为了评估化学修饰对生理盐条件下结合能的影响,制备了43个核苷酸的crRNA,其有20个核苷酸的引导序列。将crRNA与互补的40个核苷酸的DNA寡核苷酸混合形成双链体,所述DNA寡核苷酸的每个末端上有突出的10个核苷酸。测量双链体的解链温度(“T<sub>m</sub>”)。制备了另外七个43-nt crRNA,其中在取样和锁定区中具有若干连续的修饰,并且在20-nt引导序列中的位置10处具有一个间插或“间隔子”核苷酸。图6A显示了修饰的类型和布局。核酸外切酶抗性修饰是2'-O-甲基-3'-PACE(“MP”),间隔子是一个未修饰的RNA核苷酸。将每个修饰的crRNA 601与互补的40-nt DNA寡核苷酸603混合来一一形成双链体,并对于在RNA链中含有修饰的每个RNA/DNA双链体输入序列605,测量其解链温度,以定量各种修饰在生理盐条件下对结合能的影响。

[0253] 通过测量解链曲线的斜率,可以定性地评估修饰对双链体解链的协作性的影响:斜率越大指示在去结合中协作性越大(根据已确立的平衡可逆性原理,逆向过程中的结合也是如此)。图6B显示了未修饰的gRNA的解链曲线,并且图6C至6F显示了在20-nt引导序列

部分中核苷酸6至9的各种修饰类型的解链曲线。2'-O-甲基(“M”)和2'-O-甲基硫代磷酸(“MS”)的添加使结合能增加约0.2°C/修饰,而2'-O-甲基-3'-PACE(“MP”)的添加使 $T_m$ 或结合能降低1°C/修饰,并且2'-O-甲基-3'-硫代PACE(“MSP”)的添加使 $T_m$ 降低1.4°C/修饰。图7是显示20个碱基对的gRNA/DNA双链体随着gRNA被截短至17、18或19个核苷酸或延长至21或22个核苷酸长度的解链温度变化的图。在生理盐浓度下进行测量。测量结果表明,针对20个核苷酸的引导序列的延长或截短,解链温度的变化约为2°C/碱基对。

## 实施例2

[0254] 使用实验crRNA进行进一步的解链温度测量,实验crRNA在crRNA的引导序列中的核苷酸9处含有接头或接头样修饰。所述接头或接头样修饰包括ULNA(解锁核酸)、脱碱基间隔子、包含 $-PO_4Y-(CR^3_2)_m-PO_4Y-$ 的亚烷基间隔子、或包含 $(-PO_4Y-(CR^3_2CR^3_2O)_n-PO_3Y-)$ 的乙二醇间隔子,其中m是2、3或4,n是1、2或3,每个 $R^3$ 独立地选自下组:H、烷基和取代的烷基,并且每个Y为H或负电荷。例如,在某些实施方案中,亚烷基间隔物是“C3间隔子”,其中m是3。在其他实施方案中,乙二醇间隔物是“Tri-PEG间隔子”,其中n是3。这些修饰还通过赋予链更大的柔性降低了杂交的协作性,从而降低了结合能。



[0255] 这些解链温度测量的结果显示在表1中:

表1

条目	crRNA名称	$T_m$
1	对照_1XMP	50.2
2	rA-ULNA_9_1XMP	44.8
3	C3-间隔子_9_1XMP	40.9
4	三-PEG-间隔子_9_1XMP	38.3

## 实施例3-7

[0256] 为了评估化学合成的引导RNA结合和切割DNA靶序列的能力,并且为了评估各种修饰对Cas9介导的切割的影响,使用了体外切割测定法。简言之,通过对质粒上携带的各种靶基因的人源序列进行制备型PCR扩增,制备包含表2和表7中列出的靶多核苷酸序列(中靶)或脱靶多核苷酸序列(脱靶)的PAM可寻址的DNA构建体。作为用于演示本组合物和方法选择性地编辑基因的能力的示例性基因,使用了人网格蛋白轻链CLTA基因、人血红蛋白 $\beta$  (HBB)基因、人白细胞介素2受体亚基 $\gamma$  (IL2RG)基因、及人细胞毒性T淋巴细胞相关蛋白4 (CTLA4)基因作为靶基因。这些可以代表本文中公开的用于评估和编辑靶基因的一般方法。

[0257] 表3至表6列出了合成的引导RNA。在这些引导RNA的大部分中,5'端的前20个核苷酸与靶DNA中的靶序列互补——这些互补核苷酸构成引导序列。在一些引导RNA中,在引导序列的5'端有不与靶序列互补的突出端或延伸部分。在一些引导RNA中,引导序列的5'端被截短,因而核苷酸1、或1和2、或1和2和3不存在。中靶构建体(“中靶”)包含20-nt的靶序列。脱靶构建体(“脱靶”)包含与靶DNA相同的20个核苷酸中的大多数,同时有1、2或3个核苷酸差异。因此,gRNA大体上、但不完全与脱靶构建体的序列互补。脱靶构建体基于已知在人类基因组中存在的基因序列。

[0258] 根据Dellinger等人(2011)J.Am.Chem.Soc.,133,11540-56描述的程序,使用2'-O-硫代氨基甲酸酯保护的核苷亚磷酰胺,在ABI 394合成仪(Life Technologies, Carlsbad,CA,USA)上合成gRNA。在与2'-O-硫代氨基甲酸酯保护的亚磷酰胺相同的条件下,将2'-O-甲基亚磷酰胺掺入RNA寡聚体中。大体上根据公开的方法,合成了2'-O-甲基-3'-O-(二-异-丙基氨基)膦基乙酸-1,1-二甲基氨基乙酯-5'-O-三甲氧基三苯甲基核苷,其用于合成硫代膦酰基乙酸(thioPACE)修饰的RNA。参见Dellinger等人(2003)J.Am.Chem.Soc.125,940-50;和Threlfall等人(2012)Org.Biomol.Chem.10,746-54。

[0259] 使用反相高效液相色谱(HPLC)纯化所有的寡核苷酸,并通过使用Agilent 1290 Infinity系列LC系统的液相色谱-质谱(LC-MS)进行分析,所述LC系统与Agilent 6520 Q-TOF(时间飞行)质谱仪(Agilent Technologies,Santa Clara,CA,USA)偶联。使用从LC-MS衍生的总离子色谱图获得的质谱的反卷积来估计sgRNA的合成和纯化的产率。100聚体sgRNA的化学合成通常产生25%-35%的全长产物,其来自标称的1微摩尔规模合成。使用离子配对缓冲液条件,反相HPLC纯化通常从粗产物中给出20%的产率,估计的最终sgRNA纯度在90%至95%的范围内。

[0260] DNA靶构建体包含表2中列出的靶序列(也称为中靶序列或标识为“中靶”)和脱靶序列(“脱靶”),其中脱靶序列与靶序列的不同之处以粗斜体表示,并且PAM序列(如图所示)带下划线:

表2

中靶或脱靶位 点	DNA序列	SEQ ID NO.
CLTA1中靶	AGTCCTCATCTCCCTCAA GCAGG	1
CLTA1脱靶1	AGTCCTCAACTCCCTCAA GCAGG	2
CLTA1脱靶3	ACTCCTCATCCCCCTCAA GCCGG	3
CLTA4中靶	GCAGATGTAGTGTTTCCA CAGGG	4
CLTA4脱靶1	GCAGATGTAGTATTTCCA CAGGG	5
CLTA4脱靶2	CCAGATGTAGCGTTTCCA CAGGG	6
CLTA4脱靶3	GCAGATGTGTGTGTTTCCA CAGGG	7
HBB中靶	CTTGCCCCACAGGGCAG TAACGG	8
HBB脱靶1	TCAGCCCCACAGGGCAG TAAGGG	9
IL2RG中靶	TGGTAATGATGGCTTCAA CATGG	10

IL2RG脱靶2	TGGT <b>GAG</b> GATGGCTTCA ACACGG	191
IL2RG脱靶3	TGGTAATGATG <b>ACT</b> TCAA CATAG	11
VEGFA中靶	GGTGAGTGAGTGTGTGC GTGTGG	192
VEGFA脱靶2	TGTGGGTGAGTGTGTGC GTG <b>AGG</b>	193

[0261] 注意,“HBB脱靶1与HBB中靶在潜在靶序列的前3个核苷酸中不同。因此,在gRNA的20-nt引导序列的5'端将gRNA截短3个核苷酸以提供17-nt引导序列,这不能区分HBB中靶和HBB脱靶1靶序列。实施例3-7中使用的DNA靶构建体的完整序列在下表8中列出。

[0262] 为了比较目的,纳入将gRNA的引导序列从20个核苷酸截短至18或17个核苷酸的引导RNA,评估了靶基因位点的切割与已知的脱靶位点的切割的比值。可以看出,与Yanfang等人(2014)的教导和结论相反,截短对特定的脱靶位点的切割有影响,但对其他脱靶位点处的切割没有影响。尽管有Yanfang等人(2014)的教导,本发明人还是寻找并鉴定了用于CRISPR-Cas切割、切割或结合靶多核苷酸的新化合物和方法,其具有增强的特异性且没有截短引导序列。

[0263] 在20μL的反应体积中,将2.5nM线性化的DNA靶标在50nM sgRNA、40nM重组纯化的Cas9蛋白(酿脓链球菌;Agilent)、pH 7.6的0.8mM MgCl<sub>2</sub>存在下37℃温育1小时。完成后,添加0.5μL RNase It(Agilent),并在37℃下继续温育5分钟,然后在70℃下温育15分钟。将粗产物上样到DNA 1000或DNA 7500LabChip中,在Agilent Bioanalyzer 2200上进行分析,或者将其上样到Genomic DNA ScreenTape或D5000 ScreenTape上,在Agilent TapeStation 2200或4200上进行分析。处理步骤用于将Cas9从与靶DNA的结合中释放出来,测定其切割情况。

[0264] 通过下式计算切割产率: $a/(a+b) \times 100$ ,其中a是两个切割产物的条带强度的和,b是剩余的未切割的DNA(如果存在的话)。切割百分比100%表示在检测限内所有的靶DNA构建体被切割。

### 实施例3

[0265] 制备一系列32个sgRNA用于靶向人CLTA基因中的“CLTA1”基因座。简言之,合成单独的RNA链并将其通过HPLC纯化。基于通过HPLC分析的全长链纯度和通过质谱分析的化学组成,所有寡核苷酸均获质量控制批准。表3列出了各个CLTA1 sgRNA的序列。表3显示了作为条目1至31的sgRNA的序列,而且该表格公开了本发明gRNA的含有一个或多个增强特异性的修饰的某些实施方案。条目1是未修饰的,用作比较例。条目2在引导序列的核苷酸1、2和3处含有MS修饰。条目3在引导序列的核苷酸1、2和3处含有MSP修饰。条目4在引导序列的核苷酸1处含有MSP修饰。条目5和6是可比较的实例,其中gRNA的20-nt引导序列分别在5'端被截短至18-核苷酸引导序列或17-核苷酸引导序列。条目7和8是可比较的实例,其中gRNA未修

饰,分别在20-nt引导序列的5'端具有一个或两个核苷酸的突出端。条目9、10、11、12和13分别在引导序列中的核苷酸1、核苷酸1-2、核苷酸1-3、核苷酸1-4和核苷酸1-5处含有MP修饰。条目14和15分别在gRNA的tracrRNA区中从sgRNA的3'端开始计数的核苷酸2-5处或从sgRNA的3'端开始计数的核苷酸2-6处含有MP修饰,注意核苷酸通常是从多核苷酸的5'端开始计数的。因此,对条目14和15描述的计数是对一般规则的例外。条目16和17在20-nt引导序列中核苷酸1-2处含有MP修饰,且分别具有MP修饰的C或G核苷酸突出端。条目18和19在20-nt引导序列中的核苷酸1-3处含有MP修饰,且分别具有MP修饰的UC或AG二核苷酸突出端。条目20和21在20-nt引导序列中的核苷酸1-4处包含MP修饰,且分别具有MP修饰的CUC或GAG三核苷酸突出端,并且在gRNA的tracrRNA区中在sgRNA的3'端具有MP修饰。条目22-25分别在引导序列中的核苷酸20、19、18或17处含有MP修饰。条目26在引导序列中核苷酸18和17处含有MP修饰。条目27-29分别在引导序列中的核苷酸19、18或17处含有M修饰。条目30在引导序列中的核苷酸18和17处含有M修饰。条目31在引导序列中的核苷酸1-20处含有M修饰。条目32在20-nt引导序列中的核苷酸1-7、9-11、13-14和20处含有M修饰,并且在sgRNA序列的其余部分的若干选择位置处具有M修饰,特别是在核苷酸30-31、33、35-36、39、42、45、47、50、60、65-66、70、71、76-77、80-82、90、93、95-96、100-101、104和106-112处。





计算切割的靶多核苷酸与切割的脱靶多核苷酸的比值。还计算了通过将每个比值乘以每个测定的sgRNA的对应中靶切割百分比而获得的特异性评分。图8B的条目6、12、13、20、21和31中的阴影值是由于靶多核苷酸序列(中靶1)的切割产率略微降低至显著降低所致。重要的是,所述降低对于脱靶多核苷酸序列之一或两者更大,产生了大于2.0的中靶:脱靶切割比。阴影值表明可以获得至少两倍的特异性改善。在本实验中测试的各种化学修饰和组合中,2'-O-甲基-3'-PACE("MP")修饰产生了最大的减少脱靶切割的期望效果,同时保留高水平的中靶切割,特别是当它在引导序列中最佳数目的位置处掺入时。这些例子用作本教导的实施方案。

#### 实施例4

[0267] 对于图9A中表示的例子,将单个或三个MP修饰在整个20-nt引导序列上“步移”,以观察哪些修饰位置能够产生特异性的改善,如通过中靶与脱靶切割活性判断的。如表4中列出的,制备了一系列28个sgRNA(SEQ ID NO:43-70),用于靶向人CLTA基因中的“CLTA4”基因座,其中实验sgRNA在引导序列中一个或多个位置处含有2'-O-甲基-3'-PACE("MP")修饰,此外在核苷酸1处以及在sgRNA的3'端的tracrRNA区(包括最后一个(即最3'端)核苷酸间键)中的倒数第二个核苷酸处具有MP修饰。因此,将末端核苷酸间键中的修饰设计为保护这样的修饰的sgRNA免于被核酸外切酶降解。合成了单独的RNA链并将其通过HPLC纯化。基于通过HPLC分析的全长链纯度和通过质谱分析的化学组成,所有寡核苷酸均获质量控制批准。将CLTA4靶多核苷酸序列(中靶)和可比较的脱靶多核苷酸序列(脱靶1、脱靶2和脱靶3)的切割百分比以数字和条形图显示在图9A中。图9B是从图9A中的结果导出的,其中对测定的每个合成sgRNA(SEQ ID NO:43-70)计算了切割的靶多核苷酸与切割的脱靶多核苷酸的比值。记录为“大”的比值表明在那些特定测定中未检测到脱靶DNA多核苷酸的切割。如上面在图8B的描述中所述,通过将比值乘以其对应的中靶切割百分比来计算特异性评分。在由图9B表示的例子中, $\geq 2.0$ 的特异性评分用阴影表示,其指示相对于无阴影的评分的特异性改善。阴影指示在引导序列中的哪些MP位置提供了至少两倍的特异性改善。条目1-18中MP步移的结果表明,步移MP修饰的布置对特异性有影响,并且这个趋势对于每个测定的脱靶位点的各组特异性评分是明显的,表明在引导序列5'端附近的MP修饰比引导序列中其他位置的MP更能增强特异性,如在条目1和2相对于条目3-18可见。这个趋势与针对靶向至CLTA1靶序列的gRNA中的MP修饰观察到的特异性增强趋势一致,其中添加到CLTA1 sgRNA的5'端的MP修饰增强了特异性,如图8B中的条目12-13和20-21中的阴影评分指示的。虽然也许并非在每种情况下都有效,但是改善特异性的一般策略是在gRNA中的引导序列的5'端的连续磷酸二酯核苷酸间键处掺入1、2、3、4或5个MP修饰。



表 4

条目	sgRNA 名	5' → 3'	SEQ. ID NO.
1	CLTA4_5-7MP_1xMP	G* <u>CAGA</u> <u>U</u> * <u>G</u> *UAGUGUUUCCACAGUUUAGAGCUAUGCGUUAACAGCAUAGCAAGUUUAAAUAAAGGCUAGUCGCUUUAACAACUUGAAAGUUGGCACCCGAGUCGGUGCUUUUUU*U	43
2	CLTA4_6-8MP_1xMP	G* <u>CAGAU</u> <u>G</u> * <u>G</u> *U*AGUGUUUCCACAGUUUAGAGCUAUGCGUUAACAGCAUAGCAAGUUUAAAUAAAGGCUAGUCGCUUUAACAACUUGAAAGUUGGCACCCGAGUCGGUGCUUUUUU*U	44
3	CLTA4_7-9MP_1xMP	G* <u>CAGAU</u> <u>G</u> * <u>U</u> *A*GUGUUUCCACAGUUUAGAGCUAUGCGUUAACAGCAUAGCAAGUUUAAAUAAAGGCUAGUCGCUUUAACAACUUGAAAGUUGGCACCCGAGUCGGUGCUUUUUU*U	45
4	CLTA4_8-10MP_1xMP	G* <u>CAGAU</u> <u>G</u> * <u>A</u> *G*UGUUUCCACAGUUUAGAGCUAUGCGUUAACAGCAUAGCAAGUUUAAAUAAAGGCUAGUCGCUUUAACAACUUGAAAGUUGGCACCCGAGUCGGUGCUUUUUU*U	46
5	CLTA4_9-11MP_1xMP	G* <u>CAGAU</u> <u>G</u> * <u>A</u> *G*U*GUUUCCACAGUUUAGAGCUAUGCGUUAACAGCAUAGCAAGUUUAAAUAAAGGCUAGUCGCUUUAACAACUUGAAAGUUGGCACCCGAGUCGGUGCUUUUUU*U	47
6	CLTA4_10-12MP_1xMP	G* <u>CAGAU</u> <u>G</u> * <u>A</u> *G*U*U*UCCACAGUUUAGAGCUAUGCGUUAACAGCAUAGCAAGUUUAAAUAAAGGCUAGUCGCUUUAACAACUUGAAAGUUGGCACCCGAGUCGGUGCUUUUUU*U	48
7	CLTA4_11-13MP_1xMP	G* <u>CAGAU</u> <u>G</u> * <u>A</u> *G*U*U*UCCACAGUUUAGAGCUAUGCGUUAACAGCAUAGCAAGUUUAAAUAAAGGCUAGUCGCUUUAACAACUUGAAAGUUGGCACCCGAGUCGGUGCUUUUUU*U	49
8	CLTA4_12-14MP_1xMP	G* <u>CAGAU</u> <u>G</u> * <u>A</u> *G*U*U*U*UCCACAGUUUAGAGCUAUGCGUUAACAGCAUAGCAAGUUUAAAUAAAGGCUAGUCGCUUUAACAACUUGAAAGUUGGCACCCGAGUCGGUGCUUUUUU*U	50
9	CLTA4_13-15MP_1xMP	G* <u>CAGAU</u> <u>G</u> * <u>A</u> *G*U*U*U*U*UCCACAGUUUAGAGCUAUGCGUUAACAGCAUAGCAAGUUUAAAUAAAGGCUAGUCGCUUUAACAACUUGAAAGUUGGCACCCGAGUCGGUGCUUUUUU*U	51
10	CLTA4_14-16MP_1xMP	G* <u>CAGAU</u> <u>G</u> * <u>A</u> *G*U*U*U*U*U*UCCACAGUUUAGAGCUAUGCGUUAACAGCAUAGCAAGUUUAAAUAAAGGCUAGUCGCUUUAACAACUUGAAAGUUGGCACCCGAGUCGGUGCUUUUUU*U	52
11	CLTA4_4MP_1xMP	G* <u>CAG</u> *AUGUAGUGUUUCCACAGUUUAGAGCUAUGCGUUAACAGCAUAGCAAGUUUAAAUAAAGGCUAGUCGCUUUAACAACUUGAAAGUUGGCACCCGAGUCGGUGCUUUUUU*U	53
12	CLTA4_5MP_1xMP	G* <u>CAGA</u> <u>A</u> *UGUAGUGUUUCCACAGUUUAGAGCUAUGCGUUAACAGCAUAGCAAGUUUAAAUAAAGGCUAGUCGCUUUAACAACUUGAAAGUUGGCACCCGAGUCGGUGCUUUUUU*U	54
13	CLTA4_6MP_1xMP	G* <u>CAGAU</u> <u>A</u> *GUGUGUUUCCACAGUUUAGAGCUAUGCGUUAACAGCAUAGCAAGUUUAAAUAAAGGCUAGUCGCUUUAACAACUUGAAAGUUGGCACCCGAGUCGGUGCUUUUUU*U	55
14	CLTA4_7MP_1xMP	G* <u>CAGAU</u> <u>G</u> *UAGUGUUUCCACAGUUUAGAGCUAUGCGUUAACAGCAUAGCAAGUUUAAAUAAAGGCUAGUCGCUUUAACAACUUGAAAGUUGGCACCCGAGUCGGUGCUUUUUU*U	56
15	CLTA4_8MP_1xMP	G* <u>CAGAU</u> <u>G</u> *U*AGUGUUUCCACAGUUUAGAGCUAUGCGUUAACAGCAUAGCAAGUUUAAAUAAAGGCUAGUCGCUUUAACAACUUGAAAGUUGGCACCCGAGUCGGUGCUUUUUU*U	57
16	CLTA4_9MP_1xMP	G* <u>CAGAU</u> <u>G</u> *A*GUGUUUCCACAGUUUAGAGCUAUGCGUUAACAGCAUAGCAAGUUUAAAUAAAGGCUAGUCGCUUUAACAACUUGAAAGUUGGCACCCGAGUCGGUGCUUUUUU*U	58
17	CLTA4_10MP_1xMP	G* <u>CAGAU</u> <u>G</u> *U*GUGUUUCCACAGUUUAGAGCUAUGCGUUAACAGCAUAGCAAGUUUAAAUAAAGGCUAGUCGCUUUAACAACUUGAAAGUUGGCACCCGAGUCGGUGCUUUUUU*U	59
18	CLTA4_11MP_1xMP	G* <u>CAGAU</u> <u>G</u> *U*G*U*UCCACAGUUUAGAGCUAUGCGUUAACAGCAUAGCAAGUUUAAAUAAAGGCUAGUCGCUUUAACAACUUGAAAGUUGGCACCCGAGUCGGUGCUUUUUU*U	60
19	CLTA4_12MP_1xMP	G* <u>CAGAU</u> <u>G</u> *U*G*U*U*UCCACAGUUUAGAGCUAUGCGUUAACAGCAUAGCAAGUUUAAAUAAAGGCUAGUCGCUUUAACAACUUGAAAGUUGGCACCCGAGUCGGUGCUUUUUU*U	61
20	CLTA4_13MP_1xMP	G* <u>CAGAU</u> <u>G</u> *U*G*U*U*U*UCCACAGUUUAGAGCUAUGCGUUAACAGCAUAGCAAGUUUAAAUAAAGGCUAGUCGCUUUAACAACUUGAAAGUUGGCACCCGAGUCGGUGCUUUUUU*U	62
21	CLTA4_14MP_1xMP	G* <u>CAGAU</u> <u>G</u> *U*G*U*U*U*U*UCCACAGUUUAGAGCUAUGCGUUAACAGCAUAGCAAGUUUAAAUAAAGGCUAGUCGCUUUAACAACUUGAAAGUUGGCACCCGAGUCGGUGCUUUUUU*U	63
22	CLTA4_15MP_1xMP	G* <u>CAGAU</u> <u>G</u> *U*G*U*U*U*U*U*UCCACAGUUUAGAGCUAUGCGUUAACAGCAUAGCAAGUUUAAAUAAAGGCUAGUCGCUUUAACAACUUGAAAGUUGGCACCCGAGUCGGUGCUUUUUU*U	64
23	CLTA4_16MP_1xMP	G* <u>CAGAU</u> <u>G</u> *U*G*U*U*U*U*U*U*UCCACAGUUUAGAGCUAUGCGUUAACAGCAUAGCAAGUUUAAAUAAAGGCUAGUCGCUUUAACAACUUGAAAGUUGGCACCCGAGUCGGUGCUUUUUU*U	65
24	CLTA4_17MP_1xMP	G* <u>CAGAU</u> <u>G</u> *U*G*U*U*U*U*U*U*U*UCCACAGUUUAGAGCUAUGCGUUAACAGCAUAGCAAGUUUAAAUAAAGGCUAGUCGCUUUAACAACUUGAAAGUUGGCACCCGAGUCGGUGCUUUUUU*U	66
25	CLTA4_18MP_1xMP	G* <u>CAGAU</u> <u>G</u> *U*G*U*U*U*U*U*U*U*U*UCCACAGUUUAGAGCUAUGCGUUAACAGCAUAGCAAGUUUAAAUAAAGGCUAGUCGCUUUAACAACUUGAAAGUUGGCACCCGAGUCGGUGCUUUUUU*U	67
26	CLTA4_19MP_1xMP	G* <u>CAGAU</u> <u>G</u> *U*G*U*U*U*U*U*U*U*U*U*UCCACAGUUUAGAGCUAUGCGUUAACAGCAUAGCAAGUUUAAAUAAAGGCUAGUCGCUUUAACAACUUGAAAGUUGGCACCCGAGUCGGUGCUUUUUU*U	68
27	CLTA4_20MP_1xMP	G* <u>CAGAU</u> <u>G</u> *U*G*U*U*U*U*U*U*U*U*U*U*UCCACAGUUUAGAGCUAUGCGUUAACAGCAUAGCAAGUUUAAAUAAAGGCUAGUCGCUUUAACAACUUGAAAGUUGGCACCCGAGUCGGUGCUUUUUU*U	69
28	CLTA4_unmodif	GCAGAUAGUGUUUCCACAGUUUAGAGCUAUGCGUUAACAGCAUAGCAAGUUUAAAUAAAGGCUAGUCGCUUUAACAACUUGAAAGUUGGCACCCGAGUCGGUGCUUUUUU	70

图注

Ns=核苷酸 N 的 2'-O-甲基-3'-硫代磷酸酯修饰  
N'=核苷酸 N 的 2'-O-甲基-3'-PACE 修饰  
Ns\*=核苷酸 N 的 2'-O-甲基-3'-硫代 PACE 修饰  
Ns\*=突出端核苷酸 N 的 2'-O-甲基-3'-PACE 修饰(其中 N 位于与 sgRNA 的引导序列共价连接的 5'突出端序列中)  
N=核苷酸 N 的 2'-O-甲基修饰

实施例5

[0268] 在由图10表示的例子中,将单个或三个MP修饰在整个20-nt引导序列上“步移”,以观察哪些修饰位置能够产生特异性的改善,如通过中靶与脱靶切割活性判断的。如表5中列

出,制备了一系列的三十个sgRNA(SEQ ID NO:71-86和173-186),用于靶向人IL2RG基因中的基因座。这些实验sgRNA在引导序列中一个或多个位置处含有2'-O-甲基-3'-PACE("MP")修饰,另外在核苷酸1处以及sgRNA的3'端的tracrRNA区中的倒数第二个核苷酸处(包括最后一个(即大多数的3')核苷酸间键)也具有MP修饰。因此,末端核苷酸间键中的修饰被设计为保护sgRNA免受核酸外切酶影响。合成了单独的RNA链并通过HPLC纯化。基于通过HPLC分析确定的全长链纯度和通过质谱分析确定的化学组成,所有寡核苷酸均获得了质量控制批准。图10中以数字形式显示了IL2RG靶多核苷酸序列(中靶)和可比较的脱靶多核苷酸序列(脱靶3)的切割百分比。这个图还显示了为每个受测的合成sgRNA(SEQ ID NO:71-86和173-186)计算的切割的靶多核苷酸与切割的脱靶多核苷酸的比值。通过将比值乘以其对应的中靶切割百分比来计算特异性评分。将 $\geq 2.0$ 的特异性评分用阴影表示,以指示相对于无阴影的评分的特异性改善,因此阴影指示了引导序列中的哪些MP位置提供了至少两倍的特异性改善。在图10中,将引导序列中产生特异性最大改善的位置用更暗的阴影指示。来自位置7、14或16的MP修饰的结果分别在条目6、13和15中示出。具有这样的组成的引导RNA相对于其他常用的组成(如缺乏MP修饰的gRNA)可以增强特异性性能,对于靶向临床上重要的IL2RG基因座的gRNA的各种CRISPR-Cas应用而言尤其如此。由图10表示的例子还例示了在gRNA的引导序列部分上“步移”MP修饰的本发明新方法,以鉴定哪个或哪些位置由于步移MP修饰的位置而产生了特异性增强。通过针对每个测试位置的中靶与脱靶切割比来评估特异性增强的幅度,而且,当从业者决定哪个或那些MP修饰的位置可能有益于gRNA的整体性能时,可以将这些值与针对每种设计所测得的中靶切割百分比一起考虑。单个MP在整个引导序列上的渐进式步移还可以鉴定该序列中有可能协同性地改善特异性的位置:这样的改善是步移所测试的位置处的一个或多个MP修饰的组合造成的。



表 5

条目	sgRNA 名	5' → 3'	SEQ ID NO
1	IL2RG_2MP_1xMP	U*G*GUAUAUGGCUUCAAAGUUUAGAGCUAGAAUAGCAAGUUAUUAAUAAAGGCUAGUCCGUUAUCAAUUUAGAAAGGCGACCGAGUCGGUGCUU*U	176
2	IL2RG_3MP_1xMP	U*GG*UAAUAGUAGGCUUCAAAGUUUAGAGCUAGAAUAGCAAGUUAUUAAUAAAGGCUAGUCCGUUAUCAAUUUAGAAAGGCGACCGAGUCGGUGCUU*U	177
3	IL2RG_4MP_1xMP	U*GGU*AAUAGUAGGCUUCAAAGUUUAGAGCUAGAAUAGCAAGUUAUUAAUAAAGGCUAGUCCGUUAUCAAUUUAGAAAGGCGACCGAGUCGGUGCUU*U	178
4	IL2RG_5MP_1xMP	U*GGUAA*UAGUAGGCUUCAAAGUUUAGAGCUAGAAUAGCAAGUUAUUAAUAAAGGCUAGUCCGUUAUCAAUUUAGAAAGGCGACCGAGUCGGUGCUU*U	179
5	IL2RG_6MP_1xMP	U*GGUAA*U*UAGUAGGCUUCAAAGUUUAGAGCUAGAAUAGCAAGUUAUUAAUAAAGGCUAGUCCGUUAUCAAUUUAGAAAGGCGACCGAGUCGGUGCUU*U	180
6	IL2RG_7MP_1xMP	U*GGUAAU*U*UAGUAGGCUUCAAAGUUUAGAGCUAGAAUAGCAAGUUAUUAAUAAAGGCUAGUCCGUUAUCAAUUUAGAAAGGCGACCGAGUCGGUGCUU*U	181
7	IL2RG_8MP_1xMP	U*GGUAAU*U*U*UAGUAGGCUUCAAAGUUUAGAGCUAGAAUAGCAAGUUAUUAAUAAAGGCUAGUCCGUUAUCAAUUUAGAAAGGCGACCGAGUCGGUGCUU*U	182
8	IL2RG_9MP_1xMP	U*GGUAAU*U*U*U*UAGUAGGCUUCAAAGUUUAGAGCUAGAAUAGCAAGUUAUUAAUAAAGGCUAGUCCGUUAUCAAUUUAGAAAGGCGACCGAGUCGGUGCUU*U	183
9	IL2RG_10MP_1xMP	U*GGUAAU*U*U*U*U*UAGUAGGCUUCAAAGUUUAGAGCUAGAAUAGCAAGUUAUUAAUAAAGGCUAGUCCGUUAUCAAUUUAGAAAGGCGACCGAGUCGGUGCUU*U	184
10	IL2RG_11MP_1xMP	U*GGUAAU*U*U*U*U*U*UAGUAGGCUUCAAAGUUUAGAGCUAGAAUAGCAAGUUAUUAAUAAAGGCUAGUCCGUUAUCAAUUUAGAAAGGCGACCGAGUCGGUGCUU*U	185
11	IL2RG_12MP_1xMP	U*GGUAAU*U*U*U*U*U*U*UAGUAGGCUUCAAAGUUUAGAGCUAGAAUAGCAAGUUAUUAAUAAAGGCUAGUCCGUUAUCAAUUUAGAAAGGCGACCGAGUCGGUGCUU*U	186
12	IL2RG_13MP_1xMP	U*GGUAAU*U*U*U*U*U*U*U*UAGUAGGCUUCAAAGUUUAGAGCUAGAAUAGCAAGUUAUUAAUAAAGGCUAGUCCGUUAUCAAUUUAGAAAGGCGACCGAGUCGGUGCUU*U	78
13	IL2RG_14MP_1xMP	U*GGUAAU*U*U*U*U*U*U*U*U*UAGUAGGCUUCAAAGUUUAGAGCUAGAAUAGCAAGUUAUUAAUAAAGGCUAGUCCGUUAUCAAUUUAGAAAGGCGACCGAGUCGGUGCUU*U	79
14	IL2RG_15MP_1xMP	U*GGUAAU*U*U*U*U*U*U*U*U*U*UAGUAGGCUUCAAAGUUUAGAGCUAGAAUAGCAAGUUAUUAAUAAAGGCUAGUCCGUUAUCAAUUUAGAAAGGCGACCGAGUCGGUGCUU*U	80
15	IL2RG_16MP_1xMP	U*GGUAAU*U*U*U*U*U*U*U*U*U*U*UAGUAGGCUUCAAAGUUUAGAGCUAGAAUAGCAAGUUAUUAAUAAAGGCUAGUCCGUUAUCAAUUUAGAAAGGCGACCGAGUCGGUGCUU*U	81
16	IL2RG_17MP_1xMP	U*GGUAAU*U*U*U*U*U*U*U*U*U*U*U*UAGUAGGCUUCAAAGUUUAGAGCUAGAAUAGCAAGUUAUUAAUAAAGGCUAGUCCGUUAUCAAUUUAGAAAGGCGACCGAGUCGGUGCUU*U	82
17	IL2RG_18MP_1xMP	U*GGUAAU*U*U*U*U*U*U*U*U*U*U*U*U*UAGUAGGCUUCAAAGUUUAGAGCUAGAAUAGCAAGUUAUUAAUAAAGGCUAGUCCGUUAUCAAUUUAGAAAGGCGACCGAGUCGGUGCUU*U	83
18	IL2RG_19MP_1xMP	U*GGUAAU*U*U*U*U*U*U*U*U*U*U*U*U*U*UAGUAGGCUUCAAAGUUUAGAGCUAGAAUAGCAAGUUAUUAAUAAAGGCUAGUCCGUUAUCAAUUUAGAAAGGCGACCGAGUCGGUGCUU*U	84
19	IL2RG_20MP_1xMP	U*GGUAAU*U*U*U*U*U*U*U*U*U*U*U*U*U*U*UAGUAGGCUUCAAAGUUUAGAGCUAGAAUAGCAAGUUAUUAAUAAAGGCUAGUCCGUUAUCAAUUUAGAAAGGCGACCGAGUCGGUGCUU*U	85
20	IL2RG_1xMP (control)	U*GGUAAU*U*U*U*U*U*U*U*U*U*U*U*U*U*U*U*UAGUAGGCUUCAAAGUUUAGAGCUAGAAUAGCAAGUUAUUAAUAAAGGCUAGUCCGUUAUCAAUUUAGAAAGGCGACCGAGUCGGUGCUU*U	86
21	IL2RG_4-6MP_1xMP	U*GGU*U*U*U*U*U*U*U*U*U*U*U*U*U*U*U*U*UAGUAGGCUUCAAAGUUUAGAGCUAGAAUAGCAAGUUAUUAAUAAAGGCUAGUCCGUUAUCAAUUUAGAAAGGCGACCGAGUCGGUGCUU*U	173
22	IL2RG_5-7MP_1xMP	U*GGUAA*U*U*U*U*U*U*U*U*U*U*U*U*U*U*U*U*U*UAGUAGGCUUCAAAGUUUAGAGCUAGAAUAGCAAGUUAUUAAUAAAGGCUAGUCCGUUAUCAAUUUAGAAAGGCGACCGAGUCGGUGCUU*U	174
23	IL2RG_6-8MP_1xMP	U*GGUAA*U*U*U*U*U*U*U*U*U*U*U*U*U*U*U*U*U*U*UAGUAGGCUUCAAAGUUUAGAGCUAGAAUAGCAAGUUAUUAAUAAAGGCUAGUCCGUUAUCAAUUUAGAAAGGCGACCGAGUCGGUGCUU*U	175
24	IL2RG_7-9MP_1xMP	U*GGUAAU*UAGUAGGCUUCAAAGUUUAGAGCUAGAAUAGCAAGUUAUUAAUAAAGGCUAGUCCGUUAUCAAUUUAGAAAGGCGACCGAGUCGGUGCUU*U	71
25	IL2RG_8-10MP_1xMP	U*GGUAAU*UAGUAGGCUUCAAAGUUUAGAGCUAGAAUAGCAAGUUAUUAAUAAAGGCUAGUCCGUUAUCAAUUUAGAAAGGCGACCGAGUCGGUGCUU*U	72
26	IL2RG_9-11MP_1xMP	U*GGUAAU*UAGUAGGCUUCAAAGUUUAGAGCUAGAAUAGCAAGUUAUUAAUAAAGGCUAGUCCGUUAUCAAUUUAGAAAGGCGACCGAGUCGGUGCUU*U	73
27	IL2RG_10-12MP_1xMP	U*GGUAAU*UAGUAGGCUUCAAAGUUUAGAGCUAGAAUAGCAAGUUAUUAAUAAAGGCUAGUCCGUUAUCAAUUUAGAAAGGCGACCGAGUCGGUGCUU*U	74
28	IL2RG_11-13MP_1xMP	U*GGUAAU*UAGUAGGCUUCAAAGUUUAGAGCUAGAAUAGCAAGUUAUUAAUAAAGGCUAGUCCGUUAUCAAUUUAGAAAGGCGACCGAGUCGGUGCUU*U	75
29	IL2RG_12-14MP_1xMP	U*GGUAAU*UAGUAGGCUUCAAAGUUUAGAGCUAGAAUAGCAAGUUAUUAAUAAAGGCUAGUCCGUUAUCAAUUUAGAAAGGCGACCGAGUCGGUGCUU*U	76
30	IL2RG_13-15MP_1xMP	U*GGUAAU*UAGUAGGCUUCAAAGUUUAGAGCUAGAAUAGCAAGUUAUUAAUAAAGGCUAGUCCGUUAUCAAUUUAGAAAGGCGACCGAGUCGGUGCUU*U	77

图注

N\*=核苷酸 N 的 2'-O-甲基-3'-PACE 修饰

实施例6

[0269] 进行修饰“步移”以观察在靶向如图11A中所示的人HBB基因的引导序列中的各个位点是否可以显著切割中靶位点并减少脱靶位点的切割,修饰“步移”通过在这条20-nt的

引导序列各个位置上渐进地引入修饰来进行。制备了一系列65个sgRNA,用于靶向图6中列出的HBB基因,其中实验sgRNA在引导序列中一个或多个内部核苷酸位置处含有2'-O-甲基-3'-PACE("MP")修饰,另外在核苷酸1处以及在sgRNA的3'端的tracrRNA区中的倒数第二个核苷酸处(包括最后一个(即大多数的3')核苷酸间键)具有MP修饰。因此,将末端核苷酸间键中的修饰设计为保护sgRNA免于被核酸外切酶降解。合成了单独的RNA链并将其通过HPLC纯化。基于通过HPLC分析的全长链纯度和通过质谱分析的化学组成,所有寡核苷酸均获质量控制批准。将HBB靶多核苷酸序列(中靶)和可比较的脱靶多核苷酸序列(脱靶1)的切割百分比以数字显示在图11A中。该图还显示了为每个受测的合成sgRNA(SEQ ID NO:87-103)计算的切割的靶多核苷酸与切割的脱靶多核苷酸的比值。通过将比值乘以其对应的中靶切割百分比来计算特异性评分。将 $\geq 2.0$ 的特异性评分用阴影表示,以指示相对于无阴影的评分的特异性改善,因此阴影指示在引导序列中的哪些MP位置提供了至少两倍的特异性改善。在图11A中,引导序列中产生特异性最大改善的位置由较暗的阴影指示,该改善由于分别如条目2、6和8中所示的位置5、9或11的MP修饰所致。具有这样的组成的引导RNA相对于其他常用的组成(如缺乏MP修饰的gRNA)可以增强特异性,对于靶向临床上重要的HBB基因座的gRNA的各种CRISPR-Cas应用而言尤其如此。图11A所示的实例还例示了我们在gRNA的整个引导序列部分上“步移”MP修饰,以鉴定哪个或哪些位置由于步移MP修饰的位置而产生了特异性增强的方法。通过针对每个测试位置的中靶与脱靶切割比来评估特异性增强的幅度,并且当决定哪个或那些MP修饰的位置可能有益于gRNA的整体性能时,从业者可以将这些值与每种设计测得的中靶切割百分比一起考虑。通过在单一MP在整个引导序列上的渐进步移,还可以鉴定该序列中潜在的可协同性改善特异性的位置,所述协同性改善由该步移所测试的位置上的MP修饰的一种或多种组合所致。

[0270] 在一项相关实验中,使用具有相同20-nt引导序列的HBB sgRNA,评估了对K562细胞中基因组HBB靶基因座的编辑的特异性,其中体内20个碱基对的靶序列与图11A的体外多核苷酸构建体中测试的相同。图11B显示了靶向K562细胞中的人HBB基因的sgRNA中各种类型的修饰的影响,其中将每个合成sgRNA(SEQ ID NO:187-190)与Cas9 mRNA一起共转染,并测量靶基因座和三个脱靶基因座的切割,所述脱靶基因座包括与图11A中评估的相同的脱靶序列脱靶1。针对每个受测的合成sgRNA,计算了切割的靶(中靶)与切割的脱靶多核苷酸(脱靶1)的比值。将比值乘以其对应的中靶切割百分比来计算特异性评分。结果显示,引导序列中的PACE修饰产生了显著的特异性改善,如对所测试的各种类型的修饰通过特异性评分所评估的,特别是相对于主要脱靶活性(在脱靶1)而言。

表6



条目	sgRNA 名	5' → 3'	SEQ. ID. NO.
1	HBB_4MP_1xMP	C*UUG*CCCCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	87
2	HBB_5MP_1xMP	C*UUG*CCCCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	88
3	HBB_6MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	89
4	HBB_7MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	90
5	HBB_8MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	91
6	HBB_9MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	92
7	HBB_10MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	93
8	HBB_11MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	94
9	HBB_12MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	95
10	HBB_13MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	96
11	HBB_14MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	97
12	HBB_15MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	98
13	HBB_16MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	99
14	HBB_17MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	100
15	HBB_18MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	101
16	HBB_19MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	102
17	HBB_1xMP [control]	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	103
18	HBB_16.17MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	125
19	HBB_14.17MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	126
20	HBB_13.17MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	127
21	HBB_10.17MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	128
22	HBB_9.17MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	129
23	HBB_8.17MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	130
24	HBB_7.17MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	131
25	HBB_6.17MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	132
26	HBB_5.17MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	133
27	HBB_14.16MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	134
28	HBB_13.16MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	135
29	HBB_10.16MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	136
30	HBB_9.16MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	137
31	HBB_8.16MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	138
32	HBB_7.16MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	139
33	HBB_6.16MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	140
34	HBB_5.16MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	141
35	HBB_13.14MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	142
36	HBB_10.14MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	143
37	HBB_9.14MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	144
38	HBB_8.14MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	145
39	HBB_6.14MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	146
40	HBB_7.14MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	147
41	HBB_6.14MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	148
42	HBB_10.13MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	149
43	HBB_9.13MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	150
44	HBB_8.13MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	151
45	HBB_7.13MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	152
46	HBB_6.13MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	153
47	HBB_5.13MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	154
48	HBB_6.11MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	155
49	HBB_5.11MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	156
50	HBB_9.10MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	157
51	HBB_8.10MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	158
52	HBB_7.10MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	159
53	HBB_6.10MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	160
54	HBB_5.10MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	161
55	HBB_8.9MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	162
56	HBB_7.9MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	163
57	HBB_6.9MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	164
58	HBB_5.9MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	165
59	HBB_6.8MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	166
60	HBB_5.8MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	167
61	HBB_6.7MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	168
62	HBB_5.7MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	169
63	HBB_5.6MP_1xMP	C*UUGCC*CCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU*U	170
64	HBB_unmodified	CUUGCCCCACAGGCGAGUAGUUUUAGAGCUAGAAUAGCAAGUUAUUAAUAGGCUAGUCCGUUAUACUUGAAAAAGUGGACCGAGUCGGUGCUUU	171

图注

N\* = 核苷酸N的2'-O-甲基-3'-PACE修饰

Ns\* = 核苷酸N的2'-O-甲基-3'-硫代PACE修饰

## 实施例7

[0271] 使用体外切割测定法在溶液中或在两种类型的培养的人类细胞中进行进一步实验,以确定由于MP修饰所致的中靶与脱靶特异性增强。两种细胞类型是K562细胞和诱导的多能干细胞(也称为iPS细胞或iPSC)。在各个孔中用与重组Cas9蛋白预复合的MP修饰的gRNA转染培养的细胞。针对基因组HBB靶基因座的编辑评估了特异性,其中培养细胞中的20个碱基对的靶序列与体外多核苷酸构建体中测定的相同,如图11A和12A中所评估。图12A的条目1-17(SEQ ID NO:87-103)显示与图11A相同的结果。将图12A中的数据根据特异性评分从最高到最低排列。条目18-64(SEQ ID NO:125-171)显示了在靶向人HBB基因中的序列的gRNA中的不同位置处的其他MP修饰对于称为“中靶”(“ON”)的靶多核苷酸的体外切割、以及另行测定的称为“脱靶1”(“OFF1”)的脱靶多核苷酸的切割的影响,分别表示为HBB中靶(ON-target)和HBB脱1-靶(OFF1-target)。对于所有条目,针对测定的每个合成sgRNA计算切割



的靶(中靶)与切割的脱靶多核苷酸(脱靶1)的比值。通过将比值乘以其对应的中靶切割百分比来计算特异性评分。将条目18-64根据特异性评分从最高到最低排列。具有最高评分的条目用阴影表示。

[0272] 图12B显示了基因组HBB靶的编辑结果,所述靶具有与在图12A的体外多核苷酸构建体中测试的相同的20个碱基对的序列。对于图12B,基因组HBB靶位点在测试的人K562细胞和iPS细胞中是内源的。根据在20-nt引导序列中掺入的MP修饰的数目将结果分组。图12A中的条目1-17和图12B中的条目1-12和22-33表示在20-nt引导序列中的不同位置处的单个内部MP的测试,另外在每个sgRNA的5'端的第一个核苷酸间键处具有MP修饰且在3'端的最后一个核苷酸间键处具有另一个MP,以保护其免受核酸外切酶降解。对于图12B的条目1-12和22-33,测试的sgRNA是SEQ ID NO:88、90、92-94、96、97、99、100、103、171和172。对于图12B的条目13-19和34-40,测试的sgRNA是128、129、133、141、160、165和168。这些分组中包括缺乏内部MP修饰的对照的结果,如图12A中的条目14和图12B中的条目7-9和29-31中所示。对在20-nt引导序列中不同位置处具有两个内部MP修饰的sgRNA进行了结果的其他分组,所述sgRNA在每一末端具有一个MP修饰以抑制核酸外切酶,如图12A中的条目18-64和图12B中的条目13-19和34-40所示。不同分组在图中用粗黑线隔开。这些图显示,在体外产生最大的特异性增强的内部MP位置与在两种细胞类型中给出最大增强的那些位置相同,如图12A中的条目1-2和图12B中的条目1-2和22-23所示。由内部MP修饰导致的低一层次的特异性增强用较浅的阴影指示,如图12A中的条目3-5和图12B中的条目3-5和24-26所示。由HBB sgRNA中的内部MP修饰的位置引起的特异性增强的相对等级在体外和体内测定中惊人地一致,如图12A的条目1-17和图12B的条目1-12和22-33所示。

[0273] 将一对内部MP修饰的结果单独地图12A中的条目18-64中和图12B中的条目13-19和34-40中分组,这些体外和体内设计的相对性能,按照“特异性评分”排名,在不同的测定和细胞类型之间引人注目地一致。一种略有不同的评估特异性增强的方式是只考虑中靶与脱靶比,对于图12A和12B中针对HBB的实例,其为“中靶:脱靶1”。使用这一替代的度量,将图12B中的分组根据测量比值重新排名显示在图12C中,所述比值按照分组的最高比值到最低比值排序。图中的不同分组用粗黑线隔开。与我们对图12B中的特异性评分的观察结果相当的是,图12C中呈现的更为简单的特异性评估方式——借助于中靶:脱靶比的更简单的特异性评估在两种细胞类型中显示出相似的结果。实施例6和7证明,特异性增强依赖于在靶向HBB基因的20-nt引导序列中的MP修饰的位置。

[0274] 在使用表5的一些IL2RG sgRNA的实验中,针对K562细胞中的基因组IL2RG靶基因座的编辑评估了特异性。同样,使用表7的VEGFA sgRNA(如下所示),针对K562细胞中的基因组VEGFA靶基因座的编辑评估了特异性。

表 7

条目	sgRNA名称	5' 3'	SEQ.ID. NO.
1	VEGFA_11MP_1xMP	G*GUGAGUGAGU*GUGUGCGUGGUUUUAGAGCUAGAAAUAAGCAAGUUAAAAUAAAGGCUA GUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUU*U	194
2	VEGFA_10MP_1xMP	G*GUGAGUGAG*UGUGUGCGUGGUUUUAGAGCUAGAAAUAAGCAAGUUAAAAUAAAGGCUA GUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUU*U	195
3	VEGFA_9MP_1xMP	G*GUGAGUGA*GUGUGUGCGUGGUUUUAGAGCUAGAAAUAAGCAAGUUAAAAUAAAGGCUA GUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUU*U	196
4	VEGFA_7MP_1xMP	G*GUGAGU*GAGUGUGCGUGGUUUUAGAGCUAGAAAUAAGCAAGUUAAAAUAAAGGCUA GUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUU*U	197
5	VEGFA_5MP_1xMP	G*GUGA*GUGAGUGUGCGUGGUUUUAGAGCUAGAAAUAAGCAAGUUAAAAUAAAGGCUA GUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUU*U	198
6	VEGFA_1xMP(对照)	G*GUGAGUGAGUGUGCGUGGUUUUAGAGCUAGAAAUAAGCAAGUUAAAAUAAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUU*U	199

[0275] 以这种方式,在体内测试了20-nt引导序列。图14显示了IL2RG sgRNA和VEGFA sgRNA中的各种类型的修饰在K562细胞中的影响。将每个合成sgRNA与Cas9 mRNA一起共转

染,并测量靶基因座和脱靶基因座脱靶2的切割。针对测定的每个合成sgRNA计算切割的靶(中靶)与切割的脱靶多核苷酸(脱靶2)的比值。通过将比值乘以其对应的中靶切割百分比来计算特异性评分。

[0276] 图14中的结果显示,在靶向IL2RG和VEGFA的合成gRNA的引导序列中的修饰产生了显著的特异性改善,如对所测试的各种类型的修饰通过特异性评分所评估的,特别是相对于脱靶活性(在脱靶2)而言。与条目1相反,条目2显示出由IL2RG sgRNA两端的MP修饰产生的显著特异性增强,其中5'和3'末端核苷酸间键被修饰。与条目1和2相反,条目3和4显示在IL2RG引导序列的内部位置处(即分别在位置5或11处)具有MP修饰的IL2RG sgRNA的令人印象深刻的特异性增强。此外,位置5的MP修饰在图14中的条目1至4中给出了最大的特异性增强。

[0277] 与条目5相反,条目6至10显示在VEGFA引导序列的内部位置处(即分别在位置5、7、9、10或11处)具有MP修饰的VEGFA sgRNA的显著特异性增强。当比较在条目5至10中的增强时,值得注意的是条目6在20-nt引导序列的位置5处的MP修饰给出了针对VEGFA的最大特异性增强。

[0278] 如实施例4至7呈现的,通过在靶向不同的DNA序列的gRNA中在引导序列上进行MP步移并在体外测定中靶和脱靶多核苷酸的切割,在图13中显示了产生特异性增强的在各个引导序列中的MP修饰位置的复合图谱。具体而言,图13显示了来自图9A、9B、10和11A的体外切割结果的比较。图13显示了若干个重要趋势。首先,如条目12中所示,与图13中的所有其他条目相反,在其20-nt引导序列中的位置15处具有MP修饰的gRNA遭受了Cas9介导的切割活性的显著丧失。这表明位置15特别不耐受针对Cas9介导的切割的MP修饰。其次,在各处的条目1-2中存在特异性增强的趋势,如用阴影特异性评分所指示的。这个趋势表明,为了增强gRNA的靶特异性,在各个20-nt引导序列的5'端或其附近掺入MP修饰可能是通常有用的设计策略。对于在图8A和8B中呈现的关于在体外靶向CLTA1基因座的gRNA的5'端的MP修饰的结果,该方法的实用性的另一个例子论述于上面的实施例3中。关于如图11B所示在转染的人K562细胞中靶向HBB基因的gRNA的5'端的MSP修饰,一般方法的实用性的一个另外的例子论述于上面的实施例6中。关于MP修饰的特异性效应的第三个例子在图13的条目3中是显而易见的,其中由不同靶向的20-nt引导序列中的位置6的MP修饰产生的特异性评分构成了每MP步移的最低评分,而不考虑上述位置15的异常效应。对于涉及CLTA4脱靶3、IL2RG脱靶3和HBB脱靶1的每个系列的MP步移,由于位置6的MP修饰所致的特异性增强的最低值相比于由相同引导序列中的相邻位置(即位置4、5或7)的MP修饰产生的特异性评分(分别在各处的条目1、2或4中显示为阴影评分)是特别显著的。

[0279] 实施例3至7证明,活性Cas蛋白和gRNA:Cas蛋白复合物耐受了在引导序列中的特定位置含有修饰的gRNA,因为在gRNA中的许多序列位置处的修饰没有阻止中靶多核苷酸的靶特异性切割。实施例3至7还证明,本发明的gRNA、复合物和方法可以实现至少1.2、或者至少1.5、或者至少2、或者至少2.5、或者至少3、或者至少3.5、或者至少4、或者至少4.5、或者至少5的中靶与脱靶比,和/或该比值高达10、12、15或20。在许多情况下,2'-O-甲基-3'-PACE("MP")修饰通过降低脱靶切割水平同时保持高水平的中靶切割而对比值具有正面影响,正如特异性增强所期望的。

[0280] 作为这些实验的结果,计算出掺入多个降低引导序列上特定定位点的结合能的修饰

将降低已知和未知的脱靶切割事件。

[0281] 如上所述,实施例3-7中使用的DNA构建体的完整序列在表8中列出。靶多核苷酸或脱靶序列以及PAM序列以粗体显示,并且PAM序列还带有下列线。

表8

**CLTA1中1-靶(SEQ ID NO:104):**

GCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAA  
AGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTT  
TTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGAT  
ACATATTTGAATGTATTTAGAAAAATAACAAATAGGGGTTCGCGCA  
CATTTCCCCGAAAAGTGCCACCTAAATTGTAAGCGTTAATATTTTGTT  
AAAATTCGCGTTAAATTTTTGTAAATCAGCTCATTTTTTAACCAATAG  
GCCGAAATCGGCAAAATCCCTTATAAATCAAAGAATAGACCGAGAT  
AGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGA  
ACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGA  
TGGCCCACTACGTGAACCATCACCTAATCAAGTTTTTTGGGGTCGA  
GGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCGATTT  
AGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGG  
AAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCG  
GTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCT  
ACAGGGCGCGTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAG  
GGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGG  
GGATGTGCTGCAAGGCGATTAAGTTGGGTAAACGCCAGGGTTTTCCA  
GTCACGACGTTGTAAAACGACGGCCAGTGAGCGCGCGTAATACGAC  
TCACTATAGGGCGAATTGGGTACGATCGATGCGGCCTCGCAGGCCAA  
AGATGTCTCCCGCATGCGCTC**AGTCCTCATCTCCCTCAAGCAGGCC**  
CTGCTGGTGCACTGAAGAGCCACCCTGTGCGCGTGATATGCAGCTCC  
AGCTTTTGTTCCCTTTAGTGAGGGTTAATTGCGCGCTTGGCGTAATCA  
TGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCA  
CACAACATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCTT  
AATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTT  
TCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAA  
CGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTC  
GCTCACTGACTCGCTGCGCTCGGTCGTTCCGGCTGCGGCGAGCGGTAT

CAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGAT  
AACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGG  
AACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCC  
CCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGA  
AACCCGACAGGACTATAAAGATAACCAGGCGTTTCCCCCTGGAAGCTC  
CCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTC  
CGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTG  
TAGGTATCTCAGTTCGGTGTAGGTCGTTTCGCTCCAAGCTGGGCTGTGT  
GCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACT  
ATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCA  
GCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTG  
CTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGG  
ACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAA  
AAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGC  
GGTGGTTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGG  
ATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTG  
GAACGAAAACCTCACGTTAAGGGATTTTGGTTCATGAGATTATCAAAA  
GGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAA  
TCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAAT  
CAGTGAGGCACCTATCTCAGCGATCTGTCTATTTTCGTTTCATCCATAGTT  
GCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACC  
ATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGG  
CTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGC  
AGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGT  
TGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAA  
CGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCGTTTGG  
TATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATG  
ATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGAT  
CGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGC  
AGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCT  
GTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCG



GCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGC  
CACATAGC

**CLTA1 脱 1-靶(SEQ ID NO:105):**

GCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAA  
AGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTT  
TTTCAATATTATTGAAGCATTATCAGGGTTATTGTCTCATGAGCGGAT  
ACATATTTGAATGTATTTAGAAAAATAACAAATAGGGGTTCCGCGCA  
CATTTCCCCGAAAAGTGCCACCTAAATTGTAAGCGTTAATATTTTGTT  
AAAATTCGCGTTAAATTTTGTAAATCAGCTCATTTTTTAACCAATAG  
GCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGAT  
AGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGA  
ACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGA  
TGGCCCACTACGTGAACCATCACCTAATCAAGTTTTTTGGGGTCGA  
GGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTT  
AGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGG  
AAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCG  
GTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCT  
ACAGGGCGCGTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAG  
GGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGG  
GGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCA  
GTCACGACGTTGTAAAACGACGGCCAGTGAGCGCGCGTAATACGAC  
TCACTATAGGGCGAATTGGGTACGATCGATGCGGCCTCGCAGGGCAA  
AGAGGTCTCCTGTATGCACTCAGTCCTCAACTCCCTCAAGCAGGCG  
ACCCTTGGTGCACTGACAAACCGCTCCTGCGCGTGATATGCAGCTCC  
AGCTTTTGTTCCCTTTAGTGAGGGTTAATTGCGCGCTTGGCGTAATCA  
TGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCA  
CACAACATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCCT  
AATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTT  
TCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAA  
CGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTC

GCTCACTGACTCGCTGCGCTCGGTCGTTTCGGCTGCGGCGAGCGGTAT  
CAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGAT  
AACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGG  
AACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCC  
CCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGA  
AACCCGACAGGACTATAAAGATAACCAGGCGTTTCCCCCTGGAAGCTC  
CCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTC  
CGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTG  
TAGGTATCTCAGTTCGGTGTAGGTCGTTTCGCTCCAAGCTGGGCTGTGT  
GCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACT  
ATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCA  
GCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTG  
CTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGG  
ACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAA  
AAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGC  
GGTGGTTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGG  
ATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTG  
GAACGAAAACCTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAA  
GGATCTTCACCTAGATCCTTTTAAATTAATAAATGAAGTTTTAAATCAA  
TCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAAT  
CAGTGAGGCACCTATCTCAGCGATCTGTCTATTTTCGTTTCATCCATAGTT  
GCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACC  
ATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGG  
CTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGC  
AGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGT  
TGCCGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAA  
CGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCGTTTGG  
TATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATG  
ATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGAT  
CGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGC  
AGCACTGCATAATTCTCTTACTGTTCATGCCATCCGTAAGATGCTTTTCT



GTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCG  
GCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGC  
CACATAGC

**CLTA1 脱 3-靶(SEQ ID NO:106):**

GCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAA  
AGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTT  
TTTCAATATTATTGAAGCATTATCAGGGTTATTGTCTCATGAGCGGAT  
ACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCA  
CATTTCCCCGAAAAGTGCCACCTAAATTGTAAGCGTTAATATTTTGTT  
AAAATTCGCGTTAAATTTTGTAAATCAGCTCATTTTTTAACCAATAG  
GCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGAT  
AGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGA  
ACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGA  
TGGCCCACTACGTGAACCATCACCTAATCAAGTTTTTTGGGGTCGA  
GGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTT  
AGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGG  
AAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCG  
GTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCT  
ACAGGGCGCGTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAG  
GGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGG  
GGATGTGCTGCAAGGCGATTAAGTTGGGTAAACGCCAGGGTTTTCCA  
GTCACGACGTTGTAAAACGACGGCCAGTGAGCGCGCGTAATACGAC  
TCACTATAGGGCGAATTGGGTACGATCGATGCGGCCTCAGGAGAGGG  
AGCCATGCTCATCTCCAGCCCACTCCTCATCCCCCTCAAGCCGGTC  
CCAGGCTGAGAGGCTAAAGCTTGTCTTTGCGCGTGATATGCAGCTCC  
AGCTTTTGTTCCCTTTAGTGAGGGTTAATTGCGCGCTTGGCGTAATCA  
TGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCA  
CACAACATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCTT  
AATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCCGCTT  
TCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAA

CGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTC  
GCTCACTGACTCGCTGCGCTCGGTCTGTCGGCTGCGGCGAGCGGTAT  
CAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGAT  
AACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGG  
AACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCC  
CCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGA  
AACCCGACAGGACTATAAAGATAACCAGGCGTTTCCCCCTGGAAGCTC  
CCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTC  
CGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTG  
TAGGTATCTCAGTTCGGTGTAGGTCGTTTCGCTCCAAGCTGGGCTGTGT  
GCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACT  
ATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCA  
GCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTG  
CTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGG  
ACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAA  
AAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGC  
GGTGGTTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAGG  
ATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTG  
GAACGAAAACCTACGTTAAGGGATTTTGGTCATGAGATTATCAAAAA  
GGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTAAATCAA  
TCTAAAGTATATATGAGTAACTTGGTCTGACAGTTACCAATGCTTAAT  
CAGTGAGGCACCTATCTCAGCGATCTGTCTATTTTCGTTTCATCCATAGTT  
GCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACC  
ATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGG  
CTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGC  
AGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGT  
TGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAA  
CGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCGTTTGG  
TATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATG  
ATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGAT  
CGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGC

AGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCT  
GTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCG  
GCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGC  
CACATAGC

**CLTA4中-靶(SEQ ID NO:107):**

GCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAA  
AGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTT  
TTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGAT  
ACATATTTGAATGTATTTAGAAAAATAACAAATAGGGGTTCGCGCA  
CATTTCCCGAAAAGTGCCACCTAAATTGTAAGCGTTAATATTTTGTT  
AAAATTCGCGTTAAATTTTTGTAAATCAGCTCATTTTTTAACCAATAG  
GCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGAT  
AGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGA  
ACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGA  
TGGCCCACTACGTGAACCATCACCTAATCAAGTTTTTTGGGGTCGA  
GGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTT  
AGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGG  
AAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCG  
GTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCT  
ACAGGGCGCGTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAG  
GGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGG  
GGATGTGCTGCAAGGCGATTAAGTTGGGTAAACGCCAGGGTTTTCCA  
GTCACGACGTTGTAAAACGACGGCCAGTGAGCGCGCGTAATACGAC  
TCACTATAGGGCGAATTGGGTACGATCGATGCGGCCTCAAGAGCTTC  
ACTGAGTAGGATTAAGATATTGCAGATGTAGTGTTCACAGGGTG  
GCTCTTCAGTGCACCAGCGGAACCTGCTGCGCGTGATATGCAGCTCC  
AGCTTTTGTTCCCTTTAGTGAGGGTTAATTGCGCGCTTGGCGTAATCA  
TGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCA  
CACAACATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCCT  
AATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCCGCTT

TCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAA  
CGCGCGGGGAGAGGCGGTTTGGCGTATTGGGCGCTCTTCCGCTTCCTC  
GCTCACTGACTCGCTGCGCTCGGTCGTTTCGGCTGCGGCGAGCGGTAT  
CAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGAT  
AACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGG  
AACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCC  
CCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGA  
AACCCGACAGGACTATAAAGATAACCAGGCGTTTCCCCCTGGAAGCTC  
CCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTC  
CGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTG  
TAGGTATCTCAGTTCGGTGTAGGTCGTTTCGCTCCAAGCTGGGCTGTGT  
GCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACT  
ATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCA  
GCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTG  
CTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGG  
ACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAA  
AAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGC  
GGTGGTTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGG  
ATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTG  
GAACGAAAACCTACGTTAAGGGATTTTGGTTCATGAGATTATCAAAAA  
GGATCTTCACCTAGATCCTTTTAAATTAATAAATGAAGTTTTAAATCAA  
TCTAAAGTATATATGAGTAACTTGGTCTGACAGTTACCAATGCTTAAT  
CAGTGAGGCACCTATCTCAGCGATCTGTCTATTTTCGTTTCATCCATAGTT  
GCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACC  
ATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGG  
CTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGC  
AGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGT  
TGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAA  
CGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCGTTTGG  
TATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATG  
ATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGAT



CGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGC  
AGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCT  
GTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCG  
GCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGC  
CACATAGC

**CLTA4脱1-靶(SEQ ID NO:108):**

GCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAA  
AGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTT  
TTTCAATATTATTGAAGCATTATCAGGGTTATTGTCTCATGAGCGGAT  
ACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCA  
CATTTCCCGAAAAGTGCCACCTAAATTGTAAGCGTTAATATTTTGTT  
AAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAG  
GCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGAT  
AGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGA  
ACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGA  
TGGCCCACTACGTGAACCATCACCTAATCAAGTTTTTTGGGGTCGA  
GGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTT  
AGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGG  
AAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCG  
GTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCT  
ACAGGGCGCGTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAG  
GGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGG  
GGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCA  
GTCACGACGTTGTAAAACGACGGCCAGTGAGCGCGCGTAATACGAC  
TCACTATAGGGCGAATTGGGTACGATCGATGCGGCCTCAAGAGCTTC  
ACTGAGTAGGATTAAGATATTGCAGATGTAGTATTTCCACAGGGTG  
GCTCTTCAGTGCACCAGCGGAACCTGCTGCGCGTGATATGCAGCTCC  
AGCTTTTGTTCCCTTTAGTGAGGGTTAATTGCGCGCTTGGCGTAATCA  
TGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCA  
CACAACATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCCT

AATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCCGCTT  
TCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAA  
CGCGCGGGGAGAGGCGGTTTTCGTATTGGGCGCTCTTCCGCTTCCTC  
GCTCACTGACTCGCTGCGCTCGGTTCGTTCGGCTGCGGCGAGCGGTAT  
CAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGAT  
AACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGG  
AACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCC  
CCCTGACGAGCATCACAAAATCGACGCTCAAGTCAGAGGTGGCGA  
AACCCGACAGGACTATAAAGATAACAGGCGTTTCCCCCTGGAAGCTC  
CCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTC  
CGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTG  
TAGGTATCTCAGTTCGGTGTAGGTCGTTTCGCTCCAAGCTGGGCTGTGT  
GCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACT  
ATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCA  
GCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTG  
CTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGG  
ACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAA  
AAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGC  
GGTGGTTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGG  
ATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTG  
GAACGAAAACCTCACGTAAAGGGATTTTGGTCATGAGATTATCAAAAA  
GGATCTTCACCTAGATCCTTTTAAATTAATAAATGAAGTTTTAAATCAA  
TCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAAT  
CAGTGAGGCACCTATCTCAGCGATCTGTCTATTTTCGTTTCATCCATAGTT  
GCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACC  
ATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGG  
CTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGC  
AGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGT  
TGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAA  
CGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCGTTTGG  
TATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATG

ATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGAT  
CGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGC  
AGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCT  
GTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCG  
GCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGC  
CACATAGC

**CLTA4脱2-靶(SEQ ID NO:109):**

GCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAA  
AGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCT  
TTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCG  
GATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTC  
GCGCACATTTCCCCGAAAAGTGCCACCTAAATTGTAAGCGTTAATAT  
TTTGTTAAAATTCGCGTTAAATTTTTGTAAATCAGCTCATTTTTTAA  
CCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAGAATAG  
ACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCAC  
TATTAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTA  
TCAGGGCGATGGCCCACTACGTGAACCATCACCTAATCAAGTTTTT  
TGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAG  
CCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGA  
AAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCA  
AGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTA  
ATGCGCCGCTACAGGGCGCGTCCCATTGCGCCATTCAGGCTGCGCAA  
CTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAG  
CTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGC  
CAGGGTTTTCCAGTCACGACGTTGTAAAACGACGGCCAGTGAGCG  
CGCGTAATACGACTCACTATAGGGCGAATTGGGTACGATCGATGCG  
GCCTCAAGAGCTTCACTGAGTAGGATTAAGATATTCCAGATGTAGC  
**GTTTCCACAGGGT**GGCTCTTCAGTGCACCAGCGGAACCTGCTGCGC  
GTGATATGCAGCTCCAGCTTTTGTTCCTTTAGTGAGGGTTAATTGC  
GCGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTT



ATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTG  
TAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCG  
TTGCGCTCACTGCCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCT  
GCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATT  
GGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGT  
TCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGG  
TTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCA  
AAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTG  
GCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAAATCG  
ACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATAC  
CAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGAC  
CCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCG  
TGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAG  
GTCGTTGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCC  
CGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGG  
TAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGAT  
TAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGG  
TGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCG  
CTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGA  
TCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTTGTTTGCAA  
GCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTG  
ATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACCTCACGTTA  
AGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCC  
TTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAG  
TAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTAT  
CTCAGCGATCTGTCTATTTTCGTTTCATCCATAGTTGCCTGACTCCCCGT  
CGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGT  
GCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTAT  
CAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTC  
CTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAA  
GCTAGAGTAAGTAGTTCCGCCAGTTAATAGTTTGCGCAACGTTGTTGC

CATTGCTACAGGCATCGTGGTGTACGCTCGTCGTTTGGTATGGCTT  
CATTGAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCC  
ATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGT  
CAGAAGTAAGTTGGCCGCAGTGTATCACTCATGGTTATGGCAGCA  
CTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTG  
ACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGC  
GACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCC  
ACATAGC

**CLTA4脱3-靶(SEQ ID NO:110):**

GCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAA  
AGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCT  
TTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCG  
GATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCC  
GCGCACATTTCCCCGAAAAGTGCCACCTAAATTGTAAGCGTTAATAT  
TTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAA  
CCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAGAATAG  
ACCGAGATAGGGTTGAGTGTGTTCCAGTTTGGAACAAGAGTCCAC  
TATTAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTA  
TCAGGGCGATGGCCCACTACGTGAACCATCACCTAATCAAGTTTTT  
TGGGGTGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAG  
CCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGA  
AAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCA  
AGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTA  
ATGCGCCGCTACAGGGCGCGTCCCATTTCGCCATTTCAGGCTGCGCAA  
CTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAG  
CTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGC  
CAGGGTTTTCCAGTCACGACGTTGTAAAACGACGGCCAGTGAGCG  
CGCGTAATACGACTCACTATAGGGCGAATTGGGTACGATCGATGCG  
GCCTCAAGAGCTTCACTGAGTAGGATTAAGATATTGCAGATGTTGT  
GTTTCCACAGGGTGGCTCTTCAGTGCACCAGCGGAACCTGCTGCGC

GTGATATGCAGCTCCAGCTTTTGTTCCTTTAGTGAGGGTTAATTGC  
GCGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTT  
ATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTG  
TAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCG  
TTGCGCTCACTGCCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCT  
GCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATT  
GGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGT  
TCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGG  
TTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCA  
AAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTG  
GCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAATCG  
ACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATAC  
CAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGAC  
CCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCG  
TGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAG  
GTCGTTGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCC  
CGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGG  
TAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGAT  
TAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGG  
TGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCG  
CTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGA  
TCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTTGTTTGCAA  
GCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTG  
ATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACCTCACGTTA  
AGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCC  
TTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAG  
TAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTAT  
CTCAGCGATCTGTCTATTTTCGTTTCATCCATAGTTGCCTGACTCCCCGT  
CGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGT  
GCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTAT  
CAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTC

CTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAA  
GCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGC  
CATTGCTACAGGCATCGTGGTGTACGCTCGTCGTTTGGTATGGCTT  
CATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCC  
ATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGT  
CAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCA  
CTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTG  
ACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGC  
GACCGAGTTGCTCTTGCCCCGGCGTCAATACGGGATAATACCGCGCC  
ACATAGC

**HBB 中-靶(SEQ ID NO:111):**

GCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAA  
AGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTT  
TTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGAT  
ACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCA  
CATTTCCCCGAAAAGTGCCACCTAAATTGTAAGCGTTAATATTTTGTT  
AAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAG  
GCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGAT  
AGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGA  
ACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGA  
TGGCCCACTACGTGAACCATCACCTAATCAAGTTTTTTGGGGTCGA  
GGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATT  
AGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGG  
AAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCG  
GTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCT  
ACAGGGCGCGTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAG  
GGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGG  
GGATGTGCTGCAAGGCGATTAAGTTGGGTAAACGCCAGGGTTTTCCA  
GTCACGACGTTGTAAAACGACGGCCAGTGAGCGCGCGTAATACGAC  
TCACTATAGGGCGAATTGGGTACGATCGATGCGGCCTCGGCCTCACC

ACCAACTTCATCCACGTTACCTTGCCCCACAGGGCAGTAACCGGC  
AGACTTCTCCTCAGGAGTCAGATGCACCAGCGCGTGATATGCAGCTC  
CAGCTTTTGTTCCTTTAGTGAGGGTTAATTGCGCGCTTGGCGTAATC  
ATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCC  
ACACAACATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCC  
TAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCCGCT  
TTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCA  
ACGCGCGGGGAGAGGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCT  
CGCTCACTGACTCGCTGCGCTCGGTCGTTTCGGCTGCGGCGAGCGGTA  
TCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGA  
TAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAG  
GAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCC  
CCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCG  
AAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCT  
CCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGT  
CCGCCTTTCTCCCTTCGGGAAGCGTGCGCTTTCTCATAGCTCACGCT  
GTAGGTATCTCAGTTCGGTGTAGGTCGTTTCGCTCCAAGCTGGGCTGT  
GTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAA  
CTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGG  
CAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGG  
TGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAA  
GGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGA  
AAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACACCGCTGGTA  
GCGGTGGTTTTTTTGTGTTGCAAGCAGCAGATTACGCGCAGAAAAAA  
GGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAG  
TGGAACGAAAACCTCACGTAAAGGGATTTTGGTCATGAGATTATCAA  
AAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTAAATC  
AATCTAAAGTATATATGAGTAACTTGGTCTGACAGTTACCAATGCTT  
AATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTTCGTTTCATCCATA  
GTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTT  
ACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCAC



CGGCTCCAGATTTATCAGCAATAAACAGCCAGCCGGAAGGGCCGA  
GCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAA  
TTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGC  
GCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCG  
TTTGGTATGGCTTCATTACAGCTCCGGTTCCCAACGATCAAGGCGAGTT  
ACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCC  
TCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGT  
TATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATG  
CTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGT  
GTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAAT  
ACCGCGCCACATAGC

**HBB 脱 1-靶(SEQ ID NO:112):**

GCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAA  
AGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTT  
TTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGAT  
ACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCA  
CATTTCCCCGAAAAGTGCCACCTAAATTGTAAGCGTTAATATTTTGT  
AAAATTCGCGTTAAATTTTTGTAAATCAGCTCATTTTTTAACCAATAG  
GCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGAT  
AGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGA  
ACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGA  
TGGCCCACTACGTGAACCATCACCTAATCAAGTTTTTTGGGGTCGA  
GGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTT  
AGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGG  
AAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCG  
GTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCT  
ACAGGGCGCGTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAG  
GGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGG  
GGATGTGCTGCAAGGCGATTAAGTTGGGTAAACGCCAGGGTTTTCCA  
GTCACGACGTTGTAAAACGACGGCCAGTGAGCGCGCGTAATACGAC

TCACTATAGGGCGAATTGGGTACGATCGATGCGGCCTCGGATAGGAA  
AGGTGAAGTCAGAGCAGTGCTTCAGCCCCACAGGGCAGTAAGGG  
CAGCCTTCCTCTAAATACCAGATTCCCAAAGCGCGTGATATGCAGCTC  
CAGCTTTTGTTCCTTTAGTGAGGGTTAATTGCGCGCTTGGCGTAATC  
ATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCC  
ACACAACATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCC  
TAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCCGCT  
TTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCA  
ACGCGCGGGGAGAGGCGGTTTTCGTATTGGGCGCTCTTCCGCTTCCT  
CGCTCACTGACTCGCTGCGCTCGGTCGTTTCGGCTGCGGCGAGCGGTA  
TCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGA  
TAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAG  
GAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCC  
CCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCG  
AAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCT  
CCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGT  
CCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCT  
GTAGGTATCTCAGTTCGGTGTAGGTCGTTTCGCTCCAAGCTGGGCTGT  
GTGCACGAACCCCCCGTTTCAGCCCGACCGCTGCGCCTTATCCGGTAA  
CTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGG  
CAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGG  
TGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAA  
GGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGA  
AAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTA  
GCGGTGGTTTTTTTGTGTTGCAAGCAGCAGATTACGCGCAGAAAAAAA  
GGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAG  
TGGAACGAAAACCTCACGTTAAGGGATTTTGGTCATGAGATTATCAAA  
AAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATC  
AATCTAAAGTATATATGAGTAACTTGGTCTGACAGTTACCAATGCTT  
AATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTTCGTTTCATCCATA  
GTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTT



ACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCAC  
CGGCTCCAGATTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGA  
GCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAA  
TTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGC  
GCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCG  
TTTGGTATGGCTTCATTCAGCTCCGGTTCCTAACGATCAAGGCGAGTT  
ACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCC  
TCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGT  
TATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATG  
CTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGT  
GTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAAT  
ACCGCGCCACATAGC

**IL2RG 中-靶(SEQ ID NO:113):**

GCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAA  
AGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTT  
TTTCAATATTATTGAAGCATTATCAGGGTTATTGTCTCATGAGCGGAT  
ACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCA  
CATTTCCCCGAAAAGTGCCACCTAAATTGTAAGCGTTAATATTTTGTT  
AAAATTCGCGTTAAATTTTTGTAAATCAGCTCATTTTTTAACCAATAG  
GCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGAT  
AGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGA  
ACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGA  
TGGCCCACTACGTGAACCATCACCTAATCAAGTTTTTTGGGGTCGA  
GGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTT  
AGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGG  
AAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCG  
GTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCT  
ACAGGGCGCGTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAG  
GGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGG  
GGATGTGCTGCAAGGCGATTAAGTTGGGTAAACGCCAGGGTTTTCCA

GTCACGACGTTGTAAAACGACGGCCAGTGAGCGCGCGTAATACGAC  
TCACTATAGGGCGAATTGGGTACGATCGATGCGGCCTCGGGCAGCTG  
CAGGAATAAGAGGGATGTGAATGGTAATGATGGCTTCAACATGGC  
GCTTGCTCTTCATTCCCTGGGTGTAGTCTGCGCGTGATATGCAGCTCC  
AGCTTTTGTTCCTTTAGTGAGGGTTAATTGCGCGCTTGGCGTAATCA  
TGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCA  
CACAACATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCCT  
AATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCCGCTT  
TCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAA  
CGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTC  
GCTCACTGACTCGCTGCGCTCGGTCGTTTCGGCTGCGGCGAGCGGTAT  
CAGCTCACTCAAAGGCGGTAAACGGTTATCCACAGAATCAGGGGAT  
AACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGG  
AACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCC  
CCCTGACGAGCATCACAAAATCGACGCTCAAGTCAGAGGTGGCGA  
AACCCGACAGGACTATAAAGATAACAGGCGTTTCCCCCTGGAAGCTC  
CCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTC  
CGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTG  
TAGGTATCTCAGTTCGGTGTAGGTCGTTTCGCTCCAAGCTGGGCTGTGT  
GCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACT  
ATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCA  
GCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTG  
CTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGG  
ACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAA  
AAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGC  
GGTGGTTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAGG  
ATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTG  
GAACGAAAACCTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAA  
GGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAA  
TCTAAAGTATATATGAGTAACTTGGTCTGACAGTTACCAATGCTTAAT  
CAGTGAGGCACCTATCTCAGCGATCTGTCTATTTTCGTTTCATCCATAGTT

GCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACC  
ATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGG  
CTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGC  
AGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGT  
TGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAA  
CGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCGTTTGG  
TATGGCTTCATTAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATG  
ATCCCCCATGTTGTGCAAAAAGCGGTTAGCTCCTTCGGTCCTCCGAT  
CGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGC  
AGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCT  
GTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCG  
GCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGC  
CACATAGC

**IL2RG 脱 3-靶(SEQ ID NO:114):**

GCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAA  
AGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTT  
TTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGAT  
ACATATTTGAATGTATTTAGAAAAATAACAAATAGGGGTTCCGCGCA  
CATTTCCCCGAAAAGTGCCACCTAAATTGTAAGCGTTAATATTTTGTT  
AAAATTCGCGTTAAATTTTGTAAATCAGCTCATTTTTTAACCAATAG  
GCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGAT  
AGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGA  
ACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGA  
TGGCCCACTACGTGAACCATCACCTAATCAAGTTTTTTGGGGTCGA  
GGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCGATTT  
AGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGG  
AAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCG  
GTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCT  
ACAGGGCGCGTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAG  
GGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGG

GGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCA  
GTCACGACGTTGTAAAACGACGGCCAGTGAGCGCGCGTAATACGAC  
TCACTATAGGGCGAATTGGGTACGATCGATGCGGCCTCCAATATTGAG  
AGTGAATGAAAAGTGTCAGCTGGTAATGATGACTTCAACATAGTCA  
GAACTCTTTGGGCTGTTCCAAACATCAGCGCGTGATATGCAGCTCCA  
GCTTTTGTTCCTTTAGTGAGGGTTAATTGCGCGCTTGGCGTAATCAT  
GGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCAC  
ACAACATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCCTA  
ATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCCGCTTT  
CCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAAC  
GCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCG  
CTCACTGACTCGCTGCGCTCGGTCGTTCCGGCTGCGGCGAGCGGTATC  
AGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATA  
ACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGA  
ACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCC  
CCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAA  
ACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCC  
CTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCC  
GCCTTTCTCCCTTCGGGAAGCGTGCGCTTTCTCATAGCTCACGCTGT  
AGGTATCTCAGTTCGGTGTAGGTCGTTGCTCCAAGCTGGGCTGTGT  
GCACGAACCCCCGTTTCAGCCCGACCGCTGCGCCTTATCCGGTAACT  
ATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCA  
GCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTG  
CTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGG  
ACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAA  
AAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGC  
GGTGGTTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGG  
ATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTG  
GAACGAAAACCTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAA  
GGATCTTCACCTAGATCCTTTTAAATTAATAAATGAAGTTTTAAATCAA  
TCTAAAGTATATATGAGTAACTTGGTCTGACAGTTACCAATGCTTAAT

```

CAGTGAGGCACCTATCTCAGCGATCTGTCTATTTTCGTTTCATCCATAGTT
GCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACC
ATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGG
CTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGC
AGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGT
TGCCGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAA
CGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCGTTTGG
TATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATG
ATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGAT
CGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGC
AGCACTGCATAATTCTCTTACTGTTCATGCCATCCGTAAGATGCTTTTCT
GTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCG
GCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGC
CACATAGC

```

#### 实施例8

[0282] 该实验证明了使用根据本发明的合成gRNA通过Cas9切割刺激的同源定向修复(HDR)过程。提供在两对染色体上具有正常HBB基因的K562细胞(Chr11)。还提供了用于携带镰状细胞病(SCD)突变的HDR的DNA供体模板。该实验试图编辑K562细胞中的HBB基因以添加或敲入SCD突变以产生杂合细胞系。这种细胞系可用于模拟SCD疾病。

[0283] 使用具有和不具有化学修饰的sgRNA。更具体地,使用以下20-nt合成的引导RNA进行实验:未修饰的gRNA(Unmodif),在5'和3'末端具有MP修饰的gRNA(1xMP),在5'位、5'及3'末端具有MP修饰的gRNA(5MP\_1xMP),在11位和5'和3'末端具有MP修饰的gRNA(11MP\_1xMP),和用缓冲液代替sgRNA:Cas9复合物处理的模拟转染细胞(模拟)。每个sgRNA分别与Cas9蛋白预复合,用于与ssDNA寡核苷酸一起转染K562细胞(通过电穿孔),用作HBB靶位点处HDR的修复模板。在每个组中,用供体DNA模板(图15中的供体+)和没有供体DNA模板(供体-在图15中)进行实验。靶位点被sgRNA:Cas9复合物切割以形成双链断裂。通过HDR过程修复断裂,使得供体DNA模板的序列被转染细胞中的内源DNA修复酶用于指导切割的DNA靶序列的修复。

[0284] 转染到K562细胞中作为修复模板的单链供体DNA模板的序列是:

```

TCAGGGCAGAGCCATCTATTGCTTACATITGCTTCTGACACAACCTGTGTTCACTAGCAACCTCAAACA
GACACCATGGTGCACCTGACTCCTgtaGAGAAGTCTGCGGTTACTGCCCTGTGGGCAAGGTGAACGTGGATGAAG
TTGGTGGTGAGGCC (SEQ ID NO:200) .

```

[0285] 寡核苷酸长158个核苷酸,小写“gta”是被HDR敲入的SCD突变的密码子。培养转染的细胞,随后在转染后48小时收获。分离基因组DNA,并通过PCR扩增HBB基因座及其主要脱靶基因座以进行深度测序以定量编辑结果,其包括(i)形成Cas9诱导的插入缺失(图15中的“中”靶插入缺失和“脱”靶插入缺失),(ii)部分HDR(部分“中”靶HDR),以及(iii)安放SCD密码子(“中”靶HDR和“脱”靶HDR)的完整HDR。测量中靶位点(图15中的“中”基因座)处的所有

三种类型的编辑,在Chr9上的基因组中一处主要脱靶位点检测到三种类型的编辑中的两种(图15中的“脱”基因座)。

[0286] 图15总结了来自人K562细胞中HBB基因的同源定向修复(HDR)的这些实验的结果。对于“‘中’基因座”,报告的结果(从左到右)为:中靶HDR%,部分中靶HDR%,和中靶插入缺失%。对于“‘脱’基因座”,报告的结果(从左到右)为:脱靶HDR%和脱靶插入缺失%。

[0287] 该实验证明,用于增强特异性的化学修饰,特别是在20-nt引导序列中第5位和第11位的MP修饰,不仅大大减少了脱靶插入缺失的形成,而且还显著降低了脱靶HDR活性。注意到,在该实验中,脱靶位点与ssDNA修复模板具有足够的序列相似性,从而同样受到了HDR编辑。

#### 实施例9

[0288] 在该实验中,以两种不同但互补的方式评估合成引导RNA的脱靶活性。使用上文关于实施例7和图12B所述的相同体内测定,使用了几种化学合成的引导RNA来结合并切割HBB靶位点和脱靶位点。因此,对于第一次评估,通过PCR扩增人K562细胞的基因组DNA中的由实施者指定的多核苷酸序列或基因座,对于第二次评估,通过使用市售的基因座富集试剂盒扩增基因组DNA的片段中的约1,000个多核苷酸序列或基因座(由实施者指定)的文库。更具体地,在第一次评估中,通过对基因组DNA进行PCR以扩增16个脱靶位点来评估脱靶活性,所述16个脱靶位点构成人基因组中相对于HBB中20-bp靶序列具有3个或更少错配的序列,并且在相对于靶序列的21-23位也具有NGG或NAG PAM序列;这样的PAM序列是Cas9识别所需要的。在第二次评估中,通过使用生物素化的寡核苷酸诱饵(来自Agilent Technologies的SureSelect Target Enrichment Kit)对来自K562细胞的基因组DNA捕获靶位点和脱靶位点来评估脱靶活性。诱饵被设计成可选择性捕获人类基因组中的如下所述的序列(总共大约960个序列):其相对于HBB中的20-bp靶序列具有5个或更少的错配,并且在21-23位还具有NGG或NAG PAM序列。中靶序列和脱靶序列是:



表9

HBH中靶座位和相似的靶座位	人类基因组中的靶位点序列 PAM序列加下划线；错配以斜体表示；凸出以小写字母表示	基因组坐标 在人类基因组GRCh38/hg38中	SEQ ID NO.
HBH中靶	<u>CTTGCCCCACAGGGCAGTAACGG</u>	Chr11: 5,226,968-5,226,990	201
靶1	<u>TCAGCCCCACAGGGCAGTAAGGG</u>	Chr9: 101,833,584-101,833,606	202
靶2	<u>CTTGCCCCACAGGGCAGTAAGGG</u>	Chr17: 78,243,545-78,243,567	203
靶3	<u>CCTC/CCCACAGGGCAGTAAGG</u>	Chr17: 68,628,098-68,628,120	204
靶4	<u>CTTGCCCC/AGAGGGCACTAAAGG</u>	Chr11: 76,420,075-76,420,097	205
靶5	<u>GCTGCCCCACAGGGCAGCAAGG</u>	Chr12: 124,319,282-124,319,304	206
靶6	<u>CCTG/CCCACAGGGCAGGAAGGG</u>	Chr19: 33,890,015-33,890,037	207
靶7	<u>GTGCCCCACAGGGCAGGAATGG</u>	ChrX: 75,786,405-75,786,427	208
靶8	<u>CTCGCCCC/CAAGGGCAGTAATGG</u>	Chr2: 11,637,669-11,637,691	209
靶9	<u>ATTGCCCCACAGGGCAGTGACGG</u>	Chr12: 93,155,409-93,155,431	210
靶10	<u>CTTGCAACCACAGAGCACTAAGGG</u>	Chr19: 37,048,134-37,048,156	211
靶11	<u>CTTGCCCCACAGGGC/GTTAAGG</u>	Chr2: 79,966,568-79,966,590	212
靶12	<u>CTTGCCCCACAGGGCACTGAAGGG</u>	Chr19: 48,638,948-48,638,970	213
靶13	<u>CCTGCCCCACAGGGCAGCCAGG</u>	Chr15: 87,945,584-87,945,606	214
靶14	<u>CTcTGCCCCACAGGGCCAGGAAGGG</u>	Chr5: 132,087,686-132,087,708	215
靶15	<u>CTTGcCCCCACAGGGCAGTGCCGG</u>	Chr16: 535,479-535,501	216
靶16	<u>CTTGCCCCACAGGGTCA/TAATAG</u>	Chr18: 45,734,262-45,734,284	217

[0289] 对于两种评估,使用以下20-nt合成的引导RNA:未修饰的gRNA(Unmodif),在5'和3'末端具有MP修饰的gRNA(1xMP),在第5位和5'和3'末端具有MP修饰的gRNA(5MP\_1xMP),和用缓冲液而不是sgRNA:Cas9复合物处理的模拟转染的细胞(模拟)。第一次评估还包括在11位及5'和3'末端具有MP修饰的gRNA(11MP\_1xMP)。

[0290] 使用与前述实施例中使用的相同程序,用gRNA和Cas9转染人K562细胞。更具体地,从ATCC获得人K562细胞,并在补充有10%牛生长血清(Thermo Fisher)的RPMI 1640培养基中培养。根据制造商的说明,使用Lonza 4D-Nucleofector(96孔Shuttle设备,程序FF-120)对K562细胞(在第3至9代内)进行核转染。核转染条件使用Lonza SF细胞系试剂盒(V4SC-2960),在20uL培养基中含有0.2百万个细胞,125皮摩尔化学修饰的sgRNA和50皮摩尔的重组Cas9蛋白(Thermo Fisher)。将细胞在37℃下在环境氧和5%二氧化碳中培养。转染后48

小时收获培养的细胞。

[0291] 对于第二次评估,设计包含靶富集诱饵的文库(使用来自Agilent的SureSelect富集平台)以具有重叠序列覆盖(也称为“层叠”(tiling))以捕获大致以HBB靶位点为中心的基因组DNA的1-Kbp片段。该文库还包含类似地层叠的诱饵,以捕获以人类基因组中以960个离散的脱靶位点为中心的1-Kbp区段。这960个脱靶位点被认为构成人类基因组中所有相对于HBB中CRISPR-Cas9靶向的20bp靶序列具有5个或更少的错配序列,并且还具有邻接20-bp脱靶序列的NGG或NAG PAM序列,这是Cas9识别和结合等所需要的。每次处理每个一式三份样品分离基因组DNA,并根据Agilent SureSelect HiSeq方案处理每种分离物。

[0292] 对于两种评估,使用Illumina试剂在Illumina HiSeq 4000测序仪上对扩增或捕获的DNA进行测序,用于双端(paired-end)2x150-bp测序读段。预处理HiSeq原始数据,抛弃每个中靶或脱靶序列的与中靶或脱靶位点在Cas9切割位点的任一侧上没有至少30bp重叠的读段。使用设置为默认参数的BWA-MEM(bwa-0.7.10)软件将读段定位到人类基因组。从分析中丢弃了产生不一致的双端定位、低质量定位或二次定位的读段。根据是否在Cas9切割的定位切割位点的10bp内具有插入或缺失序列(即插入缺失),将每个保留的定位读段评定为具有插入缺失或不插入缺失。将每个中靶或脱靶位点的定位读数根据它们是否具有插入缺失进行分箱(bin),并且将每个箱的读段计数用于计算中靶位点处形成的%插入缺失,同样在960个脱靶位点的每一个处形成的%插入缺失。将%插入缺失结果用于计算中靶:脱靶切割比和特异性得分(未显示)。

[0293] 图16总结了来自第一次评估的结果,其测量了HBB中靶位点处、以及16个修饰的或未修饰的sgRNA的类似的脱靶位点处的插入缺失形成。PCR扩增子的深度测序使用每次处理的分离的基因组DNA产生了17个位点(靶位点和16个脱靶位点)的序列读段。所有修饰的或未修饰的gRNA在中靶位点处均产生了高百分比的插入缺失。在脱靶1位点,未修饰的gRNA和1xMPgRNA也产生了高百分比的插入缺失;相反,5MP\_1xMP gRNA和11MP\_1xMP gRNA给出的插入缺失百分比低得多。在脱靶5位点,5MP\_1xMP gRNA和11MP\_1xMP gRNA比未修饰的gRNA和1xMP gRNA产生低得多的插入缺失百分比。在其他脱靶位点,所有gRNA产生非常低百分比的插入缺失。该评估表明,用于增强特异性的化学修饰,特别是在20-nt引导序列中第5位和第11位的MP修饰,相当显著地减少了人类基因组中如下所述的序列中脱靶插入缺失的形成:与HBB中20-bp靶序列具有3个或更少的错配,并且在21-23位也具有NGG或NAG PAM序列。

[0294] 图17总结了来自两次评估的扩增的DNA基因座(PCR)或捕获的DNA基因座(SureSelect)的深度测序分析的结果。图17显示了使用以下20-nt合成的引导RNA的实验在HBB靶位点(HBB中靶)和三个脱靶位点(脱靶1,脱靶5和脱靶9)测量的插入缺失形成百分比:未修饰的gRNA(HBB\_unmodif(PCR))和HBB\_unmodif(SureSelect)),在5'和3'末端具有MP修饰的gRNA(HBB\_1xMP(PCR)和HBB\_1xMP(SureSelect)),在第5位和5'和3'末端具有MP修饰的gRNA(HBB\_5xMP(PCR)和HBB\_5xMP(SureSelect)) and 用缓冲液代替sgRNA:Cas9复合物处理的模拟转染细胞(模拟(PCR)和模拟(SureSelect))。检测到的插入缺失的百分比在两次评估之间是一致的,并且通过SureSelect富集分析捕获的DNA发现了与对通过PCR扩增的离散DNA基因座的分析中发现的相似的插入缺失百分比。所有修饰的或未修饰的gRNA在HBB中靶位点处均产生了高百分比的插入缺失。在脱靶1位点,HBB\_unmodif(PCR)和HBB\_unmodif(SureSelect)实验以及HBB\_1xMP(PCR)和HBB\_1xMP(SureSelect)实验显示出高百分比的插

入缺失。相反,通过PCR和SureSelect评估,5MP 1xMPgRNA和HMP\_1xMP gRNA显示出低得多的插入缺失百分比。在脱靶5位点,使用5MP 1xMP gRNA(即HBB 5xMP gRNA(PCR)和HBB 5xMP gRNA(SureSelect))的两种评估均产生比未修饰的gRNA和HBB 1xMP gRNA低得多的插入缺失百分比。在脱靶9位点,所有测试的gRNA都具有非常低的插入缺失百分比。在通过SureSelect富集评估的剩余的957个脱靶基因座中,仅32个产生的插入缺失的水平高于0.1%的检测限(结果未显示),并且其中最高的插入缺失%小于0.5%。由此可见,用于增强特异性的化学修饰,特别是在20-nt引导序列中第5位的MP修饰,大大减少了脱靶插入缺失的形成。特异性的这种增强可以通过评估PCR扩增的基因组基因座或使用靶向文库富集试剂盒(如SureSelect)富集的基因组基因座来证明。

#### 示例性实施方案

根据当前公开的主题提供的示例性实施方案包括但不限于权利要求和以下实施方案:

A1.一种合成的引导RNA,其包含:

- (a) crRNA区段,其包含(i)能够与靶多核苷酸杂交的引导序列,(ii)茎序列;和
- (b) tracrRNA区段,其包含与所述茎序列部分或完全互补的核苷酸序列,

其中所述引导序列包含至少一个增强特异性的修饰,其中所述合成的引导RNA具有gRNA功能性。

A2.实施方案A1的合成的引导RNA,其中所述至少一个增强特异性的修饰减弱所述引导序列与所述靶多核苷酸之间的杂交。

A3.实施方案A1的合成的引导RNA,其中所述至少一个增强特异性的修饰减弱所述引导序列与脱靶多核苷酸之间的杂交。

A4.实施方案A1的合成的引导RNA,其中所述至少一个增强特异性的修饰增强所述引导序列与所述靶多核苷酸之间的杂交,并且减弱所述引导序列与脱靶多核苷酸之间的杂交。

A5.任何前述实施方案的合成的引导RNA,进一步包含位于所述引导RNA的5'端或/和3'端上的至少一个增强稳定性的修饰。

A6.任何前述实施方案的合成的引导RNA,其中所述引导序列包含锁定区、取样区和种子区,并且所述至少一个增强特异性的修饰存在于所述取样区和/或所述种子区中。

A7.实施方案A6的合成的引导RNA,其中所述至少一个增强特异性的修饰存在于所述种子区和/或所述取样区和/或所述锁定区中。

A8.任何前述实施方案的合成的引导RNA,其中所述至少一个增强特异性的修饰包括核苷酸间键修饰。

A9.任何前述实施方案的合成的引导RNA,其中所述至少一个增强特异性的修饰包括化学修饰的核碱基。

A10.任何前述实施方案的合成的引导RNA,其中所述至少一个增强特异性的修饰位于至少一个核苷酸糖模块中。

A11.任何前述实施方案的合成的引导RNA,其中所述至少一个增强特异性的修饰包括3'-磷酸基乙酸核苷酸间键、3'-磷酸基乙酸甲酯核苷酸间键、3'-甲基磷酸酯核苷酸间键、3'-硫代磷酸基乙酸核苷酸间键、3'-甲基硫代磷酸酯核苷酸间键、3'-硼烷磷酸核苷酸

间键,或其组合。

A12.任何前述实施方案的合成的引导RNA,其包含赋予C3'-内糖皱褶的2'修饰。

A13.任何前述实施方案的合成的引导RNA,其中所述至少一个增强特异性的修饰包括:

(a) 2'-脱氧核糖、2'-脱氧-2'-氟阿拉伯呋喃糖基、2'-脱氧-2'-氟呋喃核糖基,糖类,如具有2'-0-苯基、2'-苯硫基、2'-S-苯硫基、2'-甲基、2'-乙基、2'-丙基、2'-烯丙基、2'-烯丙基苯基、2'-甲基羟基、2'-甲氧基甲基、2'-0-氨基甲酸酯、2'-0-乙基氨基、2'-0-烯丙基氨基、2'-0-丙基氨基或2'-0-取代的苯基的核糖,或其组合;

(b) 膦酰基乙酸、硫代膦酰基乙酸、膦酰基丙酸酯、膦酰基硫代丙酸酯、甲基膦酸酯、甲基硫代磷酸或硼烷膦酸;或其组合;或

(c) (a) 和 (b) 的组合。

A14.实施方案A8、A9、A11或A13中任一项的合成的引导RNA,进一步包含2'-0-甲基修饰。

A15.任何前述实施方案的合成的引导RNA,其中所述至少一个增强特异性的修饰是非结构化核酸("UNA")、解锁核酸('ULNA')、脱碱基核苷酸、包含- $\text{PO}_4\text{Y}-(\text{CR}^3_2)_m-\text{PO}_4\text{Y}-$ 的亚烷基间隔子、或包含- $\text{PO}_4\text{Y}-(\text{CR}^3_2\text{CR}^3_2\text{O})_n-\text{PO}_3\text{Y}-$ 的乙二醇间隔子,其中m是2、3或4,n是1、2或3,每个 $\text{R}^3$ 独立地选自下组:H、烷基和取代的烷基,并且每个Y为H或负电荷。

A16.实施方案A1至A9、A11至A13或A15中任一项的合成的引导RNA,其中所述至少一个增强特异性的修饰不包括核碱基修饰。

A17.实施方案A1至A15中任一项的合成的引导RNA,其中所述至少一个增强特异性的修饰是选自下组的核碱基:2-硫代U、2-硫代C、4-硫代U、6-硫代G、2-氨基A、2-氨基嘌呤、次黄嘌呤、7-脱氮鸟嘌呤、7-脱氮-8-氮杂鸟嘌呤、7-脱氮腺嘌呤、7-脱氮-8-氮杂腺嘌呤、5-甲基C、5-甲基U、5-羟甲基胞嘧啶、5-羟甲基尿嘧啶、5,6-脱氢尿嘧啶、5-乙炔基胞嘧啶、5-氨基烯丙基U、5-氨基烯丙基C、脱碱基核苷酸、UNA碱基、异胞嘧啶(isoC)、异鸟嘌呤(isoG)、5-甲基-嘧啶、x(A、G、C、T、U)、y(A、G、C、T、U)及其组合。

A18.任何前述实施方案的合成的引导RNA,其中所述至少一个增强特异性的修饰是选自下组的核苷酸或核苷酸类似物:5-硝基吡啶、水粉蕈素、肌苷、二氨基嘌呤、脱碱基连接、和脱碱基荧光团连接,如3-0-基-2-(4-丁基酰胺基荧光素)丙基-1-0-基-磷酸二酯。

A19.任何前述实施方案的合成的引导RNA,其中所述至少一个增强特异性的修饰包括这样的修饰:相对于没有该增强特异性的修饰的双链体的解链温度( $T_m$ )而言,该修饰降低由合成的引导RNA和靶多核苷酸形成的第一DNA/RNA双链体的 $T_m$ 。

A20.实施方案A19的合成的引导RNA,其中所述至少一个增强特异性的修饰使 $T_m$ 降低约0.5°C/修饰,或者约0.5-1.0°C/修饰,或者约1.0-2.0°C/修饰,或者2-8°C/修饰。

A21.任何前述实施方案的合成的引导RNA,其中所述引导序列包含20个核苷酸并且任选地包含5'-突出端序列。

A22.实施方案A21的合成的引导RNA,其中所述引导序列包含或为从所述引导序列的5'端计数的核苷酸1至20,并且在核苷酸1处、或者在核苷酸1和2处、或者在核苷酸1、2和3处、或者在核苷酸1、2、3和4处;或者在核苷酸1、2、3、4和5处具有至少一个增强特异性的修饰。

A23. 实施方案A21的合成的引导RNA, 其中所述引导序列由从所述引导序列的5'端计数的核苷酸1至20-N组成, 其中N是-10至10之间的整数(任选地在-10至6之间), 并且所述至少一个增强特异性的修饰在4-N至20-N的核苷酸内, 或者在5-N至20-N的核苷酸内, 或者在10-N至20-N的核苷酸内, 或者在13-N至20-N的核苷酸内, 或者在13-N至14-N或16-N至19-N的核苷酸内, 或者在13-N至14-N或16-N至18-N的核苷酸内。

A24. 实施方案A21的合成的引导RNA, 其中所述引导序列由从所述引导序列的5'端计数的核苷酸1至20-N组成, 其中N是-10至10之间的正整数或负整数(任选地在-10至6之间), 并且所述引导序列包含一个增强特异性的修饰, 其位于核苷酸11-N、12-N、13-N、14-N、16-N、17-N、18-N、19-N或20-N处, 或者位于核苷酸13-N、14-N、16-N、17-N、18-N、19-N或20-N处, 或者位于核苷酸13-N、14-N、16-N、17-N或18-N处。

A25. 任何前述实施方案的合成的引导RNA, 其进一步包含在引导RNA的5'端的核苷酸1开始的至少一个增强稳定性的修饰以及在3'端的五个末端核苷酸内的至少一个增强稳定性的修饰。

B1. 一种gRNA:Cas蛋白复合物, 其包含任何前述实施方案的合成的引导RNA和Cas蛋白, 其能够切割、切刻或结合靶多核苷酸, 或具有切割活性、切口活性和/或结合活性。

C1. 一种切割、切刻或结合靶多核苷酸的方法, 其包括:

使所述靶多核苷酸与实施方案B1的gRNA:Cas蛋白复合物接触, 和切割、切刻或结合所述靶多核苷酸。

C2. 实施方案C1的方法, 其中与未修饰的gRNA相比, 包含至少一个增强特异性的修饰的合成的引导RNA降低脱靶多核苷酸的切割、切刻或结合。

C3. 实施方案C2的方法, 其中所述至少一个脱靶多核苷酸被CRISPR相关蛋白切割、切刻或结合, 并且其中切割、切刻或结合的靶多核苷酸与切割、切刻或结合的脱靶多核苷酸的比值为至少1.2、或者至少1.5、或者至少2、或者至少2.5、或者至少3、或者至少3.5、或者至少4、或者至少4.5、或者至少5、或者至少10、12、15或20。

D1. 一种制备合成的引导RNA的方法, 其包括:

选择基因组中的靶多核苷酸, 其中所述靶多核苷酸包含相应的核苷酸序列; 鉴定基因组中的至少一个脱靶多核苷酸, 其中所述脱靶多核苷酸包含相应的核苷酸序列;

将所述靶多核苷酸的序列与所述脱靶多核苷酸的序列比对以鉴定这两个序列的一个或多个相同的部分;

设计合成的引导RNA, 其包含:

(a) crRNA区段, 其包含(i)能够与靶多核苷酸杂交的引导序列, (ii)茎序列; 和

(b) tracrRNA区段, 其包含与所述茎序列部分或完全互补的核苷酸序列, 其中所述引导序列在其序列的一部分内包含增强特异性的修饰, 所述部分与所述鉴定的相同部分之一互补。

D2. 实施方案D1的方法, 其中所述至少一个增强特异性的修饰减弱所述引导序列与所述靶多核苷酸之间的杂交。

D3. 实施方案D1的方法, 其中所述至少一个增强特异性的修饰减弱所述引导序列与脱靶多核苷酸之间的杂交。

D4. 实施方案D1的方法, 其中所述至少一个增强特异性的修饰增强所述引导序列

与所述靶多核苷酸之间的杂交,并且减弱所述引导序列与脱靶多核苷酸之间的杂交。

D5.任何前述实施方案的方法,进一步包括合成所设计的引导RNA。

D6.任何前述实施方案的方法,其中通过如下方式鉴定脱靶多核苷酸:预测脱靶位点的算法,如在<http://www.rgenome.net/Cas-OFFinder>; <https://cm.jefferson.edu/Off-Spotter>;或<http://crispr.mit.edu>中可找到的那些;或其他用于在实际情况下鉴定和定量脱靶位点激活的技术,如在Tsai等人(2015)Nat.Biotechnol.33,187-97;Ran等人(2015)Nature520,186-91;和/或Frock等人(2015)Nat.Biotechnol.33,179-86中所公开的。

D7.任何前述实施方案的方法,其进一步包括:

鉴定所述靶多核苷酸和所述脱靶多核苷酸之间的至少一个区别核苷酸位置,其中所述靶多核苷酸和所述脱靶多核苷酸在所述至少一个区别位置具有不同的核苷酸,和

使合成的引导RNA中包括与所述靶多核苷酸中的所述至少一个区别位置处的核苷酸匹配的核苷酸。

D8.任何前述实施方案的方法,其中相对于没有该增强特异性的修饰的双链体的解链温度(“ $T_m$ ”)而言,所述增强特异性的修饰降低由所述合成的引导RNA的引导序列和所述靶多核苷酸形成的第一DNA/RNA双链体的 $T_m$ 。

D9.实施方案D8的方法,其中所述增强特异性的修饰使第一DNA/RNA双链体的 $T_m$ 降低约0.5°C/修饰,或者约0.5-1°C/修饰,或者约1-2.0°C/修饰,或者约2-8°C/修饰。

D10.实施方案D8的方法,其中所述增强特异性的修饰使第一DNA/RNA双链体的 $T_m$ 降低至少约1°C、至少约2°C、至少约3°C、至少约4°C、至少约5°C、和/或至多约6°C,或者至多约8°C、或者至多约10°C、或者至多约13°C,例如使 $T_m$ 从降低约1°C至降低约13°C,或者降低约1°C至降低约6°C。

D11.实施方案D8的方法,其中所述增强特异性的修饰降低由合成的引导RNA的引导序列和至少一个脱靶多核苷酸形成的第二DNA/RNA双链体的 $T_m$ 。

D12.实施方案D8的方法,其中所述第一DNA/RNA双链体的 $T_m$ 高于所述第二DNA/RNA双链体的 $T_m$ ,例如至少高至少约0.5°C,或者高至少约1°C。

E1.一种切割、切刻或结合靶多核苷酸的方法,其包括:

选择基因组中的靶多核苷酸;

提供和/或设计包含以下区段的合成的引导RNA:

(a) crRNA区段,其包含(i)能够与靶多核苷酸杂交的引导序列,(ii)茎序列;和

(b) tracrRNA区段,其包含与所述茎序列部分或完全互补的核苷酸序列,

其中所述引导序列包括增强特异性的修饰;

计算所述引导序列和所述靶多核苷酸的DNA/RNA双链体的解链温度( $T_m$ );和

形成包含Cas蛋白和所述合成的引导RNA的gRNA:Cas蛋白复合物;

在距 $T_m$  10°C以内、或者距 $T_m$  5°C以内、或者在大约 $T_m$ 的温度下使所述靶多核苷酸与gRNA:Cas蛋白复合物接触;和

通过所述接触来切割、切刻或结合所述靶多核苷酸。

F1.一种切割、切刻或结合靶多核苷酸的方法,其包括:

选择基因组中的靶多核苷酸;



提供和/或设计包含以下区段的合成的引导RNA:

(a) crRNA区段,其包含(i)能够与靶多核苷酸杂交的引导序列,(ii)茎序列;和

(b) tracrRNA区段,其包含与所述茎序列部分或完全互补的核苷酸序列,其中所述引导序列包括至少一个增强特异性的修饰;选择所述修饰或修饰的组合使得所述引导序列和所述靶多核苷酸的DNA/RNA双链体的解链温度( $T_m$ ) 在25℃至49℃之间;

形成包含Cas蛋白和所述设计的合成的引导RNA的gRNA:Cas蛋白复合物;在 $T_m$  12℃以内、或者在距 $T_m$  8℃以内、或者在距 $T_m$  5℃以内、或者在大约 $T_m$ 的温度下使所述靶多核苷酸与gRNA:Cas蛋白复合物接触;和

切割、切刻或结合所述靶多核苷酸。

G1. 一种切割、切刻或结合靶多核苷酸的方法,其包括:

选择基因组中的靶多核苷酸;

提供和/或设计包含以下区段的合成的引导RNA:

(a) crRNA区段,其包含(i)能够与靶多核苷酸杂交的引导序列,(ii)茎序列;和

(b) tracrRNA区段,其包含与所述茎序列部分或完全互补的核苷酸序列,其中所述引导序列包括至少一个增强特异性的修饰;

选择或在所述引导序列中包括所述修饰或修饰组合,使得所述引导序列和所述靶多核苷酸的DNA/RNA双链体的解链温度( $T_m$ )比未修饰的gRNA/靶双链体的 $T_m$ 低至少0.5-1℃、或低至少1-3℃或低至少1-12℃;和

形成包含Cas蛋白和所述设计的合成的引导RNA的gRNA:Cas蛋白复合物;在距 $T_m$  12℃以内、或者在距 $T_m$  8℃以内、或者在距 $T_m$  5℃以内、或者在大约 $T_m$ 的温度下使所述靶多核苷酸与gRNA:Cas蛋白复合物接触;和切割、切刻或结合所述靶多核苷酸。

EFG2. 任何前述实施方案E1、F1或G1的方法,其中所述至少一个增强特异性的修饰减弱所述引导序列与所述靶多核苷酸之间的杂交。

EFG3. 实施方案E1、F1或G1中任一项的合成的引导RNA,其中所述至少一个增强特异性的修饰减弱所述引导序列与脱靶多核苷酸之间的杂交。

EFG4. 实施方案E1、F1或G1中任一项的合成的引导RNA,其中所述至少一个增强特异性的修饰增强所述引导序列与所述靶多核苷酸之间的杂交,并且减弱所述引导序列与脱靶多核苷酸之间的杂交。

H1. 任何前述实施方案的方法,其中所述切割、切刻或结合在体外发生。

H2. 任何前述实施方案的方法,其中所述切割、切刻或结合在细胞中发生。

H3. 实施方案H2的方法,其中在使所述靶多核苷酸与所述gRNA:Cas蛋白复合物接触之前将细胞从多细胞来源分离。

H4. 实施方案H3的方法,其中所述来源是植物、动物、多细胞原生生物或真菌。

H5. 实施方案H2至H4中任一项的方法,其中在使所述靶多核苷酸与所述gRNA:Cas蛋白复合物接触后,将所述细胞或由其衍生的细胞返回至所述来源。

H6. 任何前述实施方案的方法,其中所述切割、切刻或结合在体内发生。

H7. 任何前述实施方案中任一项实施方案的方法,其中所述Cas蛋白是Cas9。

H8. 任何前述实施方案的方法,其中所述切割或切刻导致基因编辑。

H9. 任何前述实施方案的方法,其中所述切割、切刻或结合导致基因表达的改变。

H10.任何前述实施方案的方法,其中所述切割、切刻或结合导致靶基因的功能性敲除。

H11.任何前述实施方案的方法,进一步包括用外源或内源模板多核苷酸通过同源定向修复来修复切割的靶多核苷酸。

H12.实施方案H11的方法,其中所述外源或内源模板多核苷酸包含至少一个与切割位点任一侧上的序列具有实质序列同一性的序列。

H13.任何前述实施方案的方法,进一步包括通过非同源末端连接修复切割的靶多核苷酸。

H14.任何前述实施方案的方法,其中所述修复步骤产生所述靶多核苷酸的一个或多个核苷酸的插入、缺失或取代。

H15.实施方案H14的方法,其中所述插入、缺失或取代导致从包含所述靶多核苷酸的基因表达的蛋白质中的一个或多个氨基酸变化。

I1.合成的引导RNA分子组或文库,其包含两个或更多个如任何前述实施方案所述的合成的引导RNA。

J1.一种试剂盒,包含任何前述实施方案的合成的引导RNA以及一种或多种其他组分。

K1.RNA分子的阵列,其包含两个或更多个如任何前述实施方案所述的合成的引导RNA。

L1.任何前述实施方案的合成的引导RNA、方法、组或文库、试剂盒或阵列,其包含含有crRNA区段和tracrRNA区段两者的单一RNA链。

L2.任何前述实施方案的合成的引导RNA、方法、组或文库、试剂盒或阵列,其包含两条RNA链,并且所述crRNA区段和所述tracrRNA区段在不同的RNA链中。

L3.任何前述实施方案中的合成的引导RNA、方法、组或文库、试剂盒或阵列,其中所述合成的引导RNA是单引导RNA,其中所述crRNA区段与所述tracrRNA区段通过环L连接。

L4.实施方案L3的合成的引导RNA、方法、组或文库、试剂盒或阵列,其中环L包含1、2、3、4、5、6、7、8、9或10个核苷酸。

L5.实施方案L3或L4的合成的引导RNA、方法、组或文库、试剂盒或阵列,其中环L包含核苷酸序列GNRA,其中N表示A、C、G或U,并且R表示A或G。

L6.实施方案L3至L5中任一项的合成的引导RNA、方法、组或文库、试剂盒或阵列,其中环L包含核苷酸序列GAAA。

L7.实施方案L3至L6中任一项的合成的引导RNA、方法、组或文库、试剂盒或阵列,其中环L包含一个或多个修饰的核苷酸。

L8.实施方案L3至L7中任一项的合成的引导RNA、方法、组或文库、试剂盒或阵列,其中环L包含荧光染料。

L9.任何前述实施方案的合成的引导RNA、方法、组或文库、试剂盒或阵列,其包含一个或多个同位素标记物。

L10.任何前述实施方案的合成的引导RNA、方法、组或文库、试剂盒或阵列,其包含一个或多个荧光标记物。

L11.任何前述实施方案的合成的引导RNA、方法、组或文库、试剂盒或阵列,其中所

述合成的引导RNA包含至少一个2'-O-甲基-3'-磷酸基乙酸(2'-O-甲基-3')-PACE)核苷酸。

L12.任何前述实施方案的合成的引导RNA、方法、组或文库、试剂盒或阵列,其中所述合成的引导RNA包含至少一个2-硫代U。

L13.任何前述实施方案的合成的引导RNA、方法、组或文库、试剂盒或阵列,其中所述合成的引导RNA包含至少一个6-硫代G。

L14.任何前述实施方案的合成的引导RNA、方法、组或文库、试剂盒或阵列,其中所述合成的引导RNA包含至少一个2-硫代C。

L15.任何前述实施方案的合成的引导RNA、方法、组或文库、试剂盒或阵列,其中所述合成的引导RNA包含至少一个2'-脱氧-2'-氟阿拉伯呋喃糖基修饰。

L16.任何前述实施方案的合成的引导RNA、方法、组或文库、试剂盒或阵列,其中所述合成的引导RNA包含至少一个2'-脱氧-2'-氟呋喃核糖基修饰。

L17.任何前述实施方案的合成的引导RNA、方法、组或文库、试剂盒或阵列,其中所述合成的引导RNA包含至少一个2'-O-苯基核糖。

L18.任何前述实施方案的合成的引导RNA、方法、组或文库、试剂盒或阵列,其中所述合成的引导RNA包含至少一个2'-苯硫基核糖。

L19.任何前述实施方案的合成的引导RNA、方法、组或文库、试剂盒或阵列,其中所述合成的引导RNA包含至少一个2'-S-苯硫基核糖。

L20.任何前述实施方案的合成的引导RNA、方法、组或文库、试剂盒或阵列,其中所述合成的引导RNA包含至少一个2'-甲基核糖。

L21.任何前述实施方案的合成的引导RNA、方法、组或文库、试剂盒或阵列,其中所述合成的引导RNA包含至少一个2'-乙基核糖。

L22.任何前述实施方案的合成的引导RNA、方法、组或文库、试剂盒或阵列,其中所述合成的引导RNA包含至少一个2'-丙基核糖。

L23.任何前述实施方案的合成的引导RNA、方法、组或文库、试剂盒或阵列,其中所述合成的引导RNA包含至少一个2'-烯丙基核糖。

L24.任何前述实施方案的合成的引导RNA、方法、组或文库、试剂盒或阵列,其中所述合成的引导RNA包含至少一个2'-烯丙基苯基核糖。

L25.任何前述实施方案的合成的引导RNA、方法、组或文库、试剂盒或阵列,其中所述合成的引导RNA包含至少一个2'-甲基羟基核糖。

L26.任何前述实施方案的合成的引导RNA、方法、组或文库、试剂盒或阵列,其中所述合成的引导RNA包含至少一个2'-甲氧基甲基核糖。

L27.任何前述实施方案的合成的引导RNA、方法、组或文库、试剂盒或阵列,其中所述合成的引导RNA包含至少一个2'-O-氨基甲酸酯核糖。

L28.任何前述实施方案的合成的引导RNA、方法、组或文库、试剂盒或阵列,其中所述合成的引导RNA包含至少一个2'-O-乙基氨基核糖。

L29.任何前述实施方案的合成的引导RNA、方法、组或文库、试剂盒或阵列,其中所述合成的引导RNA包含至少一个2'-O-烯丙基氨基核糖。

L30.任何前述实施方案的合成的引导RNA、方法、组或文库、试剂盒或阵列,其中所述合成的引导RNA包含至少一个2'-O-丙基氨基核糖。

L31.任何前述实施方案的合成的引导RNA、方法、组或文库、试剂盒或阵列,其中所述合成的引导RNA包含至少一个2'-O-取代的苯基核糖。

L32.任何前述实施方案的合成的引导RNA、方法、组或文库、试剂盒或阵列,其中所述合成的引导RNA的引导序列由从所述引导序列的5'端计数的核苷酸1至20-N组成,其中N是-10至10之间的整数(任选地在-10至6之间),并且核苷酸6-N至14-N的区域包含增强特异性的修饰。

M1.一种切割、切刻或结合靶HBB多核苷酸的方法,其包括:

选择靶多核苷酸内HBB基因座中的靶序列;

提供包含以下区段的合成的引导RNA:

(a) crRNA区段,其包含(i)能够与靶HBB多核苷酸杂交的引导序列,(ii)茎序列;和

(b) tracrRNA区段,其包含与所述茎序列部分或完全互补的核苷酸序列,其中所述引导序列由20-N个核苷酸组成,其中N是-10至10之间的整数,并且所述引导序列在选自位置4-N、5-N、7-N、9-N、10-N和11-N的核苷酸处包含至少一个增强特异性的修饰;和

形成包含Cas蛋白和所述合成的引导RNA的gRNA:Cas蛋白复合物;

使所述靶HBB多核苷酸与所述gRNA:Cas蛋白复合物接触;和

切割、切刻或结合所述靶多核苷酸。

M2.实施方案M1的方法,其中所述至少一个增强特异性的修饰在核苷酸11-N处。

M3.实施方案M1或M2中任一项的方法,其中所述至少一个增强特异性的修饰在核苷酸5-N处。

M4.实施方案M1至M3中任一项的方法,其中所述至少一个增强特异性的修饰在核苷酸7-N处。

M5.实施方案M1至M4中任一项的方法,其中所述至少一个增强特异性的修饰在核苷酸10-N处。

M6.实施方案M1至M5中任一项的方法,其中所述至少一个增强特异性的修饰在核苷酸9-N处。

M7.实施方案M1至M6中任一项的方法,其中所述至少一个增强特异性的修饰在核苷酸4-N处。

M8.实施方案M1至M7中任一项的方法,其中所述至少一个增强特异性的修饰包括具有2'-O-甲基-3'-磷酸基乙酸(MP)、2'-O-甲基-3'-硫代磷酸基乙酸(MSP)、2'-脱氧-3'-磷酸基乙酸(DP)、2'-脱氧-3'-硫代磷酸基乙酸(DSP)或其组合的修饰。

M9.实施方案M1至M7中任一项的方法,其中所述至少一个增强特异性的修饰选自赋予C3'-内糖皱褶的2'-修饰、磷酸基乙酸或硫代磷酸基乙酸键修饰及其组合。

M10.实施方案M9的方法,其中所述2'-修饰选自2'-F和2'-O-(2-甲氧基乙基)。

M11.实施方案M1至M10中任一项的方法,其中所述切割、切刻或结合在体外发生。

M12.实施方案M1至M10中任一项的方法,其中所述切割、切刻或结合在细胞中发生。

M13.实施方案M1至M12中任一项的方法,其中所述引导序列进一步在其5'端或/和3'端处包含修饰,任选地一个或多个增强稳定性的修饰选自2'-O-甲基-3'-磷酸基乙酸(MP)、2'-O-甲基-3'-硫代磷酸基乙酸(MSP)、2'-O-甲基-3'-硫代磷酸(MS)、2'-脱氧-3'-磷

酰基乙酸(DP)、2'-脱氧-3'-硫代磷酸酯(DSP)、2'-氟-3'-磷酸基乙酸(FP)、2'-氟-3'-硫代磷酸基乙酸(FSP)、2'-氟-3'-硫代磷酸(FS)或其组合。

M14. 实施方案M1至M13中任一项的方法,其中所述Cas蛋白是Cas9或Cpf1。

M15. 实施方案M1至M14中任一项的方法,其中所述引导RNA是合成的单引导RNA。

M16. 实施方案M1至M15中任一项的方法,其中所述方法包括切割所述靶HBB多核苷酸。

M17. 实施方案M1至M15中任一项的方法,其中所述方法包括切割所述靶HBB多核苷酸。

M18. 实施方案M1至M15中任一项的方法,其中所述方法包括结合所述靶HBB多核苷酸。

M19. 实施方案M1至M18中任一项的方法,其中所述至少一个增强特异性的修饰增加了所述方法的特异性。

N1. 一种合成的引导RNA,其包含:

(a) crRNA区段,其包含(i)能够与靶HBB多核苷酸杂交的引导序列,(ii)茎序列;和

(b) tracrRNA区段,其包含与所述茎序列部分或完全互补的核苷酸序列,其中所述引导序列由20-N个核苷酸组成,其中N是-10至10之间的整数,并且所述引导序列在选自位置4-N、5-N、7-N、9-N、10-N和11-N的核苷酸处包含至少一个增强特异性的修饰,并且其中所述合成的引导RNA具有gRNA功能性。

N2. 实施方案N1的合成的引导RNA,其中所述至少一个增强特异性的修饰在核苷酸11-N处。

N3. 实施方案N1或N2的合成的引导RNA,其中所述至少一个增强特异性的修饰在核苷酸5-N处。

N4. 实施方案N1至N3中任一项的合成的引导RNA,其中所述至少一个增强特异性的修饰在核苷酸7-N处。

N5. 实施方案N1至N4中任一项的合成的引导RNA,其中所述至少一个增强特异性的修饰在核苷酸10-N处。

N6. 实施方案N1至N5中任一项的合成的引导RNA,其中所述至少一个增强特异性的修饰在核苷酸9-N处。

N7. 实施方案N1至N6中任一项的合成的引导RNA,其中所述至少一个增强特异性的修饰在核苷酸4-N处。

N8. 实施方案N1至N7中任一项的合成的引导RNA,其中所述至少一个增强特异性的修饰包括2'-O-甲基-3'-磷酸基乙酸(MP)、2'-O-甲基-3'-硫代磷酸基乙酸(MSP)、2'-脱氧-3'-磷酸基乙酸(DP)、2'-脱氧-3'-硫代磷酸基乙酸(DSP)或其组合。

N9. 实施方案N1至N8中任一项的合成的引导RNA,其中所述至少一个增强特异性的修饰选自:赋予C3'-内糖皱褶的2'-修饰、磷酸基乙酸或硫代磷酸基乙酸键修饰、及其组合。

N10. 实施方案N9的合成的引导RNA,其中所述2'-修饰选自2'-F和2'-O-(2-甲氧基乙基)。

N11. 实施方案N1至N10中任一项的合成的引导RNA,其中所述至少一个增强特异性的修饰减弱所述引导序列和所述靶HBB多核苷酸之间的杂交。

N12. 实施方案N1至N10中任一项的合成的引导RNA, 其中所述至少一个增强特异性的修饰减弱所述引导序列与脱靶多核苷酸之间的杂交。

N13. 实施方案N1至N10中任一项的合成的引导RNA, 其中所述至少一个增强特异性的修饰增强所述引导序列和所述靶HBB多核苷酸之间的杂交, 并且减弱所述引导序列与脱靶多核苷酸之间的杂交。

N14. 实施方案N1至N13中任一项的合成的引导RNA, 其中所述引导RNA是合成的单引导RNA。

N15. 实施方案N1至N14中任一项的方法, 其中所述引导序列由从所述引导序列的5'端计数的核苷酸1至20组成。

01. 一种切割、切刻或结合靶多核苷酸的方法, 其包括:

选择靶多核苷酸;

提供包含以下区段的合成的引导RNA:

(a) crRNA区段, 其包含 (i) 能够与靶多核苷酸杂交的引导序列, (ii) 茎序列; 和

(b) tracrRNA区段, 其包含与所述茎序列部分或完全互补的核苷酸序列, 其中所述引导序列由20-N个核苷酸组成, 其中N是-10至10之间的整数, 并且所述引导序列在选自位置4-N、5-N、7-N、9-N、10-N和11-N的核苷酸处包含至少一个增强特异性的修饰; 和

形成包含Cas蛋白和所述合成的引导RNA的gRNA: Cas蛋白复合物;

使所述靶多核苷酸与所述gRNA: Cas蛋白复合物接触; 和

切割、切刻或结合所述靶多核苷酸。

02. 实施方案01的方法, 其中所述靶多核苷酸选自下组: VEGFA多核苷酸、IL2RG多核苷酸、CLTA1多核苷酸和CLTA4多核苷酸。

03. 实施方案01或02的方法, 其中所述至少一个增强特异性的修饰在核苷酸11-N处。

04. 实施方案01至03中任一项的方法, 其中所述至少一个增强特异性的修饰在核苷酸5-N处。

05. 实施方案01至04中任一项的方法, 其中所述至少一个增强特异性的修饰在核苷酸7-N处。

06. 实施方案01至05中任一项的方法, 其中所述至少一个增强特异性的修饰在核苷酸10-N处。

07. 实施方案01至06中任一项的方法, 其中所述至少一个增强特异性的修饰在核苷酸9-N处。

08. 实施方案01至07中任一项的方法, 其中所述至少一个增强特异性的修饰在核苷酸4-N处。

09. 实施方案01至08中任一项的方法, 其中所述至少一个增强特异性的修饰包括具有2'-O-甲基-3'-磷酸基乙酸 (MP)、2'-O-甲基-3'-硫代磷酸基乙酸 (MSP)、2'-脱氧-3'-磷酸基乙酸 (DP)、2'-脱氧-3'-硫代磷酸基乙酸 (DSP) 或其组合的修饰。

010. 实施方案01至08中任一项的方法, 其中所述至少一个增强特异性的修饰包括赋予C3'-内糖皱褶的2'-修饰和磷酸基乙酸或硫代磷酸基乙酸键修饰。

011. 实施方案010的方法, 其中所述2'-修饰选自2'-F和2'-O-(2-甲氧基乙基)。



012.实施方案01至011中任一项的方法,其中所述至少一个增强特异性的修饰增加所述方法的特异性。

013.实施方案01至012中任一项的方法,其中所述切割、切刻或结合在体外发生。

014.实施方案01至012中任一项的方法,其中所述切割、切刻或结合在细胞中发生。

015.实施方案01至014中任一项的方法,其中所述Cas蛋白是Cas9或Cpf1。

016.实施方案01至015中任一项的方法,其中所述引导RNA是合成的单引导RNA。

017.实施方案01至016中任一项的方法,其中所述方法包括切割所述靶HBB多核苷酸。

018.实施方案01至016中任一项的方法,其中所述方法包括切刻所述靶HBB多核苷酸。

019.实施方案01至016中任一项的方法,其中所述方法包括结合所述靶HBB多核苷酸。

020.实施方案01至019中任一项的方法,其中所述引导序列由从所述引导序列的5'端计数的核苷酸1至20组成。

P1.一种合成的引导RNA,其包含:

(a) crRNA区段,其包含(i)能够与靶多核苷酸杂交的引导序列,(ii)茎序列;和

(b) tracrRNA区段,其包含与所述茎序列部分或完全互补的核苷酸序列,其中所述引导序列由20-N个核苷酸组成,其中N是-10至10之间的整数,并且所述引导序列包含在选自位置4-N、5-N、7-N、9-N、10-N和11-N的核苷酸处的至少一个增强特异性的修饰,并且其中所述合成的引导RNA具有gRNA功能性。

P2.实施方案P1的合成的引导RNA,其中所述靶多核苷酸选自下组:VEGFA多核苷酸、IL2RG多核苷酸、CLTA1多核苷酸和CLTA4多核苷酸。

P3.实施方案P1或P2的合成的引导RNA,其中所述至少一个增强特异性的修饰在核苷酸11-N处。

P4.实施方案P1至P3中任一项的合成的引导RNA,其中所述至少一个增强特异性的修饰在核苷酸5-N处。

P5.实施方案P1至P4中任一项的合成的引导RNA,其中所述至少一个增强特异性的修饰在核苷酸7-N处。

P6.实施方案P1至P5中任一项的合成的引导RNA,其中所述至少一个增强特异性的修饰在核苷酸10-N处。

P7.实施方案P1至P6中任一项的合成的引导RNA,其中所述至少一个增强特异性的修饰在核苷酸9-N处。

P8.实施方案P1至P7中任一项的合成的引导RNA,其中所述至少一个增强特异性的修饰在核苷酸4-N处。

P9.实施方案P1至P8中任一项的合成的引导RNA,其中所述增强特异性的修饰包括2'-O-甲基-3'-磷酸基乙酸(MP)、2'-O-甲基-3'-硫代磷酸基乙酸(MSP)、2'-脱氧-3'-磷酸基乙酸(DP)、2'-脱氧-3'-硫代磷酸基乙酸(DSP)或其组合。

P10.实施方案P1至P9中任一项的合成的引导RNA,其中所述至少一个增强特异性

的修饰包括赋予C3'-内糖皱褶的2'-修饰和磷酸基乙酸或硫代磷酸基乙酸键修饰。

P11. 实施方案P10的合成的引导RNA, 其中所述2'-修饰选自2'-F和2'-O- (2-甲氧基乙基)。

P12. 实施方案P1至P11中任一项的合成的引导RNA, 其中所述至少一个增强特异性的修饰减弱所述引导序列与所述靶多核苷酸之间的杂交。

P13. 实施方案P1至P12中任一项的合成的引导RNA, 其中所述至少一个增强特异性的修饰减弱所述引导序列与脱靶多核苷酸之间的杂交。

P14. 实施方案P1至P13中任一项的合成的引导RNA, 其中所述至少一个增强特异性的修饰增强所述引导序列与所述靶多核苷酸之间的杂交, 并且减弱所述引导序列与脱靶多核苷酸之间的杂交。

P15. 实施方案P1至P14中任一项的合成的引导RNA, 其中所述引导RNA是合成的单引导RNA。

P16. 实施方案P1至P15中任一项的合成的引导RNA, 其中所述引导序列由从所述引导序列的5'端计数的核苷酸1至20组成。

Q1. 一种切割、切刻或结合靶多核苷酸的方法, 其包括:

选择基因组中的靶多核苷酸;

提供包含以下区段的合成的引导RNA:

(a) crRNA区段, 其包含(i)能够与靶多核苷酸杂交的引导序列, (ii)茎序列; 和

(b) tracrRNA区段, 其包含与所述茎序列部分或完全互补的核苷酸序列, 其中所述引导序列包含至少两个连续的增强特异性的修饰; 和

形成包含Cas蛋白和所述合成的引导RNA的gRNA: Cas蛋白复合物;

使所述靶多核苷酸与gRNA: Cas蛋白复合物接触; 和

切割、切刻或结合所述靶多核苷酸,

其中所述至少一个增强特异性的修饰减弱所述引导序列与所述靶多核苷酸之间的杂交。

Q2. 实施方案Q1的方法, 其中所述引导序列在从所述引导序列的5'端计数的核苷酸1和2处包含修饰。

Q3. 实施方案Q1的方法, 其中所述引导序列在从所述引导序列的5'端计数的核苷酸1、2和3处包含修饰。

Q4. 实施方案Q1的方法, 其中所述引导序列在从所述引导序列的5'端计数的核苷酸1、2、3和4处包含修饰。

Q5. 实施方案Q1的方法, 其中所述引导序列在从所述引导序列的5'端计数的核苷酸1、2、3、4和5处包含修饰。

Q6. 实施方案Q1的方法, 其中所述引导序列在5'端的连续的增强特异性包含修饰。

Q7. 实施方案Q1的方法, 其中所述连续的增强特异性的修饰在从所述引导序列的5'端计数的核苷酸1、2、3或4处开始。

Q8. 实施方案Q1至Q7中任一项的方法, 其中所述靶多核苷酸选自下组: HBB多核苷酸、VEGFA多核苷酸、IL2RG多核苷酸、CLTA1多核苷酸和CLTA4多核苷酸。

Q9. 实施方案Q1至Q8中任一项的方法, 其中所述增强特异性的修饰包括2'-O-甲

基-3'-磷酸基乙酸(MP)、2'-O-甲基-3'-硫代磷酸基乙酸(MSP)、2'-脱氧-3'-磷酸基乙酸(DP)、2'-脱氧-3'-硫代磷酸基乙酸(DSP)或其组合。

Q10. 实施方案Q1至Q8中任一项的方法,其中所述至少一个增强特异性的修饰包括赋予C3'-内糖皱褶的2'-修饰和磷酸基乙酸或硫代磷酸基乙酸键修饰。

Q11. 实施方案Q10的方法,其中所述2'-修饰选自2'-F和2'-O-(2-甲氧基乙基)。

Q12. 实施方案Q1至Q11中任一项的方法,其中所述切割、切刻或结合在体外发生。

Q13. 实施方案Q1至Q11中任一项的方法,其中所述切割、切刻或结合在细胞中发生。

Q14. 实施方案Q1至Q13中任一项的方法,其中所述方法包括切割所述靶HBB多核苷酸。

Q15. 实施方案Q1至Q13中任一项的方法,其中所述方法包括切刻所述靶HBB多核苷酸。

Q16. 实施方案Q1至Q13中任一项的方法,其中所述方法包括结合所述靶HBB多核苷酸。

Q17. 实施方案Q1至Q16中任一项的方法,其中所述引导RNA是合成的单引导RNA。

Q18. 实施方案Q1至Q17中任一项的方法,其中所述引导序列包含从所述引导序列的5'端计数的核苷酸1至20或由所述核苷酸1至20组成。

R1. 一种合成的引导RNA,其包含:

(a) crRNA区段,其包含(i)能够与靶多核苷酸杂交的引导序列,(ii)茎序列;和

(b) tracrRNA区段,其包含与所述茎序列部分或完全互补的核苷酸序列,其中所述引导序列包含至少两个连续的增强特异性的修饰,并且其中所述合成的引导RNA具有gRNA功能性,并且其中所述至少一个增强特异性的修饰减弱所述引导序列与所述靶多核苷酸之间的杂交。

R2. 实施方案R1的合成的引导RNA,其中所述引导序列包含在从所述引导序列的5'端计数的核苷酸1和2处的增强特异性的修饰。

R3. 实施方案R1的合成的引导RNA,其中所述引导序列包含在从所述引导序列的5'端计数的核苷酸1、2和3处的增强特异性的修饰。

R4. 实施方案R1的合成的引导RNA,其中所述引导序列包含在从所述引导序列的5'端计数的核苷酸1、2、3和4处的增强特异性的修饰。

R5. 实施方案R1的合成的引导RNA,其中所述引导序列包含在从所述引导序列的5'端计数的核苷酸1、2、3、4和5处的增强特异性的修饰。

R6. 实施方案R1的合成的引导RNA,其中所述引导序列在5'端包含从所述引导序列的5'端计数的连续的增强特异性的修饰。

R7. 实施方案R1的合成的引导RNA,其中所述连续的增强特异性的修饰在从所述引导序列的5'端计数的核苷酸1、2、3或4处开始。

R8. 实施方案R1的合成的引导RNA,其中所述引导序列包含三个连续的增强特异性的修饰。

R9. 实施方案R1的合成的引导RNA,其中所述引导序列包含四个连续的增强特异性的修饰。

R10. 实施方案R1至R9中任一项的合成引导RNA, 其中所述靶多核苷酸选自下组: HBB多核苷酸、VEGFA多核苷酸、IL2RG多核苷酸、CLTA1多核苷酸和CLTA4多核苷酸。

R11. 实施方案的R1至R10中任一项的合成的引导RNA, 其中所述增强特异性的修饰包括2'-O-甲基-3'-磷酸基乙酸 (MP)、2'-O-甲基-3'-硫代磷酸基乙酸 (MSP)、2'-脱氧-3'-磷酸基乙酸 (DP)、2'-脱氧-3'-硫代磷酸基乙酸 (DSP) 或其组合。

R12. 实施方案R1至R10中任一项的合成的引导RNA, 其中所述至少一个增强特异性的修饰选自赋予C3'-内糖皱褶的2'-修饰、磷酸基乙酸或硫代磷酸基乙酸键修饰及其组合。

R13. 实施方案R12的合成的引导RNA, 其中所述2'-修饰选自2'-F和2'-O-(2-甲氧基乙基)。

R14. 实施方案R1至R13中任一项的合成的引导RNA, 其中所述至少一个增强特异性的修饰减弱所述引导序列与所述靶多核苷酸之间的杂交。

R15. 实施方案R1至R13中任一项的合成的引导RNA, 其中所述至少一个增强特异性的修饰减弱所述引导序列与脱靶多核苷酸之间的杂交。

R16. 实施方案R1至R13中任一项的合成的引导RNA, 其中所述至少一个增强特异性的修饰增强所述引导序列与所述靶多核苷酸之间的杂交, 并且减弱所述引导序列与脱靶多核苷酸之间的杂交。

R17. 实施方案R1至R16中任一项的合成的引导RNA, 其中所述引导RNA是合成的单引导RNA。

R18. 实施方案R1至R17中任一项的方法, 其中所述引导序列包含从所述引导序列的5'端计数的核苷酸1至20或由所述核苷酸1至20组成。

S1. 一种选择合成的引导RNA的方法, 其包括:

提供至少第一合成的引导RNA和第二合成的引导RNA, 其各自包含:

(a) crRNA区段, 其包含 (i) 能够与靶多核苷酸杂交的引导序列, (ii) 茎序列; 和

(b) tracrRNA区段, 其包含与所述茎序列部分或完全互补的核苷酸序列, 其中每个引导序列由从所述引导序列的5'端计数的20-N个核苷酸组成, 其中N是-10至10之间的整数,

其中所述第一合成的引导RNA包含在所述引导序列内的第一位置处的增强特异性的修饰, 并且所述第二合成的引导RNA包含所述引导序列内的第二位置处的增强特异性的修饰;

形成包含Cas蛋白和所述第一合成的引导RNA的第一gRNA: Cas蛋白复合物, 使所述靶多核苷酸与所述第一gRNA: Cas蛋白复合物接触, 并且切割、切刻或结合所述靶多核苷酸;

形成包含Cas蛋白和所述第二合成的引导RNA的第二gRNA: Cas蛋白复合物, 使所述靶多核苷酸与所述第二gRNA: Cas蛋白复合物接触, 并且切割、切刻或结合所述靶多核苷酸;

确定所述第一gRNA: Cas蛋白复合物和所述第二gRNA: Cas蛋白复合物在所述靶多核苷酸的切割、切刻或结合中的特异性;

鉴定所述第一gRNA: Cas蛋白复合物和所述第二gRNA: Cas蛋白复合物中的哪一个对所述靶多核苷酸具有更好的特异性。

S2. 实施方案S1的方法, 其中所述增强特异性的修饰包括2'-O-甲基-3'-磷酸基乙酸 (MP)、2'-O-甲基-3'-硫代磷酸基乙酸 (MSP)、2'-脱氧-3'-磷酸基乙酸 (DP)、2'-脱氧-3'-

硫代磷酸基乙酸(DSP)或其组合。

S3.实施方案S2的方法,其中所述增强特异性的修饰包括2'-O-甲基-3'-磷酸基乙酸(MP)或2'-O-甲基-3'-硫代磷酸基乙酸(MSP)。

S4.实施方案S1的方法,其中所述至少一个增强特异性的修饰包括赋予C3'-内糖皱褶的2'-修饰和磷酸基乙酸或硫代磷酸基乙酸键修饰。

S5.实施方案S4的方法,其中所述2'-修饰选自2'-F和2'-O-(2-甲氧基乙基)。

S6.实施方案S1至S5中任一项的方法,其中所述第一和第二合成的引导RNA包含在不同核苷酸处的增强特异性的修饰。

S7.实施方案S1至S6中任一项的方法,其中所述特异性选自中靶和/或脱靶切割、结合或切割百分比、中靶:脱靶比、特异性评分或其组合。

S8.实施方案S1至S7中任一项的方法,其中所述方法包括提供在引导序列部分中的不同核苷酸位置处包含增强特异性的修饰的第一至第二合成的引导RNA,使用每个合成的引导RNA形成gRNA:Cas蛋白复合物,使靶多核苷酸与所述gRNA:Cas蛋白复合物接触,切割、切割或结合所述靶多核苷酸并测量每个合成的引导RNA的特异性,并且鉴定提供最大特异性增强的一个或多个修饰位置。

S9.实施方案S1至S8中任一项的方法,其中所述引导序列进一步包括在5'端和3'端的具有2'-O-甲基-3'-磷酸基乙酸(MP)、2'-O-甲基-3'-硫代磷酸基乙酸(MSP)、2'-O-甲基-3'-硫代磷酸(MS)、2'-脱氧-3'-磷酸基乙酸(DP)、2'-脱氧-3'-硫代磷酸基乙酸(DSP)、2'-氟-3'-磷酸基乙酸(FP)、2'-氟-3'-硫代磷酸基乙酸(FSP)、2'-氟-3'-硫代磷酸(FS)或其组合的修饰。

S10.实施方案S1至S9中任一项的方法,其中所述方法包括切割靶HBB多核苷酸。

S11.实施方案S1至S9中任一项的方法,其中所述方法包括切割靶HBB多核苷酸。

S12.实施方案S1至S9中任一项的方法,其中所述方法包括结合靶HBB多核苷酸。

S13.实施方案S1至S12中任一项的方法,其中所述第一和第二引导RNA是合成的单引导RNA。

T1.一种用于选择合成的引导RNA的试剂盒,其包含:

包含以下区段的至少两个合成的引导RNA:

(a) crRNA区段,其包含(i)能够与靶多核苷酸杂交的引导序列,(ii)茎序列;和

(b) tracrRNA区段,其包含与所述茎序列部分或完全互补的核苷酸序列,其中所述引导序列由从所述引导序列的5'端计数的20-N个核苷酸组成,其中N是-10至10之间的整数,并且在所述引导序列中的核苷酸处包含至少一个增强特异性的修饰;并且其中所述至少两个合成的引导RNA由于在所述引导序列中具有至少一个不同的增强特异性的修饰、或者由于增强特异性的修饰位于至少一个不同位置而彼此不同;和

Cas蛋白或编码所述Cas蛋白的多核苷酸。

T2.实施方案T1的试剂盒,其中所述试剂盒包含至少二十个合成的引导RNA。

T3.实施方案T1或T2的试剂盒,其中所述Cas蛋白是Cas9或Cpf1。

T4.实施方案T1至T3中任一项的试剂盒,其中所述增强特异性的修饰包括2'-O-甲基-3'-磷酸基乙酸(MP)、2'-O-甲基-3'-硫代磷酸基乙酸(MSP)、2'-脱氧-3'-磷酸基乙酸(DP)或2'-脱氧-3'-硫代磷酸基乙酸(DSP)。

T5. 实施方案T1至T3中任一项的试剂盒,其中所述增强特异性的修饰包括赋予C3'-内糖皱褶的2'-修饰和磷酸基乙酸或硫代磷酸基乙酸键修饰。

T6. 实施方案T5的试剂盒,其中所述2'-修饰选自2'-F和2'-O-(2-甲氧基乙基)。

T7. 实施方案T1至T6中任一项的试剂盒,其中所述引导RNA是合成的单引导RNA。

U1. 任何前述实施方案的合成的引导RNA、方法或试剂盒,其中所述gRNA:Cas蛋白复合物包含增强特异性的修饰并且具有大于1、优选至少1.1、更优选至少1.5、甚至更优选至少2、甚至更优选至少5、甚至更优选至少10、或最佳至少20的特异性评分。

U2. 任何前述实施方案的合成的引导RNA、方法或试剂盒,其中所述gRNA:Cas蛋白复合物包含增强特异性的修饰并且具有约2至约60或优选约10至约60的特异性评分。

U3. 任何前述实施方案的合成的引导RNA、方法或试剂盒,其中所述gRNA:Cas蛋白复合物包含增强特异性的修饰并且具有至少30%、优选至少50%、更优选至少70%、或最佳至少90%的中靶切割。

U4. 任何前述实施方案的合成的引导RNA、方法或试剂盒,其中所述gRNA:Cas蛋白复合物包含增强特异性的修饰并且具有约25%至99.9%或优选约50%至约99.9%的中靶切割。

U5. 任何前述实施方案的合成的引导RNA、方法或试剂盒,其中所述gRNA:Cas蛋白复合物包含增强特异性的修饰并且具有大于1、优选至少1.1:1、更优选至少1.5:1、甚至更优选至少3:1、甚至更优选至少10:1、甚至更优选至少20:1、或最佳至少40:1的中靶:脱靶比。

U6. 任何前述实施方案的合成的引导RNA、方法或试剂盒,其中所述gRNA:Cas蛋白复合物包含增强特异性的修饰并且具有约1.5:1至约99.9:1或优选约10:1至约99.9:1的中靶:脱靶比。

W1. 任何前述实施方案的合成的引导RNA、方法或试剂盒,其中所述引导序列由从所述引导序列的5'端计数的核苷酸1至19组成,并且在选自位置3、4、6、8、9和10的核苷酸之一处包含至少一个化学修饰。

W2. 任何前述实施方案的合成的引导RNA、方法或试剂盒,其中所述引导序列由从所述引导序列的5'端计数的核苷酸1至18组成,并且包含在选自位置2、3、5、7、8和9的核苷酸之一的至少一个化学修饰。

W3. 任何前述实施方案的合成的引导RNA、方法或试剂盒,其中所述引导序列由从所述引导序列的5'端计数的核苷酸1至17组成,并且包含在选自位置1、2、4、6、7和8的核苷酸之一的至少一个化学修饰。

W4. 任何前述实施方案的合成的引导RNA、方法或试剂盒,其中所述引导序列由从所述引导序列的5'端计数的核苷酸1至16组成,并且包含在选自1、3、5、6和7的核苷酸位置之一的至少一个化学修饰。

W5. 任何前述实施方案的合成的引导RNA、方法或试剂盒,其中所述引导序列由从所述引导序列的5'端计数的核苷酸1至15组成,并且包含在选自位置2、4、5和6的核苷酸之一的至少一个化学修饰。

W6. 任何前述实施方案的合成的引导RNA、方法或试剂盒,其中所述引导序列由从所述引导序列的5'端计数的核苷酸1至14组成,并且包含在选自位置1、3、4或5的核苷酸处



的至少一个化学修饰。

在下面的示例性实施方案中，“X实施方案”表示其编号以X开头的所有实施方案。类似地，在下面的示例性实施方案中，“Xn实施方案”表示其编号以Xn开头的所有实施方案。例如，任何权利要求中所述的“X9实施方案”表示“X9”，以及“X9a”直至“X9z”，如果存在的话。

X.一种合成的引导RNA，其包含：

(a) crRNA区段，其包含(i)能够与靶多核苷酸杂交的引导序列，其中所述靶多核苷酸包含邻近PAM位点的靶序列，(ii)茎序列；和

(b) tracrRNA区段，其包含与所述茎序列部分或完全互补的核苷酸序列，

其中所述引导序列由20-N个核苷酸组成，其中N是-10至10之间的整数；

其中所述引导序列包含至少一个修饰。

X1.一种合成的crRNA，其包含：(i)能够与靶多核苷酸杂交的引导序列，其中所述靶多核苷酸包含邻近PAM位点的靶序列，和(ii)茎序列；

其中所述引导序列由20-N个核苷酸组成，其中N是-10至10之间的整数；其中所述引导序列包含至少一个修饰。

X1a.实施方案X或X1的合成的引导RNA或crRNA，其中所述至少一个修饰包括在所述引导序列的1-N位的修饰。

X1b.前述X实施方案中任一项的合成的引导RNA或crRNA，其中所述至少一个修饰包括在所述引导序列的2-N位的修饰。

X1c.前述X实施方案中任一项的合成的引导RNA或crRNA，其中所述至少一个修饰包括在所述引导序列的3-N位的修饰。

X2.前述X实施方案中任一项的合成的引导RNA或crRNA，其中所述至少一个修饰包括在所述引导序列的4-N位的修饰。

X3.前述X实施方案中任一项的合成的引导RNA或crRNA，其中所述至少一个修饰包括在所述引导序列的5-N位的修饰。

X3a.前述X实施方案中任一项的合成的引导RNA或crRNA，其中所述至少一个修饰包括在所述引导序列的6-N位的修饰。

X4.前述X实施方案中任一项的合成的引导RNA或crRNA，其中所述至少一个修饰包括在所述引导序列的7-N位的修饰。

X4a.前述X实施方案中任一项的合成的引导RNA或crRNA，其中所述至少一个修饰包括在所述引导序列的8-N位的修饰。

X5.前述X实施方案中任一项的合成的引导RNA或crRNA，其中所述至少一个修饰包括在所述引导序列的9-N位的修饰。

X6.前述X实施方案中任一项的合成的引导RNA或crRNA，其中所述至少一个修饰包括在所述引导序列的10-N位的修饰。

X7.前述X实施方案中任一项的合成的引导RNA或crRNA，其中所述至少一个修饰包括在所述引导序列的11-N位的修饰。

X7a.前述X实施方案中任一项的合成的引导RNA或crRNA，其中所述至少一个修饰包括在所述引导序列的12-N位的修饰。

X7b. 前述X实施方案中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包括在所述引导序列的13-N位的修饰。

X7c. 前述X实施方案中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包括在所述引导序列的14-N位的修饰。

X7d. 前述X实施方案中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包括在所述引导序列的15-N位的修饰。

X7e. 前述X实施方案中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包括在所述引导序列的16-N位的修饰。

X7f. 前述X实施方案中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包括在所述引导序列的17-N位的修饰。

X7g. 前述X实施方案中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包括在所述引导序列的18-N位的修饰。

X7h. 前述X实施方案中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包括在所述引导序列的19-N位的修饰。

X7i. 前述X实施方案中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包括在所述引导序列的20-N位的修饰。

X8. 前述X实施方案中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包括磷酸基羧酸酯核苷酸间键, 任选地为磷酸基乙酸核苷酸间键(P)。

X8a. 前述X实施方案中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包括磷酸基羧酸酯核苷酸间键, 任选地为磷酸基乙酸核苷酸间键。

X8b. 前述X实施方案中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包括磷酸丙酸酯核苷酸间键。

X9. 前述X实施方案中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包括硫代磷酸基羧酸酯核苷酸间键, 任选地硫代磷酸基乙酸核苷酸间键(SP)。

X9a. 前述X实施方案中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包含硫代磷酸基丙酸酯核苷酸间键。

X9b. 前述X实施方案中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包含硫代磷酸基羧酸酯核苷酸间键, 任选地硫代磷酸基乙酸核苷酸间键。

X9c. 前述X实施方案中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包含硫代磷酸核苷酸间键。

X9d. 前述X实施方案中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包含手性硫代磷酸核苷酸间键。

X9e. 前述X实施方案中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包括二硫代磷酸核苷酸间键。

X9f. 前述X实施方案中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包括硼烷磷酸核苷酸键。

X9g. 前述X实施方案中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包括C<sub>1-4</sub>烷基磷酸核苷酸间键。

X9h. X9g实施方案的合成的引导RNA或crRNA, 其中所述至少一个修饰包括甲基磷

酸酯核苷酸间键。

X9i. 前述X实施方案中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包括甲基硫基磷酸酯核苷酸间键。

X10. 前述X实施方案中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包括赋予C3'-内糖皱褶构型的2'-修饰。

X10a. 实施方案X10的合成的引导RNA或crRNA, 其中所述至少一个修饰是2'-脱氧核糖(2'-脱氧)。

X10b. 实施方案X10的合成的引导RNA或crRNA, 其中所述至少一个修饰是2'-NH<sub>2</sub>。

X10c. 实施方案X10的合成的引导RNA或crRNA, 其中所述至少一个修饰是2'-阿拉伯呋喃糖基(2'-阿拉伯糖基)。

X10d. 实施方案X10的合成的引导RNA或crRNA, 其中所述至少一个修饰是2'-脱氧-2'-氟阿拉伯呋喃糖基(2'-F-阿拉伯糖基)。

X10e. 实施方案X10的合成的引导RNA或crRNA, 其中所述至少一个修饰是2'-LNA。

X10f. 实施方案X10的合成的引导RNA或crRNA, 其中所述至少一个修饰是2'-ULNA。

X10g. 实施方案X10的合成的引导RNA或crRNA, 其中所述至少一个修饰是4'-硫代核糖基。

X10h. 实施方案X10的合成的引导RNA或crRNA, 其中所述至少一个修饰是2'-O-C<sub>1-4</sub>烷基。

X10i. 实施方案X10的合成的引导RNA或crRNA, 其中所述至少一个修饰是2'-O-C<sub>1-3</sub>烷基-O-C<sub>1-3</sub>烷基。

X10j. 实施方案X10的合成的引导RNA或crRNA, 其中所述至少一个修饰选自2'-O-苯基, 2'-噻吩基, 2'-S-噻吩基, 2'-甲基, 2'-乙基, 2'-丙基, 2'-烯丙基, 2'-烯丙基苯基, 2'-甲基羟基, 2'-甲氧基甲基, 2'-O-氨基甲酸酯, 2'-O-乙基氨基, 2'-O-烯丙基氨基, 2'-O-丙基氨基和2'-O-取代的苯基。

X10k. 前述X实施方案中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包含赋予C3'-内糖皱褶构型的2'-修饰。

X11. 实施方案X10k的合成的引导RNA或crRNA, 其中所述2'-修饰选自2'-O-甲基、2'-氟和2'-O-(2-甲氧基乙基)。

X11a. 实施方案X11的合成的引导RNA或crRNA, 其中所述2'-修饰是2'-O-甲基。

X11b. 实施方案X11的合成的引导RNA或crRNA, 其中所述2'-修饰是2'-脱氧-2'-氟呋喃核糖基(2'-氟)。

X11c. 实施方案X11的合成的引导RNA或crRNA, 其中所述2'-修饰是2'-O-(2-甲氧基乙基)。

X12. 前述X实施方案中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包括2'-O-甲基-3'-磷酸基乙酸(MP)。

X13. 前述X实施方案中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包括2'-O-甲基-3'-硫代磷酸基乙酸(MSP)。

X14. 前述X实施方案中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包括2'-脱氧-3'-磷酸基乙酸(DP)。

X15. 前述X实施方案中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包括2'-O-脱氧-3'-硫代磷酸基乙酸(DSP)。

X16. 实施方案X或X1的合成的引导RNA或crRNA, 其中所述至少一个修饰包括在所述引导序列的位置4-N至20-N的任何位置处的磷酸基羧酸酯或硫代磷酸基羧酸酯核苷酸间键修饰, 并且其中所述修饰不在所述引导序列的位置15-N处。

X16a. 实施方案X或X1的合成的引导RNA或crRNA, 其中所述至少一个修饰包括核苷酸间键修饰, 并且其中15-N位不包含2'-O-甲基-3'-磷酸基乙酸(MP)或2'-O-甲基-3'-硫代磷酸基乙酸(MSP)。

X17. 实施方案X、X1、X16-X16a中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包括在所述引导序列的位置4-N处的磷酸基羧酸酯或硫代磷酸基羧酸酯核苷酸间键修饰。

X18. 实施方案X、X1、X16至X17中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包括在所述引导序列的位置5-N处的磷酸基羧酸酯或硫代磷酸基羧酸酯核苷酸间键修饰。

X19. 实施方案X、X1和X16-X18中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包括在所述引导序列的位置6-N处的磷酸基羧酸酯或硫代磷酸基羧酸酯核苷酸间键修饰。

X20. 实施方案X、X1和X16-X19中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包括在所述引导序列的位置7-N处的磷酸基羧酸酯或硫代磷酸基羧酸酯核苷酸间键修饰。

X21. 实施方案X、X1和X16-X20中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包括在所述引导序列的位置8-N处的磷酸基羧酸酯或硫代磷酸基羧酸酯核苷酸间键修饰。

X22. 实施方案X、X1和X16-X21中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包括在所述引导序列的位置9-N处的磷酸基羧酸酯或硫代磷酸基羧酸酯核苷酸间键修饰。

X23. 实施方案X、X1和X16-X22中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包括在所述引导序列的位置10-N处的磷酸基羧酸酯或硫代磷酸基羧酸酯核苷酸间键修饰。

X24. 实施方案X、X1和X16-X23中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包括在所述引导序列的位置11-N处的磷酸基羧酸酯或硫代磷酸基羧酸酯核苷酸间键修饰。

X25. 实施方案X、X1和X16-X24中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包括在所述引导序列的位置12-N处的磷酸基羧酸酯或硫代磷酸基羧酸酯核苷酸间键修饰。

X26. 实施方案X、X1和X16-X25中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包括在所述引导序列的位置13-N处的磷酸基羧酸酯或硫代磷酸基羧酸酯核苷酸间键修饰。

X27. 实施方案X、X1和X16-X26中任一项的合成的引导RNA或crRNA, 其中所述至少

一个修饰包括在所述引导序列的位置14-N处的磷酰基羧酸酯或硫代磷酰基羧酸酯核苷酸间键修饰。

X27a. 实施方案X、X1和X16a-X26中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包括在所述引导序列的位置15-N处的磷酰基羧酸酯或硫代磷酰基羧酸酯核苷酸间键修饰。

X28. 实施方案X、X1和X16-X27中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包括在所述引导序列的位置16-N处的磷酰基羧酸酯或硫代磷酰基羧酸酯核苷酸间键修饰。

X29. 实施方案X、X1和X16-X28中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包括在所述引导序列的位置17-N处的磷酰基羧酸酯或硫代磷酰基羧酸酯核苷酸间键修饰。

X30. 实施方案X、X1和X16-X29中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包括在所述引导序列的位置18-N处的磷酰基羧酸酯或硫代磷酰基羧酸酯核苷酸间键修饰。

X31. 实施方案X、X1和X16-X30中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包括在所述引导序列的位置19-N处的磷酰基羧酸酯或硫代磷酰基羧酸酯核苷酸间键修饰。

X32. 实施方案X、X1和X16-X31中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包括在所述引导序列的位置20-N处的磷酰基羧酸酯或硫代磷酰基羧酸酯核苷酸间键修饰。

X33. 实施方案X、X1和X16-X32中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包括磷酰基羧酸酯核苷酸间键。

X34. 实施方案X、X1和X16-X32中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包括硫代磷酰基羧酸酯核苷酸间键。

X35. 实施方案X、X1和X16-X33中任一项的合成的引导RNA或crRNA, 其中所述磷酰基羧酸酯核苷酸间键是磷酰基乙酸键(P)。

X36. 实施方案X、X1和X16-X33中任一项的合成的引导RNA或crRNA, 其中所述磷酰基羧酸酯核苷酸间键是硫代磷酰基乙酸键(SP)。

X37. 前述X实施方案中任一项的合成的引导RNA或crRNA, 进一步包含在所述引导RNA的5'端或/和3'端处的至少一个修饰。

X38. 实施方案X37的合成的引导RNA或crRNA, 其中在5'端或/和3'端处的所述至少一个修饰独立地选自2'-O-甲基(M)、硫代磷酸核苷酸间键(S)、磷酰基乙酸核苷酸间键(P)、硫代磷酰基乙酸核苷酸间键(SP)、2'-O-甲基-3'-硫代磷酸(MS)、2'-O-甲基-3'-磷酰基乙酸(MP)和2'-O-甲基-3'-硫代磷酰基乙酸(MSP)或其组合。

X38a. 实施方案X37的合成的引导RNA或crRNA, 其中在5'端或/和3'端处的所述至少一个修饰独立地选自2'-脱氧-3'-磷酰基乙酸(DP)、2'-O-脱氧-3'-硫代磷酰基乙酸(DSP)、或其组合。

X39. 实施方案X-X38a中任一项的合成的引导RNA或crRNA, 其中所述靶多核苷酸位于HBB多核苷酸内。

X40. 实施方案X-X38a中任一项的合成的引导RNA或crRNA, 其中所述靶多核苷酸位于IL2RG多核苷酸内。

X41. 实施方案X-X38a中任一项的合成的引导RNA或crRNA, 其中所述靶多核苷酸位于VEGFA多核苷酸内。

X42. 实施方案X-X38a中任一项的合成的引导RNA或crRNA, 其中所述靶多核苷酸位于CLTA1多核苷酸内。

X43. 实施方案X-X38a中任一项的合成的引导RNA或crRNA, 其中所述靶多核苷酸位于CLTA4多核苷酸内。

X44. 实施方案X39的合成的引导RNA或crRNA, 其中所述靶多核苷酸包含GCCCCACAGGGCAGTAA。

X45. 实施方案X40的合成的引导RNA或crRNA, 其中所述靶多核苷酸包含TAATGATGGCTTCAACA。

X46. 实施方案X41的合成的引导RNA或crRNA, 其中所述靶多核苷酸包含GAGTGAGTGTGTGCGTG。

X47. 实施方案X42的合成的引导RNA或crRNA, 其中所述靶多核苷酸包含CCTCATCTCCCTCAAGC。

X48. 实施方案X43的合成的引导RNA或crRNA, 其中所述靶多核苷酸包含GATGTAGTGTTCACACA。

X48a. 前述X实施方案中任一项的合成的引导RNA或crRNA, 其中所述至少一个修饰包括修饰的碱基。

X48b. 实施方案X48a的合成的引导RNA或crRNA, 其中修饰的碱基选自2-硫代U、2-硫代C、4-硫代U、6-硫代G、2-氨基A、2-氨基嘌呤、假尿嘧啶、次黄嘌呤、7-脱氮鸟嘌呤、7-脱氮-8-氮杂鸟嘌呤、7-脱氮腺嘌呤、7-脱氮-8-氮杂腺嘌呤、5-甲基C、5-甲基U、5-羟甲基胞嘧啶、5-羟甲基尿嘧啶、5,6-脱氢尿嘧啶、5-丙炔基胞嘧啶、5-丙炔基尿嘧啶、5-乙炔基胞嘧啶、5-乙炔基尿嘧啶、5-烯丙基U、5-烯丙基C、5-氨基烯丙基U、5-氨基烯丙基C、脱碱基核苷酸、UNA碱基, isoC, isoG, 5-甲基-嘧啶, x (A, G, C, T, U), y (A, G, C, T, U) 及其组合。

X48c. 实施方案X48a的合成的引导RNA或crRNA, 其中修饰的碱基选自2-硫代U、2-硫代C、4-硫代U和6-硫代G。

X48d. 实施方案X48a的合成的引导RNA或crRNA, 其中修饰的碱基选自2-氨基A和2-氨基嘌呤。

X48e. 实施方案X48a的合成的引导RNA或crRNA, 其中修饰的碱基选自假尿嘧啶和次黄嘌呤。

X48f. 实施方案X48a的合成的引导RNA或crRNA, 其中修饰的碱基是肌苷。

X48g. 实施方案X48a的合成的引导RNA或crRNA, 其中修饰的碱基选自7-脱氮鸟嘌呤, 7-脱氮-8-氮杂鸟嘌呤, 7-脱氮腺嘌呤和7-脱氮-8-氮杂腺嘌呤。

X48h. 实施方案X48a的合成的引导RNA或crRNA, 其中修饰的碱基选自5-甲基-胞嘧啶, 5-甲基-尿嘧啶和5-甲基-嘧啶。

X48i. 实施方案X48a的合成的引导RNA或crRNA, 其中修饰的碱基选自5-羟甲基胞嘧啶, 5-羟基甲基尿嘧啶, 5,6-脱氢尿嘧啶, 5-丙炔基胞嘧啶, 5-丙炔基尿嘧啶, 5-乙炔基胞



嘧啶和5-乙炔基尿嘧啶。

X48j. 实施方案X48a的合成的引导RNA或crRNA, 其中修饰的碱基选自5-烯丙基尿嘧啶, 5-烯丙基胞嘧啶, 5-氨基-尿嘧啶和5-氨基烯丙基胞嘧啶。

X48k. 实施方案X48a的合成的引导RNA或crRNA, 其中修饰的碱基是脱碱基核苷酸。

X48l. 实施方案X48a的合成的引导RNA或crRNA, 其中修饰的碱基选自isoC和isoG。

X48m. 实施方案X48a的合成的引导RNA或crRNA, 其中修饰的碱基是UNA碱基。

X49. 一种增强CRISPR功能的特异性的方法, 其包括:

选择靶多核苷酸;

提供至少一个如前述X实施方案中任一项所述的合成的引导RNA或crRNA; 形成包含Cas蛋白和所述合成的引导RNA的gRNA: Cas蛋白复合物; 和

使所述靶多核苷酸与所述gRNA: Cas蛋白复合物接触。

X49a. 一种增强CRISPR功能的特异性的方法, 包括:

选择靶多核苷酸;

提供至少一个如前述X实施方案中任一项所述的合成crRNA;

形成包含Cas蛋白和合成crRNA的crRNA: Cas蛋白复合物; 和

使靶多核苷酸与crRNA: Cas蛋白复合物接触。

X49b. 一种分析CRISPR功能特异性的方法, 该方法包括:

提供样品, 其包含 (1) 包含靶序列的靶多核苷酸; 和 (2) 一个或多个包含脱靶序列的脱靶多核苷酸;

为靶序列提供至少一个如前述任何X实施方案所述的合成的引导RNA或crRNA;

形成包含Cas蛋白和所述合成的引导RNA或crRNA的gRNA: Cas蛋白复合物, 其中所述Cas蛋白作为蛋白质或编码所述Cas蛋白的多核苷酸提供;

在导致所述gRNA或crRNA与所述靶多核苷酸杂交且发挥CRISPR功能的条件下, 使样品与所述gRNA: Cas蛋白复合物接触;

添加寡核苷酸诱饵文库, 所述诱饵包括设计用于与所述靶多核苷酸杂交的诱饵和设计用于与所述一个或多个脱靶多核苷酸杂交的诱饵, 以捕获靶多核苷酸和所述一个或多个脱靶多核苷酸;

分离并分析该捕获的靶多核苷酸和该捕获的脱靶多核苷酸以鉴定由CRISPR功能引起的变化, 并确定靶多核苷酸和脱靶多核苷酸的相对变化, 从而评估所述CRISPR功能的特异性。

X49c. 实施方案X49b的方法, 其中所述样品包含两个或更多个靶多核苷酸, 每个靶多核苷酸包含靶序列。

X49d. 实施方案X49b的方法, 其中所述指导RNA是任何前述X实施方案的合成crRNA, 并且所述gRNA: Cas蛋白复合物是crRNA: Cpf1复合物。

X50. 任何X49实施方案的方法, 其中所述接触在细胞中进行。

X51. 实施方案X50的方法, 其中所述细胞是原代细胞。

X52. 实施方案X49-51中任一项的方法, 其中将Cas蛋白作为蛋白质或表达Cas蛋白的多核苷酸提供。

X52a. 实施方案X49-52中任一项的方法, 其中Cas蛋白是Cas9。

X52b. 实施方案X49-52中任一项的方法, 其中Cas蛋白是Cpf1。

X53. 实施方案X52的方法, 其中所述多核苷酸是mRNA。

X54. 前述X实施方案中任一项的方法, 其中所述形成在细胞外进行。

X55. 前述X实施方案中任一项的方法, 其中所述CRISPR功能是基因编辑。

X55a. 前述X实施方案中任一项的方法, 其中CRISPR功能是基因编辑, 并且进一步包括提供单链或双链多核苷酸作为供体模板用于同源定向修复。

X56. 前述X实施方案中任一项的方法, 其中所述CRISPR功能是CRISPRa。

X57. 前述X实施方案中任一项的方法, 其中所述CRISPR功能是CRISPRi。

X57a. 实施方案X55至X57中任一项的方法, 其包括分析捕获的多核苷酸, 其中所述分析通过测序进行。

X57b. 实施方案X56或X57的方法, 其进一步包括通过测量基因表达水平, 任选地通过qRT-PCR测量, 来分析所述CRISPR功能。

X57c. 实施方案X57b的方法, 其中所述测量基因表达水平在微阵列上进行。

X58. 前述X实施方案中任一项的方法, 其中CRISPR功能是基因表达的调控。

X59. 实施方案X-X58中任一项的合成的引导RNA或crRNA或方法, 其中N等于0, 并且所述引导序列由20个核苷酸组成。

X60. 实施方案X-X58中任一项的合成的引导RNA或crRNA或方法, 其中N等于1, 并且所述引导序列由19个核苷酸组成。

X61. 实施方案X-X58中任一项的合成的引导RNA或crRNA或方法, 其中N等于2, 并且所述引导序列由18个核苷酸组成。

X62. 实施方案X-X58中任一项的合成的引导RNA或crRNA或方法, 其中N等于3, 并且所述引导序列由17个核苷酸组成。

X63. 实施方案X-X58中任一项的合成的引导RNA或crRNA或方法, 其中N等于4, 并且所述引导序列由16个核苷酸组成。

X64. 实施方案X-X58中任一项的合成的引导RNA或crRNA或方法, 其中N等于5, 并且所述引导序列由15个核苷酸组成。

X65. 实施方案X-X58中任一项的合成的引导RNA或crRNA或方法, 其中N等于-1, 并且所述引导序列由21个核苷酸组成。

X66. 实施方案X-X58中任一项的合成的引导RNA或crRNA或方法, 其中N等于-2, 并且所述引导序列由22个核苷酸组成。

X67. 实施方案X-X58中任一项的合成的引导RNA或crRNA或方法, 其中N等于-3, 并且所述引导序列由23个核苷酸组成。

X68. 实施方案X-X58中任一项的合成的引导RNA或crRNA或方法, 其中N等于-4, 并且所述引导序列由24个核苷酸组成。

X69. 实施方案X-X58中任一项的合成的引导RNA或crRNA或方法, 其中N等于-5, 并且所述引导序列由25个核苷酸组成。

X70. 前述X实施方案中任一项的合成的引导RNA或crRNA或方法, 其中所述引导序列中不超过10个核苷酸包含修饰。

X71. 前述X实施方案中任一项的合成的引导RNA或crRNA或方法, 其中所述引导序

列中不超过9个核苷酸包含修饰。

X72. 前述X实施方案中任一项的合成的引导RNA或crRNA或方法, 其中所述引导序列中不超过8个核苷酸包含修饰。

X73. 前述X实施方案中任一项的合成的引导RNA或crRNA或方法, 其中所述引导序列中不超过7个核苷酸包含修饰。

X74. 前述X实施方案中任一项的合成的引导RNA或crRNA或方法, 其中所述引导序列中不超过6个核苷酸包含修饰。

X75. 前述X实施方案中任一项的合成的引导RNA或crRNA或方法, 其中所述引导序列中不超过5个核苷酸包含修饰。

X76. 前述X实施方案中任一项的合成的引导RNA或crRNA或方法, 其中所述引导序列中不超过4个核苷酸包含修饰。

X77. 前述X实施方案中任一项的合成的引导RNA或crRNA或方法, 其中所述引导序列中不超过3个核苷酸包含修饰。

X78. 前述X实施方案中任一项的合成的引导RNA或crRNA或方法, 其中所述引导序列中不超过2个核苷酸包含修饰。

X79. 前述X实施方案中任一项的合成的引导RNA或crRNA或方法, 其中, 如果所述引导序列的前4个核苷酸全部被修饰, 则它们不是以相同的方式被修饰。

X80. 前述X实施方案中任一项的合成的引导RNA或crRNA或方法, 其中, 如果所述引导序列的前5个核苷酸全部被修饰, 则它们不是以相同的方式被修饰。

X81. 前述X实施方案中任一项的合成的引导RNA或方法, 其中所述引导RNA是单引导RNA。

X82. 前述X实施方案中任一项的合成的引导RNA或crRNA或方法, 其中所述引导RNA比没有修饰的相应引导RNA对靶多核苷酸具有更高的特异性或具有更高的gRNA功能性。

X83. 前述X实施方案中任一项的方法, 其中将至少两个不同的合成的引导RNA提供给两个不同的靶多核苷酸。

X84. 实施方案X83的方法, 其中所述两个不同的靶多核苷酸位于同一个基因中。

X85. 前述X实施方案中任一项的合成的引导RNA或crRNA或方法, 其中N在-10至6之间, 并且所述引导序列由14至30个核苷酸组成。

X86. 前述X实施方案中任一项的合成的引导RNA或crRNA或方法, 其中N在-5至6之间, 并且所述引导序列由14至25个核苷酸组成。

X87. 前述X实施方案中任一项的合成的引导RNA或crRNA或方法, 其中N在-4至3之间, 并且所述引导序列由17至24个核苷酸组成。

X88. 前述X实施方案中任一项的合成的引导RNA或crRNA或方法, 其中所述至少一个修饰包括2'-脱氧-3'-磷酸基乙酸(DP)。

X89. 前述X实施方案中任一项的合成的引导RNA或crRNA或方法, 其中所述至少一个修饰包括2'-O-脱氧-3'-硫代磷酸基乙酸(DSP)。

X90. 前述X实施方案中任一项的合成的引导RNA或crRNA或方法, 包括在5'端或/和3'端处的至少一个修饰, 所述修饰包括2'-脱氧-3'-磷酸基乙酸(DP)。

X91. 前述X实施方案中任一项的合成的引导RNA或crRNA或方法, 包括在5'端或/和

3'端的至少一个修饰,所述修饰包括2'-O-脱氧-3'-硫代磷酸基乙酸(DSP)。

Y1.一种合成的引导RNA,其包含:

(a) crRNA区段,其包含(i)能够与靶多核苷酸杂交的引导序列,其中所述靶多核苷酸包含邻近PAM位点的靶序列,(ii)茎序列;和

(b) tracrRNA区段,其包含与所述茎序列部分或完全互补的核苷酸序列,

其中所述引导序列由20-N个核苷酸组成,其中N是-10至10之间的整数,任选地在-10至6之间;其中所述引导序列进一步包含位于所述引导序列的位置4-N、5-N、7-N、9-N、10-N或11-N处或其组合处的至少一个修饰。

Y2.实施方案Y1的合成的引导RNA,其中所述引导RNA是单引导RNA。

Y3.实施方案Y1的合成的引导RNA,其中所述至少一个修饰选自磷酸基乙酸核苷酸间键(P)、硫代磷酸基乙酸核苷酸间键(SP)和赋予C3'-内糖皱褶构型的2'-修饰,或其组合。

Y4.实施方案Y3的合成的引导RNA,其中所述2'-修饰选自2'-O-甲基、2'-氟和2'-O-(2-甲氧基乙基)。

Y5.实施方案Y1的合成的引导RNA,其中所述至少一个修饰选自2'-O-甲基-3'-磷酸基乙酸(MP)和2'-O-甲基-3'-硫代磷酸基乙酸(MSP)。

Y6.实施方案Y5的合成的引导RNA,其中所述至少一个修饰位于所述引导序列的位置5-N或11-N处或其组合处。

Y7.实施方案Y1的合成的引导RNA,进一步包含在所述引导RNA的5'端或/和3'端处的至少一个修饰。

Y8.实施方案Y7的合成的引导RNA,其中在5'端或/和3'端处的所述至少一个修饰独立地选自2'-O-甲基(M)、硫代磷酸核苷酸间键(S)、磷酸基乙酸核苷酸间键(P)、硫代磷酸基乙酸核苷酸间键(SP)、2'-O-甲基-3'-硫代磷酸(MS)、2'-O-甲基-3'-磷酸基乙酸(MP)和2'-O-甲基-3'-硫代磷酸基乙酸(MSP)或其组合。

Y9.实施方案Y5的合成的引导RNA,进一步包括在5'端或/和3'端处的至少一个修饰,所述至少一个修饰独立地选自2'-O-甲基(M)、硫代磷酸核苷酸间键(S)、磷酸基乙酸核苷酸间键(P)、硫代磷酸基乙酸核苷酸间键(SP)、2'-O-甲基-3'-硫代磷酸(MS)、2'-O-甲基-3'-磷酸基乙酸(MP)和2'-O-甲基-3'-硫代磷酸基乙酸(MSP)或其组合。

Y10.实施方案Y5的合成的引导RNA,其中所述靶多核苷酸位于HBB基因、IL2RG基因或VEGFA基因内。

Y11.实施方案Y10的合成的引导RNA,其中所述靶多核苷酸包含HBB基因的GCCCCACAGGGCAGTAA、IL2RG基因的TAATGATGGCTTCAACA或VEGFA基因的GAGTGAGTGTGTGCGTG。

Y12.实施方案Y10的合成的引导RNA,其中所述引导RNA是单引导RNA,并且其中在所述引导RNA的5'端或/和3'端处的所述至少一个修饰独立地选自2'-O-甲基(M)、硫代磷酸核苷酸间键(S)、磷酸基乙酸核苷酸间键(P)、硫代磷酸基乙酸核苷酸间键(SP)、2'-O-甲基-3'-硫代磷酸(MS)、2'-O-甲基-3'-磷酸基乙酸(MP)和2'-O-甲基-3'-硫代磷酸基乙酸(MSP)或其组合。

Y13.一种合成的引导RNA,其包含:

(a) crRNA区段,其包含(i)能够与靶多核苷酸杂交的引导序列,其中所述靶多核苷酸包含邻近PAM序列的靶序列,(ii)茎序列;和

(b) tracrRNA区段,其包含与所述茎序列部分或完全互补的核苷酸序列,

其中所述引导序列由20-N个核苷酸组成,其中N是-10至10之间的整数,任选地在-10至6之间;

其中所述引导序列进一步包含在所述引导序列的位置4-N至20-N的任何位置处的至少一个磷酸基羧酸酯或硫代磷酸基羧酸酯核苷酸间键修饰,并且其中所述修饰不在所述引导序列的位置15-N处。

Y14.实施方案Y13的合成的引导RNA,其中所述磷酸基羧酸酯核苷酸间键选自磷酸基乙酸键(P)和硫代磷酸基乙酸键(SP)。

Y15.实施方案Y13的合成的引导RNA,进一步包含在所述引导RNA的5'端或/和3'端处的至少一个修饰。

Y16.实施方案Y13的合成的引导RNA,其中所述引导RNA是单引导RNA。

Y17.实施方案Y15的合成的引导RNA,其中在5'端或/和3'端处的所述至少一个修饰独立地选自2'-O-甲基(M)、硫代磷酸核苷酸间键(S)、磷酸基乙酸核苷酸间键(P)、硫代磷酸基乙酸核苷酸间键(SP)、2'-O-甲基-3'-硫代磷酸(MS)、2'-O-甲基-3'-磷酸基乙酸(MP)和2'-O-甲基-3'-硫代磷酸基乙酸(MSP)或其组合。

Y18.一种合成的crRNA,其包含能够与靶多核苷酸杂交的引导序列,所述靶多核苷酸包含邻近PAM位点的靶序列,其中所述引导序列由20-N个核苷酸组成,其中N是-10至10之间的整数,任选地在-10至6之间;其中所述引导序列进一步包含在所述引导序列的位置4-N至20-N的任何位置处的至少一个磷酸基羧酸酯或硫代磷酸基羧酸酯核苷酸间键修饰,并且其中所述修饰不在所述引导序列的位置15-N处。

Y19.实施方案Y18的合成crRNA,进一步包含在所述引导crRNA的5'端或/和3'端处的至少一个修饰。

Y20.一种增强CRISPR功能的特异性的方法,其包括:

选择靶多核苷酸;

提供至少一个如实施方案Y13所述的合成的引导RNA;

形成包含Cas蛋白和所述合成的引导RNA的gRNA:Cas蛋白复合物;和

使所述靶多核苷酸与所述gRNA:Cas蛋白复合物接触;

其中将所述Cas蛋白作为蛋白质或编码所述Cas蛋白的多核苷酸提供。

Y21.实施方案Y20的方法,其中所述引导RNA是单引导RNA。

Y22.实施方案Y20的方法,其中所述引导RNA进一步包含在所述引导RNA的5'端或/和3'端处的至少一个修饰。

Y23.实施方案Y22的方法,其中在5'端或/和3'端处的所述至少一个修饰独立地选自2'-O-甲基(M)、硫代磷酸核苷酸间键(S)、磷酸基乙酸核苷酸间键(P)、硫代磷酸基乙酸核苷酸间键(SP)、2'-O-甲基-3'-硫代磷酸(MS)、2'-O-甲基-3'-磷酸基乙酸(MP)和2'-O-甲基-3'-硫代磷酸基乙酸(MSP),或其组合。

Y24.实施方案Y20的方法,其中所述引导RNA包含选自磷酸基乙酸核苷酸间键和硫代磷酸基乙酸核苷酸间键或其组合的至少一个修饰。

Y25.实施方案Y20的方法,其中所述引导RNA包含选自2'-O-甲基-3'-磷酸基乙酸(MP)和2'-O-甲基-3'-硫代磷酸基乙酸(MSP)的至少一个修饰。

Y26.实施方案Y20的方法,其中所述多核苷酸靶与所述gRNA:Cas蛋白复合物的所述接触在细胞中进行。

Y27.实施方案Y26的方法,其中所述Cas蛋白是Cas9蛋白。

Y28.实施方案Y27的方法,其中所述靶多核苷酸位于HBB基因、IL2RG基因或VEGFA基因内。

Y29.实施方案Y28的方法,其中所述靶多核苷酸包含HBB基因的GCCCCACAGGGCAGTAA、IL2RG基因的TAATGATGGCTTCAACA或VEGFA基因的GAGTGAGTGTGTGCGTG。

Y30.实施方案Y20的方法,其中所述形成在细胞外进行。

前述示例性或优选实施方案的说明应被认为是展示而不是限制由权利要求书所限定的本发明。正如将易于理解的,可以利用以上陈述的特征的众多变化形式和组合,而不脱离如权利要求书中所陈述的本发明。这样的变化形式不应被视为脱离本发明的范围,并且所有这样的变化形式预期都被包含在以下权利要求的范围内。将所有在本文中引用的这些参考文献通过提述完整并入本文。



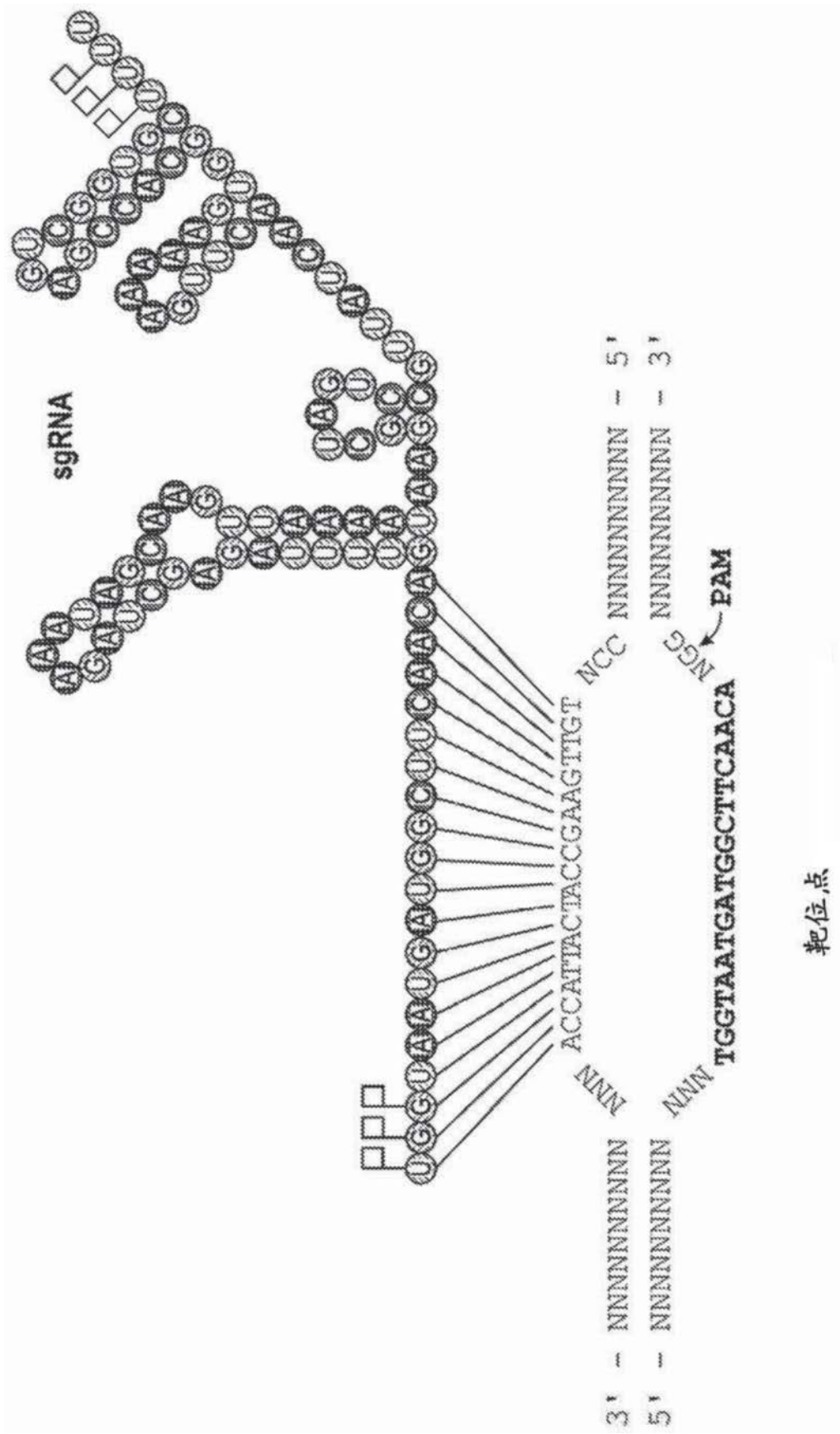


图1



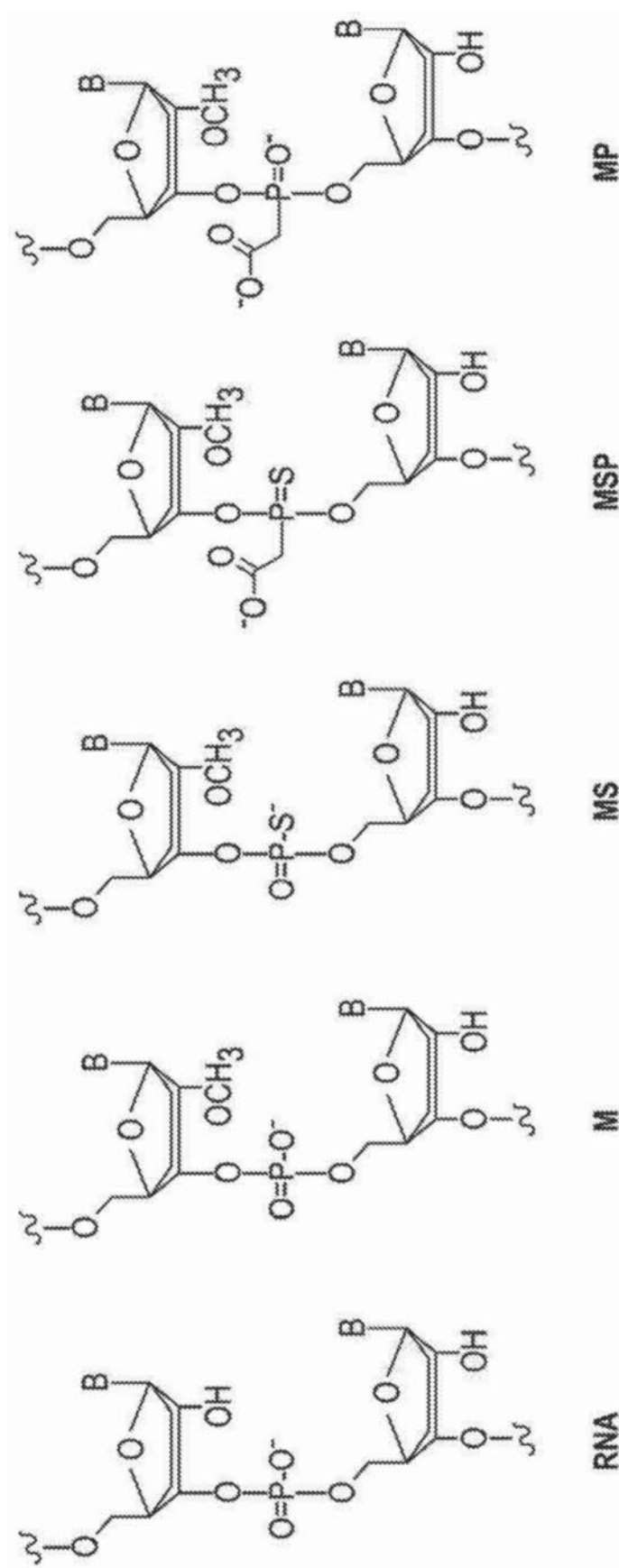


图3

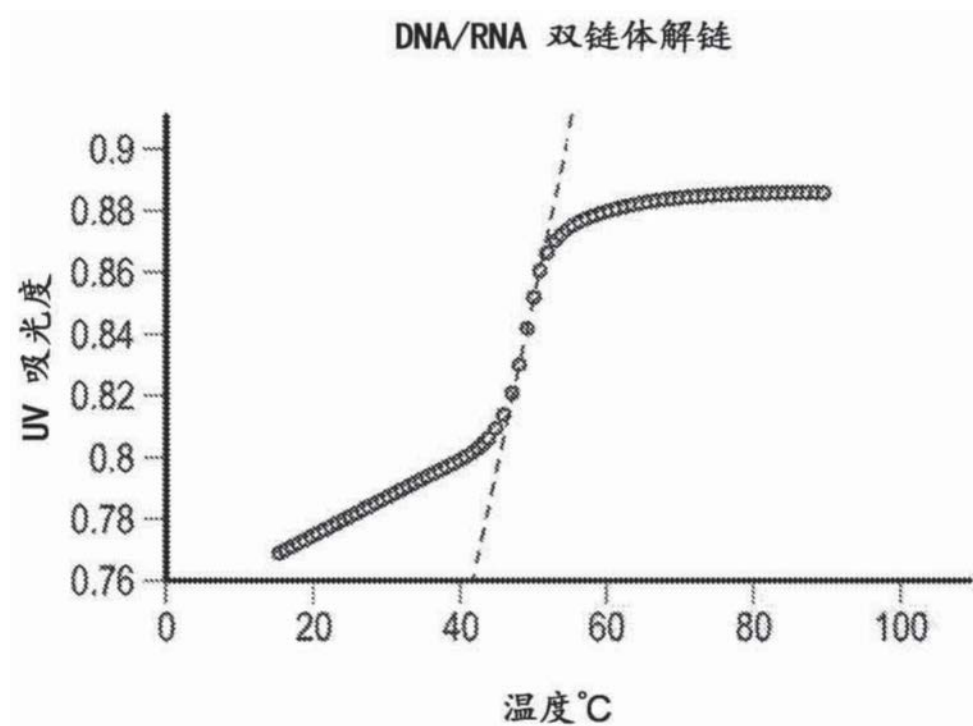


图4

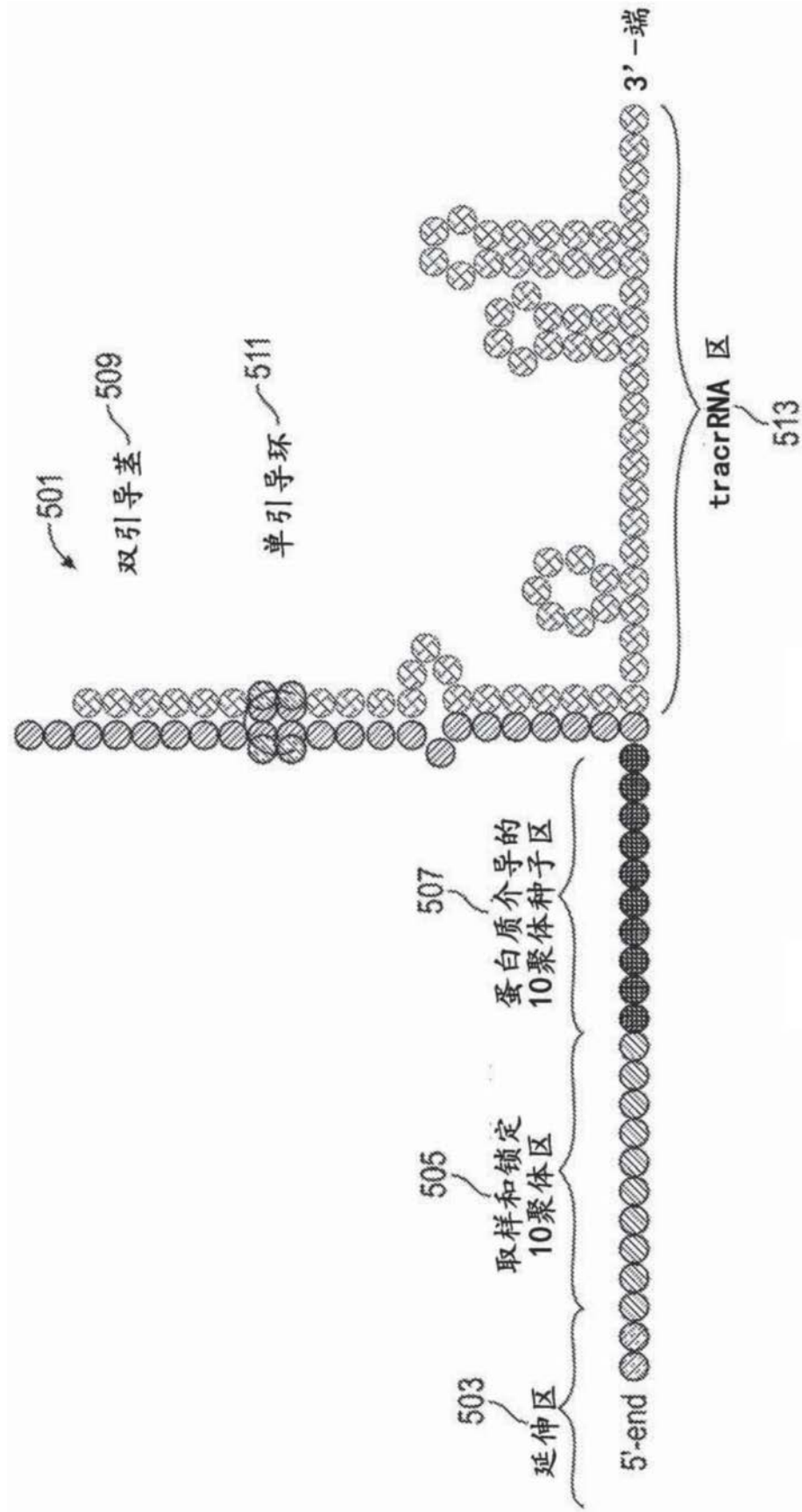


图5A

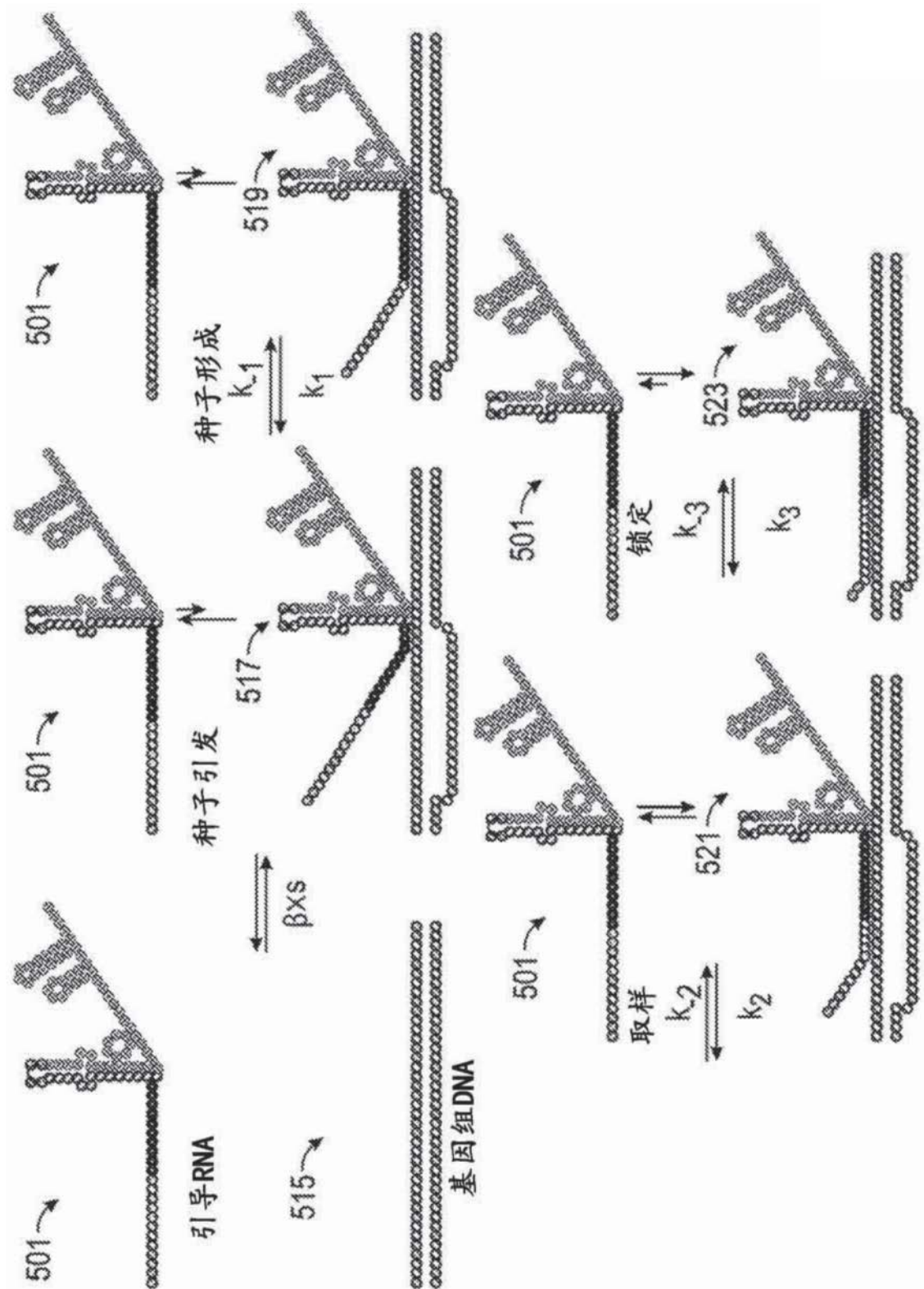


图5B



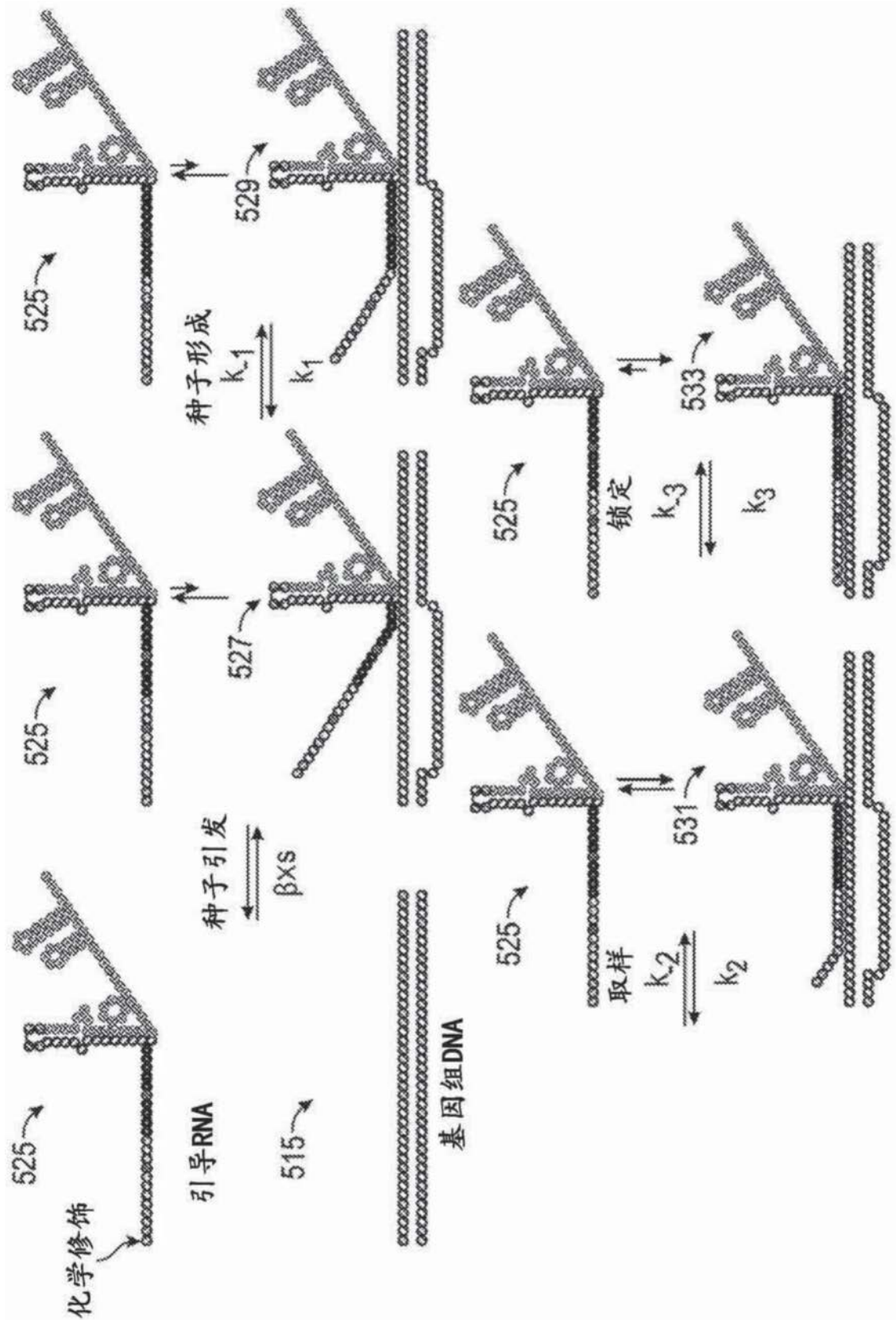


图5C

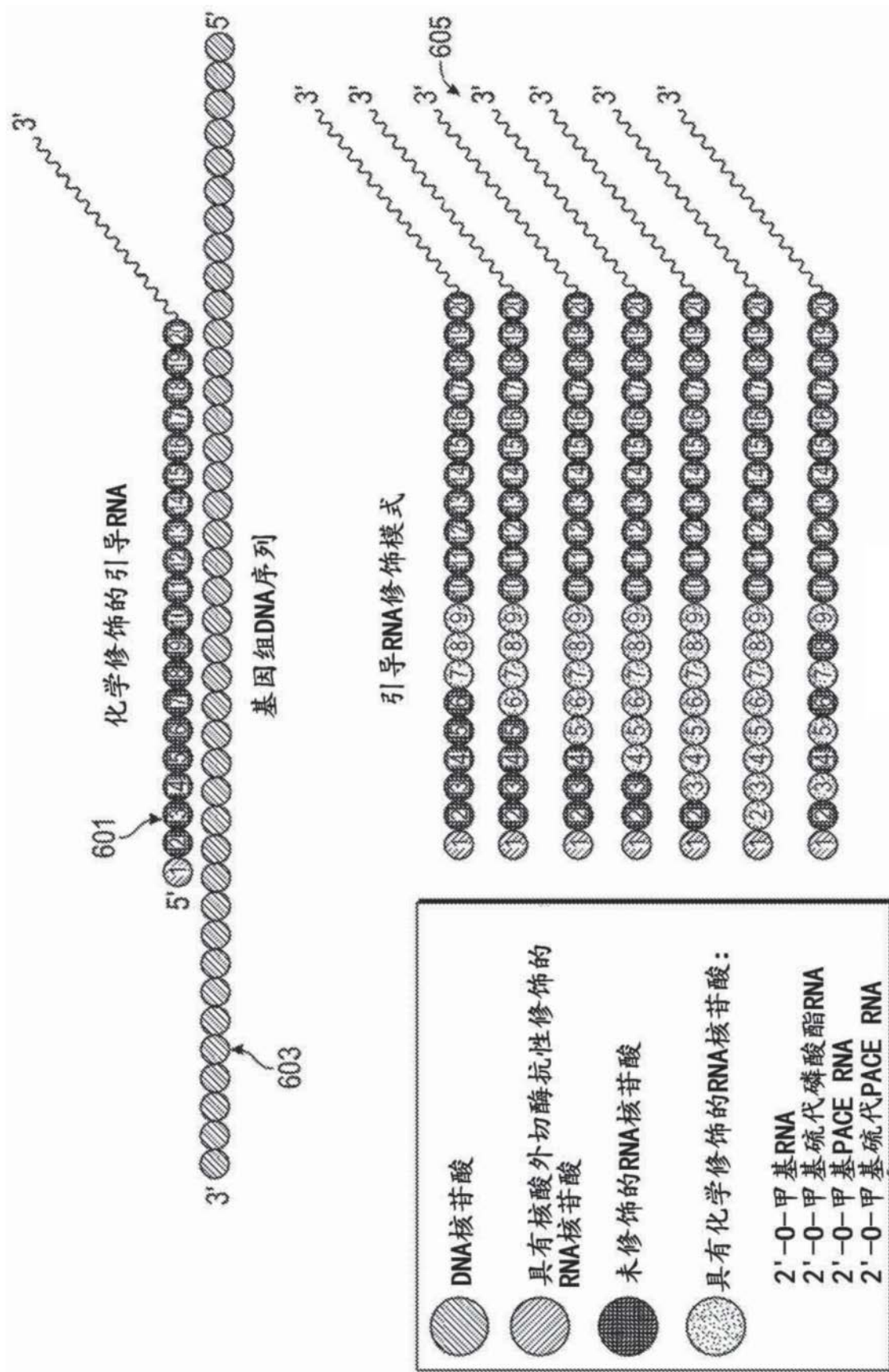


图6A

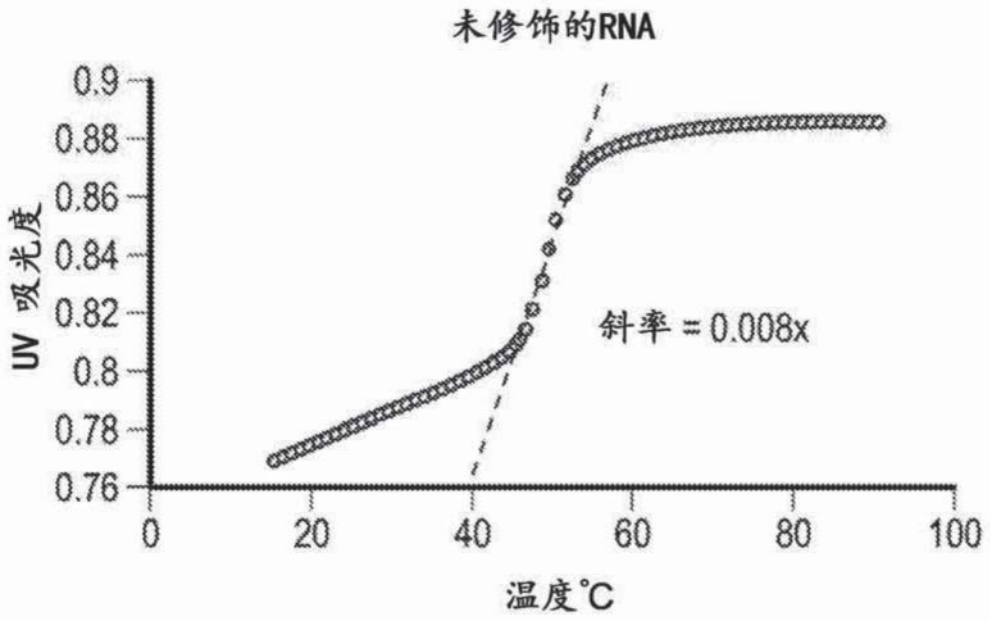


图6B

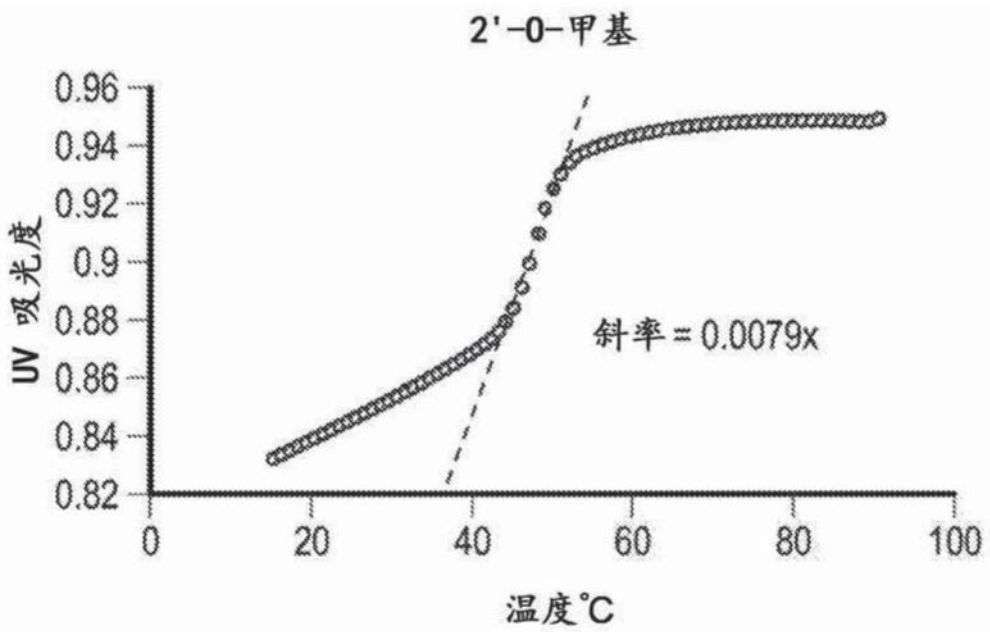


图6C

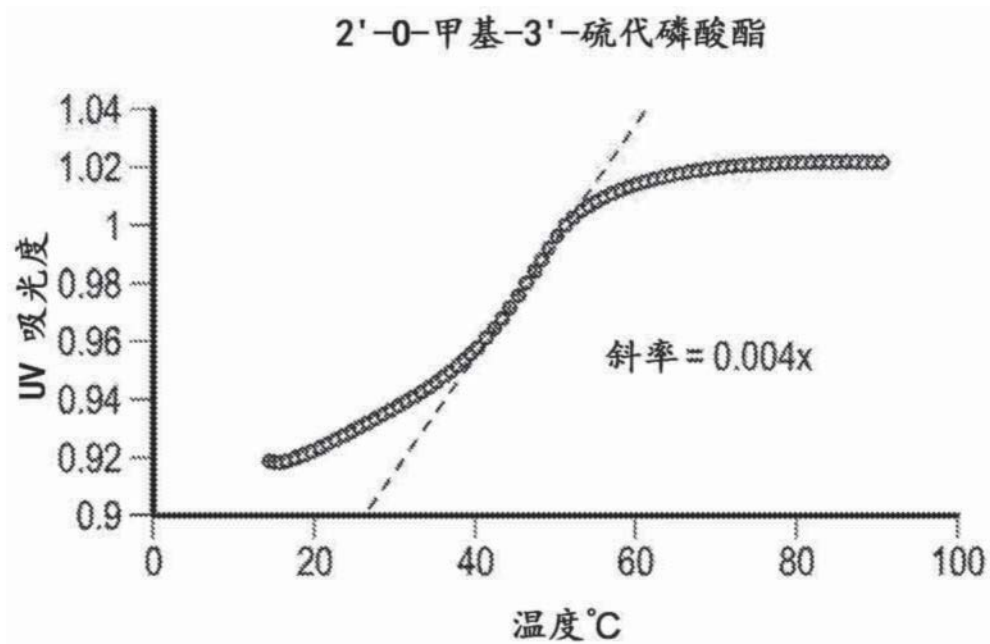


图6D

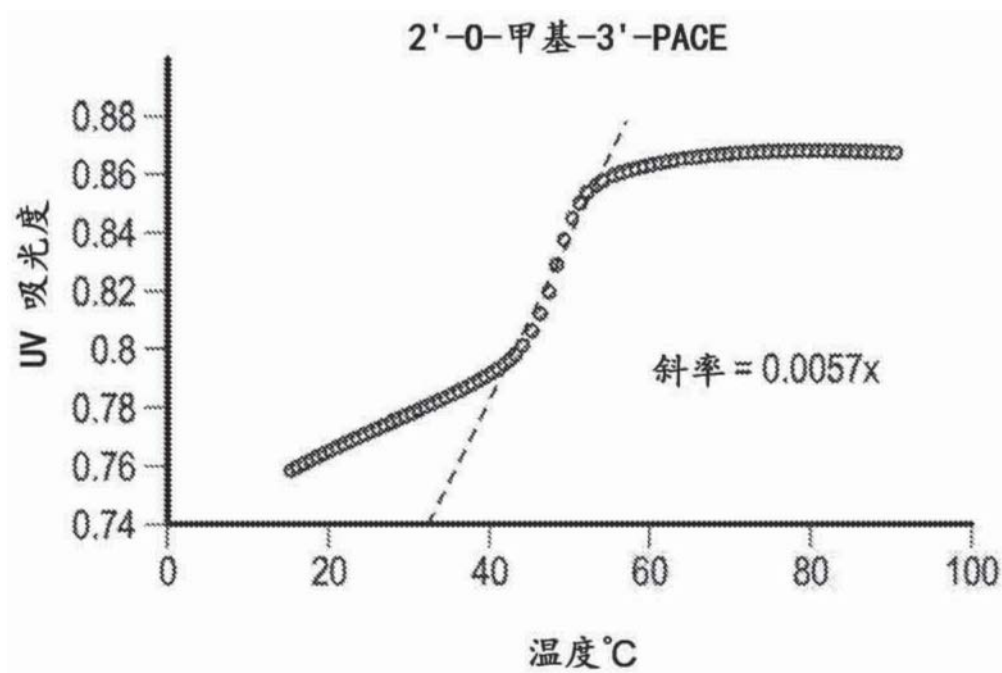


图6E

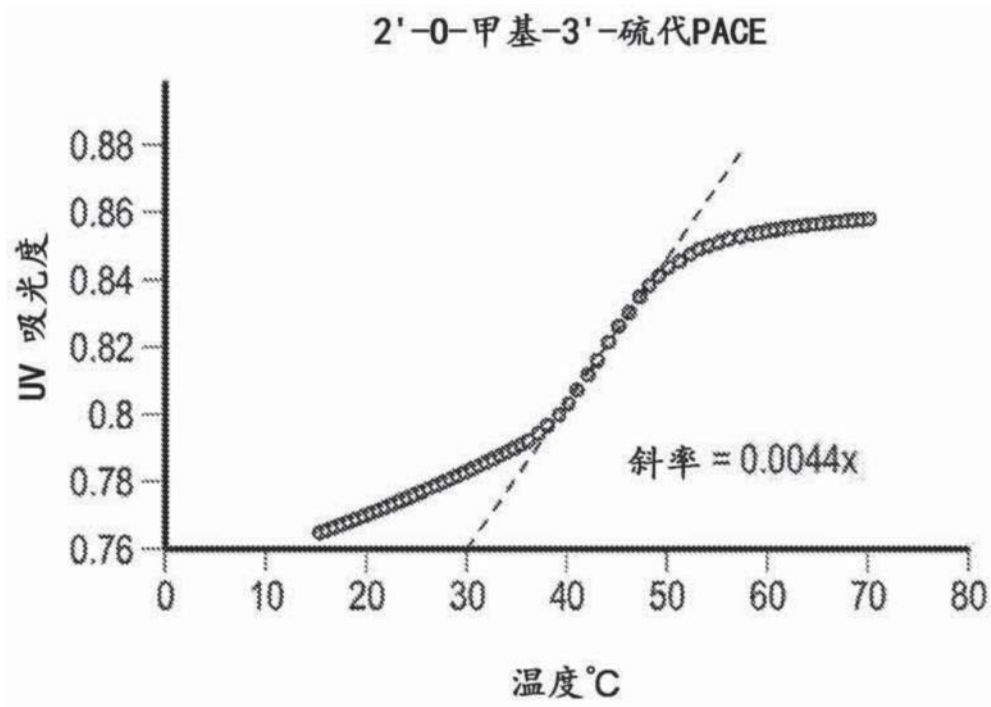


图6F

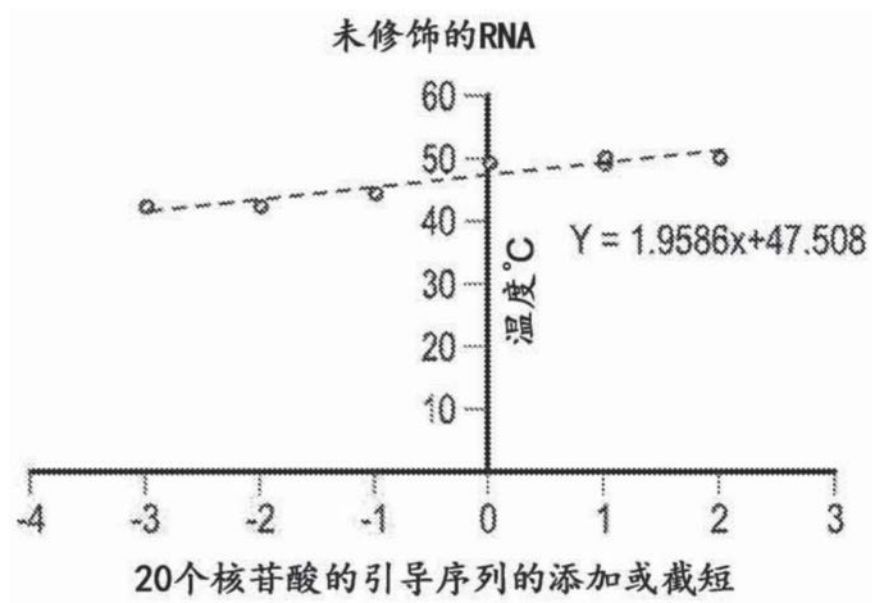


图7



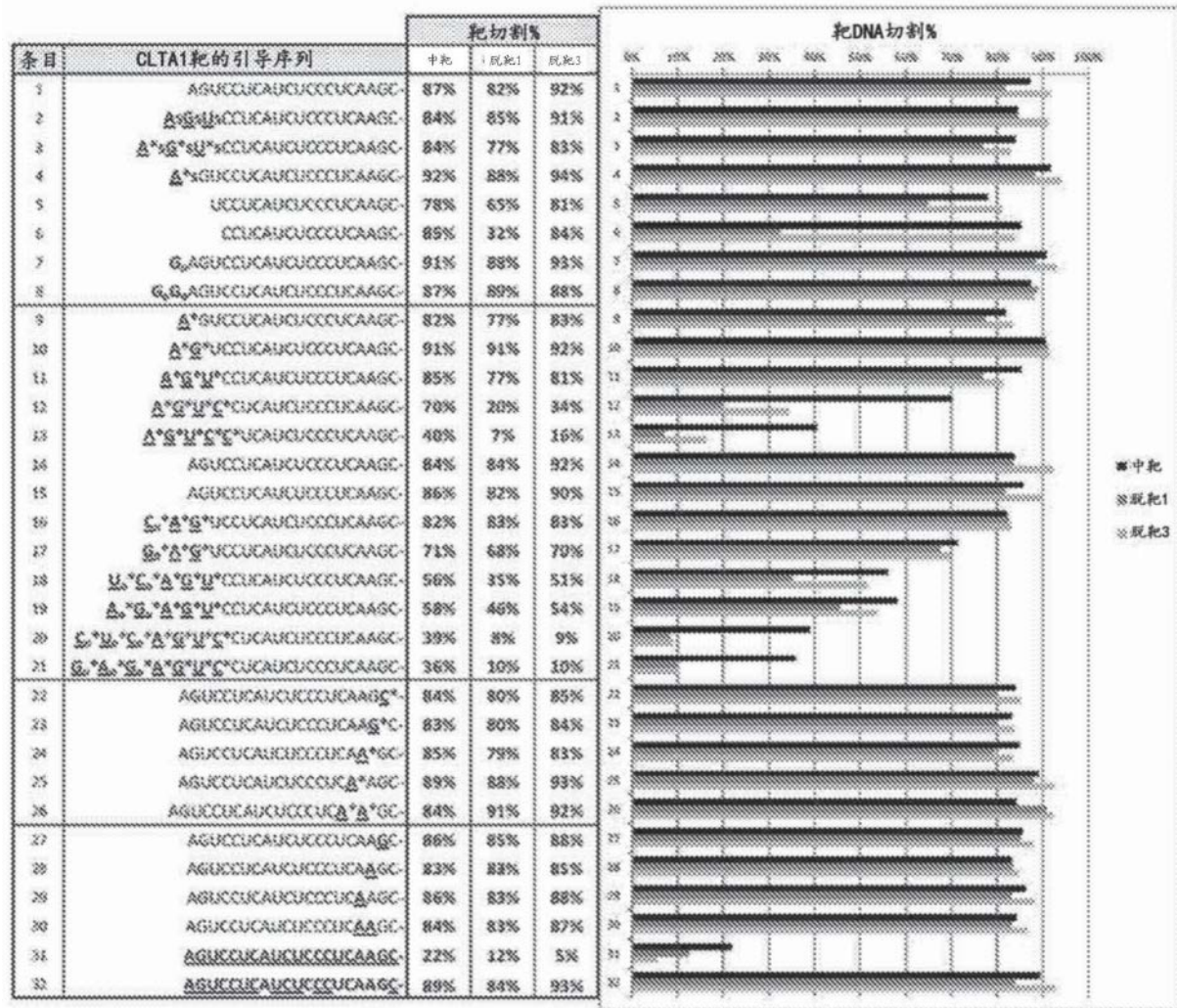


图8A



CLTA1体外切割的中靶与脱靶切割比和特异性评分					
条目	CLTA1靶的引导序列	中靶： 脱靶1比	特异性 评分	中靶： 脱靶3比	特异性 评分
1	AGUCCUCAUCUCCCUCAAGC-	1.07	0.93	0.95	0.83
2	<u>A</u> s <u>G</u> s <u>U</u> sCCUCAUCUCCCUCAAGC-	1.00	0.84	0.93	0.78
3	<u>A</u> *s <u>G</u> *s <u>U</u> *sCCUCAUCUCCCUCAAGC-	1.09	0.92	1.01	0.85
4	<u>A</u> *sGUCCUCAUCUCCCUCAAGC-	1.04	0.95	0.97	0.89
5	UCCUCAUCUCCCUCAAGC-	1.20	0.94	0.96	0.75
6	CCUCAUCUCCCUCAAGC-	2.63	2.24	1.02	0.87
7	G <sub>0</sub> AGUCCUCAUCUCCCUCAAGC-	1.03	0.93	0.98	0.89
8	G <sub>0</sub> G <sub>0</sub> AGUCCUCAUCUCCCUCAAGC-	0.98	0.86	0.99	0.87
9	<u>A</u> *GUCCUCAUCUCCCUCAAGC-	1.06	0.87	0.98	0.81
10	<u>A</u> * <u>G</u> *UCCUCAUCUCCCUCAAGC-	0.99	0.90	0.99	0.90
11	<u>A</u> * <u>G</u> * <u>U</u> *CCUCAUCUCCCUCAAGC-	1.11	0.95	1.05	0.90
12	<u>A</u> * <u>G</u> * <u>U</u> * <u>C</u> *CUCAUCUCCCUCAAGC-	3.56	2.50	2.04	1.43
13	<u>A</u> * <u>G</u> * <u>U</u> * <u>C</u> * <u>C</u> *UCAUCUCCCUCAAGC-	5.75	2.32	2.49	1.01
14	AGUCCUCAUCUCCCUCAAGC-	1.00	0.84	0.91	0.76
15	AGUCCUCAUCUCCCUCAAGC-	1.05	0.90	0.95	0.81
16	<u>C</u> <sub>0</sub> * <u>A</u> * <u>G</u> *UCCUCAUCUCCCUCAAGC-	0.99	0.82	0.99	0.81
17	<u>G</u> <sub>0</sub> * <u>A</u> * <u>G</u> *UCCUCAUCUCCCUCAAGC-	1.06	0.75	1.02	0.73
18	<u>U</u> <sub>0</sub> * <u>C</u> <sub>0</sub> * <u>A</u> * <u>G</u> * <u>U</u> *CCUCAUCUCCCUCAAGC-	1.59	0.89	1.09	0.61
19	<u>A</u> <sub>0</sub> * <u>G</u> <sub>0</sub> * <u>A</u> * <u>G</u> * <u>U</u> *CCUCAUCUCCCUCAAGC-	1.27	0.74	1.08	0.63
20	<u>C</u> <sub>0</sub> * <u>U</u> <sub>0</sub> * <u>C</u> <sub>0</sub> * <u>A</u> * <u>G</u> * <u>U</u> * <u>C</u> *CUCAUCUCCCUCAAGC-	4.70	1.84	4.32	1.69
21	<u>G</u> <sub>0</sub> * <u>A</u> <sub>0</sub> * <u>G</u> <sub>0</sub> * <u>A</u> * <u>G</u> * <u>U</u> * <u>C</u> *CUCAUCUCCCUCAAGC-	3.69	1.33	3.50	1.26
22	AGUCCUCAUCUCCCUCAAG <u>C</u> *-	1.05	0.88	0.99	0.83
23	AGUCCUCAUCUCCCUCAAG <u>C</u> *-	1.04	0.87	0.99	0.83
24	AGUCCUCAUCUCCCUCA <u>A</u> *GC-	1.07	0.91	1.02	0.86
25	AGUCCUCAUCUCCCUCA <u>A</u> *AGC-	1.01	0.90	0.96	0.86
26	AGUCCUCAUCUCCCUCA <u>A</u> * <u>A</u> *GC-	0.93	0.78	0.91	0.77
27	AGUCCUCAUCUCCCUCAAG <u>C</u> -	1.01	0.86	0.97	0.83
28	AGUCCUCAUCUCCCUCAAG <u>C</u> -	1.00	0.83	0.98	0.82
29	AGUCCUCAUCUCCCUCAAG <u>C</u> -	1.04	0.90	0.98	0.84
30	AGUCCUCAUCUCCCUCAAG <u>C</u> -	1.01	0.85	0.97	0.82
31	<u>AGUCCUCAUCUCCCUCAAGC</u> -	1.76	0.39	4.11	0.90
32	<u>AGUCCUCAUCUCCCUCAAGC</u> -	1.06	0.95	0.96	0.86

图8B

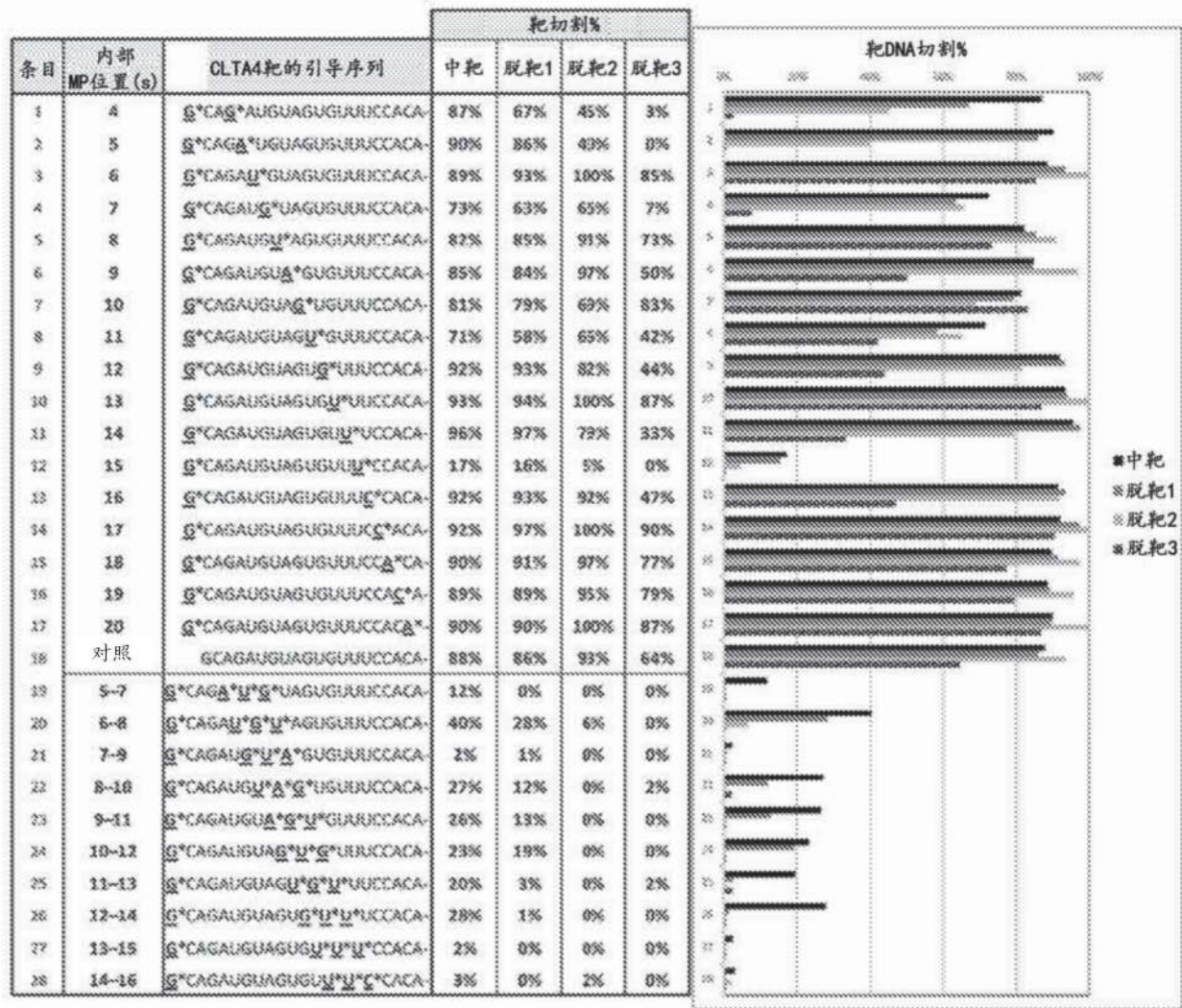


图9A

CLTA4体外切割的中靶与脱靶切割比和特异性评分								
条目	内部MP位置(s)	CLTA4靶的引导序列	中靶： 脱靶1比	特异性 评分	中靶： 脱靶2比	特异性 评分	中靶： 脱靶3比	特异性 评分
1	4	<u>G</u> *CAG <u>A</u> *AUGUAGUGUUCCACA-	1.3	1.1	1.9	1.7	34.5	30.0
2	5	<u>G</u> *CAGA <u>A</u> *UGUAGUGUUCCACA-	1.1	0.9	2.3	2.0	大	165.9
3	6	<u>G</u> *CAGAU <u>G</u> *GUAGUGUUCCACA-	1.0	0.8	0.9	0.8	1.0	0.9
4	7	<u>G</u> *CAGAU <u>G</u> *UAGUGUUCCACA-	1.1	0.8	1.1	0.8	9.7	7.1
5	8	<u>G</u> *CAGAU <u>G</u> *AGUGUUCCACA-	1.0	0.8	0.9	0.7	1.1	0.9
6	9	<u>G</u> *CAGAU <u>G</u> *GUGUUCCACA-	1.0	0.9	0.9	0.7	1.7	1.4
7	10	<u>G</u> *CAGAU <u>G</u> *UGUUCCACA-	1.0	0.8	1.2	1.0	1.0	0.8
8	11	<u>G</u> *CAGAU <u>G</u> *GUUCCACA-	1.2	0.9	1.1	0.8	1.7	1.2
9	12	<u>G</u> *CAGAU <u>G</u> *UUUCCACA-	1.0	0.9	1.1	1.0	2.1	1.9
10	13	<u>G</u> *CAGAU <u>G</u> *UCCACA-	1.0	0.9	0.9	0.9	1.1	1.0
11	14	<u>G</u> *CAGAU <u>G</u> *UCCACA-	1.0	0.9	1.2	1.2	2.9	2.8
12	15	<u>G</u> *CAGAU <u>G</u> *UCCACA-	1.1	0.2	3.7	0.6	大	6.0
13	16	<u>G</u> *CAGAU <u>G</u> *UCCACA-	1.0	0.9	1.0	0.9	2.0	1.8
14	17	<u>G</u> *CAGAU <u>G</u> *UCCACA-	0.9	0.9	0.9	0.8	1.0	0.9
15	18	<u>G</u> *CAGAU <u>G</u> *UCCACA-	1.0	0.9	0.9	0.8	1.2	1.0
16	19	<u>G</u> *CAGAU <u>G</u> *UCCACA-	1.0	0.9	0.9	0.8	1.1	1.0
17	20	<u>G</u> *CAGAU <u>G</u> *UCCACA-	1.0	0.9	0.9	0.8	1.0	0.9
18	对照	GCAGAU <u>G</u> *UCCACA-	1.0	0.9	0.9	0.8	1.4	1.2
19	5-7	<u>G</u> *CAGA <u>A</u> * <u>G</u> *UAGUGUUCCACA-	大	2.8	大	2.8	大	2.8
20	6-8	<u>G</u> *CAGAU <u>G</u> * <u>G</u> *UAGUGUUCCACA-	1.4	0.6	6.4	2.6	大	33.5
21	7-9	<u>G</u> *CAGAU <u>G</u> * <u>A</u> *GUGUUCCACA-	2.5	0.1	大	0.1	大	0.1
22	8-10	<u>G</u> *CAGAU <u>G</u> * <u>A</u> *GUGUUCCACA-	2.3	0.6	大	14.9	12.4	3.3
23	9-11	<u>G</u> *CAGAU <u>G</u> * <u>G</u> *GUUCCACA-	2.1	0.6	大	14.3	大	14.3
24	10-12	<u>G</u> *CAGAU <u>G</u> * <u>G</u> *UUUCCACA-	1.2	0.3	大	11.1	大	11.1
25	11-13	<u>G</u> *CAGAU <u>G</u> * <u>G</u> *UCCACA-	7.6	1.5	大	7.8	8.3	1.6
26	12-14	<u>G</u> *CAGAU <u>G</u> * <u>G</u> *UCCACA-	20.8	5.8	大	15.6	大	15.6
27	13-15	<u>G</u> *CAGAU <u>G</u> * <u>U</u> *UCCACA-	大	0.1	大	0.1	大	0.1
28	14-16	<u>G</u> *CAGAU <u>G</u> * <u>U</u> *UCCACA-	大	0.2	1.4	0.0	大	0.2

图9B



IL2RG体外切割的中靶与脱靶切割比和特异性评分						
条目	内部MP位置(s)	IL2RG靶的引导序列	靶切割%		比值 中靶:脱靶3	特异性 评分
			中靶	脱靶3		
1	2	<u>U</u> * <u>G</u> *GUAUGAUGGCUUCAACA-	80%	75%	1.1	0.9
2	3	<u>U</u> * <u>G</u> <u>G</u> *UAAUGAUGGCUUCAACA-	85%	62%	1.4	1.2
3	4	<u>U</u> * <u>G</u> <u>G</u> <u>U</u> *AAUGAUGGCUUCAACA-	63%	8%	8.4	5.2
4	5	<u>U</u> * <u>G</u> <u>G</u> <u>U</u> <u>A</u> *AUGAUGGCUUCAACA-	76%	31%	2.5	1.9
5	6	<u>U</u> * <u>G</u> <u>G</u> <u>U</u> <u>A</u> <u>A</u> *UGAUGGCUUCAACA-	75%	69%	1.1	0.8
6	7	<u>U</u> * <u>G</u> <u>G</u> <u>U</u> <u>A</u> <u>A</u> <u>U</u> *GAUGGCUUCAACA-	65%	3%	18.8	12.3
7	8	<u>U</u> * <u>G</u> <u>G</u> <u>U</u> <u>A</u> <u>A</u> <u>G</u> *AUGGCUUCAACA-	73%	53%	1.4	1.0
8	9	<u>U</u> * <u>G</u> <u>G</u> <u>U</u> <u>A</u> <u>A</u> <u>G</u> <u>A</u> *UGGCUUCAACA-	73%	32%	2.3	1.6
9	10	<u>U</u> * <u>G</u> <u>G</u> <u>U</u> <u>A</u> <u>A</u> <u>G</u> <u>A</u> <u>U</u> *GGCUUCAACA-	64%	48%	1.3	0.9
10	11	<u>U</u> * <u>G</u> <u>G</u> <u>U</u> <u>A</u> <u>A</u> <u>G</u> <u>A</u> <u>G</u> *GCUUCAACA-	65%	18%	3.7	2.4
11	12	<u>U</u> * <u>G</u> <u>G</u> <u>U</u> <u>A</u> <u>A</u> <u>G</u> <u>A</u> <u>G</u> <u>G</u> *CUUCAACA-	84%	10%	8.3	7.0
12	13	<u>U</u> * <u>G</u> <u>G</u> <u>U</u> <u>A</u> <u>A</u> <u>G</u> <u>A</u> <u>G</u> <u>G</u> <u>C</u> *UUCAACA-	95%	21%	4.5	4.2
13	14	<u>U</u> * <u>G</u> <u>G</u> <u>U</u> <u>A</u> <u>A</u> <u>G</u> <u>A</u> <u>G</u> <u>G</u> <u>C</u> <u>U</u> *UCAACA-	96%	10%	9.7	9.3
14	15	<u>U</u> * <u>G</u> <u>G</u> <u>U</u> <u>A</u> <u>A</u> <u>G</u> <u>A</u> <u>G</u> <u>G</u> <u>C</u> <u>U</u> <u>U</u> *CAACA-	29%	8%	3.8	1.1
15	16	<u>U</u> * <u>G</u> <u>G</u> <u>U</u> <u>A</u> <u>A</u> <u>G</u> <u>A</u> <u>G</u> <u>G</u> <u>C</u> <u>U</u> <u>C</u> *AACA-	96%	9%	10.7	10.2
16	17	<u>U</u> * <u>G</u> <u>G</u> <u>U</u> <u>A</u> <u>A</u> <u>G</u> <u>A</u> <u>G</u> <u>G</u> <u>C</u> <u>U</u> <u>C</u> <u>A</u> *ACA-	95%	12%	7.7	7.3
17	18	<u>U</u> * <u>G</u> <u>G</u> <u>U</u> <u>A</u> <u>A</u> <u>G</u> <u>A</u> <u>G</u> <u>G</u> <u>C</u> <u>U</u> <u>C</u> <u>A</u> <u>A</u> *CA-	92%	17%	5.4	4.9
18	19	<u>U</u> * <u>G</u> <u>G</u> <u>U</u> <u>A</u> <u>A</u> <u>G</u> <u>A</u> <u>G</u> <u>G</u> <u>C</u> <u>U</u> <u>C</u> <u>A</u> <u>C</u> *A-	92%	25%	3.6	3.3
19	20	<u>U</u> * <u>G</u> <u>G</u> <u>U</u> <u>A</u> <u>A</u> <u>G</u> <u>A</u> <u>G</u> <u>G</u> <u>C</u> <u>U</u> <u>C</u> <u>A</u> <u>C</u> <u>A</u> *-	87%	61%	1.4	1.2
20	对照	<u>U</u> * <u>G</u> <u>G</u> <u>U</u> <u>A</u> <u>A</u> <u>G</u> <u>A</u> <u>G</u> <u>G</u> <u>C</u> <u>U</u> <u>C</u> <u>A</u> <u>C</u> <u>A</u> -	84%	68%	1.2	1.0
21	4-6	<u>U</u> * <u>G</u> <u>G</u> <u>U</u> * <u>A</u> * <u>A</u> *UGAUGGCUUCAACA-	19%			
22	5-7	<u>U</u> * <u>G</u> <u>G</u> <u>U</u> <u>A</u> * <u>A</u> * <u>U</u> *GAUGGCUUCAACA-	11%			
23	6-8	<u>U</u> * <u>G</u> <u>G</u> <u>U</u> <u>A</u> * <u>U</u> * <u>G</u> *AUGGCUUCAACA-	4%			
24	7-9	<u>U</u> * <u>G</u> <u>G</u> <u>U</u> <u>A</u> <u>U</u> * <u>G</u> * <u>A</u> *UGGCUUCAACA-	2%	3%	0.5	0.0
25	8-10	<u>U</u> * <u>G</u> <u>G</u> <u>U</u> <u>A</u> <u>U</u> <u>G</u> * <u>A</u> * <u>U</u> *GGCUUCAACA-	23%	4%	5.7	1.3
26	9-11	<u>U</u> * <u>G</u> <u>G</u> <u>U</u> <u>A</u> <u>U</u> <u>G</u> <u>A</u> * <u>U</u> * <u>G</u> *GCUUCAACA-	50%	4%	12.1	6.0
27	10-12	<u>U</u> * <u>G</u> <u>G</u> <u>U</u> <u>A</u> <u>U</u> <u>G</u> <u>A</u> <u>U</u> * <u>G</u> * <u>G</u> *CUUCAACA-	39%	6%	6.6	2.6
28	11-13	<u>U</u> * <u>G</u> <u>G</u> <u>U</u> <u>A</u> <u>U</u> <u>G</u> <u>A</u> <u>U</u> <u>G</u> * <u>G</u> * <u>C</u> *UUCAACA-	35%	3%	10.7	3.7
29	12-14	<u>U</u> * <u>G</u> <u>G</u> <u>U</u> <u>A</u> <u>U</u> <u>G</u> <u>A</u> <u>U</u> <u>G</u> <u>G</u> * <u>C</u> * <u>U</u> *UCAACA-	19%	3%	5.6	1.1
30	13-15	<u>U</u> * <u>G</u> <u>G</u> <u>U</u> <u>A</u> <u>U</u> <u>G</u> <u>A</u> <u>U</u> <u>G</u> <u>G</u> <u>C</u> * <u>U</u> * <u>U</u> *CAACA-	4%	4%	1.1	0.0

图10

HBB体外切割的中靶与脱靶切割比和特异性评分						
条目	内部MP位置(s)	HBB靶的引导序列	靶切割%		比值	特异性评分
			中靶	脱靶1	中靶:脱靶1	
1	4	<u>C</u> *UUG*CCCCACAGGGCAGUAA-	30%	3%	10.1	3.0
2	5	C*UUG <u>C</u> *CCCACAGGGCAGUAA-	53%	3%	16.7	8.9
3	6	C*UUGC <u>C</u> *CCACAGGGCAGUAA-	73%	37%	2.0	1.4
4	7	C*UUGCC <u>C</u> *CACAGGGCAGUAA-	39%	3%	14.1	5.5
5	8	C*UUGCCC <u>C</u> *ACAGGGCAGUAA-	40%	7%	6.1	2.4
6	9	C*UUGCCCC <u>A</u> *CAGGGCAGUAA-	59%	5%	11.9	7.0
7	10	C*UUGCCCCA <u>C</u> *AGGGCAGUAA-	58%	6%	10.1	5.9
8	11	C*UUGCCCCACA <u>A</u> *GGGCAGUAA-	53%	3%	20.3	10.9
9	12	C*UUGCCCCACAG <u>G</u> *GGCAGUAA-	75%	21%	3.6	2.7
10	13	C*UUGCCCCACAGG <u>G</u> *GCAGUAA-	66%	20%	3.3	2.2
11	14	C*UUGCCCCACAGGG <u>G</u> *CAGUAA-	76%	19%	3.9	3.0
12	15	C*UUGCCCCACAGGGC <u>C</u> *AGUAA-	0%	2%	0.0	0.0
13	16	C*UUGCCCCACAGGGCA <u>A</u> *GUAA-	65%	27%	2.4	1.6
14	17	C*UUGCCCCACAGGGCAG <u>G</u> *UAA-	81%	47%	1.7	1.4
15	18	C*UUGCCCCACAGGGCAGU <u>U</u> *AA-	76%	34%	2.2	1.7
16	19	C*UUGCCCCACAGGGCAGUA <u>A</u> *A-	62%	23%	2.7	1.7
17	对照	C*UUGCCCCACAGGGCAGUAA-	75%	38%	2.0	1.5

图11A

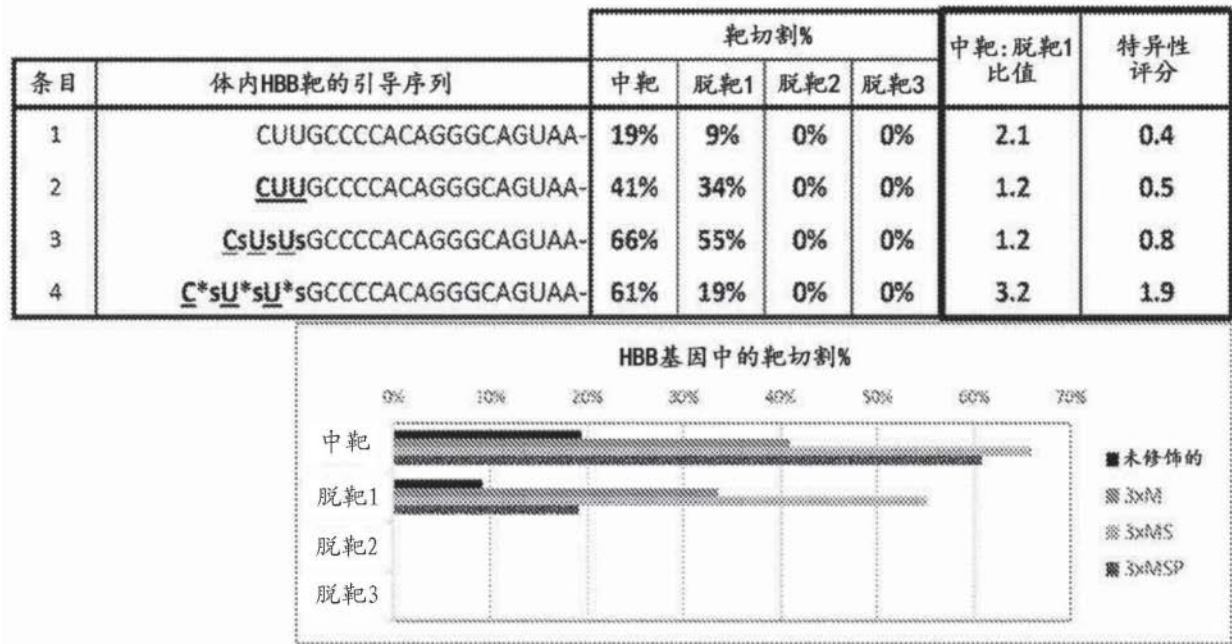


图11B



条目	sgRNA 名称	中靶	脱靶	中靶： 脱靶1比	特异性 评分
1	HBB_11MP_1xMP	53%	3%	20.3	10.9
2	HBB_5MP_1xMP	53%	3%	16.7	8.9
3	HBB_9MP_1xMP	59%	5%	11.9	7.0
4	HBB_10MP_1xMP	58%	6%	10.1	5.9
5	HBB_7MP_1xMP	39%	3%	14.1	5.5
6	HBB_14MP_1xMP	76%	19%	3.9	3.0
7	HBB_4MP_1xMP	30%	3%	10.1	3.0
8	HBB_12MP_1xMP	75%	21%	3.6	2.7
9	HBB_8MP_1xMP	40%	7%	6.1	2.4
10	HBB_13MP_1xMP	66%	20%	3.3	2.2
11	HBB_18MP_1xMP	76%	34%	2.2	1.7
12	HBB_19MP_1xMP	62%	23%	2.7	1.7
13	HBB_16MP_1xMP	65%	27%	2.4	1.6
14	HBB_1xMP (对照)	75%	38%	2.0	1.5
15	HBB_6MP_1xMP	73%	37%	2.0	1.4
16	HBB_17MP_1xMP	81%	47%	1.7	1.4
17	HBB_15MP_1xMP	0%	2%	0.0	0.0
18	HBB_6,7MP_1xMP	50%	1%	50.0	25.0
19	HBB_10,17MP_1xMP	50%	1%	49.5	24.5
20	HBB_5,17MP_1xMP	46%	1%	46.1	21.3
21	HBB_5,16MP_1xMP	45%	1%	45.1	20.4
22	HBB_6,10MP_1xMP	61%	2%	34.5	20.9
23	HBB_10,16MP_1xMP	40%	1%	39.6	15.7
24	HBB_5,9MP_1xMP	65%	3%	23.5	15.3
25	HBB_9,16MP_1xMP	38%	1%	38.1	14.6
26	HBB_8,17MP_1xMP	37%	1%	37.5	14.0
27	HBB_6,8MP_1xMP	45%	2%	26.7	12.0
28	HBB_9,10MP_1xMP	50%	2%	24.0	12.0
29	HBB_10,13MP_1xMP	51%	2%	22.2	11.4
30	HBB_9,14MP_1xMP	34%	1%	33.5	11.3
31	HBB_9,17MP_1xMP	54%	3%	20.4	11.1
32	HBB_13,17MP_1xMP	32%	1%	31.9	10.2
33	HBB_7,14MP_1xMP	48%	2%	20.7	9.8
34	HBB_5,14MP_1xMP	46%	2%	19.8	9.2
35	HBB_5,6MP_1xMP	58%	4%	13.5	7.9
36	HBB_9,13MP_1xMP	54%	4%	12.8	7.0
37	HBB_6,11MP_1xMP	43%	3%	16.7	7.2
38	HBB_13,14MP_1xMP	75%	8%	9.6	7.2
39	HBB_6,9MP_1xMP	61%	6%	10.2	6.2
40	HBB_14,17MP_1xMP	23%	1%	23.3	5.4
41	HBB_7,17MP_1xMP	53%	6%	9.3	4.9
42	HBB_16,17MP_1xMP	21%	1%	20.9	4.4
43	HBB_8,14MP_1xMP	36%	2%	16.0	5.7
44	HBB_8,9MP_1xMP	27%	2%	17.5	4.7
45	HBB_8,13MP_1xMP	42%	4%	9.4	3.9
46	HBB_7,13MP_1xMP	39%	4%	10.2	3.9
47	HBB_8,16MP_1xMP	32%	3%	11.8	3.8
48	HBB_10,14MP_1xMP	26%	2%	14.0	3.7
49	HBB_7,16MP_1xMP	37%	4%	9.0	3.3
50	HBB_7,9MP_1xMP	32%	3%	10.1	3.3
51	HBB_5,8MP_1xMP	17%	1%	16.9	2.9
52	HBB_6,14MP_1xMP	64%	14%	4.5	2.9
53	HBB_6,13MP_1xMP	69%	16%	4.2	2.9
54	HBB_5,13MP_1xMP	35%	5%	7.7	2.7
55	HBB_6,16MP_1xMP	62%	17%	3.6	2.2
56	HBB_7,10MP_1xMP	21%	2%	8.4	1.7
57	HBB_6,17MP_1xMP	68%	37%	1.8	1.3
58	HBB_8,10MP_1xMP	16%	2%	7.8	1.2
59	HBB_1xMP (对照)	69%	44%	1.6	1.1
60	HBB_13,16MP_1xMP	10%	1%	9.9	1.0
61	HBB_14,16MP_1xMP	9%	1%	9.0	0.8
62	HBB_5,10MP_1xMP	16%	3%	5.1	0.8
63	HBB_5,7MP_1xMP	18%	6%	2.9	0.5
64	HBB_5,11MP_1xMP	4%	3%	1.5	0.1

图12A

在K562细胞中:

条目	sgRNA名称 或对照类型	中靶	脱靶 靶	中靶: 脱靶1比	特异性 评分
1	HBB_11MP_1xMP	69.6%	0.9%	80.2	55.8
2	HBB_5MP_1xMP	81.2%	3.0%	26.9	21.8
3	HBB_7MP_1xMP	53.5%	3.1%	17.2	9.2
4	HBB_10MP_1xMP	66.7%	9.1%	7.3	4.9
5	HBB_9MP_1xMP	65.0%	9.0%	7.2	4.7
6	HBB_13MP_1xMP	61.5%	19.6%	3.1	1.9
7	HBB_3xMS (对照)	92.3%	52.6%	1.8	1.6
8	HBB_unmodif (对照)	84.3%	46.5%	1.8	1.5
9	HBB_1xMP (对照)	78.5%	44.4%	1.8	1.4
10	HBB_14MP_1xMP	34.9%	12.3%	2.8	1.0
11	HBB_17MP_1xMP	34.4%	22.0%	1.6	0.5
12	HBB_16MP_1xMP	9.1%	5.5%	1.7	0.2
13	HBB_6_10MP_1xMP	64.4%	4.4%	14.6	9.4
14	HBB_5_17MP_1xMP	25.8%	1.5%	17.4	4.5
15	HBB_6_7MP_1xMP	29.5%	2.7%	10.8	3.2
16	HBB_10_17MP_1xMP	24.9%	3.0%	8.3	2.1
17	HBB_5_16MP_1xMP	6.5%	0.2%	27.6	1.8
18	HBB_5_9MP_1xMP	12.6%	2.0%	6.3	0.8
19	HBB_9_17MP_1xMP	6.5%	0.8%	7.8	0.5
20	GFP转染对照	0.1%	0.1%		
21	模拟转染对照	0.2%	0.0%		

在iPS细胞中:

条目	sgRNA名称 或对照类型	中靶	脱靶 靶	中靶: 脱靶1比	特异性 评分
22	HBB_11MP_1xMP	45.1%	0.5%	96.1	43.3
23	HBB_5MP_1xMP	51.8%	1.8%	29.3	15.2
24	HBB_7MP_1xMP	39.2%	2.0%	19.9	7.8
25	HBB_10MP_1xMP	37.9%	6.1%	6.3	2.4
26	HBB_9MP_1xMP	32.5%	4.9%	6.6	2.1
27	HBB_13MP_1xMP	37.1%	19.7%	1.9	0.7
28	HBB_14MP_1xMP	25.2%	14.3%	1.8	0.4
29	HBB_3xMS (对照)	31.3%	22.6%	1.4	0.4
30	HBB_1xMP (对照)	34.0%	31.8%	1.1	0.4
31	HBB_unmodif (对照)	25.3%	19.6%	1.3	0.3
32	HBB_16MP_1xMP	7.8%	6.2%	1.3	0.1
33	HBB_17MP_1xMP	16.4%	17.3%	1.0	0.2
34	HBB_6_10MP_1xMP	35.7%	2.0%	17.4	6.2
35	HBB_5_17MP_1xMP	16.2%	0.6%	29.1	4.7
36	HBB_5_16MP_1xMP	6.3%	0.1%	59.1	3.7
37	HBB_6_7MP_1xMP	21.2%	2.1%	10.8	2.1
38	HBB_10_17MP_1xMP	11.4%	1.1%	10.1	1.1
39	HBB_5_9MP_1xMP	9.0%	1.4%	6.5	0.6
40	HBB_9_17MP_1xMP	3.3%	0.6%	5.2	0.2
41	GFP转染对照	0.0%	0.0%		
42	模拟转染对照	0.0%	0.0%		

图12B

在K562细胞中:

条目	sgRNA名称 或对照类型	中靶	脱靶 靶	中靶: 脱靶1比	特异性 评分
1	HBB_11MP_1xMP	69.6%	0.9%	80.2	55.8
2	HBB_5MP_1xMP	81.2%	3.0%	26.9	21.8
3	HBB_7MP_1xMP	53.5%	3.1%	17.2	9.2
4	HBB_10MP_1xMP	66.7%	9.1%	7.3	4.9
5	HBB_9MP_1xMP	65.0%	9.0%	7.2	4.7
6	HBB_13MP_1xMP	61.5%	19.6%	3.1	1.9
7	HBB_14MP_1xMP	34.9%	12.3%	2.8	1.0
8	HBB_3xMS (对照)	92.3%	52.6%	1.8	1.6
9	HBB_unmodif (对照)	84.3%	46.5%	1.8	1.5
10	HBB_1xMP (对照)	78.5%	44.4%	1.8	1.4
11	HBB_16MP_1xMP	9.1%	5.5%	1.7	0.2
12	HBB_17MP_1xMP	34.4%	22.0%	1.6	0.5
13	HBB_5_16MP_1xMP	6.5%	0.2%	27.6	1.8
14	HBB_5_17MP_1xMP	25.8%	1.5%	17.4	4.5
15	HBB_6_10MP_1xMP	64.4%	4.4%	14.6	9.4
16	HBB_6_7MP_1xMP	29.5%	2.7%	10.8	3.2
17	HBB_10_17MP_1xMP	24.9%	3.0%	8.3	2.1
18	HBB_9_17MP_1xMP	6.5%	0.8%	7.8	0.5
19	HBB_5_9MP_1xMP	12.6%	2.0%	6.3	0.8
20	GFP转染对照	0.1%	0.1%		
21	模拟转染对照	0.2%	0.0%		

在iPS细胞中:

条目	sgRNA名称 或对照类型	中靶	脱靶 靶	中靶: 脱靶1比	特异性 评分
22	HBB_11MP_1xMP	45.1%	0.5%	96.1	43.3
23	HBB_5MP_1xMP	51.8%	1.8%	29.3	15.2
24	HBB_7MP_1xMP	39.2%	2.0%	19.9	7.8
26	HBB_9MP_1xMP	32.5%	4.9%	6.6	2.1
25	HBB_10MP_1xMP	37.9%	6.1%	6.3	2.4
27	HBB_13MP_1xMP	37.1%	19.7%	1.9	0.7
28	HBB_14MP_1xMP	25.2%	14.3%	1.8	0.4
29	HBB_3xMS (对照)	31.3%	22.6%	1.4	0.4
31	HBB_unmodif (对照)	25.3%	19.6%	1.3	0.3
32	HBB_16MP_1xMP	7.8%	6.2%	1.3	0.1
30	HBB_1xMP (对照)	34.0%	31.8%	1.1	0.4
33	HBB_17MP_1xMP	16.4%	17.3%	1.0	0.2
34	HBB_5_16MP_1xMP	6.3%	0.1%	59.1	3.7
35	HBB_5_17MP_1xMP	16.2%	0.6%	29.1	4.7
36	HBB_6_10MP_1xMP	35.7%	2.0%	17.4	6.2
37	HBB_10_17MP_1xMP	11.4%	1.1%	10.1	1.1
38	HBB_6_7MP_1xMP	21.2%	2.1%	10.8	2.1
39	HBB_5_9MP_1xMP	9.0%	1.4%	6.5	0.6
40	HBB_9_17MP_1xMP	3.3%	0.6%	5.2	0.2
41	GFP转染对照	0.0%	0.0%		
42	模拟转染对照	0.0%	0.0%		

图12C

[illegible]

图13



条目	靶基因	sgRNA名称 或对照类型	中靶	脱2 靶	中靶:脱靶2 比	特异性 评分
1	<i>IL2RG</i>	IL2RG-未修饰的(对照)	98.5%	5.228%	18.8	18.6
2	<i>IL2RG</i>	IL2RG_1xMP	97.8%	0.094%	1,040	1,020
3	<i>IL2RG</i>	IL2RG_5MP_1xMP	93.2%	0.005%	18,600	17,400
4	<i>IL2RG</i>	IL2RG_11MP_1xMP	87.1%	0.008%	10,900	9,500
5	<i>VEGFA</i>	VEGFA_1xMP	74.1%	10.82%	6.8	5.1
6	<i>VEGFA</i>	VEGFA_5MP_1xMP	75.7%	0.19%	398	302
7	<i>VEGFA</i>	VEGFA_7MP_1xMP	16.5%	0.02%	825	136
8	<i>VEGFA</i>	VEGFA_9MP_1xMP	42.0%	0.22%	191	80
9	<i>VEGFA</i>	VEGFA_10MP_1xMP	23.7%	0.08%	296	70
10	<i>VEGFA</i>	VEGFA_11MP_1xMP	23.9%	0.11%	217	52

图14

用于K562细胞中的HBB靶向的中靶和脱靶HDR和插入缺失形成

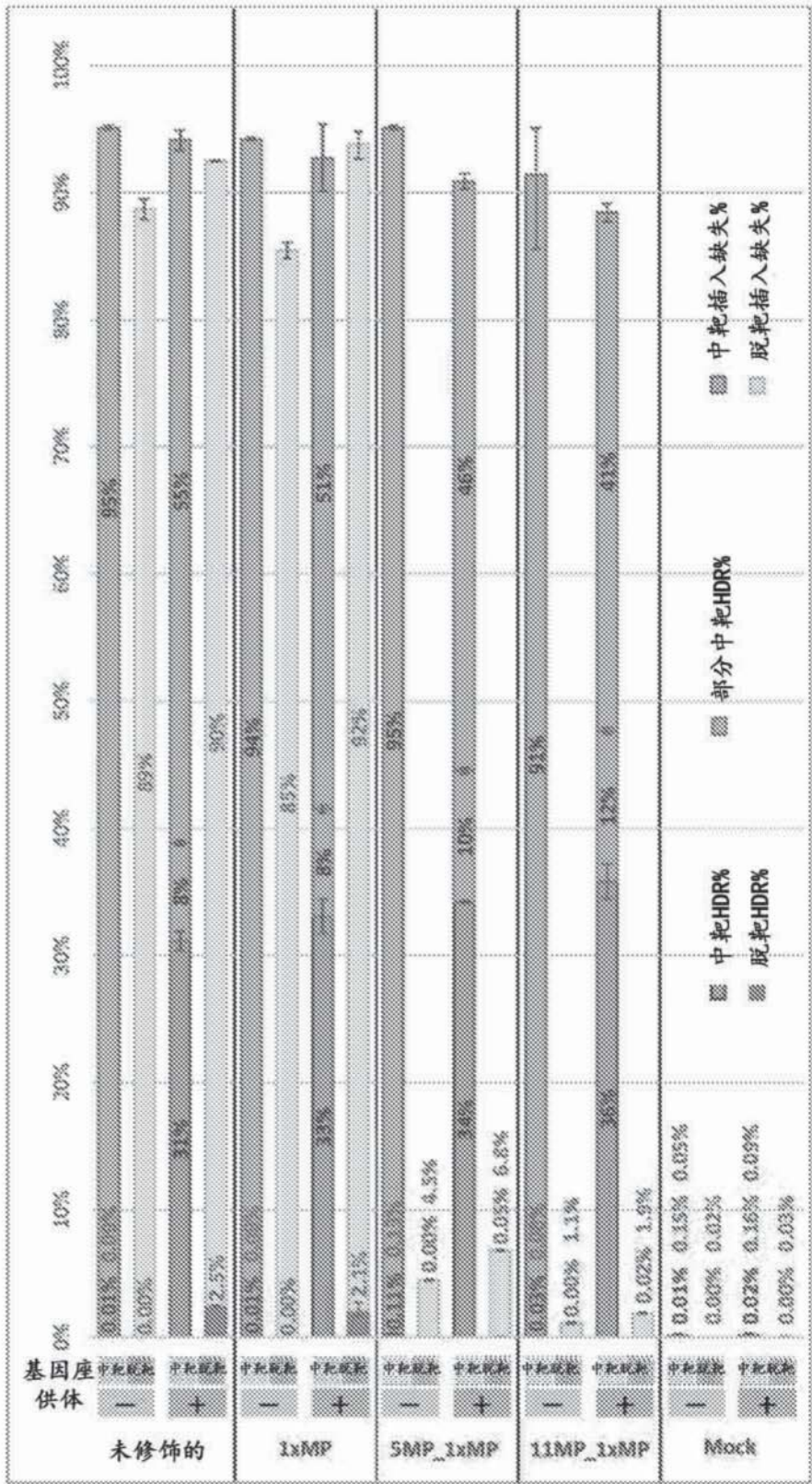


图15



图16



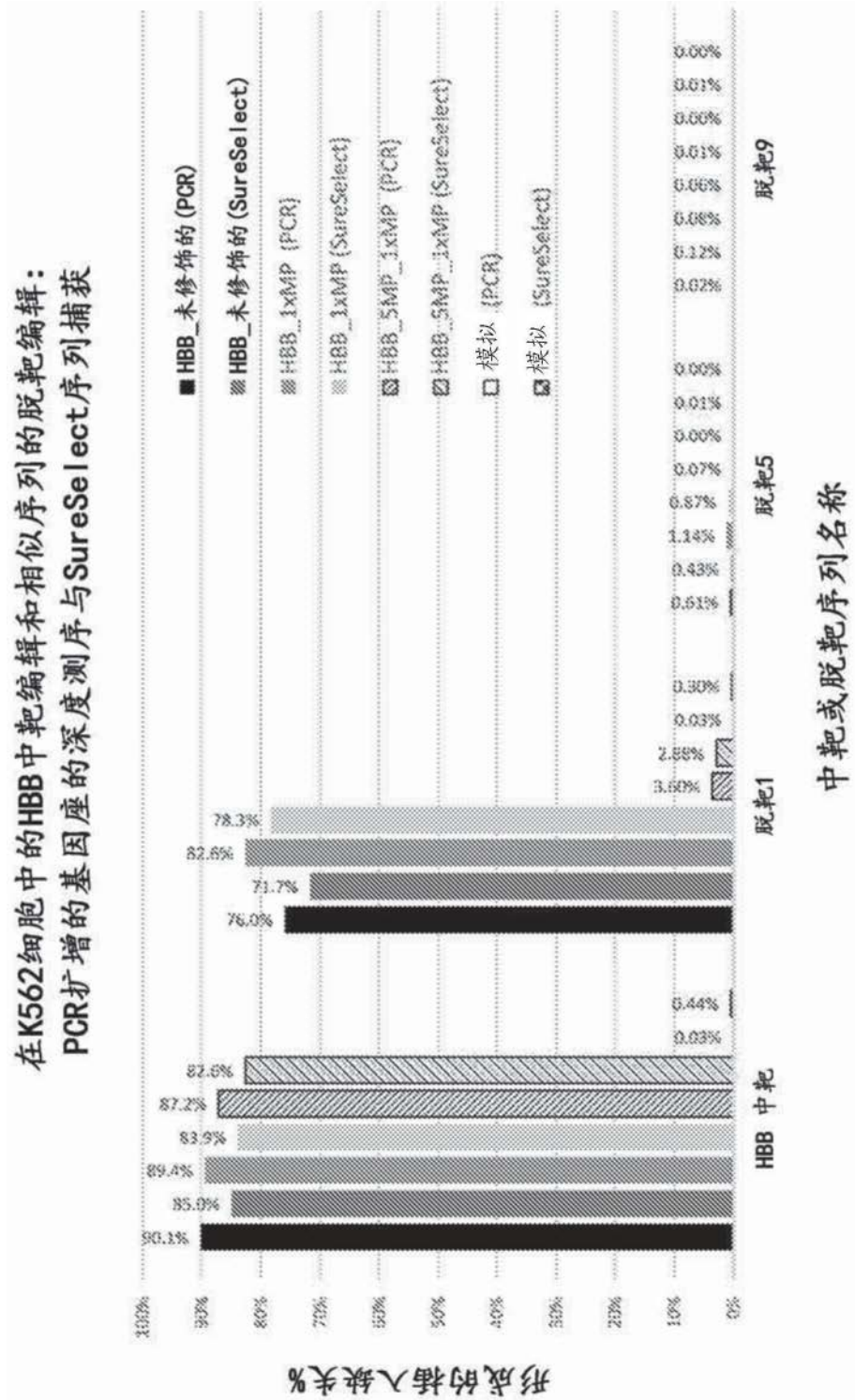


图17