(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 114231527 A
(43)申请公布日 2022.03.25

(21)申请号 202111196588.9

(22)申请日 2016.04.05

(30)优先权数据
62/143,729 2015.04.06 US
62/160,545 2015.05.12 US

(62)分案原申请数据
201680032647.X 2016.04.05

(71)申请人 里兰斯坦福初级大学理事会
地址 美国加利福尼亚州
申请人 安捷伦科技有限公司

(72)发明人 M·H·波蒂厄斯 A·亨德尔
J·克拉克 R·O·巴克
D·E·赖安 D·J·德林杰
R·凯撒 J·迈尔森

(74)专利代理机构 北京坤瑞律师事务所 11494
代理人 罗天乐

(51)Int.Cl.
C12N 15/113(2010.01)
C12N 15/85(2006.01)
C12N 9/22(2006.01)
C12N 15/55(2006.01)
C12N 5/10(2006.01)
A61K 31/7105(2006.01)
A61K 48/00(2006.01)
A61P 7/06(2006.01)
A61P 7/04(2006.01)
A61P 35/00(2006.01)
A61P 31/18(2006.01)
A61P 25/18(2006.01)
A61P 27/02(2006.01)
A61P 25/34(2006.01)
A61P 25/28(2006.01)
A61P 25/16(2006.01)
A61P 7/00(2006.01)
A61P 29/00(2006.01)
A61P 37/00(2006.01)
A61P 3/00(2006.01)
A61P 1/16(2006.01)
A61P 13/12(2006.01)
A61P 21/00(2006.01)
A61P 19/08(2006.01)
A61P 25/00(2006.01)
A61P 9/00(2006.01)
A61P 11/00(2006.01)

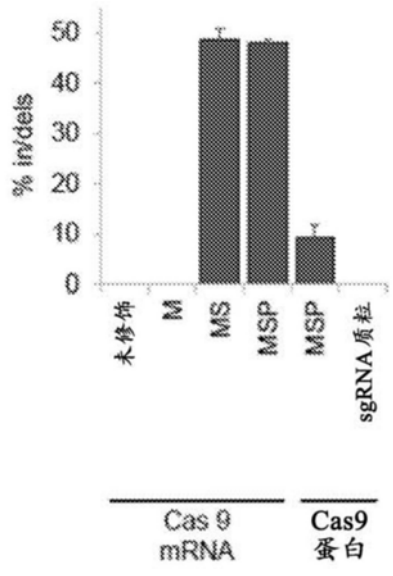
权利要求书1页 说明书63页 附图23页

(54)发明名称

用于CRISPR/CAS介导的基因调控的化学修饰的引导RNA

(57)摘要

本文提供用于诱导细胞中靶核酸(例如靶DNA或靶RNA)的基于CRISPR/Cas的基因调控(例如,基因组编辑或基因表达)的方法。所述方法包括使用修饰的单引导RNA(sgRNA),所述sgRNA可增强用于离体治疗的原代细胞中或用于离体治疗的受试者细胞中的靶核酸的基因调节。另外,本文提供了通过施用足够量的修饰的sgRNA以纠正靶基因中与遗传疾病相关的突变来预防或治疗受试者的遗传疾病的方法。



CN 114231527 A

1.一种诱导原代细胞中靶核酸的基因调控的方法,所述方法包括:

向原代细胞中导入:

(a)修饰的单引导RNA(sgRNA),其包含与靶核酸互补的第一核苷酸序列和与CRISPR相关蛋白(Cas)多肽相互作用的第二核苷酸序列,其中第一核苷酸序列和/或第二核苷酸序列中的一个或多个核苷酸是修饰的核苷酸;和

(b)Cas多肽、编码Cas多肽的mRNA、和/或包含编码Cas多肽的核苷酸序列的重组表达载体,

其中所述修饰的sgRNA将Cas多肽引导至靶核酸,

其中所述修饰的sgRNA以相对于相应的未修饰的sgRNA增强的活性诱导所述靶核酸的基因调控。

2.权利要求1的方法,其中所述增强的活性包括:修饰的sgRNA的稳定性增加,和/或修饰的sgRNA对靶核酸的特异性增加。

3.权利要求1或2的方法,其中所述靶核酸包含靶DNA或靶RNA。

4.权利要求3的方法,其中所述基因调控包括对靶DNA的基因组编辑。

5.权利要求4的方法,其中所述基因组编辑包括对靶DNA的同源性指导修复(HDR)或非同源末端连接(NHEJ)。

6.权利要求4或5的方法,还包括将重组供体修复模板导入原代细胞。

7.权利要求6的方法,其中所述重组供体修复模板包含两个核苷酸序列,所述两个核苷酸序列包含靶DNA的两个非重叠的、同源的部分,其中这两个核苷酸序列位于与要进行基因组编辑的靶DNA对应的核苷酸序列的5'端和3'端。

8.权利要求6的方法,其中所述重组供体修复模板包含合成单链寡聚脱氧核苷酸(ssODN)模板,该模板包含:编码用于纠正单核苷酸多态性(SNP)的突变的核苷酸序列,和包含靶DNA的两个非重叠的、同源的部分的两个核苷酸序列,其中这两个核苷酸序列位于所述编码突变的核苷酸序列的5'端和3'端。

9.权利要求3的方法,其中所述基因调控包括使用内切核酸酶缺陷的Cas多肽来抑制或激活靶DNA或靶RNA的基因表达。

10.权利要求1至9中任一项的方法,其中所述原代细胞在将所述修饰的sgRNA和Cas多肽导入原代细胞之前从多细胞生物体中分离。

用于CRISPR/CAS介导的基因调控的化学修饰的引导RNA

[0001] 本申请是2016年4月5日提交的申请号为201680032647.X、发明名称为“用于CRISPR/CAS介导的基因调控的化学修饰的引导RNA”的发明专利申请 的分案申请。

[0002] 对相关申请的交叉引用

[0003] 本申请要求2015年4月6日提交的美国临时申请No.62/143,729和2015年5 月12日提交的美国临时申请No.62/160,545的优先权,将其全部公开内容通过 引用并入本文中用于所有目的。

[0004]

[0005]

[0006] 发明背景

[0007] 用工程化的核酸酶编辑基因组是一项可用于修饰几乎任何感兴趣的基 因组序列的突破性技术(Porteus,M.H.&Carroll,D.,Nature Biotechnology 23, 967-973(2005))。该技术利用工程化核酸酶产生位点特异性双链断裂(DSB), 然后通过内源性细胞修复机制解决DSB。结果可能是特定位点通过诱变性非 同源末端连接(NHEJ)而被突变,从而在突变位点产生插入或缺失(in/del),或 者也可能是在外源导入的供体模板的情况下,基因组序列通过同源重组(HR) 而被精确改变(Hendel等,Trends in Biotechnology 33,132-140(2015))。此平台 近来增加的一个重要成员是成簇规则间隔的回文重复(CRISPR)/Cas系统,该 系统由RNA引导的核酸酶(Cas)和短引导RNA(sgRNA)组成Jinek,M.et al., Science 337,816-821(2012),Mali,P.et al.,Science 339,823-826(2013),Cong, L.et al., Science 339,819-823(2013),Hsu et al.,Cell 157,1262-1278(2014))。引导RNA由两条RNA组成,分别称为CRISPR RNA(crRNA)和反式激活 crRNA(tracrRNA),二者通常融合在嵌合的单引导RNA(sgRNA)中。

[0008] 用于基因组编辑的sgRNA可由100个核苷酸(nt)组成,其中5’末端的20nt 通过Watson-Crick碱基配对与靶DNA序列杂交,并引导Cas内切核酸酶切割靶 基因组DNA。

[0009] CRISPR/Cas系统也已被适用于序列特异性的基因表达控制,例如抑制或 激活基因表达。使用缺乏内切核酸酶活性的特定Cas9多肽变体,可以阻遏或 激活靶基因(Qi et al.,Cell,2013,152(5):1173-7783,Perez-Pinera et al.,Nat Methods,2013,10(10):973-976,Maeder et al.,Nat Methods,2013,10(10):977- 979,Gilbert et al.,Cell,2014,159:647-661,O’Connell et al.,Nature,2014, 516:263-266)。

[0010] 不幸的是,使用CRISPR/Cas系统进行基因组编辑和调节基因表达仍然是 低效的,尤其是在原代细胞中。因此,本领域仍需要更好的基于CRISPR/Cas 系统的组合物和方法,其可用于基因调控,例如基因组编辑,抑制基因表达 和激活基因表达。本发明满足了这种需要,并提供了更多的优点。

发明内容

[0011] 本发明提供用于诱导(例如启动、调变、增强等)细胞中靶核酸的基因调 控的方法。本发明包括使用这样的单引导RNA(sgRNA):它们可在原代细胞 中(例如在体外培养用

于离体疗法的)或受试者(例如人)体内的细胞中(例如用 于体内疗法的)增强靶核酸的基因编辑和/或抑制或激活靶核酸的基因表达。本发明还提供了通过增强精确的基因组编辑以纠正与疾病相关的靶基因中 的突变来预防或治疗受试者的疾病的方法。本发明可以与任何细胞类型以及 任何适合于核酸酶介导的基因组编辑技术的基因座位一起使用。

[0012] 在第一方面,本发明提供了用于诱导原代细胞中靶核酸的基因调节的方 法,所述方法包括:

[0013] 向原代细胞中导入:

[0014] (a)修饰的单引导RNA(sgRNA),其包含与靶核酸互补的第一核苷酸序 列和与CRISPR相关蛋白(Cas)多肽相互作用的第二核苷酸序列,其中第一核 苷酸序列和/或第二核苷酸序列中的一个或多个核苷酸是修饰的核苷酸;和

[0015] (b)Cas多肽,编码Cas多肽的mRNA、和/或包含编码Cas多肽的核苷酸序 列的重组表达载体,

[0016] 其中修饰的sgRNA将Cas多肽引导至靶核酸,

[0017] 其中修饰的sgRNA以相对于相应的未修饰的sgRNA增强的活性诱导靶 核酸的基因调节。

[0018] 在一个相关的方面,本发明提供了用于增强原代细胞中靶DNA的基因组 编辑的方法,所述方法包括:

[0019] 向原代细胞中导入:

[0020] (a)修饰的单引导RNA(sgRNA),其包含与靶核酸互补的第一核苷酸序 列和与CRISPR相关蛋白(Cas)多肽相互作用的第二核苷酸序列,其中第一核 苷酸序列和/或第二核苷酸序列中的一个或多个核苷酸是修饰的核苷酸;和

[0021] (b)Cas多肽、编码Cas多肽的mRNA、和/或包含编码Cas多肽的核苷酸序 列的重组表达载体,

[0022] 其中修饰的sgRNA将Cas多肽引导至靶DNA,

[0023] 其中修饰的sgRNA相对于相应的未修饰的sgRNA增强靶DNA的基因组 编辑(例如,藉由修饰的sgRNA的增加的稳定性和/或修饰的sgRNA对于靶 DNA的增加的特异性)。

[0024] 在第二方面中,本发明提供预防或治疗受试者的遗传疾病的方法,所述 方法包括:

[0025] 向受试者施用足够量的修饰的单引导RNA(sgRNA)以纠正靶基因中与 遗传疾病相关的突变,其中修饰的sgRNA包含与靶基因互补的第一核苷酸序 列和与CRISPR相关蛋白(Cas)多肽相互作用的第二核苷酸序列,并且其中第 一核苷酸序列和/或第二核苷酸序列中的一个或多个核苷酸是修饰的核苷酸。

[0026] 根据下面的详细描述和附图,本发明的其它目的、特征和优点对于本领 域技术人员会是明显的。

附图说明

[0027] 图1A-G显示合成的和化学修饰的sgRNA有助于在体外高水平的DNA切 割和在人类细胞系(K562)中高频率的in/del。图1A显示装载入Cas9并结合到 其基因组靶位点的IL2RG sgRNA的二级结构的序列和示意图。具有化学修饰 的核苷酸用白旗标记。图1B描绘了在

sgRNA的化学合成期间掺入的化学修饰 的结构(序列见表1)。当存在时,包括:2'-O-甲基(M)、2'-O-甲基、3'硫代磷酸 酯(MS)或2'-O-甲基、3'硫代PACE(MSP)(灰色显示)的化学修饰分别掺入在5' 和3'末端的三个末端核苷酸处。图1C显示通过对PCR扩增子的深度测序而测 量的诱变性NHEJ所致的基因破坏。图1D显示了由Cas9与合成sgRNA的组合 诱导的K562细胞中三个基因座IL2RG、HBB和CCR5上的HR基因添加。每100 万个细胞递送1μg(浅色)或20μg(深色)的合成sgRNA。Cas9是从质粒(2μg)表 达的,对于HR实验,加入了5μg编码GFP的供体质粒。作为阳性对照,使用 2μg编码sgRNA和Cas9蛋白二者的sgRNA质粒(灰色条形)。各条形代表平均值 +SEM,n=3。图1E显示了合成的sgRNA介导的靶向切割的特异性,20μg sgRNA,如图1C中所示地进行。通过靶定的基因组基因座、以及每个基因的 三个生物信息学预测的脱靶基因座的PCR扩增子进行深度测序来测量In/del 频率。各条形代表平均值+SEM,n=3。图1F显示使用电穿孔将15μg Cas9 mRNA和10μg IL2RG合成sgRNA时差递送(staggered delivery)到100万个 K562细胞中。条形代表平均in/del频率+SEM,n=3,通过对跨越sgRNA靶位 点的PCR扩增子的TIDE分析来测量,使用模拟物处理样品作为参考对照。在 图1G中,将Cas9蛋白与2.5倍摩尔过量的标示的合成IL2RG sgRNA预先复合, 并以标示的量核转染(nucleofect)至1百万个K562细胞中。通过如上的TIDE分 析测量In/del频率,各条形代表平均in/del频率+SEM,n=3。

[0028] 图2A-2C显示化学修饰的sgRNA促进原代人T细胞和$CD34^+$造血干细胞 和祖细胞(HSPC)中的高速率的基因破坏。将100万个原代人T细胞用10μg所 标示的合成sgRNA、以及15μg Cas9 mRNA或1μg Cas9编码质粒(图2A)进行核 转染。加入1μg编码sgRNA和Cas9蛋白二者的质粒用于比较。条形表示三个 不同供体的平均in/del频率+SEM,n=6,通过对跨越sgRNA靶位点的PCR扩 增子的TIDE分析来测量,使用模拟物处理样品作为参考对照。在图2B中,如 上所述地对经刺激的T细胞进行核转染,但是用2.5倍摩尔过量的所标示的合 成CCR5 sgRNA与15μg Cas9蛋白预复合。In/del频率由上面的TIDE分析测量。 条形代表三个不同的供体的平均in/del频率+SEM,n=6。在图2C中,用10μg 标示的靶向IL2RG或HBB的合成sgRNA、以及15μg Cas9 mRNA或1μg Cas9质 粒核转染500,000个动员的人外周血$CD34^+$ HSPC。纳入1μg编码sgRNA和Cas9 蛋白二者的sgRNA质粒用于比较。条形表示平均in/del频率+SEM,n=3,通 过T7核酸内切酶切割测定法测量。在图2D中,用15μg Cas9 mRNA和10μg标示的合成CCR5 sgRNA核转染100万个经刺激的T细胞或动员的人外周血 CD34+HSPC。当组合使用时,每种sgRNA的量是5μg。单sgRNA的样品的 In/del频率通过如上所述的TIDE分析测量,两个sgRNA的样品的in/del频率通 过对克隆的PCR产物测序(参见图18A-B)测量。条形代表平均in/del频率+ SEM,n=3。

[0029] 图3显示在体外由化学修饰的sgRNA所指导的Cas9对dsDNA靶标的切割。条形表示用Cas9蛋白和sgRNA处理的靶DNA片段(参见图4)的切割产物的百 分比产量。显示了每个sgRNA的三次独立合成的平均值±SEM。

[0030] 图4显示在体外由化学修饰的sgRNA所指导的Cas9对dsDNA靶标的切割。 在Bioanalyzer 2200上使用DNA 7500LabChips测定来自dsDNA靶标的生物化 学切割的切割产物。示出了每个靶标的代表性凝胶,更多的重复包括在图3 中绘制的结果中。各样品如下:(L):分子量标准;(1):未修饰的sgRNA+靶 DNA(-Cas9模拟物处理),(2):靶DNA +Cas9蛋白(-sgRNA模拟物处理),(3): 未修饰的sgRNA+靶DNA+Cas9蛋白,(4):M sgRNA+靶DNA+Cas9

蛋白，(5):MS sgRNA+靶DNA +Cas9蛋白,(6):MSP sgRNA+靶DNA +Cas9蛋白。

[0031] 图5显示了由合成的sgRNA介导的靶向切割的特异性。靶标特异性如图 1E中那样使用Illumina深度测序进行评估,但使用来自图1的样品,用1μg sgRNA核转染。通过对来自每个sgRNA的三个生物信息学预测的脱靶基因座 的PCR扩增子进行深度测序来测量In/del频率。条形代表平均值+SEM,n= 3,以对数尺度显示。in/del百分比参见表5。

[0032] 图6显示了在K562细胞中靶向IL2RG和Cas9mRNA的MSP sgRNA的滴定。 测量的in/del频率是三次重复的平均值,值在热图中标示。为了清楚起见,没 有标示重复样品(n=3)的SEM,但SEM都小于测量的标示值的4%。In/del频 率是通过对跨越sgRNA靶位点的PCR扩增子的TIDE分析测量的,使用模拟物 处理的样品作为参考对照。

[0033] 图7提供了时差递送sgRNA和Cas9 mRNA的示意性实验流程。得出图1G 的数据的实验的示意性概述。在标示的时间点用Cas9 mRNA和/或靶向IL2RG 的sgRNA对K562细胞进行核转染。在最后一个组分核转染后72小时提取基因 组DNA,并通过TIDE测量in/del频率,使用模拟处理的样品作为参考对照。

[0034] 图8比较来自不同供应商的Cas9蛋白。用与2.5倍摩尔过量的MS sgRNA 预复合过的15μg Cas9蛋白质核转染一百万个K562细胞,三天后提取基因组 DNA,通过TIDE测量in/del频率,使用模拟处理的样品作为参考对照。条形 代表平均值+SEM,n=3。

[0035] 图9A-9B显示靶向切割的特异性由合成的IL2RG sgRNA和Cas9质粒、mRNA或蛋白介导。靶标特异性如图1E和图5所示进行评估,使用Illumina深 度测序并以线性标度(图9A)和对数标度(图9B)显示。用(i)2μg Cas9质粒+20μg sgRNA,(ii)15μg Cas9 mRNA+10μg sgRNA,或(iii)与7.6μg sgRNA预先复合 的15μgCas9蛋白质(蛋白质:sgRNA摩尔比=1:2.5)核转染100万个K562细胞。Cas9质粒结果与图1E所示相同。条形代表平均in/del频率+SEM,n=3。

[0036] 图10显示了原代人T细胞中的高RNA核转染效率。在刺激后三天,将来 自三个不同供体的受刺激的T细胞用GFP mRNA核转染。通过流式细胞术在 核转染三天后测量GFP的表达。显示核转染细胞(灰色)中的GFP表达相对于 模拟转染的细胞(黑色)。

[0037] 图11显示在T细胞核转染中增加CCR5 sgRNA和Cas9mRNA量产生了类 似的in/del频率。将受刺激的T细胞用标示量的MSP CCR5 sgRNA和Cas9 mRNA进行核转染。通过对跨越sgRNA靶位点的PCR扩增子的TIDE分析来测 量In/del频率,并使用模拟处理的样品作为参考对照。显示了平均in/del频率+ SEM,n=6。

[0038] 图12显示了CD4+、CD8+和总T细胞群体中的相似的in/del频率。将受刺 激的T细胞用CCR5 MSP sgRNA和Cas9 mRNA核转染,随后分选成CD4+和 CD8+亚群。In/del频率通过TIDE测量,并与大群体中的in/del频率进行比较。 条形代表一个T细胞供体的平均in/del频率+SEM,n=8。

[0039] 图13显示:原代人T细胞中的in/del频率随时间稳定。将受刺激的T细胞 用CCR5 MSP sgRNA和Cas9 mRNA核转染。在标示的时间点从一部分细胞中 提取gDNA,并通过跨越sgRNA靶位点的PCR扩增子的TIDE分析来测量in/del 频率,使用模拟处理的样品作为参考对照。显示了三个不同T细胞供体的平 均in/del频率+/-SEM,n=6。

[0040] 图14显示了与Cas9质粒相比,用合成的sgRNA和Cas9 mRNA核转染的T 细胞中较低的细胞死亡频率。将100万个受刺激的T细胞用10μg的标示的合成 sgRNA和15μg Cas9 mRNA

核转染,或用1μg Cas9编码质粒核转染。纳入1μg 编码CCR5 sgRNA和Cas9蛋白质的质粒用于比较(sgRNA质粒)。核转染三天 后,细胞用活/死(LIVE/DEAD)细胞染色进行染色。条形代表三个不同T细胞 供体的死细胞的平均百分比+SEM,n=6。

[0041] 图15显示了将合成的sgRNA核转染入T细胞后的增殖测定的结果。在第 0天将来自两个不同供体的受刺激的T细胞核转染,并且使用CellTiter Glo测 定来监测细胞增殖。为了清楚,没有标出重复实验的SEM,但是所有SEM均 小于所示值的15%。

[0042] 图16显示未刺激的T细胞中的CCR5破坏。来自三个不同供体的未受刺激 的人T细胞在分离当天用MS sgRNA和Cas9 mRNA进行核转染。核转染后3天 提取gDNA,并通过TIDE测量in/del频率,使用模拟处理的样品作为参考对照。条形代表平均值±SEM,n=2。

[0043] 图17显示了动员的PB CD34+HSPC中IL2RG和HBB的in/del频率。在 CD34+HSPC核转染后三天,提取基因组DNA,并通过T7测定法测量in/del频 率。分别显示了IL2RG和HBB的三个生物学重复的一个代表性凝胶。+SEM, n=2。

[0044] 图18A-B显示使用两种sgRNA在原代人T细胞和CD34+HSPC中CCR5基 因的高修饰频率。对于来自三个不同供体的受刺激的T细胞及PB动员的 CD34+HSPC,用'D'和'Q'sgRNA与Cas9 mRNA一式三份共同核转染。在核转 染后三天提取gDNA,并使用一对引物PCR扩增CCR5的修饰区,该对引物对 未修饰的等位基因可产生2.1kb的扩增子(图18A)。将PCR扩增子亚克隆到用 于转化的质粒中,对代表单个等位基因的单个菌落进行测序,与预期的基因 组序列参比,并根据等位基因型进行分类(图18B)。

[0045] 图19显示,在CD34+HSPC中,MS修饰的sgRNA比未修饰的sgRNA表现 更好。对于CD34+HSPC,用30μg Cas9蛋白与2.5倍摩尔过量的标示的合成 HBB sgRNA的复合物进行核转染。核转染后4天通过TIDE分析测量Indel频率。 条形代表三个不同的供体的平均indel频率+SEM,n=3。**p<0.01,Student's t检验。

[0046] 图20显示,修饰的sgRNA可以用于有效的多重化基因组编辑。将100万个 K562细胞用15μg Cas9 mRNA以及5μg CCR5、HBB或IL2RG MS修饰的sgRNA, 或全部三种sgRNA(多重化的)(3x 5μg)核转染。在核转染三天后,通过TIDE 分析测量三个基因座中每一个上的Indel频率。条形代表平均indel频率+SEM, n=3。***p<0.001,n.s.=p≥0.05,Student's t检验。

[0047] 图21显示了使用化学修饰的CCR5 sgRNA、Cas9 mRNA和CCR5 ssODN 进行同源重组后跨越CCR5靶位点的PCR产物。

具体实施方式

[0048] I.简介

[0049] 本文提供了用于在体外细胞(例如用于离体疗法的原代细胞)或体内细胞 (例如受试者诸如人的器官或组织中的细胞)中基于CRISPR/Cas的基因组编 辑和/或基因表达调变的方法。具体而言,本文提供的方法利用化学修饰的单 引导RNA(sgRNA),这些化学修饰的sgRNA与相应的未修饰的sgRNA相比, 在基因调控(例如,基因组编辑、基因表达的抑制、基因表达的激活)过程中 具有增强的活性。在某些方面,本发明提供对细胞中的靶核酸进行基因调控 的方法,其手段是导入与靶核酸杂交的化学修饰的sgRNA及:Cas核酸酶(例 如Cas9多肽)或其变体或片段;或编码Cas核酸酶(例如Cas9多肽)或其变体或 片段的mRNA;抑

或包含编码Cas核酸酶(例如Cas9多肽)或其变体或片段的核 苷酸序列的重组表达载体。在某些其他方面,本发明提供了预防或治疗受试 者的遗传疾病的方法,其手段是施用足够量的化学修饰的sgRNA以纠正与疾 病相关的基因突变。

[0050] II.一般说明

[0051] 除非另有说明,否则本发明的实施使用在本领域技术范围内的免疫学、生物化学、化学、分子生物学、微生物学、细胞生物学、基因组学和重组DNA 的常规技术。参见Sambrook,Fritsch and Maniatis,Molecular Cloning:A Laboratory Manual,2nd edition(1989),Current Protocols in Molecular Biology (F.M.Ausubel,et al.eds.,(1987)),the series Methods in Enzymology(Academic Press,Inc.):PCR 2:A Practical Approach(M.J.MacPherson,B.D.Hames and G.R.Taylor eds.(1995)),Harlow and Lane,eds.(1988)Antibodies,A Laboratory Manual,and Animal Cell Culture (R.I.Freshney,ed.(1987))。

[0052] 不可商购的寡核苷酸可以化学合成,例如,根据由Beaucage and Caruthers,Tetrahedron Lett.22:1859-1862(1981)首次描述的固相亚磷酰胺三酯方法。使 用自动合成仪,如Van Devanter et.al.,Nucleic Acids Res.12:6159-6168(1984) 中所述。寡核苷酸的纯化使用任何本领域公认的策略进行,例如Pearson and Reanier,J.Chrom.255:137-149(1983)。

[0053] III.定义

[0054] 除非特别指出,否则本文使用的所有技术和科学术语具有与本发明所属 领域的普通技术人员通常理解的相同的含义。另外,在本发明的实施中可以 使用与本文所述的方法或材料相似或等效的任何方法或材料。为了本发明的 目的,定义下列术语。

[0055] 如本文所使用的术语“一”或“该”不仅包括具有一个成员的方面,而且还 包括具有多于一个成员的方面。例如,单数形式“一”和“该”包括复数指称, 除非上下文另有明确说明。因此,例如,提述“一个细胞”包括多个这样的细 胞,并且对“作用剂”的提述包括对本领域技术人员已知的一种或多种作用剂 的提述,等等。

[0056] 术语“原代细胞”是指直接从多细胞生物体分离的细胞。与连续的(肿瘤或 人工永生的)细胞系相比,原代细胞通常经历过极少的群体倍增,因此更加能 够代表它们所来源的组织的主要功能成分。在一些情况下,原代细胞是已经 被分离并立即使用的细胞。在其他情况下,原代细胞不能无限分裂,因此不 能在体外长时间培养。

[0057] 术语“基因组编辑”是指使用一种或多种核酸酶和/或切口酶将DNA插入 靶DNA(例如细胞的基因组)、从靶DNA置换、或从靶DNA移除的基因工程类 型。核酸酶在基因组的期望位置产生特定的双链断裂(DSB),并利用细胞的 内源机制,通过同源性指导修复(HDR)(例如同源重组)或通过非同源末端连 接(NHEJ)来修复诱导的断裂。切口酶在基因组中的期望位置产生特定的单链 断裂。在一个非限制性实例中,可使用两个切口酶在靶DNA的两条相对链上 产生两个单链断裂,由此产生平末端或粘性末端。可以将任何合适的核酸酶 导入细胞以诱导靶DNA序列的基因组编辑,这样的核酸酶包括但不限于 CRISPR相关蛋白(Cas)核酸酶、锌指核酸酶(ZFN)、转录激活物样效应物核酸 酶(TALEN)、大范围核酸酶、其他内切或外切核酸酶、它们的变体、它们的 片段、及上述者的组合。在具体的实施方案中,可以利用本文所述的修饰的 单引导RNA(sgRNA)与Cas核酸酶(例如Cas9多肽或Cas9 mRNA)组合来

“诱导” 或“调变”(例如增强)核酸酶介导的靶核酸序列的基因组编辑,例如通过同源 性指导修复(HDR)来提高精确基因组编辑的效率。

[0058] 术语“同源性指导修复”或“HDR”是指细胞中一种准确、精密地修复双链 DNA断裂的机制,其使用同源模板来引导修复。HDR最常见的形式是同源重 组(HR),这是一种基因重组,其中在两个相似或相同的DNA分子之间交换核 苷酸序列。

[0059] 术语“非同源末端连接”或“NHEJ”是指修复双链DNA断裂的途径,其中 断裂末端直接连接而不需要同源模板。

[0060] 术语“核酸”、“核苷酸”或“多核苷酸”是指单链、双链或多链形式的脱氧 核糖核酸(DNA)、核糖核酸(RNA)及其聚合物。该术语包括但不限于单链、 双链或多链DNA或RNA、基因组DNA、cDNA、DNA-RNA杂合体、或包含 嘌呤和/或嘧啶碱基或其他天然的、化学修饰的、生物化学修饰的、非天然的、 合成的、或衍生化的核苷酸碱基的聚合物。在一些实施方案中,核酸可以包 含DNA、RNA及其类似物的混合物。除非特别限定,否则该术语涵盖这样的 核酸:它们含有天然核苷酸的已知类似物,具有与参照核酸相似的结合特性, 并以类似于天然存在的核苷酸的方式被代谢。除非另外指出,否则具体的核 酸序列也隐含地涵盖其保守修饰的变体(例如简并密码子取代),等位基因、 直向同源物、单核苷酸多态性(SNP)和互补序列以及明确指出的序列。具体 而言,简并密码子取代可以这样实现:产生这样的序列,其中一个或多个选 定(或全部)密码子的第三位置被混合碱基和/或脱氧肌苷残基取代(Batzer et al.,Nucleic Acid Res.19:5081(1991);Ohtsuka et al.,J.Biol.Chem.260:2605-2608(1985);和Rossolini et al.,Mol.Cell.Probes 8:91-98(1994))。术语“核酸” 可以与由基因编码的基因、cDNA和mRNA互换使用。

[0061] 术语“核苷酸类似物”或“修饰的核苷酸”是指这样的核苷酸,其在核苷的 含氮碱基中(例如胞嘧啶(C),胸腺嘧啶(T)或尿嘧啶(U)、腺嘌呤(A)、或鸟嘌 呤(G))、核苷的糖部分中(例如,核糖、脱氧核糖、修饰的核糖、修饰的脱氧 核糖、六元糖类似物、或开链糖类似物)、或磷酸上具有一个或多个化学修饰 (例如取代)。

[0062] 术语“基因”或“编码多肽的核苷酸序列”意指涉及多肽链的产生的DNA 区段。DNA片段可以包括位于编码区之前和之后、涉及基因产物的转录/翻译 和转录/翻译的调控的区域(前导区和尾随区),以及各个编码区段(外显子)之 间的间插序列(内含子))。

[0063] 术语“多肽”、“肽”和“蛋白质”在本文中可互换使用,是指氨基酸残基的 聚合物。这些术语适用于其中一个或多个氨基酸残基是相应天然存在的氨基 酸的人造化学模拟物的氨基酸聚合物,以及天然存在的氨基酸聚合物和非天 然存在的氨基酸聚合物。如本文所用的,这些术语涵盖任何长度的氨基酸链, 包括全长蛋白质,其中氨基酸残基通过共价肽键连接。

[0064] 术语“变体”是指生物体、菌株、基因、多核苷酸、多肽或特征的与自然 界中出现的形式不同的形式。

[0065] 术语“互补性”是指核酸与另一核酸序列通过传统的Watson-Crick型或其 他非传统类型形成氢键的能力。百分比互补性表示某一核酸分子中可与第二 个核酸序列形成氢键(例如Watson-Crick碱基配对)的残基数的百分比(例如、10个中有5、6、7、8、9、10个分别为50%、60%、70%、80%、90%、100% 互补)。“完全互补”是指某一核酸序列的所有连续残基将与第二个核酸序列中 相同数目的连续残基氢键键合。如本文所用,“基本上互补”是指

在8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、30个、35个、40个、45个、50个或更多个 核苷酸的区域上至少60％、65％、70％、75％、80％、85％、90％、95％、97％、98％、99％、或100％的互补程度，或者指在严格条件下杂交的两个核酸。

[0066] 用于杂交的术语“严格条件”是指这样的条件，在该条件下与靶序列具有 互补性的核酸绝大部分与靶序列杂交，并且基本上不与非靶序列杂交。严格 的条件通常是依赖于序列的，并且取决于许多因素而变化。通常，序列越长， 序列与其靶序列特异性杂交的温度越高。严格条件的非限制性实例详细描述 于Tijssen(1993)，Laboratory Techniques In Biochemistry And Molecular Biology- Hybridization With Nucleic Acid Probes，第一部分第二章“Overview of principles of hybridization and the strategy of nucleic acid probe assay”，Elsevier，NY

[0067] 术语“杂交”是指其中一个或多个多核苷酸反应形成通过核苷酸残基碱 基之间的氢键稳定化的复合物的反应。氢键键合可以通过Watson Crick碱基 配对，Hoogstein结合、或任何其它序列特异性方式发生。所述复合物可以包 含形成双链体结构的两条链、形成多链复合物的三条或更多条链、自我杂交 的单链、或它们的任何组合。

[0068] “重组表达载体”是重组产生或合成产生的核酸构建体，其具有允许特定 多核苷酸序列在宿主细胞中转录的一系列特定的核酸元件。表达载体可以是 质粒、病毒基因组、或核酸片段的一部分。通常，表达载体包括待转录的多 核苷酸及与之可操作地连接的启动子。“可操作地连接”在本文中是指两个或 更多个遗传元件，如多核苷酸编码序列和启动子，被放置在一定的相对位置， 从而允许所述元件正常地发挥生物功能，例如启动子引导编码序列的转录。术语“启动子”在本文中用于指引导核酸转录的一组核酸控制序列。如本文所 用的，启动子包括转录起始位点附近的必要核酸序列，例如就聚合酶II型启 动子而言是TATA元件。启动子还任选地包含远处的增强子或阻遏子元件，它 们可以位于距转录起始位点多达几千个碱基对处。其它可以存在于表达载体 中的元件包括可增强转录的元件(例如，增强子)和可终止转录的元件(例如， 终止子)，以及可赋予从该表达载体产生的重组蛋白一定的结合亲和力或抗 原性的元件。

[0069] “重组(的)”是指遗传修饰的多核苷酸、多肽、细胞、组织或生物体。例 如，重组多核苷酸(或重组多核苷酸的拷贝或互补物)是使用众所周知的方法 操作过的重组多核苷酸。包含与第二多核苷酸(例如，编码序列)可操作连接 的启动子的重组表达盒可以包括作为人为操作的结果相对于第二多核苷酸 为异源的启动子(例如，通过Sambrook等，Molecular Cloning-A Laboratory Manual，Cold Spring Harbor Laboratory，Cold Spring Harbor，New York，(1989) 或Current Protocols in Molecular Biology Volume 1-3，John Wiley&Sons，Inc。(1994-1998))。重组表达盒(或表达载体)通常包含天然不存在的多核苷酸组合。例如，人工操纵的限制性位点或质粒载体序列可以侧翼于启动子或者将启动 子与其他序列分开。重组蛋白质是由重组多核苷酸表达的蛋白质；重组细胞、重组组织和重组生物体是包含重组序列(多核苷酸和/或多肽)的细胞、组织和 生物体。

[0070] 术语“单核苷酸多态性”或“SNP”是指多核苷酸内(包括等位基因内)单个 核苷酸的改变。这可以包括一个核苷酸取代另一个核苷酸，以及单个核苷酸 的缺失或插入。最典型的是，SNP是双等位标记(biallelic markers)，尽管三等 位和四等位标记也可以存在。

举一非限制性实例,包含SNP A\C的核酸分子 在多态性位置上可以包含C或A。

[0071] 当提到细胞培养物本身或者细胞培养过程时,术语“培养”,“培育”、“维 持”,“扩增”等可以互换使用,以表示将细胞(例如,原代细胞)维持在其正 常环境之外的受控条件下,例如在适于存活的条件下。培养的细胞被允许存 活,且培养可导致细胞生长、停滞、分化或分裂。该术语并不暗示着培养中 的所有细胞都存活、成长或分裂,因为有些细胞会自然死亡或衰老。通常将 细胞培养在培养基中,培养基在培养过程中可以更换。

[0072] 术语“受试者”、“患者”、和“个体”在本文中可互换地使用,包括人或动 物。例如,动物受试者可以是哺乳动物、灵长类动物(例如猴),家畜(例如马, 牛,绵羊,猪或山羊),伴侣动物(例如狗,猫),实验室试验动物(例如小鼠, 大鼠,豚鼠,鸟),具有兽医意义的动物或具有经济意义的动物。

[0073] 如本文所用,术语“施用”包括口服施用、局部接触、作为栓剂施用、静 脉内施用、腹膜内施用、肌内施用、病灶内施用、鞘内施用、鼻内施用或皮 下施用给受试者。施用可通过任何途径,包括肠胃外和透粘膜(例如,口腔含 服、舌下、腭、齿龈、鼻、阴道、直肠或透皮)。肠胃外施用包括例如静脉内、 肌内、小动脉内、皮内、皮下、腹膜内、心室内和颅内。其他递送模式包括 但不限于使用脂质体制剂、静脉内输注、透皮贴剂等

[0074] 术语“治疗”是指获得有益的或期望的结果,包括但不限于治疗效益和/或 预防效益的方法。治疗效益是指被治疗一种或多种疾病、病症或状况的任何 治疗相关的改善或影响。为了获得预防效益,可将组合物施用给有发生特定 疾病、病症或状况的风险的受试者,或施用给报告有疾病的一种或多种生理 症状的受试者,即使疾病、病症或状况可能还没有表现出来。

[0075] 术语“有效量”或“足够量”是指作用剂(例如,Cas核酸酶,修饰的单引导 RNA等)的足以实现有益的或期望的结果的量。治疗有效量可根据以下一项 或多项而变化:正在治疗的受试者和疾病状况、受试者的体重和年龄、疾病 状况的严重程度、施用方式等,这些可容易地由本领域的普通技术人员确定。 具体的量可以随以下的一种或多种因素而变化:所选择的具体药剂、靶细胞 类型、受试者中靶细胞的位置、所遵循的给药方案、是否与其它作用剂联合 施用、施用时机、以及运载它的物理投递系统。

[0076] 术语“药学上可接受的载体”是指有助于向细胞、生物体或受试者施用作 用剂(例如,Cas核酸酶,修饰的单引导RNA等)的物质。“药学上可接受的载 体”指可包含在组合物或制剂中并且不会对患者造成显著的不利毒理学作用 的载体或赋形剂。药学上可接受的载体的非限制性实例包括水、NaCl、生理 盐水溶液、乳酸林格氏液、生理蔗糖(normal sucrose)、生理葡萄糖(normal glucose)、粘合剂、填充剂、崩解剂、润滑剂、包衣、甜味剂、香味剂和着色 剂等。本领域技术人员将认识到,其他药物载体可用于本发明。

[0077] 就CRISPR系统的组分而言,术语“增加稳定性”是指稳定CRISPR系统的 任何分子组分的结构的修饰。该术语包括减少、抑制、减小、或降低CRISPR 系统的任何分子组分的降解的修饰。

[0078] 就CRISPR系统的组分而言,术语“增加特异性”是指增加CRISPR系统的 任何分子组分的特异性活性(例如,中靶(on-target)活性)的修饰。该术语包括 减少、抑制、减小、或降低CRISPR系统的任何分子组分的非特异性活性(例 如,脱靶(off-target)活性)的修饰。

[0079] 就CRISPR系统的组分而言,术语“降低毒性”是指减少、抑制、减小、或 降低CRISPR

系统的任何分子组分对细胞、生物体、受试者等的毒性作用的修 饰。

[0080] 就CRISPR系统的组分而言，同时在基因调控的语境下，术语“增强的活 性”是指诱导、调变、调控或控制基因组编辑和/或基因表达的效率和/或频率 的增加或改善。

[0081] 与参考数值相关的术语“约”可以包括从该值加上或减去10％的值的范围。例如、“约10”的量包括从9到11的量、包括参考数字9、10和11在内。与参考 数值相关的术语“约”还可以包括从该值加上或减去10％、9％、8％、7％、6％、 5％、4％、3％、2％或1％的值的范围。

[0082] IV.实施方案的说明

[0083] 本发明提供诱导细胞中靶核酸的基因调控的方法。本发明包括使用这样 的修饰的单引导RNA(sgRNA)：其可增强原代细胞中(例如体外培养用于离体 疗法)、或者受试者诸如人体中的细胞中(例如用于体内治疗)靶核酸的基因组 编辑和/或抑制或激活靶核酸的基因表达。本发明还提供了通过增强精确的基 因组编辑以纠正与疾病相关的靶基因中的突变来预防或治疗受试者的疾病 的方法。本发明可以用于任何细胞类型以及任何适合于核酸酶介导的基因组 编辑技术的基因座位。

[0084] 在第一方面，本发明提供用于诱导(例如启动、调变、增强等)原代细胞 中靶核酸的基因调控的方法，所述方法包括：

[0085] 向原代细胞中导入：

[0086] (a)修饰的单引导RNA(sgRNA)，其包含与靶核酸互补的第一核苷酸序 列和与CRISPR相关蛋白(Cas)多肽相互作用的第二核苷酸序列，其中第一核 苷酸序列和/或第二核苷酸序列的一个或多个核苷酸是修饰的核苷酸；和

[0087] (b)Cas多肽、编码Cas多肽的mRNA、和/或包含编码Cas多肽的核苷酸序 列的重组表达载体，

[0088] 其中修饰的sgRNA将Cas多肽引导至靶核酸，

[0089] 其中修饰的sgRNA以相对于相应的未修饰sgRNA增强的活性诱导靶核 酸的基因调控。

[0090] 在一些实施方案中，所述增强的活性包括修饰的sgRNA的稳定性增加， 和/或修饰的sgRNA对靶核酸的特异性增加。

[0091] 在一些实施方案中，靶核酸包括靶DNA或靶RNA。靶核酸的基因调控涵 盖任何被细胞用来增加或减少靶核酸的特定基因产物(例如蛋白质或RNA) 的产生的机制，包括靶核酸的基因组编辑或靶核酸的基因表达的调变(例如， 抑制或活化)。在一些情况下，基因调控包括靶DNA的基因组编辑。基因组编 辑可以是目标DNA的同源性指导修复(HDR)或非同源末端连接(NHEJ)。在其 他情况下，基因调控包括使用核酸内切酶缺陷型Cas多肽调变(例如抑制或激 活)靶DNA或靶RNA的基因表达。

[0092] 在一些实施方案中，该方法进一步包括将重组供体修复模板导入到原代 细胞中。在某些情况下，重组供体修复模板包括两个核苷酸序列，它们包含 靶DNA的两个非重叠的同源部分，其中这些核苷酸序列位于对应于要进行基 因组编辑的靶DNA的核苷酸序列的5’和3’端。在其它情况下，重组供体修复 模板包含合成单链寡聚脱氧核苷酸(ssODN)模板，所述模板包含编码修饰单 核苷酸多态性(SNP)的突变的核苷酸序列和包含两个非重叠的同源部分的两 个核苷酸序列靶DNA，其中核苷酸序列位于编码突变的核苷酸序列的5’和3’

端。

[0093] 在一些实施方案中,在将修饰的sgRNA和Cas多肽导入原代细胞之前,从 多细胞生物中分离原代细胞。多细胞生物可以是植物、多细胞原生生物、多 细胞真菌或动物如哺乳动物(例如人)。在某些情况下,原代细胞选自干细胞、 免疫细胞及其组合。干细胞的非限制性实例包括造血干细胞和祖细胞(HSPC), 如CD34+HSPC、间充质干细胞、神经干细胞、器官干细胞及其组合。免疫 细胞的非限制性实例包括T细胞(例如CD3+T细胞,CD4+T细胞,CD8+T细 胞,肿瘤浸润细胞(TIL),记忆T细胞,记忆干T细胞,效应T细胞)、天然杀伤 细胞、单核细胞、外周血单个核细胞(PBMC)、外周血淋巴细胞(PBL)及其组 合。在其他实施方案中,在将修饰的sgRNA和Cas多肽导入原代细胞后,将原 代细胞或其后代(例如衍生自原代细胞的细胞)返回(例如通过任何可接受的 递送系统和递送途径施用给)多细胞生物体(例如人)。

[0094] 在一些实施方案中,原代细胞包含原代细胞群体。在一些情况下,修饰 的sgRNA在原代细胞群体的至少约5%,10%,15%,20%,25%,30%,35%, 40%,45%,50%,55%,60%,65%,70%,75%,80%,85%,90%或95% 中诱导出靶核酸的基因调控。在其他情况下,原代细胞群体包含至少约10、10^2、10^3、10^4、10^5、10^6、10^7、10^8或10^9个原代细胞。在某些情况下,基因调 控包括原代细胞群体中靶DNA的基因组编辑(例如HDR或NHEJ)。在某些其 他情况下,基因调控包括使用核酸内切酶缺陷型Cas多肽调变(例如抑制或激 活)原代细胞群体中靶DNA或靶RNA的基因表达。作为非限制性实例,在将 修饰的sgRNA与Cas多肽(例如,作为核糖核蛋白(RNP)复合物)或编码Cas多肽 的mRNA一道导入原代T细胞后,修饰的sgRNA可以原代T细胞群体的至少约 30%、35%、40%、45%、50%、55%或60%中诱导HDR(例如in/del频率)。作 为另一个非限制性实例,在将修饰的sgRNA与Cas多肽(例如,作为RNP复合物)或编码Cas多肽的mRNA一道导入原代造血干细胞和祖细胞(HSPC)之后, 修饰的sgRNA可以在原代HSPC群体的至少约10%、15%、20%、25%、30%、 35%、40%、45%、50%、55%或60%中诱导HDR(例如in/del频率)。

[0095] 在一些实施方案中,修饰的sgRNA的第一核苷酸序列和/或第二核苷酸序 列中的一个或多个修饰的核苷酸在核糖基团、磷酸基团、核碱基或其组合中 包含修饰。核糖基团中的修饰可以是在核糖基团的2'位处的修饰。在一些情 况下,核糖基的2'位的修饰下组:2'-O-甲基、2'-氟、2'-脱氧、2'-O-(2-甲氧基 乙基)。磷酸基团中的修饰可以是硫代磷酸酯修饰。

[0096] 在具体的实施方案中,修饰的sgRNA的第一核苷酸序列和/或第二核苷酸 序列中的一个或多个修饰的核苷酸包括:2'-O-甲基(M)核苷酸、2'-O-甲基3'- 硫代磷酸酯(MS)核苷酸、2'-O-甲基3'-硫代PACE(MSP)核苷酸、或其组合。 在一些情况下,修饰的sgRNA在第一核苷酸序列和/或第二核苷酸序列中包含 一个或多个MS核苷酸。在一些情况下,修饰的sgRNA在第一核苷酸序列和/ 或第二核苷酸序列中包含一个或多个MSP核苷酸。在一些情况下,修饰的 sgRNA在第一核苷酸序列和/或第二核苷酸序列中包含一个或多个MS核苷酸 和一个或多个MSP核苷酸。在一些情况下,修饰的sgRNA在第一核苷酸序列 和/或第二核苷酸序列中不包含M核苷酸。在一些情况下,修饰的sgRNA仅包 含MS核苷酸和/或MSP核苷酸作为第一核苷酸序列和/或第二核苷酸序列中 的修饰的核苷酸。在其他情况下,修饰的sgRNA在第一核苷酸序列和/或第二 核苷酸序列中包含一个或多个MS核苷酸和/或一个或多个MSP核苷酸,并且 可以在第一核苷酸序列和/或第二核苷酸序列中进一步包含一个或多个M个

核苷酸。

[0097] 在一些实施方案中，修饰的sgRNA的第一核苷酸序列长约20个核苷酸。在一些情况下，第一核苷酸序列中的至少2、3、4、5、6、7、8、9、10个或 更多个核苷酸是修饰的核苷酸。在某些情况下，第一核苷酸序列(例如，约20 个核苷酸长的第一核苷酸序列)中约2、3、4、5、6、7、8、9、10、11、12、 13、14、15、16、17、18、19或20个核苷酸是修饰的核苷酸。在其他情况下， 第一核苷酸序列(例如，约20个核苷酸长的第一核苷酸序列)中的所有核苷酸 均是修饰的核苷酸。在一些情况下，修饰的核苷酸位于第一核苷酸序列的5’ 末端(例如5’末端的末端核苷酸)或5’末端附近(例如距5’末端1、2、3、4或5个 核苷酸以内的核苷酸)和/或第一核苷酸序列内的内部位置。在其他情况下， 第一核苷酸序列中约10％至约30％的核苷酸是修饰的核苷酸。

[0098] 在一些实施方案中，修饰的sgRNA的第二核苷酸序列的长度为约80个核 苷酸。在一些情况下，第二核苷酸序列中的至少2、3、4、5、6、7、8、9、 10个或更多个核苷酸是修饰的核苷酸。在某些情况下，第二核苷酸序列(例如， 约80个核苷酸长的第二核苷酸序列)中的大约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、50、51、52、53、54、55、56、57、58、59、60、61、 62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、 79或80个核苷酸是修饰的核苷酸。在其它情况下，第二核苷酸序列(例如约80 个核苷酸长的第二核苷酸序列)中的所有核苷酸均是修饰的核苷酸。在一些 情况下，修饰的核苷酸位于第二核苷酸序列的3’末端(例如，3’末端的末端核 苷酸)或3’末端附近(例如距3’末端1、2、3、4或5个核苷酸以内)和/或第二核苷 酸序列内的内部位置。在其他情况下，第二核苷酸序列中约1％至约10％的核 苷酸是修饰的核苷酸。

[0099] 在某些实施方案中，修饰的sgRNA自第一核苷酸序列的5’端(例如5’末端 的末端核苷酸)或5’端附近(例如，距5’-末端的末端核苷酸1、2、3、4或5个核 苷酸内)开始包含一个，两个或三个连续或非连续的修饰核苷酸，并且自第二 核苷酸序列的3’末端(例如3’末端的末端核苷酸)或3’末端附近(例如距3’末端1、2、3、4或5个核苷酸内)开始包含一个，两个或三个连续或非连续的修饰的核 苷酸。

[0100] 在一些情况下，修饰的sgRNA在第一核苷酸序列的5’末端(例如5’末端的 末端核苷酸)或5’末端附近(例如距5’末端1、2、3、4或5个核苷酸内)包含一个 修饰的核苷酸，并在第二核苷酸序列的3’末端(例如3’末端的末端核苷酸)或3’ 末端附近(例如距3’末端1、2、3、4或5个核苷酸内)包含一个修饰的核苷酸。

[0101] 在其它情况下，在其它情况下，修饰的sgRNA在第一核苷酸序列的5’末 端(例如5’末端的末端核苷酸)或5’末端附近(例如距5’末端1、2、3、4或5个核 苷酸内)包含两个连续的或非连续的修饰的核苷酸，并在第二核苷酸序列的3’ 末端(例如3’末端的末端核苷酸)或3’末端附近(例如距3’末端1、2、3、4或5个 核苷酸内)包含两个连续的或非连续的修饰的核苷酸。。

[0102] 在又一些情况下，在其它情况下，修饰的sgRNA在第一核苷酸序列的5’ 末端(例如5’末端的末端核苷酸)或5’末端附近(例如距5’末端1、2、3、4或5个 核苷酸内)包含三个连续的或非连续的修饰的核苷酸，并在第二核苷酸序列 的3’末端(例如3’末端的末端核苷酸)或3’末端附近(例如距3’末端1、2、3、4或 5个核苷酸内)包含三个连续的或非连续的修

饰的核苷酸。

[0103] 在具体的实施方案中,修饰的sgRNA在第一核苷酸序列的5'-末端包含三 个连续的修饰的核苷酸,在第二核苷酸序列的3'-末端包含三个连续的修饰的 核苷酸。

[0104] 在一些实施方案中,修饰的sgRNA是化学合成的。在其他实施方案中, 用于诱导基因调控的方法包括使用多个修饰的sgRNA对单个靶核酸序列或 不同靶核酸序列进行多重化基因调控(例如,基因组编辑或调变基因表达)。 在具体的实施方案中,相对于使用相应的未修饰的sgRNA,多重化基因调控 更有效和/或一致。在某些情况下,所述多个修饰的sgRNA包含至少2、3、4、 5、10、15、20个,或更多个不同的修饰的sgRNA,其中每个修饰的sgRNA针 对不同的靶核酸。在其它情况下,所述多个修饰的sgRNA包含至少2、3、4、 5、10、15、20个,或更多个不同的修饰的sgRNA,其中每个修饰的sgRNA针 对相同的靶核酸。

[0105] 在某些实施方案中,Cas多肽是Cas多肽变体或Cas多肽片段。在具体的 实施方案中,Cas多肽是Cas9多肽、其变体或其片段。在其他实施方案中,导 入原代细胞的步骤包括电穿孔(例如,通过核转染)原代细胞。

[0106] 在第二方面,本发明提供了用于预防或治疗受试者中的遗传疾病的方法, 所述方法包括:

[0107] 向受试者施用足够量的修饰的单引导RNA(sgRNA)以纠正与遗传疾病 相关的靶基因中的突变,其中修饰的sgRNA包含与靶基因互补的第一核苷酸 序列和与CRISPR相关蛋白(Cas)多肽相互作用的第二核苷酸序列,并且其中 第一核苷酸序列和/或第二核苷酸序列中的一个或多个核苷酸是修饰的核苷 酸。

[0108] 在一些实施方案中、所述遗传疾病选自下组:X连锁严重联合免疫缺陷、 镰状细胞性贫血、地中海贫血、血友病、瘤形成、癌症、年龄相关性黄斑变 性、精神分裂症、三核苷酸重复疾病、X染色体易裂症、朊病毒相关病症、 肌萎缩性侧索硬化、药物成瘾、自闭症、阿尔茨海默病、帕金森病、囊性纤 维化、血液和凝血疾病或病症、炎症、免疫相关疾病或病症、代谢疾病、肝 脏疾病和病症、肾脏疾病和病症、肌肉/骨骼疾病和病症、神经和神经元疾病 和病症、心血管疾病和病症、肺部疾病和病症、眼部疾病和病症以及病毒感 染(例如HIV感染)。

[0109] 在一些实施方案中,所述方法进一步包括向受试者施用Cas多肽、编码 Cas多肽的mRNA、和/或包含编码Cas多肽的核苷酸序列的重组表达载体。

[0110] 在一些实施方案中,该方法进一步包括向受试者施用重组供体修复模板。在某些情况下,重组供体修复模板包含两个核苷酸序列,所述核苷酸序列包 含靶基因的两个非重叠的同源部分,其中所述核苷酸序列位于对应于要进行 基因组编辑的靶基因的核苷酸序列的5'和3'端。在其他情况下,重组供体修复 模板包含合成的单链寡聚脱氧核苷酸(ssODN)模板,所述模板包含:编码用 于矫正靶基因中的单核苷酸多态性(SNP)的突变的核苷酸序列,以及包含靶 基因的两个非重叠同源部分的两个核苷酸序列,其中这两个核苷酸序列位于 前述编码突变的核苷酸序列的5'和3'端。

[0111] 在某些实施方案中,与施用相应的未修饰的sgRNA相比,施用修饰的 sgRNA增强Cas多肽矫正靶基因中的突变的作用。上文描述了与用于在受试 者中预防或治疗遗传疾病的方法中使用的修饰的sgRNA有关的非限制性实 施方案。

[0112] 在一些实施方案中,将修饰的sgRNA、Cas多肽和/或重组供体修复模板 与药学上

可接受的载体一起施用给受试者。

[0113] 在一些实施方案中,通过选自下组的递送系统将修饰的sgRNA、Cas多肽 和/或重组供体修复模板给予受试者:纳米颗粒、脂质体、胶束、病毒微体 (virosome)、核酸复合物、以及它们的组合。在某些情况下,核酸复合物包含 与Cas多肽复合的修饰的sgRNA。

[0114] 在一些实施方案中、通过选自下组的递送途径将修饰的sgRNA、Cas多肽 和/或重组供体修复模板给予受试者:口服、静脉内、腹膜内、肌肉内、皮内、 皮下、肌内、皮内、肌内、皮下、肌肉内、小动脉、脑室内、颅内、病灶内、 鞘内、局部、经粘膜、鼻内及其组合。

[0115] 预期本文所述的化学修饰的sgRNA可与任何CRISPR相关技术一起使用, 这些技术包括但不限于:用于抑制基因表达的CRISPRi、用于激活基因表达 的CRISPRa、用于基因组座位动态可视化的CRISPR成像工具、以及CRISPR 介导的RNA识别和切割。在一些情况下,sgRNA可用于将小分子或蛋白质成 像和/或递送到体外细胞或体内细胞中。因此,sgRNA可以用于研究和治疗应 用。

[0116] A.CRISPR/Cas系统

[0117] 基因组修饰的CRISPR/Cas系统包括:Cas核酸酶(例如,Cas9核酸酶)或其 变体或片段;靶向DNA的RNA(例如修饰的sgRNA),其含有将Cas9核酸酶靶 向到靶基因组DNA的引导序列;以及与Cas核酸酶相互作用的支架序列(例如 tracrRNA),以及任选的供体修复模板。在一些情况下,可以使用包含一个或 多个以下突变的Cas核酸酶变体,例如Cas9突变体:D10A、H840A、D839A 和H863A;或者可以使用Cas9切口酶。在其他情况下,可以使用具有期望特 性(例如,能够产生单链或双链断裂和/或调变基因表达)的Cas核酸酶或其变 体的片段。供体修复模板可以包括编码报道多肽(例如荧光蛋白或抗生素抗 性标记)的核苷酸序列,以及与靶DNA同源且位于基因修饰位点侧翼的同源 臂。或者,供体修复模板可以是单链寡聚脱氧核苷酸(ssODN)。

[0118] 1.Cas核酸酶及其变体

[0119] CRISPR(成簇规则间隔短回文重复)/Cas(CRISPR相关蛋白)核酸酶系统 是一种基于细菌系统的工程化核酸酶系统,可用于基因组工程。它是以许多 细菌和古菌的适应性免疫应答一部分为基础的。当病毒或质粒侵入细菌时, 侵入者的DNA的区段被“免疫”应答转化为CRISPR RNA(crRNA)。然后, crRNA通过具有部分互补性的区域与另一种类型的称为tracrRNA的RNA缔 合,以将Cas(例如Cas9)核酸酶引导至靶DNA中与crRNA同源的区域(称为“原 型间隔子”(protospacer))。Cas(例如Cas9)核酸酶切割DNA以在双链断裂处产 生平端,切割的部位由crRNA转录物内包含的20个核苷酸的引导序列所规定。Cas(例如Cas9)核酸酶的位点特异性DNA识别和切割需要crRNA和tracrRNA 二者。这个系统现在已经被改造,使得crRNA和tracrRNA可以组合成一个分 子(“单引导RNA”或“sgRNA”),并且单引导RNA的crRNA等价部分可以被改 造,从而可以引导Cas(例如Cas9)核酸酶靶向任何期望的序列(参见例如Jinek et al.(2012)Science,337:816-821;Jinek et al.(2013)eLife,2:e00471;Segal(2013) eLife,2:e00563))。因此,CRISPR/Cas系统可以被改造成在细胞基因组中的期 望靶标处产生双链断裂,并利用细胞的内源性机制通过同源性指导修复 (HDR)或非同源末端连接(NHEJ)来修复诱导的断裂。

[0120] 在一些实施方案中,所述Cas核酸酶具有DNA切割活性。Cas核酸酶可以 指导在靶DNA序列中的某个位置的切割一条或两条链。例如,Cas核酸酶可 以是具有一个或多个失活

的催化结构域的切口酶，其切割靶DNA序列的单链。

[0121] Cas核酸酶的非限制性实例包括Cas1、Cas1B、Cas2、Cas3、Cas4、Cas5、Cas6、Cas7、Cas8、Cas9(又称为Csn1和Csx12)、Cas10、Csy1、Csy2、Csy3、Cse1、Cse2、Csc1、Csc2、Csa5、Csn2、Csm2、Csm3、Csm4、Csm5、Csm6、Cmr1、Cmr3、Cmr4、Cmr5、Cmr6、Csb1、Csb2、Csb3、Csx17、Csx14、Csx10、Csx16、CsaX、Csx3、Csx1、Csx15、Csf1、Csf2、Csf3、Csf4、其同源物，其变体，其片段，其突变体、及其衍生物。有三种主要类型的Cas核酸酶(I型、II型、III型)和10种亚型，包括5种I型蛋白、3种II型蛋白、和2种III型蛋白(参见例如Hochstrasser and Doudna,Trends Biochem Sci,2015:40(1):58-66)。II型Cas核酸酶包括Cas1、Cas2、Csn2和Cas9。这些Cas核酸酶是本领域技术人员已知的。例如，酿脓链球菌野生型Cas9多肽的氨基酸序列在例如NBCI Ref.Seq.No.NP_269215中列出，而嗜热链球菌野生型Cas9多肽的氨基酸序列在例如NBCI Ref.Seq.No.WP_011681470中列出。可用于本发明的CRISPR相关核酸内切酶在例如美国申请公开号2014/0068797、2014/0302563和2014/0356959中公开。

[0122] Cas核酸酶，例如，Cas9多肽，可以衍生自多种细菌物种，包括但不限于：非典型韦荣氏球菌(Veillonella atypical)，具核梭杆菌(Fusobacterium nucleatum)，龈沟产线菌(Filifactor alocis)，Solobacterium moorei，灵巧粪球菌(Coprococcus catus)，齿垢密螺旋体(Treponema denticola)，Peptoniphilus duerdenii，Catenibacterium mitsuokai，变异链球菌(Streptococcus mutans)，无毒李斯特菌(Listeria innocua)，伪中间葡萄球菌(Staphylococcus pseudintermedius)，肠氨基酸球菌(Acidaminococcus intestine)，树胶欧尔森氏菌(Olsenella uli)，北原氏酒球菌(Oenococcus kitaharae)，双歧双歧杆菌(Bifidobacterium bifidum)，鼠李糖乳杆菌(Lactobacillus rhamnosus)，格氏乳杆菌(Lactobacillus gasseri)，大芬戈尔德菌(Finegoldia magna)，运动支原体(Mycoplasma mobile)，鸡毒支原体(Mycoplasma gallisepticum)，绵羊肺炎支原体(Mycoplasma ovipneumoniae)，犬支原体(Mycoplasma canis)，滑液支原体(Mycoplasma synoviae)，直肠真杆菌(Eubacterium rectale)，嗜热链球菌(Streptococcus thermophilus)，细长真杆菌(Eubacterium dolichum)，棒状乳杆菌扭曲亚种(Lactobacillus coryniformis subsp.Torquens)，多养型泥杆菌(Ilyobacter polytropus)，白色瘤胃球菌(Ruminococcus albus)，Akkermansia muciniphila，解纤维热酸菌(Acidothermus cellulolyticus)，长双歧杆菌(Bifidobacterium longum)，齿双歧杆菌(Bifidobacterium dentium)，白喉棒状杆菌(Corynebacterium diphtheria)，Elusimicrobium minutum，卤水硝酸盐裂解菌(Nitratifractor salsuginis)，Sphaerochaeta globus，产琥珀酸丝状杆菌产琥珀酸亚种(Fibrobacter succinogenes subsp.Succinogenes)，脆弱拟杆菌(Bacteroides fragilis)，黄褐二氧化碳嗜纤维菌(Capnocytophaga ochracea)，沼泽红假单胞菌(Rhodopseudomonas palustris)，彩虹普雷沃菌(Prevotella micans)，栖瘤胃普雷沃菌(Prevotella ruminicola)，柱状黄杆菌(Flavobacterium columnare)，少食氨基单胞菌(Aminomonas paucivorans)，深红红螺菌(Rhodospirillum rubrum)，Candidatus Puniceispirillum marinum，Verminephrobacter eiseniae，蒲桃雷尔氏菌(Ralstonia syzygii)，恒雄芝氏沟鞭藻玫瑰杆菌(Dinoroseobacter shibae)，固氮螺菌属(Azospirillum)，汉堡硝化杆菌(Nitrobacter hamburgensis)，慢生根瘤菌属(Bradyrhizobium)，产琥珀酸沃林氏菌

(Wolinella succinogenes),空肠弯 曲杆菌空肠亚种(Campylobacter jejuni subsp.Jejuni),伶鼬鼠螺杆菌 (Helicobacter mustelae),蜡样芽孢杆菌(Bacillus cereus),Acidovorax ebreus, 产气荚膜梭菌(Clostridium perfringens),食清洁剂细小棒菌(Parvibaculum lavamentivorans),肠道罗斯拜瑞氏菌(Roseburia intestinalis),脑膜炎奈瑟氏菌 (Neisseria meningitidis),多杀巴斯德菌多杀亚种(Pasteurella multocida subsp. Multocida),华德萨特菌(Sutterella wadsworthensis),变形菌(proteobacterium), 嗜肺军团菌(Legionella pneumophila),Parasutterella excrementihominis,产琥 珀酸沃廉氏菌(Wolinella succinogenes),和新泽西弗朗西斯菌(Francisella novicida)。

[0123] “Cas9”是指RNA引导的、结合双链DNA的核酸酶蛋白或切口酶蛋白。野 生型Cas9核酸酶具有切割不同DNA链的两个功能结构域,例如RuvC和HNH。当两个功能域都有活性时,Cas9可以诱导基因组DNA(靶DNA)中的双链断裂。所述Cas9酶可包含来自属于下组的细菌的Cas9蛋白的一个或多个催化结构 域:棒状杆菌属(Corynebacter),萨特菌属(Sutterella),军团菌属(Legionella), 密螺旋体属(Treponema),产线菌属(Filifactor),真杆菌属(Eubacterium),链球 菌属(Streptococcus),乳杆菌属(Lactobacillus),支原体属(Mycoplasma),拟杆 菌属(Bacteroides),Flaviivola属,黄杆菌属(Flavobacterium),Sphaerochaeta属, 固氮螺菌属(Azospirillum),葡糖醋杆菌属(Gluconacetobacter),奈瑟氏菌属 (Neisseria),罗斯拜瑞氏菌属(Roseburia),细小棒菌属(Parvibaculum),葡萄球 菌属(Staphylococcus),硝酸盐裂解菌属(Nitratifractor)、和弯曲杆菌属 (Campylobacter)。在一些实施方案中,两个催化结构域来源于不同的细菌物 种。

[0124] Cas9核酸酶的有用变体可以包括单个失活的催化结构域,如$RuvC^-$或 HNH^-酶或切口酶。Cas9切口酶仅具有一个活性功能结构域,仅能够切割靶 DNA的一条链,由此产生单链断裂或缺口(nick)。在一些实施方案中,具有至 少D10A突变的突变Cas9核酸酶是Cas9切口酶。在其他实施方案中,具有至少 H840A突变的突变Cas9核酸酶是Cas9切口酶。其他存在于Cas9切口酶中的突 变的例子包括但不限于N854A和N863A。如果使用至少两个靶向DNA的RNA, 这些RNA靶向相反的DNA链,则可以使用Cas9切口酶导入双链断裂。双切口 诱导的双链断裂(double-nicked induced double-strand break)可以通过NHEJ或 HDR被修复(Ran等,2013,Cell,154:1380-1389)。这种基因编辑策略有利于 HDR,并降低脱靶DNA位点上的indel突变的频率。Cas9核酸酶或切口酶的非 限制性实例描述于例如美国专利8,895,308、8889418、8,865,406,以及美国申 请公开号2014/0356959、2014/0273226、2014/0186919。Cas9核酸酶或切口酶 可针对靶细胞或靶生物体进行密码子优化。

[0125] 在一些实施方案中,所述Cas核酸酶可以是这样的Cas9多肽,其包含RuvC1和HNH核酸酶结构域的两个沉默突变(D10A和H840A),该多肽被称为 dCas9(Jinek et al,Science,2012,337:816-821;Qi et al,Cell,152(5):1173-1183)。 在一个实施方案中,来自酿脓链球菌的dCas9多肽在位置D10、G12、G17、 E762、H840、N854、N863、H982、H983、A984、D986、A987或上述位置 的任何组合上包含至少一个突变。这种dCas9多肽及其变体在例如在国际专 利公开号WO 2013/176772中有描述。dCas9酶可以在D10,E762,H983或D986 处含有突变,以及在H840或N863处有突变。在一些情况下,dCas9酶含有 D10A或D10N突变。而且,

dCas9酶可以包括H840A、H840Y或H840N。在一 些实施方案中，本发明的dCas9酶包含下述取代：D10A和H840A；D10A和 H840Y；D10A和H840N；D10N和H840A；D10N和H840Y；或D10N和H840N。取代可以是使得Cas9多肽无催化活性并能够与靶DNA结合的保守取代或非 保守取代。

[0126] 在某些实施方案中，dCas9多肽是无催化活性的，例如是核酸酶活性缺 陷的。在一些情况下，dCas9酶或其变体或片段可以阻断靶序列的转录，并且 在一些情况下阻断RNA聚合酶。在其他情况下，dCas9酶或其变体或片段可 以激活靶序列的转录。

[0127] 对于基因组编辑方法，Cas核酸酶可以是Cas9融合蛋白，例如这样的多 肽：其包含IIS型限制酶的催化结构域FokI及与之连接的dCas9。FokI-dCas9融 合蛋白(fCas9)可以使用两个结合到靶DNA的一条单链的引导RNA，以产生双 链断裂。

[0128] 对于基因调控(例如，调变靶DNA的转录)，核酸酶缺陷型Cas蛋白(例如 但不限于dCas9)可用于转录激活或转录抑制。使用无核酸酶的Cas蛋白使基 因表达失活的方法描述于例如Larson et al.Nat.Protoc.,2013,8(11):2180-2196。

[0129] 在一些实施方案中，编码Cas核酸酶的核苷酸序列存在于重组表达载体 中。在某些情况下、重组表达载体是病毒构建体、例如重组腺伴随病毒构建 体、重组腺病毒构建体、重组慢病毒构建体等。例如，病毒载体可以基于痘 苗病毒、脊髓灰质炎病毒、腺病毒、腺伴随病毒、SV40、单纯疱疹病毒、人 免疫缺陷病毒等。逆转录病毒载体可以基于鼠白血病病毒、脾脏坏死病毒、 以及来自逆转录病毒的载体，例如劳氏肉瘤病毒、哈维肉瘤病毒、禽白血病 病毒、慢病毒、人免疫缺陷病毒、骨髓增殖性肉瘤病毒、乳腺肿瘤病毒，等 等。有用的表达载体是本领域技术人员已知的，并且许多是可商购的。作为 示例提供了以下用于真核宿主细胞的载体：pXT1、pSG5、pSVK3、pBPV、 pMSG和pSVLSV40。但是，任何其他载体如果与宿主细胞兼容均可以使用。 例如，含有编码Cas9酶的核苷酸序列的有用表达载体可从例如Addgene、Life Technologies、Sigma-Aldrich和Origene商购获得。

[0130] 取决于使用的靶细胞/表达系统，可以在表达载体中使用许多转录和翻译 控制元件中的任何一种，包括启动子、转录增强子、转录终止子等。有用的 启动子可以来自病毒或任何生物体，例如原核生物体或真核生物体。合适的 启动子包括但不限于：SV40早期启动子；小鼠乳房肿瘤病毒长末端重复序列 (LTR)启动子；腺病毒主要晚期启动子(Ad MLP)；单纯疱疹病毒(HSV)启动子； 巨细胞病毒(CMV)启动子如CMV立即早期启动子区(CMVIE)；劳斯肉瘤病毒 (RSV)启动子；人U6小核启动子(U6)；增强型U6启动子；人H1启动子(H1)等 等。

[0131] 可以将Cas核酸酶及其变体或片段以下述形式导入细胞(例如体外细胞， 例如用于离体疗法的原代细胞；或体内细胞，例如患者体内的细胞)：Cas多 肽或其变体或片段、编码Cas多肽或其变体或片段的mRNA、或包含编码Cas 多肽或其变体或片段的核苷酸序列的重组表达载体。

[0132] 2.修饰的单引导RNA(sgRNA)

[0133] 供CRISPR/Cas基因组修饰系统中使用的修饰的sgRNA通常包括：与靶核 酸序列互补的引导序列(例如crRNA)，以及与Cas核酸酶(例如Cas9多肽)或其 变体或片段序列相互作用的支架序列(例如tracrRNA)。本发明人已经发现， 包含一种或多种化学修饰的修饰的sgRNA当在原代细胞(例如T细胞或造血 干细胞和祖细胞)中用于基于CRISPR的基因调控(例如基因组编辑或调变基 因表达)时，与相应的未修饰的sgRNA相比，可以增加修饰的

sgRNA的活性、稳定性和特异性,和/或减少修饰的sgRNA的毒性。与现有技术相比,修饰的sgRNA的优点包括但不限于:更容易递送到靶细胞例如原代细胞中、以及稳 定性增加、活性持续时间增加、和靶细胞的毒性减少。在某些情况下,与其 他系统相比,作为CRISPR/Cas系统的一部分的修饰的sgRNA提供更高的中靶 (on-target)基因调控频率。在其他情况下,与其未修饰的序列等同物相比,修 饰的sgRNA提供改善的活性和/或特异性。

[0134] 在某些情况下,修饰的sgRNA与Cas核酸酶(例如Cas9多肽)或其变体或片 段复合以形成基于核糖核蛋白(RNP)的递送系统,用于导入细胞(例如体外细 胞诸如用于体外治疗的原代细胞或诸如患者体内细胞)。在其他情况下,将修 饰的sgRNA与编码Cas核酸酶的mRNA(例如,编码一种或多种核酸酶的 mRNA)一起导入细胞(例如体外细胞,例如离体疗法的原代细胞,或体内细 胞,例如患者体内)Cas9多肽)或其变体或片段。在其它情况下,将修饰的 sgRNA用包含核苷酸序列的重组表达载体导入细胞(例如体外细胞,例如离 体疗法的原代细胞,或体内细胞,例如患者体内)编码Cas核酸酶(例如Cas9多 肽)或其变体或片段。

[0135] 在一些情况下,可以使用多个修饰的sgRNA以实现靶细胞(例如原代细 胞)中的多重化的基于CRISPR的基因调控(例如,基因组编辑或调控基因表 达)。所述多个修饰的sgRNA可以包括至少约2、3、4、5、6、7、8、9、10、 15、20、25、30、35、40、45、50个或更多个修饰的sgRNA,它们可与相同 的靶核酸序列或不同的靶核酸序列杂交。所述多个修饰的sgRNA可以作为与 Cas核酸酶(例如Cas9)或其变体或片段的复合物、或者作为编码Cas核酸酶(例如Cas9多肽)或其变体或片段的核苷酸序列(例如,mRNA或重组表达载体)导 入细胞(例如体外细胞,例如离体疗法的原代细胞;或体内细胞,例如患者体 内的细胞)。

[0136] 修饰的sgRNA的核酸序列可以是任何这样的多核苷酸序列:其与靶多核 苷酸序列(例如靶DNA序列)具有足够互补性,从而与靶序列杂交并指导 CRISPR复合物与靶序列序列特异性地结合。在一些实施方案中,修饰的 sgRNA的引导序列及其相应的靶序列之间的互补程度,当任选地将二者使用 合适的比对算法实现最佳比对时,约为或大于约50%、60%、75%、80%、85%、 90%、95%、97.5%、99%或更多。最佳比对可通过使用任何合适的序列比对 算法来确定,其非限制性的例子包括:Smith-Waterman算法、Needleman- Wunsch算法、基于Burrows-Wheeler变换的算法(例如Burrows Wheeler比对器)、ClustalW、Clustal X、BLAT、Novoalign(Novocraft Technologies,ELAND (Illumina,San Diego,Calif.)、SOAP(可在soap.genomics.org.cn访问)、以及Maq (可在maq.sourceforge.net访问)。在一些实施例中,引导序列的长度为约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、35、40、45、50、75或更多个核苷 酸。在一些情况下,引导序列的长度为约20个核苷酸。在其它情况下,引导 序列的长度为约15个核苷酸。在其它情况下,引导序列的长度为约25个核苷 酸。可以通过任何合适的测定法来评估一个引导序列指导CRISPR复合物序 列特异性地结合靶序列的能力。例如,可以将足以形成CRISPR复合物的 CRISPR系统组分,包括要测试的引导序列在内,提供给具有相应的靶序列的 宿主细胞(例如通过用编码CRISPR序列的组分的载体进行转染),随后评估靶 序列内的优先切割。类似地,可以如下所述在试管中评估靶多核苷酸序列的 切割:提供靶序列、CRISPR复合物的组分(包括要测试的引导序列和与测试 引导序列不同的对照引导序列),并比较测试引导序列和对照引导序列反应 在靶序列处的结合或切割速率。

[0137] 可以使用上述任何基于网络的软件来选择修饰的sgRNA的核苷酸序列。 选择DNA

靶向性RNA的考虑因素包括:要使用的CAS核酸酶(例如Cas9多肽) 的PAM序列,以及最小化脱靶性修饰的策略。工具,如CRISPR设计工具,可 提供用于制备修饰的sgRNA的序列、用于评估靶修饰效率序列、和/或用于评 估脱靶位点处的切割的序列。选择修饰的sgRNA的序列的另一个考虑因素包 括降低引导序列内的二级结构的程度。二级结构可以通过任何合适的多核苷 酸折叠算法确定。一些程序是基于计算最小吉布斯自由能。合适的算法的例 子包括MFOLD(Zuker and Stiegler,Nucleic Acids Res,9(1981),133-148)、UNAFOLD工具包(Markham et al.,Methods Mol Biol,2008,453:3-31)、和 ViennaRNa工具包中的RNAfold。

[0138] 修饰的sgRNA的引导序列中的一个或多个核苷酸和/或支架序列中的一 个或多个核苷酸可以是修饰的核苷酸。例如,长度为约20个核苷酸的引导序 列可以具有1个或多个,例如1,2,3,4,5,6,7,8,9,10,11,12,13, 14,15,16,17,18,19,20个或更多个修饰的核苷酸。在一些情况下,引 导序列包括至少2,3,4,5,6,7,8,9,10或更多个修饰的核苷酸。在其他情况下,引导序列包括至少2,3,4,5,6,7,8,9,10,11,12,13, 14,15,16,17,19,20,或更多个修饰的核苷酸。修饰的核苷酸可以位于 引导序列的任何核酸位置。换句话说,修饰的核苷酸可以在引导序列的第一 个和/或最后一个核苷酸处或附近,和/或二者之间的任何位置。例如,对于 长度为20个核苷酸的引导序列,该一个或多个经修饰的核苷酸可以位于核酸 第1位,第2位,第3位,第4位,第5位,第6位,第7位,第8位,第9位,第10 位,第11位,第12位,第13位,第14位,第15位,第16位,第17位,第18位, 第19位和/或引导序列的第20位。在某些情况下,引导序列的约10%至约30%, 例如,约10%至约25%,约10%至约20%,约10%至约15%,约15%至约30%, 约20%至约30%,或约25%至约30%可以包含修饰的核苷酸。在其他情况下, 引导序列的约10%至约30%,例如约10%,约11%,约12%,约13%,约14%, 约15%,约16%,约17%,约18%,约19%,约20%,约21%,约22%,约23%, 约24%,约25%,约26%,约27%,约28%,约29%,或约30%可以包含修饰 的核苷酸。

[0139] 在一些实施方案中,修饰的sgRNA的支架序列含有一个或多个修饰的核 苷酸。例如,长度为约80个核苷酸的支架序列可以具有1个或多个,例如1, 2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20, 21,22,23,24,25,26,27,28,29,30,35,40,45,50,55,60,65, 70,75,76,77,78,79,80个或更多个修饰的核苷酸。在一些情况下,支 架序列包括至少2,3,4,5,6,7,8,9,10个或更多个修饰的核苷酸。在 其它情况下,所述支架序列包括至少2,3,4,5,6,7,8,9,10,11,12, 13,14,15,16,17,19,20个或更多个修饰的核苷酸。修饰的核苷酸可以 位于支架序列的任何核酸位置。例如,修饰的核苷酸可以在支架序列的第一个和/或最后一个核苷酸处或附近,和/或二者之间的任何位置。例如,对于 长度为约80个核苷酸的支架序列,一个或多个修饰的核苷酸可位于所述序列 的核酸位置第1位,第2位,第3位,第4位,第5位,第6位,第7位,第8位, 第9位,第10位,第11位,第12位,第13位,第14位,第15位,第16位,第17 位,第18位,第19位,第20位,第21位,第22位,第23位,第24位,第25位, 第26位,第27位,第28位,第29位,第30位,第31位,第32位,第33位,第 34位,第35位,第36位,第37位,第38位,第39位,第40位,第41位,第42 位,第43位,第44位,第45位,第46位,第47位,第48位,第49位,第50位, 第51位,第52位,第53位,第54位,第55位,第56位,第57位,第58位,第 59位,第60位,第61位,第62位,第63位,第64位,第65位,第66位,第67 位,第68位,第69位,第70位,第71位,第72位,第73位,第74位,第75位, 第76位,第77位,第78

位,第79位和/或第80位。在一些情况下,支架序列的 约1%至约10%,例如约1%至约8%,约1%至约5%,约5%至约10%,或约3% 至约7%可以包含修饰的核苷酸。在其它实例中,支架序列的约1%至约10%, 例如,约1%,约2%,约3%,约4%,约5%,约6%,约7%,约8%,约9%, 或约10%可包含修饰的核苷酸。

[0140] sgRNA的修饰核苷酸可以包括核糖(例如糖)基团、磷酸基团、核碱基或 其任何组合中的修饰。在一些实施方案中,核糖基团中的修饰包括核糖的2’ 位的修饰。

[0141] 在一些实施方案中,修饰的核苷酸包括2’-氟-阿拉伯糖核酸、三环-DNA (tc-DNA)、肽核酸、环己烯核酸(CeNA)、锁核酸(LNA)、乙烯桥核酸(ethylene- bridged nucleic acid,ENA)、磷酸二酰胺吗啉代(phosphodiamidate morpholino)、或其组合。

[0142] 修饰的核苷酸或核苷酸类似物可以包括糖修饰和/或骨架修饰的核糖核 苷酸(即,包括对磷酸-糖骨架的修饰)。例如,天然或自然RNA的磷酸二酯键 可以被修饰以包括氮或硫杂原子中的至少一个。在一些骨架修饰的核糖核苷 酸中,连接到相邻核糖核苷酸的磷酸酯基团可以被替换为修饰的基团,例如 硫代磷酸酯基团。在优选的糖修饰的核糖核苷酸中,2’部分是选自H、OR、 R、卤素、SH、SR、NH_2、NHR、NR_2或ON的基团,其中R是C_1-C_6烷基、烯基或炔基,F,Cl,Br或I。

[0143] 在一些实施方案中,修饰的核苷酸含有糖修饰。糖修饰的非限制性实例 包括2’-脱氧-2’-氟-寡核糖核苷酸(2’-氟-2’-脱氧胞苷-5’-三磷酸,2’-氟-2’-脱氧尿 苷-5’-三磷酸),2’-脱氧-2’-脱氨基寡核糖核苷酸(2’-氨基-2’-脱氧胞苷-5’-三磷酸, 2’-氨基-2’-脱氧尿苷-5’-三磷酸),2’-O-烷基寡核糖核苷酸,2’-脱氧-2’-C-烷基寡 核糖核苷酸(2’-O-甲基胞苷-5’-三磷酸,2’-甲基尿苷-5’-三磷酸),2’-C-烷基寡核 糖核苷酸,及其异构体(2’-阿糖胞苷-5’-三磷酸,2’-阿糖尿苷-5’-三磷酸),叠氮 三磷酸(2’-叠氮-2’-脱氧胞苷-5’-三磷酸,2’-叠氮-2’-脱氧尿苷-5’-三磷酸)及其组 合。

[0144] 在一些实施方案中,修饰的sgRNA含有一个或多个2’-氟、2’-氨基和/或2’- 硫修饰。在一些情况下,修饰是2’-氟-胞苷,2’-氟-尿苷,2’-氟-腺苷,2’-氟-鸟 苷,2’-氨基-胞苷,2’-氨基-尿苷,2’-氨基-腺苷,2’-氨基-鸟苷,2,6-二氨基嘌 呤,4-硫代-尿苷,5-氨基-烯丙基-尿苷,5-溴-尿苷,5-碘-尿苷,5-甲基-胞苷, 核糖-胸苷,2-氨基嘌呤,2’-氨基-丁酰基-苤-尿苷,5-氟-胞苷和/或5-氟-尿苷。

[0145] 在哺乳动物RNA上发现超过96个天然存在的核苷修饰。参见例如 Limbach等,Nucleic Acids Research,22(12):2183-2196(1994)。核苷酸和修 饰的核苷酸和核苷的制备在本领域中是公知的,并在例如美国专利号 4,373,071、4,458,066、4,500,707、4,668,777、4,973,679、5,047,524、5,132,418、5,153,319、5,262,530和5,700,642中有所描述。适用于本文所述的许多修饰的 核苷和修饰的核苷酸是可商购的。核苷可以是天然存在的核苷的类似物。在 一些情况下,类似物是二氢尿苷,甲基腺苷,甲基胞苷,甲基尿苷,甲基假尿苷,硫尿苷,脱氧胞嘧啶和脱氧尿苷。

[0146] 在一些情况下,本文描述的修饰的sgRNA包括核碱基修饰的核糖核苷酸, 即包含至少一个非天然存在的核碱基而不是天然存在的核碱基的核糖核苷 酸。可以掺入修饰的核苷和修饰的核苷酸的修饰的核碱基的非限制性实例包 括:m5C(5-甲基胞苷),m5U(5-甲基尿苷),m6A(N6-甲基腺苷),s2U(2-硫尿 苷),Um(2’-O-甲基尿苷),m1A(1-甲基腺苷)(2-1-O-甲基腺苷),ms2m6A(2-甲 硫基-N6-甲基腺苷),i6A(N6-异戊烯基腺苷),ms216A(2-甲

硫基-N6异戊烯基 腺苷),io6A(N6-(顺式-羟基异戊烯基)腺苷),ms2io6A(2-甲硫基-N6-(顺式- 羟基异戊烯基)腺苷),g6A(N6-甘氨酰氨基甲酰基腺苷),t6A(N6-苏氨酰氨基 甲酰腺苷),ms2t6A(2-甲硫基-N6-苏氨酰氨基甲酰腺苷),m6t6A(N6-甲基-N6- 苏氨酰氨基甲酰基腺苷),hn6A(N6-羟基神经氨酰基氨基甲酰基腺苷), ms2hn6A(2-甲硫基-N6-羟基香叶基氨甲酰腺苷),Ar(p)(2'-O-核糖腺苷(磷酸)), I(肌苷),m11(1-甲基肌苷),m'Im(1,2′-O-二甲基肌苷),m3C(3-甲基胞苷),Cm (2T-O-甲基胞苷),s2C(2-硫代胞苷),ac4C(N4-乙酰胞苷),f5C(5-甲酰化胞苷), m5Cm(5,2-O-二甲基肌苷),ac4Cm(N4乙酰基2TO甲基胞苷),k2C(甲咪唑啉), m1G(1-甲基鸟苷),m2G(N2-甲基鸟苷),m7G(7-甲基鸟苷),Gm(2′-O-甲基鸟苷),m22G(N2,N2-二甲基鸟苷),m2Gm(N2,2′-O-二甲基鸟苷),m22Gm (N2,N2,2′-O-三甲基鸟苷),Gr(p)(2′-O-核糖基鸟苷(磷酸)),yW(怀丁苷),o2yW (过氧怀丁苷),OHyW(羟基怀丁苷),OHyW*(修饰不足的羟基怀丁苷),imG (怀俄苷),mimG(甲基鸟苷),Q(Q核苷),oQ(环氧Q核苷),galQ(半乳糖基-Q核 苷),manQ(甘露糖基-Q核苷),preQo(7-氰基-7-脱氮鸟苷),preQi(7-氨基甲基- 7-脱氮鸟苷),G(古嘌苷),D(二羟尿苷),m5Um(5,2′-O-二甲基尿苷),s4U(4-硫 代尿苷),m5s2U(5-甲基-2-硫代尿苷),s2Um(2-硫-2′-O-甲基尿苷),acp3U(3-(3-氨基-3-羧基丙基)尿苷),ho5U(5-羟基尿苷),mo5U(5-甲氧基尿苷),cmo5U (尿苷5-羟乙酸),mcmo5U(尿苷5-羟乙酸甲酯),chm5U(5-(羧基羟基甲基)尿 苷)),mchm5U(5-(羧基羟基甲基)尿苷甲酯),mcm5U(5-甲氧基羰基甲基尿苷), mcm5Um(S-甲氧基羰基甲基-2-O-甲基尿苷),mcm5s2U(5-甲氧基羰基甲基- 2-硫代尿苷),nm5s2U(5-氨基甲基-2-硫代尿苷),mnm5U(5-甲基氨基甲基尿 苷),mnm5s2U(5-甲基氨基甲基-2-硫代尿苷),mnm5se2U(5-甲基氨基甲基-2- 硒代尿苷),ncm5U(5-氨甲酰基甲基尿苷),ncm5Um(5-氨甲酰基甲基-2′-O-甲基尿苷),cmnm5U(5-羧基甲基氨基甲基尿苷),cnmm5Um(5-羧基甲基氨基甲 基-2-L-O甲基尿苷),cmnm5s2U(5-羧基甲基氨基甲基-2-硫代尿苷),m62A (N6,N6-二甲基腺苷),Tm(2′-O-甲基肌苷),m4C(N4-甲基胞苷),m4Cm(N4,2- O-二甲基胞苷),hm5C(5-羟基甲基胞苷),m3U(3-甲基尿苷),cm5U(5-羧基甲 基尿苷),m6Am(N6,T-O-二甲基腺苷),rn62Am(N6,N6,O-2-三甲基腺苷), m2′7G(N2,7-二甲基鸟苷),m2′2′7G(N2,N2,7-三甲基鸟苷),m3Um(3,2T-O-二 甲基尿苷),m5D(5-甲基二羟尿苷),f5Cm(5-甲酰基-2′-O-甲基胞苷),m1Gm (1,2′-O-二甲基鸟苷),m'Am(1,2-O-二甲基腺苷)亚氨基甲基尿苷),tm5s2U(S- 牛磺胆酸甲基-2-硫代尿苷)),imG-14(4-脱甲基鸟苷),imG2(异鸟苷),或ac6A (N6-乙酰腺苷),次黄嘌呤,肌苷,8-氧代-腺嘌呤,其7-取代的衍生物,2-尿嘧 啶,假尿嘧啶,2-硫尿嘧啶,4-硫尿嘧啶,5-氨基尿嘧啶,5-(C1-C6)-烷基尿 嘧啶,5-甲基尿嘧啶,5-(C2-C6)-烯基尿嘧啶,5-(C2-C6)-炔基尿嘧啶,5-(羟 基甲基)尿嘧啶,5-氯尿嘧啶,5-氟尿嘧啶,5-溴尿嘧啶,5-羟基胞嘧啶,5- (C1-C6)-烷基胞嘧啶,5-甲基胞嘧啶,5-(C2-C6)-烯基胞嘧啶,5-(C2-C6)-炔基 胞嘧啶,5-氯胞嘧啶,5-氟胞嘧啶,5-溴胞嘧啶,N2-二甲基鸟嘌呤,7-脱氮鸟 嘌呤,8-氮鸟嘌呤,7-脱氮-7-取代鸟嘌呤,7-脱氮-7-(C2-C6)炔基鸟嘌呤,7- 脱氮-8-取代鸟嘌呤,8-羟基鸟嘌呤,6-硫鸟嘌呤,8-羟基鸟嘌呤,2-氨基嘌呤, 2-氨基-6-氯嘌呤,2,4-二氨基嘌呤,2,6-二氨基嘌呤,8-氮杂嘌呤,取代的7-脱 氮嘌呤,7-脱氮-7-取代的嘌呤,7-脱氮-8-取代的嘌呤,及其组合。

[0147] 在一些实施方案中,修饰的sgRNA的磷酸骨架被改变。修饰的sgRNA可 包括一个或多个硫代磷酸酯、氨基磷酸酯(例如N3'-P5'-氨基磷酸酯(NP))、2'- O-甲氧基-乙基(2'

MOE)、2'-O-甲基-乙基(2ME)和/或甲基膦酸酯联接。

[0148] 在具体的实施方案中,修饰的sgRNA的引导序列的一个或多个修饰的核 苷酸和/或修饰的sgRNA的支架序列的一个或多个修饰的核苷酸包括2'-O-甲 基(M)核苷酸,2'-O-甲基3'-硫代磷酸酯(MS)核苷酸,2'-O-甲基3'硫代 PACE(MSP)核苷酸,或其组合。在一些情况下,修饰的sgRNA包含一个或多 个MS核苷酸。在其他情况下,修饰的sgRNA包含一个或多个MSP核苷酸。在 其他情况下,修饰的sgRNA包含一个或多个MS核苷酸和一个或多个MSP核苷酸。在进一步的例子中,修饰的sgRNA不包含M个核苷酸。在某些情况下, 修饰的sgRNA包含一个或多个MS核苷酸和/或一个或多个MSP核苷酸,还包 括一个或多个M个核苷酸。在某些其他情况下,MS核苷酸和/或MSP核苷酸 是修饰的sgRNA中存在的唯一修饰的核苷酸。

[0149] 应该注意的是,在修饰的sgRNA的引导序列和/或支架序列中可以组合和 纳入任何本文所述的修饰。

[0150] 在一些情况下,修饰的sgRNA还包括结构修饰,例如茎环,例如M2茎环 或四环。

[0151] 修饰的sgRNA可以通过本领域普通技术人员已知的任何方法合成。在一 些实施方案中,修饰的sgRNA是化学合成的。修饰的sgRNA可以使用2'-O-硫 代氨基甲酸酯保护的核苷亚磷酰胺来合成。方法在例如Dellinger et al., J.American Chemical Society 133, 11540-11556(2011);Threlfall et al.,Organic &Biomolecular Chemistry 10,746-754(2012);and Dellinger et al.,J.American Chemical Society 125,940-950(2003)中有记载。

[0152] 化学修饰的sgRNA可以与任何CRISPR相关技术(例如RNA引导技术)一 起使用。如本文所述,修饰的sgRNA可以用作任何Cas核酸酶或其变体或片 段(包括任何工程化或人造Cas9多肽)的引导。修饰的sgRNA可以靶向分离的 原代细胞中的DNA和/或RNA分子用于体外治疗或体内(例如在动物中)。本文 公开的方法可以应用于基因组编辑,基因调控,成像以及任何其他基于 CRISPR的应用。

[0153] 3.供体修复模板

[0154] 在一些实施方案中,本发明提供重组供体修复模板,其包含两条同源臂, 所述同源臂与靶DNA序列(例如靶基因或基因座)上位于Cas核酸酶(例如Cas9 核酸酶)切割位点任一侧的部分同源。在某些情况下,重组供体修复模板包含 报道盒,所述报告盒包含:编码报道多肽(例如,可检测多肽,荧光多肽或选 择标记)的核苷酸序列,和位于报道盒两侧、且与靶DNA上位于Cas核酸酶切 割位点任一侧的部分同源的两条同源臂。报告盒可以进一步包含编码自切割 肽、一个或多个核定位信号、和/或荧光多肽,例如superfolder GFP(sfGFP)的 序列。

[0155] 在一些实施方案中,同源臂具有相同的长度。在其他实施方案中,同源 臂是不同的长度。同源臂可以是至少约10个碱基对(bp),例如至少约10bp、 15bp、20bp、25bp、30bp、35bp、45bp、55bp、65bp、75bp、85bp、95 bp、100bp、150bp、200bp、250bp、300bp、350bp、400bp、450bp、500 bp、550bp、600bp、650bp、700bp、750bp、800bp、850bp、900bp、950 bp、1000bp、1.1千碱基(kb)、1.2kb、1.3kb、1.4kb、1.5kb、1.6kb、1.7kb、 1.8kb、1.9kb、2.0kb、2.1kb、2.2kb、2.3kb、2.4kb、2.5kb、2.6kb、2.7 kb、2.8kb、2.9kb、3.0kb、3.1kb、3.2kb、3.3kb、3.4kb、3.5kb、3.6kb、 3.7kb、3.8kb、3.9kb、4.0kb,或更长。同源臂可为约10bp至约4kb,例如 约10bp至约20bp、约10bp至约50bp、约10bp至约100bp、约10bp至约200 bp、约

10bp至约500bp、约10bp至约1kb、约10bp至约2kb、约10bp至约4 kb、约100bp至约200bp、约100bp至约500bp、约100bp至约1kb、约100bp 至约2kb、约100bp至约4kb、约500bp至约1kb、约500bp至约2kb、约500 bp至约4kb、约1kb至约2kb、约1kb至约2kb、约1kb至约4kb、或约2kb至 约4kb。

[0156] 供体修复模板可被克隆到表达载体中。可以使用本领域普通技术人员已 知的常规的基于病毒和非病毒的表达载体。

[0157] 单链寡聚脱氧核苷酸(ssODN)供体模板可以代替重组供体修复模板用于 同源重组介导的修复。ssODN可用于在靶DNA中导入短的修饰。例如,ssODN 适用于精确纠正SNP等基因突变。ssODNs可以在Cas核酸酶切割的靶位点的 每一侧包含两个侧翼的同源序列,并且可以相对于靶DNA呈有义或反义取向。 每个侧翼序列可以是至少约10个碱基对(bp),例如至少约10bp、15bp、20bp、 25bp、30bp、35bp、40bp、45bp、50bp、55bp、60bp、65bp、70bp、75bp、80bp、85bp、90bp、95bp、100bp、150bp、200bp、250bp、300bp、 350bp、400bp、450bp、500bp、550bp、600bp、650bp、700bp、750bp、 800bp、850bp、900bp、950bp、1kb、2kb、4kb,或更长。在一些实施方 案中,每条同源臂为约10bp至约4kb,例如,约10bp至约20bp、约10bp至约50bp、约10bp至约100bp、约10bp至约200bp、约10bp至约500bp、约10 bp至约1kb、约10bp至约2kb、约10bp至约4kb、约100bp至约200bp、约100 bp至约500bp、约100bp至约1kb、约100bp至约2kb、约100bp至约4kb、约 500bp至约1kb、约500bp至约2kb、约500bp至约4kb、约1kb至约2kb、约 1kb至约2kb、约1kb至约4kb,或约2kb至约4kb。ssODN的长度可为至少 约25个核苷酸(nt),例如至少约25nt、30nt、35nt、40nt、45nt、50nt、55 nt、60nt、65nt、70nt、75nt、80nt、85nt、90nt、95nt、100nt、150nt、200nt、250nt、300nt,或更长。在一些实施方案中,ssODN的长度为约25至 约50;约50至约100;约100至约150;约150至约200;约200至约250;约250 至约300;或约25nt至约300nt。

[0158] 在一些实施方案中,所述ssODN模板包含至少一个,例如1、2、3、4、 5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、200、300、 400、500个或更多个本文所描述的修饰的核苷酸。在一些情况下,ssODN的 序列的至少1%、5%、10%、20%、30%、40%、50%、60%、70%、80%、90% 或99%包括修饰核苷酸。在一些实施方案中,修饰的核苷酸位于ssODN的一 个或两个末端。修饰的核苷酸可以是自末端起的第一,第二,第三,第四, 第五,第六,第七,第八,第九或第十个核苷酸,或其任何组合。例如,修 饰的核苷酸可以位于ssODN模板两端的三个末端核苷酸处。另外,修饰的核 苷酸可以位于末端的内部。

[0159] 4.靶DNA

[0160] 在CRISPR/Cas系统中,靶DNA序列可以与靶向DNA的RNA(例如,修饰 的sgRNA)的片段互补,并且其后面可以紧跟着原型间隔子邻接基序(PAM)序 列。靶DNA位点可能直接位于PAM序列的5'处,此PAM序列对于所使用的 Cas9的细菌种类是特异性的。例如,来自酿脓链球菌的Cas9的PAM序列是 NGG;来自脑膜炎奈瑟氏球菌的Cas9的PAM序列是NNNNGATT;来自嗜热 链球菌的Cas9的PAM序列是NNAGAA;而来自密螺旋体的Cas9的PAM序列 是NAAAAC。在一些实施方案中,PAM序列可以是5'-NGG,其中N是任何核 苷酸;5'-NRG,其中N是任何核苷酸,R是嘌呤;或5'-NNGRR,其中N是任 何核苷酸,R是嘌呤。对于酿脓链球菌系统,选择的靶DNA序列应该紧接在 5'NGG PAM之前(例如位于5'),其中N是任何核苷酸,使得靶向DNA的RNA 的引导序列(例如,修饰的sgRNA)与相对链碱基配对,以介导PAM序列上游 约3个碱基

对处的切割。

[0161] 在一些实施方案中，当将靶向DNA的RNA(例如，修饰的sgRNA)的引导 序列与其相应的靶DNA序列使用合适的比对算法最佳比对时，二者之间的互 补性程度为大约，或多于大约50％、55％、60％、65％、70％、75％、80％、 85％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或更多。可以使用用于比对序列的任何合适算法来确定最佳比对，此类算法的非限制 性实例包括Smith-Waterman算法、Needleman-Wunsch算法、基于Burrows- Wheeler变换的算法(例如Burrows Wheeler Aligner)、ClustalW、Clustal X、BLAT、Novoalign(Novocraft Technologies，Selangor，Malaysia)和ELAND (Illumina，San Diego，CA)。

[0162] 可以使用基于网络的软件在预定的基因组序列(基因)中选择靶DNA位 点，此类软件例如ZiFiT Targeter软件(Sander et al.，2007，Nucleic Acids Res, 35:599-605；Sander et al.，2010，Nucleic Acids Res，38:462-468)、E-CRISP (Heigwer et al.，2014，Nat Methods，11:122-123)、RGEN Tools(Bae et al.，2014, Bioinformatics，30(10):1473-1475)、CasFinder(Aach et al.，2014，bioRxiv)、 DNA2.0 gNRADesign Tool (DNA2.0，Menlo Park，CA)以及CRISPR Design Tool(Broad Institute，Cambridge，MA)。这样的工具分析基因组序列(例如感兴 趣的基因或基因座)并鉴定用于基因编辑的合适靶位点。为了评估每种靶向 DNA的RNA(例如，修饰的sgRNA)的脱靶基因修饰，基于对碱基配对错配同 一性、位置和分布的定量特异性分析，进行脱靶位点的计算预测。

[0163] 5.调变基因表达

[0164] 调控基因表达(诸如抑制基因表达或激活基因表达)的CRISPR/Cas系统 可包括野生型或天然Cas核酸酶(例如Cas9多肽变体或片段)的变体或片段，以 及和DNA靶向性sgRNA或RNA靶向性sgRNA。作为一个非限制性实例，包含 Cas9变体或片段与sgRNA的复合物，其中sgRNA能够结合与sgRNA的一部分 互补的靶DNA序列，可以阻断或阻碍RNA聚合酶的转录起始和/或延伸。这 进而可以抑制或阻遏靶DNA的基因表达。或者，包含另一种不同Cas9变体或 片段与sgRNA的复合物，其中sgRNA能够结合与sgRNA的一部分互补的靶 DNA序列，可以诱导或激活靶DNA的基因表达。

[0165] 关于进行CRISPR干扰(CRISPRi)以灭活或减少基因表达的方法的详细 描述见于例如Larson et al.，Nature Protocols，2013，8(11):2180-2196，和Qi et al., Cell，152，2013，1173-1183。在CRISPRi中，sgRNA-Cas9变体复合物可以结合 蛋白质编码区域的非模板DNA链，并阻断转录延伸。在一些情况下，当 sgRNA-Cas9变体复合物结合基因的启动子区域时，复合物阻止或阻碍转录 起始。

[0166] 关于进行CRISPR激活以增加基因表达的方法的详细描述可见，例如：Cheng et al.，Cell Research，2013，23:1163-1171，Konerman et al.，Nature，2015, 517:583-588，以及美国专利8，697，359。

[0167] 关于以调变基因表达为目的的基于CRISPR的RNA结合和/或切割的详 细描述可见，例如，O'Connell et al.，Nature，2014，516:263-266，和美国专利 申请公布号2014/0302563。

[0168] 对于基于CRISPR的基因表达控制，可以使用Cas核酸酶(例如Cas9多肽) 的缺乏核酸内切酶活性的催化失活变体。在一些实施方案中，Cas核酸酶是 在RuvC样和HNH核酸酶结

构域中含有至少两个点突变的Cas9变体。在一些 实施方案中,Cas9变体具有D10A和H840A氨基酸取代,其被称为dCas9(Jinek et al.,Science,2012,337:816-821;Qi et al.,Cell,152(5):1173-1183)。在一些情 况下,来自酿脓链球菌的dCas9多肽在位置D10、G12、G17、E762、H840、 N854、N863、H982、H983、A984、D986、A987或它们的任何组合上包含至 少一个突变。关于此类dCas9多肽及其变体的描述可见例如在国际专利公开 号WO 2013/176772。dCas9酶可以在D10、E762、H983或D986处含有突变, 以及在H840或N863处有突变。在某些情况下,dCas9酶含有D10A或D10N突 变。而且,dCas9酶可以包括H840A、H840Y或H840N。在一些情况下,dCas9 酶包含D10A和H840A;D10A和H840Y;D10A和H840N;D10N和H840A; D10N和H840Y;或D10N和H840N替换。取代可以是保守取代或非保守取代, 以使得Cas9多肽催化失活并能够与靶DNA结合。

[0169] 在某些实施方案中,dCas9多肽是催化失活的,例如核酸酶活性缺陷。在 一些情况下,dCas9酶或其变体或片段可以阻断靶序列的转录,并且在一些 情况下,可以阻断RNA聚合酶。在其他情况下,dCas9酶或其变体或片段可 以激活靶序列的转录。

[0170] 在某些实施方案中,缺乏核酸内切酶活性的Cas9变体(例如,dCas9)可以 与转录阻遏结构域[例如Kruppel关联盒(KRAB)结构域]或转录激活结构域(例 如VP16反式激活结构域)融合。在一些实施方案中,Cas9变体是包含dCas9和 转录因子(例如RNA聚合酶ω因子、休克因子1、或其片段)的融合多肽。在其 他实施方案中,Cas9变体是包含dCas9和DNA甲基化酶、组蛋白乙酰化酶、 或其片段的融合多肽。

[0171] 对于通过RNA结合和/或RNA切割介导的基于CRISPR的基因表达控制, 可以使用合适的Cas核酸酶(例如Cas9多肽)的具有内切核糖核酸酶活性的变 体,例如O'Connell et al.,Nature,2014,516:263-266中描述的。其他有用的Cas 核酸酶(例如Cas9)变体描述于例如美国专利申请公开号2014/0302563中。其 它能够切割RNA的CRISPR相关酶包括Csy4内切核糖核酸酶、CRISPR相关的 Cas6酶、Cas5家族成员酶、Cas6家族成员酶、I型CRISPR系统内切核糖核酸 酶、II型CRISPR系统内切核糖核酸酶、III型CRISPR系统内切核糖核酸酶及其变体。

[0172] 在基于CRISPR的RNA切割的一些实施方案中,与本文所述的修饰的 sgRNA和Cas核酸酶(例如Cas9)变体一并使用含有PAM序列的DNA寡核苷酸 (例如,PAMmer)并切割单链RNA转录物。合适的PAMmer序列的详细描述参 见例如O'Connell et al.,Nature,2014,516:263-266。

[0173] Cas核酸酶(例如Cas9多肽)或其变体或片段可以以多肽、编码多肽的 mRNA、或包含编码多肽的核苷酸序列的重组表达载体的形式提供。其他细 节可见前文。

[0174] 在一些实施方案中,使用多个修饰的sgRNA来靶向靶基因的不同区域以 调控该靶基因的基因表达。与每个修饰的sgRNA单独相比,多个修饰的 sgRNA可以提供对单一靶基因的基因表达的协同调变(例如抑制或活化)。在 其他实施方案中,使用多个修饰的sgRNA来调控至少两种靶基因的基因表达。

[0175] B.原代细胞

[0176] 本发明可用于诱导任何感兴趣的原代细胞中靶核酸的基因调控。原代细 胞可以是从任何多细胞生物体分离的细胞,例如植物细胞(例如稻细胞、小麦 细胞、番茄细胞、拟南芥细胞、玉米细胞等)、来自多细胞原生生物的细胞、 来自多细胞真菌的细胞、动物细胞

如来自无脊椎动物(例如果蝇、海胆、棘皮 动物、线虫等)的细胞或来自脊椎动物(例如鱼、两栖动物、爬行动物、鸟类、 哺乳动物等)的细胞、来自人的细胞、来自健康人的细胞、来自人患者的细胞、 来自癌症患者的细胞等。在一些情况下,具有被诱导的基因调控的原代细胞 可以移植到受试者(例如患者)。例如,原代细胞可以来自待治疗的受试者(例 如,患者)。

[0177] 任何类型的原代细胞可能是感兴趣的,诸如干细胞,例如胚胎干细胞、 诱导的多能干细胞、成体干细胞(例如间充质干细胞、神经干细胞、造血干细 胞、器官干细胞)、祖细胞、体细胞(例如成纤维细胞、肝细胞、心脏细胞、 肝细胞、胰细胞、肌细胞、皮肤细胞、血细胞、神经细胞、免疫细胞)和身体 (例如人体)的任何其它细胞。细胞可以是源自受试者(例如动物受试者或人受 试者)的原代细胞或原代细胞培养物,并允许在体外生长有限的传代数。在一 些实施方案中,细胞是疾病细胞或来源于患有疾病的受试者。例如,细胞可 以是癌症或肿瘤细胞。

[0178] 原代细胞可以通过任何标准方法从受试者中收获。例如,来自组织如皮 肤、肌肉、骨髓、脾、肝、肾、胰腺、肺、肠、胃等的细胞可通过组织活检 或细针抽吸收集。血细胞和/或免疫细胞可以从全血、血浆或血清中分离。在 一些情况下,合适的原代细胞包括外周血单个核细胞(PBMC),外周血淋巴细 胞(PBL)和其它血细胞亚群,例如但不限于T细胞、天然杀伤细胞、单核细胞、 天然杀伤细胞T细胞、单核细胞前体细胞、造血干细胞和祖细胞(HSPC)如 CD34+HSPC或非多能干细胞。在一些情况下,细胞可以是任何免疫细胞, 包括但不限于任何T细胞,例如肿瘤浸润细胞(TIL)、CD3+T细胞、CD4+T细 胞、CD8+T细胞或任何其他类型的T细胞。T细胞还可以包括记忆T细胞、 记忆干细胞T细胞或效应T细胞。T细胞也可以向特定的群体和表型漂移 (skew)。例如,T细胞可以漂移为包括CD45RO(-)、CCR7(+)、CD45RA(+)、 CD62L(+)、CD27(+)、CD28(+)和/或IL-7Rα(+)表型。可以选择合适的细胞, 它们包含选自下组的一种或多种标志物:CD45RO(-)、CCR7(+)、CD45RA(+)、 CD62L(+)、CD27(+)、CD28(+)和/或IL-7Rα(+)。诱导的多能干细胞可以根据 例如美国专利号7,682,828、8,058,065、8,530,238、8,871,504、8,900,871和 8,791,248中所述的标准方案由分化的细胞产生。

[0179] 在一些实施方案中,原代细胞是体外的。在其他实施方案中,原代细胞 是离体的。

[0180] C.离体疗法

[0181] 本文所述的方法可用于离体疗法。离体疗法可以包括给受试者(例如患 者)施用在生物体外部产生或修饰的组合物(例如细胞)。在一些实施方案中, 组合物(例如细胞)可以通过本文公开的方法产生或修饰。例如,离体疗法可 以包括将生物体外部产生或修饰的原代细胞给予受试者(例如患者),其中所 述原代细胞已经根据本发明的方法体外培养,所述方法包括使原代细胞中的 靶核酸接触一种或多种本文中所述的修饰的siRNA,以及Cas核酸酶(例如 Cas9多肽)或其变体或片段,或编码Cas核酸酶(例如Cas9多肽)或其变体或片段的mRNA,或包含编码Cas核酸酶(例如Cas9多肽)或其变体或片段的核苷酸 序列的重组表达载体。

[0182] 在一些实施方案中,组合物(例如细胞)可以源自要通过离体疗法治疗的 受试者(例如患者)。在一些实施方案中,离体疗法可以包括基于细胞的疗法, 例如过继免疫疗法。

[0183] 在一些实施方案中,体外治疗中使用的组合物可以是细胞。细胞可以是 原代细胞,包括但不限于外周血单个核细胞(PBMC)、外周血淋巴细胞(PBL) 和其他血细胞亚群。原

代细胞可以是免疫细胞。原代细胞可以是T细胞(例如 CD3+T细胞、CD4+T细胞和/或CD8+T细胞)、天然杀伤细胞、单核细胞、天 然杀伤T细胞、单核细胞前体细胞、造血干细胞细胞或非多能干细胞、干细 胞、或祖细胞。原代细胞可以是造血干细胞或祖细胞(HSPC),例如CD34+HSPC。原代细胞可以是人类细胞。原代细胞可以被分离、选择和/或培养。 原代细胞可以离体扩增。原代细胞可以在体内扩增。原代细胞可以是 CD45RO(-),CCR7(+),CD45RA(+),CD62L(+),CD27(+),CD28(+)和/或IL- 7Rα(+)。原代细胞对于需要的受试者可以是自体的。原代细胞对于需要的受 试者可以是非自体的。原代细胞可以是符合良好生产规范(GMP)的试剂。原 发性细胞可以是在需要的受试者中治疗疾病(包括癌症、感染、自身免疫性疾 病或移植物抗宿主病(GVHD))的组合疗法的一部分。

[0184] 作为离体疗法的非限制性实例,可以在使原代细胞内的靶核酸与Cas核 酸酶和修饰的sgRNA接触之前从多细胞生物体(例如,植物、多细胞原生生物、 多细胞真菌、无脊椎动物、脊椎动物等)分离原代细胞。在将靶核酸与Cas核 酸酶和修饰的sgRNA接触之后,可以将原代细胞或其后代(例如来源于原代 细胞的细胞)返回到多细胞生物体中。

[0185] D.将核酸导入靶细胞的方法

[0186] 用于将多肽和核酸导入靶细胞(宿主细胞)的方法是本领域已知的,并且 可以使用任何已知的方法将核酸酶或核酸[例如,编码核酸酶的核苷酸序列、靶向DNA的RNA(例如修饰的单引导RNA),用于同源性指导修复(HDR)的供 体修复模板等]导入细胞,例如原代细胞如干细胞、祖细胞或分化的细胞。合 适方法的非限制性实例包括电穿孔、病毒或噬菌体感染、转染、接合、原生 质体融合、脂质体转染、磷酸钙沉淀、聚乙烯亚胺(PEI)介导的转染、DEAE- 葡聚糖介导的转染、脂质体介导的转染、粒子枪技术、磷酸钙沉淀、直接显 微注射、纳米颗粒介导的核酸递送等。

[0187] 在一些实施方案中,可以使用递送系统将CRISPR/Cas介导的基因调控的 组分导入细胞中。在某些情况下,递送系统包含纳米颗粒、微粒(例如聚合物 微聚合物)、脂质体、胶束、病毒微体、病毒颗粒、核酸复合物、转染剂、电 穿孔剂(例如使用NEON转染系统)、核转染试剂、脂质转染试剂和/或包括下 述组分的缓冲系统:核酸酶组分(作为多肽或由表达构建体编码)和一种或多 种核酸组分[如靶向DNA的RNA(例如修饰的单引导RNA)]和/或供体修复模 板。例如,可以将各组分与脂质转染剂混合,使得它们被包封或包装成阳离 子亚微米水包油乳液。或者,可以在没有递送系统的情况下递送组分,例如 作为水溶液。

[0188] 制备脂质体和在脂质体中包封多肽和核酸的方法在例如Methods and Protocols,Volume 1:Pharmaceutical Nanocarriers:Methods and Protocols.(ed.Weissig).Humana Press,2009,和Heyes et al.(2005)J Controlled Release 107:276-87中有描述。制备微粒和包封多肽和核酸的方法描述于例如 Functional Polymer Colloids and Microparticles volume 4(Microspheres, microcapsules&liposomes).(eds.Arshady&Guyot).Citus Books,2002,以及 Microparticulate Systems for the Delivery of Proteins and Vaccines.(eds.Cohen &Bernstein).CRC Press,1996。

[0189] E.评估基因组编辑效率的方法

[0190] 为了功能性地测试正确的基因组编辑修饰的存在,可以通过本领域技术 人员已知的标准方法分析靶DNA。例如,可以使用SURVEYOR®突变检测试 剂盒(Integrated DNA Technologies,Coralville,IA)或Guide-it™Indel鉴定试剂 盒(Clontech,Mountain

View,CA)进行测序来鉴定indel突变。同源性指导修 复(HDR)可以通过基于PCR的方法,并结合测序或RFLP分析来加以检测。基 于PCR的试剂盒的非限制性实例包括Guide-it突变检测试剂盒(Clontech)和 GeneArt®基因组切割检测试剂盒(Life Technologies,Carlsbad,CA)。也可以 使用深度测序,特别是对于大量样品或潜在的靶/脱靶位点而言。

[0191] 在某些实施方案中,基因组编辑的效率(例如特异性)对应于:中靶(on- target)的基因组切割事件的数目或百分比相对于所有基因组切割事件(包括 中靶的事件和脱靶的事件)的数量或百分比。

[0192] 在一些实施方案中,相对于相应的未修饰的sgRNA,本文描述的修饰的 sgRNA能够增强细胞(例如原代细胞)中靶DNA序列的基因组编辑。基因组编 辑可包括同源性指导修复(HDR)(例如,插入,缺失或点突变)或非同源末端连 接(NHEJ)。

[0193] 在某些实施方案中,与相应的未修饰的sgRNA序列相比,在存在本文描 述的修饰的sgRNA的情况下,细胞中靶DNA序列的核酸酶介导的基因组编辑 效率增加至少约0.5倍,0.6倍,0.7倍,0.8倍,0.9倍,1倍,1.1倍,1.2倍,1.3 倍,1.4倍,1.5倍,2倍,2.5倍,3倍,3.5倍,4倍,4.5倍,5倍,5.5倍,6倍, 6.5倍,7倍,7.5倍,8倍,8.5倍,9倍,9.5倍,10倍,15倍,20倍,25倍,30 倍,35倍,40倍,45倍,50倍或更大。

[0194] F.用于细胞中靶核酸的基因调控的方法

[0195] 本文提供了用于诱导基因调控,例如基因组编辑和/或调变(例如抑制或 激活)细胞中靶核酸的基因表达的方法。细胞可以是体外的(例如,用于离体 疗法的原代细胞)或体内的(例如人类器官或组织中的细胞)。

[0196] 诱导基因组编辑的方法包括向细胞中导入本文所述的修饰的sgRNA和 Cas核酸酶(例如Cas9多肽)或其变体或片段、编码Cas核酸酶(例如Cas9多肽) 或变体或其片段的mRNA,或包含编码Cas核酸酶(例如Cas9多肽)或其变体或 片段的核苷酸序列的重组表达载体。修饰的sgRNA将Cas核酸酶(例如Cas9多 肽)或其变体或片段引导至靶核酸(例如靶DNA)。与相应的未修饰的sgRNA序 列相比,修饰的sgRNA具有增强的活性、稳定性和/或对靶DNA的特异性。在 某些情况下,基因组编辑是靶DNA的非同源末端连接(NHEJ)。在其他情况下, 基因组编辑是靶DNA的同源性指导修复(HDR)。在HDR的一些实施方案中, 将重组供体修复模板添加到细胞中。重组供体修复模板可以包括两个核苷酸 序列,它们包含靶DNA的两个非重叠的同源部分,其中所述核苷酸序列位于 对应于靶DNA的核苷酸序列的5'和3'端。在一些实施方案中,重组供体修复 模板包含合成的单链寡聚脱氧核苷酸(ssODN)模板,该模板包含:编码修饰 单核苷酸多态性(SNP)的突变的核苷酸序列、以及包含靶的两个非重叠同源 部分的两个核苷酸序列DNA,其中核苷酸序列位于编码突变的核苷酸序列的 5'和3'端。使用任何合适的方法,例如通过电穿孔,将修饰的sgRNA和/或Cas 核酸酶(例如Cas9多肽)或其变体或片段、编码Cas核酸酶(例如Cas9多肽)或其 变体或片段的mRNA、或包含编码Cas核酸酶(例如Cas9多肽)或其变体或片段 的核苷酸序列的重组表达载体导入细胞中。

[0197] 用于调变(例如抑制或激活)细胞中的靶核酸(例如靶DNA)的基因表达的 方法包括向细胞中导入(例如,电穿孔)本文所述的经修饰的sgRNA和Cas核酸 酶(例如Cas9多肽)或其变体或片段、编码Cas核酸酶(例如Cas9多肽)或其变体 或片段的mRNA、或包含编码Cas核酸酶(例如Cas9多肽)或变体或片段的核苷 酸序列的重组表达载体。在一些实施方案中,

Cas核酸酶(例如Cas9)变体是内 切核酸酶缺陷型Cas(例如dCas9)多肽。在一些情况下,与野生型Cas9多肽相 比,Cas9变体可以具有两个或更多个氨基酸取代。在其他情况下,Cas9变体 不能切割双链DNA。Cas核酸酶变体可以是Cas(例如dCas9)融合多肽。在一 些实施方案中,融合多肽包括转录阻抑结构域、转录激活结构域、转录因子、 组蛋白修饰酶(例如组蛋白脱乙酰酶、组蛋白甲基转移酶、组蛋白乙酰转移酶), DNA修饰酶(例如DNA甲基转移酶)等等。

[0198] 用于调变(例如抑制或激活)细胞中靶核酸(例如靶RNA)的基因表达的方 法包括向细胞中导入(例如,电穿孔)本文所述的修饰的sgRNA以及Cas核酸酶 (例如Cas9多肽)或其变体或片段、编码Cas核酸酶(例如Cas9多肽)或其变体或 片段的mRNA、或包含编码Cas核酸酶(例如Cas9多肽)的核苷酸序列或其变体 或片段的重组表达载体。在一些实施方案中,Cas核酸酶(例如Cas9)变体已经 降低或缺乏内切核酸酶活性。Cas核酸酶变体可以含有两个或更多个氨基酸 取代,使得多肽不能切割双链DNA。Cas核酸酶(例如Cas9)变体可具有内切核 糖核酸酶活性并可切割靶RNA。

[0199] G.预防或治疗受试者的遗传疾病的方法

[0200] 本文描述的修饰的sgRNA可用于调控基因调控的效率。例如,相对于相 应的未修饰的sgRNA,修饰的sgRNA可以以增强的活性诱导基因调控。在一 些情况下,增强的活性包括修饰的sgRNA的增加的稳定性和/或修饰的sgRNA 对于靶核酸的增加的特异性。作为另一个实例,相对于相应的未修饰的 sgRNA,修饰的sgRNA可以以降低的细胞毒性诱导基因调控。

[0201] 修饰的sgRNA可以应用于遗传疾病的基于靶向核酸酶的治疗剂。目前用 于精确纠正原代患者细胞基因组中的基因突变的方法效率非常低(可以被精 确编辑的细胞少于1%)。本文提供的经修饰的sgRNA可增强基因组编辑的活 性、并增加基于基因组编辑的疗法的功效。在具体的实施方案中,修饰的 sgRNA可以用于具有遗传疾病的受试者中的基因的体内基因编辑。修饰的 sgRNA可经由任何合适的施用途径、以足以增强基于核酸酶的疗法的效果 (例如,提高基因组编辑效率)的剂量或量施用于受试者。

[0202] 本文提供了通过纠正与疾病相关的基因突变来预防或治疗有需要的受 试者的遗传疾病的方法。该方法包括以足以纠正突变的量向受试者施用本文 所述的修饰的sgRNA。本文还提供了本文所述的修饰的sgRNA在制备用于通 过纠正与疾病相关的基因突变来预防或治疗有需要的受试者的遗传疾病的 药物中的用途。修饰的sgRNA可以包含组合物中,所述组合物还包含Cas核 酸酶(例如Cas9多肽)或其变体或片段、编码Cas核酸酶(例如Cas9多肽)或其变 体或片段的mRNA、或包含编码Cas核酸酶(例如Cas9多肽)或其变体或片段的核苷酸序列的重组表达载体。在一些情况下,修饰的sgRNA被包括在上述递 送系统中。

[0203] 可通过该方法纠正的遗传疾病包括但不限于X连锁严重联合免疫缺陷、 镰状细胞性贫血、地中海贫血、血友病、瘤形成、癌症、年龄相关性黄斑变 性、精神分裂症、三核苷酸重复疾病、X染色体易裂症、朊病毒相关病症、 肌萎缩性侧索硬化、药物成瘾、自闭症、阿尔茨海默病、帕金森病、囊性纤 维化、血液和凝血疾病或病症、炎症、免疫相关疾病或病症、代谢疾病、肝 疾病和病症、肾脏疾病和病症、肌肉/骨骼疾病和病症(例如肌营养不良、杜 兴肌营养不良)、神经和神经元疾病和病症、心血管疾病和病症、肺部疾病和 病症、眼部疾病和病症、病毒感染(例如、HIV感染)等。

[0204] V.实施例

[0205] 提供以下实施例来说明而不是限制要求保护的发明。

[0206] 实施例1.化学修饰的引导RNA增强人原代细胞中的CRISPR/Cas基因组编辑。

[0207] CRISPR/Cas介导的基因组编辑依赖于引导RNA,引导RNA在Cas内切核 酸酶的促进下指导位点特异性DNA切割。在这里,我们报告了化学合成的单 引导RNA(sgRNA)的应用,并显示,通过化学改变修饰sgRNA,显著增强了 人原代T细胞和CD34+造血干细胞和祖细胞的基因组编辑。这种方法为基因 组编辑的流线化提供了一种简便、高效的途径,并且为加速CRISPR/Cas技术 在生物技术和治疗上的广泛应用提供了可能。

[0208] 利用工程化核酸酶编辑基因组是一项基本上可以修饰任何感兴趣的基 因组序列的突破性技术(Porteus,M.H.&Carroll,D.,Nature biotechnology 23,967-973(2005))。该技术利用工程核酸酶产生位点特异性双链断裂(DSB), 然后通过内源性细胞修复机制解决DSB。其结果或是特定位点通过诱变性非 同源末端连接(NHEJ)而被突变,在突变位点产生插入或缺失(in/dels),或是基 因组序列通过同源重组(HR)使用外源导入的供体模板而被精确改变(Hendel 等,Trends in Biotechnology 33,132-140(2015))。该项技术最近增加的一个主 要成员是成簇规则间隔的回文重复(CRISPR)/Cas系统,该系统由RNA引导的核酸酶(Cas)和短引导RNA(sgRNA)组成Jinek,M.et al.,Science 337,816-821 (2012),Mali,P.et al.,Science 339,823-826(2013),Cong,L.et al.,Science 339, 819-823(2013),Hsu et al.,Cell 157,1262-1278(2014))。引导RNA由两种RNA 组成,它们称为CRISPR RNA(crRNA)和反式激活crRNA(tracrRNA),为了基 因编辑目的,二者通常融合成嵌合的单引导RNA(sgRNA)。sgRNA由100个 核苷酸(nt)组成,其5'末端20nt可通过Watson-Crick碱基配对与靶DNA序列杂 交,并引导Cas内切核酸酶切割靶基因组DNA(图1A)。sgRNA可以以RNA的 形式递送到细胞中,例如,通过体外转录制备,或通过使用具有由RNA聚合 酶III启动子表达的sgRNA的DNA载体制备。尽管使用CRISPR/Cas系统的基 因组编辑在人细胞系中效率高,但是在原代人细胞中的CRISPR/Cas基因组编 辑通常更有难度。这种活性降低的原因仍然难以确定,但是转染率、启动子 活性、核酸外切酶活性、递送核酸时的天然干扰素免疫应答以及DNA修复保 真度等方面的差异可能是相关的促成因素。在这里,我们证明了化学合成的 sgRNA可以诱导高水平的基因组编辑,并进一步显示,化学改变的sgRNA可以显著增强人类原代T细胞和CD34+造血干细胞和祖细胞(HSPC)中的基因组 编辑。在化学修饰的sgRNA在这些细胞类型中导致基因组编辑增加的基础上, 通过以mRNA或蛋白质、而不是DNA表达质粒的形式递送Cas9,进一步提高 了该增加的效果,从而为CRISPR/Cas系统生成了一个简单而完整的基于 RNA或核糖核蛋白(RNP)的递送系统。

[0209] 结果

[0210] RNA合成技术的最新进展使得化学合成长度超过100nt的RNA变得可行(Dellinger,D.J.et al.,Journal of the American Chemical Society 133,11540-11556 (2011))。为了测试化学合成的sgRNA用于基因组编辑的效用,我们根据先前 述的程序(Dellinger,D.J.et al.,Journal of the American Chemical Society 133, 11540-11556(2011)),使用ABI 394合成仪和2'-O-硫代氨基甲酸酯保护的核苷 亚磷酰胺合成了100nt的全长sgRNA。我们还合成了在两个末端具有各种化学 修饰的sgRNA,以评估它们对功效的影响(图1A和1B以及表1)。当存在时,包 含2'-O-甲基(M)、2'-O-甲基3'硫代磷酸酯

(MS)、或2'-O-甲基3'硫代PACE(MSP) 的化学修饰物被纳入5'末端和3'末端的三个末端核苷酸中。选择这三种修饰 用于评估是因为它们先前被报道对血清和蛇毒磷酸二酯酶稳定，而且据报道 它们对核酸的免疫刺激性质有多种多样的影响(Deleavey,G.F.&Damha,M.J., Journal of the American Chemical Society 125,940-950(2003),Hendel et al.,Cell Reports 7,293-305(2014))。我们选择了先前报道的在细胞系中对以下各种人 基因产生高基因编辑率的三种sgRNA(Hendel et al.,Cell Reports 7,293-305 (2014),Cradick et al.,Nucleic Acids Research 41,9584-9592(2013)):(i)IL2RG，该基因中的突变是造成先天性原发性免疫缺陷SCID-X1的原因,(ii)HBB,该 基因中的突变导致镰状细胞性贫血和地中海贫血,和(iii)CCR5,该基因编码 HIV的一种共受体,目前正在研究其作为抗HIV临床试验中治疗性基因编辑 的靶标(Tebas,P.et al.,The New England Journal of Medicine 370,901-910 (2014))。我们首先测试了在纯化的重组Cas9蛋白和每种合成sgRNA的存在下 的体外靶DNA切割。所有化学合成的和修饰的sgRNA都在Cas9蛋白的存在下有效地切割了它们的DNA靶标(图3和4)。

[0211] 我们接下来检查了合成的sgRNA是否能够在人类细胞系中诱导生成可 指示诱变性NHEJ和基因破坏的目标in/del。我们通过核转染将每个sgRNA与 编码Cas9的DNA质粒一起递送到K562细胞中,并分析in/del频率。投递1μg合 成的未修饰的靶向IL2RG基因座的sgRNA后,产生了2.4%的in/del频率,证明 了其是有功能的(图1c,浅阴影条)。对于M修饰的sgRNA,我们观察到in/del 频率略微增加到13.5%,这表明稳定性比未修饰的sgRNA略有提高。引人注 目的是,相同数量的化学修饰的sgRNA使in/del频率分别增加到68.0%(MS修饰的sgRNA)和75.7%(MSP修饰的sgRNA)。将修饰的sgRNA的量增加到20倍 后,所有in/del频率均进一步增加,使MS和MSP sgRNA分别达到75.3%和83.3% (图1C,深色阴影条形),这与通过质粒表达CRISPR/Cas系统获得的频率可以 比较。对于HBB和CCR5靶标获得了非常相似的结果(图1C),这证明了化学修 饰的sgRNA具有一般性的在人类细胞中指导高水平的靶向基因突变的能力。我们接下来确定这些合成的sgRNA是否可以通过HR刺激基因靶向。我们为3 个基因座一一设计了靶向载体,其中CRISPR切割位点的5'和3'有约0.8kb的同 源臂。在同源臂之间,我们加入了一个GFP表达盒,在HR成功时,其可以在 稳定地整合在靶基因座上(Lombard.et al.,Nature Methods,8,861-869(2011); Voit et al.,Nucleic Acids Research 42,1365-1378(2014))。在所有三个靶基因座 上,MS和MSP sgRNA刺激的HR水平显著高于未修饰的sgRNA及M修饰的 sgRNA(图1D)。在更高的sgRNA水平(20μg)下,我们测得IL2RG,HBB和CCR5 处MSP sgRNA的HR率分别为20.6%,25.5%和50.0%。这些频率与完全由质粒 表达CRISPR/Cas系统获得的频率相当或更高。

[0212] 为了研究化学修饰的sgRNA是否影响脱靶活性,我们使用深度测序来测 量每个sgRNA的三个不同位点的脱靶突变频率。通过计算机预测工具(Hsu et al.,Nature Biotechnology 31,827-832(2013),Cradick et al.,Molecular therapy. Nucleic acids,3,e214(2014))预测了这些脱靶位点中的八个,而对于HBB sgRNA,我们纳入了一个脱靶位点,先前显示sgRNA对该脱靶位点具有高水 平脱靶活性(Cradick et al.,Nucleic Acid Research,42,9584-9592,(2013))。对于 八个预测位点中的四个,我们发现所有化学合成的sgRNA脱靶活性均接近背 景,即使检测到修饰的sgRNA的高水平的中靶活性(图1E和5,表5)。对于其 他四个预测位点,sgRNA的化学修饰倾向于导致更高的脱靶活性,但是活性

水平变化不定。在某些情况下,修饰的sgRNAs的中靶与脱靶的in/del频率之 比改善,提示相对特异性改善。这些结论的细节总结如下。对于未修饰的 sgRNA我们只检测到两个脱靶位点(IL2RG‘脱靶2’和HBB‘脱靶1’,先前已证 明HBB sgRNA对其显示出显著的脱靶活性)处的活性,这使我们能够比较未 修饰和修饰的sgRNA之间的中靶:脱靶比率。在IL2RG位点,未修饰的sgRNA 的中靶:脱靶比率与MSP sgRNA相比好5.8倍(图1E和表5),而对于HBB位点,MSP sgRNA的中靶:脱靶比率与未修饰的sgRNA相比好2.6倍和1.5倍(分别对 于1μg和20μg)(图1E和5,表5)。将修饰的sgRNA的中靶:脱靶比率与sgRNA质 粒的中靶:脱靶比率相比较,在CCR5‘脱靶2’和IL2RG‘脱靶2’处sgRNA质粒的 比率更好,而对于HBB‘脱靶1’,MSP sgRNA的中靶:脱靶比率更好(图1E和表 5)。除了HBB脱靶1位点之外,测得的最高脱靶频率是使用IL2RG sgRNA在 “脱靶2”处,当使用1μg和20μg MS sgRNA时,产生的in/del频率分别为1.0%和 7.8%。有趣的是,与MS sgRNA相比,相同位点的IL2RG MSP sgRNA尽管具 有更高的中靶活性,但脱靶活性分别低2.7倍和2.8倍。同样,MSP sgRNA在 HBB脱靶-1位点处——先前HBB sgRNA在该位点显示显著的脱靶活性—— 的脱靶活性相比MS sgRNA较好。在CCR5’脱靶2’处观察到相反的情况,与MS sgRNA相比,MSP sgRNA具有更高的脱靶活性。总的看来,这些结果提示, 通常情况下化学修饰的sgRNA保持高特异性。观察到的中靶:脱靶比率的差异 表明对sgRNA的化学改变可能具有调变中靶:脱靶靶活性的潜能,然而,特定 的化学改变的影响似乎是序列依赖性的,并且还可能取决于在其他因素,如 细胞类型和递送条件。这些观察结果是否可推广到针对不同物种中不同基因 座的其他sgRNAs,将需要进一步的研究。

[0213] 为了进一步探索化学修饰的sgRNA在人类细胞系中的性能,我们转向了 一种“全RNA”递送平台,其共同递送sgRNA和编码Cas9的mRNA。我们使用 不同量的Cas9 mRNA(1-15μg)以及不同量的MSP sgRNA(1-20μg)(图6),测量 了IL2RG基因座上的in/del频率。我们观察到,所有测试的浓度下均类似地具 有81-90%的高in/del比率,例外是当使用1μg的Cas9 mRNA和MSP sgRNA时, 比率稍微降低至70%,这显示这种“全RNA”方法的高效率。为了探索化学修 饰是否对sgRNA活性的半衰期有影响,我们进行了一个实验,其中我们通过 核转染共同递送或顺序递送Cas9 mRNA和各种合成的靶向IL2RG的 sgRNA(图7)。Cas9 mRNA与MS或MSP sgRNA的共同递送分别导致87%和84% 的in/dels的高编辑频率(图1F)。M sgRNA产生66%in/del,而未修饰的sgRNA 仅产生7.0%,这清楚地表明当与Cas9 mRNA共递送时sgRNA修饰的重要性。 有趣的是,对于所有四种sgRNA,先递送Cas9 mRNA,4或8小时后再递送sgRNA,产生了可比较的高水平in/del(83.1%-92.4%),效率在未修饰和化学 修饰的sgRNA之间没有差异(图1F)。相比之下,当先递送sgRNA,4、8、12 或24小时后再递送Cas9mRNA时,到4小时的时间点时用未修饰的sgRNA获 得的in/del频率已经下降到接近背景水平。对于M sgRNA,我们也观察到in/del 频率降低到接近背景水平,但是这个下降有稍微延迟,表明稳定性比未修饰 的sgRNA略有改善。对于MS sgRNA,直到24小时的时间点我们才观察到 in/del频率下降,此时in/del频率下降到53%。引人注目的是,对于MSP sgRNA,即使在sgRNA和Cas9递送之间的24小时延迟之后,我们也没有检测到活性显 著下降。这些结果符合我们的模型:sgRNA末端修饰提高细胞内稳定性,因 此可增加当Cas9 mRNA和sgRNA共同递送到人类细胞中时基因组编辑的功 效。当先递送Cas9时未修饰的sgRNA指导的in/del频率未降低这一事实提示 Cas9蛋白可保护sgRNA免于降解。为了研究这个假设,

我们将未修饰的 IL2RG sgRNA或MS IL2RG sgRNA与重组Cas9蛋白复合,然后以两种不同的量将这种活性RNP电穿孔到K562细胞中(图1G和8)。在此低RNP含量下,当 使用MS sgRNA时,我们观察到与未修饰的sgRNA相比3.8倍高的in/del频率 (35.7%比9.5%),而在高RNP量下,这种增加是2.3倍(81.0%比35.9%)。然而, MS和未修饰的sgRNA之间的这个比例显著小于当用sgRNA与Cas9 mRNA共 同递送时观察到的比例(与图1F'共递送'比较),表明Cas9蛋白确实可部分地保 护未修饰的sgRNA免于降解。尽管如此,当递送与Cas9蛋白预先复合的 sgRNA时,对sgRNA进行修饰仍能提高基因编辑效率。接下来,我们比较了 未修饰的IL2RG sgRNA和MS IL2RG sgRNA在和Cas9质粒、Cas9 mRNA一起 递送或与Cas9蛋白预先复合递送时,在三个先前研究的脱靶位点上的脱靶活 性。当使用未修饰的sgRNA时,我们在所有三个Cas9来源和全部三个脱靶位 点上均检测到低水平的in/del(<0.37%)(图9A-B)。对于MS sgRNA,我们观察 到,当递送Cas9质粒时,与递送Cas9 mRNA相比,在所有三个脱靶位点处的 中靶:脱靶比均有改善(2.6-3.0倍)。值得注意的是,在递送与MS sgRNA形成 的Cas9 RNP时,与Cas9质粒和mRNA相比,我们在所有三个位点上均检测到 显著更好的中靶-脱靶比,其中在两个位点处的脱靶活性接近背景。总之,当 与Cas9 mRNA共递送或作为RNP递送时,化学修饰的sgRNA与未修饰的 sgRNA相比在人细胞系中显示出显著的基因编辑优势。

[0214] 接下来我们在原代细胞中测试了化学修饰的sgRNA。目前正在临床试验 中探索使用锌指核酸酶(ZFN)破坏CD4+T细胞中的CCR5基因作为抗HIV治 疗方法(Tebas et al.New Eng Jour of Med,370,901-910(2014))。然而,最近的 一项研究发现,人类原代T细胞的基因组固有地难以使用单sgRNA的 CRISPR/Cas9质粒系统来编辑,而且能在细胞系中引起高频率等位基因修饰 的sgRNA在T细胞中的效力大大降低,即使对转染细胞加以富集也是如此 (Mandal et al.,Cell Stem Cell,15,643-652(2014))。我们在受刺激的人原代T细 胞中将经化学修饰的CCR5 sgRNA与编码Cas9的mRNA共递送来测试该 sgRNA。值得注意的是,在使用GFP mRNA时,通过流式细胞术测量,我们 始终观察到T细胞中超过98%的核转染效率(图10),因此无需富集转染细胞。 用编码sgRNA和Cas9二者的sgRNA质粒进行核转染并未导致高于背景的等 位基因修饰率(图2A)。然而,将MSP sgRNA与Cas9的DNA表达质粒共转染能 够将活性挽救到9.3%的in/del频率。用Cas9 mRNA与未修饰的sgRNA或M sgRNA进行核转染未导致高于背景的等位基因修饰率。相反,用MS或MSP sgRNA与Cas9 mRNA一起核转染分别产生了引人注目的48.7%和47.9%的 in/del频率。增加sgRNA或Cas9 mRNA的量未产生更高的in/del频率(图11)。对 于用Cas9 mRNA和MSP sgRNA核转染的细胞,我们在CD4+、CD8+及总T细 胞群中发现了相近的等位基因修饰频率(图12)。重要的是,对于MSP sgRNA,观察到的高修饰频率在历时21天的时间测量时是稳定的(图13),而且我们用 修饰的sgRNA进行核转染后对仅观察到对细胞生存力和增殖的微小影响,与 之形成对照的是,质粒核转染导致显著的细胞死亡和增殖潜力下降(图14和 15)。我们还在未受刺激的T细胞中测试了MS sgRNA,已经证明未刺激的T细 胞比受刺激的T细胞更难编辑(Yi et al.,Molecular Therapy Nucleic Acids,3, e198(2014))。与受刺激的T细胞形成对照的是,未受刺激的T细胞中的in/del 频率范围为6.6%至22.2%,显示更高的供体变异性(图16)。未受刺激的T细胞 中的这些编辑频率虽然低于受刺激的T细胞,但仍然可能有用处,特别是在 T细胞的工程化中,因为T细胞工程化中的活化和延长培养可能影响细胞的后 续生物功能。我们接

下来使用未修饰的sgRNA和MS sgRNA在受刺激的原代 T细胞中测试了RNP递送(图2B)。与在K562细胞中获得的结果类似,我们观 察到MS sgRNA的in/del频率比未修饰的sgRNA提高了2.4倍(30.7%对12.8%),这进一步支持了化学修饰的sgRNA以与Cas9蛋白预复合的形式递送时的用 处。

[0215] 人们对HSPC中的基因疗法用于治疗遗传性或获得性造血系统疾病已经 进行了大量探索。我们测试了靶向IL2RG和HBB的化学修饰sgRNA在从经动 员的外周血分离的CD34+HSPC中是否有功能。同样,表达sgRNA与Cas9二者 的sgRNA质粒未产生可检测的in/del频率,而用MSP sgRNA与Cas9 DNA表达 质粒共转染将对IL2RG和HBB的in/del频率分别挽救至3.2%和5.2%(图2C和 17)。与我们在T细胞中的观察结果相似,当使用未修饰的sgRNA或MsgRNA 与Cas9 mRNA共转染时,我们未在任一基因座处检测到in/dels,但MS和MSP sgRNA各自产生了高的in/del频率:对于IL2RG,MS sgRNA和MSP sgRNA的 in/del频率分别为17.5%和17.7%;而对于HBB的频率更高,分别为23.4%和 22.0%。进一步的研究可以阐明化学修饰的sgRNA对HSPC的遗传修饰是否影 响其多能性能力,或者长期再增殖干细胞(long-term repopulating stem cells)和 谱系定型的祖细胞(lineage-committed progenitor cells)之间编辑效率是否有所 不同。

[0216] 最近的一项研究显示同时使用两种sgRNA可改善人原代T细胞和 CD34+HSPC中的基因破坏(Mandal et al.,Cell Stem Cell,15,643-652(2014))。我们化学合成了MS和MSP CCR5 sgRNA,它们具有在该研究中报道的序列 (称为"D"和"Q"),而二者切割位置相隔205bp。我们将它们与Cas9 mRNA共 递送,在T细胞和CD34+HSPC中测试了它们。当两个sgRNA单独使用时,我 们使用TIDE分析定量了等位基因修饰频率。当使用两者时,我们通过对克隆 的PCR产物进行测序来定量等位基因修饰频率,这也使得我们能够定量全部 的编辑事件,包括之前报道的两个sgRNA靶位点之间序列缺失的高发生率。当在原代T细胞中单独使用时,"D"sgRNA分别对MS和MSP sgRNA产生56.0% 和56.3%的in/del,"Q"sgRNA分别产生62.6%和69.6%的in/del(图2D和18A)。当组合使用时,等位基因修饰的频率增加,因为我们分别观察到MS和MSP sgRNA分别为73.9%和93.1%,其中大部分修饰事件是两个sgRNA靶位点之间 的缺失(图18B)。在CD34+HSPC中,我们的观察结果相似,尽管总体频率较 低。因此,对于'D'sgRNA,我们观察到MS和MSP sgRNA分别有9.8%和11.2% 的等位基因修饰频率,'Q'sgRNA分别有17.8%和19.2%(图2D和18B)。当组合 使用时,频率对于MS和MSP sgRNA而言分别增加到37.8%和43.0%。我们得 出结论:使用两种化学修饰的sgRNA是一种高度有效的促进原代人T细胞和 CD34+HSPC中的基因破坏的方法。

[0217] 图19提供了另外的实验数据,其显示MS修饰的sgRNA在CD34+HSPC中 比相应的未修饰的sgRNA表现更好。图20显示修饰的sgRNA可以用于高效的 多重化基因组编辑。

[0218] 在该研究中,我们显示化学合成的sgRNA可以有效地用于靶向基因组编 辑,并证明化学修饰的sgRNA显著增强人原代T细胞和CD34+HSPC中的基因 组编辑效率。化学合成的sgRNA和修饰的sgRNA提供相比于表达的或体外转 录的sgRNAs提供了优势,包括:(1)功效增加;(2)稳健、可规模放大地生 产供生物技术和治疗应用的高纯度sgRNA;(3)sgRNA设计的灵活度更大,这 与通常用于质粒表达或体外转录sgRNA的U6或T7启动子对第一转录核苷酸 施加的约束形成了对照;(4)能够采用高活性的"仅RNA"或RNP CRISPR平台, 它们在原代细胞中具有比基于DNA质粒的系统更低的细胞毒性。将 CRISPR/Cas系统简化为纯RNA或

RNP CRISPR系统,使其适合于配制在不同 的纳米颗粒载体中供体内递送,使得核酸酶不会像其作为病毒载体的一部分 递送时那样连续表达。此外,我们预计化学修饰的sgRNA将增强多重基因组 编辑以及高通量多孔实验。我们也预计修饰的sgRNA将会完善多种CRISPR相关技术,例如用于抑制和激活基因表达的CRISPRi/CRISPRa系统(Gilbert, L.A.et al., Cell 159,647-661(2014));用于基因组座位动态可视化的CRISPR 成像工具(Chen,B.et al.,Cell 155,1479-1491(2013));和CRISPR介导的RNA 识别和切割(O'Connell,M.R.et al.,Nature 516,263-266(2014))。可进一步发 展该技术以改善对目标细胞或组织的细胞内递送,并为与各种分子缀合用于 成像和生化研究提供可能。未来的研究可以研究更多种类的化学修饰,探索 用于合理设计优化的sgRNA的不同位置的修饰,以及修饰sgRNA活性增强的 机制。总之,我们相信化学修饰的sgRNA,如本文中提供的这些,将会显著 改善多种CRISPR/Cas生物技术和治疗应用。

[0219] 材料和方法

[0220] 1.sgRNA合成

[0221] 所有RNA寡聚物均使用2'-O-硫羰氨基甲酸酯保护的核苷亚磷酰胺 (Sigma-Aldrich,St.Louis,MO,USA或Thermo Fisher,USA)在ABI 394合成 仪(Life Technologies,Carlsbad,CA,USA)上根据前人描述的程序(Dellinger et al.,Journal of the American Chemical Society 133,11540-11556(2011))合成。2'-O-甲基亚磷酰胺购自Thermo Scientific,Grand Island,NY,并在与2'-O-硫代 氨基甲酸酯保护的亚磷酰胺相同的条件下掺入RNA寡聚物中。基本上根据公 开的方法(Threlfall et al.,Organic& biomolecular chemistry,10,746-754(2012); Dellinger et al.,Journal of the American Chemical Society,125,940-950(2003)) 合成用于合成硫代膦酰基乙酸酯(thioPACE)修饰的RNA的2'-O-甲基-3'-O-(二 异丙基氨基)膦基乙酸-1,1-二甲基氰乙基酯-5'-O-二甲氧基三苯甲基核苷。对 于含硫代磷酸酯的寡聚物,将偶联反应后的碘氧化步骤用硫化步骤代替,使 用3-((N,N-二甲基氨基甲叉基)氨基)-3H-1,2,4-二噻唑-5-酮在吡啶-乙腈(3:2) 混合物中的0.05M溶液反应6分钟。除非另有说明,用于固相RNA合成的试剂购自Glen Research(Sterling,VA,USA)。

[0222] 使用反相HPLC纯化所有寡核苷酸,并使用与Agilent 6520Q-TOF质谱仪 (Agilent Technologies,Santa Clara,CA,USA)偶联的Agilent 1290Infinity系 列LC系统通过LC-MS进行分析。表1显示了使用的所有sgRNA的序列以及从 所发现的峰的电荷状态系列解卷积获得的质量。解卷积使用Mass Hunter定性 分析(版本B.06.00)软件(Agilent)完成。

[0223] 表1用于本研究的所有sgRNA的总览

[0224]

SEQ ID NO:	名称	序列	计算质量 (Da)	观察质量(Da)
SEQ ID NO: 4	HBB未修饰的sgRNA	CUUGCCCCACAGGGCAGUAAGUUUUAGAGCUAGAAAUAGC AAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGU GGCACCGAGUCGGUGCUUUU	32187.42	32187.84
SEQ ID NO: 5	HBB M sgRNA	CUUGCCCCACAGGGCAGUAAGUUUUAGAGCUAGAAAUAGC AAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGU GGCACCGAGUCGGUGCUUUU	32271.54	32271.35
SEQ ID NO: 6	HBB MS sgRNA	C•U•U•GCCCCACAGGGCAGUAAGUUUUAGAGCUAGAAAU AGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAA AGUGGCACCGAGUCGGUGCU•U•U•U	32367.42	32367.31
SEQ ID NO: 7	HBB MSP sgRNA	C♦U♦U♦GCCCCACAGGGCAGUAAGUUUUAGAGCUAGAAAU AGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAA AGUGGCACCGAGUCGGUGCU♦U♦U♦U	32619.93	32619.39
SEQ ID NO: 8	IL2G未修饰的sgRNA	UGGUAAUGAUGGCUUCAACAGUUUUAGAGCUAGAAAUAGC AAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGU GGCACCGAGUCGGUGCUUUU	32214.40	32214.37
SEQ ID NO: 9	IL2G M sgRNA	UGGUAAUGAUGGCUUCAACAGUUUUAGAGCUAGAAAUAGC AAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGU GGCACCGAGUCGGUGCUUUU	32298.52	32297.01
SEQ ID NO: 10	IL2G MS sgRNA	U•G•G•UAAUGAUGGCUUCAACAGUUUUAGAGCUAGAAAU AGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAA AGUGGCACCGAGUCGGUGCU•U•U•U	32394.4	32395.43
SEQ ID NO: 11	IL2G MSP sgRNA	U♦G♦G♦UAAUGAUGGCUUCAACAGUUUUAGAGCUAGAAAU AGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAA AGUGGCACCGAGUCGGUGCU♦U♦U♦U	32646.91	32645.39
SEQ ID NO: 12	CCR5未修饰的sgRNA	GGCAGCAUAGUGAGCCCAGAGUUUUAGAGCUAGAAAUAGC AAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGU GGCACCGAGUCGGUGCUUUU	32290.51	32289.07
SEQ ID NO: 13	CCR5 M sgRNA	GGCAGCAUAGUGAGCCCAGAGUUUUAGAGCUAGAAAUAGC AAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGU GGCACCGAGUCGGUGCUUUU	32374.63	32375.3
SEQ ID NO: 14	CCR5 MS sgRNA	G•G•C•AGCAUAGUGAGCCCAGAGUUUUAGAGCUAGAAAU AGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAA AGUGGCACCGAGUCGGUGCU•U•U•U	32470.51	32469.92
SEQ ID NO: 15	CCR5 MSP sgRNA	G♦G♦C♦AGCAUAGUGAGCCCAGAGUUUUAGAGCUAGAAAU AGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAA AGUGGCACCGAGUCGGUGCU♦U♦U♦U	32723.02	32721.96
SEQ ID NO: 16	CCR5 'D' MS sgRNA	U•C•A•CUAUGCUGCCGCCCAGUGUUUUAGAGCUAGAAAU AGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAA AGUGGCACCGAGUCGGUGCU•U•U•U	32281.32	32282.52

[0225]

SEQ ID NO: 17	CCR5 'D' MSP sgRNA	U♦C♦A♦CUAUGCUGCCGCCCAGUGUUUUAGAGCUAGAAAU AGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAA AGUGGCACCGAGUCGGUGCU♦U♦U♦U	32533.83	32533.55
SEQ ID NO: 18	CCR5 'Q' MS sgRNA	G•C•U•GUGUUUGCGUCUCUCCCGUUUUAGAGCUAGAAAU AGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAA AGUGGCACCGAGUCGGUGCU•U•U•U	32252.22	32253.21
SEQ ID NO: 19	CCR5 'Q' MSP sgRNA	G♦C♦U♦GUGUUUGCGUCUCUCCCGUUUUAGAGCUAGAAAU AGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAA AGUGGCACCGAGUCGGUGCU♦U♦U♦U	32504.73	32504.63

[0226] 标出了sgRNA序列以及计算和观察到的分子量。具有2'-O-甲基修饰的核 苷酸标有下划线。磷酸盐骨架的修改用●(MS)和◆(MSP)表示。

[0227] 2.体外切割测定

[0228] 通过对质粒携带的人序列进行制备性PCR扩增来制备4kb的PAM可寻址 靶物。在20μL反应体积中,在50nM sgRNA,39nM重组纯化的Cas9蛋白 (Agilent)和10mM $MgCl_2$,pH7.6

存在下,将50fmoles的线性化DNA靶物在37℃ 温育30分钟。完成后,加入0.5μL的RNace It (Agilent),并在37℃下继续温育 5分钟,然后在70℃下温育15分钟。随后加入0.5μL的蛋白酶K(分子生物学级, NEB)并在37℃温育15分钟。将等分试样加载到DNA 7500LabChip中,并在 Bioanalyzer 2200上分析。通过下式计算切割频率:(a/b)×100,其中'a'是两种 切割产物的条带强度的总和,b是切割的和未切割的DNA的条带强度的总和。

[0229] 3.质粒

[0230] 通过将20bp寡核苷酸靶序列克隆到含有人密码子优化的SpCas9表达盒 和驱动嵌合sgRNA表达的人U6启动子的px330(Addgene质粒#42230)中来构 建sgRNA表达载体(参见表1的sgRNA序列)。

[0231] 所有三种质粒靶向载体携带约2×800bp的同源臂,同源臂是利用从K562 细胞分离的基因组DNA PCR扩增相应基因座而产生的。然后使用标准克隆方 法将同源臂克隆到一个基于pBluescript SK+的约2900个碱基对的载体中。HBB和CCR5供体在两条同源臂之间均含有驱动GFP表达的EF1α启动子。IL2RG供体缺乏启动子,依赖于IL2RG基因的内源活性来驱动GFP表达。IL2RG靶向载体的核酸序列如SEQ ID NO:1所示。HBB靶向载体的核酸序列如SEQ ID NO:2所示。CCR5靶向载体的核酸序列如SEQ ID NO:3所示。

[0232] 4.细胞培养和核转染

[0233] 在37℃,5%CO_2和环境氧水平下培养K562和T细胞。在37℃,5%CO_2和 5%O_2下培养CD34+造血干/祖细胞(HSPC)。将K562细胞维持在补充有10%牛 生长血清、100mg/ml链霉素、100单位/ml青霉素和2mM L-谷氨酰胺的RPMI 1640(HyClone)中。使用Lonza Nucleofector 2b(程序T-016)和含有100mM KH_2PO_4、15mM $NaHCO_3$、12mM $MgCl_2\times6H_2O$、8mM ATP、2mM葡萄糖(pH 7.4)的核转染缓冲液核转染K562细胞。核转染条件:将100μL核转染溶液,10^6个细胞,1至20μg化学修饰的sgRNA,1至15μg Cas9 mRNA(Cas9mRNA, 5meC,Ψ,产品编号:L-6125,TriLink BioTechnologies,San Diego,CA, USA),2μg sgRNA/Cas9编码质粒或5μgHR供体质粒。使用人泛T细胞分离试 剂盒(Miltenyi Biotec,San Diego,CA,USA)从自斯坦福医学院血液中心获 得的血沉棕黄层分离CD3+T细胞。将CD3+细胞维持在补充有5%人血清 (Sigma-Aldrich,St.Louis,MO,USA)、100IU/mL人重组IL-2(Peprotech,Rocky Hill,NJ,USA)、和10ng/mL人重组IL-7(BD Biosciences,San Jose,CA,USA) 的X-VIVO15 (Lonza,Walkersville,MD,USA)中。在核转染之前,用固定的 抗CD3抗体(克隆:OKT3,eBioscience,San Diego,CA,USA)和可溶性抗 CD28抗体(克隆:CD28.2,eBioscience)活化T细胞三天。对于未激活的CD3+ T细胞,分离后立即核转染细胞。使用Lonza Nucleofector 2b(程序U-014)和人 T细胞Nucleofector试剂盒(VPA-1002,Lonza)对T细胞进行核转染。核转染条 件:将100μL核转染溶液、10^6个细胞、10-20μg化学修饰的sgRNA、15-30μg Cas9(或15μg eGFP mRNA,TriLink BioTechnologies,San Diego,CA,USA), 1μg sgRNA/Cas9编码质粒。经动员的人外周血CD34+HSPCs购自AllCells并 根据制造商的说明解冻。在补充有SCF(100ng/ml)、TPO(100ng/ml)、Flt3- 配体(100ng/ml)、IL-6(100ng/ml)、和StemRegenin1(0.75mM)的X-VIVO15 (Lonza)中维持CD34+HSPC。使用Lonza 4D-Nucleofector(程序EO-100)和P3 原代细胞4D-核转染试剂盒(V4XP-3024)对CD34+HSPC进行核转染。核转染 条件:100μL核转染溶液,5×10^5个细胞,10μg化学修饰的sgRNA,15μg Cas9mRNA,1μg质粒。对于Cas9 RNP实验,Cas9蛋白购自PNA Bio(Thousand Oaks,CA,USA)或

Life Technologies(Carlsbad,CA,USA)。对于除了图8之 外的所有RNP实验,使用来自PNA Bio的Cas9蛋白。将Cas9蛋白与sgRNA以 Cas9:sgRNA摩尔比1:2.5在25℃复合10分钟。如上所述用含有10^6个细胞的100 μL的相应核转染溶液将RNP核转染入K562细胞或T细胞中。对于双sgRNA实 验,总sgRNA量为10μg(单独使用时为10μg,一起使用时为2x 5μg)。对于T 细胞和CD34+HSPC二者,均将sgRNA与15μg Cas9 mRNA核转染至10^6个细 胞中。T细胞核转染如上所述进行,而CD34+HSPCs的核转染与T细胞核转染 相似,使用Lonza Nucleofector 2b(程序U-014)和人T细胞Nucleofector试剂盒 (VPA-1002,Lonza)进行。核转染后直接将CD34+HSPC在30℃温育24小时, 然后将其转移至37℃直至收获基因组DNA。

[0234] 5.流式细胞术和荧光激活细胞分选(FACS)

[0235] 对于HR实验,取决于细胞类型,在核转染后2-3周,当没有剩余的来自 附加型质粒的eGFP表达时对细胞进行分析。在Accuri C6流式细胞仪(BD Biosciences,San Jose,CA,USA)上测量eGFP表达。使用LIVE/DEAD Fixable Red Dead Cell Stain Kit(Life Technologies,Carlsbad,CA,USA)根据制造商 的说明测量细胞死亡,并在Accuri C6流式细胞仪上分析细胞。为了将CD3+ T细胞分选成CD4+和CD8+细胞群,用PE-Cy7抗人CD4(克隆:RPA-T4,Tonbo Biosciences,San Diego,CA,USA)和APC抗人CD8a(克隆:RPA-T8,Tonbo Biosciences)的混合物染色细胞,并在FACS Aria II SORP上分选这两个群体。

[0236] 6.使用TIDE和T7测定法测量等位基因修饰频率

[0237] 对于TIDE和T7测定,在核转染三天后(如果未另外指出的话)使用 QuickExtract DNA提取溶液(Epicentre,Madison,WI,USA)按照制造商的说 明从细胞中提取gDNA。使用iProof高保真母预混物(Bio-Rad,Hercules,CA, USA)和以下引物对产生跨越sgRNA基因组靶位点的PCR扩增子:

[0238] IL2RG_fw(SEQ ID NO:84):5'- TCACACAGCACATATTTGCCACACCCTCTG-3',

[0239] IL2RG_RV(SEQ ID NO:85):5'-TGCCCACATGATTGTAATGGCCAGTGG- 3',

[0240] HBB_fw(SEQ ID NO:86):5'-CCAACTCCTAAGCCAGTGCCAGAAGAG-3',

[0241] HBB_rv(SEQ ID NO:87)5'-AGTCAGTGCCTATCAGAAACCCAAGAG-3',

[0242] CCR5_fw(SEQ ID NO:88):5'-GCACAGGGTGGAACAAGATGG-3',

[0243] CCR5_rv(SEQ ID NO:89):5'-CACCACCCCAAAGGTGACCGT-3'。

[0244] 对于T7测定,纯化PCR扩增子,取200ng变性并在热循环仪中重新退火, 并根据制造商的方案用T7核酸内切酶I(New England Biolabs,Waltham,MA, USA)消化。将经过消化的DNA在4-20%TBE聚丙烯酰胺凝胶上电泳,用 Diamond Nucleic Acid Dye(Promega,Madison,WI,USA)染色,并在ChemiDoc XRS+(Bio-Rad)上可视化。使用Image Lab Software(Bio-Rad)分析条带强度, 并使用以下公式计算等位基因修饰频率:100×(1-(1-切割的分数)^0.5)。为了 使用TIDE(Tracking by In/dels by Decomposition,Brinkman et al,Nucleic Acids Research 42,e168(2014))分析等位基因修饰频率,使用两种引物对纯化的PCR 产物进行Sanger测序,并且在线TIDE软件(可在网站tide.nki.nl.访问)分析每个序列的色谱图。分析使用来自模拟转染的样品的参考序列进行。参数被设置 为:最大in/del大小为10个核苷酸,并且分解窗口覆盖具有高质量迹线的最大 可能窗口。所有低于3.5%的检测灵敏度的TIDE分析被设定为0%。对于TOPO 克隆的PCR片段的测序,使用iProof高保真母预混物和引物5'- GGCTGTTGTCATCTATGACCTTCCC-3'(SEQ ID NO:90)和5'-

TGTAAACTGAGCTTGCTCGCTCG-3’(SEQ ID NO:91)产生跨越切割位点的 2.1kb扩增子进行25个循环,每个循环包括72℃退火温度和2分钟延伸时间。使用Zero Blunt TOPO PCR克隆试剂盒(Life Technologies)根据制造商的方案 将PCR反应产物直接亚克隆到质粒中。将TOPO反应物转化到XL-1Blue感受 态细胞中,用卡那霉素涂布在琼脂平板上,由McLab(South San Francisco, CA,USA)直接从平板测序单个菌落,方法是滚环扩增后使用引物5’- GCACAGGGTGGAACAAGATGG-3’(SEQ ID NO:92)测序。

[0245] 7.增殖测定

[0246] 为了测量核转染后T细胞的增殖,使用CellTiter-Glo 2.0测定法(Promega,Madison,WI,USA)。在核转染之后即时,将T细胞以3×10^4个细胞/孔转移至 多个96孔U底96孔板。在核转染之后即时,以及每隔24小时,将100μL培养基 中的细胞转移至白色96孔板中的,并且按照制造商的指导添加100μL CellTiter-Glo 2.0。在Tecan Infinite 200PRO(Tecan,Männedorf,Switzerland)上 使用1秒的积分时间读取荧光。

[0247] 8.深度测序以定量基因组修饰的效率和特异性

[0248] 对于每个基因靶向实验,在转染48小时后从各种CRISPR处理K562细胞 和对照K562细胞中提取基因组DNA。通过两轮PCR扩增CRISPR靶标和三个 脱靶(表2)的两侧的基因组区域,以附接(处理特异性的)条形码和Illumina测序 衔接子(表3)。将加有条形码的PCR扩增子等摩尔汇集,通过旋转柱纯化并在 Illumina MiSeq DNA测序仪平台上测序。更多的细节参见下文的“9.生成 CRISPR中靶和预测的脱靶扩增子供深度测序”一节。

[0249] 表2提供了通过PCR扩增子的深度测序询问的中靶和脱靶的基因座的列 表。对于靶向CCR5、HBB和IL2RG的CRISPR实验,提供了预期的基因组靶 序列(“ON”)和三个计算预测的OFF-靶序列(“OFF1-3”),以及它们的基因组位 置(人类基因组构建组装体hg38)。所选择的脱靶序列是由COSMID(Cradick et al.,Molecular therapy.Nucleic acids 3,e214(2014))和Optimized CRISPR Design (Hsu et al.,Nature Biotechnology,31,827-832(2013))网上工具(MIT设计)共同 预测的得分最高者,例外是HBB-OFF3,它只被COSMID预测为具有显著活 性。预测的目标活性随着COSMID得分值的增加和“MIT设计”得分值的降低而增加,脱靶序列中的PAM位点和错配分别用红色和粗体文本表示。

[0250] 表2.通过PCR扩增子的深度测序询问的中靶及脱靶基因座的列表。

[0251]

SEQ ID NO:	靶ID	靶位点序列(5'→ 3')	基因组位置	链	靶得分	
					COSMID	MIT设计
	*CCR5*靶向引导RNA					
SEQ ID NO: 20	ON	GGCAGCATAGTGAGCCCAGA**AGG**	Chr3:46373153-46373175	-	0	56
SEQ ID NO: 21	OFF1	ATCATCATAGTGAGCCCAGA**GAG**	Chr3:15440658-15440680	+	0.44	2.4
SEQ ID NO: 22	OFF2	ACCAGCAGAGTGAGCCCAGA**GGG**	Chr4:3744369-3744391	+	0.52	2.6
SEQ ID NO: 23	OFF3	AGGAGCAGAGTGAGCCCAGA**GAG**	Chr15:92469456-92469478	+	0.54	2.6
	*HBB*靶向引导RNA					
SEQ ID NO: 24	ON	CTTGCCCCACAGGGCAGTAA**CGG**	Chr11:5226968-5226990	+	0	65
SEQ ID NO: 25	OFF1	TCAGCCCCACAGGGCAGTAA**GGG**	Chr9:101833584-101833606	+	0.4	2.3
SEQ ID NO: 26	OFF2	CCTCTCCCACAGGGCAGTAA**AGG**	Chr17:68628098-68628120	-	0.49	2.4
SEQ ID NO: 27	OFF3	TTTTCCCCAAAGGGCAGTAA**TAG**	Chr13:109165988-109166010	+	0.79	N/A
	*IL2RG*靶向引导RNA					
SEQ ID NO: 28	ON	TGGTAATGATGGCTTCAACA**TGG**	ChrX:71111519-71111541	+	0	49
SEQ ID NO: 29	OFF1	TGGGAAGGATGGCTTCAACA**CAG**	Chr7:151485304-151485326	-	0.4	3.9
SEQ ID NO: 30	OFF2	TGGTGAGGATGGCTTCAACA**CGG**	Chr1:167730172-167730194	-	0.42	3.7

[0252]

SEQ ID NO: 31	OFF3	TGGTAATGATGACTTCAACA**TAG**	Chr3:72764801-72764823	-	0.8	49.2

[0253] 表3提供了用于产生中靶与脱靶扩增子以通过深度测序定量in/del频率的 寡核苷酸引物的列表。分别用下划线和粗体显示了基因特异性扩增子引物的 基因特异性杂交序列和Illumina条形码引物的条形码。

[0254] 表3.用于产生中靶与脱靶扩增子以通过深度测序定量in/del频率的寡核 苷酸引物的列表。

[0255]

SEQ ID NO:	引物名称	引物序列(5'→ 3')
基因特异性扩增子引物		
SEQ ID NO: 32	CCR5_ON-fwd	CGACAGGTTCAGAGTTCTACAGTCCGACGATCCAAACACAGCATGGACGACA
SEQ ID NO: 33	CCR5_ON-rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGCTGAAGAGCATGACTGACA
SEQ ID NO: 34	CCR5_OFF1-fwd	CGACAGGTTCAGAGTTCTACAGTCCGACGATCGGGGAAGCAGTCTGGACTTAGA
SEQ ID NO: 35	CCR5_OFF1-rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCTGCCATTAAATCCACCAAA
SEQ ID NO: 36	CCR5_OFF2-fwd	CGACAGGTTCAGAGTTCTACAGTCCGACGATCAGCGAGTCGAGTTCAGGTG
SEQ ID NO: 37	CCR5_OFF2-rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGCTACCTACCCCAGGTTCT
SEQ ID NO: 38	CCR5_OFF3-fwd	CGACAGGTTCAGAGTTCTACAGTCCGACGATCTCTCACCAACACTGCCGAAT
SEQ ID NO: 39	CCR5_OFF3-rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTGGCATATAGTGCTCCCCACT
SEQ ID NO: 40	HBB_ON-fwd	CGACAGGTTCAGAGTTCTACAGTCCGACGATCTCTGTCTCCACATGCCCAGT
SEQ ID NO: 41	HBB_ON-rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAGGGCAGAGCCATCTATTG
SEQ ID NO: 42	HBB_OFF1-fwd	CGACAGGTTCAGAGTTCTACAGTCCGACGATCTCCCGTTCTCCACCCAATAG

[0256]

SEQ ID NO: 43	HBB_OFF1-rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGATTTCCAGGCTATGCTTCCA
SEQ ID NO: 44	HBB_OFF2-fwd	CGACAGGTTCAGAGTTCTACAGTCCGACGATCGTTGGCAGGGAGACTTACCA
SEQ ID NO: 45	HBB_OFF2-rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCCATGGTACGACTGTTCTCA
SEQ ID NO: 46	HBB_OFF3-fwd	CGACAGGTTCAGAGTTCTACAGTCCGACGATCTGGGGCCTTCAAGTGTTCTT
SEQ ID NO: 47	HBB_OFF3-rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTGTGCTCCTATGCCTGGTT
SEQ ID NO: 48	IL2RG_ON-fwd	CGACAGGTTCAGAGTTCTACAGTCCGACGATCCATTGGGCGTCAGAATTGTC
SEQ ID NO: 49	IL2RG_ON-rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAGGTTCTTTCCACCGGAAGC
SEQ ID NO: 50	IL2RG_OFF1-fwd	CGACAGGTTCAGAGTTCTACAGTCCGACGATCTCCGGAAGTTATTCAAGTCTGA
SEQ ID NO: 51	IL2RG_OFF1-rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTCTGGCATCAGAGCACAAA
SEQ ID NO: 52	IL2RG_OFF2-fwd	CGACAGGTTCAGAGTTCTACAGTCCGACGATCCCTGGGCCATATCAAGAGAC
SEQ ID NO: 53	IL2RG_OFF2-rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTCTTTGGGGTGATGTTTGTG
SEQ ID NO: 54	IL2RG_OFF3-fwd	CGACAGGTTCAGAGTTCTACAGTCCGACGATCCACAACAGTTGACCCAGGAA
SEQ ID NO: 55	IL2RG_OFF3-rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCCAACCCAGGTCTCTGAAC
Illumina条形码引物		
SEQ ID NO: 56	P5-BC_A-fwd	AATGATACGGCGACCACCGAGATCTACAC**TAGAGCTC**CGACAGGTTCAGAGTTCTACAGTCCGACGATC
SEQ ID NO: 57	P5-BC_B-fwd	AATGATACGGCGACCACCGAGATCTACAC**TCATAGCG**

[0257]

		CGACAGGTTCAGAGTTCTACAGTCCGACGATC
SEQ ID NO: 58	P5-BC_C-fwd	AATGATACGGCGACCACCGAGATCTACAC**GTAGCACT** CGACAGGTTCAGAGTTCTACAGTCCGACGATC
SEQ ID NO: 59	P5-BC_D-fwd	AATGATACGGCGACCACCGAGATCTACAC**GCTCATAG** CGACAGGTTCAGAGTTCTACAGTCCGACGATC
SEQ ID NO: 60	P5-BC_E-fwd	AATGATACGGCGACCACCGAGATCTACAC**ATGCATCG** CGACAGGTTCAGAGTTCTACAGTCCGACGATC
SEQ ID NO: 61	P5-BC_F-fwd	AATGATACGGCGACCACCGAGATCTACAC**AGTCGATC** CGACAGGTTCAGAGTTCTACAGTCCGACGATC
SEQ ID NO: 62	P5-BC-G_fwd	AATGATACGGCGACCACCGAGATCTACAC**CACTGTGA** CGACAGGTTCAGAGTTCTACAGTCCGACGATC
SEQ ID NO: 63	P5-BC-H_fwd	AATGATACGGCGACCACCGAGATCTACAC**CGAGTATC** CGACAGGTTCAGAGTTCTACAGTCCGACGATC
SEQ ID NO: 64	P7-BC-A_rev	CAAGCAGAAGACGGCATACGAGAT**TAGAGCTC**GTGAC TGGAGTTCAGACGTGTGCTCTTCCGATCT
SEQ ID NO: 65	P7-BC_B rev	CAAGCAGAAGACGGCATACGAGAT**TCATAGCG**GTGAC TGGAGTTCAGACGTGTGCTCTTCCGATCT
SEQ ID NO: 66	P7-BC_C rev	CAAGCAGAAGACGGCATACGAGAT**GTAGCACT**GTGAC TGGAGTTCAGACGTGTGCTCTTCCGATCT
SEQ ID NO: 67	P7-BC_D rev	CAAGCAGAAGACGGCATACGAGAT**GCTCATAG**GTGAC TGGAGTTCAGACGTGTGCTCTTCCGATCT
SEQ ID NO: 68	P7-BC_E_rev	CAAGCAGAAGACGGCATACGAGAT**ATGCATCG**GTGAC TGGAGTTCAGACGTGTGCTCTTCCGATCT
SEQ ID NO: 69	P7-BC_F_rev	CAAGCAGAAGACGGCATACGAGAT**AGTCGATC**GTGAC TGGAGTTCAGACGTGTGCTCTTCCGATCT
SEQ ID NO: 70	P7-BC-G_ rev	CAAGCAGAAGACGGCATACGAGAT**CACTGTGA**GTGAC TGGAGTTCAGACGTGTGCTCTTCCGATCT
SEQ ID NO: 71	P7-BC-H_rev	CAAGCAGAAGACGGCATACGAGAT**CGAGTATC**GTGAC TGGAGTTCAGACGTGTGCTCTTCCGATCT

[0258] 分别用下划线和粗体显示了基因特异性扩增子引物的基因特异性杂交 序列,以及Illumina条形码引物的条形码。

[0259] 为了分析测序数据,将来自不同处理的读段按照其相应的处理条形码进 行分拣,并使用BWA-MEM(Li et al.,Bioinformatics 26,589-595(2010))(bwa- 0.7.10)以默认参

数将它们映射至基因组。从分析中过滤掉不一致配对的末端 映射(间隔>1Kbp)以及由BWA-MEM中的默认参数定义的低质量映射读段和 二次比对读段。对于每个中靶和脱靶区域,我们计算了在预测的切割位点周 围具有in/del的读段的%,其中预测的切割位点估计位于PAM位点上游的第3 个和第4个位置之间。通过使用in/del[i]=(I[i]+D[i])/C[i]来计算基因组中 每个位置的in/del%,其中D[i]和I[i]分别表示在位置i具有任意大小的删除或插入的读段的数目,并且C[i]指示映射到任何包含位置i的基因组区间的读段 的数目。然后通过in/del[c]计算每个靶的in/del%,其中c是预测的剪切位点, 位于PAM位点上游第四个位置。在第4位置具有同型核苷酸序列,在IL2Rg中 为"AA",CCR5中为"CCC",以及插入事件的情况下,BWA_MEM比对器 无法解析插入的核苷酸的位置。这是通过用比对报告中的相关位置取代第4 位来加以纠正,具体而言,针对IL2Rg和CCR5分别采取位置3和位置6作为脱 靶位置。

[0260] 9.生成CRISPR中靶和预测的脱靶扩增子供深度测序

[0261] 对于每个靶基因座,用1μg或20μg合成的引导RNA(未修饰的、M、MS、或MSP),以及2μg的Cas9表达质粒(PX330)、2μg的编码sgRNA与Cas9蛋白二 者的sgRNA质粒(阳性对照)、或仅2μg的PX330(仅Cas9,阴性对照)转染10^6个K562细胞。此外,对于靶向IL2RG基因座的实验,用15μg Cas9 mRNA和 10μg合成的引导RNA(未修饰的或MS)、与7.6μg合成引导RNA(未修饰的或 MS)预先复合的15μg Cas9蛋白质转染10^6个K562细胞。在转染后72小时,使 用QuickExtract™DNA提取溶液(Epicentre,Madison,WI)根据制造商的说明 提取来自这些样品和模拟转染样品(第二阴性对照)的基因组DNA。使用40ng 基因组DNA作为模板来PCR扩增中靶的基因座和三个计算机预测的脱靶基 因座(表2),PCR扩增使用PfuUltra II HS 2x母预混物(Agilent Technologies, Santa Clara,CA)和基因特异性引物,这些引物让扩增子的末端带上MiSeq (Illumina,San Diego,CA)平台深度测序中使用的测序引物(表3)。对中靶和脱 靶扩增子(表4)进行第二次PCR反应以附接上额外的Illumina测序衔接子(例如P5、P7)和定制的双8-bp条形码,从而唯一地标识相应的转染处理。在第二次PCR之后,对带条形码的扩增子用Agilent D1000 TapeStation进行定量,以等 摩尔浓度汇集,然后用QIAquick PCR纯化试剂盒(Qiagen,Valencia,CA)根据 制造商的说明纯化。将经纯化的文库在Illumina MiSeq DNA测序仪上进行测 序,2×20^1个循环,通过NGS DNA测序服务(Seqmatic,Fremont,CA)进行双 重索引。

[0262] 表4提供了为了对in/del频率进行深度测序分析而产生的CCR5,HBB和 IL2RG中靶及脱靶扩增子的列表。扩增子(未经编辑的基因组DNA的)大小的 范围是183-220bp,从靶位点到基因特异性引物的杂交序列的距离最小为50 bp。下划线和粗体文本分别表示用于从基因组DNA产生扩增子的杂交序列和 推定的CRISPR-靶位点。

[0263] 表4.为进行in/del频率的深度测序分析而产生的CCR5、HBB和IL2RG中靶 及脱靶扩增子列表。

[0264]

SEQ ID NO:	靶ID	靶扩增子序列(5'→ 3')
	靶向*CCR5*的引导RNA	
SEQ ID NO: 72	ON	CAAACACAGCATGGACGACAGCCAGGTACCTATCGATTGTCAGGAGG ATGATGAAGAAGATTCCAGAGAAGAAGCCTATAAAATAGAGCCCTGT CAAGAGTTGACACATTGTATTTCCAAAGTCCCACTGGGC**GGCAGCAT AGTGAGCCCAGAAGG**GGACAGTAAGAAGGAAAAACAGGTCAGAGA TGGCCAGGTTGAGCAGGTAGATGTCAGTCATGCTCTTCAGCC
SEQ ID NO: 73	OFF1	GGGGAAGCAGTCTGGACTTAGAAAGGAAATAGGTGGTCTGTCATAGG GGCTTTCATTAGAGTTAAACTTCATAGAGTCAACTGTTTC**ATCATCAT AGTGAGCCCAGAGAG**CCACTGCCCAGCAGCATGCTCACACCACCTA CCCTAGTGTAGGTAATAGGTCTACGCTAGGACCCCGTGCTGGGCTCTC AGCCCATCATGAGATTTTGGTGGATTTAATGGCAGG
SEQ ID NO: 74	OFF2	AGCGAGTCGAGTTCAGGTGGGAGCAGAGGGCGCCCACCAGCAGAGC GAGTCGAGTCCAGGCGGGAGCAGAGGGCGCAC**ACCAGCAGAGTGA GCCCAGAGGG**TTTAAAGAAGGGGCGGTCTCTACGGTATGGGTAGAG TCAGGGGAACTAGGAAAGGACAGAGCAGAACCTGGGGTAGGTAGCC
SEQ ID NO: 75	OFF3	TCTCACCAACACTGCCGAATGTCATCTCTTCTCATCTTTATCTCTATTC TTTGCTTCCTGTCTTCAGGGCTCTTCCCTTGGCATTCACC**AGGAGCAG**

[0265]

		AGTGAGCCCAGAGAGCTGAGTGGTATCCCTTCTTCTTGGGTCCCTGA GCCCTGACCTGGAGCAATGCTGTGAGACAGCAGGAAAGGAGGGGAG TGTGGAGTGGGGAGCACTATATGCCA
	靶向*HBB*的引导RNA	
SEQ ID NO: 76	ON	TCTGTCTCCACATGCCCAGTTTCTATTGGTCTCCTTAAACCTGTCTTGT AACCTTGATACCAACCTGCCCAGGGCCTCACCACCAACTTCATCCACG TTCAC**CTTGCCCCACAGGGCAGTAACGG**CAGACTTCTCCTCAGGAG TCAGATGCACCATGGTGTCTGTTTGAGGTTGCTAGTGAACACAGTTGT GTCAGAAGCAAATGTAAGCAATAGATGGCTCTGCCCTG
SEQ ID NO: 77	OFF1	TCCCGTTCTCCACCCAATAGCTATGGAAAGGGGAAGATCCCAGAGAA CTTGGATAGGAAAGGTGAAGTCAGAGCAGTGCT**TCAGCCCCACAGG GCAGTAAGGG**CAGCCTTCCTCTAAATACCAGATTCCCAAATCTGGCT GTGCTTTCAATTTGGGAGTTGGACATACTGCTAAACTATAATTTCTTA GGCCGTACCTAAAATATATTATGGAAGCATAGCCTGGAAATC
SEQ ID NO: 78	OFF2	GTTGGCAGGGAGACTTACCAGCTTCCCGTATCTCCCTCCACATGGAGG CAGGACACGCTCTGGCCTTGCCCACCCTCCCACTAG**CCTCTCCCACA GGGCAGTAAAGG**TGAGTCTGGGAGAAAGAACCGGTCAGACTTAGTT CAGCTCCACCCTTTCCTCCTGGGAGTGAGTCTTTCCAAGACAGAGCAT GTTTTTTTCTACCCCTCAGTGAGAACAGTCGTACCATGGG
SEQ ID NO: 79	OFF3	TGGGGCCTTCAAGTGTTCTTCCCAAGAGTCAGAGTGAACCAGAACCA AGAACCATGTTGAGTTGCCCAGATGTAACCAGGCCTACAGGTACCTG GGAGAAACACGTGTACA**TTTTCCCCAAAGGGCAGTAATAG**CATCCT AGGCTTCAAAACATTCATAGAAACCATTTTTCAAATGCAAAGTCCAA CACAGTTAGAAATAACCAGGCATAGGAGCACAC
	靶向*IL2RG*的引导RNA	
SEQ ID NO: 80	ON	CATTGGGCGTCAGAATTGTCGTGTTCAGCCCCACTCCCAGCAGGGGC AGCTGCAGGAATAAGAGGGATGTGAA**TGGTAATGATGGCTTCAACA TGG**CGCTTGCTCTTCATTCCCTGGGTGTAGTCTGTCTGTGTCAGGAAC CTGGGTCCCTCACCCACTACCCCTCCCCACCCACACGTTTCCTCTGTC ATAGCTTCCGGTGGAAAGAACCT

[0266]

SEQ ID NO: 81	OFF1	TCCGGAAGTTATTCAAGTCTGATTTTCTTTCCTCCCTTGTCAGGGAAA AGAAGTTGTGACAAATTGCTTGGATCCTTAAGCTTAAG**TGGGAAGGA** **TGGCTTCAACACAG**AACATCTGTTTCATTGCTGTTTTATCCGTCAGTA AAACTGTTACTTCTTTTATGTACTAAAAGTTCTTAGCACTTAACTAAT ATTAGCTCTTTGTGCTCTGATGCCAGAC
SEQ ID NO: 82	OFF2	CCTGGGCCATATCAAGAGACTCTGCCTCAAAAAAGAAAAGAAAGAA AGAAAAAGAAAAAAAAAAGAACATCATTAAAAATCCCTGAGGAGCA TTTAGAGTATTGGGTGGCACAAACAGATTCTGCATGAT**TGGTGAGGA** **TGGCTTCAACACGG**CAGCTTTATTCCTCTTTAACAGAGTCAGCAGCA TCAAGGCATGAGGGATCTTGGCACAAACATCACCCCAAAGA
SEQ ID NO: 83	OFF3	CACAACAGTTGACCCAGGAACAGGGGGAACCTCCCACCATTCCCATC CCACTGTTTGATCAGATCCAAGAATCCACAATATTGAGAGTGAATGA AAAGTGTCAGC**TGGTAATGATGACTTCAACATAG**TCAGAACTCTTT GGGGTGTTCCAAACATCATGGTGCATATGTATTACCTGGGAGTCTTGT TAAAAAGACTCCTGTTCAGAGACCTGGGTTGGG

[0267] 扩增子大小(未经编辑的基因组DNA的)的范围从183-220bp,从靶位点到 基因特异性引物的杂交序列的距离最小为50bp。下划线和粗体文本分别表示 用于从基因组DNA产生扩增子的杂交序列和推定的CRISPR靶位点。

[0268] 10.化学修饰的sgRNAs的中靶和脱靶活性

[0269] 表5示出图1C、图1E和图5所依据的数字。如图1C中所进行的由合成 sgRNA介导的靶向切割的特异性,使用2μg Cas9质粒和1μg sgRNA(上图)或 20μg sgRNA(下图)sgRNA。通过对来自靶向基因组基因座的PCR扩增子以及 每个基因的三个生物信息学预测的脱靶基因座进行深度测序来测量In/del频 率。显示了平均值+/-SEM,n=3。

[0270] 表5

[0271]

2 μg Cas9 质粒和 1 μg sgRNA						
		模拟	**未修饰**	**M**	**MS**	**MSP**
IL2RG	**中靶**	0.16 ±0.03	2.43 ±0.23	13.47 ±0.54	68 ±1.06	75.73 ±1.30
	脱靶 1	0.00	0.00	0.00	0.00	0.00

[0272]

	脱靶 2	0.00	0.00	0.10 ±0.00	1.00 ±0.10	0.37 ±0.03
	脱靶 3	0.00	0.00	0.00	0.00	0.00
HBB	中靶	0.15 ±0.05	1.3 ±0.08	12.4 ±0.28	48.17 ±0.39	38.68 ±0.85
	脱靶 1	0.01 ±0.01	0.84 ±0.07	7.01 ±0.36	27.31 ±0.60	9.51 ±0.25
	脱靶 2	0.00	0.00	0.01 ±0.00	0.00	0.00
	脱靶 3	0.00	0.00	0.00	0.00	0.00
CCR5	中靶	0.02 ±0.01	4.26 ±0.18	3.41 ±0.23	24.6 ±1.25	22.78 ±1.76
	脱靶 1	0.01 ±0.00	0.01 ±0.00	0.01 ±0.00	0.01 ±0.00	0.02 ±0.00
	脱靶 2	0.01 ±0.01	0.01 ±0.00	0.01 ±0.00	0.08 ±0.03	0.54 ±0.10
	脱靶 3	0.01 ±0.00	0.01 ±0.00	0.01 ±0.00	0.01 ±0.00	0.01 ±0.00

[0273]

2 μg Cas9 质粒和 20 μg sgRNA						
		未修饰	M	MS	MSP	sgRNA质粒
IL2RG	中靶	23.4 ±1.8	48.13 ±0.4	75.3 ±5.1	83.27 ±0.7	70.53 ±0.01
	脱靶 1	0.00	0.03 ±0.03	0.13 ±0.03	0.00	0.00
	脱靶 2	0.13 ±0.03	1.20 ±0.06	7.83 ±0.58	2.77 ±0.24	0.07 ±0.00
	脱靶 3	0.00	0.10 ±0.00	0.30 ±0.00	0.10 ±0.00	0.17 ±0.00
HBB	中靶	19.42 ±0.27	40.99 ±1.4	65.91 ±0.62	60.71 ±0.25	31.11 ±0.04
	脱靶 1	9.2 ±0.38	33.56 ±2.3	55.1 ±0.97	19.08 ±0.49	25.44 ±0.04
	脱靶 2	0.00	0.00	0.00	0.00	0.00
	脱靶 3	0.00	0.00	0.00	0.00	0.00
CCR5	中靶	11.04 ±2.5	14.87 ±2.1	56.3 ±3.3	52.19 ±4.5	45.86 ±0.03
	脱靶 1	0.02 ±0.01	0.01 ±0.00	0.01 ±0.00	0.01 ±0.00	0.01 ±0.00
	脱靶 2	0.01 ±0.00	0.02 ±0.00	1.75 ±0.18	5.32 ±0.68	0.02 ±0.00
	脱靶 3	0.01 ±0.00	0.01 ±0.00	0.02 ±0.01	0.02 ±0.00	0.01 ±0.00

[0274] 实施例2.使用化学修饰的引导RNA和合成的单链寡聚脱氧核苷酸(ssODN)模板的基于CRISPR/Cas9的同源重组。

[0275] 将来自三个不同供体的受刺激的人原代T细胞用10μg CCR5 sgRNA(未 修饰的或MS),15μg Cas9 mRNA和2.81μg的183nt CCR5 ssODN(两端的三 个末端核苷酸之间有或者

没有硫代磷酸酯(“PS”)联接)核转染。ssODN含有一 个不存在于WT CCR5序列中的中央HindIII限制性位点。在核转染三天后提取 基因组DNA(gDNA),产生跨越靶位点(在与ssODN同源的序列的外部)的PCR 产物,并用HindIII消化。在2%TBE琼脂糖凝胶上分析限制片段并计算HDR频 率。

[0276] 图21示出了来自HDR实验的琼脂糖凝胶。当使用修饰的sgRNA时,HDR 频率有所增加。另外,当ssODN两端的三个末端核苷酸之间包含硫代磷酸酯 联接时,HDR频率进一步增加。

[0277] 尽管为了清楚理解的目的,已经通过举例说明和实施例的方式较详细地 描述了前述发明,但是本领域技术人员将会理解,在所附权利要求的范围内 可以实施某些改变和修改。此外,本文提供的每篇参考文献通过引用整体并 入,如同每篇参考文献单独导入作为参考一样。

[0278] 序列表

[0279] SEQ ID NO:1

[0280] 靶向IL2RG的载体序列

ctaaattgtaagcgttaatattttgttaaaattcgcgttaaatttttgttaaatcagctcatttttaaccaatag

gccgaaatcggcaaaatcccttataaatcaaaagaatagaccgagatagggttgagtgttgttccagtttggaaca

agagtccactattaaagaacgtggactccaacgtcaaagggcgaaaaaccgtctatcagggcgatggcccactacg

tgaaccatcaccctaatcaagtttttttggggtcgaggtgccgtaaagcactaaatcggaaccctaaagggagcccc

cgatttagagcttgacggggaaagccggcgaacgtggcgagaaaggaagggaagaaagcgaaaggagcgggcgcta

gggcgctggcaagtgtagcggtcacgctgcgcgtaaccaccacacccgccgcgcttaatgcgccgctacagggcgc

gtcccattcgccattcaggctgcgcaactgttgggaagggcgatcggtgcgggcctcttcgctattacgccagctg

gcgaaaggggatgtgctgcaaggcgattaagttgggtaacgccagggttttcccagtcacgacgttgtaaaacga

cggccagtgagcgcgcgtaatacgactcactatagggcgaattgggtacctaactataacggtcctaaggtagcga

ttaattaa***tgggagaaacaccacagaagcagagtgggttatattctctgggtgagagaggggagaaattgaagct***

gattctgaggtttcaagtctgggtgactgagagggtgacgataccattgactgaggtggggaaggcaggaagagaa

gcagagttgggggaagatgggaagcttgaagctagtattgttgttcctccatttctagaatattttgtattataa

gtcacacttcctcgccagtctcaacagggacccagctcaggcagcagctaagggtgggtattctggtttggattag

atcagaggaaagacagctgtatatgtgcccacaggagccaagacggtattttccatcctcccaaaacagtagagct

ttgacagagatttaagggtgaccaagtcaaggaagaggcatggcatagaacggtgatgtcgggggtgggggttcag

aacttccattatagaaggtaatgatttagaggagaaggtggttgagaatggtgctagtggtagtgaacagatcctt

cccaggatctaggtgggctgaggatttttgagtctgtgacactattgtatatccagctttagtttctgtttaccac

cttacagcagcacctaatctcctagaggacttagcccgtgtcacacagcacatatttgccacaccctctgtaaagc

cctggtttataaggttctttccaccggaagctatgacagaggaaacgtgtgggtggggagggggtagtgggtgaggg

acccaggttcctgacacagacagactacacccagggaatgaagagcaagcgcctctagaattaccctgttatccct

aaggtaagtatcaaggttacaagacaggtttaaggaggccaatagaaactgggcttgtcgagacagagaagattct

tgcgtttctgataggcacctattggtcttactgacatccactttgcctttctctccacaggggtaccgaagccgct

agcgctaccggtcgccacc*ATGCCCGCCATGAAGATCGAGTGCCGCATCACCGGCACCCTGAACGGCGTGGAGTTC*

GAGCTGGTGGGCGGCGGAGAGGGCACCCCCGAGCAGGGCCGCATGACCAACAAGATGAAGAGCACCAAAGGCGCCC

TGACCTTCAGCCCCTACCTGCTGAGCCACGTGATGGGCTACGGCTTCTACCACTTCGGCACCTACCCCAGCGGCTA

CGAGAACCCCTTCCTGCACGCCATCAACAACGGCGGCTACACCAACACCCGCATCGAGAAGTACGAGGACGGCGGC

GTGCTGCACGTGAGCTTCAGCTACCGCTACGAGGCCGGCCGCGTGATCGGCGACTTCAAGGTGGTGGGCACCGGCT

TCCCCGAGGACAGCGTGATCTTCACCGACAAGATCATCCGCAGCAACGCCACCGTGGAGCACCTGCACCCCATGGG

CGATAACGTGCTGGTGGGCAGCTTCGCCCGCACCTTCAGCCTGCGCGACGGCGGCTACTACAGCTTCGTGGTGGAC

AGCCACATGCACTTCAAGAGCGCCATCCACCCCAGCATCCTGCAGAACGGGGGCCCCATGTTCGCCTTCCGCCGCG

TGGAGGAGCTGCACAGCAACACCGAGCTGGGCATCGTGGAGTACCAGCACGCCTTCAAGACCCCCATCGCCTTCGC

*CAGATCTCGAGCTCG*Atgactcgagggcgcgccccgctgatcagcctcgactgtgccttctagttgccagccatct

gttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcctaataaaatgagg

aaattgcatcgcattgtctgagtaggtgtcattctattctgggggggtggggtggggcaggacagcaaggggggagga

ttgggaagacaatagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaacgtttaaacac

tagt***gaagagcaagcgccatgttgaagccatcattaccattcacatccctcttattcctgcagctgcccctgctgg***

gagtggggctgaacacgacaattctgacgcccaatgggaatgaagacaccacagctggtgggaaatctgggactgg

aggggctggtgagaagggtggctgtgggaaggggccgtacagagatctggtgcctgccactggccattacaatca

tgtgggcagaattgaaaagtggagtgggaagggcaaggggagggttccctgcctcacgctacttcttctttcttt

cttgtttgtttgtttctttctttctttgaggcagggtctcactatgttgcctaggctggtctcaaactcctggcc

tctagtgatcctcctgcctcagcctttcaaagcaccaggattacagacatgagccaccgtgcttggcctcctcctt

ctgaccatcatttctctttccctccctgccttcattttctccccaatctagatttcttcctgaccactatgcccac

[0281] ***tgactccctcagtgtttccactctgcccctcccagaggttcagtgttttgtgttcaatgtcgagtacatgaattgc***

acttggaacagcagctctgagccccagcctaccaacctcactctgcattattggtatgagaagggacgaggggggag

gggatgaagaagaggtgggttggatcagagaccaagagagagggtagcaagtctcccaggtaccccactgttttct

cctggggtaagtcataagtcggtgccatttcattacctctttctccgcacccgacatagatgagctccagcttttg

ttccctttagtgagggttaattgcgcgcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccg

ctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactca

cattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggcca

acgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgtt

cggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaa

agaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggct

ccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagatac

caggcgtttcccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcct

ttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctc

caagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtcc

aacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcg

gtgctacagagttcttgaagtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgct

gaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttt

tttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctg

acgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatcct

tttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgctta

atcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataa

ctacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccaga

tttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccag

tctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgcta

[0282] caggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttac

atgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgca

gtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtga

ctggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacg

ggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctca

aggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactt

tcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatg

ttgaatactcatactcttccttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacata

tttgaatgtatttagaaaaataaacaaatagggttccgcgcacatttccccgaaaagtgccac

[0283] IL2RG同源臂以加粗及斜体表示。嵌合内含子加下划线。GFP序列以大写字 母及斜体表示。BGH多聚(A)序列加双下划线。

[0284] SEQ ID NO:2

[0285] 靶向HBB的载体序列

ctaaattgtaagcgttaatattttgttaaaattcgcgttaaatttttgttaaatcagctcatttttaaccaatag

gccgaaatcggcaaaatcccttataaatcaaaagaatagaccgagatagggttgagtgttgttccagtttggaaca

agagtccactattaaagaacgtggactccaacgtcaaagggcgaaaaaccgtctatcagggcgatggcccactacg

tgaaccatcaccctaatcaagttttttggggtcgaggtgccgtaaagcactaaatcggaaccctaaagggagcccc

cgatttagagcttgacggggaaagccggcgaacgtggcgagaaaggaagggaagaaagcgaaaggagcgggcgcta

gggcgctggcaagtgtagcggtcacgctgcgcgtaaccaccacacccgccgcgcttaatgcgccgctacagggcgc

gtcccattcgccattcaggctgcgcaactgttgggaagggcgatcggtgcgggcctcttcgctattacgccagctg

gcgaaaggggatgtgctgcaaggcgattaagttgggtaacgccagggttttcccagtcacgacgttgtaaaacga

cggccagtgagcgcgcgtaatacgactcactatagggcgaattgggtacctaactataacggtcctaaggtagcga

ttaattaa***aggcagaaacagttagatgtccccagttaacctcctatttgacaccactgattaccccattgatagtc***

acactttgggttgtaagtgacttttttatttatttgtatttttgactgcattaagaggtctctagttttttatctct

tgtttcccaaaacctaataagtaactaatgcacagagcacattgatttgtatttattctatttttagacataattt

attagcatgcatgagcaaattaagaaaaacaacaacaaatgaatgcatatatatgtatatgtatgtgtgtatatat

acacacatatatatatatttttttcttttcttaccagaaggttttaatccaaataaggagaagatatgcttagaa

ccgaggtagagttttcatccattctgtcctgtaagtattttgcatattctggagacgcaggaagagatccatctac

atatcccaaagctgaattatggtagacaaaactcttccacttttagtgcatcaacttcttatttgtgtaataagaa

aattgggaaaacgatcttcaatatgcttaccaagctgtgattccaaatattacgtaaatacacttgcaaaggagga

tgtttttagtagcaatttgtactgatggtatggggccaagagatatatcttagagggagggctgagggtttgaagt

ccaactcctaagccagtgccagaagagccaaggacaggtacggctgtcatcacttagacctcaccctgtggagcca

caccctagggttggccaatctactcccaggagcagggagggcaggagccagggctgggcataaaagtcagggcaga

gccatctattgcttacatttgcttctgacacaactgtgttcactagcaacctcaaacagacaccatggtgcacctg

actcctgaggagaagtctgccgttactgcccagatctgggctccggtgcccgtcagtgggcagagcgcacatcgcc

cacagtccccgagaagttgggggggagggtcggcaattgaaccggtgcctagagaaggtggcgcggggtaaactgg

gaaagtgatgtcgtgtactggctccgccttttcccgagggtgggggagaaccgtatataagtgcagtagtcgccg

tgaacgttcttttcgcaacgggtttgccgccagaacacaggtaagtgccgtgtgtggttcccgcgggcctggcct

ctttacgggttatggcccttgcgtgccttgaattacttccacctggctgcagtacgtgattcttgatcccgagctt

cgggttggaagtgggtgggagagttcgaggccttgcgcttaaggagcccttcgcctcgtgcttgagttgaggcct

ggcctgggcgctggggccgccgcgtgcgaatctggtggcaccttcgcgcctgtctcgctgctttcgataagtctct

agccatttaaaatttttgatgacctgctgcgacgctttttttctggcaagatagtcttgtaaatgcgggccaagat

ctgcacactggtatttcggtttttggggccgcgggcggcgacggggcccgtgcgtcccagcgcacatgttcggcga

ggcggggcctgcgagcgcggccaccgagaatcggacgggggtagtctcaagctggccggcctgctctggtgcctgg

cctcgcgccgccgtgtatcgccccgccctgggcggcaaggctggcccggtcggcaccagttgcgtgagcggaaaga

tggccgcttcccggccctgctgcagggagctcaaaatggaggacgcggcgctcgggagagcgggcgggtgagtcac

ccacacaaaggaaaagggcctttccgtcctcagccgtcgcttcatgtgactccacggagtaccgggcgccgtccag

gcacctcgattagttctcgagcttttggagtacgtcgtctttaggttggggggaggggttttatgcgatggagttt

ccccacactgagtgggtggagactgaagttaggccagcttggcacttgatgtaattctccttggaatttgcccttt

ttgagtttggatcttggttcattctcaagcctcagacagtggttcaaagttttttttcttccatttcaggtgtcgtg

aggtaccgagctcttcgaaggatccatcgccacc*ATGCCCGCCATGAAGATCGAGTGCCGCATCACCGGCACCCTG*

AACGGCGTGGAGTTCGAGCTGGTGGGCGGCGGAGAGGGCACCCCCGAGCAGGGCCGCATGACCAACAAGATGAAGA

GCACCAAAGGCGCCCTGACCTTCAGCCCCTACCTGCTGAGCCACGTGATGGGCTACGGCTTCTACCACTTCGGCAC

CTACCCCAGCGGCTACGAGAACCCCTTCCTGCACGCCATCAACAACGGCGGCTACACCAACACCCGCATCGAGAAG

TACGAGGACGGCGGCGTGCTGCACGTGAGCTTCAGCTACCGCTACGAGGCCGGCCGCGTGATCGGCGACTTCAAGG

TGGTGGGCACCGGCTTCCCCGAGGACAGCGTGATCTTCACCGACAAGATCATCCGCAGCAACGCCACCGTGGAGCA

[0286] *CCTGCACCCCATGGGCGATAACGTGCTGGTGGGCAGCTTCGCCCGCACCTTCAGCCTGCGCGACGGCGGCTACTAC*

AGCTTCGTGGTGGACAGCCACATGCACTTCAAGAGCGCCATCCACCCCAGCATCCTGCAGAACGGGGGCCCCATGT

TCGCCTTCCGCCGCGTGGAGGAGCTGCACAGCAACACCGAGCTGGGCATCGTGGAGTACCAGCACGCCTTCAAGAC

*CCCCATCGCCTTCGCCAGATCTCGAGTC*tagctcgagggcgcgcccgctgatcagcctcgacctgtgccttctagt

tgccagccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcct

aataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctgggggggtggggtggggcaggacag

caaggggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaaga

acgtttcgcgcc***tgtggggcaaggtgaacgtggatgaagttggtggtgaggccctgggcaggttggtatcaaggtt***

acaagacaggtttaaggagaccaatagaaactgggcatgtggagacagagaagactcttgggtttctgataggcac

tgactctctctgcctattggtctattttcccacccttaggctgctggtggtctacccttggacccagaggttcttt

gagtcctttggggatctgtccactcctgatgctgttatgggcaaccctaaggtgaaggctcatggcaagaaagtgc

tcggtgcctttagtgatggcctggctcacctggacaacctcaagggcacctttgccacactgagtgagctgcactg

tgacaagctgcacgtggatcctgagaacttcagggtgagtctatgggacgcttgatgttttctttccccttctttt

ctatggttaagttcatgtcataggaaggggataagtaacagggtacagtttagaatgggaaacagacgaatgattg

catcagtgtggaagtctcaggatcgttttagtttcttttatttgctgttcataacaattgttttcttttgtttaat

tcttgctttctttttttttcttctccgcaatttttactattatacttaatgccttaacattgtgtataacaaaagg

aaatatctctgagatacattaagtaacttaaaaaaaaactttacacagtctgcctagtacattactatttggaata

tatgtgtgcttatttgcatattcataatctccctactttatttctttttatttttaattgatacataatcattata

catatttatgggttaaagtgtaatgttttaatatgtgtacacatattgaccaaatcagggtaattttgcatttgta

attttaaaaaaatgctttcttcttttaatatacttttttgtttatcttatttctaatactttccctaatctctttct

ttcagggcaataatgatacaatgtatcatgcctctttgcaccattctaaagaatcctgcaggggagtgagcagctg

taagatttgaggggcgactccgattagtttatcttcccacggactagagttggtgtcgaggtgccatttcattacc

tctttctccgcacccgacatagatgagctccagcttttgttccctttagtgagggttaattgcgcgcttggcgtaa

tcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataa

agtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtc

gggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctct

tccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcgg

taatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaa

ccgtaaaaaggccgcgttgctggcgtttttccataggctccgccccctgacgagcatcacaaaaatcgacgctca

agtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctc

ctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctc

[0287] acgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaacccccgttcagccc

gaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcag

ccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacgg

ctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctct

tgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaag

gatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggatttt

ggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagt

atatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttc

gttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgc

tgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgag

cgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagtt

cgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggc

ttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctcc

ttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataatt

ctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtg

tatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtg

ctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaac

ccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggca

[0288] aaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttccttttcaatattattga

agcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggttc

cgcgcacatttccccgaaaagtgccac

[0289] HBB同源臂加粗并以斜体表示。EF1α启动子加下划线。GFP序列大写并以斜 体表示。BGH多聚(A)序列加双下划线。

[0290] SEQ ID NO:3

[0291] 靶向CCR5的载体序列

ctaaattgtaagcgttaatattttgttaaaattcgcgttaaatttttgttaaatcagctcatttttaaccaatag

gccgaaatcggcaaaatcccttataaatcaaaagaatagaccgagatagggttgagtgttgttccagtttggaaca

agagtccactattaaagaacgtggactccaacgtcaaagggcgaaaaaccgtctatcagggcgatggcccactacg

tgaaccatcaccctaatcaagttttttggggtcgaggtgccgtaaagcactaaatcggaaccctaaagggagcccc

cgatttagagcttgacggggaaagccggcgaacgtggcgagaaaggaagggaagaaagcgaaaggagcgggcgcta

gggcgctggcaagtgtagcggtcacgctgcgcgtaaccaccacacccgccgcgcttaatgcgccgctacagggcgc

gtcccattcgccattcaggctgcgcaactgttgggaagggcgatcggtgcgggcctcttcgctattacgccagctg

gcgaaaggggatgtgctgcaaggcgattaagttgggtaacgccagggttttcccagtcacgacgttgtaaaacga

cggccagtgagcgcgcgtaatacgactcactatagggcgaattgggtacctaactataacggtcctaaggtagcga

ttaattaa***gaacagtgattggcatccagtatgtgccctcgaggcctcttaattattactggcttgctcatagtgca***

[0292] ***tgttctttgtgggctaactctagcgtcaataaaaatgttaagactgagttgcagccgggcatggtggctcatgcct***

gtaatcccagcattctaggaggctgaggcaggaggatcgcttgagcccaggagttcgagaccagcctgggcaacat

agtgtgatcttgtatctataaaaataaacaaaattagcttggtgtggtggcgcctgtagtccccagccacttggag

gggtgaggtgagaggattgcttgagcccgggatggtccaggctgcagtgagccatgatcgtgccactgcactccag

cctgggcgacagagtgagaccctgtctcacaacaacaacaacaacaacaacaaaaaggctgagctgcaccatgcttgac

ccagtttcttaaaattgttgtcaaagcttcattcactccatggtgctatagagcacaagattttatttggtgagat

ggtgctttcatgaattcccccaacagagccaagctctccatctagtggacagggaagctagcagcaaaccttccct

tcactacaaaacttcattgcttggccaaaaagagagttaattcaatgtagacatctatgtaggcaattaaaaacct

attgatgtataaaacagtttgcattcatggagggcaactaaatacattctaggactttataaaagatcacttttta

tttatgcacagggtggaacaagatggattatcaagtgtcaagtccaatctatgacatcaattattatagcgcctct

agaattaccctgttatccctaagatctgggctccggtgcccgtcagtgggcagagcgcacatcgcccacagtcccc

gagaagttgggggggaggggtcggcaattgaaccggtgcctagagaaggtggcgcggggtaaactgggaaagtgatg

tcgtgtactggctccgccttttttcccgagggtgggggagaaccgtatataagtgcagtagtcgccgtgaacgttct

ttttcgcaacgggtttgccgccagaacacaggtaagtgccgtgtgtggttcccgcgggcctggcctctttacgggt

tatggcccttgcgtgccttgaattacttccacctggctgcagtacgtgattcttgatcccgagcttcgggttggaa

gtgggtgggagagttcgaggccttgcgcttaaggagcccttcgcctcgtgcttgagttgaggcctggcctgggcg

ctggggccgccgcgtgcgaatctggtggcaccttcgcgcctgtctcgctgctttcgataagtctctagccatttaa

aatttttgatgacctgctgcgacgctttttttctggcaagatagtcttgtaaatgcgggccaagatctgcacactg

gtatttcggttttggggccgcgggcggcgacggggcccgtgcgtcccagcgcacatgttcggcgaggcggggcct

gcgagcgcggccaccgagaatcggacgggggtagtctcaagctggccggcctgctctggtgcctggcctcgcgccg

ccgtgtatcgccccgccctgggcggcaaggctggcccggtcggcaccagttgcgtgagcggaaagatggccgcttc

ccggccctgctgcagggagctcaaaatggaggacgcggcgctcgggagagcgggcgggtgagtcacccacacaaag

gaaaagggcctttccgtcctcagccgtcgcttcatgtgactccacggagtaccgggcgccgtccaggcacctcgat

tagttctcgagcttttggagtacgtcgtctttaggttgggggggagggggttttatgcgatggagtttccccacactg

agtgggtggagactgaagttaggccagcttggcacttgatgtaattctccttggaatttgccctttttgagtttgg

[0293] atcttggttcattctcaagcctcagacagtggttcaaagtttttttcttccatttcaggtgtcgtgaggtaccgag

ctcttcgaaggatccatcgccaccatg*GTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCG*

AGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCT

GACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCTTCGGCTACGGC

CTGATGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACG

TCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACAC

CCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTAC

AACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCC

ACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCT

GCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTC

*CTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAG*taactcgagggcgcgccta

atcaacctctggattacaaaatttgtgaaagattgactggtattcttaactatgttgctccttttacgctatgtgg

atacgctgctttaatgcctttgtatcatgctattgcttcccgtatggctttcattttctcctccttgtataaatcc

tggttgctgtctctttatgaggagttgtggcccgttgtcaggcaacgtggcgtggtgtgcactgtgtttgctgacg

caacccccactggttggggcattgccaccacctgtcagctcctttccgggactttcgctttcccctccctattgc

cacggcggaactcatcgccgcctgccttgcccgctgctggacaggggctcggctgttgggcactgacaattccgtg

gtgttgtcggggaaatcatcgtcctttccttggctgctcgcctgtgttgccacctggattctgcgcgggacgtcct

tctgctacgtcccttcggccctcaatccagcggaccttccttcccgcggcctgctgccggctctgcggcctcttcc

gcgtcttcgccttcgccctcagacgagtcggatctccctttgggccgcctccccgcctggcgcgccccgctgatca

gcctcgactgtgccttctagttgccagccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggtg

ccactcccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctggg

gggtggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctct

atggcttctgaggcggaaagaacgtttaaacactagt***ctatgacatcaattattatacatcggagccctgccaaaa***

aatcaatgtgaagcaaatcgcagcccgcctcctgcctccgctctactcactggtgttcatctttggttttgtgggc

aacatgctggtcatcctcatcctgataaactgcaaaaggctgaagagcatgactgacatctacctgctcaacctgg

ccatctctgacctgtttttccttcttactgtccccttctgggctcactatgctgccgcccagtgggactttggaaa

tacaatgtgtcaactcttgacagggctctattttataggcttcttctctggaatcttcttcatcatcctcctgaca

atcgataggtacctggctgtcgtccatgctgtgtttgctttaaaagccaggacggtcacctttggggtggtgacaa

gtgtgatcacttgggtggtggctgtgtttgcgtctctcccaggaatcatctttaccagatctcaaaaagaaggtct

tcattacacctgcagctctcattttccatacagtcagtatcaattctggaagaatttccagacattaaagatagtc

atcttggggctggtcctgccgctgcttgtcatggtcatctgctactcgggaatcctaaaaactctgcttcggtgtc

[0294] ***gaaatgagaagaagaggcacagggctgtgaggcttatcttcaccatcatgattgtttatttctcttctgggctcc***

ctacaacattgtccttctcctgaacaccttccaggaattctttggcctgaataattgcagtagctctaacaggttg

gaccaagctatgcaggtgactgccatttcattacctctttctccgcacccgacatagatgagctccagcttttgtt

ccctttagtgagggttaattgcgcgcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgct

cacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcaca

ttaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaac

gcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcg

gctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaag

aacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctcc

gcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagatacca

ggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgccttt

ctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctcca

agctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaa

cccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggt

gctacagagttcttgaagtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctga

agccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggttttt

[0295]
tgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgac
gctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttt
taaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaat
cagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataact
acgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatt
tatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtc
tattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctaca
ggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacat
gatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagt
gttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgact
ggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacggg
ataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaag
gatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttc
accagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgtt
gaatactcatactcttccttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatt
tgaatgtatttagaaaaataaacaaatagggttccgcgcacatttccccgaaaagtgccac

[0296] HBB同源臂加粗并以斜体表示。EF1α启动子加下划线。GFP序列大写并以斜 体表示。WPRE序列加下划短线。BGH多聚(A)序列加双下划线。

[0297] SEQ ID NOS:4-19

[0298] 见表1

[0299] SEQ ID NOS:20-31

[0300] 见表2

[0301] SEQ ID NOS:32-71

[0302] 见表3

[0303] SEQ ID NOS:72-83

[0304] 见表4

[0305] SEQ ID NO:84

[0306] IL2RG_fw:5’-TCACACAGCACATATTTGCCACACCCTCTG-3’,

[0307] SEQ ID NO:85

[0308] IL2RG_RV:5’-TGCCCACATGATTGTAATGGCCAGTGG-3’

[0309] SEQ ID NO:86

[0310] HBB_fw:5’-CCAACTCCTAAGCCAGTGCCAGAAGAG-3’

[0311] SEQ ID NO:87

[0312] HBB_rv:5’-AGTCAGTGCCTATCAGAAACCCAAGAG-3’

[0313] SEQ ID NO:88

[0314] CCR5_fw:5’-GCACAGGGTGGAACAAGATGG-3’

[0315] SEQ ID NO:89

[0316] CCR5_rv:5’-CACCACCCCAAAGGTGACCGT-3’

[0317] SEQ ID NO:90

[0318] 5’-GGCTGTTGTCATCTATGACCTTCCC-3’

[0319] SEQ ID NO:91

[0320] 5’-TGTAAACTGAGCTTGCTCGCTCG-3’

[0321] SEQ ID NO:92

[0322] 5’-GCACAGGGTGGAACAAGATGG-3’

[0323] 示例实施方案

[0324] 1.一种诱导原代细胞中靶核酸的基因调控的方法,所述方法包括:

[0325] 向原代细胞中导入:

[0326] (a)修饰的单引导RNA(sgRNA),其包含与靶核酸互补的第一核苷酸序 列和与CRISPR相关蛋白(Cas)多肽相互作用的第二核苷酸序列,其中第一核 苷酸序列和/或第二核苷酸序列中的一个或多个核苷酸是修饰的核苷酸;和

[0327] (b)Cas多肽、编码Cas多肽的mRNA、和/或包含编码Cas多肽的核苷酸序 列的重组表达载体,

[0328] 其中所述修饰的sgRNA将Cas多肽引导至靶核酸,

[0329] 其中所述修饰的sgRNA以相对于相应的未修饰的sgRNA增强的活性诱 导所述靶核酸的基因调控。

[0330] 2.实施方案1的方法,其中所述增强的活性包括:修饰的sgRNA的稳定 性增加,和/或修饰的sgRNA对靶核酸的特异性增加。

[0331] 3.实施方案1或2的方法,其中所述靶核酸包含靶DNA或靶RNA。

[0332] 4.实施方案3的方法,其中所述基因调控包括对靶DNA的基因组编辑。

[0333] 5.实施方案4的方法,其中所述基因组编辑包括对靶DNA的同源性指导 修复(HDR)或非同源末端连接(NHEJ)。

[0334] 6.实施方案4或5的方法,还包括将重组供体修复模板导入原代细胞。

[0335] 7.实施方案6的方法,其中所述重组供体修复模板包含两个核苷酸序列, 所述两个核苷酸序列包含靶DNA的两个非重叠的、同源的部分,其中这两个 核苷酸序列位于与要进行基因组编辑的靶DNA对应的核苷酸序列的5’端和3’ 端。

[0336] 8.实施方案6的方法,其中所述重组供体修复模板包含合成单链寡聚脱 氧核苷酸(ssODN)模板,该模板包含:编码用于纠正单核苷酸多态性(SNP)的 突变的核苷酸序列,和包含靶DNA的两个非重叠的、同源的部分的两个核苷 酸序列,其中这两个核苷酸序列位于所述编码突变的核苷酸序列的5’端和3’ 端。

[0337] 9.实施方案3的方法,其中所述基因调控包括使用内切核酸酶缺陷的Cas 多肽来抑制或激活靶DNA或靶RNA的基因表达。

[0338] 10.实施方案1至9中任一项的方法,其中所述原代细胞在将所述修饰的 sgRNA和Cas多肽导入原代细胞之前从多细胞生物体中分离。

[0339] 11.实施方案10的方法,其中多细胞生物是植物、多细胞原生生物、多 细胞真菌或动物。

[0340] 12.实施方案1至11中任一项的方法,其中原代细胞选自下组:干细胞、免疫细胞、及其组合。

[0341] 13.实施方案12的方法,其中所述干细胞选自下组:造血干细胞和祖细 胞(HSPC)、间充质干细胞、神经干细胞、器官干细胞、及其组合。

[0342] 14.实施方案12的方法,其中所述免疫细胞选自下组:T细胞、自然杀伤 细胞、单核细胞、外周血单个核细胞(PBMC)、外周血淋巴细胞(PBL)、和它 们的组合。

[0343] 15.实施方案10至14中任一项的方法,其中在将所述修饰的sgRNA和所 述Cas多肽导入原代细胞中之后,该原代细胞或其后代被返回至所述多细胞 生物体。

[0344] 16.实施方案1至15中任一项的方法,其中所述原代细胞包括原代细胞群 体。

[0345] 17.实施方案16的方法,其中所述修饰的sgRNA在所述原代细胞群体的 至少约30%中诱导靶核酸的基因调控。

[0346] 18.实施方案16的方法,其中所述修饰的sgRNA在所述原代细胞群体的 至少约40%中诱导靶核酸的基因调控。

[0347] 19.实施方案16的方法,其中所述修饰的sgRNA在所述原代细胞群体的 至少约50%中诱导靶核酸的基因调控。

[0348] 20.实施方案16的方法,其中所述修饰的sgRNA在所述原代细胞群体的 至少约60%中诱导靶核酸的基因调控。

[0349] 21.实施方案1至20中任一项的方法,其中所述修饰的核苷酸在核糖基团、磷酸基团、核碱基或其组合中包含修饰。

[0350] 22.实施方案21的方法,其中核糖基团中的修饰包括核糖基团的2'位上 的修饰。

[0351] 23.实施方案22的方法,其中核糖基团的2'位上的修饰选自下组:2'-O-甲 基、2'-氟、2'-脱氧、2'-O-(2-甲氧基乙基)、及其组合。

[0352] 24.实施方案21的方法,其中磷酸基团中的修饰包括硫代磷酸酯修饰。

[0353] 25.实施方案1至24中任一项的方法,其中所述修饰的核苷酸选自下组: 2'-O-甲基(M)核苷酸,2'-O-甲基3'-硫代磷酸酯(MS)核苷酸,2'-O-甲基3'-硫代 PACE(MSP)核苷酸、及其组合。

[0354] 26.实施方案1至25中任一项的方法,其中所述第一核苷酸序列的长度为 约20个核苷酸。

[0355] 27.实施方案1至26中任一项的方法,其中所述第一核苷酸序列中的至少 2,3,4,5,6,7,8,9,10个或更多个核苷酸是修饰的核苷酸。

[0356] 28.实施方案1至27中任一项的方法,其中所述修饰的核苷酸位于所述第 一核苷酸序列的5'端或其附近,和/或在所述第一核苷酸序列内的内部位置。

[0357] 29.实施方案1至28中任一项的方法,其中所述第一核苷酸序列中的约10% 至约30%的核苷酸是修饰的核苷酸。

[0358] 30.实施方案1至29中任一项的方法,其中所述第二核苷酸序列的长度为 约80个核苷酸。

[0359] 31.实施方案1至30中任一项的方法,其中所述第二核苷酸序列中的至少 2,3,4,5,6,7,8,9,10个或更多个核苷酸是修饰的核苷酸。

[0360] 32.实施方案1至31中任一项的方法,其中所述修饰的核苷酸位于所述第 二核苷

酸序列的3'端或其附近,和/或在所述第二核苷酸序列内的内部位置。

[0361] 33.实施方案1至32中任一项的方法,其中所述第二核苷酸序列中的约1% 至约10%的核苷酸是修饰的核苷酸。

[0362] 34.实施方案1至33中任一项的方法,其中所述修饰的sgRNA在所述第一 核苷酸序列的5'末端或其附近包含1个、2个或3个连续或不连续的修饰的核苷 酸,并且在所述第二核苷酸序列的3'末端或其附近包含1个、2个或3个连续或 不连续的修饰的核苷酸。

[0363] 35.实施方案34的方法,其中所述修饰的sgRNA在所述第一核苷酸序列 的5'末端包含3个连续的修饰的核苷酸,并且在所述第二核苷酸序列的3'末端 包含3个连续的修饰的核苷酸。

[0364] 36.实施方案1至35中任一项的方法,其中所述修饰的sgRNA是化学合成 的。

[0365] 37.实施方案1至36中任一项的方法,其中所述修饰的sgRNA包含至少两 种不同的修饰的sgRNA,其中每种修饰的sgRNA针对不同的靶核酸。

[0366] 38.实施方案1至37中任一项的方法,其中所述Cas多肽是编码Cas多肽的 mRNA。

[0367] 39.实施方案1至38中任一项的方法,其中所述Cas多肽是Cas9多肽、其 变体或其片段。

[0368] 40.实施方案1至39中任一项的方法,其中导入原代细胞的步骤包括将原 代细胞电穿孔。

[0369] 41.一种预防或治疗受试者的遗传疾病的方法,所述方法包括:

[0370] 向受试者施用足够量的修饰的单引导RNA(sgRNA)以纠正与遗传疾病 相关的靶基因中的突变,其中修饰的sgRNA包含与靶基因互补的第一核苷酸 序列和与CRISPR相关蛋白(Cas)多肽相互作用的第二核苷酸序列,并且其中 第一核苷酸序列和/或第二核苷酸序列中的一个或多个核苷酸是修饰的核苷 酸。

[0371] 42.实施方案41的方法,其中所述遗传疾病选自下组:X-连锁重症联合 免疫缺陷,镰状细胞贫血,地中海贫血,血友病,瘤形成,癌症,HIV感染, 年龄相关性黄斑变性,精神分裂症,三核苷酸重复疾病,X染色体易裂症, 朊病毒相关疾病,肌萎缩性侧索硬化,药物成瘾,孤独症,阿尔茨海默病, 帕金森氏病,囊性纤维化,血液和凝血疾病或病症,炎症,免疫相关性疾病 或病症,代谢性疾病,肝脏疾病和病症,肾脏疾病和病症,肌肉/骨骼疾病和 病症,神经和神经元疾病和病症,心血管疾病和病症,肺部疾病和病症,以 及眼部疾病和病症。

[0372] 43.实施方案41或42的方法,还包括向受试者施用Cas多肽、编码Cas多 肽的mRNA、和/或包含编码Cas多肽的核苷酸序列的重组表达载体。

[0373] 44.实施方案43的方法,其中Cas多肽是Cas9多肽、其变体或其片段。

[0374] 45.实施方案41至44中任一项的方法,还包括向受试者施用重组供体修 复模板。

[0375] 46.实施方案41至45中任一项的方法,其中相比于施用相应的未修饰的 sgRNA,施用修饰的sgRNA增强所述Cas多肽纠正所述靶基因中的突变的效 果。

[0376] 47.实施方案45或46的方法,其中修饰的sgRNA、Cas多肽和/或重组供 体修复模板与药学上可接受的载体一起施用给受试者。

[0377] 48.实施方案45至47中任一项的方法,其中修饰的sgRNA、Cas多肽和/ 或重组供体修复模板经由选自下组的递送系统施用给受试者:纳米颗粒、脂 质体、胶束、病毒微体、核

酸复合物、及其组合。

[0378] 49.实施方案48的方法，其中所述核酸复合物包含与所述Cas多肽复合的 所述修饰的sgRNA。

[0379] 50.实施方案45至49中任一项的方法，其中所述修饰的sgRNA、Cas多肽 和/或重组供体修复模板经由选自下组的途径施用给受试者：口服、静脉内、 腹膜内、肌肉内、皮内、皮下、小动脉内、心室内、颅内、病灶内、鞘内、 局部、透粘膜、鼻内及其组合。

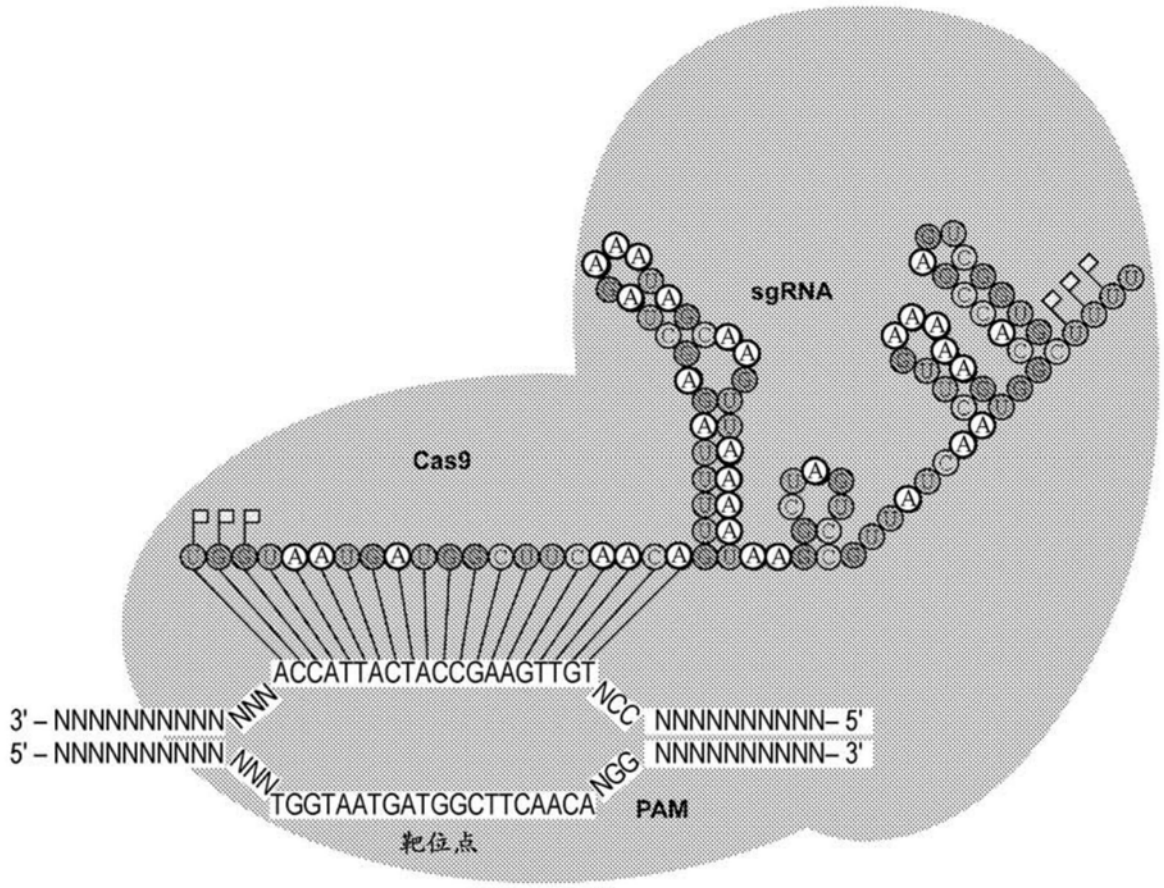

sgRNA
Cas9
ACCATTACTACCGAAGTTGT
3' – NNNNNNNNNN NNN
5' – NNNNNNNNNN NNN
NCC
NGG
NNNNNNNNNN– 5'
NNNNNNNNNN– 3'
TGGTAATGATGGCTTCAACA
PAM
靶位点

图1A

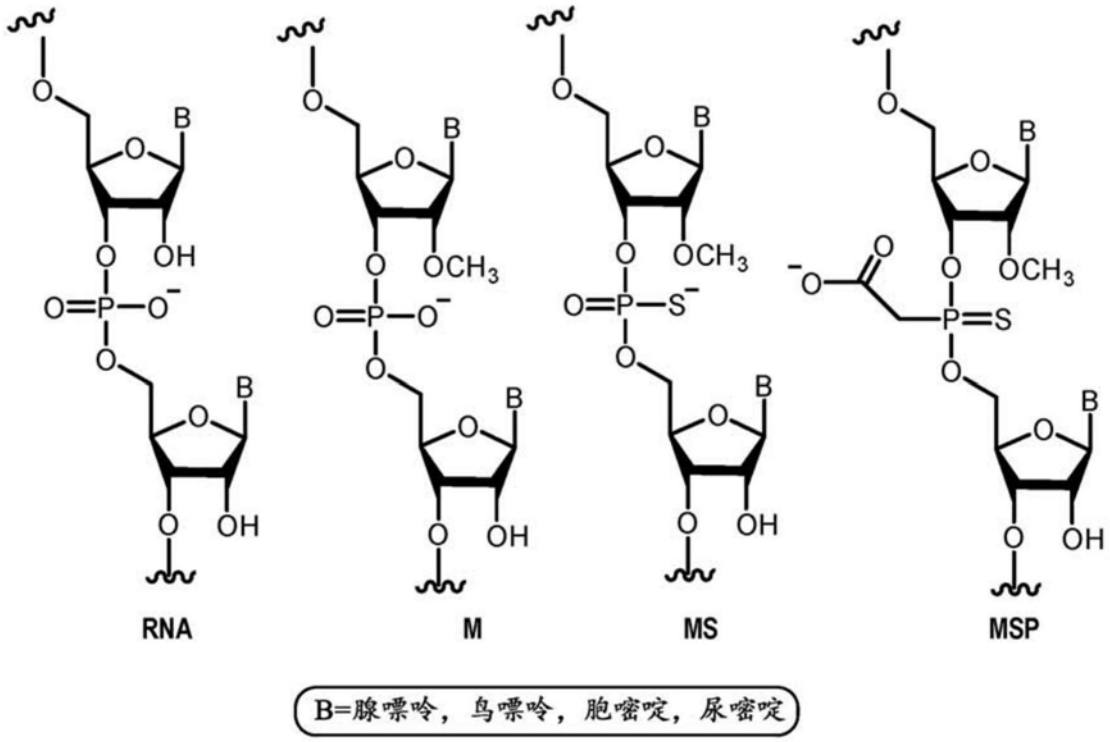

RNA
M
MS
MSP
B=腺嘌呤，鸟嘌呤，胞嘧啶，尿嘧啶

图1B

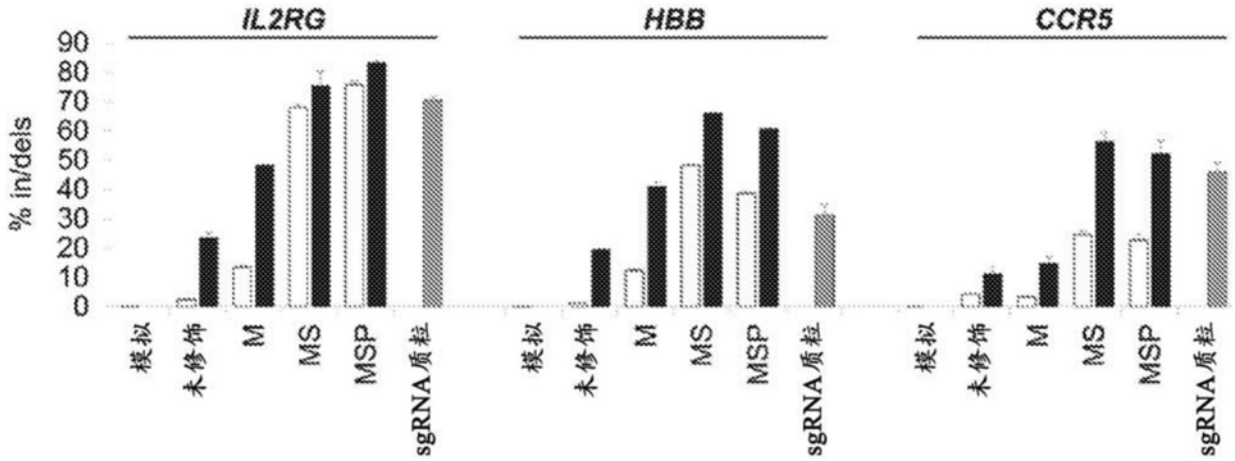

IL2RG
HBB
CCR5
% in/dels
90
80
70
60
50
40
30
20
10
0
模拟
未修饰
M
MS
MSP
sgRNA质粒

图1C

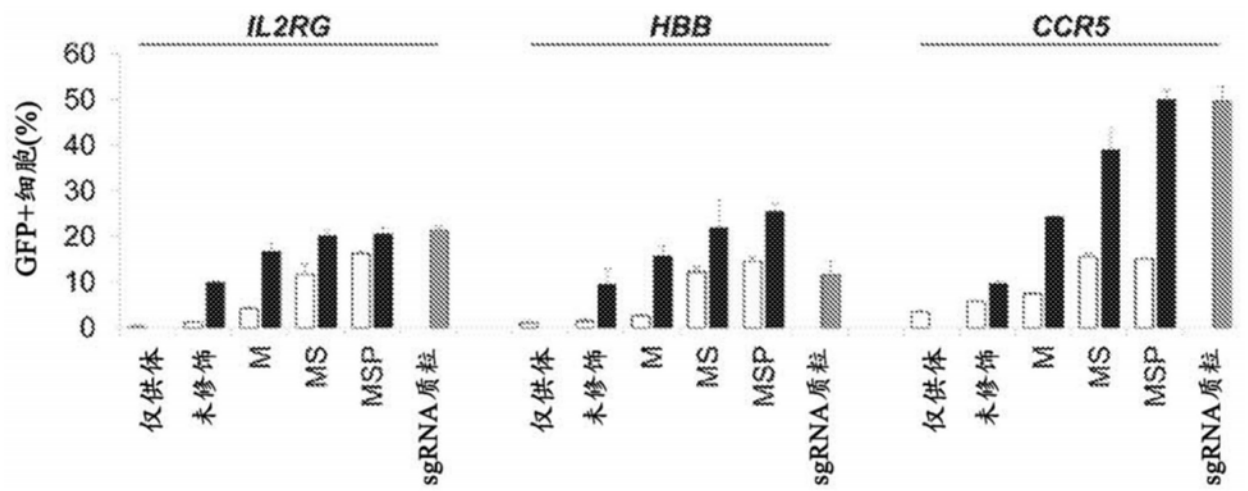


图1D

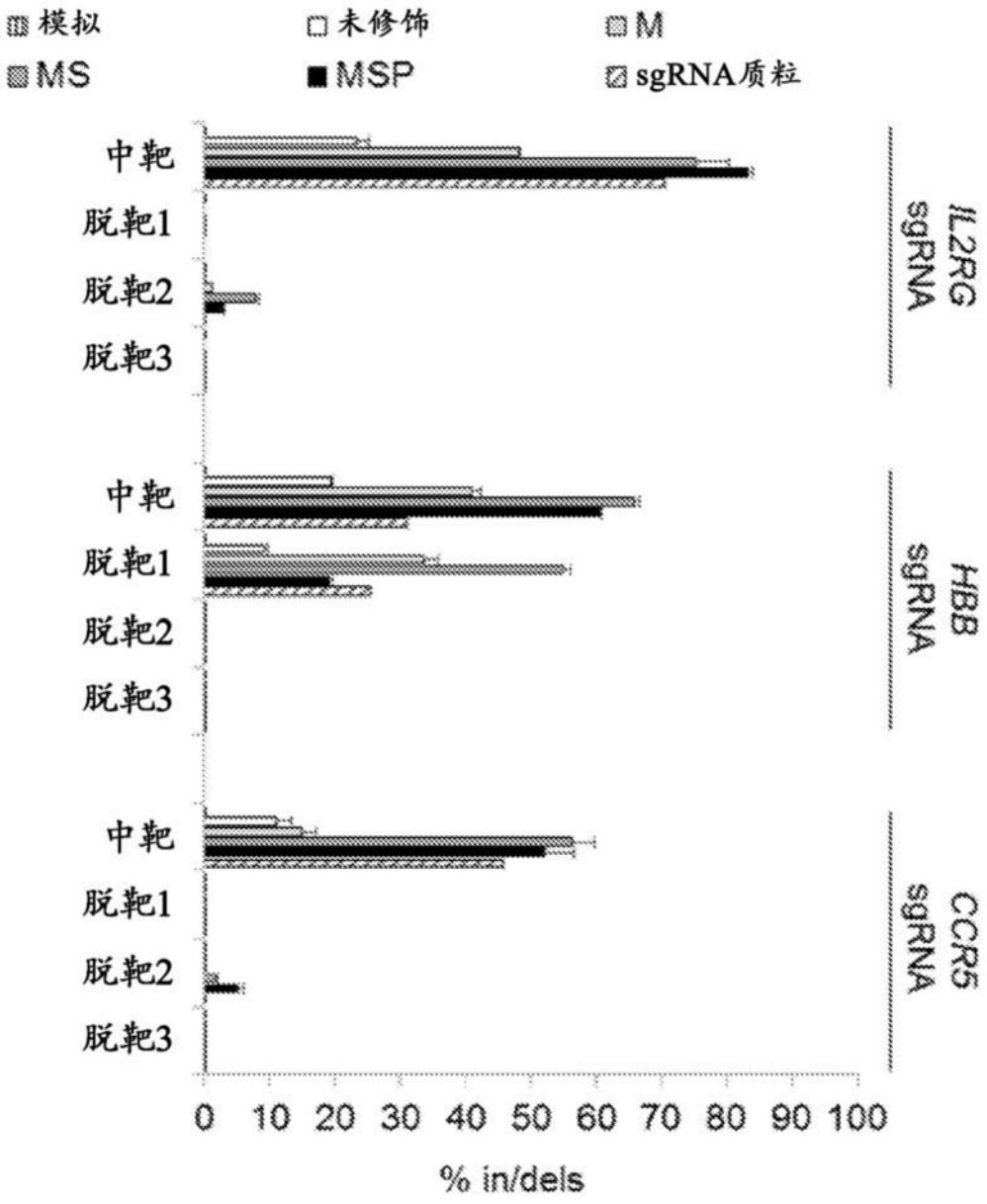

模拟
未修饰
M
MS
MSP
sgRNA质粒
中靶
脱靶1
脱靶2
脱靶3
IL2RG sgRNA
HBB sgRNA
CCR5 sgRNA
0 10 20 30 40 50 60 70 80 90 100
% in/dels

图1E

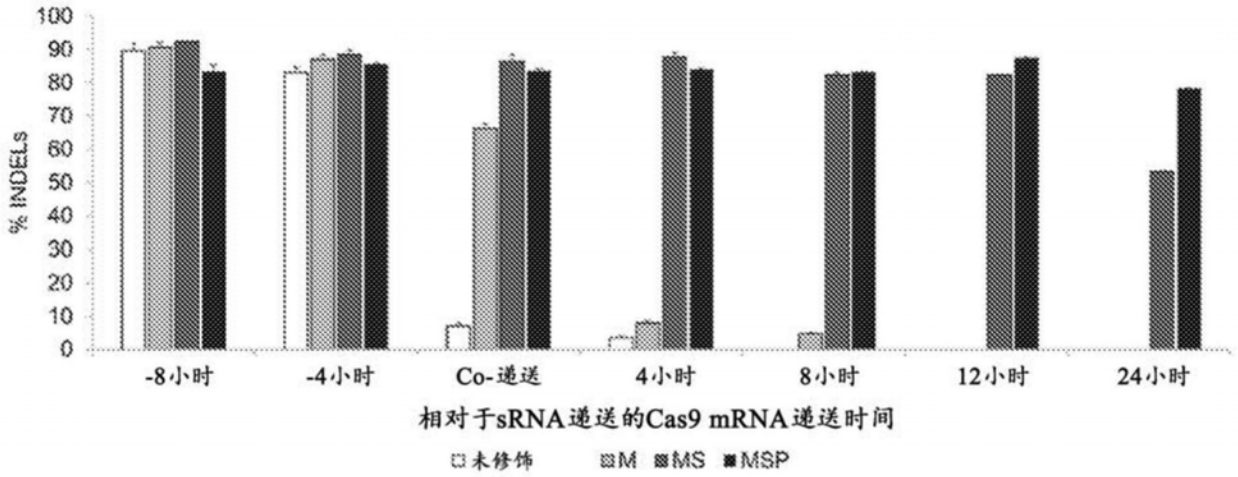

% INDELs
100
90
80
70
60
50
40
30
20
10
0
-8小时
-4小时
Co-递送
4小时
8小时
12小时
24小时
相对于sRNA递送的Cas9 mRNA递送时间
未修饰
M
MS
MSP

图1F

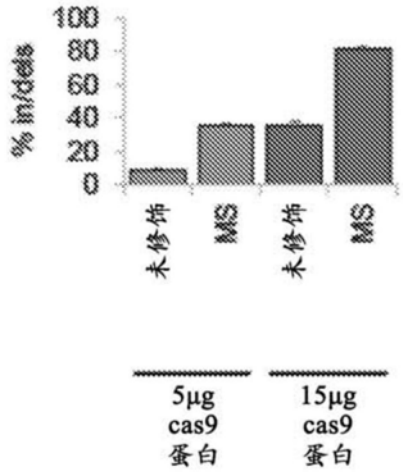

% in/dels
100
80
60
40
20
0
未修饰
MS
未修饰
MS
5μg cas9 蛋白
15μg cas9 蛋白

图1G

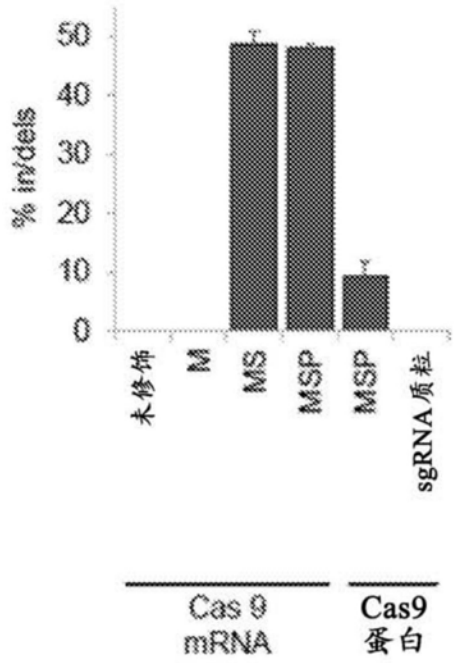

% in/dels
50
40
30
20
10
0
未修饰
M
MS
MSP
MSP
sgRNA/质粒
Cas 9 mRNA
Cas9 蛋白

图2A

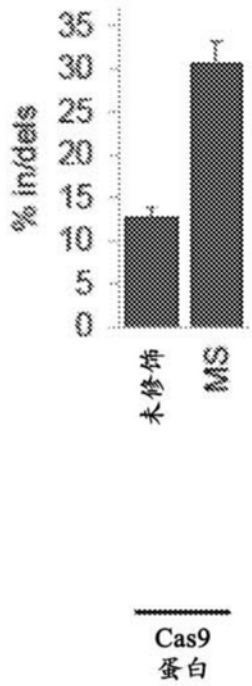

% in/dels
35
30
25
20
15
10
5
0
未修饰
MS
Cas9 蛋白

图2B

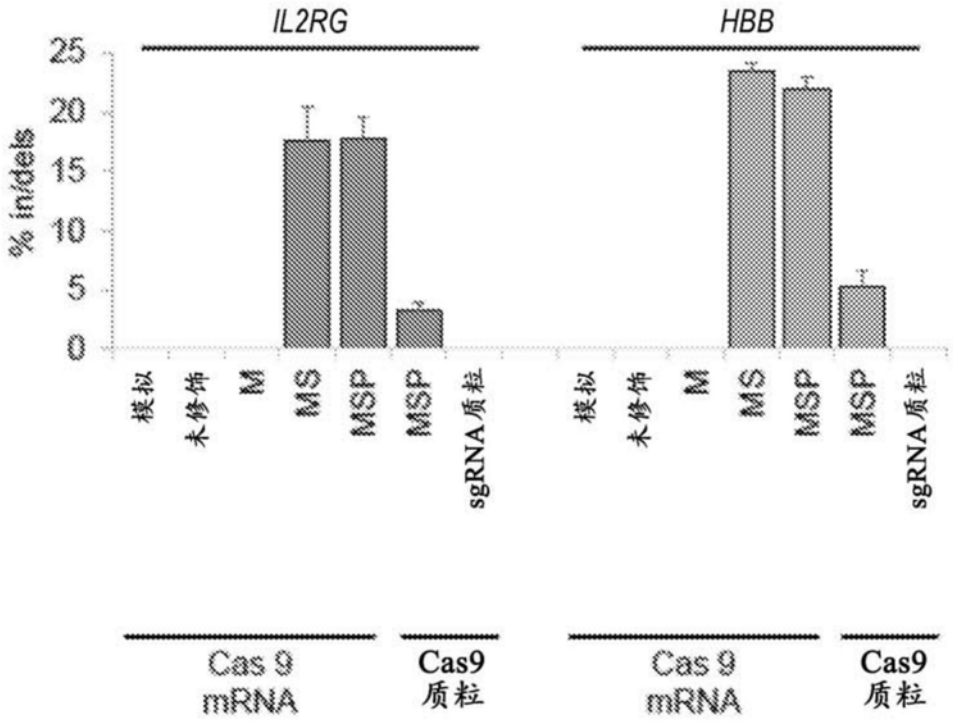

IL2RG
HBB
% indels
25
20
15
10
5
0
模拟
未修饰
M
MS
MSP
MSP
sgRNA质粒
Cas 9 mRNA
Cas9 质粒

图2C

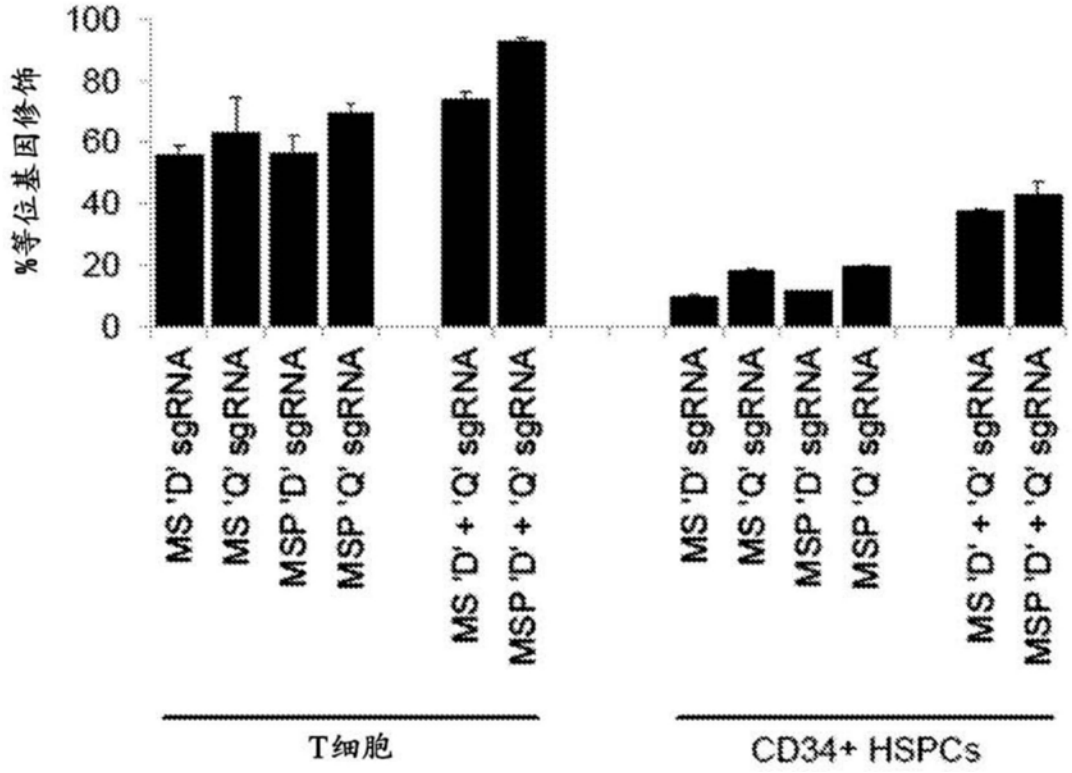

%等位基因修饰
100
80
60
40
20
0
MS 'D' sgRNA
MS 'Q' sgRNA
MSP 'D' sgRNA
MSP 'Q' sgRNA
MS 'D' + 'Q' sgRNA
MSP 'D' + 'Q' sgRNA
T细胞
CD34+ HSPCs

图2D

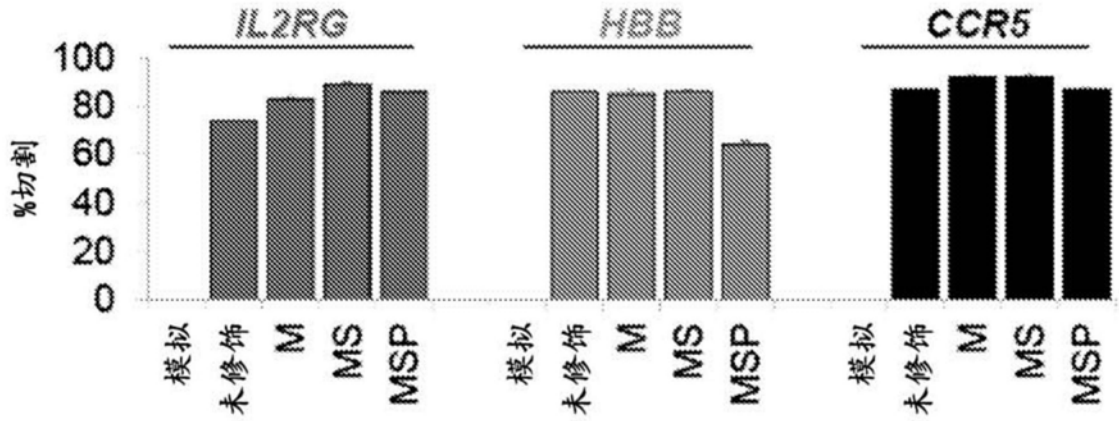


图3

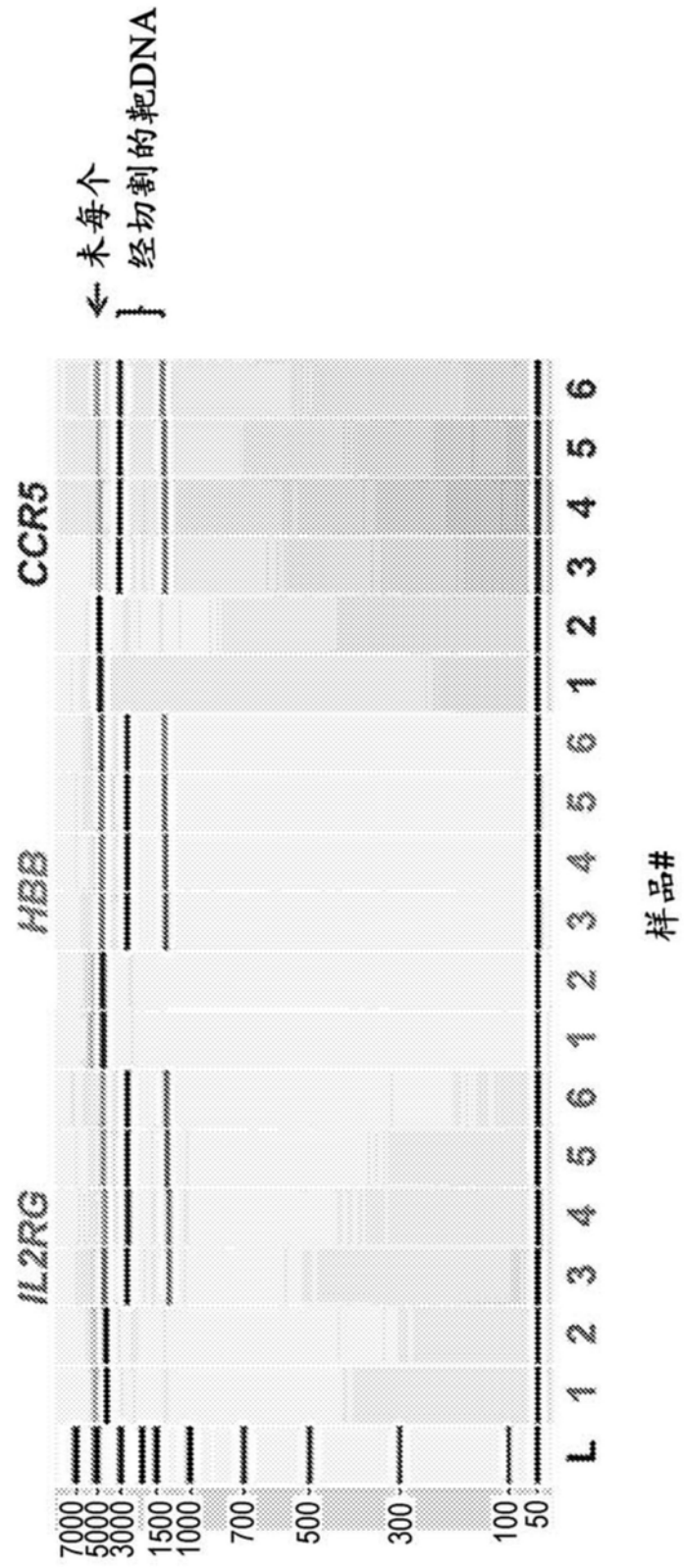

未每个
经切割的靶DNA
IL2RG
HBB
CCR5
L 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6
样品#
7000
5000
3000
1500
1000
700
500
300
100
50

图4

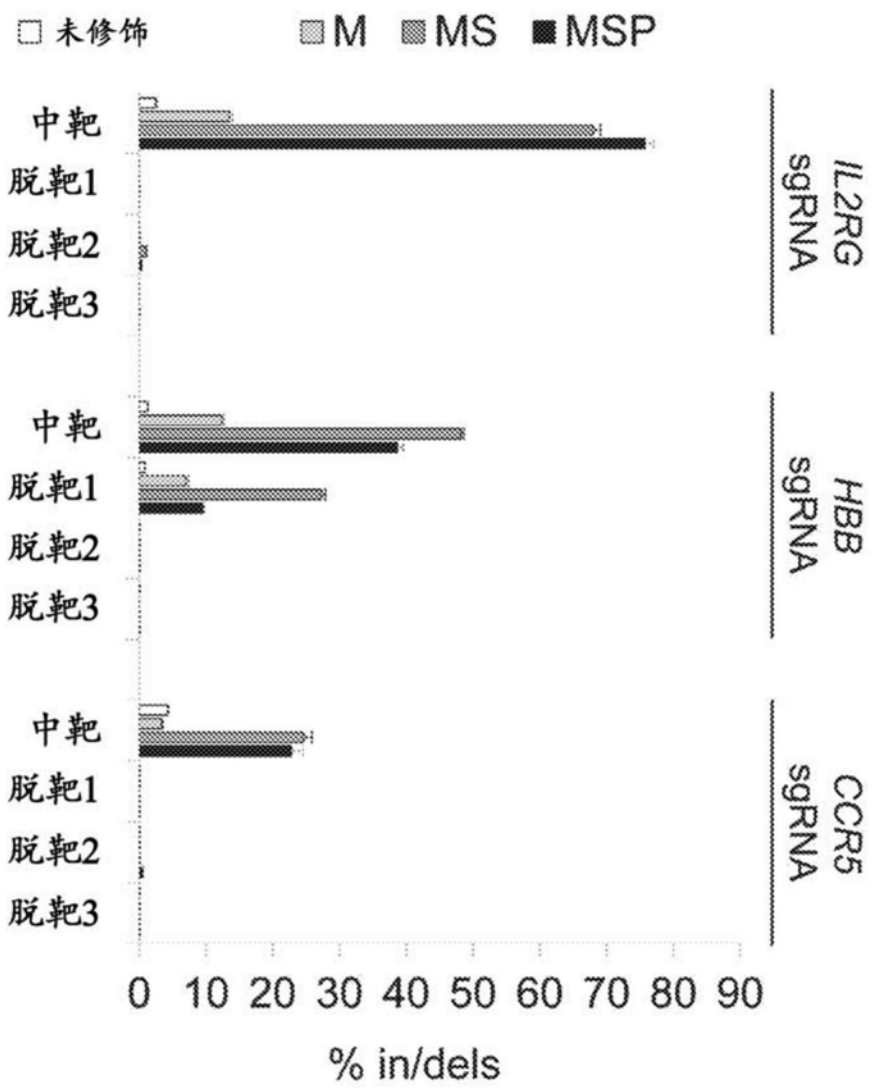


图5

sgRNA (μg) \ Cas9 (μg)	1	5	10	15
1	70	86	86	85
2.5	81	86	82	89
5	84	85	87	88
10	83	88	89	90
20	86	89	89	90

图6

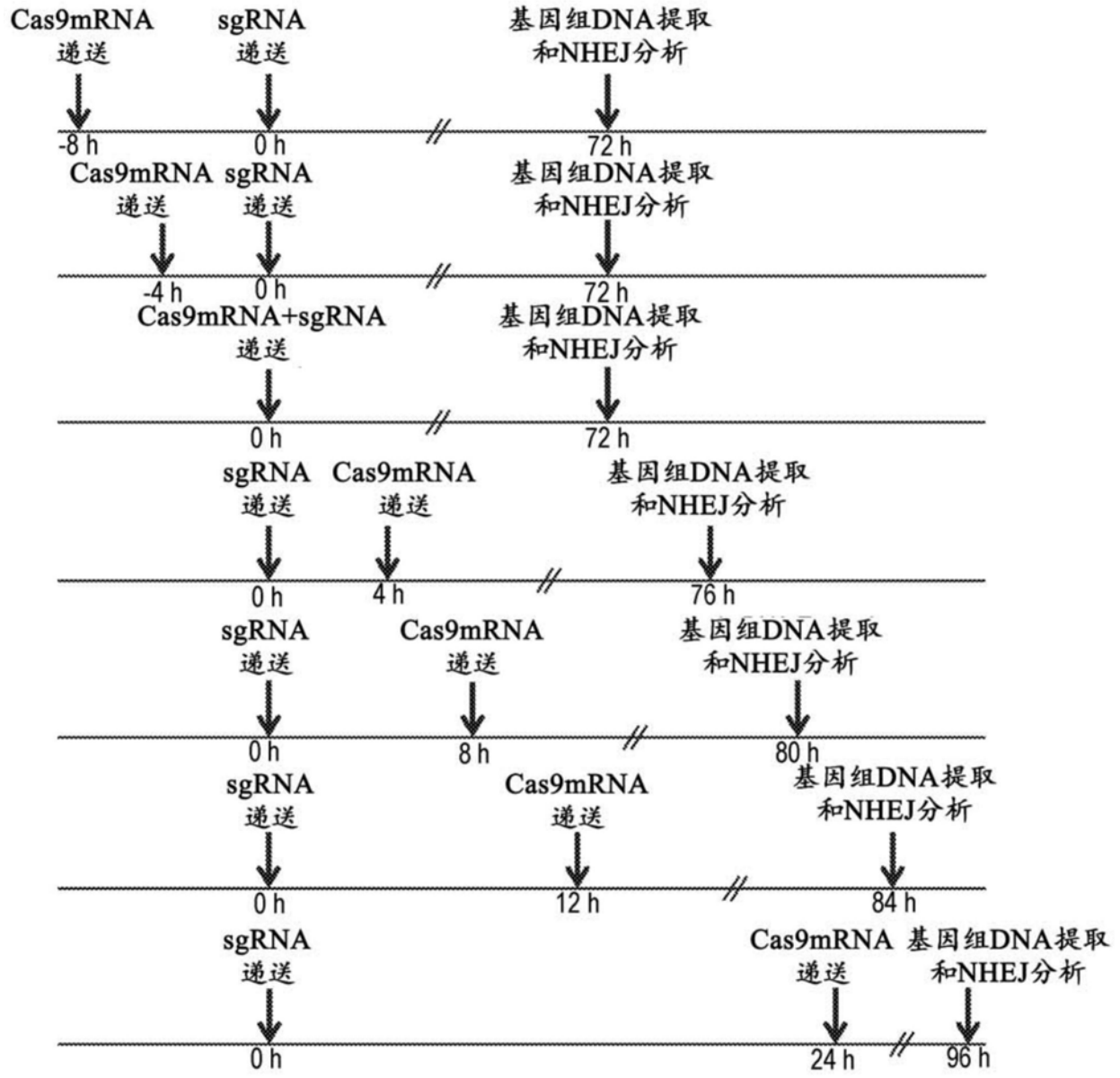

Cas9mRNA
递送
sgRNA
递送
基因组DNA提取
和NHEJ分析
-8 h
0 h
72 h
Cas9mRNA
递送
sgRNA
递送
基因组DNA提取
和NHEJ分析
-4 h
0 h
72 h
Cas9mRNA+sgRNA
递送
基因组DNA提取
和NHEJ分析
0 h
72 h
sgRNA
递送
Cas9mRNA
递送
基因组DNA提取
和NHEJ分析
0 h
4 h
76 h
sgRNA
递送
Cas9mRNA
递送
基因组DNA提取
和NHEJ分析
0 h
8 h
80 h
sgRNA
递送
Cas9mRNA
递送
基因组DNA提取
和NHEJ分析
0 h
12 h
84 h
sgRNA
递送
Cas9mRNA
递送
基因组DNA提取
和NHEJ分析
0 h
24 h
96 h

图7

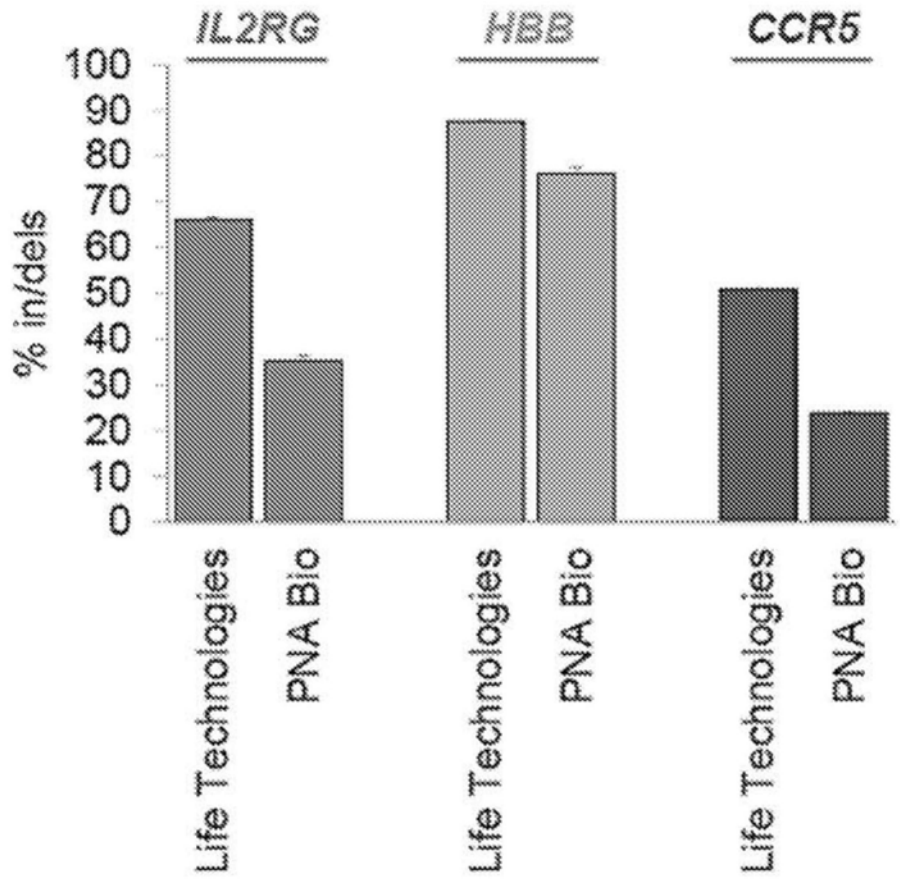

IL2RG
HBB
CCR5
% in/dels
100
90
80
70
60
50
40
30
20
10
0
Life Technologies
PNA Bio
Life Technologies
PNA Bio
Life Technologies
PNA Bio

图8

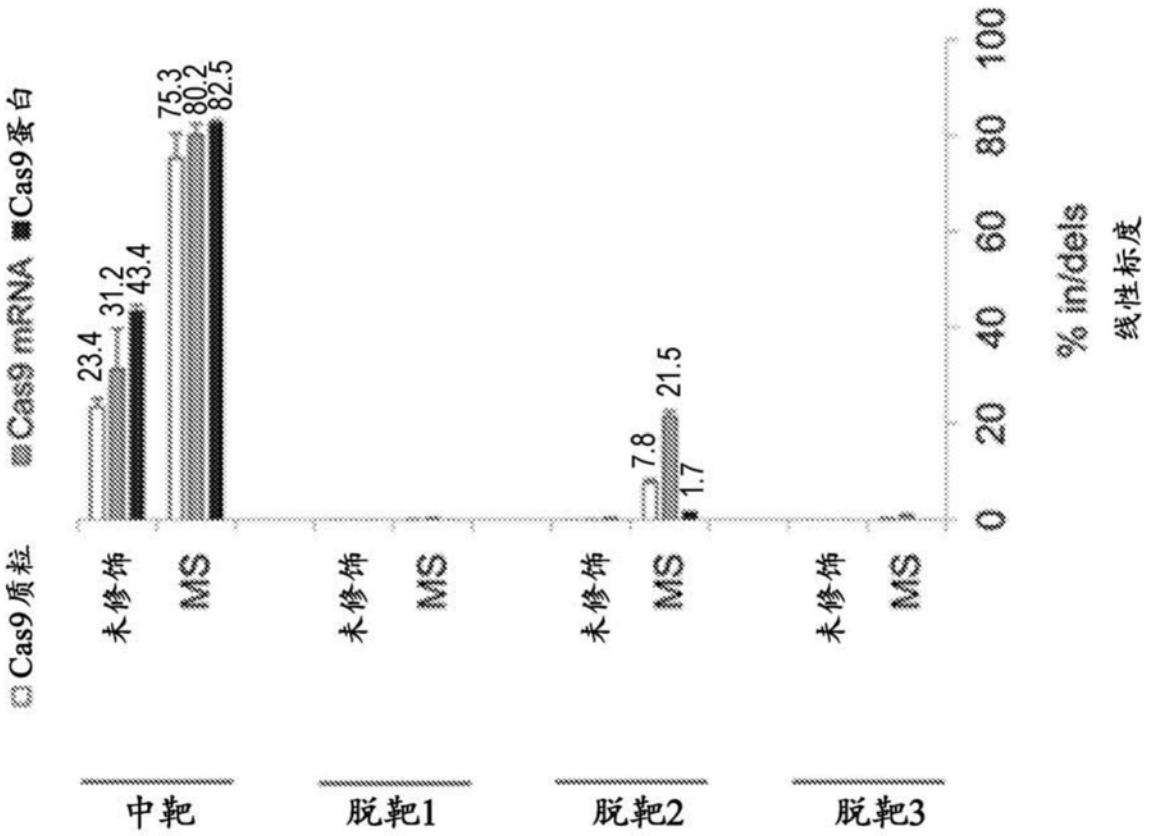

Cas9质粒
Cas9 mRNA
Cas9蛋白
% in/dels
线性标度
0
20
40
60
80
100
23.4
31.2
43.4
75.3
80.2
82.5
7.8
21.5
1.7
未修饰
MS
中靶
脱靶1
脱靶2
脱靶3

图9A

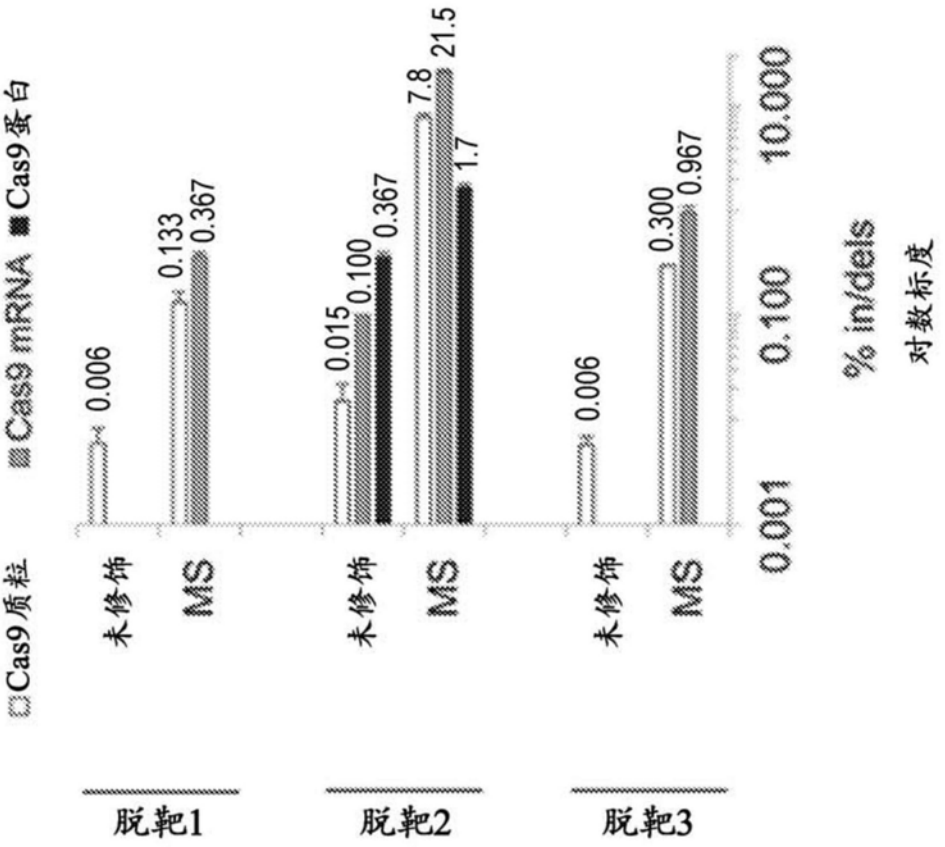

Cas9质粒
Cas9 mRNA
Cas9蛋白
0.006
0.133
0.367
0.015
0.100
0.367
7.8
21.5
1.7
0.006
0.300
0.967
未修饰
MS
未修饰
MS
未修饰
MS
脱靶1
脱靶2
脱靶3
10,000
0.100
0.001
% in/dels
对数标度

图9B

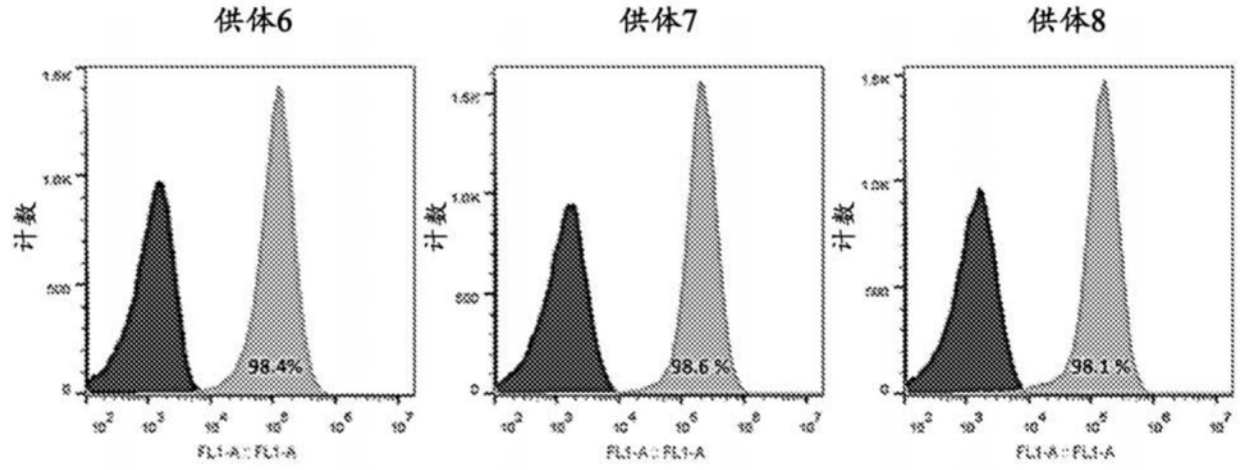

供体6
供体7
供体8
计数
98.4%
98.6 %
98.1 %

图10

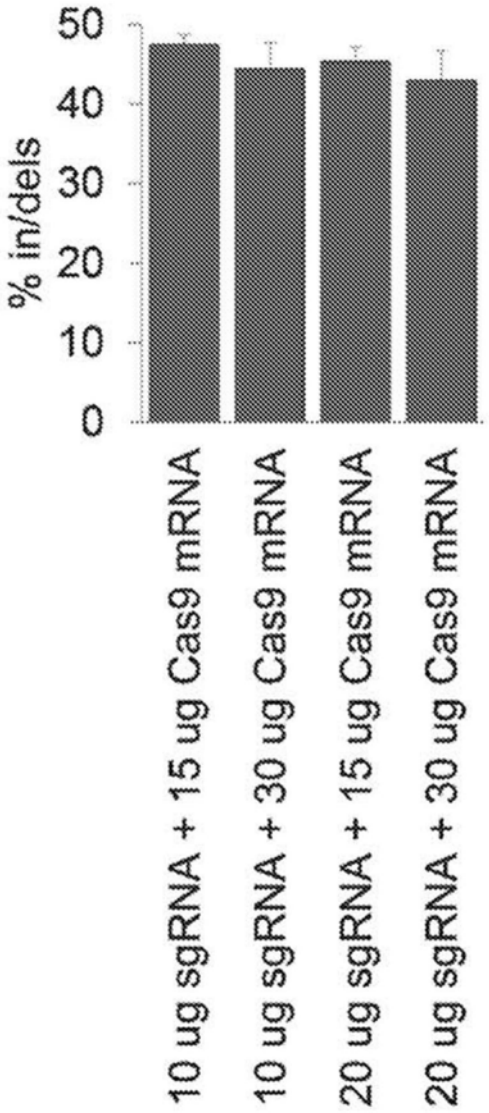

50
40
30
20
10
0
% in/dels
10 ug sgRNA + 15 ug Cas9 mRNA
10 ug sgRNA + 30 ug Cas9 mRNA
20 ug sgRNA + 15 ug Cas9 mRNA
20 ug sgRNA + 30 ug Cas9 mRNA

图11

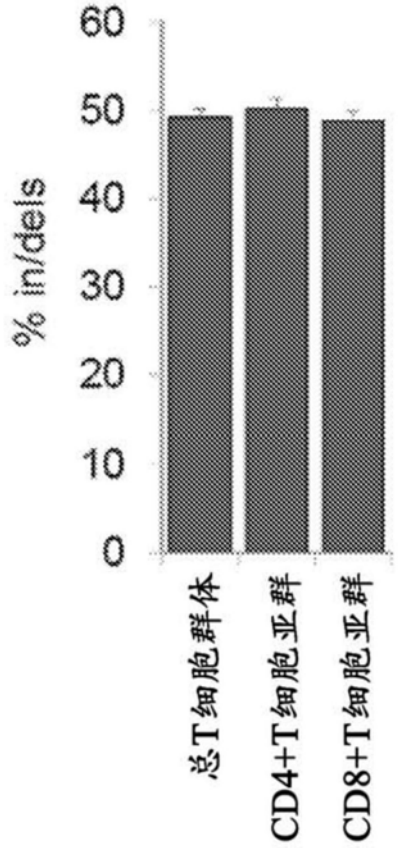


图12

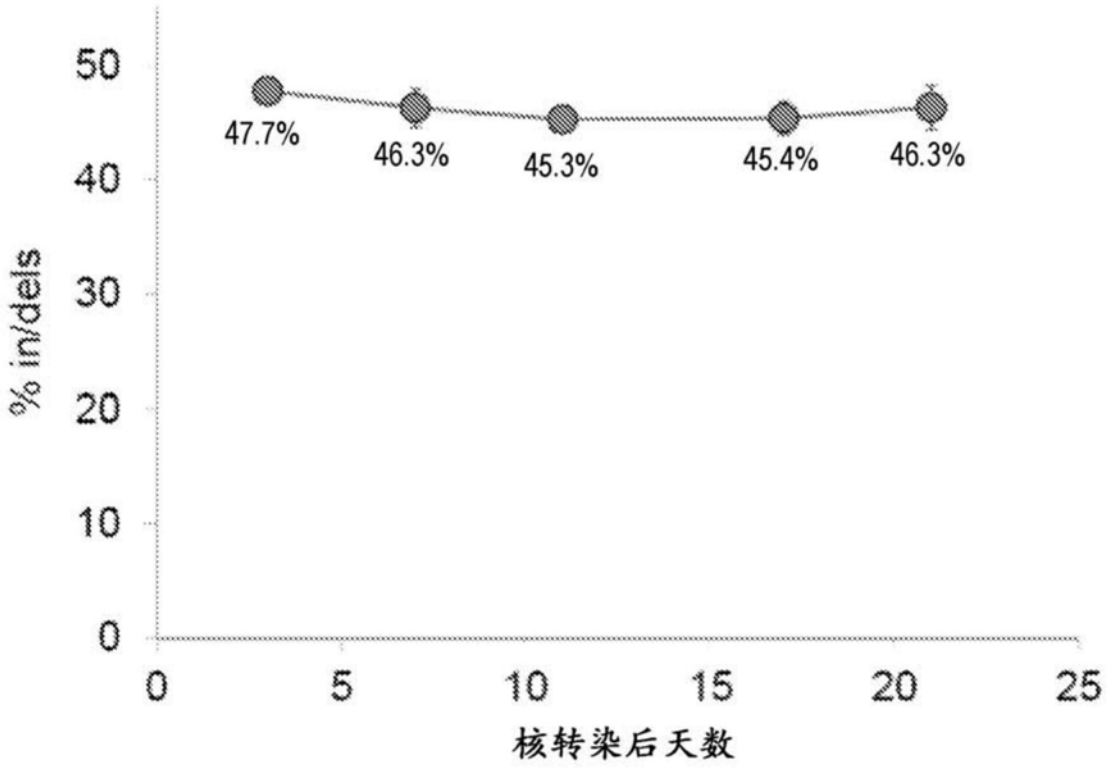


图13

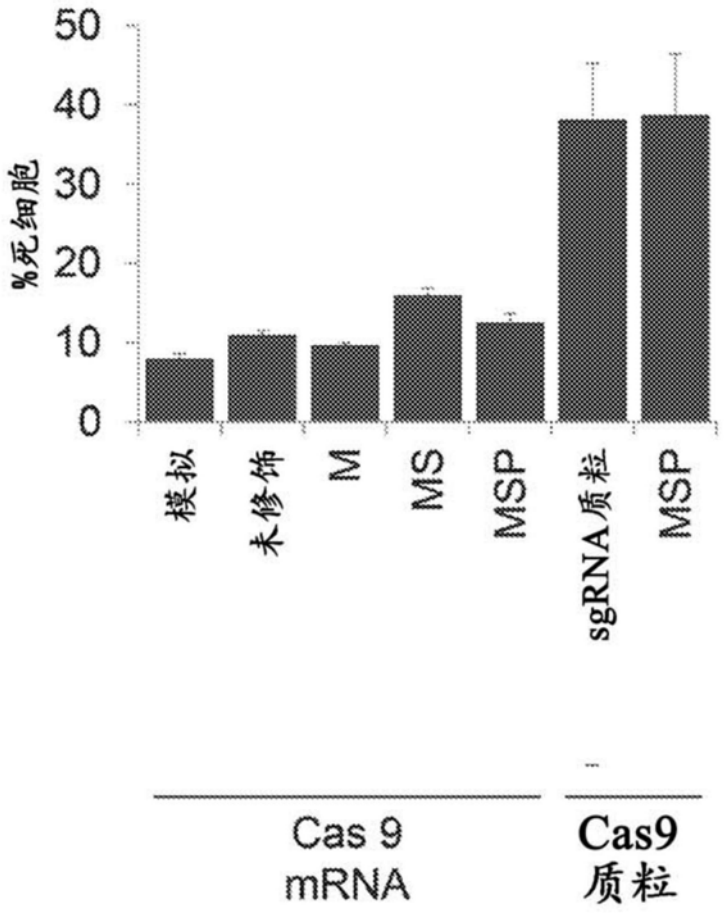


图14

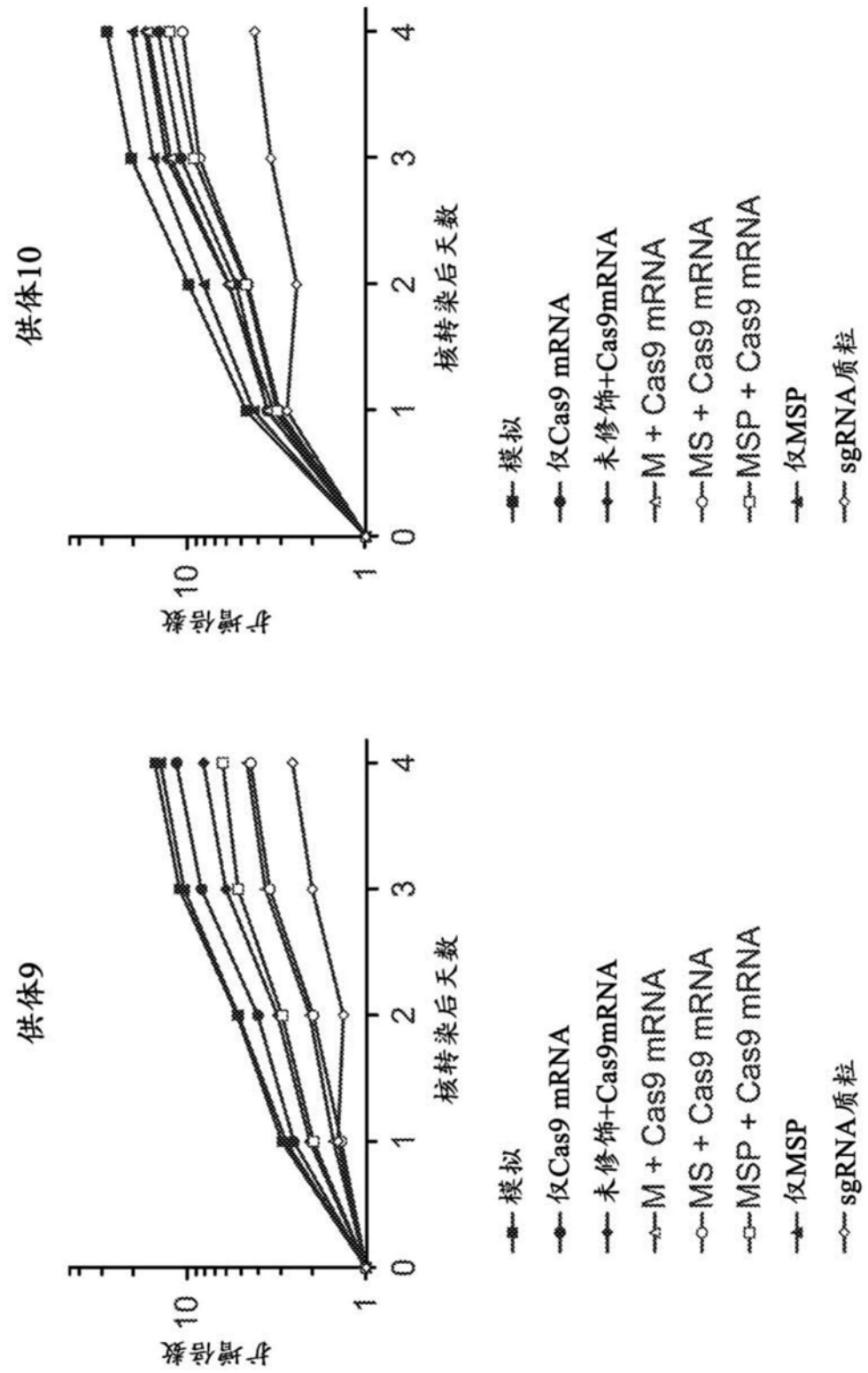

供体10
扩增倍数
核转染后天数
模拟
仅Cas9 mRNA
未修饰+Cas9mRNA
M + Cas9 mRNA
MS + Cas9 mRNA
MSP + Cas9 mRNA
仅MSP
sgRNA质粒
供体9
扩增倍数
核转染后天数
模拟
仅Cas9 mRNA
未修饰+Cas9mRNA
M + Cas9 mRNA
MS + Cas9 mRNA
MSP + Cas9 mRNA
仅MSP
sgRNA质粒

图15

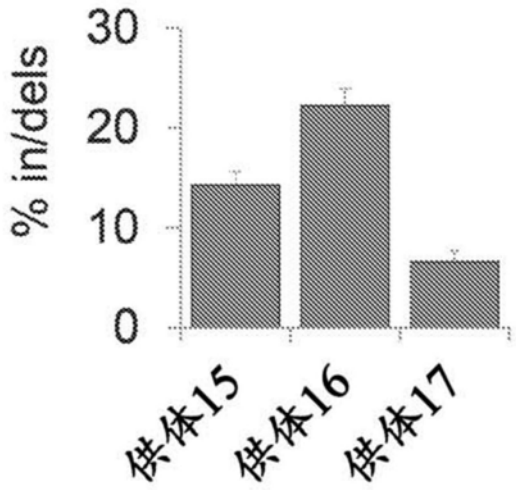

% in/dels
30
20
10
0
供体15
供体16
供体17

图16

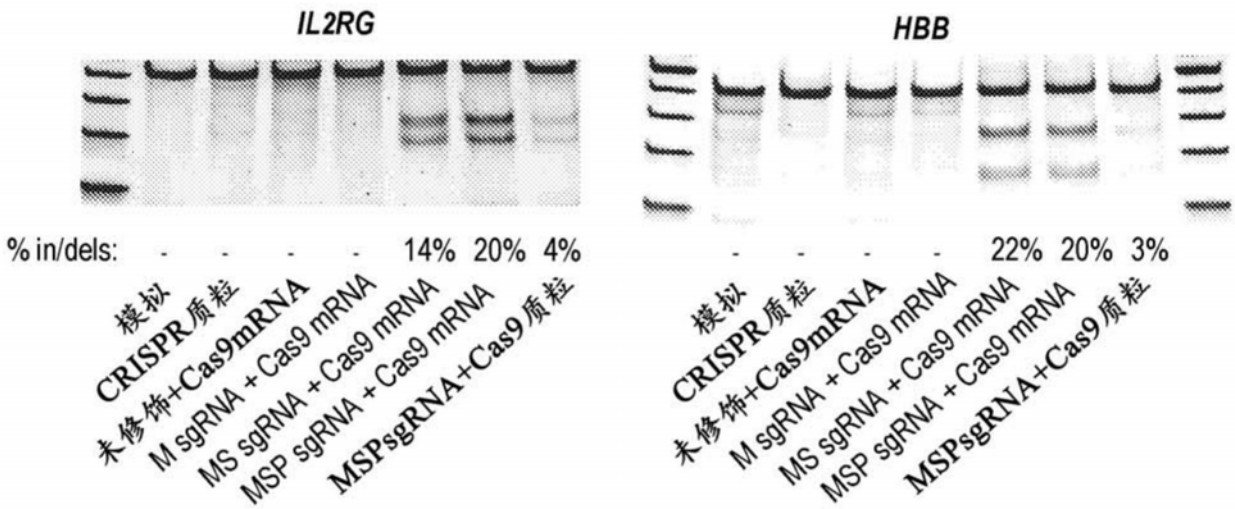

IL2RG
HBB
% in/dels: - - - - 14% 20% 4%
- - - - 22% 20% 3%
模拟
CRISPR质粒
未修饰+Cas9mRNA
M sgRNA + Cas9 mRNA
MS sgRNA + Cas9 mRNA
MSP sgRNA + Cas9 mRNA
MSPsgRNA+Cas9质粒

图17

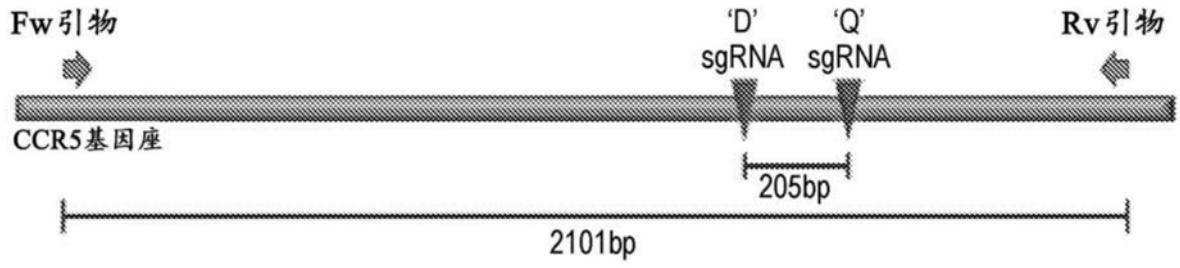

Fw引物
'D' sgRNA
'Q' sgRNA
Rv引物
CCR5基因座
205bp
2101bp

图18A

	T细胞		CD34+ HSPCs	
分析的克隆PCR片段总数	181	186	222	216
%等位基因修饰±SEM	73.9 (±2.8)	93.1 (±1.1)	37.8 (±0.5)	43.0 (±4.3)
%野生型等位基因±SEM	26.1 (±2.8)	6.9 (±1.1)	62.2 (±0.5)	57.0 (±4.3)
%在剪切位点之间具有缺失的等位基因±SEM	53.4 (±4.2)	69.7 (±2.1)	21.1 (±1.8)	23.1 (±3.1)
%在剪切位点之间具有倒序序列的等位基因±SEM	5.3 (±2.2)	9.6 (±2.0)	5.9 (±2.6)	6.9 (±2.1)
%在两个位点有in/del±SEM	12.7 (±2.4)	12.6 (±1.6)	3.6 (±1.2)	7.9 (±1.9)
%仅在右位点有in/del±SEM	2.4 (±1.7)	1.1 (±0.6)	3.6 (±0.9)	2.3 (±1.2)
%仅在左位点有in/del ±SEM	0	0	3.6 (±0.5)	2.8 (±0.8)

图18B

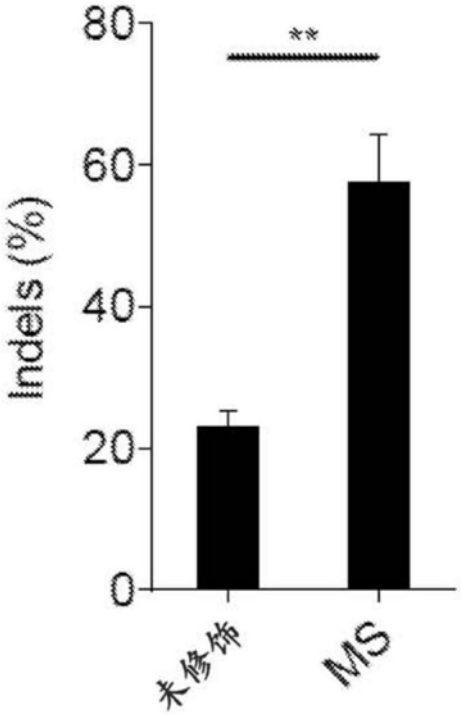


图19

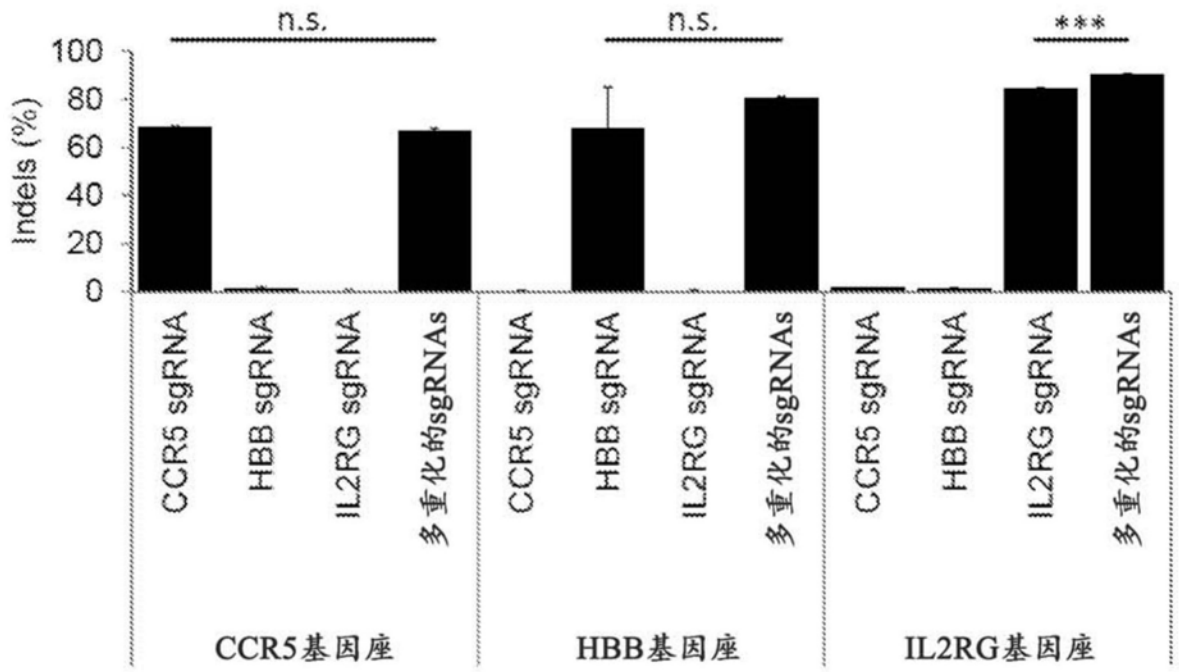

n.s.
n.s.
***
100
80
60
40
20
0
Indels (%)
CCR5 sgRNA
HBB sgRNA
IL2RG sgRNA
多重化的sgRNAs
CCR5 sgRNA
HBB sgRNA
IL2RG sgRNA
多重化的sgRNAs
CCR5 sgRNA
HBB sgRNA
IL2RG sgRNA
多重化的sgRNAs
CCR5基因座
HBB基因座
IL2RG基因座

图20

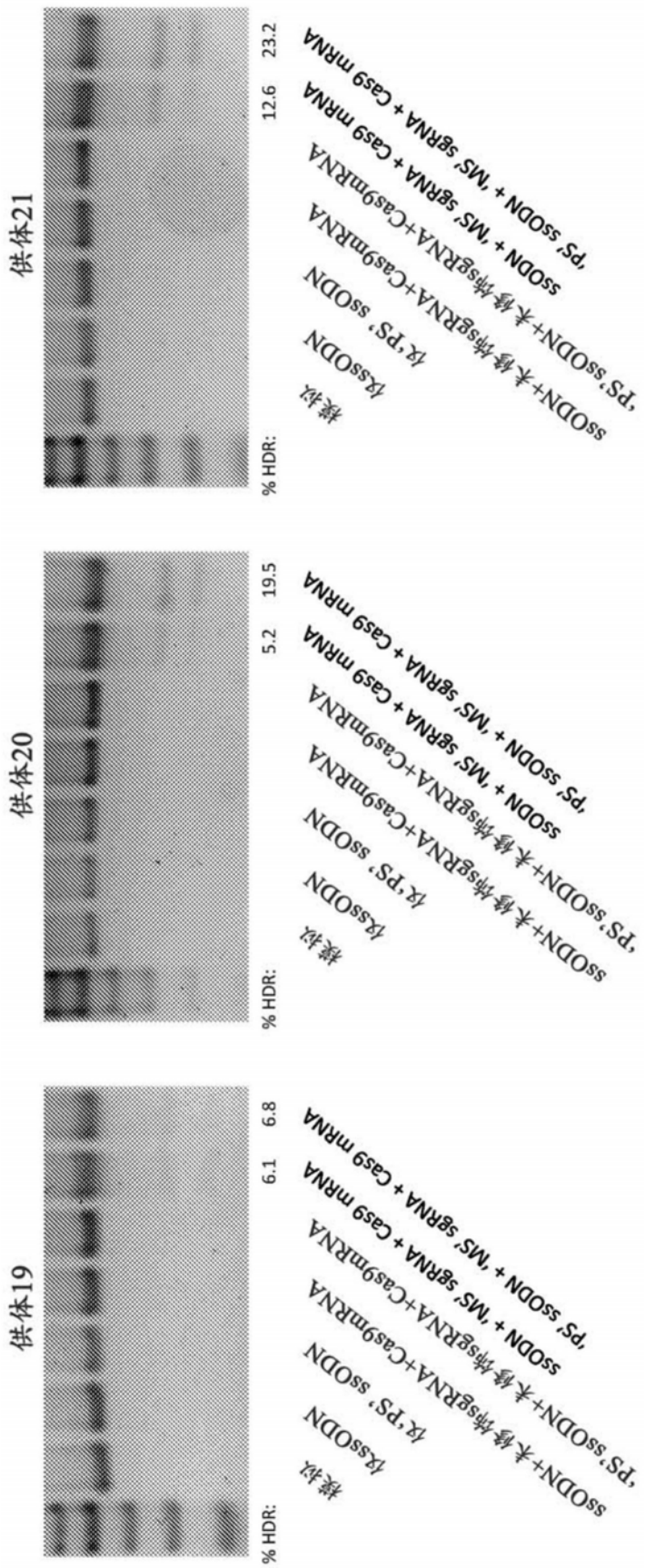

供体21
% HDR:
12.6
23.2
模拟
仅ssODN
仅ssODN，Sd，
ssODN+未修饰sgRNA+Cas9mRNA
ssODN+未修饰sgRNA+Cas9mRNA
ssODN + MS，sgRNA + Cas9 mRNA
ssODN + MS，sgRNA + Cas9 mRNA
ssODN，Sd，+ MS，sgRNA + Cas9 mRNA
ssODN，Sd，+ MS，sgRNA + Cas9 mRNA
供体20
% HDR:
5.2
19.5
模拟
仅ssODN
仅ssODN，Sd，
ssODN+未修饰sgRNA+Cas9mRNA
ssODN+未修饰sgRNA+Cas9mRNA
ssODN + MS，sgRNA + Cas9 mRNA
ssODN + MS，sgRNA + Cas9 mRNA
ssODN，Sd，+ MS，sgRNA + Cas9 mRNA
ssODN，Sd，+ MS，sgRNA + Cas9 mRNA
供体19
% HDR:
6.1
6.8
模拟
仅ssODN
仅ssODN，Sd，
ssODN+未修饰sgRNA+Cas9mRNA
ssODN+未修饰sgRNA+Cas9mRNA
ssODN + MS，sgRNA + Cas9 mRNA
ssODN + MS，sgRNA + Cas9 mRNA
ssODN，Sd，+ MS，sgRNA + Cas9 mRNA
ssODN，Sd，+ MS，sgRNA + Cas9 mRNA

图21