(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 112020558 A
(43)申请公布日 2020.12.01

(21)申请号 201980025285.5

(22)申请日 2019.03.14

(30)优先权数据
62/643,168 2018.03.14 US
62/767,488 2018.11.14 US
62/773,073 2018.11.29 US

(85)PCT国际申请进入国家阶段日
2020.10.12

(86)PCT国际申请的申请数据
PCT/US2019/022374 2019.03.14

(87)PCT国际申请的公布数据
WO2019/178426 EN 2019.09.19

(71)申请人 爱迪塔斯医药公司
地址 美国马萨诸塞州

(72)发明人 J·L·戈里
E·奥佩潘·德·拉莫特-德赫兹
J·希斯 J·A·左瑞斯 K·常

(74)专利代理机构 北京市君合律师事务所 11517
代理人 张怡 杜丹

(51)Int.Cl.
C12N 15/113(2006.01)
A61K 31/712(2006.01)
A61K 31/7125(2006.01)
C12N 5/10(2006.01)

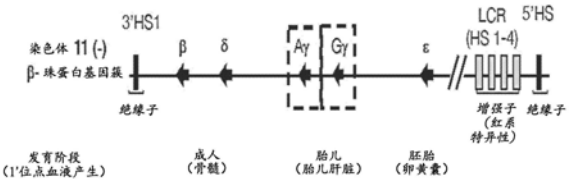
权利要求书1页 说明书122页
序列表（电子公布） 附图104页

(54)发明名称

治疗血红蛋白病的系统和方法

(57)摘要

提供了基因组编辑系统、指导RNA和CRISPR介导的方法，用于改变细胞中HBG1和HBG2基因座的部分、BCL11A基因的红系特异性增强子的部分、或其组合，并增加胎儿血红蛋白的表达。



CN 112020558 A

1.一种RNP复合物,其包含来自普雷沃菌属(Prevotella)和弗朗西斯菌属(Franciscella)的CRISPR 1(Cpf1)RNA指导的核酸酶或其变体和gRNA,其中所述gRNA能够结合细胞中的HBG基因启动子中的靶位点。

2.如权利要求1所述的RNP复合物,其中所述gRNA经修饰或未经修饰。

3.一种改变细胞中HBG基因启动子的方法,所述方法包括使所述细胞与如权利要求1所述的RNP复合物接触。

4.如权利要求3所述的方法,其中所述改变包含所述HBG基因启动子的CCAAT盒靶区域内的indel。

5.如权利要求1所述的RNP复合物,其中所述靶位点包含位于Chr 11(NC_000011.10)5,249,904-5,249,927(表17,区域6)和Chr 11(NC_000011.10)5,254,879-5,254,909(表17,区域16)之间的核苷酸;或其组合。

6.一种分离的细胞,所述分离细胞包含HBG基因启动子中的改变,所述改变通过将RNP复合物递送至如权利要求1所述的细胞而产生。

7.一种增加人细胞中胎儿血红蛋白(HbF)水平的离体方法,所述方法是通过使用包含gRNA和Cpf1 RNA指导的核酸酶或其变体的RNP复合物进行基因组编辑,以实现HBG基因启动子中的改变,从而增加HbF的表达。

8.一种CD34+或造血干细胞群体,其中所述群体中的一个或多个细胞包含HBG基因启动子中的改变,所述改变是通过将RNP复合物递送至所述CD34+或造血干细胞群体产生的,所述RNP复合物包含gRNA和Cpf1 RNA指导的核酸酶或其变体。

9.一种减轻有需要的受试者中镰状细胞病的一种或多种症状的方法,所述方法包括:

a)从所述受试者中分离CD34+或造血干细胞群体;

b)通过以下来修饰离体分离的细胞群体:将包含gRNA和Cpf1 RNA指导的核酸酶或其变体的RNP复合物递送至分离的细胞群体,从而实现所述群体中的一个或多个细胞中HBG基因启动子中的改变;和

c)对所述受试者施用经修饰的细胞群体,

从而减轻所述受试者的镰状细胞病的一种或多种症状。

10.如权利要求9所述的方法,其中所述改变包含所述HBG基因启动子的CCAAT盒靶区域内的indel。

11.如权利要求9所述的方法,其中所述RNP复合物使用电穿孔递送。

12.如权利要求9所述的经修饰的细胞群体,其中所述细胞群体中至少约5%、至少约10%、至少约20%、至少约30%、至少约40%、至少约50%、至少约60%、至少约70%、至少约80%或至少约90%的细胞包含有效indel。

13.一种gRNA,其包含5'末端和3'末端,并且在所述5'末端包含DNA延伸并且在所述3'末端包含2'-O-甲基-3'-硫代磷酸酯修饰,其中所述gRNA包括能够与靶位点杂交的RNA区段和能够与Cpf1 RNA指导的核酸酶缔合的RNA区段。

14.如权利要求13所述的gRNA,其中所述DNA延伸包含SEQ ID NO:1235-1250所示的序列。

15.一种RNP复合物,其包含Cpf1 RNA指导的核酸酶和如权利要求14所述的gRNA。

治疗血红蛋白病的系统和方法

相关申请的引用

[0001] 本申请要求于2018年3月14日提交的美国临时申请号62/643,168、2018年11月14日提交的美国临时申请号62/767,488和2018年11月29日提交的美国临时申请号62/773,073的权益,其内容特此于此通过引用以其整体并入。

序列表

[0002] 本申请含有以ASCII格式经由EFS-Web提交的序列表,并且将其通过引用以其整体并入本文。于2019年3月14日创建的ASCII副本名为2019-03-14_EM115PCT_Editas_8013WO02_Sequence_Listing.txt,并且大小为811KB。

技术领域

[0003] 本公开涉及用于基因组编辑系统和改变靶核酸序列或调节靶核酸序列的表达的方法,及其与编码血红蛋白亚基的基因的改变和/或血红蛋白病的治疗有关的应用。

背景技术

[0004] 血红蛋白(Hb)将红细胞或红血细胞(RBC)中的氧从肺部运送到组织。在胚胎发育期间和出生后不久,血红蛋白以胎儿血红蛋白(HbF)的形式存在,胎儿血红蛋白是由两个α-珠蛋白链和两个γ-珠蛋白链组成的四聚体蛋白质。HbF在很大程度上被成人血红蛋白(HbA)替换,成人血红蛋白是一种四聚体蛋白质,其中HbF的γ-珠蛋白链通过称为珠蛋白转换的过程被β-珠蛋白链替换。普通成人的总血红蛋白小于1%的HbF(Thein 2009)。所述α-血红蛋白基因位于第16号染色体上,而β-血红蛋白基因(HBB),Aγ(Aγ)珠蛋白链(HBG1,也称为γ珠蛋白A)和Gγ(Gγ)珠蛋白链(HBG2,也称为γ珠蛋白G)位于第11号染色体上的珠蛋白基因簇(也称为珠蛋白基因座)内。

[0005] HBB中的突变可引起血红蛋白异常(即血红蛋白病),所述血红蛋白病包括镰状细胞病(SCD)和β地中海贫血(β-Thal)。美国大约93,000人被诊断患有血红蛋白病。全球每年有300,000儿童出生患有血红蛋白病(Angastiniotis 1998)。因为这些病症与HBB突变有关,所以他们的症状典型地直到珠蛋白从HbF转换为HbA后才表现。

[0006] SCD是美国最常见的遗传性血液病,影响大约80,000人(Brousseau2010)。SCD在非洲血统的人中最常见,SCD的流行率为500人中有1人。在非洲,SCD的流行率为1500万(Aliyu,2008)。SCD在印度人,沙特阿拉伯人和地中海人后裔中也更常见。在西班牙裔美国人后裔中,镰状细胞病的流行率为1000人中有1人(Lewis 2014)。

[0007] SCD由HBB基因中的单个纯合突变引起,c.17A>T(HbS突变)。镰状突变是HBB上的点突变(GAG>GTG),其导致缬氨酸取代外显子1中氨基酸位置6的谷氨酸。β-血红蛋白链位置6的缬氨酸是疏水性的,并且当β-珠蛋白不与氧气结合时,引起β-珠蛋白构象的变化。这种构象变化引起HbS蛋白在没有氧的情况下聚合,导致RBC变形(即,镰状)。SCD以常染色体隐性方式遗传,因此只有具有两个HbS等位基因的患者才患有所述疾病。杂合的受试者具有镰状细胞性状,如果他们严重脱水或缺氧,可能患有贫血和/或痛性危象。

[0008] 镰状形状的RBC引起多种症状，包括贫血、镰状细胞危象、血管阻塞性危象，再生障碍性危象和急性胸部综合征。镰状形状的RBC比野生型RBC弹性小，因此不能容易地通过毛细血管床并导致阻塞和缺血(即，血管阻塞)。当镰状细胞阻塞器官毛细血管床中的血流导致疼痛、缺血和坏死时，就会发生血管阻塞性危象。这些发作典型地持续5天至7天。脾脏在清除功能失调的RBC中起作用，因此典型地在儿童早期期间扩大并且频繁发生血管阻塞性危象。到儿童期结束时，SCD患者的脾经常梗塞，导致自体脾切除。溶血是SCD的一个恒定特征并引起贫血。镰状细胞在循环中存活10天至20天，而健康的RBC存活90天至120天。必要时输血SCD受试者以维持足够的血红蛋白水平。频繁的输血使受试者有感染HIV、乙型肝炎、和丙型肝炎的风险。受试者还可能患有急性胸部危象和四肢、终末器官、和中枢神经系统的梗塞。

[0009] 患有SCD的受试者的生命期望降低。通过对危象和贫血进行认真的，终身的管理，SCD患者的预后正在稳定地改进。在2001年，患有镰状细胞病的受试者的平均生命期望是50岁中后期。目前对SCD的治疗涉及危象期间的水合和疼痛管理，以及根据需要进行输血以校正贫血。

[0010] 地中海贫血(例如，β-Thal、δ-Thal、和β/δ-Thal)引起慢性贫血。估计β-Thal影响全球大约100,000人中有1人。它在某些群体中的流行率较高，包括欧洲后裔的群体，其流行率大约为10,000人中有1人。除非通过终身输血和螯合疗法治疗，否则重型β-Thal是疾病的更严重形式，是危及生命的。在美国，大约有3,000名患有重型β-Thal的受试者。中间型β-Thal不需要输血，但可能引起生长延迟和显著的全身异常，并且频繁地需要终身螯合疗法。尽管HbA构成成人RBC中大多数血红蛋白，但大约3％的成人血红蛋白是HbA2形式，HbA变体是两个γ-珠蛋白链被两个δ(Δ)-珠蛋白链替换。δ-Thal与引起HBD表达损失的Δ血红蛋白基因(HBD)突变有关。HBD突变的共遗传可以通过将HbA2水平降低至正常范围来掩盖β-Thal(即，β/δ-Thal)的诊断(Bouva 2006)。β/δ-Thal通常由两个等位基因中HBB和HBD序列的缺失引起。在纯合的(δo/δoβo/βo)患者中，表达HBG，导致单独产生HbF。

[0011] 与SCD一样，β-Thal是由HBB基因突变引起的。导致β-Thal的最常见HBB突变为:c.-136C>G、c.92+1G>A、c.92+6T>C、c.93-21G>A、c.118C>T、c.316-106C>G、c.25_26delAA、c.27_28insG、c.92+5G>C、c.118C>T、c.135delC、c.315+1G>A、c.-78A>G、c.52A>T、c.59A>G、c.92+5G>C、c.124_127delTTCT、c.316-197C>T、c.-78A >G、c.52A>T、c.124_127delTTCT、c.316-197C>T、c.-138C>T、c.-79A>G、c.92+5G>C、c.75T>A、c.316-2A>G和c.316-2A>C。与β-Thal有关的这些和其他突变引起β-珠蛋白链的突变或缺失，这导致正常Hbα-血红蛋白与β-血红蛋白比率的破坏。过量的α-珠蛋白链在骨髓中的红系前体中沉淀。

[0012] 在重型β-Thal中，HBB的两个等位基因都含有无意义突变、移码突变、或剪接突变，导致完全不存在β-珠蛋白产生(表示为β^0/β^0)。重型β-Thal导致β-珠蛋白链的严重减少，导致RBC中α-珠蛋白链的显著沉淀和更严重的贫血。

[0013] 中间型β-Thal由HBB的5’或3’非翻译区域突变、启动子区域突变、或HBB多聚腺苷酸化信号或HBB基因内的剪接突变导致。患者基因型表示为βo/β+或β+/β+。βo代表不存在β-珠蛋白链的表达;β+代表功能失调但存在的β-珠蛋白链。表型表达因患者而异。由于存在一些β-珠蛋白的产生，中间型β-Thal导致红系前体中α-珠蛋白链的沉淀较少，并且与重型β-Thal相比导致较少的严重贫血。然而，继发于慢性贫血的红系谱系扩增有更显著的后果。

[0014] 具有重型β-Thal的受试者存在于6个月和2岁之间，并且患有未能茁壮成长、发热、肝脾大和腹泻。足够的治疗包括定期输血。重型β-Thal疗法还包括脾切除术和羟基脲治疗。如果患者定期输血，他们将正常发育，直到第二个十年开始。那时，他们需要螯合疗法(除了继续输血)以防止铁超过载的并发症。铁过载可能表现为生长延迟或性成熟延迟。在成人期，不充分的螯合疗法可能导致心肌病、心律失常、肝纤维化和/或肝硬化、糖尿病、甲状腺和甲状旁腺异常、血栓症和骨质疏松症。频繁的输血还使受试者有感染HIV、乙型肝炎、和丙型肝炎的风险。

[0015] 中间型β-Thal受试者通常存在于2岁至6岁之间。他们通常不需要输血。然而，由于红系谱系的慢性肥大而发生骨异常以补偿慢性贫血。由于骨质疏松症，受试者可能有长骨骨折。髓外红细胞生成是常见的并且导致脾，肝和淋巴结的扩大。它还可能引起脊髓压缩和神经系统问题。受试者还患有下肢溃疡并且血栓形成事件的风险增加，包括中风、肺栓塞和深静脉血栓形成。中间型β-Thal的治疗包括脾切除术、叶酸补充、羟基脲疗法和髓外肿块的放射治疗。螯合疗法用于发生铁过载的受试者。

[0016] β-Thal患者的生命期望通常会降低。患有重型β-Thal且未接收输血疗法的受试者通常在其第二或第三个十年死亡。接收常规输血和足够螯合疗法的重型β-Thal受试者可以活到第五个十年甚至更长时间。继发于铁毒性的心脏衰竭是由于铁毒性导致的重型β-Thal受试者死亡的主要原因。

[0017] 目前正在发育SCD和β-Thal的各种新治疗。目前正在临床试验中研究经由基因疗法递送抗镰状化HBB基因。然而，这种途径的长期功效和安全性尚不清楚。已经证明用来自HLA匹配的异基因干细胞供体的造血干细胞(HSC)移植治疗SCD和β-Thal，但是所述方法涉及风险，包括与切除疗法有关的风险，这要求准备受试者进行移植，增加威胁生命的机会性感染的风险，和移植后移植物抗宿主疾病的风险。另外，通常无法识别匹配的异基因供体。因此，需要改进的管理这些和其他血红蛋白病的方法。

发明内容

[0018] 本文提供了基因组编辑系统、核糖核蛋白(RNP)复合物、指导RNA、Cpf1蛋白(包括经修饰的Cpf1蛋白(Cpf1变体))以及用于改变一个或多个γ-珠蛋白基因(例如HBG1、HBG2或HBG1和HBG2)启动子区域并增加胎儿血红蛋白(HbF)的表达的CRISPR介导方法。在某些实施例中，RNP复合物可以包括与野生型Cpf1或经修饰的Cpf1 RNA指导的核酸酶(经修饰的Cpf1蛋白)复合的指导RNA(gRNA)。在某些实施例中，gRNA可包含表13、表18、或表19所示的序列。在某些实施例中，gRNA可包含gRNA靶向结构域。在某些实施例中，gRNA靶向结构域可包含选自由SEQ ID NO:1002、1254、1258、1260、1262和1264组成的组的序列。在某些实施例中，gRNA可包含表19所示的gRNA序列。在某些实施例中，gRNA可包含选自由SEQ ID NO:1022、1023、1041-1105组成的组的序列。在某些实施例中，RNP复合物可包含表21所示的RNP复合物。例如，RNP复合物可以包括包含SEQ ID NO:1051所示的序列的gRNA和由SEQ ID NO:1097所示的序列编码的经修饰的Cpf1蛋白(RNP32，表21)。

[0019] 在本文中发明人发现，包括与经修饰的Cpf1蛋白复合的gRNA的RNP复合物的递送可能导致靶核酸的编辑增加。在某些实施例中，所述经修饰的Cpf1蛋白可含有一个或多个修饰。在某些实施例中，所述一个或多个修饰可以包括但不限于野生型Cpf1氨基酸序列中

的一个或多个突变、野生型Cpf1核酸序列中的一个或多个突变、一个或多个核定位信号(NLS)、一个或多个纯化标签(例如,His标签)或其组合。在某些实施例中,经修饰的Cpf1可以由SEQ ID NO:1000、1001、1008-1018、1032、1035-39、1094-1097、1107-09(Cpf1多肽序列)或SEQ ID NO:1019-1021、1110-17(Cpf1多核苷酸序列)所示的序列编码。在某些实施例中,gRNA可以是经修饰的或未经修饰的gRNA。在某些实施例中,gRNA可包含表13、表18或表19所示的序列。在某些实施例中,RNP复合物可包含表21所示的RNP复合物。例如,RNP复合物可以包括包含SEQ ID NO:1051所示的序列的gRNA和由SEQ ID NO:1097所示的序列编码的经修饰的Cpf1蛋白(RNP32,表21)。在某些实施例中,包含经修饰的Cpf1蛋白的RNP复合物可以增加靶核酸的编辑。在某些实施例中,包含经修饰的Cpf1蛋白的RNP复合物可以增加编辑,导致有效indel的增加。在各种实施例中,可以通过技术人员已知的任何方式来评估靶核酸的编辑增加,例如但不限于,靶核酸的PCR扩增和随后的测序分析(例如,桑格测序,下一代测序)。

[0020] 在本文中发明人还发现,包括与未经修饰或经修饰的Cpf1蛋白复合的经修饰的gRNA的RNP复合物的递送可能导致靶核酸的编辑增加。在某些实施例中,所述经修饰的gRNA可包含一个或多个如下修饰,所述修饰包括硫代磷酸酯键修饰,二硫代磷酸酯(PS2)键修饰、2'-O-甲基修饰、一个或多个或一段脱氧核糖核酸(DNA)碱基(在本文中也称为"DNA延伸")、一个或多个或一段核糖核酸(RNA)碱基(在本文中也称为"RNA延伸")或其组合。在某些实施例中,DNA延伸可包含表24所示的序列。例如,在某些实施例中,DNA延伸可包含SEQ ID NO:1235-1250所示的序列。在某些实施例中,RNA延伸可包含表24所示的序列。例如,在某些实施例中,RNA延伸可包含SEQ ID NO:1231-1234、1251-1253所示的序列。在某些实施例中,gRNA可包含表13、表18或表19所示的序列。在某些实施例中,RNP复合物可包含表21所示的RNP复合物。例如,RNP复合物可以包括包含SEQ ID NO:1051所示的序列的gRNA和由SEQ ID NO:1097所示的序列编码的经修饰的Cpf1蛋白(RNP32,表21)。在某些实施例中,包含经修饰的gRNA的RNP复合物可以增加靶核酸的编辑。在某些实施例中,包含经修饰的gRNA的RNP复合物可以增加编辑,导致有效indel的增加。

[0021] 在某些实施例中,包含经修饰的gRNA和经修饰的Cpf1蛋白的RNP复合物可以增加靶核酸的编辑。在某些实施例中,包含经修饰的gRNA和经修饰的Cpf1蛋白的RNP复合物可增加编辑,导致有效indel的增加。

[0022] 发明人还发现包含与具有"加强元件"的Cpf1分子复合的gRNA的RNP复合物(例如,"gRNA-Cpf1-RNP")的共递可以导致靶核酸的编辑增加。在某些实施例中,RNP复合物可包含表21所示的RNP复合物。例如,RNP复合物可以包括包含SEQ ID NO:1051所示的序列的gRNA和由SEQ ID NO:1097所示的序列编码的经修饰的Cpf1蛋白(RNP32,表21)。如本文所用,术语"加强元件"是指以下元件,所述元件当与包含跟RNA引导的核酸酶复合的gRNA的RNP复合物("gRNA-核酸酶-RNP")共递送时与没有所述加强元件情况下的靶核酸编辑相比增加所述靶核酸的编辑。在某些实施例中,可以将一种或多种加强元件与gRNA-核酸酶-RNP复合物共递送以增加靶核酸的编辑。在某些实施例中,加强元件的共递送可以增加编辑,从而导致有效indel的增加。在各种实施例中,可以通过技术人员已知的任何方式来评估靶核酸的编辑增加,例如但不限于,靶核酸的PCR扩增和随后的测序分析(例如,桑格测序,下一代测序)。

[0023] 在某些实施例中,gRNA-核酸酶-RNP可包含gRNA-Cpf1-RNP。在某些实施例中,

gRNA-Cpf1-RNP复合物的Cpf1分子可以是野生型Cpf1或经修饰的Cpf1。在某些实施例中，gRNA-Cpf1-RNP的Cpf1分子可以由SEQ ID NO:1000、1001、1008-1018、1032、1035-39、1094-1097、1107-09(Cpf1多肽序列)或SEQ ID NO:1019-1021、1110-17(Cpf1多核苷酸序列)所示的序列编码。在某些实施例中，gRNA-Cpf1-RNP复合物可包含gRNA，所述gRNA包含表13或表18所示的靶向结构域。在某些实施例中，gRNA-Cpf1-RNP复合物可包含gRNA，所述gRNA包含表19所示的序列。在某些实施例中，gRNA可以是经修饰的或未经修饰的gRNA。

[0024] 在某些实施例中，加强元件可包含失活RNP，所述失活RNP包含与RNA指导的核酸酶分子复合的失活gRNA分子(“失活gRNA-核酸酶-RNP”)。在某些实施例中，失活gRNA-核酸酶-RNP可包含与野生型(WT)Cas9分子复合的失活gRNA(“失活gRNA-Cas9-RNP”)，与Cas9切口酶分子复合的失活gRNA(“失活gRNA-切口酶-RNP”)或与酶失活(ei)Cas9分子复合的失活gRNA(“失活gRNA-eiCas9-RNP”)。在某些实施例中，失活gRNA-核酸酶-RNP复合物可具有降低的活性或缺乏核酸酶活性。在某些实施例中，失活gRNA-核酸酶-RNP复合物的失活gRNA可包含本文所述的任何失活gRNA。例如，失活gRNA可包含靶向结构域，所述靶向结构域可以与表10或表15所示的失活gRNA靶向结构域相同或差异不超过3个核苷酸。在某些实施例中，失活gRNA可以包括靶向结构域，所述靶向结构域包括gRNA靶向结构域的截短。在某些实施例中，待截短的gRNA靶向结构域可以是表2、表10或表15所示的gRNA靶向结构域。在某些实施例中，失活gRNA可以是经修饰的或未经修饰的失活gRNA。如本文中所示，gRNA-Cpf1-RNP与失活gRNA-Cas9-RNP(即，包含与WT Cas9复合的失活gRNA的RNP)的共递送或gRNA-Cpf1-RNP与失活gRNA-切口酶-RNP(即，包含与Cas9切口酶(即，Cas9 D10A切口酶)复合的失活gRNA的RNP)的共递送导致总编辑增加，高于单独递送gRNA-Cpf1-RNP(参见，例如，实例，15、17、18)后所观察到的水平。失活gRNA分子可包含与靶核酸中的靶区域(例如，CCAAT盒靶区域、13nt靶区域、近端HBG1/2启动子靶序列和/或BCL11Ae中的GATA1结合基序)近端或内的区域互补的靶向结构域。在某些实施例中，“近端”可表示靶区域(例如，CCAAT盒靶区域、13nt靶区域、近端HBG1/2启动子靶序列、和/或BCL11Ae中的GATA1结合基序)的10、25、50、100或200个核苷酸内的区域。在某些实施例中，一种或多种加强元件可以由有待与gRNA-Cpf1-RNP共递送的一种或多种失活gRNA-核酸酶-RNP(例如，失活gRNA-Cas9-RNP、失活gRNA-切口酶-RNP、失活gRNA-eiCas9-RNP)构成。在某些实施例中，失活gRNA-核酸酶-RNP的共递送不改变gRNA-Cpf1-RNP的indel谱。

[0025] 在某些实施例中，加强元件可包含RNP复合物，所述RNP复合物包含与RNA指导的核酸酶切口酶分子复合的gRNA分子(“gRNA-切口酶-RNP”)。在某些实施例中，RNA指导的核酸酶切口酶分子可以是Cas9切口酶分子，例如Cas9D10A切口酶。在某些实施例中，所述gRNA-切口酶-RNP的gRNA可包含本文所示的任何gRNA。例如，gRNA可包含表2、表10或表15所示的gRNA靶向结构域。在某些实施例中，gRNA可以是经修饰的或未经修饰的gRNA。如本文所示，gRNA-Cpf1-RNP与gRNA-切口酶-RNP复合物(包含与Cas9 D10A切口酶分子复合的指导RNA的RNP)的共递送导致总编辑增加，高于在递送单独gRNA-Cpf1-RNP后观察到的水平(参见，例如，实例15、16)。另外，gRNA-切口酶-RNP复合物与gRNA-Cpf1-RNP复合物的共递送改变gRNA-Cpf1-RNP的indel谱的方向性、长度和/或位置。在某些实施例中，可以使用加强元件来提供期望的编辑结果，例如以增加有效indel的比率。在某些实施例中，gRNA-切口酶-RNP复合物与gRNA-Cpf1-RNP复合物的共递送可以改变gRNA-Cpf1-RNP的indel谱。

[0026] 在某些实施例中,加强元件可包含单链寡脱氧核苷酸(ssODN)或双链寡脱氧核苷酸(dsODN)。在某些实施例中,ssODN可以是本文公开的任何ssODN。在某些实施例中,ssODN可包含表11所示的序列。例如,在某些实施例中,ssODN可包含SEQ ID NO:1040所示的序列。

[0027] 一方面,本公开涉及RNP复合物,其包含来自普雷沃菌属(Prevotella)和弗朗西斯菌属(Franciscella)的CRISPR 1(Cpf1)RNA指导的核酸酶或其变体和gRNA,其中所述gRNA能够结合细胞中的HBG基因启动子的靶位点。在某些实施例中,所述gRNA可以是经修饰的或未经修饰的。在某些实施例中,gRNA可包含一个或多个修饰,所述修饰包括硫代磷酸酯键修饰、二硫代磷酸酯(PS2)键修饰、2'-O-甲基修饰、DNA延伸、RNA延伸或其组合。在某些实施例中,DNA延伸可包含表24所示的序列。在某些实施例中,RNA延伸可包含表24所示的序列。在某些实施例中,gRNA可包含表13、表18或表19所示的序列。在某些实施例中,RNP复合物可包含表21所示的RNP复合物。例如,RNP复合物可以包括包含SEQ ID NO:1051所示的序列的gRNA和由SEQ ID NO:1097所示的序列编码的Cpf1变体蛋白(RNP32,表21)。在某些实施例中,Cpf1变体蛋白可含有一个或多个修饰。在某些实施例中,所述一个或多个修饰可以包括但不限于野生型Cpf1氨基酸序列中的一个或多个突变、野生型Cpf1核酸序列中的一个或多个突变、一个或多个核定位信号(NLS)、一个或多个纯化标签(例如,His标签)或其组合。在某些实施例中,Cpf1变体蛋白可以由SEQ ID NO:1000、1001、1008-1018、1032、1035-39、1094-1097、1107-09(Cpf1多肽序列)或SEQ ID NO:1019-1021、1110-17(Cpf1多核苷酸序列)所示的序列编码。

[0028] 一方面,本公开涉及改变细胞中HBG基因的启动子的方法,所述方法包括使细胞与本文公开的RNP复合物接触。在某些实施例中,所述改变可包含表17所示的一个或多个区域内的indel。在某些实施例中,所述改变可包含HBG基因的启动子的CCAAT盒靶区域内的indel。例如,在某些实施例中,所述改变可包含Chr 11(NC_000011.10):5,249,955-5,249,987(表17,区域6)、Chr 11(NC_000011.10):5,254,879-5,254,909(表17,区域16)内或其组合内的indel。在某些实施例中,RNP复合物可包含gRNA和Cpf1蛋白。在某些实施例中,gRNA可包含表19所示的RNA靶向结构域。在某些实施例中,gRNA靶向结构域可包含选自由SEQ ID NO:1002、1254、1258、1260、1262和1264组成的组的序列。在某些实施例中,gRNA可包含表19所示的gRNA序列。在某些实施例中,gRNA可包含选自由SEQ ID NO:1022、1023、1041-1105组成的组的序列。在某些实施例中,可以将gRNA配置为在如下处提供编辑事件:Chr11:5249973,Chr11:5249977(HBG1);Chr11:5250042,Chr11:5250046(HBG1);Chr11:5250055,Chr11:5250059(HBG1);Chr11:5250179,Chr11:5250183(HBG1);Chr11:5254897,Chr11:5254901(HBG2);Chr11:5254897,Chr11:5254901(HBG2);Chr11:5254966,5254970(HBG2);Chr11:5254979,5254983(HBG2)(表22,表23)。在某些实施例中,细胞可以进一步与加强元件接触。在某些实施例中,加强元件可包含单链寡脱氧核苷酸(ssODN)或双链寡脱氧核苷酸(dsODN)。在某些实施例中,ssODN可以是本文公开的任何ssODN。在某些实施例中,ssODN可包含表11所示的序列。例如,在某些实施例中,ssODN可包含SEQ ID NO:1040所示的序列。

[0029] 一方面,本公开涉及分离的细胞,所述分离的细胞包含通过将RNP复合物递送至细胞而产生的HBG基因的启动子中的改变。在某些实施例中,RNP复合物可包含gRNA和Cpf1蛋白。在某些实施例中,所述gRNA可以是经修饰的或未经修饰的。在某些实施例中,gRNA可包含一个或多个修饰,所述修饰包括硫代磷酸酯键修饰、二硫代磷酸酯(PS2)键修饰、2'-O-甲

基修饰、DNA延伸、RNA延伸或其组合。在某些实施例中，DNA延伸可包含表24所示的序列。在某些实施例中，RNA延伸可包含表24所示的序列。在某些实施例中，gRNA可包含表13、表18或表19所示的序列。在某些实施例中，RNP复合物可包含表21所示的RNP复合物。例如，RNP复合物可以包括包含SEQ ID NO:1051所示的序列的gRNA和由SEQ ID NO:1097所示的序列编码的Cpf1变体蛋白(RNP32，表21)。在某些实施例中，Cpf1变体蛋白可含有一个或多个修饰。在某些实施例中，所述一个或多个修饰可以包括但不限于野生型Cpf1氨基酸序列中的一个或多个突变、野生型Cpf1核酸序列中的一个或多个突变、一个或多个核定位信号(NLS)、一个或多个纯化标签(例如，His标签)或其组合。在某些实施例中，Cpf1变体蛋白可以由SEQ ID NO:1000、1001、1008-1018、1032、1035-39、1094-1097、1107-09(Cpf1多肽序列)或SEQ ID NO:1019-1021、1110-17(Cpf1多核苷酸序列)所示的序列编码。在某些实施例中，加强元件可以与RNP复合物共递送。在某些实施例中，加强元件可包含单链寡脱氧核苷酸(ssODN)或双链寡脱氧核苷酸(dsODN)。在某些实施例中，ssODN可以是本文公开的任何ssODN。在某些实施例中，ssODN可包含表11所示的序列。例如，在某些实施例中，ssODN可包含SEQ ID NO:1040所示的序列。

[0030] 一方面，本公开涉及增加人细胞中胎儿血红蛋白(HbF)水平的离体方法，所述方法是通过使用包含gRNA和Cpf1 RNA指导的核酸酶或其变体的RNP复合物进行基因组编辑，以实现HBG基因启动子的改变，从而增加HbF的表达。在某些实施例中，所述gRNA可以是经修饰的或未经修饰的。在某些实施例中，gRNA可包含一个或多个修饰，所述修饰包括硫代磷酸酯键修饰、二硫代磷酸酯(PS2)键修饰、2'-O-甲基修饰、DNA延伸、RNA延伸或其组合。在某些实施例中，DNA延伸可包含表24所示的序列。在某些实施例中，RNA延伸可包含表24所示的序列。在某些实施例中，gRNA可包含表13、表18或表19所示的序列。在某些实施例中，RNP复合物可包含表21所示的RNP复合物。例如，RNP复合物可以包括包含SEQ ID NO:1051所示的序列的gRNA和由SEQ ID NO:1097所示的序列编码的Cpf1变体蛋白(RNP32，表21)。在某些实施例中，Cpf1变体蛋白可含有一个或多个修饰。在某些实施例中，所述一个或多个修饰可以包括但不限于野生型Cpf1氨基酸序列中的一个或多个突变、野生型Cpf1核酸序列中的一个或多个突变、一个或多个核定位信号(NLS)、一个或多个纯化标签(例如，His标签)或其组合。在某些实施例中，Cpf1变体蛋白可以由SEQ ID NO:1000、1001、1008-1018、1032、1035-39、1094-1097、1107-09(Cpf1多肽序列)或SEQ ID NO:1019-1021、1110-17(Cpf1多核苷酸序列)所示的序列编码。在某些实施例中，加强元件可以与RNP复合物共递送。在某些实施例中，加强元件可包含单链寡脱氧核苷酸(ssODN)或双链寡脱氧核苷酸(dsODN)。在某些实施例中，ssODN可以是本文公开的任何ssODN。在某些实施例中，ssODN可包含表11所示的序列。例如，在某些实施例中，ssODN可包含SEQ ID NO:1040所示的序列。

[0031] 一方面，本公开涉及CD34+或造血干细胞的群体，其中群体中的一个或多个细胞包含HBG基因的启动子的改变，所述改变是通过将RNP复合物递送至CD34+或造血干细胞群体产生的，所述RNP复合物包含gRNA和Cpf1 RNA指导的核酸酶或其变体。在某些实施例中，所述gRNA可以是经修饰的或未经修饰的。在某些实施例中，gRNA可包含一个或多个修饰，所述修饰包括硫代磷酸酯键修饰、二硫代磷酸酯(PS2)键修饰、2'-O-甲基修饰、DNA延伸、RNA延伸或其组合。在某些实施例中，DNA延伸可包含表24所示的序列。在某些实施例中，RNA延伸可包含表24所示的序列。在某些实施例中，gRNA可包含表13、表18或表19所示的序列。在某

些实施例中,RNP复合物可包含表21所示的RNP复合物。例如,RNP复合物可以包括包含SEQ ID NO:1051所示的序列的gRNA和由SEQ ID NO:1097所示的序列编码的Cpf1变体蛋白(RNP32,表21)。在某些实施例中,Cpf1变体蛋白可含有一个或多个修饰。在某些实施例中,所述一个或多个修饰可以包括但不限于野生型Cpf1氨基酸序列中的一个或多个突变、野生型Cpf1核酸序列中的一个或多个突变、一个或多个核定位信号(NLS)、一个或多个纯化标签(例如,His标签)或其组合。在某些实施例中,Cpf1变体蛋白可以由SEQ ID NO:1000、1001、1008-1018、1032、1035-39、1094-1097、1107-09(Cpf1多肽序列)或SEQ ID NO:1019-1021、1110-17(Cpf1多核苷酸序列)所示的序列编码。在某些实施例中,加强元件可以与RNP复合物共递送。在某些实施例中,加强元件可包含单链寡脱氧核苷酸(ssODN)或双链寡脱氧核苷酸(dsODN)。在某些实施例中,ssODN可以是本文公开的任何ssODN。在某些实施例中,ssODN可包含表11所示的序列。例如,在某些实施例中,ssODN可包含SEQ ID NO:1040所示的序列。

[0032] 一方面,本公开涉及减轻有需要的受试者中镰状细胞病的一种或多种症状的方法,所述方法包括:a)从所述受试者中分离CD34+或造血干细胞的群体;b)通过以下修饰离体分离的细胞的群体:将包含gRNA和Cpf1 RNA指导的核酸酶或其变体的RNP复合物递送至分离的细胞的群体,从而实现所述群体中的一个或多个细胞中HBG基因启动子的改变;以及c)将经修饰的细胞群体施用于所述受试者,从而减轻所述受试者中镰状细胞病的一种或多种症状。在某些实施例中,所述改变可包含HBG基因的启动子的CCAAT盒靶区域内的indel。在某些实施例中,可以使用电穿孔递送所述RNP复合物。在某些实施例中,细胞群中至少约5%、至少约10%、至少约20%、至少约30%、至少约40%、至少约50%、至少约60%、至少约70%、至少约80%或至少约90%的细胞包含有效indel。

[0033] 一方面,本公开涉及gRNA,所述gRNA包含5'末端和3'末端,并且在所述5'末端包含DNA延伸,并且在所述3'末端包含2'-O-甲基-3'-硫代磷酸酯修饰,其中所述gRNA包括能够与靶位点杂交的RNA区段和能够与Cpf1 RNA指导的核酸酶缔合的RNA区段。在某些实施例中,DNA延伸可包含SEQ ID NO:1235-1250所示的序列。在某些实施例中,所述gRNA可以是经修饰的或未经修饰的。在某些实施例中,gRNA可包含一个或多个修饰,所述修饰包括硫代磷酸酯键修饰、二硫代磷酸酯(PS2)键修饰、2'-O-甲基修饰、DNA延伸、RNA延伸或其组合。在某些实施例中,DNA延伸可包含表24所示的序列。在某些实施例中,RNA延伸可包含表24所示的序列。在某些实施例中,gRNA可包含表13、表18或表19所示的序列。

[0034] 一方面,本公开涉及RNP复合物,其包含如本文公开的Cpf1 RNA指导的核酸酶和如本文公开的gRNA。

[0035] 本文还提供了基因组编辑系统、指导RNA和CRISPR介导的方法,其用于改变一种或多种γ-珠蛋白基因(例如HBG1、HBG2或HBG1和HBG2)、BCL11A基因(BCL11Ae)的红系特异性增强子、或其组合,并增加胎儿血红蛋白(HbF)的表达。在某些实施例中,包含表18或表19所示的序列的一种或多种gRNA可用于在HBG基因的启动子区域中引入改变。在某些实施例中,基因组编辑系统、指导RNA和CRISPR介导的方法可以改变13个核苷酸(nt)靶区域,所述13个核苷酸(nt)靶区域是HBG1、HBG2、或HBG1和HBG2基因的转录位点的5'("13nt靶区域")。在某些实施例中,基因组编辑系统、指导RNA和CRISPR介导的方法可以改变CCAAT盒靶区域,所述CCAAT盒靶区域是HBG1、HBG2、或HBG1和HBG2基因的转录位点的5'("CCAAT盒靶区域")。在某些实施例中,CCAAT盒靶区域可以是在远端CCAAT盒处或附近的区域,并且包括远端CCAAT盒

的核苷酸和远端CCAAT盒的上游(5')25个核苷酸和下游(3')25个核苷酸(即HBG1/2c.-86至-140)。在某些实施例中,CCAAT盒靶区域可以是在远端CCAAT盒处或附近的区域,并且包括远端CCAAT盒的核苷酸和远端CCAAT盒的上游(5')5个核苷酸和下游(3')5个核苷酸(即HBG1/2c.-106至-120)。在某些实施例中,CCAAT盒靶区域可包含18nt靶区域、13nt靶区域、11nt靶区域、4nt靶区域、1nt靶区域、-117G>A靶区域、或如本文所公开的其组合。在某些实施例中,所述改变可以是HBG1、HBG2、或HBG1和HBG2基因的18nt缺失、13nt缺失、11nt缺失、4nt缺失、1nt缺失、在c.-117处从G到A的取代、或其组合。在某些实施例中,改变可以是非天然发生的改变或天然发生的改变。在某些实施例中,包含SEQ ID NO:251-901或940-942所示的靶向结构域的一种或多种gRNA可以用于在13nt靶区域中引入改变。在某些实施例中,包含SEQ ID NO:251-901、940-942、996、997、970、971、1002或1003所示的序列的一种或多种gRNA可以用于在CCAAT盒靶区域中引入改变。在某些实施例中,基因组编辑系统、指导RNA和CRISPR介导的方法可以改变BCL11Ae中的GATA1结合基序(其位于BCL11A基因内含子2的+58DNA酶I超敏位点(DHS)区域中)("BCL11Ae中的GATA1结合基序")。在某些实施例中,包含SEQ ID NO:952-955所示的靶向结构域的一种或多种gRNA可以用于在BCL11Ae中的GATA1结合基序中引入改变。在某些实施例中,一种或多种gRNA可以用于在BCL11Ae中的GATA1结合基序中引入改变,并且一种或多种gRNA可以用于在HBG1和/或HBG2的13nt靶区域中引入改变。

[0036] 在某些实施例中,本文还提供了任选的基因组编辑系统组分,例如模板核酸(寡核苷酸供体模板)的用途。在某些实施例中,用于靶向CCAAT盒靶区域的模板核酸可以包括但不限于编码CCAAT盒靶区域的改变的模板核酸。在某些实施例中,CCAAT盒靶区域可包含18nt靶区域、11nt靶区域、4nt靶区域、1nt靶区域、或其组合。在某些实施例中,模板核酸可以是单链寡脱氧核苷酸(ssODN)或双链寡脱氧核苷酸(dsODN)。在某些实施例中,下面(例如,SEQ ID NO:904-909、974-995)也提供了5'和3'同源臂,以及编码CCAAT盒靶区域处的改变的示例性全长供体模板。在某些实施例中,模板核酸可以是正链或负链。在某些实施例中,ssODN可包含5'同源臂、替代序列和3'同源臂。在某些实施例中,5'同源臂可以是长度约25至约200个核苷酸或更多,例如,长度至少约25、50、75、100、125、150、175或200个核苷酸;所述替代序列可包含长度为0个核苷酸;并且3'同源臂可以是长度约25至约200个核苷酸或更多,例如,长度至少约25、50、75、100、125、150、175或200个核苷酸。在某些实施例中,ssODN可包含一个或多个硫代磷酸酯。

[0037] 在某些实施例中,用于改变一种或多种γ-珠蛋白基因(例如,HBG1,HBG2或HBG1和HBG2)的基因组编辑系统、指导RNA、CRISPR介导的方法可包括RNA指导的核酸酶。在某些实施例中,RNA指导的核酸酶可以是Cas9或经修饰的Cas 9。在某些实施例中,RNA指导的核酸酶可以是本文公开的Cpf1或经修饰的Cpf1。

[0038] 一方面,本公开涉及组合物,所述组合物包括以下:通过以上公开的方法产生的多个细胞,其中至少20%、30%、40%、50%、60%、70%、80%或90%的细胞包括人HBG1或HBG2基因的13nt靶区域的序列的改变;或通过上述方法产生的多个细胞,其中至少20%、30%、40%、50%、60%、70%、80%或90%的细胞包括人HBG1或HBG2基因的13nt靶区域的序列的改变并且至少20%、30%、40%、50%、60%、70%、80%或90%的细胞包括BCL11Ae中GATA1结合基序的序列的改变。在某些实施例中,多个细胞中的至少一部分可以在红系谱系内。在

某些实施例中，多个细胞的特征在于相对于未经修饰的多个细胞而言，胎儿血红蛋白表达水平的增加。在某些实施例中，胎儿血红蛋白的水平可以增加至少20％、30％、40％、50％、60％、70％、80％或90％。在某些实施例中，组合物可以进一步包括药学上可接受的载剂。

[0039] 一方面，本公开涉及改变细胞的方法，所述方法包括解旋细胞中核酸的靶区域内或近端的染色质区段，并在所述核酸的靶区域内产生双链断裂(DSB)，从而改变所述靶区域。在某些实施例中，解旋染色质区段的步骤可以包括使染色质区段与RNA指导的解旋酶接触。在某些实施例中，解旋染色质的步骤不包括将外源反式作用因子募集至染色质区段。RNA指导的解旋酶可以是RNA指导的核酸酶，并且RNA指导的核酸酶可以与包括长度为15个或更少的核苷酸的第一靶向结构域序列的失活指导RNA(dgRNA)复合。在某些实施例中，dgRNA可以包括在5’或3’末端的修饰，所述修饰包括但不限于在RNA的5’末端的抗反向帽类似物(ARCA)，在RNA的3’末端的聚A尾，或两者。在某些实施例中，RNA指导的解旋酶可以是酶活性的RNA指导的核酸酶，或者可以被配置为缺乏核酸酶活性。在某些实施例中，dgRNA的靶向结构域序列可以与靶区域近端的序列互补。在本文的一些实施例中，“近端”可以意指靶区域的10、25、50、100或200个核苷酸内。在某些实施例中，解旋染色质区段的步骤可以不包括在染色质区段内的核酸中形成单链或双链断裂。在某些实施例中，在靶区域内产生DSB的步骤可以包括使染色质区段与具有核酸酶活性的RNA指导的核酸酶接触。在某些实施例中，具有核酸酶活性的RNA指导的核酸酶可以与包括被配置为与靶区域重叠的靶向结构域的gRNA复合。在某些实施例中，具有核酸酶活性的RNA指导的核酸酶可以是Cpf1分子。

[0040] 本公开的另一方面包括诱导核酸的靶区域的可及性以在细胞中进行编辑的方法，所述方法包括使细胞与RNA指导的解旋酶和dgRNA接触，并用RNA指导的解旋酶解旋靶区域内或近端的DNA，从而诱导靶区域的可及性以进行编辑。在各种情况下，RNA指导的解旋酶和dgRNA可以被配置为在靶区域内或近端缔合。在某些实施例中，可以配置dgRNA，使其不提供RNA指导的核酸酶切割事件。在某些实施例中，RNA指导的解旋酶和dgRNA可以复合形成缺乏切割活性的失活核糖核蛋白(RNP)。在某些实施例中，dgRNA可包括长度为15个或更少核苷酸的靶向结构域序列。在某些实施例中，RNA指导的解旋酶可以是RNA指导的核酸酶。在某些实施例中，RNA指导的核酸酶和dgRNA未配置为将外源反式作用因子募集至靶区域。在某些实施例中，RNA指导的核酸酶可以是Cas9或Cas9融合蛋白。在某些实施例中，Cas9可以是酶活性的Cas9或酶失活的Cas9。在某些实施例中，Cas9可以是切口酶，例如Cas9 D10A。在某些实施例中，解旋DNA的步骤不包括在DNA中形成单链或双链断裂。在某些实施例中，具有核酸酶活性的RNA指导的核酸酶可以与包括被配置为与靶区域重叠的靶向结构域的gRNA复合。在某些实施例中，具有核酸酶活性的RNA指导的核酸酶可以是Cpf1分子。

[0041] 在另一方面，本公开涉及增加核酸中indel形成率的方法，所述方法包括使用RNA指导的解旋酶(被配置缔合到靶区域内或近端)解旋所述核酸的靶区域内或近端的双链DNA，并在靶区域内生成DSB。在某些实施例中，在靶区域内产生DSB导致在靶区域形成indel。在某些实施例中，可以以在靶区域形成indel的方式修复DSB。在某些实施例中，与不使用RNA指导的解旋酶实现的基因中的indel形成率相比，使用RNA指导的解旋酶实现的基因中的indel形成率增加。在某些实施例中，RNA指导的解旋酶可以与被配置为在靶区域内或近端缔合的dgRNA形成RNP复合物。在某些实施例中，dgRNA可包括长度为15个核苷酸或更短的靶向结构域序列。在某些实施例中，RNA指导的解旋酶可以是RNA指导的核酸酶。在某些

实施例中,RNA指导的核酸酶可以是Cas9或Cas9融合蛋白。在某些实施例中,Cas9可以是酶活性的Cas9或酶失活的Cas9。在某些实施例中,Cas9可以是切口酶,例如Cas9 D10A。在某些实施例中,RNA指导的核酸酶和dgRNA未配置为将外源反式作用因子募集至靶区域。在某些实施例中,解旋双链DNA的步骤不包括在DNA中形成单链或双链断裂。

[0042] 另一方面,本公开涉及缺失细胞中靶核酸的区段的方法,所述方法包括使所述细胞与RNA指导的解旋酶接触并在靶区域内产生DSB,由此靶核酸的区段酸被缺失。在某些实施例中,可以以缺失靶核酸的区段的方式修复DSB。在某些实施例中,RNA指导的解旋酶可以被配置为在靶核酸的靶区域内或近端缔合并解旋靶区域内或近端的双链DNA(dsDNA)。在某些实施例中,RNA指导的解旋酶可以与被配置为在靶区域内或近端缔合的dgRNA形成核糖核蛋白复合物。在某些实施例中,dgRNA可包括长度为15个核苷酸或更短的靶向结构域序列。在某些实施例中,RNA指导的解旋酶可以是RNA指导的核酸酶。在某些实施例中,RNA指导的核酸酶可以是Cas9或Cas9融合蛋白。在某些实施例中,Cas9可以是酶活性的Cas9或酶失活的Cas9。在某些实施例中,RNA指导的核酸酶和dgRNA未配置为将外源反式作用因子募集至靶区域。在某些实施例中,靶核酸可以是基因的启动子区、基因的编码区、基因的非编码区、基因的内含子、或基因的外显子。在某些实施例中,靶核酸的区段可以是至少约15、20、25、30、35、40、45、50、55、60、65、70、75或100个碱基对长度。

[0043] 本公开还涉及包含靶向结构域的失活gRNA(dgRNA)分子,所述靶向结构域包括gRNA靶向结构域的截短。在某些实施例中,待截短的gRNA靶向结构域可以是表2、表10或表15所示的gRNA靶向结构域。在某些实施例中,可以将gRNA靶向结构域从gRNA靶向结构域的5’末端截短。在某些实施例中,dgRNA可包括长度为15个核苷酸或更短的靶向结构域序列。在某些实施例中,第一靶向结构域可以与表10或表15所示的dgRNA靶向结构域相同或差异不超过3个核苷酸。

[0044] 本公开的另一个方面涉及包括至少一种编码多种gRNA的多核苷酸和RNA指导的解旋酶的组合物,其中至少一种gRNA可以是被配置使得它不提供RNA指导的核酸酶切割事件的dgRNA。在某些实施例中,dgRNA可包括长度为15个核苷酸或更短的靶向结构域序列。在某些实施例中,RNA指导的解旋酶可以是RNA指导的核酸酶。在某些实施例中,RNA指导的核酸酶可以是Cas9或Cas9融合蛋白。在某些实施例中,Cas9可以是酶活性的Cas9或酶失活的Cas9。在某些实施例中,Cas9可以是切口酶,例如Cas9 D10A。在某些实施例中,RNA指导的核酸酶和dgRNA未配置为将外源反式作用因子募集至靶区域。在某些实施例中,所述组合物还包括被配置为提供切割事件的第二RNA指导的核酸酶。在某些实施例中,所述组合物还包括被配置为提供切割事件第二gRNA。

[0045] 在另一方面,本公开涉及基因组编辑系统,所述系统包括RNA指导的核酸酶和RNA指导的解旋酶,其被配置为与靶核酸的靶区域近端的靶核酸缔合并诱导靶区域中的构象变化,由此促进靶区域对于RNA指导的核酸酶的可及性,从而在靶区域中形成断裂。本公开还涉及基因组编辑系统,所述基因组编辑系统包括dgRNA(其包括长度为15个核苷酸或更少的靶向结构域序列)、第一RNA指导的核酸酶、和RNA指导的解旋酶。在某些实施例中,基因组编辑系统进一步包括gRNA。在某些实施例中,gRNA和第一RNA指导的核酸酶可以与靶核酸中的靶区域缔合。在某些实施例中,第一RNA指导的核酸酶可以是Cpf1分子。在某些实施例中,gRNA和第一RNA指导的核酸酶可以与靶核酸中的第一PAM序列缔合,其中所述第一PAM序列

面向外。在某些实施例中，RNA指导的解旋酶可以是第二RNA指导的核酸酶。在某些实施例中，第二RNA指导的核酸酶和dgRNA未配置为将外源反式作用因子募集至靶区域。在某些实施例中，dgRNA和第二RNA指导的核酸酶在靶核酸中的靶区域内或近端缔合。在某些实施例中，第一RNA指导的核酸酶和第二RNA指导的核酸酶可以分别与gRNA和dgRNA复合，形成第一和第二核糖核蛋白复合物。

[0046] 另一方面，本公开涉及基因组编辑系统，所述基因组编辑系统包括dgRNA(其包含长度为15个核苷酸或更少的靶向结构域序列)、第一RNA指导的核酸酶、和RNA指导的解旋酶。在某些实施例中，gRNA和第一RNA指导的核酸酶可以与靶核酸中的靶区域缔合。在某些实施例中，第一RNA指导的核酸酶可以是Cpf1分子。在某些实施例中，gRNA和第一RNA指导的核酸酶可以与靶核酸中的第一原间隔子邻近基序(PAM)序列缔合。在某些实施例中，第一PAM序列可以面向外。在某些实施例中，RNA指导的解旋酶可以是第二RNA指导的核酸酶。在某些实施例中，第二RNA指导的核酸酶和dgRNA未配置为将外源反式作用因子募集至靶区域。在某些实施例中，dgRNA和第二RNA指导的核酸酶可以在靶核酸中的靶区域内或近端缔合。在某些实施例中，第一RNA指导的核酸酶和第二RNA指导的核酸酶可以分别与gRNA和dgRNA复合，形成第一和第二核糖核蛋白复合物。在某些实施例中，dgRNA和第二RNA指导的核酸酶可以与靶核酸中的第二PAM序列缔合，其中第二PAM序列可以面向外。

[0047] 另一方面，本公开涉及基因组编辑系统，所述基因组编辑系统包括dgRNA、第一gRNA(其包含长度大于17个核苷酸的第二靶向结构域序列)、第一RNA指导的核酸酶、和第二RNA指导的核酸酶。在某些实施例中，第一RNA指导的核酸酶和dgRNA可以被配置为在靶核酸的第一靶区域内缔合。在某些实施例中，第二RNA指导的核酸酶和第一gRNA可以被配置为在第二靶区域内缔合并在靶核酸中产生双链断裂(DSB)，从而在第一靶区域和第二靶区域之间产生indel。在某些实施例中，第二RNA指导的核酸酶可以是Cpf1分子。在某些实施例中，dgRNA可包含长度为15个核苷酸或更短的第一靶向结构域序列。在某些实施例中，dgRNA具有降低的或没有RNA指导的核酸酶切割活性。在某些实施例中，可以配置dgRNA，使其不提供RNA指导的核酸酶切割事件。在某些实施例中，dgRNA和第一RNA指导的核酸酶可以与靶核酸中的第一原间隔子邻近基序(PAM)序列缔合。在某些实施例中，第一PAM序列可以面向外。在某些实施例中，第一gRNA和第二RNA指导的核酸酶可与靶核酸中的第二PAM序列缔合。在某些实施例中，第二PAM序列可以面向外。

[0048] 在另一方面，本公开涉及改变细胞的方法，所述方法包括使所述细胞与dgRNA、第一gRNA(其包含长度大于17个核苷酸的第二靶向结构域序列)、第一RNA指导的核酸酶、和第二RNA指导的核酸酶接触。在某些实施例中，第一RNA指导的核酸酶和dgRNA可以被配置为在靶核酸的第一靶区域内缔合。在某些实施例中，第二RNA指导的核酸酶和第一gRNA可以在第二靶区域内缔合并在靶核酸中产生双链断裂(DSB)，从而在第一靶区域和第二靶区域之间产生indel。在某些实施例中，第二RNA指导的核酸酶可以是Cpf1分子。在某些实施例中，dgRNA可包含长度为15个核苷酸或更短的第一靶向结构域序列。在某些实施例中，dgRNA具有降低的或没有RNA指导的核酸酶切割活性。在某些实施例中，可以配置dgRNA，使其不提供RNA指导的核酸酶切割事件。在某些实施例中，dgRNA和第一RNA指导的核酸酶可以与靶核酸中的第一原间隔子邻近基序(PAM)序列缔合。在某些实施例中，第一PAM序列可以面向外。在某些实施例中，第一gRNA和第二RNA指导的核酸酶可与靶核酸中的第二PAM序列缔合。在某

些实施例中，第二PAM序列可以面向外。

[0049] 本文的公开还涉及改变细胞的方法，所述方法包括使所述细胞与本文公开的任何基因组编辑系统接触。在某些实施例中，接触细胞的步骤可以包括使细胞与包含第一和第二核糖核蛋白复合物的溶液接触。在某些实施例中，使细胞与溶液接触的步骤还包含电穿孔细胞，从而将第一和第二核糖核蛋白复合物引入细胞。

[0050] 在另一方面，本公开涉及使用本文公开的方法改变的细胞。本文还公开了包括导致HbF表达的有效indel的细胞。在某些实施例中，indel可通过以下来产生：使细胞与dgRNA、第一gRNA(其包括长度大于17个核苷酸的第二靶向结构域序列)、第一RNA指导的核酸酶、和第二RNA指导的核酸酶接触。在某些实施例中，第一RNA指导的核酸酶和dgRNA可以被配置为在靶核酸的第一靶区域内缔合。在某些实施例中，第二RNA指导的核酸酶和第一gRNA可以与第二靶区域缔合并在靶核酸中产生双链断裂(DSB)，从而在第一靶区域和第二靶区域之间产生indel。在某些实施例中，本文公开的细胞能够分化为成红细胞、红细胞或红细胞的前体或成红细胞的前体。在某些实施例中，细胞可以是CD34+细胞。

[0051] 包括上述所有特征中的任何一个的基因组编辑系统或方法可以包括包含人HBG1、HBG2基因或其组合的靶核酸。在某些实施例中，靶区域可以是人HBG1、HBG2基因或其组合的CCAAT盒靶区域。在某些实施例中，第一靶向结构域序列可以与人HBG1、HBG2基因或其组合的CCAAT盒靶区域一侧的第一序列互补，其中所述第一序列任选地与人HBG1、HBG2基因或其组合的CCAAT盒靶区域重叠。在某些实施例中，第二靶向结构域序列可以与人HBG1、HBG2基因或其组合的CCAAT盒靶区域一侧的第二序列互补，其中所述第二序列任选地与人HBG1、HBG2基因或其组合的CCAAT盒靶区域重叠。在某些实施例中，第一靶向结构域可包含gRNA靶向结构域的截短。在某些实施例中，gRNA靶向结构域可以包括表2、表10或表15所示的gRNA，并且gRNA靶向结构域已经从gRNA靶向结构域的5’末端截短。在某些实施例中，第一靶向结构域可以与表10或表15所示的dgRNA靶向结构域相同或差异不超过3个核苷酸。在某些实施例中，第二靶向结构域与表2、表10或表15所示的gRNA靶向结构域差异不超过3个核苷酸。在某些实施例中，indel可以改变CCAAT盒靶区域indel。在某些实施例中，indel可以是导致胎儿血红蛋白表达水平升高的有效indel。在某些实施例中，gRNA、dgRNA或两者可以体外合成或化学合成。

[0052] 在某些实施例中，细胞可以包括通过本文公开的改变细胞的任何方法产生的HBG基因座的至少一个经修饰的等位基因，其中HBG基因座的经修饰的等位基因包含人HBG1、HBG2基因或其组合的改变。

[0053] 在某些实施例中，可以通过本文公开的改变细胞的任何方法来修饰分离的细胞群体，其中所述细胞群体可以包括indel分布，所述indel分布可能与具有相同细胞类型的未通过所述方法修饰的分离的细胞群体或其后代不同。

[0054] 在某些实施例中，可通过本文公开的改变细胞的任何方法产生多个细胞，其中至少20％、30％、40％、50％、60％、70％、80％或90％的细胞可以包括人HBG1基因、HBG2基因或其组合的CCAAT盒靶区域中的序列的改变。

[0055] 在某些实施例中，本文公开的细胞可以用于药物。在某些实施例中，所述细胞可以用于治疗β-血红蛋白病。在某些实施例中，β-血红蛋白病可以选自由镰状细胞病和β-地中海贫血组成的组。

[0056] 一方面，本公开涉及组合物，所述组合物包括以下：通过包括上述dgRNA的方法产生的多个细胞，其中至少20％、30％、40％、50％、60％、70％、80％或90％的细胞包括人HBG1或HBG2基因的CCAAT盒靶区域的序列的改变；或通过上述方法产生的多个细胞，其中至少20％、30％、40％、50％、60％、70％、80％或90％的细胞包括人HBG1或HBG2的CCAAT盒靶区域的序列的改变。在某些实施例中，多个细胞中的至少一部分可以在红系谱系内。在某些实施例中，多个细胞的特征在于相对于未经修饰的多个细胞而言，胎儿血红蛋白表达水平的增加。在某些实施例中，胎儿血红蛋白的水平可以增加至少20％、30％、40％、50％、60％、70％、80％或90％。在某些实施例中，组合物可以进一步包括药学上可接受的载剂。

[0057] 一方面，本公开涉及通过基因组编辑系统修饰的细胞群体，所述基因编辑系统包括上述dgRNA，其中相对于未经所述基因编辑系统修饰的细胞群体，经所述基因编辑系统修饰的细胞群体包含更高百分比的有效indel。本公开还涉及通过包括上述dgRNA的基因组编辑系统修饰的细胞群体，其中相对于未经基因组编辑系统修饰的细胞群体，更高百分比的细胞群体能够分化为表达HbF的红系谱系细胞群体。在某些实施例中，更高百分比可以至少约15％、至少约20％、至少约25％、至少约30％或至少约40％或更高。在某些实施例中，所述细胞可以是造血干细胞。在某些实施例中，所述细胞能够分化为成红细胞、红细胞或红细胞的前体或成红细胞的前体。在某些实施例中，indel可以通过除微同源性介导的末端连接(MMEJ)修复以外的修复机制产生。

[0058] 本公开还涉及本文公开的任何细胞在制备用于治疗受试者的β-血红蛋白病的药物中的用途。

[0059] 一方面，本公开涉及在有需要的受试者中治疗β-血红蛋白病的方法，所述方法包括向所述受试者施用本文公开的细胞。在某些实施例中，在有需要的受试者中治疗β-血红蛋白病的方法可包括向所述受试者施用经修饰的造血细胞群体，其中根据本文公开的改变细胞的方法已经改变了一个或多个细胞。

[0060] 一方面，本公开涉及基因组编辑系统，其包括：RNA指导的核酸酶；以及第一指导RNA，其中所述第一指导RNA可包含与人HBG1、HBG2基因或其组合的CCAAT盒靶区域一侧的第一序列互补的第一靶向结构域，其中所述第一序列任选地与人HBG1、HBG2基因或其组合的CCAAT盒靶区域重叠。在某些实施例中，基因组编辑系统可进一步包含模板核酸，所述模板核酸编码人HBG1、HBG2基因或其组合的CCAAT盒靶区域的改变。在某些实施例中，模板核酸可以是单链寡脱氧核苷酸(ssODN)或双链寡脱氧核苷酸(dsODN)。在某些实施例中，ssODN可包含5’同源臂、替代序列和3’同源臂。在某些实施例中，同源臂在长度方面可以是对称的。在某些实施例中，同源臂在长度方面可以是不对称的。在某些实施例中，ssODN可包含一个或多个硫代磷酸酯修饰。在某些实施例中，一个或多个硫代磷酸酯修饰可以在5’末端、3’末端或其组合。在某些实施例中，ssODN可以是正链或负链。在某些实施例中，所述改变可以是非天然发生的改变。在某些实施例中，所述改变可包含CCAAT盒靶区域的缺失。在某些实施例中，所述缺失可包含18nt缺失、11nt缺失、4nt缺失、1nt缺失或其组合。在某些实施例中，CCAAT盒靶区域可包含18nt靶区域、11nt靶区域、4nt靶区域、1nt靶区域、或其组合。在某些实施例中，5’同源臂的长度可以是约25至约200个或更多个核苷酸，例如，长度是至少约25、50、75、100、125、150、175或200个核苷酸；所述替代序列可包含长度为0个核苷酸；并且3’同源臂的长度可以是约25至约200个或更多个核苷酸，例如长度是至少约25、50、75、100、125、

150、175或200个核苷酸。在某些实施例中，5’同源臂可包含约50至100bp，例如55至95、60至90、70至90或80至90bp与18nt靶区域的5’同源，3’同源臂可包含约50至100bp，例如55至95、60至90、70至90或80至90bp与18nt靶区域的3’同源。在某些实施例中，ssODN可包含以下、可以基本上由以下组成或由以下组成：SEQ ID NO:974或SEQ ID NO:975。在某些实施例中，5’同源臂可包含约50至100bp，例如55至95、60至90、70至90或80至90bp与11nt靶区域的5’同源，3’同源臂可包含约50至100bp，例如55至95、60至90、70至90或80至90bp与11nt靶区域的3’同源。在某些实施例中，ssODN可包含以下、可以基本上由以下组成或由以下组成：SEQ ID NO:976或SEQ ID NO:978。在某些实施例中，5’同源臂可包含约50至100bp，例如55至95、60至90、70至90或80至90bp与4nt靶区域的5’同源，3’同源臂可包含约50至100bp，例如55至95、60至90、70至90或80至90bp与4nt靶区域的3’同源。在某些实施例中，ssODN可包含以下、可以基本上由以下组成或由以下组成：选自由SEQ ID NO:984、SEQ ID NO:985、SEQ ID NO:986、SEQ ID NO:987、SEQ ID NO:988、SEQ ID NO:989、SEQ ID NO:990、SEQ ID NO:991、SEQ ID NO:992、SEQ ID NO:993、SEQ ID NO:994和SEQ ID NO:995组成的组的序列。在某些实施例中，5’同源臂可包含约50至100bp，例如55至95、60至90、70至90或80至90bp与1nt靶区域的5’同源，3’同源臂可包含约50至100bp，例如55至95、60至90、70至90或80至90bp与1nt靶区域的3’同源。在某些实施例中，同源臂在长度方面可以是对称的。在某些实施例中，ssODN可包含以下、可以基本上由以下组成或由以下组成：SEQ ID NO:982或SEQ ID NO:983。在某些实施例中，所述改变可以是天然发生的改变。在某些实施例中，所述改变可包含CCAAT盒靶区域的缺失或突变。在某些实施例中，所述CCAAT盒靶区域可包含13nt靶区域、-117G>A靶区域、或其组合。在某些实施例中，所述改变可包含在13nt靶区域的13nt缺失或在-117>A靶区域的从G至A的取代，或其组合。在某些实施例中，5’同源臂可包含约50至100bp，例如55至95、60至90、70至90或80至90bp与13nt靶区域的5’同源，3’同源臂可包含约50至100bp，例如55至95、60至90、70至90或80至90bp与13nt靶区域的3’同源。在某些实施例中，ssODN可包含以下、可以基本上由以下组成或由以下组成：SEQ ID NO:977或SEQ ID NO:979。在某些实施例中，5’同源臂可包含约50至100bp，例如55至95、60至90、70至90或80至90bp与13nt靶区域的5’同源，3’同源臂可包含约50至100bp，例如55至95、60至90、70至90或80至90bp与13nt靶区域的3’同源。在某些实施例中，ssODN可包含以下、可以基本上由以下组成或由以下组成：SEQ ID NO:980或SEQ ID NO:981。在某些实施例中，RNA指导的核酸酶可以是化脓链球菌(S.pyogenes)Cas9。在某些实施例中，RNA指导的核酸酶可以是本文公开的Cpf1变体。在某些实施例中，第一靶向结构域可以与表7、表18、表19列出的靶结构域或表12、表19中的gRNA差异不超过3个核苷酸。在某些实施例中，基因组编辑系统可以进一步包含第二指导RNA，其中所述第二指导RNA可包含第二靶向结构域，所述第二靶向结构域可以与人HBG1、HBG2基因或其组合的CCAAT盒靶区域一侧的第二序列互补，其中所述第二序列任选地与人HBG1、HBG2基因或其组合的CCAAT盒靶区域重叠。在某些实施例中，RNA指导的核酸酶可以是切口酶，并且任选地缺乏RuvC活性。在某些实施例中，基因组编辑系统可包含第一和第二RNA指导的核酸酶。在某些实施例中，第一和第二RNA指导的核酸酶可以分别与第一和第二指导RNA复合，形成第一和第二核糖核蛋白复合物。在某些实施例中，基因组编辑系统可进一步包含第三指导RNA；以及任选的第四指导RNA，其中所述第三和第四指导RNA可包含与第三和第四序列互补的第三和第四靶向结构域，所述第三和第四序列在人BCL11A基因的BCL11A红系增强子

(BCL11Ae)中GATA1结合基序的位置的相对侧,其中所述第三和第四序列之一或两者任选地与人BCL11A基因的BCL11Ae中的GATA1结合基序重叠。在某些实施例中,基因组编辑系统可以进一步包含编码BCL11Ae中的GATA1结合基序的缺失的核酸模板。在某些实施例中,RNA指导的核酸酶可以是化脓链球菌Cas9。在某些实施例中,RNA指导的核酸酶可以是切口酶,并且任选地缺乏RuvC活性。在某些实施例中,第三靶向结构域可与BCL11Ae中GATA1结合基序上游1000个核苷酸内的序列互补。在某些实施例中,第三靶向结构域可与BCL11Ae中GATA1结合基序上游100个核苷酸内的序列互补。在某些实施例中,第三和第四靶向结构域之一可以与BCL11Ae中GATA1结合基序下游100个核苷酸内的序列互补。在某些实施例中,第四靶向结构域可以与BCL11Ae中GATA1结合基序下游的50个核苷酸内的序列互补。在某些实施例中,第三和第四靶向结构域中的至少一个可以与表9中列出的靶向结构域差异不超过3个核苷酸。在某些实施例中,基因组编辑系统可包含第一和第二RNA指导的核酸酶。在某些实施例中,第一和第二RNA指导的核酸酶可以分别与第三和第四指导RNA复合,形成第三和第四核糖核蛋白复合物。

[0061] 一方面,本公开涉及改变细胞的方法,所述方法包括使细胞与基因组编辑系统接触。在某些实施例中,使细胞与基因组编辑系统接触的步骤可包含使细胞与包含第一和第二核糖核蛋白复合物的溶液接触。在某些实施例中,使细胞与溶液接触的步骤可以进一步包括电穿孔细胞,从而将第一和第二核糖核蛋白复合物引入细胞中。在某些实施例中,改变细胞的方法可以进一步包括使细胞与基因组编辑系统接触,其中使细胞与基因组编辑系统接触的步骤可包含使细胞与包含第一、第二、第三和任选的第四核糖核蛋白复合物的溶液接触。在某些实施例中,使细胞与溶液接触的步骤可以进一步包括电穿孔细胞,从而将第一、第二、第三和任选的第四核糖核蛋白复合物引入细胞中。在某些实施例中,所述细胞能够分化为成红细胞、红细胞或红细胞的前体或成红细胞的前体。在某些实施例中,细胞可以是$CD34^+$细胞。

[0062] 一方面,本公开涉及CRISPR介导的改变细胞的方法,所述方法包括:在人HBG1或HBG2基因的c.-106至-120位之间的细胞基因组中引入第一DNA单链断裂(SSB)或双链断裂(DSB);以及任选地在人HBG1或HBG2基因的c.-106至-120位之间的细胞基因组中引入第二SSB或DSB,其中所述第一和第二SSB或DSB可以由细胞以改变人HBG1或HBG2基因的CCAAT盒靶区域的方式修复。在某些实施例中,第一和第二SSB或DSB可以由细胞以导致人HBG1或HBG2基因的CCAAT盒靶区域的改变的方式修复。在某些实施例中,CRISPR介导的方法可以进一步包含编码人HBG1、HBG2基因或其组合的CCAAT盒靶区域的改变的模板核酸。在某些实施例中,所述模板核酸可以是单链寡脱氧核苷酸(ssODN)。在某些实施例中,ssODN可包含5'同源臂、替代序列和3'同源臂。在某些实施例中,所述ssODN可以是正链或负链。在某些实施例中,所述改变可以是非天然发生的改变。在某些实施例中,第一和第二SSB或DSB可以由细胞以下述方式修复,所述方式导致在人HBG1或HBG2基因的CCAAT盒靶区域中indel、缺失或插入中的至少一种的形成。在某些实施例中,CCAAT盒靶区域可包含18nt靶区域、11nt靶区域、4nt靶区域、1nt靶区域、或其组合。在某些实施例中,5'同源臂可以是长度约25至约200个核苷酸或更多,例如,长度至少约25、50、75、100、125、150、175或200个核苷酸;所述替代序列可包含长度为0个核苷酸;并且3'同源臂可以是长度约25至约200个核苷酸或更多,例如,长度至少约25、50、75、100、125、150、175或200个核苷酸。在某些实施例中,5'同源臂可包含约

50至100bp，例如55至95、60至90、70至90或80至90bp与18nt靶区域、11nt靶区域、4nt靶区域或1nt靶区域的5’同源，并且3’同源臂可包含约50至100bp，例如55至95、60至90、70至90或80至90bp与18nt靶区域、11nt靶区域、4nt靶区域或1nt靶区域的3’同源。在某些实施例中，所述ssODN可包含以下、可以基本上由以下组成或由以下组成：选自由SEQ ID NO:974、SEQ ID NO:975、SEQ ID NO:976、SEQ ID NO:978、SEQ ID NO:984、SEQ ID NO:985、SEQ ID NO:986、SEQ ID NO:987、SEQ ID NO:988、SEQ ID NO:989、SEQ ID NO:990、SEQ ID NO:991、SEQ ID NO:992、SEQ ID NO:993、SEQ ID NO:994、SEQ ID NO:995、SEQ ID NO:982和SEQ ID NO:983组成的组的序列。在某些实施例中，所述改变可以是非天然发生的改变。在某些实施例中，第一和第二SSB或DSB可以由细胞以下述方式修复，所述方式导致在人HBG1或HBG2基因的CCAAT盒靶区域中indel、缺失或插入中的至少一种的形成。在某些实施例中，所述CCAAT盒靶区域可包含13nt靶区域、-117G>A靶区域、或其组合。在某些实施例中，所述改变可包含在13nt靶区域的13nt缺失或在-117G>A靶区域的从G至A的取代，或其组合。在某些实施例中，5’同源臂可包含约50至100bp，例如55至95、60至90、70至90或80至90bp与13nt靶区域或-117G>A靶区域的5’同源，并且3’同源臂可包含约50至100bp，例如55至95、60至90、70至90或80至90bp与13nt靶区域或-117G>A靶区域的3’同源。在某些实施例中，ssODN可包含以下、可以基本上由以下组成或由以下组成：选自由SEQ ID NO:977或SEQ ID NO:979、SEQ ID NO:980或SEQ ID NO:981组成的组的序列。

[0063] 在一个方面，本公开涉及组合物，所述组合物可包含通过改变本文公开的改变细胞的方法产生的多个细胞，其中至少20％、30％、40％、50％、60％、70％、80％或90％的细胞可包含人HBG1基因、HBG2基因或其组合的CCAAT盒靶区域的序列的改变。在某些实施例中，所述改变可包含18nt缺失，11nt缺失，4nt缺失，1nt缺失，13nt缺失，人HBG1基因、HBG2基因或其组合在c.-117处从G到A的取代。在某些实施例中，多个细胞中的至少一部分可以在红系谱系内。在某些实施例中，多个细胞的特征在于相对于未经修饰的多个细胞而言，胎儿血红蛋白表达水平的增加。在某些实施例中，胎儿血红蛋白的水平可以增加至少20％、30％、40％、50％、60％、70％、80％或90％。在某些实施例中，组合物可以进一步包含药学上可接受的载剂。

[0064] 一方面，本公开涉及包含通过本文公开的改变细胞的方法产生的合成基因型的细胞，其中所述细胞可包含18nt缺失，11nt缺失，4nt缺失，1nt缺失，13nt缺失，人HBG1基因、HBG2基因或其组合在-117处从G到A的取代。

[0065] 在一个方面，本公开涉及细胞，所述细胞包含通过本文公开的改变细胞的方法产生的HBG基因座的至少一个等位基因，其中所述细胞可以编码18nt缺失、11nt缺失、4nt缺失、1nt缺失，13nt缺失，人HBG1基因、HBG2基因或其组合在-117处从G到A的取代。

[0066] 一方面，本公开涉及包含模板核酸的AAV载体，所述模板核酸编码人HBG1、HBG2基因或其组合的CCAAT盒靶区域的非天然发生的改变。在某些实施例中，所述模板核酸可以是单链寡脱氧核苷酸(ssODN)。在某些实施例中，CCAAT盒靶区域可包含18nt靶区域、11nt靶区域、4nt靶区域、1nt靶区域、或其组合。在某些实施例中，ssODN可包含5’同源臂、替代序列和3’同源臂。在某些实施例中，5’同源臂的长度可以是约25至约200个或更多个核苷酸，例如，长度是至少约25、50、75、100、125、150、175或200个核苷酸；所述替代序列可包含长度为0个核苷酸；并且3’同源臂的长度可以是约25至约200个或更多个核苷酸，例如长度是至少约

25、50、75、100、125、150、175或200个核苷酸。在某些实施例中，5’同源臂可包含约50至100bp，例如55至95、60至90、70至90或80至90bp与18nt靶区域、11nt靶区域、4nt靶区域或1nt靶区域的5’同源，并且3’同源臂可包含约50至100bp，例如55至95、60至90、70至90或80至90bp与18nt靶区域、11nt靶区域、4nt靶区域或1nt靶区域的3’同源。在某些实施例中，ssODN可包含以下、可以基本上由以下组成或由以下组成：选自由SEQ ID NO:974-976、SEQ ID NO:978、SEQ ID NO:982-995组成的组的序列。

[0067] 一方面，本公开涉及包含模板核酸的核苷酸序列，所述模板核酸编码人HBG1、HBG2基因或其组合的CCAAT盒靶区域的非天然发生的改变。在某些实施例中，模板核酸可以是包含所述改变的单链寡脱氧核苷酸(ssODN)或双链寡脱氧核苷酸(dsODN)。在某些实施例中，CCAAT盒靶区域可包含18nt靶区域、11nt靶区域、4nt靶区域、1nt靶区域、或其组合。在某些实施例中，ssODN可包含5’同源臂、替代序列和3’同源臂。在某些实施例中，5’同源臂的长度可以是约25至约200个或更多个核苷酸，例如，长度是至少约25、50、75、100、125、150、175或200个核苷酸；所述替代序列可包含长度为0个核苷酸；并且3’同源臂的长度可以是约25至约200个或更多个核苷酸，例如长度是至少约25、50、75、100、125、150、175或200个核苷酸。在某些实施例中，5’同源臂可包含约50至100bp，例如55至95、60至90、70至90或80至90bp与18nt靶区域、11nt靶区域、4nt靶区域或1nt靶区域的5’同源，并且3’同源臂可包含约50至100bp，例如55至95、60至90、70至90或80至90bp与18nt靶区域、11nt靶区域、4nt靶区域或1nt靶区域的3’同源。在某些实施例中，ssODN可包含以下、可以基本上由以下组成或由以下组成：选自由SEQ ID NO:974-976、SEQ ID NO:978、SEQ ID NO:982-995组成的组的序列。

[0068] 一方面，本公开涉及包含合成的基因型的细胞，其中所述细胞可包含18nt缺失，11nt缺失，4nt缺失，1nt缺失，13nt缺失，人HBG1基因、HBG2基因或其组合在-117处从G到A的取代。

[0069] 一方面，本公开涉及组合物，所述组合物包含通过本文所述的改变细胞的方法产生的多个细胞，其中相对于未经修饰的细胞群体而言，所述细胞包含人HBG1基因、HBG2基因或其组合的CCAAT盒靶区域的序列的更高的改变频率。在某些实施例中，更高的频率是高至少约10％、20％、30％、40％、50％、60％、70％、80％或90％。在某些实施例中，所述改变包含18nt缺失，11nt缺失，4nt缺失，1nt缺失，13nt缺失，人HBG1基因、HBG2基因或其组合在c.-117处从G到A的取代。在某些实施例中，细胞群体至少一部分在红系谱系内。

[0070] 该清单旨在是示例性和说明性的，而不是全面的和限制性的。其他方面和实施例可以在本公开和权利要求书的其余部分中阐明或从中显而易见。

附图说明

[0071] 附图旨在提供本公开的某些方面和实施例的说明性和示意性而非综合性的实例。图式并不旨在限制或结合于任何具体理论或模型，并且不一定成比例。不限制前文，可将核酸和多肽描绘为线性序列，或描绘为示意性二维或三维结构；这些描绘旨在具有说明性，而不是限制或束缚于关于其结构的任何具体模型或理论。

[0072] 图1以示意图形式在人第11号染色体上的β-珠蛋白基因簇的背景下描绘了一个或多个HBG1和HBG2基因。图1β-珠蛋白基因簇中的每个基因均受近端启动子的转录调控。尽管不希望受到任何特定理论的束缚，但通常认为A_γ和/或G_γ表达是通过近端启动子与远端强

红系特异性增强子(基因座控制区(LCR))之间的结合而激活的。LCR的长程反式激活被认为是通过染色质构型/确认的改变来介导的。LCR的特征是4个红细胞特异性DNA酶I超敏位点(HS1-4)和2个远端增强子元件(5'HS和3'HS1)。β样基因珠蛋白基因的表达以发育阶段特异性的方式调节,并且珠蛋白基因的表达变化与血液产生的主要部位的变化一致。

[0073] 图2A-2B描绘了HBG1和HBG2基因,编码序列(CDS)以及HBG1和HBG2近端启动子中和上游的小缺失和点突变(已在患者中鉴定出并与胎儿血红蛋白(HbF)升高有关)。近端启动子(CAAT盒,13nt序列)内的核心元件,其在某些具有遗传持续性胎儿血红蛋白(HPFH)的患者中已被缺失。还可以鉴定每个基因座的"靶序列"区域,所述区域已针对gRNA结合靶位点进行了筛选。

[0074] 图3显示来自用于在人K562红白血病细胞中掺入13nt缺失的gRNA筛选的数据。图3A通过T7E1内切核酸酶测定分析(可互换地称为"T7E1分析")确定的HBG1和HBG2基因座特异性PCR产物的基因编辑,所述PCR产物由编码化脓性链球菌特异性gRNA的DNA和编码化脓链球菌Cas9的质粒DNA进行电穿孔后,从K562细胞中提取的基因组DNA进行扩增而得到。图3B由PCR产物的DNA序列分析确定的基因编辑,所述PCR产物通过用编码所示gRNA和Cas9质粒的DNA电穿孔后,从K562细胞中提取的基因组DNA的HBG1基因座扩增而得到。图3C由PCR产物的DNA序列分析确定的基因编辑,所述PCR产物通过用编码所示gRNA和Cas9质粒的DNA电穿孔后,从K562细胞中提取的基因组DNA的HBG2基因座扩增而得到。对于图3B-C,编辑事件(插入,缺失)的类型和缺失的亚型(13nt靶部分地[12nt HPFH]或完全地[13nt至26nt HPFH]缺失,其他序列缺失[其他缺失])由不同的阴影/图案条指示。

[0075] 图4描绘了在用与体外转录的化脓链球菌gRNA(靶向用于缺失的特定13nt序列(HBG sgRNA Sp35和Sp37))复合的RNP进行电穿孔后,人脐带血(CB)和成人$CD34^+$细胞中的基因编辑结果。图4A描绘了通过T7E1分析检测到的HBG1和HBG2特异性PCR产物的indel的百分比,所述特异性PCR产物通过用指定RNP处理的CB $CD34^+$细胞或供体匹配的未经处理对照细胞(n=3CB $CD34^+$细胞,3个独立的实验)中提取的gDNA扩增而得到。示出的数据代表平均值,误差条相应于3个独立供体/实验的标准差。图4B描绘了通过T7E1分析检测到的HBG2特异性PCR产物的indel的百分比,所述特异性PCR产物通过用指定RNP处理的CB $CD34^+$细胞或成人$CD34^+$细胞或供体匹配的未经处理对照细胞(n=3CB $CD34^+$细胞,n=3动员的外周血(mPB)$CD34^+$细胞,3个独立的实验)中提取的gDNA扩增而得到。示出的数据代表平均值,误差条相应于3个独立供体/实验的标准差。图4C(顶部分图)描绘了通过T7E1分析检测到的用HBG Sp35 RNP或HBG Sp37 RNP+/-ssODN(未经修饰或PhTx修饰的5'和3'末端)电穿孔的人CB $CD34^+$细胞中提取的gDNA扩增的HBG2 PCR产物的indel。左下分图显示了通过桑格DNA序列分析确定的用HBG Sp37 RNP和ssODN编辑的细胞中gDNA的基因编辑水平。右下分图显示了在总缺失中检测到的特定缺失类型。

[0076] 图5描绘了用编码13nt缺失的HBG Sp37 RNP+/-ssODN电穿孔mPB $CD34^+$细胞后成人动员的外周血(mPB)$CD34^+$细胞中HBG的基因编辑和RNP处理的细胞的红系后代中胎儿血红蛋白的诱导。图5A描绘了从用RNP处理的mPB $CD34^+$细胞或供体匹配的未经处理对照细胞提取的gDNA扩增的HBG2 PCR产物的T7E1分析检测到的indel百分比。图5B描绘了从用RNP处理的和未处理的供体匹配的对照mPB $CD34^+$细胞分化的第7天成红细胞中HBG mRNA表达的倍数变化。将mRNA水平标准化至GAPDH,并校准至相应的分化天在未经处理的对照中检测到

的水平。

[0077] 图6描绘了来自相同供体的RNP处理的和未经处理的mPB CD34⁺细胞的离体分化潜能。图6A示出了造血骨髓/红系集落形成细胞(CFC)潜能,其中指示了集落的数量和亚型(GEMM:粒细胞-红系-单核细胞-巨噬细胞集落、E:红系集落、GM:粒细胞-巨噬细胞集落、M:巨噬细胞集落、G:粒细胞集落)。图6B描绘了在指示的时间点和指示的样品通过流式细胞术分析确定的红系分化时间过程中表达的血型糖蛋白A的百分比。

[0078] 图7A通过T7E1分析检测到的从用HBG RNP(D10A成对切口酶)处理的人mPB CD34⁺细胞提取的gDNA扩增的HBG PCR产物的indel。对于样品的子集,细胞还接受了编码13nt缺失的ssODN加上沉默的SNP来监测HDR(ssODN)。图7B描绘了用于选择图7A中所示的样品的子集的DNA测序分析。根据indel的类型(插入、13nt缺失或其他缺失)将indel细分。

[0079] 图8A描绘了用与D10A切口酶复合的指定gRNA对和WT RNP对电穿孔mPB CD34⁺细胞后HBG靶位点处的indel。图8B描绘了用与D10A切口酶复合的指定gRNA对和WT RNP电穿孔mPB CD34⁺细胞后的大缺失事件(例如HBG2的缺失)。图8C描绘了DNA测序分析和在来自用成对D10A切口酶对处理的mPB CD34⁺细胞的gDNA中检测到的事件亚型(插入,缺失)。图8D描绘了DNA测序分析和在来自用成对WT RNP对处理的mPB CD34⁺细胞的gDNA中检测到的事件亚型(插入,缺失)。

[0080] 图9描绘了用图7和图8中所示的实验中靶向HBG的成对RNP处理的mPB CD34⁺细胞的后代中的HbF蛋白和mRNA表达的总结。在RNP处理的人mPB CD34⁺细胞的红系后代中评估HbF蛋白(通过HPLC分析)和HbF mRNA表达(ddPCR分析)(从RNP处理的CD34⁺细胞后代中检测到的水平中减去在供体匹配的未经处理的对照中检测到的HbF背景水平)。

[0081] 图10描绘了在用不同浓度(0、2.5、3.7μM)的成对D10A切口酶RNP(SpA+Sp85)处理后,CD34⁺细胞的indel频率以及离体和体内短期造血潜能。通过T7E1分析(图10A)和通过Illumina测序分析(插入和缺失,图10B)评估indel。图10C描绘了通过HPLC分析检测到的HbF蛋白的%(%HbF=100%x HbF/(HbF+HbA)。图10D描绘了在集落形成细胞(CFC)测定中RNP处理的和供体匹配的未经处理对照CD34⁺细胞的造血活性。所示的CFC为每千接种的CD34⁺细胞。图10E描绘了供体匹配的人mPB CD34⁺(其未经处理(0μM)或用D10A RNP和成对gRNA的两种剂量(2.5和3.75μM)之一处理)移植后1个月免疫缺陷小鼠(NSG)中外周血的人血CD45⁺细胞重构。图10F描绘了移植2个月后免疫缺陷小鼠(NSG)中外周血的人血CD45⁺细胞重构。图10G和10H描绘了在1个月(图10G)和2个月(图10H)时的NSG小鼠的人CD45⁺血细胞重构之后的谱系分布。

[0082] 图11A对于两个D10A切口酶RNP对(SP37+SPB和SP37+SPA)以指定浓度递送至mPB CD34⁺细胞而言,通过HPLC测定的HbF水平和通过T7E1分析评估的indel频率。在经编辑的CD34⁺细胞的红系后代(第18天)中分析了HbF水平。从经编辑的样品中减去在供体匹配的未经处理对照中检测到的HbF蛋白。图11B描绘了与用指示的D10A切口酶对处理的CD34⁺细胞或未经处理的对照相关的造血集落形成细胞(CFC)活性重叠的indel比率。图11C描绘了在用给定浓度的D10 RNP切口酶对处理的mPB CD34⁺细胞或供体匹配的未经处理对照移植后一个月,免疫缺陷NSG小鼠的人CD45⁺血细胞重构。图11D描绘了在移植后一个月的小鼠外周血中的人CD45⁺级分中检测到的人血谱系分布。

[0083] 图12描述了针对HbF抑制的靶位点,BCL11A(BCL11Ae)(基因组座标:chr2:60,495,

265至60,495,270)的+58DNA酶I超敏位点(DHS)红细胞特异性增强子的GATA1基序。

[0084] 图13A描绘了通过T7E1内切核酸酶分析检测到的BCL11A PCR产物indel百分比,所述PCR产物是从用指示的RNP+/-ssODN处理的CB CD34$^+$细胞或供体匹配的未经处理对照细胞中提取的gDNA扩增得到。显示的数据代表三个3个独立供体/实验的平均值。图13B描绘了通过T7E1内切核酸酶分析检测到的BCL11A PCR产物的indel,以及每种情况下细胞的造血活性,所述PCR产物是用指示的WT RNP(与具有RuvC和HNH活性的WT化脓性链球菌Cas9复合的靶向BCL11A红系细胞增强子的单gRNA)或成对切口酶RNP(与具有相同的HNH单链剪切活性的化脓链球菌Cas9切口酶(例如D10A)复合的靶向BCL11A红系增强子的成对gRNA)处理的CB CD34$^+$细胞提取的gDNA中扩增而得到。

[0085] 图14A描绘了成人BM CD34$^+$细胞中BCL11Ae的编辑频率(使用靶向GATA1基序的单gRNA方法)。图14B描绘了通过DNA测序分析确定的在造血集落(GEMM,BCL11Ae RNP处理的CD34$^+$细胞的克隆后代)中检测到的单等位基因和双等位基因编辑。图14C描绘了成红细胞成熟的动力学(通过流式细胞术分析检测到的DRAQ5$^-$细胞所确定的去核)。图14D描绘了在分化的对照和经RNP处理的BM CD34$^+$细胞中的红系表型(血型糖蛋白A$^+$细胞)的获得,而图14E显示相对于未经处理的供体匹配对照样品中的HbF+细胞,通过流式细胞术分析确定的HbF$^+$细胞的增加倍数。

[0086] 图15描述了在成人mPB CD34$^+$细胞中BCL11Ae的基因编辑以及在用BCL11Ae RNP+非特异性ssODN电穿孔mPB CD34$^+$细胞后,RNP和ssODN处理的细胞的红系后代中胎儿血红蛋白的诱导。图15A描绘了从用BCL11Ae RNP和非特异性ssODN处理的mPB CD34$^+$细胞或供体匹配的未经处理对照细胞提取的gDNA扩增的HBG2 PCR产物的通过T7E1分析检测到的indel百分比。图15B描绘了从BCL11Ae RNP处理的和未经处理的供体匹配的对照mPB CD34$^+$细胞分化的第10天成红细胞中HBG mRNA表达的倍数变化(将mRNA水平标准化至GAPDH,并校准至相应的分化天在未经处理的对照中检测到的水平)。图15C描绘了在指示的时间点和针对指示的样品通过流式细胞术分析确定的+/-用BCL11Ae RNP和非特异性ssODN处理的mPB CD34$^+$细胞的红系分化过程中表达的血型糖蛋白A的百分比。

[0087] 图16描绘了从HBG PCR产物的下一代测序(NGS)检测到的indel的百分比,所述HBG PCR产物是从与浓度为16μM的化学合成的指导RNA OLI7066(SEQ ID NO:970,表10)复合的Cas9("OLI7066-RNP")处理的造血干/祖细胞(HSPC)提取的gDNA扩增而得到。鉴定出各种indel,包括HBGΔ-102:-121、HBGΔ-114:-124、HBGΔ-116、HBG-114+T、HBG-116+G;HBGΔ-112:-115、HBGΔ-113:-115、HBGΔ-114:-115、HBGΔ-115,HBGΔ-102:-114(天然发生的13nt缺失)。

[0088] 图17A-D描绘了与通过UPLC分析在用与gRNA OLI7066(SEQ ID NO:970)(OLI7066-RNP)(表10)复合的Cas9电穿的单个HSPC的红系后代中的HBG1或HBG2处携带的相对indel进行比较,由[γ链]/[全部γ链+β链]确定的Gγ(Gγ)-珠蛋白、Aγ(Aγ)-珠蛋白链(或由4.9kb缺失产生的AGγ(AGγ)-珠蛋白)的表达水平。图17A描绘了对于在相应的HBG1等位基因上携带指示的indel(HBG1Δ-115、HBG1Δ-114:-115、HBG1Δ-113:-115、HBG1Δ-112:-115、HBG1Δ-102:-114、HBG1Δ-104:-121、HBG1Δ-116)的克隆,由[AγT-珠蛋白链]/[全部γ链+β链]确定的AγT(AγT)-珠蛋白链表达。图17B描绘了对于在HBG2等位基因上携带指示的indel(HBG2Δ-115、HBG2Δ-114:-115、HBG2Δ-113:-115、HBG2Δ-112:-115、HBG2Δ-

102:-114、HBG2Δ-104:-121、HBG2Δ-116)的克隆，由[Gγ-γ链]/[全部γ链+β链]确定的Gγ(Gγ)-珠蛋白链表达。为确保对Gγ(Gγ)-珠蛋白诱导的分析是单个经编辑的等位基因的结果，仅分析具有HBG2单等位基因编辑或具有HBG2等位基因之一缺失(导致4.9kb缺失)的克隆。图17C描绘了对于在相应的HBG1/2等位基因上携带指示的indel(HBG1/2Δ-115、HBG1/2Δ-114:-115、HBG1/2Δ-113:-115、HBG1/2Δ-112:-115、HBG1/2Δ-102:-114、HBG1/2Δ-104:-121、HBG1/2Δ-116)的克隆，由[AGγT-γ链]/[全部γ链+β链]确定的AGγT(AGγT)-珠蛋白链表达。图17D描绘了对于在相应的HBG1/2等位基因上携带指示的indel(HBG1/2Δ-115、HBG1/2Δ-114:-115、HBG1/2Δ-113:-115、HBG1/2Δ-112:-115、HBG1/2Δ-102:-114、HBG1/2Δ-104:-121、HBG1/2Δ-116)的克隆，由[AGγI-γ链]/[全部γ链+β链]确定的AGγI(AGγI)-珠蛋白链表达。

[0089] 图18以示意图形式在人第11号染色体上的β-珠蛋白基因簇的背景下描绘了一个或多个HBG1和HBG2基因。该示意图显示了HBG1和HBG2上的CCAAT盒靶位点。由于该区域内的同源性，与RNA核酸酶(例如，Cas9)复合的单指导RNA，例如OLI8394(SEQ ID NO:971)将在HBG1和HBG2处剪切。与OLI8394(“OLI8394-RNP”)复合的Cas9递送后的编辑结果会有所不同，并导致大小不同的缺失或插入。设计单链寡脱氧核苷酸(ssODN)以提供可在CCAAT盒(表11)处复制所需的indel签名的模板。ssODN使用缺失序列侧翼的同源性臂“编码”相应的缺失，从而形成完全缺失。

[0090] 图19A-G描绘了成人CD34+细胞(来自用2μM OLI8394-RNP或OLI7066-RNP和2.5μM各种ssODN(表11)电穿孔的mPB)(“mPB CD34+细胞”)的HBG CCAAT盒靶区域的基因编辑结果。图19A描绘了通过用OLI8394-RNP和ssODN OLI16413(“-11nt+链”)或ssODN OLI16411(“-11nt-链”)电穿孔后72小时，对HBG PCR产物进行测序检测到的indel的百分比。图19B描绘了通过用OLI8394-RNP和ssODN OLI16413(“-11nt+链”)或ssODN OLI16411(“-11nt-链”)电穿孔后72小时，对HBG PCR产物进行测序检测到的精确“-11nt缺失”占总indel的百分比。图19C描绘了通过用OLI8394-RNP和ssODN OLI16430(“-4nt+链”)或ssODN OLI16424(“-4nt-链”)电穿孔后72小时，对HBG PCR产物进行测序检测到的indel的百分比。精确的-4nt缺失(即Δ-112:-115)的百分比与其他indel区分开来。图19D描绘了通过用OLI8394-RNP和ssODN OLI16418(“-1nt+链”)或ssODN OLI16417(“-1nt-链”)电穿孔后72小时，对HBG PCR产物进行测序检测到的indel的百分比。精确的-1nt缺失(即Δ-116)的百分比与其他indel区分开来。图19E描绘了通过用OLI8394-RNP和ssODN OLI16409(“-18nt+链”)或ssODN OLI16410(“-18nt-链”)电穿孔后72小时，对HBG PCR产物进行测序检测到的indel的百分比。精确的-18nt缺失(即Δ-104:-121)的百分比与其他indel区分开来。图19F描绘了通过用OLI7066-RNP和ssODN OLI16414(“-13nt+链”)或ssODN OLI16412(“-13nt-链”)电穿孔后72小时，对HBG PCR产物进行测序检测到的indel的百分比。精确的-13nt缺失(即Δ-102:-114)的百分比与其他indel区分开来。图19G描绘了通过用OLI8394-RNP和ssODN OLI16416(“-117G>A+链”)或ssODN OLI16415(“-117G>A -链”)电穿孔后72小时，对HBG PCR产物进行测序检测到的indel的百分比。带有-117G>A取代的读段的百分比(带有或不带有indel)与其他读段区分开来。

[0091] 图20A-B描绘了通过UPLC分析对用OLI8394-RNP或OLI7066-RNP和各种ssODN(表11)电穿孔的mPB CD34+细胞的红系后代所测得的γ-珠蛋白链相对于总β样珠蛋白链(γ

链/[γ链+β链])的表达水平。图20A描绘了通过UPLC测量的用以下项进行电穿孔后的γ-珠蛋白链相对于总β样珠蛋白链(γ链/[γ链+β链])的百分比:(i)OLI8394-RNP和OLI7066-RNP单独,(ii)OLI8394-RNP和ssODN OLI16430("-4nt+链"),ssODN OLI16424("-4nt-链"),ssODN OLI16413("-11nt+链")或ssODN OLI16411("-11nt-链"),和(iii)OLI7066-RNP和ssODN OLI16414("-13nt+链")或ssODN OLI16412("-13nt-链")。图20B描绘了用OLI8394-RNP和ssODN OLI16418("-1nt+链")、ssODN OLI16417("-1nt-链")、ssODN OLI16416("-117G>A+链")、ssODN OLI16415("-117G>A-链")、ssODN OLI16409("-18nt+链")、或ssODN OLI16410("-18nt-链")电穿孔后,通过UPLC测量的γ-珠蛋白链相比于总β样珠蛋白链(γ链/[γ链+β链])的百分比。

[0092] 图21A-D描绘了来自用2μM OLI8394-RNP和ssODN OLI16424("-4nt-链")(表11)以0.625μM至10μM剂量范围电穿孔的mPB CD34+细胞的基因编辑结果。图21A描绘了在电穿孔后72小时通过对HBG PCR产物进行测序检测到的indel的百分比。图21B描绘了在电穿孔后,通过ddPCR在HBG1和HBG2之间检测到4.9kb缺失的频率。图21C描绘了电穿孔后48小时,来自mPB的成人CD34+细胞的生存力百分比。图21D描绘了通过UPLC分析经电穿孔的细胞的红系后代的细胞裂解物测得的γ-珠蛋白链相比于总β样珠蛋白链(γ链/[γ链+β链])的百分比。

[0093] 图22A-D描绘了用指示剂量的OLI8394-RNP和指示剂量的ssODN OLI16424("-4nt-链")(表11)电穿孔的mPB CD34+细胞的基因编辑结果。图22A描绘了用指示剂量(2、4、或8μM)的OLI8394-RNP和指示剂量(0、1.25、2.5或5μM)的ssODN OLI16424("-4nt-链")电穿孔后72小时,对HBG PCR产物进行测序而检测到的indel的百分比。图22B描绘了用指示剂量(2、4、或8μM)的OLI8394-RNP和指示剂量(0、1.25、2.5或5μM)的ssODN OLI16424("-4nt-链")电穿孔后,对HBG PCR产物进行下一代测序(NGS)检测到的-4nt缺失("-112:-115缺失")的百分比。图22C描绘了用指示剂量(2、4、或8μM)的OLI8394-RNP和指示剂量(0、1.25、2.5或5μM)的ssODN OLI16424("-4nt-链")电穿孔后,HBG1和HBG2之间的4.9kb缺失的频率。通过ddPCR测量缺失。图22D描绘了用指示剂量(2、4、或8μM)的OLI8394-RNP和指示剂量(0、1.25、2.5或5μM)的ssODN OLI16424("-4nt-链")电穿孔后,通过UPLC分析mPB CD34+细胞红系后代的细胞裂解物测量的γ-珠蛋白链相比于总β样珠蛋白链(γ链/[γ链+β链])的百分比。

[0094] 图23A-B描绘了具有对称和不对称同源臂的ssODN模板提供的结果的和示意图。图23A描绘了由OLI8394(SEQ ID NO:971)和OLI7066(SEQ ID NO:970)靶向的HBG1和HBG2处的CCAAT盒靶位点。设计具有对称或不对称臂的ssODN,以提供复制HBG1和HBG2(表11)的-4nt缺失(HBG-112:-115)的模板。ssODN使用缺失序列侧翼的同源性臂"编码"相应的缺失,从而在HBG-112:-115处产生完全缺失。图23B描绘了用"编码"4nt缺失(HBG-112:-115)的2μM OLI8394-RNP和2.5μM各种ssODN、OLI16424("90/90")、OLI16419("40/80")或OLI16421("50/50")(表11)电穿孔mPB CD34+细胞后72小时,对HBG PCR产物进行测序检测到的indel的百分比。

[0095] 图24描绘了用D10ACas9(单独与Sp37和SpA gRNA(表12)复合("sp37-D10A-RNP+spA-D10A-RNP")或与ssODN OLI16424("-4nt-链")(表11)复合)电穿孔mPB CD34+细胞后72小时,对HBG PCR产物进行测序检测到的indel的百分比。精确的-4nt缺失(即Δ-112:-115)的百分比与其他indel区分开来。

[0096] 图25描绘了用氨基酸球菌属物种Cpf1的变体(“AsCpf1”)电穿孔的mPB CD34+细胞的HBG的基因编辑,所述变体即His-AsCpf1-nNLS(SEQ ID NO:1000)和His-AsCpf1-sNLS-sNLS(SEQ ID NO:1001),与指导RNA HBG1-1(OLI13620)(表13)复合(“His-AsCpf1-nNLS_HBG1-1 RNP”和“His-AsCpf1-sNLS-sNLS_HBG1-1 RNP”)。RNP以5μM或20μM电穿孔。

[0097] 图26A-C描绘了用His-AsCpf1-sNLS-sNLS_HBG1-1 RNP单独或与各种ssODN一起电穿孔的mPB CD34+细胞的HBG的基因编辑。图26A描绘了在用His-AsCpf1-sNLS-sNLS_HBG1-1 RNP单独或与OLI164324(“-4nt-链”)、OLI16430(“-4nt+链”)、OLI16410(“-18nt-链”)或OLI16409(“-18nt+链”)一起电穿孔后72小时,对HBG PCR产物进行测序检测到的indel的百分比。图26B描绘了在用His-AsCpf1-sNLS-sNLS_HBG1-1 RNP单独或与OLI16410(“-18nt链”)或OLI16409(“-18nt+链”)一起电穿孔后72小时,对HBG PCR产物进行测序检测到的精确18核苷酸缺失indel的百分比。图26C描绘了在用His-AsCpf1-sNLS-sNLS_HBG1-1 RNP电穿单独或与OLI16410(“-18nt-链”)或OLI16409(“-18nt+链”)一起电穿孔后72小时,对HBG PCR产物进行测序检测到的在所有indel中精确18nt缺失的百分比。

[0098] 图27A-F描绘了HBG1-1靶区域和化脓链球菌Cas9 gRNA对组合使用的示意图。图27A显示了HBG1-1 gRNA的靶区域(包含SEQ ID NO:1002所示的RNA靶向结构域,表15)。灰色框表示HBG启动子的远端CCAAT盒(即HBG1/2c.-111至-115)。图27B显示了HBG1-1的靶区域、HBG的远端CCAAT盒和靶区域SpA gRNA(包含SEQ ID NO:941的靶向结构域,表15)。图27C显示了HBG1-1的靶区域、HBG的远端CCAAT盒和靶区域SpG gRNA(包含SEQ ID NO:359的靶向结构域,表15)。图27D显示了HBG1-1的靶区域、HBG的远端CCAAT盒、tSpA失活gRNA(“dgRNA”)的靶区域(包含SEQ ID NO:326的靶向结构域,表15),以及Sp182dgRNA的靶区域(包含SEQ ID NO:1028的靶向结构域,表15)。图27E显示了HBG1-1的靶区域、HBG的远端CCAAT盒和tSpA dgRNA(包含SEQ ID NO:326的靶向结构域,表15)。图27F显示了HBG1-1的靶区域、HBG的远端CCAAT盒和Sp182 dgRNA的靶区域(包含SEQ ID NO:1028的靶向结构域,表15)。

[0099] 图28A-B描绘了通过使用靶向HBG启动子区域的HBG1-1-AsCpf1-RNP离体编辑mPB CD34+细胞实现的HbF表达。图28A显示了在mPB CD34+细胞中通过Amaxa电穿孔递送5μM或20μM HBG1-1-AsCpf1-RNP(“HBG-1-1”)后在HBG启动子区域进行编辑的结果。通过Amaxa电穿孔递送20μM HBG1-1-AsCpf1-RNP可获得高达约43%的编辑率和21%的HbF诱导(高于背景水平)。HbF水平用黑圈表示,该黑圈描绘了通过mPB CD34+细胞的类红细胞后代的UPLC分析测得的γ珠蛋白链占总β样珠蛋白链(γ链/[γ链+β链])的表达水平。灰色柱表示在电穿孔后72小时,通过HBG PCR产物的下一代测序(NGS)检测到的indel的百分比。图28B显示了在mPB CD34+细胞中通过MaxCyte电穿孔递送5μM或20μM HBG1-1-AsCpf1-RNP(“HBG-1-1”)后在HBG启动子区域进行编辑的结果。通过MaxCyte电穿孔递送20μM HBG1-1-AsCpf1-RNP可获得高达约16%的编辑率和7%的HbF诱导(高于背景水平)。HbF水平用黑圈表示,该黑圈描绘了通过mPB CD34+细胞的类红细胞后代的UPLC分析测得的γ珠蛋白链占总β样珠蛋白链(γ链/[γ链+β链])的表达水平。灰色柱表示在电穿孔后72小时,通过HBG PCR产物的NGS检测到的indel的百分比。

[0100] 图29描绘了在Maxcyte设备上通过共递送各种化脓链球菌Cas9 WT或Cas9D10A RNP,在HBG启动子区域处HBG1-1-AsCpf1H800A-RNP进行的增强的编辑。“化脓链球菌D10A”代表Cas9 D10A切口酶蛋白,并且“化脓链球菌WT”代表Cas9 WT蛋白。测试的RNP包括SpA-

D10A-RNP、SpG-D10A-RNP、tSpA-Cas9-RNP、Sp182-Cas9-RNP、和tSpA-Cas9-RNP+Sp182-Cas9-RNP(表14)。描绘了HBG1-1-AsCpf1H800A-RNP(“HBG1-1”)单独或与化脓链球菌的Cas9 RNP或RNP对组合一起递送后,在HBG启动子区域处的总编辑(灰色柱)和相关的HbF蛋白诱导(黑圈)。HbF水平用黑圈表示,该黑圈描绘了通过mPB CD34+细胞的类红细胞后代的UPLC分析测得的γ珠蛋白链占总β样珠蛋白链(γ链/[γ链+β链])的表达水平。灰色柱描绘了由HBG PCR产物的NGS检测到的indel的百分比。

[0101] 图30描绘了HBG1-1-AsCpf1H800A-RNP(“HBG1-1”)单独或与各种化脓链球菌Cas9 WT或Cas9D10A RNP组合一起MaxCyte递送后,mPB CD34+细胞的生存力。“化脓链球菌D10A”代表Cas9 D10A切口酶蛋白,并且“化脓链球菌WT”代表Cas9 WT蛋白。测试的RNP包括SpA-D10A-RNP、SpG-D10A-RNP、tSpA-Cas9-RNP、Sp182-Cas9-RNP、和tSpA-Cas9-RNP+Sp182-Cas9-RNP(表14)。在电穿孔后24小时通过DAPI染色和流式细胞术分析来测量生存力。

[0102] 图31A-B描绘了靶区域的HBG1-1-AsCpf1H800A-RNP和D10A-Cas9RNP的切割位点,以及HBG1-1-AsCpf1H800A-RNP与D10A RNP共递送所产生的编辑谱。图31A描绘了靶区域(浅灰色箭头)的每条链上的HBG1-1-AsCpf1H800A-RNP剪切位点的位置,以及由SpG-D10A-RNP和SpA-D10A-RNP靶向的切口位点的位置(黑色箭头)(表14)。图31B描绘了通过HBG PCR产物的NGS分析检测到的在电穿孔后72小时,在mPB CD34+中共递送HBG1-1-AsCpf1H800A-RNP(“HBG1-1RNP”)与SpG-D10A-RNP(“spG RNP”)或SpA-D10A-RNP(“spA RNP”)产生的编辑谱(表14)。X轴代表相对于HBG1-1-AsCpf1H800A-RNP正链切割位点的indel中心的基因组位置。Y轴表示indel的长度,其中缺失表示为负值,并且插入表示为正值。每个indel的总频率由符号的面积表示。描绘了以等于或大于0.1%的频率出现的indel。SpG和SpA靶位点用虚线表示。

[0103] 图32A-B描绘了通过电穿孔后72小时的HBG PCR产物的NGS分析检测到,Sp182-Cas9-RNP与HBG1-1-AsCpf1H800A-RNP共递送导致总indel以及远端CCAAT盒破坏性indel的加强(indel谱并未发生实质性改变)。图32A显示了用HBG1-1-AsCpf1H800A-RNP(“HBG1-1RNP”)单独或与Sp182-Cas9-RNP(sp182RNP)组合一起进行编辑后的indel谱。X轴代表相对于HBG1-1-AsCpf1H800A-RNP正链切割位点的indel中心的基因组位置。Y轴表示indel的长度,其中缺失表示为负值,并且插入表示为正值。每个indel的总频率由符号的面积表示。描绘了以等于或大于0.1%的频率出现的indel。Sp182靶位点由虚线表示。图32B描绘了破坏远端CCAAT盒序列中的无、1nt、2nt、3nt、4nt或全部5nt的indel的频率。

[0104] 图33描绘了与Sp182-Cas9-RNP共递送的HBG1-1-AsCpf1H800A-RNP的最佳剂量导致总编辑和HbF生产的增加(表14)。将HBG1-1-AsCpf1H800A-RNP(“HBG1-1”)作为与Sp182-Cas9-RNP(“Sp182”)一起的RNP对进行递送,在HBG启动子区域实现了>92%编辑(灰色柱),高达34%HbF诱导(高于背景)(黑色圆圈)。当Sp182-Cas9-RNP单独以12μM递送时,未观察到编辑。HbF水平用黑圈表示,该黑圈描绘了通过mPB CD34+细胞的类红细胞后代的UPLC分析测得的γ珠蛋白链占总β样珠蛋白链(γ链/[γ链+β链])的表达水平。灰色柱描绘了由HBG PCR产物的NGS检测到的indel的百分比。

[0105] 图34描绘了通过UPLC测量的用HBG1-1-AsCpf1H800A-RNP与Sp182-Cas9-RNP组合(表14)在HBG启动子区域编辑的单一人mPB CD34+细胞的克隆红系后代中γ链表达相比于总β样链(γ链/[γ链+β链])的水平分布。每个黑圈代表在通过在电穿孔后48小时通过FACS

分选分离的单一细胞衍生的克隆红系群体中检测到的γ-珠蛋白水平。

[0106] 图35A-C描绘了RNP(含有与His-AsCpf1-sNLS-sNLS H800A(SEQ ID NO:1032,表14)复合的经修饰的HBG1-1 gRNA(SEQ ID NO:1041,表14)("His-AsCpf1-sNLS-sNLS H800A_HBG1-1 RNP",在图35A-C中表示为"HBG1-1"))与增加浓度的ssODN OLI16431(SEQ ID NO:1040,表11)(在图35A-C中表示为"OLI16431")(表11)的共递送后的总编辑、HbF产生、生存力和集落形成潜力。图35A描绘了共递送6μM His-AsCpf1-sNLS-sNLS H800A_HBG1-1 RNP及增加浓度的ssODN OLI16431后,在远端CAATT盒处的编辑(灰色柱)和HbF诱导(黑圈)。HbF水平用黑圈表示,该黑圈描绘了通过mPB CD34+细胞的类红细胞后代的UPLC分析测得的γ珠蛋白链占总β样珠蛋白链(γ链/[γ链+β链])的表达水平。灰色柱描绘了由HBG PCR产物的NGS检测到的indel的百分比。图35B描绘了His-AsCpf1-sNLS-sNLS H800A_HBG1-1 RNP单独或与增加剂量的ssODN OLI16431组合递送后,mPB CD34+细胞的生存力。通过电穿孔后72小时的DAPI排除来测量生存力。图35C描绘了在集落形成细胞(CFC)测定中"HBG1-1"RNP和ssODN OLI16431处理的和供体匹配的未经处理对照$CD34^+$细胞的造血活性。所示的CFC为每800接种的$CD34^+$细胞。指示了集落的数量和亚型(GEMM:粒细胞-红系-单核细胞-巨噬细胞集落(黑色),GM:粒细胞-巨噬细胞集落(深灰色),E:红系集落(浅灰色))。

[0107] 图36A-B描绘了使用不同浓度的RNP(含有与His-AsCpf1-sNLS-sNLS H800A(SEQ ID NO:1032,表14)复合的未经修饰的HBG1-1 gRNA(SEQ ID NO:1022,表14)("His-AsCpf1-sNLS-sNLS H800A_HBG1-1 RNP",在图36A-B中表示为"HBG1-1"))与不同浓度的ssODN OLI16431(SEQ ID NO:1040,表11)(在图36A-B中表示为"OLI16431")(表11)共递送的总编辑、HbF产生、和生存力结果。图36A描绘了在共递送His-AsCpf1-sNLS-sNLS H800A_HBG1-1RNP("HBG1-1")和不同浓度的ssODN OLI16431之后,在远端CAATT盒(黑柱(48小时)和浅灰色柱(红系培养14天)处的编辑和HbF诱导(黑圈)。HbF水平用黑圈表示,该黑圈描绘了通过mPB CD34+细胞的类红细胞后代的UPLC分析测得的γ珠蛋白链占总β样珠蛋白链(γ链/[γ链+β链])的表达水平。柱描绘在48小时(黑色)和14天红系培养(浅灰色)时间点,由HBG PCR产物的NGS检测到的indel的百分比。图36B描绘了His-AsCpf1-sNLS-sNLS H800A_HBG1-1 RNP("HBG1-1")单独、ssODN OLI16431单独、或与不同剂量的HBG1-1和OLI16431组合递送后,mPB CD34+细胞的生存力。通过电穿孔后48小时和在红系培养中14天后的DAPI排除来测量生存力。在对mPB CD34+细胞进行编辑后,进行了18天的离体分化为红系谱系(Giarratana 2011)。在培养的第14天,分离了细胞的子集,并进行生存力(DAPI排除)和编辑(HBG PCR产物的NGS)测量。

[0108] 图37描绘了RNP在mPB CD34+细胞中的编辑。RNP包括与表21所示的Cpf1蛋白复合的gRNA。电穿孔后72小时,对分离的基因组DNA进行Illumina测序。

[0109] 图38A-B描绘了在电穿孔后48小时通过Illumina测序确定的混合CD34+细胞群体(黑色柱)、祖细胞(浅灰色柱)和HSC(深灰色柱)的编辑。图38A描绘了单独递送或与ssODN OLI16431(SEQ ID NO:1040,表11)共同递送的RNP33(表21)。图38B描绘了单独递送或与Sp182 RNP(包含与化脓链球菌Cas9(SEQ ID NO:1033)复合的SEQ ID NO:1027(表14)的失活gRNA)共递送的RNP33(表21)。

[0110] 图39描绘了在电穿孔后48小时通过Illumina测序确定的混合CD34+细胞群体(黑色柱)、祖细胞(浅灰色柱)和HSC(深灰色柱)的编辑。RNP34、RNP33和RNP43(表21)单独或与

Sp182 RNP(包含与化脓链球菌Cas9(SEQ ID NO:1033)复合的SEQ ID NO:1027(表14)的失活gRNA)或ssODN OLI16431(SEQ ID NO:1040,表11)组合递送。

[0111] 图40描绘了在电穿孔后72小时通过Illumina测序确定的CD34+细胞中的编辑。RNP64、RNP63和RNP45(表21)以2或4的化学计量(gRNA:Cpf1复合比)递送,其中gRNA摩尔过量。

[0112] 图41描绘了在电穿孔后72小时通过Illumina测序确定的CD34+细胞中的编辑。RNP33、RNP64、RNP63、和RNP45(表21)单独或与Sp182 RNP(包含与化脓链球菌Cas9(SEQ ID NO:1033)复合的SEQ ID NO:1027(表14)的失活gRNA)或ssODN OLI16431(SEQ ID NO:1040,表11)组合递送。

[0113] 图42描绘了通过Illumina测序确定的CD34+细胞中的编辑。包含与具有不同5'DNA延伸的gRNA复合的Cpf1(SEQ ID:1094)的RNP(表21)单独递送或与8μM OLI16431(SEQ ID NO:1040,表11)组合递送。

[0114] 图43描绘了通过Illumina测序确定的CD34+细胞中的编辑。将包含具有匹配的5'末端的gRNA的RNP(RNP49相比于RNP58和RNP59相比于RNP60,表21)递送至CD34+细胞,以评估3'修饰的影响。在这两个比较中,带有3'PS-OMe的gRNA在电穿孔后24小时的表现优于未经修饰的3'版本。

[0115] 图44A-B描绘了在电穿孔后24和48小时通过Illumina测序确定的CD34+细胞中的编辑。将RNP58(表21)以2:1、1:1或0.5:1摩尔比的化学计量(gRNA:Cpf1复合比)递送至CD34+细胞。在所有测试剂量下,当RNP以2:1的比例复合时,编辑效果最佳。

[0116] 图45A-B描绘了通过Illumina测序确定的CD34+细胞中的编辑。图45A和45B描绘了包含具有匹配的5'末端但不同的3'修饰的gRNA的RNP(表21),其递送至CD34+细胞以评估3'修饰或延伸的影响。

[0117] 图46A-C描绘了在递送靶向HBG基因座内的各种剪切位点的RNP之后,CD34+细胞及其红系后代中的编辑以及红系后代中的HbF水平。图46A描绘了包含指导RNA的RNP,所述指导RNA含有未经修饰的5'和在3'末端的1x PS-Ome(表21)。图46B描绘了包含指导RNA的RNP,所述指导RNA含有在5'的2PS+2ODNA延伸和在3'末端的1x PS-Ome(表21)。图46C描绘了包含指导RNA的RNP,所述指导RNA含有在5'的25DNA延伸和在3'末端的1x PS-Ome(表21)。

[0118] 图47描绘了在以1μM、2μM和4μM递送RNP58之后,CD34+细胞的红系后代中的编辑和HbF水平。

[0119] 图48描绘了在RNP的Maxcyte电穿孔后CD34+细胞中的编辑。将包含与不同的Cpf1蛋白(SEQ IDs:1094、1096、1107、1108)复合的gRNA SEQ ID:1051的RNP58、RNP26、RNP27和RNP28(表21)递送至CD34+细胞。在电穿孔后24和48小时通过Illumina测序确定编辑。

[0120] 图49描绘了在RNP的Maxcyte电穿孔后CD34+细胞中的编辑。将包含与具有各种5'延伸的指导RNA复合的Cpf1蛋白SEQ ID:1094的RNP58、RNP29、RNP30和RNP31(表21)递送到CD34+细胞中。在电穿孔后24和48小时通过Illumina测序确定编辑。由于细胞数限制,未在1μM下测试RNP30(nt)。

[0121] 图50描绘了在电穿孔后48小时通过Illumina测序确定的混合CD34+细胞群体(黑色柱)、祖细胞(深灰色柱)和HSC(浅灰色柱)的编辑。将RNP58、RNP27和RNP26(表21)以2μM或4μM递送至CD34+细胞。

[0122] 图51描绘了在电穿孔后48小时通过Illumina测序确定的混合CD34+细胞群体(黑色柱)、祖细胞(深灰色柱)和HSC(浅灰色柱)的编辑。将RNP61、RNP62和RNP34(表21)(8μM)与ssODN OLI16431(SEQ ID NO:1040,表11)(8μM)共递送至CD34+细胞。

[0123] 图52描绘了在电穿孔后48小时通过Illumina测序确定的混合CD34+细胞群体(黑色柱)、祖细胞(深灰色柱)和HSC(浅灰色柱)的编辑。将RNP58和RNP32(表21)以2μM递送至CD34+细胞。

[0124] 图53描绘了在电穿孔后48小时通过Illumina测序确定的混合CD34+细胞群体(黑色柱)、祖细胞(深灰色柱)和HSC(浅灰色柱)的编辑。将RNP58和RNP1(表21)以2μM、4μM或8μM递送至CD34+细胞。未对使用2μM RNP1编辑的细胞进行分选(N.S),并且因此无法使用编辑数据。

[0125] 图54描绘了经电穿孔的细胞输注后8周,来自"NBSGW"小鼠BM的植入的mPB CD34+细胞的indel。将RNP34和RNP33(表21)(8μM)与ssODN OLI16431(SEQ ID NO:1040,表11)(6μM)共递送至CD34+细胞。

[0126] 图55A-B描绘了经电穿孔的细胞输注后8周,植入的mPB CD34+细胞的indel和来自"NBSGW"小鼠的嵌合BM的红系细胞的HbF表达。RNP33或RNP34(表21)与Sp182 RNP(包含与化脓链球菌Cas9(SEQ ID NO:1033)复合的SEQ ID NO:1027(表14)的失活gRNA)(16μM总RNP)共递送至CD34+细胞。图55A描绘了在模拟转染(未添加RNP)或RNP转染的细胞中未分级的骨髓或按流分选的CD15+、CD19+、GlyA+和Lin-CD34+细胞的单个群体中的indel频率。Lin-CD34+细胞定义为以下CD34+细胞,其对CD3、CD14、CD15、CD16、CD19、CD20和CD56是阴性的,来自未经辐照的用模拟(无RNP)或RNP转染的mPB CD34+细胞输注的NOD,B6.SCID Il2rγ-/-Kit(W41/W41)("NBSGW")小鼠的骨髓(BM)。通过Illumina测序确定每个细胞群体的indel。图55B描绘了来自总嵌合BM的18天红系分化培养后,红系细胞裂解物的HbF表达,通过UPLC计算为γ/β样(%)。

[0127] 图56A-B描绘了经电穿孔的细胞输注后8周,植入的mPB CD34+细胞的indel和来自"NBSGW"小鼠的嵌合BM的红系细胞的HbF表达。将RNP61或RNP62(表21)(8μM)与ssODN OLI16431(SEQ ID NO:1040,表11)(8μM)共递送至CD34+细胞。图56A描绘了在模拟转染(未添加RNP)或RNP转染的细胞中未分级的骨髓或按流分选的CD15+、CD19+、GlyA+和Lin-CD34+细胞的单个群体的indel。Lin-CD34+细胞定义为以下CD34+细胞,其对CD3、CD14、CD15、CD16、CD19、CD20和CD56是阴性的,来自未经辐照的用模拟(无RNP)或RNP转染的mPB CD34+细胞输注的NBSGW小鼠的骨髓(BM)。通过Illumina测序确定每个细胞群体的indel。图56B描绘了来自总嵌合BM的18天红系分化培养后,红系细胞的HbF表达,通过UPLC计算为γ/β样(%)。

[0128] 图57描绘了在模拟(不添加RNP)mPB CD34+细胞或用RNP1(4或8μM)或RNP58(2、4或8μM)(表21)编辑的mPB CD34+细胞输注后8周骨髓内的人嵌合性。通过流式细胞术确定了BM中的人嵌合性和谱系重构(CD45+、CD14+、CD19+、血型糖蛋白A(GlyA、CD235a+)、谱系和CD34+以及小鼠CD45+标记表达)。

[0129] 图58描绘了在模拟(不添加RNP)mPB CD34+细胞或用RNP1(4或8μM)或RNP58(2、4或8μM)(表21)编辑的mPB CD34+细胞输注后8周未分选的混合骨髓内的indel。通过Illumina测序确定indel。

[0130] 图59描绘了在模拟转染(未添加RNP)或RNP转染的细胞中未分级的骨髓或按流分选的CD15+、CD19+、GlyA+和Lin-CD34+细胞的单个群体中的indel频率。Lin-CD34+细胞定义为以下CD34+细胞,其对CD3、CD14、CD15、CD16、CD19、CD20和CD56是阴性的,来自未经辐照的用模拟(无RNP)或RNP转染的mPB CD34+细胞输注的NOD,B6.SCID Il2rγ-/-Kit(W41/W41)("NBSGW")小鼠的骨髓(BM)。通过Illumina测序确定每个细胞群体的indel。

[0131] 图60描绘了模拟(不添加RNP)mPB CD34+细胞或经RNP58(2、4或8μM)(表21)编辑的mPB CD34+细胞输注后8周从骨髓分离的GlyA+级分中的HbF。

[0132] 图61描绘了在模拟的或经编辑的人动员的CD34+细胞输注后8周来自获取的骨髓冲洗物的细胞的集落形成潜力。指示了集落的数量和亚型(GEMM:粒细胞-红系-单核细胞-巨噬细胞集落(黑色),GM:粒细胞-巨噬细胞集落(深灰色),E:红系集落(浅灰色))。

[0133] 图62描述了表20所示的Cpf1蛋白变体的序列。核定位序列显示为粗体字母,六组氨酸序列显示为带下划线的字母。NLS序列的同一性和N末端/C末端位置的另外的排列(例如附加两个或更多个nNLS序列或nNLS和sNLS序列(或其他NLS序列)的组合至N末端/C末端位置,以及有和没有纯化序列(例如六组氨酸序列))的序列在本公开的主题范围内。

具体实施方式

定义和缩写

[0134] 除非另外指定,否则以下术语中的每一个具有此章节中与其相关的含义。

[0135] 不定冠词"一个"("a"和"an")是指至少一个相关名词,并且可与术语"至少一个"和"一个或多个"互换使用。例如,"一个模块"意指至少一个模块,或一个或多个模块。

[0136] 连词"或"和"和/或"可作为非排他析取词互换使用。

[0137] "结构域"用于描述蛋白质或核酸的区段。除非另外指明,否则不需要结构域具有任何特定功能特性。

[0138] 术语"外源反式作用因子"是指基因组编辑系统中的任何肽或核苷酸组分,两者(a)都通过修饰(例如肽或核苷酸插入或融合,针对RNA指导的核酸酶或gRNA)与RNA指导的核酸酶或gRNA相互作用,和(b)与靶DNA相互作用以改变其螺旋结构。肽或核苷酸的插入或融合可包括但不限于RNA指导的核酸酶或gRNA与外源反式作用因子之间的直接共价键,和/或由RNA/蛋白质相互作用结构域例如MS2环和蛋白质/蛋白质相互作用结构域(例如PDZ,Lim或SH1、2或3结构域)的插入或融合介导的非共价键。其他特定的RNA和氨基酸相互作用基序对于本领域技术人员将是熟悉的。反作用因子通常可以包括转录激活子。

[0139] 术语"加强元件"是指以下元件,所述元件当与包含跟RNA引导的核酸酶复合的gRNA的核糖核蛋白(RNP)复合物("gRNA-核酸酶-RNP")共递送时与没有所述加强元件情况下的靶核酸编辑相比增加所述靶核酸的编辑。在某些实施例中,共递送可以是顺序的或同时的。在某些实施例中,加强元件可以是RNP复合物,其由与WT Cas9蛋白、Cas9切口酶蛋白(例如,Cas9 D10A蛋白)或酶失活的Cas9(eiCas9)蛋白复合的失活指导RNA构成。在某些实施例中,加强元件可以是RNP复合物,其由与Cas9切口酶蛋白(例如,Cas9 D10A蛋白)或酶失活的Cas9(eiCas9)蛋白复合的指导RNA构成。在某些实施例中,加强元件可以是单链或双链供体模板DNA。在某些实施例中,可以将一种或多种加强元件与gRNA-核酸酶-RNP共递送以增加靶核酸的编辑。在某些实施例中,加强元件可以与包含跟Cpf1分子复合的gRNA的RNP

("gRNA-Cpf1-RNP")共递送以增加靶核酸的编辑。

[0140] "有效indel"是指导致HbF表达的indel(缺失和/或插入)。在某些实施例中,有效indel可以诱导HbF表达。在某些实施例中,有效indel可以导致HbF表达水平升高。

[0141] "indel"是核酸序列中的插入和/或缺失。indel可为DNA双链断裂的修复产物,该DNA双链断裂例如通过本公开的基因组编辑系统形成的双链断裂。indel最常在通过"错误倾向"修复路径(例如下文所述的NHEJ路径)修复断裂时形成。

[0142] "基因转变"是指通过并入内源同源序列(例如基因阵列内的同源序列)改变DNA序列。"基因修正"是指通过并入外源同源序列(例如外源单链或双链供体模板DNA)改变DNA序列。基因转变和基因修正是通过HDR路径(例如下文所述的那些)修复DNA双链断裂的产物。

[0143] Indel、基因转变、基因修正和其他基因组编辑结果典型地通过测序(最常通过"新一代(next-gen)"或"边合成边测序(sequencing-by-synthesis)"方法进行,但仍可使用Sanger测序)来评价,并且通过所有测序读段之间所关注位点处的数值变化(例如,±1、±2或更多个碱基)的相对频率来定量。测序用DNA样品可通过本领域中已知的多种方法来制备,并且可包括通过聚合酶链式反应(PCR)扩增所关注位点、捕捉通过双链断裂产生的DNA末端,如在Tsai 2016(通过引用并入本文)中所述的GUIDEseq方法中,或通过本领域熟知的其他手段。基因组编辑结果也可通过原位杂交法(例如FiberComb™系统,由基因组视觉公司(Genomic Vision)(法国巴涅)商品化)和通过本领域中已知的任何其他适宜方法来评价。

[0144] "alt-HDR"、"替代性同源定向修复"或"替代性HDR"可互换使用,是指使用同源核酸(例如,内源同源序列(例如姐妹染色单体)或外源核酸(例如模板核酸))修复DNA损伤的过程。alt-HDR与经典HDR的不同之处在于,该过程利用与经典HDR不同的路径,并且可以被经典HDR介体RAD51和BRCA2抑制。alt-HDR的不同之处还在于涉及单链或带切口同源核酸模板,而经典HDR通常涉及双链同源模板。

[0145] "经典HDR"、"经典同源定向修复"或"cHDR"是指使用同源核酸(例如,内源同源序列(例如,姐妹染色单体)或外源核酸(例如,模板核酸))修复DNA损伤的过程。当在双链断裂处已有显著切除时,典型HDR通常起作用,形成DNA的至少一个单链部分。在正常细胞中,cHDR通常涉及一系列步骤,诸如断裂的识别、断裂的稳定、切除、单链DNA的稳定、DNA交叉中间体的形成、交叉中间体的拆分以及连接。该过程需要RAD51和BRCA2,并且同源核酸典型地为双链。

[0146] 除非另外指明,否则如本文所用的术语"HDR"涵盖经典HDR和alt-HDR两者。

[0147] "非同源末端连接"或"NHEJ"是指连接介导的修复和/或非模板介导的修复,包括经典NHEJ(cNHEJ)和替代性NHEJ(altNHEJ),替代性NHEJ又包括微同源介导的末端连接(MMEJ)、单链退火(SSA)和合成依赖性微同源介导的末端连接(SD-MMEJ)。

[0148] 在关于分子的修饰(例如核酸或蛋白质)使用时,"替代"或"替代的"不需要方法限制,但仅指示替代实体是存在的。

[0149] "受试者"是指人、小鼠或非人灵长类动物。人受试者可为任何年龄(例如,婴儿、儿童、青年人或成年人),并且可患有疾病,并且可能实际上具有基因改变。

[0150] "治疗(Treat)"、"治疗(treating)"和"治疗(treatment)"意指治疗受试者(例如,人受试者)的疾病,包括以下各项中的一种或多种:抑制疾病,即,阻止或预防其发展或进展;缓解疾病,即,引起疾病状态消退;减轻疾病的一种或多种症状;和治愈疾病。

[0151] “预防(prevent)”、“预防(preventing)”和“预防(prevention)”是指预防受试者的疾病,包括:(a)避免或预先排除疾病;(b)影响朝向疾病的倾向;或(c)预防或延迟疾病的至少一种症状的发作。

[0152] “试剂盒”是指两种或更多种组分的任何集合,该两种或更多种组分一起构成可用于特殊目的的功能单元。通过说明(而不是限制),根据本公开的一个试剂盒可以包括与RNA指导的核酸酶复合或能够与该核酸酶复合的指导RNA,并且伴有(例如悬浮于,或可悬浮于)药学上可接受的载体。在某些实施例中,试剂盒可以包括加强元件。该试剂盒可用于将复合物引入例如细胞或受试者中,用于在这种细胞或受试者中引起所需基因组改变的目的。试剂盒的组分可以包装在一起,或者这些组分可分开包装。根据本公开的试剂盒还任选地包括使用说明书(DFU),其描述例如根据本公开的方法使用该试剂盒。DFU可以物理方式与试剂盒包装在一起,或者可以使试剂盒的使用者能获得该DFU,例如通过电子方式获得。

[0153] 术语“多核苷酸”、“核苷酸序列”、“核酸”、“核酸分子”、“核酸序列”和“寡核苷酸”是指DNA和RNA中的一系列核苷酸碱基(也称为“核苷酸”),并且意指两个或更多个核苷酸的任何链。多核苷酸、核苷酸序列、核酸等可以是嵌合混合物或其衍生物或经修饰的形式,单链或双链。它们可以在碱基部分、糖部分或磷酸骨架处经修饰,例如以改进分子的稳定性、其杂交参数等。核苷酸序列典型地携带遗传信息,包括但不限于细胞器用于制造蛋白质和酶的信息。这些术语包括双链或单链基因组DNA、RNA、任何合成的和遗传操作的多核苷酸,以及正义和反义多核苷酸二者。这些术语还包括含有经修饰碱基的核酸。

[0154] 常规IUPAC表示法用于本文所呈现的核苷酸序列中,如下表1中所示(还参见Cornish-Bowden A,Nucleic Acids Res.[核酸研究]1985年5月10日;13(9):3021-30,通过引用并入本文)。然而应注意,在序列可能由DNA或者RNA编码的那些情况下,例如在gRNA靶向结构域中,“T”表示“胸腺嘧啶或尿嘧啶”。

表1:IUPAC核酸表示法

符号	碱基
A	腺嘌呤
T	胸腺嘧啶或尿嘧啶
G	鸟嘌呤
C	胞嘧啶
U	尿嘧啶
K	G或T/U
M	A或C
R	A或G
Y	C或T/U
S	C或G
W	A或T/U
B	C、G或T/U
V	A、C或G
H	A、C或T/U
D	A、G或T/U

N	A、C、G或T/U

[0155] 术语“蛋白质”、“肽”和“多肽”可互换使用，是指通过肽键连接在一起的氨基酸的连续链。这些术语包括个别蛋白质、缔合在一起的蛋白质的组或复合物，以及此类蛋白质的片段或部分、变体、衍生物和类似物。肽序列使用常规表示法呈现于本文中，在左侧以氨基或N末端开始，并且前进至右侧的羧基或C末端。可以使用标准单字母或三字母缩写。

[0156] 符号“CCAAT盒靶区域”等是指HBG1和/或HBG2基因的转录起始位点(TSS)的5’的序列。CCAAT盒是α样和β样珠蛋白基因启动子区域内高度保守的基序。CCAAT盒内或附近的区域在珠蛋白基因调控中起重要作用。例如，γ-珠蛋白远端CCAAT盒与胎儿血红蛋白的遗传持久性有关。据报道，许多转录因子与γ-珠蛋白启动子的复制的CCAAT盒区域结合，例如NF-Y、COUP-TFII(NF-E3)、CDP、GATA1/NF-E1和DRED(Martyn 2017)。尽管不希望受到理论的束缚，但据信转录激活子NF-Y的结合位点在γ-珠蛋白启动子处与转录阻遏子重叠。存在于远端γ-珠蛋白启动子区域内(例如，在CCAAT盒内或附近)的HPFH突变可改变那些因子的竞争性结合，并因此促使γ-珠蛋白表达增加和HbF水平升高。本文为HBG1和HBG2提供的基因组位置基于NCBI参考序列NC_000011，“智人第11号染色体，GRCh38.p12初级组装”(版本NC_000011.10)中提供的坐标。HBG1和HBG2的远端CCAAT盒位于HBG1和HBG2 c.-111至-115(基因组位置分别为Hg38 Chr11:5,249,968至Chr11:5,249,972和Hg38 Chr11:5,254,892至Chr11:5,254,896)。HBG1 c.-111至-115区域示例在SEQ ID NO:902(HBG1)中的位置2823-2827，并且HBG2 c.-111至-115区域示例在SEQ ID NO:903(HBG2)中的位置2747-2751。在某些实施例中，“CCAAT盒靶区域”表示在远端CCAAT盒处或附近的区域，并且包括远端CCAAT盒的核苷酸和远端CCAAT盒的上游(5’)25个核苷酸和下游(3’)25个核苷酸(即HBG1/2c.-86至-140)(基因组位置分别为Hg38 Chr11:5249943至Hg38 Chr11:5249997和Hg38Chr11:5254867至Hg38 Chr11:5254921)。HBG1 c.-86至-140区域示例在SEQ ID NO:902(HBG1)中的位置2798-2852，并且HBG2 c.-86至-140区域示例在SEQ ID NO:903(HBG2)中的位置2723-2776。在其他实施例中，“CCAAT盒靶区域”表示在远端CCAAT盒处或附近的区域，并且包括远端CCAAT盒的核苷酸和远端CCAAT盒的上游(5’)5个核苷酸和下游(3’)5个核苷酸(即HBG1/2c.-106至-120(基因组位置为Hg38 Chr11:5249963至Hg38 Chr11:5249977(分别是HGB1和Hg38 Chr11:5254887至Hg38 Chr11:5254901))。HBG1 c.-106至-120区域示例在SEQ ID NO:902(HBG1)中的位置2818-2832，并且HBG2 c.-106至-120区域示例在SEQ ID NO:903(HBG2)中的位置2742-2756。术语“CCAAT盒靶位点改变”等是指CCAAT盒靶区域的一个或多个核苷酸的改变(例如，缺失、插入、突变)。示例性CCAAT盒靶区域改变的实例包括但不限于1nt缺失、4nt缺失、11nt缺失、13nt缺失和18nt缺失以及-117G>A改变。如本文所使用的，术语“CCAAT盒”和“CAAT盒”可以互换使用。

[0157] 符号“c.-114至-102区域”、“c.-102至-114区域”、“-102:-114”、“13nt靶区域”等是指分别位于基因组位置Hg38 Chr11:5,249,959至Hg38Chr11:5,249,971和Hg38 Chr11:5,254,883至Hg38 Chr11:5,254,895的HBG1和/或HBG2基因的转录起始位点(TSS)的5’的序列。HBG1 c.-102至-114区域示例在SEQ ID NO:902(HBG1)中的位置2824-2836，并且HBG2 c.-102至-114区域示例在SEQ ID NO:903(HBG2)中的位置2748-2760。术语“13nt缺失”等是指13nt靶区域的缺失。

[0158] 符号“c.-121至-104区域”、“c.-104至-121区域”、“-104:-121”、“18nt靶区域”等

是指分别位于基因组位置Hg38 Chr11:5,249,961至Hg38Chr11:5,249,978和Hg38 Chr11:5,254,885至Hg38 Chr11:5,254,902的HBG1和/或HBG2基因的转录起始位点(TSS)的5'的序列。HBG1 c.-104至-121区域示例在SEQ ID NO:902(HBG1)中的位置2817-2834,并且HBG2 c.-104至-121区域示例在SEQ ID NO:903(HBG2)中的位置2741-2758。术语"18nt缺失"等是指18nt靶区域的缺失。

[0159] 符号"c.-105至-115区域"、"c.-115至-105区域"、"-105:-115"、"11nt靶区域"等是指分别位于基因组位置Hg38 Chr11:5,249,962至Hg38Chr11:5,249,972和Hg38 Chr11:5,254,886至Hg38 Chr11:5,254,896的HBG1和/或HBG2基因的转录起始位点(TSS)的5'的序列。HBG1 c.-105至-115区域示例在SEQ ID NO:902(HBG1)中的位置2823-2833,并且HBG2 c.-105至-115区域示例在SEQ ID NO:903(HBG2)中的位置2747-2757。术语"11nt缺失"等是指11nt靶区域的缺失。

[0160] 符号"c.-115至-112区域"、"c.-112至-115区域"、"-112:-115"、"4nt靶区域"等是指分别位于基因组位置Hg38 Chr11:5,249,969至Hg38Chr11:5,249,972和Hg38 Chr11:5,254,893至Hg38 Chr11:5,254,896的HBG1和/或HBG2基因的转录起始位点(TSS)的5'的序列。HBG1 c.-112至-115区域示例在SEQ ID NO:902中的位置2823-2826,并且HBG2 c.-112至-115区域示例在SEQ ID NO:903(HBG2)中的位置2747-2750。术语"4nt缺失"等是指4nt靶区域的缺失。

[0161] 符号"c.-116区域"、"HBG-116"、"1nt靶区域"等是指分别位于基因组位置Hg38 Chr11:5,249,973和Hg38 Chr11:5,254,897的HBG1和/或HBG2基因的转录起始位点(TSS)的5'的序列。HBG1 c.-116区域示例在SEQ ID NO:902中的位置2822,并且HBG2 c.-116区域示例在SEQ ID NO:903(HBG2)中的位置2746。术语"1nt缺失"等是指1nt靶区域的缺失。

[0162] 符号"c.-117G>A区域"、"HBG-117G>A"、"-117G>A靶区域"等是指分别位于基因组位置Hg38 Chr11:5,249,974至Hg38 Chr11:5,249,974和Hg38Chr11:5,254,898至Hg38 Chr11:5,254,898的HBG1和/或HBG2基因的转录起始位点(TSS)的5'的序列。HBG1 c.-117G>A区域通过在SEQ ID NO:902中的位置2821处鸟嘌呤(G)取代为腺嘌呤(A)而示例,并且HBG2 c.-117G>A区域通过在SEQ ID NO:903(HBG2)中的位置2745处G取代为A而示例。术语"-117G>A改变"等是指在-117G>A靶区域处G取代为A。

[0163] 术语"近端HBG1/2启动子靶序列"表示包括13nt靶区域的近端HBG1/2启动子序列的50、100、200、300、400或500bp内的区域。根据本公开的基因组编辑系统的改变有利于(例如引起、促进或倾向于增加红系后代中的HbF产生的上调)。

[0164] 术语"BCL11Ae中的GATA1结合基序"是指作为BCL11A(BCL11Ae)红系特异性增强子中的GATA1结合基序(其在BCL11A基因的内含子2的+58DNA酶I超敏位点(DHS)区域中)的序列。BCL11Ae中GATA1结合基序的基因组坐标为chr2:60,495,265至60,495,270。+58DHS位点包含115个碱基对(bp)序列,如SEQ ID NO:968所示。SEQ ID NO:969中示出了+58DHS位点序列,包括约500bp上游和约200bp下游。

[0165] 在本文提供范围的情况下,括端点。此外,应理解的是,除非另外指出或从上下文和/或本领域普通技术人员的理解中显而易见,否则以范围表示的值在本发明的不同实施例中可以假定为所述范围内的任何特定值,至所述范围的下限的单位的十分之一,除非上下文另有明确规定。还应理解,除非另外指出或从上下文和/或本领域普通技术人员的理解

中显而易见，否则表示为范围的值可以假定给定范围内的任何子范围，其中所述子范围的端点表示为与所述范围的下限的单位的十分之一相同的精度程度。

概述

[0166] 本公开的各种实施例通常涉及基因组编辑系统，其被配置为将改变(例如，缺失或插入或其他突变)引入染色体DNA中，以增强HBG1和/或HBG2基因的转录，所述基因分别编码血红蛋白的Aγ和Gγ亚基。在某些实施例中，使用本文提供的方法增加的一种或多种γ-珠蛋白基因(例如，HBG1、HBG2)的表达导致相对于HbA优先形成HbF和/或增加的HbF水平(作为总血红蛋白的百分比)。在某些实施例中，本公开总体上涉及包含与Cpf1分子复合的gRNA的RNP复合物的用途。在某些实施例中，gRNA可以是未经修饰的或经修饰的，Cpf1分子可以是野生型Cpf1蛋白或经修饰的Cpf1蛋白。在某些实施例中，gRNA可包含表13、表18或表19所示的序列。在某些实施例中，经修饰的Cpf1可以由SEQ ID NO:1000、1001、1008-1018、1032、1035-39、1094-1097、1107-09(Cpf1多肽序列)或SEQ ID NO:1019-1021、1110-17(Cpf1多核苷酸序列)所示的序列编码。在某些实施例中，RNP复合物可包含表21所示的RNP复合物。例如，RNP复合物可以包括包含SEQ ID NO:1051所示的序列的gRNA和由SEQ ID NO:1097所示的序列编码的经修饰的Cpf1蛋白(RNP32，表21)。

[0167] 先前已经证明患有病症遗传持续性胎儿血红蛋白(HPFH)的患者在γ-珠蛋白调节元件中含有突变，导致胎儿γ-珠蛋白在整个生命中表达，而不是在出生时受到阻遏(Martyn 2017)。这导致胎儿血红蛋白(HbF)表达升高。HPFH突变可以是缺失的或非缺失的(例如，点突变)。患有HPFH的受试者展现出HbF的终身表达，即，他们不经历或仅经历部分珠蛋白转换，没有贫血症状。

[0168] 可以通过与天然存在的HPFH变体相关的γ-珠蛋白调节元件中的点突变来诱导HbF表达，所述变体包括例如HBG1 c.-114C>T;c.-117G>A;c.-158C>T;c.-167C>T;c.-170G>A;c.-175T>G;c.-175T>C;c.-195C>G;c.-196C>T;c.-197C>T;c.-198T>C;c.-201C>T;c.-202C>T;c.-211C>T、c.-251T>C;或c.-499T>A;或HBG2 c.-109G>T;c.-110A>C;c.-114C>A;c.-114C>T;c.-114C>G;c.-157C>T;c.-158C>T;c.-167C>T;c.-167C>A;c.-175T>C;c.-197C>T;c.-200+C;c.-202C>G;c.-211C>T;c.-228T>C;c.-255C>G;c.-309A>G;c.-369C>G;或c.-567T>G。

[0169] 在HBG1和/或HBG2基因的启动子内发现的远端CCAAT盒基序处的天然存在突变(即HBG1/2c.-111至-115)也已显示出导致持续的γ-珠蛋白表达和HPFH病症。据认为，CCAAT盒的改变(突变或缺失)可能会破坏一个或多个转录阻遏子的结合，导致γ-珠蛋白基因的持续表达和HbF表达的升高(Martyn 2017)。例如，已经显示天然存在的13个碱基对的del c.-114至-102(“13nt缺失”)与升高的HbF水平相关(Martyn 2017)。远端CCAAT盒可能与负调节转录因子(其在成人期表达并阻遏HBG)的CCAAT盒内和周围的结合基序重叠(Martyn 2017)。

[0170] 本文公开的基因编辑策略是通过破坏远端CCAAT盒中和/或远端CCAAT盒周围的一个或多个核苷酸来增加HbF表达。在某些实施例中，“CCAAT盒靶区域”可以是在远端CCAAT盒处或附近的区域，并且包括远端CCAAT盒的核苷酸和远端CCAAT盒的上游(5')25个核苷酸和下游(3')25个核苷酸(即HBG1/2c.-86至-140)。在其他实施例中，“CCAAT盒靶区域”可以是在远端CCAAT盒处或附近的区域，并且包括远端CCAAT盒的核苷酸和远端CCAAT盒的上游

(5')5个核苷酸和下游(3')5个核苷酸(即HBG1/2c.-106至-120)。本文公开了CCAAT盒靶区域的独特的非天然存在的改变,所述改变诱导HBG表达,包括但不限于HBG del c.-104至-121("18nt缺失")、HBG del c.-105至-115("11nt缺失")、HBG del c.-112至-115("4nt缺失")和HBG del c.-116("1nt缺失")。在某些实施例中,本文公开的基因组编辑系统可用于将改变引入HBG1和/或HBG2的CCAAT盒靶区域。在某些实施例中,基因组编辑系统可以包括一种或多种DNA供体模板,其编码CCAAT盒靶区域中的改变(例如缺失、插入或突变)。在某些实施例中,改变可以是非天然发生的改变或天然发生的改变。在某些实施例中,供体模板可以编码1nt缺失、4nt缺失、11nt缺失、13nt缺失、18nt缺失或c.-117G>A改变。在某些实施例中,基因组编辑系统可以包括RNA指导的核酸酶,其包括Cas9、经修饰的Cas 9、Cpf1或经修饰的Cpf1。在某些实施例中,基因组编辑系统可以包括包含gRNA和Cpf1分子的RNP。在某些实施例中,gRNA可以是未经修饰的或经修饰的,Cpf1分子可以是野生型Cpf1蛋白或经修饰的Cpf1蛋白,或其组合。在某些实施例中,gRNA可包含表13、表18或表19所示的序列。在某些实施例中,经修饰的Cpf1可以由SEQ ID NO:1000、1001、1008-1018、1032、1035-39、1094-1097、1107-09(Cpf1多肽序列)或SEQ ID NO:1019-1021、1110-17(Cpf1多核苷酸序列)所示的序列编码。在某些实施例中,RNP复合物可包含表21所示的RNP复合物。例如,RNP复合物可以包括包含SEQ ID NO:1051所示的序列的gRNA和由SEQ ID NO:1097所示的序列编码的经修饰的Cpf1蛋白(RNP32,表21)。

[0171] 还可以通过靶向破坏转录阻遏子BCL11A(其编码使HBG1和HBG2沉默的阻遏子)的红系细胞特异性表达来诱导HbF表达(Canvers 2015)。本文公开的另一种基因编辑策略是通过靶向破坏BCL11A(BCL11Ae)的红系细胞特异性增强子来增加HbF表达(也在Friedland等人("Friedland")共同转让的国际专利公开号WO 2015/148860中讨论,公开于2015年10月1日,将其通过引用以其整体并入本文)。在某些实施例中,用于靶向破坏的BCL11Ae区域可以是BCL11Ae中的GATA1结合基序。在某些实施例中,本文公开的基因组编辑系统可以用于将改变引入BCL11Ae中的GATA1结合基序中、引入CCAAT盒靶区域中、引入HBG1和/或HBG2的13nt靶区域中,或其组合。

[0172] 本公开的基因组编辑系统可以包括RNA指导的核酸酶(例如Cas9或Cpf1)以及一种或多种具有与靶区域内或附近的序列互补的靶向结构域的gRNA,以及任选地一种或多种DNA供体模板(其编码靶区域内或附近的特定突变(例如缺失或插入),和/或增强产生此类突变的效率的试剂(其包括但不限于随机寡核苷酸、参与DNA修复或DNA损伤应答的基因产物的小分子激动剂或拮抗剂、或肽试剂)。

[0173] 在本公开的实施例中可以采用多种将突变引入CCAAT盒靶区域、13nt靶区域、近端HBG1/2启动子靶序列、和/或BCL11Ae中的GATA1结合基序的方法。在一种方法中,在CCAAT盒靶区域、13nt靶区域、近端HBG1/2启动子靶序列、和/或BCL11Ae中的GATA1结合基序内进行单个改变,例如双链断裂,并以破坏区域功能的方式修复,例如通过形成indel或通过并入编码所述区域缺失的供体模板序列。在第二方法中,在所述区域的任一侧进行两个或更多个改变,导致间插序列(包括CCAAT盒靶区域、13nt靶区域和/或BCL11Ae中的GATA1结合基序)的缺失。

[0174] 通过基因疗法和/或基因组编辑对血红蛋白病的治疗因以下事实而变得复杂:因疾病在表型方面受影响的细胞,红细胞或红细胞被去核,并且不包含编码异常血红蛋白

(Hb)亚基的遗传材料也不含有上述示例性基因组编辑方法中靶向的Aγ或Gγ亚基。在本公开的某些实施例中,通过改变能够分化为红细胞或以其他方式产生红细胞的细胞来解决该复杂情况。根据本公开的各种实施例改变的红系谱系内的细胞包括但不限于造血干细胞和祖细胞(HSC)、成红细胞(包括嗜碱性、多染性和/或正染性成红细胞)、原成红细胞、多染性红细胞或网织红细胞、胚胎细胞干(ES)细胞和/或诱导性多能干(iPSC)细胞。这些细胞可以原位(例如在受试者的组织内)或离体地被改变。用于原位和离体改变细胞的基因组编辑系统的实施描述于以下标题"基因组编辑系统的实施:递送、配制和施用途径"下。

[0175] 在某些实施例中,通过使用基因组编辑系统获得导致Aγ和/或Gγ表达诱导的改变,所述基因组编辑系统包含RNA指导的核酸酶和至少一种具有靶向结构域的gRNA,所述靶向结构域与HBG1和/或HBG2的CCAAT盒靶区域内的或其近端(例如,在CCAAT盒靶区域的10、20、30、40或50、100、200、300、400或500个碱基内)的序列互补。如下面更详细地讨论的,RNA指导的核酸酶和gRNA形成复合物,所述复合物能够缔合并改变CCAAT盒靶区域或其近端区域。用于本文公开的实施例中的针对HBG1和/或HBG2的CCAAT盒靶区域或其近端的合适的gRNA和gRNA靶向结构域的实例包括但不限于SEQ ID NO:251-901、940-942、970、971、996、997、1002和1004所示的那些。

[0176] 在某些实施例中,通过使用基因组编辑系统获得导致Aγ和/或Gγ表达诱导的改变,所述基因组编辑系统包含RNA指导的核酸酶和至少一种具有靶向结构域的gRNA,所述靶向结构域与HBG1和/或HBG2的13nt靶区域内的或其近端(例如,在13nt靶区域的10、20、30、40或50、100、200、300、400或500个碱基内)的序列互补。如下面更详细地讨论的,RNA指导的核酸酶和gRNA形成复合物,所述复合物能够缔合并改变13nt靶区域或其近端区域。用于本文公开的实施例中的针对HBG1和/或HBG2的13nt靶区域或其近端的合适的gRNA和gRNA靶向结构域的实例包括但不限于SEQ ID NO:251-901、940-942、970、971、996、997、1002和1004所示的那些。

[0177] 在某些实施例中,通过使用基因组编辑系统获得导致HbF表达诱导的改变,所述基因组编辑系统包含RNA指导的核酸酶和至少一种具有靶向结构域的gRNA,所述靶向结构域与BCL11Ae中的GATA1结合基序内的或其近端(例如在BCL11Ae中的GATA1结合基序的10、20、30、40或50、100、200、300、400或500个碱基内)的序列互补。在某些实施例中,RNA针对的核酸酶和gRNA形成复合物,所述复合物能够缔合并改变BCL11Ae中的GATA1结合基序。用于本文公开的实施例中的针对BCL11Ae中的GATA1结合基序的合适的靶向结构域的实例包括但不限于SEQ ID NO:952-955所示的那些。

[0178] 基因组编辑系统可以以多种方式实现,如下面详细讨论的。例如,本公开的基因组编辑系统可以作为核糖核蛋白复合物或使用多种gRNA的多种复合物来实施。可以使用本领域已知的方法,包括电穿孔法将该核糖核蛋白复合物引入靶细胞,如共同转让的国际专利公开号WO 2016/182959(Jennifer Gori("Gori"),公开于2016年11月17日)所述,将其通过引用以其整体并入本文。

[0179] 通过本领域已知的方法将这些组合物中的核糖核蛋白复合物引入靶细胞中,所述方法包括但不限于电穿孔(例如,使用瑞士巴塞尔(Basel,Switzerland)龙沙公司商业化的Nucleofection™技术或由马里兰州盖瑟斯堡(Gaithersburg,Maryland)的Maxcyte Inc.商业化的类似技术)和脂转染(例如,使用马萨诸塞州沃尔瑟姆(Waltham Massachusetts)赛

默飞世尔科技公司(Thermo Fisher Scientific)公司商业化的Lipofectamine™试剂)。可替代地或另外地,在引入编码RNA指导的核酸酶和/或gRNA的核酸之后,在靶细胞自身内形成核糖核蛋白复合物。这些和其他递送方式在下面和Gori中进行了概述。

[0180] 根据本公开离体改变的细胞可以在递送给受试者之前被操作(例如,扩增、传代、冷冻、分化、去分化、用转基因转导等)。将细胞不同地递送至获得它们的受试者(以"自体"移植),或递送到在免疫学上不同于细胞供体的接受体(以"同种异体"移植)。

[0181] 在一些情况下,自体移植包括以下步骤:从受试者获得多个细胞,所述多个细胞在外周血中或在骨髓或其他组织(例如脾脏,皮肤等)中循环,并且对这些细胞操作以富集红系谱系中的细胞(例如,通过诱导产生iPSC,纯化表达某些细胞表面标记(例如CD34、CD90、CD49f)和/或不表达非红系谱系特征表面标记(例如CD10、CD14、CD38等)的细胞)。任选地或另外地,将细胞扩增,用转基因转导,暴露于细胞因子或其他肽或小分子试剂,和/或在用靶向CCAAT盒靶区域、13nt靶区域、近端HBG1/2启动子靶序列、和/或BCL11Ae中的GATA1结合基序的基因组编辑系统转导前冷冻/融化。基因组编辑系统可以以任何合适的形式实施或递送至细胞,包括作为核糖核蛋白复合物、作为分离的蛋白质和核酸组分、和/或作为编码基因组编辑系统的组分的核酸。

[0182] 在某些实施例中,使用本文公开的基因组编辑方法编辑的CD34+造血干细胞和祖细胞(HSPC)可用于治疗有需要的受试者中的血红蛋白病。在某些实施例中,血红蛋白病可以是严重镰状细胞病(SCD)或地中海贫血,如β-地中海贫血、δ-地中海贫血或β/δ-地中海贫血。在某些实施例中,用于治疗血红蛋白病的示例性方案可以包括从有需要的受试者中收获CD34+HSPC,使用本文公开的基因组编辑方法离体编辑自体CD34+HSPC,然后将编辑的自体CD34+HSPC再输注到受试者中。在某些实施例中,用编辑的自体CD34+HSPC治疗可导致HbF诱导增加。

[0183] 在收获CD34+HSPC之前,在某些实施例中,受试者可以中断羟基脲治疗(如果适用的话),并接受输血以维持足够的血红蛋白(Hb)水平。在某些实施例中,受试者可经静脉内施用普乐沙福(plerixafor)(例如,0.24mg/kg)以将CD34+HSPC从骨髓动员至外周血。在某些实施例中,受试者可经历一个或多个白细胞去除周期(例如,周期之间约一个月,一个周期定义为连续几天进行两次普乐沙福动员的白细胞去除收集)。在某些实施例中,为受试者进行的白细胞去除周期的数量可以是实现编辑的自体CD34+HSPC重新注入回受试者所需的剂量(例如,≥2x 10^6细胞/kg,≥3x 10^6细胞/kg,≥4x 10^6细胞/kg,≥5x 10^6细胞/kg,2x 10^6细胞/kg至3x 10^6细胞/kg,3x 10^6细胞/kg至4x 10^6细胞/kg,4x 10^6细胞/kg至5x 10^6细胞/kg)连同实现用于备份存储的未编辑的自体CD34+HSPC/kg的剂量(例如,≥1.5x 10^6细胞/kg)的数量。在某些实施例中,可以使用本文公开的任何基因组编辑方法编辑从受试者收获的CD34+HSPC。在某些实施例中,本文公开的任何一种或多种gRNA和一种或多种RNA指导的核酸酶可以在基因组编辑方法中使用。

[0184] 在某些实施例中,治疗可以包括自体干细胞移植。在某些实施例中,受试者可以用白消安调理进行清髓性调理(例如,基于第一剂量药物代谢动力学分析调整剂量,测试剂量为1mg/kg)。在某些实施例中,调理可以连续发生四天。在某些实施例中,三天白消安洗脱期后,可以将编辑的自体CD34+HSPC(例如,≥2x 10^6细胞/kg,≥3x 10^6细胞/kg,≥4x 10^6细胞/kg,≥5x 10^6细胞/kg,2x 10^6细胞/kg至3x 10^6细胞/kg,3x 10^6细胞/kg至4x 10^6细胞/

kg,4x 10^6细胞/kg至5x 10^6细胞/kg)重新注入受试者(例如重新注入外周血)。在某些实施例中,可以制造编辑的自体CD34+HSPC并为特定受试者冷冻保存。在某些实施例中,受试者可以在连续的清髓性调理方案和经编辑的自体CD34+细胞输注后获得中性粒细胞移植。中性粒细胞移植可以定义为三次连续测量的ANC≥0.5x 10^9/L。

[0185] 无论如何实施,基因组编辑系统可包括以下或可与以下一起共递送:一种或多种在编辑过程中和编辑后改善细胞生存力的因子,包括但不限于芳烃受体拮抗剂,例如StemRegenin-1(SR1)、UM171,LGC0006、α-萘并黄酮和CH-223191,和/或先天性免疫应答拮抗剂,例如环孢菌素A、地塞米松、白藜芦醇(reservatrol)、MyD88抑制肽、靶向Myd88的RNAi剂、B18R重组蛋白、糖皮质激素、OxPAPC、TLR拮抗剂、雷帕霉素、BX795和RLR shRNA。可在编辑过程中和编辑后改善细胞生存力的这些和其他因子描述于Gori中标题"I.干细胞的优化"(第36页至第61页)下,将其通过引用并入本文。

[0186] 递送基因组编辑系统后,任选地对细胞进行操作,例如以富集红系谱系中的HSC和/或细胞和/或富集经编辑的细胞,以使其扩增、冷冻/融化,或以其他方式使细胞返回受试者。然后将经编辑的细胞返回给受试者,例如通过静脉内递送在循环系统中或递送至实体组织例如骨髓中。

[0187] 在功能上,使用本公开的组合物、方法和基因组编辑系统改变CCAAT盒靶区域,13nt靶区域,近端HBG1/2启动子靶序列和/或BCL11Ae中的GATA1结合基序导致在表达血红蛋白的细胞中显著诱导Aγ和/或Gγ亚基(可互换地称为HbF表达),例如相对于未经修饰的对照,至少5%、10%、15%、20%、25%、30%、35%、40%、45%、50%或更高的Aγ和/或Gγ亚基表达诱导。这种蛋白表达的诱导通常是经处理的多个细胞中的一些或全部中的CCAAT盒靶区域、13nt靶区域、近端HBG1/2启动子靶序列、和/或BCL11Ae中GATA1结合基序发生改变的结果(表达,例如关于多个细胞中包含indel突变的总基因组的百分比),例如所述多个细胞中至少5%、10%、15%、20%、25%、30%、35%、40%、45%、50%包含至少一个等位基因,所述等位基因包含在CCAAT盒靶区域、13nt靶区域、近端HBG1/2启动子靶序列、和/或BCL11Ae中GATA1结合基序内或附近的序列改变,所述序列改变包括但不限于indel、插入或缺失。

[0188] 可以以许多合适的方式评估由本公开的基因组编辑系统和方法引起或促进的改变的功能作用。例如,可以在蛋白质或mRNA水平上评估改变对胎儿血红蛋白表达的作用。HBG1和HBG2 mRNA的表达可通过数字液滴PCR(ddPCR)进行评估,ddPCR可对通过从经处理的或未经处理的样品收获的mRNA的逆转录而获得的cDNA样品进行。HBG1、HBG2、HBB和/或HBA的引物可使用本领域已知的方法单独使用或多重使用。例如,可以使用由Bio Rad(海格立斯(Hercules),加利福尼亚)商业化的QX200™ddPCR系统以及由BioRad公开的相关方案进行样品的ddPCR分析。胎儿血红蛋白可以通过以下进行评估:高压液相色谱(HPLC)(例如,根据Chang 2017第143-44页上讨论的方法(通过引用并入本文)),或快速蛋白液相色谱(FPLC)(如本领域已知的,使用离子交换和/或反相柱来解析HbF、HbB和HbA和/或Aγ和Gγ珠蛋白链)。

[0189] 应当指出,CCAAT盒靶区域(例如18nt、11nt、4nt、1nt、c.-117G>A靶区域)、13nt靶区域、近端HBG1/2启动子靶序列和/或BCL11Ae中的GATA1结合基序在靶细胞中被改变的比率可以通过使用任选的基因组编辑系统组分(例如寡核苷酸供体模板)进行修饰。供体模板

设计在下面的“供体模板设计”标题下进行了概述。用于靶向13nt靶区域的供体模板可以包括但不限于编码HBG1 c.-114至-102(对应于SEQ ID NO:902的核苷酸2824-2836)、HBG1 c.-225至-222(对应于SEQ ID NO:902的核苷酸2716-2719)和/或HBG2 c.-114至-102(对应于SEQ ID NO:903的核苷酸2748-2760)的改变(例如缺失)的供体模板。示例性5’和3’同源臂,以及编码缺失(例如c.-114至-102)的示例性全长供体模板也在下面给出(SEQ ID NO:904-909)。在某些实施例中,用于靶向18nt靶区域的供体模板可以包括但不限于编码HBG1c.-104至-121、HBG2 c.-104至-121、或其组合的改变(例如缺失)的供体模板。编码缺失(例如c.-104至-121)的示例性全长供体模板包括SEQ ID NO:974和975。在某些实施例中,用于靶向11nt靶区域的供体模板可以包括但不限于编码HBG1 c.-105至-115、HBG2 c.-105至-115、或其组合的改变(例如缺失)的供体模板。编码缺失(例如c.-105至-115)的示例性全长供体模板包括SEQ ID NO:976和978。在某些实施例中,用于靶向4nt靶区域的供体模板可以包括但不限于编码HBG1 c.-112至-115、HBG2 c.-112至-115、或其组合的改变(例如缺失)的供体模板。编码缺失(例如c.-112至-115)的示例性全长供体模板包括SEQ ID NO:984-995。在某些实施例中,用于靶向1nt靶区域的供体模板可以包括但不限于编码HBG1 c.-116、HBG2 c.-116、或其组合的改变(例如缺失)的供体模板。编码缺失(例如c.-116)的示例性全长供体模板包括SEQ ID NO:982和983。在某些实施例中,用于靶向c.-117G>A靶区域的供体模板可以包括但不限于编码HBG1 c.-117G>A、HBG2 c.-117G>A、或其组合的改变(例如缺失)的供体模板。编码缺失(例如c.-117G>A)的示例性全长供体模板包括SEQ ID NO:980和981。在某些实施例中,供体模板可以是正链或负链。

[0190] 本文使用的供体模板可以是与靶序列内或附近的DNA区域不同源的非特异性模板。在某些实施例中,用于靶向13nt靶区域的供体模板可包括但不限于与13nt靶区域内或附近的DNA区域不同源的非靶特异性模板。例如,用于靶向13nt靶区域的非特异性供体模板可以与13nt靶区域内或附近的DNA区域不同源,并且可包含编码HBG1 c.-225至-222(对应于SEQ ID NO:902的核苷酸2716-2719)的缺失的供体模板。在某些实施例中,用于靶向BCL11Ae中的GATA1结合基序的供体模板可以包括但不限于与BCL11Ae靶序列中的GATA1结合基序内或附近的DNA区域不同源的非靶特异性模板。用于靶向BCL11Ae的其他供体模板可以包括但不限于包括BCL11Ae(包括但不限于BCL11Ae中的GATA1基序)的改变(例如,缺失)的供体模板。

[0191] 本文描述的实施例可用于所有类别的脊椎动物,包括但不限于灵长类动物、小鼠、大鼠、兔、猪、狗和猫。

[0192] 该概述集中在少数示例性实施例上,这些实施例说明了基因组编辑系统和CRISPR介导的细胞改变方法的原理。然而,为清楚起见,本公开包含以上未明确解决的但对于本领域技术人员而言显而易见的修改和变化。考虑到这一点,以下公开旨在更概括地说明基因组编辑系统的操作原理。以下内容不应理解为限制性的,而应是基因组编辑系统和利用这些系统的CRISPR介导方法的某些原理的说明,这与本公开相结合,将在其范围内为本领域技术人员提供有关另外的实施和修改的信息。

RNA指导的解旋酶、指导RNA和失活的指导RNA

[0193] 本公开的各个实施例还通常涉及基因组编辑系统(其被配置为改变核酸的螺旋结构以增强核酸中的靶区域(例如,CCAAT盒靶区域、13nt靶区域、近端HBG1/2启动子靶序列

和/或BCL11Ae中的GATA1结合基序)的基因组编辑)、及其方法和组合物。许多实施例涉及以下观察结果:将改变DNA的螺旋结构的事件定位在核酸中的靶区域内或附近可以改善针对此类靶区域的基因组编辑系统的活性。不希望受到任何理论的束缚,认为DNA靶区域内或近端的螺旋结构的改变(例如,通过解旋)可以诱导或增加基因组编辑系统对靶区域的可及性,从而导致基因组编辑系统对靶区域的增加的编辑。

[0194] CRISPR核酸酶主要是为了保护细菌抵抗病毒病原体而进化,病毒病原体的基因组不天然地组织成染色质。相比之下,当真核生物基因组被组织成核小体单位时,其包含盘绕组蛋白的基因组DNA区段。已发现来自多个细菌家族的CRISPR核酸酶对真核DNA的编辑无活性,这表明编辑与核小体结合的DNA的能力可能因酶而异(Ran 2015)。生化证据表明,化脓链球菌Cas9可以在核小体边缘有效切割DNA,但是当靶位点位于核小体二分体中心附近时,其活性降低(Hinz 2016)。

[0195] 在许多细胞类型中,目的靶位点可与核小体牢固结合,或者对于在核小体存在下无法有效编辑的酶而言可以仅具有相邻的PAM。在这种情况下,有问题的核小体可以首先通过使用更靠近核小体边缘的相邻靶位点来置换或被更有效地结合核小体DNA的酶结合。然而,在这些相邻位点的切割可能不利于治疗策略。因此,具有结合这些相邻位点但不切割的可编程酶可以实现更有效的功能编辑。

[0196] 相关策略利用外源反式作用因子的募集以促进核小体置换。然而,本公开的系统和方法优于该策略,因为它们不需要靶向结构域截短之外的gRNA修饰,不需要募集外源反式作用因子,并且也不需要转录激活来实现更高的编辑率。

[0197] 在本公开的各种实施例中采用了多种用于解旋和改变核酸的方法。一种方法包括解旋(或打开)细胞中核酸的靶区域(例如CCAAT盒靶区域、13nt靶区域、近端HBG1/2启动子靶序列、和/或BCL11Ae中的GATA1结合基序)内或近端的染色质区段,并且在核酸的靶区域内产生双链断裂(DSB),从而改变靶区域。在某些实施例中,可以以改变靶区域的方式修复DSB。使用本文提供的方法解旋染色质区段可以促进增加催化活性RNP(例如,催化活性RNA指导的核酸酶和gRNA)进入染色质,以允许DNA的更有效编辑。例如,这些方法可用于编辑染色质中难以被核糖核蛋白(例如,与gRNA复合的RNA指导的核酸酶)进入(因为染色质被核小体占据,例如封闭的染色质)的靶区域。在某些实施例中,染色质区段的解旋是通过RNA指导的解旋酶活性发生的。在某些实施例中,解旋步骤不需要将外源反式作用因子募集至染色质区段。在某些实施例中,解旋染色质区段的步骤不包括在染色质区段内的核酸中形成单链或双链断裂。

[0198] 在上述办法和方法的某些实施例中,DNA螺旋结构的改变是通过“RNA指导的解旋酶”的作用实现的,术语“RNA指导的解旋酶”通常用于指分子,典型地是肽,其(a)与gRNA相互作用(例如,复合),并且(b)与gRNA一起,缔合并解旋靶位点。在某些实施例中,RNA指导的解旋酶可以包含被配置为缺乏核酸酶活性的RNA指导的核酸酶。然而,发明人已经观察到,甚至具有切割能力的RNA指导的核酸酶也可以通过将其与具有长度为15个或更少的核苷酸的截短的靶向结构域的失活gRNA复合而适于用作RNA指导的解旋酶。野生型RNA指导的核酸酶与失活gRNA的复合物显示减少的或消除的RNA切割活性,但似乎保留了解旋酶活性。RNA指导的解旋酶和失活gRNA的详细描述如下。

[0199] 关于RNA指导的解旋酶,根据本公开,RNA指导的解旋酶可包含本文公开的和下文

“RNA指导的核酸酶”标题下的任何RNA指导的核酸酶，包括但不限于Cas9或Cpf1 RNA指导的核酸酶。这些RNA指导的核酸酶的解旋酶活性允许解旋DNA，从而增加基因组编辑系统组分(例如但不限于催化活性RNA指导的核酸酶和gRNA)进入所期望的有待编辑的靶区域(例如CCAAT盒靶区域、13nt靶区域、近端HBG1/2启动子靶序列和/或BCL11Ae中的GATA1结合基序)。在某些实施例中，RNA指导的核酸酶可以是具有核酸酶活性的催化活性RNA指导的核酸酶。在某些实施例中，RNA指导的解旋酶可以被配置为缺乏核酸酶活性。例如，在某些实施例中，RNA指导的解旋酶可以是缺乏核酸酶活性的催化失活性RNA指导的核酸酶，例如催化失活性Cas9分子，其仍然提供解旋酶活性。在某些实施例中，RNA指导的解旋酶可以与失活gRNA形成复合物，从而形成不能切割核酸的失活RNP。在其他实施例中，RNA指导的解旋酶可以是与失活gRNA复合的催化活性RNA指导的核酸酶，形成不能切割核酸的失活RNP。在某些实施例中，RNA指导的核酸酶未配置为将外源反式作用因子募集至所期望的有待的编辑的靶区域(例如，CCAAT盒靶区域、13nt靶区域，近端HBG1/2启动子靶标序列、和/或BCL11Ae中的GATA1结合基序)。

[0200] 关于失活gRNA，这些包括本文讨论的以及下文“失活gRNA分子”标题下的任何失活gRNA。可以通过以下来产生失活gRNA(在本文中也称为“dgRNA”)：截短gRNA靶向结构域序列的5’末端，从而导致长度为15个核苷酸或更短的靶向结构域序列。在某些实施例中，可以通过截短本文表2或表10中公开的gRNA靶向结构域序列中任一个的5’末端来产生dgRNA。根据本公开的失活指导RNA分子包括具有降低的、低的或不可检测的切割活性的失活指导RNA分子。与活性指导RNA的靶向结构域序列相比，失活指导RNA的靶向结构域序列的长度可以短1、2、3、4、5、6、7、8、9或10个核苷酸。失活gRNA分子可包含与靶核酸中的靶区域(例如，CCAAT盒靶区域、13nt靶区域、近端HBG1/2启动子靶序列和/或BCL11Ae中的GATA1结合基序)近端或内的区域互补的靶向结构域。在某些实施例中，“近端”可表示靶区域(例如，CCAAT盒靶区域、13nt靶区域、近端HBG1/2启动子靶序列、和/或BCL11Ae中的GATA1结合基序)的10、25、50、100或200个核苷酸内的区域。在某些实施例中，失活gRNA包含与DNA的转录链或非转录链互补的靶向结构域。在某些实施例中，失活指导RNA未配置为将外源反式作用因子募集至靶区域(例如，CCAAT盒靶区域、13nt靶区域，近端HBG1/2启动子靶标序列、和/或BCL11Ae中的GATA1结合基序)。

[0201] 本文还提供了通过以下来增加靶核酸中indel形成率的方法：使用RNA指导的解旋酶解旋靶区域(例如，CCAAT盒靶区域，13nt靶区域，近端HBG1/2启动子靶序列和/或BCL11Ae中的GATA1结合基序)内或近端的DNA，在靶区域内生成DSB，并通过修复DSB在靶区域中形成indel。与不使用解旋酶的形成indel的方法相比，使用RNA指导的解旋酶解旋DNA的步骤提供了增加的indel形成。

[0202] 本公开还包括缺失细胞中靶核酸区段的方法，所述方法包括使细胞与RNA指导的解旋酶接触并在靶区域(例如，CCAAT盒靶区域、13nt靶区域、近端HBG1/2启动子靶序列和/或BCL11Ae中的GATA1结合基序)内产生双链断裂(DSB)。在某些实施例中，RNA指导的解旋酶被配置为在靶核酸的靶区域内或近端缔合并解旋靶区域内或近端的双链DNA(dsDNA)。在某些实施例中，靶核酸是基因的启动子区、基因的编码区、基因的非编码区、基因的内含子、或基因的外显子。在某些实施例中，有待缺失的靶核酸的区段可以是至少约15、20、25、30、35、40、45、50、55、60、65、70、75或100个碱基对长度。在某些实施例中，以缺失靶核酸的区段的

方式修复DSB。

[0203] 如下面详细讨论的,可以以多种方式来实施被配置为引入螺旋结构的改变的基因组编辑系统。例如,本公开的基因组编辑系统可以作为核糖核蛋白复合物或使用多种gRNA的多种复合物来实施。在某些实施例中,基因组编辑系统的核糖核蛋白复合物可以是与失活指导RNA复合的RNA指导的解旋酶。可以使用本领域已知的方法,包括电穿孔法,将核糖核蛋白复合物引入靶细胞,如Gori所述。掺入了RNA指导的解旋酶的基因组编辑系统也可以以任何合适的方式进行修饰,包括但不限于在靶区域中或附近包含一个或多个编码特定突变(例如缺失或插入)的DNA供体模板,和/或增强产生此类突变的效率的试剂(其包括但不限于随机寡核苷酸、参与DNA修复或DNA损伤应答的基因产物的小分子激动剂或拮抗剂、或肽试剂)。这些修饰将在下面的“基因组编辑策略”标题下进行详细说明。为了清楚起见,本公开包括包含一种或多种gRNA、失活gRNA、RNA指导的解旋酶、RNA指导的核酸酶或其组合的组合物。

[0204] 尽管以上几个示例性实施例集中于DNA解旋,但应注意,其他螺旋改变也在本公开的范围内。这些包括但不限于,在靶区域内或近端的DNA链上的过度旋绕、旋绕不足、扭转应变的增加或减少(例如,通过拓扑异构酶活性)、变性或链分离和/或导致染色质结构修饰的其他合适的改变。这些改变中的每一个都可以由RNA指导的活性或内源性因子募集到靶区域来催化。

[0205] 本文还提供了基因组编辑系统和改变由活性指导物产生的一个或多个indel(例如,indel签名)的方法。如发明人在本文中发现的,失活RNP(dRNP)(即与RNA指导的核酸酶复合的失活指导RNA)和活性RNP(即与RNA指导的核酸酶复合的活性指导RNA)的配对可导致仅由活性RNP(无dRNP)产生的indel(例如indel签名)方向性改变。如以下实例所示,使用失活指导RNA可导致从活性指导RNA切割位点向失活指导RNA结合位点延伸的较大缺失的频率增加。因此,失活指导RNA可以用于有效地将缺失编辑“定向”朝向期望的靶位点。在某些实施例中,将失活指导RNA与活性指导RNA一起使用可以增加与微同源性不相关的缺失的频率。

[0206] 尽管下文实例部分中公开的实例针对CCAAT盒靶区域的改变,但是熟练的技术人员会想到,本文所述的基因组编辑系统、方法、细胞和组合物可用于改变任何其他靶区域,例如,非限制性地,增加在靶区域处的缺失的频率,增加在靶区域处与微同源性不相关的缺失的频率(例如,不通过MMEJ修复)。

[0207] 该概述集中在少数示例性实施例上,这些实施例说明了基因组编辑系统和CRISPR介导的细胞改变方法的原理。然而,为清楚起见,本公开包含以上未明确解决的但对于本领域技术人员而言显而易见的修改和变化。考虑到这一点,以下公开旨在更概括地说明基因组编辑系统的操作原理。以下内容不应理解为限制性的,而应是基因组编辑系统和利用这些系统的CRISPR介导方法的某些原理的说明,这与本公开相结合,将在其范围内为本领域技术人员提供有关另外的实施和修改的信息。

基因组编辑系统

[0208] 术语“基因组编辑系统”是指具有RNA指导的DNA编辑活性的任何系统。本公开的基因组编辑系统包括至少两种从天然存在的CRISPR系统改适的组分:指导RNA(gRNA)和RNA指导的核酸酶。这两种组分形成复合物,所述复合物能够与特定核酸序列结合并在所述核酸

序列中或其周围编辑DNA，例如通过制备一条或多条单链断裂(SSB或切口)、双链断裂(DSB)和/或点突变。

[0209] 在某些实施例中，本公开中的基因组编辑系统可包括用于解旋DNA的解旋酶。在某些实施例中，解旋酶可以是RNA指导的解旋酶。在某些实施例中，RNA指导的解旋酶可以是如本文所述的RNA指导的核酸酶，例如Cas9或Cpf1分子。在某些实施例中，RNA指导的核酸酶未配置为将外源反式作用因子募集到靶区域。在某些实施例中，RNA指导的核酸酶可以被配置为缺乏核酸酶活性。在某些实施例中，RNA指导的解旋酶可以与本文公开的失活指导RNA复合。例如，失活指导RNA(dgRNA)可包含长度少于15个核苷酸的靶向结构域序列。在某些实施例中，失活指导RNA未配置为将外源反式作用因子募集至靶区域。

[0210] 天然存在的CRISPR系统进化性地组织化为两个类别和五种类型(Makarova 2011，通过引用并入本文)，并且虽然本公开的基因组编辑系统可改适任一类型或类别的天然存在的CRISPR系统的组分，但本文所呈现的实施例通常是从2类和II型或V型CRISPR系统改适。2类系统涵盖II型和V型，其特征为相对较大的多结构域RNA指导的核酸酶蛋白(例如，Cas9或Cpf1)以及一个或多个指导RNA(例如，crRNA和任选地tracrRNA)，它们形成核糖核蛋白(RNP)复合物，该复合物缔合(即，靶向)并切割与crRNA的靶向(或间隔序列)序列互补的特定基因座。根据本公开的基因组编辑系统类似地靶向并编辑细胞DNA序列，但与自然界中存在的CRISPR系统显著不同。例如，本文所述的单分子指导RNA在自然界中不存在，并且根据本公开的指导RNA和RNA指导的核酸酶二者可并入任一数目的非天然存在的修饰。

[0211] 基因组编辑系统可以用多种方式来实施(例如施用于或递送至细胞或受试者)，并且不同的实施可适合于不同应用。例如，在某些实施例中，基因组编辑系统作为蛋白质/RNA复合物(核糖核蛋白或RNP)进行实施，所述复合物可以被包含在药物组合物中，所述药物组合物任选地包括药学上可接受的载剂和/或包囊剂，例如但不限于脂质或聚合物的微颗粒或纳米颗粒、胶束或脂质体。在某些实施例中，基因组编辑系统作为编码上述RNA指导的核酸酶和指导RNA组分的一种或多种核酸(任选地具有一种或多种另外的组分)进行实施；在某些实施例中，基因组编辑系统作为包含此类核酸的一种或多种载体(例如病毒载体，例如腺相关病毒(参见下文“基因组编辑系统的实施：递送、配制和施用途径”标题下的部分))进行实施；并且在某些实施例中，基因组编辑系统是作为前述任一种的组合来实施。根据本文所述原理操作的其他或经修改的实施将为技术人员所了解并且在本公开的范围内。

[0212] 应注意，本公开的基因组编辑系统可靶向单一特定核苷酸序列，或者可通过使用两个或更多个指导RNA靶向(并且能平行编辑)两个或更多个特定核苷酸序列。在本公开通篇中，多种gRNA的使用称为“多重化”，并且可用于靶向多个无关的所关注靶序列，或用于在单一靶结构域内形成多个SSB或DSB，并且在一些情形中，用于在这种靶结构域内产生特定编辑。例如，Maeder等人的国际专利公开号WO 2015/138510(“Maeder”)(将其通过引用并入本文)描述用于修正人CEP290基因中的点突变(C.2991+1655A至G)的基因组编辑系统，所述点突变导致产生隐蔽剪接位点，这又降低或消除该基因的功能。Maeder的基因组编辑系统利用两个指导RNA，这些指导RNA靶向该点突变任一侧上的(即，侧接)序列，并且形成侧接该突变的DSB。这又促进了包括突变在内的间插序列的缺失，由此消除隐蔽剪接位点并恢复正常的基因功能。

[0213] 作为另一实例，Cotta-Ramusino等人的WO 2016/073990(“Cotta-Ramusino”)(将

其通过引用并入本文)描述利用两个gRNA与Cas9切口酶(制造单链切口的Cas9,例如化脓链球菌D10A)的基因组编辑系统,该布置称为“双重切口酶系统”。Cotta-Ramusino的双重切口酶系统经配置以在所关注序列的相对链上制造两个偏移一个或多个核苷酸的切口,所述切口组合产生具有悬突(在Cotta-Ramusino情形中是5’悬突,但3’悬突也有可能)的双链断裂。在一些情况下,该悬突又可以促进同源定向修复事件。并且作为另一实例,Palestrant等人的WO2015/070083(通过引用并入本文)描述靶向编码Cas9的核苷酸序列的gRNA(称为“管理RNA”),其可以包括于基因组编辑系统中,所述基因组编辑系统包含一个或多个另外的gRNA以容许Cas9的瞬时表达,所述Cas9可能原本例如在一些经病毒转导的细胞中是组成型表达的。这些多重化应用旨在具有示例性而不是限制性,并且技术人员将了解,其他多重化应用通常与本文所述的基因组编辑系统相容。

[0214] 如本文所公开,在某些实施例中,基因组编辑系统可包含多种gRNA,其可用于将突变引入BCL11Ae中的GATA1结合基序或HBG1和/或HBG2的13nt靶区域。在某些实施例中,本文公开的基因组编辑系统可包含用于将突变引入BCL11Ae中的GATA1结合基序和HBG1和/或HBG2的13nt靶区域的多种gRNA。

[0215] 在一些情况下,基因组编辑系统可以形成双链断裂,这些双链断裂是通过细胞DNA双链断裂机制例如NHEJ或HDR来修复。这些机制描述于文献中(参见例如Davis&Maizels 2014(描述Alt-HDR);Frit 2014(描述Alt-NHEJ);Iyama&Wilson2013(一般性描述经典HDR和NHEJ途径))。

[0216] 如果基因组编辑系统通过形成DSB来操作,那么此类系统任选地包括促进或有助于特定双链断裂修复模式或特定修复结果的一个或多个组分。例如,Cotta-Ramusino还描述其中添加单链寡核苷酸“供体模板”的基因组编辑系统;将供体模板并入细胞DNA的靶区域中,该靶区域由基因组编辑系统切割,并且可以导致靶序列中的变化。

[0217] 在某些实施例中,基因组编辑系统在不引起单链或双链断裂的情况下修饰靶序列,或修饰靶序列中或附近的基因的表达。例如,基因组编辑系统可包括融合至作用于DNA的功能结构域的RNA指导的核酸酶,由此修饰靶序列或其表达。作为一个实例,RNA指导的核酸酶可连接至(例如融合至)胞苷脱氨酶功能结构域,并且可通过产生所靶向的C至A取代来操作。示例性核酸酶/脱氨酶融合体描述于Komor 2016中,将其通过引用并入本文。可替代地,基因组编辑系统可利用切割失活的(即,“死”)核酸酶,例如失活Cas9(dCas9),并且可通过在细胞DNA的一个或多个所靶向区域上形成稳定复合物来操作,由此干扰涉及一个或多个所靶向区域的功能,包括但不限于mRNA转录、染色质重塑等。在某些实施例中,基因组编辑系统可以包括RNA指导的解旋酶,其解旋靶序列内或近端的DNA,而不会引起单链或双链断裂。例如,基因组编辑系统可以包括RNA指导的解旋酶,其被配置为在靶序列内或附近缔合以解旋DNA并诱导对靶序列的可及性。在某些实施例中,RNA指导的解旋酶可以与被配置为缺乏切割活性的失活指导RNA复合,从而允许DNA解旋而不引起DNA断裂。

指导RNA(gRNA)分子

[0218] 术语“指导RNA”和“gRNA”是指促进RNA指导的核酸酶(例如Cas9或Cpf1)与细胞中的靶序列(例如基因组或附加体序列)的特异性缔合(或“靶向”)的任何核酸。gRNA可以是单分子(包含单一RNA分子,并且可替代地称为嵌合)或模块(包含多于一个、并且典型地两个单独的RNA分子,例如crRNA和tracrRNA,其通常例如通过双链化彼此缔合)。在整个文献中

描述了gRNA及其组分(例如,参见Briner 2014,将其通过引用并入;Cotta-Ramusino)。可以根据本文的实施例使用的模块性和单分子gRNA的实例包括但不限于SEQ ID NO:29-31和38-51所示的序列。可以根据本文的实施例使用的gRNA近端和尾结构域的实例包括但不限于SEQ ID NO:32-37所示的序列。

[0219] 在细菌和古细菌中,II型CRISPR系统通常包含RNA指导的核酸酶蛋白(例如Cas9)、包括与外来序列互补的5'区的CRISPR RNA(crRNA)和包括与crRNA的3'区互补并形成双链体的5'区的反式激活crRNA(tracrRNA)。虽然并非旨在受限于任何理论,但认为此双链体有助于形成Cas9/gRNA复合物,并且是所述复合物的活性所需的。在II型CRISPR系统经改适用于基因编辑中时,发现crRNA和tracrRNA可以接合成单一单分子或嵌合指导RNA,在一个非限制性实例中借助桥接crRNA(在其3'末端)和tracrRNA(在其5'末端)的互补区的四核苷酸(例如GAAA)"四环(tetraloop)"或"接头"序列来接合。(Mali 2013;Jiang 2013;Jinek 2012;全部通过引用并入本文)。

[0220] 指导RNA不论是单分子或模块都包括"靶向结构域",所述靶向结构域与靶序列内的靶结构域完全或部分互补,所述靶序列例如期望编辑的细胞基因组中的DNA序列。靶向结构域在文献中通过多种名称来提及,包括但不限于"指导序列"(Hsu 2013,通过引用并入本文)、"互补性区域"(Cotta-Ramusino)、"间隔子"(Briner 2014)和一般性地称为"crRNA"(Jiang)。不论给予其何种名称,靶向结构域典型地长度为10-30个核苷酸,并且在某些实施例中长度为16-24个核苷酸(例如,长度为16、17、18、19、20、21、22、23或24个核苷酸),并且在Cas9 gRNA情形中位于5'末端处或5'末端附近,并且在Cpf1 gRNA情形中位于3'末端处或3'末端附近。

[0221] 除了靶向结构域以外,gRNA典型地(但不一定,如下文所讨论)包括多个可影响gRNA/Cas9复合物的形成或活性的域。例如,如上文所提及,通过gRNA的第一和第二互补性结构域形成的双链化结构(也称为重复:抗重复双链体)与Cas9的识别(REC)叶相互作用,并且可以介导Cas9/gRNA复合物的形成(Nishimasu 2014;Nishimasu 2015;两者均通过引用并入本文)。应注意,第一和/或第二互补性结构域可含有一个或多个多聚腺苷酸段,其可以由RNA聚合酶识别为终止信号。因此,第一和第二互补性结构域的序列任选地经修饰以消除这些段,并促进gRNA的体外转录的完成,例如通过使用如Briner 2014中所述的A-G交换或A-U交换来修饰。对第一和第二互补性结构域的这些和其他类似修饰在本公开的范围内。

[0222] 与第一和第二互补性结构域一起,Cas9 gRNA典型地包括两个或更多个其他双链化区域,该两个或更多个另外的双链化区域在体内但不一定在体外参与核酸酶活性。(Nishimasu 2015)。在第二互补性结构域的3'部分附近的第一茎环1被不同地称为"近端结构域"(Cotta-Ramusino)、"茎环1"(Nishimasu 2014和2015)以及"连结(nexus)"(Briner 2014)。一个或多个其他茎环结构通常存在于gRNA的3'末端附近,其数目依物种而变:化脓链球菌gRNA典型地包括2个3'茎环(总计4个茎环结构,包括重复:抗重复双链体),而金黄色葡萄球菌和其他物种仅具有一个(总计3个茎环结构)。对根据物种组织化的保守茎环结构(更通常地,和gRNA结构)的描述于Briner 2014中提供。

[0223] 虽然前述说明集中于用于Cas9的gRNA,但应了解,存在其他RNA指导的核酸酶,其利用在一些方面与针对这一点描述的那些gRNA不同的gRNA。例如,Cpf1("来自普雷沃菌属和弗朗西斯菌属1的CRISPR")是最近发现的RNA指导的核酸酶,其发挥功能不需要

tracrRNA。(Zetsche 2015,通过引用并入本文)。用于Cpf1基因组编辑系统的gRNA通常包括靶向结构域和互补性结构域(替代性地称为“把手”)。还应注意,在用于Cpf1的gRNA中,靶向结构域通常存在于3’末端处或附近,而不是如上文结合Cas9 gRNA所述的5’末端(把手位于Cpf1 gRNA的5’末端处或5’末端附近)。Cpf1 gRNA的示例性靶向结构域在表13和表18中示出。

[0224] 但本领域技术人员将了解,虽然在来自不同原核物种的gRNA之间或在Cpf1与Cas9 gRNA之间可能存在结构差异,但gRNA的操作原理通常是一致的。因为这种操作一致性,gRNA可在广义上通过其靶向结构域序列来定义,并且技术人员将了解,可将给定靶向结构域序列并入任何适宜gRNA中,包括单分子或嵌合gRNA,或包括一种或多种化学修饰和/或序列修饰(取代、另外的核苷酸、截短等)的gRNA。因此,为了便于呈现本公开,gRNA可仅在其靶向结构域序列方面加以描述。

[0225] 更通常地,技术人员将了解,本公开的一些方面涉及可以使用多种RNA指导的核酸酶实施的系统、方法和组合物。为此,除非另外指定,否则术语gRNA应理解为不仅涵盖那些与Cas9或Cpf1的特定种类相容的gRNA,还涵盖可以用于任何RNA指导的核酸酶的任何适宜gRNA。通过说明,在某些实施例中,术语gRNA可以包括与存在于2类CRISPR系统(例如II型或V型或CRISPR系统)中的任何RNA指导的核酸酶或从其衍生或改适的RNA指导的核酸酶一起使用的gRNA。

gRNA设计

[0226] 先前已经描述了用于选择和验证靶序列以及脱靶分析的方法(参见,例如,Mali 2013;Hsu 2013;Fu 2014;Heigwer 2014;Bae 2014;Xiao 2014)。这些参考文献中的每一篇都是通过引用并入本文。作为非限制性实例,gRNA设计可包括使用软件工具来优化对应于使用者的靶序列的潜在靶序列的选择,例如以使全基因组的总脱靶活性降至最低。虽然脱靶活性不限于切割,但每个脱靶序列处的切割效率可以例如使用实验衍生的加权方案来预测。这些和其他指导选择方法详细描述于Maeder和Cotta-Ramusino中。

[0227] 关于针对HBG1/2靶位点(例如13nt靶区域)的gRNA靶向结构域序列的选择,使用了计算机gRNA靶结构域鉴定工具,并将命中分为四个等级。对于化脓链球菌,基于(1)从靶位点任一末端上游或下游的距离(即,HBG1/2 13nt靶区域),特异地是在靶位点任一末端的400bp内,(2)高水平的正交性,和(3)5’G的存在来选择等级1靶向结构域。基于(1)从靶位点任一末端上游或下游的距离(即,HBG1/213nt靶区域),特异地是在靶位点任一末端的400bp内,和(2)高水平的正交性来选择等级2靶向结构域。基于(1)从靶位点任一末端上游或下游的距离(即,HBG1/2 13nt靶区域),特异地是在靶位点任一末端的400bp内,和(2)5’G的存在来选择等级3靶向结构域。基于从靶位点任一末端上游或下游的距离(即,HBG1/2 13nt靶区域),特异地是在靶位点任一末端的400bp内来选择等级4靶向结构域。

[0228] 对于金黄色葡萄球菌,基于(1)从靶位点任一末端上游或下游的距离(即,HBG1/2 13nt靶区域),特异地是在靶位点任一末端的400bp内,(2)高水平的正交性,(3)5’G的存在,和(4)具有序列NNGRRT(SEQ ID NO:204)的PAM来选择等级1靶向结构域。基于(1)从靶位点任一末端上游或下游的距离(即,HBG1/213nt靶区域),特异地是在靶位点任一末端的400bp内,(2)高水平的正交性,和(3)具有序列NNGRRT(SEQ ID NO:204)的PAM来选择等级2靶向结构域。基于(1)从靶位点任一末端上游或下游的距离(即,HBG1/2 13nt靶区域),特异地是在

靶位点任一末端的400bp内，和(2)具有序列NNGRRT(SEQ ID NO:204)的PAM来选择等级3靶向结构域。基于(1)从靶位点任一末端上游或下游的距离(即，HBG1/2 13nt靶区域)，特异地是在靶位点任一末端的400bp内，和(2)具有序列NNGRRV(SEQ ID NO:205)的PAM来选择等级4靶向结构域。

[0229] 下表2呈现了化脓链球菌和金黄色葡萄球菌gRNA的靶向结构域，分为(a)等级(1、2、3或4)和(b)HBG1或HBG2。

表2:HBG1/2靶位点的gRNA靶向结构域序列

	HBG1	HBG2
化脓链球菌等级1	251-256	760-764
等级2	257-274	765-781
等级3	275-300	275-281, 283-300
等级4	301-366	301-311, 313-342, 344-348, 350-366, 782, 783

	HBG1	HBG2
金黄色葡萄球菌等级1	367-376	784-791
等级2	343, 377-393	778, 792-803
等级3	357, 365, 394-461	357, 365, 394-461
等级4	252-254, 256, 268, 272-274, 292, 295, 347, 348, 353, 360-362, 366, 598-759	292, 295, 347, 348, 353, 360-362, 366, 462-468 476-481, 489-587, 601-607, 614-620, 640-666, 674-679, 687-693, 708-714, 733-753, 762-764, 775, 779-781, 804-901

[0230] 设计为靶向改变CCAAT盒靶区域的另外的gRNA序列包括但不限于SEQ ID NO:970和971所示的序列。设计为靶向破坏CCAAT盒靶区域的gRNA的靶向结构域序列包括但不限于SEQ ID NO:1002。设计为靶向破坏CCAAT盒靶区域的gRNA的靶向结构域序列加PAM(UUUG)包括但不限于SEQ ID NO:1004。在某些实施例中，包含SEQ ID NO:1002和/或1004所示序列的gRNA可以与Cpf1蛋白或经修饰的Cpf1蛋白复合，以在CCAAT盒靶区域中产生改变。在某些实施例中，可以将包含表15、表18或表19所示的任何Cpf1 gRNA的gRNA与Cpf1蛋白或经修饰的Cpf1蛋白复合以形成RNP(“gRNA-Cpf1-RNP”)以在CCAAT盒靶区域中产生改变。在某些实施例中，经修饰的Cpf1蛋白可以是His-AsCpf1-nNLS(SEQ ID NO:1000)或His-AsCpf1-sNLS-sNLS(SEQ ID NO:1001)。在某些实施例中，gRNA-Cpf1-RNP的Cpf1分子可以由SEQ ID NO:1000、1001、1008-1018、1032、1035-39(Cpf1多肽序列)或SEQ ID NO:1019-1021所示的序列编码(Cpf1多核苷酸序列)。

[0231] gRNA可以被设计以靶向BCL11A(BCL11Ae)的红系特异性增强子，以破坏转录阻遏子BCL11A(在Friedland中描述，其通过引用并入本文)的表达。gRNA被设计以靶向BCL11A的红系特异性增强子中的内含子2的+58DHS区域中的GATA1结合基序(即BCL11Ae中的GATA1结合基序)，其中+58DHS增强子区包含SEQ ID NO:968所示的序列。设计以靶向破坏BCL11Ae中

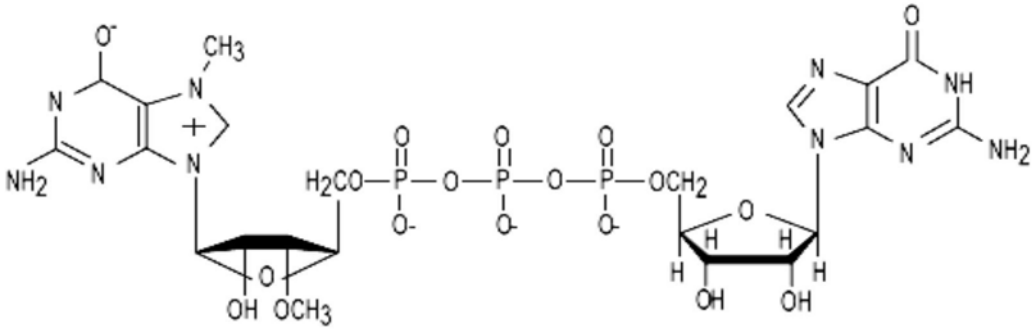
的GATA1结合基序的gRNA的靶向结构域序列包括但不限于SEQ ID NO:952-955所示的序列。设计以靶向破坏BCL11Ae中的GATA1结合基序的gRNA的靶向结构域序列加上PAM(NGG)包括但不限于SEQ ID NO:960-963所示的序列。

gRNA修饰

[0232] gRNA的活性、稳定性或其他特征可以通过并入某些修饰来改变。作为一个实例,瞬时表达或递送的核酸可能易于被例如细胞核酸酶降解。因此,本文所述的gRNA可以含有引入针对核酸酶的稳定性的一个或多个经修饰核苷或核苷酸。虽然不希望受理论束缚,但还应相信,本文所述的某些经修饰gRNA在引入细胞中时可展现降低的先天性免疫反应。本领域技术人员将了解一般在细胞(例如哺乳动物细胞)中响应外源核酸(尤其那些病毒或细菌来源的核酸)观察到的某些细胞反应。此类反应可以包括诱导细胞因子表达和释放以及细胞死亡,可通过本文所呈现的修饰来减少或完全消除。

[0233] 这个章节中讨论的某些示例性修饰可以包括于gRNA序列内的任一位置,包括但不限于在5'末端处或5'末端附近(例如,在5'末端的1-10、1-5或1-2个核苷酸内)和/或在3'末端处或3'末端附近(例如,在3'末端的1-10、1-5或1-2个核苷酸内)。在一些情形中,修饰定位于功能基序内,例如Cas9 gRNA的重复-抗重复双链体、Cas9或Cpf1 gRNA的茎环结构和/或gRNA的靶向结构域。

[0234] 作为一个实例,gRNA的5'末端可以包括真核mRNA帽结构或帽类似物(例如,G(5')ppp(5')G帽类似物、m7G(5')ppp(5')G帽类似物或3'-O-Me-m7G(5')ppp(5')G抗反向帽类似物(ARCA)),如下文所示:



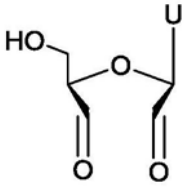
该帽或帽类似物可以在gRNA的化学合成或者体外转录期间被包括。

[0235] 按类似方式,gRNA的5'末端可能缺少5'三磷酸基团。例如,体外转录的gRNA可能经磷酸酶处理(例如,使用小牛肠碱性磷酸酶)以移除5'三磷酸基团。

[0236] 另一常见修饰包括在gRNA的3'末端添加多个(例如,1-10、10-20或25-200个)腺嘌呤(A)残基,称为polyA段。可在化学合成期间,在体外转录后使用多聚腺苷聚合酶(例如,大肠杆菌(E.coli)多聚(A)聚合酶),或在体内借助多聚腺苷酸化序列,将polyA段添加至gRNA,如Maeder中所述。

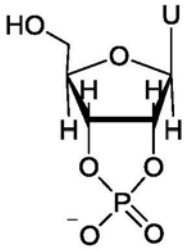
[0237] 应注意,本文所述的修饰可以按任一适宜方式组合,例如不论是在体内从DNA载体转录的gRNA,或者是在体外转录的gRNA,都可以包括5'帽结构或帽类似物中的一种或两种以及3'polyA段。

[0238] 指导RNA可以在3'末端U核糖处经修饰。例如,U核糖的两个末端羟基可以被氧化为醛基,并且伴随核糖环的打开,以提供如下所示的经修饰核苷:



其中“U”可以是未经修饰或经修饰的尿苷。

[0239] 3’末端U核糖可以经如下所示的2’3’环状磷酸酯修饰：



其中“U”可以是未经修饰或经修饰的尿苷。

[0240] 指导RNA可以例如通过并入本文所述的一个或多个经修饰核苷酸而含有可以针对降解稳定的3’核苷酸。在某些实施例中，尿苷可以由经修饰尿苷(例如，5-(2-氨基)丙基尿苷和5-溴代尿苷)或由本文所述的任何经修饰尿苷替代；腺苷和鸟苷可以由经修饰腺苷和鸟苷(例如，在8位处具有修饰，例如8-溴代鸟苷)或由本文所述的任何经修饰腺苷和鸟苷替代。

[0241] 在某些实施例中，可以将糖修饰的核糖核苷酸并入gRNA中，例如，其中2’OH-基团由选自以下的基团替代：H、-OR、-R(其中R可以是例如烷基、环烷基、芳基、芳烷基、杂芳基或糖)、卤基、-SH、-SR(其中R可以是例如烷基、环烷基、芳基、芳烷基、杂芳基或糖)、氨基(其中氨基可以是例如NH_2；烷基氨基、二烷基氨基、杂环基、芳基氨基、二芳基氨基、杂芳基氨基、二杂芳基氨基或氨基酸)；或氰基(-CN)。在某些实施例中，磷酸骨架可以如本文所述经修饰，例如经硫代磷酸酯(PhTx)基团修饰。在某些实施例中，gRNA的一个或多个核苷酸可以各自独立地是经修饰或未经修饰的核苷酸，包括但不限于2’-糖修饰的，例如2’-O-甲基、2’-O-甲氧基乙基，或2’-氟修饰的，包括例如，2’-F或2’-O-甲基腺苷(A)、2’-F或2’-O-甲基胞苷(C)、2’-F或2’-O-甲基尿苷(U)、2’-F或2’-O-甲基胸苷(T)、2’-F或2’-O-甲基鸟苷(G)、2’-O-甲氧基乙基-5-甲基尿苷(Teo)、2’-O-甲氧基乙基腺苷(Aeo)、2’-O-甲氧基乙基-5-甲基胞苷(m5Ceo)及其任何组合。

[0242] 指导RNA还可以包括“锁定”核酸(LNA)，其中2’OH-基团可以例如通过C1-6亚烷基C1-6亚杂烷基桥连接至同一核糖的4’碳。可使用任何适宜部分来提供此类桥，包括但不限于亚甲基、亚丙基、醚或氨基桥；O-氨基(其中氨基可以是例如NH_2；烷基氨基、二烷基氨基、杂环基、芳基氨基、二芳基氨基、杂芳基氨基或二杂芳基氨基、乙二胺或聚氨基)和氨基烷氧基或O$(CH_2)_n$-氨基(其中氨基可以是例如NH_2；烷基氨基、二烷基氨基、杂环基、芳基氨基、二芳基氨基、杂芳基氨基或二杂芳基氨基、乙二胺或聚氨基)。

[0243] 在某些实施例中，gRNA可以包括经修饰核苷酸，其为多环状(例如，三环；和“解锁”形式，如二醇核酸(GNA)(例如，R-GNA或S-GNA，其中核糖被附接至磷酸二酯键的二醇单元置换)，或苏糖核酸(TNA，其中核糖被α-L-苏呋喃糖基(threofuranosyl)-(3’→2’)置换)。

[0244] 通常，gRNA包括糖基核糖，其为具有氧的5元环。示例性的经修饰gRNA可以包括但

不限于核糖中氧的替代(例如,经硫(S)、硒(Se)或亚烷基,例如亚甲基或亚乙基);双键的添加(例如,以用环戊烯基或环己烯基替代核糖);核糖的缩环(例如,以形成环丁烷或氧杂环丁烷的4元环);核糖的扩环(例如,以形成具有另外的碳或杂原子的6元环或7元环,例如脱水己糖醇、阿卓糖醇、甘露醇、环己烷基、环己烯基以及吗啉代,其也具有氨基磷酸酯骨架)。尽管大多数的糖类似物改变位于2'位,但其他位点也适于修饰,包括4'位。在某些实施例中,gRNA包含4'-S、4'-Se或4'-C-氨基甲基-2'-O-Me修饰。

[0245] 在某些实施例中,可以将脱氮核苷酸(例如,7-脱氮-腺苷)并入gRNA中。在某些实施例中,可以将O-烷基化和N-烷基化核苷酸(例如,N6-甲基腺苷)并入gRNA中。在某些实施例中,gRNA分子中的一个或多个或所有核苷酸是脱氧核苷酸。

[0246] 在某些实施例中,本文使用的gRNA可以是经修饰的或未经修饰的gRNA。在某些实施例中,gRNA可以包括一个或多个修饰。在某些实施例中,一个或多个修饰可包括硫代磷酸酯键修饰、二硫代磷酸酯(PS2)键修饰、2'-O-甲基修饰或其组合。在某些实施例中,一个或多个修饰可以在gRNA的5'末端、在gRNA的3'末端或其组合。

[0247] 在某些实施例中,gRNA修饰可包含一个或多个二硫代磷酸酯(PS2)键修饰。

[0248] 在一些实施例中,本文使用的gRNA包括一个或多个或一段脱氧核糖核酸(DNA)碱基,在本文中也称为"DNA延伸"。在一些实施例中,本文使用的gRNA在gRNA的5'末端、gRNA的3'末端或其组合处包括DNA延伸。在某些实施例中,DNA延伸可以是1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、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、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99或100个DNA碱基长。例如,在某些实施例中,DNA延伸可以是1、2、3、4、5、10、15、20或25个DNA碱基长。在某些实施例中,DNA延伸可以包括一个或多个选自腺嘌呤(A)、鸟嘌呤(G)、胞嘧啶(C)或胸腺嘧啶(T)的DNA碱基。在某些实施例中,DNA延伸包括相同的DNA碱基。例如,DNA延伸可以包括一段腺嘌呤(A)碱基。在某些实施例中,DNA延伸可以包括一段胸腺嘧啶(T)碱基。在某些实施例中,DNA延伸包括不同DNA碱基的组合。在某些实施例中,DNA延伸可包含表24所示的序列。例如,DNA延伸可包含SEQ ID NO:1235-1250所示的序列。在某些实施例中,本文使用的gRNA包括DNA延伸以及一个或多个硫代磷酸酯键修饰、一个或多个二硫代磷酸酯(PS2)键修饰、一个或多个2'-O-甲基修饰或其组合。在某些实施例中,一个或多个修饰可以在gRNA的5'末端、在gRNA的3'末端或其组合。在某些实施例中,包括DNA延伸的gRNA可包含表19中所示的包括DNA延伸的序列。在一个特定的实施例中,包括DNA延伸的gRNA可包含SEQ ID NO:1051所示的序列。在某些实施例中,包括DNA延伸的gRNA可包含选自下组的序列,该组由以下组成:SEQ ID NO:1046-1060、1067、1068、1074、1075、1078、1081-1084、1086-1087、1089-1090、1092-1093、1098-1102和1106。不希望被理论所束缚,可以预期的是,本文中可以使用任何DNA延伸,只要它不与gRNA靶向的靶核酸杂交并且还相对于不包含此类DNA延伸的gRNA在靶核酸位点处表现出编辑增加。

[0249] 在一些实施例中,本文使用的gRNA包括一个或多个或一段核糖核酸(RNA)碱基,在本文中也称为"RNA延伸"。在一些实施例中,本文使用的gRNA在gRNA的5'末端、gRNA的3'末端或其组合处包括RNA延伸。在某些实施例中,RNA延伸可以是1、2、3、4、5、6、7、8、9、10、11、

12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、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、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99或100个RNA碱基长。例如，在某些实施例中，RNA延伸可以是1、2、3、4、5、10、15、20或25个RNA碱基长。在某些实施例中，RNA延伸可以包括一个或多个选自腺嘌呤(rA)、鸟嘌呤(rG)、胞嘧啶(rC)或尿嘧啶(rU)的RNA碱基，其中“r”代表RNA，2’-羟基。在某些实施例中，RNA延伸包括相同的RNA碱基。例如，RNA延伸可以包括一段腺嘌呤(rA)碱基。在某些实施例中，RNA延伸包括不同RNA碱基的组合。在某些实施例中，RNA延伸可包含表24所示的序列。例如，RNA延伸可包含1231-1234、1251-1253所示的序列。在某些实施例中，本文使用的gRNA包括RNA延伸以及一个或多个硫代磷酸酯键修饰、一个或多个二硫代磷酸酯(PS2)键修饰、一个或多个2’-O-甲基修饰或其组合。在某些实施例中，一个或多个修饰可以在gRNA的5’末端、在gRNA的3’末端或其组合。在某些实施例中，包括RNA延伸的gRNA可包含表19所示的包括RNA延伸的序列。在gRNA的5’末端包含RNA延伸的gRNA可包含选自由SEQ ID NO:1042-1045、1103-1105组成的组的序列。在gRNA的3’末端包括RNA延伸的gRNA可包含选自由SEQ ID NO:1070-1075、1079、1081、1098-1100组成的组的序列。

[0250] 预期本文中使用的gRNA也可以包括RNA延伸和DNA延伸。在某些实施例中，RNA延伸和DNA延伸两者可以都在gRNA的5’末端、gRNA的3’末端或其组合处。在某些实施例中，RNA延伸在gRNA的5’末端，并且DNA延伸在gRNA的3’末端。在某些实施例中，RNA延伸在gRNA的3’末端，并且DNA延伸在gRNA的5’末端。

[0251] 在一些实施例中，将在3’末端包括硫代磷酸酯修饰以及在5’末端包括DNA延伸的gRNA与RNA指导的核酸酶例如Cpf1复合以形成RNP，然后将所述RNP用于离体(即在此类细胞来源的受试者体外)编辑造血干细胞(HSC)或CD34+细胞的HBG基因座。

[0252] 本文所用的gRNA的实例包括SEQ ID NO:1051所示的序列。

失活gRNA分子

[0253] 根据本公开的失活指导RNA(dgRNA)分子包括包含降低的、低的或不可检测的切割活性的失活指导RNA分子。与活性指导RNA的靶向结构域序列相比，失活指导RNA的靶向结构域序列的长度短1、2、3、4、5、6、7、8、9或10个核苷酸。在某些实施例中，失活指导RNA分子可包含靶向结构域，其包含15个或更少核苷酸的长度、14个或更少核苷酸的长度、13个或更少核苷酸的长度、12个或更少核苷酸的长度或11个或更少核苷酸的长度。在一些实施例中，对失活指导RNA进行配置，使得它们不提供RNA指导的核酸酶切割事件。可以通过去除gRNA靶向结构域序列的5’末端来生成失活指导RNA，从而导致截短的靶向结构域序列。例如，如果配置为提供切割事件(即，长度为17个核苷酸或更多)的gRNA序列具有长度为20个核苷酸的靶向结构域序列，则可以通过从gRNA序列的5’末端去除5个核苷酸来产生失活指导RNA。例如，本文中使用的dgRNA可包含表2或表10所示的靶向结构域，所述靶向结构域已从gRNA序列的5’末端截短并且长度为15个核苷酸或更少。在某些实施例中，dgRNA可以被配置为结合(或缔合)在有待编辑的靶区域(例如，CCAAT盒靶区域、13nt靶区域、近端HBG1/2启动子靶序列、和/或BCL11Ae中的GATA1结合基序)内或近端的核酸序列。例如，表10所示的任何dgRNA可以用于结合在13nt靶区域或CCAAT盒靶区域近端的核酸序列。在某些实施例中，“近端”可表示靶区域(例如，CCAAT盒靶区域、13nt靶区域、近端HBG1/2启动子靶序列、和/或BCL11Ae

中的GATA1结合基序)的10、25、50、100或200个核苷酸内的区域。在某些实施例中,失活指导RNA未配置为将外源反式作用因子募集至靶区域。在某些实施例中,dgRNA被配置为使得当与RNA指导的核酸酶复合时,其不提供DNA切割事件。技术人员将理解,可以将失活指导RNA分子设计为包含与靶核酸中靶区域近端或内的区域互补的靶向结构域。在某些实施例中,失活指导RNA包含与双链DNA的转录链或非转录链互补的靶向结构域序列。如本文在"gRNA修饰"部分中针对指导RNA所描述的,本文的dgRNA可以在dgRNA的5'和3'末端包括修饰。例如,在某些实施例中,失活指导RNA可以在RNA的5'末端包括抗反向帽类似物(ARCA)。在某些实施例中,dgRNA可在3'末端包括聚A尾。

[0254] 在某些实施例中,将失活指导RNA与本文公开的基因组编辑系统和方法一起使用可增加活性指导RNA的总编辑水平。在某些实施例中,将失活指导RNA与本文公开的基因组编辑系统及其方法一起使用可增加缺失的频率。在某些实施例中,所述缺失可以从活性指导RNA的剪切位点向失活指导RNA结合位点延伸。以此方式,失活指导RNA可以改变活性指导RNA的方向性,并使编辑朝向期望的靶区域定向。

[0255] 如本文所用,术语"失活gRNA"和"截短的gRNA"可互换使用。

RNA指导的核酸酶

[0256] 根据本公开的RNA指导的核酸酶包括但不限于天然存在的2类的CRISPR核酸酶,例如Cas9和Cpf1,以及由其衍生或获得的其他核酸酶。还显示出某些RNA指导的核酸酶,例如Cas9,也具有解旋酶活性,使它们能够解旋核酸。在某些实施例中,根据本公开的RNA指导的解旋酶可以是本文所述的和上文标题"RNA指导的核酸酶"部分中所述的任何RNA核酸酶。在某些实施例中,RNA指导的核酸酶未配置为将外源反式作用因子募集到靶区域。在某些实施例中,RNA指导的解旋酶可以是被配置为缺乏核酸酶活性的RNA指导的核酸酶。例如,在某些实施例中,RNA指导的解旋酶可以是缺乏核酸酶活性但仍保留其解旋酶活性的催化失活性RNA指导的核酸酶。在某些实施例中,可将RNA指导的核酸酶突变以消除其核酸酶活性(例如,失活Cas9),从而产生无法切割核酸但仍可解旋DNA的催化失活性RNA指导的核酸酶。在某些实施例中,RNA指导的解旋酶可以与本文所述的任何失活指导RNA复合。例如,催化活性的RNA指导的解旋酶(例如,Cas9或Cpf1)可以与失活指导RNA形成RNP复合物,从而导致催化失活性死RNP(dRNP)。在某些实施例中,催化失活性RNA指导的解旋酶(例如,失活Cas9)和失活指导RNA可以形成dRNP。这些dRNP尽管不能提供切割事件,但仍保留其解旋酶活性,这对于解旋核酸很重要。

[0257] 在功能方面,RNA指导的核酸酶被定义为如下的那些核酸酶:(a)与gRNA相互作用(例如复合);和(b)与gRNA一起缔合或任选地切割或修饰DNA的靶区域,所述靶区域包括(i)与gRNA的靶向结构域互补的序列,以及任选地,(ii)称为"原型间隔子邻近基序"或"PAM"的另一序列,其更详细地描述于下文中。在说明以下实例时,RNA指导的核酸酶可以在广义上依据其PAM特异性和切割活性来定义,即使在共享相同PAM特异性或切割活性的个别RNA指导的核酸酶之间可能存在变异。技术人员将了解,本公开的一些方面涉及可以使用具有某种PAM特异性和/或切割活性的任何适宜RNA指导的核酸酶实施的系统、方法和组合物。为此,除非另外指定,否则术语RNA指导的核酸酶应理解为通用术语,并不限于RNA指导的核酸酶的任何特定类型(例如,Cas9与Cpf1)、物种(例如,化脓链球菌与金黄色葡萄球菌)或变异(例如,全长与截短或分裂;天然存在的PAM特异性与工程改造的PAM特异性等)。

[0258] 多种RNA指导的核酸酶可能需要PAM与原型间隔子之间的不同顺序关系。通常，Cas9识别原型间隔子3'的PAM序列。另一方面，Cpf1通常识别原型间隔子5'的PAM序列。

[0259] 除了识别PAM和原型间隔子的特定顺序定向以外，RNA指导的核酸酶还可以识别特定PAM序列。例如，金黄色葡萄球菌Cas9识别NNGRRT或NNGRRV的PAM序列，其中N个残基紧靠由gRNA靶向结构域所识别区域的3'。化脓链球菌Cas9识别NGG PAM序列。并且新凶手弗朗西斯菌(F.novicida)Cpf1识别TTN PAM序列。已经识别了多种RNA指导的核酸酶的PAM序列，并且Shmakov 2015描述了用于识别新的PAM序列的策略。还应注意，工程改造的RNA指导的核酸酶可具有与参考分子的PAM特异性不同的PAM特异性(例如，在工程改造的RNA指导的核酸酶的情形中，参考分子可以是衍生出RNA指导的核酸酶的天然存在的变体，或与工程改造的RNA指导的核酸酶具有最大氨基酸序列同源的天然存在的变体)。可以根据本文的实施例使用的PAM的实例包括但不限于SEQ ID NO:199-205所示的序列。

[0260] 除了其PAM特异性以外，RNA指导的核酸酶可通过其DNA切割活性来表征：天然存在的RNA指导的核酸酶典型地在靶核酸中形成DSB，但是已产生仅生成SSB的工程改造的变体(如以上讨论和Ran和Hsu 2013中讨论，通过引用并入本文)，或完全不剪切的工程改造的变体。

Cas9

[0261] 已确定化脓链球菌Cas9的晶体结构(Jinek 2014)以及与单分子指导RNA和靶DNA复合的金黄色葡萄球菌Cas9的晶体结构(Nishimasu 2014;Anders2014;和Nishimasu 2015)。

[0262] 天然存在的Cas9蛋白包含两个叶：识别(REC)叶和核酸酶(NUC)叶；每一叶包含特定的结构和/或功能域。REC叶包含富精氨酸桥螺旋(BH)结构域，以及至少一个REC结构域(例如REC1结构域和任选地REC2结构域)。REC叶与其他已知蛋白不共享结构相似性，指示其是独特的功能域。不希望受限于任何理论，突变分析提出BH和REC域的特殊功能作用：BH域似乎在gRNA:DNA识别中起作用，而REC域被认为与gRNA的重复:抗重复双链体相互作用并且介导Cas9/gRNA复合物的形成。

[0263] NUC叶包含RuvC结构域、HNH结构域和PAM相互作用(PI)域。RuvC结构域与逆转录病毒整合酶超家族成员共享结构相似性，并且切割靶核酸的非互补(即底部)链。其可从两个或更多个分裂RuvC基序(例如酿脓链球菌和金黄色葡萄球菌中的RuvC I、RuvCII和RuvCIII)形成。同时，HNH结构域在结构上与HNN内切核酸酶基序类似，并且切割靶核酸的互补(即顶部)链。顾名思义，PI结构域有助于PAM特异性。可以根据本文的实施例使用的编码Cas9 RuvC样和Cas9 HNH样结构域的多肽序列的实例在SEQ ID NO:15-23、52-123(RuvC样结构域)和SEQ ID NO:24-28、124-198(HNH样结构域)中示出。

[0264] 虽然Cas9的某些功能与上文所述的特定结构域相关(但不一定完全取决于所述结构域)，但这些和其他功能可以由其他Cas9结构域或任一叶上的多个结构域来介导或影响。例如，在化脓链球菌Cas9中，如在Nishimasu 2014中所述，gRNA的重复:抗重复双链体落在REC叶与NUC叶之间的沟中，并且双链体中的核苷酸与BH、PI和REC结构域中的氨基酸相互作用。第一茎环结构中的一些核苷酸也与多个结构域(PI、BH和REC1)中的氨基酸相互作用，第二和第三茎环(RuvC和PI结构域)中的一些核苷酸也是如此。可以根据本文的实施例使用的编码Cas9分子的多肽序列的实例在SEQ ID NO:1-2、4-6、12和14中示出。

Cpf1

[0265] 与crRNA和包括TTTN PAM序列的双链(ds)DNA靶复合的氨基酸球菌属(Acidaminococcus sp.)Cpf1的晶体结构已经由Yamano 2016(通过引用并入本文)解析。Cpf1像Cas9一样,具有两个叶:REC(识别)叶和NUC(核酸酶)叶。REC叶包括REC1和REC2结构域,其缺少与任何已知蛋白质结构的相似性。同时,NUC叶包括三个RuvC结构域(RuvC-I、-II和-III)和BH结构域。然而,与Cas9相反,Cpf1 REC叶缺少HNH结构域,并且包括同样缺少与已知蛋白质结构的相似性的其他结构域:结构上独特的PI结构域、三个楔形(WED)结构域(WED-I、-II和-III)和核酸酶(Nuc)结构域。

[0266] 虽然Cas9和Cpf1共享结构和功能的相似性,但应了解,某些Cpf1活性是由与任何Cas9域不同的结构域介导的。例如,靶DNA的互补链的切割似乎是由Nuc结构域介导,所述Nuc结构域在顺序上和空间上与Cas9的HNH结构域不同。另外,Cpf1 gRNA的非靶向部分(把手)采用假结(pseudoknot)结构,而不是Cas9gRNA中由重复:抗重复双链体形成的茎环结构。

[0267] 在某些实施例中,Cpf1蛋白可以是经修饰的Cpf1蛋白。在某些实施例中,经修饰的Cpf1蛋白可包括一个或多个修饰。在某些实施例中,修饰可以是但不限于Cpf1核苷酸序列或Cpf1氨基酸序列中的一个或多个突变、一个或多个另外的序列例如His标签或核定位信号(NLS)或其组合。在某些实施例中,经修饰的Cpf1在本文中也可以称为Cpf1变体。

[0268] 在某些实施例中,Cpf1蛋白可衍生自选自下组的Cpf1蛋白,该组由以下组成:氨基酸球菌属菌株BV3L6 Cpf1蛋白(AsCpf1)、毛螺菌科细菌ND2006 Cpf1蛋白(LbCpf1)和毛螺菌科细菌MA2020(Lb2Cpf1)。在某些实施例中,Cpf1蛋白可包含选自由以下组成的组的序列:SEQ ID NO:1016-1018,其分别具有SEQ ID NO:1019-1021的密码子优化的核酸序列。

[0269] 在某些实施例中,经修饰的Cpf1蛋白可包含核定位信号(NLS)。例如,但并非进行限制,可用于结合本文公开的方法和组合物的NLS序列将包含能够促进蛋白质导入细胞核的氨基酸序列。可用于结合本文公开的方法和组合物的NLS序列为本领域中已知。此类NLS序列的实例包括具有如下氨基酸序列的核质蛋白NLS:KRPAATKKAGQAKKKK(SEQ ID NO:1006)和具有氨基酸序列PKKKRKV(SEQ ID NO:1007)的猿猴病毒40“SV40”NLS。

[0270] 在某些实施例中,修饰的Cpf1蛋白的NLS序列定位于Cpf1蛋白序列的C末端处或定位在Cpf1蛋白序列的C末端附近。例如,但并非进行限制,修饰的Cpf1蛋白可以选自以下:His-AsCpf1-nNLS(SEQ ID NO.1000);His-AsCpf1-sNLS(SEQ ID NO.1008)和His-AsCpf1-sNLS-sNLS(SEQ ID NO.1001),其中“His”是指六组氨酸纯化序列,“AsCpf1”是指氨基酸球菌属物种Cpf1蛋白序列,“nNLS”是指核质蛋白NLS,并且“sNLS”是指SV40 NLS。NLS序列的同一性和C末端位置的另外的排列(例如附加两个或更多个nNLS序列或nNLS和sNLS序列(或其他NLS序列)的组合以及有和没有纯化序列(例如六个组氨酸序列)的序列)在本公开的主题范围内。

[0271] 在某些实施例中,修饰的Cpf1蛋白的NLS序列可以定位于Cpf1蛋白序列的N末端处或定位在Cpf1蛋白序列的N末端附近。例如,但并非进行限制,经修饰的Cpf1蛋白可以选自以下:His-sNLS-AsCpf1(SEQ ID NO.1009)、His-sNLS-sNLS-AsCpf1(SEQ ID NO.1010)和sNLS-sNLS-AsCpf1(SEQ ID NO.1011)。NLS序列的同一性和N末端位置的另外的排列(例如附加两个或更多个nNLS序列或nNLS和sNLS序列(或其他NLS序列)的组合以及有和没有纯化

序列(例如六个组氨酸序列)的序列)在本公开的主题范围内。

[0272] 在某些实施例中,经修饰的Cpf1蛋白可包含定位于Cpf1蛋白序列的N末端和C末端处或定位在Cpf1蛋白序列的N末端和C末端附近的NLS序列。例如,但并非进行限制,经修饰的Cpf1蛋白可以选自以下:His-sNLS-AsCpf1-sNLS(SEQ ID NO.1012)和His-sNLS-sNLS-AsCpf1-sNLS-sNLS(SEQ ID NO.1013)。NLS序列的同一性和N末端/C末端位置的另外的排列(例如附加两个或更多个nNLS序列或nNLS和sNLS序列(或其他NLS序列)的组合至N末端/C末端位置,以及有和没有纯化序列(例如六组氨酸序列))的序列在本公开的主题范围内。

[0273] 在某些实施例中,修饰的Cpf1蛋白可包含Cpf1蛋白序列的一个或多个半胱氨酸残基处的改变(例如,缺失或取代)。例如,但并非进行限制,经修饰的Cpf1蛋白可包含在选自由以下组成的组的位置处的改变:C65、C205、C334、C379、C608、C674、C1025和C1248。在某些实施例中,经修饰的Cpf1蛋白可包含一个或多个半胱氨酸残基取代丝氨酸或丙氨酸。在某些实施例中,经修饰的Cpf1蛋白可包含选自由以下组成的组的改变:C65S、C205S、C334S、C379S、C608S、C674S、C1025S和C1248S。在某些实施例中,经修饰的Cpf1蛋白可包含选自由以下组成的组的改变:C65A、C205A、C334A、C379A、C608A、C674A、C1025A和C1248A。在某些实施例中,经修饰的Cpf1蛋白可包含在位置C334和C674或C334、C379、和C674处的改变。在某些实施例中,经修饰的Cpf1蛋白可包含以下改变:C334S和C674S,或C334S、C379S、和C674S。在某些实施例中,经修饰的Cpf1蛋白可包含以下改变:C334A和C674A、或C334A、C379A、和C674A。在某些实施例中,经修饰的Cpf1蛋白可包含一个或多个半胱氨酸残基改变以及一个或多个NLS序列(例如,His-AsCpf1-nNLS Cys-less(SEQ ID NO.1014)或His-AsCpf1-nNLS Cys-low(SEQ ID NO.1015))的引入。在各种实施例中,在一个或多个半胱氨酸残基中包含缺失或取代的Cpf1蛋白展现出减少的聚集。

[0274] 在某些实施例中,本领域已知的其他经修饰的Cpf1蛋白可与本文所述的方法和系统一起使用。例如,在某些实施例中,经修饰的Cpf1可以是含有突变S542R/K548V/N552R(“Cpf1 RVR”)的Cpf1。Cpf1 RVR已显示切割具有TATV PAM的靶位点。在某些实施例中,经修饰的Cpf1可以是含有突变S542R/K607R(“Cpf1RR”)的Cpf1。Cpf1 RR已显示切割具有TYCV/CCCC PAM的靶位点。

[0275] 在一些实施例中,本文使用了Cpf1变体,其中所述Cpf1变体包含AsCpf1(氨基酸球菌属物种BV3L6)的选自下组的一个或多个残基处的突变,该组由以下组成:11、12、13、14、15、16、17、34、36、39、40、43、46、47、50、54、57、58、111、126、127、128、129、130、131、132、133、134、135、136、157、158、159、160、161、162、163、164、165、166、167、168、169、170、171、172、173、174、175、176、177、178、532、533、534、535、536、537、538、539、540、541、542、543、544、545、546、547、548、549、550、551、552、553、554、555、556、565、566、567、568、569、570、571、572、573、574、575、592、593、594、595、596、597、598、599、600、601、602、603、604、605、606、607、608、609、610、611、612、613、614、615、616、617、618、619、620、626、627、628、629、630、631、632、633、634、635、636、637、638、642、643、644、645、646、647、648、649、651、652、653、654、655、656、676、679、680、682、683、684、685、686、687、688、689、690、691、692、693、707、711、714、715、716、717、718、719、720、721、722、739、765、768、769、773、777、778、779、780、781、782、783、784、785、786、870、871、872、873、874、875、876、877、878、879、880、881、882、883、884或1048或AsCpf1直向同源物、同源物或变体的相应位置。

[0276] 在某些实施例中，本文使用的Cpf1变体可以包括Zhang等人的国际公开号WO 2017/184768 A1(“768公开”(其通过引用并入本文))中描述的任何Cpf1蛋白。

[0277] 在某些实施例中，本文中使用的经修饰的Cpf1蛋白(也称为Cpf1变体)可以由SEQ ID NO:1000、1001、1008-1018、1032、1035-39、1094-1097、1107-09(Cpf1多肽序列)或SEQ ID NO:1019-1021、1110-17(Cpf1多核苷酸序列)所示的任何序列编码。表20列出了示例性的Cpf1变体氨基酸和核苷酸序列。这些序列在图62中示出，其中详细说明了六组氨酸序列(带下划线的字母)和NLS序列(粗体字母)的位置。NLS序列的同一性和N末端/C末端位置的另外的排列(例如附加两个或更多个nNLS序列或nNLS和sNLS序列(或其他NLS序列)的组合至N末端/C末端位置，以及有和没有纯化序列(例如六组氨酸序列))的序列在本公开的主题范围内。

[0278] 在某些实施例中，本文公开的任何Cpf1蛋白或经修饰的Cpf1蛋白可与包含SEQ ID NO 1002和/或1004所示的靶向结构域的一种或多种gRNA复合，以改变CCAAT盒靶区域。在某些实施例中，本文公开的任何Cpf1蛋白或经修饰的Cpf1蛋白可以与包含表18或表19所示的序列的一种或多种gRNA复合。在某些实施例中，经修饰的Cpf1蛋白可以是His-AsCpf1-nNLS(SEQ ID NO:1000)或His-AsCpf1-sNLS-sNLS(SEQ ID NO:1001)。在某些实施例中，本文中使用的经修饰的Cpf1蛋白可以由SEQ ID NO:1000、1001、1008-1018、1032、1035-39、1094-1097、1107-09(Cpf1多肽序列)或SEQ ID NO:1019-1021、1110-17(Cpf1多核苷酸序列)所示的任何序列编码。在某些实施例中，经修饰的Cpf1蛋白可包含SEQ ID NO:1097所示的序列。

RNA指导的核酸酶的修饰

[0279] 上述RNA指导的核酸酶具有可用于多种应用的活性和特性，但技术人员将了解，RNA指导的核酸酶在某些情况下也可以经修饰，以改变切割活性、PAM特异性或其他结构或功能特征。

[0280] 首先参看改变切割活性的修饰，上文已描述降低或消除NUC叶内结构域活性的突变。可在RuvC结构域中、在Cas9 HNH结构域中或在Cpf1 Nuc结构域中进行的示例性突变描述于Ran&Hsu 2013和Yamano 2016中，以及Cotta-Ramusino中。通常，降低或消除两个核酸酶结构域之一中的活性的突变导致具有切口酶活性的RNA指导的核酸酶，但应注意，切口酶活性的类型根据哪个结构域失活而变化。作为一个实例，Cas9的RuvC结构域的失活将导致切割如下所示的互补或顶部链的切口酶(其中C表示切割位点)。

[0281] 另一方面，Cas9 HNH结构域的失活导致切割底部或非互补链的切口酶。

[0282] 对于化脓链球菌(Kleinstiver 2015a)和金黄色葡萄球菌(Kleinstiver2015b)，相对于天然存在的Cas9参考分子的PAM特异性的修饰已经由Kleinstiver等人描述。Kleinstiver等人还描述了改进Cas9的靶向保真性的修饰(Kleinstiver 2016)。这些参考文献中的每一篇都是通过引用并入本文。

[0283] 如Zetsche 2015和Fine 2015(均通过引用并入本文)所述，RNA指导的核酸酶已分为两个或更多个部分。

[0284] 在某些实施例中，RNA指导的核酸酶可以是尺寸经优化的或截短的，例如通过一个或多个缺失来进行，所述缺失减小核酸酶的尺寸，同时仍保留gRNA关联、靶标和PAM识别以及切割活性。在某些实施例中，RNA指导的核酸酶以共价或非共价方式，任选地通过接头结合至另一多肽、核苷酸或其他结构。示例性的结合的核酸酶和接头由Guilinger 2014(通过

引用结合在此用于所有目的)描述。

[0285] RNA指导的核酸酶还任选地包括标签,例如但不限于核定位信号,以促进RNA指导的核酸酶蛋白移动至细胞核中。在某些实施例中,RNA指导的核酸酶可以并入C末端和/或N末端核定位信号。核定位序列是本领域中已知的并且描述于Maeder和其他文献中。

[0286] 前述修饰列表旨在具有示例性,并且技术人员根据本公开将了解,在某些应用中,其他修饰可以是可能的或期望的。因此,为简洁起见,参考特定RNA指导的核酸酶呈现本公开的示例性系统、方法和组合物,但应理解,所用RNA指导的核酸酶可以不改变其操作原理的方式经修饰。此类修饰在本公开的范围内。

编码RNA指导的核酸酶的核酸

[0287] 本文提供编码RNA指导的核酸酶(例如Cas9、Cpf1或其功能片段)的核酸。可以根据本文的实施例使用的编码Cas9分子的核酸序列的实例在SEQ ID NO:3、7-11、13中示出。先前已描述编码RNA指导的核酸酶的示例性核酸(参见,例如,Cong 2013;Wang 2013;Mali 2013;Jinek 2012)。

[0288] 在一些情形中,编码RNA指导的核酸酶的核酸可以是合成的核酸序列。例如,合成核酸分子可以经化学修饰。在某些实施例中,编码RNA指导的核酸酶的mRNA将具有一种或多种(例如,所有)以下特性:其可以被加帽;多聚腺苷酸化;以及经5-甲基胞苷和/或假尿苷取代。

[0289] 合成的核酸序列也可以经密码子优化,例如,至少一个非常见密码子或较不常见的密码子已经常见密码子替代。例如,合成核酸可以引导经优化信使mRNA的合成,例如,针对在哺乳动物表达系统中的表达进行优化,例如,本文所述。密码子优化的Cas9编码序列的实例呈现于Cotta-Ramusino中。

[0290] 另外,或可替代地,编码RNA指导的核酸酶的核酸可包含核定位序列(NLS)。核定位序列是本领域中已知的。

候选分子的功能分析

[0291] 候选RNA指导的核酸酶、gRNA和其复合物可以通过本领域中已知的标准方法来评估。参见,例如Cotta-Ramusino。RNP复合物的稳定性可通过差示扫描荧光法来评估,如下文所述。

差示扫描荧光法(DSF)

[0292] 包含gRNA和RNA指导的核酸酶的核糖核蛋白(RNP)复合物的热稳定性可以通过DSF来测量。DSF技术测量蛋白质的热稳定性,所述热稳定性可以在有利条件下(例如添加结合RNA分子,例如gRNA)增加。

[0293] DSF测定可以根据任何适宜方案来进行,并且可以用于任何适宜环境中,包括但不限于(a)测试不同条件(例如gRNA:RNA指导的核酸酶蛋白的不同化学计量比,不同缓冲溶液等)以鉴别RNP形成的最佳条件;和(b)测试RNA指导的核酸酶和/或gRNA的修饰(例如化学修饰、序列改变等)以鉴别改进RNP形成或稳定性的那些修饰。DSF测定的一个读出是RNP复合物的熔融温度的位移;相对高的位移表明,相对于特征为较低位移的参考RNP复合物,RNP复合物更稳定(并且可能因此具有更高活性或更有利的形成动力学、降解动力学或另一功能特征)。在作为筛选工具布置DSF测定时,可指定阈值熔融温度位移,使得输出是熔融温度位移等于或高于阈值的一种或多种RNP。例如,阈值可以是5℃-10℃(例如5°、6°、7°、8°、9°、

10°)或更高,并且输出可以是一种或多种特征为熔融温度位移大于或等于该阈值的RNP。

[0294] DSF测定条件的两个非限制性实例陈述于下文中:

[0295] 为了确定形成RNP复合物的最佳溶液,将水+10x SYPROOrange®(生命技术公司(Life Techonologies)目录号S-6650)中固定浓度(例如2μM)的Cas9分配至384孔板中。然后添加稀释于具有不同pH和盐的溶液中的等摩尔量的gRNA。在室温下孵育10分钟并短暂离心以移除任何气泡后,使用Bio-Rad CFX384™Real-Time System C1000 Touch™热循环仪和Bio-Rad CFX Manager软件运行从20℃至90℃的梯度,每10秒温度增加1℃。

[0296] 第二测定由以下步骤组成:在来自上文测定1的最适缓冲液中混合不同浓度的gRNA与固定浓度(例如2μM)的Cas9,并在384孔板中孵育(例如,在室温下10分钟)。添加等体积的最适缓冲液+10x SYPROOrange®(生命技术公司目录号S-6650),并且将板用Microseal® B粘合剂(MSB-1001)密封。在短暂离心以移除任何气泡后,使用Bio-Rad CFX384™Real-Time System C1000 Touch™热循环仪和Bio-Rad CFX Manager软件运行从20℃至90℃的梯度,每10秒温度增加1℃。

基因组编辑策略

[0297] 在本公开的各个实施例中,上述基因组编辑系统用于在细胞内或从细胞获得的DNA的靶向区中产生编辑(即改变)。本文描述产生特定编辑的各种策略,并且这些策略通常是在所需修复结果、个别编辑(例如SSB或DSB)的数目和定位以及此类编辑的靶位点方面加以描述。

[0298] 涉及SSB或DSB的形成的基因组编辑策略是通过修复结果来表征,这些修复结果包括:(a)所靶向区域的全部或部分的缺失;(b)所靶向区域的全部或部分中的插入或其替代;或(c)所靶向区域的全部或部分的中断。此分组并不旨在具有限制性,或结合于任何具体理论或模型,而仅是为了便于呈现而提供。技术人员将了解,所列结果不会相互排斥,并且一些修复可导致其他结果。除非另外指定,否则对特定编辑策略或方法的描述不应理解为需要特定的修复结果。

[0299] 所靶向区域的替代通常涉及用同源序列替代所靶向区域内现有序列的全部或部分,例如通过基因修正或基因转变来进行,两种修复结果是通过HDR路径介导。HDR是通过使用供体模板来促进,所述供体模板可以是单链或双链的,如下文更详细地描述。单链或双链模板可以是外源的,在这种情形中其将促进基因修正,或者所述模板可以是内源的(例如细胞基因组内的同源序列),以促进基因转变。外源模板可以具有不对称悬突(即,模板中与DSB位点互补的部分可在3'或5'方向上偏移,而不是位于供体模板内的中心),例如如Richardson 2016(通过引用并入本文)所述。在模板为单链的情况下,其可以对应于所靶向区域的互补(顶部)或非互补(底部)链。

[0300] 在一些情形中,通过在所靶向区域中或周围形成一个或多个切口来促进基因转变和基因修正,如在Ran&Hsu 2013和Cotta-Ramusino中所述。在一些情形中,双重切口酶策略用于形成两个偏移SSB,其又形成具有悬突(例如5'悬突)的单一DSB。

[0301] 所靶向序列的全部或部分的中断和/或缺失可通过多种修复结果来实现。作为一个实例,可通过同时产生两个或更多个侧接所靶向区域的DSB使序列缺失,然后在修复DSB时切除所述所靶向区域,如在Maeder中针对LCA10突变所述。作为另一实例,可在修复前通过以下方式产生的缺失来中断序列:形成具有单链悬突的双链断裂,之后对悬突进行核酸

外切加工。

[0302] 靶序列中断的一个特定子集是通过在所靶向序列内形成indel来介导,其中修复结果典型地是通过NHEJ路径(包括Alt-NHEJ)来介导。NHEJ由于其与indel突变的关联而称为“错误倾向”修复路径。然而,在一些情形中,DSB是通过NHEJ修复,并且不改变其周围的序列(所谓的“完美”或“无瘢痕”修复);这通常需要DSB的两端完美连接。同时,Indel被认为是从游离DNA末端在连接前的酶加工产生,其在一个或两个游离末端的一条或两条链中添加和/或移除核苷酸。

[0303] 由于游离DSB末端的酶加工可具有随机性,indel突变往往是可变的,沿分布发生,并且可能受到多种因素影响,包括特定靶位点、所用细胞类型、所用基因组编辑策略等。即使如此,有可能引起关于indel形成的有限泛化:通过修复单一DSB形成的缺失最常在1-50bp范围内,但是可能达到大于100-200bp。通过修复单一DSB形成的插入往往较短,并且常包括紧密围绕断裂位点的序列的短重复。然而,有可能获得大的插入,并且在这些情形中,插入的序列通常已经被追溯至基因组的其他区域或追溯至存在于细胞中的质粒DNA。

[0304] indel突变和经配置以产生indel的基因组编辑系统可用于例如在不需要产生特定的最终序列时和/或在可耐受移码突变的情况下中断靶序列。其还可以用于偏好特定序列的环境中,只要某些所需序列往往优先通过给定位点处的SSB或DSB的修复而发生即可。indel突变还是可用于评估或筛选特定基因组编辑系统和其组分的活性的工具。在这些和其他环境中,indel可以通过以下各项来表征:(a)其在与基因组编辑系统接触的细胞的基因组中的相对和绝对频率,和(b)相对于未编辑序列的数值差异的分布,例如±1、±2、±3等。作为一个实例,在先导发现(lead-finding)环境中,可基于在受控条件下的indel读出筛选多种gRNA,以鉴别最有效驱动靶位点处的切割的那些gRNA。可以选择以阈值频率或以高于阈值的频率产生indel或产生indel的特定分布的指导以供进一步研究和开发。Indel频率和分布还可以用作读出,用于评估不同的基因组编辑系统实施或配置和递送方法,例如通过保持gRNA不变并改变某些其他反应条件或递送方法。

多重策略

[0305] 根据本公开的基因组编辑系统也可以用于多重基因编辑以在相同基因座或不同基因座中产生两个或更多个DSB。本文公开的任何RNA指导的核酸酶和gRNA都可以在基因组编辑系统中用于多重基因编辑。涉及形成多个DSB或SSB的编辑策略描述于例如Cotta-Ramusino中。

[0306] 如本文所公开的,可以在基因组编辑系统中使用多种gRNA,以将改变(例如,缺失,插入)引入HBG1和/或HBG2的13nt靶区域。在某些实施例中,包含SEQ ID NO:251-901、940-942所示的靶向结构域的一种或多种gRNA可以用于在HBG1和/或HBG2的13nt靶区域中引入改变。在其他实施例中,可以在基因组编辑系统中使用多种gRNA,以将改变引入CCAAT盒靶区域。在某些实施例中,包含SEQ ID NO:970、971、996、997所示的序列的一种或多种gRNA可以用于在CCAAT盒靶区域中引入改变。在某些实施例中,包含SEQ ID NO:1002、1004所示的靶向结构域的一种或多种gRNA可以用于在CCAAT盒靶区域中引入改变。在其他实施例中,可以在基因组编辑系统中使用多种gRNA,以将改变引入BCL11Ae中的GATA1结合基序中。在某些实施例中,包含SEQ ID NO:952-955所示的靶向结构域的一种或多种gRNA可以用于在BCL11Ae中的GATA1结合基序中引入改变。多种gRNA也可以用于基因组编辑系统中,以将改

变引入BCL11Ae中的GATA1结合基序、CCAAT盒靶区域、HBG1和/或HBG2的13nt靶区域、或其组合。在某些实施例中，可以使用包含SEQ ID NO:952-955所示的靶向结构域的一种或多种gRNA在BCL11Ae中GATA1结合基序中引入改变，以及可以使用包含SEQ ID NO:251-901、940-942所示的靶向结构域的一种或多种gRNA在HBG1和/或HBG2的13nt靶区域中引入改变。在某些实施例中，可使用包含SEQ ID NO:952-955所示的靶向结构域的一种或多种gRNA在BCL11Ae中GATA1结合基序中引入改变，以及使用包含SEQ ID NO:970、971、996、997所示的靶向结构域的一种或多种gRNA或gRNA在CCAAT盒靶区域中引入改变。在某些实施例中，可使用包含SEQ ID NO:952-955所示的靶向结构域的一种或多种gRNA在BCL11Ae中GATA1结合基序中引入改变，以及使用包含SEQ ID NO:1002、1004所示的靶向结构域的一种或多种gRNA在CCAAT盒靶区域中引入改变。

[0307] 在某些实施例中，可以在基因组编辑系统中使用多种gRNA和RNA指导的核酸酶，以将改变(例如，缺失，插入)引入HBG1和/或HBG2的CCAAT盒靶区域。在某些实施例中，RNA指导的核酸酶可以是Cas9、经修饰的Cas9(例如，D10A)、Cpf1或经修饰的Cpf1。

供体模板设计

[0308] 供体模板设计详细描述于文献中，例如Cotta-Ramusino中。DNA寡聚体供体模板(寡脱氧核苷酸或ODN)可以是单链(ssODN)或双链(dsODN)的，可以用于促进基于HDR的DSB修复或加强整体编辑率，并且尤其可用于将改变引入靶DNA序列中、将新序列插入靶序列中、或完全替代靶序列。

[0309] 不论是单链或双链，供体模板通常包括与待切割的靶序列内或附近(例如侧接或邻近)的DNA区域同源的区域。这些同源区域在本文中称作“同源臂”，并且示意性地示于下文中：

[5’同源臂]-[替代序列]-[3’同源臂]。

[0310] 同源臂可具有任何适宜长度(如果仅使用一个同源臂，包括0个核苷酸)，并且3’和5’同源臂可以具有相同长度或可以具有不同长度。适当同源臂长度的选择可能受到多种因素影响，例如对避免与某些序列(例如Alu重复序列或其他极为常见的元件)的同源性或微同源性的期望。例如，可以缩短5’同源臂以避免序列重复元件。在其他实施例中，可以缩短3’同源臂以避免序列重复元件。在一些实施例中，可以同时缩短5’和3’同源臂以避免包括某些序列重复元件。另外，一些同源臂设计可以改进编辑效率或增加所需修复结果的频率。例如，Richardson 2016(通过引用并入本文)发现，单链供体模板的3’和5’同源臂的相对不对称性影响修复率和/或结果。

[0311] 供体模板中的替代序列已在其他地方进行了描述，包括在Cotta-Ramusino中。替代序列可以是任何适宜长度(如果所需修复结果是缺失，那么包括0个核苷酸)，并且相对于需要编辑的细胞内的天然存在的序列，典型地包括1、2、3或更多个序列修饰。一种常见序列修饰涉及改变天然存在的序列以修复突变，所述突变与需要治疗的疾病或病症相关。另一常见序列修饰涉及改变一个或多个序列，所述序列互补或然后RNA指导的核酸酶的PAM序列或用于产生SSB或DSB的一种或多种gRNA的靶向结构域，以在将替代序列并入靶位点中之后减少或消除靶位点的重复切割。

[0312] 如果使用线性ssODN，其可以经配置以(i)退火至靶核酸的带切口链，(ii)退火至靶核酸的完整链，(iii)退火至靶核酸的正链，和/或(iv)退火至靶核酸的负链。ssODN可以

具有任何合适的长度，例如大约、至少或不超过80-200个核苷酸(例如80、90、100、110、120、130、140、150、160、170、180、190，或200个核苷酸)。

[0313] 应注意，模板核酸也可以是核酸载体，例如病毒基因组或环状双链DNA，例如质粒。包含供体模板的核酸载体可以包括其他编码或非编码元件。例如，模板核酸可以作为病毒基因组的部分来递送(例如在AAV或慢病毒基因组中)，其包括某些基因组骨架元件(例如在AAV基因组情形中，末端反向重复序列)并且任选地包括编码gRNA和/或RNA指导的核酸酶的其他序列。在某些实施例中，供体模板可以邻近或侧接由一种或多种gRNA识别的靶位点，以促进在供体模板的一端或两端上形成游离DSB，所述供体模板可以参与使用相同gRNA修复在细胞DNA中形成的相应SSB或DSB。适合用作供体模板的示例性核酸载体描述于Cotta-Ramusino中，其通过引用并入本文。

[0314] 不论使用何种形式，模板核酸都可以设计为避免不期望的序列。在某些实施例中，可以缩短一个或两个同源臂以避免与某些序列重复元件(例如，Alu重复、LINE元件等)重叠。

[0315] 在某些实施例中，ssODN供体模板中可以包括沉默的非致病性的SNP，以允许鉴定基因编辑事件。

[0316] 在某些实施例中，供体模板可以是与待切割的靶序列之内或附近的DNA区域非同源的非特异性模板。在某些实施例中，用于靶向BCL11Ae中的GATA1结合基序的供体模板可以包括但不限于与BCL11Ae中的GATA1结合基序内或附近的DNA区域不同源的非靶特异性模板。在某些实施例中，用于靶向13nt靶区域的供体模板可包括但不限于与13nt靶区域内或附近的DNA区域不同源的非靶特异性模板。

[0317] 如本文所用的术语，供体模板或模板核酸是指可与RNA核酸酶分子和一种或多种gRNA分子结合使用以改变(例如，缺失、破坏或修饰)靶DNA序列的核酸序列。在某些实施例中，模板核酸导致HBG1和/或HBG2的CCAAT盒靶区域的改变(例如缺失)。在某些实施例中，所述改变是非天然发生的改变。在某些实施例中，HBG1和/或HBG2的CCAAT靶区域的非天然发生的改变可包括18nt靶区域、11nt靶区域、4nt靶区域或1nt靶区域、或其组合。在某些实施例中，所述改变是天然发生的改变。在某些实施例中，HBG1和/或HBG2的CCAAT盒靶区域的天然发生的改变可包括13nt靶区域、c.-117G>A靶区域或其组合。在某些实施例中，模板核酸是ssODN。在某些实施例中，ssODN是正链或负链。

[0318] 例如，用于在18nt靶区域(HBG1 c.-104至-121、HBG2 c.-104至-121或其组合)处引入18nt缺失的模板核酸可包含5’同源臂、替代序列和3’同源臂，其中替代序列为0个核苷酸或0bp。在某些实施例中，5’同源臂可以是长度约25至约200个核苷酸或更多，例如，长度至少约25、50、75、100、125、150、175或200个核苷酸。在某些实施例中，5’同源臂包含约50至100bp，例如55至95、60至90，70至90或80至90bp与18nt靶区域的5’同源。在某些实施例中，3’同源臂可以是长度约25至约200个核苷酸或更多，例如，长度至少约25、50、75、100、125、150、175或200个核苷酸。在某些实施例中，3’同源臂包含约50至100bp，例如55至95、60至90，70至90或80至90bp与18nt靶区域的3’同源。在某些实施例中，5’和3’同源臂的长度是对称的。在某些实施例中，5’和3’同源臂的长度是不对称的。在某些实施例中，模板核酸是ssODN。在某些实施例中，ssODN是正链。在某些实施例中，ssODN是负链。在某些实施例中，ssODN包含以下、可以基本上由以下组成或由以下组成：SEQ ID NO:974(OLI16409)或SEQ

ID NO:975(OLI16410)。

[0319] 在某些实施例中,用于在11nt靶区域(HBG1 c.-105至-115、HBG2 c.-105至-115或其组合)处引入11nt缺失的模板核酸可包含5'同源臂、替代序列和3'同源臂,其中替代序列为0个核苷酸或0bp。在某些实施例中,5'同源臂可以是长度约25至约200个核苷酸,例如,长度至少约25、50、75、100、125、150、175或200个核苷酸。在某些实施例中,5'同源臂包含约50至100bp,例如55至95、60至90,70至90或80至90bp与11nt靶区域的5'同源。在某些实施例中,3'同源臂可以是长度约25至约200个核苷酸,例如,长度至少约25、50、75、100、125、150、175或200个核苷酸。在某些实施例中,3'同源臂包含约50至100bp,例如55至95、60至90,70至90或80至90bp与11nt靶区域的3'同源。在某些实施例中,5'和3'同源臂的长度是对称的。在某些实施例中,5'和3'同源臂的长度是不对称的。在某些实施例中,模板核酸是ssODN。在某些实施例中,ssODN是正链。在某些实施例中,ssODN是负链。在某些实施例中,ssODN可以包含以下、可以基本上由以下组成或由以下组成:SEQ ID NO:976(OLI16411)或SEQ ID NO:978(OLI16413)。

[0320] 在某些实施例中,用于在4nt靶区域(HBG1 c.-112至-115、HBG2 c.-112至-115或其组合)处引入4nt缺失的模板核酸可包含5'同源臂、替代序列和3'同源臂,其中替代序列为0个核苷酸或0bp。在某些实施例中,5'同源臂可以是长度约25至约200个核苷酸,例如,长度至少约25、50、75、100、125、150、175或200个核苷酸。在某些实施例中,5'同源臂包含约50至100bp,例如55至95、60至90,70至90或80至90bp与4nt靶区域的5'同源。在某些实施例中,3'同源臂可以是长度约25至约200个核苷酸,例如,长度至少约25、50、75、100、125、150、175或200个核苷酸。在某些实施例中,3'同源臂包含约50至100bp,例如55至95、60至90,70至90或80至90bp与4nt靶区域的3'同源。在某些实施例中,5'和3'同源臂的长度是对称的。在某些实施例中,5'和3'同源臂的长度是不对称的。在某些实施例中,模板核酸是ssODN。在某些实施例中,ssODN是正链。在某些实施例中,ssODN是负链。在某些实施例中,ssODN包含以下、基本上由以下组成或由以下组成:SEQ ID NO:984(OLI16419)、SEQ ID NO:985(OLI16420)、SEQ ID NO:986(OLI16421)、SEQ ID NO:987(OLI16422)、SEQ ID NO:988(OLI16423)、SEQ ID NO:989(OLI16424)、SEQ ID NO:990(OLI16425)、SEQ ID NO:991(OLI16426)、SEQ ID NO:992(OLI16427)、SEQ ID NO:993(OLI16428)、SEQ ID NO:994(OLI16429)、或SEQ ID NO:995(OLI16430)。

[0321] 在某些实施例中,用于在1nt靶区域(HBG1 c.-116、HBG2 c.-116或其组合)处引入1nt缺失的模板核酸可包含5'同源臂、替代序列和3'同源臂,其中替代序列为0个核苷酸或0bp。在某些实施例中,5'同源臂可以是长度约25至约200个核苷酸,例如,长度至少约25、50、75、100、125、150、175或200个核苷酸。在某些实施例中,5'同源臂包含约50至100bp,例如55至95、60至90,70至90或80至90bp与1nt靶区域的5'同源。在某些实施例中,3'同源臂可以是长度约25至约200个核苷酸,例如,长度至少约25、50、75、100、125、150、175或200个核苷酸。在某些实施例中,3'同源臂包含约50至100bp,例如55至95、60至90,70至90或80至90bp与1nt靶区域的3'同源。在某些实施例中,5'和3'同源臂的长度是对称的。在某些实施例中,5'和3'同源臂的长度是不对称的。在某些实施例中,模板核酸是ssODN。在某些实施例中,ssODN是正链。在某些实施例中,ssODN是负链。在某些实施例中,ssODN可以包含以下、可以基本上由以下组成或由以下组成:SEQ ID NO:982(OLI16417)或SEQ ID NO:983

(OLI16418)。

[0322] 在某些实施例中,CCAAT盒靶区域处的改变重现或类似于天然发生的改变,例如13nt缺失。在某些实施例中,用于在13nt靶区域(HBG1 c.-116、HBG2c.-116或其组合)处引入13nt缺失的模板核酸可包含5'同源臂、替代序列和3'同源臂,其中替代序列为0个核苷酸或0bp。在某些实施例中,5'同源臂可以是长度约25至约200个核苷酸,例如,长度至少约25、50、75、100、125、150、175或200个核苷酸。在某些实施例中,5'同源臂包含约50至100bp,例如55至95、60至90,70至90或80至90bp与13nt靶区域的5'同源。在某些实施例中,3'同源臂可以是长度约25至约200个核苷酸,例如,长度至少约25、50、75、100、125、150、175或200个核苷酸。在某些实施例中,3'同源臂包含约50至100bp,例如55至95、60至90,70至90或80至90bp与13nt靶区域的3'同源。在某些实施例中,5'和3'同源臂的长度是对称的。在某些实施例中,5'和3'同源臂的长度是不对称的。在某些实施例中,模板核酸是ssODN。在某些实施例中,ssODN是正链。在某些实施例中,ssODN是负链。在某些实施例中,ssODN可以包含以下、可以基本上由以下组成或由以下组成:SEQ ID NO:979(OLI16414)或SEQ ID NO:977(OLI16412)。

[0323] 在某些实施例中,CCAAT盒靶区域处的改变重现或类似于天然发生的改变,例如在-117G>A靶区域从G取代为A。在某些实施例中,用于在-117G>A靶区域(HBG1 c.-117G>A,HBG2 c.-117G>A或其组合)处引入-117G>A取代的模板核酸可包含5'同源臂、替代序列和3'同源臂,其中替代序列为0个核苷酸或0bp。在某些实施例中,5'同源臂可以是长度约100至约200个核苷酸,例如长度至少约100、125、150、175或200个核苷酸。在某些实施例中,5'同源臂包含约50至100bp,例如55至95、60至90、70至90或80至90bp与-117G>A靶区域的5'同源。在某些实施例中,3'同源臂可以是长度约25至约200个核苷酸,例如,长度至少约25、50、75、100、125、150、175或200个核苷酸。在某些实施例中,3'同源臂包含约50至100bp,例如55至95、60至90、70至90或80至90bp与-117G>A靶区域的3'同源。在某些实施例中,5'和3'同源臂的长度是对称的。在某些实施例中,5'和3'同源臂的长度是不对称的。在某些实施例中,模板核酸是ssODN。在某些实施例中,ssODN是正链。在某些实施例中,ssODN是负链。在某些实施例中,ssODN可以包含以下、可以基本上由以下组成或由以下组成:SEQ ID NO:980(OLI16415)或SEQ ID NO:981(OLI16416)。

[0324] 在某些实施例中,5'同源臂包含5'硫代磷酸酯(PhTx)修饰。在某些实施例中,3'同源臂包含3'PhTx修饰。在某些实施例中,模板核酸包含5'和3'PhTx修饰。

[0325] 在某些实施例中,用于在CCAAT盒靶区域处引入改变(例如缺失)的ssODN可以与RNA核酸酶和靶向CCAAT靶区域的一种或多种gRNA结合使用,例如,表7、表10、表12、表13、表18、表19中公开的gRNA。

靶细胞

[0326] 根据本公开的基因组编辑系统可以用于操作或改变细胞,例如以编辑或改变靶核酸。在各个实施例中,操作可在体内进行或离体进行。

[0327] 可以根据本公开的实施例操作或改变多种细胞类型,并且在一些情形中,例如在体内应用中,例如通过将根据本公开的基因组编辑系统递送至多个细胞类型来改变或操作多个细胞类型。然而,在其他情形中,可能需要将操作或改变限制于特定的一个或多个细胞类型。例如,在一些情况下可能需要编辑具有有限分化潜力的细胞或最终分化细胞,例如在

Maeder情形中的感光细胞，其中预期基因型的修饰会导致细胞表型变化。然而，在其他情形中，可能需要编辑分化程度较低的、多潜能或多能性的干细胞或祖细胞。举例来说，细胞可以是胚胎干细胞、诱导型多能干细胞(iPSC)、造血干细胞/祖细胞(HSPC)或其他干细胞或祖细胞类型，其分化为与给定应用或适应证相关的细胞类型。

[0328] 作为推论，根据所靶向的一种或多种细胞类型和/或所需编辑结果，所改变或操作的细胞不同地为分裂细胞或非分裂细胞。

[0329] 在离体操作或改变细胞时，细胞可以立即使用(例如施用于受试者)，或可以维持或储存细胞以供将来使用。本领域技术人员将了解，可以使用本领域已知的任何适宜方法将细胞维持在培养中或储存(例如冷冻于液氮中)。

基因组编辑系统的实施:递送、配制和施用途径

[0330] 如上文所讨论的，本公开的基因组编辑系统可以用任何适宜方式来实施，意味着此类系统的组分(包括但不限于RNA指导的核酸酶、gRNA和可选供体模板核酸)可以用任何适宜形式或形式的组合来递送、配制或施用，从而在细胞、组织或受试者中导致基因组编辑系统的转导、表达或引入和/或引起所需的修复结果。表3和4展示基因组编辑系统实施方式的若干非限制性实例。但本领域技术人员将了解，这些列表不是综合性的，并且其他实施是可能的。尤其参照表3，该表示例了包含单gRNA和任选供体模板的基因组编辑系统的若干示例性实施方式。然而，根据本公开的基因组编辑系统可以并入多种gRNA、多种RNA指导的核酸酶以及其他组分，例如蛋白质，并且基于该表中所示的原理，多种实施将为技术人员所了解。在该表中，[N/A]指示，基因组编辑系统不包括所指示的组分。

表3

基因组编辑系统组分			
RNA 指导的核酸酶	**gRNA**	**供体模板**	**注解**
蛋白质	RNA	[N/A]	与gRNA分子复合的RNA指导的核酸酶蛋白（RNP 复合物）
蛋白质	RNA	DNA	如上所述的RNP复合物加单链或双链供体模板。
蛋白质	DNA	[N/A]	RNA 指导的核酸酶蛋白加从 DNA 转录的 gRNA。
蛋白质	DNA	DNA	RNA 指导的核酸酶蛋白加 gRNA 编码 DNA 和单独的 DNA 供体模板。
蛋白质	DNA		RNA指导的核酸酶蛋白以及编码 gRNA 和供体模板两者的单一 DNA。

DNA			编码 RNA 指导的核酸酶、gRNA 和供体模板的 DNA 或 DNA 载体。
DNA	DNA	[N/A]	两个单独的 DNA 或两个单独的 DNA 载体，其分别编码 RNA 指导的核酸酶和 gRNA。
DNA	DNA	DNA	三个单独的 DNA 或三个单独的 DNA 载体，其分别编码 RNA 指导的核酸酶、gRNA 和供体模板。
DNA		[N/A]	编码 RNA 指导的核酸酶和 gRNA 的 DNA 或 DNA 载体
DNA		DNA	编码 RNA 指导的核酸酶和 gRNA 的第一 DNA 或 DNA 载体，以及编码供体模板的第二 DNA 或 DNA 载体。
DNA	DNA		编码 RNA 指导的核酸酶的第一 DNA 或 DNA 载体以及编码 gRNA 和供体模板的第二 DNA 或 DNA 载体。
DNA DNA			编码 RNA 指导的核酸酶和供体模板的第一 DNA 或 DNA 载体，以及编码 gRNA 的第二 DNA 或 DNA 载体
DNA RNA			编码 RNA 指导的核酸酶和供体模板的 DNA 或 DNA 载体，以及 gRNA
RNA		[N/A]	RNA 或 RNA 载体，其编码 RNA 指导的核酸酶并包含 gRNA
RNA		DNA	编码 RNA 指导的核酸酶并包含 gRNA 的 RNA 或 RNA 载体，以及编码供体模板的 DNA 或 DNA 载体。

[0331] 表4总结了用于如本文所述的基因组编辑系统的组分的各种递送方法。同样，该列表旨在是示例性而不是限制性的。

表4

递送载体/模式		递送到未分裂细胞中	表达的持续时间	基因组整合	所递送的分子类型
物理的（例如，电穿孔、基因枪、磷酸钙转染、细胞压缩或挤压）		是	瞬时的	NO	核酸和蛋白质
病毒的	逆转录病毒	NO	稳定的	是	RNA
	慢病毒	是	稳定的	是/否具有修饰	RNA
	腺病毒	是	瞬时的	NO	DNA
	腺相关病毒（AAV）	是	稳定的	NO	DNA
	痘苗病毒	是	非常瞬时的	NO	DNA
	单纯疱疹病毒	是	稳定的	NO	DNA
非病毒的	阳离子脂质体	是	瞬时的	取决于递送什么	核酸和蛋白质
	聚合物纳米颗粒	是	瞬时的	取决于递送什么	核酸和蛋白质
生物非病毒递送媒介物	减毒细菌	是	瞬时的	NO	核酸
	工程化的噬菌体	是	瞬时的	NO	核酸
	哺乳动物病毒样颗粒	是	瞬时的	NO	核酸
	生物脂质体：红细胞血影和外泌体	是	瞬时的	NO	核酸

基因组编辑系统的基于核酸的递送

[0332] 可以通过本领域已知的方法或如本文所述将编码根据本公开的基因组编辑系统的各种元件的核酸施用于受试者或递送至细胞。例如，可以通过例如载体(例如，病毒或非病毒载体)、非载体基方法(例如，使用裸DNA或DNA复合物)或其组合递送编码RNA指导的核酸酶的DNA和/或编码gRNA的DNA、以及供体模板核酸。

[0333] 编码基因组编辑系统或其组分的核酸可以作为裸DNA或RNA直接递送至细胞，例如借助转染或电穿孔来递送，或者可以缀合至促进靶细胞(例如，红血球、HSC)摄取的分子(例如，N-乙酰半乳糖胺)。也可以使用核酸载体，如表4中总结的载体。

[0334] 核酸载体可以包含编码基因组编辑系统组分(例如RNA指导的核酸酶、gRNA和/或供体模板)的一个或多个序列。载体还可以包含编码信号肽(例如，用于核定位、核仁定位或线粒体定位)的序列，其与编码蛋白质的序列缔合(例如插入其中或与其融合)。作为一个实

例,核酸载体可以包括Cas9编码序列,其包括一个或多个核定位序列(例如,来自SV40的核定位序列)。

[0335] 核酸载体还可以包括任何适宜数目的调节/控制元件,例如,启动子、增强子、内含子、多聚腺苷酸化信号、科扎克(Kozak)共有序列或内部核糖体进入位点(IRES)。这些元件为本领域中所熟知,并且描述于Cotta-Ramusino中。

[0336] 根据本公开的核酸载体包括重组病毒载体。示例性病毒载体展示于表4中,并且其他适宜病毒载体以及其使用和产生描述于Cotta-Ramusino中。还可以使用本领域已知的其他病毒载体。另外,可以使用病毒颗粒来递送呈核酸和/或肽形式的基因组编辑系统组分。例如,可以组装"空"病毒颗粒以含有任何适宜负荷。病毒载体和病毒颗粒也可以经工程改造以并入靶向配体,从而改变靶组织特异性。

[0337] 除了病毒载体以外,可以使用非病毒载体来递送编码根据本公开的基因组编辑系统的核酸。非病毒核酸载体的一个重要分类是纳米颗粒,其可以是有机的或无机的。纳米颗粒为本领域所熟知,并且概述于Cotta-Ramusino中。可以使用任何适宜的纳米颗粒设计来递送基因组编辑系统组分或编码此类组分的核酸。例如,在本公开的某些实施例中,有机(例如脂质和/或聚合物)纳米颗粒可以适合用作递送媒介物。用于纳米颗粒配制品和/或基因转移的示例性脂质显示于表5中,并且表6列示用于基因转移和/或纳米颗粒配制品的示例性聚合物。

表5:用于基因转移的脂质

脂质	缩写	特征
1,2-二油酰基-sn-甘油基-3-磷脂酰胆碱	DOPC	辅助
1,2-二油酰基-sn-甘油基-3-磷脂酰乙醇胺	DOPE	辅助
胆固醇		辅助

N-[1-(2,3-二油烯基氧基)丙基]*N,N,N*-三甲基氯化铵	DOTMA	阳离子的
1,2-二油烯基氧基-3-三甲基铵-丙烷	DOTAP	阳离子的
二十八烷基酰胺基甘氨酰精胺	DOGS	阳离子的
N-(3-氨基丙基)-*N,N*-二甲基-2,3-双(十二烷氧基)-1-溴化丙铵	GAP-DLRIE	阳离子的
溴化十六烷基三甲基铵	CTAB	阳离子的
6-月桂氧己基鸟氨酸	LHON	阳离子的
1-(2,3-二油酰基氧基丙基)-2,4,6-三甲基吡啶鎓	2Oc	阳离子的
2,3-二油烯基氧基-*N*-[2(精胺甲酰胺基-乙基]-*N,N*-二甲基-1-三氟乙酸丙铵	DOSPA	阳离子的
1,2-二油烯基-3-三甲基铵-丙烷	DOPA	阳离子的
N-(2-羟基乙基)-*N,N*-二甲基-2,3-双(十四烷氧基)-1-溴化丙铵	MDRIE	阳离子的
二肉豆蔻酰氧基丙基二甲基羟基乙基溴化铵	DMRI	阳离子的
3β-[*N*-(*N'*,*N'*-二甲基氨基乙烷)-氨甲酰基]胆固醇	DC-Chol	阳离子的
双-胍鎓-三氨乙基胺-胆固醇	BGTC	阳离子的
1,3-双脱氧基-2-(6-羧基-精胺基)-丙酰胺	DOSPER	阳离子的
二甲基十八烷基溴化铵	DDAB	阳离子的
双十八烷基氨基甘氨酰基亚精胺	DSL	阳离子的
外消旋-[(2,3-双十八烷氧基丙基)(2-羟基乙基)]-二甲基氯化铵	CLIP-1	阳离子的
外消旋-[2(2,3-双十六烷氧基丙基-氧基甲氧基)乙基]三甲基溴化铵	CLIP-6	阳离子的
乙基二肉豆蔻酰磷脂酰胆碱	EDMPC	阳离子的
1,2-二硬脂氧基-*N,N*-二甲基-3-氨基丙烷	DSDMA	阳离子的
1,2-二肉豆蔻酰基-三甲基铵丙烷	DMTAP	阳离

		子的
O,O'-二肉豆蔻基-N-赖氨酰基天冬氨酸酯	DMKE	阳离子的
1,2-二硬脂酰基-sn-甘油基-3-乙基磷酸胆碱	DSEPC	阳离子的
N-棕榈酰 D-赤-鞘氨苷氨甲酰基-精胺	CCS	阳离子的
N-叔丁基-*N*0-十四烷基-3-十四烷基氨基丙脒	二 C14-脒	阳离子的
十八碳烯醇基氧基[乙基-2-十七烯基-3 羟基乙基]氯化咪唑啉鎓	DOTIM	阳离子的
*N*1-胆固醇氧基羰基-3,7-二氮杂壬烷-1,9-二胺	CDAN	阳离子的
2-(3-[双(3-氨基-丙基)-氨基]丙胺基)-*N*-双十四烷基氨基甲酰基甲-乙基-乙酰胺	RPR209120	阳离子的
1,2-二亚油基氧基-3-二甲基氨基丙烷	DLinDMA	阳离子的
2,2-二亚油基-4-二甲基氨基乙基-[1,3]-二氧戊环	DLin-KC2-DMA	阳离子的
二亚油基-甲基-4-二甲基氨基丁酸酯	DLin-MC3-DMA	阳离子的

表6:用于基因转移的聚合物

聚合物	缩写
聚乙二醇	PEG
聚乙烯亚胺	PEI
二硫代双(琥珀酰亚胺基丙酸酯)	DSP
二甲基-3,3'-二硫代双丙亚氨酸酯	DTBP
聚(乙烯亚胺)双氨基甲酸酯	PEIC
聚(L-赖氨酸)	PLL
组氨酸修饰的 PLL	
聚(*N*-乙烯基吡咯烷酮)	PVP
聚(丙烯亚胺)	PPI
聚(酰胺基胺)	PAMAM
聚(酰胺基乙烯亚胺)	SS-PAEI
三亚乙基四胺	TETA
聚(β-氨基酯)	
聚(4-羟基-L-脯氨酸酯)	PHP
聚(烯丙基胺)	
聚(α-[4-氨基丁基]-L-乙醇酸)	PAGA

聚(D,L-乳酸-共-乙醇酸)	PLGA
聚(*N*-乙基-4-乙烯基溴化吡啶鎓)	
聚(磷腈)	PPZ
聚(磷酸酯)	PPE
聚(氨基磷酸酯)	PPA
聚(*N*-2-羟基丙基甲基丙烯酰胺)	pHPMA
聚(2-(二甲基氨基)甲基丙烯酸乙酯)	pDMAEMA
聚(2-氨乙基丙烯磷酸酯)	PPE-EA
壳聚糖	
半乳糖基化壳聚糖	
N-十二烷基化的壳聚糖	
组蛋白	
胶原	
葡聚糖-精胺	D-SPM

[0338] 非病毒载体任选地包括靶向修饰以改进摄取和/或选择性靶向某些细胞类型。这些靶向修饰可以包括例如细胞特异性抗原、单克隆抗体、单链抗体、适体、聚合物、糖(例如,N-乙酰半乳糖胺(GalNAc))和细胞穿透肽。此类载体还任选地使用致融性和核内体去稳定肽/聚合物,经历酸触发的构象变化(例如,加速负荷的核内体逃逸),和/或并入刺激可裂解的聚合物,例如用于在细胞区室中释放。例如,可以使用在还原性细胞环境中裂解的基于二硫化物的阳离子型聚合物。

[0339] 在某些实施例中,除了基因组编辑系统的组分(例如,本文所述的RNA指导的核酸酶组分和/或gRNA组分)以外,递送一种或多种核酸分子(例如,DNA分子)。在某些实施例中,该核酸分子是与基因组编辑系统的一个或多个组分同时递送。在某些实施例中,该核酸分子是在递送基因组编辑系统的一个或多个组分之前或之后(例如,小于约30分钟、1小时、2小时、3小时、6小时、9小时、12小时、1天、2天、3天、1周、2周或4周)递送。在某些实施例中,该核酸分子是通过与递送基因组编辑系统的一个或多个组分(例如,RNA指导的核酸酶组分和/或gRNA组分)不同的方式来递送。该核酸分子可以通过本文所述的任何递送方法来递送。例如,核酸分子可以通过病毒载体(例如,整合缺陷型慢病毒)来递送,并且RNA指导的核酸酶分子组分和/或gRNA组分可以通过电穿孔来递送,例如,使得可以降低由核酸(例如,DNA)引起的毒性。在某些实施例中,该核酸分子编码治疗性蛋白质,例如,本文所述的蛋白质。在某些实施例中,该核酸分子编码RNA分子,例如,本文所述的RNA分子。

RNP和/或编码基因组编辑系统组分的RNA的递送

[0340] 可以通过本领域已知的方法将RNP(gRNA与RNA指导的核酸酶的复合物)和/或编码RNA指导的核酸酶和/或gRNA的RNA递送至细胞中或施用给受试者,其中一些方法描述于Cotta-Ramusino中。在体外,编码RNA指导的核酸酶的和/或编码gRNA的RNA可以通过例如显微注射、电穿孔、瞬时细胞压缩或挤压来递送(参见,例如,Lee 2012)。还可以使用脂质介导的转染、肽介导的递送、GalNAc或其他缀合物介导的递送和其组合进行体外和体内递送。保护性、交互式、非冷凝(PINC)系统可用于递送。

[0341] 在体外,通过电穿孔递送包含将细胞与编码RNA指导的核酸酶和/或gRNA的RNA(具有或不具有供体模板核酸分子)在盒、室或比色皿中混合,并且施加一个或多个限定持续时间和幅度的电脉冲。用于电穿孔的系统和方案为本领域中已知,并且任何适宜的电穿孔工

具和/或方案可以结合本公开的各个实施例使用。

施用途径

[0342] 可以通过任何适宜模式或途径(局部或全身)将基因组编辑系统或使用此类系统改变或操作的细胞施用于受试者。全身施用模式包括口服和肠胃外途径。肠胃外途径包括例如静脉内、骨髓内、动脉内、肌内、真皮内、皮下、鼻内和腹膜内途径。全身施用的组分可以经修饰或配制以靶向例如HSC(造血干细胞/祖细胞)或红系祖细胞或前体细胞。

[0343] 局部施用模式包括例如骨髓内注射至骨小梁中或股骨内注射到髓隙中,以及输注至门静脉中。在某些实施例中,与全身施用(例如,静脉内)时相比,在局部施用(例如,直接施用至骨髓中)时,显著较少量的组分(与全身方法相比)可以发挥作用。局部施用模式可以降低或消除在全身性施用治疗有效量的组分时可能发生的潜在毒副作用的发生率。

[0344] 施用可以作为定期推注(例如静脉内)或作为持续输注从内部储库或从外部储库(例如从静脉内注射袋或可植入泵)提供。组分可以局部施用,例如,通过从持续释放药物递送装置中连续释放。

[0345] 另外,组分可以经配制以容许经延长时段释放。释放系统可以包括生物可降解材料或通过扩散释放所并入组分的材料的基质。所述组分可以在释放系统中均匀或者非均匀分配。多种释放系统可以是有用的,但适当系统的选择将取决于具体应用所需要的释放速率。不可降解和可降解的释放系统均可以被使用。适宜释放系统包括聚合物和聚合基质、非聚合基质或无机和有机赋形剂和稀释剂,例如但不限于碳酸钙和糖(例如,海藻糖)。释放系统可以是天然的或合成的。然而,合成释放系统是优选的,因为其通常更可靠、更具可重现性并且产生更明确的释放曲线。可以选择释放系统材料,使得具有不同分子量的组分是通过穿过材料扩散或通过材料降解而释放。

[0346] 代表性的合成生物可降解聚合物包括例如:聚酰胺,例如聚(氨基酸)和聚(肽);聚酯,例如聚(乳酸)、聚(乙醇酸)、聚(乳酸-共-乙醇酸)和聚(己内酯);聚(酸酐);聚原酸酯;聚碳酸酯;和其化学衍生物(取代、添加化学基团,例如,烷基、亚烷基、羟化、氧化和本领域技术人员常规进行的其他修饰)、共聚物和其混合物。代表性的合成非生物可降解聚合物包括例如:聚醚,例如聚(氧化乙烯)、聚(乙二醇)和聚(环氧丁烷);乙烯基聚合物-聚丙烯酸酯和聚甲基丙烯酸酯,例如甲基、乙基、其他烷基、甲基丙烯酸羟乙酯、丙烯酸和甲基丙烯酸和其他,例如聚(乙烯醇)、聚(乙烯基吡咯烷酮)和聚(乙酸乙烯酯);聚(氨酯);纤维素和其衍生物,例如烷基、羟烷基、醚、酯、硝化纤维素和各种醋酸纤维素;聚硅氧烷;和其任何化学衍生物(取代、添加化学基团,例如,烷基、亚烷基、羟化、氧化和本领域技术人员常规进行的其他修饰)、共聚物和其混合物。

[0347] 也可以使用聚(丙交酯共乙交酯)微球。典型地,微球是由乳酸和乙醇酸的聚合物构成,其经结构化以形成空心球体。球体的直径可以是约15-30微米,并且可以加载本文所述的组分。在一些实施例中,基因组编辑系统、系统组分和/或编码系统组分的核酸与嵌段共聚物例如泊洛沙姆或泊洛沙明一起递送。

组分的多模或差别递送

[0348] 技术人员根据本公开将了解,本文所公开的基因组编辑系统的不同组分可以一起或分开并且同时或不同时递送。可能尤其期望基因组编辑系统组分的分开和/或异步递送以提供对基因组编辑系统功能的时间或空间控制并限制由其活性引起的某些效应。

[0349] 如本文所用,不同或差别模式是指递送模式,这些递送模式赋予受试组分分子(例如,RNA指导的核酸酶分子、gRNA、模板核酸或有效负载)不同的药效学或药物代谢动力学特性。例如,递送模式可以导致不同的组织分布,不同的半衰期或不同的时间分布,例如,在所选区室、组织或器官中。

[0350] 一些递送的模式(例如,通过例如通过自主复制或插入细胞核酸中而在细胞或细胞后代中持续存在的核酸载体来递送)导致组分更持久的表达和存在。实例包括病毒(例如,AAV或慢病毒)递送。

[0351] 举例来说,基因组编辑系统的组分(例如,RNA指导的核酸酶和gRNA)可以通过在所递送组分在体内或在特定区室、组织或器官中的所得半衰期或持久性方面不同的模式进行递送。在某些实施例中,gRNA可以通过此类模式进行递送。RNA指导的核酸酶分子组分可以通过导致在身体或特定区室或组织或器官中更低持久性或更少暴露的模式进行递送。

[0352] 更通常地,在某些实施例中,使用第一递送模式来递送第一组分,并且使用第二递送模式来递送第二组分。第一递送模式赋予第一药效学或药物代谢动力学特性。第一药效学特性可以是例如组分或编码该组分的核酸在体内、区室、组织或器官中的分布、持久性或暴露。第二递送模式赋予第二药效学或药物代谢动力学特性。第二药效学特性可以是例如组分或编码该组分的核酸在体内、区室、组织或器官中的分布、持久性或暴露。

[0353] 在某些实施例中,第一药效学或药物代谢动力学特性(例如,分布、持久性或暴露)比第二药效学或药物代谢动力学特性更受限。

[0354] 在某些实施例中,第一递送模式经选择以优化(例如,最小化)药效学或药物代谢动力学特性(例如,分布、持久性或暴露)。

[0355] 在某些实施例中,第二递送模式经选择以优化(例如,最大化)药效学或药物代谢动力学特性(例如,分布、持久性或暴露)。

[0356] 在某些实施例中,第一递送模式包含使用相对持久的元件,例如,核酸,例如,质粒或病毒载体,例如,AAV或慢病毒。由于此类载体相对持久,从其转录的产物将相对持久。

[0357] 在某些实施例中,第二递送模式包含相对瞬时元件,例如,RNA或蛋白质。

[0358] 在某些实施例中,第一组分包含gRNA,并且递送模式相对持久,例如,gRNA是从质粒或病毒载体(例如,AAV或慢病毒)转录。这些基因的转录将具有极小生理学意义,因为这些基因不编码蛋白质产物,并且这些gRNA不能够单独起作用。第二组分(RNA指导的核酸酶分子)是以瞬时方式递送,例如,作为mRNA或作为蛋白质递送,从而确保完整的RNA指导的核酸酶分子/gRNA复合物仅在短时段内存在并有活性。

[0359] 此外,这些组分可以不同的分子形式或用不同的互为补充以增强安全性和组织特异性的递送载体进行递送。

[0360] 差别递送模式的使用可以增强性能、安全性和/或功效,例如,可以减少最终脱靶修饰的可能性。通过较不持久的模式递送免疫原性组分(例如,Cas9分子)可以降低免疫原性,因为来自细菌源Cas酶的肽通过MHC分子展示于细胞表面上。两部分式递送系统可以改善这些缺点。

[0361] 差别递送模式可以用于将组分递送至不同但重叠的靶区域。在靶区域的重叠以外,活性复合物的形成被最小化。因此,在某些实施例中,第一组分(例如,gRNA)是通过第一递送模式进行递送,其导致第一空间(例如,组织)分布。第二组分(例如,RNA指导的核酸酶

分子)是通过第二递送模式进行递送,其导致第二空间(例如,组织)分布。在某些实施例中,第一模式包含选自脂质体、纳米颗粒(例如,聚合纳米颗粒)和核酸的第一元件,例如,病毒载体。第二模式包含选自该组的第二元件。在某些实施例中,第一递送模式包含第一靶向元件,例如,细胞特异性受体或抗体,并且第二递送模式不包括该元件。在某些实施例中,第二递送模式包含第二靶向元件(例如,第二细胞特异性受体或第二抗体)。

[0362] 当在病毒递送载体、脂质体或聚合纳米颗粒中递送RNA指导的核酸酶分子时,存在递送至多个组织并且在多个组织中具有治疗活性的可能性,但此时可能希望仅靶向单一组织。两部分式递送系统可以解决这一挑战并且增强组织特异性。如果将gRNA分子和RNA指导的核酸酶分子包装于具有不同但重叠的组织向性的分开的递送媒介物中,那么完全功能复合物仅在两种载体所靶向的组织中形成。

实例

[0363] 通过以下非限制性实例进一步说明上述原理和实施例:

实例1:作为核糖核蛋白复合物递送至K562细胞用于在HBG1和HBG2调控区域中引起13nt缺失的化脓链球菌gRNA的筛选

[0364] 通过标准方法设计了靶向26nt片段的gRNA,所述片段跨并包括HBG1和HBG2的13nt靶区域处的13个核苷酸。在电脑模拟和分级后设计gRNA后,选择一部分gRNA并筛选人K562细胞的活性和特异性。选择用于筛选的gRNA在表7中示出。简而言之,gRNA先进行体外转录,然后将其与化脓链球菌野生型(Wt)Cas9蛋白进行复合以形成核糖核蛋白复合物(RNP)。与化脓链球菌Cas9蛋白复合的gRNA是经修饰的sgRNA((例如,5'ARCA加帽和3'聚A(20A)尾;表7)并靶向HBG1和HBG2调控区域。为了直接比较这些RNP在K562细胞和人$CD34^+$细胞中的活性,首先将RNP通过电穿孔(Amaxa Nucleofector)递送到K562细胞。

[0365] RNP电穿孔后3天,从K562细胞中提取gDNA,然后从gDNA中PCR扩增HBG1和HBG2基因座。通过T7E1内切核酸酶测定分析在PCR产物中评估基因编辑。九个RNP中有八个支持高百分比的NHEJ。Sp37 RNP是唯一示出于人$CD34^+$细胞中有活性的gRNA(在$CD34^+$细胞中<10%编辑),在K562细胞中具有高活性,其中在HBG1和HBG2处均检测到>60%的indel并且在启动子序列中的HBG1和HBG2靶向区域中都有八个剪切(图3A)。

表7:用于在K562细胞或$CD34^+$细胞中筛选的选择的gRNA

gRNA ID	靶向结构域序列(RNA)	靶向结构域序列(DNA)	靶向结构域序列加上PAM(NGG)(RNA)	靶向结构域序列加上PAM(NGG)(DNA)	正义
Sp9	GGCUAUUGGUCA	GGCTATTGGTC	GGCUAUUGGUCA	GGCTATTGGTCA	反义

	AGGCA (SEQ ID NO: 277)	AAGGCA（SEQ ID NO: 910）	AGGCAAGG(SEQ ID NO: 920)	AGGCAAGG (SEQ ID NO: 930)	
Sp36	CAAGGCUAUUGG UCAAGGCA（SEQ ID NO: 338）	CAAGGCTATTG GTCAAGGCA (SEQ ID NO: 911)	CAAGGCUAUUGG UCAAGGCAAGG (SEQ ID NO: 921)	CAAGGCTATTGG TCAAGGCAAGG (SEQ ID NO: 931)	反义
Sp40	UGCCUUGUCAAG GCUAU (SEQ ID NO: 327)	TGCCTTGTCAA GGCTAT（SEQ ID NO: 912）	UGCCUUGUCAAG GCUAUUGG (SEQ ID NO: 922)	TGCCTTGTCAAG GCTATTGG (SEQ ID NO: 932)	反义
Sp42	GUUUGCCUUGUC AAGGCUAU（SEQ ID NO: 299）	GTTTGCCTTGT CAAGGCTAT (SEQ ID NO: 913)	GUUUGCCUUGUC AAGGCUAUUGG (SEQ ID NO: 923)	GTTTGCCTTGTC AAGGCTATTGG (SEQ ID NO: 933)	反义
Sp38	GACCAAUAGCCU UGACA (SEQ ID NO: 276)	GACCAATAGCC TTGACA（SEQ ID NO: 914）	GACCAAUAGCCU UGACAAGG (SEQ ID NO: 924)	GACCAATAGCCT TGACAAGG (SEQ ID NO: 934)	正义
Sp37	CUUGACCAAUAG CCUUGACA（SEQ ID NO: 333）	CTTGACCAATA GCCTTGACA (SEQ ID NO: 915)	CUUGACCAAUAG CCUUGACAAGG(S EQ ID NO: 925)	CTTGACCAATAG CCTTGACAAGG (SEQ ID NO: 935)	正义
Sp43	GUCAAGGCUAUU GGUCA (SEQ ID NO: 278)	GTCAAGGCTAT TGGTCA (SEQ ID NO: 916)	GUCAAGGCUAUU GGUCAAGG (SEQ ID NO: 926)	GTCAAGGCTATT GGTCAAGG (SEQ ID NO: 936)	反义
Sp35	CUUGUCAAGGCU AUUGGUCA（SEQ ID NO: 339）	CTTGTCAAGG CTATTGGTCA (SEQ ID NO: 917)	CUUGUCAAGGCU AUUGGUCAAGG (SEQ ID NO: 927)	CTTGTCAAGGCT ATTGGTCAAGG (SEQ ID NO: 937)	反义
Sp41	UCAAGUUUGCCU UGUCA (SEQ ID NO: 310)	TCAAGTTTGCC TTGTCA（SEQ ID NO: 918）	UCAAGUUUGCCU UGUCAAGG (SEQ ID NO: 928)	TCAAGTTTGCCT TGTCAAGG (SEQ ID NO: 938)	反义
Sp34	UGGUCAAGUUUG CCUUGUCA（SEQ ID NO: 340）	TGGTCAAGTTT GCCTTGTCA(S EQ ID NO: 919)	UGGUCAAGUUUG CCUUGUCAAGG (SEQ ID NO: 929)	TGGTCAAGTTTG CCTTGTCAAGG (SEQ ID NO: 939)	反义
Sp85	AGUAUCCAGUGA GGCCA (SEQ ID NO: 940)	AGTATCCAGTG AGGCCA (SEQ ID NO: 943)	AGUAUCCAGUGA GGCCAGGG (SEQ ID NO: 946)	AGTATCCAGTGA GGCCAGGG (SEQ ID NO: 949)	反义
SpA	GGCAAGGCUGGC CAACCCAU (SEQ ID NO: 941)	GGCAAGGCTG GCCAACCCAT (SEQ ID NO: 944)	GGCAAGGCUGGC CAACCCAUGGG (SEQ ID NO: 947)	GGCAAGGCTGG CCAACCCATGGG (SEQ ID NO: 950)	正义
SpB	UAUUUGCAUUGA GAUAGUGU (SEQ ID NO: 942)	TATTTGCATTG AGATAGTGT (SEQ ID NO: 945)	UAUUUGCAUUGA GAUAGUGUGGG(S EQ ID NO: 948)	TATTTGCATTGAG ATAGTGTGGG (SEQ ID NO: 951)	正义

[0366] 然后通过DNA测序分析来分析用八种活性sgRNA靶向的K562细胞的HBG1和HBG2 PCR产物，并对检测到的插入和缺失进行评分。缺失被细分为靶位点的精确的13nt缺失、包括13nt靶位点和近端小缺失(18-26nt)、13nt靶位点的12nt缺失(即，部分缺失)、跨越部分HPFH靶位点的>26nt缺失、和其他缺失，例如，邻近HPFH靶位点但在HPFH靶位点外的缺失。八种sgRNA中的七种针对HBG1靶向13nt的缺失(HPFH突变诱导)(图3B)。八种sgRNA中的至少五种也支持HBG2启动子区域中13nt的靶向缺失(图3C)。需注意，用HBG Sp34 sgRNA处理的细胞中HBG2的DNA序列结果不可用。这些数据表明，Cas9和sgRNA支持13nt缺失的精确诱导。图3B-3C描绘了在HBG1中的靶序列中观察到的缺失类型的实例。

实例2:含有靶向13nt缺失突变的gRNA的Cas9 RNP支持人造血干/祖细胞中的基因编辑

[0367] 在人脐带血(CB)CD34$^+$细胞中测试的含有不同gRNA的RNP中，只有Sp37在HBG1和

HBG2启动子的靶位点处导致可检测的编辑，如通过HBG1和HBG2特异性PCR产物(从三个脐带血供体的电穿孔CB CD34$^+$细胞中提取的gDNA扩增产物)中的indel的T7E1分析所确定的(图4A)。在用与Sp37复合的Cas9蛋白电穿孔的细胞中检测到的平均编辑水平在HBG1处为5%±2%indel，在HBG2处检测为3%±1%indel(3个独立的实验和CB供体)。

[0368] 接下来，将三个化脓链球菌gRNA(其靶位点在HBG启动子内)(Sp35、Sp36、Sp37)与野生型化脓链球菌Cas9蛋白复合以形成核糖核蛋白复合物。将这些HBG靶向的RNPS电穿孔到CB CD34$^+$细胞(n=3供体)和成人动员的外周血(mPB)CD34$^+$细胞供体(n=3供体)中。然后，在Cas9 RNP递送后大约三天，通过从样品中提取的基因组DNA扩增的HBG2PCR产物的T7E1内切核酸酶分析来分析靶位点处的插入/缺失水平。这些RNP中的每一个仅在3个供体和3个独立的实验的CB和成人CD34$^+$细胞中支持低水平基因编辑(图4B)。

[0369] 为了增加基因编辑并增加靶位点13nt缺失的发生，产生单链脱氧核苷酸供体修复模板(ssODN)，其在靶向缺失位点的每一侧编码87nt和89nt的同源性。ssODN，在末端未经修饰(即，ssODN1，SEQ ID NO:906，表8)，或在5'和3'末端经修饰为含有硫代磷酸酯(PhTx)(即，PhTx ssODN1，SEQ ID NO:909，表8)。ssODN设计为'编码'13nt缺失，其中序列同源臂被工程改造侧翼于缺失序列的两侧，从而形成完全缺失。

表8:单链脱氧核苷酸供体修复模板(ssODN)

ssODN ID	SEQ ID NO	序列
ssODN1	904	**GGGTGCTTCCTTTTATTCTTCATCCCTAGCCAGCCGCCGGCCCCTG**
5'同源臂		**GCCTCACTGGATACTCTAAGACTATTGGTCAAGTTTGCCTT**
ssODN1 3'同源臂	905	GTCAAGGCAAGGCTGGCCAACCCATGGGTGGAGTTTAGCCAGGGACCG TTTCAGACAGATATTTGCATTGAGATAGTGTGGGGAAGGGG
ssODN1	906	**GGGTGCTTCCTTTTATTCTTCATCCCTAGCCAGCCGCCGGCCCCTG GCCTCACTGGATACTCTAAGACTATTGGTCAAGTTTGCCTT**GTCAAG GCAAGGCTGGCCAACCCATGGGTGGAGTTTAGCCAGGGACCGTTTCAG ACAGATATTTGCATTGAGATAGTGTGGGGAAGGGG
PhTx ssODN1 5'同源臂	907	**G*GGTGCTTCCTTTTATTCTTCATCCCTAGCCAGCCGCCGGCCCCT GGCCTCACTGGATACTCTAAGACTATTGGTCAAGTTTGCCTT**
PhTx ssODN1 3'同源臂	908	GTCAAGGCAAGGCTGGCCAACCCATGGGTGGAGTTTAGCCAGGGACCG TTTCAGACAGATATTTGCATTGAGATAGTGTGGGGAAGGG*G
PhTx ssODN1	909	**G*GGTGCTTCCTTTTATTCTTCATCCCTAGCCAGCCGCCGGCCCCT GGCCTCACTGGATACTCTAAGACTATTGGTCAAGTTTGCCTT**GTCAA GGCAAGGCTGGCCAACCCATGGGTGGAGTTTAGCCAGGGACCGTTTCA GACAGATATTTGCATTGAGATAGTGTGGGGAAGGG*G
在缺失侧翼的同源臂用粗体[5'同源臂]和下划线[3'同源臂]指示。请注意，ssODN1 和 PhTx ssODN1 中不存在 13 bp 序列。 *表示被硫代磷酸酯修饰。		

[0370] ssODN1和PhTx ssODN1与靶向HBG的RNP共递送至CB CD34$^+$细胞，所述HBG含有Sp37 gRNA(HBG Sp37 RNP)或HBG Sp35(HBG Sp35 RNP)。如通过HBG2 PCR产物的T7E1分析确定，编码13nt缺失的ssODN供体与HBG Sp35 RNP或HBG Sp37 RNP共递送导致分别在靶位点的基因编辑中有6倍和5倍的增加(图4C)。HBG2 PCR产物的DNA测序分析(桑格测序)指示了在用HBG Sp37 RNP和PhTx修饰的ssODN1处理的细胞中20%的基因编辑，具有15%的缺失和5%的插入(图4C，左下小图)。对靶位点缺失的特异性类型和尺寸的进一步分析揭示，检测到的

总缺失的75％中含有13nt缺失(其中包括近端启动子中的CAAT盒的c.-110处的缺失)，所述缺失与HbF表达的升高有关(图4C，右下小图)。剩余的1/4缺失是部分缺失，没有跨越完全的13nt缺失。这些数据指示，工程改造具有缺失的同源ssODN的共递送支持在人$CD34^+$细胞中HBG的精确基因编辑(缺失)。

实例3:靶向13nt缺失突变的Cas9 RNP支持成人的动员的具有成红细胞后裔中HBG表达增加的外周血造血干细胞/祖细胞中的基因编辑。

[0371] 为了确定在HBG启动子中用与Sp37 gRNA或Sp35 gRNA(即，靶向与HPFH有关的13nt缺失的gRNA)复合Cas9 RNP编辑HBG支持编辑的$CD34^+$细胞的红系后裔中HBG表达的增加，用RNP电穿孔来自动员外周血(mPB)的成人$CD34^+$细胞。简而言之，在StemSpan SFEM中用人细胞因子和PGE2预刺激mPB $CD34^+$细胞2天，然后分别用预复合到Sp35和Sp37的Cas9蛋白电穿孔。HBG PCR产物的T7E1分析指示，用与Sp37复合的RNP处理的mPB $CD34^+$细胞检测到约3％的indel，而未检测到用与Sp35复合的RNP处理的细胞的编辑(图5A)。

[0372] 为了增加靶位点的基因编辑并增加靶位点处13nt缺失的发生，将PhTx ssODN1(SEQ ID NO:909)与靶向HBG的含有Sp37 gRNA的预复合RNP共递送。编码13nt缺失的ssODN供体的共递送导致靶位点的基因编辑增加近2倍(图5A)。为了确定编辑HBG是否增加编辑的成人$CD34^+$细胞的红系后裔中胎儿血红蛋白的产生，在人细胞因子(促红细胞生成素，SCF，IL3)、人血浆(Octoplas)、和其他补充剂(氢化可的松、肝素、转铁蛋白)存在下，通过培养持续长达18天将细胞分化成成红细胞。在分化的时间过程中，收集mRNA以评估RNP处理的mPB $CD34^+$细胞和供体匹配的阴性(未处理的)对照的红系后裔中的HBG基因表达。到分化的第七天，用HBG Sp37 RNP处理的人$CD34^+$细胞的成红细胞后裔并且编码13nt缺失的ssODN(通过T7E1分析在来自混合细胞群的gDNA中检测到约5％的indel)展现出HBG mRNA的产生增加2倍(图5B)。重要的是，用HBG RNP电穿孔的$CD34^+$细胞维持其离体造血活性(即，与未处理的供体匹配的$CD34^+$细胞阴性对照相比，红系和骨髓集落的量或多样性没有差异)，如在造血集落形成细胞(CFC)测定中确定的(图6A)。此外，通过用于获得红系表型(％血型糖蛋白A^+细胞)的流式分析确定，从RNP处理的$CD34^+$细胞分化的成红细胞维持对于供体匹配的未处理对照细胞观察到的分化动力学(图6B)。这些数据指示HBG1/HBG2近端启动子区域的靶向破坏支持RNP处理的成人造血干细胞/祖细胞的红系后裔中HBG表达的增加而不改变分化潜能。

实例4:靶向HPFH突变的Cas9 RNP支持成人的动员的具有成红细胞后裔中HBG表达增加的外周血造血干细胞/祖细胞中的基因编辑

[0373] 为了确定靶向HBG的成对切口酶RNP的共递送是否会增加对近端HBG启动子的靶向破坏，将mPB $CD34^+$细胞与人细胞因子和PGE2在StemSpan SFEM中培养2天，然后用与两种gRNA(其靶向13nt缺失位点侧翼的位点)预复合的化脓链球菌D10A Cas9蛋白电穿孔。在该实例中，在切口酶对中使用的gRNA的靶向结构域序列(包括但不限于SpA、Sp85和SpB)示于表7中。选择D10A切口酶对，以使对于靶的PAM朝外定向，并且剪切位点之间的距离<100nt。gRNA与D10A Cas9蛋白复合形成RNP复合物，然后使人$CD34^+$细胞和成对的切口酶经受电穿孔。为了确定编码13nt缺失的ssODN的共递送是否增加突变的编辑和向细胞中的引入，在一些实验中，在电穿孔之前将ssODN1添加到细胞RNP混合物中。电穿孔后约3天，从RNP处理的细胞中提取gDNA，并通过以下进行分析:T7E1内切核酸酶测定和/或从提取的gDNA扩增得到

的HBG2 PCR产物的桑格DNA测序。在测试的三个D10A切口酶对中，对于一个包括ssODN1的切口酶对(gRNA SpA+Sp85)样品，通过T7E1内切核酸酶分析检测到的indel增加(图7A)。DNA测序分析是对有限的样品进行，如图7A所示。DNA测序分析显示在靶位点的indel高达约27％，插入被检测为主要indel，随后是靶向的区域的缺失(成对切口酶剪切位点之间的区域)，并且当编码缺失的ssODN1被共递送时也以2％-3％的频率检测到13nt缺失突变(图7B)。沉默的非致病性的SNP包括在ssODN1供体模板中，并在含有13nt缺失的序列中检测到，表明HFPH突变的产生是通过HDR事件发生的。

实例5:电穿孔进入成人CD34+细胞的D10A成对RNP支持在红系后代中诱导HbF蛋白。

[0374] 为了进一步优化在mPB $CD34^+$细胞中靶位点处的编辑条件并评估在另外的人细胞供体中的编辑，将人mPB $CD34^+$细胞用靶向HBG的D10A Cas9和WT Cas9成对RNP电穿孔。对于D10A Cas9和WT Cas9 RNP而言，最有效的指导对是Sp37+SpA，通过HBG2 PCR产物的T7E1内切核酸酶分析确定，其支持大于30％indel(图8A)。鉴于在HBG1和HBG2两者上进行编辑均可能导致HBG2的大缺失和HBG2与HBG1之间的基因间区域的大缺失，因此对indel进行了进一步表征，以便通过T7E1内切核酸酶分析和测序捕获局部indel，并通过ddPCR分析捕获大缺失。对于D10A Cas9和WT Cas9 RNP切口酶对，在所有样品中均以可变频率检测到大的缺失(图8B)。与通过T7E1分析确定的indel相关的indel的Illumina测序分析(图8C-8D)。

[0375] 为了确定在HBG启动子处用双切口酶编辑的$CD34^+$细胞是否产生了具有升高的HbF表达的红系后代，将供体匹配的RNP处理和未经处理的对照诱导向红系分化，并且然后评估在分化过程中indel的维持和HbF mRNA与蛋白质的表达。在去核之前，在RNP处理的细胞后代中，在红系分化的前2周中评估编辑水平(通过T7E1内切核酸酶测定确定)。在每个测定的时间点在红系后代中检测到indel，这表明$CD34^+$细胞中发生的编辑在红系分化过程中维持，并且经编辑的$CD34^+$细胞维持了红系分化潜力。

[0376] 分别通过ddPCR和HPLC分析(根据Chang 2017中第143-44页描述的HPLC方法，通过引用并入本文)来定量HBG mRNA(分化的第10天)和HbF蛋白(分化的第20-23天)的水平(图9)。观察到HBG:HBA比率(数据未显示)和HbF/HbF+HbA比率(即，HBG mRNA/HGB+HBB mRNA)约增加了2倍(与未经编辑的供体匹配对照相比HBG转录本增加+40％)，增加到比供体匹配的未处理对照样品中检测到的水平高30％。

[0377] 对于D10A Cas9切口酶对，在红系后代中检测到HbF mRNA和蛋白的上调(图9)。关于HbF蛋白质分析，对于两个D10A切口酶对，两个对支持20％HbF诱导。在WT Cas9 RNP处理的$CD34^+$细胞的红系后代中未检测到HbF上调(数据未显示)。

实例6:增加RNP的剂量提高HBG基因座处的成人$CD34^+$细胞的总编辑效率。

[0378] 对于切口酶对SpA+Sp85，D10A Cas9 RNP浓度增加(标准浓度2.5μM和3.7μM)，并通过电穿孔递送至mPB $CD34^+$细胞。如通过CD34+细胞电穿孔后3天提取的gDNA扩增的HBG PCR产物的T7E1内切核酸酶分析确定的，RNP浓度增加支持HBG靶位点的indel增加至>30％(图10A)。测序分析表明，增加RNP浓度会增加插入(图10B)。RNP处理的$CD34^+$细胞的红系细胞后代的HbF蛋白产生也增加(图10C)。重要的是，在编辑后维持了造血集落的形成潜力(图10D)。然后将这些细胞移植到免疫缺陷小鼠中，并在移植后1个月(图10E)和2个月(图10F)通过对外周血取样并测量人$CD45^+$细胞的百分比来评估其植入。早期的植入数据表明，在供体匹配的未治疗对照(0μM RNP)的接受者队列与移植有RNP处理的细胞的小鼠之间在植入

方面无差异。此外，在所指示的时间点在队列之间在人CD45+级分中的人血谱系分布(髓样、B细胞、T细胞)没有差异(图10G-H)。

[0379] 在成人mPB CD34+细胞(Sp37+SpA、Sp37+SpB)的RNP剂量应答研究中，还测试了另外的两个D10A切口酶对。在这里，将mPB CD34+细胞用D10A成对切口酶(以0、2.5和3.75μM的总RNP递送)进行电穿孔。经RNP处理的细胞分化为红系后代，并通过HPLC分析HbF蛋白水平(%HbF/HbF+HbA)。将在CD34+细胞中检测到的indel频率与在红系细胞后代中检测到的HbF水平作图，以使编辑和HbF诱导关联(图11A)。经RNP处理的和未经处理的对照mPB CD34+细胞也分化为集落，以评估离体造血活性。对于RNP处理的和供体匹配的未处理的对照CD34+细胞的后代，维持了集落形成细胞(CFC)活性(图11B)。对于暴露于以不同剂量的不同D10A RNP对的细胞而言，与未经处理的供体匹配的对照CD34+细胞相比，移植后1个月小鼠外周血中人CD45+细胞的百分比没有差异，并且血谱系分布也没有差异(图11C-D)。

实例7:共递送靶向BCL11A的红系细胞特异性增强子的RNP和非特异性(N)单链脱氧核苷酸序列或成对RNP增加人CD34+细胞中的基因编辑并支持在红系后代中胎儿血红蛋白表达的诱导

[0380] 胎儿血红蛋白的表达可以通过转录抑制因子BCL11A的红系细胞特异性表达的靶向破坏来诱导(Canvers 2015)。通过基因编辑策略增加HbF表达的一种潜在策略是多重基因编辑，以引入与HBG近端启动子相关的13nt缺失，以及靶向破坏BCL11A红系细胞特异性增强子中的GATA1结合基序(在BCL11A基因的内含子2的+58DHS区域中)(图12)。为了完成通过多重基因编辑增加HbF表达的多重策略，必须首先将BCL11A红系细胞增强子(BCL11Ae)的破坏的作用确定为单个编辑事件。

[0381] 在该实验中，将CB CD34+细胞用与体外转录的sgRNA(靶向BCL11A基因的内含子2的+58DHS区域中的GATA1基序)(gRNA SpK，表9)复合的化脓链球菌WT Cas9电穿孔(图13A)。为了确定共递送非靶特异性ssODN是否增加对靶序列的编辑，在CB CD34+细胞中，将BCL11Ae RNP与ssODN(其与BCL11Ae靶序列不同源)共递送。从用BCL11Ae RNP处理的CB CD34+细胞提取的gDNA的BCL11A红系细胞增强PCR产物的T7E1分析表明，实现了约5%indel(图13A)。如通过T7E1内切核酸酶分析检测，BCL11Ae RNP与非靶特异性ssODN的共递送使indel增加5倍至20%。Illumina测序分析表明，>90%的编辑具有BCL11A基因的内含子2中+DHS58区域增强子中的GATA1基序的破坏(数据未显示)。为了增加编辑，将人CB CD34+细胞用WT Cas9 RNP(与WT Cas9复合的单gRNA)或用WT Cas9成对RNP(与WT Cas9复合的成对gRNA)进行电穿孔，以使每个对中的剪切位点位于用于切除GATA1基序的靶位点的侧翼(gRNA SpC、SpK、SpM、SpN)(表9)。如通过T7E1内切核酸酶分析确定，两个单gRNA和两个对具有>50%indel(图13B)。

表9:选择靶向BCL11A红系细胞增强子的gRNA序列用于在CD34+细胞中进行筛选

gRNA ID	靶向结构域序列（RNA）	靶向结构域序列（DNA）	靶向结构域序列加上 PAM（NGG）（RNA）	靶向结构域序列加上 PAM（NGG）（DNA）	正义
SpK	CUAACAGUUG CUUUUAUCAC (SEQ ID NO: 952)	CTAACAGTTG CTTTTATCAC (SEQ ID NO: 956)	CUAACAGUUGC UUUUAUCACAG G (SEQ ID NO: 960)	CTAACAGTTGCT TTTATCACAGG (SEQ ID NO: 964)	反义
SpM	GGGCGUGGGU GGGGUAGAAG (SEQ ID NO: 953)	GGGCGTGGG TGGGGTAGA AG (SEQ ID NO: 957)	GGGCGUGGGUG GGGUAGAAGAG G (SEQ ID NO: 961)	GGGCGTGGGTG GGGTAGAAGAG G (SEQ ID NO: 965)	反义
SpN	CUCUUAGACA UAACACACCA (SEQ ID NO: 954)	CTCTTAGACA TAACACACC A (SEQ ID NO: 958)	CUCUUAGACAU AACACACCAGG G (SEQ ID NO: 962)	CTCTTAGACATA ACACACCAGGG (SEQ ID NO: 966)	反义
SpC	AUCAGAGGCC AAACCCUUCC (SEQ ID NO: 955)	ATCAGAGGC CAAACCCTTC C (SEQ ID NO: 959)	AUCAGAGGCCA AACCCUUCCUG G (SEQ ID NO: 963)	ATCAGAGGCCA AACCCTTCCTG G (SEQ ID NO: 967)	正义

[0382] 接下来，用BCL11Ae RNP电穿孔成人骨髓CD34$^+$细胞。从骨髓CD34$^+$细胞提取的gDNA扩增的BCL11A PCR产物的DNA测序分析表明15％基因编辑(包含插入和缺失)(图14A)。重要的是，所有缺失均导致GATA1基序缺失，并且所有插入都通过在基序中添加少量bp破坏了GATA1基序。将CD34$^+$细胞接种到集落形成测定中，并挑选出对应于CD34$^+$细胞克隆的混合造血集落(GEMM)。分离gDNA并通过Illumina测序进行分析，以量化靶位点的单等位基因和双等位基因破坏。从CD34$^+$细胞克隆中分化的大多数GEMM具有单等位基因破坏，也检测到了双等位基因破坏，其中总indel率与混合CD34+细胞群体中检测到的相比高约2/3(图14B)。这很可能反映了普通髓系祖细胞(CMP)的百分比，所述普通髓系祖细胞产生GEMM，所述GEMM占异源CD34$^+$细胞的大部分(相对于存在的其他谱系)，但在短期CFC测定中未捕获/分化。与供体匹配的未经处理的对照细胞相比，经RNP处理的骨髓CD34$^+$细胞还维持着相似的红系成熟(去核，图14C)和分化(表型获得，图14D)动力学。如通过流式细胞术分析所确定的，经编辑的骨髓CD34$^+$细胞的红系细胞后代显示出在HbF诱导方面约5倍增加(图14E)。

[0383] 还在成人mPB CD34$^+$细胞中评估了基因编辑和胎儿血红蛋白的诱导。BCL11Ae RNP和非特异性ssODN的共递送支持在靶位点处约20％indel(图15A)。为了评估经编辑的细胞的红系后代中胎儿血红蛋白的早期诱导，将mPB CD34$^+$细胞分化为成红细胞，并通过qRT-PCR分析评估胎儿血红蛋白转录(HBG mRNA)的诱导。与未经处理的对照相比，BCL11A e RNP处理的CD34$^+$细胞的红系后代表现出HBG mRNA的2倍诱导，表明胎儿血红蛋白表达的诱导(图15B)。与供体匹配的未经处理的对照细胞相比，经RNP处理的骨髓CD34$^+$细胞还维持着相似的分化(表型获得，图15C)动力学。

实例8:靶向人造血干/祖细胞中HBG启动子处远端CCAAT盒的Cas9 RNP的电穿孔产生几个缺失，所述缺失在红系分化后促进HBG表达

[0384] 在携带与HBG远端CCAAT盒重叠的13nt缺失的患者中观察到胎儿血红蛋白的遗传性持久性(HPFH表型)。靶向HBG远端CCAAT盒的Cas9-RNP可用于造血干/祖细胞(HSPC)中以重现HPFH表型，这可能是通过破坏阻遏HBG表达的转录因子的结合位点。Cas9 RNP产生的

DNA双链断裂(DSB)可以导致多种修复结果,包括在RNP剪切位点近端的插入和缺失。引入的indel中的一些可能不太有效地破坏阻遏因子的结合。可以预见,ssODN供体模板可用于通过指导针对导致HBG基因表达的特定缺失的修复来改善重现HPFH表型的indel的频率。

[0385] 为了评估靶向HBG启动子处远端CCAAT盒的Cas9-RNP的修复结果,将Cas9与化学合成的指导RNA OLI7066(SEQ ID NO:970,表10)(OLI7066-RNP)进行复合并且将RNP以16μM电穿孔到HSPC中。在电穿孔后第2天进行的测序分析(下一代测序)表明,有23.7%的等位基因携带与自然发生的HPFH突变相同的13nt缺失(图16,“Δ-102:-114”所示的13nt缺失)。在OLI7066-RNP剪切位点附近也观察到了几个其他频繁缺失(图16)。

表10:靶向CCAAT盒的化学合成的gRNA的序列,带有或不带有末端修饰。

名称	Oli-ID	gRNA 序列（RNA）	gRNA 序列（DNA）
未经修饰的 Sp35	OLI7066	CUUGUCAAGGCUAUUGGUCAG UUUUAGAGCUAGAAAUAGCAA GUUAAAAUAAGGCUAGUCCGU UAUCAACUUGAAAAAGUGGCA CCGAGUCGGUGCUUUU (SEQ ID NO: 970)	CTTGTCAAGGCTATTGGTCA GTTTTAGAGCTAGAAATAGC AAGTTAAAATAAGGCTAGTC CGTTATCAACTTGAAAAAGT GGCACCGAGTCGGTGCTTTT (SEQ ID NO: 972)
Sp35 5'-3' PSOMe	OLI8394	mC*mU*mU*GUCAAGGCUAUUG GUCAGUUUUAGAGCUAGAAAU AGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAG UGGCACCGAGUCGGUGCmU*m U*mU*U (SEQ ID NO: 971)	mC*mT*mT*GTCAAGGCTATT GGTCAGTTTTAGAGCTAGAA ATAGCAAGTTAAAATAAGGC TAGTCCGTTATCAACTTGAA AAAGTGGCACCGAGTCGGT GCmT*mT*mT*T (SEQ ID NO: 973)
*: 代表硫代磷酸酯修饰 m: 代表 2-o-甲基修饰			

[0386] 进行了单细胞实验以评估由靶向远端CCAAT盒的RNP(Cas9与gRNA OLI7066(SEQ ID NO:970)复合(“OLI7066-RNP”))产生的最频繁缺失所诱导的HbF表达的水平。简而言之,将以靶向远端CCAAT盒的OLI7066-RNP复合物电穿孔的单个HSPC接种在单个孔中,并在人细胞因子(促红细胞生成素,SCF,IL3)、人血浆(Octoplas公司)和其他补充剂(氢化可的松、肝素、转铁蛋白、胰岛素)存在下培养18天,分化为红系细胞。14天后,将来自每个单个细胞的一部分红系细胞后代进行分离,以提取gDNA。对来自HBG1和HBG2的PCR产物进行测序分析,以鉴定每个克隆群体的基因型。还对每个克隆群体的基因组DNA进行了ddPCR分析,以检测HBG1和HBG2中指导RNA靶位点之间4.9kb片段的缺失。在第18天,裂解衍生自每个克隆的红系细胞,并通过超高效液相色谱(UPLC)确定珠蛋白链的相对表达。最后,将由[γ链]/[全γ链+β链]确定的G-γ(Gγ)-珠蛋白,A-γ(Aγ)-珠蛋白链表达水平(或由4.9kb缺失产生的AG-γ(AGγ)-珠蛋白)水平分别相对于HBG1和HBG2携带的indel进行比较。重现HPFH基因型的13nt缺失显示诱导HBG表达。另外,鉴定出了以可比较的水平诱导HBG表达的几个独特的非天然缺失。这些包括HBGΔ-112:-115(“4nt缺失”)、HBGΔ-104:-121(“18nt缺失”)、HBGΔ-116(“1nt缺失”)(图17A-D)。

实例9:靶向远端CCAAT盒处或附近的区域的Cas9 RNP与ssODN供体的共递送支持人造血干/祖细胞中的精确基因编辑和红系后代中的γ-珠蛋白表达的增加

[0387] 为了改善HSPC中HbF诱导indel的频率,对“编码”经鉴定的HbF诱导缺失(例如1nt(HBGΔ-116)和4nt(HBGΔ-112:-115)缺失)的单链脱氧核苷酸供体修复模板(ssODN)进行了设计(图18)。用于诱导这些缺失的ssODN如下:“1nt”ssODN:OLI16417-18;“4nt”ssODN:

OLI16424,OLI16430,表11)。ssODN设计具有90nt同源臂(位于该缺失序列的侧翼)以产生完全缺失。修饰ssODN,使其在5'和3'末端含有硫代磷酸酯(PhTx)。设计另外的ssODN,以编码在具有HPFH的患者中观察到的天然发生的突变(例如HBG Δ-102:-114(13nt缺失,图18)和HBG-117G>A突变(HBG-117G>A)。用于诱导这些缺失的ssODN如下:"13nt"ssODN:OLI16412、OLI16414和"-117G>A"ssODN:OLI16415-16,表11)。最后,为便于评估基因校正,ssODN(OLI16411、OLI16413,表11)被设计为编码与远端CCAAT盒重叠的11nt缺失,其在单独OLI8394-RNP的电穿孔后以低频率发生(0.1%,参见图19B、图18)。

表11:单链脱氧核苷酸供体修复模板"编码"CCAAT盒处或附近的缺失或突变

ssODN ID	OLI-ID	序列
Ptx ssODN - 正链 -18nt (180)	OLI16409	G*GTGCTTCCTTTTATTCTTCATCCCTAGCCAGCCGCCGGCCC CTGGCCTCACTGGATACTCTAAGACTATTGGTCAAGTTTGCCT TGTCAAGGCTGGCCAACCCATGGGTGGAGTTTAGCCAGGGA CCGTTTCAGACAGATATTTGCATTGAGATAGTGTGGGGAAGG GGCCCCCAAGA*G (SEQ ID NO: 974)
Ptx ssODN - 负链 -18nt (180)	OLI16410	C*TCTTGGGGGCCCCTTCCCCACACTATCTCAATGCAAATATC TGTCTGAAACGGTCCCTGGCTAAACTCCACCCATGGGTTGGC CAGCCTTGACAAGGCAAACTTGACCAATAGTCTTAGAGTATC CAGTGAGGCCAGGGGCCGGCGGCTGGCTAGGGATGAAGAAT AAAAGGAAGCAC*C (SEQ ID NO: 975)
Ptx ssODN - 负链 -11nt (180)	OLI16411	G*GCCCCTTCCCCACACTATCTCAATGCAAATATCTGTCTGAA ACGGTCCCTGGCTAAACTCCACCCATGGGTTGGCCAGCCTTG CCTTGAGACAAGGCAAACTTGACCAATAGTCTTAGAGTATCC AGTGAGGCCAGGGGCCGGCGGCTGGCTAGGGATGAAGAATA AAAGGAAGCACC*C (SEQ ID NO: 976)
Ptx ssODN - 负链 -13nt (180)	OLI16412	G*GCCCCTTCCCCACACTATCTCAATGCAAATATCTGTCTGAA ACGGTCCCTGGCTAAACTCCACCCATGGGTTGGCCAGCCTTG CCTTGACAAGGCAAACTTGACCAATAGTCTTAGAGTATCCAG TGAGGCCAGGGGCCGGCGGCTGGCTAGGGATGAAGAATAAA AGGAAGCACCCT*T (SEQ ID NO: 977)
Ptx ssODN - 正链 -11nt (180)	OLI16413	G*GGTGCTTCCTTTTATTCTTCATCCCTAGCCAGCCGCCGGCC CCTGGCCTCACTGGATACTCTAAGACTATTGGTCAAGTTTGCC TTGTCTCAAGGCAAGGCTGGCCAACCCATGGGTGGAGTTTAG CCAGGGACCGTTTCAGACAGATATTTGCATTGAGATAGTGTG GGGAAGGGGC*C (SEQ ID NO: 978)

Ptx ssODN - 正链 -13nt (180)	OLI16414	A*AGGGTGCTTCCTTTTATTCTTCATCCCTAGCCAGCCGCCGG CCCCTGGCCTCACTGGATACTCTAAGACTATTGGTCAAGTTTG CCTTGTCAAGGCAAGGCTGGCCAACCCATGGGTGGAGTTTAG CCAGGGACCGTTTCAGACAGATATTTGCATTGAGATAGTGTG GGGAAGGGGC*C (SEQ ID NO: 979)
Ptx ssODN - 负链 117: G > A (180)	OLI16415	G*GCCCCTTCCCCACACTATCTCAATGCAAATATCTGTCTGAA ACGGTCCCTGGCTAAACTCCACCCATGGGTTGGCCAGCCTTG CCTTAACCAATAGCCTTGACAAGGCAAACTTGACCAATAGTC TTAGAGTATCCAGTGAGGCCAGGGGCCGGCGGCTGGCTAGG GATGAAGAATAA*A (SEQ ID NO: 980)
Ptx ssODN - 正链 117: G > A (180)	OLI16416	T*TTATTCTTCATCCCTAGCCAGCCGCCGGCCCCTGGCCTCAC TGGATACTCTAAGACTATTGGTCAAGTTTGCCTTGTCAAGGCT ATTGGTTAAGGCAAGGCTGGCCAACCCATGGGTGGAGTTTAG CCAGGGACCGTTTCAGACAGATATTTGCATTGAGATAGTGTG GGGAAGGGGC*C (SEQ ID NO: 981)
Ptx ssODN - 负链 -1nt (180)	OLI16417	G*GCCCCTTCCCCACACTATCTCAATGCAAATATCTGTCTGAA ACGGTCCCTGGCTAAACTCCACCCATGGGTTGGCCAGCCTTG CCTTGCCAATAGCCTTGACAAGGCAAACTTGACCAATAGTCT TAGAGTATCCAGTGAGGCCAGGGGCCGGCGGCTGGCTAGGG ATGAAGAATAAA*A (SEQ ID NO: 982)
Ptx ssODN - 正链 -1nt (180)	OLI16418	T*TTTATTCTTCATCCCTAGCCAGCCGCCGGCCCCTGGCCTCA CTGGATACTCTAAGACTATTGGTCAAGTTTGCCTTGTCAAGGC TATTGGCAAGGCAAGGCTGGCCAACCCATGGGTGGAGTTTAG CCAGGGACCGTTTCAGACAGATATTTGCATTGAGATAGTGTG GGGAAGGGGC*C (SEQ ID NO: 983)
Ptx ssODN - 负链 -4nt 40/80 (120)	OLI16419	T*GGCTAAACTCCACCCATGGGTTGGCCAGCCTTGCCTTGATA GCCTTGACAAGGCAAACTTGACCAATAGTCTTAGAGTATCCA GTGAGGCCAGGGGCCGGCGGCTGGCTAGGGATGAA*G (SEQ ID NO: 984)
Ptx ssODN - 负链 -4nt 30/70 (100)	OLI16420	C*CACCCATGGGTTGGCCAGCCTTGCCTTGATAGCCTTGACA AGGCAAACTTGACCAATAGTCTTAGAGTATCCAGTGAGGCCA GGGGCCGGCGGCTGGC*T (SEQ ID NO: 985)
Ptx ssODN - 负链 -4nt (100)	OLI16421	A*AACGGTCCCTGGCTAAACTCCACCCATGGGTTGGCCAGCC TTGCCTTGATAGCCTTGACAAGGCAAACTTGACCAATAGTCTT AGAGTATCCAGTGAG*G (SEQ ID NO: 986)
Ptx ssODN - 负链 -4nt (120)	OLI16422	T*ATCTGTCTGAAACGGTCCCTGGCTAAACTCCACCCATGGGT TGGCCAGCCTTGCCTTGATAGCCTTGACAAGGCAAACTTGAC CAATAGTCTTAGAGTATCCAGTGAGGCCAGGGGCC*G (SEQ ID NO: 987)
Ptx ssODN - 负链 -4nt (140)	OLI16423	T*CAATGCAAATATCTGTCTGAAACGGTCCCTGGCTAAACTCC ACCCATGGGTTGGCCAGCCTTGCCTTGATAGCCTTGACAAGG CAAACTTGACCAATAGTCTTAGAGTATCCAGTGAGGCCAGGG GCCGGCGGCTGGC*T (SEQ ID NO: 988)
Ptx ssODN - 负链 -4nt (180)	OLI16424	G*GCCCCTTCCCCACACTATCTCAATGCAAATATCTGTCTGAA ACGGTCCCTGGCTAAACTCCACCCATGGGTTGGCCAGCCTTG CCTTGATAGCCTTGACAAGGCAAACTTGACCAATAGTCTTAG AGTATCCAGTGAGGCCAGGGGCCGGCGGCTGGCTAGGGATG AAGAATAAAAGG*A (SEQ ID NO: 989)
Ptx ssODN - 正链 -4nt 40/80 (120)	OLI16425	T*ACTCTAAGACTATTGGTCAAGTTTGCCTTGTCAAGGCTATC AAGGCAAGGCTGGCCAACCCATGGGTGGAGTTTAGCCAGGG ACCGTTTCAGACAGATATTTGCATTGAGATAGTGTG*G (SEQ ID NO: 990)
Ptx ssODN - 正链 -4nt 30/70 (100)	OLI16426	C*TATTGGTCAAGTTTGCCTTGTCAAGGCTATCAAGGCAAGGC TGGCCAACCCATGGGTGGAGTTTAGCCAGGGACCGTTTCAGA CAGATATTTGCATTG*A (SEQ ID NO: 991)

Ptx ssODN - 正链 -4nt (100)	OLI16427	C*CTCACTGGATACTCTAAGACTATTGGTCAAGTTTGCCTTGT CAAGGCTATCAAGGCAAGGCTGGCCAACCCATGGGTGGAGT TTAGCCAGGGACCGTT*T (SEQ ID NO: 992)
Ptx ssODN - 正链 -4nt (120)	OLI16428	C*GGCCCCTGGCCTCACTGGATACTCTAAGACTATTGGTCAAG TTTGCCTTGTCAAGGCTATCAAGGCAAGGCTGGCCAACCCAT GGGTGGAGTTTAGCCAGGGACCGTTTCAGACAGAT*A (SEQ ID NO: 993)
Ptx ssODN - 正链 -4nt (140)	OLI16429	A*GCCAGCCGCCGGCCCCTGGCCTCACTGGATACTCTAAGAC TATTGGTCAAGTTTGCCTTGTCAAGGCTATCAAGGCAAGGCT GGCCAACCCATGGGTGGAGTTTAGCCAGGGACCGTTTCAGA CAGATATTTGCATTG*A (SEQ ID NO: 994)
Ptx ssODN - 正链 -4nt (180)	OLI16430	T*CCTTTTATTCTTCATCCCTAGCCAGCCGCCGGCCCCTGGCC TCACTGGATACTCTAAGACTATTGGTCAAGTTTGCCTTGTCAA GGCTATCAAGGCAAGGCTGGCCAACCCATGGGTGGAGTTTAG CCAGGGACCGTTTCAGACAGATATTTGCATTGAGATAGTGTG GGGAAGGGGC*C (SEQ ID NO: 995)
Ptx ssODN - 正链 -18nt (100)	OLI16431	C*CCTGGCCTCACTGGATACTCTAAGACTATTGGTCAAGTTTG CCTTGTCAAGGCTGGCCAACCCATGGGTGGAGTTTAGCCAGG GACCGTTTCAGACAG*A (SEQ ID NO: 1040)
*: 代表硫代磷酸酯修饰		

[0388] 将ssODN供体OLI16409-OLI16418、OLI16424和OLI16430(表11)跟与OLI8394预复合的Cas9蛋白(OLI8394-RNP)或与OLI7066预复合的Cas9蛋白(OLI7066-RNP)(靶向HBG远端CCAAT盒)一起共递送至成人mPB CD34+细胞(表10)。简而言之,将mPB CD34+细胞用X-Vivo-10中的人细胞因子预刺激2天,然后用由2.5μM的ssODN供体和2μM的OLI8394-RNP或OLI7066-RNP构成的混合物电穿孔。电穿孔后三天后,提取基因组DNA,并对HBG PCR产物进行下一代测序。编辑水平从单独使用OLI8394-RNP时的62.1%提高到在共递送“-11”ssODN(即OLI16411或OLI16413)时的78.73%(图19A)。ssODN模板至人HSPC的共递送介导的基因校正水平显示对总indel的贡献高达42.4%,这是通过来自用OLI8394-RNP+“11”ssODN电穿孔的HSPC的HBG PCR产物的测序所检测到的总indel中“-11”缺失的频率而确定(图19B)。跨其他ssODN模板和OLI8394-RNP(图19C-E、G)或OLI7066-RNP(图19F)观察到有效的基因校正。

[0389] 将用与ssODN模板共递送的OLI8394-RNP或OLI7066-RNP电穿孔的mPB CD34+细胞分化为红系细胞,以评估基因编辑产生的γ-珠蛋白表达的水平。为了确定与ssODN共递送RNP是否增加经编辑的成人CD34+细胞的红系后代中胎儿血红蛋白的产生,通过在存在人细胞因子(促红细胞生成素、SCF、IL3)、人血浆(Octoplas)和其他补充剂(氢化可的松、肝素、转铁蛋白、胰岛素)的情况下培养18天,将细胞分化为红系细胞。在第18天,通过UPLC测量γ-珠蛋白链相对于总β样珠蛋白链(γ链/[γ链+β链])的相对表达水平。在共递送ssODN后检测到高达39.1%的γ-珠蛋白,而不是当单独使用OLI8394-RNP时的29%(图20A-B)。

[0390] 使用ssODN OLI16424评估了ssODN剂量的效果。将mPB CD34+细胞用X-Vivo-10中的人细胞因子预刺激2天,然后用由0.625μM至10μM剂量范围的OLI16424-ssODN和2μM的OLI8394-RNP构成的混合物电穿孔。电穿孔后三天后,提取基因组DNA,并对HBG PCR产物进行下一代测序。将ssODN的剂量增加高达5μM,这导致基因校正的增加和HBG1和HBG2之间的4.9kb缺失频率的降低(通过ddPCR测定),同时维持总体编辑水平且不影响细胞生存力(图21A-C)。如通过红系培养18天后的细胞裂解物的UPLC分析测量,在5μM剂量的ssODN下,γ-珠蛋白水平增加达至总β样链的39.5%(图21D)。

[0391] 通过改变ssODN和RNP的相对剂量，进一步优化OLI8394-RNP和OLI16424 ssODN的共递送。将mPB CD34+细胞用X-Vivo-10中的人细胞因子预刺激2天，然后用由1.25μM至5μM剂量范围的OLI16424-ssODN和2μM至8μM的OLI8394-RNP构成的混合物电穿孔。电穿孔后三天后，提取基因组DNA，并对HBG PCR产物进行下一代测序。用较低的RNP剂量与较高的ssODN剂量组合实现改善的基因编辑结果，即，HBG1和HBG2基因之间4.9kb缺失的频率较低，并且基因校正水平较高(图22A-C)。在红系分化后，这导致了最高水平的γ-珠蛋白表达(图22D)。

实例10：可以使用具有对称和不对称同源臂的短ssODN模板实现人造血干/祖细胞的基因校正

[0392] 为了确定是否可以使用比在实例9中使用的先前测试的180nt长度更短的模板来实现基因校正，设计并合成了编码4nt缺失(HBGΔ-112:-115)的具有相对于缺失的序列对称或不对称的较短同源臂的ssODN(OLI16419-OLI16423和OLI16425-OLI16429，表11，图23A)。简而言之，将mPB CD34+细胞用X-Vivo-10中的人细胞因子预刺激2天，然后用由2.5μM的ssODN供体(OLI16424、OLI16419或OLI16421)和2μM的与OLI8394预复合的Cas9蛋白(OLI8394-RNP)构成的混合物电穿孔。电穿孔后三天后，提取基因组DNA，并对HBG PCR产物进行下一代测序。跨具有对称的90nt同源性臂、对称的50nt同源性臂或40/80非对称性臂的ssODN，总编辑和基因校正水平相似(图23B)。

实例11：ssODN供体与成对D10A-Cas9 RNP共递送至成人CD34+细胞支持HBG远端CCAAT盒处的基因校正

[0393] 为了确定D10A切口酶是否支持ssODN介导的基因校正以在mPB $CD34^+$细胞中引入CCAAT盒破坏性缺失(如“4nt”缺失)，将CD34+细胞用靶向HBG的D10A Cas9 RNP电穿孔(在具有或不具有ssODN(OLI16424)的情况下)(表11)。简要地，化学合成Sp37和SpA gRNA(分别为OLI7075和OLI7074，表12)，并与D10A-Cas9切口酶突变蛋白复合。在补充有人细胞因子的培养基中将成人mPB CD34+细胞预刺激48小时。接下来，将包含Sp37+SpA(2μM+2μM)的D10A-Cas9RNP对单独或与OLI16424组合地递送至CD34+细胞，并在电穿孔后3天提取基因组DNA，用于Illumina测序分析。

表12：用于在D10A Cas9情况下靶向$CD34^+$细胞中CCAAT盒的gRNA序列

gRNA ID	OLI-ID	gRNA 序列（RNA）	gRNA 序列（DNA）
Sp37	OLI17075	CUUGACCAAUAGCCUUGACA GUUUUAGAGCUAGAAAUAGC AAGUUAAAAUAAGGCUAGUC CGUUAUCAACUUGAAAAAGU GGCACCGAGUCGGUGCUUUU (SEQ ID NO: 996)	CTTGACCAATAGCCTTGACAGTTT TAGAGCTAGAAATAGCAAGTTAAA ATAAGGCTAGTCCGTTATCAACTT GAAAAAGTGGCACCGAGTCGGTG CTTTT (SEQ ID NO: 998)
SpA	OLI17074	GGCAAGGCUGGCCAACCCAU GUUUUAGAGCUAGAAAUAGC AAGUUAAAAUAAGGCUAGUC CGUUAUCAACUUGAAAAAGU GGCACCGAGUCGGUGCUUUU (SEQ ID NO: 997)	GGCAAGGCTGGCCAACCCATGTT TTAGAGCTAGAAATAGCAAGTTAA AATAAGGCTAGTCCGTTATCAACT TGAAAAAGTGGCACCGAGTCGGT GCTTTT (SEQ ID NO: 999)

[0394] ssODN的共递送支持约65％indel，而不是当单独递送D10A-Cas9 RNP对时的约51％，如通过对基因组DNA的HBG PCR产物进行测序确定的(图24)。详细的测序分析还表明，

当共递送OLI16424 ssODN供体时，有16％的等位基因带有精确的4nt缺失(HBG:Δ-112:-115)，而当单独递送RNP对时，未检测到这种缺失。这表明在ssODN存在下，约25％的indel是通过精确的基因校正而发生的。

实例12:含有靶向CCAAT盒的gRNA的gpf的Cpf1 RNP支持人造血干/祖细胞中的基因编辑

[0395] 设计了Cpf1指导RNA，HBG1-1(即，OLI13620(表13))以靶向CCAAT盒。为了确定在CD34+细胞中递送氨基酸球菌属物种(Acidaminococcus sp.)Cpf1(“AsCpf1”)的最佳核定位信号配置，将HBG1-1 gRNA与AsCpf1的两个核定位信号(NLS)变体(即His-AsCpf1-nNLS(SEQ ID NO:1000)和His-AsCpf1-sNLS-sNLS(SEQ ID NO:1001))复合。“His”是指六组氨酸的纯化序列，“AsCpf1”是指氨基酸球菌属物种Cpf1蛋白序列，“nNLS”是指核质蛋白NLS，并且“sNLS”是指SV40 NLS。

[0396] 简而言之，将mPB CD34+细胞用X-Vivo-10中的人细胞因子预刺激2天，然后用5μM或20μM的RNP电穿孔。电穿孔三天后提取基因组DNA，并对HBG PCR产物进行下一代测序。与测试的Cpf1 NLS变体(“His-AsCpf1-sNLS-sNLS_HBG1-1 RNP”或“His-AsCpf1-nNLS_HBG1-1 RNP”)复合的HBG1-1 gRNA支持CD34+细胞的13nt靶位点处的编辑。在测试的最高剂量下，His-AsCpf1-sNLS-sNLS_HBG1-1 RNP产生了60.6％的经编辑的等位基因，并且His-AsCpf1-nNLS_HBG1-1 RNP产生了51.1％的经编辑的等位基因(图25A)。

表13:用于靶向$CD34^+$细胞中CCAAT盒的HBG1-1 gRNA序列

gRNA ID	OLI-ID	靶向结构域序列（RNA）	靶向结构域序列（DNA）	靶向结构域序列加上PAM（UUUG）（RNA）	靶向结构域序列加上PAM（TTTG）（DNA）	正义
HBG1-1	OLI13620	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO: 1002)	CCTTGTCAAGGCTATTGGTC (SEQ ID NO: 1003)	UUUGCCUUGUCAAGGCUAUUGGUC (SEQ ID NO: 1004)	TTTGCCTTGTCAAGGCTATTGGTC (SEQ ID NO: 1005)	反义

实例13:靶向CCAAT盒的Cpf1 RNP与ssODN供体的共递送支持人造血干/祖细胞中的基因编辑

[0397] 将包含与His-AsCpf1-sNLS-sNLS变体复合的HBG1-1 gRNA的RNP(“His-AsCpf1-sNLS-sNLS_HBG1-1 RNP”)跟单链寡脱氧核苷酸供体修复模板(ssODN)通过电穿孔共递送到mPB CD34+细胞。如以上实例中所讨论的，OLI16430和OLI16424 ssODN被设计为“编码”4个核苷酸缺失，并且OLI16409和OLI16410ssODN被设计为“编码”18个核苷酸缺失(表11)。4nt和18nt缺失都破坏了HBG远端CCAAT盒，并与HBG表达的诱导有关。ssODN包括90个核苷酸长的同源臂(位于编码的缺失序列的侧翼)以产生完全缺失。修饰ssODN，使其在5’和3’末端含有硫代磷酸酯(PhTx)(OLI16430、OLI16424、OLI16409和OLI16410，表11)。简而言之，将在补充有人细胞因子的培养基中预刺激两天的成人mPB CD34+细胞用5μM RNP电穿孔，所述RNP包含与HBG1-1 gRNA(“His-AsCpf1-sNLS-sNLS_HBG1-1RNP”)复合的His-AsCpf1-sNLS-sNLS蛋白，单独或与2.5μM的ssODN供体(OLI16430、OLI164324、OLI16409或OLI16410)之一组合。RNP和编码18nt缺失的具有正链同源臂的ssODN供体(OLI16409)的共递送将编辑频率从没有供体情况下的39.5％提高到73.3％，如通过电穿孔后72小时对提取的基因组DNA的HBG

PCR产物的测序分析确定的(图26A)。

[0398] 对靶位点处缺失的特定类型和大小的进一步分析表明,在存在18nt正链供体(OLI16409)的情况下,57.3%的等位基因(相比之下,当单独递送His-AsCpf1-sNLS-sNLS_HBG1-1 RNP时是7.5%的等位基因)携带18nt缺失(图26B)。ssODN共递送介导了DNA DSB的精确修复,因为18nt缺失代表ssODN OLI16409和His-AsCpf1-sNLS-sNLS_HBG1-1 RNP共递送产生的所有indel的71.2%,然而,当单独递送His-AsCpf1-sNLS-sNLS_HBG1-1 RNP时,18nt缺失仅代表产生的所有indel的19.0%(图26C)。

[0399] 尽管图26A-26C中的数据最初表明,Cpf1 RNP与ssODN供体的共递送增加了基因组DNA的精确基因编辑,导致18nt缺失,但随后进行的重复实例13实验的测试证实了这些数据是实验的假象。尽管如此,从随后使用指导RNA HBG1-1和ssODN OLI16409测试AsCpf1 RNP的重复实验中获得的数据表明,ssODN供体的共递送支持人mPB-CD34+细胞中总编辑的增加,尽管与针对"-18nt缺失"的DNA DSB的精确修复无关(参见实例21)。

实例14:含有靶向HBG启动子的远端CCAAT盒区域的gRNA的Cpf1 RNP支持人造血干/祖细胞中的基因编辑,所述基因编辑促进红系后代中HbF蛋白表达的诱导

[0400] 具有包含SEQ ID NO:1002的靶向结构域(表15)的指导RNA HBG1-1(表14)靶向HBG启动子内的位点(图27A)。HBG1-1 gRNA(SEQ ID NO:1022)与野生型AsCpf1(AsCpf1-sNLS-sNLS,SEQ ID NO:1001)复合形成RNP("AsCpf1-HBG1-1-RNP",表14)。然后使用Amaxa核苷酸转染仪(龙沙公司(Lonza))或MaxCyte GT(MaxCyte,Inc.)电穿孔设备将这种复合物(5μM或20μM)电穿孔到动员的外周血(mPB)衍生的CD34+细胞中。在电穿孔后三天,通过对PCR扩增的靶位点的Illumina测序(NGS)分析靶位点处的插入/缺失水平。AsCpf1-HBG1-1-RNP在Amaxa和MaxCyte电穿孔系统上分别支持43%和17%的中靶编辑(图28A(Amaxa)和28B(MaxCyte))。在对mPB CD34+细胞进行编辑后,进行了18天的离体分化为红系谱系(Giarratana 2011)。然后,通过UPLC测量γ-珠蛋白链的相对表达水平(相比于总β样珠蛋白链)(图28A(Amaxa)和28B(MaxCyte))。AsCpf1-HBG1-1-RNP导致γ-珠蛋白表达增加,比在模拟电穿孔的mPB CD34+细胞衍生的细胞中检测到的水平高多达21%(图28A)。

表14:RNP组分

gRNA ID	蛋白变体	RNP 组分	RNP ID	gRNA 修饰	gRNA 序列（RNA）
HBG1-1 gRNA	AsCpf1-sNLS-sNLS (SEQ ID NO: 1001)	gRNA + AsCpf1 WT	HBG1-1-AsCpf1-RNP	合成的未经修饰的	UAAUUUCUACUCUUGUAGAUCCUUGUCAAGGCUAUUGGUC (SEQ ID NO: 1022)
HBG1-1 gRNA	His-AsCpf1-sNLS-sNLS H800A (SEQ ID NO: 1032)	gRNA + AsCpf1 H800A	HBG1-1-AsCpf1H800A-RNP	合成的未经修饰的	UAAUUUCUACUCUUGUAGAUCCUUGUCAAGGCUAUUGGUC (SEQ ID NO: 1022)
HBG1-1 gRNA	His-AsCpf1-sNLS-sNLS H800A (SEQ ID NO: 1032)	gRNA + AsCpf1 H800A	HBG1-1-AsCpf1H800A-RNP	5'OMe-PS2 OMe	UAAUUUCUACUCUUGUAGAUCCUUGUCAAGGCUAUUGGmU/PS2/mC (SEQ ID NO: 1041)
tSpA 失活的 gRNA	SpCas9WT (SEQ ID NO: 1033)	失活的 gRNA + SpCas9 WT	tSpA-Cas9-RNP	合成的 5'-3' 3xPSOMe	mG*mG*mC*UGGCCAACCCAUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCmU*mU*mU*U (SEQ ID NO: 1024)

SpA gRNA	His-NLS-SpCas 9D10A (SEQ ID NO: 1034)	gRNA + SpCas9 切口酶	SpA-D10A-RNP	合成的 5'-3' 3xPSOMe	mG*mG*mC*AAG GCUGGCCAACCC AUGUUUUAGAGC UAGAAAUAGCAA GUUAAAAUAAGG CUAGUCCGUUAU CAACUUGAAAAA GUGGCACCGAGU CGGUGCmU*mU* mU*U (SEQ ID NO: 1025)
SpG gRNA	His-NLS-SpCas 9D10A (SEQ ID NO: 1034)	gRNA + SpCas9 切口酶	SpG-D10A-RNP	合成的 5'-3' 3xPSOMe	mU*mA*mG*UCU UAGAGUAUCCAG UGGUUUUAGAGC UAGAAAUAGCAA GUUAAAAUAAGG CUAGUCCGUUAU CAACUUGAAAAA GUGGCACCGAGU CGGUGCmU*mU* mU*U (SEQ ID NO: 1026)
Sp182 失活的 gRNA	SpCas9WT (SEQ ID NO: 1033)	失活的 gRNA + SpCas9 WT	Sp182-Cas9 -RNP	合成的 5'-3' 3xPSOMe	mU*mU*mA*GAG UAUCCAGUGGUU UUAGAGCUAGAA AUAGCAAGUUAA AAUAAGGCUAGU CCGUUAUCAACU UGAAAAAGUGGC ACCGAGUCGGUG CmU*mU*mU*U (SEQ ID NO: 1027)
"His" 是指六组氨酸的纯化序列 AsCpf1 是指氨基酸球菌属物种 Cpf1 蛋白序列 "nNLS" 是指核质蛋白 NLS，并且 "sNLS" 是指 SV40 NLS "SpCas9" 是指化脓链球菌 Cas9 蛋白序列 "*" 代表硫代磷酸酯修饰 "PS" 代表硫代磷酸酯修饰 "PS2" 代表二硫代磷酸酯修饰 "OMe" 代表 2-o-甲基修饰 "m" 表示 2-o-甲基修饰					

表15:Cpf1指导RNA靶向序列

gRNA ID	gRNA 靶向结构域序列（RNA）	gRNA 靶向结构域序列（DNA）
HBG1-1 gRNA	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO: 1002)	CCTTGTCAAGGCTATTGGTC (SEQ ID NO: 1003)
Sp182 dgRNA	UUAGAGUAUCCAGUG (SEQ ID NO: 1028)	TTAGAGTATCCAGTG (SEQ ID NO: 1029)
SpG gRNA	UAGUCUUAGAGUAUCCAGUG (SEQ ID NO: 359)	TAGTCTTAGAGTATCCAGTG (SEQ ID NO: 1030)
SpA gRNA	GGCAAGGCUGGCCAACCCAU (SEQ ID NO: 941)	GGCAAGGCTGGCCAACCCAT (SEQ ID NO: 944)
tSpA dgRNA	GGCUGGCCAACCCAU (SEQ ID NO: 326)	GGCTGGCCAACCCAT (SEQ ID NO: 1031)

实例15:当与化脓链球菌Cas9 RNP共递送时,在HBG启动子的CCAAT盒区域处的Cpf1

RNP编辑效率提高,而对生存力没有影响

[0401] 假设Cpf1-RNP与近端靶向性化脓链球菌Cas9 RNP的共递送,无论是催化失活还是引入单个切口,都可以增强靶位点处的编辑水平。为了增强HBG启动子的远端CCAAT盒区域处的Cpf1-RNP介导的编辑,将HBG1-1-AsCpf1H800A-RNP(由His-AsCpf1-sNLS-sNLS-H800A(SEQ ID NO:1032)与如表14所示的具有3'修饰的HBG1-1 gRNA(SEQ ID NO:1041)复合构成)与以下共递送到mPB CD34+细胞中:(1)化脓链球菌Cas9 D10A RNP,其含有与选自以下的全长指导RNA(100mer)复合的Cas9 D10A蛋白(His-NLS-SpCas9D10A,SEQ ID NO:1034):(a)SpA gRNA(表14,15;图27B)("SpA-D10A-RNP",表14)或(b)SpG gRNA(表14,15;图27C)("SpG-D10A-RNP",表14));(2)化脓链球菌WT Cas9 RNP,其含有与选自以下的截短的失活指导RNA(95mer-具有缩短的原间隔子区域)复合的WT Cas9蛋白(SpCas9WT,SEQ ID NO:1033):(a)tSpA dgRNA(表14、15;图27E)("tSpA-Cas9-RNP",表14);(b)Sp182 dgRNA(表14,15;图27F)("Sp182-Cas9-RNP",表14);或(c)tSpA-Cas9-RNP和Sp182-Cas9-RNP(表14,15,图27D)。使用MaxCyte GT设备(MaxCyte,Inc)对RNP复合物进行电穿孔。然后在电穿孔后三天,通过对PCR扩增的靶位点的Illumina测序(NGS)分析靶位点处的插入/缺失水平。在所有测试的组合中,无论使用的化脓链球菌Cas9酶(D10A或WT)和PAM方向如何,总编辑增加,超过在单独HBG1-1-AsCpf1H800A-RNP的Maxcyte递送后观察到的水平(图29和表16)。此外,共递送化脓链球菌Cas9 RNP时对mPB CD34+细胞的生存力没有有害影响(图30)。

表16:在测试的最大RNP剂量下,每种RNP组合的indel和γ-珠蛋白表达的汇总

Cpf1 RNP	Cas9 RNP1	Cas9 RNP2	Indel (NGS)	γ-珠蛋白 (相比总β样链)
-	-	-	0%	13.5%
HBG1-1-AsCpf1H800A-RNP	-	-	8.0%	20.0%
HBG1-1-AsCpf1H800A-RNP	SpG-D10A-RNP	-	88.6%	30.2%
HBG1-1-AsCpf1H800A-RNP	SpA-D10A-RNP	-	42.6%	32.5%
HBG1-1-AsCpf1H800A-RNP	tSpA-Cas9-RNP	-	32.2%	19.8%
HBG1-1-AsCpf1H800A-RNP	Sp182-Cas9-RNP	-	85.4%	42.8%
HBG1-1-AsCpf1H800A-RNP	tSpA-Cas9-RNP	Sp182-Cas9-RNP	80.1%	33.7%

实例16:Cpf1编辑谱可以用化脓链球菌Cas9-D10A-RNP的共递送操作

[0402] 除了将HBG1-1-AsCpf1H800A-RNP与化脓链球菌Cas9-D10A-RNP共递送至mPB CD34+细胞时观察到的总编辑增加外(图29),还观察到indel谱的变化。通过D10A酶在Cpf1-RNP靶位点近端引入单链断裂(图31A)改变了indel的方向性、长度和/或位置(图31B)。例如,Sp182-D10A-RNP强烈地将谱移向D10A RNP引入的切口位点,其中各种长度的缺失从Cpf1剪切位点向上游切口位点延伸(Sp182-D10A-RNP,图31B)。SpA-D10A-RNP也促进了缺失的产生

从HBG1-1-AsCpf1H800A-RNP剪切位点延伸到D10A酶引入的切口位点(此处位于下游),但在这种情况下,还产生了明显来自切口位点但并未完全延伸到HBG1-1-AsCpf1H800A-RNP剪切位点的另外缺失(SpA-D10A-RNP,图31B)。

[0403] 化脓链球菌Cas9-D10A-RNP靶位点相对于Cpf1-RNP靶位点的方向、靶链和距离可能导致由另外DNA缺口促进的突变的位置和长度方面的差异(图31A)。应当注意,在某些应用中,indel谱的这种定向操作(即,Cpf1-RNP和化脓链球菌Cas9 D10A-RNP结合位点之间发生indel的频率增加)可用于有利于所期望的编辑结果,从而提高有效indel率(例如,会破坏靶位点的indel)。在这种情况下,为了破坏dCCAAT靶区域并诱导HbF蛋白表达,HBG1-1-AsCpf1H800A-RNP与化脓链球菌D10A-RNP的共递送导致γ-珠蛋白水平增加16.7%(SpG-D10A-RNP)或19.0%(SpA-D10A-RNP),高于在模拟电穿孔细胞中检测到的水平,如电穿孔后红系分化18天后通过UPLC分析检测到的(图29和表16)。将HBG1-1-AsCpf1H800A-RNP与SpA-D10A-RNP配对时(相比于SpG-D10A-RNP)有效indel的频率更高,因为从较低的编辑水平(SpA-D10A-RNP情况下为42.6%,相比于SpG-D10A-RNP情况下为88.6%)实现了较高的HbF水平增加(Cas9D10A-SpA-RNP情况下为19%,相比于SpG-D10A-RNP情况下为16.7%)(图29和表16)。

实例17:Cpf1编辑在含有失活指导RNA的化脓链球菌Cas9 WT RNP存在下增加,而不改变编辑谱

[0404] 当使用MaxCyte电穿孔仪设备将HBG1-1-AsCpf1H800A-RNP(固定剂量为6μM)与增加剂量的Sp182-Cas9-RNP(与截短的95mer失活gRNA,Sp182复合)共递送至mPB CD34+细胞时,观察到编辑加强高达大于10倍(最高剂量的Sp182-Cas9-RNP编辑情况下为85.4%,相比之下,单独递送HBG1-1-AsCpf1H800A-RNP情况下为8.0%)(图29)。本文通过共递送Sp182-Cas9-RNP达到的编辑水平也大大超过了当以高达20uM的剂量单独递送HBG1-1-AsCpf1H800A-RNP时使用MaxCyte电穿孔仪设备获得的编辑水平(图28B)。与通过共递送D10A-RNP观察到的结果相反(在实例16中),通过共递送HBG1-1-AsCpf1H800A-RNP与Sp182-Cas9-RNP(与截短的gRNA复合的Cas9-WT)实现的编辑增加与对indel谱的明显影响无关。在电穿孔细胞中检测到的indel以近似对称的方式以Cpf1剪切位点为中心,并且在Sp182-Cas9-RNP靶位点未检测到indel(图32A)。值得注意的是,总indel的96%破坏了远端CCAAT盒,而总indel的79.5%破坏了CCAAT盒基序的3个或更多个核苷酸(图32B)。使用HBG1-1-AsCpf1H800A-RNP(由His-AsCpf1-sNLS-sNLS-H800A(SEQ ID NO:1032)与HBG1-1 gRNA(SEQ ID NO:1022)复合构成)和Sp182-Cas9-RNP的进一步剂量优化使得在电穿孔后120小时mPB-CD34+细胞中的编辑水平高达92%(使用MaxCyte装置),并导致红系衍生细胞中的γ链表达水平比背景高32%(在经处理的细胞中41%的γ链,在模拟处理的细胞中8%)(图33)。这些结果证明由WT化脓链球菌Cas9蛋白与截短的gRNA复合构成的RNP可以增加近端结合的Cpf1-RNP的编辑,而不会在其自身的结合位点引入可检测的编辑水平,也不会明显影响Cpf1-RNP产生的indel的长度或方向。在这里,Sp182-Cas9-RNP提供的编辑增强使得能在HBG启动子处对HBG1-1-AsCpf1H800A-RNP进行高编辑,其中高频率的indel破坏远端CCAAT盒靶基序并导致电穿孔细胞的混合红系后代中治疗相关水平的γ链表达。

实例18:用靶向远端CCAAT盒的Cpf1 gRNA编辑的细胞群内的克隆HbF分布

[0405] 接下来进行单细胞实验以评估在用HBG1-1-AsCpf1H800A-RNP(由His-AsCpf1-

sNLS-sNLS-H800A(SEQ ID NO:1032)与表14显示的具有3'修饰的HBG1-1 gRNA(SEQ ID NO:1041)复合构成)+Sp182-Cas9-RNP组合电穿孔的mPB CD34+细胞衍生的红系细胞中γ链表达水平的分布(图27F,表14)。在电穿孔后48小时过程中使其恢复后,通过荧光激活细胞分选术(FACS)以1个细胞/孔在非组织培养处理的384孔板中分选细胞。然后使细胞分化并克隆扩增18天成为红系谱系(改编自Giarratana 2011)。进行了UPLC分析,以确定衍生自最初编辑的mPB-CD34+细胞总群体的细胞的克隆性红系后代中γ链表达水平的分布(γ链/[总β样链]的百分比)。认为为了获得功能性益处并减轻与镰状细胞病相关的症状,约30%的红系细胞的胎儿血红蛋白水平应大于30%。在分析的克隆中,83.1%的克隆的γ链水平高于从模拟电穿孔的mPB CD34+细胞衍生的红系细胞克隆中检测到的中值水平(中值=18.8%),64.2%的克隆的γ链水平超过了总β样链的30%并且35.8%的克隆的γ链水平超过了总β样链的48.8%(对照细胞中值水平为30%+18.8%)(图34)。

实例19:靶向HBG启动子区的Cpf1 gRNA的筛选

[0406] 为鉴定可以用作单个RNP的组分或与"增强元件"组合以增加CD34+细胞中HBG启动子区的编辑并诱导修饰的细胞的红系后代中胎儿血红蛋白表达的其他AsCpf1 gRNA,设计了靶向HBG启动子的几个结构域(表17)的His-AsCpf1-NLS-NLS("AsCpf1",SEQ ID NO:1000);AsCpf1 S542R/K607R("AsCpf1 RR",SEQ ID NO:1036);或AsCpf1 S542R/K548V/N552R("AsCpf1RVR",SEQ ID NO:1037)gRNA序列(在表18中列出)。AsCpf1 RR和AsCpf1 RVR是工程改造的AsCpf1变体,其分别识别TYCV/ACCC/CCCC和TATV/RATR PAM(Gao2017)。

表17:HBG基因组区的亚结构域

HbG*的基因组坐标	核苷酸	区域的名称
Chr 11 (NC_000011.10): 5,247,883-5,248,186	TCCTAAAGCT TGGAACACTT TCCCTTCCTT AAGAACCATC CTTGCTACTC AGCTGCAATC AATCCAGCCC CCAGGTCTTC ACTGAACCTT TTCCCATCTC TTCCAAAACA TCTGTTTCTG AGAAGTCCTG TCCTATAGAG GTCTTTCTTC CCACCGGATT TCTCCTACAC CATTTACTCC CACTTGCAGA ACTCCCGTGT ACAAGTGTCT TTACTGCTTT TATTTGCTCA TCAAAATGCA CATCTCATAT AAAAATAAAT GAGGAGCATG CACACACCAC AAACACAAAC AGGCATGCAGAAAT (SEQ ID NO: 1118)	区域 1：*HBG1* 的下游
Chr 11 (NC_000011.10): 5,248,509 - 5,249,173	ATAAAGATGA ACCCATAGTG AGCTGAGAGC TCCAGCCTGG CCTCCAGATA ACTACACACC AAGCTTCCAC CCAGAATCAA GCCTATGTTA ACTTCCCTCA AAGCCTGAGA TTTTGCCTTC CCATTAAATG CAGGTAGTTG TTCCCCTTCA AGCACTAGTC ACTGGCCATA ATTTAAATCT TGCTATCTTC TTGCCACCAT GAACCCTGTA TGTTGTAGGC TGAAGACGTT AAAAGAAACA CACGCTGACA CACACACACA CACGCGCGCG CGCACACACA CACACACACA CAGAGCTGAC TTTCAAAATC TACTCCAGCC CAAATGTTTC AATTGTTCCT CACCCCTGGA CATACTTTGC CCCCATCTGG AATTAAAGGA TATAAGTTTG TAATGAAGCA TTAGCAGCAT TTTATATGTG TCCAGCTGAT ATAGGAATAG CCTTAGCAAT GTATGTTTGG CCACCAAAGT TCCCCACTTT GACTGAGCCA ATATATGCCT TCTGCCTGCA TCTTTTTAAC GACCATACTT GTCCTGCCTC CAGATAGATG TTTTAAAACA ACAAAAATGA GGGAAAGATG AAAGTTCTTT CTACTGGAAT CTAATAAAGA AAAGTCATTT TCCTCATTTC CACCTCTCTT TTCTCAAAGT CAAAATTGTC CATCT (SEQ ID NO: 1119)	区域 2: *HBG1* 内含子 2 - A
Chr 11 (NC_000011.10): 5,249,198 - 5,249,362	CCCTAAAACA TTACCACTGG GTCTCAGCCC AGTTAGTCCT CTGCAGTTTC TTCACCCCCA	区域 3: *HBG1* 内含子 2 - B

	ACCCCAGTAT CTTCAAACAG CTCACACCCT GCTGTGCTCA GATCAATACT CCGTTGTCTA AGTTGCCTCG AGACTAAAGG CAACAGGGCT GAAACATCTC CTGGA (SEQ ID NO: 1120)	
Chr 11 (NC_000011.10): 5,249,591- 5,249,712	CTGTGAGATT GACAAGAACA GTTTGACAGT CAGAAGGTGC CACAAATCCT GAGAAGCGAC CTGGACTTTT GCCAGGCACA GGGTCCTTCC TTCCCTCCCT TGTCCTGGTC ACCAGAGCCT AC (SEQ ID NO: 1121)	区域 4: *HBG1* 内含子 1
Chr 11 (NC_000011.10): 5,249,904 - 5,249,927	GCCGCCGGCC CCTGGCCTCA CTGG (SEQ ID NO: 1122)	区域 5: *HBG1* 的从转录起始位点（TSS）起的 -60 nt 区
Chr 11 (NC_000011.10): 5,249,955 - 5,249,987	CCTTGTCAAG GCTATTGGTC AAGGCAAGGC TGG (SEQ ID NO: 1123)	区域 6: *HBG1* 的从 TSS 起的 -110 nt 区
Chr 11 (NC_000011.10): 5,250,040 - 5,250,075	TGAGATAGTG TGGGGAAGGG GCCCCC AAGAGGATAC (SEQ ID NO: 1124)	区域 7: *HBG1* 的从 TSS 起的 -200 nt 区
Chr 11 (NC_000011.10): 5,250,089 - 5,250,129	TATAGCCTTT GCCTTGTTCC GATTCAGTCA TTCCAGTTTT T (SEQ ID NO: 1125)	区域 8: *HBG1* 的从 TSS 起的 -250 nt 区
Chr 11 (NC_000011.10): 5,250,141 - 5,250,254	TCTTCCCTTT AGCTAGTTTC CTTCTCCCAT CATAGAGGAT ACCAGGACTT CTTTTGTCAG CCGTTTTTTA CCTTCTTGTC TCTAGCTCCA GTGAGGCCTG TAGTTTAAAG CTAA (SEQ ID NO: 1126)	区域 9: *HBG1* 的从 TSS 起的 -333 nt 区
Chr 11 (NC_000011.10): 5,250,464 - 5,250,549	CCACAGTTTC AGCGCAGTAA TAGATTAGTG TTACATAATA TAAGACCTAA TGCTTACCTC AATATCTACT TATCCGTACC TATTTG (SEQ ID NO: 1127)	区域 10: *HBG1* 的从 TSS 起的-650 nt 区
Chr 11 (NC_000011.10): 5,250,594 - 5,250,735	TATTCAGGTA TGTATGTATA CACCAGATGA TGTGTATTTA CCACTGGATA AGTGTGTGTG CTGGCTGATG ACCCAGGGTT TTGGCGTAGC TCTTCTATGC TCAGTAAAGA TGATGGTAGA ATGTTCTTTG GCAGGTACTG TG (SEQ ID NO: 1128)	区域 11: *HBG1* 的从 TSS 起的-800 nt 区
Chr 11 (NC_000011.10): 5,253,425 - 5,254,121	CAATAAAGAT GAACCCATAG TGAGCTGAGA GCTCCAGCCT GGCCTCCAGA TAACTACACA CCAAGCTTCC ACCCAGAATC AAGCCTATGT TAACTTCCCT CAAAGCCTGA GATTTTGCTT TCCCATTAAA TGCAGGTAGT TGTTCTTCTT GCAGCACTAG TCACTGGCCA TAATTTAAAT CTTGTTATCT TCTTGCCACC ATGAACCCTG TATGCTGTAG GCTGAAAACG TTAAAAGAAA	区域 12: *HBG2* 内含子 2 - A

	CACACGCTCT CACACACACA CAAACACACG CGCGCACACA CACACACACA CACACAGAGC TGACTTTCAA AATCTACTCC AGCCCAAATG TTTCAATTGT TCCTCACCCC TGGACATACT TTGCCCCCAT CTGGAATTAA AGGATATAAG TTTGTAATGA AGCATTAGCA GCATTTTATA TGTGTCCAGC TGATATAGGA ATAGCCTTAG CAATGTATGT TTGGCCACCA AAGTTCCCCA CTTTGACTGA GCCAATATAT GCCTTCTGCC TGCATCTTTT TAATGACCAT ACTTGTCCTG CCTCCAGATA GATGTTTTAA AACGAATAAC AAAAATAGGG GAAAGGTGAA AGTTCTTTCT ACCGAAATCT AATAAAGAAA AGTCATTTTC CTCATTTCCA CCTCTCTTTT CTCAAAGTCA AAGTTGTCCA TCTAGATTTT CAGAGGCACT CCTTAGG (SEQ ID NO: 1129)	
Chr 11 (NC_000011.10): 5,254,122 - 5,254,306	CCCTAAAACA TTGCCACTGG GTCTCAGCCC AGTTAGTCCT CTGCAGTTTC TTCACTCCCA ACCCCAGTAT CTTCAAACAG CTCACACCCT GCTGTGCTCA GATCAATACT CAGTTGTCTA AGTTGCCTCG AGACTAAAGG CAACAGTGCT GAAACATCTC CTGGACTCAC CTTGAAGTTC TCAGG (SEQ ID NO: 1130)	区域 13: *HBG2* 内含子 2 - B
Chr 11 (NC_000011.10): 5,254,511 - 5,254,648	AGCCTGTGAG ATTGACAAGA ACAGTTTGAC AGTCAGAAGG TGCCACAAAT CCTGAGAAGC GACCTGGACT TTTGCCAGGC ACAGGGTCCT TCCTTCCCTC CCTTGTCCTG GTCACCAGAG CCTACCTTCC CAGGGTT (SEQ ID NO: 1131)	区域 14: *HBG2* 内含子 1
Chr 11 (NC_000011.10): 5,254,829 - 5,254,866	CCGCCGGCCC CTGGCCTCAC TGGATACTCT AAGACTAT (SEQ ID NO: 1132)	区域 15: *HBG2* 的从 TSS 起的-60 nt 区
Chr 11 (NC_000011.10): 5,254,879 - 5,254,909	CCTTGTCAAG GCTATTGGTC AAGGCAAGGC T (SEQ ID NO: 1133)	区域 16: *HBG2* 的从 TSS 起-110 nt 区
Chr 11 (NC_000011.10): 5,254,935 5,255,009	CAGGGACCGT TTCAGACAGA TATTTGCATT GAGATAGTGT GGGGAAGGGG CCCCCAAGAG GATACTGCTG CTTAA (SEQ ID NO: 1134)	区域 17: *HBG2* 的从 TSS 起的-200 nt 区

Chr 11 (NC_000011.10): 5,255,025 - 5,255,053	TTGCCTTGTT CCGATTCAGT CATTCCAAT (SEQ ID NO: 1135)	区域 18: *HBG2* 的从 TSS 起的-250 nt 区
Chr 11 (NC_000011.10): 5,255,076 - 5,255,179	TTTAGCTAGT TTTCTTCTCC CACCATAGAA GATACCAGGA CTTCTTTTGT CAGCCGTTTT TCACCTTCTT GTCTGTAGCT CCAGTGAGGC CTGTAGTTTA AAGT (SEQ ID NO: 1136)	区域 19: *HBG2* 的从 TSS 起的-330 nt 区
Chr 11 (NC_000011.10): 5,255,255 - 5,255,292	GGACACGTCT TAGTCTCATT TAGTAAGCAT TGGTTTCC (SEQ ID NO: 1137)	区域 20: *HBG2* 的从 TSS 起的-500 nt 区
Chr 11 (NC_000011.10): 5,255,518 - 5,255,641	TTTTTTATAT TCAGGTATGT ATGTAGGCAC CCGATGATGT GTATTTATCA CTGGATAAGT GTATGTGCTG GCTGATGACC CAGGGTTTTG GTGTAGCTCT TCTATGCTCG GTAAAGATGA TGGT (SEQ ID NO: 1138)	区域 21: *HBG2* 的从 TSS 起的-800 nt 区
*NCBI 参考序列 NC_000011，使用基于一个的坐标系统（One-based coordinate system）报告坐标，“智人染色体 11，GRCh38.p12 初级组装”（版本 NC_000011.10）。		

表18:Cpf1指导RNA

gRNA 靶向结构域序列 ID*	gRNA 靶向结构域序列（RNA）	gRNA 靶向结构域序列（DNA）	在 HbG1**处的基因组坐标	在 HbG2**处的基因组坐标	%编辑	链
AsCpf1 HBG1 启动子-1	AGACAGAUA UUUGCAUUG AG (SEQ ID NO: 1139)	AGACAGATAT TTGCATTGAG (SEQ ID NO: 1140)	Chr11:5250024:5250043	Chr11:5254948:5254967	5.43	+
AsCpf1 HBG1 启动子-2	CAUUGAGAU AGUGUGGGG AA (SEQ ID NO: 1141)	CATTGAGATA GTGTGGGGAA (SEQ ID NO: 1142)	Chr11:5250037:5250056	Chr11:5254961:5254980	8.30	+
AsCpf1 HBG1 启动子-3	UAGCCUUUGC CUUGUUCCGA (SEQ ID NO: 1143)	TAGCCTTTGCC TTGTTCCGA (SEQ ID NO: 1144)	Chr11:5250091:5250110	Chr11:5255019:5255038	0.23	+
AsCpf1 HBG1 启动子-4	CCUUGUUCCG AUUCAGUCA U (SEQ ID NO: 1145)	CCTTGTTCCGA TTCAGTCAT (SEQ ID NO: 1146)	Chr11:5250100:5250119	Chr11:5255028:5255047	1.15	+
AsCpf1 HBG1 启动子-5	UCUAAUUUA UUCUUCCCUU U (SEQ ID NO: 1147)	TCTAATTTATT CTTCCCTTT (SEQ ID NO: 1148)	Chr11:5250131:5250150	Chr11:5255059:5255078	0.16	+
AsCpf1 HBG1 启动子-6	CUUCUCCCAU CAUAGAGGA U (SEQ ID NO: 1149)	CTTCTCCCATC ATAGAGGAT (SEQ ID NO: 1150)	Chr11:5250161:5250180		12.73	+
AsCpf1 HBG2 启动子-7	UUCUCCCACC AUAGAAGAU	TTCTCCCACCA TAGAAGATA		Chr11:5255090:5255109	8.11	+

	A (SEQ ID NO: 1151)	(SEQ ID NO: 1152)				
AsCpf1 HBG1 启动子-8	CCACUGGAUA AGUGUGUGU G (SEQ ID NO: 1153)	CCACTGGATA AGTGTGTGTG (SEQ ID NO: 1154)	Chr11:525063 4:5250653		13.33	+
AsCpf1 HBG1 启动子-9	GCGUAGCUCU UCUAUGCUCA (SEQ ID NO: 1155)	GCGTAGCTCT TCTATGCTCA (SEQ ID NO: 1156)	Chr11:525067 7:5250696		13.48	+
AsCpf1 HBG1 启动子-10	CUGAGCAUA GAAGAGCUA CG (SEQ ID NO: 1157)	CTGAGCATAG AAGAGCTACG (SEQ ID NO: 1158)	Chr11:525067 8:5250697		10.73	-
AsCpf1 HBG2 启动子-11	UCACUGGAU AAGUGUAUG UG (SEQ ID NO: 1159)	TCACTGGATA AGTGTATGTG (SEQ ID NO: 1160)		Chr11:52555 65:5255584	0.43	+
AsCpf1 HBG2 启动子-12	GUGUAGCUC UUCUAUGCUC G (SEQ ID NO: 1161)	GTGTAGCTCTT CTATGCTCG (SEQ ID NO: 1162)		Chr11:52556 08:5255627	5.78	+
AsCpf1 HBG2 启动子-13	CCGAGCAUAG AAGAGCUACA (SEQ ID NO: 1163)	CCGAGCATAG AAGAGCTACA (SEQ ID NO: 1164)		Chr11:52556 09:5255628	3.24	-
HBG1-1 AsCpf1	CCUUGUCAAG GCUAUUGGU C (SEQ ID NO: 1002)	CCTTGTCAAG GCTATTGGTC (SEQ ID NO: 1003)	Chr11:524995 5:5249974	Chr11:52548 79:5254898	17.96	+
AsCpf1 RR HBG1 启动子-1	GACAGAUAU UUGCAUUGA GA (SEQ ID NO: 1167)	GACAGATATTT GCATTGAGA (SEQ ID NO: 1168)	Chr11:525002 5:5250044	Chr11:52549 49:5254968	8.48	+
AsCpf1 RR HBG1 启动子-2	ACACUAUCUC AAUGCAAAU A (SEQ ID NO: 1169)	ACACTATCTCA ATGCAAATA (SEQ ID NO: 1170)	Chr11:525003 1:5250050	Chr11:52549 55:5254974	0.09	-
AsCpf1 RR HBG1 启动子-3	CACACUAUCU CAAUGCAAA U (SEQ ID NO: 1171)	CACACTATCTC AATGCAAAT (SEQ ID NO: 1172)	Chr11:525003 2:5250051	Chr11:52549 56:5254975	2.10	-
AsCpf1 RR HBG1 启动子-4	CCACACUAUC UCAAUGCAAA (SEQ ID NO: 1173)	CCACACTATCT CAATGCAAA (SEQ ID NO: 1174)	Chr11:525003 3:5250052	Chr11:52549 57:5254976	2.52	-
AsCpf1 RR HBG1 启动子-5	UUCCCCACAC UAUCUCAAU G (SEQ ID NO: 1175)	TTCCCCACACT ATCTCAATG (SEQ ID NO: 1176)	Chr11:525003 7:5250056	Chr11:52549 61:5254980	0.05	-
AsCpf1 RR HBG1 启动子-6	GAUUCAGUC AUUCCAGUU UU (SEQ ID NO: 1177)	GATTCAGTCAT TCCAGTTTT (SEQ ID NO: 1178)	Chr11:525010 9:5250128		0.77	+
AsCpf1 RR	AUUCAGUCA	ATTCAGTCATT	Chr11:525011		0.24	+

HBG1 启动子-7	UUCCAGUUUUU (SEQ ID NO: 1179)	CCAGTTTTT (SEQ ID NO: 1180)	0:5250129			
AsCpf1 RR HBG1 启动子-8	GUCAUUCCAGUUUUUCUCUA (SEQ ID NO: 1181)	GTCATTCCAGTTTTTCTCTA (SEQ ID NO: 1182)	Chr11:5250115:5250134		1.00	+
AsCpf1 RR HBG1 启动子-9	AGUUUUUCUCUAAUUUAUUC (SEQ ID NO: 1183)	AGTTTTTCTCTAATTTATTC (SEQ ID NO: 1184)	Chr11:5250123:5250142		0.15	+
AsCpf1 RR HBG1 启动子-10	GUUUUUCUCUAAUUUAUUCU (SEQ ID NO: 1185)	GTTTTTCTCTAATTTATTCT (SEQ ID NO: 1186)	Chr11:5250124:5250143		0.15	+
AsCpf1 RR HBG2 启动子-11	GAUUCAGUCAUUCCAAUUUU (SEQ ID NO: 1187)	CAAGAGGATACTGCTGCTTA (SEQ ID NO: 1188)		Chr11:5254989:5255008	***	+
AsCpf1 RR HBG2 启动子-12	AUUCAGUCAUUCCAAUUUUU (SEQ ID NO: 1189)	AAGAGGATACTGCTGCTTAA (SEQ ID NO: 1190)		Chr11:5254990:5255009	***	+
AsCpf1 RR HBG2 启动子-13	GUCAUUCCAAUUUUUCUCUA (SEQ ID NO: 1191)	GATTCAGTCATTCCAATTTT (SEQ ID NO: 1192)		Chr11:5255037:5255056	0.25	+
AsCpf1 RR HBG2 启动子-14	AAUUUUUCUCUAAUUUAUUC (SEQ ID NO: 1193)	ATTCAGTCATTCCAATTTTT (SEQ ID NO: 1194)		Chr11:5255038:5255057	0.13	+
AsCpf1 RR HBG2 启动子-15	AUUUUUCUCUAAUUUAUUCU (SEQ ID NO: 1195)	GTCATTCCAATTTTTCTCTA (SEQ ID NO: 1196)		Chr11:5255043:5255062	0.32	+
AsCpf1 RR HBG2 启动子-16	UUCUCCCAUCAUAGAGGAUA (SEQ ID NO: 1197)	AATTTTTCTCTAATTTATTC (SEQ ID NO: 1198)		Chr11:5255051:5255070	0.14	+
AsCpf1 RR HBG2 启动子-17	AUCAUAGAGGAUACCAGGAC (SEQ ID NO: 1199)	ATTTTTCTCTAATTTATTCT (SEQ ID NO: 1200)		Chr11:5255052:5255071	0.10	+
AsCpf1 RR HBG1 启动子-18	ACCAUAGAAGAUACCAGGAC (SEQ ID NO: 1201)	TTCTCCCATCATAGAGGATA (SEQ ID NO: 1202)	Chr11:5250162:5250181		1.40	+
AsCpf1 RR HBG1 启动子-19	CAGUACCUGCCAAAGAACAU (SEQ ID NO: 1203)	ATCATAGAGGATACCAGGAC (SEQ ID NO: 1204)	Chr11:5250169:5250188		7.88	+
AsCpf1 RR HBG2 启动子-20	UAGUAUCUGGUAAAGAGCAU (SEQ ID NO: 1205)	ACCATAGAAGATACCAGGAC (SEQ ID NO: 1206)		Chr11:5255097:5255116	13.03	+

AsCpf1 RR HBG1 启动子 -21	UCAAUGCAA AUAUCUGUC UG (SEQ ID NO: 1207)	CAGTACCTGC CAAAGAACAT (SEQ ID NO: 1208)	Chr11:525071 4:5250733		13.31	-
AsCpf1 RR HBG2 启动子 -22	CUCUUGGGG GCCCCUUCCC C (SEQ ID NO: 1209)	TAGTATCTGGT AAAGAGCAT (SEQ ID NO: 1210)		Chr11:52556 45:5255664	4.07	-
AsCpf1 RVR HBG1 启动子-1	GAUUCAGUC AUUCCAAUU UU (SEQ ID NO: 1211)	TCAATGCAAAT ATCTGTCTG (SEQ ID NO: 1212)	Chr11:525002 3:5250042	Chr11:52549 47:5254966	0.15	-
AsCpf1 RVR HBG1 启动子-2	AUUCAGUCA UUCCAAUUU UU (SEQ ID NO: 1213)	CTCTTGGGGG CCCCTTCCCC (SEQ ID NO: 1214)	Chr11:525005 1:5250070	Chr11:52549 75:5254994	1.09	-
AsCpf1 RVR HBG1 启动子-3	AAAAAAAUU AGCAGUAUCC U (SEQ ID NO: 1215)	AAAAAAATTA GCAGTATCCT (SEQ ID NO: 1216)	Chr11:525006 9:5250088		***	-
AsCpf1 RVR HBG1 启动子-4	GCCUUUGCCU UGUUCCGAU U(SEQ ID NO: 1217)	GCCTTTGCCTT GTTCCGATT (SEQ ID NO: 1218)	Chr11:525009 3:5250112	Chr11:52550 21:5255040	3.96	+
AsCpf1 RVR HBG2 启动子-5	AAAAAAAUU AAGCAGCAG UA (SEQ ID NO: 1219)	AAAAAAATTA AGCAGCAGTA (SEQ ID NO: 1220)		Chr11:52549 97:5255016	***	-
AsCpf1 RVR HBG1 启动子-6	CUCAGUAAA GAUGAUGGU AG (SEQ ID NO: 1221)	CTCAGTAAAG ATGATGGTAG (SEQ ID NO: 1222)	Chr11:525069 3:5250712		5.32	+
AsCpf1 RVR HBG2 启动子-7	ACUGGAUAA GUGUAUGUG CU (SEQ ID NO: 1223)	ACTGGATAAG TGTATGTGCT (SEQ ID NO: 1224)		Chr11:52555 67:5255586	9.78	+
AsCpf1 RVR HBG2 启动子-8	UGCUGGCUG AUGACCCAGG G (SEQ ID NO: 1225)	TGCTGGCTGA TGACCCAGGG (SEQ ID NO: 1226)	Chr11:525065 2:5250671	Chr11:52555 83:5255602	0.24	+
AsCpf1 RVR HBG2 启动子-9	CUCGGUAAA GAUGAUGGU AG (SEQ ID NO: 1227)	CTCGGTAAAG ATGATGGTAG (SEQ ID NO: 1228)		Chr11:52556 24:5255643	5.75	+
AsCpf1 RVR HBG2 启动子 -10	UGGUAAAGA GCAUUCUACC A (SEQ ID NO: 1229)	TGGTAAAGAG CATTCTACCA (SEQ ID NO: 1230)		Chr11:52556 38:5255657	8.55	-

*gRNA ID 名称提供了 RNP 复合物中使用的特定 Cpf1 分子
**NCBI 参考序列 NC_000011，使用基于一个的坐标系统报告坐标，“智人染色体 11，GRCh38.p12 初级组装”（版本 NC_000011.10）。
***代表未测试的 gRNA。

[0407] 将与包含来自表18的gRNA靶向结构域的单gRNA(对于使用的特定Cpf1分子,参见gRNA ID名称)复合的含有AsCpf1蛋白(SEQ ID NO:1000)、AsCpf1RR蛋白(SEQ ID NO:1036)或AsCpf1 RVR(SEQ ID NO:1037)的RNP(5μM)使用Amaxa电穿孔仪设备(龙沙公司)递送至动员的外周血(mPB)CD34+细胞。72小时后,从细胞中提取基因组DNA,并通过PCR扩增的靶位点的Illumina测序(NGS)分析靶位点处的插入/缺失水平。每种gRNA的编辑百分比(indel=缺失和插入)显示于上表18中。在某些实施例中,包含表18中所示的一种或多种gRNA的Cpf1 RNP可以用于靶向表17中列出的区域以诱导HbF表达,并且可以与“加强元件”共递送以与单

独的Cpf1 RNP的编辑水平相比实现更高的编辑水平。

实例20:靶向CCAAT盒的HBG1-1-Cpf1 RNP与ssODN的共递送支持人造血干/祖细胞中基因编辑的增加

[0408] 已经证明与Cas9 RNP共递送的编码“4nt缺失”(HBGΔ-112:-115)的100nt ssODN(50/50同源)供体模板产生与更长的180nt模板可比较的结果(实例10和图23),将产生“18nt缺失”(HBGΔ-104:-121)的100nt ssODN(即ssODN OLI16431(SEQ ID NO:1040),表11)与Cpf1 RNP共递送,以进一步研究编辑结果。

[0409] 简而言之,将在补充了人细胞因子的培养基中预刺激两天的成人mPB CD34+细胞用RNP电穿孔,所述RNP包含与经修饰的HBG1-1 gRNA(SEQ ID NO:1041,表14)复合的His-AsCpf1-sNLS-sNLS H800A蛋白(SEQ ID NO:1032,表14)(“His-AsCpf1-sNLS-sNLS H800A_HBG1-1 RNP”)(单独的(6μM)或跟OLI16431(0.5-6μM)组合)。RNP和编码18nt缺失的具有正链同源臂的ssODN供体(OLI16431)的共递送将编辑频率从没有供体情况下的8.0%提高到75.6%,如通过电穿孔后72小时对提取的基因组DNA的HBG PCR产物的测序分析确定的(图35A)。这种编辑水平使HbF水平比背景值增加约24%(图35A)。另外,应当注意,在任何测试剂量下,通过DAPI排除或集落形成潜力(图35C)测得的观察到的细胞生存力未降低(图35B)。

实例21:优化HBG1-1-Cpf1 RNP和OLI16431的剂量,以最大化RNP靶向的远端CCAAT盒位点的编辑

[0410] 当共递送ssODN与RNP时,已显示出增加的编辑(实例9-11、13、20),使用相同的方法来优化每个组分的剂量,以使总编辑最大化。简而言之,用RNP建立剂量矩阵,所述RNP包含与未经修饰的HBG1-1 gRNA(SEQ ID NO:1022,表14)复合的His-AsCpf1-sNLS-sNLS H800A蛋白(SEQ ID NO:1032,表14)(“His-AsCpf1-sNLS-sNLS H800A_HBG1-1 RNP”),与0-12μM的OLI16431(SEQ ID NO:1040,表11)以0-12μM共递送。使用与8μM OLI16431共递送的8μM剂量的His-AsCpf1-sNLS-sNLS H800A_HBG1-1 RNP,当通过对从红系培养14天后从细胞中提取的基因组DNA的HBG PCR产物进行测序分析进行测量时,实现89.4%的indel最大值(图36A)。在这些剂量条件下,这与HbF水平增加到比背景高>32%一致(图36A)。该样品(His-AsCpf1-sNLS-sNLS H800A_HBG1-1 RNP[8μM]+OLI16431[8μM]的生存力在电穿孔后48小时及红系培养14天后均与未经处理的对照样品相当(图36B)。

[0411] 然而,如在实例13中所避开的那样,当单独将>8μM的OLI16431电穿孔进入细胞而没有RNP时,观察到假象。在这些条件下,电穿孔后48小时进行PCR扩增时会产生假阳性结果,这很可能是由于系统中过多的ssODN(参见图36A,对于只有OLI16431的情况在48小时处为8μM和12μM(黑条))。在较低剂量的ssODN和以后的时间点,这种假阳性现象不再明显(图36A,对于只有OLI16431的情况在48小时处为4μM和6μM(黑条)并且对于只有OLI16431的情况在14天处为4μM、6μM、8μM和12μM(浅灰色条))。另外,在单独使用ssODN的组中观察到的细胞生存力下降(图36B,在仅使用12μM OLI16431情况下在48小时处下降至约57%(黑条))可能是复合因素。ssODN剂量为4μM和6μM时没有假阳性结果,表明该假象可能是由于过量的ssODN。这些因素,以及分别在只有8μM和12μM OLI16431的组中观察到的基线HbF水平(图36A,灰点,如与阴性对照相比)提供了把握,即随着添加ssODN,RNP编辑的实质性和持续性增强是真实的,而不是假像。

实例22:含有多种Cpf1和靶向HBG启动子区域的gRNA的RNP支持人造血干/祖细胞中的基因编辑,所述基因编辑促进红系后代中HbF蛋白表达的诱导

[0412] 将包含SEQ ID NO:1022、1023、1041-1093、1098-1106(表19)的指导RNA与各种Cpf1变体酶(SEQ ID NO:1032、1094-1097、1107-1109,表20)复合,形成各种RNP复合物(表21)。RNP含有具有gRNA的5'末端修饰和/或3'末端的修饰的gRNA(表19)。包含SEQ ID NO:1022、1023、1041-1084、1098-1106(表19)的指导RNA在远端CCAAT盒靶区域具有相同的预期切割位点,相关的靶向结构域含有SEQ ID NO:1002(HBG1-1)、SEQ ID NO:1254(HBG1-1-21mer)、SEQ ID NO:1256(HBG1-1-22mer)、SEQ ID NO:1258(HBG1-1-23mer)(分别针对包含20mer、21mer、22mer或23mer原间隔子序列的gRNA)所示序列(表22,表23)。在某些情况下,还测试了靶向HBG启动子内其他位置的gRNA,包括指导RNA SEQ ID NO:1085-1096,其包含含有SEQ ID NO:1260(AsCpf1 HBG1启动子1(21mer))、SEQ ID NO:1262(AsCpf1 HBG1启动子2(21mer))或SEQ ID NO:1264(AsCpf1 HBG1启动子6(21mer))所示的序列的靶向结构域(表22,表23)。表21提供了在实例22-24中测试的每种RNP的列表以及形成每种RNP复合物的gRNA的SEQ ID NO和Cpf1变体的SEQ ID NO。有关每种gRNA和Cpf1变体的其他信息,分别可以在表19和表20中找到。

[0413] 实例22和23中使用的gRNA是化学合成的。用于寡核苷酸合成的化学品购自生物自形成公司(BioAutomation)、格兰研究公司(Glen Research)、密理博西格玛公司(Millipore Sigma)、西格玛奥德里奇公司(Sigma-Aldrich)、化学基因公司(ChemGenes)和赛默飞世尔科技公司(Thermo Fisher Scientific)。用于合成的固体支持物是ChemGenes的Unylinker2000 ÅCPG树脂、2'-TBDMS rU2000 ÅCPG树脂或2'-O-甲基腺苷(N-Bz)Icaa 2000 ÅCPG树脂。RNA(TBDMS保护的)和DNA亚磷酰胺购自赛默飞世尔科技公司。在某些实施例中,在硫化步骤期间用来自格兰研究公司的DDTT(3-((二甲基氨基-亚甲基)氨基)-3H-1,2,4-二噻唑-3-硫酮)溶液引入硫代磷酸酯。使用标准RNA和DNA亚磷酰胺化学方法,在BioAutomation MerMade 12合成仪或GE ÄktaOligoPilot 100合成仪上合成寡核苷酸。合成后,将寡核苷酸从固体支持物上切割下来,并使用氢氧化铵/甲胺和TEA-3HF在两步过程中脱保护。脱盐后,在制备型HPLC上使用反相色谱法纯化寡核苷酸。

[0414] 首先,测试了使用包含gRNA(具有gRNA的5'末端修饰)的RNP进行编辑的效果。简而言之,在补充有SCF、TPO和FLT3的X-Vivo 10培养基中进行预刺激48小时后,通过MaxCyte电穿孔将RNP复合物(6.0μM和12μM)递送至1x 10^6mPB CD34+细胞。在电穿孔后72小时,通过对PCR扩增的靶位点的Illumina测序(NGS)分析靶位点处的插入/缺失水平。结果证明,包含不具有修饰的gRNA的RNP(RNP33,也称为HBG1-1-AsCpf1 RNP,表14)或包含具有5'末端修饰(包括添加5nt RNA(RNP37),10nt RNA(RNP38),25nt RNA(RNP39),60nt RNA(RNP40),5nt DNA(RNP41),10nt DNA(RNP42),25nt DNA(RNP43)和60nt DNA(RNP44))的gRNA的RNP(表21)支持中靶编辑(图37)。

[0415] 接下来,测试了Cpf1 RNP与Sp182失活RNP(包含与化脓链球菌Cas9(SEQ ID NO:1033)复合的SEQ ID NO:1027(表14)的失活gRNA)或ssODN OLI16431(SEQ ID NO:1040,表11)的共递送的效果。简而言之,在补充有SCF、TPO和FLT3的X-Vivo 10培养基中进行48小时预刺激后,经由MaxCyte电穿孔将具有不同浓度(6μM、8μM、8μM和12μM)的RNP33(无5'或3'

gRNA修饰,表21)复合物的给药矩阵与Sp182 RNP(8μM、8μM、6μM和4μM)或ssODN OLI16431(8μM、8μM、6μM和4μM)共递送到5.25x 10^6mPB CD34+细胞。电穿孔后,将细胞放回培养另外48小时。然后,将一部分CD34细胞分离,以进行gDNA提取并在混合细胞群体中进行indel定量。此外,为了调查表型祖细胞和表型造血干细胞(HSC)的编辑,通过免疫表型分型(Notta 2011)表征了HSPC亚群(在其余的CD34细胞中),并通过荧光激活细胞分选(FACS)进行分离。通过用针对hCD34、hCD38、hCD45RA、hCD90和hCD123的抗体染色细胞,在电穿孔后的48小时进行免疫表型分型。表型HSC定义为hCD34亮hCD38 hCD90+hCD45RA-,祖细胞定义为hCD34亮CD38+。通过FACS对这两个群体进行分选,并提取DNA以确定这些亚群中的编辑水平。通过PCR扩增的靶的Illumina测序(NGS)分析靶标位点处的插入/缺失水平。结果表明,RNP33与"加强元件"Sp182 RNP或ssODN OLI16431的共递送提供支持中靶编辑(图38A-B),其水平高于单独递送RNP33时观察到的水平(图37)。

[0416] 接下来,测试了RNP33(无5'或3'gRNA修饰,表21)、RNP43(+25DNA5'gRNA修饰,表21)或RNP34(1xPS2-OMe+1xOMe 3'gRNA,表21)与Sp182 dgRNA(表15)或ssODN OLI16431(表11)的共递送的效果。简而言之,在补充有SCF、TPO和FLT3的X-Vivo 10培养基中进行预刺激48小时后,通过MaxCyte电穿孔将RNP33、RNP34或RNP43复合物(8μM)与Sp182 RNP(8μM)或ssODN OLI16431(8μM)共递送至5.25x 10^6mPB CD34+细胞。在电穿孔后,将细胞放回培养另外48小时,然后分类为表型祖细胞和表型HSC级分以进行indel定量。在电穿孔后48小时,通过对PCR扩增的靶位点的Illumina测序(NGS)分析靶位点处的插入/缺失水平。结果证明,与"加强元件"Sp182 RNP或ssODN OLI16431共递送的含有带有5'修饰或3'修饰的gRNA的RNP支持中靶编辑(图39)。测试的两个加强元件均提供了与RNP34和RNP33可比较的编辑水平。当与不具有加强元件情况下的编辑水平进行比较时,虽然两个加强元件都增强了RNP33编辑,但是RNP43编辑仅在添加Sp182 RNP情况下增加(图37、39)。

[0417] 接下来,测试了不同的Cpf1蛋白对RNP编辑的影响。用包含各种Cpf1蛋白的RNP进行了化学计量比较(gRNA :Cpf1)。简而言之,RNP(RNP64、RNP63、RNP45,表21)以2或4的化学计量(gRNA:Cpf1复合比)递送,其中gRNA摩尔过量。在补充有SCF、TPO和FLT3的X-Vivo 10培养基中进行预刺激48小时后,所有RNP均通过MaxCyte电穿孔以8μM递送至1x 10^6CD34+细胞。电穿孔后,在indel定量之前,将细胞放回培养另外72小时。结果表明,在测试浓度下,化学计量比改变时编辑率没有差异(图40)。RNP45(包含Cpf1蛋白SEQ ID NO:1094)优于RNP63(包含Cpf1蛋白SEQ ID NO:1095)和RNP64(包含Cpf1 SEQ ID NO:1109)。

[0418] 接下来,测试了Sp182 RNP或ssODN OLI16431与含有不同Cpf1蛋白的各种RNP的共递送的效果。简而言之,RNP(RNP33、RNP64、RNP63、RNP45,表21)单独递送或与Sp182 RNP或ssODN OLI16431组合递送。在补充有SCF、TPO和FLT3的X-Vivo 10培养基中进行预刺激48小时后,所有试剂均通过MaxCyte电穿孔以8μM递送至1x 10^6CD34+细胞。电穿孔后,在indel定量之前,将细胞放回培养另外72小时。结果表明,对于所有测试的RNP,"加强元件"Sp182 RNP或ssODN OLI16431增强编辑(图41)。

[0419] 测试了使用含有具有各种5'DNA延伸的gRNA的RNP或不带5'DNA延伸的RNP(RNP45)与或不与ssODN OLI16431共递送情况下进行编辑的效果。简而言之,在补充有SCF、TPO和FLT3的X-Vivo 10培养基中进行48小时预刺激后,RNP(RNP45(无5'修饰),RNP46、RNP47、RNP48、RNP49、RNP50、RNP51、RNP52、RNP53、RNP54、RNP55、RNP56和RNP57,表21)以8μM的浓度

单独或与8μM ssODN OLI16431一起通过MaxCyte电穿孔递送至1x 10^6CD34+细胞。电穿孔后,在indel定量之前将细胞放回培养。结果证明,所有RNP都支持中靶编辑(图42)。

[0420] 测试了使用RNP(包括具有相同5'DNA延伸但不同3'gRNA修饰的gRNA)进行编辑的效果,以评估3'gRNA修饰的影响。简而言之,在补充有SCF、TPO和FLT3的X-Vivo 10培养基中进行48小时预刺激后,通过MaxCyte电穿孔以8μM的浓度将包含具有匹配的5'末端但不同的3'gRNA修饰的gRNA的RNP(RNP49相比于RNP58和RNP59相比于RNP60)递送至1x 10^6个CD34+细胞。电穿孔后,在indel定量之前将细胞放回培养。在两个比较中,在电穿孔后24小时,含有具有3'PS-OMe的gRNA的RNP优于未经修饰的3'版本(图43)。

[0421] 接下来,针对RNP58测试了不同浓度的Cpf1和gRNA。简而言之,将RNP58(+25DNA 5'gRNA修饰和1xPS-OMe 3'gRNA修饰,表21)通过MaxCyte电穿孔以2:1、1:1或0.5:1的化学计量比(gRNA :Cpf1复合比)在预刺激48小时后递送至1x 10^6个CD34+细胞。电穿孔后,在indel定量之前将细胞放回培养。在所有测试剂量下,当RNP以2:1的比例复合时,编辑效果最佳(图44A-B)。

[0422] 进一步测试了使用RNP(包括具有相同5'DNA延伸但不同3'gRNA修饰的gRNA)进行编辑的效果,以评估3'gRNA修饰的影响。简而言之,预刺激48小时后,通过MaxCyte电穿孔将包含具有匹配的5'末端但不同的3'gRNA修饰的gRNA的RNP(RNP58、RNP2、RNP3、RNP4、RNP5、RNP6、RNP7、RNP8、RNP9、RNP10)以不同的浓度(1μM、2μM、4μM)递送至1x 10^6个CD34+细胞。电穿孔后,在indel定量之前将细胞放回培养。结果证明,所有RNP都支持按中靶编辑(图45A-B)。

[0423] 接下来,测试了由RNP进行的编辑,所述RNP包括靶向HBG启动子各个区域的具有3'gRNA修饰或5'和3'修饰的gRNA。这些包括指导RNA SEQ ID NO:1085-1096,其包含靶向结构域SEQ ID NO:1260(AsCpf1 HBG1启动子-1(21mer))、1262(AsCpf1 HBG1启动子-2(21mer))、1264(AsCpf1 HBG1启动子-6)(21mer))(表22,表23)。代替远端CCAAT盒靶区域,将这些gRNA配置为在选自表17的区域内提供编辑事件。简而言之,在补充有SCF、TPO和FLT3的X-Vivo 10培养基中进行48小时预刺激后,将包括含有未经修饰的5'gRNA和1xPS-OMe 3'gRNA修饰的gRNA的RNP(RNP11、RNP16、RNP19和RNP22,表21)、包括含有+20DNA+2xPS 5'gRNA修饰和1xPS-OMe 3'gRNA修饰的gRNA的RNP(RNP12、RNP21和RNP24,表21)、包括含有+25DNA 5'gRNA修饰和1xPS-OMe 3'gRNA修饰的gRNA的RNP(RNP58和RNP20,表21)通过MaxCyte电穿孔以8μM的浓度递送至1x 10^6个CD34+细胞。在对mPB CD34+细胞进行编辑后,进行了18天的离体分化为红系谱系(Giarratana 2011),其中在培养的第14天分离了gDNA用于indel分析。简而言之,如上所述,将RNP递送至CD34+细胞。在补充有SCF、TPO和FLT3的X-Vivo 10培养基中恢复48小时后,对经处理的细胞进行计数并转移至红系分化培养基中,其中在第4、7、10和14天进行细胞计数和补料,并在第18天进行红系收集。然后对这些CD34+衍生的红系细胞进行计数,并在HPLC级水中溶解,然后过滤除去细胞碎片。然后,通过UPLC测量每个样品的γ-珠蛋白链的相对表达水平(相比于总β样珠蛋白链),其中通过逐渐增加含0.1%三氟乙酸的乙腈与含0.1%三氟乙酸的水的比例来实现蛋白分离(图46A-C)。图46A-C显示了所有RNP支持中靶编辑。但是,仅在某些靶位点进行编辑会导致HBF表达增加。使用不同浓度(0、1μM、2μM、4μM)的RNP58进行了相同的实验(图47)。

[0424] 接下来,测试了不同的Cpf1蛋白对RNP编辑的影响。简而言之,在补充有SCF、TPO和

FLT3的X-Vivo 10培养基中进行48小时预刺激后，将包括与不同Cpf1蛋白复合的gRNA(包含SEQ ID NO:1051(+25DNA 5'gRNA修饰和1xPS-OMe3'gRNA修饰，表19))的RNP58、RNP26、RNP27和RNP28(表21)通过MaxCyte电穿孔以8μM递送至1x 10^6个CD34+细胞。结果证明，所有RNP都支持中靶编辑(图48)。

[0425] 接下来，测试了含有具有不同5'gRNA修饰和相同3'修饰的gRNA的不同RNP的效果。简而言之，在补充有SCF、TPO和FLT3的X-Vivo 10培养基中进行48小时预刺激后，将RNP(RNP58(+25DNA 5'gRNA修饰和1xPS-OMe 3'gRNA修饰)、RNP29(+25DNA+2xPS 5'gRNA修饰和1xPS-OMe 3'gRNA修饰)、RNP30(聚A RNA+2xPS 5'gRNA修饰和1xPS-OMe 3'gRNA修饰)和RNP31(聚U RNA+2xPS 5'gRNA修饰和1xPS-OMe 3'gRNA修饰)(表21)通过MaxCyte电穿孔以1μM、2μM或4μM递送至1x 10^6个CD34+细胞(由于细胞的可用性，未在1μM下测试RNP30)。结果证明，所有RNP都支持中靶编辑(图49)。

[0426] 接下来，测试了不同的Cpf1蛋白对RNP编辑的影响。简而言之，在补充有SCF、TPO和FLT3的X-Vivo 10培养基中进行48小时预刺激后，将包括与不同Cpf1蛋白复合的gRNA(包含SEQ ID NO:1051(+25DNA 5'gRNA修饰和1xPS-OMe3'gRNA修饰，表19))的RNP58、RNP27和RNP26(表21)通过MaxCyte电穿孔以2μM或4μM递送至1x 10^6个CD34+细胞。为了研究混合CD34、表型祖细胞和表型造血干细胞(HSC)的编辑，对HSPC亚群进行了表征。因此，在电穿孔后，将细胞放回培养另外48小时，然后分类为祖细胞和HSC级分以进行indel定量。在电穿孔后48小时，通过对PCR扩增的靶位点的Illumina测序(NGS)分析靶位点处的插入/缺失水平。结果证明，所有RNP都支持混合CD34、表型祖细胞和表型造血干细胞中的中靶编辑(图50)。

[0427] 接下来，测试了RNP与Sp182 RNP(包含与化脓链球菌Cas9(SEQ ID NO:1033)复合的SEQ ID NO:1027(表14)的失活gRNA)或ssODN OLI16431(SEQ ID NO:1040，表11)的共递送对含有不同Cpf1蛋白的RNP的效果。简而言之，在补充有SCF、TPO和FLT3的X-Vivo 10培养基中进行预刺激48小时后，通过MaxCyte电穿孔将RNP61、RNP62，RNP34(表21)以8μM与Sp182 RNP(8μM)或ssODN OLI16431(8μM)共递送至25x 10^6个mPB CD34+细胞。为了研究混合CD34、表型祖细胞和表型造血干细胞(HSC)的编辑，对HSPC亚群进行了表征。因此，在电穿孔后，将细胞放回培养另外48小时，然后分类为祖细胞和HSC级分以进行indel定量。在电穿孔后48小时，通过对PCR扩增的靶位点的Illumina测序(NGS)分析靶位点处的插入/缺失水平。结果证明，与“加强元件”Sp182 RNP或ssODN OLI16431共递送的RNP支持中靶编辑(图51)。在电穿孔后24小时还将这些细胞的一部分冷冻保存，以在体内植入模型中进一步表征(参见实例23)。

[0428] 接下来，测试了RNP58和RNP32编辑的效果。简而言之，在补充有SCF、TPO和FLT3的X-Vivo 10培养基中进行预刺激48小时后，通过MaxCyte电穿孔将RNP58和RNP32(表21)以2μM递送至6x 10^6个mPB CD34+细胞。为了研究混合CD34、表型祖细胞和表型HSC的编辑，对HSPC亚群进行了表征。因此，在电穿孔后，将细胞放回培养另外48小时，然后分类为祖细胞和HSC级分以进行indel定量。在电穿孔后48小时，通过对PCR扩增的靶位点的Illumina测序(NGS)分析靶位点处的插入/缺失水平。结果证明RNP58和RNP32支持中靶编辑(图52)。

[0429] 在补充有SCF、TPO和FLT3的X-Vivo 10培养基中进行预刺激48小时后，通过MaxCyte电穿孔将RNP58和RNP1以2μM至8μM的浓度递送至25x 10^6个mPB CD34+细胞(表21)。为了研究混合CD34、表型祖细胞和表型造血干细胞(HSC)的编辑，对HSPC亚群进行了表征。

因此,在电穿孔后,将细胞放回培养另外48小时,然后分类为祖细胞和HSC级分以进行indel定量。在电穿孔后48小时,通过对PCR扩增的靶位点的Illumina测序(NGS)分析靶位点处的插入/缺失水平。结果证明测试的RNP支持中靶编辑(图53)。在电穿孔后24小时还将这些细胞的一部分冷冻保存,以在体内植入模型中进一步表征(参见实例23)。

实例23:将经编辑的mPB CD34+细胞输注到NOD,B6.SCID Il2rγ-/- Kit(W41/W41)小鼠中导致长期植入和HbF诱导。

[0430] 为了确定与ssODN OLI16431(SEQ ID NO:1040,表11)共递送的RNP34和RNP33(表21)的递送是否在长期繁殖的造血干细胞中实现编辑,将来自动员外周血(mPB)的成人CD34+细胞输注入未辐照的NOD,B6.SCID Il2rγ-/-Kit(W41/W41)(杰克森(Jackson)实验室储备名称:NOD.Cg-Kit<W-41J>Tyr<+>Prkdc<scid>Il2rg<tm1Wjl>/ThomJ)("NBSGW")小鼠。简而言之,在补充有SCF、TPO和FLT3的X-Vivo 10培养基中进行48小时预刺激后,将mPB CD34+细胞用8μM剂量的RNP和6μM ssODN OLI16431通过MaxCyte电穿孔以62.5x 10^6/mL进行电穿孔。24小时后,将mCD34+细胞冷冻保存。五天后,将mPB CD34+细胞解冻,并通过静脉内尾静脉注射以每只小鼠100万个细胞注入NBSGW小鼠。八周后,对小鼠实施安乐死,并从股骨、胫骨和骨盆骨中收集骨髓(BM)。通过FACS分离分别鉴定为CD15+、CD19+、糖蛋白A(GlyA,CD235a+)和谱系阴性CD34+的细胞的骨髓亚群,并提取DNA以确定每个这些级分的编辑水平。图54描绘了由下一代测序确定的未分级BM或按流分选的各单独BM亚群的indel的频率。

[0431] 接下来,为了确定与Sp182 RNP(包含与化脓链球菌Cas9(SEQ ID NO:1033)复合的SEQ ID NO:1027(表14)的失活gRNA)共递送的RNP34或RNP33(表21)的递送是否在长期繁殖的造血干细胞中实现编辑,将来自动员的外周血(mPB)的成人CD34+细胞输注入未辐照的NBSGW小鼠中。简而言之,在补充有SCF、TPO和FLT3的X-Vivo 10培养基中进行48小时预刺激后,将mPB CD34+细胞用不同剂量的RNP和不同剂量的Sp182 RNP通过MaxCyte电穿孔以62.5x 10^6/mL进行电穿孔。24小时后,将mCD34+细胞冷冻保存。五天后,将mPB CD34+细胞解冻,并通过静脉内尾静脉注射以每只小鼠100万个细胞注入NBSGW小鼠。八周后,对小鼠实施安乐死,并从股骨、胫骨和骨盆骨中收集骨髓(BM)。通过FACS分离分别鉴定为CD15+、CD19+、糖蛋白A(GlyA,CD235a+)和谱系阴性CD34+的细胞的骨髓亚群,并提取DNA以确定每个这些级分的编辑水平。图55A描绘了由下一代测序确定的未分级BM或按流分选的各单独BM亚群的indel的频率。

[0432] 最后,分析了衍生自BM的CD34+细胞的长期HbF诱导。在红系分化条件下将BM细胞的等分试样培养18天(Giarratana 2011),并通过UPLC评估HbF表达。简而言之,将在输注后8周从小鼠中提取的未分级的BM细胞置于红系培养条件下18天。在第7、10和14天进行细胞计数和补料,在第18天进行红系收集。然后对这些骨髓衍生的红系细胞进行计数,并在HPLC级水中溶解,然后过滤除去细胞碎片。然后,通过UPLC测量每个样品的γ-珠蛋白链的相对表达水平(相比于总β样珠蛋白链),其中通过逐渐增加含0.1%三氟乙酸的乙腈与含0.1%三氟乙酸的水的比例来实现蛋白分离(图54B)。这些数据证明,通过与Sp182 RNP共递送的RNP34和RNP33编辑人CD34+细胞实现了稳健的长期HbF诱导。

[0433] 接下来,为了确定与ssODN OLI16431(SEQ ID NO:1040,表11)共递送的RNP61或RNP62(表21)的递送是否在长期繁殖的造血干细胞中实现编辑,将来自mPB的成人CD34+细胞输注入未辐照的NBSGW小鼠中。简而言之,在补充有SCF、TPO和FLT3的X-Vivo 10培养基中

进行48小时预刺激后,将mPB CD34+细胞用8μM RNP和8μM ssODN OLI16431通过MaxCyte电穿孔以62.5x 10^6/mL进行电穿孔。24小时后,将mCD34+细胞冷冻保存。四天后,将mPB CD34+细胞解冻,并通过静脉内尾静脉注射以每只小鼠100万个细胞注入NBSGW小鼠。八周后,对小鼠实施安乐死,并从股骨、胫骨和骨盆骨中收集骨髓(BM)。通过FACS分离分别鉴定为CD15+、CD19+、糖蛋白A(GlyA,CD235a+)和谱系阴性CD34+的细胞的骨髓亚群,并提取DNA以确定每个这些级分的编辑水平。图56A描绘了由下一代测序确定的未分级BM或按流分选的各单独BM亚群的indel的频率。

[0434] 最后,分析了衍生自BM的CD235a+(GlyA+)红系细胞的长期HbF诱导。在红系分化条件下将BM细胞的等分试样培养18天,并通过UPLC评估HbF表达。简而言之,将在输注后8周从小鼠中提取的未分级的BM细胞置于红系培养条件下18天。图56B描绘了红系细胞的HbF表达,计算为γ/β样链(%)。这些数据证明,通过与ssODN OLI16431共递送的RNP61和RNP62编辑人CD34+细胞实现了稳健的长期HbF诱导。

[0435] 接下来,为了确定RNP1或RNP58(表21)的递送是否在长期繁殖的造血干细胞中实现编辑,将来自mPB的成人CD34+细胞输注入未辐照的NBSGW小鼠中。简而言之,在补充有SCF、TPO和FLT3的X-Vivo 10培养基中进行48小时预刺激后,将mPB CD34+细胞用4μM或8μM RNP1或2μM、4μM或8μM RNP58通过MaxCyte电穿孔以62.5x 10^6/mL进行电穿孔。24小时后,将mCD34+细胞冷冻保存。四天后,将mPB CD34+细胞解冻,并通过静脉内尾静脉注射以每只小鼠100万个细胞注入NBSGW小鼠。八周后,对小鼠实施安乐死,并从股骨、胫骨和骨盆骨中收集骨髓(BM)。通过流式细胞术确定了BM中的人嵌合性和谱系重构(CD45+、CD14+、CD19+、糖蛋白A(GlyA、CD235a+)、谱系和CD34+以及小鼠CD45+标记表达)并进行了分析(图57)。GlyA+细胞的频率计算为BM中GlyA+细胞/总细胞。人嵌合性被定义为人CD45/(人CD45+mCD45)。

[0436] 与模拟转染的mPB CD34+细胞相比,RNP转染的mPB CD34+细胞在移植后8周获得了相似的嵌合性和谱系分布,这证明编辑与保留造血干细胞的植入潜力是相当的。图58描绘了在输注后8周,由下一代测序确定的未分级BM的indel。图59显示了在具有2μM、4μM或8μM RNP58的BM、CD15+、CD19+、糖蛋白A(GlyA、CD235a+)和lin-CD34+细胞中indel率的频率。

[0437] 还分析了来自经编辑的CD34+细胞的CD235a+(GlyA+)红系细胞的长期HbF诱导。通过荧光激活细胞分选术(FACS)分离从骨髓获得的GlyA+细胞,收集并在HPLC级水中裂解。然后通过UPLC评估裂解物的HbF表达。图60描绘了GlyA+细胞的HbF表达,计算为γ/β样链(%)。这些数据证明,在各种RNP58浓度下,对人CD34+细胞进行RNP58编辑实现稳健的长期HbF诱导。

[0438] 重要的是,用不同浓度的RNP1或RNP58电穿孔的$CD34^+$细胞维持其离体造血活性(即,与未处理的供体匹配的$CD34^+$细胞阴性对照相比,红系和骨髓集落的量或多样性没有差异),如在造血集落形成细胞(CFC)测定中确定的(图61)。

实例24:使用经编辑的造血干细胞治疗β-血红蛋白病。

[0439] 本文公开的方法和基因组编辑系统可用于在有需要的患者中治疗β-血红蛋白病,例如镰状细胞病或β-地中海贫血。例如,可以以自体程序对衍生自患者的细胞进行基因组编辑。对患者细胞进行离体校正和将细胞重新引入患者体内可能会导致HbF表达增加并治疗β-血红蛋白病。

[0440] 例如,HSC可使用本领域技术人员众所周知的技术从患有β-血红蛋白病的患者的

骨髓中提取。HSC可以使用本文公开的用于基因组编辑的方法进行修饰。例如，由靶向HBG基因中的一个或多个区域的与RNA指导的核酸酶复合的指导RNA(gRNA)构成的RNP可用于编辑HSC。在某些实施例中，RNA指导的核酸酶可以是Cpf1蛋白。在某些实施例中，Cpf1蛋白可以是经修饰的Cpf1蛋白。在某些实施例中，经修饰的Cpf1蛋白可以由SEQ ID NO:1000、1001、1008-1018、1032、1035-39、1094-1097、1107-09(Cpf1多肽序列)或SEQ ID NO:1019-1021、1110-17(Cpf1多核苷酸序列)所示的序列编码。例如，经修饰的Cpf1蛋白可以由SEQ ID NO:1097所示的序列编码。在某些实施例中，gRNA可以是经修饰的或未经修饰的gRNA。在某些实施例中，gRNA可包含表13、表18或表19所示的序列。例如，在某些实施例中，gRNA可包含SEQ ID NO:1051所示的序列。在某些实施例中，RNP复合物可包含表21所示的RNP复合物。例如，RNP复合物可以包括包含SEQ ID NO:1051所示的序列的gRNA和由SEQ ID NO:1097所示的序列编码的经修饰的Cpf1蛋白(RNP32，表21)。在某些实施例中，相对于未经修饰的HSC，经修饰的HSC在人HBG1基因、HBG2基因、或两者中的indel的频率或水平增加。在某些实施例中，经修饰的HSC可以分化为表达增加水平的HbF的红系细胞。可以选择经修饰的HSC的群体，以通过输注或本领域技术人员已知的其他方法重新引入患者体内。可以基于例如经修饰的HSC的红系后代的HbF表达增加或经修饰的HSC的indel频率增加来选择用于重新引入的经修饰的HSC的群体。在一些实施例中，在细胞重新引入之前的任何形式的消融可用于增强经修饰的HSC的植入。在其他实施例中，可以使用技术人员熟知的技术(例如，单采血液分离术(apheresis)或白细胞单采术(leukapheresis))从患有β-血红蛋白病的患者中提取外周血干细胞(PBSC)，并且可以从PBSC中去除干细胞。如上所述的基因组编辑方法可以在干细胞上进行，并且经修饰的干细胞可以如上所述被重新引入患者体内。

表20:Cpf1蛋白变体

Cpf1 变体 ID	Cpf1 变体氨基酸 SEQ ID NO	Cpf1 变体核苷酸 SEQ ID NO
His-AsCpf1-sNLS-sNLS H800A	SEQ ID NO: 1032	SEQ ID NO: 1110
Cpf1-1	SEQ ID NO: 1094	SEQ ID NO: 1111
Cpf1-2	SEQ ID NO: 1095	SEQ ID NO: 1112
Cpf1-3	SEQ ID NO: 1096	SEQ ID NO: 1113
Cpf1-4	SEQ ID NO: 1097	SEQ ID NO: 1114
Cpf1-5	SEQ ID NO: 1107	SEQ ID NO: 1115
Cpf1-6	SEQ ID NO: 1108	SEQ ID NO: 1116
Cpf1-7	SEQ ID NO: 1109	SEQ ID NO: 1117

表21:RNP复合物

RNP 名称	gRNA SEQ ID NO*	Cpf1 氨基酸 SEQ ID NO**
RNP64	SEQ ID NO: 1022	SEQ ID NO: 1109
RNP1	SEQ ID NO: 1051	SEQ ID NO: 1095
RNP2	SEQ ID NO: 1098	SEQ ID NO: 1094
RNP3	SEQ ID NO: 1099	SEQ ID NO: 1094
RNP4	SEQ ID NO: 1100	SEQ ID NO: 1094
RNP5	SEQ ID NO: 1101	SEQ ID NO: 1094
RNP6	SEQ ID NO: 1068	SEQ ID NO: 1094
RNP7	SEQ ID NO: 1102	SEQ ID NO: 1094

RNP8	SEQ ID NO: 1081	SEQ ID NO: 1094
RNP9	SEQ ID NO: 1082	SEQ ID NO: 1094
RNP10	SEQ ID NO: 1083	SEQ ID NO: 1094
RNP11	SEQ ID NO: 1069	SEQ ID NO: 1094
RNP12	SEQ ID NO: 1084	SEQ ID NO: 1094
RNP16	SEQ ID NO: 1085	SEQ ID NO: 1094
RNP19	SEQ ID NO: 1088	SEQ ID NO: 1094
RNP20	SEQ ID NO: 1089	SEQ ID NO: 1094
RNP21	SEQ ID NO: 1090	SEQ ID NO: 1094
RNP22	SEQ ID NO: 1091	SEQ ID NO: 1094
RNP23	SEQ ID NO: 1048	SEQ ID NO: 1097
RNP24	SEQ ID NO: 1093	SEQ ID NO: 1094
RNP26	SEQ ID NO: 1051	SEQ ID NO: 1096
RNP27	SEQ ID NO: 1051	SEQ ID NO: 1107
RNP28	SEQ ID NO: 1051	SEQ ID NO: 1108
RNP29	SEQ ID NO: 1103	SEQ ID NO: 1094
RNP30	SEQ ID NO: 1104	SEQ ID NO: 1094
RNP31	SEQ ID NO: 1105	SEQ ID NO: 1094
RNP32	SEQ ID NO: 1051	SEQ ID NO: 1097
RNP33	SEQ ID NO: 1022	SEQ ID NO: 1032
RNP34	SEQ ID NO: 1041	SEQ ID NO: 1032
RNP37	SEQ ID NO: 1042	SEQ ID NO: 1032
RNP38	SEQ ID NO: 1043	SEQ ID NO: 1032
RNP39	SEQ ID NO: 1044	SEQ ID NO: 1032
RNP40	SEQ ID NO: 1045	SEQ ID NO: 1032
RNP41	SEQ ID NO: 1046	SEQ ID NO: 1032
RNP42	SEQ ID NO: 1047	SEQ ID NO: 1032
RNP43	SEQ ID NO: 1048	SEQ ID NO: 1032
RNP44	SEQ ID NO: 1049	SEQ ID NO: 1032
RNP45	SEQ ID NO: 1022	SEQ ID NO: 1094
RNP46	SEQ ID NO: 1046	SEQ ID NO: 1094
RNP47	SEQ ID NO: 1047	SEQ ID NO: 1094
RNP48	SEQ ID NO: 1052	SEQ ID NO: 1094
RNP49	SEQ ID NO: 1048	SEQ ID NO: 1094
RNP50	SEQ ID NO: 1053	SEQ ID NO: 1094
RNP51	SEQ ID NO: 1054	SEQ ID NO: 1094
RNP52	SEQ ID NO: 1055	SEQ ID NO: 1094
RNP53	SEQ ID NO: 1056	SEQ ID NO: 1094
RNP54	SEQ ID NO: 1057	SEQ ID NO: 1094
RNP55	SEQ ID NO: 1058	SEQ ID NO: 1094
RNP56	SEQ ID NO: 1059	SEQ ID NO: 1094
RNP57	SEQ ID NO: 1060	SEQ ID NO: 1094
RNP58	SEQ ID NO: 1051	SEQ ID NO: 1094
RNP59	SEQ ID NO: 1067	SEQ ID NO: 1094
RNP60	SEQ ID NO: 1068	SEQ ID NO: 1094
RNP61	SEQ ID NO: 1041	SEQ ID NO: 1095

RNP62	SEQ ID NO: 1041	SEQ ID NO: 1094
RNP63	SEQ ID NO: 1022	SEQ ID NO: 1095
*参见表 19 **参见表 20		

表22:Cpf1 HBG1靶向结构域和预期的切割位点

gRNA ID	gRNA 靶向结构域 (RNA)	gRNA 靶向结构域 (DNA)	HBG1 处的靶向结构域坐标*	HBG1 处的预期的切割位点坐标#	链	靶向结构域长度
HBG1-1	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO: 1002)	CCTTGTCAAGGCTATTGGTC (SEQ ID NO: 1003)	Chr11:5249954-5249974	Chr11:5249973, Chr11:5249977	+	20
HBG1-1 (21mer)	CCUUGUCAAGGCUAUUGGUCA (SEQ ID NO: 1254)	CCTTGTCAAGGCTATTGGTCA (SEQ ID NO: 1255)	Chr11:5249954-Chr11:5249975	Chr11:5249973, Chr11:5249977	+	21
HBG1-1 (22mer)	CCUUGUCAAGGCUAUUGGUCAA(SEQ ID NO: 1256)	CCTTGTCAAGGCTATTGGTCAA(SEQ ID NO: 1257)	Chr11:5249954-Chr11:5249976	Chr11:5249973, Chr11:5249977	+	22
HBG1-1 (23mer)	CCUUGUCAAGGCUAUUGGUCAAG(SEQ ID NO: 1258)	CCTTGTCAAGGCTATTGGTCAAG(SEQ ID NO: 1259)	Chr11:5249954-Chr11:5249976	Chr11:5249973, Chr11:5249977	+	23
AsCpf1 HBG1 启动子-1	AGACAGAUAUUUGCAUUGAG (SEQ ID NO: 1139)	AGACAGATATTTGCATTGAG (SEQ ID NO: 1140)	Chr11:5250023-Chr11:5250043	Chr11:5250042, Chr11:5250046	+	20
AsCpf1 HBG1 启动子-1 (21mer)	AGACAGAUAUUUGCAUUGAGA (SEQ ID NO: 1260)	AGACAGATATTTGCATTGAGA(SEQ ID NO: 1261)	Chr11:5250023-Chr11:5250044	Chr11:5250042, Chr11:5250046	+	21
AsCpf1 HBG1 启动子-2	CAUUGAGAUAGUGUGGGGAA (SEQ ID NO: 1141)	CATTGAGATAGTGTGGGGAA (SEQ ID NO: 1142)	Chr11:5250036-Chr11:5250056	Chr11:5250055, Chr11:5250059	+	20
AsCpf1 HBG1 启动子-2 (21mer)	CAUUGAGAUAGUGUGGGGAAG (SEQ ID NO: 1262)	CATTGAGATAGTGTGGGGAAG (SEQ ID NO: 1263)	Chr11:5250036-Chr11:5250057	Chr11:5250055, Chr11:5250059	+	21
AsCpf1 HBG1 启动子-6	CUUCUCCCAUCAUAGAGGAU (SEQ ID NO: 1149)	CTTCTCCCATCATAGAGGAT (SEQ ID NO: 1150)	Chr11:5250160-Chr11:5250180	Chr11:5250179, Chr11:5250183	+	20
AsCpf1 HBG1 启动子-6 (21mer)	CUUCUCCCAUCAUAGAGGAUA (SEQ ID NO: 1264)	CTTCTCCCATCATAGAGGATA (SEQ ID NO: 1265)	Chr11:5250160-Chr11:5250181	Chr11:5250179, Chr11:5250183	+	21

*NCBI 参考序列 NC_000011，使用基于一个的坐标系统（One-based coordinate system）报告坐标，“智人染色体 11，GRCh38.p12 初级组装”（版本 NC_000011.10）。
#基于 Zetsche 等人 2015 的预期切割位点，使用基于零的坐标来报告坐标。

表23:Cpf1 HBG2靶向结构域和预期的切割位点

gRNA ID	gRNA 靶向结构域 (RNA)	gRNA 靶向结构域 (DNA)	HBG2 处的靶向结构域坐标*	HBG2 处的预期的切割位点坐标#	链	靶向结构域长度
HBG1-1	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO: 1002)	CCTTGTCAAGGCTATTGGTC (SEQ ID NO: 1003)	Chr11:5254878 Chr11:5254898	Chr11:5254897, Chr11:5254901	+	20
HBG1-1 21mer	CCUUGUCAAGGCUAUUGGUCA (SEQ ID NO: 1254)	CCTTGTCAAGGCTATTGGTCA (SEQ ID NO: 1255)	Chr11:5254878 Chr11:5254899	Chr11:5254897, Chr11:5254901	+	21
HBG1-1 22mer	CCUUGUCAAGGCUAUUGGUCAA (SEQ ID NO: 1256)	CCTTGTCAAGGCTATTGGTCAA (SEQ ID NO: 1257)	Chr11:5254878 Chr11:5254900	Chr11:5254897, Chr11:5254901	+	22
HBG1-1 23mer	CCUUGUCAAGGCUAUUGGUCAAG (SEQ ID NO: 1258)	CCTTGTCAAGGCTATTGGTCAAG (SEQ ID NO: 1259)	Chr11:5254878 Chr11:5254901	Chr11:5254897, Chr11:5254901	+	23
AsCpf1 HBG1 启动子-1	AGACAGAUAUUUGCAUUGAG (SEQ ID NO: 1139)	AGACAGATATTTGCATTGAG (SEQ ID NO: 1140)	Chr11:5254947 Chr11:5254967	Chr11:5254966 Chr11:5254970	+	20
AsCpf1 HBG1 启动子-1 (21mer)	AGACAGAUAUUUGCAUUGAGA (SEQ ID NO: 1260)	AGACAGATATTTGCATTGAGA (SEQ ID NO: 1261)	Chr11:5254947 Chr11:5254968	Chr11:5254966 Chr11:5254970	+	21
AsCpf1 HBG1 启动子-2	CAUUGAGAUAGUGUGGGGAA (SEQ ID NO: 1141)	CATTGAGATAGTGTGGGGAA (SEQ ID NO: 1142)	Chr11:5254960 Chr11:5254980	Chr11:5254979 Chr11:5254983	+	20
AsCpf1 HBG1 启动子-2 (21mer)	CAUUGAGAUAGUGUGGGGAAG (SEQ ID NO: 1262)	CATTGAGATAGTGTGGGGAAG (SEQ ID NO: 1263)	Chr11:5254960 Chr11:5254981	Chr11:5254979 Chr11:5254983	+	21
*NCBI 参考序列 NC_000011，使用基于一个的坐标系统（One-based coordinate system）报告坐标，“智人染色体 11，GRCh38.p12 初级组装”（版本 NC_000011.10）。 #基于（Zetsche 等人 2015）的预期切割位点，使用基于零的坐标来报告坐标。						

表24:gRNA 5’延伸

5'延伸序列 ID No:	5'延伸序列	5'修饰
1231	rCrUrUrUrU	+5 RNA
1232	rArArGrArCrCrUrUrUrU	+10 RNA
1233	rArUrGrUrGrUrUrUrUrUrGrUrCrArArArArArGrArCrCrUrUrUrU	+25 RNA
1234	rArGrGrCrCrArGrCrUrUrGrCrCrGrGrUrUrUrUrUrUrArGrUrCrGrUrGrCrUrGrCrUrUrCrArUrGrUrGrUrUrUrUrUrGrUrCrArArArArArGrArCrCrUrUrUrU	+60 RNA
1235	CTTTT	+5 DNA
1236	AAGACCTTTT	+10 DNA
1237	ATGTGTTTTTGTCAAAAGACCTTTT	+25 DNA
1238	AGGCCAGCTTGCCGGTTTTTTAGTCGTGCTGCTTCATGTGTTTTTGTCAAAAGACCTTTT	+60 DNA
1239	TTTTTGTCAAAAGACCTTTT	+20 DNA
1240	GCTTCATGTGTTTTTGTCAAAAGACCTTTT	+30 DNA
1241	GCCGGTTTTTTAGTCGTGCTGCTTCATGTGTTTTTGTCAAAAGACCTTTT	+50 DNA
1242	TAGTCGTGCTGCTTCATGTGTTTTTGTCAAAAGACCTTTT	+40 DNA
1243	C*C*GAAGTTTTCTTCGGTTTT	+20 DNA + 2xPS
1244	T*T*TTTCCGAAGTTTTCTTCGGTTTT	+25 DNA + 2xPS
1245	A*A*CGCTTTTTCCGAAGTTTTCTTCGGTTTT	+30 DNA + 2xPS
1246	G*C*GTTGTTTTCAACGCTTTTTCCGAAGTTTTCTTCGGTTTT	+41 DNA + 2xPS
1247	G*G*CTTCTTTTGAAGCCTTTTTGCGTTGTTTTCAACGCTTTTTCCGAAGTTTTCTTCGGTTTT	+62 DNA + 2xPS
1248	A*T*GTGTTTTTGTCAAAAGACCTTTT	+25 DNA + 2xPS
1249	AAAAAAAAAAAAAAAAAAAAAAAAA	+25 A
1250	TTTTTTTTTTTTTTTTTTTTTTTTT	+25 T
1251	mA*mU*rGrUrGrUrUrUrUrGrUrCrArArArArGrArCrCrUrUrUrU	+25 RNA + 2xPS
1252	mA*mA*rArArArArArArArArArArArArArArArArArArArArArArArA	PolyA RNA + 2xPS
1253	mU*mU*rUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrU	聚 U RNA + 2xPS
所有碱基均为大写 小写字母"r"代表 RNA，2'-羟基；没有被"r"修饰的碱基是 DNA 所有碱基均通过标准磷酸二酯键连接，除另有说明外： "*"代表硫代磷酸酯修饰 "PS"代表硫代磷酸酯修饰		

表19:Cpf1指导RNA

gRNA 序列 SEQ ID NO.	gRNA 序列	5'修饰**	3'修饰	crRNA + gRNA 靶向结构域的长度	gRNA 靶向结构域的长度	gRNA 靶向结构域 (RNA)
1022	rUrArArUrUrUrCrUrArCrUrCrUrUrGrUrA	-	-	40	20	CCUUGUCAAGGCUAUUGGUC

	rGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC					(SEQ ID NO: 1002)
1023	rUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUmC/PS2/mA	-	1xPS2-OMe + 1xOMe	41	21	CCUUGUCAAGGCUAUUGGUCA (SEQ ID NO: 1254)
1041	rUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGmU/PS2/mC	-	1xPS2-OMe + 1xOMe	40	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO: 1002)
1042	rCrUrUrUrUrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC	+5 RNA	-	45	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO: 1002)
1043	rArArGrArCrCrUrUrUrUrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC	+10 RNA	-	50	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO: 1002)
1044	rArUrGrUrGrUrUrUrUrUrGrUrCrArArArArGrArCrCrUrUrUrUrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC	+25 RNA	-	65	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO: 1002)
1045	rArGrGrCrCrArGrCrUrUrGrCrCrGrGrUrUrUrUrUrArGrUrCrGrUrGrCrUrGrCrCrUrUrCrArUrGrUrGrUrUrUrUrUrGrUrCrArArArArGrArCrCrUrUrUrUrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC	+60 RNA	-	100	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO: 1002)
1046	CTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC	+5 DNA	-	45	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO: 1002)
1047	AAGACCTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUC	+10 DNA	-	50	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO:

	UrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC					1002)
1048	ATGTGTTTTTGTCAAAAGACCTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC	+25 DNA	-	65	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO: 1002)
1049	AGGCCAGCTTGCCGGTTTTTTAGTCGTGCTGCTTCATGTGTTTTGTCAAAAGACCTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrGrCrUrArUrUrGrGrUrC	+60 DNA	-	100	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO: 1002)
1050	ATGTGTTTTTGTCAAAAGACCTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUmC/PS2/mA	+25 DNA	1xPS2-OMe + 1xOMe	66	21	CCUUGUCAAGGCUAUUGGUCA (SEQ ID NO: 1254)
1051	ATGTGTTTTTGTCAAAAGACCTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC*mA	+25 DNA	1xPS-OMe	66	21	CCUUGUCAAGGCUAUUGGUCA (SEQ ID NO: 1254)
1052	TTTTTGTCAAAAGACCTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC	+20 DNA	-	60	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO: 1002)
1053	GCTTCATGTGTTTTTGTCAAAAGACCTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC	+30 DNA	-	70	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO: 1002)
1054	GCCGGTTTTTTAGTCGTGCTGCTTCATGTGTTTTTGTCAAAAGACCTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrG	+50 DNA	-	90	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO: 1002)

	rGrUrC					
1055	TAGTCGTGCTGCTTCATGTGTTTTTGTCAAAAGACCTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC	+40 DNA	-	80	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO: 1002)
1056	C*C*GAAGTTTTCTTCGGTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC	+20 DNA + 2xPS	-	60	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO: 1002)
1057	T*T*TTTCCGAAGTTTTCTTCGGTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC	+25 DNA + 2xPS	-	65	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO: 1002)
1058	A*A*CGCTTTTTCCGAAGTTTTCTTCGGTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC	+30 DNA + 2xPS	-	70	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO: 1002)
1059	G*C*GTTGTTTTCAACGCTTTTTCCGAAGTTTTCTTCGGTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC	+41 DNA + 2xPS	-	81	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO: 1002)
1060	G*G*CTTCTTTTGAAGCCTTTTTGCGTTGTTTTCAACGCTTTTTCCGAAGTTTTCTTCGGTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC	+62 DNA + 2xPS	-	102	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO: 1002)
1061	mUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC	1xOMe	-	40	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO: 1002)
1062	mU*rArArUrUrUrCrUrArCrUrCrUrUrGrU	1xPS-OMe	-	40	20	CCUUGUCAAGGCUAUUGGUC

	rArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC					(SEQ ID NO: 1002)
1063	mUmArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC	2xOMe	-	40	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO: 1002)
1064	mU*mA*rArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC	2xPS-OMe	-	40	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO: 1002)
1065	mUmAmArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC	3xOMe	-	40	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO: 1002)
1066	mU*mA*mA*rUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC	3xPS-OMe	-	40	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO: 1002)
1067	A*T*GTGTTTTTGTCAAAAGACCTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC	+25 DNA + 2xPS	-	65	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO: 1002)
1068	A*T*GTGTTTTTGTCAAAAGACCTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC*mA	+25 DNA + 2xPS	1xPS-OMe	66	21	CCUUGUCAAGGCUAUUGGUCA (SEQ ID NO: 1254)
1069	rUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC*mA	-	1xPS-OMe	41	21	CCUUGUCAAGGCUAUUGGUCA (SEQ ID NO: 1254)
1070	rUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrCmA*rU	-	1xPS-OMe + rU	42	21	CCUUGUCAAGGCUAUUGGUCA (SEQ ID NO: 1254)
1071	rUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrCrU	-	rU	41	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO: 1002)
1072	rUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGr	-	rU	42	21	CCUUGUCAAGGCUAUUGGUCA (SEQ ID NO:

	UrCrArArGrGrCrUrA rUrUrGrGrUrCrArU					1254)
1073	rUrArArUrUrUrCrUr ArCrUrCrUrUrGrUrA rGrArUrCrCrUrUrGr UrCrArArGrGrCrUrA rUrUrGrGrUrCrArAr GrU	-	rU	44	23	CCUUGUCAAGG CUAUUGGUCAA G (SEQ ID NO: 1258)
1074	A*T*GTGTTTTTGT CAAAAGACCTTTTr UrArArUrUrUrCrUrA rCrUrCrUrUrGrUrAr GrArUrCrCrUrUrGrU rCrArArGrGrCrUrAr UrUrGrGrUrCrU	+25 DNA + 2xPS	rU	66	20	CCUUGUCAAGG CUAUUGGUC (SEQ ID NO: 1002)
1075	A*T*GTGTTTTTGT CAAAAGACCTTTTr UrArArUrUrUrCrUrA rCrUrCrUrUrGrUrAr GrArUrCrCrUrUrGrU rCrArArGrGrCrUrAr UrUrGrGrUrCrArArG rU	+25 DNA + 2xPS	rU	69	23	CCUUGUCAAGG CUAUUGGUCAA G (SEQ ID NO: 1258)
1076	rUrArArUrUrUrCrUr ArCrUrCrUrUrGrUrA rGrArUrCrCrUrUrGr UrCrArArGrGrCrUrA rUrUrGrGrUrCrA	-	-	41	21	CCUUGUCAAGG CUAUUGGUCA (SEQ ID NO: 1254)
1077	rUrArArUrUrUrCrUr ArCrUrCrUrUrGrUrA rGrArUrCrCrUrUrGr UrCrArArGrGrCrUrA rUrUrGrGrUrCrArAr G	-	-	43	23	CCUUGUCAAGG CUAUUGGUCAA G (SEQ ID NO: 1258)
1078	A*T*GTGTTTTTGT CAAAAGACCTTTTr UrArArUrUrUrCrUrA rCrUrCrUrUrGrUrAr GrArUrCrCrUrUrGrU rCrArArGrGrCrUrAr UrUrGrGrUrCrArArG	+25 DNA + 2xPS	-	68	23	CCUUGUCAAGG CUAUUGGUCAA G (SEQ ID NO: 1258)
1079	rUrArArUrUrUrCrUr ArCrUrCrUrUrGrUrA rGrArUrCrCrUrUrGr UrCrArArGrGrCrUrA rUrUrGrGrUrCrUrUr UrU	-	4xrU	44	20	CCUUGUCAAGG CUAUUGGUC (SEQ ID NO: 1002)
1080	rUrArArUrUrUrCrUr ArCrUrCrUrUrGrUrA rGrArUrCrCrUrUrGr UrCrArArGrGrCrUrA rUrUrGrGrUrCmU*m U*mU*rU	-	3xPS-O Me-U + rU	44	20	CCUUGUCAAGG CUAUUGGUC (SEQ ID NO: 1002)
1081	A*T*GTGTTTTTGT CAAAAGACCTTTTr UrArArUrUrUrCrUrA	+25 DNA + 2xPS	4xrU	69	20	CCUUGUCAAGG CUAUUGGUC (SEQ ID NO:

	rCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrCrUrUrUrU					1002)
1082	AAAAAAAAAAAAAAAAAAAAArUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrUrCrArArGrGrCrUrArArUrUrGrGrUrC*mA	+25 A	1xPS-OMe	66	21	CCUUGUCAAGGCUAUUGGUCA (SEQ ID NO: 1254)
1083	TTTTTTTTTTTTTTTTTTTTTTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC*mA	+25 T	1xPS-OMe	66	21	CCUUGUCAAGGCUAUUGGUCA (SEQ ID NO: 1254)
1084	C*C*GAAGTTTTCTTCGGTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrArGrGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC*mA	+20 DNA + 2xPS	1xPS-OMe	61	21	CCUUGUCAAGGCUAUUGGUCA (SEQ ID NO: 1254)
1085	rUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrArGrArCrArGrArUrArUrUrUrGrCrArUrUrGrArG*mA	-	1xPS-OMe	41	21	AGACAGAUAUUUGCAUUGAGA (SEQ ID NO: 1260)
1086	ATGTGTTTTTGTCAAAAGACCTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrArGrArCrArGrArArUrArUrUrUrGrCrArUrUrGrArG*mA	+25 DNA	1xPS-OMe	66	21	AGACAGAUAUUUGCAUUGAGA (SEQ ID NO: 1260)
1087	C*C*GAAGTTTTCTTCGGTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrArGrArCrArGrArArUrArUrUrUrGrCrArUrUrGrArG*mA	+20 DNA + 2xPS	1xPS-OMe	61	21	AGACAGAUAUUUGCAUUGAGA (SEQ ID NO: 1260)
1088	rUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrArUrUrGrArGrArUrArGrUrGrUrGrGrGrArA*mG	-	1xPS-OMe	41	21	CAUUGAGAUAGUGUGGGGAAG (SEQ ID NO: 1262)
1089	ATGTGTTTTTGTCAAAAGACCTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrArUrUrGrArGrArUrArGrUrGrUrGrGrGrArA	+25 DNA	1xPS-OMe	66	21	CAUUGAGAUAGUGUGGGGAAG (SEQ ID NO: 1262)

	GrGrGrArA*mG					
1090	C*C*GAAGTTTTCT TCGGTTTTrUrArAr UrUrUrCrUrArCrUrC rUrUrGrUrArGrArUr CrArUrUrGrArGrArU rArGrUrGrUrGrGrGr GrArA*mG	+20 DNA + 2xPS	1xPS-O Me	61	21	CAUUGAGAUA GUGUGGGGAA G (SEQ ID NO: 1262)
1091	rUrArArUrUrUrCrUr ArCrUrCrUrUrGrUrA rGrArUrCrUrUrCrUr CrCrCrArUrCrArUrA rGrArGrGrArU*mA	-	1xPS-O Me	41	21	CUUCUCCCAUC AUAGAGGAUA (SEQ ID NO: 1264)
1092	ATGTGTTTTTGTCA AAAGACCTTTTrUr ArArUrUrUrCrUrArC rUrCrUrUrGrUrArGr ArUrCrUrUrCrUrCrC rCrArUrCrArUrArGr ArGrGrArU*mA	+25 DNA	1xPS-O Me	66	21	CUUCUCCCAUC AUAGAGGAUA (SEQ ID NO: 1264)
1093	C*C*GAAGTTTTCT TCGGTTTTrUrArAr UrUrUrCrUrArCrUrC rUrUrGrUrArGrArUr CrUrUrCrUrCrCrCrAr UrCrArUrArGrArGrG rArU*mA	+20 DNA + 2xPS	1xPS-O Me	61	21	CUUCUCCCAUC AUAGAGGAUA (SEQ ID NO: 1264)
1098	A*T*GTGTTTTTGT CAAAAGACCTTTTr UrArArUrUrUrCrUrA rCrUrCrUrUrGrUrAr GrArUrCrCrUrUrGrU rCrArArGrGrCrUrAr UrUrGrGrUrCmA*rU	+25 DNA + 2xPS	1xPS-O Me + rU	67	21	CCUUGUCAAGG CUAUUGGUCA (SEQ ID NO: 1254)
1099	A*T*GTGTTTTTGT CAAAAGACCTTTTr UrArArUrUrUrCrUrA rCrUrCrUrUrGrUrAr GrArUrCrCrUrUrGrU rCrArArGrGrCrUrAr UrUrGrGrUrCmA*m A*rU	+25 DNA + 2xPS	2xPS-O Me + rU	68	22	CCUUGUCAAGG CUAUUGGUCAA (SEQ ID NO: 1256)
1100	A*T*GTGTTTTTGT CAAAAGACCTTTTr UrArArUrUrUrCrUrA rCrUrCrUrUrGrUrAr GrArUrCrCrUrUrGrU rCrArArGrGrCrUrAr UrUrGrGrUmC*rU	+25 DNA + 2xPS	1xPS-O Me + rU	66	20	CCUUGUCAAGG CUAUUGGUC (SEQ ID NO: 1002)
1101	A*T*GTGTTTTTGT CAAAAGACCTTTTr UrArArUrUrUrCrUrA rCrUrCrUrUrGrUrAr GrArUrCrCrUrUrGrU rCrArArGrGrCrUrAr UrUrGrGrU*mC	+25 DNA + 2xPS	1xPS-O Me	65	20	CCUUGUCAAGG CUAUUGGUC (SEQ ID NO: 1002)

1102	A*T*GTGTTTTTGT CAAAAGACCTTTTr UrArArUrUrUrCrUrA rCrUrCrUrUrGrUrAr GrArUrCrCrUrUrGrU rCrArArGrGrCrUrAr UrUrGrGrUrC*mA*m A	+25 DNA + 2xPS	2xPS-O Me	67	22	CCUUGUCAAGG CUAUUGGUCAA (SEQ ID NO: 1256)
1103	mA*mU*rGrUrGrUr UrUrUrUrGrUrCrArA rArArGrArCrCrUrUr UrUrUrArArUrUrUrC rUrArCrUrCrUrUrGr UrArGrArUrCrCrUrU rGrUrCrArArGrGrCr UrArUrUrGrGrUrC* mA	+25 RNA + 2xPS	1xPS-O Me	66	21	CCUUGUCAAGG CUAUUGGUCA (SEQ ID NO: 1254)
1104	mA*mA*rArArArAr ArArArArArArArArA rArArArArArArArAr ArArUrArArUrUrUrC rUrArCrUrCrUrUrGr UrArGrArUrCrCrUrU rGrUrCrArArGrGrCr UrArUrUrGrGrUrC* mA	PolyA RNA + 2xPS	1xPS-O Me	66	21	CCUUGUCAAGG CUAUUGGUCA (SEQ ID NO: 1254)
1105	mU*mU*rUrUrUrUr UrUrUrUrUrUrUrUrU rUrUrUrUrUrUrUrUr UrUrUrArArUrUrUrC rUrArCrUrCrUrUrGr UrArGrArUrCrCrUrU rGrUrCrArArGrGrCr UrArUrUrGrGrUrC* mA	聚 U RNA + 2xPS	1xPS-O Me	66	21	CCUUGUCAAGG CUAUUGGUCA (SEQ ID NO: 1254)
所有碱基均为大写 小写字母“r”代表 RNA，2’-羟基；没有被“r”修饰的碱基是 DNA 所有碱基均通过标准磷酸二酯键连接，除另有说明外： “*”代表硫代磷酸酯修饰 “PS”代表硫代磷酸酯修饰 “PS2”代表二硫代磷酸酯修饰 “OMe”代表 2’-o-甲基修饰 “m”代表 2’-o-甲基修饰 **表 24 提供了 gRNA 5’延伸的序列的列表						

序列

[0441] 根据本公开内容的基因组编辑系统组分(包括但不限于,RNA指导的核酸酶、指导RNA、供体模板核酸、编码核酸酶或指导RNA的核酸、以及任何前述的部分或片段),用序列表中表示的核苷酸和氨基酸序列例示。序列表中表示的序列不旨在是限制性的,而是说明性的基因组编辑系统及其组分部分的某些原理,组合本公开内容,将通知本领域技术人员另外的本公开内容范围内的实施和修饰。呈现的序列的列表在以下表25中提供。

表25:序列表中呈现的序列:

SEQ ID NO:	描述
1-2、4-6、12、14	Cas9 多肽
3、7-11、13	Cas9 编码序列
15-23、52-123	Cas9 RuvC 样结构域
24-28, 124-198	Cas9 HNH 样结构域
29-31、38-51	全长模块化和单分子 gRNA
32-37	gRNA 近端和尾部结构域
199-205	PAM 序列
251-901, 1002	gRNA 靶向结构域(RNA)-参见表 2, 19, 22, 23
910-919, 943-945, 956-959, 1003	gRNA 靶向结构域(DNA)-参见表 7, 9
920-929, 946-948, 960-963, 1004	gRNA 靶向结构域加上 PAM (NGG) (RNA)-参见表 7, 9
930-939, 949-951, 964-967, 1005	gRNA 靶向结构域加上 PAM (NGG) (DNA)-参见表 7, 9
970, 971, 996, 997	gRNA 序列(DNA)-参见表 10, 12
972, 973, 998, 999	gRNA 序列(DNA)-参见表 10, 12
902, 903	人 *HBG1* 和 *HBG2* 启动子序列，包括 HPFH 缺失位点
904-909, 974-995	寡核苷酸供体序列和同源臂 - 参见表 8, 11
968-969	*BCL11Ae* 序列
1006, 1007	核定位信号序列
1000, 1001, 1008-1018 1032-1039 1094-1097, 1107-1109	Cpf1 多肽序列
1019-1021 1110-1117	Cpf1 核苷酸序列

通过引用并入

[0442] 本文提及的所有出版物、专利和专利申请都通过引用以其整体而特此并入，如同每一单独的出版物、专利或专利申请具体且单独地指明通过引用而并入一样。在有冲突的

情况下,以本申请(包括本文的任何定义)为准。

等效物

[0443] 本领域技术人员仅使用常规实验就将认识到或能够确定本文所述的具体实施例的许多等效物。此类等效物旨在由以下权利要求涵盖。

参考文献

Ahern等人,Br J Haematol[英国血液学杂志]25(4):437-444(1973)

Akinbami,Hemoglobin[血红蛋白]40:64-65(2016)

Aliyu等人,Am J Hematol[美国血液学杂志]83:63-70(2008)

Anders等人,Nature[自然]513(7519):569-573(2014)

Angastiniotis和Modell,Ann N Y Acad Sci[纽约科学院年报]850:251-269(1998)

Bae等人,Bioinformatics[生物信息学]30(10):1473-1475(2014)

Barbosa等人,Braz J Med Bio Res[巴西医学和生物研究杂志]43(8):705-711(2010)

Bouva,Hematologica[血液学]91(1):129-132(2006)

Briner等人Mol Cell[分子细胞]56(2):333-339(2014)

Brousseau,Am J Hematol[美国血液学杂志]85(1):77-78(2010)

Caldecott,Nat Rev Genet[自然遗传学评论]9(8):619-631(2008)

Canvers等人Nature[自然]527(12):192-197(2015)

Chang等人Mol Ther Methods Clin Dev[分子治疗方法与临床开发]4:137-148(2017)

Chassanidis,Ann Hematol[血液学年鉴]88(6):549-555(2009)

Chylinski等人,RNA Biol[RNA生物学]10(5):726-737(2013)

Cong等人,Science[科学]399(6121):819-823(2013)

Costa等人,Cad Saude Publica 18(5):1469-1471(2002)

Davis&Maizels 2 Proc Natl Acad Sci USA[美国国家科学院院刊]111(10):E924-932(2014)

Fine等人,Sci Rep.[科学报告]5:10777(2015)

Frit等人DNA Repair[DNA修复](Amst)17:81-97(2014)

Fu等人Nat Biotechnol[自然生物技术]32:279-284(2014)

Gao等人Nat Biotechnol.[自然生物技术]:35(8):789-792(2017)

Giarratana等人Blood[血液]:118,5071-5079(2011)

Guilinger等人,Nat Biotechnol[自然生物技术]32:577-582(2014)

Heigwer等人Nat Methods[自然方法]11(2):122-123(2014)

Hsu等人Nat Biotechnol[自然生物技术]31(9):827-832(2013)

Iyama&Wilson DNA Repair[DNA修复](Amst)12(8):620-636(2013)

Jiang等人Nat Biotechnol[自然生物技术]31(3):233-239(2013)

Jinek等人,Science[科学]337(6096):816-821(2012)

Jinek等人,Science[科学]343(6176):1247997(2014)

Kleinstiver等人,Nature[自然]523(7561):481-485(2015a)

Kleinstiver等人,Nat Biotechnol[自然生物技术]33(12):1293-1298(2015b)

Kleinstiver等人,Nature[自然]529(7587):490-495(2016)

Komor等人Nature[自然]533(7603):420-424(2016)

Lee等人,Nano Lett[纳米快报]12(12):6322-6327(2012)

Lewis“Medical-Surgical Nursing:Assessment and Management of Clinical Problems”[医学外科护理:临床问题的评估与管理](2014)

Li,Cell Res[细胞研究]18(1):85-98(2008)

Makarova等人Nat Rev Microbiol[自然微生物学评论]9(6):467-477(2011)

Mali等人Science[科学]339(6121):823-826(2013)

Mantovani等人,Nucleic Acids Res[核酸研究]16(16):7783-7797(1988)

Marteijn等人,Nat Rev Mol Cell Biol[自然分子细胞生物学评论]15(7):465-481(2014)

Martyn等人,Biochim Biophys Acta[生物化学与生物物理学学报]1860(5):525-536(2017)

Nishimasu等人,Cell[细胞]156(5):935-949(2014)

Nishimasu等人Cell[细胞]162:1113-1126(2015)

Notta等人Science[科学]333(6039):218-21(2011)

Ran&Hsu Cell[细胞]154(6):1380-1389(2013)

Richardson等人Nat Biotechnol[自然生物技术]34:339-344(2016)

Shmakov等人,Molecular Cell[分子细胞]60(3):385-397(2015)

Sternberg等人Nature[自然]507(7490):62-67(2014)

Superti-Furga等人,EMBO J[欧洲分子生物学杂志]7(10):3099-3107(1988)

Thein,Hum Mol Genet[人类分子遗传学]18(R2):R216-223(2009)

Tsai等人Nat Biotechnol[自然生物技术]34(5):483(2016)

Waber等人,Blood[血液]67(2):551-554(1986)

Wang等人,Cell[细胞]153(4):910-918(2013)

Xiao等人Bioinformatics[生物信息学]30(8):1180-1182(2014)

Xu等人,Genes Dev[基因与发育]24(8):783-798(2010)

Yamano等人,Cell[细胞]165(4):949-962(2016)

Zetsche等人,Nat Biotechnol[自然生物技术]33(2):139-42(2015)

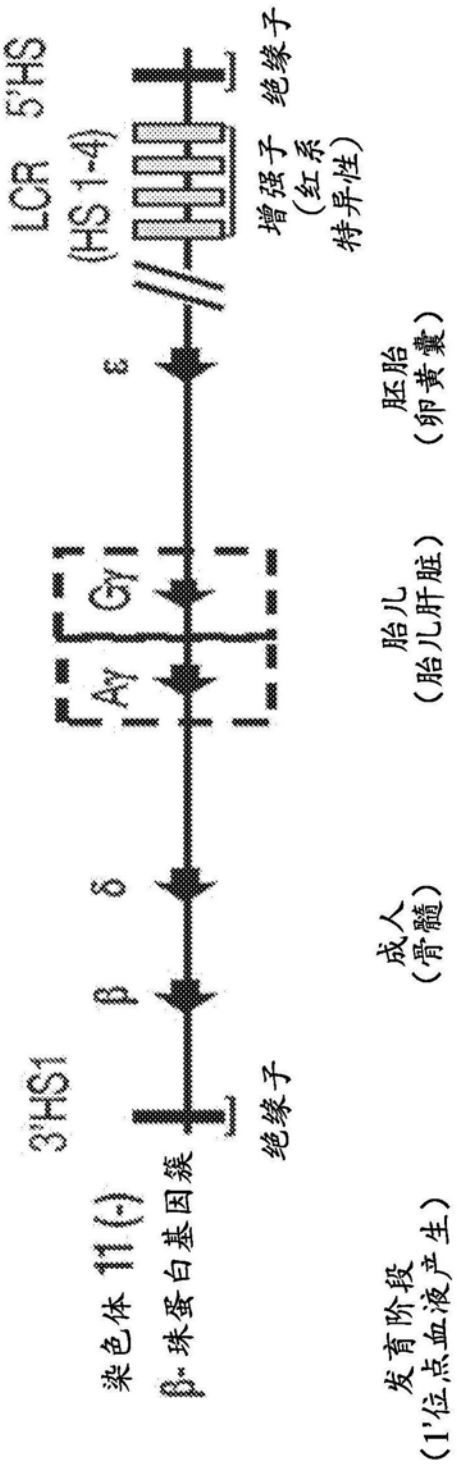

染色体 11(-)
β-珠蛋白基因簇
5'HS
LCR
(HS 1-4)
3'HS1
β
δ
Aγ
Gγ
ε
绝缘子
绝缘子
增强子
(红系
特异性)
发育阶段
(1'位点血液产生)
成人
(骨髓)
胎儿
(胎儿肝脏)
胚胎
(卵黄囊)

图1

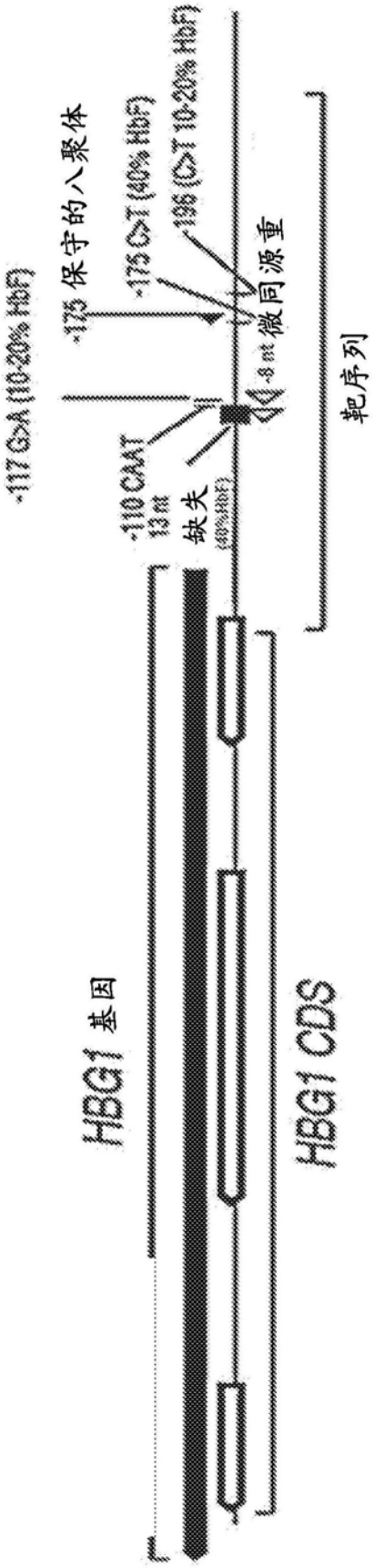

-117 G>A (10-20% HbF)
-175 保守的八聚体
-175 C>T (40% HbF)
-196 (C>T 10-20% HbF)
-110 CAAT
13 nt
缺失
(40% HbF)
-8 nt 微同源重
靶序列
HBG1 基因
HBG1 CDS

图2A

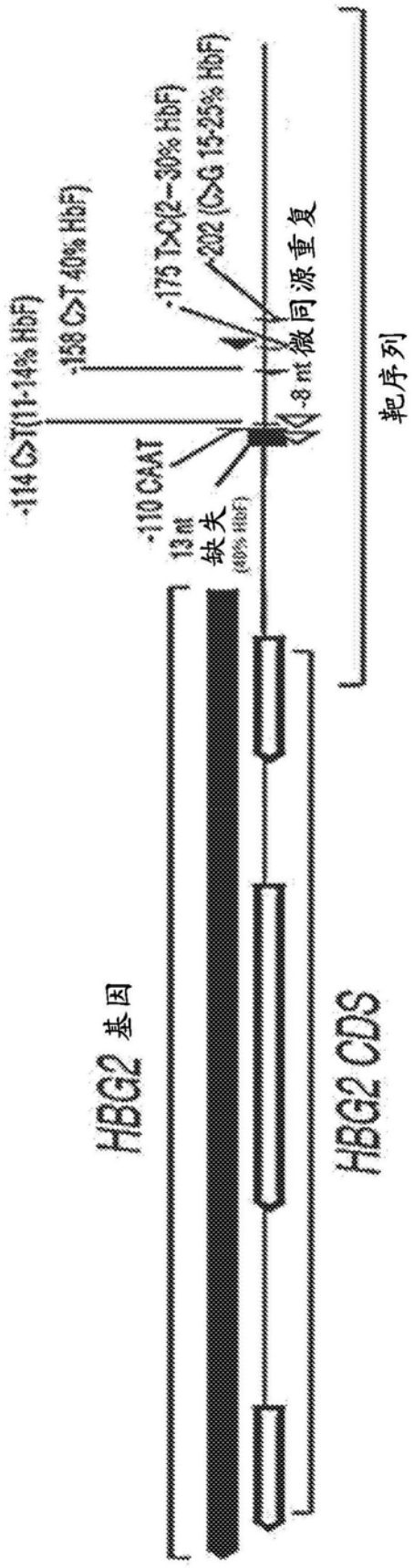

-114 C>T(11-14% HbF)
-158 C>T 40% HbF
-175 T>C(2-30% HbF)
-202 (C>G 15-25% HbF)
-110 CMAT
13 nt
缺失
(40% HbF)
(1-8 nt 微同源重复
靶序列
HBG2 基因
HBG2 CDS

图2B

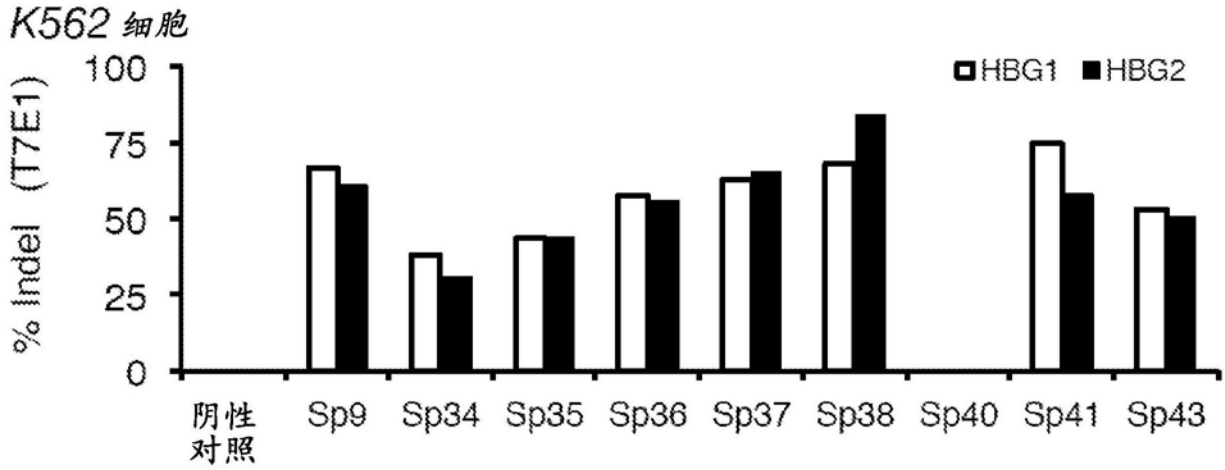

K562 细胞
% Indel (T7E1)
100
75
50
25
0
HBG1
HBG2
阴性对照
Sp9
Sp34
Sp35
Sp36
Sp37
Sp38
Sp40
Sp41
Sp43

图3A

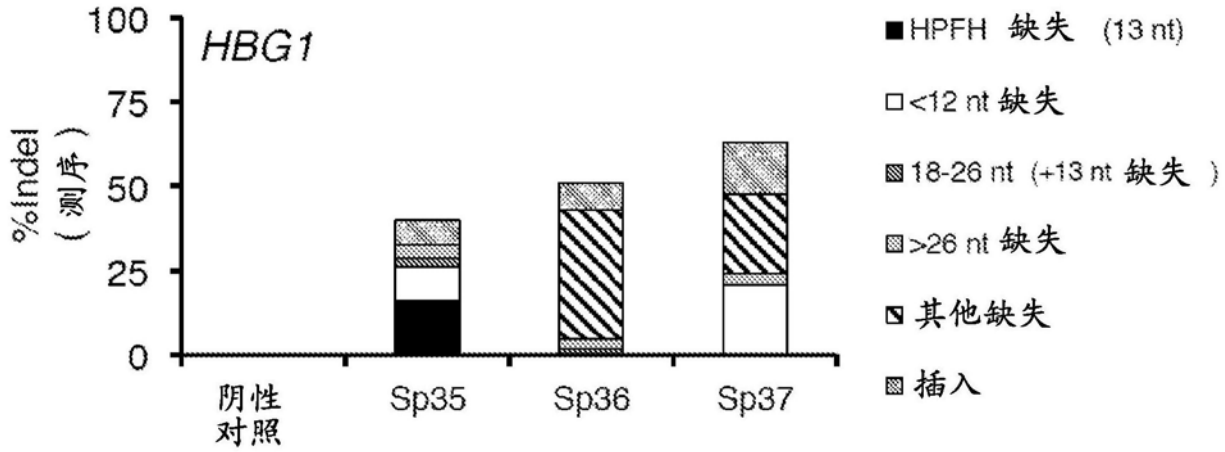

HBG1
%indel (测序)
100
75
50
25
0
阴性对照
Sp35
Sp36
Sp37
HPFH 缺失 (13 nt)
<12 nt 缺失
18-26 nt (+13 nt 缺失)
>26 nt 缺失
其他缺失
插入

图3B

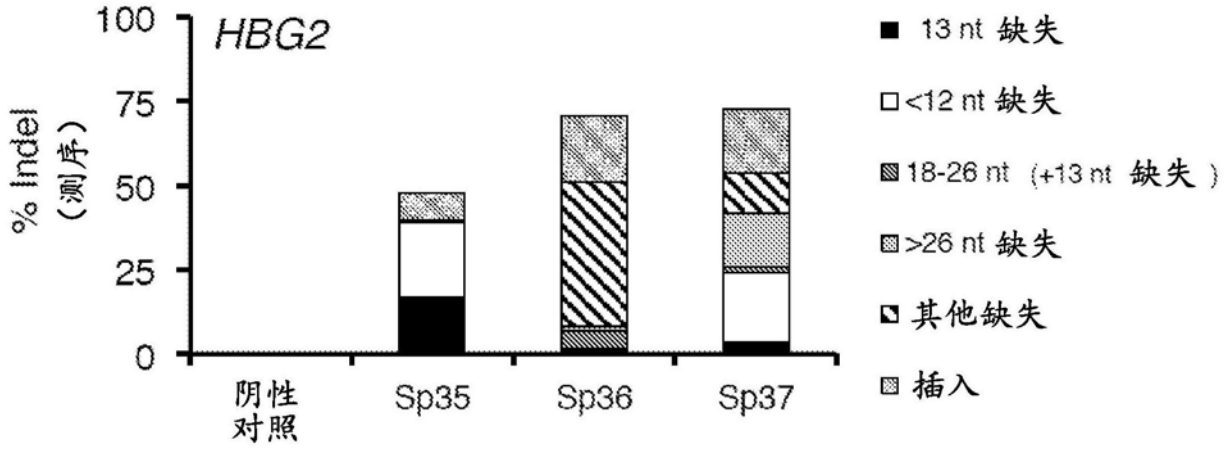

HBG2
% Indel (测序)
100
75
50
25
0
阴性对照
Sp35
Sp36
Sp37
13 nt 缺失
<12 nt 缺失
18-26 nt (+13 nt 缺失)
>26 nt 缺失
其他缺失
插入

图3C

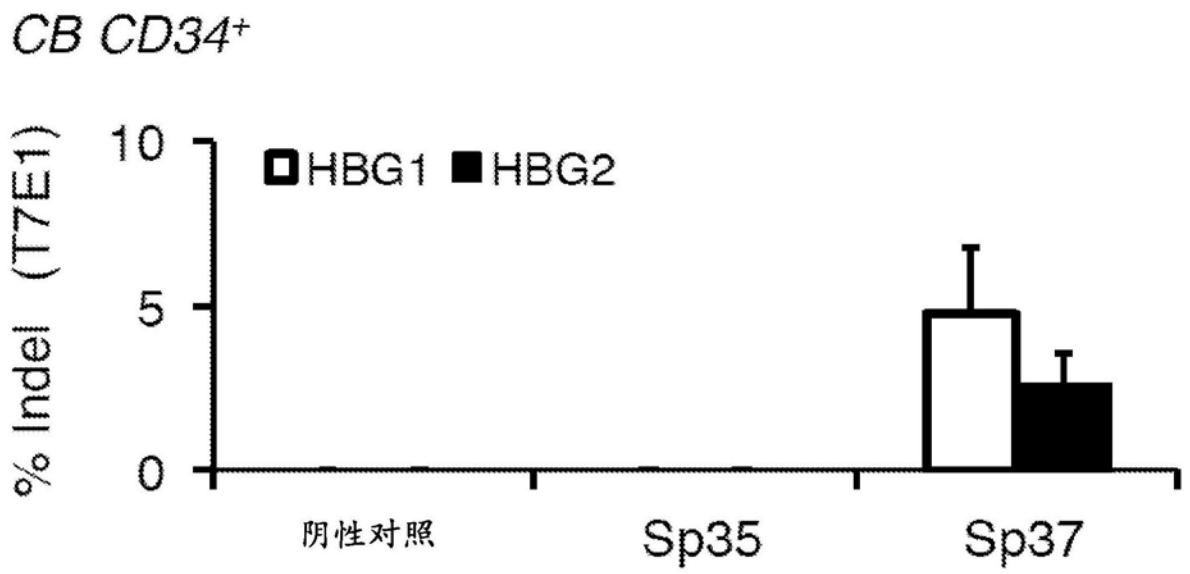

CB CD34+
% Indel (T7E1)
10
5
0
HBG1
HBG2
阴性对照
Sp35
Sp37

图4A

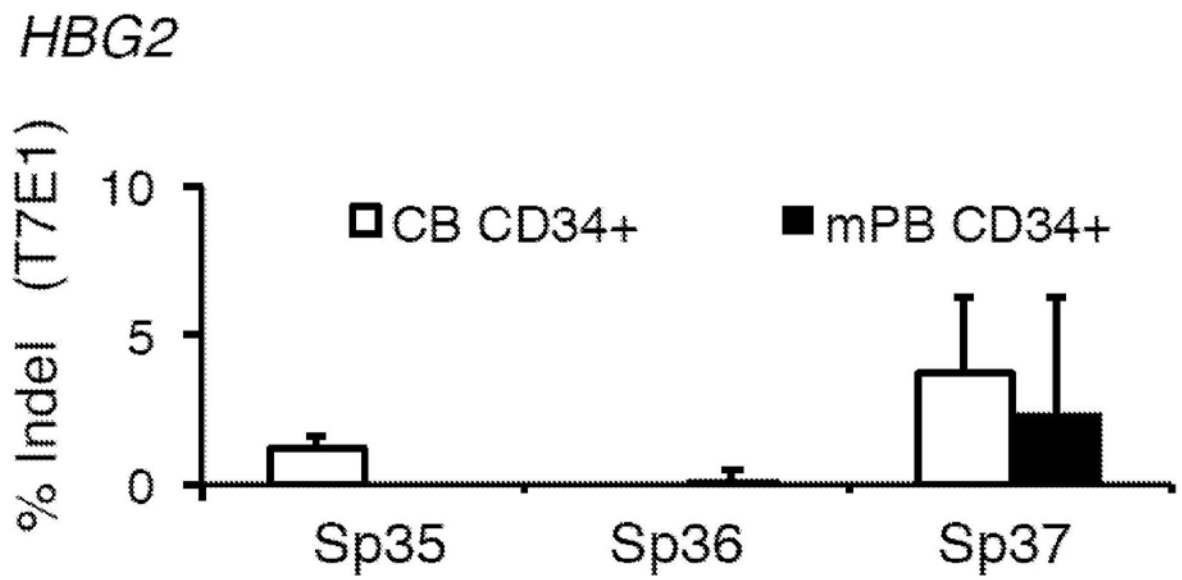

HBG2
% Indel (T7E1)
10
5
0
CB CD34+
mPB CD34+
Sp35
Sp36
Sp37

图4B

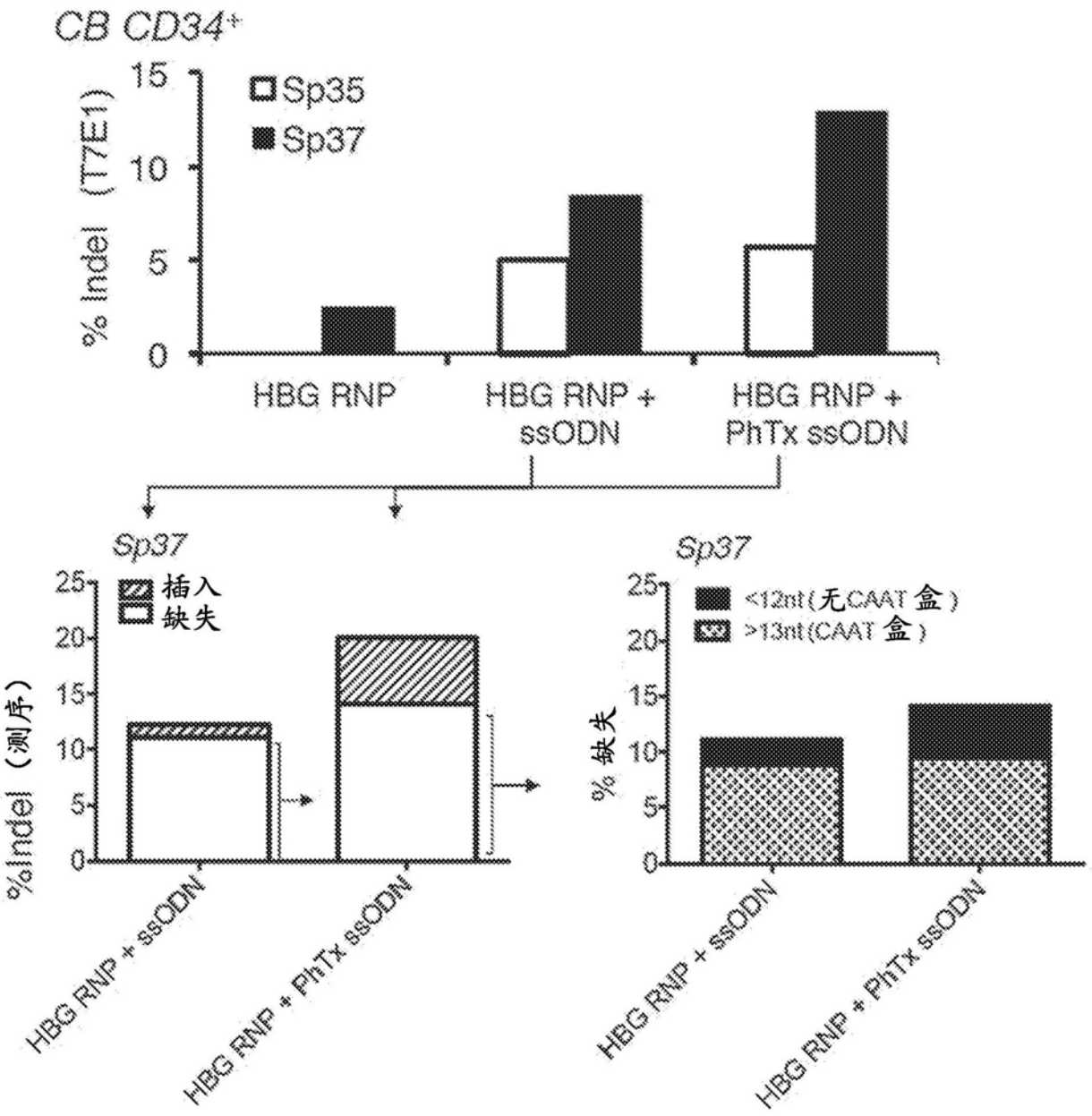

CB CD34+
% Indel (T7E1)
15
10
5
0
Sp35
Sp37
HBG RNP
HBG RNP + ssODN
HBG RNP + PhTx ssODN
Sp37
插入
缺失
%Indel（测序）
25
20
15
10
5
0
HBG RNP + ssODN
HBG RNP + PhTx ssODN
Sp37
<12nt(无CAAT 盒)
>13nt(CAAT 盒)
% 缺失
25
20
15
10
5
0
HBG RNP + ssODN
HBG RNP + PhTx ssODN

图4C

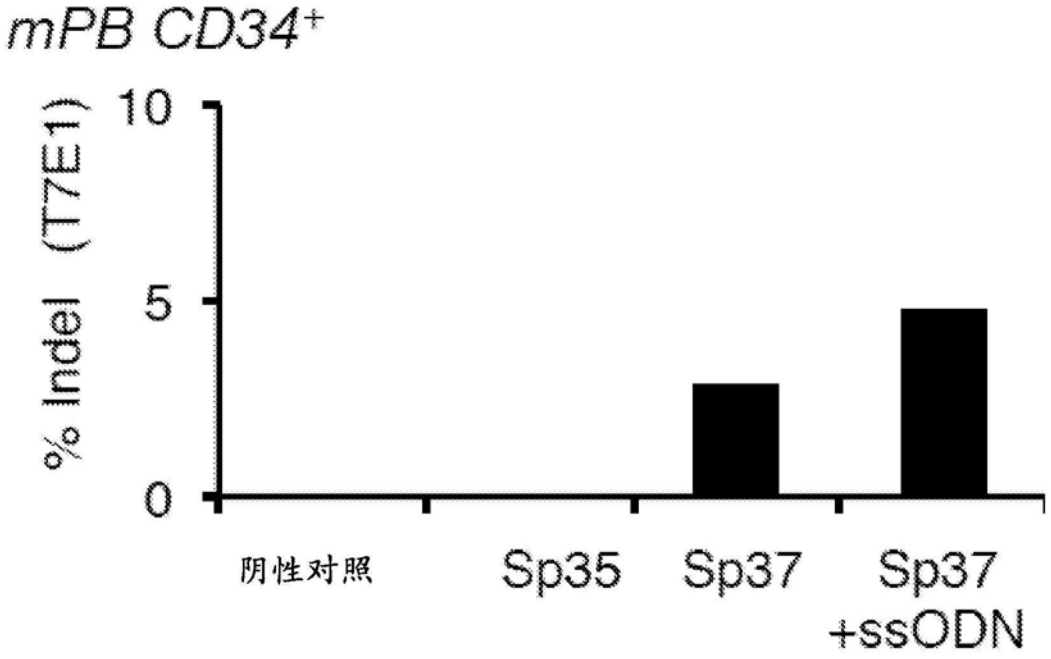

mPB CD34+
% Indel (T7E1)
10
5
0
阴性对照
Sp35
Sp37
Sp37
+ssODN

图5A

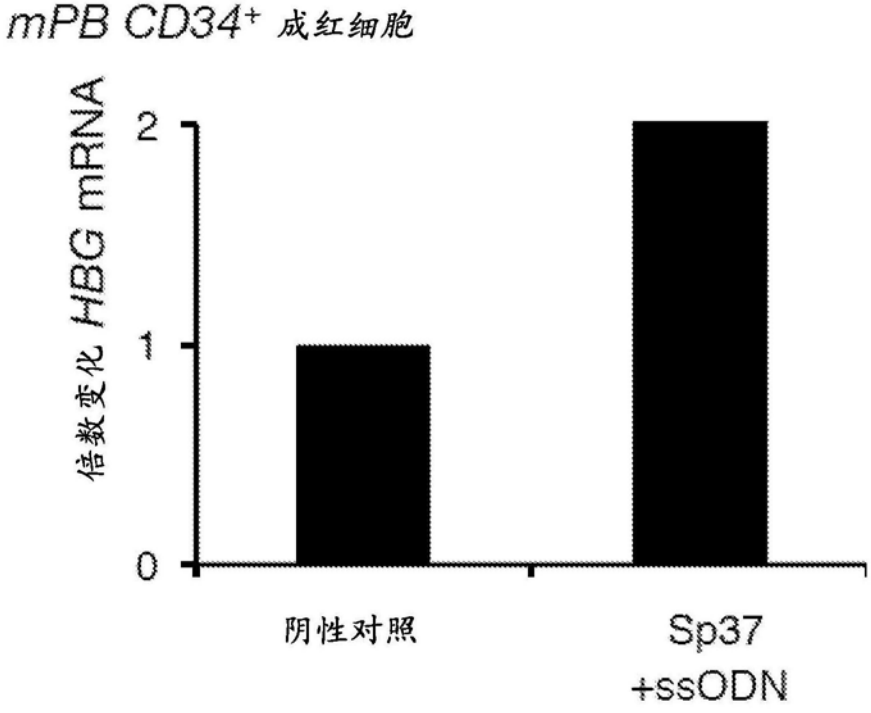

mPB CD34+ 成红细胞
倍数变化 HBG mRNA
2
1
0
阴性对照
Sp37
+ssODN

图5B

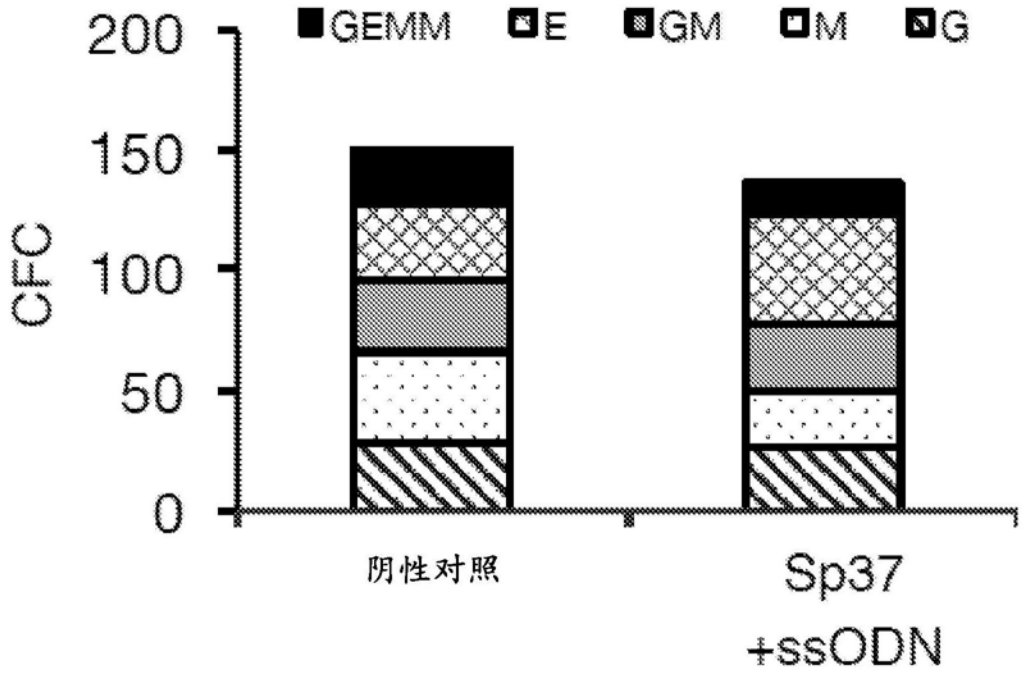

GEMM
E
GM
M
G
200
150
100
50
0
CFC
阴性对照
Sp37
+ssODN

图6A

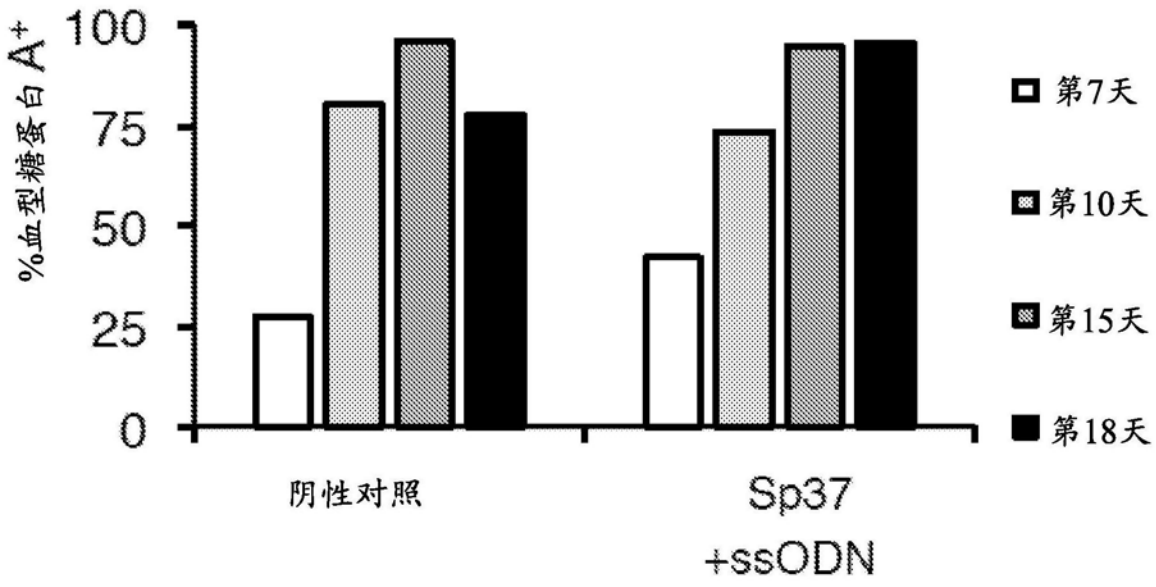

100
75
50
25
0
%血型糖蛋白A+
第7天
第10天
第15天
第18天
阴性对照
Sp37
+ssODN

图6B

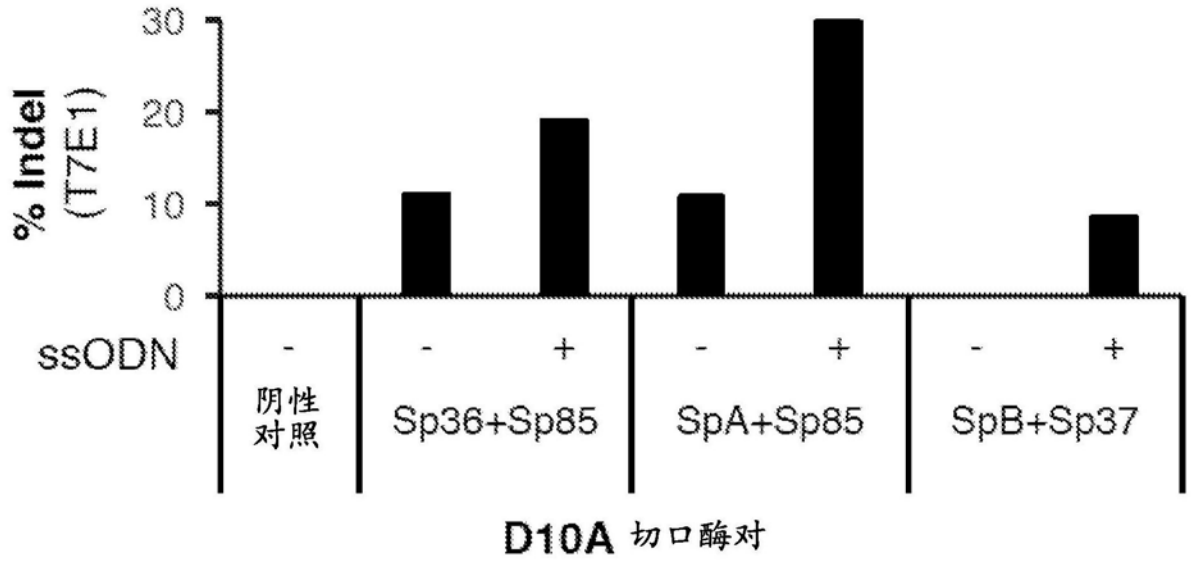

30
20
10
0
% Indel
(T7E1)
ssODN
-
-
+
-
+
-
+
阴性
对照
Sp36+Sp85
SpA+Sp85
SpB+Sp37
D10A 切口酶对

图7A

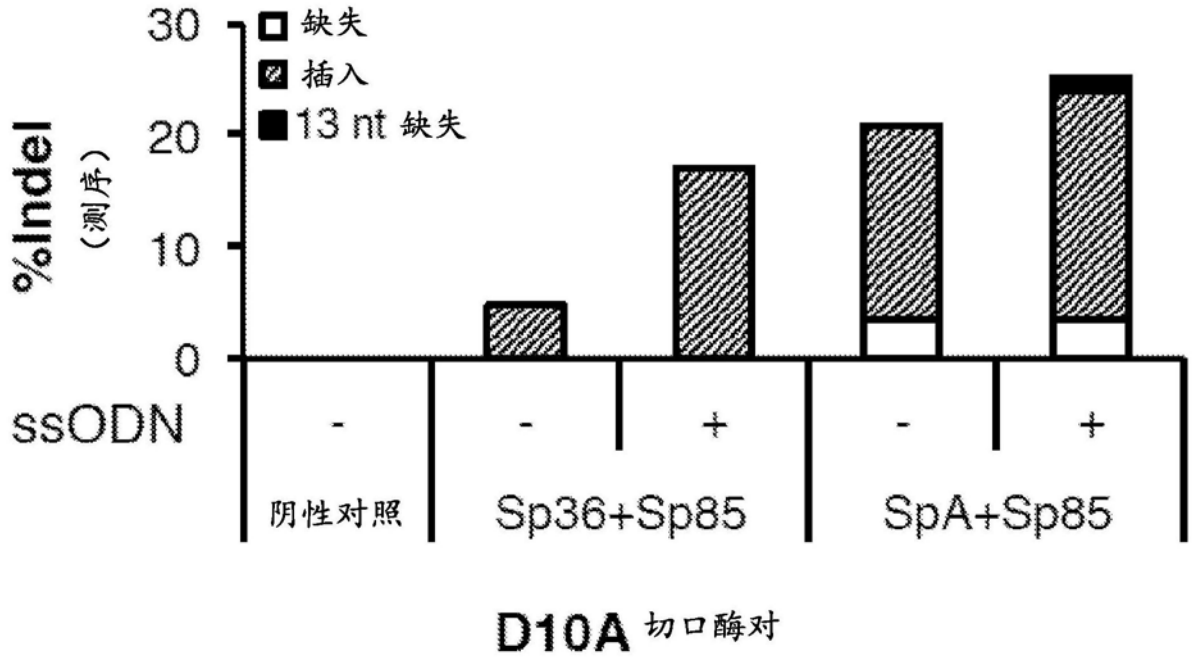

30
缺失
插入
13 nt 缺失
20
10
0
%Indel
(测序)
ssODN
-
-
+
-
+
阴性对照
Sp36+Sp85
SpA+Sp85
D10A 切口酶对

图7B

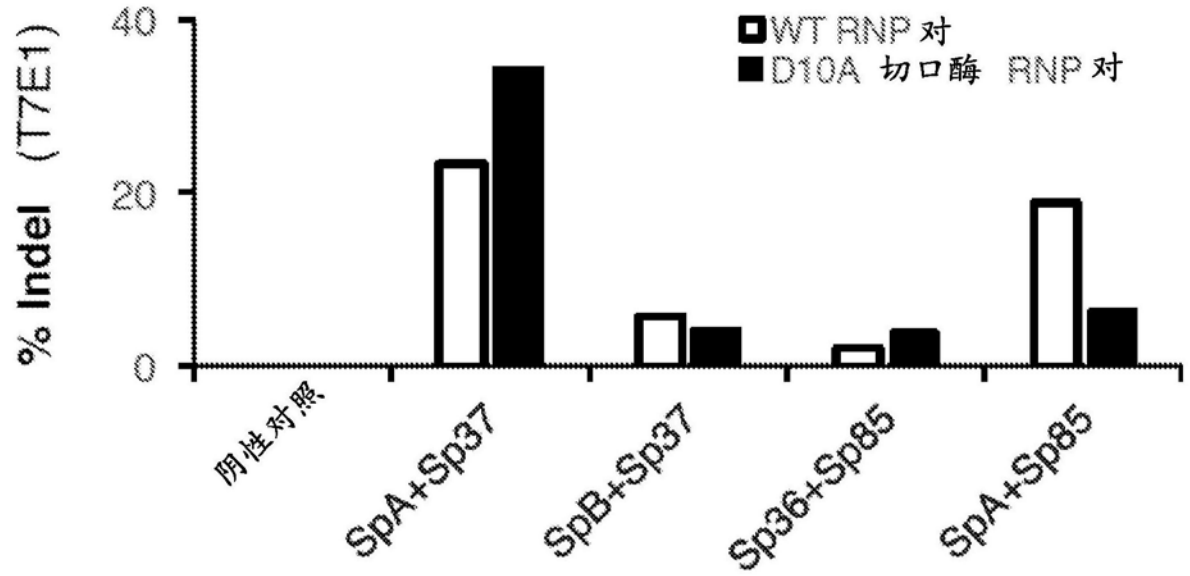

% Indel (T7E1)
40
20
0
WT RNP 对
D10A 切口酶 RNP 对
阴性对照
SpA+Sp37
SpB+Sp37
Sp36+Sp85
SpA+Sp85

图8A

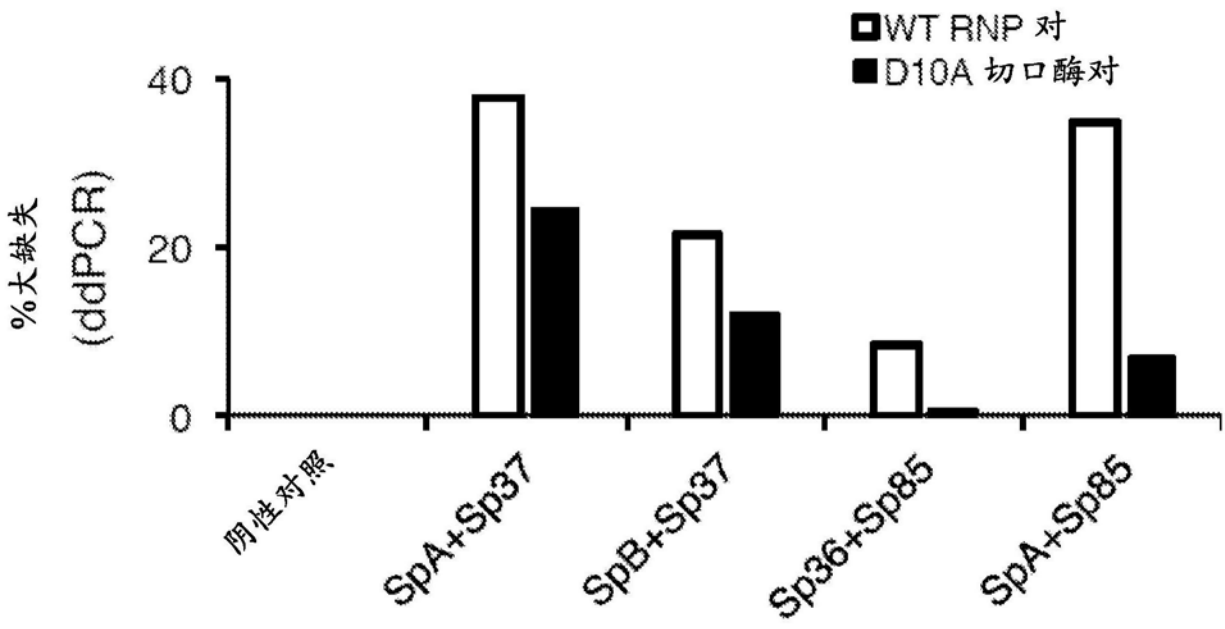

%大缺失
(ddPCR)
40
20
0
WT RNP 对
D10A 切口酶对
阴性对照
SpA+Sp37
SpB+Sp37
Sp36+Sp85
SpA+Sp85

图8B

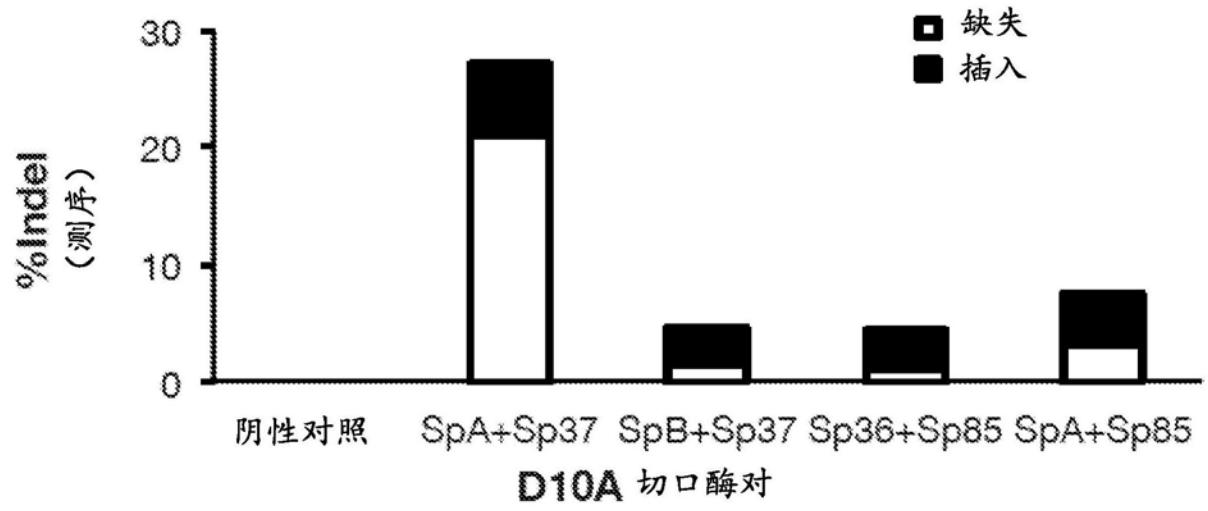

缺失
插入
%Indel
（测序）
30
20
10
0
阴性对照
SpA+Sp37
SpB+Sp37
Sp36+Sp85
SpA+Sp85
D10A 切口酶对

图8C

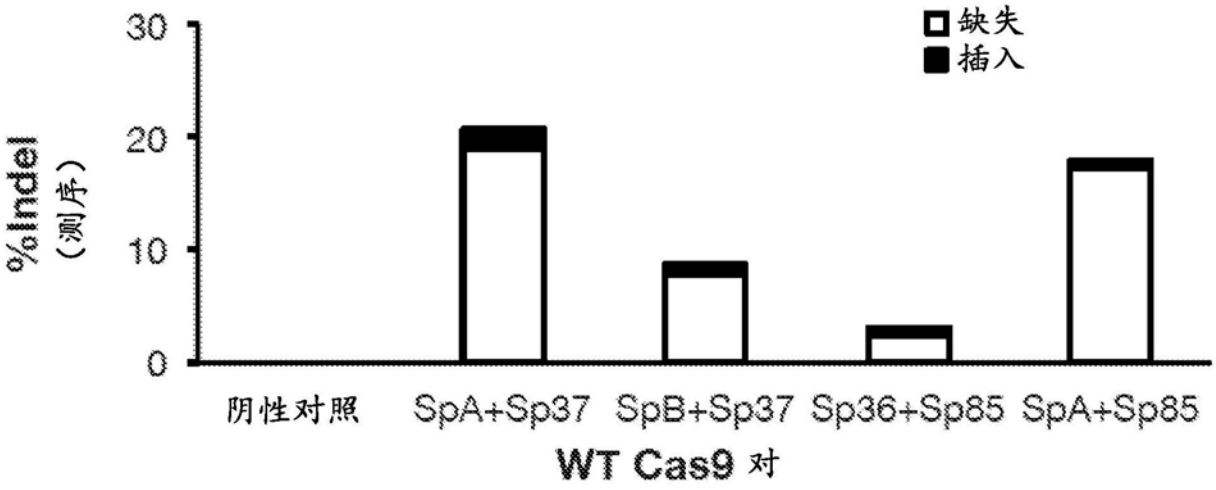

缺失
插入
%Indel
（测序）
30
20
10
0
阴性对照
SpA+Sp37
SpB+Sp37
Sp36+Sp85
SpA+Sp85
WT Cas9 对

图8D

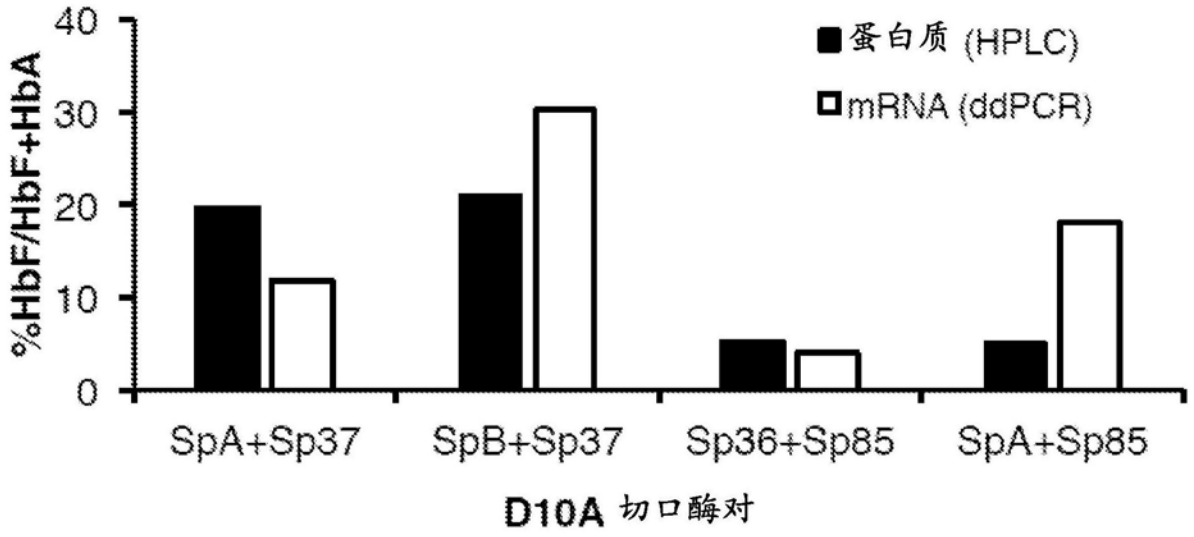

%HbF/HbF+HbA
40
30
20
10
0
蛋白质 (HPLC)
mRNA (ddPCR)
SpA+Sp37
SpB+Sp37
Sp36+Sp85
SpA+Sp85
D10A 切口酶对

图9

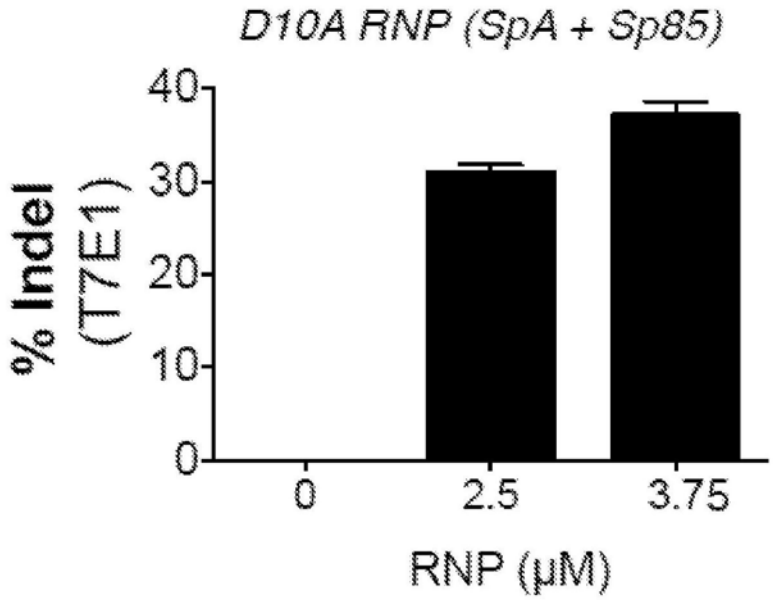

D10A RNP (SpA + Sp85)
% Indel
(T7E1)
40
30
20
10
0
0
2.5
3.75
RNP (μM)

图10A

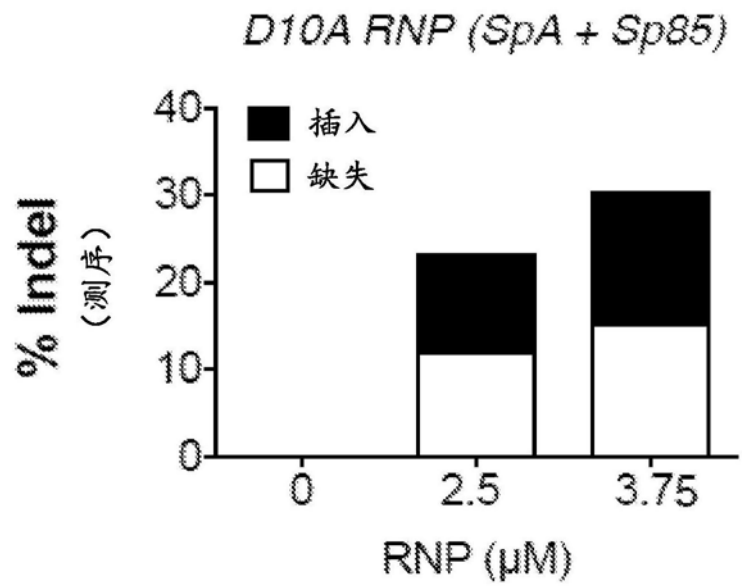

D10A RNP (SpA + Sp85)
插入
缺失
% Indel
(测序)
40
30
20
10
0
0
2.5
3.75
RNP (μM)

图10B

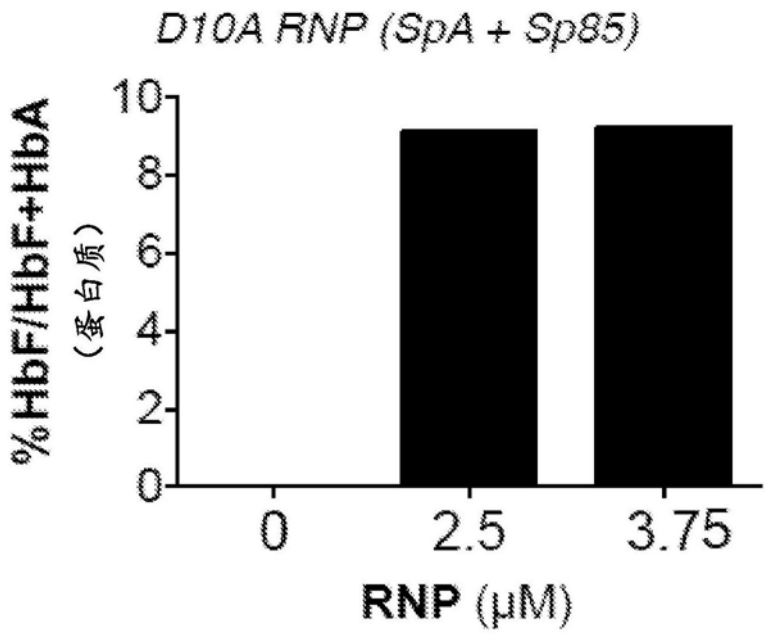

D10A RNP (SpA + Sp85)
%HbF/HbF+HbA
(蛋白质)
10
8
6
4
2
0
0
2.5
3.75
RNP (μM)

图10C

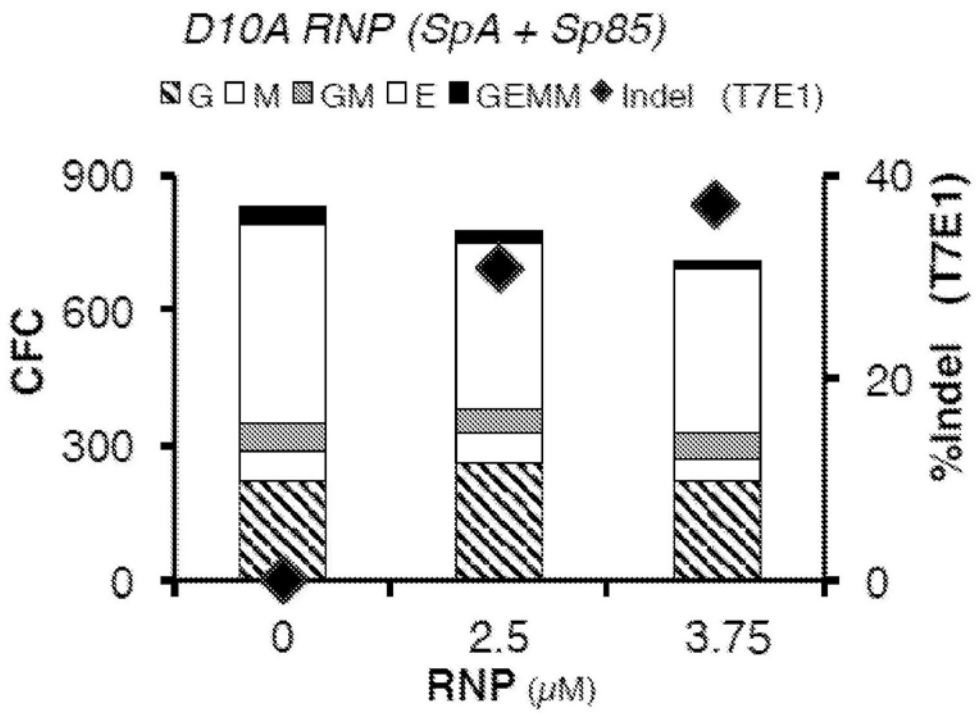

D10A RNP (SpA + Sp85)
G M GM E GEMM Indel (T7E1)
CFC
900
600
300
0
40
20
0
%Indel (T7E1)
0
2.5
3.75
RNP (μM)

图10D

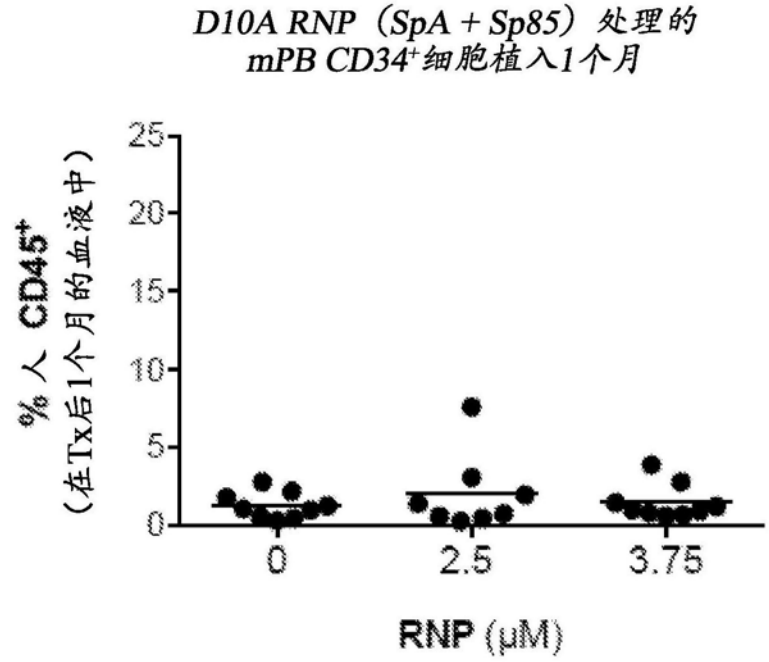

D10A RNP（SpA + Sp85）处理的
mPB CD34+细胞植入1个月
% 人 CD45+
（在Tx后1个月的血液中）
25
20
15
10
5
0
0
2.5
3.75
RNP (μM)

图10E

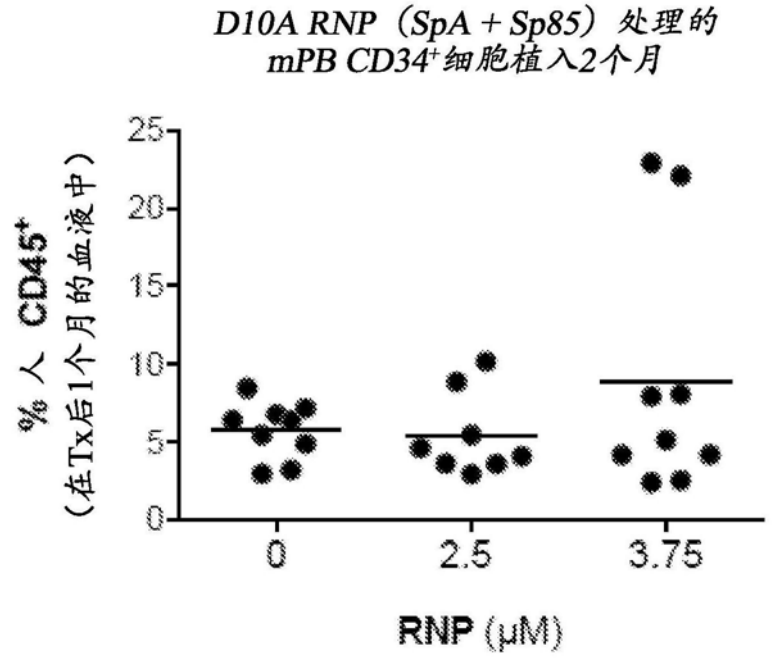

D10A RNP（SpA + Sp85）处理的
mPB CD34+细胞植入2个月
% 人 CD45+
（在Tx后1个月的血液中）
25
20
15
10
5
0
0
2.5
3.75
RNP (μM)

图10F

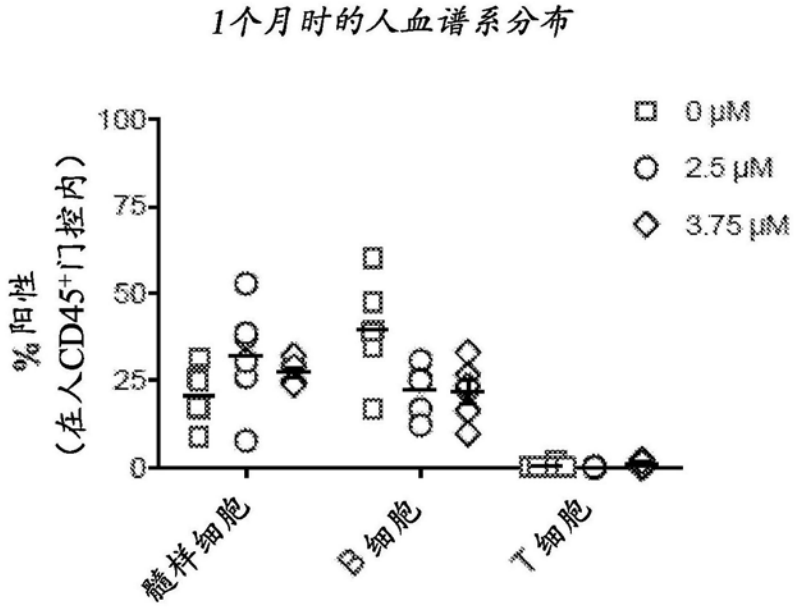

1个月时的人血谱系分布
% 阳性
（在人CD45+门控内）
100
75
50
25
0
0 μM
2.5 μM
3.75 μM
髓样细胞
B 细胞
T 细胞

图10G

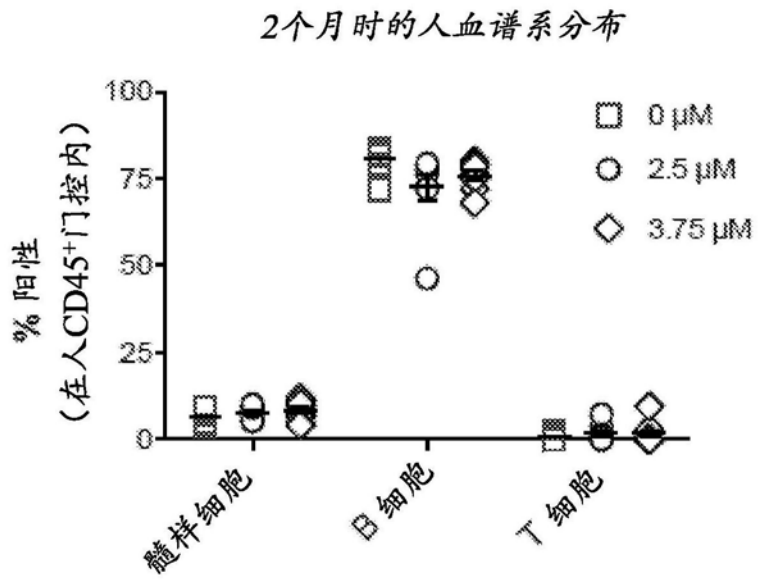

2个月时的人血谱系分布
% 阳性
（在人CD45+门控内）
100
75
50
25
0
0 μM
2.5 μM
3.75 μM
髓样细胞
B细胞
T细胞

图10H

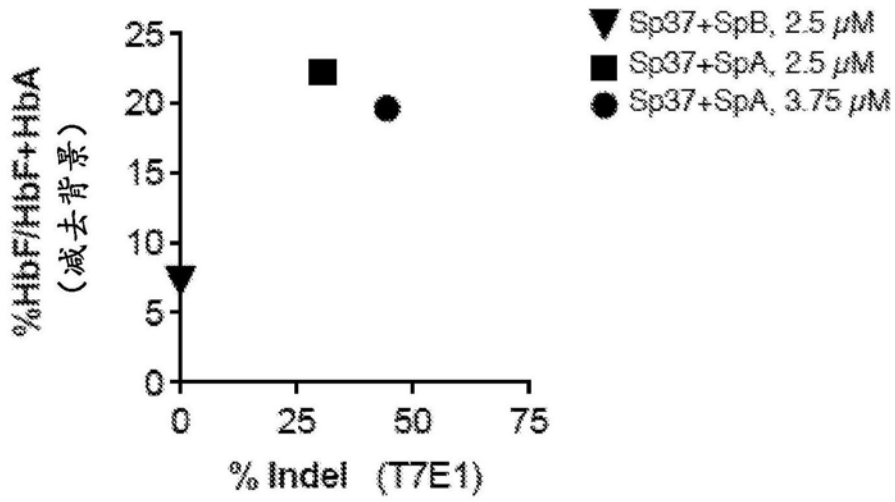

%HbF/HbF+HbA
（减去背景）
25
20
15
10
5
0
0
25
50
75
% Indel (T7E1)
Sp37+SpB, 2.5 μM
Sp37+SpA, 2.5 μM
Sp37+SpA, 3.75 μM

图11A

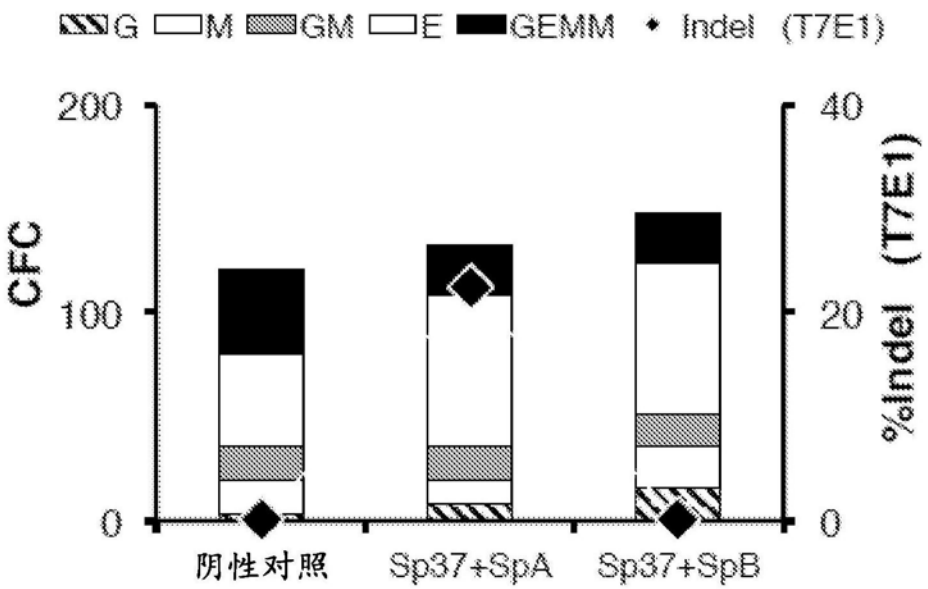

G M GM E GEMM Indel (T7E1)
200
100
0
CFC
40
20
0
%Indel (T7E1)
阴性对照
Sp37+SpA
Sp37+SpB

图11B

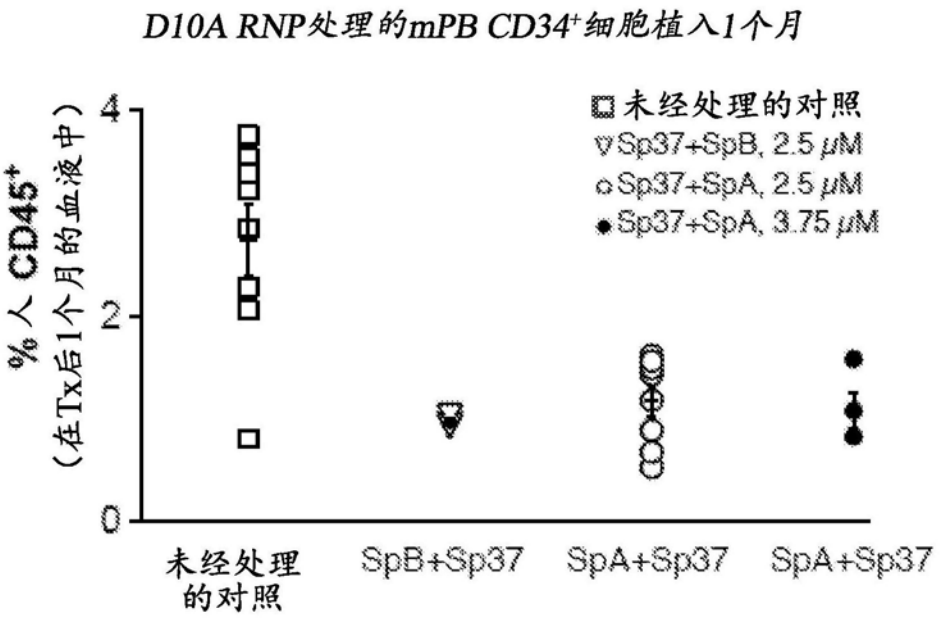

D10A RNP处理的mPB CD34+细胞植入1个月
% 人 CD45+
（在Tx后1个月的血液中）
4
2
0
未经处理的对照
Sp37+SpB, 2.5 μM
Sp37+SpA, 2.5 μM
Sp37+SpA, 3.75 μM
未经处理的对照
SpB+Sp37
SpA+Sp37
SpA+Sp37

图11C

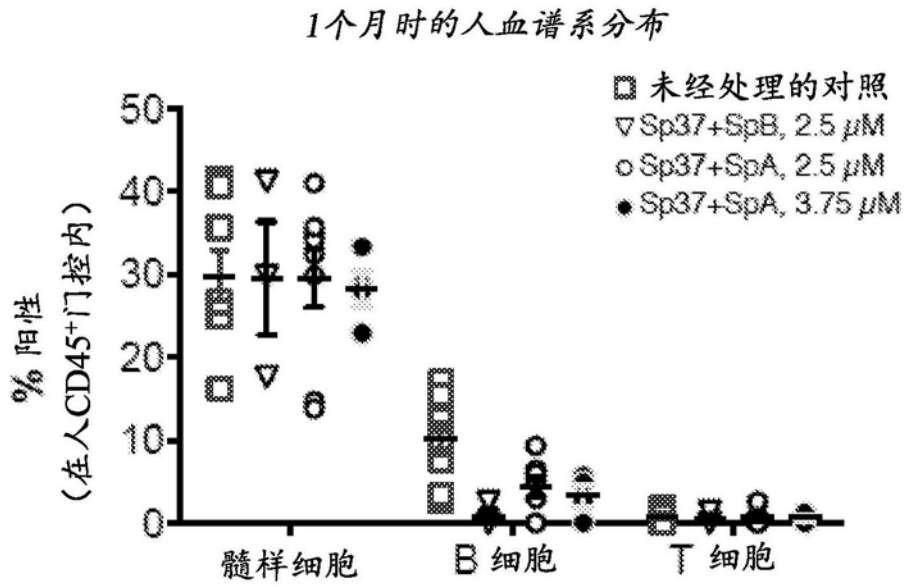

1个月时的人血谱系分布
未经处理的对照
Sp37+SpB, 2.5 μM
Sp37+SpA, 2.5 μM
Sp37+SpA, 3.75 μM
% 阳性
(在人CD45+门控内)
50
40
30
20
10
0
髓样细胞
B 细胞
T 细胞

图11D

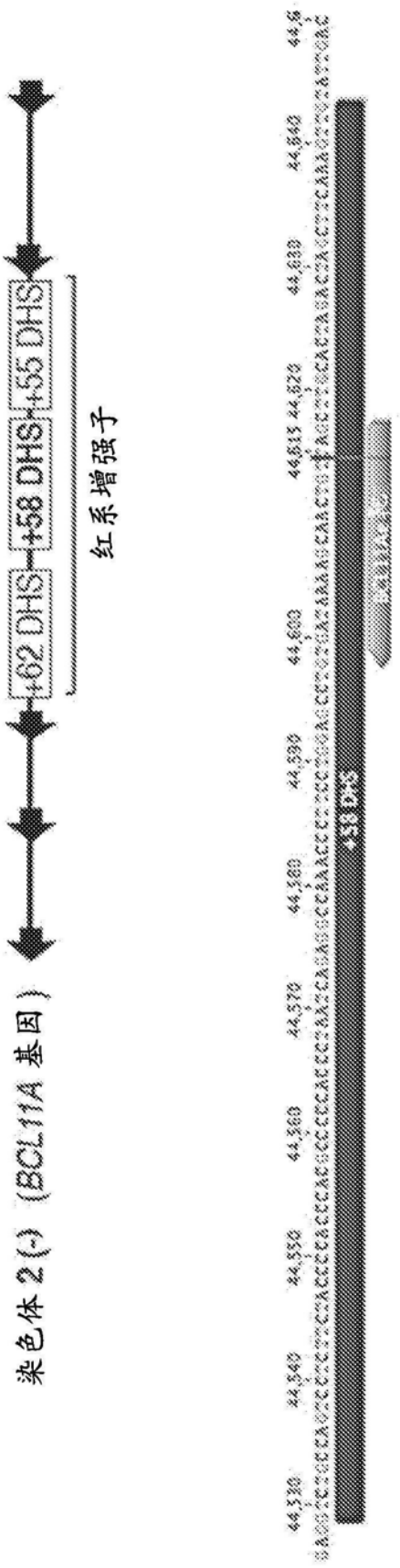


图12

转让（Attorn）

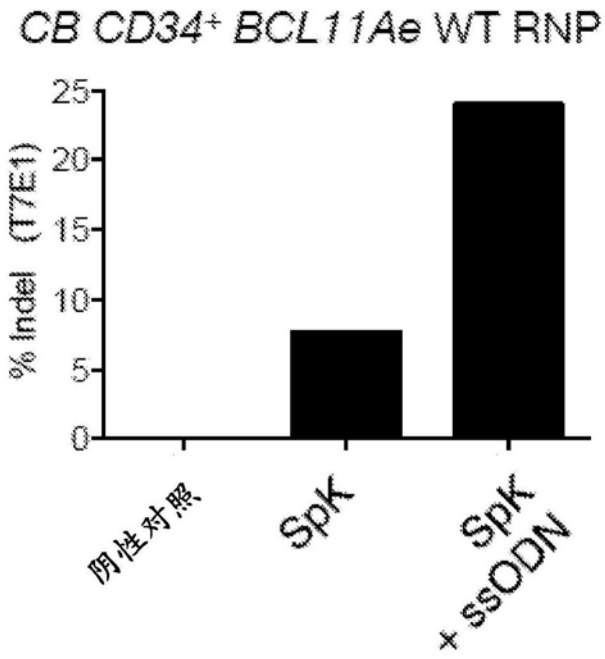


图13A

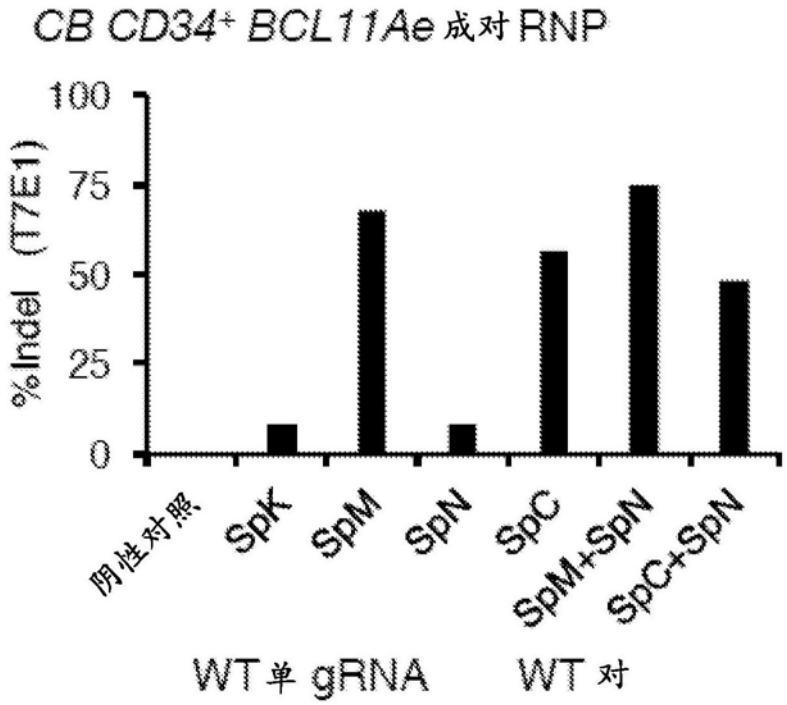


图13B

转让（Attorn）

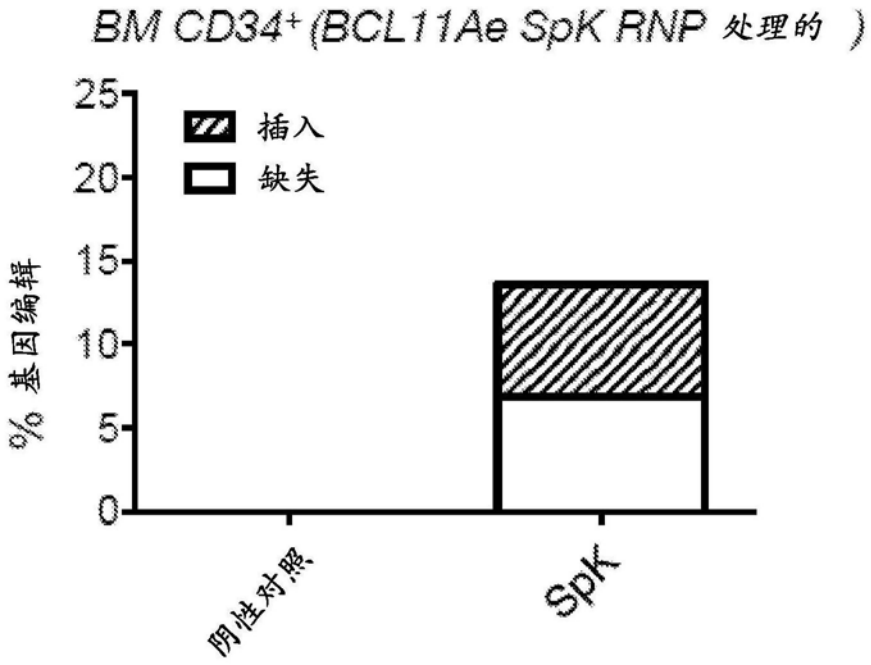


图14A

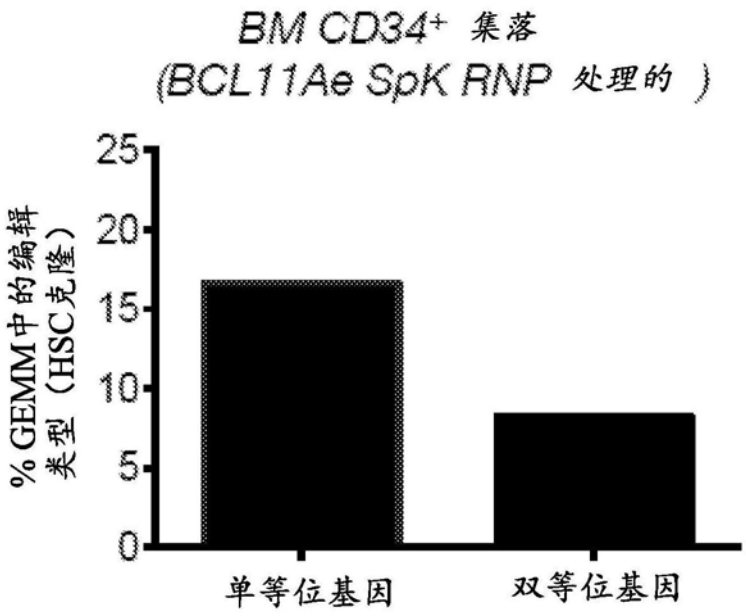


图14B

转让（Attorn）

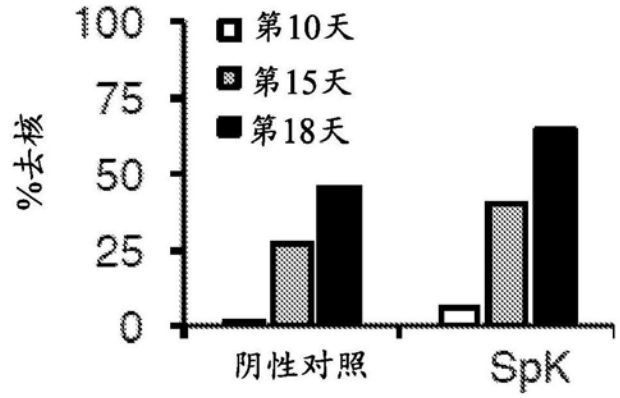


图14C

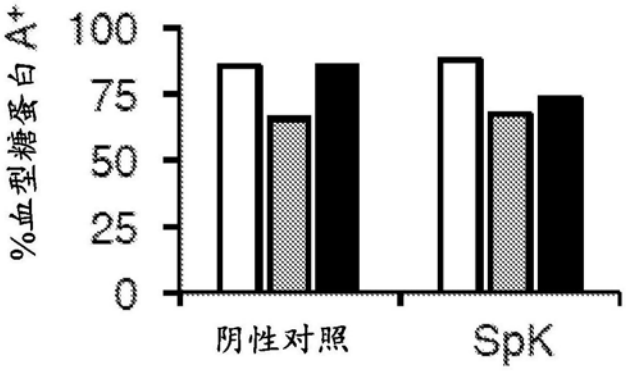


图14D

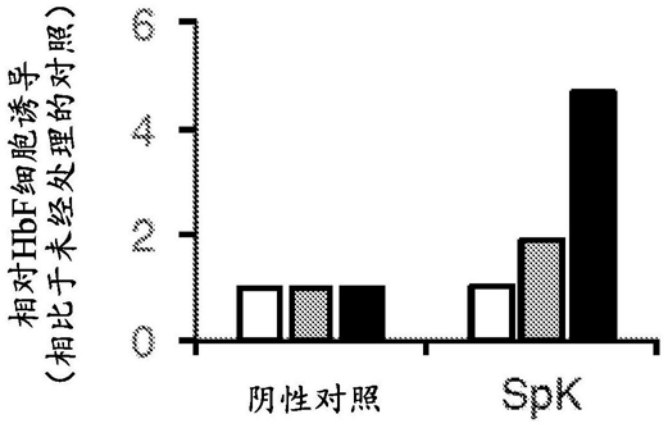


图14E

转让（Attorn）

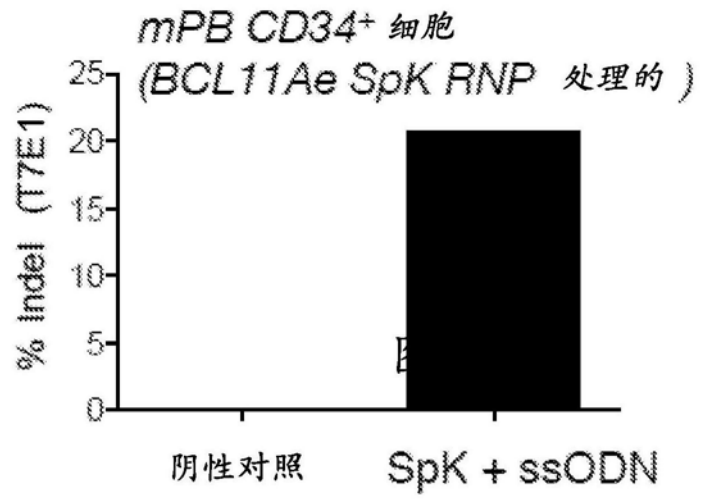


图15A

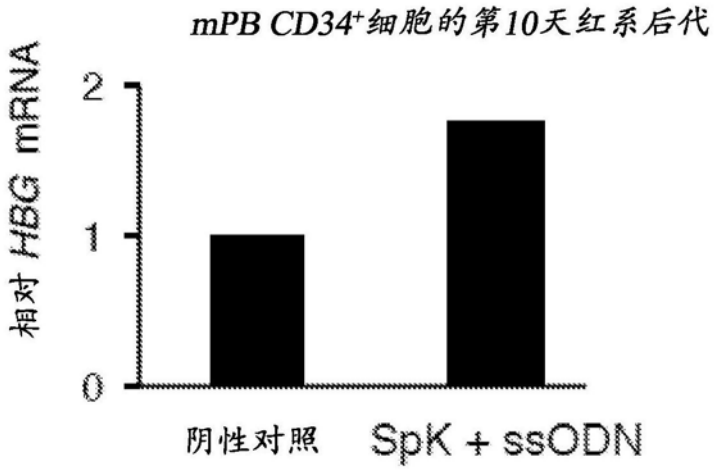


图15B

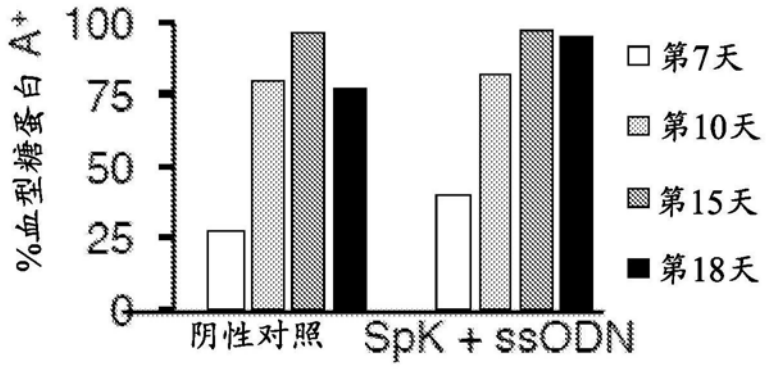

%血型糖蛋白 A+
100
75
50
25
0
阴性对照
SpK + ssODN
第7天
第10天
第15天
第18天

图15C

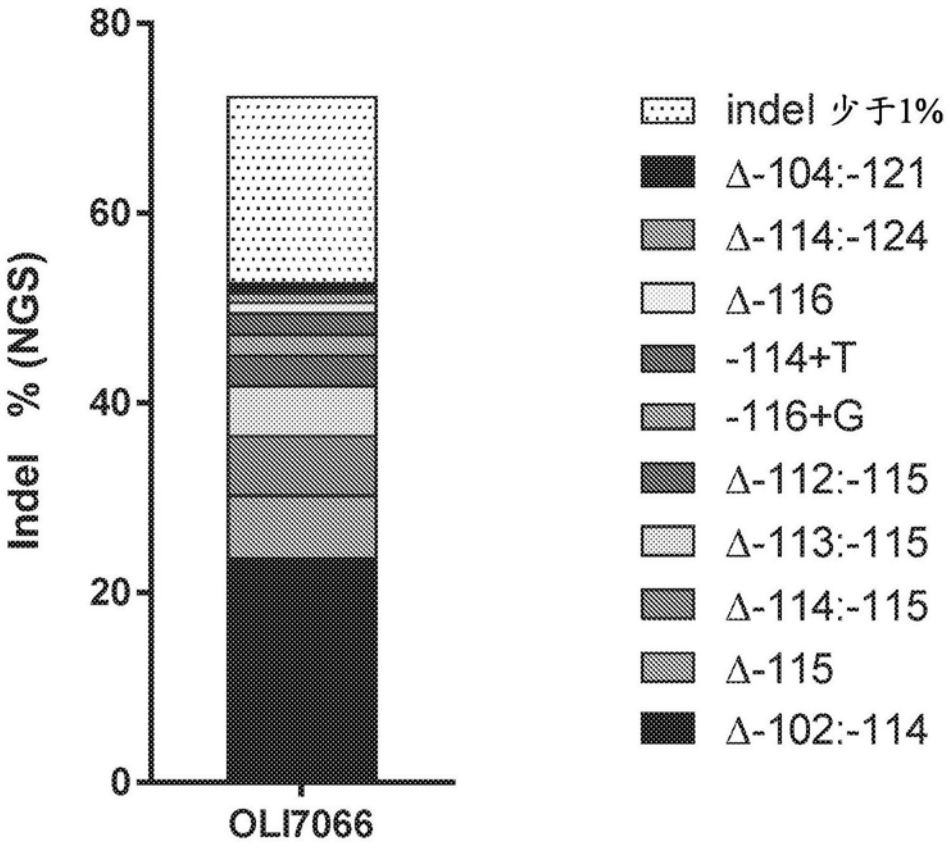

Indel % (NGS)
80
60
40
20
0
OLI7066
indel 少于1%
Δ-104:-121
Δ-114:-124
Δ-116
-114+T
-116+G
Δ-112:-115
Δ-113:-115
Δ-114:-115
Δ-115
Δ-102:-114

图16

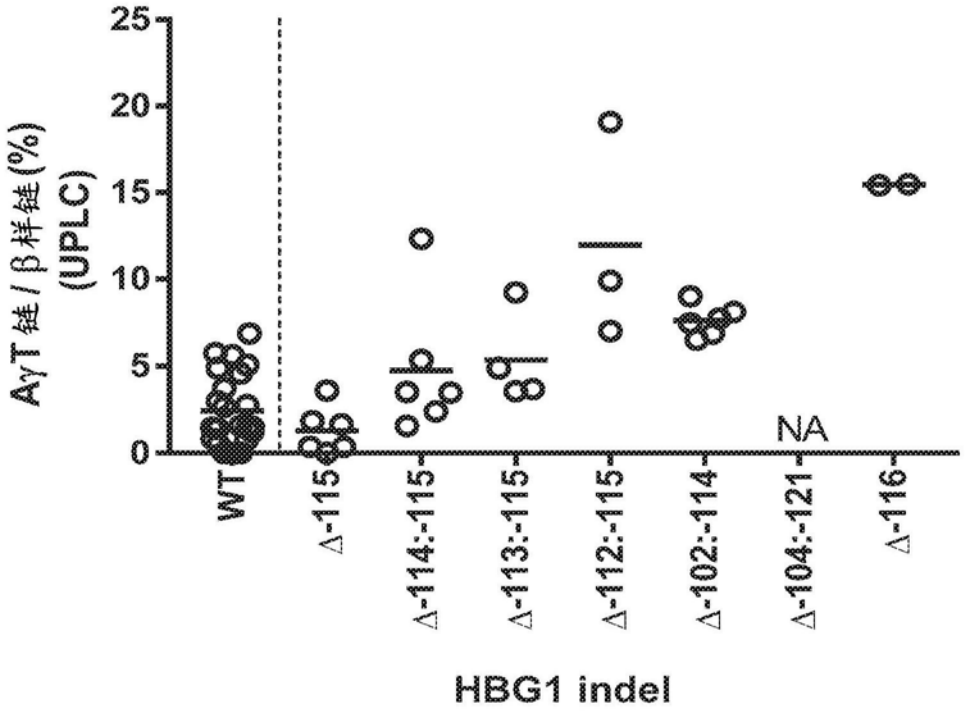

Aγ T 链 / β 样链 (%)
(UPLC)
25
20
15
10
5
0
NA
WT
Δ-115
Δ-114:-115
Δ-113:-115
Δ-112:-115
Δ-102:-114
Δ-104:-121
Δ-116
HBG1 indel

图17A

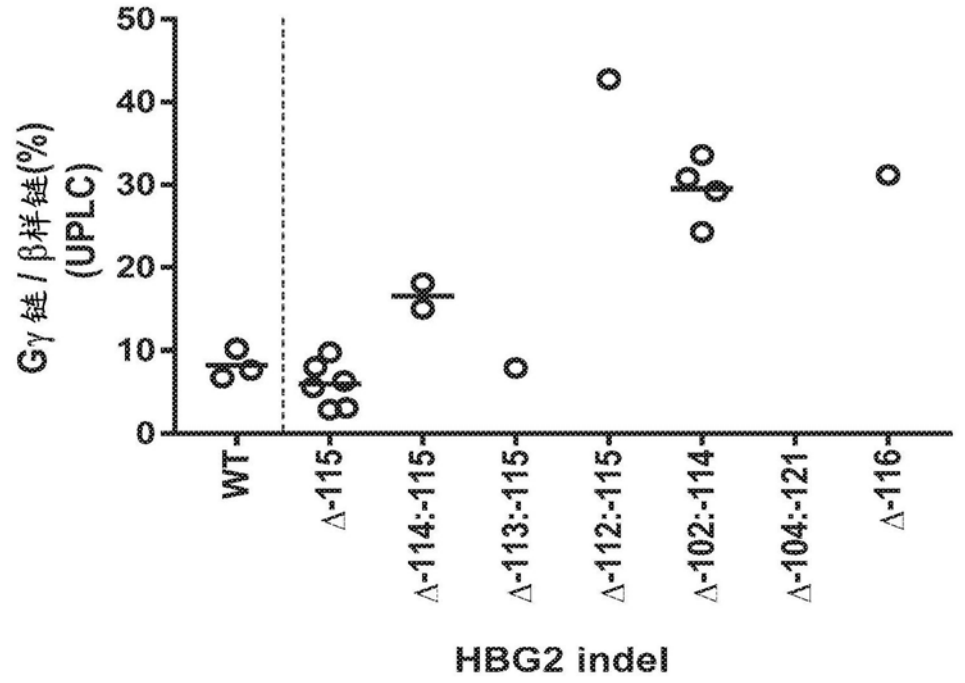

Gγ 链 / β 样链 (%)
(UPLC)
50
40
30
20
10
0
WT
Δ-115
Δ-114:-115
Δ-113:-115
Δ-112:-115
Δ-102:-114
Δ-104:-121
Δ-116
HBG2 indel

图17B

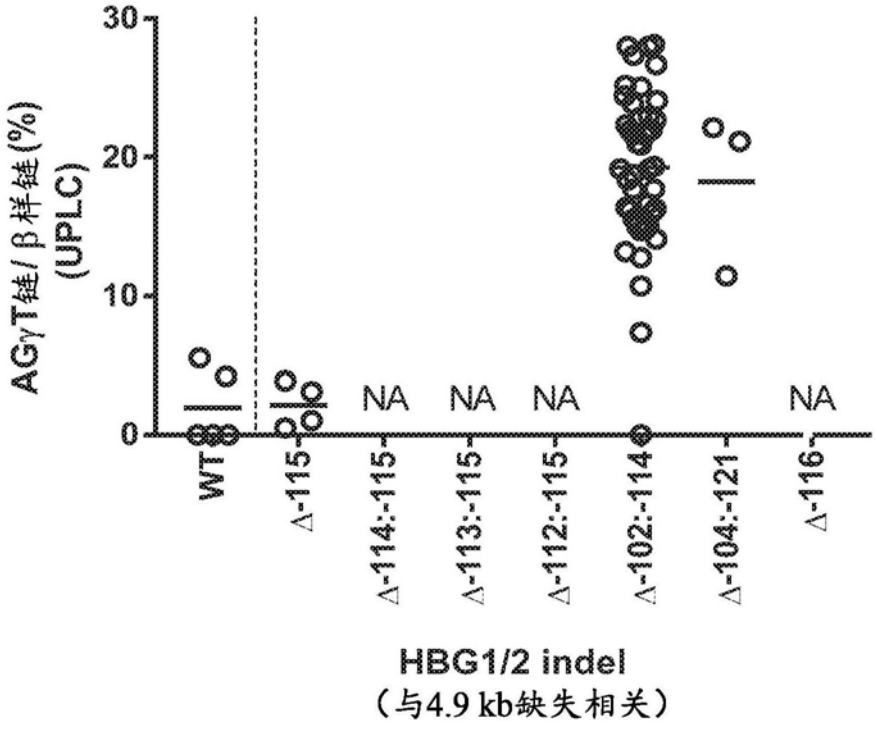

AGγT链/β样链(%)
(UPLC)
30
20
10
0
NA
NA
NA
NA
WT
Δ-115
Δ-114:-115
Δ-113:-115
Δ-112:-115
Δ-102:-114
Δ-104:-121
Δ-116
HBG1/2 indel
（与4.9 kb缺失相关）

图17C

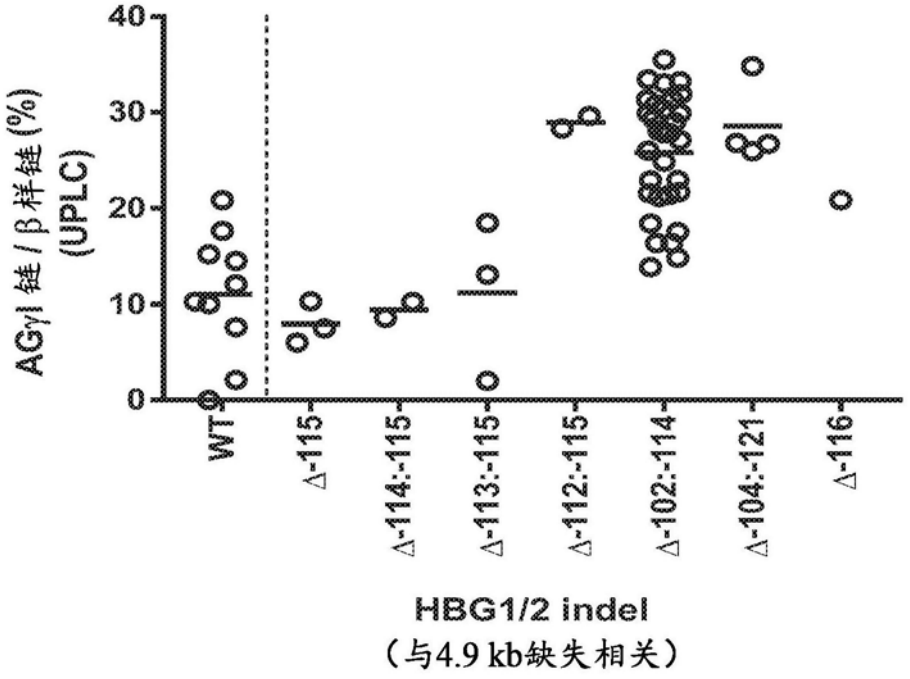

AGγI 链 / β样链 (%)
(UPLC)
40
30
20
10
0
WT
Δ-115
Δ-114:-115
Δ-113:-115
Δ-112:-115
Δ-102:-114
Δ-104:-121
Δ-116
HBG1/2 indel
（与4.9 kb缺失相关）

图17D

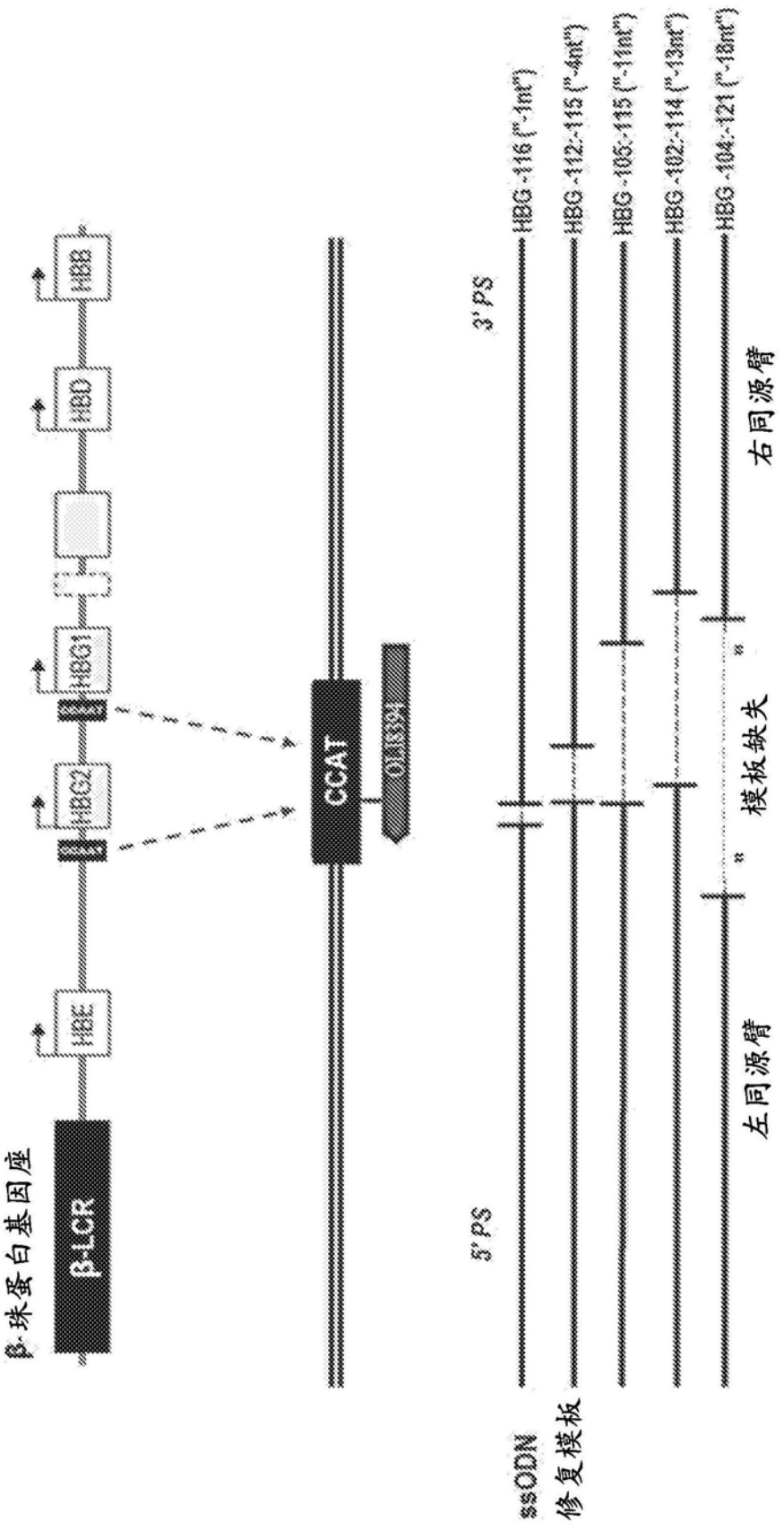

β-珠蛋白基因座
β-LCR
HBE
HBG2
HBG1
HBD
HBB
CCAT
OR3B3
5' PS
3' PS
ssODN
修复模板
HBG -116 ("-1nt")
HBG -112:-115 ("-4nt")
HBG -105:-115 ("-11nt")
HBG -102:-114 ("-13nt")
HBG -104:-121 ("-18nt")
左同源臂
模板缺失
右同源臂

图18

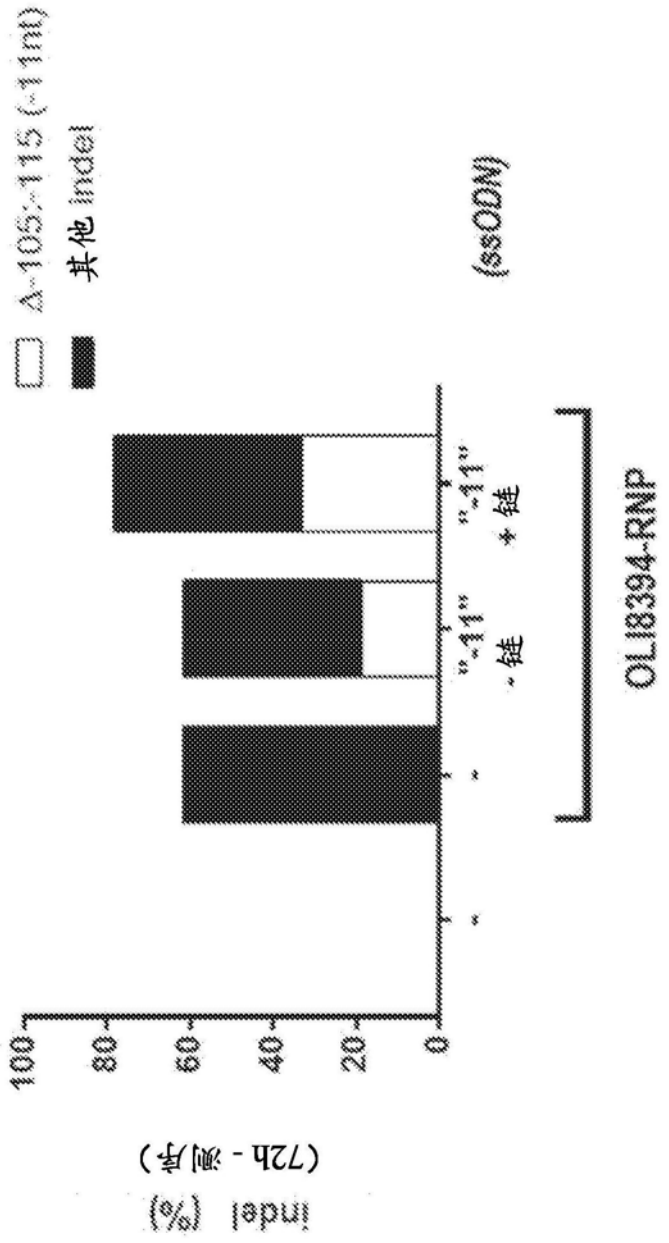

Δ-105:-115 (-11nt)
其他 indel
(ssODN)
"-11" + 链
"-11" - 链
OLI8394-RNP
100
80
60
40
20
0
indel (%)
(72h - 测序)

图19A

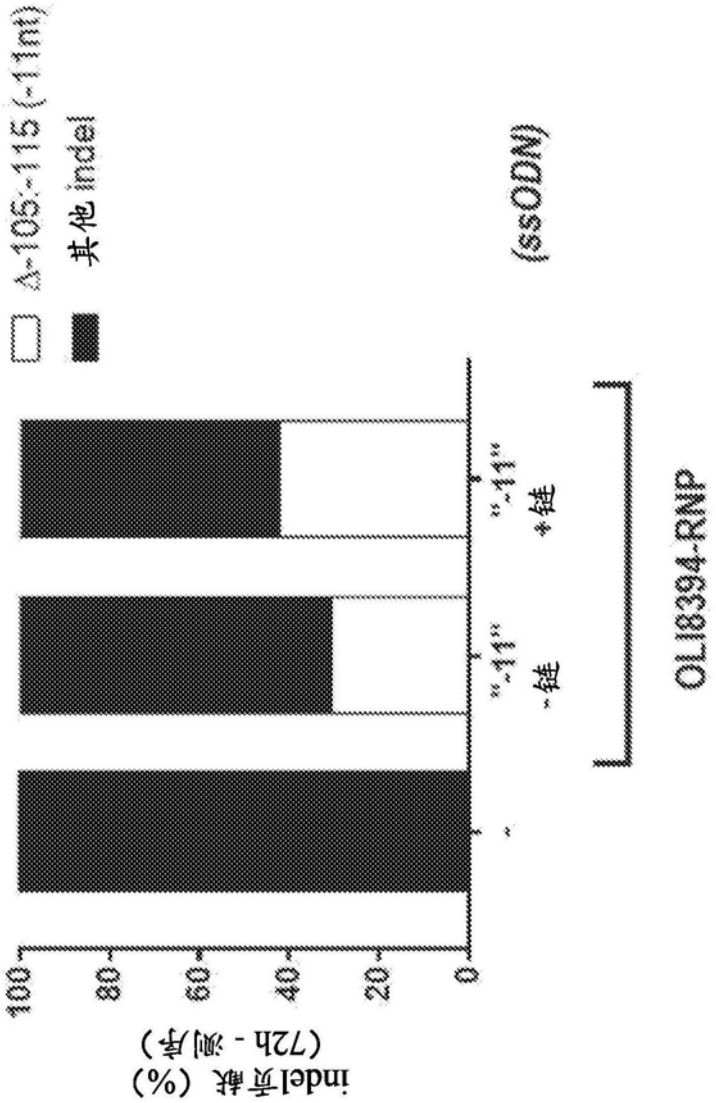

Δ-105:-115 (-11nt)
其他 indel
(ssODN)
"-11"
+链
"-11"
-链
OLI8394-RNP
100
80
60
40
20
0
indel贡献 (%)
(72h - 测序)

图19B

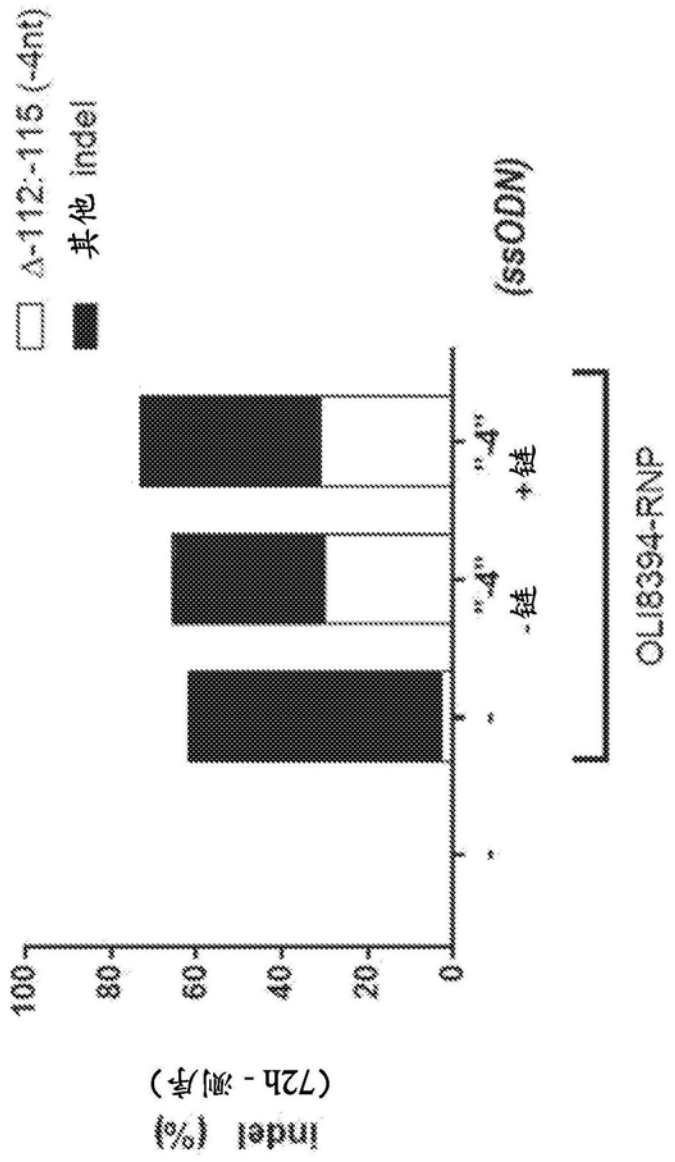

Δ-112:-115 (-4nt)
其他 indel
(ssODN)
"-4"
+链
"-4"
-链
OLI8394-RNP
100
80
60
40
20
0
indel (%)
(72h - 测序)

图19C

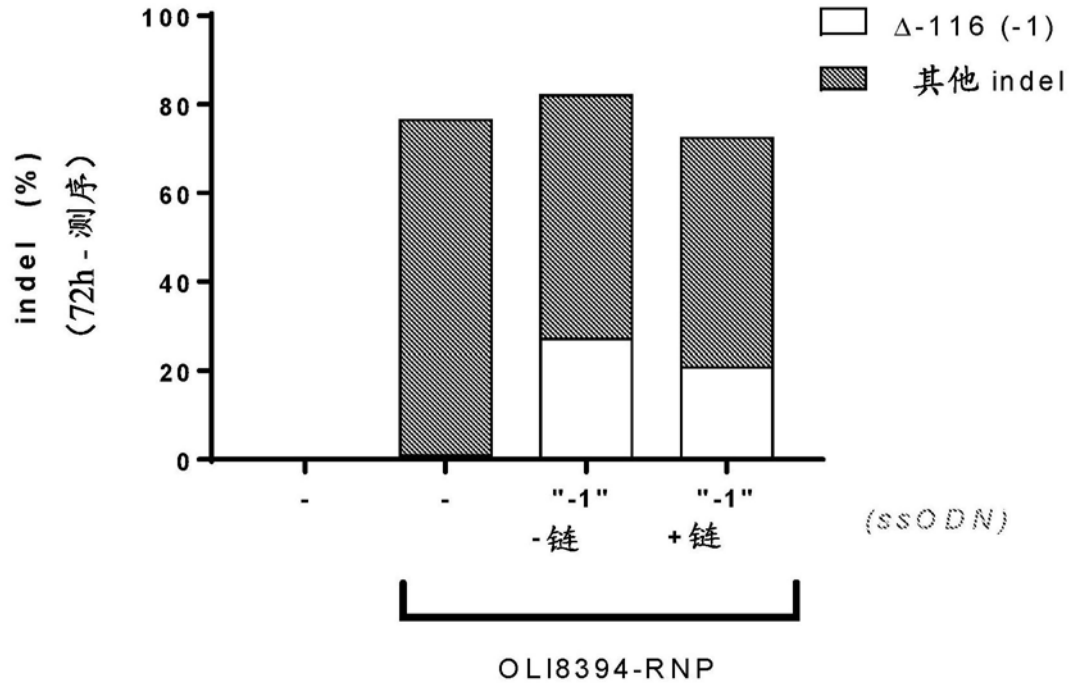

100
80
60
40
20
0
indel (%)
(72h-测序)
Δ-116 (-1)
其他 indel
-
-
"-1"
-链
"-1"
+链
(ssODN)
OLI8394-RNP

图19D

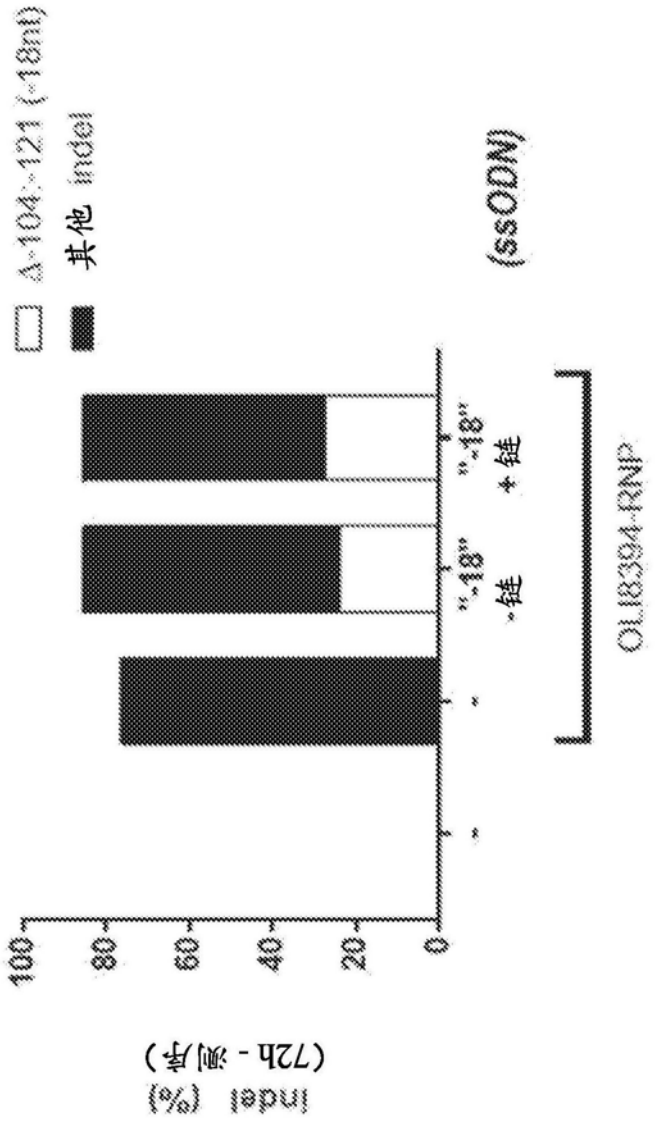

Δ-104:-121 (-18nt)
其他 indel
(ssODN)
"-18"
+链
"-18"
-链
OLI8394-RNP
100
80
60
40
20
0
indel (%)
(72h - 测序)

图19E

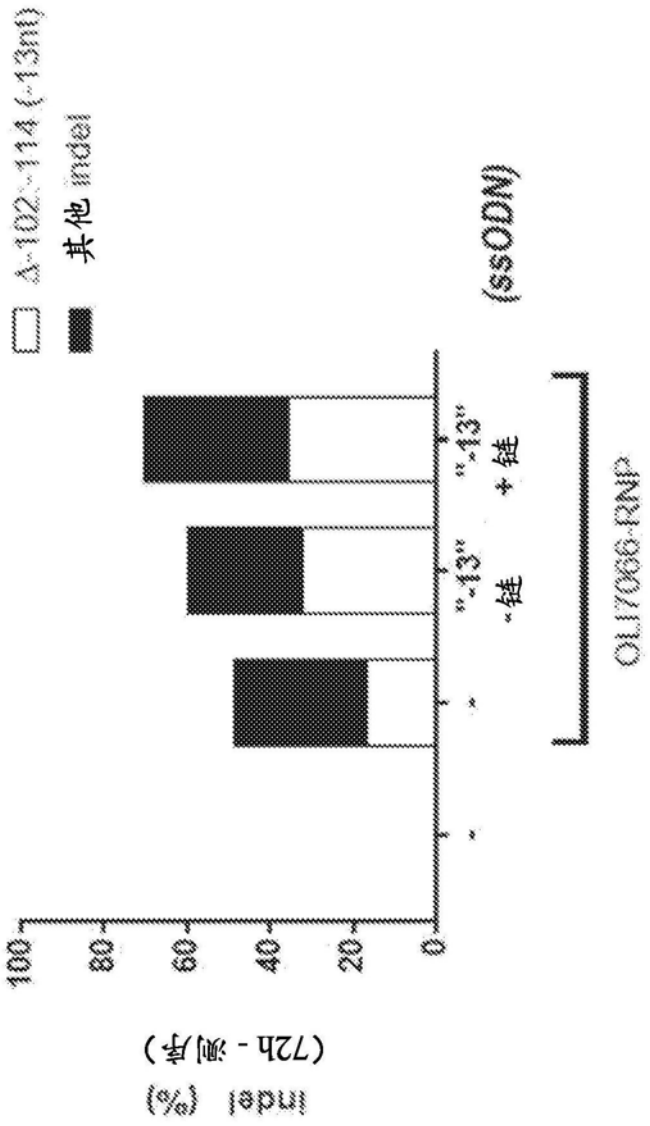

Δ-102~114 (-13nt)
其他 indel
(ssODN)
"-13" + 链
"-13" - 链
OLI7066-RNP
100
80
60
40
20
0
indel (%)
(72h - 测序)

图19F

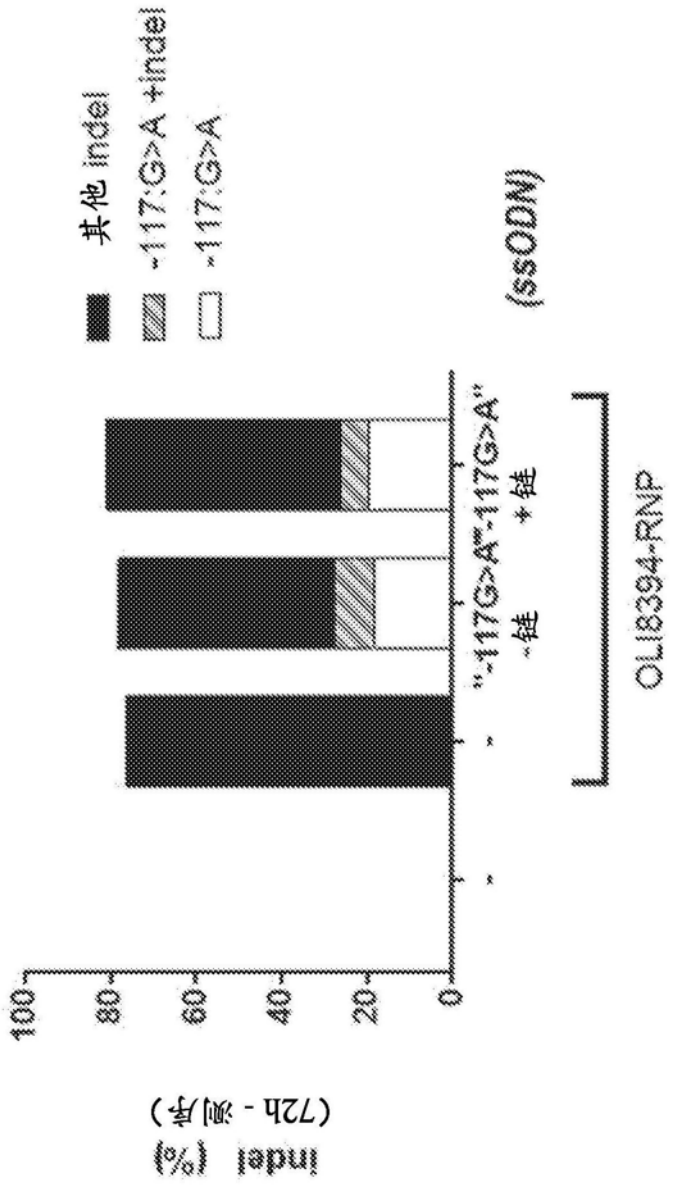

其他 indel
-117:G>A +indel
-117:G>A
(ssODN)
"-117G>A"-117G>A"
-链 +链
OLI8394-RNP
100
80
60
40
20
0
indel (%)
(72h - 测序)

图19G

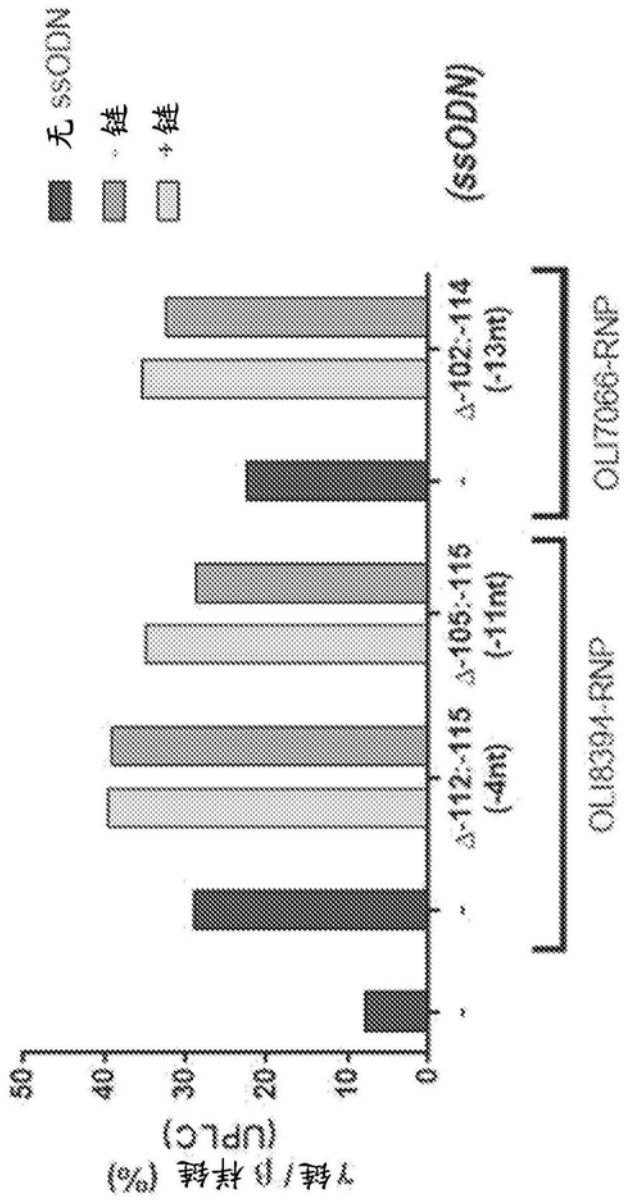

无 ssODN
- 链
+ 链
(ssODN)
Δ-102:-114 (-13nt)
-
Δ-105:-115 (-11nt)
Δ-112:-115 (-4nt)
-
-
OLI7066-RNP
OLI8304-RNP
50
40
30
20
10
0
γ链/β样链 (%)
(UPLC)

图20A

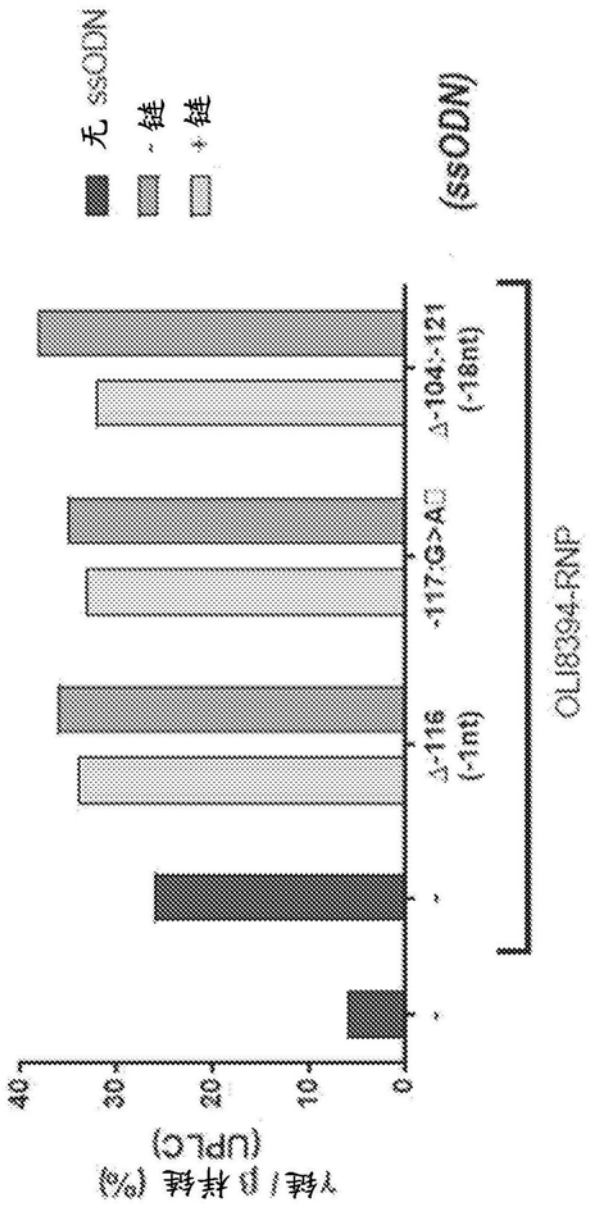

无 ssODN
- 链
+ 链
(ssODN)
Δ-104:-121
(-18nt)
-117:G>A
Δ-116
(-1nt)
-
-
OLI8394-RNP
40
30
20
10
0
γ链 / β 样链 (%)
(UPLC)

图20B

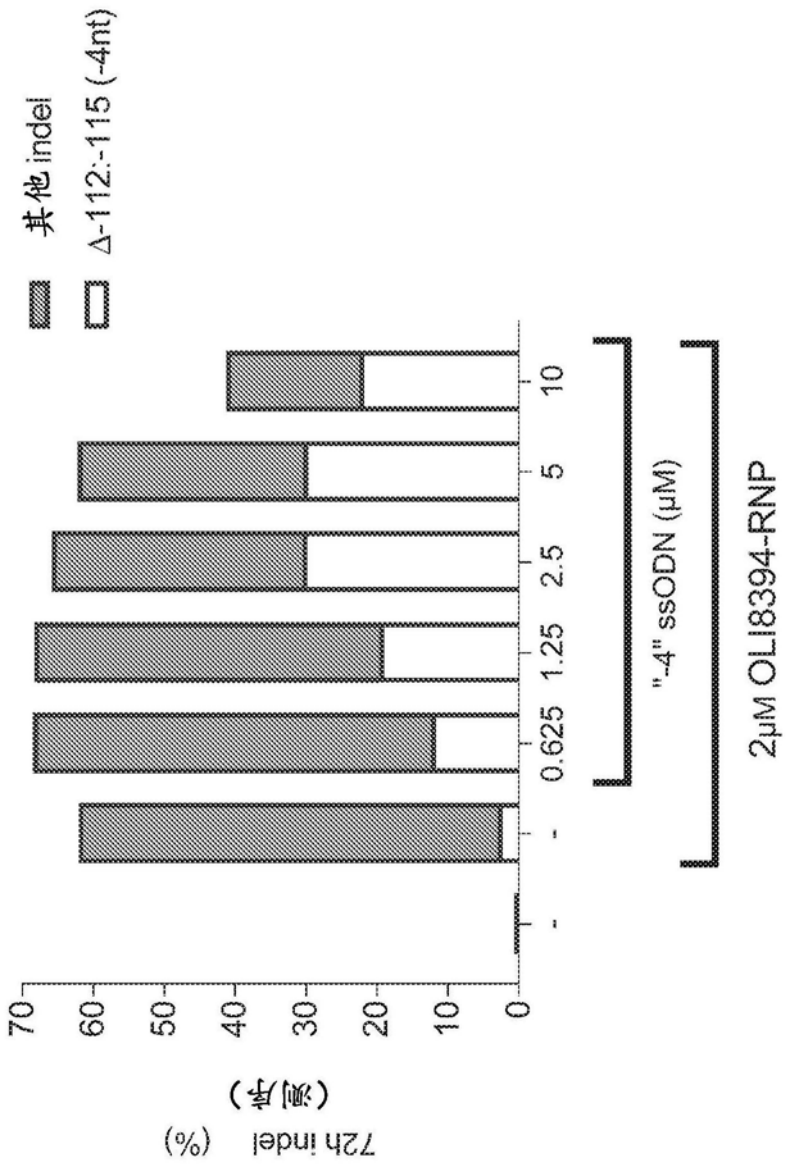

其他 indel
Δ-112:-115 (-4nt)
72h indel (%)
(测序)
70
60
50
40
30
20
10
0
-
-
0.625
1.25
2.5
5
10
"-4" ssODN (μM)
2μM OLI8394-RNP

图21A

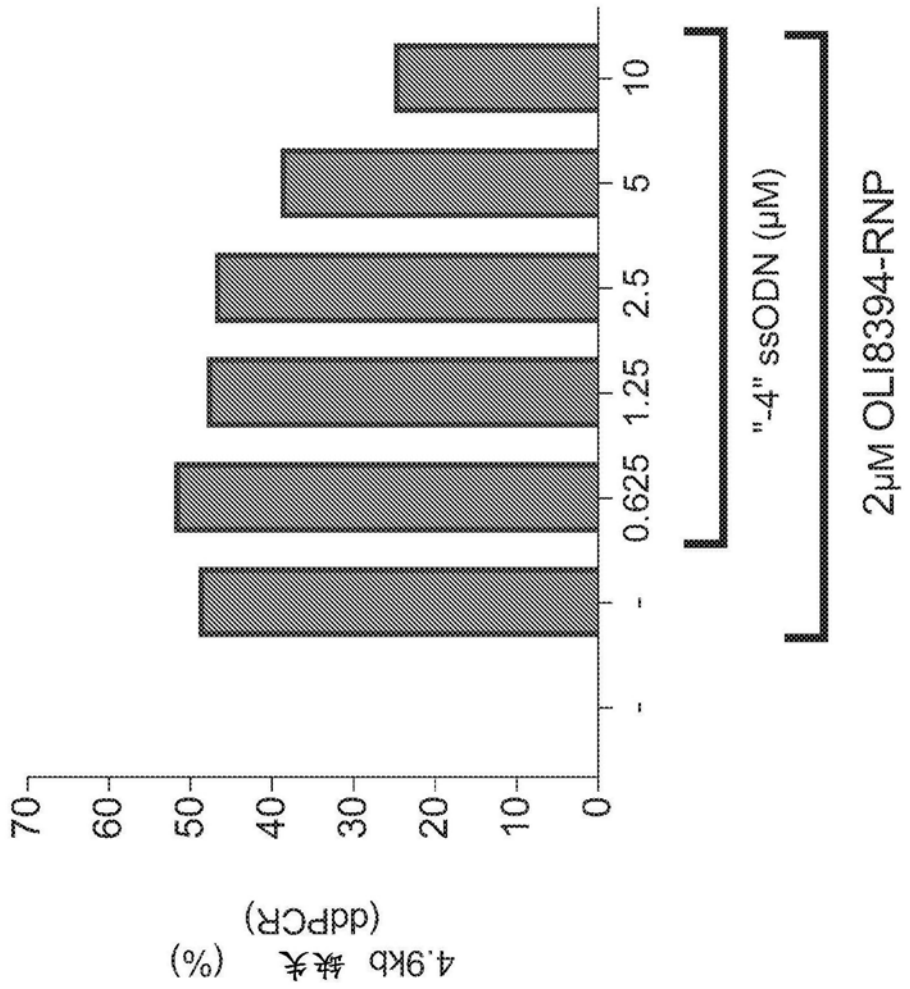

4.9kb 缺失 (%)
(ddPCR)
70
60
50
40
30
20
10
0
-
0.625
1.25
2.5
5
10
"-4" ssODN (μM)
2μM OLI8394-RNP

图21B

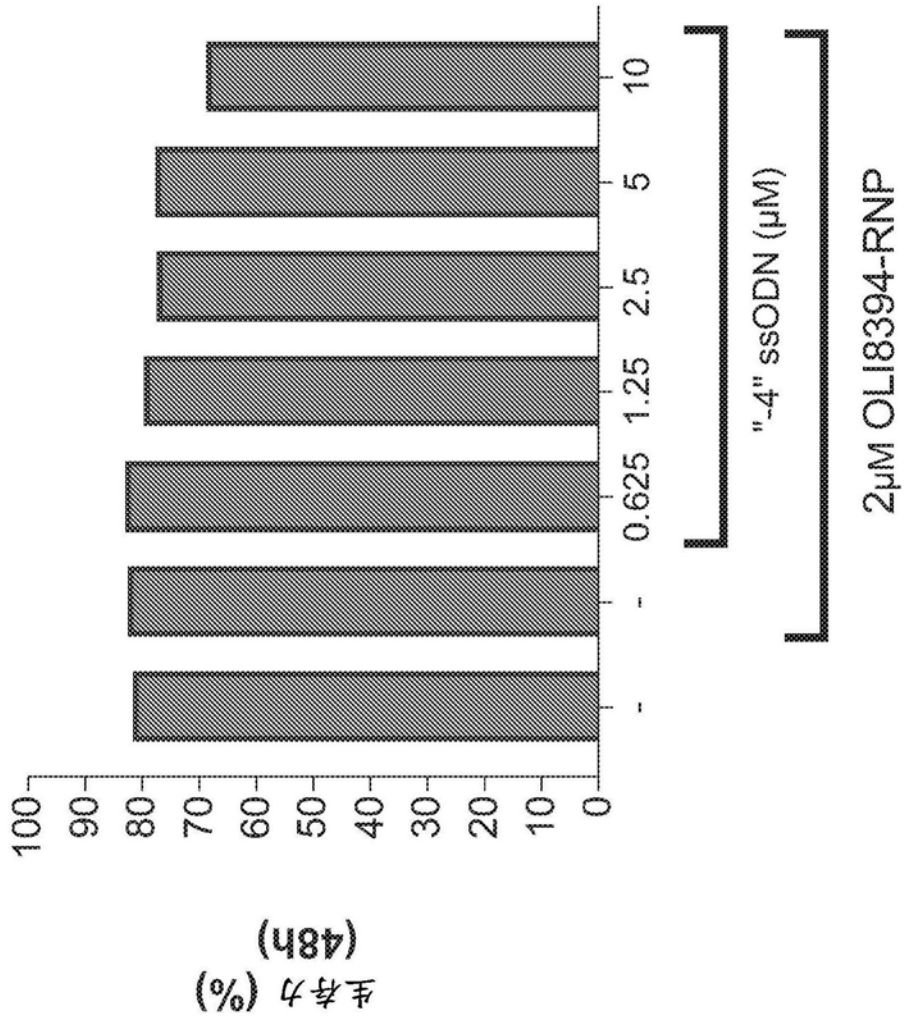

生存力 (%)
(48h)
100
90
80
70
60
50
40
30
20
10
0
-
0.625
1.25
2.5
5
10
"-4" ssODN (μM)
2μM OLI8394-RNP

图21C

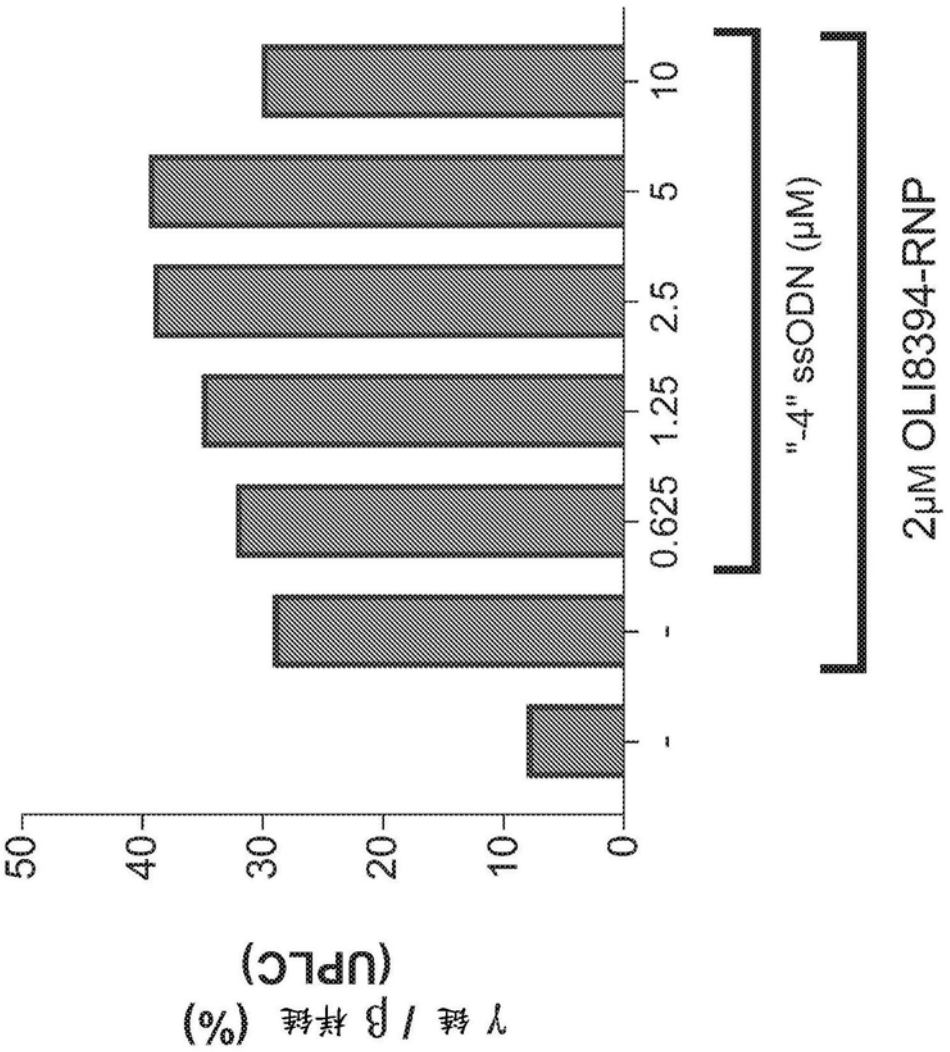

γ 链 / β 样链 (%)
(UPLC)
50
40
30
20
10
0
-
0.625
1.25
2.5
5
10
"-4" ssODN (μM)
2μM OLI8394-RNP

图21D

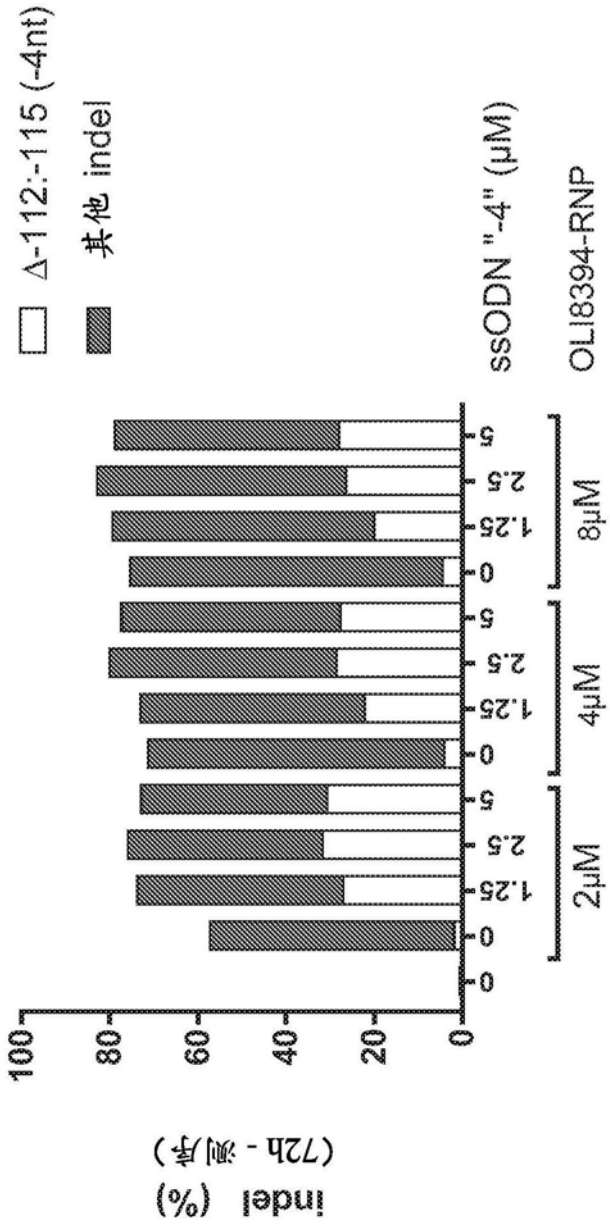

Δ-112:-115 (-4nt)
其他 indel
ssODN "-4" (μM)
OLI8394-RNP
indel (%)
(72h - 测序)
100
80
60
40
20
0
0
0
1.25
2.5
5
2μM
0
1.25
2.5
5
4μM
0
1.25
2.5
5
8μM

图22A

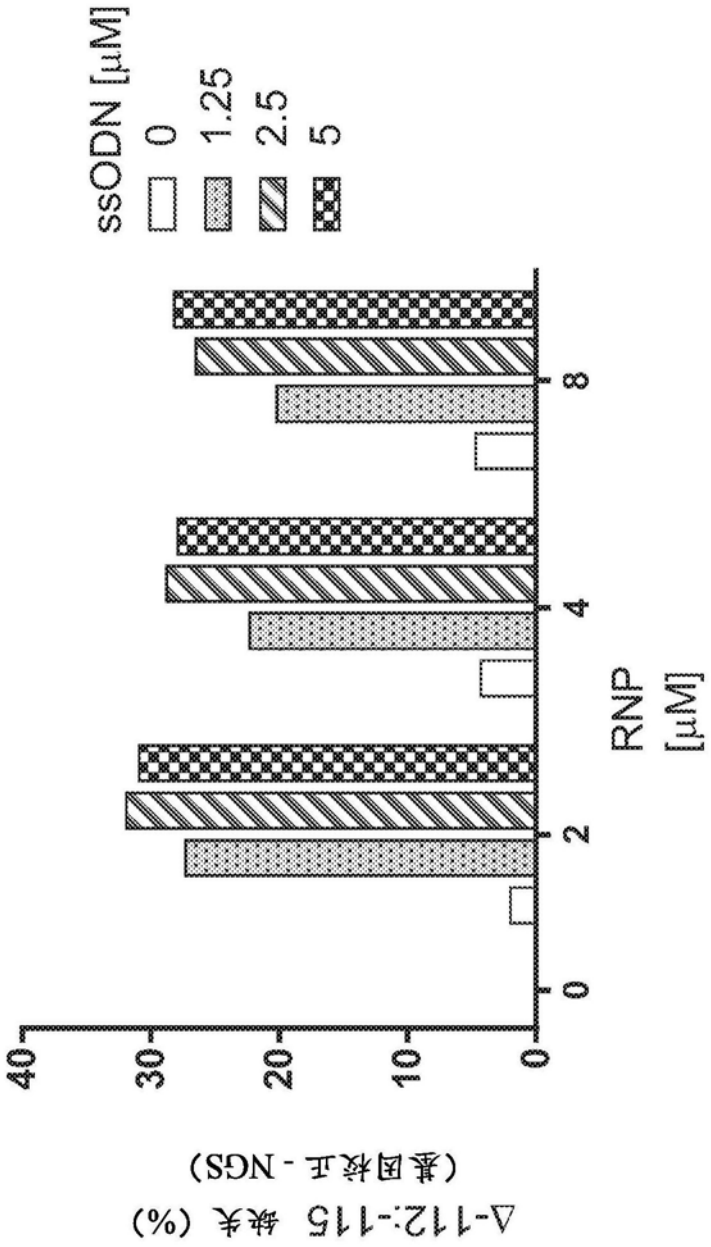


图22B

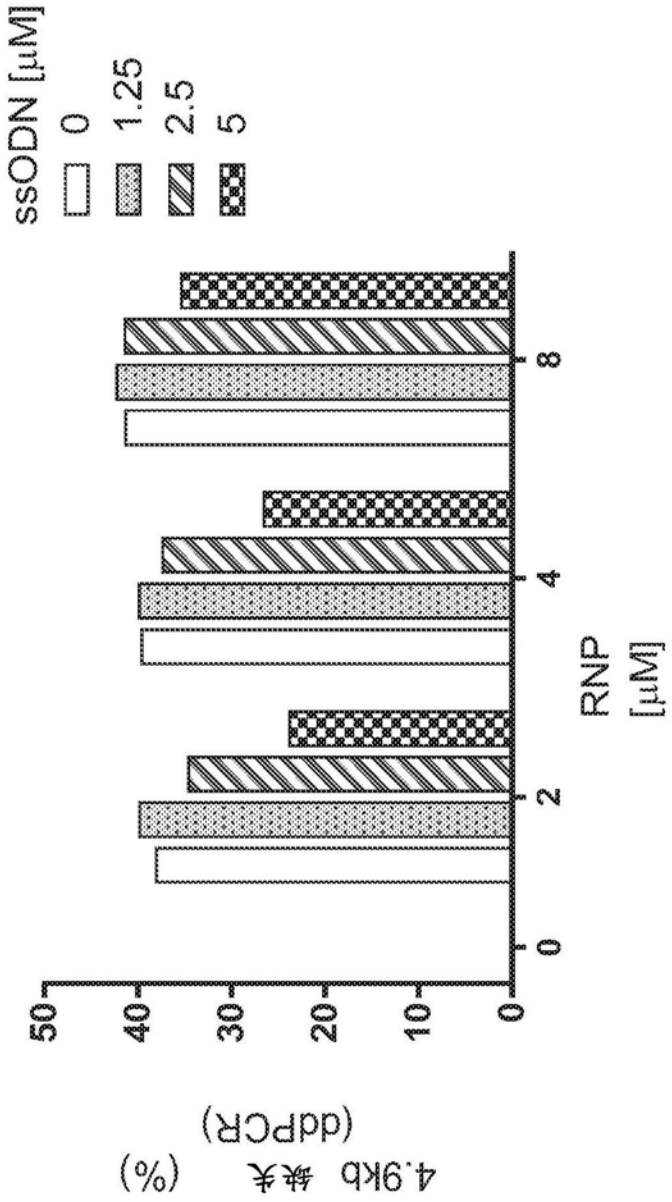


图22C

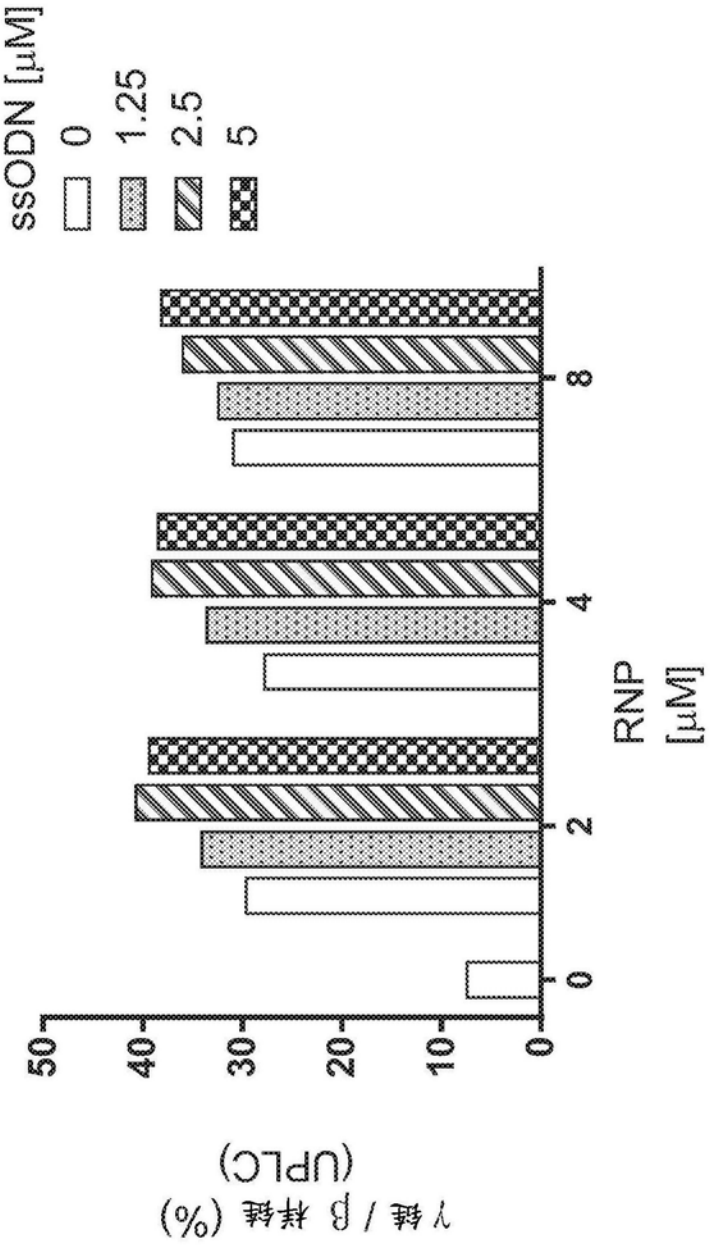


图22D

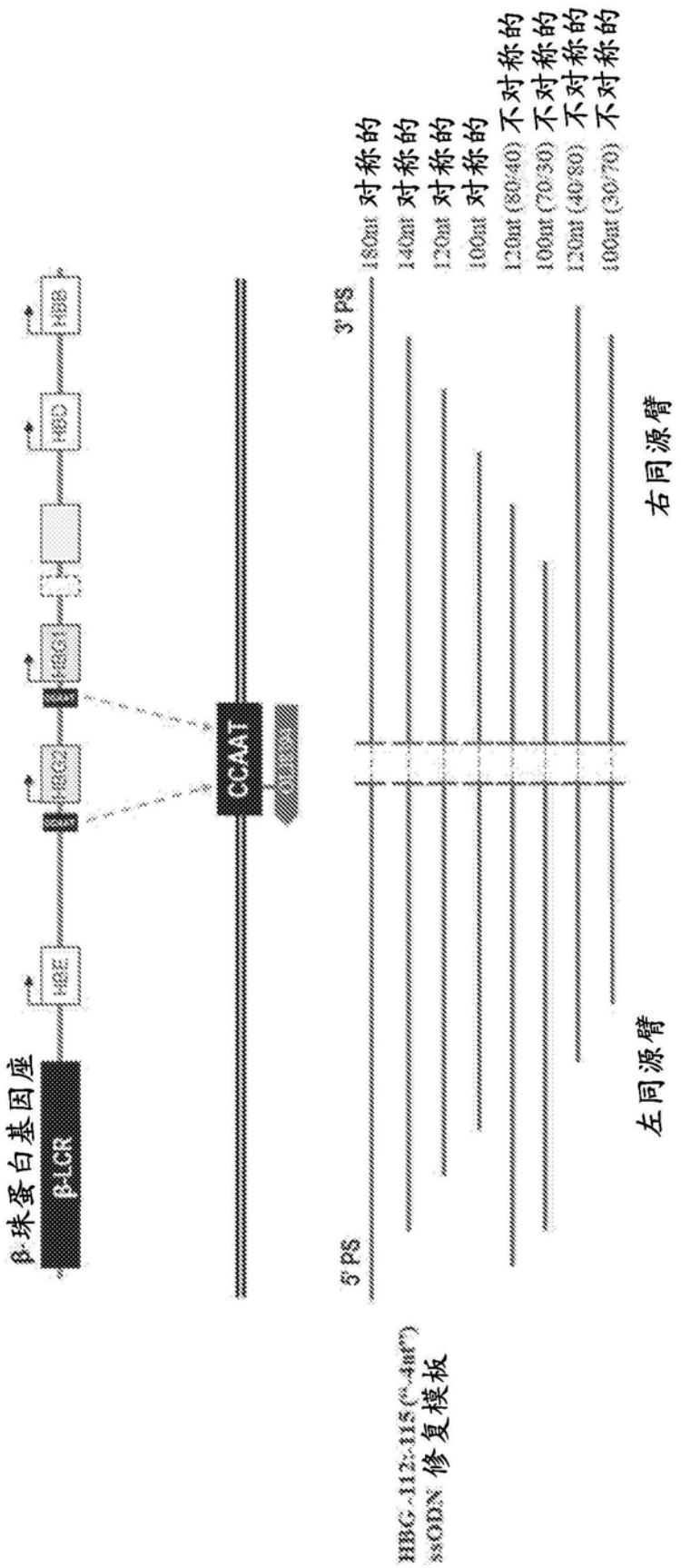

β-珠蛋白基因座
HBE
HBG2
HBG1
HBD
HBB
CCAAT
5' PS
3' PS
180nt 对称的
140nt 对称的
120nt 对称的
100nt 对称的
120nt (80:40) 不对称的
100nt (70:30) 不对称的
120nt (40:80) 不对称的
100nt (30:70) 不对称的
左同源臂
右同源臂
HBG -112:-115("-4nt")
ssODN 修复模板

图23A

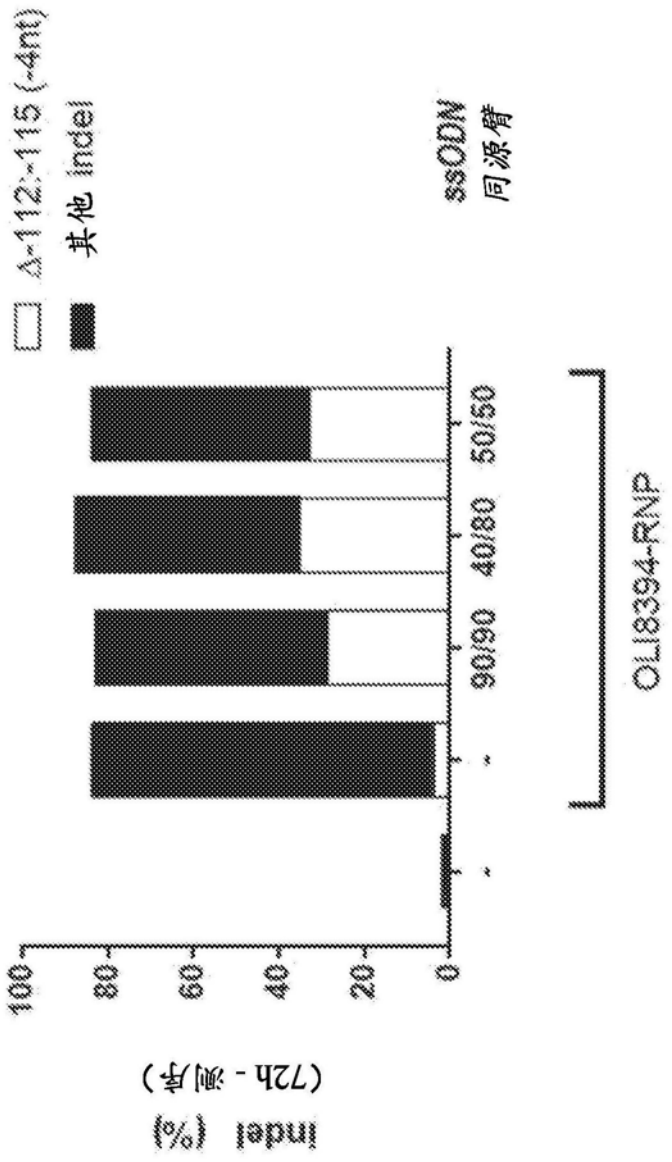

Δ-112:-115 (-4nt)
其他 indel
ssODN
同源臂
50/50
40/80
90/90
OLI8394-RNP
100
80
60
40
20
0
Indel (%)
(72h - 测序)

图23B

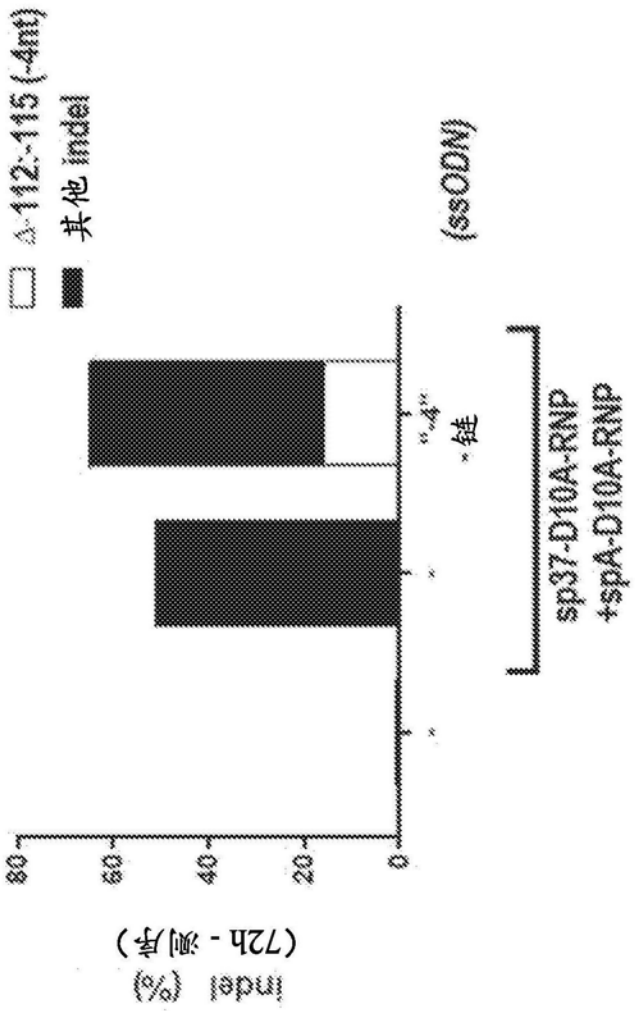

△-112:-115 (-4nt)
其他 indel
(ssODN)
"-"
+
sp37-D10A-RNP
+spA-D10A-RNP
80
60
40
20
0
indel (%)
(72h - 测序)

图24

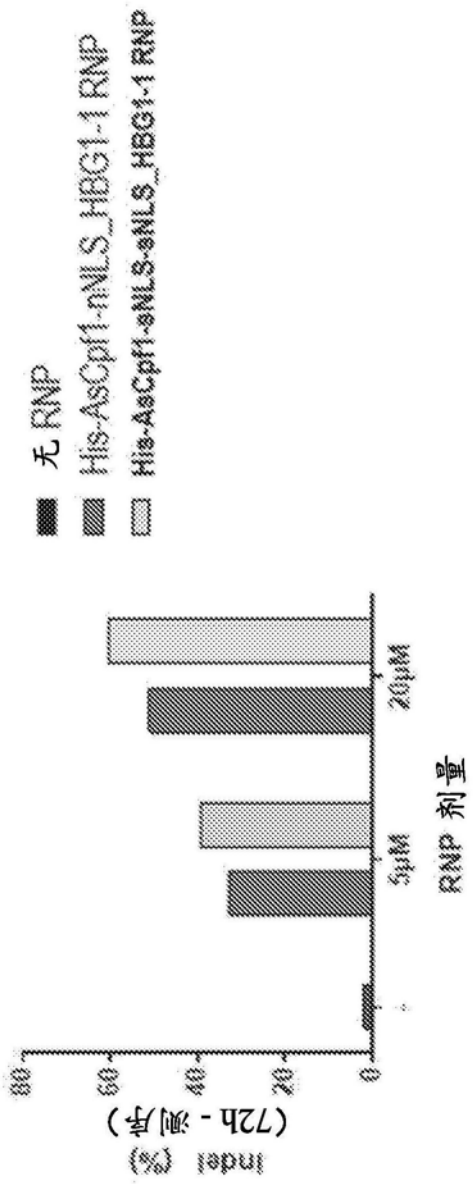

无 RNP
His-AsCpf1-nNLS_HBG1-1 RNP
His-AsCpf1-sNLS-sNLS_HBG1-1 RNP
indel (%)
(72h - 测序)
0
20
40
60
80
5μM
20μM
RNP 剂量

图25

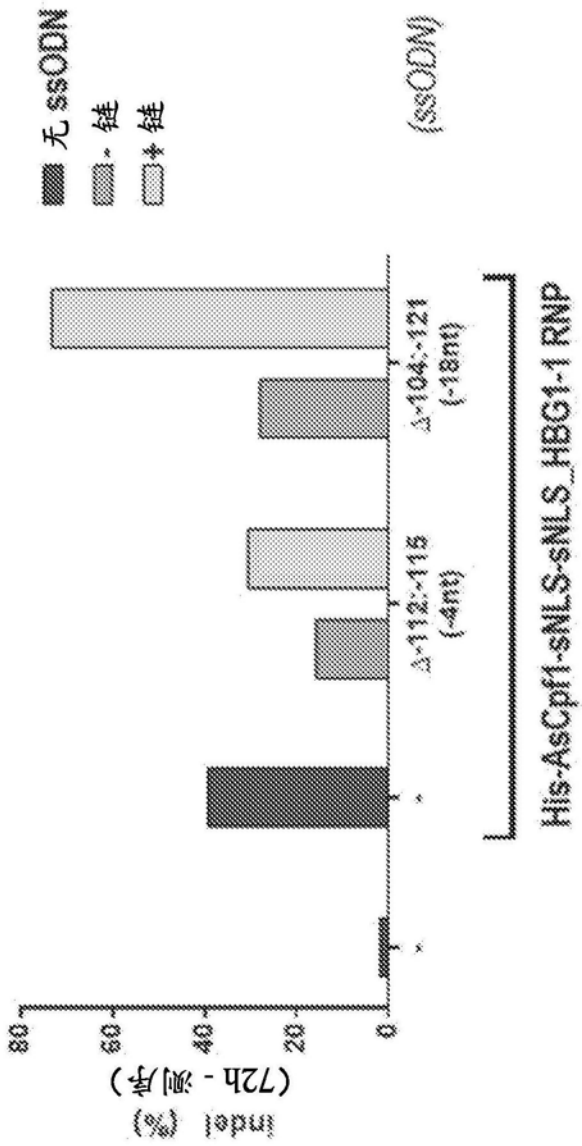

无 ssODN
- 链
+ 链
(ssODN)
Δ-104,-121
(-18nt)
Δ-112,-115
(-4nt)
His-AsCpf1-sNLS-sNLS_HBG1-1 RNP
0
20
40
60
80
indel (%)
(72h - 测序)

图26A

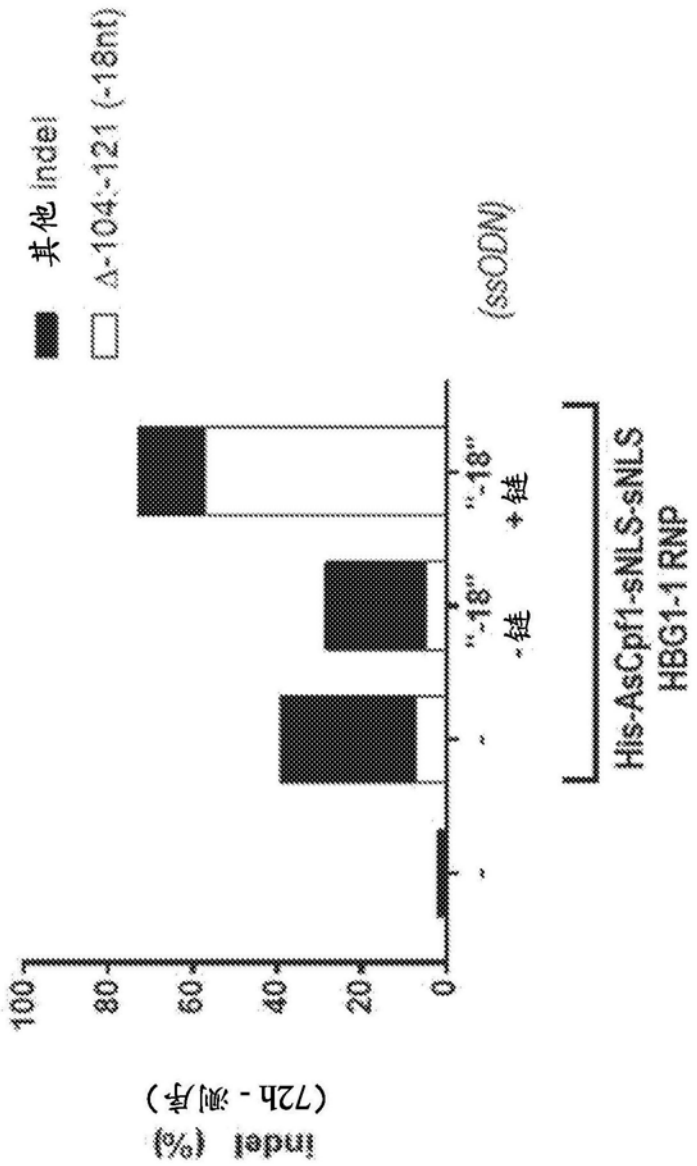

其他 indel
Δ-104,-121 (-18nt)
(ssODN)
"-18"
+链
"-18"
-链
His-AsCpf1-sNLS-sNLS
HBG1-1 RNP
100
80
60
40
20
0
indel (%)
(72h - 测序)

图26B

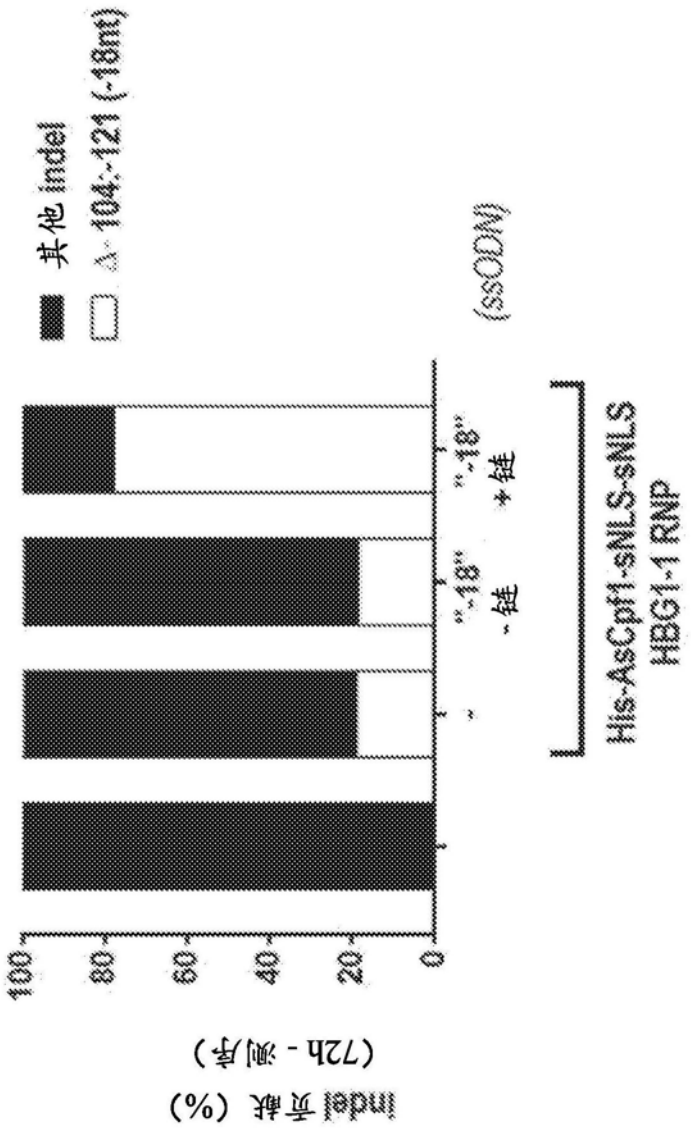

其他 indel
Δ -104,-121 (-18nt)
(ssODN)
"-18"
+链
"-18"
-链
-
His-AsCpf1-sNLS-sNLS
HBG1-1 RNP
100
80
60
40
20
0
indel 贡献 (%)
(72h - 测序)

图26C

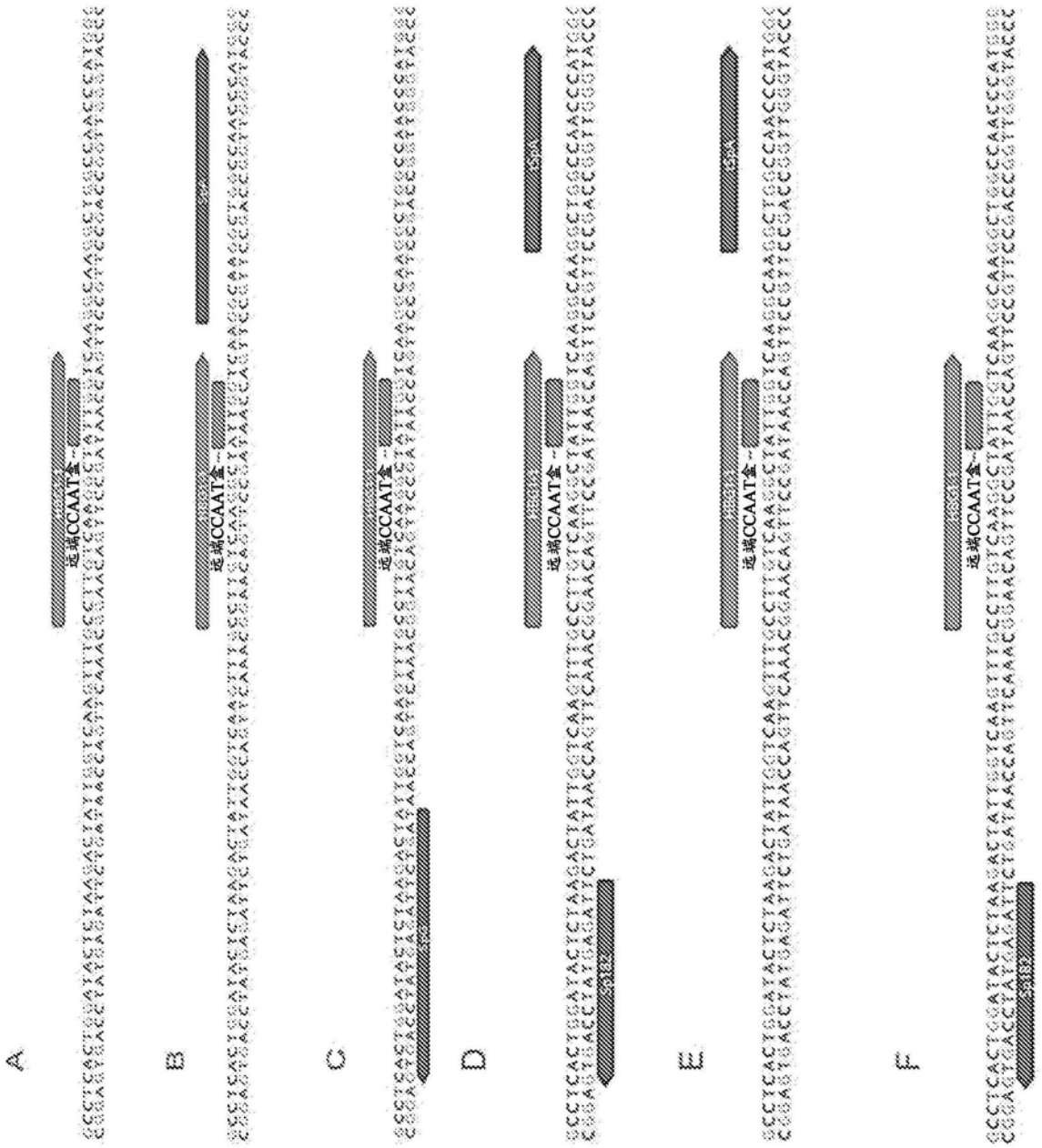


图27A

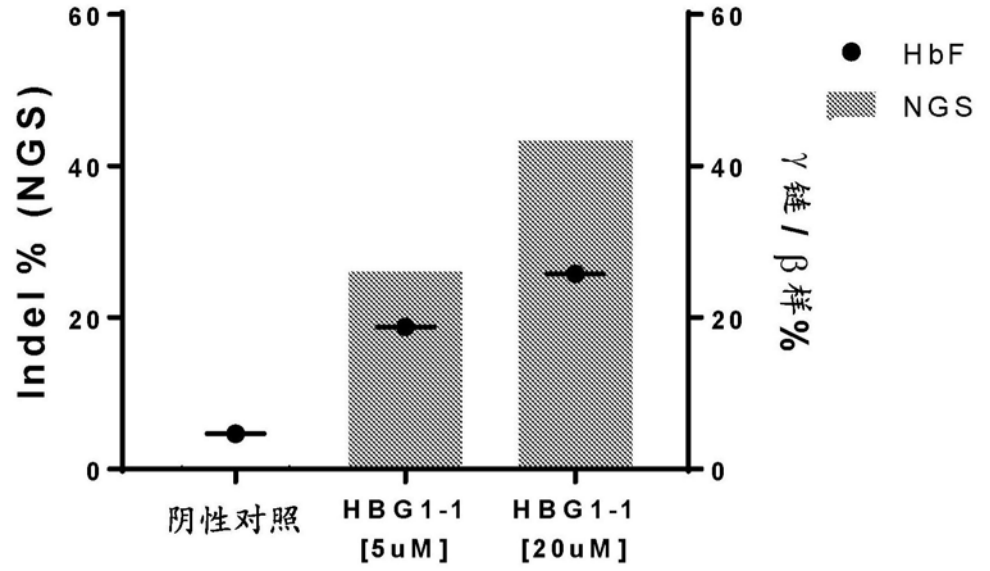

60
40
20
0
Indel % (NGS)
60
40
20
0
γ 链 / β 样 %
HbF
NGS
阴性对照
HBG1-1 [5uM]
HBG1-1 [20uM]

图28A

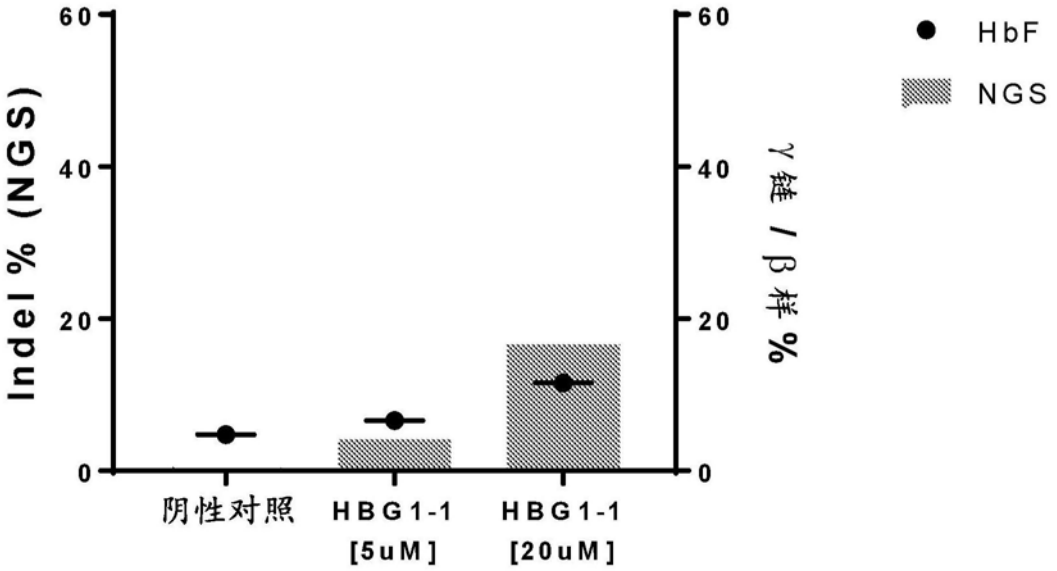

60
40
20
0
Indel % (NGS)
60
40
20
0
γ 链 / β 样 %
HbF
NGS
阴性对照
HBG1-1 [5uM]
HBG1-1 [20uM]

图28B

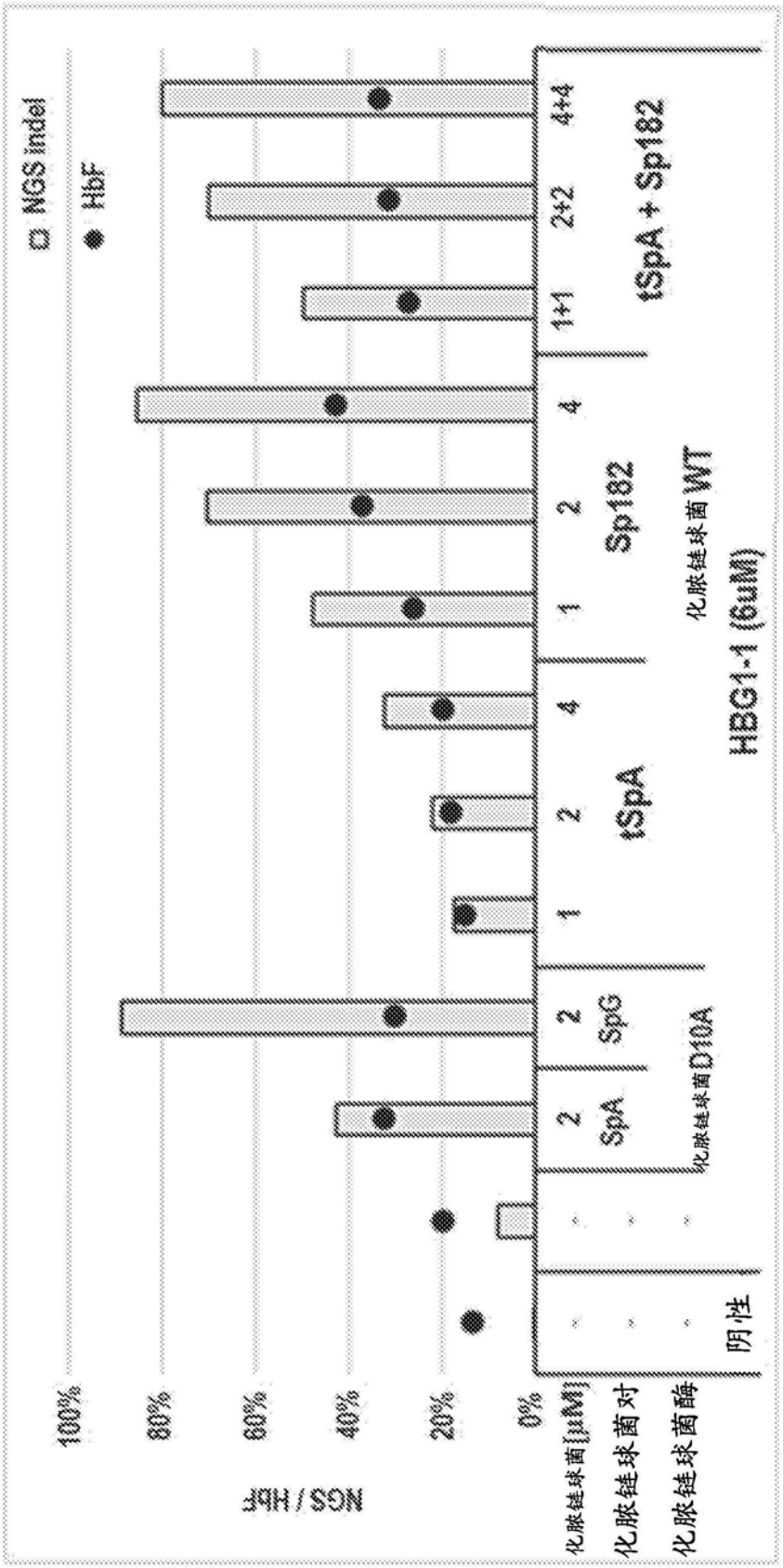

NGS indel
HbF
NGS / HbF
100%
80%
60%
40%
20%
0%
化脓链球菌[μM]
化脓链球菌对
化脓链球菌酶
阴性
2
SpA
2
SpG
化脓链球菌D10A
1
2
4
tSpA
1
2
4
Sp182
化脓链球菌 WT
HBG1-1 (6uM)
1+1
2+2
4+4
tSpA + Sp182

图29

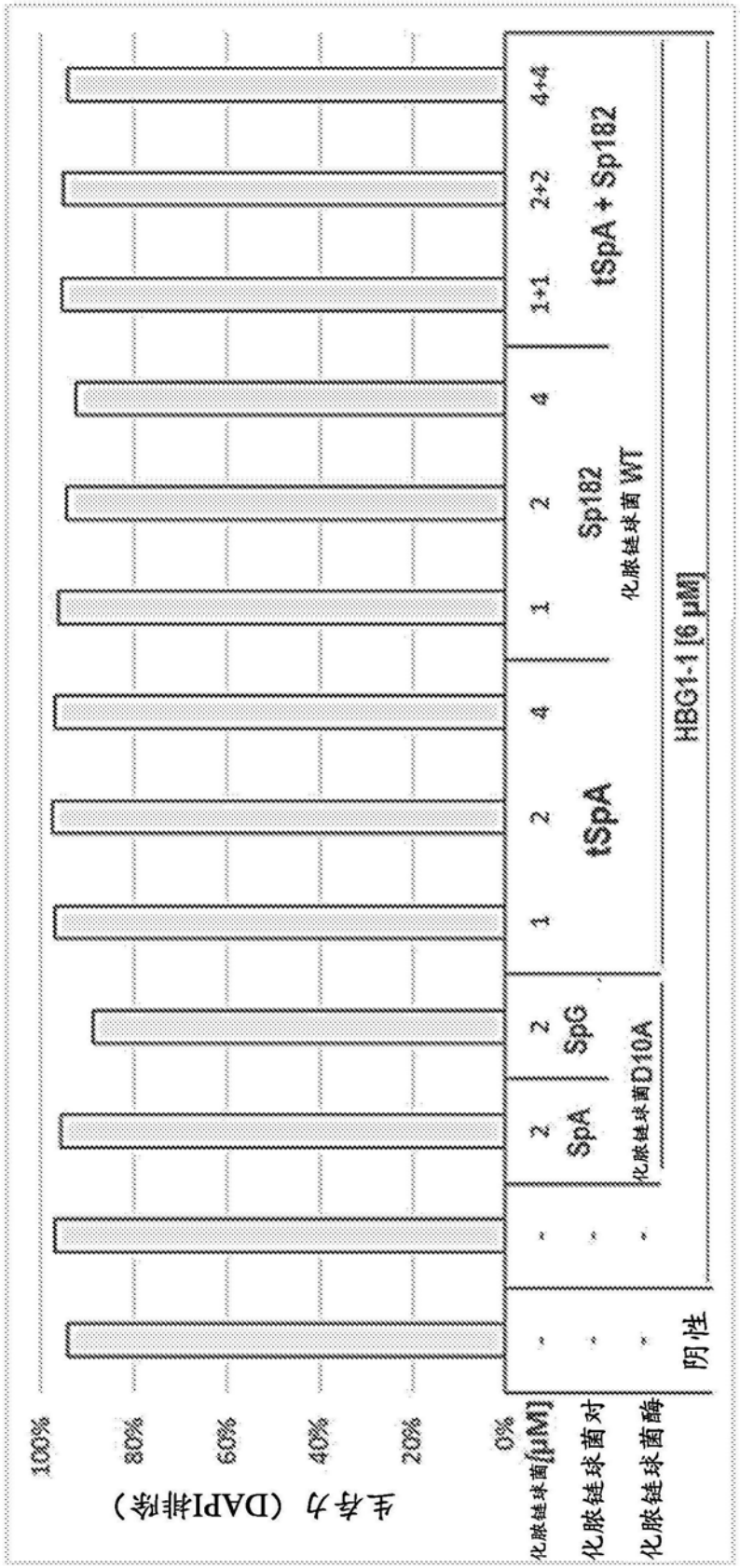

生存力（DAPI排除）
100%
80%
60%
40%
20%
0%
化脓链球菌[μM]
化脓链球菌对
化脓链球菌酶
-
-
-
阴性
-
-
-
2
SpA
2
SpG
1
2
4
tSpA
1
2
4
Sp182
化脓链球菌WT
1+1
2+2
4+4
tSpA + Sp182
HBG1-1 [6 μM]
化脓链球菌D10A

图30

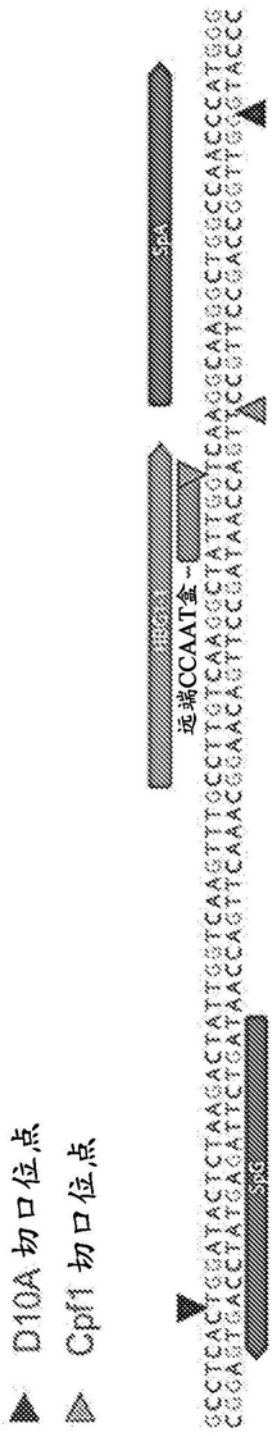


图31A

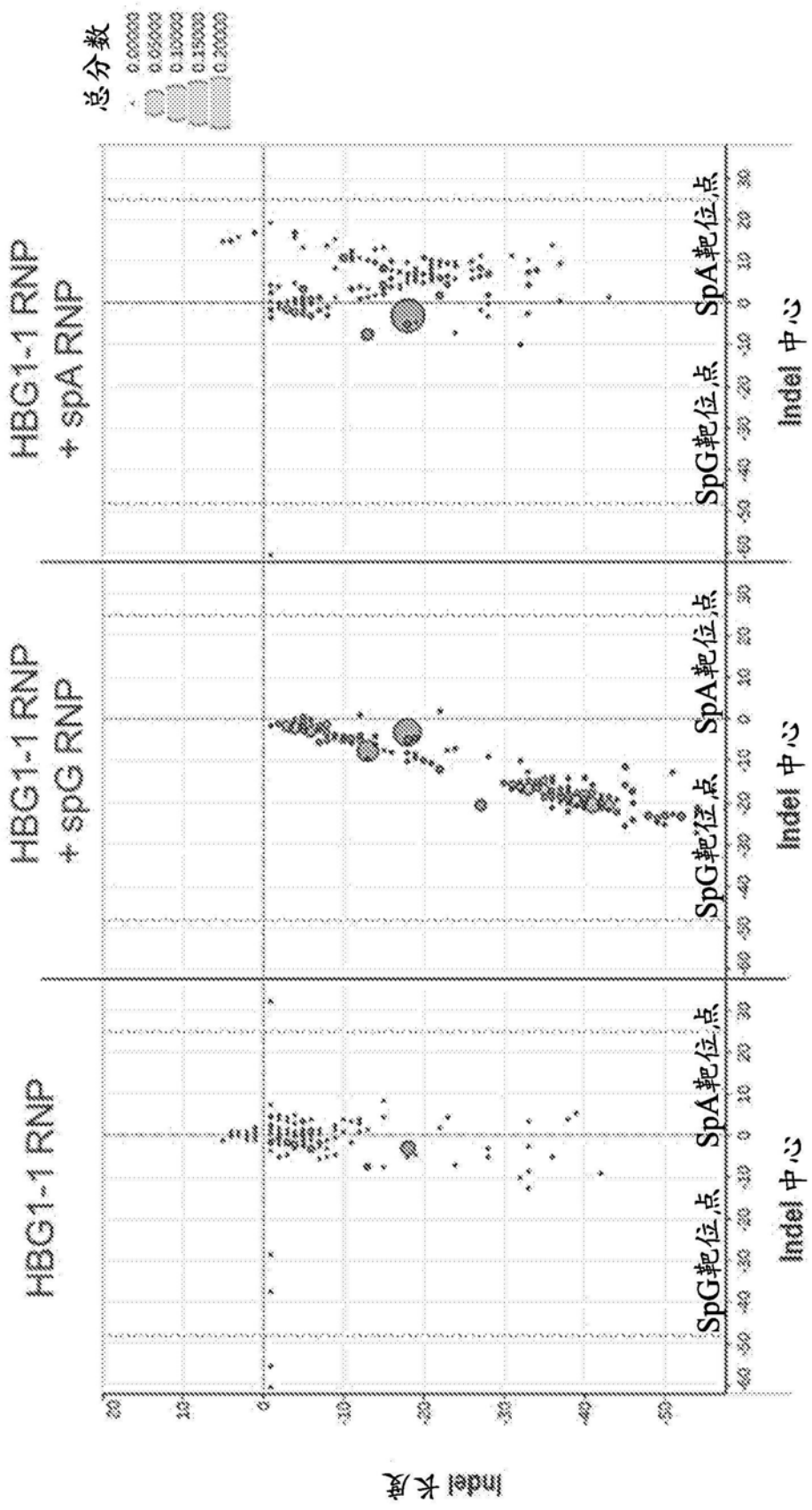

总分数
0.00000
0.05000
0.10000
0.15000
0.20000
HBG1-1 RNP
HBG1-1 RNP + spG RNP
HBG1-1 RNP + spA RNP
SpG靶位点
SpA靶位点
Indel 中心
Indel 长度

图31B

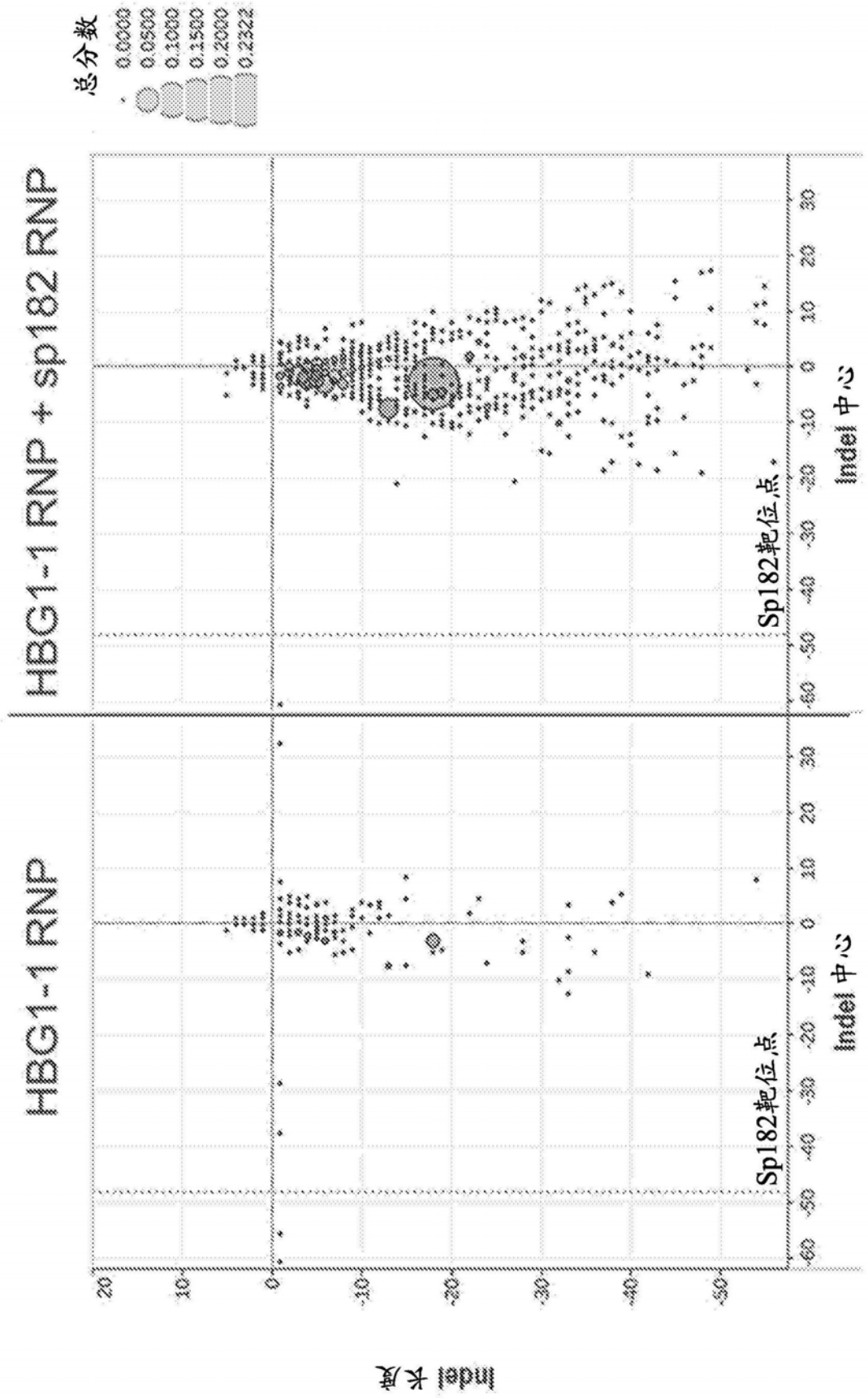

总分数
0.0000
0.0500
0.1000
0.1500
0.2000
0.2322
HBG1-1 RNP + sp182 RNP
HBG1-1 RNP
Sp182靶位点
Indel 中心
Indel 长度

图32A

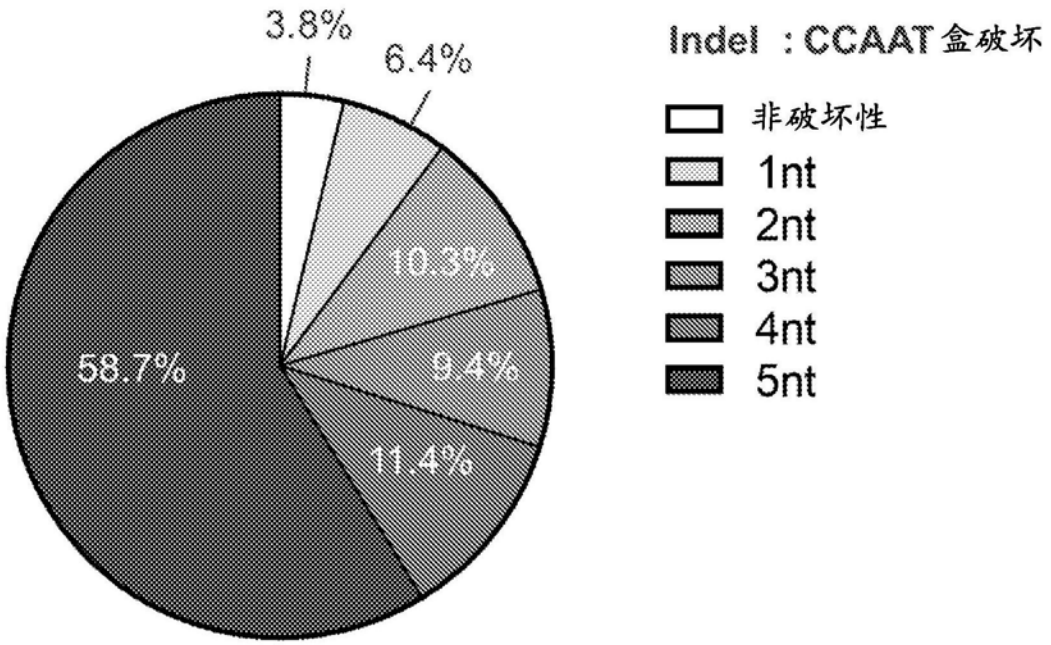


图32B

转让（Attorn）

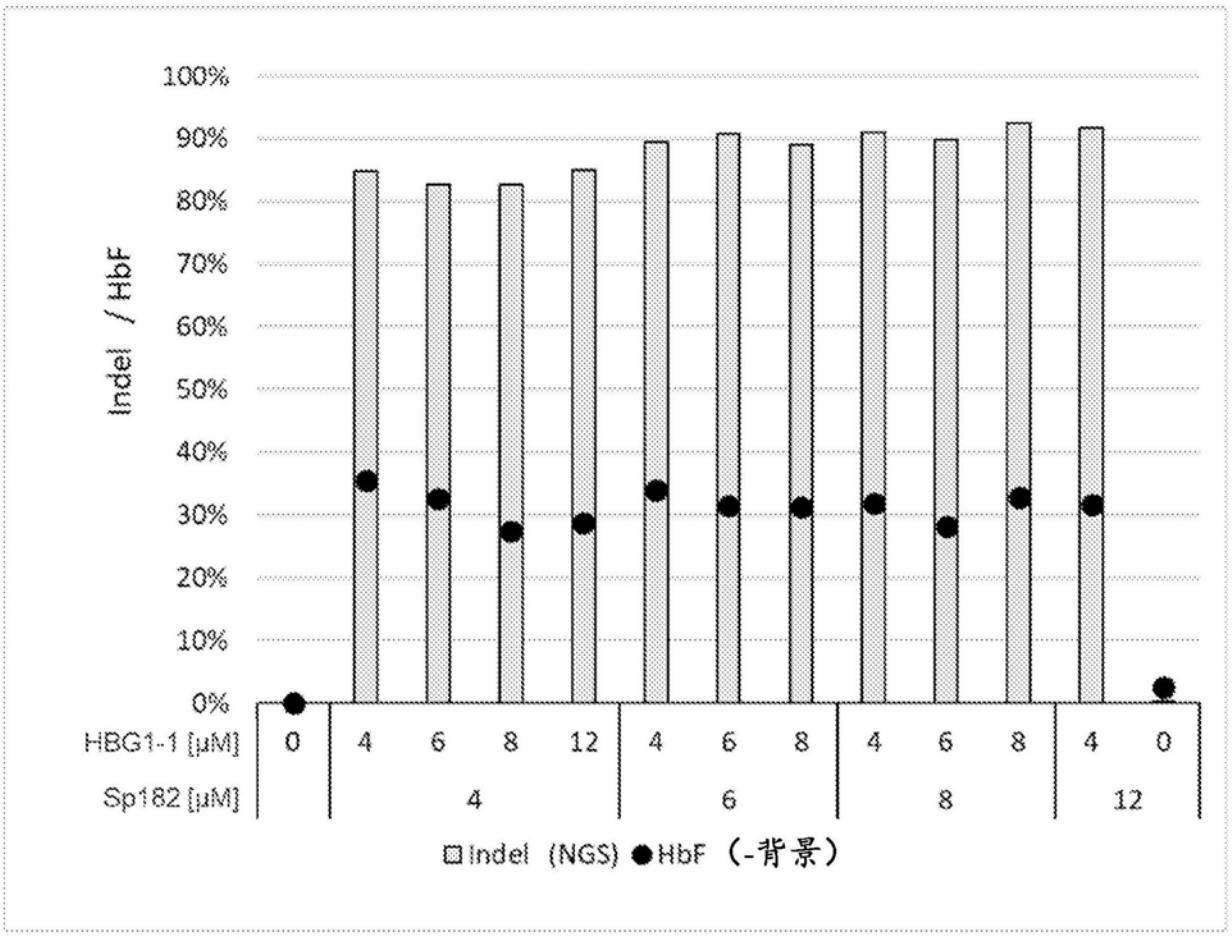


图33

转让（Attorn）

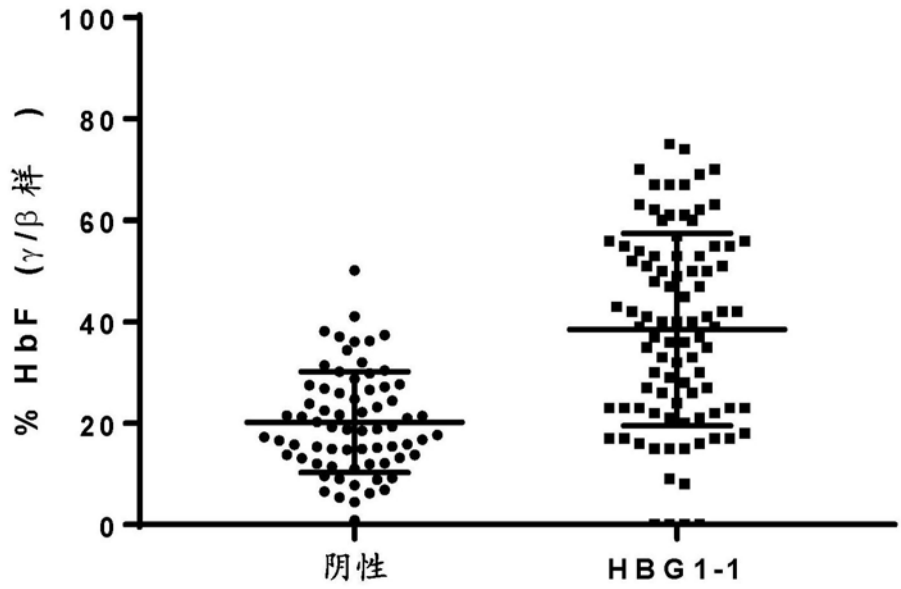


图34

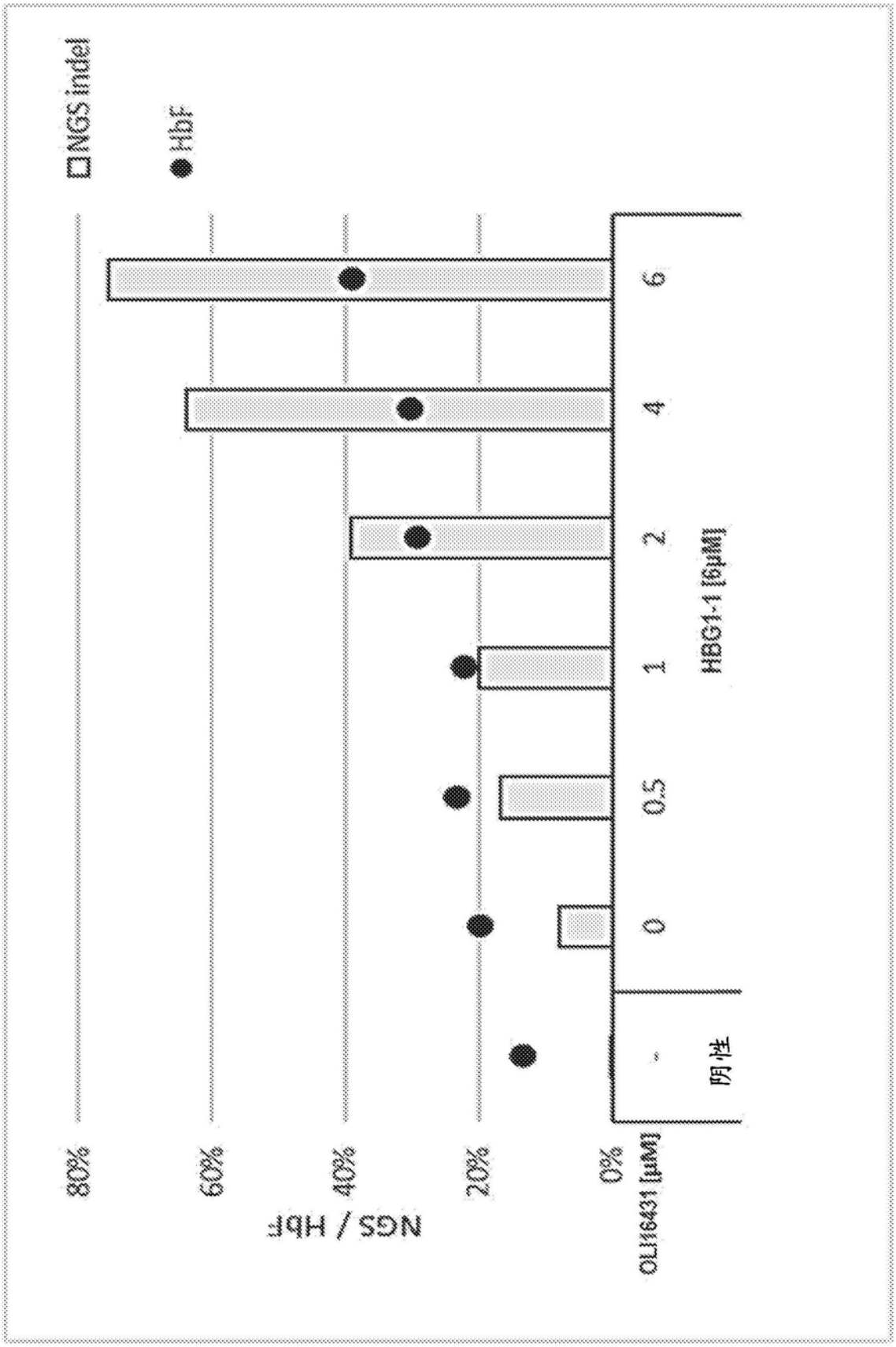

NGS indel
HbF
NGS / HbF
80%
60%
40%
20%
0%
OLI16431 [μM]
-
0
0.5
1
2
4
6
HBG1-1 [6μM]
阴性

图35A

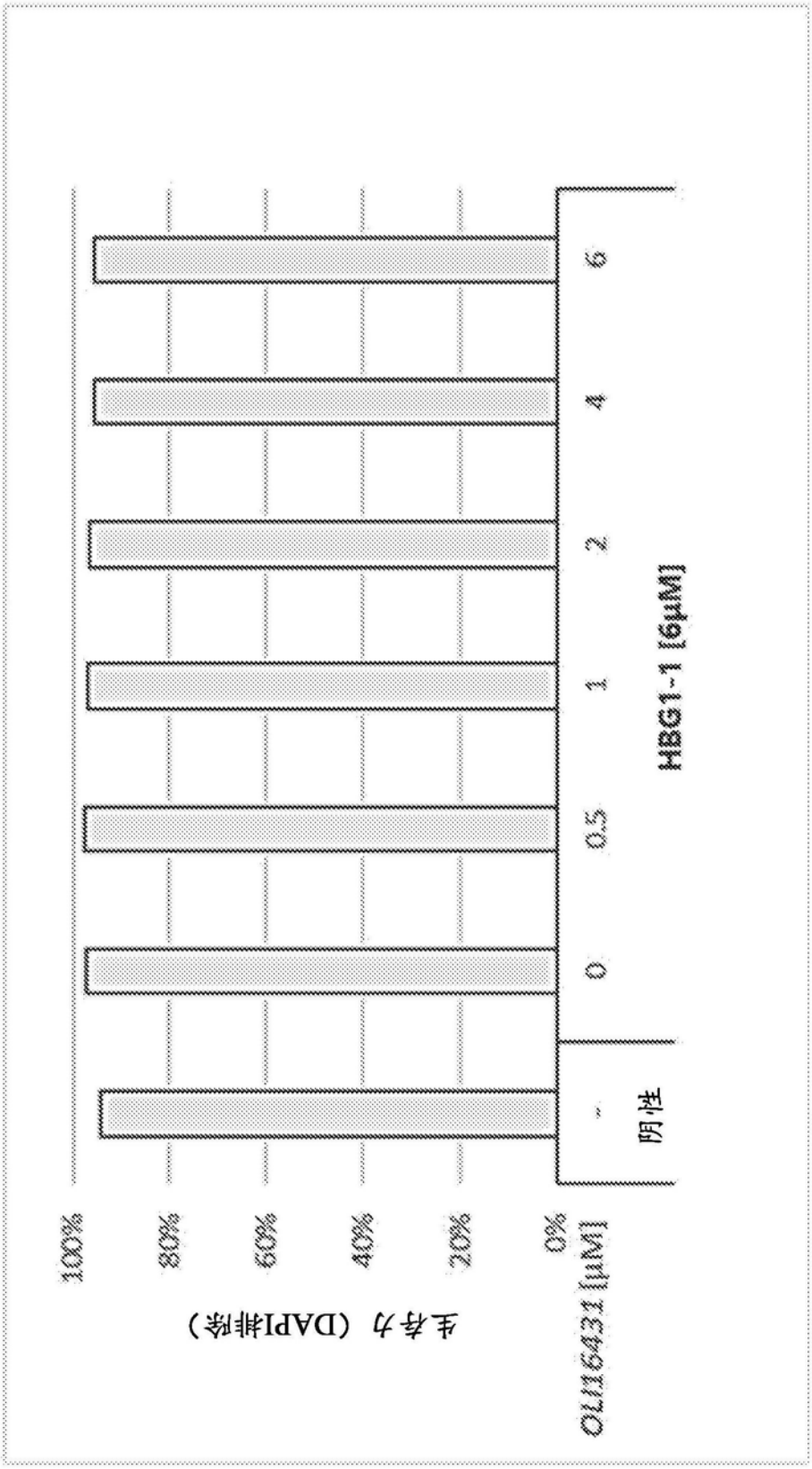


图35B

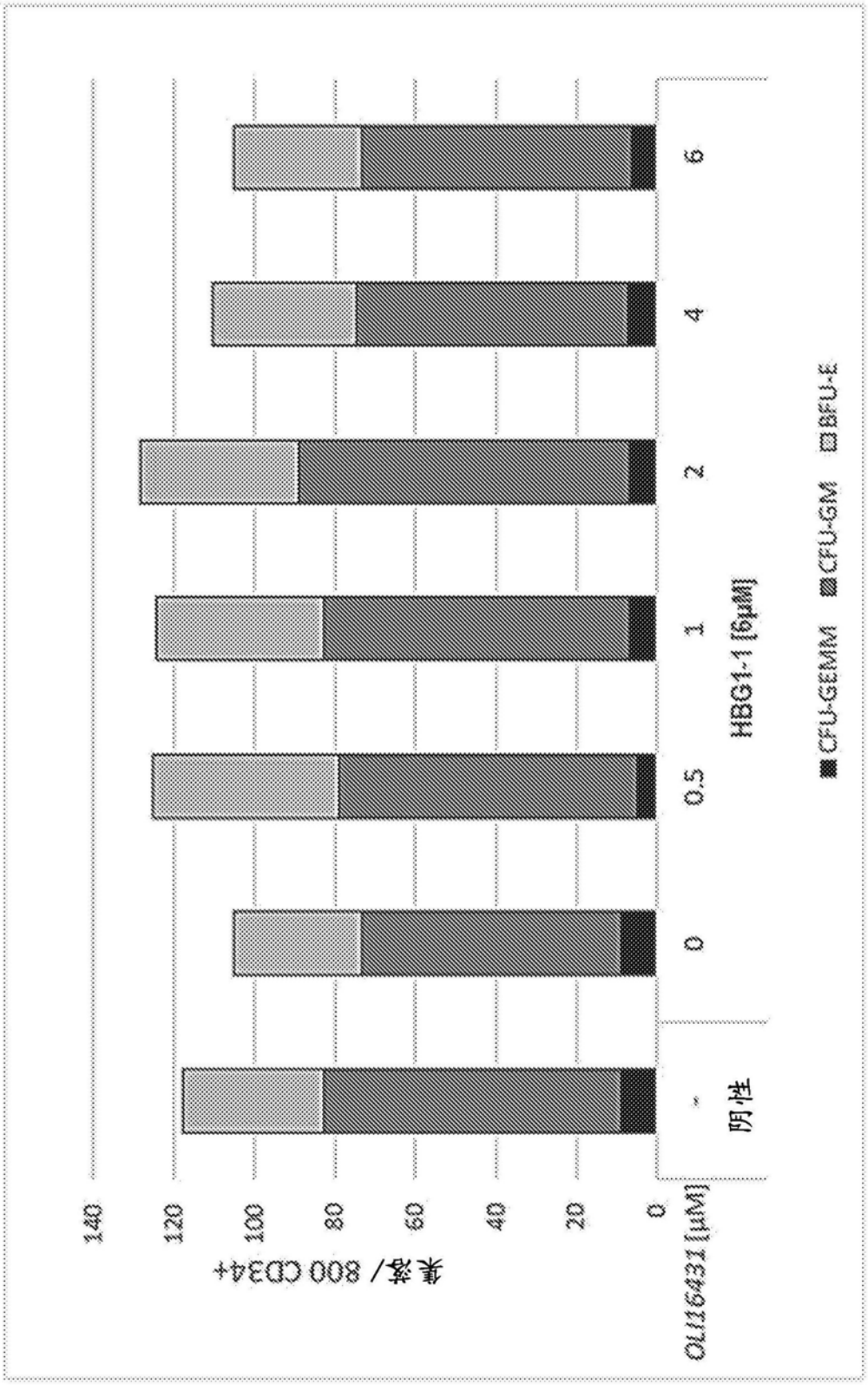

集落/ 800 CD34+
140
120
100
80
60
40
20
0
OU16431 [μM]
-
阴性
0
0.5
1
2
4
6
HBG1-1 [6μM]
CFU-GEMM
CFU-GM
BFU-E

图35C

转让（Attorn）

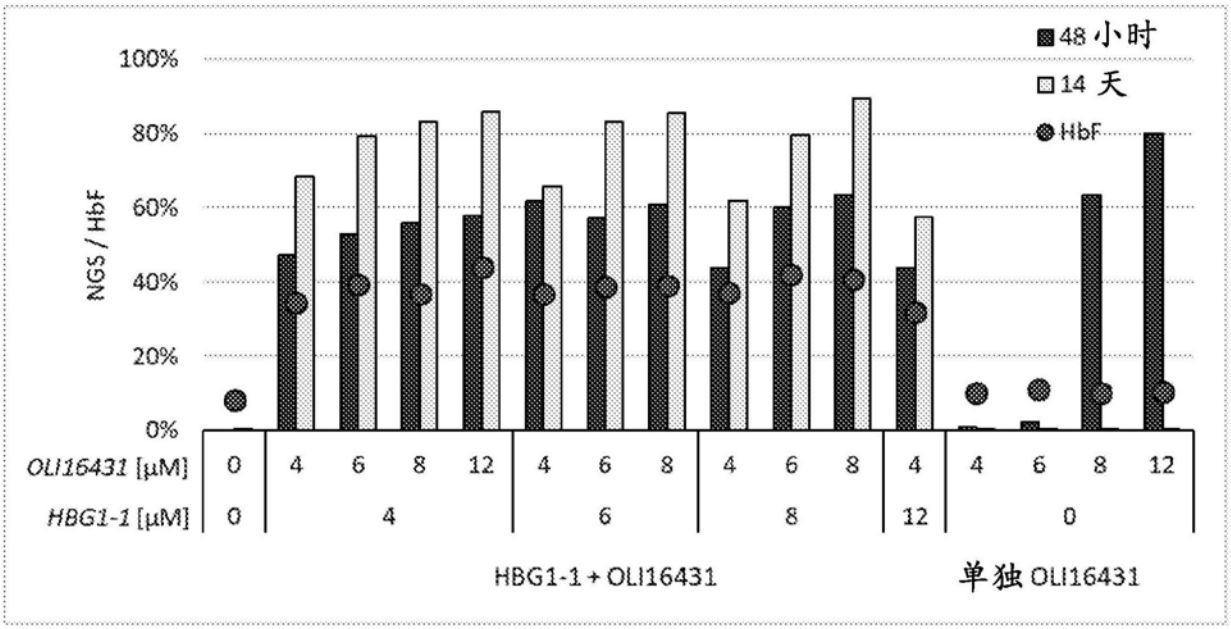


图36A

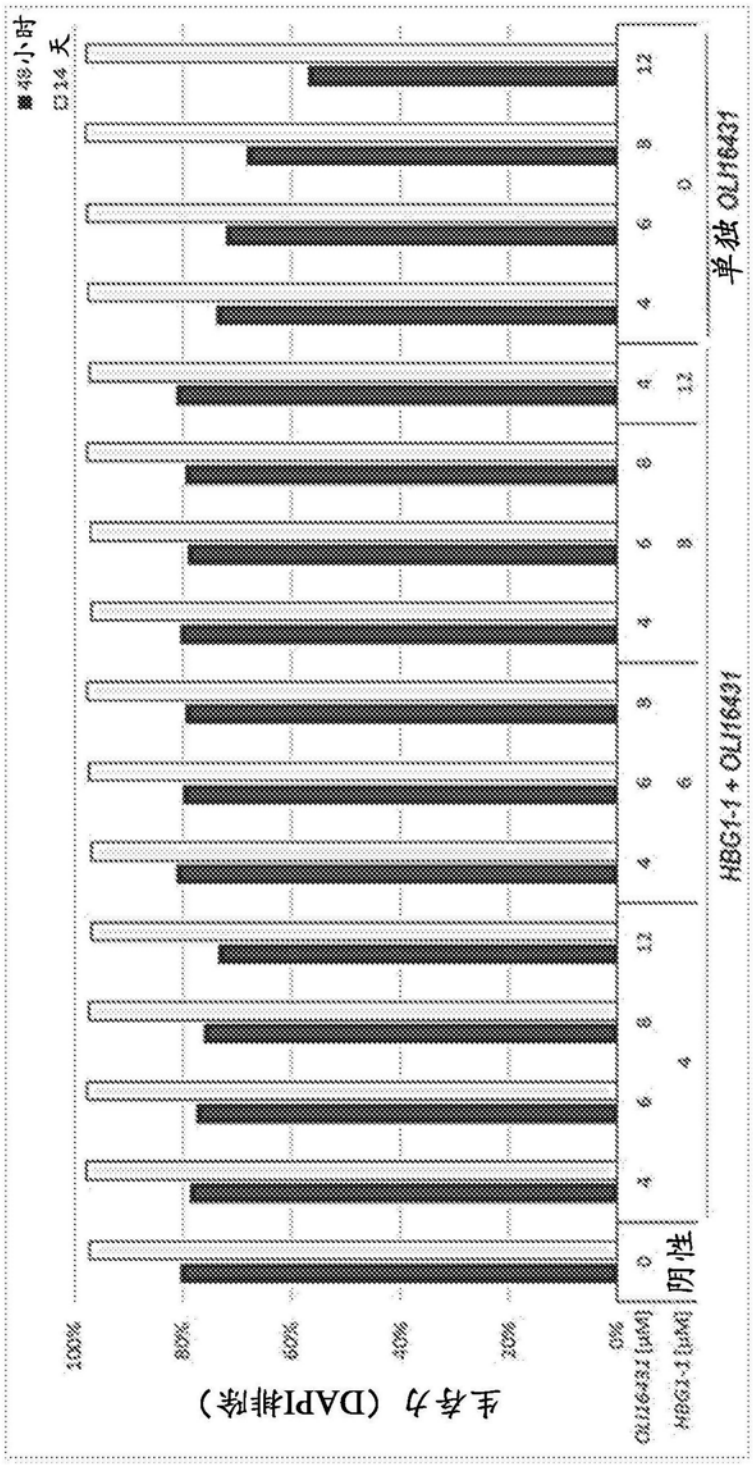

48小时
14天
生存力（DAPI排除）
100%
80%
60%
40%
20%
0%
OLI16431 (μM)
HBG1-1 (μM)
阴性
0
4
6
8
12
4
6
8
12
4
6
8
12
4
6
8
12
HBG1-1 + OLI16431
单独 OLI16431

图36B

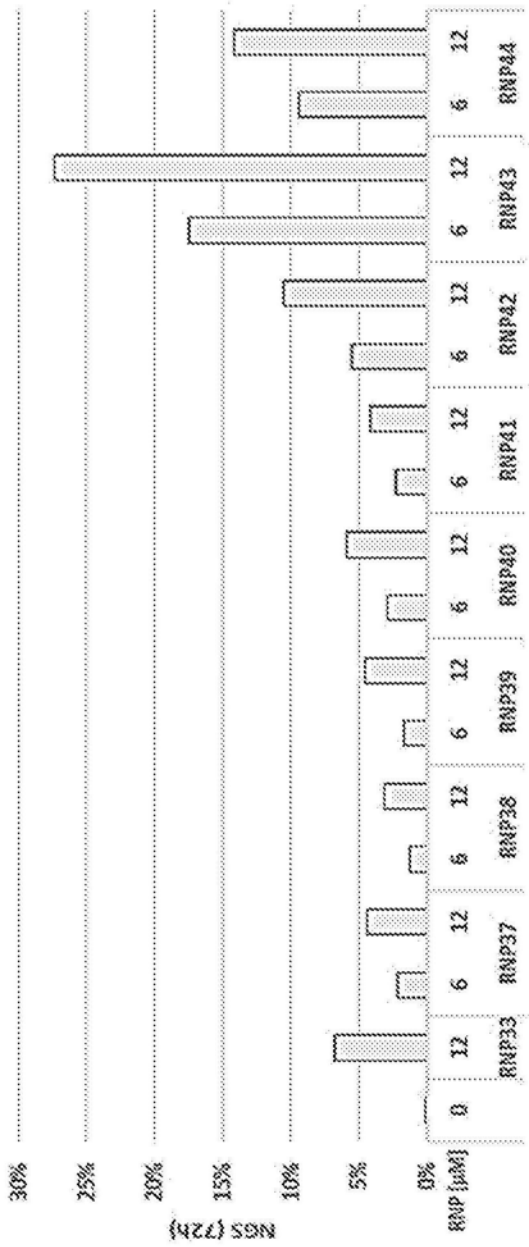

NGS (72h)
30%
25%
20%
15%
10%
5%
0%
RNP (μM)
0
12
6
12
RNP33
6
12
RNP37
6
12
RNP38
6
12
RNP39
6
12
RNP40
6
12
RNP41
6
12
RNP42
6
12
RNP43
6
12
RNP44

图37

转让（Attorn）

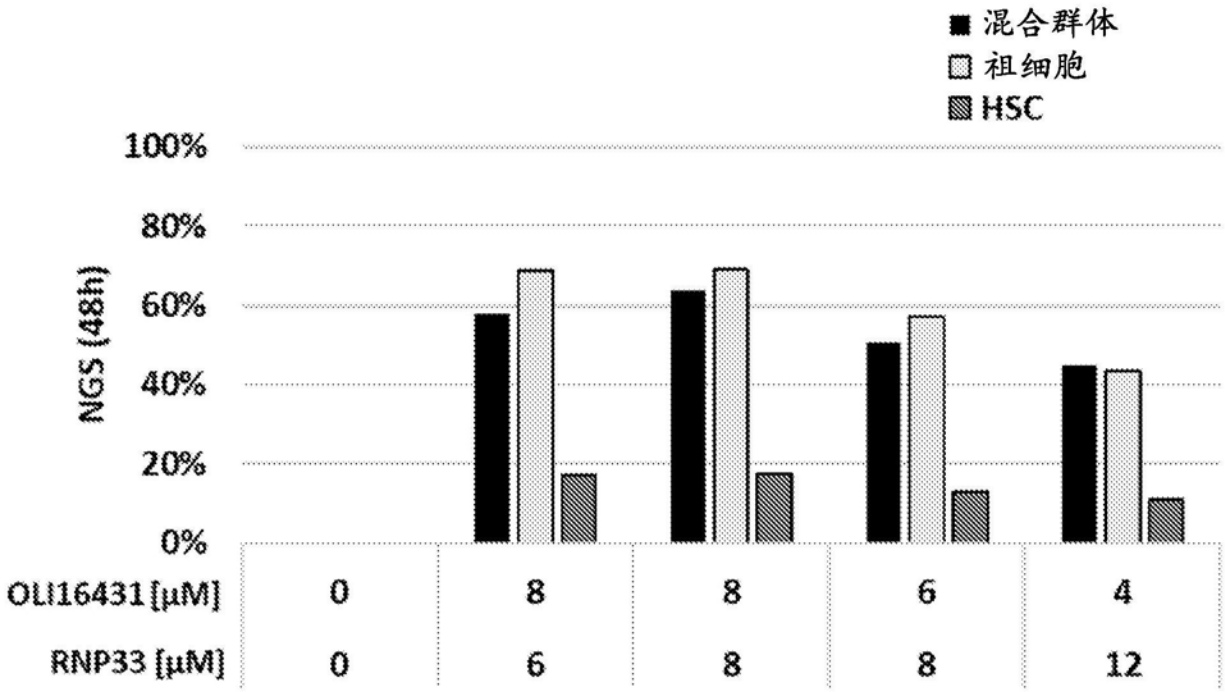


图38A

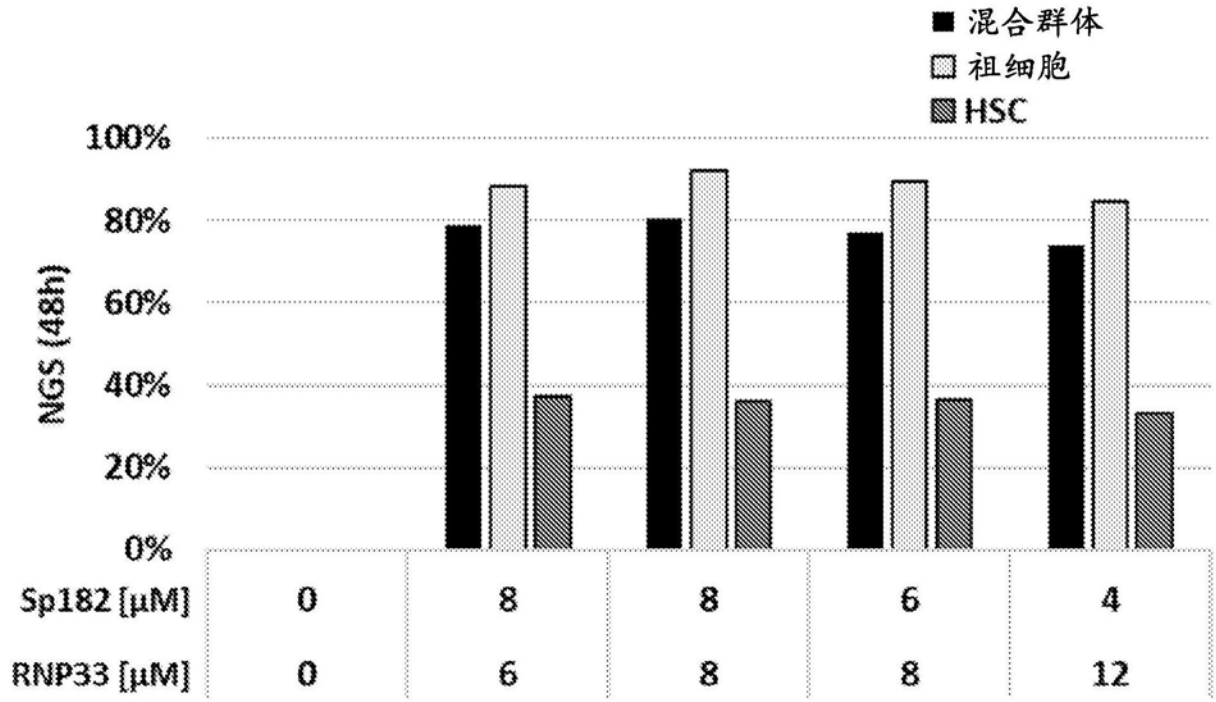


图38B

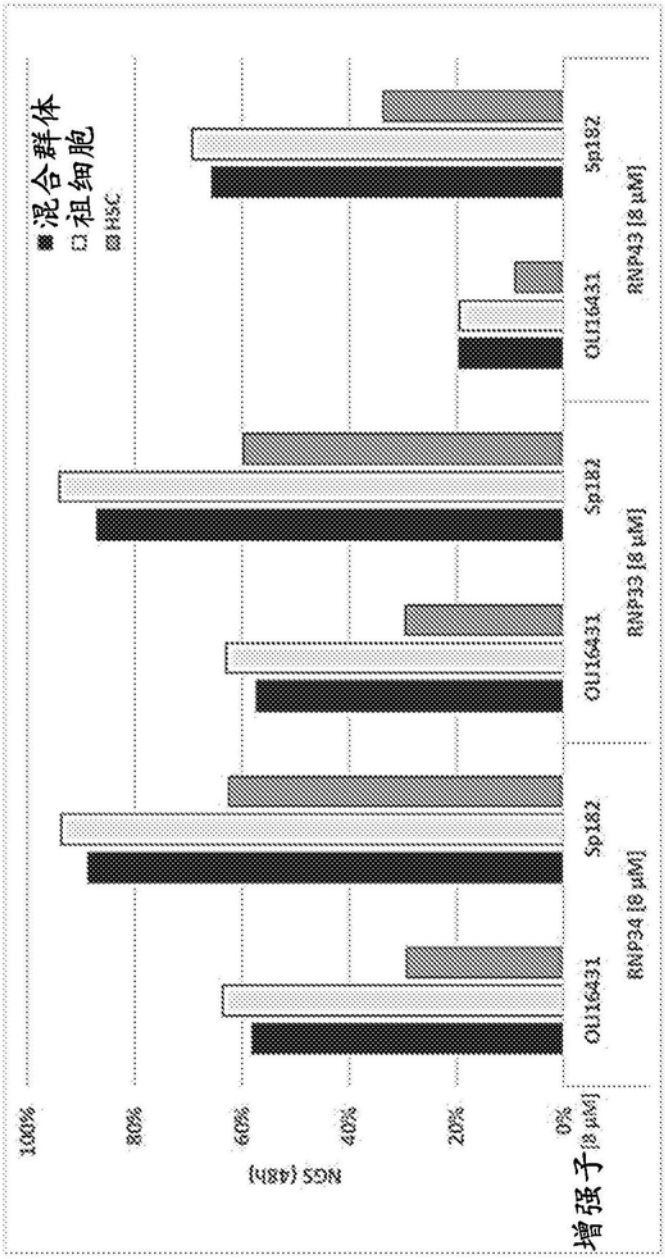

混合群体
祖细胞
HSC
NGS (48h)
100%
80%
60%
40%
20%
0%
增强子 [8 μM]
OU16431
Sp182
RNP34 [8 μM]
OU16431
Sp182
RNP33 [8 μM]
OU16431
Sp182
RNP43 [8 μM]

图39

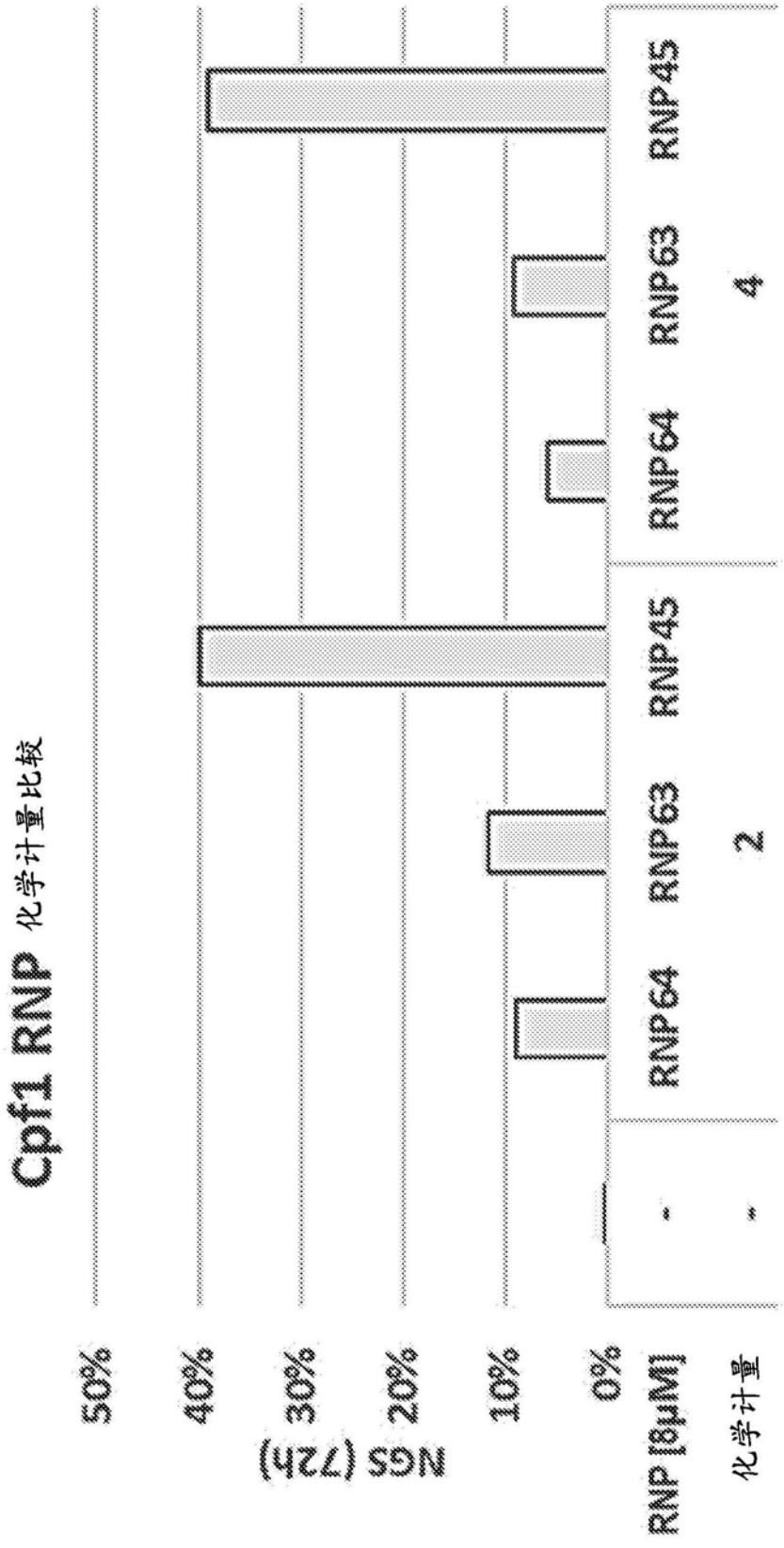

Cpf1 RNP 化学计量比较
NGS (72h)
50%
40%
30%
20%
10%
0%
RNP [8μM]
化学计量
-
-
RNP64
RNP63
RNP45
2
RNP64
RNP63
RNP45
4

图40

转让（Attorn）

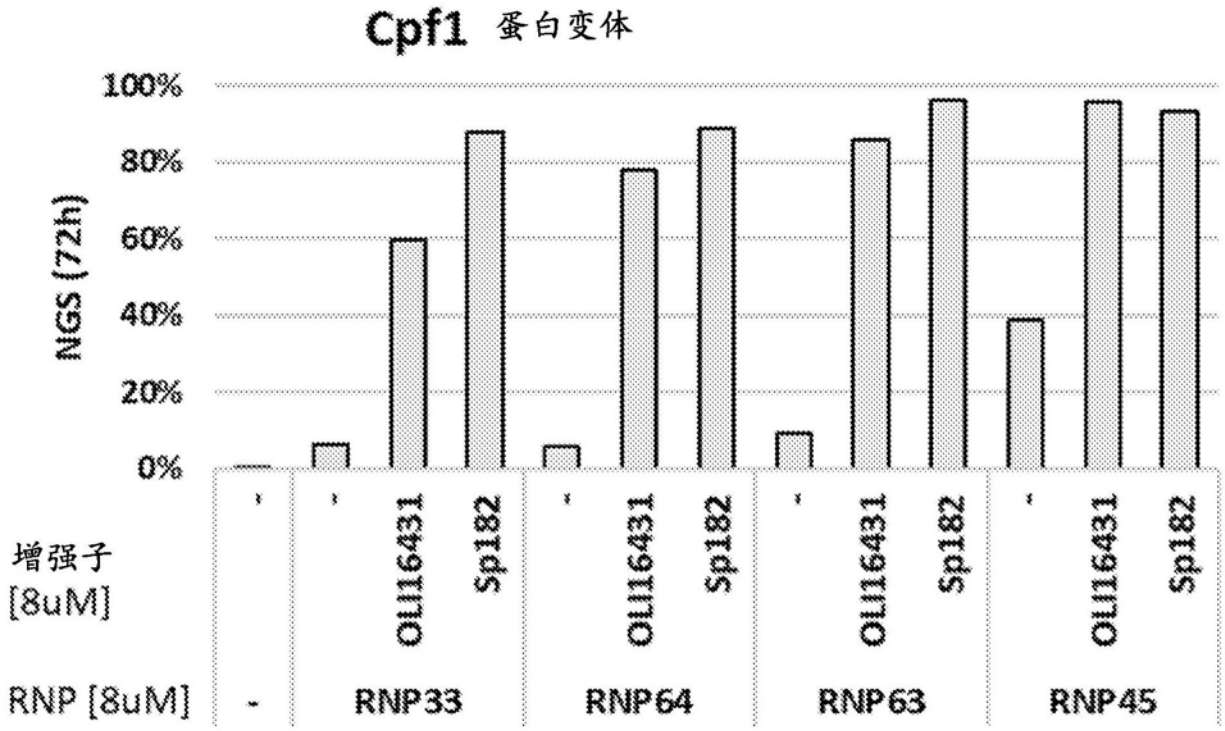


图41

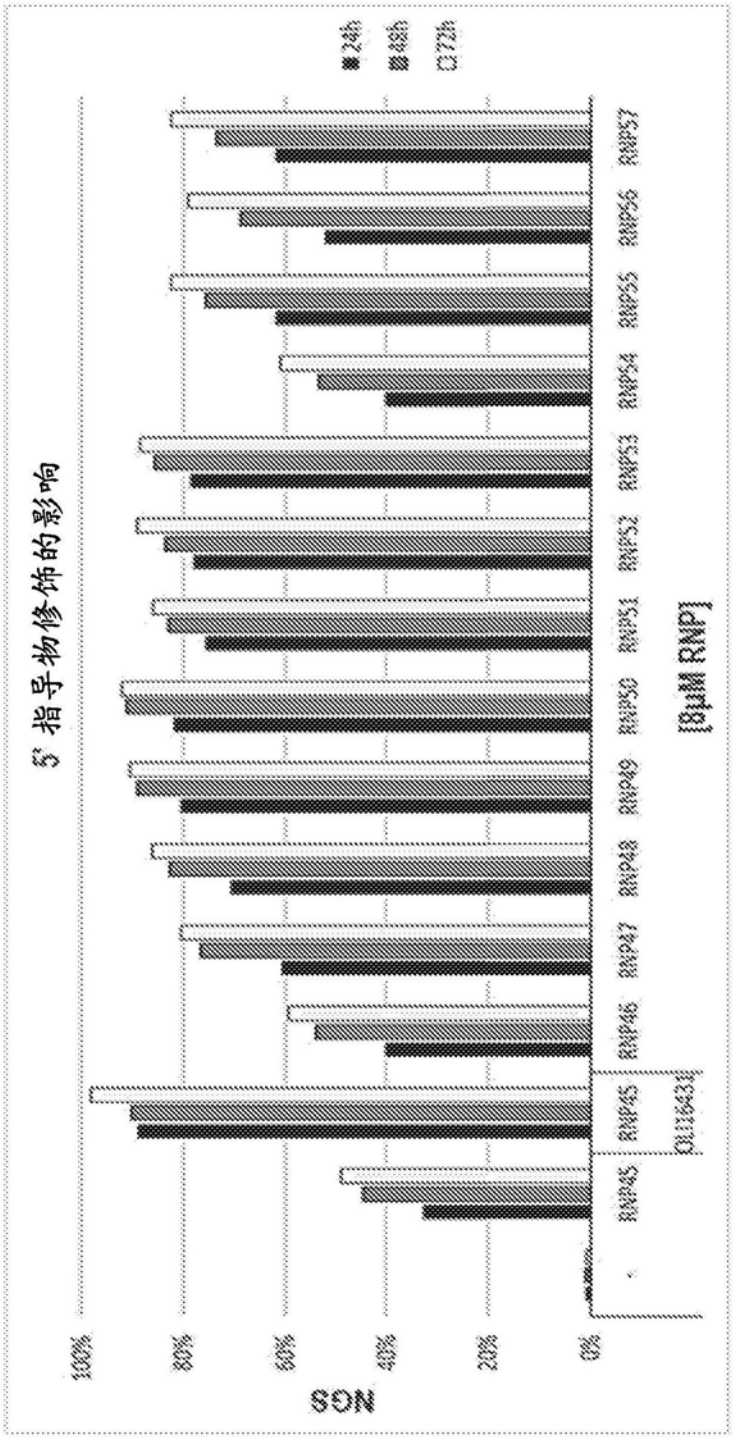

24h
48h
72h
5' 指导物修饰的影响
RNP57
RNP56
RNP55
RNP54
RNP53
RNP52
RNP51
RNP50
RNP49
RNP48
RNP47
RNP46
RNP45
RNP45
-
[8μM RNP]
NOS
100%
80%
60%
40%
20%
0%

图42

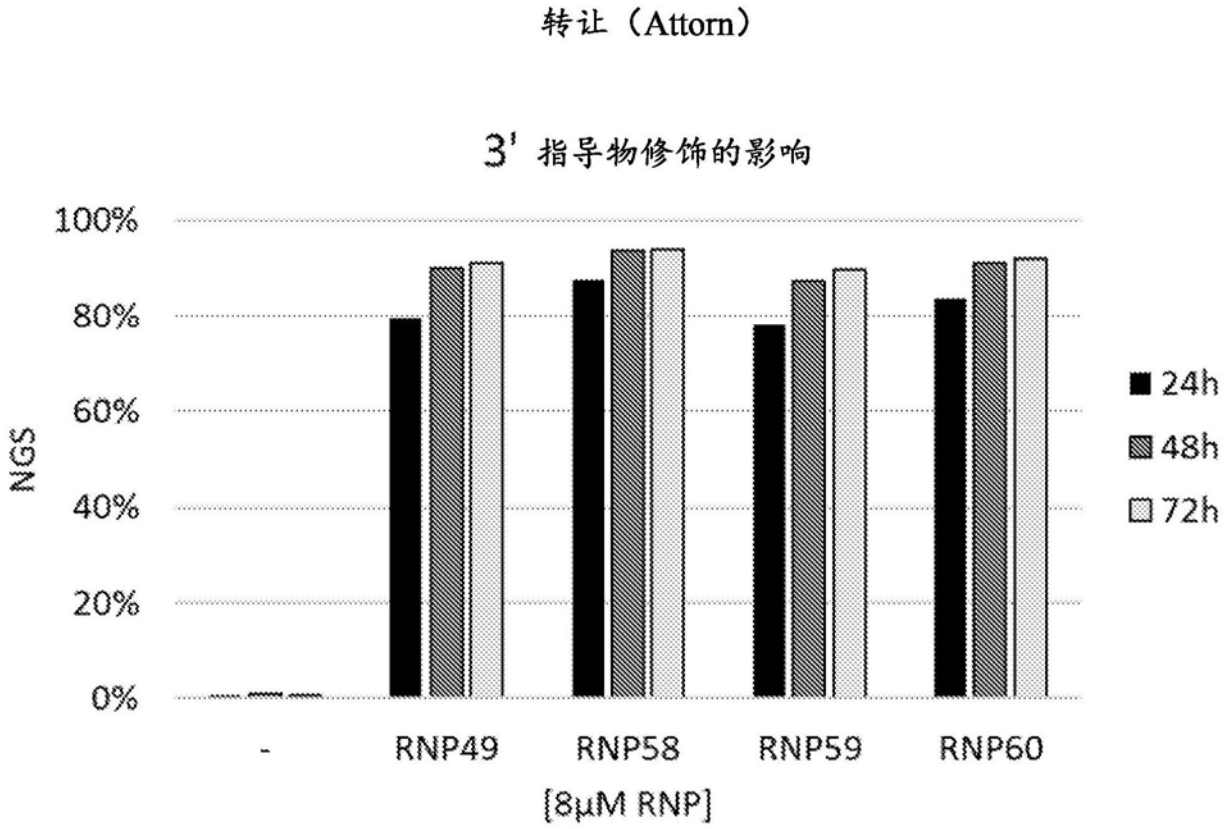

转让（Attorn）
3' 指导物修饰的影响
100%
80%
60%
40%
20%
0%
NGS
24h
48h
72h
-
RNP49
RNP58
RNP59
RNP60
[8μM RNP]

图43

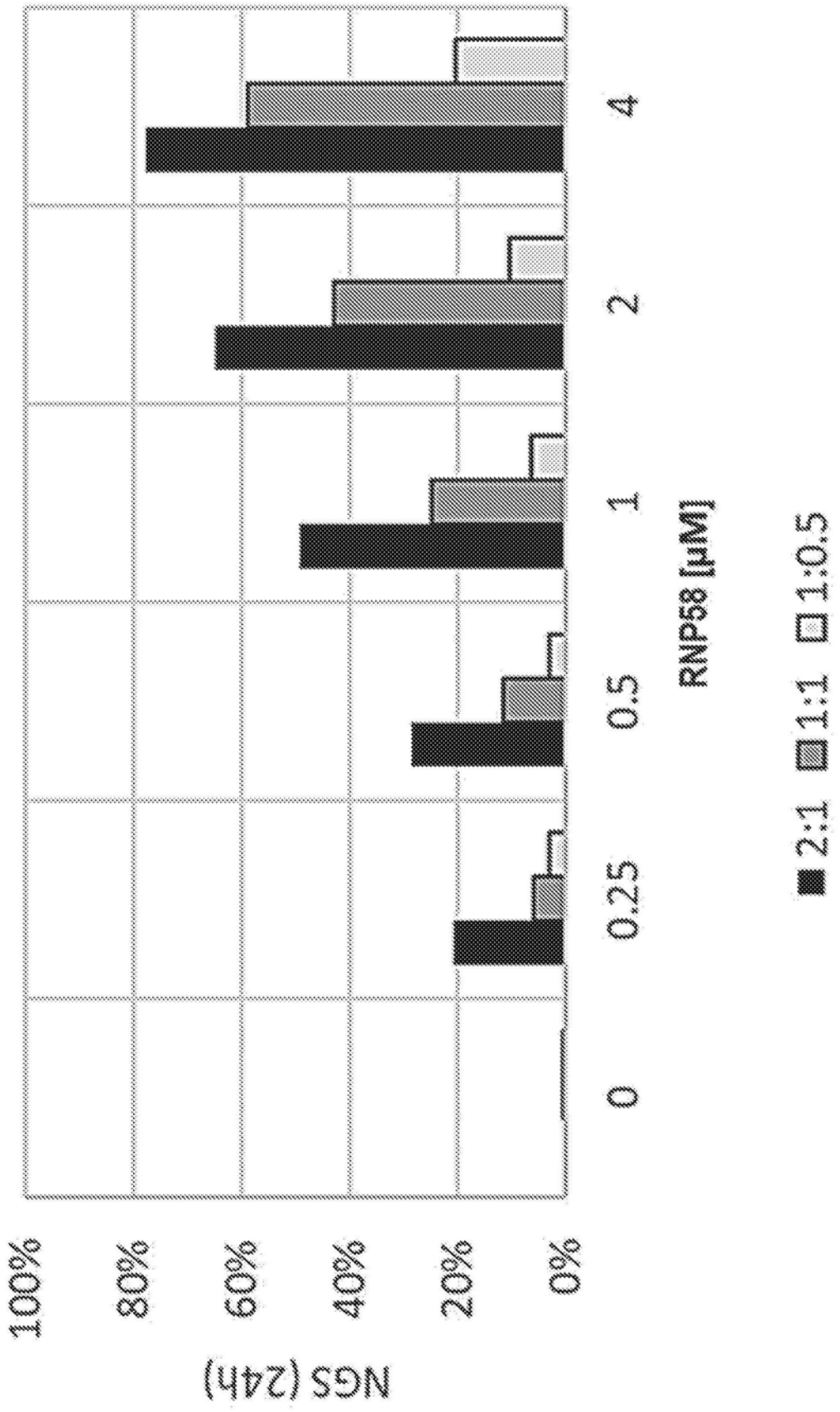


图44A

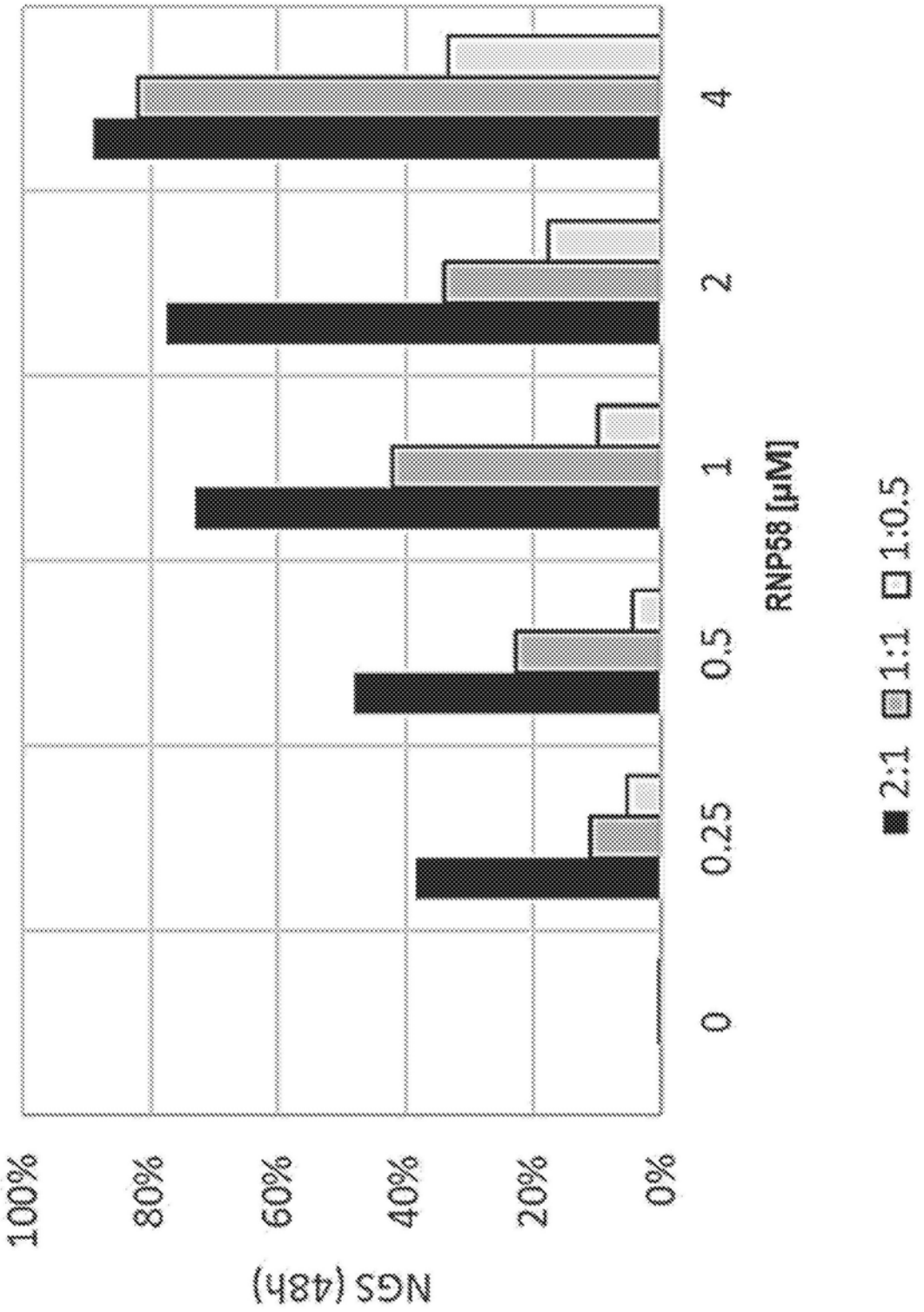


图44B

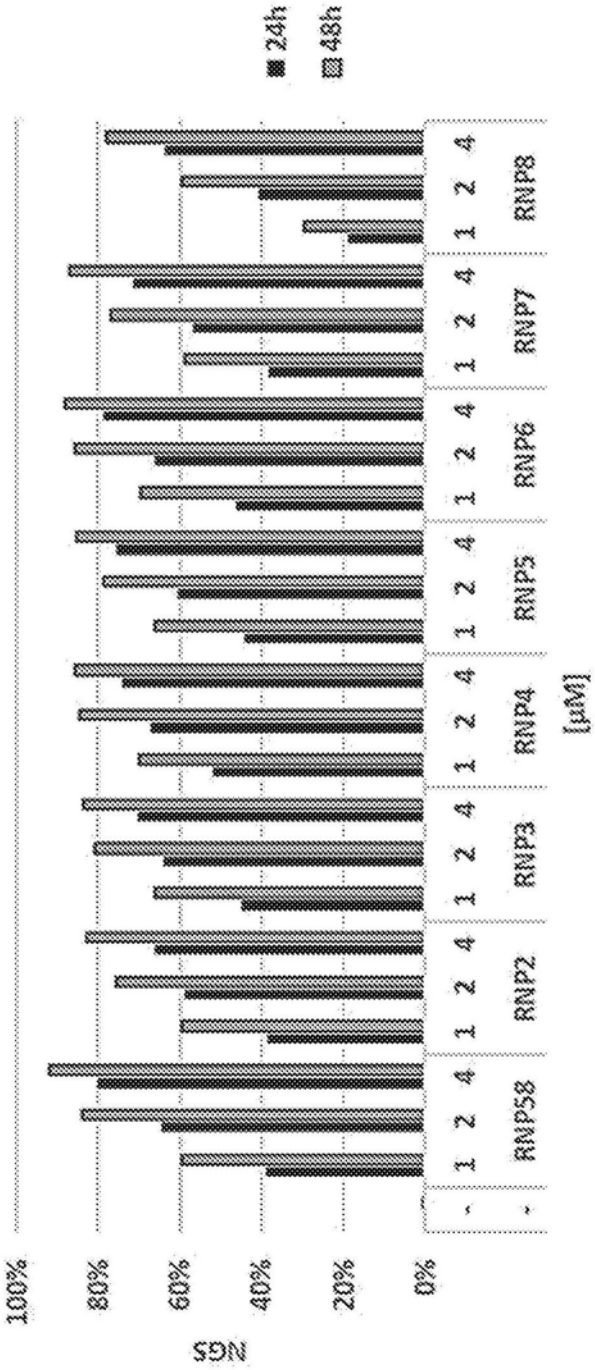

24h
48h
100%
80%
60%
40%
20%
0%
NGS
-
-
RNP58
1 2 4
RNP2
1 2 4
RNP3
1 2 4
RNP4
1 2 4
RNP5
1 2 4
RNP6
1 2 4
RNP7
1 2 4
RNP8
1 2 4
[μM]

图45A

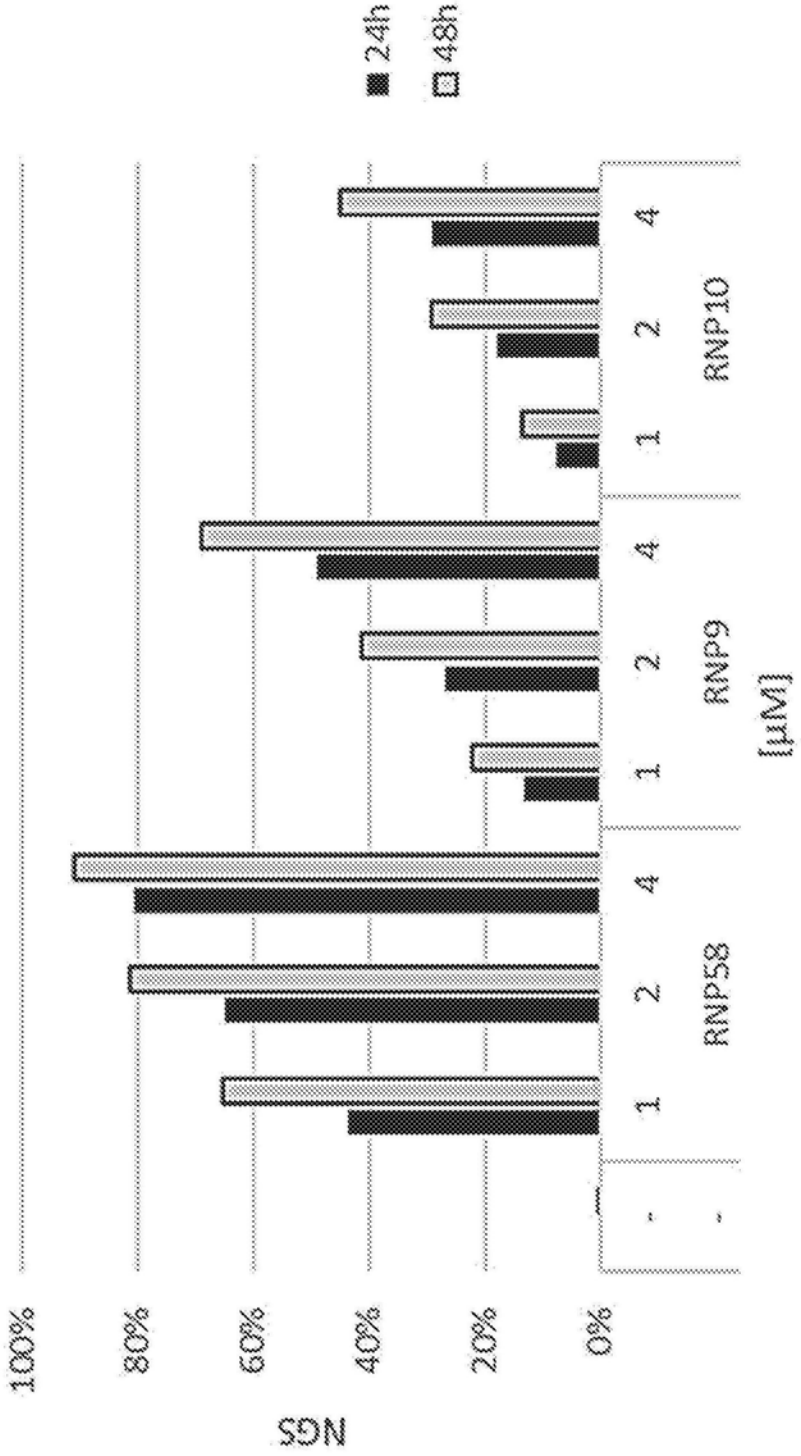


图45B

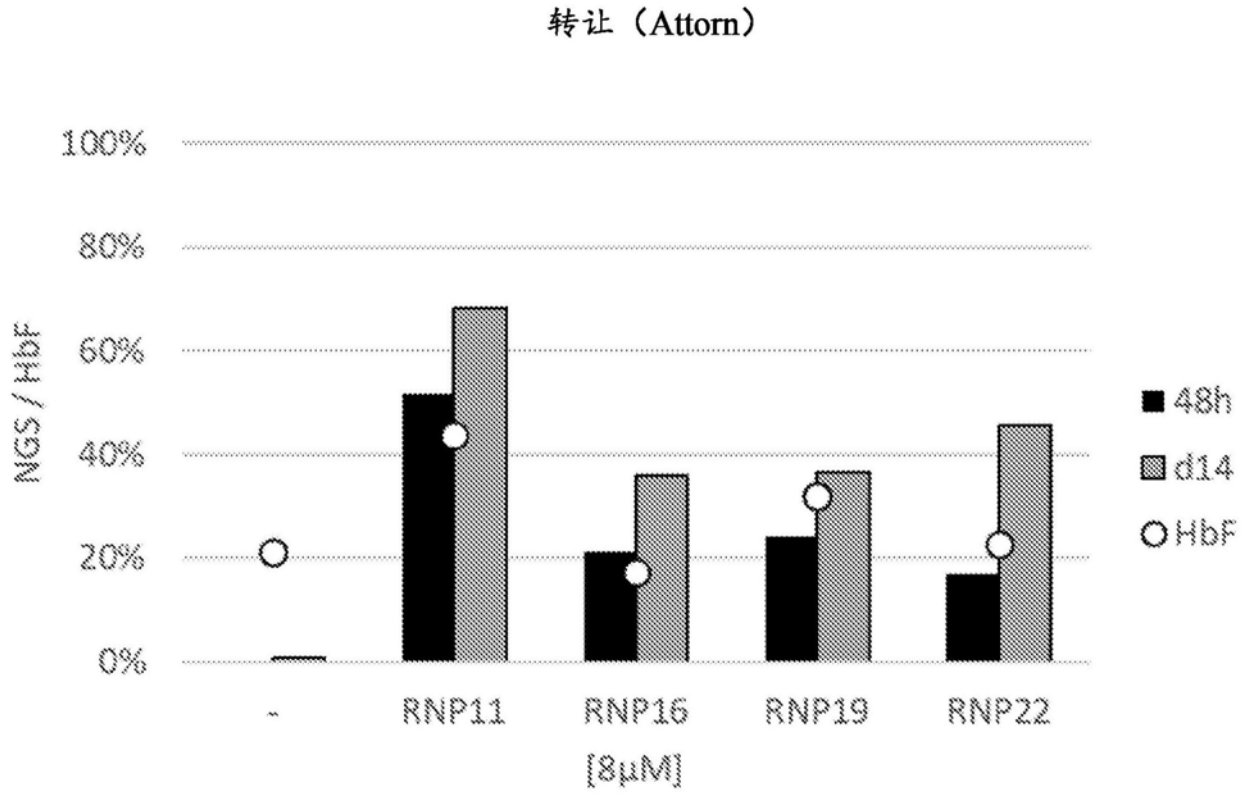

转让（Attorn）
100%
80%
60%
40%
20%
0%
NGS / HbF
48h
d14
HbF
-
RNP11
RNP16
RNP19
RNP22
[8μM]

图46A

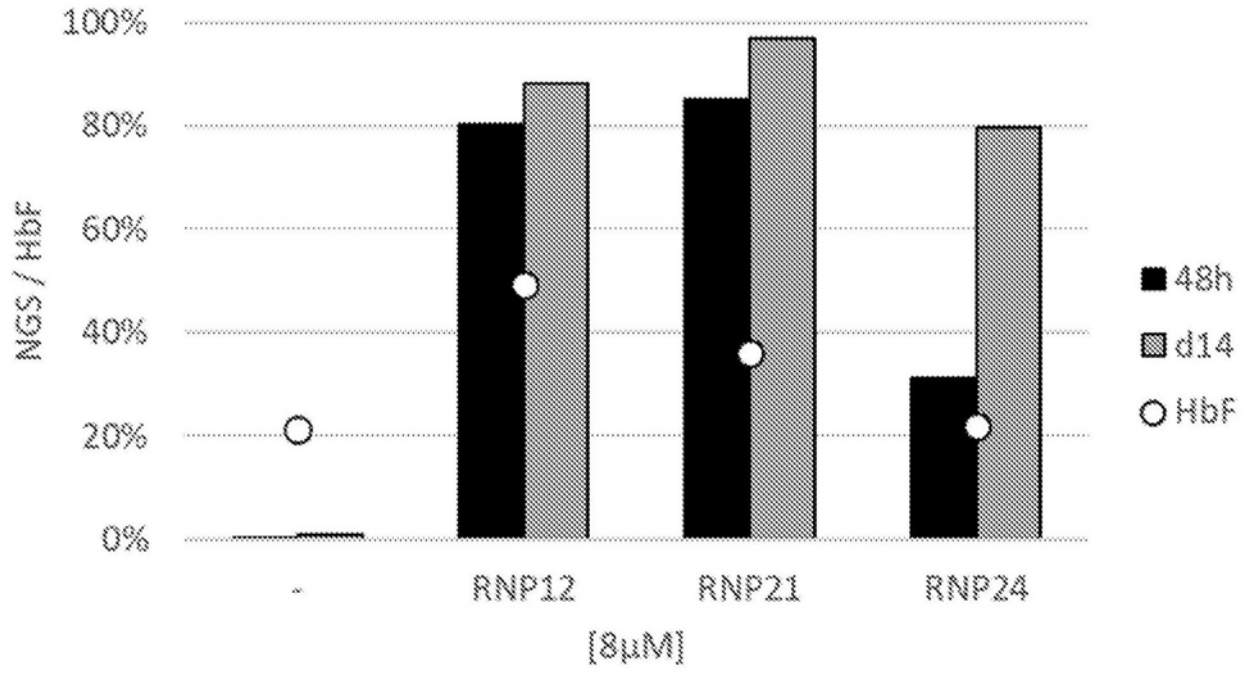

100%
80%
60%
40%
20%
0%
NGS / HbF
48h
d14
HbF
-
RNP12
RNP21
RNP24
[8μM]

图46B

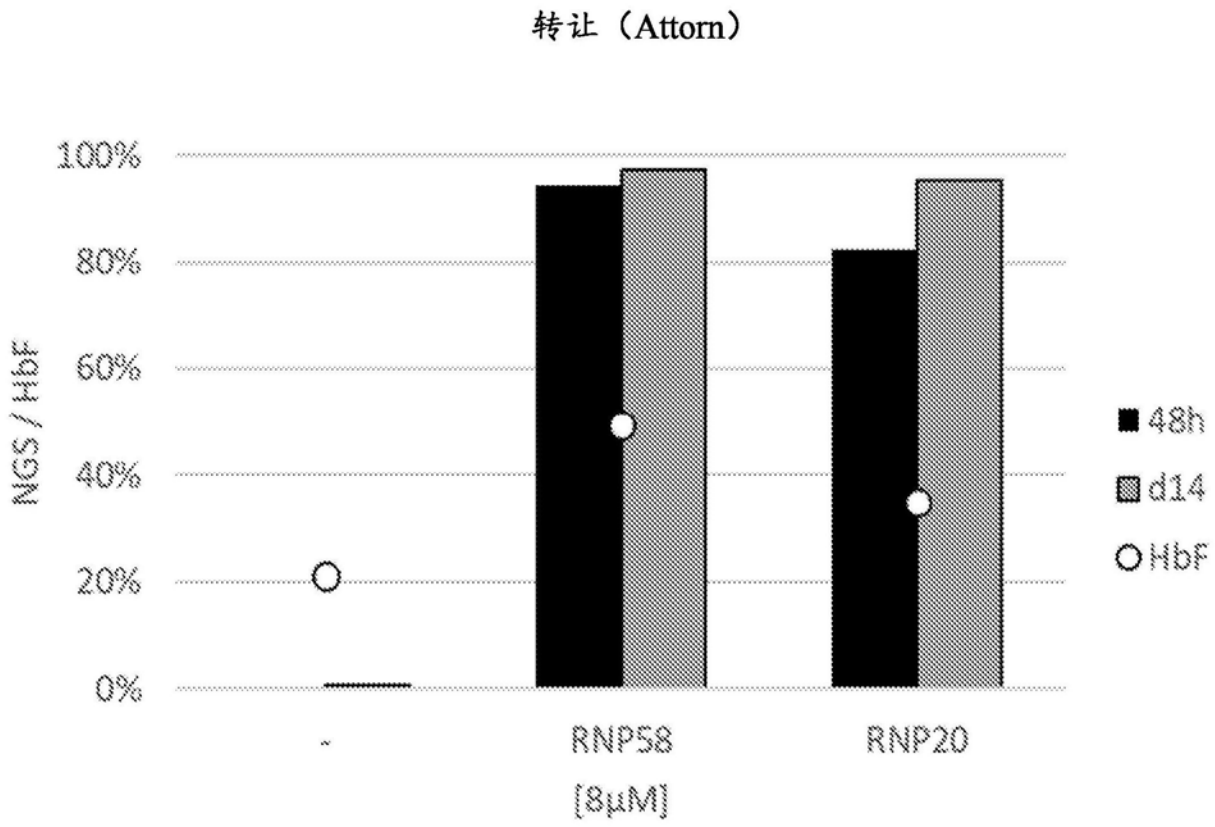


图46C

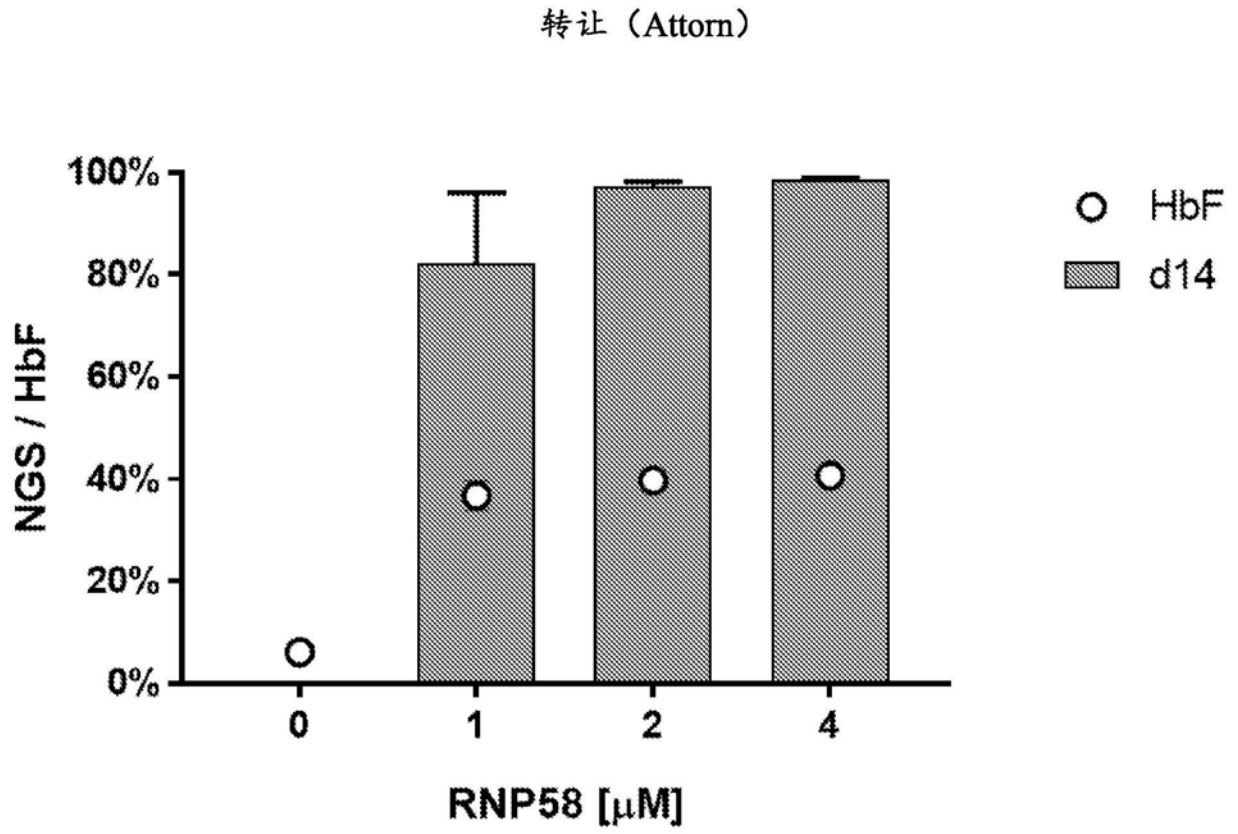

转让（Attorn）
100%
80%
60%
40%
20%
0%
NGS / HbF
0
1
2
4
RNP58 [μM]
HbF
d14

图47

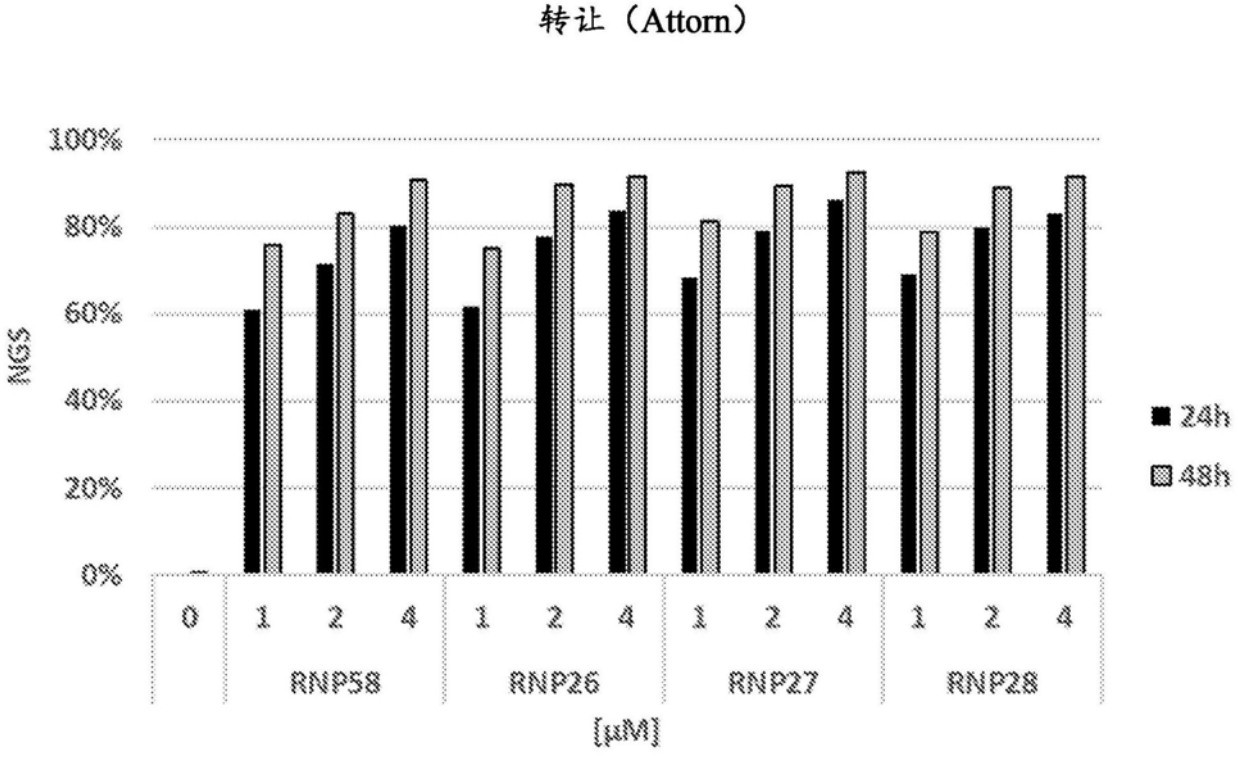

转让（Attorn）
100%
80%
60%
40%
20%
0%
NGS
0
1
2
4
1
2
4
1
2
4
1
2
4
RNP58
RNP26
RNP27
RNP28
[μM]
24h
48h

图48

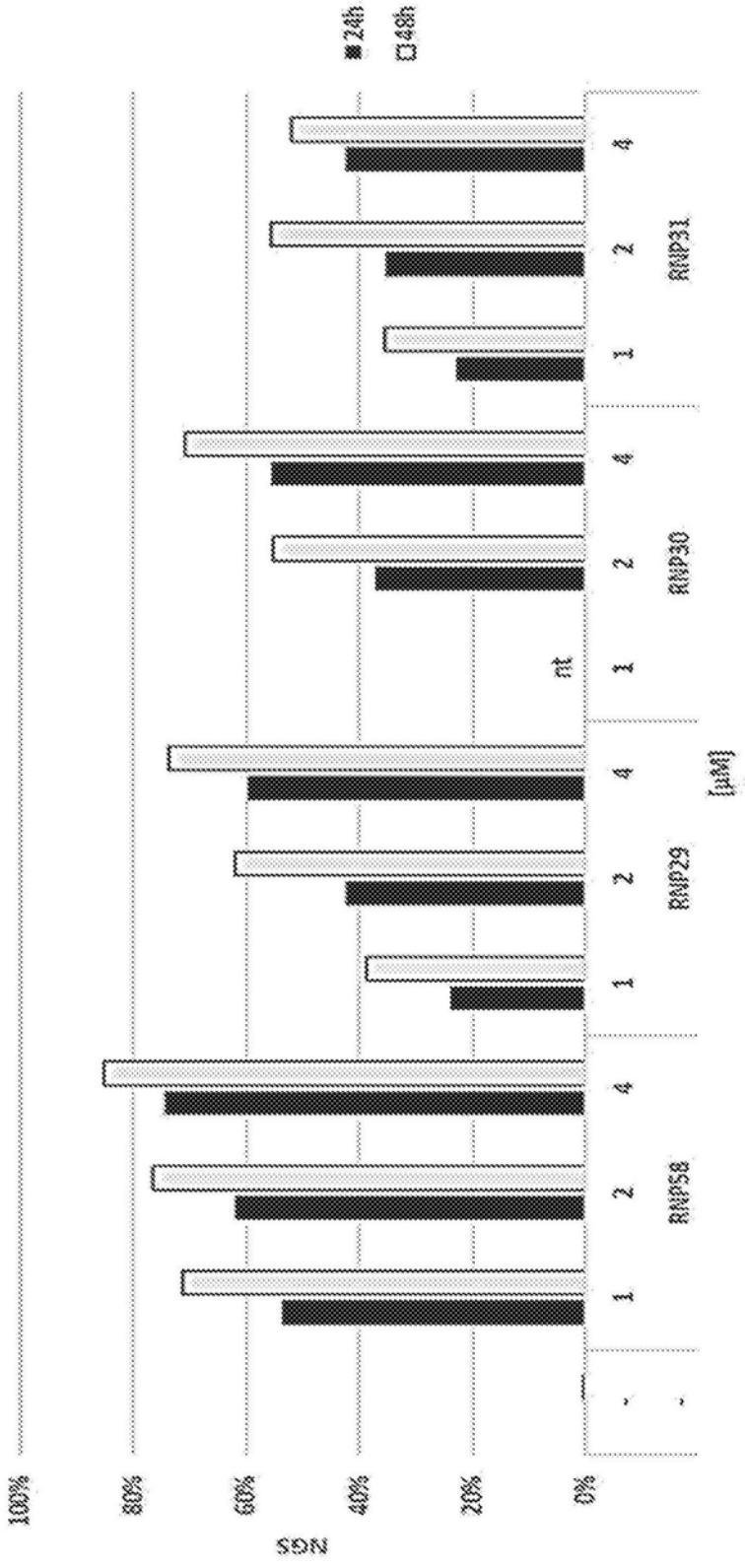

24h
48h
100%
80%
60%
40%
20%
0%
NGS
-
-
RMP58
1
2
4
RMP29
1
2
4
RMP30
1
2
4
nt
RMP31
1
2
4
[μM]

图49

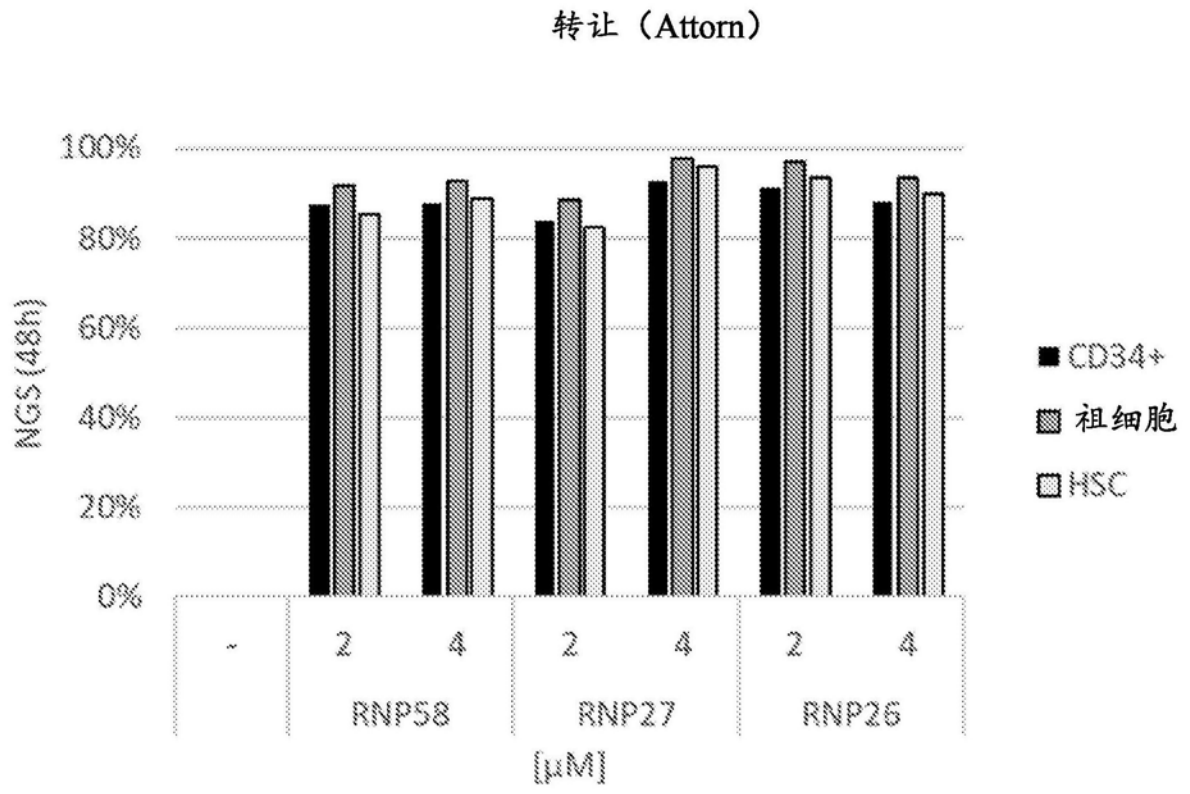

转让（Attorn）
100%
80%
60%
40%
20%
0%
NGS(48h)
-
2
4
2
4
2
4
RNP58
RNP27
RNP26
[μM]
CD34+
祖细胞
HSC

图50

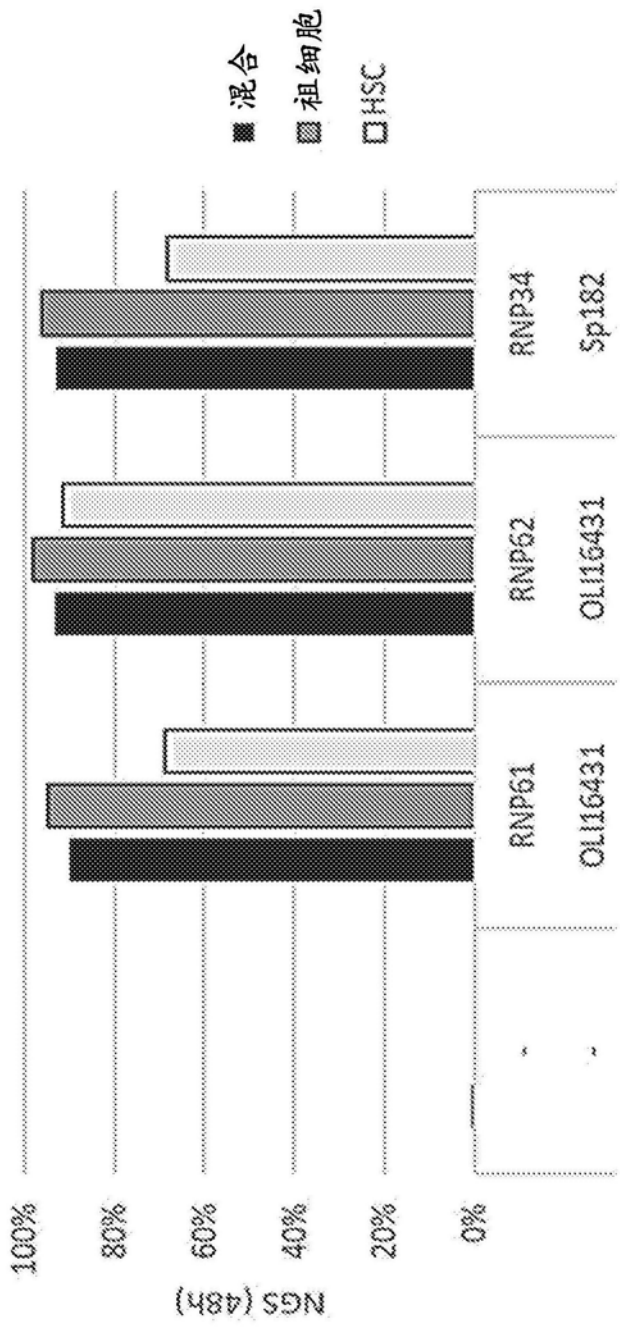

混合
祖细胞
HSC
RNP34
Sp182
RNP62
OLI16431
RNP61
OLI16431
-
-
100%
80%
60%
40%
20%
0%
NGS (48h)

图51

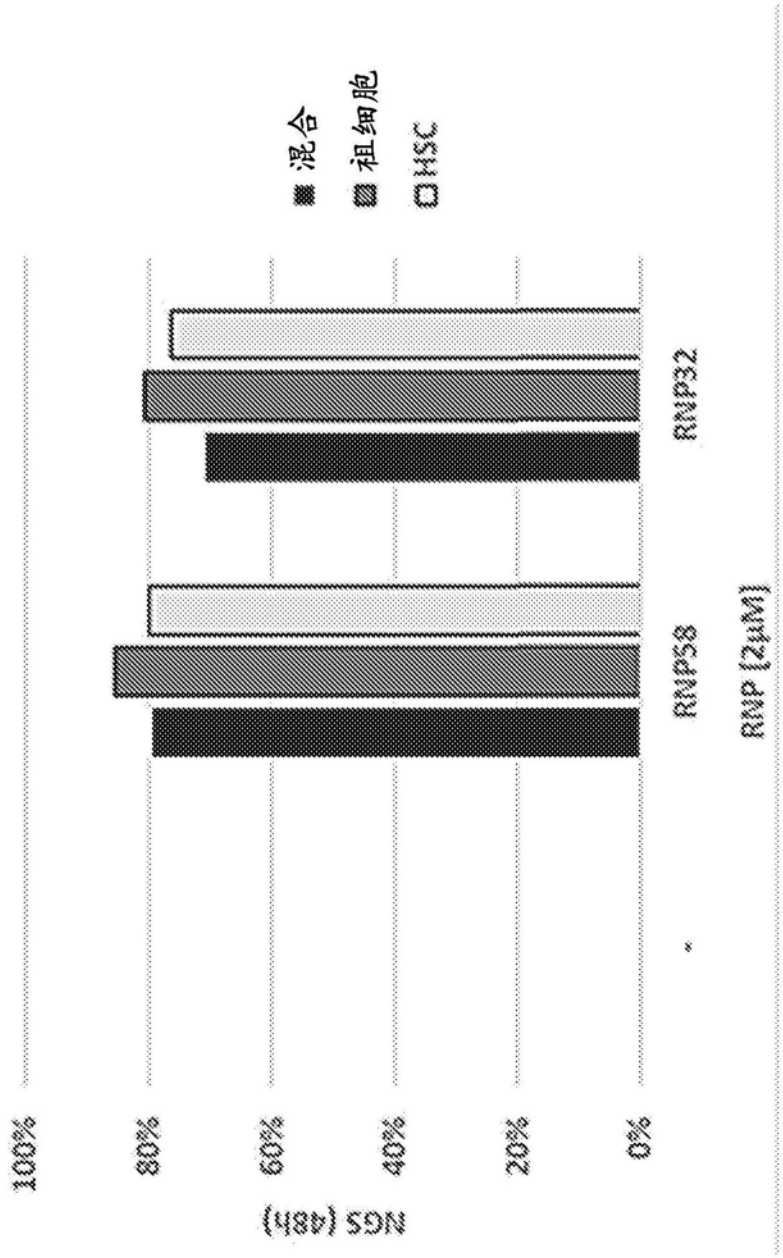


图52

转让（Attorn）

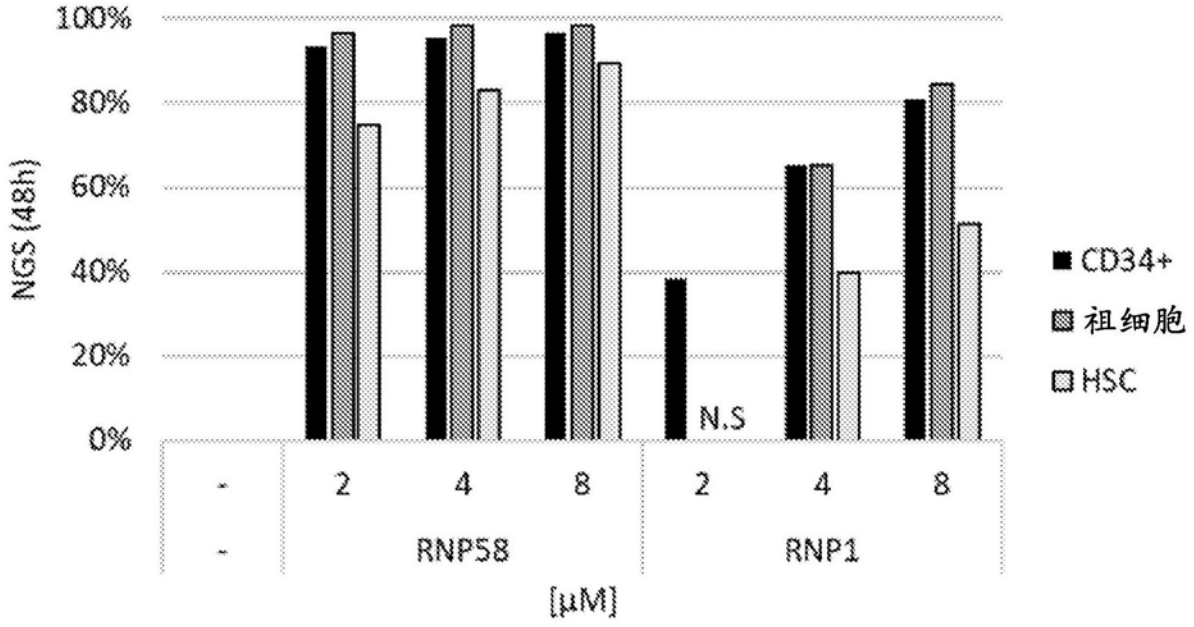


图53

转让（Attorn）

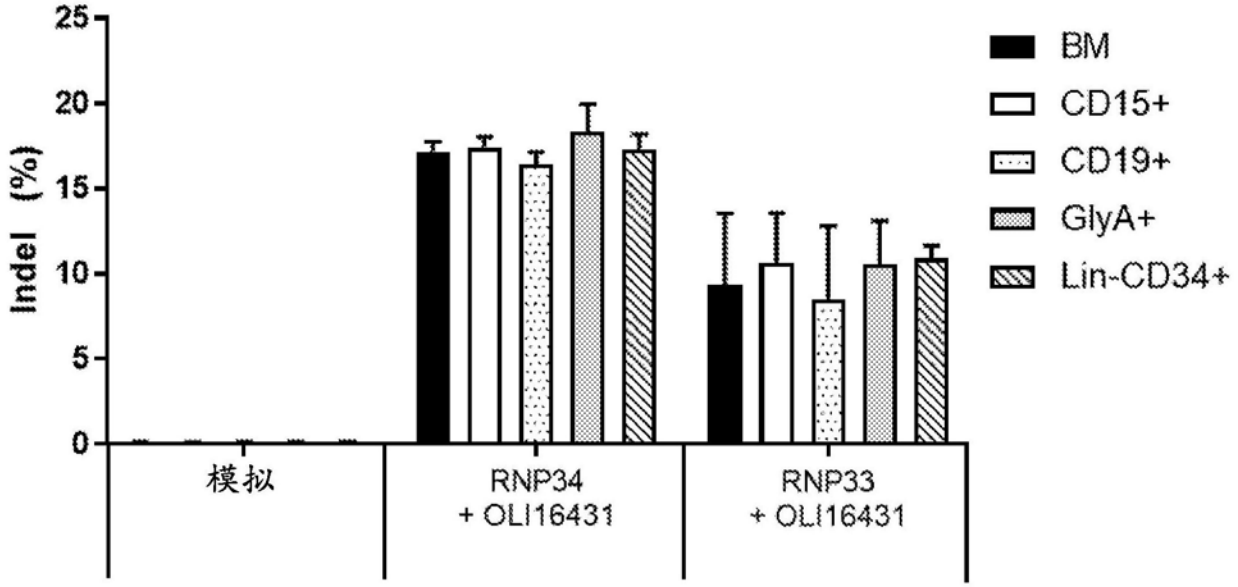


图54

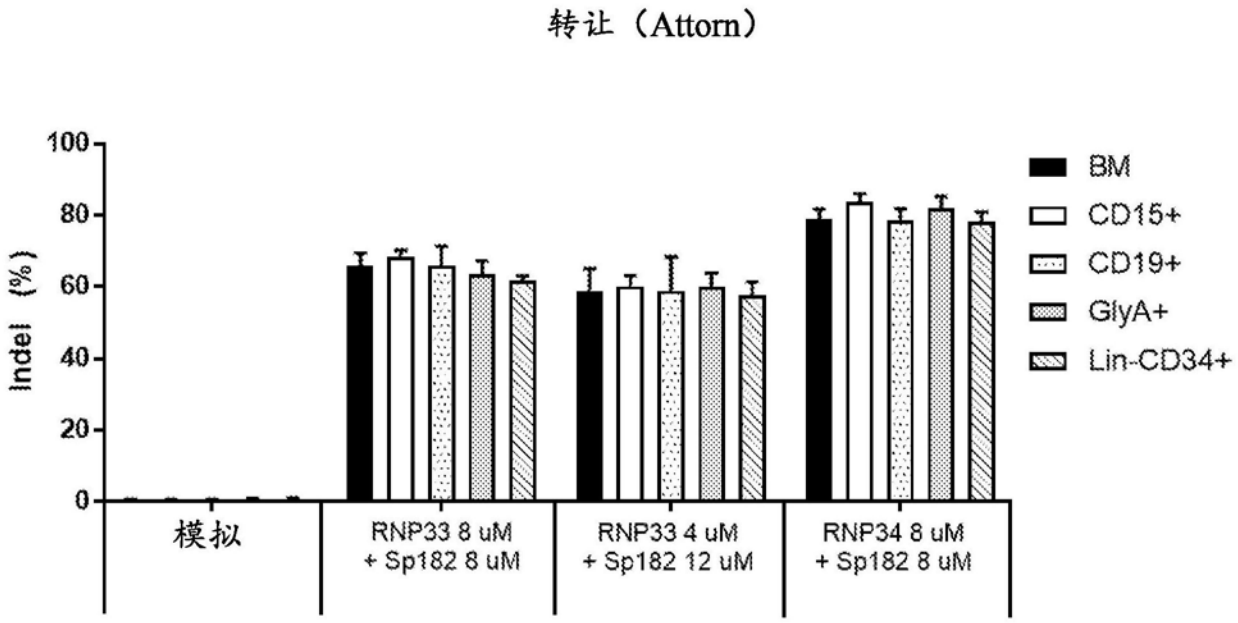

转让（Attorn）
Indel (%)
100
80
60
40
20
0
BM
CD15+
CD19+
GlyA+
Lin-CD34+
模拟
RNP33 8 uM
+ Sp182 8 uM
RNP33 4 uM
+ Sp182 12 uM
RNP34 8 uM
+ Sp182 8 uM

图55A

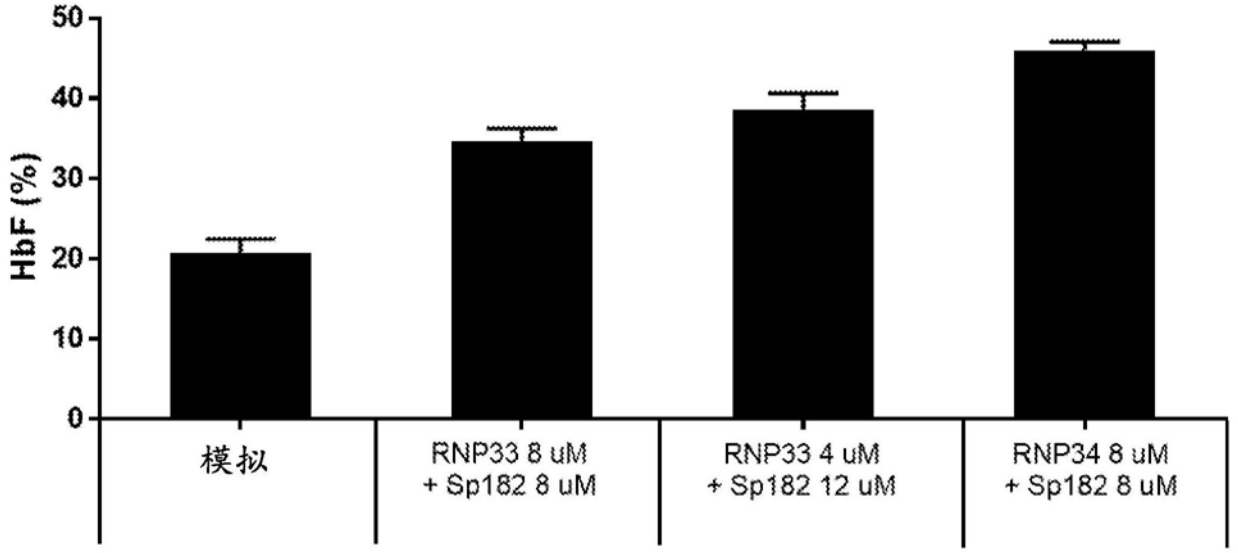

HbF (%)
50
40
30
20
10
0
模拟
RNP33 8 uM
+ Sp182 8 uM
RNP33 4 uM
+ Sp182 12 uM
RNP34 8 uM
+ Sp182 8 uM

图55B

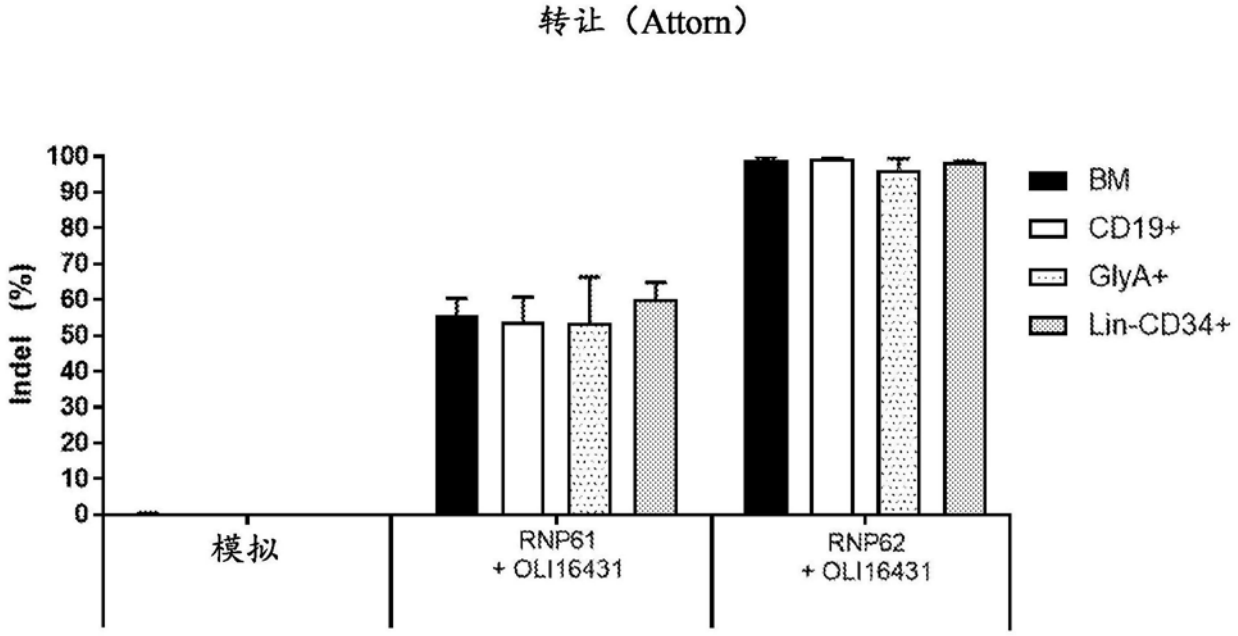

转让（Attorn）
Indel (%)
100
90
80
70
60
50
40
30
20
10
0
模拟
RNP61
+ OLI16431
RNP62
+ OLI16431
BM
CD19+
GlyA+
Lin-CD34+

图56A

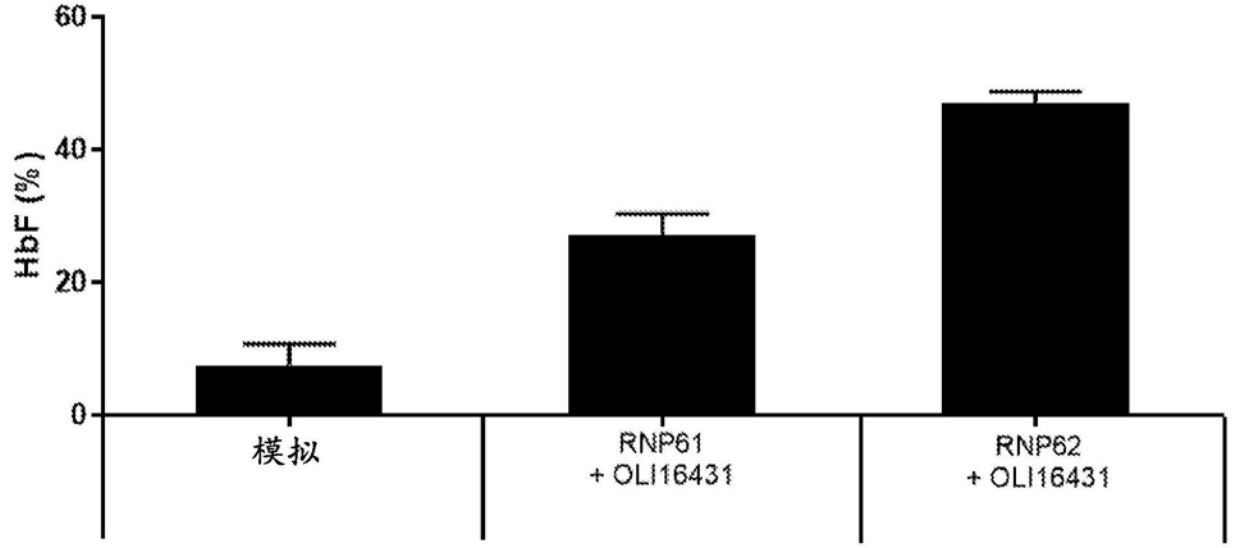

HbF (%)
60
40
20
0
模拟
RNP61
+ OLI16431
RNP62
+ OLI16431

图56B

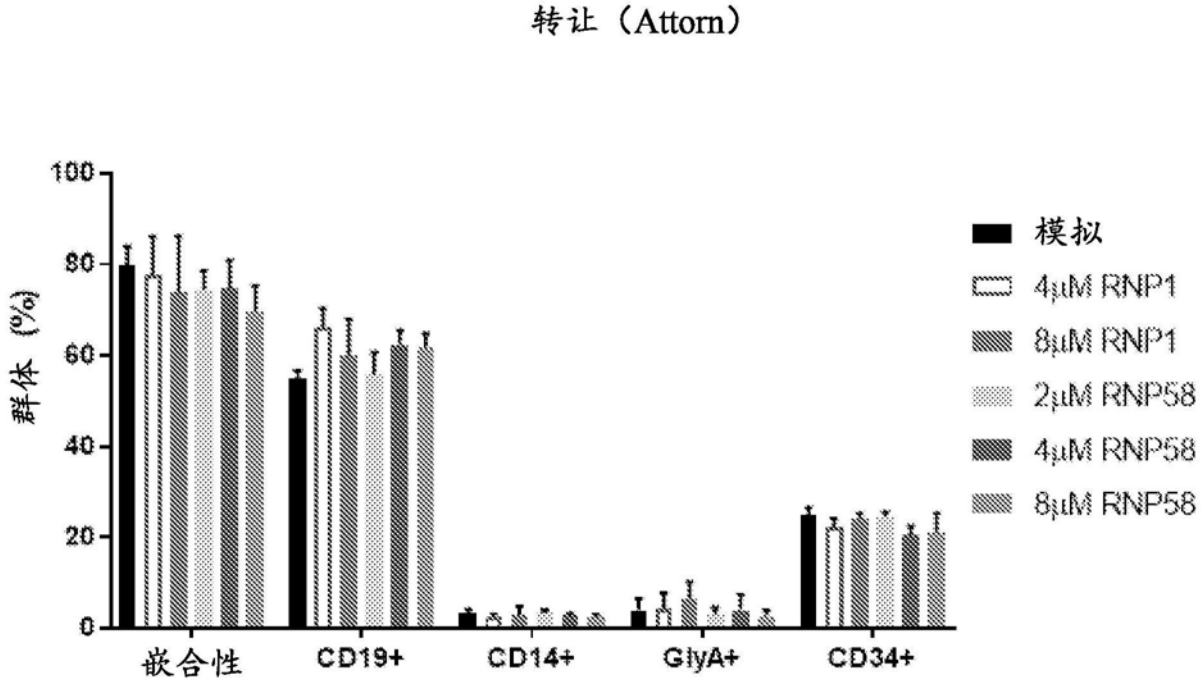

转让（Attorn）
100
80
60
40
20
0
群体（%）
嵌合性
CD19+
CD14+
GlyA+
CD34+
模拟
4μM RNP1
8μM RNP1
2μM RNP58
4μM RNP58
8μM RNP58

图57

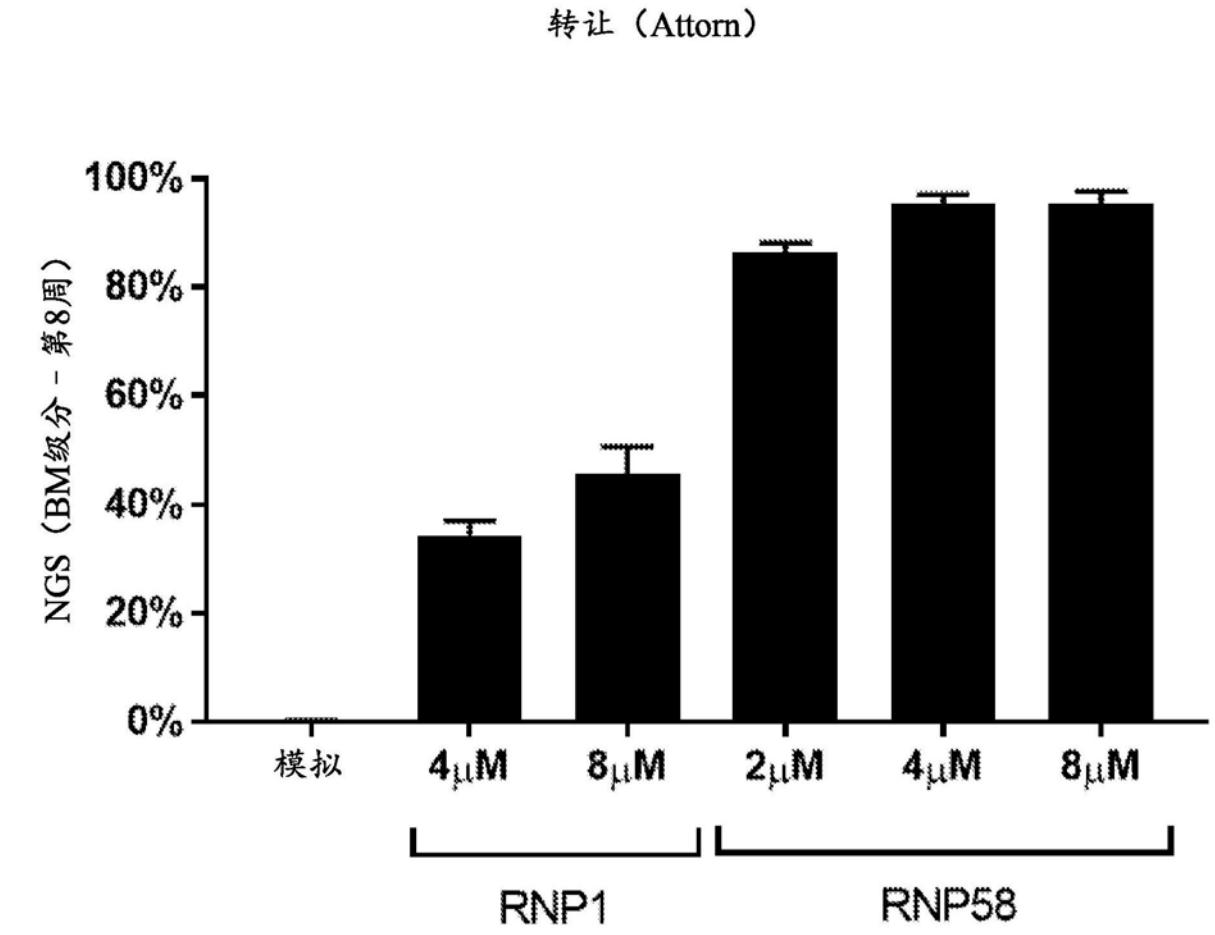

转让（Attorn）
NGS（BM级分 - 第8周）
100%
80%
60%
40%
20%
0%
模拟
4μM
8μM
2μM
4μM
8μM
RNP1
RNP58

图58

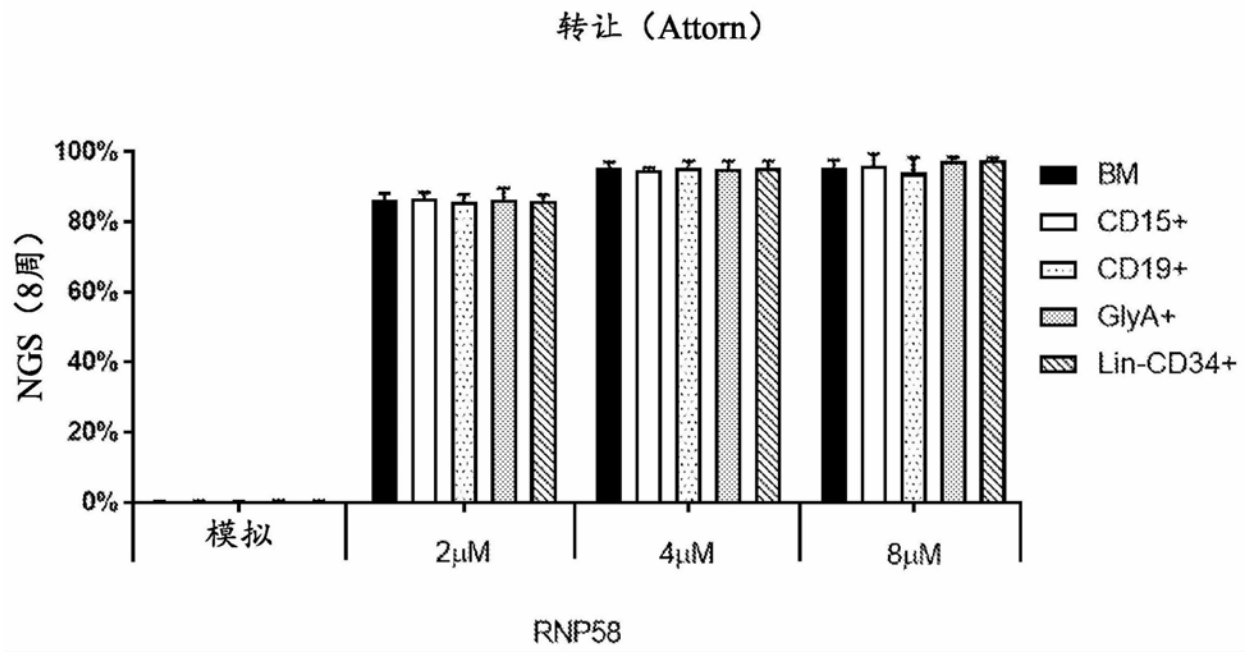

转让（Attorn）
NGS（8周）
100%
80%
60%
40%
20%
0%
模拟
2μM
4μM
8μM
BM
CD15+
CD19+
GlyA+
Lin-CD34+
RNP58

图59

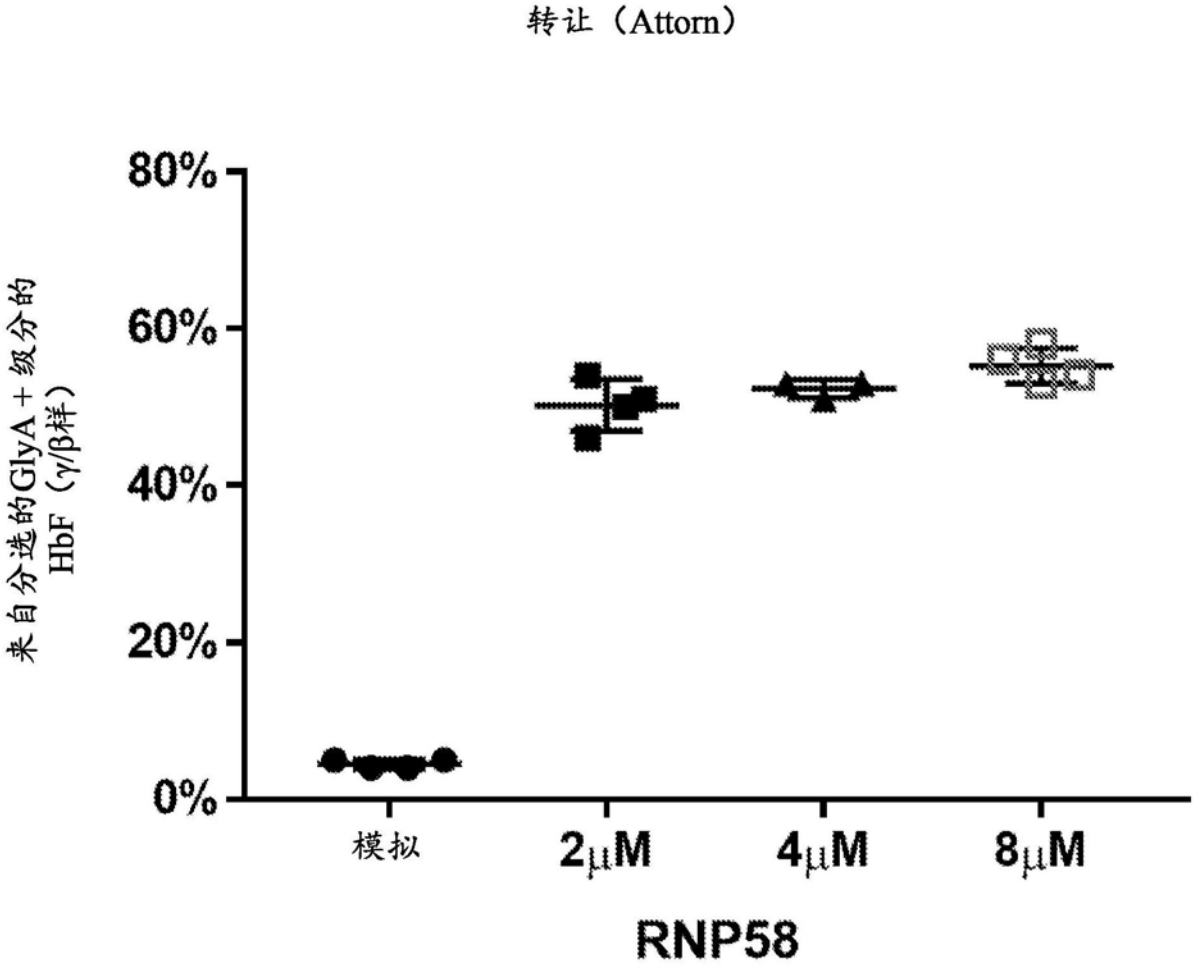

转让（Attorn）
来自分选的GlyA＋级分的HbF（γ/β样）
80%
60%
40%
20%
0%
模拟
2μM
4μM
8μM
RNP58

图60

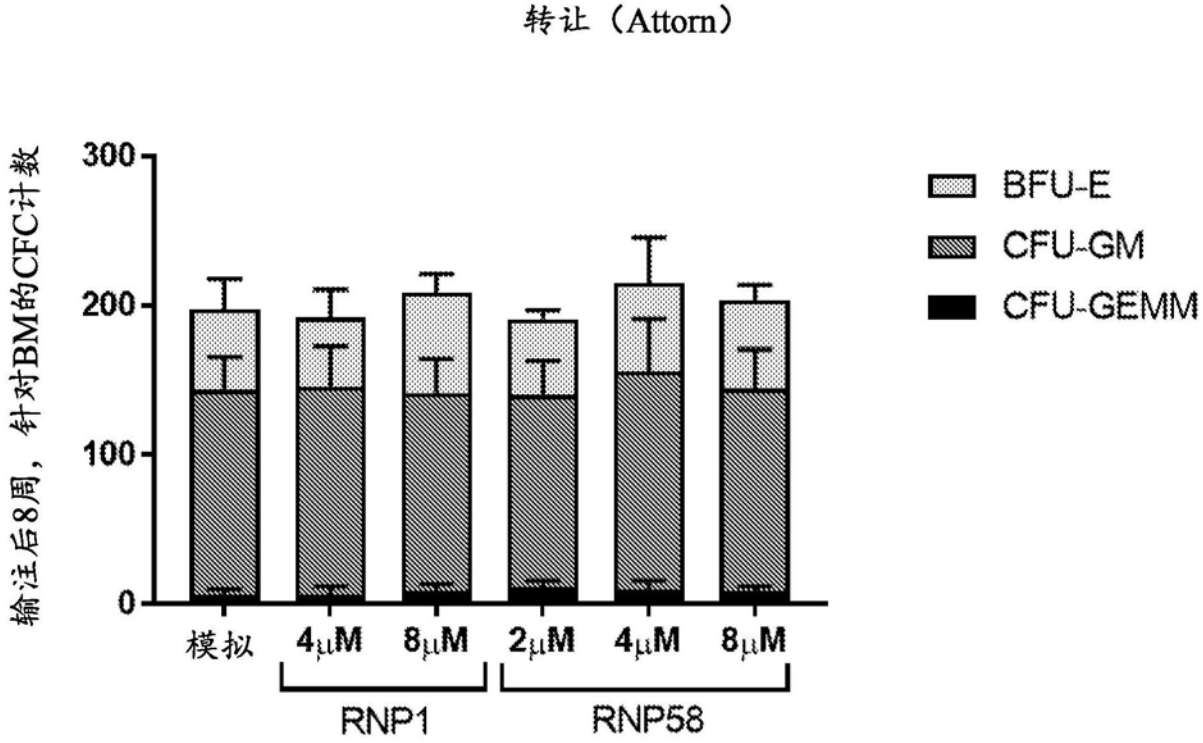

转让（Attorn）
输注后8周，针对BM的CFC计数
300
200
100
0
BFU-E
CFU-GM
CFU-GEMM
模拟
4μM
8μM
2μM
4μM
8μM
RNP1
RNP58

图61

转让（Attorn）

His-AsCpf1-sNLS-sNLS H800A 氨基酸序列 (SEQ ID NO:1032)

MGHHHHHHGSTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTYAD
QCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKA
ELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIF
TRLITAVPSLREHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNL
AIQKNDETAHIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSI
DLTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEIISAAGKELSE
AFKQKTSEILSHAHAALDQPLPTTLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKL
EMEPSLSFYNKARNYATKKPYSVEKFKLNFQMPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGR
YKALSFEPTEKTSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNN
PEKEPKKFQTAYAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLL
YHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYR
PKSRMKRMAARLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKD
RRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQ
QFDYQKKLDNREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFKSKRTGI
AEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPL
TGFVDPFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQF
DAKGTPFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIRS
VLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHLKESKDLKLQ
NGISNQDWLAYIQELRNGS**PKKKRKV**GS**PKKKRKV**

His-AsCpf1-sNLS-sNLS H800A 核苷酸序列 (SEQ ID NO:1110)

ATGGGCCATCATCATCACCATCACGGTAGCACCCAGTTTGAAGGCTTTACCAATCTGTATCAGGTTAGC
AAAACCCTGCGTTTTGAACTGATTCCGCAGGGTAAAACCCTGAAACATATTCAAGAACAGGGCTTCAT
CGAAGAGGATAAAGCACGTAACGATCACTACAAAGAACTGAAACCGATTATCGACCGCATCTATAAA
ACGTACGCAGATCAGTGTCTGCAGCTGGTTCAGCTGGATTGGGAAAATCTGAGCGCAGCAATTGATAG
TTATCGCAAAGAGAAAACCGAAGAAACCCGTAATGCACTGATTGAAGAACAGGCAACCTATCGTAAT
GCCATCCATGATTATTTCATTGGTCGTACCGATAATCTGACCGATGCAATTAACAAACGTCACGCCGA
AATCTATAAAGGCCTGTTTAAAGCCGAACTGTTTAATGGCAAAGTTCTGAAACAGCTGGGCACCGTTA
CCACCACCGAACATGAAAATGCACTGCTGCGTAGCTTTGATAAATTCACCACCTATTTCAGCGGCTTTT
ATGAGAATCGCAAAAACGTGTTTAGCGCAGAAGATATTAGCACCGCAATTCCGCATCGTATTGTGCAG
GATAATTTCCCGAAATTCAAAGAGAACTGCCACATTTTTACCCGTCTGATTACCGCAGTTCCGAGCCTG
CGTGAACATTTTGAAAACGTTAAAAAAGCCATCGGCATCTTTGTTAGCACCAGCATTGAAGAAGTTTT
TAGCTTCCCGTTTTACAATCAGCTGCTGACCCAGACACAGATTGATCTGTATAACCAGCTGTTAGGTGG
TATTAGCCGTGAAGCAGGCACCGAAAAAATCAAAGGTCTGAATGAAGTGCTGAATCTGGCCATTCAG
AAAAATGATGAAACCGCACATATTATTGCAAGCCTGCCGCATCGTTTTATTCCGCTGTTTAAGCAGATT
CTGAGCGATCGTAATACCCTGAGCTTTATTCTGGAAGAATTCAAATCCGATGAAGAGGTGATTCAGAG
CTTCTGCAAATACAAAACGCTGCTGCGCAATGAAAATGTTCTGGAAACTGCCGAAGCACTGTTTAACG
AACTGAATAGCATTGATCTGACCCACATCTTTATCAGCCACAAAAAACTGGAAACCATTTCAAGCGCA
CTGTGTGATCATTGGGATACCCTGCGTAATGCCCTGTATGAACGTCGTATTAGCGAACTGACCGGTAA
AATTACCAAAAGCGCGAAAGAAAAAGTTCAGCGCAGTCTGAAACATGAGGATATTAATCTGCAAGAG
ATTATTAGCGCAGCCGGTAAAGAACTGTCAGAAGCATTTAAACAGAAAACCAGCGAAATTCTGAGCC
ATGCACATGCAGCACTGGATCAGCCGCTGCCGACCACACTGAAAAAACAAGAAGAAAAAGAGATCCT
GAAAAGCCAGCTGGATAGCCTGCTGGGTCTGTATCATCTGCTGGACTGGTTTGCAGTTGATGAAAGCA
ATGAAGTTGATCCGGAATTTAGCGCACGTCTGACCGGCATTAAGCTGGAAATGGAACCGAGCCTGAGT
TTTTACAATAAAGCCCGTAATTATGCCACCAAAAAACCGTATAGCGTGGAAAAATTCAAGCTGAACTT
CCAGATGCCGACACTGGCAAGCGGTTGGGATGTTAATAAAGAAAAGAACAATGGTGCCATCCTGTTCG
TGAAAAATGGCCTGTATTATCTGGGTATCATGCCGAAACAGAAAGGTCGTTATAAAGCACTGAGCTTT
GAACCGACAGAAAAAACCAGTGAAGGCTTCGATAAAATGTACTATGATTACTTTCCGGACGCAGCCA
AAATGATTCCGAAATGCAGCACCCAGCTGAAAGCAGTTACCGCACATTTTCAGACCCATACCACACCG
ATTCTGCTGAGCAATAACTTTATTGAACCGCTGGAAATCACCAAAGAGATCTACGATCTGAATAACCC

转让（Attorn）

GGAAAAAGAGCCGAAAAAGTTTCAGACCGCCTATGCAAAAAAAACCGGTGATCAGAAAGGTTATCGT
GAAGCGCTGTGTAAATGGATTGATTTCACCCGTGATTTTCTGAGCAAATACACCAAAACCACCAGTAT
CGATCTGAGCAGCCTGCGTCCGAGCAGCCAGTATAAAGATCTGGGCGAATATTATGCAGAACTGAATC
CGCTGCTGTATCATATTAGCTTTCAGCGTATTGCCGAGAAAGAAATAATGGATGCAGTTGAAACCGGT
AAGCTGTACCTGTTCCAGATTTATAACAAAGACTTTGCCAAAGGCCATCATGGTAAACCGAATCTGCA
TACCCTGTATTGGACCGGTCTGTTTAGCCCTGAAAATCTGGCAAAAACCAGCATTAAACTGAATGGTC
AGGCGGAACTGTTTTATCGTCCGAAAAGCCGTATGAAACGTATGGCAGCACGTCTGGGTGAAAAAATG
CTGAATAAAAAGCTGAAGGATCAGAAAACCCCGATTCCGGATACACTGTATCAAGAGCTGTATGATTA
TGTGAATCATCGTCTGAGTCATGATCTGAGTGATGAAGCACGTGCCCTGCTGCCGAATGTTATTACCAA
AGAAGTTAGCCACGAGATCATTAAAGATCGTCGTTTTACCAGCGACAAGTTTTTTTTCCATGTTCCGAT
TACGCTGAATTATCAGGCAGCAAATAGCCCGAGCAAATTTAACCAGCGTGTGAATGCATATCTGAAAG
AACATCCGGAAACACCGATTATTGGTATTGATCGTGGTGAACGCAACCTGATTTATATCACCGTTATTG
ATAGCACCGGCAAGATCCTGGAACAGCGTAGCCTGAATACCATTCAGCAGTTTGATTATCAGAAAAAG
CTGGACAACCGTGAAAAAGAACGTGTTGCAGCCCGTCAGGCATGGTCAGTTGTTGGTACAATTAAAGA
CCTGAAACAGGGTTATCTGAGCCAGGTTATTCATGAAATTGTGGATCTGATGATTCATTACCAGGCAG
TTGTTGTGCTGGAAAACCTGAATTTTGGCTTTAAAAGCAAACGTACCGGCATTGCGGAAAAAGCAGTT
TATCAGCAGTTCGAAAAGATGCTGATCGATAAACTGAACTGTCTGGTGCTGAAAGATTATCCGGCTGA
AAAAGTTGGCGGTGTTCTGAATCCGTATCAGCTGACCGATCAGTTTACCAGCTTTGCAAAAATGGGCA
CCCAGAGCGGATTTCTGTTTTATGTTCCGGCACCGTATACGAGCAAAATTGATCCGCTGACCGGTTTTG
TTGATCCGTTTGTTTGGAAAACCATCAAAAACCATGAAAGCCGCAAACATTTTCTGGAAGGTTTTGATT
TCCTGCATTACGACGTTAAGACCGGTGATTTTATTCTGCACTTTAAGATGAATCGGAACCTGTCATTTC
AGCGTGGTCTGCCTGGTTTTATGCCTGCATGGGATATTGTGTTTGAGAAAAACGAAACACAGTTCGAT
GCAAAAGGCACCCCGTTTATTGCAGGTAAACGTATTGTTCCGGTGATTGAAAATCATCGTTTCACCGGT
CGTTATCGCGATCTGTATCCGGCAAATGAACTGATCGCACTGCTGGAAGAGAAAGGTATTGTTTTTCGT
GATGGCTCAAACATTCTGCCGAAACTGTTAGAGAATGATGATAGCCATGCCATTGATACCATGGTTGC
CCTGATTCGTAGCGTTCTGCAGATGCGTAATAGCAATGCAGCAACCGGTGAAGATTACATTAATAGTC
CGGTTCGTGATCTGAATGGTGTTTGTTTTGATAGCCGTTTTCAGAATCCGGAATGGCCGATGGATGCGG
ATGCAAATGGTGCATATCATATTGCACTGAAAGGACAGCTGCTGCTGAACCACCTGAAAGAAAGTAA
AGATTTGAAGCTGCAGAATGGCATTAGCAATCAGGATTGGCTGGCATATATCCAAGAACTGCGTAACG
GTTCT**CCTAAGAAAAAGCGCAAAGTT**GGCAGT**CCGAAGAAGAAACGTAAGGTT**

Cpf1-1 氨基酸序列 (SEQ ID NO:1094)

MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTYADQCLQLVQLD
WENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKAELFNGKVLK
QLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSL
REHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETA
HIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSIDLTHIFISHK
KLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEIL
SHAHAALDQPLPTTLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYN
KARNYATKKPYSVEKFKLNFQRPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEK
TSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTA
YAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEI
MDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAH
RLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFLFH
VPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDN
REKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFKSKRTGIAEKAVYQQFE
KMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWK
TIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIAG
KRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIRSVLQMRNSNA
ATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWL
AYIQELRNGR**SSDDEATADSQHAAPPKKKRKV**GGSGGSGGSGGSGGSGGSGGSGGSLE<u>HHHHHH</u>

Cpf1-1 核苷酸序列 (SEQ ID NO:1111)

转让（Attorn）

ATGACCCAGTTTGAAGGTTTCACCAATCTGTATCAGGTTAGCAAAACCCTGCGTTTTGAACTGATTCCG
CAGGGTAAAACCCTGAAACATATTCAAGAACAGGGCTTCATCGAAGAGGATAAAGCACGTAACGATC
ACTACAAAGAACTGAAACCGATTATCGACCGCATCTATAAAACCTATGCAGATCAGTGTCTGCAGCTG
GTTCAGCTGGATTGGGAAAATCTGAGCGCAGCAATTGATAGTTATCGCAAAGAAAAAACCGAAGAAA
CCCGTAATGCACTGATTGAAGAACAGGCAACCTATCGTAATGCCATCCATGATTATTTCATTGGTCGTA
CCGATAATCTGACCGATGCAATTAACAAACGTCACGCCGAAATCTATAAAGGCCTGTTTAAAGCCGAA
CTGTTTAATGGCAAAGTTCTGAAACAGCTGGGCACCGTTACCACCACCGAACATGAAAATGCACTGCT
GCGTAGCTTTGATAAATTCACCACCTATTTCAGCGGCTTTTATGAGAATCGCAAAAACGTGTTTAGCGC
AGAAGATATTAGCACCGCAATTCCGCATCGTATTGTGCAGGATAATTTCCCGAAATTCAAAGAGAACT
GCCACATTTTTACCCGTCTGATTACCGCAGTTCCGAGCCTGCGTGAACATTTTGAAAACGTTAAAAAAG
CCATCGGCATCTTTGTTAGCACCAGCATTGAAGAAGTTTTTAGCTTCCCGTTTTACAATCAGCTGCTGA
CCCAGACCCAGATTGATCTGTATAACCAACTGCTGGGTGGTATTAGCCGTGAAGCAGGCACCGAAAAA
ATCAAAGGTCTGAATGAAGTGCTGAATCTGGCCATTCAGAAAAATGATGAAACCGCACATATTATTGC
AAGCCTGCCGCATCGTTTTATTCCGCTGTTCAAACAAATTCTGAGCGATCGTAATACCCTGAGCTTTAT
TCTGGAAGAATTCAAATCCGATGAAGAGGTGATTCAGAGCTTTTGCAAATACAAAACGCTGCTGCGCA
ATGAAAATGTTCTGGAAACTGCCGAAGCACTGTTTAACGAACTGAATAGCATTGATCTGACCCACATC
TTTATCAGCCACAAAAAACTGGAAACCATTTCAAGCGCACTGTGTGATCATTGGGATACCCTGCGTAA
TGCCCTGTATGAACGTCGTATTAGCGAACTGACCGGTAAAATTACCAAAAGCGCGAAAGAAAAAGTTC
AGCGCAGTCTGAAACATGAGGATATTAATCTGCAAGAGATTATTAGCGCAGCCGGTAAAGAACTGTCA
GAAGCATTTAAACAGAAAACCAGCGAAATTCTGTCACATGCACATGCAGCACTGGATCAGCCGCTGCC
GACCACCCTGAAAAAACAAGAAGAAAAAGAAATCCTGAAAAGCCAGCTGGATAGCCTGCTGGGTCTG
TATCATCTGCTGGACTGGTTTGCAGTTGATGAAAGCAATGAAGTTGATCCGGAATTTAGCGCACGTCT
GACCGGCATTAAACTGGAAATGGAACCGAGCCTGAGCTTTTATAACAAAGCCCGTAATTATGCCACCA
AAAAACCGTATAGCGTCGAAAAATTCAAACTGAACTTTCAGCGTCCGACCCTGGCAAGCGGTTGGGAT
GTTAATAAAGAAAAAAACAACGGTGCCATCCTGTTCGTGAAAAATGGCCTGTATTATCTGGGTATTAT
GCCGAAACAGAAAGGTCGTTATAAAGCGCTGAGCTTTGAACCGACGGAAAAAACCAGTGAAGGTTTT
GATAAAATGTACTACGACTATTTTCCGGATGCAGCCAAAATGATTCCGAAATGTAGCACCCAGCTGAA
AGCAGTTACCGCACATTTTCAGACCCATACCACCCCGATTCTGCTGAGCAATAACTTTATTGAACCGCT
GGAAATCACCAAAGAGATCTACGATCTGAATAACCCGGAAAAAGAGCCGAAAAAATTCCAGACCGCA
TATGCAAAAAAAACCGGTGATCAGAAAGGTTATCGTGAAGCGCTGTGTAAATGGATTGATTTCACCCG
TGATTTTCTGAGCAAATACACCAAAACCACCAGTATCGATCTGAGCAGCCTGCGTCCGAGCAGCCAGT
ATAAAGATCTGGGCGAATATTATGCAGAACTGAATCCGCTGCTGTATCATATTAGCTTTCAGCGTATTG
CCGAGAAAGAAATCATGGACGCAGTTGAAACCGGTAAACTGTACCTGTTCCAGATCTACAATAAAGAT
TTTGCCAAAGGCCATCATGGCAAACCGAATCTGCATACCCTGTATTGGACCGGTCTGTTTAGCCCTGAA
AATCTGGCAAAAACCTCGATTAAACTGAATGGTCAGGCGGAACTGTTTTATCGTCCGAAAAGCCGTAT
GAAACGTATGGCACATCGTCTGGGTGAAAAAATGCTGAACAAAAAACTGAAAGACCAGAAAACCCCG
ATCCCGGATACACTGTATCAAGAACTGTATGATTATGTGAACCATCGTCTGAGCCATGATCTGAGTGA
TGAAGCACGTGCCCTGCTGCCGAATGTTATTACCAAAGAAGTTAGCCACGAGATCATTAAAGATCGTC
GTTTTACCAGCGACAAATTCCTGTTTCATGTGCCGATTACCCTGAATTATCAGGCAGCAAATAGCCCGA
GCAAATTTAACCAGCGTGTTAATGCATATCTGAAAGAACATCCAGAAACGCCGATTATTGGTATTGAT
CGTGGTGAACGTAACCTGATTTATATCACCGTTATTGATAGCACCGGCAAAATCCTGGAACAGCGTAG
CCTGAATACCATTCAGCAGTTTGATTACCAGAAAAAACTGGATAATCGCGAGAAAGAACGTGTTGCAG
CACGTCAGGCATGGTCAGTTGTTGGTACAATTAAAGACCTGAAACAGGGTTATCTGAGCCAGGTTATT
CATGAAATTGTGGATCTGATGATTCACTATCAGGCCGTTGTTGTGCTGGAAAACCTGAATTTTGGCTTT
AAAAGCAAACGTACCGGCATTGCAGAAAAAGCAGTTTATCAGCAGTTCGAGAAAATGCTGATTGACA
AACTGAATTGCCTGGTGCTGAAAGATTATCCGGCTGAAAAAGTTGGTGGTGTTCTGAATCCGTATCAG
CTGACCGATCAGTTTACCAGCTTTGCAAAAATGGGCACCCAGAGCGGATTTCTGTTTTATGTTCCGGCA
CCGTATACGAGCAAAATTGATCCGCTGACCGGTTTTGTTGATCCGTTTGTTTGGAAAACCATCAAAAAC
CATGAAAGCCGCAAACATTTTCTGGAAGGTTTCGATTTTCTGCATTACGACGTTAAAACGGGTGATTTC
ATCCTGCACTTTAAAATGAATCGCAATCTGAGTTTTCAGCGTGGCCTGCCTGGTTTTATGCCTGCATGG
GATATTGTGTTTGAGAAAAACGAAACACAGTTCGATGCAAAAGGCACCCCGTTTATTGCAGGTAAACG
TATTGTTCCGGTGATTGAAAATCATCGTTTCACCGGTCGTTATCGCGATCTGTATCCGGCAAATGAACT
GATCGCACTGCTGGAAGAGAAAGGTATTGTTTTTCGTGATGGCTCAAACATTCTGCCGAAACTGCTGG
AAAATGATGATAGCCATGCAATTGATACCATGGTTGCACTGATTCGTAGCGTTCTGCAGATGCGTAAT
AGCAATGCAGCAACCGGTGAAGATTACATTAATAGTCCGGTTCGTGATCTGAATGGTGTTTGTTTTGAT
AGCCGTTTTCAGAATCCGGAATGGCCGATGGATGCAGATGCAAATGGTGCATATCATATTGCACTGAA

转让（Attorn）

AGGACAGCTGCTGCTGAACCACCTGAAAGAAAGCAAAGATCTGAAACTGCAAAACGGCATTAGCAAT
CAGGATTGGCTGGCATATATCCAAGAACTGCGTAACGGTCGT**AGCAGTGATGATGAAGCAACCGCA**
GATAGCCAGCATGCAGCACCGCCTAAAAAGAAACGTAAAGTTGGTGGTAGCGGTGGTTCAGGTGG
TAGTGGCGGTAGTGGTGGCTCAGGGGGTTCTGGTGGCTCTGGTGGTAGCCTCGAGCACCACCACCACC
ACCAC

Cpf1-2 氨基酸序列 (SEQ ID NO:1095)

MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTYADQCLQLVQLD
WENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKAELFNGKVLK
QLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSL
REHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETA
HIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSIDLTHIFISHK
KLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEIL
SHAHAALDQPLPTTLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYN
KARNYATKKPYSVEKFKLNFQMPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTE
KTSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQT
AYAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEK
EIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMA
HRLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFFF
HVPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLD
NREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFKSKRTGIAEKAVYQQF
EKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVW
KTIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIA
GKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIRSVLQMRNSN
AATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDW
LAYIQELRNGR**SSDDEATADSQHAAPPKKKRKV**GGSGGSGGSGGSGGSGGSGGSGGSLEHHHHHH

Cpf1-2 核苷酸序列 (SEQ ID NO:1112)

ATGACCCAGTTTGAAGGTTTCACCAATCTGTATCAGGTTAGCAAAACCCTGCGTTTTGAACTGATTCCG
CAGGGTAAAACCCTGAAACATATTCAAGAACAGGGCTTCATCGAAGAGGATAAAGCACGTAACGATC
ACTACAAAGAACTGAAACCGATTATCGACCGCATCTATAAAACCTATGCAGATCAGTGTCTGCAGCTG
GTTCAGCTGGATTGGGAAAATCTGAGCGCAGCAATTGATAGTTATCGCAAAGAAAAAACCGAAGAAA
CCCGTAATGCACTGATTGAAGAACAGGCAACCTATCGTAATGCCATCCATGATTATTTCATTGGTCGTA
CCGATAATCTGACCGATGCAATTAACAAACGTCACGCCGAAATCTATAAAGGCCTGTTTAAAGCCGAA
CTGTTTAATGGCAAAGTTCTGAAACAGCTGGGCACCGTTACCACCACCGAACATGAAAATGCACTGCT
GCGTAGCTTTGATAAATTCACCACCTATTTCAGCGGCTTTTATGAGAATCGCAAAAACGTGTTTAGCGC
AGAAGATATTAGCACCGCAATTCCGCATCGTATTGTGCAGGATAATTTCCCGAAATTCAAAGAGAACT
GCCACATTTTTACCCGTCTGATTACCGCAGTTCCGAGCCTGCGTGAACATTTTGAAAACGTTAAAAAAG
CCATCGGCATCTTTGTTAGCACCAGCATTGAAGAAGTTTTTAGCTTCCCGTTTTACAATCAGCTGCTGA
CCCAGACCCAGATTGATCTGTATAACCAACTGCTGGGTGGTATTAGCCGTGAAGCAGGCACCGAAAAA
ATCAAAGGTCTGAATGAAGTGCTGAATCTGGCCATTCAGAAAAATGATGAAACCGCACATATTATTGC
AAGCCTGCCGCATCGTTTTATTCCGCTGTTCAAACAAATTCTGAGCGATCGTAATACCCTGAGCTTTAT
TCTGGAAGAATTCAAATCCGATGAAGAGGTGATTCAGAGCTTTTGCAAATACAAAACGCTGCTGCGCA
ATGAAAATGTTCTGGAAACTGCCGAAGCACTGTTTAACGAACTGAATAGCATTGATCTGACCCACATC
TTTATCAGCCACAAAAAACTGGAAACCATTTCAAGCGCACTGTGTGATCATTGGGATACCCTGCGTAA
TGCCCTGTATGAACGTCGTATTAGCGAACTGACCGGTAAAATTACCAAAAGCGCGAAAGAAAAAGTTC
AGCGCAGTCTGAAACATGAGGATATTAATCTGCAAGAGATTATTAGCGCAGCCGGTAAAGAACTGTCA
GAAGCATTTAAACAGAAAACCAGCGAAATTCTGTCACATGCACATGCAGCACTGGATCAGCCGCTGCC
GACCACCCTGAAAAAACAAGAAGAAAAAGAAATCCTGAAAAGCCAGCTGGATAGCCTGCTGGGTCTG
TATCATCTGCTGGACTGGTTTGCAGTTGATGAAAGCAATGAAGTTGATCCGGAATTTAGCGCACGTCT
GACCGGCATTAAACTGGAAATGGAACCGAGCCTGAGCTTTTATAACAAAGCCCGTAATTATGCCACCA
AAAAACCGTATAGCGTCGAAAAATTCAAACTGAACTTTCAGATGCCGACCCTGGCAAGCGGTTGGGAT
GTTAATAAAGAAAAAAACAACGGTGCCATCCTGTTCGTGAAAAATGGCCTGTATTATCTGGGTATTAT
GCCGAAACAGAAAGGTCGTTATAAAGCGCTGAGCTTTGAACCGACGGAAAAAACCAGTGAAGGTTTT

转让（Attorn）

GATAAAATGTACTACGACTATTTTCCGGATGCAGCCAAAATGATTCCGAAATGTAGCACCCAGCTGAA
AGCAGTTACCGCACATTTTCAGACCCATACCACCCCGATTCTGCTGAGCAATAACTTTATTGAACCGCT
GGAAATCACCAAAGAGATCTACGATCTGAATAACCCGGAAAAAGAGCCGAAAAAATTCCAGACCGCA
TATGCAAAAAAAACCGGTGATCAGAAAGGTTATCGTGAAGCGCTGTGTAAATGGATTGATTTCACCCG
TGATTTTCTGAGCAAATACACCAAAACCACCAGTATCGATCTGAGCAGCCTGCGTCCGAGCAGCCAGT
ATAAAGATCTGGGCGAATATTATGCAGAACTGAATCCGCTGCTGTATCATATTAGCTTTCAGCGTATTG
CCGAGAAAGAAATCATGGACGCAGTTGAAACCGGTAAACTGTACCTGTTCCAGATCTACAATAAAGAT
TTTGCCAAAGGCCATCATGGCAAACCGAATCTGCATACCCTGTATTGGACCGGTCTGTTTAGCCCTGAA
AATCTGGCAAAAACCTCGATTAAACTGAATGGTCAGGCGGAACTGTTTTATCGTCCGAAAAGCCGTAT
GAAACGTATGGCACATCGTCTGGGTGAAAAAATGCTGAACAAAAAACTGAAAGACCAGAAAACCCCG
ATCCCGGATACACTGTATCAAGAACTGTATGATTATGTGAACCATCGTCTGAGCCATGATCTGAGTGA
TGAAGCACGTGCCCTGCTGCCGAATGTTATTACCAAAGAAGTTAGCCACGAGATCATTAAAGATCGTC
GTTTTACCAGCGACAAATTCTTTTTTCATGTGCCGATTACCCTGAATTATCAGGCAGCAAATAGCCCGA
GCAAATTTAACCAGCGTGTTAATGCATATCTGAAAGAACATCCAGAAACGCCGATTATTGGTATTGAT
CGTGGTGAACGTAACCTGATTTATATCACCGTTATTGATAGCACCGGCAAAATCCTGGAACAGCGTAG
CCTGAATACCATTCAGCAGTTTGATTACCAGAAAAAACTGGATAATCGCGAGAAAGAACGTGTTGCAG
CACGTCAGGCATGGTCAGTTGTTGGTACAATTAAAGACCTGAAACAGGGTTATCTGAGCCAGGTTATT
CATGAAATTGTGGATCTGATGATTCACTATCAGGCCGTTGTTGTGCTGGAAAACCTGAATTTTGGCTTT
AAAAGCAAACGTACCGGCATTGCAGAAAAAGCAGTTTATCAGCAGTTCGAGAAAATGCTGATTGACA
AACTGAATTGCCTGGTGCTGAAAGATTATCCGGCTGAAAAAGTTGGTGGTGTTCTGAATCCGTATCAG
CTGACCGATCAGTTTACCAGCTTTGCAAAAATGGGCACCCAGAGCGGATTTCTGTTTTATGTTCCGGCA
CCGTATACGAGCAAAATTGATCCGCTGACCGGTTTTGTTGATCCGTTTGTTTGGAAAACCATCAAAAAC
CATGAAAGCCGCAAACATTTTCTGGAAGGTTTCGATTTTCTGCATTACGACGTTAAAACGGGTGATTTC
ATCCTGCACTTTAAAATGAATCGCAATCTGAGTTTTCAGCGTGGCCTGCCTGGTTTTATGCCTGCATGG
GATATTGTGTTTGAGAAAAACGAAACACAGTTCGATGCAAAAGGCACCCCGTTTATTGCAGGTAAACG
TATTGTTCCGGTGATTGAAAATCATCGTTTCACCGGTCGTTATCGCGATCTGTATCCGGCAAATGAACT
GATCGCACTGCTGGAAGAGAAAGGTATTGTTTTTCGTGATGGCTCAAACATTCTGCCGAAACTGCTGG
AAAATGATGATAGCCATGCAATTGATACCATGGTTGCACTGATTCGTAGCGTTCTGCAGATGCGTAAT
AGCAATGCAGCAACCGGTGAAGATTACATTAATAGTCCGGTTCGTGATCTGAATGGTGTTTGTTTTGAT
AGCCGTTTTCAGAATCCGGAATGGCCGATGGATGCAGATGCAAATGGTGCATATCATATTGCACTGAA
AGGACAGCTGCTGCTGAACCACCTGAAAGAAAGCAAAGATCTGAAACTGCAAAACGGCATTAGCAAT
CAGGATTGGCTGGCATATATCCAAGAACTGCGTAACGGTCGT**AGCAGTGATGATGAAGCAACCGCA**
GATAGCCAGCATGCAGCACCGCCTAAAAAGAAACGTAAAGTTGGTGGTAGCGGTGGTTCAGGTGG
TAGTGGCGGTAGTGGTGGCTCAGGGGGTTCTGGTGGCTCTGGTGGTAGCCTCGAGCACCACCACCACC
ACCAC

Cpf1-3 氨基酸序列 (SEQ ID NO:1096)

MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTYADQCLQLVQLD
WENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKAELFNGKVLK
QLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSL
REHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETA
HIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSIDLTHIFISHK
KLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEIL
SHAHAALDQPLPTTLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYN
KARNYATKKPYSVEKFKLNFQRPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEK
TSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTA
YAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEI
MDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAA
RLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFLFH
VPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDN
REKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFKSKRTGIAEKAVYQQFE
KMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWK
TIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIAG
KRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIRSVLQMRNSNA
ATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWL
AYIQELRNGR**SSDDEATADSQHAAPPKKKRKV**GGSGGSGGSGGSGGSGGSGGSGGSLEHHHHHH

转让（Attorn）

Cpf1-3 核苷酸序列 (SEQ ID NO:1113)

ATGACCCAGTTTGAAGGTTTCACCAATCTGTATCAGGTTAGCAAAACCCTGCGTTTTGAACTGATTCCG
CAGGGTAAAACCCTGAAACATATTCAAGAACAGGGCTTCATCGAAGAGGATAAAGCACGTAACGATC
ACTACAAAGAACTGAAACCGATTATCGACCGCATCTATAAAACCTATGCAGATCAGTGTCTGCAGCTG
GTTCAGCTGGATTGGGAAAATCTGAGCGCAGCAATTGATAGTTATCGCAAAGAAAAAACCGAAGAAA
CCCGTAATGCACTGATTGAAGAACAGGCAACCTATCGTAATGCCATCCATGATTATTTCATTGGTCGTA
CCGATAATCTGACCGATGCAATTAACAAACGTCACGCCGAAATCTATAAAGGCCTGTTTAAAGCCGAA
CTGTTTAATGGCAAAGTTCTGAAACAGCTGGGCACCGTTACCACCACCGAACATGAAAATGCACTGCT
GCGTAGCTTTGATAAATTCACCACCTATTTCAGCGGCTTTTATGAGAATCGCAAAAACGTGTTTAGCGC
AGAAGATATTAGCACCGCAATTCCGCATCGTATTGTGCAGGATAATTTCCCGAAATTCAAAGAGAACT
GCCACATTTTTACCCGTCTGATTACCGCAGTTCCGAGCCTGCGTGAACATTTTGAAAACGTTAAAAAAG
CCATCGGCATCTTTGTTAGCACCAGCATTGAAGAAGTTTTTAGCTTCCCGTTTTACAATCAGCTGCTGA
CCCAGACCCAGATTGATCTGTATAACCAACTGCTGGGTGGTATTAGCCGTGAAGCAGGCACCGAAAAA
ATCAAAGGTCTGAATGAAGTGCTGAATCTGGCCATTCAGAAAAATGATGAAACCGCACATATTATTGC
AAGCCTGCCGCATCGTTTTATTCCGCTGTTCAAACAAATTCTGAGCGATCGTAATACCCTGAGCTTTAT
TCTGGAAGAATTCAAATCCGATGAAGAGGTGATTCAGAGCTTTTGCAAATACAAAACGCTGCTGCGCA
ATGAAAATGTTCTGGAAACTGCCGAAGCACTGTTTAACGAACTGAATAGCATTGATCTGACCCACATC
TTTATCAGCCACAAAAAACTGGAAACCATTTCAAGCGCACTGTGTGATCATTGGGATACCCTGCGTAA
TGCCCTGTATGAACGTCGTATTAGCGAACTGACCGGTAAAATTACCAAAAGCGCGAAAGAAAAAGTTC
AGCGCAGTCTGAAACATGAGGATATTAATCTGCAAGAGATTATTAGCGCAGCCGGTAAAGAACTGTCA
GAAGCATTTAAACAGAAAACCAGCGAAATTCTGTCACATGCACATGCAGCACTGGATCAGCCGCTGCC
GACCACCCTGAAAAAACAAGAAGAAAAAGAAATCCTGAAAAGCCAGCTGGATAGCCTGCTGGGTCTG
TATCATCTGCTGGACTGGTTTGCAGTTGATGAAAGCAATGAAGTTGATCCGGAATTTAGCGCACGTCT
GACCGGCATTAAACTGGAAATGGAACCGAGCCTGAGCTTTTATAACAAAGCCCGTAATTATGCCACCA
AAAAACCGTATAGCGTCGAAAAATTCAAACTGAACTTTCAGCGTCCGACCCTGGCAAGCGGTTGGGAT
GTTAATAAAGAAAAAAACAACGGTGCCATCCTGTTCGTGAAAAATGGCCTGTATTATCTGGGTATTAT
GCCGAAACAGAAAGGTCGTTATAAAGCGCTGAGCTTTGAACCGACGGAAAAAACCAGTGAAGGTTTT
GATAAAATGTACTACGACTATTTTCCGGATGCAGCCAAAATGATTCCGAAATGTAGCACCCAGCTGAA
AGCAGTTACCGCACATTTTCAGACCCATACCACCCCGATTCTGCTGAGCAATAACTTTATTGAACCGCT
GGAAATCACCAAAGAGATCTACGATCTGAATAACCCGGAAAAAGAGCCGAAAAAATTCCAGACCGCA
TATGCAAAAAAAACCGGTGATCAGAAAGGTTATCGTGAAGCGCTGTGTAAATGGATTGATTTCACCCG
TGATTTTCTGAGCAAATACACCAAAACCACCAGTATCGATCTGAGCAGCCTGCGTCCGAGCAGCCAGT
ATAAAGATCTGGGCGAATATTATGCAGAACTGAATCCGCTGCTGTATCATATTAGCTTTCAGCGTATTG
CCGAGAAAGAAATCATGGACGCAGTTGAAACCGGTAAACTGTACCTGTTCCAGATCTACAATAAAGAT
TTTGCCAAAGGCCATCATGGCAAACCGAATCTGCATACCCTGTATTGGACCGGTCTGTTTAGCCCTGAA
AATCTGGCAAAAACCTCGATTAAACTGAATGGTCAGGCGGAACTGTTTTATCGTCCGAAAAGCCGTAT
GAAACGTATGGCAGCTCGTCTGGGTGAAAAAATGCTGAACAAAAAACTGAAAGACCAGAAAACCCCG
ATCCCGGATACACTGTATCAAGAACTGTATGATTATGTGAACCATCGTCTGAGCCATGATCTGAGTGA
TGAAGCACGTGCCCTGCTGCCGAATGTTATTACCAAAGAAGTTAGCCACGAGATCATTAAAGATCGTC
GTTTTACCAGCGACAAATTCCTGTTTCATGTGCCGATTACCCTGAATTATCAGGCAGCAAATAGCCCGA
GCAAATTTAACCAGCGTGTTAATGCATATCTGAAAGAACATCCAGAAACGCCGATTATTGGTATTGAT
CGTGGTGAACGTAACCTGATTTATATCACCGTTATTGATAGCACCGGCAAAATCCTGGAACAGCGTAG
CCTGAATACCATTCAGCAGTTTGATTACCAGAAAAAACTGGATAATCGCGAGAAAGAACGTGTTGCAG
CACGTCAGGCATGGTCAGTTGTTGGTACAATTAAAGACCTGAAACAGGGTTATCTGAGCCAGGTTATT
CATGAAATTGTGGATCTGATGATTCACTATCAGGCCGTTGTTGTGCTGGAAAACCTGAATTTTGGCTTT
AAAAGCAAACGTACCGGCATTGCAGAAAAAGCAGTTTATCAGCAGTTCGAGAAAATGCTGATTGACA
AACTGAATTGCCTGGTGCTGAAAGATTATCCGGCTGAAAAAGTTGGTGGTGTTCTGAATCCGTATCAG
CTGACCGATCAGTTTACCAGCTTTGCAAAAATGGGCACCCAGAGCGGATTTCTGTTTTATGTTCCGGCA
CCGTATACGAGCAAAATTGATCCGCTGACCGGTTTTGTTGATCCGTTTGTTTGGAAAACCATCAAAAAC
CATGAAAGCCGCAAACATTTTCTGGAAGGTTTCGATTTTCTGCATTACGACGTTAAAACGGGTGATTTC
ATCCTGCACTTTAAAATGAATCGCAATCTGAGTTTTCAGCGTGGCCTGCCTGGTTTTATGCCTGCATGG
GATATTGTGTTTGAGAAAAACGAAACACAGTTCGATGCAAAAGGCACCCCGTTTATTGCAGGTAAACG
TATTGTTCCGGTGATTGAAAATCATCGTTTCACCGGTCGTTATCGCGATCTGTATCCGGCAAATGAACT
GATCGCACTGCTGGAAGAGAAAGGTATTGTTTTTCGTGATGGCTCAAACATTCTGCCGAAACTGCTGG
AAAATGATGATAGCCATGCAATTGATACCATGGTTGCACTGATTCGTAGCGTTCTGCAGATGCGTAAT

转让（Attorn）

AGCAATGCAGCAACCGGTGAAGATTACATTAATAGTCCGGTTCGTGATCTGAATGGTGTTTGTTTTGAT
AGCCGTTTTCAGAATCCGGAATGGCCGATGGATGCAGATGCAAATGGTGCATATCATATTGCACTGAA
AGGACAGCTGCTGCTGAACCACCTGAAAGAAAGCAAAGATCTGAAACTGCAAAACGGCATTAGCAAT
CAGGATTGGCTGGCATATATCCAAGAACTGCGTAACGGTCGT**AGCAGTGATGATGAAGCAACCGCA**
GATAGCCAGCATGCAGCACCGCCTAAAAAGAAACGTAAAGTTGGTGGTAGCGGTGGTTCAGGTGG
TAGTGGCGGTAGTGGTGGCTCAGGGGGTTCTGGTGGCTCTGGTGGTAGCCTCGAGCACCACCACCACC
ACCAC

Cpf1-4 氨基酸序列 (SEQ ID NO:1097)

MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTYADQCLQLVQLD
WENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKAELFNGKVLK
QLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSL
REHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETA
HIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSIDLTHIFISHK
KLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEIL
SHAHAALDQPLPTTLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYN
KARNYATKKPYSVEKFKLNFQRPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEK
TSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTA
YAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEI
MDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAA
RLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFLFH
VPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDN
REKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFKSKRTGIAEKAVYQQFE
KMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWK
TIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIAG
KRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIRSVLQMRNSNA
ATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWL
AYIQELRNGR**SSDDEATADSQHAAPPKKKRKV**

Cpf1-4 核苷酸序列 (SEQ ID NO:1114)

ATGACCCAGTTTGAAGGTTTCACCAATCTGTATCAGGTTAGCAAAACCCTGCGTTTTGAACTGATTCCG
CAGGGTAAAACCCTGAAACATATTCAAGAACAGGGCTTCATCGAAGAGGATAAAGCACGTAACGATC
ACTACAAAGAACTGAAACCGATTATCGACCGCATCTATAAAACCTATGCAGATCAGTGTCTGCAGCTG
GTTCAGCTGGATTGGGAAAATCTGAGCGCAGCAATTGATAGTTATCGCAAAGAAAAAACCGAAGAAA
CCCGTAATGCACTGATTGAAGAACAGGCAACCTATCGTAATGCCATCCATGATTATTTCATTGGTCGTA
CCGATAATCTGACCGATGCAATTAACAAACGTCACGCCGAAATCTATAAAGGCCTGTTTAAAGCCGAA
CTGTTTAATGGCAAAGTTCTGAAACAGCTGGGCACCGTTACCACCACCGAACATGAAAATGCACTGCT
GCGTAGCTTTGATAAATTCACCACCTATTTCAGCGGCTTTTATGAGAATCGCAAAAACGTGTTTAGCGC
AGAAGATATTAGCACCGCAATTCCGCATCGTATTGTGCAGGATAATTTCCCGAAATTCAAAGAGAACT
GCCACATTTTTACCCGTCTGATTACCGCAGTTCCGAGCCTGCGTGAACATTTTGAAAACGTTAAAAAAG
CCATCGGCATCTTTGTTAGCACCAGCATTGAAGAAGTTTTTAGCTTCCCGTTTTACAATCAGCTGCTGA
CCCAGACCCAGATTGATCTGTATAACCAACTGCTGGGTGGTATTAGCCGTGAAGCAGGCACCGAAAAA
ATCAAAGGTCTGAATGAAGTGCTGAATCTGGCCATTCAGAAAAATGATGAAACCGCACATATTATTGC
AAGCCTGCCGCATCGTTTTATTCCGCTGTTCAAACAAATTCTGAGCGATCGTAATACCCTGAGCTTTAT
TCTGGAAGAATTCAAATCCGATGAAGAGGTGATTCAGAGCTTTTGCAAATACAAAACGCTGCTGCGCA
ATGAAAATGTTCTGGAAACTGCCGAAGCACTGTTTAACGAACTGAATAGCATTGATCTGACCCACATC
TTTATCAGCCACAAAAAACTGGAAACCATTTCAAGCGCACTGTGTGATCATTGGGATACCCTGCGTAA
TGCCCTGTATGAACGTCGTATTAGCGAACTGACCGGTAAAATTACCAAAAGCGCGAAAGAAAAAGTTC
AGCGCAGTCTGAAACATGAGGATATTAATCTGCAAGAGATTATTAGCGCAGCCGGTAAAGAACTGTCA
GAAGCATTTAAACAGAAAACCAGCGAAATTCTGTCACATGCACATGCAGCACTGGATCAGCCGCTGCC
GACCACCCTGAAAAAACAAGAAGAAAAAGAAATCCTGAAAAGCCAGCTGGATAGCCTGCTGGGTCTG
TATCATCTGCTGGACTGGTTTGCAGTTGATGAAAGCAATGAAGTTGATCCGGAATTTAGCGCACGTCT
GACCGGCATTAAACTGGAAATGGAACCGAGCCTGAGCTTTTATAACAAAGCCCGTAATTATGCCACCA
AAAAACCGTATAGCGTCGAAAAATTCAAACTGAACTTTCAGCGTCCGACCCTGGCAAGCGGTTGGGAT
GTTAATAAAGAAAAAAACAACGGTGCCATCCTGTTCGTGAAAAATGGCCTGTATTATCTGGGTATTAT

转让（Attorn）

GCCGAAACAGAAAGGTCGTTATAAAGCGCTGAGCTTTGAACCGACGGAAAAAACCAGTGAAGGTTTT
GATAAAATGTACTACGACTATTTTCCGGATGCAGCCAAAATGATTCCGAAATGTAGCACCCAGCTGAA
AGCAGTTACCGCACATTTTCAGACCCATACCACCCCGATTCTGCTGAGCAATAACTTTATTGAACCGCT
GGAAATCACCAAAGAGATCTACGATCTGAATAACCCGGAAAAAGAGCCGAAAAAATTCCAGACCGCA
TATGCAAAAAAAACCGGTGATCAGAAAGGTTATCGTGAAGCGCTGTGTAAATGGATTGATTTCACCCG
TGATTTTCTGAGCAAATACACCAAAACCACCAGTATCGATCTGAGCAGCCTGCGTCCGAGCAGCCAGT
ATAAAGATCTGGGCGAATATTATGCAGAACTGAATCCGCTGCTGTATCATATTAGCTTTCAGCGTATTG
CCGAGAAAGAAATCATGGACGCAGTTGAAACCGGTAAACTGTACCTGTTCCAGATCTACAATAAAGAT
TTTGCCAAAGGCCATCATGGCAAACCGAATCTGCATACCCTGTATTGGACCGGTCTGTTTAGCCCTGAA
AATCTGGCAAAAACCTCGATTAAACTGAATGGTCAGGCGGAACTGTTTTATCGTCCGAAAAGCCGTAT
GAAACGTATGGCAGCTCGTCTGGGTGAAAAAATGCTGAACAAAAAACTGAAAGACCAGAAAACCCCG
ATCCCGGATACACTGTATCAAGAACTGTATGATTATGTGAACCATCGTCTGAGCCATGATCTGAGTGA
TGAAGCACGTGCCCTGCTGCCGAATGTTATTACCAAAGAAGTTAGCCACGAGATCATTAAAGATCGTC
GTTTTACCAGCGACAAATTCCTGTTTCATGTGCCGATTACCCTGAATTATCAGGCAGCAAATAGCCCGA
GCAAATTTAACCAGCGTGTTAATGCATATCTGAAAGAACATCCAGAAACGCCGATTATTGGTATTGAT
CGTGGTGAACGTAACCTGATTTATATCACCGTTATTGATAGCACCGGCAAAATCCTGGAACAGCGTAG
CCTGAATACCATTCAGCAGTTTGATTACCAGAAAAAACTGGATAATCGCGAGAAAGAACGTGTTGCAG
CACGTCAGGCATGGTCAGTTGTTGGTACAATTAAAGACCTGAAACAGGGTTATCTGAGCCAGGTTATT
CATGAAATTGTGGATCTGATGATTCACTATCAGGCCGTTGTTGTGCTGGAAAACCTGAATTTTGGCTTT
AAAAGCAAACGTACCGGCATTGCAGAAAAAGCAGTTTATCAGCAGTTCGAGAAAATGCTGATTGACA
AACTGAATTGCCTGGTGCTGAAAGATTATCCGGCTGAAAAAGTTGGTGGTGTTCTGAATCCGTATCAG
CTGACCGATCAGTTTACCAGCTTTGCAAAAATGGGCACCCAGAGCGGATTTCTGTTTTATGTTCCGGCA
CCGTATACGAGCAAAATTGATCCGCTGACCGGTTTTGTTGATCCGTTTGTTTGGAAAACCATCAAAAAC
CATGAAAGCCGCAAACATTTTCTGGAAGGTTTCGATTTTCTGCATTACGACGTTAAAACGGGTGATTTC
ATCCTGCACTTTAAAATGAATCGCAATCTGAGTTTTCAGCGTGGCCTGCCTGGTTTTATGCCTGCATGG
GATATTGTGTTTGAGAAAAACGAAACACAGTTCGATGCAAAAGGCACCCCGTTTATTGCAGGTAAACG
TATTGTTCCGGTGATTGAAAATCATCGTTTCACCGGTCGTTATCGCGATCTGTATCCGGCAAATGAACT
GATCGCACTGCTGGAAGAGAAAGGTATTGTTTTTCGTGATGGCTCAAACATTCTGCCGAAACTGCTGG
AAAATGATGATAGCCATGCAATTGATACCATGGTTGCACTGATTCGTAGCGTTCTGCAGATGCGTAAT
AGCAATGCAGCAACCGGTGAAGATTACATTAATAGTCCGGTTCGTGATCTGAATGGTGTTTGTTTTGAT
AGCCGTTTTCAGAATCCGGAATGGCCGATGGATGCAGATGCAAATGGTGCATATCATATTGCACTGAA
AGGACAGCTGCTGCTGAACCACCTGAAAGAAAGCAAAGATCTGAAACTGCAAAACGGCATTAGCAAT
CAGGATTGGCTGGCATATATCCAAGAACTGCGTAACGGTCGT**AGCAGTGATGATGAAGCAACCGCA**
GATAGCCAGCATGCAGCACCGCCTAAAAAGAAACGTAAAGTT

Cpf1-5 氨基酸序列 (SEQ ID NO:1107)

STQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTYADQCLQLVQLD
WENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKAELFNGKVLK
QLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSL
REHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETA
HIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSIDLTHIFISHK
KLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEIL
SHAHAALDQPLPTTLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYN
KARNYATKKPYSVEKFKLNFQRPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEK
TSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTA
YAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEI
MDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAH
RLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFLFH
VPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDN
REKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFKSKRTGIAEKAVYQQFE
KMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWK
TIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIAG
KRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIRSVLQMRNSNA
ATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWL
AYIQELRNGR**SSDDEATADSQHAAPPKKKRKV**

转让（Attorn）

Cpf1-5 核苷酸序列 (SEQ ID NO:1115)

AGCACCCAGTTTGAAGGTTTCACCAATCTGTATCAGGTTAGCAAAACCCTGCGTTTTGAACTGATTCCG
CAGGGTAAAACCCTGAAACATATTCAAGAACAGGGCTTCATCGAAGAGGATAAAGCACGTAACGATC
ACTACAAAGAACTGAAACCGATTATCGACCGCATCTATAAAACCTATGCAGATCAGTGTCTGCAGCTG
GTTCAGCTGGATTGGGAAAATCTGAGCGCAGCAATTGATAGTTATCGCAAAGAAAAAACCGAAGAAA
CCCGTAATGCACTGATTGAAGAACAGGCAACCTATCGTAATGCCATCCATGATTATTTCATTGGTCGTA
CCGATAATCTGACCGATGCAATTAACAAACGTCACGCCGAAATCTATAAAGGCCTGTTTAAAGCCGAA
CTGTTTAATGGCAAAGTTCTGAAACAGCTGGGCACCGTTACCACCACCGAACATGAAAATGCACTGCT
GCGTAGCTTTGATAAATTCACCACCTATTTCAGCGGCTTTTATGAGAATCGCAAAAACGTGTTTAGCGC
AGAAGATATTAGCACCGCAATTCCGCATCGTATTGTGCAGGATAATTTCCCGAAATTCAAAGAGAACT
GCCACATTTTTACCCGTCTGATTACCGCAGTTCCGAGCCTGCGTGAACATTTTGAAAACGTTAAAAAAG
CCATCGGCATCTTTGTTAGCACCAGCATTGAAGAAGTTTTTAGCTTCCCGTTTTACAATCAGCTGCTGA
CCCAGACCCAGATTGATCTGTATAACCAACTGCTGGGTGGTATTAGCCGTGAAGCAGGCACCGAAAAA
ATCAAAGGTCTGAATGAAGTGCTGAATCTGGCCATTCAGAAAAATGATGAAACCGCACATATTATTGC
AAGCCTGCCGCATCGTTTTATTCCGCTGTTCAAACAAATTCTGAGCGATCGTAATACCCTGAGCTTTAT
TCTGGAAGAATTCAAATCCGATGAAGAGGTGATTCAGAGCTTTTGCAAATACAAAACGCTGCTGCGCA
ATGAAAATGTTCTGGAAACTGCCGAAGCACTGTTTAACGAACTGAATAGCATTGATCTGACCCACATC
TTTATCAGCCACAAAAAACTGGAAACCATTTCAAGCGCACTGTGTGATCATTGGGATACCCTGCGTAA
TGCCCTGTATGAACGTCGTATTAGCGAACTGACCGGTAAAATTACCAAAAGCGCGAAAGAAAAAGTTC
AGCGCAGTCTGAAACATGAGGATATTAATCTGCAAGAGATTATTAGCGCAGCCGGTAAAGAACTGTCA
GAAGCATTTAAACAGAAAACCAGCGAAATTCTGTCACATGCACATGCAGCACTGGATCAGCCGCTGCC
GACCACCCTGAAAAAACAAGAAGAAAAAGAAATCCTGAAAAGCCAGCTGGATAGCCTGCTGGGTCTG
TATCATCTGCTGGACTGGTTTGCAGTTGATGAAAGCAATGAAGTTGATCCGGAATTTAGCGCACGTCT
GACCGGCATTAAACTGGAAATGGAACCGAGCCTGAGCTTTTATAACAAAGCCCGTAATTATGCCACCA
AAAAACCGTATAGCGTCGAAAAATTCAAACTGAACTTTCAGCGTCCGACCCTGGCAAGCGGTTGGGAT
GTTAATAAAGAAAAAAACAACGGTGCCATCCTGTTCGTGAAAAATGGCCTGTATTATCTGGGTATTAT
GCCGAAACAGAAAGGTCGTTATAAAGCGCTGAGCTTTGAACCGACGGAAAAAACCAGTGAAGGTTTT
GATAAAATGTACTACGACTATTTTCCGGATGCAGCCAAAATGATTCCGAAATGTAGCACCCAGCTGAA
AGCAGTTACCGCACATTTTCAGACCCATACCACCCCGATTCTGCTGAGCAATAACTTTATTGAACCGCT
GGAAATCACCAAAGAGATCTACGATCTGAATAACCCGGAAAAAGAGCCGAAAAAATTCCAGACCGCA
TATGCAAAAAAAACCGGTGATCAGAAAGGTTATCGTGAAGCGCTGTGTAAATGGATTGATTTCACCCG
TGATTTTCTGAGCAAATACACCAAAACCACCAGTATCGATCTGAGCAGCCTGCGTCCGAGCAGCCAGT
ATAAAGATCTGGGCGAATATTATGCAGAACTGAATCCGCTGCTGTATCATATTAGCTTTCAGCGTATTG
CCGAGAAAGAAATCATGGACGCAGTTGAAACCGGTAAACTGTACCTGTTCCAGATCTACAATAAAGAT
TTTGCCAAAGGCCATCATGGCAAACCGAATCTGCATACCCTGTATTGGACCGGTCTGTTTAGCCCTGAA
AATCTGGCAAAAACCTCGATTAAACTGAATGGTCAGGCGGAACTGTTTTATCGTCCGAAAAGCCGTAT
GAAACGTATGGCACATCGTCTGGGTGAAAAAATGCTGAACAAAAAACTGAAAGACCAGAAAACCCCG
ATCCCGGATACACTGTATCAAGAACTGTATGATTATGTGAACCATCGTCTGAGCCATGATCTGAGTGA
TGAAGCACGTGCCCTGCTGCCGAATGTTATTACCAAAGAAGTTAGCCACGAGATCATTAAAGATCGTC
GTTTTACCAGCGACAAATTCCTGTTTCATGTGCCGATTACCCTGAATTATCAGGCAGCAAATAGCCCGA
GCAAATTTAACCAGCGTGTTAATGCATATCTGAAAGAACATCCAGAAACGCCGATTATTGGTATTGAT
CGTGGTGAACGTAACCTGATTTATATCACCGTTATTGATAGCACCGGCAAAATCCTGGAACAGCGTAG
CCTGAATACCATTCAGCAGTTTGATTACCAGAAAAAACTGGATAATCGCGAGAAAGAACGTGTTGCAG
CACGTCAGGCATGGTCAGTTGTTGGTACAATTAAAGACCTGAAACAGGGTTATCTGAGCCAGGTTATT
CATGAAATTGTGGATCTGATGATTCACTATCAGGCCGTTGTTGTGCTGGAAAACCTGAATTTTGGCTTT
AAAAGCAAACGTACCGGCATTGCAGAAAAAGCAGTTTATCAGCAGTTCGAGAAAATGCTGATTGACA
AACTGAATTGCCTGGTGCTGAAAGATTATCCGGCTGAAAAAGTTGGTGGTGTTCTGAATCCGTATCAG
CTGACCGATCAGTTTACCAGCTTTGCAAAAATGGGCACCCAGAGCGGATTTCTGTTTTATGTTCCGGCA
CCGTATACGAGCAAAATTGATCCGCTGACCGGTTTTGTTGATCCGTTTGTTTGGAAAACCATCAAAAAC
CATGAAAGCCGCAAACATTTTCTGGAAGGTTTCGATTTTCTGCATTACGACGTTAAAACGGGTGATTTC
ATCCTGCACTTTAAAATGAATCGCAATCTGAGTTTTCAGCGTGGCCTGCCTGGTTTTATGCCTGCATGG
GATATTGTGTTTGAGAAAAACGAAACACAGTTCGATGCAAAAGGCACCCCGTTTATTGCAGGTAAACG
TATTGTTCCGGTGATTGAAAATCATCGTTTCACCGGTCGTTATCGCGATCTGTATCCGGCAAATGAACT
GATCGCACTGCTGGAAGAGAAAGGTATTGTTTTTCGTGATGGCTCAAACATTCTGCCGAAACTGCTGG
AAAATGATGATAGCCATGCAATTGATACCATGGTTGCACTGATTCGTAGCGTTCTGCAGATGCGTAAT

转让（Attorn）

AGCAATGCAGCAACCGGTGAAGATTACATTAATAGTCCGGTTCGTGATCTGAATGGTGTTTGTTTTGAT
AGCCGTTTTCAGAATCCGGAATGGCCGATGGATGCAGATGCAAATGGTGCATATCATATTGCACTGAA
AGGACAGCTGCTGCTGAACCACCTGAAAGAAAGCAAAGATCTGAAACTGCAAAACGGCATTAGCAAT
CAGGATTGGCTGGCATATATCCAAGAACTGCGTAACGGTCGT**AGCAGTGATGATGAAGCAACCGCA**
GATAGCCAGCATGCAGCACCGCCTAAAAAGAAACGTAAAGTT

Cpf1-6 氨基酸序列 (SEQ ID NO:1108)

STQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTYADQCLQLVQLD
WENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKAELFNGKVLK
QLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSL
REHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETA
HIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSIDLTHIFISHK
KLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEIL
SHAHAALDQPLPTTLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYN
KARNYATKKPYSVEKFKLNFQRPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEK
TSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTA
YAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEI
MDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAH
RLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFLFH
VPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDN
REKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFKSKRTGIAEKAVYQQFE
KMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWK
TIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIAG
KRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIRSVLQMRNSNA
ATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWL
AYIQELRNGR**SSDDEATADSQHAAPPKKKRKV**GGSGGSGGSGGSGGSGGSGGSLE<u>HHHHHH</u>

Cpf1-6 核苷酸序列 (SEQ ID NO:1116)

AGCACCCAGTTTGAAGGTTTCACCAATCTGTATCAGGTTAGCAAAACCCTGCGTTTTGAACTGATTCCG
CAGGGTAAAACCCTGAAACATATTCAAGAACAGGGCTTCATCGAAGAGGATAAAGCACGTAACGATC
ACTACAAAGAACTGAAACCGATTATCGACCGCATCTATAAAACCTATGCAGATCAGTGTCTGCAGCTG
GTTCAGCTGGATTGGGAAAATCTGAGCGCAGCAATTGATAGTTATCGCAAAGAAAAAACCGAAGAAA
CCCGTAATGCACTGATTGAAGAACAGGCAACCTATCGTAATGCCATCCATGATTATTTCATTGGTCGTA
CCGATAATCTGACCGATGCAATTAACAAACGTCACGCCGAAATCTATAAAGGCCTGTTTAAAGCCGAA
CTGTTTAATGGCAAAGTTCTGAAACAGCTGGGCACCGTTACCACCACCGAACATGAAAATGCACTGCT
GCGTAGCTTTGATAAATTCACCACCTATTTCAGCGGCTTTTATGAGAATCGCAAAAACGTGTTTAGCGC
AGAAGATATTAGCACCGCAATTCCGCATCGTATTGTGCAGGATAATTTCCCGAAATTCAAAGAGAACT
GCCACATTTTTACCCGTCTGATTACCGCAGTTCCGAGCCTGCGTGAACATTTTGAAAACGTTAAAAAAG
CCATCGGCATCTTTGTTAGCACCAGCATTGAAGAAGTTTTTAGCTTCCCGTTTTACAATCAGCTGCTGA
CCCAGACCCAGATTGATCTGTATAACCAACTGCTGGGTGGTATTAGCCGTGAAGCAGGCACCGAAAAA
ATCAAAGGTCTGAATGAAGTGCTGAATCTGGCCATTCAGAAAAATGATGAAACCGCACATATTATTGC
AAGCCTGCCGCATCGTTTTATTCCGCTGTTCAAACAAATTCTGAGCGATCGTAATACCCTGAGCTTTAT
TCTGGAAGAATTCAAATCCGATGAAGAGGTGATTCAGAGCTTTTGCAAATACAAAACGCTGCTGCGCA
ATGAAAATGTTCTGGAAACTGCCGAAGCACTGTTTAACGAACTGAATAGCATTGATCTGACCCACATC
TTTATCAGCCACAAAAAACTGGAAACCATTTCAAGCGCACTGTGTGATCATTGGGATACCCTGCGTAA
TGCCCTGTATGAACGTCGTATTAGCGAACTGACCGGTAAAATTACCAAAAGCGCGAAAGAAAAAGTTC
AGCGCAGTCTGAAACATGAGGATATTAATCTGCAAGAGATTATTAGCGCAGCCGGTAAAGAACTGTCA
GAAGCATTTAAACAGAAAACCAGCGAAATTCTGTCACATGCACATGCAGCACTGGATCAGCCGCTGCC
GACCACCCTGAAAAAACAAGAAGAAAAAGAAATCCTGAAAAGCCAGCTGGATAGCCTGCTGGGTCTG
TATCATCTGCTGGACTGGTTTGCAGTTGATGAAAGCAATGAAGTTGATCCGGAATTTAGCGCACGTCT
GACCGGCATTAAACTGGAAATGGAACCGAGCCTGAGCTTTTATAACAAAGCCCGTAATTATGCCACCA
AAAAACCGTATAGCGTCGAAAAATTCAAACTGAACTTTCAGCGTCCGACCCTGGCAAGCGGTTGGGAT
GTTAATAAAGAAAAAAACAACGGTGCCATCCTGTTCGTGAAAAATGGCCTGTATTATCTGGGTATTAT
GCCGAAACAGAAAGGTCGTTATAAAGCGCTGAGCTTTGAACCGACGGAAAAAACCAGTGAAGGTTTT
GATAAAATGTACTACGACTATTTTCCGGATGCAGCCAAAATGATTCCGAAATGTAGCACCCAGCTGAA

转让（Attorn）

AGCAGTTACCGCACATTTTCAGACCCATACCACCCCGATTCTGCTGAGCAATAACTTTATTGAACCGCT
GGAAATCACCAAAGAGATCTACGATCTGAATAACCCGGAAAAAGAGCCGAAAAAATTCCAGACCGCA
TATGCAAAAAAAACCGGTGATCAGAAAGGTTATCGTGAAGCGCTGTGTAAATGGATTGATTTCACCCG
TGATTTTCTGAGCAAATACACCAAAACCACCAGTATCGATCTGAGCAGCCTGCGTCCGAGCAGCCAGT
ATAAAGATCTGGGCGAATATTATGCAGAACTGAATCCGCTGCTGTATCATATTAGCTTTCAGCGTATTG
CCGAGAAAGAAATCATGGACGCAGTTGAAACCGGTAAACTGTACCTGTTCCAGATCTACAATAAAGAT
TTTGCCAAAGGCCATCATGGCAAACCGAATCTGCATACCCTGTATTGGACCGGTCTGTTTAGCCCTGAA
AATCTGGCAAAAACCTCGATTAAACTGAATGGTCAGGCGGAACTGTTTTATCGTCCGAAAAGCCGTAT
GAAACGTATGGCACATCGTCTGGGTGAAAAAATGCTGAACAAAAAACTGAAAGACCAGAAAACCCCG
ATCCCGGATACACTGTATCAAGAACTGTATGATTATGTGAACCATCGTCTGAGCCATGATCTGAGTGA
TGAAGCACGTGCCCTGCTGCCGAATGTTATTACCAAAGAAGTTAGCCACGAGATCATTAAAGATCGTC
GTTTTACCAGCGACAAATTCCTGTTTCATGTGCCGATTACCCTGAATTATCAGGCAGCAAATAGCCCGA
GCAAATTTAACCAGCGTGTTAATGCATATCTGAAAGAACATCCAGAAACGCCGATTATTGGTATTGAT
CGTGGTGAACGTAACCTGATTTATATCACCGTTATTGATAGCACCGGCAAAATCCTGGAACAGCGTAG
CCTGAATACCATTCAGCAGTTTGATTACCAGAAAAAACTGGATAATCGCGAGAAAGAACGTGTTGCAG
CACGTCAGGCATGGTCAGTTGTTGGTACAATTAAAGACCTGAAACAGGGTTATCTGAGCCAGGTTATT
CATGAAATTGTGGATCTGATGATTCACTATCAGGCCGTTGTTGTGCTGGAAAACCTGAATTTTGGCTTT
AAAAGCAAACGTACCGGCATTGCAGAAAAAGCAGTTTATCAGCAGTTCGAGAAAATGCTGATTGACA
AACTGAATTGCCTGGTGCTGAAAGATTATCCGGCTGAAAAAGTTGGTGGTGTTCTGAATCCGTATCAG
CTGACCGATCAGTTTACCAGCTTTGCAAAAATGGGCACCCAGAGCGGATTTCTGTTTTATGTTCCGGCA
CCGTATACGAGCAAAATTGATCCGCTGACCGGTTTTGTTGATCCGTTTGTTTGGAAAACCATCAAAAAC
CATGAAAGCCGCAAACATTTTCTGGAAGGTTTCGATTTTCTGCATTACGACGTTAAAACGGGTGATTTC
ATCCTGCACTTTAAAATGAATCGCAATCTGAGTTTTCAGCGTGGCCTGCCTGGTTTTATGCCTGCATGG
GATATTGTGTTTGAGAAAAACGAAACACAGTTCGATGCAAAAGGCACCCCGTTTATTGCAGGTAAACG
TATTGTTCCGGTGATTGAAAATCATCGTTTCACCGGTCGTTATCGCGATCTGTATCCGGCAAATGAACT
GATCGCACTGCTGGAAGAGAAAGGTATTGTTTTTCGTGATGGCTCAAACATTCTGCCGAAACTGCTGG
AAAATGATGATAGCCATGCAATTGATACCATGGTTGCACTGATTCGTAGCGTTCTGCAGATGCGTAAT
AGCAATGCAGCAACCGGTGAAGATTACATTAATAGTCCGGTTCGTGATCTGAATGGTGTTTGTTTTGAT
AGCCGTTTTCAGAATCCGGAATGGCCGATGGATGCAGATGCAAATGGTGCATATCATATTGCACTGAA
AGGACAGCTGCTGCTGAACCACCTGAAAGAAAGCAAAGATCTGAAACTGCAAAACGGCATTAGCAAT
CAGGATTGGCTGGCATATATCCAAGAACTGCGTAACGGTCGT**AGCAGTGATGATGAAGCAACCGCA**
GATAGCCAGCATGCAGCACCGCCTAAAAAGAAACGTAAAGTTGGTGGTAGCGGTGGTTCAGGTGG
TAGTGGCGGTAGTGGTGGCTCAGGGGGTTCTGGTGGCTCTGGTGGTAGCCTCGAGCACCACCACCACC
ACCAC

Cpf1-7 氨基酸序列 (SEQ ID NO:1109)

MGRDPGKPIPNPLLGLDSTAPKKKRKVGIHGVPAATQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEE
DKARNDHYKELKPIIDRIYKTYADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFI
GRTDNLTDAINKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVFSA
EDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLY
NQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFC
KYKTLLRNENVLETAEALFNELNSIDLTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKE
KVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTLKKQEEKEILKSQLDSLLGLYHL
LDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNFQMPTLASGWDVNKEK
NNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQ
THTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTAYAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSI
DLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLY
WTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRL
SHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGID
RGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDNREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIV
DLMIHYQAVVVLENLNFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFT
SFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNR
NLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFR
DGSNILPKLLENDDSHAIDTMVALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADA
NGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWLAYIQELRN**PKKKRKV**KLAAALEHHHHHH

转让（Attorn）

Cpf1-7 核苷酸序列 (SEQ ID NO:1117)

ATGGGTCGGGATCCAGGTAAACCGATTCCGAATCCGCTGCTGGGTCTGGATAGCACCGCACCGAAAAA
AAAACGTAAAGTTGGTATTCATGGTGTTCCGGCAGCAACCCAGTTTGAAGGTTTCACCAATCTGTATC
AGGTTAGCAAAACCCTGCGTTTTGAACTGATTCCGCAGGGTAAAACCCTGAAACATATTCAAGAACAG
GGCTTCATCGAAGAGGATAAAGCACGTAACGATCACTACAAAGAACTGAAACCGATTATCGACCGCA
TCTATAAAACCTATGCAGATCAGTGTCTGCAGCTGGTTCAGCTGGATTGGGAAAATCTGAGCGCAGCA
ATTGATAGTTATCGCAAAGAAAAAACCGAAGAAACCCGTAATGCACTGATTGAAGAACAGGCAACCT
ATCGTAATGCCATCCATGATTATTTCATTGGTCGTACCGATAATCTGACCGATGCAATTAACAAACGTC
ACGCCGAAATCTATAAAGGCCTGTTTAAAGCCGAACTGTTTAATGGCAAAGTTCTGAAACAGCTGGGC
ACCGTTACCACCACCGAACATGAAAATGCACTGCTGCGTAGCTTTGATAAATTCACCACCTATTTCAGC
GGCTTTTATGAGAATCGCAAAAACGTGTTTAGCGCAGAAGATATTAGCACCGCAATTCCGCATCGTAT
TGTGCAGGATAATTTCCCGAAATTCAAAGAGAACTGCCACATTTTTACCCGTCTGATTACCGCAGTTCC
GAGCCTGCGTGAACATTTTGAAAACGTTAAAAAAGCCATCGGCATCTTTGTTAGCACCAGCATTGAAG
AAGTTTTTAGCTTCCCGTTTTACAATCAGCTGCTGACCCAGACCCAGATTGATCTGTATAACCAACTGC
TGGGTGGTATTAGCCGTGAAGCAGGCACCGAAAAAATCAAAGGTCTGAATGAAGTGCTGAATCTGGC
CATTCAGAAAAATGATGAAACCGCACATATTATTGCAAGCCTGCCGCATCGTTTTATTCCGCTGTTCAA
ACAAATTCTGAGCGATCGTAATACCCTGAGCTTTATTCTGGAAGAATTCAAATCCGATGAAGAGGTGA
TTCAGAGCTTTTGCAAATACAAAACGCTGCTGCGCAATGAAAATGTTCTGGAAACTGCCGAAGCACTG
TTTAACGAACTGAATAGCATTGATCTGACCCACATCTTTATCAGCCACAAAAAACTGGAAACCATTTC
AAGCGCACTGTGTGATCATTGGGATACCCTGCGTAATGCCCTGTATGAACGTCGTATTAGCGAACTGA
CCGGTAAAATTACCAAAAGCGCGAAAGAAAAAGTTCAGCGCAGTCTGAAACATGAGGATATTAATCT
GCAAGAGATTATTAGCGCAGCCGGTAAAGAACTGTCAGAAGCATTTAAACAGAAAACCAGCGAAATT
CTGTCACATGCACATGCAGCACTGGATCAGCCGCTGCCGACCACCCTGAAAAAACAAGAAGAAAAAG
AAATCCTGAAAAGCCAGCTGGATAGCCTGCTGGGTCTGTATCATCTGCTGGACTGGTTTGCAGTTGAT
GAAAGCAATGAAGTTGATCCGGAATTTAGCGCACGTCTGACCGGCATTAAACTGGAAATGGAACCGA
GCCTGAGCTTTTATAACAAAGCCCGTAATTATGCCACCAAAAAACCGTATAGCGTCGAAAAATTCAAA
CTGAACTTTCAGATGCCGACCCTGGCAAGCGGTTGGGATGTTAATAAAGAAAAAAACAACGGTGCCAT
CCTGTTCGTGAAAAATGGCCTGTATTATCTGGGTATTATGCCGAAACAGAAAGGTCGTTATAAAGCGC
TGAGCTTTGAACCGACGGAAAAAACCAGTGAAGGTTTTGATAAAATGTACTACGACTATTTTCCGGAT
GCAGCCAAAATGATTCCGAAATGTAGCACCCAGCTGAAAGCAGTTACCGCACATTTTCAGACCCATAC
CACCCCGATTCTGCTGAGCAATAACTTTATTGAACCGCTGGAAATCACCAAAGAGATCTACGATCTGA
ATAACCCGGAAAAAGAGCCGAAAAAATTCCAGACCGCATATGCAAAAAAAACCGGTGATCAGAAAG
GTTATCGTGAAGCGCTGTGTAAATGGATTGATTTCACCCGTGATTTTCTGAGCAAATACACCAAAACC
ACCAGTATCGATCTGAGCAGCCTGCGTCCGAGCAGCCAGTATAAAGATCTGGGCGAATATTATGCAGA
ACTGAATCCGCTGCTGTATCATATTAGCTTTCAGCGTATTGCCGAGAAAGAAATCATGGACGCAGTTG
AAACCGGTAAACTGTACCTGTTCCAGATCTACAATAAAGATTTTGCCAAAGGCCATCATGGCAAACCG
AATCTGCATACCCTGTATTGGACCGGTCTGTTTAGCCCTGAAAATCTGGCAAAAACCTCGATTAAACTG
AATGGTCAGGCGGAACTGTTTTATCGTCCGAAAAGCCGTATGAAACGTATGGCACATCGTCTGGGTGA
AAAAATGCTGAACAAAAAACTGAAAGACCAGAAAACCCCGATCCCGGATACACTGTATCAAGAACTG
TATGATTATGTGAACCATCGTCTGAGCCATGATCTGAGTGATGAAGCACGTGCCCTGCTGCCGAATGTT
ATTACCAAAGAAGTTAGCCACGAGATCATTAAAGATCGTCGTTTTACCAGCGACAAATTCTTTTTTCAT
GTGCCGATTACCCTGAATTATCAGGCAGCAAATAGCCCGAGCAAATTTAACCAGCGTGTTAATGCATA
TCTGAAAGAACATCCAGAAACGCCGATTATTGGTATTGATCGTGGTGAACGTAACCTGATTTATATCA
CCGTTATTGATAGCACCGGCAAAATCCTGGAACAGCGTAGCCTGAATACCATTCAGCAGTTTGATTAC
CAGAAAAAACTGGATAATCGCGAGAAAGAACGTGTTGCAGCACGTCAGGCATGGTCAGTTGTTGGTA
CAATTAAAGACCTGAAACAGGGTTATCTGAGCCAGGTTATTCATGAAATTGTGGATCTGATGATTCAC
TATCAGGCCGTTGTTGTGCTGGAAAACCTGAATTTTGGCTTTAAAAGCAAACGTACCGGCATTGCAGA
AAAAGCAGTTTATCAGCAGTTCGAGAAAATGCTGATTGACAAACTGAATTGCCTGGTGCTGAAAGATT
ATCCGGCTGAAAAAGTTGGTGGTGTTCTGAATCCGTATCAGCTGACCGATCAGTTTACCAGCTTTGCAA
AAATGGGCACCCAGAGCGGATTTCTGTTTTATGTTCCGGCACCGTATACGAGCAAAATTGATCCGCTG
ACCGGTTTTGTTGATCCGTTTGTTTGGAAAACCATCAAAAACCATGAAAGCCGCAAACATTTTCTGGA
AGGTTTCGATTTTCTGCATTACGACGTTAAAACGGGTGATTTCATCCTGCACTTTAAAATGAATCGCAA
TCTGAGTTTTCAGCGTGGCCTGCCTGGTTTTATGCCTGCATGGGATATTGTGTTTGAGAAAAACGAAAC
ACAGTTCGATGCAAAAGGCACCCCGTTTATTGCAGGTAAACGTATTGTTCCGGTGATTGAAAATCATC
GTTTCACCGGTCGTTATCGCGATCTGTATCCGGCAAATGAACTGATCGCACTGCTGGAAGAGAAAGGT

转让（Attorn）

ATTGTTTTTCGTGATGGCTCAAACATTCTGCCGAAACTGCTGGAAAATGATGATAGCCATGCAATTGAT
ACCATGGTTGCACTGATTCGTAGCGTTCTGCAGATGCGTAATAGCAATGCAGCAACCGGTGAAGATTA
CATTAATAGTCCGGTTCGTGATCTGAATGGTGTTTGTTTTGATAGCCGTTTTCAGAATCCGGAATGGCC
GATGGATGCAGATGCAAATGGTGCATATCATATTGCACTGAAAGGACAGCTGCTGCTGAACCACCTGA
AAGAAAGCAAAGATCTGAAACTGCAAAACGGCATTAGCAATCAGGATTGGCTGGCATATATCCAAGA
ACTGCGTAAC**CCTAAAAAAAAAACGCAAAGTG**AAGCTTGCGGCCGCACTCGAGCACCACCACCACCAC
CAC

图62