

用于治疗血红蛋白疾病的组合物和方法

技术领域

本申请涉及生物医药领域，具体的涉及一种用于治疗血红蛋白疾病的分子、包含该分子的组合物及其方法和用途。

背景技术

血红蛋白病是遗传性贫血症中异质性很强的一类，根据缺乏的珠蛋白链分类，可包括 α -血红蛋白病和 β -血红蛋白病，以及血红蛋白定量异常。 β -地中海贫血症是一种异质性常染色体隐性遗传性贫血症，其特征是 β -珠蛋白链合成减少或缺失导致与血红蛋白 A (HbA) 相关的疾病。A 型血红蛋白链过量会导致红细胞生成障碍、红细胞前体髓内凋亡和溶血性贫血。 β -地中海贫血在临床上分为两种类型，根据症状的严重程度加以区分： β -地中海贫血重型（或称 β^0 ，在这种情况下， β -珠蛋白链的生成因突变而停止）和 β -地中海贫血中型（或称 β^+ ，在这种情况下， β -珠蛋白链的生成减少）。重型 β 地中海贫血是一种严重的内科疾病，需要常规输血。镰状细胞病 (SCD) 也是一种因 β -珠蛋白链生成缺陷而导致的遗传性血液疾病，其包括镰状细胞性贫血、镰状血红蛋白 C 病 (HbSC)、镰状 β^+ 地中海贫血 (HbS/ β^+) 和镰状 β^0 地中海贫血 (HbS/ β^0)。

研究表明，所有形式的 β -血红蛋白病都是 β -珠蛋白结构基因 (HBB) 突变引起的。 β -地中海贫血是由 200 多种不同的 β -珠蛋白基因突变引起的。SCD 是由第六个密码子中的 A 到 T 点突变引起的，该突变导致有缺陷的 β -珠蛋白 (β^s) 中出现 E6V 替换。人类 β -珠蛋白基因座由位于第 11 号染色体短区内的五个基因组成，负责产生 HAb 的 β -珠蛋白链。该基因座除包含 β -珠蛋白基因外，还包含 δ 、 γ -A (HBG1)、 γ -G (HBG2) 和 ϵ 珠蛋白。一个 16 kb 长的基因座控制区 (LCR) 位于 HBB 基因座上游~40-60 kb 处，被认为在整个发育过程中调控类 β -珠蛋白基因的不同表达。在新生儿后期，胎儿的 γ -A、 γ -G 表达被抑制，而成人的 β -珠蛋白基因则启动表达；这一过程被称为 γ 到 β 珠蛋白表达转换。胎儿血红蛋白 (HbF) 遗传性持续存在 (HPFH) 是一种良性病变，成年后仍有大量 HbF 生成。HPFH 和 β -血红蛋白病的杂合子患者临床症状轻微，而这些疾病的纯合子患者症状严重。HPFH 通常是由 β -珠蛋白基因簇或 γ -珠蛋白启动子区域的突变引起的，这些突变要么为红细胞激活剂创造了新的结合位点，要么破坏了抑制因子的结合位点。位于 HBG1 和 HBG2 近端启动子的点突变分为两个不同的

群组：分别位于复制的 γ -珠蛋白基因转录起始位点（TSS）上游约 115bp 和 200bp 处。在 γ -珠蛋白基因沉默中起主要作用的抑制因子是与-115bp 处结合的 BCL11A。最近，ZBTB7A；又称 LRF 或 FBI-1 被确定为与-200bp 处的位点结合的第二个主要抑制因子。这些发现表明，将人工 HFPH 样突变引入启动子区域以重新激活 γ -珠蛋白基因表达，是针对 β -血红蛋白病的一种具有应用前景的治疗策略。

在中国，尽管临床治疗的改善降低了患儿的死亡率，但对大多数 β -血红蛋白病患者的治疗仍以支持治疗为主。目前的治疗方法旨在缓解症状和治疗并发症，包括定期输血、抑制红细胞生成、祛铁和止痛。人类白细胞抗原（HLA）相同的兄弟姐妹骨髓移植（BMT）或 HLA 匹配的捐献者脐带血移植（CBT）是另一种可行的治疗方法，但存在移植物抗宿主病（GVHD）的风险。然而，异基因 BMT 仅适用于一小部分 β -血红蛋白病患者，原因是大多数患者无法找到完全匹配的非亲缘供体。

利用自体转基因 HSPCs 的新型治疗方法可避免对完全匹配的骨髓或脐带血供体的需求，从而避免移植后发生 GVHD 的风险。全基因组关联研究显示，位于 BCL11A 基因座红细胞特异性增强子上的一些单核苷酸多态性（SNPs）与 BCL11a 表达下调、成年期 γ -珠蛋白持续表达以及轻度 β -血红蛋白病临床表现有关。使用校正过的 β -珠蛋白进行基因治疗也被认为是一种可能治愈 β -血红蛋白病的方法。无 HIV-1 复制元件的慢病毒载体（LV）能够将外源基因片段转移到静止的造血干细胞中。用编码全功能 β A-T87Q-珠蛋白的 BB305 慢病毒载体（ZYNTEGLO® - betibeglogene autotemcel）进行自体移植的 HSPCs 基因疗法，使严重的 β -血红蛋白病患者摆脱定期输血。然而，上述基因治疗的安全性令人担忧，将 LV 载体中的载体基因随机整合到基因组中，同时引入强启动子和增强子元件，可能会引起对激活原癌基因或由这种效应引发的其他有害事件。

总体而言，尽管最近在开发基因方法以解决 β -血红蛋白病方面取得了进展，但在为这些遗传疾病提供安全有效的治疗方面，仍有一些临床需求尚未得到满足。

发明内容

为了克服现有技术治疗血红蛋白病的局限性，本申请提供了通过基因组编辑提高人 CD34+HSPCs 的人红细胞后代中 HbF 的水平的方法，HbF 的两条 γ -珠蛋白链由 HBG1 和 HBG2 表达，本申请的分子、组合物、细胞（包括可输注给患有严重血红蛋白病的患者的自体 CD34+ HSPCs）、试剂盒以及方法等作用于 HBG1 和 HBG2 的转录调控区域，可用于治疗血红蛋白病，如 β -地中海贫血和镰状细胞病。

一方面，本申请提供一种指导 RNA（gRNA）分子，所述 gRNA 分子包含与靶序列互补的靶向结构域，所述靶序列位于 HBG1 启动子区域内或 HBG2 启动子区域内，所述 HBG1 启动子区域位于人基因组 hg19 的 chr11: 5,248,269 至 5,249,857 之间，所述 HBG2 启动子区域位于人基因组 hg19 的 chr11: 5,253,188-5,254,781 之间。

在一些实施方案中，所述靶序列位于所述 HBG1 和所述 HBG2 的转录起始位点上游约 210bp 至上游约 100bp 之间。

在一些实施方案中，所述 gRNA 包含 tracr 序列和 crRNA 序列。

在一些实施方案中，所述 gRNA 为双组分向导 RNA 分子。

在一些实施方案中，所述 gRNA 为单组分向导 RNA 分子（sgRNA）。

在一些实施方案中，其中（1）使包含所述 gRNA 分子的 CRISPR/Cas 系统接触所述靶序列，或（2）将包含所述 gRNA 分子的 CRISPR/Cas 系统引入包含所述靶序列的基因所在的细胞中，能够在所述靶序列处或附近形成的碱基插入或缺失（indel）。

在一些实施方案中，所述 CRISPR/Cas 系统包含 Cas12b 蛋白和/或其功能活性片段。

在一些实施方案中，所述 CRISPR/Cas 系统包含 Cas12i 蛋白和/或其功能活性片段。

在一些实施方案中，所述核酸酶的 N 端和/或 C 端与一个或一个以上的核定位信号连接。

在一些实施方案中，所述核定位信号与 His 标签连接。

在一些实施方案中，所述 tracr 序列包含 SEQ ID NO:60 所述的核苷酸序列。例如，所述 tracr 序列包含如下所示序列（m 表示 2-O-甲基修饰，*表示 3'硫代磷酸酯修饰）：

mG*mG*mT*CGTCTATAGGACGGCGAGTTTTTCAACGGGTGTGCCAATGGCCACTTTCCA
GGTGGCAAAGCCCGTTGAACTTCTCAAAAAGAACGCTCGCTCAGTGTTCT*mG*mA*mC

在一些实施方案中，所述 crRNA 序列包含选自 SEQ ID NOs: 61-66 中任一项所述的核苷酸序列。例如，所述 crRNA 序列包含如下表所示序列：

编号	种类	核苷酸序列（m 表示 2-O-甲基修饰，*表示 3'硫代磷酸酯修饰）
12b-crRN A1-18	RNA	mG*mT*mC*GGATCACTGAGCGAGCGATCTGAGAAGTGGCACAGAT AGTGTGGGGAA*mG*mG*mG
12b-crRN A1-20	RNA	mG*mT*mC*GGATCACTGAGCGAGCGATCTGAGAAGTGGCACAGAT AGTGTGGGGAAGG*mG*mG*mC
12b-crRN A1-23	RNA	mG*mT*mC*GGATCACTGAGCGAGCGATCTGAGAAGTGGCACAGAT AGTGTGGGGAAGGGGC*mC*mC*mC

12b-crRNA1-25	RNA	mG*mT*mC*GGATCACTGAGCGAGCGATCTGAGAAGTGGCACAGATAGTGTGGGGAAGGGGCCC*mC*mC*mA
12b-crRNA1-28	RNA	mG*mT*mC*GGATCACTGAGCGAGCGATCTGAGAAGTGGCACAGATAGTGTGGGGAAGGGGCCCCCA*mA*mG*mA
12b-crRNA1-31	RNA	mG*mT*mC*GGATCACTGAGCGAGCGATCTGAGAAGTGGCACAGATAGTGTGGGGAAGGGGCCCCCAAGA*mG*mG*mA

在一些实施方案中，其包含选自 SEQ ID NOs: 12-59 和 67-78 中任一项所述的核苷酸序列。例如，所述 gRNA 序列包含如下表所示序列：

编号	种类	核苷酸序列（m 表示 2-O-甲基修饰，*表示 3'硫代磷酸酯修饰）
12b-gRNA1-18	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTGCCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACAGATAGTGTGGGGAA*mG*mG*mG
12b-gRNA1-20	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTGCCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACAGATAGTGTGGGGAAGG*mG*mG*mC
12b-gRNA1-21	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTGCCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACAGATAGTGTGGGGAAGGG*mG*mC*mC
12b-gRNA1-22	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTGCCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACAGATAGTGTGGGGAAGGGG*mC*mC*mC
12b-gRNA1-23	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTGCCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACAGATAGTGTGGGGAAGGGGC*mC*mC*mC
12b-gRNA1-24	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTGCCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACAGATAGTGTGGGGAAGGGGCC*mC*mC*mC
12b-	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG

gRNA1 -25		CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACAGATAGTGTGGGGAAGGGGCCCC*mC*mC*mA
12b- gRNA1 -26	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACAGATAGTGTGGGGAAGGGGCCCC*mC*mA*mA
12b- gRNA1 -27	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACAGATAGTGTGGGGAAGGGGCCCCC*mA*mA*mG
12b- gRNA1 -28	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACAGATAGTGTGGGGAAGGGGCCCCCA*mA*mG*mA
12b- gRNA1 -29	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACAGATAGTGTGGGGAAGGGGCCCCCAA*mG*mA*mG
12b- gRNA1 -30	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACAGATAGTGTGGGGAAGGGGCCCCCAAG*mA*mG*mG
12b- gRNA1 -31	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACAGATAGTGTGGGGAAGGGGCCCCCAAGA*mG*mG*mA
12b- gRNA2 -18	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACCATTGAGATAGTGTG*mG*mG*mG
12b- gRNA2 -20	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACCATTGAGATAGTGTGGG*mG*mA*mA
12b- gRNA2 -21	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACCATTGAGATAGTGTGGGG*mA*mA*mG
12b-	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG

gRNA2 -22		CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACCATTGAGATAGTGTGGGGA*mA*mG*mG
12b- gRNA2 -23	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACCATTGAGATAGTGTGGGGA*mG*mG*mG
12b- gRNA2 -24	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACCATTGAGATAGTGTGGGGAAG*mG*mG*mG
12b- gRNA2 -25	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACCATTGAGATAGTGTGGGGAAGG*mG*mG*mC
12b- gRNA2 -26	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACCATTGAGATAGTGTGGGGAAGGG*mG*mC*mC
12b- gRNA2 -27	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACCATTGAGATAGTGTGGGGAAGGGG*mC*mC*mC
12b- gRNA2 -28	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACCATTGAGATAGTGTGGGGAAGGGGC*mC*mC*mC
12b- gRNA2 -29	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACCATTGAGATAGTGTGGGGAAGGGGCC*mC*mC*mC
12b- gRNA2 -30	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACCATTGAGATAGTGTGGGGAAGGGGCCC*mC*mC*mA
12b- gRNA2 -31	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACCATTGAGATAGTGTGGGGAAGGGGCCCC*mC*mA*mA
12b-	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG

gRNA3 -18		CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACGCATTGAGATAGTGT*mG*mG*mG
12b- gRNA3 -20	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACGCATTGAGATAGTGTGG*mG*mG*mA
12b- gRNA3 -21	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACGCATTGAGATAGTGTGGG*mG*mA*mA
12b- gRNA3 -22	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACGCATTGAGATAGTGTGGGG*mA*mA*mG
12b- gRNA3 -23	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACGCATTGAGATAGTGTGGGGA*mA*mG*mG
12b- gRNA3 -24	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACGCATTGAGATAGTGTGGGGAA*mG*mG*mG
12b- gRNA3 -25	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACGCATTGAGATAGTGTGGGGAAG*mG*mG*mG
12b- gRNA3 -26	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACGCATTGAGATAGTGTGGGGAAGG*mG*mG*mC
12b- gRNA3 -27	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACGCATTGAGATAGTGTGGGGAAGGG*mG*mC*mC
12b- gRNA3 -28	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACGCATTGAGATAGTGTGGGGAAGGGG*mC*mC*mC
12b-	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG

gRNA3 -29		CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACGCATTGAGATAGTGTGGGGAAGGGGC*mC*mC*mC
12b- gRNA3 -30	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACGCATTGAGATAGTGTGGGGAAGGGGCC*mC*mC*mC
12b- gRNA3 -31	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACGCATTGAGATAGTGTGGGGAAGGGGCCC*mC*mC*mA
12b- gRNA4	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACGCCAGCCTTGCCCTTGAC*mC*mA*mA
12b- gRNA5	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACCCCTTGTC AAGGCTATTG*mG*mT*mC
12b- gRNA6	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACGCCTTGTC AAGGCTATT*mG*mG*mT
12b- gRNA7	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACCCAGACAGATATTTGCAT*mT*mG*mA
12b- gRNA8	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACCCACACTATCTCAATG*mC*mA*mA
12b- gRNA9	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACAGACAGATATTTGCATT*mG*mA*mG
12b- gRNA1 0	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACTCAAGGCTATTGGTCAA*mG*mG*mC
12b-	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG

gRNA1 1		CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACAGCAGCAGTATCCTCTT*mG*mG*mG
12b- gRNA1 2	RNA	mG*mC*mA*GGTCGTCTATAGGACGGCGAGTTTTTCAACGGGTGTG CCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAA GTGGCACGTCAAGTTTGCCTTGTC*mA*mA*mG
12b- Aa13tra crRNA	RNA	mG*mG*mT*CGTCTATAGGACGGCGAGTTTTTCAACGGGTGTGCCA ATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAACTTCTCAAAAAGA ACGCTCGCTCAGTGTCT*mG*mA*mC
12b- crRNA 1-18	RNA	mG*mT*mC*GGATCACTGAGCGAGCGATCTGAGAAGTGGCACAGAT AGTGTGGGGAA*mG*mG*mG
12b- crRNA 1-20	RNA	mG*mT*mC*GGATCACTGAGCGAGCGATCTGAGAAGTGGCACAGAT AGTGTGGGGAAGG*mG*mG*mC
12b- crRNA 1-23	RNA	mG*mT*mC*GGATCACTGAGCGAGCGATCTGAGAAGTGGCACAGAT AGTGTGGGGAAGGGGC*mC*mC*mC
12b- crRNA 1-25	RNA	mG*mT*mC*GGATCACTGAGCGAGCGATCTGAGAAGTGGCACAGAT AGTGTGGGGAAGGGGCCC*mC*mC*mA
12b- crRNA 1-28	RNA	mG*mT*mC*GGATCACTGAGCGAGCGATCTGAGAAGTGGCACAGAT AGTGTGGGGAAGGGGCCCCA*mA*mG*mA
12b- crRNA 1-31	RNA	mG*mT*mC*GGATCACTGAGCGAGCGATCTGAGAAGTGGCACAGAT AGTGTGGGGAAGGGGCCCCAAGA*mG*mG*mA
12i- gRNA- 1	RNA	mA*mG*mA*AATCCGTCTTTCATTGACGGAGATAGTGTGGGGAAGG *mG*mG*mC
12i-	RNA	mA*mG*mA*AATCCGTCTTTCATTGACGGCATTGAGATAGTGTGGG*

gRNA- 2		mG*mA*mA
12i- gRNA- 3	RNA	mA*mG*mA*AATCCGTCTTTCATTGACGGGCATTGAGATAGTGTGG* mG*mG*mA
12i- gRNA- 4	RNA	mA*mG*mA*AATCCGTCTTTCATTGACGGGCCAGCCTTGCCTTGAC* mC*mA*mA
12i- gRNA- 5	RNA	mA*mG*mA*AATCCGTCTTTCATTGACGGCCTTGTCAAGGCTATTG* mG*mT*mC
12i- gRNA- 6	RNA	mA*mG*mA*AATCCGTCTTTCATTGACGGGCCTTGTCAAGGCTATT* mG*mG*mT
12i- gRNA- 7	RNA	mA*mG*mA*AATCCGTCTTTCATTGACGGCAGACAGATATTTGCAT* mT*mG*mA
12i- gRNA- 8	RNA	mA*mG*mA*AATCCGTCTTTCATTGACGGCCCACACTATCTCAATG* mC*mA*mA
12i- gRNA- 9	RNA	mA*mG*mA*AATCCGTCTTTCATTGACGGAGACAGATATTTGCATT* mG*mA*mG
12i- gRNA- 10	RNA	mA*mG*mA*AATCCGTCTTTCATTGACGGTCAAGGCTATTGGTCAA* mG*mG*mC
12i- gRNA- 11	RNA	mA*mG*mA*AATCCGTCTTTCATTGACGGAGCAGCAGTATCCTCTT* mG*mG*mG
12i-	RNA	mA*mG*mA*AATCCGTCTTTCATTGACGGGTCAAGTTTGCCTTGTC*

gRNA- 12		mA*mA*mG
-------------	--	----------

另一方面，本申请提供一种核酸分子，其编码本申请所述的 gRNA 分子。

另一方面，本申请提供一种载体，其包含本申请所述的核酸分子。

在一些实施方案中，所述载体选自慢病毒载体、腺病毒载体、腺相关病毒(AAV)载体、单纯疱疹病毒(HSV)载体、质粒、微环、纳米质粒和 RNA 载体。

另一方面，本申请提供一种核糖核蛋白（RNP）复合物，其包含本申请所述的 gRNA 分子和 CRISPR/Cas 系统的核酸酶。

在一些实施方案中，所述核酸酶为 II 类 V 型 Cas 酶。

在一些实施方案中，所述核酸酶为 Cas12b 核酸酶。

在一些实施方案中，所述 Cas12b 核酸酶包含其变体、其直系同源物和/或其功能活性片段。

在一些实施方案中，所述 Cas12b 核酸酶为来自 *Alicyclobacillus acidiphilus* 的 Cas12b 核酸酶(AaCas12b)。

在一些实施方案中，所述 AaCas12b 核酸酶包含下列相对于野生型 AaCas12b 核酸酶的突变中的一种或多种：（1）用带正电荷的氨基酸残基取代所述野生型 AaCas12b 核酸酶中与前间隔序列邻近基序(PAM)相互作用的一个或多个氨基酸残基；（2）用具有芳环的氨基酸残基取代所述野生型 AaCas12b 核酸酶中参与打开 DNA 双链的一个或多个氨基酸残基；和/或（3）用带正电荷的氨基酸残基或疏水性氨基酸残基取代所述野生型 AaCas12b 核酸酶的 RuvC 结构域中与单链 DNA 底物相互作用的一个或多个氨基酸残基，其中，所述野生型 AaCas12b 核酸酶的氨基酸序列如 SEQ ID NO: 91 所述。

在一些实施方案中，所述与 PAM 相互作用的所述一个或多个氨基酸残基位于以下的一个或多个位置：116、123、130、132、144、145、153、173、222、395、400 和 475；且其中氨基酸残基根据 SEQ ID NO:91 编号。

在一些实施方案中，所述带正电荷的氨基酸残基取代所述野生型 AaCas12b 核酸酶中与 PAM 相互作用的一个或多个氨基酸残基是以下取代中的一种或多种：D116R 和 E475R；且其中所述氨基酸残基根据 SEQ ID NO:91 编号。

在一些实施方案中，所述参与打开所述 DNA 双链的一个或多个氨基酸残基位于以下的一个或多个位置：118，和 119；且其中所述氨基酸残基根据 SEQ ID NO:91 编号。

在一些实施方案中，所述用具有芳环的氨基酸残基取代所述野生型 AaCas12b 核酸酶中

参与打开所述 DNA 双链的一个或多个氨基酸残基是 Q119Y、Q119F 或 Q119W 取代；且其中所述氨基酸残基根据 SEQ ID NO:91 编号。

在一些实施方案中，所述位于 RuvC 结构域中并与单链 DNA 底物相互作用的一个或多个氨基酸残基位于以下的一个或多个位置：300、301、304、329、636、639、647、682、757、758、761、764、768、852、854、856、857、858、860、862、863、865、866、867、869、938、956、957、958、994、1093 和 1097；且其中所述氨基酸残基根据 SEQ ID NO:91 编号。

在一些实施方案中，所述取代所述野生型 AaCas12b 核酸酶中位于 RuvC 结构域中并与单链 DNA 底物相互作用的一个或多个氨基酸残基是以下取代中的一种或多种：E636R、Q639R、T647R、Q682R、I757R、E758R、E761R、Q854R、N857K、D858R、I994R、Q1093R 和 W1097R；且其中所述氨基酸残基根据 SEQ ID NO:91 编号。

在一些实施方案中，所述 gRNA 分子包含 SEQ ID NOs: 12-59 中任一项所述的核苷酸序列。

在一些实施方案中，所述核酸酶为 Cas12i 核酸酶。

在一些实施方案中，所述 Cas12i 核酸酶包含其变体、其直系同源物和/或其功能活性片段。

在一些实施方案中，所述 Cas12i 核酸酶包含下列相对于野生型 Cas12i2 核酸酶的突变中的一种或多种：（1）将所述野生型 Cas12i2 核酸酶中与 PAM 相互作用的一个或多个氨基酸替换为带正电的氨基酸；（2）将所述野生型 Cas12i2 核酸酶中参与打开 DNA 双链的一个或多个氨基酸替换为带芳香环的氨基酸；（3）将所述野生型 Cas12i2 核酸酶中位于 RuvC 结构域并与单链 DNA 底物相互作用的一个或多个氨基酸替换为带正电的氨基酸；和/或（4）将所述野生型 Cas12i2 核酸酶中与 DNA-RNA 双螺旋相互作用的一个或多个氨基酸替换为带正电的氨基酸，其中，所述野生型 Cas12i2 核酸酶的氨基酸序列如 SEQ ID NO: 92 所述。

在一些实施方案中，其中所述与 PAM 相互作用的一个或多个氨基酸位于下述的一个或多个位置：176、238、447、和/或 563；且其中所述位置根据 SEQ ID NO:92 编号。

在一些实施方案中，其中（1）所述带正电的氨基酸是 R 或 K。

在一些实施方案中，其中所述 Cas12i 核酸酶包含下述任何一种突变或突变组合：（1）E563R；（2）E176R、T447R、E176R 和 E563R；（3）K238R 和 E563R；（4）E176R、K238R 和 T447R；（5）E176R、K238R 和 E563R；（6）E176R、T447R 和 E563R；和/或（7）E176R、K238R、T447R 和 E563R；且其中氨基酸残基根据 SEQ ID NO:92 编号。

在一些实施方案中，其中所述参与打开 DNA 双链的一个或多个氨基酸位于下述的一个

或多个位置：163 和/或 164；且其中所述位置根据 SEQ ID NO:92 编号。

在一些实施方案中，其中（2）所述带芳香环的氨基酸是 F、Y 或 W。

在一些实施方案中，其中所述参与打开 DNA 双链的一个或多个氨基酸替换为带芳香环的氨基酸是指：Q163F、Q163Y、Q163W、和/或 N164F 或 N164Y；且其中氨基酸残基根据 SEQ ID NO:92 编号。

在一些实施方案中，其中所述位于 RuvC 结构域并与单链 DNA 底物相互作用的一个或多个氨基酸位于下述的一个或多个位置：323、362、425、925、926、391、424 和/或 929；且其中所述位置根据 SEQ ID NO:92 编号。

在一些实施方案中，其中（3）所述带正电的氨基酸是 R 或 K。

在一些实施方案中，其中所述 Cas12i 核酸酶包含下述任何一种突变或突变组合：（1）E323R；（2）D362R；（3）Q425R；（4）N925R；（5）I926R；（6）E323R 和 D362R；（7）E323R 和 Q425R；（8）E323R 和 I926R；（9）Q425R 和 I926R；（10）D362R 和 I926R；（11）N925R 和 I926R；（12）E323R、D362R 和 Q425R；（13）E323R、D362R 和 I926R；（14）E323R、Q425R 和 I926R；（15）D362R、N925R 和 I926R；和/或（16）E323R、D362R、Q425R 和 I926R；且其中氨基酸残基根据 SEQ ID NO:92 编号。

在一些实施方案中，其中所述与 DNA-RNA 双螺旋相互作用的一个或多个氨基酸位于下述的一个或多个位置：116、117、159、161、319、343 和/或 958；且其中所述位置根据 SEQ ID NO:92 编号。

在一些实施方案中，其中（4）所述带正电的氨基酸是 R 或 K。

在一些实施方案中，其中所述 Cas12i 核酸酶包含下述任何一种突变或突变组合：G116R、E117R、T159R、S161R、E319R、E343R、和/或 D958R；且其中氨基酸残基根据 SEQ ID NO:92 编号。

在一些实施方案中，所述 Cas12i 核酸酶进一步包含一个或多个柔性区的突变，所述突变能够使所述 Cas12i 核酸酶的柔性区相比于所述野生型 Cas12i2 核酸酶具有增加的柔性。

在一些实施方案中，所述柔性区选自对应于以下的区的组：氨基酸残基 228-232、氨基酸残基 439-443、氨基酸残基 478-482、氨基酸残基 500-504、氨基酸残基 775-779 和氨基酸残基 925-929；且其中所述氨基酸残基根据 SEQ ID NO: 92 编号。

在一些实施方案中，所述一个或多个柔性区的突变包括在所述柔性区中插入一个或多个 G 残基。

在一些实施方案中，所述一个或多个 G 残基插入在所述柔性区中柔性氨基酸残基的 N 末

端，其中所述柔性氨基酸残基选自下组：G、S、N、D、H、M、T、E、Q、K、R、A 和 P。

在一些实施方案中，所述一个或多个柔性区的突变包括用 G 残基替换所述柔性区中的疏水氨基酸残基，其中所述疏水氨基酸残基选自下组：L、I、V、C、Y、F 和 W。

在一些实施方案中，其中所述一个或多个柔性区的突变为：(1) I926G；和/或(2) 439G 或 439GG，且其中氨基酸残基根据 SEQ ID NO:92 编号。

在一些实施方案中，所述 Cas12i 核酸酶包含突变：E176R、K238R、T447R、E563R、N164Y、E323R 和 D362R；且其中氨基酸残基根据 SEQ ID NO:92 编号。

在一些实施方案中，所述 gRNA 分子包含 SEQ ID NOs: 67-78 中任一项所述的核苷酸序列。

在一些实施方案中，所述核酸酶的 N 端和/或 C 端与一个或一个以上的核定位信号连接。

在一些实施方案中，所述核定位信号与 His 标签连接。

在一些实施方案中，所述核酸酶包含选自 SEQ ID NOs: 1-11 中任一项所述的氨基酸序列。

另一方面，本申请提供一种组合物，其包含：(1) 一种或多种如权利要求 1-11 中任一项所述的 gRNA 分子，和 CRISPR/Cas 系统的核酸酶；(2) 编码一种或多种如权利要求 1-11 中任一项所述的 gRNA 分子的核酸，和 CRISPR/Cas 系统的核酸酶；(3) 一种或多种如权利要求 1-11 中任一项所述的 gRNA 分子，和编码 CRISPR/Cas 系统的核酸酶的核酸；或(4) 编码一种或多种如权利要求 1-11 中任一项所述的 gRNA 分子的核酸，和编码 CRISPR/Cas 系统的核酸酶的核酸。

在一些实施方案中，其(1)中所述一种或多种 gRNA 分子和所述 CRISPR/Cas 系统的核酸酶存在于核糖核蛋白(RNP)复合物中。

在一些实施方案中，所述核酸酶为 II 类 V 型 Cas 酶。

在一些实施方案中，所述核酸酶为 Cas12b 核酸酶。

在一些实施方案中，所述 Cas12b 核酸酶包含其变体、其直系同源物和/或其功能活性片段。

在一些实施方案中，所述 Cas12b 核酸酶为来自 *Alicyclobacillus acidiphilus* 的 Cas12b 核酸酶(AaCas12b)。

在一些实施方案中，所述 AaCas12b 包含下列相对于野生型 AaCas12b 核酸酶的突变中的一种或多种：(1) 用带正电荷的氨基酸残基取代所述野生型 AaCas12b 核酸酶中与前间隔序列邻近基序(PAM)相互作用的一个或多个氨基酸残基；(2) 用具有芳环的氨基酸残基取代所述野生型 AaCas12b 核酸酶中参与打开 DNA 双链的一个或多个氨基酸残基；和/或(3) 用带正电

荷的氨基酸残基或疏水性氨基酸残基取代所述野生型 AaCas12b 核酸酶的 RuvC 结构域中与单链 DNA 底物相互作用的一个或多个氨基酸残基，其中，所述野生型 AaCas12b 核酸酶的氨基酸序列如 SEQ ID NO: 91 所述。

在一些实施方案中，所述与 PAM 相互作用的所述一个或多个氨基酸残基位于以下的一个或多个位置：116、123、130、132、144、145、153、173、222、395、400 和 475；且其中所述氨基酸残基根据 SEQ ID NO: 91 编号。

在一些实施方案中，所述带正电荷的氨基酸残基取代所述野生型 AaCas12b 核酸酶中与 PAM 相互作用的一个或多个氨基酸残基是以下取代中的一种或多种：D116R 和 E475R；且其中所述氨基酸残基根据 SEQ ID NO: 91 编号。

在一些实施方案中，所述参与打开所述 DNA 双链的一个或多个氨基酸残基位于以下的一个或多个位置：118，和 119；且其中所述氨基酸残基根据 SEQ ID NO: 87 编号。

在一些实施方案中，所述用具有芳环的氨基酸残基取代所述野生型 AaCas12b 核酸酶中参与打开所述 DNA 双链的一个或多个氨基酸残基是 Q119Y、Q119F 或 Q119W 取代；且其中所述氨基酸残基根据 SEQ ID NO: 91 编号。

在一些实施方案中，所述位于 RuvC 结构域中并与单链 DNA 底物相互作用的一个或多个氨基酸残基位于以下的一个或多个位置：300、301、304、329、636、639、647、682、757、758、761、764、768、852、854、856、857、858、860、862、863、865、866、867、869、938、956、957、958、994、1093 和 1097；且其中所述氨基酸残基根据 SEQ ID NO: 91 编号。

在一些实施方案中，所述取代所述野生型 AaCas12b 核酸酶中位于 RuvC 结构域中并与单链 DNA 底物相互作用的一个或多个氨基酸残基是以下取代中的一种或多种：E636R、Q639R、T647R、Q682R、I757R、E758R、E761R、Q854R、N857K、D858R、I994R、Q1093R 和 W1097R；且其中所述氨基酸残基根据 SEQ ID NO: 91 编号。

在一些实施方案中，所述 gRNA 分子包含 SEQ ID NOs: 12-59 中任一项所述的核苷酸序列。

在一些实施方案中，所述核酸酶为 Cas12i 核酸酶。

在一些实施方案中，所述 Cas12i 核酸酶包含其变体、其直系同源物和/或其功能活性片段。

在一些实施方案中，所述 Cas12i 核酸酶包含下列相对于野生型 Cas12i2 核酸酶的突变中的一种或多种：（1）将所述野生型 Cas12i2 核酸酶中与 PAM 相互作用的一个或多个氨基酸替换为带正电的氨基酸；（2）将所述野生型 Cas12i2 核酸酶中参与打开 DNA 双链的一个或多个

氨基酸替换为带芳香环的氨基酸；(3) 将所述野生型 Cas12i2 核酸酶中位于 RuvC 结构域并与单链 DNA 底物相互作用的一个或多个氨基酸替换为带正电的氨基酸；和/或(4) 将所述野生型 Cas12i2 核酸酶中与 DNA-RNA 双螺旋相互作用的一个或多个氨基酸替换为带正电的氨基酸，其中，所述野生型 Cas12i2 核酸酶的氨基酸序列如 SEQ ID NO: 92 所述。

在一些实施方案中，其中所述与 PAM 相互作用的一个或多个氨基酸位于下述的一个或多个位置：176、238、447、和/或 563；且其中所述位置根据 SEQ ID NO:92 编号。

在一些实施方案中，其中(1) 所述带正电的氨基酸是 R 或 K。

在一些实施方案中，其中所述 Cas12i 核酸酶包含下述任何一种突变或突变组合：(1) E563R；(2) E176R、T447R、E176R 和 E563R；(3) K238R 和 E563R；(4) E176R、K238R 和 T447R；(5) E176R、K238R 和 E563R；(6) E176R、T447R 和 E563R；和/或(7) E176R、K238R、T447R 和 E563R；且其中氨基酸残基根据 SEQ ID NO:92 编号。

在一些实施方案中，其中所述参与打开 DNA 双链的一个或多个氨基酸位于下述的一个或多个位置：163 和/或 164；且其中所述位置根据 SEQ ID NO:92 编号。

在一些实施方案中，其中(2) 所述带芳香环的氨基酸是 F、Y 或 W。

在一些实施方案中，其中所述参与打开 DNA 双链的一个或多个氨基酸替换为带芳香环的氨基酸是指：Q163F、Q163Y、Q163W、和/或 N164F 或 N164Y；且其中氨基酸残基根据 SEQ ID NO: 92 编号。

在一些实施方案中，其中所述位于 RuvC 结构域并与单链 DNA 底物相互作用的一个或多个氨基酸位于下述的一个或多个位置：323、362、425、925、926、391、424 和/或 929；且其中所述位置根据 SEQ ID NO:92 编号。

在一些实施方案中，其中(3) 所述带正电的氨基酸是 R 或 K。

在一些实施方案中，其中所述 Cas12i 核酸酶包含下述任何一种突变或突变组合：(1) E323R；(2) D362R；(3) Q425R；(4) N925R；(5) I926R；(6) E323R 和 D362R；(7) E323R 和 Q425R；(8) E323R 和 I926R；(9) Q425R 和 I926R；(10) D362R 和 I926R；(11) N925R 和 I926R；(12) E323R、D362R 和 Q425R；(13) E323R、D362R 和 I926R；(14) E323R、Q425R 和 I926R；(15) D362R、N925R 和 I926R；和/或(16) E323R、D362R、Q425R 和 I926R；且其中氨基酸残基根据 SEQ ID NO:92 编号。

在一些实施方案中，其中所述与 DNA-RNA 双螺旋相互作用的一个或多个氨基酸位于下述的一个或多个位置：116、117、159、161、319、343 和/或 958；且其中所述位置根据 SEQ ID NO:92 编号。

在一些实施方案中，其中（4）所述带正电的氨基酸是 R 或 K。

在一些实施方案中，其中所述 Cas12i 核酸酶包含下述任何一种突变或突变组合：G116R、E117R、T159R、S161R、E319R、E343R、和/或 D958R；且其中氨基酸残基根据 SEQ ID NO:92 编号。

在一些实施方案中，所述 Cas12i 核酸酶进一步包含一个或多个柔性区的突变，所述突变能够使所述 Cas12i 核酸酶的柔性区相比于所述野生型 Cas12i2 核酸酶具有增加的柔性。

在一些实施方案中，所述柔性区选自对应于以下的区的组：氨基酸残基 228-232、氨基酸残基 439-443、氨基酸残基 478-482、氨基酸残基 500-504、氨基酸残基 775-779 和氨基酸残基 925-929；且其中所述氨基酸残基根据 SEQ ID NO: 92 编号。

在一些实施方案中，所述一个或多个柔性区的突变包括在所述柔性区中插入一个或多个 G 残基。

在一些实施方案中，所述一个或多个 G 残基插入在所述柔性区中柔性氨基酸残基的 N 末端，其中所述柔性氨基酸残基选自下组：G、S、N、D、H、M、T、E、Q、K、R、A 和 P。

在一些实施方案中，所述一个或多个柔性区的突变包括用 G 残基替换所述柔性区中的疏水氨基酸残基，其中所述疏水氨基酸残基选自下组：L、I、V、C、Y、F 和 W。

在一些实施方案中，其中所述一个或多个柔性区的突变为：（1）I926G；和/或（2）439G 或 439GG，且其中氨基酸残基根据 SEQ ID NO:92 编号。

在一些实施方案中，所述 Cas12i 核酸酶包含突变：E176R、K238R、T447R、E563R、N164Y、E323R 和 D362R；且其中氨基酸残基根据 SEQ ID NO:92 编号。

在一些实施方案中，所述 gRNA 分子包含 SEQ ID NOs: 67-78 中任一项所述的核苷酸序列。

在一些实施方案中，所述核酸酶的 N 端和/或 C 端与一个或一个以上的核定位信号连接。

在一些实施方案中，所述核定位信号与 His 标签连接。

在一些实施方案中，所述核酸酶包含选自 SEQ ID NOs: 1-11 中任一项所述的氨基酸序列。

在一些实施方案中，所述组合物还任选地包含药学上可接受的载体。

另一方面，本申请提供一种细胞，其包含本申请所述的 gRNA 分子、本申请所述的核酸分子、本申请所述的载体、本申请所述的核糖核蛋白（RNP）复合物、和/或本申请所述的组合物。

另一方面，本申请提供细胞群或其后代，其包含施用本申请所述的核糖核蛋白（RNP）复合物、本申请所述的组合物、本申请所述的细胞所产生的基因产物。

在一些实施方案中，所述基因产物包含被修饰的 HBG 基因。

在一些实施方案中，所述修饰为位于靶序列处或附近的碱基插入或缺失 (indel)；其中所述靶序列位于 HBG1 启动子区域内或 HBG2 启动子区域内，所述 HBG1 启动子区域位于人基因组 hg19 的 chr11: 5,248,269 至 5,249,857 之间，所述 HBG2 启动子区域位于人基因组 hg19 的 chr11: 5,253,188-5,254,781 之间。

在一些实施方案中，所述靶序列位于所述 HBG1 和所述 HBG2 的转录起始位点上游约 210bp 至上游约 100bp 之间。

在一些实施方案中，所述细胞群能够分化成为红系细胞谱系的分化细胞，且所述分化细胞相比于 HBG 基因未经修饰的细胞群具有增加的胎儿血红蛋白表达水平。

另一方面，本申请提供一种试剂盒，其包含本申请所述的 gRNA 分子、本申请所述的核酸分子、本申请所述的载体、本申请所述的核糖核蛋白 (RNP) 复合物、本申请所述的组合物、本申请所述的细胞、和/或本申请所述的细胞群或其后代。

另一方面，本申请提供一种调控细胞、细胞群或其后代中胎儿血红蛋白 (HbF) 表达的方法，所述方法包括施用本申请所述的核糖核蛋白 (RNP) 复合物、本申请所述的组合物、本申请所述的细胞、本申请所述的细胞群或其后代、和/或本申请所述的试剂盒。

在一些实施方案中，所述细胞或细胞群为造血干/祖细胞 (HSPC)。

在一些实施方案中，所述细胞或细胞群为 CD34+HSPC。

另一方面，本申请提供一种治疗、预防或缓解血红蛋白病和/或其相关病症的方法，所述方法包括向有需要的受试者施用有效量的本申请所述的核糖核蛋白 (RNP) 复合物、本申请所述的组合物、本申请所述的细胞、本申请所述的细胞群或其后代、和/或本申请所述的试剂盒。

在一些实施方案中，所述血红蛋白并和/或其相关病症选自：镰状细胞病、镰状细胞性贫血、血红蛋白 C 病、血红蛋白 C 性状、血红蛋白 S/C 病、血红蛋白 D 病、血红蛋白 E 病、地中海贫血、氧亲和力和下降血红蛋白病、和不稳定血红蛋白病。

另一方面，本申请提供本申请所述的 gRNA 分子、本申请所述的核酸分子、本申请所述的载体、本申请所述的核糖核蛋白 (RNP) 复合物、本申请所述的组合物、本申请所述的细胞、和/或本申请所述的细胞群或其后代用于制备药物的用途，所述药物用于治疗血红蛋白病和/或其相关病症。

另一方面，本申请提供本申请所述的 gRNA 分子、本申请所述的核酸分子、本申请所述的载体、本申请所述的核糖核蛋白 (RNP) 复合物、本申请所述的组合物、本申请所述的细

胞、本申请所述的细胞群或其后代、和/或本申请所述的试剂盒，其用于治疗血红蛋白病和/或其相关病症。

本领域技术人员能够从下文的详细描述中容易地洞察到本申请的其它方面和优势。下文的详细描述中仅显示和描述了本申请的示例性实施方式。如本领域技术人员将认识到的，本申请的内容使得本领域技术人员能够对所公开的具体实施方式进行改动而不脱离本申请所涉及发明的精神和范围。相应地，本申请的附图和说明书中的描述仅仅是示例性的，而非为限制性的。

附图说明

本申请所涉及的发明的具体特征如所附权利要求书所显示。通过参考下文中详细描述 of 示例性实施方式和附图能够更好地理解本申请所涉及发明的特点和优势。对附图简要说明书如下：

图 1A 显示的是 HBG 启动区域的关键模序分布以及 gRNA 靶向位置。图 1B-1C 显示的是结合区域长度为 20nt 的 gRNA 筛选结果；图 1D-1E 显示的是不同结合区域长度的 crRNA 和 sgRNA 筛选结果。

图 2A 显示的是八种 Cas12b^{Max} 构建体的结构示意图。图 2B-2C 显示的是 Cas12b^{Max} 构建体 M1-M7 在 sgRNA1 引导下的编辑效率和 HbF+红细胞诱导效率。图 2D 显示的是 Cas12b^{Max} 构建体 M1、M2 相比于 SpCas9 对于 HBG 启动区域的剪切活性。图 2E 显示不同供应商来源的构建体 M1 与构建体 M8，在不同 RNP 剂量条件下的对造血干细胞目标基因编辑效率。

图 3A-3C 显示的是引导 Cas12b^{Max} 的 sgRNA 单分子骨架的优化方案以及筛选结果。

图 4A-4D 显示的分别是 Cas12b^{Max}-sgRNA G1 复合物 (RNP) 编辑造血干细胞时，RNP 剂量与 Cas 酶：sgRNA 比例的筛选结果、健康对照与地贫患者的造血干细胞编辑效率、目标基因组区域的主要编辑序列以及编辑序列长度的分布统计结果。

图 5A-5B 显示的是 Cas12b^{Max}-sgRNA G1 复合物 (RNP) 编辑造血干细胞时的脱靶效应评估。

图 6A-6D 显示的是 Cas12b^{Max}-sgRNA G1 复合物 (RNP) 对 γ -珠蛋白以及胎儿血红蛋白的 mRNA 水平、HbF 阳性红细胞水平和蛋白水平的诱导结果。

图 7A-7B 显示的是 Cas12b^{Max}-sgRNA G1 复合物 (RNP) 体外编辑造血干细胞时对其分化潜能的影响和对细胞固有免疫相关的影响。

图 8A-8F 显示的时 Cas12b^{Max}-sgRNA G1 复合物 (RNP) 编辑的造血干细胞的体内药效评

估。

具体实施方式

以下由特定的具体实施例说明本申请发明的实施方式，熟悉此技术的人士可由本说明书所公开的内容容易地了解本申请发明的其他优点及效果。

术语定义

在本申请中，术语“向导 RNA”可以与“引导 RNA（分子）”、“gRNA（分子）”互换地使用，其通常是指促进将 RNA 引导的核酸酶或其他效应分子(通常与 gRNA 分子复合)特异性引导到靶序列上的一组核酸分子。在一些实施例中，通过将 gRNA 的部分与 DNA 杂交(例如通过 gRNA 靶向结构域)以及通过将 gRNA 分子的部分与 RNA 引导的核酸酶或其他效应分子结合(例如，至少通过 tracr RNA)来实现所述指导。在某些实施例中，gRNA 分子由单个连续的多核苷酸分子组成，在本申请中称为“单组分向导 RNA”或“sgRNA”等。在其他实施例中，gRNA 分子由多个、通常是两个多核苷酸分子组成，这些多核苷酸分子本身能够缔合，通常通过杂交，在本文中称为“双组分向导 RNA”或“dgRNA”等。gRNA 分子通常包含靶向结构域和 tracr。在实施例中，靶向结构域和 tracr 被布置于单个多核苷酸上。在其他实施例中，靶向结构域和 tracr 被布置于分开的多核苷酸上。

在本申请中，术语“靶序列”通常是指与 gRNA 靶向结构域互补，例如完全互补的核酸序列。在实施例中，靶序列被布置于基因组 DNA 上。在一个实施例中，靶序列与由具有核酸酶或其他效应子活性的蛋白质识别的前间区邻近基序(PAM)序列(例如由 Cas 酶识别的 PAM 序列)相邻(在 DNA 的相同链上或在 DNA 的互补链上)。在实施例中，靶序列是影响珠蛋白基因表达，例如影响 β -珠蛋白或胎儿血红蛋白(HbF)表达的基因或基因座内的靶序列。在实施例中，靶序列是非缺失型 HPFH 区内的例如在 HBG1 和/或 HBG2 启动子区内的靶序列。而在本申请中，术语“靶向结构域”(当该术语与 gRNA 结合使用时)是指 gRNA 分子的部分，其识别靶序列(例如细胞核酸内的靶序列，例如在基因内)，例如与其互补。

在本申请中，术语“crRNA”(当该术语与 gRNA 分子结合使用时)通常是指 gRNA 分子的部分，其包含靶向结构域和与 tracr 相互作用以形成 dgRNA 结构域。

在本申请中，术语“tracr”(当该术语与 gRNA 分子结合使用时)通常是指 gRNA 的与核酸酶或其他效应分子结合的部分。在实施例中，tracr 包含与 Cas 酶特异性结合的核酸序列。在实施例中，tracr 包含形成 dgRNA 的部分的核酸序列。

术语“CRISPR 系统”、“Cas 系统”或“CRISPR/Cas 系统”是指包含 RNA 指导的核酸酶或其他效应分子和 gRNA 分子的一组分子，它们一起对于向导和实现由 RNA 指导的核酸酶或其他效应分子在靶序列处对核酸的修饰是必需且足够的。在一个实施例中，CRISPR 系统包含 gRNA 和 Cas 蛋白(例如 Cas9 蛋白)，此时包含 Cas9 或经修饰的 Cas9 分子的此类系统被称为 Cas9 系统”或“CRISPR/Cas9 系统”。在一个实例中，gRNA 分子和 Cas 分子可以复合，以形成核糖核蛋白(RNP)复合物。

在本申请中，术语“indel”通常是指相对于参考核酸包含一个或多个核苷酸插入、一个或多个核苷酸缺失、或核苷酸插入和缺失的組合的核酸，其在暴露于包含 gRNA 分子的组合物(例如 CRISPR 系统)之后产生。插入/缺失可以通过在暴露于包含 gRNA 分子的组合物之后对核酸进行测序来确定，例如通过 NGS。关于插入/缺失的位点，如果插入/缺失包含参考位点的约 10、约 20、约 30、约 40、约 50、约 60、约 70、约 80、约 90 或约 100 个核苷酸内的至少一个插入或缺失，或者插入/缺失与所述参考位点的部分或全部重叠(例如，包含至少一个与如本文所述的 gRNA 分子的靶向结构域互补的位点的约 10、约 20、约 30、约 40、约 50、约 60、约 70、约 80、约 90 或约 100 个核苷酸重叠的插入或缺失，或至少一个在如本文所述的 gRNA 分子的靶向结构域互补的位点的约 10、约 20、约 30、约 40、约 50、约 60、约 70、约 80、约 90 或约 100 个核苷酸内的插入或缺失)，则称插入/缺失位于参考位点(例如，与 gRNA 分子的靶向结构域互补的位点)“处或附近”。例如，本申请所述的 indel 可以是在靶序列处或附近插入 10 个、11 个、12 个、13 个、14 个、15 个、16 个、17 个、18 个、19 个或 20 个核苷酸；或者可以是在靶序列处或附近缺失 10 个、11 个、12 个、13 个、14 个、15 个、16 个、17 个、18 个、19 个、20 个、21 个、22 个、23 个、24 个、25 个、26 个、27 个、28 个、29 个、30 个、31 个、32 个、33 个、34 个、35 个、36 个、37 个、38 个、39 个、40 个、41 个、42 个、43 个、44 个、45 个、46 个、47 个、48 个、49 个、50 个、51 个、52 个、53 个、54 个、55 个、56 个、57 个、58 个、59 个或 60 个核苷酸。

在本申请中，术语“功能活性片段”通常是指具有全长蛋白质或核酸的部分区域，但保留或部分保留全长蛋白质或核酸的生物活性或功能的片段。例如，功能活性片段可以保留或部分保留全长蛋白质结合另一种分子的能力。

在本申请中，术语“核定位序列”或“核定位信号”或“NLS”通常是指将蛋白导向至细胞核的肽。例如，NLS 可以位于肽链上的任何位置。NLS 在长度和/或序列上可以不同，并且很多特异性 NLS 序列已经有说明。例如，现已发现，包含带正电荷的氨基酸(尤其是赖氨酸(K)、精氨酸(R)和/或组氨酸(H))的肽一般可以用作 NLS。因此，一个示例性的 NLS 可以是

例如 4 至 20 个氨基酸、更具体而言 4 至 15 个、4 至 12 个、4 至 10 个或 4 至 8 个氨基酸的肽，其中至少 4 个氨基酸(更具体而言是 NLS 肽中的 60%、70%、75%、80%、85%或 90% 的氨基酸残基)是带有正电荷的氨基酸，这些氨基酸优选选自 K、R 或 H。可用于结合本申请所述的核酸酶和/或其构建体的 NLS 序列为本领域中已知。此类 NLS 序列的非限制性实例包括具有如下氨基酸序列的核质蛋白 (nucleoplasmin) NLS (KRPAATKKAGQAKKKK)、猿猴病毒 40“SV40”NLS (PKKKRKV)、c-Myc NLS、和/或 BP SV40 NLS。

在本申请中，术语“His 标签”通常是指一种由 6 至 10 个组氨酸残基组成的表位标签，很容易被标签特异性抗体识别。His 标签是蛋白纯化和检测的常用标签之一，由于其分子量小 (只有 0.84KD)，融合到重组蛋白后，不会影响标记蛋白质的结构和生化特性。His 抗体可以用于检测和 His 标签融合表达蛋白的表达、细胞内定位，以及纯化、定性或定量检测 His 融合表达蛋白等，应用十分广泛。

在本申请中，术语“核酸”与“多核苷酸”、“核苷酸”、“核苷酸序列”和“寡核苷酸”可互换地使用，其通常是指核苷酸 (例如，脱氧核糖核苷酸或核糖核苷酸) 和其呈单链、双链或多链形式的聚合物或其互补物。例如，核苷酸可以为核糖核苷酸、脱氧核糖核苷酸或其修饰版本。例如，核苷酸可以为单链和双链 DNA、单链和双链 RNA 以及具有单链和双链 DNA 和 RNA 的混合物的杂交分子。例如，核苷酸可以包括但不限于任何类型的 RNA，例如 mRNA、siRNA、miRNA、sgRNA 和引导 RNA，以及任何类型的 DNA、基因组 DNA、质粒 DNA 和微环 DNA 以及其任何片段。所述术语还涵盖含有已知核苷酸类似物或经修饰的主链残基或键的核酸，所述核酸为合成的、天然存在的和非天然存在的。

在本申请中，术语“编码……的序列”或“编码……的核酸”通常是指包含编码蛋白质的核苷酸序列的核酸 (RNA 或 DNA 分子)。编码序列还可包括与调控元件可操作地连接的起始和终止信号，所述调控元件包含能够在对其施用了核酸的个体或哺乳动物的细胞中指导表达的启动子和多腺苷酸化信号。可对编码序列进行密码子优化。

在本申请中，术语“药学上可接受的载体”通常是指给予治疗剂，例如抗体或多肽、基因和其它治疗剂的载体。该术语指本身不诱导对接受组合物的个体有害的抗体产生并且可以给予而不产生过度毒性的任何药物载体。合适的载体可以是大的、代谢缓慢的大分子，例如蛋白质、多糖、聚乳酸、聚乙醇酸、多聚氨基酸、氨基酸共聚物、脂质聚集物和灭活的病毒颗粒。本领域技术人员熟知这些载体。治疗组合物中药学上可接受的载体可包括液体，例如水、盐水、甘油和乙醇。这些载体中也可存在辅助物质，例如润湿剂或乳化剂、pH 缓冲物质等。

在本申请中，术语“试剂盒”通常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)，根据本披露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。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。根据本披露的试剂盒还任选地包括使用说明书，其描述例如根据本申请的方法使用该试剂盒。使用说明书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使用说明书，例如通过电子方式获得。

在本申请中，术语“细胞群”通常是指从生物来源(例如血液制品或组织)分离并且衍生自多于一种细胞的真核哺乳动物(优选人)细胞。

在本申请中，术语“造血干细胞和祖细胞”或“HSPC”可互换使用，并且是指包含造血干细胞(“HSC”)和造血祖细胞(“HPC”)两者的细胞群。此类细胞被表征为例如 CD34+。在示例性实施例中，HSPC 自骨髓分离。在其他示例性实施例中，HSPC 自外周血分离。在其他示例性实施例中，HSPC 自脐带血分离。在一个实施例中，HSPC 被表征为 CD34+/CD38-/CD90+/CD45RA-。在实施例中，HSPC 被表征为 CD34+/CD90+/CD49f+细胞。在实施例中，HSPC 被表征为 CD34+细胞。在实施例中，HSPC 被表征为 CD34+/CD90+细胞。在实施例中，HSPC 被表征为 CD34+/CD90+/CD45RA-细胞。

在本申请中，术语“治疗”，例如疾病，意指当施用例如本文所述的复合物、组合物、细胞等时，与从未施用该复合物、组合物、细胞等时相比，患有疾病、有患疾病的风险和/或经历疾病症状的受试者(例如，人)在一个实施方案中将经历症状更轻和/或将更快恢复。在一些实施例中，“治疗”可指减少或缓解病症例如血红蛋白病的进展、严重性和/或持续时间，或者缓解病症例如血红蛋白病的一种或多种症状(优选地，一种或多种可辨别的症状)，这由施用一种或多种疗法(例如一种或多种治疗剂，如本发明的 gRNA 分子、CRISPR 系统或经修饰的细胞)引起。在具体的实施例中，术语“治疗”是指改善血红蛋白病病症的至少一种可测量的物理参数，这是患者无法辨别的。在其他实施例中，术语“治疗”是指通过例如稳定可辨别的症状来物理地，通过例如稳定物理参数来生理地，或通过两者，抑制病症的进展。在其他实施例中，术语“治疗”是指减少或稳定血红蛋白病例如镰状细胞疾病或 β -地中海贫血的症状。

在本申请中，术语“受试者”通常是指动物，通常是哺乳动物，诸如人、非人灵长类动物(猴、长臂猿、大猩猩、黑猩猩、猩猩、猕猴)、家畜(狗和猫)、农场动物(家禽如鸡和鸭、马、

牛、山羊、绵羊、猪)和实验动物(小鼠、大鼠、兔、豚鼠)。人受试者包括胎儿、新生儿、婴儿、青少年和成人受试者。受试者包括动物疾病模型,例如小鼠和血液凝固疾病(诸如 HeA)的其它动物模型,和本领域技术人员已知的其它动物模型。

在本申请中,术语“包含”通常是指包括明确指定的特征,但不排除其他要素。

在本申请中,术语“选自”通常是指包括选择的对象以及其所有组合。例如“选自(:)A、B 和 C”意指包括 A、B 和 C 的所有组合,例如, A、B、C、A+B、A+C、B+C 或 A+B+C。

不欲被任何理论所限,下文中的实施例仅仅是为了阐释本申请的融合蛋白、制备方法和用途等,而不适用于限制本申请发明的范围。

实施例

实验材料与方法:

(1) 细胞培养

人 CD34+HSPCs 取自去鉴定健康供者的动员外周血,β-地中海贫血患者 CD34+HSPCs 从普乐沙福动员外周血中分离。使用 Miltenyi CD34 Microbead 试剂盒(Miltenyi Biotec)富集 CD34+HSPCs。CD34+HSPCs 在 StemSpan™ SFEM II (Stemcell Technologies)中维持和扩增。电穿孔的 HSPCs 按照无血清悬浮三步法方案进行体外红细胞分化。红细胞表型和 γ-珠蛋白诱导在第 18 天进行评估。

用于集落形成单位检测的半固态细胞培养:将 CD34+ HSPCs 种子加入 MethoCult™培养基(Stemcell Technologies)并在培养皿中培养。在半固体培养的第 14 天,对 HSPC 衍生的菌落进行计数或摘取,以便进一步评估。

(2) RNP 电穿孔

化学合成的 sgRNAs (前 3 个核苷酸和后 3 个核苷酸经 2-O-甲基和 3'硫代磷酸酯修饰)购自 IDT Technologies。AaCas12b 变体由 Cellnuo Biotec 公司表达和纯化(原核表达序列 SEQ ID NOs:79-86 所示)。在用 20μl Nucleocuvette Strips 进行电穿孔时,将 Cas12b(60 至 180 pmol)和 sgRNA 或 dgRNA 以特定的比例混合并在室温下孵育 15 分钟,然后制备 RNP 复合物。将 5×10^4 至 5×10^5 CD34+HSPCs 重悬于 20 μl P3 溶液中并与 RNP 混合。将 20 μl HSPCs/RNP 混合物转移到比色皿中,用 DZ-100 程序进行电穿孔。将电穿孔细胞转移到 SFEM II 培养基中过夜,然后进行体外分化或移植实验。

(3) 用新一代测序技术(NGS)测量 Indel 频率

使用小鼠尾部 DNA 提取试剂盒(Beyotime, D7283M)提取 HSPC 基因组 DNA。用 Phanta Max Super-Fidelity DNA 聚合酶(Vazyme, P525-03)和引物以两步法扩增 HBG1 和 HBG2 转

录起始位点上游的启动子区。对 PCR 产物进行 NGS 测序。测序数据被导入 CRISPR RGEN Tools Cas-Analyzer 在线软件, 以 70 bp 的分解窗口进行 indel 频率测量和 indel 模式识别。

(4) RT-qPCR 定量

使用 FastPure Cell/Tissue Total RNA Isolation Kit V2 (Vazyme, RC112-01) 分离 HSPCs 的总 RNA, 然后使用 HiScript II 1st Strand cDNA Synthesis Kit (Vazyme, R212-01) 进行反转录。使用 AceQ Universal SYBR qPCR Aster Mix (Vazyme, Q511-02) 进行 RT- qPCR。

(5) 用流式细胞仪测量 HbF+细胞

用人 CD235a+抗体 (克隆 GA-R2 with PE; BD Pharmingen) 染色后, 用 4%多聚甲醛固定体外分化的细胞 10 分钟, 然后用 0.1% Triton X-100 在室温下渗透 5 分钟。随后, 用人 HbF 抗体 (克隆 HBF-1 与 APC 结合; 购自 Invitrogen 公司) 在黑暗中染色细胞 30 分钟。在进行 FACS 分析之前, 先将细胞清洗两次以去除未结合的抗体。

(6) 用 HPLC 测定血红蛋白

使用溶血试剂从 CD34+HSPCs 体外分化后的红细胞制备溶血。裂解物用 D-10 血红蛋白分析仪 (Bio-Rad) 分析, 并使用临床校准的人类血红蛋白做为标准品。

实施例 1

Cas12b/Cas12i 构建体及其 gRNA 靶向结合区域的筛选

将 Cas12b^{Max} 或 Cas12i2^{Max} 的表达质粒和 gRNA 的表达质粒共转染到 HEK293 细胞中, 以此筛选出能够有效靶向位于 HBG1 和 HBG2 TSS 上游顺式元件的 gRNA, 来用于切换成熟红细胞中 β/γ 珠蛋白的表达。图 1A 显示了 CRISPR-Cas12 DNA 内切酶在 HBG1 和 HBG2 (人类 γ -珠蛋白基因座) 启动子区域产生 HPFH 样突变的目标位点的基因组位置, 其中预期的切割位点用箭头标出。Cas12b^{Max} 的 sgRNA 序列筛选结果如 SEQ ID NOs:12-59 所示, 图 1B 则显示 Cas12b^{Max} gRNA 在 293HEK 细胞系中筛选出的针对 HBG1/2 原基区域的高效编辑 gRNA。Cas12b^{Max} 双分子 gRNA 由 tracrRNA 和 crRNA 组成, 它们的序列筛选结果如 SEQ ID NOs:60-66 所示, 图 1C 则显示 Cas12i2^{Max} gRNA 在 293HEK 细胞系中针对 HBG1/2 原体区域进行高效编辑的筛选结果。Cas12i2^{Max} 的 sgRNA 序列筛选结果如 SEQ ID NOs:67-78 所示。

另外, 为了提高负责红细胞 γ -珠蛋白表达的目标基因组的编辑效率, 利用体外转录产生的不同间隔长度的单分子 gRNA 或双分子 gRNA 来优化 RNP 介导的对人原代 CD34+HSPCs 中 HBG1/2 启动子区 LRF 结合模序的破坏。NGS 数据显示, 在 RNP 电穿孔的 HSPC 中, 靶标 indel 频率与 sgRNA 的间隔长度有关 (图 1D 和 1E)。

进一步优化引入了一组工程化的 Cas 蛋白构建体, 其末端的 NLS 有不同的排列和组合。

这些构建体的氨基酸序列被识别为 SEQ ID NOs:1-7。实验结果显示,将包含不同 NLS 排列组合的 RNPs 电穿孔到 HSPC 中,在 3 μ M 剂量下,其对目标基因组的编辑效率各不相同,从 30%到 85%不等(图 2B)。电穿孔 HSPCs 体外分化的红细胞之间通过 FACS 检测到类似的 HbF 诱导幅度(图 2C)。本实施例还选取了对靶基因组编辑效率较高的 M1 和 M2 构建体与 SpCas9 进行了对比,结果显示本申请提供的 RNP 要优于 Cas9 的编辑效率(图 2D)。基于上述蛋白层面的结果与 RNA 层面的结果,并结合比较 Cas 蛋白纯化方案的回收率,我们进一步引入了优化的 Cas 蛋白构建体 M8,它的氨基酸序列被识别为 SEQ ID NO:8。本实施例选取了 M1 构建体与新构建体 M8 进行比较,结果显示在不同的 RNP 电转剂量条件下,M1 与 M8 的 RNP 对供试造血干细胞有类似的目标基因编辑效率。(图 2E)

实施例 2

Cas12b^{Max} 的 sgRNA 单分子骨架的优化和筛选结果

本实施例测试了靶向 HBG 启动区域的 Cas12b^{Max} 不同 sgRNA 骨架的 Cas12b^{Max} 复合物在体外的剪切活性(图 3A),以及不同 Artsg13 骨架序列的末端延长方案,对于靶向 HBG 启动区域的 Cas12b^{Max} 复合物体外剪切活性存在不同影响(图 3B)。最终得到优选的 Artsg13 骨架序列末端延长方案(双端延长 113 nt,或单端延长 116 nt,与未优化方案的编辑效率对比如图 3C 所示。图 3C 的 110nt-nm 序列以及 113nt 序列如 SEQ ID NOs: 13 与 90 (SEQ ID NO: 90 所示序列经化学修饰: mG*mG*mT*CGTCTATAGGACGGCGAGTTTTTCAACGGGTGTGCCAA TGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCAAAGAAGTGGCACAGATAGTGT GGGGAAGGGGC*mC*mC*mC) 所示。

实施例 3

选定的 Cas12b^{Max} 与 sgRNA-G1 对造血干细胞的基因编辑

在电穿孔 HSPCs 实验中,设置 Cas : sgRNA 摩尔比为 1:1 至 1:3 时,观察不同 RNP 剂量对编辑效率的依赖性影响。观察结果显示,当 RNP 剂量为 180 pmol, Cas : sgRNA 摩尔比为 1:3 时,目标 LRF 结合基序上的 indel 频率超过 85%(图 4A)。NGS 结果表明,在来自健康供体和 β -地中海贫血($\beta^0\beta^0$)患者的 CD34+ HSPCs 中都实现了高效编辑(图 4B),并显示最常见的 indel 是长度超过 10bp 的缺失,由微同源介导的末端连接修复(MMEJ)产生(图 4C 和图 4D)。

实施例 4

选定的 Cas12b^{Max} 与 sgRNA-G1 对造血干细胞的基因编辑脱靶效应评估

为了评估 Cas12b^{Max}-RNP 高效靶上编辑的特异性,本实施例采用了两种无偏见的全基因组方法,即改良的 Site-seq 和 Tag-seq,分别筛选裸基因组 DNA 和细胞环境中易被 RNP 切除的非靶序列;并结合生物信息学预测,共发现了 4 个潜在的脱靶位点(OTS)与脱靶序列筛选出的列表重叠。这 4 个潜在的脱靶位点和 23 个靶上序列高度相似的位点在高度编辑的 CD34⁺ HSPC 中进行了扩增片段深度测序验证,其靶向编辑 indel 频率 $\geq 90\%$ 。每个候选 OTS 都没有检测到 RNP 介导的突变,其 indel 频率低于 NGS 检测限的 0.1% (图 5B)。总之,这些数据表明,本申请提供的 CD34⁺ HSPCs 编辑方法对目标序列具有高度特异性,且未具检测出遗传或基因组相关毒性。

脱靶效应分析方法:

Site-seq 实验: 简言之,纯化的基因组 DNA 在反应缓冲液中被 Cas12b^{Max}-RNP 消化。消化后的 gDNA 与生物素化、粘性末端、Illumina 兼容的适配体杂交、连接。适配体连接的 gDNA 被片段化、经末端修复并连接到第二个适配体上。纯化双适配体连接的 DNA 后,进行两步 PCR,构建测序文库。测序数据由 Perl 软件包进行分析。

Tag-seq 实验: 用 lipo3000 将 Cas12b^{Max}、编码 gRNA 的质粒与 dsTag 共同转染 HEK293 细胞。纯化基因组 DNA,剪切至平均长度为 300 bp,末端修复,A 尾连接至半适配体,并加入 8-nt 随机分子标签。两轮嵌套锚定 PCR 被用来构建目标富集测序文库。测序数据由开源 Python 软件包进行分析。

为了通过 NGS 验证编辑 HSPC 的脱靶效应,采用两步 PCR 流程扩增潜在的脱靶位点。使用 MiSeq 测序系统(Illumina)对扩增子进行 2 $\times$ 150 成对末端读数测序。深度测序数据由离线 CRISPResso 软件进行分析。

实施例 5

选定的 Cas12b^{Max} 与 sgRNA-G1 对 γ 珠蛋白胎儿血红蛋白的诱导

为了测试 HBG 基因启动子区 LRF 结合基序的破坏是否会导致显著的 γ -珠蛋白再激活,用 RNPs 电穿孔后,将来自健康供体或 β -地中海贫血患者的 HSPCs 体外诱导成熟的红细胞。结果表明,与来自未编辑 HSPCs 的红细胞相比,HBG 启动子区 LRF 基序破坏率高的 HSPCs 红细胞后代中的 γ -珠蛋白 mRNA 水平升高, β -珠蛋白 mRNA 水平下降(图 6A)。通过高效液相色谱法(HPLC)或低倍细胞计数法分析 HbF 蛋白的表达。在体外生成的红细胞中,HBG

启动子区 LRF 结合基序的破坏显著提高了 F 细胞的数量 (HbF/CD235a 双阳性) (图 6B)。从结果还观察到, 从 HBG 启动子区 LRF 基序被破坏的红细胞中提取的总血红蛋白中, HbF 蛋白含量相比于健康供体的对照组显著升高 (图 6C 和 6D)。

实施例 6

Cas12b^{Max} 与 sgRNA-G1 基因编辑对造血干细胞的其他影响

通过集落形成单位检测分析电穿孔 Cas12bMax-RNP 的 CD34+ HSPCs 的多系分化潜能, 图 7A 的结果显示该基因编辑过程对造血干细胞分化潜能影响较小。另外, 通过 RT-PCR 分析电穿孔 Cas12bMax-RNP 后 CD34+HSPCs 中 P21 表达的动态, 如图 7B 所示。

实施例 7

Cas12b^{Max} 与 sgRNA-G1 编辑的造血干细胞的体内药效

将来自重型 β^0 -地中海贫血 ($\beta^0\beta^0$) 患者的经过编辑的 CD34+HSPCs 注入免疫缺陷的 NCG-X 小鼠体内, 以测试本申请的编辑方法对其移植能力和血细胞分化潜能的影响。免疫缺陷的 NCG-X 小鼠支持造血干细胞移植, 无需进行清髓药物处理。与天然 CD34+HSPCs 移植组相比, RNP 电穿孔 hCD34+ HSPCs 移植组在移植后外周循环中的 hCD45+细胞嵌合率较低。在第 20 周时, 两组受试者骨髓中的人类淋巴和骨髓 (hCD45+) 或红细胞 (hCD235a+) 重组水平相似 (图 8A、图 8B 和图 8D)。接受 RNP 电穿孔 hCD34+ HSPCs 的受体小鼠骨髓中的 hCD34+细胞数量较少 (图 8C)。在第 20 周, 经编辑的 hCD34+ HSPCs 输注小鼠骨髓中的 hCD235+细胞诱导了显著的 HbF+细胞 (图 8E)。与刚注入编辑过的 HSPCs 时相比, 从外周血或骨髓中提取的人类多系细胞中检测到的靶向诱变频率降低了 (图 8F)。

人 CD34+HSPC 的异种移植:

NCG-X 小鼠购自 GemPharatech 有限公司。未接受辐照的 NCG-X 雌性小鼠 (4-5 周龄) 通过尾静脉注射 $0.5-0.6 \times 10^6$ 来自健康供体的 CD34 + HSPCs。用于移植实验的 CD34 + HSPCs 在电穿孔 RNPs 后体外培养过夜。移植后 20 周, 分离骨髓进行人体细胞嵌合和 HbF+ 细胞分析。通过尾静脉注射来自初代受试者骨髓的 hCD34+HSPCs 进行第二次移植。在对骨髓进行流式细胞术分析时, 首先对骨髓细胞进行孵育, 采用: V450 小鼠抗-人 CD45 克隆 HI30 (560367, BD Biosciences)、PE-eFluor 610 mCD45 单克隆抗体 (30-F11) (61-0451-82, Thermo Fisher)、FITC 抗人 CD235a 抗体 (349104, BioLegend)、PE 抗人 CD33 抗体 (366608, BioLegend)、APC 抗人 CD19 抗体 (302212, BioLegend) 和用于活体/死体染色的可固定活力染料 eFluor 780 (650865-14, Thermo Fisher)。人体移植百分比的计算公式为 $\text{hCD45+cells} / (\text{hCD45+cells} +$

mCD45+cells)×100。为了评估人 CD34+HSPCs 的移植能力，将 CD34+ HSPCs 与抗人 CD34 抗体（343512，Biolegend）、抗小鼠 CD34（303508，Biolegend）和染料 eFluor 780（650865-14，Thermo Fisher）孵育，进行活体/死体染色。分析 HbF+细胞是通过使用 hCD234a+ 微珠试剂盒（Miltenyi）从骨髓中分离出人 CD235a+细胞。HbF+细胞的 FACS 分析方法与上述流式细胞仪测量 HbF+细胞的方法相同。

权 利 要 求 书

1. 一种指导 RNA (gRNA) 分子, 所述 gRNA 分子包含与靶序列互补的靶向结构域, 所述靶序列位于 HBG1 启动子区域内或 HBG2 启动子区域内, 所述 HBG1 启动子区域位于人基因组 hg19 的 chr11: 5,248,269 至 5,249,857 之间, 所述 HBG2 启动子区域位于人基因组 hg19 的 chr11: 5,253,188-5,254,781 之间。
2. 根据权利要求 1 所述的 gRNA 分子, 所述靶序列位于所述 HBG1 和所述 HBG2 的转录起始位点上游约 210bp 至上游约 100bp 之间。
3. 根据权利要求 1-2 中任一项所述的 gRNA 分子, 其包含 tracr 序列和 crRNA 序列。
4. 根据权利要求 1-3 中任一项所述的 gRNA 分子, 其为双组分向导 RNA 分子。
5. 根据权利要求 1-3 中任一项所述的 gRNA 分子, 其为单组分向导 RNA 分子 (sgRNA)。
6. 根据权利要求 1-5 中任一项所述的 gRNA 分子, 其中 (1) 使包含所述 gRNA 分子的 CRISPR/Cas 系统接触所述靶序列, 或 (2) 将包含所述 gRNA 分子的 CRISPR/Cas 系统引入包含所述靶序列的基因所在的细胞中, 能够在所述靶序列处或附近形成的碱基插入或缺失 (indel)。
7. 根据权利要求 6 所述的 gRNA 分子, 所述 CRISPR/Cas 系统包含 Cas12b 核酸酶和/或其功能活性片段。
8. 根据权利要求 6 所述的 gRNA 分子, 所述 CRISPR/Cas 系统包含 Cas12i 核酸酶和/或其功能活性片段。
9. 根据权利要求 3 所述的 gRNA 分子, 所述 tracr 序列包含 SEQ ID NO:60 所述的核苷酸序列。
10. 根据权利要求 3 所述的 gRNA 分子, 所述 crRNA 序列包含选自 SEQ ID NOs: 61-66 中任一项所述的核苷酸序列。
11. 根据权利要求 1-10 中任一项所述的 gRNA 分子, 其包含选自 SEQ ID NOs: 12-59 和 67-78 中任一项所述的核苷酸序列。
12. 一种核酸分子, 其编码权利要求 1-11 中任一项所述的 gRNA 分子。
13. 一种载体, 其包含权利要求 12 所述的核酸分子。
14. 根据权利要求 13 所述的载体, 其选自慢病毒载体、腺病毒载体、腺相关病毒(AAV)载体、单纯疱疹病毒(HSV)载体、质粒、微环、纳米质粒和 RNA 载体。
15. 一种核糖核蛋白 (RNP) 复合物, 其包含权利要求 1-11 中任一项所述的 gRNA 分子和 CRISPR/Cas 系统的核酸酶。
16. 根据权利要求 15 所述的 RNP 复合物, 所述核酸酶为 II 类 V 型 Cas 酶。
17. 根据权利要求 16 所述的 RNP 复合物, 所述核酸酶为 Cas12b 核酸酶。

18. 根据权利要求 17 所述 RNP 复合物, 所述 Cas12b 核酸酶包括其变体、其直系同源物和/或其功能活性片段。
19. 根据权利要求 17 或 18 所述的 RNP 复合物, 所述 Cas12b 核酸酶为来自 *Alicyclobacillus acidiphilus* 的 Cas12b 核酸酶(AaCas12b)。
20. 根据权利要求 19 所述的 RNP 复合物, 所述 AaCas12b 核酸酶相对于野生型 AaCas12b 核酸酶包含下列突变中的一种或多种:
 - (1) 用带正电荷的氨基酸残基取代所述野生型 AaCas12b 核酸酶中与前间隔序列邻近基序(PAM)相互作用的一个或多个氨基酸残基;
 - (2) 用具有芳环的氨基酸残基取代所述野生型 AaCas12b 核酸酶中参与打开 DNA 双链的一个或多个氨基酸残基; 和/或
 - (3) 用带正电荷的氨基酸残基或疏水性氨基酸残基取代所述野生型 AaCas12b 核酸酶的 RuvC 结构域中与单链 DNA 底物相互作用的一个或多个氨基酸残基,其中, 所述野生型 AaCas12b 核酸酶的氨基酸序列如 SEQ ID NO: 91 所述。
21. 根据权利要求 20 所述的 RNP 复合物, 所述与 PAM 相互作用的所述一个或多个氨基酸残基位于以下的一个或多个位置: 116、123、130、132、144、145、153、173、222、395、400 和 475; 且其中氨基酸残基根据 SEQ ID NO:91 编号。
22. 根据权利要求 20 或 21 所述的 RNP 复合物, 所述带正电荷的氨基酸残基取代所述野生型 AaCas12b 核酸酶中与 PAM 相互作用的一个或多个氨基酸残基是以下取代中的一种或多种: D116R 和 E475R; 且其中所述氨基酸残基根据 SEQ ID NO:91 编号。
23. 根据权利要求 20 所述的 RNP 复合物, 所述参与打开所述 DNA 双链的一个或多个氨基酸残基位于以下的一个或多个位置: 118, 和 119; 且其中所述氨基酸残基根据 SEQ ID NO:91 编号。
24. 根据权利要求 20 或 23 所述的 RNP 复合物, 所述用具有芳环的氨基酸残基取代所述野生型 AaCas12b 核酸酶中参与打开所述 DNA 双链的一个或多个氨基酸残基是 Q119Y、Q119F 或 Q119W 取代; 且其中所述氨基酸残基根据 SEQ ID NO:91 编号。
25. 根据权利要求 20 所述的 RNP 复合物, 所述位于 RuvC 结构域中并与单链 DNA 底物相互作用的一个或多个氨基酸残基位于以下的一个或多个位置: 300、301、304、329、636、639、647、682、757、758、761、764、768、852、854、856、857、858、860、862、863、865、866、867、869、938、956、957、958、994、1093 和 1097; 且其中所述氨基酸残基根据 SEQ ID NO:91 编号。
26. 根据权利要求 20 或 25 所述的 RNP 复合物, 所述取代所述野生型 AaCas12b 核酸酶中位

于 RuvC 结构域中并与单链 DNA 底物相互作用的一个或多个氨基酸残基是以下取代中的一种或多种：E636R、Q639R、T647R、Q682R、I757R、E758R、E761R、Q854R、N857K、D858R、I994R、Q1093R 和 W1097R；且其中所述氨基酸残基根据 SEQ ID NO:91 编号。

27. 根据权利要求 17-26 中任一项所述的 RNP 复合物，所述 gRNA 分子包含 SEQ ID NOs: 12-59 中任一项所述的核苷酸序列。

28. 根据权利要求 16 所述的 RNP 复合物，所述核酸酶为 Cas12i 核酸酶。

29. 根据权利要求 28 所述的 RNP 复合物，所述 Cas12i 核酸酶包括其变体、其直系同源物和/或其功能活性片段。

30. 根据权利要求 28 或 29 所述的 RNP 复合物，所述 Cas12i 核酸酶相对于野生型 Cas12i2 核酸酶包含下列突变中的一种或多种：

(1) 将所述野生型 Cas12i2 核酸酶中与 PAM 相互作用的一个或多个氨基酸替换为带正电的氨基酸；

(2) 将所述野生型 Cas12i2 核酸酶中参与打开 DNA 双链的一个或多个氨基酸替换为带芳香环的氨基酸；

(3) 将所述野生型 Cas12i2 核酸酶中位于 RuvC 结构域并与单链 DNA 底物相互作用的一个或多个氨基酸替换为带正电的氨基酸；和/或

(4) 将所述野生型 Cas12i2 核酸酶中与 DNA-RNA 双螺旋相互作用的一个或多个氨基酸替换为带正电的氨基酸，

其中，所述野生型 Cas12i2 核酸酶的氨基酸序列如 SEQ ID NO: 92 所述。

31. 根据权利要求 30 所述的 RNP 复合物，其中所述与 PAM 相互作用的一个或多个氨基酸位于下述的一个或多个位置：176、238、447、和/或 563；且其中所述位置根据 SEQ ID NO:92 编号。

32. 根据权利要求 30 所述的 RNP 复合物，其中 (1) 所述带正电的氨基酸是 R 或 K。

33. 根据权利要求 28-32 中任一项所述的 RNP 复合物，其中所述 Cas12i 核酸酶包含下述任何一种突变或突变组合：(1) E563R；(2) E176R、T447R、E176R 和 E563R；(3) K238R 和 E563R；(4) E176R、K238R 和 T447R；(5) E176R、K238R 和 E563R；(6) E176R、T447R 和 E563R；和/或 (7) E176R、K238R、T447R 和 E563R；且其中氨基酸残基根据 SEQ ID NO:92 编号。

34. 根据权利要求 30 所述的 RNP 复合物，其中所述参与打开 DNA 双链的一个或多个氨基酸位于下述的一个或多个位置：163 和/或 164；且其中所述位置根据 SEQ ID NO:92 编号。

35. 根据权利要求 30 所述的 RNP 复合物，其中 (2) 所述带芳香环的氨基酸是 F、Y 或 W。

36. 根据权利要求 34 或 35 所述的 RNP 复合物，其中所述参与打开 DNA 双链的一个或多个氨基酸替换为带芳香环的氨基酸是指：Q163F、Q163Y、Q163W、和/或 N164F 或 N164Y；且其中氨基酸残基根据 SEQ ID NO:92 编号。
37. 根据权利要求 30 所述的 RNP 复合物，其中所述位于 RuvC 结构域并与单链 DNA 底物相互作用的一个或多个氨基酸位于下述的一个或多个位置：323、362、425、925、926、391、424 和/或 929；且其中所述位置根据 SEQ ID NO:92 编号。
38. 根据权利要求 30 所述的 RNP 复合物，其中 (3) 所述带正电的氨基酸是 R 或 K。
39. 根据权利要求 37 或 38 所述的 RNP 复合物，其中所述 Cas12i 核酸酶包含下述任何一种突变或突变组合：(1) E323R；(2) D362R；(3) Q425R；(4) N925R；(5) I926R；(6) E323R 和 D362R；(7) E323R 和 Q425R；(8) E323R 和 I926R；(9) Q425R 和 I926R；(10) D362R 和 I926R；(11) N925R 和 I926R；(12) E323R、D362R 和 Q425R；(13) E323R、D362R 和 I926R；(14) E323R、Q425R 和 I926R；(15) D362R、N925R 和 I926R；和/或 (16) E323R、D362R、Q425R 和 I926R；且其中氨基酸残基根据 SEQ ID NO:92 编号。
40. 根据权利要求 30 所述的 RNP 复合物，其中所述与 DNA-RNA 双螺旋相互作用的一个或多个氨基酸位于下述的一个或多个位置：116、117、159、161、319、343 和/或 958；且其中所述位置根据 SEQ ID NO:92 编号。
41. 根据权利要求 30 所述的 RNP 复合物，其中 (4) 所述带正电的氨基酸是 R 或 K。
42. 根据权利要求 40 或 41 所述的 RNP 复合物，其中所述 Cas12i 核酸酶包含下述任何一种突变或突变组合：G116R、E117R、T159R、S161R、E319R、E343R、和/或 D958R；且其中氨基酸残基根据 SEQ ID NO:92 编号。
43. 根据权利要求 30-42 中任一项所述的 RNP 复合物，所述 Cas12i 核酸酶进一步包含一个或多个柔性区的突变，所述突变能够使所述 Cas12i 核酸酶的柔性区相比于所述野生型 Cas12i2 核酸酶具有增加的柔性。
44. 根据权利要求 43 所述的 RNP 复合物，所述柔性区选自对应于以下的区的组：氨基酸残基 228-232、氨基酸残基 439-443、氨基酸残基 478-482、氨基酸残基 500-504、氨基酸残基 775-779 和氨基酸残基 925-929；且其中所述氨基酸残基根据 SEQ ID NO: 92 编号。
45. 根据权利要求 43 或 44 所述的 RNP 复合物，所述一个或多个柔性区的突变包括在所述柔性区中插入一个或多个 G 残基。
46. 根据权利要求 45 所述的 RNP 复合物，所述一个或多个 G 残基插入在所述柔性区中柔性氨基酸残基的 N 末端，其中所述柔性氨基酸残基选自下组：G、S、N、D、H、M、T、E、Q、K、R、A 和 P。

47. 根据权利要求 43 或 44 所述的 RNP 复合物, 所述一个或多个柔性区的突变包括用 G 残基替换所述柔性区中的疏水氨基酸残基, 其中所述疏水氨基酸残基选自: L、I、V、C、Y、F 和 W。
48. 根据权利要求 43-47 中任一项所述的 RNP 复合物, 其中所述一个或多个柔性区的突变为:
(1) I926G; 和/或 (2) 439G 或 439GG, 且其中氨基酸残基根据 SEQ ID NO:92 编号。
49. 根据权利要求 43-48 中任一项所述的 RNP 复合物, 所述 Cas12i 核酸酶包含突变: E176R、K238R、T447R、E563R、N164Y、E323R 和 D362R; 且其中氨基酸残基根据 SEQ ID NO:92 编号。
50. 根据权利要求 28-49 中任一项所述的 RNP 复合物, 所述 gRNA 分子包含 SEQ ID NOs: 67-78 中任一项所述的核苷酸序列。
51. 根据权利要求 15-49 中任一项所述的 RNP 复合物, 所述核酸酶的 N 端和/或 C 端与一个或一个以上的核定位信号连接。
52. 根据权利要求 51 所述的 RNP 复合物, 所述核定位信号与 His 标签连接。
53. 根据权利要求 15-52 中任一项所述的 RNP 复合物, 所述核酸酶包含选自 SEQ ID NOs: 1-11 中任一项所述的氨基酸序列。
54. 一种组合物, 其包含:
(1) 一种或多种如权利要求 1-11 中任一项所述的 gRNA 分子, 和 CRISPR/Cas 系统的核酸酶;
(2) 编码一种或多种如权利要求 1-11 中任一项所述的 gRNA 分子的核酸, 和 CRISPR/Cas 系统的核酸酶;
(3) 一种或多种如权利要求 1-11 中任一项所述的 gRNA 分子, 和编码 CRISPR/Cas 系统的核酸酶的核酸; 或
(4) 编码一种或多种如权利要求 1-11 中任一项所述的 gRNA 分子的核酸, 和编码 CRISPR/Cas 系统的核酸酶的核酸。
55. 根据权利要求 54 所述的组合物, 其 (1) 中所述一种或多种 gRNA 分子和所述 CRISPR/Cas 系统的核酸酶存在于核糖核蛋白 (RNP) 复合物中。
56. 根据权利要求 55 所述的组合物, 所述核酸酶为 II 类 V 型 Cas 酶。
57. 根据权利要求 56 所述的组合物, 所述核酸酶为 Cas12b 核酸酶。
58. 根据权利要求 57 所述的组合物, 所述 Cas12b 核酸酶包括其变体、其直系同源物和/或其功能活性片段。
59. 根据权利要求 57 或 58 所述的组合物, 所述 Cas12b 核酸酶为来自 *Alicyclobacillus acidiphilus*

的 Cas12b 核酸酶(AaCas12b)。

60. 根据权利要求 59 所述的组合物, 所述 AaCas12b 相对于野生型 AaCas12b 核酸酶包含下列突变中的一种或多种:
- (1) 用带正电荷的氨基酸残基取代所述野生型 AaCas12b 核酸酶中与前间隔序列邻近基序(PAM)相互作用的一个或多个氨基酸残基;
 - (2) 用具有芳环的氨基酸残基取代所述野生型 AaCas12b 核酸酶中参与打开 DNA 双链的一个或多个氨基酸残基; 和/或
 - (3) 用带正电荷的氨基酸残基或疏水性氨基酸残基取代所述野生型 AaCas12b 核酸酶的 RuvC 结构域中与单链 DNA 底物相互作用的一个或多个氨基酸残基,
- 其中, 所述野生型 AaCas12b 核酸酶的氨基酸序列如 SEQ ID NO: 91 所述。
61. 根据权利要求 60 所述的组合物, 所述与 PAM 相互作用的所述一个或多个氨基酸残基位于以下的一个或多个位置: 116、123、130、132、144、145、153、173、222、395、400 和 475; 且其中所述氨基酸残基根据 SEQ ID NO:91 编号。
62. 根据权利要求 60 或 61 所述的组合物, 所述带正电荷的氨基酸残基取代所述野生型 AaCas12b 核酸酶中与 PAM 相互作用的一个或多个氨基酸残基是以下取代中的一种或多种: D116R 和 E475R; 且其中所述氨基酸残基根据 SEQ ID NO: 91 编号。
63. 根据权利要求 60 所述的组合物, 所述参与打开所述 DNA 双链的一个或多个氨基酸残基位于以下的一个或多个位置: 118, 和 119; 且其中所述氨基酸残基根据 SEQ ID NO: 91 编号。
64. 根据权利要求 60 或 63 所述的组合物, 所述用具有芳环的氨基酸残基取代所述野生型 AaCas12b 核酸酶中参与打开所述 DNA 双链的一个或多个氨基酸残基是 Q119Y、Q119F 或 Q119W 取代; 且其中所述氨基酸残基根据 SEQ ID NO: 91 编号。
65. 根据权利要求 60 所述的组合物, 所述位于 RuvC 结构域中并与单链 DNA 底物相互作用的一个或多个氨基酸残基位于以下的一个或多个位置: 300、301、304、329、636、639、647、682、757、758、761、764、768、852、854、856、857、858、860、862、863、865、866、867、869、938、956、957、958、994、1093 和 1097; 且其中所述氨基酸残基根据 SEQ ID NO: 91 编号。
66. 根据权利要求 60 或 65 所述的组合物, 所述取代所述野生型 AaCas12b 核酸酶中位于 RuvC 结构域中并与单链 DNA 底物相互作用的一个或多个氨基酸残基是以下取代中的一种或多种: E636R、Q639R、T647R、Q682R、I757R、E758R、E761R、Q854R、N857K、D858R、I994R、Q1093R 和 W1097R; 且其中所述氨基酸残基根据 SEQ ID NO:91 编号。

67. 根据权利要求 57-66 中任一项所述的组合物，所述 gRNA 分子包含 SEQ ID NOs: 12-59 中任一项所述的核苷酸序列。
68. 根据权利要求 56 所述的组合物，所述核酸酶为 Cas12i 核酸酶。
69. 根据权利要求 68 所述的组合物，所述 Cas12i 核酸酶包括其变体、其直系同源物和/或其功能活性片段。
70. 根据权利要求 68 或 69 所述的组合物，所述 Cas12i 核酸酶相对于野生型 Cas12i2 核酸酶包含下列突变中的一种或多种：
- (1) 将所述野生型 Cas12i2 核酸酶中与 PAM 相互作用的一个或多个氨基酸替换为带正电的氨基酸；
 - (2) 将所述野生型 Cas12i2 核酸酶中参与打开 DNA 双链的一个或多个氨基酸替换为带芳香环的氨基酸；
 - (3) 将所述野生型 Cas12i2 核酸酶中位于 RuvC 结构域并与单链 DNA 底物相互作用的一个或多个氨基酸替换为带正电的氨基酸；和/或
 - (4) 将所述野生型 Cas12i2 核酸酶中与 DNA-RNA 双螺旋相互作用的一个或多个氨基酸替换为带正电的氨基酸，
- 其中，所述野生型 Cas12i2 核酸酶的氨基酸序列如 SEQ ID NO: 92 所述。
71. 根据权利要求 70 所述的组合物，其中所述与 PAM 相互作用的一个或多个氨基酸位于下述的一个或多个位置：176、238、447、和/或 563；且其中所述位置根据 SEQ ID NO:92 编号。
72. 根据权利要求 70 所述的组合物，其中 (1) 所述带正电的氨基酸是 R 或 K。
73. 根据权利要求 71 或 72 所述的组合物，其中所述 Cas12i 核酸酶包含下述任何一种突变或突变组合：(1) E563R；(2) E176R、T447R、E176R 和 E563R；(3) K238R 和 E563R；(4) E176R、K238R 和 T447R；(5) E176R、K238R 和 E563R；(6) E176R、T447R 和 E563R；和/或 (7) E176R、K238R、T447R 和 E563R；且其中氨基酸残基根据 SEQ ID NO:92 编号。
74. 根据权利要求 70 所述的组合物，其中所述参与打开 DNA 双链的一个或多个氨基酸位于下述的一个或多个位置：163 和/或 164；且其中所述位置根据 SEQ ID NO:92 编号。
75. 根据权利要求 70 所述的组合物，其中 (2) 所述带芳香环的氨基酸是 F、Y 或 W。
76. 根据权利要求 74 或 75 所述的组合物，其中所述参与打开 DNA 双链的一个或多个氨基酸替换为带芳香环的氨基酸是指：Q163F、Q163Y、Q163W、和/或 N164F 或 N164Y；且其中氨基酸残基根据 SEQ ID NO: 92 编号。

77. 根据权利要求 70 所述的组合物，其中所述位于 RuvC 结构域并与单链 DNA 底物相互作用的一个或多个氨基酸位于下述的一个或多个位置：323、362、425、925、926、391、424 和/或 929；且其中所述位置根据 SEQ ID NO:92 编号。
78. 根据权利要求 70 所述的组合物，其中 (3) 所述带正电的氨基酸是 R 或 K。
79. 根据权利要求 77 或 78 所述的组合物，其中所述 Cas12i 核酸酶包含下述任何一种突变或突变组合：(1) E323R；(2) D362R；(3) Q425R；(4) N925R；(5) I926R；(6) E323R 和 D362R；(7) E323R 和 Q425R；(8) E323R 和 I926R；(9) Q425R 和 I926R；(10) D362R 和 I926R；(11) N925R 和 I926R；(12) E323R、D362R 和 Q425R；(13) E323R、D362R 和 I926R；(14) E323R、Q425R 和 I926R；(15) D362R、N925R 和 I926R；和/或 (16) E323R、D362R、Q425R 和 I926R；且其中氨基酸残基根据 SEQ ID NO:92 编号。
80. 根据权利要求 70 所述的组合物，其中所述与 DNA-RNA 双螺旋相互作用的一个或多个氨基酸位于下述的一个或多个位置：116、117、159、161、319、343 和/或 958；且其中所述位置根据 SEQ ID NO:92 编号。
81. 根据权利要求 70 所述的组合物，其中 (4) 所述带正电的氨基酸是 R 或 K。
82. 根据权利要求 80 或 81 所述的组合物，其中所述 Cas12i 核酸酶包含下述任何一种突变或突变组合：G116R、E117R、T159R、S161R、E319R、E343R、和/或 D958R；且其中氨基酸残基根据 SEQ ID NO:92 编号。
83. 根据权利要求 70-82 中任一项所述的组合物，所述 Cas12i 核酸酶进一步包含一个或多个柔性区的突变，所述突变能够使所述 Cas12i 核酸酶的柔性区相比于所述野生型 Cas12i2 核酸酶具有增加的柔性。
84. 根据权利要求 83 所述的组合物，所述柔性区选自对应于以下的区的组：氨基酸残基 228-232、氨基酸残基 439-443、氨基酸残基 478-482、氨基酸残基 500-504、氨基酸残基 775-779 和氨基酸残基 925-929；且其中所述氨基酸残基根据 SEQ ID NO: 92 编号。
85. 根据权利要求 83 或 84 所述的组合物，所述一个或多个柔性区的突变包括在所述柔性区中插入一个或多个 G 残基。
86. 根据权利要求 85 所述的组合物，所述一个或多个 G 残基插入在所述柔性区中柔性氨基酸残基的 N 末端，其中所述柔性氨基酸残基选自下组：G、S、N、D、H、M、T、E、Q、K、R、A 和 P。
87. 根据权利要求 85 或 86 所述的组合物，所述一个或多个柔性区的突变包括用 G 残基替换所述柔性区中的疏水氨基酸残基，其中所述疏水氨基酸残基选自下组：L、I、V、C、Y、F 和 W。

88. 根据权利要求 83-87 中任一项所述的组合物，其中所述一个或多个柔性区的突变为：(1) I926G；和/或 (2) 439G 或 439GG，且其中氨基酸残基根据 SEQ ID NO:92 编号。
89. 根据权利要求 68-88 中任一项所述的组合物，所述 Cas12i 核酸酶包含突变：E176R、K238R、T447R、E563R、N164Y、E323R 和 D362R；且其中氨基酸残基根据 SEQ ID NO:92 编号。
90. 根据权利要求 68-89 中任一项所述的组合物，所述 gRNA 分子包含 SEQ ID NOs: 67-78 中任一项所述的核苷酸序列。
91. 根据权利要求 54-90 中任一项所述的组合物，所述核酸酶的 N 端和/或 C 端与一个或一个以上的核定位信号连接。
92. 根据权利要求 91 所述的组合物，所述核定位信号与 His 标签连接。
93. 根据权利要求 54-92 中任一项所述的组合物，所述核酸酶包含选自 SEQ ID NOs: 1-11 中任一项所述的氨基酸序列。
94. 根据权利要求 54-93 中任一项所述的组合物，所述组合物还任选地包含药学上可接受的载体。
95. 一种细胞，其包含权利要求 1-11 中任一项所述的 gRNA 分子、权利要求 12 所述的核酸分子、权利要求 13-14 中任一项所述的载体、权利要求 15-53 中任一项所述的核糖核蛋白 (RNP) 复合物、和/或权利要求 54-94 中任一项所述的组合物。
96. 细胞群或其后代，其包含施用权利要求 15-53 中任一项所述的核糖核蛋白 (RNP) 复合物、权利要求 54-94 中任一项所述的组合物、权利要求 95 所述的细胞所产生的基因产物。
97. 根据权利要求 96 所述的细胞群或其后代，所述基因产物包含被修饰的 HBG 基因。
98. 根据权利要求 97 所述的细胞群或其后代，所述修饰为位于靶序列处或附近的碱基插入或缺失 (indel)；其中所述靶序列位于 HBG1 启动子区域内或 HBG2 启动子区域内，所述 HBG1 启动子区域位于人基因组 hg19 的 chr11: 5,248,269 至 5,249,857 之间，所述 HBG2 启动子区域位于人基因组 hg19 的 chr11: 5,253,188-5,254,781 之间。
99. 根据权利要求 98 所述的细胞群或其后代，所述靶序列位于所述 HBG1 和所述 HBG2 的转录起始位点上游约 210bp 至上游约 100bp 之间。
100. 根据权利要求 96-99 中任一项所述的细胞群或其后代，所述细胞群能够分化成为红系细胞谱系的分化细胞，且所述分化细胞相比于 HBG 基因未经修饰的细胞群具有增加的胎儿血红蛋白表达水平。
101. 一种试剂盒，其包含权利要求 1-11 中任一项所述的 gRNA 分子、权利要求 12 所述的核酸分子、权利要求 13-14 中任一项所述的载体、权利要求 15-53 中任一项所述的核糖核蛋白 (RNP) 复合物、权利要求 54-94 中任一项所述的组合物、权利要求 95 所述的细胞、

和/或权利要求 96-100 中任一项所述的细胞群或其后代。

102. 一种调控细胞、细胞群或其后代中胎儿血红蛋白（HbF）表达的方法，所述方法包括施用权利要求 15-53 中任一项所述的核糖核蛋白（RNP）复合物、权利要求 54-94 中任一项所述的组合物、权利要求 95 所述的细胞、权利要求 96-100 中任一项所述的细胞群或其后代、和/或权利要求 101 所述的试剂盒。
103. 根据权利要求 102 所述的方法，所述细胞或细胞群为造血干/祖细胞（HSPC）。
104. 根据权利要求 103 所述的方法，所述细胞或细胞群为 CD34+HSPC。
105. 一种治疗、预防或缓解血红蛋白病和/或其相关病症的方法，所述方法包括向有需要的受试者施用有效量的权利要求 15-53 中任一项所述的核糖核蛋白（RNP）复合物、权利要求 54-94 中任一项所述的组合物、权利要求 95 所述的细胞、权利要求 96-100 中任一项所述的细胞群或其后代、和/或权利要求 101 所述的试剂盒。
106. 根据权利要求 105 所述的方法，所述血红蛋白并和/或其相关病症选自：镰状细胞病、镰状细胞性贫血、血红蛋白 C 病、血红蛋白 C 性状、血红蛋白 S/C 病、血红蛋白 D 病、血红蛋白 E 病、地中海贫血、氧亲和力下降血红蛋白病、和不稳定血红蛋白病。