激活γ -珠蛋白基因表达的方法和组合物

技术领域

本发明涉及基因编辑技术领域，涉及一种用于激活γ -珠蛋白基因表达的方法、组合物及其应用。该方法采用含有 GATA 或其反义互补序列 TATC 的单链寡核苷酸(ssODN)作为指导信息，在 γ-珠蛋白基因启动子区进行基因编辑形成含有 GATA 的红系增强子元件，能够促进 γ-珠蛋白基因在成熟红细胞中表达。

背景技术

血红蛋白（hemoglobin，简称 Hb）是红细胞内携带和运输氧气的特殊蛋白质。血红蛋白由珠蛋白（globin）和血红素构成。成人体内的血红蛋白主要是由 2 个 α-珠蛋白和 2 个 β-珠蛋白组成的四聚体（α2β2），成为成人血红蛋白（HbA）。β-珠蛋白的基因（HBB）的突变可引起 β-血红蛋白障碍(也称 β-血红蛋白病)，包括镰状细胞病（Sickle Cell Disease，SCD）和 β-地中海贫血（β-thalassemia，简称 β-thal）。镰状细胞性贫血是由 β-珠蛋白结构基因的点突变引起的，导致产生异常的血红蛋白(HbS)。β-地中海贫血是由 β-珠蛋白基因表达的部分或完全缺陷引起的，导致成人血红蛋白(HbA)缺陷或缺失。血红蛋白的珠蛋白链减少或缺失会导致血红蛋白结构异常，这种含有异常血红蛋白的红细胞变形性降低，寿命减短，在骨髓中可能出现原位溶血，进入到外周血液循环后被脾脏等脏器提前破坏，致使贫血、体内铁沉积甚至发育异常。血红蛋白病影响全球数百万人口，目前每年大约有 330,000 儿童出生患有血红蛋白病，严重威胁人类的健康甚至生命。地中海贫血和镰状细胞病患者主要通过规范性长期输血和去铁治疗来缓解病症，但这种方式不但不能治愈，同时还会带来较高的安全隐患。异体造血干细胞移植是目前唯一能根治地中海贫血和镰状细胞病的治疗技术，然而由于骨髓配型成功率低和免疫排斥风险等因素的限制，在临床上难以大范围应用。因此迫切需要发展新的安全和有效的治疗方法。

人类发育过程中，血红蛋白并不是一直以 2 个 α-珠蛋白和 2 个 β-珠蛋白的形式组成的。在胚胎发育期间和出生后不久，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的形式存在(α2γ2)，称之为胎儿血红蛋白(HbF)，它相比于 HbA 有更强的氧亲和力。随着胎儿的发育，γ-珠蛋白基因逐渐沉默而不表达，而相同基因组位点附近的 β-珠蛋白基因逐渐升高表达，在婴儿出生后约半年后，血液中血红蛋白的组成和比例逐渐稳定下来，HbF 被成人血红蛋白(HbA)替换，只剩余极低水平 HbF（占总血红蛋白的 1%以下）。

研究发现，有些人群中一些特殊的位点出现突变，γ-珠蛋白基因转录被激活，导致血红蛋白中 HbF 比例升高，被称之为遗传性胎儿血红蛋白持续增多（hereditary persistence of fetal hemoglobin，HPFH）。比如，γ-珠蛋白基因启动子区发现了一些不同类型的突变，如-114 至-102 13bp del c .、4bp del c .-225 至-222、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 .-198T>C、c .-201C>T 等，这些天然存在的突变不同程度地提高 HbF 在总血红蛋白的比例。临床研究发现，少数 β-地中海贫血或者镰刀型贫血症患者由于恰好基因组中存在 HPFH 突变，HbF 的表达弥补了 HbA 的不足，在一定程度上减轻或者缓解了病人的贫血症状，或减少了输血需求。受到该发现的启发，科学家们不断探索各种能够诱导 HbF 表达的方法来达到治疗 β-血红蛋白病的目的。如目前临床上有用羟基脲等药物诱导 HbF 表达治疗 β-血红蛋白病，然而大多数病人中诱导 HbF 的水平较低，不能够彻底改善病人的状况。

基因编辑技术为血红蛋白病等遗传病的治疗带来了新希望和新方法。近几年基因编辑技术获得突破性的发展，使得人为改变基因组中特定位点的碱基序列成为可能，并且越来越容易。目前发展相对成熟的基因编辑技术有 ZFN(锌指核酸酶)、TALEN(转录激活因子样效应物核酸酶)和 CRISPR(规律成簇间隔短回文重复)/Cas 系统（CRISPR-Cas 系统）。通过设计特异的基因编辑系统，将活细胞中目标基因组 DNA 位点切

开形成双链断裂(DSB)，细胞可以利用非同源末端连接（non-homologous end joining，NHEJ）修复机制对DNA 缺口进行修复，可随机形成序列的插入、缺失、碱基替换等突变。如在进行基因编辑的同时提供人为设计的 DNA 供体模板（donor template），细胞可利用同源介导的修复（homology-directed repair，HDR）机制完成修复，从而在目标位点引入预期的碱基突变形式。

目前，通过基因编辑增强红细胞中 HbF 表达的治疗策略，国际上有几种不同的方法。（1）删除或者破坏人类第 2 号染色体上 BCL11A 基因内含子中的红细胞系增强子元件，减少红细胞中 BCL11A 的表达，解除 BCL11A 对 γ-珠蛋白基因的转录抑制作用。（2）删除或者破坏 γ-珠蛋白基因启动子区中一些序列，阻止转录抑制因子的结合，或者引入天然存在的 HPFH 突变类型，如-114 至-102 位点 13 个碱基对删除（13bp del c.）。（3）制造 3.5kb 至 13.6kb 大片段的 DNA 删除，删除 γ-珠蛋白基因和 β-珠蛋白之间的含有不明抑制子的序列，以促使 γ-珠蛋白基因与远端 LCR 增强子结合。这些方法显示出激活 HbF 表达的效果，但是这些方法的长期功效或/和安全性尚不清楚，需要进一步发展新的方法。

发明内容

本发明提供一种通过基因编辑激活 γ-珠蛋白基因转录的方法。本发明利用含有 GATA 或其反义互补序列 TATC 的单链寡核苷酸(ssODN)，通过基因编辑技术（例如 CRISPR-Cas 基因编辑系统），在 γ-珠蛋白基因启动子区恰当的位置制造少数碱基的突变（删除、替换、插入等），被编辑后在 γ-珠蛋白基因启动子区的正义链或者反义链上形成含有 GATA 的红系增强子元件，如 NTG - N(7-8) – WGATAR 或 NAG - N(7-8) – WGATAR 的序列结构，能够促进 γ-珠蛋白基因在成熟红细胞中转录，使红细胞中 HbF 表达升高。其中，W 为 T 或 A；R 为 A 或 G；N 为 A、G、C 或 T，更优选为 C 或 T；N(7-8)是指 7 至 8 个任意碱基。

如图 1 所示，本发明的原理是基于 CRISPR-Cas 等基因编辑技术，结合本发明设计的单链寡核苷酸(ssODN)，对 CD34+造血干细胞等具有红细胞分化能力的干细胞、祖细胞进行高效基因编辑，在编码人类 γ-珠蛋白的 HBG1 和 HBG2 基因（位于人类 11 号染色体）启动子中制造突变，在启动子的正义链或者反义链上人为创造一个 NTG - N(7-8) – WGATAR 或 NAG - N(7-8) – WGATAR 的序列结构，该序列作为增强子，可以在目的细胞分化为红细胞后招募 GATA1 等激活因子促进 γ-珠蛋白表达。

本发明提供一种含有 GATA 或 TATC 序列的 ssODN，该 ssODN 可用于 γ-珠蛋白基因的基因编辑。本发明中 ssODN 结构上包含 5′ 同源臂、替换序列、3′ 同源臂。其中 GATA 或 TATC 序列可以位于 ssODN 上任意位置，包括替换序列、5′ 同源臂、3′ 同源臂，以及 5′ 同源臂与替换序列连接处、3′ 同源臂与替换序列连接处。单侧同源臂的长度可以是 20 至 300 nt，在本发明的实施例中，ssODN 同源臂的长度是 40 nt 左右。在本发明的一些实施例中，含有 GATA 或 TATC 序列的 ssODN 选自 SEQ ID NO:40～SEQ ID NO:65 所示序列。在其中一些实施例中，ssODN 的 5′ 和 3′ 是硫代磷酸（Phosphorothioate）修饰的。

本发明公开了一种用于γ-珠蛋白基因编辑的组合物。该组合物中包含 ssODN，sgRNA，以及 CRISPR-Cas 核酸酶。在本发明的一些实施例中，所述基因编辑系统 sgRNA 靶向的位点位于 SEQ ID NO：32 至 SEQ ID NO:39 所示序列中。在本发明的一些实施例中，所述 sgRNA 是 SEQ ID NO：69～SEQ ID NO:76 所示序列。在本发明的一些实施例中，所述 Cas9 蛋白为来源于 Streptococcus pyogenes 的 Cas9 蛋白。

本发明公开了一种用于 γ-珠蛋白基因编辑的电转方法。该方法利用电穿孔技术，将 ssODN，sgRNA，以及 CRISPR-Cas 核酸酶形成的组合物高效转染入人类造血干细胞中，介导高效的基因编辑。

本发明还公开了一种造血干细胞，采用电转方法将上述包含 ssODN、sgRNA、以及 CRISPR-Cas 核酸酶的组合物转入造血干细胞中，该细胞中 γ-珠蛋白基因启动子区含有激活性 GATA 元件，能够在分化成红细胞时提升 γ-珠蛋白基因的表达。

与现有技术相比，本发明具有以下有益效果：

本发明的 ssODN 与基因编辑系统（例如 CRISPR-Cas、TALEN、ZFN 等）一起导入细胞中，可以指导

γ-珠蛋白基因启动子区的基因编辑，在特定的位点突变形成具有激活作用的 GATA 元件。本发明的一些实施例中，利用电转技术将 ssODN 和 CRISPR-Cas 基因编辑系统导入人类造血干细胞，在 γ-珠蛋白基因启动子区获得高效的基因编辑，在特定的位点引入突变形成了预期的含有 GATA 的红系增强子元件，该造血干细胞分化成红细胞后，γ-珠蛋白基因表达显著提高，血红蛋白中 HbF 比例显著提高。

附图说明

图 1 示意本发明的技术原理。

图 2A 显示基因编辑靶向区域为 HBG1 和 HBG2 基因-92 至-66 之间的野生型序列，所用 gRNA 靶向序列的位置，以及基因编辑后期望得到的突变序列。

图 2B 显示基因编辑靶向区域为 HBG1 和 HBG2 基因-129 至-98 之间的野生型序列，所用 gRNA 靶向序列的位置，以及基因编辑后期望得到的突变序列。

图 2C 显示基因编辑靶向区域为 HBG1 和 HBG2 基因-175 至-153 之间的野生型序列，所用 gRNA 靶向序列的位置，以及基因编辑后期望得到的突变序列。

图 2D 显示基因编辑靶向区域为 HBG1 和 HBG2 基因-192 至-160 之间的野生型序列，所用 gRNA 靶向序列的位置，以及基因编辑后期望得到的突变序列。

图 2E 显示基因编辑靶向区域为 HBG1 基因-428 至-406 和 HBG2 基因-432 至-410 之间的野生型序列，所用 gRNA 靶向序列的位置，以及基因编辑后期望得到的突变序列。

图 3 显示 293T 细胞中转染 CRISPR-Cas 质粒对 HBG1 和 HBG2 基因组的编辑效率数据，所述数据来自于 293T 细胞使用 Lipofectamine 2000™ 转染 CRISPR-Cas 质粒四天后基因组 DNA 的 INDEL 检测结果，图中显示了总体插入缺失突变所占百分比（总突变%）。

图 4 显示 293T 细胞中同时转染 CRISPR-Cas 质粒及 ssODN 对 HBG2 基因组的编辑效率数据，所述数据来自于 293T 细胞使用 Lipofectamine 2000™ 转染 CRISPR-Cas 质粒及 ssODN 四天后 HBG2 的 INDEL 检测结果。

图 5 显示 K562 细胞中同时转染 CRISPR-Cas 质粒及 ssODN 对 HBG2 基因组的编辑效率数据，所述数据来自于 K562 细胞使用 Lonza 4D 电穿孔转染 CRISPR-Cas 质粒及 ssODN 四天后 HBG2 的 INDEL 检测结果。

图 6 显示 K562 细胞中电转导入核糖核蛋白（RNP）及 ssODN 后 HBG2 基因组的编辑效率数据，所述数据来自于 ssODN 和 CRISPR-Cas/sgRNA RNP 通过电穿孔转入 K562 细胞四天后 HBG2 的 INDEL 检测结果。

图 7A-C 显示动员外周血 CD34+细胞（mPBSC）中电转导入 RNP 及 ssODN 后 HBG2 基因编辑效率数据，所述数据来自于 ssODN 和 CRISPR-Cas/sgRNA RNP 通过电穿孔转入 mPBSC 后，诱导红细胞分化 5 天后基因组 DNA 的 INDEL 检测结果。图 7A 和 7B 是 sanger 测序图谱，图 7C 是经 synthego 软件分析后的 INDEL 比例结果。

图 8 显示动员外周血 CD34+细胞（mPBSC）基因编辑后 γ-珠蛋白 mRNA 表达数据，所述数据来自于 ssODN 和 CRISPR-Cas/sgRNA RNP 通过电穿孔转入 mPBSC 后，诱导红细胞分化 18 天后 mRNA 的 qRT-PCR 检测结果，图中 γ-珠蛋白 mRNA 的表达以 GAPDH 表达水平作为内参，以不编辑的细胞中 γ-珠蛋白 mRNA 的表达作为参照进行标准化，n=3。

图 9 显示动员外周血 CD34+细胞（mPBSC）基因编辑后 HbF 表达数据，所述数据来自于 ssODN 和 CRISPR-Cas/sgRNA RNP 通过电穿孔转入 mPBSC 后，诱导红细胞分化 18 天后 HPLC 检测血红蛋白检测结果，图中 HbF 表达是其占总血红蛋白的百分比。

具体实施方式

为了便于理解本发明，下面将参照相关附图对本发明进行更全面的描述。附图中给出了本发明的较佳实施例。但是，本发明可以以许多不同的形式来实现，并不限于本文所描述的实施例。相反地，提供这些实施例的目的是使对本发明的公开内容的理解更加透彻全面。

除非另有定义，本文所使用的所有的技术和科学术语与属于本发明的技术领域的技术人员通常理解的含义相同。本文中在本发明的说明书中所使用的术语只是为了描述具体的实施例的目的，不是旨在于限制本发明。本文所使用的术语“和／或”包括一个或多个相关的所列项目的任意的和所有的组合。

以下实施例所用原料，均为市售购得。

本文提供了使用基因编辑技术（例如 CRISPR-Cas 介导基因编辑技术，结合本发明设计的单链寡核苷酸 (ssODN)，对 CD34+造血干细胞等具有红细胞分化能力的干细胞、祖细胞进行高效基因编辑，在编码人类 γ-珠蛋白的 HBG1 和 HBG2 基因（位于人类 11 号染色体）启动子中制造突变，在启动子的正义链或者反义链上人为创造一个 NTG - N(7-8) – WGATAR 或 NAG - N(7-8) – WGATAR 的序列结构（图 1），该序列作为增强子，可以在目的细胞分化为红细胞后招募 GATA1 等转录激活因子促进 γ-珠蛋白表达。

GATA-1 是一种在红细胞系统中特异高表达的转录调控因子, 在红细胞分化过程中发挥着重要作用。红细胞中 α-和 β-珠蛋白、亚铁血红素生物合成酶等红细胞中特异性高表达的基因均受到 GATA-1 的直接转录激活。GATA-1 作为转录因子，选择性地结合于染色质 DNA 上的 WGATAR 序列元件（其中 W 代表 T 或者 A，R 代表 A 或者 G，也可以用 T/A(GATA)A/G 表示；其反义互补序列为 YTATCW，其中 Y 代表 T 或者 C，也可用 T/C(TATC)T/A 表示）上，比如人β -珠蛋白基因的调控区中也发现了多个 GATA 位点。值得注意的是，并不是含有 GATA 序列的基因都会被 GATA-1 激活。这是因为，GATA 序列在基因组上分布非常广泛，往往需要与上下游的序列（顺式元件）形成一定形式的序列组合（元件组合）来发挥作用。比如，GATA 序列附近有 CACCC 或 GC 盒等顺式调控元件，通常可招募 GATA-1 与转录因子 EKLF 共同结合。如 GATA 序列附近有 E-box 元件（CAGNTG，N 代表任意碱基），或 half E-box 元件(NTG)，则招募 GATA-1 与转录因子 TAL1 共同结合。ChIP-seq 研究基因组水平 GATA-1 和 TAL1 的结合位点表明，大约四分之三的 GATA-1/TAL1 共结合位点存在一个明显的序列规律，WGATA 序列上游 7-8 个碱基处往往存在 half E-box 元件(NTG)，可以用 NTG-N（7-8）-WGATA 表示。有实验证明，这种序列是激活性 GATA 元件，可促进基因在红细胞中的表达。如 BCL11A 内含子中的+58 增强子中含有 CTG-N（7）-TGATAA 序列，破坏该序列中的 GATA 基序可以显著降低红细胞中 BCL11A 的表达。

HBG1 和 HBG2 基因转录起始位点上游 1400 个碱基的序列相似度非常高，被认为是启动子区域。本申请的发明人研究发现，γ-珠蛋白基因启动子区本身含有多个 GATA 序列（或 TATC），同时也含有非常多的 TG 序列，但是 γ-珠蛋白在成熟红细胞中几乎不表达。本申请的发明人通过深入研究发现，γ-珠蛋白基因启动子上的 GATA（或 TATC）和 TG 序列在位置和方向上没有形成有效的激活性 GATA 元件（比如 TG-N（7-8）-WGATAR），因此不能招募 GATA-1/TAL1 等在成熟红细胞中高表达的转录激活因子结合（共结合）到 γ-珠蛋白基因启动子区，因此 γ-珠蛋白在成熟红细胞中被沉默，几乎不表达。

本申请的发明人通过深入研究发现，HBG1 和 HBG2 启动子区域存在一些位点，其正义链或者反义链中的序列如果经过少数几个碱基的替换、删除、插入等改变，便可以形成 NTG - N(7-8) – WGATAR 或 NAG - N(7-8) – WGATAR 的序列结构。本申请的发明人认为，如果利用基因编辑系统对这些位点进行切割形成 DSB，通过添加供体 DNA 模板（如单链寡核苷酸 ssODN）指导这些位点进行 HDR 修复，可能将一定比例的基因组编辑成预期的序列结构，最终 γ-珠蛋白基因启动子区由于形成了如 NTG - N(7-8) – WGATAR 或 NAG - N(7-8) – WGATAR 的序列，可以作为增强子招募 GATA-1/TAL1 等在成熟红细胞中高表达的转录激活因子结合（共结合）到 γ-珠蛋白基因启动子区，从而促进 γ-珠蛋白在成熟红细胞中高表达。

基因编辑从目的上来讲，可以分为两大类，一类是破坏靶位点的序列或结构，称之为敲除（knock out）。

另一类目的是要在靶位点引入特定的序列突变，称之为敲入（knock in）。要引入特定的序列突变，最常见的方法是同时引入一个 DNA 供体模板，细胞可以利用同源重组的机制，将模板中的序列替换到基因组中，实现敲入。用于 HDR 的 DNA 模板可以是质粒 DNA，线性化的双链 DNA，AAV 递送的 DNA，或者单链寡核苷酸(ssODN)。ssODN 在基因编辑中的应用越来越多，其效率高、合成方便，特别适用于体外培养细胞中的基因编辑。

ssODN 结构上包含 5′ 同源臂、替换序列、3′ 同源臂。同源臂是与待编辑的靶位点两侧的 DNA 区域同源的序列，用于定位到染色体上目标位点。 5′ 同源臂和 3′ 同源臂在染色体上对应的序列在位置上通常是不连续的，中间有一定的间隔，如 1-20 个核苷酸的间隔，这些间隔序列是基因编辑的靶标（拟编辑序列），即希望通过基因编辑被替换掉的序列。ssODN 上 5′ 同源臂的 3′ 末端与 3′ 同源臂的 5′ 末端之间的“替换序列”是基因编辑希望得到的结果。基因编辑的同时添加 ssODN，可以诱导基因编辑时进行同源重组修复，修复的结果使细胞基因组中拟编辑序列被修复成替换序列。ssODN 中替换序列的长度可以小于拟编辑的序列，这种情况下基因编辑的结果是原基因组中拟编辑序列的删除或者/和替换。ssODN 中替换序列的长度可以是 0 碱基，这种情况下基因组上拟编辑序列被删除。ssODN 中替换序列的长度可以大于拟编辑的序列，这种情况下基因编辑的结果是拟编辑序列被替换或者/和插入序列。

本发明适用于“CRISPR-Cas”、TALEN、ZFN 等位点特异性核酸酶介导的基因编辑系统。“CRISPR-Cas”是一种基因编辑技术，包括但不限于各种自然存在或人工设计的 CRISPR-Cas 系统，如 CRISPR-Cas9 系统、CRISPR-Cas12 系统等。CRISPR-Cas9 的工作原理是 crRNA(CRISPR-derived RNA) 通过碱基配对与 tracrRNA(trans-activating RNA)结合形成 tracrRNA/crRNA 复合物，此复合物引导核酸酶 Cas9 蛋白在与 crRNA 配对的序列靶位点剪切双链 DNA。tracrRNA 和 crRNA 的作用也可以由一条人为合成的具有引导作用的 sgRNA 来替代。使用其它 CRISPR-Cas 系统时需要设计与之对应的 sgRNA 或者 crRNA。使用 TALEN 或 ZFN 等系统时需要根据本发明公布的编辑位点设计对应的 TALEN 或 ZFN 核酸酶。

实施例 1：用于将激活性 GATA 元件引入 γ-珠蛋白基因启动子的 ssODN 和 CRISPR-Cas 系统筛选

通过对 HBG1 和 HBG2 转录起始位点上游约 1.4 kb 内序列（SEQ ID NO: 1 和 SEQ ID NO:2）的分析，发明人发现了很多区段的序列满足经过少数几个碱基的替换、删除、插入等改变，便可以形成 NTG - N(7-8) – WGATAR 或 NAG - N(7-8) – WGATAR 的序列结构。通过分析这些区段中目标突变碱基附近是否具备 NGG（spCas9 识别的 PAM）序列，以及切点是否正中所需编辑的目标序列，最终获得了五个最有可能实现高效基因编辑的候选靶向区域（图 2A 至 2E），按照距离转录起始位点由近至远分别是：一号区域，HBG1 和 HBG2 启动子-92 至-66 之间（图 2A）；二号区域，HBG1 和 HBG2 启动子-129 至-98 之间（图 2B）；三号区域，HBG1 和 HBG2 启动子-175 至-153 之间（图 2C）；四号区域，HBG1 基因和 HBG2 基因-192 至-160 之间（图 2D）；五号区域，HBG1 基因-428 至-406 和 HBG2 基因-432 至-410 之间（图 2E）。

根据五个候选区域内本身的序列特征，以及 CRISPR-Cas9 基因编辑系统预测的切点位置，分析得到每个区域内序列比较适合的突变目标类型，如表 1 所示（SEQ ID NO:3 至 SEQ ID NO:31）。其中下划线表示需要替换的碱基，是基因编辑后希望得到的结果。本发明中替换碱基长度是 0-6 个碱基，实现的效果是删除、替换、插入，或者删除、替换、插入的组合。

表 1. HBG1 和 HBG2 启动子上野生型序列及编辑后获得的序列

SEQ ID	编号	目标位点及编辑后的序列（5′- 3′）
SEQ ID NO:3	一号区，野生型	TTGACCAATAGTCTTAGAGTATCCAGT
SEQ ID NO:4	突变型_01	TTGACCAATAGTGATAGAGTATCCAGT
SEQ ID NO:5	突变型_02	TTGACCAATAGAGATAGAGTATCCAGT
SEQ ID NO:6	突变型_03	TTGACCAATAAGATAGAGTATCCAGT

SEQ ID NO:7	突变型_04	TTGACCAATATGATAGAGTATCCAGT
SEQ ID NO:8	二号区，野生型	CAGCCTTGCCTTGACCAATAGCCTTGACAAGG
SEQ ID NO:9	突变型_05	CAGCCTTGCCTTGACAGATAGCCTTGACAAGG
SEQ ID NO:10	突变型_06	CAGCCTTGCCTTGACTGATAGCCTTGACAAGG
SEQ ID NO:11	突变型_07	CAGCCTTGCCTTGAGAGATAGCCTTGACAAGG
SEQ ID NO:12	突变型_08	CAGCCTTGCCTTGAGTGATAGCCTTGACAAGG
SEQ ID NO:13	突变型_09	CAGCCTTGCCTTGAAAGATAGCCTTGACAAGG
SEQ ID NO:14	突变型_10	CAGCCTTGCCTTGAATGATAGCCTTGACAAGG
SEQ ID NO:15	突变型_11	CAGCCTTGCCTTGATAGATAGCCTTGACAAGG
SEQ ID NO:16	突变型_12	CAGCCTTGCCTTGATTGATAGCCTTGACAAGG
SEQ ID NO:17	突变型_13	CAGCCTTGCCTTGATAAATAGCCTTGACAAGG
SEQ ID NO:18	突变型_14	CAGCCTTGCCTTGATAAAATAGCCTTGACAAGG
SEQ ID NO:19	突变型_15	CAGCCTTGCCTTGATAATAGCCTTGACAAGG
SEQ ID NO:20	突变型_16	CAGCCTTGCCTTGATAGCCTTGACAAGG
SEQ ID NO:21	突变型_17	CAGCCTTGCCTTGATAAGG
SEQ ID NO:22	突变型_18	CAGCCTTGCCTTGATAAAGG
SEQ ID NO:23	三号区，野生型	TATCTGTCTGAAACGGTCCCTGG
SEQ ID NO:24	突变型_19	TATCTGTCTGAAACGGTAGATAACCCTGG
SEQ ID NO:25	四号区，野生型	CACTATCTCAATGCAAATATCTGTCTGAAACGG
SEQ ID NO:26	突变型_20	CACTATCTCAATGCAACAGAAACGG
SEQ ID NO:27	五号区，野生型	TCCCTGAACTTTTCAAAAATTGG
SEQ ID NO:28	突变型_21	TCCCTGAACTTTTCAGATAATTGG
SEQ ID NO:29	突变型_22	TCCCTGAACTTTTCAGATAAATTGG
SEQ ID NO:30	突变型_23	TCCCTGAACTTTTCAGATAAAATTGG
SEQ ID NO:31	突变型_24	TCCCTGAACTTTTCAGATAGAATTGG

为了达到在表1所选区域内编辑位点附近切开DNA双链形成DSB，本发明选择切割效率较高的spCas9 CRISPR-Cas系统（来源于Streptococcus pyogenes），分析切割效率可能较高的靶位点，选择site_1至site_8作为候选靶位点，其识别的靶标位点DNA序列及PAM序列（NGG）如表2所示（SEQ ID NO:32 至SEQ ID NO:39）。

表2. HBG1和HBG2启动子上sgRNA识别的靶位点

SEQ ID	位点	sgRNA识别位点及PAM（5′-3′）	DNA链
SEQ ID NO:32	site_1	ACTGGATACTCTAAGACTAT TGG	反义链
SEQ ID NO:33	site_2	CTTGTCAAGGCTATTGGTCA AGG	反义链
SEQ ID NO:34	site_3	GTTTGCCTTGTCAAGGCTAT TGG	反义链
SEQ ID NO:35	site_4	TGGTCAAGTTTGCCTTGTCA AGG	反义链
SEQ ID NO:36	site_5	CTTGACCAATAGCCTTGACA AGG	正义链
SEQ ID NO:37	site_6	TATCTGTCTGAAACGGTCCC TGG	正义链
SEQ ID NO:38	site_7	ATGCAAATATCTGTCTGAAA CGG	正义链
SEQ ID NO:39	site_8	TCCCTGAACTTTTCAAAAAT TGG	正义链

根据表 1 所示的期望突变类型，以及表 2 中 sgRNA 识别的 DNA 链（正义链或反义链），设计对应的指导基因编辑修复 ssODN（表 3, SEQ ID NO:40 至 SEQ ID NO:65）。本发明中 ssODN 结构上包含 5′ 同源臂、替换序列、3′ 同源臂。其中 GATA 或 TATC 序列可以位于 ssODN 上任意位置，包括替换序列、5′ 同源臂、3′ 同源臂，以及 5′ 同源臂与替换序列连接处、3′ 同源臂与替换序列连接处。ssODN 两侧同源臂可以是对称的，或者是非对称的（两侧长短不一），为了系统对比的方便性，本发明中统一使用对称同源臂的 ssODN 进行实验。ssODN 两侧同源臂长度可以是 20-300nt，为了系统对比的方便性，本发明中实施例使用两侧同源臂长度约为 40nt 的 ssODN 进行示例。ssODN 可以是编辑区域 DNA 的正义链或者是反义链（同源臂序列与对应的正义链相同或者与反义链相同），本发明中 ssODN 统一选择与 sgRNA 识别位点相同的 DNA 链作为 ssODN 母版序列。ssODN 中的序列除两侧同源臂序列外，替换碱基可以是 0 个、1 个、2 个、3 个、4 个、5 个、6 个碱基，实现的效果是删除、替换、插入，或者删除、替换、插入的组合。实现的效果可以是改变基因组上 1 个以上的碱基，比如 1-20 个碱基的删除或者替换。如表 3 所示序列（SEQ ID NO:40 至 SEQ ID NO:65），下划线标示的碱基是两侧同源臂，长度在 40 nt 左右，两侧同源臂之间的未用下划线标示的碱基是替换碱基，其中 ssODN_16 的两侧同源臂之间的替换碱基是 0 个。ssODN 的两端前三个核苷酸使用硫代磷酸（Phosphorothioate）修饰以增强 ssODN 的稳定性和提高其在基因编辑中的活性。

表 3. ssODN 序列及与之配合的 gRNA 编号

SEQ ID	ssODN	ssODN 序列（5′-3′）	适用区域
SEQ ID NO:40	ssODN_1	CCTAGCCAGCCGCCGGCCCCTGGCCTCACTGGATACTCTATCACTATTGGTCAAGTTTGCCTTGTCAAGGCTATTGGTCAAGGC	一号区域
SEQ ID NO:41	ssODN_2	CCTAGCCAGCCGCCGGCCCCTGGCCTCACTGGATACTCTATCTCTATTGGTCAAGTTTGCCTTGTCAAGGCTATTGGTCAAGGC	一号区域
SEQ ID NO:42	ssODN_3	CCTAGCCAGCCGCCGGCCCCTGGCCTCACTGGATACTCTATCTTATTGGTCAAGTTTGCCTTGTCAAGGCTATTGGTCAAGGC	一号区域
SEQ ID NO:43	ssODN_4	CCTAGCCAGCCGCCGGCCCCTGGCCTCACTGGATACTCTATCATATTGGTCAAGTTTGCCTTGTCAAGGCTATTGGTCAAGGC	一号区域
SEQ ID NO:44	ssODN_5	CTCTAAGACTATTGGTCAAGTTTGCCTTGTCAAGGCTATCTGTCAAGGCAAGGCTGGCCAACCCATGGGTGGAGTTTAGCCA	二号区域
SEQ ID NO:45	ssODN_6	CTCTAAGACTATTGGTCAAGTTTGCCTTGTCAAGGCTATCAGTCAAGGCAAGGCTGGCCAACCCATGGGTGGAGTTTAGCCA	二号区域
SEQ ID NO:46	ssODN_7	CTCTAAGACTATTGGTCAAGTTTGCCTTGTCAAGGCTATCTCTCAAGGCAAGGCTGGCCAACCCATGGGTGGAGTTTAGCCA	二号区域
SEQ ID NO:47	ssODN_8	CTCTAAGACTATTGGTCAAGTTTGCCTTGTCAAGGCTATCACTCAAGGCAAGGCTGGCCAACCCATGGGTGGAGTTTAGCCA	二号区域
SEQ ID NO:48	ssODN_9	CTCTAAGACTATTGGTCAAGTTTGCCTTGTCAAGGCTATCTTTCAAGGCAAGGCTGGCCAACCCATGGGTGGAGTTTAGCCA	二号区域
SEQ ID NO:49	ssODN_10	CTCTAAGACTATTGGTCAAGTTTGCCTTGTCAAGGCTATCATTCAAGGCAAGGCTGGCCAACCCATGGGTGGAGTTTAGCCA	二号区域
SEQ ID NO:50	ssODN_11	CTCTAAGACTATTGGTCAAGTTTGCCTTGTCAAGGCTATCTATCAAGGCAAGGCTGGCCAACCCATGGGTGGAGTTTAGCCA	二号区域

SEQ ID NO:51	ssODN_12	CTCTAAGACTATTGGTCAAGTTTGCCTTGTCAAGGCTATCAATCAAGGCAAGGCTGGCCAACCCATGGGTGGAGTTTAGCCA	二号区域
SEQ ID NO:52	ssODN_13	CTCTAAGACTATTGGTCAAGTTTGCCTTGTCAAGGCTATTTATCAAGGCAAGGCTGGCCAACCCATGGGTGGAGTTTAGCCA	二号区域
SEQ ID NO:53	ssODN_14	CTCTAAGACTATTGGTCAAGTTTGCCTTGTCAAGGCTATTTTATCAAGGCAAGGCTGGCCAACCCATGGGTGGAGTTTAGCCA	二号区域
SEQ ID NO:54	ssODN_15	CTCTAAGACTATTGGTCAAGTTTGCCTTGTCAAGGCTATTATCAAGGCAAGGCTGGCCAACCCATGGGTGGAGTTTAGCCA	二号区域
SEQ ID NO:55	ssODN_16	CTCTAAGACTATTGGTCAAGTTTGCCTTGTCAAGGCTATCAAGGCAAGGCTGGCCAACCCATGGGTGGAGTTTAGCCA	二号区域
SEQ ID NO:56	ssODN_17	CCTCACTGGATACTCTAAGACTATTGGTCAAGTTTGCCTTATCAAGGCAAGGCTGGCCAACCCATGGGTGGAGTTTAGCCA	二号区域
SEQ ID NO:57	ssODN_18	CCTCACTGGATACTCTAAGACTATTGGTCAAGTTTGCCTTTATCAAGGCAAGGCTGGCCAACCCATGGGTGGAGTTTAGCCA	二号区域
SEQ ID NO:58	ssODN_19	TGGCTAAACTCCACCCATGGGTTGGCCAGCCTTGCCTTGATAAGGCAAACTTGACCAATAGTCTTAGAGTATCCAGTGAGG	二号区域
SEQ ID NO:59	ssODN_20	TGGCTAAACTCCACCCATGGGTTGGCCAGCCTTGCCTTGATAAAGGCAAACTTGACCAATAGTCTTAGAGTATCCAGTGAGG	二号区域
SEQ ID NO:60	ssODN_21	TCCCCACACTATCTCAATGCAAATATCTGTCTGAAACGGTAGATAACCCTGGCTAAACTCCACCCATGGGTTGGCCAGCCTTGCCT	三号区域
SEQ ID NO:61	ssODN_22	TATCCTCTTGGGGGCCCCTTCCCCACACTATCTCAATGCAACAGAAACGGTCCCTGGCTAAACTCCACCCATGGGTTGGCCAGC	四号区域
SEQ ID NO:62	ssODN_23	CCTATGCCTAAAACACATTTCACAATCCCTGAACTTTTCAGATAATTGGTACATGCTTTAACTTTAAACTACAGGCCTCACTG	五号区域
SEQ ID NO:63	ssODN_24	CCTATGCCTAAAACACATTTCACAATCCCTGAACTTTTCAGATAAATTGGTACATGCTTTAACTTTAAACTACAGGCCTCACTG	五号区域
SEQ ID NO:64	ssODN_25	CCTATGCCTAAAACACATTTCACAATCCCTGAACTTTTCAGATAAAATTGGTACATGCTTTAACTTTAAACTACAGGCCTCACTG	五号区域
SEQ ID NO:65	ssODN_26	CCTATGCCTAAAACACATTTCACAATCCCTGAACTTTTCAGATAGAATTGGTACATGCTTTAACTTTAAACTACAGGCCTCACTG	五号区域

使用 sgRNA-spCas9 载体（PX459，pSpCas9(BB)-2A-Puro）表达表 2 中靶位点对应的 sgRNA，该载体表达 spCas9 蛋白，同时可表达长度为 100 nt 的 sgRNA，其 5′端为 20 nt 的指导序列，后面 80nt 为通用的 sgRNA 骨架序列（SEQ ID NO:66, GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU）。克隆 sgRNA 的方法是：分别合成 gRNA 对应的指导链序列 oligo（同表 2 中 20nt 靶位点序列）和互补链 oligo (指导链序列的 5’- 端加 cacc，若指导链 5’-端第一个核苷酸不是鸟嘌呤 G，则在指导链的 5’-端加 caccG；在互补链的 5’-端加 aaac，若指导链 5’-端加了一个 G，则在互补链的 3’-端加 C 与之互补配对)。将两种 oligo 等量混合，95℃热变形退火处理，然后连入采用限制性内切酶 Bbs I 消化过的 PX459 载体，并通过菌落 PCR 及 sanger 测序鉴定是否克隆成功。PCR 使用 SEQ ID NO:67（5′- 3′: GGACTATCATATGCTTACCGTAAC）和 SEQ ID NO:68（5′- 3′: GGCGGGCCATTTACCGTAAG）作为引物。克隆成功后即获得 8 种质粒，分别表达 SEQ ID NO:69 至 SEQ ID NO:76 的 sgRNA（表 4）。

表 4. PX459 载体表达的 sgRNA

SEQ ID	sgRNA	sgRNA 序列（5′- 3′）
SEQ ID NO:69	sgRNA-1	ACUGGAUACUCUAAGACUAU GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUG AAAAAGUGGCACCGAGUCGGUGCUUUU
SEQ ID NO:70	sgRNA-2	CUUGUCAAGGCUAUUGGUCA GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUG AAAAAGUGGCACCGAGUCGGUGCUUUU
SEQ ID NO:71	sgRNA-3	GUUUGCCUUGUCAAGGCUAU GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUG AAAAAGUGGCACCGAGUCGGUGCUUUU
SEQ ID NO:72	sgRNA-4	UGGUCAAGUUUGCCUUGUCA GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUG AAAAAGUGGCACCGAGUCGGUGCUUUU
SEQ ID NO:73	sgRNA-5	CUUGACCAAUAGCCUUGACA GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUG AAAAAGUGGCACCGAGUCGGUGCUUUU
SEQ ID NO:74	sgRNA-6	UAUCUGUCUGAAACGGUCCC GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUG AAAAAGUGGCACCGAGUCGGUGCUUUU
SEQ ID NO:75	sgRNA-7	AUGCAAAUAUCUGUCUGAAA GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUG AAAAAGUGGCACCGAGUCGGUGCUUUU
SEQ ID NO:76	sgRNA-8	UCCCUGAACUUUUCAAAAAU GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUG AAAAAGUGGCACCGAGUCGGUGCUUUU

将表达 8 种质粒分别转染 293T 细胞，方法参考 Lipofectamine 2000™ 试剂盒的方法。转染后的细胞培养 4 天，收集细胞提取基因组 DNA。分别使用 SEQ ID NO:77 和 SEQ ID NO:78，以及 SEQ ID NO:79 和 SEQ ID NO:78，采用 KOD Plus 高保真酶 PCR 法扩增 HBG1 和 HBG2 启动子基因片段，得到的 PCR 产物纯化后分别使用 SEQ ID NO:77 和 SEQ ID NO:79 进行测序。测序结果得到的文件经 synthego 软件分析基因编辑效率，如图 3 所示，除 sgRNA-8 的编辑效率相对较低外（约 20%），其它组编辑总体突变（插入缺失，Indel）效率在 40-50%左右。

表 5. PCR 及 sanger 测序引物

SEQ ID	位点	sgRNA 识别位点及 PAM（5′- 3′）
SEQ ID NO:77	HBG1_F	TACTGCGCTGAAACTGTGGCTT
SEQ ID NO:78	HBG1/2_R	CTTCCCAGGGTTTCTCCTCCAG
SEQ ID NO:79	HBG2_F	TGCACTGAAACTGTTGCTTTATAGGA

实施例 2： ssODN 与 sgRNA-spCas9 载体共转 293T 细胞进行基因编辑

将上述 8 种 sgRNA-spCas9 载体分别与恰当的 ssODN 共转染入 293T 细胞中以检测添加 ssODN 后是否可以诱导基因编辑后形成预期的突变类型（发生 HDR 修复）。根据 sgRNA 切点的位置，将 8 种 sgRNA-spCas9 载体按照表 6 中所示的 sgRNA + ssODN 组合方法，分别形成 30 种组合。利用 Lipofectamine 2000™ 试剂，在转染质粒的同时添 ssODN，将 ssODN 和质粒一同递送到细胞中。转染后的细胞培养 4 天，收集细胞提取基因组 DNA。由于 HBG1 和 HBG2 的基因编辑效率几乎是正相关的，而且往往 HBG2 的基因编辑效率稍高于 HBG1（图 3），因此后续实验只选择 HBG2 基因组基因编辑效率进行对比，按照实施例 1 中的方法，PCR 扩增 HBG2 启动子基因片段，纯化进行测序。测序结果经 synthego 软件分析基因编辑效率显示，添加 ssODN 与 sgRNA-spCas9 载体共转染可以提升编辑效率，同时大多数组诱导约 10-20%的基因编辑形成预期的突变类型（图 4）。

表 6. ssODN 与 sgRNA-spCas9 载体共转组合

组合	转染复合物	组合	转染复合物
组合 1	sgRNA-1 + ssODN_1	组合 16	sgRNA-2 + ssODN_16
组合 2	sgRNA-1 + ssODN_2	组合 17	sgRNA-2 + ssODN_17
组合 3	sgRNA-1 + ssODN_3	组合 18	sgRNA-2 + ssODN_18
组合 4	sgRNA-1 + ssODN_4	组合 19	sgRNA-3 + ssODN_17
组合 5	sgRNA-2 + ssODN_5	组合 20	sgRNA-3 + ssODN_18
组合 6	sgRNA-2 + ssODN_6	组合 21	sgRNA-4 + ssODN_17
组合 7	sgRNA-2 + ssODN_7	组合 22	sgRNA-4 + ssODN_18
组合 8	sgRNA-2 + ssODN_8	组合 23	sgRNA-5 + ssODN_19
组合 9	sgRNA-2 + ssODN_9	组合 24	sgRNA-5 + ssODN_20
组合 10	sgRNA-2 + ssODN_10	组合 25	sgRNA-6 + ssODN_21
组合 11	sgRNA-2 + ssODN_11	组合 26	sgRNA-7 + ssODN_22
组合 12	sgRNA-2 + ssODN_12	组合 27	sgRNA-8 + ssODN_23
组合 13	sgRNA-2 + ssODN_13	组合 28	sgRNA-8 + ssODN_24
组合 14	sgRNA-2 + ssODN_14	组合 29	sgRNA-8 + ssODN_25
组合 15	sgRNA-2 + ssODN_15	组合 30	sgRNA-8 + ssODN_26

实施例 3： 电穿孔将 ssODN 与 sgRNA-spCas9 载体共转入 K562 细胞进行基因编辑

利用脂质体（如 Lipofectamine 2000™ 试剂）将质粒转入悬浮细胞的效率比较低，通常转染悬浮细胞时一般采用电穿孔法导入。取实施例 2 中部分 sgRNA 与 ssODN 的组合，通过 lonza-4D 电转仪导入 K562 细胞，使用仪器自带的 K562 细胞转染程序。转染后的细胞培养 4 天，收集细胞提取基因组 DNA，PCR 扩增 HBG2 启动子基因片段进行测序。测序结果经 synthego 软件分析基因编辑效率显示，K562 中电转 sgRNA-1、sgRNA-2、sgRNA-5 与 ssODN 的组合可达到较高编辑效率（图 5），略高于 293T 细胞，电转 ssODN 与 sgRNA-spCas9 载体诱导形成预期的突变类型也略微高于 293T 细胞。

实施例 4： 电穿孔将 ssODN 与 spCas9 蛋白/sgRNA RNP 共转入 K562 细胞实现高效基因编辑

相对于质粒表达 CRISPR-Cas 系统，以核糖核蛋白（RNP）的形式递送 CRISPR-Cas 系统被证实基因编辑效率高而且不易诱导细胞凋亡。利用商业化的 spCas9 蛋白及化学合成的 sgRNA-1、sgRNA-2、sgRNA-5（序列同表 4 中对应的 sgRNA，但末端进行化学修饰），分别在体外孵育包装 RNP-1、RNP-2、RNP-5。通过电穿孔将 RNP 与部分 ssODN 组合递送入 K562 细胞。转染后的细胞培养 4 天，收集细胞提取基因组 DNA，PCR 扩增 HBG2 启动子基因片段进行测序。测序结果经 synthego 软件分析基因编辑效率显示（图 6），相对于没有添加 ssODN 的组合，K562 中电转 RNP 时加入 ssODN 显著地提高总体基因编辑效率约两倍（图 6）。同时，电转 ssODN 和 RNP 可以诱导更高比例的基因编辑形成预期的突变类型（图 6）。

实施例 5： 电穿孔将 ssODN 与 spCas9 蛋白/sgRNA RNP 导入人类造血干/祖细胞实现高效基因编辑并显著提高 HBG 表达

动员外周血来源的人 CD34 阳性细胞（mPBSC）复苏后，在富含人类细胞因子（SCF、TPO、Flt3L, 各 100ng/ml）的 StemSpan 无血清扩增培养基(SFEM)中培养两天，采用 lonza-4D 电转仪自带的 EO-100 程序将 ssODN 与 spCas9/sgRNA RNP 递送到 mPBSC 细胞中。电转完成后在富含人类细胞因子（SCF/TPO/Flt3L, 各 100ng/ml）的 SFEM 中恢复培养一天，然后分成三阶段诱导 mPBSC 向红细胞分化。第一阶段：在含有 EPO、SCF、人 AB 血清、胰岛素、转铁蛋白、肝素、IL-3、氢化可的松等添加物的 IMDM 培养液中培养 7 天。在诱导分化第 5 天取出 2 x 10e5 细胞提取基因组 DNA，然后 PCR 扩增 HBG2 启动子片段送 sanger 测序。与 K562 细胞中的结果类似，mPBSC 细胞中电转 ssODN 与 spCas9/sgRNA RNP 可产生高效的基因编辑，其中部分组基因编辑形成预期的突变类型达到 30%（图 7A-C）。

电转过的 mPBSC 细胞及未处理的 mPBSC 细胞（对照，NC）经过第一阶段分化诱导 7 天后，在含有 EPO、SCF、人 AB 血清、胰岛素、转铁蛋白、肝素等添加物的 IMDM 培养液中培养 4 天，最后在在含有 EPO、人 AB 血清、胰岛素、转铁蛋白、肝素等添加物的 IMDM 培养液中培养 7 天。诱导分化 18 天后，电转过的 mPBSC 细胞及未处理的 mPBSC 细胞离心后均可见明显深红色，其中含有大量红细胞，表明电转加工过的 mPBSC 细胞可正常分化为红细胞。

在诱导分化接近结束的时候，取部分细胞提取 mRNA，用逆转录-实时定量 PCR（RT-qPCR）检测 HBG mRNA 的表达水平。相对于没有转染的细胞（NC），RNP-2 与 ssODN_6、ssODN_13、ssODN_14、ssODN_15、ssODN_17 共转染均提高了 HBG mRNA 的表达水平（图 8），HBG mRNA 升高最高可达 3 倍以上。

使用糖化血红蛋白 A1c 检测系统（VARIANT II TURBO HbA1c kit-2.0），对 RNP-2 组电转的造血干细胞分化成红细胞后的血红蛋白进行 HPLC 法检测，结果显示，相对于未进行基因编辑的细胞，添加 ssODN 进行基因编辑后，胎儿血红蛋白的含量（HbF）都有显著提高，在 27%-36%之间，相对于对照组 13.3% 有 2 倍以上的上调（图 9）。

上述实施例的结果表明，本发明提供了一种与现有技术不同的增强 γ-珠蛋白基因表达的基因编辑方法。本发明公布的 ssODN 用于 γ-珠蛋白基因编辑时，可以指导基因编辑后在 HBG1 和 HBG2 基因调控区预定的位点形成预期的突变，从而形成了激活性 GATA 元件，该元件在基因编辑的细胞（例如造血干细胞）分化成红细胞后发挥正调控作用，激活或显著提升 γ-珠蛋白基因表达。因此，该发明在 β-血红蛋白病的基因治疗中有潜在应用价值。

以上所述实施例的各技术特征可以进行任意的组合，为使描述简洁，未对上述实施例中的各个技术特征所有可能的组合都进行描述，然而，只要这些技术特征的组合不存在矛盾，都应当认为是本说明书记载的范围。

以上所述实施例仅表达了本发明的几种实施方式，其描述较为具体和详细，但并不能因此而理解为对发明专利范围的限制。应当指出的是，对于本领域的普通技术人员来说，在不脱离本发明构思的前提下，

还可以做出若干变形和改进，这些都属于本发明的保护范围。因此，本发明专利的保护范围应以所附权利要求为准。

序列表

<110> 广州瑞风生物科技有限公司

<120> 用于激活γ-珠蛋白基因表达的方法和组合物

<130> 2020

<160> 79

<170> PatentIn version 3.3

<210> 1
<211> 1500
<212> DNA
<213> Homo sapiens

<400> 1
tctttctggc cttttagcca tctgtatcaa tgagcagata taagctttac acaggatcat 60
gaaggatgaa agaatttcac caatattata ataatttcaa tcaacctgat agcttagggg 120
ataaactaat ttgaagatac agcttgcctc cgataagcca gaattccaga gcttctggca 180
ttataatcta gcaaggttag agatcatgga tcactttcag agaaaaacaa aaacaaacta 240
accaaaagca aaacagaacc aaaaaaccac cataaatact tcctaccctg ttaatggtcc 300
aatatgtcag aaacagcact gtgttagaaa taaagctgtc taaagtacac taatattcga 360
gttataatag tgtgtggact attagtcaat aaaaacaacc cttgcctctt tagagttgtt 420
ttccatgtac acgcacatct tatgtcttag agtaagattc cctgagaagt gaacctagca 480
tttatacaag ataattaatt ctaatccaca gtacctgcca aagaacattc taccatcatc 540
tttactgagc atagaagagc tacgccaaaa ccctgggtca tcagccagca cacacactta 600
tccagtggta aatacacatc atctggtgta tacatacata cctgaatatg gaatcaaata 660
tttttctaag atgaaacagt catgatttat ttcaaatagg tacggataag tagatattga 720
ggtaagcatt aggtcttata ttatgtaaca ctaatctatt actgcgctga aactgtggct 780
ttatagaaat tgttttcact gcactattga gaaattaaga gataatggca aaagtcacaa 840
agagtatatt caaaaagaag tatagcactt tttccttaga aaccactgct aactgaaaga 900
gactaagatt tgtcccgtca aaaatcctgg acctatgcct aaaacacatt tcacaatccc 960
tgaacttttc aaaaattggt acatgcttta gctttaaact acaggcctca ctggagctag 1020
agacaagaag gtaaaaaacg gctgacaaaa gaagtcctgg tatcctctat gatgggagaa 1080
ggaaactagc taaagggaag aataaattag agaaaaactg gaatgactga atcggaacaa 1140
ggcaaaggct ataaaaaaaa ttagcagtat cctcttgggg gccccttccc cacactatct 1200
caatgcaaat atctgtctga aacggtccct ggctaaactc cacccatggg ttggccagcc 1260
ttgccttgac caatagcctt gacaaggcaa acttgaccaa tagtcttaga gtatccagtg 1320
aggccagggg ccggcggctg gctagggatg aagaataaaa ggaagcaccc ttcagcagtt 1380
ccacacactc gcttctggaa cgtctgaggt tatcaataag ctcctagtcc agacgccatg 1440
ggtcatttca cagaggagga caaggctact atcacaagcc tgtggggcaa ggtgaatgtg 1500

<210> 2
<211> 1500
<212> DNA

<213> Homo sapiens

<400> 2
atctttctgg tcttttagcc gcctaacact ttgagcagat ataagcctta cacaggatta 60
tgaagtctga aaggattcca ccaatattat tataattcct atcaacctga taggttaggg 120
gaaggtagag ctctcctcca ataagccaga tttccagagt ttctgacgtc ataatctacc 180
aaggtcatgg atcgagttca gagaaaaaac aaaagcaaaa ccaaacctac caaaaaataa 240
aaatcccaaa gaaaaaataa agaaaaaaac agcatgaata cttcctgcca tgttaagtgg 300
ccaatatgtc agaaacagca ctgagttaca gataaagatg tctaaactac agtgacatcc 360
cagctgtcac agtgtgtgga ctattagtca ataaaacagt ccctgcctct taagagttgt 420
tttccatgca aatacatgtc ttatgtctta gaataagatt ccctaagaag tgaacctagc 480
atttatacaa gataattaat tctaatccat agtatctggt aaagagcatt ctaccatcat 540
ctttaccgag catagaagag ctacaccaaa accctgggtc atcagccagc acatacactt 600
atccagtgat aaatacacat catcgggtgc ctacatacat acctgaatat aaaaaaaata 660
cttttgctga gatgaaacag gcgtgattta tttcaaatag gtacggataa gtagatattg 720
aagtaaggat tcagtcttat attatattac ataacattaa tctattcctg cactgaaact 780
gttgctttat aggatttttc actacactaa tgagaactta agagataatg gcctaaaacc 840
acagagagta tattcaaaga taagtatagc acttcttatt tggaaaccaa tgcttactaa 900
atgagactaa gacgtgtccc atcaaaaatc ctggacctat gcctaaaaca catttcacaa 960
tccctgaact tttcaaaaat tggtacatgc tttaacttta aactacaggc ctcactggag 1020
ctacagacaa gaaggtgaaa aacggctgac aaaagaagtc ctggtatctt ctatggtggg 1080
agaagaaaac tagctaaagg gaagaataaa ttagagaaaa attggaatga ctgaatcgga 1140
acaaggcaaa ggctataaaa aaaattaagc agcagtatcc tcttgggggc cccttcccca 1200
cactatctca atgcaaatat ctgtctgaaa cggtccctgg ctaaactcca cccatgggtt 1260
ggccagcctt gccttgacca atagccttga caaggcaaac ttgaccaata gtcttagagt 1320
atccagtgag gccaggggcc ggcggctggc tagggatgaa gaataaaagg aagcaccctt 1380
cagcagttcc acacactcgc ttctggaacg tctgaggtta tcaataagct cctagtccag 1440
acgccatggg tcatttcaca gaggaggaca aggctactat cacaagcctg tggggcaagg 1500

<210> 3
<211> 27
<212> DNA
<213> Homo sapiens

<400> 3
ttgaccaata gtcttagagt atccagt 27

<210> 4
<211> 27
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 4
ttgaccaata gtgatagagt atccagt 27

<210> 5
<211> 27

<212> DNA
<213> 人工序列(Artificial Sequence)
<400> 5
ttgaccaata gagatagagt atccagt 27

<210> 6
<211> 26
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 6
ttgaccaata agatagagta tccagt 26

<210> 7
<211> 26
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 7
ttgaccaata tgatagagta tccagt 26

<210> 8
<211> 32
<212> DNA
<213> Homo sapiens

<400> 8
cagccttgcc ttgaccaata gccttgacaa gg 32

<210> 9
<211> 32
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 9
cagccttgcc ttgacagata gccttgacaa gg 32

<210> 10
<211> 32
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 10
cagccttgcc ttgactgata gccttgacaa gg 32

<210> 11
<211> 32
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 11
cagccttgcc ttgagagata gccttgacaa gg 32

<210> 12
<211> 32
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 12
cagccttgcc ttgagtgata gccttgacaa gg 32

<210> 13
<211> 32
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 13
cagccttgcc ttgaaagata gccttgacaa gg 32
<210> 14
<211> 32
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 14
cagccttgcc ttgaatgata gccttgacaa gg 32

<210> 15
<211> 32
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 15
cagccttgcc ttgatagata gccttgacaa gg 32

<210> 16
<211> 32

<212> DNA
<213> 人工序列(Artificial Sequence)
<400> 16
cagccttgcc ttgattgata gccttgacaa gg 32

<210> 17
<211> 32
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 17
cagccttgcc ttgataaata gccttgacaa gg 32

<210> 18
<211> 33
<212> DNA
<213> 人工序列(Artificial Sequence)
<400> 18
cagccttgcc ttgataaaat agccttgaca agg 33

<210> 19
<211> 31
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 19
cagccttgcc ttgataatag ccttgacaag g 31

<210> 20
<211> 28
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 20
cagccttgcc ttgatagcct tgacaagg 28

<210> 21
<211> 19
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 21
cagccttgcc ttgataagg 19

<210> 22
<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 22
cagccttgcc ttgataaagg 20

<210> 23
<211> 23
<212> DNA
<213> Homo sapiens
<400> 23
tatctgtctg aaacggtccc tgg 23

<210> 24
<211> 29
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 24
tatctgtctg aaacggtaga taaccctgg 29

<210> 25
<211> 33
<212> DNA
<213> Homo sapiens

<400> 25
cactatctca atgcaaatat ctgtctgaaa cgg 33

<210> 26
<211> 25
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 26
cactatctca atgcaacaga aacgg 25

<210> 27
<211> 23

<212> DNA
<213> Homo sapiens
<400> 27
tccctgaact tttcaaaaat tgg 23

<210> 28
<211> 24
<212> DNA
<213> 人工序列(Artificial Sequence)
<400> 28
tccctgaact tttcagataa ttgg 24

<210> 29
<211> 25
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 29
tccctgaact tttcagataa attgg 25

<210> 30
<211> 26
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 30
tccctgaact tttcagataa aattgg 26

<210> 31
<211> 26
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 31
tccctgaact tttcagatag aattgg 26

<210> 32
<211> 23
<212> DNA
<213> Homo sapiens

<400> 32
actggatact ctaagactat tgg 23

<210> 33
<211> 23
<212> DNA
<213> Homo sapiens

<400> 33
cttgtcaagg ctattggtca agg 23

<210> 34
<211> 23
<212> DNA
<213> Homo sapiens

<400> 34
gtttgccttg tcaaggctat tgg 23

<210> 35
<211> 23
<212> DNA
<213> Homo sapiens

<400> 35
tggtcaagtt tgccttgtca agg 23

<210> 36
<211> 23
<212> DNA
<213> Homo sapiens

<400> 36
cttgaccaat agccttgaca agg 23

<210> 37
<211> 23
<212> DNA
<213> Homo sapiens

<400> 37
tatctgtctg aaacggtccc tgg 23

<210> 38

<211> 23
<212> DNA
<213> Homo sapiens

<400> 38
atgcaaatat ctgtctgaaa cgg 23

<210> 39
<211> 23
<212> DNA
<213> Homo sapiens

<400> 39
tccctgaact tttcaaaaat tgg 23

<210> 40
<211> 84
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 40
cctagccagc cgccggcccc tggcctcact ggatactcta tcactattgg tcaagtttgc 60
cttgtcaagg ctattggtca aggc 84

<210> 41
<211> 84
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 41
cctagccagc cgccggcccc tggcctcact ggatactcta tctctattgg tcaagtttgc 60
cttgtcaagg ctattggtca aggc 84

<210> 42
<211> 83
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 42
cctagccagc cgccggcccc tggcctcact ggatactcta tcttattggt caagtttgcc 60
ttgtcaaggc tattggtcaa ggc 83

<210> 43

<211> 83
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 43
cctagccagc cgccggcccc tggcctcact ggatactcta tcatattggt caagtttgcc 60
ttgtcaaggc tattggtcaa ggc 83

<210> 44
<211> 82
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 44
ctctaagact attggtcaag tttgccttgt caaggctatc tgtcaaggca aggctggcca 60
acccatgggt ggagtttagc ca 82

<210> 45
<211> 82
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 45
ctctaagact attggtcaag tttgccttgt caaggctatc agtcaaggca aggctggcca 60
acccatgggt ggagtttagc ca 82

<210> 46
<211> 82
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 46
ctctaagact attggtcaag tttgccttgt caaggctatc tctcaaggca aggctggcca 60
acccatgggt ggagtttagc ca 82

<210> 47
<211> 82
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 47
ctctaagact attggtcaag tttgccttgt caaggctatc actcaaggca aggctggcca 60
acccatgggt ggagtttagc ca 82

<210> 48
<211> 82
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 48
ctctaagact attggtcaag tttgccttgt caaggctatc tttcaaggca aggctggcca 60
acccatgggt ggagtttagc ca 82

<210> 49
<211> 82
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 49
ctctaagact attggtcaag tttgccttgt caaggctatc attcaaggca aggctggcca 60
acccatgggt ggagtttagc ca 82

<210> 50
<211> 82
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 50
ctctaagact attggtcaag tttgccttgt caaggctatc tatcaaggca aggctggcca 60
acccatgggt ggagtttagc ca 82

<210> 51
<211> 82
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 51
ctctaagact attggtcaag tttgccttgt caaggctatc aatcaaggca aggctggcca 60
acccatgggt ggagtttagc ca 82

<210> 52
<211> 82
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 52
ctctaagact attggtcaag tttgccttgt caaggctatt tatcaaggca aggctggcca 60
acccatgggt ggagtttagc ca 82

<210> 53
<211> 83
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 53
ctctaagact attggtcaag tttgccttgt caaggctatt ttatcaaggc aaggctggcc 60
aacccatggg tggagtttag cca 83

<210> 54
<211> 81
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 54
ctctaagact attggtcaag tttgccttgt caaggctatt atcaaggcaa ggctggccaa 60
cccatgggtg gagtttagcc a 81

<210> 55
<211> 78
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 55
ctctaagact attggtcaag tttgccttgt caaggctatc aaggcaaggc tggccaaccc 60
atgggtggag tttagcca 78

<210> 56
<211> 81
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 56
cctcactgga tactctaaga ctattggtca agtttgcctt atcaaggcaa ggctggccaa 60
cccatgggtg gagtttagcc a 81

<210> 57
<211> 82
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 57
cctcactgga tactctaaga ctattggtca agtttgcctt tatcaaggca aggctggcca 60

acccatgggt ggagtttagc ca 82

<210> 58
<211> 81
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 58
tggctaaact ccacccatgg gttggccagc cttgccttga taaggcaaac ttgaccaata 60
gtcttagagt atccagtgag g 81

<210> 59
<211> 82
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 59
tggctaaact ccacccatgg gttggccagc cttgccttga taaaggcaaa cttgaccaat 60
agtcttagag tatccagtga gg 82

<210> 60
<211> 86
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 60
tccccacact atctcaatgc aaatatctgt ctgaaacggt agataaccct ggctaaactc 60
cacccatggg ttggccagcc ttgcct 86

<210> 61
<211> 84
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 61
tatcctcttg ggggcccctt ccccacacta tctcaatgca acagaaacgg tccctggcta 61
aactccaccc atgggttggc cagc 84

<210> 62
<211> 83
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 62
cctatgccta aaacacattt cacaatccct gaacttttca gataattggt acatgcttta 60

actttaaact acaggcctca ctg 83

<210> 63
<211> 84
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 63
cctatgccta aaacacattt cacaatccct gaacttttca gataaattgg tacatgcttt 60
aactttaaac tacaggcctc actg 84

<210> 64
<211> 85
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 64
cctatgccta aaacacattt cacaatccct gaacttttca gataaaattg gtacatgctt 60
taactttaaa ctacaggcct cactg 85

<210> 65
<211> 85
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 65
cctatgccta aaacacattt cacaatccct gaacttttca gatagaattg gtacatgctt 60
taactttaaa ctacaggcct cactg 85

<210> 66
<211> 80
<212> RNA
<213> 人工序列(Artificial Sequence)

<400> 66
guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc cguuaucaac uugaaaaagu 60
ggcaccgagu cggugcuuuu 80

<210> 67
<211> 24
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 67
ggactatcat atgcttaccg taac 24

<210> 68
<211> 20
<212> DNA
<213> 人工序列(Artificial Sequence)

<400> 68
ggcgggccat ttaccgtaag 24

<210> 69
<211> 100
<212> RNA
<213> 人工序列(Artificial Sequence)

<400> 69
acuggauacu cuaagacuau guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60
cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu 100

<210> 70
<211> 100
<212> RNA
<213> 人工序列(Artificial Sequence)

<400> 70
cuugucaagg cuauugguca guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60
cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu 100

<210> 71
<211> 100
<212> RNA
<213> 人工序列(Artificial Sequence)

<400> 71
guuugccuug ucaaggcuau guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60
cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu 100

<210> 72
<211> 100
<212> RNA
<213> 人工序列(Artificial Sequence)

<400> 72
uggucaaguu ugccuuguca guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60
cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu 100

<210> 73
<211> 100
<212> RNA
<213> 人工序列(Artificial Sequence)

<400> 73
cuugaccaau agccuugaca guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60
cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu 100

<210> 74
<211> 100
<212> RNA
<213> 人工序列(Artificial Sequence)

<400> 74
uaucugucug aaacgguccc guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60
cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu 100

<210> 75
<211> 100
<212> RNA
<213> 人工序列(Artificial Sequence)

<400> 75
augcaaauau cugucugaaa guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60
cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu 100

<210> 76
<211> 100
<212> RNA
<213> 人工序列(Artificial Sequence)

<400> 76
ucccugaacu uuucaaaaau guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60
cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu 100

<210> 77
<211> 22
<212> DNA
<213> Homo sapiens

<400> 77
tactgcgctg aaactgtggc tt 22

<210> 78
<211> 22
<212> DNA
<213> Homo sapiens

<400> 78
cttcccaggg tttctcctcc ag 22

<210> 79
<211> 26
<212> DNA
<213> Homo sapiens

<400> 79
tgcactgaaa ctgttgcttt atagga 26