



(12) 发明专利申请

(10) 申请公布号 CN 104704110 A

(43) 申请公布日 2015. 06. 10

(21) 申请号 201380045574. 4

(22) 申请日 2013. 08. 29

(30) 优先权数据

61/694, 693 2012. 08. 29 US

(85) PCT国际申请进入国家阶段日

2015. 02. 28

(86) PCT国际申请的申请数据

PCT/US2013/057214 2013. 08. 29

(87) PCT国际申请的公布数据

W02014/036219 EN 2014. 03. 06

(71) 申请人 桑格摩生物科学股份有限公司

地址 美国加利福尼亚州

(72) 发明人 格雷戈里·J·科斯特

菲利普·D·格雷戈里

德米特里·格斯钦

迈克尔·C·福尔摩斯

杰弗里·C·米勒 戴维·帕绍恩

爱德华·J·雷巴尔

安德里亚斯·赖克

费奥多·乌尔诺夫 张雷

(74) 专利代理机构 上海胜康律师事务所 31263

代理人 樊英如 李献忠

(51) Int. Cl.

C12N 5/071(2006. 01)

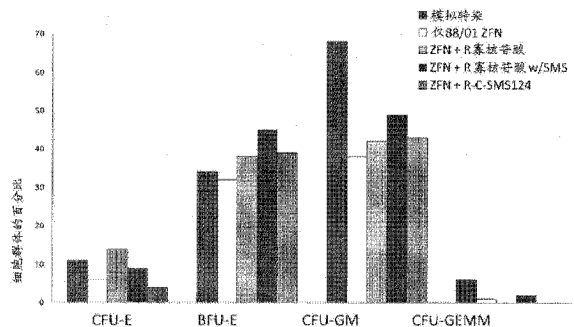
权利要求书1页 说明书46页 附图14页

(54) 发明名称

用于治疗遗传性病状的方法和组合物

(57) 摘要

本文提供用于遗传性疾病的方法和组合物。



1. 一种锌指蛋白,其包含4、5或6个包含识别螺旋区域的锌指结构域,其中所述锌指蛋白包含以表1的单行所示顺序的所述识别螺旋区域。

2. 一种融合蛋白,其包含权利要求1的锌指蛋白和野生型或工程改造的裂解结构域或野生型或工程改造的裂解半结构域。

3. 一种多核苷酸,其编码一种或多种权利要求1或权利要求2的蛋白。

4. 一种分离细胞,其包含一种或多种根据权利要求2的融合蛋白或一种或多种根据权利要求3的多核苷酸。

5. 根据权利要求4所述的细胞,其中所述细胞选自由红血细胞(RBC)或前体细胞组成的组。

6. 根据权利要求5所述的细胞,其中所述前体细胞是CD34+造血干细胞。

7. 一种试剂盒,其包含根据权利要求1或权利要求2的蛋白或根据权利要求3的多核苷酸。

8. 一种改变细胞中的珠蛋白基因表达的方法,所述方法包括:

在一定条件下向所述细胞中引入一种或多种根据权利要求1或权利要求2的蛋白或一种或多种根据权利要求3的多核苷酸,使得所述一种或多种蛋白表达并且改变所述珠蛋白基因的表达。

9. 根据权利要求8所述的方法,其中所述蛋白增加珠蛋白基因的表达。

10. 根据权利要求9所述的方法,其中所述珠蛋白基因是 $\gamma$ 珠蛋白或 $\beta$ 珠蛋白基因。

11. 根据权利要求8至10中任一项所述的方法,其进一步包括将供体序列整合至所述细胞的基因组中。

12. 根据权利要求11所述的方法,其中使用病毒载体、寡核苷酸或质粒将所述供体序列引入所述细胞中。

13. 根据权利要求8至12中任一项所述的方法,其中所述细胞是红血细胞(RBC)前体细胞。

14. 根据权利要求13所述的细胞,其中所述RBC前体细胞是CD34+造血干细胞。

15. 根据权利要求8至14中任一项所述的方法,其中所述供体序列包含在内源启动子控制下的转基因。

16. 根据权利要求8至14中任一项所述的方法,其中所述供体序列包含在外源启动子控制下的转基因。

## 用于治疗遗传性病状的方法和组合物

[0001] 相关申请的交叉引用

[0002] 本申请要求 2012 年 8 月 29 日提交的美国临时申请号 61/694,693 的权益,所述临时申请的公开内容特此以全文引用的方式并入。

### 技术领域

[0003] 本公开属于造血干细胞的基因组工程领域,尤其是用于治疗血红蛋白病。

[0004] 背景

[0005] 基因疗法对于人类医学的新时代保持巨大的潜力。这些方法将允许治疗迄今尚未被标准的医学实践解决的病状。特别有前途的一个领域是对细胞进行基因工程改造以使得该细胞表达先前没有在该细胞中产生的产物的能力。这种技术的用途的实例包括插入编码新治疗蛋白的基因、插入编码在细胞中或个体中缺乏的蛋白的编码序列、在含有突变基因序列的细胞中插入野生型基因,和插入编码结构核酸例如微 RNA 或 siRNA 的序列。

[0006] 转基因可以通过多种方式递送至细胞,以使得转基因整合至细胞自身的基因组中并且保持在其中。近年来,已开发了转基因整合策略,其使用位点特异性核酸酶裂解以靶向插入至所选基因组基因座中(参见例如共同拥有的美国专利 7,888,121)。可利用对靶向基因具有特异性的核酸酶以使得转基因构建体通过同源介导的修复(HDR)插入或在非同源末端接合(NHEJ)驱动的方法期间通过末端捕捉插入。靶向基因座包括“安全港”基因座,例如 CCR5 基因、CXCR4 基因、PPP1R12C(也称为 AAVS1)基因、白蛋白基因或 Rosa 基因。参见例如美国专利公布号 20080299580;20080159996;201000218264;20110301073;20130177983 和 20130177960 以及美国临时申请号 61/823,689。与依赖于转基因的随机整合的经典的整合方法相比,核酸酶介导的整合提供改进的转基因表达、增加的安全性和表达持久性的前景,因为它允许精确的转基因定位,使得基因沉默或附近致癌基因活化的风险最小。

[0007] 红血细胞(RBC)或红细胞是血液的主要细胞组分。实际上,RBC在人类中占细胞的四分之一。人类中的成熟RBC没有细胞核和许多其它细胞器,并且全部是血红蛋白,血红蛋白是存在于RBC中的金属蛋白,其功能是携带氧气至组织以及携带二氧化碳离开组织并返回到肺部以去除。所述蛋白构成RBC干重的约97%并且它使血液的携氧能力提高约70倍。血红蛋白是包含两条 $\alpha$ 样珠蛋白链和两条 $\beta$ 样珠蛋白链和4个血红素基团的异四聚体。在成人中, $\alpha_2\beta_2$ 四聚体被称为血红蛋白A(HbA)或成人血红蛋白。通常, $\alpha$ 和 $\beta$ 珠蛋白链以近似1:1比率合成并且这个比率在血红蛋白和RBC稳定化方面似乎是关键的。实际上,在一类珠蛋白基因不足表达的一些情况下(参见下文),降低另一类型的珠蛋白的表达(例如使用特异性siRNA),从而恢复这种1:1比率,减轻了突变细胞表型的一些方面(参见Voon等,(2008)Haematologica 93(8):1288)。在发育中的胎儿中,产生不同形式的血红蛋白(胎儿血红蛋白(HbF)),其相比于血红蛋白A对氧具有更高的结合亲和力,以使得氧可经由母体血流递送至婴儿系统。胎儿血红蛋白也含有两条 $\alpha$ 珠蛋白链,但代替成人 $\beta$ 珠蛋白链,其具有两条胎儿 $\gamma$ 珠蛋白链(即,胎儿血红蛋白是 $\alpha_2\gamma_2$ )。在妊娠约30周时,胎儿中 $\gamma$ 珠蛋白的合成开始下降,而 $\beta$ 珠蛋白的产生增加。至大约10月龄时,新生儿的

血红蛋白几乎全部是  $\alpha 2 \beta 2$ , 但一些 HbF 持续到成年期 (总血红蛋白的约 1-3%)。从产生  $\gamma$  至产生  $\beta$  的开关的调节相当复杂, 并且主要涉及  $\gamma$  珠蛋白的表达下调与  $\beta$  珠蛋白表达的同时上调。

[0008] 在编码血红蛋白链的序列中的遗传缺陷可造成多种疾病, 称为血红蛋白病, 包括镰状细胞贫血和地中海贫血。在大多数患有血红蛋白病的患者中, 仍存在编码  $\gamma$  珠蛋白的基因, 但由于如上所述在围产期发生的正常基因抑制, 表达相对较低。

[0009] 据估计, 在美国有 1/5000 的人群患有镰状细胞疾病 (SCD), 大多是撒哈拉以南非洲裔人群。镰状细胞杂合性似乎对于防止疟疾存在益处, 因此这种特性可能已经随着时间推移经过选择, 使得据估计在撒哈拉以南非洲, 有三分之一的人口具有镰状细胞特性。镰状细胞疾病是由  $\beta$  珠蛋白基因中的突变造成, 其中氨基酸 #6 的谷氨酸被缬氨酸取代 (在 DNA 水平上是 GAG 至 GTG), 其中所产生的血红蛋白被称为“血红蛋白 S”或“HbS”。在较低氧条件下, HbS 的脱氧形式的构象转变暴露出蛋白上 E 与 F 螺旋之间的疏水补丁。血红蛋白中  $\beta$  链位置 6 的缬氨酸的疏水残基能够与所述疏水补丁缔合, 造成 HbS 分子聚集并形成纤维状沉淀物。这些聚集体又造成 RBC 的异常或‘镰状化’, 导致细胞的柔性损失。镰状化 RBC 不再能够挤入毛细血管床中并且可以导致镰状细胞患者中的血管闭塞性危机。另外, 镰状 RBC 比正常 RBC 更脆弱, 并且易于溶血, 最终导致患者中出现贫血。

[0010] 镰状细胞患者的治疗和管理是终身主张, 涉及抗生素治疗、疼痛管理和在急性发作期间的输血。一种方法是使用羟基脲, 其通过增加  $\gamma$  珠蛋白产生而部分地发挥其作用。然而, 长期羟基脲疗法的长期副作用仍然未知, 并且治疗引起不希望的副作用并且在不同患者之间可能具有可变的功效。尽管镰状细胞治疗的功效增加, 但患者的预期寿命仍然仅为 55 至 59 岁并且疾病的相关发病率对患者的生活质量具有深远的影响。

[0011] 地中海贫血也是与血红蛋白有关的疾病并且通常涉及珠蛋白链的表达降低。这可以通过基因的调节区域中的突变或由导致表达降低的珠蛋白编码序列中的突变发生。 $\alpha$  地中海贫血与西非和南亚裔人群有关, 并且可能赋予抗疟疾性。 $\beta$  地中海贫血与通常来自希腊以及土耳其和意大利的沿海地区的地中海裔人群有关。地中海贫血的治疗通常涉及输血和铁螯合疗法。如果可以鉴定适当的供体, 则骨髓移植也被用于治疗患有严重地中海贫血的人, 但这种程序可能具有显著风险。

[0012] 已提出的一种用于治疗 SCD 和  $\beta$  地中海贫血的方法是增加  $\gamma$  珠蛋白的表达, 目的在于用 HbF 在功能上替换异常成人血红蛋白。如上所述, 用羟基脲治疗 SCD 患者被认为是成功的, 这在一定程度上是由于其对于增加的  $\gamma$  珠蛋白表达的影响。发现影响 HbF 再激活活性的第一组化合物是细胞毒性药物。在实验动物中使用 5-氮杂胞苷首次显示通过药理操纵重新合成  $\gamma$  珠蛋白的能力 (DeSimone (1982) Proc Natl Acad Sci USA 79 (14): 4428-31)。后续研究证实 5-氮杂胞苷在具有  $\beta$  地中海贫血和镰状细胞疾病的患者中增加 HbF 的能力 (Ley 等, (1982) N. Engl. J. Medicine, 307: 1469-1475; 和 Ley 等, (1983) Blood 62: 370-380)。另外, 短链脂肪酸 (例如丁酸酯和衍生物) 已显示在实验系统中增加 HbF (Constantoulakis 等, (1988) Blood 72 (6): 1961-1967)。此外, 存在一部分的具有被称为‘遗传性胎儿血红蛋白持续存在症’ (HPFH) 的病状的人类群体, 其中在成年期中持续存在升高量的 HbF (HPFH 杂合子中的 10-40%) (参见 Thein 等, (2009) Hum. Mol. Genet 18 (R2): R216-R223)。这是一种罕见的病状, 但在不存在任何相关  $\beta$  珠蛋白异常的情况下,



它不与任何显著临床表现相关联,即使当个体的血红蛋白的 100%是 HbF 时。当具有  $\beta$  地中海贫血的个体也具有共同发作的 HPFH 时, HbF 的表达可减轻疾病的严重程度。此外,镰状细胞疾病的自然过程的严重程度在不同患者之间可能显著变化,并且这种可变性在某种程度上可以追溯到如下事实:一些具有较轻疾病的个体表达较高水平的 HbF。

[0013] 一种提高 HbF 表达的方法涉及鉴定其产物在  $\gamma$  珠蛋白表达调节中起作用的基因。一种此类基因是 BCL11A,首次鉴定是由于其在淋巴细胞发育中的作用。BCL11A 编码锌指蛋白,该蛋白被认为是参与  $\gamma$  珠蛋白表达的阶段特异性调节。BCL11A 表达于成人红系前体细胞中并且其表达下调导致  $\gamma$  珠蛋白表达增加。另外,似乎 BCL11A mRNA 的剪接经过发育调节。在胚胎细胞中,似乎主要表达较短的 BCL11A mRNA 变体,称为 BCL11A-S 和 BCL11A-XS,而在成体细胞中,主要表达较长的 BCL11A-L 和 BCL11A-XL mRNA 变体。参见 Sankaran 等, (2008) *Science* 322 第 1839 页。BCL11A 蛋白似乎与  $\beta$  珠蛋白基因座相互作用以改变其构象以及因此其在不同发育阶段的表达。另外,另一种调节蛋白 KLF1 似乎参与  $\gamma$  珠蛋白表达的调节。已发现 KLF1 水平与 BCL11A 水平成正比,并且两者都与  $\gamma$  珠蛋白水平成反比。例如,在具有 HbF 的持续表达的马耳他家族中,该家族携带 KLF1 基因的杂合突变 (Borg 等, (2010) *Nat Genet*, 42(9):801-805)。KLF1 基因产物似乎在体内直接结合至 BCL11A 基因,并且因此可造成其上调 (参见 Borg 等,同上;Bieker (2010) *Nat Genet* 42(9):733-734; Zhou 等, (2010) *Nat Genet* 42(9):742-744)。因此,如果 KLF1 刺激 BCL11A 表达,则所诱导的 BCL11A 的作用将导致  $\gamma$  珠蛋白和 HbF 产生的抑制。已提出了靶向 BCL11A 基因的抑制性 RNA 的用途 (参见例如美国专利公开 20110182867),但这种技术具有若干潜在缺点,即可能无法实现完全敲低,这类 RNA 的递送可能有问题并且 RNA 必须连续存在,从而需要终生多次治疗。

[0014]  $\alpha$  地中海贫血也流行于人类群体,特别是在亚洲,并且一些类型的  $\alpha$  珠蛋白异常被认为是人类中最常见的遗传性病症。在世界上的热带和亚热带地区, $\alpha$  珠蛋白病症发现于群体的 80-90% 中 (参见 Harteveld 和 Higgs (2010) *Orphanet Journal of Rare Diseases* 5:13)。

[0015] 人类在 16 号染色体上携带串联的 2 个拷贝的  $\alpha$  珠蛋白基因 ( $\alpha 1$  和  $\alpha 2$ ),因此在正常的二倍体细胞中总共存在 4 个拷贝。对于  $\alpha$  珠蛋白 mRNA, $\alpha 2$  基因通常是  $\alpha 1$  基因的 2-3 倍。这两种基因的串联构造可能与  $\alpha$  地中海贫血患者中  $\alpha$  珠蛋白基因大量缺失的高发病率有关,其中非功能性  $\alpha$  珠蛋白基因的数目通常与任何  $\alpha$  地中海贫血的严重程度直接相关 (参见 Chui 等, (2003) *Blood* 101(3):791)。一个拷贝的缺失似乎相当常见 (30% 的非裔美国人和 60-80% 的居住在沙特阿拉伯、印度和泰国的人),并且通常在个体中不明显,除非进行基因测试。两个拷贝的缺失,无论是在相同染色体上 (顺式) 或者每个染色体一个 (反式),都可造成患病个人具有轻度贫血。当三个  $\alpha$  珠蛋白基因缺失时,使得个体仅具有一个正常工作的  $\alpha$  珠蛋白基因,测得中度贫血,但更重要的是,关键的  $\alpha$  珠蛋白与  $\beta$  珠蛋白的比率被破坏。通常在仅具有一个功能性  $\alpha$  珠蛋白基因的患者中观察到包含四个  $\beta$  珠蛋白链的  $\beta 4$  四聚体,这是被称为 HbH 的病状。所述  $\beta 4$  四聚体能够结合氧,但不会将其释放至外周,从而造成所谓的 HbH 疾病。患有 HbH 疾病的个体具有半衰期缩短的 RBC 并且其容易经历溶血,从而导致贫血增加。失去全部四个  $\alpha$  珠蛋白基因在子宫内通常是致命的。

[0016] 因此,仍需要可用于基因组编辑的另外的方法和组合物,以校正异常基因或改变其它基因的表达,例如以治疗血红蛋白病例如镰状细胞疾病和地中海贫血。

[0017] 发明概述

[0018] 本文公开用于改变表达或用于校正一种或多种编码在遗传性疾病例如镰状细胞疾病或地中海贫血中所涉及的蛋白的基因(例如,产生疾病中缺乏、不足或异常的蛋白和/或调节这些蛋白的蛋白)的方法和组合物。这些蛋白的改变可以导致这些遗传性疾病的治疗。特别是,使用基因组编辑来校正异常基因,插入野生型基因或改变内源基因的表达。作为非限制性实例,可将编码 $\beta$ 珠蛋白的野生型基因插入细胞中以产生由有缺陷的 $\beta$ 珠蛋白造成的血红蛋白病中所缺乏的蛋白和/或治疗由有缺陷的 $\beta$ 珠蛋白造成的血红蛋白病。在一些情况下,可将野生型基因插入安全港基因座中或已知在目标组织中高度表达的基因座例如红系细胞中的 $\beta$ 珠蛋白基因座。基因组编辑可类似地通过将野生型 $\alpha$ 珠蛋白基因插入安全港中而用于产生 $\alpha$ 地中海贫血中缺乏的蛋白(并从而治疗 $\alpha$ 地中海贫血)。另一种方法涉及使用基因校正,其中靶向有缺陷的内源 $\alpha$ 或 $\beta$ 珠蛋白基因并且置换突变序列。可选地,可改变或敲除 $\gamma$ 珠蛋白的抑制中所涉及的调节基因(例如,以通过灭活和/或降低抑制蛋白的量来增加 $\gamma$ 珠蛋白的表达)和/或可改变 $\gamma$ 珠蛋白基因上游或 $\beta$ 珠蛋白基因座的其它区域中的调节结合位点,以使得调节子不能在 $\gamma$ 珠蛋白基因座适当地相互作用并且产生HbF,从而消除由异常 $\beta$ 珠蛋白基因造成的作用(即SCD或 $\beta$ 地中海贫血)。一种方法进一步涉及使用干细胞(例如,造血干细胞或RBC前体)的修饰,该干细胞然后可用于植入患者中,以治疗血红蛋白病。

[0019] 在一个方面,本文描述结合于基因组中目标区域(例如, $\beta$ 珠蛋白、 $\alpha$ 珠蛋白或安全港基因,或调节基因或其DNA靶标例如BCL11A、 $\gamma$ 珠蛋白或KLF1)中的靶位点的锌指蛋白(ZFP),其中该ZFP包含一个或多个工程改造的锌指结合结构域。在一个实施方案中,ZFP是裂解目标靶基因组区域的锌指核酸酶(ZFN),其中该ZFN包含一个或多个工程改造的锌指结合结构域和核酸酶裂解结构域或裂解半结构域。裂解结构域和裂解半结构域可以获自例如各种限制核酸内切酶和/或归巢核酸内切酶。在一个实施方案中,裂解半结构域是源自IIS型限制核酸内切酶(例如,Fok I)。在某些实施方案中,锌指结构域识别珠蛋白或安全港基因中的靶位点。在某些实施方案中,锌指结构域包含5或6个锌指结构域并识别珠蛋白基因中的靶位点(例如,具有5或6个含识别螺旋区域的指的锌指蛋白示于表1A中)。在另一个实施方案中,锌指结构域识别BCL11A、KLF1、 $\alpha$ 、 $\beta$ 或 $\gamma$ 珠蛋白基因或其调节元件中的靶位点。在某些实施方案中,锌指结构域包含5或6个锌指结构域并识别BCL11A、KLF1、 $\alpha$ 、 $\beta$ 或 $\gamma$ 珠蛋白基因中或其调节元件中的靶位点(例如,具有5或6个含识别螺旋区域的指的锌指蛋白示于表1A中)。

[0020] 在另一个方面,本文描述结合于基因组中目标区域(例如, $\alpha$ 或 $\beta$ 珠蛋白或安全港基因,或调节基因或其DNA靶标例如BCL11A、 $\gamma$ 珠蛋白或KLF1)中的靶位点的TALE蛋白(类转录激活因子),其中该TALE包含一个或多个工程改造的TALE结合结构域。在一个实施方案中,TALE是裂解目标靶基因组区域的核酸酶(TALEN),其中该TALEN包含一个或多个工程改造的TALE DNA结合结构域和核酸酶裂解结构域或裂解半结构域。裂解结构域和裂解半结构域可以获自例如各种限制核酸内切酶和/或归巢核酸内切酶。在一个实施方案中,裂解半结构域是源自IIS型限制核酸内切酶(例如,Fok I)。在某些实施方案中,TALE DNA

结合结构域识别珠蛋白或安全港基因中的靶位点。在其它实施方案中,TALE DNA 结合结构域识别 BCL11A、KLF1、 $\alpha$ 、 $\beta$  或  $\gamma$  珠蛋白基因或其调节元件(例如,表 3 中所示例的 TALEN 蛋白)中的靶位点。

[0021] 在另一个方面,本文描述结合于基因组中目标区域(例如,高度表达的基因、疾病相关基因或安全港基因)中的靶位点的 CRISPR/Cas 系统,其中该 CRISPR/Cas 系统包含 CRISPR/Cas 核酸酶和工程改造的 crRNA/tracrRNA(或单导向 RNA)。在某些实施方案中,CRISPR/Cas 系统识别高度表达的基因、疾病相关基因或安全港基因中的靶位点。在某些实施方案中,CRISPR/Cas 系统识别珠蛋白、白蛋白、CCR5、CXCR4、AAVS1、Rosa 或 HPRT 基因中的靶标。

[0022] 如本文所述的 ZFN、TALEN 和 / 或 CRISPR/Cas 系统可结合和 / 或裂解基因内或基因相邻的编码或非编码区域中的目标区域,例如前导序列、尾随序列或内含子,或非转录区域内,在编码区域上游或下游的目标区域。在某些实施方案中,ZFN、TALEN 和 / 或 CRISPR/Cas 系统结合和 / 或裂解珠蛋白基因。在其它实施方案中,ZFN、TALEN 和 / 或 CRISPR/Cas 系统结合和 / 或裂解安全港基因,例如 CCR5 基因、CXCR4 基因、PPP1R12C(也称为 AAVS1) 基因、白蛋白基因或 Rosa 基因。参见例如美国专利公布号 20080299580 ;20080159996 ;201000218264 ;20110301073 ;20130177983 和 20130177960 以及美国临时申请号 61/823,689。另外,为了辅助选择,可使用 HPRT 基因座(参见美国专利公布号 20130122591)。在另一个方面,本文描述包含一种或多种如本文所述的锌指和 / 或 TALE 核酸酶或 CRISPR/Cas 系统的组合物。在一些实施方案中,ZFN、TALEN 和 / 或 CRISPR/Cas 系统结合并裂解 BCL11A、KLF1、 $\alpha$ 、 $\beta$  或  $\gamma$  珠蛋白基因或裂解其调节元件。在另一个方面,本文描述包含一种或多种如本文所述的锌指、TALE 或 Cas 核酸酶的组合物。

[0023] 在另一个方面,本文描述编码一种或多种如本文所述的 ZFN、TALEN 和 / 或 CRISPR/Cas 系统的多核苷酸。多核苷酸可能是例如 mRNA。在一些方面,mRNA 可被化学修饰(参见例如 Kormann 等,(2011)Nature Biotechnology 29(2):154-157)。

[0024] 在另一个方面,本文描述包含可操作地连接至启动子的编码一种或多种本文所述的 ZFN、TALEN 和 / 或 CRISPR/Cas 系统的多核苷酸的 ZFN、TALEN 和 / 或 CRISPR/Cas 系统表达载体。在一个实施方案中,表达载体是病毒载体。

[0025] 在一个方面,本文描述用于裂解靶 DNA 的 ZFN、TALEN 和 / 或 CRISPR/Cas 系统蛋白。

[0026] 在其它方面,遗传修饰的 RBC 前体(造血干细胞,称为“HSC”)提供于骨髓移植中并且 RBC 在体内分化并成熟。在一些实施方案中,在 G-CSF 诱导的动员后分离 HSC,并且在其它情况下,细胞是从人类骨髓或脐带中分离。在一些方面,通过经设计以敲除珠蛋白表达调节子(例如,BCL11A 或 KLF1)的核酸酶处理来编辑 HSC。在其它方面,用工程改造的核酸酶和供体核酸修饰 HSC,以使得插入并表达野生型基因(例如,珠蛋白基因)和 / 或校正内源异常基因。在一些情况下,用于插入的野生型基因序列编码野生型  $\beta$  珠蛋白或野生型  $\alpha$  珠蛋白。在其它情况下,内源异常基因是  $\beta$  珠蛋白或  $\alpha$  珠蛋白基因。在一些实施方案中,在轻度骨髓清除性预调节后,将修饰的 HSC 施用至患者。在其它方面,在完全骨髓清除后施用 HSC,以使得在植入后,100%的造血细胞是源自修饰的 HSC。

[0027] 在另一个方面,本文描述用于裂解 RBC 前体细胞中的内源基因(例如,其失活导致

$\gamma$  珠蛋白表达增加的基因,例如 BCL11A 或 KLF1) 的方法,该方法包括:向细胞中引入一种或多种编码一种或多种在一定条件下结合于一种或多种内源基因中的靶位点的 ZFN、TALEN 和 / 或 CRISPR/Cas 系统的多核苷酸,以使得表达所述 ZFN、TALEN 和 / 或 CRISPR/Cas 系统并且裂解所述一种或多种基因。在另一个方面,本文描述用于裂解细胞中的 BCL11A 或 KLF1 基因的方法,该方法包括:向细胞中引入一种或多种编码一种或多种在一定条件下结合于一种或多种 BCL11A 或 KLF1 基因中的靶位点的 ZFN、TALEN 和 / 或 CRISPR/Cas 系统的多核苷酸,以使得表达所述 ZFN、TALEN 和 / 或 CRISPR/Cas 系统并且裂解所述一种或多种 BCL11A 或 KLF1 基因。在某些实施方案中,锌指结构域包含 5 或 6 个锌指结构域并识别珠蛋白基因中的靶位点(例如,具有 5 至 6 个含识别螺旋区域的指的锌指蛋白示于表 1A 中)。在其它实施方案中,TALEN 识别  $\beta$  珠蛋白、 $\alpha$  珠蛋白、 $\gamma$  珠蛋白、KLF 或 BCL11A 序列中的靶位点(示例于表 3 中)。在其它实施方案中,CRISPR/Cas 系统识别  $\beta$  珠蛋白、 $\alpha$  珠蛋白、 $\gamma$  珠蛋白、KLF 或 BCL11A 序列中的靶位点,其中单导向 RNA 被工程改造以识别目标靶基因中的所需靶位点。裂解基因可能失活(敲除),例如敲除一种或多种基因,其产物可抑制基因(例如,珠蛋白基因)的表达,或破坏这些蛋白的 DNA 上的调节靶位点。在一些实施方案中,失活基因或其靶序列是参与抑制胎儿血红蛋白表达的那些。细胞(例如,干细胞)在分化时含有胎儿血红蛋白并且可提供至有需要的患者。在一些实施方案中,珠蛋白基因被敲除。例如,当  $\beta$  珠蛋白不良表达时, $\alpha$  珠蛋白基因可被敲除以恢复  $\alpha$  珠蛋白与  $\beta$  珠蛋白的比率,或者可伴随着野生型  $\beta$  珠蛋白的插入而敲除编码 HbS 的  $\beta$  珠蛋白基因。细胞(例如,干细胞)在分化时将含有 HbA 血红蛋白并且可提供至有需要的患者。

[0028] 在另一个方面,本文描述用于将序列插入细胞(例如干细胞)中的内源基因(例如, $\beta$  珠蛋白、 $\alpha$  珠蛋白和 / 或安全港基因)中的方法,该方法包括使用一种或多种核酸酶来裂解内源基因并且将序列插入裂解位点中。在某些实施方案中,例如使用如本文所述的 ZFN 或 TALEN 对或 CRISPR/Cas 系统(或编码所述 ZFN、TALEN 和 / 或 CRISPR/Cas 系统的载体)和在用 ZFN、TALEN 和 / 或 CRISPR/Cas 系统靶向裂解后插入基因中的“供体”序列(也称为“转基因”)置换任何靶基因中的基因组序列。供体序列可存在于 ZFN 或 TALEN 载体中,存在于独立的载体(例如,Ad、AAV 或 LV 载体)中,或者可选地,可使用不同的核酸递送机制引入细胞中。所述将供体核苷酸序列插入靶基因座(例如,珠蛋白基因、其它安全港基因等)导致转基因在靶基因座(例如珠蛋白)的基因控制元件的控制下表达。在一些实施方案中,转基因编码非编码 RNA(例如,shRNA)。在 RBC 成熟之前的转基因表达将产生含有目标非编码 RNA 的 RBC。

[0029] 在其它实施方案中,转基因包含功能蛋白,例如珠蛋白(例如,野生型  $\beta$  和 / 或野生型  $\gamma$ ) 蛋白。在一些实施方案中,将目标转基因插入内源基因(例如,珠蛋白基因),导致完整外源蛋白序列的表达并缺乏任何由内源基因编码的序列。在其它实施方案中,所表达的外源蛋白是融合蛋白并包含由转基因和珠蛋白基因(例如,来自内源靶基因座,或者可选地来自转基因上的珠蛋白编码序列)编码的氨基酸。在一些情况下,珠蛋白基因是  $\beta$  珠蛋白,而在其它情况下,珠蛋白基因是  $\alpha$  珠蛋白。在其它情况下,珠蛋白基因是  $\gamma$  珠蛋白基因。当存在时,内源珠蛋白序列可存在于外源蛋白的氨基(N)末端部分和 / 或外源蛋白的羧基(C)末端部分上。珠蛋白序列可包括全长野生型或突变型珠蛋白序列,或者可选地可包括部分珠蛋白编码序列。在一些实施方案中,珠蛋白-转基因融合体位于细胞内的

内源基因座处,而在其它实施方案中,将珠蛋白-转基因编码序列插入基因组内的安全港中。在一些方面,安全港选自 CCR5 基因、CXCR4 基因、PPP1R12C(也称为 AAVS1) 基因、白蛋白基因或 Rosa 基因。参见例如美国专利公布号 20080299580 ;20080159996 ;201000218264 ;20110301073 ;20130177983 和 20130177960 以及美国临时申请号 61/823,689。另外,为辅助选择,可使用 HPRT 基因座(参见美国专利公布号 20130122591)。

[0030] 在另一个方面,本文提供细胞系和/或转基因动物模型(系统)。在一些实施方案中,转基因细胞和/或动物包括编码人类基因的转基因。在一些情况下,转基因动物包含在对应于外源转基因的内源基因座处的敲除(例如,敲除小鼠珠蛋白基因并且将人类珠蛋白基因插入小鼠中),从而使得开发出其中可孤立地研究人类蛋白的体内系统。这些转基因模型可用于筛选目的以鉴定可与目标人类蛋白相互作用或修饰目标人类蛋白的小分子或大的生物分子或其它实体。在一些方面,将转基因整合至通过任何本文所述方法获得的干细胞(例如,胚胎干细胞、诱导多能性干细胞、造血干细胞等)或动物胚胎中的所选基因座(例如,珠蛋白或安全港)中,然后植入胚胎以使得活体动物出生。在其它方面,干细胞含有内源基因座例如 BCL11A、KLF1 或  $\gamma$  珠蛋白基因或其组合处的基因组变化,以使得  $\gamma$  珠蛋白表达升高。在一些实施方案中,相比于未编辑的干细胞, $\gamma$  珠蛋白表达的升高改变了细胞中  $\gamma$  珠蛋白与  $\beta$  珠蛋白的比率。然后将动物饲养至性成熟并使其产生后代,其中至少一些后代包含编辑的内源基因序列或整合的转基因。

[0031] 在另一个方面,本文提供一种用于将核酸序列位点特异性整合至染色体的内源基因座(例如,珠蛋白或安全港基因)中、例如胚胎的染色体中的方法。在某些实施方案中,所述方法包括:(a) 向胚胎中注入(i)至少一种 DNA 载体,其中该 DNA 载体包含侧接待整合的核酸序列的上游序列和下游序列,和(ii)至少一种编码锌指、TALE 或 Cas9 核酸酶的 RNA 分子。在使用 Cas9 蛋白的情况下,还引入了工程改造的 sgRNA。核酸酶或核酸酶系统识别靶基因座(例如,珠蛋白或安全港基因座)中的靶位点,然后(b)培养胚胎以使其表达锌指或 TALE 核酸酶和/或 CRISPR/Cas 系统,其中通过锌指核酸酶、TALEN 或 CRISPR/Cas 系统将双链断裂引入靶标中,然后经由与 DNA 载体同源重组来修复,以将核酸序列整合至染色体中。

[0032] 在任何本文所述的方法中,编码锌指核酸酶、TALEN 和/或 CRISPR/Cas 系统的多核苷酸可包含 DNA、RNA 或其组合。在某些实施方案中,多核苷酸包含质粒。在其它实施方案中,编码核酸酶的多核苷酸包含 mRNA。

[0033] 此外提供了包含本发明的 ZFP、TALEN 和/或 CRISPR/Cas 系统的试剂盒。该试剂盒可包含编码 ZFP、TALEN 或 CRISPR/Cas 系统的核酸(例如合适的表达载体中所含的 RNA 分子或编码 ZFP、TALEN 或 Cas9 的基因)和工程改造的 sg RNA(如果需要的话),或核酸酶蛋白、供体分子、合适的宿主细胞系的等分试样、用于实施本发明方法的说明书等。

[0034] 鉴于整个公开内容,这些和其它方面对于本领域技术人员将是显而易见的。

[0035] 附图简述

[0036] 图 1 描绘镰状细胞疾病突变周围区域中的  $\beta$  样珠蛋白基因序列的比对(示于图上)。示出了(顶线至底线)具有镰状细胞突变的血红蛋白  $\beta$  序列(HBB-镰状,SEQ ID NO:1);血红蛋白  $\beta$  (HBB,SEQ ID NO:2);血红蛋白  $\delta$  (HBD,SEQ ID NO:3); $\beta$  血红蛋白假基因(HBBP1,SEQ ID NO:4);血红蛋白  $\epsilon$  (HBE1,SEQ ID NO:5)、血红蛋白  $\gamma 1$  (HBG1,SEQ ID NO:6) 和血红蛋白  $\gamma 2$  (HBG2,SEQ ID NO:7)。下面示出关于所示五个 ZFN 对的比对的 Cel1

I 活性分析结果。

[0037] 图 2 的 A 图和 B 图是描绘由 CD34<sup>+</sup> 细胞中的  $\beta$  珠蛋白供体指定序列的插入的凝胶。图 2A (RFLP) 描绘由  $\beta$  珠蛋白供体指定序列的插入, 其中插入通过供体 DNA 上存在的新限制位点的存在来证实。图 2B 描绘 CeI-1 错配测定法 (Surveyor<sup>TM</sup>, Transgenomic) 的结果, 显示产生错配的新序列的存在。带有突变的等位基因的百分比 (% NHEJ) 示于凝胶右侧的文字中 (第一栏对应于凝胶上的泳道编号)。数字是指 ZFN 组合; unt: 未转染对照物。

[0038] 图 3 是描绘 KLF1 和 BCL11A 在  $\beta$  和  $\gamma$  珠蛋白基因表达的调节中所起作用的图形。KLF1 的表达刺激 BCL11a 和  $\beta$  珠蛋白基因的表达。BCL11A 蛋白抑制  $\gamma$  珠蛋白表达。

[0039] 图 4 的 A 图至 E 图描绘示出在用所示的 BCL11A 特异性 ZFN (图 4A)、KLF1 特异性 ZFN (图 4C 和图 4D) 或 HPRT 特异性 ZFN (图 4B) 处理 HSC 后以及用短暂低温冲击 (30°) 或在标准条件 (37°) 下处理转导细胞后, 如上文所述的 CeI-1 测定法的结果的凝胶。在转染后的 3 天收获 DNA。图 4E 描绘利用在 HSC 转染后的 3 天或在 17 天的红系分化后收获的样品进行的相同类型的 CeI-1 分析。带有突变的等位基因的百分比 (% NHEJ) 示于泳道底部, 并且所用的 ZFN 对的身份示于各图中。

[0040] 图 5 的 A 图和 B 图描绘在通过 Taqman<sup>®</sup> 程序分析的分化后的 7 天或 17 天, 相比于  $\beta$  珠蛋白的  $\gamma$  珠蛋白 (图 5A) 或用 18s RNA 标准物校正的  $\gamma$  珠蛋白 mRNA (图 5B) 的表达。 $\gamma$  珠蛋白 mRNA 相比于  $\gamma + \beta$  珠蛋白 mRNA 的百分比示于图 5A 中各条柱上方。在图 5B 中, 如通过 18s RNA 所归一化的  $\gamma$  珠蛋白的相对水平示于条柱上方, 并且显示在已经用 BCL11A 特异性 ZFN 处理的细胞中, 关于 18S 的  $\gamma$  珠蛋白 mRNA 的水平较高。

[0041] 图 6 描绘取决于细胞的基因型, 来源于 HSC 的甲基纤维素集落中的  $\gamma$  珠蛋白 mRNA 的量。相比于具有单一 BCL11A 敲除等位基因的细胞 (“Bb”) 或已敲除 (“敲除”) 两种等位基因的细胞, 其中两种 BCL11A 基因都是野生型的细胞 (“BB”) 产生最低量的  $\gamma$  珠蛋白 mRNA。条柱上方的数字指示从总  $\beta$  珠蛋白产生的  $\gamma$  珠蛋白的百分比。

[0042] 图 7 示出在用 K562 细胞中的  $\gamma$  珠蛋白特异性 ZFN 处理后,  $\gamma$  珠蛋白基因上游区域的一系列 DNA 序列 (SEQ ID NO:140 至 148)。所述序列具有多个插入和缺失, 包括 13bp 缺失 (“ $\Delta$  13bp”), 其与和 HPFH 相关的人类基因型之一同一。顶部的 ‘参考’ 序列是  $\gamma$  珠蛋白的野生型 5' 调节区的序列。ZFN 对的结合位点以红色强调, 天然存在的 13bp 缺失以下划线示出。

[0043] 图 8 的 A 图至 C 图描绘来源于用靶向  $\gamma$  珠蛋白启动子的 ZFN 处理且接着涂在甲基纤维素集落上的 HSC 的红系集落的 Taqman 分析。集落编号指示在各条柱底部, 基因型也是如此。图 8A 示出相对的  $\gamma / \beta$  珠蛋白 mRNA 比率; 图 8B 示出由 18s RNA 水平校正的  $\gamma$  珠蛋白 mRNA 水平并且图 8C 示出由 18s 校正的  $\beta$  珠蛋白水平的相应分析。野生型和突变型集落的比率的平均值的比较指示  $\gamma$  珠蛋白启动子中具有 ZFN 诱导突变的集落中的  $\gamma$  珠蛋白水平升高。

[0044] 图 9 示出  $\gamma$  珠蛋白基因的启动子区域 (SEQ ID NO:149-152)。比对两个  $\gamma$  珠蛋白等位基因 (HBG1 和 HBG2)。用灰色方框指示两个等位基因的序列中的差异。另外, 用黑色轮廓指示与 HPFH 相关的突变。起始 ATG 指示为外显子 1 边界。与每一突变相关的胎儿珠蛋白水平的增加由其上方的数字指示。

[0045] 图 10 的 A 图和 B 图描绘使用所示锌指核酸酶和寡核苷酸供体在 CD34<sup>+</sup> 细胞中的

NHEJ 量（即，由核酸酶靶向断裂的基于 NHEJ 的修复事件产生的靶向基因座破坏）和针对  $\beta$  珠蛋白基因检测的基因校正。

[0046] 图 11 描绘 CD34+ 细胞中的 NHEJ 量和供体核苷酸的靶向整合，其中针对供体的同源臂改变。

[0047] 图 12 描绘已经用 ZFN 和寡核苷酸供体处理的干细胞的红系衍生物中的基因编辑的持续存在。在由分化产生的四种类型的细胞群体，即红系集落形成单位（“CFU-E”）、红系爆发集落形成单位（“BFU-E”）、粒细胞 / 巨噬细胞集落形成单位（“CFU-GM”）和粒细胞 / 红细胞 / 单核细胞 / 巨噬细胞集落形成单位（“CFU-GEMM”）中分析基因修饰。

[0048] 图 13 描绘  $\beta$  珠蛋白基因的基因修饰随时间的稳定性。

[0049] 发明详述

[0050] 本文公开用于研究和治疗遗传性疾病例如血红蛋白病的方法和组合物。本发明描述靶细胞的基因组编辑以使得在一种或多种珠蛋白基因的表达中存在有利的变化，其又导致治疗有需要的受试者中的血红蛋白病例如镰状细胞疾病或地中海贫血。珠蛋白基因表达中的有利变化包括（但不限于）在具有异常  $\beta$  珠蛋白的受试者中提供  $\gamma$  珠蛋白基因；和 / 或校正异常的  $\alpha$  或  $\beta$  珠蛋白基因序列。另外，在被改变以表达所需蛋白产物的移植植物中递送改变的造血干细胞可类似地有利于治疗血红蛋白病例如镰状细胞贫血或地中海贫血。此外描述了具有改变的珠蛋白表达的细胞系和动物。

[0051] 因此，本发明的方法和组合物可用于改变一种或多种珠蛋白基因（例如， $\gamma$ 、 $\alpha$  和 / 或  $\beta$ ）在细胞（例如，红系前体细胞）中的表达。这些方法和组合物可用于破坏  $\gamma$  珠蛋白抑制中所涉及的基因（例如，BCL11A 或 KLF1），以使得在编辑后，细胞将以较高水平表达  $\gamma$  珠蛋白，并且可以产生 HbF。可选地，编辑可用于破坏基因上的结合位点（例如，破坏  $\beta$  珠蛋白基因座中的 BCL11A 结合、破坏在  $\gamma$  珠蛋白启动子处  $\gamma$  珠蛋白转录的阻遏子的结合）以禁止基因的抑制。可选地或除这些改变之外，所述方法和组合物可用于校正异常的内源  $\alpha$  和 / 或  $\beta$  珠蛋白基因或将野生型基因插入细胞基因组中的所需位置（例如，插入 HSC 中）。前体细胞可源自有需要的受试者，经离体修饰，然后于骨髓移植植物中置回受试者。

[0052] 概况

[0053] 除非另外说明，否则本文公开的方法的实践以及组合物的制备和利用使用分子生物学、生物化学、染色质结构和分析、计算化学、细胞培养、重组 DNA 和如本领域技术范围的相关领域的常规技术。这些技术在文献中得到充分解释。参见例如 Sambrook 等，MOLECULAR CLONING: A LABORATORY MANUAL, 第二版, Cold Spring Harbor Laboratory Press, 1989, 和第三版, 2001; Ausubel 等, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, 1987 和定期更新; 丛书 METHODS IN ENZYMOLOGY, Academic Press, San Diego; Wolffe, CHROMATIN STRUCTURE AND FUNCTION, 第三版, Academic Press, San Diego, 1998; METHODS IN ENZYMOLOGY, 第 304 卷, “Chromatin” (P. M. Wassarman 和 A. P. Wolffe 编), Academic Press, San Diego, 1999; 以及 METHODS IN MOLECULAR BIOLOGY, 第 119 卷, “Chromatin Protocols” (P. B. Becker 编) Humana Press, Totowa, 1999。

[0054] 定义

[0055] 术语“核酸”、“多核苷酸”和“寡核苷酸”可互换使用并且是指呈线性或环状构象以及呈单链或双链形式的脱氧核糖核苷酸或核糖核苷酸聚合物。对于本公开的目的，这些术

语不应被理解为在聚合物长度方面具有限制。所述术语可以涵盖天然核苷酸以及在碱基、糖和 / 或磷酸酯部分 (例如, 硫代磷酸酯主链) 中被修饰的核苷酸的已知类似物。一般来说, 特定核苷酸的类似物具有相同的碱基配对特异性; 即, A 的类似物将与 T 进行碱基配对。

[0056] 术语“多肽”、“肽”和“蛋白”可互换使用以指代氨基酸残基的聚合物。该术语还适用于其中一个或多个氨基酸是相应的天然存在的氨基酸的化学类似物或修饰衍生物的氨基酸聚合物。

[0057] “结合”是指大分子之间 (例如, 蛋白与核酸之间) 的序列特异性、非共价相互作用。结合相互作用的组分并非全部都需要具序列特异性 (例如, 接触 DNA 主链中的磷酸酯残基), 只要相互作用整体具序列特异性即可。这些相互作用的一般特征在于, 解离常数 ( $K_d$ ) 为  $10^{-6}M^{-1}$  或更低。“亲和力”是指结合强度: 结合亲和力增加与较低的  $K_d$  相关。

[0058] “结合蛋白”是能够非共价结合至另一分子的蛋白。结合蛋白可结合于例如 DNA 分子 (DNA- 结合蛋白)、RNA 分子 (RNA- 结合蛋白) 和 / 或蛋白分子 (蛋白- 结合蛋白)。在蛋白- 结合蛋白的情况下, 其可结合于自身 (以形成同型二聚体、同型三聚体等) 和 / 或其可结合于一种或多种不同蛋白的一个或多个分子。结合蛋白可具有一种以上类型的结合活性。例如, 锌指蛋白具有 DNA 结合、RNA 结合和蛋白结合活性。

[0059] “锌指 DNA 结合蛋白”(或结合结构域) 是蛋白, 或较大蛋白内的结构域, 其以序列特异性方式通过一个或多个锌指结合 DNA, 所述锌指是通过锌离子的配位稳定化的结构的结合结构域内的氨基酸序列的区域。术语锌指 DNA 结合蛋白通常缩写为锌指蛋白或 ZFP。

[0060] “TALE DNA 结合结构域”或“TALE”是包含一个或多个 TALE 重复结构域 / 单元的多肽。重复结构域参与 TALE 与其同源靶 DNA 序列的结合。单一“重复单元”(也称为“重复序列”) 通常长度为 33-35 个氨基酸并展现与天然存在的 TALE 蛋白内的其它 TALE 重复序列的至少一些序列同源性。参见例如美国专利公布号 20110301073。

[0061] 锌指和 TALE 结合结构域可以“工程改造”以结合于预定核苷酸序列, 例如经由天然存在的锌指或 TALE 蛋白的识别螺旋区域的工程改造 (改变一个或多个氨基酸)。因此, 工程改造的 DNA 结合蛋白 (锌指或 TALE) 是非天然存在的蛋白。用于工程改造 DNA- 结合蛋白的方法的非限制性实例是设计和选择。所设计的 DNA 结合蛋白是并非天然存在的蛋白, 其设计 / 组成主要源于合理的设计标准。合理的设计标准包括应用替换规则和用于处理储存现有的 ZFP 和 / 或 TALE 设计和结合数据的信息的数据库中的信息的计算化算法。参见例如美国专利 6, 140, 081 ; 6, 453, 242 ; 和 6, 534, 261 ; 还参见 WO 98/53058 ; WO 98/53059 ; WO 98/53060 ; WO 02/016536 和 WO 03/016496 以及美国公布号 20110301073。

[0062] “所选的”锌指蛋白或 TALE 是自然界中不存在的蛋白, 其产生主要是由于实验过程例如噬菌体展示、相互作用捕集或杂交选择。参见例如 US 5, 789, 538 ; US 5, 925, 523 ; US 6, 007, 988 ; US 6, 013, 453 ; US 6, 200, 759 ; WO 95/19431 ; WO 96/06166 ; WO 98/53057 ; WO 98/54311 ; WO 00/27878 ; WO 01/60970 ; WO 01/88197 ; WO 02/099084 以及美国公布号 20110301073。

[0063] “重组”是指两种多核苷酸之间的遗传信息的交换过程。为本公开的目的, “同源重组”(HR) 是指例如在经由同源介导的修复机制在细胞中的双链断裂修复期间发生的这种交换的特殊形式。这个过程需要核苷酸序列同源性, 使用“供体”分子来模板化“靶标”分子 (即, 经历双链断裂者) 的修复, 并且被不同地称为“非交叉基因转换”或“短道基因转换”,



因为其导致遗传信息从供体转移至靶标。不希望受任何特定理论束缚,这种转移可涉及在断裂的靶标和供体之间形成的异源双链 DNA 的错配校正,和/或“合成依赖性链粘接”,其中使用供体来再次合成将变成靶标和/或相关过程的一部分的遗传信息。这种特殊化 HR 通常导致靶分子的序列改变,以使得供体多核苷酸序列的一部分或全部被并入靶多核苷酸中。

[0064] 在本公开的方法中,一种或多种如本文所述的靶向核酸酶在靶序列(例如,细胞染色质)中预定位点处产生双链断裂,并且与断裂区域中的核苷酸序列具有同源性的“供体”多核苷酸可被并入细胞中。双链断裂的存在已证实促进供体序列的整合。供体序列可被物理整合,或者可选地,使用供体多核苷酸作为模板以经由同源重组修复断裂,从而导致如供体中的全部或部分核苷酸序列引入细胞染色质中。因此,细胞染色质中的第一序列可改变,并且在某些实施方案中,可转化成供体多核苷酸中存在的序列。因此,术语“置换(replace/replacement)”的使用可理解为表示一个核苷酸序列被另一个置换(即,在信息意义上的序列置换),并且未必需要一个多核苷酸被另一个多核苷酸物理或化学置换。

[0065] 在任何本文所述的方法中,其它的锌指或 TALEN 蛋白可用于对细胞内其它靶位点的另外的双链裂解。另外,可类似地使用 CRISPR/Cas 系统来诱导其它双链断裂。

[0066] 在用于细胞染色质中目标区域中的序列的靶向重组和/或置换和/或改变的方法的某些实施方案中,通过与外源“供体”核苷酸序列同源重组来改变染色体序列。如果存在与断裂区域同源的序列,则通过细胞染色质中双链断裂的存在来刺激这种同源重组。

[0067] 在任何本文所述的方法中,外源核苷酸序列(“供体序列”或“转基因”)可含有与目标区域中的基因组序列同源、但不同一的序列,从而刺激同源重组以将非同源序列插入目标区域中。因此,在某些实施方案中,与目标区域中的序列同源的供体序列的一部分展现与被置换的基因组序列的约 80% 至 99% (或其间的任何整数)序列同一性。在其它实施方案中,例如如果供体与具有 100 个以上连续碱基对的基因组序列之间仅相差 1 个核苷酸,则供体与基因组序列之间的同源性高于 99%。在某些情况下,供体序列的非同源部分可能含有不存在于目标区域中的序列,以使得新序列被引入目标区域中。在这些情况下,非同源序列一般侧接具有 50-1,000 个碱基对(或介于其间的任何整数值)或大于 1,000 的任何数目的碱基对的序列,其与目标区域中的序列同源或同一。在其它实施方案中,供体序列与第一序列非同源,并且通过非同源重组机制插入基因组中。

[0068] 任何本文所述的方法可用于通过破坏目标基因的表达的供体序列的靶向整合来部分或完全灭活细胞中一个或多个靶序列。还提供了具有部分或完全灭活的基因的细胞系。

[0069] 此外,如本文所述的靶向整合的方法还可用于整合一种或多种外源序列。外源核酸序列可包含例如一种或多种基因或 cDNA 分子,或任何类型的编码或非编码序列,以及一种或多种控制元件(例如,启动子)。另外,外源核酸序列可产生一种或多种 RNA 分子(例如,小发夹 RNA(shRNA)、抑制 RNA(RNAi)、微 RNA(miRNA)等)。

[0070] “裂解”是指 DNA 分子的共价主链的断裂。裂解可通过多种方法起始,包括(但不限于)磷酸二酯键的酶促或化学水解。单链裂解和双链裂解都是可能的,并且双链裂解可能由于两种不同的单链裂解事件而发生。DNA 裂解可导致产生钝端或交错端。在某些实施方案中,将融合多肽用于靶向双链 DNA 裂解。

[0071] “裂解半结构域”是一种多肽序列,其结合第二多肽(同一或不同)形成具有裂解

活性（优选双链裂解活性）的复合物。术语“第一和第二裂解半结构域”、“+ 和 - 裂解半结构域”和“右和左裂解半结构域”可互换使用以指代二聚合的裂解半结构域对。

[0072] “工程改造的裂解半结构域”是已经被修饰以与另一个裂解半结构域（例如，另一个工程改造的裂解半结构域）形成专性异二聚体的裂解半结构域。也参见美国专利公布号 2005/0064474、2007/0218528、2008/0131962 和 2011/0201055，其以全文引用的方式并入本文中。

[0073] 术语“序列”是指任何长度的核苷酸序列，其可以是 DNA 或 RNA；可以是线性、环状或分枝的并且可以是单链或双链。术语“供体序列”是指插入基因组中的核苷酸序列。供体序列可以具有任何长度，例如长度为 2 至 10,000 个核苷酸（或其间或其上的任何整数值），优选长度为约 100 至 1,000 个核苷酸（或其间的任何整数），更优选长度为约 200 至 500 个核苷酸。

[0074] “疾病相关基因”是在单基因疾病中在一些方式下有缺陷的基因。单基因疾病的非限制性实例包括重度联合免疫缺陷、囊性纤维化、溶酶体贮存疾病（例如戈谢病 (Gaucher's)、赫尔勒病 (Hurler's)、亨特病 (Hunter's)、法布里病 (Fabry's)、尼曼 - 匹克病 (Neimann-Pick)、泰萨二氏病 (Tay-Sach's) 等)、镰状细胞贫血和地中海贫血。

[0075] “染色质”是包含细胞基因组的核蛋白结构。细胞染色质包含核酸，主要是 DNA；和蛋白，包括组蛋白和非组蛋白染色体蛋白。大多数真核细胞染色质以核小体形式存在，其中核小体核心包含与包含各两个的组蛋白 H2A、H2B、H3 和 H4 的八聚体相关的约 150 个碱基对的 DNA；并且接头 DNA（取决于生物体，具有可变长度）在核小体核心之间延伸。组蛋白 H1 的分子一般与接头 DNA 相关。为本公开的目的，术语“染色质”意在涵盖所有类型的原核和真核细胞核蛋白。细胞染色质包括染色体和游离体染色质。

[0076] “染色体”是包含细胞基因组的全部或一部分的染色质复合物。细胞的基因组通常特征在于其核型，它是包含细胞基因组的所有染色体的集合。细胞基因组可包含一个或多个染色体。

[0077] “游离体”是复制核酸、核蛋白复合物或包含并非细胞的染色体核型的一部分的核酸的其它结构。游离体的实例包括质粒和某些病毒基因组。

[0078] “靶位点”或“靶序列”是定义结合分子将结合的核酸的一部分的核酸序列，提供存在结合的充分条件。

[0079] “外源”分子是通常不存在于细胞中、但可以通过一种或多种遗传、生物化学或其它方法引入细胞中的分子。“正常存在于细胞中”是在细胞的特定发育阶段和环境条件方面测定。因此，例如，仅在肌肉的胚胎发育期间存在的分子是关于成人肌肉细胞的外源分子。类似地，由热休克诱导的分子是关于非热休克细胞的外源分子。外源分子可以包含例如不正常工作的内源分子的正常工作型式或正常工作的内源分子的不正常工作型式。

[0080] 其中，外源分子可以是小分子，例如通过组合化学方法产生；或者大分子例如蛋白、核酸、碳水化合物、脂质、糖蛋白、脂蛋白、多糖、上述分子的任何修饰衍生物，或任何包含一种或多种上述分子的复合物。核酸包括 DNA 和 RNA，可以是单链或双链的；可以是线性、分枝或环状的；并且可以具有任何长度。核酸包括能够形成双链体的核酸，以及形成三链体的核酸。参见例如美国专利号 5,176,996 和 5,422,251。蛋白包括（但不限于）DNA-结合蛋白、转录因子、染色质重塑因子、甲基化 DNA 结合蛋白、聚合酶、甲基化酶、去甲基化酶、乙

酰化酶、去乙酰化酶、激酶、磷酸酶、整合酶、重组酶、连接酶、拓扑异构酶、促旋酶和解螺旋酶。

[0081] 外源分子可以是与内源分子相同类型的分子,例如,外源蛋白或核酸。例如,外源核酸可以包含感染的病毒基因组、引入细胞中的质粒或游离体、或通常不存在于细胞中的染色体。用于将外源分子引入细胞中的方法是本领域技术人员已知的并且包括(但不限于)脂质介导的转移(即,脂质体,包括中性和阳离子性脂质)、电穿孔、直接注射、细胞融合、粒子轰击、磷酸钙共沉淀、DEAE-葡聚糖介导的转移和病毒载体介导的转移。外源分子也可以是与内源分子相同类型的分子,但源自与细胞来源不同的物种。例如,可将人类核酸序列引入最初来源于小鼠或仓鼠的细胞系。

[0082] 相比之下,“内源”分子是通常在特定环境条件下在特定发育阶段存在于特定细胞中的分子。例如,内源核酸可以包含染色体、线粒体、叶绿体或其它细胞器的基因组,或天然存在的游离体核酸。其它的内源分子可包括蛋白,例如,转录因子和酶。

[0083] “融合”分子是其中优先共价地连接两个或更多个亚基分子的分子。亚基分子可以是相同化学类型的分子,或者可以是不同化学类型的分子。第一类型的融合分子的实例包括(但不限于)融合蛋白(例如,ZFP 或 TALE DNA 结合结构域与一个或多个活化结构域之间的融合体)和融合核酸(例如,编码上述融合蛋白的核酸)。第二类型的融合分子的实例包括(但不限于)形成三链体的核酸与多肽之间的融合体,和小沟结合物与核酸之间的融合体。

[0084] 融合蛋白在细胞中的表达可由融合蛋白向细胞的递送或通过编码融合蛋白的多核苷酸向细胞的递送产生,其中所述多核苷酸被转录,并且将转录物翻译,以生成融合蛋白。反式剪接、多肽裂解和多肽连接也可参与蛋白在细胞中的表达。用于将多核苷酸和多肽递送至细胞的方法呈现于本公开中其它地方。

[0085] 为本公开的目的,“基因”包括编码基因产物(参见上文)的 DNA 区域,以及调节基因产物的产生的所有 DNA 区域,无论这些调节序列是否与编码和/或转录序列相邻。因此,基因包括(但未必限于)启动子序列、终止子、翻译调节序列例如核糖体结合位点和内部核糖体进入位点、增强子、沉默子、绝缘子、边界元件、复制起点、基质连接位点和基因座控制区。

[0086] “基因表达”是指基因中所含的信息转化成基因产物。基因产物可以是基因的直接转录产物(例如,mRNA、tRNA、rRNA、反义 RNA、核糖酶、结构 RNA 或任何其它类型的 RNA)或通过 mRNA 的翻译产生的蛋白。基因产物还包括通过例如封端、多腺苷酸化、甲基化和编辑等方法修饰的 RNA,和通过例如甲基化、乙酰化、磷酸化、泛素化、ADP-核糖基化、十四烷基化和糖基化修饰的蛋白。

[0087] 基因表达的“调节”是指基因活性的变化。表达的调节可以包括(但不限于)基因活化和基因抑制。可以使用基因组编辑(例如,裂解、改变、失活、随机突变)来调节表达。基因失活是指与不包括如本文所述的 ZFP 或 TALEN 的细胞相比,基因表达的任何降低。因此,基因失活可以是部分的或完全的。

[0088] “目标区域”是细胞染色质的任何区域,例如,需要结合外源分子的基因或者基因内或相邻的非编码序列。结合可用于靶向 DNA 裂解和/或靶向重组的目的。目标区域可存在于例如染色体、游离体、细胞器基因组(例如,线粒体、叶绿体)或感染的病毒基因组中。

目标区域可在基因的编码区域内、转录的非编码区域内,例如,前导序列、尾随序列或内含子,或非转录区域内,在编码区域上游或下游。目标区域的长度可能小至单个核苷酸对或高达 2,000 个核苷酸对,或任何整数值的核苷酸对。

[0089] “真核”细胞包括(但不限于)真菌细胞(例如酵母)、植物细胞、动物细胞、哺乳动物细胞和人类细胞(例如,T 细胞)。

[0090] “红血细胞”(RBC)或红细胞是源自造血干细胞的终末分化细胞。它们缺乏核酸酶和大部分细胞器。RBC 含有血红蛋白以从肺部向外周组织输送氧气。实际上,个体 RBC 的 33%是血红蛋白。它们也携带在代谢期间由细胞产生的 CO<sub>2</sub> 离开组织并返回到肺部以在呼气期间释放。在骨髓中响应于血缺氧产生 RBC,其通过由肾脏释放促红细胞生成素(EPO)来介导。EPO 造成原成红细胞数目增加并缩短完全 RBC 成熟所需的时间。在约 120 天后,由于 RBC 不含细胞核或任何其它再生能力,因此通过肝脏、脾脏和淋巴结中巨噬细胞的吞噬活性(约 90%)或通过血浆中的溶血(约 10%)从循环中去除细胞。在巨噬细胞吞没后,由于溶酶体酶的作用,RBC 的化学组分在巨噬细胞的液泡内分解。

[0091] “分泌组织”是动物中从个体细胞中分泌产物进入通常源自上皮的一些类型的内腔中的那些组织。位于胃肠道的分泌组织的实例包括肠、胰腺和胆囊的内衬细胞。其它分泌组织包括肝脏、与眼睛和粘膜相关的组织例如唾液腺、乳腺、前列腺、垂体腺和内分泌系统的其它成员。另外,分泌组织包括能够分泌的组织类型的个体细胞。

[0092] 术语“操作性连接”和“操作性地连接”(或可操作地连接)可关于两种或更多种组分(例如序列元件)的并置互换使用,其中所述组分被布置以使得两种组分正常发挥作用并允许至少一种组分可以介导施加在至少一种其它组分上的功能的可能性。通过说明的方式,如果转录调节序列响应于一种或多种转录调节因子的存在或不存在来控制编码序列的转录水平,则转录调节序列例如启动子是操作性地连接至编码序列。转录调节序列一般操作性地与编码序列顺式连接,但无需与其直接相邻。例如,增强子是操作性地连接至编码序列的转录调节序列,即使它们是不连续的。

[0093] 关于融合多肽,术语“操作性地连接”可以指每一组分执行与其将不会如此连接的其他组分连接的相同功能。例如,关于其中 ZFP、TALE 或 Cas DNA 结合结构域与活化结构域融合的融合多肽,所述 ZFP、TALE 或 Cas DNA 结合结构域和所述结构域是操作性连接,如果在融合多肽中,则 ZFP、TALE 或 Cas DNA 结合结构域部分能够结合其靶位点和/或其结合位点,而活化结构域能够上调基因表达。当其中 ZFP 或 TALE DNA 结合结构域与裂解结构域融合的融合多肽时,所述 ZFP 或 TALE DNA 结合结构域和所述裂解结构域是操作性连接,如果在融合多肽中,则 ZFP 或 TALE DNA 结合结构域部分能够结合其靶位点和/或其结合位点,而裂解结构域能够裂解靶位点附近的 DNA。

[0094] 蛋白、多肽或核酸的“功能片段”是序列与全长蛋白、多肽或核酸不同一、但保持与全长蛋白、多肽或核酸相同的功能的蛋白、多肽或核酸。功能片段可以具有与相应的天然分子相比更多、更少或相同数目的残基,和/或可含有一个或多个氨基酸或核苷酸取代。用于测定核酸功能(例如,编码功能、与另一个核酸杂交的能力)的方法是本领域中公知的。类似地,用于测定蛋白功能的方法是公知的。例如,可例如通过过滤器结合、电泳迁移率变动或免疫沉淀测定法来测定多肽的 DNA 结合功能。可通过凝胶电泳来测定 DNA 裂解。参见 Ausubel 等,同上。可例如通过共免疫沉淀、双杂交测定法或遗传性和生物化学互补来测定

蛋白与另一种蛋白相互作用的能力。参见例如 Fields 等, (1989) Nature 340:245-246 ;美国专利号 5, 585, 245 和 PCT WO 98/44350。

[0095] “载体”能够转移基因序列至靶细胞。通常,“载体构建体”、“表达载体”和“基因转移载体”意指能够引导目标基因的表达并且可以转移基因序列至靶细胞的任何核酸构建体。因此,所述术语包括克隆和表达媒介物,以及整合载体。

[0096] “报告基因”或“报告序列”是指产生容易测量的蛋白产物、但优选未必用常规测定法的任何序列。合适的报告基因包括(但不限于)编码介导抗生素抗性(例如,氨苄青霉素抗性、新霉素抗性、G418 抗性、嘌呤霉素抗性)的序列、编码有色或荧光或发光蛋白(例如,绿色荧光蛋白、增强型绿色荧光蛋白、红色荧光蛋白、荧光素酶)的序列,和介导增强的细胞生长和/或基因扩增的蛋白(例如,二氢叶酸还原酶)。表位标签包括例如一个或多个拷贝的 FLAG、His、myc、Tap、HA 或任何可检测的氨基酸序列。“表达标签”包括编码可操作地连接至所需基因序列以监测目标基因的表达的报告基因的序列。

[0097] 术语“受试者”和“患者”可互换使用并且是指哺乳动物例如人类患者和非人类灵长类动物,以及实验动物例如兔、狗、猫、大鼠、小鼠和其它动物。因此,如本文所用的术语“受试者”或“患者”意指可施用本发明的改变的 RBC(或干细胞)的任何哺乳动物患者或受试者。本发明的受试者包括已经暴露于一种或多种化学毒素(包括例如神经毒素)的那些。

[0098] 核酸酶

[0099] 本文描述可用于靶向基因以用于血红蛋白病的组合物、特别是核酸酶。在某些实施方案中,核酸酶是天然存在的。在其它实施方案中,核酸酶是非天然存在的,即,在 DNA 结合结构域和/或裂解结构域中工程改造。例如,天然存在的核酸酶的 DNA 结合结构域可被改变以结合至所选靶位点(例如,已经工程改造以结合于不同于同源结合位点的位点的大范围核酸酶(meganuclease))。在其它实施方案中,核酸酶包含异源 DNA 结合和裂解结构域(例如,锌指核酸酶;TAL-效应子核酸酶;具有异源裂解结构域的大范围核酸酶 DNA 结合结构域),或由特异性导向 RNA 引导的通用核酸酶(例如 CRPISR/Cas)。

[0100] A. DNA 结合结构域

[0101] 在某些实施方案中,核酸酶是大范围核酸酶(归巢核酸内切酶)。天然存在的大范围核酸酶识别 15-40 个碱基对裂解位点并且通常分成四个家族:LAG LIDADG 家族、GIY-YIG 家族、His-Cyst box 家族和 HNH 家族。示例性归巢核酸内切酶包括 I-SceI、I-CeuI、PI-PspI、PI-Sce、I-SceIV、I-CsmI、I-PanI、I-SceII、I-PpoI、I-SceIII、I-CreI、I-TevI、I-TevII 和 I-TevIII。其识别序列是已知的。也参见美国专利号 5, 420, 032 ;美国专利号 6, 833, 252 ;Belfort 等, (1997) Nucleic Acids Res. 25:3379-3388 ;Dujon 等, (1989) Gene 82:115-118 ;Perler 等, (1994) Nucleic Acids Res. 22, 1125-1127 ;Jasin(1996) Trends Genet. 12:224-228 ;Gimble 等, (1996) J. Mol. Biol. 263:163-180 ;Argast 等, (1998) J. Mol. Biol. 280:345-353 以及新英格兰生物实验室目录 (New England Biolabs catalogue)。

[0102] 在某些实施方案中,核酸酶包含工程改造的(非天然存在的)归巢核酸内切酶(大范围核酸酶)。归巢核酸内切酶和大范围核酸酶例如 I-SceI、I-CeuI、PI-PspI、PI-Sce、I-SceIV、I-CsmI、I-PanI、I-SceII、I-PpoI、I-SceIII、I-CreI、I-TevI、I-TevII 和 I-TevIII 的识别序列是已知的。也参见美国专利号 5, 420, 032 ;美国专利号

6, 833, 252 ;Belfort 等, (1997) *Nucleic Acids Res.* 25:3379-3388 ;Dujon 等, (1989) *Gene* 82:115-118 ;Perler 等, (1994) *Nucleic Acids Res.* 22,1125-1127 ;Jasin(1996) *Trends Genet.* 12:224-228 ;Gimble 等, (1996) *J. Mol. Biol.* 263:163-180 ;Argast 等, (1998) *J. Mol. Biol.* 280:345-353 以及新英格兰生物实验室目录。另外, 归巢核酸内切酶和大范围核酸酶的 DNA 结合特异性可以被工程改造以结合非天然靶位点。参见例如 Chevalier 等, (2002) *Molec. Cell* 10:895-905 ;Epinat 等, (2003) *Nucleic Acids Res.* 31:2952-2962 ;Ashworth 等, (2006) *Nature* 441:656-659 ;Paques 等, (2007) *Current Gene Therapy* 7:49-66 ;美国专利公布号 20070117128。在核酸酶总体情形下, 归巢核酸内切酶和大范围核酸酶的 DNA 结合结构域可改变 (即, 使得核酸酶包括同源裂解结构域) 或可与异源裂解结构域融合。

[0103] 在其它实施方案中, DNA 结合结构域包含天然存在或工程改造 (非天然存在) 的 TAL 效应 DNA 结合结构域。参见例如美国专利公布号 20110301073, 其以全文引用的方式并入本文中。已知黄单胞菌 (*Xanthomonas*) 属的植物病原细菌在重要的作物植物中产生许多疾病。黄单胞菌的致病性取决于保守的 III 型分泌 (T3S) 系统, 其将 25 种以上不同的效应蛋白注入植物细胞中。在这些注入蛋白中是转录活化子样效应子 (TALE), 其模拟植物转录活化子并操纵植物转录组 (参见 Kay 等, (2007) *Science* 318:648-651)。这些蛋白含有 DNA 结合结构域和转录活化结构域。一种最充分表征的 TALE 是来自通用型番茄疮痂病辣椒斑点病菌 (*Xanthomonas campestris* pv. *Vesicatoria*) 的 AvrBs3 (参见 Bonas 等, (1989) *Mol Gen Genet* 218:127-136 和 W02010079430)。TALE 含有串联重复序列的中心化结构域, 每个重复序列含有约 34 个氨基酸, 这对于这些蛋白的 DNA 结合特异性是关键。另外, 它们含有核定位序列和酸性转录活化结构域 (关于综述, 参见 Schornack S 等, (2006) *J Plant Physiol* 163(3):256-272)。另外, 在植物病原细菌青枯雷尔氏菌 (*Ralstonia solanacearum*) 中, 已经发现两种基因, 称为 brg11 和 hpx17, 其与青枯雷尔氏菌生物 1 型菌株 GMI1000 和生物 4 型菌株 RS1000 中的黄单胞菌的 AvrBs3 家族同源 (参见 Heuer 等, (2007) *Appl and Envir Micro* 73(13):4379-4384)。这些基因的核苷酸序列彼此具 98.9% 同一性, 但差异是在 hpx17 的重复结构域中缺失 1,575bp。然而, 两种基因产物与黄单胞菌的 AvrBs3 家族蛋白具有小于 40% 的序列同一性。

[0104] 因此, 在一些实施方案中, 结合于靶基因座 (例如, 珠蛋白或安全港) 中的靶位点的 DNA 结合结构域是来自与源自植物病原体黄单胞菌的那些类似的 TAL 效应子的工程改造的结构域 (参见 Boch 等, (2009) *Science* 326:1509-1512 ;以及 Moscou 和 Bogdanove, (2009) *Science* 326:1501) ;和雷尔氏菌 (参见 Heuer 等, (2007) *Applied and Environmental Microbiology* 73(13):4379-4384) ;美国专利号 8,420,782 和 8,440,431 和美国专利公布号 20110301073。

[0105] 在某些实施方案中, DNA 结合结构域包含锌指蛋白 (例如, 结合于珠蛋白或安全港基因中的靶位点的锌指蛋白)。优选地, 锌指蛋白是非天然存在的, 因为它经过工程改造以结合于所选的靶位点。参见例如 Beerli 等, (2002) *Nature Biotechnol.* 20:135-141 ;Pabo 等, (2001) *Ann. Rev. Biochem.* 70:313-340 ;Isalan 等, (2001) *Nature Biotechnol.* 19:656-660 ;Segal 等, (2001) *Curr. Opin. Biotechnol.* 12:632-637 ;Choo 等, (2000) *Curr. Opin. Struct. Biol.* 10:411-416 ;美国专利号 6,453,242 ;6,534,261 ;

6, 599, 692 ;6, 503, 717 ;6, 689, 558 ;7, 030, 215 ;6, 794, 136 ;7, 067, 317 ;7, 262, 054 ;7, 070, 934 ;7, 361, 635 ;7, 253, 273 ;和美国专利公布号 2005/0064474 ;2007/0218528 ;2005/0267061,全部以全文引用的方式并入本文中。

[0106] 相比于天然存在的锌指蛋白,工程改造的锌指结合或 TALE 结构域可具有新结合特异性。工程改造方法包括(但不限于)合理的设计和各種类型的选择。合理的设计包括例如使用包含三重(或四重)核苷酸序列和单独的锌指氨基酸序列的数据库,其中每个三重或四重核苷酸序列与结合特定三重或四重序列的锌指的一个或多个氨基酸序列缔合。参见例如共同拥有的美国专利 6, 453, 242 和 6, 534, 261,其以全文引用的方式并入本文中。

[0107] 示例性选择方法(包括噬菌体展示和双杂交系统)公开于美国专利 5, 789, 538 ;5, 925, 523 ;6, 007, 988 ;6, 013, 453 ;6, 410, 248 ;6, 140, 466 ;6, 200, 759 ;和 6, 242, 568 ;以及 WO 98/37186 ;WO 98/53057 ;WO 00/27878 ;WO 01/88197 和 GB 2, 338, 237。另外,锌指结合结构域的结合特异性的增强已经描述于例如共同拥有的 WO 02/077227 中。

[0108] 另外,如这些和其它参考文献中所公开,DNA 结构域(例如,多指锌指蛋白或 TALE 结构域)可使用任何合适的接头序列(包括例如长度为 5 个或更多个氨基酸的接头)连接在一起。关于长度为 6 个或更多个氨基酸的示例性接头序列,还参见美国专利号 6, 479, 626 ;6, 903, 185 ;和 7, 153, 949。本文描述的 DNA 结合蛋白可包括蛋白的单独的锌指之间的合适接头的任何组合。另外,锌指结合结构域的结合特异性的增强已经描述于例如共同拥有的 WO 02/077227 中。

[0109] 用于设计和构建融合蛋白(和编码其的多核苷酸)的靶位点、DNA 结合结构域的选择和方法是本领域技术人员已知的并且详细地描述于美国专利号 6, 140, 0815 ;789, 538 ;6, 453, 242 ;6, 534, 261 ;5, 925, 523 ;6, 007, 988 ;6, 013, 453 ;6, 200, 759 ;WO 95/19431 ;WO 96/06166 ;WO 98/53057 ;WO 98/54311 ;WO 00/27878 ;WO 01/60970 ;WO 01/88197 ;WO 02/099084 ;WO 98/53058 ;WO 98/53059 ;WO 98/53060 ;WO 02/016536 和 WO 03/016496 以及美国公布号 20110301073 中。

[0110] 另外,如这些和其它参考文献中所公开,DNA 结合结构域(例如,多指锌指蛋白)可使用任何合适的接头序列(包括例如长度为 5 个或更多个氨基酸的接头)连接在一起。关于长度为 6 个或更多个氨基酸的示例性接头序列,还参见美国专利号 6, 479, 626 ;6, 903, 185 ;和 7, 153, 949。本文所述的蛋白可包括蛋白的单独的锌指之间的合适接头的任何组合。

#### [0111] B. 裂解结构域

[0112] 任何合适的裂解结构域可以操作性地连接至 DNA 结合结构域以形成核酸酶。例如,ZFP DNA 结合结构域已经与核酸酶结构域融合以产生 ZFN,它是能够通过其工程改造的(ZFP)DNA 结合结构域识别其预期核酸靶标并造成 DNA 经由核酸酶活性在 ZFP 结合位点附近被切断的功能实体。参见例如 Kim 等,(1996)Proc Nat'l Acad Sci USA 93(3):1156-1160。最近,ZFN 已经被用于多种生物体中的基因组修饰。参见例如美国专利公开 20030232410 ;20050208489 ;20050026157 ;20050064474 ;20060188987 ;20060063231 ;和国际公布 WO 07/014275。同样,TALE DNA 结合结构域已经与核酸酶结构域融合以产生 TALEN。参见例如美国公布号 20110301073。

[0113] 如上文所示,裂解结构域可与 DNA 结合结构域异源,例如锌指 DNA 结合结构域和来

自核酸酶的裂解结构域,或TALEN DNA结合结构域和裂解结构域,或大范围核酸酶DNA结合结构域和来自不同核酸酶的裂解结构域。异源裂解结构域可获自任何核酸内切酶或核酸外切酶。可获得裂解结构域的示例性核酸内切酶包括(但不限于)限制核酸内切酶和归巢核酸内切酶。参见例如2002-2003Catalogue,New England Biolabs,Beverly,MA;和Belfort等,(1997)*Nucleic Acids Res.* 25:3379-3388。裂解DNA的其它酶是已知的(例如,S1核酸酶;绿豆核酸酶;胰腺DNA酶I;微球菌核酸酶;酵母HO核酸内切酶;也参见Linn等(编)*Nucleases*,Cold Spring Harbor Laboratory Press,1993)。一种或多种的这些酶(或其功能片段)可用作裂解结构域和裂解半结构域的来源。

[0114] 类似地,裂解半结构域可源自如上文所述的任何核酸酶或其部分,其需要二聚合以提供裂解活性。一般来说,如果融合蛋白包含裂解半结构域,则裂解需要两个融合蛋白。可选地,可使用包含两个裂解半结构域的单蛋白。两个裂解半结构域可源自相同的核酸内切酶(或其功能片段),或每个裂解半结构域可源自不同的核酸内切酶(或其功能片段)。另外,两个融合蛋白的靶位点优选关于彼此布置,以使得两个融合蛋白与其相应的靶位点的结合将裂解半结构域置于彼此空间关系中以允许裂解半结构域例如通过二聚形成功能性裂解结构域。因此,在某些实施方案中,靶位点的附近边缘被5-8个核苷酸或15-18个核苷酸隔开。然而,任何整数数目的核苷酸或核苷酸对可以插入两个靶位点之间(例如,2至50个核苷酸对或更多个)。一般来说,裂解位点位于靶位点之间。

[0115] 限制核酸内切酶(限制酶)存在于许多物种中并且能够序列特异性结合于DNA(在识别位点),并且在结合位点处或附近裂解DNA。某些限制酶(例如,IIS型)在从识别位点去除的位点裂解DNA并且具有可分离的结合和裂解结构域。例如,IIS型酶Fok I催化DNA的双链裂解,在一条链上从其识别位点开始的9个核苷酸处以及在另一条链上从其识别位点开始的13个核苷酸处。参见例如美国专利5,356,802;5,436,150和5,487,994;以及Li等,(1992)*Proc. Natl. Acad. Sci. USA* 89:4275-4279;Li等,(1993)*Proc. Natl. Acad. Sci. USA* 90:2764-2768;Kim等,(1994a)*Proc. Natl. Acad. Sci. USA* 91:883-887;Kim等,(1994b)*J. Biol. Chem.* 269:31,978-31,982。因此,在一个实施方案中,融合蛋白包含来自至少一种IIS型限制酶的裂解结构域(或裂解半结构域)和一个或多个锌指结合结构域,其可能被工程改造或可能不被工程改造。

[0116] 示例性IIS型限制酶(其裂解结构域可与结合结构域分离)是Fok I。这种特定的酶作为二聚体形式是具活性的。Bitinaite等,(1998)*Proc. Natl. Acad. Sci. USA* 95:10,570-10,575。因此,为本公开的目的,在所公开的融合蛋白中使用的Fok I酶的部分被视为裂解半结构域。因此,对于使用锌指-Fok I融合体的细胞序列的靶向双链裂解和/或靶向置换,可使用各包含FokI裂解半结构域的两个融合蛋白来重构催化活性裂解结构域。可选地,还可以使用含有DNA结合结构域和两个Fok I裂解半结构域的单多肽分子。

[0117] 裂解结构域或裂解半结构域可以是保持裂解活性或保持多聚合(例如,二聚合)以形成功能性裂解结构域的能力的蛋白的任何部分。

[0118] 示例性IIS型限制酶描述于以全文引用的方式并入本文中的国际公布WO 07/014275中。其它限制酶也含有可分离的结合和裂解结构域,并且这些由本公开所涵盖。参见例如Roberts等,(2003)*Nucleic Acids Res.* 31:418-420。



[0119] 在某些实施方案中,裂解结构域包含一个或多个减少或预防同型二聚合的工程改造的裂解半结构域(也称为二聚合结构域突变体),如例如美国专利公布号 20050064474; 20060188987 和 20080131962 中所述,其全部的公开内容都以全文引用的方式并入本文中。Fok I 的位置 446、447、479、483、484、486、487、490、491、496、498、499、500、531、534、537 和 538 处的氨基酸残基都是影响 Fok I 裂解半结构域的二聚合的靶标。

[0120] 形成专性的异二聚体的 Fok I 的示例性工程改造的裂解半结构域包括一对其中第一裂解半结构域包括在 Fok I 的位置 490 和 538 的氨基酸残基处的突变并且第二裂解半结构域包括在氨基酸残基 486 和 499 处的突变。

[0121] 因此,在一个实施方案中,490 处的突变用 Lys(K) 置换 Glu(E);538 处的突变用 Lys(K) 置换 Iso(I);486 处的突变用 Glu(E) 置换 Gln(Q);并且位置 499 处的突变用 Lys(K) 置换 Iso(I)。特别地,通过在一个裂解半结构域中的位置 490(E → K) 和 538(I → K) 突变以产生工程改造的裂解半结构域(称为“E490K:I538K”)以及通过在另一个裂解半结构域中的位置 486(Q → E) 和 499(I → L) 突变以产生工程改造的裂解半结构域(称为“Q486E:I499L”)来制备本文所述的工程改造的裂解半结构域。本文所述的工程改造的裂解半结构域是其中异常裂解被最小化或消除的专性异二聚体突变体。参见例如美国专利公布号 2008/0131962,其公开内容出于所有目的以全文引用的方式并入。

[0122] 在某些实施方案中,工程改造的裂解半结构域包含在位置 486、499 和 496(相对于野生型 FokI 进行编号)处的突变,例如用 Glu(E) 残基置换位置 486 处的野生型 Gln(Q) 残基、用 Leu(L) 残基置换位置 499 处的野生型 Iso(I) 残基和用 Asp(D) 或 Glu(E) 残基置换位置 496 处的野生型 Asn(N) 残基的突变(也分别称为“ELD”和“ELE”结构域)。在其它实施方案中,工程改造的裂解半结构域包含在位置 490、538 和 537(相对于野生型 FokI 进行编号)处的突变,例如用 Lys(K) 残基置换位置 490 处的野生型 Glu(E) 残基、用 Lys(K) 残基置换位置 538 处的野生型 Iso(I) 残基和用 Lys(K) 残基或 Arg(R) 残基置换位置 537 处的野生型 His(H) 残基的突变(也分别称为“KKK”和“KKR”结构域)。在其它实施方案中,工程改造的裂解半结构域包含在位置 490 和 537(相对于野生型 FokI 进行编号)处的突变,例如用 Lys(K) 残基置换位置 490 处的野生型 Glu(E) 残基和用 Lys(K) 残基或 Arg(R) 残基置换位置 537 处的野生型 His(H) 残基的突变(也分别称为“KIK”和“KIR”结构域)。(参见美国专利公布号 20110201055,以引用的方式并入本文中)。本文描述的工程改造的裂解半结构域可使用任何合适的方法制备,例如,通过野生型裂解半结构域(Fok I)的定点突变诱发,如美国专利公布号 20050064474;20080131962;和 20110201055 中所述。

[0123] 可选地,可使用所谓的“裂解酶”技术在体内核酸靶位点处组装核酸酶(参见例如美国专利公布号 20090068164)。这些裂解酶的组分可表达在单独的表达构建体上,或者可连接于一个开放阅读框中,其中例如通过自裂解 2A 肽或 IRES 序列分离单独的组分。组分可能是单独的锌指结合结构域或大范围核酸酶核酸结合结构域的结构域。

[0124] 可在使用前针对活性筛选核酸酶,例如在基于酵母的染色体系统中,如 WO 2009/042163 和 20090068164 中所述。可使用本领域中已知的方法容易地设计核酸酶表达构建体。参见例如美国专利公开 20030232410;20050208489;20050026157;20050064474;20060188987;20060063231;和国际公布 WO 07/014275。核酸酶的表达可在组成型启动子或诱导型启动子的控制下,例如半乳糖激酶启动子,其在棉子糖和/或半乳糖的存在下活

化（去抑制）并在葡萄糖存在下抑制。

#### [0125] CRISPR/Cas 系统

[0126] 最近对于古菌中 RNA 介导的基因组防御途径的存在已经出现了令人信服的证据，并且已经假设许多细菌类似于真核 RNAi 途径（关于综述，参见 Godde 和 Bickerton, 2006. *J. Mol. Evol.* 62:718-729; Lillestol 等, 2006. *Archaea* 2:59-72; Makarova 等, 2006. *Biol. Direct* 1:7. ; Sorek 等, 2008. *Nat. Rev. Microbiol.* 6:181-186)。称为 CRISPR-Cas 系统或原核 RNAi (pRNAi)，提出从以下两种进化性和通常物理上连接的基因座产生的途径：CRISPR（规律成簇间隔短回文重复）基因座，其编码系统的 RNA 组分，以及 cas（CRISPR 相关）基因座，其编码蛋白（Jansen 等, 2002. *Mol. Microbiol.* 43:1565-1575; Makarova 等, 2002. *Nucleic Acids Res.* 30:482-496; Makarova 等, 2006. *Biol. Direct* 1:7; Haft 等, 2005. *PLoS Comput. Biol.* 1:e60)。微生物宿主中的 CRISPR 基因座含有 CRISPR 相关 (Cas) 基因以及能够编程 CRISPR 介导的核酸裂解的特异性的非编码 RNA 元件的组合。单独的 Cas 蛋白与真核 RNAi 机制的蛋白组分不共有显著序列相似性，但具有类似的预测功能（例如，RNA 结合、核酸酶、解螺旋酶等）（Makarova 等, 2006. *Biol. Direct* 1:7）。CRISPR 相关 (cas) 基因通常与 CRISPR 重复序列 - 间隔序列阵列相关。已经描述了 40 种以上不同的 Cas 蛋白家族。在这些蛋白家族中，在不同的 CRISPR/Cas 系统中，Cas1 似乎普遍存在。已经使用 cas 基因和重复结构的特定组合来定义 8 种 CRISPR 亚型 (Ecoli、Ypest、Nmeni、Dvulg、Tneap、Hmari、Apern 和 Mtube)，其中一些与另外的编码重复序列相关的神秘蛋白 (RAMP) 的基因模块相关。一种以上的 CRISPR 亚型可存在于单一基因组中。CRISPR/Cas 亚型的零星分布表明在微生物进化期间系统经受水平基因转移。

[0127] II 型 CRISPR（由 Cas9 示例）是最充分表征的系统之一并且在四个连续步骤中进行靶向 DNA 双链断裂。首先，从 CRISPR 基因座转录两种非编码 RNA，前体 crRNA 阵列和 tracrRNA。其次，tracrRNA 与前体 crRNA 的重复区域杂交并介导前体 crRNA 加工成含有单独的间隔序列的成熟 crRNA。第三，成熟的 crRNA:tracrRNA 复合物引导 Cas9 以经由 crRNA 上的间隔序列与和原型间隔序列相邻基序 (PAM) 紧邻的靶 DNA 上的原型间隔序列之间的 Watson-Crick 碱基配对来靶向 DNA，这是靶标识别的另一个要求。最后，Cas9 介导靶 DNA 的裂解以在原型间隔序列内产生双链断裂。CRISPR/Cas 系统的活性包含三个步骤：(i) 在所谓的‘适应’过程中，将外来 DNA 序列插入 CRISPR 阵列中以防止未来攻击，(ii) 相关蛋白的表达，以及阵列的表达和加工，接着 (iii) RNA 介导的外来核酸干扰。因此，在细菌细胞中，涉及具有 CRISPR/Cas 系统的天然功能的若干种所谓的‘Cas’蛋白。

[0128] CRISPR 基因座的主要产物似乎是含有侵染靶向序列的短 RNA，并且基于其在途径中的假设作用而称为导向 RNA 或原核沉默 RNA (psiRNA) (Makarova 等, (2006) *Biol. Direct* 1:7; Hale 等, (2008) *RNA* 14:2572-2579)。RNA 分析指示 CRISPR 基因座转录物在重复序列内裂解以释放含有单独的侵染靶向序列和侧接重复序列片段的约 60 至 70-nt RNA 中间物 (Tang 等, (2002) *Proc. Natl. Acad. Sci.* 99:7536-7541; Tang 等, (2005) *Mol. Microbiol.* 55:469-481; Lillestol 等, (2006) *Archaea* 2:59-72; Brouns 等, (2008) *Science* 321:960-964; Hale 等, (2008) *RNA* 14:2572-2579)。在古菌激烈火球菌 (*Pyrococcus furiosus*) 中，这些中间物 RNA 被进一步加工成丰富的稳定的约 35 至 45-nt 成熟 psiRNA (Hale 等, (2008) *RNA* 14:2572-2579)。

[0129] Cas 蛋白

[0130] “Cas1”多肽是指 CRISPR 相关 (Cas) 蛋白 1。Cas1 (蛋白分类系统的相邻类的聚簇中的 COG1518) 是 CRISPR 相关系统 (CASS) 的最佳标志物。基于系统发育比较,已经鉴定出七种不同型式的 CRISPR 相关免疫系统 (CASS1-7)。

[0131] 本文所述方法中使用的 Cas1 多肽可以是任何原核生物中存在的任何 Cas1 多肽。在某些实施方案中, Cas1 多肽是古菌微生物的 Cas1 多肽。在某些实施方案中, Cas1 多肽是广古菌门微生物的 Cas1 多肽。在某些实施方案中, Cas1 多肽是泉古菌微生物的 Cas1 多肽。在某些实施方案中, Cas1 多肽是细菌的 Cas1 多肽。在某些实施方案中, Cas1 多肽是革兰氏阴性或革兰氏阳性细菌的 Cas1 多肽。在某些实施方案中, Cas1 多肽是铜绿假单胞菌 (*Pseudomonas aeruginosa*) 的 Cas1 多肽。在某些实施方案中, Cas1 多肽是超嗜热菌 (*Aquifexaeolicus*) 的 Cas1 多肽。在某些实施方案中, Cas1 多肽是作为 CASS1-7 之一成员的 Cas1 多肽。在某些实施方案中, Cas1 多肽是作为 CASS3 成员的 Cas1 多肽。在某些实施方案中, Cas1 多肽是作为 CASS7 成员的 Cas1 多肽。在某些实施方案中, Cas1 多肽是作为 CASS3 或 CASS7 的成员的 Cas1 多肽。

[0132] 在一些实施方案中, Cas1 多肽是由在 GenBank 中在例如以下 GeneID 编号 2781520、1006874、9001811、947228、3169280、2650014、1175302、3993120、4380485、906625、3165126、905808、1454460、1445886、1485099、4274010、888506、3169526、997745、897836 或 1193018 下提供的核苷酸序列和 / 或与由这些多核苷酸编码的氨基酸展现同源性 (例如, 大于 80%、90% 至 99%, 包括 91%、92%、93%、94%、95%、96%、97%、98% 或 99%) 的氨基酸序列编码并且所述多肽充当 Cas1 多肽。

[0133] Cas6 是另一种 Cas 多肽, 并且核糖核酸内切酶活性在本文被称为 Cas6 核糖核酸内切酶活性。合适的 Cas6 多肽的非限制性实例描述在 Genbank 登记号 AAL81255。Cas6 多肽可从具有 CRISPR 基因座和 cas (CRISPR 相关) 基因座的微生物 (例如但不限于激烈火球菌) 富集、分离或纯化, 或者可使用重组技术产生, 或者使用常规方法化学或酶促合成。在一些方面, Cas6 多肽可不具有 CRISPR 基因座的微生物富集、分离或纯化。Cas6 多肽含有至少一个可在催化中起作用的残基, 或其保守性取代。Cas6 多肽可含有也可在催化中起作用的其它残基, 或其保守性取代。预期在催化中起作用的残基可位于在蛋白的 3D 结构中含有 Cas6 标记基序的富 G 环附近。Cas6 多肽可包括在 TIGRFAM 数据库中在登记号 TIGR01877 和 PF01881 下存在的结构域。TIGRFAM 数据库包括功能保守的多肽家族 (Haft 等, (2003) *Nucl. Acids Res.* 31:371-373; Bateman 和 Haft (2002) *Briefings Bioinformatics*, 3:236-245; 和 Haft 等, (2005) *PLoS Computational Biol.* 1(6):e60)。

[0134] 本文提供的 Cas6 多肽的其它实例包括在具有 CRISPR 基因座和 cas 基因座的原核微生物中存在的那些。可在包括 CRISPR 基因座的任何微生物中容易地鉴定 Cas6 多肽。编码 Cas6 多肽的编码区域通常在紧邻 CRISPR 基因座定位的 cas 基因座中。Haft 等, (2005) *PLoS Computational Biol.* 1(6):e60) 评述了 Cas 蛋白家族, 并且产生用于鉴定 CRISPR/Cas 系统的特定亚型的规则。Haft 等描述了编码 Cas6 多肽的编码区, 如可见于结合至少四种单独的 CRISPR/Cas 亚型 (Tneap、Hmari、Apern 和 Mtube), 并且通常是最远离 CRISPR 基因座定位的 cas 编码区域。可使用在 JCVI Comprehensive Microbial Resource 获得的资源鉴定 Cas6 多肽。因此, 本领域技术人员可使用常规方法鉴定可用于本文所述的方法中的

Cas6 多肽。

[0135] 具有含有预期编码 Cas6 多肽的编码区域的已知完整基因组序列的原核微生物的实例包括海栖热袍菌 (*Thermotogamaritima*)MSB8、胎儿弯曲杆菌 (*Campylobacter fetus*) 胎儿亚种 82-40、具核梭杆菌 (*Fusobacteriumnucleatum*)ATCC 25586、嗜热链球菌 (*Streptococcus thermophilus*)LMG 18311、腾冲嗜热厌氧菌 (*Thermoanaerobactert engcongensis*)MB4(T)、热醋穆尔氏菌 (*Mororellathermoacetica*)ATCC 39073、哥本哈根脱亚硫酸菌 (*Desulfitobacteriumhafniense*)Y51、破伤风梭菌 (*Clostridium tetani*)E88、产气荚膜梭菌 (*Clostridium perfringens*)SM101、艰难梭状芽孢杆菌 (*Clostridium difficile*)QCD-32g58、肉毒杆菌 (*Clostridium botulinum*)Hall A Sanger、肉毒杆菌 F Langel and、肉毒杆菌 B1 株 Okra、肉毒杆菌 A3 株 Loch Maree、肉毒杆菌 A Hall、肉毒杆菌 A ATCC 19397、生氢一氧化碳嗜热菌 (*Carboxydothemushydrogen oformans*)Z-2901、表皮葡萄球菌 (*Staphylococcus epidermidis*)RP62A、嗜热栖热菌 (*Thermusthermophilus*)HB8、嗜热栖热菌 HB27、念珠藻属 (*Nostoc sp.*)PCC 7120、多变鱼腥藻 (*Anabaena variabilis*)ATCC 29413、聚球蓝细菌属 (*Synechococcus sp.*)OS 型 B prime、聚球蓝细菌属 OS 型 A、牙龈卟啉单胞菌 (*Porphyromonasgingivalis*)W83、脆弱类杆菌 (*Bacteroidesfragilis*)YCH46、脆弱类杆菌 NCTC9343、超嗜热菌 VF5、喜木聚糖红色杆菌 (*Rubrobacterxylophilus*)DSM 9941、结核分枝杆菌 (*Mycobacterium tuberculosis*)H37Rv (实验室菌株)、结核分枝杆菌 CDC1551、牛分枝杆菌 (*Mycobacterium bovis*) 牛亚种 AF2122/97、桉木弗兰克氏菌 (*Frankiaalni*)ACN14a、火山热原体 (*Thermoplasmavolcanium*)GSS1、嗜苦古菌 (*Picrophilustorridus*)DSM 9790、超耐热菌 (*Thermococcuskodakarensis*)KOD1、深海嗜热古菌 (*Pyrococcushorikoshii*)OT3、激烈火球菌 (*Pyrococcusfuriosus*)DSM 3638、古核生物深海热球菌 (*Pyrococcusabyssi*)GE5、巴氏甲烷八叠球菌 (*Methanosarcinabarkerifusaro*)、常温产甲烷菌 (*Methanosarcinaacetivorans*)C2A、布氏拟甲烷球菌 (*Methanococcoidesburtonii*)DSM 6242、詹氏甲烷球菌 (*Methanococcusjannaschii*)DSM2661、嗜热自养甲烷杆菌 (*Methanobacteriumthermoautotrophicum*) $\delta$ H、嗜盐死海古菌 (*Haloarculamarismortui*)ATCC 43049、闪烁古生球菌 (*Archaeoglobusfulgidus*)DSM4304、好氧火球菌 (*Pyrobaculum aerophilum*)IM2、托氏硫化叶菌 (*Sulfolobustokodaii*)株 7、硫磺矿硫化叶菌 (*Sulfolobussolfataricus*)P2、嗜酸热硫化裂片菌 (*Sulfolobusacidocaldarius*)DSM 639、敏捷气热菌 (*Aeropyrumpernix*)K1。Cas6 多肽的其它实例为本领域技术人员已知,参见例如多肽的 COG1583 组成员 (可通过美国国家生物技术信息中心互联网网站访问蛋白相邻类的聚簇 (COG) 网页,也参见 Tatusov 等, (1997) *Science* 278:631-637 和 Tatusov 等, (2003) *BMC Bioinformatics* 4(1):41), 具有登记号 IPR010156 的 InterPro 家族成员 (Makarova 等, (2002) *Nuc. Acids Res.* 30:482-496; 和 Haft 等, (2005) *PLoS Comput. Biol.* 1(6):e60, 474-483)。

[0136] 存在三种类型的 CRISPR/Cas 系统,其都并有 RNA 和 Cas 蛋白。I 型和 III 型都具有加工前体 crRNA 的 Cas 核酸内切酶,当完全加工成 crRNA 时,组装能够裂解与 crRNA 互补的核酸的多 Cas 蛋白复合物。

[0137] 在 II 型 CRISPR/Cas 系统中,使用不同机制产生 crRNA,其中与前体 crRNA 中的重复序列互补的反式激活 RNA (tracrRNA) 在 Cas9 蛋白存在下通过双链特异性 RNA 酶 III 触发

加工。然后 Cas9 能够裂解与成熟 crRNA 互补的靶 DNA，然而被 Cas 9 裂解取决于 crRNA 与靶 DNA 之间的碱基配对，并且取决于被称为 PAM 序列（原型间隔序列相邻基序（protospacer adjacent motif））的 crRNA 中的短基序的存在（参见 Qi 等，(2013)Cell 152:1173）。另外，tracrRNA 还必须以其在其 3' 端与 crRNA 的碱基对形式存在，并且这种缔合触发 Cas9 活性。

[0138] Cas9 蛋白具有至少两个核酸酶结构域：一个核酸酶结构域类似于 HNH 核酸内切酶，而另一个类似于 Ruv 核酸内切酶结构域。HNH 型结构域似乎负责裂解与 crRNA 互补的 DNA 链，而 Ruv 结构域裂解非互补链。

[0139] 可通过使用包含通常通过 crRNA 与 tracrRNA 的粘接形成的发夹的工程改造的“单导向 RNA”(sgRNA) 来避免 crRNA-tracrRNA 复合物的需求（参见 Jinek 等，(2012)Science 337:816 和 Cong 等，(2013)Scienceexpress/10.1126/science.1231143）。在化脓性链球菌 (*S. pyogenes*) 中，当 Cas 相关 RNA 和靶 DNA 之间形成双链 RNA:DNA 异二聚体时，工程改造的 tracrRNA:crRNA 融合体或 sgRNA 引导 Cas9 以裂解靶 DNA。包含 Cas9 蛋白和工程改造的含有 PAM 序列的 sgRNA 的这种系统已用于 RNA 引导的基因组编辑（参见 Ramalingam, 同上）并且已经可用于体内斑马鱼胚胎基因组编辑（参见 Hwang 等，(2013)Nature Biotechnology 31(3):227），其中编辑效率类似于 ZFN 和 TALE N。

[0140] 在某些实施方案中，Cas 蛋白可能是天然存在的 Cas 蛋白的“功能衍生物”。天然序列多肽的“功能衍生物”是具有与天然序列多肽相同的定性生物特性的化合物。“功能衍生物”包括（但不限于）天然序列的片段和天然序列多肽的衍生物及其片段，条件是其具有与相应的天然序列多肽相同的生物活性。本文涵盖的生物活性是功能衍生物将 DNA 底物水解成片段的能力。术语“衍生物”涵盖多肽的氨基酸序列变体、共价修饰和其融合体。

[0141] “Cas 多肽”涵盖全长 Cas 多肽、Cas 多肽的酶促活性片段，和 Cas 多肽或其片段的酶促活性衍生物。Cas 多肽或其片段的合适的衍生物包括（但不限于）Cas 蛋白或其片段的突变体、融合体、共价修饰。

[0142] Cas 蛋白和 Cas 多肽可获自细胞或者通过化学合成或通过这两种程序的组合合成。细胞可以是天然产生 Cas 蛋白的细胞，或者是天然产生 Cas 蛋白并被遗传工程改造以在较高表达水平下产生内源 Cas 蛋白或从外源引入核酸产生 Cas 蛋白的细胞，所述核酸编码与内源 Cas 相同或不同的 Cas。在一些情况下，细胞不会天然产生 Cas 蛋白并且被遗传工程改造以产生 Cas 蛋白。

[0143] CRISPR/Cas 系统也可用于抑制基因表达。Lei 等，(2013)Cell 152(5):1173-1183 已经证实，催化死亡的 Cas9 缺乏核酸内切酶活性，当与导向 RNA 共同表达时，产生可特异性干扰转录伸长、RNA 聚合酶结合或转录因子结合的 DNA 识别复合物。这种系统（称为 CRISPR 干扰 (CRISPRi)）可以有效地抑制靶向基因的表达。

[0144] 另外，已经开发了 Cas 蛋白，在其裂解结构域中包含突变以使得它们不能诱导 DSB，而是将切口引入靶 DNA 中（“Cas9 切口酶”，参见 Cong 等，同上）。

[0145] 本发明的 Cas 蛋白可被突变以改变功能性。示例性选择方法（包括噬菌体展示和双杂交系统）公开于美国专利 5,789,538;5,925,523;6,007,988;6,013,453;6,410,248;6,140,466;6,200,759;和 6,242,568;以及 WO 98/37186;WO 98/53057;WO 00/27878;WO 01/88197 和 GB 2,338,237。

[0146] CRISPR/Cas 的 RNA 组分

[0147] Cas9 相关 CRISPR/Cas 系统包含两种 RNA 非编码组分：含有由同一的同向重复序列 (DR) 隔开的核酸酶引导序列 (间隔序列) 的 tracrRNA 和前体 crRNA 阵列。为了使用 CRISPR/Cas 系统来实现基因组工程改造, 这些 RNA 的两种功能必须存在 (参见 Cong 等, (2013) Scienceexpress 1/10. 1126/science 1231143)。在一些实施方案中, 经由单独的表达构建体或以单独的 RNA 形式提供 tracrRNA 和前体 crRNA。在其它实施方案中, 构建嵌合 RNA, 其中将工程改造的成熟 crRNA (赋予靶标特异性) 与 tracrRNA (提供与 Cas9 相互作用) 融合以产生嵌合的 cr-RNA-tracrRNA 杂合体 (也称为单导向 RNA)。(参见 Jinek, 同上; 和 Cong, 同上)。

[0148] 嵌合或 sgRNA 可被工程改造以包含与任何所需靶标互补的序列。RNA 包含与靶标的 22 个碱基互补性并具有形式 G[n19], 接着是形式 NGG 的原型间隔序列相邻基序 (PAM)。因此, 在一个方法中, 可通过在目标基因中利用已知 ZFN 靶标来设计 sgRNA, 其通过: (i) 比对 ZFN 异二聚体的识别序列与相关基因组 (人类、小鼠或特定植物物种) 的参考序列; (ii) 鉴定 ZFN 半位点之间的间隔序列区域; (iii) 鉴定最接近于间隔序列区域的基序 G[N20]GG 的位置 (当一个以上的这种基序与间隔序列重叠时, 选择相对于间隔序列居中的基序); (iv) 使用所述基序作为 sgRNA 的核心。这种方法有利地依赖于证实的核酸酶靶标。可选地, sgRNA 可以经过设计以简单地通过鉴定符合 G[n20]GG 式的合适的靶序列来靶向任何目标区域。

[0149] 靶位点

[0150] 如上文所详细描述, DNA 结合结构域可以经过工程改造以结合于基因座例如珠蛋白或安全港基因中的任何所选序列。相比于天然存在的 DNA 结合结构域, 工程改造的 DNA 结合结构域可具有新结合特异性。工程改造方法包括 (但不限于) 合理的设计和类型的选择。合理的设计包括例如使用包含三重 (或四重) 核苷酸序列和单独的 (例如, 锌指) 氨基酸序列的数据库, 其中每个三重或四重核苷酸序列与结合特定三重或四重序列的 DNA 结合结构域的一个或多个氨基酸序列缔合。参见例如共同拥有的美国专利 6, 453, 242 和 6, 534, 261, 其以全文引用的方式并入本文中。也可以进行 TAL-效应子结构域的合理设计。参见例如美国专利公布号 20110301073。

[0151] 可应用于 DNA 结合结构域的示例性选择方法 (包括噬菌体展示和双杂交系统) 公开于美国专利 5, 789, 538 ; 5, 925, 523 ; 6, 007, 988 ; 6, 013, 453 ; 6, 410, 248 ; 6, 140, 466 ; 6, 200, 759 ; 和 6, 242, 568 ; 以及 WO 98/37186 ; WO 98/53057 ; WO 00/27878 ; WO 01/88197 和 GB 2, 338, 237 中。

[0152] 用于设计和构建融合蛋白 (和编码其的多核苷酸) 的靶位点、核酸酶的选择和方法是本领域技术人员已知的并且详细地描述于美国专利申请公布号 20050064474 和 20060188987, 其以全文引用的方式并入本文中。

[0153] 另外, 如这些和其它参考文献中所公开, DNA 结合结构域 (例如, 多指锌指蛋白) 可使用任何合适的接头序列 (包括例如长度为 5 个或更多个氨基酸的接头) 连接在一起。关于长度为 6 个或更多个氨基酸的示例性接头序列, 参见例如美国专利号 6, 479, 626 ; 6, 903, 185 ; 和 7, 153, 949。本文所述的蛋白可包括蛋白的单独的 DNA 结合结构域之间的合适接头的任何组合。也参见美国专利公布号 20110287512。

[0154] 供体

[0155] 如上文所示,插入外源序列(也称为“供体序列”或“供体”或“转基因”),例如以校正突变基因或用于增加野生型基因的表达。将显而易见的是,供体序列通常不与其所安置的基因组序列同一。供体序列可含有非同源序列,其侧接两个同源区域以允许在目标位置处的有效HDR。另外,供体序列可包含含有不与细胞染色质中的目标区域同源的序列的载体分子。供体分子可含有若干个不连续的与细胞染色质同源的区域。例如,对于通常不存在于目标区域中的序列的靶向插入,所述序列可存在于供体核酸分子中并侧接与目标区域中的序列同源的区域。

[0156] 本文描述用于插入所选位置中的任何多核苷酸的靶向插入方法。用于插入的多核苷酸也可以称为“外源”多核苷酸、“供体”多核苷酸或分子或“转基因”。供体多核苷酸可以是DNA或RNA,单链和/或双链,并且可以线性或环状形式引入细胞中。参见例如美国专利公布号20100047805、20110281361、20110207221和美国申请号13/889,162。供体序列可包含在DNA MC内,其可以环状或线性形式引入细胞中。如果以线性形式引入,则可通过本领域技术人员已知的方法来保护供体序列的末端(例如,免于核酸外切降解)。例如,将一个或多个二脱氧核苷酸残基添加至线性分子的3'端和/或将自互补寡核苷酸连接至一端或两端。参见例如Chang等,(1987)Proc. Natl. Acad. Sci. USA84:4959-4963;Nehls等,(1996)Science 272:886-889。用于保护外源多核苷酸免于降解的其它方法包括(但不限于)添加末端氨基和使用修饰的核苷酸间键,例如,硫代磷酸酯、氨基磷酸酯和O-甲基核糖或脱氧核糖残基。

[0157] 可将多核苷酸引入细胞中作为具有其它序列例如复制起点、启动子和编码抗生素抗性的基因的载体分子的一部分。此外,供体多核苷酸可以裸核酸形式、与例如脂质体或泊洛沙姆(poloxamer)等药剂复合的核酸形式引入,或者可通过病毒(例如,腺病毒、AAV、疱疹病毒、逆转录病毒、慢病毒和整合酶缺陷慢病毒(IDLV))递送。

[0158] 在某些实施方案中,双链供体包括长度大于1kb、例如2至200kb、2至10kb(或介于其间的任何数值)的序列(例如,编码序列,也称为转基因)。双链供体也包括例如至少一个核酸酶靶位点。在某些实施方案中,供体包括至少1个靶位点,例如,用于CRISPR/Cas,或2个靶位点,例如用于一对ZFN或TALEN。通常,核酸酶靶位点是在转基因序列外部,例如,转基因序列的5'和/或3',以裂解转基因。核酸酶裂解位点可用于任何核酸酶。在某些实施方案中,双链供体中所含的核酸酶靶位点是针对用于裂解内源靶标的相同核酸酶,在所述内源靶标中已经由同源非依赖性方法整合裂解供体。

[0159] 通常插入供体以使得其表达受到整合位点处的内源启动子驱动,所述内源启动子即驱动已插入供体中的内源基因的表达的启动子(例如,珠蛋白、AAVS1等)。然而,将显而易见的是,供体可包含启动子和/或增强子,例如组成型启动子或者诱导型或组织特异性启动子。

[0160] 可将供体分子插入内源基因中以使得所有、一些或无内源基因被表达。例如,如本文所述的转基因可插入珠蛋白基因座中以使得一些或无内源珠蛋白序列被表达,例如以与转基因的融合体形式。在其它实施方案中,将转基因(例如,具有或不具有珠蛋白编码序列)整合至任何内源基因座,例如安全港基因座。参见例如美国专利公开20080299580;20080159996和201000218264。

[0161] 当用转基因表达其它（例如，珠蛋白序列、内源或转基因的一部分）时，另外的（例如，珠蛋白）序列可为全长序列（野生型或突变型）或部分序列。其它序列优选具功能性。这些全长或部分的其它序列、例如珠蛋白编码序列的功能的非限制性实例包括增加由转基因（例如，治疗基因）表达的多肽的血清半衰期和 / 或充当载体。

[0162] 此外，尽管不是表达所需，但外源序列还可包括转录或翻译调节序列，例如，启动子、增强子、绝缘子、内部核糖体进入位点、编码 2A 肽的序列和 / 或多腺苷酸化信号。

[0163] 在本文所述的供体序列上携带的转基因可使用本领域中已知的标准技术例如 PCR 从质粒、细胞或其他来源分离。供使用的供体可包括不同类型的拓扑结构，包括超螺旋环状、松弛环状、线性等。可选地，它们可使用标准寡核苷酸合成技术来化学合成。另外，供体可被甲基化或缺乏甲基化。供体可呈细菌或酵母人工染色体 (BAC 或 YAC) 形式。

[0164] 本文所述的双链供体多核苷酸可包括一个或多个非天然碱基和 / 或主链。特别是，可使用本文所述的方法进行具有甲基化胞嘧啶的供体分子的插入以在目标区域中实现转录休眠状态。

[0165] 外源（供体）多核苷酸可包含任何目标序列（外源序列）。示例性外源序列包括（但不限于）任何多肽编码序列（例如，cDNA）、启动子序列、增强子序列、表位标签、标记基因、裂解酶识别位点和各种类型的表达构建体。标记基因包括（但不限于）编码介导抗生素抗性（例如，氨苄青霉素抗性、新霉素抗性、G418 抗性、嘌呤霉素抗性）的序列、编码有色或荧光或发光蛋白（例如，绿色荧光蛋白、增强型绿色荧光蛋白、红色荧光蛋白、荧光素酶）的序列，和介导增强的细胞生长和 / 或基因扩增的蛋白（例如，二氢叶酸还原酶）。表位标签包括例如一个或多个拷贝的 FLAG、His、myc、Tap、HA 或任何可检测的氨基酸序列。

[0166] 在一个优选的实施方案中，外源序列（转基因）包含编码需要在细胞中表达的任何多肽的多核苷酸，所述多肽包括（但不限于）抗体、抗原、酶、受体（细胞表面或核）、激素、淋巴因子、细胞因子、报告多肽、生长因子和任何上述物质的功能片段。编码序列可以是例如 cDNA。

[0167] 在某些实施方案中，外源序列可包含允许选择已经靶向整合的细胞的标记基因（上文所述），和编码其它功能性的连接序列。标记基因的非限制性实例包括 GFP、药物选择标记等。

[0168] 可插入的其它基因序列可包括例如置换突变序列的野生型基因。例如，野生型  $\beta$  珠蛋白基因序列可插入干细胞基因组中，其中基因的内源拷贝被突变。野生型拷贝可插入在内源基因座处，或者可能可选地靶向安全港基因座。

[0169] 遵循本说明书的教导，这些表达盒的构建利用分子生物领域中公知的方法（参见例如 Ausubel 或 Maniatis）。在使用表达盒来产生转基因动物之前，可通过将表达盒引入合适的细胞系（例如，原代细胞、转化细胞或永生化细胞系）来测试表达盒对于与所选控制元件相关的应力诱导物的反应性。

[0170] 此外，尽管不是表达所需，但外源序列还可包括转录或翻译调节序列，例如，启动子、增强子、绝缘子、内部核糖体进入位点、编码 2A 肽的序列和 / 或多腺苷酸化信号。此外，目标基因的控制元件可以可操作地连接至报告基因以产生嵌合基因（例如，报告表达盒）。

[0171] 也可能实现非编码核酸序列的靶向插入。编码反义 RNA、RNAi、shRNA 和微 RNA (miRNA) 的序列也可用于靶向插入。



[0172] 在其它实施方案中,供体核酸可包含对于其它核酸酶设计的特异性靶位点的非编码序列。随后,其它核酸酶可表达于细胞中以使得原始供体分子裂解并通过插入另一种目标供体分子来修饰。以这种方式,可产生供体分子的反复整合,从而使得在特定的目标基因座处或在安全港基因座处特性堆叠。

[0173] 递送

[0174] 本文描述的核酸酶、编码这些核酸酶的多核苷酸、供体多核苷酸和包含所述蛋白和 / 或多核苷酸的组合物可通过任何合适的方式在体内或离体递送。

[0175] 如本文所述的递送核酸酶的方法描述于例如美国专利号 6,453,242 ;6,503,717 ;6,534,261 ;6,599,692 ;6,607,882 ;6,689,558 ;6,824,978 ;6,933,113 ;6,979,539 ;7,013,219 ;和 7,163,824 中,其全部的公开内容以全文引用的方式并入本文中。

[0176] 如本文所述的核酸酶和 / 或供体构建体也可使用含有编码一个或多个锌指或 TALEN 蛋白的序列的载体递送。可使用任何载体系统,包括(但不限于)质粒载体、逆转录病毒载体、慢病毒载体、腺病毒载体、痘病毒载体;疱疹病毒载体和腺相关病毒载体等。也参见美国专利号 6,534,261 ;6,607,882 ;6,824,978 ;6,933,113 ;6,979,539 ;7,013,219 ;和 7,163,824,其以全文引用的方式并入本文中。此外,将显而易见的是,任何这些载体可包含一种或多种治疗所需的序列。因此,当将一种或多种核酸酶和供体构建体引入细胞中时,核酸酶和 / 或供体多核苷酸可携带在相同载体上或不同载体上。当使用多个载体时,每个载体可包含编码一个或多个核酸酶和 / 或供体构建体的序列。

[0177] 可使用常规的基于病毒和非病毒的基因转移方法将编码酶的核酸和供体构建体引入细胞(例如,哺乳动物细胞)和靶组织中。非病毒载体递送系统包括 DNA 质粒、裸核酸,和与递送媒介物例如脂质体或泊洛沙姆复合的核酸。病毒载体递送系统包括 DNA 和 RNA 病毒,其在递送至细胞后具有游离体或整合的基因组。对于基因治疗程序的综述,参见 Anderson, *Science* 256:808-813(1992); Nabel 和 Felgner, *TIBTECH* 11:211-217(1993); Mitani 和 Caskey, *TIBTECH* 11:162-166(1993); Dillon, *TIBTECH* 11:167-175(1993); Miller, *Nature* 357:455-460(1992); Van Brunt, *Biotechnology* 6(10):1149-1154(1988); Vigne, *Restorative Neurology and Neuroscience* 8:35-36(1995); Kremer 和 Perricaudet, *British Medical Bulletin* 51(1):31-44(1995); Haddada 等, *Current Topics in Microbiology and Immunology* Doerfler 和 Böhm (编)(1995); 以及 Yu 等, *Gene Therapy* 1:13-26(1994)。

[0178] 核酸的非病毒递送方法包括电穿孔、脂转染、显微注射、基因枪法、病毒颗粒、脂质体、免疫脂质体、聚阳离子或脂质:核酸缀合物、裸 DNA、人工病毒粒子和 DNA 的药剂增强吸收。使用例如 Sonitron 2000 系统(Rich-Mar)进行的声孔作用(Sonoporation)还可用于递送核酸。

[0179] 其它示例性核酸递送系统包括由 Amaxa Biosystems(Cologne, Germany)、Maxcyte, Inc.(Rockville, Maryland)、BTX Molecular Delivery Systems(Holliston, MA) 和 Copernicus Therapeutics Inc 所提供的那些(参见例如 US6008336)。脂转染描述于例如美国专利号 5,049,386 ;4,946,787 ;和 4,897,355 中并且脂转染试剂可商业销售(例如, Transfectam™ 和 Lipofectin™)。适于多核苷酸的有效受体识别脂转染的阳离子性和中性脂质包括 Felgner、WO 91/17424、WO 91/16024 的那些。

[0180] 脂质：核酸复合物（包括靶向脂质体，例如免疫脂质复合物）的制备是本领域技术人员所公知的（参见例如 Crystal, Science 270:404-410(1995) ;Blaese 等, Cancer Gene Ther. 2:291-297(1995) ;Behr 等, Bioconjugate Chem. 5:382-389(1994) ;Remy 等, Bioconjugate Chem. 5:647-654(1994) ;Gao 等, Gene Therapy 2:710-722(1995) ;Ahmad 等, Cancer Res. 52:4817-4820(1992) ;美国专利号 4,186,183、4,217,344、4,235,871、4,261,975、4,485,054、4,501,728、4,774,085、4,837,028 和 4,946,787）。

[0181] 其它递送方法包括使用待递送至 EnGeneIC 递送媒介物 (EDV) 中的核酸的包装。使用双特异性抗体将这些 EDV 特别递送至靶组织，其中所述抗体的一个臂对靶组织具有特异性并且另一个对 EDV 具有特异性。抗体将 EDV 送至靶细胞表面，然后通过内吞使 EDV 进入细胞中。一旦在细胞中，就释放出内含物（参见 MacDiarmid 等, (2009) Nature Biotechnology 27(7):643）。

[0182] 使用基于 RNA 或 DNA 病毒的系统来递送编码工程改造的 ZFP 的核酸，其利用用于将病毒靶向体内的特异性细胞的高度进化方法以及向细胞核运输病毒有效载荷。病毒载体可直接施用于受试者（体内）或者它们可用于在体外处理细胞并且将经过修饰的细胞施用于受试者（离体）。常规的用于递送 ZFP 的基于病毒的系统包括（但不限于）逆转录病毒、慢病毒、腺病毒、腺相关病毒、牛痘和单纯疱疹病毒载体以用于基因转移。利用逆转录病毒、慢病毒和腺相关病毒基因转移方法整合于宿主基因组中是可能的，通常导致所插入转基因的长期表达。另外，已经在许多不同的细胞类型和靶组织中观察到高转导效率。

[0183] 可通过并入外来包膜蛋白来改变逆转录病毒的趋向性，从而扩大靶细胞的潜在靶标群体。慢病毒载体是能够转导或感染非分裂细胞的逆转录病毒载体并且通常产生高病毒滴度。逆转录病毒基因转移系统的选择取决于靶组织。逆转录病毒载体由具有至多 6-10kb 的外来序列的包装容量的顺式作用的长末端重复序列构成。最小的顺式作用 LTR 足以用于载体的复制和包装，其然后用于将治疗基因整合至靶细胞中以提供永久的转基因表达。广泛使用的逆转录病毒载体包括基于鼠类白血病病毒 (MuLV)、长臂猿白血病病毒 (GaLV)、猿免疫缺陷病毒 (SIV)、人类免疫缺陷病毒 (HIV) 和其组合的那些（参见例如 Buchscher 等, J. Virol. 66:2731-2739(1992) ;Johann 等, J. Virol. 66:1635-1640(1992) ;Sommerfelt 等, Virol. 176:58-59(1990) ;Wilson 等, J. Virol. 63:2374-2378(1989) ;Miller 等, J. Virol. 65:2220-2224(1991) ;PCT/US94/05700）。

[0184] 在其中优选短暂表达的应用中，可使用基于腺病毒的系统。基于腺病毒的载体能够在许多细胞类型中具有非常高的转导效率并且不需要细胞分裂。利用这些载体，已经获得高滴度和高水平表达。这种载体可大量产生于相对简单的系统中。腺相关病毒（“AAV”）载体还用于转导具有靶核酸的细胞，例如，在核酸和肽的体外产生中，以及用于体内和离体基因疗法程序（参见例如 West 等, Virology 160:38-47(1987) ;美国专利号 4,797,368 ;WO 93/24641 ;Kotin, Human Gene Therapy 5:793-801(1994) ;Muzyczka, J. Clin. Invest. 94:1351(1994)。重组 AAV 载体的构建描述于多个出版物中，包括美国专利号 5,173,414 ;Tratschin 等, Mol. Cell. Biol. 5:3251-3260(1985) ;Tratschin 等, Mol. Cell. Biol. 4:2072-2081(1984) ;Hermonat 和 Muzyczka, PNAS 81:6466-6470(1984) ;以及 Samulski 等, J. Virol. 63:03822-3828(1989)。

[0185] 至少六种病毒载体方法目前可用于临床试验中的基因转移，其利用涉及通过插入

辅助细胞系中的基因来补充有缺陷的载体以产生转导剂的方法。

[0186] pLASN 和 MFG-S 是已用于临床试验中的逆转录病毒载体的实例 (Dunbar 等, Blood 85:3048-305(1995); Kohn 等, Nat. Med. 1:1017-102(1995); Malech 等, PNAS 94:2212133-12138(1997))。PA317/pLASN 是用于基因疗法试验中的第一治疗载体。(Blaese 等, Science 270:475-480(1995))。对于 MFG-S 包装载体已观察到 50% 或更大的转导效率。(Ellem 等, Immunol Immunother. 44(1):10-20(1997); Dranoff 等, Hum. Gene Ther. 1:111-2(1997))。

[0187] 重组腺相关病毒载体 (rAAV) 是一种有前途的可选基因递送系统, 其基于有缺陷的和非致病性细小病毒腺相关 2 型病毒。所有载体都源自仅保留侧接转基因表达盒的 AAV 145bp 反向末端重复序列的质粒。由于整合至转导细胞的基因组中的有效基因转移和稳定转基因递送是这种载体系统的关键特征。(Wagner 等, Lancet 351:91171702-3(1998); Kearns 等, Gene Ther. 9:748-55(1996))。也可以根据本发明使用其它 AAV 血清型, 包括 AAV1、AAV3、AAV4、AAV5、AAV6、AAV8、AAV9 和 AAVrh10 以及其所有变体。

[0188] 复制缺陷型重组腺病毒载体 (Ad) 可以高滴度产生并且容易感染多种不同的细胞类型。大多数腺病毒载体经过工程改造以使得转基因置换 Ad E1a、E1b 和 / 或 E3 基因; 随后复制缺陷型载体在提供反式缺失基因功能的人类 293 细胞中增殖。Ad 载体可在体内转导多种类型的组织, 包括非分裂、分化的细胞, 例如在肝脏、肾脏和肌肉中发现的那些。常规的 Ad 载体具有大的携带能力。在临床试验中使用 Ad 载体的实例涉及用肌肉内注射进行抗肿瘤免疫的多核苷酸疗法 (Sterman 等, Hum. Gene Ther. 7:1083-9(1998))。在临床试验中使用腺病毒载体进行基因转移的其它实例包括 Rosenecker 等, Infection 24:15-10(1996); Sterman 等, Hum. Gene Ther. 9:71083-1089(1998); Welsh 等, Hum. Gene Ther. 2:205-18(1995); Alvarez 等, Hum. Gene Ther. 5:597-613(1997); Topf 等, Gene Ther. 5:507-513(1998); Sterman 等, Hum. Gene Ther. 7:1083-1089(1998)。

[0189] 使用包装细胞形成能够感染宿主细胞的病毒粒子。这些细胞包括 293 细胞, 其包装腺病毒; 以及  $\psi$  2 细胞或 PA317 细胞, 其包装逆转录病毒。通常通过将核酸载体包装至病毒粒子中的产生剂细胞系来生成基因疗法中所用的病毒载体。所述载体通常含有包装所需的最小病毒序列以及随后整合至宿主中 (如果适用的话), 其它病毒序列被编码待表达的蛋白的表达盒置换。通过包装细胞系反式提供缺少的病毒功能。例如, 在基因疗法中使用的 AAV 载体通常仅具有来自 AAV 基因组的反向末端重复 (ITR) 序列, 其为包装和整合至宿主基因组中所需。将病毒 DNA 包装在细胞系中, 其含有编码其它 AAV 基因 (即 rep 和 cap) 的辅助质粒, 但缺乏 ITR 序列。细胞系也用腺病毒感染作为辅助。辅助病毒促进 AAV 载体的复制和来自辅助质粒的 AAV 基因的表达。由于缺乏 ITR 序列, 辅助质粒不以显著量包装。可以通过例如热处理来降低腺病毒污染, 腺病毒比 AAV 对热处理更敏感。

[0190] 在许多基因疗法应用中, 希望基因疗法载体以高特异性程度递送至特定组织类型。因此, 通过在病毒的外表面上表达呈与病毒外壳蛋白的融合蛋白形式的配体, 病毒载体可被修饰以对特定细胞类型具有特异性。所述配体被选择为对已知在目标细胞类型上存在的受体具有亲和力。例如, Han 等, Proc. Natl. Acad. Sci. USA 92:9747-9751(1995) 报道了莫洛尼鼠类白血病病毒可以被修饰以表达与 gp70 融合的人类调蛋白, 并且重组病毒感染某些表达人类表皮生长因子受体的人类乳腺癌细胞。这种原理可以扩展到其它病毒-靶细

胞对,其中靶细胞表达受体并且病毒表达包含细胞表面受体的配体的融合蛋白。例如,丝状噬菌体可以被工程改造以显示对几乎任何所选细胞受体具有特异性结合亲和力的抗体片段(例如,FAB或Fv)。尽管上面的描述主要适用于病毒载体,但相同的原理可适用于非病毒载体。这些载体可以被工程改造以含有有利于被特异性靶细胞吸收的特异性吸收序列。

[0191] 基因疗法载体可通过施用至个体受试者,通常通过如下所述的全身施用(例如,静脉内、腹膜内、肌肉内、皮下或颅内输注)或局部施用在体内递送。可选地,载体可离体递送至细胞,例如从个体患者外植的细胞(例如,淋巴细胞、骨髓抽吸物、组织活检)或普遍供体造血干细胞,接着将细胞再植入患者中,通常在选择已并有载体的细胞后。

[0192] 还可以直接向用于转导细胞的生物体体内直接施用含有核酸酶和/或供体构建体的载体(例如,逆转录病毒、腺病毒、脂质体等)。可选地,可施用裸DNA。施用是通过通常用于将分子引入与血液或组织细胞最终接触的任何途径,包括(但不限于)注射、输注、局部施用和电穿孔。适于施用这些核酸的方法是可用的并且为本领域技术人员所公知,并且尽管可使用一种以上途径来施用特定组合物,但特定途径通常可相比于另一种途径提供更直接和更有效的反应。

[0193] 适于引入本文所述的多核苷酸的载体包括非整合慢病毒载体(IDLV)。参见例如Ory等,(1996)Proc.Natl.Acad.Sci.USA93:11382-11388;Dull等,(1998)J.Virol.72:8463-8471;Zuffery等,(1998)J.Virol.72:9873-9880;Follenzi等,(2000)Nature Genetics 25:217-222;美国专利公布号2009/054985。

[0194] 药学上可接受的载体部分地通过所施用的特定组合物,以及通过用于施用组合物的特定方法确定。因此,存在广泛多种合适的如下所述可用的药物组合物的制剂(参见例如Remington's Pharmaceutical Sciences,第17版,1989)。

[0195] 将显而易见的是,可使用相同或不同的系统来递送编码核酸酶的序列和供体构建体。例如,供体多核苷酸可由质粒携带,而一种或多种核酸酶可由AAV载体携带。此外,不同的载体可藉由相同或不同的途径(肌肉内注射、尾静脉注射、其它静脉内注射、腹膜内施用和/或肌肉内注射施用。载体可同时或以任何连续顺序递送。

[0196] 因此,本公开包括适于插入编码治疗蛋白的转基因的疾病和病状的体内或离体治疗,例如经由编码珠蛋白的基因的核酸酶介导的整合来治疗血红蛋白病。组合物以可在血清或靶器官或细胞中有效获得所需治疗多肽浓度的量施用至人类患者。施用可以通过将多核苷酸递送至所需靶细胞的任何方式。例如,涵盖体内和离体方法。静脉内注射至门静脉是优选的施用方法。其它体内施用模式包括例如直接注射至肝叶或胆管和远侧静脉内注射至肝脏,包括经由肝动脉,直接注射到肝实质,经由肝动脉注射,和/或经由胆管树逆行注射。离体施用模式包括切除肝细胞或肝脏的其它细胞的体外转导,接着输注转导的切除肝细胞回到人类患者的门脉管、肝实质或胆管树,参见例如Grossman等,(1994)Nature Genetics,6:335-341。

[0197] 待施用的核酸酶和供体的有效量在不同患者之间并根据目标治疗多肽将有所改变。因此,最好由施用组合物的医师来确定有效量并且本领域的普通技术人员可容易地确定适当剂量。在允许足够的时间来整合和表达(例如,通常4-15天)后,治疗多肽的血清或其它组织水平的分析和与施用之前的初始水平的比较将确定所施用的量是否过低、在正确范围内或过高。对于初始和后续施用的合适方案也是可变的,但典型的是进行初始施用,

接着在必要时进行后续施用。后续施用可以不同的间隔施用,范围是每天一次至每年一次至每几年一次。本领域技术人员将理解,可建议适当的免疫抑制技术来避免递送载体的免疫抑制对转导的抑制或阻断,参见例如 Vilquin 等, (1995) Human Gene Ther. 6:1391-1401。

[0198] 用于离体和体内施用的制剂包括于液体或乳化液体中的悬浮液。活性成分通常与赋形剂混合,所述赋形剂是药学上可接受的并且与所述活性成分相容。合适的赋形剂包括例如水、盐水、右旋糖、甘油、乙醇等,以及它们的组合。另外,组合物可含有少量的辅助物质,例如,润湿或乳化剂、pH 缓冲剂、稳定剂或其它增强药物组合物的有效性的试剂。

[0199] 应用

[0200] 本文公开的方法和组合物是用于调节蛋白的表达,或校正对遗传性疾病例如镰状细胞疾病或地中海贫血中所表达的蛋白进行编码的异常基因序列。因此,所述方法和组合物提供这些遗传性疾病的治疗和 / 或预防。例如干细胞的基因组编辑被用于校正异常基因,插入野生型基因,或改变内源基因的表达。通过非限制性实例,可将例如编码至少一种珠蛋白(例如,  $\alpha$  和 / 或  $\beta$  珠蛋白)的野生型基因插入细胞中以提供细胞中不足和 / 或缺乏的珠蛋白,并且从而治疗由有缺陷的珠蛋白表达造成的遗传性疾病,例如血红蛋白病。可选地或另外,在存在或不存在适当供体施用下的基因组编辑可校正有缺陷的内源基因,例如,校正  $\alpha$  血红蛋白或  $\beta$  血红蛋白中的点突变,恢复基因的表达和 / 或治疗遗传性疾病,例如镰状细胞疾病和 / 或任何直接或间接珠蛋白调节基因的敲除或改变(过度表达或抑制)(例如  $\gamma$  珠蛋白调节基因 BCL11A 或 BCL11A 调节子 KLF1 的失活)。

[0201] 本发明的方法和组合物还可用于需要提供编码一种或多种治疗剂的转基因的任何情形中,以便所述治疗剂在 RBC 和 / 或造血干细胞中产生以使得源自这些细胞的成熟 RBC 含有治疗剂。

[0202] 下列实施例涉及本公开的示例性实施方案,其中核酸酶包含锌指核酸酶(ZFN)或 TALEN。应理解,这只是为了示例目的并且可使用其它核酸酶,例如具有工程改造的 DNA 结合结构域的归巢核酸内切酶(大范围核酸酶)和 / 或天然存在的工程改造的归巢核酸内切酶(大范围核酸酶)DNA 结合结构域和包含工程改造的单导向 RNA 的异源裂解结构域和 / 或 CRISPR/Cas 系统的融合体。

## 实施例

[0203] 实施例 1: 锌指蛋白核酸酶(ZFN)的设计、构建和一般表征

[0204] 对锌指蛋白进行设计并且并入质粒、AAV 或腺病毒载体中,基本上如 Urnov 等, (2005) Nature 435(7042):646-651; Perez 等, (2008) Nature Biotechnology 26(7):808-816 中所述,以及如美国专利号 6,534,261 中所述。关于对人类  $\beta$  珠蛋白基因座和人类 HPRT 基因座具有特异性的 ZFN 和 TALEN,参见共同拥有的美国专利号 7,888,121 和美国专利公布号 20130137104 和 20130122591。关于对人类 AAVS1 具有特异性的核酸酶,参见共同拥有的美国专利号 8,110,379。关于对 CCR5 具有特异性的核酸酶,参见共同拥有的美国专利号 7,951,925。关于对白蛋白具有特异性的核酸酶,参见美国专利公布号 20130177983 和 20139177968。

[0205] 实施例 2: 珠蛋白特异性 ZFN 的活性

[0206] 使用靶向人类珠蛋白基因座或  $\beta$  样珠蛋白基因表达的调节子的 ZFN 对来测试这

些 ZFN 在特定靶位点诱导 DSB 的能力。所示 ZFN 的每个指的识别螺旋区域的氨基酸序列与全部靶位点一起示于下表 1A 中 (DNA 靶位点以大写字母指示 ; 非接触核苷酸以小写字母指示)。

[0207] 表 1A : 锌指核酸酶

[0208]

SBS #, 靶标	设计					
人类β血红蛋白特异性 ZFN						
	F1	F2	F3	F4	F5	F6
<b>SBS#33511</b> ggGCAGTAACGGC AGACttctcctca gg (SEQ ID NO:8)	DRSNLSR (SEQ ID NO:9)	QSSDLRR (SEQ ID NO:10)	RSDTLA (SEQ ID NO:11)	QSGALAR (SEQ ID NO:12)	QSGDLTR (SEQ ID NO:13)	N/A
<b>SBS#33533</b> tgGGGCAAGGTGA ACGTGGAtgaagt tg (SEQ ID NO:14)	QSAHRKN (SEQ ID NO:15)	LKHHLTD (SEQ ID NO:16)	QRSNLVR (SEQ ID NO:17)	TSGHLSR (SEQ ID NO:18)	QSNHLTE (SEQ ID NO:19)	RSHHLKA (SEQ ID NO:20)
<b>SBS#35256</b> agAGTCAGGTGCA CCATggtgtctgt tt (SEQ ID NO:21)	TNQRIT (SEQ ID NO:22)	DRSNRRT (SEQ ID NO:23)	RNASRTR (SEQ ID NO:24)	RSDNLSE (SEQ ID NO:25)	RSQHRKT (SEQ ID NO:26)	N/A
<b>SBS#35263</b> gtGGAGAAGTctG CCGTtactgcct gt (SEQ ID NO:27)	TSGLSR (SEQ ID NO:28)	DRSDLR (SEQ ID NO:29)	DRSALAR (SEQ ID NO:30)	QSSNLAR (SEQ ID NO:31)	QSGHLSR (SEQ ID NO:32)	N/A
<b>SBS#34770</b> acAGGAGTCAGGT GCACcatggtgtc tg (SEQ ID NO:33)	DQSNLRA (SEQ ID NO:34)	RNASRTR (SEQ ID NO:24)	RSDNLSE (SEQ ID NO:25)	RSQHRKT (SEQ ID NO:26)	RSDHLTQ (SEQ ID NO:35)	N/A
<b>SBS#34791</b> gaGAAGTctGCCG TTACTgocctgtg gg (SEQ ID NO:36)	ARSTRTN (SEQ ID NO:37)	TSGLSR (SEQ ID NO:28)	DRSDLR (SEQ ID NO:29)	DRSARTR (SEQ ID NO:38)	QSGNLAR (SEQ ID NO:39)	N/A
<b>SBS#34805</b> taACGGCAGACtT CTCCAcaggagtc ag (SEQ ID NO:40)	QSGDLTR (SEQ ID NO:13)	SSSDRKK (SEQ ID NO:41)	DRSNLSR (SEQ ID NO:9)	QSADRTK (SEQ ID NO:42)	RSDTLA (SEQ ID NO:11)	N/A
<b>SBS#34826</b> gcCCTGTGGGCA	LRHHLTR (SEQ ID	QSGNLHV (SEQ ID	RSAHLSR (SEQ ID	RSDVLST (SEQ ID	RKQDLRT (SEQ ID	N/A

[0209]

AGGTgaacgtgga tg (SEQ ID NO:43)	NO:44)	NO:45)	NO:46)	NO:47)	NO:48)	
<b>SBS#35301</b> ggGCAGTAAACGGC AGACttctcctca gg (SEQ ID NO:8)	DRSNLSR (SEQ ID NO:9)	QSGDLTR (SEQ ID NO:13)	RSDTLSA (SEQ ID NO:11)	QSGALAR (SEQ ID NO:12)	QSGDLTR (SEQ ID NO:13)	N/A
<b>SBS#35328</b> tgGGGCAAGGTGA ACGTggatgaagt tg (SEQ ID NO:14)	MSHHLRD (SEQ ID NO:49)	QRSNLVR (SEQ ID NO:17)	TSGHLSR (SEQ ID NO:18)	QSNHLTE (SEQ ID NO:19)	RSHHLKA (SEQ ID NO:20)	N/A
<b>SBS#35497</b> caCAGGGCAGTAA CgGCAGACTtctc ct (SEQ ID NO:50)	DRSNLSR (SEQ ID NO:9)	QSGDLTR (SEQ ID NO:13)	DRSNLSR (SEQ ID NO:9)	LKHHLTD (SEQ ID NO:16)	DRSHLTR (SEQ ID NO:51)	RSDNLRE (SEQ ID NO:52)
<b>SBS#35506</b> ggCAAGGTGAACG TGGAtgaagttgg tg (SEQ ID NO:53)	QSGHLAR (SEQ ID NO:54)	VSHHLRD (SEQ ID NO:55)	QSGNLAR (SEQ ID NO:39)	LRHHLTR (SEQ ID NO:44)	QSGNLHV (SEQ ID NO:45)	N/A
<b>β珠蛋白IVS.1</b>						
<b>SBS#43545</b> atCAAGGTTACAA GACAGGTttaagg ag (SEQ ID NO:158)	LRHHLTR (SEQ ID NO:44)	QSGTRKT (SEQ ID NO:153)	RSDNLST (SEQ ID NO:154)	DSANRIK (SEQ ID NO:155)	LRHHLTR (SEQ ID NO:44)	QSGNLHV (SEQ ID NO:45)
<b>SBS#43544</b> aaTCTGCCAGGG CCTCaccaccaac tt (SEQ ID NO:159)	AMQTLRV (SEQ ID NO:156)	DRSHLAR (SEQ ID NO:76)	RSDNLSE (SEQ ID NO:25)	ASKTRKN (SEQ ID NO:77)	RNSDRTR (SEQ ID NO:157)	N/A
<b>人类BCL11A特异性ZEN 外显子2</b>						
<b>SBS#39172</b> ctCCAGAAGGGGA TCATGACctctc ac (SEQ ID NO:160)	DRSNLSR (SEQ ID NO:9)	LRQNLIM (SEQ ID NO:161)	TSANLTV (SEQ ID NO:162)	RSDHLRS (SEQ ID NO:94)	QSGNLAR (SEQ ID NO:39)	QRNDRKS (SEQ ID NO:163)
<b>SBS#43490</b> ctCCAGAAGGGGA TCATGACctctc ac (SEQ ID NO:160)	DRSNLSR (SEQ ID NO:9)	LRQNLIM (SEQ ID NO:161)	LQSQLNR (SEQ ID NO:164)	RSDHLRS (SEQ ID NO:94)	QSGNLAR (SEQ ID NO:39)	QRNDRKS (SEQ ID NO:163)
<b>SBS#44642</b> ctCCAGAAGGGGA TCATGACctctc ac (SEQ ID NO:160)	DRANLSR (SEQ ID NO:165)	LRQNLIM (SEQ ID NO:161)	LQSQLNR (SEQ ID NO:164)	RSDHLRS (SEQ ID NO:94)	QSGNLAR (SEQ ID NO:39)	QRNDRKS (SEQ ID NO:163)
<b>SBS#45148</b> ctCCAGAAGGGGA TCATGACctctc ac (SEQ ID NO:160)	DRSNLSR (SEQ ID NO:9)	TSSNRNH (SEQ ID NO:166)	HSGNLTK (SEQ ID NO:167)	RSDHLRS (SEQ ID NO:94)	QSGNLAR (SEQ ID NO:39)	QKVDLSR (SEQ ID NO:168)
<b>SBS#45147</b> ctCCAGAAGGGGA TCATGACctctc ac (SEQ ID	DRSNLSR (SEQ ID NO:9)	TSSNRNH (SEQ ID NO:166)	QANNLKV (SEQ ID NO:169)	RSDHLRS (SEQ ID NO:94)	QSGNLAR (SEQ ID NO:39)	QKVDLSR (SEQ ID NO:168)

[0210]

NO:160)						
<b>SBS#39145</b> ccCAACGGGCCGT GGTCTGGttcatc at (SEQ ID NO:170)	RSDHLSA (SEQ ID NO:59)	DRSALAR (SEQ ID NO:30)	RSDSLSR (SEQ ID NO:171)	DRSVRTK (SEQ ID NO:172)	RSDHLSA{ SEQ ID NO:59)	QRSNLKV (SEQ ID NO:173)
<b>SBS#44490</b> ccCAACGGGCCGT GGTCTGGttcatc at (SEQ ID NO:170)	RSDHLTQ (SEQ ID NO:35)	DRSALAR (SEQ ID NO:30)	RSDSLSR (SEQ ID NO:171)	DRSVRTK (SEQ ID NO:172)	RSDHLSA{ SEQ ID NO:59)	QRSNLKV (SEQ ID NO:173)
<b>SBS#44489</b> ccCAACGGGCCGT GGTCTGGttcatc at (SEQ ID NO:170)	RSDHLTT (SEQ ID NO:174)	DRSALAR (SEQ ID NO:30)	RSDSLSR (SEQ ID NO:171)	DRSVRTK (SEQ ID NO:172)	RSDHLSA{ SEQ ID NO:59)	QRSNLKV (SEQ ID NO:173)
<b>SBS#45081</b> ccCAACGGGCCGT GGTCTGGttcatc at (SEQ ID NO:170)	RSDHLSA (SEQ ID NO:59)	WAFADR (SEQ ID NO:175)	RSDSLSR (SEQ ID NO:171)	HTKLSLR (SEQ ID NO:176)	RSDHLSA{ SEQ ID NO:59)	QRSNLKV (SEQ ID NO:173)
<b>SBS#44493</b> ccCAACGGGCCGT GGTCTGGttcatc at (SEQ ID NO:170)	RS AHLTQ (SEQ ID NO:177)	DRSVLRR (SEQ ID NO:178)	RSDSLSR (SEQ ID NO:171)	DRSVRTK (SEQ ID NO:172)	RSDHLSA{ SEQ ID NO:59)	QRSNLKV (SEQ ID NO:173)
<b>SBS#29527</b> atCCCATGGAGAG GTGGCTGggaagg ac (SEQ ID NO:56)	RSDVLSE (SEQ ID NO:57)	RNQRKT (SEQ ID NO:58)	RSDHLSA (SEQ ID NO:59)	RSANLTR (SEQ ID NO:60)	RSDVLSN{ SEQ ID NO:61)	DRSTRIT (SEQ ID NO:62)
<b>SBS#29528</b> atATTGCAGACAA TAACccctttaac ct (SEQ ID NO:63)	DRSNLSR (SEQ ID NO:9)	HRQHLVT (SEQ ID NO:64)	DRSNLTR (SEQ ID NO:65)	QSGDLTR (SEQ ID NO:13)	HRSSLN{ SEQ ID NO:66)	N/A
<b>SBS# 29525</b> caTCCAGGCCGTG GGGAtttagagctc ca (SEQ ID NO:66)	QSGHLSR (SEQ ID NO:32)	RSDHLST (SEQ ID NO:67)	RSADLSR (SEQ ID NO:68)	RSDNLSQ (SEQ ID NO:69)	ASNDRKK{ SEQ ID NO:70)	N/A
<b>SBS#29526</b> gtGCAGAAATATGC CCCGCAGggtatt tg (SEQ ID NO:71)	RSDNLSA (SEQ ID NO:72)	RNDRKT (SEQ ID NO:73)	DRSDLSR (SEQ ID NO:29)	TSSNRKT (SEQ ID NO:74)	QSGNLR{ SEQ ID NO:39)	QSGDLTR (SEQ ID NO:13)
<b>外显子4</b>						
<b>SBS#34678</b> atATTGCAGACAA TAACccctttaac ct (SEQ ID NO:179)	DRSNLSR (SEQ ID NO:9)	HRQHLVT (SEQ ID NO:64)	DRSNLTR (SEQ ID NO:65)	QSGDLTR (SEQ ID NO:13)	HRWLSN{ SEQ ID NO:180)	N/A
<b>SBS#34642</b> atCCCATGgAGAG GTGGCTGggaagg ac (SEQ ID NO:56)	RSDHLSQ (SEQ ID NO:99)	DSSHRTR (SEQ ID NO:181)	LRHHLTR (SEQ ID NO:44)	QSAHLKA (SEQ ID NO:182)	RSDVLSN{ SEQ ID NO:61)	DRSTRIT (SEQ ID NO:62)
<b>Bcl11a-XL</b>						
<b>SBS#44889</b> ctCACTGTCCACA	RSANLAR (SEQ ID	RLDNRTA (SEQ ID	QSNLNS (SEQ ID	WRSSLKT (SEQ ID	DRSNRKT (SEQ ID	N/A

[0211]



GGAGaagccacac gg (SEQ ID NO:183)	NO:184)	NO:185)	NO:186)	NO:187)	NO:188)	
<b>SBS#44888</b> ttGCTIACAGTTCT TGAAGACtttccc ac (SEQ ID NO:189)	DRSNLSR (SEQ ID NO:9)	QSGNLR (SEQ ID NO:39)	YKHLSD (SEQ ID NO:190)	TSGSLTR (SEQ ID NO:191)	QSGDLTR (SEQ ID NO:13)	LKDTLRR (SEQ ID NO:192)
<b>SBS#44905*</b> gaGAAGCCACACG GGCGAAAggcctt at (SEQ ID NO:193)	QSGNLDS (SEQ ID NO:194)	RSADLSR (SEQ ID NO:68)	RSDHLE (SEQ ID NO:78)	QNATRIN (SEQ ID NO:195)	WNSDLRK (SEQ ID NO:196)	QSGNLR (SEQ ID NO:39)
<b>SBS#44904*</b> tgGACAGTGAGAT TGCTacagttctt ga (SEQ ID NO:197)	QSSDLR (SEQ ID NO:88)	YKWTLRN (SEQ ID NO:198)	RSANLTR (SEQ ID NO:60)	TSTKLRT (SEQ ID NO:199)	DRSNLTR (SEQ ID NO:65)	N/A
<b>SBS#44911</b> gcCACACGGGCGA AaGGCCTTataaa tg (SEQ ID NO:200)	AMQTLRV (SEQ ID NO:156)	DRSHLAR (SEQ ID NO:76)	QRSNLVR (SEQ ID NO:17)	DRSHLAR (SEQ ID NO:76)	RSDTLST (SEQ ID NO:201)	DSSNRIN (SEQ ID NO:202)
<b>SBS#44910</b> ctCCTGTGGACAG TGAGATTgetaca gt (SEQ ID NO:203)	NDFLYL (SEQ ID NO:204)	RSANLTR (SEQ ID NO:60)	TSTKLRT (SEQ ID NO:199)	DRSNLTR (SEQ ID NO:65)	RSDSLSV (SEQ ID NO:205)	HNDRKRN (SEQ ID NO:206)
<b>SBS#44945</b> aaGCTCACCCAGGC ACATGAAAacgca tg (SEQ ID NO:207)	QSGNLR (SEQ ID NO:39)	CRQNLAN (SEQ ID NO:208)	YQVLTR (SEQ ID NO:209)	RSDNLRE (SEQ ID NO:52)	DRSNRTT (SEQ ID NO:23)	HRSSLRR (SEQ ID NO:210)
<b>SBS#44944</b> ctACTCTGgGCAC AGGCATAGttgca ca (SEQ ID NO:211)	RSDNLST (SEQ ID NO:154)	QSSDLRR (SEQ ID NO:10)	RSDALSE (SEQ ID NO:212)	QNATRTK (SEQ ID NO:115)	RSDTLSE (SEQ ID NO:84)	ARSTRTN (SEQ ID NO:37)
<b>SBS#44947**</b> ctCACCAGGCACA TGAAAACgcatgg cc (SEQ ID NO:213)	GSSALTO (SEQ ID NO:214)	QSGNLR (SEQ ID NO:39)	TASHLKE (SEQ ID NO:215)	QNATRTK (SEQ ID NO:115)	RSDNLSE (SEQ ID NO:25)	SSRNLAS (SEQ ID NO:216)
<b>SBS#44946**</b> ctACTCTGgGCAC AGGCATAGttgca ca (SEQ ID NO:211)	RSDNLST (SEQ ID NO:154)	QSSDLRR (SEQ ID NO:10)	RSDALSE (SEQ ID NO:212)	QNATRTK (SEQ ID NO:115)	RSDTLSE (SEQ ID NO:84)	ARSTRTN (SEQ ID NO:37)
<b>人类KLF1特异性ZEN</b>						
<b>KLF-外显子1</b>						
<b>SBS#36004</b> ggGAAGGGGCCCA GGGCGGTcagtg gc (SEQ ID NO:75)	TSGHLR (SEQ ID NO:18)	DRSHLAR (SEQ ID NO:76)	RSDNLSQ (SEQ ID NO:69)	ASNDRKK (SEQ ID NO:70)	RSDHLE (SEQ ID NO:78)	QSGNLR (SEQ ID NO:39)
<b>SBS#36021</b> acACACAGGATGA Cttcctcaaggtg gg (SEQ ID NO:79)	DRSNLTR (SEQ ID NO:65)	TSANLSR (SEQ ID NO:217)	RSDHLE (SEQ ID NO:78)	QSASRKN (SEQ ID NO:81)	NA	NA

[0212]

<b>SBS#33237</b> ggGAAGGGGCCCA GGGCGGtcaagtgt gc (SEQ ID NO:75)	TSGHLR (SEQ ID NO:18)	DRSHLR (SEQ ID NO:76)	RSDNLSE (SEQ ID NO:25)	ASKTRKN (SEQ ID NO:77)	RSDHLE (SEQ ID NO:78)	QSGNLR (SEQ ID NO:39)
<b>SBS#33238</b> acACACAGGATGA Cttcctcaagtg gg (SEQ ID NO:79)	DRSNLRS (SEQ ID NO:9)	TSGNLTR (SEQ ID NO:80)	RSDHLE (SEQ ID NO:78)	QSASRKN (SEQ ID NO:81)	N/A	N/A
<b>SBS#33257</b> cgCCACCGGGCTC CGGGcccgaagaag tt (SEQ ID NO:82)	RSAHLR (SEQ ID NO:46)	DSSDRKK (SEQ ID NO:83)	DRSHLR (SEQ ID NO:76)	RSDTLSE (SEQ ID NO:84)	QSGDLTR (SEQ ID NO:13)	N/A
<b>SBS#33258</b> ccCCAGACcTGCG CTCTGGCGcccag cg (SEQ ID NO:85)	RSDSLR (SEQ ID NO:86)	RLDWLPV (SEQ ID NO:87)	QSSDLR (SEQ ID NO:88)	AASNRSK (SEQ ID NO:89)	DRSNLRS (SEQ ID NO:9)	QSGDLTR (SEQ ID NO:13)
<b>SBS#33269</b> ggCTCGGGgGCGG GGGCTGGAgccag gg (SEQ ID NO:90)	QSSHLTR (SEQ ID NO:91)	QSSDLTR (SEQ ID NO:92)	RSDHLE (SEQ ID NO:78)	HSRTRTK (SEQ ID NO:93)	RSDHLRS (SEQ ID NO:94)	DRSARNS (SEQ ID NO:95)
<b>SBS#33270</b> aaGGCGCTGGCGC TgCAACCGgtata cc (SEQ ID NO:96)	RSDTLSE (SEQ ID NO:84)	QSHNRTK (SEQ ID NO:97)	QSSDLR (SEQ ID NO:88)	DRSHLR (SEQ ID NO:76)	QSSDLRS (SEQ ID NO:88)	DRSHLR (SEQ ID NO:76)
<b>SBS#33271</b> ttGCAGCGCCAGC GCCTTGGgctcgg gg (SEQ ID NO:98)	RSDHLSQ (SEQ ID NO:99)	HRSSLGD (SEQ ID NO:100)	RSDDLTR (SEQ ID NO:101)	QRSTLSS (SEQ ID NO:102)	RSADLTR (SEQ ID NO:103)	QSGDLTR (SEQ ID NO:13)
<b>SBS#33272</b> cgGTGTACCCGGG GCCCGggcggc tc (SEQ ID NO:104)	DRSDLSR (SEQ ID NO:29)	RSTHLVR (SEQ ID NO:105)	RSDSLST (SEQ ID NO:106)	DSSDRTK (SEQ ID NO:107)	RSALARS (SEQ ID NO:108)	N/A
<b>KLF-外显子2</b>						
<b>SBS#36071</b> ggTGAGGAGGAGA TCCAggtcccagg tg (SEQ ID NO:218)	NNRDLIN (SEQ ID NO:219)	TSSNLSR (SEQ ID NO:220)	QSGHLR (SEQ ID NO:32)	QSGHLAR (SEQ ID NO:54)	QRTHLNS (SEQ ID NO:221)	N/A
<b>SBS#36085</b> ctTCTCGGGCCCCG GaCCCCGgtggcg cg (SEQ ID NO:222)	RSDHLE (SEQ ID NO:78)	HSRTRTK (SEQ ID NO:93)	RSDHLE (SEQ ID NO:78)	HSRTRTK (SEQ ID NO:93)	RSDHLE (SEQ ID NO:78)	RKSDRIK (SEQ ID NO:223)

[0213]

人类7珠蛋白5'调节区ZFN						
调节区(-175)						
<b>SBS#34360</b> ttGCATTGAGATA GTGTGGGgaagg gc (SEQ ID NO:109)	RSDHLSV (SEQ ID NO:110)	RSDVRKT (SEQ ID NO:111)	RSDYLSK (SEQ ID NO:112)	TSSVRTT (SEQ ID NO:113)	RPYTLRL (SEQ ID NO:114)	QNATRTK (SEQ ID NO:115)
<b>SBS#34363</b> atCTGTCTGAAAC GGTCcoctggctaa ac (SEQ ID NO:116)	DRSALAR (SEQ ID NO:30)	RRDILHQ (SEQ ID NO:117)	QSGNLAR (SEQ ID NO:39)	LAYDRRK (SEQ ID NO:118)	RSDVLSE (SEQ ID NO:57)	N/A
<b>SBS#34398</b> ttTGCATTGAGAT AGTgtggggaagg gg (SEQ ID NO:119)	RSDSLLR (SEQ ID NO:86)	QSCARNV (SEQ ID NO:120)	RSDNLAR (SEQ ID NO:121)	HRNTLLG (SEQ ID NO:122)	MRNRLNR (SEQ ID NO:123)	N/A
<b>SBS#34400</b> ctGTCTGAaACGG TCcCTGGCTaaac tc (SEQ ID NO:124)	QSSDLR (SEQ ID NO:88)	RRDALLM (SEQ ID NO:125)	DRSALAR (SEQ ID NO:30)	RRDILHQ (SEQ ID NO:117)	QNAHRKT (SEQ ID NO:126)	DRSALAR (SEQ ID NO:30)
<b>SBS#31160</b> taTTGCATtTGAG ATAGTGTGgggaa gg (SEQ ID NO:127)	RSDSLLR (SEQ ID NO:86)	LQHHLTD (SEQ ID NO:128)	TSGNLTR (SEQ ID NO:80)	TSTHLHI (SEQ ID NO:129)	QSGDLTR (SEQ ID NO:13)	HKWVLRQ (SEQ ID NO:130)
<b>SBS#34365</b> ctGTCTGAaACGG TCcCTGGCTaaac tc (SEQ ID NO:124)	QSSDLR (SEQ ID NO:88)	RRDALLM (SEQ ID NO:131)	DRSALAR (SEQ ID NO:30)	RRDILHQ (SEQ ID NO:117)	QNAHRKT (SEQ ID NO:126)	DRSALAR (SEQ ID NO:30)
调节区(-110)						
<b>SBS#34539</b> tgGTCAAGGCAAG GCTGgccaaccca tg (SEQ ID NO:224)	RSDVLSE (SEQ ID NO:57)	RNQHRKT (SEQ ID NO:58)	QSGDLTR (SEQ ID NO:13)	RSDHLST (SEQ ID NO:67)	DRSALAR (SEQ ID NO:30)	NA
<b>SBS#34574</b> gcCTTGACAAGGC AAACttgaccaat ag (SEQ ID NO:225)	DRSNRTT (SEQ ID NO:23)	QSGSLTR (SEQ ID NO:226)	RSDNLSV (SEQ ID NO:227)	DRSNLSR (SEQ ID NO:9)	LKFALAN (SEQ ID NO:228)	NA
<b>SBS#43865</b> gcCTTGACAAGGC AAACttgaccaat ag (SEQ ID NO:225)	NPANLTR (SEQ ID NO:229)	QNATRTK (SEQ ID NO:115)	RSDNLSV (SEQ ID NO:227)	DRSNLSR (SEQ ID NO:9)	LKFALAN (SEQ ID NO:228)	NA
<b>SBS#43852</b> tgGTCAAGGCAAG GCTGgccaaccca tg (SEQ ID NO:224)	RSDVLSE (SEQ ID NO:57)	RNQHRKT (SEQ ID NO:58)	QSGDLTR (SEQ ID NO:13)	RSDNLST (SEQ ID NO:154)	DSSARKK (SEQ ID NO:230)	NA

[0214] 注释:用单星号(\*)或用双星号(\*\*)标记的BCL11A XL特异性ZFN对分别含有新接头L7a和L8p。参见实施例6。

[0215] 使用如Perez等,(2008)Nat. Biotechnol. 26:808-816和Guschin等,(2010)Methods Mol Biol. 649:247-56)中所述的Cel-I测定法(Surveyor<sup>TM</sup>, Transgenomics)来检测K562或HSC中靶基因的ZFN诱导修饰。在这种测定法中,靶位点的PCR扩增之后,使用错配检测酶Cel-I(Yang等,(2000)Biochemistry 39:3533-3541)定量插入和缺失(indel),其提供DSB频率的下限估计值。ZFN表达载体在标准条件(37℃)下或使用低温

冲击 (30°C, 参见共同拥有的美国专利公布号 20110041195) 转染后的三天, 使用 DNeasy 试剂盒 (Qiagen) 从 K562 细胞中分离基因组 DNA。

[0216] 来自 *Cel-I* 测定法的结果证实 ZFN 能够诱导在其相应的靶位点处的裂解 (也参见共同拥有的美国临时申请号 61/556, 691)。结果示于图 1 中并且指示在  $\beta$  珠蛋白基因中的大部分靶基因座存在活性蛋白。

[0217] 实施例 3:  $\beta$  珠蛋白基因座的编辑

[0218] 如下使用人类  $\beta$  珠蛋白基因 (HBB) 特异性 ZFN (表 1) 将供体 DNA 引入  $\beta$  珠蛋白基因座。供体 DNA 被设计成使得编码 HBB 基因序列的序列侧接与围绕  $\beta$  珠蛋白基因中的 ZFN 裂解位点的区域同源 (同源臂) 的序列。所述同源臂长度为约 500-600 个碱基对。HBB 供体序列缺乏任何非编码序列以使得当插入  $\beta$  珠蛋白靶位点中时, 通过  $\beta$  珠蛋白启动子和任何其它  $\beta$  珠蛋白调节序列来调节供体的表达。当插入时, HBB 供体与内源珠蛋白序列同框融合并产生融合蛋白。另外, HBB 供体寡核苷酸 (oligo) 被设计成用于在 ZFN 处理后捕捉至裂解 HBB 基因中。所述寡核苷酸含有限制位点以使得在插入寡核苷酸后, 新限制位点被引入 HBB 基因中, 其随后可能裂解。

[0219] 如图 2A 中所示, 将  $\beta$  珠蛋白寡核苷酸供体插入适当的基因座中, 如通过存在于供体 DNA 上的新限制位点的存在所证实。此外, 如图 2B 中所示, *Cel-I* 分析显示若干个 ZFN 对能够裂解 DNA, 但寡核苷酸仅存在于泳道 8 中的样品中。

[0220] 为了使转基因 CD34<sup>+</sup> 细胞分化成成熟的 RBC, 使用本领域中已知的方法。例如, 使用 Ficoll-Paque (GE Healthcare) 和 CD34<sup>+</sup> 微珠 (Miltenyi Biotec) 根据制造商的说明纯化 SCD CD34<sup>+</sup> 细胞。在生长因子存在下将 CD34<sup>+</sup> 细胞与 BIT95000 (StemCell Technologies) 一起在 Iscove 的 MDM 中培养。使用 2 阶段液体培养模型将细胞分化成红系谱系。在前 6 天期间 (第一阶段), 用 SCF (100ng/ml)、Flt3-L (100ng/ml) 和 IL-3 (20ng/ml) 使 CD34<sup>+</sup> 细胞扩增。然后将扩增细胞定型并用 Epo (2U/ml) 和 SCF (50ng/ml) 分化成红系谱系 (第二阶段)。参见 Giarratana 等, (2011) Blood 118(19):5071-9。

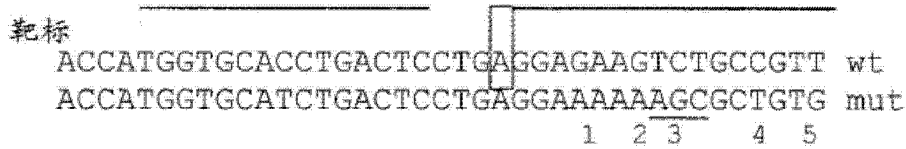
[0221] 实施例 4:  $\beta$  珠蛋白中的突变的基因校正。

[0222] 为了校正镰状  $\beta$  珠蛋白基因中的人类镰状细胞突变, 在  $\beta$  珠蛋白基因座中用 ZFN 进行双链断裂, 接着使用外源校正型寡核苷酸作为模板 (“供体寡核苷酸”) 进行 DNA 修复。为了避免核酸酶裂解校正的珠蛋白基因的可能性 (其中供体寡核苷酸已引导内源 HBB 基因中的镰状突变的校正), 供体寡核苷酸被设计成通过翻译方式将沉默突变共同引入 HBB 编码序列中, 以使得校正的等位基因将缺乏 ZFN 靶序列之一。以这种方式, 将观察到所需基因校正等位基因的频率增加。为了设计最佳的寡核苷酸供体, 研究 ZFN 靶序列中的若干突变以及同源臂的长度。

[0223] 下文示出围绕镰状突变的序列并且各种突变以数字指示。因此, 突变 1 = G 至 A 变化, 突变 2 = G 至 A, 突变 3 = TCT 至 AGC, 突变 4 = C 至 T 以及突变 5 = T 至 G。产生包含突变的各种组合的寡核苷酸。野生型序列 (“wt”) 示于顶部 (SEQ ID NO:231) 并且具有突变的序列 (“mut”) 示于下文 (SEQ ID NO:232)。

[0224] 核酸酶的靶位点 (“靶标”) 由粗线指示, 并且镰状突变的位点被加上方框。根据突变标记寡核苷酸, 因此, 例如, 寡核苷酸 SMS1 仅存在沉默突变位点 1, 而 SMS 124 存在沉默突变位点 1、2 和 4。

[0225]



[0226] 将各种寡核苷酸以‘正义’或正向链（指示为‘F’）或‘反义’或反向链（指示为‘R’）形式的单链分子形式递送至 CD34+ 细胞。经由用具有或不具有核酸酶的 BTX ECM 830 Square Wave 装置转染来递送寡核苷酸。除非另外说明，否则递送 3 μg 核酸酶。通过 HBB 基因的 PCR 扩增子的高通量 DNA 测序来测量基因编辑。指示通过非同源末端接合（“NHEJ”，由 ZFN 诱导裂解后的 DNA 中的双链断裂的修复造成）或在 ZFN 裂解后的寡核苷酸的靶向整合（“基因校正”）的基因修饰百分比（参见图 9）。结果指示突变的一些组合能够增强细胞中的基因校正以使得至多 20% 的细胞在镰状基因座显示基因校正。

[0227] 为了研究同源臂长度对基因校正百分比的影响，使用在镰状突变位点的任一侧上具有 41 和 46 个核苷酸（88bp 供体寡核苷酸）或 50 和 50 个同源核苷酸（101bp 供体寡核苷酸）的 SMS12 和 SMS124 寡核苷酸。结果（参见图 10）指示较长同源臂对于引起基因校正更有效，其中至多 40% 的等位基因并有由寡核苷酸指定的变化。所用的寡核苷酸示于下文：

[0228] SMS124, 88bp, R (SEQ ID NO :233) :

[0229] 5' CGTTCACCTGCCCCACAGGGCAGTAACAGCAGATTTTCCTCAGGAGTCAGGTGCACCA

[0230] TGGTGTCTGTTTGAGGTTGCTAGTGAAC

[0231] SMS12, 88bp, R (SEQ ID NO:234) :

[0232] 5' CGTTCACCTGCCCCACAGGGCAGTAACAGCAGATTTTCCTCAGGAGTCAGGTGCACCA

[0233] TGGTGTCTGTTTGAGGTTGCTAGTGAAC

[0234] SMS124, 101bp, R (SEQ ID NO :235) :

[0235] 5' CGTTCACCTGCCCCACAGGGCAGTAACAGCAGATTTTCCTCAGGAGTCAGGTGCACCA

[0236] GGTGCACCATGGTGTCTGTTTGAGGTTGCTAGTGAACACAG

[0237] 为了研究在 CD34+ 细胞分化期间的分化能力和基因校正的持久性，使用 Stemcell Technologies 的 Methocult 甲基纤维素培养基根据制造商的说明诱导 ZFN 修饰的 CD34+ 细胞的库进行分化。通过测定由 Methocult 诱导的分化产生的集落类型来分析分化：红系集落形成单位（“CFU-E”）；红系爆发集落形成单位（“BFU-E”）；粒细胞 / 巨噬细胞集落形成单位（“CFU-GM”）和粒细胞 / 红细胞 / 单核细胞 / 巨噬细胞集落形成单位（“CFU-GEMM”）。结果指示 ZFN 处理的细胞保持与模拟转染细胞相同的分化能力。从板上挑取单独的 BFU-E 集落并且在 HBB 进行基因分型。结果指示在集落分化期间维持 ZFN 诱导的修饰（参见图 11）。此外，修饰的 BFU-E 集落的频率与起始库中修饰的等位基因的频率类似，证实在 BFU-E 形成期间针对编辑的细胞无偏差。另外，在液态培养体外红血细胞分化过程中，针对基因修饰测定整个细胞群体。在整个至少 18 天红血细胞分化过程中，修饰是稳定的（参见图 12）。

[0238] 与 β 地中海贫血相关的 β 珠蛋白基因中的另一个常见突变被称为 IVS1.1。这种 G→A 突变位于 β 珠蛋白基因的内含子 1 的第一碱基对内，并且其在基因中的存在导致 β 珠蛋白前体 mRNA 的有缺陷的剪接。因此，将一对 ZFN 工程改造以识别并裂解区域，基本上出于模型目的概括这种突变。对这些 ZFN 的测试发现，它们能够裂解 β 珠蛋白基因中的位

点,在 CD34+ 细胞中产生 52.63% NHEJ。

[0239] 实施例 5 :将  $\beta$  珠蛋白供体插入安全港基因座中

[0240] 为了将野生型  $\beta$  珠蛋白基因插入安全港基因座中,以使得来自转基因的表达将校正 HSC 中缺少的  $\beta$  珠蛋白,将对所述安全港基因座具有特异性的核酸酶与供体核酸一起引入细胞中。将对 HPRT (参见共同拥有的美国专利公布号 20130137104 和 20130122591)、AAVS1 (参见美国专利 8,110,379)、CCR5 (参见美国专利 7,951,925) 或  $\beta$  珠蛋白 (参见表 1A) 具有特异性的核酸酶引入患者来源的 CD34+ 干细胞中。可通过本领域中已知的任何方法引入,例如 mRNA 电穿孔。供体 DNA 被设计成含有转基因、野生型  $\beta$  珠蛋白和侧接与围绕安全港靶标的区域具有足够同源性的转基因的同源性区域以允许 HDR (在每一侧上通常 500bp)。可选地,可提供供体构建体,无论是缺乏还是含有同源性区域,经由末端捕捉将其整合至 ZFN 或 TALEN 靶向基因座 (参见美国申请号 13/889,162)。在引入 ZFN 之前、期间或之后,将供体共同引入 CD34+ 细胞中。将经过修饰的 CD34+ 细胞再次引入患者中,并且在植入后,产生足够水平的  $\beta$  血红蛋白以允许产生治疗相关量的血红蛋白。

[0241] 实施例 6 :BCL11A 和 KLF1 的失活

[0242] 将对 BCL11A 和 KLF1 具有特异性的核酸酶 (例如, ZFN, 如表 1A 中所示) 引入如上所述的 HSC 中以造成  $\gamma$  珠蛋白表达上调 (参见图 3) 并且通过如上所述的 Cel 1 测定法 (Perez 等, (2008), 同上) 分析细胞的基因组。

[0243] 如图 4 中所示,在用所示 KLF1 特异性 ZFN 处理 HSC 后,所述 ZFN 成功地修饰 KLF1 基因座 (图 4C 和 4D)。同样, BCL11A 特异性 ZFN 修饰 BCL11A 基因座 (图 4A)。将一对靶向 HPRT 基因座的 ZFN (参见共同拥有的美国临时申请 61/552,309) 用作对照物并且还证实成功的裂解 (图 4B)。比较 CD34+ 细胞转导后第 3 天与第 17 天分化培养的信号 (图 4E),证实基因编辑百分比 (% NHEJ) 随时间变化是稳定的。在图 4E 中所示的每一凝胶中,缺乏鉴定的泳道是阴性对照。

[0244] 类似地测试靶向 BCL11A 外显子 2 或外显子 4 的另一对 ZFN。对于这些研究,通过如先前所述的 Amaxa 将候选 ZFN 对引入 K562 细胞或引入 CD34+ 细胞。对于 CD34+ 转导,使用具有 2mm 间隙比色皿的 BTX ECM830 装置。使用 mMessageMachine T7Ultra 试剂盒 (#AM1345, Ambion) 制备来自细胞的 mRNA。使人类 CD34+ 细胞与 1xCC110 (Stem cell Technology) 在非组织培养处理板中在 x-vivo10 培养基 (Lonza) 中生长。在室温下对细胞计数并通过在 1200rpm 下离心 10 分钟来收集。用室温 PBS 洗涤细胞 1-2 次。每次转染使用 200,000 个细胞,并且将它们再次悬浮在 100  $\mu$  L BTEXpress 溶液中。每次转染加入 2-4  $\mu$  g mRNA 并且将混合物转移至比色皿。在转移后不久,在 250V 下将混合物电穿孔 5 毫秒。将预升温培养基加入比色皿并且将培养基与细胞转移至 48 孔非组织培养处理板,然后在 37 $^{\circ}$ C 下孵育。

[0245] 在指定数目的天数后,然后使用 Illumina MiSeq 对细胞进行基因组分析。为了定量所编辑等位基因的百分比,使用加入标准的 Illumina 测序接头序列的引物对目标基因组区域进行 PCR 扩增。进行第二组 13 轮的 PCR 以向两端添加条码和桥联接头序列。根据制造商的扩增子测序方案对 Illumina MiSeq 进行测序。MiSeq 产生配对末端读数,使用标准的比对软件将其合并并且进行接头修整。然后使用自定义脚本经由条码序列对由样品多路分离读数。然后经由实施 Needleman-Wunsch 算法 (Needleman, Saul B.; 和 Wunsch, Christian D. (1970). Jour Mol Bio 48(3):443-53) 将扩增子序列与参考序列进行整体比

对。将比对中的间隙或插入计算为% NHEJ 事件,并且与未被处理的对照样品序列进行比较以测定序列特异性背景率。

[0246] 对于靶向整合的计算,经由 biopython 实施 Needleman-Wunsch 算法 (Needleman, Saul B. ;和 Wunsch,Christian D. 同上) 来整体比对扩增子序列与参考序列。对经由实验处理产生的序列变化进行搜寻、计数并与对照样品中的计数进行比较。在这个过程中可能遮蔽已知单特征多态性 (SNP) 并且从进一步计数中排除 (例如,接近于 ZFN 靶位点的 1-bp 缺失 SNP)。通过测定含有插入或缺失的序列的百分比来计算 NHEJ% (也称为 indel)。使用仅用 GFP 载体处理的样品来评估 PCR 和基于测序误差的插入和缺失的背景频率。观察到小于 1% 的背景频率。

[0247] 代表性数据集示于下表 1B 中并且证实这些核酸酶蛋白可有效裂解其靶标。另外,在一些核酸酶处理细胞中监测到  $\gamma$  珠蛋白的表达。为了进行这种分析,根据标准程序使用实时 RT-qPCR (“Taqman”) (参见下文)。来自代表性数据集的结果显示为与 GFP 处理对照细胞相比的  $\gamma$  珠蛋白表达的倍增。 $\gamma$  值计算为  $\gamma$  珠蛋白与  $\alpha$  珠蛋白的比率,因此下文显示的任何观察到的增加代表核酸酶处理细胞中  $\gamma$  与  $\alpha$  的比率相比于 GFP 载体处理细胞中  $\gamma$  与  $\alpha$  的比率的增加。

[0248] 表 1B :BCL11A 外显子 2 和外显子 4ZFN 对的活性

[0249]

靶标	ZFN 对	% indel , K562	% indel, CD34+	$\gamma$ mRNA 倍增
外显子 2	39145/39172		69.78	3.65X
	39145/43490	19.88		nd
	39145/44642	38.52		nd
	39145/41548	42.26		nd
	39145/41547	35.63		nd
	44490/39172	29.38		nd
	44489/39172	24.34		nd
	45081/39172	27.80		nd
	44493/39172	25.68		nd
外显子 4	34678/34642		82.24	3.52X

[0250] 也对 BCL11A 的外显子 2 和外显子 4 区域做出 TALEN。使用规范的 TALE 编码和 ‘+17’ TALEN 主链,如先前所述构建 TALEN (参见共同拥有的美国专利公开 20110301073)。表 1C 示出 TALEN 的靶序列以及 DNA 结合结构域中的 RVD 序列。

[0251] 表 1C :针对 BCL11A 的 TALEN 对

[0252]

SBS编号 (外显子)	靶序列5'→3'	RVD序列(N→C)
101291 (外显子2)	ctGTGGGGCAGTGCCAGATga (SEQ ID NO:236)	NN-NG-NN-NN-NN-HD-NI-NN-NG-NN-HD- HD-NI-NN-NI-NG (SEQ ID NO:237)
101292 (外显子2)	ctCGATAAAAAATAAGAATgt (SEQ ID NO:238)	HD-NN-NI-NG-NI-NI-NI-NI-NI-NG-NI- NI-NN-NI-NI-NG (SEQ ID NO:239)
101301 (外显子4)	atGTCCTTCCCAGCCACCTct (SEQ ID NO:240)	NN-NG-HD-HD-NG-NG-HD-HD-HD-NI-NN- HD-HD-NI-HD-HD-NG (SEQ ID NO:241)
101304 (外显子4)	gtTAAAGGGGTTATTGTct (SEQ ID NO:242)	NG-NI-NI-NI-NN-NN-NN-NN-NG-NG-NI- NG-NG-NN-NG (SEQ ID NO:243)

[0253] 上文示出的 TALEN 对被引入细胞中并显示裂解活性。如通过在 K562 细胞中的 Cel-1 测定法所测量,101291/101292 对得到 0.8% indel 的值。TALEN 对 101301/101304 在 CD34+ 细胞中得到 35.7% indel 形成值,并且通过上述 RT-PCR 测定法测得以诱导 γ 珠蛋白 mRNA 表达增加约 2.31 倍。

[0254] 也做出 ZFN 对以靶向 BCL11A-XL 剪接变体的 ‘XL’ 部分。在 K562 细胞中测试这些蛋白并且代表性数据集示于下表 1D 中。BCL11A 的 ‘XL’ 同种型含有 3 个另外的天然锌指 (指 4-6),因此所采取的方法涉及破坏这个区域中的 BCL11A 基因以造成潜在锌指 4、5 和 / 或 6 以及其组合的伸展 (XL 区域内的编号 1 至 3)。ZFN 也被工程改造以避免相关 BCL11B 基因序列的裂解。一个 ZFN 对 44888/44889 靶向 BCL11A 的第四锌指,而两个对 44904/44905 和 44910/44911 靶向第四指的上游 (XL 区域内的编号 1),而两个其它对 44946/44947 和 44945/44944 靶向第五指 (XL 区域内的编号 2)。在 K562 细胞中测试这些蛋白并且代表性数据集示于下表 1D 中。两个 ZFN 对在 ZFP DNA 结合结构域与 FokI 核酸酶结构域之间含有新接头序列。44904/44905 对都含有 L7a 接头序列 (参见美国专利申请号 20090305419) 并且 44946/44947 对都含有 L8p 接头序列,它们都示于下文。也参见美国临时申请号 61/871, 219 :

[0255] L7a:HTK1HLRGSQLVKSKSEAAAR (SEQ ID NO :244)

[0256] L8p :HTKIHLRGSYAPMPPLALASP (SEQ ID NO :245)。

[0257] 表 1D :对 BCL11A XL 具有特异性的 ZFN 对的活性

[0258]

ZFN 对	% indel, K562
44889/44888	35.14
44905/44904	25.45
44911/44910	36.43
44945/44944	24.03
44947/44946	34.22

[0259] 然后在 CD34+ 细胞中测试 BCL11A XL 对并且其具活性。γ 珠蛋白的表达的测量结



果证实 BCL11A XL 的修饰导致  $\gamma$  珠蛋白表达相对于  $\alpha$  珠蛋白增加。

[0260] 在 CD34+ 细胞中测试其他对的 KLF1 特异性 ZFN 的活性, 并且对这些细胞分析  $\gamma$  珠蛋白表达的任何变化。代表性数据集示于下表 1E 中。

[0261] 表 1E :KLF 特异性 ZFN 对的活性

[0262]

靶标	ZFN 对	% indel, CD34+	$\gamma$ mRNA 倍增
KLF 外显子 1	36004/36021	44.4	2.2X
KLF 外显子 2	36071/36085	22.6	3.17X

[0263] 通过 Taqman 分析在多个时间点直至 ZFN 引入后 17 天测定在 HSC 中处理 BCL11A 或 KLF1 特异性核酸酶后编码  $\gamma$  珠蛋白和  $\beta$  珠蛋白的 mRNA 的比率, 并且也将  $\beta$  珠蛋白 mRNA 水平归一化为 18S rRNA 水平。在已经用 BCL11A 或 KLF1 特异性核酸酶处理的那些细胞中,  $\gamma$  珠蛋白表达水平增加 (图 5)。遵循由制造商 (Applied Biosystems) 提供的方案并使用基因特异性测定法, 利用标准的 Taqman 分析进行分析。

[0264] 也分析 BCL11A ZFN 修饰细胞以测定其中一个等位基因被 ZFN 修饰 (“Bb”) 的细胞群体与其中两个等位基因被 ZFN 修饰 (“敲除”) 和野生型修饰 (“BB”) 的细胞之间的  $\gamma / \beta$  mRNA 比率。

[0265] 如图 6 中所示, 不同细胞之间的  $\gamma / \beta$  mRNA 比率是不同的, 其中 BCL11A 敲除仅发生在一个等位基因 (Bb, 从左侧起 6-10 条) 或其中两个等位基因已经被敲除 (敲除, 最右侧 5 条, 从左侧起 11-15 条), 并且两种细胞的库不同于野生型 (BB, 前 5 条)。

[0266] 实施例 7 :  $\gamma$  珠蛋白基因的调节区的修饰

[0267] 在增加  $\gamma$  珠蛋白表达的另一种方法中, 在  $\gamma$  珠蛋白基因的调节区中做出突变以模拟 HPFH 突变 (参见图 9)。下文示出相对于  $\gamma$  珠蛋白基因中的 ATG 从 -202 至 -102 的区域。在这个序列上是指示已显示与 HPFH 相关的区域的灰色方框, 以及下划线序列, 其在缺失时也已经与 HPFH 相关 (参见 A Syllabus of Thalassemia Mutations(1997) Titus H. J. Huisman, Marianne F. H. Carver, 和 Erol Baysal, 由 Augusta, GA, USA 的 Sickle Cell Anemia Foundation 出版. 版权 © 1997 Titus H. J. Huisman) :

[0268]

-202

GCCTGCCTTACCAATGGCCTTGAC(SEQ ID NO:132)

GGCCTTGCCCTTACCAATGGCCTTGAC

-102

[0269] 核酸酶是如实施例 1 中所述设计并且示于表 1A 中以结合在这些 HPFH 相关突变的区域中以诱导野生型区域中的突变。通过 Cel I 分析 (参见 Perez 等 (2008), 同上) 在 K562 细胞中检测到的编辑等位基因百分比 (% NHEJ) 示于下表 2 中。另外, 如上所述在 CD34+ 细胞中测试一些对并且通过如上所述的 MiSeq 测序进行分析。对于一些对, 分析细胞的  $\gamma$  珠

蛋白表达的任何变化。下表 2 示出代表性数据集：

[0270] 表 2:通过  $\gamma$  珠蛋白特异性 ZFN 对进行编辑

[0271]

ZFN 对 (位置)	% NHEJ K562	% NHEJ	$\gamma$ mRNA 倍增
------------	----------------	-----------	---------------------

[0272]

		CD34+	
34360/34363 (-175)	39		
34398/34400 (-175)	54		
31160/34365 (-175)	53	45.22	1.63X
34539/34574 (-110)		45.71	5.38X
43865/43852 (-110)		56.13	

[0273] 在这种测定法中测试的前三对靶向  $\gamma$  启动子区域中 -175 周围的区域,而最后两个靶向  $\gamma$  珠蛋白启动子中的 -110 区域。

[0274] 已经编辑的 K562 细胞中的  $\gamma$  启动子区域经过测序以分析产生的突变。首先对区域进行 PCR 扩增,然后对 PCR 产物测序,并且观察到多种不同的突变,包括缺失和插入(图 8)。在这个实验中,42%的等位基因突变,并且 20%携带与 HPFH 相关的从 -114 至 -102 的 13bp 缺失。

[0275] 也使用两对靶向  $\gamma$  珠蛋白启动子的 ZFN 与寡核苷酸供体组合来处理细胞,所述寡核苷酸供体被设计成在具有 HPFH 的受试者中再产生最常见的突变。遵循上文关于供体 BTX 装置使用所述相同的方案,其中加入 3  $\mu$ L 的 100  $\mu$ M 供体寡核苷酸溶液。下文示出寡核苷酸供体的序列。通常,在这些实验中使用正向寡核苷酸供体,但反向供体也起作用:

[0276] HBG\_d13 正向:

[0277] acactatctcaatgcaaatactgtetgaaacgggtccctggctaaactccacccatg

[0278] ggttggccagccttgccottgacaaggcaaacttgaccaatagtcttagagtatccagtgaggccagg (124mer, SEQ ID NO :246)

[0279] HEG\_d13\_反向:

[0280] cctggcctcactggatactetaagactattgggtcaagtttgccctgtcaaggcaagg

[0281] ctggccaacccatgggtggagtttagccagggacctttcagcagatatttgcatt

[0282] gagatagtgt (124mer, SEQ ID NO :247)

[0283] 对于在供体存在下的 ZFN 对 34539/34574,从  $\gamma$  珠蛋白基因的 mRNA 产生相比于用 GFP 载体处理的细胞增加 6.38 倍,而对于 ZFN 对 31160/34365,相比于用 GFP 载体处理的细胞,  $\gamma$  mRNA 增加 6.13 倍。

[0284] 将核酸酶处理的 HSC 涂在甲基纤维素上。在通过 PCR 测序对单独的集落进行基因分型后,通过 RT-PCR 测量野生型和突变型集落的  $\gamma$  珠蛋白、 $\beta$  珠蛋白和 18s rRNA 对照物的 mRNA 水平(图 8)。平均来说,  $\gamma$  珠蛋白启动子突变体具有相比于野生型细胞更高的  $\gamma$

珠蛋白 /  $\beta$  珠蛋白比率信息, 并且通过 18srRNA 信号校正指示, 突变型集落中的  $\gamma$  珠蛋白 /  $\beta$  珠蛋白比率增加是由这些集落中的  $\gamma$  珠蛋白 mRNA 水平增加而非  $\beta$  珠蛋白 mRNA 水平降低造成的。

[0285] 实施例 8: 靶向  $\gamma$  珠蛋白启动子的 TALE 核酸酶

[0286] 也做出 TALE 核酸酶以靶向  $\gamma$  珠蛋白启动子区域的 -200 区域或 -110 区域 (上文所述)。使用规范的 TALE 编码和 '+17' TALEN 主链, 如先前所述构建 TALEN (参见共同拥有的美国专利公开 20110301073)。

[0287] 表 3:  $\gamma$  珠蛋白启动子特异性 TALEN

[0288]

SBS 编号	序列 5'→3'	RVD 序列(N→C)
102314	gtATCCTCTTGGGGGcc (SEQ ID NO:133)	NI-NG-HD-HD-NG-HD-NG-NG-NN-NN-NN-NN-NK (SEQ ID NO:134)
102318	atATTTGCATTGAGATAGT gt (SEQ ID NO:135)	NI-NG-NG-NG-NN-HD-NI-NG-NG-NN-NI-NN-NI- NG-NI-NN-NG (SEQ ID NO:136)
102315	gtATCCTCTTGGGGGcc (SEQ ID NO:133)	NI-NG-HD-HD-NG-HD-NG-NG-NN-NN-NN-NN-HD (SEQ ID NO:137)
102320	atATTTGCATTGAGATAgT (SEQ ID NO:135)	NI-NG-NG-NG-NN-HD-NI-NG-NG-NN-NI-NN-NI- NG-NI (SEQ ID NO:136)
102316	gtATCCTCTTGGGGGCCcc (SEQ ID NO:133)	NI-NG-HD-HD-NG-HD-NG-NG-NN-NN-NN-NN-NN- HD-HD (SEQ ID NO:138)
102321	atATTTGCATTGAGATag (SEQ ID NO:135)	NI-NG-NG-NG-NN-HD-NI-NG-NG-NN-NI-NN-NI-NG (SEQ ID NO:139)
102566 (-110)	gtTGGCCAGCCTTGCCTTG ac (SEQ ID NO:248)	NG-NN-NN-HD-HD-NI-NN-HD-HD-NG-NG-NN-HD- HD-NG-NG-NK (SEQ ID NO:249)
102568 (-110)	ttGGTCAAGTTGCCTTGT ca (SEQ ID NO:250)	NN-NN-NG-HD-NI-NI-NN-NG-NG-NG-NN-HD-HD- NG-NG-NN-NG (SEQ ID NO:251)

[0289] 然后成对使用 TALEN 以测试 K562 细胞中的裂解并通过如先前所述的 Cell 测定法测定并且配对结果示于下表 4 中。另外, 针对 CD34+ 细胞测试 TALEN 对 102566/102568 并测得其具有如通过 MiSeq 分析所测量的 51.39% NHEJ。

[0290] 也对两对 TALEN 测试  $\gamma$  珠蛋白 mRNA 表达, 如通过  $\gamma$  珠蛋白与  $\alpha$  珠蛋白 mRNA 的比率所测量。测得 102566/102568 对相比于用 GFP 载体处理的 CD34+ 细胞使  $\gamma$  珠蛋白表达增加 6.25 倍, 并且 102318/102314 对相比于用 GFP 载体处理的 CD34+ 细胞使  $\gamma$  珠蛋白增加 2.14 倍。也对上述供体寡核苷酸测试 102566/102568 对, 并且测得所得细胞相比于用 GFP 载体处理的 CD34+ 细胞使  $\gamma$  珠蛋白表达增加 9.13 倍。

[0291] 表 4: 用 TALEN 编辑  $\gamma$  珠蛋白启动子区域

[0292]

TALEN 对	% NHEJ '+17'
102314:102318	41.6
102315:102320	47.9
102316:102321	46.6

[0293] 实施例 9 :CD34+ 干细胞中的  $\gamma$  珠蛋白编辑

[0294] 然后将对  $\gamma$  珠蛋白启动子区域具有特异性的核酸酶用于患者来源的 CD34+ 细胞中。用核酸酶处理细胞,然后通过 Cel 1 分析来分析成功的编辑。进一步分析细胞以检查  $\gamma$  珠蛋白与  $\beta$  珠蛋白的比率并且证实  $\gamma$  珠蛋白的表达增加。对于增加的  $\gamma$  珠蛋白表达测得的代表性数据位于上述不同方法的实验区中。

[0295] 实施例 10 :小鼠中编辑的 CD34+ 植入

[0296] 核酸酶处理的 CD34+ 细胞 (人类干细胞 / 祖细胞 HSPC) 保持植入 NOD/SCID/IL2r  $\gamma$  (裸) 小鼠的能力并产生多克隆多谱系后代,其中参与调节  $\gamma$  珠蛋白的基因被永久破坏 (参见 Holt 等, (2010) Nat Biotechnol. 8 月 ;28(8):839-47)。类似地,在  $\beta$  珠蛋白基因座处编辑的 CD34+ 或 HSPC 能够植入并产生带有所需基因组编辑的多谱系后代,在所述  $\beta$  珠蛋白基因座中校正突变,或者供体  $\beta$  珠蛋白基因被插入安全港基因座中,或用核酸酶处理以改变  $\gamma$  珠蛋白的表达。少数经过编辑的 HSPC 可以繁殖具有经过编辑的后代的动物,这种论证支持使用核酸酶修饰的自体同源造血干细胞作为治疗血红蛋白病的临床方法。

[0297] 本文提及的所有专利、专利申请和公开特此以全文引用的方式并入。

[0298] 尽管已出于理解清楚的目的通过说明和实例详细地提供了公开内容,但对于本领域技术人员将显而易见的是,可在不偏离本公开的精神或范围的情况下实施各种变化和修改。因此,前述描述和实施例不应理解为具限制性。

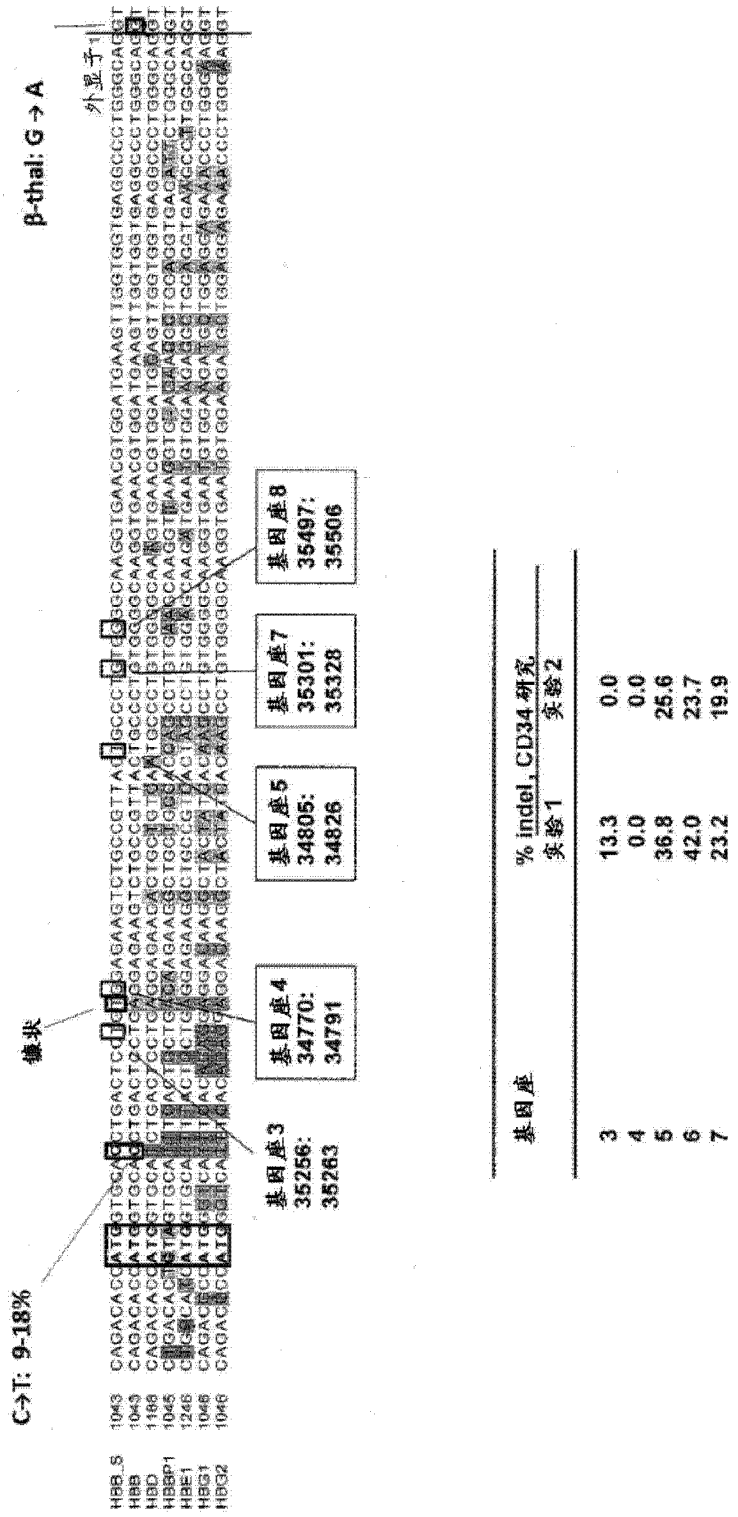


图 1

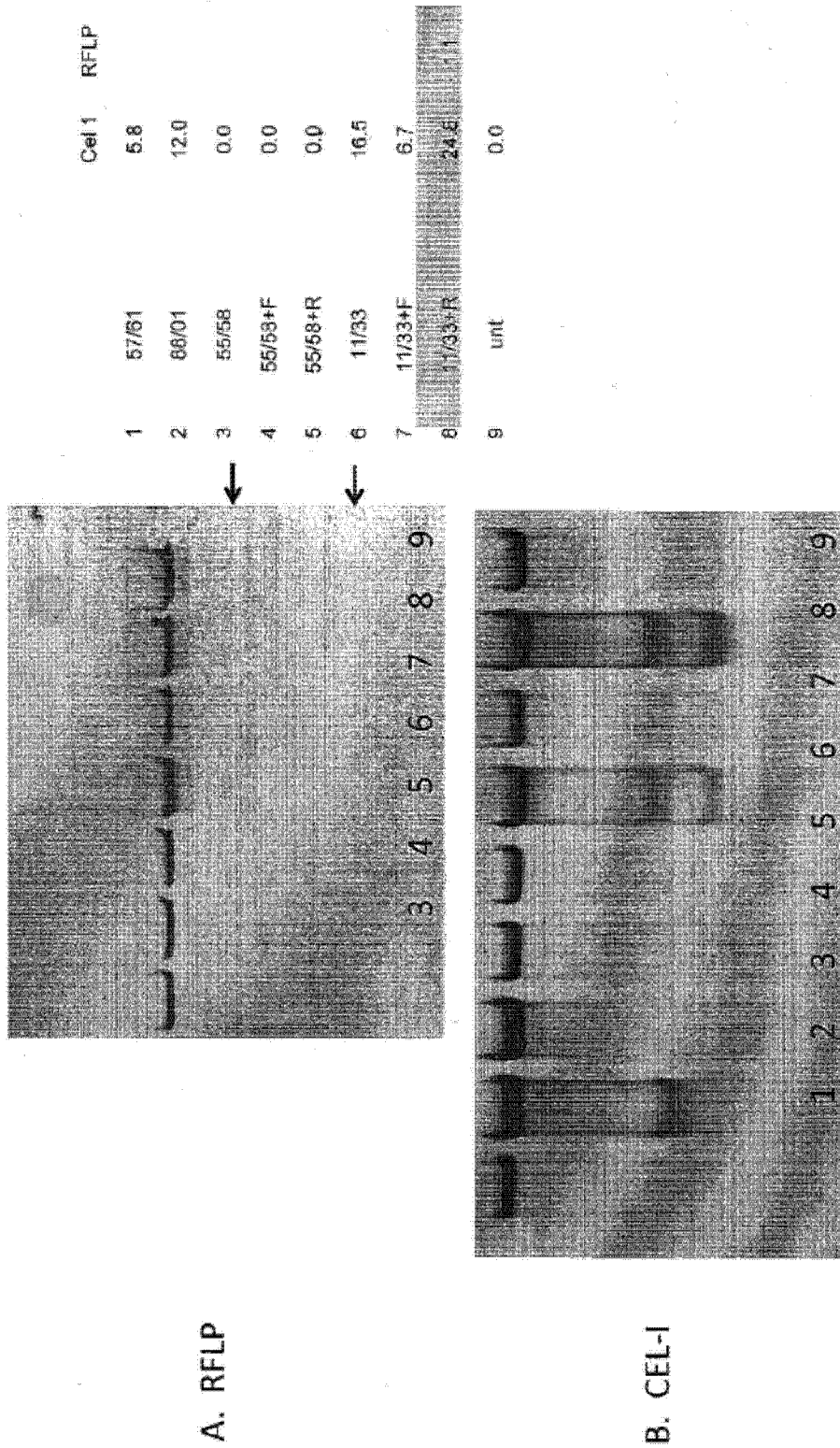


图 2

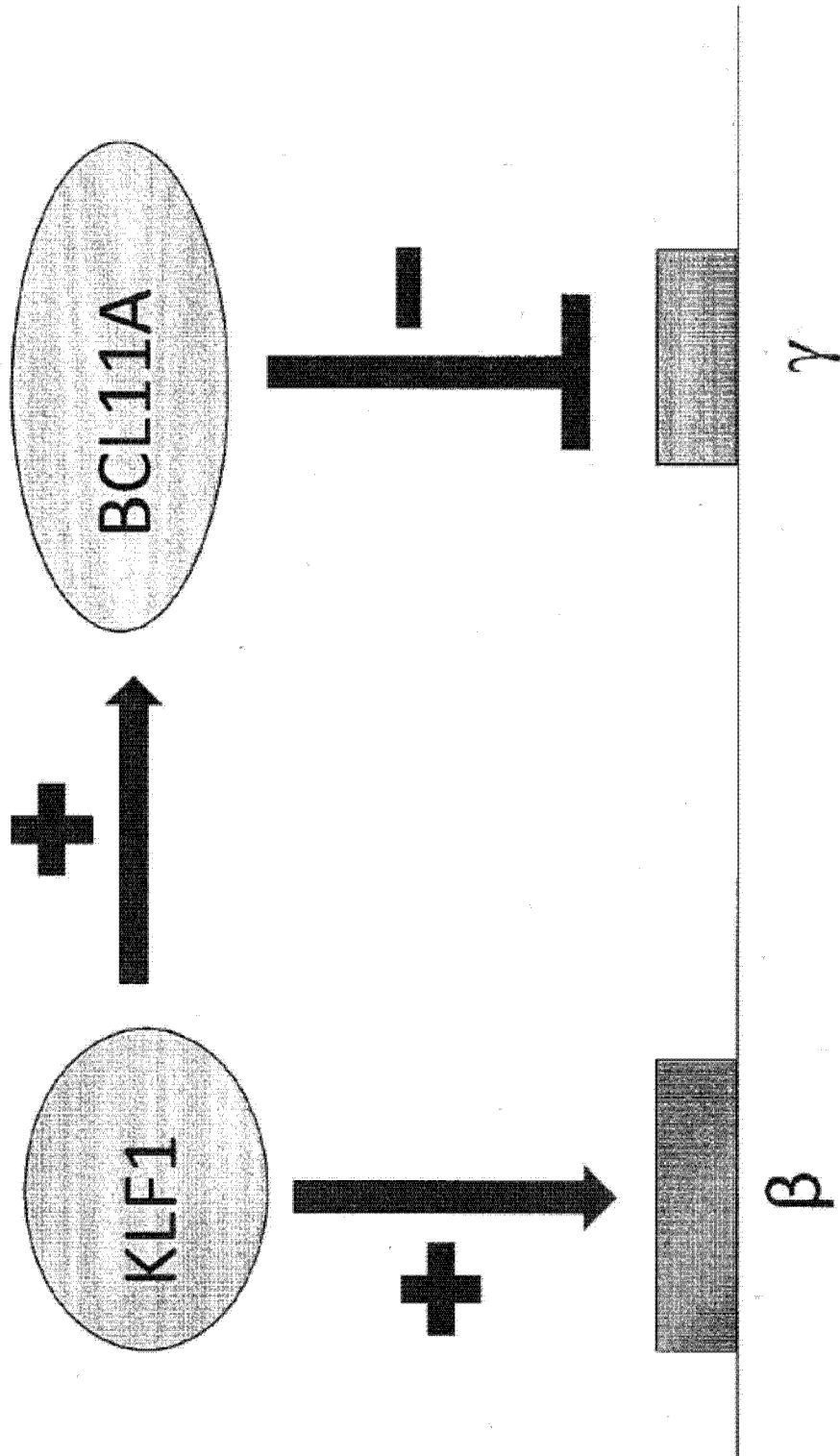


图 3

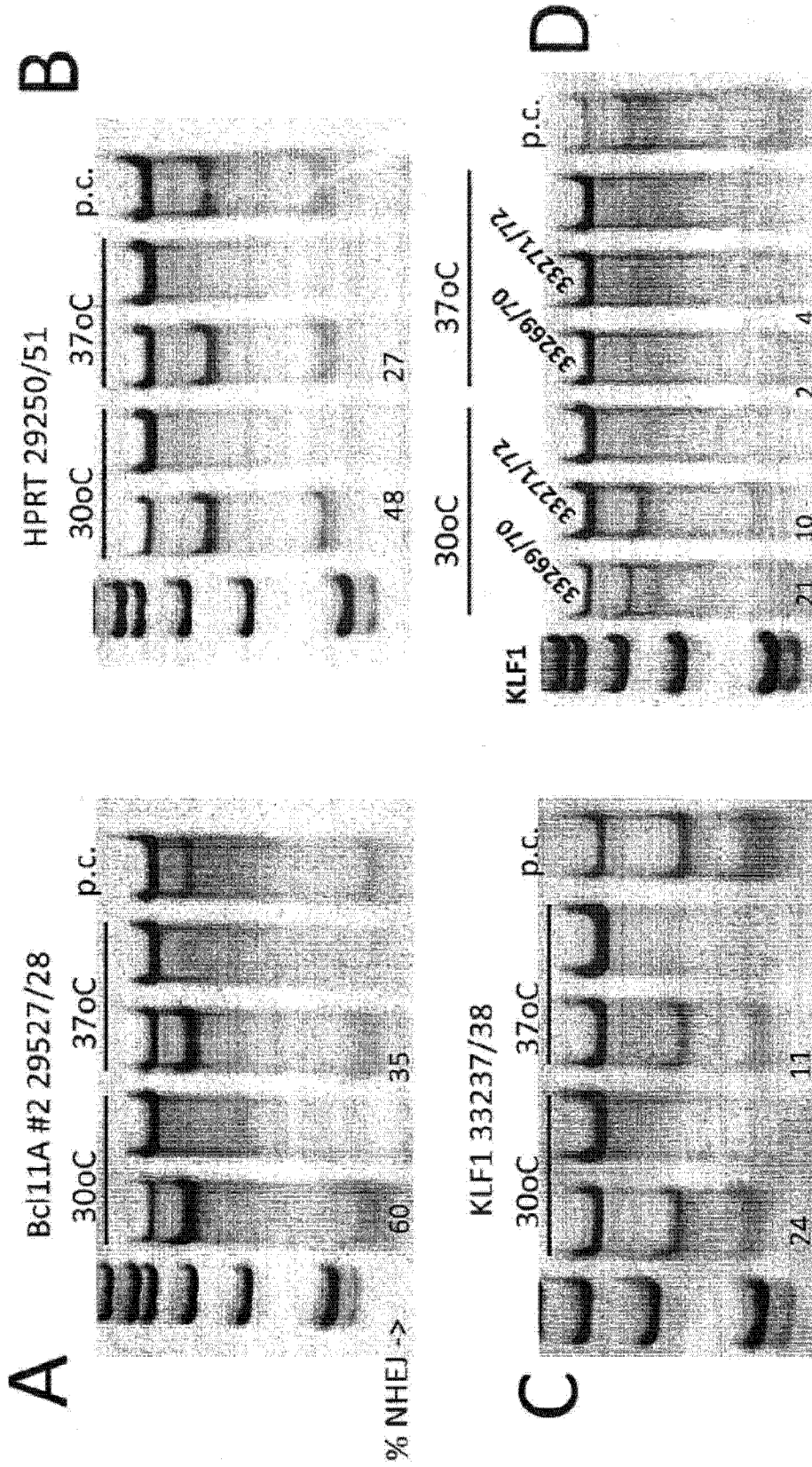


图 4



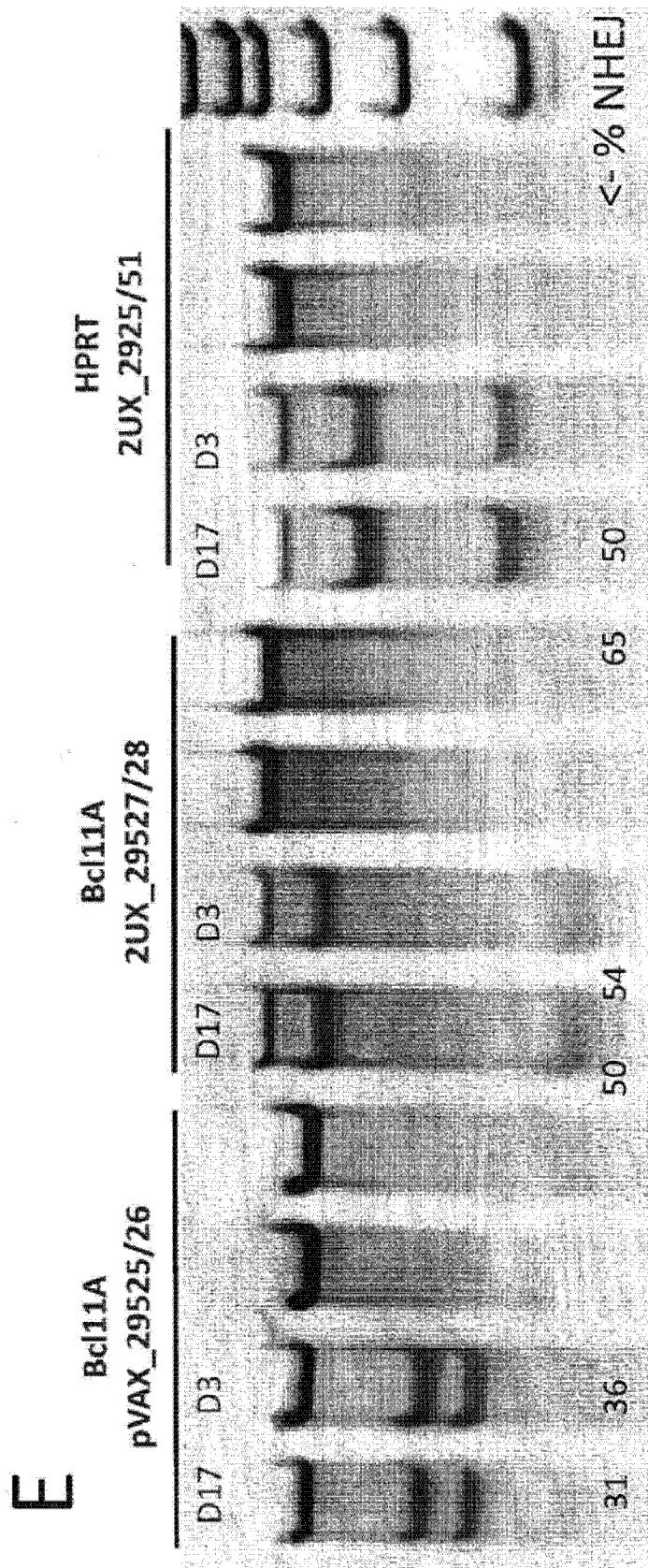


图 4

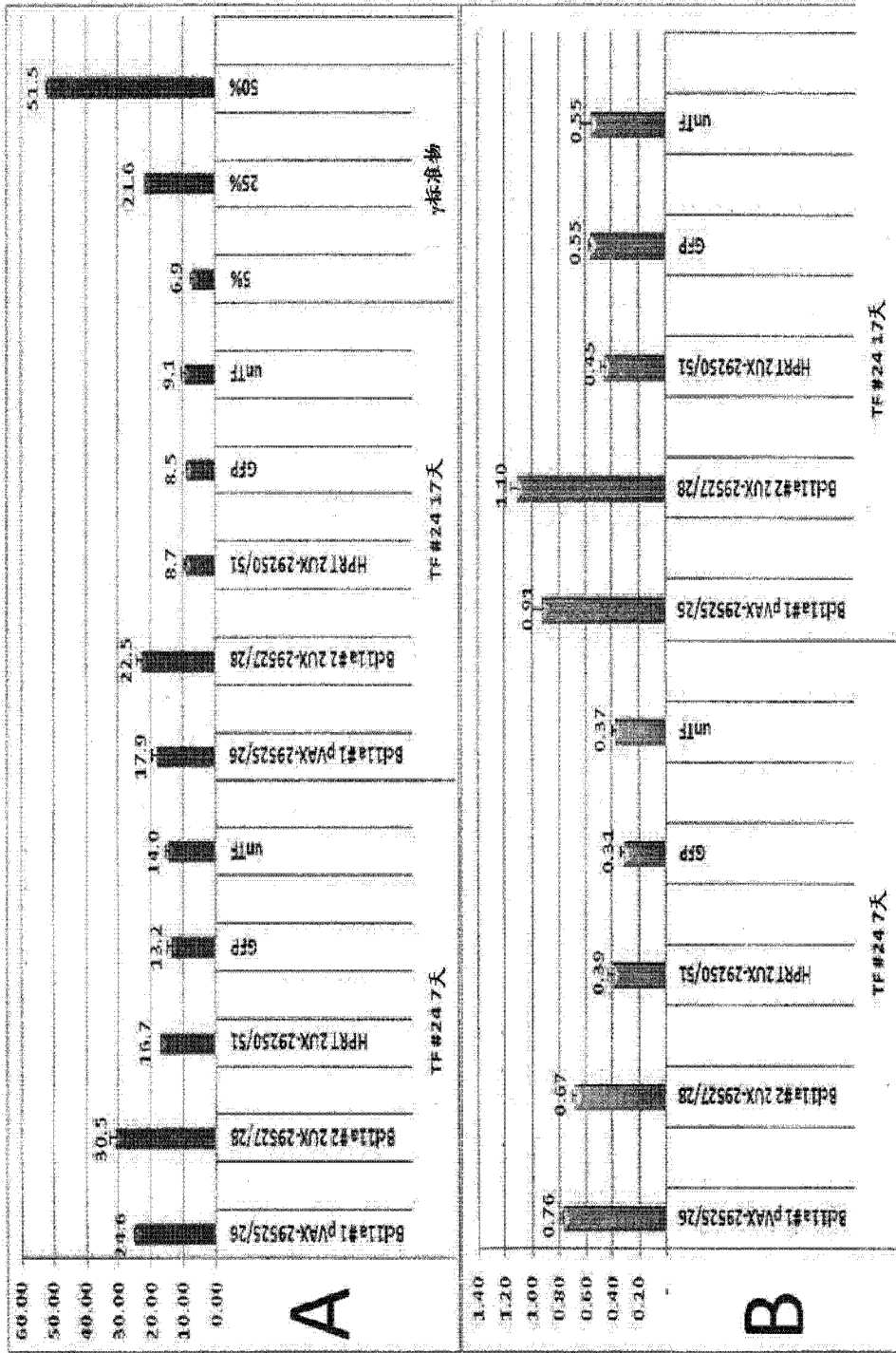


图 5

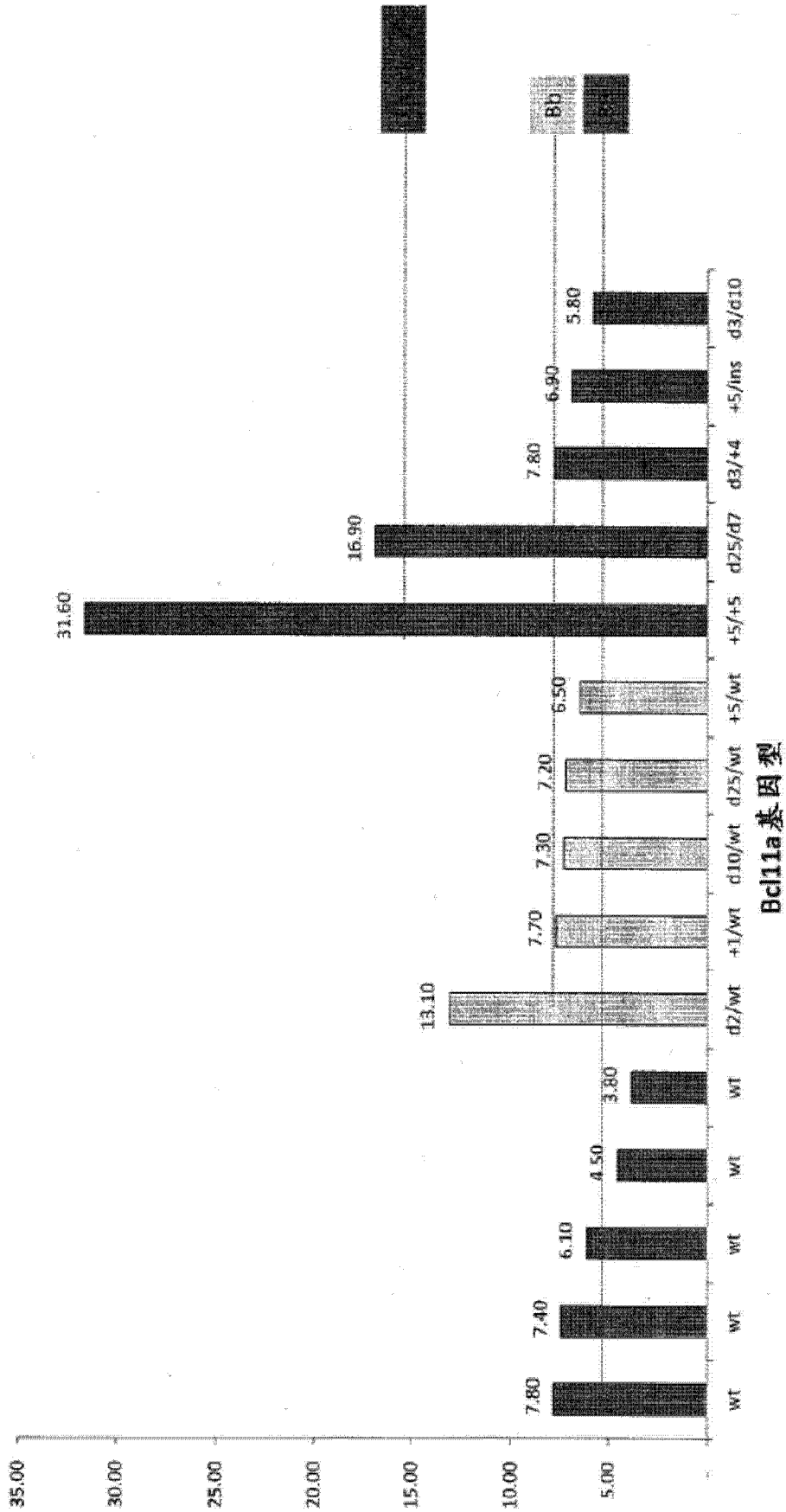


图 6

参考物	GGCCAGCCCTTGCCCTTGACCCAATAGCCCTTGACCAAGGCAAACTTGACCAATAG
wt (26)	GGCCAGCCCTTGCCCTTGACCCAATAGCCCTTGACCAAGGCAAACTTGACCAATAG
Δ1bp (1)	GGCCAGCCCTTGCCCTTGACC-ATAGCCCTTGACCAAGGCAAACTTGACCAATAG
Δ4bp (1)	GGCCAGCCCTTGCCCTT---AATAGCCCTTGACCAAGGCAAACTTGACCAATAG
Δ6bp (2)	GGCCAGCCCTTGCCCTTGAC-----CCTTGACCAAGGCAAACTTGACCAATAG
Δ6bp (1)	GGCCAGCCCTTGCCCTTGACCCAATAGC-----AAGGCAAACTTGACCAATAG
Δ13bp (9)	GGCCAGCCCTTGCCCTTGAC-----AAGGCAAACTTGACCAATAG
+1bp (1)	GGCCAGCCCTTGCCCTTGACCCAATAAGCCTTGACCAAGGCAAACTTGACCAATA
+3bp (1)	GGCCAGCCCTTGCCCTTGACCCAATAGCAGCCTTGACCAAGGCAAACTTGACCCAA
+5bp (3)	GGCCAGCCCTTGCCCTTGACCCAATAGATAGCCCTTGACCAAGGCAAACTTGACCC

图 7

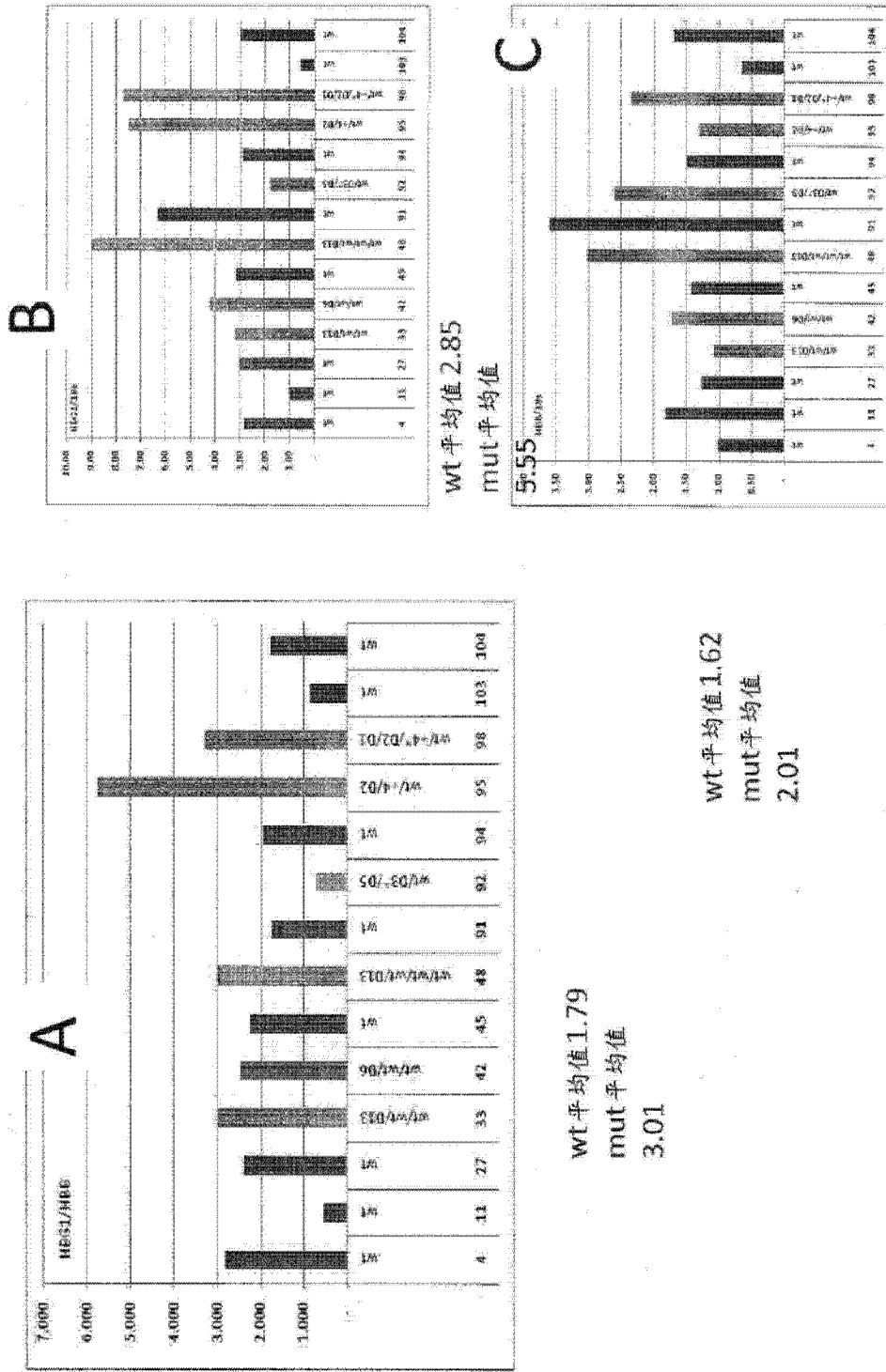


图 8

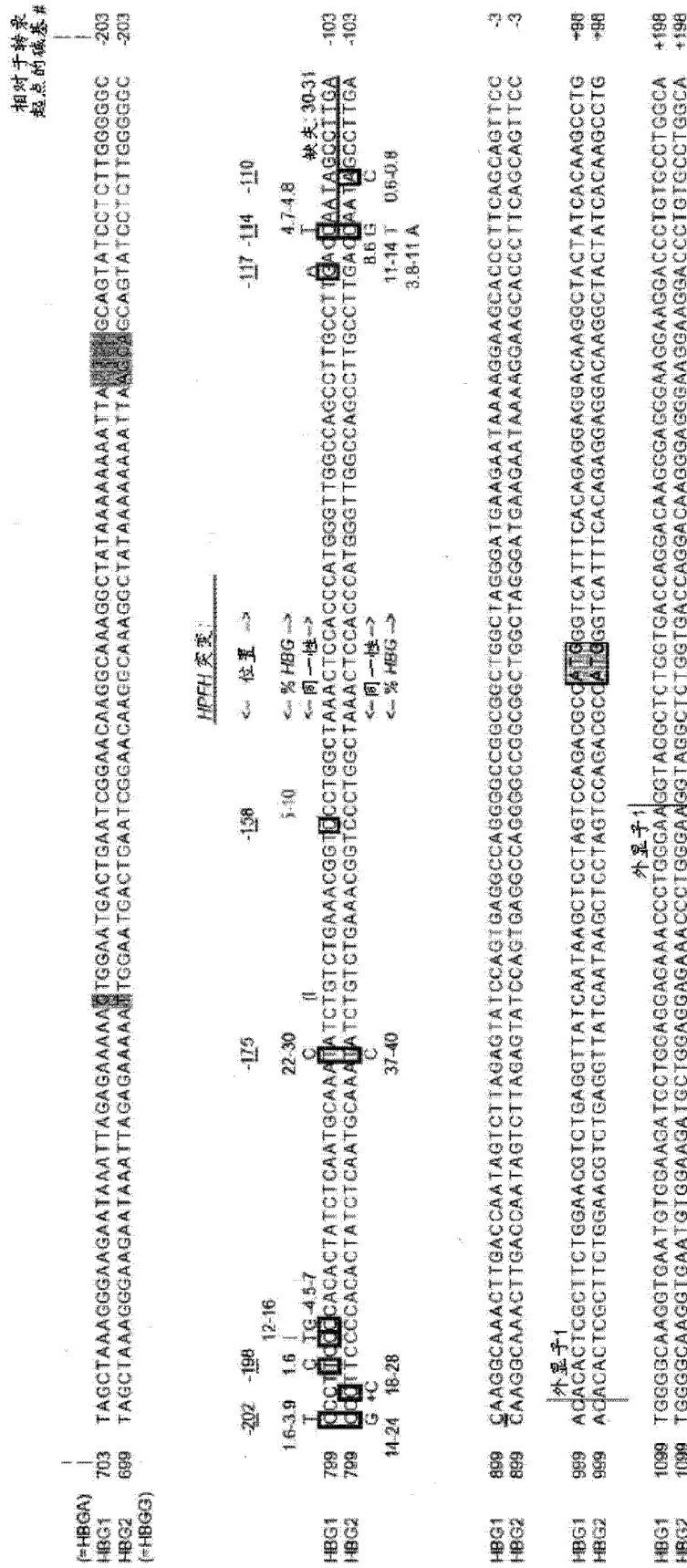


图 9

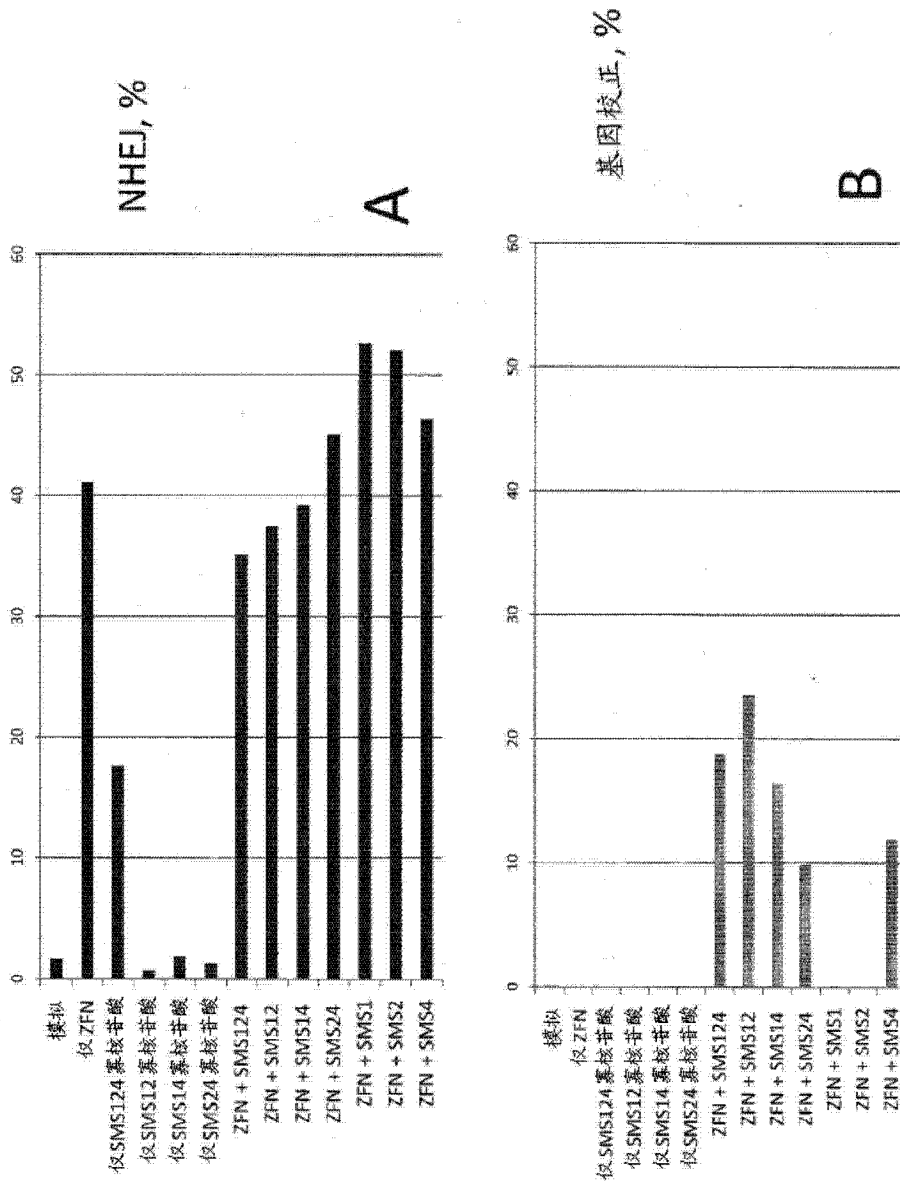


图 10

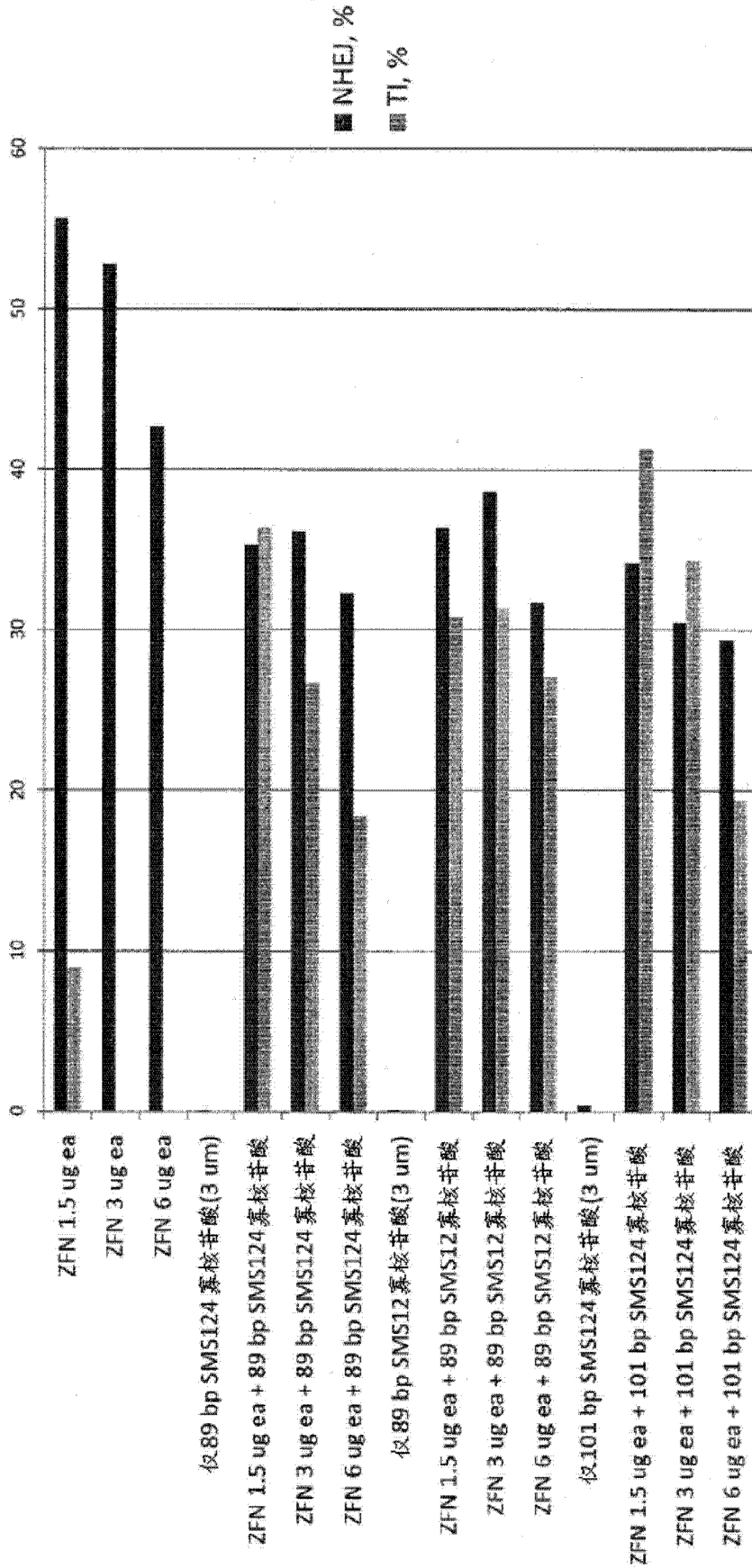


图 11



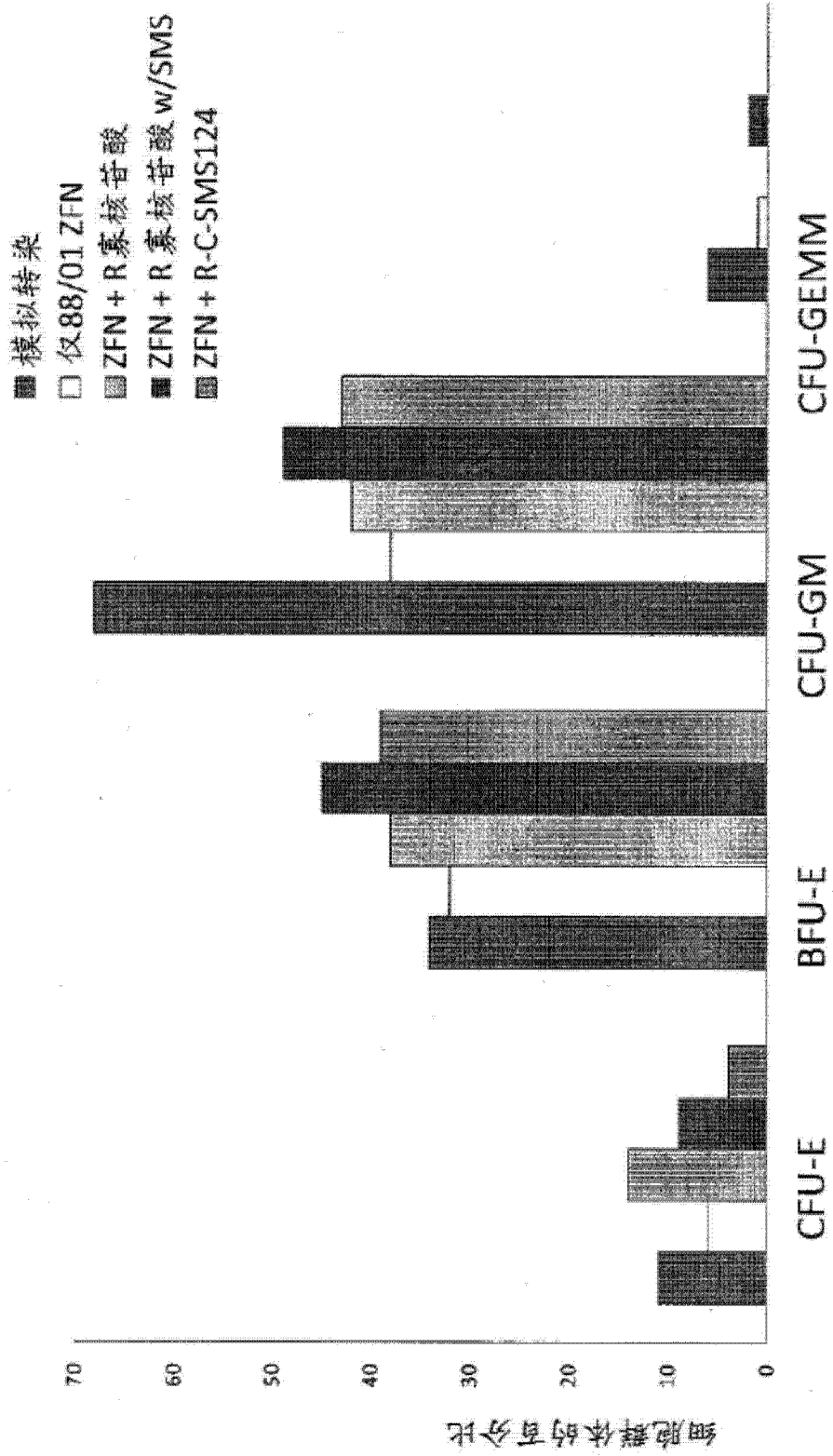


图 12

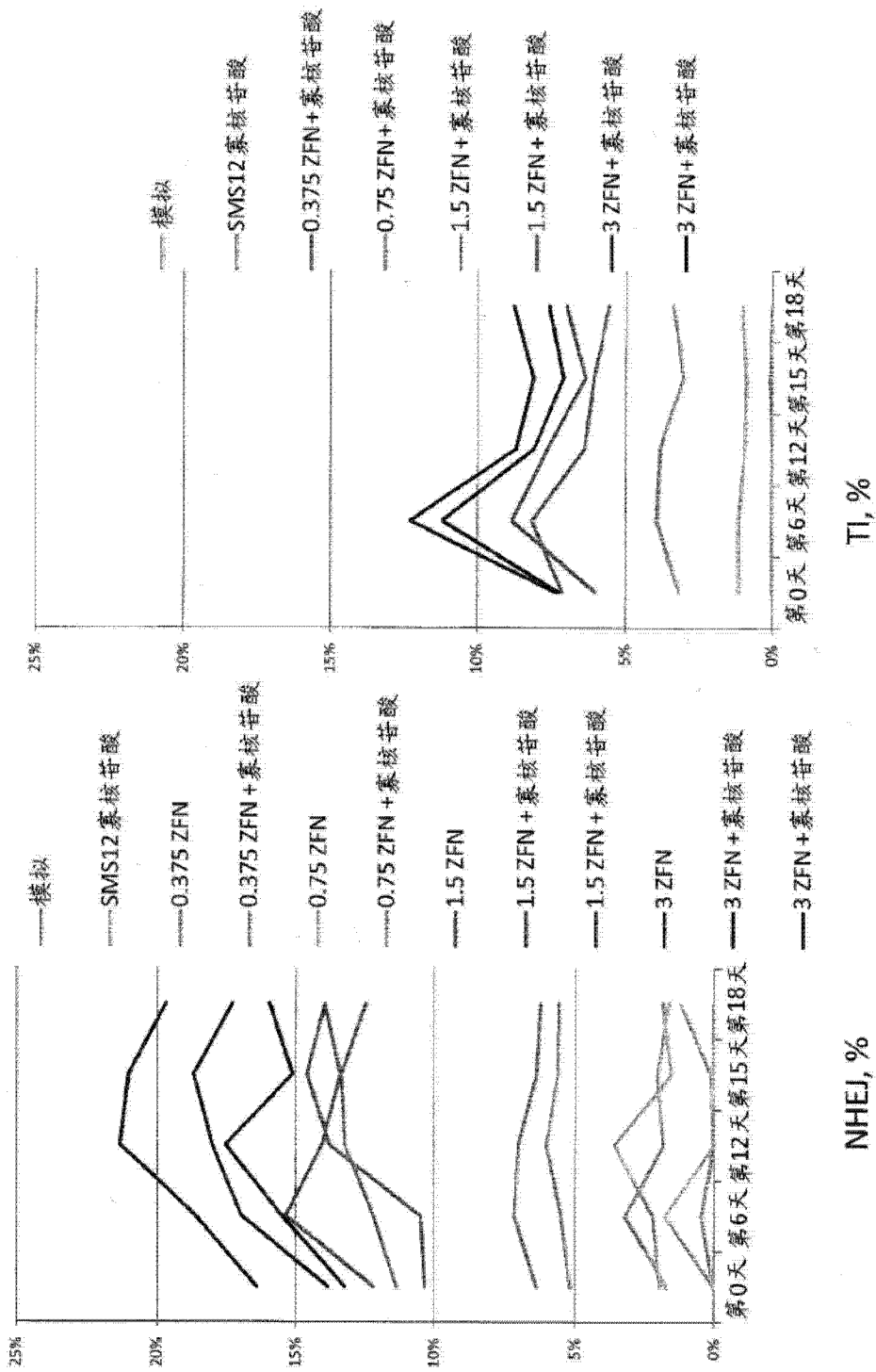


图 13