



## (12) 发明专利申请

(10) 申请公布号 CN 113699113 A

(43) 申请公布日 2021.11.26

(21) 申请号 202110258721.2

C12N 9/22 (2006.01)

(22) 申请日 2015.09.16

A61K 35/28 (2015.01)

(30) 优先权数据

62/051,159 2014.09.16 US

A61P 7/06 (2006.01)

(62) 分案原申请数据

201580049971.8 2015.09.16

(71) 申请人 桑格摩治疗股份有限公司

地址 美国加利福尼亚州

(72) 发明人 格雷戈里·J·科斯特

杰弗里·C·米勒 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 邱晓敏

(51) Int.Cl.

C12N 5/10 (2006.01)

C12N 15/12 (2006.01)

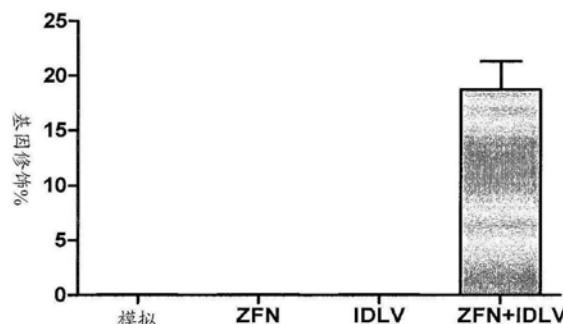
权利要求书1页 说明书40页 附图35页

(54) 发明名称

用于造血干细胞中核酸酶介导的基因组工程化和校正的方法和组合物

(57) 摘要

本公开是在基因组工程领域内，特别是造血干细胞的基因组的靶向修饰。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)中所涉及的蛋白的基因(例如，产生疾病中的缺乏、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。本发明描述了用于基因治疗和基因组工程中的组合物和方法。



1.一种包含经遗传修饰的细胞的组合物,其包含:

-一对锌指核酸酶,其与SEQ ID NO:23和SEQ ID NO:24结合,并在内源性人β-血红蛋白(Hbb)基因中形成双链断裂;和

-在被该对锌指核酸酶双链断裂后,Hbb基因的SEQ ID NO:23或SEQ ID NO:24的1-300个碱基对内的基因组修饰,

其中该对锌指核酸酶包含表1中的命名为33488和33501;33488和47817;47773和33501;或47773和47817的锌指蛋白,其中所述基因组修饰包括插入和/或缺失。

2.权利要求1的组合物,其中所述细胞是干细胞,例如造血干细胞,任选地是CD34+细胞。

3.权利要求1的组合物,其中:

所述经遗传修饰的细胞是HSPC,并且所述对锌指核酸酶包含命名为33488和33501的锌指蛋白;和

所述组合物还包含源自所述经遗传修饰的HSPC细胞的经遗传修饰的分化的红细胞(RBC)。

4.权利要求1或2的组合物,其中所述对锌指核酸酶包含表1中命名为33488和33501或47773和47817的锌指蛋白。

5.根据权利要求1至4中任一项所述的组合物,其中一个或多个多核苷酸编码所述对锌指核酸酶。

6.权利要求1-5中任一项的组合物,其中所述插入包括整合编码转基因的供体多核苷酸,任选地其中所述供体多核苷酸纠正镰状细胞突变。

7.根据权利要求1-6中任一项所述的组合物,其中所述供体多核苷酸通过非整合型慢病毒载体(IDLV)递送至所述细胞。

8.一种包含经遗传修饰的细胞的组合物,所述经遗传修饰的细胞在被一对锌指核酸酶双链断裂后,在内源性人β-血红蛋白(Hbb)基因的SEQ ID NO:23或SEQ ID NO:24的1-300个碱基对内包含基因组修饰,该对锌指核酸酶包含表1中命名为33488和33501;33488和47817;47773和33501;或47773和47817的锌指蛋白,其中所述基因组修饰包括整合编码纠正镰刀细胞突变的转基因的供体多核苷酸,并且其中供体多核苷酸的整合引入了如图4C所示的一个或多个沉默突变,该突变阻断了锌指核酸酶重新切割。

9.根据权利要求1至8中任一项所述的组合物,其用于治疗血红蛋白病,例如镰状细胞病,其中所述细胞是干细胞(例如造血干细胞或红细胞前体)并被移植到病人。

## 用于造血干细胞中核酸酶介导的基因组工程化和校正的方法 和组合物

本申请是申请号为201580049971.8,申请日为2015年9月16日,申请人为桑格摩治疗股份有限公司,发明创造名称为“用于造血干细胞中核酸酶介导的基因组工程化和校正的方法和组合物”的发明专利申请的分案申请。

### 相关申请的交叉引用

[0001] 本申请要求2014年9月16日提交的美国临时申请No.62/051,159 的权益,所述临时申请的公开内容在此通过引用全文并入。

### 技术领域

[0002] 本公开属于造血干细胞的基因组工程领域,特别是造血细胞的基因组的靶向修饰。

### 背景技术

[0003] 在用于大量疾病的基因治疗中最有希望的方法之一涉及使用干细胞的体外遗传修饰,然后移植并在患者体内植入经修饰的细胞。当引入的干细胞显示长期持久性和多谱系分化时是特别有希望的。造血干细胞(最常见的是基于CD34细胞表面标记物的表达的富集的细胞形式)是最特别有用的细胞群,因为它们可以容易获得并含有长期造血干细胞(LT-HSC),其可以在移植后重建整个造血谱系。

[0004] 已经描述了用于基因组DNA的靶向裂解(cleavage)的各种方法和组合物。这种靶向裂解事件可用于例如诱导靶向诱变,诱导细胞DNA序列的靶向缺失,并促进来自任何生物体的细胞中预定染色体基因座处的靶向重组。参见,例如,美国专利No.9,045,763;美国专利No.9,005,973;美国专利 No.8,956,828;美国专利No.8,945,868;美国专利No.8,586,526;美国专利 No.6,534,261;美国专利No.6,599,692;美国专利No.6,503,717;美国专利No.6,689,558;美国专利No.7,067,317;美国专利No.7,262,054;美国专利 No.7,888,121;美国专利No.7,972,854;美国专利No.7,914,796;美国专利 No.7,951,925;美国专利No.8,110,379;美国专利No.8,409,861;美国专利公布20030232410;美国专利公布20050208489;美国专利公布20050026157;美国专利公布20050064474;美国专利公布20060063231;美国专利公布 20080159996;美国专利公布201000218264;美国专利公布20120017290;美国专利公布20110265198;美国专利公布20130137104;美国专利公布 20130122591;美国专利公布20130177983与美国专利公布20130177960和美国专利公布20150056705以及美国申请No.14/706,747,其公开内容通过引用整体并入本文以用于所有目的。这些方法通常涉及使用工程改造裂解系统来诱导靶DNA序列中的双链断裂(DSB)或切口,使得通过错误产生过程导致的断裂的修复(如非同源末端连接(NHEJ))或使用修复模板的修复(同源性定向修复或HDR)可以导致基因的敲除或目的序列的插入(靶向整合)。随后的修复途径(NHEJ与HDR或两者)通常取决于修复模板的存在和几种竞争性修复途径的活性。

[0005] 在不存在外部供应的修复模板(例如供体)的情况下引入的双链断裂通常用于经

由通过细胞NHEJ途径引入的突变使靶基因失活。NHEJ途径可以基于微观同源介导的末端连接而分离成标准Ku依赖性途径和替代的Ku独立途径,该微观同源介导的末端连接利用了在裂解位点附近的直接重复的短轨道。通过这两个NHEJ途径进行基因编辑后的插入和/或缺失(“indel”)模式不同,这可能导致突变的功能结果方面的差异,具体取决于应用。

[0006] 在携带与双链断裂的旁侧序列同源的段(stretch)的外部供应的供体的存在下,使用供体分子的同源性定向基因修复(HDR)可以用于改变单碱基或小段的序列(“基因校正”),或者在另一方面,用于将整个表达盒(“基因添加”)靶向插入预定的基因组位置。

[0007] 裂解可以通过使用特异性核酸酶,例如工程改造的锌指核酸酶(ZFN)、转录激活因子样效应物核酸酶(TALEN),或使用具有工程改造的crRNA/tracr RNA的CRISPR/Cas系统(‘单导向RNA’)以指导特异性裂解。参见例如美国专利No.8,697,359和No.8,932,814以及美国专利公布No.20150056705。此外,靶向核酸酶正在基于Argonaute系统开发(例如,来自嗜热栖热菌(*T. thermophilus*),称为“TtAgo”,参见Swarts等人(2014)Nature 507(7491):258-261),其也可能具有在基因组编辑和基因治疗中的用途。

[0008] 可以利用使用上述核酸酶系统中的一种的靶向裂解来使用HDR或NHEJ介导的方法将核酸插入特定的靶位置。然而,将核酸酶系统和供体两者都递送到细胞可能是有问题的。例如,通过质粒转导来递送供体或核酸酶到细胞中可能对受体细胞有毒,特别是对于作为原代细胞的细胞是有毒的,因此不如来自细胞系的细胞那样健康。

[0009] CD34+干细胞或祖细胞是特征在于其自我更新和/或分化成淋巴谱系细胞(例如T细胞、B细胞、NK细胞)和髓细胞谱系细胞(例如单核细胞、红细胞、嗜酸性粒细胞、嗜碱性粒细胞和嗜中性粒细胞)的能力的异质的成组的细胞。它们的异质性源于CD34+干细胞群体内有多个亚组,亚组通常反映特定组的多能性(不管是否是谱系定型的(lineage committed))。例如,属于CD38-的CD34+细胞是更原始的、未成熟的CD34+祖细胞(也称为长期造血祖细胞),而属于CD34+CD38+(短期造血祖细胞)是谱系定型的(参见Stella et al (1995) Hematologica 80:367-387)。当该群体随后沿着分化途径进一步向下进展时,CD34标记丧失。CD34+干细胞在临床细胞治疗中具有巨大的潜力。

[0010] 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他细胞器,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便去除。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 $\beta$ 样珠蛋白链以及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的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的比率,减轻了突变细胞表型的一些方面(参见Voon et al (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对氧具有更大的结合亲和力,以使得氧可经由母体血流递送至婴儿的系统。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。

[0011] 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,因而表达相对较低。

[0012] 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’,导致细胞的柔韧性损失。镰状化RBC不再能够挤入毛细血管床中并且可能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。

[0013] 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 $\gamma$ 珠蛋白的生产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。

[0014] 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他基因的表达,例如以治疗血红蛋白病,例如镰状细胞疾病。

## 发明内容

[0015] 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)中所涉及的蛋白的基因(例如,产生疾病中的缺乏、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。本发明描述了用于基因治疗和基因组工程中的组合物和方法。

[0016] 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。一种方法涉及使用基因校正,其中有缺陷的内源性 $\beta$ 珠蛋白基因被靶向并且突变体序列被替换。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。

[0017] 在一个方面,本文描述了经遗传修饰的细胞或细胞系,例如与相同类型的细胞或细胞系的野生型基因组序列相比而言。在某些实施方案中,细胞包含经遗传修饰的RBC前体(称为“HSC”的造血干细胞)。细胞或细胞系对于修饰可以是杂合的或纯合的。修饰可包括插入、缺失和/或其组合。在某些实施方案中,用工程改造的核酸酶和供体核酸修饰HSC,使得插入和表达野生型基因(例如珠蛋白基因)和/或校正内源性异常基因。在某些实施方案中,修饰(例如插入)在核酸酶结合和/或裂解位点处或附近,包括但不限于位于裂解和/或结

合位点的位点的上游或下游的1-300个碱基对(或其间的任何数目的碱基对)内的序列进行修饰;在结合和/或裂解位点的任一侧的1-100个碱基对(或其间的任何数目的碱基对)内的修饰;在结合和/或裂解位点的任一侧上的1至50个碱基对(或其间的任何数目的碱基对)内的修饰;和/或对核酸酶结合位点和/或裂解位点的一个或多个碱基对的修饰。在某些实施方案中,修饰是在或接近(例如,1-300个碱基对或其间的任何数目的碱基对)SEQ ID NO:23或SEQ ID NO:24。在其他实施方案中,修饰是SEQ ID NO:23或SEQ ID NO:24的1-100个碱基对(或其间的任何数目的碱基对)。在某些实施方案中,修饰是在SEQ ID NO:23和/或SEQ ID NO:24内,例如在SEQ ID NO:23 或SEQ ID NO:24的1个或多个碱基对的修饰。在一些情况下,用于插入的野生型基因序列编码野生型β珠蛋白。在其他情况下,内源性异常基因是β珠蛋白基因,例如一个或多个校正内源性异常人β-血红蛋白(Hbb)基因中的至少一个突变的基因组修饰。在一些方面,β珠蛋白基因的修饰允许转录的 RNA的适当剪接。在其他方面,β珠蛋白基因的修饰校正镰状等位基因中的突变,使得编码多肽序列中第六位置的错误缬氨酸氨基酸的密码子GTG被改变为GAG,其编码谷氨酸,如在野生型多肽中所发现的。还提供了从如本文所述的经遗传修饰的干细胞的部分或完全分化的细胞(例如,RBC或RBC前体细胞)。还提供了组合物,例如包含本文所述的经遗传修饰的细胞的药物组合物。

[0018] 在另一个方面,本文描述结合于基因组中目标区域(例如,β珠蛋白基因)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,示于表1A中的具有5或6个含识别螺旋区域的指的锌指蛋白)。

[0019] 在另一个方面,本文描述结合于基因组中目标区域(例如,β珠蛋白基因)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)。

[0020] 在另一个方面,本文描述结合于基因组中目标区域(例如,疾病相关基因)中的靶位点的CRISPR/Cas或TtAgo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或TtAgo系统识别珠蛋白基因中的靶标。

[0021] 在一些实施方案中,促进细胞的基因组工程改造的核酸酶作为肽递送,而在其他实施方案中,其作为编码核酸酶的核酸递送。在一些实施方案中,多于一种的核酸酶被使用,并且可以被以核酸形式、蛋白形式或其组合形式递送。在一些优选的实施方案中,编码核酸酶的核酸是mRNA,并且在一些情况下,mRNA被保护。在进一步优选的实施方案中,mRNA可以包含ARCA帽和/或可以包含经修饰和未修饰的核苷酸的混合物。核酸酶可以包括锌指

核酸酶 (ZFN)、TALE-核酸酶 (TALEN)、TtAgo或CRISPR/Cas核酸酶系统或其组合。在优选的实施方案中,编码核酸酶的核酸通过电穿孔递送。在另一方面,本文描述了用于在细胞中核酸酶介导的裂解后增加靶向整合(例如,经由HDR)的方法。在某些实施方案中,所述方法包括以下步骤:(i) 将一种或多种核酸酶(和/或mRNA或表达构建体,如果需要的话,其表达核酸酶和一种或多种单导向RNA)与一个或多个供体分子导入宿主细胞中,和(ii) 在引入核酸酶之前、期间和/或之后,将影响和/或增加干细胞扩增而不损失干细胞或分化的一种或多种因子引入该细胞和/或含有该细胞的培养基中。供体可以在编码核酸酶的核酸之前、之后或与其同时递送。

[0022] 供体可以通过病毒和/或非病毒基因转移方法递送。在其他实施方案中,供体通过腺相关病毒 (AAV) 递送至细胞。在一些情况下,AAV包含与衣壳血清型相比属于异源血清型的LTR。在其他实施方案中,供体通过慢病毒 (lentivirus) 递送至细胞。在一些情况下,慢病毒是整合酶缺陷型慢病毒 (IDLV)。

[0023] 在某些实施方案中,供体核酸通过非同源依赖性方法(例如, NHEJ)整合。如上所述,在体内裂解时,供体序列可以以靶向方式整合到在 DSB位置处的细胞基因组中。供体序列可以包括用于产生DSB的一种或多种核酸酶的一个或多个相同的靶位点。因此,供体序列可以被用于裂解内部期望整合的内源基因的一种或多种相同的核酸酶裂解。在某些实施方案中,供体序列包括来自用于诱导DSB的核酸酶的不同核酸酶靶位点。靶细胞的基因组中的DSB可以通过任何机制产生。在某些实施方案中,DSB由一个或多个锌指核酸酶 (ZFN)、包含锌指结合结构域(其被工程改造以结合目标区域内的序列)的融合蛋白和裂解结构域或裂解半结构域产生。在其他实施方案中,DSB由与核酸酶结构域 (TALEN) 融合的一个或多个TALE DNA结合结构域 (天然存在的或非天然存在的) 产生。在另外的实施方案中,使用CRISPR/Cas 核酸酶系统产生DSB,其中工程改造的单导向RNA或其功能等同物用于将核酸酶引导至基因组中的靶位点。在另外的实施方案中,使用TtAgo系统产生DSB。

[0024] 在另一方面,本发明提供了用于在核酸酶介导的细胞基因组裂解后增加基因破坏和/或靶向整合(例如基因校正)的试剂盒(例如ZFN、TAL-效应结构域核酸酶融合蛋白、TtAgo系统、或工程化的归巢内切核酸酶或具有 CRISPR/Cas系统的工程化导向RNA)。试剂盒通常包括:一种或多种结合靶位点的核酸酶,一种或多种影响干细胞扩增和/或分化的因子,以及用于将核酸酶和干细胞影响因子导入细胞使得核酸酶介导的基因破坏和/或靶向整合增强的说明书。任选地,含有核酸酶的靶位点的细胞也可以包括在本文所述的试剂盒中。在某些实施方案中,试剂盒包含至少一种具有靶基因和能够在靶基因内进行裂解的已知核酸酶的构建体。此类试剂盒可用于优化多种不同宿主细胞类型中的裂解条件。本发明考虑的其他试剂盒可以包括能够在基因组内的已知靶基因座内进行裂解的已知核酸酶,并且可以另外包含供体核酸。在一些方面,供体DNA可编码多肽、调节区域或结构核酸。在一些实施方案中,多肽是报告基因(例如GFP)。这样的试剂盒对于优化是有用的。这样的试剂盒可用于优化供体整合或构建特异性修饰的细胞、细胞系和含有基因破坏或靶向插入物的动物的条件。

[0025] 在其他方面,描述了将如本文所述的经遗传修饰的干细胞施用给受试者的方法。本文所述的经遗传修饰的血细胞前体 (“HSC/PC”) 通常被提供于骨髓移植植物中并且HSC/PC 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/PC,并且在其他情

况下,细胞是从人类骨髓或脐带中分离的。在一些方面,通过经设计以敲除特异性基因或调节序列的核酸酶处理来编辑HSC/PC。在其他方面,用工程改造的核酸酶和供体核酸修饰HSC/PC,以使得插入并表达野生型基因或其他目标基因和/或校正内源异常基因。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/PC施用至患者。在其他方面,在完全骨髓清除后施用HSC/PC,以使得在植入后,100%的造血细胞是源自经修饰的HSC/PC。此外,细胞可以在细胞周期的G2期停止。

[0026] 在一些实施方案中,转基因HSC/PC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,从而使得能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他实体。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,安全港)中,然后植入胚胎以使得活体动物出生。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。

[0027] 鉴于作为整体的公开内容,这些和其他方面对于本领域技术人员将是显而易见的。

## 附图说明

[0028] 图1A至1E显示使用IDLV供体在CD34+细胞中的 $\beta$ -珠蛋白基因座处的裂解和校正。图1A (SEQ ID N0:3) 显示了 $\beta$ -珠蛋白的外显子I DNA 序列的一部分,其显示ZFN靶位点(下划线)在起始密码子(粗体)和镰状突变(粗体,斜体)上。图1B是显示人脐带血CD34+细胞中 $\beta$ -珠蛋白基因座处的靶向裂解的凝胶。细胞在用编码ZFN的体外转录的mRNA电穿孔后三天进行分析。<sup>1</sup>模拟(Mock)<sup>1</sup>代表未处理的CD34+细胞。箭头表示PCR扩增和用 Surveyor® Nuclease(Transgenomic®)酶切后的切割条带。图1C是在 $\beta$ 珠蛋白基因中的相对于外显子1和2的镰状突变处的位点特异性基因校正的示意图。还示出了供体构建体和在由ZFN裂解并通过HDR修复后得到的基因组 DNA 的细节。指示了镰状突变和HhaI RFLP(星号)的位置。用于Surveyor Nuclease和RFLP测定的引物位置如下所示。外显子按比例;斜划线标记表示不按比例的区域。图1D是在ZFN和供体存在下靶向基因修饰 $\beta$ -珠蛋白的 RFLP凝胶。脐带血(CB) CD34+细胞用体外转录的ZFN mRNA电穿孔和/或用包含如凝胶顶部所示的IDLV的供体转导。细胞在处理后4天收获,从供体区域外部进行PCR扩增,用HhaI酶酶切,并在琼脂糖凝胶上分离。箭头显示裂解的产物,表明将RFLP掺入靶位点的基因组中。图1E是显示在指定条件下CD34+细胞中基因修饰百分比的图形。CB CD34+细胞用体外转录的 ZFN mRNA电穿孔并用供体IDLV转导。细胞在处理后三天收获,从供体区域外进行PCR扩增并且qPCR用设计用于特异性检测产生HhaI RFLP的沉默碱基变化的掺入的引物完成,以及归一化为结合在扩增子中的 $\beta$ -珠蛋白基因座的外显子2中的引物,(对于所有条件,n=4)。在用ZFN和供体处理的约 18%的细胞中发现基因修饰。

[0029] 图2A至2D显示了优化用于转导的ZFN和IDLV剂量的结果。图2A、图2B和图2C是显示滴定ZFN mRNA后的结果的图形。用增加量的体外转录的ZFN mRNA对CD34+细胞进行电穿孔,并用恒定量的供体IDLV 转导CD34+细胞。模拟(Mock) 表示未处理的细胞,ZFN表示仅用30 $\mu$

g/mL 的每种 ZFN 电穿孔的样品, IDLV 表示仅用脉冲处理并用 2E+07 TU/mL 供体 IDLV 转导的样品。图 2A 显示了通过用于 RFLP 的 qPCR 测定的电穿孔后 3 天的基因修饰率。图 2B 显示了电穿孔后 1 天的扩增倍数, 图 2C 显示了通过台盼 (trypan) 排除染料确定的在电穿孔后 1 天的细胞活力, (对于所有条件, n = 8)。图 2D 显示了供体整合酶缺陷型慢病毒载体 (IDLV) 的示意图: 相对于长末端重复 (LTR) 中的启动子元件, 将 β-珠蛋白的 1.1kb 区域反向克隆到慢病毒骨架中, 其中 3'LTR 含有在反转录过程中转移到前病毒 DNA 的 5'LTR 中的自我失活 (SIN) 缺失。UTR: 非翻译区, RRE: rev- 反应元件, Int: 内含子。

[0030] 图 3A 至 3C 是显示供体 IDLV 的滴定的图。用恒定量的体外转录的 ZFN mRNA 对 CD34+ 细胞进行电穿孔, 并用递增量的供体 IDLV 转导 CD34+ 细胞。模拟表示未处理的细胞, ZFN 表示仅用 10 μg/mL 的每种 ZFN 电穿孔的样品, IDLV 表示仅用脉冲处理并用 6E+07 TU/mL 供体 IDLV 转导的样品。图 3A 显示了通过用于 RFLP 的 qPCR 测定的电穿孔后 3 天的基因修饰率。图 3B 显示了电穿孔后 1 天的扩增倍数, 图 3C 显示了通过台盼排除染料确定的在电穿孔后 1 天的细胞活力, (对于所有条件 n=6)。

[0031] 图 4A-4F 显示了使用寡核苷酸供体在 CD34+ 细胞中的 β 珠蛋白基因座处使用 ZFN 对 33488/33501 进行校正。图 4A 显示了寡核苷酸指导的基因修饰 (SEQ ID NO:4 和 SEQ ID NO:5) 的示意图。图 4B 是 β- 珠蛋白基因座的 AvrII 酶切的 PCR 扩增子的凝胶。该片段含有天然 AvrII 位点, 其裂解作为 AvrII 酶切的内部对照 (凝胶上的较低条带)。箭头指示 AvrII 裂解产物。在寡聚物供体和 ZFN 的存在下, 存在 15% 的基因校正率。图 4C 显示 SBS# 33501 ZFN 结合位点 (SEQ ID NO:53 (对于 WT) 和 SEQ ID NO:54 (对于 SMS)) 的沉默突变的六个可能位点。镰状突变用斜体显示, 可能的沉默突变位点 (SMS) 以粗体显示。图 4D 显示沉默突变增加了 β- 珠蛋白的基因校正。用 ZFN 和所示的供体寡核苷酸转染 mPB CD34+ 细胞。通过高通量测序测定相关沉默突变的引入。白色条表示插入和缺失 (indel); 灰色条表示基因校正。图 4E 显示沉默突变阻断 ZFN 再裂解。检查具有 indel 的等位基因以用作同源性介导的修饰的证据。显示了还具有 NHEJ 驱动的 indel 的证据的具有基因修饰的等位基因的百分比。图 4F 显示了 ZFN 浓度和供体类型的优化。通过高通量 DNA 测序测定 NHEJ 驱动的 indel (白色条) 和基因校正 (灰色条)。鉴于高通量 DNA 测序的深度, 预期测量误差非常低。WT: 野生型; SMS: 沉默突变位点; oligo: 寡核苷酸。

[0032] 图 5 显示供体长度和链的优化。用设计为将三碱基对序列引入 β- 珠蛋白基因座的 ZFN 和寡核苷酸的组合转染 mPB CD34+ 细胞。收获细胞, 并通过高通量 DNA 测序 (HTS) 测定该短序列的 NHEJ 驱动的 indel 和靶向整合 (“TI”) 的水平。每个样品获得约 4-9,000 个序列读数。

[0033] 图 6A 和 6B 描述部分 HDR 事件, 并揭示在寡核苷酸介导的基因修饰期间的合成从左到右进行。图 6A 显示来自图 4F 的 ZFN 加上 SMS124 供体的高通量 DNA 测序数据, 其中测定了部分 HDR 产物 (SMS1、SMS12、SMS24 和 SMS4 等位基因) 的证据。在该实验中约 41% 的基因修饰中, 6% 的修饰来自 SMS12 事件, 而 SMS24 等位基因几乎完全缺失 (0.03%)。这种不对称意味着 DNA 合成从左到右进行。图 6B 证实了这一点, 显示了使用来自所有三种 ZFN 浓度的数据的平均值, 其中从 30 μg/mL 和 60 μg/mL 样品的事件向上归一化以匹配在 15 μg/mL 样品中观察到的约 41% 的基因修饰。

[0034] 图 7A-7D 显示 β- 珠蛋白基因簇上的靶标裂解分析。图 7A 是 GFP 捕获的说明如下:

K562红白血病细胞用体外转录的mRNA电穿孔，并用表达GFP的IDLV载体转导。将细胞培养60天以稀释所有未捕获的GFP，之后使用荧光激活细胞分选(FACS)分选GFP阳性细胞。对样品进行nrLAM- PCR，并完成载体整合位点分析。显示了使用聚类分析确定的 $\beta$ -珠蛋白和 $\delta$ -珠蛋白的最常见整合位点的概述。HBB：野生型人血红蛋白 $\beta$ ；HBD：人血红蛋白 $\delta$ 。图7B显示在ZFN切割位点周围的 $\beta$ -珠蛋白基因簇中的染色体11上的部分珠蛋白基因座的比对(SEQ ID NO:6至SEQ ID NO:11)。ZFN结合位点加下划线，镰状突变加粗，斜体。HBB：野生型人血红蛋白 $\beta$ ；HBD：人血红蛋白 $\delta$ ；HBE1：人血红蛋白 $\epsilon$ ；HBBP1：人血红蛋白 $\beta$ 假基因1；HBG1：人血红蛋白 $\gamma$ A；HBG2：人血红蛋白 $\gamma$ G。图7C是显示用ZFN体外转录的mRNA处理的或未处理(模拟)的CD34+细胞的Surveyor® Nuclease测定的凝胶。每个指定的 $\beta$ 珠蛋白簇基因在与靶位点具有最高同源性的区域周围扩增，并用Surveyor® Nuclease裂解。通过光密度法对条带进行定量。图7D是显示如图7B所示的SCD患者骨髓样品的Surveyor® Nuclease测定的凝胶。 $\gamma$ -珠蛋白基因中的单核苷酸多态性的杂合性产生背景裂解，即使在未暴露于ZFN的样品(细箭头)中也是如此。

[0035] 图8是显示ZFN和寡核苷酸修饰的HSPC的菌落形成潜力的图形。将mPB CD34+细胞模拟转染，用10 $\mu$ g/mL ZFN转染，或用ZFN和寡核苷酸供体(3 $\mu$ M)转染。培养5天后，将细胞接种在含甲基纤维素和细胞因子的培养基中并分化。生长十五天后，将菌落形成单位(CFU)计数并在形态学上分类为红细胞(E)、爆发形成单位-红细胞(BFU-E)、粒细胞/巨噬细胞(GM)或粒细胞/红细胞/巨噬细胞/巨核细胞(GEMM)。

[0036] 图9A和9B显示了在红细胞分化期间修饰的细胞的稳定性。图9显示mPB CD34+细胞用0 $\mu$ g/mL、3.75 $\mu$ g/mL、7.5 $\mu$ g/mL、15 $\mu$ g/mL或30 $\mu$ g/mL的ZFN和3 $\mu$ M的寡核苷酸供体转染或仅用寡核苷酸供体转染。诱导细胞分化成红细胞，并在0天、6天、12天、15天和18天后取出等分试样。从细胞制备基因组DNA，并通过高通量测序测定经基因修饰的细胞的频度。每个时间点每个样品获得约10-50,000个序列读数。图9B显示工程细胞中珠蛋白的HPLC分析。

[0037] 图10A-10E显示了经ZFN和供体处理过的细胞移植到NSG小鼠中。图10显示了用ZFN+IDLV处理并体外培养的整体移植细胞在电穿孔后7天通过用于RFLP的qPCR确定的基因修饰率。模拟细胞未被处理，(n=3次独立实验)。图10B是显示通过对ZFN+IDLV修饰的CD34+细胞的高通量测序评价的镰状碱基的修饰的图形。 $\beta$ -珠蛋白基因座的测序结果显示包括镰状碱基(T)的改变的野生型以及在切割位点的插入和缺失(indel)的全部对准读数的百分比。与图10A中的样品相同。改变的碱基用白色表示；indel用灰色表示。图10C是显示在移植后5周和8周经移植的小鼠的外周血中的移植效果(engraftment)的图形。通过对来自接受模拟处理的或经ZFN+IDLV处理的细胞的小鼠的细胞进行流式细胞术，将人移植效果确定作为全部hCD45+和mCD45+细胞中的hCD45+细胞的百分比。模拟用空心的菱形表示；ZFN+IDLV用实心的菱形表示。(n=3次独立实验；模拟：n=5，ZFN+IDLV：n=12)。图10D是显示CD34+细胞用01igo、ZFN或ZFN+01igo电穿孔并在体外培养，然后移植入NSG小鼠的图形。镰状碱基和indel的修饰率(n=1)如图10C所示。图10E是显示来自接受经01igo处理的、ZFN处理的或ZFN+01igo处理的细胞的小鼠的细胞的在如图10C的外周血中的移植效果的图形。01igo用圆表示；ZFN用三角形表示；ZFN+01igo用菱形表示。(01igo：n=8，ZFN：n=7，ZFN+01igo：n=9)，n.s.指不显著，\*表示p<0.05，\*\*表示p<0.01。

[0038] 图11A和11B显示了在第8周移植NSG小鼠的谱系分析。在移植后8周，对用模拟

(mock) 和 ZFN 和供体处理的细胞移植的 NSG 小鼠的外周血进行免疫表型分析。使用流式细胞术列举对 B 细胞 (CD19)、T 细胞 (CD3)、造血祖细胞 (CD34)、骨髓祖细胞 (CD33) 和自然杀伤细胞 (CD56) 的标记物呈阳性的人 CD45+ 细胞的百分比。图 11A 是显示来自用模拟处理的或经 ZFN+IDLV 处理的细胞移植的小鼠的细胞的图形, (n=3 次独立实验; 模拟: n=5, ZFN+IDLV: n=12)。图 11B 是显示来自用经 Oligo 处理的、ZFN 处理的或 ZFN+Oligo 处理的细胞移植的小鼠的细胞的图形, (Oligo: n=8, ZFN: n=7, ZFN+Oligo: n=9)。n.s. 指不显著; 星号表示显著性,\* 表示 p<0.05,\*\* 表示 p<0.01。

[0039] 图 12A 至 12D 显示了最终外周血移植效果和谱系分析。图 12A 是显示移植后 16 周移植的小鼠的外周血中的移植效果的图形。通过对来自接受模拟或 ZFN+IDLV 处理的细胞的小鼠的细胞的流式细胞术, 将人移植效果确定为全部 hCD45+ 和 mCD45+ 细胞中的 hCD45+ 细胞的百分比。模拟用空心的菱形表示; ZFN+IDLV 用实心的菱形表示。图 12B 是显示移植后 16 周时移植 NSG 小鼠的外周血的免疫表型分析的图形。使用对来自接受模拟或 ZFN+IDLV 处理的细胞的小鼠的细胞的流式细胞术, 列举了对 B 细胞 (CD19)、T 细胞 (CD3)、造血祖细胞 (CD34)、骨髓祖细胞 (CD33) 和自然杀伤细胞 (CD56) 的标记物呈阳性的 hCD45+ 细胞的百分比。模拟用空心的菱形表示; (n=3 次独立实验; 模拟: n=5, ZFN+IDLV: n=12)。图 12C 是显示来自接受 Oligo 处理的、ZFN 处理的或 ZFN+Oligo 处理的细胞的小鼠的细胞的在如图 12A 的外周血中的移植效果的图形。Oligo 用圆表示; ZFN 用三角形表示; ZFN+Oligo 用菱形表示。图 12D 是显示来自接受 Oligo 处理的、ZFN 处理的或 ZFN+Oligo 处理的细胞的小鼠的细胞的在如图 12B 的外周血中的谱系分析的图形。Oligo 用圆表示; ZFN 用三角形表示; ZFN+Oligo 用菱形表示; (Oligo: n=8, ZFN: n=7, ZFN+Oligo: n=9)。n.s. 指不显著; \* 表示 p<0.05。

[0040] 图 13A-13C 显示基因修饰的细胞存留在 NSG 小鼠中。图 13A 是显示在 16 周时的移植的小鼠的骨髓和脾脏中的、来自接受模拟或 ZFN+IDLV 处理的细胞的小鼠的细胞中的基因修饰率的图形。收获组织, 提取并扩增基因组 DNA, 且用 qPCR 分析并入的 RFLP。模拟用空心的菱形表示; ZFN+IDLV 用实心的菱形表示。图 13B 是显示在通过对图 13A 中描述的样品的高通量测序评价的镰状突变处的校正的图。 $\beta$ -珠蛋白基因座的测序结果显示含有在镰状突变处的修饰碱基的全部对准读数的百分比。模拟用空心的菱形表示; ZFN+IDLV 用实心的菱形表示。 (n=3 次独立实验; 模拟: n=5, ZFN+IDLV: n=12)。图 13C 是显示来自如图 13B 所述的接受 Oligo 处理的、ZFN 处理的、或 ZFN+Oligo 处理的细胞的小鼠的细胞中的在镰状突变处的校正的图形。模拟用空心的菱形表示; ZFN 用三角形表示; ZFN+IDLV 用实心的菱形表示。 (Oligo: n=8, ZFN: n=7, ZFN+Oligo: n=9)。n.s. 指不显著; 星号表示显著性,\* 表示 p<0.05,\*\* 表示 p<0.01。

[0041] 图 14A 至 14D 显示了最终骨髓移植效果和谱系分析。图 14A 是显示移植后 16 周移植的小鼠的骨髓中的移植效果的图形。通过对来自接受模拟或 ZFN+IDLV 处理的细胞的小鼠的细胞的流式细胞术, 将人移植效果确定为全部 hCD45+ 和 mCD45+ 细胞中的 hCD45+ 细胞的百分比。模拟用空心的菱形表示; ZFN+IDLV 用实心的菱形表示。图 14B 是显示移植后 16 周时移植 NSG 小鼠的骨髓的免疫表型分析的图形。使用对来自接受模拟或 ZFN+IDLV 处理的细胞的小鼠的细胞的流式细胞术, 列举了对 B 细胞 (CD19)、T 细胞 (CD3)、造血祖细胞 (CD34)、骨髓祖细胞 (CD33) 和自然杀伤细胞 (CD56) 的标记物呈阳性的 hCD45+ 细胞的百分比。模拟用空心的菱形表示; (n=3 次独立实验; 模拟: n=5, ZFN+IDLV: n=12)。图 14C 是显示来自接受 Oligo 处

理的、ZFN处理的或ZFN+Oligo处理的细胞的小鼠的细胞的在如图14A的骨髓中的移植效果的图形。Oligo用圆表示；ZFN用三角形表示；ZFN+Oligo用菱形表示。图14D是显示来自接受Oligo处理的、ZFN处理的或ZFN+Oligo处理的细胞的小鼠的细胞的在如图12B的骨髓中的谱系分析的图形。Oligo用圆表示；ZFN用三角形表示；ZFN+Oligo用菱形表示；(Oligo:n=8, ZFN: n=7, ZFN+Oligo:n=9)。n.s.指不显著；\*表示p<0.05。

[0042] 图15A至15C显示了镰状骨髓CD34+细胞的分化。图15A是显示SCD患者骨髓CD34+细胞的扩增倍数的图形，所述SCD患者骨髓CD34+细胞用体外转录的ZFN mRNA电穿孔，并用在镰状位置携带WT碱基的供体 IDLV转导，且在红细胞条件下生长。图15B是表示针对接种的每200个细胞识别的每种类型的造血菌落的全部菌落形成单位的图形。图15C是显示形成的菌落相对于接种的细胞总数的百分比的图形。CFU：菌落形成单位；BFU：造粒单元；GEMM：粒细胞、红细胞、单核细胞和巨噬细胞；E：红细胞，GM：粒细胞和巨噬细胞；G：粒细胞；M：单核细胞，星号表示显著性，\*表示p<0.05,\*\*表示p<0.01,(n=2个独立实验；模拟:n=3；仅ZFN: n=4；仅IDLV:n=3,ZFN+IDLV:n=6)。

[0043] 图16A至16D显示镰状骨髓CD34+细胞的功能校正。镰状细胞病患者骨髓CD34+细胞用体外转录的ZFN mRNA电穿孔，并用在镰状位置携带WT碱基的供体 IDLV转导，且在红细胞条件下生长。图16A是显示在电穿孔后12天通过用于并入的RFLP的qPCR分析的基因修饰率的图形。图 16B是显示通过高通量测序评价的镰状突变处的校正的图形。 $\beta$ -珠蛋白基因座的测序结果显示含有在镰状突变处的经校正的WT碱基(A)以及在切割位点的插入和缺失(indel)处的全部对准读数的百分比。经校正的碱基用白色表示；indel用灰色表示。图16C显示了对在培养终止时分化的红系细胞的 HPLC分析。将细胞沉淀(pelleted)，裂解，并通过高效液相色谱(HPLC)分析上清液。左图显示了SCD模拟样品，右图显示了SCD ZFN+IDLV样品。阴影表示HbA:WT成人血红蛋白峰。图16D是由主峰表示的曲线下的总面积中 HbA的百分比的定量图。HbA:WT成人血红蛋白,HbF:胎儿血红蛋白, HbS:镰状血红蛋白。n.s.:不显著,(n=2个独立实验；模拟:n=3；仅 ZFN:n=4；仅IDLV:n=3,ZFN+IDLV:n=6)。

[0044] 图17A和17B显示了使用寡核苷酸供体和不同浓度的锌指核酸酶对47773/47817在CD34+细胞中的 $\beta$ -珠蛋白基因座处的校正。图17A显示了 NHEJ DNA修复的频度。图17B显示了同源性依赖性DNA修复的频度。除了镰状细胞突变之外，寡核苷酸供体包括图4C所示的SMS12突变(黑色柱) 或SMS012突变(灰色柱)。

## 具体实施方式

[0045] 本文公开了用于转导用于基因治疗或基因组工程中的细胞的组合物和方法。特别地，通过靶向分裂进行的外源序列的核酸酶介导的(即ZFN, TALEN、TtAgo或CRISPR/Cas系统)靶向整合或基因组改变在细胞中有效地实现。对于HSC/PC的转导和工程化特别有用的所述方法和组合物也可用于其他细胞类型，以提供包含在珠蛋白基因中的插入和/或缺失的经遗传修饰的细胞。具体地，本发明的方法和组合物可用于通过校正疾病相关等位基因来编辑珠蛋白基因，以治疗和预防镰状细胞病。

### 概述

[0046] 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内

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.Wasserman 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。

### 定义

[0047] 术语“核酸”、“多核苷酸”和“寡核苷酸”可交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。

[0048] 术语“多肽”、“肽”和“蛋白”可交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。

[0049] “结合”是指大分子之间(例如,蛋白与核酸之间)的序列特异性、非共价的相互作用。结合相互作用的组分并非全部都需要具有序列特异性(例如,接触DNA主链中的磷酸酯残基),只要相互作用作为整体具有序列特异性即可。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。

[0050] “结合蛋白”是能够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。

[0051] “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,所述锌指是结合结构域内的氨基酸序列的区域,其结构通过锌离子的配位稳定化。术语锌指 DNA结合蛋白通常缩写为锌指蛋白或ZFP。

[0052] “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蛋白内的其他TALE重复序列的至少一些序列同源性。

[0053] 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)进行。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; 6,534,261 和 8,586,526; 还参见 WO 98/53058; WO 98/53059; WO 98/53060; WO 02/016536 和 WO 03/016496。

[0054] “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。参见例如美国专利No.5,789,538; 美国专利No.5,925,523; 美国专利No.6,007,988; 美国专利No.6,013,453; 美国专利No.6,200,759; 美国专利No.8,586,526; WO 95/19431; WO 96/06166; WO 98/53057; WO 98/54311; WO 00/27878; WO 01/60970; WO 01/88197; WO 02/099084。

[0055] “TtAgo”是被认为参与基因沉默的原核Argonaute蛋白。TtAgo 源自嗜热栖热菌 (*Thermus thermophilus*)。参见，例如，Swarts et al, ibid, G. Sheng et al., (2013) Proc. Natl. Acad. Sci. U.S.A. 111, 652)。“TtAgo系统”是所需的所有组件，包括例如，用于通过TtAgo酶裂解的导向DNA。

[0056] “重组”是指两种多核苷酸之间的遗传信息的交换过程，包括但不限于通过非同源末端连接 (NHEJ) 和同源重组的供体捕获。为了本公开的目的，“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“合成依赖性链复性”，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通常导致靶标分子的序列改变，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。

[0057] 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 (DSB)。DSB可以通过同源定向修复或通过非同源定向修复机制导致缺失和/或插入。缺失可以包括任何数目的碱基对。类似地，插入可以包括任何数目的碱基对，包括例如可选地与断裂区域中的核苷酸序列具有同源性的“供体”多核苷酸的整合。供体序列可被物理整合，或者替代地，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一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 or 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。

[0058] 在任何本文所述的方法中，其他的锌指蛋白、TALEN、TtAgo和/或CRISPR/Cas系统可用于对细胞内其他靶位点的另外的双链裂解。

[0059] 本文所述的任何方法可用于插入任何大小的供体和/或通过靶向整合破坏目标基因表达的供体序列来部分或完全地灭活细胞中的一个或多个靶序列。还提供了具有部分或完全失活的基因的细胞系。

[0060] 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、但不同的序列，从而刺激同源重组以将非同一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有约80%至99% (或其间的任何整数) 序列同一性。在其他实施方案

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。在其他实施方案中,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。

[0061] “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。

[0062] “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。

[0063] “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No. 7,888,121; No. 7,914,796; No. 8,034,598和No. 8,823,618,其全文通过引用并入本文中。

[0064] 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,例如长度为介于2个和100,000,000个之间的核苷酸(或其间或其上的任何整数值),优选长度为介于约100个和100,000个之间的核苷酸(或其间的任何整数),更优选长度为介于约2000个和20,000个之间的核苷酸(或其间的任何值),并且甚至更优选介于约5kb和15kb之间(或其间的任何值)。

[0065] “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。为了本公开的目的,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。

[0066] “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,其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。

[0067] “游离体(episome)”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他结构。游离体的实例包括质粒和某些病毒基因组。

[0068] “可接近区域”是细胞染色质中的位点,其中存在于核酸中的靶位点可以通过识别靶位点的外源分子结合。不希望受任何特定理论的束缚,相信可接近区域是不被包装到核小体结构中的区域。可接近区域的不同结构通常可以通过其对化学和酶探针(例如核酸酶)的敏感性来检测。

[0069] “靶位点”或“靶序列”是定义结合分子将结合的核酸的一部分(假如存在结合的充

分条件的话)的核酸序列。

[0070] “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、生物化学或其他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形式或正常工作的内源分子的不正常工作形式。

[0071] 外源分子尤其可以是小分子,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No.5,176,996和No.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。

[0072] 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,可将人核酸序列引入最初来源于小鼠或仓鼠的细胞系。将外源分子引入植物细胞的方法是本领域技术人员已知的,并且包括但不限于原生质体转化、碳化硅(例如,WHISKERS<sup>TM</sup>)、农杆菌介导的转化、脂质介导的转移(即包括中性和阳离子脂质的脂质体)、电穿孔,直接注射、细胞融合、粒子轰击(例如使用“基因枪”)、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。

[0073] 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、叶绿体或其他细胞器的基因组,或天然存在的游离体核酸。其他的内源分子可包括蛋白,例如,转录因子和酶。

[0074] 如本文所使用的,术语“外源核酸的产物”包括多核苷酸和多肽产物,例如转录产物(多核苷酸,如RNA)和翻译产物(多肽)。

[0075] “融合”分子是其中优选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。

[0076] 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涉及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他地方。

[0077] 为了本公开的目的,“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。

[0078] “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或任何其他类型的RNA)或通过mRNA的翻译产生的蛋白。基因产物还包括通过例如加帽、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、肉豆蔻酰化(myristylation)、和糖基化修饰的蛋白。

[0079] 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、TALE、TtAgo或CRISPR/Cas系统的细胞相比,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。

[0080] “目标区域”是细胞染色质的任何区域,诸如,例如,需要结合外源分子的基因或者该基因内或与该基因相邻的非编码序列。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、尾随序列或内含子内,或在编码区域上游或下游的非转录区域内。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。

[0081] “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),包括干细胞(多潜能的和多能的)。

[0082] 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他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被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,尽管它们是不连续的。

[0083] 关于融合多肽,术语“操作性地连接”可以涉及如下事实:每一组分与其他组分连接执行与其不如此连接将执行的功能相同的功能。例如,关于其中ZFP、TALE、TtAgo或Cas DNA结合结构域与活化结构域融合的融合多肽,如果在融合多肽中,所述ZFP、TALE、TtAgo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,则所述ZFP、TALE、TtAgo或Cas DNA结合结构域和所述活化结构域是操作性连接的。当融合多肽中的ZFP、TALE、TtAgo或Cas DNA结合结构域与裂解结构域融合时,如果在融合多肽中所述ZFP、TALE、TtAgo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,而裂解结构域能够裂解靶位点附近的DNA,则所述ZFP、TALE、TtAgo或Cas DNA结合结构域和所述裂解结构域是操作性连接的。

[0084] 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No.5,585,245和PCT WO 98/44350。

[0085] “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。

[0086] 术语“受试者”和“患者”可交换使用并且是指哺乳动物,例如人患者和非人灵长类动物,以及实验动物,例如兔、狗、猫、大鼠、小鼠和其他动物。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核酸酶、供体和/或经遗传修饰的细胞的任何哺乳动物患者或受试者。本发明的受试者包括具有病症的那些。

[0087] “干性(Stemness)”是指任何细胞以干细胞样方式作用的相对能力,即,任何特定干细胞可能有的全能性、多能性或寡能性以及扩展或无限自我更新的程度。

#### 增强干细胞扩增的因子

[0088] 可以在本发明的实践中使用增强干细胞扩增和/或分化的任何一个或多个因子。可以将因子直接引入细胞(例如,作为编码因子的基因)和/或可以引入周围的培养基(包括饲养层和其他固体基质)以影响细胞。这些因子例如在诱导核酸酶介导的修饰之前、期间或之后的培养条件中的使用增加核酸酶介导的干细胞修饰的速率。

[0089] 可以使用的因子的非限制性实例包括SR1(一种芳基烃受体拮抗剂)、dmPGE2(一种前列腺素)、在文库筛选中识别的化合物UM171和UM729(参见Pabst et al (2014) Nat Meth 11:436-442)、雷帕霉素(参见Wang et al (2014) Blood pii:blood-2013-12-546218)、血管生成素样蛋白(“Angptls”,例如Notch/delta/ANGPTL5(参见Zhang Zhang et al (2008) Blood.111 (7):3415-3423),Angptl2,Angptl3,Angptl5,Angptl7,和Mfap4)、铜螯合剂四亚乙基戊胺(TEPA,参见de Lima et al (2008) Bone Mar Trans 41:771-778)、组蛋白脱乙酰酶(HDAC)抑制剂例如丙戌酸(参见Chaurasia et al (2014) J Clin Invest 124:2378-2395)、IGF结合蛋白2(IGFBP2)、烟酰胺(参见Horwitz et al (2014) J. Clin. Invest 124:3121-3128)、Tat-myc(参见WO2010025421)和tat-Bc12(参见 WO2014015312)融合蛋白、MAPK14/p38a Ly2228820(参见Baudet et al (2012) Blood 119 (26):6255-6258)、自我更新基因的产物如HOXB4、OCT3/4脐带血和/或MSC衍生的饲养层或称为E4+EC的离体血管隙状共培养系统(参见Butler et al (2012) Blood.120 (6):1344-1347)、细胞因子、(非限制性实例,Stemspan<sup>TM</sup>CC110,CC100, and/or H3000 (Stemcell<sup>TM</sup>technologies),Flt-3ligand,SCF,TPO)。

[0090] 在一些实施方案中,因子包括StemRegenin(SR1,参见例如美国专利No.8,741,640;Boitano et al, (2010) Science 329 (5997):1345-1348),一种芳烃受体(AhR)拮抗剂,其促进离体CD34+细胞的扩增使用在本文所述的方法和组合物中。在其他实施方案中,所述因子包括作为干细胞更新的激动剂的UM171(参见Fares et al (2013) Blood:122 (21))。在

其他方面,所述因子包括一种或多种前列腺素,例如dmPGE2。参见,例如,美国专利No. 8,168,428;North et al (2007) Nature 447 (7147) :1007-1011)。在一些方面,所述因子包括一种或多种激素,例如血管生成素样蛋白(“Angptls”,例如 Angpt12、Angpt13、Angpt15、Angpt17和Mfap4)和IGF结合蛋白2 (IGFBP2)。参见例如美国专利No.7,807,464;Zhang et al (2008) 111 (7) :3415-3423)。在其他方面,所述因子包括自我更新基因的一种或多种蛋白产物,例如HOXB4 或使用的OCT。参见例如美国专利No.8,735,153;Watts et al (2012) Exp Hematol.40 (3) :187-196)。替代地,这些基因可以在培养基和/或干细胞中瞬时表达。

[0091] 影响干细胞扩增的因子还可以包括细胞支持方法,包括但不限于衍生自基质细胞和/或MSC衍生的细胞的饲养层。参见例如Breems et al (1998) Blood 91 (1) :111-117和Magin et al., (2009) Stem Cells Dev.2009Jan- Feb;18 (1) :173-86。

[0092] 可以使用任何合适量的一种或多种增强干细胞扩增的因子,只要其有效增加核酸酶活性和核酸酶介导的基因组修饰即可。所使用的具体浓度可以容易地由本领域技术人员确定。在某些实施方案中,使用介于0.1nM至 100μM之间的浓度,例如使用介于0.5μM至25μM之间的浓度,包括其间的任何量(例如,1μM至20μM,3μM至10μM等)。

### 融合分子

[0093] 本文描述可用于在细胞中(特别是干细胞中)的所选的靶基因的裂解的组合物,例如核酸酶。在某些实施方案中,融合分子的一种或多种组分(例如核酸酶)是天然存在的。在其他实施方案中,融合分子的一种或多种组分(例如核酸酶)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的。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))。在其他实施方案中,核酸酶包含异源DNA结合和裂解结构域(例如,锌指核酸酶;TAL-效应子结构域DNA结合蛋白;具有异源裂解结构域的大范围核酸酶DNA结合结构域)。

#### A.DNA结合分子

[0094] 在某些实施方案中,本文所述的组合物和方法采用用于结合供体分子和/或结合细胞基因组中的目标区域的大范围核酸酶(归巢内切核酸酶)DNA结合分子(也称为DNA结合结构域)。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:LAGLIDADG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No.5,420,032;美国专利No.6,833,252;Belfort et al. (1997) Nucleic Acids Res.25:3379-3388;Dujon et al. (1989) Gene 82:115-118;Perler et al. (1994) Nucleic Acids Res.22, 1125-1127;Jasin (1996) Trends Genet.12:224-228;Gimble et al. (1996) J.Mol.Biol.263:163-180;Argast et al. (1998) J.Mol. Biol.280:345-353以及新英格兰生物实验室目录(New England Biolabs catalogue)。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 et al. (2002) Molec. Cell 10:895-905;Epinat et al. (2003) Nucleic Acids Res.31:2952-2962;Ashworth et al. (2006) Nature 441:656-659;Paques et al. (2007) Current Gene Therapy 7:49- 66;美国专利公布No.20070117128。在核酸酶作为总体的情形下,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。

[0095] 在其他实施方案中,在本文所述的方法和组合物中使用的一种或者多种核酸酶的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No.8,586,526,其全文通过引用并入本文中。已知黄单胞菌(*Xanthomonas*)属的植物病原细菌导致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(TAL)效应子,其模拟植物转录活化子并操纵植物转录组(参见 Kay et al, (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TAL效应子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 et al, (1989) *Mol Gen Genet* 218: 127-136和W02010079430)。TAL效应子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结合特异性是关键的。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 et al, (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 et al, (2007) *Appl and Envir Micro* 73(13):4379-4384)。这些基因的核苷酸序列彼此具有98.9%同一性,但差异是在hpx17的重复结构域中缺失1,575个碱基对。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。参见美国专利No.8,586,526,其全文通过引用并入本发明。

[0096] 这些TAL效应子的特异性取决于在串联重复序列中发现的序列。重复的序列包含约102个碱基对,重复序列通常彼此具有91-100%的同源性(Bonas et al,同上)。重复序列的多态性通常位于位置12和13,并且在12和 13位的高变双残基(RVD)的同一性与TAL效应子的靶序列中的连续核苷酸的同一性之间似乎存在一对一的对应关系(参见Moscou和Bogdanove, (2009) *Science* 326:1501和Boch et al (2009) *Science* 326:1509-1512)。实验上,已经确定用于这些TAL效应子的DNA识别的天然编码,使得位置12和13处的HD 序列导致与胞嘧啶(C)结合,NG结合T,NI与A、C、G或T结合,NN结合 A或G,而ING结合T。这些DNA结合重复序列已组装成具有重复序列的新组合和数目的蛋白,以制备能够与新序列相互作用并激活植物细胞内的非内源性报道基因的表达的人工转录因子(Boch et al.,同上)。工程改造的TAL蛋白已经连接到FokI裂解半结构域以产生TAL效应子结构域核酸酶融合体(TALEN)。参见例如美国专利No.8,586,526;Christian et al ((2010) <Geneticsepub 10.1534/genetics.110.120717)。在某些实施方案中, TALE结构域包含如在美国专利No.8,586,526中所述的N帽和/或C帽。

[0097] 在某些实施方案中,用于细胞基因组的体外裂解和/或靶向裂解的一种或多种核酸酶的DNA结合结构域包含锌指蛋白。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No. 6,453,242; No. 6,534,261; No. 6,599,692; No. 6,503,717; No. 6,689,558; No. 7,030,215; No. 6,794,136; No. 7,067,317; No. 7,262,054; No. 7,070,934; No. 7,361,635; No. 7,253,273; 和美国专利公布No. 2005/0064474; No. 2007/0218528; No. 2005/0267061, 全部通过引用全文并入本文中。

[0098] 相比于天然存在的锌指蛋白, 工程改造的锌指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)合理的设计和各种类型的选择。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结合。参见例如共同拥有的美国专利 6,453,242和6,534,261, 其全文通过引用并入本文中。

[0099] 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。

[0100] 另外, 如这些和其他参考文献中所公开的, 锌指结构域和/或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No. 6,479,626; No. 6,903,185; 和No. 7,153,949。本文描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。

[0101] 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No. 6,140,081; No. 5789, 538; No. 6,453,242; No. 6,534,261; No. 5,925,523; No. 6,007,988; No. 6,013,453; No. 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。

[0102] 另外, 如这些和其他参考文献中所公开的, 锌指结构域和/或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No. 6,479,626; No. 6,903,185; 和No. 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。

[0103] 在某些实施方案中, DNA结合结构域是CRISPR/Cas核酸酶系统的一部分, 包括例如单导向RNA(sgRNA)。参见例如美国专利No. 8,697,359和美国专利公布No. 20150056705。编码系统的RNA组分的CRISPR(规律成簇间隔短回文重复序列)基因座和编码蛋白的cas(CRISPR相关)基因座(Jansen et al., 2002. Mol. Microbiol. 43:1565-1575; Makarova et al., 2002. Nucleic Acids Res. 30:482-496; Makarova et al., 2006. Biol. Direct 1:7; Haft et al., 2005. PLoS Comput. Biol. 1:e60)构成CRISPR/Cas核酸酶系统的基因序列。微生物宿主中的CRISPR基因座含有CRISPR相关(Cas)基因以及能够对CRISPR介导的核酸裂解的特异性编程的非编码RNA元件的组合。

[0104] II型CRISPR是最好表征的系统之一, 并且在四个连续步骤中进行靶向DNA双链断裂。首先, 两个非编码RNA, 即pre-crRNA阵列和tracrRNA, 从CRISPR基因座转录。第二, tracrRNA与pre-crRNA的重复区杂交, 并将pre-crRNA的处理介导到含有单独间隔序列的成熟crRNA中。第三, 成熟的crRNA:tracrRNA复合物通过在crRNA上的间隔区和在靶DNA上的前

间区 (protospacer) 之间的Watson-Crick碱基配对将Cas9引导到靶DNA，靶DNA邻近前间区序列邻近基序 (PAM)，对于靶识别的额外要求。最后，Cas9介导靶DNA的裂解以在前间区内产生双链断裂。CRISPR/Cas系统的活化包括三个步骤：(i) 将外来DNA序列插入到CRISPR阵列中以防止未来在称为“适应”的过程中的攻击，(ii) 相关蛋白的表达，以及阵列的表达和加工，随后是 (iii) RNA介导的对外来核酸的干扰。因此，在细菌细胞中，几种所谓的“Cas”蛋白参与CRISPR/Cas系统的天然功能，并且在诸如插入外来DNA 等之类的功能中起作用。

[0105] 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以及其片段，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。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。Cas蛋白，其包括Cas蛋白或其片段，以及Cas蛋白或其片段的衍生物，可以从细胞获得或化学合成或通过这两种方法的组合来获得。细胞可以是天然产生Cas蛋白的细胞，或是天然产生Cas蛋白并且被遗传工程改造以便以更高的表达水平产生内源性Cas蛋白或从外源导入的核酸产生Cas蛋白的细胞，所述核酸编码与内源性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。

[0106] 在一些实施方案中，DNA结合结构域是TtAgo系统的一部分 (参见Swarts et al., 同上；Sheng et al., 同上)。在真核生物中，基因沉默由 Argonaute (Ago) 家族的蛋白介导。在这个范例中，Ago与小的 (19-31nt) RNA 结合。这种蛋白-RNA沉默复合物通过小RNA和靶标之间的Watson-Crick碱基配对识别靶RNA，并通过核内溶解方式 (endonucleolytically) 裂解靶 RNA (Vogel (2014) Science 344:972-973)。相反，原核Ago蛋白结合小的单链 DNA片段并且可能起到检测和去除外源(通常是病毒)DNA的作用 (Yuan et al., (2005) Mol. Cell 19, 405; Olovnikov, et al. (2013) Mol. Cell 51, 594; Swarts et al., 同上)。示例性的原核生物Ago蛋白包括来自风产液菌 (*Aquifex aeolicus*)、类球红细菌 (*Rhodobacter sphaeroides*) 和嗜热栖热菌 (*Thermus thermophilus*) 的那些。

[0107] 最好表征的原核生物Ago蛋白之一是来自嗜热栖热菌的原核生物Ago蛋白 (TtAgo；Swarts et al., 同上)。TtAgo与15nt或13-25nt的具有5' 磷酸基团的单链DNA片段结合。由TtAgo结合的这种“导向DNA”用于引导蛋白-DNA复合物结合DNA的第三方分子中的Watson-Crick互补DNA序列。一旦这些导向DNA中的序列信息已使得能识别靶DNA，则TtAgo- 导向DNA复合物裂解靶DNA。这样的机制也由TtAgo- 导向DNA复合物的结构支持，同时与其靶DNA结合 (G. Sheng et al., 同上)。来自类球红细菌 (RsAgo) 的Ago具有相似的性质 (Olivnikov et al., 同上)。

[0108] 任意DNA序列的外源导向DNA可以加载到TtAgo蛋白上 (Swarts et al., 同上)。由于TtAgo裂解的特异性由导向DNA引导，因此由外源的、研究者特定的导向DNA形成的TtAgo-DNA复合物将引导TtAgo靶 DNA裂解至互补的研究者特定的靶DNA。以这种方式，可以在DNA中产生靶向的双链断裂。使用TtAgo- 导向DNA系统(或来自其他生物体的直向同源 Ago- 导向DNA系统)使得在细胞内的基因组DNA能靶向裂解。这种裂解可以是单链的或双链的。对于哺乳动物基因组DNA的裂解，优选使用为了在哺乳动物细胞中表达而优化的TtAgo密码子的

版本。此外,可能优选的是,使用体外形成的TtAgo-DNA复合物处理细胞,其中TtAgo蛋白与细胞穿透肽融合。此外,可能优选的是,使用已通过诱变改变的TtAgo蛋白的版本,以在37摄氏度具有改善的活性。Ago-RNA介导的DNA裂解可以用于影响大量的包括基因敲除、靶向基因添加、基因校正、本领域的用于开发DNA断裂的标准的靶向基因缺失使用技术等在内的结果。

[0109] 因此,核酸酶包含DNA结合结构域,其特异性结合到期望插入供体(转基因)的任何基因中的靶位点。

#### B. 裂解结构域

[0110] 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,其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,包括用于多种生物体中的基因组修饰。参见例如美国专利公布20030232410; 20050208489; 20050026157; 20050064474; 20060188987; 20060063231; 和国际公布WO 07/014275。同样,TALEN DNA结合结构域已经与核酸酶结构域融合以产生TALEN。参见例如美国公开No.5,886,526。也已经描述了包含结合 DNA并与裂解结构域(例如Cas结构域)缔合以诱导靶向裂解的单导向 RNA(sgRNA)的CRISPR/Cas核酸酶系统。参见例如美国专利No.8,697,359和 No.8,932,814以及美国专利公布No.20150056705。

[0111] 如上文所示,裂解结构域可与DNA结合结构域是异源的,例如:锌指DNA结合结构域和来自核酸酶的裂解结构域;TALEN DNA结合结构域和来自核酸酶的裂解结构域;sgRNA DNA结合结构域和来自核酸酶的裂解结构域(CRISPR/Cas);和/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。异源裂解结构域可源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。裂解DNA的其他酶是已知的(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。这些酶中的一种或多种(或其功能片段)可用作裂解结构域和裂解半结构域的来源。

[0112] 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。替代地,可使用包含两个裂解半结构域的单一蛋白。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彼此之间相互布置,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的空间方位中以使得裂解半结构域例如通过二聚合能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。

[0113] 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 et

al., (1992) Proc.Natl.Acad.Sci.USA 89:4275-4279; Li et al., (1993) Proc.Natl. Acad.Sci.USA 90:2764-2768; Kim et al., (1994a) Proc.Natl.Acad.Sci.USA 91:883-887; Kim et al. 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。

[0114] 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 et al., (1998) Proc.Natl. Acad.Sci.USA 95:10, 570-10,575。因此,为了本公开的目的,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Fok I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,还可以使用含有锌指结合结构域和两个Fok I裂解半结构域的单一多肽分子。用于使用锌指Fok I融合物的靶向裂解和靶序列改变的参数在本公开内容的其他地方提供。

[0115] 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。

[0116] 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。其他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 et al., (2003) Nucleic Acids Res.31:418-420。

[0117] 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No.20050064474; No.20060188987; No.20070305346 和 20080131962 中所述,其全部的公开内容都全文通过引用并入本文中。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。

[0118] 可以使用多于一个的突变的裂解结构域,例如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”);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);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Fok I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);和/或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Fok I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)。参见例如美国专利No.7,914,796; No.8,034,598 和 No.8,623,618,其公开内容通过引用整体并入本文以用于所有目的。在其他实施方案中,工程改造的裂解半结构域包含“Sharkey”和/或“Sharkey”突变(参见Guo et al., (2010) J.Mol.Biol.400 (1) :96-107)。

[0119] 替代地,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No.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。

[0120] 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,如美国专利No.8,563,314中所述。

[0121] Cas9相关CRISPR/Cas系统包含两种RNA非编码组分:含有由同一的定向重复序列(DR)隔开的核酸酶引导序列(间隔序列)的tracrRNA和 pre-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 et al, (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pre-crRNA。在其他实施方案中,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)。

#### 靶位点

[0122] 如上文所详细描述的,DNA结合结构域可以经过工程改造以结合于基因座任何所选序列。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。

[0123] 合适的靶基因的非限制性实例是贝塔( $\beta$ )珠蛋白基因(HBB)、伽马( $\gamma$ )珠蛋白基因(HBG1)、B细胞淋巴瘤/白血病11A(BCL11A)基因、Kruppel样因子1(KLF1)基因、CCR5基因、CXCR4基因、PPP1R12C(AAVS1) 基因、次黄嘌呤磷酸核糖基转移酶(HPRT)基因、白蛋白基因、因子VIII基因、因子IX基因、富含亮氨酸的重复激酶2(LRRK2)基因、亨廷顿(Htt)基因、视紫红质(RHO)基因、囊性纤维化跨膜传导调节剂(CFTR)基因、表面活性蛋白B基因(SFTPB)、T细胞受体 $\alpha$ (TRAC)基因、T细胞受体 $\beta$ (TRBC)基因、程序性细胞死亡1(PD1)基因、细胞毒性T淋巴细胞抗原4(CTLA-4)基因、人白细胞抗原(HLA)A基因、HLA B基因、HLA C基因、HLA-DPA基因、HLA-DQ基因、HLA-DRA基因、LMP7基因、与抗原加工(TAP)1基因相关的转运体、TAP2基因、tapasin基因(TAPBP)、II类主要组织相容性复合体反式激活因子(CITA)基因、肌营养不良蛋白基因(DMD)、糖皮质激素受体基因(GR)、IL2RG基因、Rag-1基因、RFX5基因、FAD2基因、FAD3基因、ZP15基因、KASII基因、MDH基因和/或EPSPS基因。

[0124] 在某些实施方案中,核酸酶靶向“安全港”基因座,例如人类细胞中的AAVS1、HPRT、白蛋白和CCR5基因,和鼠细胞中的Rosa266(参见例如美国专利No.7,888,121;No.7,972,854;No.7,914,796;No.7,951,925; No.8,110,379;No.8,409,861;No.8,586,526;美国专利公布20030232410; 20050208489;20050026157;20060063231;20080159996;201000218264; 20120017290;20110265198;20130137104;20130122591;20130177983和20130177960)和植物中的Zp15基因座(参见美国专利US 8,329,986)。

#### 供体

[0125] 在某些实施方案中,本公开涉及干细胞的基因组的核酸酶介导的修饰。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”),例如以缺失指定区域和/或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或可以通过非同源性定向修复机制整合。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。此外,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。

[0126] 与核酸酶一样,可以将供体以任何形式引入。在某些实施方案中,以mRNA形式引入供体以消除修饰的细胞中的残留病毒。在其他实施方案中,可以通过本领域已知的方法使用DNA和/或病毒载体引入供体。参见例如美国专利公布No. 20100047805和No. 20110207221。供体可以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 et al. (1987) Proc. Natl. Acad. Sci. USA 84: 4959- 4963; Nehls et al. (1996) Science 272: 886-889。用于保护外源多核苷酸免于降解的其他方法包括但不限于添加末端氨基和使用经修饰的核苷酸连锁,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。

[0127] 在某些实施方案中,供体包括长度大于1kb、例如介于2kb和 200kb之间、介于2kb和10kb之间(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)。供体也可以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2个靶位点,例如用于一对ZFN、TALEN、TtAgo或CRISPR/Cas核酸酶。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。

[0128] 可以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内部已插入供体的内源基因的表达的启动子。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。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。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。

[0129] 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。替代地,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。

[0130] 本文所述的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。

[0131] 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。

[0132] 在一些实施方案中，供体还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。

[0133] 在某些实施方案中，外源序列可包含使得能选择已经历靶向整合的细胞的标记基因(上文所述)，和编码其他功能性的连接序列。标记基因的非限制性实例包括GFP、药物选择标记等。

[0134] 在某些实施方式中，转基因可包括例如置换突变内源序列的野生型基因。例如，野生型(或者其他功能型)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。转基因可插入在内源基因座处，或者可以可选地靶向安全港基因座。

[0135] 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。

[0136] 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/或多腺昔酸化信号。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)。示例性剪接受体位点序列是本领域技术人员已知的，并且仅举例而言，包括CTGACCTCTTCTCTCCTCCCCACAG, (SEQ ID NO:1) (来自人HBB基因) 和 TTTCTCTCCACAG (SEQ ID NO:2) (来自人免疫珠蛋白- $\gamma$ 基因)。

[0137] 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(mirRNA)的序列也可用于靶向插入。

[0138] 在其他实施方案中，供体核酸可包含对于其他核酸酶设计是特异性靶位点的非编码序列。随后，其他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，从而使得能在特定的目标基因座处或在安全港基因座处特性堆叠。

### 细胞

[0139] 因此，本文提供了经遗传修饰的细胞，例如包含失活基因和/或转基因的干细胞，包括通过本文所述的方法产生的细胞。使用一种或多种核酸酶将转基因以靶向方式整合到细胞的基因组中。与随机整合不同，靶向整合确保转基因整合到指定基因中。转基因可以整合在靶基因的任何位置。在某些实施方案中，转基因整合在核酸酶结合和/或裂解位点处或附近，例如整合在裂解和/或结合位点的上游或下游且介于1个-300个碱基对(或其间的任何数目的碱基对)之间，优选在裂解和/或结合位点的任一侧介于1个至100个碱基对(或其间的任何数目的碱基对)之间，甚至更优选在裂解和/或结合位点的任一侧介于1至50个碱基对之间。在某些实施方案中，整合序列不包括任何载体序列(例如，病毒载体序列)。在某些实施方案中，细胞包含由本文所述的核酸酶制备的修饰(例如插入和/或缺失)，使得修饰在 $\beta$ -珠蛋白基因的外显子内，例如在外显子1、2或3内。在某些实施方案中，修饰校正 $\beta$ 珠蛋白基因的外显子1中的镰状细胞突变。在某些实施方案中，修饰是在或接近(例如，1个-300个碱基对或其间的任何数目的碱基对) SEQ ID NO:23或SEQ ID NO:24。在其他实施方案中，修饰是SEQ ID NO:23或SEQ ID NO:24的1个-100个(或其间的任何数目的碱基对)。在某些

实施方案中,修饰是在SEQ ID NO:23和/或SEQ ID NO:24内,例如是在SEQ ID NO:23或SEQ ID NO:24 内的1个或多个碱基对的修饰。

[0140] 任何细胞类型可以如本文所述进行遗传修饰以包含转基因,包括但不限于细胞和细胞系。本文所述的经遗传修饰的细胞的其他非限制性实例包括T细胞(例如CD4+、CD3+、CD8+等);树突细胞;B细胞;自体细胞(例如,患者衍生的)。在某些实施方案中,细胞是干细胞,包括异源多潜能(pluripotent)干细胞、全能干细胞或多能(multipotent)干细胞(例如,CD34+细胞、诱导多潜能干细胞(iPSC)、胚胎干细胞等)。在某些实施方案中,本文所述的细胞是源自患者的干细胞。

[0141] 本文所述的细胞可用于治疗和/或预防患有病症的受试者的所述疾症,例如通过离体治疗进行。核酸酶修饰的细胞可以扩增,然后使用标准技术重新引入患者。参见,例如 Tebas et al (2014) New Eng J Med 370 (10) :901。在干细胞的情况下,在输入受试者后,也发生这些前体体内分化成表达功能性蛋白(来自插入的供体)的细胞。还提供了包含本文所述细胞的药物组合物。此外,细胞可以在施用给患者之前冷冻保存。

#### 递送

[0142] 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递送。在某些实施方案中,核酸酶和/或供体在体内递送。在其他实施方案中,将核酸酶和/或供体递送至分离的细胞(例如,自体或异源干细胞),以提供用于离体递送至患者的经修饰细胞。

[0143] 如本文所述的递送核酸酶的方法描述于例如美国专利No. 6,453,242;No.6,503,717;No.6,534,261;No.6,599,692;No.6,607,882; No.6,689,558;No.6,824,978;No.6,933,113;No.6,979,539;No.7,013,219;和 No.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。

[0144] 如本文所述的核酸酶和/或供体构建体也可使用任何核酸递送机制,包括裸DNA和/或RNA(例如mRNA)以及含有编码组分中的一种或多种的序列的载体递送。可使用任何载体系统,包括但不限于质粒载体、DNA小环、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以及其组合。也参见美国专利No.6,534,261; No. 6,607,882;No.6,824,978;No.6,933,113;No.6,979,539;No.7,013,219;和No. 7,163,824,以及美国专利申请No.14/271,008,其以全文引用的方式并入本文中。此外,将显而易见的是,任何这些系统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递送系统上或不同递送系统上。当使用多个系统时,每个递送系统可包含编码一个或多个核酸酶和/或供体构建体的序列(例如,编码一种或多种核酸酶的mRNA和/或携带一种或多种供体构建体的mRNA或AAV)。

[0145] 可使用常规的基于病毒和非病毒的基因转移方法来将编码核酸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、DNA小环、裸核酸,和与递送媒介物(例如脂质体、脂质纳米粒子、聚乳酸-乙醇酸纳米粒子、聚胺络合剂或泊洛沙姆)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Science 256:808- 813 (1992);Nabel&Felgner,TIBTECH 11:211-217 (1993);Mitani&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&Perricaudet, British Medical Bulletin 51 (1) :31-44 (1995) ;Haddada et al.,in Current Topics in Microbiology and Immunology Doerfler and Böhm (eds.) (1995) ;and Yu et al., Gene Therapy 1:13-26 (1994) 。

[0146] 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、裸RNA、加帽RNA、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。

[0147] 其他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(参见例如美国专利No.6,008,336)。脂转染描述于例如美国专利No.5,049,386;No. 4,946,787;和No.4,897,355中并且脂转染试剂可商业销售(例如, Transfectam<sup>TM</sup>和Lipofectin<sup>TM</sup>)。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。

[0148] 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 et al.,Cancer Gene Ther.2:291-297 (1995) ;Behr et al.,Bioconjugate Chem. 5:382-389 (1994) ;Remy et al.,Bioconjugate Chem.5:647-654 (1994) ;Gao et al., Gene Therapy 2:710-722 (1995) ;Ahmad et al.,Cancer Res.52:4817-4820 (1992) ;美国专利No.4,186,183, No.4,217,344, No.4,235,871, No.4,261,975, No. 4,485,054, No.4,501,728, No.4,774,085, No.4,837,028, 和No.4,946,787)。

[0149] 其他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 et al (2009) Nature Biotechnology 27 (7) :643)。

[0150] 使用基于RNA或DNA病毒的系统来递送编码工程改造的 CRISPR/Cas系统的核酸,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至受试者(离体)。常规的用于递送CRISPR/Cas系统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。

[0151] 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)、猿猴免疫缺陷病毒(SIV)、人免疫缺陷病毒(HIV)和其组合的那些(参见例如Buchscher et al., J.Virol.66:2731-2739(1992); Johann et al., J.Virol. 66:1635-1640(1992); Sommerfelt et al., Virol.176:58-59(1990); Wilson et al., J. Virol.63:2374-2378(1989); Miller et al., J.Virol.65:2220-2224(1991); PCT/US94/05700)。

[0152] 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 et al., Virology 160:38-47(1987);美国专利No.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No.5,173,414; Tratschin et al., Mol.Cell.Biol.5:3251-3260(1985); Tratschin, et al., Mol.Cell.Biol.4:2072-2081(1984); Hermonat&Muzyczka, PNAS 81:6466-6470(1984); 以及Samulski et al., J.Virol.63:03822-3828 (1989)。可以使用任何AAV血清型,包括AAV1、AAV3、AAV4、AAV5、AAV6和AAV8、AAV 8.2、AAV9和AAV rh10以及假型AAV,例如AAV2/8,AAV2/5和AAV2/6。

[0153] 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。

[0154] 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 et al., Blood 85:3048-305 (1995); Kohn et al., Nat.Med.1:1017-102 (1995); Malech et al., PNAS 94:22 12133-12138 (1997))。PA317/pLASN是用于基因疗法试验中的第一治疗载体。(Blaese et al., 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 et al., Immunol Immunother.44 (1) :10-20 (1997); Dranoff et al., Hum.Gene Ther.1:111-2 (1997))。

[0155] 重组腺相关病毒载体(rAAV)是基于有缺陷的和非致病性细小病毒腺相关2型病毒的一种有前途的可选基因递送系统。所有载体都源自仅保留侧接转基因表达盒的AAV 145碱基对(bp)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而导致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 et al., Lancet 351:9117 1702-3 (1998), Kearns et al., Gene Ther.9:748-55 (1996))。也可以根据本发明使用其他AAV血清型,包括AAV1、AAV3、AAV4、AAV5、AAV6、AAV8、AAV9和AAVrh10以及其所有变体。

[0156] 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载体具有大的携载能力。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 et al., 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他实例包括Rosenecker et al., Infection 24:1 5-10 (1996); Sterman et al., Hum. Gene Ther. 9:7 1083-1089 (1998); Welsh et al., Hum. Gene Ther. 2:205-18 (1995); Alvarez et al., Hum. Gene Ther. 5:597-613 (1997); Topf et al., Gene Ther. 5:507-513 (1998); Sterman et al., Hum. Gene Ther. 7:1083-1089 (1998)。

[0157] 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以及(如果适用的话)随后整合至宿主中的最小病毒序列,其他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(ITS)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,其含有编码其他AAV基因(即rep和cap)的辅助质粒,但缺乏ITS序列。细胞系也被用腺病毒作为辅助而感染。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ITS序列,因此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。

[0158] 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与病毒外壳蛋白作为融合蛋白的配体,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 et al., 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融合的人调蛋白(Heregulin),并且重组病毒感染某些表达人表皮生长因子受体的人乳腺癌细胞。这种原理可以扩展到其他病毒-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。

[0159] 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,通常在选择已合并有载体的细胞后。

[0160] 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。

[0161] 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 et al. (1996) Proc. Natl. Acad. Sci. USA 93:11382-11388; Dull et al. (1998) J. Virol. 72:8463-8471; Zuffery et al. (1998) J. Virol. 72:9873-9880; Follenzi et

a1. (2000) Nature Genetics 25:217-222; 美国专利公布No. 2009/054985。

[0162] 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,存在广泛的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)。

[0163] 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AAV携带,而一种或多种核酸酶可由mRNA携带。此外,不同的系统可通过相同或不同的途径(肌肉内注射、尾静脉注射、其他静脉内注射、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。

[0164] 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、稳定剂或其他增强药物组合物的有效性的试剂。

#### 试剂盒

[0165] 还提供用于执行上述方法中的任何一种的试剂盒。试剂盒通常含有编码一种或多种核酸酶、一种或多种影响干细胞扩增的因子和/或如本文所述的供体多核苷酸的多核苷酸以及用于施用所述因子的说明书,所述因子影响干细胞进入内部引入了核酸酶和/或供体多核苷酸(或周围培养基)的细胞。试剂盒还可以含有细胞、用于转化细胞的缓冲液、用于细胞的培养基和/或用于进行测定的缓冲液。通常,试剂盒还包含标签,所述标签包括任何材料,例如附接或以其他方式伴随试剂盒的其他组分的说明书、包装或广告传单。

[0166] 下列实施例涉及本公开的示例性实施方案,其中核酸酶包含一种或者多种ZFN或一种或者多种TALEN。应理解,这只是为了示例目的并且可使用其他核酸酶,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结合结构域与异源裂解结构域、大型TAL、紧密TALEN以及核酸酶系统(如使用工程改造的单导向RNA的TtAgo和CRISPR/Cas)的融合体。

#### 实施例

##### 实施例1:锌指蛋白核酸酶(ZFN)的设计、构建和一般表征

[0167] 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 et al. (2005) Nature 435 (7042) :646-651,Perez et al (2008) Nature Biotechnology 26 (7) :808-816中所述,以及如美国专利No. 6,534,261中所述并且被测试以供结合。关于对人 $\beta$ 珠蛋白基因座具有特异性的ZFN和 TALEN,参见共同拥有的美国专利No.7,888,121和美国专利公布 No.20130137104与No.20130122591。

##### 实施例2:珠蛋白特异性ZFN的活性

[0168] 使用靶向人珠蛋白基因座的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示于下表1。靶位点(DNA靶位点以大写字母指示;非接触核苷酸以小写字母指示)示于表2。

表1:人 $\beta$ 珠蛋白特异性锌指蛋白识别螺旋设计

SBS #	设计					
	F1	F2	F3	F4	F5	F6
<b>33488</b>	LRHHLTR (SEQ ID NO:12)	LRHNLRA (S SEQ ID NO:13)	DQSNLRA (S EQ ID NO:14)	RNASRTR (S SEQ ID NO:15)	RSDNLSE (S SEQ ID NO:16)	RSQHRKT (SEQ ID NO:17)
<b>33501</b>	TSGSLSR (S SEQ ID NO:18)	DRSDLSR (S SEQ ID NO:19)	DRSALAR (S EQ ID NO:20)	QSSNLAR (S SEQ ID NO:21)	QSGHLSR (S SEQ ID NO:22)	NA
<b>47773</b>	LRHHLTR (SEQ ID NO:12)	LKQNLLDA (SEQ ID NO:55)	DQSNLRA (S EQ ID NO:14)	RNCARLR (S SEQ ID NO:56)	RSDNLSE (S SEQ ID NO:16)	RNQTRLN (SEQ ID NO:57)
<b>47817</b>	HHHSLKR (SEQ ID NO:58)	DRSDLSR (S SEQ ID NO:19)	TNSELDR (SEQ ID NO:59)	QSSNLAR (S SEQ ID NO:21)	QSGHLSR (S SEQ ID NO:22)	NA

表2:人β珠蛋白特异性锌指的靶位点

SBS #	靶位点
<b>33488, 47773</b>	ggAGTCAGGTGCACCATGGTgtctgttt (SEQ ID NO:23)
<b>33501, 47817</b>	gtGGAGAAGTctGCCGTTactgccctgt (SEQ ID NO:24)

[0169] 获得人类CD34+细胞，并如下进行处理。所有脐血(CB)标本根据加州大学批准的指南获得，并且被认为是免除IRB审查的匿名医疗废物。在收集的48小时内处理细胞。来自具有SCD的志愿者供体的骨髓(BM)抽吸物在其对UCLAIRB协议#10-001399知情同意的情况下获得。使用Ficoll Hypaque(Stem Cell Technologies)密度离心从BM和CB分离单核细胞(MNC)。然后通过将MNC与抗CD34微珠(Miltenyi Biotec Inc.)在4°C下温育30分钟，使用免疫磁性柱分离来富集CD34+细胞。接着使细胞通过磁性柱，收集CD34+细胞，并置于具有冷冻培养基(10%二甲亚砜(Sigma Aldrich), 90%FBS)的冷冻管中，并冷冻保存在液氮中。动员后CD34+细胞(mPB)购自Allcells。

[0170] 为了产生用于电穿孔的mRNA，将编码表1所示ZFN的质粒用SpeI(New England Biolabs)线性化，并在用作用于体外转录的模板之前通过苯酚:氯仿纯化。根据制造商的方案使用mMessage mMachine® T7 ULTRA转录试剂盒(Life Technologies)来产生体外转录的mRNA，并用RNeasy® MinElute® 清洁试剂盒(Qiagen)净化。为了电穿孔，遵循以下方案：将CB CD34+细胞在37°C下解冻，在补充有20%胎牛血清(Gemini Bio-products)和(1×谷氨酰胺、青霉素和链霉素)的Iscove改良Dulbecco培养基(Iscove's Modified Dulbecco's Medium:IMDM;Life Technologies))中洗涤，并在含有谷氨酰胺、青霉素、链霉素、50ng/ml的SCF、50ng/ml的Flt-3和50ng/ml的TP0(Peprotech)的X-VIVO15培养基(Lonza)中预刺激48小时。为了电穿孔，将每个反应的200,000个细胞以90g旋转15分钟，重悬于100μl的BTXpress缓冲液(Harvard Apparatus)中，与指定量的ZFN mRNA和/或寡核苷酸(如果适用)混合，并在BTX ECM830方波电穿孔仪(HarvardApparatus)中在250V下5毫秒脉冲一次。

[0171] 电穿孔后，细胞在室温下静置10分钟，然后加入培养基并转移到总共500μl的平板中。供体IDLV以针对适当样品所述的浓度存在于最终培养基中。

[0172] 基因校正和基因破坏分析如下:使用 Surveyor® Nuclease Assay (Cel-1) 来测定 ZFN 诱导的位点特异性等位基因破坏。使用 Cel1Fwd (5' - gacaggtacggctgtcatca-3' SEQ ID NO:25) 和 Cel1Rev (5' - cagcctaagggtggaaaat-3' SEQ ID NO:26), 利用 Accuprime Taq Hi-Fi (Life Technologies), 从 200ng 基因组 DNA 来 PCR 扩增围绕 ZFN 结合位点的 410bp 区域。完成变性、再退火、酶切、电泳和光密度测定分析, 如所描述的那样 (例如 Joglekat et al (2013) Mol. Ther: J of the Amer Soc Gene Ther 21:1705-17)。通过限制性片段长度多态性 (Restriction Fragment Length Polymorphism: RFLP) 检测位点特异性基因修饰。

[0173] 使用引物 BgloOuterFwd (5' - atgcttagaaccgaggttagagttt-3', SEQ ID NO:27) 和 BgloOuterRev (5' - cctgagacttccacactgatg-3', SEQ ID NO:28) 和 Accuprime Taq Hi-Fi (Life Technologies) 对围绕 ZFN 结合位点的 1.1kb 区域进行 PCR 扩增。PCR 产物使用 PCR 清除试剂盒 (Life Technologies) 纯化, 并使用 10 单位 HhaI (New England Biolabs) 在 37°C 下酶切 3.5 小时。酶切产物在用 GelGreen (Biotium) 预染色的 1.0% 的 TBE-琼脂糖凝胶上分离, 并在 Typhoon FLA 9000 生物分子成像仪 (GE Healthcare) 上成像。

[0174] 为了定量基因修饰, 使用基于定量 PCR 的测定。使用上述的 1.1kb PCR 产物作为模板进行一组两个 PCR 反应。将未纯化的 PCR 模板以 1:5,000 稀释, 以下每个 25μl 反应中使用其中的 1μl。使用引物 HhaIFwd (5' - gaagtctgcgttactgcg-3', SEQ ID NO:29) 和 HhaIRev (5' - cccagttctattggtctcc-3', SEQ ID NO:30) 进行第一次 PCR 以扩增经修饰的基因组。使用引物 ExonIIFwd (5' - ctcgggcctttagtgatgg-3', SEQ ID NO:31) 和 ExonIIRev (5' - gactcaccctgaagttctc-3', SEQ ID NO:32) 进行第二次 PCR 以使输入模板标准化。这些 PCR 都使用 Power SYBR Green PCR Master Mix (Life Technologies) 进行定量, 并在 ViiA7 (Life Technologies) 上获得。使用两个反应之间的 Ct (循环至阈值) 差异和质粒标准曲线确定基因修饰的频度。发现所有 ZFN 都是有活性的。

[0175] 在 Illumina MiSeq 机器上 PCR 扩增和深度测序珠蛋白旁系同源物 (paralogs) 以分析脱靶标修饰。用于 PCR 的引物如下: HBB: 5' - acacgacgcttccgatctnnnngggctggcataaaagtcag-3' (SEQ ID NO:33) 和 5' - gacgtgtgcttccgatcttccacatgccagtttctatt-3' (SEQ ID NO:34); HBD: 5' - acacgacgcttccgatctnnntaaaaggcagggcagagtcga-3' (SEQ ID NO:35) 和 5' - gacgtgtgcttccgatctacatgccagttccattgc-3' (SEQ ID NO:36); HBE1: 5' - acacgacgcttcgcatctnnnnctgttccgacacagctgcaa-3' (SEQ ID NO:37) 和 5' - gacgtgtgcttccgatcttcacccttccatccatgc-3' (SEQ ID NO:38); HBG1 和 HBG2: 5' - acacgacgcttccgatctnnnnggaacgtctgaggttatcaat-3' (SEQ ID NO:39) 和 5' - gacgtgtgcttccgatcttccctcccttgc-3' (SEQ ID NO:40)。使用扩增子内的基因座特异性 SNP 将 HBG1 和 HBG2 共扩增, 并将序列读数分配给 HBG1 或 HBG2。正向引物内的混合碱基允许测序期间的簇群去旋 (cluster deconvolution)。

[0176] 根据细胞类型, ZFN 诱导在 β-珠蛋白基因座处的 35-65% 的等位基因破坏 (indel), 而在 β 珠蛋白旁系同源物 (paralogs) 处的检测到的修饰较低, 如下表 3 所示。

表 3: 人 β 珠蛋白及其在 K562 或 CD34+ 细胞中的旁系同源物的 ZFN 相关修饰 (33488/33501)

	通过 Illumina 测序检测的 NHEJ (%)	
基因	K562 基因	人类 CD34+ 细胞
HBB	66.600	37.500
HBD	34.000	28.600
HBE1	0.000	0.000
HBG1	0.006	0.000
HBG2	0.013	0.005
HBBP1	0.024	0.005

[0177] 将编码 ZFN 的体外转录的 mRNA 电穿孔到从人脐带血 (CB) 以及动员后外周血 (mPB) 分离的 CD34+ 中，导致靶基因座的有效裂解，如通过 Surveyor Nuclease 测定法所确定的 (图 1B)。

[0178] 结果显示 ZFN 驱动在  $\beta$ -珠蛋白基因座处的高水平的基因修饰。

#### 实施例 4：在镰状细胞碱基处的 ZFN 驱动的校正

[0179] 在靶向的  $\beta$ -珠蛋白基因座成功裂解后，我们试图确定使用同源供体模板是否可能校正该位点的镰状碱基。为此，同时设计并测试了两种类型的基因校正模板：短 DNA 寡核苷酸和 IDLV。将克隆到 IDLV 中的 1.1kb 人  $\beta$ -珠蛋白基因片段供体模板设计为包括镰状突变的校正变化以及沉默限制性片段长度多态性 (RFLP)，以产生用于同源性重组的替代分析的 HhaI 限制性位点 (图 1C)。供体模板通过 IDLV 递送 (参见图 2D)，使得能以最小的细胞毒性有效转导 CD34+HSPC，从而以最小的基因组整合瞬时产生高模板拷贝数。

[0180] 用 ZFN (对 33488/33501) 加上 IDLV 供体处理的 CD34+ 细胞中的基因修饰水平最初通过 HhaI RFLP 酶切 (图 1D) 和基于定量 PCR (qPCR) 的测定确定在平均  $18.0 \pm 2.2\%$  的等位基因 (图 1E)。ZFN mRNA 和 IDLV 供体浓度的优化被执行 (图 2 和 图 3)，并且证明了在目标细胞类型中滴定 ZFN 试剂以实现高水平修饰同时保持细胞数量和生存力的重要性。

[0181] 使用寡核苷酸作为基因修饰/校正模板将具有速度、成本、可重复性和实验简易方面的优点。使用来自 mPB 的 HSPC 中的寡核苷酸供体模板的引入了被设计用于在  $\beta$ -珠蛋白基因 (HBB) 中产生 AvrII RFLP 的 3 碱基对序列的初始实验产生 15% 的基因修饰 (参见图 4)。

[0182] 为了改进寡核苷酸供体的用途，测试了一组寡核苷酸，其对应于以 ZFN 裂解位点为中心的一条或另一条链和对称增加的长度 (参见下表 4)，其中粗体表示相对于野生型序列突变的碱基。用 's' 标记的寡核苷酸包含设计为将镰状突变引入野生型基因的序列。如表中所示，在标记为 "S" 的列中，野生型序列在该位置包含 A，而镰状突变包含 T。

表 4：寡聚核苷酸供体

SMS 位置		S	0		1		2		3		4		5	SEQ ID NO					
WT	G	A	G	G	A	G	A	A	G	T	C	T	G	C	C	G	T	T	60
SMS1	G	A	G	G	A	<b>A</b>	A	A	G	T	C	T	G	C	C	G	T	T	61
SMS2	G	A	G	G	A	G	A	A	<b>A</b>	T	C	T	G	C	C	G	T	T	62
SMS4	G	A	G	G	A	G	A	A	G	T	C	T	G	C	<b>T</b>	G	T	T	63
SMS12	G	A	G	G	A	<b>A</b>	A	A	<b>A</b>	T	C	T	G	C	C	G	T	T	64
SMS14	G	A	G	G	A	<b>A</b>	A	A	G	T	C	T	G	C	<b>T</b>	G	T	T	65
SMS24	G	A	G	G	A	G	A	A	<b>A</b>	T	C	T	G	C	<b>T</b>	G	T	T	66
SMS124	G	A	G	G	A	<b>A</b>	A	A	<b>A</b>	T	C	T	G	C	<b>T</b>	G	T	T	67
SMS012	G	A	<b>A</b>	G	A	<b>A</b>	A	A	<b>A</b>	T	C	T	G	C	C	G	T	T	68
SMS12s	G	T	G	G	A	<b>A</b>	A	A	<b>A</b>	T	C	T	G	C	C	G	T	T	69
SMS012s	G	T	<b>A</b>	G	A	<b>A</b>	A	A	<b>A</b>	T	C	T	G	C	C	G	T	T	70

[0183] 使用更长的反向链寡核苷酸得到最高水平的基因修饰(参见图 5)。由于在SCD突变位点处的基因修饰或校正不会改变ZFN结合位点的序列或间隔,因此沉默突变被设计到供体模板中并共同导入染色体中以防止或减少ZFN结合和经修饰的等位基因的再裂解。使用反向链供体与在3'ZFN结合位点引入沉默突变是相容的(图4C)。

[0184] 将沉默突变的各种组合引入供体寡核苷酸,并测定基因修饰的频度。沉默突变位点(SMS)一和二的组合(SMS12)产生最高水平的HBB修饰,并用于所有子序列后续实验(图4D)。经基因修饰的HBB等位基因和经修饰然后重新裂解的等位基因之间的反向关系表明SMS供体确实可用于阻断 ZFN再裂解(图4E)。检查在HSPC库和爆发形成单位中的HBB转录物,红细胞菌落(BFU-E)没有发现改变来自任何SMS等位基因的mRNA剪接的证据。

[0185] 在高ZFN mRNA(对47773/47817)浓度下,使用含有SMS012的供体寡核苷酸比使用SMS12供体产生更多的经HBB-修饰的等位基因(图 17B)。

[0186] 在用ZFN和SMS124供体处理后HBB的分子结果的综合分析仅显示了较小水平的意外的DNA修复事件,例如来自同源性定向和基于NHEJ的基因修饰的组合(1.5%)或来自寡核苷酸的基于NHEJ的捕获(0.4%) 的那些(参见下表5)。

表5:用ZFN和寡核苷酸供体处理HSPC的分子结果

结果	频度 (%)
没有变化	46.1
NHEJ-介导的缺失	30.5
寡核苷酸模板化基因修饰	18.8
NHEJ-介导的插入	2.8
半基因修饰,半NHEJ	1.5
供体寡核苷酸的捕获	0.4

[0187] 最后,部分同源性指导的基因修饰事件(SMS12对SMS24)的过度丰富提供了对以下构思的强有力的支持:在寡核苷酸模板化的基因修饰过程中的新的DNA合成使用切除的DSB的左手3'单链末端进行(参见图6)。ZFN、寡核苷酸和HSPC供体的最佳组合允许30-40%的等位基因的基因修饰(参见图4F)。

[0188] 因此,当与同源供体模板一起递送时,ZFN能够诱导初级 HSPC中的高水平的靶向DSB以及高水平的位点特异性基因修饰。

#### 实施例5:ZFN特异性

[0189] 由于β-珠蛋白与其他珠蛋白基因( $\delta$ -, $\epsilon$ -, $\alpha\gamma$ -, $\gamma\gamma$ -和伪 $\beta$ -珠蛋白)具有高同源

性,因此设计ZFN以避免在这些区域中的裂解(图6A)。使用Surveyor核酸酶测定法并通过高通量DNA测序,测定使用33488/33501对的同源珠蛋白基因的裂解速率。对这些区域中的每一个的分析显示仅在mPB、CB和SCD BM CD34+细胞中的高度同源的 $\beta$ -珠蛋白基因处的脱靶修饰(图6B,表3)。我们补充这些脱靶ZFN裂解的直接测试与更全面、无倾向性的全基因组方法来识别脱靶位点。该测定基于非同源IDLV在双链断裂位点被捕获的倾向,并且使用针对其对IDLV捕获的允许而选择的K562细胞中的整合的下游聚类整合位点分析(CLIS)。所遵循的程序概述如下:

[0190] 用以下引物PCR扩增 $\beta$ -珠蛋白基因簇位点:HBD:5' - gttcattttcattctcaca-3'(SEQ ID NO:41)和5' - gtaatctgaggtaggaaaac-3'(SEQ ID NO:42);HBBP1:5' - cacccttgaccaatagattc-3'(SEQ ID NO:52)和5' - gagactgtggatagtcata-3'(SEQ ID NO:43);HBE1:5' - cattatcacaaacttagtgtcc-3'(SEQ ID NO:44)和5' - agtctatgaaatgacaccatat-3'(SEQ ID NO:45);HBG1:5' - gcaaaggctataaaaaaaattagc-3'(SEQ ID NO:46)和5' - gagatcatccaggtgcttg-3'(SEQ ID NO:47);HBG2:5' - ggcaaaggctataaaaaaaattaagca-3'(SEQ ID NO:48)和5' - gagatcatccaggtgcctta-3'(SEQ ID NO:49),并通过如上所述的核酸酶进行分析。

[0191] K562细胞用15 $\mu$ g/mL的ZFN mRNA电穿孔,并用2E+08 TU/mL(MOI=250)的含有由MND启动子驱动的GFP转基因的IDLV转导(参见Challita,P.M.et al.(1995) J of Virol 69, 748-55),且与 $\beta$ -珠蛋白基因座缺乏任何同源性以确保双链断裂处的整合是由于NHEJ介导的捕获而实现的(参见图7A)。完成了ZFN+IDLV的四个单独的生物复制以维持对潜在的全基因组靶外位点的高敏感性,同时完成两个对照以确定背景整合的水平。对照细胞接受GFP IDLV,但没有ZFN mRNA以检测捕获IDLV而无ZFN作用的天然存在的DSB的位点。将K562细胞在具有10%胎牛血清(Gemini Bio-products) 和1%青霉素/链霉素/L-谷氨酰胺(Gemini Bio-products) 的RPMI1640(Cellegro) 中培养60天。在此期间,定期进行GFP阳性的流式细胞术分析以确保任何非整合的IDLV被稀释出样品。在第60天,通过荧光激活细胞分选(FACS) 分选样品以分离GFP+细胞,扩增5天,提取基因组DNA,并通过非限制性线性扩增介导的PCR35(nonrestrictive Linear Amplification Mediated-PCR35) 制备用于载体整合位点测序的样品。如先前概述的(Gabriel,R.,et al(2011) Nat Biotech.29 (9):816-23)用500个碱基对的CLIS窗口进行聚集整合位点(CLIS) 分析以揭示ZFN裂解的位点。分析所有CLIS以确定每个位点与ZFN对序列的同源性水平。同源性百分比定义为在间隔物(spacer) 长度高达20个核苷酸的 ZFN结合(同源二聚体和异源二聚体) 的所有重复的位点处的匹配碱基对的数目。每个CLIS的最高同源性百分比记录在下表6中。对于该表,列出所有确定的CLIS,其具有在其整合位点的200个碱基对内的与ZFN靶位点的最高同源性百分比。评价了ZFN结合(异源二聚体和同源二聚体) 的所有可能的重复,并分析了高达20nt的间隔物长度。裂解的两个ZFN位点以粗体突出显示。CLIS确定的所有其他脆弱位点都是斜体。Chr:染色体;CLIS:聚类整合位点;L:“左”ZFN单体(SBS#33488);R:‘右’ZFN单体(SBS# 33501);而ZFN二聚体中的数目对应于给出与ZFN靶位点有最大百分比的同源性的间隔物大小。使用此方法分析20,000个随机基因组位点以确定跨基因组的ZFN同源性的背景水平。

表6:CLIS和与ZFN靶位点的同源性的百分比

Chr	位置	最近的 RefSeq 基因 (在 500 kb 内)	同一性 (%)	CLIS 的编 号 (#)	ZFN 二聚体	样本
11	5248238	HBB	100%	18	L_5_R	ZFN+GFP
11	5255649	HBD	91%	10	L_5_R	ZFN+GFP
7	12238685	TMEM106b	52.94%	3	L_12_R	ZFN+GFP
19	58024409	ZNF773	62.50%	6	L_18_L	ZFN+GFP; 仅 GFP
1	121484829	EMBP1	55.88%	6	R_14_L	ZFN+GFP; 仅 GFP
10	117392565	ATRNL1	55.88%	3	L_15_R	ZFN+GFP; 仅 GFP
19	27733086	-	56.25%	3	L_17_L	仅 GFP
12	112392669	TMEM116	52.94%	2	L_4_R	仅 GFP
12	693139	NINJ2	35.29%	2	L_1_R	仅 GFP
16	46428309	ANKRD26P1	53.12%	2	L_3_L	仅 GFP
5	144123495	KCTD16	53.12%	2	L_4_L	仅 GFP
3	183525263	YEATS2	53.13%	2	L_15_L	仅 GFP
5	53018665	NDUFS4	58.82%	2	R_9_L	仅 GFP
3	19674417	EFHB	56.25%	2	L_9_L	仅 GFP
9	127395708	NR6A1	55.88%	2	L_6_R	仅 GFP
<hr/>						
7	137218714	DGKI	52.94%	2	L_12_R	仅 GFP

[0192] 如所预测的, CLIS分析显示在β-珠蛋白和δ-珠蛋白的捕获(图 7,表6)。在CLIS分析中没有发现与ZFN结合位点具有显著同源性的其他推定的脱靶位点。

[0193] 这些结果表明该对ZFN在其全基因组规模上对其预期靶位点的高水平的特异性。

#### 实施例6:经修饰的HSPC的体外分化

[0194] 为了确保经基因修饰的HSPC能够有正常的、广谱的红细胞和髓样分化,将经处理的细胞作为单细胞且在大量培养物中进行测定。监测单细胞在含有促进HSPC分化的细胞因子的甲基纤维素培养基中的菌落形成潜能。简言之,在电穿孔后一天,在基于MethoCult™ Optimum甲基纤维素的培养基(Stem Cell Technologies)中,将100个活细胞和300个活细胞在每个 35mm细胞培养皿中按一式两份铺板。对于使用寡核苷酸供体的实验,在电穿孔后

5天将细胞铺板接种。在5%CO<sub>2</sub>、37℃和潮湿气氛下培养两周后,基于菌落的形态表征和计数不同类型的菌落。单独用ZFN(对33488/33501)处理的HSPC或用ZFN和寡核苷酸供体处理的HSPC产生类似的数量和模式的红细胞和骨髓克隆。

[0195] 用于SCDBM样品的体外红细胞分化技术改编自Giarratana et al. ((2005) Nat Biotech 23:69-74) 并描述于Romero和Urbinati等((2013) J Clin Invest 123 (8) :3317-3330) 中。在电穿孔后一天,将细胞置于红细胞培养物中。为了获得大量的细胞以用于下游分析,在方案的第12天而非第8天将细胞置于共培养物中,以允许增加的红细胞扩增。细胞在实验的第22天收获,沉淀,并进行HPLC分析(见下文)。还使用5'-acatttgcttctgacacaac-3'(外显子1, SEQ ID N0:50) 和5'-gaaattggacagcaagaaagc-3'(外显子3, SEQ ID N0:51) 通过 RT-PCR进行HBB剪接的测定,没有观察到外显子跳跃,也没有观察到内含子保留。

[0196] 单个红细胞菌落的基因分型证实在预期频度存在预期的编辑。此外,在大量培养物中,HSPC被诱导分化成红细胞(Giarratana,M. et al. (2011) Blood 118, 5071-9)。在整个红细胞分化中对细胞进行取样,并通过高通量DNA测序来测量编辑的细胞的频度。与我们对在甲基纤维素中生长的红细胞菌落的单细胞分析一致,不管在基因库(pool)中的基因修饰的起始水平如何,在大量培养物中分化为红细胞期间,发现经修饰的细胞的频度基本上没有变化(图9A)。

[0197] 为了分析珠蛋白链,执行HPLC(图9B)。对于SCD BM样品,对通过终末期红细胞产生的血红蛋白(Hb)物质执行HPLC,如Wilber,A. et al. ((2011) Blood 117:2817-26) 所描述的,但略有修改。在红系细胞培养结束(第 22天)时收获红系细胞(1-2E+07),沉淀,并在-80℃冷冻储存直至裂解。解冻后,将细胞重悬浮于25μL溶血试剂(Helena Laboratories)中,冷冻过夜,然后在4℃以20,800×g离心10分钟以除去细胞碎片。将澄清的上清液用于通过HPLC使用阳离子交换柱(Ultra2 Variant Resolution Analyzer;Primus Diagnostics)和人血红蛋白的校准样品来表征Hb的产生。基于保留时间,利用伴随HPLC仪器的软件来识别对应于Hb物质的峰。基于包括乙酰化胎儿血红蛋白,HbFAC;胎儿血红蛋白,HbF;野生型血红蛋白,HbA;和镰状血红蛋白HbS的每个初级血红蛋白峰的曲线下面积的总和计算针对每个样品产生的HbA的相对百分比。

[0198] 对于mPB样品,在红细胞培养的第18天收获红系细胞。将细胞沉淀并在室温下在水中裂解10分钟。在20,000g离心5分钟以除去细胞碎片后,将细胞裂解物在-80℃冷冻储存。解冻后,细胞裂解物在流动相A中以 1:10稀释,并使用弱阳离子交换柱(PolyCAT ATM, PolyLCINC.) 通过 HPLC(Infinity1260,Agilent) 表征。FASC参考物质(Trinity Biotech) 用于限定常见血红蛋白(HbF、HbA、HbS、HbC) 的洗脱时间。使用OpenLAB CDS Chemstation软件进行分析和峰积分。基于包括乙酰化胎儿血红蛋白, HbFAC;胎儿血红蛋白,HbF;野生型血红蛋白,HbA;和镰状血红蛋白 HbS的每个血红蛋白峰的曲线下面积的总和计算针对每个样品产生的HbA的相对百分比。

[0199] 在平行实验中,我们将14%的HBB等位基因转换为镰状形式,并通过HPLC测定红系分化18天后存在的珠蛋白链。这些细胞产生的约18%的血红蛋白是镰状血红蛋白,证明了基因修饰事件在蛋白水平的影响(图 9B)。

#### 实施例7:经修饰的细胞植入小鼠

[0200] 为了确定经ZFN修饰的细胞是否维持其造血再增殖能力,将经 ZFN和供体处理或

模拟处理的CB来源的HSPC异种移植到免疫缺陷型NSG 小鼠中。

[0201] 简而言之,在用ZFNmRNA电穿孔并用供体 IDLV (2E+07TU/mL; MOI=50) 转导之前,预刺激来自多个健康个体的新鲜 CBCD34+细胞2天。使用包含含有谷氨酰胺、青霉素、链霉素, 50ng/mlSCF、50ng/mlFlt-3和50ng/ml TPO (Peprotech) 的X-VIVO15培养基 (Lonza) 的预刺激培养基。恢复一天后,通过尾静脉注射将含有0.1% BSA (Sigma) 的PBS (Corning) 中的1E+06活细胞移植到6至8周龄的用250cGy 全身照射后的NSG小鼠 (The Jackson Laboratory) 中。使用平行培养但未暴露于mRNA、电穿孔或IDLV (模拟处理) 的CD34+细胞的对照样品。将小等分试样在体外培养用于多重分析。将mPB样品在电穿孔后1天冷冻,在之后的日子解冻,并允许移植前如上所述进行一天的恢复。

[0202] 使用V450-缀合的抗人CD45与APC-缀合的抗鼠CD45在移植后第5周通过流式细胞术评价人类细胞的移植。在用抗体孵育后,用BD FACS裂解液 (BD Biosciences) 裂解红细胞。移植效果百分比定义为% huCD45+/ (% huCD45++% muCD45+)。使用V450-缀合的抗人CD45、V500- 缀合的抗-鼠CD45、FITC-缀合的抗-人CD33、PerCP-缀合的抗-人CD3、PE 缀合的抗人CD56, PE-Cy7缀合的抗人CD19和APC缀合的抗人CD34 (所有抗体BD Biosciences) 在移植后8周和16周通过流式细胞术评价外周血和BM 中的移植效果和谱系分布。

[0203] 包括结果变量的平均值和标准偏差的描述性统计数据都被报告并以图形方式呈现。对于与基因修饰水平、植入和谱系相关的定量结果,如果存在超过两个组,则通过不配对t检验或在单因素方差分析 (ANOVA) 的框架内进行成对比较。进行剂量-反应分析以评价ZFN和供体比优化。具体来说,我们使用线性混合模型方法,但处理剂量作为连续和固定效应,实验作为随机效应。通过重复测量ANOVA评价SCD患者BM细胞随时间推移的扩增倍数。在我们的统计调查中,假设检验是双侧的,使用用于p值的显著性阈值0.05。所有统计分析使用SAS版本9.3 (SAS Institute Inc. 2012) 进行。

[0204] 体外平行培养细胞表明,通过qPCR RFLP分析,基因修饰率在 5%-20%范围内,平均值为 $14.5 \pm 6.4\%$  (图10)。高通量测序显示 $10.5 \pm 4.0\%$ 的等位基因在镰状位置含有交换的碱基,在 $32.0 \pm 9.9\%$ 中看到插入或缺失 (indel)。此外,在体外菌落形成测定中评价这些细胞的造血潜能;相对于所有可分析谱系中的未经处理的模拟样品,细胞保持了它们的广谱菌落形成能力。

[0205] 在注射后5和8周,评价经移植的小鼠,人HSPC的移植效果测量为人CD45+细胞占全部CD45+细胞(来自人和鼠)的百分比。经ZFN和 IDLV处理的HSPC的植入水平与未经处理的对照的植入水平相当,在第5周平均为 $14.8 \pm 11.4\%$ ,在第8周增加至45% (图10C)。小鼠的外周血的谱系分析 (T细胞的CD3, 骨髓细胞的CD33, HSPC的CD34, B细胞的CD19和 NK细胞的CD56) 是如同在该模型中预期的 (图11)。对在接受用ZFN和寡核苷酸供体处理过的细胞的小鼠中的人类细胞的分析揭示了与经ZFN和IDLV 处理的HSPC的HSPC分化光谱类似的HSPC分化光谱 (图10E)。

[0206] 在移植后16周对小鼠实施安乐死,以使得能评价人类细胞移植效果和基因修饰水平。对于接受经ZFN和IDLV处理的细胞的小鼠,对外周血的评价显示高水平的移植效果和预期的谱系分布。骨髓(BM) 区室的分析是相似的 (图12)。类似地,在接收经ZFN和寡核苷酸两者处理的细胞中,移植水平和谱系分布与在外周血和BM中接受经ZFN和IDLV处理的HSPC的小鼠组中观察到的相似。

[0207] 为了确定小鼠中存在的细胞是否确实是通过ZFN试剂修饰的那些细胞,从每只小鼠的BM和脾组织中分离基因组DNA,并且研究人 $\beta$ -珠蛋白基因。通过在接受经ZFN和IDLV处理的细胞的小鼠的组织中的qPCR分析HhaI RFLP揭示了这些小鼠中存在基因修饰,但是比针对输入细胞观察到的10.5%基因修饰水平,频度更低(对于BM和脾分别为0.14±0.32%和0.16±0.18%) (图13)。序列分析证实在BM和脾中的人类细胞中的镰状突变处的基因修饰分别具有0.21±0.39%和0.27±0.31%的类似的平均值 (图13B)。对在切割位点通过NHEJ引起的插入和缺失(indel)的分析显示相比于通过HDR引起的有更高水平的变化,来自含indel的BM的所有序列读数的4.8±7.8%和来自脾的所有序列读数的3.8±3.7% (图14)。

[0208] 对接受经ZFN和寡核苷酸处理的细胞的小鼠的BM和脾的基因组分析也表明维持基因修饰,所述细胞在施用之前在镰状碱基具有17.3%的基因修饰且在主群体中有19.8%的indel。对来自这些组织的DNA进行高通量测序,分别在BM和脾中显示0.85±0.81%和2.11±1.19%的靶向基因修饰 (图13C)。在这些组织中的indel的评价显示在BM中为3.34±2.65%的水平以及在脾中为5.86±2.30%的水平(图14)。

[0209] 这些结果表明已经经历通过ZFN和同源定向的基因修饰进行的位点特异性DNA裂解的CD34+细胞能够移植并进行多谱系分化。

#### 实施例8:镰状骨髓中的基因校正

[0210] 由于镰状细胞疾病患者不是利用G-CSF的干细胞动员的候选者,因此我们从这些患者的BM吸出物获得CD34+HSPC。使用ZFN mRNA 和 IDLV供体进行位点特异性基因校正。将细胞置于红细胞扩增培养基中,然后使用已建立的方法分化 (Romero et al, 同上, Giarratana et al, 同上)。此外,评价一部分细胞的菌落形成潜能。与模拟、非电穿孔的对照样品相比,用 ZFN+IDLV (对33488/33501) 处理的细胞显示适度较低的(35%) 菌落形成能力 (图15)。

[0211] 在初始红系扩增后(但在去核前),收获细胞用于基因组分析。通过这些样品的qPCR进行的RFLP分析显示平均为20.1±8.8%的基因修饰水平(图16A)。这些结果通过深度测序证实,以18.4±6.7%的读数显示SCD 突变的校正(图16B)。此外,测序证实,包含SCD突变的校正的大多数等位基因也在HhaI位置处包含碱基变化,从而表明大多数HDR驱动的事件涵盖这两个碱基之间的至少22bp的距离。红系培养结束后,收集样品,以通过高压液相色谱(HPLC) 分析珠蛋白四聚体(图16C)。

[0212] 来源于经ZFN和IDLV处理的样品的红细胞中HbA峰的存在表明基因校正导致 $\beta$ S等位基因功能性转化为 $\beta$ A等位基因。在来源于模拟处理的细胞的红细胞中没有观察到这样的峰。由于经ZFN处理的样品中HbS 峰的减少,因而HbF峰显示相对增加。在经ZFN和IDLV处理的样品中HbA 的相对诱导率平均为5.3±0.02%,其中蛋白校正水平高达10.0% (图16D)。

[0213] 这些结果证明了ZFN与IDLV供体组合以校正来自镰状细胞疾病的骨髓的HSPC表型的能力。

[0214] 这里显示的结果表明使用锌指核酸酶在体外人造血干/祖细胞镰状细胞疾病突变的基因校正的高水平。在本研究中,我们设计了ZFN对来裂解 $\beta$ -珠蛋白基因座。这些ZFN与同源供体模板(作为寡核苷酸或通过整合酶缺陷型慢病毒载体递送)组合,能够在祖细胞中以高水平诱导同源定向修复。

[0215] ZFN裂解位点的分析揭示大部分核酸酶活性发生在β-珠蛋白基因座靶位点处,在同源δ-珠蛋白基因处裂解的比例较小。由于在成年红系细胞中的低水平的δ-珠蛋白表达(<所有珠蛋白的3.5%),在细胞亚群中的该基因处的ZFN活性不太可能是有害的。使用IDLV末端捕获对在K562细胞中的ZFN靶外裂解位点的无倾向性的全基因组评价证实了ZFN的高特异性,仅发现背景整合到天然脆弱位点内。

[0216] 当ZFN和经供体处理的细胞移植到小鼠中时,细胞的移植效率和谱系分布在经模拟处理的和经ZFN和供体处理的样品之间是相等的。然而,尽管在移植前的体外样品中基因修饰的平均水平为10-20%,但在植入16周后,在小鼠的脾脏和BM中的人类细胞中的基因校正水平显著降低。这些发现与本领域最近公开的工作一致(Genovese et al (2014) Nature 510:235- 240),并且可能意味着虽然更成熟的祖细胞被有效地校正,但是通过修饰早期的更原始的造血干细胞提供持续的益处(例如临床益处)。

[0217] 与位点特异性基因破坏的效率(参见Holt et al (2010) Nat biotech:839-47 regarding CCR5 disruption)相比,在更原始HSPC中的同源定向的基因校正的效率可能是由于在使用含有校正碱基的供体模板中的差异,其是共递送的并且细胞DNA损伤修复途径必须使用HDR而不是NHEJ来分辨DSB。由于NHEJ倾向于静止、原始的HSC(参见Mohrin et al (2010) Cell stem cell 7:174-85),修复途径选择的倾向性可能限制原始HSC中的基因校正。因而,不受一种理论的束缚,ZFN在成熟和原始细胞群体中以相似的效率起作用是可能的,但是每种细胞类型中的活性修复途径不同,使得HDR在更成熟的细胞中更具活性,并且在原始HSC中较不具活性。特别地,移植到小鼠中的细胞与它们保持基因修饰的输入水平相比在较大程度上保持了indel的输入水平(对于IDLV和寡聚实验,相比于43.9和11.7倍的基因修饰变化,相应的indel变化为7.4和4.3倍)。

[0218] 来自镰状细胞疾病的BM CD34+细胞中的实验提供了临床翻译的前景。对于这些实验,典型镰状形突变(至少在红系祖细胞中)的校正水平平均为18%,并且在分化后,这些细胞产生校正的野生型血红蛋白(HbA)。基于来自用于SCD的同种异体造血干细胞移植的数据,由于正常的供体来源的红细胞的选择优势,因此10-30%的供体嵌合率可导致显著的临床改善(Andreani et al (2011) Chimerism2 (1) :21-22和Walters et al (2001) Am Soc Blood and Marrow Transpl 7:665-73)。此外,SCD突变的杂合子通常不经历疾病的症状。因此,校正每个HSC中的仅一个等位基因可证明足以减轻与SCD相关的大部分症状。

[0219] 尽管基于慢病毒的用于血红蛋白病,并且特别是用于SCD的基因治疗最近有进展(参见Romero et al,同上,以及Chandrakasan and Malik (2014) Hematol/oncol Clin of North Amer 28:199-216),但是由于需要治疗性转基因的长期和适当调节的表达,因而潜在的并发症仍然存在。在HSC中使用靶向核酸酶的经典A至T致病镰状颠换的位点特异性校正提供了维持β珠蛋白在其内源启动子和基因座控制区下的表达的独特能力。此外,基因组校正试剂仅需要对细胞进行一次性的、瞬时的离体处理而导致永久校正。

[0220] 总之,这些数据支持作为SCD潜在治疗的HSPC中基因组编辑的持续发展。

[0221] 本文提及的所有专利、专利申请和公开特此通过引用全文并入。

[0222] 虽然为了清楚理解的目的,通过说明和实施例相当详细地提供了公开内容,但是对于本领域技术人员显而易见的是,在不脱离本发明的精神或范围的情况下,可以实践各种改变和修改方案。因此,前述描述和实施例不应被解释为是限制性的。

5'-TCAACAGACACCATGGTGCATCGACTCCTGAGGAGAAGTCTGCCGTACTGCCCTGTGGGGC

图1A

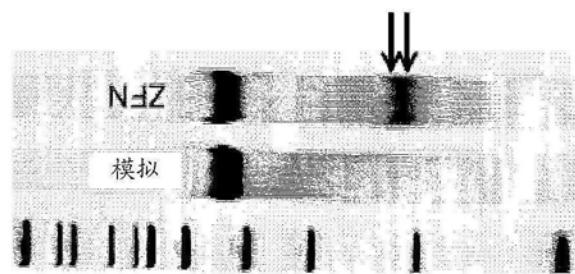


图1B

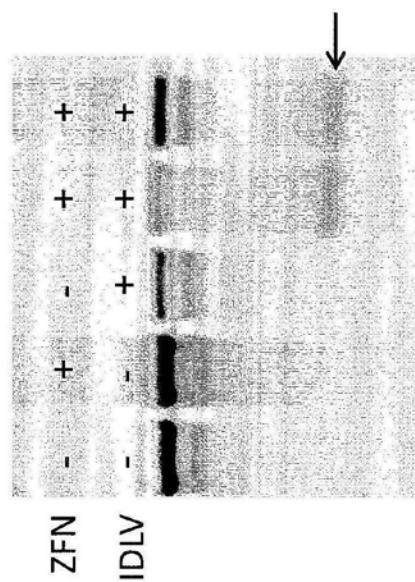


图1D

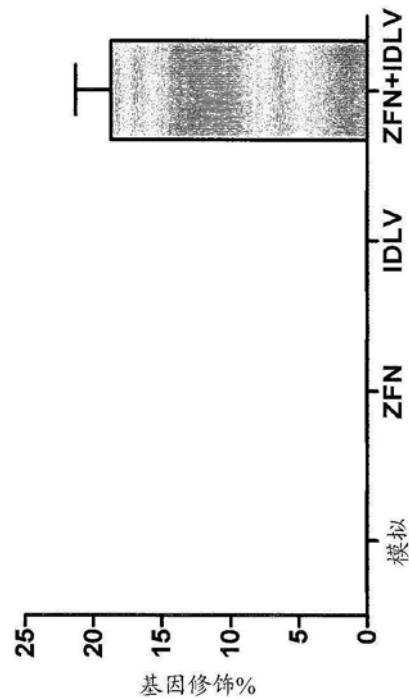


图1E

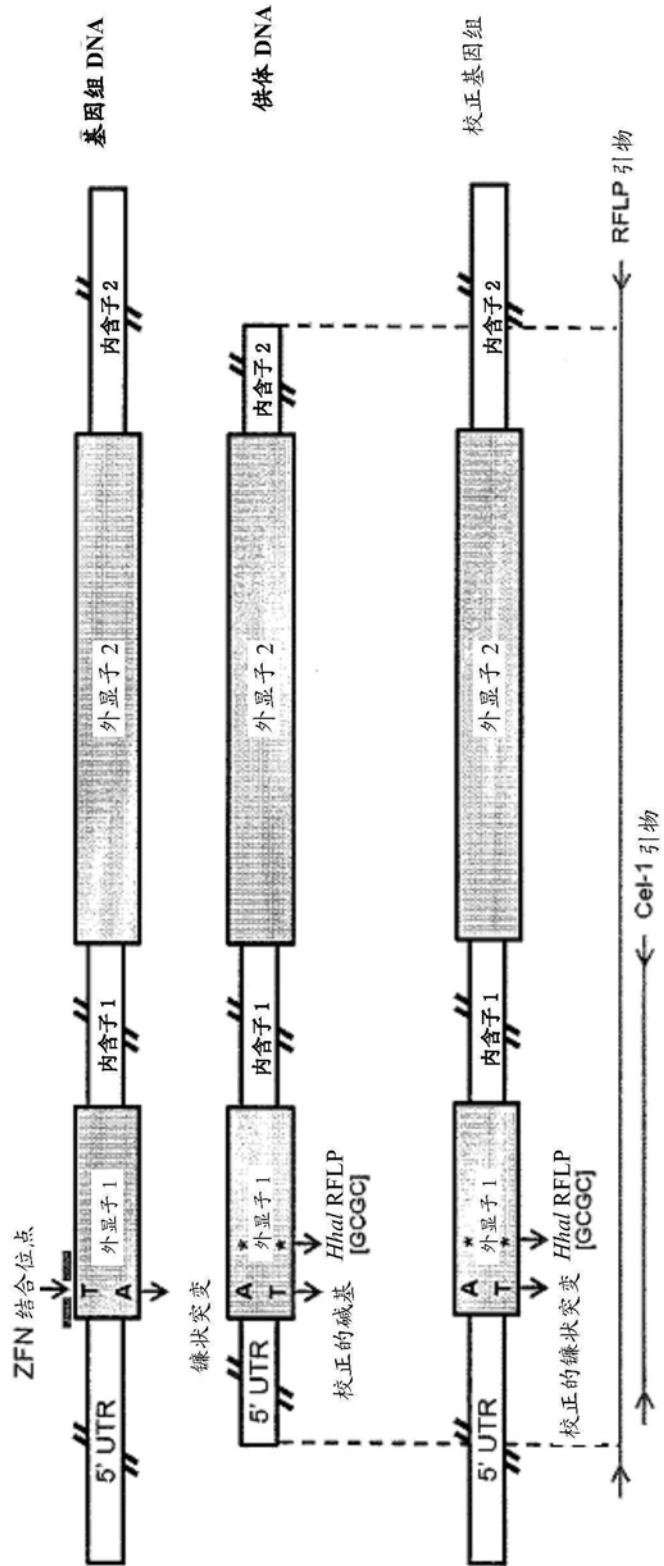


图1C

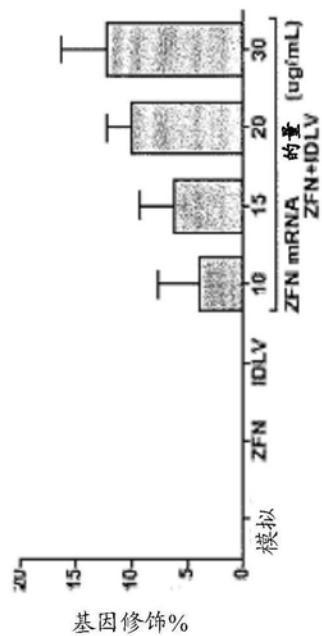


图2A

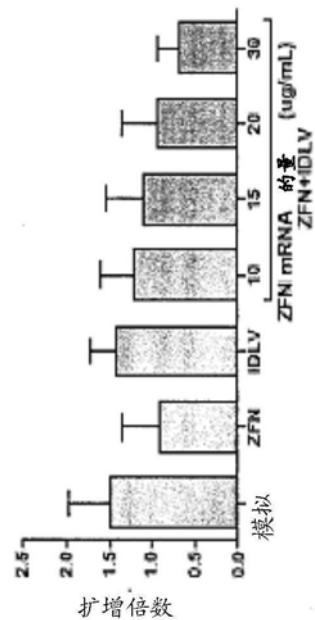


图2B

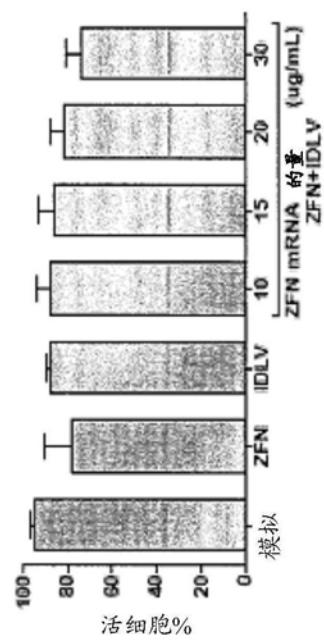


图2C

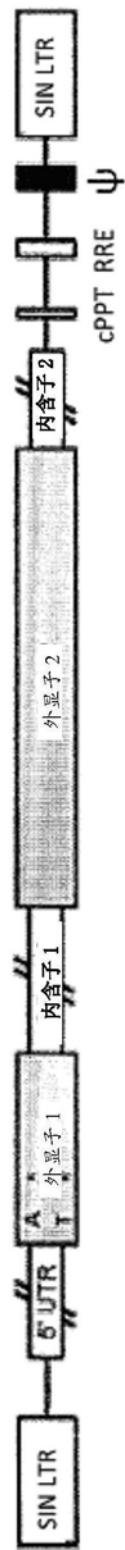


图2D

图 3A

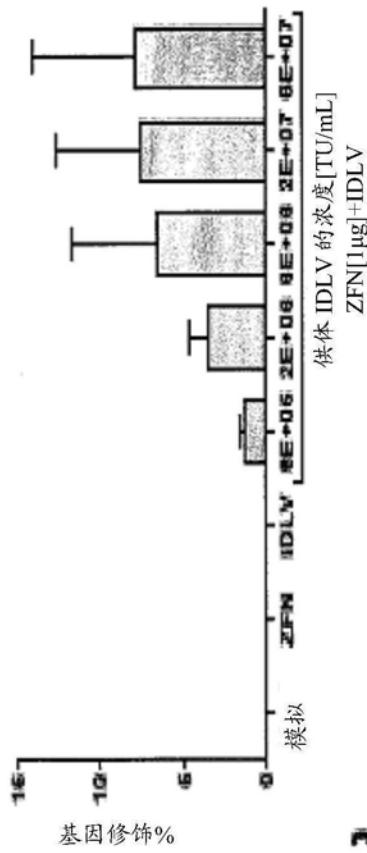


图 3B

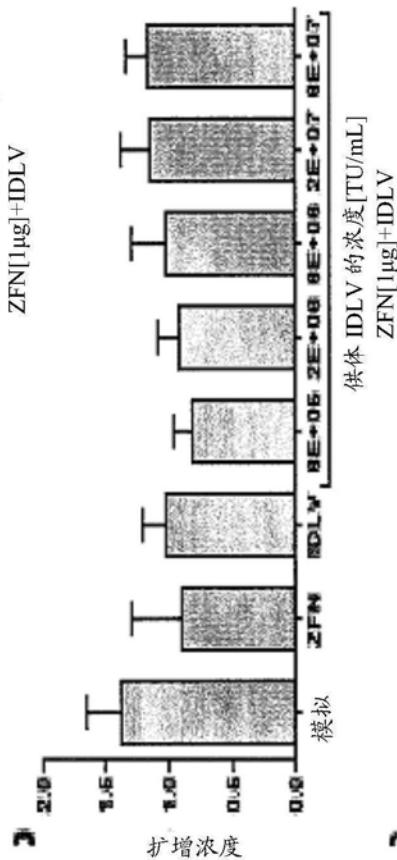


图 3C

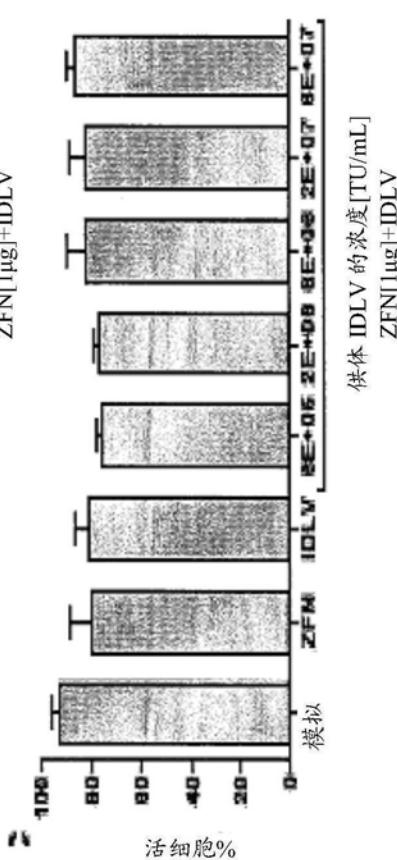


图 4A

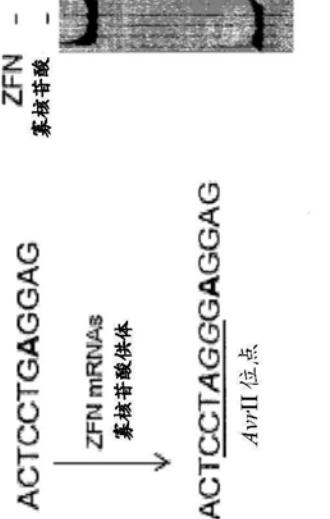


图 4B

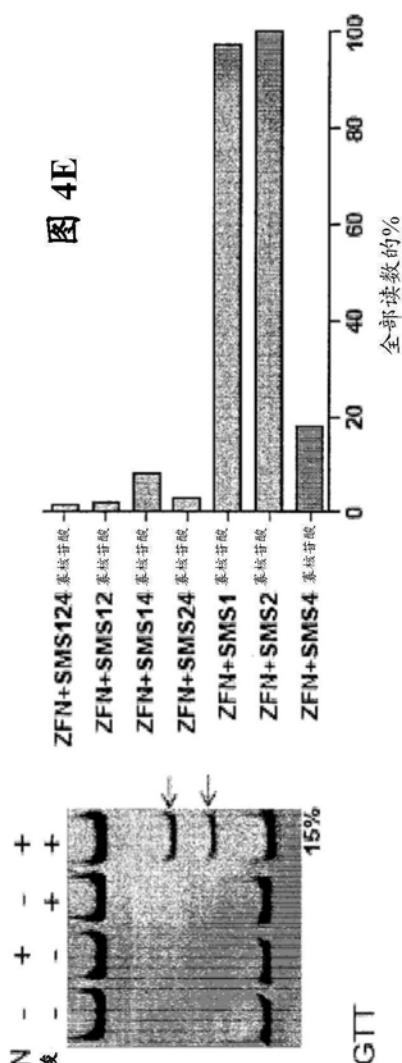


图 4E

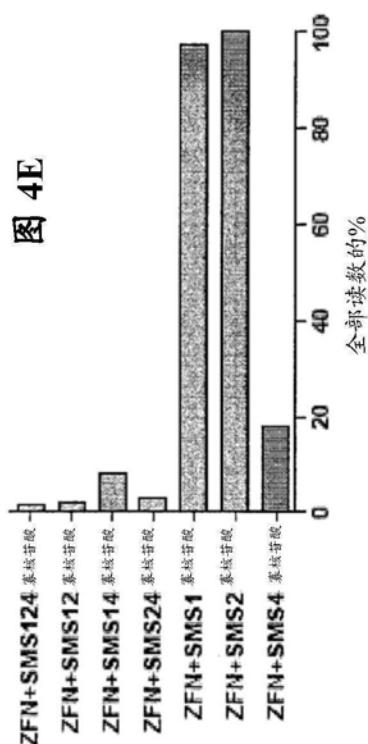


图 4F

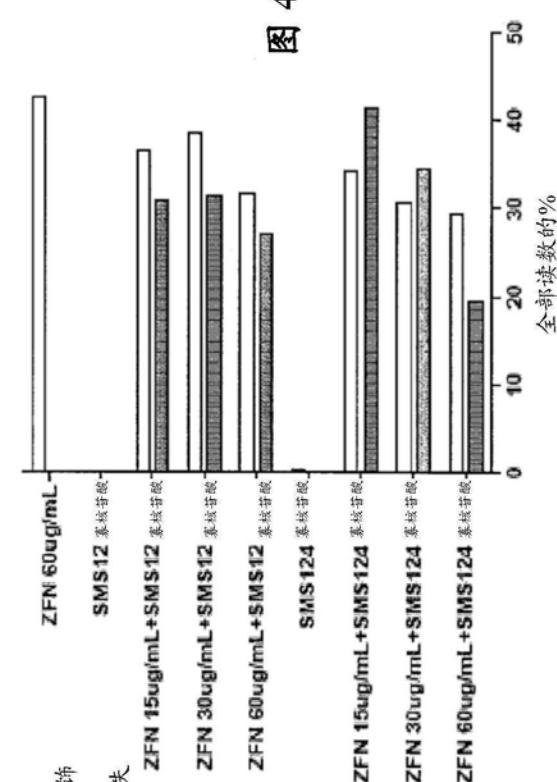
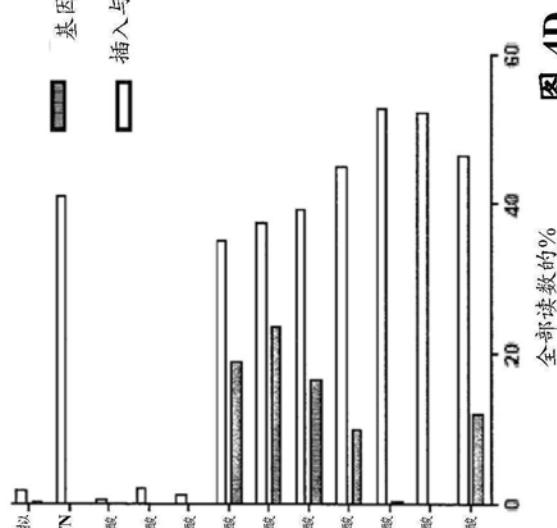


图 4D



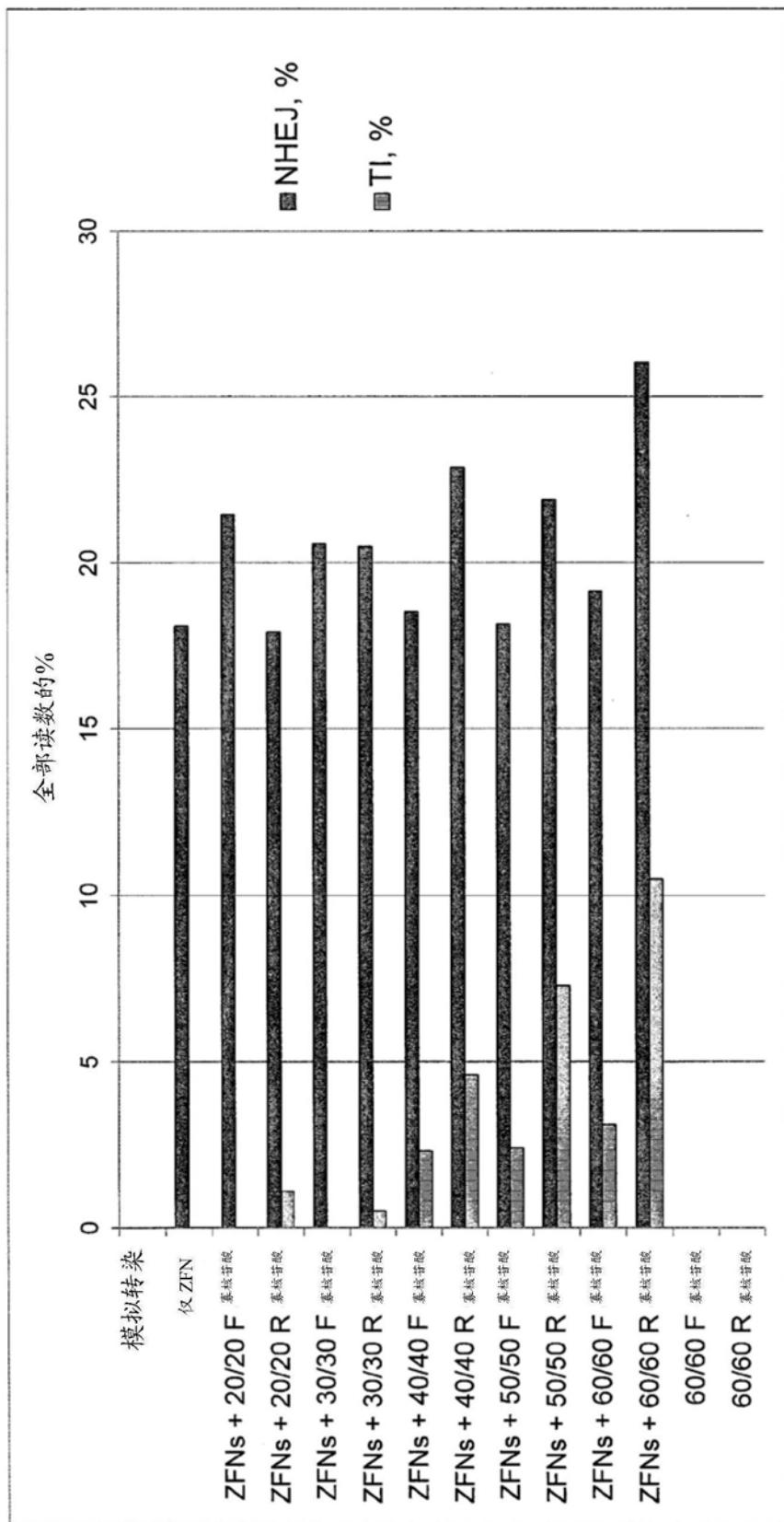


图5

	从左到右		从右到左
SMS1	SMS12	SMS124	SMS4
0.03	6.0	35.2	0.2
		0.03	
		35.2	

图6A

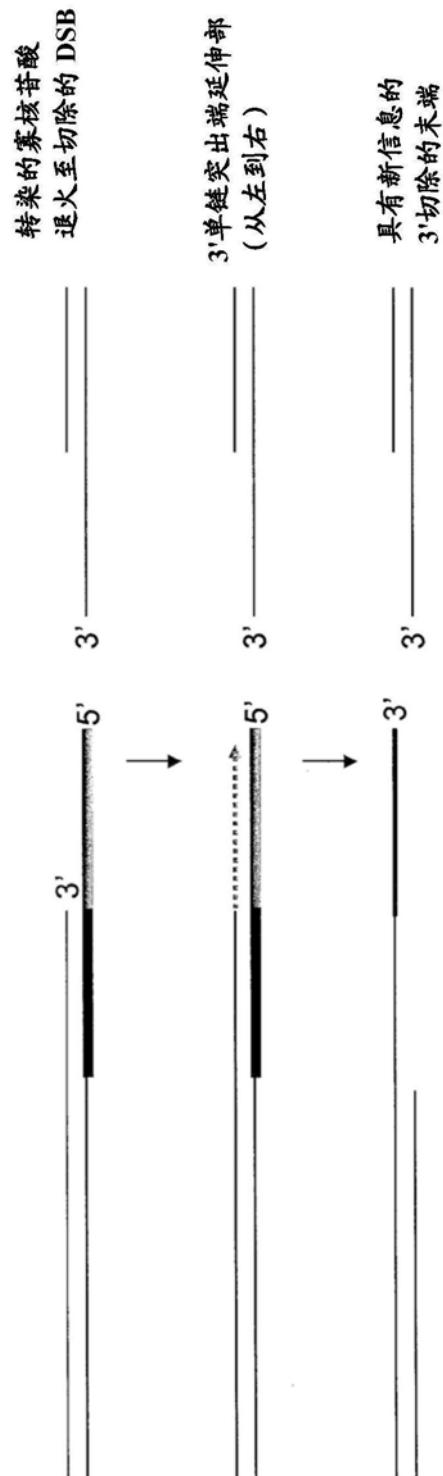


图6B

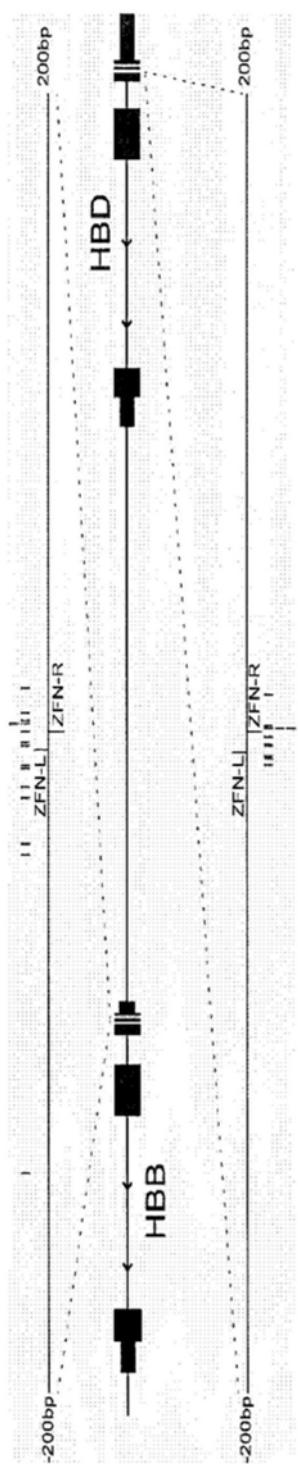


图7A

HBB	C	A	C	C	A	T	G	G	T	G	C	T	G	C	G	T	T
HBD	C	A	C	C	A	T	G	G	T	G	C	T	G	C	T	G	T
HBE1	C	A	T	C	A	T	G	G	T	G	C	T	G	C	G	T	C
HBBP1	C	A	C	T	G	T	A	G	T	G	C	T	G	C	T	G	C
HBG1	C	G	C	C	A	T	G	G	G	T	C	A	T	G	G	C	T
HBG2	C	G	C	C	A	T	G	G	G	T	C	A	T	G	G	C	T

图7B

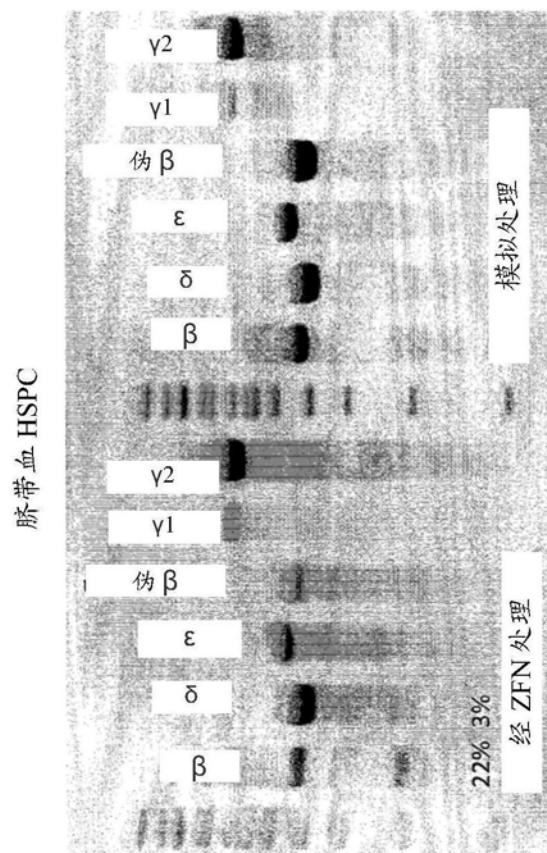


图7C

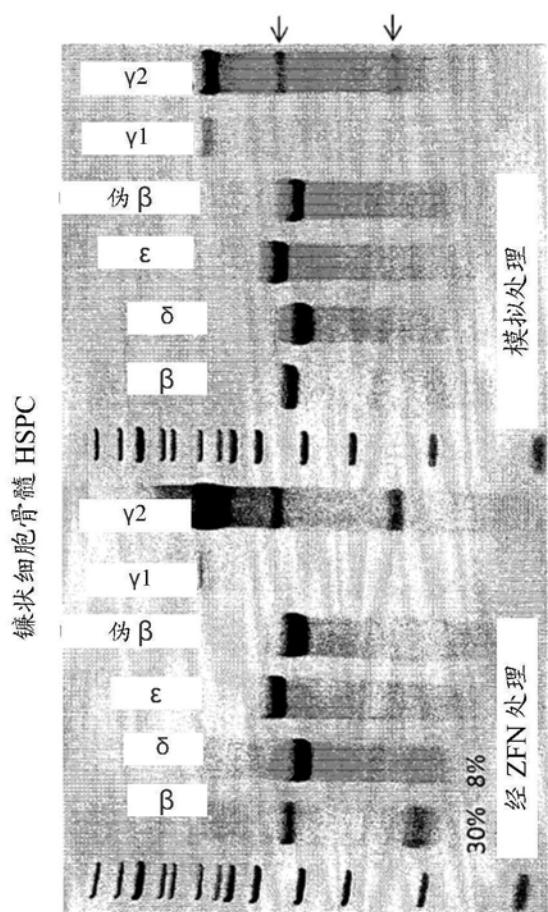


图7D

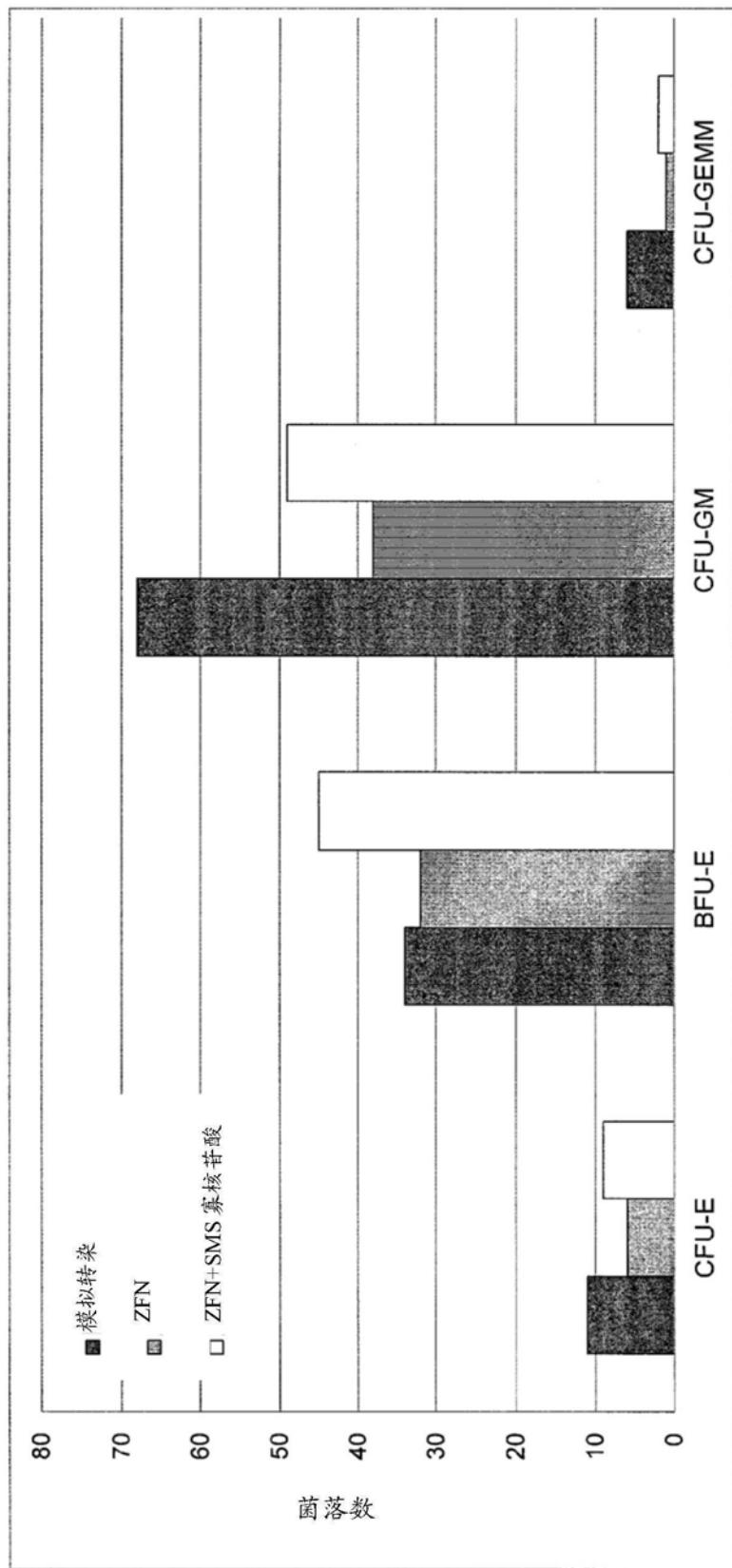


图8

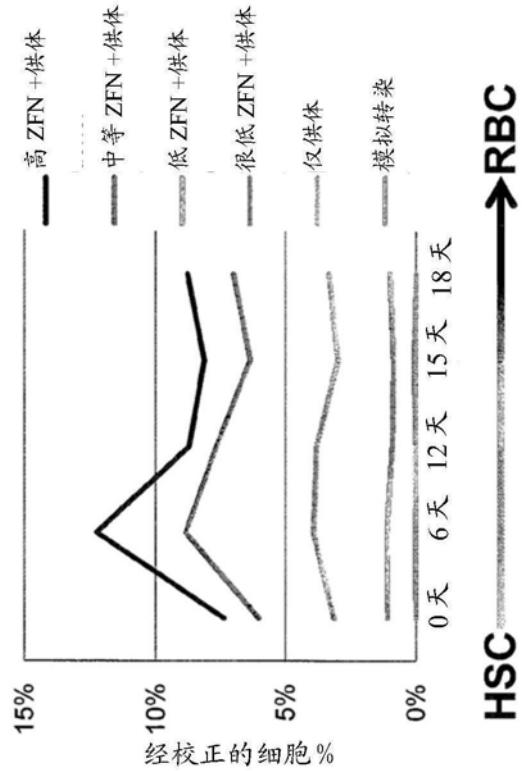


图9A

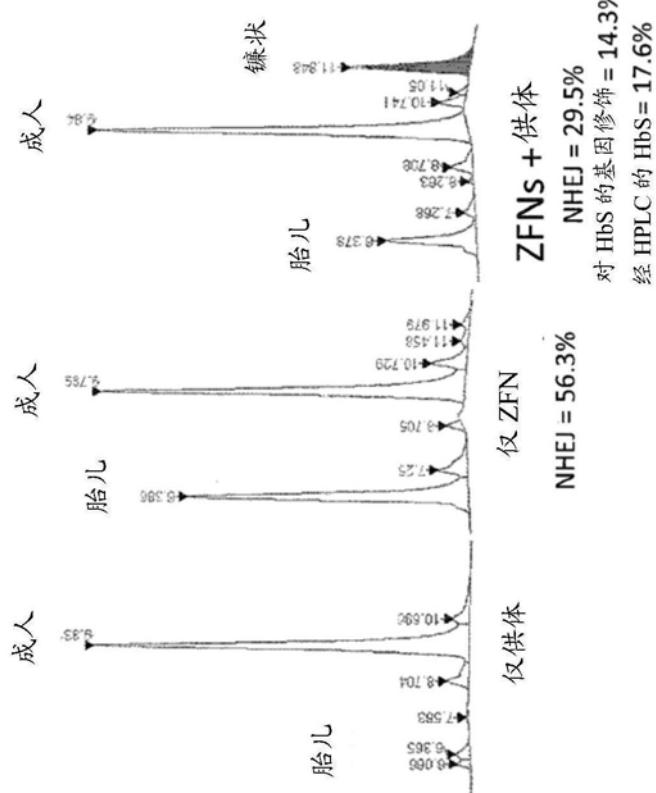


图9B

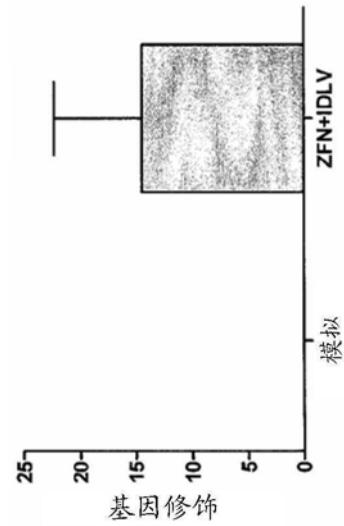


图10A

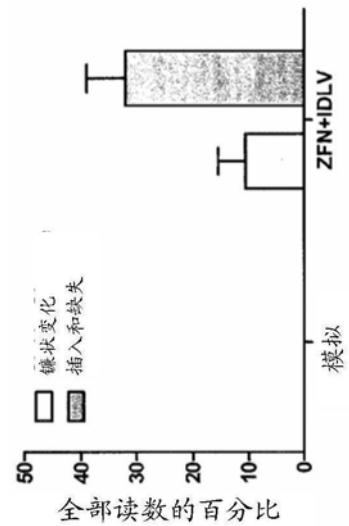


图10B

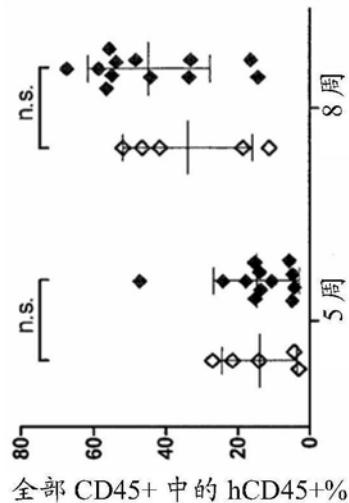


图10C

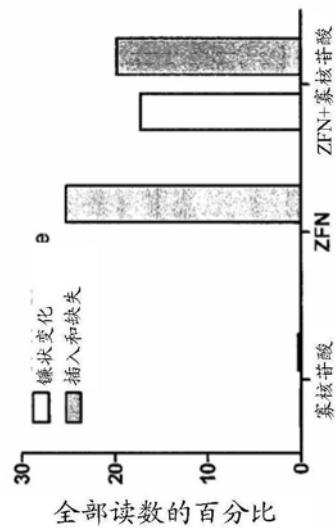


图 10D

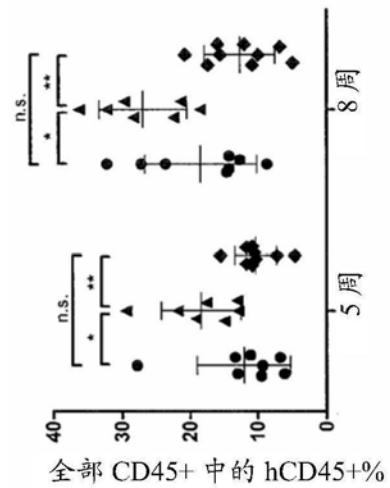


图 10E

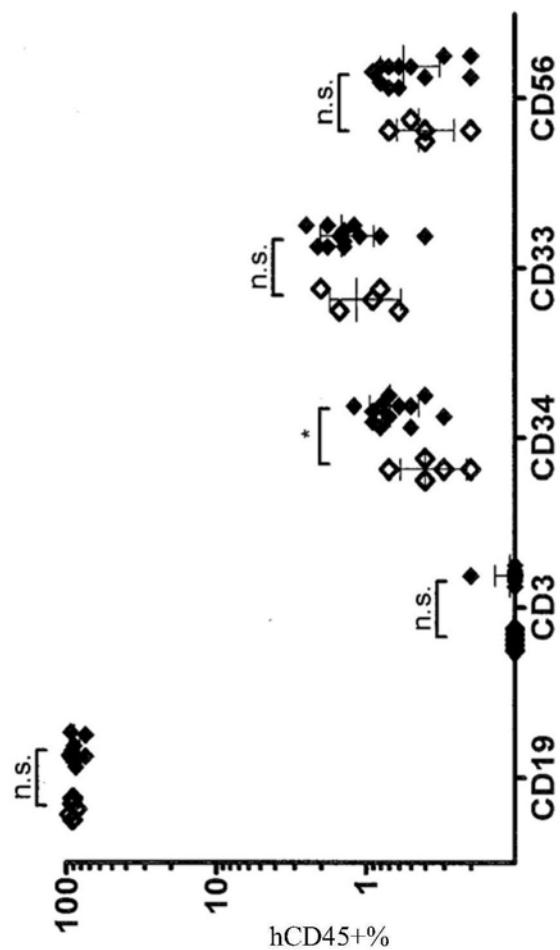


图11A

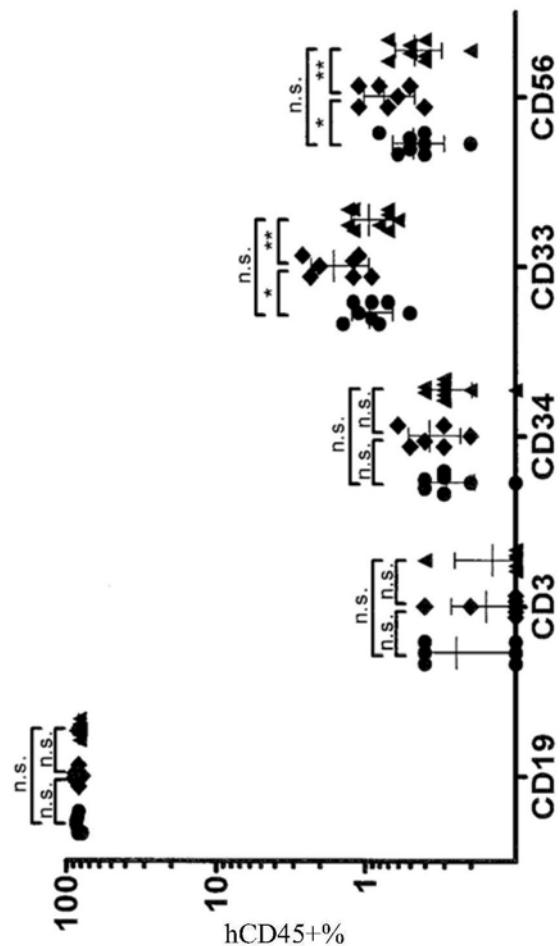


图11B

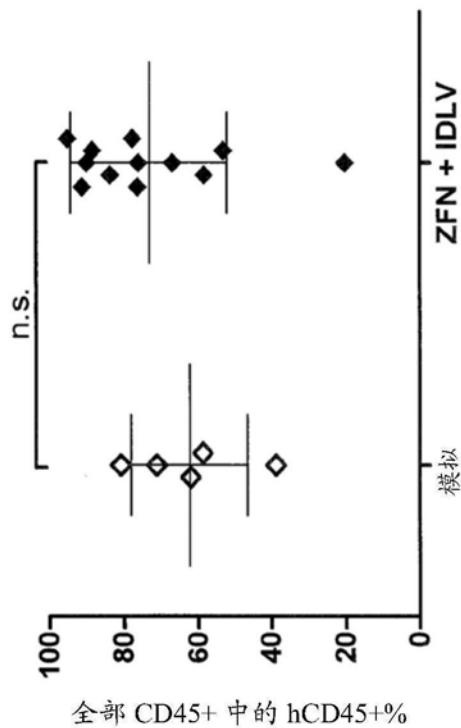


图12A

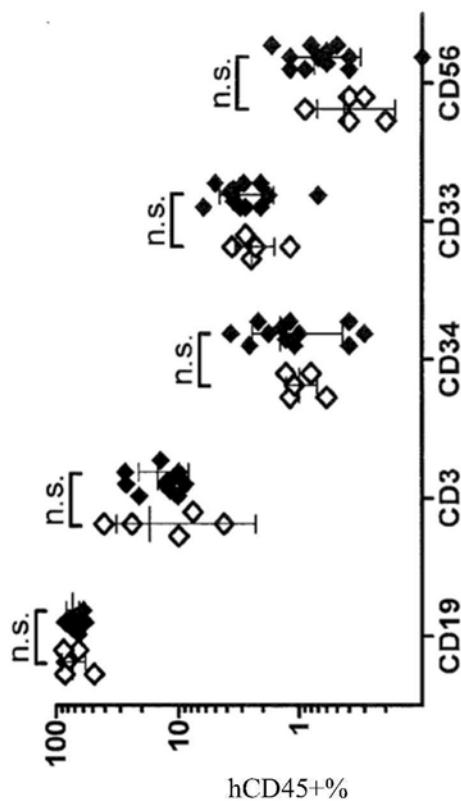


图12B

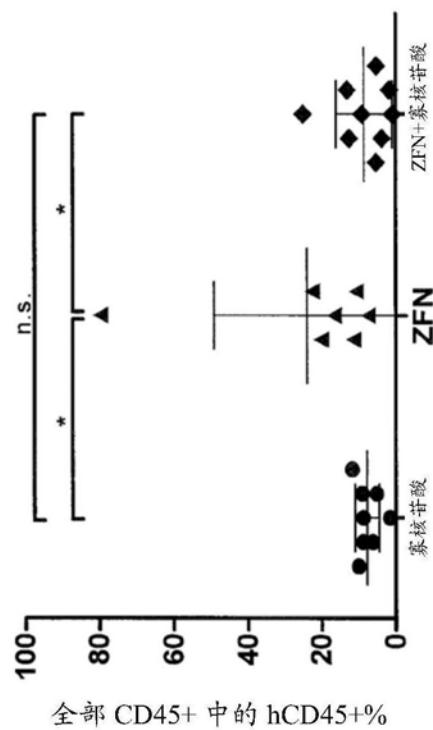


图12C

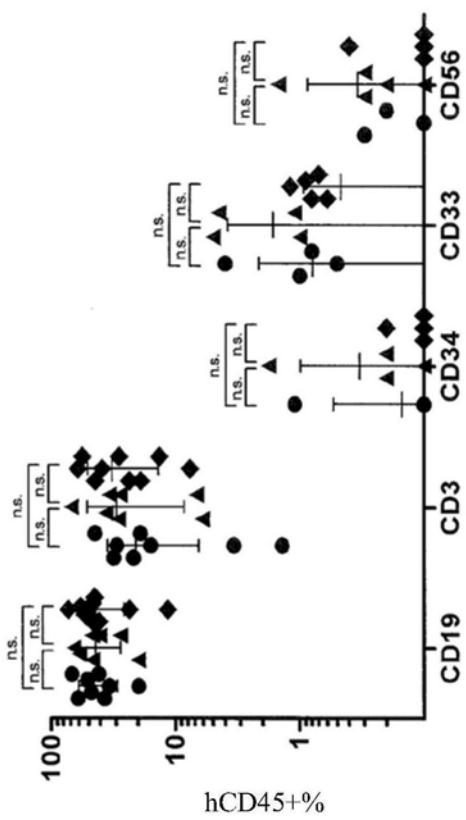


图12D

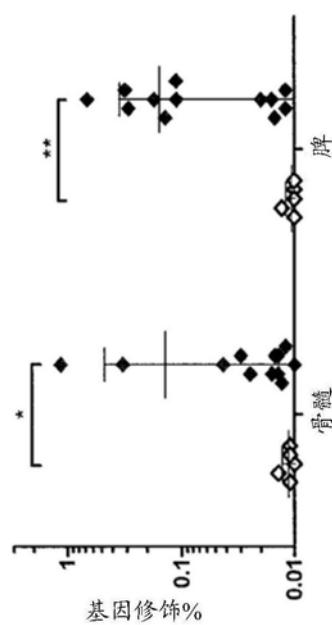


图13A

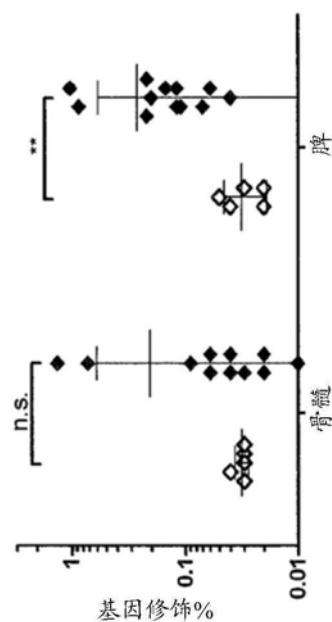


图13B

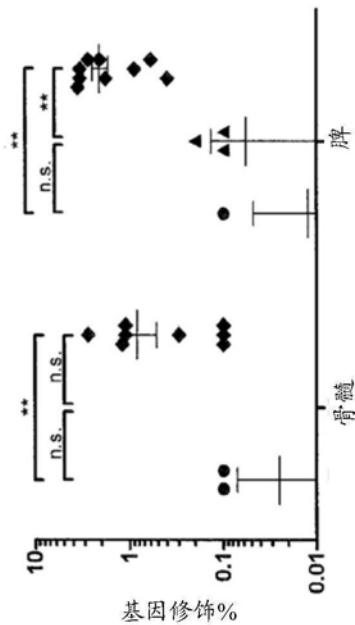


图13C

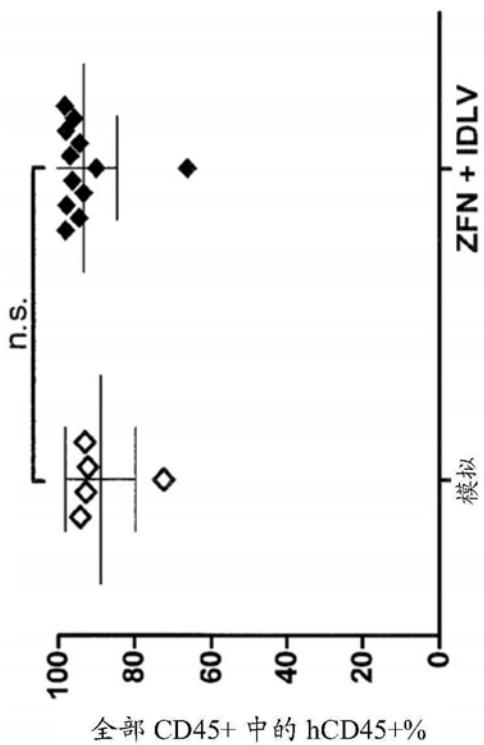


图14A

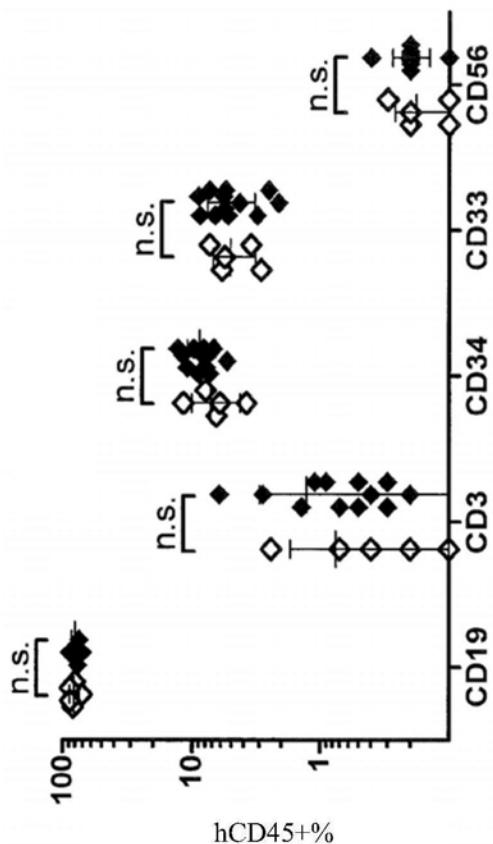


图14B

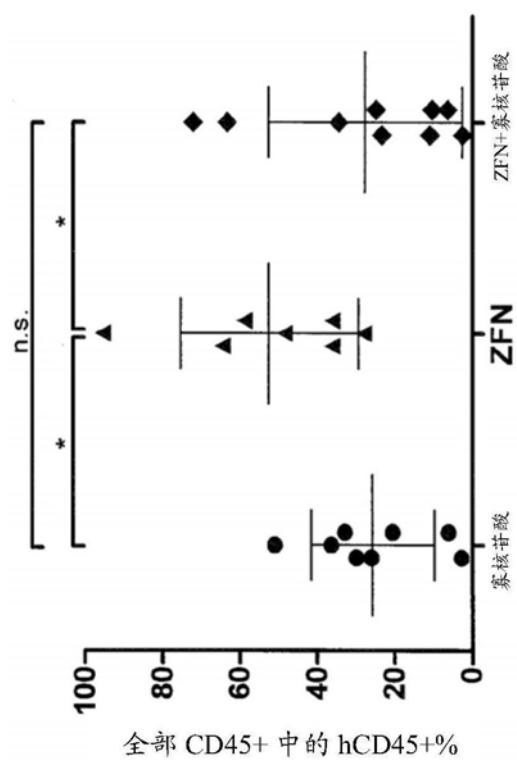


图14C

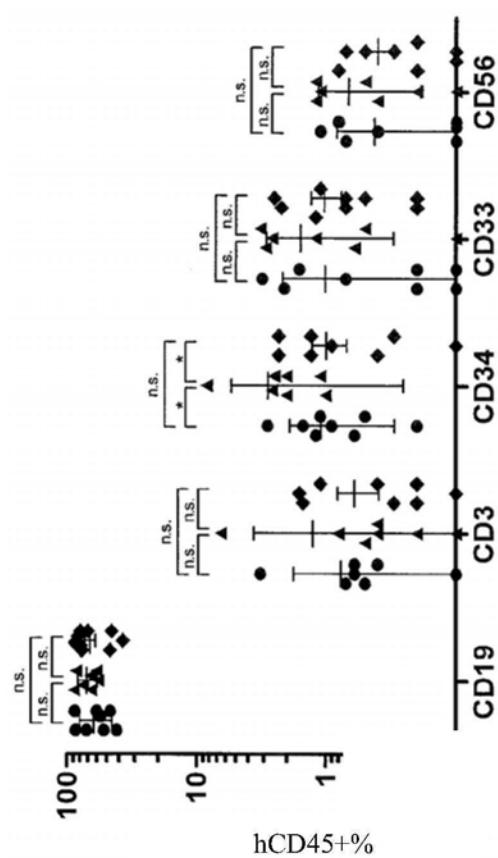


图14D

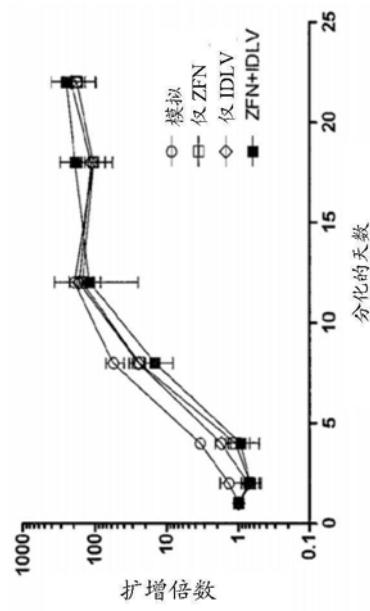


图15A

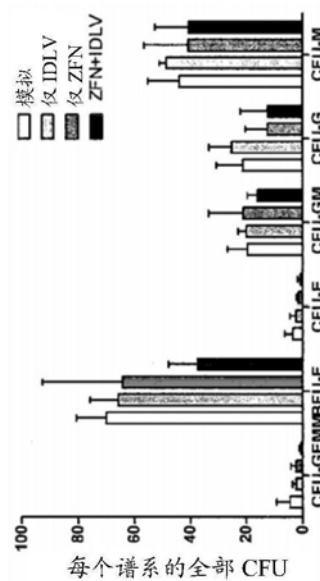


图15B

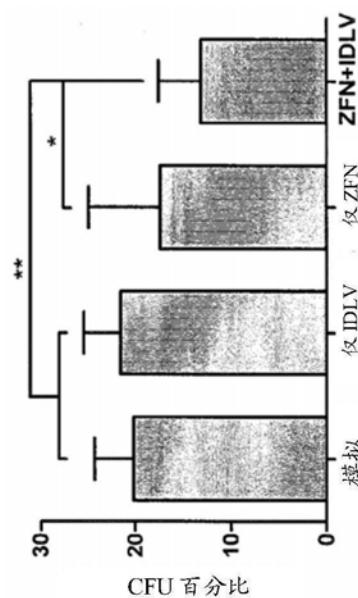


图15C

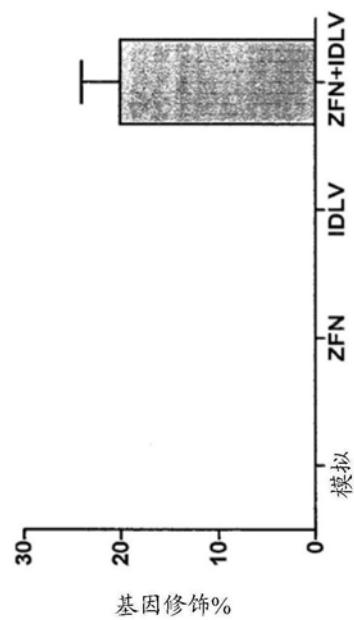


图16A

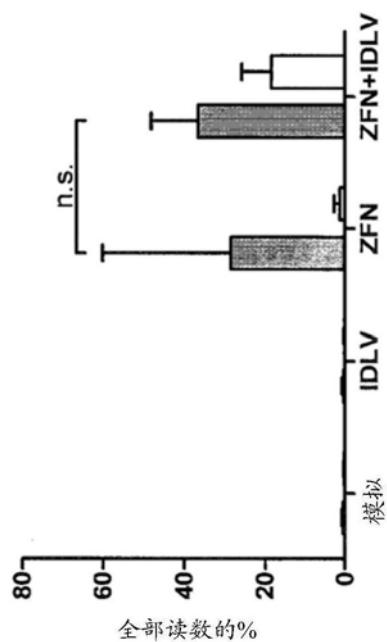


图16B

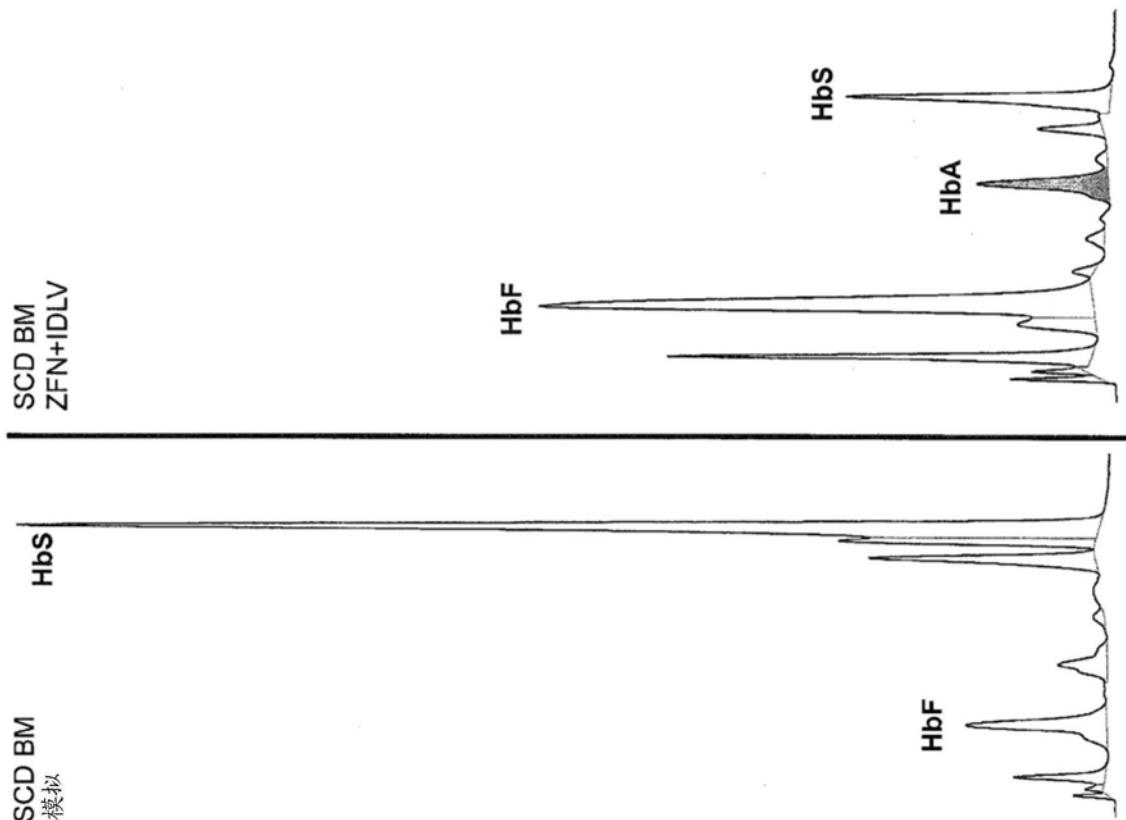


图16C

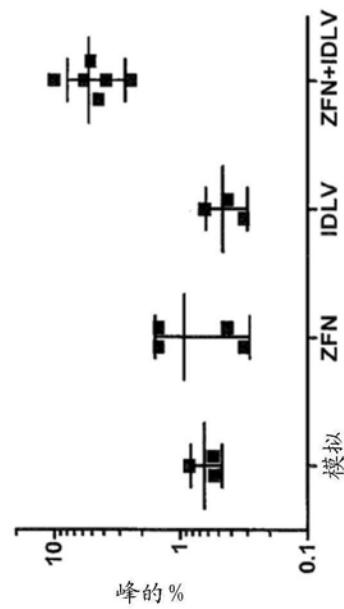


图16D

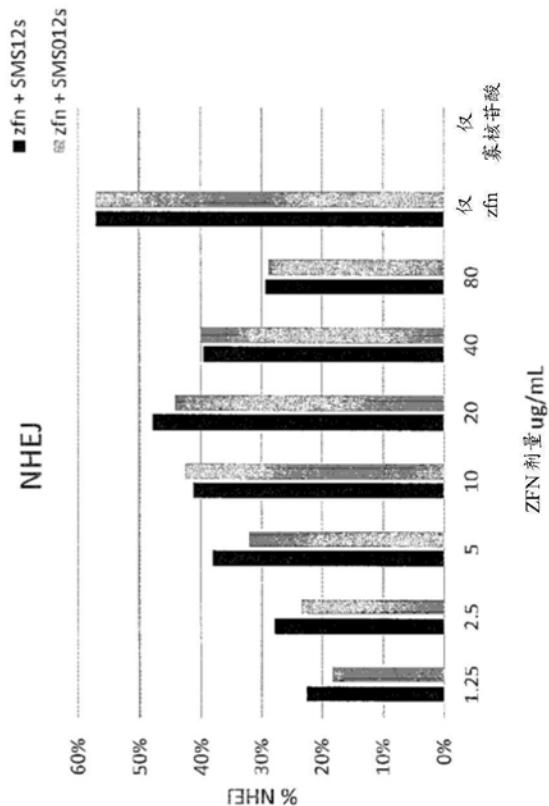


图17A

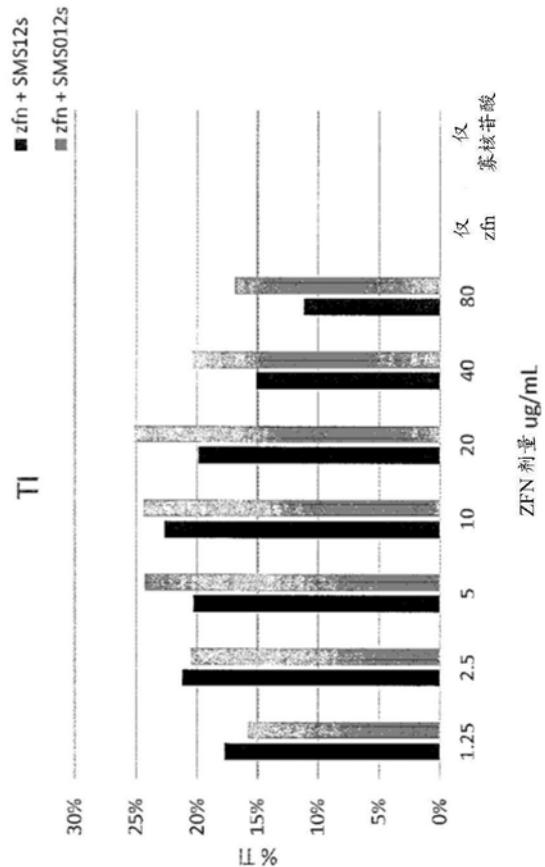


图17B