(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 118127006 A
(43)申请公布日 2024.06.04

(21)申请号 202311860300.2
(22)申请日 2017.03.14
(30)优先权数据
62/308,190 2016.03.14 US
62/456,615 2017.02.08 US
(62)分案原申请数据
201780029929.9 2017.03.14
(71)申请人 爱迪塔斯医药公司
地址 美国马萨诸塞州
(72)发明人 J·L·戈里 L·A·巴雷拉
(74)专利代理机构 北京市君合律师事务所 11517
专利代理师 张怡 张璐
(51)Int.Cl.
C12N 15/113(2010.01)
C12N 9/22(2006.01)
C12N 15/85(2006.01)
C12N 15/90(2006.01)
C12N 5/10(2006.01)
A61K 31/713(2006.01)
A61K 48/00(2006.01)
A61K 38/46(2006.01)
A61P 43/00(2006.01)
A61P 7/06(2006.01)

权利要求书1页 说明书134页
序列表（电子公布） 附图28页

(54)发明名称

用于治疗β-血红蛋白病的CRISPR/CAS相关方法和组合物

(57)摘要

披露了用于治疗β血红蛋白病的CRISPR/CAS相关组合物和方法。

CN 118127006 A

1.一种gRNA分子,所述gRNA分子包含靶向结构域,所述靶向结构域包含与完全或部分位于HBG1或HBG2调控区域内的靶结构域互补或部分互补的核苷酸序列。

2.一种gRNA分子,所述gRNA分子包含靶向结构域,所述靶向结构域包含与完全或部分位于HBG1或HBG2调控区域内的靶结构域互补或部分互补的核苷酸序列,其中所述靶向结构域包含与SEQ ID NO:251-901中任一个所示的核苷酸序列相同或相差不超过1、2、3、4、或5个核苷酸的核苷酸序列。

3.一种核酸组合物,所述核酸组合物包含:(a)编码包含靶向结构域的gRNA分子的核苷酸序列,所述靶向结构域包含与完全或部分位于HBG1或HBG2调控区域内的靶结构域互补或部分互补的核苷酸序列,其中所述gRNA分子是如权利要求1或2所述的gRNA分子。

4.一种组合物,所述组合物包含(a)如权利要求1或2所述的gRNA分子。

5.一种改变细胞的方法,所述方法包括使所述细胞与以下接触:

(a)如权利要求1或2所述的gRNA分子;以及

(b)RNA指导的核酸酶。

6.一种治疗有需要的受试者中的β-血红蛋白病的方法,所述方法包括使所述受试者或来自所述受试者的细胞与以下接触:

(a)如权利要求1或2所述的gRNA分子;以及

(b)RNA指导的核酸酶。

7.一种反应混合物,所述反应混合物包含:

(a)如权利要求1或2所述的gRNA分子、如权利要求3所述的核酸组合物、或如权利要求4所述的组合物;以及

来自患有β-血红蛋白病的受试者的细胞。

8.一种试剂盒,所述试剂盒包含,

(a)如权利要求1或2所述的gRNA分子、或编码所述gRNA分子的核酸组合物、以及以下中的一个或多个:

(b)RNA指导的核酸酶;

(c)包含靶向结构域的第二gRNA分子,所述靶向结构域包含与完全或部分位于HBG1或HBG2调控区域内的靶结构域互补或部分互补的核苷酸序列;以及

(d)编码(b)和(c)中的一个或多个的核酸组合物。

9.一种基因组编辑系统,所述基因组编辑系统包含:

(a)如权利要求1或2所述的gRNA分子;以及

(b)RNA指导的核酸酶。

用于治疗β-血红蛋白病的CRISPR/CAS相关方法和组合物

[0001] 本申请是申请号为201780029929.9、申请日为2017年3月14日、发明名称为“用于治疗β-血红蛋白病的CRISPR/CAS相关方法和组合物”的中国发明专利申请的分案申请,申请号为201780029929.9的中国发明专利申请为国际申请号为PCT/US2017/022377的PCT国际申请的中国国家阶段申请,该国际申请要求2016年3月14日提交的申请号为62/308,190的美国临时专利申请和2017年2月8日提交的申请号为62/456,615的美国临时专利申请的优先权,其全文内容在此以引用方式并入本文。

[0002] 相关申请的引用

[0003] 本申请要求2016年3月14日提交的美国临时申请号62/308,190和2017年2月8日提交的美国临时申请号62/456,615的权益,将其各自内容通过引用以其全文结合在此。

[0004] 序列表

[0005] 本申请含有对应于以ASCII格式经由EFS-Web提交的文件的序列表,并且通过引用以其全文结合在此。该以ASCII格式经由EFS-Web提交的文件是在2017年3月14日创建的所述ASCII副本被命名为8009WO00_SequenceListing.txt,并且大小为335KB。

技术领域

[0006] 本发明涉及用于编辑靶核酸序列或调控靶核酸序列表达的CRISPR/Cas相关方法和组分,及其与包括镰状细胞病和β-地中海贫血的β-血红蛋白病的结合应用。

背景技术

[0007] 血红蛋白(Hb)将氧气从肺部携带到红细胞或红血细胞(RBC)中的组织。在胚胎发育期间和出生后不久,血红蛋白以胎儿血红蛋白(HbF)的形式存在,胎儿血红蛋白是由两个α-珠蛋白链和两个γ-珠蛋白链组成的四聚体蛋白质。HbF在很大程度上被成人血红蛋白(HbA)替换,成人血红蛋白是一种四聚体蛋白质,其中HbF的γ-珠蛋白链通过称为珠蛋白转换的过程被β-珠蛋白链替换。HbF在携带氧气方面比HbA更有效。普通成人的总血红蛋白小于1%的HbF(Thein 2009)。α-血红蛋白基因位于16号染色体上,而β-血红蛋白基因(HBB)、γ(γ^A)-珠蛋白链(HBG1,也称为γ珠蛋白A)和Gγ(γ^G)-珠蛋白链(HBG2,也称为γ珠蛋白G)位于珠蛋白基因簇内的11号染色体上(即,珠蛋白基因座)。

[0008] HBB的突变可引起血红蛋白障碍(即,血红蛋白病),包括镰状细胞病(SCD)和β-地中海贫血(β-Thal)。美国大约93,000人被诊断患有血红蛋白病。全球每年有300,000儿童出生患有血红蛋白病(Angastiniotis 1998)。因为这些病症与HBB突变有关,所以他们的症状典型地直到珠蛋白从HbF转换为HbA后才表现。

[0009] SCD是美国最常见的遗传性血液病,影响大约80,000人(Brousseau 2010)。SCD在非洲血统的人中最常见,SCD的流行率为500人中有1人。在非洲,SCD的流行率为1500万(Aliyu,2008)。SCD在印度人,沙特阿拉伯人和地中海人后裔中也更常见。在西班牙裔美国人后裔中,镰状细胞病的流行率为1000人中有1人(Lewis 2014)。

[0010] SCD由HBB基因中的单个纯合突变引起,c.17A>T(HbS突变)。镰状突变是HBB上的点

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

[0011] 镰状形状的RBC引起多种症状,包括贫血、镰状细胞危象、血管阻塞性危象,再生障碍性危象和急性胸部综合征。镰状形状的RBC比野生型RBC弹性小,因此不能容易地通过毛细血管床并导致阻塞和缺血(即,血管阻塞)。当镰状细胞阻塞器官毛细血管床中的血流导致疼痛、缺血和坏死时,就会发生血管阻塞性危象。这些发作典型地持续5天至7天。脾脏在清除功能失调的RBC中起作用,因此典型地在儿童早期期间扩大并且频繁发生血管阻塞性危象。到儿童期结束时,SCD患者的脾经常梗塞,导致自体脾切除。溶血是SCD的一个恒定特征并引起贫血。镰状细胞在循环中存活10天至20天,而健康的RBC存活90天至120天。必要时输血SCD受试者以维持足够的血红蛋白水平。频繁的输血使受试者有感染HIV、乙型肝炎、和丙型肝炎的风险。受试者还可能患有急性胸部危象和四肢、终末器官、和中枢神经系统的梗塞。

[0012] 患有SCD的受试者的生命期望降低。通过对危象和贫血进行认真的,终身的管理,SCD患者的预后正在稳定地改进。在2001年,患有镰状细胞病的受试者的平均生命期望是50岁中后期。目前对SCD的治疗涉及危象期间的水合和疼痛管理,以及根据需要进行输血以校正贫血。

[0013] 地中海贫血(例如,β-Thal、δ-Thal、和β/δ-Thal)引起慢性贫血。估计β-Thal影响全球大约100,000人中有1人。它在某些群体中的流行率较高,包括欧洲后裔的群体,其流行率大约为10,000人中有1人。除非通过终身输血和螯合疗法治疗,否则重型β-Thal是疾病的更严重形式,是危及生命的。在美国,大约有3,000名患有重型β-Thal的受试者。中间型β-Thal不需要输血,但可能引起生长延迟和显著的全身异常,并且频繁地需要终身螯合疗法。尽管HbA构成成人RBC中大多数血红蛋白,但大约3%的成人血红蛋白是HbA_2形式,HbA变体是两个γ-珠蛋白链被两个δ(Δ)-珠蛋白链替换。δ-Thal与引起HBD表达损失的Δ血红蛋白基因(HBD)突变有关。HBD突变的共遗传可以通过将HbA_2水平降低至正常范围来掩盖β-Thal(即,β/δ-Thal)的诊断(Bouva 2006)。β/δ-Thal通常由两个等位基因中HBB和HBD序列的缺失引起。在纯合的($\delta^o/\delta^o\beta^o/\beta^o$)患者中,表达HBG,导致单独产生HbF。

[0014] 与SCD一样,β-Thal是由HBB基因突变引起的。导致β-Thal的最常见的HBB突变是:c.-136C>G、c.92+1G>A、c.92+6T>C、c.93-21G>A、c.118C>T、c.316-106C>G、c.25_26delAA、c.27_28insG、c.92+5G>C、c.118C>T、c.135delC、c.315+1G>A、c.-78A>G、c.52A>T、c.59A>G、c.92+5G>C、c.124_127delTTCT、c.316-197C>T、c.-78A>G、c.52A>T、c.124_127delTTCT、c.316-197C>T、c.-138C>T、c.-79A>G、c.92+5G>C、c.75T>A、c.316-2A>G、和c.316-2A>C。与β-Thal有关的这些和其他突变引起β-珠蛋白链的突变或缺失,这导致正常Hbα-血红蛋白与β-血红蛋白比率的破坏。过量的α-珠蛋白链在骨髓中的红系前体中沉淀。

[0015] 在重型β-Thal中,HBB的两个等位基因都含有无意义突变、移码突变、或剪接突变,导致完全不存在β-珠蛋白产生(表示为β^o/β^o)。重型β-Thal导致β-珠蛋白链的严重减少,导致红系细胞中α-珠蛋白链的显著沉淀和更严重的贫血。

[0016] 中间型β-Thal由HBB的5′或3’非翻译区域突变、启动子区域突变、或HBB多聚腺苷酸化信号或HBB基因内的剪接突变导致。患者基因型表示为β^0/β^+或β^+/β^+。β^0代表不存在β-珠蛋白链的表达;β^+代表功能失调但存在的β-珠蛋白链。表型表达因患者而异。由于存在一些β-珠蛋白的产生,中间型β-Thal导致红系前体中α-珠蛋白链的沉淀较少,并且与重型β-Thal相比导致较少的严重贫血。然而,继发于慢性贫血的红系谱系扩增有更显著的后果。

[0017] 具有重型β-Thal的受试者存在于6个月和2岁之间,并且患有未能茁壮成长、发热、肝脾大和腹泻。足够的治疗包括定期输血。重型β-Thal疗法还包括脾切除术和羟基脲治疗。如果患者定期输血,他们将正常发育,直到第二个十年开始。那时,他们需要螯合疗法(除了继续输血)以防止铁超过载的并发症。铁过载可能表现为生长延迟或性成熟延迟。在成人期,不充分的螯合疗法可能导致心肌病、心律失常、肝纤维化和/或肝硬化、糖尿病、甲状腺和甲状旁腺异常、血栓症和骨质疏松症。频繁的输血还使受试者有感染HIV、乙型肝炎、和丙型肝炎的风险。

[0018] 中间型β-Thal受试者通常存在于2岁至6岁之间。他们通常不需要输血。然而,由于红系谱系的慢性肥大而发生骨异常以补偿慢性贫血。由于骨质疏松症,受试者可能有长骨骨折。髓外红细胞生成是常见的并且导致脾,肝和淋巴结的扩大。它还可能引起脊髓压缩和神经系统问题。受试者还患有下肢溃疡并且血栓形成事件的风险增加,包括中风、肺栓塞和深静脉血栓形成。中间型β-Thal的治疗包括脾切除术、叶酸补充、羟基脲疗法和髓外肿块的放射治疗。螯合疗法用于发生铁过载的受试者。

[0019] β-Thal患者的生命期望通常会降低。患有重型β-Thal且未接收输血疗法的受试者通常在其第二或第三个十年死亡。接收常规输血和足够螯合疗法的重型β-Thal受试者可以活到第五个十年甚至更长时间。继发于铁毒性的心脏衰竭是由于铁毒性导致的重型β-Thal受试者死亡的主要原因。

[0020] 目前正在发育SCD和β-Thal的各种新治疗。目前正在临床试验中研究经由基因疗法递送校正的HBB基因。然而,这种途径的长期功效和安全性尚不清楚。已经证明用来自HLA匹配的异基因干细胞供体的造血干细胞移植治疗SCD和β-Thal,但是所述方法涉及风险,包括与切除疗法有关的风险,以准备移植受试者和移植后移植物抗宿主疾病的风险。另外,通常无法识别匹配的异基因供体。因此,需要改进的管理这些和其他血红蛋白病的方法。

发明内容

[0021] 本文在某些实施例中提供了使用基因组编辑系统(例如,CRISPR/Cas介导的基因组编辑系统)增加受试者或细胞中一种或多种γ-珠蛋白基因(例如,HBG1、HBG2、或HBG1和HBG2)的表达(即,转录活性)的方法。在某些实施例中,这些方法可利用任何修复机制来改变(例如,删除、破坏或修饰)一种或多种γ-珠蛋白基因调控元件的全部或部分。在某些实施例中,这些方法可利用DNA修复机制,例如,NHEJ或HDR来删除或破坏一种或多种γ-珠蛋白基因调控元件(例如,沉默子,增强子,启动子或隔离子)。在某些实施例中,这些方法利用DNA修复机制,例如,HDR,来改变,包括突变、插入、删除或破坏γ-珠蛋白基因调控元件中的一个或多个核苷酸的序列(例如,沉默子、增强子、启动子、或隔离子)。在某些实施例中,这些方法利用一种或多种DNA修复机制的组合,例如,NHEJ和HDR。在某些实施例中,这些方法导致与天然存在的HPFH变体有关的γ-珠蛋白调控元件的突变或变异,包括,例如,HBG1

13bp del c.-114至-102、4bp del c.-225至-222、c.-114C>T、c.-117G>A、c.-158C>T、c.-167C>T、c.-170G>A、c.-175T>G、c.-175T>C、c.-195C>G、c.-196C>T、c.-198T>C、c.-201C>T、c.-251T>C、或c.-499T>A、或HBG2 13bp del c.-114至-102、c.-109G>T、c.-114C>A、c.-114C>T、c.-157C>T、c.-158C>T、c.-167C>T、c.-167C>A、c.-175T>C、c.-202C>G、c.-211C>T、c.-228T>C、c.-255C>G、c.-309A>G、c.-369C>G、或c.-567T>G。

[0022] 本文在某些实施例中提供了使用CRISPR/Cas介导的基因组编辑来治疗有需要的受试者中的β-血红蛋白病以增加一种或多种γ-珠蛋白基因(例如,HBG1、HBG2、或HBG1和HBG2)的表达(即,转录活性)的方法。在某些实施例中,这些方法利用DNA修复机制,例如,NHEJ或HDR来删除或破坏一种或多种γ-珠蛋白基因调控元件(例如,沉默子,增强子,启动子或隔离子)。在某些实施例中,这些方法利用DNA修复机制,例如,HDR,来改变,包括突变、插入、删除或破坏γ-珠蛋白基因调控元件中的一个或多个核苷酸的序列(例如,沉默子、增强子、启动子、或隔离子)。在某些实施例中,这些方法利用一种或多种DNA修复机制的组合,例如,NHEJ和HDR。在某些实施例中,这些方法导致与天然存在的HPFH变体有关的γ-珠蛋白调控元件的突变或变异,包括,例如HBG1 13bp del c.-114至-102、4bp del c.-225至-222、c.-114C>T、c.-117G>A、c.-158C>T、c.-167C>T、c.-170G>A、c.-175T>G、c.-175T>C、c.-195C>G、c.-196C>T、c.-198T>C、c.-201C>T、c.-251T>C、或c.-499T>A、或HBG2 13bp del c.-114至-102、c.-109G>T、c.-114C>A、c.-114C>T、c.-157C>T、c.-158C>T、c.-167C>T、c.-167C>A、c.-175T>C、c.-202C>G、c.-211C>T、c.-228T>C、c.-255C>G、c.-309A>G、c.-369C>G、或c.-567T>G。在某些实施例中,β-血红蛋白病是SCD或β-Thal。

[0023] 本文在某些实施例中提供了用于CRISPR/Cas介导的增加一种或多种γ-珠蛋白基因(例如HBG1、HBG2、或HBG1和HBG2)的表达(即,转录活性)的方法的gRNA。在某些实施例中,这些gRNA包含靶向结构域,所述结构域包含SEQ ID NO:251-901所示的核苷酸序列。在某些实施例中,这些gRNA进一步包含第一互补结构域、第二互补结构域、连接结构域、5'延伸结构域,近端结构域或尾部结构域中的一个或多个。在另一个实施例中,gRNA是模块化gRNA。在其他实施例中,所述gRNA是单分子(或嵌合的)gRNA。

附图说明

[0024] 图1A-1I是若干示例性gRNA的表示。

[0025] 图1A描绘了部分来源于(或部分地在序列上建模)化脓链球菌(S.pyogenes)呈双链体结构的模块化gRNA分子(按照出现次序分别是SEQ ID NO:39和40);

[0026] 图1B描绘了部分来源于化脓链球菌呈双链体结构的单分子gRNA分子(SEQ ID NO:41);

[0027] 图1C描绘了部分来源于化脓链球菌呈双链体结构的单分子gRNA分子(SEQ ID NO:42);

[0028] 图1D描绘了部分来源于化脓链球菌呈双链体结构的单分子gRNA分子(SEQ ID NO:43);

[0029] 图1E描绘了部分来源于化脓链球菌呈双链体结构的单分子gRNA分子(SEQ ID NO:44);

[0030] 图1F描绘了部分来源于嗜热链球菌(S.thermophilus)呈双链体结构的模块化

gRNA分子(按照出现次序分别是SEQ ID NO:45和46);

[0031] 图1G描绘了化脓链球菌和嗜热链球菌的模块化gRNA分子(按照出现次序分别是SEQ ID NO:39、45、47和46)的比对。

[0032] 图1H-1I描绘了单分子gRNA分子的另外示例性结构。

[0033] 图1H示出了部分来源于化脓链球菌呈双链体结构的单分子gRNA分子(SEQ ID NO:42)的示例性结构。

[0034] 图1I示出了部分来源于金黄色葡萄球菌(S.aureus)呈双链体结构的单分子gRNA分子(SEQ ID NO:38)的示例性结构。

[0035] 图2A-2G描绘了Cas9序列的比对(Chylinski 2013)。N-末端RuvC样结构域加框并且以“Y”指示。其他两个RuvC样结构域加框并且以“B”指示。HNH样结构域被加框并以“G”指示。Sm:变形链球菌(SEQ ID NO:1)、Sp:化脓链球菌(SEQ ID NO:2)、St:嗜热链球菌(SEQ ID NO:4)、和Li:无害利斯特菌(SEQ ID NO:5)。“基序”(SEQ ID NO:14)是基于四个序列的共有序列。所有四个序列中保守的残基用单字母氨基酸缩写指示;“*”指示在四个序列的任一者的相应位置中发现的任何氨基酸;并且“-”指示不存在。

[0036] 图3A-3B示出了来自披露于Chylinski 2013中的Cas9分子(SEQ ID NO:52-95、120-123)的N-末端RuvC样结构域的比对。图3B的最后一行识别了4个高度保守的残基。

[0037] 图4A-4B示出了来自披露于Chylinski 2013中的除去序列异常值的Cas9分子(SEQ ID NO:52-123)的N-末端RuvC样结构域的比对。图4B的最后一行识别了3个高度保守的残基。

[0038] 图5A-5C示出了来自披露于Chylinski 2013中的Cas9分子(SEQ ID NO:124-198)的HNH样结构域的比对。图5C的最后一行识别了保守残基。

[0039] 图6A-6B示出了来自披露于Chylinski 2013中的除去序列异常值的Cas9分子(SEQ ID NO:124-141、148、149、151-153、162、163、166-174、177-187、194-198)的HNH样结构域的比对。图6B的最后一行识别了3个高度保守的残基。

[0040] 图7示出了使用示例性gRNA序列(SEQ ID NO:42)的gRNA结构域命名法。

[0041] 图8A和8B提供了化脓链球菌Cas9的结构域组织的示意性表示。图8A参照Cas9的两种叶片(识别(REC)叶片和核酸酶(NUC)叶片)示出了Cas9结构域的组织,包括氨基酸位置。图8B示出了83个Cas9直向同源物中每个结构域的百分比同源性。

[0042] 图9A至9C提供了在珠蛋白基因座的背景下HBG1和HBG2基因的示意图。指示了编码序列(CDS)、mRNA区域和基因。(A)示出了靶向gRNA设计的区域(虚线和括号,指示HBG1和HBG2基因的近端遗传区域)。(B)指示核心启动子元件。(C)指示转录激活因子和转录阻遏蛋白可以结合以调控基因表达的基因调控区域中的基序。注意基序和靶向gRNA设计的基因组区域之间的重叠。指示了引起HPFH的HBG1和HBG2基因调控区域中的缺失的实例,以及与每种有关的%HbF。

[0043] 图10A至10F示出了来自gRNA筛选的数据,所述数据用于在人类K562红白血病细胞中掺入13bp的del c.-114至-102HPFH突变。(A)通过T7E1核酸内切酶测定分析确定的用编码化脓链球菌特异性gRNA的DNA和编码化脓链球菌Cas9的质粒DNA电穿孔后从K562细胞提取的基因组DNA扩增的HBG1和HBG2基因座特异性PCR产物的基因编辑。(B)通过用编码指示的gRNA和Cas9质粒的DNA电穿孔后从K562细胞中提取的基因组DNA中的HBG1基因座扩增的

PCR产物的DNA序列分析确定的基因编辑。(C)通过用编码指示的gRNA和Cas9质粒的DNA电穿孔后从K562细胞中提取的基因组DNA中的HBG2基因座扩增的PCR产物的DNA序列分析确定的基因编辑。对于(B)和(C),缺失的编辑事件(插入,缺失)和亚型的类型(13nt靶向部分[12nt HPFH]或完全[13nt至26nt HPFH]缺失,其他序列缺失[其他缺失])由不同的阴影/图案条指示。(D)-(F)HBG1基因调控区域缺失的实例。

[0044] 图11A至11C描绘了电穿孔后人类脐血(CB)和人类成人$CD34^+$细胞中基因编辑的结果,其中复合的RNP体外转录的化脓链球菌gRNA靶向特异性的13nt序列进行缺失(HBG gRNAs Sp35(包含SEQ ID NO:339)和Sp37(包含SEQ ID NO:333))。图11A描绘了从用指示的RNP或供体匹配的未处理的对照细胞(n=3CB $CD34^+$细胞,3个独立的实验)处理的CB $CD34^+$细胞提取的gDNA扩增的HBG1和HBG2特异性PCR产物的T7E1分析检测到的indel百分比。示出的数据代表平均值,误差条相应于三个独立供体/实验的标准差。图11B描绘了从用指示的RNP或供体匹配的未处理的对照细胞(n=3CB $CD34^+$细胞,n=3mPB $CD34^+$细胞,3个独立的实验)处理的CB $CD34^+$细胞或成人CD34+细胞提取的gDNA扩增的HBG2特异性PCR产物的T7E1分析检测到的indel百分比。示出的数据代表平均值,误差条相应于三个独立供体/实验的标准差。图11C(顶部组)描绘了通过T7E1分析对用HBG Sp35 RNP或HBG Sp37 RNP+/-ssODN1(SEQ ID NO:906)或PhTx ssODN1(SEQ ID NO:909)电穿孔的人类CB $CD34^+$细胞中提取的gDNA扩增的HBG2PCR产物进行检测的编辑。图11C(左下部组)示出了通过Sanger DNA序列分析从用HBG Sp37 RNP和ssODN1和PhTx ssODN1编辑的细胞的gDNA确定的基因编辑水平。图11C(右下部组)示出了从左下部组中存在的数据中检测到的总缺失中检测到的特定类型的缺失。

[0045] 图12A至12C描绘了K562红白血病细胞中HBG1和HBG2的基因编辑。图12A描绘了通过T7E1分析检测到的NHEJ(indel),其中HBG1和HBG2 PCR产物在核转染后三天从用K562细胞提取的gDNA扩增,其中RNP与指示的gRNA复合。图12B描绘了从HBG1基因座扩增的PCR产物的Sanger DNA序列分析,所述PCR产物用于与靶向13nt HPFH序列(Sp35(包含SEQ ID NO:339)、Sp36(包含SEQ ID NO:338)、Sp37(包含SEQ ID NO:333)的gRNA复合的Cas9蛋白进行核转染的细胞。图12C描绘了从HBG2基因座扩增的PCR产物的Sanger DNA序列分析,所述PCR产物用于与靶向13bp HPFH序列(Sp35、Sp36、Sp37)的gRNA复合的Cas9蛋白进行核转染的细胞。对于图12B和图12C将缺失细分为含有13bp靶向缺失(HPFH缺失、18nt-26nt缺失、>26nt缺失)和不含有13bp缺失(<12nt缺失、其他缺失、插入)的缺失。

[0046] 图13描绘了用编码13bp缺失的HBG Sp37 RNP+/-ssODN电穿孔mPB $CD34^+$细胞后成人人类动员的外周血(mPB)$CD34^+$细胞中HBG的基因编辑和RNP处理的细胞的红系后裔中胎儿血红蛋白的诱导。图13A描绘了从用RNP或供体匹配的未处理对照细胞处理的mPB $CD34^+$细胞提取的gDNA扩增的HBG2 PCR产物的T7E1分析检测到的编辑百分比。图13B描绘了在与RNP处理的和未处理的供体匹配的对照mPB $CD34^+$细胞分化的第7天成红细胞中HBG mRNA表达的倍数变化。将mRNA水平归一化为GAPDH并校准至在相应的分化天数中未处理对照中检测到的水平。

[0047] 图14描绘了来自相同供体的RNP处理和未处理的mPB $CD34^+$细胞的离体分化潜能。图14A示出了造血骨髓/红系集落形成细胞(CFC)潜能,其中指示了集落的数量和亚型(GEMM:粒细胞-红系-单核细胞-巨噬细胞集落、E:红系集落、GM:粒细胞-巨噬细胞集落、M:

巨噬细胞集落、G:粒细胞集落)。图14B描绘了在指示的时间点和指示的样品通过流式细胞术分析确定的红系分化时间过程中表达的血型糖蛋白A的百分比。

具体实施方式

[0048] 定义

[0049] 如本文所用的“结构域”是用于描述蛋白质或核酸的区段。除非另外指明,结构域不需要具有任何特定功能特性。

[0050] 如下进行两个序列之间的同源性或序列一致性(这些术语在本文可互换地使用)的计算。将这些序列进行比对用于最优比较的目的(例如,用于最优比对,可以在第一和第二氨基酸或核酸序列中的一个或两个中引入空位,并且出于比较的目的,可以不考虑非同源序列)。使用具有Blossum 62打分矩阵(其中空位罚分为12,空位延伸罚分为4,并且移码空位罚分为5)的GCG软件包中的GAP程序,将最优比对确定为最佳评分。然后比较相应的氨基酸位置或核苷酸位置处的氨基酸残基或核苷酸。当第一序列中的位置被与在第二序列中的相应位置相同的氨基酸残基或核苷酸占据时,则所述分子在那个位置是一致的。两个序列之间的百分比一致性是由所述序列共享的相同位置的数量的函数。

[0051] 如本文所用的“多肽”是指具有少于100个氨基酸残基的氨基酸的聚合物。在实施例中,它具有少于50、20、或10个氨基酸残基。

[0052] 如本文所用的“alt-HDR”、“替代性同源定向修复”或“替代性HDR”是指使用同源核酸(例如,内源性同源序列(例如姐妹染色单体)或外源核酸(例如模板核酸))修复DNA损伤的过程。alt-HDR与典型HDR的不同之处在于,所述过程利用与典型HDR不同的途径,并且可以被典型HDR介体RAD51和BRCA2抑制。此外,alt-HDR使用单链或有切口的同源核酸来修复断裂。

[0053] 如本文所用的“典型HDR”或典型同源定向修复是指使用同源核酸(例如,内源性同源序列(例如,姐妹染色单体)或外源核酸(例如,模板核酸))修复DNA损伤的过程。当在双链断裂处已有显著切除时,典型HDR通常起作用,形成DNA的至少一个单链部分。在正常细胞中,HDR通常涉及一系列步骤,诸如断裂的识别、断裂的稳定、切除、单链DNA的稳定、DNA交叉中间体的形成、交叉中间体的拆分以及连接。所述过程需要RAD51和BRCA2,并且同源核酸通常是双链的。

[0054] 除非另有说明,如本文所用的术语“HDR”涵盖典型HDR和alt-HDR两者。

[0055] 如本文所用的“非同源末端连接”或“NHEJ”是指连接介导的修复和/或非模板介导的修复,包括典型NHEJ(cNHEJ)、替代性NHEJ(altNHEJ)、微同源性介导的末端连接(MMEJ)、单链退火(SSA)以及合成依赖性微同源性介导的末端连接(SD-MMEJ)。

[0056] 如本文所用的“参考分子”是指经修饰或候选分子与其作比较的分子。例如,参考Cas9分子是指经修饰或候选Cas9分子与其作比较的Cas9分子。同样地,参考gRNA是指经修饰或候选gRNA分子与其作比较的gRNA分子。经修饰或候选分子可以基于序列(例如,经修饰或候选分子可以与参考分子具有X%序列一致性或同源性)、或活性(例如,经修饰或候选分子可以具有参考分子的X%的活性)与参考分子进行比较。例如,在参考分子是Cas9分子的情况下,经修饰或候选分子可以表征为具有不多于参考Cas9分子的核酸酶活性的10%。参考Cas9分子的实例包括天然存在的未经修饰的Cas9分子,例如来自化脓链球菌、金黄色葡

菊球菌、嗜热链球菌或脑膜炎奈瑟氏菌的天然存在的Cas9分子。在某些实施例中,参考Cas9分子是具有和与其进行比较的经修饰或候选Cas9分子最接近序列一致性或同源性的天然存在的Cas9分子。在某些实施例中,参考Cas9分子是具有天然存在或已知的序列的亲本分子,其上已经进行突变以得到经修饰或候选Cas9分子。

[0057] 术语“基因组编辑系统”是指具有RNA指导的DNA编辑活性的任何系统。本披露内容的基因组编辑系统包括至少两种适应天然存在的CRISPR系统的组分:指导RNA(gRNA)和RNA指导的核酸酶。这两种组分形成复合物,所述复合物能够与特定核酸序列结合并在所述核酸序列中或其周围编辑DNA,例如通过制备一条或多条单链断裂(SSB或切口)、双链断裂(DSB)和/或点突变。

[0058] 如本文关于分子的修饰所用的“替换”或“替换的”不需要方法限制,但仅表明替换实体是存在的。

[0059] 如本文所用的“受试者”可以意指人或人类、小鼠、或非人类灵长动物。

[0060] 如本文所使用的,“治疗(treat、treating和treatment)”意指治疗受试者中(例如,在人类中)的疾病,包括(a)抑制所述疾病,即抑制或防止其发育或进展;(b)缓解所述疾病,即,导致疾病状态的消退;(c)缓解一种或多种疾病症状;以及(d)治愈所述疾病。例如,“治疗”SCD或β-Thal可以指除了其他可能性之外,防止SCD或β-Thal的发育或进展,缓解SCD或β-Thal的一种或多种症状(例如,贫血、镰状细胞危象、血管阻塞性危象),或治愈SCD或β-Thal。

[0061] 如本文所用的“防止”(prevent、preventing和prevention)意指预防受试者(例如,人类)的疾病,包括(a)避免或排除疾病;(b)影响对疾病的诱因;以及(c)防止或延迟疾病的至少一种症状的发作。

[0062] 如本文所用的在氨基酸序列的语境下的“X”是指任何氨基酸(例如,二十种天然氨基酸中的任何一种),除非另外说明。

[0063] 如本文所用的“调控区域”是指包含一种或多种控制或调控基因表达的调控元件(例如,沉默子,增强子,启动子或隔离子)的DNA序列。例如,γ-珠蛋白基因调控区域包含一个或多个控制或调控γ-珠蛋白基因表达的调控元件。在某些实施例中,调控区域与控制的或调控的基因相邻。例如,γ-珠蛋白基因调控区域可以与γ-珠蛋白基因相邻或有关。在其他实施例中,调控区域可以与另一基因相邻或有关,其表达可以导致控制的或调控的基因的上调或下调。例如,γ-珠蛋白基因调控区域可以与表达γ-珠蛋白基因表达的阻遏物的基因相邻。对于HBG1,调控区域至少包含SEQ ID NO:902中的核苷酸1-2990。对于HBG2,调控区域至少包含SEQ ID NO:903中的核苷酸1-2914。

[0064] 如本文所用,“HBG靶位置”是指HBG1或HBG2调控区域中的位置(分别为“HBG1靶位置”和“HBG2靶位置”),其含有靶位点(例如,待删除或突变的靶序列),当被改变的(例如,通过引入DNA修复机制介导的(例如,NHEJ或HDR介导的)破坏或缺失)插入或缺失,通过DNA修复机制介导的(例如,HDR介导的)序列改变进行修饰))导致HBG1或HBG2基因产物(即,γ-珠蛋白)的表达增加(例如,去阻遏)。在某些实施例中,HBG靶位置在与HBG1或HBG2相邻的调控区域中的HBG1或HBG2调控元件(例如,沉默子,增强子,启动子或隔离子)中。在某些这些实施例中,HBG靶位置的改变导致阻遏物结合降低,即,去阻遏,导致HBG1或HBG2的表达增加。在其他实施例中,HBG靶位置在HBG1或HBG2以外的基因的调控元件中,其编码参与控制HBG1

或HBG2基因表达的基因产物(例如,HBG1或HBG2基因表达的阻遏物)。在某些实施例中,HBG靶位置是HBG1或HBG2调控区域的区域,其具有最大密度的结合基序,所述基序参与HBG1或HBG2表达的调控。在某些实施例中,本文提供的方法同时或顺序地靶向多个HBG靶位置。

[0065] 如本文所用的“靶序列”是指包含HBG靶位置的核酸序列。

[0066] 如本文所用的“Cas9分子”或“Cas9多肽”分别是指可以与gRNA分子相互作用,并且与gRNA分子一起定位至包含靶结构域(并且在某些实施例中,是PAM序列)的位点的分子或多肽。Cas9分子和Cas9多肽包括天然存在的Cas9分子和Cas9多肽,以及工程化的、改变的或经修饰的Cas9分子或Cas9多肽,它们与参考序列(例如最相似的天然存在的Cas9分子)相差例如至少一个氨基酸残基。

[0067] 概述

[0068] 本文提供了使用基因组编辑系统(例如,CRISPR/Cas介导的基因组编辑)增加一种或多种γ-珠蛋白基因(例如,HBG1、HBG2、或HBG1和HBG2)的表达(即,转录活性)的方法。这些方法利用基因组编辑系统(例如,CRISPR/Cas介导的基因组编辑)来改变(例如,缺失、破坏或修饰)一个或多个γ-珠蛋白基因调控区域以增加(例如,去阻遏、增强)γ-珠蛋白基因表达。在某些这些实施例中,所述方法改变与γ-珠蛋白靶向的基因有关的一种或多种调控元件(例如,沉默子,增强子,启动子或隔离子)。在其他实施例中,所述方法改变除靶向的γ-珠蛋白基因之外的基因中的一种或多种调控元件(例如,编码γ-珠蛋白基因阻遏物的基因)。在某些实施例中,基因组编辑系统(例如,CRISPR/Cas介导的基因组编辑)用于改变HBG1、HBG2、或HBG1和HBG2两者的调控元件(例如,沉默子、增强子、启动子或隔离子)。在某些实施例中,基因组编辑系统(例如,CRISPR/Cas介导的基因组编辑)导致与天然存在的HPFH变体有关的γ-珠蛋白调控元件的突变或变异,包括,例如,HBG1 13bp del c.-114至-102、4bp del c.-225至-222、c.-114C>T、c.-117G>A、c.-158C>T、c.-167C>T、c.-170G>A、c.-175T>G、c.-175T>C、c.-195C>G、c.-196C>T、c.-198T>C、c.-201C>T、c.-251T>C、或c.-499T>A、或HBG2 13bp del c.-114至-102、c.-109G>T、c.-114C>A、c.-114C>T、c.-157C>T、c.-158C>T、c.-167C>T、c.-167C>A、c.-175T>C、c.-202C>G、c.-211C>T、c.-228T>C、c.-255C>G、c.-309A>G、c.-369C>G、或c.-567T>G。

[0069] 在一些实施例中,使用本文所述的基因组编辑系统(例如,CRISPR/Cas介导的基因组编辑)的方法可利用任何修复机制来改变(例如,缺失、破坏、或修饰)一种或多种γ-珠蛋白基因调控元件的全部或部分。在某些实施例中,所述方法利用DNA修复机制介导的(例如,NHEJ或HDR介导的)插入或缺失来破坏一种或多种γ-珠蛋白基因调控元件的全部或部分。例如,所述方法可以利用DNA修复机制(例如,NHEJ或HDR)来缺失γ-珠蛋白基因负调控元件(例如,沉默子)的全部或部分,导致负调控元件的失活(例如,沉默子和阻遏物之间的结合丧失)和γ-珠蛋白基因的表达增加。在其他实施例中,所述方法利用DNA修复机制介导的(例如,NHEJ或HDR介导的)插入或缺失来破坏与编码γ-珠蛋白基因阻遏物的基因有关的一种或多种调控元件的全部或部分。例如,所述方法可利用DNA修复机制(例如,NHEJ或HDR)来缺失γ-珠蛋白阻遏物基因的正调控元件(例如,启动子)的全部或部分,导致阻遏物的表达降低、阻遏物与γ-珠蛋白基因沉默子的结合降低、γ-珠蛋白基因的表达增加。在其他实施例中,所述方法利用DNA修复机制(例如,HDR)来修饰一种或多种γ-珠蛋白基因调控元件的序列(例如,在对应于天然存在的HPFH突变的HBG1和/或HBG2调控元件或缺失全部或部分的

HBG1和/或HBG2调控元件中插入突变)。在一些实施例中,所述方法可使用一种或多种DNA修复机制的组合,例如,NHEJ和HDR。在某些实施例中,所述方法在受试者中产生HbF的持久性。本文还提供了用于这些方法的组合物(例如,gRNA、Cas9多肽和分子、模板核酸、载体)和试剂盒。

[0070] 从γ-珠蛋白基因(即,HBG1,HBG2)的表达到HBB的表达(即,珠蛋白转换)的转变与β-血红蛋白病(包括SCD和β-Thal)的症状的发作有关。因此,在某些实施例中,本文提供了用于治疗或防止包括SCD和β-Thal的β-血红蛋白病的方法、组合物、和试剂盒,其使用CRISPR/Cas介导的基因组编辑来增加一种或多种γ-珠蛋白基因的表达(例如,HBG1、HBG2、或HBG1和HBG2)。在某些这些实施例中,所述方法改变与γ-珠蛋白靶向的基因有关的一种或多种调控元件(例如,沉默子,增强子,启动子或隔离子)。在其他实施例中,所述方法改变除靶向的γ-珠蛋白基因之外的基因中的一种或多种调控元件(例如,编码γ-珠蛋白基因阻遏物的基因)。在某些实施例中,CRISPR/Cas介导的基因组编辑用于改变HBG1、HBG2、或HBG1和HBG2两者的调控元件(例如,沉默子、增强子、启动子或隔离子)。在一些实施例中,所述方法利用DNA修复机制介导的(例如,NHEJ或HDR介导的)插入或缺失来破坏一种或多种γ-珠蛋白基因调控元件的全部或部分。例如,所述方法可以利用DNA修复机制(例如,NHEJ或HDR)来缺失γ-珠蛋白基因负调控元件(例如,沉默子)的全部或部分,导致负调控元件的失活(例如,沉默子和阻遏物之间的结合丧失)和γ-珠蛋白基因的表达增加。在其他实施例中,所述方法利用DNA修复机制介导的(例如,NHEJ或HDR介导的)插入或缺失来破坏与编码γ-珠蛋白基因阻遏物的基因有关的一种或多种调控元件的全部或部分。例如,所述方法可利用DNA修复机制(例如,NHEJ或HDR)来缺失γ-珠蛋白阻遏物基因的正调控元件(例如,启动子)的全部或部分,导致阻遏物的表达降低、阻遏物与γ-珠蛋白基因沉默子的结合降低、γ-珠蛋白基因的表达增加。在其他实施例中,所述方法利用DNA修复机制(例如,HDR)来修饰一种或多种γ-珠蛋白基因调控元件的序列(例如,在对应于天然存在的HPFH突变的HBG1和/或HBG2调控元件或缺失全部或部分的HBG1和/或HBG2调控元件中插入突变)。在一些实施例中,所述方法可使用一种或多种DNA修复机制的组合(例如,NHEJ和HDR)。在某些实施例中,所述方法在受试者中产生HbF的持久性。

[0071] 在某些实施例中,使用本文提供的方法增加的一种或多种γ-珠蛋白基因(例如,HBG1、HBG2)的表达导致优先形成HbF而不是HbA和/或增加的HbF水平作为总血红蛋白的百分比。因此,本文进一步提供了使用CRISPR/Cas介导的基因组编辑来增加总HbF水平,以总血红蛋白水平的百分比增加HbF水平,或通过增加一个或多个γ-珠蛋白基因(例如,HBG1,HBG2或HBG1和HBG2)的表达来增加受试者中HbF与HbA的比率的方法。类似地,在某些实施例中,一种或多种γ-珠蛋白基因的表达增加导致HbF相对于HbS的优先形成和/或降低的HbS百分比作为总血红蛋白的百分比。因此,本文进一步提供了使用CRISPR/Cas介导的基因组编辑来降低总HbS水平、以总血红蛋白水平的百分比降低HbS水平、或通过增加一个或多个γ-珠蛋白基因(例如,HBG1、HBG2、或HBG1和HBG2)的表达来增加受试者中HbF与HbS的比率的方法。

[0072] 本文在某些实施例中提供了用于本文披露的方法的gRNA。在某些实施例中,这些gRNA包含与HBG靶位置中或附近的靶结构域互补或部分互补的靶向结构域。在某些实施例中,靶向结构域包含SEQ ID NO:251-901之一所示的核苷酸序列、由其组成、或基本上由其

组成。

[0073] 基因组研究已经识别了几种调控珠蛋白转换的基因,包括BCL11A、Kruppel样因子1(KLF1)、MYB和β珠蛋白基因座内的基因。某些这些基因的突变可能导致抑制或不完全的珠蛋白转换,也称为遗传性胎儿血红蛋白持续存在症(HPFH)。HPFH突变可以是缺失的或非缺失的(例如,点突变)。患有HPFH的受试者展现出HbF的终身表达,即,他们不经历或仅经历部分珠蛋白转换,没有贫血症状。杂合的受试者展现出20%-40%的全细胞型HbF,并且共遗传导致β血红蛋白病的缓解(Thein 2009;Akinbami 2016)。血红蛋白病和HPFH的复合杂合子,例如,作为SCD和HPFH、β-Thal和HPFH、镰状细胞性状和HPFH、或Δ-β-Thal和HPFH的复合杂合子的受试者,相对于没有HPFH突变的受试者具有更轻度的疾病和症状。共遗传HPFH突变的HbS纯合子患者,例如,通过去阻遏HBG1或HBG2诱导HbF表达的突变,不会发育SCD症状或β-Thal症状(Steinberg等人,Disorders of Hemoglobin[血红蛋白障碍],剑桥大学出版社(Cambridge Univ.Press),2009年,第570页)。HPFH临床上是良性的(Chassanidis 2009)。

[0074] 虽然HPFH的发生在全球人群中很稀有,但在血红蛋白病流行率较高的人群中更为常见,包括南欧、南美、和非洲后裔。在这些人群中,HPFH的流行率在1,000人中可以达到1-2人(Costa 2002;Ahern 1973)。从理论上讲,HPFH突变在这些人群中持续存在,因为他们改善了患有血红蛋白病的受试者的疾病。

[0075] 最常见的天然存在的HPFH突变是β珠蛋白基因座内的缺失。缺失HPFH突变的常见实例包括法国HPFH(23kb缺失)、高加索HPFH(19kb缺失)、HPFH-1(84kb缺失)、HPFH-2(84kb缺失)、和HPFH-3(50kb缺失)。在具有这些突变的受试者中,β-珠蛋白合成减少,并且γ-珠蛋白合成第二次地增加。

[0076] 其他HPFH突变位于γ-珠蛋白基因调控区域。一个这样的突变是13个核苷酸缺失(13个碱基对(bp)del c.-114至-102;CAATAGCCTTGAC del,基于HBG1/HBG2的反向互补序列)位于HBG1和HBG2基因的上游。所述缺失破坏了通常防止HBG1/HBG2表达的沉默子元件,并且所述缺失杂合的成人受试者展现出大约30%的HbF。另一个HPFH突变是4个核苷酸缺失(4个碱基对(bp)del c.-225至-222(AGCA del))。在HBG1和HBG2调控元件中发现的其他HPFH突变包括,例如,非缺失点突变(非del HPFH),例如c.-114C>T、c.-158C>T、c.-167C>T、和c.-175T>C。

[0077] 与HBG1调控元件有关的非del HPFH突变包括,例如,c.-117G>A、c.-170G>A、c.-175T>G、c.-195C>G、c.-196C>T、c.-198T>C、c.-201C>T、c.-251T>C、和c.-499T>A。

[0078] 与HBG2调控元件有关的非del HPFH突变包括,例如,c.-109G>T、c.-114C>A、c.-157C>T、c.-167C>A、c.-202C>G、c.-211C>T、c.-228T>C、c.-255C>G、和c.-567T>G。

[0079] HBG1和HBG2启动子区域的另外的多态性已在一组巴西SCD患者中识别,其校正HbF水平>5%(Barbosa 2010)。这些包括HBG2启动子中的c.-309A>G和c.-369C>G。

[0080] 可以改变以重建HPFH突变的HBG1和HBG2启动子元件,包括,例如,红系Kruppel样因子(EKLF-2)和胎儿Kruppel样因子(FKLF)转录因子结合基序(CTCCACCCA)、CP1/Coup TFII结合基序(CCAATAGC)、GATA1结合基序(CTATCT、ATATCT)、或阶段选择元件(SSE)结合基序。可以改变以重建HPFH突变的HBG1和HBG2增强子元件包括,例如,SOX结合基序,例如SOX14、SOX2、或SOX1(CCAATAGCCTTGA)。

[0081] 在本文提供的方法的某些实施例中,CRISPR/Cas介导的改变用于改变γ-珠蛋白

基因调控区域中的一个调控元件或基序,例如,HBG1或HBG2调控区域中的沉默子序列、或启动子或与编码HBG1或HBG2阻遏物的基因有关的增强子序列。在其他实施例中,CRISPR/Cas介导的改变用于改变γ-珠蛋白基因调控区域中的两个或更多个(例如,三个、四个、或五个或更多个)调控元件或基序,例如,HBG1或HBG2沉默子序列和HBG1或HBG2增强子序列;HBG1或HBG2沉默子序列和与编码HBG1或HBG2阻遏物的基因有关的启动子或增强子序列;或HBG1或HBG2沉默子序列和与编码HBG1或HBG2阻遏物的基因有关的启动子或增强子序列。将多个变体引入单个基因的调控区域或将一个变体引入两个或更多个基因的调控区域中在本文中是指"复合"。因此,复合构成了(a)在相同一个细胞或多个细胞中一个基因调控区域的多于一个位置的修饰或(b)多于一个基因调控区域中一个位置的修饰。

[0082] 在本文提供的方法的某些实施例中,CRISPR/Cas介导的一种或多种γ-珠蛋白基因调控元件的改变产生与天然存在的HPFH突变有关的相同或相似的表型。在某些实施例中,CRISPR/Cas介导的改变导致包含相应于天然存在的HPFH突变的γ-珠蛋白基因调控元件的改变。在其他实施例中,一种或多种γ-珠蛋白基因调控元件的改变导致在天然存在的HPFH突变(即,非天然存在的变体)中未观察到的改变。

[0083] 在本文提供的方法的某些实施例中,CRISPR/Cas介导的一种或多种γ-珠蛋白基因调控元件的改变产生与天然存在的HPFH变体有关的γ-珠蛋白调控元件的突变或变异,包括,例如,HBG1 13bp del c.-114至-102、4bp del c.-225至-222、c.-114C>T、c.-117G>A、c.-158C>T、c.-167C>T、c.-170G>A、c.-175T>G、c.-175T>C、c.-195C>G、c.-196C>T、c.-198T>C、c.-201C>T、c.-251T>C、或c.-499T>A、或HBG2 13bp del c.-114至-102、c.-109G>T、c.-114C>A、c.-114C>T、c.-157C>T、c.-158C>T、c.-167C>T、c.-167C>A、c.-175T>C、c.-202C>G、c.-211C>T、c.-228T>C、c.-255C>G、c.-309A>G、c.-369C>G、或c.-567T>G。

[0084] 在某些实施例中,本文提供的方法包括改变γ-珠蛋白基因调控元件中的一个或多个转录因子结合基序(例如,基因调控基序)。这些转录因子结合基序包括,例如,在HBG1和/或HBG2的启动子区域内被转录因子(TF)、TF复合物、和转录阻遏物占据的结合基序。在本文提供的方法的某些实施例中,在一个或多个γ-珠蛋白基因调控元件中引入CRISPR/Cas介导的改变来改变转录因子(例如,阻遏物)在1个、2个、3个、或多于三个基序的结合。在某些实施例中,在一个或多个γ-珠蛋白基因调控元件中引入CRISPR/Cas介导的改变导致RNA聚合酶II在γ-珠蛋白基因启动子区域附近或在γ-珠蛋白基因启动子区域处的转录起始增加,例如,通过增加转录因子结合增强子区域,例如通过减少阻遏物在沉默子区域的结合。

[0085] 在某些实施例中,本文提供的方法利用DNA修复机制介导的(例如,NHEJ-或HDR-介导的)缺失来缺失HBG1、HBG2、或HBG1和HBG2的一个或两个等位基因中的全部或部分核苷酸-114至-102,导致与天然存在的13bp del c.-114至-102突变有关的相同或相似的HPFH表型。在其他实施例中,利用DNA修复机制介导的(例如,NHEJ-或HDR-介导的)缺失来缺失HBG1的一个或两个等位基因中的全部或部分核苷酸-225至-222,导致与天然存在的HBG1 4bp del-225至-222突变有关的相同或相似的HPFH表型。在其他实施例中,利用DNA修复机制介导的(例如,NHEJ-或HDR-介导的)缺失来缺失HBG2的一个或两个等位基因的全部或部分核苷酸-225至-222。

[0086] 在某些实施例中,本文提供的方法利用DNA修复机制介导的(例如,NHEJ-或HDR-介

导的)缺失来缺失HBG1的一个或两个等位基因和HBG2的一个或两个等位基因的全部或部分核苷酸-114至-102。

[0087] 在某些实施例中,本文提供的方法利用DNA修复机制介导的(例如,NHEJ或HDR-介导的)缺失来缺失HBG1的一个或两个等位基因的全部或部分核苷酸-225至-222和HBG2的一个或两个等位基因的全部或部分核苷酸-114至-102。在其他实施例中,利用DNA修复机制介导的(例如,NHEJ-或HDR-介导的)缺失来缺失HBG1的一个或两个等位基因的全部或部分核苷酸-225至-222和HBG1的一个或两个等位基因的全部或部分核苷酸-114至-102。

[0088] 在其中DNA修复机制介导的(例如,NHEJ-或HDR-介导的)缺失用于从HBG1、HBG2、或HBG1和HBG2调控元件中缺失一个或多个核苷酸的那些实施例中,缺失可以与那些在天然存在的HPFH突变中观察到的一致,即,缺失可以由HBG1或HBG2的核苷酸-114至-102或HBG1的核苷酸-225至-222组成。在其他实施例中,DNA修复机制介导的(例如,NHEJ-或HDR-介导的)缺失导致仅去除这些核苷酸的一部分,例如,缺失落入HBG1或HBG2的-114至-102内的12个或更少的核苷酸,或HBG1的-225至-222内的三个或更少的核苷酸。在某些实施例中,除了在天然存在的缺失边界内的全部或部分核苷酸之外,还可以在天然存在的HPFH突变缺失边界的任一侧(即,在-114至-102或-225至-222之外)敲除一个或多个核苷酸。

[0089] 在某些实施例中,本文提供的方法利用DNA修复机制介导的(例如,NHEJ-或HDR-介导的)插入将一个或多个核苷酸插入跨越HBG1调控区域的核苷酸-114至-102的区域、HBG2调控区域、或HBG1和HBG2调控区域、或跨越HBG1调控区域的核苷酸-225至-222的区域,以破坏阻遏物结合位点。

[0090] 在某些实施例中,本文提供的方法利用DNA修复机制(例如,HDR)产生相应于与HPFH有关的天然存在的突变的单核苷酸改变(即,非缺失突变体)。例如,在某些实施例中,所述方法利用DNA修复机制(例如,HDR)在HBG1调控区域中产生相应于与HPFH有关的天然存在的突变的单核苷酸改变,包括,例如,c.-114C>T、c.-117G>A、c.-158C>T、c.-167C>T、c.-170G>A、c.-175T>G、c.-175T>C、c.-195C>G、c.-196C>T、c.-198T>C、c.-201C>T、c.-251T>C、或c.-499T>A。例如,在其他实施例中,利用DNA修复机制(例如,HDR)在HBG2调控区域中产生相应于与HPFH有关的天然存在的突变的单核苷酸改变,包括,例如,c.-109G>T、c.-114C>A、c.-114C>T、c.-157C>T、c.-158C>T、c.-167C>T、c.-167C>A、c.-175T>C、c.-202C>G、c.-211C>T、c.-228T>C、c.-255C>G、c.-309A>G、c.-369C>G、c.-567。

[0091] 在某些实施例中,利用DNA修复机制(例如,HDR)在HBG1调控区域中产生相应于在HBG2调控区域但不是HBG1调控区域中发现的天然存在的HPFH突变的单核苷酸改变。这种改变包括,例如,c.-109G>T、c.-114C>A、c.-157C>T、c.-167C>A、c.-202C>G、c.-211C>T、c.-228T>C、c.-255C>G、c.-309A>G、c.-369C>G、或c.-567T>G。

[0092] 同样地,在某些实施例中,利用DNA修复机制(例如,HDR)在HBG2调控区域中产生相应于在HBG1调控区域但不是HBG2调控区域中发现的天然存在的HPFH突变的单核苷酸改变。这种改变包括,例如,c.-117G>A、c.-170G>A、c.-175T>G、c.-195C>G、c.-196C>T、c.-198T>C、c.-201C>T、c.-251T>C、或c.-499T>A。

[0093] 在某些实施例中,本文提供的方法包括通过DNA修复机制(例如,HDR)将非缺失HPFH变体c.-114C>T插入HBG1和/或HBG2调控区域。

[0094] 在某些实施例中,本文提供的方法包括通过DNA修复机制(例如,HDR)将非缺失

HPFH变体c.-158C>T(即,rs7482144或XmnI-HBG2变体)插入HBG1和/或HBG2调控区域。

[0095] 在某些实施例中,本文提供的方法包括通过DNA修复机制(例如,HDR)将非缺失HPFH变体c.-167C>T插入HBG1和/或HBG2调控区域。

[0096] 在某些实施例中,本文提供的方法包括通过DNA修复机制(例如,HDR)将非缺失HPFH变体c.-175T>C(即,在保守的八核苷酸[ATGCAAAT]序列中的c.-175位置的T→C取代)插入HBG1调控区域。这种与40% HbF有关的变体已显示可以消除普遍存在的八核苷酸结合核蛋白以结合HBG启动子片段的能力,同时增加3-5倍两种红系特异性蛋白质结合相同片段的能力(Mantovani 1988)。

[0097] 在某些实施例中,本文提供的方法包括通过DNA修复机制(例如,HDR)将非缺失HPFH变体c.-175T>C插入HBG2调控区域。所述变体与20%-30%HbF表达有关。

[0098] 在某些实施例中,本文提供的方法包括通过DNA修复机制(例如,HDR)将非缺失HPFH变体c.-117G>A插入HBG1调控区域。这种变体,是指“希腊类型”,是最常见的非缺失HPFH突变体,并在远端CCAAT盒上游映射两个核苷酸(Waber1986)。HBG1 c.-117G>A极大地降低了红系特异性因子的结合,但不是普遍存在的蛋白质与CCAAT盒区域片段的结合,并且与10%-20% HbF有关(Mantovani 1988)。所述突变被认为干扰核因子E(NF-E)的结合,其可能在成人红系细胞中阻遏γ-珠蛋白转录中起作用(Superti-Furga 1988)。在其他实施例中,本文提供的方法包括将非缺失HPFH变体c.-117G>A插入HBG2调控区域,产生非天然存在的HPFH变体。

[0099] 在某些实施例中,本文提供的方法包括通过DNA修复机制(例如,HDR)将非缺失HPFH变体c.-170G>A插入HBG1调控区域。

[0100] 在某些实施例中,本文提供的方法包括通过DNA修复机制(例如,HDR)将非缺失HPFH变体c.-175T>G插入HBG1调控区域。

[0101] 在某些实施例中,本文提供的方法包括将非缺失HPFH变体c.-195C>G插入HBG1调控区域。

[0102] 在某些实施例中,本文提供的方法包括通过DNA修复机制(例如,HDR)将非缺失HPFH变体c.-196C>T插入HBG1调控区域。所述变体与10%-20% HbF有关。

[0103] 在某些实施例中,本文提供的方法包括通过DNA修复机制(例如,HDR)将非缺失HPFH变体c.-198T>C插入HBG1调控区域。所述变体与18%-21% HbF有关。

[0104] 在某些实施例中,本文提供的方法包括将非缺失HPFH变体c.-201C>T插入HBG1调控区域。

[0105] 在某些实施例中,本文提供的方法包括通过DNA修复机制(例如,HDR)将非缺失HPFH变体c.-251T>C插入HBG1调控区域。

[0106] 在某些实施例中,本文提供的方法包括通过DNA修复机制(例如,HDR)将非缺失HPFH变体c.-499T>A插入HBG1调控区域。

[0107] 在某些实施例中,本文提供的方法包括通过DNA修复机制(例如,HDR)将非缺失HPFH变体c.-109G>T(“希腊突变”)插入HBG2调控区域。所述突变位于启动子区域的HBG2 CCAAT盒的3′端(Chassanidis 2009)。

[0108] 在某些实施例中,本文提供的方法包括通过DNA修复机制(例如,HDR)将非缺失HPFH变体c.-114C>A插入HBG2调控区域。

[0109] 在某些实施例中,本文提供的方法包括通过DNA修复机制(例如,HDR)将非缺失HPFH变体c.-157C>T插入HBG2调控区域。

[0110] 在某些实施例中,本文提供的方法包括通过DNA修复机制(例如,HDR)将非缺失HPFH变体c.-167C>A插入HBG2调控区域。

[0111] 在某些实施例中,本文提供的方法包括通过DNA修复机制(例如,HDR)将非缺失HPFH变体c.-202C>G插入HBG2调控区域。所述变体与15%-25% HbF表达有关。

[0112] 在某些实施例中,本文提供的方法包括通过DNA修复机制(例如,HDR)将非缺失HPFH变体c.-211C>T插入HBG2调控区域。

[0113] 在某些实施例中,本文提供的方法包括通过DNA修复机制(例如,HDR)将非缺失HPFH变体c.-228T>C插入HBG2调控区域。

[0114] 在某些实施例中,本文提供的方法包括通过DNA修复机制(例如,HDR)将非缺失HPFH变体c.-255C>G插入HBG2调控区域。

[0115] 在某些实施例中,本文提供的方法包括通过DNA修复机制(例如,HDR)将非缺失HPFH变体c.-309A>G插入HBG2调控区域。

[0116] 在某些实施例中,本文提供的方法包括通过DNA修复机制(例如,HDR)将非缺失HPFH变体c.-369C>G插入HBG2调控区域。

[0117] 在某些实施例中,本文提供的方法包括通过DNA修复机制(例如,HDR)将非缺失HPFH变体c.-567T>G插入HBG2调控区域。

[0118] 在某些实施例中,本文提供的方法包括位于相对于HBG1和/或HBG2的位置c.-56和/或γ-珠蛋白基因调控区域中的另一位置的BCL11a核心结合基序(即,GGCCGG)的缺失、破坏或突变。

[0119] 在某些实施例中,本文提供的方法包括改变GATA(例如,GATA1)基序中的一个或多个核苷酸。在某些这些实施例中,使用DNA修复机制(例如,HDR)将T>C突变插入序列AAATATCTGT内的HBG1 GATA结合基序中,导致序列AAACATCTGT的改变。这种天然存在的T>C HPFH突变与40% HbF有关。

[0120] 在某些实施例中,本文提供的方法利用一种或多种DNA修复机制(例如,NHEJ和HDR)途径。例如,在某些实施例中,所述方法利用NHEJ介导的缺失,例如,将13bp del c.-114至-102引入HBG1和/或HBG2的一个或两个等位基因中,和/或将4bp del c.-225至-222引入HBG1的一个或两个等位基因中,结合HDR介导的单核苷酸改变,例如引入一个或多个c.-109G>T、c.-114C>A、c.-114C>T、c.-117G>A、c.-157C>T、c.-158C>T、c.-167C>T、c.-167C>A、c.-170G>A、c.-175T>C、c.-175T>G、c.-195C>G、c.-196C>T、c.-198T>C、c.-201C>T、c.-202C>G、c.-211C>T、c.-228T>C、c.-251T>C、c.-255C>G、c.-309A>G、c.-369C>G、c.-499T>A、或c.-567T>G进入HBG1和/或HBG2的一个或两个等位基因。

[0121] 在某些实施例中,所述方法利用HDR介导的缺失,例如,将13bp del c.-114至-102引入HBG1和/或HBG2的一个或两个等位基因中,和/或将4bp del c.-225至-222引入HBG1的一个或两个等位基因中,结合HDR介导的单核苷酸改变,例如引入一个或多个c.-109G>T、c.-114C>A、c.-114C>T、c.-117G>A、c.-157C>T、c.-158C>T、c.-167C>T、c.-167C>A、c.-170G>A、c.-175T>C、c.-175T>G、c.-195C>G、c.-196C>T、c.-198T>C、c.-201C>T、c.-202C>G、c.-211C>T、c.-228T>C、c.-251T>C、c.-255C>G、c.-309A>G、c.-369C>G、c.-499T>A、或c.-567T>

G进入HBG1和/或HBG2的一个或两个等位基因。

[0122] 虽然不希望受理论束缚,但将4bp del c.-225至-222引入HBG1基因调控区域可逆转70%γ^A-珠蛋白(HBG1基因的γ-珠蛋白产物)与30%γ^G-珠蛋白(HBG2基因的γ-珠蛋白产物)的正常比例,使γ-珠蛋白产生大约30%的γ^A-珠蛋白和70%的γ^G-珠蛋白。虽然不希望受理论束缚,但γ^G-珠蛋白和γ^A-珠蛋白比率的逆转导致受试者中γ^G-珠蛋白的产生增加。虽然不希望受理论束缚,但将4bp del c.-225至-222引入HBG1基因调控区域并将13bp del c.-114至-102伴随引入HBG2基因调控区域导致HBG2转录活性的增加、γ^G-珠蛋白产生的增加、和受试者中HbF的增加。虽然不希望受理论束缚,但将(a)4bp del c.-225至-222伴随引入HBG1基因调控区域,例如,通过NHEJ-或HDR介导的缺失,和(b)非缺失HPFH变体,例如,通过HDR,例如,c.-109G>T、c.-114C>T、c.-114C>A、c.-157C>T、c.-158C>T、c.-167C>T、c.-167C>A、c.-175T>C、c.-202C>G、c.-211C>T、c.-228T>C、c.-255C>G、c.-309A>G、c.-369C>G、c.-567T>G,引入HBG2基因调控区域导致受试者中HBG2的转录活性增加、γ^G-珠蛋白的产生增加和HbF增加。

[0123] 虽然不希望受理论束缚,但是相对于γ^A-珠蛋白(HBG1基因的γ-珠蛋白产物)的产生,将4bp del c.-225至-222引入HBG2基因调控区域可以减少γ^G-珠蛋白(HBG2基因的γ-珠蛋白产物)的产生,产生的γ^A-珠蛋白比γ^G-珠蛋白更多。虽然不希望受理论束缚,但将4bp del c.-225至-222引入HBG2基因调控区域并将13bp del c.-114至-102伴随引入HBG1基因调控区域可导致HBG1转录活性的增加、γ^A-珠蛋白产生的增加和受试者中HbF的增加。虽然不希望受理论束缚,但将(a)4bp del c.-225至-222伴随引入HBG2基因调控区域,例如,通过NHEJ-或HDR介导的缺失,和(b)非缺失HPFH变体,例如,通过HDR,例如,c.-114C>T、c.-117G>A、c.-158C>T、c.-167C>T、c.-170G>A、c.-175T>G、c.-175T>C、c.-195C>G、c.-196C>T、c.-198T>C、c.-201C>T、c.-251T>C、或c.-499T>A,引入HBG1基因调控区域可导致受试者中HBG1的转录活性增加、γ^A-珠蛋白的产生增加和HbF增加。

[0124] 虽然不希望受理论束缚,但将(a)13bp del c.-114至-102伴随引入HBG1基因调控区域,例如,通过NHEJ-或HDR介导的缺失,和(b)非缺失HPFH变体,例如,通过HDR,例如,c.-109G>T、c.-114C>T、c.-114C>A、c.-157C>T、c.-158C>T、c.-167C>A、c.-167C>T、c.-175T>C、c.-202C>G、c.-211C>T、c.-228T>C、c.-255C>G、c.-309A>G、c.-369C>G、或c.-567T>G,引入HBG2基因调控区域导致受试者中HBG2的转录活性增加、γ^G-珠蛋白的产生增加和HbF增加。

[0125] 虽然不希望受理论束缚,但将(a)13bp del c.-114至-102伴随引入HBG2基因调控区域,例如,通过NHEJ-或HDR介导的缺失,和(b)非缺失HPFH变体,例如,通过HDR,例如,c.-114C>T、c.-117G>A、c.-158C>T、c.-167C>T、c.-170G>A、c.-175T>G、c.-175T>C、c.-195C>G、c.-196C>T、c.-198T>C、c.-201C>T、c.-251T>C、或c.-499T>A,引入HBG1基因调控区域导致受试者中HBG1的转录活性增加、γ^A-珠蛋白的产生增加和HbF增加。

[0126] 伴随的(a)通过siRNA敲低BCL11A和(b)通过siRNA敲低SOX6导致HBG1和HBG2的表达增加(Xu 2010)。在某些实施例中,本文提供的方法包括使用HBG1的DNA修复机制(例如,HDR、NHEJ、或NHEJ和HDR)修饰和HBG2启动子区域和BCL11A的红系特异性增强子,单独或平行破坏BCL11A、SOX6、或BCL11A和SOX6对HBG1和HBG2表达的作用。在某些实施例中,本文提供的方法包括通过NHEJ和HDR破坏其内含子红系特异性增强子的功能来降低BCL11A表达,并同时诱导HPFH突变以对HbF产生的协同效应。

[0127] 本文描述的实施例可用于所有类别的脊椎动物,包括但不限于灵长类动物、小鼠、大鼠、兔、猪、狗和猫。

[0128] 时间和受试者选择

[0129] 使用本文披露的方法起始治疗可以在疾病发作之前发生,例如在基于遗传测试、家族史或者其他因素,例如被认为有发育β-血红蛋白病(例如,SCD,β-Thal)的风险的但尚未显示出疾病的任何表现或症状的受试者中。在某些这些实施例中,可以在天然存在的珠蛋白转换之前,即,在从主要是HbF转变为主要是HbA之前起始治疗。在其他实施例中,可以在发生天然存在的珠蛋白转换后起始治疗。

[0130] 在某些实施例中,在疾病发作后起始治疗,例如SCD或β-Thal或与其有关的一种或多种症状发作后的1、2、3、4、5、6、7、8、9、10、12、16、24、36、或48个月或更多个月。在某些这些实施例中,治疗在疾病进展的早期阶段起始,例如,当受试者仅显示轻微症状或仅显示一部分症状时。示例性症状包括但不限于贫血、腹泻、发热、未能茁壮成长、镰状细胞危象、血管阻塞性危象、再生障碍性危象、急性胸部综合征性贫血、血管阻塞、肝肿大、血栓症、肺栓塞、中风、腿部溃疡、心肌病、贲门心律失常、脾肿大、骨质生长延迟和/或青春期、以及髓外红细胞生成的证据。在其他实施例中,在SCD或β-Thal发作后治疗在疾病发作后或在疾病进展的更晚期阶段开始,例如,1、2、3、4、5、6、7、8、9、10、12、16、24、36或48个月或更长时间。虽然不希望受理论束缚,但据信如果受试者很好地存在于疾病过程中,这种治疗将是有效的。

[0131] 在某些实施例中,本文提供的方法防止或减缓与所治疗疾病有关的一种或多种症状的发育。在某些实施例中,与未接受疗法的受试者相比,本文提供的方法导致疾病进展的防止或延迟。在某些实施例中,本文提供的方法导致疾病完全治愈。

[0132] 在某些实施例中,本文提供的方法是基于一次性进行的。在其他实施例中,本文提供的方法利用多剂量疗法。

[0133] 在某些实施例中,使用本文提供的方法治疗的受试者是输血依赖性的。

[0134] 在某些实施例中,本文提供的方法包括在体内细胞中使用CRISPR/Cas介导的基因组编辑改变一种或多种γ-珠蛋白基因(例如,HBG1、HBG2)的表达。在其他实施例中,本文提供的方法包括在离体细胞中使用CRISPR/Cas介导的基因组编辑改变一种或多种γ-珠蛋白基因的表达。在某些这些实施例中,细胞最初来自受试者。在某些实施例中,经历改变的细胞是成人红系细胞。在其他实施例中,细胞是造血干细胞(HSC)。

[0135] 在某些实施例中,本文提供的方法包括向细胞递送一种或多种gRNA分子和一种或多种Cas9多肽或编码Cas9多肽的核酸序列。在某些实施例中,所述方法进一步包括递送一种或多种核酸,例如,HDR供体模板。

[0136] 在某些实施例中,使用一种或多种AAV载体、慢病毒载体、纳米颗粒或其组合递送这些组分中的一种或多种(即,一种或多种gRNA分子、一种或多种Cas9多肽或编码Cas9多肽的核酸序列、和一种或多种核酸,例如,HDR供体模板)。

[0137] 在某些实施例中,本文提供的方法在具有HBB基因中的一个或多个突变的受试者上进行,包括一种或多种与β-血红蛋白病有关的突变,例如SCD或β-Thal。这种突变的实例包括,但不限于,c.17A>T、c.-136C>G、c.92+1G>A、c.92+6T>C、c.93-21G>A、c.118C>T、c.316-106C>G、c.25_26delAA、c.27_28insG、c.92+5G>C、c.118C>T、c.135delC、c.315+1G>A、c.-78A>G、c.52A>T、c.59A>G、c.92+5G>C、c.124_127delTTCT、c.316-197C>T、c.-78A>G、

c.52A>T、c.124_127delTTCT、c.316-197C>T、c.-138C>T、c.-79A>G、c.92+5G>C、c.75T>A、c.316-2A>G、和c.316-2A>C。

[0138] NHEJ介导的γ-珠蛋白基因调控元件中indel的引入

[0139] 在某些实施例中,本文提供的方法利用NHEJ介导的插入或缺失来破坏γ-珠蛋白基因调控元件的全部或部分,以增加γ-珠蛋白基因(例如,HBG1、HBG2、或HBG1和HBG2)的表达。

[0140] 在某些实施例中,本文提供的利用NHEJ的方法包括经由NHEJ缺失或破坏HBG1或HBG2沉默子元件的全部或部分,导致沉默子的失活和随后HBG1和/或HBG2表达的增加。在某些实施例中,NHEJ介导的缺失导致在HBG1的一个或两个等位基因中去除c.-114至-102或-225至-222的全部或部分,和/或去除在HBG2的一个或两个等位基因中c.-114至-102的全部或部分。在某些这些实施例中,这些区域的一个或多个核苷酸5'或3'也被缺失。

[0141] 在某些实施例中,本文提供的利用NHEJ的方法包括在γ-珠蛋白基因调控区域内引入一个或多个断裂(例如,单链断裂或双链断裂),并且在这些实施例的某些中,所述一个或多个断裂位于足够接近HBG靶位置,可以合理地预期断裂诱导的indel跨越全部或部分HBG靶位置。

[0142] 在某些实施例中,第一gRNA分子的靶向结构域被配置为提供切割事件,例如,双链断裂或单链断裂,足够接近HBG靶位置以允许NHEJ介导的插入或在HBG靶位置的缺失。在某些实施例中,所述gRNA靶向结构域被配置成使得切割事件(例如,双链或单链断裂)被定位在HBG靶位置的1、2、3、4、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、150、200、250、300、350、400、450或500个核苷酸内。断裂,例如双链或单链断裂可以位于HBG靶位置的上游或下游。

[0143] 在某些实施例中,包含第二靶向结构域的第二gRNA分子被配置为提供切割事件,例如双链断裂或单链断裂,足够接近HBG靶位置以允许在HBG靶位置的NHEJ介导的插入或缺失,单独或与由所述第一gRNA分子位置的断裂组合。在某些实施例中,所述第一和第二gRNA分子的靶向结构域被配置成使得切割事件(例如,双链或单链断裂)对于所述gRNA分子中的每者而言独立地被定位在靶位置的1、2、3、4、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、150、200、250、300、350、400、450或500个核苷酸内。在某些实施例中,断裂(例如,双链或单链断裂)位于HBG靶位置的核苷酸的任一侧。在其他实施例中,断裂(例如,双链或单链断裂)均位于HBG靶位置的核苷酸的一侧,例如,上游或下游。

[0144] 在某些实施例中,单链断裂伴随有由第二gRNA分子位置的另外的单链断裂,如下文所讨论的。例如,所述gRNA靶向结构域可被配置成使得切割事件(例如,两个单链断裂)被定位在HBG靶位置的1、2、3、4、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、150、200、250、300、350、400、450或500个核苷酸内。在某些实施例中,所述第一和第二gRNA分子被配置成使得在指导Cas9切口酶时,单链断裂将伴随有由第二gRNA定位的彼此足够接近的另外的单链断裂,以引起HBG靶位置的改变。在某些实施例中,所述第一和第二gRNA分子被配置成使得例如,当Cas9是切口酶时,由所述第二gRNA定位的单链断裂在由所述第一gRNA分子定位的断裂的10、20、30、40或50个核苷酸内。在某些实施例中,所述两个gRNA分子被配置成将切口定位在相同位置、或彼此相距在几个核苷酸之内、在不同链上,例如,基本上模拟了双链断裂。

[0145] 在某些实施例中,双链断裂可以伴随有由第二gRNA分子定位的另外的双链断裂,如下文所讨论的。例如,第一gRNA分子的靶向结构域被配置成使得双链断裂被定位在HBG靶位置的上游,例如,在靶位置的1、2、3、4、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、150、200、250、300、350、400、450或500个核苷酸内;并且第二gRNA分子的靶向结构域被配置成使得双链断裂被定位在HBG靶位置的下游,例如在靶位置的1、2、3、4、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、150、200、250、300、350、400、450或500个核苷酸内。

[0146] 在某些实施例中,双链断裂可以伴随有由第二gRNA分子和第三gRNA分子定位的两个另外的单链断裂。例如,第一gRNA分子的靶向结构域被配置成使得双链断裂被定位在HBG靶位置的上游,例如,在靶位置的1、2、3、4、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、150、200、250、300、350、400、450或500个核苷酸内;并且第二和第三gRNA分子的靶向结构域被配置成使得两个单链断裂被定位在HBG靶位置的下游,例如在靶位置的1、2、3、4、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、150、200、250、300、350、400、450或500个核苷酸内。在某些实施例中,所述第一、第二和第三gRNA分子的靶向结构域被配置成使得切割事件(例如,双链或单链断裂)对于所述gRNA分子中的每者而言独立地被定位。

[0147] 在某些实施例中,第一和第二单链断裂可以伴随有由第三和第四gRNA分子定位的两个另外的单链断裂。例如,第一gRNA分子和第二gRNA分子的靶向结构域被配置成使得两个单链断裂被定位在HBG靶位置的上游,例如在靶位置的1、2、3、4、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、150、200、250、300、350、400、450或500个核苷酸内;并且第三和第四gRNA分子的靶向结构域被配置成使得两个单链断裂被定位在HBG靶位置的下游,例如在靶位置的1、2、3、4、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、150、200、250、300、350、400、450或500个核苷酸内。

[0148] 在某些实施例中,本文提供的方法包括引入NHEJ介导的包括HBG靶位置的基因组序列的缺失。在某些实施例中,所述方法包括引入两个双链断裂,一个5'至另一个3'至(即侧翼)HBG靶位置。两种gRNA,例如单分子(或嵌合)或模块化gRNA分子,被配置成将两个双链断裂定位在HBG靶位置的相对侧。在某些实施例中,第一双链断裂位于突变的上游,第二双链断裂位于突变的下游。在某些实施例中,定位两个双链断裂以去除HBG1 c.-114至-102、HBG1 4bp del-225至-222的全部或部分。在一个实施例中,定位断裂(即,两个双链断裂)以避免不想要的靶向染色体元件,例如重复元件,例如Alu重复,或内源剪接位点。

[0149] 在其他实施例中,所述方法包括引入两组断裂,一条双链断裂和一对单链断裂。两组位于HBG靶位置的侧翼,即,一组是5'到另一组是HBG靶位置的3'。两种gRNA,例如,单分子(或嵌合)或模块化gRNA分子,被配置成在HBG靶位置的相对侧定位两组断裂(双链断裂或一对单链断裂)。在一个实施例中,定位断裂(即,两组断裂(双链断裂或一对单链断裂)以避免不想要的靶向染色体元件,例如重复元件,例如Alu重复,或内源剪接位点。

[0150] 在其他实施例中,所述方法包括引入两对单链断裂,一个5′至另一个3'至(即,侧翼)HBG靶位置。两种gRNA,例如单分子(或嵌合)或模块化gRNA分子,被配置成将两组断裂定位在HBG靶位置的相对侧。在某些实施例中,定位断裂(即,两对单链断裂)以避免不想要的靶向染色体元件,例如重复元件,例如Alu重复,或内源剪接位点。

[0151] HDR介导的γ-珠蛋白基因调控元件中序列改变的引入

[0152] 在某些实施例中,本文提供的方法利用HDR以修饰在γ-珠蛋白基因调控元件中的

一个或多个核苷酸,以增加γ-珠蛋白基因(例如,HBG1、HBG2、或HBG1和HBG2)的表达。在某些这些实施例中,利用HDR掺入相应于与HPFH有关的天然存在的突变的一个或多个核苷酸修饰。例如,在某些实施例中,HDR用于将一种或多种下列单核苷酸改变掺入HBG1调控区域:c.-114C>T、c.-117G>A、c.-158C>T、c.-167C>T、c.-170G>A、c.-175T>C、c.-175T>G、c.-195C>G、c.-196C>T、c.-198T>C、c.-201C>T、c.-251T>C、或c.-499T>A。在其他实施例中,HDR用于将一种或多种下列单核苷酸改变掺入HBG2调控区域:c.-109G>T、c.-114C>A、c.-114C>T、c.-157C>T、c.-158C>T、c.-167C>T、c.-167C>A、c.-175T>C、c.-202C>G、c.-211C>T、c.-228T>C、c.-255C>G、c.-309A>G、c.-369C>G、c.-567T>G。

[0153] 在某些实施例中,本文提供的方法利用HDR-介导的改变(例如,插入或缺失)以破坏γ-珠蛋白基因调控元件的全部或部分,以增加γ-珠蛋白基因(例如,HBG1、HBG2、或HBG1和HBG2)的表达。

[0154] 在某些实施例中,本文提供的利用HDR的方法包括经由HDR缺失或破坏HBG1或HBG2沉默子元件的全部或部分,导致沉默子的失活和随后HBG1和/或HBG2表达的增加。在某些实施例中,HDR-介导的缺失导致在HBG1的一个或两个等位基因中去除c.-114至-102或-225至-222的全部或部分,和/或去除在HBG2的一个或两个等位基因中c.-114至-102的全部或部分。在某些这些实施例中,这些区域的一个或多个核苷酸5'或3'也被缺失。

[0155] 在某些实施例中,本文提供的利用HDR的方法包括在γ-珠蛋白基因调控区域内引入一个或多个断裂(例如,单链断裂或双链断裂),并且在这些实施例的某些中,所述一个或多个断裂位于足够接近HBG靶位置,可以合理地预期断裂诱导的改变跨越全部或部分HBG靶位置。

[0156] 在某些实施例中,HDR介导的改变可包括使用模板核酸。

[0157] 在某些实施例中,将HDR介导的遗传改变掺入一个γ-珠蛋白基因等位基因(例如,HBG1和/或HBG2的一个等位基因)中。在另一个实施例中,遗传改变掺入两个等位基因(例如,HBG1和/或HBG2的两个等位基因)。在任一情况下,经处理的受试者展现出增加的γ-珠蛋白基因表达(例如,HBG1、HBG2、或HBG1和HBG2表达)。

[0158] 在某些实施例中,本文提供的利用HDR的方法包括引入足够接近(例如,5'或3'至)HBG靶位置的一个或多个断裂(例如,单链断裂或双链断裂)以允许与目标位置的HDR有关的更改。

[0159] 在某些实施例中,第一gRNA分子的靶向结构域被配置为提供切割事件,例如,双链断裂或单链断裂,足够接近HBG靶位置以允许与目标位置的HDR有关的更改。在某些实施例中,所述gRNA靶向结构域被配置成使得切割事件(例如,双链或单链断裂)被定位在HBG靶位置的1、2、3、4、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、150、200、250、300、350、400、450或500个核苷酸内。断裂,例如双链或单链断裂可以位于HBG靶位置的上游或下游。

[0160] 在某些实施例中,第二、第三、和/或第四gRNA分子被配置为提供切割事件,例如,双链断裂或单链断裂,足够接近(例如,5′或3'至)HBG靶位置以允许与目标位置的HDR有关的更改。在某些实施例中,所述gRNA靶向结构域被配置成使得切割事件(例如,双链或单链断裂)被定位在HBG靶位置的1、2、3、4、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、150、200、250、300、350、400、450或500个核苷酸内。断裂,例如双链或单链断裂可以位于靶位置的上游或下游。

[0161] 在某些实施例中,单链断裂伴随有由第二、第三和/或第四gRNA分子定位的另外的单链断裂。例如,所述gRNA靶向结构域可被配置成使得切割事件(例如,两个单链断裂)被定位在HBG靶位置的1、2、3、4、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、150、200、250、300、350、400、450或500个核苷酸内。在某些实施例中,所述第一和第二gRNA分子被配置成使得在指导Cas9切口酶时,单链断裂将伴随有由第二gRNA定位的足够接近第一条链断裂的另外的单链断裂,以导致HBG靶位置的改变。在某些实施例中,所述第一和第二gRNA分子被配置成使得例如,当Cas9是切口酶时,由所述第二gRNA定位的单链断裂在由所述第一gRNA分子定位的断裂的10、20、30、40、50、100、200、300、400、500、600、700、800、900或1000个核苷酸内。在某些实施例中,所述两个gRNA分子被配置成将切口定位在相同位置、或彼此相距在几个核苷酸之内、在不同链上,例如,基本上模拟了双链断裂。

[0162] 在某些实施例中,双链断裂可伴随有由第二、第三和/或第四gRNA分子定位的另外的双链断裂。例如,第一gRNA分子的靶向结构域可被配置成使得双链断裂被定位在HBG靶位置的上游,例如,在靶位置的1、2、3、4、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、150、200、250、300、350、400、450或500个核苷酸内;并且第二gRNA分子的靶向结构域可被配置成使得双链断裂被定位在HBG靶位置的下游,例如在靶位置的1、2、3、4、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、150、200、250、300、350、400、450或500个核苷酸内。

[0163] 在某些实施例中,双链断裂可以伴随有由第二和第三gRNA分子定位的两个另外的单链断裂。例如,第一gRNA分子的靶向结构域可被配置成使得双链断裂被定位在HBG靶位置的上游,例如,在靶位置的1、2、3、4、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、150、200、250、300、350、400、450或500个核苷酸内;并且第二和第三gRNA分子的靶向结构域可被配置成使得两个单链断裂被定位在靶位置的下游,例如,在靶位置的1、2、3、4、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、150、200、250、300、350、400、450或500个核苷酸内。在某些实施例中,所述第一、第二和第三gRNA分子的靶向结构域被配置成使得切割事件(例如,双链或单链断裂)对于所述gRNA分子中的每者而言独立地被定位。

[0164] 在某些实施例中,第一和第二单链断裂可以伴随有由第三gRNA分子和第四gRNA分子定位的两个另外的单链断裂。例如,第一gRNA分子和第二gRNA分子的靶向结构域可被配置成使得两个单链断裂被定位在HBG靶位置的上游,例如在靶位置的1、2、3、4、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、150、200、250、300、350、400、450或500个核苷酸内;并且第三和第四gRNA分子的靶向结构域可被配置成使得两个单链断裂被定位在HBG靶位置的下游,例如,在靶位置的1、2、3、4、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、150、200、250、300、350、400、450或500个核苷酸内。

[0165] <u>指导RNA(gRNA)分子</u>

[0166] 如该术语在本文使用的,gRNA分子是指促进gRNA分子/Cas9分子复合物向靶核酸特异性靶向或归巢的核酸。gRNA分子可以是单分子的(具有单RNA分子)(例如,嵌合的)、或模块化的(包含多于一种并且典型地两种分开的RNA分子)。本文提供的gRNA分子包含靶向结构域,所述靶向结构域包含与靶结构域完全或部分互补的核酸序列,由其组成或基本上由其组成。在某些实施例中,所述gRNA分子进一步包含一个或多个另外的结构域,包括例如第一互补结构域、连接结构域、第二互补结构域、近端结构域、尾部结构域和5'延伸结构域。下面详细讨论了这些结构域中的每一者。在某些实施例中,gRNA分子中的一个或多个结构

域包含与例如来自化脓链球菌、金黄色葡萄球菌或嗜热链球菌的天然存在的序列相同或与其共享序列同源性的核苷酸序列。

[0167] 图1A-1I中提供了若干示例性gRNA结构。关于gRNA的三维形式、或活化形式的链内或链间相互作用,高度互补的区域在图1A-1I和本文提供的其他描绘中有时显示为双链体。图7说明了使用SEQ ID NO:42的gRNA序列的gRNA结构域命名法,所述gRNA序列在tracrRNA衍生区域中含有一个发夹环。在某些实施例中,gRNA可以在该区域中含有多于一个(例如,两个、三个或更多个)发夹环(参见例如,图1H-1I)。

[0168] 在某些实施例中,单分子或嵌合gRNA包含,优选地从5'到3′:

[0169] 与γ-珠蛋白基因调控区域中的靶结构域互补的靶向结构域,例如来自SEQ ID NO:251-SEQ ID NO:901中任一个的靶向结构域;

[0170] 第一互补结构域;

[0171] 连接结构域;

[0172] 第二互补结构域(其与所述第一互补结构域互补);

[0173] 近端结构域;以及

[0174] 任选地,尾部结构域。

[0175] 在某些实施例中,模块化gRNA包含:

[0176] 第一链,其包含,优选地从5'到3':

[0177] 与γ-珠蛋白基因调控区域中的靶结构域互补的靶向结构域,例如来自SEQ ID NO:251-SEQ ID NO:901中任一个的靶向结构域;和

[0178] 第一互补结构域;以及

[0179] 第二条链,其包含,优选地从5'到3':

[0180] 任选地,5'延伸结构域;

[0181] 第二互补结构域;

[0182] 近端结构域;以及

[0183] 任选地,尾部结构域。

[0184] 靶向结构域

[0185] 靶向结构域(有时可替代地称为指导序列或互补区域)包含γ-珠蛋白基因调控区域中与靶核酸互补或部分互补的核酸序列,由其组成或基本上由其组成。γ-珠蛋白基因调控区域中全部或部分靶向结构域与其互补或部分互补的核酸序列在本文中称为靶结构域。在某些实施例中,靶结构域包含HBG靶位置。在其他实施例中,HBG靶位置位于靶结构域的外部(即,上游或下游)。在某些实施例中,靶结构域完全位于γ-珠蛋白基因调控区域内,例如,在与γ-珠蛋白基因有关的调控元件中或与编码γ-珠蛋白基因表达阻遏物的基因有关的调控元件中。在其他实施例中,靶结构域的全部或部分位于γ-珠蛋白基因调控区域之外,例如,在HBG1或HBG2编码区域、外显子、或内含子中。

[0186] 用于选择靶向结构域的方法本领域是已知的(参见例如,Fu 2014;Sternberg2014)。用于本文所述的方法、组合物和试剂盒中的合适的靶向结构域的实例包括SEQ ID No:251-901中示出的那些。

[0187] 包含靶结构域的靶核酸的链在本文中称为互补链,因为其与靶向结构域序列互补。由于靶向结构域是gRNA分子的一部分,所以它包含碱基尿嘧啶(U)而非胸腺嘧啶(T);相

反,编码gRNA分子的任何DNA分子将包含胸腺嘧啶而非尿嘧啶。在靶向结构域/靶结构域对中,靶向结构域中的尿嘧啶碱基将与靶结构域中的腺嘌呤碱基配对。在某些实施例中,靶向结构域与靶结构域之间的互补程度足以允许Cas9分子靶向靶核酸。

[0188] 在某些实施例中,靶向结构域包括核心结构域和任选的第二结构域。在这些实施例的某些中,核心结构域位于第二结构域的3′处,并且在这些实施例的某些中,核心结构域位于靶向结构域的3′端或其附近。在这些实施例的某些中,核心结构域由靶向结构域的3’端处的约8至约13个核苷酸组成或基本上由其组成。在某些实施例中,只有核心结构域与靶结构域的相应部分互补或部分互补,并且在这些实施例的某些中,核心结构域与靶结构域的相应部分完全互补。在其他实施例中,第二结构域也与靶结构域的一部分互补或部分互补。在某些实施例中,核心结构域与靶结构域中的核心结构域靶互补或部分互补,而第二结构域与靶结构域中的第二结构域靶互补或部分互补。在某些实施例中,核心结构域和第二结构域与靶结构的它们各自对应的部分域具有相同的互补程度。在其他实施例中,核心结构域与其靶之间的互补程度和第二结构域与其靶之间的互补程度可以不同。在这些实施例的某些中,核心结构域可以对其靶具有比第二结构域更高的互补程度,而在其他实施例中,第二结构域可以具有比核心结构域更高的互补程度。

[0189] 在某些实施例中,靶向结构域和/或靶向结构域内的核心结构域的长度为3至100、5至100、10至100或20至100个核苷酸,并且在这些实施例的某些中,靶向结构域或核心结构域的长度为3至15、3至20、5至20、10至20、15至20、5至50、10至50或20至50个核苷酸。在某些实施例中,靶向结构域和/或靶向结构域内的核心结构域的长度为6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25或26个核苷酸。在某些实施例中,靶向结构域和/或靶向结构域内的核心结构域的长度为6+/-2、7+/-2、8+/-2、9+/-2、10+/-2、10+/-4、10+/-5、11+/-2、12+/-2、13+/-2、14+/-2、15+/-2、或16+-2、20+/-5、30+/-5、40+/-5、50+/-5、60+/-5、70+/-5、80+/-5、90+/-5、或100+/-5个核苷酸。

[0190] 在靶向结构域包括核心结构域的某些实施例中,核心结构域的长度为3至20个核苷酸,并且在这些实施例的某些中,核心结构域的长度为5至15或8至13个核苷酸。在靶向结构域包括第二结构域的某些实施例中,第二结构域的长度为0、1、2、3、4、5、6、7、8、9、10、11、12、13、14或15个核苷酸。在靶向结构域包含长度为8至13个核苷酸的核心结构域的某些实施例中,各自地,靶向结构域的长度为26、25、24、23、22、21、20、19、18、17、或16个核苷酸,并且第二结构域的长度为13至18、12至17、11至16、10至15、9至14、8至13、7至12、6至11、5至10、4至9、或3至8个核苷酸。

[0191] 在某些实施例中,靶向结构域与靶结构域完全互补。同样地,在靶向结构域包含核心结构域和/或第二结构域的情况下,在某些实施例中,核心结构域和第二结构域中的一个或两个与靶结构域的相应部分完全互补。在其他实施例中,靶向结构域与靶结构域部分互补,并且在靶向结构域包含核心结构域和/或第二结构域的这些实施例的某些中,核心结构域和第二结构域中的一个或两个与靶结构域的相应部分部分地互补。在这些实施例的某些中,靶向结构域或靶向结构域内的核心结构域或第二结构域的核酸序列与靶结构域或靶结构域的相应部分至少80％、85％、90％或95％互补。在某些实施例中,靶向结构域和/或靶向结构域内的核心或第二结构域包括与靶结构域或其部分不互补的一个或多个核苷酸,并且在这些实施例的某些中,靶向结构域和/或靶向结构域内的核心或第二结构域包括与靶结

构域不互补的1、2、3、4、5、6、7或8个核苷酸。在某些实施例中,核心结构域包括与靶结构域的相应部分不互补的1、2、3、4或5个核苷酸。在靶向结构域包括与靶结构域不互补的一个或多个核苷酸的某些实施例中,所述非互补核苷酸中的一个或多个位于靶向结构域的5'或3'端的五个核苷酸内。在这些实施例的某些中,靶向结构域包括在其5′端、3′端或其5′和3′端的五个核苷酸内的与靶结构域不互补的1、2、3、4或5个核苷酸。在靶向结构域包括与靶结构域不互补的两个或更多个核苷酸的某些实施例中,所述非互补核苷酸中的两个或更多个彼此相邻,并且在这些实施例的某些中,所述两个或更多个连续的非互补核苷酸位于靶向结构域的5′或3'端的五个核苷酸内。在其他实施例中,所述两个或更多个连续的非互补核苷酸都位于离靶向结构域的5'和3'端超过五个核苷酸处。

[0192] 在某些实施例中,靶向结构域、核心结构域和/或第二结构域不包含任何修饰。在其他实施例中,靶向结构域、核心结构域和/或第二结构域或其中的一个或多个核苷酸具有修饰,包括但不限于以下阐述的修饰。在某些实施例中,靶向结构域、核心结构域和/或第二结构域的一个或多个核苷酸可以包含2'修饰(例如,在核糖上2'位置处的修饰),例如2-乙酰化,例如2′甲基化。在某些实施例中,可以用硫代磷酸酯修饰靶向结构域的骨架。在某些实施例中,对靶向结构域、核心结构域和/或第二结构域的一个或多个核苷酸的修饰使得靶向结构域和/或包含靶向结构域的gRNA不易降解或更生物相容,例如更低的免疫原性。在某些实施例中,靶向结构域和/或核心或第二结构域包括1、2、3、4、5、6、7或8个或更多个修饰,并且在这些实施例的某些中,靶向结构域和/或核心或第二结构域包括它们各自5'端的五个核苷酸内的1、2、3或4个修饰,和/或它们各自3'端的五个核苷酸内的1、2、3或4个修饰。在某些实施例中,靶向结构域和/或核心或第二结构域包括在两个或更多个连续核苷酸处的修饰。

[0193] 在靶向结构域包括核心和第二结构域的某些实施例中,核心和第二结构域含有相同数量的修饰。在这些实施例的某些中,这两个结构域都不含修饰。在其他实施例中,核心结构域包括比第二结构域更多的修饰,或反之亦然。

[0194] 在某些实施例中,选择对靶向结构域(包括核心或第二结构域)中的一个或多个核苷酸的修饰以不干扰靶向功效,这可以通过使用如下阐述的系统测试候选修饰来评价。具有候选靶向结构域的gRNA可以使用如下阐述的系统进行评价,所述候选靶向结构域具有选定的长度、序列、互补程度、或修饰程度。所述候选靶向结构域可以被单独地或与一种或多种其他候选变化放置在已知与选定的靶具有功能性的gRNA分子/Cas9分子系统中并且进行评价。

[0195] 在某些实施例中,全部的修饰核苷酸互补于并且能够杂交到靶结构域中存在的相应核苷酸上。在另一个实施例中,1、2、3、4、5、6、7或8个或更多个修饰核苷酸不互补于或不能够杂交到靶结构域中存在的相应核苷酸上。

[0196] 图1A-1I提供了靶向结构域在gRNA分子内的放置的实例。

[0197] <u>第一和第二互补结构域</u>

[0198] 第一和第二互补(有时可替代地分别称为crRNA衍生的发夹序列和tracrRNA衍生的发夹序列)结构域互相完全或部分互补。在某些实施例中,互补程度足以使所述两个结构域在至少一些生理条件下形成双链体区域。在某些实施例中,第一与第二互补结构域之间的互补程度与gRNA的其他特性一起足以允许Cas9分子靶向靶核酸。第一和第二互补结构域

的实例在图1A-1G中进行了阐述。

[0199] 在某些实施例中(参见例如,图1A-1B),第一和/或第二互补结构域包括与相应互补结构域缺乏互补性的一个或多个核苷酸。在某些实施例中,第一和/或第二互补结构域包括不与相应互补结构域互补的1、2、3、4、5或6个核苷酸。例如,第二互补结构域可以含有与第一互补结构域中的相应核苷酸不配对的1、2、3、4、5或6个核苷酸。在某些实施例中,第一或第二互补结构域上不与相应互补结构域互补的核苷酸从在第一与第二互补结构域之间形成的双链体环出。在这些实施例的某些中,未配对环出位于第二互补结构域上,并且在这些实施例的某些中,未配对区域从离第二互补结构域的5'端1、2、3、4、5或6个核苷酸处开始。

[0200] 在某些实施例中,第一互补结构域的长度为5至30、5至25、7至25、5至24、5至23、7至22、5至22、5至21、5至20、7至18、7至15、9至16、或10至14个核苷酸,并且在这些实施例的某些中,第一互补结构域的长度为5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、或25个核苷酸。在某些实施例中,第二互补结构域的长度为5至27、7至27、7至25、5至24、5至23、5至22、5至21、7至20、5至20、7至18、7至17、9至16、或10至14个核苷酸,并且在这些实施例的某些中,第二互补结构域的长度为5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25或26个核苷酸。在某些实施例中,第一和第二互补结构域的长度各自独立地为6+/-2、7+/-2、8+/-2、9+/-2、10+/-2、11+/-2、12+/-2、13+/-2、14+/-2、15+/-2、16+/-2、17+/-2、18+/-2、19+/-2、或20+/-2、21+/-2、22+/-2、23+/-2、或24+/-2个核苷酸。在某些实施例中,第二互补结构域长于第一互补结构域(例如,长出2、3、4、5、或6个核苷酸)。

[0201] 在某些实施例中,第一和/或第二互补结构域各自独立地包含三个亚结构域,其按5′到3'方向是:5'亚结构域、中央亚结构域、和3'亚结构域。在某些实施例中,第一互补结构域的5' 亚结构域和3' 亚结构域分别与第二互补结构域的3' 亚结构域和5' 亚结构域完全或部分互补。

[0202] 在某些实施例中,第一互补结构域的5'亚结构域的长度为4至9个核苷酸,并且在这些实施例的某些中,5'结构域的长度为4、5、6、7、8或9个核苷酸。在某些实施例中,第二互补结构域的5'亚结构域的长度为3至25、4至22、4至18、或4至10个核苷酸,并且在这些实施例的某些中,5′结构域的长度为3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24或25个核苷酸。在某些实施例中,第一互补结构域的中央亚结构域的长度为1、2或3个核苷酸。在某些实施例中,第二互补结构域的中央亚结构域的长度为1、2、3、4或5个核苷酸。在某些实施例中,第一互补结构域的3′亚结构域的长度为3至25、4至22、4至18、或4至10个核苷酸,并且在这些实施例的某些中,3'亚结构域的长度为3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24或25个核苷酸。在某些实施例中,第二互补结构域的3'亚结构域的长度为4至9个(例如4、5、6、7、8或9个)核苷酸。

[0203] 第一和/或第二互补结构域可以与天然存在的或参考的第一和/或第二互补结构域共享同源性,或者从其衍生。在这些实施例的某些中,第一和/或第二互补结构域与天然存在的或参考的第一和/或第二互补结构域具有至少50%、60%、70%、80%、85%、90%或95%的同源性或与其相差不多于1、2、3、4、5或6个核苷酸。在这些实施例的某些中,第一和/或第二互补结构域可以与来自化脓链球菌或金黄色葡萄球菌的第一和/或第二互补结构域

具有至少50％、60％、70％、80％、85％、90％或95％的同源性。

[0204] 在某些实施例中,第一和/或第二互补结构域不包含任何修饰。在其他实施例中,第一和/或第二互补结构域或其中的一个或多个核苷酸具有修饰,包括但不限于以下阐述的修饰。在某些实施例中,第一和/或第二互补结构域的一个或多个核苷酸可以包含2′修饰(例如在核糖上的2’位置处的修饰),例如2-乙酰化,例如2′甲基化。在某些实施例中,可以用硫代磷酸酯修饰靶向结构域的骨架。在某些实施例中,对第一和/或第二互补结构域的一个或多个核苷酸的修饰使得第一和/或第二互补结构域和/或包含第一和/或第二互补结构域的gRNA不易降解或更生物相容,例如更低的免疫原性。在某些实施例中,第一和/或第二互补结构域各自独立地包括1、2、3、4、5、6、7或8个或更多个修饰,并且在这些实施例的某些中,第一和/或第二互补结构域各自独立地包括它们各自的5’端、3’端或它们的5’和3’端的五个核苷酸内的1、2、3或4个修饰。在其他实施例中,第一和/或第二互补结构域各自独立地在它们各自的5’端、3’端或它们的5’和3′端的五个核苷酸内不含修饰。在某些实施例中,第一和第二互补结构域中的一个或两个包含在两个或更多个连续核苷酸处的修饰。

[0205] 在某些实施例中,选择对第一和/或第二互补结构域中的一个或多个核苷酸的修饰以不干扰靶向功效,这可以通过在如下面阐述的系统中测试候选修饰来评价。具有候选第一或第二互补结构域的gRNA可以在如下阐述的系统中进行评价,所述第一或第二互补结构域具有选定的长度、序列、互补程度、或修饰程度。所述候选互补结构域可以被单独地或与一种或多种其他候选变化放置在已知与选定的靶具有功能性的gRNA分子/Cas9分子系统中并且进行评价。

[0206] 在某些实施例中,由第一和第二互补结构域形成的双链体区域的长度为例如6bp、7bp、8bp、9bp、10bp、11bp、12bp、13bp、14bp、15bp、16bp、17bp、18bp、19bp、20bp、21bp或22bp,排除任何环出的或未配对的核苷酸。

[0207] 在某些实施例中,当成双链体时,第一和第二互补结构域包含11个配对的核苷酸(参见例如,SEQ ID NO:48的gRNA)。在某些实施例中,当成双链体时,第一和第二互补结构域包含15个配对的核苷酸(参见例如,SEQ ID NO:50的gRNA)。在某些实施例中,当成双链体时,第一和第二互补结构域包含16个配对的核苷酸(参见例如,SEQ ID NO:51的gRNA)。在某些实施例中,当成双链体时,第一和第二互补结构域包含21个配对的核苷酸(参见例如,SEQ ID NO:29的gRNA)。

[0208] 在某些实施例中,在第一与第二互补结构域之间交换一个或多个核苷酸以去除聚-U束。例如,可以交换SEQ ID NO:48的gRNA的核苷酸23和48或核苷酸26和45,以分别产生SEQ ID NO:49或31的gRNA。类似地,SEQ ID NO:29的gRNA的核苷酸23和39可以与核苷酸50和68交换以产生SEQ ID NO:30的gRNA。

[0209] <u>连接结构域</u>

[0210] 连接结构域被布置在单分子gRNA或嵌合gRNA中的第一和第二互补结构域之间并用于连接第一和第二互补结构域。图1B-1E提供了连接结构域的实例。在某些实施例中,连接结构域的一部分来自crRNA衍生区域,并且另一部分来自tracrRNA衍生区域。

[0211] 在某些实施例中,连接结构域共价连接第一和第二互补结构域。在这些实施例的某些中,连接结构域由共价键组成或包含共价键。在其他实施例中,连接结构域非共价连接第一和第二互补结构域。在某些实施例中,连接结构域的长度为十个或更少的核苷酸,例如

1、2、3、4、5、6、7、8、9或10个核苷酸。在其他实施例中,连接结构域的长度为大于10个核苷酸,例如11、12、13、14、15、16、17、18、19、20、21、22、23、24或25个或更多个核苷酸。在某些实施例中,连接结构域的长度为2至50、2至40、2至30、2至20、2至10、2至5、10至100、10至90、10至80、10至70、10至60、10至50、10至40、10至30、10至20、10至15、20至100、20至90、20至80、20至70、20至60、20至50、20至40、20至30或20至25个核苷酸。在某些实施例中,连接结构域的长度为10+/-5、20+/-5、20+/-10、30+/-5、30+/-10、40+/-5、40+/-10、50+/-5、50+/-10、60+/-5、60+/-10、70+/-5、70+/-10、80+/-5、80+/-10、90+/-5、90+/-10、100+/-5或100+/-10个核苷酸。

[0212] 在某些实施例中,连接结构域与天然存在的序列(例如,对所述第二互补结构域是5'的tracrRNA的序列)共享同源性,或从其衍生。在某些实施例中,连接结构域与本文所披露的连接结构域(例如,图1B-1E的连接结构域)具有至少50%、60%、70%、80%、90%或95%的同源性或与其相差不多于1、2、3、4、5或6个核苷酸。

[0213] 在某些实施例中,连接结构域不包含任何修饰。在其他实施例中,连接结构域或其中的一个或多个核苷酸具有修饰,包括但不限于以下阐述的修饰。在某些实施例中,连接结构域的一个或多个核苷酸可以包含2'修饰(例如在核糖上的2′位置处的修饰),例如2-乙酰化,例如2'甲基化。在某些实施例中,可以用硫代磷酸酯修饰连接结构域的骨架。在某些实施例中,对连接结构域的一个或多个核苷酸的修饰使得连接结构域和/或包含连接结构域的gRNA不易降解或更生物相容,例如更低的免疫原性。在某些实施例中,连接结构域包括1、2、3、4、5、6、7或8个或更多个修饰,并且在这些实施例的某些中,连接结构域包括其5'和/或3'端的五个核苷酸内的1、2、3或4个修饰。在某些实施例中,连接结构域包含在两个或更多个连续核苷酸处的修饰。

[0214] 在某些实施例中,选择对连接结构域中的一个或多个核苷酸的修饰以不干扰靶向功效,这可以通过在如下阐述的系统中测试候选修饰来评价。具有候选连接结构域的gRNA可以在如下阐述的系统中进行评价,所述候选连接结构域具有选定的长度、序列、互补程度、或修饰程度。所述候选连接结构域可以被单独地或与一种或多种其他候选变化放置在已知与选定的靶具有功能性的gRNA分子/Cas9分子系统中并且进行评价。

[0215] 在某些实施例中,连接结构域包含典型地邻近于第一互补结构域的3'端和/或第二互补结构域的5'端或在其1、2或3个核苷酸内的双链体区域。在这些实施例的某些中,连接区域的双链体区域的长度为10+/-5、15+/-5、20+/-5、20+/-10或30+/-5bp。在某些实施例中,连接结构域的双链体区域的长度为1、2、3、4、5、6、7、8、9、10、11、12、13、14或15bp。在某些实施例中,形成连接结构域的双链体区域的序列是完全互补的。在其他实施例中,形成双链体区域的一个或两个序列含有与其他双链体序列不互补的一个或多个核苷酸(例如,1、2、3、4、5、6、7或8个核苷酸)。

[0216] 5'延伸结构域

[0217] 在某些实施例中,如本文所披露的模块化gRNA包含5′延伸结构域,即第二互补结构域的5'的一个或多个另外的核苷酸(参见例如,图1A)。在某些实施例中,5'延伸结构域的长度为2至10或更多、2至9、2至8、2至7、2至6、2至5或2至4个核苷酸,并且在这些实施例的某些中,5'延伸结构域的长度为2、3、4、5、6、7、8、9或10个或更多个核苷酸。

[0218] 在某些实施例中,5'延伸结构域核苷酸不包含修饰,例如以下提供的类型的修饰。

然而,在某些实施例中,5’延伸结构域包含一个或多个修饰,例如,使其较不易降解或更生物相容(例如,更低的免疫原性)的修饰。作为举例,5’延伸结构域的骨架可以用硫代磷酸酯、或如下阐述的其他一个或多个修饰来修饰。在某些实施例中,5’延伸结构域的核苷酸可以包含2’修饰(例如,在核糖上的2’位置处的修饰),例如2-乙酰化,例如2’甲基化,或如下阐述的其他一个或多个修饰。

[0219] 在某些实施例中,5’延伸结构域可以包含多达1、2、3、4、5、6、7或8个修饰。在某些实施例中,5’延伸结构域包含其5’端的5个核苷酸内的多达1、2、3或4个修饰,例如在模块化gRNA分子中。在某些实施例中,5’延伸结构域包含其3’端的5个核苷酸内的多达1、2、3或4个修饰,例如在模块化gRNA分子中。

[0220] 在某些实施例中,5’延伸结构域包含在两个连续核苷酸处的修饰,例如5’延伸结构域的5’端的5个核苷酸内、5’延伸结构域的3’端的5个核苷酸内、或远离5’延伸结构域的一端或两端超过5个核苷酸的两个连续核苷酸。在某些实施例中,在5’延伸结构域的5’端的5个核苷酸内、5’延伸结构域的3’端的5个核苷酸内、或在远离5’延伸结构域的一端或两端超过5个核苷酸的区域内没有两个连续核苷酸被修饰。在某些实施例中,在5’延伸结构域的5’端的5个核苷酸内、5’延伸结构域的3’端的5个核苷酸内、或在远离5’延伸结构域的一端或两端超过5个核苷酸的区域内没有核苷酸被修饰。

[0221] 可以选择5’延伸结构域中的修饰以便不干扰gRNA分子功效,这可以通过在如下阐述的系统中测试候选修饰来评价。具有候选5’延伸结构域的gRNA可以在如下阐述的系统中进行评价,所述候选5’延伸结构域具有选定的长度、序列、互补程度、或修饰程度。候选5’延伸结构域可以被单独地或与一种或多种其他候选变化放置在已知与选定的靶具有功能性的gRNA分子/Cas9分子系统中并且进行评价。

[0222] 在某些实施例中,5’延伸结构域与参考5’延伸结构域(例如天然存在的(例如,化脓链球菌、金黄色葡萄球菌或嗜热链球菌)5’延伸结构域)、或本文所述的5′延伸结构域(例如,来自图1A-1G))具有至少60%、70%、80%、85%、90%或95%的同源性,或与其相差不多于1、2、3、4、5、或6个核苷酸。

[0223] <u>近端结构域</u>

[0224] 图1A-1G提供了近端结构域的实例。

[0225] 在某些实施例中,近端结构域的长度为5至20个或更多个核苷酸,例如长度为5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25或26个核苷酸。在这些实施例的某些中,近端结构域的长度为6+/-2、7+/-2、8+/-2、9+/-2、10+/-2、11+/-2、12+/-2、13+/-2、14+/-2、14+/-2、16+/-2、17+/-2、18+/-2、19+/-2、或20+/-2个核苷酸。在某些实施例中,近端结构域的长度为5至20、7至18、9至16、或10至14个核苷酸。

[0226] 在某些实施例中,近端结构域可以与天然存在的近端结构域共享同源性,或从其衍生。在这些实施例的某些中,近端结构域与本文所披露的近端结构域(例如化脓链球菌、金黄色葡萄球菌或嗜热链球菌)近端结构域,包括图1A-1G所阐述的那些)具有至少50%、60%、70%、80%、85%、90%或95%的同源性,或与其相差不多于1、2、3、4、5或6个核苷酸。

[0227] 在某些实施例中,近端结构域不包含任何修饰。在其他实施例中,近端结构域或其中的一个或多个核苷酸具有修饰,包括但不限于本文所阐述的修饰。在某些实施例中,近端结构域的一个或多个核苷酸可以包含2′修饰(例如在核糖上的2′位置处的修饰),例如2-乙

酰化,例如2′甲基化。在某些实施例中,可以用硫代磷酸酯修饰近端结构域的骨架。在某些实施例中,对近端结构域的一个或多个核苷酸的修饰使得近端结构域和/或包含近端结构域的gRNA不易降解或更生物相容,例如更低的免疫原性。在某些实施例中,近端结构域包括1、2、3、4、5、6、7或8个或更多个修饰,并且在这些实施例的某些中,近端结构域包括其5′和/或3′端的五个核苷酸内的1、2、3或4个修饰。在某些实施例中,近端结构域包含在两个或更多个连续核苷酸处的修饰。

[0228] 在某些实施例中,选择对近端结构域中的一个或多个核苷酸的修饰以不干扰靶向功效,这可以通过在如下阐述的系统中测试候选修饰来评价。具有候选近端结构域的gRNA可以在如下阐述的系统中进行评价,所述候选近端结构域具有选定的长度、序列、互补程度、或修饰程度。所述候选近端结构域可以被单独地或与一种或多种其他候选变化放置在已知与选定的靶具有功能性的gRNA分子/Cas9分子系统中并且进行评价。

[0229] 尾部结构域

[0230] 广谱的尾部结构域适于在本文所披露的gRNA分子中使用。图1A和1C-1G提供了这种尾部结构域的实例。

[0231] 在某些实施例中,不存在尾部结构域。在其他实施例中,尾部结构域的长度为1至100个或更多个核苷酸,例如长度为1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90或100个核苷酸。在某些实施例中,尾部结构域的长度为1至5、1至10、1至15、1至20、1至50、10至100、20至100、10至90、20至90、10至80、20至80、10至70、20至70、10至60、20至60、10至50、20至50、10至40、20至40、10至30、20至30、20至25、10至20或10至15个核苷酸。在某些实施例中,尾部结构域的长度为5+/-5、10+/-5、20+/-10、20+/-5、25+/-10、30+/-10、30+/-5、40+/-10、40+/-5、50+/-10、50+/-5、60+/-10、60+/-5、70+/-10、70+/-5、80+/-10、80+/-5、90+/-10、90+/-5、100+/-10或100+/-5个核苷酸。

[0232] 在某些实施例中,尾部结构域可以与天然存在的尾部结构域或天然存在的尾部结构域的5′端共享同源性,或从其衍生。在这些实施例的某些中,尾部结构域与本文所披露的天然存在的尾部结构域(例如化脓链球菌、金黄色葡萄球菌或嗜热链球菌尾部结构域,包括图1A和1C-1G所阐述的那些)具有至少50%、60%、70%、80%、85%、90%或95%的同源性,或与其相差不多于1、2、3、4、5或6个核苷酸。

[0233] 在某些实施例中,尾部结构域包括彼此互补,并且在至少一些生理条件下形成双链体区域的序列。在这些实施例的某些中,尾部结构域包含尾部双链体结构域,其可以形成尾部双链体区域。在某些实施例中,尾部双链体区域的长度为3、4、5、6、7、8、9、10、11或12bp。在某些实施例中,尾部结构域包含不形成双链体的尾部双链体结构域的3′的单链结构域。在这些实施例的某些中,单链结构域的长度为3至10个核苷酸(例如3、4、5、6、7、8、9、10个)或长度为4至6个核苷酸。

[0234] 在某些实施例中,尾部结构域不包含任何修饰。在其他实施例中,尾部结构域或其中的一个或多个核苷酸具有修饰,包括但不限于本文所阐述的修饰。在某些实施例中,尾部结构域的一个或多个核苷酸可以包含2′修饰(例如在核糖上的2′位置处的修饰),例如2-乙酰化,例如2′甲基化。在某些实施例中,可以用硫代磷酸酯修饰尾部结构域的骨架。在某些实施例中,对尾部结构域的一个或多个核苷酸的修饰使得尾部结构域和/或包含尾部结构域的gRNA不易降解或更生物相容,例如更低的免疫原性。在某些实施例中,尾部结构域包括

1、2、3、4、5、6、7或8个或更多个修饰,并且在这些实施例的某些中,尾部结构域包括其5′和/或3′端的五个核苷酸内的1、2、3或4个修饰。在某些实施例中,尾部结构域包含在两个或更多个连续核苷酸处的修饰。

[0235] 在某些实施例中,选择对尾部结构域中的一个或多个核苷酸的修饰以不干扰靶向功效,这可以通过测试如下阐述的候选修饰来评价。具有候选尾部结构域的gRNA可以使用如下阐述的系统进行评价,所述候选尾部结构域具有选定的长度、序列、互补程度、或修饰程度。所述候选尾部结构域可以被单独地或与一种或多种其他候选变化放置在已知与选定的靶具有功能性的gRNA分子/Cas9分子系统中并且进行评价。

[0236] 在某些实施例中,尾部结构域包括在3’端与体外或体内转录方法相关的核苷酸。当将T7启动子用于gRNA的体外转录时,这些核苷酸可以是DNA模板的3’端前存在的任何核苷酸。当将U6启动子用于体内转录时,这些核苷酸可以是序列UUUUUU。当将H1启动子用于转录时,这些核苷酸可以是序列UUUU。当使用替代的pol-III启动子时,这些核苷酸可以是各种数量的尿嘧啶碱基,这取决于例如pol-III启动子的终止信号,或者它们可以包括替代碱基。

[0237] 在某些实施例中,所述近端结构域和尾部结构域一起包含SEQ ID NO:32、33、34、35、36或37所示的序列,由其组成或基本上由其组成。

[0238] 示例性单分子/嵌合gRNA

[0239] 在某些实施例中,如本文所披露的单分子或嵌合gRNA具有结构:5′[靶向结构域]-[第一互补结构域]-[连接结构域]-[第二互补结构域]-[近端结构域]-[尾部结构域]-3′,其中:

[0240] 靶向结构域包含核心结构域和任选地第二结构域,并且长度为10至50个核苷酸;

[0241] 第一互补结构域的长度为5至25个核苷酸,并且在某些实施例中,与本文所披露的参考第一互补结构域具有至少50%、60%、70%、80%、85%、90%或95%的同源性;

[0242] 连接结构域的长度为1至5个核苷酸;

[0243] 第二互补结构域的长度为5至27个核苷酸,并且在某些实施例中,与本文所披露的参考第二互补结构域具有至少50%、60%、70%、80%、85%、90%或95%的同源性;

[0244] 近端结构域的长度为5至20个核苷酸,并且在某些实施例中,与本文所披露的参考近端结构域具有至少50%、60%、70%、80%、85%、90%或95%的同源性;并且

[0245] 尾部结构域是不存在的或是长度为1至50个核苷酸的核苷酸序列,并且在某些实施例中,与本文所披露的参考尾部结构域具有至少50%、60%、70%、80%、85%、90%或95%的同源性。

[0246] 在某些实施例中,如本文所披露的单分子gRNA包含,优选地从5’到3′:

[0247] 靶向结构域,其例如包含10-50个核苷酸;

[0248] 第一互补结构域,其例如包含15、16、17、18、19、20、21、22、23、24、25、或26个核苷酸;

[0249] 连接结构域;

[0250] 第二互补结构域;

[0251] 近端结构域;以及

[0252] 尾部结构域,

[0253] 其中,

[0254] (a)当一起考虑时,近端结构域和尾部结构域包含至少15、18、20、25、30、31、35、40、45、49、50、或53个核苷酸;

[0255] (b)第二互补结构域的最后一个核苷酸的3'存在至少15、18、20、25、30、31、35、40、45、49、50、或53个核苷酸;或

[0256] (c)第二互补结构域的最后一个核苷酸的3′存在至少16、19、21、26、31、32、36、41、46、50、51、或54个核苷酸,所述核苷酸与第一互补结构域的相应核苷酸互补。

[0257] 在某些实施例中,来自(a)、(b)、和/或(c)的序列与天然存在的gRNA的相应序列或与本文所述的gRNA具有至少50%、60%、70%、75%、80%、85%、90%、95%或99%的同源性。

[0258] 在某些实施例中,当一起考虑时,近端结构域和尾部结构域包含至少15、18、20、25、30、31、35、40、45、49、50、或53个核苷酸。

[0259] 在某些实施例中,所述第二互补结构域的最后一个核苷酸的3′存在至少15、18、20、25、30、31、35、40、45、49、50、或53个核苷酸。

[0260] 在某些实施例中,第二互补结构域的最后一个核苷酸的3'存在至少16、19、21、26、31、32、36、41、46、50、51或54个核苷酸,所述核苷酸与第一互补结构域的相应核苷酸互补。

[0261] 在某些实施例中,靶向结构域由16、17、18、19、20、21、22、23、24、25或26个与靶结构域或其部分互补或部分互补的核苷酸(例如,16、17、18、19、20、21、22、23、24、25或26个连续核苷酸)组成,基本上由其组成或包含其,例如,靶向结构域的长度为16、17、18、19、20、21、22、23、24、25或26个核苷酸。在这些实施例的某些中,靶向结构域在靶向结构域的整个长度、靶结构域的整个长度或两者上与靶结构域互补。

[0262] 在某些实施例中,本文所披露的单分子或嵌合gRNA分子(包含靶向结构域、第一互补结构域、连接结构域、第二互补结构域、近端结构域以及任选地尾部结构域)包含SEQ ID NO:42所示的核苷酸序列,其中靶向结构域被列为20个N(残基1-20)但长度范围可以从16至26个核苷酸,并且其中最后的六个残基(残基97-102)表示U6启动子的终止信号,但可以不存在或数目更少。在某些实施例中,所述单分子或嵌合gRNA分子是化脓链球菌gRNA分子。

[0263] 在某些实施例中,本文所披露的单分子或嵌合gRNA分子(包含靶向结构域、第一互补结构域、连接结构域、第二互补结构域、近端结构域以及任选地尾部结构域)包含SEQ ID NO:38所示的核苷酸序列,其中靶向结构域被列为20个N(残基1-20)但长度范围可以从16至26个核苷酸,并且其中最后的六个残基(残基97-102)表示U6启动子的终止信号,但可以不存在或数目更少。在某些实施例中,所述单分子或嵌合gRNA分子是金黄色葡萄球菌gRNA分子。

[0264] 示例性嵌合gRNA的序列和结构也示于图1H-1I中。

[0265] 示例性模块化gRNA

[0266] 在某些实施例中,本文所披露的模块化gRNA包含:

[0267] 第一链,其包含,优选地从5'到3':

[0268] 靶向结构域,其例如包含15、16、17、18、19、20、21、22、23、24、25、或26个核苷酸;

[0269] 第一互补结构域;以及

[0270] 第二条链,其包含,优选地从5'到3':

[0271] 任选地,5'延伸结构域;

[0272] 第二互补结构域;

[0273] 近端结构域;以及

[0274] 尾部结构域,

[0275] 其中:

[0276] (a)当一起考虑时,近端结构域和尾部结构域包含至少15、18、20、25、30、31、35、40、45、49、50、或53个核苷酸;

[0277] (b)第二互补结构域的最后一个核苷酸的3'存在至少15、18、20、25、30、31、35、40、45、49、50、或53个核苷酸;或

[0278] (c)第二互补结构域的最后一个核苷酸的3'存在至少16、19、21、26、31、32、36、41、46、50、51、或54个核苷酸,所述核苷酸与第一互补结构域的相应核苷酸互补。

[0279] 在某些实施例中,来自(a)、(b)、或(c)的序列与天然存在的gRNA的相应序列或与本文所述的gRNA具有至少60%、75%、80%、85%、90%、95%或99%的同源性。

[0280] 在某些实施例中,当一起考虑时,近端结构域和尾部结构域包含至少15、18、20、25、30、31、35、40、45、49、50、或53个核苷酸。

[0281] 在某些实施例中,所述第二互补结构域的最后一个核苷酸的3'存在至少15、18、20、25、30、31、35、40、45、49、50、或53个核苷酸。

[0282] 在某些实施例中,第二互补结构域的最后一个核苷酸的3'存在至少16、19、21、26、31、32、36、41、46、50、51或54个核苷酸,所述核苷酸与第一互补结构域的相应核苷酸互补。

[0283] 在某些实施例中,靶向结构域包含、具有、或由16、17、18、19、20、21、22、23、24、25或26个与靶结构域具有互补性的核苷酸(例如,16、17、18、19、20、21、22、23、24、25或26个连续核苷酸)组成,例如,靶向结构域的长度为16、17、18、19、20、21、22、23、24、25或26个核苷酸。

[0284] 在某些实施例中,靶向结构域由16、17、18、19、20、21、22、23、24、25或26个与靶结构域或其部分互补的核苷酸(例如,16、17、18、19、20、21、22、23、24、25或26个连续核苷酸)组成,基本上由其组成或包含其。在这些实施例的某些中,靶向结构域在靶向结构域的整个长度、靶结构域的整个长度或两者上与靶结构域互补。

[0285] 在某些实施例中,靶向结构域包含16个与靶结构域具有互补性的核苷酸(例如,16个连续核苷酸),由其组成或基本上由其组成,例如,靶向结构域的长度为16个核苷酸。在这些实施例的某些实施例中,(a)当一起考虑时,近端结构域和尾部结构域包含至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸;(b)第二互补结构域的最后一个核苷酸的3'存在至少15、18、20、25、30、31、35、40、45、49、50、或53个核苷酸;并且/或(c)第二互补结构域的最后一个核苷酸的3'存在至少16、19、21、26、31、32、36、41、46、50、51、或54个核苷酸,所述核苷酸与第一互补结构域的相应核苷酸互补。

[0286] 在某些实施例中,靶向结构域包含17个与靶结构域具有互补性的核苷酸(例如,17个连续核苷酸),由其组成或基本上由其组成,例如,靶向结构域的长度为17个核苷酸。在这些实施例的某些中,(a)当一起考虑时,近端结构域和尾部结构域包含至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸;(b)第二互补结构域的最后一个核苷酸的3'存在至少15、18、20、25、30、31、35、40、45、49、50、或53个核苷酸;并且/或(c)第二互补结构域的最

后一个核苷酸的3’存在至少16、19、21、26、31、32、36、41、46、50、51、或54个核苷酸，所述核苷酸与第一互补结构域的相应核苷酸互补。

[0287] 在某些实施例中，靶向结构域包含18个与靶结构域具有互补性的核苷酸(例如，18个连续核苷酸)，由其组成或基本上由其组成，例如，靶向结构域的长度为18个核苷酸。在这些实施例的某些中，(a)当一起考虑时，近端结构域和尾部结构域包含至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸；(b)第二互补结构域的最后一个核苷酸的3’存在至少15、18、20、25、30、31、35、40、45、49、50、或53个核苷酸；并且/或(c)第二互补结构域的最后一个核苷酸的3’存在至少16、19、21、26、31、32、36、41、46、50、51、或54个核苷酸，所述核苷酸与第一互补结构域的相应核苷酸互补。

[0288] 在某些实施例中，靶向结构域包含19个与靶结构域具有互补性的核苷酸(例如，19个连续核苷酸)，由其组成或基本上由其组成，例如，靶向结构域的长度为19个核苷酸。在这些实施例的某些中，(a)当一起考虑时，近端结构域和尾部结构域包含至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸；(b)第二互补结构域的最后一个核苷酸的3’存在至少15、18、20、25、30、31、35、40、45、49、50、或53个核苷酸；并且/或(c)第二互补结构域的最后一个核苷酸的3’存在至少16、19、21、26、31、32、36、41、46、50、51、或54个核苷酸，所述核苷酸与第一互补结构域的相应核苷酸互补。

[0289] 在某些实施例中，靶向结构域包含20个与靶结构域具有互补性的核苷酸(例如，20个连续核苷酸)，由其组成或基本上由其组成，例如，靶向结构域的长度为20个核苷酸。在这些实施例的某些中，(a)当一起考虑时，近端结构域和尾部结构域包含至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸；(b)第二互补结构域的最后一个核苷酸的3’存在至少15、18、20、25、30、31、35、40、45、49、50、或53个核苷酸；并且/或(c)第二互补结构域的最后一个核苷酸的3’存在至少16、19、21、26、31、32、36、41、46、50、51、或54个核苷酸，所述核苷酸与第一互补结构域的相应核苷酸互补。

[0290] 在某些实施例中，靶向结构域包含21个与靶结构域具有互补性的核苷酸(例如，21个连续核苷酸)，由其组成或基本上由其组成，例如，靶向结构域的长度为21个核苷酸。在这些实施例的某些中，(a)当一起考虑时，近端结构域和尾部结构域包含至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸；(b)第二互补结构域的最后一个核苷酸的3’存在至少15、18、20、25、30、31、35、40、45、49、50、或53个核苷酸；并且/或(c)第二互补结构域的最后一个核苷酸的3’存在至少16、19、21、26、31、32、36、41、46、50、51、或54个核苷酸，所述核苷酸与第一互补结构域的相应核苷酸互补。

[0291] 在某些实施例中，靶向结构域包含22个与靶结构域具有互补性的核苷酸(例如，22个连续核苷酸)，由其组成或基本上由其组成，例如，靶向结构域的长度为22个核苷酸。在这些实施例的某些中，(a)当一起考虑时，近端结构域和尾部结构域包含至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸；(b)第二互补结构域的最后一个核苷酸的3’存在至少15、18、20、25、30、31、35、40、45、49、50、或53个核苷酸；并且/或(c)第二互补结构域的最后一个核苷酸的3’存在至少16、19、21、26、31、32、36、41、46、50、51、或54个核苷酸，所述核苷酸与第一互补结构域的相应核苷酸互补。

[0292] 在某些实施例中，靶向结构域包含23个与靶结构域具有互补性的核苷酸(例如，23个连续核苷酸)，由其组成或基本上由其组成，例如，靶向结构域的长度为23个核苷酸。在这

些实施例的某些中,(a)当一起考虑时,近端结构域和尾部结构域包含至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸;(b)第二互补结构域的最后一个核苷酸的3'存在至少15、18、20、25、30、31、35、40、45、49、50、或53个核苷酸;并且/或(c)第二互补结构域的最后一个核苷酸的3'存在至少16、19、21、26、31、32、36、41、46、50、51、或54个核苷酸,所述核苷酸与第一互补结构域的相应核苷酸互补。

[0293] 在某些实施例中,靶向结构域包含24个与靶结构域具有互补性的核苷酸(例如,24个连续核苷酸),由其组成或基本上由其组成,例如,靶向结构域的长度为24个核苷酸。在这些实施例的某些中,(a)当一起考虑时,近端结构域和尾部结构域包含至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸;(b)第二互补结构域的最后一个核苷酸的3'存在至少15、18、20、25、30、31、35、40、45、49、50、或53个核苷酸;并且/或(c)第二互补结构域的最后一个核苷酸的3'存在至少16、19、21、26、31、32、36、41、46、50、51、或54个核苷酸,所述核苷酸与第一互补结构域的相应核苷酸互补。

[0294] 在某些实施例中,靶向结构域包含25个与靶结构域具有互补性的核苷酸(例如,25个连续核苷酸),由其组成或基本上由其组成,例如,靶向结构域的长度为25个核苷酸。在这些实施例的某些中,(a)当一起考虑时,近端结构域和尾部结构域包含至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸;(b)第二互补结构域的最后一个核苷酸的3'存在至少15、18、20、25、30、31、35、40、45、49、50、或53个核苷酸;并且/或(c)第二互补结构域的最后一个核苷酸的3'存在至少16、19、21、26、31、32、36、41、46、50、51、或54个核苷酸,所述核苷酸与第一互补结构域的相应核苷酸互补。

[0295] 在某些实施例中,靶向结构域包含26个与靶结构域具有互补性的核苷酸(例如,26个连续核苷酸),由其组成或基本上由其组成,例如,靶向结构域的长度为26个核苷酸。在这些实施例的某些中,(a)当一起考虑时,近端结构域和尾部结构域包含至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸;(b)第二互补结构域的最后一个核苷酸的3'存在至少15、18、20、25、30、31、35、40、45、49、50、或53个核苷酸;并且/或(c)第二互补结构域的最后一个核苷酸的3'存在至少16、19、21、26、31、32、36、41、46、50、51或54个核苷酸,所述核苷酸与该第一互补结构域的相应核苷酸互补。

[0296] gRNA递送

[0297] 在本文提供的方法的某些实施例中,所述方法包括递送如本文所述的一种或多种(例如,两种、三种或四种)gRNA分子。在某些这些实施例中,通过静脉内注射、肌内注射、皮下注射或吸入递送gRNA分子。

[0298] 用于设计gRNA的方法

[0299] 提供了用于选择、设计和验证用于如本文所述的gRNA的靶向结构域的方法。本文还提供了用于掺入gRNA的示例性靶向结构域。

[0300] 先前已经描述了用于靶序列的选择和验证以及脱靶分析的方法(参见例如,Mali 2013;Hsu 2013;Fu 2014;Heigwer 2014;Bae 2014;Xiao 2014)。例如,软件工具可以用来优化与使用者的靶序列对应的潜在靶向结构域的选择,例如,以跨基因组最小化总脱靶活性。脱靶活性可以不同于切割。对于使用化脓链球菌Cas9的每个可能的靶向结构域选择,所述工具可以鉴别跨基因组所有脱靶序列(以上的NAG或NGG PAM),所述脱靶序列含有高达一定数量(例如,1、2、3、4、5、6、7、8、9或10)的错配碱基对。在每个脱靶序列处的切割效率是可

以预测的,例如,使用实验衍生的加权方案。然后将每个可能的靶向结构域进行排序,根据其总的预测的脱靶切割;最高排名的靶向结构域表示可能具有最大中靶切割和最少脱靶切割的那些。其他功能(例如,用于CRISPR构建的自动化试剂设计、用于中靶Surveyor测定的引物设计、和用于高通量检测以及经由下一代测序对脱靶切割进行定量的引物设计)也可以被包括在所述工具中。可以使用本领域已知和/或本文所阐述的方法对候选靶向结构域和包含那些靶向结构域的gRNA进行功能评价。

[0301] 作为非限制性实例,使用DNA序列检索算法鉴定用于与化脓链球菌Cas9和金黄色葡萄球菌Cas9一起使用的gRNA中使用的靶向结构域。针对化脓链球菌靶向,设计了17-mer和20-mer靶向结构域,同时针对金黄色葡萄球菌靶向,设计了18-mer、19-mer、20-mer、21-mer、22-mer、23-mer、和24-mer靶向结构域。使用基于公共工具cas-offinder(Bae 2014)的定制gRNA设计软件进行gRNA设计。该软件在计算指导物的全基因组脱靶倾向之后为指导物打分。典型地,针对长度范围从17至24的指导物考虑范围从完美匹配至7个错配的匹配。一旦经计算确定了脱靶位点,便计算每种指导物的总分并且使用web界面以表格输出总结。除鉴定与PAM序列邻近的潜在靶位点之外,所述软件还鉴定与所选靶位点相差1、2、3个或多于3个核苷酸的所有PAM邻近序列。从UCSC基因组浏览器获得HBG1和HBG2调控区域基因组DNA序列,并且使用可公开获得的RepeatMasker程序针对重复元件对序列进行筛选。RepeatMasker针对重复元件和具有低复杂性的区域对输入DNA序列进行检索。输出是存在于给定查询序列中的重复的详细注释。

[0302] 识别后,基于他们到靶位点的距离、他们的正交性、和5′G的存在(基于在人类基因组中含有相关PAM的靠近匹配的识别,例如在化脓链球菌的情况下是NGG PAM,在金黄色葡萄球菌的情况下是NNGRRT(SEQ ID NO:204)或NNGRRV(SEQ ID NO:205)PAM),将靶向结构域排成等级。正交性是指在人类基因组中含有与靶序列最小数量的错配的序列的数量。“高水平的正交性”或“良好的正交性”可以例如是指20-mer靶向结构域,其除预期靶之外在人类基因组中既没有一致序列,又没有含有靶序列中的一个或两个错配的任何序列。具有良好正交性的靶向结构域被选择用于最小化脱靶DNA切割。

[0303] 针对单-gRNA核酸酶切割并且针对双-gRNA成对的“切口酶”策略两者对靶向结构域进行鉴定。用于选择靶向结构域并且确定哪些靶向结构域可以用于双-gRNA成对的“切口酶”策略的标准是基于两个考虑:

[0304] (1)靶向结构域对应该在DNA上定向成使得PAM朝外并且用D10A Cas9切口酶切割将产生5’突出端;和

[0305] (2)假设用双切口酶对切割将以合理频率导致整个插入序列的缺失。然而,用双切口酶对切割还可以仅在所述gRNA之一的位点处导致indel突变。可以针对它们如何有效地去除整个序列对比引起一个靶向结构域的靶位点处的indel突变对候选对成员进行测试。

[0306] 靶向结构域用于缺失HBG1 c.-114至-102

[0307] 识别了结合本文披露的方法用于在gRNA中用于缺失HBG1的c.-114至-102的靶向结构域并将其针对化脓链球菌和金黄色葡萄球菌排序成4个等级。

[0308] 对于化脓链球菌,基于(1)从靶位点任一末端上游或下游的距离(即,HBG1c.-114至-102),特异地是在靶位点任一末端的400bp内,(2)高水平的正交性,和(3)5’G的存在来选择等级1靶向结构域。基于(1)从靶位点任一末端上游或下游的距离(即,HBG1 c.-114至-

102),特异地是在靶位点任一末端的400bp内,和(2)高水平的正交性来选择等级2靶向结构域。基于(1)从靶位点任一末端上游或下游的距离(即,HBG1 c.-114至-102),特异地是在靶位点任一末端的400bp内,和(2)5'G的存在来选择等级3靶向结构域。基于从靶位点任一末端上游或下游的距离(即,HBG1c.-114至-102),特异地是在靶位点任一末端的400bp内来选择等级4靶向结构域。

[0309] 对于金黄色葡萄球菌,基于(1)从靶位点任一末端上游或下游的距离(即,HBG1c.-114至-102),特异地是在靶位点任一末端的400bp内,(2)高水平的正交性,(3)5'G的存在,和(4)具有序列NNGRRT(SEQ ID NO:204)的PAM来选择等级1靶向结构域。基于(1)从靶位点任一末端上游或下游的距离(即,HBG1 c.-114至-102),特异地是在靶位点任一末端的400bp内,(2)高水平的正交性,和(3)具有序列NNGRRT(SEQ ID NO:204)的PAM来选择等级2靶向结构域。基于(1)从靶位点任一末端上游或下游的距离(即,HBG1 c.-114至-102),特异地是在靶位点任一末端的400bp内,和(2)具有序列NNGRRT(SEQ ID NO:204)的PAM来选择等级3靶向结构域。基于(1)从靶位点任一末端上游或下游的距离(即,HBG1 c.-114至-102),特异地是在靶位点任一末端的400bp内,和(2)具有序列NNGRRV(SEQ ID NO:205)的PAM来选择等级4靶向结构域。

[0310] 需注意等级是非包容性的(对于所述策略,每种靶向结构域仅被列出一次)。在某些情况下,基于具体等级的标准未识别出靶向结构域。识别的靶向结构域总结在表6中。

[0311] 表6:化脓链球菌和金黄色葡萄球菌靶向结构域的核苷酸序列

[0312]

	化脓链球菌	金黄色葡萄球菌
等级1	SEQ ID NO: 251-SEQ ID NO: 256	SEQ ID NO: 367-SEQ ID NO: 376
等级2	SEQ ID NO: 257-SEQ ID NO: 274	SEQ ID NO: 343、SEQ ID NO: 377-SEQ ID NO: 393
等级3	SEQ ID NO: 275-SEQ ID NO:	SEQ ID NO: 357、SEQ ID

[0313]

	300	NO: 365、SEQ ID NO: 394-SEQ ID NO: 461
等级4	SEQ ID NO: 301-SEQ ID NO: 366	SEQ ID NO: 252-SEQ ID NO: 254、SEQ ID NO: 256、SEQ ID NO: 268、SEQ ID NO: 272-SEQ ID NO: 274、SEQ ID NO: 292、SEQ ID NO: 295、SEQ ID NO: 347、SEQ ID NO: 348、SEQ ID NO: 353、SEQ ID NO: 360-SEQ ID NO: 362、SEQ ID NO: 366、SEQ ID NO: 598-SEQ ID NO: 759

[0314] 靶向结构域用于缺失HBG2 c.-114至-102

[0315] 识别了结合本文披露的方法用于在gRNA中用于缺失HBG2的c.-114至-102的靶向结构域并将其针对化脓链球菌和金黄色葡萄球菌排序成4个等级。

[0316] 对于化脓链球菌,基于(1)从靶位点任一末端上游或下游的距离(即,HBG2c.-114至-102),特异地是在靶位点任一末端的400bp内,(2)高水平的正交性,和(3)5'G的存在来选择等级1靶向结构域。基于(1)从靶位点任一末端上游或下游的距离(即,HBG2 c.-114至-102),特异地是在靶位点任一末端的400bp内,和(2)高水平的正交性来选择等级2靶向结构域。基于(1)从靶位点任一末端上游或下游的距离(即,HBG2 c.-114至-102),特异地是在靶位点任一末端的400bp内,和(2)5′G的存在来选择等级3靶向结构域。基于从靶位点任一末端上游或下游的距离(即,HBG2c.-114至-102),特异地是在靶位点任一末端的400bp内来选择等级4靶向结构域。

[0317] 对于金黄色葡萄球菌,基于(1)从靶位点任一末端上游或下游的距离(即,HBG2c.-114至-102),特异地是在靶位点任一末端的400bp内,(2)高水平的正交性,(3)5'G的存在,和(4)具有序列NNGRRT(SEQ ID NO:204)的PAM来选择等级1靶向结构域。基于(1)从靶位点任一末端上游或下游的距离(即,HBG2 c.-114至-102),特异地是在靶位点任一末端的400bp内,(2)高水平的正交性,和(3)具有序列NNGRRT(SEQ ID NO:204)的PAM来选择等级2靶向结构域。基于(1)从靶位点任一末端上游或下游的距离(即,HBG2 c.-114至-102),特异地是在靶位点任一末端的400bp内,和(2)具有序列NNGRRT(SEQ ID NO:204)的PAM来选择等级3靶向结构域。基于(1)从靶位点任一末端上游或下游的距离(即,HBG2 c.-114至-102),特异地是在靶位点任一末端的400bp内,和(2)具有序列NNGRRV(SEQ ID NO:205)的PAM来选择等级4靶向结构域。

[0318] 需注意等级是非包容性的(对于所述策略,每种靶向结构域仅被列出一次)。在某些情况下,基于具体等级的标准未识别出靶向结构域。识别的靶向结构域总结在表7中。

[0319] 表7:化脓链球菌和金黄色葡萄球菌靶向结构域的核苷酸序列

[0320]

	化脓链球菌	金黄色葡萄球菌
等级1	SEQ ID NO: 760-SEQ ID NO: 764	SEQ ID NO: 784-SEQ ID NO: 791
等级2	SEQ ID NO: 765-SEQ ID NO: 781	SEQ ID NO: 778、SEQ ID NO: 792-SEQ ID NO: 803
等级3	SEQ ID NO: 275-SEQ ID NO: 281、SEQ ID NO: 283-SEQ ID NO: 300	SEQ ID NO: 357、SEQ ID NO: 365、SEQ ID NO: 394-SEQ ID NO: 461
等级4	SEQ ID NO: 301-SEQ ID NO: 311、SEQ ID NO: 313-SEQ ID NO: 342、SEQ ID NO: 344-SEQ ID NO: 348、SEQ ID NO: 350-SEQ ID NO: 366、SEQ ID NO: 782、SEQ ID NO: 783	SEQ ID NO: 292、SEQ ID NO: 295、SEQ ID NO: 347、SEQ ID NO: 348、SEQ ID NO: 353、SEQ ID NO: 360-SEQ ID NO: 362、SEQ ID NO: 366、SEQ ID NO: 462-SEQ ID NO: 468、SEQ ID NO: 476-SEQ ID NO: 481、SEQ ID NO: 489-SEQ ID NO: 587、SEQ ID NO: 601-SEQ ID NO: 607、SEQ ID NO: 614-SEQ ID NO: 620、SEQ ID NO: 640-SEQ ID NO: 666、SEQ ID NO: 674-SEQ ID NO: 679、SEQ ID NO: 687-SEQ ID NO: 693、SEQ ID NO: 708-SEQ ID NO: 714、

[0321]

		SEQ ID NO: 733-SEQ ID NO: 753、SEQ ID NO: 762-SEQ ID NO: 764、SEQ ID NO: 775、SEQ ID NO: 779-SEQ ID NO: 781、SEQ ID NO: 804-SEQ ID NO: 901

[0322] 在某些实施例中,两种或更多种(例如,三种或四种)gRNA分子与一种Cas9分子一起使用。在另一个实施例中,当两种或更多种(例如,三种或四种)gRNA与两种或更多种Cas9分子一起使用时,至少一种Cas9分子来自与其他一种或多种Cas9分子不同的物种。例如,当两个gRNA分子与两个Cas9分子一起使用时,一个Cas9分子可以来自一种物种并且另一个Cas9分子可以来自不同的物种。根据需要,两种Cas9种类均用于产生单链或双链断裂。

[0323] 本文所述表格中的任何靶向结构域可与产生单链断裂的Cas9分子(即,化脓链球菌或金黄色葡萄球菌Cas9切口酶)或产生双链断裂的Cas9分子一起使用(即,化脓链球菌或金黄色葡萄球菌Cas9核酸酶)。

[0324] 当两个gRNA设计用于两个Cas9分子时,两个Cas9分子可能是不同的物种。根据希

望,两种Cas9种类均可用于产生单链或双链断裂。

[0325] 本文中考虑了任何上游gRNA可与本文所述的任何下游gRNA配对。当设计用于一种Cas9的上游gRNA与设计用于不同种类的Cas9的下游gRNA配对时,根据希望,两种Cas9种都用于产生单链或双链断裂。

[0326] RNA指导的核酸酶

[0327] 根据本披露内容的RNA指导的核酸酶包括但不限于天然存在的2类CRISPR核酸酶,例如Cas9和Cpf1,以及由其衍生或获得的其他核酸酶。在功能上,RNA指导的核酸酶定义为以下核酸酶:(a)与gRNA相互作用(例如,复合);和(b)与gRNA一起,与DNA的靶向区域有关,并任选地切割或修饰,所述DNA靶向区域包括(i)与gRNA的靶向结构域互补的序列,和任选地,(ii)PAM。RNA指导的核酸酶可以在广义上定义为他们的PAM特异性和切割活性,即使在具有相同PAM特异性或切割活性的各个RNA指导的核酸酶之间可能存在变异。技术人员可以理解,本披露内容的一些方面涉及可以使用具有某些PAM特异性和/或切割活性的任何合适的RNA指导的核酸酶实施的系统、方法和组合物。因此,除非另有说明,术语RNA指导的核酸酶应理解为通用术语,并不限于任何特定类型(例如,Cas9与Cpf1),物种(例如,化脓链球菌与金黄色葡萄球菌)或RNA指导的核酸酶的变异(例如,全长与截短的或分裂;天然存在的PAM特异性与工程化PAM特异性等)。

[0328] PAM序列的名称来源于与gRNA靶向结构域(或“间隔子”)互补的“原型间隔区域”序列的顺序关系。与原型间隔子序列一起,PAM序列定义特定RNA指导的核酸酶/gRNA组合的靶向结构域或序列。

[0329] 各种RNA指导的核酸酶可能需要PAM和原型间隔子之间的不同顺序关系。通常,Cas9s识别相对于顶部或互补链可视化的原型间隔子3′的PAM序列:

[0330] 5′----------[原型间隔子]----------3’

[0331] 3’----------[PAM]----------5’

[0332] 另一方面,Cpf1通常识别原型间隔子5′的PAM序列:

[0333] 5′----------[原型间隔子]----------3’

[0334] 3′----------[PAM]----------5’

[0335] 除了识别PAM和原型间隔子的特异顺序定向外,RNA指导的核酸酶还可以识别特异的PAM序列。例如,金黄色葡萄球菌Cas9识别NNGRRT或NNGRRV的PAM序列,其中N个残基紧邻gRNA靶向结构域识别的结构域的3′。化脓链球菌Cas9识别NGG PAM序列。并且新凶手弗朗西斯菌(F.novicida)Cpf1识别TTN PAM序列。已经识别了多种RNA指导的核酸酶的PAM序列,并且Shmakov 2015描述了用于识别新的PAM序列的策略。还应该注意的是,工程化的RNA指导的核酸酶可具有不同于参考分子的PAM特异性的PAM特异性(例如,在工程化的RNA指导的核酸酶的情况下,参考分子可以是天然存在的变体,其中RNA指导的核酸酶是衍生的,或与工程化的RNA指导的核酸酶具有最大的氨基酸序列同源性的天然存在的变体)。

[0336] 除了他们的PAM特异性外,RNA指导的核酸酶还可以通过他们的DNA切割活性来表征:天然存在的RNA指导的核酸酶典型地在靶核标酸中形成DSB,但是已经产生仅产生SSB的工程化的变体(如上所讨论的)Ran 2013年,通过引用结合在此),或完全不切割的那些。

[0337] Cas9分子

[0338] 多个物种的Cas9分子可以用于本文所述的方法和组合物中。尽管化脓链球菌和金

黄色葡萄球菌Cas9分子是本披露的大部分的主题,也可以使用在此列出的其他物种的Cas9蛋白质的、从中衍生的、或基于其的Cas9分子。这些包括例如来自以下项的Cas9分子:燕麦食酸菌(Acidovorax avenae)、胸膜肺炎放线杆菌(Actinobacillus pleuropneumoniae)、琥珀酸放线杆菌(Actinobacillus succinogenes)、猪放线杆菌(Actinobacillus suis)、放线菌属(Actinomyces sp.)、cycliphilus denitrificans、Aminomonas paucivorans、蜡样芽孢杆菌(Bacillus cereus)、史氏芽孢杆菌(Bacillus smithii)、苏云金芽孢杆菌(Bacillus thuringiensis)、拟杆菌属(Bacteroides sp.)、Blastopirellula marina、慢生根瘤菌属(Bradyrhizobium sp.)、侧孢短芽孢杆菌(Brevibacillus laterosporus)、结肠弯曲菌(Campylobacter coli)、空肠弯曲菌(Campylobacter jejuni)、红嘴鸥弯曲杆菌(Campylobacter lari)、Candidatus puniceispirillum、解纤维梭菌(Clostridium cellulolyticum)、产气荚膜梭菌(Clostridium perfringens)、拥挤棒杆菌(Corynebacterium accolens)、白喉棒状杆菌(Corynebacterium diphtheria)、Corynebacterium matruchotii、恒雄芝氏沟鞭藻玫瑰杆菌(Dinoroseobacter shibae)、细长真杆菌(Eubacterium dolichum)、γ-变形菌纲(gamma proteobacterium)、重氮营养葡糖酸醋杆菌(Gluconacetobacter diazotrophicus)、副流感嗜血杆菌(Haemophilus parainfluenzae)、嗜血弯曲杆菌(Haemophilus sputorum)、Helicobacter canadensis、同性恋螺杆菌(Helicobacter cinaedi)、鼬鼠螺杆菌(Helicobacter mustelae)、Ilyobacter polytropus、金氏金氏菌(Kingella kingae)、卷曲乳酸杆菌(Lactobacillus crispatus)、伊氏李斯特菌(Listeria ivanovii)、单核细胞增生李斯特菌(Listeria monocytogenes)、李斯特氏菌科菌(Listeriaceae bacterium)、甲基孢囊菌属(Methylocystis sp.)、甲烷氧化菌(Methylosinus trichosporium)、羞怯动弯杆菌(Mobiluncus mulieris)、奈瑟氏杆菌(Neisseria bacilliformis)、灰色奈瑟球菌(Neisseria cinerea)、浅黄色奈瑟氏菌(Neisseria flavescens)、乳糖奈瑟氏菌(Neisseria lactamica)、奈瑟氏菌属(Neisseria sp.)、瓦茨瓦尔西奈瑟氏菌(Neisseria wadsworthii)、亚硝化单胞菌属(Nitrosomonas sp.)、食清洁剂细小棒菌(Parvibaculum lavamentivorans)、出血败血性巴士杆菌(Pasteurella multocida)、Phascolarctobacterium succinatutens、Ralstonia syzygii、沼泽红假单胞菌(Rhodopseudomonas palustris)、小红卵菌属(Rhodovulum sp.)、米氏西蒙斯氏菌(Simonsiella muelleri)、鞘氨醇单胞菌属(Sphingomonas sp.)、Sporolactobacillus vineae、路邓葡萄球菌(Staphylococcus lugdunensis)、链球菌属(Streptococcus sp.)、Subdoligranulum sp.、运动替斯崔纳菌(Tistrella mobilis)、密螺旋体属(Treponema sp.)或Verminephrobacter eiseniae。

[0339] Cas9结构域

[0340] 已经确定了两种不同的天然存在的细菌Cas9分子(Jinek 2014)和具有指导RNA(例如,crRNA和tracrRNA的合成融合体)的化脓链球菌Cas9(Nishimasu 2014;Anders 2014)的晶体结构。

[0341] 天然存在的Cas9分子包含两种叶片:识别(REC)叶片和核酸酶(NUC)叶片;其各自进一步包含本文所述的结构域。图8A-8B提供了重要的Cas9结构域的一级结构的组织的示意图。贯穿本披露使用的由每个结构域所涵盖的结构域命名和氨基酸残基编号是如先前(Nishimasu 2014)所描述的。氨基酸残基的编号是参照来自化脓链球菌的Cas9。

[0342] REC叶片包含富精氨酸的桥螺旋(BH)、REC1结构域、和REC2结构域。REC叶片与其他已知蛋白不享有结构相似性,指示它是Cas9特异性功能结构域。BH结构域是长的α螺旋且富精氨酸的区域并且包含化脓链球菌Cas9(SEQ ID NO:2)的氨基酸60-93。REC1结构域对于例如gRNA或tracrRNA的重复:抗重复双链体的识别而言是重要的,并且因此对于识别靶序列的Cas9活性而言是关键的。REC1结构域在化脓链球菌Cas9(SEQ ID NO:2)的氨基酸94至179和308至717处包含两个REC1基序。尽管在线性一级结构中被REC2结构域分开,这两个REC1结构域在三级结构中组装以形成REC1结构域。REC2结构域、或其部分在重复:抗重复双链体的识别中也可以发挥作用。REC2结构域包含化脓链球菌Cas9(SEQ ID NO:2)的氨基酸180-307。

[0343] NUC叶片包含RuvC结构域、HNH结构域和PAM相互作用(PI)结构域。RuvC结构域与逆转录病毒整合酶超家族成员享有结构相似性,并且切割靶核酸分子的单链(例如,非互补链)。RuvC结构域由化脓链球菌Cas9(SEQ ID NO:2)的分别在氨基酸1-59、718-769和909-1098处的三个分割RuvC基序(RuvCI、RuvCII和RuvCIII,在本领域其通常分别称为RuvCI结构域或N-末端RuvC结构域、RuvCII结构域和RuvCIII结构域)组装而来。与REC1结构域类似,这三个RuvC基序被一级结构中的其他结构域线性分开。然而,在三级结构中,这三个RuvC基序组装并形成RuvC结构域。HNH结构域与HNH内切核酸酶享有结构相似性,并且切割靶核酸分子的单链(例如,非互补链)。HNH结构域位于RuvC II-III基序之间并且包含化脓链球菌Cas9(SEQ ID NO:2)的氨基酸775-908。PI结构域与靶核酸分子的PAM相互作用,并且包含化脓链球菌Cas9(SEQ ID NO:2)的氨基酸1099-1368。

[0344] RuvC样结构域和HNH样结构域

[0345] 在某些实施例中,Cas9分子或Cas9多肽包含HNH样结构域和RuvC样结构域,并且在这些实施例的某些中,切割活性取决于RuvC样结构域和HNH样结构域。Cas9分子或Cas9多肽可以包含RuvC样结构域和HNH样结构域中的一者或多者。在某些实施例中,Cas9分子或Cas9多肽包含RuvC样结构域(例如,如下所述的RuvC样结构域)和/或HNH样结构域(例如,如下所述的HNH样结构域)。

[0346] RuvC样结构域

[0347] 在某些实施例中,RuvC样结构域切割靶核酸分子的单链(例如,非互补链)。Cas9分子或Cas9多肽可以包括多于一个RuvC样结构域(例如,一个、两个、三个或更多个RuvC样结构域)。在某些实施例中,RuvC样结构域的长度为至少5、6、7、8个氨基酸但长度不多于20、19、18、17、16或15个氨基酸。在某些实施例中,Cas9分子或Cas9多肽包含长度为约10至20个氨基酸(例如,约15个氨基酸)的N-末端RuvC样结构域。

[0348] N-末端RuvC样结构域

[0349] 一些天然存在的Cas9分子包含多于一个RuvC样结构域,其中切割取决于N-末端RuvC样结构域。因此,Cas9分子或Cas9多肽可以包含N-末端RuvC样结构域。示例性N-末端RuvC样结构域如下所述。

[0350] 在某些实施例中,Cas9分子或Cas9多肽包含N-末端RuvC样结构域,所述N-末端RuvC样结构域包含具有化学式I的氨基酸序列:

[0351] D-X_1-G-X_2-X_3-X_4-X_5-G-X_6-X_7-X_8-X_9(SEQ ID NO:20),

[0352] 其中

[0353] X_1选自I、V、M、L和T(例如,选自I、V和L);

[0354] X_2选自T、I、V、S、N、Y、E和L(例如,选自T、V和I);

[0355] X_3选自N、S、G、A、D、T、R、M和F(例如,A或N);

[0356] X_4选自S、Y、N和F(例如,S);

[0357] X_5选自V、I、L、C、T和F(例如,选自V、I和L);

[0358] X_6选自W、F、V、Y、S和L(例如,W);

[0359] X_7选自A、S、C、V和G(例如,选自A和S);

[0360] X_8选自V、I、L、A、M和H(例如,选自V、I、M和L);并且

[0361] X_9选自任何氨基酸或是不存在的(例如,选自T、V、I、L、Δ、F、S、A、Y、M和R,或例如,选自T、V、I、L和Δ)。

[0362] 在某些实施例中,N-末端RuvC样结构域与SEQ ID NO:20的序列相差多达1个但不多于2、3、4或5个残基。

[0363] 在某些实施例中,N-末端RuvC样结构域是有切割能力的。在其他实施例中,N-末端RuvC样结构域是无切割能力的。

[0364] 在某些实施例中,Cas9分子或Cas9多肽包含N-末端RuvC样结构域,所述N-末端RuvC样结构域包含具有化学式II的氨基酸序列:

[0365] D-X_1-G-X_2-X_3-S-X_5-G-X_6-X_7-X_8-X_9(SEQ ID NO:21),

[0366] 其中

[0367] X_1选自I、V、M、L和T(例如,选自I、V和L);

[0368] X_2选自T、I、V、S、N、Y、E和L(例如,选自T、V和I);

[0369] X_3选自N、S、G、A、D、T、R、M和F(例如,A或N);

[0370] X_5选自V、I、L、C、T和F(例如,选自V、I和L);

[0371] X_6选自W、F、V、Y、S和L(例如,W);

[0372] X_7选自A、S、C、V和G(例如,选自A和S);

[0373] X_8选自V、I、L、A、M和H(例如,选自V、I、M和L);并且

[0374] X_9选自任何氨基酸或是不存在的(例如,选自T、V、I、L、Δ、F、S、A、Y、M和R,或选自例如,T、V、I、L和Δ)。

[0375] 在某些实施例中,N-末端RuvC样结构域与SEQ ID NO:21的序列相差多达1个但不多于2、3、4或5个残基。

[0376] 在某些实施例中,N-末端RuvC样结构域包含具有化学式III的氨基酸序列:

[0377] D-I-G-X_2-X_3-S-V-G-W-A-X_8-X_9(SEQ ID NO:22),

[0378] 其中

[0379] X_2选自T、I、V、S、N、Y、E和L(例如,选自T、V和I);

[0380] X_3选自N、S、G、A、D、T、R、M和F(例如,A或N);

[0381] X_8选自V、I、L、A、M和H(例如,选自V、I、M和L);并且

[0382] X_9选自任何氨基酸或是不存在的(例如,选自T、V、I、L、Δ、F、S、A、Y、M和R,或选自例如,T、V、I、L和Δ)。

[0383] 在某些实施例中,N-末端RuvC样结构域与SEQ ID NO:22的序列相差多达1个但不多于2、3、4或5个残基。

[0384] 在某些实施例中,N-末端RuvC样结构域包含具有化学式IV的氨基酸序列:

[0385] D-I-G-T-N-S-V-G-W-A-V-X(SEQ ID NO:23),

[0386] 其中

[0387] X是非极性烷基氨基酸或羟基氨基酸,例如,X选自V、I、L和T(例如,Cas9分子可以包含示于图2A-2G中的N-末端RuvC样结构域(描绘为Y))。

[0388] 在某些实施例中,N-末端RuvC样结构域与SEQ ID NO:23的序列相差多达1个但不多于2、3、4或5个残基。

[0389] 在某些实施例中,N-末端RuvC样结构域与本文(例如,在图3A-3B中)所披露的N-末端RuvC样结构域的序列相差多达1个但不多于2、3、4或5个残基。在一个实施例中,在图3A-3B中鉴定出的高度保守残基中的1个、2个、3个或全部是存在的。

[0390] 在某些实施例中,N-末端RuvC样结构域与本文(例如,在图4A-4B中)所披露的N-末端RuvC样结构域的序列相差多达1个但不多于2、3、4或5个残基。在一个实施例中,在图4A-4B中识别出的高度保守残基中的1个、2个或全部是存在的。

[0391] 另外的RuvC样结构域

[0392] 除了N-末端RuvC样结构域之外,Cas9分子或Cas9多肽可以包含一个或多个另外的RuvC样结构域。在某些实施例中,Cas9分子或Cas9多肽可以包含两个另外的RuvC样结构域。优选地,另外的RuvC样结构域的长度为至少5个氨基酸,并且例如长度为小于15个氨基酸,例如长度为5至10个氨基酸,例如长度为8个氨基酸。

[0393] 另外的RuvC样结构域可以包含具有化学式V的氨基酸序列:

[0394] I-X_1-X_2-E-X_3-A-R-E(SEQ ID NO:15),

[0395] 其中

[0396] X_1是V或H;

[0397] X_2是I、L或V(例如,I或V);并且

[0398] X_3是M或T。

[0399] 在某些实施例中,另外的RuvC样结构域包含具有化学式VI的氨基酸序列:

[0400] I-V-X_2-E-M-A-R-E(SEQ ID NO:16),

[0401] 其中

[0402] X_2是I、L或V(例如,I或V)(例如,Cas9分子或Cas9多肽可以包含示于图2A-2G中的另外的RuvC样结构域(描绘为B))。

[0403] 另外的RuvC样结构域可以包含具有化学式VII的氨基酸序列:

[0404] H-H-A-X_1-D-A-X_2-X_3(SEQ ID NO:17),

[0405] 其中

[0406] X_1是H或L;

[0407] X_2是R或V;并且

[0408] X_3是E或V。

[0409] 在某些实施例中,另外的RuvC样结构域包含如下氨基酸序列:H-H-A-H-D-A-Y-L(SEQ ID NO:18)。

[0410] 在某些实施例中,另外的RuvC样结构域与SEQ ID NO:15-18的序列相差多达1个但不多于2、3、4或5个残基。

[0411] 在某些实施例中,N-末端RuvC样结构域侧翼的序列具有化学式VIII的氨基酸序列:

[0412] K-X_1'-Y-X_2'-X_3'-X_4'-Z-T-D-X_9'-Y(SEQ ID NO:19),

[0413] 其中

[0414] X_1'选自K和P;

[0415] X_2'选自V、L、I和F(例如,V、I和L);

[0416] X_3'选自G、A和S(例如,G);

[0417] X_4'选自L、I、V和F(例如,L);

[0418] X_9'选自D、E、N和Q;并且

[0419] Z是N末端RuvC样结构域,例如,如上所述,例如具有5至20个氨基酸。

[0420] HNH样结构域

[0421] 在某些实施例中,HNH样结构域切割双链核酸分子的单链互补结构域(例如,互补链)。在某些实施例中,HNH样结构域的长度为至少15、20或25个氨基酸但长度为不多于40、35或30个氨基酸,例如长度为20至35个氨基酸,例如长度为25至30个氨基酸。示例性的HNH样结构域如下所述。

[0422] 在一个实施例中,Cas9分子或Cas9多肽包含HNH样结构域,所述HNH样结构域具有化学式IX的氨基酸序列:

[0423] X_1-X_2-X_3-H-X_4-X_5-P-X_6-X_7-X_8-X^9-X^{10}-X^{11}-X^{12}-X^{13}-X^{14}-X^{15}-N-X^{16}-X^{17}-X^{18}-X^{19}-X_{20}-X_{21}-X_{22}-X_{23}-N(SEQ ID NO:25),

[0424] 其中

[0425] X_1选自D、E、Q和N(例如,D和E);

[0426] X_2选自L、I、R、Q、V、M和K;

[0427] X_3选自D和E;

[0428] X_4选自I、V、T、A和L(例如,A、I和V);

[0429] X_5选自V、Y、I、L、F和W(例如,V、I和L);

[0430] X_6选自Q、H、R、K、Y、I、L、F和W;

[0431] X_7选自S、A、D、T和K(例如,S和A);

[0432] X_8选自F、L、V、K、Y、M、I、R、A、E、D和Q(例如,F);

[0433] X_9选自L、R、T、I、V、S、C、Y、K、F和G;

[0434] X_{10}选自K、Q、Y、T、F、L、W、M、A、E、G和S;

[0435] X_{11}选自D、S、N、R、L和T(例如,D);

[0436] X_{12}选自D、N和S;

[0437] X_{13}选自S、A、T、G和R(例如,S);

[0438] X_{14}选自I、L、F、S、R、Y、Q、W、D,K和H(例如,I、L和F);

[0439] X_{15}选自D、S、I、N、E、A、H、F、L、Q、M、G、Y和V;

[0440] X_{16}选自K、L、R、M、T和F(例如,L、R和K);

[0441] X_{17}选自V、L、I、A和T;

[0442] X_{18}选自L、I、V和A(例如,L和I);

[0443] X_{19}选自T、V、C、E、S和A(例如,T和V);

[0444] X_{20}选自R、F、T、W、E、L、N、C、K、V、S、Q、I、Y、H和A;

[0445] X_{21}选自S、P、R、K、N、A、H、Q、G和L;

[0446] X_{22}选自D、G、T、N、S、K、A、I、E、L、Q、R和Y;并且

[0447] X_{23}选自K、V、A、E、Y、I、C、L、S、T、G、K、M、D和F。

[0448] 在某些实施例中,HNH样结构域与SEQ ID NO:25的序列相差至少一个但不多于2、3、4、或5个残基。

[0449] 在某些实施例中,HNH样结构域是有切割能力的。在某些实施例中,HNH样结构域是无切割能力的。

[0450] 在某些实施例中,Cas9分子或Cas9多肽包含HNH样结构域,所述HNH样结构域包括具有化学式X的氨基酸序列:

[0451] X_1-X_2-X_3-H-X_4-X_5-P-X_6-S-X_8-X_9-X_{10}-D-D-S-X_{14}-X_{15}-N-K-V-L-X_{19}-X_{20}-X_{21}-X_{22}-X_{23}-N (SEQ ID NO:26),

[0452] 其中

[0453] X_1选自D和E;

[0454] X_2选自L、I、R、Q、V、M和K;

[0455] X_3选自D和E;

[0456] X_4选自I、V、T、A和L(例如,A、I和V);

[0457] X_5选自V、Y、I、L、F和W(例如,V、I和L);

[0458] X_6选自Q、H、R、K、Y、I、L、F和W;

[0459] X_8选自F、L、V、K、Y、M、I、R、A、E、D和Q(例如,F);

[0460] X_9选自L、R、T、I、V、S、C、Y、K、F和G;

[0461] X_{10}选自K、Q、Y、T、F、L、W、M、A、E、G和S;

[0462] X_{14}选自I、L、F、S、R、Y、Q、W、D,K和H(例如,I、L和F);

[0463] X_{15}选自D、S、I、N、E、A、H、F、L、Q、M、G、Y和V;

[0464] X_{19}选自T、V、C、E、S和A(例如,T和V);

[0465] X_{20}选自R、F、T、W、E、L、N、C、K、V、S、Q、I、Y、H和A;

[0466] X_{21}选自S、P、R、K、N、A、H、Q、G和L;

[0467] X_{22}选自D、G、T、N、S、K、A、I、E、L、Q、R和Y;并且

[0468] X_{23}选自K、V、A、E、Y、I、C、L、S、T、G、K、M、D和F。

[0469] 在某些实施例中,HNH样结构域与SEQ ID NO:26的序列相差1、2、3、4或5个残基。

[0470] 在某些实施例中,Cas9分子或Cas9多肽包含HNH样结构域,所述HNH样结构域包括具有化学式XI的氨基酸序列:

[0471] X_1-V-X_3-H-I-V-P-X_6-S-X_8-X_9-X_{10}-D-D-S-X_{14}-X_{15}-N-K-V-L-T-X_{20}-X_{21}-X_{22}-X_{23}-N (SEQ ID NO:27),

[0472] 其中

[0473] X_1选自D和E;

[0474] X_3选自D和E;

[0475] X_6选自Q、H、R、K、Y、I、L和W;

[0476] X_8选自F、L、V、K、Y、M、I、R、A、E、D和Q(例如,F);

[0477] X_9选自L、R、T、I、V、S、C、Y、K、F和G;

[0478] X_{10}选自K、Q、Y、T、F、L、W、M、A、E、G和S;

[0479] X_{14}选自I、L、F、S、R、Y、Q、W、D,K和H(例如,I、L和F);

[0480] X_{15}选自D、S、I、N、E、A、H、F、L、Q、M、G、Y和V;

[0481] X_{20}选自R、F、T、W、E、L、N、C、K、V、S、Q、I、Y、H和A;

[0482] X_{21}选自S、P、R、K、N、A、H、Q、G和L;

[0483] X_{22}选自D、G、T、N、S、K、A、I、E、L、Q、R和Y;并且

[0484] X_{23}选自K、V、A、E、Y、I、C、L、S、T、G、K、M、D和F。

[0485] 在某些实施例中,HNH样结构域与SEQ ID NO:27的序列相差1、2、3、4或5个残基。

[0486] 在某些实施例中,Cas9分子或Cas9多肽包含HNH样结构域,所述HNH样结构域具有化学式XII的氨基酸序列:

[0487] D-X_2-D-H-I-X_5-P-Q-X_7-F-X_9-X_{10}-D-X_{12}-S-I-D-N-X_{16}-V-L-X_{19}-X_{20}-S-X_{22}-X_{23}-N(SEQ ID NO:28),

[0488] 其中

[0489] X_2选自I和V;

[0490] X_5选自I和V;

[0491] X_7选自A和S;

[0492] X_9选自I和L;

[0493] X_{10}选自K和T;

[0494] X_{12}选自D和N;

[0495] X_{16}选自R、K和L;

[0496] X_{19}选自T和V;

[0497] X_{20}选自S和R;

[0498] X_{22}选自K、D和A;并且

[0499] X_{23}选自E、K、G和N(例如,Cas9分子或Cas9多肽可以包含如本文所描述的HNH样结构域)。

[0500] 在一个实施例中,HNH样结构域与SEQ ID NO:28的序列相差多达1个但不多于2、3、4或5个残基。

[0501] 在某些实施例中,Cas9分子或Cas9多肽包含具有化学式XIII的氨基酸序列:

[0502] L-Y-Y-L-Q-N-G-X_1'-D-M-Y-X_2'-X_3'-X_4'-X_5'-L-D-I-X_6'-X_7'-L-S-X_8'-Y-Z-N-R-X_9'-K-X_{10}'-D-X_{11}'-V-P(SEQ ID NO:24),

[0503] 其中

[0504] X_1'选自K和R;

[0505] X_2'选自V和T;

[0506] X_3'选自G和D;

[0507] X_4'选自E、Q和D;

[0508] X_5'选自E和D;

[0509] X_6'选自D、N和H;

[0510] X_7'选自Y、R和N;

[0511] X_8'选自Q、D和N;

[0512] X_9'选自G和E;

[0513] X_{10}'选自S和G;

[0514] X_{11}'选自D和N;并且

[0515] Z是HNH样结构域,例如,如以上所描述的。

[0516] 在某些实施例中,Cas9分子或Cas9多肽包含与SEQ ID NO:24的序列相差多达1个但不多于2、3、4或5个残基的氨基酸序列。

[0517] 在某些实施例中,HNH样结构域与本文(例如,在图5A-5C中)所披露的HNH样结构域的序列相差多达1个但不多于2、3、4或5个残基。在某些实施例中,在图5A-5C中识别出的高度保守残基中的1个或两者是存在的。

[0518] 在某些实施例中,HNH样结构域与本文(例如,在图6A-6B中)所披露的HNH样结构域的序列相差多达1个但不多于2、3、4或5个残基。在一个实施例中,在图6A-6B中识别出的高度保守残基中的1个、2个或所有3个是存在的。

[0519] Cas9活性

[0520] 在某些实施例中,Cas9分子或Cas9多肽能够切割靶核酸分子。典型地,野生型Cas9分子切割靶核酸分子的两条链。Cas9分子和Cas9多肽可以被工程化以改变核酸酶切割(或其他特性),例如以提供作为切口酶、或缺乏切割靶核酸能力的Cas9分子或Cas9多肽。能够切割靶核酸分子的Cas9分子或Cas9多肽在本文中称为eaCas9(酶促活性Cas9)分子或eaCas9多肽。

[0521] 在某些实施例中,eaCas9分子或eaCas9多肽包括以下酶活性中的一种或多种:

[0522] (1)切口酶活性,即切割核酸分子的单链(例如,非互补链或互补链)的能力;

[0523] (2)双链核酸酶活性,即切割双链核酸的两条链并且产生双链断裂的能力,其在一个实施例中是在两种切口酶活性的存在下;

[0524] (3)内切核酸酶活性;

[0525] (4)外切核酸酶活性;以及

[0526] (5)解旋酶活性,即解旋双链核酸的螺旋结构的能力。

[0527] 在某些实施例中,eaCas9分子或eaCas9多肽切割两条DNA链,并且产生双链断裂。在某些实施例中,eaCas9分子或eaCas9多肽仅切割一条链,例如,gRNA杂交到的链、或互补于与gRNA杂交的链的链。在一个实施例中,eaCas9分子或eaCas9多肽包含与HNH结构域相关的切割活性。在一个实施例中,eaCas9分子或eaCas9多肽包含与RuvC结构域相关的切割活性。在一个实施例中,eaCas9分子或eaCas9多肽包含与HNH结构域相关的切割活性以及与RuvC结构域相关的切割活性。在一个实施例中,eaCas9分子或eaCas9多肽包含有活性、或有切割能力的HNH结构域以及无活性、或无切割能力的RuvC结构域。在一个实施例中,eaCas9分子或eaCas9多肽包含无活性、或无切割能力的HNH结构域和有活性、或有切割能力的RuvC结构域。

[0528] 靶向和PAM

[0529] Cas9分子或Cas9多肽可以与gRNA分子相互作用并且与所述gRNA分子一起定位至包含靶结构域(并且在某些实施例中,是PAM序列)的位点。

[0530] 在某些实施例中,eaCas9分子或eaCas9多肽与靶核酸相互作用并且切割靶核酸的

能力是PAM序列依赖性的。PAM序列是在靶核酸中的序列。在一个实施例中,靶核酸的切割发生在PAM序列的上游。来自不同细菌物种的eaCas9分子可以识别不同序列基序(例如,PAM序列)。在一个实施例中,化脓链球菌的eaCas9分子识别序列基序NGG并且指导切割靶核酸序列的在该序列的上游的1至10(例如,3至5)bp(参见例如,Mali 2013)。在一个实施例中,嗜热链球菌的eaCas9分子识别序列基序NGGNG(SEQ ID NO:199)和/或NNAGAAW(W=A或T)(SEQ ID NO:200)并且指导切割靶核酸序列的在这些序列上游的1至10(例如,3至5)bp(参见例如,Horvath 2010;Deveau 2008)。在一个实施例中,变形链球菌的eaCas9分子识别序列基序NGG和/或NAAR(R=A或G)(SEQ ID NO:201)并且指导切割靶核酸序列的在该序列上游的1至10(例如,3至5)bp(参见例如,Deveau 2008)。在一个实施例中,金黄色葡萄球菌的eaCas9分子识别序列基序NNGRR(R=A或G)(SEQ ID NO:202)并且指导切割靶核酸序列的在该序列的上游的1至10(例如,3至5)bp。在一个实施例中,金黄色葡萄球菌的eaCas9分子识别序列基序NNGRRN(R=A或G)(SEQ ID NO:203)并且指导切割靶核酸序列的在该序列的上游的1至10(例如,3至5)bp。在一个实施例中,金黄色葡萄球菌的eaCas9分子识别序列基序NNGRRT(R=A或G)(SEQ ID NO:204)并且指导切割靶核酸序列的在该序列的上游的1至10(例如,3至5)bp。在一个实施例中,金黄色葡萄球菌的eaCas9分子识别序列基序NNGRRV(R=A或G,V=A、G或C)(SEQ ID NO:205)并且指导切割靶核酸序列的在该序列的上游的1至10(例如,3至5)bp。Cas9分子识别PAM序列的能力可以例如使用如先前所述的转化测定(Jinek 2012)来确定。在上述实施例(即SEQ ID NO:199-205)中的每一个中,N可以是任何核苷酸残基,例如A、G、C或T中的任一项。

[0531] 如本文所讨论的,Cas9分子可以被工程化以改变Cas9分子的PAM特异性。

[0532] 先前已经描述了示例性的天然存在的Cas9分子(参见例如,Chylinski 2013)。此类Cas9分子包括以下项的Cas9分子:簇1细菌家族、簇2细菌家族、簇3细菌家族、簇4细菌家族、簇5细菌家族、簇6细菌家族、簇7细菌家族、簇8细菌家族、簇9细菌家族、簇10细菌家族、簇11细菌家族、簇12细菌家族、簇13细菌家族、簇14细菌家族、簇15细菌家族、簇16细菌家族、簇17细菌家族、簇18细菌家族、簇19细菌家族、簇20细菌家族、簇21细菌家族、簇22细菌家族、簇23细菌家族、簇24细菌家族、簇25细菌家族、簇26细菌家族、簇27细菌家族、簇28细菌家族、簇29细菌家族、簇30细菌家族、簇31细菌家族、簇32细菌家族、簇33细菌家族、簇34细菌家族、簇35细菌家族、簇36细菌家族、簇37细菌家族、簇38细菌家族、簇39细菌家族、簇40细菌家族、簇41细菌家族、簇42细菌家族、簇43细菌家族、簇44细菌家族、簇45细菌家族、簇46细菌家族、簇47细菌家族、簇48细菌家族、簇49细菌家族、簇50细菌家族、簇51细菌家族、簇52细菌家族、簇53细菌家族、簇54细菌家族、簇55细菌家族、簇56细菌家族、簇57细菌家族、簇58细菌家族、簇59细菌家族、簇60细菌家族、簇61细菌家族、簇62细菌家族、簇63细菌家族、簇64细菌家族、簇65细菌家族、簇66细菌家族、簇67细菌家族、簇68细菌家族、簇69细菌家族、簇70细菌家族、簇71细菌家族、簇72细菌家族、簇73细菌家族、簇74细菌家族、簇75细菌家族、簇76细菌家族、簇77细菌家族、或簇78细菌家族。

[0533] 示例性的天然存在的Cas9分子包括簇1细菌家族的Cas9分子。实例包括以下项的Cas9分子:金黄色葡萄球菌、化脓链球菌(例如,菌株SF370、MGAS10270、MGAS10750、MGAS2096、MGAS315、MGAS5005、MGAS6180、MGAS9429、NZ131、SSI-1)、嗜热链球菌(例如,菌株LMD-9)、假豕链球菌(S.pseudoporcinus)(例如,菌株SPIN 20026)、变形链球菌(例如,菌株

UA159、NN2025)、猕猴链环菌(S.macacae)(例如,菌株NCTC11558)、解没食子酸链球菌(S.gallolyticus)(例如,菌株UCN34、ATCC BAA-2069)、马链球菌(S.equinus)(例如,菌株ATCC 9812、MGCS124)、停乳链球菌(S.dysdalactiae)(例如,菌株GGS124)、牛链球菌(S.bovis)(例如,菌株ATCC 700338)、咽峡炎链球菌(S.anginosus)(例如,菌株F0211)、无乳链球菌(S.agalactiae)(例如,菌株NEM316、A909)、单核细胞增生李斯特菌(Listeria monocytogenes)(例如,菌株F6854)、无害利斯特菌(Listeria innocua)(无害利斯特菌(L.innocua)(例如,菌株Clip11262))、意大利肠道球菌(Enterococcus italicus)(例如,菌株DSM 15952)、或屎肠球菌(Enterococcus faecium)(例如,菌株1,231,408)。

[0534] 在某些实施例中,Cas9分子或Cas9多肽包含以下氨基酸序列:

[0535] 与本文所述的任何Cas9分子序列或天然存在的Cas9分子序列(例如,来自本文列出的(例如,SEQ ID NO:1、2、4-6或12)或描述于Chylinski 2013中的物种的Cas9分子)具有60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%、或99%的同源性;

[0536] 当与其相比时,与其相差不多于2%、5%、10%、15%、20%、30%或40%的氨基酸残基;

[0537] 与其相差至少1、2、5、10或20个氨基酸但相差不多于100、80、70、60、50、40或30个氨基酸;或

[0538] 与其相同。在一个实施例中,Cas9分子或Cas9多肽包含以下一种或多种活性:切口酶活性;双链切割活性(例如,内切核酸酶和/或外切核酸酶活性);解旋酶活性;或连同gRNA分子定位至靶核酸的能力。

[0539] 在某些实施例中,Cas9分子或Cas9多肽包含图2A-2G的共有序列的任何氨基酸序列,其中“*”指示在化脓链球菌、嗜热链球菌、变形链球菌或无害利斯特菌的Cas9分子的氨基酸序列中的相应位置中发现的任何氨基酸,并且“-”指示不存在。在一个实施例中,Cas9分子或Cas9多肽与披露于图2A-2G中的共有序列的序列相差至少1个但不多于2、3、4、5、6、7、8、9或10个氨基酸残基。在某些实施例中,Cas9分子或Cas9多肽包含SEQ ID NO:2的氨基酸序列。在其他实施例中,Cas9分子或Cas9多肽与SEQ ID NO:2的序列相差至少1个但不多于2、3、4、5、6、7、8、9或10个氨基酸残基。

[0540] 多个Cas9分子的序列的比较表明某些区域是保守的。将这些进行识别如下:

[0541] 区域1(残基1至180,或在区域1’的情况下,残基120至180);

[0542] 区域2(残基360至480);

[0543] 区域3(残基660至720);

[0544] 区域4(残基817至900);以及

[0545] 区域5(残基900至960)。

[0546] 在某些实施例中,Cas9分子或Cas9多肽包含区域1-5,连同足够的另外的Cas9分子序列以提供生物活性分子(例如,具有至少一种本文所述的活性的Cas9分子)。在某些实施例中,区域1-5各自独立地与本文所述的Cas9分子或Cas9多肽(例如,来自图2A-2G的序列(SEQ ID NO:1、2、4、5、14))的相应残基具有50%、60%、70%、80%、85%、90%、95%、96%、97%、98%或99%的同源性。

[0547] 在某些实施例中,Cas9分子或Cas9多肽包含以下称为区域1的氨基酸序列:

[0548] 与化脓链球菌的Cas9的氨基酸序列(SEQ ID NO:2)的氨基酸1-180具有50%、

60％、70％、80％、85％、90％、95％、96％、97％、98％或99％的同源性(编号是根据图2中的基序序列;图2A-2G中的四个Cas9序列中的52％的残基是保守的)；

[0549] 与化脓链球菌、嗜热链球菌、变形链球菌或无害利斯特菌的Cas9的氨基酸序列(分别是SEQ ID NO:2、4、1和5)的氨基酸1-180相差至少1、2、5、10或20个氨基酸但相差不多于90、80、70、60、50、40或30个氨基酸;或

[0550] 与化脓链球菌、嗜热链球菌、变形链球菌或无害利斯特菌的Cas9的氨基酸序列(分别是SEQ ID NO:2、4、1和5)的氨基酸1-180相同。

[0551] 在某些实施例中,Cas9分子或Cas9多肽包含以下称为区域1'的氨基酸序列：

[0552] 与化脓链球菌、嗜热链球菌、变形链球菌或无害利斯特菌的Cas9的氨基酸序列(分别是SEQ ID NO:2、4、1和5)的氨基酸120-180具有55％、60％、65％、70％、75％、80％、85％、90％、95％、96％、97％、98％或99％的同源性(图2中的四个Cas9序列中的55％的残基是保守的)；

[0553] 与化脓链球菌、嗜热链球菌、变形链球菌或无害利斯特菌的Cas9的氨基酸序列(分别是SEQ ID NO:2、4、1和5)的氨基酸120-180相差至少1、2或5个氨基酸但相差不多于35、30、25、20或10个氨基酸;或

[0554] 与化脓链球菌、嗜热链球菌、变形链球菌或无害利斯特菌的Cas9的氨基酸序列(分别是SEQ ID NO:2、4、1和5)的氨基酸120-180相同。

[0555] 在某些实施例中,Cas9分子或Cas9多肽包含以下称为区域2的氨基酸序列：

[0556] 与化脓链球菌、嗜热链球菌、变形链球菌或无害利斯特菌的Cas9的氨基酸序列(分别是SEQ ID NO:2、4、1和5)的氨基酸360-480具有50％、55％、60％、65％、70％、75％、80％、85％、90％、95％、96％、97％、98％或99％的同源性(图2中的四个Cas9序列中的52％的残基是保守的)；

[0557] 与化脓链球菌、嗜热链球菌、变形链球菌或无害利斯特菌的Cas9的氨基酸序列(分别是SEQ ID NO:2、4、1和5)的氨基酸360-480相差至少1、2或5个氨基酸但相差不多于35、30、25、20或10个氨基酸;或

[0558] 与化脓链球菌、嗜热链球菌、变形链球菌或无害利斯特菌的Cas9的氨基酸序列(分别是SEQ ID NO:2、4、1和5)的氨基酸360-480相同。

[0559] 在某些实施例中,Cas9分子或Cas9多肽包含以下称为区域3的氨基酸序列：

[0560] 与化脓链球菌、嗜热链球菌、变形链球菌或无害利斯特菌的Cas9的氨基酸序列(分别是SEQ ID NO:2、4、1和5)的氨基酸660-720具有55％、60％、65％、70％、75％、80％、85％、90％、95％、96％、97％、98％或99％的同源性(图2中的四个Cas9序列中的56％的残基是保守的)；

[0561] 与化脓链球菌、嗜热链球菌、变形链球菌或无害利斯特菌的Cas9的氨基酸序列(分别是SEQ ID NO:2、4、1和5)的氨基酸660-720相差至少1、2或5个氨基酸但相差不多于35、30、25、20或10个氨基酸;或

[0562] 与化脓链球菌、嗜热链球菌、变形链球菌或无害利斯特菌的Cas9的氨基酸序列(分别是SEQ ID NO:2、4、1和5)的氨基酸660-720相同。

[0563] 在某些实施例中,Cas9分子或Cas9多肽包含以下称为区域4的氨基酸序列：

[0564] 与化脓链球菌、嗜热链球菌、变形链球菌或无害利斯特菌的Cas9的氨基酸序列(分

别是SEQ ID NO:2、4、1和5)的氨基酸817-900具有50％、55％、60％、65％、70％、75％、80％、85％、90％、95％、96％、97％、98％或99％的同源性(图2A-2G中的四个Cas9序列中的55％的残基是保守的)；

[0565] 与化脓链球菌、嗜热链球菌、变形链球菌或无害利斯特菌的Cas9的氨基酸序列(分别是SEQ ID NO:2、4、1和5)的氨基酸817-900相差至少1、2或5个氨基酸但相差不多于35、30、25、20或10个氨基酸;或

[0566] 与化脓链球菌、嗜热链球菌、变形链球菌或无害利斯特菌的Cas9的氨基酸序列(分别是SEQ ID NO:2、4、1和5)的氨基酸817-900相同。

[0567] 在某些实施例中,Cas9分子或Cas9多肽包含以下称为区域5的氨基酸序列:

[0568] 与化脓链球菌、嗜热链球菌、变形链球菌或无害利斯特菌的Cas9的氨基酸序列(分别是SEQ ID NO:2、4、1和5)的氨基酸900-960具有50％、55％、60％、65％、70％、75％、80％、85％、90％、95％、96％、97％、98％或99％的同源性(图2A-2G中的四个Cas9序列中的60％的残基是保守的)；

[0569] 与化脓链球菌、嗜热链球菌、变形链球菌或无害利斯特菌的Cas9的氨基酸序列(分别是SEQ ID NO:2、4、1和5)的氨基酸900-960相差至少1、2或5个氨基酸但相差不多于35、30、25、20或10个氨基酸;或

[0570] 与化脓链球菌、嗜热链球菌、变形链球菌或无害利斯特菌的Cas9的氨基酸序列(分别是SEQ ID NO:2、4、1和5)的氨基酸900-960相同。

[0571] 工程化的或改变的Cas9

[0572] 本文所述的Cas9分子和Cas9多肽(可以具有多种特性中的任一种,包括:核酸酶活性(例如,内切核酸酶和/或外切核酸酶活性);解旋酶活性;在功能上与gRNA分子相关联的能力;以及靶向(或定位至)核酸上的位点的能力(例如,PAM识别和特异性)。在某些实施例中,Cas9分子或Cas9多肽可以包括这些特性的全部或子集。在一个典型的实施例中,Cas9分子或Cas9多肽具有与gRNA分子相互作用,并且与所述gRNA分子一起定位至核酸中的位点的能力。其他活性(例如,PAM特异性、切割活性、或解旋酶活性)在Cas9分子和Cas9多肽中可以更广泛地变化。

[0573] Cas9分子包括工程化的Cas9分子和工程化的Cas9多肽(如在此背景下使用的,工程化的仅仅意指所述Cas9分子或Cas9多肽不同于参考序列,并且没有暗示过程或来源限制)。工程化的Cas9分子或Cas9多肽可以包含改变的酶特性,例如改变的核酸酶活性(与天然存在的或其他参考Cas9分子相比)或改变的解旋酶活性。如本文所讨论的,工程化的Cas9分子或Cas9多肽可以具有切口酶活性(与双链核酸酶活性相反)。在某些实施例中,工程化的Cas9分子或Cas9多肽可以具有改变其尺寸的改变,例如减小其尺寸的氨基酸序列缺失,例如对一种或多种Cas9活性没有显著影响。在某些实施例中,工程化的Cas9分子或Cas9多肽可以包含影响PAM识别的改变,例如,工程化的Cas9分子可被改变以识别不同于由内源性野生型PI结构域识别的PAM序列的PAM序列。在某些实施例中,Cas9分子或Cas9多肽在序列上可以不同于天然存在的Cas9分子,但是在一种或多种Cas9活性上没有显著改变。

[0574] 具有所希望特性的Cas9分子或Cas9多肽可以通过多种方式制成,例如,通过改变亲本(例如天然存在的)Cas9分子或Cas9多肽,以提供具有所希望特性的改变的Cas9分子或Cas9多肽。例如,可以相对于亲本Cas9分子(例如,天然存在的或工程化的Cas9分子)引入一

个或多个突变或差异。此类突变和差异包括:取代(例如,保守取代或非必需氨基酸的取代);插入;或缺失。在一个实施例中,Cas9分子或Cas9多肽相对于参考(例如,亲本)Cas9分子可以包含一个或多个突变或差异,例如,至少1、2、3、4、5、10、15、20、30、40或50个突变但少于200、100或80个突变。

[0575] 在某些实施例中,一个突变或多个突变对Cas9活性(例如,本文所述的Cas9活性)不具有实质影响。在其他实施例中,一个突变或多个突变对Cas9活性(例如,本文所述的Cas9活性)具有实质影响。

[0576] 非切割和经修饰的切割Cas9

[0577] 在一个实施例中,Cas9分子或Cas9多肽包含不同于天然存在的Cas9分子(例如,不同于具有最接近同源性的天然存在的Cas9分子)的切割特性。例如,Cas9分子或Cas9多肽可以与天然存在的Cas9分子(例如,化脓链球菌的Cas9分子)有如下区域别:例如,相比于天然存在的Cas9分子(例如,化脓链球菌的Cas9分子),它调控(例如,降低或增加)对双链核酸切割的能力(内切核酸酶和/或外切核酸酶活性);例如,相比于天然存在的Cas9分子(例如,化脓链球菌的Cas9分子),它调节(例如,降低或增加)对核酸的单链(例如,核酸分子的非互补链或核酸分子的互补链)切割的能力(切口酶活性);或切割核酸分子(例如,双链或单链核酸分子)的能力可以被消除。

[0578] 在某些实施例中,eaCas9分子或eaCas9多肽包含以下活性中的一种或多种:与N-末端RuvC-样结构域有关的切割活性;与HNH样结构域相关的切割活性;与HNH样结构域相关的切割活性和与N-末端RuvC样结构域相关的切割活性。

[0579] 在某些实施例中,eaCas9分子或eaCas9多肽包含有活性、或有切割能力的HNH样结构域(例如,本文所述的HNH样结构域,例如,SEQ ID NO:24-28)和无活性、或无切割能力的N-末端RuvC样结构域。示例性的无活性、或无切割能力的N-末端RuvC样结构域可以在N-末端RuvC样结构域中具有天冬氨酸的突变(例如,披露于图2A-2G中的共有序列的位置9处的天冬氨酸或SEQ ID NO:2的位置10处的天冬氨酸例如可以被丙氨酸取代)。在一个实施例中,eaCas9分子或eaCas9多肽与野生型的区别在于N-末端RuvC样结构域并且不切割靶核酸、或以显著低于参考Cas9分子的切割活性(例如,低于20%、10%、5%、1%或0.1%)的效率进行切割,如通过本文所述的测定所测量的。参比Cas9分子可以是天然存在的未经修饰的Cas9分子,例如,天然存在的Cas9分子,如化脓链球菌、金黄色葡萄球菌或嗜热链球菌的Cas9分子。在一个实施例中,参考Cas9分子是具有最接近序列一致性或同源性的天然存在的Cas9分子。

[0580] 在一个实施例中,eaCas9分子或eaCas9多肽包含无活性、或无切割能力的HNH结构域和有活性、或有切割能力的N-末端RuvC样结构域(例如,本文所述的N-末端RuvC样结构域,例如SEQ ID NO:15-23)。示例性的无活性、或无切割能力的HNH样结构域可以具有在以下一处或多处的突变:HNH样结构域中的组氨酸(例如,在披露于图2A-2G中的共有序列的位置856处所示的组氨酸例如可以被丙氨酸取代);以及HNH样结构域中的一个或多个天冬酰胺(例如,在披露于图2A-2G中的共有序列的位置870处和/或披露于图2A-2G中的共有序列的位置879处所示的天冬酰胺例如可以被丙氨酸取代)。在一个实施例中,eaCas9与野生型的区别在于HNH样结构域并且不切割靶核酸、或以显著低于参考Cas9分子的切割活性(例如,低于20%、10%、5%、1%或0.1%)的效率进行切割,如通过本文所述的测定所测量的。

参比Cas9分子可以是天然存在的未经修饰的Cas9分子,例如天然存在的Cas9分子,如化脓链球菌、金黄色葡萄球菌或嗜热链球菌的Cas9分子。在一个实施例中,参考Cas9分子是具有最接近序列一致性或同源性的天然存在的Cas9分子。

[0581] 在某些实施例中,示例性Cas9活性包括PAM特异性、切割活性、和解旋酶活性中的一项或多项。一个或多个突变可以存在于,例如:一个或多个RuvC样结构域(例如,N-末端RuvC样结构域);HNH结构域;RuvC结构域和HNH结构域之外的区域中。在一个实施例中,一个或多个突变存在于RuvC结构域中。在一个实施例中,一个或多个突变存在于HNH结构域中。在一个实施例中,突变存在于RuvC结构域和HNH结构域二者中。

[0582] 可以参照化脓链球菌Cas9序列在RuvC结构域或HNH结构域中进行的示例性突变包括:D10A、E762A、H840A、N854A、N863A和/或D986A。可参考金黄色葡萄球菌Cas9序列在RuvC结构域中进行的示例性突变包括N580A(参见例如,SEQ ID NO:11)。

[0583] 无论具体序列(例如,取代)是否可以影响一种或多种活性(如靶向活性、切割活性等),例如可以通过评价所述突变是否是保守的来评价或预测。在一个实施例中,“非必需”氨基酸残基,如在Cas9分子的背景下所使用的,是可以改变自Cas9分子的野生型序列(例如,天然存在的Cas9分子(例如,eaCas9分子))的残基,不会消除或更优选地不会实质上改变Cas9活性(例如,裂解活性),而改变“必需”氨基酸残基导致活性(例如,裂解活性)的实质性损失。

[0584] 在一个实施例中,Cas9分子包含不同于天然存在的Cas9分子(例如,不同于具有最接近同源性的天然存在的Cas9分子)的切割特性。例如,Cas9分子可以与天然存在的Cas9分子(例如,金黄色葡萄球菌或化脓链球菌的Cas9分子)有如下区域别:例如,相比于天然存在的Cas9分子(例如,金黄色葡萄球菌或化脓链球菌的Cas9分子),它调控(例如,降低或增加)对双链断裂切割的能力(内切核酸酶和/或外切核酸酶活性);例如,相比于天然存在的Cas9分子(例如,金黄色葡萄球菌或化脓链球菌的Cas9分子),它调控(例如,降低或增加)对核酸的单链(例如,核酸分子的非互补链或核酸分子的互补链)切割的能力(切口酶活性);或切割核酸分子(例如,双链或单链核酸分子)的能力可以被消除。在某些实施例中,切口酶是包含SEQ ID NO:10(D10A)或SEQ ID NO:11(N580A)(Friedland 2015)的序列的金黄色葡萄球菌Cas9衍生的切口酶。

[0585] 在一个实施例中,改变的Cas9分子是包含以下一种或多种活性的eaCas9分子:与RuvC结构域有关的切割活性;与HNH结构域相关的切割活性;与HNH结构域相关的切割活性和与RuvC结构域相关的切割活性。

[0586] 在某些实施例中,改变的Cas9分子或Cas9多肽包含如下序列,其中:

[0587] 对应于披露于图2A-2G中的共有序列的固定序列的序列与披露于图2A-2G中的共有序列中的固定残基的不多于1%、2%、3%、4%、5%、10%、15%或20%有区别;并且

[0588] 相应于披露于图2A-2G中的共有序列中由“*”识别的残基的序列与来自天然存在的Cas9分子(例如,化脓链球菌、嗜热链球菌、变形链球菌或无害利斯特菌Cas9分子)的相应序列的“*”残基的不多于1%、2%、3%、4%、5%、10%、15%、20%、25%、30%、35%或40%有区域别。

[0589] 在一个实施例中,改变的Cas9分子或Cas9多肽是包含披露于图2A-2G中的化脓链球菌Cas9的氨基酸序列(SEQ ID NO:2)的eaCas9分子或eaCas9多肽,其中在由披露于图2A-

2G中的共有序列(SEQ ID NO:14)中的“*”表示的一个或多个残基(例如,2、3、5、10、15、20、30、50、70、80、90、100、200个氨基酸残基)处具有不同于化脓链球菌的序列的一个或多个氨基酸(例如,取代)。

[0590] 在一个实施例中,改变的Cas9分子或Cas9多肽是包含披露于图2A-2G中的嗜热链球菌Cas9的氨基酸序列(SEQ ID NO:4)的eaCas9分子或eaCas9多肽,其中在由披露于图2A-2G中的共有序列(SEQ ID NO:14)中的“*”表示的一个或多个残基(例如,2、3、5、10、15、20、30、50、70、80、90、100、200个氨基酸残基)处具有不同于嗜热链球菌的序列的一个或多个氨基酸(例如,取代)。

[0591] 在一个实施例中,改变的Cas9分子或Cas9多肽是包含披露于图2A-2G中的变形链球菌Cas9的氨基酸序列(SEQ ID NO:1)的eaCas9分子或eaCas9多肽,其中在由披露于图2A-2G中的共有序列(SEQ ID NO:14)中的“*”表示的一个或多个残基(例如,2、3、5、10、15、20、30、50、70、80、90、100、200个氨基酸残基)处具有不同于变形链球菌的序列的一个或多个氨基酸(例如,取代)。

[0592] 在一个实施例中,改变的Cas9分子或Cas9多肽是包含披露于图2A-2G中的无害利斯特菌Cas9的氨基酸序列(SEQ ID NO:5)的eaCas9分子或eaCas9多肽,其中在由披露于图2A-2G中的共有序列(SEQ ID NO:14)中的“*”表示的一个或多个残基(例如,2、3、5、10、15、20、30、50、70、80、90、100、200个氨基酸残基)处具有不同于无害利斯特菌的序列的一个或多个氨基酸(例如,取代)。

[0593] 在某些实施例中,改变的Cas9分子或Cas9多肽(例如,eaCas9分子或eaCas9多肽)可以是例如多种不同Cas9分子(例如,不同物种的两种或更多种天然存在的Cas9分子)中的两种的融合体。例如,可以将一个物种的天然存在的Cas9分子的片段融合到第二物种的Cas9分子的片段上。作为实例,可以将包含N-末端RuvC样结构域的化脓链球菌的Cas9分子的片段融合到包含HNH样结构域的不同于化脓链球菌的物种(例如,嗜热链球菌)的Cas9分子的片段上。

[0594] 具有改变的PAM识别或无PAM识别的Cas9

[0595] 天然存在的Cas9分子可以识别特异性PAM序列,例如如上针对例如化脓链球菌、嗜热链球菌、变形链球菌和金黄色葡萄球菌描述的PAM识别序列。

[0596] 在某些实施例中,Cas9分子或Cas9多肽具有与天然存在的Cas9分子相同的PAM特异性。在其他实施例中,Cas9分子或Cas9多肽具有与天然存在的Cas9分子不相关的PAM特异性、或与它与之具有最接近序列同源性的天然存在的Cas9分子不相关的PAM特异性。例如,可以改变天然存在的Cas9分子,例如,以改变PAM识别,例如以改变Cas9分子或Cas9多肽识别的PAM序列以减少脱靶位点和/或改进特异性;或消除PAM识别需要的PAM序列。在某些实施例中,可以改变Cas9分子或Cas9多肽,例如以增加PAM识别序列的长度和/或提高Cas9对高水平一致性(例如,gRNA与PAM序列之间的98%、99%或100%匹配)的特异性,例如以减少脱靶位点和/或增加特异性。在某些实施例中,PAM识别序列的长度在长度上是至少4、5、6、7、8、9、10或15个氨基酸。在一个实施例中,Cas9特异性需要gRNA与PAM序列之间的至少90%、95%、96%、97%、98%、99%或更多的同源性。可以使用定向进化产生识别不同PAM序列和/或具有降低的脱靶活性的Cas9分子或Cas9多肽。描述了可以用于Cas9分子定向进化的示例性方法和系统(参见例如,Esvelt 2011)。可以例如通过下文所述的方法对候选Cas9

分子进行评价。

[0597] 尺寸优化的Cas9

[0598] 本文所描述的工程化的Cas9分子和工程化的Cas9多肽包括包含减小分子的尺寸但仍保留所希望Cas9特性(例如,基本上天然的构象、Cas9核酸酶活性、和/或靶核酸分子识别)的缺失的Cas9分子或Cas9多肽。本文提供了包含一个或多个缺失和任选地一个或多个接头的Cas9分子或Cas9多肽,其中接头被布置于在所述缺失的侧翼的氨基酸残基之间。用于识别参考Cas9分子中的适合缺失的方法、用于产生具有缺失和接头的Cas9分子的方法、以及使用此类Cas9分子的方法在审查本文献后对于本领域的普通技术人员应是清楚的。

[0599] 具有缺失的Cas9分子(例如,金黄色葡萄球菌或化脓链球菌Cas9分子)比相应的天然存在的Cas9分子小,例如,具有减少数量的氨基酸。Cas9分子的较小尺寸允许提高递送方法的灵活性,并且由此增加基因组编辑的实用性。Cas9分子可以包含一个或多个不会实质上影响或降低本文所描述的所得Cas9分子的活性的缺失。包含如本文所描述的缺失的Cas9分子中所保留的活性包括以下项中的一种或多种:

[0600] 切口酶活性,即切割核酸分子的单链(例如,非互补链或互补链)的能力;双链核酸酶活性,即切割双链核酸的两条链并且产生双链断裂的能力,其在一个实施例中是在两种切口酶活性的存在下;

[0601] 内切核酸酶活性;

[0602] 外切核酸酶活性;

[0603] 解旋酶活性,即,解旋双链核酸的螺旋结构的能力;

[0604] 以及核酸分子(例如,靶核酸或gRNA)的识别活性。

[0605] 可以使用本文所述的或本领域的活性测定来评估本文所述的Cas9分子的活性。

[0606] 识别适于缺失的区域

[0607] 可以通过多种方法识别Cas9分子的适于缺失的区域。可以在化脓链球菌Cas9的晶体结构上建模来自不同细菌物种的天然存在的直向同源Cas9分子(例如,在表1中所列出的那些中的任一种)(Nishimasu 2014),以便相对于所述蛋白的三维构象跨所选的Cas9直向同源物检查保守水平。在空间定位上远离在Cas9活性中所涉及的区域(例如,与靶核酸分子和/或gRNA相互作用)的较不保守的或不保守的区域代表作为用于缺失而不实质上影响或降低Cas9活性的候选物的区域或结构域。

[0608] 编码Cas9分子的核酸

[0609] 本文提供了编码Cas9分子或Cas9多肽(例如,eaCas9分子或eaCas9多肽)的核酸。先前已经描述了编码Cas9分子或Cas9多肽的示例性核酸(参见例如,Cong2013;Wang 2013;Mali 2013;Jinek 2012)。

[0610] 在一个实施例中,编码Cas9分子或Cas9多肽的核酸可以是合成核酸序列。例如,合成核酸分子可以进行化学修饰,例如如本文所述。在一个实施例中,Cas9 mRNA具有以下一种或多种(例如,所有)特性:它被5-甲基胞苷和/或假尿苷加帽、聚腺苷酸化、取代。

[0611] 另外或可替代地,可以对合成核酸序列进行密码子优化,例如至少一个非常见密码子或低不常见密码子已经被常见密码子取代。例如,合成的核酸可以指导优化的信使mRNA的合成(例如,针对在哺乳动物表达系统(例如,本文描述的)中的表达进行优化)。

[0612] 另外或可替代地,编码Cas9分子或Cas9多肽的核酸可以包含核定位序列(NLS)。核

定位序列在本领域是已知的。

[0613] 编码化脓链球菌的Cas9分子的示例性密码子优化的核酸序列示于SEQ ID NO:3中。化脓链球菌Cas9分子的相应氨基酸序列示于SEQ ID NO:2中。

[0614] 编码金黄色葡萄球菌的Cas9分子的示例性密码子优化的核酸序列示于SEQ ID NO:7-9中。金黄色葡萄球菌Cas9分子的氨基酸序列示于SEQ ID NO:6中。

[0615] 如果任何上述Cas9序列与肽或多肽在C-末端处融合,则应理解的是终止密码子将被去除。

[0616] 其他Cas分子和Cas多肽

[0617] 不同类型的Cas分子或Cas多肽可以用来实践本文所披露的发明。在一些实施例中,使用II型Cas系统的Cas分子。在其他实施例中,使用其他Cas系统的Cas分子。例如,可以使用I型或III型Cas分子。先前已经描述了示例性Cas分子(和Cas系统)(参见例如,Haft 2005和Makarova 2011)。示例性Cas分子(和Cas系统)也示于表2中。

[0618] 表2:Cas系统

[0619]

基因名称‡	系统类型或亚型	来自 Haft 2005§的名称	所编码蛋白的结构(PDB 登录)¶	所编码蛋白的家族（和超家族）#**	代表
cas1	• I 型 • II 型 • III 型	*cas1*	3GOD、3LFX 和 2YZS	COG1518	SERP2463、SPy1047 和 *ygbT*
cas2	• I 型 • II 型 • III 型	*cas2*	2IVY、2I8E 和 3EXC	COG1343 和 COG3512	SERP2462、SPy1048、SPy1723(N-末端结构域）和 *ygbF*
cas3'	• I 型‡‡	*cas3*	NA	COG1203	APE1232 和 *ygcB*
cas3''	• 亚型 I-A • 亚型 I-B	NA	NA	COG2254	APE1231 和 BH0336
cas4	• 亚型 I-A • 亚型 I-B • 亚型 I-C • 亚型 I-D • 亚型 II-B	*cas4* 和 *csa1*	NA	COG1468	APE1239 和 BH0340
cas5	• 亚型 I-A • 亚型 I-B • 亚型 I-C • 亚型 I-E	*cas5a*、*cas5d*、*cas5e*、*cas5h*、*cas5p*、*cas5t* 和 *cmx5*	3KG4	COG1688（RAMP）	APE1234、BH0337、*devS* 和 *ygcI*
cas6	• 亚型 I-A • 亚型 I-B • 亚型 I-D • 亚型 III-A• 亚型 III-B	*cas6* 和 *cmx6*	3I4H	COG1583 和 COG5551（RAMP）	PF1131 和 slr7014
cas6e	• 亚型 I-E	*cse3*	1WJ9	（RAMP）	*ygcH*
cas6f	• 亚型 I-F	*csy4*	2XLJ	（RAMP）	y1727
cas7	• 亚型 I-A	*csa2*、*csd2*、	NA	COG1857	*devR* 和 *ygcJ*

基因名称‡	系统类型或亚型	来自 Haft 2005§的名称	所编码蛋白的结构(PDB 登录)¶	所编码蛋白的家族（和超家族）#**	代表
	• 亚型 I-B • 亚型 I-C • 亚型 I-E	*cse4*、*csh2*、*csp1* 和 *cst2*		和 COG3649（RAMP）	
cas8a1	• 亚型 I-A‡‡	*cmx1*、*cst1*、*csx8*、*csx13* 和 CXXC-CXXC	NA	BH0338 样	LA3191§§和 PG2018§§
cas8a2	• 亚型 I-A‡‡	*csa4* 和 *csx9*	NA	PH0918	AF0070、AF1873、MJ0385、PF0637、PH0918 和 SSO1401
[0620] *cas8b*	• 亚型 I-B‡‡	*csh1* 和 TM1802	NA	BH0338 样	MTH1090 和 TM1802
cas8c	• 亚型 I-C‡‡	*csd1* 和 *csp2*	NA	BH0338 样	BH0338
cas9	• II 型‡‡	*csn1* 和 *csx12*	NA	COG3513	FTN_0757 和 SPy1046
cas10	• III 型‡‡	*cmr2*、*csm1* 和 *csx11*	NA	COG1353	MTH326、Rv2823c§§ 和 TM1794§§
cas10d	• 亚型 I-D‡‡	*csc3*	NA	COG1353	slr7011
csy1	• 亚型 I-F‡‡	*csy1*	NA	y1724 样	y1724
csy2	• 亚型 I-F	*csy2*	NA	（RAMP）	y1725
csy3	• 亚型 I-F	*csy3*	NA	（RAMP）	y1726
cse1	• 亚型 I-E‡‡	*cse1*	NA	YgcL 样	*ygcL*
cse2	• 亚型 I-E	*cse2*	2ZCA	YgcK 样	*ygcK*
csc1	• 亚型 I-D	*csc1*	NA	alr1563 样（RAMP）	alr1563

[0621]

基因名称‡	系统类型或亚型	来自 Haft 2005§的名称	所编码蛋白的结构(PDB 登录)¶	所编码蛋白的家族（和超家族）#**	代表
csc2	• 亚型 I-D	*csc1* 和 *csc2*	NA	COG1337（RAMP）	slr7012
csa5	• 亚型 I-A	*csa5*	NA	AF1870	AF1870、MJ0380、PF0643 和 SSO1398
csn2	• 亚型 II-A	*csn2*	NA	SPy1049 样	SPy1049
csm2	• 亚型 III-A‡‡	*csm2*	NA	COG1421	MTH1081 和 SERP2460
csm3	• 亚型 III-A	*csc2* 和 *csm3*	NA	COG1337（RAMP）	MTH1080 和 SERP2459
csm4	• 亚型 III-A	*csm4*	NA	COG1567（RAMP）	MTH1079 和 SERP2458
csm5	• 亚型 III-A	*csm5*	NA	COG1332（RAMP）	MTH1078 和 SERP2457
csm6	• 亚型 III-A	APE2256 和 *csm6*	2WTE	COG1517	APE2256 和 SSO1445
cmr1	• 亚型 III-B	*cmr1*	NA	COG1367（RAMP）	PF1130
cmr3	• 亚型 III-B	*cmr3*	NA	COG1769（RAMP）	PF1128
cmr4	• 亚型 III-B	*cmr4*	NA	COG1336（RAMP）	PF1126
cmr5	• 亚型 III-B‡‡	*cmr5*	2ZOP 和 2OEB	COG3337	MTH324 和 PF1125
cmr6	• 亚型 III-B	*cmr6*	NA	COG1604（RAMP）	PF1124
csb1	• 亚型 I-U	GSU0053	NA	（RAMP）	Balac_1306 和 GSU0053

[0622]

基因名称‡	系统类型或亚型	来自 Haft 2005§的名称	所编码蛋白的结构(PDB登录)¶	所编码蛋白的家族(和超家族)#**	代表
csb2	• 亚型 I-U§§	NA	NA	(RAMP)	Balac_1305 和 GSU0054
csb3	• 亚型 I-U	NA	NA	(RAMP)	Balac_1303§§
csx17	• 亚型 I-U	NA	NA	NA	Btus_2683
csx14	• 亚型 I-U	NA	NA	NA	GSU0052
csx10	• 亚型 I-U	*csx10*	NA	(RAMP)	Caur_2274
csx16	• 亚型 III-U	VVA1548	NA	NA	VVA1548
csaX	• 亚型 III-U	*csaX*	NA	NA	SSO1438
csx3	• 亚型 III-U	*csx3*	NA	NA	AF1864
csx1	• 亚型 III-U	*csa3*、*csx1*、*csx2*、DXTHG、NE0113 和 TIGR02710	1XMX 和 2I71	COG1517 和 COG4006	MJ1666、NE0113、PF1127 和 TM1812
csx15	• 未知	NA	NA	TTE2665	TTE2665
csf1	• U 型	*csf1*	NA	NA	AFE_1038
csf2	• U 型	*csf2*	NA	(RAMP)	AFE_1039
csf3	• U 型	*csf3*	NA	(RAMP)	AFE_1040
csf4	• U 型	*csf4*	NA	NA	AFE_1037

[0623] Cpf1分子

[0624] 与crRNA复合的氨基酸球菌属物种(Acidaminococcus sp.)Cpf1和包括TTTN PAM序列的双链(ds)DNA靶向的晶体结构已由Yamano 2016解决,通过引用结合在此。与Cas9一样,Cpf1具有两个叶片:REC(识别)叶片和NUC(核酸酶)叶片。REC叶片包括REC1和REC2结构域,其与任何已知的蛋白质结构缺乏相似性。同时,NUC叶片包括三个RuvC结构域(RuvC-I、-II和-III)和BH结构域。然而,与Cas9相反,Cpf1 REC叶片缺乏HNH结构域,并且包括与已知蛋白质结构缺乏相似性的其他结构域:结构上独特的PI结构域/三个Wedge(WED)结构域(WED-I、-II和-III)、和核酸酶(Nuc)结构域。

[0625] 虽然Cas9和Cpf1在结构和功能上具有相似性,但应当理解,某些Cpf1活性是由与任何Cas9结构域不类似的结构域介导的。例如,靶DNA的互补链的切割似乎由Nuc结构域介导,其在顺序上和空间上与Cas9的HNH结构域不同。另外,Cpf1 gRNA的非靶向部分(柄)采用假结结构,而不是由Cas9 gRNA中的重复:抗重复双链体形成的茎环结构。

[0626] RNA指导的核酸酶的修饰

[0627] 上述RNA指导的核酸酶具有可用于多种应用的活性和特性,但技术人员将理解,

RNA指导的核酸酶也可在某些情况下被修饰,以改变切割活性,PAM特异性或其他结构或功能特征。

[0628] 首先转向改变切割活性的修饰,上面已经描述了降低或消除NUC叶片内结构域活性的突变。可以在RuvC结构域、Cas9 HNH结构域、或Cpf1 Nuc结构域中进行的示例性突变描述于Ran 2013和Yamano 2016,以及Cotta-Ramusino 2016中。通常,降低或消除两个核酸酶结构域之一中活性的突变导致具有切口酶活性的RNA指导的核酸酶,但应注意,切口酶活性的类型根据哪个结构域失活而变化。作为一个实例,Cas9的RuvC结构域的失活将导致切割互补链或顶链的切口酶,如下示出(其中C表示切割位点):

[0629] 5′----------[原型间隔子]--[C]----------3′

[0630] 3′----------5’

[0631] 另一方面,Cas9 HNH结构域的失活导致切割底部或非互补链的切口酶:

[0632] 5’----------[原型间隔子]----------3’

[0633] 3’----------[C]----------5’

[0634] Kleinstiver 2015a已经描述了对于化脓链球菌和金黄色葡萄球菌(Kleinstiver 2015b)的相对于天然存在的Cas9参考分子的PAM特异性的修饰。Kleinstiver等人还描述了改进Cas9的靶向保真性的修改(Kleinstiver 2016)。这些参考中的每一篇都通过引用结合在此。

[0635] RNA指导的核酸酶已被分成两个或更多个部分,如Zetsche 2015(通过引用结合)和Fine 2015(通过引用结合)所述。

[0636] 在某些实施例中,RNA指导的核酸酶可以是尺寸优化的或截短的,例如经由一种或多种缺失,其减少核酸酶的尺寸,同时仍保留gRNA关联、靶向和PAM识别以及切割活性。在某些实施例中,RNA指导的核酸酶任选地通过接头与另一种多肽、核苷酸或其他结构共价或非共价结合。示例性结合的核酸酶和接头由Guilinger 2014描述,其通过引用结合在此用于所有目的。

[0637] RNA指导的核酸酶还任选地包括标签,例如,但不限于核定位信号,以促进RNA指导的核酸酶蛋白质迁移入细胞核。在某些实施例中,RNA指导的核酸酶可以掺入C-和/或N-末端核定位信号。核定位序列是本领域已知的并且在Maeder 2015和其他地方描述。

[0638] 前述修改列表本质上是示例性的,并且鉴于本披露内容,技术人员将理解,在某些应用中可能或希望其他修饰。因此,为简洁起见,参考特定的RNA指导的核酸酶表示本披露内容的示例性系统、方法和组合物,但应理解所使用的RNA指导的核酸酶可以以不改变其操作原理的方式进行修饰。这些修饰在本披露内容的范围内。

[0639] <u>编码RNA指导的核酸酶的核酸</u>

[0640] 本文提供了编码RNA指导的核酸酶的核酸,例如Cas9、Cpf1或其功能片段。先前已经描述了编码RNA指导的核酸酶的示例性核酸(参见例如,Cong 2013;Wang 2013;Mali 2013;Jinek 2012)。

[0641] 在一些情况下,编码RNA指导的核酸酶的核酸可以是合成的核酸序列。例如,合成核酸分子可以进行化学修饰。在某些实施例中,编码RNA指导的核酸酶的mRNA将具有一种或多种(例如,所有)以下特性:它可以被5-甲基胞苷和/或假尿苷加帽、聚腺苷酸化、取代。

[0642] 合成核酸序列也可以是密码子优化的,例如,至少一个非常见密码子或较不常见

的密码子已被共同密码子替换。例如,合成的核酸可以指导优化的信使mRNA的合成(例如,针对在哺乳动物表达系统(例如,本文描述的)中的表达进行优化)。密码子优化的Cas9编码序列的实例存在于Cotta-Ramusino 2016中。

[0643] 另外,或可替代地,编码RNA指导的核酸可以包含核定位序列(NLS)。核定位序列在本领域是已知的。

[0644] 候选分子的功能分析

[0645] 可以通过本领域已知或如在此描述的方法来评价候选Cas9分子、候选gRNA分子,候选Cas9分子/gRNA分子复合物。例如,先前已经描述了用于评价Cas9分子的内切核酸酶活性的示例性方法(Jinek 2012)。

[0646] 结合和切割测定:测试Cas9分子的核酸内切酶活性

[0647] 可以在质粒切割测定中对Cas9分子/gRNA分子复合物结合至并且切割靶核酸的能力进行评价。在这个测定中,在反应之前通过加热至95℃并且缓慢冷却至室温,将合成或体外转录的gRNA分子预退火。在37℃,将天然或限制酶切消化-线性化的质粒DNA(300ng(约8nM))用纯化的Cas9蛋白分子(50nM-500nM)和gRNA(50nM-500nM,1:1)在具有或不具有10mM $MgCl_2$的Cas9质粒切割缓冲液(20mM HEPES pH 7.5、150mM KCl、0.5mM DTT、0.1mM EDTA)中孵育60分钟。用5X DNA加样缓冲液(30%甘油、1.2% SDS、250mM EDTA)终止反应,通过0.8%或1%琼脂糖凝胶电泳进行解析并且通过溴化乙锭染色进行可视化。所得切割产物指示Cas9分子是否切割两条DNA链、或仅切割两条链中的一条。例如,线性DNA产物指示两条DNA链的切割,而有缺口的开放圆形产物指示两条链中只有一条被切割。

[0648] 可替代地,可以在寡核苷酸DNA切割测定中对Cas9分子/gRNA分子复合物结合至并且切割靶核酸的能力进行评价。在这个测定中,在37℃,在50μL反应中,通过用在1X T4多核苷酸激酶反应缓冲液中的5单位T4多核苷酸激酶以及约3-6pmol(约20mCi-40mCi)[γ-32P]-ATP孵育30分钟,对DNA寡核苷酸(10pmol)进行放射性标记。在热灭活后(65℃持续20min),通过柱对反应进行纯化以去除未结合的标记。通过在95℃用等摩尔量的未标记的互补寡核苷酸退火标记的寡核苷酸持续3分钟,随后缓慢冷却至室温而生成双链体底物(100nM)。对于切割测定,通过加热至95℃持续30秒,随后缓慢冷却至室温对gRNA分子进行退火。在9μL的总体积中,将Cas9(500nM终浓度)与退火的gRNA分子(500nM)在切割测定缓冲液(20mM HEPES pH 7.5、100mM KCl、5mM MgCl2、1mM DTT、5%甘油)中进行预孵育。通过添加1μL靶DNA(10nM)开始反应并在37℃孵育1小时。将反应通过添加20μL的加样染料(5mM EDTA、0.025% SDS、5%甘油,在甲酰胺中)淬灭并加热至95℃持续5分钟。将切割产物在含有7M尿素的12%变性聚丙烯酰胺凝胶上进行解析,并且通过磷成像进行可视化。所得切割产物指示互补链、非互补链、或两者是否被切割。

[0649] 这些测定中的一个或两个可以用于评价候选gRNA分子或候选Cas9分子的适合性。

[0650] 结合测定:测试Cas9分子与靶DNA的结合

[0651] 先前已经描述了用于评价Cas9分子与靶DNA的结合的示例性方法(Jinek 2012)。

[0652] 例如,在电泳迁移率变动测定中,通过在去离子水中混合每条链(10nmol),加热至95℃持续3分钟并且缓慢冷却至室温而形成靶DNA双链体。将所有DNA在含有1X TBE的8%非变性凝胶上进行纯化。将DNA条带通过UV遮蔽进行可视化,切除,并且通过将凝胶片浸泡在DEPC处理的H_2O中进行洗脱。将洗脱的DNA进行乙醇沉淀并且溶解在DEPC处理的H_2O中。在37

℃,使用T4多核苷酸激酶将DNA样品用[γ-^{32}P]-ATP进行5′端标记持续30分钟。将多核苷酸激酶在65℃热变性持续20分钟,并且使用柱去除未结合的放射性标记。在10μL的总体积中,在含有20mM HEPES pH 7.5、100mM KCl、5mM $MgCl_2$、1mM DTT以及10%甘油的缓冲液中进行结合测定。用等摩尔量的预退火的gRNA分子对Cas9蛋白分子进行程序化,并且从100pM滴定至1μM。将放射性标记的DNA添加至20pM的终浓度。将样品在37℃孵育1小时并且在4℃下在含有1X TBE和5mM $MgCl_2$的8%天然聚丙烯酰胺凝胶上进行解析。将凝胶干燥并且通过感光成像进行DNA可视化。

[0653] 差示扫描荧光测定法(DSF)

[0654] 可以经由DSF测量Cas9-gRNA核糖核蛋白(RNP)复合物的热稳定性。这种技术测量蛋白质的热稳定性,其可以在有利条件下(如添加结合型RNA分子,例如,gRNA)增加。

[0655] 可以使用两种不同的方案进行测定,一种方案用于测试gRNA:Cas9蛋白的最佳化学计量比,并且另一种方案用于确定RNP形成的最佳溶液条件。

[0656] 为了确定形成RNP复合物的最好的的溶液条件,将Cas9的2μM溶液置于水与10x SYPRO Orange®(生命技术公司(Life Techonologies)目录#S-6650)中并分配到384孔板中。然后添加稀释于溶液中的具有不同pH和盐的等摩尔量的gRNA。在室温,孵育10分钟并短暂离心以去除任何气泡之后,使用带有Bio-Rad CFX Manager软件的Bio-Rad CFX384™Real-Time System C1000 Touch™热循环仪运行从20℃至90℃的梯度,其中温度每10秒增加1℃。

[0657] 第二个测定由在来自上述测定1的缓冲液中混合不同浓度的gRNA分子与2μM Cas9并在384孔板中于室温孵育10分钟组成。添加等体积的最适缓冲液和10x SYPRO Orange®(生命技术公司目录#S-6650)并且将板用Microseal®B粘合剂(MSB-1001)密封。短暂离心以去除任何气泡后,使用带有Bio-Rad CFX Manager软件的Bio-Rad CFX384™Real-Time System C1000 Touch™热循环仪运行从20℃至90℃的梯度,其中温度每10秒增加1℃。

[0658] 用于基因靶向的NHEJ方法

[0659] 在本文提供的方法的某些实施例中,NHEJ介导的缺失用于缺失γ-珠蛋白基因(例如HBG1,HBG2)负调控元件(例如,沉默子)的全部或部分。如本文所述,核酸酶诱导的NHEJ可用于以靶向特异性方式敲除全部或部分调控元件。在其他实施例中,NHEJ介导的插入用于将序列插入γ-珠蛋白基因负调控元件中,导致调控元件的失活。

[0660] 虽然不希望受理论束缚,但认为在某些实施例中,与本文所描述的方法相关的基因组改变依赖于核酸酶诱导的NHEJ以及NHEJ修复途径的易错性质。NHEJ通过将两端连接在一起修复DNA中的双链断裂;然而,通常,只有两个相容末端(完全如他们通过双链断裂所形成的)是完全连接的,原始序列才被恢复。在末端重新连接之前,双链断裂的DNA末端常常是酶加工的受试者,在一条或两条链处产生核苷酸的添加或去除。这使得NHEJ修复位点处的DNA序列中存在插入和/或缺失(indel)突变。典型地,这些突变中的三分之二改变阅读框并且因此产生非功能蛋白。另外,维持阅读框但插入或缺失大量的序列的突变可以破坏蛋白质的功能性。这是基因座依赖性的,因为关键功能结构域中的突变可能比蛋白质的非关键区域中的突变耐受性低。

[0661] 由NHEJ产生的indel突变在性质上是不可预测的;然而,在给定的断裂位点处,某些indel序列是有利的并且在群体中过度表达,这可能归因于微同源性的小区域。缺失的长

度可以广泛变化;它们最常见地在1bp-50bp范围内,但是可以达到大于100bp-200bp。插入往往是较短的并且常常包括紧密围绕断裂位点的序列的短的重复。然而,有可能获得大插入,并且在这些情况中,插入序列通常已经被追溯至基因组的其他区域或至存在于细胞中的质粒DNA。

[0662] 因为NHEJ是诱变的过程,所以它还可以用于缺失小序列基序(例如,长度上小于或等于50个核苷酸的基序),只要不需要产生特异性最终序列。如果双链断裂被靶向靶序列附近,则由NHEJ修复导致的缺失突变常常跨越并且因此去除不想要的核苷酸。对于较大的DNA区段的缺失,引入两个双链断裂(序列的每侧上一个双链断裂)可以在末端之间产生NHEJ,其中去除了整个间插序列。以这种方式,可以缺失大至几百千碱基的DNA区段。这两种方法都可以用于缺失特异性DNA序列;然而,NHEJ的易出错性质仍可能在修复位点产生indel突变。

[0663] 双链切割性eaCas9分子和单链、或切口酶,eaCas9分子均可以用于本文所描述的方法和组合物以产生NHEJ介导的indel。NHEJ介导的indel靶向感兴趣的调控区域可用于破坏或缺失靶向调控元件。

[0664] 双链或单链断裂相对于靶位置的布置

[0665] 在某些实施例中,为了诱导NHEJ介导的indel的目的,其中gRNA和Cas9核酸酶生成双链断裂,gRNA(例如,单分子(或嵌合)或模块化gRNA分子)被配置成将一个双链断裂定位在极接近于靶位置的核苷酸之处。在一个实施例中,切割位点在远离靶位置0-30bp之间处(例如,离靶位置小于30bp、25bp、20bp、15bp、10bp、9bp、8bp、7bp、6bp、5bp、4bp、3bp、2bp或1bp)。

[0666] 在某些实施例中,为了诱导NHEJ介导的indel的目的,其中与Cas9切口酶复合的两个gRNA诱导两个单链断裂,两个gRNA(例如,独立地为单分子(或嵌合)或模块化gRNA)被配置成将两个单链断裂定位以提供NHEJ修复靶位置的核苷酸。在某些实施例中,所述gRNA被配置成将切口定位在相同位置、或彼此相距在几个核苷酸之内、在不同链上,实质上模拟了双链断裂。在某些实施例中,较近的切口在远离靶位置0-30bp(例如,离靶位置小于30bp、25bp、20bp、15bp、10bp、9bp、8bp、7bp、6bp、5bp、4bp、3bp、2bp或1bp)之间处,并且这两个切口彼此在25bp-55bp之内(例如,在25bp至50bp、25bp至45bp、25bp至40bp、25bp至35bp、25bp至30bp、50bp至55bp、45bp至55bp、40bp至55bp、35bp至55bp、30bp至55bp、30bp至50bp、35bp至50bp、40bp至50bp、45bp至50bp、35bp至45bp、或40bp至45bp之间)并且彼此远离不超过100bp(例如,不超过90bp、80bp、70bp、60bp、50bp、40bp、30bp、20bp或10bp)。在某些实施例中,所述gRNA被配置成将单链断裂布置在靶位置的核苷酸的任一侧上。

[0667] 双链切割性eaCas9分子和单链、或切口酶,eaCas9分子均可以用于在此所描述的方法和组合物中以在靶位置的两侧均产生断裂。可以在靶位置的两侧产生双链的或成对的单链断裂,以去除两个切口之间的核酸序列(例如,两个断裂之间的区域被缺失)。在某些实施例中,两个gRNA(例如,独立地为单分子(或嵌合)或模块化gRNA)被配置成将双链断裂定位在靶位置的两侧。在其他实施例中,三个gRNA(例如,独立地为单分子(或嵌合)或模块化gRNA)被配置成将双链断裂(即,一个gRNA与cas9核酸酶复合)和两个单链断裂或成对单链断裂(即,两个gRNA与Cas9切口酶复合)定位在靶位置的任一侧。在又其他实施例中,四个gRNA(例如,独立地为单分子(或嵌合)或模块化gRNA)被配置成在靶位置的任一侧产生两对

单链断裂(即,两对的两个gRNA与Cas9切口酶复合物)。理想地,一个或多个双链断裂或成对的两个单链切口中更近者将在靶位置的0-500bp之内(例如,离靶位置不超过450bp、400bp、350bp、300bp、250bp、200bp、150bp、100bp、50bp或25bp)。当使用切口酶时,成对的两个切口彼此在25bp-55bp(例如,在25bp至50bp、25bp至45bp、25bp至40bp、25bp至35bp、25bp至30bp、50bp至55bp、45bp至55bp、40bp至55bp、35bp至55bp、30bp至55bp、30bp至50bp、35bp至50bp、40bp至50bp、45bp至50bp、35bp至45bp、或40bp至45bp之间)之内并且彼此远离不超过100bp(例如、不超过90bp、80bp、70bp、60bp、50bp、40bp、30bp、20bp或10bp)。

[0668] HDR修复、HDR介导的敲入、敲除、或缺失,以及模板核酸

[0669] 在本文提供的方法的某些实施例中,HDR介导的序列改变用于使用外源提供的模板核酸(在本文中也称为供体构建体)改变(例如,缺失、破坏或修饰)γ-珠蛋白基因(例如,HBG1,HBG2)调控区域中的一个或多个核苷酸的序列。虽然希望不受理论束缚,但据信HDR-介导的γ-珠蛋白基因调控区域内HBG靶位置的改变通过HDR与外源提供的供体模板或模板核酸发生。例如,供体构建体或模板核酸提供了HBG靶位置的改变。在此考虑了可以将质粒供体用作用于同源重组的模版。在此进一步考虑了可以通过在靶序列和供体模板之间的HDR(例如,单链退火)的替代方法,将单链供体模板用作用于改变HBG靶位置的模板。通过供体模板实现的HBG靶位置的改变取决于Cas9分子的切割。经由Cas9切割可以包括双链断裂或两个单链断裂。

[0670] 在本文提供的方法的某些实施例中,HDR介导的改变用于敲除或缺失γ-珠蛋白基因(例如HBG1,HBG2)负调控元件(例如,沉默子)的全部或部分。如本文所述,HDR可用于以靶向特异性方式敲除或缺失全部或部分调控元件。

[0671] 在其他实施例中,HDR介导的序列改变用于不使用外源提供的模板核酸改变γ-珠蛋白基因(例如,HBG1,HBG2)调控区域中的一个或多个核苷酸的序列。虽然不希望受理论束缚,但认为HBG靶位置的改变通过用内源基因组供体序列进行HDR而发生。例如,内源性基因组供体序列提供了HBG靶位置的改变。考虑了在一个实施例中,内源基因组供体序列位于与靶序列相同的染色体上。进一步考虑了,在其他的实施例中,内源基因组供体序列位于与靶序列不同的染色体上。通过内源基因组供体序列改变HBG靶位置取决于Cas9分子的切割。经由Cas9切割可以包括双链断裂或两个单链断裂。

[0672] 在本文提供的方法的某些实施例中,HDR介导的改变用于改变γ-珠蛋白基因调控区域中的单个核苷酸。这些实施例可以利用一个双链断裂或两个单链断裂。在某些实施例中,单个核苷酸改变可以通过以下项来掺入:(1)一个双链断裂,(2)两个单链断裂,(3)两个双链断裂,其中断裂发生在靶位置的每一侧,(4)一个双链断裂和两个单链断裂,其中双链断裂和两个单链断裂发生在靶位置的每一侧,(5)四个单链断裂,其中一对单链断裂发生在靶位置的每一侧,或(6)一个单链断裂。

[0673] 在使用单链模板核酸的某些实施例中,靶位置可以通过替代的HDR来改变。

[0674] 在本文提供的方法的某些实施例中,HDR介导的改变用于引入γ-珠蛋白基因调控区域中的一个或多个核苷酸的改变(例如,缺失)。在某些实施例中,γ-珠蛋白基因调控区域可以是HBG靶位置。在某些实施例中,可以在HBG靶位置内的靶位点处引入改变(例如,缺失)。在某些实施例中,改变(例如,缺失)可选自HBG1 13bp del c.-114至-102、HBG1 4bp del c.-225至-222、和HBG1 13bp del c.-114至-102中的一种或多种。在某些实施例中,靶

位点可选自HBG1 c.-114至-102(例如,SEQ ID NO:902(HBG1)的核苷酸2824-2836)、HBG1 c.-225至-222(例如,SEQ ID NO:902(HBG1)的核苷酸2716-2719)、和HBG2 c.-114至-102(例如,SEQ ID NO:903(HBG2)的核苷酸2748-2760)中的一种或多种。

[0675] 通过供体模板实现的HBG靶位置的改变取决于Cas9分子的切割。经由Cas9切割可以包括切口、双链断裂或两个单链断裂(例如靶核酸的每条链上的一个断裂)。在靶核酸上引入断裂之后,在断裂端发生切除,产生单链的突出DNA区域。

[0676] 在典型HDR中,引入包含与靶核酸同源的序列的双链供体模板,其将会直接掺入靶核酸中或用作模板以改变靶核酸序列。在断裂处切除后,修复可以通过不同的途径进行,例如通过双霍利迪连接模型(或双链断裂修复(DSBR)途径)或合成依赖性链退火(SDSA)途径。在双霍利迪连接模型中,发生由靶核酸的两个单链突出端进行链入侵到供体模板中的同源序列,导致形成具有两个霍利迪连结的中间体。当从入侵链的末端合成新的DNA以填充由切除产生的空位时,结点迁移。将新合成的DNA的末端连接到切除的末端,并且结点被分解,导致靶核酸的改变,例如将供体模板的HPFH突变体序列掺入相应的HBG靶位置。与供体模板的交叉可能会在结点分解时发生。在SDSA途径中,只有一个单链突出端侵入供体模板,并从入侵链的末端合成新的DNA以填补由切除产生的空位。然后新合成的DNA与剩余的单链突出端退火,合成新的DNA以填补空位,并将链连接以产生改变的DNA双链体。

[0677] 在可替代的HDR中,引入单链供体模板,例如模板核酸。用于改变所希望的HBG靶位置的靶核酸处的切口、单链断裂或双链断裂由例如本文所述的Cas9分子介导,并且在断裂处发生切除以显示单链突出端。掺入模板核酸序列以校正或改变HBG靶位置典型地通过如上所述的SDSA途径而发生。

[0678] 在国际申请PCT/US2014/057905中的标题为“模板核酸”的第IV节中提供了关于模板核酸的另外的细节。

[0679] 在某些实施例中,双链切割是由具有与HNH样结构域有关的切割活性以及与RuvC样结构域(例如,N端RuvC样结构域)有关的切割活性的Cas9分子(例如,野生型Cas9)实现的。这样的实施例仅需要单一的gRNA。

[0680] 在某些实施例中,一个单链断裂或切口由具有切口酶活性的Cas9分子例如,本文所述的Cas9切口酶实现。带切口的靶核酸可以是alt-HDR的底物。

[0681] 在其他实施例中,两个单链断裂或切口是由具有切口酶活性(例如,与HNH样结构域有关的切割活性或与N端RuvC样结构域有关的切割活性)的Cas9分子实现的。这样的实施例通常需要两个gRNA,各用于布置每个单链断裂。在实施例中,具有切口酶活性的Cas9分子切割gRNA所杂交的链,但不是互补于与gRNA杂交的链的链。在实施例中,具有切口酶活性的Cas9分子不切割gRNA所杂交的链,而是切割互补于与gRNA杂交的链的链。

[0682] 在某些实施例中,所述切口酶具有HNH活性,例如,具有RuvC活性失活的Cas9分子(例如,在D10处具有突变(例如,D10A突变)的Cas9分子)(参见,例如,SEQ ID NO:10)。D10A使RuvC失活;因此,Cas9切口酶(仅)具有HNH活性并且将切割gRNA所杂交的链(例如,互补链,其上并不具有NGG PAM)。在其他的实施例中,具有H840(例如,H840A)突变的Cas9分子可以用作切口酶。H840A使HNH失活;因此,Cas9切口酶(仅)具有RuvC活性并且切割非互补链(例如,具有NGG PAM并且其序列与gRNA相同的链)。在其他的实施例中,具有N863突变(例如,N863A)突变的Cas9分子可以用作切口酶。N863A灭活HNH,因此Cas9切口酶(仅)具有RuvC

活性并且切割非互补链(具有NGG PAM并且其序列与gRNA相同的链)。

[0683] 在某些实施例中,其中将一种切口酶和两种gRNA用于定位两个单链切口,一个切口在靶核酸的+链上而一个切口在-链上。PAM可以面向外部。可以选择gRNA,这样使得通过从约0-50、0-100、或0-200个核苷酸中分离gRNA。在实施例中,互补于两个gRNA的靶向结构域的靶序列之间没有重叠。在实施例中,这些gRNA不重叠并且是通过多至50、100、或200个核苷酸分离的。在一个实施例中,使用两个gRNA可以增加特异性,例如,通过降低脱靶结合(Ran 2013)。

[0684] 在某些实施例中,单一切口可以用于诱导HDR,例如alt-HDR。在此考虑了可以将单一切口用于增加HR与NHEJ在给定切割位点的比率。在一个实施例中,单链断裂形成于所述靶核酸的与所述gRNA的靶向结构域互补的链中。在其他实施例中,单链断裂形成于所述靶核酸的不同于与所述gRNA的靶向结构域互补的链的链中。

[0685] 双链或单链断裂相对于靶位置的布置

[0686] 其中一条链中的双链断裂或单链断裂应足够接近HBG靶位置,即在所希望区域产生改变,例如,掺入HPFH突变。在某些实施例中,所述距离不超过HBG靶位置的50、100、200、300、350或400个核苷酸。虽然希望不受理论束缚,在某些实施例中,认为断裂应当足够接近于HBG靶位置,以使得靶位置在末端切除过程中在经受外切核酸酶介导的去除的区域内。如果HBG靶位置与断裂之间的距离太大,则所希望被改变的序列可能不包括在末端切除中,并且因此可能不会被改变为供体序列,外源提供的供体序列或内源基因组供体序列,在一些实施例中,仅用于在末端切除区域内改变序列。

[0687] 在某些实施例中,本文所述的方法在HGB1和/或HGB2基因的γ-珠蛋白基因调控区域(例如,增强子区域,例如,沉默子区域,例如,启动子区域)附近引入一个或多个断裂。在某些这些实施例中,引入位于调控区域的至少一部分的侧翼的两个或更多个断裂,例如,HGB1和/或HGB2基因的增强子区域,例如,HGB1和/或HGB2基因的沉默子区域。两个或更多个断裂去除(例如,缺失)包括至少一部分γ-珠蛋白基因调控区域的基因组序列,例如,HGB1和/或HGB2基因的增强子区域,例如,HGB1和/或HGB2基因的沉默子区域。本文描述的所有方法导致调控区域的改变,例如,HGB1和/或HGB2基因的增强子区域,例如,HGB1和/或HGB2基因的沉默子区域。

[0688] 在某些实施例中,靶向结构域被配置为使得切割事件(例如双链或单链断裂)位于所希望被改变(例如,突变)的区域的1、2、3、4、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、150或200个核苷酸中。断裂,例如双链或单链断裂可以位于所希望被改变(例如,突变)的区域的上游或下游。在一些实施例中,断裂位于所希望被改变的区域内,例如,位于由至少两个突变体核苷酸限定的区域内。在一些实施例中,断裂被定位在紧邻所希望被改变的区域,例如,直接位于突变的上游或下游。

[0689] 在某些实施例中,单链断裂伴随有由第二gRNA分子位置的另外的单链断裂,如下文所讨论的。例如,靶向结构域被配置成,使得切割事件(例如两条单链断裂)位于HBG靶位置的1、2、3、4、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、150或200个核苷酸中。在一个实施例中,所述第一和第二gRNA分子被配置成使得在指导Cas9切口酶时,单链断裂将伴随有由第二gRNA定位的彼此足够接近的另外的单链断裂,以引起希望的区域的改变。在一个实施例中,所述第一和第二gRNA分子被配置成使得例如,当Cas9是切口酶时,由所述

第二gRNA定位的单链断裂在由所述第一gRNA分子定位的断裂的10、20、30、40、或50个核苷酸内。在一个实施例中,所述两个gRNA分子被配置成将切口定位在相同位置、或彼此相距在几个核苷酸之内、在不同链上,例如,基本上模拟了双链断裂。

[0690] 在某些实施例中,为了诱导HDR介导的序列改变的目的,其中gRNA(单分子(或嵌合)或模块化gRNA)和Cas9核酸酶诱导双链断裂,切割位点在远离HBG靶位置0-200bp(例如,0至175、0至150、0至125、0至100、0至75、0至50、0至25、25至200、25至175、25至150、25至125、25至100、25至75、25至50、50至200、50至175、50至150、50至125、50至100、50至75、75至200、75至175、75至150、75至125、75至100bp)处。在某些实施例中,切割位点在远离HBG靶位置0-100bp(例如,0至75、0至50、0至25、25至100、25至75、25至50、50至100、50至75或75至100bp)处。

[0691] 在HBG靶位置实施例中,可以通过使用切口酶来产生与突出端的断裂来促进HDR。虽然不希望受理论束缚,突出端的单链本质可以通过HDR增强细胞修复破坏的可能性,与例如,NHEJ相反。具体而言,在一些实施例中,通过选择将第一切口酶靶向第一靶序列的第一gRNA和将第二切口酶靶向第二靶序列的第二gRNA来促进HDR,所述第二靶序列位于与第一靶序列相对的DNA链上并且从第一个切口处偏移。

[0692] 在某些实施例中,gRNA分子的靶向结构域被配置成将切割事件定位为离预选择的核苷酸足够远,所述核苷酸不被改变。在某些实施例中,gRNA分子的靶向结构域被配置成将内含子切割事件定位为离内含子/外显子边界、或天然存在的剪接信号足够远,以避免所述外显子序列的改变或不想要的剪接事件。所述gRNA分子可以是如在此所描述的第一、第二、第三和/或第四gRNA分子。

[0693] 第一断裂和第二断裂的彼此相对布置

[0694] 在某些实施例中,双链断裂可以伴随有由第二gRNA分子定位的另外的双链断裂,如下文所讨论的。

[0695] 在某些实施例中,双链断裂可以伴随有由第二gRNA分子和第三gRNA分子定位的两个另外的单链断裂。

[0696] 在某些实施例中,第一和第二单链断裂可以伴随有由第三gRNA分子和第四gRNA分子定位的两个另外的单链断裂。

[0697] 当两种或更多种gRNA用于在靶核酸中定位两个或更多个切割事件(例如,双链或单链断裂)时,考虑了所述两个或更多个切割事件可以由相同或不同Cas9蛋白产生。例如,当两个gRNA用于定位两个双链断裂时,单Cas9核酸酶可以用于产生两个双链断裂。当两种或更多种gRNA用于定位两个或更多个单链断裂(切口)时,单Cas9切口酶可以用于产生所述两个或更多个切口。当两种或更多种gRNA用于定位至少一个双链断裂和至少一个单链断裂时,可以使用两种Cas9蛋白,例如,一种Cas9核酸酶和一种Cas9切口酶。考虑了当使用两种或更多种Cas9蛋白时,可以顺序地递送所述两种或更多种Cas9蛋白,以控制靶核酸中的希望位置处的双链断裂对比单链断裂的特异性。

[0698] 在一些实施例中,所述第一gRNA分子的靶向结构域和所述第二gRNA分子的靶向结构域与靶核酸分子的相对链互补。在一些实施例中,所述gRNA分子和所述第二gRNA分子被配置成使得PAM朝外定向。

[0699] 在某些实施例中,选择两个gRNA以指导在距离彼此预选距离的两个位置处的Cas9

介导的切割。在某些实施例中,两个切割点位于靶核酸的相反链上。在一些实施例中,两个切割点形成平端的断裂,并且在其他实施例中,他们被偏移以使得DNA末端包含一个或两个突出端(例如,一个或多个5′突出端和/或一个或多个3′突出端)。在一些实施例中,每个断裂事件是一个切口。在一个实施例中,切口足够接近以致他们形成由双链断裂机器识别的断裂(而不是被例如,SSBr机器识别)。在某些实施例中,切口足够分开,使得他们产生作为HDR底物的突出端,即,断裂的布置模仿经历了一些切除的DNA底物。例如,在一些实施例中,切口被间隔开以产生作为用于进行性切除的底物的突出端。在一些实施例中,两个断裂彼此间隔着25-65个核苷酸。两个断裂可以是例如,彼此的约25、30、35、40、45、50、55、60或65个核苷酸。两个断裂可以是例如,彼此的至少约25、30、35、40、45、50、55、60或65个核苷酸。两个断裂可以是例如,彼此的至多约30、35、40、45、50、55、60或65个核苷酸。在某些实施例中,两个断裂是彼此的约25-30、30-35、35-40、40-45、45-50、50-55、55-60、或60-65个核苷酸。

[0700] 在一些实施例中,模拟切除的断裂的断裂包括3’突出端(例如,由DSB和切口产生,其中切口留下了3’突出端)、5’突出端(例如,由DSB和缺口产生,其中切口留下了5′突出端)、3’和5’突出端(例如,由三个切割产生)、两个3’突出端(例如,由彼此偏移的两个切口产生)或两个5’突出端(例如,由彼此偏移的两个切口产生)。

[0701] 在某些实施例中,为了诱导HDR介导的改变的目的,其中与Cas9切口酶复合的两个gRNA(独立地为单分子(或嵌合)或模块化gRNA)诱导两个单链断裂,较近的切口在远离HBG靶位置0-200bp(例如,0至175、0至150、0至125、0至100、0至75、0至50、0至25、25至200、25至175、25至150、25至125、25至100、25至75、25至50、50至200、50至175、50至150、50至125、50至100、50至75、75至200、75至175、75至150、75至125、或75至100bp)之间处,并且理想地这两个切口彼此将在25-65bp之内(例如,25至50、25至45、25至40、25至35、25至30、30至55、30至50、30至45、30至40、30至35、35至55、35至50、35至45、35至40、40至55、40至50、40至45bp、45至50bp、50至55bp、55至60bp、或60至65bp)并且彼此远离不超过100bp(例如,彼此远离不超过90、80、70、60、50、40、30、20、10或5bp)。在某些实施例中,切割位点在远离HBG靶位置0-100bp(例如,0至75、0至50、0至25、25至100、25至75、25至50、50至100、50至75或75至100bp)之间处。

[0702] 在一些实施例中,两个gRNA(例如,独立地为单分子(或嵌合)或模块化gRNA)被配置成将双链断裂定位在靶位置的两侧。在其他实施例中,三个gRNA(例如,独立地为单分子(或嵌合)或模块化gRNA)被配置成将双链断裂(即,一个gRNA与cas9核酸酶复合)和两个单链断裂或成对单链断裂(即,两个gRNA与Cas9切口酶复合)定位在靶位置的任一侧。在其他实施例中,四个gRNA(例如,独立地为单分子(或嵌合)或模块化gRNA)被配置成在靶位置的任一侧产生两对单链断裂(即,两对的两个gRNA与Cas9切口酶复合物)。理想地,一个或多个双链断裂或成对的两个单链切口中更近者将在HBG靶位置的0-500bp之内(例如,离靶位置不超过450、400、350、300、250、200、150、100、50或25bp)。当使用切口酶时,成对的两个切口在某些实施例中彼此在25-65bp(例如,在25-55、25至50、25至45、25至40、25至35、25至30、50至55、45至55、40至55、35至55、30至55、30至50、35至50、40至50、45至50、35至45、40至45bp、45至50bp、50至55bp、55至60bp、或60至65bp之间)之内并且彼此远离不超过100bp(例如,不超过90、80、70、60、50、40、30、或20或10bp)。

[0703] 当使用两个gRNA靶向Cas9分子至断裂时,可以设想Cas9分子的不同组合。在一些实施例中,使用第一gRNA将第一Cas9分子靶向第一靶位置,并且使用第二gRNA将第二Cas9分子靶向第二靶位置。在一些实施例中,第一Cas9分子在靶核酸的第一条链上产生切口,并且第二Cas9分子在相对链上产生切口,导致双链断裂(例如,平端的切割或突出端的切割)。

[0704] 可以选择切口酶的不同组合来将一条单链断裂靶向一条链,并且将第二条单链断裂靶向相反链。当选择组合时,可以考虑到具有一个活性RuvC样结构域的切口酶和具有一个活性HNH结构域的切口酶。在某些实施例中,RuvC样结构域切割靶核酸分子的非互补链。在某些实施例中,HNH样结构域切割双链核酸分子的单链互补结构域(例如,互补链)。一般而言,如果两个Cas9分子具有相同的活性结构域(例如,两者均具有活性RuvC结构域或两者都具有活性HNH结构域),则将选择两个与靶的相反链结合的gRNA。更详细地,在一些实施例中,第一gRNA与靶核酸的第一条链互补并结合具有活性RuvC样结构域的切口酶并引起切口酶切割与第一gRNA不互补的链,即,靶核酸的第二条链;并且第二gRNA与靶核酸的第二条链互补并结合具有活性RuvC样结构域的切口酶并引起切口酶切割与第二gRNA不互补的链,即,靶核酸的第一条链。相反地,在一些实施例中,第一gRNA与靶核酸的第一条链互补并结合具有活性HNH结构域的切口酶并引起切口酶切割与第一gRNA互补的链,即,靶核酸的第一条链;并且第二gRNA与靶核酸的第二条链互补并结合具有活性HNH结构域的切口酶并引起切口酶切割与第二gRNA互补的链,即,靶核酸的第二条链。在另一种排列中,如果一个Cas9分子具有活性RuvC样结构域并且另一个Cas9分子具有活性HNH结构域,则两个Cas9分子的gRNA可以与靶核酸的相同链互补,使得具有活性的RuvC样结构域的Cas9分子将切割非互补链并且具有HNH结构域的Cas9分子将切割互补链,导致双链断裂。

[0705] 供体模板的同源臂

[0706] 同源臂应该至少延伸远至可发生末端切除的区域,例如,以便允许所切除的单链突出端在供体模板内找到互补区域。总长度可由如质粒大小或病毒包装限制之类的参数进行限制。在一个实施例中,同源臂不延伸到重复元件(例如,Alu重复或LINE重复)中。

[0707] 示例性同源臂长度包括至少50、100、250、500、750、1000、2000、3000、4000、或5000个核苷酸。在一些实施例中,同源臂长度是50-100、100-250、250-500、500-750、750-1000、1000-2000、2000-3000、3000-4000、或4000-5000个核苷酸。

[0708] 模板核酸,如该术语在此使用的,是指可以与Cas9分子和gRNA分子结合用于改变(例如,缺失、破坏、或修饰)HBG靶位置的结构的核酸序列。在某些实施例中,HBG靶位置可以是在其中添加进一个或多个核苷酸的靶核酸上两个核苷酸(例如,相邻的核苷酸)之间的位点。可替代地,HBG靶位置可以包含一个或多个通过模板核酸改变的核苷酸。在某些实施例中,可以在HBG靶位置内的靶位点处引入改变(例如,缺失)。在某些实施例中,改变(例如,缺失)可选自HBG1 13bp del c.-114至-102、HBG1 4bp del c.-225至-222、和HBG1 13bp del c.-114至-102中的一种或多种。在某些实施例中,靶位点可选自HBG1 c.-114至-102(例如,SEQ ID NO:902(HBG1)的核苷酸2824-2836)、HBG1 c.-225至-222(例如,SEQ ID NO:902(HBG1)的核苷酸2716-2719)、和HBG2 c.-114至-102(例如,SEQ ID NO:903(HBG2)的核苷酸2748-2760)中的一种或多种。

[0709] 在某些实施例中,对靶核酸进行修饰以具有模板核酸的一些或全部序列,典型地在一个或多个切割位点处或附近。在某些实施例中,模板核酸是单链的。在其他实施例中,

模板核酸是双链的。在某些实施例中,模板核酸是DNA(例如,双链DNA)。在其他实施例中,模板核酸是单链DNA。在一个实施例中,模板核酸如Cas9和gRNA在相同的载体骨架例如,AAV基因组,质粒DNA上编码。在某些实施例中,模板核酸在体内从载体骨架上切除,例如,其侧翼为gRNA识别序列。在某些实施例中,模板核酸包含内源基因组序列。

[0710] 在某些实施例中,模板核酸通过参与HDR事件来改变靶位置的结构。在某些实施例中,模板核酸改变靶位置的序列。在某些实施例中,模板核酸导致修饰的或非天然存在的碱基掺入靶核酸中。

[0711] 在某些实施例中,模板核酸导致靶核酸的一个或多个核苷酸的缺失。在某些实施例中,模板核酸导致HBG靶位置的一个或多个核苷酸的缺失。在某些实施例中,可以在HBG靶位置内的靶位点处引入改变(例如,缺失)。在某些实施例中,改变(例如,缺失)可选自HBG1 13bp del c.-114至-102、HBG1 4bp del c.-225至-222、和HBG113bp del c.-114至-102中的一种或多种。在某些实施例中,靶位点可选自HBG1 c.-114至-102(例如,SEQ ID NO:902(HBG1)的核苷酸2824-2836)、HBG1 c.-225至-222(例如,SEQ ID NO:902(HBG1)的核苷酸2716-2719)、和HBG2 c.-114至-102(例如,SEQ ID NO:903(HBG2)的核苷酸2748-2760)中的一种或多种。

[0712] 典型地,模板序列经历断裂介导或催化的与靶序列的重组。在某些实施例中,模板核酸包括与通过eaCas9介导的切割事件切割的靶序列上的位点相相应的序列。在某些实施例中,模板核酸包括与在第一Cas9介导的事件中被切割的靶序列上的第一位点、以及在第二Cas9介导的事件中被切割的靶序列上的第二位点二者相相应的序列。

[0713] 与γ-珠蛋白基因调控区域中的HBG靶位置具有同源性的模板核酸可用于改变调控区域的结构。例如,与γ-珠蛋白基因调控区域中HBG靶位置的5'和3′区域具有同源性的模板核酸可用于缺失HBG靶位置的一个或多个核苷酸。

[0714] 模板核酸典型地包括以下组分:

[0715] [5'同源臂]-[替换序列]-[3'同源臂]。

[0716] 同源臂提供用于重组到染色体中,因此用替换序列替换了不希望的元件(例如,突变或标签)。同源臂是与待切割的靶核酸内或附近(例如,侧翼或邻接)的DNA区域同源的区域。在某些实施例中,同源臂侧翼于最远端切割位点。

[0717] 在某些实施例中,模板核酸可用于去除(例如,缺失)包括至少一部分γ-珠蛋白基因调控区域的基因组序列,例如,HGB1和/或HGB2基因的增强子区域,例如,HGB1和/或HGB2基因的沉默子区域。在某些实施例中,模板核酸可用于缺失HBG靶位置的一个或多个核苷酸,即,将改变(例如,缺失)引入HBG靶位置。在某些实施例中,可以在HBG靶位置内的靶位点处引入改变(例如,缺失)。在某些实施例中,改变(例如,缺失)可选自HBG1 13bp del c.-114至-102、HBG1 4bp del c.-225至-222、和HBG1 13bp del c.-114至-102中的一种或多种。在某些实施例中,靶位点可选自HBG1 c.-114至-102(例如,SEQ ID NO:902(HBG1)的核苷酸2824-2836)、HBG1 c.-225至-222(例如,SEQ ID NO:902(HBG1)的核苷酸2716-2719)、和HBG2c.-114至-102(例如,SEQ ID NO:903(HBG2)的核苷酸2748-2760)中的一种或多种。

[0718] 供体模板中的替换序列已在别处描述,包括在Cotta-Ramusino 2016中,其通过引用结合在此。替换序列可以是任何合适的长度。在某些实施例中,替换序列可包括1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20或相对于希望编辑的细胞内的天然存在

的序列进行更多的序列修饰。

[0719] 在某些实施例中,当希望的修复结果是靶核酸的缺失时,替换序列可以是0个核苷酸或0bp。在某些实施例中,模板核酸省略了与待缺失的靶核酸序列同源的序列。如果替换序列是0个核苷酸或0bp,那么位于5′同源臂和3′同源臂与模板核酸退火的位置之间的靶核酸序列将被缺失。

[0720] 在某些实施例中,5′同源臂的3′端是紧邻替换序列的5′端的位置。在某些实施例中,5′同源臂可以从替换序列的5′端延伸至少10、20、30、40、50、100、200、300、400、500、600、700、800、900、1000、1500、2000、3000、4000或5000个核苷酸5′。在某些实施例中,当替换序列是0个核苷酸或0bp时,5'同源臂的3'端是紧邻3′同源臂的5'端的位置。在某些实施例中,当替换序列是0个核苷酸或0bp时,5'同源臂可以从3'同源臂的5'端延伸至少10、20、30、40、50、100、200、300、400、500、600、700、800、900、1000、1500、2000、3000、4000或5000个核苷酸5′。

[0721] 在某些实施例中,3′同源臂的5′端是紧邻替换序列的3′端的位置。在一个实施例中,3' 同源臂可以从替换序列的3' 端延伸至少10、20、30、40、50、100、200、300、400、500、600、700、800、900、1000、1500、2000、3000、4000或5000个核苷酸3' 。在某些实施例中,当替换序列是0个核苷酸或0bp时,3'同源臂的5′端是紧邻5'同源臂的3'端的位置。在一个实施例中,3′同源臂可以从5′同源臂的3′端延伸至少10、20、30、40、50、100、200、300、400、500、600、700、800、900、1000、1500、2000、3000、4000或5000个核苷酸3'。

[0722] 在某些实施例中,为了改变HBG靶位置处的一个或多个核苷酸,同源臂(例如5'和3'同源臂)可各自包含位于最远端gRNA侧翼的约1000bp序列(例如,HBG靶位置任一侧的1000bp序列)。

[0723] 在此考虑了可以将一个或两个同源臂缩短以避免包括某些序列重复元件(例如,Alu重复或LINE元件)。例如,可以将5' 同源臂缩短以避免序列重复元件。在其他的实施例中,可以将3' 同源臂缩短以避免序列重复元件。在一些实施例中,可以将5' 和3' 同源臂均缩短以避免包括某些序列重复元件。

[0724] 本文中考虑了可以将用于改变HBG靶位置的序列设计成用作单链寡核苷酸,例如单链,寡脱氧核苷酸(ssODN)。当使用ssODN时,5'和3′同源臂在长度上范围可以高达约200个核苷酸(例如,在长度上至少25、50、75、100、125、150、175、或200bp)。针对ssODN还考虑了更长的同源臂作为继续有待完成的寡核苷酸合成中的改进。在一些实施例中,通过化学合成以外的方法制备较长的同源臂,例如,通过变性长双链核酸并纯化一条链,例如,通过对锚定于固体底物的链特异性序列的亲和力。

[0725] 虽然不希望受理论束缚,但在某些实施例中,当模板核酸与切口(即,带切口的链的5'方向)或靶位点(即,靶位点的5'方向)具有延伸的同源性5′时alt-HDR更有效地进行。相应地,在一些实施例中,模板核酸具有较长的同源臂和较短的同源臂,其中较长的同源臂可以使切口或靶位点的5'退火。在一些实施例中,可以与切口或靶位点5′退火的臂是距离切口或靶位点或替换序列的5'或3'端至少25、50、75、100、125、150、175、或200、300、400、500、600、700、800、900、1000、1500、2000、3000、4000、或5000个核苷酸。在一些实施例中,可以与切口或靶位点5'退火的臂比可以使切口或靶位点3'退火的臂长至少10%、20%、30%、40%、或50%。在一些实施例中,可以与切口或靶位点5'退火的臂比可以使切口或靶位点3′

退火的臂长至少2倍、3倍、4倍或5倍。取决于ssDNA模板是否可以与完整链或带切口或靶位点的链退火,与缺口5′退火的同源臂可以分别位于ssDNA模板的5’端或ssDNA模板的3’端。

[0726] 类似地,在一些实施例中,模板核酸具有5′同源臂、替换序列和3’同源臂,使得模板核酸与切口的5’具有延伸的同源性。例如,5′同源臂和3′同源臂可以具有基本上相同的长度,但是替换序列比切口的3′可以更远地延伸切口的5′。在一些实施例中,替换序列比切口的3’端至少10％、20％、30％、40％、50％、2倍、3倍、4倍、或5倍更远地延伸至切口的5’端。

[0727] 虽然不希望受理论束缚,但在一些实施例中,当模板核酸以切口或靶位点为中心时,alt-HDR更有效地进行。因此,在一些实施例中,模板核酸具有两个基本上相同大小的同源臂。例如,模板核酸的第一同源臂可以具有在模板核酸的第二同源臂的10％、9％、8％、7％、6％、5％、4％、3％、2％或1％之内的长度。

[0728] 类似地,在一些实施例中,模板核酸具有5′同源臂、替换序列和3′同源臂,使得模板核酸在切口或靶位点的任一侧上延伸基本上相同的距离。例如,同源臂可以具有不同的长度,但是可以选择替换序列以补偿这一点。例如,替换序列比切口的3′可以从切口的5′延伸更远,但是切口的同源臂5’比切口的同源臂3’更短,以补偿。相反的也是可能的,例如,替换序列比切口的5’可以从切口的3’延伸更远,但是切口的同源臂3′比切口的同源臂5’更短,以补偿。

[0729] 示例性模板核酸

[0730] 在某些实施例中,模板核酸是双链的。在其他实施例中,模板核酸是单链的。在某些实施例中,模板核酸包含单链部分和双链部分。在某些实施例中,模板核酸包含在切口、靶位点、和/或替换序列的任一侧上的约50至100bp,例如,55至95、60至90、65至85、或70至80bp的同源性。在某些实施例中,模板核酸包含切口、靶位点、或替换序列的5’、切口、靶位点、或替换序列的3’、或切口、靶位点、或替换序列的5′和3′的约50、55、60、65、70、75、80、85、90、95、或100bp同源性。

[0731] 在某些实施例中,模板核酸包含在切口、靶位点、和/或替换序列3′的约150至200bp,例如155至195、160至190、165至185、或170至180bp的同源性。在某些实施例中,模板核酸包含切口、靶位点、或替换序列3′的约150、155、160、165、170、175、180、185、190、195、或200bp同源性。在某些实施例中,模板核酸包含切口、靶位点、或替换序列5′的小于约100、90、80、70、60、50、40、30、20、15、或10bp同源性。

[0732] 在某些实施例中,模板核酸包含在切口、靶位点、和/或替换序列5′的约150至200bp,例如,155至195、160至190、165至185、或170至180bp的同源性。在某些实施例中,模板核酸包含切口、靶位点、或替换序列5′的约150、155、160、165、170、175、180、185、190、195、或200bp同源性。在某些实施例中,模板核酸包含切口、靶位点、或替换序列3′的小于约100、90、80、70、60、50、40、30、20、15、或10bp同源性。

[0733] 在某些实施例中,模板核酸包含例如,一个或多个核苷酸的核苷酸序列,其将被添加到或将模拟靶核酸中的变化。在其他实施例中,模板核酸包含可用于修饰靶位置的核苷酸序列。在其他实施例中,模板核酸包含可用于缺失HBG靶位置的一个或多个核苷酸的核苷酸序列。

[0734] 该模板核酸可以包含替代序列。在一些实施例中,模板核酸包含5′同源臂。在其他的实施例中,该模板核酸包含3′同源臂。

[0735] 模板核酸可包含5′同源臂,0个核苷酸或0bp的替换序列,和3′同源臂。

[0736] 在某些实施例中,模板核酸是线性双链DNA。长度可以是例如约150bp-200bp,例如,约150bp、160bp、170bp、180bp、190bp或200bp。长度可以是例如,至少150bp、160bp、170bp、180bp、190bp或200bp。在一些实施例中,长度不大于150bp、160bp、170bp、180bp、190bp或200bp。在一些实施例中,双链模板核酸具有约160bp的长度,例如,约155bp-165bp、150bp-170bp、140bp-180bp、130bp-190bp、120bp-200bp、110bp-210bp、100bp-220bp、90bp-230bp或80bp-240bp。

[0737] 该模板核酸可以是线性单链DNA。在某些实施例中,模板核酸是(i)可退火至靶核酸的带切口的链的线性单链DNA,(ii)可退火至靶核酸的完整链的线性单链DNA,(iii)可退火至靶核酸的正链的线性单链DNA,(iv)可退火至靶核酸的负链的线性单链DNA,或多于一种前述的线性单链DNA。长度可以是例如,约150-200个核苷酸,例如,约150、160、170、180、190或200个核苷酸。长度可以是例如,至少150、160、170、180、190或200个核苷酸。在一些实施例中,长度不大于150、160、170、180、190或200个核苷酸。在一些实施例中,单链模板核酸具有约160个核苷酸的长度,例如,约155-165、150-170、140-180、130-190、120-200、110-210、100-220、90-230或80-240个核苷酸。

[0738] 在一些实施例中,模板核酸是环状双链DNA,例如,质粒。在一些实施例中,模板核酸在替换序列、靶位点、和/或切口的任一侧上包含约500至1000bp的同源性。在一些实施例中,模板核酸包含切口、靶位点、或替换序列的5′、切口、靶位点、或替换序列的3′、或切口、靶位点、或替换序列的5’和3’的约300、400、500、600、700、800、900、1000、1500、或2000bp同源性。在一些实施例中,模板核酸包含切口、靶位点、或替换序列的5’、切口、靶位点、或替换序列的3’、或切口、靶位点、或替换序列的5′和3′的至少300、400、500、600、700、800、900、1000、1500、或2000bp同源性。在一些实施例中,模板核酸包含切口、靶位点、或替换序列的5′、切口、靶位点、或替换序列的3’、或切口、靶位点、或替换序列的5’和3’的不超过300、400、500、600、700、800、900、1000、1500、或2000bp同源性。

[0739] 在某些实施例中,可以将一个或两个同源臂缩短以避免包括某些序列重复元件(例如,Alu重复、LINE元件)。例如,可以缩短5’同源臂以避免序列重复元件,而可以缩短3′同源臂以避免序列重复元件。在一些实施例中,可以将5’和3’同源臂均缩短以避免包括某些序列重复元件。

[0740] 在一些实施例中,模板核酸是腺病毒载体,例如,AAV载体,例如,长度和序列允许其包装在AAV衣壳中的ssDNA分子。载体可以是例如,小于5kb,并且可以含有促进包装到衣壳中的ITR序列。载体可能是整合缺陷的。在一些实施例中,模板核酸在替换序列、靶位点、和/或切口的任一侧上包含约150至1000个核苷酸的同源性。在一些实施例中,模板核酸包含切口、靶位点、或替换序列的5′、切口、靶位点、或替换序列的3′、或切口、靶位点、或替换序列的5’和3′的约100、150、200、300、400、500、600、700、800、900、1000、1500、或2000个核苷酸。在一些实施例中,模板核酸包含切口、靶位点、或替换序列的5′、切口、靶位点、或替换序列的3’、或切口、靶位点、或替换序列的5’和3’的至少100、150、200、300、400、500、600、700、800、900、1000、1500、或2000个核苷酸。在一些实施例中,模板核酸包含切口、靶位点、或替换序列的5’、切口、靶位点、或替换序列的3’、或切口、靶位点、或替换序列的5’和3’的至多100、150、200、300、400、500、600、700、800、900、1000、1500、或2000个核苷酸。

[0741] 在一些实施例中,模板核酸是慢病毒载体,例如,IDLV(整合缺陷型慢病毒)。在一些实施例中,模板核酸在替换序列、靶位点、和/或切口的任一侧上包含约500至1000bp的同源性。在一些实施例中,模板核酸包含切口、靶位点、或替换序列的5'、切口、靶位点、或替换序列的3′、或切口、靶位点、或替换序列的5'和3'的约300、400、500、600、700、800、900、1000、1500、或2000bp同源性。在一些实施例中,模板核酸包含切口、靶位点、或替换序列的5'、切口、靶位点、或替换序列的3′、或切口或替换序列的5'和3'的至少300、400、500、600、700、800、900、1000、1500、或2000bp同源性。在一些实施例中,模板核酸包含切口、靶位点、或替换序列的5'、切口、靶位点、或替换序列的3'、或切口、靶位点、或替换序列的5′和3'的不超过300、400、500、600、700、800、900、1000、1500、或2000bp同源性。

[0742] 在一个实施例中,模板核酸包含阻止Cas9识别和切割模板核酸的一个或多个突变,例如,沉默突变。相对于待改变的细胞基因组中的相应序列,模板核酸可包含例如,至少1、2、3、4、5、10、20或30个沉默突变。在某些实施例中,模板核酸相对于要改变的细胞的基因组中的相应序列包含至多2、3、4、5、10、20、30或50个沉默突变。在一个实施例中,cDNA包含阻止Cas9识别和切割模板核酸的一个或多个突变,例如,沉默突变。相对于待改变的细胞基因组中的相应序列,模板核酸可包含例如,至少1、2、3、4、5、10、20或30个沉默突变。在某些实施例中,模板核酸相对于要改变的细胞的基因组中的相应序列包含至多2、3、4、5、10、20、30或50个沉默突变。

[0743] 在本文提供的方法的某些实施例中,HDR介导的改变用于引入γ-珠蛋白基因调控区域中的一个或多个核苷酸的改变(例如,缺失)。在某些实施例中,γ-珠蛋白基因调控区域可以是HBG靶位置。在某些实施例中,可以在HBG靶位置内的靶位点处引入改变(例如,缺失)。在某些实施例中,改变(例如,缺失)可选自HBG1 13bp del c.-114至-102、HBG1 4bp del c.-225至-222、和HBG1 13bp del c.-114至-102中的一种或多种。在某些实施例中,靶位点可选自HBG1 c.-114至-102(例如,SEQ ID NO:902(HBG1)的核苷酸2824-2836)、HBG1 c.-225至-222(例如,SEQ ID NO:902(HBG1)的核苷酸2716-2719)、和HBG2 c.-114至-102(例如,SEQ ID NO:903(HBG2)的核苷酸2748-2760)中的一种或多种。

[0744] 在某些实施例中,用于在HBG靶位置内的靶位点(即,HBG1或HBG2调控区域)引入改变(例如,缺失)的模板核酸包含从5'到3'方向的5同源臂、替换序列、和3′同源臂,其中替换序列是0个核苷酸或0bp。在某些实施例中,所述模板核酸可以是单链寡脱氧核苷酸(ssODN)。在某些实施例中,5′同源臂可以是任何本文所述的5′同源臂。在某些实施例中,3′同源臂可以是任何本文所述的3′同源臂。在某些实施例中,可以在HBG靶位置内的靶位点处引入改变(例如,缺失)。在某些实施例中,改变(例如,缺失)可选自HBG1 13bp del c.-114至-102、HBG1 4bp del c.-225至-222、和HBG1 13bp del c.-114至-102中的一种或多种。在某些实施例中,靶位点可选自HBG1 c.-114至-102(例如,SEQ ID NO:902(HBG1)的核苷酸2824-2836)、HBG1 c.-225至-222(例如,SEQ ID NO:902(HBG1)的核苷酸2716-2719)、和HBG2 c.-114至-102(例如,SEQ ID NO:903(HBG2)的核苷酸2748-2760)中的一种或多种。

[0745] 例如,用于在靶位点HBG1 c.-114至-102(例如,SEQ ID NO:902(HBG1)的核苷酸2824-2836)处引入改变HBG1 13bp del c.-114至-102的模板核酸可以包含5′同源臂、替换序列、和3′同源臂,其中替换序列是0个核苷酸或0bp。在某些实施例中,5′同源臂的长度为包含约200个核苷酸,例如,长度为至少25、50、75、100、125、150、175、或200个核苷酸。在某

些实施例中,5′同源臂包含约50至100bp,例如55至95、60至90、70至90、或80至90bp,靶位点HBG1 c.-114至-102(例如,SEQ ID NO:902(HBG1)的核苷酸2824-2836)的同源性5'。在某些实施例中,5'同源臂包含SEQ ID NO:904(ssODN1 5′同源臂),基本上由其组成或由其组成。在某些实施例中,5'同源臂包含SEQ ID NO:907(PhTx ssODN1 5'同源臂),基本上由其组成或由其组成。在某些实施例中,3′同源臂的长度为包含约200个核苷酸,例如,长度为至少25、50、75、100、125、150、175、或200个核苷酸。在某些实施例中,3′同源臂包含约50至100bp,例如55至95、60至90、70至90、或80至90bp,靶位点HBG1 c.-114至-102(例如,SEQ ID NO:902(HBG1)的核苷酸2824-2836)的同源性3′。在某些实施例中,3'同源臂包含SEQ ID NO:905(ssODN1 3'同源臂),基本上由其组成或由其组成。在某些实施例中,3′同源臂包含SEQ ID NO:908(PhTx ssODN1 3′同源臂),基本上由其组成或由其组成。在某些实施例中,模板核酸包含SEQ ID NO:906,基本上由其组成或由其组成。在某些实施例中,模板核酸包含SEQ ID NO:909(PhTx ssODN1),基本上由其组成或由其组成。

[0746] 在另一个实例中,用于在靶位点HBG2 c.-114至-102(例如,SEQ ID NO:903(HBG2)的核苷酸2748-2760)处引入改变HBG2 13bp del c.-114至-102的模板核酸可以包含5'同源臂、替换序列、和3'同源臂,其中替换序列是0个核苷酸或0bp。在某些实施例中,5'同源臂的长度为包含约200个核苷酸,例如,长度为至少25、50、75、100、125、150、175、或200个核苷酸。在某些实施例中,5'同源臂包含约50至100bp,例如55至95、60至90、70至90、或80至90bp,靶位点HBG2 c.-114至-102(例如,SEQ ID NO:903(HBG2)的核苷酸2748-2760)的同源性5'。在某些实施例中,5'同源臂包含SEQ ID NO:904(ssODN1 5'同源臂),基本上由其组成或由其组成。在某些实施例中,5'同源臂包含SEQ ID NO:907(PhTx ssODN1 5′同源臂),基本上由其组成或由其组成。在某些实施例中,3'同源臂的长度为包含约200个核苷酸,例如,长度为至少25、50、75、100、125、150、175、或200个核苷酸。在某些实施例中,3′同源臂包含约50至100bp,例如55至95、60至90、70至90、或80至90bp,靶位点HBG2 c.-114至-102(例如,SEQ ID NO:903(HBG2)的核苷酸2748-2760)的同源性3'。在某些实施例中,3'同源臂包含SEQ ID NO:905(ssODN1 3′同源臂),基本上由其组成或由其组成。在某些实施例中,3′同源臂包含SEQ ID NO:908(PhTx ssODN1 3′同源臂),基本上由其组成或由其组成。在某些实施例中,模板核酸包含SEQ ID NO:906,基本上由其组成或由其组成。在某些实施例中,模板核酸包含SEQ ID NO:909(PhTx ssODN1),基本上由其组成或由其组成。

[0747] 在另一个实例中,用于在靶位点HBG1 c.-225至-222(例如,SEQ ID NO:902(HBG1)的核苷酸2716-2719)处引入改变HBG1 4bp del c.-225至-222的模板核酸可以包含5'同源臂、替换序列、和3'同源臂,其中替换序列是0个核苷酸或0bp。在某些实施例中,5′同源臂的长度为包含约200个核苷酸,例如,长度为至少25、50、75、100、125、150、175或200个核苷酸。在某些实施例中,5'同源臂包含约50至100bp,例如55至95、60至90、70至90、或80至90bp,靶位点HBG1 c.-225至-222(例如,SEQ ID NO:902(HBG1)的核苷酸2716-2719)的同源性5′。在某些实施例中,3'同源臂的长度为包含约200个核苷酸,例如,长度为至少25、50、75、100、125、150、175或200个核苷酸。在某些实施例中,3′同源臂包含约50至100bp,例如55至95、60至90、70至90、或80至90bp,靶位点HBG1 c.-225至-222(例如,SEQ ID NO:902(HBG1)的核苷酸2716-2719)的同源性3'。

[0748] 在某些实施例中,5'同源臂包含5′硫代磷酸酯(PhTx)修饰。在某些实施例中,3′同

源臂包含3'PhTx修饰。在某些实施例中,模板核酸包含5'和3'PhTx修饰。

[0749] 在某些实施例中,用于改变γ-珠蛋白基因(例如,HBG1,HBG2)调控区域中的单个核苷酸的模板核酸包含5'至3'方向的5'同源臂、替换序列、和3'同源臂,其中替换被设计成掺入单核苷酸改变。例如,其中掺入的改变是HBG1 c.-114C>T、c.-158C>T、c.-167C>T、c.-196C>T、或c.-201C>T或HBG2 c.-109G>T、c.-114C>T、c.-157C>T、c.-158C>T、c.-167C>T、c.-211C>T,替换序列可以包含单个核苷酸T,并且任选地在所述T的一侧或两侧包含一个或多个核苷酸。类似地,如果掺入的改变是HBG1 c.-117G>A、c.-170G>A、或c.-499T>A或HBG2 c.-114C>A或c.-167C>A,替换序列可以包含单个核苷酸A,并且任选地在所述A的一侧或两侧包含一个或多个核苷酸;其中掺入的改变是HBG1 c.-175T>G或c.-195C>G或HBG2 c.-202C>G、c.-255C>G、c.-309A>G、c.-369C>G、或c.-567T>G,替换序列可以包含单个核苷酸G,并且任选地在所述G的一侧或两侧包含一个或多个核苷酸;并且其中掺入的改变是HBG1 c.-175T>C、c.-198T>C、或c.-251T>C或HBG2 c.-175T>C或c.-228T>C,替换序列可以包含单个核苷酸C,并且任选地在所述C的一侧或两侧包含一个或多个核苷酸。

[0750] 在某些实施例中,5'和3'同源臂各自包含与替换序列相应的核苷酸侧翼的一段序列长度。在某些实施例中,模板核酸包含侧翼为5′同源臂和3'同源臂的替换序列,所述臂各自独立地包含10个或更多、20个或更多、50个或更多、100个或更多、150个或更多、200个或更多、250个或更多、300个或更多、350个或更多、400个或更多、450个或更多、500个或更多、550个或更多、600个或更多、650个或更多、700个或更多、750个或更多、800个或更多、850个或更多、900个或更多、1000个或更多、1100个或更多、1200个或更多、1300个或更多、1400个或更多、1500个或更多、1600个或更多、1700个或更多、1800个或更多、1900个或更多、或2000个或更多核苷酸。在某些实施例中,模板核酸包含侧翼为5'同源臂和3'同源臂的替换序列,所述臂各自独立地包含至少50个、100个、或150个核苷酸,但不足以长到包括重复元件。在某些实施例中,模板核酸包含侧翼为5′同源臂和3′同源臂的替换序列,所述臂各自独立地包含5至100个、10至150个、或20至150个核苷酸。在某些实施例中,替换序列任选地包含启动子和/或polyA信号。

[0751] 单链退火

[0752] 单链退火(SSA)是另一种修复存在于靶核酸中的两个重复序列之间的双链断裂的DNA修复过程。由SSA途径所利用的重复序列的长度通常大于30个核苷酸。在断裂端处发生切除,以显示靶核酸的两条链上的重复序列。切除之后,将含有重复序列的单链突出端用RPA蛋白涂覆,以防止重复序列不适当地退火,例如退火至自身。RAD52结合至突出端上的重复序列中的每者上并且将所述序列对齐以使得互补的重复序列能够退火。退火之后,切割突出端的单链翼。新的DNA合成填充任何空位,并且连接恢复DNA双链体。作为所述处理的结果,两个重复之间的DNA序列被缺失。缺失的长度可以取决于很多因素,包括所利用的两个重复的位置、以及切除的途径或持续进行能力。

[0753] 与HDR途径相反,SSA不需要模板核酸来改变靶核酸序列。而是利用互补的重复序列。

[0754] 其他DNA修复途径

[0755] SSBR(单链断裂修复)

[0756] 基因组中的单链断裂(SSB)由SSBR途径来修复,所述SSBR途径是不同于上文讨论

的DSB修复机制的机制。SSBR途径具有四个主要阶段:SSB检测、DNA末端处理、DNA空位填充、以及DNA连接。更详细的解释给出于Caldecott 2008中,并且这里给出概述。

[0757] 在第一阶段中,当形成SSB时,PARP1和/或PARP2识别断裂并募集修复机器。DNA断裂处的PARP1结合和活性是瞬时的,并且它似乎通过促进损伤处的SSBr蛋白复合物的病灶积累或稳定性而加速SSBr。可论证地,这些SSBr蛋白中最重要的是XRCC1,它作为分子支架起作用,所述分子支架与SSBr过程的多种酶组分(包括负责清除DNA 3'和5'端的蛋白质)相互作用,使所述酶组分稳定化,并且刺激所述酶组分。例如,XRCC1与促进末端处理的若干蛋白质(DNA聚合酶β、PNK、和三种核酸酶APE1、APTX和APLF)相互作用。APE1具有内切核酸酶活性。APLF展示出内切核酸酶和3'到5'外切核酸酶活性。APTX具有内切核酸酶和3'到5'外切核酸酶活性。

[0758] 这种末端处理是SSBR的重要阶段,因为大部分(若非全部)SSB的3'-和/或5'-末端是'被损伤的'。末端处理通常涉及将被损伤的3'-端恢复到羟基化状态和/或将被损伤的5'端恢复成磷酸酯部分,这样使得所述末端变得有连接能力。可以处理被损伤的3'末端的酶包括PNKP、APE1、和TDP1。可以处理被损伤的5'末端的酶包括PNKP、DNA聚合酶β、和APTX。LIG3(DNA连接酶III)也可以参与末端处理。一旦将末端清除,便可发生空位填充。

[0759] 在DNA空位填充阶段,典型存在的蛋白质是PARP1、DNA聚合酶β、XRCC1、FEN1(翼内切核酸酶1)、DNA聚合酶δ/ε、PCNA、以及LIG1。存在两种空位填充方式,短补丁修复(short patch repair)和长补丁修复(long patch repair)。短补丁修复涉及插入丢失的单核苷酸。在一些SSB处,"空位填充"可能继续取代两个或更多个核苷酸(已经报道了多达12个碱基的取代)。FEN1是去除被取代的5'-残基的内切核酸酶。多种DNA聚合酶(包括Polβ)涉及在SSB的修复中,其中DNA聚合酶的选择受SSB的来源和类型的影响。

[0760] 在第四阶段中,DNA连接酶如LIG1(连接酶I)或LIG3(连接酶III)催化末端连接。短补丁修复使用连接酶III,并且长补丁修复使用连接酶I。

[0761] 有时,SSBR是与复制偶联的。这条途径可以涉及CtIP、MRN、ERCC1、和FEN1中的一者或多者。可以促进SSBR的另外的因子包括:aPARP、PARP1、PARP2、PARG、XRCC1、DNA聚合酶b、DNA聚合酶d、DNA聚合酶e、PCNA、LIG1、PNK、PNKP、APE1、APTX、APLF、TDP1、LIG3、FEN1、CtIP、MRN、以及ERCC1。

[0762] MMR(错配修复)

[0763] 细胞含有三条切除修复途径:MMR、BER、和NER。所述切除修复途径具有的共同特点在于它们典型地识别DNA一条链上的损伤,然后外切/内切核酸酶去除所述损伤并且留下随后被DNA聚合酶填充的1-30个核苷酸的空位并且最终用连接酶密封。更完整的图片给出于Li 2008中,并且这里提供了概述。

[0764] 错配修复(MMR)在错配的DNA碱基上运行。

[0765] MSH2/6或MSH2/3两种复合物都具有在错配识别和修复启动中发挥重要作用的ATP酶活性。MSH2/6优先识别碱基-碱基错配并且识别1或2个核苷酸的错配,而MSH2/3优先识别较大的ID错配。

[0766] hMLH1与hPMS2杂二聚化,以形成hMutLα,其具有ATP酶活性并且对于MMR的多个步骤而言是重要的。它具有PCNA/复制因子C(RFC)依赖性内切核酸酶活性,所述活性在涉及EXO1(EXO1是HR和MMR两者中的参与者)的3′切口指导的MMR中发挥重要作用。它调节错配引

起的切除的终止。连接酶I是这条途径的相关连接酶。可以促进MMR的另外的因子包括：EXO1、MSH2、MSH3、MSH6、MLH1、PMS2、MLH3、DNA Pol d、RPA、HMGB1、RFC、以及DNA连接酶I。

[0767] 碱基切除修复(BER)

[0768] 碱基切除修复(BER)途径贯穿细胞周期是激活的；它主要负责从基因组中去除小的、非螺旋扭曲碱基损伤。相比之下，相关的核苷酸切除修复途径(在下一部分中讨论)修复庞大的螺旋扭曲损伤。更详细的解释给出于Caldecott 2008中，并且这里给出概述。

[0769] DNA碱基损伤后，碱基切除修复(BER)启动并且所述过程可以被简化为五个主要步骤：(a)去除被损伤的DNA碱基；(b)切开后续碱基位点；(c)清理DNA末端；(d)将希望的核苷酸(例如，HPFH突变体)插入修复空位；以及(e)连接DNA骨架中的剩余切口。这些最后的步骤类似于SSBR。

[0770] 在第一步中，损伤特异性DNA糖基化酶通过切割将碱基连接至糖磷酸骨架上的N-糖苷键而切除被损伤的碱基。然后具有相关裂解酶活性的AP内切核酸酶-1(APE1)或双功能DNA糖基化酶切开磷酸二酯骨架以产生DNA单链断裂(SSB)。BER的第三步涉及清理DNA末端。BER中的第四步由Polβ进行，它将新的互补核苷酸添加到修复空位中，并且在最终步骤中，XRCC1/连接酶III密封DNA骨架中的剩余切口。这完成了短补丁BER途径，其中大多数(约80％)的被损伤的DNA碱基得到修复。然而，如果在步骤3中在通过Polβ插入一个核苷酸后5′-端对末端处理活性有抗性，则将聚合酶换为复制型DNA聚合酶Polδ/ε，所述复制型DNA聚合酶然后再将约2-8个核苷酸添加到DNA修复空位中。这产生了5′翼结构，其被与持续合成能力因子增殖细胞核抗原(PCNA)相关的翼内切核酸酶-1(FEN-1)识别并切除。DNA连接酶I然后密封DNA骨架中的剩余切口并完成长补丁BER。可以促进BER途径的另外的因子包括：DNA糖基化酶、APE1、Polb、Pold、Pole、XRCC1、连接酶III、FEN-1、PCNA、RECQL4、WRN、MYH、PNKP、以及APTX。

[0771] 核苷酸切除修复(NER)

[0772] 核苷酸切除修复(NER)是从DNA中去除庞大的螺旋扭曲损伤的重要切除机制。关于NER的另外的细节在Marteijn 2014中给出，并且这里给出概述。NER是涵盖两条更小途径的宽途径：全基因组NER(GG-NER)和转录偶联修复NER(TC-NER)。GG-NER和TC-NER使用不同的因子来识别DNA损伤。然而，它们利用相同的机器进行损伤切开、修复、和连接。

[0773] 一旦识别出损伤，细胞去除含有所述损伤的短的单链DNA区段。内切核酸酶XPF/ERCC1和XPG(由ERCC5编码)通过切割损伤任一侧上的被损伤的链而去除损伤，产生22-30个核苷酸的单链空位。接着，细胞进行DNA空位填充合成和连接。在这个过程中涉及的是：PCNA、RFC、DNA Polδ、DNA Polε或DNA Polκ、以及DNA连接酶I或XRCC1/连接酶III。复制型细胞倾向于使用DNA polε和DNA连接酶I进行连接步骤，而非复制型细胞倾向于使用DNA Polδ、DNA Polκ、和XRCC1/连接酶III复合物进行连接步骤。

[0774] NER可以涉及以下因子：XPA-G、POLH、XPF、ERCC1、XPA-G、以及LIG1。转录偶联NER(TC-NER)可以涉及以下因子：CSA、CSB、XPB、XPD、XPG、ERCC1、以及TTDA。可以促进NER修复途径的另外的因子包括XPA-G、POLH、XPF、ERCC1、XPA-G、LIG1、CSA、CSB、XPA、XPB、XPC、XPD、XPF、XPG、TTDA、UVSSA、USP7、CETN2、RAD23B、UV-DDB、CAK子复合物、RPA、以及PCNA。

[0775] 链间交联(ICL)

[0776] 称为ICL修复途径的专用途径修复链间交联。可以在复制或转录期间在不同DNA链

中的碱基之间发生链间交联、或共价交联。ICL修复涉及多个修复过程的协作,具体地溶核活性、跨损伤合成(TLS)、和HDR。核酸酶被募集以切除被交联的碱基的任一侧上的ICL,同时TLS和HDR协作以修复被切割的链。ICL修复可以涉及以下因子:内切核酸酶(例如,XPF和RAD51C)、内切核酸酶(例如RAD51)、跨损伤聚合酶(例如,DNA聚合酶ζ和Rev1)、以及范科尼贫血(FA)蛋白(例如,FancJ)。

[0777] 其他途径

[0778] 在哺乳动物体内存在若干其他DNA修复途径。

[0779] 跨损伤合成(TLS)是用于修复有缺陷复制事件之后留下的单链断裂的途径,并且涉及跨损伤聚合酶(例如,DNA polβ和Rev1)。

[0780] 无误复制后修复(PRR)是另一条用于修复有缺陷复制事件之后留下的单链断裂的途径。

[0781] 基因组编辑方法中gRNA的实例

[0782] 如在此描述的gRNA分子可以与产生双链断裂或单链断裂的Cas9分子一起使用,以改变靶核酸的序列,例如靶位置或靶基因标签。下文描述了在这些方法中有用的gRNA分子。

[0783] 在某些实施例中,所述gRNA(例如,一种嵌合gRNA)被配置成使得它包含以下特性中的一种或多种:

[0784] (a)它可以将双链断裂(例如,当靶向产生双链断裂的Cas9分子时)定位(i)在靶位置的50、100、150、200、250、300、350、400、450、或500个核苷酸内、或(ii)足够接近使得所述靶位置在末端切除的区域内;

[0785] (b)它具有至少16个核苷酸的靶向结构域,例如(i)16、(ii)17、(iii)18、(iv)19、(v)20、(vi)21、(vii)22、(viii)23、(ix)24、(x)25、或(xi)26个核苷酸的靶向结构域;以及

[0786] (c)(i)当一起考虑时,所述近端和尾部结构域包含来自以下项的至少15、18、20、25、30、31、35、40、45、49、50、或53个核苷酸,例如至少15、18、20、25、30、31、35、40、45、49、50,或53个核苷酸:天然存在的化脓链球菌或金黄色葡萄球菌尾部和近端结构域、或一个与之相差不多于1、2、3、4、5、6、7、8、9或10个核苷酸的序列;

[0787] (c)(ii)所述第二互补结构域的最后一个核苷酸的3'存在来自以下项的至少15、18、20、25、30、31、35、40、45、49、50、或53个核苷酸,例如至少15、18、20、25、30、31、35、40、45、49、50,或53个核苷酸:天然存在的化脓链球菌或金黄色葡萄球菌gRNA的相应序列、或一个与之相差不多于1、2、3、4、5、6、7、8、9或10个核苷酸的序列;

[0788] (c)(iii)所述第二互补结构域的最后一个核苷酸的3'存在来自以下项的至少16、19、21、26、31、32、36、41、46、50、51、或54个核苷酸(这些核苷酸与所述第一互补结构域的相应核苷酸互补)例如至少16、19、21、26、31、32、36、41、46、50、51、或54个核苷酸:天然存在的化脓链球菌或金黄色葡萄球菌gRNA的相应序列、或一个与之相差不多于1、2、3、4、5、6、7、8、9或10个核苷酸的序列;

[0789] (c)(iv)所述尾部结构域在长度上是至少10、15、20、25、30、35或40个核苷酸,例如,它包含来自以下项的至少10、15、20、25、30、35或40个核苷酸:天然存在的化脓链球菌或金黄色葡萄球菌尾部结构域、或一个与之相差不多于1、2、3、4、5、6、7、8、9或10个核苷酸的序列;或

[0790] (c)(v)所述尾部结构域包含天然存在的尾部结构域(例如,天然存在的化脓链球

菌或金黄色葡萄球菌尾部结构域)的相应部分的15、20、25、30、35、40个核苷酸或全部。

[0791] 在某些实施例中,所述gRNA被配置成使得它包含特性:a和b(i);a和b(ii);a和b(iii);a和b(iv);a和b(v);a和b(vi);a和b(vii);a和b(viii);a和b(ix);a和b(x);a和b(xi);a和c;a、b和c;a(i)、b(i)和c(i);a(i)、b(i)和c(ii);a(i)、b(ii)和c(i);a(i)、b(ii)和c(ii);a(i)、b(iii)和c(i);a(i)、b(iii)和c(ii);a(i)、b(iv)和c(i);a(i)、b(iv)和c(ii);a(i)、b(v)和c(i);a(i)、b(v)和c(ii);a(i)、b(vi)和c(i);a(i)、b(vi)和c(ii);a(i)、b(vii)和c(i);a(i)、b(vii)和c(ii);a(i)、b(viii)和c(i);a(i)、b(viii)和c(ii);a(i)、b(ix)和c(i);a(i)、b(ix)和c(ii);a(i)、b(x)和c(i);a(i)、b(x)和c(ii);a(i)、b(xi)或c(i);a(i)、b(xi)和c(ii)。

[0792] 在某些实施例中,所述gRNA(例如,一种嵌合gRNA)被配置成使得它包含以下特性中的一种或多种:

[0793] (a)gRNA中的一者或两者可以将单链断裂(例如,当靶向产生单链断裂的Cas9分子时)定位(i)在靶位置的50、100、150、200、250、300、350、400、450、或500个核苷酸内、或(ii)足够接近使得所述靶位置在末端切除的区域内;

[0794] (b)一者或两者具有至少16个核苷酸的靶向结构域,例如(i)16、(ii)17、(iii)18、(iv)19、(v)20、(vi)21、(vii)22、(viii)23、(ix)24、(x)25、或(xi)26个核苷酸的靶向结构域;以及

[0795] (c)(i)当一起考虑时,所述近端和尾部结构域包含来自以下项的至少15、18、20、25、30、31、35、40、45、49、50、或53个核苷酸,例如,至少15、18、20、25、30、31、35、40、45、49、50,或53个核苷酸:天然存在的化脓链球菌或金黄色葡萄球菌尾部和近端结构域、或一个与之相差不多于1、2、3、4、5、6、7、8、9或10个核苷酸的序列;

[0796] (c)(ii)所述第二互补结构域的最后一个核苷酸的3′存在来自以下项的至少15、18、20、25、30、31、35、40、45、49、50、或53个核苷酸,例如,至少15、18、20、25、30、31、35、40、45、49、50,或53个核苷酸:天然存在的化脓链球菌或金黄色葡萄球菌gRNA的相应序列、或一个与之相差不多于1、2、3、4、5、6、7、8、9或10个核苷酸的序列;

[0797] (c)(iii)所述第二互补结构域的最后一个核苷酸的3'存在来自以下项的至少16、19、21、26、31、32、36、41、46、50、51、或54个核苷酸(这些核苷酸与所述第一互补结构域的相应核苷酸互补)例如,至少16、19、21、26、31、32、36、41、46、50、51、或54个核苷酸:天然存在的化脓链球菌或金黄色葡萄球菌gRNA的相应序列、或一个与之相差不多于1、2、3、4、5、6、7、8、9或10个核苷酸的序列;

[0798] (c)(iv)所述尾部结构域在长度上是至少10、15、20、25、30、35或40个核苷酸,例如,它包含来自以下项的至少10、15、20、25、30、35或40个核苷酸:天然存在的化脓链球菌或金黄色葡萄球菌尾部结构域、或一个与之相差不多于1、2、3、4、5、6、7、8、9或10个核苷酸的序列;或

[0799] (c)(v)所述尾部结构域包含天然存在的尾部结构域(例如,天然存在的化脓链球菌或金黄色葡萄球菌尾部结构域)的相应部分的15、20、25、30、35、或40个核苷酸或全部。

[0800] 在某些实施例中,所述gRNA被配置成使得它包含特性:a和b(i);a和b(ii);a和b(iii);a和b(iv);a和b(v);a和b(vi);a和b(vii);a和b(viii);a和b(ix);a和b(x);a和b(xi);a和c;a、b和c;a(i)、b(i)和c(i);a(i)、b(i)和c(ii);a(i)、b(ii)和c(i);a(i)、b(ii)

和c(ii);a(i)、b(iii)和c(i);a(i)、b(iii)和c(ii);a(i)、b(iv)和c(i);a(i)、b(iv)和c(ii);a(i)、b(v)和c(i);a(i)、b(v)和c(ii);a(i)、b(vi)和c(i);a(i)、b(vi)和c(ii);a(i)、b(vii)和c(i);a(i)、b(vii)和c(ii);a(i)、b(viii)和c(i);a(i)、b(viii)和c(ii);a(i)、b(ix)和c(i);a(i)、b(ix)和c(ii);a(i)、b(x)和c(i);a(i)、b(x)和c(ii);a(i)、b(xi)和c(i);a(i)、b(xi)和c(ii)。

[0801] 在某些实施例中,所述gRNA与具有HNH活性的Cas9切口酶分子一起使用,例如,具有RuvC活性失活的Cas9分子(例如,在D10处具有突变(例如,D10A突变)的Cas9分子)。

[0802] 在一个实施例中,所述gRNA与具有RuvC活性的Cas9切口酶分子一起使用,例如,具有HNH活性失活的Cas9分子(例如,在840处具有突变(例如,H840A)的Cas9分子)。

[0803] 在一个实施例中,所述gRNA与具有RuvC活性的Cas9切口酶分子一起使用,例如HNH活性失活的Cas9分子(例如,在N863处具有突变(例如,N863A突变)的Cas9分子)。

[0804] 在一个实施例中,一对gRNA(例如,一对嵌合gRNA),其包含第一和第二gRNA,被配置成使得他们包含以下特性中的一种或多种:

[0805] (a)gRNA中的一者或两者可以将单链断裂(例如,当靶向产生单链断裂的Cas9分子时)定位(i)在靶位置的50、100、150、200、250、300、350、400、450、或500个核苷酸内、或(ii)足够接近使得所述靶位置在末端切除的区域内;

[0806] (b)一者或两者具有至少16个核苷酸的靶向结构域,例如(i)16、(ii)17、(iii)18、(iv)19、(v)20、(vi)21、(vii)22、(viii)23、(ix)24、(x)25、或(xi)26个核苷酸的靶向结构域;

[0807] (c)(i)当一起考虑时,所述近端和尾部结构域包含来自以下项的至少15、18、20、25、30、31、35、40、45、49、50、或53个核苷酸,例如至少15、18、20、25、30、31、35、40、45、49、50,或53个核苷酸:天然存在的化脓链球菌或金黄色葡萄球菌尾部和近端结构域、或一个与之相差不多于1、2、3、4、5、6、7、8、9或10个核苷酸的序列;

[0808] (c)(ii)所述第二互补结构域的最后一个核苷酸的3'存在来自以下项的至少15、18、20、25、30、31、35、40、45、49、50、或53个核苷酸,例如,至少15、18、20、25、30、31、35、40、45、49、50,或53个核苷酸:天然存在的化脓链球菌或金黄色葡萄球菌gRNA的相应序列、或一个与之相差不多于1、2、3、4、5、6、7、8、9或10个核苷酸的序列;

[0809] (c)(iii)所述第二互补结构域的最后一个核苷酸的3'存在来自以下项的至少16、19、21、26、31、32、36、41、46、50、51、或54个核苷酸(这些核苷酸与所述第一互补结构域的相应核苷酸互补)例如,至少16、19、21、26、31、32、36、41、46、50、51、或54个核苷酸:天然存在的化脓链球菌或金黄色葡萄球菌gRNA的相应序列、或一个与之相差不多于1、2、3、4、5、6、7、8、9或10个核苷酸的序列;

[0810] (c)(iv)所述尾部结构域在长度上是至少10、15、20、25、30、35或40个核苷酸,例如,它包含来自以下项的至少10、15、20、25、30、35或40个核苷酸:天然存在的化脓链球菌或金黄色葡萄球菌尾部结构域、或一个与之相差不多于1、2、3、4、5、6、7、8、9或10个核苷酸的序列;或

[0811] (c)(v)所述尾部结构域包含天然存在的尾部结构域(例如,天然存在的化脓链球菌或金黄色葡萄球菌尾部结构域)的相应部分的15、20、25、30、35、或40个核苷酸或全部;

[0812] (d)所述gRNA被配置成使得,当杂交到靶核酸上时,他们被0-50、0-100、0-200、至

少10、至少20、至少30或至少50个核苷酸分开;

[0813] (e)由所述第一gRNA和第二gRNA产生的断裂在不同的链上;以及

[0814] (f)这些PAM面向外部。

[0815] 在某些实施例中,所述gRNA中的一者或两者被配置成使得它包含特性:a和b(i);a和b(ii);a和b(iii);a和b(iv);a和b(v);a和b(vi);a和b(vii);a和b(viii);a和b(ix);a和b(x);a和b(xi);a和c;a、b和c;a(i)、b(i)和c(i);a(i)、b(i)和c(ii);a(i)、b(i)c和d;a(i)、b(i)、c和e;a(i)、b(i)、c、d和e;a(i)、b(ii)和c(i);a(i)、b(ii)和c(ii);a(i)、b(ii)、c和d;a(i)、b(ii)、c和e;a(i)、b(ii)、c、d和e;a(i)、b(iii)和c(i);a(i)、b(iii)和c(ii);a(i)、b(iii)、c和d;a(i)、b(iii)、c和e;a(i)、b(iii)、c、d和e;a(i)、b(iv)和c(i);a(i)、b(iv)和c(ii);a(i)、b(iv)、c和d;a(i)、b(iv)、c和e;a(i)、b(iv)、c、d和e;a(i)、b(v)和c(i);a(i)、b(v)和c(ii);a(i)、b(v)、c和d;a(i)、b(v)、c和e;a(i)、b(v)、c、d和e;a(i)、b(vi)和c(i);a(i)、b(vi)和c(ii);a(i)、b(vi)、c和d;a(i)、b(vi)、c和e;a(i)、b(vi)、c、d和e;a(i)、b(vii)和c(i);a(i)、b(vii)和c(ii);a(i)、b(vii)、c和d;a(i)、b(vii)、c和e;a(i)、b(vii)、c、d和e;a(i)、b(viii)和c(i);a(i)、b(viii)和c(ii);a(i)、b(viii)、c和d;a(i)、b(viii)、c和e;a(i)、b(viii)、c、d和e;a(i)、b(ix)和c(i);a(i)、b(ix)和c(ii);a(i)、b(ix)、c和d;a(i)、b(ix)、c和e;a(i)、b(ix)、c、d和e;a(i)、b(x)和c(i);a(i)、b(x)和c(ii);a(i)、b(x)、c和d;a(i)、b(x)、c和e;a(i)、b(x)、c、d和e;a(i)、b(xi)和c(i);a(i)、b(xi)和c(ii);a(i)、b(xi)、c和d;a(i)、b(xi)、c和e;a(i)、b(xi)、c、d和e。

[0816] 在某些实施例中,所述gRNA与具有HNH活性的Cas9切口酶分子一起使用,例如,具有RuvC活性失活的Cas9分子(例如,在D10处具有突变(例如,D10A突变)的Cas9分子)。

[0817] 在某些实施例中,所述gRNA与具有RuvC活性的Cas9切口酶分子一起使用,例如,HNH活性失活的Cas9分子(例如,在H840处具有突变(例如,H840突变)的Cas9分子)。

[0818] 在某些实施例中,所述gRNA与具有RuvC活性的Cas9切口酶分子一起使用,例如,HNH活性失活的Cas9分子(例如,在N863处具有突变(例如,N863A突变)的Cas9分子)。

[0819] 靶细胞

[0820] 在多种细胞中,可以将Cas9分子和gRNA分子(例如,Cas9分子/gRNA分子复合物)用于改变(例如,引入突变或缺失)靶核酸,例如,γ-珠蛋白基因(例如,HBG1,HBG2)调控区域)。在某些实施例中,靶向细胞中靶核酸的改变可以在体外、离体或体内进行。

[0821] 可以将本文描述的Cas9和gRNA分子递送至靶细胞。在某些实施例中,靶向细胞是红系细胞,例如,成红细胞。在某些实施例中,优先靶向红系细胞,例如,至少约90%、95%、96%、97%、98%、99%或100%的靶细胞是红系细胞。例如,在体内递送的情况下,优先靶向红系细胞,并且如果细胞离体处理并返回受试者,则优先修饰红系细胞。

[0822] 在某些实施例中,靶向细胞是循环血细胞,例如,网织红细胞、巨核细胞红系祖细胞(MEP)、髓样祖细胞(CMP/GMP)、淋巴样祖细胞(LP)、造血干细胞/祖细胞(HSC)、或内皮细胞(EC)。在某些实施例中,靶向细胞是骨髓细胞(例如,网织红细胞、红系细胞(例如,成红细胞)、MEP细胞、髓样祖细胞(CMP/GMP)、LP细胞、红系祖细胞(EP)、HSC、多能祖细胞(MPP)、内皮细胞(EC)、造血内皮细胞(HE)、或间充质干细胞)。在某些实施例中,靶向细胞是髓样祖细胞(例如,普通髓样祖细胞(CMP)或粒细胞巨噬细胞集落刺激因子祖细胞(GMP))。在某些实施例中,靶向细胞是淋巴样祖细胞,例如,淋巴共同祖细胞(CLP)。在某些实施例中,靶向细

胞是红系祖细胞(例如,MEP细胞)。在某些实施例中,靶向细胞是造血干细胞/祖细胞(例如,长期HSC(LT-HSC)、短期HSC(ST-HSC)、MPP细胞、或谱系限制性祖细胞(LRP))。在某些实施例中,靶向细胞是$CD34^+$细胞、$CD34^+CD90^+$细胞、$CD34^+CD38^-$细胞、$CD34^+CD90^+CD49f^+CD38^-CD45RA^-$细胞、$CD105^+$细胞、$CD31^+$、或$CD133^+$细胞、或$CD34^+CD90^+CD133^+$细胞。在某些实施例中,靶向细胞是脐血$CD34^+$HSPC、脐带静脉内皮细胞、脐带动脉内皮细胞、羊水$CD34^+$细胞、羊水内皮细胞、胎盘内皮细胞或胎盘造血$CD34^+$细胞。在某些实施例中,靶向细胞是动员的外周血造血$CD34^+$细胞(在患者用动员剂例如,G-CSF或普乐沙福(Plerixafor)治疗之后)。在某些实施例中,靶向细胞是外周血内皮细胞。

[0823] 在某些实施例中,通过编辑γ-珠蛋白基因调控区域离体操纵靶向细胞,然后将靶细胞施用于受试者。用于离体操纵的靶向细胞的来源可以包括,例如,受试者的血液、骨髓或脐血。其他用于离体操纵的靶向细胞的来源可以包括,例如,异源供体血液、脐血、或骨髓。在某些实施例中,从受试者中去除红细胞,将其如上所述离体操纵,并返回至受试者中。在某些实施例中,从受试者中去除造血干细胞,将其如上所述离体操纵,并返回至受试者中。在某些实施例中,从受试者中去除红系祖细胞,将其如上所述离体操纵,并返回至受试者中。在某些实施例中,从受试者中去除髓样祖细胞,将其如上所述离体操纵,并返回至受试者中。在某些实施例中,从受试者中去除多能祖细胞(MPP),将其如上所述离体操纵,并返回至受试者中。在某些实施例中,从受试者中去除造血干细胞/祖细胞(HSC),将其如上所述离体操纵,并返回至受试者中。在某些实施例中,从受试者中去除$CD34^+$HSC,将其如上所述离体操纵,并返回至受试者中。

[0824] 在某些实施例中,离体产生的修饰的HSC在没有清髓性预调理的情况下施用于受试者。在其他实施例中,在轻度清髓性调理后施用修饰的HSC,使得在植入后,一些造血细胞衍生自修饰的HSC。在又其他实施例中,在完全清髓性调理后施用修饰的HSC,使得在植入后,100%造血细胞衍生自修饰的HSC。

[0825] 合适的细胞还可包括干细胞,例如,胚胎干细胞、诱导性多能干细胞,造血干细胞、或造血内皮细胞(HE)(造血干细胞和内皮细胞的前体)。在某些实施例中,细胞是诱导的多能干细胞(iPS)或衍生自iPS细胞的细胞,例如,使用本文公开的方法修饰的受试者产生的iPS细胞,并分化成临床相关细胞,例如,红细胞。在某些实施例中,将AAV用于转导靶细胞。

[0826] 在某些实施例中,可以制备用于如本文所述的基因编辑的干细胞,用于根据Gori2016中的实例中描述的方法,例如,第219-223页、第223-224页、第227-231页、第231-236页、第235-238页、第240-241页、第242-244页,其通过引用结合在此。干细胞可以以本领域技术人员合适和已知的任何方式培养和扩增。

[0827] 可以即刻使用由本文所述方法产生的细胞。可替代地,细胞可以被冷冻(例如在液氮中)并储存以备后用。通常细胞将被冷冻在10%二甲基亚砜(DMSO)、50%血清、40%缓冲介质或本领域常用的一些其他此类溶液中,以在这样的冷冻温度下保存细胞,并以本领域通常已知用于解冻冷冻的培养细胞的方式解冻。细胞也可以进行热稳定化,以便在4℃下长期存储。

[0828] <u>递送、配制品和给予途径</u>

[0829] 基因组编辑系统组分,例如,RNA指导的核酸酶分子,例如,Cas9分子、gRNA分子(例如,Cas9分子/gRNA分子复合物)、和供体模板核酸或全部三种,能以各种形式被递送、配制

或施用,参见例如,表3和表4。

[0830] 在某些实施例中,例如通过AAV载体递送一种Cas9分子和两种或更多种(例如,2、3、4种或更多种)不同的gRNA分子。在某些实施例中,编码Cas9分子的序列和编码两种或更多种(例如,2、3、4种或更多种)不同的gRNA分子的一个或多个序列存在于相同的核酸分子上,例如,AAV载体。当递送一种Cas9或gRNA组分被编码在DNA中时,所述DNA将典型地包括控制区域(例如,包含启动子)以实现表达。对于Cas9分子序列有用的启动子包括CMV、SFFV、EFS、EF-1a、PGK、CAG和CBH启动子或血细胞特异性启动子。在实施例中,启动子是组成型启动子。在另一个实施例中,启动子是组织特异性启动子。对于gRNA有用的启动子包括T7.H1、EF-1a、U6、U1和tRNA启动子。可以选择具有类似或不同强度的启动子来调谐组分的表达。编码Cas9分子的序列可以包括核定位信号(NLS),例如,SV40 NLS。在一个实施例中,编码Cas9分子的序列包含至少两个核定位信号。在实施例中,用于Cas9分子或gRNA分子的启动子可以独立地是诱导型、组织特异性、或细胞特异性的。

[0831] 表3提供了可以如何配制、递送、或施用所述组分的实例。

[0832] 表3

[0833]

元件			
一种或多种 Cas9 分子	一种或多种 gRNA 分子	供体模板核酸	注解
DNA	DNA	DNA	在这个实施例中，Cas9 分子（典型地是 eaCas9 分子）、和 gRNA 转录自 DNA。在这个实施例中，它们在单独的分子上进行编码。在这个实施例中，提供所述供体模板作为单独的 DNA 分子。
DNA	DNA		在这个实施例中，Cas9 分子（典型地是 eaCas9 分子）、和 gRNA 转录自 DNA。在这个实施例中，它们在单独的分子上进行编码。在这个实施例中，在编码所述 gRNA 的相同 DNA 分子上提供所述供体模板。
DNA		DNA	在这个实施例中，Cas9 分子（典型地是 eaCas9 分子）、和 gRNA 转录自 DNA，这里转录自单分子。在这个实施例中，提供所述供体模板作为单独的 DNA 分子。
DNA	DNA		在这个实施例中，Cas9 分子（典型地是 eaCas9 分子）、和 gRNA 转录自 DNA。在这个实施例中，它们在单独的分子上进行编

			码。在这个实施例中，在编码所述 Cas9 的相同 DNA 分子上提供所述供体模板。
DNA	RNA	DNA	在这个实施例中，Cas9 分子（典型地是 eaCas9 分子）转录自 DNA，并且 gRNA 被提供为体外转录或合成的 RNA。在这个实施例中，提供所述供体模板作为单独的 DNA 分子。
DNA	RNA		在这个实施例中，Cas9 分子（典型地是 eaCas9 分子）转录自 DNA，并且 gRNA 被提供为体外转录或合成的 RNA。在这个实施例中，在编码所述 Cas9 的相同 DNA 分子上提供所述供体模板。
mRNA	RNA	DNA	在这个实施例中，Cas9 分子（典型地是 eaCas9 分子）翻译自体外转录的 mRNA，并且 gRNA 被提供为体外转录或合成的 RNA。在这个实施例中，提供所述供体模板作为 DNA 分子。
mRNA	DNA	DNA	在这个实施例中，Cas9 分子（典型地是 eaCas9 分子）翻译自体外转录的 mRNA，并且 gRNA 转录自 DNA。在这个实施例中，提供所述供体模板作为单独的 DNA 分子。
mRNA	DNA		在这个实施例中，Cas9 分子（典型地是 eaCas9 分子）翻译自体外转录的 mRNA，并且 gRNA 转录自 DNA。在这个实施例中，在编码所述 gRNA 的相同 DNA 分子上提供所述供体模板。
蛋白质	DNA	DNA	在这个实施例中，Cas9 分子（典型地是 eaCas9 分子）被提供为蛋白质，并且 gRNA 转录自 DNA。在这个实施例中，提供所述供体模板作为单独的 DNA 分子。
蛋白质	DNA		在这个实施例中，Cas9 分子（典型地是 eaCas9 分子）被提供为蛋白质，并且 gRNA 转录自 DNA。在这个实施例中，在编码所述 gRNA 的相同 DNA 分子上提供所述供体模板。
蛋白质	RNA	DNA	在这个实施例中，eaCas9 分子被提供为蛋白质，并且 gRNA 被提供为转录或合成的 RNA。在这个实施例中，提供所述供体模板作为 DNA 分子。

[0834]

[0835]

[0836] 表4概括了Cas系统的组分(例如,如本文所述的Cas9分子组分和gRNA分子组分)的

各种递送方法。

[0837] 表4

[0838]

递送载体/方式		递送到非分裂细胞中	表达的持续时间	基因组整合	所递送分子的类型
物理的（例如，电穿孔、基因枪、磷酸钙转染、细胞压缩或挤压）		是	瞬时的	否	核酸和蛋白
病毒	逆转录病毒	否	稳定的	是	RNA
	慢病毒	是	稳定的	是/否具有修饰	RNA
	腺病毒	是	瞬时的	否	DNA
	腺相关病毒（AAV）	是	稳定的	否	DNA
	痘苗病毒	是	非常瞬时的	否	DNA
	单纯疱疹病毒	是	稳定的	否	DNA
非病毒	阳离子脂质体	是	瞬时的	取决于递送什么	核酸和蛋白
	聚合物纳米粒子	是	瞬时的	取决于递送什么	核酸和蛋白
生物非病毒递送运载体	减毒细菌	是	瞬时的	否	核酸
	工程化的噬菌体	是	瞬时的	否	核酸
	哺乳动物病毒样粒子	是	瞬时的	否	核酸
	生物脂质体：	是	瞬时的	否	核酸

[0839]

	红细胞血影和外泌体				

[0840] 基于DNA的RNA指导的核酸酶和或一个或多个gRNA分子的递送

[0841] 编码RNA指导的核酸酶，例如，Cas9分子（例如，eaCas9分子）、gRNA分子的核酸、供体模板核酸或其任何组合（例如，两种或全部）可以通过本领域已知的方法或如在本文中所述施用于受试者或递送至细胞中。例如，编码Cas9和/或编码gRNA的DNA、以及供体模板核酸可以通过，例如，载体（例如，病毒或非病毒载体）、非基于载体的方法（例如，使用裸DNA或DNA复合物）、或其组合进行递送。

[0842] 编码Cas9分子（例如，eaCas9分子）和/或gRNA分子的核酸可以与促进靶细胞（例

如,红细胞,HSC)的摄取的分子(例如,N-乙酰半乳糖胺)进行缀合。供体模板分子同样可以与促进靶细胞(例如,红细胞,HSC)摄取的分子(例如,N-乙酰半乳糖胺)偶联。

[0843] 在一些实施例中,编码Cas9和/或gRNA的DNA是通过载体(例如,病毒载体/病毒或质粒)进行递送的。

[0844] 载体可包含编码Cas9分子和/或gRNA分子的序列和/或与靶向区域(例如,靶向序列)具有高度同源性的供体模板。在某些实施例中,供体模板包含靶序列的全部或部分。示例性供体模板是修复模板,例如基因校正模板或基因突变模板,例如点突变(例如单核苷酸(nt)取代)模板。载体还可以包括编码融合到例如Cas9分子序列上的信号肽(例如,用于核定位、核仁定位、线粒体定位)的序列。例如,载体可以包括融合到编码Cas9分子的序列上的核定位序列(例如,来自SV40)。

[0845] 可以在载体中包括一个或多个调控/控制元件,例如启动子、增强子、内含子、多聚腺苷酸化信号、Kozak共有序列或内部核糖体进入位点(IRES)。在一些实施例中,启动子由RNA聚合酶II识别(例如,CMV启动子)。在其他实施例中,启动子被RNA聚合酶III识别(例如,U6启动子)。在一些实施例中,启动子是受调控的启动子(例如,诱导型启动子)。在其他的实施例中,启动子是组成型启动子。在一些实施例中,启动子是组织特异性启动子。在一些实施例中,启动子是病毒启动子。在其他的实施例中,启动子是非病毒启动子。

[0846] 在一些实施例中,载体是病毒载体(例如,用于产生重组病毒)。在一些实施例中,病毒是DNA病毒(例如,dsDNA或ssDNA病毒)。在其他实施例中,所述病毒是RNA病毒(例如,ssRNA病毒)。在一些实施例中,病毒感染分裂细胞。在其他的实施例中,病毒感染非分裂细胞。示例性病毒载体/病毒包括,例如,逆转录病毒、慢病毒、腺病毒、腺相关病毒(AAV)、痘苗病毒、痘病毒、以及单纯疱疹病毒。

[0847] 在一些实施例中,病毒感染分裂和非分裂细胞两者。在一些实施例中,病毒可以整合到宿主基因组中。在一些实施例中,病毒被工程化以具有降低的免疫性(例如,在人类中)。在一些实施例中,病毒是有复制能力的。在其他的实施例中,病毒是复制缺陷型的(例如,另外多轮的病毒粒子复制和/或包装所需的基因的一个或多个编码区域被其他基因替换或缺失)。在一些实施例中,病毒引起Cas9分子和/或gRNA分子的瞬时表达。在其他的实施例中,病毒引起Cas9分子和/或gRNA分子的持久(例如,至少1周、2周、1个月、2个月、3个月、6个月、9个月、1年、2年、或永久)表达。病毒的包装能力可以在,例如,至少约4kb到至少约30kb(例如,至少约5kb、10kb、15kb、20kb、25kb、30kb、35kb、40kb、45kb或50kb)之间变化。

[0848] 在一个实施例中,病毒载体识别特定的细胞类型或组织。例如,病毒载体可以用不同/替代的病毒包膜糖蛋白进行假型包装;用细胞类型特异性受体进行工程化(例如,对一种或多种病毒包膜糖蛋白进行一个或多个遗传修饰以结合靶向配体(例如肽配体、单链抗体或生长因子));和/或进行工程化以具有双重特异性的分子桥,其中一端识别病毒糖蛋白而另一端识别靶细胞表面的部分(例如,配体-受体、单克隆抗体、亲和素-生物素和化学缀合)。

[0849] 在一些实施例中,通过重组逆转录病毒递送编码Cas9和/或gRNA的核酸序列。在一些实施例中,逆转录病毒(例如,莫洛尼鼠白血病病毒)包括(例如,允许整合进宿主基因组中的)逆转录酶。在一些实施例中,逆转录病毒是有复制能力的。在其他的实施例中,逆转录病毒是复制缺陷型的(例如,另外多轮的病毒粒子复制和包装所需的基因的一个或多个编

码区域被其他基因替换或缺失)。

[0850] 在一些实施例中,编码Cas9和/或gRNA的核酸序列是通过重组慢病毒进行递送的。在一个实施例中,供体模板核酸通过重组逆转录病毒递送。例如,慢病毒是复制缺陷型的(例如,不包含一种或多种病毒复制所需的基因)。

[0851] 在一个实施例中,通过重组慢病毒递送编码Cas9和/或gRNA的核酸序列。在一个实施例中,供体模板核酸通过重组慢病毒递送。例如,慢病毒是复制缺陷型的(例如,不包含一种或多种病毒复制所需的基因)。

[0852] 在一些实施例中,编码Cas9和/或gRNA的核酸序列是通过重组腺病毒进行递送的。在一个实施例中,供体模板核酸通过重组腺病毒递送。在一些实施例中,腺病毒被工程化以在人类中具有减少的免疫性。

[0853] 在一些实施例中,编码Cas9和/或gRNA的核酸序列是通过重组AAV进行递送的。在一个实施例中,供体模板核酸通过重组AAV递送。在一些实施例中,AAV不将其基因组整合到宿主细胞,例如,本文所述的靶细胞的基因组中。在一些实施例中,AAV可以将其基因组结合到宿主细胞的基因组中。在一些实施例中,AAV是自我互补腺有关病毒(scAAV)(例如,对一起退火以形成双链DNA的两条链进行包装的scAAV)。

[0854] 在一个实施例中,可用于本文所述方法的AAV衣壳是来自血清型AAV1、AAV2、AAV3、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAV.rh8、AAV.rh10、AAV.rh32/33、AAV.rh43、AAV.rh64R1、或AAV7m8的衣壳序列。

[0855] 在一个实施例中,编码Cas9和/或gRNA的DNA以再工程化的AAV衣壳递送,其例如,与来自血清型AAV1、AAV2、AAV3、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAV.rh8、AAV.rh10、AAV.rh32/33、AAV.rh43、或AAV.rh64R1的衣壳序列具有50%或更大,例如,60%或更大、70%或更大、80%或更大、90%或更大、或95%或更大的序列同源性。

[0856] 在一个实施例中,编码Cas9和/或gRNA的DNA通过嵌合AAV衣壳递送。在一个实施例中,供体模板核酸通过嵌合AAV衣壳递送。示例性的嵌合AAV衣壳包括但不限于AAV9i1、AAV2i8、AAV-DJ、AAV2G9、AAV2i8G9、或AAV8G9。

[0857] 在实施例中,AAV是自我互补腺相关病毒(scAAV)(例如,对一起退火以形成双链DNA的两条链进行包装的scAAV)。

[0858] 在一些实施例中,编码Cas9和/或gRNA的DNA是通过混合病毒(例如,在此描述的一种或多种病毒的混合物)进行递送的。在一个实施例中,杂合病毒是AAV(例如,任何AAV血清型)与人类博卡病毒、B19病毒、猪AAV、鹅AAV、猫AAV、犬AAV或MVM的杂合体。

[0859] 使用包装细胞形成能够感染靶细胞的病毒粒子。示例性包装细胞包括293细胞,其可以包装腺病毒,和ψ2或PA317细胞,其可以包装逆转录病毒。在基因治疗中使用的病毒载体通常由将核酸载体包装进病毒粒子的生产者细胞系产生。载体典型地含有包装以及随后整合进宿主或靶细胞(如果适用的话)所需的最低量病毒序列,而其他病毒序列由编码有待表达的蛋白质(例如,Cas9)的表达盒替换。例如,在基因治疗中使用的AAV载体典型地仅具有来自AAV基因组的反向末端重复(ITR)序列,所述序列为包装并在宿主或靶细胞中基因表达所需。如在“三重转染方案”中所述,缺失的病毒功能可以通过包装细胞系和/或含有来自腺病毒的E2A、E4和VA基因的质粒以及编码来自AAV的Rep和Cap基因的质粒来反式地提供。此后,将病毒DNA包装进以下细胞系中,所述细胞系包含编码辅助性质粒的其他AAV基因,即

rep和cap,但缺少ITR序列。在某些实施例中,将病毒DNA包装在含有来自腺病毒的E1A和/或E1B基因的生产者细胞系中。所述细胞系还被作为辅助者的腺病毒感染。辅助性病毒(例如,腺病毒或HSV)或辅助性质粒促进AAV载体的复制和从有ITR的辅助质粒表达AAV基因。由于缺少ITR序列,未以显著的量包装所述辅助性质粒。可以通过例如与AAV相比腺病毒更加敏感的热处理减少腺病毒的污染。

[0860] 在某些实施例中,病毒载体能够进行细胞类型和/或组织类型识别。例如,病毒载体可以用不同/替代的病毒包膜糖蛋白进行假型包装;用细胞类型特异性受体进行工程化(例如,对病毒包膜糖蛋白进行遗传修饰以结合靶向配体(例如,肽配体、单链抗体、或生长因子));和/或进行工程化以具有双重特异性的分子桥,其中一端识别病毒糖蛋白而另一端识别靶细胞表面的部分(例如,配体-受体、单克隆抗体、亲和素-生物素和化学缀合)。

[0861] 在某些实施例中,病毒载体实现细胞类型特异性表达。例如,可以构建组织特异性启动子以仅在靶细胞中限制转基因(Cas9和gRNA)的表达。载体的特异性也可以由转基因表达的微小RNA依赖性控制所介导。在实施例中,病毒载体具有增加的病毒载体和靶细胞膜的融合效率。例如,可以结合融合蛋白(例如,融合感受态血球凝集素(HA))以增加病毒摄取进入细胞中。在实施例中,病毒载体具有核定位的能力。例如,可以将需要分解核膜(在细胞分裂期间)并且因此将不感染非分裂细胞的病毒改变成结合病毒的基质蛋白中的核定位肽,由此能够实现非增殖细胞的转导。

[0862] 在一些实施例中,编码Cas9和/或gRNA的DNA是通过非基于载体的方法(例如,使用裸DNA或DNA复合物)进行递送的。例如,DNA可以例如,通过有机改性的二氧化硅或硅酸盐(Ormosil)、电穿孔、瞬时细胞压缩或挤压(参见例如,Lee2012)、基因枪、声致穿孔、磁转染、脂质介导的转染、树枝状聚合物、无机纳米粒子、磷酸钙、或其组合进行递送。

[0863] 在实施例中,经由电穿孔递送包括将细胞与编码Cas9和/或gRNA的DNA在盒、室或比色皿中混合并且施加一个或多个限定持续时间和幅度的电脉冲。在一个实施例中,使用如下系统进行经由电穿孔的递送,在所述系统中将细胞与编码Cas9和/或gRNA的DNA在连接至装置(例如,泵)的容器中混合,所述装置向盒、室或比色皿中给料混合物,在所述盒、室或比色皿中施加一个或多个限定持续时间和幅度的电脉冲,之后将细胞递送至第二容器。

[0864] 在一些实施例中,编码Cas9和/或gRNA的DNA是通过载体和非基于载体的方法的组合进行递送的。在一个实施例中,供体模板核酸通过基于载体和非载体的方法的组合来递送。例如,病毒体将脂质体与灭活的病毒(例如,HIV或流感病毒)结合,其可导致比单独的病毒或脂质体方法更有效的基因转移,例如,在呼吸道上皮细胞中。

[0865] 在某些实施例中,递送载体是非病毒载体,并且在某些这些实施例中,非病毒载体是无机纳米颗粒。示例性无机纳米颗粒包括,例如,磁性纳米颗粒(例如,Fe_3MnO_2)、或二氧化硅。可以将纳米颗粒的外表面与带正电荷的聚合物(例如,聚乙烯亚胺、聚赖氨酸、聚丝氨酸)缀合,这允许有效载荷的附接(例如,缀合或截留)。在实施例中,非病毒载体是有机纳米颗粒(例如,截留纳米颗粒内的有效载荷)。示例性有机纳米颗粒包括,例如,包含阳离子脂质连同中性辅助脂质的SNALP脂质体,其涂覆有聚乙二醇(PEG)和鱼精蛋白以及涂覆有脂质包衣的核酸复合物。

[0866] 用于基因转移的示例性脂质示于下表1中。

[0867] 表1:用于基因转移的脂质

[0868]

脂质	缩写	特征
1,2-二油酰基-sn-甘油基-3-磷脂酰胆碱	DOPC	辅助细胞
1,2-二油酰基-sn-甘油基-3-磷脂酰乙醇胺	DOPE	辅助细胞
胆固醇		辅助细胞
N-[1-(2,3-二油烯基氧基)丙基]N,N,N-三甲基氯化铵	DOTMA	阳离子的
1,2-二油烯基氧基-3-三甲基铵-丙烷	DOTAP	阳离子的
双十八烷基氨基甘氨酰精胺	DOGS	阳离子的
N-(3-氨基丙基)-*N,N*-二甲基-2,3-双(十二烷氧基)-1-溴化丙铵	GAP-DLRIE	阳离子的
溴化十六烷基三甲基铵	CTAB	阳离子的
6-月桂氧己基乌氨酸	LHON	阳离子的
1-(2,3-二油酰基氧基丙基)-2,4,6-三甲基吡啶	2Oc	阳离子的
2,3-二油烯基氧基-*N*-[2(精胺甲酰胺基-乙基]-*N,N*-二甲基-1-三氟乙酸丙铵	DOSPA	阳离子的
1,2-二油烯基-3-三甲基铵-丙烷	DOPA	阳离子的
N-(2-羟基乙基)-*N,N*-二甲基-2,3-双(十四烷氧基)-1-溴化丙铵	MDRIE	阳离子的
二肉豆蔻酰氧丙基 二甲基 羟乙基 溴化铵	DMRI	阳离子的
3β-[*N*-(*N'*,*N'*-二甲基氨基乙烷)-氨甲酰基]胆固醇	DC-Chol	阳离子的
双-胍鎓-三氨乙基胺-胆固醇	BGTC	阳离子的

1,3-双脱氧-2-(6-羧基-精胺基)-丙酰胺	DOSPER	阳离子的
二甲基十八烷基溴化铵	DDAB	阳离子的
双十八烷基氨基甘氨酰基亚精胺	DSL	阳离子的
外消旋-[(2,3-双十八烷氧基丙基)(2-羟乙基)]-二甲基氯化铵	CLIP-1	阳离子的
外消旋-[2(2,3-双十六烷氧基丙基-氧甲基氧甲基)乙基]三甲基溴化铵	CLIP-6	阳离子的
乙基二肉豆蔻酰磷脂酰胆碱	EDMPC	阳离子的
1,2-二硬脂氧基-*N,N*-二甲基-3-氨基丙烷	DSDMA	阳离子的
1,2-二肉豆蔻酰基-三甲基铵丙烷	DMTAP	阳离子的
O,O'-二肉豆蔻基-N-赖氨酰-天冬氨酸酯	DMKE	阳离子的
[0869] 1,2-二硬脂酰-sn-甘油基-3-乙基胆碱磷酸	DSEPC	阳离子的
N-棕榈酰 D-赤-鞘氨苷氨甲酰基-精胺	CCS	阳离子的
N-叔丁基-*N*0-十四烷基-3-十四烷基氨基丙脒	二 C14-脒	阳离子的
十八碳烯醇基氧基[乙基-2-十七烯基-3 羟乙基]氯化咪唑啉鎓	DOTIM	阳离子的
*N*1-胆固醇氧基羰基-3,7-二氮杂壬烷-1,9-二胺	CDAN	阳离子的
2-(3-[双(3-氨基-丙基)-氨基]丙胺基)-*N*-双十四烷基氨基甲酰基甲-乙基-乙酰胺	RPR209120	阳离子的
1,2-二亚油基氧基-3-二甲基氨基丙烷	DLinDMA	阳离子的
2,2-二亚油基-4-二甲基氨基乙基-[1,3]-二氧戊环	DLin-KC2-DMA	阳离子的
二亚油基-甲基-4-二甲基氨基丁酸酯	DLin-MC3-DMA	阳离子的

[0870] 用于基因转移的示例性聚合物示于下表5。

[0871] 表5:用于基因转移的聚合物

聚合物	缩写
聚乙二醇	PEG
聚乙烯亚胺	PEI
二硫代双(琥珀酰亚胺基丙酸酯)	DSP
[0872] 二甲基-3,3'-二硫代双丙亚氨酸酯	DTBP
聚(乙烯亚胺)双氨基甲酸酯	PEIC
聚(L-赖氨酸)	PLL
组氨酸修饰的 PLL	
聚(*N*-乙烯基吡咯烷酮)	PVP

[0873]

聚(丙烯亚胺)	PPI
聚(酰胺胺)	PAMAM
聚(酰胺基乙烯亚胺)	SS-PAEI
三亚乙基四胺	TETA
聚(β-氨基酯)	
聚(4-羟基-L-脯氨酸酯)	PHP
聚(烯丙基胺)	
聚(α-[4-氨基丁基]-L-乙醇酸)	PAGA
聚(D,L-乳酸-共-乙醇酸)	PLGA
聚(*N*-乙基-4-乙烯基吡啶鎓溴化物)	
聚(磷腈)	PPZ
聚(磷酸酯)	PPE
聚(氨基磷酸酯)	PPA
聚(*N*-2-羟基丙基甲基丙烯酰胺)	pHPMA
聚(2-(二甲氨基)甲基丙烯酸乙酯)	pDMAEMA
聚(2-氨乙基丙烯磷酸酯)	PPE-EA
壳聚糖	
半乳糖基化壳聚糖	
N-十二烷基化的壳聚糖	
组蛋白	
胶原	
葡聚糖-精胺	D-SPM

[0874] 在一个实施例中,载体具有靶向修饰以增加靶细胞摄入纳米粒子和脂质体(例如,细胞特异性抗原、单克隆抗体、单链抗体、适配体、聚合物、糖(例如,N-乙酰半乳糖胺(GalNAc))和细胞穿透肽)。在实施例中,载体使用融合肽和内体去稳定肽/聚合物。在实施例中,载体经历酸触发的构象变化(例如,以加速负荷物的内体逃逸)。在实施例中,使用刺激可切割的聚合物,例如,用于在细胞区室中释放。例如,可以使用在还原性细胞环境中被切割的基于二硫化物的阳离子型聚合物。

[0875] 在实施例中,递送运载体是生物非病毒递送运载体。在实施例中,运载体是减毒细菌(例如,天然或人工工程化成待侵入性的,但减毒以防止发病原和表达转基因(例如,单核细胞增生李斯特菌、某些沙门氏菌菌株、长双歧杆菌、和修饰的大肠杆菌)、具有营养和组织特异性向性以靶向特定组织的细菌、具有修饰的表面蛋白以改变靶组织特异性的细菌)。在实施例中,运载体是转基因噬菌体(例如,具有大包装能力、较少免疫原性、含有哺乳动物质粒维持序列并且具有结合的靶向配体的工程化噬菌体)。在实施例中,运载体是哺乳动物病毒样颗粒。例如,可以产生修饰的病毒颗粒(例如,通过纯化“空心”颗粒,随后用希望的负荷物离体组装病毒)。运载体也可以被工程化以结合靶向配体从而改变靶组织特异性。在实施例中,运载体是生物脂质体。例如,生物脂质体是衍生自人类细胞的基于磷脂的颗粒(例如,红细胞血影,其是所述红血细胞分解成衍生自受试者的球状结构(例如,可以通过附接不同组织或细胞特异性配体来实现组织靶向)、或分泌外泌体-受试者(即,患者)衍生的内吞起源的膜结合纳米运载体(30nm-100nm)(例如,可以产生自不同细胞类型并且因此可以被细

胞吸收,而不需要靶向配体)。

[0876] 在一个实施例中,递送除了Cas系统的所述组分(例如,Cas9分子组分和/或在此描述的gRNA分子组分)外的一种或多种核酸分子(例如,DNA分子)。在实施例中,该核酸分子是在递送Cas系统一个或多个组分的同时进行递送的。在实施例中,该核酸分子是在递送Cas系统一个或多个组分之前或之后(例如,小于约30分钟、1小时、2小时、3小时、6小时、9小时、12小时、1天、2天、3天、1周、2周、或4周)进行递送的。在实施例中,该核酸分子是通过不同于递送Cas系统的一个或多个组分(例如,Cas9分子组分和/或gRNA分子组分)的方式进行递送的。该核酸分子可以通过任何在此描述的递送方法进行递送。例如,该核酸分子可以通过病毒载体(例如,整合缺陷型慢病毒)进行递送,并且Cas9分子组分和/或gRNA分子组分可以通过电穿孔(例如,这样可以降低由核酸(例如,DNA)引起的毒性)进行递送。在实施例中,该核酸分子编码治疗性蛋白质(例如,在此描述的蛋白质)。在实施例中,该核酸分子编码RNA分子(例如,在此描述的RNA分子)。

[0877] 编码RNA指导的核酸酶的RNA的递送

[0878] 编码RNA指导的核酸酶(例如,Cas9分子)和/或gRNA分子的RNA可以通过本领域已知的方法或如本文所述递送到细胞中,例如,本文所述的靶向细胞。例如,Cas9编码和/或gRNA编码RNA可以例如,通过显微注射、电穿孔、瞬时细胞压缩或挤压(参见例如,Lee 2012)、脂质介导的转染、肽介导的递送或其组合进行递送。编码Cas9的和/或编码gRNA的RNA可以与分子偶联,从而促进靶细胞(例如,本文所述的靶细胞)的摄取。

[0879] 在一个实施例中,经由电穿孔递送包括将细胞与编码Cas9分子和/或gRNA分子(具有或不具有供体模板核酸分子)在盒、室或比色皿中混合并且施加一个或多个限定持续时间和幅度的电脉冲。在一个实施例中,使用如下系统进行经由电穿孔的递送,在所述系统中将细胞与RNA编码Cas9分子和/或gRNA分子、与或不与供体模板核酸分子在连接至装置(例如,泵)的容器中混合,所述装置向盒、室或比色皿中供给混合物,在所述盒、室或比色皿中施加一个或多个限定持续时间和幅度的电脉冲,之后将细胞递送至第二容器。编码Cas9和/或编码gRNA的RNA可以与分子偶联以促进靶细胞(例如,本文所述的靶细胞)的摄取。

[0880] RNA指导的核酸酶的递送

[0881] RNA指导的核酸酶,例如,Cas9分子可以通过领域已知的方法或如本文所述的递送到细胞中。例如,Cas9蛋白分子可以例如通过显微注射、电穿孔、瞬时细胞压缩或挤压(参见例如,Lee 2012)、脂质介导的转染、肽介导的递送或其组合进行递送。递送可以与编码gRNA的DNA或与gRNA相伴。Cas9蛋白可以与促进靶细胞(例如本文所述的靶细胞)摄取的分子进行缀合。

[0882] 在一个实施例中,经由电穿孔递送包括将细胞与Cas9分子和/或gRNA分子,与或不与供体核酸在盒、室或比色皿中混合并且施加一个或多个限定持续时间和幅度的电脉冲。在一个实施例中,使用如下系统进行经由电穿孔的递送,在所述系统中将细胞与Cas9分子和/或gRNA分子、与或不与供体核酸在连接至装置(例如,泵)的容器中混合,所述装置向盒、室或比色皿中供给混合物,在所述盒、室或比色皿中施加一个或多个限定持续时间和幅度的电脉冲,之后将细胞递送至第二容器。编码Cas9和/或编码gRNA的RNA可以与分子偶联以促进靶细胞(例如,本文所述的靶细胞)的摄取。

[0883] 基因组编辑系统组分的施用途径

[0884] 全身施用模式包括口服和肠胃外途径。肠胃外途径包括,举例来说,静脉内、骨髓内、动脉内、肌内、皮内、皮下、鼻内以及腹膜内途径。可以修饰或配制全身施用的组分以靶向例如,HSC,造血干细胞/祖细胞或红系祖细胞或前体细胞。

[0885] 举例来说,局部施用模式包括,骨髓内注射到骨小梁中或股骨内注射到髓隙中,以及输注到门静脉中。在一个实施例中,与当全身施用(例如,静脉内)相比时,显著较少量的组分(与全身途径相比)可以在局部施用(例如,直接进入骨髓)时发挥作用。局部施用模式可以降低或消除潜在毒副作用的发生率,当全身性施用治疗有效量的组分时可能发生毒副作用。

[0886] 可以以周期性推注(例如静脉内)的形式,或者从内部储库或外部储库(例如从静脉注射袋或可植入泵)持续输注提供施用。组分可以局部施用,例如,通过从持续释放药物递送装置中持续释放。

[0887] 此外,可以将组分配制成允许在延长时段内释放。释放系统可以包括生物降解材料或通过扩散释放所掺入的组分的材料的基质。所述组分可以在释放系统中均匀或者非均匀分配。各种释放系统可能是有用的,可根据特定应用所需的释放速率选择合适的系统。不可降解和可降解的释放系统均可以被使用。适合的释放系统包括聚合物和聚合物基质、非聚合物基质、或无机和有机赋形剂和稀释剂(如,但不局限于,碳酸钙和糖(例如,海藻糖))。释放系统可以是天然的或合成的。然而,合成的释放系统是优选的,因为它们通常更可靠、更具可重现性并且产生更多定义的释放曲线。可以选择释放系统材料以使得具有不同分子量的组分是通过扩散或通过材料的降解而释放的。

[0888] 代表性合成的、可生物降解的聚合物包括,例如:聚酰胺(例如,聚(氨基酸)和聚(肽));聚酯(如聚(乳酸)、聚(乙醇酸)、聚(乳酸-共-乙醇酸)、以及聚(己内酯));聚(酸酐);聚原酸酯;聚碳酸酯;及其化学衍生物(化学基团的取代、添加,例如,烷基、亚烷基、羟化、氧化、以及其他由本领域的技术人员常规做出的修饰)、共聚物及其混合物。代表性合成的,不可降解聚合物包括,例如:聚醚(例如,聚(氧化乙烯)、聚(乙二醇)、以及聚(四氢呋喃));乙烯基聚合物-聚丙烯酸酯和聚甲基丙烯酸酯(如甲基、乙基、其他烷基、甲基丙烯酸羟基乙酯、丙烯酸和甲基丙烯酸、以及其他例如聚(乙烯醇)、聚(乙烯基吡咯烷酮)、以及聚(乙酸乙烯酯);聚(氨酯);纤维素及其衍生物(如烷基、羟烷基、醚、酯、硝化纤维素、以及不同的醋酸纤维素);聚硅氧烷;及其任何化学衍生物(化学基团的取代、添加,例如,烷基、亚烷基、羟化、氧化、以及其他由本领域的技术人员常规做出的修饰)、共聚物及其混合物。

[0889] 也可使用聚丙交酯乙交酯共聚物微球。典型地,微球是由乳酸和乙醇酸的聚合物构成的,其被结构化形成空心球体。球体的直径可以是大约15-30微米并且可以加载在此描述的组分。

[0890] <u>基因组编辑系统组分的双模或差别递送</u>

[0891] Cas系统组分的分别递送,例如,Cas9分子组分和gRNA分子组分,并且更具体地,通过不同模式递送这些组分,可以通过,例如,改进组织特异性和安全性来增强性能。

[0892] 在某些实施例中,Cas9分子和gRNA分子是通过不同模式(或有时在此称为差别模式)进行递送的。如在此使用的,不同或差别模式是指递送的模式,所述递送模式赋予受试组分分子,例如,Cas9分子、gRNA分子、模板核酸或有效负载不同的药效动力学或药物代谢动力学特性。例如,递送的模式可以导致不同的组织分布,不同的半衰期、或不同的时间分

布(例如,在选定的区室、组织、或器官中)。

[0893] 一些递送的模式(例如,通过例如通过自主复制或插入进细胞核酸中而持续存在于细胞、或细胞子代中的核酸载体的递送)导致组分更为持续的表达和存在。实例包括病毒,例如,AAV或慢病毒递送。

[0894] 通过举例,这些组分,例如,Cas9分子和gRNA分子,可以在所递送组分在体内、或在特定区域室、组织、或器官中的所得半衰期或持续性方面不同的模式进行递送。在实施例中,gRNA分子可以通过此类模式进行递送。Cas9分子组分可以通过这样一种模式进行递送,所述模式导致更少持久性或更少暴露于身体或特定区域室或组织或器官。

[0895] 更一般地说,在实施例中,第一递送模式被用于递送第一组分并且第二递送模式被用于递送第二组分。第一递送模式赋予第一药效动力学或药物代谢动力学特性。第一药效动力学特性可以是,例如,组分或编码该组分的核酸在体内、区室、组织、或器官中的分布、持久性、或暴露。第二递送模式赋予第二药效动力学或药物代谢动力学特性。第二药效动力学特性可以是,例如,组分或编码该组分的核酸在体内、区室、组织、或器官中的分布、持久性、或暴露。

[0896] 在某些实施例中,第一药效动力学或药物代谢动力学特性(例如,分布、持久性或暴露)比第二药效动力学或药物代谢动力学特性更为有限。

[0897] 在某些实施例中,选择第一递送模式以优化(例如,最小化)药效动力学或药物代谢动力学特性(例如,分布、持久性或暴露)。

[0898] 在某些实施例中,选择第二递送模式以优化(例如,最大化)药效动力学或药物代谢动力学特性(例如,分布、持久性或暴露)。

[0899] 在某些实施例中,第一递送模式包括使用较为持久的元件(例如,核酸(例如,质粒或病毒载体(例如,AAV或慢病毒)))。由于此类载体是较为持久的,从它们中转录的产物将是较为持久的。

[0900] 在某些实施例中,第二递送模式包括较为短暂的元件(例如,RNA或蛋白)。

[0901] 在某些实施例中,第一组分包含gRNA,并且所述递送模式是较为持久的(例如,gRNA转录自质粒或病毒载体(例如,AAV或慢病毒))。这些基因的转录将具有很小的生理学意义,因为基因并不编码蛋白产物,并且这些gRNA不能够单独起作用。第二组分(Cas9分子)是以瞬时方式进行递送的(例如,作为mRNA或作为蛋白),从而确保完全Cas9分子/gRNA分子复合物仅在短时段内存在和有活性。

[0902] 此外,这些组分可以不同的分子形式或用不同的互为补充以增强安全性和组织特异性的递送载体进行递送。

[0903] 差别递送模式的使用可以增强性能、安全性和/或功效,例如,可以减少最终脱靶修饰的可能性。通过较不持久的模式递送免疫原性组分(例如,Cas9分子)可以降低免疫原性,因为来自细菌衍生的Cas酶的肽通过MHC分子被展示在细胞表面上。两部分式递送系统可以改善这些缺点。

[0904] 差别递送模式可以用于将组分递送至不同,但是重叠的靶区域。在靶区域的重叠之外形成活性复合物被最小化。因此,在实施例中,第一组分(例如,gRNA分子)通过第一递送模式进行递送,其导致第一空间(例如,组织)分布。第二组分(例如,Cas9分子)通过第二递送模式进行递送,其导致第二空间(例如,组织)分布。在一个实施例中,第一模式包括选

自脂质体、纳米颗粒(例如,聚合物纳米颗粒)、以及核酸(例如,病毒载体)的第一元件。第二模式包括选自下组的第二元件。在实施例中,第一递送模式包括第一靶向元件(例如,细胞特异性受体或抗体),并且第二递送模式不包括该元件。在某些实施例中,第二递送模式包括第二靶向元件(例如,第二细胞特异性受体或第二抗体)。

[0905] 当Cas9分子在病毒递送载体、脂质体、或聚合物纳米颗粒中进行递送时,存在递送至多个组织并且在多个组织具有治疗活性的可能性,这是当可能希望仅靶向单个组织时。两部分式递送系统可以解决这一挑战并且增强组织特异性。如果gRNA分子和Cas9分子被包装在具有不同但重叠的组织嗜性的分离的递送运载体中,完全功能性复合物仅在被两种载体靶向的组织中形成。

[0906] Cas系统组分的离体递送

[0907] 在某些实施例中,将表3中描述的Cas系统组分引入细胞中,然后将其引入受试者中。引入组分的方法可以包括例如表4中所述的任何递送方法。

[0908] 修饰的核苷、核苷酸和核酸

[0909] 修饰的核苷和修饰的核苷酸可以存在于核酸中,例如特别是gRNA,但是还有其他形式的RNA,例如mRNA、RNAi或siRNA。如本文所描述的,“核苷”被定义为包含五碳糖分子(戊糖或核糖)或其衍生物以及有机碱(嘌呤或嘧啶)或其衍生物的化合物。如本文所描述的,“核苷酸”被定义为进一步包含磷酸基团的核苷。

[0910] 修饰的核苷和核苷酸可以包括以下项中的一项或多项:

[0911] (i)磷酸二酯骨架键联中的一个或两个非连接磷酸氧和/或一个或多个连接磷酸氧的改变,例如置换;

[0912] (ii)核糖的组分(例如,核糖上的2’羟基)的改变,例如置换;

[0913] (iii)“脱磷酸”接头对磷酸部分的完全置换;

[0914] (iv)天然存在的核碱基的修饰或置换;

[0915] (v)核糖-磷酸骨架的置换或修饰;

[0916] (vi)寡核苷酸的3’端或5’端的修饰,例如,末端磷酸基团的去除、修饰或置换或部分的结合;以及

[0917] (vii)糖的修饰。

[0918] 以上列出的修饰可以组合,以提供可以具有两个、三个、四个或更多个修饰的修饰的核苷和核苷酸。例如,修饰的核苷或核苷酸可以具有修饰的糖和修饰的核碱基。在一个实施例中,修饰gRNA的每个碱基,例如,所有碱基都具有修饰的磷酸酯基团,例如,所有修饰的磷酸酯基团都是硫代磷酸酯基团。在一个实施例中,单分子(或嵌合)或模块化gRNA分子的所有或基本上所有磷酸酯基团被硫代磷酸酯基团替换。

[0919] 在一个实施例中,可以将修饰的核苷酸(例如,具有如本文描述的修饰的核苷酸)掺入核酸中,例如“修饰的核酸”。在一个实施例中,修饰的核酸包含一个、两个、三个或更多个修饰的核苷酸。在一个实施例中,修饰的核酸中的至少5%(例如,至少约5%、至少约10%、至少约15%、至少约20%、至少约25%、至少约30%、至少约35%、至少约40%、至少约45%、至少约50%、至少约55%、至少约60%、至少约65%、至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、至少约95%或约100%)的位置是修饰的核苷酸。

[0920] 未修饰的核酸可以易于被例如细胞核酸酶降解。例如,核酸酶可以水解核酸磷酸

二酯键。因此,在一个方面中,本文描述的修饰的核酸可以含有一个或多个修饰的核苷或核苷酸,例如,以引入对核酸酶的稳定性。

[0921] 在一个实施例中,本文描述的修饰的核苷、修饰的核苷酸和修饰的核酸当被在体内和离体地引入细胞群体中时都可以展现出减少的先天性免疫应答。术语“先天性免疫应答”包括对外源核酸的细胞应答,外源核酸包括通常是病毒或细菌来源的单链核酸,所述细胞应答涉及细胞因子(特别是干扰素)表达与释放以及细胞死亡的诱导。在一个实施例中,本文描述的修饰的核苷、修饰的核苷酸和修饰的核酸可以破坏大沟相互作用配偶体与核酸的结合。在一个实施例中,本文描述的修饰的核苷、修饰的核苷酸和修饰的核酸当被在体内和离体地引入细胞群体中时都可以展现出减少的先天性免疫应答,并且还破坏大沟相互作用配偶体与核酸的结合。

[0922] 化学基团的定义

[0923] 如本文所用的,“烷基”意在指直链的或支链的饱和烃基。示例性烷基基团包括甲基(Me)、乙基(Et)、丙基(例如,正丙基和异丙基)、丁基(例如,正丁基、异丁基、叔丁基)、戊基(例如,正戊基、异戊基、新戊基)等。烷基基团可以包含从1至约20、从2至约20、从1至约12、从1至约8、从1至约6、从1至约4或从1至约3个碳原子。

[0924] 如在此所使用的,“芳基”是指单环或多环(例如,具有2、3或4个稠环)的芳香烃,例如像苯基、萘基、蒽基、菲基、茚满基、茚基等。在一个实施例中,芳基基团具有从6至约20个碳原子。

[0925] 如本文所用的,“烯基”是指包含至少一个双键的脂肪族基团。

[0926] 如本文所用的,“炔基”是指含有2-12个碳原子并且特征在于具有一个或多个三键的直链的或支链的烃链。炔基基团的实例包括但不限于乙炔基、炔丙基和3-己炔基。

[0927] 如本文所用的,“芳基烷基”或“芳烷基”是指烷基氢原子被芳基基团替换的烷基部分。芳烷基包括一个以上氢原子已经被芳基基团置换的基团。“芳基烷基”或“芳烷基”的实例包括苄基、2-苯基乙基、3-苯基丙基、9-芴基、二苯甲基以及三苯甲基基团。

[0928] 如本文所用的,“环烷基”是指具有3至12个碳的环状的、二环的、三环的或多环的非芳香烃基团。环烷基部分的实例包括但不限于环丙基、环戊基和环己基。

[0929] 如本文所用的,“杂环基”是指杂环系统的单价基。代表性杂环基包括但不限于四氢呋喃基、四氢噻吩基、吡咯烷基、吡咯烷酮基、哌啶基、吡咯啉基、哌嗪基、二噁烷基、二氧戊环基、二氮杂卓基、氧氮杂卓基、硫氮杂卓基以及吗啉基。

[0930] 如本文所用的,“杂芳基”是指杂芳香环系统的单价基。杂芳基部分的实例包括但不限于咪唑基、噁唑基、噻唑基、三唑基、吡咯基、呋喃基、吲哚基、苯硫基、吡唑基、吡啶基、吡嗪基、哒嗪基、嘧啶基、吲嗪基、嘌呤基、萘啶基、喹啉基以及蝶啶基。

[0931] 磷酸骨架修饰

[0932] 磷酸酯基团

[0933] 在一个实施例中,可以通过用不同取代基置换一个或多个氧来修饰经修饰的核苷酸的磷酸酯基团。此外,修饰的核苷酸(例如,存在于修饰的核酸中的修饰的核苷酸)可以包括如本文描述的修饰的磷酸酯对未修饰的磷酸酯部分的完全置换。在一个实施例中,磷酸骨架的修饰可以包括产生不带电接头或具有不对称电荷分布的带电接头的改变。

[0934] 修饰的磷酸酯基团的实例包括硫代磷酸酯、硒代磷酸酯(phosphoroselenate)、硼

磷酸酯(borano phosphate)、硼磷酸酯(borano phosphate ester)、氢膦酸酯、磷酰胺酯(phosphoroamidate)、烷基或芳基膦酸酯和磷酸三酯。在一个实施例中,磷酸骨架部分中的非桥连磷酸氧原子之一可以被以下基团中的任一项替换:硫(S)、硒(Se)、BR_3(其中R可以是例如氢、烷基或芳基)、C(例如,烷基基团、芳基基团等)、H、NR_2(其中R可以是例如氢、烷基或芳基)或OR(其中R可以是例如烷基或芳基)。未修饰的磷酸酯基团中的磷原子是非手性的。然而,以上原子或原子的基团之一对非桥连磷酸氧之一的置换可以使得磷原子是手性的;也就是说以这种方式修饰的磷酸酯基团中的磷原子是立构中心。立构磷原子可以具有“R”构型(本文是Rp)或“S”构型(本文是Sp)。

[0935] 二硫代磷酸酯具有两个被硫置换的非桥连氧。二硫代磷酸酯中的磷中心是非手性的,这阻止寡核糖核苷酸非对映异构体的形成。在一个实施例中,对一个或两个非桥连氧的修饰还可以包括用以下基团替换非桥连氧,所述基团独立地选自S、Se、B、C、H、N以及OR(R可以是,例如,烷基或芳基)。

[0936] 还可以通过用氮(桥连的磷酰胺酯)、硫(桥连的硫代磷酸酯)和碳(桥连的亚甲基膦酸酯)置换桥连氧(即,将磷酸连接至核苷的氧)来修饰磷酸酯接头。置换可以发生在连接氧或发生在两个连接氧处。

[0937] 磷酸酯基团的置换

[0938] 磷酸酯基团可以被不含磷连接物置换。在一个实施例中,带电磷酸酯基团可以被中性部分替换。

[0939] 可以替换磷酸酯基团的部分的实例可以包括但不限于例如甲基膦酸酯、羟氨基、硅氧烷、碳酸酯、羧甲基、氨基甲酸酯、酰胺、硫醚、环氧乙烷接头、磺酸酯、磺酰胺、硫代甲缩醛(thioformacetal)、甲缩醛(formacetal)、肟、亚甲亚氨基、亚甲甲基亚氨基、亚甲肼基、亚甲二甲基肼基以及亚甲氧基甲基亚氨基。

[0940] 核糖磷酸骨架的置换

[0941] 还可以构建可以模拟核酸的支架,其中磷酸酯接头和核糖被核酸酶抗性核苷或核苷酸替代物置换。在一个实施例中,可以通过替代骨架拴住核碱基。实例可以包括但不限于吗啉代、环丁基、吡咯烷和肽核酸(PNA)核苷替代物。

[0942] 糖修饰

[0943] 修饰的核苷和修饰的核苷酸可以包括对糖基的一种或多种修饰。例如,2’羟基基团(OH)可以被多种不同的“氧基”或“脱氧”取代基修饰或替换。在一个实施例中,对2’羟基基团的修饰可以增强核酸的稳定性,因为羟基不再可以被去质子化以形成2’-醇盐离子。2’-醇盐可以通过接头磷原子上的分子内亲核攻击而催化降解。

[0944] “氧基”-2’羟基基团修饰的实例可以包括烷氧基或芳氧基(OR,其中“R”可以是,例如,烷基、环烷基、芳基、芳烷基、杂芳基或糖);聚乙二醇(PEG),$O(CH_2CH_2O)_nCH_2CH_2OR$,其中R可以是例如H或任选取代的烷基,并且n可以是从0至20的整数(例如,从0至4、从0至8、从0至10、从0至16、从1至4、从1至8、从1至10、从1至16、从1至20、从2至4、从2至8、从2至10、从2至16、从2至20、从4至8、从4至10、从4至16以及从4至20)。在一个实施例中,“氧基”-2’羟基基团修饰可以包括“锁”核酸(LNA),其中2’羟基可以,例如,通过C_{1-6}亚烷基或C_{1-6}杂亚烷基桥连接至同一核糖的4’碳,其中示例性桥可以包括亚甲基、亚丙基、醚或氨基桥;O-氨基(其中氨基可以是例如NH_2;烷氨基、二烷氨基、杂环基、芳氨基、二芳氨基、杂芳氨基或二杂芳氨

基、乙二胺或聚氨基)和氨基烷氧基O$(CH_2)_n$-氨基(其中氨基可以是例如NH_2;烷氨基、二烷氨基、杂环基、芳氨基、二芳氨基、杂芳氨基或二杂芳氨基、乙二胺或聚氨基)。在一个实施例中,“氧基”-2'羟基基团修饰可以包括甲氧基乙基基团(MOE)($OCH_2CH_2OCH_3$,例如,PEG衍生物)。

[0945] “脱氧”修饰可以包括氢(即脱氧核糖,例如在部分ds RNA的突出端部分);卤素(例如,溴、氯、氟或碘);氨基(其中氨基可以是例如NH_2;烷氨基、二烷氨基、杂环基、芳氨基、二芳氨基、杂芳氨基、二杂芳氨基或氨基酸);$NH(CH_2CH_2NH)_nCH_2CH_2$-氨基(其中氨基可以是,例如,如本文描述的),-NHC(O)R(其中R可以是,例如,烷基、环烷基、芳基、芳烷基、杂芳基或糖),氰基;巯基;烷基-硫代-烷基;硫代烷氧基;以及烷基、环烷基、芳基、烯基和炔基,其可以任选地被例如,如本文描述的氨基取代。

[0946] 糖基还可以包含一个或多个具有与核糖中的对应碳相反的立体化学构型的碳。因此,修饰的核酸可以包括含有例如阿拉伯糖作为糖的核苷酸。核苷酸“单体”可以在糖的1'位置处具有α键联,例如α-核苷。修饰的核酸还可以包括“无碱基”糖,其在C-1'处缺乏核碱基。这些无碱基糖还可以在一个或多个构成性糖原子处被进一步修饰。修饰的核酸还可以包括一种或多种处于L型的糖,例如L-核苷。

[0947] 通常,RNA包括糖基核糖,它是具有氧的5元环。示例性修饰的核苷和修饰的核苷酸可以包括但不限于核糖中氧的置换(例如,用硫(S)、硒(Se)或亚烷基,例如像亚甲基或亚乙基);双键的添加(例如,以用环戊烯基或环己烯基置换核糖);核糖的缩环(例如,以形成环丁烷或氧杂环丁烷的4元环);核糖的扩环(例如,以形成具有另外的碳或杂原子的6元或7元环,例如像脱水己糖醇、阿卓糖醇、甘露醇、环己烷基、环己烯基以及吗啉代,其也具有氨基磷酸酯骨架)。在一个实施例中,修饰的核苷酸可以包括多环形式(例如,三环;和“解锁”形式,如二醇核酸(GNA)(例如,R-GNA或S-GNA,其中核糖被附接至磷酸二酯键的二醇单元替换),或苏糖核酸(TNA,其中核糖被α-L-苏呋喃糖基-(3'→2')替换)。

[0948] <u>核碱基上的修饰</u>

[0949] 能够被掺入修饰的核酸中的本文描述的修饰的核苷和修饰的核苷酸可以包括修饰的核碱基。核碱基的实例包括但不限于腺嘌呤(A)、鸟嘌呤(G)、胞嘧啶(C)以及尿嘧啶(U)。这些核碱基可以被修饰或全部替换,以提供可以被掺入修饰的核酸中的修饰的核苷和修饰的核苷酸。核苷酸的核碱基可以独立地选自嘌呤、嘧啶、嘌呤或嘧啶类似物。在一个实施例中,核碱基可以包括,例如,天然存在的碱基及其合成的衍生物。

[0950] 尿嘧啶

[0951] 在一个实施例中,修饰的核碱基是修饰的尿嘧啶。具有修饰的尿嘧啶的示例性核碱基和核苷包括但不限于假尿苷(ψ)、吡啶-4-酮核糖核苷、5-氮杂-尿苷、6-氮杂-尿苷、2-硫代-5-氮杂-尿苷、2-硫代-尿苷(s2U)、4-硫代-尿苷(s4U)、4-硫代-假尿苷、2-硫代-假尿苷、5-羟基-尿苷(ho^5U)、5-氨基烯丙基-尿苷、5-卤代-尿苷(例如,5-碘代-尿苷或5-溴代-尿苷)、3-甲基-尿苷(m^3U)、5-甲氧基-尿苷(mo^5U)、尿苷5-氧基乙酸(cmo^5U)、尿苷5-氧基乙酸甲酯($mcmo^5U$)、5-羧甲基-尿苷(cm^5U)、1-羧甲基-假尿苷、5-羧基羟甲基-尿苷(chm^5U)、5-羧基羟甲基-尿苷甲酯($mchm^5U$)、5-甲氧羰基甲基-尿苷(mcm^5U)、5-甲氧羰基甲基-2-硫代-尿苷(mcm^5s2U)、5-氨甲基-2-硫代-尿苷(nm^5s2U)、5-甲基氨甲基-尿苷(mnm^5U)、5-甲基氨甲基-2-硫代-尿苷(mnm^5s2U)、5-甲基氨甲基-2-硒代-尿苷(mnm^5se^2U)、5-氨甲酰基甲基-尿苷

(ncm^{5}U)、5-羧甲基氨基甲基-尿苷(cmnm^{5}U)、5-羧甲基氨基甲基-2-硫代-尿苷(cmnm ^{5}s2U)、5-丙炔基-尿苷、1-丙炔基-假尿苷、5-牛磺酸甲基-尿苷(τcm^{5}U)、1-牛磺酸甲基-假尿苷、5-牛磺酸甲基-2-硫代-尿苷(τm^{5}s2U)、1-牛磺酸甲基-4-硫代-假尿苷、5-甲基-尿苷(m^{5}U,即具有核碱基脱氧胸腺嘧啶)、1-甲基-假尿苷(m^{1}ψ)、5-甲基-2-硫代-尿苷(m^{5}s2U)、1-甲基-4-硫代-假尿苷(m^{1}s^{4}ψ)、4-硫代-1-甲基-假尿苷、3-甲基-假尿苷(m^{3}ψ)、2-硫代-1-甲基-假尿苷、1-甲基-1-去氮杂-假尿苷、2-硫代-1-甲基-1-去氮杂-假尿苷、二氢尿苷(D)、二氢假尿苷、5,6-二氢尿苷、5-甲基-二氢尿苷(m^{5}D)、2-硫代-二氢尿苷、2-硫代-二氢假尿苷、2-甲氧基-尿苷、2-甲氧基-4-硫代-尿苷、4-甲氧基-假尿苷、4-甲氧基-2-硫代-假尿苷、N1-甲基-假尿苷、3-(3-氨基-3-羧丙基)尿苷(acp^{3}U)、1-甲基-3-(3-氨基-3-羧丙基)假尿苷(acp^{3}ψ)、5-(异戊烯基氨甲基)尿苷(inm^{5}U)、5-(异戊烯基氨甲基)-2-硫代-尿苷(inm^{5}s2U)、α-硫代-尿苷、2'-O-甲基-尿苷(Um)、5,2′-O-二甲基-尿苷(m^{5}Um)、2'-O-甲基-假尿苷(ψm)、2-硫代-2'-O-甲基-尿苷(s2Um)、5-甲氧基羰甲基-2'-O-甲基-尿苷(mcm ^{5}Um)、5-氨甲酰基甲基-2'-O-甲基-尿苷(ncm ^{5}Um)、5-羧甲基氨甲基-2'-O-甲基-尿苷(cmnm ^{5}Um)、3,2'-O-二甲基-尿苷(m^{3}Um)、5-(异戊烯基氨甲基)-2'-O-甲基-尿苷(inm ^{5}Um)、1-硫代-尿苷、脱氧胸苷、2'-F-阿糖(ara)-尿苷、2′-F-尿苷、2'-OH-阿糖-尿苷、5-(2-甲氧甲酰基乙烯基)尿苷、5-[3-(1-E-丙烯基氨基)尿苷、吡唑并[3,4-d]嘧啶、黄嘌呤以及次黄嘌呤。

[0952] 胞嘧啶

[0953] 在一个实施例中,修饰的核碱基是修饰的胞嘧啶。具有修饰的胞嘧啶的示例性核碱基和核苷包括但不限于5-氮杂-胞苷、6-氮杂-胞苷、假异胞苷、3-甲基-胞苷(m^{3}C)、N4-乙酰基-胞苷(act)、5-甲酰基-胞苷(f^{5}C)、N4-甲基-胞苷(m^{4}C)、5-甲基-胞苷(m^{5}C)、5-卤代-胞苷(例如,5-碘代-胞苷)、5-羟甲基-胞苷(hm^{5}C)、1-甲基-假异胞苷、吡咯并-胞苷、吡咯并-假异胞苷、2-硫代-胞苷(s2C)、2-硫代-5-甲基-胞苷、4-硫代-假异胞苷、4-硫代-1-甲基-假异胞苷、4-硫代-1-甲基-1-去氮杂-假异胞苷、1-甲基-1-去氮杂-假异胞苷、泽布拉林(zebularine)、5-氮杂-泽布拉林、5-甲基-泽布拉林、5-氮杂-2-硫代-泽布拉林、2-硫代-泽布拉林、2-甲氧基-胞苷、2-甲氧基-5-甲基-胞苷、4-甲氧基-假异胞苷、4-甲氧基-1-甲基-假异胞苷、赖西丁(k^{2}C)、α-硫代-胞苷、2′-O-甲基-胞苷(Cm)、5,2′-O-二甲基-胞苷(m^{5}Cm)、N4-乙酰基-2′-O-甲基-胞苷(ac^{4}Cm)、N4,2′-O-二甲基-胞苷(m^{4}Cm)、5-甲酰基-2′-O-甲基-胞苷(f ^{5}Cm)、N4,N4,2′-O-三甲基-胞苷(m$^4{}_2$Cm)、1-硫代-胞苷、2′-F-阿糖-胞苷、2′-F-胞苷以及2′-OH-阿糖-胞苷。

[0954] 腺嘌呤

[0955] 在一个实施例中,修饰的核碱基是修饰的腺嘌呤。具有修饰的腺嘌呤的示例性核碱基和核苷包括但不限于2-氨基-嘌呤、2,6-二氨基嘌呤、2-氨基-6-卤代-嘌呤(例如,2-氨基-6-氯代-嘌呤)、6-卤代-嘌呤(例如,6-氯代-嘌呤)、2-氨基-6-甲基-嘌呤、8-叠氮基-腺苷、7-去氮杂-腺苷、7-去氮杂-8-氮杂-腺苷、7-去氮杂-2-氨基-嘌呤、7-去氮杂-8-氮杂-2-氨基-嘌呤、7-去氮杂-2,6-二氨基嘌呤、7-去氮杂-8-氮杂-2,6-二氨基嘌呤、1-甲基-腺苷(m^{1}A)、2-甲基-腺苷(m^{2}A)、N6-甲基-腺苷(m^{6}A)、2-甲硫基-N6-甲基-腺苷(ms2m^{6}A)、N6-异戊烯基-腺苷(i^{6}A)、2-甲硫基-N6-异戊烯基-腺苷(ms^{2}i^{6}A)、N6-(顺羟基异戊烯基)腺苷(io^{6}A)、2-甲硫基-N6-(顺羟基异戊烯基)腺苷(ms2io^{6}A)、N6-缩水甘油基氨甲酰基-腺苷(g^{6}A)、N6-苏氨酰基氨甲酰基-腺苷(t^{6}A)、N6-甲基-N6-苏氨酰基氨甲酰基-腺苷(m^{6}t^{6}A)、2-

甲硫基-N6-苏氨酰基氨甲酰基-腺苷(ms^2g^6A)、N6,N6-二甲基-腺苷($m^6{}_2A$)、N6-羟基正缬氨酰基氨甲酰基-腺苷(hn^6A)、2-甲硫基-N6-羟基正缬氨酰基氨甲酰基-腺苷($ms2hn^6A$)、N6-乙酰基-腺苷(ac^6A)、7-甲基-腺苷、2-甲硫基-腺苷、2-甲氧基-腺苷、α-硫代-腺苷、2'-O-甲基-腺苷(Am)、N^6,2'-O-二甲基-腺苷(m^6Am)、N^6-甲基-2'-脱氧腺苷、N6,N6,2'-O-三甲基-腺苷($m^6{}_2Am$)、1,2'-O-二甲基-腺苷(m^1Am)、2'-O-核糖基腺苷(磷酸盐)(Ar(p))、2-氨基-N6-甲基-嘌呤、1-硫代-腺苷、8-叠氮基-腺苷、2'-F-阿糖-腺苷、2'-F-腺苷、2′-Oh-阿糖-腺苷以及N6-(19-氨基-五氧杂十九烷基)-腺苷。

[0956] 鸟嘌呤

[0957] 在一个实施例中,修饰的核碱基是修饰的鸟嘌呤。具有修饰的鸟嘌呤的示例性核碱基和核苷包括但不限于肌苷(I)、1-甲基-肌苷(m^1I)、怀俄苷(imG)、甲基怀俄苷(mimG)、4-去甲基-怀俄苷(imG-14)、异怀俄苷(imG2)、怀丁苷(yW)、过氧怀丁苷(o_2yW)、羟基怀丁苷(OHyW)、修饰不足的羟基怀丁苷(OHyW*)、7-去氮杂-鸟苷、辫苷(Q)、环氧辫苷(oQ)、半乳糖基-辫苷(galQ)、甘露糖基-辫苷(manQ)、7-氰基-7-去氮杂-鸟苷($preQ_0$)、7-氨基甲基-7-去氮杂-鸟苷($preQ_1$)、古嘌苷(G^+)、7-去氮杂-8-氮杂-鸟苷、6-硫代-鸟苷、6-硫代-7-去氮杂-鸟苷、6-硫代-7-去氮杂-8-氮杂-鸟苷、7-甲基-鸟苷(m^7G)、6-硫代-7-甲基-鸟苷、7-甲基-肌苷、6-甲氧基-鸟苷、1-甲基-鸟苷(m'G)、N2-甲基-鸟苷(m^2G)、N2,N2-二甲基-鸟苷($m^2{}_2G$)、N2,7-二甲基-鸟苷(m^2,7G)、N2,N2,7-二甲基-鸟苷(m^2,2,7G)、8-氧代-鸟苷、7-甲基-8-氧代-鸟苷、1-甲基-6-硫代-鸟苷、N2-甲基-6-硫代-鸟苷、N2,N2-二甲基-6-硫代-鸟苷、α-硫代-鸟苷、2'-O-甲基-鸟苷(Gm)、N2-甲基-2'-O-甲基-鸟苷(m^2Gm)、N2,N2-二甲基-2'-O-甲基-鸟苷($m^2{}_2Gm$)、1-甲基-2'-O-甲基-鸟苷(m'Gm)、N2,7-二甲基-2'-O-甲基-鸟苷(m^2,7Gm)、2′-O-甲基-肌苷(Im)、1,2'-O-二甲基-肌苷(m'Im)、O^6-苯基-2'-脱氧肌苷、2′-O-核糖基鸟苷(磷酸盐)(Gr(p))、1-硫代-鸟苷、O^6-甲基-鸟苷、O^6-甲基-2'-脱氧鸟苷、2'-F-阿糖-鸟苷以及2'-F-鸟苷。

[0958] 示例性修饰的gRNA

[0959] 在一些实施例中,修饰的核酸可以是修饰的gRNA。应理解,本文所述的任何gRNA可根据本节进行修饰,包括包含来自SEQ ID NO:251-SEQ ID NO:901的靶向结构域的任何gRNA。

[0960] 如以上所讨论的,瞬时表达的或递送的核酸可以易于被例如,细胞核酸酶降解。因此,在一个方面中,在此所描述的修饰的gRNA可以含有一个或多个修饰的核苷或核苷酸,其引入对核酸酶的稳定性。虽然不希望受理论束缚,但还据信当引入细胞群,特别是本发明的细胞时,本文所述的某些修饰的gRNA可展现出减少的先天性免疫应答。如上所述的,术语"先天性免疫应答"包括对外源核酸的细胞应答,外源核酸包括通常是病毒或细菌来源的单链核酸,所述细胞应答涉及细胞因子(特别是干扰素)表达与释放以及细胞死亡的诱导。

[0961] 虽然本部分中讨论的一些示例性修饰可以包括在gRNA序列内的任何位置,但是在一些实施例中,gRNA在其5'端或其附近包含修饰(例如,在其5'端的1-10、1-5或1-2个核苷酸内)。在一些实施例中,gRNA在其3'端处或附近(例如,在其3'端的1-10、1-5或1-2个核苷酸内)包含修饰。在一些实施例中,gRNA既在其5'端处或附近包含修饰又在其3'端处或附近包含修饰。

[0962] 在一个实施例中,gRNA的5′端通过包含真核mRNA帽结构或帽类似物(例如,G(5')

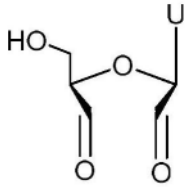
ppp(5')G帽类似物、m7G(5')ppp(5')G帽类似物、或3'-O-Me-m7G(5')ppp(5')G抗反向帽类似物(ARCA))进行修饰。所述帽或帽类似物可以在gRNA的化学合成或体外转录期间被包括。

[0963] 在实施例中,体外转录的gRNA通过用磷酸酶(例如,牛小肠碱性磷酸酶)处理进行修饰,以去除5'三磷酸酯基团。

[0964] 在一个实施例中,gRNA的3'端通过添加一个或多个(例如,25-200个)腺嘌呤(A)残基进行修饰。聚A束可以被包含在编码gRNA的核酸(例如,质粒、PCR产物、病毒基因组)中,或者可以在化学合成期间、或在使用聚腺苷聚合酶(例如,大肠杆菌聚(A)聚合酶)体外转录后被添加到gRNA上。

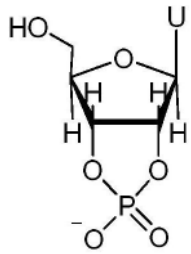
[0965] 在实施例中,体外转录的gRNA既含有5'帽结构或帽类似物又含有3'聚A束。在实施例中,体外转录的gRNA通过用磷酸酶(例如,牛小肠碱性磷酸酶)处理进行修饰以去除5'三磷酸酯基团,并且包含3'聚A束。

[0966] 在一些实施例中,gRNA可以在3'末端U核糖处被修饰。例如,U核糖的两个末端羟基基团可以被氧化为醛基基团和核糖环的伴随开口,以提供如下所示的修饰的核苷:

[0967] 

[0968] 其中"U"可以是未修饰的或修饰的尿苷。

[0969] 在另一个实施例中,可以用如下所示的2'3'环状磷酸酯修饰3'末端U:

[0970] 

[0971] 其中"U"可以是未修饰的或修饰的尿苷。

[0972] 在一些实施例中,gRNA分子可以含有3'核苷酸,其可以例如,通过掺入本文描述的一个或多个修饰的核苷酸而相对于降解进行稳定化。在这个实施例中,例如,尿苷可以被修饰的尿苷(例如,5-(2-氨基)丙基尿苷和5-溴代尿苷)或被本文描述的任何修饰的尿苷置换;腺苷和鸟苷可以被修饰的腺苷和鸟苷(例如,在8-位具有修饰,例如8-溴代鸟苷)或被本文描述的任何修饰的腺苷和鸟苷置换。

[0973] 在一些实施例中,可以向gRNA中掺入糖-修饰的核糖核苷酸,例如,其中2'OH-基团被选自以下项的基团替换:H、-OR、-R(其中R可以是,例如,烷基、环烷基、芳基、芳烷基、杂芳基或糖)、卤素、-SH、-SR(其中R可以是,例如,烷基、环烷基、芳基、芳烷基、杂芳基或糖)、氨基(其中氨基可以是,例如,NH_2;烷氨基、二烷氨基、杂环基、芳氨基、二芳氨基、杂芳氨基、二杂芳氨基或氨基酸);或氰基(-CN)。在一些实施例中,可以例如,用硫代磷酸酯基团如本文描述的修饰磷酸骨架。在一些实施例中,gRNA的一个或多个核苷酸可以各自独立地是修饰的或未修饰的核苷酸,包括但不限于2'-糖修饰的如2'-O-甲基、2'-O-甲氧基乙基,或2'-氟修饰的,包括例如,2'-F或2'-O-甲基腺苷(A)、2'-F或2'-O-甲基胞苷(C)、2'-F或2'-O-甲基

尿苷(U)、2'-F或2'-O-甲基胸苷(T)、2'-F或2'-O-甲基鸟苷(G)、2'-O-甲氧基乙基-5-甲基尿苷(Teo)、2'-O-甲氧基乙基腺苷(Aeo)、2'-O-甲氧基乙基-5-甲基胞苷(m5Ceo)、及其任何组合。

[0974] 在一些实施例中,gRNA可以包括"锁"核酸(LNA),其中2'OH-基团可以例如通过C1-6亚烷基或C1-6杂亚烷基桥连接至同一核糖的4'碳,其中示例性桥可以包括亚甲基、亚丙基、醚或氨基桥;O-氨基(其中氨基可以是例如NH_2;烷氨基、二烷氨基、杂环基、芳氨基、二芳氨基、杂芳氨基或二杂芳氨基、乙二胺或聚氨基)和氨基烷氧基或$O(CH_2)_n$-氨基(其中氨基可以是例如NH_2;烷氨基、二烷氨基、杂环基、芳氨基、二芳氨基、杂芳氨基或二杂芳氨基、乙二胺或聚氨基)。

[0975] 在一些实施例中,gRNA可以包括修饰的核苷酸,其是多环的(例如,三环;和"解锁"形式,如二醇核酸(GNA)(例如,R-GNA或S-GNA,其中核糖被附接至磷酸二酯键的二醇单元替换),或苏糖核酸(TNA,其中核糖被α-L-苏呋喃糖基-(3'→2')替换)。

[0976] 通常,gRNA分子包括糖基核糖,它是具有氧的5元环。示例性修饰的gRNA可以包括但不限于核糖中氧的置换(例如,用硫(S)、硒(Se)或亚烷基,例如像亚甲基或亚乙基);双键的添加(例如,以用环戊烯基或环己烯基置换核糖);核糖的缩环(例如,以形成环丁烷或氧杂环丁烷的4元环);核糖的扩环(例如,以形成具有另外的碳或杂原子的6元环或7元环,例如像脱水己糖醇、阿卓糖醇、甘露醇、环己烷基、环己烯基以及吗啉代,其也具有氨基磷酸酯骨架)。尽管大多数的糖类似物改变被定位至2'位,其他位点也适于修饰,包括4'位。在一个实施例中,gRNA包含4'-S、4'-Se或4′-C-氨基甲基-2'-O-Me修饰。

[0977] 在一些实施例中,可以将去氮杂核苷酸(例如,7-去氮杂-腺苷)掺入gRNA中。在一些实施例中,可以将O-和N-烷基化的核苷酸(例如,N6-甲基腺苷)掺入gRNA中。在一些实施例中,gRNA分子中的一个或多个或所有核苷酸是脱氧核苷酸。

[0978] miRNA结合位点

[0979] 微小RNA(或miRNA)是天然存在的19-25个核苷酸长的细胞非编码RNA。它们结合至例如在mRNA的3'UTR中具有适当miRNA结合位点的核酸分子,并且下调基因表达。虽然不希望受理论束缚,但认为该下调是通过降低核酸分子稳定性或通过抑制翻译而发生。本文所披露的RNA种类(例如,编码Cas9的mRNA)可以例如在其3'UTR中包含miRNA结合位点。miRNA结合位点可以被选择为促进所选细胞类型中的表达下调。

[0980] 本公开还提供了以下的实施方式。

[0981] 实施方式1.一种gRNA分子,所述gRNA分子包含靶向结构域,所述靶向结构域包含与完全或部分位于HBG1或HBG2调控区域内的靶结构域互补或部分互补的核苷酸序列。

[0982] 实施方式2.如实施方式1所述的gRNA分子,其中所述HBG1或HBG2调控区域分别与HBG1或HBG2基因相邻。

[0983] 实施方式3.如实施方式2所述的gRNA分子,其中所述HBG1调控区域位于跨越SEQ ID NO:902的核苷酸1-2990的区域。

[0984] 实施方式4.如实施方式2所述的gRNA分子,其中所述HBG2调控区域位于跨越SEQ ID NO:903的核苷酸1-2914的区域。

[0985] 实施方式5.如实施方式1-4中任一项所述的gRNA分子,其中所述靶向结构域被配置成在HBG靶位置的500、400、300、200、100、50、25、或10个核苷酸内提供选自双链断裂和单

链断裂的切割事件。

[0986] 实施方式6.如实施方式1所述的gRNA分子,其中所述靶结构域完全位于HBG1或HBG2调控区域内。

[0987] 实施方式7.如实施方式1-6中任一项所述的gRNA分子,其中所述靶向结构域被配置成靶向HBG1或HBG2调控区域中的转录调控元件。

[0988] 实施方式8.如实施方式7所述的gRNA分子,其中所述转录调控元件是启动子。

[0989] 实施方式9.如实施方式8所述的gRNA分子,其中所述启动子控制HBG1和HBG2中的一种或多种的转录。

[0990] 实施方式10.如实施方式7所述的gRNA分子,其中所述转录调控元件是沉默子。

[0991] 实施方式11.如实施方式1-10中任一项所述的gRNA分子,其中所述靶向结构域包含与SEQ ID NO:251-901中任一个所示的核苷酸序列相同或相差不超过1、2、3、4、或5个核苷酸的核苷酸序列。

[0992] 实施方式12.如实施方式11所述的gRNA分子,其中所述靶向结构域包含与SEQ ID NO:251-901中任一个所示的核苷酸序列相同的核苷酸序列。

[0993] 实施方式13.如实施方式1-12中任一项所述的gRNA分子,其中所述gRNA分子是模块化gRNA分子。

[0994] 实施方式14.如实施方式1-12中任一项所述的gRNA分子,其中所述gRNA分子是单分子gRNA分子。

[0995] 实施方式15.如实施方式1-12中任一项所述的gRNA分子,其中所述gRNA分子是嵌合gRNA分子。

[0996] 实施方式16.如实施方式1-12中任一项所述的gRNA分子,其中所述靶向结构域在长度上是16个或更多个核苷酸。

[0997] 实施方式17.如实施方式1-12中任一项所述的gRNA分子,其中所述靶向结构域在长度上是17个或更多个核苷酸。

[0998] 实施方式18.如实施方式1-12中任一项所述的gRNA分子,其中所述靶向结构域在长度上是18个或更多个核苷酸。

[0999] 实施方式19.如实施方式1-12中任一项所述的gRNA分子,其中所述靶向结构域在长度上是19个或更多个核苷酸。

[1000] 实施方式20.如实施方式1-12中任一项所述的gRNA分子,其中所述靶向结构域在长度上是20个或更多个核苷酸。

[1001] 实施方式21.如实施方式1-12中任一项所述的gRNA分子,其中所述靶向结构域在长度上是21个或更多个核苷酸。

[1002] 实施方式22.如实施方式1-12中任一项所述的gRNA分子,其中所述靶向结构域在长度上是22个或更多个核苷酸。

[1003] 实施方式23.如实施方式1-12中任一项所述的gRNA分子,其中所述靶向结构域在长度上是23个或更多个核苷酸。

[1004] 实施方式24.如实施方式1-12中任一项所述的gRNA分子,其中所述靶向结构域在长度上是24个或更多个核苷酸。

[1005] 实施方式25.如实施方式1-12中任一项所述的gRNA分子,其中所述靶向结构域在

长度上是25个或更多个核苷酸。

[1006] 实施方式26.如实施方式1-12中任一项所述的gRNA分子,其中所述靶向结构域在长度上是26个或更多个核苷酸。

[1007] 实施方式27.如实施方式1-12中任一项所述的gRNA分子,所述gRNA分子进一步包含第一互补结构域、连接结构域、第二互补结构域、近端结构域、5′延伸结构域、和尾部结构域中的一个或多个。

[1008] 实施方式28.如实施方式27所述的gRNA分子,所述gRNA分子从5'到3'包含:靶向结构域;第一互补结构域;连接结构域;第二互补结构域;和近端结构域。

[1009] 实施方式29.如实施方式28所述的gRNA分子,所述gRNA分子进一步包含尾部结构域。

[1010] 实施方式30.如实施方式1-29中任一项所述的gRNA分子,所述gRNA分子包含:包含不多于25个核苷酸的连接结构域;包含至少20个核苷酸的一起考虑的近端结构域和尾部结构域;和由17或18个核苷酸组成的靶向结构域。

[1011] 实施方式31.如实施方式1-29中任一项所述的gRNA分子,所述gRNA分子包含:包含不多于25个核苷酸的连接结构域;包含至少25个核苷酸的一起考虑的近端结构域和尾部结构域;和由17或18个核苷酸组成的靶向结构域。

[1012] 实施方式32.如实施方式1-29中任一项所述的gRNA分子,所述gRNA分子包含:包含不多于25个核苷酸的连接结构域;包含至少30个核苷酸的一起考虑的近端结构域和尾部结构域;和由17个核苷酸组成的靶向结构域。

[1013] 实施方式33.如实施方式1-37中任一项所述的gRNA分子,所述gRNA分子包含:包含不多于25个核苷酸的连接结构域;包含至少40个核苷酸的一起考虑的近端结构域和尾部结构域;和由17个核苷酸组成的靶向结构域。

[1014] 实施方式34.一种核酸组合物,所述核酸组合物包含:(a)编码包含靶向结构域的gRNA分子的核苷酸序列,所述靶向结构域包含与完全或部分位于HBG1或HBG2调控区域内的靶结构域互补或部分互补的核苷酸序列。

[1015] 实施方式35.如实施方式34所述的核酸组合物,其中所述gRNA分子是如实施方式1-33中任一项所述的gRNA分子。

[1016] 实施方式36.如实施方式34或35所述的核酸组合物,其中所述靶向结构域被配置成在HBG靶位置的500、400、300、200、100、50、25、或10个核苷酸内提供选自双链断裂和单链断裂的切割事件。

[1017] 实施方式37.如实施方式34-36中任一项所述的核酸组合物,其中所述靶向结构域包含与SEQ ID NO:251-901中任一个所示的核苷酸序列相同或相差不超过1、2、3、4、或5个核苷酸的核苷酸序列。

[1018] 实施方式38.如实施方式37所述的核酸组合物,其中所述靶向结构域包含与SEQ ID NO:251-901中任一个所示的核苷酸序列相同的核苷酸序列。

[1019] 实施方式39.如实施方式34-38中任一项所述的核酸组合物,其中所述gRNA分子是模块化gRNA分子。

[1020] 实施方式40.如实施方式34-38中任一项所述的核酸组合物,其中所述gRNA分子是单分子gRNA分子。

[1021] 实施方式41.如实施方式34-38中任一项所述的核酸组合物,其中所述gRNA分子是嵌合gRNA分子。

[1022] 实施方式42.如实施方式34-41中任一项所述的核酸组合物,其中所述靶向结构域在长度上是16个或更多个核苷酸。

[1023] 实施方式43.如实施方式34-41中任一项所述的核酸组合物,其中所述靶向结构域在长度上是17个或更多个核苷酸。

[1024] 实施方式44.如实施方式34-41中任一项所述的核酸组合物,其中所述靶向结构域在长度上是18个或更多个核苷酸。

[1025] 实施方式45.如实施方式34-41中任一项所述的核酸组合物,其中所述靶向结构域在长度上是19个或更多个核苷酸。

[1026] 实施方式46.如实施方式34-41中任一项所述的核酸组合物,其中所述靶向结构域在长度上是20个或更多个核苷酸。

[1027] 实施方式47.如实施方式34-41中任一项所述的核酸组合物,其中所述靶向结构域在长度上是21个或更多个核苷酸。

[1028] 实施方式48.如实施方式34-41中任一项所述的核酸组合物,其中所述靶向结构域在长度上是22个或更多个核苷酸。

[1029] 实施方式49.如实施方式34-41中任一项所述的核酸组合物,其中所述靶向结构域在长度上是23个或更多个核苷酸。

[1030] 实施方式50.如实施方式34-41中任一项所述的核酸组合物,其中所述靶向结构域在长度上是24个或更多个核苷酸。

[1031] 实施方式51.如实施方式34-41中任一项所述的核酸组合物,其中所述靶向结构域在长度上是25个或更多个核苷酸。

[1032] 实施方式52.如实施方式34-41中任一项所述的核酸组合物,其中所述靶向结构域在长度上是26个或更多个核苷酸。

[1033] 实施方式53.如实施方式34-52中任一项所述的核酸组合物,其中所述gRNA分子进一步包含靶向结构域;第一互补结构域;连接结构域;第二互补结构域;和近端结构域中的一个或多个。

[1034] 实施方式54.如实施方式53所述的核酸组合物,其中所述gRNA分子从5'到3′包含:靶向结构域;第一互补结构域;连接结构域;第二互补结构域;和近端结构域。

[1035] 实施方式55.如实施方式54所述的核酸组合物,其中所述gRNA进一步包含尾部结构域。

[1036] 实施方式56.如实施方式34-55中任一项所述的核酸组合物,其中所述gRNA分子包含:包含不多于25个核苷酸的连接结构域;包含至少20个核苷酸的一起考虑的近端结构域和尾部结构域;和具有17或18个核苷酸的靶向结构域。

[1037] 实施方式57.如实施方式34-55中任一项所述的核酸组合物,其中所述gRNA分子包含:包含不多于25个核苷酸的连接结构域;包含至少25个核苷酸的一起考虑的近端结构域和尾部结构域;和在长度上有17或18个核苷酸的靶向结构域。

[1038] 实施方式58.如实施方式34-55中任一项所述的核酸组合物,其中所述gRNA分子包含:包含不多于25个核苷酸的连接结构域;包含至少30个核苷酸的一起考虑的近端结构域

和尾部结构域;和在长度上有17个核苷酸的靶向结构域。

[1039] 实施方式59.如实施方式34-55中任一项所述的核酸组合物,其中所述gRNA分子包含:包含不多于25个核苷酸的连接结构域;包含至少40个核苷酸的一起考虑的近端结构域和尾部结构域;和在长度上有17个核苷酸的靶向结构域。

[1040] 实施方式60.如实施方式34-59中任一项所述的核酸组合物,所述核酸组合物进一步包含:(b)编码RNA指导的核酸酶的核苷酸序列。

[1041] 实施方式61.如实施方式60所述的核酸组合物,其中所述RNA指导的核酸酶是Cas9分子或Cas9-融合蛋白。

[1042] 实施方式62.如实施方式61所述的核酸组合物,其中所述Cas9分子是具有酶活性的Cas9(eaCas9)分子。

[1043] 实施方式63.如实施方式62所述的核酸组合物,其中所述eaCas9分子包含切口酶分子。

[1044] 实施方式64.如实施方式62或63所述的核酸组合物,其中所述eaCas9分子在靶核酸中形成双链断裂。

[1045] 实施方式65.如实施方式62或63所述的核酸组合物,其中所述eaCas9分子在靶核酸中形成单链断裂。

[1046] 实施方式66.如实施方式65所述的核酸组合物,其中所述单链断裂形成于所述靶核酸的与所述gRNA分子的靶向结构域互补的链中。

[1047] 实施方式67.如实施方式66所述的核酸组合物,其中所述单链断裂形成于所述靶核酸的不同于与所述gRNA分子的靶向结构域互补的链的链中。

[1048] 实施方式68.如实施方式61-67中任一项所述的核酸组合物,其中所述eaCas9分子包含HNH样结构域切割活性,但没有或具有不显著的N-末端RuvC样结构域切割活性。

[1049] 实施方式69.如实施方式68所述的核酸组合物,其中所述eaCas9分子是HNH样结构域切口酶。

[1050] 实施方式70.如实施方式68或69所述的核酸组合物,其中所述eaCas9分子在D10处包含突变。

[1051] 实施方式71.如实施方式65-70中任一项所述的核酸组合物,其中所述eaCas9分子包含N端RuvC样结构域切割活性,但没有或具有不显著的HNH样结构域切割活性。

[1052] 实施方式72.如实施方式70所述的核酸组合物,其中所述eaCas9分子是N端RuvC样结构域切口酶。

[1053] 实施方式73.如实施方式71或72所述的核酸组合物,其中所述eaCas9分子在H840或N863处包含突变。

[1054] 实施方式74.如实施方式34-73中任一项所述的核酸组合物,所述核酸组合物进一步包含:(c)编码包含靶向结构域的第二gRNA分子的核苷酸序列,所述靶向结构域包含与完全或部分位于HBG1或HBG2调控区域内的靶结构域互补或部分互补的核苷酸序列。

[1055] 实施方式75.如实施方式74所述的核酸组合物,其中所述第二gRNA分子是如实施方式1-33中任一项所述的gRNA分子。

[1056] 实施方式76.如实施方式74或75所述的核酸组合物,其中所述第二gRNA分子的所述靶向结构域被配置成在HBG靶位置的500、400、300、200、100、50、25、或10个核苷酸内提供

选自双链断裂和单链断裂的切割事件。

[1057] 实施方式77.如实施方式74-76中任一项所述的核酸组合物,其中所述第二gRNA分子的所述靶向结构域包含与SEQ ID NO:251-901中任一个所示的核苷酸序列相同或相差不超过1、2、3、4、或5个核苷酸的核苷酸序列。

[1058] 实施方式78.如实施方式77所述的核酸组合物,其中所述第二gRNA分子的所述靶向结构域包含与SEQ ID NO:251-901中任一个所示的核苷酸序列相同的核苷酸序列。

[1059] 实施方式79.如实施方式74-78中任一项所述的核酸组合物,其中所述第二gRNA分子是单分子gRNA分子。

[1060] 实施方式80.如实施方式74-78中任一项所述的核酸组合物,其中所述第二gRNA分子是模块化gRNA分子。

[1061] 实施方式81.如实施方式74-78中任一项所述的核酸组合物,其中所述第二gRNA分子是嵌合gRNA分子。

[1062] 实施方式82.如实施方式74-81中任一项所述的核酸组合物,其中所述第二gRNA分子的所述靶向结构域在长度上是16个或更多个核苷酸。

[1063] 实施方式83.如实施方式74-81中任一项所述的核酸组合物,其中所述第二gRNA分子的所述靶向结构域在长度上是17个或更多个核苷酸。

[1064] 实施方式84.如实施方式74-81中任一项所述的核酸组合物,其中所述第二gRNA分子的所述靶向结构域在长度上是18个或更多个核苷酸。

[1065] 实施方式85.如实施方式74-81中任一项所述的核酸组合物,其中所述第二gRNA分子的所述靶向结构域在长度上是19个或更多个核苷酸。

[1066] 实施方式86.如实施方式74-81中任一项所述的核酸组合物,其中所述第二gRNA分子的所述靶向结构域在长度上是20个或更多个核苷酸。

[1067] 实施方式87.如实施方式74-81中任一项所述的核酸组合物,其中所述第二gRNA分子的所述靶向结构域在长度上是21个或更多个核苷酸。

[1068] 实施方式88.如实施方式74-81中任一项所述的核酸组合物,其中所述第二gRNA分子的所述靶向结构域在长度上是22个或更多个核苷酸。

[1069] 实施方式89.如实施方式74-81中任一项所述的核酸组合物,其中所述第二gRNA分子的所述靶向结构域在长度上是23个或更多个核苷酸。

[1070] 实施方式90.如实施方式74-81中任一项所述的核酸组合物,其中所述第二gRNA分子的所述靶向结构域在长度上是24个或更多个核苷酸。

[1071] 实施方式91.如实施方式74-81中任一项所述的核酸组合物,其中所述第二gRNA分子的所述靶向结构域在长度上是25个或更多个核苷酸。

[1072] 实施方式92.如实施方式74-81中任一项所述的核酸组合物,其中所述第二gRNA分子的所述靶向结构域在长度上是26个或更多个核苷酸。

[1073] 实施方式93.如实施方式74-92中任一项所述的核酸组合物,其中所述第二gRNA分子进一步包含靶向结构域;第一互补结构域;连接结构域;第二互补结构域;和近端结构域中的一个或多个。

[1074] 实施方式94.如实施方式93所述的核酸组合物,其中所述第二gRNA分子从5′到3’包含:靶向结构域;第一互补结构域;连接结构域;第二互补结构域;和近端结构域。

[1075] 实施方式95.如实施方式94所述的核酸组合物,其中所述第二gRNA进一步包含尾部结构域。

[1076] 实施方式96.如实施方式74-95中任一项所述的核酸组合物,其中所述第二gRNA分子包含:包含不多于25个核苷酸的连接结构域;包含至少20个核苷酸的一起考虑的近端结构域和尾部结构域;和具有17或18个核苷酸的靶向结构域。

[1077] 实施方式97.如实施方式74-95中任一项所述的核酸组合物,其中所述第二gRNA分子包含:包含不多于25个核苷酸的连接结构域;包含至少25个核苷酸的一起考虑的近端结构域和尾部结构域;和在长度上有17或18个核苷酸的靶向结构域。

[1078] 实施方式98.如实施方式74-95中任一项所述的核酸组合物,其中所述第二gRNA分子包含:包含不多于25个核苷酸的连接结构域;包含至少30个核苷酸的一起考虑的近端结构域和尾部结构域;和在长度上有17个核苷酸的靶向结构域。

[1079] 实施方式99.如实施方式74-95中任一项所述的核酸组合物,其中所述第二gRNA分子包含:包含不多于25个核苷酸的连接结构域;包含至少40个核苷酸的一起考虑的近端结构域和尾部结构域;和在长度上有17个核苷酸的靶向结构域。

[1080] 实施方式100.如实施方式74-99中任一项所述的核酸组合物,所述核酸组合物进一步包含:(d)编码包含靶向结构域的第三gRNA分子的核苷酸序列,所述靶向结构域包含与完全或部分位于HBG1或HBG2调控区域内的靶结构域互补或部分互补的核苷酸序列。

[1081] 实施方式101.如实施方式100所述的核酸组合物,所述核酸组合物进一步包含:(f)编码包含靶向结构域的第四gRNA分子的核苷酸序列,所述靶向结构域包含与完全或部分位于HBG1或HBG2调控区域内的靶结构域互补或部分互补的核苷酸序列。

[1082] 实施方式102.如实施方式34-101中任一项所述的核酸组合物,所述核酸组合物进一步包含(g)模板核酸。

[1083] 实施方式103.如实施方式102所述的核酸组合物,其中所述模板核酸是单链寡脱氧核苷酸(ssODN)。

[1084] 实施方式104.如实施方式103所述的核酸组合物,其中所述模板核酸包含5'同源臂、替换序列、和3'同源臂。

[1085] 实施方式105.如实施方式104所述的核酸组合物,其中所述5'同源臂在长度上包含约200个核苷酸,例如,长度为至少25、50、75、100、125、150、175、或200个核苷酸;所述替换序列在长度上包含0个核苷酸;并且所述3'同源臂在长度上包含约200个核苷酸,例如,长度为至少25、50、75、100、125、150、175、或200个核苷酸。

[1086] 实施方式106.如实施方式105所述的核酸组合物,其中所述5'同源臂包含约50bp至100bp,例如,55bp至95bp、60bp至90bp、70bp至90bp、或80bp至90bp、HBG靶位置内的靶位点同源性5′,并且所述3'同源臂包含约50bp至100bp,例如,55bp至95bp、60bp至90bp、70bp至90bp、或80bp至90bp、HBG靶位置内的靶位点同源性3'。

[1087] 实施方式107.如实施方式106所述的核酸组合物,其中所述靶位点选自下组,该组由以下组成:HBG1 c.-114至-102(例如,SEQ ID NO:902(HBG1)的核苷酸2824-2836)、HBG1 c.-225至-222(例如,SEQ ID NO:902(HBG1)的核苷酸2716-2719)、和HBG2 c.-114至-102(例如,SEQ ID NO:903(HBG2)的核苷酸2748-2760)。

[1088] 实施方式108.如实施方式107所述的核酸组合物,其中所述靶位点是HBG1c.-114

至-102(例如,SEQ ID NO:902(HBG1)的核苷酸2824-2836),并且所述5′同源臂包含约50bp至100bp,例如,55bp至95bp、60bp至90bp、70bp至90bp、或80bp至90bp,HBG1 c.-114至-102(例如,SEQ ID NO:902(HBG1)的核苷酸2824-2836)的同源性5'。

[1089] 实施方式109.如实施方式108所述的核酸组合物,其中所述5'同源臂包含SEQ ID NO:904(即,ssODN1 5'同源臂),基本上由其组成或由其组成。

[1090] 实施方式110.如实施方式108或109所述的核酸组合物,其中所述3′同源臂包含约50bp至100bp,例如,55bp至95bp、60bp至90bp、70bp至90bp、或80bp至90bp,HBG1 c.-114至-102(例如,SEQ ID NO:902(HBG1)的核苷酸2824-2836)的同源性3'。

[1091] 实施方式111.如实施方式108-110中任一项所述的核酸组合物,其中所述3′同源臂包含SEQ ID NO:905(即,ssODN1 3'同源臂),基本上由其组成或由其组成。

[1092] 实施方式112.如实施方式107所述的核酸组合物,其中所述靶位点是HBG2c.-114至-102(例如,SEQ ID NO:903(HBG2)的核苷酸2748-2760),并且所述5′同源臂包含约50bp至100bp,例如,55bp至95bp、60bp至90bp、70bp至90bp、或80bp至90bp,HBG2 c.-114至-102(例如,SEQ ID NO:903(HBG2)的核苷酸2748-2760)的同源性5′。

[1093] 实施方式113.如实施方式112所述的核酸组合物,其中所述5′同源臂包含SEQ ID NO:904(即,ssODN1 5′同源臂),基本上由其组成或由其组成。

[1094] 实施方式114.如实施方式112或113所述的核酸组合物,其中所述3′同源臂包含约50bp至100bp,例如,55bp至95bp、60bp至90bp、70bp至90bp、或80bp至90bp,HBG2 c.-114至-102(例如,SEQ ID NO:903(HBG2)的核苷酸2748-2760)的同源性3′。

[1095] 实施方式115.如实施方式112-114中任一项所述的核酸组合物,其中所述3′同源臂包含SEQ ID NO:905(即,ssODN1 3′同源臂),基本上由其组成或由其组成。

[1096] 实施方式116.如实施方式108-115中任一项所述的核酸组合物,其中所述模板核酸包含SEQ ID NO:906(即,ssODN1),基本上由其组成或由其组成。

[1097] 实施方式117.如实施方式107所述的核酸组合物,其中所述靶位点是HBG1c.-225至-222(例如,SEQ ID NO:902(HBG1)的核苷酸2716-2719),并且所述5′同源臂包含约50bp至100bp,例如,55bp至95bp、60bp至90bp、70bp至90bp、或80bp至90bp,HBG1 c.-225至-222(例如,SEQ ID NO:902(HBG1)的核苷酸2716-2719)的同源性5′。

[1098] 实施方式118.如实施方式117所述的核酸组合物,其中所述3′同源臂包含约50bp至100bp,例如,55bp至95bp、60bp至90bp、70bp至90bp、或80bp至90bp,HBG1 c.-225至-222(例如,SEQ ID NO:902(HBG1)的核苷酸2716-2719)的同源性3′。

[1099] 实施方式119.如实施方式103-118中任一项所述的核酸组合物,其中所述ssODN包含5′硫代磷酸酯修饰。

[1100] 实施方式120.如实施方式103-118中任一项所述的核酸组合物,其中所述ssODN包含3′硫代磷酸酯修饰。

[1101] 实施方式121.如实施方式103-118中任一项所述的核酸组合物,其中所述ssODN包含5'硫代磷酸酯修饰和3′硫代磷酸酯修饰。

[1102] 实施方式122.如实施方式34-121中任一项所述的核酸组合物,其中所述核酸组合物不包含(c)编码第二gRNA分子的核苷酸序列、(d)编码第三gRNA分子的核苷酸序列、或(e)编码第四gRNA分子的核苷酸序列。

[1103] 实施方式123.如实施方式34-122中任一项所述的核酸组合物,其中(a)和(b)存在于一个核酸分子上。

[1104] 实施方式124.如实施方式101-122中任一项所述的核酸组合物,其中(a)、(b)、和(g)存在于一个核酸分子上。

[1105] 实施方式125.如实施方式123或124所述的核酸组合物,其中所述核酸分子是AAV载体或LV载体。

[1106] 实施方式126.如实施方式34-122中任一项所述的核酸组合物,其中:(a)存在于第一核酸分子上;并且(b)存在于第二核酸分子上。

[1107] 实施方式127.如实施方式126所述的核酸组合物,其中所述第一和第二核酸分子是AAV载体或LV载体。

[1108] 实施方式128.如实施方式74-122中任一项所述的核酸组合物,其中(a)和(c)存在于一个核酸分子上。

[1109] 实施方式129.如实施方式101-122中任一项所述的核酸组合物,其中(a)和(g)存在于一个核酸分子上。

[1110] 实施方式130.如实施方式128或129所述的核酸组合物,其中所述核酸分子是AAV载体或LV载体。

[1111] 实施方式131.如实施方式74-122中任一项所述的核酸组合物,其中:(a)存在于第一核酸分子上;并且(c)存在于第二核酸分子上。

[1112] 实施方式132.如实施方式101-122中任一项所述的核酸组合物,其中:(a)存在于第一核酸分子上;并且(g)存在于第二核酸分子上。

[1113] 实施方式133.如实施方式130或131所述的核酸组合物,其中所述第一和第二核酸分子是AAV载体或LV载体。

[1114] 实施方式134.如实施方式60-122中任一项所述的核酸组合物,其中(a)、(b)和(c)存在于一个核酸分子上。

[1115] 实施方式135.如实施方式101-122中任一项所述的核酸组合物,其中:(a)、(b)、和(g)存在于一个核酸分子上。

[1116] 实施方式136.如实施方式134或135所述的核酸组合物,其中所述核酸分子是AAV载体或LV载体。

[1117] 实施方式137.如实施方式60-122中任一项所述的核酸组合物,其中:

[1118] (a)、(b)、和(c)之一在第一核酸分子上进行编码;并且

[1119] (a)、(b)、和(c)中的第二个和第三个在第二核酸分子上进行编码。

[1120] 实施方式138.如实施方式101-122中任一项所述的核酸组合物,其中:

[1121] (a)、(b)、和(g)之一在第一核酸分子上进行编码;以及

[1122] (a)、(b)、和(g)中的第二个和第三个在第二核酸分子上进行编码。

[1123] 实施方式139.如实施方式137或138所述的核酸组合物,其中所述第一和第二核酸分子是AAV载体或LV载体。

[1124] 实施方式140.如实施方式137或139所述的核酸组合物,其中:(a)存在于第一核酸分子上;并且(b)和(c)存在于第二核酸分子上。

[1125] 实施方式141.如实施方式138或139所述的核酸组合物,其中:(a)存在于第一核酸

分子上;并且(b)和(g)存在于第二核酸分子上。

[1126] 实施方式142.如实施方式140或141所述的核酸组合物,其中所述第一和第二核酸分子是AAV载体或LV载体。

[1127] 实施方式143.如实施方式137或139所述的核酸组合物,其中:(b)存在于第一核酸分子上;并且(a)和(c)存在于第二核酸分子上。

[1128] 实施方式144.如实施方式138或139所述的核酸组合物,其中:(b)存在于第一核酸分子上;并且(a)和(g)存在于第二核酸分子上。

[1129] 实施方式145.如实施方式143或144所述的核酸组合物,其中所述第一和第二核酸分子是AAV载体或LV载体。

[1130] 实施方式146.如实施方式137或139所述的核酸组合物,其中:(c)存在于第一核酸分子上;并且(b)和(a)存在于第二核酸分子上。

[1131] 实施方式147.如实施方式138或139所述的核酸组合物,其中:(g)存在于第一核酸分子上;并且(b)和(a)存在于第二核酸分子上。

[1132] 实施方式148.如实施方式146或147所述的核酸组合物,其中所述第一和第二核酸分子是AAV载体或LV载体。

[1133] 实施方式149.如实施方式126、131、132、137、139、140、141、143、144、146、或147中任一项所述的核酸组合物,其中所述第一核酸分子不同于AAV载体并且所述第二核酸分子是AAV载体。

[1134] 实施方式150.如实施方式34-149中任一项所述的核酸组合物,其中所述核酸组合物包含可操作地连接至(a)上的启动子。

[1135] 实施方式151.如实施方式74-149中任一项所述的核酸组合物,其中所述核酸组合物包含可操作地连接至(c)上的第二启动子。

[1136] 实施方式152.如实施方式151中任一项所述的核酸组合物,其中所述启动子和第二启动子彼此不同。

[1137] 实施方式153.如实施方式151中任一项所述的核酸组合物,其中所述启动子和第二启动子相同。

[1138] 实施方式154.如实施方式60-149中任一项所述的核酸组合物,其中所述核酸组合物包含可操作地连接至(b)上的启动子。

[1139] 实施方式155.一种组合物,所述组合物包含(a)如实施方式1-33中任一项所述的gRNA分子。

[1140] 实施方式156.如实施方式155所述的组合物,所述组合物进一步包含(b)RNA指导的核酸酶。

[1141] 实施方式157.如实施方式156所述的组合物,其中所述RNA指导的核酸酶是Cas9分子或Cas9-融合蛋白。

[1142] 实施方式158.如实施方式157所述的组合物,其中所述Cas9分子是具有酶活性的Cas9(eaCas9)分子。

[1143] 实施方式159.如实施方式158所述的组合物,其中所述eaCas9分子包含切口酶分子。

[1144] 实施方式160.如实施方式158或159所述的组合物,其中所述eaCas9分子在靶核酸

中形成双链断裂。

[1145] 实施方式161.如实施方式158或159所述的组合物,其中所述eaCas9分子在靶核酸中形成单链断裂。

[1146] 实施方式162.如实施方式161所述的组合物,其中所述单链断裂形成于所述靶核酸的与所述gRNA分子的靶向结构域互补的链中。

[1147] 实施方式163.如实施方式161所述的组合物,其中所述单链断裂形成于所述靶核酸的不同于与所述gRNA分子的靶向结构域互补的链的链中。

[1148] 实施方式164.如实施方式158-163中任一项所述的组合物,其中所述eaCas9分子包含HNH样结构域切割活性,但没有或具有不显著的N-末端RuvC样结构域切割活性。

[1149] 实施方式165.如实施方式164所述的组合物,其中所述eaCas9分子是HNH样结构域切口酶。

[1150] 实施方式166.如实施方式164或165所述的组合物,其中所述eaCas9分子在D10处包含突变。

[1151] 实施方式167.如实施方式158-163中任一项所述的组合物,其中所述eaCas9分子包含N端RuvC样结构域切割活性,但没有或具有不显著的HNH样结构域切割活性。

[1152] 实施方式168.如实施方式167所述的组合物,其中所述eaCas9分子是N端RuvC样结构域切口酶。

[1153] 实施方式169.如实施方式167或168所述的组合物,其中所述eaCas9分子在H840或N863处包含突变。

[1154] 实施方式170.如实施方式156-169中任一项所述的组合物,所述组合物进一步包含(c)包含靶向结构域的第二gRNA分子,所述靶向结构域包含与完全或部分位于HBG1或HBG2调控区域内的靶结构域互补或部分互补的核苷酸序列。

[1155] 实施方式171.如实施方式170所述的组合物,其中所述第二gRNA分子是如实施方式1-33中任一项所述的gRNA分子。

[1156] 实施方式172.如实施方式156-171中任一项所述的组合物,所述组合物进一步包含(d)第三gRNA分子。

[1157] 实施方式173.如实施方式172所述的组合物,所述组合物进一步包含(e)第四gRNA分子。

[1158] 实施方式174.如实施方式155-173中任一项所述的组合物,所述组合物进一步包含(g)模板核酸。

[1159] 实施方式175.如实施方式174所述的组合物,其中所述模板核酸是单链寡脱氧核苷酸(ssODN)。

[1160] 实施方式176.如实施方式175所述的组合物,其中所述模板核酸包含5′同源臂、替换序列、和3’同源臂。

[1161] 实施方式177.如实施方式176所述的组合物,其中所述5’同源臂在长度上包含约200个核苷酸,例如,长度为至少25、50、75、100、125、150、175、或200个核苷酸;所述替换序列在长度上包含0个核苷酸;并且所述3′同源臂在长度上包含约200个核苷酸,例如,长度为至少25、50、75、100、125、150、175、或200个核苷酸。

[1162] 实施方式178.如实施方式177所述的组合物,其中所述5′同源臂包含约50bp至

100bp,例如,55bp至95bp、60bp至90bp、70bp至90bp、或80bp至90bp、HBG靶位置内的靶位点同源性5',并且所述3'同源臂包含约50bp至100bp,例如,55bp至95bp、60bp至90bp、70bp至90bp、或80bp至90bp、HBG靶位置内的靶位点同源性3′。

[1163] 实施方式179.如实施方式178所述的组合物,其中所述靶位点选自下组,该组由以下组成:HBG1 c.-114至-102(例如,SEQ ID NO:902(HBG1)的核苷酸2824-2836)、HBG1 c.-225至-222(例如,SEQ ID NO:902(HBG1)的核苷酸2716-2719)、和HBG2c.-114至-102(例如,SEQ ID NO:903(HBG2)的核苷酸2748-2760)。

[1164] 实施方式180.如实施方式179所述的组合物,其中所述靶位点是HBG1 c.-114至-102(例如,SEQ ID NO:902(HBG1)的核苷酸2824-2836),并且所述5′同源臂包含约50bp至100bp,例如,55bp至95bp、60bp至90bp、70bp至90bp、或80bp至90bp,HBG1 c.-114至-102(例如,SEQ ID NO:902(HBG1)的核苷酸2824-2836)的同源性5′。

[1165] 实施方式181.如实施方式180所述的组合物,其中所述5′同源臂包含SEQ ID NO:904(即,ssODN1 5'同源臂),基本上由其组成或由其组成。

[1166] 实施方式182.如实施方式180或181所述的组合物,其中所述3′同源臂包含约50bp至100bp,例如,55bp至95bp、60bp至90bp、70bp至90bp、或80bp至90bp,HBG1 c.-114至-102(例如,SEQ ID NO:902(HBG1)的核苷酸2824-2836)的同源性3'。

[1167] 实施方式183.如实施方式179-182中任一项所述的组合物,其中所述3'同源臂包含SEQ ID NO:905(即,ssODN1 3'同源臂),基本上由其组成或由其组成。

[1168] 实施方式184.如实施方式179所述的组合物,其中所述靶位点是HBG2 c.-114至-102(例如,SEQ ID NO:903(HBG2)的核苷酸2748-2760),并且所述5'同源臂包含约50bp至100bp,例如,55bp至95bp、60bp至90bp、70bp至90bp、或80bp至90bp,HBG2 c.-114至-102(例如,SEQ ID NO:903(HBG2)的核苷酸2748-2760)的同源性5'。

[1169] 实施方式185.如实施方式184所述的组合物,其中所述5′同源臂包含SEQ ID NO:904(即,ssODN1 5'同源臂),基本上由其组成或由其组成。

[1170] 实施方式186.如实施方式184或185所述的组合物,其中所述3'同源臂包含约50bp至100bp,例如,55bp至95bp、60bp至90bp、70bp至90bp、或80bp至90bp,HBG2 c.-114至-102(例如,SEQ ID NO:903(HBG2)的核苷酸2748-2760)的同源性3'。

[1171] 实施方式187.如实施方式184-186中任一项所述的组合物,其中所述3′同源臂包含SEQ ID NO:905(即,ssODN1 3'同源臂),基本上由其组成或由其组成。

[1172] 实施方式188.如实施方式175-187中任一项所述的组合物,其中所述模板核酸包含SEQ ID NO:906(即,ssODN1),基本上由其组成或由其组成。

[1173] 实施方式189.如实施方式179所述的组合物,其中所述靶位点是HBG1 c.-225至-222(例如,SEQ ID NO:902(HBG1)的核苷酸2716-2719),并且所述5'同源臂包含约50bp至100bp,例如,55bp至95bp、60bp至90bp、70bp至90bp、或80bp至90bp,HBG1 c.-225至-222(例如,SEQ ID NO:902(HBG1)的核苷酸2716-2719)的同源性5'。

[1174] 实施方式190.如实施方式189所述的组合物,其中所述3′同源臂包含约50bp至100bp,例如,55bp至95bp、60bp至90bp、70bp至90bp、或80bp至90bp,HBG1c.-225至-222(例如,SEQ ID NO:902(HBG1)的核苷酸2716-2719)的同源性3'。

[1175] 实施方式191.如实施方式175-190中任一项所述的组合物,其中所述ssODN包含5'

硫代磷酸酯修饰。

[1176] 实施方式192.如实施方式175-190中任一项所述的组合物,其中所述ssODN包含3′硫代磷酸酯修饰。

[1177] 实施方式193.如实施方式175-190中任一项所述的组合物,其中所述ssODN包含5’硫代磷酸酯修饰和3’硫代磷酸酯修饰。

[1178] 实施方式194.一种改变细胞的方法,所述方法包括使所述细胞与以下接触:

[1179] (a)如实施方式1-33中任一项所述的gRNA分子;以及

[1180] (b)RNA指导的核酸酶。

[1181] 实施方式195.如实施方式194所述的方法,所述方法进一步包括使所述细胞与以下接触:(c)包含靶向结构域的第二gRNA分子,所述靶向结构域包含与完全或部分位于HBG1或HBG2调控区域内的靶结构域互补或部分互补的核苷酸序列。

[1182] 实施方式196.如实施方式194或195所述的方法,其中所述RNA指导的核酸酶是Cas9分子或Cas9-融合蛋白。

[1183] 实施方式197.如实施方式196所述的方法,其中所述第二gRNA分子是如实施方式1-33中任一项所述的gRNA分子。

[1184] 实施方式198.如实施方式195-196中任一项所述的方法,其中所述Cas9分子是具有酶活性的Cas9(eaCas9)分子。

[1185] 实施方式199.如实施方式198所述的方法,其中所述eaCas9分子包含切口酶分子。

[1186] 实施方式200.如实施方式198或199所述的方法,其中所述eaCas9分子在靶核酸中形成双链断裂。

[1187] 实施方式201.如实施方式198或199所述的方法,其中所述eaCas9分子在靶核酸中形成单链断裂。

[1188] 实施方式202.如实施方式201所述的方法,其中所述单链断裂形成于所述靶核酸的与所述gRNA分子的靶向结构域互补的链中。

[1189] 实施方式203.如实施方式201所述的方法,其中所述单链断裂形成于所述靶核酸的不同于与所述gRNA分子的靶向结构域互补的链的链中。

[1190] 实施方式204.如实施方式198-203中任一项所述的组合方法物,其中所述eaCas9分子包含HNH样结构域切割活性,但没有或具有不显著的N-末端RuvC样结构域切割活性。

[1191] 实施方式205.如实施方式204所述的方法,其中所述eaCas9分子是HNH样结构域切口酶。

[1192] 实施方式206.如实施方式204或205所述的方法,其中所述eaCas9分子在D10处包含突变。

[1193] 实施方式207.如实施方式198-203中任一项所述的方法,其中所述eaCas9分子包含N端RuvC样结构域切割活性,但没有或具有不显著的HNH样结构域切割活性。

[1194] 实施方式208.如实施方式207所述的方法,其中所述eaCas9分子是N端RuvC样结构域切口酶。

[1195] 实施方式209.如实施方式207或208所述的方法,其中所述eaCas9分子在H840或N863处包含突变。

[1196] 实施方式210.如实施方式196-209中任一项所述的方法,所述方法进一步包括使

所述细胞与(d)第三gRNA分子接触。

[1197] 实施方式211.如实施方式210所述的方法,所述方法进一步包括使所述细胞与(e)第四gRNA分子接触。

[1198] 实施方式212.如实施方式194-211中任一项所述的方法,其中所述细胞来自患有β-血红蛋白病的受试者。

[1199] 实施方式213.如实施方式212所述的方法,其中所述β-血红蛋白病选自由SCD和β-Thal组成的组。

[1200] 实施方式214.如实施方式194-213中任一项所述的方法,其中所述细胞是红系细胞。

[1201] 实施方式215.如实施方式214所述的方法,其中所述细胞是成红细胞。

[1202] 实施方式216.如实施方式194-215中任一项所述的方法,其中所述接触步骤在体内进行。

[1203] 实施方式217.如实施方式194-216中任一项所述的方法,所述方法包括获取所述细胞中的所述HBG靶位置的序列的信息。

[1204] 实施方式218.如实施方式194-217中任一项所述的方法,所述方法包括向HBG靶位置引入indel。

[1205] 实施方式219.如实施方式218所述的方法,其中所述indel选自下组,该组由以下组成:HBG1 13bp del-114至-102、HBG1 4bp del-225至-222、和HBG2 13bp del-114至-102。

[1206] 实施方式220.如实施方式218或219所述的方法,其中所述indel是使用NHEJ引入的。

[1207] 实施方式221.如实施方式194-220中任一项所述的方法,所述方法包括向HBG靶位置引入单核苷酸改变。

[1208] 实施方式222.如实施方式221所述的方法,其中所述单核苷酸改变选自下组,该组由以下组成:HBG1 c.-114C>T、c.-117G>A、c.-158C>T、c.-167C>T、c.-170G>A、c.-175T>G、c.-175T>C、c.-195C>G、c.-196C>T、c.-198T>C、c.-201C>T、c.-251T>C、或c.-499T>A、和HBG2 c.-109G>T、c.-114C>A、c.-114C>T、c.-157C>T、c.-158C>T、c.-167C>T、c.-167C>A、c.-175T>C、c.-202C>G、c.-211C>T、c.-228T>C、c.-255C>G、c.-309A>G、c.-369C>G、或c.-567T>G。

[1209] 实施方式223.如实施方式221或222所述的方法,其中所述单核苷酸改变是使用HDR引入的。

[1210] 实施方式224.如实施方式194-223中任一项所述的方法,所述方法包括向HBG靶位置的靶位点引入改变。

[1211] 实施方式225.如实施方式224所述的方法,其中所述改变选自下组,该组由以下组成:HBG1 13bp del-114至-102、HBG1 4bp del-225至-222、和HBG2 13bp del-114至-102。

[1212] 实施方式226.如实施方式224或225所述的方法,其中所述改变是使用HDR引入的。

[1213] 实施方式227.如实施方式226所述的方法,所述方法进一步包括使所述细胞与(g)模板核酸接触。

[1214] 实施方式228.如实施方式227所述的方法,其中所述模板核酸是单链寡脱氧核苷

酸(ssODN)。

[1215] 实施方式229.如实施方式228所述的方法,其中所述模板核酸包含5′同源臂、替换序列、和3'同源臂。

[1216] 实施方式230.如实施方式229所述的方法,其中所述5'同源臂在长度上包含约200个核苷酸,例如,长度为至少25、50、75、100、125、150、175、或200个核苷酸;所述替换序列在长度上包含0个核苷酸;并且所述3'同源臂在长度上包含约200个核苷酸,例如,长度为至少25、50、75、100、125、150、175、或200个核苷酸。

[1217] 实施方式231.如实施方式230所述的方法,其中所述5′同源臂包含约50bp至100bp,例如,55bp至95bp、60bp至90bp、70bp至90bp、或80bp至90bp、靶位点同源性5′,并且所述3'同源臂包含约50bp至100bp,例如,55bp至95bp、60bp至90bp、70bp至90bp、或80bp至90bp、靶位点同源性3'。

[1218] 实施方式232.如实施方式231所述的方法,其中所述改变是HBG1 13bp del-114至-102并且所述靶位点是HBG1 c.-114至-102(例如,SEQ ID NO:902(HBG1)的核苷酸2824-2836)。

[1219] 实施方式233.如实施方式232所述的方法,其中所述5′同源臂包含约50bp至100bp,例如,55bp至95bp、60bp至90bp、70bp至90bp、或80bp至90bp,HBG1c.-114至-102(例如,SEQ ID NO:902(HBG1)的核苷酸2824-2836)的同源性5′。

[1220] 实施方式234.如实施方式230-233中任一项所述的方法,其中所述3'同源臂包含约50bp至100bp,例如,55bp至95bp、60bp至90bp、70bp至90bp、或80bp至90bp,HBG1 c.-114至-102(例如,SEQ ID NO:902(HBG1)的核苷酸2824-2836)的同源性3′。

[1221] 实施方式235.如实施方式234所述的方法,其中所述改变是HBG2 13bp del-114至-102并且靶位点是HBG2 c.-114至-102(例如,SEQ ID NO:903(HBG2)的核苷酸2748-2760)。

[1222] 实施方式236.如实施方式235所述的方法,其中所述5'同源臂包含约50bp至100bp,例如,55bp至95bp、60bp至90bp、70bp至90bp、或80bp至90bp,HBG2c.-114至-102(例如,SEQ ID NO:903(HBG2)的核苷酸2748-2760)的同源性5'。

[1223] 实施方式237.如实施方式234或235所述的方法,其中所述3′同源臂包含约50bp至100bp,例如,55bp至95bp、60bp至90bp、70bp至90bp、或80bp至90bp,HBG2 c.-114至-102(例如,SEQ ID NO:903(HBG2)的核苷酸2748-2760)的同源性3'。

[1224] 实施方式238.如实施方式232-237中任一项所述的方法,其中所述5'同源臂包含SEQ ID NO:904(ssODN1 5′同源臂),基本上由其组成或由其组成。

[1225] 实施方式239.如实施方式232-238中任一项所述的方法,其中所述3′同源臂包含SEQ ID NO:905(ssODN1 3′同源臂),基本上由其组成或由其组成。

[1226] 实施方式240.如实施方式232-239中任一项所述的方法,其中所述模板核酸包含SEQ ID NO:906(ssODN1),基本上由其组成或由其组成。

[1227] 实施方式241.如实施方式231所述的方法,其中所述改变是HBG1 4bp del-225至-222并且靶位点是HBG1 c.-225至-222(例如,SEQ ID NO:902(HBG1)的核苷酸2716-2719)。

[1228] 实施方式242.如实施方式241所述的方法,其中所述5′同源臂包含约50bp至100bp,例如,55bp至95bp、60bp至90bp、70bp至90bp、或80bp至90bp,HBG1c.-225至-222(例

如,SEQ ID NO:902(HBG1)的核苷酸2716-2719)的同源性5′。

[1229] 实施方式243.如实施方式241或242所述的方法,所述3′同源臂包含约50bp至100bp,例如,55bp至95bp、60bp至90bp、70bp至90bp、或80bp至90bp,HBG1c.-225至-222(例如,SEQ ID NO:902(HBG1)的核苷酸2716-2719)的同源性3′。

[1230] 实施方式244.如实施方式228-243中任一项所述的方法,其中所述ssODN包含5′硫代磷酸酯修饰。

[1231] 实施方式245.如实施方式228-243中任一项所述的方法,其中所述ssODN包含3′硫代磷酸酯修饰。

[1232] 实施方式246.如实施方式228-243中任一项所述的方法,其中所述ssODN包含5′硫代磷酸酯修饰和3′硫代磷酸酯修饰。

[1233] 实施方式247.如实施方式195-246中任一项所述的方法,其中所述接触步骤包含使所述细胞与编码(a)、(b)、和(c)中的至少一个的核酸组合物接触。

[1234] 实施方式248.如实施方式227-246中任一项所述的方法,其中所述接触步骤包含使所述细胞与编码(a)、(b)、(g)、和任选地(c)的核酸组合物接触。

[1235] 实施方式249.如实施方式248或249所述的方法,其中所述接触步骤包含使所述细胞与如实施方式34-154中任一项所述的核酸组合物接触。

[1236] 实施方式250.如实施方式195-249中任一项所述的方法,其中所述接触步骤包含向所述细胞递送所述(b)和编码(a)的核酸组合物。

[1237] 实施方式251.如实施方式250所述的方法,其中所述核酸组合物进一步编码(c)。

[1238] 实施方式252.如实施方式250或251所述的方法,其中所述核酸组合物进一步编码(g)。

[1239] 实施方式253.如实施方式195-251中任一项所述的方法,其中所述接触步骤包含向所述细胞递送(a)和(b)。

[1240] 实施方式254.如实施方式195-251中任一项所述的方法,其中所述接触步骤包含向所述细胞递送(a)和编码(b)的核酸组合物。

[1241] 实施方式255.如实施方式253或254所述的方法,其中所述接触步骤进一步包含向所述细胞递送(c)。

[1242] 实施方式256.如实施方式227-255中任一项所述的方法,其中所述接触步骤进一步包含向所述细胞递送(g)。

[1243] 实施方式257.一种治疗有需要的受试者中的β-血红蛋白病的方法,所述方法包括使所述受试者或来自所述受试者的细胞与以下接触:

[1244] (a)如实施方式1-33中任一项所述的gRNA分子;以及

[1245] (b)RNA指导的核酸酶。

[1246] 实施方式258.如实施方式257所述的方法,其中所述RNA指导的核酸酶是Cas9分子或Cas9-融合蛋白。

[1247] 实施方式259.如实施方式257或258所述的方法,其中所述β-血红蛋白病选自由SCD和β-Tha1组成的组。

[1248] 实施方式260.如实施方式257-259中任一项所述的方法,所述方法进一步包括使所述受试者或来自所述受试者的所述细胞与以下接触:(c)包含靶向结构域的第二gRNA分

子,所述靶向结构域包含与完全或部分位于HBG1或HBG2调控区域内的靶结构域互补或部分互补的核苷酸序列。

[1249] 实施方式261.如实施方式260所述的方法,其中所述第二gRNA分子是如实施方式1-33中任一项所述的gRNA分子。

[1250] 实施方式262.如实施方式258-261中任一项所述的方法,其中所述Cas9分子是具有酶活性的Cas9(eaCas9)分子。

[1251] 实施方式263.如实施方式262所述的方法,其中所述eaCas9分子包含切口酶分子。

[1252] 实施方式264.如实施方式262或263所述的方法,其中所述eaCas9分子在靶核酸中形成双链断裂。

[1253] 实施方式265.如实施方式262或263所述的方法,其中所述eaCas9分子在靶核酸中形成单链断裂。

[1254] 实施方式266.如实施方式265所述的方法,其中所述单链断裂形成于所述靶核酸的与所述gRNA分子的靶向结构域互补的链中。

[1255] 实施方式267.如实施方式265所述的方法,其中所述单链断裂形成于所述靶核酸的不同于与所述gRNA分子的靶向结构域互补的链的链中。

[1256] 实施方式268.如实施方式263-267中任一项所述的组合方法物,其中所述eaCas9分子包含HNH样结构域切割活性,但没有或具有不显著的N-末端RuvC样结构域切割活性。

[1257] 实施方式269.如实施方式268所述的方法,其中所述eaCas9分子是HNH样结构域切口酶。

[1258] 实施方式270.如实施方式268或269所述的方法,其中所述eaCas9分子在D10处包含突变。

[1259] 实施方式271.如实施方式262-270中任一项所述的方法,其中所述eaCas9分子包含N端RuvC样结构域切割活性,但没有或具有不显著的HNH样结构域切割活性。

[1260] 实施方式272.如实施方式271所述的方法,其中所述eaCas9分子是N端RuvC样结构域切口酶。

[1261] 实施方式273.如实施方式271或272中任一项所述的方法,其中所述eaCas9分子在H840或N863处包含突变。

[1262] 实施方式274.如实施方式257-273中任一项所述的方法,所述方法进一步包括使所述受试者或来自所述受试者的所述细胞与(d)第三gRNA分子接触。

[1263] 实施方式275.如实施方式274所述的方法,所述方法进一步包括使所述受试者或来自所述受试者的所述细胞与第四gRNA分子接触。

[1264] 实施方式276.如实施方式257-275中任一项所述的方法,所述方法包括向HBG靶位置引入单核苷酸改变。

[1265] 实施方式277.如实施方式276所述的方法,其中所述单核苷酸改变选自下组,该组由以下组成:HBG1 c.-114C>T、c.-117G>A、c.-158C>T、c.-167C>T、c.-170G>A、c.-175T>G、c.-175T>C、c.-195C>G、c.-196C>T、c.-198T>C、c.-201C>T、c.-251T>C、或c.-499T>A、和HBG2 c.-109G>T、c.-114C>A、c.-114C>T、c.-157C>T、c.-158C>T、c.-167C>T、c.-167C>A、c.-175T>C、c.-202C>G、c.-211C>T、c.-228T>C、c.-255C>G、c.-309A>G、c.-369C>G、或c.-567T>G。

[1266] 实施方式278.如实施方式276或277所述的方法,其中所述单核苷酸改变是使用HDR引入的。

[1267] 实施方式279.如实施方式257-278中任一项所述的方法,所述方法包括向HBG靶位置内的靶位点引入indel。

[1268] 实施方式280.如实施方式279所述的方法,其中所述indel选自下组,该组由以下组成:HBG1 13bp del-114至-102、HBG1 4bp del-225至-222、和HBG2 13bp del-114至-102。

[1269] 实施方式281.如实施方式279或280所述的方法,其中所述改变是使用HDR引入的。

[1270] 实施方式282.如实施方式281所述的方法,所述方法进一步包括使所述受试者或来自所述受试者的所述细胞与(g)模板核酸接触。

[1271] 实施方式283.如实施方式282所述的方法,其中所述模板核酸是单链寡脱氧核苷酸(ssODN)。

[1272] 实施方式284.如实施方式283所述的方法,其中所述模板核酸包含5′同源臂、替换序列、和3'同源臂,其中所述替换序列是0个核苷酸。

[1273] 实施方式285.如实施方式284所述的方法,其中所述5'同源臂在长度上包含约200个核苷酸,例如,长度为至少25、50、75、100、125、150、175、或200个核苷酸;所述替换序列在长度上包含0个核苷酸;并且所述3'同源臂在长度上包含约200个核苷酸,例如,长度为至少25、50、75、100、125、150、175、或200个核苷酸。

[1274] 实施方式286.如实施方式285所述的方法,其中所述5′同源臂包含约50bp至100bp,例如,55bp至95bp、60bp至90bp、70bp至90bp、或80bp至90bp、HBG靶位置内的靶位点同源性5',并且所述3′同源臂包含约50bp至100bp,例如,55bp至95bp、60bp至90bp、70bp至90bp、或80bp至90bp、HBG靶位置内的靶位点同源性3′。

[1275] 实施方式287.如实施方式286所述的方法,其中所述indel是HBG1 13bp del-114至-102并且靶位点是SEQ ID NO:902的HBG1 -114至-102。

[1276] 实施方式288.如实施方式287所述的方法,其中所述5′同源臂包含约50bp至100bp,例如,55bp至95bp、60bp至90bp、70bp至90bp、或80bp至90bp,HBG1c.-114至-102(例如,SEQ ID NO:902(HBG1)的核苷酸2824-2836)的同源性5'。

[1277] 实施方式289.如实施方式287或288所述的方法,其中所述3'同源臂包含约50bp至100bp,例如,55bp至95bp、60bp至90bp、70bp至90bp、或80bp至90bp,HBG1 c.-114至-102(例如,SEQ ID NO:902(HBG1)的核苷酸2824-2836)的同源性3'。

[1278] 实施方式290.如实施方式286所述的方法,其中所述indel是HBG2 13bp del-114至-102并且靶位点是HBG2 c.-114至-102(例如,SEQ ID NO:903(HBG2)的核苷酸2748-2760)。

[1279] 实施方式291.如实施方式290所述的方法,其中所述5′同源臂包含约50bp至100bp,例如,55bp至95bp、60bp至90bp、70bp至90bp、或80bp至90bp,HBG2c.-114至-102(例如,SEQ ID NO:903(HBG2)的核苷酸2748-2760)的同源性5'。

[1280] 实施方式292.如实施方式290或291所述的方法,其中所述3'同源臂包含约50bp至100bp,例如,55bp至95bp、60bp至90bp、70bp至90bp、或80bp至90bp,HBG2 c.-114至-102(例如,SEQ ID NO:903(HBG2)的核苷酸2748-2760)的同源性3'。

[1281] 实施方式293.如实施方式290-292中任一项所述的方法,其中所述5′同源臂包含SEQ ID NO:904(ssODN1 5′同源臂),基本上由其组成或由其组成。

[1282] 实施方式294.如实施方式290-293中任一项所述的方法,其中所述3'同源臂包含SEQ ID NO:905(ssODN1 3'同源臂),基本上由其组成或由其组成。

[1283] 实施方式295.如实施方式283-294中任一项所述的方法,其中所述模板核酸包含SEQ ID NO:906(ssODN1),基本上由其组成或由其组成。

[1284] 实施方式296.如实施方式256所述的方法,其中所述indel是HBG1 4bp del-225至-222并且靶位点是HBG1 c.-225至-222(例如,SEQ ID NO:902(HBG1)的核苷酸2716-2719)。

[1285] 实施方式297.如实施方式296所述的方法,其中所述5'同源臂包含约50bp至100bp,例如,55bp至95bp、60bp至90bp、70bp至90bp、或80bp至90bp,HBG1c.-225至-222(例如,SEQ ID NO:902(HBG1)的核苷酸2716-2719)的同源性5′。

[1286] 实施方式298.如实施方式296或297所述的方法,所述3'同源臂包含约50bp至100bp,例如,55bp至95bp、60bp至90bp、70bp至90bp、或80bp至90bp,HBG1c.-225至-222(例如,SEQ ID NO:902(HBG1)的核苷酸2716-2719)的同源性3′。

[1287] 实施方式299.如实施方式283-298中任一项所述的方法,其中所述ssODN包含5′硫代磷酸酯修饰。

[1288] 实施方式300.如实施方式283-298中任一项所述的方法,其中所述ssODN包含3′硫代磷酸酯修饰。

[1289] 实施方式301.如实施方式258-298中任一项所述的方法,其中所述ssODN包含5′硫代磷酸酯修饰和3′硫代磷酸酯修饰。

[1290] 实施方式302.如实施方式257-301中任一项所述的方法,其中所述接触步骤在体内进行。

[1291] 实施方式303.如实施方式257-302中任一项所述的方法,其中所述接触步骤包含静脉内注射。

[1292] 实施方式304.如实施方式260-303中任一项所述的方法,其中所述接触步骤包含使所述受试者或来自所述受试者的所述细胞与编码(a)、(b)、和(c)中的至少一个的核酸组合物接触。

[1293] 实施方式305.如实施方式282-303中任一项所述的方法,其中所述接触步骤包含使所述受试者或来自所述受试者的所述细胞与编码(a)、(b)、(c)、和(g)中的至少一个的核酸组合物接触。

[1294] 实施方式306.如实施方式257-305中任一项所述的方法,其中所述接触步骤包含使所述受试者或来自所述受试者的所述细胞与如实施方式34-154中任一项所述的核酸组合物接触。

[1295] 实施方式307.如实施方式257-305中任一项所述的方法,其中所述接触步骤包含向所述受试者或来自所述受试者的所述细胞递送(b)和编码(a)的核酸组合物。

[1296] 实施方式308.如实施方式307所述的方法,其中所述核酸组合物进一步编码(c)。

[1297] 实施方式309.如实施方式307或308所述的方法,其中所述核酸组合物进一步编码(g)。

[1298] 实施方式310.如实施方式257-305中任一项所述的方法,其中所述接触步骤包含向所述受试者或来自所述受试者的所述细胞递送(a)和(b)。

[1299] 实施方式311.如实施方式257-305中任一项所述的方法,其中所述接触步骤包含向使所述受试者或来自所述受试者的所述细胞递送(a)和编码(b)的核酸组合物。

[1300] 实施方式312.如实施方式310或311所述的方法,其中所述接触步骤进一步包含向使所述受试者或来自所述受试者的所述细胞递送(c)。

[1301] 实施方式313.如实施方式282-312中任一项所述的方法,其中所述接触步骤进一步包含向所述受试者或来自所述受试者的所述细胞递送(g)。

[1302] 实施方式314.一种反应混合物,所述反应混合物包含:

[1303] (a)如实施方式1-33中任一项所述的gRNA分子、如实施方式34-154中任一项所述的核酸组合物、或如实施方式155-193中任一项所述的组合物;以及

[1304] 来自患有β-血红蛋白病的受试者的细胞。

[1305] 实施方式315.一种试剂盒,所述试剂盒包含,

[1306] (a)如实施方式1-33中任一项所述的gRNA分子、或编码所述gRNA分子的核酸组合物、以及以下中的一个或多个:

[1307] (b)RNA指导的核酸酶;

[1308] (c)包含靶向结构域的第二gRNA分子,所述靶向结构域包含与完全或部分位于HBG1或HBG2调控区域内的靶结构域互补或部分互补的核苷酸序列;以及

[1309] (d)编码(b)和(c)中的一个或多个的核酸组合物。

[1310] 实施方式316.如实施方式315所述的试剂盒,其中所述RNA指导的核酸酶是Cas9分子或Cas9-融合蛋白。

[1311] 实施方式317.如实施方式315或316所述的试剂盒,其中所述第二gRNA分子是如实施方式1-33中任一项所述的gRNA分子。

[1312] 实施方式318.如实施方式315-317中任一项所述的试剂盒,所述试剂盒包含编码(a)、(b)、和(c)中的一个或多个的核酸组合物。

[1313] 实施方式319.如实施方式315-318中任一项所述的试剂盒,所述试剂盒进一步包含第三gRNA分子,所述第三gRNA分子包含靶向结构域,所述靶向结构域包含与完全或部分位于HBG1或HBG2调控区域内的靶结构域互补或部分互补的核苷酸序列。

[1314] 实施方式320.如实施方式319所述的试剂盒,所述试剂盒进一步包含第四gRNA分子,所述第四gRNA分子包含靶向结构域,所述靶向结构域包含与完全或部分位于HBG1或HBG2调控区域内的靶结构域互补或部分互补的核苷酸序列。

[1315] 实施方式321.如实施方式315-320中任一项所述的试剂盒,所述试剂盒进一步包含(g)模板核酸。

[1316] 实施方式322.如实施方式1-33中任一项所述的gRNA分子,用于治疗对其有需要的受试者的β-血红蛋白病。

[1317] 实施方式323.如实施方式291所述的gRNA分子,其中所述gRNA分子与(b)RNA指导的核酸酶组合使用。

[1318] 实施方式324.如实施方式323所述的gRNA分子,其中所述RNA指导的核酸酶是Cas9分子或Cas9-融合蛋白。

[1319] 实施方式325如实施方式322-324中任一项所述的gRNA分子,其中所述gRNA分子与下组合使用:(c)包含靶向结构域的第二gRNA分子,所述靶向结构域包含与完全或部分位于HBG1或HBG2调控区域内的靶结构域互补或部分互补的核苷酸序列。

[1320] 实施方式326.如实施方式322-325中任一项所述的gRNA分子,其中所述gRNA分子与(g)模板核酸组合使用。

[1321] 实施方式327.如实施方式1-33中任一项所述的gRNA分子在生产用于治疗对其有需要的受试者的β-血红蛋白病的药物中的用途。

[1322] 实施方式328.如实施方式327所述的用途,其中所述药物进一步包含(b)RNA指导的核酸酶。

[1323] 实施方式329.如实施方式328所述的用途,其中所述RNA指导的核酸酶是Cas9分子。

[1324] 实施方式330.如实施方式327-329中任一项所述的用途,其中所述药物进一步包含(c)包含靶向结构域的第二gRNA分子,所述靶向结构域包含与完全或部分位于HBG1或HBG2调控区域内的靶结构域互补或部分互补的核苷酸序列。

[1325] 实施方式331.如实施方式327-330中任一项所述的用途,其中所述药物进一步包含(g)模板核酸。

[1326] 实施方式332.一种基因组编辑系统,所述基因组编辑系统包含:

[1327] (a)如实施方式1-33中任一项所述的gRNA分子;以及

[1328] (b)RNA指导的核酸酶。

[1329] 实施方式333.如实施方式332所述的基因组编辑系统,其中所述RNA指导的核酸酶是Cas9分子或Cas9-融合蛋白。

[1330] 实施方式334.如实施方式333所述的基因组编辑系统,其中所述Cas9分子是具有酶活性的Cas9(eaCas9)分子。

[1331] 实施方式335.如实施方式334所述的基因组编辑系统,其中所述eaCas9分子包含切口酶分子。

[1332] 实施方式336.如实施方式334或335所述的基因组编辑系统,其中所述eaCas9分子在靶核酸中形成双链断裂。

[1333] 实施方式337.如实施方式334或335所述的基因组编辑系统,其中所述eaCas9分子在靶核酸中形成单链断裂。

[1334] 实施方式338.如实施方式337所述的基因组编辑系统,其中所述单链断裂形成于所述靶核酸的与所述gRNA分子的靶向结构域互补的链中。

[1335] 实施方式339.如实施方式337所述的基因组编辑系统,其中所述单链断裂形成于所述靶核酸的不同于与所述gRNA分子的靶向结构域互补的链的链中。

[1336] 实施方式340.如实施方式334-339中任一项所述的基因组编辑系统,其中所述eaCas9分子包含HNH样结构域切割活性,但没有或具有不显著的N-末端RuvC样结构域切割活性。

[1337] 实施方式341.如实施方式340所述的基因组编辑系统,其中所述eaCas9分子是HNH样结构域切口酶。

[1338] 实施方式342.如实施方式340或341所述的基因组编辑系统,其中所述eaCas9分子

在D10处包含突变。

[1339] 实施方式343.如实施方式334-342中任一项所述的基因组编辑系统,其中所述eaCas9分子包含N端RuvC样结构域切割活性,但没有或具有不显著的HNH样结构域切割活性。

[1340] 实施方式344.如实施方式343所述的基因组编辑系统,其中所述eaCas9分子是N端RuvC样结构域切口酶。

[1341] 实施方式345.如实施方式343或344所述的基因组编辑系统,其中所述eaCas9分子在H840或N863处包含突变。

[1342] 实施方式346.如实施方式332-345中任一项所述的基因组编辑系统,所述基因组编辑系统进一步包含(c)包含靶向结构域的第二gRNA分子,所述靶向结构域包含与完全或部分位于HBG1或HBG2调控区域内的靶结构域互补或部分互补的核苷酸序列。

[1343] 实施方式347.如实施方式34所述的基因组编辑系统,其中所述第二gRNA分子是如实施方式1-33中任一项所述的gRNA分子。

[1344] 实施方式348.如实施方式332-347中任一项所述的基因组编辑系统,其中所述基因组编辑系统进一步包含(d)第三gRNA分子。

[1345] 实施方式349.如实施方式348所述的基因组编辑系统,其中所述基因组编辑系统进一步包含(e)第四gRNA分子。

[1346] 实施方式350.如实施方式332-349中任一项所述的基因组编辑系统,其中所述基因组编辑系统进一步包含(g)模板核酸。

[1347] 实施方式351.如实施方式350所述的基因组编辑系统,其中所述模板核酸是单链寡脱氧核苷酸(ssODN)。

[1348] 实施方式352.如实施方式351所述的基因组编辑系统,其中所述模板核酸包含5'同源臂、替换序列、和3'同源臂。

[1349] 实施方式353.如实施方式352所述的基因组编辑系统,其中所述5'同源臂在长度上包含约200个核苷酸,例如,长度为至少25、50、75、100、125、150、175、或200个核苷酸;所述替换序列在长度上包含0个核苷酸;并且所述3'同源臂在长度上包含约200个核苷酸,例如,长度为至少25、50、75、100、125、150、175、或200个核苷酸。

[1350] 实施方式354.如实施方式353所述的基因组编辑系统,其中所述5'同源臂包含约50至100bp,例如,55至95、60至90、70至90、或80至90bp、HBG靶位置内的靶位点同源性5',并且所述3'同源臂包含约50至100bp,例如,55至95、60至90、70至90、或80至90bp、HBG靶位置内的靶位点同源性3′。

[1351] 实施方式355.如实施方式354所述的基因组编辑系统,其中所述靶位点选自下组,该组由以下组成:HBG1 c.-114至-102(例如,SEQ ID NO:902(HBG1)的核苷酸2824-2836)、HBG1 c.-225至-222(例如,SEQ ID NO:902(HBG1)的核苷酸2716-2719)、和HBG2 c.-114至-102(例如,SEQ ID NO:903(HBG2)的核苷酸2748-2760)。

[1352] 实施方式356.如实施方式355所述的基因组编辑系统,其中所述靶位点是HBG1c.-114至-102(例如,SEQ ID NO:902(HBG1)的核苷酸2824-2836),并且所述5′同源臂包含约50bp至100bp,例如,55bp至95bp、60bp至90bp、70bp至90bp、或80bp至90bp,HBG1 c.-114至-102(例如,SEQ ID NO:902(HBG1)的核苷酸2824-2836)的同源性5′。

[1353] 实施方式357.如实施方式356所述的基因组编辑系统,其中所述5′同源臂包含SEQ ID NO:904(即,ssODN1 5'同源臂),基本上由其组成或由其组成。

[1354] 实施方式358.如实施方式180或181所述的基因组编辑系统,其中所述3'同源臂包含约50至100bp,例如,55至95、60至90、70至90、或80至90bp,HBG1 c.-114至-102(例如,SEQ ID NO:902(HBG1)的核苷酸2824-2836)的同源性3′。

[1355] 实施方式359.如实施方式355-358中任一项所述的基因组编辑系统,其中所述3′同源臂包含SEQ ID NO:905(即,ssODN1 3′同源臂),基本上由其组成或由其组成。

[1356] 实施方式360.如实施方式355所述的基因组编辑系统,其中所述靶位点是HBG2c.-114至-102(例如,SEQ ID NO:903(HBG2)的核苷酸2748-2760),并且所述5′同源臂包含约50bp至100bp,例如,55bp至95bp、60bp至90bp、70bp至90bp、或80bp至90bp,HBG2 c.-114至-102(例如,SEQ ID NO:903(HBG2)的核苷酸2748-2760)的同源性5'。

[1357] 实施方式361.如实施方式360所述的基因组编辑系统,其中所述5'同源臂包含SEQ ID NO:904(即,ssODN1 5'同源臂),基本上由其组成或由其组成。

[1358] 实施方式362.如实施方式360或361所述的基因组编辑系统,其中所述3′同源臂包含约50bp至100bp,例如,55bp至95bp、60bp至90bp、70bp至90bp、或80bp至90bp,HBG2 c.-114至-102(例如,SEQ ID NO:903(HBG2)的核苷酸2748-2760)的同源性3'。

[1359] 实施方式363.如实施方式356-362中任一项所述的基因组编辑系统,其中所述3'同源臂包含SEQ ID NO:905(即,ssODN1 3′同源臂),基本上由其组成或由其组成。

[1360] 实施方式364.如实施方式356-363中任一项所述的基因组编辑系统,其中所述模板核酸包含SEQ ID NO:906(即,ssODN1),基本上由其组成或由其组成。

[1361] 实施方式365.如实施方式355所述的基因组编辑系统,其中所述靶位点是HBG1c.-225至-222(例如,SEQ ID NO:902(HBG1)的核苷酸2716-2719),并且所述5′同源臂包含约50bp至100bp,例如,55bp至95bp、60bp至90bp、70bp至90bp、或80bp至90bp,HBG1 c.-225至-222(例如,SEQ ID NO:902(HBG1)的核苷酸2716-2719)的同源性5'。

[1362] 实施方式366.如实施方式365所述的基因组编辑系统,其中所述3'同源臂包含约50bp至100bp,例如,55bp至95bp、60bp至90bp、70bp至90bp、或80bp至90bp,HBG1 c.-225至-222(例如,SEQ ID NO:902(HBG1)的核苷酸2716-2719)的同源性3′。

[1363] 实施方式367.如实施方式351-366中任一项所述的基因组编辑系统,其中所述ssODN包含5'硫代磷酸酯修饰。

[1364] 实施方式368.如实施方式351-366中任一项所述的基因组编辑系统,其中所述ssODN包含3′硫代磷酸酯修饰。

[1365] 实施方式369.如实施方式351-366中任一项所述的基因组编辑系统,其中所述ssODN包含5'硫代磷酸酯修饰和3'硫代磷酸酯修饰。

[1366] 实例

[1367] 以下实例仅仅是说明性的,并不旨在以任何方式限制本发明的范围或内容。

[1368] 实例1:用于将13bp del c.-114至-102插入HBG1和HBG2调控区域的化脓链球菌gRNA的筛选

[1369] 如本文所阐述的设计的化脓链球菌gRNA靶向跨越并包括HBG1的c.-114至-102处的13个核苷酸的26nt片段(例如,SEQ ID NO:902(HBG1)的核苷酸2824-2836,导致HBG1

13bp del c.-114至-102的改变)和HBG2(例如,SEQ ID NO:903(HBG2)的核苷酸2748-2760,导致HBG1 13bp del c.-114至-102的改变)。在电脑模拟和分级后设计gRNA后,选择一部分gRNA并筛选人类K562细胞的活性和特异性。选择用于筛选的gRNA含有表8中阐述的靶向结构域序列。将编码U6启动子和每种化脓链球菌gRNA的DNA与编码化脓链球菌Cas9的质粒DNA共电穿孔(Amaxa核转染仪)到人类K562细胞中。实验条件通常与本领域已知的那些条件一致(例如,Gori 2016,其通过引用结合在此)。电穿孔后3天,从K562细胞中提取gDNA,然后从gDNA中PCR扩增HBG1和HBG2基因座。通过T7E1内切核酸酶测定分析在PCR产物中评估基因编辑。在筛选的10个sgRNA中,8个切割启动子序列中的HBG1和HBG2靶向区域(图10A)。

[1370] 然后通过DNA测序分析来分析用8种活性sgRNA靶向的K562细胞的HBG1和HBG2 PCR产物,并对检测到的插入和缺失进行打分。缺失被细分为HPFH位点的精确的13nt缺失、HPFH包容性的和近端小缺失(18-26nt)、HPFH靶位点12nt缺失(即,部分缺失)、跨越部分HPFH靶位点的>26nt缺失、和其他缺失,例如,邻近HPFH靶位点但在HPFH靶位点外的缺失。八个sgRNA中的七个HBG1的13nt靶向缺失(HPFH突变诱导)(图10B)。八个sgRNA中的至少五个也支持HBG2启动子区域中13nt(HPFH突变诱导)靶向缺失(图10C)。需注意,用HBG Sp34 sgRNA处理的细胞中HBG2的DNA序列结果不可用。这些数据表明Cas9和sgRNA支持精确诱导13bp del c.-114至-102HPFH突变。图10D-10F描绘了在HBG1中靶向序列中观察到的缺失类型的实例。在每个具体实例中使用的gRNA以黑色示出,并且在每个组的实例中未靶向的其他gRNA以白色示出。

[1371] 表8:选择用于在K562细胞中筛选的gRNA列表

[1372]

gRNA ID	靶向结构域序列(RNA)	靶向结构域序列(DNA)	靶向结构域序列加上3 NT化脓链球菌PAM序列(NGG)(RNA)	靶向结构域序列加上3 NT化脓链球菌PAM序列(NGG)(DNA)	正义	指导类别
Sp9	GGCUAUUGGUCAAGGCA(SEQ ID NO:277)	GGCTATTGGTCAAGGCA(SEQ ID NO: 910)	GGCUAUUGGUCAAGGCAAGG(SEQ ID NO:920)	GGCTATTGGTCAAGGCAAGG(SEQ ID NO:930)	反义	spy17
Sp36	CAAGGCUA	CAAGGCTA	CAAGGCUA	CAAGGCTA	反义	spy20

[1373]

	UUGGUCAAGGCA（SEQ ID NO: 338）	TTGGTCAAGGCA（SEQ ID NO: 911）	UUGGUCAAGGCAAGG（SEQ ID NO:921）	TTGGTCAAGGCAAGG（SEQ ID NO:931）		
Sp40	UGCCUUGUCAAGGCUAU（SEQ ID NO:327）	TGCCTTGTCAAGGCTAT（SEQ ID NO: 912）	UGCCUUGUCAAGGCUAUUGG（SEQ ID NO:922）	TGCCTTGTCAAGGCTATTGG（SEQ ID NO:932）	反义	spy17
Sp42	GUUUGCCUUGUCAAGGCUAU（SEQ ID NO: 299）	GTTTGCCTTGTCAAGGCTAT（SEQ ID NO: 913）	GUUUGCCUUGUCAAGGCUAUUGG（SEQ ID NO:923）	GTTTGCCTTGTCAAGGCTATTGG（SEQ ID NO:933）	反义	spy20
Sp38	GACCAAUAGCCUUGACA（SEQ ID NO:276）	GACCAATAGCCTTGACA（SEQ ID NO: 914）	GACCAAUAGCCUUGACAAGG（SEQ ID NO:924）	GACCAATAGCCTTGACAAGG（SEQ ID NO:934）	正义	spy17
Sp37	CUUGACCAAUAGCCUUGACA（SEQ ID NO: 333）	CTTGACCAATAGCCTTGACA（SEQ ID NO: 915）	CUUGACCAAUAGCCUUGACAAGG（SEQ ID NO:925）	CTTGACCAATAGCCTTGACAAGG（SEQ ID NO:935）	正义	spy20
Sp43	GUCAAGGCUAUUGGUCA（SEQ ID NO:278）	GTCAAGGCTATTGGTCA（SEQ ID NO:916）	GUCAAGGCUAUUGGUCAAGG（SEQ ID NO:926）	GTCAAGGCTATTGGTCAAGG（SEQ ID NO:936）	反义	spy17
Sp35	CUUGUCAAGGCUAUUGGUCA（SEQ ID NO: 339）	CTTGTCAAGGCTATTGGTCA（SEQ ID NO: 917）	CUUGUCAAGGCUAUUGGUCAAGG（SEQ ID NO:927）	CTTGTCAAGGCTATTGGTCAAGG（SEQ ID NO:937）	反义	spy20
Sp41	UCAAGUUUGCCUUGUCA（SEQ ID NO:310）	TCAAGTTTGCCTTGTCA（SEQ ID NO: 918）	UCAAGUUUGCCUUGUCAAGG（SEQ ID NO:928）	TCAAGTTTGCCTTGTCAAGG（SEQ ID NO:938）	反义	spy17
Sp34	UGGUCAAGUUUGCCUUGUCA（SEQ ID NO: 340）	TGGTCAAGTTTGCCTTGTCA（SEQ ID NO: 919）	UGGUCAAGUUUGCCUUGUCAAGG（SEQ ID NO:929）	TGGTCAAGTTTGCCTTGTCAAGG（SEQ ID NO:939）	反义	spy20

[1374] 实例2:含有靶向HPFH突变的gRNA的Cas9RNP支持人类造血干/祖细胞中的基因编辑

[1375] 用人类细胞因子(干细胞因子(SCF)、血小板生成素(TPO)、Flt3配体(FL))和小分子(前列腺素E2(PGE2)、StemRegenin 1(SR1))预刺激人类脐带血(CB)$CD34^+$细胞两天。实验条件通常根据提供于Gori 2016第240-241页中的方法,其通过引入结合在此。将CB $CD34^+$

细胞用含有(例如,5′ARCA加帽和3’polyA(20A)尾部)靶向HBG1和HBG2调控区域的sgRNA(表8)的化脓链球菌Cas9 RNP电穿孔(Amaxa核转染仪)。电穿孔后3天,从RNP处理的CB $CD34^+$细胞中提取gDNA,并通过T7E1测定和DNA测序分析基因编辑。

[1376] 在CB $CD34^+$细胞中测试的含有不同gRNA的RNP中,只有Sp37 gRNA(包含SEQ ID NO:333)在HBG1和HBG2启动子的靶位点处导致可检测的编辑,如通过HBG1和HBG2特异性PCR产物(从三个脐带血供体的电穿孔CB $CD34^+$细胞中提取的gDNA扩增产物)中的indel的T7E1分析所确定的(图11A)。在用与Sp37复合的Cas9蛋白电穿孔的细胞中检测到的平均编辑水平在HBG1处为5％±2％indel,在HBG2处检测为3％±1％indel(三个独立的实验和CB供体)。

[1377] 接下来,将三个化脓链球菌gRNA(其靶位点在HBG启动子内)(Sp35(包含SEQ ID NO:339)、Sp36(包含SEQ ID NO:338)、Sp37(包含SEQ ID NO:333))与野生型化脓链球菌Cas9蛋白复合以形成核糖核蛋白复合物。将这些HBG靶向的RNPS电穿孔到CB $CD34^+$细胞(n＝3供体)和成人动员的外周血(mPB)$CD34^+$细胞供体(n＝3供体)中。根据上述方法和Gori 2016第240-241页制备CB $CD34^+$细胞。除了不添加SR1之外,以与CB $CD34^+$细胞基本相同的方式制备成人mPB $CD34^+$细胞。在Cas9 RNP递送后大约三天,通过从样品中提取的基因组DNA扩增的HBG2PCR产物的T7E1内切核酸酶分析来分析靶位点处的插入/缺失水平。这些RNP中的每一个仅在三个供体和三个独立的实验的CB和成人$CD34^+$细胞中支持低水平基因编辑(图11B)。

[1378] 为了增加靶位点的基因编辑并增加靶位点13bp缺失的发生,单链脱氧核苷酸供体修复模板(ssODNs)在5′编码87bp和89bp的同源性并产生HBG1和HBG2的靶向缺失位点的3′侧。构建体ssODN1(SEQ ID NO:906,表9),包括5′和3′同源臂,被设计为“编码”13bp缺失,其中序列同源臂工程化侧接所述不存在的序列以产生完美缺失。5′同源臂(SEQ ID NO:904,表9)包括与HBG1和HBG2的c.-114至-102的序列5′同源的核苷酸(即,与SEQ ID NO:902(HBG1)的核苷酸2824-2836的序列5′同源的核苷酸,和与SEQ ID NO:903(HBG2)的核苷酸2748-2760的序列5′同源的核苷酸)。3′同源臂(SEQ ID NO:905,表9)包括与HBG1和HBG2的c.-114至-102的3′区域同源的核苷酸(即,与SEQ ID NO:902(HBG1)的核苷酸2824-2836的序列3′同源的核苷酸,和与SEQ ID NO:903(HBG2)的核苷酸2748-2760的序列3′同源的核苷酸)。在末端修饰ssODN1构建体以在5′和3′端含有硫代磷酸酯(PhTx)(SEQ ID NO:909,表9)以形成PhTx ssODN1。

[1379] 表9:单链脱氧核苷酸供体修复模板(ssODN)

[1380]

ssODN ID	SEQ ID NO	序列
ssODN1 5'同源臂	904	**GGGTGCTTCCTTTTATTCTTCATCCCTAGCCAGCCG CCGGCCCCTGGCCTCACTGGATACTCTAAGACTATT GGTCAAGTTTGCCTT**
ssODN1 3'同源臂	905	GTCAAGGCAAGGCTGGCCAACCCATGGGTGGAGTTTA GCCAGGGACCGTTTCAGACAGATATTTGCATTGAGATA GTGTGGGGAAGGGG
ssODN1	906	**GGGTGCTTCCTTTTATTCTTCATCCCTAGCCAGCCG CCGGCCCCTGGCCTCACTGGATACTCTAAGACTATT GGTCAAGTTTGCCTT**GTCAAGGCAAGGCTGGCCAAC CCATGGGTGGAGTTTAGCCAGGGACCGTTTCAGACAGA TATTTGCATTGAGATAGTGTGGGGAAGGGG
PhTx ssODN1 5'同源臂	907	***GGGTGCTTCCTTTTATTCTTCATCCCTAGCCAGCC GCCGGCCCCTGGCCTCACTGGATACTCTAAGACTAT TGGTCAAGTTTGCCTT**
PhTx ssODN1 3'同源臂	908	GTCAAGGCAAGGCTGGCCAACCCATGGGTGGAGTTTA GCCAGGGACCGTTTCAGACAGATATTTGCATTGAGATA GTGTGGGGAAGGGG*
PhTx ssODN1	909	***GGGTGCTTCCTTTTATTCTTCATCCCTAGCCAGCC GCCGGCCCCTGGCCTCACTGGATACTCTAAGACTAT TGGTCAAGTTTGCCTT**GTCAAGGCAAGGCTGGCCAA
		CCCATGGGTGGAGTTTAGCCAGGGACCGTTTCAGACA GATATTTGCATTGAGATAGTGTGGGGAAGGGG*

[1381]

侧翼缺失的同源臂用粗体[5'同源臂]和下划线[3'同源臂]指示。注意 ssODN1 和 PhTx ssODN1 中不存在 13 bp 序列。
*代表硫代磷酸酯的修饰。

[1382] 根据上述方法和Gori 2016第240-241页制备CB $CD34^+$细胞。ssODN(即,ssODN1和PhTx ssODN1)与靶向HBG的RNP共递送至CB $CD34^+$细胞,所述HBG含有Sp37gRNA(HBG Sp37 RNP)或HBG Sp35(HBG Sp35 RNP)。

[1383] 通过HBG2 PCR产物的T7E1分析确定,编码13bp缺失的ssODN1和PhTx ssODN1供体模板与含有Sp35 gRNA的RNP(即,HBG Sp35 RNP)或含有Sp37 gRNA的RNP(即,HBG Sp37 RNP)的共递送导致分别在靶位点的基因编辑中有6倍和5倍的增加(图11C)。HBG2 PCR产物的DNA测序分析(Sanger测序)指示了在用HBG Sp37 RNP和PhTx ssODN1处理的细胞中20%的基因编辑,具有15%的缺失和5%的插入(图11C,左下部组)。对靶位点缺失的特异性类型和尺寸的进一步分析揭示,检测到的总缺失的75%中有3/4含有HPFH 13bp缺失(其中包括缺失近端启动子中的CAAT盒),所述缺失与HbF表达的升高有关(图11C,右下部组)。剩余的1/4缺失是部分缺失,没有跨越完全的13bp缺失。这些数据指示,工程化具有缺失的同源ssODN的共递送支持在人类$CD34^+$细胞中HBG的精确基因编辑(缺失)。

[1384] 实例3:筛选作为核糖核蛋白复合物递送至K562细胞的化脓链球菌gRNA,用于引起13bpdelc.-114至-102进入HBG1和HBG2调控区域

[1385] 如实例1(图10)所述,通过将Cas9和gRNA DNA电穿孔进入K562细胞中筛选的指导RNA进行体外转录,然后与化脓链球菌Wt Cas9蛋白复合形成核糖核蛋白复合物(RNP)。为了比较这些RNP的活性与通过将Cas9和gRNA DNA递送至K562细胞(即,实例1)和通过RNP递送

至人类CD34$^+$细胞(即,实例2)所观察到的活性,此处通过电穿孔(Amaxa核转染仪)将RNP递送至K562细胞。与化脓链球菌Cas9蛋白复合的gRNA是修饰的gRNA((例如,5'ARCA加帽和3' polyA(20A)尾部;表8)并靶向HBG1和HBG2调控区域。

[1386] 电穿孔后3天,从K562细胞中提取gDNA,然后通过PCR扩增HBG1和HBG2启动子区域,随后对PCR产物进行T7E1分析。(图12A)。九个RNP中有八个支持高百分比的NHEJ。Sp37 RNP是唯一示出于人类CD34+细胞中有活性的gRNA(在CD34$^+$细胞中<10%编辑),在K562细胞中具有高活性,在HBG1和HBG2处均检测到>60%的indel(图12A)。靶向HPFH缺失突变位点Sp35的其他gRNA支持HBG1和HBG2的43%的编辑(图12A)。

[1387] 对来自用Cas9处理复合的细胞的gDNA的一部分PCR产物进行DNA测序分析,所述Cas9复合至最接近靶向HPFH位点的gRNA。对DNA序列进行打分以检测插入和缺失。缺失被细分为HPFH位点的精确的13nt缺失、HPFH包容性的和近端小缺失(18-26nt)、HPFH靶位点12nt缺失(即,部分缺失)、跨越部分HPFH靶位点的>26nt缺失、和其他缺失,例如,邻近HPFH靶位点但在HPFH靶位点外的缺失。在用与HBG1/HBG2的gRNA Sp35和37(HPFH突变诱导)(图12B)复合的RNP处理的细胞中检测到13nt缺失,这些数据指示Cas9和sgRNA(Sp35和Sp37)作为核糖核蛋白复合物递送至造血细胞引起c.-114至-102HPFH突变。

[1388] 实例4:靶向HPFH突变的Cas9 RNP支持人类成人动员的具有成红细胞后裔中HBG表达增加的外周血造血干细胞/祖细胞中的基因编辑

[1389] 为了确定在HBG启动子中编辑HBG与Cas9 RNP复合到Sp37 gRNA或Sp35gRNA(即,靶向与HPFH有关的13bp缺失的gRNA)的HBG支持编辑的CD34$^+$细胞的红系后裔中HBG表达的增加,用RNP电穿孔来自动员外周血(mPB)的人类成人CD34$^+$细胞。简而言之,在StemSpan无血清扩增培养基(SFEM)中用人类细胞因子和PGE2预刺激mPB CD34$^+$细胞2天,然后分别用预复合到Sp35和Sp37的Cas9蛋白电穿孔。参见Gori 2016。HBG PCR产物的T7E1分析指示,用与Sp37复合的RNP处理的mPB CD34$^+$细胞检测到约3%的indel,而未检测到用与Sp35复合的RNP处理的细胞的编辑(图13A)。

[1390] 为了增加靶位点的基因编辑并增加靶位点处13bp缺失的发生,将PhTx ssODN1与靶向含有Sp37 gRNA的HBG的预复合RNP共递送。编码13bp缺失的PhTx ssODN1供体的共递送导致靶位点的基因编辑增加近2倍(图13A)。

[1391] 为了确定编辑HBG是否增加编辑的成人CD34$^+$细胞的红系后裔中胎儿血红蛋白的产生,在人类细胞因子(促红细胞生成素,SCF,IL3)、人类血浆(Octoplas)、和其他补充剂(氢化可的松、肝素、转铁蛋白)存在下,通过培养持续长达18天将细胞分化成成红细胞。在分化的时间过程中,收集mRNA以评估RNP处理的mPB CD34$^+$细胞和供体匹配的阴性(未处理的)对照的红系后裔中的HBG基因表达。到分化的第7天,用HBG Sp37 RNP处理的人类CD34$^+$细胞的成红细胞后裔和编码ssODN1的13bp HPFH缺失(通过T7E1分析在来自大量细胞群的gDNA中检测到约5%的indel)展现出HBG mRNA的产生增加2倍(图13B)。此外,通过用于获得红系表型(%血型糖蛋白A$^+$细胞)的流式分析确定,从RNP处理的CD34+细胞分化的成红细胞维持观察到供体匹配的未处理对照细胞的分化动力学(图14A)。重要的是,用HBG Sp37RNP和ssODN1电穿孔的CD34$^+$细胞维持其离体造血活性(即,与未处理的供体匹配的CD34$^+$细胞阴性对照相比,红系和骨髓集落的量或多样性没有差异),如在造血集落形成细胞(CFC)测定中确定的(图14B)。这些数据指示HBG1/HBG2近端启动子区域的靶向破坏支持RNP处理的成

人造血干细胞/祖细胞的红系后裔中HBG表达的增加而不改变分化潜能。

[1392] 序列

[1393] 根据本披露内容的基因组编辑系统组件(包括但不限于,RNA指导的核酸酶、指导RNA、供体模板核酸、编码核酸酶或指导RNA的核酸、以及任何前述的部分或片段),用序列表中表示的核苷酸和氨基酸序列例示。序列表中表示的序列不旨在是限制性的,而是说明性的基因组编辑系统及其组分部分的某些原理,组合本披露内容,将通知本领域技术人员另外的本披露内容范围内的实施和修饰。表示序列的列表在以下表10中提供。

[1394] 表10:序列表中表示的序列:

[1395]

SEQ ID NO:	描述
1-2、4-6、12、14	Cas9 多肽
3、7-11、13	Cas9 编码序列
15-23、52-123	Cas9 RuvC 样结构域
24-28、124-198	Cas9 HNH 样结构域
29-31、38-51	全长模块化和单分子 gRNA
32-37	gRNA 近端和尾部结构域
199-205	PAM 序列
251-901	gRNA 靶向结构域(RNA)- 参见表 6-8
910-919	gRNA 靶向结构域(DNA)- 参见表 8
920-929	gRNA 靶向结构域加上 PAM (NGG) (RNA) - 参见表 8
930-939	gRNA 靶向结构域加上 PAM (NGG) (DNA) - 参见表 8
902、903	人类 HBG1、HBG2 启动子序列包括 HPFH 缺失位点
904-909	寡核苷酸供体序列和同源臂 - 参见表 9

[1396] 通过引用结合

[1397] 本文提及的所有出版物、专利和专利申请都通过引用以其全文而特此结合,如同每一单独的出版物、专利或专利申请具体且单独地指明通过引用而结合一样。在有冲突的情况下,以本申请(包括本文的任何定义)为准。

[1398] 等效物

[1399] 本领域的普通技术人员仅使用常规实验就应认识到或能够确定本文描述的本发

明的具体实施例的许多等效物。此类等效物旨在由以下权利要求书涵盖。

[1400] 参考文献

[1401] Ahern等人,Br J Haematol[英国血液学杂志]25(4):437-444(1973)

[1402] Akinbami,Hemoglobin[血红蛋白]40:64-65(2016)

[1403] Aliyu等人,Am J Hematol[美国血液学杂志]83:63-70(2008)

[1404] Anders等人,Nature[自然]513(7519):569-573(2014)

[1405] Angastiniotis和Modell,Ann N Y Acad Sci[纽约科学院年报]850:251-269(1998)

[1406] Bae等人,Bioinformatics[生物信息学]30(10):1473-1475(2014)

[1407] Barbosa等人,Braz J Med Bio Res[巴西医学和生物研究杂志]43(8):705-711(2010)

[1408] Bouva,Hematologica[血液学]91(1):129-132(2006)

[1409] Brousseau,Am J Hematol[美国血液学杂志]85(1):77-78(2010)

[1410] Caldecott,Nat Rev Genet[自然遗传学评论]9(8):619-631(2008)

[1411] Chassanidis,Ann Hematol[血液学年鉴]88(6):549-555(2009)

[1412] Chylinski等人,RNA Biol[RNA生物学]10(5):726-737(2013)

[1413] Cong等人,Science[科学]399(6121):819-823(2013)

[1414] Costa等人,Cad Saude Publica 18(5):1469-1471(2002)

[1415] Cotta-Ramusino等人,国际专利公开号WO 2016/073990(2016)

[1416] Fine等人,Sci Rep.[科学报告]5:10777(2015)

[1417] Friedland等人,Genome Biol[基因组生物学]16:257(2015)

[1418] Fu等人,Nat Biotechnol[自然生物技术]32:279-284(2014)

[1419] Gori等人,国际专利公开号WO 2016/182959 A1(2016)

[1420] Guilinger等人,Nat Biotechnol[自然生物技术]32:577-582(2014)

[1421] Jinek等人,Science[科学]337(6096):816-821(2012)

[1422] Jinek等人,Science[科学]343(6176):1247997(2014)

[1423] Kleinstiver等人,Nature[自然]523(7561):481-485(2015a)

[1424] Kleinstiver等人,Nat Biotechnol[自然生物技术]33(12):1293-1298(2015b)

[1425] Kleinstiver等人,Nature[自然]529(7587):490-495(2016)

[1426] Lee等人,Nano Lett[纳米快报]12(12):6322-6327(2012)

[1427] Lewis,"Medical-Surgical Nursing:Assessment and Management of Clinical Problems"[医学外科护理:临床问题的评估与管理](2014)

[1428] Li,Cell Res[细胞研究]18(1):85-98(2008)

[1429] Maeder等人,国际专利公开号WO 2015/138510(2015)

[1430] Mali等人,Science[科学]339(6121):823-826(2013)

[1431] Mantovani等人,Nucleic Acids Res[核酸研究]16(16):7783-7797(1988)

[1432] Marteijn等人,Nat Rev Mol Cell Biol[自然分子细胞生物学评论]15(7):465-481(2014)

[1433] Nishimasu等人,Cell[细胞]156(5):935-949(2014)

[1434] Ran等人,Cell[细胞]154(6):1380-1389(2013)
[1435] Shmakov等人,Molecular Cell[分子细胞]60(3):385-397(2015)
[1436] Sternberg等人,Nature[自然]507(7490):62-67(2014)
[1437] Superti-Furga等人,EMBO J[欧洲分子生物学杂志]7(10):3099-3107(1988)
[1438] Thein,Hum Mol Genet[人类分子遗传学]18(R2):R216-223(2009)
[1439] Waber等人,Blood[血液]67(2):551-554(1986)
[1440] Wang等人,Cell[细胞]153(4):910-918(2013)
[1441] Xu等人,Genes Dev[基因与发育]24(8):783-798(2010)
[1442] Yamano等人,Cell[细胞]165(4):949-962(2016)
[1443] Zetsche等人,Nat Biotechnol[自然生物技术]33(2):139-42(2015)

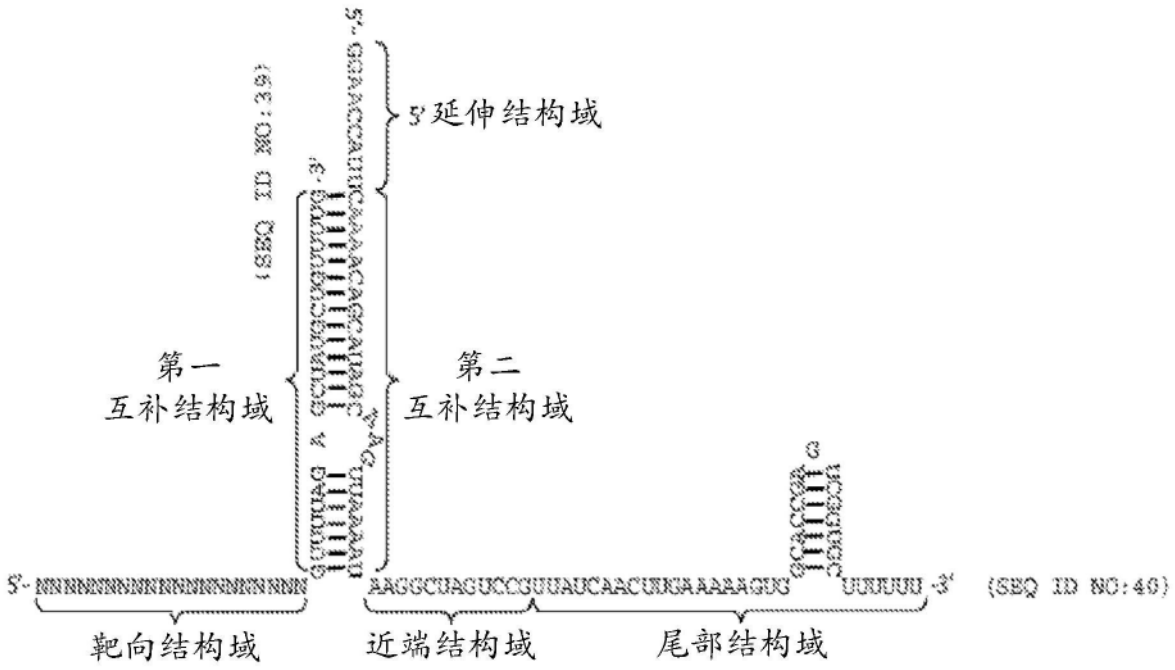

5'延伸结构域
第一
互补结构域
第二
互补结构域
(SEQ ID NO:39)
(SEQ ID NO:40)
靶向结构域
近端结构域
尾部结构域

图1A

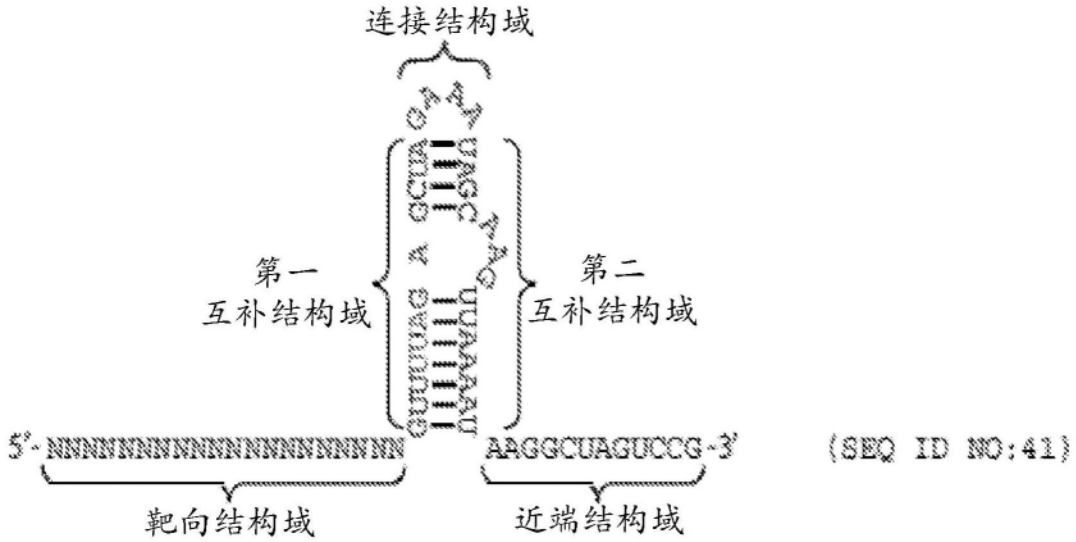

连接结构域
第一
互补结构域
第二
互补结构域
(SEQ ID NO:41)
靶向结构域
近端结构域

图1B

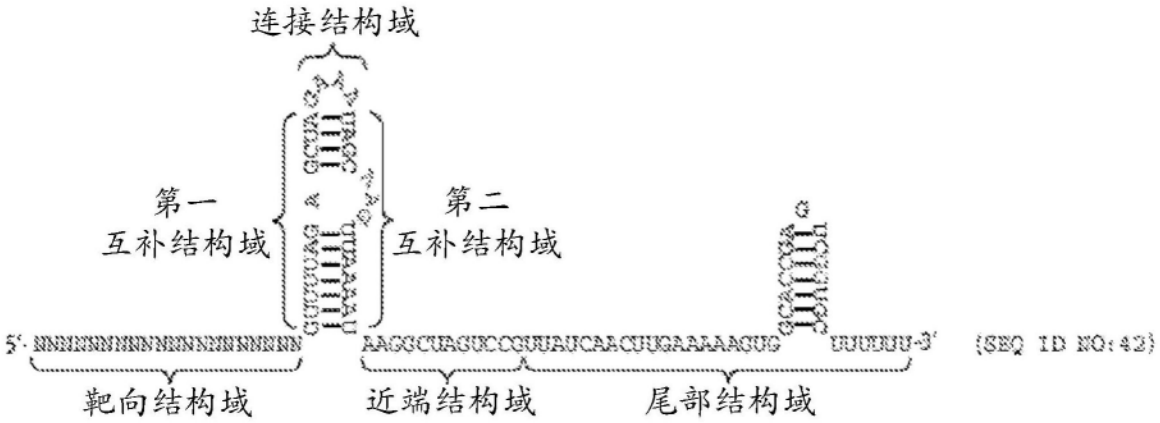

连接结构域
第一
互补结构域
第二
互补结构域
5'-NNNNNNNNNNNNNNNNNNNN
AAGGCUAGUCCGUUAUCAACUUGAAAAAGUG
UUUUUU-3'
(SEQ ID NO:42)
靶向结构域
近端结构域
尾部结构域

图1C

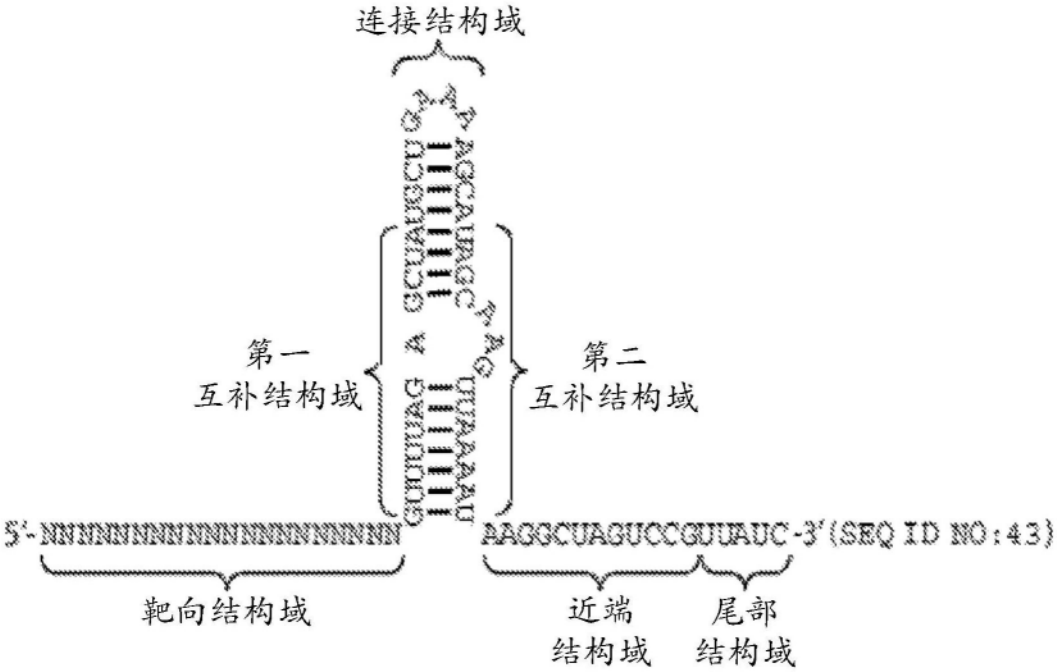

连接结构域
第一
互补结构域
第二
互补结构域
5'-NNNNNNNNNNNNNNNNNNNN
AAGGCUAGUCCGUUAUC-3'(SEQ ID NO:43)
靶向结构域
近端
结构域
尾部
结构域

图1D

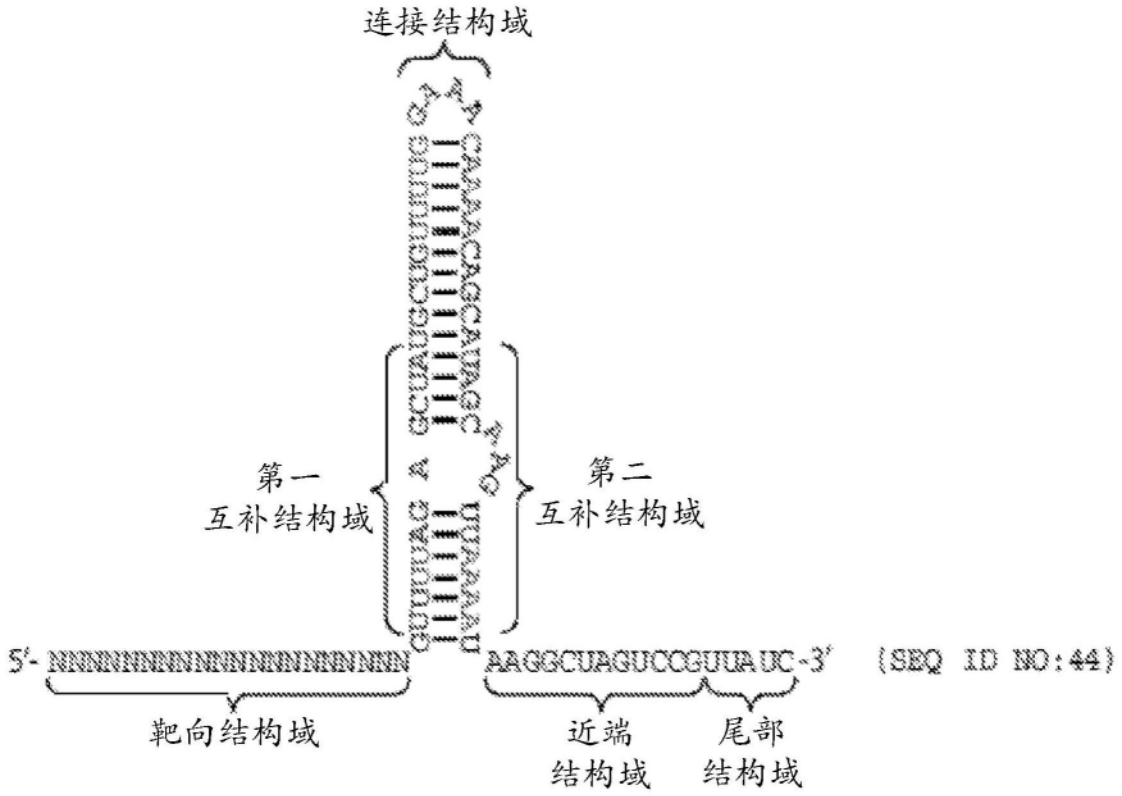

连接结构域
第一
互补结构域
第二
互补结构域
5'- NNNNNNNNNNNNNNNNNNNN AAGGCUAGUCCGUUAUC-3' (SEQ ID NO:44)
靶向结构域
近端
结构域
尾部
结构域

图1E

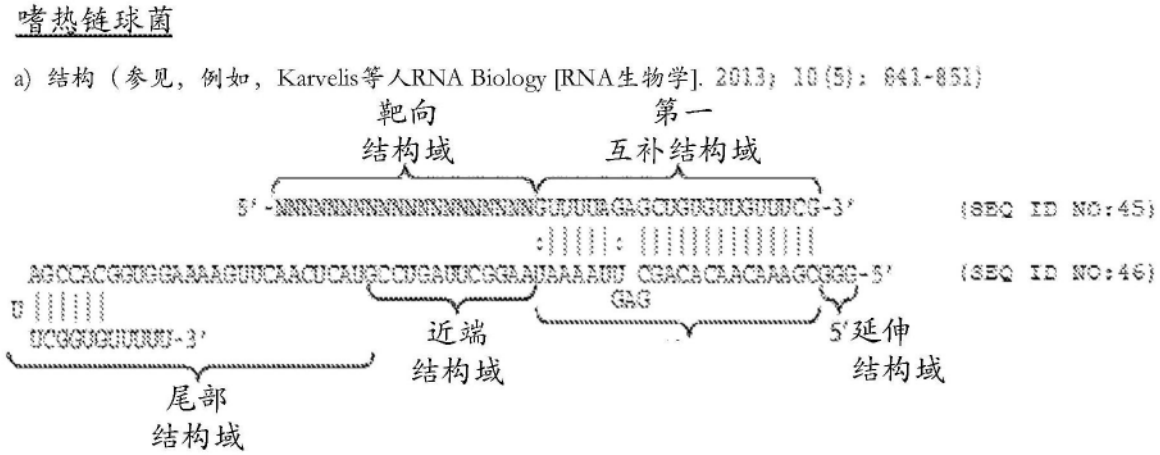

嗜热链球菌
a) 结构（参见，例如，Karvelis等人RNA Biology [RNA生物学]. 2013; 10(5): 841-851)
靶向
结构域
第一
互补结构域
(SEQ ID NO:45)
(SEQ ID NO:46)
近端
结构域
5'延伸
结构域
尾部
结构域

图1F

比对

嗜热链球菌 5'-NNNNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAGCUGUUUG-3' (SEQ ID NO:39)
化脓链球菌 5'-NNNNNNNNNNNNNNNNNNNNGUUUUAGAGCUGUGUUGUUUCG-3' (SEQ ID NO:45)
********** ** ***** *

嗜热链球菌 5'-GAACCAUCAAAACAGCAUAGCAAGUUAAAAUAAGGC-UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU-3'
化脓链球菌 5'------GCCCGAAACAACACAGCCAGUUAAAAUAAGGCUUAGUCCGUACUCAACUUGAAAAGCUGGCACCGAUUCCGUGUUUUU----3'
* ***** ** *** ********** ******** ********** ********** ****** ****

嗜热链球菌 - cont (SEQ ID NO:47)
化脓链球菌 - cont (SEQ ID NO:46)

图1G

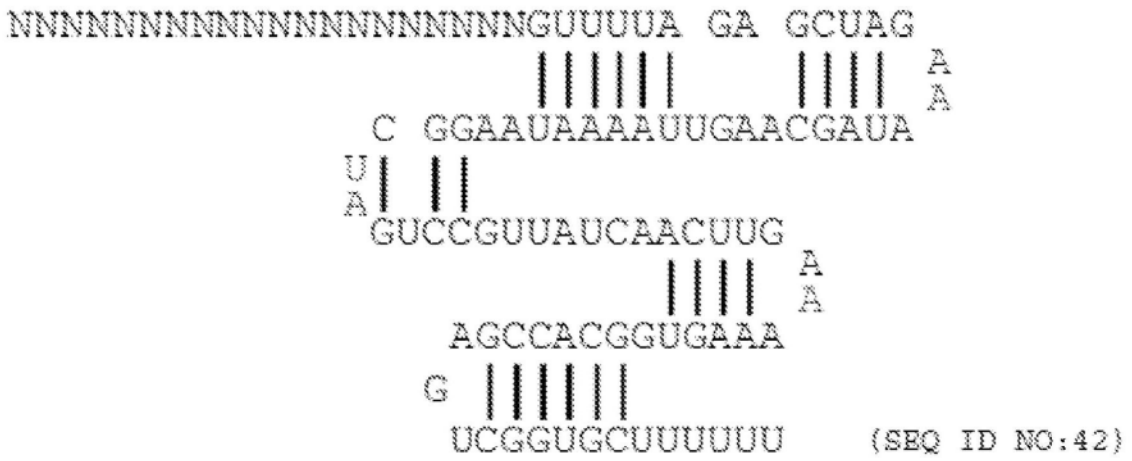


图1H

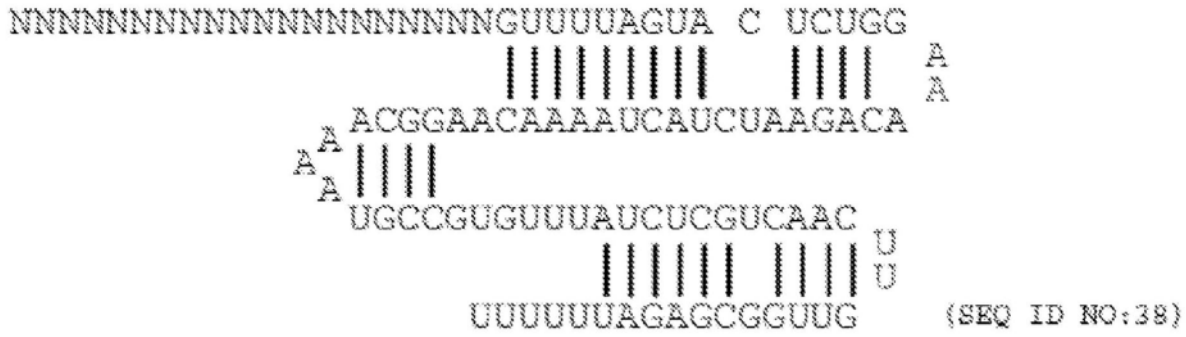


图1I

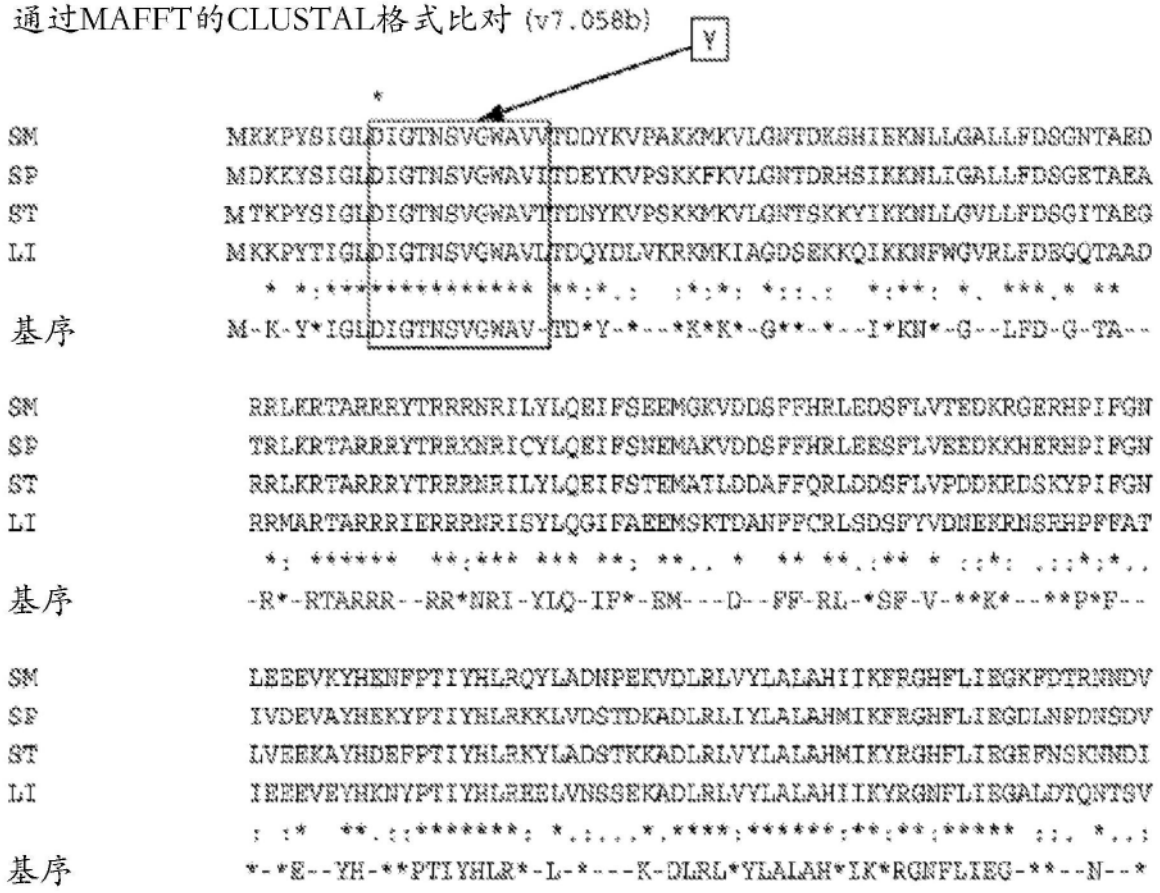

通过MAFFT的CLUSTAL格式比对 (v7.058b)
Y
*
SM MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIEKNLLGALLFDSGNTAED
SP MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEA
ST MTKPYSIGLDIGTNSVGWAVTTDNYKVPSKKMKVLGNTSKKYIKKNLLGVLLFDSGITAEG
LI MKKPYTIGLDIGTNSVGWAVLTDQYDLVKRKMKIAGDSEKKQIKKNFWGVRLFDEGQTAAD
基序 M-K-Y*IGLDIGTNSVGWAV-TD*Y-*--*K*K*-G**-*--I*KN*-G--LFD-G-TA--
SM RRLKRTARRRYTRRRNRILYLQEIFSEEMGKVDDSFFHRLEDSFLVTEDKRGERHPIFGN
SP TRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGN
ST RRLKRTARRRYTRRRNRILYLQEIFSTEMATLDDAFFQRLDDSFLVPDDKRDSKYPIFGN
LI RRMARTARRRIERRRNRISYLQGIFAEEMSKTDANFFCRLSDSFYVDNEKRNSRHPFFAT
基序 -R*-RTARRR--RR*NRI-YLQ-IF*-EM---D--FF-RL-*SF-V-**K*--**P*F--
SM LEEEVKYHENFPTIYHLRQYLADNPEKVDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDV
SP IVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDV
ST LVEEKAYHDEFPTIYHLRKYLADSTKKADLRLVYLALAHMIKYRGHFLIEGEFNSKNNDI
LI IEEEVEYHKNYPTIYHLREELVNSSEKADLRLVYLALAHIIKYRGNFLIEGALDTQNTSV
基序 *-*E--YH-**PTIYHLR*-L-*---K-DLRL*YLALAH*IK*RGNFLIEG-**--N--*

图2A

```
SM      QRLFQEFLAVYDNTFENSS------LQEQNVQVEEILTDKISKSAKKDEVLKLFPNEKSN
SP      DKLFIQLVQTYNQLFEENP------INASGVDAKAILSARLSKSRRLENLIAQLPGEKKN
ST      QKNFQDFLDTYNAIFESDL-------SLENSKQLEEIVKDKISKLEKKDRILKLFPGEENS
LI      DGIYKQFIQTYNQVFASGIEDGSLKKLEDNKDVAKILVEKVTRKEKLERILKLYPGEKSA
        :  : ::: .*:  * ..          .. :   *:  ::::  : :.::   *.**.
基序    *--*-***--Y*--f-----------------*---I*--****--*-*-**---P-EK--

SM      GRFAEFLKLIVGNQADFKKHFELEEKAPLQFSKDTYEEELEVLLAQIGDNYAELFLSAKK
SP      GLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKN
ST      GIFSEFLKLIVGNQADFRKCFNLDEKASLHFSKESYDEDLETLLGYIGDDYSDVFLKAKK
LI      GMFAQFISLIVGSKGNFQKPFDLIEKSDIECAKDSYEEDLESLLALIGDEYAELFVAAKN
        * *,::: * :*   :*:. *:* *.: :. :*::*:::*: **. ***:*:::*: **:
基序    G-F-***-L-*G---*F*--F*L-E-*-*--*K**Y***L*-LL--IGD*Y***F*-AK*

SM      LYDSILLSGILTVTDVGTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSDVS
SP      LSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS
ST      LYDAILLSGFLTVTDNETEAPLSSAMIKRYNEHKEDLALLKEYIRNISLKTYNEVFKDDT
LI      AYSAVVLSSIITVAETETNAKLSASMIERFDTHEEDLGELKAFIKLHLPKHYEEIFSNTE
          .:::**.:: *    *:* **::**:*:: *. **  **  ::    . *:*:* :
基序    ---***LS-**-V----T*A-LS**MI*R**-H--DL--LK--**-------Y*E*F-*--

SM      KDGYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIEREDFLRKQRTFDNGSIPHQIH
SP      KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIH
ST      KNGYAGYIDGKTNQEDFYVYLKKLLAEFEGADYFLEKIDREDFLRKQRTFDNGSIPYQIH
LI      KHGYAGYIDGKTKQADFYKYMKMTLENIEGADYFIAKIEKENFLRKQRTFDNGAIPHQLH
        *.******** :.*  ** ::*  * :::*:  :: *:::*::**********;**;*;*
基序    K-GYAGYIDG-*-Q--FY-**K--L-***G*---**-K***E**LRKQRTFDNG*IP*Q*H
```

图2B

```
SM          LQEMRAIIRRQAEFYPFLADNQDRIEKLLTFRIPYYVGPLARGKSDFAWLSRKSADKITP
SP          LGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP
ST          LQEMRAILDKQAKFYPFLAKNKERIEKILTFRIPYYVGPLARGNSDFAWSIRKRNEKITP
LI          LEELEAILHQQAKYYPFLKENYDKIKSLVTFRIPYFVGPLANGQSEFAWLTRKADGRIRP
            * *;.**: :* .:**** .* ::*:.::******:*****.*;* ***  **    * *
基序        L-E*-AI*-*Q--*YPFL--N-**I*-**TFRIPY*VGPLA-G*S-FAW--RK----I-P

SM          WNFDRIVDKESSAEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTE-QG
SP          WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMR
ST          WNFEDVIDKESSAEAFINRMTSFDLYLPEEKVLPKHSLLYETFNVYNELTKVRFIAESMR
LI          WNIEEKVDFGKSAVDFIEKMTNKDTYLPKENVLPKHSLCYQKYLVYNELTKVRYIND-QG
            **::: :*   **  **;:**. *  **:::******* *: : ********::  :
基序        WN***-*D---SA--FI**MT--D--LP***VLPKHSL-Y*-*-VYNELTKV**--*---

SM          KTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKVFNASYG
SP          KPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDR---FNASLG
ST          DYQFLDSKQKKDIVRLYFKDKRKVTDKDIIEYL-HAIYGYDGIELKGIEKQ---FNSSLS
LI          KTSYFSGQEKEQIFNDLFKQKRKVKKKDLELFL-RNMSHVESPTIEGLEDS---FNSSYS
            .  ;;,.: *: *.   **  ***. ..:     : :       ; *::.    **:* .
基序        ---**--*-K*-I----FK--RKV-----*------*-*--------*-G**------FN*S--

SM          TYHDLCKIL-DKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLENYSDLLTKEQVKKLE
SP          TYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLK
ST          TYHDLLNIINDKEFLDDSSNEAIIEEIIHTLTIFEDREMIKQRLSKFENIFDKSVLKKLS
LI          TYHDLLKVGIKQEILDNFVNTEMLENIVKILTVFEDKRMIKEQLQQFSDVLDGVVLKKLE
            ***** ::  .:::**;  *  ::*:*:  **;***;.**.::*. : .::    ;*;*.
基序        TYHDL-**---***LD*--N--**E*I*--LT*FED*-MI-**L--*--**----*K*L-
```

图2C

图2D

图2E

```
SM      GNSDK-LIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIM
SP      RNSDE-LIARKKD---WDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIM
ST      PNSNENLVGAKEY---LDPKKYGGYAGISNSFTVLVKGTIEKGAKKKITNVLEFQGISIL
LI      GNSSK-LIPRKTN---WDPMKYGGLDSPNMAYAVVI--EYAKGKN-KLVFEKKIIRVTIM
         **.: *:  *      *. ****  .   ::::::     ** . *:     :  ::*:
基序    -NS-*-L*--K------D--KYGG------******------KG---K*------*--**I*

SM      EKMTFEERDPVAFLERKGYRNVQEENIIKLPKYSLFKLENGRKRLLAS-------ARELQK
SP      ERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS-------AGELQK
ST      DRINYRKDKLNFLLEKGYKDI--ELIIELPKYSLFELSDGSRRMLASILSTNNKRGEIHK
LI      ERKAFEKDEKAFLEEQGYRQP--KVLAKLPKYTLYECEEGRRRMLAS-------ANEAQK
        ::  :.::   **  :**::   . : :****:*:: .:* :*:***         * :*
基序    **--*-**---FL--*GY**------*-*LPKY*L**--*G-*R*LAS---------E-*K

SM      GNETVLPNHLGTLLYHAKNIHKV------DEPKHLDYVDKHKDEFKELLDVVSNFSKKYT
SP      GNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL-FVEQHKHYLDELIEQISEFSKRVI
ST      GNQIFLSQKFVKLLYHAKRISNT------INENHRKYVENHKKEFEELFYYILEFNENYV
LI      GNQQVLPNHLVTLLHHAANCEVS-------DGKSLDYIESNREMFAELLAHVSEFAKRYT
        **:  *..:  .:*: * .           : :   :::.::. : *::  : :* :.
基序    GN*--L--*---*L*-A-------------*-*---***-**--*-E**--*-*F-*---

SM      LAEGNLEKIKELYAQNNGEDLKELASSFI--------NLLTFTAIGAPATFKFFDKNIDR
SP      LADANLDKVLSAYNKHRDKPIREQAENII--------NLFTLTNLGAPAAFKYFDTTIDR
ST      GAKKNGKLLNSAFQSWQNHSIDELCSSFIGPTGSERKGLFELTSRGSAADFEFLGVKIFR
LI      LAEANLNKINQLFEQNKEGDIKAIAQSFV--------DLMAFNAMGAPASFKFFETTIER
         *. * . : . : . .   :   ...::         *: :.  *:.* *:::  .* *
基序    -A--N---*---*-------*------**---------L*-*---G*-A-F***---I-R
```

图2F

```
SM      KR-YTSTTEILNATLIHQSITGLYETRIDLNKLGGD   (SEQ ID NO:1)
SP      KR-YTSTKEVLDATLIHQSITGLYETRIDLSQLGGD   (SEQ ID NO:2)
ST      YRDYTPSSLLKDATLIHQSVTGLYETRIDLAKLGEG   (SEQ ID NO:4)
LI      KR-YNNLKELLNSTIIYQSITGLYESRKRLD----D   (SEQ ID NO:5)
         * *.  . : ::*:*:**:*****:*  *     .
基序    -R-Y-----*-**T*I*QS*TGLYE*R--L------   (SEQ ID NO:14)
```

图2G

披露于Chylinski等人N端RuvC样结构域的比对（排除序列异常值）
(通过MAFFT的CLUSTAL格式比对 (v7.058b))

1,12	DIGTNSVGWAVT	(SEQ ID NO:120)
3,20	DVGTNSVGWAVT	(SEQ ID NO:121)
15	DMGTNSVGWAVT	(SEQ ID NO:122)
4	DVGTSSVGWAVT	(SEQ ID NO:123)
7	DIGTASVGWAVT	(SEQ ID NO:52)
6	DVGTGSVGWAVT	(SEQ ID NO:53)
9	DIGTNSVGWAVV	(SEQ ID NO:54)
10	DIGTNSVGWAVI	(SEQ ID NO:55)
11	DIGTNSVGWAVL	(SEQ ID NO:56)
42	DLGTNSIGWAVV	(SEQ ID NO:57)
48	DLGTNSIGWAI-	(SEQ ID NO:58)
43	DLGTNSIGWALV	(SEQ ID NO:59)
2	DIGTNSVGWCVT	(SEQ ID NO:60)
14	DIGTNSVGYAVT	(SEQ ID NO:61)
5	DMGTGSLGWAVT	(SEQ ID NO:62)
16	DIGTSSVGWAAI	(SEQ ID NO:63)
8	DLGTGSVGWAVV	(SEQ ID NO:64)
22	DLGVGSVGWAIV	(SEQ ID NO:65)
23	DLGIASIGWAII	(SEQ ID NO:66)
24	DLGIASVGWAIV	(SEQ ID NO:67)
25	DLGVASVGWSIV	(SEQ ID NO:68)
26	DIGIASVGWAIL	(SEQ ID NO:69)
28	DLGISSVGWSVI	(SEQ ID NO:70)
32	DIGIASVGWSVI	(SEQ ID NO:71)
33	DVGIGSIGWAVI	(SEQ ID NO:72)
39	DLGVGSIGFAIV	(SEQ ID NO:73)
34	DIGYASIGWAVI	(SEQ ID NO:74)
47	DTGTNSLGWAIV	(SEQ ID NO:75)
50	DLGTNSIGWCLL	(SEQ ID NO:76)
49	DIGTDSLGWAVF	(SEQ ID NO:77)
18	DIGSNSIGFAVV	(SEQ ID NO:78)
41	DLGVGSIGVAVA	(SEQ ID NO:79)
45	DLGIASCGWGVV	(SEQ ID NO:80)

图3A

21	DLGIASVGWCLT	(SEQ ID NO:81)
27	DIGIGSVGVGIL	(SEQ ID NO:82)
29	DIGITSVGYGLI	(SEQ ID NO:83)
30	DIGITSVGFGII	(SEQ ID NO:84)
31	DVGITSTGYAVL	(SEQ ID NO:85)
40	DLGITSFGYAIL	(SEQ ID NO:86)
17	DIGNASVGWSAF	(SEQ ID NO:87)
19	DVGTNSCGWVAM	(SEQ ID NO:88)
35	DVGERSIGLAAV	(SEQ ID NO:89)
36	DVGLNSVGLAAV	(SEQ ID NO:90)
37	DVGLMSVGLAAI	(SEQ ID NO:91)
38	DVGTFSVGLAAI	(SEQ ID NO:92)
13	DIGTGSVGYACM	(SEQ ID NO:92)
44	DLGTTSIGFAHI	(SEQ ID NO:94)
46	DLGTNSIGSSVR	(SEQ ID NO:95)
	* * * *	

图3B

披露于Chylinski等人N端RuvC样结构域的比对
(通过MAFFT的CLUSTAL格式比对(v7.058b))

1,12	D----IGTNSVGWAVT	(SEQ ID NO:120)
3,20	D----VGTNSVGWAVT	(SEQ ID NO:121)
15	D----MGTNSVGWAVT	(SEQ ID NO:122)
4	D----VGTSSVGWAVT	(SEQ ID NO:123)
7	D----IGTASVGWAVT	(SEQ ID NO:52)
6	D----VGTGSVGWAVT	(SEQ ID NO:53)
9	D----IGTNSVGWAVV	(SEQ ID NO:54)
10	D----IGTNSVGWAVI	(SEQ ID NO:55)
53	D----IGTNSIGWAVI	(SEQ ID NO:96)
11	D----IGTNSVGWAVL	(SEQ ID NO:56)
42	D----LGTNSIGWAVV	(SEQ ID NO:57)
48	D----LGTNSIGWAI-	(SEQ ID NO:58)
43	D----LGTNSIGWALV	(SEQ ID NO:59)
2	D----IGTNSVGWCVT	(SEQ ID NO:60)
14	D----IGTNSVGYAVT	(SEQ ID NO:61)
5	D----MGTGSLGWAVT	(SEQ ID NO:62)
16	D----IGTSSVGWAAI	(SEQ ID NO:63)
8	D----LGTGSVGWAVV	(SEQ ID NO:64)
22	D----LGVGSVGWAIV	(SEQ ID NO:65)
23	D----LGIASIGWAII	(SEQ ID NO:66)
24	D----LGIASVGWAIV	(SEQ ID NO:67)
58	D----LGIASVGWAVV	(SEQ ID NO:97)
25	D----LGVASVGWSIV	(SEQ ID NO:68)
26	D----IGIASVGWAIL	(SEQ ID NO:69)
66	D----IGIASVGWAVL	(SEQ ID NO:98)
59	D----IGIASIGWAVI	(SEQ ID NO:99)
61	D----IGIASVGWAII	(SEQ ID NO:100)
64	D----VGIASVGWAVI	(SEQ ID NO:101)
62	D----IGIASVGWAL-	(SEQ ID NO:102)
67	D----IGIASVGWAMV	(SEQ ID NO:103)
32	D----IGIASVGWSVI	(SEQ ID NO:71)
28	D----LGISSVGWSVI	(SEQ ID NO:70)
63	D----IGITSVGWAVI	(SEQ ID NO:104)

图4A

33	D----VGIGSIGWAVI	(SEQ ID NO:72)
57	D----LGISSLGWAIV	(SEQ ID NO:105)
39	D----LGVGSIGFAIV	(SEQ ID NO:73)
34	D----IGYASIGWAVI	(SEQ ID NO:74)
50	D----LGTNSIGWCLL	(SEQ ID NO:76)
54	D----LGTNSIGWGLL	(SEQ ID NO:106)
47	D----TGTNSLGWAIV	(SEQ ID NO:75)
49	D----IGTDSLGWAVF	(SEQ ID NO:77)
51	D----LGSTSLGWAIF	(SEQ ID NO:107)
58	D----IGISSIGWAFS	(SEQ ID NO:108)
21	D----LGIASVGWCLT	(SEQ ID NO:81)
45	D----LGIASCGWGVV	(SEQ ID NO:80)
18	D----IGSNSIGFAVV	(SEQ ID NO:78)
65	D----IGTTSIGFSVI	(SEQ ID NO:109)
29	D----IGITSVGYGLI	(SEQ ID NO:83)
30	D----IGITSVGFGII	(SEQ ID NO:84)
44	D----LGTTSIGFAHI	(SEQ ID NO:94)
27	D----IGIGSVGVGIL	(SEQ ID NO:82)
41	D----LGVGSIGVAVA	(SEQ ID NO:79)
31	D----VGITSTGYAVL	(SEQ ID NO:85)
40	D----LGITSFGYAIL	(SEQ ID NO:86)
53	D----IGTSSIGWWLY	(SEQ ID NO:110)
55	D----LGSNSLGWFVT	(SEQ ID NO:111)
56	D----LGANSLGWFVV	(SEQ ID NO:112)
17	D----IGNASVGWSAF	(SEQ ID NO:87)
19	D----VGTNSCGWVAM	(SEQ ID NO:88)
35	D----VGERSIGLAAV	(SEQ ID NO:89)
36	D----VGLNSVGLAAV	(SEQ ID NO:90)
37	D----VGLMSVGLAAI	(SEQ ID NO:91)
38	D----VGTFSVGLAAI	(SEQ ID NO:92)
13	D----IGTGSVGYACM	(SEQ ID NO:93)
46	D----LGTNSIGSSVR	(SEQ ID NO:95)
60	DIGLRIGITSCGWSI-	(SEQ ID NO:113)
69	D----MGAKYTGVFYA	(SEQ ID NO:114)
73	D----LGGKNTGFFSF	(SEQ ID NO:115)
74	D----LGVKNTGVFSA	(SEQ ID NO:116)
70	D----LGAKFTGVALY	(SEQ ID NO:117)
71	D----LGGKFTGVCLS	(SEQ ID NO:118)
72	D----LGGTYTGTFIT	(SEQ ID NO:119)

图4B

披露于Chylinski等人HNH样结构域的比对
(通过MAFFT的CLUSTAL格式比对 (v7.058b))

```
1          YDIDHIYPRS-LTKD------DSF-DNLVLCERTAN    (SEQ ID NO:196)
2          -DIDHIYPRSKVIKD------DSF-DNLVLVLKNEN    (SEQ ID NO:197)
3          -DRDHIYPQS-KIKD------DSI-DNLVLVNKTYN    (SEQ ID NO:198)
4          -DIDHIYPRS-KIKD------DSI-TNRVLVEKDIN    (SEQ ID NO:195)
6          -DIDHIYPQS-KIKD------DSI-SNRVLVCSSCN    (SEQ ID NO:124)
5          -DIDHIYPQS-KTMD------DSL-NNRVLVKKNYN    (SEQ ID NO:125)
7          -DQDHIYPKS-KIYD------DSL-ENRVLVKKNLN    (SEQ ID NO:126)
8          -QIDHIVPQS-LVKD------DSF-DNRVLVVPSEN    (SEQ ID NO:127)
9          -DIDHIIPQA-FIKD------NSI-DNRVLTSSKEN    (SEQ ID NO:128)
12         -DIDHIIPQA-FLKD------NSI-DNKVLVSSASN    (SEQ ID NO:129)
16         -DIDHIIPQA-YTKD------NSL-DNRVLVSNITN    (SEQ ID NO:130)
11         -DIDHIVPQS-FITD------NSI-DNLVLTSSAGN    (SEQ ID NO:131)
10         -DVDHIVPQS-FLKD------DSI-DNKVLTRSDKN    (SEQ ID NO:132)
14         -NIDHIYPQS-MVKD------DSL-DNKVLVQSEIN    (SEQ ID NO:133)
18         -DIDHILPQS-LIKD------DSL-DNRVLVNATIN    (SEQ ID NO:134)
19         -DIDHILPQS-FIKD------DSL-ENRVLVKKAVN    (SEQ ID NO:135)
13         -EVDHIFPRS-FIKD------DSI-DNKVLVIKKMN    (SEQ ID NO:136)
15         -EVDHIIPRS-YIKD------DSF-ENKVLVYREEN    (SEQ ID NO:137)
17         -DIDHIIPQA-VTQN------DSI-DNRVLVARAEN    (SEQ ID NO:138)
22         -EIDHIIPYS-ISFD------DSS-SNKLLVLAESN    (SEQ ID NO:139)
24         -EIDHIIPYS-LCFD------DSS-ANKVLVHKQSN    (SEQ ID NO:140)
32         -DIDHIIPYS-RSMD------DSY-SNKVLVLSGEN    (SEQ ID NO:141)
63         -DIDHIIPYS-KSMD------DSF-NNKVLCLAEEN    (SEQ ID NO:142)
59         -EIDHIYPYS-RSFD------DSY-MNKVLVFTKQN    (SEQ ID NO:143)
65         -QIDHIYPYS-RSMD------DSY-MNKVLVLTDEN    (SEQ ID NO:144)
64         -EIDHIIPFS-RSFD------DSL-SNKILVLGSEN    (SEQ ID NO:145)
68         -EIDHALPFS-RTWD------DSF-NNKVLVLGSEN    (SEQ ID NO:146)
69         -EIDHALPFS-RTWD------DSF-NNKVLVLASEN    (SEQ ID NO:147)
28         -EIDHIIPIS-ISLD------DSI-NNKVLVLSKAN    (SEQ ID NO:148)
30         -EVDHIIPIS-ISLD------DSI-TNKVLVTHREN    (SEQ ID NO:149)
62         -QVDHALPYS-RSYD------DSK-NNKVLVLTHEN    (SEQ ID NO:150)
27         -EVDHILPLS-ITFD------DSL-ANKVLVYATAN    (SEQ ID NO:151)
26         -EIDHIIPRS-ISFD------DAR-SNKVLVYRSEN    (SEQ ID NO:152)
```

图5A

```
29    -EVDHIIPRS-VSFD------NSY-HNKVLVKQSEN    (SEQ ID NO:153)
67    -DIDHILPYS-ITFD------DSF-RNKVLVTSQEN    (SEQ ID NO:154)
58    -EIDHILPRS-RSAD------DSF-ANKVLCLARAN    (SEQ ID NO:155)
61    -EIEHLLPFS-LTLD------DSM-ANKTVCFRQAN    (SEQ ID NO:156)
55    -DIDHILPFS-VSLD------DSA-ANKVVCLREAN    (SEQ ID NO:157)
57    -DIDHLIPFS-ISWD------DSA-ANKVVCMRYAN    (SEQ ID NO:158)
56    -DIDHILPVA-MTLD------DSP-ANKIICMRYAN    (SEQ ID NO:159)
54    -DVDHILPYS-RTLD------DSF-PNRTLCLREAN    (SEQ ID NO:160)
52    -EIEHILPFS-RTLD------DSL-NNRTVAMRRAN    (SEQ ID NO:161)
31    -EVDHIIPYS-ISWD------DSY-TNKVLTSAKCN    (SEQ ID NO:162)
45    -QVDHILPWS-RFGD------DSY-LNKTLCTARSN    (SEQ ID NO:163)
53    -QVDHILPFS-KTLD------DSF-ANKVLAQHDAN    (SEQ ID NO:164)
60    -QIDHAFPLS-RSLD------DSQ-SNKVLCLTSSN    (SEQ ID NO:165)
21    -DIDHIVPRS-ISFD------DSF-SNLVIVNKLDN    (SEQ ID NO:166)
23    -EIEHIIPYS-MSYD------NSQ-ANKILTEKAEN    (SEQ ID NO:167)
25    -EIDHVIPYS-KSAD------DSW-FNKLLVKKSTN    (SEQ ID NO:168)
49    -EMDHILPYS-RSLD------NGW-HNRVLVHGKDN    (SEQ ID NO:169)
33    -EVDHIVPYS-LILD------NTI-NNKALVYAEEN    (SEQ ID NO:170)
42    -EIEHVIPQS-LYFD------DSF-SNKVICEAEVN    (SEQ ID NO:171)
43    -DIEHIIPQA-RLFD------DSF-SNKTLEARSVN    (SEQ ID NO:172)
44    -EIEHIVPKA-RVFD------DSF-SNKTLTPHRIN    (SEQ ID NO:173)
20    -DKDHIIPQS-MKKD------DSIINNLVLVNKNAN    (SEQ ID NO:174)
66    -EVEHIWPRS-RSFD------NSP-RNKTLCRKDVN    (SEQ ID NO:175)
61    -IVNHIIPYN-RSFD------DTY-HNRVLTLTETK    (SEQ ID NO:176)
46    -DMEHTIPKS-ISFD------NSD-QNLTLCESYYN    (SEQ ID NO:177)
47    -DIEHTIPRS-AGGD------STK-MNLTLCSSRFN    (SEQ ID NO:178)
48    -DIEHTIPRS-ISQD------NSQ-MNKTLCSLKFN    (SEQ ID NO:179)
50    -DIDHVIPLA-RGGR------DSL-DNMVLCQSDAN    (SEQ ID NO:180)
39    -DIEHLFPIA-ESED------NGR-NNLVISHSACN    (SEQ ID NO:181)
41    -DVDHIFPRD-DTAD------NSY-GNKVVAHRQCN    (SEQ ID NO:182)
40    -DIEHIVPQS-LGGL------STD-YNTIVTLKSVN    (SEQ ID NO:183)
35    -ELDHIVPRT-DGGS------NRH-ENLAITCGACN    (SEQ ID NO:184)
36    -EMDHIVPRKGVGST------NTR-TNFAAVCABCN    (SEQ ID NO:185)
37    -EMDHIVPRKGVGST------NTR-VNLAAACAACN    (SEQ ID NO:186)
38    -EMDHIVPRAGQGST------NTR-ENLVAVCHRCN    (SEQ ID NO:187)
70    -EIDHILPRS-LIKDARGIVFNAE-PNLIYASSRGN    (SEQ ID NO:188)
71    -EIDHIIPRS-LTGRTKKTVFNSE-ANLIYCSSKGN    (SEQ ID NO:189)
73    -EIDHIIPRS-LTLKKSESIYNSE-VNLIFVSAQGN    (SEQ ID NO:190)
```

图5B

```
72    -EIDHIYPRS-LSKKHPGVIFNSE-VNLIYCSSQGN    (SEQ ID NO:191)
74    -EIDHILPRS-HTLKIYGTVFNPE-GNLIYVHQKCN    (SEQ ID NO:192)
75    -ELDHIIPRS-HKKY---GTLNDE-ANLICVTRGDN    (SEQ ID NO:193)
34    -ELEHIVPHS-FRQS------NAL-SSLVLTWPGVN    (SEQ ID NO:194)
         :*  *              .    .        :
```

图5C

披露于Chylinski等人HNH样结构域的比对（排除序列异常值）
(通过MAFFT的CLUSTAL格式比对(v7.058b))

```
1          YDIDHIYPRS-LTKDDS-FDNLVLCERTAN     (SEQ ID NO:196)
2          -DIDHIYPRSKVIKDDS-FDNLVLVLKNEN     (SEQ ID NO:197)
3          -DEDHIYPQS-KIKDDS-IDNLVLVNKTYN     (SEQ ID NO:198)
4          -DIDHIYPRS-KIKDDS-ITNRVLVEKDIN     (SEQ ID NO:195)
6          -DIDHIYPQS-KIKDDS-ISNRVLVCSSCN     (SEQ ID NO:124)
5          -DIDHIYPQS-KTMDDS-LNNRVLVKKNYN     (SEQ ID NO:125)
7          -DQDHIYPKS-KIYDDS-LENRVLVKKNLN     (SEQ ID NO:126)
8          -QIDHIVPQS-LVKDDS-FDNRVLVVPSEN     (SEQ ID NO:127)
9          -DIDHIIPQA-FIKDNS-IDNRVLTSSKEN     (SEQ ID NO:128)
12         -DIDHIIPQA-FLKDNS-IDNKVLVSSASN     (SEQ ID NO:129)
16         -DIDHIIPQA-YTKDNS-LDNRVLVSNITN     (SEQ ID NO:130)
11         -DIDHIVPQS-FITDNS-IDNLVLTSSAGN     (SEQ ID NO:131)
10         -DVDHIVPQS-FLKDDS-IDNKVLTRSDKN     (SEQ ID NO:132)
14         -NIDHIYPQS-MVKDDS-LDNKVLVQSEIN     (SEQ ID NO:133)
18         -DIDHILPQS-LIKDDS-LDNRVLVNATIN     (SEQ ID NO:134)
19         -DIDHILPQS-FIKDDS-LENRVLVKKAVN     (SEQ ID NO:135)
13         -EVDHIFPRS-FIKDDS-IDNKVLVIKKMN     (SEQ ID NO:136)
15         -EVDHIIPRS-YIKDDS-FENKVLVYREEN     (SEQ ID NO:137)
17         -DIDHIIPQA-VTQNDS-IDNRVLVARAEN     (SEQ ID NO:138)
21         -DIDHIVPRS-ISFDDS-FSNLVIVNKLDN     (SEQ ID NO:166)
22         -EIDHIIPYS-ISFDDS-SSNKLLVLAESN     (SEQ ID NO:139)
24         -EIDHIIPYS-LCFDDS-SANKVLVHKQSN     (SEQ ID NO:140)
28         -EIDHIIPIS-ISLDDS-INNKVLVLSKAN     (SEQ ID NO:148)
30         -EVDHIIPIS-ISLDDS-ITNKVLVTHREN     (SEQ ID NO:149)
27         -EVDHILPLS-ITFDDS-LANKVLVYATAN     (SEQ ID NO:151)
26         -EIDHIIPRS-ISFDDA-RSNKVLVYRSEN     (SEQ ID NO:152)
29         -EVDHIIPRS-VSFDNS-YHNKVLVKQSEN     (SEQ ID NO:153)
31         -EVDHIIPYS-ISWDDS-YTNKVLTSAKCN     (SEQ ID NO:162)
32         -DIDHIIPYS-RSMDDS-YSNKVLVLSGEN     (SEQ ID NO:141)
23         -EIEHIIPYS-MSYDNS-QANKILTEKAEN     (SEQ ID NO:167)
33         -EVDHIVPYS-LILDNT-INNKALVYAEEN     (SEQ ID NO:170)
25         -EIDHVIPYS-KSADDS-WFNKLLVKKSTN     (SEQ ID NO:168)
49         -EMDHILPYS-RSLDNG-WHNRVLVHGKDN     (SEQ ID NO:169)
42         -EIEHVIPQS-LYFDDS-FSNKVICEAEVN     (SEQ ID NO:171)
43         -DIEHIIPQA-RLFDDS-FSNKTLEARSVN     (SEQ ID NO:172)
```

图6A

```
44    -EIEHIVPKA-RVFDDS-FSNKTLTFHRIN     (SEQ ID NO:173)
20    -DKDHIIPQS-MKKDDSIINNLVLVNKNAN     (SEQ ID NO:174)
45    -QVDHILPWS-RFGDDS-YLNKTLCTARSN     (SEQ ID NO:163)
50    -DIDHVIPLA-RGGRDS-LDNMVLCQSDAN     (SEQ ID NO:180)
46    -DMEHTIPKS-ISFDNS-DQNLTLCESYYN     (SEQ ID NO:177)
47    -DIEHTIPRS-AGGDST-KMNLTLCSSRFN     (SEQ ID NO:178)
48    -DIEHTIPRS-ISQDNS-QMNKTLCSLKFN     (SEQ ID NO:179)
39    -DIEHLFPIA-ESEDNG-RNNLVISHSACN     (SEQ ID NO:181)
41    -DVDHIFPRD-DTADNS-YGNKVVAHRQCN     (SEQ ID NO:182)
40    -DIEHIVPQS-LGGLST-DYNTIVTLKSVN     (SEQ ID NO:183)
35    -ELDHIVPRT-DGGSNR-HENLAITCGACN     (SEQ ID NO:184)
36    -EMDHIVPRKGVGSTNT-RTNFAAVCAECN     (SEQ ID NO:185)
37    -EMDHIVPRKGVGSTNT-RVNLAAACAACN     (SEQ ID NO:186)
38    -EMDHIVPRAGQGSTNT-RENLVAVCHRCN     (SEQ ID NO:187)
34    -ELEHIVPHS-FRQSNA-LSSLVLTWPGVN     (SEQ ID NO:194)
```

图6B

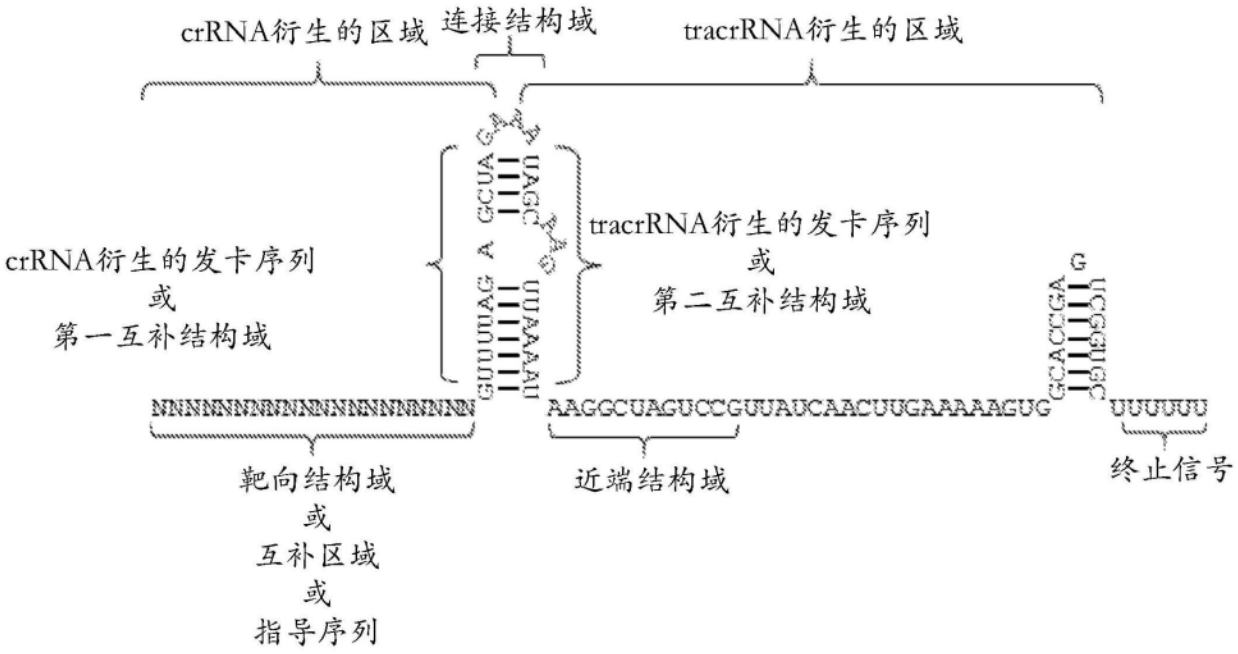

crRNA衍生的区域
连接结构域
tracrRNA衍生的区域
crRNA衍生的发卡序列
或
第一互补结构域
tracrRNA衍生的发卡序列
或
第二互补结构域
NNNNNNNNNNNNNNNNNNNN
AAGGCUAGUCCGUUAUCAACUUGAAAAAGUG
UUUUUU
靶向结构域
或
互补区域
或
指导序列
近端结构域
终止信号

图7

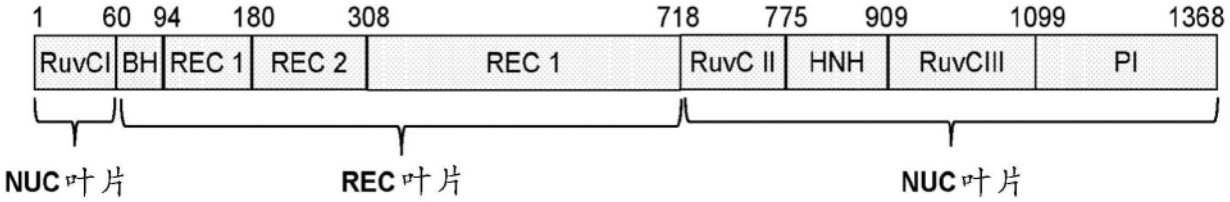

1
60
94
180
308
718
775
909
1099
1368
RuvCI
BH
REC 1
REC 2
REC 1
RuvC II
HNH
RuvCIII
PI
NUC叶片
REC叶片
NUC叶片

图8A

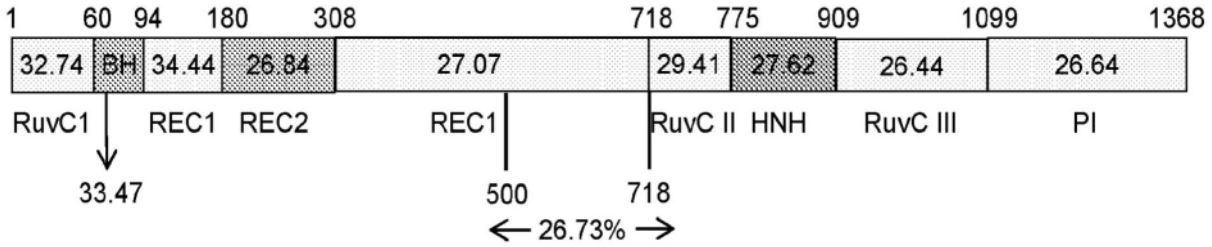

1
60
94
180
308
718
775
909
1099
1368
32.74
BH
34.44
26.84
27.07
29.41
27.62
26.44
26.64
RuvC1
REC1
REC2
REC1
RuvC II
HNH
RuvC III
PI
33.47
500
718
26.73%

图8B

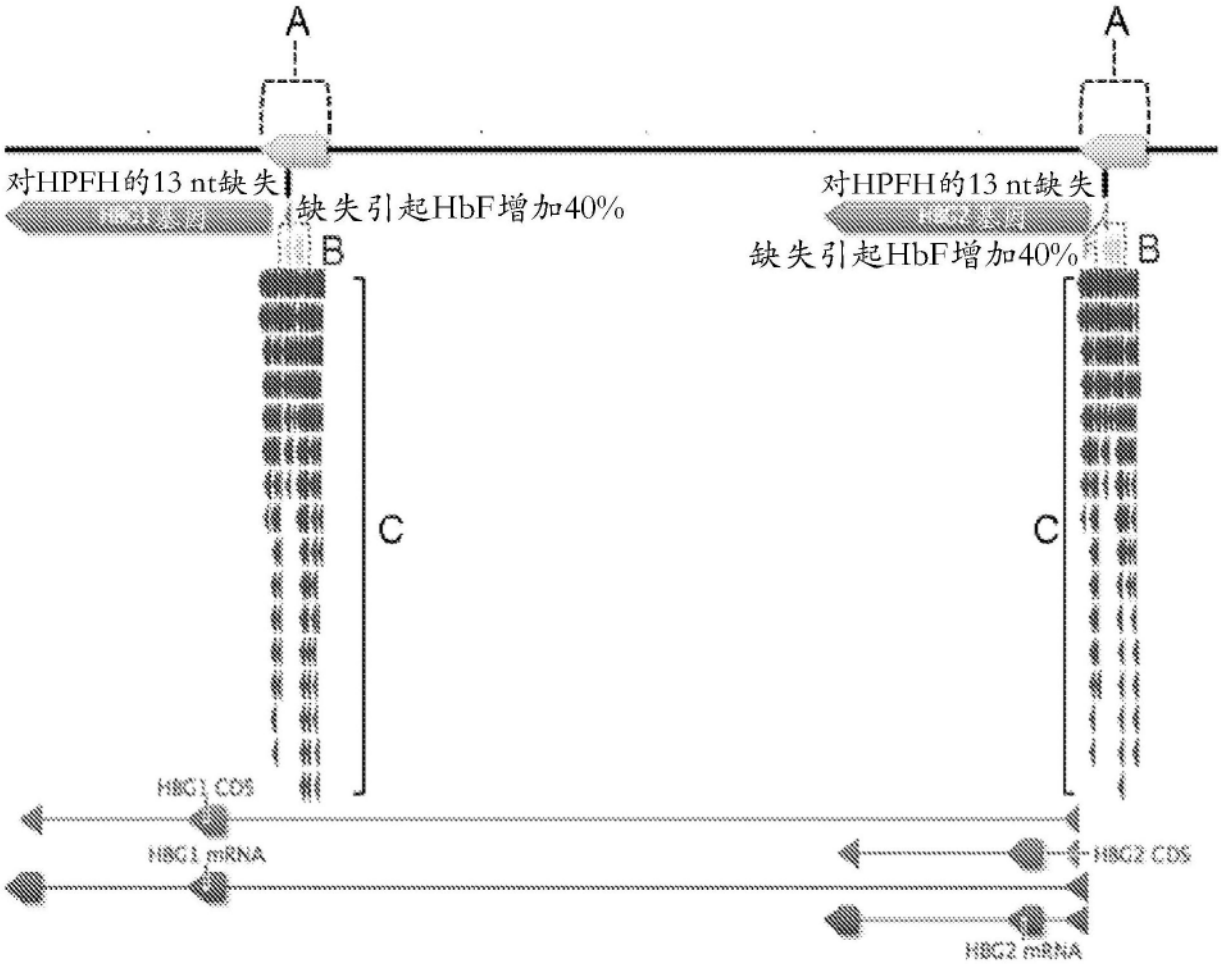

A
A
对HPFH的13 nt缺失
缺失引起HbF增加40%
B
对HPFH的13 nt缺失
缺失引起HbF增加40%
B
C
C
HBG1 CDS
HBG1 mRNA
HBG2 CDS
HBG2 mRNA

图9

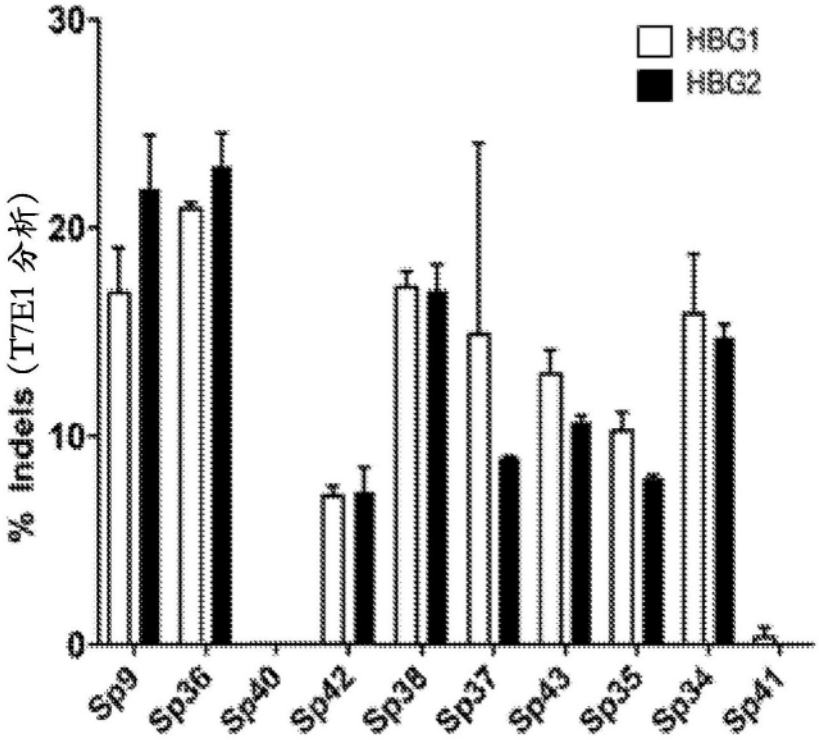


图10A

基因编辑：K562细胞中的HBG1基因座

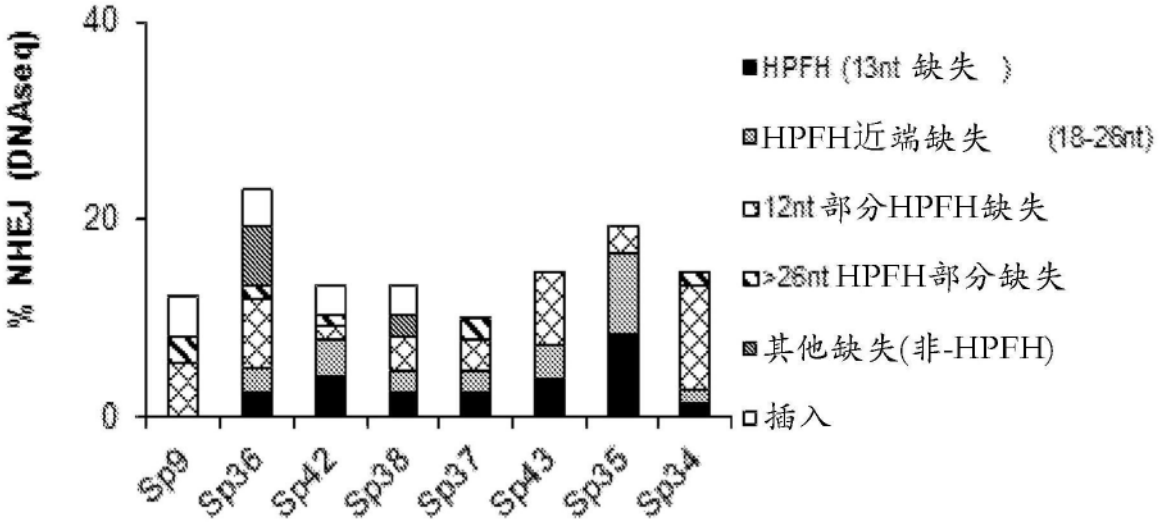


图10B

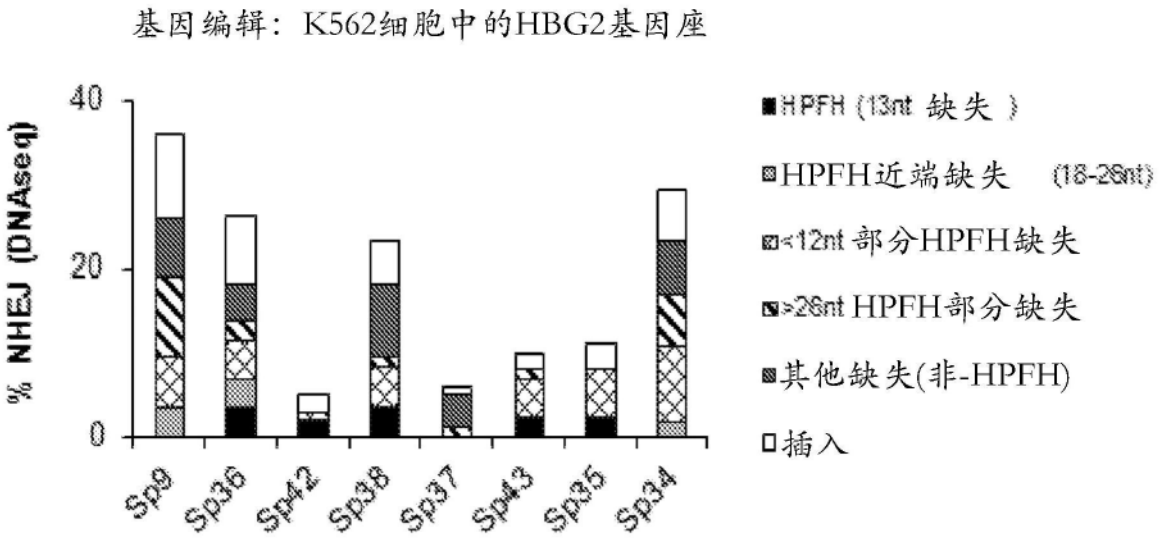

基因编辑：K562细胞中的HBG2基因座
% NHEJ (DNAseq)
40
20
0
Sp9
Sp36
Sp42
Sp38
Sp37
Sp43
Sp35
Sp34
HPFH (13nt 缺失)
HPFH近端缺失 (18-26nt)
<12nt 部分HPFH缺失
>26nt HPFH部分缺失
其他缺失(非-HPFH)
插入

图10C

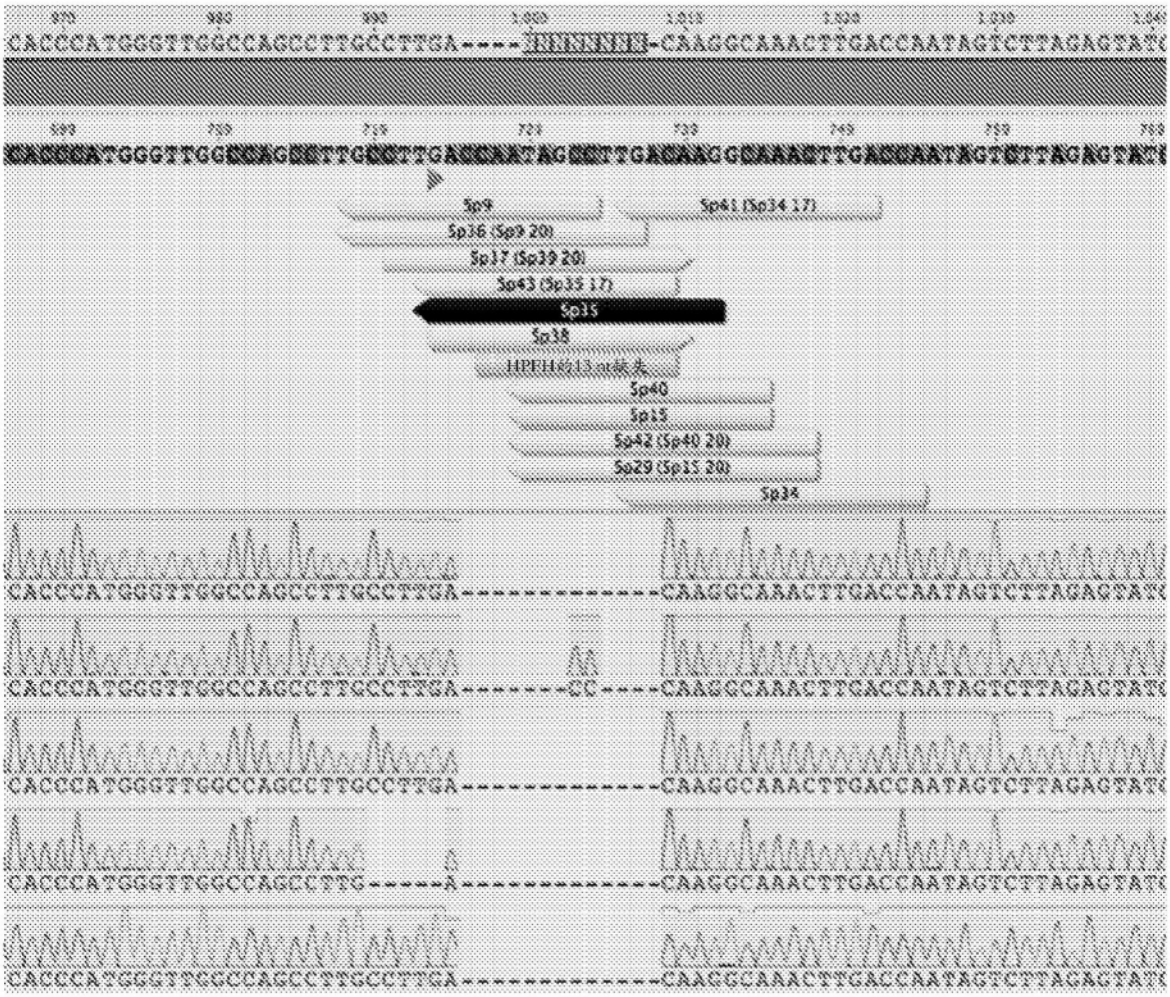

Sp9
Sp41 (Sp34 17)
Sp36 (Sp9 20)
Sp17 (Sp39 20)
Sp43 (Sp35 17)
Sp15
Sp38
HPFH的13 nt缺失
Sp40
Sp15
Sp42 (Sp40 20)
Sp29 (Sp15 20)
Sp34

图10D

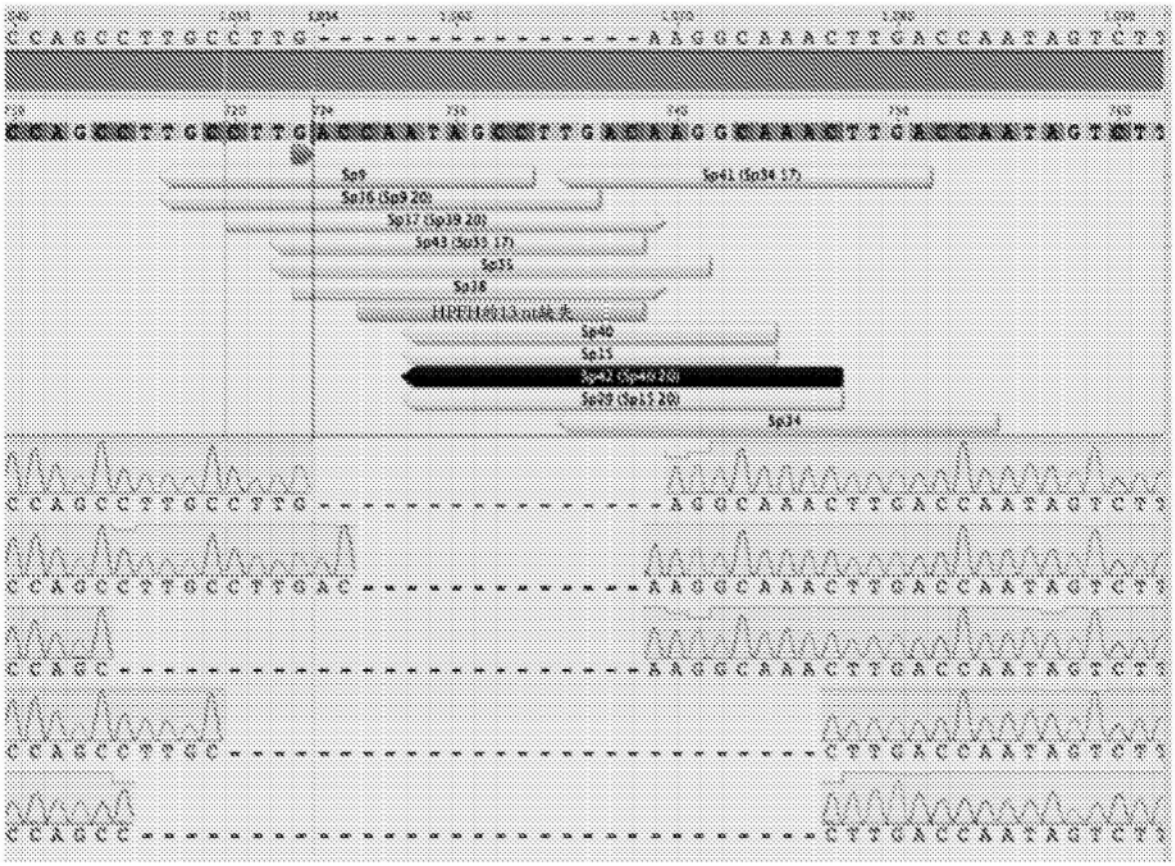

Sp9
Sp41 (Sp34 17)
Sp16 (Sp9 20)
Sp17 (Sp19 20)
Sp43 (Sp33 17)
Sp38
Sp40
Sp15
Sp34

图10E

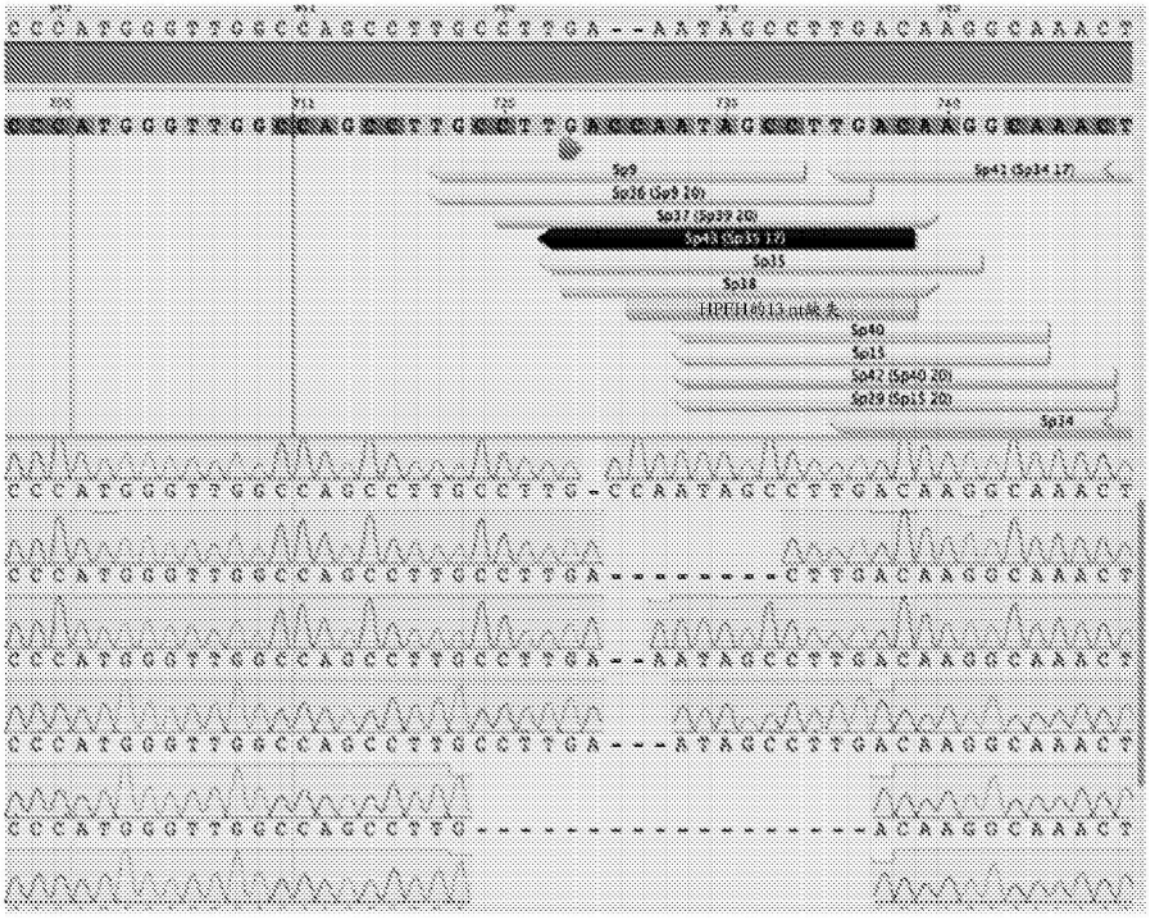

Sp9
Sp41 (Sp34 17)
Sp36 (Sp9 20)
Sp37 (Sp39 20)
Sp43 (Sp35 17)
Sp35
Sp38
HPFH的13 nt缺失
Sp40
Sp15
Sp42 (Sp40 20)
Sp29 (Sp15 20)
Sp34

图10F

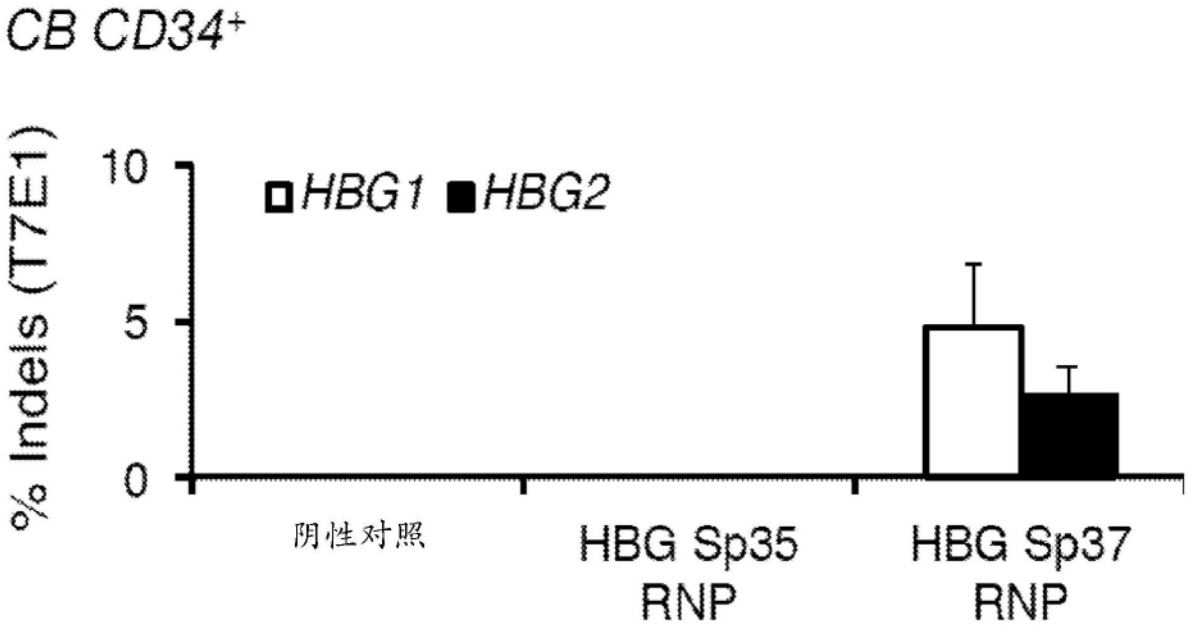

CB CD34+
% Indels (T7E1)
10
5
0
HBG1
HBG2
阴性对照
HBG Sp35 RNP
HBG Sp37 RNP

图11A

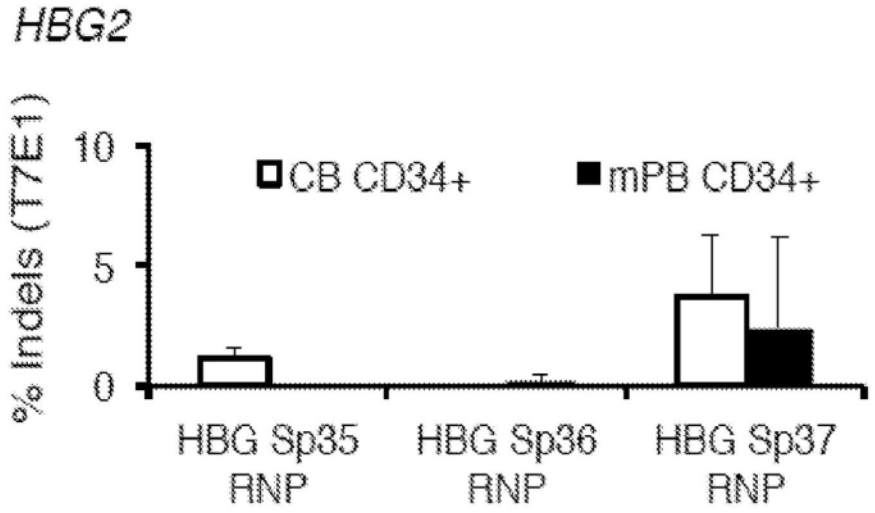

HBG2
% Indels (T7E1)
10
5
0
CB CD34+
mPB CD34+
HBG Sp35 RNP
HBG Sp36 RNP
HBG Sp37 RNP

图11B

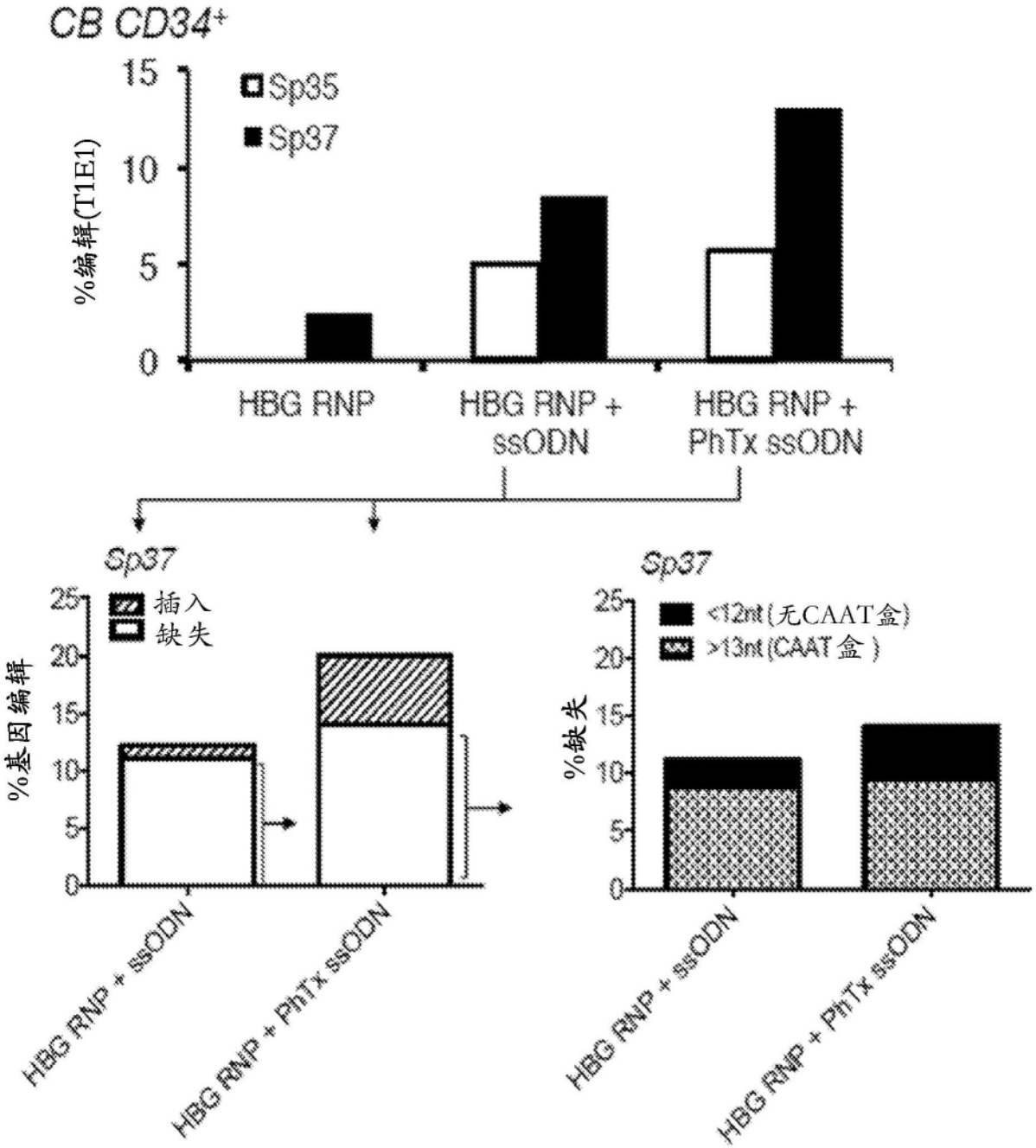

CB CD34+
Sp35
Sp37
%编辑(T1E1)
15
10
5
0
HBG RNP
HBG RNP + ssODN
HBG RNP + PhTx ssODN
Sp37
插入
缺失
%基因编辑
25
20
15
10
5
0
HBG RNP + ssODN
HBG RNP + PhTx ssODN
Sp37
<12nt(无CAAT盒)
>13nt(CAAT盒)
%缺失
25
20
15
10
5
0
HBG RNP + ssODN
HBG RNP + PhTx ssODN

图11C

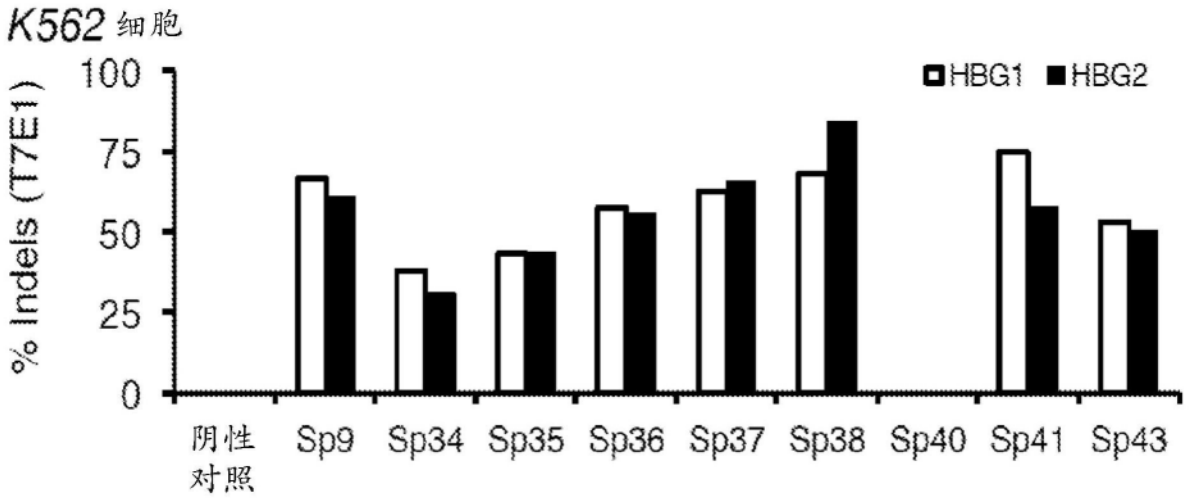


图12A

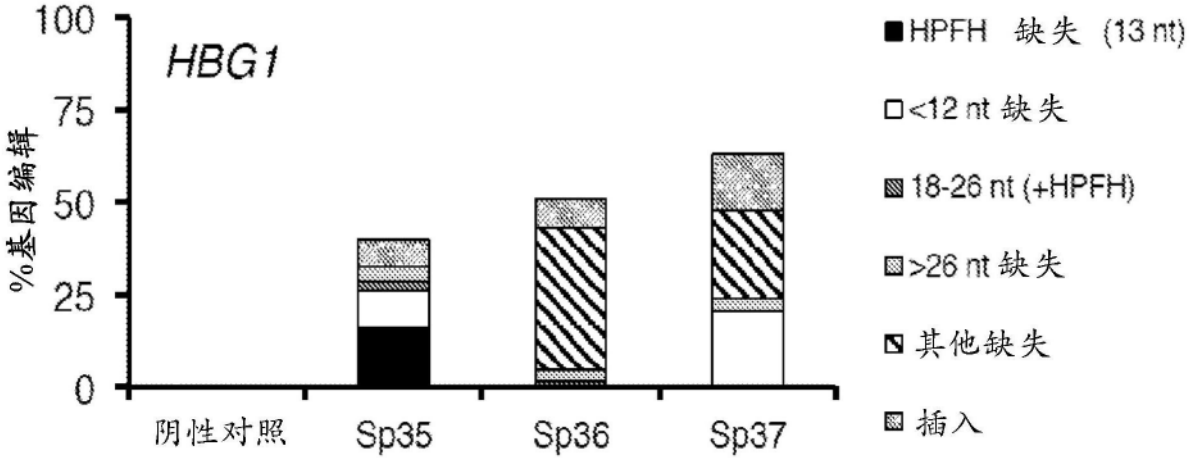


图12B

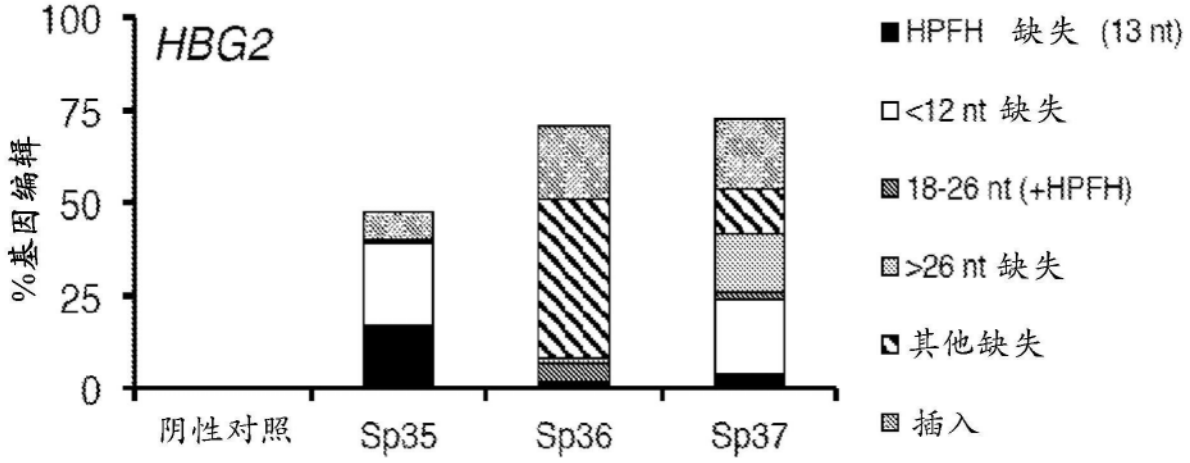


图12C

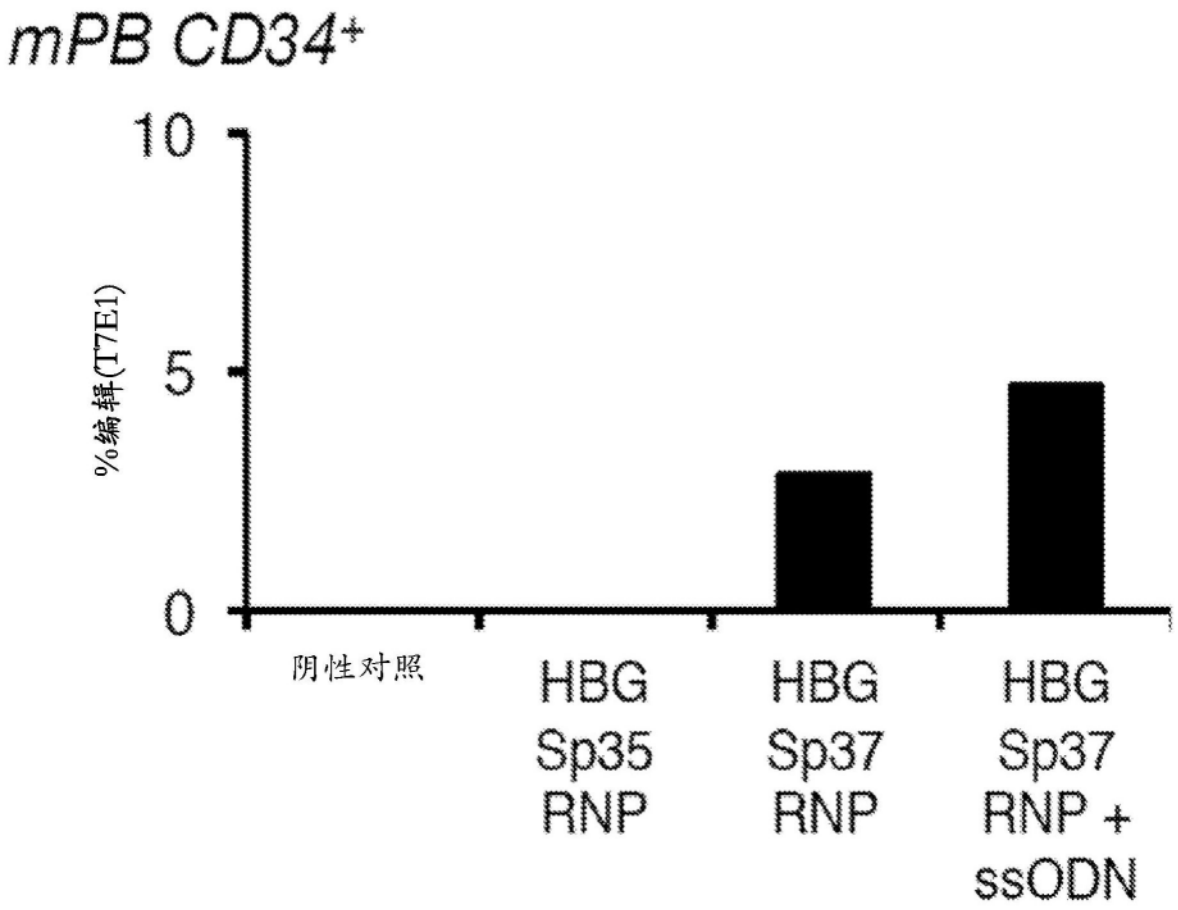

mPB CD34+
10
5
0
%编辑(T7E1)
阴性对照
HBG Sp35 RNP
HBG Sp37 RNP
HBG Sp37 RNP + ssODN

图13A

mPB CD34+ 幼红细胞

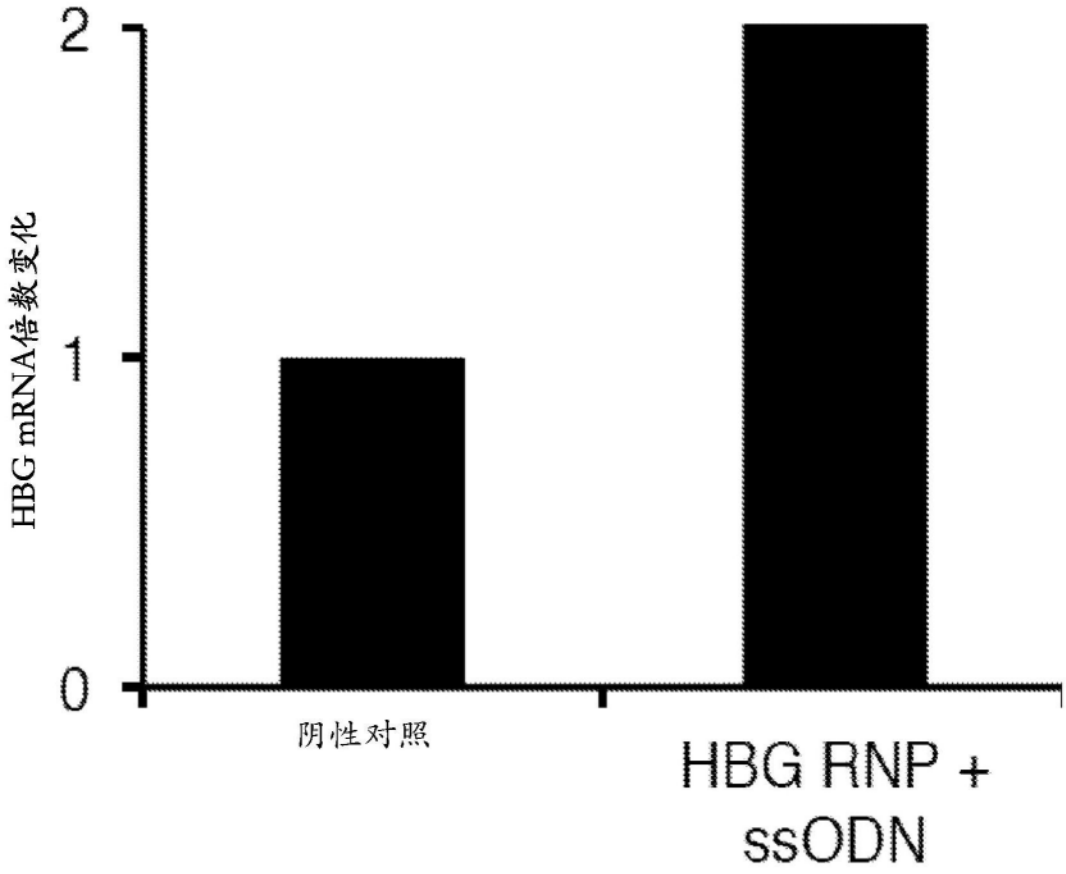


图13B

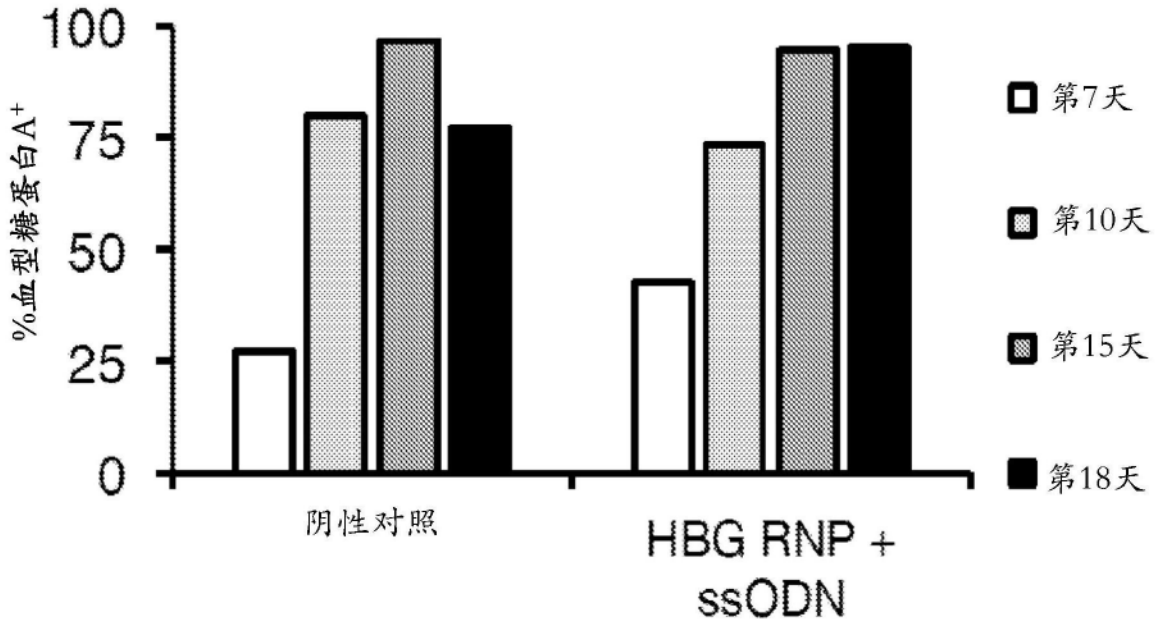

100
75
50
25
0
%血型糖蛋白A+
第7天
第10天
第15天
第18天
阴性对照
HBG RNP +
ssODN

图14A

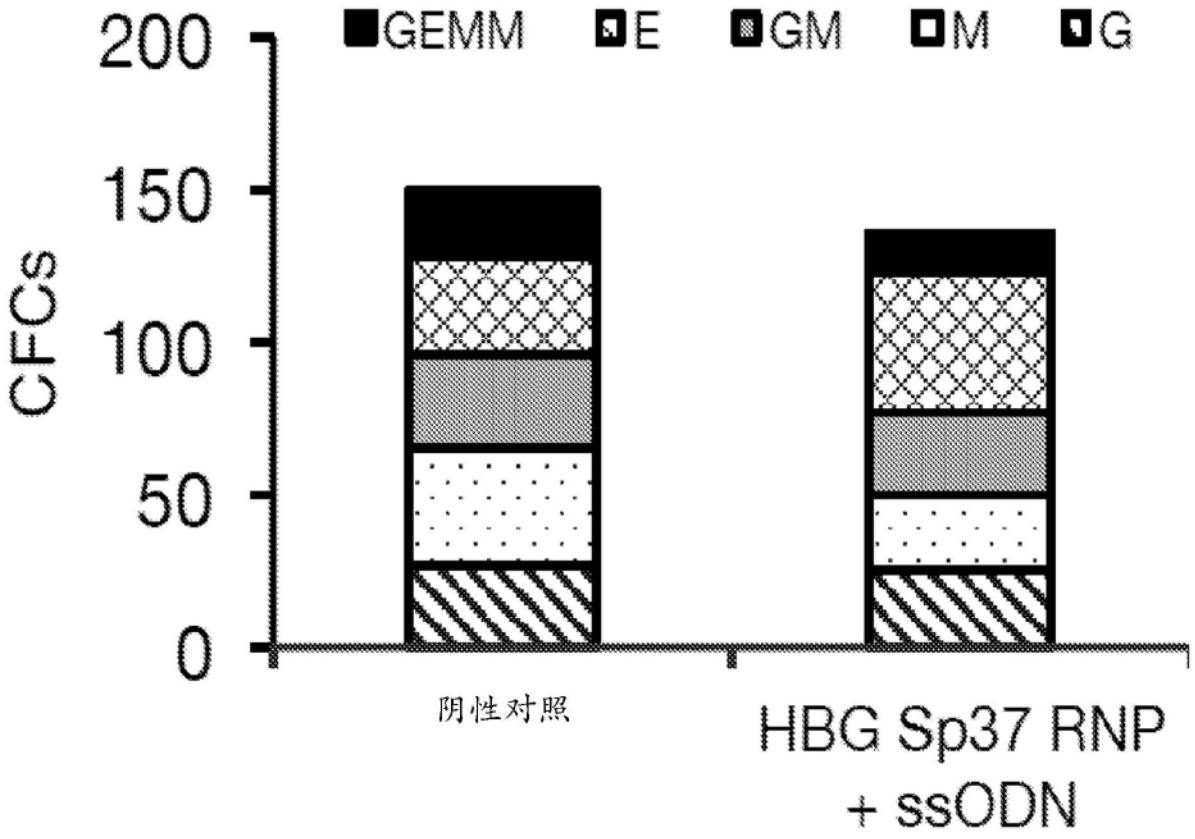

GEMM
E
GM
M
G
200
150
100
50
0
CFCs
阴性对照
HBG Sp37 RNP
+ ssODN

图14B